Existence and Enumeration of Spanning Structures in Sparse Graphs and Hypergraphs

Vom Promotionsausschuss der
Technischen Universität Hamburg-Harburg
zur Erlangung des akademischen Grades
Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von
Julia Ehrenmüller

aus
Linz, Österreich

2016

1. Gutachter: Prof. Dr. Anusch Taraz
2. Gutachter: Prof. Tibor Szabó, Ph.D.

Tag der mündlichen Prüfung: 30.06.2016

Abstract

This thesis examines the robustness of sparse graphs and hypergraphs with respect to containing copies of given spanning subgraphs. In particular, we prove analogues of the bandwidth theorem for random and pseudorandom graphs, as well as a Dirac-type theorem for Hamilton Berge cycles in random r-uniform hypergraphs. Furthermore, we determine conditions for the existence of rainbow matchings in edge-coloured multigraphs and study the number of spanning trees in graphs chosen uniformly at random from subfamilies of series-parallel graphs.

Diese Dissertation beschäftigt sich mit der Robustheit von dünnen Graphen und Hypergraphen bezüglich des Auftretens gegebener aufspannender Subgraphen. Insbesondere werden Analoga des Bandweitentheorems für zufällige und pseudozufällige Graphen und ein sogenanntes Dirac-artiges Theorem für Berge-Hamiltonkreise in zufälligen r-uniformen Hypergraphen bewiesen. Zudem werden Bedingungen, die die Existenz von Regenbogenmatchings in kantengefärbten Multigraphen sichern, bestimmt und die Anzahl der Spannbäume in Graphen, die zufällig aus Unterfamilien von serien-parallelen Graphen gewählt werden, studiert.

Zusammenfassung

Ein zentrale Fragestellung in der extremalen Graphentheorie ermittelt, unter welchen Bedingungen bestimmte große oder sogar aufspannende Substrukturen in Graphen erzwungen werden. Für Graphen, die über eine solche Substruktur verfügen, stellt sich dann unmittelbar die Frage, wie robust sich diese Eigenschaft bei ihnen zeigt.

Diese Robustheit lässt sich auf mehrere Arten messen. Eine Möglichkeit ist zu bestimmen, wie viel Prozent der inzidenten Kanten ein fiktiver Gegenspieler an jedem Knoten des Graphen mindestens löschen muss, damit der entstehende Graph die gewünschte Substruktur nicht mehr aufweist. Dieses Konzept wird häufig als *lokale Resilienz* bezeichnet. Alternativ lässt sich die Robustheit eines Graphen G bezüglich des Auftretens einer Kopie eines gegebenen Graphen H ermitteln, indem man die Subgraphen von G zählt, die isomorph zu H sind. In der vorliegenden Dissertation werden solche Probleme sowohl in dünnen Graphen als auch in dünnen Hypergraphen studiert.

Zunächst beschäftigen wir uns mit der lokalen Resilienz von zufälligen Hypergraphen sowie von zufälligen und pseudozufälligen Graphen. Insbesondere beweisen wir für diese Strukturen Analoga zu den folgenden zwei Theoremen. Der bekannte Satz von Dirac gibt Auskunft über die lokale Resilienz des vollständigen Graphen bezüglich eines Hamiltonkreises. Ein allgemeineres Resultat, das Bandweitentheorem von Böttcher, Schacht und Taraz, ermittelt diese bezüglich des Auftretens aller aufspannenden Graphen mit beschränktem Minimalgrad und sublinearer Bandweite.

Das erste der beiden Theoreme erweitern wir, indem wir den vollständigen Graphen durch einen zufälligen r-uniformen Hypergraphen ersetzen und die lokale Resilienz bezüglich eines Berge-Hamiltonkreises bestimmen. Der Beweis basiert auf der von Rödl, Ruciński und Szemerédi entwickelten Absorbtionsmethode. Dies ist das erste bislang bekannte Ergebnis über die lokale Resilienz von zufälligen Hypergraphen. Vor diesem Hintergrund diskutieren wir auch den Zusammenhang zwischen lokalen-Resilienz-Resultaten und kombinatorischen Spielen.

Für das zweite oben genannte Resultat, das Bandweitentheorem, zeigen wir analoge Sätze für dünne zufällige und pseudozufällige Graphen. Unsere Beweise beruhen auf der Regularitätsmethode und verwenden verschiedene Blow-up-Lemmata für dünne Graphen, die von Allen, Böttcher, Hàn, Kohayakawa und Person bewiesen wurden.

Ausgehend von einer Vermutung von Ryser, Brualdi und Stein über eine Bedingung, die das Auftreten von perfekten oder fast perfekten Matchings in 3-partiten Hypergraphen garantieren soll, und einer allgemeineren Vermutung von Aharoni und Berger untersuchen wir kantengefärbte Multigraphen. Wir zeigen eine asymptotisch bestmögliche Bedingung an die Größen der Farbklassen, die jeweils Vereinigungen von Cliquen induzieren, sodass der Multigraph ein perfektes Matching besitzt, dessen Kanten mit paarweise verschiedenen Farben gefärbt sind. Dieses Resultat ist eine Annäherung an die Vermutung von Aharoni und Berger und bestätigt asymptotisch eine bisher offene, von Grinblat gestellte Frage aus der Algebra.

Im letzten der Teil der Dissertation beschäftigen wir uns mit der Anzahl an Spannbäumen in Graphen bestimmter Familien. Schranken und Schätzungen für diese Zahl wurden unter anderem für (zufällig gezogene) planare Karten und Graphen gegebener Gradsequenzen untersucht. Mit Methoden der analytischen Kombinatorik, insbesondere der symbolischen Methode und der Singularitätsanalyse, ermitteln wir die asymptotische erwartete Anzahl an Spannbäumen in einem Graphen, der zufällig aus allen zusammenhängenden serien-parallen Graphen gezogen wurde. Ferner erhalten wir ähnliche Resultate für Unterfamilien von serien-parallelen Graphen, wie etwa für all jene mit maximal vielen oder besonders wenigen Kanten.

Contents

1 Introduction

1.1 Overview of the thesis

A central problem in extremal graph theory is to determine conditions that force a graph to contain a given large or spanning substructure. Knowing that a graph contains such a substructure, it is natural to ask how robust the graph is with respect to that containment. This vaguely formulated question can be approached from various different directions depending on how the robustness of a graph is measured.

A first possibility is to study how many edges a fictitious adversary has to delete from a graph G to destroy the property of containing a copy of a given large graph H. However, vertices of G can become isolated while only a few edges are removed. Therefore, it makes sense to impose the restriction that the adversary may delete only a certain fraction of the incident edges at every vertex. The minimum fraction that is necessary to obtain a graph that does no longer contain a copy of H is known under the name *local resilience*.

Instead of evaluating the robustness of a graph with regard to adversarial edge deletion, one can also examine how rich a graph G is concerning the number of copies of H that G contains. In other words, a possible choice of measurement is counting the subgraphs of G that are isomorphic to H. The higher this number is, the more likely it is that removing a few edges randomly from G does not destroy the property of containing a copy of H.

A classic theorem by Dirac [63] implies that the local resilience of K_n with respect to the containment of a Hamilton cycle is $1/2 + o(1)$. Asymptotically this remains true if K_n is replaced by a much sparser graph. For instance, it was shown by Lee and Sudakov [124] that the random graph $G(n,p)$, which is defined on n vertices with each pair of vertices forming an edge independently with probability p, satisfies the following with high probability if $p = \Omega(\log n/n)$: whichever edges an adversary removes from $G(n,p)$ respecting that $1/2+o(1)$ of the incident edges remain at every vertex, the resulting graph is still Hamiltonian.

More generally, rather than requiring that a graph G contains a copy of one single given graph, one can also ask whether and how strongly G contains any or even all graphs of a given graph family. The *bandwidth theorem* by Böttcher, Schacht, and Taraz [41] provides a sufficient minimum degree condition for graphs to contain all maximum degree bounded subgraphs with sublinear bandwidth, which is asymptotically tight. In particular, this implies a local resilience result for K_n with respect to such containment.

In this thesis we extend the above mentioned results in the following ways. First, in

Chapter 3 we prove analogues of the bandwidth theorem with K_n being replaced by a sparse random or pseudorandom graph, as well as a variant for the containment of degenerate graphs in sparse random graphs. All of these results are universal, in the sense that the simultaneous containment of all members of a given graph family is implied, and they are asymptotically optimal with respect to the minimum degree condition. The proofs of these results are based on the regularity method and use powerful blow-up lemmas for sparse graphs, proved by Allen, Böttcher, Hàn, Kohayakawa, and Person [9].

Second, in Chapter 4 we prove a local resilience result for sparse random hypergraphs with respect to Berge Hamiltonicity. The random hypergraph model that we use is $H^{(r)}(n,p)$, which is a natural extension of $G(n,p)$ to hypergraphs, i.e. $H^{(r)}(n,p)$ is defined on n vertices with each r-tuple forming a hyperedge independently from each other with probability p. Our result is the first known local resilience result for sparse random hypergraphs. It is asymptotically best possible in terms of the local resilience and the bound on the edge probability is optimal up to possibly a polylogarithmic factor. The proof is based on the absorbing method developed by Rödl, Ruciński, and Szemerédi [144]. We also discuss minimum vertex degree conditions for r-uniform hypergraphs to contain weak Hamilton cycles and Hamilton Berge cycles, as well as connections between local resilience results of hypergraphs and positional games. Furthermore, we investigate monotone and strict Avoider-Enforcer games played on the edge set of a complete 3-uniform hypergraph and prove bounds on the biases for which Avoider can keep his hypergraph (almost) Berge-acyclic.

Another substructure that we are particularly interested in are hypergraph matchings. A famous conjecture suggests that every balanced 3-partite 3-uniform hypergraph H on $3n$ vertices, where each pair of vertices from different partition classes lies in exactly one hyperedge, contains a perfect matching if n is odd, and a matching of size $(n-1)$ if n is even. This conjecture was originally formulated in terms of Latin squares by Ryser [146] (for odd n) and by Brualdi [46] and Stein [148] (for all n with a weaker statement for odd n).

This problem has also been studied in the setting of edge-coloured graphs as there is a one-to-one correspondence between hypergraphs with the property just described, and complete bipartite graphs $K_{n,n}$ whose edges are coloured with n colours such that adjacent edges receive different colours. A matching in the hypergraph corresponds then to a so-called rainbow matching in the edge-coloured graph and vice versa. Aharoni and Berger [2] posed the following generalisation of the conjecture of Ryser, Brualdi, and Stein: every multigraph whose edges are coloured with n colours, where each colour class induces a matching of size $n+1$, contains a rainbow matching of size n.

As a result towards this conjecture, we prove in Chapter 5 that if each of the n colour classes induces a matching of size $\big(3/2+o(1)\big)n$, then the edge-coloured multigraph contains a rainbow matching of size n. In fact we show a stronger theorem; we consider edge-coloured multigraphs where each of the colour classes induces a disjoint union of cliques and we prove an asymptotically tight bound on the sizes of the colour classes that guarantees the existence of a rainbow matching that uses every colour. The result also affirms asymptotically an algebraic question by Grinblat [85] on sets not belonging to algebras.

As already mentioned, another possibility to measure the robustness of a graph G with respect to the containment of graphs from a given family $\mathcal{F}$ is to count the subgraphs of G that are isomorphic to any graph from $\mathcal{F}$. In fact, graph enumeration is an extensively studied field of graph theory and dates back to the mid-19th-century. One of the earliest, classic results in this area is attributed to Cayley [48]. It states that the number of spanning trees of the complete graph on n vertices is n^{n-2}. A generalisation of Cayley's formula is Kirchhoff's matrix tree theorem [104], which provides the number of spanning trees in any fixed graph G as the determinant of a matrix that is associated with G.

Restricting to a specific class of graphs, it is interesting to know how many spanning trees a graph from that class roughly contains, without having to compute this number for each such graph. For instance, one can establish lower and upper bounds on these numbers or estimate how many spanning trees one would expect in a graph chosen uniformly at random from all graphs of the given class. Problems of this flavour have been studied for regular graphs, graphs with given degree sequences, and rooted planar maps [13, 112, 126, 127, 132].

In Chapter 6 we address this problem for different subclasses of series-parallel graphs. In particular we prove a precise asymptotic estimate for the number of spanning trees in a graph chosen uniformly at random from all connected series-parallel graphs on a given number of vertices. We obtain analogous results for random edge-maximal series-parallel graphs, which are called 2-trees, and for random connected series-parallel graphs with fixed excess, which means that their number of edges and their number of vertices differ by a constant. Furthermore, we analyse the growth constant of the number of spanning trees in 2-connected series-parallel graphs chosen uniformly at random as a function of their edge densities. Our proofs are based on analytic combinatorics, in particular on the symbolic method, generating functions, and singularity analysis.

Before stating and discussing the results of this thesis more precisely in Section 1.3, we summarise in Section 1.2 previous relevant results in the areas that we are concerned with.

1.2 Historical background

This thesis addresses questions from extremal graph theory, probabilistic graph theory, and analytic combinatorics. In this section we collect previous results from these areas that are related to the questions that we treat in the subsequent chapters. In Subsection 1.2.1 we survey local resilience results of graphs, random graphs, pseudorandom graphs, and hypergraphs, and discuss universality results of random graphs and random hypergraphs. In the same subsection we also elaborate on a relation between local resilience and positional games. Subsection 1.2.2 is devoted to the study of hypergraph matchings, Latin squares, and rainbow matchings, all with regard to the above mentioned conjectures of Ryser, Brualdi, Stein, and Aharoni and Berger. Finally, in Subsection 1.2.3 we summarise results from enumerative combinatorics that deal with the enumeration of spanning trees and with properties of series-parallel graphs.

1.2.1 Local resilience

In Chapter 3 we prove local resilience results for random and pseudorandom graphs with respect to containing maximum degree bounded spanning subgraphs with sublinear bandwidth, and in Chapter 4 for random hypergraphs with respect to Berge Hamiltonicity. In view of these results, the purpose of this subsection is to summarise known local resilience results for graphs, random and pseudrandom graphs, and hypergraphs as well as thresholds of random graphs and random hypergraphs for the properties we are interested in. We also discuss a relation between local resilience of random graphs and biased Maker-Breaker games, which we extend to hypergraphs in Chapter 4.

Local resilience of graphs

As already mentioned, a typical question in extremal graph theory is which lower bound on the minimum degree suffices to guarantee that graphs respecting this condition contain a given spanning subgraph. The prototypical example is the following classic theorem by Dirac.

Theorem 1.1 (Dirac [63]). *Let G be a graph on $n \geq 3$ vertices. If $\delta(G) \geq \frac{n}{2}$ then G is Hamiltonian.*

The lower bound in Dirac's theorem is tight, as for each $m \in \mathbb{N}$ the complete bipartite graph $K_{m,m+1}$ and the graph consisting of two disjoint cliques each of size m do not contain a Hamilton cycle.

From an edge deletion perspective, Dirac's theorem says that an adversary may delete up to $(n/2-1)$ incident edges at every vertex of K_n without destroying Hamiltonicity. In other words, one needs to delete at least $n/2$ edges at a vertex of K_n such that the graph obtained in this way does not contain a Hamilton cycle. The latter perspective is the spirit of the notion of *local resilience*, which was first introduced by Sudakov and Vu [149] for a systematic study of minimum degree results. We use the following definition of local resilience, where we say that a graph property is *monotone increasing* if it is preserved under edge addition.

Definition 1.2 (Local resilience). *Let $\mathcal{P}$ be a monotone increasing graph property and let G be a graph with property $\mathcal{P}$. The local resilience of G with respect to $\mathcal{P}$ is the minimum number $\rho \in \mathbb{R}$ such that by deleting at each vertex $v \in V(G)$ at most $\rho \cdot \deg(v)$ edges one can obtain a graph without property $\mathcal{P}$.*

Using this terminology, Dirac's theorem implies that the local resilience of K_n with respect to Hamiltonicity is $\frac{1}{2} + o(1)$.

Let us mention at this point that there is a related concept called *global resilience*, which is defined as follows. Let G be a graph with a monotone increasing graph property $\mathcal{P}$. The global resilience of G with respect to $\mathcal{P}$ is the minimum number $\rho' \in \mathbb{R}$ such that by deleting $\rho'|E(G)|$ edges from G, one can obtain a graph that does not have property $\mathcal{P}$. In this thesis we are interested in global graph properties, such as the containment of large or spanning subgraphs. Since global graph properties can be destroyed by small, local changes, such as isolating a vertex of minimum degree, global resilience is merely used for local graph properties and hence this measurement is not suitable for our purposes.

Minimum degree conditions for the containment of large subgraphs H are known for a wide range of graphs with bounded maximum degree, such as powers of Hamilton cycles, trees, and F-factors for any fixed graph F (see e.g. the survey [120] by Kühn and Osthus and the references therein). The following more general result, which confirms a conjecture by Bollobás and Komlós, was proved by Böttcher, Schacht, and Taraz and is known under the name *bandwidth theorem*. The bandwidth of a graph G is defined as the minimum integer b such that there is a labelling of the vertex set of G by integers $1, \dots, |V(G)|$ such that $|i-j| \leq b$ for every edge $\{i,j\} \in E(G)$.

Theorem 1.3 (Böttcher, Schacht, Taraz [41]). *For each $\gamma > 0$, $\Delta \geq 2$, and $k \geq 1$, there exist constants $\beta > 0$ and $n_0 \geq 1$ such that for every $n \geq n_0$ the following holds. If G is a graph on n vertices with minimum degree $\delta(G) \geq \left(\frac{k-1}{k} + \gamma\right) n$ and if H is a k-colourable graph on n vertices with $\Delta(H) \leq \Delta$, bandwidth at most βn, then G contains a copy of H.*

Neither the restriction on the bandwidth, nor the additional term γn in the minimum degree condition can be omitted in Theorem 1.3 (see e.g. [1, 41]). As proved by Böttcher, Taraz, and Würfl [42] the bound on the maximum degree of H can be relaxed to $\sqrt{n}/\log n$ if H is a-arrangeable, i.e. if there exists a labelling of its vertex set by $1, \dots, |V(H)|$ such that the size of the neighbourhood of $N(i) \cap \{i+1, \dots, |V(H)|\}$ restricted to $\{1, \dots, i-1\}$ is at most a for every $i \in [n]$.

Theorem 1.3 applies to a large family of graphs since many interesting classes of graphs have sublinear bandwidth. It is easy to verify that Hamilton cycles and their powers have constant bandwidth. Furthermore, it was proved by Böttcher, Pruessmann, Taraz, and Würfl [40] that planar graphs with bounded maximum degree have bandwidth $\mathcal{O}(n/\log n)$. More generally, they have shown that a hereditary class of bounded degree graphs has sublinear bandwidth if and only if it does not contain expanders of linear order.

The bandwidth theorem subsums therefore, up to the error term, most of the above mentioned results. In fact, for $(k-1)$-th powers of Hamilton cycles, this is only true if k divides the number of vertices since otherwise the chromatic number is $k+1$ and it is known that a minimum degree of $(k-1)n/k$ suffices for an n-vertex graph to contain the $(k-1)$-th power of a Hamilton cycle [108]. However, Böttcher, Schacht, and Taraz have actually proved a stronger version of Theorem 1.3 in [40], where H is allowed to have a few vertices coloured with an additional colour. That theorem includes in particular the cases of all powers of Hamilton cycles.

Universality of random and pseudorandom graphs

The graphs occurring in the theorems above are all *dense*, which means that they have $\Theta(n^2)$ edges if n is the number of their vertices. This leads to the question whether well-behaved *sparse* graphs also contain given spanning subgraphs. We are particularly interested in bijumbled pseudorandom graphs, which we define later, and in the Erdős-Rényi random graph $G(n,p)$, which has n vertices and each pair of vertices forms an edge independently with probability p. From now on, unless stated otherwise, the term *random graph* will always refer to $G(n,p)$.

The theory of random graphs, initiated by Erdős and Rényi [76] around 1960, is an extensively studied field. Especially the problem of determining ranges of the edge probability p, for which it is 'likely' that $G(n,p)$ contains a given subgraph has received a lot of attention. To make this more precise, given a function $p : \mathbb{N} \to [0,1]$, we say that $G(n,p)$ has a graph property $\mathcal{P}$ *asymptotically almost surely* (or a.a.s. for short) if

$$\lim_{n\to\infty} \mathbb{P}[G(n,p) \in \mathcal{P}] = 1.$$

Furthermore, the *threshold* for a monotone increasing property $\mathcal{P}$ is defined as a sequence $\hat{p} = \hat{p}(n)$ such that

$$\lim_{n\to\infty} \mathbb{P}[G(n,p) \in \mathcal{P}] = \begin{cases} 0 & \text{if } p = o(\hat{p}), \\ 1 & \text{if } p = \omega(\hat{p}). \end{cases}$$

A threshold $\hat{p}$ is *sharp* for a monotone increasing property $\mathcal{P}$ if for every constant $\varepsilon > 0$ it holds that $\lim_{n\to\infty} \mathbb{P}[G(n,p) \in \mathcal{P}] = 0$ if $p \leq (1-\varepsilon)\hat{p}$ and $\lim_{n\to\infty} \mathbb{P}[G(n,p) \in \mathcal{P}] = 1$ if $p \geq (1+\varepsilon)\hat{p}$.

It is well known that there exists a threshold for every non-trivial monotone increasing graph property, as shown by Bollobás and Thomason [37]. By now, the threshold for the containment of a specific graph is known for a wide range of classes of graphs. For instance, Bollobás [35], and independently Komlós and Szemerédi [110] (improving on earlier results by Pósa [140] and Korshunov [111]) showed that if $p \geq (\log n + \log\log n + \omega(1))/n$, then $G(n,p)$ is a.a.s. Hamiltonian. In particular, $\log n/n$ is a sharp threshold as it is well known that a.a.s. $G(n,p)$ is disconnected if $p \leq (\log n - \omega(1))/n$, where $\omega(1)$ stands for any function tending to infinity with n arbitrarily slowly (see e.g. [36]).

For general spanning trees there is a polylogarithmic gap between the lower and the best-known upper bound on the threshold. Indeed, Montgomery [129] (improving on Krivelevich [114]) showed that $G(n,p)$ a.a.s. contains a given spanning tree with maximum degree at most Δ if $p \geq \Delta \log^5 n/n$. When restricted to certain kinds of trees, optimal results have been obtained in the sense that they hold a.a.s. if $p \geq (1+\varepsilon)\log n/n$ for any $\varepsilon > 0$. Hefetz, Krivelevich, and Szabó [94] settled the cases when the spanning tree has a linear number of leaves or contains a path of linear length all of whose vertices have degree 2. The case when the spanning tree is a comb, i.e. when it contains n/k vertices, each of which has a disjoint path of length $k-1$ beginning at that vertex, was covered by Montgomery [130], improving an earlier result by Kahn, Lubetzky, and Wormald [100].

Solving a long-standing problem, the Johansson-Kahn-Vu theorem [99] determines the threshold for F-factors with F being a graph for which each proper subgraph F' of F with at least two vertices satisfies

$$d(F') := \tfrac{1}{|V(F')|-1}|E(F')| < \tfrac{1}{|V(F)|-1}|E(F)|.$$

Graphs with that property are called *strictly balanced* and include for instance complete graphs. The Johansson-Kahn-Vu theorem states that, for every strictly balanced graph F, the threshold for $G(n,p)$ to contain an F-factor is $n^{-1/d(F)}(\log n)^{1/|E(H)|}$.

Finally, a general result of Riordan [141] gives an upper bound on the threshold for the containment of spanning subgraphs from various families of graphs. For instance, it determines the threshold for the appearance of a spanning hypercube, of a spanning square lattice, as well as of the k-th power of a Hamilton cycle for $k \geq 3$. The case $k = 2$ was studied by Nenadov and Škorić [133], who (improving on Kühn and Osthus [121]) established the threshold for the appearance of the square of a Hamilton cycle up to a logarithmic factor.

Most of the above mentioned results are not *universal*, i.e. the simultaneous containment of a copy of each graph H from a given class $\mathcal{H}$ is not necessarily guaranteed. The following general universality theorem by Dellamonica, Kohayakawa, Rödl, and Ruciński [60] gives an upper bound on the edge probability such that $G(n,p)$ is a.a.s. universal for the class $\mathcal{H}(n,\Delta)$ of n-vertex graphs with maximum degree at most Δ.

Theorem 1.4 (Dellamonica, Kohayakawa, Rödl, Ruciński [60])**.** *For each $\Delta \geq 3$ there exists a constant $C > 0$ such that if $p \geq C(\log n/n)^{1/\Delta}$, then $G(n,p)$ contains a.a.s. every n-vertex graph H with maximum degree at most Δ.*

Observe that a lower bound on the edge probability p in Theorem 1.4 is given by the threshold for the appearance of a $K_{\Delta+1}$-factor, which is $n^{-2/(\Delta+1)}(\log n)^{1/\binom{\Delta+1}{2}}$ by the Johansson-Kahn-Vu theorem [99].

Let us now turn to spanning structures in pseudorandom graphs. The study of pseudorandom graphs was initiated by Thomason [150, 151] when he was investigating the question whether a property that is satisfied a.a.s. by $G(n,p)$ can be used to describe graphs such that various structural results hold for them as well as a.a.s. for $G(n,p)$. Meanwhile pseudorandom graphs have become a central subject in graph theory (see e.g. the survey [118] by Krivelevich and Sudakov).

We are mostly interested in bijumbled graphs, which are defined as follows. A graph G is called (p,ν)*-bijumbled* if for all disjoint sets $X,Y \subseteq V(G)$ we have

$$\big|e(X,Y) - p|X||Y|\big| \le \nu\sqrt{|X||Y|}\,.$$

The notion of bijumbledness is related to other common notions of pseudorandom graphs that have been studied, namely jumbled graphs and (n,d,λ)-graphs. First, bijumbled graphs and jumbled graphs are equivalent with some loss in the parameters (see e.g. [150]). Second, (n,d,λ)-graphs are in fact a subclass of bijumbled graphs. Recall that an (n,d,λ)-graph is defined as follows. For a graph G with eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$ of the adjacency matrix of G, the value $\lambda(G) := \max_{i\in\{2,\dots,n\}} |\lambda_i|$ is called the *second eigenvalue* of G. An (n,d,λ)-graph is defined as a d-regular graph on n vertices with $\lambda(G) \le \lambda$. The relation between (n,d,λ)-graphs and bijumbled graphs can be seen by the expander mixing lemma (see e.g. [15]), which says that for an (n,d,λ)-graph G it holds that

$$\big|e(X,Y) - \tfrac{d}{n}|X||Y|\big| \le \lambda\sqrt{|X||Y|}$$

for all disjoint sets $X,Y \subseteq V(G)$. An (n,d,λ)-graph is therefore in particular $(d/n,\lambda)$-bijumbled. Note that the reverse implication does not hold as bijumbled graphs are not necessarily regular. However, every (p,ν)-bijumbled d-regular graph is an (n,d,λ)-graph with $\lambda = \mathcal{O}\big(\nu\log(d/\nu)\big)$, as shown by Bilu and Linial [32].

While until recently not much was known concerning spanning subgraphs of general bijumbled graphs, there are various results for (n,d,λ)-graphs, as for instance for the existence of perfect matchings [118], Hamilton cycles [117], triangle factors [119], and, more generally, powers of Hamilton cycles [10].

The just mentioned result on powers of Hamilton cycles in (n,d,λ)-graphs was proved for pseudorandom graphs with a weaker pseudorandomness notion, which implies bijumbledness. In particular, Allen, Böttcher, Hàn, Kohayakawa, and Person [10] proved that for every $k \ge 2$ and $\beta > 0$ there is a constant $\varepsilon > 0$ such that every $(p,\varepsilon p^{3k/2}n)$-bijumbled graph with minimum degree at least βpn contains a k-th power of a Hamilton cycle and every $(p,\varepsilon p^{5/2}n)$-bijumbled graph with the same minimum degree contains a square of a Hamilton cycle.

The proof of Theorem 1.4 on the universality of $G(n,p)$ for $\mathcal{H}(n,\Delta)$ is constructive and gives a pseudorandomness condition which implies that such graphs are universal for $\mathcal{H}(n,\Delta)$. However, this condition is specialised to the proof and not one of the standard, common notions of pseudorandomness. In fact, no standard pseudorandomness condition was known to imply $\mathcal{H}(n,\Delta)$-universality. Recently, Allen, Böttcher, Hàn, Kohayakawa, and Person proved in [9] that bijumbledness does imply $\mathcal{H}(n,\Delta)$-universality. In particular, they showed that for every $\Delta \ge 2$ there exists a constant $c > 0$ such that for any $p > 0$, if $\nu \le cp^{\max\{4,3\Delta/2+1/2\}}n$, then any (p,ν)-bijumbled graph with minimum degree at least $pn/2$ is $\mathcal{H}(n,\Delta)$-universal.

Local resilience of random and pseudorandom graphs

So far we have encountered local resilience results for dense graphs and universality results of random and pseudorandom graphs, all with respect to containing spanning structures. Joining these two strands, in the present thesis we are interested in the local resilience of random and pseudorandom graphs with respect to the simultaneous containment of spanning graphs from specific graph families. The study of local resilience in random and pseudorandom graphs was initiated by Alon, Capalbo, Kohayakawa, Rödl, Ruciński, and Szemerédi [14], and named and further investigated by Sudakov and Vu [149].

Several results have been obtained concerning the local resilience of nearly spanning subgraphs. For instance, Dellamonica, Kohayakawa, Marciniszyn, and Steger [59] studied the local resilience of $G(n,p)$ and of bijumbled graphs with respect to containing nearly spanning cycles and Balogh, Csaba, and Samotij [17] investigated the local resilience of $G(n,p)$ with respect to containing nearly spanning maximum degree bounded trees. Both results are asymptotically best possible, that is, the local resilience of $G(n,p)$ with respect to these properties is a.a.s. at least $(1/2 - o(1))$ for every $p = \Omega(1/n)$. The bound on the edge probability is optimal up to the constant factor. Moreover, the constant $1/2$ cannot be improved, as one can find a.a.s. an approximately even bipartition of the vertex set of $G(n,p)$ such that each vertex v has at most $(1/2 + o(1)) \deg(v)$ neighbours in the other partition class and deleting all edges between the partition class yields a disconnected graph whose largest component has about $n/2$ vertices (see e.g. [59]).

One of the strongest results for spanning subgraphs so far is the following by Lee and Sudakov [124]. Improving on [149], they showed that a.a.s. the local resilience of $G(n,p)$ with respect to Hamiltonicity is at least $(1/2 - o(1))$ when $p = \Omega(\log n/n)$. Again, the bound on the edge probability is optimal up to the constant factor and the constant $1/2$ cannot be improved for the same reason as before. Observe that this is a generalisation of Dirac's theorem (Theorem 1.1) since $G(n,p)$ is equal to K_n if $p = 1$.

The local resilience of random graphs and pseudorandom graphs with respect to containing cycles of all possible lengths was investigated by Krivelevich, Lee and Sudakov [116]. They proved that the local resilience with respect to that property is at least $(1/2 - o(1))$ for (n,d,λ)-graphs with $d^2/n = \omega(\lambda)$ and a.a.s. for $G(n,p)$ if $p = \omega(n^{-1/2})$, which is again best possible in terms of the parameters and the constant $1/2$.

We would like to mention two more results. First, Böttcher, Kohayakawa, and Taraz [39] showed the following random graph version of the bandwidth theorem (Theorem 1.3) for the case that the graphs to be contained are bipartite and nearly spanning.

Theorem 1.5 (Böttcher, Kohayakwa, Taraz [39])**.** *For each $\eta, \gamma > 0$ and $\Delta \geq 2$ there exist constants $\beta, C > 0$ such that the following holds a.a.s. for $\Gamma = G(n,p)$ if $p \geq C(\log n/n)^{1/\Delta}$. Let G be a spanning subgraph of Γ with $\deg_G(v) \geq (1/2 + \gamma)\deg_\Gamma(v)$ for every $v \in V(G)$ and let H be a bipartite graph on $(1-\eta)n$ vertices with $\Delta(H) \leq \Delta$ and bandwidth at most βn. Then G contains a copy of H.*

Second, the Corrádi-Hajnal theorem [58] states that every graph on n vertices with minimum degree at least $2n/3$ contains a triangle factor. The following random graph analogue of this theorem was proved by Balogh, Lee, and Samotij [18].

Theorem 1.6 (Balogh, Lee, Samotij [18])**.** *For each $\gamma > 0$ there exist constants $C, D > 0$ such that if $p \geq C(\log n/n)^{1/2}$, then a.a.s. every spanning subgraph $G \subseteq G(n,p)$ with $\delta(G) \geq (2/3 + \gamma)pn$ contains a triangle factor that covers all but at most Dp^{-2} vertices.*

In Theorem 1.6 the constant $2/3$ as well as the order $O(p^{-2})$ of uncovered vertices are best possible and p is optimal up to the $(\log n)^{1/2}$ factor.

Both in the proof of Theorem 1.5 and the proof of Theorem 1.6 one of the main difficulties was to prove a special case of the blow-up lemma for sparse random graphs. The blow-up lemma is a technical result in extremal graph theory, proved by Komlós, Sárközy, and Szemerédi [109], and instrumental in the proofs of most of the extremal results discussed above. However it applies only to dense graphs. Huang, Lee, and Sudakov [95] used it to prove a version of the bandwidth theorem (Theorem 1.3) for $G(n,p)$ with $0 < p < 1$ being constant. They proved that the local resilience of $G(n,p)$ is a.a.s. at least $(1/k - o(1))$ with respect to H-containment for the graphs of Theorem 1.3 with the additional restriction that a few vertices of H are not allowed to be contained in triangles of H. In fact, they observed that there must be at least $\Omega(p^{-2})$ such vertices, corresponding to the uncovered vertices in Theorem 1.6.

Very recently, a full version of the blow-up lemma for sparse random graphs, again able to handle graphs with maximum degree Δ provided that $p \gg (\log n/n)^{1/\Delta}$, was proved by Allen, Böttcher, Hàn, Kohayakawa, and Person [9] and will be presented in Chapter 2. This result as well as versions for bijumbled graphs and for the embedding of degenerate graphs are essential in our proofs in Chapter 3.

Dirac-type theorems of hypergraphs

An *r-uniform hypergraph* is a tuple (V,E) with $E \subseteq \binom{V}{r}$ and thus the generalisation of a graph. It is therefore natural to ask for degree conditions that force a subhypergraph of the complete hypergraph to contain a copy of some given large structure. Such problems have been studied extensively in the last years, especially for different kinds of Hamilton cycles.

There are several different ways to define problems in hypergraphs analogously to Dirac's theorem (Theorem 1.1) since there are several notions of minimum degrees and cycles for hypergraphs. Given an r-uniform hypergraph $H = (V,E)$ and a set $S \subseteq V$ with $|S| \leq r-1$, the *degree* of S is defined as

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

and the *minimum d-degree* $\delta_d(H)$ of H is defined as

$$\delta_d(H) = \min_{S \subseteq V, |S|=d} \deg_H(S).$$

We simply write $\deg(v)$ to denote the vertex degree of a vertex v in a given hypergraph H and call the minimum 1-degree of H *minimum vertex degree*. The notion of resilience in graphs extends verbatim to the setting of hypergraphs.

Definition 1.7. *Let $r \geq 3$, let $\mathcal{P}$ be a monotone increasing graph property and let H be an r-uniform hypergraph with property $\mathcal{P}$. The local resilience of H with respect to $\mathcal{P}$ is the minimum number ρ such that by deleting at every vertex $v \in V(H)$ at most $\rho \cdot \deg(v)$ hyperedges one can obtain a hypergraph without property $\mathcal{P}$.*

Let us mention that it is common to call minimum degree conditions that force a hypergraph to contain a Hamilton cycle, regardless of the notion of cycles or degrees, *Dirac-type* results.

We will be interested in the local resilience of random r-uniform hypergraphs with respect to weak and Berge Hamiltonicity, which are the earliest notions of cycles in hypergraphs due to Berge [24]. They are defined as follows.

Definition 1.8 (Berge cycle)**.** *A* weak cycle *is an alternating sequence* $(v_1, e_1, v_2, \dots, v_k, e_k)$ *of distinct vertices* $v_1, \dots, v_k$ *and hyperedges* $e_1, \dots, e_k$ *such that* $\{v_1, v_k\} \subseteq e_k$ *and* $\{v_i, v_{i+1}\} \subseteq e_i$ *for every* $i \in [k-1]$*. A weak cycle is called* Berge cycle *if all its hyperedges are distinct.*

If $P = (v_1, e_1, v_2, \dots, v_n, e_n)$ is a weak cycle or a Berge cycle in a hypergraph H on n vertices, then P is called *weak Hamilton cycle* or *Hamilton Berge cycle* of H, respectively.

Other common notions of cycles are ℓ-cycles, which are defined in the following way. For an integer $1 \leq \ell \leq r$, an r-uniform hypergraph C is an ℓ-cycle if there exists a cyclic ordering of the vertices of C such that every hyperedge of C consists of r consecutive vertices and such that every pair of consecutive hyperedges intersects in precisely ℓ vertices. If $\ell = 1$, then C is called a *loose cycle* and if $\ell = r-1$, then C is called a *tight cycle*.

Surprisingly, to the best of our knowledge, the only result on the minimum vertex degree that implies the existence of a weak or a Berge Hamilton cycle is the following one due to Bermond, Germa, Heydemann, and Sotteau [26]. They proved that for every integer $r \geq 3$ and $k \geq r+1$ any r-uniform hypergraph H with minimum vertex degree $\delta_1(H) \geq \binom{k-2}{r-1} + r - 1$ contains a Berge cycle on at least k vertices. If we ask for a Berge Hamilton cycle in an r-uniform hypergraph on n vertices, where r is fixed and n is large, then the bound $\binom{n-2}{r-1} + r - 1$ is weak since it differs from the maximum possible degree by $\binom{n-2}{r-2} - r + 1$.

We would like to mention the following two local resilience results for loose and tight Hamiltonicity. Han and Zhao [89] (improving on [47]) determined that the optimal minimum vertex degree condition that guarantees a loose Hamilton cycle in 3-uniform hypergraphs is $\binom{n-1}{2} - \binom{\lfloor \frac{3}{4}n \rfloor}{2} + c$, where $c = 2$ if 4 divides n and $c = 1$ otherwise. For tight cycles the best-known bound is due to Rödl and Ruciński [143], who proved that if the minimum degree of a 3-uniform hypergraph H is at least $\left((5-\sqrt{5})/3 + \varepsilon\right)\binom{n}{2}$, then H contains a tight Hamilton cycle. Non-trivial bounds for higher uniformities are not known yet. For the various approximate and exact Dirac-type results that are known for ℓ-cycles in terms of minimum d-degrees with $d > 1$, we refer to the surveys [122] by Kühn and Osthus, and [161] by Zhao and the references therein.

Hamiltonicity thresholds of random hypergraphs

Like in the setting of graphs, an intriguing question is which sparse random hypergraphs contain a.a.s. a weak Hamilton cycle or even a Hamilton Berge cycle. By $H^{(r)}(n,p)$ we denote the random r-uniform hypergraph model on the vertex set $[n]$, where each set of r vertices forms an edge randomly and independently with probability $p = p(n)$.

While the threshold for the appearance of a Hamilton Berge cycle in $H^{(r)}(n,p)$ is not yet established, the threshold for a weak Hamilton cycle in $H^{(r)}(n,p)$ is the following one due to Poole [139].

Theorem 1.9 (Poole, [139])**.** *Let* $r \geq 3$*. Then*

$$\mathbb{P}\left[H^{(r)}(n,p) \text{ is weak Hamiltonian}\right] \to \begin{cases} 0 & \text{if } p \leq (r-1)! \frac{\log n - \omega(1)}{n^{r-1}} \\ e^{-e^{-c}} & \text{if } p = (r-1)! \frac{\log n + c_n}{n^{r-1}} \\ 1 & \text{if } p \geq (r-1)! \frac{\log n + \omega(1)}{n^{r-1}} \end{cases},$$

for all functions c_n *tending to* $c \in \mathbb{R}$ *and* $\omega(1)$ *tending to infinity, respectively.*

Since every Hamilton Berge cycle is in particular a weak Hamilton cycle, Theorem 1.9 yields that $H^{(r)}(n,p)$ a.a.s. does not contain a Hamilton Berge cycle if $p \leq (r-1)! \frac{\log n - \omega(1)}{n^{r-1}}$.

An upper bound on the threshold for $H^{(r)}(n,p)$ being Berge Hamiltonian follows from a result by Dudek and Frieze [72], who showed that e/n is the sharp threshold for the appearance of a tight Hamilton cycle in $H^{(r)}(n,p)$ if $r \geq 4$ and $1/n$ is a threshold if $r = 3$. An algorithmic proof for the case $r \geq 4$ with a weaker bound of $p \geq n^{-1+o(1)}$ was given by Allen, Böttcher, Kohayakawa, and Person [11].

Finally, let us mention that, if $p = \omega(\log n / n^{r-1})$, then a.a.s. $H^{(r)}(n,p)$ contains a loose Hamilton cycle if n is a multiple of $r-1$ and $r \geq 3$. This was shown by Dudek, Frieze, Loh, and Speiss in [73]. For thresholds for general ℓ-cycles we refer again to the survey [122] by Kühn and Osthus and the references therein.

Positional games

There is an interesting relation between the concept of local resilience of random graphs and certain positional games called *Maker-Breaker games* played on the edge set of complete graphs. Maker-Breaker games, first studied by Lehman [125], Chvátal and Erdős [52], and Beck [22], have enjoyed great popularity during the last decades (see [23] by Beck and [92] by Hefetz, Krivelevich, Stojaković, and Szabó for thorough surveys on positional games in general and Maker-Breaker games in particular). Before we describe the above mentioned connection in more detail, we first summarise the general setting of Maker-Breaker games.

Let b be a positive integer, X a finite set, and $\mathcal{F} \subseteq 2^X$ a family of subsets of X. A $(1:b)$ Maker-Breaker game $(X, \mathcal{F})$ is defined as follows. The set X and the elements of $\mathcal{F}$ are called *board* and *winning sets*, respectively. The integer b is the *bias* of the game. The game $(X, \mathcal{F})$ is played by two players, called *Maker* and *Breaker*, who alternately claim elements of the board X that have not been claimed before by either of the players. Maker occupies 1 element per turn, while Breaker claims b elements in each of his turns. The game ends when every element of the board is claimed by one of the players. Maker wins the game if he has claimed all elements of at least one winning set in $\mathcal{F}$, otherwise it is Breaker's win. In particular, it is impossible that the game ends in a draw.

If $\mathcal{P}$ is a graph property and X is the edge set of a graph or hypergraph, we write $(X, \mathcal{P})$ to denote the Maker-Breaker game that is played on X and the family of winning sets consists of the edge sets of the subgraphs or subhypergraphs of X that have property $\mathcal{P}$.

Maker-Breaker games are known to be *bias monotone*, i.e. if Maker has a winning strategy for a $(1:b)$ game $(X, \mathcal{F})$, then he also wins the $(1:b-1)$ game $(X, \mathcal{F})$ and if Breaker possesses a strategy to win a $(1:b)$ game $(X, \mathcal{F})$, then he also possesses one for the $(1:b+1)$ game $(X, \mathcal{F})$. Thus, for $(1:b)$ Maker-Breaker games $(X, \mathcal{F})$ with $\mathcal{F} \neq \varnothing$ and $|F| \geq 2$ for each $F \in \mathcal{F}$ it makes sense to study the so-called *threshold bias* $b^*_{\mathcal{F}}$, which is defined as the unique non-negative integer such that Maker has a winning strategy for the corresponding $(1:b)$ game $(X, \mathcal{F})$ if and only if $b < b^*_{\mathcal{F}}$.

The relation mentioned at the beginning of the paragraph allows to derive results for Maker-Breaker games from local resilience results as explicated in the following theorem, which is a special case of a more general one proved by Ferber, Krivelevich, and Naves in [77].

Theorem 1.10 (Ferber, Krivelevich, Naves [77])**.** *For every real* $0 < \varepsilon \leq 1/100$ *the following holds if* n *is sufficiently large. Let* $p = p(n) \in (0,1)$ *and let* $\mathcal{P}$ *be a monotone increasing graph property such that* $G(n,p)$ *has a.a.s. local resilience at least* ε *with respect to* $\mathcal{P}$*. Then Maker has a winning strategy in the* $(1 : \lfloor \varepsilon/(20p) \rfloor)$ *Maker-Breaker game* $\big(E(K_n), \mathcal{P}\big)$*.*

As a case example, let us elaborate on the application of Theorem 1.10 to the so-called *Hamiltonicity game*, a variant of which we will encounter again in Chapter 4. The $(1 : b)$ Hamiltonicity game is defined as the $(1 : b)$ Maker-Breaker game $(E(K_n), \mathcal{F}_H)$, where $\mathcal{F}_H$ is the family of the edge sets of all Hamilton cycles in K_n.

Since the local resilience of $G(n, p)$ with respect to Hamiltonicity is a.a.s. at least $\big(1/2 - o(1)\big)$ when $p = \Omega(\log n/n)$ (see [124]), applying Theorem 1.10 yields that there exists a constant $\alpha > 0$ such that for every $b \leq \alpha n/\log n$, Maker has a winning strategy in the $(1 : b)$ Hamiltonicity game if n is sufficiently large (see also [77]). Thus the local resilience of $G(n, p)$ together with Theorem 1.10 readily imply a lower bound on the threshold bias of the Hamiltonicity game. It is worth mentioning that the determination of the threshold bias $b^*_{\mathcal{F}_H}$ was an open problem for a long period of time until it was finally resolved by Krivelevich [115], who showed that $b^*_{\mathcal{F}_H} = \big(1 - o(1)\big)n/\log n$.

Finally, let us mention that there is an intriguing relation between some biased Maker-Breaker games and random graphs. Observe that, as mentioned above, the threshold for $G(n, p)$ with respect to Hamiltonicity is $\log n/n$, which is asymptotically equal to the reciprocal of $b^*_{\mathcal{F}_H}$. This reflects the so-called *Erdős paradigm* (or *random graph intuition*), which suggests that the threshold bias for a Maker-Breaker game is asymptotically the same as the threshold bias for the same game, where one assumes that both Maker and Breaker claim edges randomly (see e.g. [23, 82, 92] for more information).

1.2.2 Hypergraph matchings, Latin squares, and rainbow matchings

In Chapter 5 we study edge-coloured multigraphs, where each of the colour classes induces a disjoint union of cliques. We determine a condition on the sizes of these colour classes such that a rainbow matching that uses all colours is guaranteed. The motivation for studying this problem arises from different areas of mathematics. In this subsection we summarise previous combinatorial results related to our work. To follow up seamlessly on the previous subsection, let us start with the problem of determining conditions for sparse hypergraphs to contain perfect matchings.

Hypergraph matchings

A *matching* in a hypergraph $H = (V, E)$ is defined as a subset M of E such that all hyperedges in M are pairwise disjoint. We call a hypergraph matching *perfect* if it covers all vertices of the hypergraph.

In graphs, maximum matchings can be found efficiently, for instance using Edmonds' algorithm [74]. In the setting of hypergraphs the problem seems to be more difficult. The decision problem whether, given an integer k, a 3-uniform hypergraph H contains a matching of size at least k is known to be NP-complete [102]. Hence it is natural to search for conditions that imply the existence of a perfect matching in a hypergraph. We would like to mention the following local resilience result of hypergraphs with respect to perfect matchings. For further results on minimum degree conditions for perfect matchings in hypergraphs we refer to the survey [161] by Zhao.

It was proved by Khan [103] and independently by Kühn, Osthus, and Treglown [123] that a minimum degree strictly greater than $\binom{n-1}{2} - \binom{2n/3}{2}$ guarantees a perfect matching in a 3-uniform hypergraph on n vertices if n is divisible by 3 and sufficiently large. They showed that this bound is actually tight. However, it is believed that if one imposes an additional restriction that forces the hypergraph to look somehow 'regular', a much smaller bound on the minimum degree suffices to guarantee a perfect matching.

Indeed, one of the most intriguing conjectures on 3-partite 3-uniform hypergraphs by Ryser [146] suggests that, if n is odd, a minimum degree of n suffices to guarantee a perfect matching in a balanced 3-partite 3-uniform hypergraph on $3n$ vertices if every pair of vertices from different partition classes lies in exactly one hyperedge. If n is even, Brualdi [46] and independently Stein [148] conjectured that such a hypergraph contains a matching of size $n-1$. These conjectures were originally formulated in terms of Latin squares. Before returning to the just mentioned conjecture, we will first introduce Latin squares.

Latin squares

A *Latin square* of order n is an $n \times n$ matrix filled with n different symbols such that each symbol appears exactly once in every row and exactly once in every column. There is a one-to-one correspondence between Latin squares and balanced 3-partite hypergraphs with each pair of vertices from different partition classes being contained in exactly one hyperedge. Simply let the rows, columns, and symbols of a Latin square define the partition classes of a 3-partite hypergraph and let the set of hypergraphs consist of those triples $\{x, y, z\}$, where the symbol z appears in row x and column y. In the same way one can construct a unique Latin square for every hypergraph with the stated property.

The study of Latin squares has a long history and was first systematically developed by Euler (see e.g. [61]). Latin squares have various applications in different branches of mathematics (see [61] for an extensive survey); for instance in algebra, where Latin squares are the multiplication tables of quasigroups, and in a branch of statistics called 'design of experiments', where Latin squares are a special case of row-column designs for two blocking factors. Possibly many people come across Latin squares in recreational mathematics as completed Sudoku puzzles are Latin squares, typically of order 9.

A matching in a hypergraph H, which is associated with a Latin square L, corresponds to a so-called *partial transversal* in L. A partial transversal in a Latin square is a set of entries with distinct symbols such that from each row and each column at most one entry is contained in this set. We call a partial transversal of size n in a Latin square of order n simply *transversal*.

A motivation to study transversals are *orthogonal* Latin squares, where two Latin squares $(A_{i,j})_{i,j\in[n]}$ and $(B_{i,j})_{i,j\in[n]}$ are called orthogonal if all pairs $\{(A_{i,j}, B_{i,j})\}_{i,j\in[n]}$ are distinct. Orthogonal Latin squares, also known under the name *Graeco-Latin squares*, are used for instance in experimental design and tournament scheduling (see e.g. [61]). It can be seen fairly quickly that a Latin square has an orthogonal mate if and only if it has a decomposition into disjoint transversals.

By considering Latin squares of order 2 one can easily verify that not every Latin square has a transversal. For every even $n \in \mathbb{N}$, the cyclic Latin square of order n (i.e. the addition table of the group of integers modulo n) does not have a transversal (see e.g. [156]). For every odd integer $n \geq 11$, however, it is still an open question whether a Latin square of order n without a transversal exists. This brings us back to the famous conjecture by Ryser [146].

Conjecture 1.11 (Ryser [146]). *Every Latin square of odd order has a transversal.*

Observe that this is an equivalent form of the conjecture that we mentioned above in terms of hypergraph matchings. Conjecture 1.11 is known to be true for $n \leq 9$ (see [128]). For all

$n \in \mathbb{N}$, Brualdi [46] conjectured that every Latin square of order n has a partial transversal of size $n-1$. Independently, Stein [148] made the stronger conjecture that every $n \times n$ matrix that is filled with n symbols each appearing exactly n times contains a partial transversal of size $n-1$. Because of the similarity of these conjectures, the following one is often referred to as the Brualdi-Stein conjecture.

Conjecture 1.12 (Brualdi-Stein [46, 148])**.** *For every $n \geq 1$ any Latin square of order n has a partial transversal of size $n-1$.*

Recall that we have already mentioned above an equivalent form of Conjecture 1.12 in the setting of hypergraphs. There have been several approaches to Conjecture 1.12. For instance Hatami and Shor [90] proved that every Latin square contains a partial transversal of length $n - \mathcal{O}(\log^2 n)$. This improves on earlier results by Brouwer, de Vries, Wieringa [45] and Woolbright [158], who showed independently that every Latin square contains a partial transversal of size $n - \sqrt{n}$, and by Drake [64] and Koksma [107], who determined the lower bounds $3n/4$ and $2n/3$, respectively.

There is yet another way to rephrase Conjectures 1.11 and 1.12, namely to the study of rainbow matchings in bipartite edge-coloured graphs. This also allows a more general setting for strengthenings of the above conjectures. For more details on Latin squares we refer to the survey [155] by Wanless.

Rainbow matchings

As already indicated, a natural way to transfer Conjectures 1.11 and 1.12 to graphs is the following. Let $L = (L_{i,j})_{i,j \in [n]}$ be a Latin square of order n. We define $G_L := (A \cup B, E)$ as the complete bipartite edge-coloured graph with partition classes $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_n\}$, where $\{a_i, b_j\}$ is coloured with colour $L_{i,j}$. That is, A and B represent the columns and rows of L, respectively. Moreover, a transversal of L corresponds to a perfect matching in G_L that uses each edge colour exactly once. Such a matching is called *rainbow matching* of size n. Using this notion, Conjecture 1.12 is equivalent to the following: For every $n \geq 1$ any complete bipartite edge-coloured graph, the colour classes of which are perfect matchings, contains a rainbow matching of size $n-1$. It is believed that this is true in the more general setting of bipartite edge-coloured multigraphs. Indeed, Aharoni and Berger [2] conjectured the following generalisation of Conjecture 1.12.

Conjecture 1.13 (Aharoni, Berger [2])**.** *Let G be a bipartite multigraph, the edges of which are coloured with n colours and such that each colour class induces a matching of size $n+1$. Then there is a rainbow matching of size n.*

While Conjecture 1.13 remains open, asymptotic versions are known to be true. For instance, Barat, Gyárfás, and Sárközy [19] extended Woolbright's arguments to multigraphs proving that every bipartite edge-coloured multigraph, each of whose n colour classes has size at least n, contains a rainbow matching with $n - \sqrt{n}$ edges.

Another possibility to approach Conjecture 1.13 is to let the colour classes be bigger than n while keeping the requirement of the rainbow matching to be of size n. Aharoni, Charbit, and Howard [3] proved that sizes of $\lfloor 7n/4 \rfloor$, and Kotlar and Ziv [113] proved that sizes of $\lfloor 5n/3 \rfloor$ suffice to guarantee a rainbow matching of size n. In Chapter 5 we further improve this bound.

It is worthwhile mentioning that when the matchings induced by the colour classes are all edge-disjoint, a theorem by Häggkvist and Johansson [88] on so-called Latin rectangles implies that there is a rainbow matching of size n in case when all colour classes induce edge-disjoint perfect matchings of size $n + o(n)$. Pokrovskiy [138] provided a proof for the more general case where the matchings are not necessarily perfect and thus proved an approximate version of Conjecture 1.13 in the case when the matchings are edge-disjoint.

1.2.3 Enumerative combinatorics

In Chapter 6 we are interested in the question of how many spanning trees are expected to be in a graph chosen uniformly at random from a subfamily of series-parallel graphs. In this section we provide a summary of relevant previous results of enumerative combinatorics; first in view of counting spanning trees and then with regard to properties of series-parallel graphs chosen uniformly at random.

Enumerating spanning trees

The study of spanning trees and their enumeration is a central question in graph theory and combinatorial optimisation. The number of spanning trees of a fixed graph can be computed exactly by using for instance the following methods.

A classic result, which is attributed to Cayley [48], states that the number of labelled trees on n vertices is n^{n-2}. This number is clearly equivalent to the number of spanning trees of the complete graph on n vertices. As a generalisation, *Kirchhoff's matrix tree theorem* [104] provides the number of spanning trees in any fixed graph G. More precisely, in order to apply Kirchhoff's matrix tree theorem to a graph G one needs to determine first the Laplacian matrix associated to G, which is defined as the degree matrix of G minus the adjacency matrix of G. If $\lambda_1, \dots, \lambda_{n-1}$ denote the non-zero eigenvalues of the Laplacian matrix of G, then the number of spanning trees of G equals

$$\tfrac{1}{n} \prod_{i \in [n-1]} \lambda_i.$$

Moreover, it is well known that, given any connected graph $G = (V, E)$, the number of spanning trees of G can be computed using its *Tutte polynomial*

$$T_G(x, y) = \sum\nolimits_{F \subseteq E} (x-1)^{c(F)-1} (y-1)^{|F|+c(F)-|V|},$$

where $c(F)$ denotes the number of connected components of the graph (V, F) (see e.g. [31, Theorem 13.9]). It can be easily seen that the evaluation of $T_G(x, y)$ at $(1, 1)$ returns the number of spanning trees in the graph G. It is worth mentioning that the Tutte polynomial can equivalently be defined as a transformation of the *Potts model*, which is a model of interacting spins on crystalline lattices in statistical mechanics (see e.g. [81]).

A lot of research has been devoted to the study of estimates of the number of spanning trees in the context of restricted graph families. For instance, various results have been obtained for regular graphs and more generally for graphs with given degree sequences (see e.g. [13, 112, 126, 127]).

The enumeration of graphs with a distinguished spanning tree has also been extensively studied in the context of *planar maps*, i.e. proper embeddings of connected multigraphs in the sphere (see e.g. [38] for an introduction to this area).

The first result of this kind was obtained in 1967 by Mullin [132] for rooted planar maps, that is, planar maps with one of their edges marked and assigned an orientation. Mullin showed that the number of unlabelled rooted planar maps on n edges with a distinguished spanning tree is equal to C_nC_{n+1}, where $C_n = \frac{1}{n+1}\binom{2n}{n}$ stands for the n-th Catalan number.

Later, Cori, Dulucq, and Viennot [57] interpreted this formula by means of alternating Baxter permutations (first introduced by Baxter in [21]). Then Bernardi [27] proved a direct bijection between rooted planar maps on n edges with a distinguished spanning tree and pairs consisting of a tree on $n+1$ vertices and a *non-crossing partition* of size $n+1$, i.e. an equivalence relation $\sim$ on a linearly ordered set S of size $n+1$ such that no four elements $a < b < c < d$ of S satisfy $a \sim c$, $b \sim d$, and $a \nsim b$.

Recently, Bousquet-Mélou and Courtiel [43] investigated the enumeration of regular planar maps carrying a distinguished spanning forest, as well as the connections between their counting formulas and the Potts model. For the connection between spanning trees in maps and the Tutte polynomial see e.g. [28].

Properties of series-parallel graphs

Over the past few decades, series-parallel graphs have been extensively studied from various points of view in graph theory, electrical engineering (where they describe electrical circuits) and in computer science (which is due to the fact that many of the standard NP-complete problems can be solved in polynomial time when restricted to the class of series-parallel graphs, see e.g. [16] for a survey).

There are several common equivalent definitions of series-parallel graphs. The most concise one is probably that a graph is *series-parallel* (or SP for short) if it is K_4-minor free. Being a subclass of planar graphs and a superclass of outerplanar graphs, SP graphs turned out to serve well as a pre-stage for the analysis of problems concerning planar graphs. Indeed, the family of SP graphs constitutes the prototype of the so-called *subcritical* graph class family, where, informally speaking, a class is called *subcritical* if in a typical graph on n vertices of this family the largest inclusion-maximal 2-connected subgraph (also called a *block*) has $\mathcal{O}(\log n)$ vertices (see e.g. [71, 84]).

A further definition of SP graphs is that edge-maximal SP graphs, i.e. graphs that cease to be SP whenever an edge is added, are exactly the class of 2-*trees*. Conversely, every subgraph of a 2-tree is series-parallel. Recall that a 2-tree can be defined in the following way: a single edge is a 2-tree, and if T is not a single edge, then T is a 2-tree if and only if there exists a vertex v of degree 2 such that its neighbours are adjacent and $T - v$ is a 2-tree.

One can easily verify that the number of edges of an n-vertex 2-tree is precisely $2n-3$. Moon [131] showed that the number of labelled 2-trees on n vertices equals $\binom{n}{2}(2n-3)^{n-4}$. The enumeration of SP graphs is more involved and an exact value is not known. Bodirsky, Giménez, Kang, and Noy [33] proved the following asymptotic estimate of the number of labelled connected SP graphs on n vertices.

Theorem 1.14 (Bordirsky, Giménez, Kang, Noy [33]). *Let X_n denote the number of labelled connected SP graphs on n vertices. Then*

$$X_n = c_s n^{-5/2} \varrho_s^{-n} n!(1 + o(1)),$$

where $c_s \approx 0.00679$ and $\varrho_s \approx 0.11021$ are computable constants.

In this context *computable* means that the constants can be determined exactly by explicit formulas that appear in the proof. However, those are in general too long to be compactly stated in a theorem.

In the same paper Bodirky, Giménez, Kang, and Noy showed that the number of edges in a random connected SP graph is asymptotically normally distributed with mean asymptotically equal to κn and variance asymptotically equal to λn, where $\kappa \approx 1.61673$ and $\lambda \approx 0.2112$ are again computable constants.

Building on these results, a lot of research has been conducted in order to understand the qualitative picture emerging in the study of graphs chosen uniformly at random from all SP graphs on a fixed number of vertices. For the sake of brevity, in this context we call an object of a given family *random* if the object is chosen uniformly at random among all objects of the same size, e.g. graphs on the same number of vertices.

For instance, the expected value of the maximum degree of a random SP graph on n vertices is asymptotically equal to $c\log(n)$, where $c > 0$ is a computable constant, as shown by Drmota, Giménez, and Noy [70]. In the same paper they proved that this remains true (although with different values for c) when the problem is restricted to the classes of connected or 2-connected SP graphs or to the classes of all, connected or 2-connected outerplanar graphs, respectively.

The expected number of vertices of degree $k = k(n)$ in an n-vertex graph chosen uniformly at random from a subcritical class of graphs was studied by Bernasconi, Panagiotou, and Steger [30]. Using different techniques, Drmota, Giménez, and Noy [69] showed that the number of vertices of a given degree (not depending on n) in a random (connected or 2-connected) outerplanar or SP graph on n vertices is asymptotically normally distributed with mean and variance linear in n.

Several other extremal parameters in subcritical graph classes were investigated by Drmota and Noy [71]. They showed, for instance, that the expected diameter D_n of a random connected SP graph on n vertices satisfies $c_1\sqrt{n} \le \mathbb{E}[D_n] \le c_2\sqrt{n\log(n)}$ for some positive constants c_1 and c_2. The asymptotic estimate has been recently proved by Panagiotou, Stufler, and Weller [137] to be of order $\Theta(\sqrt{n})$.

1.3 Main results

In this section we collect the main results of the present thesis. It is worth mentioning that for the sake of readability we do not intend to optimise constants in our theorems and proofs.

1.3.1 The bandwidth theorem in random and pseudorandom graphs

As elucidated in the previous section, the bandwidth theorem (Theorem 1.3) determines an asymptotically best possible minimum degree condition for the containment of all maximum degree bounded spanning graphs of sublinear bandwidth. In other words, it provides asymptotically the local resilience of K_n with respect to such containment. In Chapter 3 we prove analogues of this theorem first by replacing K_n with $G(n,p)$ and then by replacing it with a bijumbled graph. The first main result is the following bandwidth theorem for sparse random graphs.

Theorem 3.1. *For each $\gamma > 0$, $\Delta \ge 2$, and $k \ge 1$, there exist $\beta^* > 0$ and $C^* > 0$ such that the following holds asymptotically almost surely for $\Gamma = G(n,p)$ whenever $p \ge C^*(\log n/n)^{1/\Delta}$.*

Let G be a spanning subgraph of Γ with $\delta(G) \geq \left(\frac{k-1}{k} + \gamma\right) pn$ and let H be a k-colourable graph on n vertices with $\Delta(H) \leq \Delta$, bandwidth at most $\beta^* n$, and such that there are at least $C^* \max\{p^{-2}, p^{-1} \log n\}$ vertices in $V(H)$ not contained in any triangle of H. Then G contains a copy of H.

We emphasize that the asymptotically almost sure event is that *all* subgraphs G of $G(n,p)$ respecting the stated minimum degree condition contain *all* graphs H with the given property.

As in the dense setting, neither can the minimum degree condition be decreased nor can the bandwidth restriction be omitted in Theorem 3.1. Moreover, the requirement that there have to be vertices in H not contained in any triangle is necessary. The reason for this is that for each $\varepsilon > 0$ there exists some constant $p_\varepsilon > 0$ such that for all $0 < p < p_\varepsilon$ the random graph $G(n,p)$ contains a.a.s. a spanning subgraph G with $\delta(G) > (1-\varepsilon)pn$ such that at least $\varepsilon p^{-2}/3$ vertices of G are not contained in any triangles, as proved by Huang, Lee, and Sudakov [95].

If we impose the additional restriction on H to be D-degenerate, i.e. that every subgraph of H contains a vertex of degree at most D, we can prove a version of Theorem 3.1 for $p = \Omega\big((\log n/n)^{1/(2D+1)}\big)$. However, in this case we do not only require that many vertices are not in triangles of H, but in addition that these vertices are not contained in four-cycles. This is a technical restriction of our proof method; we could remove it, but at the cost of a worse bound on the edge probability. More precisely, we prove the following theorem.

Theorem 3.15. *For each $\gamma > 0$, $\Delta \geq 2$, and $D, k \geq 1$, there exist constants $\beta^* > 0$ and $C^* > 0$ such that the following holds asymptotically almost surely for $\Gamma = G(n,p)$ if $p \geq C^*(\log n/n)^{1/(2D+1)}$.*

Let G be a spanning subgraph of Γ with $\delta(G) \geq \left(\frac{k-1}{k} + \gamma\right) pn$ and let H be a D-degenerate, k-colourable graph on n vertices with $\Delta(H) \leq \Delta$, bandwidth at most $\beta^* n$ and there are at least $C^* \max\{p^{-2}, p^{-1} \log n\}$ vertices in $V(H)$ that are not contained in any triangles or four-cycles of H. Then G contains a copy of H.

Clearly, every tree is 1-degenerate. Moreover, an n-vertex tree with maximum degree at most Δ has bandwidth at most $5n/\log_\Delta n$, as shown by Chung [51]. Therefore, as an immediate consequence of Theorem 3.15 we obtain the following first resilience result of $G(n,p)$ for the containment of maximum degree bounded spanning trees.

Corollary 3.16. *For each $\gamma > 0$ and $\Delta \geq 2$, there exists $C > 0$ such that $\Gamma = G(n,p)$ satisfies the following asymptotically almost surely if $p \geq C(\log n/n)^{1/3}$.*

Let G be a spanning subgraph of Γ with $\delta(G) \geq \big(1/2 + \gamma\big)pn$. Then G contains every spanning tree with maximum degree at most Δ.

Finally, we turn to the resilience of pseudorandom graphs, where we are interested in *bijumbled* graphs. Recall that a graph Γ is called (p,ν)-*bijumbled* if for all disjoint sets $X, Y \subseteq V(\Gamma)$ we have $\big|e(X,Y) - p|X||Y|\big| \leq \nu\sqrt{|X||Y|}$. Our result is the following analogue of the bandwidth theorem.

Theorem 3.19. *For each $\gamma > 0$, $\Delta \geq 2$, and $k \geq 1$, there exists a constant $c > 0$ such that the following holds for any $p > 0$.*

Given $\nu \leq cp^{\max\{4,(3\Delta+1)/2\}}n$, let Γ be a $\big(p,\nu\big)$-bijumbled graph and let G be a spanning subgraph of Γ with $\delta(G) \geq \left(\frac{k-1}{k} + \gamma\right) pn$. Suppose further that H is a k-colourable graph on n vertices with $\Delta(H) \leq \Delta$, bandwidth at most cn and with at least $c^{-1}p^{-6}\nu^2 n^{-1}$ vertices not contained in any triangle of H. Then G contains a copy of H.

We note that for all three theorems above we actually prove a more general statement (following [41]), which allows for a few vertices to receive a $(k+1)$-st colour. Thus we can, for example, show that the local resilience of $G(n,p)$ with respect to Hamiltonicity is a.a.s. at least $1/2 - o(1)$ for all integers n whenever $p = \Omega\big((\log n/n)^{1/2}\big)$.

The proofs of the theorems from this subsection are presented in Chapter 3. They are based on the regularity method and use sparse blow-up lemmas proved by Allen, Böttcher, Hàn, Kohayakawa, and Person [9]. In Chapter 2 we present a blow-up lemma for sparse random graphs as well as some related tools. In that chapter we also prove variants of the sparse regularity lemma and collect bounds on the tails of various probability distributions.

1.3.2 A Dirac-type theorem for Berge cycles in random hypergraphs

The study of the local resilience of random graphs naturally leads to the question of how robust random hypergraphs are with respect to the containment of spanning structures. For instance, as mentioned above, the local resilience of $G(n,p)$ with respect to Hamiltonicity is a.a.s. at least $(1/2 - o(1))$ when $p = \Omega(\log n/n)$. The following theorem can be seen as a hypergraph analogue of this result, where we use the notion of Berge cycles for the desired Hamilton cycle.

Theorem 4.1. *For every integer $r \geq 3$ and every real $\gamma > 0$ the following holds asymptotically almost surely for $\mathcal{H} = H^{(r)}(n,p)$ if $p \geq \frac{\log^{8r} n}{n^{r-1}}$. Let $H \subseteq \mathcal{H}$ be a spanning subgraph with $\delta_1(H) \geq \big(\frac{1}{2^{r-1}} + \gamma\big)\, p\binom{n}{r-1}$. Then H contains a Hamilton Berge cycle.*

As in the local resilience result for random graphs, the minimum vertex degree condition in Theorem 4.1 is also asymptotically tight. The bound on p, however, might not be best possible. But, since $(r-1)!\log n/n^{r-1}$ is the threshold for the appearance of a weak Hamilton cycle in $H^{(r)}(n,p)$ (see [139]) and hence a lower bound on the threshold for Berge Hamiltonicity, the bound on p is tight up to possibly a polylogarithmic factor.

It is worthwhile mentioning that no upper bounds for the threshold of $H^{(r)}(n,p)$ with respect to Berge Hamiltonicity were known, except for those that follow from results with other notions of cycles (see e.g. [72]). As a direct consequence of Theorem 4.1 we obtain that the threshold of $H^{(r)}(n,p)$ with respect to Berge Hamiltonicity is at most $\log^{8r} n/n^{r-1}$.

Since $H^{(r)}(n,p)$ is the complete r-uniform hypergraph if $p = 1$, Theorem 4.1 provides a sufficient minimum vertex degree condition for r-uniform hypergraphs to contain a Hamilton Berge cycle provided that the number n of vertices is large enough. In Chapter 4 we investigate this dense setting more closely, aiming to show tight minimum vertex degree conditions for every n that guarantee the appearance of a weak Hamilton cycle or a Hamilton Berge cycle, respectively. Since in weak Hamilton cycles hyperedges do not need to be distinct, the following proposition can be easily proved by replacing every hyperedge of a given hypergraph by a clique on r vertices and applying Dirac's theorem.

Proposition 4.23. *Let $r \geq 3$ and $n \geq r$ and let H be an r-uniform hypergraph on n vertices. If $\delta_1(H) > \binom{\lceil n/2\rceil - 1}{r-1}$, then H contains a weak Hamilton cycle.*

The condition in Proposition 4.23 is shown to be tight. Proving an analogue for Berge Hamilton cycles seems to be more involved. Using the ideas of the proof of Dirac's theorem we are able to show the following sufficient minimum vertex degree condition, which is significantly lower than the one given by Theorem 4.1 but probably not best possible either.

Proposition 4.24. *Let $r \geq 3$ and let H be an r-uniform hypergraph on $n > 2r-2$ vertices. If $\delta_1(H) \geq \binom{\lceil n/2 \rceil - 1}{r-1} + n - 1$ then H contains a Hamilton Berge cycle.*

To the best of our knowledge, Theorem 4.1 constitutes the first non-trivial local resilience result of random hypergraphs. Against this background, we would like to point out that a similar relation as the one in Theorem 1.10 between local resilience results of random graphs and Maker-Breaker games played on the edge set of complete graphs holds in the setting of hypergraphs. More precisely, following the lines of the proof of Theorem 1.10 in [77] one obtains the following connection of the local resilience of random hypergraphs and the threshold bias of the corresponding Maker-Breaker games played on $E(K_n^{(r)})$.

Theorem 4.25. *For every integer $r \geq 3$, and real $0 < \varepsilon \leq 1/100$ the following holds if n is sufficiently large. Let $p = p(n) \in (0,1)$ and let $\mathcal{P}$ be a monotone increasing graph property such that $H^{(r)}(n,p)$ has a.a.s. local resilience at least ε with respect to $\mathcal{P}$. Then Maker has a winning strategy in the $(1 : \lceil \frac{\varepsilon}{10rp} \rceil)$ Maker-Breaker game $\big(E(K_n^{(r)}), \mathcal{P}\big)$.*

Theorem 4.25 allows us to deduce from Theorem 4.1 a lower bound on the threshold bias of the $(1 : b)$ *Berge Hamiltonicity game*, which is defined as follows. A $(1 : b)$ Maker-Breaker game $(X, \mathcal{F})$ is called $(1 : b)$ Berge Hamiltonicity game if $X = E(K_n^{(r)})$ for an integer $r \geq 3$ and $\mathcal{F}$ is the family of the edge sets of the Hamilton Berge cycles of $K_n^{(r)}$. Using this notion, the following bound holds.

Corollary 4.26. *For every $r \geq 3$ and sufficiently large n, Maker has a winning strategy in the $(1 : b)$ Berge Hamiltonicity game played on $E(K_n^{(r)})$ if $b \leq n^{r-1}/(1000r \log^{8r} n)$.*

We also investigate games, which can be seen as misère versions of the Berge Hamiltonicity game. In *Avoider-Enforcer* games two players, whose names are *Avoider* and *Enforcer*, play according to the conventional rules of the corresponding Maker-Breaker games but their goal is now to lose these corresponding games.

Following Hefetz, Krivelevich, Stojaković, and Szabó [91], we consider two variants of Avoider-Enforcer games. Let b be a positive integer, X be a board, and $\mathcal{F} \subseteq 2^X$ be a family of subsets of X. In the original, *strict* $(1 : b)$ Avoider-Enforcer game $(X, \mathcal{F})$, Avoider occupies exactly 1 and Enforcer exactly b unclaimed elements of X per round. Unlike Maker-Breaker games, strict Avoider-Enforcer games are not bias monotone (see e.g. [93]). This means that the *lower threshold bias* $f_{\mathcal{F}}^-$, which is the largest integer such that Enforcer has a winning strategy for the $(1 : b)$ game $(X, \mathcal{F})$ for every $b \leq f_{\mathcal{F}}^-$, does not necessarily coincide with the *upper threshold bias* $f_{\mathcal{F}}^+$, which is the smallest non-negative integer such that Avoider has a winning strategy for the $(1 : b)$ game $(X, \mathcal{F})$ for every $b > f_{\mathcal{F}}^+$.

In the *monotone* $(1 : b)$ Avoider-Enforcer game $(X, \mathcal{F})$, Avoider occupies at least 1 and Enforcer at least b unclaimed elements of X per round. Games with these monotonicity rules are bias monotone (see e.g. [91]). This means that there exists a unique threshold bias $f_{\mathcal{F}}^{\mathrm{mon}}$, which is defined as the non-negative integer for which Enforcer wins the monotone $(1 : b)$ game if and only if $b \leq f_{\mathcal{F}}^{\mathrm{mon}}$.

In Chapter 4 we consider monotone and strict Avoider-Enforcer games played on the edge set of a complete 3-uniform hypergraph, where Avoider wins if by the end of the game his hypergraph is a Berge-acyclic hypergraph with at most one additional hyperedge. For these games we prove the following bound on the (upper) threshold bias.

Theorem 4.27. *For n sufficiently large and $b \geq 3000n^2 \log^2 n$, Avoider can ensure that in the monotone as well as in the strict $(1 : b)$ Avoider-Enforcer game played on $E(K_n^{(3)})$ by the end of the game Avoider's hypergraph is a Berge-acyclic hypergraph with at most one additional hyperedge.*

For the strict $(1 : b)$ game played on $E(K_n^{(3)})$, where Avoider must avoid Berge cyles, we can show that he has a winning strategy for some bias b between $3000n^2 \log^2 n$ and $3001n^2 \log^2 n$, which yields an upper bound on the lower threshold bias of the Avoider-Enforcer game under consideration.

Theorem 4.28. *For n sufficiently large, there is a bias $3000n^2 \log^2 n \leq b \leq 3001n^2 \log^2 n$ such that Avoider can ensure that in the strict $(1 : b)$ Avoider-Enforcer game played on $E(K_n^{(3)})$ Avoider's hypergraph is Berge-acyclic by the end of the game.*

We would like to mention that, to the best of our knowledge, there are no known results on Maker-Breaker nor Avoider-Enforcer games played on the edge sets of complete hypergraphs so far. We refer to the books [23] by Beck and [92] by Hefetz, Krivelevich, Stojaković, and Szabó and to the references therein for Maker-Breaker and Avoider-Enforcer games played on different boards.

In Chapter 4 we present the proof of Theorem 4.1, which is based on the absorbing method developed by Rödl, Ruciński, and Szemerédi [144], and uses tools from the proof of a Dirac-type result for random directed graphs by Ferber, Nenadov, Noever, Peter, and Škoric [78]. We also prove Propositions 4.23 and 4.24 in detail and explain how the proof of Theorem 1.10 must be modified for Theorem 4.25. The proofs of Theorems 4.27 and 4.28 are extensions of the proofs in the joint work [55] with Dennis Clemens, Yury Person, and Tuan Tran to the setting of 3-uniform hypergraphs.

1.3.3 Rainbow matchings in multigraphs

In Chapter 5 we prove a result on edge-coloured multigraphs that affirms asymptotically an algebraic question by Grinblat. It also implies a partial result towards a conjecture by Aharoni and Berger (Conjecture 1.13). For motivations and related results with regard to this conjecture see Subsection 1.2.2.

While previously the appearance of rainbow matchings was studied primarily in properly edge-coloured bipartite graphs or multigraphs, we examine general multigraphs and allow the edge-colouring to be non-proper, meaning that there can be adjacent edges of the same colour. However, we require that each colour class induces a disjoint union of cliques. We show that if each of n colour classes covers $3n + o(n)$ vertices of the multigraph, there exists a rainbow matching of size n. More precisely, we prove the following.

Theorem 5.4. *For every $\delta > 0$ there exists $n_0 = n_0(\delta) = 144/\delta^2$ such that the following holds for every $n \geq n_0$. Let G be a multigraph, the edges of which are coloured with n colours. If each subgraph of G induced by a colour class has at least $(3+\delta)n$ vertices and is the disjoint union of non-trivial cliques, then G contains a rainbow matching of size n.*

Theorem 5.4 is asymptotically best possible. This can be seen by taking all colour classes to be identical and to be the disjoint union of $(n - 1)$ triangles. Such a multigraph does not contain a rainbow matching of size n.

In the case that the multigraph G is bipartite and thus each clique has size 2 we obtain immediately the following corollary.

Corollary 5.5. *For every $\varepsilon > 0$ there exists an integer $n_0 \geq 1$ such that for every $n \geq n_0$ the following holds. Let G be a bipartite multigraph whose edges are coloured with n colours and each colour class induces a matching of size at least $\left(\frac{3}{2} + \varepsilon\right) n$. Then G contains a rainbow matching of size n.*

We note that Clemens and the current author gave an independent direct proof of Corollary 5.5 in [53] before Theorem 5.4 was proved by Clemens, Pokrovskiy, and the current author in [56]. In this thesis we will solely provide the more general second proof.

Corollary 5.5 marks a step towards Conjecture 1.13 by Aharoni and Berger, which says that if $f(n)$ is the smallest integer m such that every bipartite edge-coloured multigraph with n colour classes, each being a matching of size at least m, contains a rainbow matching of size n, then $f(n) = n + 1$ holds. Using this notion, Corollary 5.5 states that $f(n) \leq 3n/2 + o(n)$, which is asymptotically the same as the best-known bound $\lfloor 3n/2 \rfloor$ on the sizes of the colour classes in the case where one aims to find a rainbow matchings of size $n - 1$ (see e.g. [113]). It is worth noting that very recently Aharoni, Kotlar, and Ziv slightly improved our result to $f(n) \leq \lceil 3n/2 \rceil + 1$ in the preprint [5].

As mentioned above, Theorem 5.4 affirms asymptotically an algebraic question of Grinblat. In order to formulate this question we require a few definitions.

Let X be a set and let $\mathcal{P}(X)$ denote its power set. A nonempty subset $\mathcal{A} \subseteq \mathcal{P}(X)$ is an *algebra* on X if $\mathcal{A}$ is closed under complementation and under unions, i.e. if $M_1, M_2 \in \mathcal{A}$, then $X \setminus M_1 \in \mathcal{A}$ and $M_1 \cup M_2 \in \mathcal{A}$.

In a series of papers and books [85, 86, 87] Grinblat investigated sufficient conditions for countable families $\{\mathcal{A}_i\}_i$ of algebras such that $\bigcup_i A_i \neq \mathcal{P}(X)$ and $\bigcup_i A_i = \mathcal{P}(X)$, respectively. In this context, Grinblat [85] defined $\mathfrak{v} = \mathfrak{v}(n)$ as the minimal cardinal number such that the following is true:

> "Let $\mathcal{A}_1, \ldots, \mathcal{A}_n$ be algebras on a set X such that for each $i \in [n]$ there exist at least $\mathfrak{v}(n)$ pairwise disjoint sets in $\mathcal{P}(X) \setminus \mathcal{A}_i$. Then there exists a family $\{U_i^1, U_i^2\}_{i \in [n]}$ of $2n$ pairwise disjoint subsets of X such that, for each $i \in [n]$, if $Q \in \mathcal{P}(X)$ and Q contains one of the two sets U_i^1 and U_i^2 and its intersection with the other one is empty, then $Q \notin \mathcal{A}_i$."

In [85] Grinblat showed that $\mathfrak{v}(3) = 9$ and $\mathfrak{v}(n) \geq 3n - 2$ for each $n \in \mathbb{N}$ and posed the following question.

Question 5.1 (Grinblat, [85])**.** *Is it true that $\mathfrak{v}(n) = 3n - 2$ for $n \geq 4$?*

Improving on Grinblat [87], who established the upper bound $\mathfrak{v}(n) \leq 10n/3 + \sqrt{2n/3}$, Nivasch and Omri [134] proved that $\mathfrak{v}(n) \leq 16n/5 + \mathcal{O}(1)$ using an equivalent definition of $\mathfrak{v}(n)$ in the context of equivalence relations.

In Chapter 5 we argue that Theorem 5.4 is equivalent to the following one.

Theorem 5.2. *For every $\delta > 0$ there exists $n_0 = n_0(\delta) = 144/\delta^2$ such that for every $n \geq n_0$ it holds that $\mathfrak{v}(n) \leq (3 + \delta)n$.*

Since $(3n - 2)$ is a lower bound on $\mathfrak{v}(n)$, Theorem 5.2 gives an asymptotic answer to Question 5.1.

1.3.4 Enumerating spanning trees in series-parallel graphs

In Chapter 6 we study the number of spanning trees in graphs from various subfamilies of connected series-parallel graphs.

In order to determine the precise number of spanning trees in a fixed graph G, one can apply for instance Kirchhoff's matrix tree theorem or evaluate the Tutte polynomial of G (cf. Section 1.2). However, if one would like to estimate how many spanning trees one expects in a graph chosen uniformly at random from a given graph class, one can no longer apply the standard approaches as in the setting of a fixed graph. A common method in this situation is the use of analytic combinatorics, which will also be the key ingredient in the proofs of our following results.

Our first theorem provides the following asymptotic estimate of the expected number of spanning trees in a graph chosen uniformly at random from the family of connected series-parallel graphs or from the family of 2-connected series-parallel graphs.

Theorem 6.1. *Let X_n and Z_n denote the number of spanning trees in a connected and, respectively, 2-connected labelled SP graph on n vertices chosen uniformly at random. Then,*

$$\begin{aligned} \mathbb{E}[X_n] &= \quad s\varrho^{-n}(1+o(1)), \quad \text{where } s \approx 0.09063,\ \ \varrho^{-1} \approx 2.08415, \\ \mathbb{E}[Z_n] &= \quad p\varpi^{-n}(1+o(1)), \quad \text{where } p \approx 0.25975,\ \ \varpi^{-1} \approx 2.25829. \end{aligned}$$

While Theorem 6.1 deals with the family of *all* connected/2-connected SP graphs with a given number of vertices, we also analyse the growth constant of the expected number of spanning trees in graphs chosen uniformly at random from the family of all 2-connected n-vertex SP graphs with a given edge density. The study of extremal situations, i.e. when the graphs are edge-maximal or have only a few more edges than a tree, is addressed separately in more detail. First we estimate the expected number of spanning trees in an edge-maximal n-vertex SP graph chosen uniformly at random. In this case, the expected number of spanning trees is slightly bigger than the estimate provided by Theorem 6.1.

Theorem 6.8. *Let U_n denote the number of spanning trees in a labelled 2-tree on n vertices chosen uniformly at random. Then, the expected value of U_n is asymptotically equal to $s_2\varrho_2^{-n}$, where $s_2 \approx 0.14307$ and $\varrho_2^{-1} \approx 2.55561$.*

As for the other extremal case, we elaborate on the expected number of spanning trees in a graph chosen uniformly at random from the family of all connected SP graph on n vertices and excess equal to a constant k, i.e. for which the number of edges and vertices differ by a constant k.

Our result is the following polynomial estimate of the expected number of spanning trees.

Theorem 6.12. *Let $k \geq 2$. Let $X_{n,k}$ denote the number of spanning trees in a connected labelled SP graph on n vertices and with fixed excess equal to k chosen uniformly at random.*

Then for sufficiently large n we have

$$\mathbb{E}[X_{n,k}] = \tilde{c}(k)\frac{\Gamma(3k/2)}{\Gamma(2k+1/2)}\left(\frac{n}{2}\right)^{\frac{k+1}{2}}(1+o(1)),$$

where for large values of k the function $\tilde{c}(k)$ satisfies

$$\tilde{c}(k) = \tilde{c}\tilde{\gamma}^{-k}(1+o(1)), \tag{1.1}$$

with $\tilde{c} \approx 0.90959$ and $\tilde{\gamma}^{-1} \approx 2.60560$.

We would like to emphasize that the previous formulas should be understood in the following way: first we fix k and then we let n tend to infinity. Additionally, if k is sufficiently large, we can get the approximation of $\tilde{c}(k)$ stated in the second part of Theorem 6.12.

The proofs of Theorems 6.1, 6.8, and 6.12 are based on graph decompositions and analytic combinatorics, which includes in particular the symbolic method, the singularity analysis of generating functions, and transfer theorems. We present the necessary combinatorial and analytic background in Chapter 2.

1.4 Organisation

The thesis is organised as follows.

In Chapter 2 we introduce the notation and definitions that we frequently use in the thesis. Moreover, we summarise necessary background material concerning the regularity method and analytic combinatorics, and state concentration inequalities for various probability distributions.

Next in Chapter 3 we prove analogues of the bandwidth theorem for sparse random and pseudorandom graphs, as well as a version for the embedding of degenerate graphs in sparse random graphs. These results are based on joint work with Peter Allen, Julia Böttcher, and Anusch Taraz [7, 8].

Then in Chapter 4 we present the proof of a Dirac-type theorem for Hamilton Berge cycles in random r-uniform hypergraphs, investigate the problem for dense r-uniform hypergraphs, and prove bounds on the biases for which Avoider can keep his hypergraph (almost) Berge-acyclic in monotone and strict Avoider-Enforcer games played on the edge set of a complete 3-uniform hypergraph. The first part is joint work with Dennis Clemens and Yury Person [54] and the second part extends a joint work with Dennis Clemens, Yury Person, and Tuan Tran [55].

Chapter 5 is devoted to the proof of a result on edge-coloured multigraphs that affirms asymptotically an algebraic question by Grinblat and is a partial result towards a conjecture by Aharoni and Berger. This chapter is based on joint work with Dennis Clemens and Alexey Pokrovskiy [56] and with Dennis Clemens [53].

Finally in Chapter 6 we prove an asymptotic estimate of the expected number of spanning trees in a labelled connected series-parallel graph chosen uniformly at random. Furthermore, we obtain analogue results for subfamilies of series-parallel graphs such as 2-trees and connected series-parallel graphs with fixed excess. This is based on joint work with Juanjo Rué [75].

2

Tools and notation

2.1 Definitions

In this section we introduce the basic definitions that are frequently used throughout the present thesis. The terminologies that we need with regard to the regularity method and analytic combinatorics are introduced in Section 2.2 and Section 2.3, respectively. Other definitions, which occur rarely or are used in only certain parts of the thesis, are deferred to the places where they are needed. For all elementary graph theoretic concepts not defined in this section we refer the reader to e.g. [62].

2.1.1 General notions

Throughout the thesis log denotes the natural logarithm, whereas we write $\log_2$ for the logarithm to the base 2. For a positive integer n we define $[n] := \{1, \ldots, n\}$ and for reals $a, b > 0$ we write $x = a \pm b$ if and only if $a - b \leq x \leq a + b$. For the sake of brevity, when we write $x > 0$ we always assume x to be real and when we write $x \geq n$ with n being a positive integer, we always assume that x is an integer, unless stated otherwise. Given a set S and a positive integer $r \leq |S|$ we let $\binom{S}{r}$ denote the set $\{S' \subseteq S : |S'| = r\}$. For a sequence $\mathcal{A} = (a_1, \ldots, a_k)$ we write $a \in \mathcal{A}$ if there exists an index $i \in [k]$ with $a = a_i$.

Given a bivariate function $A(x, y)$, we denote the partial derivative of $A(x, y)$ with respect to x and y by $A_x(x, y)$ and $A_y(x, y)$, respectively. However, we will usually use the notation $A'(x, y) := A_x(x, y)$ for the derivative of $A(x, y)$ with respect to the first variable.

To express asymptotic behaviours we use the standard Landau notation. In particular, given two functions $f, g : \mathbb{N} \to \mathbb{R} \setminus \{0\}$ we write $f = o(g)$ if $\lim_{n\to\infty} |f(n)/g(n)| = 0$ and $f = \omega(g)$ if $\lim_{n\to\infty} |g(n)/f(n)| = 0$. If there exist constants $C > 0$ and $n_0 \in \mathbb{N}$ such that $|f(n)| \leq C|g(n)|$ for every $n \geq n_0$, we use $f = \mathcal{O}(g)$ and if there exist $c > 0$ and $n_0 \in \mathbb{N}$ such that $|f(n)| \geq c|g(n)|$, we write $f = \Omega(g)$. If $f = \Omega(g)$ and $f = \mathcal{O}(g)$, we use the notation $f = \Theta(g)$. Finally, if $\lim_{n\to\infty} f(n)/g(n) = 1$, we say that f and g are *asymptotically equal* and write $f \sim g$. In all cases, for the sake of simplicity, we may replace f by $f(n)$ or g by $g(n)$, for instance we may write $f = \mathcal{O}(n)$.

By $\mathrm{Bin}(n, p)$ we denote the binomial distribution with parameters n and p.

2.1.2 Graphs and multigraphs

Most of the graph-theoretic terminology that we use is standard and follows [62]. All graphs and hypergraphs in this thesis are finite and undirected. Graphs and hypergraphs are always simple, while multigraphs may have multiple edges and loops, unless otherwise specified.

A *graph* G is a pair (V, E) of sets V and E with $E \subseteq \binom{V}{2}$, where V is called *vertex set* and E *edge set* of G. We say that vertices $x, y \in V$ are *adjacent* if $\{x, y\} \in E(G)$, and say that a vertex $x \in V$ and an edge $e \in E$ are *incident* if $x \in e$. Two distinct edges are called *adjacent* if their intersection is nonempty. Given a graph G, we use $V(G)$ and $E(G)$ to refer to its vertex and edge set, respectively.

Let $G = (V, E)$ be a graph and let $A, B \subseteq V$ be disjoint. We denote the set of edges in A by $E_G(A) := \{e \in E : e \subseteq A\}$ and the set of those between A and B by $E_G(A, B) := \big\{\{x, y\} \in E : x \in A \text{ and } y \in B\big\}$. The cardinalities of these sets are denoted by $e_G(A) := |E_G(A)|$ and $e_G(A, B) := |E_G(A, B)|$. For a vertex $x \in V$ we write $N_G(x) := \{y \in V : \{x, y\} \in E\}$ for the *neighbourhood* of x in G and $N_G(x, A) := N_G(x) \cap A$ for the neighbourhood of x restricted to A. Given vertices $x_1, \ldots, x_k \in V$ we denote the *joint neighbourhood* of $x_1, \ldots, x_k$ restricted to A by $N_G(x_1, \ldots, x_k; A) = \bigcap_{i \in [k]} N_G(x_i, A)$.

The *degree* of a vertex $x \in V$ is the size of its neighbourhood and is denoted by $\deg_G(x) := |N_G(x)|$. Similarly, we use the notation $\deg_G(x, A) := |N_G(x, A)|$ and $\deg_G(x_1, \ldots, x_k; A) := |N_G(x_1, \ldots, x_k; A)|$ for the degree of x restricted to A in G and the size of the joint neighbourhood of $x_1, \ldots, x_k$ restricted to A in G, respectively. For short notation we may omit the subscript G when there is no risk of confusion.

The *minimum* and *maximum degree* of G are denoted by $\delta(G)$ and $\Delta(G)$, respectively. If $\delta(G) = \Delta(G)$, then G is called *regular* and for the special case $\delta(G) = \Delta(G) = 3$, we say that G is *cubic*.

Given a graph $G = (V, E)$, we say that a graph $H = (V', E')$ is a *subgraph* of G if $V' \subseteq V$ and $E' \subseteq E$. If H is a subgraph of G we write $H \subseteq G$. A subgraph $H = (V', E') \subseteq G$ is *spanning* if $V' = V$ and *induced* if for all edges $e \in E$ with $e \subseteq V'$ it holds that $e \in E'$. We write $G[V']$ for the induced subgraph of G on the vertex set V'. A set $W \subseteq V$ is called *independent* if the edge set of $G[W]$ is empty. Moreover, for a set $A \subseteq V$ we write $G - A := G[V \setminus A]$.

A *graph homomorphism* or simply *homomorphism* from a graph H to a graph G is a mapping $h : V(H) \to V(G)$ such that for every $\{x, y\} \in E(H)$ it holds that $\{h(x), h(y)\} \in E(G)$. Two graphs H and H' are *isomorphic* if there exists a bijective homomorphism h from H to G such that the inverse function h^{-1} is again a homomorphism. In this case we also say that H is a *copy* of H'. If G has a subgraph that is isomorphic to H, we say that G contains a copy of H or that there is an *embedding* of H onto G. When it is clear from the context we simply say that G contains H (as a subgraph). A graph G is called *universal* for a class of graphs $\mathcal{H}$ if G contains all graphs from $\mathcal{H}$ as subgraphs.

We say that a graph G is *connected* if for all vertices $x, y \in V(G)$ there is a path in G with endpoints x and y. A maximal connected subgraph of G is a *connected component* or simply a *component* of G. A graph G is called *k-connected* if $|V(G)| \geq k$ and if one cannot disconnect G by deleting $(k-1)$ vertices of G. A graph that does not contain any cycles is called a *forest* or *acyclic*. Connected acyclic graphs are *trees*.

A *Hamilton cycle* in a graph G is a spanning subgraph of G that is a cycle. A graph G is said to be *Hamiltonian* if G contains a Hamilton cycle. Given a graph G, the *distance*

$\text{dist}_G(x, y)$ between two vertices $x, y \in V(G)$ is the length of a shortest path in G connecting them. The *k-th power* of a graph G is the graph on the vertex set $V(G)$ such that the edge set consists of all sets $\{x, y\}$ for which $x, y \in V(G)$ and $\text{dist}_G(x, y) \leq k$. The second power of a graph G is also called *square* of G.

Given a graph H, an *H-factor* of a graph G is a subgraph of G consisting of disjoint unions of copies of H. An H-factor is called *perfect* if it is spanning. The edge set of a K_2-factor is called *matching* and the edge set of a perfect K_2-factor *perfect matching.* The size of a matching M is defined as the number of edges of M. A vertex v is said to be *saturated* by a matching M if there is an edge $e \in M$ with $v \in e$.

A graph property is called *monotone increasing* if the property is preserved under edge and vertex insertions. We say that a graph G is *edge-maximal* with respect to some graph property $\mathcal{P}$ if G has property $\mathcal{P}$ but adding any edge to G produces a graph that does no longer have property $\mathcal{P}$.

A graph H is a *minor* of a graph G if H can be obtained from G by a sequence of vertex and edge deletions as well as edge contractions. We say that a family $\mathcal{H}'$ of graphs is *$\mathcal{H}$-minor free* for a family $\mathcal{H}$ of graphs if no graph of $\mathcal{H}'$ contains a graph from $\mathcal{H}$ as a minor.

Outerplanar graphs are the family of $\{K_4, K_{3,2}\}$-minor free graphs, *series-parallel* graphs are those that exclude K_4 as a minor, and a graph is called *planar* if it neither contains K_5 nor $K_{3,3}$ as a minor. Edge-maximal series-parallel graphs are called *2-trees.* Equivalent definitions of series-parallel graphs and 2-trees are discussed in Chapters 1 and 6.

A *partition* of a graph G is a partition of its vertex set $V(G)$ into disjoint sets. These sets are referred to as *partition classes.*

A *vertex-colouring* of a graph is a map from the vertex set of a graph to a set of colours. An *edge-colouring* is defined analogously. A *colouring* is said to be *proper* if adjacent vertices or edges do not receive the same colour. A graph is said to be *k-colourable* if there exists a proper colouring of its vertex set with k colours. A *colour class* of a colouring is a maximal set of vertices or edges of the same colour.

A graph G is called *k-partite*, if G is k-colourable. For $r = 2$ we say that G is *bipartite.* If G is bipartite we may write $G = (A \cup B, E)$ to indicate that A and B are partition classes of a bipartition of G.

We let K_n denote the complete graph on n vertices and $K_{n,m}$ the complete bipartite graph with partition classes of size n and m. The complete graph K_3 is also called *triangle.*

Given a graph $G = (V, E)$, a *labelling* of its vertex set is a bijective function from V to $[|V|]$. We define a *labelled* graph to be a pair (V, E) equipped with a labelling of V. Informally, a labelled graph is a graph whose vertices bear distinct labels.

The *bandwidth* of a graph, denoted by $\text{bw}(G)$, is defined as the minimum positive integer b such that there is a labelling of $V(G)$ such that $|i - j| \leq b$ for every edge $\{i, j\} \in E(G)$.

Formally, a *multigraph* is a pair (V, E) of sets V and E with a map $E \to V \cup (V \times V)$, i.e. every edge is mapped to one or two vertices. One can think of a multigraph as a graph with multiple edges and loops. For any vertices $x, y \in V$, the number of edges that are mapped to $\{x, y\}$ is called *multiplicity* of $\{x, y\}$.

2.1.3 Hypergraphs

For every integer $r \geq 3$, an *r-uniform hypergraph* is a pair (V, E) of a vertex set V and an edge set E with $E \subseteq \binom{V}{r}$. The elements of E are referred to as *edges* or *hyperedges*. Given a hypergraph H, as in the setting of graphs, we use $V(H)$ and $E(H)$ to refer to its vertex and edge set, respectively. Also, the definition of subgraphs generalises naturally to subhypergraphs of hypergraphs.

Given an r-uniform hypergraph $H = (V, E)$ and a set $S \subseteq V$ with $|S| \leq r-1$, the *degree* of S is defined as $\deg_H(S) := |\{e \in E : S \subseteq e\}|$. The *minimum d-degree* $\delta_d(H)$ of H is defined as $\delta_d(H) := \min\{\deg_H(S) : S \subseteq V, |S| = d\}$. We call $\deg_H(\{x\})$ the *vertex degree* of x and $\delta_1(H)$ the *minimum vertex degree* of H and use $\deg_H(x) := \deg_H(\{x\})$ and $\delta(H) := \delta_1(H)$ for short notation. We may omit the subscript H whenever there is no risk of confusion.

An r-uniform hypergraph H is *r-partite* if its vertex set can be partitioned into r disjoint sets such that each hyperedge contains exactly one element from each of the r sets. A *matching* of a hypergraph H is a set of pairwise disjoint hyperedges. A matching M of H is *perfect* if every vertex of H is contained in a hyperedge of M.

A *weak cycle* is an alternating sequence $(v_1, e_1, v_2, \ldots, v_k, e_k)$ of distinct vertices $v_1, \ldots, v_k$ and hyperedges $e_1, \ldots, e_k$ such that $\{v_1, v_k\} \subseteq e_k$ and $\{v_i, v_{i+1}\} \subseteq e_i$ for every $i \in [k-1]$. A weak cycle is called *Berge cycle* if all its hyperedges are distinct. If $P = (v_1, e_1, v_2, \ldots, v_n, e_n)$ is a weak cycle or a Berge cycle in a hypergraph H on n vertices, then P is called *weak Hamilton cycle* or *Hamilton Berge cycle* of H, respectively. A hypergraph H is called *Berge-acyclic* if H does not contain any Berge cycle as a subgraph.

Given an r-uniform hypergraph $H = (V, E)$ and sets $A_1, A_2, \ldots, A_\ell \subseteq V$ as well as positive integers $r_1, r_2, \ldots, r_\ell$ with $\sum_{i \in [\ell]} r_i = r$, we write

$$E_H\big(A_1^{(r_1)}, A_2^{(r_2)}, \ldots, A_\ell^{(r_\ell)}\big) := \big\{e \in E(H) : \exists (v_1, v_2, \ldots, v_r) \in A_1^{r_1} \times A_2^{r_2} \times \ldots \times A_\ell^{r_\ell} \text{ with } e = \{v_1, v_2, \ldots, v_r\}\big\}.$$

Moreover, we define $e_H\big(A_1^{(r_1)}, A_2^{(r_2)}, \ldots, A_\ell^{(r_\ell)}\big) := |E_H\big(A_1^{(r_1)}, A_2^{(r_2)}, \ldots, A_\ell^{(r_\ell)}\big)|$. Again, we may omit the subscript H whenever it is clear from the context.

Finally, we denote the complete r-uniform hypergraph on n vertices by $K_n^{(r)}$.

2.1.4 Random graphs and hypergraphs

The random graph model that we consider is the Erdős-Rényi model $G(n, p)$, which is defined on the vertex set $[n]$ where each pair of vertices forms an edge randomly and independently with probability $p = p(n)$. The generalisation of $G(n, p)$ to hypergraphs is defined as follows. For every $r \geq 3$ we denote by $H^{(r)}(n, p)$ the random r-uniform hypergraph model on the vertex set $[n]$, where each set of r vertices forms an edge randomly and independently with probability $p = p(n)$.

Given a function $p : \mathbb{N} \to [0, 1]$, we say that $G(n, p)$ has a graph property $\mathcal{P}$ *asymptotically almost surely* (or a.a.s. for short) if $\lim_{n\to\infty} \mathbb{P}[G(n, p) \in \mathcal{P}] = 1$. Furthermore, the *threshold* for an increasing property $\mathcal{P}$ is defined as a sequence $\hat{p} = \hat{p}(n)$ such that

$$\lim_{n\to\infty} \mathbb{P}[G(n, p) \in \mathcal{P}] = \begin{cases} 0 & \text{if } p = o(\hat{p}), \\ 1 & \text{if } p = \omega(\hat{p}). \end{cases}$$

These notions are defined analogously for $H^{(r)}(n,p)$.

Finally, the notion of pseudorandomness that we use in our results is bijumbledness. We say that a graph G is (p,ν)-*bijumbled* if for all disjoint sets $X, Y \subseteq V(G)$ we have

$$\big|e(X,Y) - p|X||Y|\big| \leq \nu\sqrt{|X||Y|}\,.$$

We remark that in Chapter 6 an object chosen uniformly at random from all elements of a family is called *random object* of this family, e.g. a random connected series-parallel graph on n vertices is a graph chosen uniformly at random from the family of all connected series-parallel graphs on n vertices. However, in the other parts of the thesis the term *random graph* stands exclusively for $G(n,p)$.

2.2 The regularity method

In Chapter 3 we address the problem of finding embeddings of graphs from a given graph family into subgraphs of sparse random and pseudorandom graphs. In our proofs we use the *regularity method*, which combines powerful *regularity lemmas* and *blow-up lemmas*.

In order to prove that a graph G contains a graph H using the regularity method, one typically proceeds as follows. Roughly speaking, first one prepares the host graph G using a regularity lemma in order to obtain a partition of $V(G)$ with a constant number of clusters such that most pairs are pairwise so-called *regular*. Then one needs to find a suitable substructure in the so-called *reduced graph* corresponding to the partition. Next one applies some technical manipulations in order to achieve for instance the so-called *super-regularity* of some specific pairs and an appropriate partition of $V(H)$. Finally, after possibly embedding manually a few vertices of H into G, one applies a blow-up lemma to embed the remainder of H into super-regular substructures of the partition of $V(G)$.

In the following subsection we introduce necessary definitions concerning the regularity method and prove in particular two versions of the sparse regularity lemma. Subsection 2.2.2 is then devoted to the statements of the blow-up lemma for sparse graphs and related lemmas that are important in our proofs.

2.2.1 Sparse regular partitions

An essential concept of the regularity method is the notion of regular pairs. Let $G=(V,E)$ be a graph, and let $\varepsilon, d > 0$ and $p \in (0,1]$ be reals. Moreover, let $X, Y \subseteq V$ be two disjoint nonempty sets. The *p-density* of the pair (X,Y) is defined as

$$d_{G,p}(X,Y) := \frac{e_G(X,Y)}{p|X||Y|}.$$

We give two definitions of regularity. The first one requires a lower bound on the p-density of all subpairs of a certain size.

Definition 2.1 ((Super-)regular pairs). *The pair (X,Y) is called $(\varepsilon,d,p)_G$-*regular *if for every $X' \subseteq X$ and $Y' \subseteq Y$ with $|X'| \geq \varepsilon|X|$ and $|Y'| \geq \varepsilon|Y|$ we have*

$$d_{G,p}(X',Y') \geq d - \varepsilon.$$

*If additionally we have $|N_G(x,Y)| \geq (d-\varepsilon)p|Y|$ and $|N_G(y,X)| \geq (d-\varepsilon)p|X|$ for every $x \in X$ and $y \in Y$, then the pair (X,Y) is called $(\varepsilon,d,p)_G$-*super-regular.

The second definition additionally imposes an upper bound on the p-density of those subpairs.

Definition 2.2 ((Super-)fully-regular)**.** *The pair (X,Y) is called* $(\varepsilon,d,p)_G$-fully-regular *if there exists $d' \geq d$ such that for every $X' \subseteq X$ and $Y' \subseteq Y$ with $|X'| \geq \varepsilon|X|$ and $|Y'| \geq \varepsilon|Y|$ we have*

$$d_{G,p}(X',Y') = d' \pm \varepsilon.$$

If additionally we have $|N_G(x,Y)| \geq (d-\varepsilon)p|Y|$ and $|N_G(y,X)| \geq (d-\varepsilon)p|X|$ for every $x \in X$ and $y \in Y$, then the pair (X,Y) is called $(\varepsilon,d,p)_G$-super-fully-regular.

Most of the time we will use the first version of regularity, which is sometimes called *lower-regularity*. This is the version we have to use when dealing with subgraphs of random graphs. However, the sparse regularity lemma gives us fully-regular pairs, and we have to work with this stronger concept when dealing with subgraphs of bijumbled graphs. Note that an (ε,d,p)-fully-regular pair is in particular (ε,d,p)-regular.

Whenever there is no risk of confusion, we might omit the subscript G in $(\varepsilon,d,p)_G$-(super-)(fully-)regular, which indicates with respect to which graph a pair is (super-)(fully-)regular. A direct consequence of the definition of (ε,d,p)-regular pairs is the following proposition about the sizes of neighbourhoods in regular pairs.

Proposition 2.3. *Let (X,Y) be (ε,d,p)-regular. Then there are less than $\varepsilon|X|$ vertices $x \in X$ with $|N(x,Y)| < (d-\varepsilon)p|Y|$.* □

The next proposition asserts that small alterations of the vertex sets of an (ε,d,p)-regular pair do not destroy regularity.

Proposition 2.4. *Let (X,Y) be an (ε,d,p)-regular pair in a graph G and let $\hat{X}$ and $\hat{Y}$ be two subsets of $V(G)$ such that $|X \triangle \hat{X}| \leq \mu|X|$ and $|Y \triangle \hat{Y}| \leq \nu|Y|$ for some $0 \leq \mu,\nu \leq 1$. Then $(\hat{X},\hat{Y})$ is $(\hat{\varepsilon},d,p)$-regular, where $\hat{\varepsilon} := \varepsilon + 2\sqrt{\mu} + 2\sqrt{\nu}$. Furthermore, if for any disjoint subsets $A, A' \subseteq V(G)$ with $|A| \geq \mu|X|$ and $|A'| \geq \nu|Y|$ we have $e(A,A') \leq (1+\mu+\nu)p|A||A'|$, and (X,Y) is (ε,d,p)-fully-regular, then $(\hat{X},\hat{Y})$ is $(\hat{\varepsilon},d,p)$-fully-regular.*

Proof. Let $A \subseteq \hat{X}$ and $B \subseteq \hat{Y}$ such that $|A| \geq \hat{\varepsilon}|\hat{X}|$ and $|B| \geq \hat{\varepsilon}|\hat{Y}|$ be given. Define $A' := A \cap X$ and $B' := B \cap Y$ and note that

$$|A'| \geq |A| - \mu|X| \geq \hat{\varepsilon}|\hat{X}| - \mu|X| \geq \hat{\varepsilon}(1-\mu)|X| - \mu|X| \geq \big(\hat{\varepsilon} - 2\sqrt{\mu}\big)|X| \geq \varepsilon|X|$$

by the definition of $\hat{\varepsilon}$. Analogously, one can show that $|B'| \geq \varepsilon|Y|$. Since (X,Y) is an (ε,d,p)-regular pair, we know that $d_p(A',B') \geq d-\varepsilon$. Furthermore, we have

$$|A'| \geq |A| - \mu|X| \geq |A| - \mu\frac{|A|}{\hat{\varepsilon}} \geq \big(1-\sqrt{\mu}\big)|A|$$

and by an analogous calculation we get $|B'| \geq \big(1-\sqrt{\nu}\big)|B|$. For the number of edges between A and B we get

$$\begin{aligned} e(A,B) &\geq e(A',B') \geq (d-\varepsilon)p|A'||B'| \geq (d-\varepsilon)p\big(1-\sqrt{\mu}\big)\big(1-\sqrt{\nu}\big)|A||B| \\ &\geq \big(d-\varepsilon-2\sqrt{\mu}-2\sqrt{\nu}\big)p|A||B| \geq (d-\hat{\varepsilon})p|A||B|. \end{aligned}$$

Therefore we have

$$d_p(A,B) \geq d - \hat{\varepsilon}.$$

Now suppose that (X,Y) is (ε,d,p)-fully-regular. Let d' be such that $d_p(A',B') = d' \pm \varepsilon$ for any $A' \subseteq X$ and $B' \subseteq Y$ with $|A'| \geq \varepsilon|X|$ and $|B'| \geq \varepsilon|Y|$. Let $A \subseteq \hat{X}$ and $B \subseteq \hat{Y}$ with $|A| \geq \hat{\varepsilon}|\hat{X}|$ and $|B| \geq \hat{\varepsilon}|\hat{Y}|$ be given. As above, we obtain $e(A,B) \geq (d'-\hat{\varepsilon})p|A||B|$. We also have

$$\begin{aligned} e(A,B) &\leq e(A',B') + e(A', B\setminus B') + e(A\setminus A', B) \\ &\leq (d'+\varepsilon)p|A'||B'| + (1+\mu+\nu)p|A'|\nu|B| + (1+\mu+\nu)p\mu|A||B| \\ &\leq (d'+\hat{\varepsilon})|A||B|\,, \end{aligned}$$

so that $(\hat{X},\hat{Y})$ is $(\hat{\varepsilon},d,p)$-fully-regular, as desired. $\square$

A partition $\mathcal{V} = \{V_i\}_{i\in\{0,\dots,r\}}$ of the vertex set of G is called an $(\varepsilon,p)_G$*-(fully-)regular partition* of $V(G)$ if $|V_0| \leq \varepsilon|V(G)|$ and $(V_i,V_{i'})$ forms an (ε,p)-(fully-)regular pair in G for all but at most $\varepsilon\binom{r}{2}$ pairs $\{i,i'\} \in \binom{[r]}{2}$. The partition $\mathcal{V}$ is called (ε,d,p)*-(super-)(fully-)regular* on a graph $R = ([r],F)$ if $|V_0| \leq \varepsilon|V(G)|$ and $(V_i,V_{i'})$ is (ε,d,p)-(super-)(fully-)regular for every $\{i,i'\} \in F$. The graph R is referred to as the *reduced graph* of $\mathcal{V}$, the partition classes V_i with $i \in [r]$ as *clusters*, and V_0 as the *exceptional set*. We call $\mathcal{V}$ an *equipartition* if $|V_i| = |V_{i'}|$ for every $i,i' \in [r]$.

Analogously to Szemerédi's regularity lemma for dense graphs, the sparse regularity lemma, proved by Kohayakawa and Rödl [105, 106], asserts the existence of an (ε,p)-fully-regular partition of constant size of any sparse graph. We need two versions of this lemma in our proofs. The first one allows an initial partition with parts of different sizes to be equitably refined.

Before stating this lemma, we require one more definition. Given a partition $\{V_i\}_{i\in[s]}$ of the vertex set of a graph G, we say that a partition $\{V_{i,j}\}_{i\in[s],j\in[t]}$ is an *equitable* (ε,p)*-regular refinement* of $\{V_i\}_{i\in[s]}$ if $|V_{i,j}| = |V_{i,j'}| \pm 1$ for each $i \in [s]$ and $j,j' \in [t]$, and there are at most εs^2t^2 pairs $(V_{i,j},V_{i',j'})$ that are not (ε,p)-regular.

Lemma 2.5. *For each $\varepsilon > 0$ and $s \geq 1$ there exists $t_1 \geq 1$ such that the following holds for every $0 < p < 1$. Given any graph G, suppose $\{V_i\}_{i\in[s]}$ is a partition of $V(G)$. If $e(V_i) \leq 3p|V_i|^2$ for each $i \in [s]$, and $e(V_i,V_{i'}) \leq 2p|V_i||V_{i'}|$ for all distinct $i,i' \in [s]$, then there exist sets $V_{i,0} \subseteq V_i$ for each $i \in [s]$ with $|V_{i,0}| < \varepsilon|V_i|$, and an equitable (ε,p)-regular refinement $\{V_{i,j}\}_{i\in[s],j\in[t]}$ of $\{V_i \setminus V_{i,0}\}_{i\in[s]}$ for some $t \leq t_1$.*

The proof of Lemma 2.5 follows the proof of a sparse regularity lemma by Scott [147].

Proof of Lemma 2.5. Given $\varepsilon > 0$ and $s \geq 1$, set $L = 100s^2\varepsilon^{-1}$. Let $n_1 = 1$, and for each $j \geq 1$ let $n_{j+1} = 10000\varepsilon^{-1}n_j2^{sn_j}$. Finally, set $t_1 = n_{1000\varepsilon^{-5}(L^2+16Ls^2)+1}$.

For all $i,i' \in [s]$, we define the energy $\mathcal{E}(P,P')$ of a pair of disjoint subsets $P \subseteq V_i$ and $P' \subseteq V_{i'}$ to be

$$\mathcal{E}(P,P') = \frac{|P||P'|\min\left\{d_{G,p}(P,P')^2, 2L\cdot d_{G,p}(P,P') - L^2\right\}}{|V_i||V_{i'}|}\,.$$

Note that this quantity is convex in $d_{G,p}(P,P')$. Given a partition $\mathcal{P}$ refining $\{V_i\}_{i\in[s]}$, we define the energy $\mathcal{E}(\mathcal{P})$ of $\mathcal{P}$ to be

$$\mathcal{E}(\mathcal{P}) := \sum_{\{P,P'\}\subseteq\mathcal{P}} \mathcal{E}(P,P')\,.$$

We now construct a succession of partitions $\mathcal{P}_{j+1}$ for each $j \geq 1$, each refining $\mathcal{P}_1 := \{V_i\}_{i\in[s]}$, and claim that the following holds for every $j \geq 2$:

(R1) $\mathcal{P}_j$ partitions every set $V_i \in \{V_{i'}\}_{i'\in[s]}$ into between n_j and $\big(1 + 10^{-4}\varepsilon\big)n_j$ sets such that the largest n_j sets are equally sized.

(R2) $\mathcal{E}(\mathcal{P}_2) \geq 10^{-3}\varepsilon^5 j$.

We stop if $\mathcal{P}_j$ is $(\varepsilon/2, p)$-regular. If not, then we apply the following procedure.

For each pair of $\mathcal{P}_j$ that is not $(\varepsilon/2, p)$-regular, we take a witness of its irregularity, consisting of a subset of each side of the pair. We let $\mathcal{P}'_j$ be the union of the Venn diagrams of all witness sets in each part of $\mathcal{P}_j$. Since $\mathcal{P}_j$ is not $(\varepsilon/2, p)$-regular, there are at least $\frac{1}{2}\varepsilon s^2 n_j^2$ pairs that are not $(\varepsilon/2, p)$-regular. By choice of L and by (R1), at least $\frac{1}{4}\varepsilon s^2 n_j^2$ of these pairs have density at most $L/2$. By the defect Cauchy-Schwarz inequality, just from refining these pairs we conclude that $\mathcal{E}(\mathcal{P}'_j) \geq \mathcal{E}(\mathcal{P}_j) + 10^{-3}\varepsilon^5$ (cf. [147]). Note that, by convexity of $\mathcal{E}(P, P')$ in $d_{G,p}(P, P)$, refining the other pairs does not affect $\mathcal{E}(P'_j)$ negatively.

We now let $\mathcal{P}_{j+1}$ be obtained by splitting each set of $\mathcal{P}'_j$ within each V_i into sets of size $\frac{1000-\varepsilon}{1000 n_{j+1}}|V_i|$ plus at most one smaller set. By Jensen's inequality, we have $\mathcal{E}(\mathcal{P}_{j+1}) \geq \mathcal{E}(\mathcal{P}'_j)$ (cf. [147]), giving (R2). Since $\mathcal{P}'_j$ partitions each V_i into at most $n_j 2^{s n_j} = 10^{-4}\varepsilon n_{j+1}$, the total number of smaller sets is at most $10^{-4}\varepsilon n_{j+1}$. This gives (R1).

Observe that for any partition $\mathcal{P}$ refining $\mathcal{P}_1$, we have $\mathcal{E}(\mathcal{P}) \leq L^2 + 16Ls^2$. It follows that this procedure must terminate with $j \leq 1000\varepsilon^{-5}(L^2 + 16Ls^2) + 1$. The final partition $\mathcal{P}_j$ is thus $(\varepsilon/2, p)$-regular. For each $i \in [s]$, let $V_{i,0}$ consist of the union of all but the largest n_j parts of $\mathcal{P}_j$. Let $\mathcal{P}$ be the partition of $\bigcup_{i\in[s]} V_i \setminus V_{i,0}$ given by $\mathcal{P}_j$. This is the desired equitable (ε, p)-regular refinement of $\{V_i \setminus V_{i,0}\}_{i\in[s]}$. □

The second variant of the sparse regularity lemma that we need is the following minimum degree version.

Lemma 2.6 (Minimum degree version of the sparse regularity lemma)**.** *For each $\varepsilon > 0$ and $r_0 \geq 1$ there exists $r_1 \geq 1$ with the following property. For any $d \in [0, 1]$, any positive reals α and p, and any n-vertex graph G with minimum degree αpn such that for any disjoint $X, Y \subseteq V(G)$ with $|X|, |Y| \geq \frac{\varepsilon n}{r_1}$ we have $e(X, Y) \leq \big(1 + \frac{1}{1000}\varepsilon^2\big)p|X||Y|$, there is an $(\varepsilon, d, p)_G$-fully-regular equipartition of $V(G)$ with reduced graph R such that $\delta(R) \geq (\alpha - d - \varepsilon)|V(R)|$ and $r_0 \leq |V(R)| \leq r_1$.*

Using Lemma 2.5 we now give a proof of Lemma 2.6, which follows [39] (c.f. [105]).

Proof of Lemma 2.6. Given $\varepsilon > 0$ and $r_0 \geq 1$, without loss of generality we assume $\varepsilon \leq 1/10$. Let t_1 be returned by Lemma 2.5 for input $\varepsilon^2/(1000s)$ and $s = 100r_0/\varepsilon$. Set $r_1 = st_1$.

Given $\alpha > 0$ and $p > 0$, let G be an n-vertex graph with minimum degree αpn. Let $\{V_i\}_{i\in[s]}$ be an arbitrary partition of $V(G)$ into sets of as equal size as possible. By assumption, we have $e(V_i, V_{i'}) \leq 2p|V_i||V_{i'}|$ for distinct indices $i, i' \in [s]$. Furthermore, if V_i is a part with $e(V_i) \geq 3p|V_i|^2$, then for a maximum cut (A, A') of V_i we have $e(A, A') \geq 3p|V_i|^2/2$. Enlarging the smaller of the sets A and A' if necessary, we have a pair of subsets of $V(G)$ both of size at most $|V_i|$ between which there are at least $3p|V_i|^2/2$ edges, contradicting the assumption of Lemma 2.6. Thus G satisfies the conditions of Lemma 2.5 with input $\varepsilon^2/(1000s)$ and s. Applying that lemma, we obtain a collection $\{V_{i,0}\}_{i\in[s]}$ of sets, and an (ε, p)-regular partition

$\mathcal{P}$ of $\bigcup_{i\in[s]} V_i \setminus V_{i,0}$, which partitions each $V_i \setminus V_0$ into $t \le t_1$ sets. Note that $r_0 \le s \le |\mathcal{P}| \le r_1$ by construction.

Now let V_0' be the union of the $V_{i,0}$ for $i \in [s]$, of each set $W \in \mathcal{P}$ that lies in more than $\varepsilon st/4$ pairs which are not $(\varepsilon/1000, p)$-regular, and at most two vertices from each set $W \in \mathcal{P}$ in order that the partition of $V(G) \setminus V_0'$ induced by $\mathcal{P}$ is an equipartition. Because the total number of pairs that are not $(\varepsilon/1000, p)$-regular is at most $\varepsilon^2/(1000sr_0^2t^2)$, the number of such sets in any given V_i is at most $\varepsilon t/100$, so $|V_{i,0}'|$ has size at most $\varepsilon|V_i|/50$, and the number of parts of $\mathcal{P}$ in $V_i \setminus V_{i,0}'$ is larger than $t/2$. Thus the partition $\mathcal{P}'$ of $V(G) \setminus V_0'$ induced by $\mathcal{P}$ is an (ε, p)-regular equipartition of $V(G) \setminus V_0'$, and we have $|V_0'| \le \varepsilon n$.

We claim that the partition $\mathcal{P}'$ has all the properties we require. It remains to verify that for each $d \in [0, 1]$, the d-reduced graph of $\mathcal{P}'$ has minimum degree at least $(\alpha-d-\varepsilon)t'$. Suppose that P is a part of $\mathcal{P}'$. Now we have $e(P) \le 3p|P|^2$, since otherwise, as before, a maximum cut (A, A') of P has at least $3p|P|^2/2 < \varepsilon p|P|n/20$ edges, yielding a contradiction to the assumption on the maximum density of pairs of G. By construction, P lies in at most $\varepsilon t'/2$ pairs that are not (ε, p)-regular, and these contain at most $(1+\varepsilon/10)p|P|\big(\varepsilon t'|P|/2\big) < \frac{3}{4}\varepsilon p|P|n$ edges of G. We conclude that at least $\alpha p|P|n - \frac{7}{8}\varepsilon p|P|n$ edges of G leaving P lie in (ε, p)-regular pairs of $\mathcal{P}'$. Of these, at most $dp|P|n$ can lie in pairs of density less than p, so that the remaining at least $\big(\alpha - d - \frac{7}{8}\varepsilon\big)p|P|n$ edges lie in (ε, d, p)-regular pairs. If so many edges were in less than $(\alpha - d - \varepsilon)t'$ pairs leaving P, this would contradict our assumption on the maximum density of G, so that we conclude that P lies in at least $(\alpha - d - \varepsilon)t'$ pairs that are (ε, d, p)-regular, as desired. □

2.2.2 Blow-up lemmas for sparse graphs

A key ingredient in the proofs in Chapter 3 is the *sparse blow-up lemma* developed by Allen, Böttcher, Hàn, Kohayakawa, and Person [9]. Given a subgraph $G \subseteq \Gamma = G(n, p)$ with $p = \Omega((\log n/n)^{1/\Delta})$ and an n-vertex graph H with maximum degree at most Δ with vertex partitions $\mathcal{V}$ and $\mathcal{W}$, respectively, the sparse blow-up lemma guarantees under certain conditions a spanning embedding of H in G that respects the given partitions. In order to state this lemma we need to introduce some definitions.

Definition 2.7 ((ϑ, R')-buffer)**.** *Let R' be a graph on r vertices and let H be a graph with vertex partition $\mathcal{W} = \{W_i\}_{i\in[r]}$. We say that the family $\widetilde{\mathcal{W}} = \{\widetilde{W}_i\}_{i\in[r]}$ of subsets $\widetilde{W}_i \subseteq W_i$ is an* (ϑ, R')-buffer *for H if*

(B1) $|\widetilde{W}_i| \ge \vartheta|W_i|$ for all $i \in [r]$ and

(B2) for each $i \in [r]$ and each $x \in \widetilde{W}_i$, the first and second neighbourhood of x go along R', i.e., for each $\{x, y\}, \{y, z\} \in E(H)$ with $y \in W_j$ and $z \in W_k$ we have $\{i, j\} \in E(R')$ and $\{j, k\} \in E(R')$.

Let G and H be graphs on n vertices with partitions $\mathcal{V} = \{V_i\}_{i\in[r]}$ of $V(G)$ and $\mathcal{W} = \{W_i\}_{i\in[r]}$ of $V(H)$. We say that $\mathcal{V}$ and $\mathcal{W}$ are *size-compatible* if $|V_i| = |W_i|$ for all $i \in [r]$. If there exists an integer $m \ge 1$ such that $m \le |V_i| \le \kappa m$ for every $i \in [r]$, then we say that $(G, \mathcal{V})$ is κ-balanced. Given a graph R on r vertices, we call $(G, \mathcal{V})$ an *R-partition* if for every distinct indices $i, i' \in [R]$ and for every edge $\{x, y\} \in E(G)$ with $x \in V_i$ and $y \in V_{i'}$ we have $\{i, i'\} \in E(R)$.

Definition 2.8 (Restriction pair). *Let $\varepsilon, d > 0$, $p \in [0,1]$, and let R be a graph on r vertices. Furthermore, let G be a (not necessarily spanning) subgraph of $\Gamma = G(n,p)$ and let H be a graph given with vertex partitions $\mathcal{V} = \{V_i\}_{i\in[r]}$ and $\mathcal{W} = \{W_i\}_{i\in[r]}$, respectively, such that $(G,\mathcal{V})$ and $(H,\mathcal{W})$ are size-compatible R-partitions. Let $\mathcal{I} = \{I_x\}_{x\in V(H)}$ be a collection of subsets of $V(G)$, called* image restrictions*, and $\mathcal{J} = \{J_x\}_{x\in V(H)}$ be a collection of subsets of $V(\Gamma)\setminus V(G)$, called* restricting vertices*. For each $i \in [r]$ we define $R_i \subseteq W_i$ as the set of all vertices $x \in W_i$ for which $I_x \neq V_i$. We say that $\mathcal{I}$ and $\mathcal{J}$ are a $(\rho,\zeta,\Delta,\Delta_J)$-*restriction pair *if the following properties hold.*

(RP1) *For each $i \in [r]$ we have $|R_i| \le \rho|W_i|$,*

(RP2) *if $x \in R_i$, then $|J_x| + \deg_H(x) \le \Delta$ and if $x \in W_i \setminus R_i$, then $J_x = \varnothing$,*

(RP3) *for each $i \in [r]$ and $x \in R_i$ we have $I_x \subseteq \bigcap_{u\in J_x} N_\Gamma(u,V_i)$ and $|I_x| \ge \zeta(dp/2)^{|J_x|}|V_i|$,*

(RP4) *for each $i \in [r]$ and $x \in W_i$, we have $(p-\varepsilon p)^{|J_x|}|V_i| \le \big|\bigcap_{u\in J_x} N_\Gamma(u,V_i)\big| \le (p+\varepsilon p)^{|J_x|}|V_i|$,*

(RP5) *the pair $\Big(V_i \cap \big(\bigcap_{u\in J_x} N_\Gamma(u)\big), V_j \cap \big(\bigcap_{v\in J_y} N_\Gamma(v)\big)\Big)$ is $(\varepsilon,d,p)_G$-regular for every edge $\{x,y\} \in E(H)$ with $x \in R_i$ and $y \in W_j$,*

(RP6) *each vertex in $V(G)$ appears in at most Δ_J of the sets of $\mathcal{J}$.*

Suppose $\mathcal{V}$ is an $(\varepsilon,d,p)_G$-regular partition of $V(G)$ with reduced graph R. We say $(G,\mathcal{V})$ has *one-sided inheritance* with respect to R if for all $\{i,j\},\{j,k\} \in E(R)$ and every $v \in V_i$ the pair $\big(N_\Gamma(v,V_j),V_k\big)$ is $(\varepsilon,d,p)_G$-regular, and $V_i \in \mathcal{V}$ has *two-sided inheritance* with respect to $V_j, V_k \in \mathcal{V}$ if for every $v \in V_i$ the pair $\big(N_\Gamma(v,V_j),N_\Gamma(v,V_k)\big)$ is $(\varepsilon,d,p)_G$-regular.

Now we can finally state the sparse blow-up lemma.

Theorem 2.9 (Sparse blow-up lemma, [9]). *For each Δ, $\Delta_{R'}$, Δ_J, $\vartheta,\zeta,d > 0$, $\kappa > 1$ there exist $\varepsilon_{\mathrm{BL}},\rho > 0$ such that for all $r_1 \ge 1$ there is a $C_{\mathrm{BL}} > 0$ such that for $p \ge C_{\mathrm{BL}}(\log n/n)^{1/\Delta}$ asymptotically almost surely $\Gamma = G_{n,p}$ satisfies the following.*

Let R be a graph on $r \le r_1$ vertices and let $R' \subseteq R$ be a spanning subgraph with $\Delta(R') \le \Delta_{R'}$. Let H and $G \subseteq \Gamma$ be graphs given with κ-balanced, size-compatible vertex partitions $\mathcal{W} = \{W_i\}_{i\in[r]}$ and $\mathcal{V} = \{V_i\}_{i\in[r]}$ with parts of size at least $m \ge n/(\kappa r_1)$. Let $\mathcal{I} = \{I_x\}_{x\in V(H)}$ be a family of image restrictions, and $\mathcal{J} = \{J_x\}_{x\in V(H)}$ be a family of restricting vertices. Suppose that

(BUL1) *$\Delta(H) \le \Delta$, for every edge $\{x,y\} \in E(H)$ with $x \in W_i$ and $y \in W_j$ we have $\{i,j\} \in E(R)$, and $\widetilde{\mathcal{W}} = \{\widetilde{W}_i\}_{i\in[r]}$ is an (ϑ,R')-buffer for H,*

(BUL2) *$\mathcal{V}$ is $(\varepsilon_{\mathrm{BL}},d,p)_G$-regular on R, $(\varepsilon_{\mathrm{BL}},d,p)_G$-super-regular on R' and has one-sided inheritance on R',*

(BUL3) *for every vertex $x \in \widetilde{W}_i$ and every triangle $\{x,y,z\}$ in H with $y \in W_j$ and $z \in W_k$, the set V_i has two-sided inheritance with respect to V_j and V_k,*

(BUL4) *$\mathcal{I}$ and $\mathcal{J}$ form a $(\rho,\zeta,\Delta,\Delta_J)$-restriction pair.*

Then there is an embedding $\phi\colon V(H) \to V(G)$ such that $\phi(x) \in I_x$ for each $x \in H$.

Observe that in the blow-up lemma for dense graphs, proved by Komlós, Sárközy, and Szemerédi [109], one does not need to explicitly ask for one- and two-sided inheritance properties since they are always fulfilled by dense regular partitions. This is, however, not true in general in the sparse setting. The following two lemmas will be useful whenever we need to redistribute vertex partitions in order to achieve some regularity inheritance properties.

Lemma 2.10 (One-sided regularity inheritance, [9]). *For each $\varepsilon_{\mathrm{OSRIL}}, \alpha_{\mathrm{OSRIL}} > 0$ there exist $\varepsilon_0 > 0$ and $C > 0$ such that for any $0 < \varepsilon < \varepsilon_0$ and $0 < p < 1$ asymptotically almost surely $\Gamma = G(n, p)$ has the following property. For any disjoint sets X and Y in $V(\Gamma)$ with $|X| \geq Cp^{-2} \log n$ and $|Y| \geq Cp^{-1} \log n$, and any subgraph $G \subseteq \Gamma$ such that (X, Y) is $(\varepsilon, \alpha_{\mathrm{OSRIL}}, p)_G$-regular, there are at most $Cp^{-1} \log n$ vertices $z \in V(\Gamma)$ such that $(X \cap N_\Gamma(z), Y)$ is not $(\varepsilon_{\mathrm{OSRIL}}, \alpha_{\mathrm{OSRIL}}, p)_G$-regular.*

Lemma 2.11 (Two-sided regularity inheritance, [9]). *For each $\varepsilon_{\mathrm{TSRIL}}, \alpha_{\mathrm{TSRIL}} > 0$ there exist $\varepsilon_0 > 0$ and $C > 0$ such that for any $0 < \varepsilon < \varepsilon_0$ and $0 < p < 1$, asymptotically almost surely $\Gamma = G_{n,p}$ has the following property. For any disjoint sets X and Y in $V(\Gamma)$ with $|X|, |Y| \geq C \max\{p^{-2}, p^{-1} \log n\}$, and any subgraph $G \subseteq \Gamma$ such that (X, Y) is $(\varepsilon, \alpha_{\mathrm{TSRIL}}, p)_G$-regular, there are at most $C \max\{p^{-2}, p^{-1} \log n\}$ vertices $z \in V(\Gamma)$ such that $\big(X \cap N_\Gamma(z), Y \cap N_\Gamma(z)\big)$ is not $(\varepsilon_{\mathrm{TSRIL}}, \alpha_{\mathrm{TSRIL}}, p)_G$-regular.*

2.3 Analytic combinatorics

In enumerative combinatorics, obtaining an *exact* formula for the number of combinatorial objects seems to be often out of reach. A common method for the *approximate* quantitative study of combinatorial structures is *analytic combinatorics*, which has proved to have a vast list of applications in the analysis of algorithms, statistical physics, computational biology, information theory, and many other fields (see e.g. [80]).

The key objects in analytic combinatorics are *generating functions* that are associated with the structures to be counted. The typical asymptotic enumeration process can be divided into the following three steps. First, combinatorial constructions are used to specify the object under study. With the help of the so-called *symbolic method* these constructions are then transferred into equations of the corresponding generating functions. Finally, *singularity analysis* is usually used to determine the asymptotics of the coefficients of these functions, which immediately gives the asymptotic number of the studied structures.

In this section we review the definitions and results related to these topics that are essential in our proofs in Chapter 6. Our main reference is the ground-breaking book *Analytic Combinatorics* by Flajolet and Sedgewick [80].

2.3.1 The symbolic method

The symbolic method is an elegant and powerful tool to systematically translate set-theoretic relations between combinatorial classes into operations of generating functions. In this subsection we summarise the basic rules for such operations.

A *combinatorial class* is a set $\mathcal{A}$ of objects equipped with a *size function* $|\cdot| : \mathcal{A} \to \mathbb{N}_0$ such that $\big|\{\alpha \in \mathcal{A} : |\alpha| = n\}\big|$ is finite for every $n \in \mathbb{N}$. A combinatorial class is called *labelled* if all objects in the class are labelled, i.e. if each object α is equipped with a bijective function from $\big[|\alpha|\big]$ to the atoms of α that are counted by $|\cdot|$.

Depending on whether the class is labelled, there are two types of generating functions, namely *ordinary generating functions* for combinatorial classes of unlabelled objects, and *exponential generating functions* for labelled combinatorial classes. In this thesis we will solely encounter the latter type since we restrict our enumerative study to labelled objects. The exponential generating function (or EGF for short) of a sequence $(a_n)_{n\geq 1}$ is the formal power series

$$A(x) = \sum_{n\geq 0} a_n \frac{x^n}{n!}.$$

Let $\mathcal{A}$ be a labelled combinatorial class with a size function $|\cdot|$ and let $a_n := \big|\{\alpha \in \mathcal{A} : |\alpha| = n\}\big|$ denote the number of elements in $\mathcal{A}$ of size n. Then the EGF of $\mathcal{A}$ is the EGF of the sequence $(a_n)_{n\geq 1}$, that is,

$$A(x) = \sum_{n\geq 0} a_n \frac{x^n}{n!} = \sum_{\alpha\in\mathcal{A}} \frac{x^{|\alpha|}}{|\alpha|!}.$$

Conversely, we write $[x^n]A(x)$ to denote the n-th coefficient $a_n/n!$ of $A(x)$.

There are two basic combinatorial classes, namely the *neutral class* $\mathcal{E}$, which consists of a single object of size 0, and the *atomic class* $\mathcal{Z}$, which contains a single object of size 1. Their generating functions are hence $E(x) = 1$ and $Z(x) = x$, respectively. Starting with the neutral and the atomic class one can define, sometimes recursively, more complex classes in terms of already defined ones by using various set-theoretic constructions. As we have already mentioned before, these constructions translate into relations of the corresponding generating functions. In the following we describe the essential basic operations on labelled combinatorial classes that we will use. Afterwards we summarise the corresponding operations on their associated EGFs.

We always assume labelled objects to be *well-labelled*, that is, each object of size n receives n distinct labels from the set $[n]$. If an object α is not well-labelled, we write $\rho(\alpha)$ for the well-labelled object that we obtain from relabelling α in a canonical way, i.e. by preserving the order of the labels.

Let $\mathcal{A}$ and $\mathcal{B}$ be labelled combinatorial classes with size functions $|\cdot|_A$ and $|\cdot|_B$, respectively. If $\mathcal{A}\cap\mathcal{B} = \varnothing$, the *disjoint union* $\mathcal{A}+\mathcal{B}$ is defined as $\mathcal{A}\cup\mathcal{B}$ with the size function

$$|\alpha|_{\mathcal{A}+\mathcal{B}} := \begin{cases} |\alpha|_{\mathcal{A}} & \text{if } \alpha\in\mathcal{A}, \\ |\alpha|_{\mathcal{B}} & \text{if } \alpha\in\mathcal{B}. \end{cases}$$

The *(labelled) product* $\mathcal{C} = \mathcal{A} * \mathcal{B}$ is the class defined on the ground set $\bigcup_{\alpha\in\mathcal{A},\beta\in\mathcal{B}} \alpha * \beta$, where

$$\alpha * \beta := \{(\alpha', \beta') : (\alpha', \beta') \text{ is well-labelled, } \rho(\alpha') = \alpha \text{ and } \rho(\beta') = \beta\},$$

equipped with the size function

$$|\alpha * \beta|_C := |\alpha|_{\mathcal{A}} + |\beta|_{\mathcal{B}}.$$

For every $n \in \mathbb{N}$ let $\mathrm{SEQ}_n(\mathcal{A}) = \mathcal{A} * \cdots * \mathcal{A}$ with n factors of $\mathcal{A}$. The *(labelled) sequence* of $\mathcal{A}$ is denoted by $\mathrm{SEQ}(\mathcal{A})$ and is defined as

$$\mathrm{SEQ}(\mathcal{A}) = \mathcal{E} + \sum_{n\geq 1} \mathrm{SEQ}_n(\mathcal{A}).$$

Furthermore, for every $n \in \mathbb{N}$ we let $\mathrm{SET}_n(\mathcal{A})$ denote the collection of unordered sets in $\mathcal{A}$ with n elements and define the *(labelled) set* of $\mathcal{A}$ as

$$\mathrm{SET}(\mathcal{A}) = \mathcal{E} + \sum_{n\geq 1} \mathrm{SET}_n(\mathcal{A}).$$

The size of an element $(\alpha_1, \ldots, \alpha_n)$ in $\mathrm{SEQ}_n(\mathcal{A})$, $\mathrm{SEQ}(\mathcal{A})$, $\mathrm{SET}_n(\mathcal{A})$, or $\mathrm{SET}(\mathcal{A})$ is the sum of the sizes of $\alpha_1, \ldots, \alpha_n$. We note that in the case of sequences and sets of a class $\mathcal{A}$ one needs that $a_0 = 0$ to ensure that there are only finitely many objects of a given size. The *pointing operator* $\mathcal{A}^\bullet$ of the class $\mathcal{A}$ distinguishes for each element α of size n one of the n atoms that compounds α. Finally, the *(labelled) substitution* of $\mathcal{B}$ into $\mathcal{A}$, denoted by $\mathcal{A} \circ \mathcal{B}$, consists of all elements that are obtained by substituting each atom of every element of $\mathcal{A}$ by an element of $\mathcal{B}$, and by relabelling the resulting objects in a canonical way such that they become well-labelled.

In Table 2.1 we summarise all the constructions explained above and the corresponding operations of their associated EGFs, whose proofs can be found in e.g. [80].

Construction	Notation	EGF
Disjoint union	$\mathcal{A} + \mathcal{B}$	$A(x) + B(x)$
Product	$\mathcal{A} * \mathcal{B}$	$A(x)B(x)$
n-sequence	$\mathrm{SEQ}_n(\mathcal{A})$	$A(x)^n$
Sequence	$\mathrm{SEQ}(\mathcal{A})$	$\frac{1}{1-A(x)}$
n-set	$\mathrm{SET}_n(\mathcal{A})$	$\frac{A(x)^n}{n!}$
Set	$\mathrm{SET}(\mathcal{A})$	$\exp\big(A(x)\big)$
Pointing	$\mathcal{A}^\bullet$	$x\frac{d}{dx}A(x)$
Substitution	$\mathcal{A} \circ \mathcal{B}$	$A\big(B(x)\big)$

Table 2.1: Basic set-theoretic operations on labelled combinatorial classes and the corresponding algebraic relations of their associated EGFs.

Let $\mathcal{A}$ be a labelled combinatorial class equipped with a size function $|\cdot|$. A function $\xi : \mathcal{A} \to \mathbb{N}_0$ is called a *parameter*. Let $a_{n,k} := \big|\{\alpha \in \mathcal{A} : |\alpha| = n,\, \xi(\alpha) = k\}\big|$. The *bivariate exponential generating function* associated with $\mathcal{A}$ is defined as

$$A(x, y) = \sum_{n,m\geq 0} a_{n,m} \frac{x^n}{n!} y^m.$$

We say that x *marks* the size function $|\cdot|$ and y marks the parameter ξ. Setting $y = 1$ reduces $A(x, y)$ to the EGF of $\mathcal{A}$. The constructions and rules illustrated in Table 2.1 extend analogously to bivariate EGFs (see e.g. [80]).

2.3.2 Graph decompositions

While in the latter subsection we have collected rules for translating constructions of combinatorial classes into relations of generating functions, we concentrate in this subsection on the decompositions of specific classes of graphs and on the application of the symbolic method in these settings. In view of Chapter 6 we are interested in series and parallel networks, in a dissymmetry theorem for tree-decomposable classes, and in Tutte's decomposition for decomposing 2-connected graphs into 3-connected components.

Let $\mathcal{C}$ be a class of connected graphs with the property that a graph is in $\mathcal{C}$ if and only if all its 2-connected and 3-connected components are in $\mathcal{C}$. Observe that for instance the class of connected series-parallel graphs carrying a distinguished spanning tree shares this property. Let $c_{n,m}$ denote the number of graphs in $\mathcal{C}$ with n vertices and m edges. As explained in the latter subsection, the associated (bivariate) EGF is the formal power series

$$C(x,y) = \sum_{m,n\geq 0} c_{n,m} \frac{x^n}{n!} y^m,$$

where x and y mark vertices and edges, respectively.

Similarly, let $b_{n,m}$ denote the number of 2-connected graphs in $\mathcal{C}$ with n vertices and m edges and let $B(x,y)$ be its associated EGF. A connected graph rooted at a vertex can be obtained from a set of rooted 2-connected graphs, where the root is not labelled and where every other vertex is substituted by a connected graph rooted at a vertex. Using the symbolic method, this translates into the following relation between $C(x,y)$ and $B(x,y)$ (see also [83]):

$$xC_x(x,y) = x \exp\big(B_x\left(xC_x(x,y), y\right)\big). \tag{2.1}$$

Following Walsh [154], a *network* is defined as a simple graph with two distinguished vertices, that are called 0-pole and ∞-pole and do not bear a label, such that adding an edge between the two poles gives rise to a 2-connected multigraph. If there is an edge joining the two poles, it is called *root edge*. Let $D(x,y)$ denote the EGF associated with networks. The following equation that was shown by Walsh [154] reflects the relation between $B(x,y)$ and $D(x,y)$:

$$2(1+y)B_y(x,y) = x^2(D(x,y)+1). \tag{2.2}$$

The left-hand side in Equation (2.2) corresponds to the family of 2-connected graphs rooted at a directed edge that might not be present in the graph and the right-hand side corresponds to the family of networks, that also includes the empty network, where in addition labels have been assigned to the two poles.

A *trivial* network consists of the two poles and of the root edge. Following the ideas of [152], we further distinguish between three types of networks, namely *series*, *parallel*, and *h-networks* as follows. A series network S can be obtained from a directed cycle with a distinguished edge (which defines the two poles of the network) by replacing every other edge by a network, and finally by removing the distinguished edge. A parallel network P arises from merging at least two non-trivial networks, the root edge of each of them being not present, at their common poles. In this family, the root edge joining the two poles of P might not be present in P. Finally, an h-network is obtained from a 3-connected graph H rooted at an oriented edge by replacing every edge of H apart from the root edge by a network. As in the parallel case the root edge might not be present in an h-network.

Trakhtenbrot [152] showed that a network is either trivial, series, parallel, or an h-network, and Walsh [154] translated this decomposition into counting formulas. In series-parallel graphs the set of h-networks is empty. Therefore, we restrict our attention to series and parallel networks.

Let us mention that our definition of networks slightly differs from Trakhtenbrot's. Indeed, in [152] series networks could contain the root edge. In our work, series networks containing the root edge (in Trakhtenbrot's sense) are always considered to be parallel. This convention turns out to be helpful when dealing with spanning trees.

In Chapter 6 we aim for an asymptotic estimate for the number of spanning trees in graphs chosen uniformly at random from the family of connected series-parallel graphs. For this purpose, we will enumerate the class of connected series-parallel graphs with a distinguished spanning tree. The main idea is to give a complete analytic analysis of the generating function associated with this class using the relations to the class of 2-connected series-parallel graphs and to the class of networks, both carrying a distinguished spanning tree. Using Equation (2.2) would imply integration steps that are known to get difficult when considering enriched classes of graphs. Fortunately, Chapuy, Fusy, Kang, and Shoilekova [50] found a convenient combinatorial trick to forgo this integration step by using the dissymmetry theorem for tree-decomposable classes (Theorem 2.12) and by using the grammar for decomposing graphs into 3-connected components that they developed in [50].

A class $\mathcal{A}$ of graphs is *tree-decomposable* if for each graph $G \in \mathcal{A}$ we can define a tree $\tau(G)$ associated with G. Let $\mathcal{A}_\circ$ denote the class of graphs G in $\mathcal{A}$ where $\tau(G)$ has a distinguished vertex. Similarly, denote by $\mathcal{A}_{\circ-\circ}$ the class of all graphs G in $\mathcal{A}$ where $\tau(G)$ carries a distinguished edge. Finally, let $\mathcal{A}_{\circ\to\circ}$ be the class of all graphs G in $\mathcal{A}$ where an edge of $\tau(G)$ is directed. The dissymmetry theorem for trees by Bergeron [25] allows to express the class of unrooted trees in terms of classes of trees with a distinguished vertex, edge or with a directed edge. This theorem can be extended to tree-decomposable classes in the following way (see e.g. [50]).

Theorem 2.12 (Dissymmetry theorem for tree-decomposable classes)**.** *Let $\mathcal{A}$ be a class of graphs that is tree-decomposable. Then,*

$$\mathcal{A} + \mathcal{A}_{\circ\to\circ} \simeq \mathcal{A}_\circ + \mathcal{A}_{\circ-\circ}.$$

Finally, let us briefly summarize Tutte's decomposition [153] for decomposing 2-connected graphs into 3-connected components. For a thorough exposition we refer to [50].

Tutte's decomposition is based on split operations and the structure obtained from this process is shown to be independent of the order of the operations. Informally speaking, in every split operation we split the edge set of a graph G into two edge sets E_1 and E_2 that only coincide in exactly two vertices, say u and v, and where $G[E_1]$ is 2-connected and $G[E_2]$ is connected modulo $\{u, v\}$ (meaning that there exists no partition of E_2 into two nonempty sets E_2' and E_2'' such that $G[E_2']$ and $G[E_2'']$ only intersect in u and v). Next we add a so-called virtual edge e between these two vertices. Then we split the graph along this virtual edge, which yields two graphs G_1 and G_2 that correspond respectively to $G[E_1]$ and $G[E_2]$ with e now being a real edge. We say that G_1 and G_2 are *matched* by the virtual edge e.

The resulting structure is a collection of graphs that we call bricks. Tutte showed that there are only three types of bricks, namely *ring graphs* (*R-bricks*), *multi-edge graphs* (*M-bricks*), and 3-connected graphs with at least 4 vertices (T-bricks). The class of ring graphs is defined as the class of cyclic chains of at least 3 edges and the class of multi-edge graphs as the class of graphs with exactly two labelled vertices that are connected by at least 3 edges.

The *RMT-tree* of a graph G is defined as the graph $\tau(G)$ whose vertices are the bricks that result from Tutte's decomposition applied to G. Two vertices in $\tau(G)$ are connected when the corresponding bricks are matched by a virtual edge. It was shown by Tutte [153] that $\tau(G)$ is indeed a tree and that there are no two adjacent R-bricks nor two adjacent M-bricks.

Let $\mathcal{B}$ be the class of all 2-connected graphs with at least 3 vertices. We denote by $\mathcal{B}_R$, $\mathcal{B}_M$, and $\mathcal{B}_T$ the classes of graphs G in $\mathcal{B}$ such that the RMT-tree associated with G carries a distinguished R-vertex, M-vertex, and T-vertex, respectively. Moreover, let $\mathcal{B}_{R-M}$ denote the class of graphs G in $\mathcal{B}$ such that an edge between an R-vertex and an M-vertex in the RMT-tree associated with G is distinguished. The classes $\mathcal{B}_{R-T}$, $\mathcal{B}_{M-T}$, and $\mathcal{B}_{T-T}$ are defined analogously. Finally, let $\mathcal{B}_{T\to T}$ be the class of graphs G in $\mathcal{B}$ such that an edge between two T-vertices is directed.

Using Theorem 2.12, $\mathcal{B}$ satisfies the following equation (see [50]):

$$\mathcal{B} \simeq \mathcal{B}_R + \mathcal{B}_M + \mathcal{B}_T - \mathcal{B}_{R-M} - \mathcal{B}_{R-T} - \mathcal{B}_{M-T} - \mathcal{B}_{T\to T} + \mathcal{B}_{T-T}. \tag{2.3}$$

Since series-parallel graphs do not have 3-connected components, they do not contain T-bricks. Hence there are no T-vertices in the corresponding RMT-tree. Therefore, in the case of series-parallel graphs, Equation (2.3) is simplified to

$$\mathcal{B} \simeq \mathcal{B}_R + \mathcal{B}_M - \mathcal{B}_{R-M}. \tag{2.4}$$

2.3.3 Singularity analysis

In this subsection we introduce the necessary analytic background for the singularity analysis of generating functions that we will need in Chapter 6. In particular we state simplified versions of the transfer theorems and for the singularity analysis of systems of functional equations. For more details we refer to the books *Analytic Combinatorics* by Flajolet and Sedgewick [80] and *Random Trees* by Drmota [67].

Given an EGF

$$A(x) = \sum_{n\geq 0} a_n \frac{x^n}{n!}$$

we would like to determine an asymptotic estimate of the sequence $(a_n)_{n\geq 0}$. Pringsheim's theorem (see e.g. [80]) assures that generating functions with radius of convergence ϱ and non-negative Taylor coefficients have a singularity at ϱ, in particular a positive real dominant singularity. As shown in [80, Theorem IV.7], the exponential growth of the sequence $(a_n)_{n\geq 0}$ is therefore dictated by the smallest positive singularity ϱ of $A(x)$ in the sense that

$$[x^n]A(x) \sim \Psi(n)\varrho^{-n},$$

where $\Psi(n)$ grows subexponentially, i.e. $\limsup_{n\to\infty} |\Psi(n)|^{1/n} = 1$. The subexponential term $\Psi(n)$ results from the nature of this singularity. The so-called *transfer theorems*, developed by Flajolet and Odlyzko [79], provide a convenient way to determine the subexponential term of $[x^n]A(x)$. The following theorem is a special case of the transfer theorems stated in [80]. Before stating the result we first need the definition of dented domains (see Figure 2.1 for an illustration of a such a domain). Given $R, \zeta > 0$ with $R > \zeta$ and $0 < \phi < \pi/2$, the *domain dented* at ζ, which is denoted by $\Delta_\zeta(\phi, R)$, is defined as

$$\Delta_\zeta(\phi, R) = \{z \in \mathbb{C} : |z| < R,\, z \neq \zeta,\, |\operatorname{Arg}(z-\zeta)| > \phi\}.$$

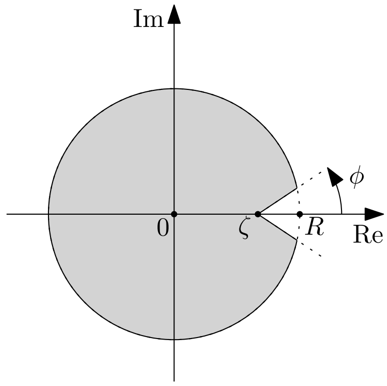

Figure 2.1: An illustration of the domain $\Delta_\zeta(\phi, R)$.

Theorem 2.13 (Transfer theorem [80], simplified version)**.** *Let $\alpha \in \mathbb{R} \setminus \mathbb{Z}^-$ and let $A(x)$ be analytic in a domain $\Delta_\varrho(\phi, R)$ dented at the smallest positive singularity ϱ of $A(x)$. If, as $x \to \varrho$ in $\Delta_\varrho(\phi, R)$,*

$$A(x) \sim c\left(1 - \frac{x}{\varrho}\right)^{-\alpha},$$

then

$$[x^n]A(x) = \frac{c}{\Gamma(\alpha)} n^{\alpha-1} \varrho^{-n}(1 + o(1)),$$

where $\Gamma(x)$ is the Euler Gamma function defined as $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$.

In this thesis the singular expansion of a generating function $A(x)$ in a domain dented at a singularity ϱ is always of the form

$$A(x) = A_0 + A_1X + A_2X^2 + \cdots + A_{2k+1}X^{2k+1} + \mathcal{O}(X^{2k+2}),$$

where $X = \sqrt{1 - x/\varrho}$. The even powers of X, being analytic functions, do not contribute to the asymptotic of $[x^n]A(x)$.

If $A_1 = A_3 = \cdots = A_{2k-1} = 0$ and $A_{2k+1} \neq 0$, then the number $(2k + 1)/2$ is called the *singular exponent*. For this situation Theorem 2.13 yields

$$[x^n]A(x) \sim \frac{c}{\Gamma(\alpha)} n^{\alpha-1} \varrho^{-n}$$

with $c = A_{2k+1}$ and $\alpha = -(2k + 1)/2$. When $A_1 \neq 0$ we say that $A(x)$ has a square-root expansion.

Let us now turn to the asymptotic analysis of systems of functional equations. The main reference for this topic is the paper [66] by Drmota. Assume that $y_1(x), \ldots, y_k(x)$ are generating functions satisfying a functional system of equations. We define $\mathbf{y} = (y_1(x), \ldots, y_k(x))$, and the system satisfied by $\mathbf{y}$ is denoted by $\mathbf{y} = \mathbf{F}(x; \mathbf{y})$, where $\mathbf{F} = (F_1, \ldots, F_k)$ is a vector of functions. The *dependency graph* $G = (V, E)$ associated with the system $\mathbf{y} = \mathbf{F}(x; \mathbf{y})$ is an oriented graph whose vertex set is $V = \{y_1, \ldots, y_k\}$ and $\overrightarrow{y_iy_j}$ is in E if and only if $\frac{\partial F_i}{\partial y_j} \neq 0$. The latter condition indicates that there is a real dependence between F_i and y_j. A dependency graph is said to be *strongly connected* if every pair of vertices can be linked by a directed path. Using this terminology we state the following shortened version of [67, Theorem 2.33].

Theorem 2.14 (Systems of functional equations [67], simplified version)**.** *Let the functional system of equations* $\boldsymbol{y} = \boldsymbol{F}(x;\boldsymbol{y})$ *where each* y_i *is analytic at* $x = 0$ *be given. Additionally, we require that each component of* $\boldsymbol{F}$ *is an entire function with positive Taylor coefficients, is not linear in the components* y_i*, and depends on* x*. Finally, we assume that* $\boldsymbol{F}(0;\boldsymbol{y}) = 0$ *and* $\boldsymbol{F}(x;\boldsymbol{0}) \neq 0$*. Assume also that the associated dependency graph is strongly connected. Denote by* $\boldsymbol{I}_k$ *the* $k \times k$ *identity matrix and by* $\mathrm{Jac}(\boldsymbol{F})$ *the Jacobian matrix associated with* $\boldsymbol{F}$ *and with respect to variables* $y_1, \dots, y_k$*. Assume that the system of equations*

$$\boldsymbol{y} = \boldsymbol{F}(x;\boldsymbol{y}), \qquad 0 = \det\left(\boldsymbol{I}_k - \mathrm{Jac}(\boldsymbol{F})\right)$$

has a unique solution $(x_0, \boldsymbol{y}_0)$ *in the region of analyticity of each component of* $\boldsymbol{F}$*. Then there is a unique solution* $\boldsymbol{y}$ *of the initial system of equations such that the components of* $\boldsymbol{y}$ *have non-negative Taylor coefficients and a square-root expansion in a domain dented at* $x = x_0$*.*

In order to obtain asymptotic estimates we need to assure that the dominant singularity is unique in a dented domain. This condition is usually satisfied whenever the counting formula $A(x)$ under consideration cannot be written in the form $A(x) = x^k f(x^r)$ for non-negative values $k \geq 0$ and $r \geq 2$. More precisely, we say that a generating function $A(x)$ is *aperiodic* if there exists a non-negative integer n_0 such that $[x^n]A(x) > 0$ for $n \geq n_0$. Observe that checking the aperiodicity condition is straightforward whenever we know that for each number of vertices there exist graphs in the family under study. The generating functions we consider in Chapter 6 (which are defined by an implicit equation, or by means of Theorem 2.14) will satisfy the aperiodicity condition by obvious combinatorial reasons. This will imply uniqueness of the dominant singularity. See [67] for more details.

2.4 Concentration inequalities of random variables

In our proofs we use the following standard bounds on deviations of random variables. Their proofs can be found in e.g. [98]. We start with Markov's inequality, which can be applied to any random variable that is almost surely positive.

Lemma 2.15 (Markov's inequality)**.** *For every random variable* X *with* $X \geq 0$ *almost surely and* $t > 0$ *we have*

$$\mathbb{P}\big[X \geq t\big] \leq \frac{\mathbb{E}[X]}{t}.$$

Next we consider random variables that follow a binomial distribution. The following two bounds belong to Chernoff's inequality, which collects different exponentially decreasing bounds on the tails of this distribution.

Theorem 2.16 (Chernoff's inequality I)**.** *For every random variable* $X \sim \mathrm{Bin}(n,p)$ *and every* $\varepsilon \leq 3/2$ *we have*

$$\mathbb{P}\big[|X - \mathbb{E}[X]| > \varepsilon\mathbb{E}[X]\big] < 2\exp\left(-\frac{\varepsilon^2\mathbb{E}[X]}{3}\right).$$

The second Chernoff's inequality that we need provides only a bound on the upper tail of the binomial distribution.

Theorem 2.17 (Chernoff's inequality II)**.** *For every random variable $X \sim \mathrm{Bin}(n,p)$ and every $t \geq 0$ we have*

$$\mathbb{P}\big[X \geq \mathbb{E}[X] + t\big] \leq \exp\left(-\frac{t^2}{2(\mathbb{E}[X] + t/3)}\right).$$

Now we turn to random variables that follow a hypergeometric distribution. Let N, m, and s be positive integers and let S and $S' \subseteq S$ be two sets with $|S| = N$ and $|S'| = m$. The *hypergeometric distribution* is the distribution of the random variable X that is defined by drawing s elements of S without replacement and counting how many of them belong to S'. It can be shown that Theorem 2.16 still holds in the case of hypergeometric distributions (see again e.g. [98]) with $\mathbb{E}[X] := ms/N$.

Theorem 2.18 (Hypergeometric inequality)**.** *Let X be a random variable that follows the hypergeometric distribution with parameters N, m, and s. Then for any $\varepsilon > 0$ and $t \geq \varepsilon ms/N$ we have*

$$\mathbb{P}\big[|X - ms/N| > t\big] < 2\exp\left(-\frac{\varepsilon^2 t}{3}\right).$$

We require the following technical lemma, which is a consequence of the hypergeometric inequality.

Lemma 2.19. *For each $\eta > 0$ and $\Delta \geq 1$ there exists $C > 0$ such that the following holds. Let $W \subseteq [n]$, let $t \leq 100n^{\Delta}$, and let $T_1, \dots, T_t$ be subsets of W. For each $m \leq |W|$ there is a set $S \subseteq W$ of size m such that*

$$|T_i \cap S| = \tfrac{m}{|W|}|T_i| \pm \big(\eta|T_i| + C\log n\big) \text{ for every } i \in [t].$$

Proof. Set $C = 30\eta^{-2}\Delta$. Let S consist of m elements of W that are drawn randomly from W without replacement. Observe that for each $i \in [t]$, the size of $T_i \cap S$ is hypergeometrically distributed. By Theorem 2.18, for each $i \in [t]$ we have

$$\mathbb{P}\Big[|T_i \cap S| \neq \tfrac{m}{|W|}|T_i| \pm \big(\eta|T_i| + C\log n\big)\Big] < 2\exp\big(-\eta^2 C\log n/3\big) < \frac{2}{n^{1+\Delta}}.$$

Taking the union bound over all $i \in [t]$ we conclude that the probability of failure is at most $2t/n^{1+\Delta} \leq 200/n$, which tends to 0 with n tending to infinity. Hence there exists a set $S \subseteq W$ of size m with the desired property. □

Finally we consider binomial random subsets. For $\Gamma = [n]$ let $\Gamma_{p_1,\dots,p_n}$ be defined by including for every $i \in [n]$ the i-th element of Γ with probability p_i independently of all other elements of Γ. For each set $S \subseteq 2^{\Gamma}$ of subsets of Γ and each set $A \in S$, we let X_A denote the indicator variable for the event $A \subseteq \Gamma_{p_1,\dots,p_n}$. Janson's inequality gives an exponentially small bound on the lower tail of the distribution of sums of such indicator variables.

Theorem 2.20 (Janson's inequality)**.** *Let Γ be a finite set and let $S \subseteq 2^{\Gamma}$ be a set of subsets of Γ. If $X = \sum_{A\in\mathcal{S}} X_A$, where X_A is an indicator variable as just defined, and $0 \leq t \leq \mathbb{E}[X]$, then*

$$\mathbb{P}\big[X \leq \mathbb{E}[X] - t\big] \leq \exp\left(-\frac{t^2}{2\overline{\Delta}}\right),$$

where

$$\overline{\Delta} = \mathbb{E}[X] + \sum_{A\in\mathcal{S}} \sum_{\substack{B\in\mathcal{S}:\\ A\cap B\neq\varnothing, A\neq B}} \mathbb{E}[X_A X_B].$$

3

The bandwidth theorem in random and pseudorandom graphs

In this chapter we study sparse random and pseudorandom graphs that contain (asymptotically almost surely) every graph of a particular class in a robust manner. The subgraphs that we are interested in are spanning and by *robust* we mean that the graph still contains the desired subgraph even after an adversary has deleted up to a certain proportion of the incident edges at every vertex. This concept is also known under the name *local resilience* (see Subsection 1.2.1 for an introduction to this area).

The bandwidth theorem by Böttcher, Schacht, and Taraz [41] states that any graph on n vertices with minimum degree at least $\big((k-1)/k+o(1)\big)n$ contains every k-colourable graph on n vertices with bounded maximum degree and sublinear bandwidth.

We prove an analogous version of this statement for random graphs. More precisely, we show that for every real $\gamma > 0$ and integers $\Delta \geq 2$ and $k \geq 1$, there exist constants $\beta > 0$ and $C > 0$ such that the following holds asymptotically almost surely for $G(n,p)$ whenever $p \geq C(\log n/n)^{1/\Delta}$. If G is any subgraph of $G(n,p)$ with minimum degree $\big((k-1)/k+\gamma\big)pn$, then G contains every n-vertex graph H with maximum degree at most Δ, bandwidth at most βn and with at least $C\max\{p^{-2}, p^{-1}\log n\}$ vertices not contained in any triangles of H.

The additional restriction that the graph H needs to have vertices not contained in any triangles is necessary in the sparse setting. We prove this theorem and a slightly stronger one where H is allowed to have a few vertices coloured with an additional colour in Section 3.1. The proof is based on the regularity method and uses the sparse blow-up lemma, proved by Allen, Böttcher, Hàn, Kohayakawa, and Person [9].

If in addition to the requirements of the theorem for random graphs mentioned above, H is also D-degenerate and does not contain any cycles of length four, we can prove a variant where a lower bound of $p = \mathcal{O}\big((\log n/n)^{1/(2D+1)}\big)$ suffices. The proof is similar to the one of the theorem mentioned above and uses a version of the sparse blow-up lemma to embed degenerate graphs into sparse random graphs [9]. We state and discuss the modifications that need to be carried out to obtain this result in Section 3.2.

We would like to emphasize that the resilience results mentioned above apply to 'pseudorandom' graphs in some sense, but not to any of the common, standard pseudorandomness conditions. However, we are also able to prove a similar resilience statement with respect to pseudorandom graphs, where we use the notion of bijumbledness. The proof of this result uses a sparse blow-up lemma for bijumbled graphs [9] and is presented in Section 3.3.

Finally, in Section 3.4 we remark on the optimality of our results and discuss some open questions and related work.

As mentioned before, the results of this chapter are joint work with Peter Allen, Julia Böttcher, and Anusch Taraz [7, 8].

3.1 The bandwidth theorem in random graphs

The goal of this section is to prove the following theorem, which can be seen as a random graph analogue of the bandwidth theorem (Theorem 1.3) by Böttcher, Schacht, and Taraz [41].

Theorem 3.1. *For each $\gamma > 0$, $\Delta \geq 2$, and $k \geq 1$, there exist $\beta^* > 0$ and $C^* > 0$ such that the following holds asymptotically almost surely for $\Gamma = G(n,p)$ if $p \geq C^*(\log n/n)^{1/\Delta}$.*

Let G be a spanning subgraph of Γ with $\delta(G) \geq \big((k-1)/k + \gamma\big)pn$ and let H be a k-colourable graph on n vertices with $\Delta(H) \leq \Delta$, bandwidth at most $\beta^* n$ and such that there are at least $C^* \max\{p^{-2}, p^{-1}\log n\}$ vertices in $V(H)$ that are not contained in any triangles of H. Then G contains a copy of H.

Theorem 3.1 is a corollary of a more general theorem (Theorem 3.3), where, subject to a few conditions, H is allowed to have vertices coloured with an additional colour. One of these requirements is that the proper $(k+1)$-colouring of the vertex set of H is *zero-free* with respect to the bandwidth labelling, where 'zero-free' is defined as follows.

Definition 3.2 (Zero-free colouring)**.** *Let H be a $(k+1)$-colourable graph on n vertices and let $\mathcal{L}$ be a labelling of its vertex set of bandwidth at most βn. A* block *of $\mathcal{L}$ is defined as a set of the form $\{(t-1)4k\beta n+1, \dots, t4k\beta n\}$ with some $t \in [1/(4k\beta)]$. A proper $(k+1)$-colouring $\sigma : V(H) \to \{0, \dots, k\}$ of its vertex set is said to be (z,β)-*zero-free *with respect to $\mathcal{L}$ if any z consecutive blocks contain at most one block with colour zero.*

With this definition in hand, we can now state Theorem 3.3.

Theorem 3.3. *For each $\gamma > 0$, $\Delta \geq 2$, and $k \geq 2$, there exist $\beta > 0$, $z > 0$, and $C > 0$ such that the following holds asymptotically almost surely for $\Gamma = G(n,p)$ if $p \geq C(\log n/n)^{1/\Delta}$.*

Let G be a spanning subgraph of Γ with $\delta(G) \geq \big((k-1)/k + \gamma\big)pn$ and let H be a graph on n vertices with $\Delta(H) \leq \Delta$ that has a labelling $\mathcal{L}$ of its vertex set of bandwidth at most βn, a $(k+1)$-colouring that is (z,β)-zero-free with respect to $\mathcal{L}$ and where the first $\sqrt{\beta}n$ vertices in $\mathcal{L}$ are not given colour zero and the first βn vertices in $\mathcal{L}$ include $C\max\{p^{-2}, p^{-1}\log n\}$ vertices that are not contained in any triangles of H. Then G contains a copy of H.

As already mentioned in Chapter 1, the minimum degree condition in Theorems 3.1 and 3.3 is tight and the bandwidth condition cannot be omitted. Moreover, the number of vertices in H that are contained in triangles needs to be restricted. However, if G is required to contain for every vertex $x \in V(G)$ a positive proportion of the copies of $K_{\Delta+1}$ containing x that occurred in $G(n,p)$, then this restriction is no longer needed. We briefly discuss this as well as the (non-)optimality of Theorem 3.1 in terms of the edge probability and of the number of vertices not in triangles in Section 3.4.

In Subsection 3.1.7 we first present the proof of Theorem 3.3 and then deduce Theorem 3.1. The proof of Theorem 3.3 uses five main lemmas, one of which is the sparse blow-up lemma (Theorem 2.9) by Allen, Böttcher, Hàn, Kohayakawa, and Person. The other four lemmas are formulated in the following subsection and are proved in Subsections 3.1.3–3.1.6.

3.1.1 Main lemmas and outline of the proof

Before we state the main lemmas that we use in the proof of Theorem 3.3 and outline roughly how they will be combined, we need to introduce some more definitions.

Let $r, k \geq 1$ and let B_r^k be the graph on kr vertices obtained from a path on r vertices by replacing every vertex by a clique of size k and by replacing every edge by a complete bipartite graph minus a perfect matching. More precisely, we define B_r^k as

$$V(B_r^k) := [r] \times [k]$$

and for all distinct $j, j' \in [k]$

$$\{(i,j),(i',j')\} \in E(B_r^k) \quad \text{if and only if} \quad i = i' \text{ or } |i - i'| = 1.$$

We call B_r^k *backbone graph* on the vertex set $[r] \times [k]$.

Let $K_r^k \subseteq B_r^k$ be the spanning subgraph of B_r^k that is the disjoint union of r complete graphs on k vertices given by the following components: the clique $K_r^k[\{(i,1),\dots,(i,k)\}]$ is the *i-th component* of K_r^k for each $i \in [r]$. See Figure 3.1 for an illustration of the clique factor K_r^k in a backbone graph B_r^k for $k = 3$.

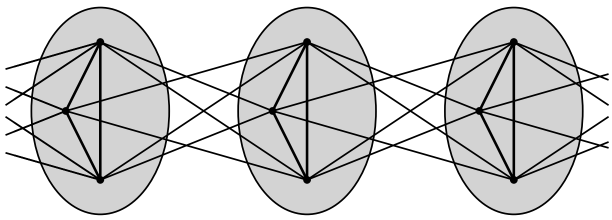

Figure 3.1: Three components of the clique factor K_r^3 in the backbone graph B_r^3.

A partition $\mathcal{V}' = \{V_{i,j}\}_{i\in[r],j\in[k]}$ is called *k-equitable* if $\big||V_{i,j}| - |V_{i,j'}|\big| \leq 1$ for every $i \in [r]$ and $j, j' \in [k]$. Similarly, an integer partition $\{n_{i,j}\}_{i\in[r],j\in[k]}$ of n (meaning that $n_{i,j} \in \mathbb{Z}_{\geq 0}$ for every $i \in [r], j \in [k]$ and $\sum_{i\in[r]j\in[k]} n_{i,j} = n$) is *$k$-equitable* if $|n_{i,j} - n_{i,j'}| \leq 1$ for every $i \in [r]$ and $j, j' \in [k]$.

Now we are ready to sketch the idea of the proof of Theorem 3.3 and state the main lemmas that we need. The goal of the proof is to find a.a.s. for every subgraph G of $G(n,p)$ with minimum degree at least $\big((k-1)/k + \gamma\big)pn$ a graph homomorphism from a given graph H, with the properties as in the statement of the theorem, to G.

The first main lemma (Lemma 3.4) is used to partition G. It states that a.a.s. $\Gamma = G(n,p)$ satisfies the following property if $p = \Omega(\log n/n)^{1/2}$. For any spanning subgraph $G \subseteq \Gamma$ with minimum degree a sufficiently large fraction of pn, there exists an $(\varepsilon, d, p)_G$-regular vertex partition $\mathcal{V}$ of $V(G)$ whose reduced graph R_r^k contains a clique factor K_r^k, on which the corresponding vertex sets of $\mathcal{V}$ are pairwise $(\varepsilon, d, p)_G$-super-regular. Furthermore, $(G, \mathcal{V})$ has one-sided and two-sided inheritance with respect to R_r^k, and the Γ-neighbourhoods of all vertices but the ones in the exceptional set of $\mathcal{V}$ have almost exactly their expected size in each cluster. The proof of Lemma 3.4 is given in Subsection 3.1.3.

Lemma 3.4 (Lemma for G)**.** *For each $\gamma > 0$, $k \geq 2$, and $r_0 \geq 1$ there exists $d > 0$ such that for every $\varepsilon \in \big(0, 1/(2k)\big)$ there exist $r_1 \geq 1$ and $C^* > 0$ such that the following holds a.a.s. for $\Gamma = G(n,p)$ if $p \geq C^* (\log n/n)^{1/2}$.*

Let $G = (V,E)$ be a spanning subgraph of Γ with $\delta(G) \geq \big((k-1)/k + \gamma\big)pn$. Then there exists an integer r with $r_0 \leq kr \leq r_1$, a subset $V_0 \subseteq V$ with $|V_0| \leq C^* \max\{p^{-2}, p^{-1}\log n\}$, a k-equitable vertex partition $\mathcal{V} = \{V_{i,j}\}_{i\in[r],j\in[k]}$ of $V(G) \setminus V_0$, and a graph R_r^k on the vertex set $[r] \times [k]$ with $K_r^k \subseteq B_r^k \subseteq R_r^k$, with $\delta(R_r^k) \geq \big((k-1)/k + \gamma/2\big)kr$, and such that the following is true:

- *(G1)* $\frac{n}{4kr} \leq |V_{i,j}| \leq \frac{4n}{kr}$ *for every $i \in [r]$ and $j \in [k]$,*
- *(G2) $\mathcal{V}$ is $(\varepsilon, d, p)_G$-regular on R_r^k and $(\varepsilon, d, p)_G$-super-regular on K_r^k,*
- *(G3) both $\big(N_\Gamma(v, V_{i,j}), V_{i',j'}\big)$ and $\big(N_\Gamma(v', V_{i,j}), N_\Gamma(v, V_{i',j'})\big)$ are $(\varepsilon, d, p)_G$-regular for every $\{(i,j),(i',j')\} \in E(R_r^k)$ and $v \in V \setminus V_0$,*
- *(G4) $|N_\Gamma(v, V_{i,j})| = (1 \pm \varepsilon)p|V_{i,j}|$ for every $i \in [r]$, $j \in [k]$ and every $v \in V \setminus V_0$.*

Furthermore, if we replace (G3) by

- *(G3') $\big(N_\Gamma(v, V_{i,j}), V_{i',j'}\big)$ is $(\varepsilon, d, p)_G$-regular for every $\{(i,j),(i',j')\} \in E(R_r^k)$ and every $v \in V \setminus V_0$,*

then we have the stronger bound $|V_0| \leq C^* p^{-1} \log n$.

After Lemma 3.4 has constructed a regular partition $\mathcal{V}$ of $V(G)$, the second main lemma deals with the graph H that we would like to find as a subgraph of G.

More precisely, Lemma 3.5 provides a homomorphism f from the graph H to the reduced graph R_r^k given by Lemma 3.4, which has among others the following properties. The edges of H are mapped to the edges of R_r^k, and the vast majority of the edges of H are assigned to edges of the clique factor $K_r^k \subseteq R_r^k$. The number of vertices of H mapped to a vertex of R_r^k only differs slightly from the size of the corresponding cluster of $\mathcal{V}$. The lemma further guarantees that each of the first $\sqrt{\beta}n$ vertices of the bandwidth ordering of $V(H)$ is mapped to $(1,j)$ with j being the colour that the vertex has received by the given colouring of H.

In case H is D-degenerate the next lemma also ensures that for every $(i,j) \in [r] \times [k]$, a constant fraction of vertices mapped to (i,j) have each at most $2D$ neighbours.

Lemma 3.5 (Lemma for H)**.** *Given $D, k, r \geq 1$ and $\xi, \beta > 0$ the following holds if $\xi \leq 1/(kr)$ and $\beta \leq 10^{-10}\xi^2/(Dk^4 r)$.*

Let H be a D-degenerate graph on n vertices, let $\mathcal{L}$ be a labelling of its vertex set of bandwidth at most βn and let $\sigma : V(H) \to \{0, \dots k\}$ be a proper $(k+1)$-colouring that is $(10/\xi, \beta)$-zero-free with respect to $\mathcal{L}$, where the colour zero does not appear in the first $\sqrt{\beta}n$ vertices of $\mathcal{L}$. Furthermore, let R_r^k be a graph on vertex set $[r] \times [k]$ with $K_r^k \subseteq B_r^k \subseteq R_r^k$ such that for every $i \in [r]$ there exists a vertex $z_i \in \big([r] \setminus \{i\}\big) \times [k]$ with $\big\{z_i, (i,j)\big\} \in E(R_r^k)$ for every $j \in [k]$.

Then, given a k-equitable integer partition $\{m_{i,j}\}_{i\in[r],j\in[k]}$ of n with $n/(10kr) \leq m_{i,j} \leq 10n/(kr)$ for every $i \in [r]$ and $j \in [k]$, there exists a mapping $f \colon V(H) \to [r] \times [k]$ and a set of special vertices $X \subseteq V(H)$ such that for every $i \in [r]$ and $j \in [k]$ we have

(H1) $m_{i,j} - \xi n \leq |f^{-1}(i,j)| \leq m_{i,j} + \xi n$,

(H2) $|X| \leq \xi n$,

(H3) $\{f(x), f(y)\} \in E(R_r^k)$ *for every* $\{x,y\} \in E(H)$,

(H4) $y, z \in \bigcup_{j' \in [k]} f^{-1}(i,j')$ *for every* $x \in f^{-1}(i,j) \setminus X$ *and* $\{x,y\},\{y,z\} \in E(H)$,

(H5) $f(x) = \big(1, \sigma(x)\big)$ *for all vertices* x *in the first* $\sqrt{\beta}n$ *vertices of* $\mathcal{L}$, *and*

(H6) $|\{x \in f^{-1}(i,j) : \deg(x) \leq 2D\}| \geq \frac{1}{24D}|f^{-1}(i,j)|$.

Lemma 3.5 is a strengthened version of [42, Lemma 8]. The proof of [42, Lemma 8] is deterministic; here we use a probabilistic argument to show the existence of a function f that also satisfies the additional property (H6), which is required for the proof of Theorem 3.15, that we give in Section 3.2. However, we still borrow ideas from the proof of [42, Lemma 8]. The proof of Lemma 3.5 is presented in Subsection 3.1.4.

So far, the vertices of the exceptional set V_0 of the regular partition $\mathcal{V}$ of $V(G)$ were disregarded. To cover them, we need to manually pre-embed vertices of H onto all vertices in V_0. For this, we use vertices in H that are not in triangles, that are pairwise far apart from each other and that are contained in the first βn vertices of the bandwidth ordering $\mathcal{L}$ of $V(H)$. Once we embedded a vertex x of H onto a vertex v of V_0, we also embed its neighbours $N_H(x)$. This creates restrictions on the vertices of G to which we can embed the second neighbours, and for the application of the sparse blow-up lemma (Theorem 2.9) we need certain conditions to be satisfied. The next lemma states that we can find vertices in $N_G(v)$ satisfying these conditions with room to spare. We prove Lemma 3.6 in Subsection 3.1.5.

Lemma 3.6 (Common neighbourhood lemma). *For each* $d > 0$, $k \geq 2$, *and* $\Delta \geq 2$ *there exists* $\alpha > 0$ *such that for every* $\varepsilon^* \in (0,1)$ *there exists* $\varepsilon_0 > 0$ *such that for every* $r \geq 1$ *and every* $0 < \varepsilon \leq \varepsilon_0$ *there exists* $C^* > 0$ *such that* $\Gamma = G(n,p)$ *a.a.s. satisfies the following if* $p \geq C^* (\log n/n)^{1/\Delta}$.

Let $G = (V,E)$ *be a (not necessarily spanning) subgraph of* Γ *and* $\{V_i \setminus W\}_{i \in [k]} \cup \{W\}$ *a vertex partition of a subset of* V *such that the following is true for every distinct* $i, i' \in [k]$:

(V1) $\frac{n}{4kr} \leq |V_i| \leq \frac{4n}{kr}$,

(V2) $(V_i, V_{i'})$ *is* $(\varepsilon, d, p)_G$*-regular,*

(V3) $|W| = 10^{-10}\frac{\varepsilon^4 pn}{k^4 r^4}$, *and*

(V4) $|N_G(w, V_i)| \geq dp|V_i|$ *for every* $w \in W$.

Then there exists a tuple $(w_1, \ldots, w_\Delta) \in \binom{W}{\Delta}$ *such that for every* $\Lambda, \Lambda^* \subseteq [\Delta]$, *and every distinct* $i, i' \in [k]$ *we have*

(W1) $|\bigcap_{j \in \Lambda} N_G(w_j, V_i)| \geq \alpha p^{|\Lambda|}|V_i|$,

(W2) $|\bigcap_{j \in \Lambda} N_\Gamma(w_j)| \leq (1+\varepsilon^*)p^{|\Lambda|}n$,

(W3) $|\bigcap_{j \in \Lambda} N_\Gamma(w_j, V_i)| = (1 \pm \varepsilon^*)p^{|\Lambda|}|V_i|$, *and*

(W4) $\big(\bigcap_{j \in \Lambda} N_\Gamma(w_j, V_i), \bigcap_{j^* \in \Lambda^*} N_\Gamma(w_{j^*}, V_{i'})\big)$ *is* $(\varepsilon^*, d, p)_G$*-regular if* $|\Lambda|, |\Lambda^*| < \Delta$ *and either* $\Lambda \cap \Lambda^* = \varnothing$ *or* $\Delta \geq 3$ *or both.*

Let H' and G' denote the subgraphs of H and G that result from removing all vertices that were used in the pre-embedding process. As a last step before finally applying the sparse blow-up lemma, the clusters in $\mathcal{V}\big|_{G'}$ need to be adjusted to the sizes of $W_{i,j}\big|_{H'}$. The next lemma states that this is possible, and that after this redistribution the regularity properties that we need for the sparse blow-up lemma (Theorem 2.9) hold. The proof of Lemma 3.7 is given in Subsection 3.1.6.

Lemma 3.7 (Balancing lemma). *For all integers $k, r_1, \Delta \geq 1$, and reals $\gamma, d > 0$ and $0 < \varepsilon < \min\{d, 1/(2k)\}$ there exist $0 < \xi < 1/(10kr_1)$ and $C^* > 0$ such that the following is true for every $p \geq C^* (\log n/n)^{1/2}$ and every $10\gamma^{-1} \leq r \leq r_1$ provided that n is large enough.*

Let Γ be a graph on the vertex set $[n]$ and let $G = (V, E) \subseteq \Gamma$ be a (not necessarily spanning) subgraph with vertex partition $\mathcal{V} = \{V_{i,j}\}_{i\in[r],j\in[k]}$ that satisfies $n/(8kr) \leq |V_{i,j}| \leq 4n/(kr)$ for each $i \in [r]$, $j \in [k]$. Let $\{n_{i,j}\}_{i\in[r],j\in[k]}$ be an integer partition of $\sum_{i\in[r],j\in[k]} |V_{i,j}|$. Let R^k_r be a graph on the vertex set $[r] \times [k]$ with minimum degree $\delta(R^k_r) \geq \big((k-1)/k + \gamma/2\big)kr$ such that $K^k_r \subseteq B^k_r \subseteq R^k_r$. Suppose that the partition $\mathcal{V}$ satisfies the following properties for each $i \in [r]$, all distinct $j, j' \in [k]$, and each $v \in V$:

- *(B1)* $n_{i,j} - \xi n \leq |V_{i,j}| \leq n_{i,j} + \xi n$,
- *(B2)* $\mathcal{V}$ *is* $\big(\frac{\varepsilon}{4}, d, p\big)_G$*-regular on* R^k_r *and* $\big(\frac{\varepsilon}{4}, d, p\big)_G$*-super-regular on* K^k_r,
- *(B3) both* $\big(N_\Gamma(v, V_{i,j}), V_{i,j'}\big)$ *and* $\big(N_\Gamma(v, V_{i,j}), N_\Gamma(v, V_{i,j'})\big)$ *are* $\big(\frac{\varepsilon}{4}, d, p\big)_G$*-regular pairs, and*
- *(B4)* $|N_\Gamma(v, V_{i,j})| = \big(1 \pm \frac{\varepsilon}{4}\big)p|V_{i,j}|$.

Then, there exists a partition $\mathcal{V}' = \{V'_{i,j}\}_{i\in[r],j\in[k]}$ of V such that the following properties hold for each $i \in [r]$, all distinct $j, j' \in [k]$, and each $v \in V$:

- *(B1')* $|V'_{i,j}| = n_{i,j}$,
- *(B2')* $|V_{i,j} \triangle V'_{i,j}| \leq 10^{-10}\varepsilon^4 k^{-2} r_1^{-2} n$,
- *(B3')* $\mathcal{V}'$ *is* $(\varepsilon, d, p)_G$*-regular on* R^k_r *and* $(\varepsilon, d, p)_G$*-super-regular on* K^k_r,
- *(B4') both* $\big(N_\Gamma(v, V'_{i,j}), V'_{i,j'}\big)$ *and* $\big(N_\Gamma(v, V'_{i,j}), N_\Gamma(v, V'_{i,j'})\big)$ *are* $(\varepsilon, d, p)_G$*-regular pairs, and*
- *(B5') for each $1 \leq s \leq \Delta$ and every collection of s vertices $v_1, \ldots, v_s \in [n]$ we have*

$$\big|N_\Gamma(v_1, \ldots, v_s; V_{i,j}) \triangle N_\Gamma(v_1, \ldots, v_s; V'_{i,j})\big| \leq 10^{-10}\varepsilon^4 k^{-2} r_1^{-2} \deg_\Gamma(v_1, \ldots, v_s) + C^* \log n\,.$$

Furthermore, if for any two disjoint vertex sets $A, A' \subseteq V(\Gamma)$ with $|A|, |A'| \geq \frac{1}{50000kr_1}\varepsilon^2\xi pn$ we have $e_\Gamma(A, A') \leq \big(1 + \frac{1}{100}\varepsilon^2\xi\big)p|A||A'|$, and if 'regular' is replaced with 'fully-regular' in (B2), and (B3), then we can replace 'regular' with 'fully-regular' in (B3') and (B4').

The last step of the proof of Theorem 3.3 is the application of the sparse blow-up lemma (Theorem 2.9) to the vertex partition of G' given by Lemma 3.7 and to the vertex partition of H given by Lemma 3.5 restricted to H' while respecting the image restrictions that resulted from the pre-embedding process.

Before we turn to the details of the proof of Theorem 3.3 in Subsection 3.1.7, we prove Lemmas 3.4–3.7 in Subsections 3.1.3–3.1.6. First of all we show in the next subsection an almost sure property of $G(n, p)$ that we need at numerous places in the proofs of this chapter.

3.1.2 Preliminaries

In this subsection we prove the following useful observation. Roughly speaking, it states that a.a.s. $G(n,p)$ fulfils that most of its vertices have approximately the expected number of neighbours within large subsets, the number of edges between disjoint large subsets is also concentrated around its mean, and the number of edges within large subsets is not significantly higher than what one expects.

Proposition 3.8. *For each $\varepsilon > 0$ there exists a constant $C > 0$ such that for every $0 < p < 1$ asymptotically almost surely $\Gamma = G(n,p)$ has the following properties. For any disjoint sets $X, Y \subseteq V(\Gamma)$ with $|X|, |Y| \geq Cp^{-1} \log n$, we have $e(X,Y) = (1 \pm \varepsilon)p|X||Y|$ and $e(X) \leq 2p|X|^2$. Furthermore, for every $Y \subseteq V(\Gamma)$ with $|Y| \geq Cp^{-1} \log n$, the number of vertices $v \in V(\Gamma)$ with $\big||N_\Gamma(v,Y)| - p|Y|\big| > \varepsilon p|Y|$ is at most $Cp^{-1} \log n$.*

Proof of Proposition 3.8. Since the statement of the proposition is stronger when ε is smaller, we may assume that $0 < \varepsilon \leq 1$. We set $C' = 100\varepsilon^{-2}$ and $C = 100C'\varepsilon^{-1}$.

We first show that $\Gamma = G(n,p)$ has a.a.s. the following two properties. For any disjoint sets $A, B \subseteq V(\Gamma)$, each of size at least $C'p^{-1} \log n$, we have $e(A,B) = (1 \pm \varepsilon/2)p|A||B|$. For any $A \subseteq V(\Gamma)$ of size at least $C'p^{-1} \log n$, we have $e(A) \leq 2p|A|^2$. Note that these two properties imply the first two conclusions of the proposition.

We estimate the failure probability of the first property by using Theorem 2.16 and applying the union bound. Assuming without loss of generality that $|A| \leq |B|$, this probability is at most

$$\sum_{|A|,|B| \leq n} n^{|A|+|B|} \cdot 2e^{-\varepsilon^2 p|A||B|/12} \leq 2n^{2+2|B|} e^{-\varepsilon^2 C'|B| \log n/12} < \sum_{|B|} 2n^{1+2|B|} n^{-4|B|} \leq 2n^{-2} .$$

Similarly, the failure probability of the second property is at most

$$\sum_{|A|} n^{|A|} \cdot 2e^{-p\binom{|A|}{2}/3} \leq \sum_{|A|} 2n^{|A|} e^{-C'|A| \log n/12} \leq 2n^{-2}.$$

We conclude that a.a.s. $G(n,p)$ enjoys both properties.

Now we condition on Γ having both of these properties. Let $Y \subseteq V(\Gamma)$ have size at least $Cp^{-1} \log n$. We first show that there are at most $C'p^{-1} \log n$ vertices in Γ that have less than $(1-\varepsilon)p|Y|$ neighbours in Y. If this was false, then we could choose a set X consisting of $C'p^{-1} \log n$ vertices in Γ where each of them has less than $(1-\varepsilon)p|Y|$ neighbours in Y. Since by choice of C we have $(1-\varepsilon)p|Y| \leq (1-\varepsilon/2)p|Y \setminus X|$, we see that $e(X, Y \setminus X) < (1-\varepsilon/2)p|X||Y \setminus X|$, which is a contradiction since $|Y \setminus X| \geq C'p^{-1} \log n$.

Next we show that there are at most $2C'p^{-1} \log n$ vertices of Γ each of them having more than $(1+\varepsilon)p|Y|$ neighbours in Y. Again, if this is not the case we can let X be a set of $2C'p^{-1} \log n$ vertices of Γ with more than $(1+\varepsilon)p|Y|$ neighbours in Y. Now $e(X) \leq 2p|X|^2 = 4C'|X| \log n$, so there are at most $|X|/2$ vertices in X that have $16C' \log n$ or more neighbours in X. Let $X' \subseteq X$ consist of those vertices with at most $16C' \log n$ neighbours in X. For each $v \in X'$ we have

$$(1+\varepsilon)p|Y| \leq \deg(v,Y) \leq \deg(v,X) + \deg(v, Y \setminus X) ,$$

and so, by choice of C, each vertex of X' has at least $(1+\varepsilon/2)p|Y \setminus X|$ neighbours in $Y \setminus X$. This is a contradiction since $|X'|, |Y \setminus X| \geq C'p^{-1} \log n$. Finally, since by choice of C we have $3C'p^{-1} \log n < Cp^{-1} \log n$, we conclude that all but at most $Cp^{-1} \log n$ vertices of Γ have $(1 \pm \varepsilon)p|Y|$ neighbours in Y, as desired. $\square$

3.1.3 The lemma for G

This subsection is devoted to the proof of the lemma for G (Lemma 3.4). We borrow ideas from the proof of [41, Proposition 17] and from the proof of [39, Lemma 9].

Our proof strategy can be summarised as follows. First we apply Lemma 2.6 to obtain an equitable partition of $V(G)$ within whose reduced graph we can find a backbone graph by Theorem 1.3. Then, starting with an empty set $Z_1 = \varnothing$, we add every vertex v to Z_1 for which there exists a cluster U such that the size of the Γ-neighbourhood of v in U is not close to $p|U|$, or for which the Γ-neihbourhood in the exceptional set is too large, or for which its Γ-neighbourhoods fail to inherit regularity. We also add a minimum number of extra vertices to maintain k-equitability and then remove temporarily the set Z_1 from the graph.

Now we consider each vertex that destroyed super-regularity on the clique factor of the backbone graph before the removal of Z_1. We redistribute these vertices as well as the vertices from the exceptional set of the partition to other clusters such that they do not destruct super-regularity anymore. The moving of the vertices may have destroyed some of the regularity inheritance, Γ-neighbourhood, and super-regularity properties we tried to obtain before. However, a vertex only witnesses failure of these properties if exceptionally many of its Γ-neighbours were moved from or to a cluster. We define Z_2 to be the set of all such vertices plus a minimum number of additional vertices to obtain k-equitability of the remaining partition, and remove Z_2 from the graph. We will see that Z_2 is so small that its removal does not significantly affect the desired properties. Hence we can set $V_0 = Z_1 \cup Z_2$ and have found a partition of $V(G)$ with the properties as demanded.

Proof of Lemma 3.4. First we fix the constants that we need in the proof. Given $\gamma > 0$, $k \geq 2$, and $r_0 \geq 1$, set $d = \gamma/32$. Let $\beta > 0$ and $n_0 \geq 1$ be returned by Theorem 1.3 for input $\gamma/2$, $3k$, and k. Let

$$r_0' = \max\left\{r_0, n_0, \tfrac{k}{d}, \tfrac{10k}{\beta}\right\}.$$

Next, let $\varepsilon_1 > 0$ and $C_1 > 0$, and $\varepsilon_2 > 0$ and $C_2 > 0$ be returned by Lemma 2.10 and Lemma 2.11, respectively, each with input $\varepsilon/2$ and d. Furthermore, let $C_3 > 0$ be returned by Proposition 3.8 with input ${\varepsilon^*}^2/(1000k^2)$. Given $\varepsilon \in \big(0, 1/(2k)\big]$, set

$$\varepsilon^* = \min\left\{\varepsilon_1, \varepsilon_2, \tfrac{\varepsilon^2\gamma}{10^{10}k^2}\right\}.$$

Then we apply Lemma 2.6 with input ε^*/k, $(k-1)/k + \gamma$, and $r_0' + k$ in order to obtain $r_1 > 0$. Finally, we set $C = \max_{i\in[3]}\{C_i\}$ and

$$C^* = \tfrac{100k^2 r_1^3 C}{\varepsilon^*}.$$

Given $p \geq C^*\big(\frac{\log n}{n}\big)^{1/2}$, it holds that $G(n,p)$ a.a.s. satisfies the good events of Lemmas 2.10 and 2.11, and of Proposition 3.8 with the parameters specified above. We condition on $\Gamma = G(n,p)$ satisfying these good events.

Given $G = (V,E) \subseteq \Gamma$ with $\delta(G) \geq \big(\frac{k-1}{k} + \gamma\big)pn$, we apply Lemma 2.6, with input ε^*/k, $\alpha = (k-1)/k + \gamma$, $r_0' + k$, and d, to G. Observe that this is possible because G is a subgraph of Γ, and we have $Cp^{-1}\log n \leq \varepsilon^* n/(kr_1)$, so that the condition of Lemma 2.6 is satisfied as we have conditioned on the good event of Proposition 3.8 for Γ. The result is an $(\varepsilon^*/k, d, p)$-regular partition $\mathcal{U}'$ of V with $t' \in [r_0' + k, r_1]$ equally sized clusters, with exceptional set U_0'

of size at most $\varepsilon^* n/k$ and whose reduced graph R has minimum degree at least

$$\delta(R) \geq \left(\tfrac{k-1}{k} + \gamma - d - \tfrac{1}{k}\varepsilon^*\right)t'.$$

We would like to work with a regular partition of V whose number of clusters is a multiple of k. For this purpose we move at most $k-1$ of the clusters of $\mathcal{U}'$ to U_0' in order to obtain a partition $\mathcal{U}$ of V with kr equally sized clusters, where $r \in \mathbb{N}$ and $r_0' \leq kr \leq t'$. By construction, $\mathcal{U}$ is an (ε^*, d, p)-regular partition with exceptional set U_0 of size at most $\varepsilon^* n$ and with a reduced graph R_r^k of minimum degree at least $\left(\frac{k-1}{k} + \gamma - d - \frac{1}{k}\varepsilon^*\right)kr - k$.

By the choice of d and ε^* as well as of r_0' we have

$$\delta(R_r^k) \geq \left(\tfrac{k-1}{k} + \gamma - d - \tfrac{1}{k}\varepsilon^*\right)kr - k \geq \left(\tfrac{k-1}{k} + \tfrac{\gamma}{2}\right)kr\,.$$

Observe that the bandwidth graph B_r^k on the vertex set $[r] \times [k]$ has bandwidth at most $2k < \beta r_0'$ and maximum degree less than $3k$. Moreover, note that $|V(B_r^k)| = kr \geq r_0' \geq n_0$ by choice of r_0'. Thus Theorem 1.3, with input $\gamma/2$, $3k$, and k, states in particular that R_r^k contains a copy of B_r^k. We fix one such copy and let its vertices $\{(i,j)\}_{i\in[r],j\in[k]}$ label the vertices of R_r^k. Similarly, for each $i \in [r]$ and $j \in [k]$, we denote the cluster of $\mathcal{U}$ corresponding to the vertex (i,j) of B_r^k by $U_{i,j}$. The partition $\mathcal{U} = \{U_{i,j}\}_{i\in[r],j\in[k]}$ is equitable and thus in particular k-equitable.

Starting with an empty set, we create a subset $Z_1 \subseteq V$ as follows. First we move to Z_1 every vertex $v \in V$ for which

- there exist pairs of indices $(i,j),(i',j') \in [r]\times[k]$ with $\{(i,j),(i',j')\} \in E(R_r^k)$ such that $\big(N_\Gamma(v,U_{i,j}),U_{i',j'}\big)$ or $\big(N_\Gamma(v,U_{i,j}),N_\Gamma(v,U_{i',j'})\big)$ is not $\big(\varepsilon/2,d,p\big)_G$-regular, or
- there exists a cluster $U_{i,j} \in \mathcal{U}$ with $\deg_\Gamma(v,U_{i,j}) \neq (1\pm\varepsilon^*)p|U_{i,j}|$, or
- $\deg_\Gamma(v,U_0) > 2\varepsilon^* pn$ holds.

We also add a minimum number of vertices to Z_1 in order to obtain k-equitability of the sets $\{U_{i,j}\setminus Z_1\}_{i\in[r],j\in[k]}$. By Lemmas 2.10 and 2.11, each with input $\varepsilon/2$ and d, and Proposition 3.8 with input $\varepsilon^*/(1000k^2) < \varepsilon^*$ we have

$$|Z_1| \leq 4kr_1^2 C \max\left\{p^{-2}, p^{-1}\log n\right\} \leq \tfrac{\varepsilon^*}{kr_1}n\,, \tag{3.1}$$

where the factor k accounts for vertices removed to maintain k-equitability.

As a next step, for every $i \in [r]$ and $j \in [k]$ we collect in $W_{i,j}$ all vertices from $U_{i,j}\setminus Z_1$ that destroyed super-regularity on the copy of K_r^k in B_r^k before the removal of Z_1. More precisely, for each $i \in [r]$ and $j \in [k]$ let $W_{i,j}$ be the set of vertices in $U_{i,j} \setminus Z_1$ that have each less than $(d-2\varepsilon^*)p|U_{i,j'}|$ neighbours in $U_{i,j'}$ for some $j' \neq j$. Since for each $i \in [r]$ and distinct $j,j' \in [k]$ the pair $(U_{i,j},U_{i,j'})$ is $(\varepsilon^*,d,p)_G$-regular, we have $|W_{i,j}| \leq k\varepsilon^*|U_{i,j}|$.

Now let $W \subseteq V$ be a set that consists of $U_0\setminus Z_1$ and $\bigcup_{i\in[r],j\in[k]} W_{i,j}$ and a minimum number of additional vertices from $V\setminus Z_1$ to obtain k-equitability of the sets $\{U_{i,j}\setminus(Z_1\cup W)\}_{i\in[r],j\in[k]}$. By construction, we have

$$|W| \leq \varepsilon^* n + kr\cdot k\varepsilon^*\tfrac{n}{kr} \leq 2k\varepsilon^* n.$$

Given any vertex $w \in W$, we have in particular that $w \notin Z_1$ and hence, for each $i \in [r]$ and $j \in [k]$, it holds that

$$\deg_\Gamma(w,U_0) \leq 2\varepsilon^* pn \quad\text{and}\quad \deg_\Gamma(w,U_{i,j}) \leq (1+\varepsilon^*)p|U_{i,j}|.$$

Now let us consider the edges in $E(G)$ that are incident to w. At most $2\varepsilon^* pn$ of these go to U_0, and, clearly, at most $2dpn$ such edges go to clusters $U_{i,j}$ with $\deg_G(w, U_{i,j}) \leq 2dp|U_{i,j}|$. Since $\deg_G(w) \geq \big(\frac{k-1}{k} + \gamma\big)pn$ and $d \leq \gamma/4$, at least $\big(\frac{k-1}{k} + \frac{\gamma}{2}\big)pn$ edges leaving w go to sets $U_{i,j}$ with $\deg_G(w, U_{i,j}) \geq 2dp|U_{i,j}|$. Since for each cluster $U_{i,j} \in \mathcal{U}$ we have $|U_{i,j}| \leq n/(kr)$ and $\deg_G(w, U_{i,j}) \leq \deg_\Gamma(w, U_{i,j}) \leq (1+\varepsilon^*)p|U_{i,j}|$ as $w \notin Z_1$, the number of sets $U_{i,j}$ with $i \in [r]$ and $j \in [k]$ and $\deg_G(w, U_{i,j}) \geq 2dp|U_{i,j}|$ is at least

$$\frac{\big(\frac{k-1}{k} + \frac{\gamma}{2}\big)pn}{(1+\varepsilon^*)p\frac{n}{kr}} \geq \big(\tfrac{k-1}{k} + \tfrac{\gamma}{4}\big)kr.$$

It follows that there are at least $\gamma r/4$ indices $i \in [r]$ with $\deg_G(w, U_{i,j}) \geq 2dp|U_{i,j}|$ for each $j \in [k]$.

To each $w \in W$ we know assign sequentially an index $c(w) \in [r] \times [k]$, where we choose $c(w) = (i,j)$ as follows. The index i is chosen minimal in $[r]$ such that we have $\deg_G(w, U_{i,j'}) \geq 2dp|U_{i,j'}|$ for each $j' \in [k]$ and at most $100r^{-1}k\varepsilon^*\gamma^{-1}n$ vertices $w' \in W$ have so far been assigned $c(w') = (i,j')$ for any $j' \in [k]$. We choose $j \in [k]$ minimising the number of vertices $w' \in W$ with $c(w') = (i,j)$. Because $|W| \leq 2k\varepsilon^* n$, this assignment is always possible.

Next, for each $i \in [r]$ and $j \in [k]$, we let

$$V'_{i,j} = U_{i,j} \setminus (Z_1 \cup W_{i,j}) \cup \big\{w \in W :\ c(w) = (i,j)\big\}.$$

By construction, we have for each $i \in [r]$ and $j \in [k]$ that

$$|U_{i,j} \triangle V'_{i,j}| \leq |Z_1| + |W_{i,j}| + \tfrac{100}{r}k\varepsilon^*\gamma^{-1}n \leq 1000k^2\varepsilon^*\gamma^{-1}|U_{i,j}|\,.$$

Finally, we let Z_2 consist of all vertices $v \in V \setminus Z_1$ with

$$\deg_\Gamma(v, U_{i,j} \triangle V'_{i,j}) \geq 2000k^2\varepsilon^*\gamma^{-1}p|U_{i,j}| \text{ for some } i \in [r] \text{ and } j \in [k],$$

together with a minimum number of additional vertices of $V \setminus Z_1$ to obtain k-equitability of the sets $V'_{i,j} \setminus Z_2$. For each $i \in [r]$ and $j \in [k]$ we set

$$V_{i,j} = V'_{i,j} \setminus Z_2 \quad \text{and } V_0 = Z_1 \cup Z_2.$$

We claim that $\mathcal{V} := \{V_{i,j}\}_{i \in [r], j \in [k]}$ is the desired partition of $V \setminus V_0$.

Note that the sets $V'_{i,j}$ and $V'_{i,j'}$ differ in size by at most one for any $i \in [r]$ and $j, j' \in [k]$, by our construction of the assignment c. By Proposition 3.8 and thanks to the choice of C^* we thus have

$$|Z_2| \leq r_1 + Ckr_1p^{-1}\log n \leq \tfrac{\varepsilon^*}{kr_1}pn\,. \tag{3.2}$$

This gives

$$|U_{i,j} \triangle V_{i,j}| \leq |U_{i,j} \triangle V'_{i,j}| + |Z_2| \leq 2000k^2\varepsilon^*\gamma^{-1}|U_{i,j}|\,. \tag{3.3}$$

Given any $v \in V \setminus V_0$, for each $i \in [r]$ and $j \in [k]$, we have $\deg_\Gamma(v, U_{i,j} \triangle V'_{i,j}) \leq 2000k^2\varepsilon^*\gamma^{-1}p|U_{i,j}|$ because $v \notin Z_2 \subseteq V_0$. We thus have

$$\deg_\Gamma(v, U_{i,j} \triangle V_{i,j}) \leq 2000k^2\varepsilon^*\gamma^{-1}p|U_{i,j}| + |Z_2| \leq 3000k^2\varepsilon^*\gamma^{-1}p|U_{i,j}|. \tag{3.4}$$

Since $v \notin Z_1 \subseteq V_0$ we have $\deg_\Gamma(v, U_{i,j}) = (1 \pm \varepsilon^*)p|U_{i,j}|$, and hence by Equations (3.3) and (3.4)

$$\deg_\Gamma(v, V_{i,j}) = \big(1 \pm 10000k^2\varepsilon^*\gamma^{-1}\big)p|V_{i,j}|\,. \tag{3.5}$$

Adding up (3.1) and (3.2), we get the following desired upper bound on the size of V_0 by choice of C^*:

$$|V_0| \leq 4kr_1^2 C \max\{p^{-2}, p^{-1}\log n\} + r_1 + Ckr_1p^{-1}\log n \leq C^* \max\{p^{-2}, p^{-1}\log n\}\,,$$

as desired. Furthermore, the partition $\mathcal{V} = \{V_{i,j}\}_{i\in[r],j\in[k]}$ is by construction k-equitable, and the graph R_r^k has minimum degree $\big((k-1)/k + \gamma/2\big)kr$ as desired.

In the remainder of the proof we check that $\mathcal{V}$ satisfies Properties (G1)–(G4) as well as the stronger bound $|V_0| \leq C^*p^{-1}\log n$ in case we require (G3') instead of (G3).

For each $i \in [r]$ and $j \in [k]$ we have $|U_{i,j}| = (1 \pm \varepsilon^*)\frac{n}{kr}$, and so by Equation (3.3) and by our choice of ε^* we have

$$\tfrac{n}{4kr} \leq (1-\varepsilon^*)(1 - 2000k^2\varepsilon_1\gamma^{-1})\tfrac{n}{kr} \leq |V_{i,j}| \leq (1+\varepsilon^*)(1 + 2000k^2\varepsilon^*\gamma^{-1})\tfrac{n}{kr} \leq \tfrac{4n}{kr},$$

which is Property (G1).

Next, if $\{(i,j),(i',j')\}$ is an edge of R_r^k, then $(U_{i,j}, U_{i',j'})$ is $(\varepsilon^*, d, p)_G$-regular by construction. By (3.3), we have $|U_{i,j''} \triangle V_{i,j''}| \leq 2000k^2\varepsilon^*\gamma^{-1}|U_{i,j''}|$ for $j'' \in \{j, j'\}$ and hence we know by Proposition 2.4 that G is (ε, d, p)-regular on $(V_{i,j}, V_{i',j'})$ since $\varepsilon^* + 4\sqrt{2000k^2\varepsilon^*\gamma^{-1}} \leq \varepsilon$. Given $i \in [r]$ and distinct indices $j, j' \in [k]$, let v be a vertex of $V_{i,j}$. Observe that since $v \in V_{i,j}$, either we have $v \in U_{i,j} \setminus W$ or $v \in W$. In the first case, since $v \notin W$, we have $\deg_G(v, U_{i,j'}) \geq (d - 2\varepsilon^*)p|U_{i,j}|$. In the other case, it holds that $c(v) = (i,j)$ and hence $\deg_G(v, U_{i,j'}) \geq dp|U_{i,j'}|$. By (3.3) and (3.4) we have

$$\deg_G(v, V_{i,j'}) \geq (d - 2\varepsilon^*)p|U_{i,j}| - 3000k^2\varepsilon^*\gamma^{-1}p|U_{i,j}| \geq (d-\varepsilon)p|V_{i,j'}|\,,$$

giving (G2).

Let $\{(i,j),(i',j')\} \in E(R_r^k)$. Then for any $v \in V \setminus V_0$, since $v \notin Z_1$, we know that the pairs $\big(N_\Gamma(v, U_{i,j}), U_{i',j'}\big)$ and $\big(N_\Gamma(v, U_{i,j}), N_\Gamma(v, U_{i',j'})\big)$ are $\big(\varepsilon/2, d, p\big)_G$-regular. By (3.4) and since $v \notin Z_1$ we have for $(i'', j'') \in \{(i,j),(i',j')\}$ that

$$|N_\Gamma(v, U_{i'',j''}) \triangle N_\Gamma(v, V_{i'',j''})| \leq 3000k^2\varepsilon^*\gamma^{-1}p|U_{i'',j''}| \leq 6000k^2\varepsilon^*\gamma^{-1}|N_\Gamma(v, U_{i'',j''})|.$$

Using this fact and Equation (3.3) we know by Proposition 2.4 that both $\big(N_\Gamma(v, V_{i,j}), V_{i',j'}\big)$ and $\big(N_\Gamma(v, V_{i,j}), N_\Gamma(v, V_{i',j'})\big)$ are $\big(\varepsilon, d, p\big)_G$-regular since $\varepsilon/2 + 4\sqrt{6000k^2\varepsilon^*\gamma^{-1}} \leq \varepsilon$. This shows Property (G3).

Finally, (G4) follows directly from (3.5) and our choice of ε^*.

If we alter the definition of Z_1 by removing the condition on $\big(N_\Gamma(v, U_{i,j}), N_\Gamma(v, U_{i',j'})\big)$, then we do not need to use Lemma 2.11 and the bound in (3.1) therefore improves to $|Z_1| \leq 3kr_1^2Cp^{-1}\log n$. Thus, if we only require (G3'), we obtain $|V_0| \leq C^*p^{-1}\log n$ as claimed. □

3.1.4 The lemma for H

Before proving the lemma for H (Lemma 3.5), let us first state McDiarmid's inequality (see e.g. [98] for a proof) that is essential in our proof.

Lemma 3.9 (McDiarmid's inequality)**.** *Let $X_1, \ldots, X_k$ be independent random variables, where X_i takes values in a finite set A_i for each $i \in [k]$. Suppose that a function $g : A_1 \times \ldots \times A_k \to \mathbb{R}$ satisfies for each $i \in [k]$*

$$\sup_{x_1,\ldots,x_k,\hat{x}_i} |g(x_1, x_2, \ldots, x_k) - g(x_1, x_2, \ldots, x_{i-1}, \hat{x}_i, x_{i+1}, \ldots, x_k)| \leq c_i.$$

Then, for any $\varepsilon > 0$, we have

$$\mathbb{P}\big[|\mathbb{E}[g(X_1,\dots,X_k)] - g(X_1,\dots,X_k)| \geq \varepsilon\big] \leq 2\exp\left(-\frac{2\varepsilon^2}{\sum_{i\in[k]} c_i^2}\right).$$

The idea of the proof of Lemma 3.5 is as follows. First, given the zero-free labelling $\mathcal{L}$ and the $(k+1)$-colouring σ of H, we split $\mathcal{L}$ into the blocks of the definition of zero-freeness. We partition the blocks into r 'sections' of consecutive blocks, such that the i-th section contains about $\sum_{j\in[k]} m_{i,j}$ vertices, and such that the 'boundary vertices, namely the first and last βn vertices of each section, do not receive colour zero.

Assigning the vertices of colour j in the i-th section to (i,j) for each $i \in [r]$ and $j \in [k]$, and the vertices of colour zero in the i-th section to z_i, yields a graph homomorphism. However, it can be very unbalanced since different colours in $[k]$ may be used with different frequencies in each section. To fix this, we replace σ with a new colouring σ', which we obtain as follows. We partition each section into 'intervals' of consecutive blocks, and for each interval except the last in each section, we pick a random permutation of $[k]$. We will show that there is a colouring σ' such that all but the first few vertices of each interval are coloured according to the permutation applied to σ, with vertices of colour zero staying coloured zero. We use this colouring σ' in place of σ to define the mapping f. We let X consist of all vertices whose distance is two or less to either boundary vertices, vertices near the start of an interval, or colour zero vertices.

To complete the proof, we show that so few vertices receive colour zero that they do not affect the desired conclusions significantly. Now the mapping f is in expectation balanced, and using McDiarmid's inequality (Lemma 3.9) we can show that it is also with high probability close to balanced. We also show that, since H is D-degenerate, in the i-th section of $\mathcal{L}$ there are many vertices of degree at most $2D$. In expectation these are distributed about evenly over $\big\{(i,j)\big\}_{j\in[k]}$ by f. Using again McDiarmid's inequality shows that with high probability the same holds. These two observations give us Properties (H1) and (H6), while the other four desired properties hold by construction.

Proof of Lemma 3.5. For given $D \geq 1$, set $\alpha = 1/(24D)$. Let $k, r \geq 1$ and $\xi, \beta > 0$ be given, where $\xi \leq 1/(kr)$ and $\beta \leq 10^{-10}\xi^2/(Dk^4 r)$. Let H and $K_r^k \subseteq B_r^k \subseteq R_r^k$ be graphs as in the statement of the lemma. Let $\mathcal{L}$ be the given labelling of $V(H)$ of bandwidth at most βn. Moreover, let $\sigma : V(H) \to \{0,\dots k\}$ be the given proper $(k+1)$-colouring of $V(H)$ that is $(10/\xi, \beta)$-zero-free with respect to $\mathcal{L}$ and such that the first $\sqrt{\beta}n$ vertices of $\mathcal{L}$ are not mapped to 0 by σ. Also, let $z_1,\dots,z_n$ be vertices such that $z_i \in \big([r]\setminus\{i\}\big)\times[k]$ with $\big\{z_i,(i,j)\big\} \in E(R_r^k)$ for every $i \in [r]$ and $j \in [k]$. Let $\{m_{i,j}\}_{i\in[r],j\in[k]}$ be the given k-equitable integer partition of n with $n/(10kr) \leq m_{i,j} \leq 10n/(kr)$ for every $i \in [r]$ and $j \in [k]$. Finally, set $b = k/\sqrt{\beta}$.

Let us now introduce the notation that we use in this proof. Recall that for every $t \in \big[1/(4k\beta)\big]$ the i-th block is defined as

$$B_t = \{(t-1)4k\beta n + 1, \dots, t4k\beta n\}.$$

Next we split the labelling $\mathcal{L}$ into r *sections*, where the first and the last block of each section are zero-free, i.e. do not contain any vertices coloured with colour zero. Each section is partitioned into *intervals*, each of which but possibly the last one consists of b *blocks.*

Since σ is $(10/\xi, \beta)$-zero-free with respect to $\mathcal{L}$, we can choose indices $0 = t_0 \le t_1 \le \ldots \le t_{r-1} \le t_r = 1/(4k\beta)$ such that B_{t_i} and B_{t_i+1} are zero-free blocks for every $i \in [r]$ and

$$\sum_{t=1}^{t_i} |B_t| \le \sum_{t=1}^{i} \sum_{j\in[k]} m_{t,j} < 12k\beta n + \sum_{t=1}^{t_i} |B_t|.$$

Since $m_{i,j} \ge n/(10kr) > 12k\beta n$, indices $t_0, \ldots, t_r$ are distinct. For every $i \in [r]$ we define the i-th section S_i as

$$\bigcup_{t=t_{i-1}+1}^{t_i} B_t.$$

This means by the choice of the indices $t_0, \ldots, t_r$ that the first and last block of each section are zero-free. Since $\{m_{i,j}\}_{i\in[r],j\in[k]}$ is a k-equitable partition, we have in particular

$$\frac{1}{k}(|S_i| - 12k\beta n) \le m_{i,j} \le \frac{1}{k}\big(|S_i| + 12k\beta n\big). \tag{3.6}$$

The last βn vertices of the blocks B_{t_i} and the first βn vertices of the blocks $B_{t_{i+1}}$ are called *boundary vertices* of H. Notice that colour zero is never assigned to boundary vertices by σ. For each $i \in [r]$, we split S_i into $s_i := \lceil (t_i - t_{i-1} - 1)/b \rceil$ *intervals*, where each of the first $(s_i - 1)$ intervals is the concatenation of exactly b blocks and the last interval consists of $t_i - t_{i-1} - 1 - b(s_i - 1) \le b$ blocks. Therefore, for every $i \in [r]$, we have

$$s_i(b-1)4k\beta n + 1 \le |S_i| \le s_i b 4k\beta n. \tag{3.7}$$

Using Equation (3.6), $b = k/\sqrt{\beta}$, and $n/(10kr) \le m_{i,j} \le 10n/(kr)$ we get, for every $i \in [r]$, the following bounds on s_i

$$\frac{1}{100rk^2\sqrt{\beta}} \le s_i \le \frac{10}{rk^2\sqrt{\beta}}.$$

We denote the intervals of the i-th section by $I_{i,1}, \ldots, I_{i,s_i}$. Let $B^{\text{sw}}_{i,\ell}$ denote the union of the first two blocks of each interval $I_{i,\ell}$. All of these blocks but $B^{\text{sw}}_{i,1}$ and B^{sw}_{i,s_i} will be used to switch colours within parts of H. Notice that we have $|B^{\text{sw}}_{i,\ell}| = 8k\beta n$ and, since σ is $(10/\xi, \beta)$-zero-free with respect to $\mathcal{L}$, at least one of the two blocks of $B^{\text{sw}}_{i,\ell}$ is zero-free. We will not use $B^{\text{sw}}_{i,1}$ and B^{sw}_{i,s_i} to switch colours because we will need that the boundary vertices do not receive colour zero. See Figure 3.2 for an illustration of the i-th section of a labelling.

For every $i \in [r]$ and every $\ell \in \{2, \ldots, s_i - 1\}$, we choose a permutation $\pi_{i,\ell} : [k] \to [k]$ uniformly at random.

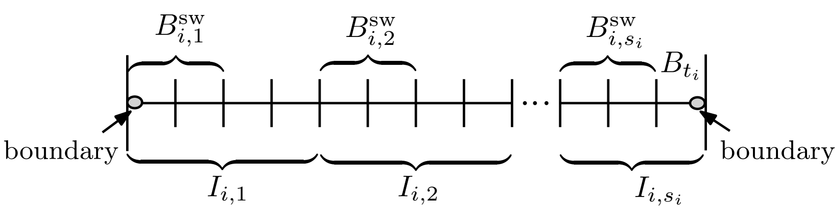

Figure 3.2: Illustration of the i-th section S_i for $b = 4$ and for I_{i,s_i} consisting of 3 blocks.

The next claim ensures that we can use zero-free blocks to obtain a proper colouring of the vertex set such that vertices before the switching block are coloured according to the original colouring and the colours of the vertices after the switching block are permuted as wished. A proof can be found in [42].

Claim 3.10 ([42])**.** *Let $\sigma : [n] \to \{0, \dots, k\}$ be a proper $(k+1)$-colouring of H, let B_t be a zero-free block and let π be any permutation of $[k]$. Then there exists a proper $(k+1)$-colouring σ' of H with $\sigma'(x) = \sigma(x)$ for all $x \in \bigcup_{i<t} B_i$ and*

$$\sigma'(x) = \begin{cases} \pi(\sigma(x)) & \text{if } \sigma(x) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

for all $x \in \bigcup_{i>t} B_i$.

We use Claim 3.10 to switch colours at the beginning of each interval except for the first and last interval of each section. More precisely, we switch colours within the sets $B^{\mathrm{sw}}_{i,\ell}$ so that the colouring of the remaining vertices in the interval $I_{i,\ell}$ matches $\pi_{i,\ell}$. Note that we can indeed use $B^{\mathrm{sw}}_{i,\ell}$ to do the switching since one of the two blocks in $B^{\mathrm{sw}}_{i,\ell}$ is zero-free. In particular, we get a proper $(k+1)$-colouring $\sigma' = \sigma'\big(\pi_{1,2}, \dots, \pi_{r,s_r-1}\big) : V(H) \to \{0, \dots k+1\}$ of H that fulfils the following. For every $x \in I_{1,1}$ we have

$$\sigma'(x) = \sigma(x),$$

for each $i \in [r]$ and $\ell \in \{2, \dots, s_i - 1\}$ and every $x \in I_{i,\ell} \setminus B^{\mathrm{sw}}_{i,\ell}$ we have that

$$\sigma'(x) = \begin{cases} \pi_{i,\ell}\big(\sigma(x)\big) & \text{if } \sigma(x) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

and for each $i \in [r]$ and every $x \in I_{i,s_i} \cup I_{i+1,1}$ (where $I_{r+1,1} := \varnothing$) we have that

$$\sigma'(x) = \pi_{i,s_i-1}\big(\sigma(x)\big).$$

While σ' is well-defined on the sets $B^{\mathrm{sw}}_{1,2}, \dots, B^{\mathrm{sw}}_{r,s_r-1}$ by Claim 3.10, the definition on these sets is rather complicated as it is depends on which of the two blocks in $B^{\mathrm{sw}}_{i,\ell}$ is zero-free and on the colourings before and after the switching. However, the precise definition on these sets is not important for the remainder of the proof. Hence, we omit it here. Observe that σ' never assigns colour zero to boundary vertices.

Using σ' we now define $f = f\big(\pi_{1,2}, \dots, \pi_{r,s_r-1}\big) : V(H) \to [r] \times [k]$ as follows. For each $i \in [r]$ and $x \in S_i$ we set

$$f(x) := \begin{cases} \big(i, \sigma'(x)\big) & \text{if } \sigma'(x) \neq 0 \\ z_i & \text{otherwise,} \end{cases}$$

where $z_i \in \big([r] \setminus \{i\}\big) \times [k]$ is the vertex defined in the statement of the lemma. Let X consist of all vertices at distance two or less from a boundary vertex of $\mathcal{L}$, from a vertex in any $B^{\mathrm{sw}}_{i,\ell}$, or from a colour zero vertex. We now show that f and X satisfy Properties (H2)–(H5) with probability 1 and Properties (H1) and (H6) with positive probability. In particular, this implies that the desired f and X exist.

We start with Property (H1). For each $i \in [r]$ let

$$S^*_i := S_i \setminus \left(\bigcup_{\ell \in [s_i]} B^{\mathrm{sw}}_{i,\ell} \cup I_{i,1} \cup I_{i,s_i} \right)$$

be the set of all vertices in S_i except for the first and last interval and the first two blocks of each interval of S_i. We will also make use of the following restricted function

$$f^* = f^*\big(\pi_{1,2}, \dots, \pi_{r,s_r}\big) := f\big|_{\bigcup_{i\in[r]} S_i^*} .$$

The basic idea of the proof of Property (H1) is to determine bounds on $|f^{*-1}(i,j)|$ that hold with positive probability and then deduce the desired bounds on $|f^{-1}(i,j)|$. Since the permutations $\pi_{i,\ell}$ were chosen uniformly at random, we have by definition of f^* that the expected number of vertices mapped to $(i,j) \in [r] \times [k]$ by f^* is

$$\mathbb{E}\big[|f^{*-1}(i,j)|\big] = \frac{1}{k}\Big[(s_i-2)(b-2)4k\beta n - \big|\{x \in S_i^* : \sigma(x) = 0\}\big|\Big] \\ + \Big| \bigcup_{\iota\in[r]\setminus\{i\}} \{x \in S_\iota^* : \sigma(x) = 0 \text{ and } z_\iota = (i,j)\}\Big| .$$

In particular, the following bounds on the expected value of $|f^{*-1}(i,j)|$ hold.

$$\mathbb{E}\big[|f^{*-1}(i,j)|\big] \le (s_i-2)(b-2)4\beta n + \frac{\xi}{10}n \tag{3.8}$$

and

$$\mathbb{E}\big[|f^{*-1}(i,j)|\big] \ge (1-\xi/10)(s_i-2)(b-2)4\beta n \ge (s_i-2)(b-2)4\beta n - \frac{\xi}{10}n. \tag{3.9}$$

If one replaced a permutation $\pi_{i,\ell}$ by some other permutation $\tilde{\pi} : [k] \to [k]$, then $|f^{*-1}(i,j)|$ would change by at most $(b-2)4k\beta n$. Hence, by McDiarmid's inequality (Lemma 3.9) we have

$$\mathbb{P}\left[\big|(s_i-2)(b-2)4\beta n - |f^{*-1}(i,j)|\big| \ge \frac{\xi}{5}n\right] \overset{(3.8),(3.9)}{\le} \\ \mathbb{P}\left[\big|\mathbb{E}[|f^{*-1}(i,j)|] - |f^{*-1}(i,j)|\big| \ge \frac{\xi}{10}n\right] \le 2\exp\left(-\frac{\xi^2 n^2}{50(s_i-2)\big((b-2)4k\beta n\big)^2}\right). \tag{3.10}$$

Taking the union bound over all $j \in [k]$ and using $s_i \le 10/(rk^2\sqrt{\beta})$ and $b = k/\sqrt{\beta}$ as well as $\beta \le 10^{-10}\xi^2/(Dk^4 r)$ yields

$$\mathbb{P}\left[\big|(s_i-2)(b-2)4\beta n - |f^{*-1}(i,j)|\big| \ge \frac{\xi}{5}n \text{ for all } j \in [k]\right] \le 2k\exp\left(-\frac{\xi^2 r}{8000k^2\sqrt{\beta}}\right) < 1.$$

Observe that $|f^{*-1}(i,j)|$ is independent of the choices for $\pi_{i',\ell}$ if $i' \ne i$. Hence, with positive probability we have, for every $i \in [r]$ and $j \in [k]$, that

$$(s_i-2)(b-2)4\beta n - \frac{\xi}{5}n \le |f^{*-1}(i,j)| \le (s_i-2)(b-2)4\beta n + \frac{\xi}{5}n.$$

From the definition of f^* it follows that $|f^{-1}(i,j)| \ge |f^{*-1}(i,j)|$ and $|f^{-1}(i,j)|$ is at most

$$|f^{*-1}(i,j)| + |I_{i,1}| + |I_{i,s_i}| + \sum_{\ell=2}^{s_i-1} |B_{i,\ell}^{\mathrm{sw}}| + \Big|\Big\{x \in \bigcup_{\iota\in[r]\setminus\{i\}} S_\iota \setminus S_\iota^* : \sigma'(x) = 0 \text{ and } z_\iota = (i,j)\Big\}\Big|.$$

Using $s_i \le 10/(rk^2\sqrt{\beta})$ and $b = k/\sqrt{\beta}$ and $\beta \le 10^{-10}\xi^2/(Dk^4r)$, with positive probability we have for every $i \in [r]$ and $j \in [k]$ that

$$\begin{aligned}
|f^{-1}(i,j)| \ge |f^{*-1}(i,j)| &\ge (s_i-2)(b-2)4\beta n - \frac{\xi}{5}n \\
&\ge (s_i-2)(b-2)4\beta n - \frac{\xi}{5}n + \left(8(s_i+b)\beta n - \frac{4}{5}\xi n\right) \\
&\ge s_i b 4\beta n + 16\beta n - \xi n \\
&\overset{(3.7)}{\ge} \frac{1}{k}\big(|S_i| + 16k\beta n\big) - \xi n \overset{(3.6)}{\ge} m_{i,j} - \xi n.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
|f^{-1}(i,j)| &\le |f^{*-1}(i,j)| + |I_{i,1}| + |I_{i,s_i}| + \sum_{\ell=2}^{s_i-1} |B^{\mathrm{sw}}_{i,\ell}| \\
&\quad + \Big|\big\{x \in \bigcup_{\iota\in[r]\setminus\{i\}} S_\iota \setminus S^*_\iota : \sigma'(x) = 0 \text{ and } z_\iota = (i,j)\big\}\Big| \\
&\le (s_i-2)(b-2)4\beta n + \frac{\xi}{5}n + 8bk\beta n + (s_i-2)8k\beta n + \frac{\xi}{10}n \\
&\le \frac{1}{k}\big((s_i-2)(b-2)4k\beta n\big) + \xi n \\
&\le \frac{1}{k}(|S_i| - 12k\beta n) + \xi n \overset{(3.6)}{\le} m_{i,j} + \xi n,
\end{aligned}$$

which shows that Property (H1) holds with positive probability.

By definition of X, since $\mathcal{L}$ is a βn-bandwidth ordering, any vertex in X is at distance at most $2\beta n$ in $\mathcal{L}$ from a boundary vertex, a vertex of some $B^{\mathrm{sw}}_{i,\ell}$, or from a vertex assigned colour zero. Since there are r sections, the boundary vertices form $(r-1)$ intervals each of length $2\beta n$, and so at most $6r\beta n$ vertices of H are at distance 2 or less from a boundary vertex. There are $\sum_{i\in[r]} s_i$ intervals and hence $\sum_{i\in[r]} s_i$ switching blocks each of size $8k\beta n$. As $s_i \le 10/(rk^2\sqrt{\beta})$ for every $i \in [r]$, there are at most $(4+8k)\beta n \cdot 10/(k^2\sqrt{\beta})$ vertices at distance 2 or less from a vertex of some switching block. Similarly, because $\mathcal{L}$ is $(10/\xi, \beta)$-zero-free, in any consecutive $10/\xi$ blocks at most one contains vertices of colour zero, and hence at most $(8+4k)\beta n$ vertices in any such $10/\xi$ consecutive blocks are at distance 2 or less from a vertex of colour zero. Thus we have

$$|X| \le 6r\beta n + (4+8k)\beta n\left(\tfrac{10}{k^2\sqrt{\beta n}}\right) + (8+4k)\beta n\big(\tfrac{n}{4k\beta n\cdot 10/\xi} + 1\big) \le 6r\beta n + \tfrac{1}{4}\xi n + \tfrac{1}{3}\xi n \le \xi n\,,$$

which gives (H2).

Since σ' is a proper colouring, and boundary vertices are not adjacent to colour zero vertices, by definition, f restricted to the boundary vertices is a graph homomorphism to B^k_r. On the other hand, on each section S_i, again since σ' is a proper colouring and since $\big\{(i,j)\big\}_{j\in[k]} \cup \{z_i\}$ forms a clique in R^k_r, f is a graph homomorphism to R^k_r. Since $\mathcal{L}$ is a βn-bandwidth ordering, any edge of H is either contained in a section or goes between two boundary vertices, and we conclude that f is a graph homomorphism from H to R^k_r, giving (H3).

Now, given $i \in [r]$ and $j \in [k]$, and $x \in f^{-1}(i,j) \setminus X$, if $\{x,y\}$ and $\{y,z\}$ are edges of H, then y and z are at distance two or less from x in H. In particular, by definition of X neither y nor z is either a boundary vertex, in any $B^{\mathrm{sw}}_{i,\ell}$, or assigned colour zero. Since boundary vertices appear in intervals of length $2\beta n$ in $\mathcal{L}$, and $\mathcal{L}$ is a βn-bandwidth ordering, it follows that y and z are both in S_i. Furthermore, suppose $x \in I_{i,\ell}$ for some ℓ. By definition $x \notin B^{\mathrm{sw}}_{i,\ell}$. Because $B^{\mathrm{sw}}_{i,\ell}$ and $B^{\mathrm{sw}}_{i,\ell+1}$ (if the latter exists) are intervals of length $8k\beta n$, both y and z are also in $I_{i,\ell} \setminus B^{\mathrm{sw}}_{i,\ell}$, and in particular both y and z are in $\bigcup_{j' \in [k]} f^{-1}(i,j')$, giving (H4).

Since $\sqrt{\beta}n \leq b4k\beta n \leq |I_{1,1}|$ and $\sigma'(x) \neq 0$ for each x in the first $\sqrt{\beta}n$ vertices of $\mathcal{L}$, it follows directly from the definition of f that $f(x) = \big(1, \sigma(x)\big)$, which shows Property (H5).

Finally, we show that Property (H6) holds with positive probability. Let $i \in [r]$ and $j \in [k]$. We define the random variable $\mathcal{E}_{i,j} := |\{x \in f^{*-1}(i,j) : \deg(x) \leq 2D\}|$. Since H is D-degenerate and $\mathcal{L}$ is a labelling of bandwidth at most βn we have

$$e\big(S_i^*, V(H)\big) \leq D|S_i^*| + D4\beta n \leq D\big(1 + 1/(4D)\big)|S_i^*|.$$

Hence it must hold that $|\{x \in S_i^* : \deg(x) \geq 2D+1\}|(2D+1) \leq 2D\big(1+1/(4D)\big)|S_i^*|$. This yields $|\{x \in S_i^* : \deg(x) \leq 2D\}| \geq |S_i^*|/(6D)$ and therefore

$$\mathbb{E}[\mathcal{E}_{i,j}] \geq \frac{1}{6kD}|S_i^*| \geq \frac{1}{6D}(s_i-2)(b-2)4\beta n.$$

By applying Chernoff's inequality (Theorem 2.16) and using Equations (3.6) and (3.7) as well as $\alpha = 1/(24D)$ we get

$$\begin{aligned}\mathbb{P}\Big[\big|\{x \in f^1(i,j) : \deg(x) \leq 2D\}\big| < \alpha|f^{-1}(i,j)|\Big] &\overset{(H1)}{\leq} \mathbb{P}\Big[\mathcal{E}_{i,j} < \alpha(s_i b4\beta n + 2\xi n)\Big] \\ &\leq \mathbb{P}\Big[\mathcal{E}_{i,j} < 2\alpha\big((s_i-2)(b-2)4\beta n\big)\Big] \\ &\leq \mathbb{P}\Big[\mathcal{E}_{i,j} < \tfrac{1}{2}\mathbb{E}[\mathcal{E}_{i,j}]\Big] < 2\exp\left(-\frac{(s_i-2)(b-2)4\beta n}{72}\right) < 1.\end{aligned}$$

Taking the union bound over all $i \in [r]$ and $j \in [k]$ yields that Property (H6) holds with positive probability. $\square$

3.1.5 Common neighbourhood lemma

In the proof of the common neighbourhood lemma (Lemma 3.6) we use a version of the sparse regularity lemma, where an initial partition is allowed to have different sizes and the final partition is an equitable regular refinement of the initial one. We stated and proved this lemma in Section 2.2 (Lemma 2.5).

The main idea of the proof of Lemma 3.6 can be summarised as follows. First we apply Lemma 2.5 to the given partition $\{V_i \setminus W\}_{i\in[k]} \cup W$ to obtain a regular refinement such that each of these $k+1$ sets is equitably partitioned into t clusters and an exceptional set. Then we show that there exist clusters $W' \subseteq W$ and $V_i' \subseteq V_i$ for each $i \in [k]$ such that (W', V_i') is regular for each $i \in [k]$. Finally we use induction to prove that we can find a tuple in W' with the desired properties. For this we only need the sets $W', V_1', \ldots, V_k'$ from the regular partition.

Proof of Lemma 3.6. First we fix all constants that we need throughout the proof. Given $d > 0$, $k \geq 1$, and $\Delta \geq 2$, let $\varepsilon^{**}_\Delta = 8^{-\Delta}\frac{1}{(k+1)^2}\left(\frac{d}{8}\right)^\Delta$. Now, for each $j = 1, \ldots, \Delta$ sequentially, choose $\varepsilon^{**}_{\Delta-j} \le \varepsilon^{**}_{\Delta-j+1}$ not larger than the ε_0 returned by Lemma 2.10 for input $\varepsilon^{**}_{\Delta-j}$ and $d/2$.

Now, Lemma 2.5 with input ε^{**}_0 and $s = k+1$ returns $t_1 \ge 1$. We set

$$\alpha = \tfrac{1}{2t_1}\big(\tfrac{d}{4}\big)^{\Delta}\,.$$

Next, given $\varepsilon^* > 0$, let $\varepsilon^*_{\Delta-1,\Delta-1} := \varepsilon^*$, and let $\varepsilon^*_{j,\Delta} = \varepsilon^*_{\Delta,j} = 1$ for each $1 \le j \le \Delta$. For each $(j,j') \in [\Delta]^2 \setminus \{(1,1)\}$ in lexicographic order sequentially, we choose

$$\varepsilon^*_{\Delta-j,\Delta-j'} \le \min\{\varepsilon^*_{\Delta-j+1,\Delta-j'}, \varepsilon^*_{\Delta-j,\Delta-j'+1}, \varepsilon^*_{\Delta-j+1,\Delta-j'+1}\}$$

not larger than the ε_0 returned by Lemma 2.10 for both input $\varepsilon^*_{\Delta-j+1,\Delta-j'}$ and d, and for input $\varepsilon^*_{\Delta-j,\Delta-j'+1}$ and d, and not larger than the ε_0 returned by Lemma 2.11 for input $\varepsilon^*_{\Delta-j+1,\Delta-j'+1}$ and d.

We choose ε_0 small enough such that $(1+\varepsilon_0)^{\Delta} \le 1+\varepsilon^*$ and $(1-\varepsilon_0)^{\Delta} \ge 1-\varepsilon^*$. Given $r \ge 1$ and ε with $0 < \varepsilon \le \varepsilon_0$, suppose that C is the maximum of the C-outputs of each of the calls to Lemmas 2.10 and 2.11 with the parameters from above, and of Proposition 3.8 with input ε_0. Finally, we set

$$C^* = 10^{12}k^4t_1r^4\varepsilon^{-4}2^{2\Delta}C\,.$$

Given $p \ge C^*\big(\frac{\log n}{n}\big)^{1/\Delta}$, we know that a.a.s. the good events of each of the above calls to Lemma 2.10 and 2.11, and to Proposition 3.8 and Lemma 2.5, occur. We condition from now on upon these events for $\Gamma = G(n,p)$.

Let $G = (V,E)$ be a (not necessarily spanning) subgraph of Γ. Suppose $\{V_i\}_{i\in[k]}$ and W satisfy the conditions of the lemma. We first apply Lemma 2.5, with the input parameters ε^{**}_0 and $s = k+1$, to $G[V_1 \cup \cdots \cup V_k \cup W]$, with input partition $\{V_i \setminus W\}_{i\in[k]} \cup \{W\}$. We can do this because $Cp^{-1}\log n < 10^{-10}\varepsilon^4 pn/(k^4r^4)$, so that the good event of Proposition 3.8 guarantees that the conditions of Lemma 2.5 are satisfied. This returns an (ε^{**}_0,p)-regular refinement such that each set of $\{V_i \setminus W\}_{i\in[k]} \cup \{W\}$ is equitably partitioned into $1 \le t \le t_1$ clusters together with an exceptional set whose size is an ε^{**}_0-fraction of the size of the set itself.

As a next step we show that there exist clusters $W' \subseteq W$ and $V'_i \subseteq V_i$ for each $i \in [k]$ such that (W',V'_i) is $(\varepsilon^{**}_0, d/2, p)_G$-regular for each $i \in [k]$. Let $W' \subseteq W$ be a cluster that is in at most $2k^2\varepsilon^{**}_0 t$ pairs with clusters in $\big(V_1 \cup \cdots \cup V_k\big) \setminus W$ that are not $(\varepsilon^{**}_0,p)_G$-regular. Observe that such a cluster exists by averaging. By Proposition 3.8 and (V1), at most $16k\varepsilon^{**}_0 p|W'|n/r$ edges lie in the irregular pairs between W' and the V_i, and by Proposition 3.8 and (V3) at most $2p|W'||W| < \varepsilon^{**}_0 p|W'|n/r$ edges leaving W' lie in W. By (V4), for each $i \in [k]$ each $w \in W'$ has at least $dp|V_i|$ neighbours in V_i, and hence there are at least $dp|V_i||W'|/2$ edges from W' to $V_i \setminus W$ that lie in $(\varepsilon^{**}_0,p)_G$-regular pairs. By averaging, for each $i \in [k]$ there exists a cluster V'_i of the partition such that (W',V'_i) is $(\varepsilon^{**}_0, d/2, p)_G$-regular. For the remainder of the proof, we will only need these $k+1$ clusters from the regular partition.

Notice that for every $i \in [k]$ we have

$$|V_i| \ge |V'_i| \overset{(V1)}{\ge} \frac{n}{8kt_1r} \ge \frac{1}{8kt_1r}(C^*)^2p^{-2}\log n \ge C^*p^{-2}\log n$$

and

$$|W'| \overset{(V3)}{\ge} 10^{-11}\frac{\varepsilon^4pn}{t_1k^4r^4} \ge 10^{-11}\frac{\varepsilon^4}{t_1k^4r^4}(C^*)^2p^{-1}\log n \ge C^*p^{-1}\log n, \tag{3.11}$$

both by the choice of C^* and p.

We choose the required Δ-tuple $(w_1, \dots, w_\Delta)$ inductively by using the following claim.

Claim 3.11. *For each $0 \le \ell \le \Delta$ there exists an ℓ-tuple $(w_1, \dots, w_\ell) \in \binom{W'}{\ell}$ such that the following holds. For every $\Lambda, \Lambda^* \subseteq [\ell]$, and all distinct indices $i, i' \in [k]$ we have*

(L1) $\big(\bigcap_{j\in\Lambda} N_\Gamma(w_j, V_i'), W'\big)$ *is* $(\varepsilon^{**}_{|\Lambda|}, \frac{d}{2}, p)_G$*-regular if* $|\Lambda| < \Delta$,

(L2) $|\bigcap_{j\in\Lambda} N_G(w_j, V_i')| \ge \big(\frac{d}{4}\big)^{|\Lambda|} p^{|\Lambda|} |V_i'|$,

(L3) $|\bigcap_{j\in\Lambda} N_\Gamma(w_j)| \le (1+\varepsilon_0)^{|\Lambda|} p^{|\Lambda|} n$,

(L4) $|\bigcap_{j\in\Lambda} N_\Gamma(w_j, V_i')| = (1\pm\varepsilon_0)^{|\Lambda|} p^{|\Lambda|} |V_i'|$,

(L5) $|\bigcap_{j\in\Lambda} N_\Gamma(w_j, V_i)| = (1\pm\varepsilon_0)^{|\Lambda|} p^{|\Lambda|} |V_i|$, *and*

(L6) $\big(\bigcap_{j\in\Lambda} N_\Gamma(w_j, V_i), \bigcap_{j^*\in\Lambda^*} N_\Gamma(w_{j^*}, V_{i'})\big)$ *is* $(\varepsilon^*_{|\Lambda|,|\Lambda^*|}, d, p)_G$*-regular if* $|\Lambda|, |\Lambda^*| < \Delta$ *and either* $\Delta \ge 3$ *or* $\Lambda \cap \Lambda^* = \varnothing$ *or both.*

We prove this claim by induction on ℓ. If $\Lambda = \varnothing$ then we define $\bigcap_{j\in\Lambda} N_\Gamma(w_j, V_i')$ to be equal to V_i', and we set $[0] = \varnothing$.

Proof of Claim 3.11. For the base case $\ell = 0$, observe that (L1) follows from our choice of W' and $\{V_i'\}_{i\in[r]}$. For all distinct indices $i, j \in [k]$, the pair (V_i, V_j) is $(\varepsilon, d, p)_G$-regular by (V2), and since $\varepsilon \le \varepsilon^*_{0,0}$, we have (L6). The remaining three properties (L2), (L4) and (L5) are tautologies for $\ell = 0$.

For the inductive step, suppose that there exists an ℓ-tuple $(w_1, \dots, w_\ell) \in \binom{W'}{\ell}$ satisfying (L1)–(L6) for some $0 \le \ell < \Delta$. We now find a vertex $w_{\ell+1} \in W'$ such that the $(\ell+1)$-tuple $(w_1, \dots, w_{\ell+1})$ still satisfies (L1)–(L6). We do this by determining, for each of these five conditions, an upper bound on the number of vertices in W' that violate them and show that the sum of these upper bounds is less than $|W'| - \ell$.

Suppose $\Lambda \subseteq [\ell]$ satisfies $|\Lambda| < \Delta - 1$, and suppose $i \in [k]$. By the choice of C^* and p we have for every $i \in [k]$

$$\Big|\bigcap_{j\in\Lambda} N_\Gamma(w_j, V_i')\Big| \overset{(L4)}{\ge} (1-\varepsilon_0)^{|\Lambda|} p^{|\Lambda|} |V_i'| \overset{|\Lambda|<\Delta-1}{\ge} (1-\varepsilon_0)^{\Delta-2} p^{\Delta-2} \frac{n}{8ktr} \ge C^* p^{-2} \log n\,. \quad (3.12)$$

We also have $|W'| \ge C^* p^{-1} \log n \ge C p^{-1} \log n$ by (3.11) and $\big(\bigcap_{j\in\Lambda} N_\Gamma(w_j, V_i'), W'\big)$ is an $(\varepsilon^{**}_{|\Lambda|}, d/2, p)_G$-regular pair by (L1). Since the good event of Lemma 2.10 with input $\varepsilon^{**}_{|\Lambda|+1}$ and $d/2$ occurs, there exist at most $Cp^{-1}\log n$ vertices w in W' such that $\big(\bigcap_{j\in\Lambda} N_\Gamma(w_j, V_i') \cap N_\Gamma(w), W'\big) = \big(\bigcap_{j\in\Lambda} N_\Gamma(w_j, V_i') \cap N_\Gamma(w, V_i'), W'\big)$ is not $(\varepsilon^{**}_{|\Lambda|+1}, \frac{d}{2}, p)_G$-regular. Summing over all possible choices of $\Lambda \subseteq [l]$ and $i \in [k]$, there are at most $2^\Delta k^2 C p^{-1} \log n$ vertices w in W' such that $(w_1, \dots, w_l, w)$ does not satisfy (L1).

Moving on to (L2), let $\Lambda \subseteq [\ell]$ and $i \in [k]$ be given. We have

$$\Big|\bigcap_{j\in\Lambda} N_G(w_j, V_i')\Big| \overset{(L2)}{\ge} \big(\tfrac{d}{4}\big)^{|\Lambda|} p^{|\Lambda|} |V_i'|$$

and

$$\Big|\bigcap_{j\in\Lambda} N_\Gamma(w_j,V_i')\Big| \overset{(L4)}{\leq} (1+\varepsilon_0)^{|\Lambda|}p^{|\Lambda|}|V_i'|\,.$$

By choice of ε_0 and $\varepsilon^{**}_{|\Lambda|}$, we thus have $\big|\bigcap_{j\in\Lambda} N_G(w_j,V_i')\big| \geq \varepsilon^{**}_{|\Lambda|}\big|\bigcap_{j\in\Lambda} N_\Gamma(w_j,V_i')\big|$. Now by (L1), the pair $\big(W',\bigcap_{j\in\Lambda} N_\Gamma(w_j,V_i')\big)$ is $\big(\varepsilon^{**}_{|\Lambda|},\frac{d}{2},p\big)_G$-regular, and thus the number of vertices $w\in W'$ such that

$$\Big|N_G(w,V_i')\cap\bigcap_{j\in\Lambda} N_G(w_j,V_i')\Big| < \big(\tfrac{d}{4}\big)^{|\Lambda|+1}p^{|\Lambda|+1}|V_i'|$$

is at most $\varepsilon^{**}_{|\Lambda|}|W'|\leq\varepsilon^{**}_{\Delta}|W'|$. Summing over the choices of $\Lambda\subseteq[\ell]$ and $i\in[k]$, the number of $w\in W'$ violating (L2) is at most $2^\Delta k\varepsilon^{**}_\Delta|W'|$.

For (L4), given $\Lambda\subseteq[\ell]$ and $i\in[k]$, by (L4) we have

$$\Big|\bigcap_{j\in\Lambda} N_\Gamma(w_j,V_i')\Big| = (1\pm\varepsilon_0)^{|\Lambda|}p^{|\Lambda|}|V_i'|\,,$$

and by choice of ε_0 and p, in particular $\big|\bigcap_{j\in\Lambda} N_\Gamma(w_j,V_i')\big| \geq Cp^{-1}\log n$. Since the good event of Proposition 3.8 occurs, the number of vertices $w\in W'$ such that $\big|N_\Gamma(w,V_i')\cap\bigcap_{j\in\Lambda} N_\Gamma(w_j,V_i')\big|$ is smaller than $(1-\varepsilon_0)^{|\Lambda|+1}p^{|\Lambda|+1}|V_i'|$ or larger than $(1+\varepsilon_0)^{|\Lambda|+1}p^{|\Lambda|+1}|V_i'|$ is at most $2Cp^{-1}\log n$. Summing over the choices of $\Lambda\subseteq[\ell]$ and of $i\in[k]$, we conclude that at most $2^{\Delta+1}kCp^{-1}\log n$ vertices of W' violate (L4). Since $n\geq|V_i|\geq|V_i'|$, the same calculation shows that a further at most $2^{\Delta+1}kCp^{-1}\log n$ vertices of W' violate (L5), and at most $2^{\Delta+1}kCp^{-1}\log n$ vertices of W' violate (L3).

Finally, we come to (L6). Suppose we are given $\Lambda,\Lambda'\subseteq[\ell]$ and distinct $i,i'\in[k]$. Suppose that $|\Lambda|\leq\Delta-2$ and $|\Lambda'|\leq\Delta-1$. We wish to show that for most vertices $w\in W'$, the pair $\big(N_\Gamma(w,V_i)\cap\bigcap_{j\in\Lambda} N_\Gamma(w_j,V_i),\bigcap_{j\in\Lambda} N_\Gamma(w_j,V_i')\big)$ is $\big(\varepsilon^*_{|\Lambda|+1,|\Lambda'|},d,p\big)_G$-regular, and furthermore, if $\Delta\geq3$ and $|\Lambda'|\leq\Delta-2$, that the pair $\big(N_\Gamma(w,V_i)\cap\bigcap_{j\in\Lambda} N_\Gamma(w_j,V_i),N_\Gamma(w,V_{i'})\cap\bigcap_{j\in\Lambda} N_\Gamma(w_j,V_i')\big)$ is $\big(\varepsilon^*_{|\Lambda|+1,|\Lambda'|+1},d,p\big)_G$-regular.

By (L5), and by choice of ε_0, C and p, we have

$$\Big|\bigcap_{j\in\Lambda} N_\Gamma(w_j,V_i)\Big| \geq (1-\varepsilon_0)^{|\Lambda|}p^{|\Lambda|}|V_i| \geq Cp^{|\Lambda|-\Delta}\log n$$

and

$$\Big|\bigcap_{j\in\Lambda'} N_\Gamma(w_j,V_{i'})\Big| \geq (1-\varepsilon_0)^{|\Lambda'|}p^{|\Lambda'|}|V_{i'}| \geq Cp^{|\Lambda'|-\Delta}\log n\,.$$

By (L6), the pair $\big(\bigcap_{j\in\Lambda} N_\Gamma(w_j,V_i),\bigcap_{j\in\Lambda} N_\Gamma(w_j,V_i')\big)$ is $\big(\varepsilon^*_{|\Lambda|,|\Lambda'|},d,p\big)_G$-regular. Since the good event of Lemma 2.10 with input $\varepsilon^*_{|\Lambda|+1,|\Lambda'|}$ and d occurs, there are at most $Cp^{-1}\log n$ vertices w of W' such that the pair $\big(N_\Gamma(w,V_i)\cap\bigcap_{j\in\Lambda} N_\Gamma(w_j,V_i),\bigcap_{j\in\Lambda} N_\Gamma(w_j,V_i')\big)$ is not $\big(\varepsilon^*_{|\Lambda|+1,|\Lambda'|},d,p\big)_G$-regular. Moreover, if $|\Lambda'|\leq\Delta-2$, then, since the good event of Lemma 2.11 with input $\varepsilon^*_{|\Lambda|+1,|\Lambda'|+1}$ and d occurs, there are at most $Cp^{-2}\log n$ vertices w of W' such that $\Big(N_\Gamma(w,V_i)\cap\bigcap_{j\in\Lambda} N_\Gamma(w_j,V_i),N_\Gamma(w,V_{i'})\cap\bigcap_{j\in\Lambda} N_\Gamma(w_j,V_i')\Big)$ is not $\big(\varepsilon^*_{|\Lambda|+1,|\Lambda'|},d,p\big)_G$-regular.

Observe that if $\Delta = 2$ the property (L6) does not require this pair to be regular. Summing over the choices of $\Lambda, \Lambda' \subseteq [\ell]$ and $i, i' \in [k]$, we conclude that if $\Delta = 2$ then at most $2^{2\Delta}k^2Cp^{-1}\log n$ vertices w of W' cause (L6) to fail, while if $\Delta \geq 3$, at most $2^{2\Delta}k^2C(p^{-1} + p^{-2})\log n$ vertices w of W' violate (L6).

Summing up, if $\Delta = 2$ then at most

$$2^{\Delta}k^2Cp^{-1}\log n + 2^{\Delta}k\varepsilon^{**}_{\Delta}|W'| + 3\cdot 2^{\Delta+1}kCp^{-1}\log n + 2^{2\Delta}k^2Cp^{-1}\log n \tag{3.13}$$

vertices w of W' cannot be chosen as $w_{\ell+1}$. By choice of C^* and $\varepsilon^{**}_{\Delta}$, and by choice of p, this is at most $|W'|/2$, so that there exists a vertex of W', which can be chosen as $w_{\ell+1}$, as desired. If on the other hand $\Delta \geq 3$, then at most

$$2^{\Delta}k^2Cp^{-1}\log n + 2^{\Delta}k\varepsilon^{**}_{\Delta}|W'| + 3\cdot 2^{\Delta+1}kCp^{-1}\log n + 2^{2\Delta}k^2C(p^{-1}+p^{-2})\log n \tag{3.14}$$

vertices of W' cannot be chosen as $w_{\ell+1}$. Again by choice of C^*, $\varepsilon^{**}_{\Delta}$ and p, this is at most $|W'|/2$, and again we therefore can choose $w_{\ell+1}$ satisfying (L1)–(L6) as desired. □

Finally, let us argue why the lemma is a consequence of Claim 3.11. Let $(w_1, \dots, w_\Delta) \in \binom{W'}{\Delta}$ be a tuple satisfying (L1)–(L6). By (L2), for any $\Lambda \subseteq [\ell]$ and $i \in [k]$ we have

$$\Big|\bigcap_{j\in\Lambda} N_G(w_j, V_i)\Big| \geq \left(\tfrac{d}{4}\right)^{|\Lambda|}p^{|\Lambda|}|V_i'| \geq \left(\tfrac{d}{4}\right)^{\Delta}p^{|\Lambda|}\tfrac{|V_i|}{2t_1} \geq \alpha p^{|\Lambda|}|V_i|\,,$$

as required for (W1). Properties (W2), (W3) and (W4) are respectively (L3), (L5) and (L6), by choice of ε_0. □

3.1.6 Balancing lemma

The statement of Lemma 3.7 gives us a partition of $V(G)$ with parts $\{V_{i,j}\}_{i,j}$, and a collection of 'target integers' $\{n_{i,j}\}_{i,j}$, where $n_{i,j}$ is close to $|V_{i,j}|$ for every $i \in [r]$ and $j \in [k]$ and where $\sum n_{i,j} = \sum |V_{i,j}|$ holds. Our aim is to find a partition of $V(G)$ with parts $\{V'_{i,j}\}_{i,j}$ such that $|V'_{i,j}| = n_{i,j}$ for each $i \in [r]$ and $j \in [k]$. This partition is required to maintain similar regularity properties as the original partition, while not substantially changing common neighbourhoods of vertices.

There are two steps to our proof. In a first step, we correct *global imbalance*, that is, we find a partition $\widetilde{\mathcal{V}}$, which maintains all the desired properties and which has the property that $\sum_i |\widetilde{V}_{i,j}| = \sum_i n_{i,j}$ for each $j \in [k]$. To do this, we identify some $j^* \in [k]$ such that $\sum_i |V_{i,j^*}| > \sum_i n_{i,j^*}$ and some index $j' \in [k]$ such that $\sum_i |V_{i,j'}| < \sum_i n_{i,j'}$. We move $\big(\sum_i |V_{i,j^*}| - n_{i,j^*}\big)$ vertices from V_{1,j^*} to some cluster $V_{i',j'}$, maintaining the desired properties, and repeat this procedure until no global imbalance remains.

In a second step, we correct *local imbalance*, that is, for each $i = 1, \dots, r-1$ sequentially, and for each $j \in [k]$, we move vertices between $\widetilde{V}_{i,j}$ and $\widetilde{V}_{i+1,j}$, maintaining the desired properties, in order to obtain a partition $\mathcal{V}'$ such that $|V'_{i,j}| = n_{i,j}$ for each $i \in [r-1]$ and $j \in [k]$. Observe that, since $\widetilde{\mathcal{V}}$ is globally balanced, once we know that $|V'_{i,j}| = n_{i,j}$ holds for each $i \in [r-1]$ and $j \in [k]$, we are guaranteed that $|V'_{r,j}| = n_{r,j}$ holds as well for each $j \in [k]$.

The proof of the lemma then comes down to showing that we can move vertices and maintain the desired properties. Because we start with a partition in which $|V_{i,j}|$ is very close to $n_{i,j}$ for each $i \in [r]$ and $j \in [k]$, the total number of vertices we move in any step is at most the sum of the differences, which is much smaller than any $n_{i,j}$. The following lemma shows that we can move any small (compared to all $n_{i,j}$) number of vertices from one part to another and maintain the desired properties.

Lemma 3.12. *For all integers $k, r_1, \Delta \geq 1$, and reals $d > 0$ and $0 < \varepsilon < 1/2k$ as well as $0 < \xi < 1/(100kr_1^3)$, there exist $C^* > 0$ such that the following holds for sufficiently large n.*

Let Γ be a graph on vertex set $[n]$, and let G be a not necessarily spanning subgraph. Let $X, Z_1, \dots, Z_{k-1} \subseteq V(G)$ be pairwise disjoint subsets, each of size at least $n/(16kr_1)$, such that (X, Z_i) is $(\varepsilon, d, p)_G$-regular for each $i \in [k-1]$. Then for each $1 \leq m \leq 2r_1^2\xi n$, there exists a set S of m vertices of X with the following properties.

(SM1) For each $v \in S$ we have $\deg_G(v, Z_i) \geq (d-\varepsilon)p|Z_i|$ for each $i \in [k-1]$, and

(SM2) for each $1 \leq s \leq \Delta$ and every collection of vertices $v_1, \dots, v_s \in [n]$ we have

$$\deg_\Gamma(v_1, \dots, v_s; S) \leq 100kr_1^3\xi \deg_\Gamma(v_1, \dots, v_s; X) + \tfrac{1}{100}C^* \log n\,.$$

Proof. Given k, r_1, Δ, d, ξ and ε. Let C be returned by Lemma 2.19 for input ξ and Δ. We set $C^* = 100C$. Given Γ, G and X, Y, $Z_1, \dots, Z_{k-1}$, let X' be the set of vertices $v \in X$ such that $\deg_G(v; Z_i) \geq (d-\varepsilon)p|Z_i|$ for each $i \in [k-1]$. Because each pair (X, Z_i) for $i \in [k-1]$ is $(\varepsilon, d, p)_G$-regular, we have $|X'| \geq |X| - k\varepsilon|X| \geq |X|/2$.

We now apply Lemma 2.19, with input ξ, Δ, $W = X'$ and the sets T_i consisting of the sets $N_\Gamma(v_1, \dots, v_s; X')$ for each $1 \leq s \leq \Delta$ and $v_1, \dots, v_s \in [n]$, to choose a set S of size $m \leq 2r_1^2\xi n \leq |X'|$ in X'. We then have

$$\begin{aligned}\deg_\Gamma(v_1, \dots, v_s; S) &\leq \left(\tfrac{2r_1^2\xi n}{|X'|} + \xi\right) \deg_\Gamma(v_1, \dots, v_s; X') + C\log n \\ &\leq 100kr_1^3\xi \deg_\Gamma(v_1, \dots, v_s; X) + \tfrac{1}{100}C^* \log n\,,\end{aligned}$$

where the final inequality is by choice of C^*, and since $|X'| \geq |X|/2 \geq n/(32kr_1)$. Thus the set S satisfies (SM2), and since $S \subseteq X'$ we have (SM1). □

We now turn to the proof of the balancing lemma.

Proof of Lemma 3.7. Given integers $k, r_1, \Delta \geq 1$ as well as reals $\gamma, d > 0$ and $0 < \varepsilon < \min\{d, 1/(2k)\}$, we set

$$\xi = 10^{-15}\varepsilon^4 d/(k^3 r_1^5).$$

Let C_1^* be returned by Lemma 3.12 with input k, r_1, Δ, d, $\varepsilon/4$, and ξ, and let C_2^* be returned by Lemma 3.12 with input k, r_1, Δ, d, $3\varepsilon/4$, and ξ. We set $C^* = \max\{C_1^*, C_2^*\}$.

Now suppose that $p \geq C^*(\log n/n)^{1/2}$, that $10\gamma^{-1} \leq r \leq r_1$, and that graphs Γ and G, a partition $\mathcal{V}$ of $V = V(G)$, and graphs R_r^k, B_r^k and K_r^k on $[r] \times [k]$ as in the statement of Lemma 3.7 are given. We divide the proof into two stages.

First stage (global imbalance):

The goal of the first stage is to move vertices between clusters of $\mathcal{V}$ such that the partition $\widetilde{\mathcal{V}} = \{\widetilde{V}_{i,j}\}_{i\in[r],j\in[k]}$ that we obtain satisfies $\sum_{i\in[r]} |\widetilde{V}_{i,j}| = \sum_{i\in[r]} n_{i,j}$ for every $j \in [k]$, such that $\mathcal{V}$ is $(\varepsilon/2, d, p)_G$-regular on R_r^k and $(\varepsilon/2, d, p)_G$-super-regular on K_r^k, and such that the sizes of the clusters and the common Γ-neighbourhoods of any at most Δ vertices restricted to any cluster has not changed much.

We use the following algorithm to create a globally balanced partition $\widetilde{\mathcal{V}}$, i.e. for which $\sum_{i\in[r]} |\widetilde{V}_{i,j}| = \sum_{i\in[r]} n_{i,j}$ holds. In Claim 3.13 we show that the properties mentioned above hold for every partition that is the outcome of the algorithm if we select the set S in the fourth step of the 'While loop' cleverly. A cluster that has not been changed by Algorithm 1 before is referred to as 'unchanged'.

Algorithm 1: Global balancing

---

while $\exists j \in [k]$ *such that* $\sum_{i\in[r]}(|V_{i,j}| - n_{i,j}) \neq 0$ **do**
    Choose $j^* \in [k]$ maximising $\sum_{i\in[r]}(|V_{i,j^*}| - n_{i,j^*})$;
    Choose $i' > 1$ such that $V_{i',j}$ is 'unchanged' and $(V_{1,j^*}, V_{i',j})$ is $\big(\varepsilon/4, d, p\big)_G$-regular for every $j \in [k]$;
    Choose $j' \in [k]$ such that $\sum_{i\in[r]}(|V_{i,j'}| - n_{i,j'}) < 0$;
    Select $S \subseteq V_{1,j^*}$ with $|S| = \sum_{i\in[r]} |V_{i,j^*}| - n_{i,j^*}$;
    Update $V_{1,j^*} := V_{1,j^*} \setminus S$ and $V_{i',j'} = V_{i',j'} \cup S$;
    Flag V_{1,j^*} and $V_{i',j'}$ as 'changed' ;
end

We use Lemma 3.12 to select S with input k, r_1, Δ, d and $\varepsilon/4$, with $X = V_{1,j^*}$ and with the $Z_1, \ldots, Z_{k-1}$ being the $\{V_{i',j''}\}_{j''\in[k]\setminus\{j'\}}$ with the notation as in Algorithm 1. See Figure 3.3 for an illustration of Algorithm 1.

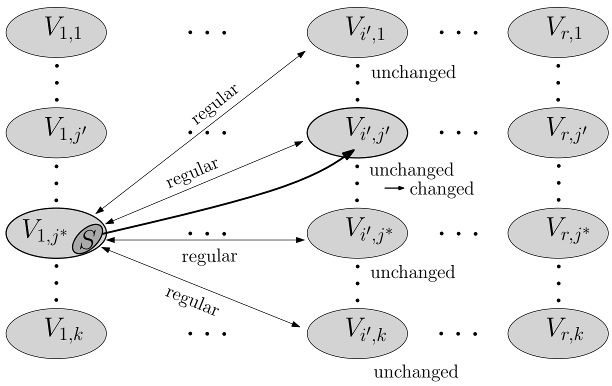

Figure 3.3: One iteration step in Algorithm 1, where vertices of V_{1,j^*} are moved to $V_{i',j'}$.

We claim that the algorithm completes successfully. In other words, we show that each of the choices is possible and that Lemma 3.12 is always applicable. In each 'While loop', since $\sum_{i\in[r],j\in[k]}(|V_{i,j}| - n_{i,j}) = 0$ and since the While condition is satisfied, the index $j^* \in [k]$ maximising $\sum_{i\in[r]}(|V_{i,j^*}| - n_{i,j})$ satisfies $\sum_{i\in[r]}(|V_{i,j^*}| - n_{i,j^*}) > 0$.

Observe that the 'While loop' is run at most k times since at the end of the 'While loop' we have achieved $\sum_{i\in[r]}(|V_{i,j}| - n_{i,j}) = 0$ for an index $j \in [k]$ and do not select this index in future iterations as j^* or j'. It follows that the number of clusters flagged as 'changed' never exceeds $2k$. Every vertex $(1, j^*) \in \{1\} \times [k]$ has degree at least $(k - 1 + \gamma k/2)r$ in R_r^k. Hence there are at least $\gamma kr/2$ indices $i \in [r]$ such that $(1, j^*)$ is adjacent to each (i, j) with $j \in [k]$ in R_r^k. Since $\gamma kr/2 > 3k$, we can choose $i' \in [r]$ in the second step of the 'While loop' such that $(1, j^*)$ is adjacent to each (i', j) with $j \in [k]$ in R_r^k and no $V_{i',j}$ is flagged as changed. Since $\{(1, j^*), (i', j)\} \in E(R_r^k)$ for each $j \in [k]$, each pair $(V_{1,j^*}, V_{i',j})$ is $\big(\varepsilon/4, d, p\big)_G$-

regular. Thus it is possible to choose i' in the second step of the 'While loop'. It is possible to choose $j' \in [k]$ in the third step of the 'While loop' since $\sum_{i\in[r],j\in[k]}(|V_{i,j}| - n_{i,j}) = 0$ and $\sum_{i\in[r]}(|V_{i,j^*}| - n_{i,j^*}) > 0$.

Finally, we need to show that Lemma 3.12 is always applicable with the given parameters. In each application, the sets denoted by $X, Z_1, \dots, Z_{k-1}$ are parts of the partition $\mathcal{V}$. This means that they have not been changed by the algorithm yet. It follows that each set has size at least $n/(8kr) > n/(16kr_1)$. Since $\mathcal{V}$ is $(\varepsilon/4, d, p)_G$-regular on B_r^k, the pairs $(X, Z_1), \dots, (X, Z_{k-1})$ are $(\varepsilon/4, d, p)_G$-regular as required. Finally, by choice of $j^* \in [k]$ we see that the sizes of the sets S that we select in each step are decreasing. Hence it is enough to show that in the first step we have $|S| \le r\xi n$, which follows from (B1). Thus Lemma 3.12 is applicable in each step, and we conclude that the algorithm indeed completes. We denote the resulting vertex partition by $\widetilde{\mathcal{V}} = \{\widetilde{V}_{i,j}\}_{i\in[r],j\in[k]}$.

Claim 3.13. *The following properties hold for $\widetilde{\mathcal{V}}$:*

(P1) $\sum_{i\in[r]} |\widetilde{V}_{i,j}| = \sum_{i\in[r]} n_{i,j}$,

(P2) for each $i \in [r]$ and $j \in [k]$ we have $\big||\widetilde{V}_{i,j}| - n_{i,j}\big| \le 2r\xi n$,

(P3) $\widetilde{\mathcal{V}}$ is $\left(\frac{\varepsilon}{2}, d, p\right)_G$-regular on R_r^k and $\left(\frac{\varepsilon}{2}, d, p\right)_G$-super-regular on K_r^k,

(P4) for each $i \in [r]$, $j \in [k]$ and $1 \le s \le \Delta$ and $v_1, \dots, v_s \in [n]$ we have

$$|N_\Gamma(v_1, \dots, v_s; \widetilde{V}_{i,j}) \triangle N_\Gamma(v_1, \dots, v_s; V_{i,j})| \le 100kr_1^3\xi \deg_\Gamma\big(v_1, \dots, v_s; V(G)\big) + \tfrac{1}{100}C^* \log n\,.$$

Proof. Property (P1) holds by construction of Algorithm 1.

Observe that vertices were removed from or added to each $V_{i,j}$ to form $\widetilde{V}_{i,j}$ at most once in the running of Algorithm 1, and the number of vertices added or removed was at most $r\xi n$. Since $|V_{i,j}|$ satisfies (B1), we conclude that (P2) holds. Furthermore, the vertices added to or removed from $V_{i,j}$ satisfy (SM2) and therefore (P4) holds.

Since $|V_{i,j}| \ge n/(8kr)$ for each $i \in [r]$ and $j \in [k]$, we can apply Proposition 2.4 with $\mu = \nu = 8kr^2\xi$ to each edge of R_r^k, concluding that $\widetilde{\mathcal{V}}$ is $(\varepsilon/2, d, p)_G$-regular on R_r^k since $\varepsilon/4 + 4\sqrt{8kr^2\xi} < \varepsilon/2$. Now for any $i \in [r]$ and $j \in [k]$, consider $v \in \widetilde{V}_{i,j}$. If $v \notin V_{i,j}$, then we applied Lemma 3.12 to select v, and at that time no $V_{i,j'}$ with $j' \in [k]$ was flagged as changed by Algorithm 1. Thus by (SM1) we have

$$\deg_G(v; \widetilde{V}_{i,j'}) = \deg_G(v; V_{i,j'}) \ge \big(d - \tfrac{\varepsilon}{4}\big)p|V_{i,j'}| = \big(d - \tfrac{\varepsilon}{4}\big)p|\widetilde{V}_{i,j'}|$$

for each $j' \in [k] \setminus \{j\}$ since $V_{i,j}$ is then flagged as changed and thus $V_{i,j'} = \widetilde{V}_{i,j'}$ for each $j' \in [k] \setminus \{j\}$. If on the other hand $v \in V_{i,j}$, then by (B2) we started with $\deg_G(v; V_{i,j'}) \ge (d - \varepsilon/4)p|V_{i,j'}|$. By (SM2) and (B4), we have

$$\deg_G(v; \widetilde{V}_{i,j'}) \ge \big(d - \tfrac{\varepsilon}{4}\big)p|V_{i,j'}| - \tfrac{\varepsilon^2}{1000kr_1}\big(1 + \tfrac{\varepsilon}{4}\big)p|V_{i,j'}| - \tfrac{1}{100}C^* \log n \ge \big(d - \tfrac{\varepsilon}{2}\big)p|\widetilde{V}_{i,j'}|\,,$$

where the final inequality follows since $|\widetilde{V}_{i,j'}| \le |V_{i,j'}| + r\xi n \le (1 + \varepsilon d/100)|V_{i,j'}|$ and since n was assumed to be sufficiently large. Therefore, $\widetilde{\mathcal{V}}$ is $(\varepsilon/2, d, p)$-super-regular on K_r^k, giving (P3). □

Second stage (local imbalance):

The first stage resulted in a partition $\widetilde{\mathcal{V}}$ with Properties (P1)–(P4) of Claim 3.13. In particular, $\widetilde{\mathcal{V}}$ is globally balanced, i.e. $\sum_{i\in[r]} |\widetilde{V}_{i,j}| = \sum_{i\in[r]} n_{i,j}$ for every $j \in [k]$. The goal of the second stage is to obtain a balanced partition $\mathcal{V}$ with the desired properties of the lemma. We use the following algorithm to correct the local imbalances in $\widetilde{\mathcal{V}}$.

Algorithm 2: Local balancing

---

foreach $i = 1, \dots, r-1$ **do**
    foreach $j = 1, \dots, k$ **do**
        if $|\widetilde{V}_{i,j}| > n_{i,j}$ **then**
            Select $S \subseteq \widetilde{V}_{i,j}$ with $|S| = |\widetilde{V}_{i,j}| - n_{i,j}$;
            Update $\widetilde{V}_{i,j} := \widetilde{V}_{i,j} \setminus S$ and $\widetilde{V}_{i+1,j} := \widetilde{V}_{i+1,j} \cup S$;
        end
        else
            Select $S \subseteq \widetilde{V}_{i+1,j}$ with $|S| = n_{i,j} - |\widetilde{V}_{i,j}|$;
            Update $\widetilde{V}_{i+1,j} := \widetilde{V}_{i+1,j} \setminus S$ and $\widetilde{V}_{i,j} := \widetilde{V}_{i,j} \cup S$;
        end
    end
end

Again, in each step when we select S we make use of Lemma 3.12 to do so. If we select S from $\widetilde{V}_{i,j}$, then we use the input k, r_1, d and $3\varepsilon/4$ with $X = \widetilde{V}_{i,j}$ and the sets $Z_1, \dots, Z_{k-1}$ being $\{\widetilde{V}_{i+1,j'}\}_{j'\in[r]\setminus\{j\}}$. If on the other hand we select S from $\widetilde{V}_{i+1,j}$, then we use the input k, r_1, d and $3\varepsilon/4$, with $X = \widetilde{V}_{i+1,j}$ and the sets $Z_1, \dots, Z_{k-1}$ being $\{\widetilde{V}_{i,j'}\}_{j'\in[r]\setminus\{j\}}$. In Figure 3.4 we sketch one iteration step of Algorithm 2.

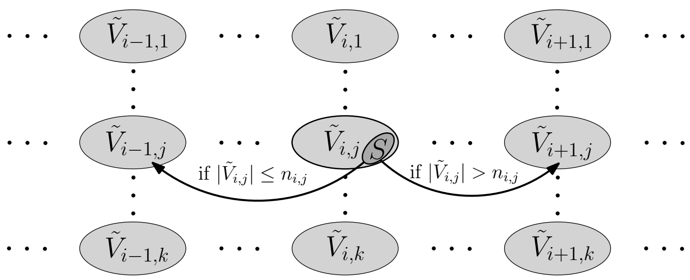

Figure 3.4: One iteration step in Algorithm 2, where vertices of $V_{i,j}$ are either moved to $V_{i-1,j}$ or to $V_{i+1,j}$.

We claim that Lemma 3.12 is always applicable. To see that this is true, observe first that the number of vertices that we move between any $\widetilde{V}_{i,j}$ and $\widetilde{V}_{i+1,j}$ in a given step is, thanks to (P2), bounded by $2r^2\xi n$. We change any given $\widetilde{V}_{i,j}$ at most twice in the running of the algorithm, so that in total at most $4r^2\xi n$ vertices are changed. In particular, we maintain $|\widetilde{V}_{i,j}| \ge n/(16kr_1)$ throughout. By Proposition 2.4, with input $\mu = \nu = 4r^2\xi n/\big(n/(16kr_1)\big) < 100r_1^3k\xi$, and by (P3) we maintain the property that any pair in R_r^k, and in particular any pair in B_r^k, is $(3\varepsilon/4, d, p)$-regular throughout. This shows that Lemma 3.12 is always applicable, and therefore the algorithm completes and returns a partition $\mathcal{V}' = \{V'_{i,j}\}_{i\in[r],j\in[k]}$.

We claim that $\mathcal{V}'$ is the desired partition. To this end, we need to verify that Properties (B1')–(B5') hold.

As for each $j \in [k]$ we have $\sum_{i\in[r]} |V'_{i,j}| = \sum_{i\in[r]} |\widetilde{V}_{i,j}| = \sum_{i\in[r]} n_{i,j}$, and as $|V'_{i,j}| = n_{i,j}$ for each $i \in [r-1]$ and $j \in [k]$, we conclude that $|V'_{i,j}| = n_{i,j}$ for all $i \in [r]$ and $j \in [k]$, giving (B1').

For the first part of (B3'), we have justified that we maintain $(3\varepsilon/4, d, p)_G$-regularity on R^k_r throughout the algorithm. For the second part, we need to show that for each $i \in [r]$ and distinct indices $j, j' \in [k]$, and each $v \in V'_{i,j}$, we have $\deg_G(v; V'_{i,j'}) \geq (d-\varepsilon)p|V'_{i,j'}|$. If $v \in \widetilde{V}_{i,j}$, then by (P3) we have $\deg_G(v; \widetilde{V}_{i,j'}) \geq (d-\varepsilon/2)p|\widetilde{V}_{i,j'}|$. We change $\widetilde{V}_{i,j'}$ at most twice to obtain $V'_{i,j'}$, both times by adding or removing vertices satisfying (SM2). As in the proof of Claim (P4) above, using (B4) and (P4) we obtain $\deg_G(v; \widetilde{V}_{i,j'}) \geq (d-\varepsilon)p|V'_{i,j}|$ as desired. If $v \notin \widetilde{V}_{i,j}$, then it was added to the set $\widetilde{V}_{i,j}$ by Algorithm 2, and $\widetilde{V}_{i,j'}$ was changed at most twice thereafter. Again, using (SM1), (SM2), (B4), and (P4) we obtain $\deg_G(v; \widetilde{V}_{i,j'}) \geq (d-\varepsilon)p|V'_{i,j}|$ as desired.

In Algorithm 1 at most $r\xi n$ vertices were removed from or added to $V_{i,j}$. Hence, we have $|V_{i,j} \triangle \widetilde{V}_{i,j}| \leq r\xi n$. In Algorithm 2 at most $4r^2\xi n$ are removed from or added to $\widetilde{V}_{i,j}$. By choice of ξ this implies $|V_{i,j} \triangle V'_{i,j'}| \leq 5r^2\xi n \leq 10^{-10}\varepsilon^4 k^{-2} r_1^{-2} n$, which is Property (B2').

To see that Property (B4') holds, observe thatthat for any given $i \in [r]$ and $j \in [k]$ we change $\widetilde{V}_{i,j}$ at most twice in the running of Algorithm 2, both times either adding or removing a set satisfying (SM2). Hence by (B4), (P4), (SM2), and by choice of ξ we have

$$\big|N_\Gamma(v; V_{i,j}) \Delta N_\Gamma(v; V'_{i,j})\big| \leq \tfrac{\varepsilon^2}{1000kr_1} \deg_\Gamma\big(v; V(G)\big) + \tfrac{1}{10} C^* \log n \leq \tfrac{\varepsilon^2}{100} \deg_\Gamma(v; V_{i,j})$$

where the final inequality follows by choice of p and of n sufficiently large. Using (B3), we can apply Proposition 2.4, with $\mu = \nu = \varepsilon^2/100$, to deduce (B4'). By a similar calculation Property (B5') holds.

Finally, suppose that for any two disjoint vertex sets $A, A' \subseteq V(\Gamma)$ with $|A|, |A'| \geq \varepsilon^2\xi pn/(50000kr_1)$ we have $e_\Gamma(A, A') \leq (1 + \varepsilon^2\xi/100)p|A||A'|$. In each application of Proposition 2.4 we have $\mu, \nu \geq \varepsilon^2\xi/200$, and if we have 'fully-regular' in place of 'regular' in (B2) and (B3), we always apply Proposition 2.4 to a fully-regular pair with sets of size at least $\varepsilon pn/(1000kr_1)$, so it returns fully-regular pairs for (B3') and (B4'), as desired. □

3.1.7 Proof of the main theorem

In this subsection we present the proofs of Theorem 3.3 and Theorem 3.1, where the latter is a corollary of the first one.

Before we start with the proof of Theorem 3.3, we first sketch the main ideas. Given the graph G, we first use the lemma for G (Lemma 3.4) to find a regular partition of $V(G)$ with a small exceptional set V_0 and such that its reduced graph R^k_r contains a spanning backbone graph B^k_r, on whose subgraph K^k_r the graph G is super-regular and has one- and two-sided inheritance. Given this, and H together with a (z, β)-zero-free $(k+1)$-colouring, we use the lemma for H (Lemma 3.5) to find a homomorphism f from $V(H)$ to R^k_r almost all of whose edges are mapped to K^k_r and in which approximately the 'right' number of vertices of H are mapped to each vertex of R^k_r. At this point, if V_0 were empty, and if the 'approximately' were exact, we would apply the sparse blow-up lemma (Theorem 2.9) to obtain an embedding of H into G.

We resolve the first of these problems in the following way. Given $v \in V_0$, we choose $x \in V(H)$ that is not in any triangles and that is far from any vertices of colour zero, and embed x to v. We then embed the neighbours of x to carefully chosen neighbours of v, which we obtain using the common neighbourhood lemma (Lemma 3.6). Here we use the fact that $N_H(x)$ is independent. This then fixes a clique of K_r^k to which $N_H^2(x)$ must be assigned, and gives image restrictions in the corresponding parts of the regular partition for these vertices. Since $N_H^2(x)$ may have been assigned by f to a different component of K_r^k, we have to adjust f to match. This will not cause any problems since x is far from vertices of colour zero.

Now the idea is to repeat the above procedure, choosing vertices of $V(H)$ to pre-embed that are widely separated in H, until we pre-embedded vertices to all vertices of V_0. We end up with a homomorphism f^* from what the remainder of $V(H)$ to R_r^k. This homomorphism still maps about the right number of vertices of H to each vertex of R_r^k since V_0 is small and each vertex has at most Δ neighbours. We now apply the balancing lemma (Lemma 3.7) to correct the sizes of the clusters of the partition of the remainder of $V(G)$ to match f^*, and complete the embedding of H using the sparse blow-up lemma (Theorem 2.9).

However, there is one difficulty with this idea. Because we perform the pre-embedding sequentially, we might use up a significant fraction of $N_G(w)$ for some $w \in V(G)$ in the pre-embedding, destroying super-regularity of G on K_r^k, or we might use up a significant fraction of some common neighbourhood that defines an image restriction for the sparse blow-up lemma. In order to avoid this, before we begin the pre-embedding, we fix a set $S \subseteq V(G)$ whose size is a small linear fraction of n. We choose S using Lemma 2.19 such that S does not have a large intersection with any common G-neighbourhood of at most Δ vertices (which could define an image restriction). We perform the pre-embedding as outlined above, except that we choose our neighbours of each $v \in V_0$ within S. We show that this procedure does not destroy super-regularity or use up image restriction sets. In order to not having to choose at some point a vertex $v \in V_0$ whose neighbourhood in S was already used up in the pre-embedding process, we choose vertices of V_0 first that have not many unused neighbours in S left.

Proof of Theorem 3.3. Let $\gamma > 0$, $\Delta \geq 2$, and $k \geq 2$ be given. Set

$$r_0 = 10/\gamma \text{ and } D = \Delta.$$

Let d be returned by Lemma 3.4, with input γ, k, and r_0. Let α be returned by Lemma 3.6 with input d, k, and Δ. Now let $\varepsilon_{\text{BL}} > 0$ and $\rho > 0$ be returned by Theorem 2.9 with input Δ, $\Delta_{R'} = 3k$, $\Delta_J = \Delta$, $\vartheta = 1/(100D)$, $\zeta = \alpha/4$, d and $\kappa := 64$. Next, putting $\varepsilon^* = \varepsilon_{\text{BL}}/8$ into Lemma 3.6 returns $\varepsilon_0 > 0$. We set

$$\varepsilon = \min\big\{\varepsilon_0, d/8, \varepsilon^*/(4D), 1/(16k)\big\}.$$

Putting ε into Lemma 3.4 returns r_1 and C_1^*. Next, Lemma 3.7, for input k, r_1, Δ, γ, d, and 8ε, returns $\xi \in \big(0, 1/(10kr_1)\big)$ and C_2^*. We set

$$\beta = 10^{-12}\xi^2/(\Delta k^4 r_1^2) \text{ and } \mu = 10^{-5}\varepsilon^2/(kr_1).$$

Let C_3^* be the maximum of the C-outputs of Theorem 2.9 with input r_1, of Proposition 3.8 with input ε and with input μ^2, and of Lemma 2.19 with input $\varepsilon\mu$ and Δ. Finally, let $C^* = \max\{C_1^*, C_2^*, C_3^*\}$ and set

$$C = 10^{10}k^2 r_1^2 \Delta^{2r_1+20} C^*/(\varepsilon^2\xi\mu^2) \text{ and } z = 10/\xi.$$

Let $p \geq C(\log n/n)^{1/\Delta}$. Then a.a.s. $G(n,p)$ satisfies the good events of Theorem 2.9, Lemmas 3.4 and 3.6, and Proposition 3.8, with the parameters stated above. Suppose that $\Gamma = G(n,p)$ satisfies these good events.

Let $G \subseteq \Gamma$ be a spanning subgraph with $\delta(G) \geq \big((k-1)/k+\gamma\big)pn$. Furthermore, let H be a graph on n vertices with $\Delta(H) \leq \Delta$ and $\mathcal{L}$ be a labelling of the vertex set of H of bandwidth at most βn such that the first βn vertices of $\mathcal{L}$ include $C\max\{p^{-2}, p^{-1}\log n\}$ vertices that are not contained in any triangles of H, and such that there exists a $(k+1)$-colouring that is (z,β)-zero-free with respect to $\mathcal{L}$, and the colour zero is not assigned to the first $\sqrt{\beta}n$ vertices.

Applying Lemma 3.4 to G, with input γ, k, r_0 and ε, we obtain an integer r with $10\gamma^{-1} \leq kr \leq r_1$, a set $V_0 \subseteq V(G)$ with $|V_0| \leq C^* \max\{p^{-2}, p^{-1}\log n\}$, a k-equitable partition $\mathcal{V} = \{V_{i,j}\}_{i\in[r],j\in[k]}$ of $V(G) \setminus V_0$, and a graph R_r^k on vertex set $[r] \times [k]$ with minimum degree $\delta(R_r^k) \geq \big((k-1)/k+\gamma/2\big)kr$, such that $K_r^k \subseteq B_r^k \subseteq R_r^k$, and such that

- (G1a) $\frac{n}{4kr} \leq |V_{i,j}| \leq \frac{4n}{kr}$ for every $i \in [r]$ and $j \in [k]$,
- (G2a) $\mathcal{V}$ is $(\varepsilon,d,p)_G$-regular on R_r^k and $(\varepsilon,d,p)_G$-super-regular on K_r^k,
- (G3a) both $\big(N_\Gamma(v,V_{i,j}), V_{i',j'}\big)$ and $\big(N_\Gamma(v,V_{i,j}), N_\Gamma(v,V_{i',j'})\big)$ are $(\varepsilon,d,p)_G$-regular pairs for every $\{(i,j),(i',j')\} \in E(R_r^k)$ and $v \in V \setminus V_0$, and
- (G4a) $|N_\Gamma(v,V_{i,j})| = (1\pm\varepsilon)p|V_{i,j}|$ for every $i \in [r]$, $j \in [k]$ and every $v \in V \setminus V_0$.

For every $i \in [r]$ and $j \in [k]$ we choose

$$m_{i,j} \in \left\{|V_{i,j}| + \left\lfloor \tfrac{1}{kr}|V_0|\right\rfloor, |V_{i,j}| + \left\lceil \tfrac{1}{kr}|V_0|\right\rceil \right\}$$

such that $\{m_{i,j}\}_{i\in[r],j\in[k]}$ forms a k-equitable integer partition of n.

Since $\delta(R_r^k) > (k-1)r$, there exists for every $i \in [r]$ a vertex $v \in V(R_r^k)$ that is adjacent to each (i,j) with $j \in [k]$. This together with our assumptions on H allow us to apply Lemma 3.5 to H, with input D, k, r, $\xi/10$, β, and $\{m_{i,j}\}_{i\in[r],j\in[k]}$. Note that since $\Delta(H) \leq \Delta$, the graph H is in particular Δ-degenerate.

Let $f\colon V(H) \to [r] \times [k]$ be the mapping returned by Lemma 3.5, let $W_{i,j} := f^{-1}(i,j)$, and let $X \subseteq V(H)$ be the set of special vertices returned by Lemma 3.5. For every $i \in [r]$ and $j \in [k]$ we have

- (H1a) $m_{i,j} - \frac{1}{10}\xi n \leq |W_{i,j}| \leq m_{i,j} + \frac{1}{10}\xi n$,
- (H2a) $|X| \leq \xi n$,
- (H3a) $\{f(x), f(y)\} \in E(R_r^k)$ for every $\{x,y\} \in E(H)$,
- (H4a) $y, z \in \bigcup_{j'\in[k]} f^{-1}(i,j')$ for every $x \in f^{-1}(i,j) \setminus X$ and $\{x,y\},\{y,z\} \in E(H)$, and
- (H5a) $f(x) = \big(1,\sigma(x)\big)$ for every x in the first $\sqrt{\beta}n$ vertices of $\mathcal{L}$.

Lemma 3.5 actually guarantees more, which we do not require for this proof. We let F be the first βn vertices of $\mathcal{L}$. By definition of $\mathcal{L}$, in F there are at least $C\max\{p^{-2}, p^{-1}\log n\}$ vertices whose neighbourhood in H is independent.

Next, we apply Lemma 2.19, with input $\varepsilon\mu$ and Δ, to choose a set $S \subseteq V(G)$ of size μn. We let the T_i of Lemma 2.19 be all sets that are common neighbourhoods in Γ of at most Δ vertices of Γ, together with the G-neighbourhoods of each vertex in G and together with the sets $V_{i,j}$ for $i \in [r]$ and $j \in [k]$. The result of Lemma 2.19 is that for any $1 \le \ell \le \Delta$ and vertices $u_1, \ldots, u_\ell$ of $V(G)$, we have for each $i \in [r]$ and $j \in [k]$ and $v \in V(G)$ that

$$\begin{aligned} \Big|S \cap \bigcap_{1 \le i' \le \ell} N_\Gamma(u_{i'})\Big| &= \mu\Big|\bigcap_{1 \le i' \le \ell} N_\Gamma(u_{i'})\Big| \pm \mu\Big(\varepsilon\Big|\bigcap_{1 \le i' \le \ell} N_\Gamma(u_{i'})\Big| + p^\ell n\Big)\,, \\ |S \cap N_G(v)| = \mu|N_G(v)| &\pm 2\mu\varepsilon|N_G(v)| \quad \text{and} \quad \big|S \cap V_{i,j}\big| \le 2\mu|V_{i,j}| \end{aligned} \tag{3.15}$$

where we use the fact $p \ge C(\log n/n)^{1/\Delta}$ and the choice of C to deduce $C^* \log n < \mu p^\Delta n$.

Our next task is to create the pre-embedding that covers the vertices of V_0. We use the following algorithm, starting with ϕ_0 as the empty partial embedding.

Algorithm 3: Pre-embedding

Set $t := 0$;
while $V_0 \setminus \mathrm{im}(\phi_t) \neq \varnothing$ **do**
1 Let $v_{t+1} \in V_0 \setminus \mathrm{im}(\phi_t)$ minimise $\big|\big(N_G(v) \cap S\big) \setminus \mathrm{im}(\phi_t)\big|$ over $v \in V_0 \setminus \mathrm{im}(\phi_t)$;
Choose $x_{t+1} \in F$ with $N_H(x)$ independent, with $\mathrm{dist}\big(x_{t+1}, \mathrm{dom}(\phi_t)\big) \ge 2r + 20$;
Let $\ell = |N_H(x_{t+1})|$ and $\{y_1, \ldots, y_\ell\} = N_H(x_{t+1})$;
2 Choose $w_1, \ldots, w_\ell \in \big(N_G(v) \cap S\big) \setminus \mathrm{im}(\phi_t)$;
Set $\phi_{t+1} = \phi_t \cup \{x_{t+1} \to v_{t+1}\} \cup \{y_1 \to w_1\} \cup \cdots \cup \{y_\ell \to w_\ell\}$;
Update $t := t + 1$;
end

Suppose this algorithm does not fail, terminating with $t = t_f$. The final ϕ_{t_f} is an embedding of some vertices of H into $V(G)$ that covers V_0 and is contained in $V_0 \cup S$. Before we specify how exactly we choose vertices at line 2, we justify that the algorithm does not fail. In other words, we need to verify that at every time t there are vertices of F whose neighbourhood is independent and that are not close to any vertices in $\mathrm{dom}(\phi_t)$, and that at every time t, the set $\big(N_G(v) \cap S\big) \setminus \mathrm{im}(\phi_t)$ is 'big'.

For the first, observe that since $|V_0| \le C^* \max\{p^{-2}, p^{-1} \log n\}$, we have $\mathrm{dom}(\phi_t) \le C^* \Delta \max\{p^{-2}, p^{-1} \log n\}$ at every step. Thus the number of vertices at distance less than $2r + 20$ from $\mathrm{dom}(\phi_t)$ is at most

$$\big(1 + \Delta + \cdots + \Delta^{2r+19}\big) C^* \Delta \max\{p^{-2}, p^{-1} \log n\} < 2C^* \Delta^{2r+20} \max\{p^{-2}, p^{-1} \log n\},$$

which by choice of C is smaller than the number of vertices in F with $N_H(x)$ independent, as there are at least $C \max\{p^{-2}, p^{-1} \log n\}$ of them.

For the second part, let t be the minimum number such that at time t a vertex v with $\big|\big(N_G(v) \cap S\big) \setminus \mathrm{im}(\phi_t)\big| < \mu^2 pn/4$ is picked. For all integers t' with $t - \frac{1}{4(\Delta+1)}\mu^2 pn \le t' < t$, we have $\big|\big(N_G(v) \cap S\big) \setminus \mathrm{im}(\phi_{t'})\big| < \mu^2 pn/2$ since $\big|\big(N_G(v) \cap S\big) \setminus \mathrm{im}(\phi_{t'})\big|$ can decrease by at most $\Delta+1$ in each step. As v was not picked for any $t - \frac{1}{4(\Delta+1)}\mu^2 pn \le t' < t$, the vertex that was chosen for t' had at most as many uncovered G-neighbours in S as v. Let Z be the set of vertices chosen at line 1 in each of these time steps. Then for each $z \in Z$ we have $\big|\big(N_G(v) \cap S\big) \setminus \mathrm{im}(\phi_t)\big| \le \mu^2 pn/2$. But by (3.15) we have $\big|N_G(z) \cap S\big| \ge 3\mu^2 pn/4$, so $\big|N_G(z) \cap \mathrm{im}(\phi_t)\big| \ge \mu^2 pn/4$ for each $z \in Z$. By choice of C, we have $|Z| = \frac{1}{4(\Delta+1)}\mu^2 pn \ge C^* p^{-1} \log n$. Since $|\mathrm{im}(\phi)| \le (\Delta+1)|V_0| \le \mu^2 n/(8r)$, by choice of C, this contradicts the good event of Proposition 3.8. Hence at each time we reach line 2 there are at least $\mu pn/4$ vertices of $\big(N_G(v) \cap S\big) \setminus \mathrm{im}(\phi)$ to choose from. In particular we verified that Algorithm 3 completes.

In order to specify how to choose the vertices $\{w_1, \dots, w_\ell\}$ in Algorithm 3, we need the following claim.

Claim 3.14. *Given any set Y of $\mu^2 pn/4$ vertices of $V(G)$, there exists $W \subseteq Y$ of size at least $\mu^2 pn/(8r)$ and an index $i \in [r]$ with the following property. For each $w \in W$ and each $j \in [k]$, we have $|N_G(w, V_{i,j})| \geq dp|V_{i,j}|$.*

Proof. First let Y' be obtained from Y by removing all vertices $y \in Y$ such that $|N_\Gamma(y, V_0)| \geq \varepsilon pn$, or for some $i \in [r]$ and $j \in [k]$ we have $\big|N_\Gamma(y, V_{i,j})\big| \neq (1 \pm \varepsilon)p|V_{i,j}|$. Because the good event of Proposition 3.8 occurs, the total number of vertices removed is at most $2krC^* p^{-1} \log n < |Y|/2$, where the inequality is by choice of C. Now given any $y \in Y'$, if for each $i \in [r]$ there exists an index $j \in [k]$ such that $\big|N_G(y, V_{i,j})\big| < dp|V_{i,j}|$, then, since $\{V_{i,j}\}_{i \in [r], j \in [k]}$ is k-equitable, we have

$$|N_G(y)| \leq \varepsilon pn + dpn + (1+\varepsilon)\tfrac{k-1}{k} pn + r < \big(\tfrac{k-1}{k} + \gamma\big)pn,$$

a contradiction. We conclude that for each $y \in Y'$ there exists $i_y \in [r]$ such that $|N_G(y, V_{i,j})| \geq dp|V_{i,j}|$ for each $j \in [k]$. We let W be a maximum subset of Y' such that $i_y = i_{y'}$ for all $y, y' \in W$. By construction we have $|W| \geq \mu^2 pn/(8r)$ and $|N_G(w, V_{i,j})| \geq dp|V_{i,j}|$ for each $w \in W$ and each $j \in [k]$. □

Now we describe how we choose the vertices $w_1, \dots, w_\ell$ at each time t in line 2 of Algorithm 3.

Let $Y = \big(N_G(v_t) \cap S\big) \setminus \mathrm{im}(\phi_t)$. Let $i_t \in [r]$ be an index, and $W_t \subseteq Y$ be a set of size $\mu^2 pn/(8r)$ such that $\big|N_G(w, V_{i_t,j})\big| \geq dpn|V_{i_t,j}|$ for each $j \in [k]$ and $w \in W_t$, whose existence is guaranteed by Claim 3.14. By construction, and by our choice of μ, we can apply Lemma 3.6 with input d, k, Δ, ε^*, r and ε, with the clusters $\big\{V_{i_t,j}\big\}_{j \in [k]}$ as the $\big\{V_i\big\}_{i \in [k]}$, and inputting a subset of W_t of size $10^{-10}\varepsilon^4 pn/(k^4 r^4)$ as required for (V3). To verify the conditions of Lemma 3.6, observe that (V1) follows from (G1a), (V2) from (G2a), and (V4) from Claim 3.14. We obtain a Δ-tuple of vertices in W_t satisfying (W1)–(W4) of Lemma 3.6. We let

$$w_1, \dots, w_\ell \in W_t \subseteq \big(N_G(v_t) \cap S\big) \setminus \mathrm{im}(\phi_t)$$

be the first ℓ vertices of this tuple, where $\ell := |N_H(x_t)|$.

Let $H' = H - \mathrm{dom}(\phi_{t_f})$. We next define image restricting vertex sets and create an updated homomorphism $f^* : V(H') \to [r] \times [k]$. For each $x \in V(H')$, set

$$J_x = \phi_{t_f}\big(N_H(x) \cap \mathrm{dom}(\phi_{t_f})\big).$$

Since the vertices $\{x_t\}_{t \in [t_f]}$ are by construction at pairwise distance at least $2r + 20$, in particular for each $y \in V(H')$ with $J_y \neq \varnothing$ there exists an index $t' \in [t_f]$ such that y is at distance two from $x_{t'}$, and at distance greater than $r + 10$ from all x_t with $t \in [t_f] \setminus \{t'\}$. In particular this means that J_y is a subset of the set $\{w_1, \dots, w_\ell\}$ that was chosen at time t'. Let $j \in [k]$ such that $f(y) = (1, j)$. Then we set

$$f^*(y) = (i_{t'}, j),$$

where $i_{t'}$ is defined as in the latter paragraph.

Next, for each $t \in [t_f]$ and each $z \in V(H)$ at distance at least 3 and at most $i_t + 1$ from x_t, we set $f^*(z)$ as follows. Recall that $f(z) = (1, j)$ for some $j \in [k]$. We set

$$f^*(z) = \big(i_t + 2 - \mathrm{dist}(x_t, z), j\big).$$

Since $\{x_t\}_{t \in [t_f]}$ are at pairwise distance at least $2r + 20$, no vertex is at distance $r + 5$ or less from any two x_t and $x_{t'}$. This means that f^* is well-defined. Because R^k_r contains B^k_r, the function f^* we constructed so far is a graph homomorphism. Furthermore, for each $t \in [t_f]$ the set of vertices z at distance $i_t + 1$ from x_t are in the first $\sqrt{\beta}n$ vertices of $\mathcal{L}$, and so by (H5a) satisfy $f^*(z) = f(z)$. We complete the construction of f^* by setting $f^*(z) = f(z)$ for each remaining $z \in V(H')$. Because f is a graph homomorphism, f^* is also a graph homomorphism whose domain is $V(H')$.

For each $i \in [r]$ and $j \in [k]$, let

$$W'_{i,j} = (f^*)^{-1}(V_{i,j})$$

and let

$$X' = X \cup \{x \in H' : \exists t \in [t_f] \text{ such that } \mathrm{dist}(x, x_t) \leq r + 10\}.$$

The total number of vertices $z \in V(H)$ at distance at most $r + 10$ from some x_t is at most $2\Delta^{r+10}|V_0| < \xi n/100$. Since $W_{i,j} \triangle W'_{i,j}$ contains only such vertices, we have

(H1b) $m_{i,j} - \frac{1}{5}\xi n \leq |W'_{i,j}| \leq m_{i,j} + \frac{1}{5}\xi n$,

(H2b) $|X'| \leq 2\xi n$,

(H3b) $\{f^*(x), f^*(y)\} \in E(R^k_r)$ for every $\{x, y\} \in E(H')$, and

(H4b) $y, z \in \bigcup_{j' \in [k]} W'_{i,j'}$ for every $x \in W'_{i,j} \setminus X'$ and $\{x, y\}, \{y, z\} \in E(H')$.

where we used Properties (H2a) and (H4a) as well as the definitions of X' and f^*.

Furthermore, we have the following properties, where (G1a)–(G4a) are repeated for convenience.

(G1a) $\frac{n}{4kr} \leq |V_{i,j}| \leq \frac{4n}{kr}$ for every $i \in [r]$ and $j \in [k]$,

(G2a) $\mathcal{V}$ is $(\varepsilon, d, p)_G$-regular on R^k_r and $(\varepsilon, d, p)_G$-super-regular on K^k_r,

(G3a) both $\big(N_\Gamma(v, V_{i,j}), V_{i',j'}\big)$ and $\big(N_\Gamma(v, V_{i,j}), N_\Gamma(v, V_{i',j'})\big)$ are $(\varepsilon, d, p)_G$-regular pairs for every $\{(i, j), (i', j')\} \in E(R^k_r)$ and $v \in V \setminus V_0$,

(G4a) $|N_\Gamma(v, V_{i,j})| = (1 \pm \varepsilon)p|V_{i,j}|$ for every $i \in [r]$, $j \in [k]$ and every $v \in V \setminus V_0$,

(G5a) $\big|\bigcap_{u \in J_x} N_G(u, V_{f^*(x)})\big| \geq \alpha p^{|J_x|}|V_{f^*(x)}|$ for each $x \in V(H')$,

(G6a) $\big|\bigcap_{u \in J_x} N_\Gamma(u, V_{f^*(x)})\big| = (1 \pm \varepsilon^*)p^{|J_x|}|V_{f^*(x)}|$ for each $x \in V(H')$,

(G7a) $\big(\bigcap_{u \in J_x} N_\Gamma(u, V_{f^*(x)}), \bigcap_{v \in J_y} N_\Gamma(v, V_{f^*(y)})\big)$ is an $(\varepsilon^*, d, p)_G$-regular pair for each edge $\{x, y\} \in E(H')$, and

(G8a) $\big|\bigcap_{u \in J_x} N_\Gamma(u)\big| \leq (1 + \varepsilon^*)p^{|J_x|}n$ for each $x \in V(H')$.

Properties (G5a), (G6a) and (G8a) are trivial when $J_x = \varnothing$, and are otherwise guaranteed by Lemma 3.6 since for every J_x there exists $t \in [t_f]$ such that J_x is a subset of the set $\{w_1, \ldots, w_\ell\}$ that was chosen at time t. Finally (G7a) follows from (G2a) when $J_x, J_y = \varnothing$, and otherwise is also guaranteed by Lemma 3.6.

For each $i \in [r]$ and $j \in [k]$, set

$$V'_{i,j} = V_{i,j} \setminus \mathrm{im}(\phi_{t_f})$$

and let $\mathcal{V}' = \{V'_{i,j}\}_{i\in[r],j\in[k]}$. Recall that $|\mathrm{im}(\phi_{t_f})| \le C^*\Delta \max\{p^{-2}, p^{-1}\log n\} \le n/(12kr)$. Since also $V_{i,j} \setminus V'_{i,j} \subseteq S$ for each $i \in [r]$ and $j \in [k]$, we obtain by Equation (3.15), Proposition 2.4 the following properties:

(G1b) $\frac{n}{6kr} \le |V'_{i,j}| \le \frac{6n}{kr}$ for every $i \in [r]$ and $j \in [k]$,

(G2b) $\mathcal{V}'$ is $(2\varepsilon, d, p)_G$-regular on R^k_r and $(2\varepsilon, d, p)_G$-super-regular on K^k_r,

(G3b) both $\big(N_\Gamma(v, V'_{i,j}), V'_{i',j'}\big)$ and $\big(N_\Gamma(v, V'_{i,j}), N_\Gamma(v, V'_{i',j'})\big)$ are $(2\varepsilon, d, p)_G$-regular pairs for every $\{(i,j),(i',j')\} \in E(R^k_r)$ and $v \in V \setminus V_0$,

(G4b) $|N_\Gamma(v, V'_{i,j})| = (1 \pm 2\varepsilon)p|V_{i,j}|$ for every $i \in [r]$, $j \in [k]$ and every $v \in V \setminus V_0$,

(G5b) $\big|V'_{f^*(x)} \cap \bigcap_{u\in J_x} N_G(u)\big| \ge \frac{1}{2}\alpha p^{|J_x|}|V'_{f^*(x)}|$,

(G6b) $\big|V'_{f^*(x)} \cap \bigcap_{u\in J_x} N_\Gamma(u)\big| = (1 \pm 2\varepsilon^*)p^{|J_x|}|V'_{f^*(x)}|$,

(G7b) $\big(\bigcap_{u\in J_x} N_\Gamma(u, V'_{f^*(x)}), \bigcap_{v\in J_y} N_\Gamma(v, V'_{f^*(y)})\big)$ is is an $(2\varepsilon^*, d, p)_G$-regular pair for each edge $\{x,y\} \in E(H')$,

(G8b) $\big|\bigcap_{u\in J_x} N_\Gamma(u)\big| \le (1 + 2\varepsilon^*)p^{|J_x|}n$ for each $x \in V(H')$.

We are now almost finished. However, we do not necessarily have $|W'_{i,j}| = |V'_{i,j}|$ for each $i \in [r]$ and $j \in [k]$. Since $|V'_{i,j}| = |V_{i,j}| \pm 2\Delta^{r+10}|V_0| = m_{i,j} \pm 3\Delta^{r+10}|V_0|$, by (H1b) we have $|V'_{i,j}| = |W'_{i,j}| \pm \xi n$. We can thus apply Lemma 3.7, with input k, r_1, Δ, γ, d, 8ε, and r. This gives us a partition $\{V''_{i,j}\}_{i\in[r],j\in[k]}$ with $|V''_{i,j}| = |W'_{i,j}|$ for each $i \in [r]$ and $j \in [k]$ by Property (B1') of Lemma 3.7. Let $\mathcal{V}'' = \{V''_{i,j}\}_{i\in[r],j\in[k]}$. Lemma 3.7 guarantees us the following since $4\varepsilon^* \ge 8\varepsilon$ and by using Properties (G1b)–(G8b) and Proposition 2.4.

(G1c) $\frac{n}{8kr} \le |V''_{i,j}| \le \frac{8n}{kr}$ for every $i \in [r]$ and $j \in [k]$,

(G2c) $\mathcal{V}''$ is $(4\varepsilon^*, d, p)_G$-regular on R^k_r and $(4\varepsilon^*, d, p)_G$-super-regular on K^k_r,

(G3c) both $\big(N_\Gamma(v, V''_{i,j}), V''_{i',j'}\big)$ and $\big(N_\Gamma(v, V''_{i,j}), N_\Gamma(v, V''_{i',j'})\big)$ are $(4\varepsilon^*, d, p)_G$-regular pairs for every $\{(i,j),(i',j')\} \in E(R^k_r)$ and $v \in V \setminus V_0$,

(G4c) we have $(1 - 4\varepsilon)p|V''_{i,j}| \le |N_\Gamma(v, V''_{i,j})| \le (1 + 4\varepsilon)p|V''_{i,j}|$ for every $i \in [r]$, $j \in [k]$ and every $v \in V \setminus V_0$.

(G5c) $\big|V''_{f^*(x)} \cap \bigcap_{u\in J_x} N_G(u)\big| \ge \frac{1}{4}\alpha p^{|J_x|}|V''_{f^*(x)}|$,

(G6c) $\big|V''_{f^*(x)} \cap \bigcap_{u\in J_x} N_\Gamma(u)\big| = (1 \pm 4\varepsilon^*)p^{|J_x|}|V'_{f^*(x)}|$, and

(G7c) $\big(V''_{f^*(x)} \cap \bigcap_{u\in J_x} N_\Gamma(u), V''_{f^*(y)} \cap \bigcap_{v\in J_y} N_\Gamma(v)\big)$ is an $(4\varepsilon^*, d, p)_G$-regular pair for each edge $\{x,y\} \in E(H')$.

Let us briefly explain why these properties hold. Property (G1c) results from Property (G1b) and Property (B2') of Lemma 3.7, while Property (G2c) comes from Property (B3') and choice of ε. Property (G3c) is guaranteed by Property (B4') of Lemma 3.7. Now, each of Properties (G4c), (G5c) and (G6c) comes from the corresponding Properties (G4b), (G5b) and (G6b) together with Property (B5') of Lemma 3.7. Finally, Property (G7c) holds due to Properties (G7b) and (G8b) together with Proposition 2.4 and Property (B5') of Lemma 3.7.

Next we define the family $\mathcal{I} = \{I_x\}_{x\in V(H')}$ of image restrictions. For each $x \in V(H')$ with $J_x = \varnothing$, set

$$I_x = V''_{f^*(x)}$$

and for each $x \in V(H')$ with $J_x \neq \varnothing$, set

$$I_x = V''_{f^*(x)} \cap \bigcap_{u\in J_x} N_G(u).$$

The last step of the proof is to use Theorem 2.9 to embed the vertices in $V(H')$ onto $V(G')$ respecting the image restrictions. By Property (G1c), the partitions $\mathcal{W}'$ and $\mathcal{V}''$ are both κ-balanced since $\kappa = 64$, and by construction they are size-compatible. While $\mathcal{W}'$ is a partition of $V(H')$, we have that $\mathcal{V}''$ is partition of $V(G) \setminus \mathrm{im}(\phi_{t_f})$. Each of their parts has size at least $n/(\kappa r_1)$ by (G1c).

Let $\widetilde{W}_{i,j} := W'_{i,j} \setminus X'$. We have to check that the following properties hold in order to apply Theorem 2.9 with the parameters defined above.

(BUL1) $\Delta(H') \leq \Delta$, for every edge $\{x, y\} \in E(H')$ with $x \in W'_{i,j}$ and $y \in W'_{i',j'}$ we have $\{(i, j), (i', j')\} \in E(R^k_r)$ and $\{\widetilde{W}_{i,j}\}_{i\in[r],j\in[k]}$ is an (ϑ, K^k_r)-buffer for H',

(BUL2) $\mathcal{V}''$ is $(\varepsilon_{\mathrm{BL}}, d, p)_G$-regular on R^k_r, $(\varepsilon_{\mathrm{BL}}, d, p)_G$-super-regular on K^k_r and has one-sided inheritance on K^k_r,

(BUL3) for every vertex $x \in \widetilde{W}_{i,j}$ and every triangle $\{x, y, z\}$ in H' with $y \in W'_{i',j'}$ and $z \in W'_{i'',j''}$, the set $V''_{i,j}$ has two-sided inheritance with respect to $V''_{i',j'}$ and $V''_{i'',j''}$,

(BUL4) $\mathcal{I}$ and $\mathcal{J}$ form a $(\rho, \zeta, \Delta, \Delta_J)$-restriction pair.

By Property (H2b), choice of ξ, and by Property (H4b) it holds that $\{\widetilde{W}_{i,j}\}_{i\in[r],j\in[k]}$ is a (ϑ, K^k_r)-buffer for H'. Furthermore since f^* is a graph homomorphism from H' to R^k_r, we have (BUL1).

By Properties (G2c), (G3c) and (G4c) we have Properties (BUL2) and (BUL3).

Finally, the pair $(\mathcal{I}, \mathcal{J}) = \big(\{I_x\}_{x\in V(H')}, \{J_x\}_{x\in V(H')}\big)$ forms a $(\rho, \alpha/4, \Delta, \Delta_J)$-restriction pair. To see this, observe that the total number of image restricted vertices in H', that is the number of vertices x with $J_x \neq \varnothing$, is at most $\Delta^2|V_0| < \rho|V_{i,j}|$ for any $i \in [r]$ and $j \in [k]$. This gives Property (RP1) of the definition of a restriction pair (Definition 2.8). Since for each vertex $x \in V(H')$ we have $|J_x| + \deg_{H'}(x) = \deg_H(x) \leq \Delta$ we have (RP2), while (RP3) follows from (G5c), and (RP4) follows from (G6c). Finally, Property (RP5) follows from (G7c), and Property (RP6) holds since $\Delta(H) \leq \Delta$. All in all, this shows that Property (BUL4) is satisfied.

As a consequence, by Theorem 2.9 there exists an embedding ϕ of H' into $G \setminus \mathrm{im}(\phi_{t_f})$, such that $\phi(x) \in I_x$ for each $x \in V(H')$. Together with ϕ_{t_f} we have thus found an embedding of H in G, as desired. $\square$

With Theorem 3.3 in hand, we can now present the proof of Theorem 3.1.

Proof of Theorem 3.1. Given γ, Δ, and k, let $\beta > 0$, $z > 0$, and $C > 0$ be returned by Theorem 3.3 with input γ, Δ, and k. Set $\beta^* = \beta/2$ and $C^* = C/\beta$. Let H be a k-colourable graph on n vertices with $\Delta(H) \leq \Delta$ such that there exists a set W of at least $C^* \max\{p^{-2}, p^{-1}\log n\}$ vertices in $V(H)$ that are not contained in any triangles of H and such that there exists a labelling $\mathcal{L}$ of its vertex set of bandwidth at most $\beta^* n$.

By the choice of C^* we find an interval $I \subseteq \mathcal{L}$ of length βn containing a subset $F \subseteq W$ with $|F| = C\max\{p^{-2}, p^{-1}\log n\}$. Now we can rearrange the labelling $\mathcal{L}$ to a labelling $\mathcal{L}'$ of bandwidth at most $2\beta^* n = \beta n$ such that F is contained in the first βn vertices in $\mathcal{L}'$.

Then, by Theorem 3.3 we know that $\Gamma = G(n,p)$ satisfies the following a.a.s. if $p \geq C(\log n/n)^{1/\Delta}$ and in particular if $p \geq C^*(\log n/n)^{1/\Delta}$. If G is a spanning subgraph of Γ with $\delta(G) \geq \big((k-1)/k + \gamma\big)pn$, then G contains a copy of H, which finishes the proof. $\square$

3.2 Local resilience of spanning degenerate subgraphs

Recall that a graph G is k-degenerate if every subgraph of G has a vertex of degree at most k. An easy upper bound on the degeneracy of a graph is by definition its maximum degree. However, there are many classes of graphs that have small degeneracy while their maximum degrees can be arbitrarily large. Clearly, trees and forests are 1-degenerate. Since every planar graph is known to have a vertex of degree 5 or less, planar graphs are 5-degenerate. Every series-parallel graphs has a vertex of degree 2. Hence, series-parallel graphs and in particular 2-trees and outerplanar graphs are 2-degenerate.

In the proof of Theorem 3.1 there are two places, where a lower bound of $(\log n/n)^{1/\Delta}$ on p is required. First, when we apply the common neighbourhood lemma (Lemma 3.6) and then when we use the sparse blow-up lemma (Theorem 2.9). If we demand H to not only have a particular number of vertices not in triangles but that the same vertices are also not contained in any 4-cycles, we can avoid having to use the common neighbourhood lemma in the embedding process. Allen, Böttcher, Hàn, Kohayakawa, and Person [9] have also proved a sparse blow-up lemma to embed degenerate graphs into sparse random graphs. This allows us to prove a variant of Theorem 3.1 to embed D-degenerate graphs with $\mathcal{O}(p^{-2}, p^{-1}\log n)$ vertices not in triangles or 4-cycles, where a lower bound of $\Omega\big((\log n)^{1/(2D+1)}\big)$ on the edge probability suffices. More precisely, we prove the following.

Theorem 3.15. *For each $\gamma > 0$, $\Delta \geq 2$, and $D, k \geq 1$, there exist constants $\beta^* > 0$ and $C^* > 0$ such that the following holds asymptotically almost surely for $\Gamma = G(n,p)$ if $p \geq C^*(\log n/n)^{1/(2D+1)}$.*

Let G be a spanning subgraph of Γ with $\delta(G) \geq \big((k-1)/k + \gamma\big)pn$ and let H be a D-degenerate, k-colourable graph on n vertices with $\Delta(H) \leq \Delta$, bandwidth at most $\beta^* n$ and there are at least $C^* \max\{p^{-2}, p^{-1}\log n\}$ vertices in $V(H)$ that are not contained in any triangles or four-cycles of H. Then G contains a copy of H.

As an immediate consequence we obtain the following resilience result for the containment of maximum degree bounded spanning trees.

Corollary 3.16. *For each $\gamma > 0$ and $\Delta \geq 2$, there exists $C > 0$ such that $\Gamma = G(n,p)$ satisfies the following asymptotically almost surely if $p \geq C(\log n/n)^{1/3}$.*

Let G be a spanning subgraph of Γ with $\delta(G) \geq \big(1/2 + \gamma\big)pn$. Then G contains every spanning tree with maximum degree at most Δ.

As with Theorem 3.1, we can deduce Theorem 3.15 from the following more general statement.

Theorem 3.17. *For each $\gamma > 0$, $\Delta \geq 2$, $D \geq 1$ and $k \geq 1$, there exist constants $\beta > 0$, $z > 0$, and $C > 0$ such that the following holds asymptotically almost surely for $\Gamma = G(n,p)$ if $p \geq C(\log n/n)^{1/(2D+1)}$.*

Let G be a spanning subgraph of Γ with $\delta(G) \geq \big((k-1)/k+\gamma\big)pn$ and let H be a graph on n vertices with $\Delta(H) \leq \Delta$ and degeneracy at most D, that has a labelling $\mathcal{L}$ of its vertex set of bandwidth at most βn, a $(k+1)$-colouring that is (z,β)-zero-free with respect to $\mathcal{L}$ and where the first $\sqrt{\beta}n$ vertices in $\mathcal{L}$ are not given colour zero and the first βn vertices in $\mathcal{L}$ include $C\max\{p^{-2}, p^{-1}\log n\}$ vertices that are not in any triangles or copies of C_4 in H. Then G contains a copy of H.

Before giving the proof of Theorems 3.15 and 3.17 in Subsection 3.2.2 we state in the following subsection the sparse blow-up lemma, which is the essential tool in the proof of Theorem 3.17.

3.2.1 Preliminaries

Given an order τ on $V(H)$ and a family $\mathcal{J} = \{J_x\}_{x\in V(H)}$ of image restricting vertices, we define $\pi^\tau(x) = |J_x| + \big|\{y \in N_H(x) : \tau(y) < \tau(x)\}\big|$. The following sparse blow-up lemma, which was proved by Allen, Böttcher, Hàn, Kohayakawa, and Person in [9], is a variant of the sparse blow-up lemma (Theorem 2.9) to embed degenerate graphs.

Theorem 3.18 (Blow-up lemma to embed degenerate graphs into random graphs, [9]). *For all non-negative integers Δ, $\Delta_{R'}$, Δ_J, D', D^{queue}, and D^{buf} and reals $\vartheta, \zeta, d > 0$, and $\kappa > 1$ there exist $\varepsilon, \rho > 0$ such that for all $r_1 \geq 1$ there exists $C > 0$ such that for $p \geq C(\log n/n)^{1/\max\{D', 2D^{\mathrm{queue}}, D^{\mathrm{buf}}\}}$ the random graph $\Gamma = G_{n,p}$ asymptotically almost surely satisfies the following.*

Let R be a graph on $r \leq r_1$ vertices and let $R' \subseteq R$ be a spanning subgraph with $\Delta(R') \leq \Delta_{R'}$. Let H and $G \subseteq \Gamma$ be graphs given with κ-balanced, size-compatible vertex partitions $\mathcal{W} = \{W_i\}_{i\in[r]}$ and $\mathcal{V} = \{V_i\}_{i\in[r]}$ with parts of size at least $m \geq N/(\kappa r_1)$. Let $\mathcal{I} = \{I_x\}_{x\in V(H)}$ be a family of image restrictions, and $\mathcal{J} = \{J_x\}_{x\in V(H)}$ be a family of restricting vertices. Suppose that

- *(DBUL1) H satisfies $\Delta(H) \leq \Delta$, for every edge $\{x,y\} \in E(H)$ with $x \in W_i$ and $y \in W_j$ we have $\{i,j\} \in E(R)$, and $\widetilde{\mathcal{W}} = \{\widetilde{W}_i\}_{i\in[r]}$ is an (ϑ, R')-buffer for H with $\deg(x) \leq D^{\mathrm{buf}}$ for each $x \in \widetilde{W}_i$ and $i \in [r]$,*
- *(DBUL2) $(G, \mathcal{V})$ is an (ε, d, p)-regular R-partition, which is (ε, d, p)-super-regular on R', and has one-sided inheritance,*
- *(DBUL3) for every vertex $x \in \widetilde{W}_i$ and every triangle $\{x,y,z\}$ in H with $y \in W_j$ and $z \in W_k$, the set V_i has two-sided inheritance with respect to V_j and V_k, and*
- *(DBUL4) $\mathcal{I}$ and $\mathcal{J}$ form a $(\rho, \zeta, \Delta, \Delta_J)$-restriction pair.*

Suppose furthermore that there is an order τ on $V(H)$, and a set $X^e \subseteq V(H)$ with $|X^e| \leq \varepsilon p^{\max\{\pi^\tau(x)\colon x\in X^e\}} n/r_1$ with the following properties.

- *(ORD1) For each $x \in V(H)$, we have*

- $\pi^\tau(x) \le D'$,
- $\pi^\tau(x) \le D' - 1$ *if there is* $y \in N_H(x)$ *with* $\tau(y) > \tau(x)$, *and*
- $\pi^\tau(x) \le D' - 2$ *if there are* $y, z \in N_H(x)$ *with* $\tau(y), \tau(z) > \tau(x)$ *and* $\{y, z\} \in E(H)$.

(ORD2) For each vertex x *in* $V(H)\backslash X^e$, *either* $\pi^\tau(x) \le D^{\mathrm{queue}}$, *or* x *is not image restricted and all neighbours of* x *appear in the* $\varepsilon p^{\pi^\tau(x)} n/r_1$ *places of* τ *before* x.

(ORD3) If $x \in N(\widetilde{W}_i)$ *for some* $i \in [r]$ *then*

- $\pi^\tau(x) \le D^{\mathrm{buf}} - 1$,
- $\pi^\tau(x) \le D^{\mathrm{buf}} - 2$ *if there is* $y \in N_H(x)$ *with* $\tau(y) > \tau(x)$, *and*
- $\pi^\tau(x) \le D^{\mathrm{buf}} - 3$ *if there are* $y, z \in N_H(x)$ *with* $\tau(y), \tau(z) > \tau(x)$ *and* $yz \in E(H)$.

(ORD4) If $x \in N(\widetilde{W}_i)$ *for some* $i \in [r]$ *then at most* $s := D^{\mathrm{buf}} - 1 - \max\{\pi^\tau(y) : \tau(y) < \tau(x), y \notin X^e\}$ *neighbours of* x *not in* X^e *appear in* τ *before* $\tau(x) - \varepsilon p^s n/r_1$.

Then there is an embedding $\psi \colon V(H) \to V(G)$ *such that* $\psi(x) \in I_x$ *for each* $x \in H$.

3.2.2 Proof of the theorem

In this subsection we prove Theorem 3.17. The proof is very similar to the one of Theorem 3.3. Therefore, we may at some points omit arguments and calculations that we provided in the proof of Theorem 3.3 in Section 3.1.

Proof of Theorem 3.17. We set up the constants quite similarly as in the proof of Theorem 3.3. Specifically, given $\gamma > 0$, $\Delta \ge 2$, $D \ge 1$, and $k \ge 2$, set $r_0 = 10/\gamma$. Let d be returned by Lemma 3.4, with input γ, k, and r_0. Set $\alpha = d/2$. Let $D' := D + 2$ if $D = 1$, and $D' := D + 3$ otherwise. Furthermore, set $D^{\mathrm{queue}} = D$ and $D^{\mathrm{buf}} = 2D + 1$.

Now let $\varepsilon_{\mathrm{BL}} > 0$ and $\rho > 0$ be returned by Theorem 3.18 with input Δ, $\Delta_{R'} = 3k$, $\Delta_J = \Delta$, D', D^{queue}, D^{buf}, $\vartheta = 1/(100D)$, $\zeta = \alpha/4$, d, and $\kappa := 64$. Set $\varepsilon^* = \varepsilon_{\mathrm{BL}}/8$. Next, let $\varepsilon_1 > 0$ be returned by Lemma 2.10 for input ε^* and d. Let $\varepsilon_0 > 0$ be small enough both for Lemma 2.11 with input ε^* and d, and for Lemma 2.10 with input ε_1 and d.

We choose

$$\varepsilon = \min\big\{\varepsilon_0, d/8, \varepsilon^*/(4D), 1/(16k)\big\}.$$

Putting ε into Lemma 3.4 returns r_1 and C_1^*. Next, Lemma 3.7, for input k, r_1, Δ, γ, d, and 8ε, returns $\xi \in \big(0, 1/(10kr_1)\big)$ and C_2^*. We set

$$\beta = 10^{-12}\xi^2/(\Delta k^4 r_1^2) \ \text{ and } \ \mu = 10^{-5}\varepsilon^2/(kr_1).$$

Let C_3^* be the maximum of the C-outputs of Theorem 2.9 with input r_1, of Proposition 3.8 with input ε and μ^2, and for Lemma 2.19 with input $\varepsilon\mu$ and Δ. Finally, let $C^* = \max\{C_1^*, C_2^*, C_3^*\}$ and set

$$C = 10^{10}k^2 r_1^2 \Delta^{2r_1+20} C^*/(\varepsilon^2\xi\mu^2) \ \text{ and } \ z = 10/\xi.$$

Given $p \ge C(\log n/n)^{1/(2D+1)}$, a.a.s. $G(n,p)$ satisfies the good events of Theorem 3.18, Lemmas 3.4, 2.10 and 2.11, and Proposition 3.8 with the inputs as specified above. We condition on these good events for $\Gamma = G(n,p)$.

Let G be a spanning subgraph of Γ with $\delta(G) \geq \big((k-1)/k+\gamma\big)pn$. Let H be any graph on n vertices with $\Delta(H) \leq \Delta$, and let $\mathcal{L}$ be a labelling of $V(H)$ of bandwidth at most βn whose first βn vertices include $C \max\{p^{-2}, p^{-1}\log n\}$ vertices that are not contained in any triangles or four-cycles of H, and such that there exists a $(k+1)$-colouring that is (z,β)-zero-free with respect to $\mathcal{L}$, and the colour zero is not assigned to the first $\sqrt{\beta}n$ vertices. Furthermore, let τ be a D-degeneracy order of $V(H)$.

Next, as in the proof of Theorem 3.3, we apply Lemma 3.4 to G, obtaining a partition of $V(G)$ with the following properties:

(G1a) $\frac{n}{4kr} \leq |V_{i,j}| \leq \frac{4n}{kr}$ for every $i \in [r]$ and $j \in [k]$,

(G2a) $\mathcal{V}$ is $(\varepsilon, d, p)_G$-regular on R^k_r and $(\varepsilon, d, p)_G$-super-regular on K^k_r,

(G3a) both $\big(N_\Gamma(v, V_{i,j}), V_{i',j'}\big)$ and $\big(N_\Gamma(v, V_{i,j}), N_\Gamma(v, V_{i',j'})\big)$ are $(\varepsilon, d, p)_G$-regular pairs for every $\{(i,j),(i',j')\} \in E(R^k_r)$ and $v \in V \setminus V_0$, and

(G4a) $|N_\Gamma(v, V_{i,j})| = (1 \pm \varepsilon)p|V_{i,j}|$ for every $i \in [r]$, $j \in [k]$ and every $v \in V \setminus V_0$.

If $D = 1$, we ask only for the following weaker condition in place of (G3a):

(G3a') $\big(N_\Gamma(v, V_{i,j}), V_{i',j'}\big)$ is an $(\varepsilon, d, p)_G$-regular pair for every $\big\{(i,j),(i',j')\big\} \in E(R^k_r)$ and $v \in V \setminus V_0$.

Thus, for $D = 1$ we have for the size of the exceptional set $|V_0| \leq C^* p^{-1}\log n$, whereas for $D \geq 2$ we have $|V_0| \leq C^* \max\{p^{-2}, p^{-1}\log n\}$.

As in the proof of Theorem 3.3, for every $i \in [r]$ and $j \in [k]$ we choose

$$m_{i,j} \in \big\{|V_{i,j}| + \big\lfloor \tfrac{1}{kr}|V_0| \big\rfloor, |V_{i,j}| + \big\lceil \tfrac{1}{kr}|V_0| \big\rceil\big\}$$

such that $\{m_{i,j}\}_{i\in[r],j\in[k]}$ forms a k-equitable integer partition of n.

Next, we apply Lemma 3.5 to obtain a partition of $V(H)$. We use the same inputs as in the proof of Theorem 3.3. The result is a function $f : V(H) \to V(R^k_r)$ and a special set X with the following properties:

(H1a) $m_{i,j} - \frac{1}{10}\xi n \leq |W_{i,j}| \leq m_{i,j} + \frac{1}{10}\xi n$,

(H2a) $|X| \leq \xi n$,

(H3a) $\{f(x), f(y)\} \in E(R^k_r)$ for every $\{x,y\} \in E(H)$,

(H4a) $y, z \in \bigcup_{j'\in[k]} f^{-1}(i,j')$ for every $x \in f^{-1}(i,j) \setminus X$ and $\{x,y\},\{y,z\} \in E(H)$, and

(H5a) $f(x) = \big(1, \sigma(x)\big)$ for every x in the first $\sqrt{\beta}n$ vertices of $\mathcal{L}$.

(H6a) $|\{x \in f^{-1}(i,j) : \deg(x) \leq 2D\}| \geq \frac{1}{24D}|f^{-1}(i,j)|$.

We now continue following the proof of Theorem 3.3, using Lemma 2.19 with input $\varepsilon\mu$ and $D+1$ (rather than $\varepsilon\mu$ and Δ), to choose a set S of size μn that satisfies for all vertices $u_1, \ldots, u_\ell$ of $V(G)$ with $1 \leq \ell \leq D+1$ the following:

$$\begin{gathered}\Big|S \cap \bigcap_{1\leq i'\leq \ell} N_\Gamma(u_{i'})\Big| = \mu\Big|\bigcap_{1\leq i'\leq \ell} N_\Gamma(u_{i'})\Big| \pm \mu\Big(\varepsilon\Big|\bigcap_{1\leq i'\leq \ell} N_\Gamma(u_{i'})\Big| + p^\ell n\Big), \\ |S \cap N_G(v)| = \mu|N_G(v)| \pm 2\mu\varepsilon|N_G(v)| \quad \text{and} \quad \big|S \cap V_{i,j}\big| \leq 2\mu|V_{i,j}|.\end{gathered} \tag{3.16}$$

We use the same pre-embedding Algorithm 3, with the exception that we choose vertices at line 2 differently. As before, given $v_{t+1} \in V_0 \setminus \mathrm{im}(\phi_t)$ we use Claim 3.14 to find a set $W \subseteq N_G(v_{t+1})$ of size at least $\mu^2 pn/(8r)$ and an index $i_{t+1} \in [r]$ such that for each $w \in W$ we have $\big|N_G(w, V_{i_{t+1},j})\big| \geq dp|V_{i_{i+1},j}|$ for each $j \in [k]$. As before, for each $t' \in [t+1]$ and each $z \in V(H)$ at distance at least 2 and at most $i_{t'}+1$ from $x_{t'}$, we set $f^*(z)$ as $f^*(z) = \big(i_t + 2 - \mathrm{dist}(x_t, z), j\big)$ if $j \in [k]$ such that $f(z) = (1, j)$. For all other $z \in V(H)$ we set $f^*(z) = f(z)$.

Rather than applying Lemma 3.6, we let $w_1, \dots, w_\ell$ be distinct vertices of W such that, if ϕ_{t+1} is as in Algorithm 3, they satisfy the following, where we let $J_x := \phi_{t+1}\big(N_H(x) \cap \mathrm{dom}(\phi_{t+1})\big)$ for each $x \in V(H) \setminus \mathrm{dom}(\phi_{t+1})$:

(G5a) $\big|\bigcap_{u \in J_x} N_G(u, V_{f^*(x)})\big| \geq \alpha p^{|J_x|}|V_{f^*(x)}|$ for each $x \in V(H) \setminus \mathrm{dom}(\phi_{t+1})$,

(G6a) $\big|\bigcap_{u \in J_x} N_\Gamma(u, V_{f^*(x)})\big| = (1 \pm \varepsilon^*)p^{|J_x|}|V_{f^*(x)}|$ for each $x \in V(H) \setminus \mathrm{dom}(\phi_{t+1})$,

(G7a) $\big(\bigcap_{u \in J_x} N_\Gamma(u, V_{f^*(x)}), \bigcap_{v \in J_y} N_\Gamma(v, V_{f^*(y)})\big)$ is an $(\varepsilon^*, d, p)_G$-regular pair for all $x, y \in V(H) \setminus \mathrm{dom}(\phi_{t+1})$ with $\{x, y\} \in E(H)$, and

(G8a) $\big|\bigcap_{u \in J_x} N_\Gamma(u)\big| \leq (1 + \varepsilon^*)p^{|J_x|}n$ for each $x \in V(H) \setminus \mathrm{dom}(\phi_{t+1})$.

Let us justify that this is possible. We choose the vertices $w_1, \dots, w_\ell$ successively. Since $x_1, \dots, x_{t+1}$ are not contained in any triangle or four-cycle of H, we have $|J_x| \leq 1$ for each $x \in V(H) \setminus \mathrm{dom}(\phi_{t+1})$. Hence, Property (G5a) is satisfied. By Proposition 3.8, Properties (G6a) and (G8a) are satisfied for all but at most $2C^* k r_1 p^{-1} \log n$ vertices of W. It remains to show that we can obtain (G7a), which we do as follows. For $s \in [\ell]$, when we come to choose w_s, we insist that for any $\big\{(i, j), (i', j')\big\} \in E(R_r^k)$, the following hold. First, $\big(N_\Gamma(w_s, V_{i,j}), V_{i',j'}\big)$ is $(\varepsilon_1, d, p)_G$-regular. Second, $\big(N_\Gamma(w_s, V_{i,j}), N_\Gamma(w_s, V_{i',j'})\big)$ is $(\varepsilon^*, d, p)_G$-regular. Third, for each $1 \leq s' \leq s-1$, $\big(N_\Gamma(w_s, V_{i,j}), N_\Gamma(w'_s, V_{i',j'})\big)$ is $(\varepsilon^*, d, p)_G$-regular. The conditions of respectively Lemma 2.10, Lemma 2.11, and Lemma 2.10 are in each case satisfied (in the last case by choice of w_t) and thus in total at most $3C^* k^2 r_1^2 \max\{p^{-2}, p^{-1} \log n\}$ vertices of W are prohibited. Since $5C^* k^2 r_1^2 \max\{p^{-2}, p^{-1} \log n\} < |W|/2 < \ell$ by choice of C, at each step there is a valid choice of w_s. Since for each $x \in V(H) \setminus \mathrm{dom}(\phi_{t+1})$ we have $|J_x| \leq 1$, this construction guarantees (G7a).

We now return to following the proof of Theorem 3.3. Let t_f be the time at which Algorithm 3 terminates. We obtain the partition $\mathcal{V}' = \{V'_{i,j}\}_{i \in [r], j \in [k]}$ by removing the images of pre-embedded vertices, and $\mathcal{V}'' = \{V''_{i,j}\}_{i \in [r], j \in [k]}$ with the following properties (as in the proof of Theorem 3.3) by applying Lemma 3.7. We have

(G1c) $\frac{n}{8kr} \leq |V''_{i,j}| \leq \frac{8n}{kr}$ for every $i \in [r]$ and $j \in [k]$,

(G2c) $\mathcal{V}''$ is $(4\varepsilon^*, d, p)_G$-regular on R_r^k and $(4\varepsilon^*, d, p)_G$-super-regular on K_r^k,

(G3c) both $\big(N_\Gamma(v, V''_{i,j}), V''_{i',j'}\big)$ and $\big(N_\Gamma(v, V''_{i,j}), N_\Gamma(v, V''_{i',j'})\big)$ are $(4\varepsilon^*, d, p)_G$-regular pairs for every $\{(i, j), (i', j')\} \in E(R_r^k)$ and $v \in V \setminus V_0$,

(G4c) we have $(1 - 4\varepsilon)p|V''_{i,j}| \leq |N_\Gamma(v, V''_{i,j})| \leq (1 + 4\varepsilon)p|V''_{i,j}|$ for every $i \in [r]$, $j \in [k]$ and every $v \in V \setminus V_0$.

(G5c) $\big|V''_{f^*(x)} \cap \bigcap_{u \in J_x} N_G(u)\big| \geq \frac{1}{4}\alpha p^{|J_x|}|V''_{f^*(x)}|$,

(G6c) $\big|V''_{f^*(x)} \cap \bigcap_{u \in J_x} N_\Gamma(u)\big| = (1 \pm 4\varepsilon^*)p^{|J_x|}|V'_{f^*(x)}|$, and

(G7c) $\big(V''_{f^*(x)} \cap \bigcap_{u\in J_x} N_\Gamma(u), V''_{f^*(y)} \cap \bigcap_{v\in J_y} N_\Gamma(v)\big)$ is an $(4\varepsilon^*, d, p)_G$-regular pair for each edge $\{x,y\} \in E(H')$.

Note that Property (B5') of Lemma 3.7 may be trivial, that is, the error term $C^*\log n$ may dominate the main term when s is large. However, we only require it for $s = 1$ to obtain (G1c)–(G7c).

Finally, we are ready to apply Theorem 3.18 to complete the embedding. We define $(\mathcal{I},\mathcal{J})$ as in the proof of Theorem 3.3. However, we let $\widetilde{W}_{i,j}$ consist of the vertices of $W'_{i,j} \setminus X$ whose degree is at most $2D \le D^{\text{buf}}$. By (H6a) there are at least $|W'_{i,j}|/(100D)$ of these, so that $\widetilde{\mathcal{W}}$ is a (ϑ, K_r^k)-buffer, giving (DBUL1). Now (DBUL2) and (DBUL3) follow from (G2c) and (G3c). Finally, $(\mathcal{I},\mathcal{J})$ is a $(\rho, \alpha/4, \Delta, \Delta)$-restriction pair, giving (DBUL4), exactly as in the proof of Theorem 3.3. However now we need to give an order τ' on $V(H')$ and a set $X^e \subseteq V(H')$. The former is simply the restriction of τ to $V(H')$, and the set X^e consists of all vertices $x \in V(H)$ with $|J_x| > 0$. We now justify that the remaining conditions of Theorem 3.18 are satisfied.

First, we claim $|X^e| \le \Delta^2|V_0| \le \varepsilon p^{\max_{x\in X^e}\pi^{\tau'}(x)} n/r_1$. Observe that $\pi^{\tau'}(x) \le \pi^{\tau}(x) + |J_x| \le D+1$. For $D = 1$, we have $|V_0| \le C^* p^{-1}\log n$, and by choice of C the desired inequality follows. For $D \ge 2$, we have $|V_0| \le C^*\max\{p^{-2}, p^{-1}\log n$, and again by choice of C we have the desired inequality. Now observe that for any vertex x of H' we have $\pi^{\tau'}(x) \le D+1$. If $D = 1$, then H' contains no triangles, and hence (ORD1) with $D' = D+2$ is satisfied. If $D \ge 2$, then we have $D' = D+3$, so that again (ORD1) is satisfied. Next, if $x \notin X^e$ then $\pi^{\tau'}(x) \le D$, so that (ORD2) holds. If $x \in N(\widetilde{W}_{i,j})$ for some $(i,j) \in R_r^k$, then by choice $x \notin X^e$, and thus $\pi^{\tau'}(x) \le D$. Since $D^{\text{buf}} = 2D+1$, if $D = 1$ then H contains no triangles and (ORD3) holds, while if $D \ge 2$ then $D^{\text{buf}} \ge D+3$ and (ORD3) holds. Finally, observe that if $x \notin X^e$ then $\pi^{\tau'}(x) \le D$, and since $D^{\text{buf}} = 2D+1$ we obtain (ORD4).

We can thus apply Theorem 3.18 to embed H' into G', completing the embedding of H into G as desired. □

The proof of Theorem 3.15 can be deduced from Theorem 3.17 verbatim as Theorem 3.1 from Theorem 3.3. Therefore we omit it here.

3.3 The bandwidth theorem in pseudorandom graphs

In this section we prove an analogue of Theorem 3.1 for bijumbled graphs. Recall that a graph G is said to be (p,ν)-*bijumbled* if for all disjoint sets $X, Y \subseteq V(G)$ it holds that $\big|e(X,Y) - p|X||Y|\big| \le \nu\sqrt{|X||Y|}$. Bijumbled graphs are so-called pseudorandom graph since their edge distribution resembles that of a random graph by definition. Indeed, the random graph $G(n,p)$ is a.a.s. $(p, \sqrt{pn})$-bijumbled (see e.g. [10]).

Recently, it was proved by Allen, Böttcher, Hàn, Kohayakawa, and Person [9] that for every $\Delta \ge 2$ there exists a constant $c > 0$ such that for all $p > 0$ every (p,ν)-bijumbled graph on n vertices with minimum degree at least $pn/2$ contains a copy of every n-vertex graph with maximum degree at most Δ whenever $\nu \le cp^{\max\{4,(3\Delta+1)/2\}}n$. In this section we prove the following theorem, which assures that such bijumbled graphs are robust with respect to the containment of all maximum degree bounded graphs with sublinear bandwidth and with a certain amount of vertices not in triangles.

Theorem 3.19. *For each $\gamma > 0$, $\Delta \geq 2$, and $k \geq 1$, there exists a constant $c > 0$ such that the following holds for any $p > 0$.*

Given $\nu \leq cp^{\max\{4,(3\Delta+1)/2\}}n$, let Γ be a $\big(p,\nu\big)$-bijumbled graph and let G be a spanning subgraph of Γ with $\delta(G) \geq \big((k-1)/k+\gamma\big)pn$. Suppose further that H is a k-colourable graph on n vertices with $\Delta(H) \leq \Delta$, bandwidth at most cn and with at least $c^{-1}p^{-6}\nu^2n^{-1}$ vertices not contained in any triangle of H. Then G contains a copy of H.

Once again, Theorem 3.19 is a consequence of the following more general theorem, in which a few vertices of H are allowed to receive an additional $(k+1)$-st colour.

Theorem 3.20. *For each $\gamma > 0$, $\Delta \geq 2$, and $k \geq 1$, there exist $c > 0$ and $z > 0$ such that the following holds for any $p > 0$.*

Given $\nu \leq cp^{\max\{4,(3\Delta+1)/2\}}n$, suppose Γ is a $\big(p,\nu\big)$-bijumbled graph, G is a spanning subgraph of Γ with $\delta(G) \geq \big((k-1)/k+\gamma\big)pn$, and H is a graph on n vertices with $\Delta(H) \leq \Delta$ and bandwidth at most cn. Suppose further that H has a labelling $\mathcal{L}$ of its vertex set of bandwidth at most cn, a $(k+1)$-colouring that is (z,c)-zero-free with respect to $\mathcal{L}$, and where the first $\sqrt{c}n$ vertices in $\mathcal{L}$ are not given colour zero, and the first cn vertices in $\mathcal{L}$ include $c^{-1}p^{-6}\nu^2n^{-1}$ vertices in $V(H)$ that are not contained in any triangles of H. Then G contains a copy of H.

The proof of Theorem 3.20 is again a modification of that of Theorem 3.3 and uses a blow-up lemma for bijumbled graphs by Allen, Böttcher, Hàn, Kohayakawa, and Person. In Subsection 3.3.1 we state this blow-up lemma as well as regularity inheritance lemmas for bijumbled graphs. We also prove an analogous version of Proposition 3.8 for bijumbled graphs in that subsection. Then in Subsection 3.3.2 we state bijumbled graph versions of the lemma for G (Lemma 3.4) and of the common neighbourhood lemma (Lemma 3.6) and sketch their proofs, which are modifications of the proofs of Lemmas 3.4 and 3.6. Finally, in Subsection 3.3.3 we present the proofs of Theorems 3.19 and 3.20.

3.3.1 Preliminaries

Since we are dealing with bijumbled graphs, we need to work with fully-regular pairs rather than with regular pairs. In order to use this concept and to work with bijumbled graphs, we need versions of Theorem 2.9, Lemmas 2.10 and 2.11, and Proposition 3.8, for fully-regular pairs and where Γ is a bijumbled graph rather than a random graph. We also require the following proposition, which gives a lower bound on the value ν for a (p,ν)-bijumbled graph with $p > 0$.

Proposition 3.21. *For every integer n and $16/n < p < 1-16/n$ there does not exist a (p,ν)-bijumbled n-vertex graph with $\nu \leq \min\big\{\sqrt{pn/32}, \sqrt{(1-p)n/32}\big\}$.*

Proof. Suppose that Γ is a (p,ν)-bijumbled graph on n vertices with $p \leq 1/2$. If Γ contains $n/2$ vertices of degree at least $4pn$, then we have $e(\Gamma) \geq pn^2$- Letting (A,B) be a maximum cut of Γ, by bijumbledness we have

$$\tfrac{1}{2}pn^2 \leq e(A,B) \leq p|A||B| + \nu\sqrt{|A||B|} \leq \tfrac{1}{4}pn^2 + \tfrac{1}{2}\nu n\,,$$

and thus $\nu \geq pn/2 \geq \sqrt{pn/32}$.

If, on the other hand, Γ contains at least $n/2$ vertices of degree less than $4pn$, then let A be a set of $1/(8p)$ such vertices, and B a set of $n/4$ vertices with no neighbours in A. By bijumbledness, we have

$$0 \geq p|A||B| - \nu\sqrt{|A||B|} = \tfrac{n}{32} - \nu\sqrt{n/(32p)}$$

and thus $\nu \geq \sqrt{pn/32}$. The same argument applied to the complement $\overline{\Gamma}$ of Γ proves the case when $p \geq 1/2$. $\square$

The following sparse blow-up lemma for bijumbled graphs was proved by Allen, Böttcher, Hàn, Kohayakawa, and Person [9].

Theorem 3.22 (Blow-up lemma for bijumbled graphs, [9])**.** *For all Δ, $\Delta_{R'}$, Δ_J, $\vartheta, \zeta, d > 0$, $\kappa > 1$ there exist $\varepsilon, \rho > 0$ such that for all $r_1 \geq 1$ there is a constant $c > 0$ such that if $p > 0$ and Γ is an n-vertex graph that is $(p, cp^{(3\Delta+1)/2}n)$-bijumbled, the following holds.*

Let R be a graph on $r \leq r_1$ vertices and let $R' \subseteq R$ be a spanning subgraph with $\Delta(R') \leq \Delta_{R'}$. Let H and $G \subseteq \Gamma$ be graphs given with κ-balanced, size-compatible vertex partitions $\mathcal{W} = \{W_i\}_{i\in[r]}$ and $\mathcal{V} = \{V_i\}_{i\in[r]}$ with parts of size at least $m \geq n/(\kappa r_1)$. Let $\mathcal{I} = \{I_x\}_{x\in V(H)}$ be a family of image restrictions and $\mathcal{J} = \{J_x\}_{x\in V(H)}$ be a family of restricting vertices. Suppose that

- *(PBUL1) H satisfies $\Delta(H) \leq \Delta$, and for every edge $\{x, y\} \in E(H)$ with $x \in W_i$ and $y \in W_j$ we have $\{i, j\} \in E(R)$, and $\widetilde{\mathcal{W}} = \{\widetilde{W}_i\}_{i\in[r]}$ is an (ϑ, R')-buffer for H,*
- *(PBUL2) $(G, \mathcal{V})$ is an (ε, d, p)-fully-regular R-partition, which is (ε, d, p)-super-fully-regular on R', and has one-sided inheritance,*
- *(PBUL3) for every vertex $x \in \widetilde{W}_i$ and every triangle $\{x, y, z\}$ in H with $y \in W_j$ and $z \in W_k$, the set V_i has two-sided inheritance with respect to V_j and V_k, and*
- *(PBUL4) $\mathcal{I}$ and $\mathcal{J}$ form a $(\rho p^\Delta, \zeta, \Delta, \Delta_J)$-restriction pair.*

Then there is an embedding $\psi \colon V(H) \to V(G)$ such that $\psi(x) \in I_x$ for each $x \in H$.

There are three main differences between Theorem 3.22 and the blow-up lemma for sparse random graphs (Theorem 2.9). First, Γ is a bijumbled graph rather than a random graph. Second, 'regular' is replaced by 'fully-regular'. Third, the number of vertices we may image restrict is smaller than in Theorem 2.9. We will see that these last two restrictions do not affect our proof substantially.

Next, we need the following regularity inheritance lemmas for bijumbled graphs. The first one deals with one-sided regularity inheritance.

Lemma 3.23 (Allen, Böttcher, Skokan, Stein [12])**.** *For each $\varepsilon', d > 0$ there are constants $\varepsilon, c > 0$ such that for all $0 < p < 1$ the following holds. Let Γ be a graph and $G \subseteq \Gamma$ be a subgraph of Γ. Let further X, Y, Z be disjoint vertex sets in $V(\Gamma)$. Assume that*

- *(X, Z) is $(p, cp^{3/2}\sqrt{|X||Z|})$-bijumbled in Γ,*
- *(X, Y) is $\Big(p, cp^2\big(\log_2 \frac{1}{p}\big)^{-1/2}\sqrt{|X||Y|}\Big)$-bijumbled in Γ, and*
- *(X, Y) is $(\varepsilon, d, p)_G$-fully-regular.*

Then, for all but at most $\varepsilon'|Z|$ vertices z of Z, the pair $\big(N_\Gamma(z) \cap X, Y\big)$ is $(\varepsilon', d, p)_G$-fully-regular.

The following lemma treats the case of two-sided regularity inheritance.

Lemma 3.24 (Allen, Böttcher, Skokan, Stein [12])**.** *For each $\varepsilon', d > 0$ there are constants $\varepsilon, c > 0$ such that for all $0 < p < 1$ the following holds. Let Γ be a graph and $G \subseteq \Gamma$ be a subgraph of Γ. Let further X, Y, Z be disjoint vertex sets in $V(\Gamma)$. Assume that*

- *(X, Z) is $(p, cp^2\sqrt{|X||Z|})$-bijumbled in Γ,*
- *(Y, Z) is $(p, cp^3\sqrt{|Y||Z|})$-bijumbled in Γ,*
- *(X, Y) is $(p, cp^{5/2}\big(\log_2 \frac{1}{p}\big)^{-\frac{1}{2}}\sqrt{|X||Y|})$-bijumbled in Γ, and*
- *(X, Y) is $(\varepsilon, d, p)_G$-fully-regular.*

Then, for all but at most $\varepsilon'|Z|$ vertices z of Z, the pair $\big(N_\Gamma(z) \cap X, N_\Gamma(z) \cap Y\big)$ is $(\varepsilon', d, p)_G$-fully-regular.

The following two lemmas, which more closely resemble Lemmas 2.10 and 2.11, are corollaries of Lemmas 3.23 and 3.24.

Lemma 3.25 (One-sided regularity inheritance for bijumbled graphs)**.** *For each $\varepsilon_{\text{OSRIL}}, \alpha_{\text{OSRIL}} > 0$ there exist $\varepsilon_0 > 0$ and $C > 0$ such that for any $0 < \varepsilon < \varepsilon_0$ and $0 < p < 1$, if Γ is any (p, ν)-bijumbled graph the following holds. For any disjoint sets X and Y in $V(\Gamma)$ with $|X| \geq Cp^{-3}\nu$ and $|Y| \geq Cp^{-2}\nu$, and any subgraph $G \subseteq \Gamma$ such that (X, Y) is $(\varepsilon, \alpha_{\text{OSRIL}}, p)_G$-regular, there are at most $Cp^{-3}\nu^2|X|^{-1}$ vertices $z \in V(\Gamma)$ such that $(X \cap N_\Gamma(z), Y)$ is not $(\varepsilon_{\text{OSRIL}}, \alpha_{\text{OSRIL}}, p)_G$-regular.*

Lemma 3.26 (Two-sided regularity inheritance for bijumbled graphs)**.** *For each $\varepsilon_{\text{TSRIL}}, \alpha_{\text{TSRIL}} > 0$ there exist $\varepsilon_0 > 0$ and $C > 0$ such that for any $0 < \varepsilon < \varepsilon_0$ and $0 < p < 1$, if Γ is any (p, ν)-bijumbled graph the following holds. For any disjoint sets X and Y in $V(\Gamma)$ with $|X|, |Y| \geq Cp^{-3}\nu$, and any subgraph $G \subseteq \Gamma$ such that (X, Y) is $(\varepsilon, \alpha_{\text{TSRIL}}, p)_G$-regular, there are at most $Cp^{-6}\nu^2/\min\big\{|X|, |Y|\big\}$ vertices $z \in V(\Gamma)$ such that $\big(X \cap N_\Gamma(z), Y \cap N_\Gamma(z)\big)$ is not $(\varepsilon_{\text{TSRIL}}, \alpha_{\text{TSRIL}}, p)_G$-regular.*

Note that the bijumbledness requirements of this lemma are such that if Y and Z are sets of size $\Theta(n)$, then X must have size $\Omega\big(p^{-6}\nu^2 n^{-1}\big)$. This will be the reason for the requirement of Theorem 3.20 on the number of vertices of H that are not allowed to be in triangles.

Finally, we provide a version of Proposition 3.8 for bijumbled graphs. The proof is similar to that of Proposition 3.8.

Proposition 3.27. *For each $\varepsilon > 0$ there exists a constant $C > 0$ such that for every $p > 0$, any graph Γ that is (p, ν)-bijumbled has the following property. For any disjoint sets $X, Y \subseteq V(\Gamma)$ with $|X|, |Y| \geq \varepsilon^{-1}p^{-1}\nu$, we have $e(X, Y) = (1 \pm \varepsilon)p|X||Y|$, and $e(X) \leq 2p|X|^2$. Furthermore, for every $Y \subseteq V(\Gamma)$ with $|Y| \geq Cp^{-1}\nu$, the number of vertices $v \in V(\Gamma)$ with $\big||N_\Gamma(v, Y)| - p|Y|\big| > \varepsilon p|Y|$ is at most $Cp^{-2}\nu^2|Y|^{-1}$.*

Proof. Given $\varepsilon > 0$, set $C' = 100/\varepsilon^2$ and $C = 200C'/\varepsilon$. Suppose Γ is (p, ν)-bijumbled.

Given disjoint subsets $X, Y \subseteq V(\Gamma)$ with $|X|, |Y| \geq \varepsilon^{-1}p^{-1}\nu$, by the (p, ν)-bijumbledness of Γ we have $e(X, Y) = p|X||Y| \pm \nu\sqrt{|X||Y|}$. Hence we need to verify that $\nu\sqrt{|X||Y|} \leq \varepsilon p|X||Y|$. This follows directly from the lower bounds on $|X|$ and $|Y|$.

For the second property, let (A, B) be a maximum cut of X. We have $e(A, B) \geq e(X)/2$, and $|A||B| \leq |X|^2/4$. By the (p, ν)-bijumbledness of Γ we conclude

$$e(X) \leq 2e(A, B) \leq 2p|A||B| + 2\nu\sqrt{|A||B|} \leq \tfrac{1}{2}p|X|^2 + \nu|X|.$$

Thus it is enough to verify $\nu|X| \leq p|X|^2$, which follows from the lower bound on $|X|$.

Now let $Y \subseteq V(\Gamma)$ have size at least $Cp^{-1}\nu$. We first show that there are at most $C'p^{-2}\nu^2|Y|^{-1}$ vertices in Γ that have each less than $(1-\varepsilon)p|Y|$ neighbours in Y. If this were false, then we could choose a set X of $C'p^{-2}\nu^2|Y|^{-1}$ vertices in Γ that have each less than $(1-\varepsilon)p|Y|$ neighbours in Y. Since by choice of C we have $(1-\varepsilon)p|Y| \leq (1-\varepsilon/2)p|Y \setminus X|$, we see that $e(X, Y \setminus X) < (1-\varepsilon/2)p|X||Y \setminus X|$. Since

$$\nu\sqrt{|X||Y|} = \nu\sqrt{C'p^{-2}\nu^2} = \sqrt{C'}\nu^2 p^{-1} < \tfrac{\varepsilon}{2}p|X||Y \setminus X|$$

this is a contradiction to the (p, ν)-bijumbleness of Γ.

Next we show that there are at most $2C'p^{-2}\nu^2|Y|^{-1}$ vertices of Γ that have each more than $(1+\varepsilon)p|Y|$ neighbours in Y. Again, if this is not the case we can let X be a set of $2C'p^{-2}\nu^2|Y|^{-1}$ vertices of Γ each with more than $(1+\varepsilon)p|Y|$ neighbours in Y.

If there are more than $|X|/2$ vertices of X with more than $\varepsilon p|Y|/2$ neighbours in X, then we have $e(X) \geq \varepsilon p|X||Y|/8$. Taking a maximum cut (A, B) of X, we have $e(A, B) \geq \varepsilon p|X||Y|/16$, and by (p, ν)-bijumbledness of Γ we therefore have

$$\tfrac{1}{16}\varepsilon p|X||Y| \leq p|A||B| + \nu\sqrt{|A||B|} \leq \tfrac{1}{4}p|X|^2 + \tfrac{1}{2}\nu|X|\,,$$

and since $|X| \leq \varepsilon|Y|/100$, we conclude $|Y| \leq 100\varepsilon^{-1}p^{-1}\nu$, a contradiction to the choice of C.

We conclude that there are $|X|/2$ vertices X' of X have at most $\varepsilon p|Y|/2$ neighbours in X, and hence at least $(1+\varepsilon/2)p|Y|$ neighbours in $Y \setminus X$. By the (p, ν)-bijumbledness of Γ we have

$$\tfrac{1}{2}|X|\big(1+\tfrac{\varepsilon}{2}\big)p|Y| \leq e(X', Y \setminus X) \leq \tfrac{1}{2}p|X||Y| + \nu\sqrt{\tfrac{1}{2}p|X||Y|}\,,$$

from which we have $\varepsilon C'p^{-1}\nu^2 \leq 2\sqrt{C'}\nu^2p^{-1}$, a contradiction to the choice of C'. □

3.3.2 Main lemmas

The idea of the proof of Theorem 3.20 is essentially the same as the one of the proof of Theorem 3.3. The lemma for H (Lemma 3.5) and the balancing lemma (Lemma 3.7) can be adopted as they stand. In place of the sparse blow-up lemma (Theorem 2.9) we use the version for pseudorandom graphs that we formulated in the previous subsection. Merely the lemma for G and the common neighbourhood lemma have to be modified for our needs and the proof of the theorem needs to be adjusted.

Briefly, the modifications that we make are replacing 'regular' with 'fully-regular' in the proofs, applying the lemmas for bijumbled graphs from above instead of their analogues for random graphs, and recalculating some error bounds. We start with stating and proving the bijumbled graph version of the lemma for G.

Lemma 3.28 (Lemma for G, bijumbled graph version)**.** *For each $\gamma > 0$ and integers $k \geq 2$ and $r_0 \geq 1$ there exists $d > 0$ such that for every $\varepsilon \in (0, 1/(2k))$ there exist $r_1 \geq 1$ and $c, C^* > 0$ such that the following holds for any n-vertex (p, ν)-bijumbled graph Γ with $\nu \leq cp^3n$ and $p > 0$.*

Let $G = (V, E)$ be a spanning subgraph of Γ with $\delta(G) \geq \big((k-1)/k + \gamma\big)pn$. Then there exists an integer r with $r_0 \leq kr \leq r_1$, a subset $V_0 \subseteq V$ with $|V_0| \leq C^* p^{-6}\nu^2 n^{-1}$, a k-equitable vertex partition $\mathcal{V} = \{V_{i,j}\}_{i\in[r],j\in[k]}$ of $V(G) \setminus V_0$, and a graph R^k_r on the vertex set $[r] \times [k]$ with $K^k_r \subseteq B^k_r \subseteq R^k_r$, with $\delta(R^k_r) \geq \big((k-1)/k + \gamma/2\big)kr$, and such that the following is true.

(G1) We have $\frac{n}{4kr} \leq |V_{i,j}| \leq \frac{4n}{kr}$ for every $i \in [r]$ and $j \in [k]$,

(G2) $\mathcal{V}$ is $(\varepsilon, d, p)_G$-fully-regular on R^k_r and $(\varepsilon, d, p)_G$-super-fully-regular on K^k_r,

(G3) $\big(N_\Gamma(v, V_{i,j}), V_{i',j'}\big)$ and $\big(N_\Gamma(v', V_{i,j}), N_\Gamma(v', V_{i',j'})\big)$ are $(\varepsilon, d, p)_G$-fully-regular pairs for every $\{(i, j), (i', j')\} \in E(R^k_r)$, $v \in V \setminus (V_0 \cup V_{i,j})$, and $v' \in V \setminus (V_0 \cup V_{i,j} \cup V_{i',j'})$,

(G4) we have $(1-\varepsilon)p|V_{i,j}| \leq |N_\Gamma(v, V_{i,j})| \leq (1+\varepsilon)p|V_{i,j}|$ for every $i \in [r]$, $j \in [k]$ and every $v \in V \setminus V_0$.

Observe that apart from the replacement of 'regular' with 'fully-regular', and 'random graph' with 'bijumbled graph', the difference between Lemma 3.4 and Lemma 3.28 is that V_0 may now be much larger. Nevertheless, the proof is very similar to the one of Lemma 3.4. Since the proof of Lemma 3.4 is quite long, but the modifications that need to be made are only a few, we concentrate on explaining the changes.

Sketch proof of Lemma 3.28. We begin the proof as in that of Lemma 3.4, setting up the constants in the same way, with the exception that we replace Lemmas 2.10 and 2.11 with Lemmas 3.25 and 3.26, respectively, and Proposition 3.8 with Proposition 3.27. We require C to be the maximum of the the C-outputs of Lemmas 3.25 and 3.26, and Proposition 3.27. We define

$$C^* = 100k^2 r_1^3 C/\varepsilon^*$$

as in the proof of Lemma 3.4, and set

$$c = 10^{-5}(\varepsilon^*)^3/(k^3 r_1^3 C^*)\,.$$

We now assume that Γ is (p, ν)-bijumbled with $\nu \leq cp^3 n$ rather than $G(n, p)$. In particular, by choice of c this implies that

$$10k^2 r_1^2 C p^{-2}\nu^2 n^{-1} \leq \varepsilon^* pn \quad \text{and} \quad 10k^2 r_1^3 C p^{-6}\nu^2 n^{-1} \leq \varepsilon^* n\,. \tag{3.17}$$

As a next step, we obtain a regular partition of $V(G)$ with a reduced graph containing B^k_r, exactly as in the proof of Lemma 3.4, using Proposition 3.27 in place of Proposition 3.8 to justify the use of Lemma 2.6. Here we need to use the fact that the regular pairs returned by Lemma 2.6 are fully-regular. The next place where we need to change something occurs in defining Z_1. In the definition of Z_1 we replace 'regular' with 'fully-regular'. Using Lemmas 3.25 and 3.26, and Proposition 3.8 with Proposition 3.27, we replace Equation (3.1) with

$$|Z_1| \leq kr_1^2 C p^{-6}\nu^2 n^{-1} + kr_1^2 C p^{-3}\nu^2 n^{-1} + 2kr_1 C p^{-2}\nu^2 n^{-1} \leq 4kr_1^2 C p^{-6}\nu^2 n^{-1} \overset{(3.17)}{\leq} \tfrac{\varepsilon^*}{kr_1} n\,.$$

Note that the final conclusion on the size of Z_1 is exactly as in Equation (3.1).

We can now continue following the proof of Lemma 3.4 until we come to estimate the size of Z_2, where we use Proposition 3.27 and replace Equation (3.2) with

$$|Z_2| \leq r_1 + kr_1 C p^{-2}\nu^2 n^{-1} \overset{(3.17)}{\leq} \tfrac{\varepsilon^*}{kr_1} pn\,.$$

Again, the final conclusion is as in Equation (3.2).

The next change we have to make is in estimating the size of V_0. We now have

$$|V_0| \le |Z_1| + |Z_2| \le 4kr_1^2 Cp^{-6}\nu^2 n^{-1} + r_1 + kr_1 Cp^{-2}\nu^2 n^{-1} \le C^* p^{-6}\nu^2 n^{-1}\,.$$

Finally, we need to assure that we have fully-regular pairs in Properties (G2) and (G3) rather than regular pairs. All other conclusions work verbatim.

We obtained fully-regular pairs from Lemma 2.6 and in the definition of Z_1, so that we only need Proposition 2.4 to return fully-regular pairs. We always apply Proposition 2.4 to pairs of sets of size at least $\varepsilon^* pn/r_1$, altering them by a factor ε^*. Now Proposition 3.27 shows that if X and Y are disjoint subsets of Γ with $|X|, |Y| \le (\varepsilon^* p)^{-1}\nu$, then $e_\Gamma(X,Y) \le (1+\varepsilon^*)p|X||Y|$, as required. By choice of c, we have $(\varepsilon^* p)^{-1}\nu \le (\varepsilon^*)^2 pn/r_1$, so that the condition of Proposition 2.4 to return fully-regular pairs is satisfied. □

Let us now turn to the second main lemma, that needs to be modified. The statement of the common neighbourhood lemma (Lemma 3.6) only changes by a replacement of 'regular' with 'fully-regular' and $G(n,p)$ with a bijumbled graph. However, the proof changes slightly more as the error bounds in the bijumbled graph versions of various lemmas are different.

Lemma 3.29 (Common neighbourhood lemma, bijumbled graph version)**.** *For each $d > 0$, $k \ge 1$, and $\Delta \ge 2$ there exists $\alpha > 0$ such that for every $\varepsilon^* \in (0,1)$ there exists $\varepsilon_0 > 0$ such that for every $r \ge 1$ and every $0 < \varepsilon \le \varepsilon_0$ there exists $c > 0$ such that the following is true. For any n-vertex (p,ν)-bijumbled graph Γ with $\nu \le cp^{\Delta+1}n$ and $p > 0$ the following holds.*

Let $G = (V,E)$ be a (not necessarily spanning) subgraph of Γ and $\{V_i \setminus W\}_{i\in[k]} \cup \{W\}$ a vertex partition of a subset of V such that the following is true for all distinct $i, i' \in [k]$.

- *(V1)* $\frac{n}{4kr} \le |V_i| \le \frac{4n}{kr}$,
- *(V2)* $(V_i, V_{i'})$ *is* $(\varepsilon, d, p)_G$*-fully-regular,*
- *(V3)* $|W| = \frac{\varepsilon pn}{16kr^2}$, *and*
- *(V4)* $|N_G(w, V_i)| \ge dp|V_i|$ *for every* $w \in W$.

Then there exists a tuple $(w_1, \dots, w_\Delta) \in \binom{W}{\Delta}$ such that for every $\Lambda, \Lambda^* \subseteq [\Delta]$, and all distinct $i, i' \in [k]$ we have

- *(W1)* $|\bigcap_{j\in\Lambda} N_G(w_j, V_i)| \ge \alpha p^{|\Lambda|}|V_i|$,
- *(W2)* $|\bigcap_{j\in\Lambda} N_\Gamma(w_j)| \le (1+\varepsilon^*)p^{|\Lambda|}n$,
- *(W3)* $(1-\varepsilon^*)p^{|\Lambda|}|V_i| \le |\bigcap_{j\in\Lambda} N_\Gamma(w_j, V_i)| \le (1+\varepsilon^*)p^{|\Lambda|}|V_i|$, *and*
- *(W4)* $\big(\bigcap_{j\in\Lambda} N_\Gamma(w_j, V_i), \bigcap_{j^*\in\Lambda^*} N_\Gamma(w_{j^*}, V_{i'})\big)$ *is* $(\varepsilon^*, d, p)_G$*-fully-regular if* $|\Lambda|, |\Lambda^*| < \Delta$ *and either* $\Lambda \cap \Lambda^* = \varnothing$ *or* $\Delta \ge 3$ *or both.*

The main modifications for the proof of Lemma 3.29 compared to the proof of Lemma 3.6 are to replace Lemmas 2.10 and 2.11 with Lemmas 3.25 and 3.26, and Proposition 3.8 with Proposition 3.27, as well as to replace all occurrences of 'regular' with 'fully-regular'. Again, we only state and explain the changes.

Sketch proof of Lemma 3.29. We begin the proof by setting up the constants as in the proof of Lemma 3.6, but appealing to Lemmas 3.25 and 3.26, and Proposition 3.27, rather than their random graph equivalents Lemmas 2.10 and 2.11, and Proposition 3.8.

Furthermore, we set

$$c = 10^{-20}2^{-2\Delta}\varepsilon^5(Ct_1kr)^{-4}.$$

Suppose that Γ is an n-vertex (p,ν)-bijumbled graph with $\nu \le cp^{\Delta+2}n$ rather than a random graph.

In order to apply Lemma 2.5 to G, we need to observe that its condition is satisfied by Proposition 3.27 and because $\varepsilon^{-1}p^{-1}\nu < 10^{-10}\varepsilon^4pn/(k^4r^4)$ by choice of c. The same inequality justifies further the use of Proposition 3.27 to find the desired set W'. Estimating the size of W', we replace (3.11) with

$$|W'| \ge 10^{-11}\frac{\varepsilon^4pn}{t_1k^4r^4} \ge 10^5Cp^{-2}\nu\,, \tag{3.18}$$

where the final inequality is by choice of c.

We only need to change the statement of Claim 3.11 by replacing 'regular' with 'fully-regular' in Properties (L1) and (L6). However we need to make more changes to its inductive proof. The base case remains trivial. In the induction step, we need to replace (3.12) with

$$\Big|\bigcap_{j\in\Lambda} N_\Gamma(w_j,V_i')\Big| \ge (1-\varepsilon_0)^{\Delta-2}p^{\Delta-2}\frac{n}{8tr} \ge 10^5Cp^{-4}\nu\,,$$

where the final inequality is by choice of c. This, together with $|W'| \ge 10^5Cp^{-2}\nu$ from (3.18), justifies that we can apply Lemma 3.25. We obtain that at most $2^\Delta k^2Cp^{-3}\nu^2(8krt_1)/n$ vertices w in W violate (L1).

The estimate on the number of vertices violating (L2) does not change.

For (L4), we need to observe that $\big|\bigcup_{j\in\Lambda} N_\Gamma(w_j,V_i')\big| = (1\pm\varepsilon_0)^{|\Lambda|}p^{|\Lambda|}|V_i'|$, and in particular by choice of ε_0 and c this quantity is at least $Cp^{-1}\nu$. Then Proposition 3.27 then gives that at most $2^{\Delta+1}kCp^{-2}\nu^2(8krt_1)/n$ vertices destroy (L4), and the same calculation gives the same bound for the number of vertices violating (L3) and (L5).

Finally, for (L6), we need to use the inequality $(1-\varepsilon_0)^{\Delta-1}p^{\Delta-1}n/(4kr) \ge Cp^{-2}\nu$, which holds by choice of c, to justify that Lemmas 3.25 and 3.26 can be applied. If $\Delta = 2$, then we only use Lemma 3.25, with an input regular pair with both sets having size at least $n/(4kr)$. Hence, the number of vertices violating (L6) in this case is at most $2^{2\Delta}k^2Cp^{-3}\nu^2(4kr)/n$. If $\Delta \ge 3$, we use both Lemma 3.25 and Lemma 3.26. The set playing the role of X in Lemma 3.25 has size at least $(1-\varepsilon_0)^{\Delta-2}p^{\Delta-2}n/(4kr)$, while we apply Lemma 3.26 with both sets of the regular pair having at least this size. As a consequence, the number of vertices violating (L6) is at most $2^{2\Delta+1}k^2Cp^{-6}\nu^2(1-\varepsilon_0)^{2-\Delta}p^{2-\Delta}(4kr)/n$ for the case $\Delta \ge 3$.

Putting this together, for the case $\Delta = 2$ we replace (3.13) with the following upper bound for the number of vertices $w \in W'$ that cannot be chosen as $w_{\ell+1}$.

$$2^\Delta k^2Cp^{-3}\nu^2\tfrac{8krt_1}{n} + 2^\Delta k\varepsilon^{**}_\Delta|W'| + 3\cdot 2^{\Delta+1}kCp^{-2}\nu^2\tfrac{8krt_1}{n} + 2^{2\Delta}k^2Cp^{-3}\nu^2\tfrac{4kr}{n} \le \tfrac{|W'|}{2},$$

where the latter inequality is by choice of c and ε^{**}_Δ. This completes the induction step for $\Delta = 2$. For $\Delta \ge 3$, we replace the upper bound (3.14) with

$$\begin{aligned}2^\Delta k^2Cp^{-3}\nu^2\tfrac{8krt_1}{n} + 2^\Delta k\varepsilon^{**}_\Delta|W'| + 3\cdot 2^{\Delta+1}kCp^{-2}\nu^2\tfrac{8krt_1}{n} +&\\ 2^{2\Delta+1}k^2Cp^{-6}\nu^2(1-\varepsilon_0)^{2-\Delta}p^{2-\Delta}\tfrac{4kr}{n} \le \tfrac{|W'|}{2},&\end{aligned}$$

where we used the choice of c and ε_0 as well as $\varepsilon^{**}_{\Delta}$. This completes the induction step for $\Delta \geq 3$.

Therefore, the modified Claim 3.11 holds, which implies the statement of Lemma 3.29 as in the proof of Lemma 3.6. □

3.3.3 Proof of the theorem

In this subsection we give the proof of Theorem 3.20, which is again similar to that of Theorem 3.3. For this reason we mainly focus on the modifications that need to be made.

Sketch proof of Theorem 3.20. We begin as in the proof of Theorem 3.3 by setting up the constants as there, but replacing Lemma 3.4 with Lemma 3.28, Lemma 3.6 with Lemma 3.29, Theorem 2.9 with Theorem 3.22, and Proposition 2.16 with Proposition 3.27. More precisely, we define the constants as follows.

Given $\gamma > 0$, $\Delta \geq 2$, and $k \geq 2$, set $r_0 = 10/\gamma$ and $D = \Delta$. Let d be returned by Lemma 3.28, with input γ, k and r_0. Let α be returned by Lemma 3.29 with input d, k and Δ. Now let $\varepsilon_{\text{BL}} > 0$ and $\rho > 0$ be returned by Theorem 3.22 with input Δ, $\Delta_{R'} = 3k$, $\Delta_J = \Delta$, $\vartheta = 1/(100D)$, $\zeta = \alpha/4$, d and $\kappa := 64$. Next, putting $\varepsilon^* = \varepsilon_{\text{BL}}/8$ into Lemma 3.29 returns $\varepsilon_0 > 0$. We set

$$\varepsilon = \min\bigl\{\varepsilon_0, d/8, \varepsilon^*/(4D), 1/(16k)\bigr\}.$$

Putting ε into Lemma 3.28 returns r_1, c_1, and C_1^* . Next, Lemma 3.7, for input k, r_1, Δ, γ, d, and 8ε, returns $\xi \in \bigl(0, 1/(10kr_1)\bigr)$ and C_2^*. We set

$$\beta = 10^{-12}\xi^2/(\Delta k^4 r_1^2) \text{ and } \mu = 10^{-5}\varepsilon^2/(kr_1).$$

Next let $C_3^* > 0$ be the maximum of the outputs of Proposition 3.27 with input ε and input μ^2 and of Lemma 2.19 with input $\varepsilon\mu$ and Δ. Let $C^* = \max\{C_1^*, C_2^*, C_3^*\}$ and set

$$C = 10^{10}k^2 r_1^2 \Delta^{2r_1+20} C^*/(\varepsilon^2\xi\mu^2) \text{ and } z = 10/\xi.$$

Let c_2 be returned by Lemma 3.29 with input r_1 and ε, and let c_3 be the outcome of Theorem 3.22 with input r_1. Finally, set

$$c = \min\{c_1, c_2, c_3, 10^{-50}\varepsilon^8\mu\rho\xi^2(\Delta k r_1 C)^{-10}\}.$$

Let Γ be an n-vertex (p,ν)-bijumbled graph with $\nu \leq cp^{\max\{4,(3\Delta+1)/2\}}n$. By Proposition 3.21 we have

$$p \geq C^*\left(\tfrac{\log n}{n}\right)^{1/2}. \tag{3.19}$$

Let H be a graph as in the statement of the theorem and observe that $c < \beta$.

We continue following the proof of Theorem 3.3. We now assume that the first βn vertices of $\mathcal{L}$ include $Cp^{-6}\nu^2 n^{-1}$ vertices that are not contained in any triangles of H. We appeal to Lemma 3.28 rather than Lemma 3.4 to obtain a partition of $V(G)$. This partition has an exceptional set of size $|V_0| \leq C^* p^{-6}\nu^2 n^{-1}$, but still satisfies (G1a) and (G4a), and (G2a) and (G3a) when 'regular' is replaced by 'fully-regular' in both statements.

The applications of Lemma 3.5 and Lemma 2.19 are identical and the deduction of (3.15) is still valid by (3.19). The pre-embedding is also identical, except that we replace each occurrence of $C^*\max\{p^{-2}, p^{-1}\log n\}$ with $C^* p^{-6}\nu^2 n^{-1}$, and that we replace the application of Proposition 3.8 justifying that at each visit to Line 1 we have at least $\mu^2 pn/4$ choices with an application of Proposition 3.27. To verify the condition of the latter, we use $\varepsilon n/(8r) \geq C^* p^{-1}\nu$, and to see that this yields a contradiction we use the inequality $|Z| \geq \mu^2 pn/\big(4(\Delta+1)\big) \geq 2C^* p^{-2}\nu^2 8r/(\varepsilon n)$. Both inequalities follow by choice of c.

Moving on, we justify Claim 3.14 by observing that $\varepsilon n/4kr_1 \geq Cp^{-1}\nu$, which allows us to apply Proposition 3.27 in place of Proposition 3.8, and that $2krC^* p^{-2}\nu^2 4kr_1/\varepsilon n \leq |Y|/2$, both inequalities following by choice of c.

Now Lemma 3.29, in place of Lemma 3.6, finds the desired vertices $w_1, \ldots, w_\ell$. Our construction of f^* and its properties are identical, while Lemma 3.29 gives (G1a)–(G8a), with 'regular' replaced by 'fully-regular' in (G2a), (G3a) and (G7a). The deduction of (G1b)–(G8b) is identical, except that we use the 'fully-regular' consequence of Proposition 2.4. To justify this, observe that each time we apply Proposition 2.4, we apply it to a regular pair with sets of size at least $(1-\varepsilon^*)p^{\Delta-1}n/(4kr)$ by (G1a) and (G6a), and we change the set sizes by a factor $(1 \pm 2\mu)$ so that Proposition 3.27 gives the required condition. To check this in turn, we need to observe that $2\mu(1-\varepsilon^*)p^{\Delta-1}n/(4kr) \geq 100\mu^{-1}p^{-1}\nu$, which follows by choice of c. We can thus replace 'regular' with 'fully-regular' in (G2b), (G3b) and (G7b).

Next, we still have $3\Delta^{r+10}|V_0| \leq \xi n/10$, so that $|V'_{i,j}| = |W'_{i,j}| \pm \xi n$ is still valid for each $i \in [r]$ and $j \in [k]$. This, together with (3.19), Proposition 3.27, and the inequality $\varepsilon^2 \xi pn/(50000kr_1) \leq 100\varepsilon^{-2}\xi^{-1}p^{-1}\nu$, justifies that we can apply Lemma 3.7 to obtain (G1c)–(G6c), with 'regular' replaced by 'fully-regular' in (G2c) and (G3c). Finally, to obtain (G7c) with 'regular' replaced by 'fully-regular', we use Proposition 2.4, with the condition to output fully-regular pairs guaranteed by the inequality $10^{-20}\varepsilon^4 k^{-3} r_1^{-3} p^{\Delta-1} n \geq 10^{20}\varepsilon^{-4}k^3 r_1^3 Cp^{-1}\nu$, which follows by choice of c, and Proposition 3.27.

Finally, we verify the conditions for Theorem 3.22. The only point where we have to be careful is with the number of image restricted vertices. The total number of image restricted vertices in H' is at most $\Delta^2|V_0| \leq \Delta^2 C^* p^{-6}\nu^2 n^{-1}$, which by choice of c and by (G1c) is smaller than $\rho p^\Delta |V_{i,j}|$ for any $i \in [r]$ and $j \in [k]$, justifying that $(\mathcal{I}, \mathcal{J})$ is indeed a $(\rho p^\Delta, \alpha/4, \Delta, \Delta)$-restriction pair. The remaining conditions of Theorem 3.22 are verified as in the proof of Theorem 3.3, and applying it we obtain an embedding ϕ of H' into $G \setminus \mathrm{im}(\phi_{t_f})$. This embedding yields together with $\phi \cup \phi_{t_f}$ the desired embedding of H into G. $\square$

Finally, we present the deduction of Theorem 3.19 by Theorem 3.20, which is essentially the same as the deduction of Theorem 3.1 by Theorem 3.3.

Proof of Theorem 3.1. Given γ, Δ, and k, let $c^* > 0$ and $z > 0$ be returned by Theorem 3.20 with input γ, Δ, and k. Set $c = (c^*)^2/2$. Let H be a k-colourable graph on n vertices with $\Delta(H) \leq \Delta$ such that there exists a set W of at least $c^{-1}p^{-6}\nu^2 n^{-1}$ vertices in $V(H)$ that are not contained in any triangles of H and such that there exists a labelling $\mathcal{L}$ of its vertex set of bandwidth at most cn.

By the choice of c we find an interval $I \subseteq \mathcal{L}$ of length $c^* n$ containing a subset $F \subseteq W$ with $|F| = (c^*)^{-1}p^{-6}\nu^2 n^{-1}$. Now we can rearrange the labelling $\mathcal{L}$ to a labelling $\mathcal{L}'$ of bandwidth at most $2cn \leq c^* n$ such that F is contained in the first $c^* n$ vertices in $\mathcal{L}'$.

Then, by Theorem 3.20 we know that for every (p, ν)-bijumbled graph the following holds if $\nu \leq c^* p^{\max\{4,(3\Delta+1)/2\}} n$ and hence in particular if $\nu \leq cp^{\max\{4,(3\Delta+1)/2\}} n$. If G is a spanning subgraph of Γ with $\delta(G) \geq \big((k-1)/k + \gamma\big)pn$, then G contains a copy of H, which finishes the proof. $\square$

3.4 Remarks on optimality

In order to conclude this chapter, we would like to collect several remarks about the optimality and non-optimality of the assumptions on the parameters in our results, which might serve as starting points for possible further improvements.

The main theorems in this chapter place restrictions on the graphs H with respect to whose containment random or pseudorandom graphs have local resilience. It was shown by Huang, Lee, and Sudakov [95] that the restrictions in terms of vertices not contained in triangles are necessary. Given $\varepsilon > 0$ and $p = o(1)$, if Γ is a random graph $G(n,p)$ or a pseudorandom graph with density p, then, if n is large enough, one can delete all edges in the neighbourhood of a given vertex of Γ without deleting (a.a.s.) more than $p^2n \le \varepsilon pn$ edges at any vertex. Thus if H is a graph all of whose vertices are in triangles, the local resilience of Γ with respect to the containment of H is a.a.s. $o(1)$ if $p = o(1)$.

While for this reason a certain number of vertices of H needs to have an independent neighbourhood, we believe that in Theorem 3.1 the bound on the number of vertices of H that may not be in triangles can be relaxed to C^*p^{-2} rather than $C^* \max\{p^{-2}, p^{-1}\log n\}$. In fact, if p is large, Theorem 3.19 gives stronger results than Theorem 3.1. Since $G(n,p)$ is a.a.s. $\big(p, O(\sqrt{pn})\big)$-bijumbled, by Theorem 3.19 we require at most C^*p^{-5} vertices of H to have independent neighbourhoods, which for large p is significantly less than $C^*p^{-1}\log n$. We note that this is also a significant improvement on the results of Huang, Lee, and Sudakov [95], who proved a bandwidth theorem for dense random graphs. The number of vertices in H that they require to have independent neighbourhood grows as a tower type function of p^{-1}. Moreover, they require that these vertices are well distributed in the bandwidth labelling.

As with Theorem 3.1, we believe that for large p the number of vertices of H which are required to not be in triangles or copies of C_4 should be C^*p^{-2}. However in this range of p, Theorem 3.15 has no advantages over Theorem 3.1.

In Theorem 3.19 the requirement that $C^*p^{-6}\nu^2n^{-1}$ vertices of H are not contained in any triangles is due to Lemma 3.24. This lemma is proved by Allen, Böttcher, Skokan, and Stein in [12], where it is conjectured that the requirement is not optimal.

Instead of restricting the number of vertices of H that are contained in triangles, one can also impose a further restriction on G in Theorem 3.1. If we require that G contains in addition a positive proportion of the copies of $K_{\Delta+1}$ in Γ at each vertex, then we can show that G contains any k-colourable bounded degree spanning subgraph H with sublinear bandwidth. More precisely, in joint work [6] with Allen, Böttcher, Schnitzer, and Taraz, we prove the following theorem.

Theorem 3.30. *For each $\gamma > 0$, $\Delta \ge 2$, $k \ge 2$ and $0 \le s \le k-1$, there exist constants $\beta^* > 0$ and $C^* > 0$ such that the following holds asymptotically almost surely for $\Gamma = G(n,p)$ whenever $p \ge C^* (\log n/n)^{1/\Delta}$.*

Let G be a spanning subgraph of Γ with $\delta(G) \ge \big((k-1)(k+\gamma\big)pn$ such that for each $v \in V(G)$ there are at least $\gamma p^{\binom{s}{2}}(pn)^s$ copies of K_s in $N_G(v)$. Let H be a graph on n vertices with $\Delta(H) \le \Delta$, bandwidth at most β^*n and suppose that there is a proper k-colouring of $V(H)$ such that there are at least $C^* \max\{p^{-2}, p^{-1}\log n\}$ vertices in $V(H)$ whose neighbourhood is coloured by at most s colours. Then G contains a copy of H.

Let us now turn to the bound on the edge probability in Theorem 3.1. For $\Delta = 2$, Theorem 3.1 is optimal up to possibly the logarithmic factor as the statement is certainly false for $p = o(n^{-1/2})$ since then $G(n,p)$ has a.a.s. local resilience $o(1)$ with respect to containing even one triangle. However, we believe that for general values of Δ, the lower bounds on p in Theorem 3.1 and Theorem 3.15 are not optimal.

The bijumbledness requirement of Theorem 3.19 is due to Theorem 3.22, which was proved in [9] by Allen, Böttcher, Hàn, Kohayakawa, and Person. It is suggested there that the statement could still hold given only $\big(p, cp^{\Delta+C}n\big)$-bijumbledness for some $C > 0$. Such an improvement would immediately improve our results correspondingly.

Finally, we would like to mention that the additional restriction that we place in Theorem 3.15 of having many vertices of H which are neither in triangles nor in four-cycles is an artefact of our proof. It would be possible to remove the stipulation regarding four-cycles since one can prove a version of Lemma 3.6 capable of embedding vertices in a degeneracy order. However this comes at the cost of a worse lower bound on p.

4

A Dirac-type theorem of Hamilton Berge cycles in random hypergraphs

In this chapter we are concerned with Berge cycles in hypergraphs. In particular, we study the robustness of complete and sparse random r-uniform hypergraphs with respect to Berge Hamiltonicity. To measure the robustness of hypergraphs we use the concept of local resilience as in the setting of graphs in Chapter 3 (see Subsection 1.2.1 for an introduction). Moreover, we investigate Maker-Breaker and Avoider-Enforcer games played on the edge set of complete r-uniform hypergraphs with respect to building and avoiding Berge cycles, respectively.

In Section 4.1 we prove the above mentioned local resilience result for Hamilton Berge cycles in random r-uniform hypergraphs by showing that for every integer $r \geq 3$ and for every real $\gamma > 0$ asymptotically almost surely every spanning subhypergraph $H \subseteq H^{(r)}(n,p)$ with minimum vertex degree $\delta_1(H) \geq \left(\frac{1}{2^{r-1}} + \gamma\right) p\binom{n}{r-1}$ contains a Hamilton Berge cycle whenever $p \geq \log^{8r}(n)/n^{r-1}$. Our proof is based on the absorbing method developed by Rödl, Ruciński, and Szemerédi [144]. Furthermore, the ideas that are used in the proof of a Dirac-type result for random directed graphs due to Ferber, Nenadov, Noever, Peter, and Škoric [78] are of particular interest as they allow us to apply the absorbing method in this sparse scenario.

We briefly study the local resilience of complete r-uniform hypergraphs with respect to weak and Berge Hamiltonicity in Section 4.2, where we prove a tight minimum degree condition for weak Hamilton cycles and an asymptotically tight one for Berge Hamilton cycles. The first result is proved by applying Dirac's theorem to an underlying graph of the hypergraph and the proof of the second one uses an extension of the proof of Dirac's theorem.

Next we study positional games played on the edge set of $K_n^{(r)}$. We first discuss a relation between local resilience of hypergraphs and Maker-Breaker games. This yields a lower bound on the threshold bias for Maker-Breaker games, where Maker aims to build a hypergraph that contains a Hamilton Berge cycle. Moreover, we examine a misère version of such games. More precisely, we prove bounds on the threshold biases for monotone and strict Avoider-Enforcer games played on $E(K_n^{(3)})$, where Avoider's task is to keep his hypergraph (almost) Berge-acyclic. Last, we present some concluding remarks and open questions in Section 4.4.

The results of Sections 4.1 and 4.2 are joint work with Dennis Clemens and Yury Person [54] and the proofs in Section 4.3 are extensions of a joint work with Dennis Clemens, Yury Person, and Tuan Tran [55] to the setting of hypergraphs.

4.1 Berge Hamiltonicity in random hypergraphs

In this section we prove a local resilience result, which can be seen as a hypergraph analogue of a result by Lee and Sudakov [124], who proved that a.a.s. the local resilience of $G(n,p)$ with respect to Hamiltonicity is at least $\big(1/2-o(1)\big)$ whenever $p=\Omega(\log n/n)$. More precisely, we prove the following theorem.

Theorem 4.1. *For every integer $r\geq 3$ and every real $\gamma>0$ the following holds asymptotically almost surely for $\mathcal{H}=H^{(r)}(n,p)$ if $p\geq \frac{\log^{8r} n}{n^{r-1}}$. Let $H\subseteq\mathcal{H}$ be a spanning subgraph with $\delta_1(H)\geq\big(\frac{1}{2^{r-1}}+\gamma\big)\,p\binom{n}{r-1}$. Then H contains a Hamilton Berge cycle.*

The minimum degree condition in Theorem 4.1 is asymptotically tight, which can be seen as follows. Roughly speaking, given $\mathcal{H}=H^{(r)}(n,p)$ together with a partition $V(\mathcal{H})=V_1\cup V_2$ with $|V_1-V_2|\leq 1$ chosen uniformly at random among all such partitions, then a.a.s. the degree of every $v\in V_i$ into V_i is at least $\big(\frac{1}{2^{r-1}}-\gamma\big)\,p\binom{n}{r-1}$ for every $\gamma>0$ and $i\in[2]$ if n is sufficiently large. Deleting all edges between V_1 and V_2 yields a hypergraph that does not contain a Hamilton Berge cycle and that satisfies $\delta_1(H)\geq\big(\frac{1}{2^{r-1}}-\gamma\big)\,p\binom{n}{r-1}$.

The bound on p is best possible up to possibly the logarithmic factor since $\log n/n^{r-1}$ is the threshold for the appearance of a weak Hamilton cycle in $H^{(r)}(n,p)$ (Theorem 1.9 by Poole [139]). To the best of our knowledge, there are no non-trivial upper bounds known for the threshold of $H^{(r)}(n,p)$ with respect to Berge Hamiltonicity, except for the ones that follow from results with other notions of cycles (see e.g. [72]). Theorem 4.1 yields immediately the following upper bound on the threshold of $H^{(r)}(n,p)$ with respect to Berge Hamiltonicity.

Corollary 4.2. *Let $r\geq 3$. Then $H^{(r)}(n,p)$ contains asymptotically almost surely a Hamilton Berge cycle if $p\geq\frac{\log^{8r} n}{n^{r-1}}$.*

This section is structured as follows. In Subsection 4.1.1 we introduce necessary definitions and prove almost sure properties of random hypergraphs. Then, in Subsection 4.1.2, we outline the proof of Theorem 4.1. In Subsections 4.1.3–4.1.5 we derive the main lemmas for the proof of Theorem 4.1, which we present in Subsection 4.1.6.

4.1.1 Preliminaries

For the proof of Theorem 4.1 we need the following definitions of weak and Berge paths. A *weak Berge path* (or simply *weak path*) is an alternating sequence $(v_1,e_1,v_2,\ldots,v_k)$ of distinct vertices $v_1,\ldots,v_k$ and (not necessarily distinct) hyperedges $e_1,\ldots,e_{k-1}$ such that $v_i,v_{i+1}\in e_i$ for every $i\in[k-1]$. A weak path is called *Berge path* if all its hyperedges are distinct.

For a weak path $P=(v_1,e_1,\ldots,e_{k-1},v_k)$ we denote by $E(P):=\{e_1,\ldots,e_{k-1}\}$ the set of hyperedges of P, by $V^*(P):=\{v_1,\ldots,v_k\}$ the set of vertices in the sequence of P, and by $V(P):=\bigcup_{i\in[k-1]}e_i$ the union of the hyperedges of P. We say that P *connects* v_1 to v_k and call v_1 and v_k *endpoints* of P. We set $\mathrm{End}(P):=\{v_1,v_k\}$.

The *length* of a weak path P is defined as $|V^*(P)|-1$. In particular, if P is a Berge path, then the length of P is exactly the number of hyperedges of P. For the sake of simplicity, we do not consider weak paths to be oriented. In other words, the weak paths $(v_1,e_1,\ldots,e_{k-1},v_k)$ and $(v_k,e_{k-1},\ldots,e_1,v_1)$ are the same. Moreover, given two weak paths $P=(v_1,e_1,\ldots,e_{k-1},v_k)$ and $Q=(v_k,e'_1,\ldots,e'_{k'-1},v'_{k'})$ with $|V^*(P)\cap V^*(Q)|=1$, we denote by $P\cdot Q$ the weak path $(v_1,e_1,\ldots,e_{k-1},v_k,e'_1,\ldots,e'_{k'-1},v'_{k'})$.

We say that two Berge paths $P = (v_1, e_1, \dots, e_{k-1}, v_k)$ and $P' = (v_1', e_1', \dots, e_{k'-1}', v_{k'}')$ are *edge-disjoint* if $e_i \neq e_j$ for all $i \in [k]$ and $j \in [k']$.

Let $H = (V, E)$ be an r-uniform hypergraph. For every vertex $v \in V$, the *link* of v in H is a subset of $\binom{V}{r-1}$ that consists of all $(r-1)$-tuples of vertices that form together with v an hyperedge of H, i.e.

$$\mathrm{link}_H(v) = \left\{ e \in \binom{V}{r-1} : e \cup \{v\} \in E \right\}.$$

We note that, in case $\mathcal{H} \sim H^{(r)}(n,p)$, the link of any vertex of $\mathcal{H}$ is distributed as $H^{(r-1)}(n,p)$. The following lemma gives an asymptotically almost sure upper bound on the maximum degree of $H^{(r)}(n,p)$.

Lemma 4.3. *Let $r \geq 3$ be an integer, and let $\varepsilon > 0$ and $p \in [0,1]$ be reals. Then $\mathcal{H} = H^{(r)}(n,p)$ satisfies the following properties with probability $\big(1 - o(n^{-1})\big)$:*

$$\Delta(\mathcal{H}) \leq 2n^{r-1}p + \log^{1.1} n.$$

Proof. For every $v \in V(\mathcal{H})$ it holds $\deg_{\mathcal{H}}(v) \sim \mathrm{Bin}\left(\binom{n-1}{r-1}, p\right)$ and thus $\mathbb{E}[\deg_{\mathcal{H}}(v)] \leq n^{r-1}p$. Setting $t = n^{r-1}p + \log^{1.1} n$ we obtain $2(\mathbb{E}[\deg_{\mathcal{H}}(v)] + t/3) \leq 3t$ for n sufficiently large. Hence Theorem 2.17 and applying the union bound yields

$$\begin{aligned}\mathbb{P}\big[\exists v \in V(\mathcal{H}) : \ \deg_{\mathcal{H}}(v) \geq \mathbb{E}[\deg_{\mathcal{H}}(v)] + t\big] &\leq n \exp\left(-\frac{t^2}{2(\mathbb{E}[X] + t/3)}\right) \\ &\leq n \exp\left(-\frac{t}{3}\right) \leq \exp\left(\log n - \frac{\log^{1.1} n}{3}\right) = o\left(\frac{1}{n}\right),\end{aligned}$$

which finishes the proof. □

The following lemma provides an upper bound on the typical number of Berge paths of length 2 in $H^{(r)}(n,p)$.

Lemma 4.4. *Let $r \geq 3$ be an integer, and let $\varepsilon > 0$ and $p \geq n^{-r}$ be reals. Then, with probability $\big(1 - o(n^{-1})\big)$, the random hypergraph $\mathcal{H} = H^{(r)}(n,p)$ contains at most $n^{2r-1+\varepsilon}p^2 + \log^2 n$ Berge paths of length 2.*

Proof. Given r, ε, and p, we set $t = \lceil 4/\varepsilon \rceil$. We consider the set

$$S_{\mathcal{H}} := \{(S_1, S_2, \dots, S_t) : \ S_1, S_2, \dots, S_t \ \text{ are edge-disjoint Berge paths of length 2 in } \mathcal{H}\}.$$

Every Berge path of length 2 covers at most $2r-1$ vertices of the hypergraph. Hence, as an upper bound on the expected value of the cardinality of $S_{\mathcal{H}}$, we obtain $\mathbb{E}[|S_{\mathcal{H}}|] \leq n^{(2r-1)t}p^{2t}$. By Markov's inequality (Lemma 2.15) we therefore have

$$\mathbb{P}\big[|S_{\mathcal{H}}| \geq n^{(2r-1)t+2}p^{2t}\big] \leq \frac{1}{n^2}.$$

In particular, using Lemma 4.3, we know that with probability $\big(1 - o(n^{-1})\big)$ the random graph $\mathcal{H}$ satisfies the following two properties:

(a) $|S_{\mathcal{H}}| \leq n^{(2r-1)t+2}p^{2t}$,

(b) $\Delta(\mathcal{H}) \leq 2n^{r-1}p + \log^{1.1} n$.

It now suffices to show that for every sufficiently large integer n and every graph $\mathcal{H}$ with Properties (a) and (b), the number N_2 of Berge paths of length 2 is at most $n^{2r-1+\varepsilon}p^2 + \log^2 n$. We may assume that $N_2 > 2rt\Delta(\mathcal{H})$ holds as otherwise we obtain $N_2 \leq n^{2r-1+\varepsilon}p^2 + \log^2 n$ immediately by Property (b) and by the choice of p, assuming n to be large enough.

Given a Berge path P of length 2 in $\mathcal{H}$, let us count how many Berge paths of length 2 there are at most such that each of these paths shares one hyperedge with P. Let $P = (v_1, e_1, v_2, e_2, v_3)$. The hyperedge e_1 can occur in at most $r\Delta(\mathcal{H})$ Berge paths of length 2. The same is true for e_2. Hence there are at most $2r\Delta(\mathcal{H})$ Berge paths that share a hyperedge with P. As a consequence we obtain

$$|S_{\mathcal{H}}| \geq N_2(N_2 - 2r\Delta(\mathcal{H}))(N_2 - 4r\Delta(\mathcal{H})) \cdots (N_2 - 2r(t-1)\Delta(\mathcal{H})) > (N_2 - 2rt\Delta(\mathcal{H}))^t.$$

Using Properties (a) and (b) we thus conclude that for n sufficiently large we have

$$N_2 \leq n^{2r-1+2/t}p^2 + 2rt\Delta(\mathcal{H}) < n^{2r-1+\varepsilon}p^2 + \log^2 n. \qquad \square$$

In the proof of Theorem 4.1 we need only a few asymptotically almost sure properties of $H^{(r)}(n,p)$ for our arguments. Therefore, rather than proving Theorem 4.1 for $H^{(r)}(n,p)$, we prove it for pseudorandom graphs that have these sufficient properties. We define pseudorandomness as follows.

Definition 4.5 ((p,ε)-pseudorandom)**.** *Let $r \geq 3$. We say that an r-uniform hypergraph H on n vertices is* (p,ε)-pseudorandom *if H satisfies the following properties.*

(H1) For every disjoint sets $A, B \subseteq V(H)$ with $|A|, |B| \geq \frac{n}{10^r \log^5 n}$ we have

$$e_H(A, B^{(r-1)}) \leq (1+\varepsilon)p|A|\binom{|B|}{r-1},$$

and for every set $C \subseteq V(H)$ with $|C| \geq \varepsilon \frac{n}{10^r \log^5 n}$ we have

$$e_H(A, B^{(r-2)}, C) \leq (1+\varepsilon)p|A|\binom{|B|}{r-2}|C|.$$

(H2) For every disjoint sets $A, B \subseteq V(H)$ with $|A|, |B| \leq \frac{n}{\log^5 n}$ and $|B| \in \{|A|, 2|A|\}$ and for every set $R \subseteq V(H)$, we have

$$e_H(A, B, R^{(r-2)}) \leq |A||B|\binom{|R|}{r-2}p + \varepsilon\frac{|A|n^{r-1}p}{\log^5 n}.$$

(H3) For every $v \in V(H)$ there are at most $n^{2r-3+\varepsilon}p^2 + \log^2 n$ Berge paths of length 2 in $\mathrm{link}_H(v)$.

Now we show that $H^{(r)}(n,p)$ is a.a.s. (p,ε)-pseudorandom for suitable values of p.

Lemma 4.6. *Let $r \geq 3$ be an integer, and let $\varepsilon > 0$ and $p \geq \log^{8r} n/n^{r-1}$ be reals. Then $\mathcal{H} = H^{(r)}(n,p)$ is a.a.s. (p,ε)-pseudorandom.*

Proof. Let $r \geq 3$, $\varepsilon > 0$, and $p \geq \log^{8r} n/n^{r-1}$ be given. We show that $\mathcal{H} = H^{(r)}(n,p)$ satisfies each of the Properties (H1)–(H3) from Definition 4.5 with probability $\big(1-o(1)\big)$. This implies then that $\mathcal{H}$ is a.a.s. (p,ε)-pseudorandom.

We start with Property (H3). Let $v \in V(\mathcal{H})$. We know from Lemma 4.4 that $\mathrm{link}_{\mathcal{H}}(v)$ contains at most $n^{2r-3+\varepsilon}p^2 + \log^2 n$ Berge paths of length 2 with probability $\big(1-o(n^{-1})\big)$. Applying the union bound, we get that $\mathcal{H}$ fulfils with probability $\big(1-o(1)\big)$ that $\mathrm{link}_{\mathcal{H}}(v)$ contains at most $n^{2r-3+\varepsilon}p^2 + \log^2 n$ Berge paths of length 2 for every vertex $v \in V(\mathcal{H})$.

Next we show that Property (H1) holds with probability $\big(1-o(1)\big)$. Let A and B be disjoint subsets of $V(\mathcal{H})$ with $|A|,|B| \geq n/(10^r \log^5 n)$. We define

$$X = e_{\mathcal{H}}(A, B^{(r-1)}).$$

Note that we have $X \sim \mathrm{Bin}\left(e_{K_n^{(r)}}(A, B^{(r-1)}), p\right)$ and therefore, since A and B are disjoint,

$$\mathbb{E}[X] = |A|\binom{|B|}{r-1}p.$$

Let $t := \varepsilon|A|\binom{|B|}{r-1}p$ and observe that $2\big(\mathbb{E}[X]+t/3\big) \leq 3t/\varepsilon$. By applying Chernoff's inequality (Theorem 2.17) we obtain

$$\mathbb{P}\big[X \geq \mathbb{E}[X]+t\big] \leq \exp\left(-\frac{t^2}{2(\mathbb{E}[X]+t/3)}\right) \leq \exp\left(-\frac{\varepsilon t}{3}\right) \leq \exp\left(-\frac{\varepsilon^2}{3}p|A|\binom{|B|}{r-1}\right).$$

Now, let $C \subseteq V(\mathcal{H})$ with $|C| \geq \varepsilon n/(10^r \log^5 n)$ and let

$$X_{A,B,C} = e_{\mathcal{H}}(A, B^{(r-2)}, C).$$

Observe that $X_{A,B,C} \sim \mathrm{Bin}\left(e_{K_n^{(r)}}(A, B^{(r-2)}, C), p\right)$ and therefore $\mathbb{E}[X_{A,B,C}] \leq |A|\binom{|B|}{r-2}|C|p$. Let $t' := \varepsilon|A|\binom{|B|}{r-2}|C|p$ and note that $2\big(\mathbb{E}[X_{A,B,C}]+t'/3\big) \leq 3t'/\varepsilon$. Therefore, by applying Chernoff's inequality (Theorem 2.17) we obtain

$$\begin{aligned}\mathbb{P}\big[X_{A,B,C} \geq \mathbb{E}[X_{A,B,C}]+t'\big] &\leq \exp\left(-\frac{(t')^2}{2(\mathbb{E}[X_{A,B,C}]+t'/3)}\right) \leq \exp\left(-\frac{\varepsilon t'}{3}\right)\\ &\leq \exp\left(-\frac{\varepsilon^2}{3}p|A|\binom{|B|}{r-2}|C|\right) \leq \exp\left(-\frac{\varepsilon^2}{r^r}p|A||B|^{r-2}|C|\right).\end{aligned}$$

By the bounds on the sizes of A, B, and C and by choice of p we have $p|A||B|^{r-2}|C| = \omega\big((|A|+|B|+|C|)\log n\big)$ and $p|A||B|^{r-1} = \omega\big((|A|+|B|)\log n\big)$. For n large enough, applying the union bound therefore leads to

$$\begin{aligned}&\mathbb{P}\big[\text{Property (H1) fails}\big]\\ &\leq \sum_{a,b,c \geq \frac{\varepsilon n}{10^r \log^5 n}} n^a n^b \left(\exp\left(-\frac{\varepsilon^2}{3}pa\binom{b}{r-1}\right) + n^c \exp\left(-\frac{\varepsilon^2}{r^r}pab^{r-2}c\right)\right)\\ &\leq \sum_{a,b,c \geq \frac{\varepsilon n}{10^r \log^5 n}} \exp\left((a+b)\log n - \frac{\varepsilon^2}{3r^r}pab^{r-1}\right) + \exp\left((a+b+c)\log n - \frac{\varepsilon^2}{r^r}pab^{r-2}c\right)\\ &\leq \sum_{a,b,c \geq \frac{\varepsilon n}{10^r \log^5 n}} \exp(-(a+b)\log n) + \exp(-(a+b+c)\log n) = o(1).\end{aligned}$$

Next we show that Property (H2) holds with probability $\big(1-o(1)\big)$. Let A and B be disjoint subsets of $V(\mathcal{H})$ with $|A|,|B| \le \frac{n}{\log^5 n}$ and $|B| \in \{|A|, 2|A|\}$, and let $R \subseteq V(\mathcal{H})$. We distinguish two cases.

First case: $|A|,|B| \le \frac{2n}{\log^{10} n}$.

Let $X_V := e_{\mathcal{H}}(A,B,V(\mathcal{H})^{(r-2)})$ and observe that we have $e_{\mathcal{H}}(A,B,R^{(r-2)}) \le X_V$ and $X_V \sim \mathrm{Bin}\left(e_{K_n^{(r)}}(A,B,V(\mathcal{H})^{(r-2)}),p\right)$. For n large enough, it holds that

$$\mathbb{E}[X_V] \le |A||B|\binom{n}{r-2}p \le \frac{\varepsilon|A|n^{r-1}p}{3\log^5 n}.$$

We set $t = \varepsilon|A|n^{r-1}p/\log^5 n$ and observe that $\mathbb{E}[X]+t/3 \le t$. By applying Theorem 2.17 we thus obtain

$$\mathbb{P}\big[X_V \ge \mathbb{E}[X_V]+t\big] \le \exp\left(-\frac{\varepsilon|A|n^{r-1}p}{2\log^5 n}\right) \le \exp\left(-\frac{\varepsilon}{2}|A|\log^{8r-5} n\right).$$

Using the union bound we obtain for n large enough that the probability of Property (H2) failing for any disjoint sets $A,B \subseteq V(\mathcal{H})$ of size at most $2n/\log^{10} n$ and with $|B| \in \{|A|,2|A|\}$ is at most

$$\sum_{a=1}^{2n/\log^{10} n} \binom{n}{a}\left(\binom{n}{a}+\binom{n}{2a}\right)\exp\left(-\frac{\varepsilon}{2}a\log^{8r-5} n\right)$$
$$\le \sum_{a=1}^{2n/\log^{10} n} 2\exp\left(a\left(3\log n - \frac{\varepsilon}{2}\log^{8r-5} n\right)\right) = o(1).$$

Second case: $\frac{n}{\log^{10} n} \le |A|,|B| \le \frac{n}{\log^5 n}$.

We now define $X_R := e_{\mathcal{H}}(A,B,R^{(r-2)})$. It holds that $X_R \sim \mathrm{Bin}\left(e_{K_n^{(r)}}(A,B,R^{(r-2)}),p\right)$ and hence

$$\mathbb{E}[X_R] \le |A||B|\binom{|R|}{r-2}p \le \frac{|A|n^{r-1}}{\log^5 n}p.$$

By Theorem 2.17 we have

$$\mathbb{P}\left[X_R \ge \mathbb{E}[X_R] + \frac{\varepsilon|A|n^{r-1}p}{\log^5 n}\right] \le \exp\left(-\left(\frac{\varepsilon|A|n^{r-1}p}{\log^{bb} n}\right)^2 \cdot \min\left\{\frac{1}{3\mathbb{E}[X_R]}, \frac{\log^5 n}{3\varepsilon|A|n^{r-1}p}\right\}\right)$$
$$\le \exp\left(-\frac{\varepsilon^2|A|n^{r-1}p}{3\log^5 n}\right) \le \exp\left(-\frac{\varepsilon^2}{3}n\log^{8r-15} n\right).$$

For n large enough, applying the union bound yields that the probability that Property (H2) fails for any disjoint sets $A,B \subseteq V(\mathcal{H})$ of size at least $n/\log^{10} n$ and any set $R \subseteq V(\mathcal{H})$ is at most

$$2^{3n} \cdot \exp\left(-\frac{\varepsilon^2}{3}n\log^{8r-15} n\right) = o(1).$$

Hence Property (H2) holds with probability $\big(1-o(1)\big)$. $\square$

We note that any spanning subhypergraph of a (p, ε)-pseudorandom hypergraph is by definition also (p, ε)-pseudorandom.

4.1.2 Outline of the proof of the main theorem

As already indicated we will prove the following theorem for (p, ε')-pseudorandom r-uniform hypergraphs, which implies immediately Theorem 4.1 using Lemma 4.6.

Theorem 4.7. *Let $r \geq 3$, and let $\gamma > 0$ and $\varepsilon' > 0$ be reals such that $\varepsilon' \leq \left(10^{-3r}\gamma^2\right)^r$. Furthermore, let $p \geq \frac{\log^{8r} n}{n^{r-1}}$ and let $\mathcal{H}$ be a (p, ε')-pseudorandom r-uniform hypergraph on n vertices. Then every spanning subhypergraph $H \subseteq \mathcal{H}$ with $\delta_1(H) \geq \left(\frac{1}{2^{r-1}} + \gamma\right) p\binom{n}{r-1}$ contains a Hamilton Berge cycle.*

We first explain the idea of the proof of Theorem 4.7 in the case of weak Hamilton cycles and sketch afterwards how we guarantee the cycle to be Berge.

Let $\mathcal{H}$ be a (p, ε')-pseudorandom r-uniform hypergraph on n vertices. Given a spanning subhypergraph $H \subseteq \mathcal{H}$ as in the statement of the theorem, using Lemma 4.8 (which we state and prove in Subsection 4.1.3) we partition the vertex set of H into disjoint sets Y, Z, and W, where Y and W are both of linear size and W contains almost all vertices of H. The set Z assumes the role of a reservoir and is of size $n/\log^{\mathcal{O}(1)} n$. Lemma 4.8 provides also a refined partition and guarantees lower bounds for every vertex v on the number of hyperedges incident to v across and into the sets of the refined partition. Also, since H is (p, ε')-pseudorandom, we have good control on the maximum number of hyperedges among various subsets of vertices.

Next, using Lemma 4.19 (which we state and prove in Subsection 4.1.5) we construct a weak path Q with $V^*(Q) \subseteq Y$ such that for every subset $M \subseteq Z$ there exists a weak path Q_M that has the same endpoints as Q and such that $V^*(Q) \cup M$ is a partition of $V^*(Q_M)$. This property will be crucial at a later stage of the proof.

Then we distribute a maximum number of vertices in $Y \setminus V^*(Q)$ among the clusters of W in the refined partition such that all of these clusters have the same size $n/\log^{\mathcal{O}(1)} n$. Informally speaking, since $|Y|$ is significantly smaller than $|W|$ and every vertex from Y is 'well-connected' to W, such partition allows us to find weak paths $P_1, \ldots, P_m$, with $m = n/\log^{\mathcal{O}(1)} n$ such that $V^*(P_1), \ldots, V^*(P_m)$ form a partition of $W \cup Y \setminus V^*(Q)$.

As a last step, we use vertices from Z to connect the paths $P_1, \ldots, P_m$ and Q into a weak cycle C (again this is possible since every vertex of H is 'well-connected' into Z). Since the unused vertices M of Z can be absorbed by the path Q into a weak path Q_M with $V^*(Q_M) = V^*(Q) \cup M$, we have found a weak Hamilton cycle in H in this way. To construct the path Q and to connect the paths $P_1, \ldots, P_m$ and Q into a cycle we will repeatedly use Lemma 4.9 (which we state and prove in Subsection 4.1.4). This will allow us to connect various vertices by paths of length $\mathcal{O}(\log n)$.

Now we describe the changes that are necessary to ensure that the cycle we actually construct is Berge. Recall that in this case, each hyperedge is allowed to appear at most once in the cycle. Hence one needs to be careful every time a path is built, be it when the paths P_i are constructed simultaneously, when we apply Lemma 4.9 to connect the paths P_i and Q, and especially when we construct the paths Q and Q_M for $M \subseteq Z$. We resolve this problem by a careful analysis whenever we build paths simultaneously and by defining for each of these main steps a different bucket set $R \subseteq V(H)$ such that the hyperedges of the paths constructed in that step only use vertices from the clusters, where the inner vertices of their sequences lie, plus vertices from R. By a careful choice of the bucket sets, it is then ensured that the Hamilton cycle that we get in the end is indeed Berge.

Before proving Theorem 4.7 in Subsection 4.1.6, we state and prove the main lemmas that we need in Subsections 4.1.3–4.1.5.

4.1.3 Partition lemma

This subsection is devoted to the following lemma, which we use at the beginning of the proof of the main theorem to partition the vertex set of the given subhypergraph.

Lemma 4.8 (Partition lemma). *Let $r \geq 3$ be an integer, and let $\gamma \in (0,1)$, $\varepsilon \in (0,1)$, and $p \geq \log^{8r} n/n^{r-1}$ be reals. Let H be an r-uniform hypergraph with the following properties. For every $v \in V(H)$ it holds that*

(a) $\deg_H(v) \geq \left(1/2^{r-1} + \gamma\right) p\binom{n}{r-1}$ *and*

(b) $\operatorname{link}_H(v)$ *contains at most* $n^{2r-3+\varepsilon}p^2 + \log^2 n$ *Berge paths of length 2.*

Then there exists a partition

$$V(H) = Y \cup Z \cup W$$

and a refined partition $\mathcal{P} = \{Y_i, Z_i, W_j\}_{i\in[\ell], j\in[t]}$ *with*

$$Y = \bigcup_{i\in[\ell]} Y_i\,, \qquad Z = \bigcup_{i\in[\ell]} Z_i\,, \qquad W = \bigcup_{j\in[t]} W_j\,,$$

where $t := \log^5 n$ and $\ell := 16\log n$ such that the following holds for every $i \in [\ell]$ and $j \in [t]$:

(P1) The sizes of the sets in the refined partition satisfy

$$\frac{\varepsilon n}{2\ell} \leq |Y_i| \leq \frac{\varepsilon n}{\ell},$$
$$\frac{n}{2\ell \log^3 n} \leq |Z_i| \leq \frac{n}{\ell \log^3 n},$$
$$\left(1 - \frac{4\varepsilon}{5}\right)\frac{n}{\log^5 n} \leq |W_j| \leq \left(1 - \frac{3\varepsilon}{5}\right)\frac{n}{\log^5 n},$$

(P2) for every set $A \in \mathcal{P}$ and every $v \in V(H) \setminus A$ it holds that

$$e_H(v, A^{(r-1)}) \geq \left(\frac{1}{2^{r-1}} + \frac{\gamma}{2}\right) p\binom{|A|}{r-1},$$

(P3) for all disjoint sets $A \in \mathcal{P}$ and

$$\varnothing \neq R \in \left\{\bigcup\nolimits_{A'\in\mathcal{F}} A' : \mathcal{F} \subseteq \{Y_i\}_{i\in[\ell]} \text{ or } \mathcal{F} \subseteq \{Z_i\}_{i\in[\ell]} \text{ or } \mathcal{F} \subseteq \{W_j\}_{j\in[t]}\right\}$$

and for every $v \in V(H) \setminus (A \cup R)$ it holds that

$$e_H(v, A, R^{(r-2)}) \geq \left(\frac{1}{2^{r-1}} + \frac{\gamma}{2}\right) p|A|\binom{|R|}{r-2}.$$

Proof of Lemma 4.8. Given $r \geq 3$, $\gamma \in (0,1)$, $\varepsilon \in (0,1)$ and $p \geq \log^{8r} n/n^{r-1}$, choose $\varepsilon' \in (0, 10^{-r}\varepsilon\gamma)$ such that

$$(1+\varepsilon')^r < \frac{1+2^{r-2}\gamma}{1+2^{r-3}\gamma}.$$

Let H be an r-uniform hypergraph as in the statement of the lemma.

Set $Y_i = Z_i = \varnothing$ and $W_j = \varnothing$ for every $i \in [\ell]$ and $j \in [t]$. Now, for every $v \in V(H)$ we add v to Y_i with probability

$$p_i = \big(3\varepsilon/(4\ell) - 3/(4\ell \log^3 n)\big)$$

for every $i \in [\ell]$, to Z_i with probability

$$p_i' = 3/(4\ell \log^3 n)$$

for every $i \in [\ell]$ and to W_j with probability

$$p_j'' = (1 - 3\varepsilon/4)/\log^5 n$$

for every $j \in [t]$. Observe that we have

$$\sum_{i\in[\ell]} (p_i + p_i') + \sum_{j\in[t]} p_j'' = \sum_{i\in[\ell]} \left(\frac{3\varepsilon}{4\ell} - \frac{3}{4\ell\log^3 n} + \frac{3}{4\ell \log^3 n}\right) + \sum_{j\in[t]} \left(1 - \frac{3\varepsilon}{4}\right) \frac{1}{\log^5 n} = 1.$$

Our aim is to show that with positive probability such a random partition

$$V(H) = \bigcup_{i\in[\ell]} Y_i \cup \bigcup_{i\in[\ell]} Z_i \cup \bigcup_{j\in[t]} W_j$$

satisfies Properties (P1)–(P3).

We first prove that each of the following properties holds with probability $\big(1-o(1)\big)$.

(P1') For every $i \in [\ell]$ and $j \in [t]$ it holds that $|Y_i| = (1\pm\varepsilon')np_i$ as well as $|Z_i| = (1\pm\varepsilon')np_i'$ and $|W_j| = (1\pm\varepsilon')np_j''$,

(P2') for every $A \in \mathcal{P}$ we have $e_H(v, A^{(r-1)}) \geq \left(\frac{1}{2^{r-1}} + \frac{3\gamma}{4}\right) p\binom{n}{r-1}(\tilde{p})^{r-1}$ for every $v \in V(H) \setminus A$, where $\tilde{p} = p_i$ if $A = Y_i$ with $i \in [\ell]$, and $\tilde{p} = p_i'$ if $A = Z_i$ with $i \in [\ell]$, and $\tilde{p} = p_j''$ if $A = W_j$ with $j \in [t]$,

(P3') for every vertex $v \in V(G)$ and every nonempty set $\mathcal{F}$ with $\mathcal{F} \subseteq \{Y_i\}_{i\in[\ell]}$ or $\mathcal{F} \subseteq \{Z_i\}_{i\in[\ell]}$ or $\mathcal{F} \subseteq \{W_j\}_{j\in[t]}$ and for every set $A \in \mathcal{P}$ that is disjoint from $R := \bigcup_{A'\in\mathcal{F}} A'$ we have

$$e_H(v, A, R^{(r-2)}) \geq \left(\frac{1}{2^{r-1}} + \frac{3\gamma}{4}\right) p\binom{n}{r-1}(r-1)\tilde{p}(sp^*)^{r-2},$$

where $s = |\mathcal{F}|$, and $\tilde{p} = p_i$ if $A = Y_i$ with $i \in [\ell]$, and $\tilde{p} = p_i'$ if $A = Z_i$ with $i \in [\ell]$, and $\tilde{p} = p_j''$ if $A = W_j$ with $j \in [t]$, and $p^* = p_i$ if $\mathcal{F} \subseteq \{Y_i\}_{i\in[\ell]}$, and $p^* = p_i'$ if $\mathcal{F} \subseteq \{Z_i\}_{i\in[\ell]}$, and $p^* = p_j''$ if $\mathcal{F} \subseteq \{W_j\}_{j\in[t]}$.

We start with the proof of Property **(P1')**. Let $i \in [\ell]$, $j \in [t]$ and set $X_1 = |Y_i|$, $X_2 = |Z_i|$, and $X_3 = |W_j|$. We have $X_1 \sim \mathrm{Bin}(n, p_i)$, $X_2 \sim \mathrm{Bin}(n, p'_i)$, and $X_3 \sim \mathrm{Bin}(n, p''_j)$. Using Chernoff's inequality (Theorem 2.16) we obtain for every $k \in [3]$ that

$$\mathbb{P}\big[|X_k - \mathbb{E}[X_k]| > \varepsilon'\mathbb{E}[X_k]\big] < 2\exp\left(-\frac{(\varepsilon')^2\mathbb{E}[X_k]}{3}\right).$$

By applying the union bound, we get that with probability $(1 - o(1))$ it holds that $|Y_i| = (1 \pm \varepsilon')np_i$ as well as $|Z_i| = (1 \pm \varepsilon')np'_i$ for every $i \in [\ell]$ and $|W_j| = (1 \pm \varepsilon')np''_j$ for every $j \in [k]$.

Let us now turn to Property **(P2')**. Let $A \in \mathcal{P}$ and $v \in V(H) \setminus A$. Set

$$\tilde{p} = \begin{cases} p_i & \text{if } A = Y_i \text{ with } i \in [\ell], \\ p'_i & \text{if } A = Z_i \text{ with } i \in [\ell], \\ p''_j & \text{if } A = W_j \text{ with } j \in [t]. \end{cases}$$

To describe the expected value of $X := e_H(v, A^{(r-1)})$ we set

$$X_M = \begin{cases} 1 & \text{if } M \subseteq A \ , \\ 0 & \text{otherwise.} \end{cases}$$

Therefore we obtain

$$X = \sum_{\substack{M \subseteq V(H)\setminus\{v\}: \\ \{v\}\cup M \in E(H)}} X_M,$$

which leads to

$$\mathbb{E}[X] = \sum_{\substack{M \subseteq V(H)\setminus\{v\}: \\ \{v\}\cup M \in E(H)}} \mathbb{P}[X_M = 1] = \deg_H(v)(\tilde{p})^{r-1}.$$

Hence, if $c_1 := (1/2^{r-1} + \gamma) \cdot r^{-r}$ and $c_2 := (1 - 3\varepsilon/4)^{r-1}c_1$, we have

$$\mathbb{E}[X] \geq \left(\frac{1}{2^{r-1}} + \gamma\right) p \binom{n}{r-1} (\tilde{p})^{r-1} \geq c_1 n^{r-1} p (\tilde{p})^{r-1} \geq c_2 \log^{3r+5} n, \tag{4.1}$$

where we used that $p \geq \log^{8r} n / n^{r-1}$ and $\tilde{p} \geq (1 - 3\varepsilon/4)/\log^5 n$. We aim to apply Janson's inequality (Theorem 2.20). With the notation from that theorem we obtain

$$\overline{\Delta} = \mathbb{E}[X] + \sum_{\substack{M \subseteq V(H)\setminus\{v\}: \\ \{v\}\cup M \in E(H)}} \sum_{\substack{M' \subseteq V(H)\setminus\{v\}: \\ \{v\}\cup M' \in E(H) \\ M \neq M',\ M \cap M' \neq \varnothing}} \mathbb{E}[X_M X_{M'}] \leq \mathbb{E}[X] + 2\left(n^{2r-3+\varepsilon}p^2 + \log^2 n\right)(\tilde{p})^r,$$

where we used that $|M \cup M'| \geq r$ and the upper bound on the number of Berge paths of length 2 in $\mathrm{link}_{\mathcal{H}}(v)$, which is given by Property (b). Thus, by applying Theorem 2.20 with

$t' := \frac{\gamma}{8}\mathbb{E}[X]$ we conclude

$$\begin{aligned}\mathbb{P}\big[X \leq \mathbb{E}[X] - t'\big] &\leq \exp\left(-\frac{(t')^2}{2\overline{\Delta}}\right)\\ &\leq \exp\left(-\frac{\gamma^2}{1000}\min\left\{\mathbb{E}[X], \frac{\mathbb{E}[X]^2}{n^{2r-3+\varepsilon}p^2(\tilde{p})^r}, \frac{\mathbb{E}[X]^2}{(\tilde{p})^r\log^2 n}\right\}\right)\\ &\leq \exp\left(-\frac{\gamma^2}{1000}\min\left\{c_2\log^{3r+5} n, c_1^2 n^{1-\varepsilon}(\tilde{p})^{r-2}, c_1^2\log^{11r} n\right\}\right)\\ &\leq \exp(-2\log n),\end{aligned}$$

as $p \geq (\log^{8r} n)/n^{r-1}$ and $\tilde{p} \geq (1-3\varepsilon/4)/\log^5 n$ and by Equation (4.1).

Applying the union bound yields that

$$\mathbb{P}\big[\text{Property (P2') fails}\big] \leq \sum_{A\in\mathcal{P}}\sum_{v\in V(H)} \exp(-2\log n) \leq 3(\log^5 n)n\exp(-2\log n) = o(1).$$

We now turn to Property **(P3')**. Let $A \in \mathcal{P}$ and

$$\varnothing \neq R \in \left\{\bigcup\nolimits_{A'\in\mathcal{F}} A' : \mathcal{F} \subseteq \{Y_i\}_{i\in[\ell]} \text{ or } \mathcal{F} \subseteq \{Z_i\}_{i\in[\ell]} \text{ or } \mathcal{F} \subseteq \{W_j\}_{j\in[t]}\right\}$$

be disjoint sets and let $v \in V(H) \setminus (A \cup R)$. As before, we set

$$\tilde{p} = \begin{cases} p_i & \text{if } A = Y_i \text{ with } i \in [\ell],\\ p_i' & \text{if } A = Z_i \text{ with } i \in [\ell],\\ p_j'' & \text{if } A = W_j \text{ with } j \in [t].\end{cases}$$

By choice of R there exists a subset $\mathcal{F} \subseteq \mathcal{P}$ such that $R = \bigcup_{A'\in\mathcal{F}} A'$. We define $s = |\mathcal{F}|$ and set

$$p^* = \begin{cases} p_i & \text{if } Y_i \in \mathcal{F} \text{ for an index } i \in [\ell],\\ p_i' & \text{if } Z_i \in \mathcal{F} \text{ for an index } i \in [\ell],\\ p_j'' & \text{if } W_j \in \mathcal{F} \text{ for an index } j \in [t].\end{cases}$$

Since $p_i = p_{i'}$ and $p_i' = p_{i'}'$ for all $i, i' \in [\ell]$ and $p_j'' = p_{j'}''$ for all $j, j' \in [\ell]$, and by definition of R, the value p^* is well defined. We consider $X := e_H(v, A, R^{(r-2)})$. To describe its expected value, we define the indicator variable

$$X_M := \begin{cases} 1 & \text{if } |M \cap A| = 1 \text{ and } |M \cap R| = r-2\ ,\\ 0 & \text{otherwise.}\end{cases}$$

So, we obtain

$$X = \sum_{\substack{M\subseteq V(H)\setminus\{v\}:\\ M\cup\{v\}\in E(H)}} X_M,$$

which leads to

$$\mathbb{E}[X] = \sum_{\substack{M\subseteq V(H)\setminus\{v\}:\\ M\cup\{v\}\in E(H)}} \mathbb{P}[X_M = 1] = \deg_H(v)(r-1)\tilde{p}(sp^*)^{r-2}.$$

With $c_1 = (1/2^{r-1} + \gamma) \cdot r^{-r}$ we have

$$\mathbb{E}[X] \geq \left(\frac{1}{2^{r-1}} + \gamma\right) p \binom{n}{r-1} (r-1)\tilde{p}(sp^*)^{r-2} \geq c_1 p n^{r-1}(r-1)\tilde{p}(sp^*)^{r-2}. \tag{4.2}$$

Again, with the notation from Theorem 2.20 we obtain

$$\begin{aligned}\overline{\Delta} &= \mathbb{E}[X] + \sum_{\substack{M \subseteq V(H)\setminus\{v\}:\\ \{v\}\cup M \in E(H)}} \sum_{\substack{M' \subseteq V(H)\setminus\{v\}:\\ \{v\}\cup M' \in E(H)\\ M \neq M',\ M\cap M' \neq \varnothing}} \mathbb{E}[X_M X_{M'}] \\ &\leq \mathbb{E}[X] + 2(n^{2r-3+\varepsilon}p^2 + \log^2 n)(r-1)\tilde{p}(sp^*)^{r-2}(\tilde{p} + sp^*),\end{aligned}$$

where we used that $|M \cup M'| \geq r$ and the upper bound on the number of Berge paths of lengh 2 in $\mathrm{link}_{\mathcal{H}}(v)$.

Thus, by applying Janson's inequality (Theorem 2.20) with $t' := \frac{\gamma}{8}\mathbb{E}[X]$ we conclude

$$\begin{aligned}&\mathbb{P}\big[X \leq \mathbb{E}[X] - t'\big] \leq \exp\left(-\frac{(t')^2}{2\overline{\Delta}}\right) \\ &\leq \exp\left(-\frac{\gamma^2}{2000}\min\left\{\mathbb{E}[X], \frac{c_1^2(r-1)\tilde{p}(sp^*)^{r-2}n^{1-\varepsilon}}{\tilde{p}+sp^*}, \frac{c_1^2(r-1)p^2n^{2r-2}\tilde{p}(sp^*)^{r-2}}{(\tilde{p}+sp^*)\log^2 n}\right\}\right) \\ &\leq \exp(-2\log^5 n),\end{aligned}$$

where we used Equation (4.2) and $p \geq (\log^{8r} n)/n^{r-1}$ and $\tilde{p}, p^* \geq (1 - 3\varepsilon/4)/\log^5 n$.

Applying the union bound yields

$$\begin{aligned}\mathbb{P}\big[\text{Property (P3') fails}\big] &\leq \sum_{A\in\mathcal{P}} \sum_{R} \sum_{\substack{v \in V(H):\\ v\notin (A\cup R)}} \exp(-2\log^5 n) \\ &\leq 9(\log^5 n)2^{\log^5 n} n \exp(-2\log^5 n) = o(1).\end{aligned}$$

Finally we need to show that Properties (P1)–(P3) hold with positive probability. Property (P1) holds a.a.s. since Property (P1') is a.a.s. satisfied. It can be seen fairly quickly that Property (P2) holds a.a.s. by conditioning on Property (P1') and using Property (P2'), the definition of $\tilde{p}$, and the choice of ε'. Similarly, Property (P3) holds a.a.s., which can be seen by conditioning on Property (P1') and using Property (P3'), the definition of $\tilde{p}$, p^*, s, and the choice of ε'.

□

4.1.4 Connecting lemma

In this subsection we prove the following lemma that is essential at various places in the proof of Theorem 4.7.

Lemma 4.9 (Connecting Lemma). *Let $r \geq 3$ be an integer and let $p \geq \log^{8r} n/n^{r-1}$. Furthermore, let $\gamma \in (0,1)$ as well as $\varepsilon \geq \varepsilon' > 0$ be reals such that $\varepsilon \leq 10^{-3}\gamma/2^{r+1}$ and $\varepsilon' \leq (\gamma\varepsilon/(10r))^r$. Let H be an n-vertex (p,ε')-pseudorandom r-uniform hypergraph given with a partition*

$$V(H) = \bigcup_{i\in[\ell]} Y_i \cup \bigcup_{i\in[\ell]} Z_i \cup \bigcup_{j\in[t]} W_j$$

with $t := \log^5 n$ and $\ell := 16 \log n$ that satisfies Properties (P1)–(P3) from Lemma 4.8 with input r, γ, ε, and p.

Let $\mathcal{A} = (A_1, \ldots, A_{8\log n})$ be a sequence of pairwise disjoint subsets of $\{Y_i, Z_i\}_{i\in[\ell]}$. Moreover, let $s \leq \min_{i\in[\ell]} \varepsilon|A_i|/4$ be a positive integer and let $\{(a_i, b_i)\}_{i\in[s]}$ be pairs of vertices in $V(H) \setminus V(\mathcal{A})$ such that $a_i \neq a_j$ and $b_i \neq b_j$ for every distinct $i, j \in [s]$. Furthermore, let

$$\varnothing \neq R \in \left\{ \bigcup\nolimits_{A'\in\mathcal{F}} A' : \mathcal{F} \subseteq \{Y_i\}_{i\in[\ell]} \text{ or } \mathcal{F} \subseteq \{W_j\}_{j\in[t]} \right\}$$

such that $R \cap \big(V(\mathcal{A}) \cup \{a_i, b_i\}_{i\in[s]}\big) = \varnothing$ and $|R| \geq \varepsilon n/10$.

Then there exist $P_1, \ldots, P_s$, each being a Berge path or a Berge cycle, such that for every $i \in [s]$ we have

- *(C1) P_i is a Berge path connecting a_i to b_i in case $a_i \neq b_i$, or a Berge cycle containing a_i in case $a_i = b_i$,*
- *(C2) $V^*(P_i) \subseteq V(\mathcal{A}) \cup \{a_i, b_i\}$,*
- *(C3) $E(P_i) \subseteq E_H\big(V(\mathcal{A}) \cup \{a_i, b_i\}, V(\mathcal{A})^{(r-1)}\big) \cup E_H\big(V(\mathcal{A}) \cup \{a_i, b_i\}, V(\mathcal{A}), R^{(r-2)}\big)$,*
- *(C4) $V^*(P_i) \cap V^*(P_{i'}) = \{a_i, b_i\} \cap \{a_{i'}, b_{i'}\}$ for every $i' \in [s] \setminus \{i\}$,*
- *(C5) $E(P_i) \cap E(P_{i'}) = \varnothing$ for every $i' \in [s] \setminus \{i\}$,*
- *(C6) $|V^*(P_i) \setminus \{a_i, b_i\}| \equiv 2 \mod 4$, and*
- *(C7) $|V^*(P_i)| \leq 8 \log n$.*

For the proof of Lemma 4.9 we need to introduce some more definitions. We start with two notions of *compatible* weak paths, which will allow us to have good control of where the vertices and hyperedges of a weak path with respect to a vertex partition lie. The first definition is that of an $\mathcal{A}$-*compatible* weak path. Informally speaking, given a sequence $\mathcal{A}$ of pairwise disjoint sets, a weak path is called $\mathcal{A}$-compatible if all vertices that are contained in any of its hyperedges except for its endpoints lie in the sets of $\mathcal{A}$ arranged in a specific manner.

Definition 4.10 ($\mathcal{A}$-compatible)**.** *Let H be an r-uniform hypergraph with $r \geq 3$, let $m \geq 1$ and let $\mathcal{A} = (A_1, \ldots, A_m)$ be a sequence of pairwise disjoint subsets of $V(H)$. We say that a weak path $P = (v_0, e_0, v_1, \ldots, v_m, e_m, v_{m+1})$ is* $\mathcal{A}$-compatible *if*

- *(1) $v_0, v_{m+1} \notin V(\mathcal{A})$,*
- *(2) $v_i \in A_i$ for every $i \in [m]$,*
- *(3) $e_0 \in E_{\mathcal{H}}\big(v_0, v_1, A_1^{(r-2)}\big)$,*
- *(4) $e_m \in E_{\mathcal{H}}\big(v_m, v_{m+1}, A_m^{(r-2)}\big)$, and*
- *(5) $e_i \in E_{\mathcal{H}}\big(v_i, v_{i+1}, A_i^{(r-2)}\big) \cup E_{\mathcal{H}}\big(v_i, v_{i+1}, A_{i+1}^{(r-2)}\big)$ for every $i \in [m-1]$.*

In the second definition of compatible paths, we have apart from a sequence $\mathcal{A}$ a bucket set R. Roughly speaking, a weak path P is called $(\mathcal{A}, R)$-*compatible* if the vertices of $V^*(P) \setminus \mathrm{End}(P)$ lie in the sets of $\mathcal{A}$ arranged in a specific manner and all other vertices of the hyperedges of P except for its endvertices are contained in R.

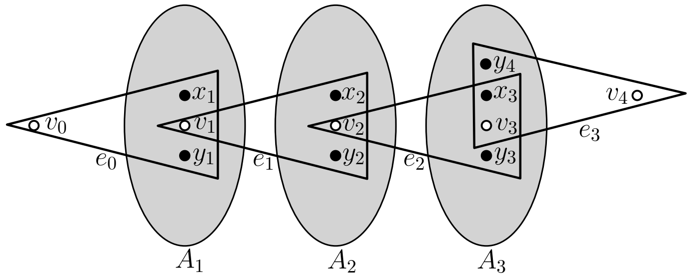

Figure 4.1: An (A_1, A_2, A_3)-compatible weak path $(v_0, e_0, v_1, e_1, v_2, e_2, v_3, e_3, v_4)$ in a 4-uniform hypergraph with $e_i = \{v_i, v_{i+1}, x_{i+1}, y_{i+1}\}$ for $i \in \{0, 1, 3\}$ and $e_2 = \{v_2, v_3, x_2, y_2\}$.

Definition 4.11 (($\mathcal{A}, R$)-compatible)**.** *Let H be an r-uniform hypergraph with $r \geq 3$ and let $m \geq 0$ be an integer. If $m = 0$, let $\mathcal{A}$ be the empty sequence and otherwise let $\mathcal{A} = (A_1, \ldots, A_m)$ be a sequence of pairwise disjoint subsets of $V(H)$. Moreover, let $R \subseteq V(H)$ be disjoint from $V(\mathcal{A})$. We say that a weak path $P = (v_0, e_0, v_1, \ldots, v_m, e_m, v_{m+1})$ is* $(\mathcal{A}, R)$-compatible *if*

(1) $v_0, v_{m+1} \notin R \cup V(\mathcal{A})$,

(2) $v_i \in A_i$ *for every* $i \in [m]$, *and*

(3) $e_i \in E_{\mathcal{H}}\big(v_i, v_{i+1}, R^{(r-2)}\big)$ *for every* $i \in \{0, \ldots, m\}$.

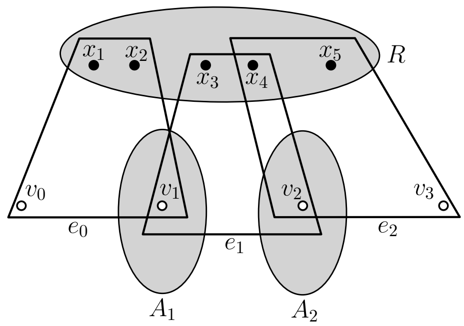

Figure 4.2: An $((A_1, A_2), R)$-compatible weak path $(v_0, e_0, v_1, e_1, v_2, e_2, v_3)$ in a 4-uniform hypergraph with $e_0 = \{v_0, v_1, x_1, x_2\}$, $e_1 = \{v_1, v_2, x_3, x_4\}$ and $\{v_2, v_3, x_4, x_5\}$.

Observe that for a weak path P that is either $\mathcal{A}$-compatible or $(\mathcal{A}, R)$-compatible for any sequence $\mathcal{A}$ of pairwise disjoint subsets of $V(H)$ and for any subset $R \subseteq V(H)$ with $R \cap V(\mathcal{A}) = \varnothing$, it holds that P is a Berge path since all its hyperedges have to be distinct by Definitions 4.10 and 4.11.

The following lemma will be useful in the proof of the connecting lemma (Lemma 4.9) when one needs to argue that certain $\mathcal{A}$-compatible and $(\mathcal{A}', R')$-compatible Berge paths are edge-disjoint.

Lemma 4.12. *Let $r \geq 3$, let H be an r-uniform hypergraph and let $k, m, m' \geq 1$. Let $\mathcal{D} = (D_1, \ldots, D_k)$ be a sequence of pairwise disjoint subsets of $V(H)$ and let $\mathcal{A} = (A_1, \ldots, A_m)$*

and $\mathcal{A}' = (A'_1 \ldots, A'_{m'})$ *be (not necessarily distinct) sequences such that for every* $j \in [m]$ *and* $j' \in [m']$ *there exist* $i, i' \in [k]$ *with* $A_j \subseteq D_i$ *and* $A'_{j'} \subseteq D_{i'}$*. Moreover, let* $R, R' \subseteq V(H)$ *be (not necessarily distinct) subsets each being disjoint from both* $V(\mathcal{A})$ *and* $V(\mathcal{A}')$*. Let* P_1 *and* P_2 *be two Berge paths such that one of the following three properties holds:*

- P_1 *is* $\mathcal{A}$*-compatible and* P_2 *is* $(\mathcal{A}', R')$*-compatible,*
- P_1 *is* $\mathcal{A}$*-compatible,* P_2 *is* $\mathcal{A}'$*-compatible and* $V^*(P_1) \cap V^*(P_2) = \varnothing$*,*
- P_1 *is* $(\mathcal{A}, R)$*-compatible,* P_2 *is* $(\mathcal{A}', R')$*-compatible,* $P_1 \neq P_2$ *and* $V^*(P_1) \cap V^*(P_2) = \mathrm{End}(P_1) \cap \mathrm{End}(P_2)$*.*

Then $E(P_1) \cap E(P_2) = \varnothing$*.*

Proof. If P_1 is $\mathcal{A}$-compatible and P_2 is $(\mathcal{A}', R')$-compatible, then it follows directly from Definitions 4.10 and 4.11 that $E(P_1) \cap E(P_2) = \varnothing$, as $R' \cap V(\mathcal{A}) = \varnothing$.

Suppose now that P_1 is $\mathcal{A}$-compatible, P_2 is $\mathcal{A}'$-compatible, and $V^*(P_1) \cap V^*(P_2) = \varnothing$. Assume for a contradiction that there exists a hyperedge $e \in E(P_1) \cap E(P_2)$. We let the Berge path P_1 be $P_1 = (v_0, e_0, v_1, \ldots, v_m, e_m, v_{m+1})$. By Properties (3)–(5) of Definition 4.10 there exists an index $i \in [m+1]$ such that $e \in E_H(v_{i-1}, v_i, A_{i-1}^{(r-2)})$ (if $i \geq 2$) or $e \in E_H(v_{i-1}, v_i, A_i^{(r-2)})$ (if $i \leq m$). Without loss of generality we may assume that $e \in E_H(v_{i-1}, v_i, A_{i-1}^{(r-2)})$. Obviously, we have $v_i \in V^*(P_1)$. By the definition of $\mathcal{A}$ and $\mathcal{A}'$ and by Properties (2)–(5) of Definition 4.10 we also get $v_i \in V^*(P_2)$, a contradiction to $V^*(P_1) \cap V^*(P_2) = \varnothing$.

Finally, suppose that P_1 is $(\mathcal{A}, R)$-compatible, P_2 is $(\mathcal{A}', R')$-compatible, and $V^*(P_1) \cap V^*(P_2) = \mathrm{End}(P_1) \cap \mathrm{End}(P_2)$ but $P_1 \neq P_2$. Assume again for a contradiction that there exists a hyperedge $e \in E(P_1) \cap E(P_2)$. By Property (3) of Definition 4.11 there exist vertices $x, y \in V^*(P_1)$ such that $e \in E_H(x, y, R^{(r-2)})$. It follows from Properties (1) and (2) and by the definition of $\mathcal{A}$ and $\mathcal{A}'$ that $x, y \in V^*(P_2)$. But then $V^*(P_1) \cap V^*(P_2) = \mathrm{End}(P_1) \cap \mathrm{End}(P_2) = \{x, y\}$ and $P_1 = (x, e, y) = P_2$, a contradiction. □

We need two more definitions. The first is the definition of the *i-th neighbourhood* of a vertex set with respect to a given sequence of length at least i.

Definition 4.13 (i-th neighbourhood)**.** *Let* H *be an* r*-uniform hypergraph, let* $m \in \mathbb{N}$*, and let* $\mathcal{A} = (A_1, \ldots, A_m)$ *be a sequence of pairwise disjoint subsets of* $V(H)$*. For every subset* $X \subseteq V(H)$ *that is disjoint from* $V(\mathcal{A})$ *and for every* $i \in [m]$ *we write*

$$\begin{aligned} N_{\mathcal{A}}^i(X) := \Big\{ v_i \in A_i : & \text{ there exist } v_0 \in X \text{ and a Berge path } (v_0, e_0, v_1, e_1, v_2, \ldots, v_{i-1}, e_{i-1}, v_i) \\ & \text{such that } v_j \in A_j \text{ for every } j \in [i-1] \text{ and} \\ & e_j \in E_H\big(v_j, v_{j+1}, A_{j+1}^{(r-2)}\big) \text{ for every } j \in \{0, \ldots, i-1\}. \Big\} \end{aligned}$$

for the i-th neighbourhood *of* X *with respect to* $\mathcal{A}$*.*

For the sake of simplicity, we write $N_{\mathcal{A}}^i(x) = N_{\mathcal{A}}^i(\{x\})$. In Figure 4.1 we have $v_i \in N^i_{(A_1,A_2,A_3)}(v_0)$ and $v_{4-i} \in N^i_{(A_3,A_2,A_1)}(v_4)$ for every $i \in [2]$ but it does not necessarily hold that $v_3 \in N^3_{(A_1,A_2,A_3)}(v_0)$ or $v_2 \in N^3_{(A_3,A_2,A_1)}(v_4)$.

Next we define $(2, R)$-*matchings* between two sets A and B with a bucket set R.

Definition 4.14 ($(2, R)$-matching)**.** *Let H be an r-uniform hypergraph and let $A, B, R \subseteq V(H)$ be pairwise disjoint subsets of. A* $(2,\ R)$-matching *between A and B that saturates A is a set $\{e_a^1, e_a^2\}_{a\in A}$ of distinct hyperedges such that for every $a, a' \in A$ and $i, j \in [2]$ we have*

(1) $e_a^1, e_a^2 \in E_H(a, B, R^{(r-2)})$ *and*

(2) $e_a^i \cap e_{a'}^j \cap B = \varnothing$.

Finally, we are in the position to prove Lemma 4.9. Before embarking on the proof, let us first explain the main idea of the proof. We give the proof for the cases that the vertices a_i and b_i are distinct for all $i \in [s]$. This means that our goal is to construct Berge paths (rather than cycles) to connect a_i to b_i. The cases when $a_i = b_i$ for at least one index $i \in [s]$ work analogously. We define $r(n) \leq 4 \log n$ to be the largest integer such that $r(n) \equiv 0 \mod 4$. The reason for this definition is to ensure that the Berge paths that we construct have the right length, i.e. satisfy Property (C6).

Informally speaking, we prove the lemma iteratively by connecting in the k-th step 2^{-k} of the given pairs of vertices by Berge paths and by identifying for each vertex x, which is not yet connected to its mate, 2^{k-1} vertices each of which can be reached via a Berge path from x using $k-1$ hyperedges. These duplicates are used in the $(k+1)$-th step to connect 2^{-k+1} not yet connected pairs of vertices.

To make this rough idea a bit clearer, we illustrate the iteration by explaining the first two steps more precisely. We start with connecting half of the given pairs of vertices by Berge paths such that the vertex sets of their sequences are pairwise disjoint and each of them is $(A_1, \dots, A_{r(n)+2})$-compatible. We store the indices of these pairs in the set I_1. Next, we find a $(2, R)$-matching between $\{a_i\}_{i\in[s]\setminus I_1}$ and $A_{4\log n+1}$ that saturates $\{a_i\}_{i\in[s]\setminus I_1}$ as well as a $(2, R)$-matching between $\{b_i\}_{i\in[s]\setminus I_1}$ and $A_{4\log n+2}$ that saturates $\{b_i\}_{i\in[s]\setminus I_1}$. We connect half of the pairs of the vertices in $A_{4\log n+1}$ and $A_{4\log n+2}$, respectively, that are touched by this matching, by Berge paths such that the vertex sets and the edge sets of their sequences are pairwise disjoint and disjoint from the already built ones and such that each of them is $(A_1, \dots, A_{r(n)})$-compatible. Hence, after this step, we can connect $3/4$ of the given pairs of vertices by Berge paths and we show that that Properties (C1)–(C7) hold for these paths.

Proof of Lemma 4.9. Let p, γ, ε, ε' be given. Furthermore, let H be a (p, ε')-pseudorandom r-uniform hypergraph on n vertices given with a partition of $V(H)$, let $\mathcal{A} = (A_1, \dots, A_{8\log n})$ be a sequence of disjoint subsets of $V(H)$, let $R \subseteq V(H)$ be a set and $\{(a_i, b_i)\}_{i\in[s]}$ pairs of vertices as in the statement of the lemma.

Let $r(n) \leq 4 \log n$ be the largest integer such that $r(n) \equiv 0 \mod 4$. Let $\ell' := 4 \log n$. For the sake of readability we assume in the proof that $a_i \neq b_i$ for every $i \in [s]$. The cases that $a_i = b_i$ for at least one index $i \in [s]$ work analogously.

We will prove the statement for every $s \geq n/\log^5 n$ for which there exists an integer N such that $s = 2^N$ and $s \leq \min_{i\in[8\log n]}\{\varepsilon|A_i|/2\}$. This gives us immediately a proof of the lemma. Indeed, let $s \leq \min_{i\in[8\log n]}\{\varepsilon|A_i|/4\}$ be given. If s is not a power of 2, let $s' \geq s$ be the smallest integer such that s' is a power of 2. Then $s' \leq 2s \leq \min_{i\in[8\log n]}\{\varepsilon|A_i|/2\}$. Hence, we can take arbitrary distinct vertices $\{a_i, b_i\}_{s+1\leq i\leq s'}$ from $V(H)\setminus\big(V(\mathcal{A})\cup\{a_i, b_i\}_{i\in[s]}\big)$ and apply the proof to $\{a_i, b_i\}_{i\in[s']}$, which gives us in particular the desired Berge paths for $\{a_i, b_i\}_{i\in[s]}$.

Let us now define the invariants that we maintain in every iteration step $k \in \{0, \dots, N+1\}$. We will explain directly afterwards what the rough meaning of them is.

(I1) $I_k \subseteq [s]$ is a subset of indices such that $|I_k| = \lceil s(1-2^{-k}) \rceil$.

(I2) $\mathcal{P}_k = \{P_i\}_{i \in I_k}$ are Berge paths such that for every distinct $i, i' \in I_k$ we have

(a) P_i connects a_i to b_i,
(b) $V^*(P_i) \subseteq \bigcup_{j \in [\ell'+2k-2]} A_j \cup \{a_i, b_i\}$,
(c) $E(P_i) \subseteq E_H\big(\bigcup_{j \in [\ell'+2k-2]} A_j \cup \{a_i, b_i\}, \big(\bigcup_{j \in [\ell'+2k-2]} A_j\big)^{(r-1)}\big)$
$\cup\, E_H\big(\bigcup_{j \in [\ell'+2k-2]} A_j \cup \{a_i, b_i\}, \bigcup_{j \in [\ell'+2k-2]} A_j, R^{(r-2)}\big)$
(d) $V^*(P_i) \cap V^*(P_{i'}) = \{a_i, b_i\} \cap \{a_{i'}, b_{i'}\}$,
(e) $E(P_i) \cap E(P_{i'}) = \varnothing$,
(f) $|V^*(P_i) \setminus \{a_i, b_i\}| \equiv 2 \mod 4$, and
(g) $|V^*(P_i) \cap A_j| \leq 1$ for every $j \in [\ell' + 2k - 2]$.

(I3) $\mathcal{K}_k^a = \{K_{i,k}^a\}_{i \in [s] \setminus I_k}$ and $\mathcal{K}_k^b = \{K_{i,k}^b\}_{i \in [s] \setminus I_k}$ are families of pairwise disjoint sets such that for every $k \leq N$ and $i \in [s] \setminus I_k$ we have

(a) $K_{i,0}^a \subseteq \{a_j\}_{j \in [s]}$ and $K_{i,0}^b \subseteq \{b_j\}_{j \in [s]}$,
(b) $K_{i,k}^a \subseteq A_{\ell'+2k-1}$ and $K_{i,k}^b \subseteq A_{\ell'+2k}$ if $k \geq 1$, and
(c) $|K_{i,k}^a| = |K_{i,k}^b| = 2^k$,

(I4) for every $i \in [s] \setminus I_k$ and $x \in K_{i,k}^a \cup K_{i,k}^b$ there exists a Berge path Q_x such that for every $j \in [s] \setminus (I_k \cup \{i\})$, every $y \in K_{j,k}^a \cup K_{j,k}^b$ and every $j' \in I_k$ we have

(a) Q_x connects a_i to x if $x \in K_{i,k}^a$,
(b) Q_x connects b_i to x if $x \in K_{i,k}^b$,
(c) Q_x is $\big((A_{\ell'+1}, A_{\ell'+3}, \dots, A_{\ell'+2k-3}), R\big)$-compatible if $x \in K_{i,k}^a$,
(d) Q_x is $\big((A_{\ell'+2}, A_{\ell'+4}, \dots, A_{\ell'+2k-2}), R\big)$-compatible if $x \in K_{i,k}^b$,
(e) $V^*(Q_x) \cap V^*(Q_y) = \mathrm{End}(Q_x) \cap \mathrm{End}(Q_y)$,
(f) $V^*(Q_x) \cap V^*(P_{j'}) = \mathrm{End}(Q_x) \cap \{a_{j'}, b_{j'}\}$, and
(g) $E(Q_x) \cap E(P_{j'}) = \varnothing$.

In the set $I_k \subseteq [s]$ we record the indices of the Berge paths $\{P_i\}_{i \in I_k}$ that we have already constructed. The set $K_{i,k}^a$ consists of 2^k vertices of $A_{\ell'+2k-1}$ such that each $x \in K_{i,k}^a$ can be reached from a_i via an $\big((A_{\ell'+1}, A_{\ell'+3}, \dots, A_{\ell'+2k-3}), R\big)$-compatible Berge path Q_x. The set $K_{i,k}^b$ contains 2^k vertices of $A_{\ell'+2k}$, where each vertex $x \in K_{i,k}^b$ is connected to b_i by an $\big((A_{\ell'+2}, A_{\ell'+4}, \dots, A_{\ell'+2k-2}), R\big)$-compatible Berge path Q_x.

Before proving that the induction works, let us first argue why then the lemma follows. Let $k = N+1$, then $I_k = [s]$ and there are Berge paths $\{P_i\}_{i \in [s]}$ with Properties (I2.a)–(I2.g), which immediately imply Properties (C1)–(C7).

For the base case, let $k = 0$. Then, $I_0 = \varnothing$, $\mathcal{P}_0 = \varnothing$, $\mathcal{K}_0^a = \{a_i\}_{i \in [s]}$, $\mathcal{K}_0^b = \{b_i\}_{i \in [s]}$, and $Q_x = (x)$ for every $x \in K_0^a \cup K_0^b$ trivially fulfil Properties (I1)–(I4).

For the inductive step, let $k \leq N$. Assume that for I_k, $\mathcal{P}_k = \{P_i\}_{i \in I_k}$, $\mathcal{K}_k^a = \{K_{i,k}^a\}_{i \in [s] \setminus I_k}$, and $\mathcal{K}_k^b = \{K_{i,k}^b\}_{i \in [s] \setminus I_k}$ Properties (I1)–(I4) hold. For every $i \in [s] \setminus I_k$, let $\{a_i^j, b_i^j\}_{j \in [2^k]}$ be a perfect matching between vertices of $K_{i,k}^a$ and vertices of $K_{i,k}^b$. Let $\{a_j', b_j'\}_{j \in [s]}$ be the union of these matchings. Let $A'_i := A_i \setminus \bigcup_{j\in I_k} V^*(P_j)$ for every $i \in [\ell']$. Using Property (I2.g) we can deduce that for every $i \in [\ell']$ we have

$$|A'_i| \geq |A_i| - |I_k| \geq |A_i| - s \geq \left(1 - \tfrac{\varepsilon}{2}\right)|A_i|.$$

Set

$$\rho(n) = \begin{cases} r(n) & \text{if } k \text{ is odd} \\ r(n)+2 & \text{if } k \text{ is even.} \end{cases}$$

The next claim ensures that we can find Berge paths that connect half of the pairs $\{a'_j, b'_j\}_{j\in[s]}$ using only vertices from $A'_1, \dots, A'_{\rho(n)}$.

Claim 4.15. *Let $\mathcal{A}' = (A'_1, \dots, A'_{\rho(n)})$ be a sequence such that $A'_i \subseteq A_i$ and $|A'_i| \geq (1-\varepsilon/2)|A_i|$ for every $i \in [\rho(n)]$. Let $\{(x_i, y_i)\}_{i\in[s]}$ be pairs of vertices from $V \setminus V(\mathcal{A}')$ such that $x_i \neq x_{i'}$ and $y_i \neq y_{i'}$ for every $i \neq i' \in [s]$.*

Then there exist a subset $I \subseteq [s]$ with $|I| = s/2$ and Berge paths $\{P'_i\}_{i\in I}$ such that for every distinct indices $i, j \in I$ we have

(1) $V^*(P'_i) \cap V^*(P'_j) = \varnothing$,

(2) P'_i connects x_i to y_i,

(3) P'_i is $(A'_1, \dots, A'_{\rho(n)})$-compatible, and

(4) $E(P'_i) \cap E(P'_j) = \varnothing$.

We defer the proof of this claim to the end of the proof of the lemma.

Let $I \subseteq [s]$ with $|I| = s/2$ be the subset and $\mathcal{P}' = \{P'_i\}_{i\in I}$ the set of Berge paths ensured by Claim 4.15 such that for every distinct indices $i, j \in I$ we have $V^*(P'_i) \cap V^*(P'_j) = E(P'_i) \cap E(P'_j) = \varnothing$, path P_i connects a'_i to b'_i and is $(A'_1, \dots, A'_{\rho(n)})$-compatible. This means that there exist a subset $I' \subseteq [s] \setminus I_k$ of size $|I'| = \max\{s2^{-k-1}, 1\}$ such that for every $i \in I'$ there exist a path $P^*_i \in \mathcal{P}'$ and vertices $x_i \in K^a_{i,k}$ and $y_i \in K^b_{i,k}$ such that P^*_i connects x_i to y_i. We define

$$P_i = Q_{x_i} \cdot P^*_i \cdot Q_{y_i}.$$

First let us argue why P_i is a Berge path. As $V^*(Q_{x_i}) \cap V^*(Q_{y_i}) = \varnothing$ by Properties (I4.a)–(I4.d) and $a_i \neq b_i$, and as $V^*(Q_{x_i}) \cap V^*(P^*_i) = \{x_i\}$ and $V^*(Q_{y_i}) \cap V^*(P^*_i) = \{y_i\}$ by using (I4.c), (I4.d) and that $A_1, \dots, A_{\ell'+2k}$ are pairwise disjoint, we have that P_i is a weak path. Since Q_{x_i}, Q_{y_i}, and P^*_i are Berge paths, we only need to ensure that there is no hyperedge that appears in two of the paths Q_{x_i}, Q_{y_i}, and P^*_i. Since Q_{x_i} is $\big((A_{\ell'+1}, A_{\ell'+3}, \dots, A_{\ell'+2k-3}), R\big)$-compatible by Property (I4.c), Q_{y_i} is $\big((A_{\ell'+2}, A_{\ell'+4}, \dots, A_{\ell'+2k-2}), R\big)$-compatible by Property (I4.d), and P^*_i is $(A'_1, \dots, A'_{\rho(n)})$-compatible by Claim 4.15 it follows from Lemma 4.12 that $E(Q_{x_i})$, $E(Q_{y_i})$, and $E(P^*_i)$ are pairwise disjoint. Therefore, P_i is indeed a Berge path for every $i \in I'$.

Let

$$I_{k+1} := I_k \cup I' \text{ and } \mathcal{P}_{k+1} := \mathcal{P}_k \cup \{P_i\}_{i\in I'}.$$

Before defining families of sets $\mathcal{K}^a_{k+1}$ and $\mathcal{K}^b_{k+1}$ as well as paths $\{Q_x : x \in \mathcal{K}^a_{k+1} \cup \mathcal{K}^b_{k+1}\}$ that satisfy Properties (I3) and (I4), respectively, we prove that I_{k+1} and $\mathcal{P}_{k+1}$ fulfil Properties (I1) and (I2), respectively.

Property (I1). Clearly, $I_{k+1} \subseteq [s]$. For $k < N$, we have $|I_{k+1}| = |I_k| + |I'| = s(1 - 2^{-k}) + s2^{-k-1} = s(1 - 2^{-k-1})$. If $k = N$, then $|I_{k+1}| = |I_k| + |I'| = s$. Hence, I_{k+1} satisfies Property (I1).

Property (I2). We have already shown that P_i is a Berge path for every $i \in I'$. The same is true by the induction hypothesis for every other index $i \in I_{k+1}$. We can also conclude using the induction hypothesis that Properties (I2.a)–(I2.g) hold for every $i \in I_k$ and $i' \in I_k$.

Let $i \in I'$. Then P_i can be decomposed into $P_i = Q_{x_i} \cdot P_i^* \cdot Q_{y_i}$ as described earlier. Property (I2.a) follows immediately by Properties (I4.a) and (I4.b) of Q_{x_i} and Q_{y_i}. Furthermore, we can deduce from Properties (I4.c) and (I4.d) for Q_{x_i} and Q_{y_i} as well as from the fact that P_i^* is $(A'_1, \ldots, A'_{\rho(n)})$-compatible that Properties (I2.b), (I2.c) and (I2.g) hold.

Next we verify Property (I2.d), that is, $V^*(P_i) \cap V^*(P_{i'}) = \{a_i, b_i\} \cap \{a_{i'}, b_{i'}\}$ for every $i' \in I_{k+1} \setminus \{i\}$. First assume that $i' \in I_k$. By Property (I4.f), by the definition of $\{A'_j\}_{j\in[\ell']}$ and since P_i^* is $(A'_1, \ldots, A'_{\rho(n)})$-compatible, we know that $V^*(Q_{x_i}) \cap V^*(P_{i'}) = \{a_i\} \cap \{a_{i'}, b_{i'}\}$, $V^*(P_i^*) \cap V^*(P_{i'}) = \varnothing$ and $V^*(Q_{y_i}) \cap V^*(P_{i'}) = \{b_i\} \cap \{a_{i'}, b_{i'}\}$. Hence, $V^*(P_i) \cap V^*(P_{i'}) = \{a_i, b_i\} \cap \{a_{i'}, b_{i'}\}$ follows. Suppose now that $i' \in I'$. Let $P_{i'} = Q_{x_{i'}} \cdot P_{i'}^* \cdot Q_{y_{i'}}$ be the decomposition of $P_{i'}$ as before. By Property (I4.e) we have $\big(V^*(Q_{x_i}) \cup V^*(Q_{y_i})\big) \cap \big(V^*(Q_{x_{i'}}) \cup V^*(Q_{y_{i'}})\big) = \{a_i, b_i\} \cap \{a_{i'}, b_{i'}\}$. Furthermore, we know by Claim 4.15 that $V^*(P_i^*) \cap V^*(P_{i'}^*) = \varnothing$. Since $V^*(P_i^*) \subseteq \bigcup_{j\in[\rho(n)]} A_j \cup \{x_i, y_i\}$ (Properties (2) and (3) of Claim 4.15) and $V^*(Q_{x_{i'}}) \cup V^*(Q_{y_{i'}}) \subseteq \bigcup_{j\in[2k-2]} A_{\ell'+j} \cup \{a_{i'}, b_{i'}, x_{i'}, y_{i'}\}$ (Properties (I4.a)–(I4.d)) we have $V^*(P_i^*) \cap \big(V^*(Q_{x_{i'}}) \cup V^*(Q_{y_{i'}})\big) = \varnothing$. Similarly one can show that $V^*(P_{i'}^*) \cap \big(V^*(Q_{x_i}) \cup V^*(Q_{y_i})\big) = \varnothing$. All in all this shows $V^*(P_i) \cap V^*(P_{i'}) = \{a_i, b_i\} \cap \{a_{i'}, b_{i'}\}$.

Let us now check Property (I2.e), i.e. $E(P_i) \cap E(P_{i'}) = \varnothing$ for every $i' \in I_{k+1} \setminus \{i\}$. If $i' \in I_k$, then we have by Property (I4.g), by Property (3) of Claim 4.15 and by the definition of $\{A'_j\}_{j\in[\ell']}$ that $\big(E(Q_{x_i}) \cup E(P_i^*) \cup E(Q_{y_i})\big) \cap E(P_{i'}) = \varnothing$, implying $E(P_i) \cap E(P_{i'}) = \varnothing$. Now suppose that $i' \in I'$. Let again $P_{i'} = Q_{x_{i'}} \cdot P_{i'}^* \cdot Q_{y_{i'}}$ be the decomposition of $P_{i'}$ as before. Using Property (4) of Claim 4.15, $E(P_i^*)$ and $E(P_{i'}^*)$ are disjoint. Using Lemma 4.12 (part 1) and Properties (I4.c)-(I4.d) and Claim 4.15 (3), we know that $E(P_{i'}^*)$ is disjoint from $E(Q_{x_i}) \cup E(Q_{y_i})$, and $E(P_i^*)$ is disjoint from $E(Q_{x_{i'}}) \cup E(Q_{y_{i'}})$. Similarly, but using Lemma 4.12 (part 3), we also see that $E(Q_{x_i}) \cup E(Q_{y_i})$ and $E(Q_{x_{i'}}) \cup E(Q_{y_{i'}})$ are disjoint. Hence, $E(P_i) \cap E(P_{i'}) = \varnothing$.

It remains to show that $|V^*(P_i) \setminus \{a_i, b_i\}| \equiv 2 \mod 4$ for every $i \in I_{k+1}$. If $i \in I_k$, then this follows by the induction hypothesis. If $i \in I'$, let $P_i = Q_{x_i} \cdot P_i^* \cdot Q_{y_i}$ as before. By Properties (I4.c) and (I4.d) we know that $|V^*(Q_{x_i}) \setminus \{a_i\}| = |V^*(Q_{y_i}) \setminus \{b_i\}| = k$ and by Property (3) of Claim 4.15 that $|V^*(P_i^*) \setminus \{x_i, y_i\}| = \rho(n)$. Hence,

$$|V^*(P_i) \setminus \{a_i, b_i\}| = 2k + \rho(n) = \begin{cases} 2k + r(n) & \text{if } k \text{ is odd} \\ 2(k+1) + r(n) & \text{if } k \text{ is even.} \end{cases}$$

This means in particular that $|V^*(P_i) \setminus \{a_i, b_i\}| \equiv 2 \mod 4$ since $r(n) \equiv 0 \mod 4$.

Let us now turn to the definition of $\mathcal{K}_{k+1}^a$ and $\mathcal{K}_{k+1}^b$. If $k = N$, then we have $[s] \setminus I_{k+1} = \varnothing$ and $\mathcal{K}_{k+1}^a = \mathcal{K}_{k+1}^b = \varnothing$ clearly fulfil (I3). Assume that $k < N$ and set $S^a := \bigcup_{i\in[s]\setminus I_{k+1}} K_{i,k}^a$. Using $|I_{k+1}| = s(1 - 2^{-k-1})$ and $|K_{i,k}^a| = 2^k$ we can deduce that $|S^a| = s/2$. Analogously one can show that for $S^b := \bigcup_{i\in[s]\setminus I_{k+1}} K_{i,k}^b$ we also have $|S^b| = s/2$. We use the following claim to find a $(2, R)$-matching between $S^a \subseteq A_{\ell'+2k-1} \cup \{a_i\}_{i\in[s]}$ and $A_{\ell'+2k+1}$ as well as a $(2, R)$-matching between $S^b \subseteq A_{\ell'+2k} \cup \{b_i\}_{i\in[s]}$ and $A_{\ell'+2k+2}$.

Claim 4.16. *Let $B \in \mathcal{A}$ and let $S \subseteq \big(V(\mathcal{A}) \setminus B\big) \cup \{a_i, b_i\}_{i\in[s]}$ of size $s/2$. Then there exists a $(2, R)$-matching between S and B that saturates S.*

Proof. Define the bipartite graph $G_S = (S \cup B, E_S)$ with

$$E_S := \big\{\{u, v\} : \ u \in S, v \in B \text{ with } e_H\big(u, v, R^{(r-2)}\big) > 0\big\}.$$

Using Hall's theorem (see e.g. [157]) the claim is proved if for every $T \subseteq S$ we have $|N_{G_S}(T)| > 2|T|$. For a contradiction let us assume that there are sets $T \subseteq S$, $N \subseteq B$ with $N_{G_S}(T) \subseteq N$ and $|N| = 2|T|$. We have

$$e_H\big(T, N, R^{(r-2)}\big) = e_H\big(T, B, R^{(r-2)}\big) \overset{(P3)}{\geq} \left(\frac{1}{2^{r-1}} + \frac{\gamma}{2}\right) p|T||B|\binom{|R|}{r-2}.$$

For an upper bound on $e_H\big(T, N, R^{(r-2)}\big)$ we distinguish two cases. First let us assume that $|N|, |T| \leq n/\log^5 n$. Then, for n large enough, it follows from Property (H2) that

$$e_H\big(T, N, R^{(r-2)}\big) \leq |T||N|\binom{|R|}{r-2}p + \varepsilon' \frac{|T|n^{r-1}p}{\log^5 n},$$

which leads to a contradiction with $|N| = 2|T| \leq 2s \leq |B|/2^{r+1}$ and $|B| \geq n/(32 \log^4 n)$ and $|R| \geq \varepsilon n/10$ as well as $\varepsilon' \leq \big(\frac{\gamma\varepsilon}{10r}\big)^r$.

Now suppose that $|N|, |T| \geq n/(2 \log^5 n)$. Then, using Property (H1) we obtain

$$e_H\big(T, N, R^{(r-2)}\big) \leq (1 + \varepsilon')p|T||N|\binom{|R|}{r-2},$$

which also leads to a contradiction with $|N| = 2|T| \leq 2s \leq |B|/2^{r+1}$ and $\varepsilon' \leq \gamma/2^{r+1}$. □

Let M^a denote a $(2, R)$-matching between S^a and $A_{\ell'+2k+1}$ that saturates S and let M^b denote a $(2, R)$-matching between S^b and $A_{\ell'+2k+2}$ that saturates S^b. For every $i \in [s] \setminus I_{k+1}$ let

$$K^a_{i,k+1} := \{x \in A_{\ell'+2k+1} : \exists\ e \in M^a \text{ with } x \in e\}$$

and

$$K^b_{i,k+1} := \{x \in A_{\ell'+2k+2} : \exists\ e \in M^b \text{ with } x \in e\}.$$

Property (I3). By definition we have for every $i \in [s] \setminus I_{k+1}$ that $K^a_{i,k+1} \subseteq A_{\ell'+2k+1}$ and $K^b_{i,k+1} \subseteq A_{\ell'+2k+2}$ and $|K^a_{i,k+1}| = |K^b_{i,k+1}| = 2|K^a_{i,k}| = 2^{k+1}$. Hence, Property (I3) holds for $\mathcal{K}^a_{k+1} := \{K^a_{i,k+1}\}_{i\in[s]\setminus I_{k+1}}$ and $\mathcal{K}^b_{k+1} := \{K^b_{i,k+1}\}_{i\in[s]\setminus I_{k+1}}$.

Let $x \in K^a_{i,k+1} \cup K^b_{i,k+1}$. If $x \in K^a_{i,k+1}$, then there exists a unique hyperedge $e_x \in M^a$ with $x \in e_x$. Let $y_x \in K^a_{i,k}$ be the unique vertex in $e_x \cap S^a$. Analogously, if $x \in K^b_{i,k+1}$, then there exists a unique hyperedge $e_x \in M^b$ that contains x. In this case, let $y_x \in K^b_{i,k}$ denote the unique vertex in $e_x \cap S^b$.

Using this notation we define for every $x \in K^a_{i,k+1} \cup K^b_{i,k+1}$ the weak path

$$Q_x := Q_{y_x} \cdot (y_x, e_x, x),$$

where Q_{y_x} is a Berge path connecting a_i and y_x (if $y_x \in K^a_{i,k}$) or b_i and y_x (if $y_x \in K^b_{i,k}$) with Property (I4). We show that Property (I4) holds for all paths Q_x with $x \in K^a_{i,k+1} \cup K^b_{i,k+1}$.

Property (I4). Obviously, if $x \in K^a_{i,k+1}$, then Q_x connects a_i to x. If $x \in K^b_{i,k+1}$, then Q_x connects b_i to x. Hence, Properties (I4.a) and (I4.b) hold for $\{Q_x : x \in K^a_{i,k+1} \cup K^b_{i,k+1},\ i \in [s] \setminus I_{k+1}\}$.

For $x \in K^a_{i,k+1}$ we have that Q_{y_x} is $\big((A_{\ell'+1}, A_{\ell'+3}, \ldots, A_{\ell'+2k-3}), R\big)$-compatible and $e_x \in E_H(A_{\ell'+2k-3}, A_{\ell'+2k-1}, R^{(r-2)})$. Hence, Q_x is $\big((A_{\ell'+1}, A_{\ell'+3}, \ldots, A_{\ell'+2k-1}), R\big)$-compatible. This shows Property (I4.c). Analogously, one can show that Property (I4.d) holds as well.

Let $x_1 \in K^a_{i,k+1} \cup K^b_{i,k+1}$ and $x_2 \in K^a_{j,k+1} \cup K^b_{j,k+1}$ for any distinct indices $i, j \in [s] \setminus I_{k+1}$. As before, let $Q_{x_1} = Q_{y_{x_1}} \cdot (y_{x_1}, e_{x_1}, x_1)$ and $Q_{x_2} = Q_{y_{x_2}} \cdot (y_{x_2}, e_{x_2}, x_2)$. By induction we have $V^*(Q_{y_{x_1}}) \cap V^*(Q_{y_{x_2}}) = \mathrm{End}(Q_{y_{x_1}}) \cap \mathrm{End}(Q_{y_{x_2}})$. Since it holds that $y_{x_1} \neq y_{x_2}$ and also that $x_1, x_2 \notin V^*(Q_{y_{x_1}}) \cup V^*(Q_{y_{x_2}})$, we conclude $V^*(Q_{x_1}) \cap V^*(Q_{x_2}) = \mathrm{End}(Q_{x_1}) \cap \mathrm{End}(Q_{x_2})$. Hence, Property (I4.e) holds.

Next we show Property (I4.f), that is, $V^*(Q_x) \cap V^*(P_j) = \mathrm{End}(Q_x) \cap \{a_j, b_j\}$ for every $x \in K^a_{i,k+1} \cup K^b_{i,k+1}$ and $j \in I_{k+1} \setminus \{i\}$. Let $Q_x = Q_{y_x} \cdot (y_x, e_x, x)$ as before. Since $x \in A_{\ell'+2k+1} \cup A_{\ell'+2k+2}$ and $V^*(P_j) \subseteq \bigcup_{j \in [\ell'+2k]} A_j \cup \{a_j, b_j\}$ (Property (I2.b)), we have $x \notin V^*(P_j)$. Therefore, it is enough to verify that $V^*(Q_{y_x}) \cap V^*(P_j) = \mathrm{End}(Q_{y_x}) \cap \{a_j, b_j\}$. If $j \in I_k$, then this follows directly by induction. If, on the other hand, $j \in I'$, let $P_j = Q_{x_j} \cdot P^*_j \cdot Q_{y_j}$ as before. By induction and since y_x, x_j, y_j are distinct, we have $\big(V^*(Q_{x_j}) \cup V^*(Q_{y_j})\big) \cap V^*(Q_{y_x}) = (\mathrm{End}(Q_{x_j}) \cup \mathrm{End}(Q_{y_j})) \cap \mathrm{End}(Q_{y_x}) = \{a_j, b_j\} \cap \mathrm{End}(Q_{y_x})$. Moreover, $V^*(Q_x) \cap \Big(\bigcup_{j' \in [\ell']} A_{j'}\Big) = \varnothing$ while $V^*(P^*_j) \subseteq \bigcup_{j' \in [\ell']} A_{j'} \cup \{x_j, y_j\}$, and $x_j, y_j \notin V^*(Q_x)$. Thus, $V^*(P_j) \cap V^*(Q_x) = \varnothing$. As a consequence, Property (I4.f) is satisfied.

Finally, it remains to show Property (I4.g), i.e. $E(Q_x) \cap E(P_j) = \varnothing$ for every $x \in K^a_{i,k+1} \cup K^b_{i,k+1}$ and $j \in I_{k+1}$. Let $Q_x = Q_{y_x} \cdot (y_x, e_x, x)$ as before. If $j \in I_k$, then $E(Q_{y_x}) \cap E(P_j) = \varnothing$ by induction. Since we also have $x \notin \bigcup_{j' \in [\ell'+2k-2]} A_{j'} \cup \{a_j, b_j\} \supseteq V(P_j)$ it follows that $E(Q_x) \cap E(P_j) = \varnothing$. If $j \in I'$, let $P_j = Q_{x_j} \cdot P^*_j \cdot Q_{y_j}$ as before. Using Properties (I4.c)–(I4.e), Property (3) of Claim 4.15 for P^*_j and $x \notin V(Q_{x_j}) \cup V(Q_{y_j})$, we deduce with Lemma 4.12 that $E(Q_x)$ is disjoint from $E(Q_{x_j})$, $E(P^*_j)$, and $E(Q_{y_j})$. Hence, $E(Q_x) \cap E(P_j) = \varnothing$ holds again, which finishes the induction.

Now let us turn to the proof of Claim 4.15, for which we need the following two claims.

Claim 4.17. *Let $A \in \mathcal{A}$ and let $B \subseteq V(H)$ be a subset of size $n/(10 \log^5 n)$ such that $A \cap B = \varnothing$. Then for every $A' \subseteq A$ with $|A'| \geq (1-\varepsilon)|A|$ we have $|N^1_{(A')}(B)| > |A|/2$.*

Proof of Claim 4.17. Assume for a contradiction that the claim is not true. Then there exist a set $A' \subseteq A$ with $|A'| \geq (1-\varepsilon)|A|$ and a set $N' \subseteq A$ of size $|A|/2$ such that $N := N^1_{(A')}(B) \subseteq N'$. Furthermore, let $D \subseteq A$ be an arbitrary subset of size $\varepsilon|A|$ with $A \setminus A' \subseteq D$. Due to the assumption on ε' we then obtain

$$\begin{aligned}
e_H(B, N^{(r-1)}) &= e_H(B, (A')^{(r-1)}) \geq e_H(B, A^{(r-1)}) - e_H(B, A \setminus A', A^{(r-2)}) \\
&\geq e_H(B, A^{(r-1)}) - e_H(B, D, A^{(r-2)}) \\
&\geq \left(\frac{1}{2^{r-1}} + \frac{\gamma}{2}\right) p|B| \binom{|A|}{r-1} - \left(1 + \varepsilon'\right) \varepsilon |B||A| \binom{|A|}{r-2} p \\
&> \left(\frac{1}{2^{r-1}} + \frac{\gamma}{20}\right) p|B| \binom{|A|}{r-1},
\end{aligned}$$

where in the second last inequality we use Properties (P2) and (H1). On the other hand, we obtain

$$\begin{aligned} e_H(B, N^{(r-1)}) \leq e_H(B, (N')^{(r-1)}) &\overset{(H1)}{\leq} \left(1+\varepsilon'\right) p|B| \binom{|N'|}{r-1} \\ &\leq \left(1+\frac{\gamma}{2^{r+1}}\right) p|B| \cdot \frac{1}{2^{r-1}} \binom{|A|}{r-1}, \end{aligned}$$

which yields a contradiction. □

Claim 4.18. *Let $m \geq \log_2 n + 1$, let $\mathcal{D} = (D_1, \ldots, D_m)$ be a subsequence of $\mathcal{A}$ and let $\mathcal{D}' = (D'_1, \ldots, D'_m)$ be a sequence with $D'_i \subseteq D_i$ and $|D'_i| \geq (1-\varepsilon)|D_i|$ for every $i \in [m]$. Let further $B \subseteq V(H) \setminus \left(\bigcup_{i \in [m]} D_i\right)$ with $|B| \geq n/(10 \log^5 n)$. Then there exists a vertex $x \in B$ such that for every $\log_2 n + 1 \leq i \leq m$ we have $|N^i_{\mathcal{D}'}(x)| > |D_i|/2$.*

Proof. We prove by induction that for every $i \in [m]$ there exists a set $B_{i-1} \subseteq B$ of size $|B_{i-1}| \leq \max\{1, \lceil |B|/2^{i-1} \rceil\}$ such that $|N^i_{\mathcal{D}'}(B_{i-1})| > |D_i|/2$. Since then $|B_{i-1}| = 1$ holds for every $i \geq \log_2 n + 1$, Claim 4.18 follows directly from this statement. For $i = 1$, Claim 4.17 ensures that for $B_0 = B$ we have $|N^1_{\mathcal{D}'}(B_0)| > |D_i|/2$.

For the inductive step, assume that the statement is true for $i < m$. Let B_{i-1} be a set such that $|N^i_{\mathcal{D}'}(B_{i-1})| > |D_i|/2$. One can easily find a subset $B_i \subseteq B_{i-1}$ such that $|B_i| = \lceil |B_{i-1}|/2 \rceil$ with $|N^i_{\mathcal{D}'}(B_i)| \geq |D_i|/4 > n/(10 \log^5 n)$. Using Claim 4.17 again, we have $|N^{i+1}_{\mathcal{D}'}(B_i)| = |N^1_{(D'_{i+1})}(N^i_{\mathcal{D}'}(B_i))| > |D_i|/2$, which completes the inductive step. □

Finally, we are in the position to prove Claim 4.15.

Proof of Claim 4.15. Let $I \subseteq [s]$ be a largest subset of $[s]$ such that there exist Berge paths $\{P'_i\}_{i \in I}$ as described in the statement of the claim. Assume for a contradiction that $|I| < s/2$. Let

$$C_j := A'_j \setminus \bigcup\nolimits_{i \in I} V^*(P'_i)$$

for every $j \in [\rho(n)]$ and let

$$\rho'(n) = \frac{\rho(n)}{2}.$$

We also define $\mathcal{C}' := (C_1, \ldots, C_{\rho'(n)})$ and $\mathcal{C}'' = (C_{\rho(n)}, \ldots, C_{\rho'(n)})$, and note that both sequence have length larger than $\log_2 n + 1$.

As P'_i is $\mathcal{A}'$-compatible and therefore $|A'_j \cap V^*(P'_i)| \leq 1$ for every $j \in [\rho(n)]$ and $i \in I$ we have $|C_j| > |A'_j| - s/2 \geq (1-\varepsilon)|A_j|$ for every $j \in [\rho(n)]$. As $|[s] \setminus I| > s/2 \geq n/(2 \log^5 n)$ we know that there are at least $|[s] \setminus I|/2 + 1$ vertices $x \in \{x_i : i \in [s] \setminus I\}$ with $|N^{\rho'(n)}_{\mathcal{C}'}(x)| > |A_{\rho'(n)}|/2$ by iteratively applying Claim 4.18. Analogously, we can find $|[s] \setminus I|/2 + 1$ vertices $y \in \{y_i : i \in [s] \setminus I\}$ with $|N^{\rho'(n)+1}_{\mathcal{C}''}(y)| > |A_{\rho'(n)}|/2$. Hence, there exists an index $i^* \in [s] \setminus I$ such that $N^{\rho'(n)}_{\mathcal{C}'}(x_{i^*}) \cap N^{\rho'(n)+1}_{\mathcal{C}''}(y_{i^*}) \cap A_{\rho'(n)} \neq \varnothing$. Let $v_{\rho'(n)} \in N^{\rho'(n)}_{\mathcal{C}'}(x_{i^*}) \cap N^{\rho'(n)+1}_{\mathcal{C}''}(y_{i^*}) \cap A_{\rho'(n)}$, then by definition there exist

- a Berge path $P_x = (x_{i^*}, e_0, v_1, e_1, v_2, \ldots, v_{\rho'(n)-1}, e_{\rho'(n)-1}, v_{\rho'(n)})$ such that $v_i \in C_i$ for every $i \in [\rho'(n)]$, $e_i \in E_H\big(v_i, v_{i+1}, C^{(r-2)}_{i+1}\big)$ for every $i \in [\rho'(n) - 1]$, and $e_0 \in E_H\big(x_{i^*}, v_1, C^{(r-2)}_1\big)$,

- a Berge path $P_y = (y_{i^*}, e_{\rho(n)+1}, v_{\rho(n)}, e_{\rho(n)}, v_{\rho(n)-1}, \ldots, v_{\rho'(n)+1}, e_{\rho'(n)+1}, v_{\rho'(n)})$ such that $v_i \in C_i$ for every $i \in \{\rho'(n), \ldots, \rho(n)\}$, $e_i \in E_H(v_{i-1}, v_i, C_{i-1}^{(r-2)})$ for every $i \in \{\rho'(n) + 1, \ldots, \rho(n)\}$, and $e_{\rho(n)+1} \in E_H(y_{i^*}, v_{\rho(n)}, C_{\rho(n)}^{(r-2)})$.

But then $P_{i^*} := P_x \cdot P_y$ is a Berge path that connects x_{i^*} and y_{i^*} and which is $(C_1, \ldots, C_{\rho(n)})$-compatible and therefore $(A'_1, \ldots, A'_{\rho(n)})$-compatible. Moreover, by the definition of the sets C_j, we conclude that $V^*(P_{i^*})$ is disjoint from $\bigcup_{i \in I} V^*(P_i)$ and $E(P_{i^*})$ is disjoint from $\bigcup_{i \in I} E(P_i)$. This, however, is a contradiction to the maximality of I. □

This finishes the proof of Lemma 4.9. □

4.1.5 Absorbing lemma

In this subsection we prove a lemma that ensures the existence of a Berge path Q in a (p, ε')-pseudorandom r-uniform hypergraph H given with a vertex partition by the partition lemma (Lemma 4.8) such that Q 'absorbs' every subset of the set Z of the vertex partition of H. More precisely, we prove the following lemma.

Lemma 4.19 (Absorbing lemma)**.** *Let $r \geq 3$ be an integer and let $p \geq \log^{8r} n / n^{r-1}$. Furthermore, let $\gamma > 0$ and $\varepsilon \geq \varepsilon' > 0$ be reals such that $\varepsilon < 10^{-3}\gamma/2^{r+1}$ and $\varepsilon' \leq \left(\frac{\gamma\varepsilon}{10r}\right)^r$. Let H be a (p, ε')-pseudorandom 3-uniform hypergraph given with a partition*

$$V(H) = \bigcup_{i \in [\ell]} Y_i \cup \bigcup_{i \in [\ell]} Z_i \cup \bigcup_{j \in [t]} W_j$$

with $t := \log^5 n$ and $\ell := 16 \log n$ that satisfies Properties (P1)–(P3) from Lemma 4.8. Set $A = \bigcup_{i \in [8 \log n]} Y_i$ and $B = \bigcup_{i \in [8 \log n]} Y_{\ell+1-i}$. Moreover, let $R_1, R_2 \subseteq \bigcup_{j \in [t]} W_j$ be two disjoint subsets each of size at least $\varepsilon n/10$.

Then H contains a Berge path Q with $V^*(Q) \subseteq Y$ such that for every $M \subseteq \bigcup_{i \in [\ell]} Z_i$ there exists a Berge path Q_M with the same endpoints as Q such that

(Q1) $V^*(Q_M) = V^*(Q) \cup M$,

(Q2) $E(Q_M) \subseteq E_H(A) \cup E_H(A^{(2)}, R_1^{(r-2)}) \cup E_H(M, A^{(r-1)}) \cup E_H(M, A, R_1^{(r-2)})$ $\cup E_H(B) \cup E_H(B^{(2)}, R_2^{(r-2)}) \cup E_H(A, B^{(r-1)}) \cup E_H(A, B, R_2^{(r-2)})$, *and*

(Q3) $|V^*(Q)| \leq 2n/\log n$.

Property (Q2) in Lemma 4.19 will be essential in the proof of Theorem 4.7, when we need to argue why certain Berge paths are edge-disjoint. Before we prove Lemma 4.19 we introduce the definitions of x-*absorbing sets* and *certifying paths*, and prove that a specific construction yields an x-aborbing set.

Definition 4.20 (x-absorbing, certifying path)**.** *Let H be an r-uniform hypergraph, let $A \subseteq V(H)$ and let $x, s_x, t_x \in A$ be distinct vertices. We say that A is x-*absorbing *with endpoints s_x and t_x if there exist two Berge paths P_x and P'_x in H with endpoints s_x and t_x such that*

(1) $x \notin V(P_x)$ *and*

(2) $V^*(P_x) \cup \{x\} = V^*(P'_x) \subseteq A$.

Additionally, we say that P_x and P'_x certify *that A is x-absorbing.*

The next lemma yields an x-absorbing set with two certifying paths constructed using a Berge cycle of length $(4k+3)$ and $2k$ Berge paths for any positive integer k. In Figure 4.3 we illustrate such a construction.

Lemma 4.21. *Let H be an r-uniform hypergraph and let $k \in \mathbb{N}$. Let C_x be a Berge cycle in H with $(x, s_x, s_2^x, t_1^x, s_3^x, t_2^x, \dots, s_{2k}^x, t_{2k-1}^x, t_x, t_{2k}^x, s_1^x)$ being the order of $V^*(C_x)$ on C_x (up to cyclic permutation). For $i \in [2k]$ let P_i^x be a Berge path in H such that the following four properties hold for all distinct $i, i' \in [2k]$:*

(1) $V^*(P_i^x) \cap V^*(P_{i'}^x) = \varnothing$,

(2) $V^*(P_i^x) \cap V^*(C_x) = \{s_i^x, t_i^x\}$,

(3) s_i^x and t_i^x are the endpoints of P_i^x,

(4) $\big(E(P_i^x) \cup E(C_x)\big) \cap E(P_{i'}^x) = \varnothing$.

Then $A_x := V^*(C_x) \cup \bigcup_{i\in[2k]} V^*(P_i^x)$ is x-absorbing with endpoints s_x and t_x, and with two certifying paths P_x and P'_x such that $E(P_x) \cup E(P'_x) = E(C_x) \cup \bigcup_{i\in[2k]} E(P_i^x)$.

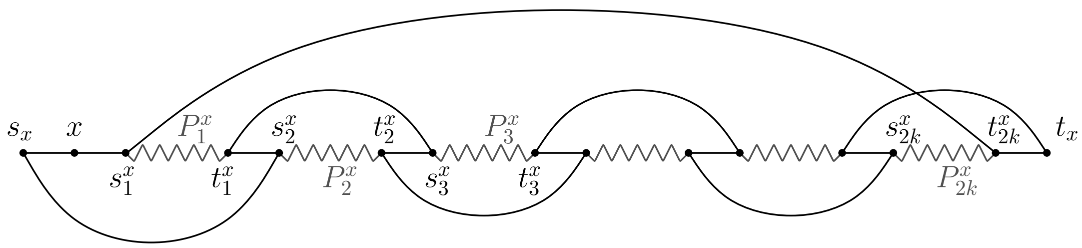

Figure 4.3: An illustration of the graph $G = (A_x, E)$, where the edge set E is defined as $E = \big\{\{x,y\} \in \binom{A_x}{2} : \exists\, e \in E(C_x) \cup \bigcup_{i\in[2k]} E(P_i^x) \text{ with } \{x,y\} \subseteq e\big\}$.

Proof of Lemma 4.21. Let

$$P'_x := (s_x, f, x, g, s_1^x) \cdot P_1^x \cdot (t_1^x, e_1, s_2^x) \cdot P_2^x \cdot (t_2^x, e_2, s_3^x) \cdot \dots \cdot (t_{2k-1}^x, e_{2k-1}, s_{2k}^x) \cdot P_{2k}^x \cdot (t_{2k}^x, h, t_x),$$

where f, g, h, e_i are the unique hyperedges in $E(C_x)$ with $\{s_x, x\} \subseteq f$, $\{x, s_1^x\} \subseteq g$, $\{t_{2k}^x, t_x\} \subseteq h$, and $\{t_i^x, s_{i+1}^x\} \subseteq e_i$ for every $i \in [2k-1]$. Furthermore, let

$$\begin{aligned} P_x := {} & (s_x, f', s_2^x) \cdot P_2^x \cdot (t_2^x, e_{(2,4)}, s_4^x) \cdot P_4^x \cdot (t_4^x, e_{(4,6)}, s_6^x) \cdot \dots \cdot (t_{2k-2}^x, e_{(2k-2,2k)}, s_{2k}^x) \cdot P_{2k}^x \\ & \cdot (t_{2k}^x, g', s_1^x) \cdot P_1^x \cdot (t_1^x, e_{(1,3)}, s_3^x) \cdot P_3^x \cdot \dots (t_{2k-3}^x, e_{(2k-3,2k-1)}, s_{2k-1}^x) \cdot P_{2k-1}^x \cdot (t_{2k-1}^x, h', t_x), \end{aligned}$$

where $f', g', h', e_{(i,i+2)}$ are the unique hyperedges in $E(C_x)$ with $\{s_x, s_2^x\} \subseteq f'$, $\{t_{2k}^x, s_1^x\} \subseteq g'$. $\{t_{2k-1}^x, t_x\} \subseteq h'$, and $\{t_i^x, s_{i+2}^x\} \subseteq e_{(i,i+2)}$ for every $i \in [2k-2]$.

Using Properties (1)–(4) one can verify quite easily that P_x and P'_x are Berge paths. Clearly, P_x and P'_x have s_x and t_x as endpoints, and $E(P_x) \cup E(P'_x) = E(C_x) \cup \bigcup_{i\in[2k]} E(P_i^x)$ holds. Furthermore, $x \notin V(P_x)$ and $V^*(P'_x) = V^*(P_x) \cup \{x\}$. $\square$

Now we turn to the proof of the absorbing lemma.

Proof of Lemma 4.19. Let $\{z_1, \dots, z_m\} := \bigcup_{i\in[\ell]} Z_i$ with $n/(32\log^3 n) \le m \le n/(16\log^3 n)$. First we aim to find z_i-absorbing sets $A_{z_i} \subseteq A$, together with certifying paths P_{z_i} and P'_{z_i}, for every $i \in [\ell]$. To this end we proceed in two steps.

For the first step, set $\mathcal{A}_1 = (Y_1, Y_2, \dots, Y_{8\log n})$ and observe that $m \le \varepsilon|Y_i|/4$ for every $i \in [8\log n]$. By Lemma 4.9 (with $\mathcal{A} = \mathcal{A}_1$ and $R = R_1$), we therefore find Berge cycles $C_1, \dots, C_m$ such that for all distinct indices $i, i' \in [m]$ we have

(S1.1) $z_i \in V^*(C_i)$,

(S1.2) $V^*(C_i) \subseteq A \cup \{z_i\}$,

(S1.3) $E(C_i) \subseteq E_H\big(A \cup \{z_i\}, A^{(r-1)}\big) \cup E_H\big(A \cup \{z_i\}, A, R_1^{(r-2)}\big)$

(S1.4) $V^*(C_i) \cap V^*(C_{i'}) = \varnothing$,

(S1.5) $E(C_i) \cap E(C_{i'}) = \varnothing$, and

(S1.6) there exists $k_i \in \mathbb{N}$ such that $|V^*(C_i)| = 4k_i + 3 \le 8\log n$.

With Lemma 4.21 in mind, let $(z_i, s_{z_i}, s_2^{z_i}, t_1^{z_i}, s_3^{z_i}, t_2^{z_i}, \dots, s_{2k_i}^{z_i}, t_{2k_i-1}^{z_i}, t_{z_i}, t_{2k_i}^{z_i}, s_1^{z_i})$ denote the ordering of $V^*(C_i)$ on the Berge cycle C_i.

For the second step, set $\mathcal{A}_2 = (Y_{8\log n+1}, Y_{8\log n+2}, \dots, Y_\ell)$. We now consider all (disjoint) pairs $(s_j^{z_i}, t_j^{z_i})$ with $j \in [k_i]$ and $i \in [m]$ as well as all (disjoint) pairs (t_{z_1}, s_{z_2}), (t_{z_2}, s_{z_3}), …, $(t_{z_{m-1}}, s_{z_m})$. These are in total at most $s := m\cdot\max_{i\in[m]}\{k_i\} + m - 1 \le 3m\log n$ pairs. Hence, $s \le \varepsilon|Y_i|/4$ for every $i \in [\ell]$. Therefore, Lemma 4.9 (with $\mathcal{A} = \mathcal{A}_2$ and $R = R_2$) ensures that we can find Berge paths P_j^i for every $j \in [k_i]$ and $i \in [m]$ as well as Berge paths P_i for every $i \in [m-1]$ such that for every distinct $i, i' \in [m-1]$ and $j, j' \in [m]$ and for every $k \in [k_j]$ and for every $(j', k') \in ([m] \times [k_{j'}]) \setminus (j, k)$ it holds that

(S2.1) P_k^j connects $s_k^{z_j}$ to $t_k^{z_j}$,

(S2.2) P_i connects t_{z_i} to $s_{z_{i+1}}$,

(S2.3) $V^*(P_k^j) \subseteq B \cup \{s_k^{z_j}, t_k^{z_j}\}$,

(S2.4) $V^*(P_i) \subseteq B \cup \{t_{z_i}, s_{z_{i-1}}\}$,

(S2.5) $E(P_k^j) \cup E(P_i) \subseteq E_H\big(A \cup B, B^{(r-1)}\big) \cup E_H\big(A \cup B, B, R_2^{(r-2)}\big)$,

(S2.6) $V^*(P_j^i)$, $V^*(P_{j'}^{i'})$, $V^*(P_i)$, and $V^*(P_{i'})$ are pairwise disjoint,

(S2.7) $E(P_j^i)$, $E(P_{j'}^{i'})$, $E(P_i)$, and $E(P_{i'})$ are pairwise disjoint, and

(S2.8) $|V^*(P_k^j)|, |V^*(P_i)| \le 8\log n$.

Using all these cycles and paths, we are able to construct z_i-absorbing sets A_{z_i}. For every $i \in [m]$, it follows from Lemma 4.21 that $A_{z_i} = V^*(C_i) \cup \bigcup_{j\in[k_i]} V^*(P_j^i)$ is z_i-absorbing with endpoints s_{z_i} and t_{z_i}. Indeed, Property (1) from Lemma 4.21 is given by (S2.6). Property (2) follows from (S1.2), (S2.1), and (S2.3). Property (3) follows from (S2.1), and Property (4) is given by (S1.3), (S2.5) and (S2.7).

Moreover, by Lemma 4.21, we find two certifying paths P_{z_i} and P'_{z_i} with endpoints s_{z_i} and t_{z_i} such that $z_i \notin V^*(P_{z_i})$, such that $V^*(P'_{z_i}) = V^*(P_{z_i}) \cup \{z_i\} \subseteq A_{z_i}$ and $E(P_{z_i}) \cup E(P'_{z_i}) = E(C_i) \cup \bigcup_{j\in[k_i]} E(P^i_j)$.

We now consider the weak path

$$Q := P_{z_1} \cdot P_1 \cdot P_{z_2} \cdot P_2 \cdot \ldots \cdot P_{m-1} \cdot P_{z_m}.$$

Observe that $V^*(Q) \cap \{z_1, \ldots, z_m\} = \varnothing$ since $V^*(P_{z_i}) = \big(V^*(C_i) \setminus \{z_i\}\big) \cup \bigcup_{j\in[k_i]} V^*(P^i_j) \subseteq Y$ for every $i \in [m]$, where we used Properties (S1.2) and (S2.3), and $V^*(P_i) \subseteq Y$ for every $i \in [m-1]$, where we used Property (S2.4).

Now, for every $i \in [m-1]$ and $i' \in [m]$, P_i is edge-disjoint from $P_{z_{i'}}$ since $E(P_{z_{i'}}) \subseteq E(C_{i'}) \cup \bigcup_{j\in[k_{i'}]} E(P^{i'}_j)$ and by Properties (S1.3), (S2.5), and (S2.7). All paths P_i with $i \in [m-1]$ are pairwise edge-disjoint by Property (S2.7), and similarly all paths P_{z_i} are pairwise edge-disjoint since $E(P_{z_i}) \subseteq E(C_i) \cup \bigcup_{j\in[k_i]} E(P^i_j)$ holds and by Properties (S1.3),(S1.5),(S2.5) and (S2.7). Hence, it follows that Q is a Berge path. By Properties (S1.6) and (S2.8) we obtain that $|V^*(Q)| \leq 32m \log^2 n \leq 2n/\log n$.

Let $M \subseteq [m]$ be given. Then we construct a path Q_M by taking the definition of Q and by replacing every path P_{z_i} by the path P'_{z_i}, whenever $i \in M$ holds. Then, its endpoints are s_{z_1} and t_{z_m}, and thus the same as of the path Q. Since $E(P'_{z_i}) \subseteq E(C_i) \cup \bigcup_{j\in[k_i]} E(P^i_j)$ holds for every $i \in M$, it follows analogously to the above discussion that Q_M is a Berge path. Moreover, we obtain $V^*(Q_M) \cap \{z_1, \ldots, z_m\} = \{z_i : \ i \in M\}$ as $V^*(P'_{z_i}) = V^*(P_{z_i}) \cup \{z_i\}$ holds for every $i \in M$. $\square$

4.1.6 Proof of the main theorem

In this subsection we first present the proof of Theorem 4.7 and then show that it directly implies Theorem 4.1 using Lemma 4.6.

Proof of Theorem 4.7. Let r, γ, ε', p, $\mathcal{H}$ be given according to the assumptions of the theorem. Set $\varepsilon = 10^{-3}\gamma/2^{r+2}$. Let H be a spanning subhypergraph of $\mathcal{H}$ with $\delta_1(H) \geq (1/2^{r-1} + \gamma)p\binom{n}{r-1}$.

Since $\mathcal{H}$ is (p, ε')-pseudorandom by assumption, we have that H is (p, ε')-pseudorandom. Hence we may apply Lemma 4.8 with input r, γ, ε, and p to H and get a partition $V(H) = Y \cup Z \cup W$ and a refined partition $\mathcal{P} = \{Y_i, Z_i, W_j\}_{i\in[\ell], j\in[t]}$ with $Y = \bigcup_{i\in[\ell]} Y_i$ and $Z = \bigcup_{i\in[\ell]} Z_i$ and $W = \bigcup_{j\in[t]} W_j$ with $\ell := 16 \log n$ and $t := \log^5 n$ such that Properties (P1)–(P3) of Lemma 4.8 are fulfilled.

Next we introduce a few definitions. First we divide Y into two sets, each consisting of $8 \log n$ clusters of $\{Y_i\}_{i\in[\ell]}$. We set

$$A = \bigcup_{i\in[8\log n]} Y_i \quad \text{and} \quad B = \bigcup_{i\in[8\log n]} Y_{\ell+1-i}.$$

We also need the following subset B' of B:

$$B' := \bigcup_{i\in[4\log n]} Y_{\ell+1-i} \subseteq B.$$

The set W is also split into two sets. Let

$$R_1 = \bigcup_{j\in[t/2]} W_j \quad \text{and} \quad R_2 = \bigcup_{j\in[t/2]} W_{t+1-j}.$$

Applying Lemma 4.19, we find a Berge path Q in H with $V^*(Q) \subseteq Y$ such that for every $M \subseteq Z$ there exists a Berge path Q_M with the same endpoints as Q such that Properties (Q1)–(Q3) from Lemma 4.19 hold with R_1 and R_2 as defined above. We denote the endpoints of Q by x_Q and y_Q.

Next we distribute all but at most $t/2$ vertices of $A \setminus V^*(Q)$ to $W_1, \dots, W_{t/2}$ such that the resulting clusters $S_1, \dots, S_{t/2}$ with $W_i \subseteq S_i$ for every $i \in [t/2]$ have equal size. Similarly, we distribute all but at most $t/2$ vertices of $B \setminus V^*(Q)$ to $W_{t/2+1}, \dots, W_t$ such that the resulting clusters $S_{t/2+1}, \dots, S_t$ with $W_{t+1-i} \subseteq S_{t+1-i}$ for every $i \in [t/2]$ have equal size and such that $B' \subseteq \bigcup_{i\in[t/2]} S_{t+1-i}$ and $(S_{t/2+1} \cup S_t) \cap B' = \varnothing$.

Let us argue why such a distribution is possible. Since $(1 - 4\varepsilon/5)n/\log^5 n \le |W_j| \le (1-3\varepsilon/5)n/\log^5 n$ for every $j \in [t]$ by Property (P1), one needs to add at most $\varepsilon n/10$ vertices to $W_1, \dots, W_{t/2}$ to extend them to equally sized clusters. The same holds for $W_{t/2+1}, \dots, W_t$. Since $|Y_i| \ge \varepsilon n/(2\ell)$ for every $i \in [\ell]$ by Property (P1) and $|V^*(Q)| \le 2n/\log n$ by Property (Q3) we have for sufficiently large n in particular $|A \setminus V^*(Q)| \ge \varepsilon n/5$ and $|B \setminus V^*(Q)| \ge \varepsilon n/5$. By definition of B' it can be seen quickly that it is possible to choose a distribution such that $B' \subseteq \bigcup_{i\in[t/2]} S_{t+1-i}$ and $(S_{t/2+1} \cup S_t) \cap B' = \varnothing$. The vertices of $A \setminus V^*(Q)$ and $B \setminus V^*(Q)$ that were not distributed are stored in the set S.

As a next step we construct edge-disjoint Berge paths such that the union of their vertex sequences cover $\bigcup_{j\in[t]} S_j$. For this we use the following claim, whose proof we defer to the end of the current proof.

Claim 4.22. *There exists a family $\mathcal{S}$ of Berge paths with $\bigcup_{P\in\mathcal{S}} V^*(P) = \bigcup_{j\in[t]} S_j$ and $|\mathcal{P}| \le 4n/\log^5 n$ such that for every distinct $P, P' \in \mathcal{S}$ the following holds:*

(1) $V^*(P) \cap V^*(P') = \varnothing$ and $E(P) \cap E(P') = \varnothing$,

(2) for every $M \subseteq Z$, we have $V^*(P) \cap V^*(Q_M) = \varnothing$ and $E(P) \cap E(Q_M) = \varnothing$, and

(3) $V(P) \cap Z = \varnothing$.

Let $\mathcal{S} = \{P_1, \dots, P_m\}$ for some $1 \le m \le 4n/\log^5 n$ be the family of Berge paths guaranteed by Claim 4.22. For every $j \in [m]$ we denote the endpoints of P_j by x_j and y_j. Moreover, let $\{s_1, \dots, s_{m'}\} = S$ for some $0 \le m' \le t$. We now apply the connecting lemma (Lemma 4.9) with $\mathcal{A} := (Z_1, \dots, Z_{8\log n})$ and $R := B'$ and the family Ω consisting of the pairs

$$(y_1, x_2), \dots, (y_{m-1}, x_m), (y_m, s_1), (s_1, s_2), \dots, (s_{m'-1}, s_{m'}), (s_{m'}, x_Q), (y_Q, x_1).$$

We can do this since the number of pairs equals $m + m' + 1 \le \varepsilon|Z_i|/4$ for every $i \in [\ell]$ and $|B'| \ge \varepsilon n/8$. For every pair $(u, v) \in \Omega$ we then obtain a Berge path $P_{(u,v)}$ such that for every other pair $(u', v') \in \Omega$ the following holds:

(O1) $P_{(u,v)}$ connects u and v,

(O2) $V^*(P_{(u,v)}) \cap V^*(P_{(u',v')}) = \{u, v\} \cap \{u', v'\}$,

(O3) $E(P_{(u,v)}) \cap E(P_{(u',v')}) = \varnothing$, and

(O4) $E(P_{(u,v)}) \subseteq E_H(V(\Omega) \cup Z, Z^{(r-1)}) \cup E_H(V(\Omega) \cup Z, Z, (B')^{(r-2)})$.

Finally, set $M = Z \setminus \bigcup_{(u,v)\in\Omega} V^*(P_{(u,v)})$. We claim that the following weak cycle defines a Berge cycle in H:

$$C := P_1 \cdot P_{(y_1,x_2)} \cdot P_2 \cdot \ldots \cdot P_{m-1} \cdot P_{(y_{m-1},x_m)} \cdot P_m \cdot P_{(y_m,s_1)} \cdot P_{(s_1,s_2)} \cdot \ldots \cdot P_{(s_{m'-1},s_{m'})} \cdot P_{(s_{m'},x_Q)} \cdot Q_M \cdot P_{(y_Q,x_1)}.$$

By construction we have that C is a weak cycle. Furthermore, observe that $V^*(C) = V(H)$ holds since $W \cup (Y \setminus V^*(Q)) = \bigcup_{i\in[m]} V^*(P_i) \cup S$ and $V^*(Q) \subseteq V^*(Q_M)$ and $Z \subseteq \bigcup_{(u,v)\in\Omega} P_{(u,v)} \cup V^*(Q_M)$. It remains to show that C is Berge. All Berge paths of the family $\{P_i\}_{i\in[m]} \cup \{Q_M\}$ are pairwise edge-disjoint according to Claim 4.22. The Berge paths $\{P_{(u,v)}\}_{(u,v)\in\Omega}$ are pairwise edge-disjoint by Property (O3). For every $i \in [m]$ and $(u,v) \in \Omega$ we have that P_i and $P_{(u,v)}$ are edge-disjoint since every hyperedge of $P_{(u,v)}$ intersects Z, while no hyperedge of P_i does. Moreover, Q_M is edge-disjoint from $P_{(u,v)}$ for every $(u,v) \in \Omega$ since every hyperedge of $P_{(u,v)}$ intersects Z but does not belong to $E_H(Z, A^{(r-1)}) \cup E(Z, A, R_1^{(r-2)})$, while every hyperedge in Q_M that intersects Z needs to belong to $E_H(Z, A^{(r-1)}) \cup E(Z, A, R_1^{(r-2)})$. As a consequence, C is a Hamilton Berge cycle in H.

Hence, in order to finish the proof, it remains to prove Claim 4.22.

Proof of Claim 4.22. By construction of the sets S_j, for every $j \in [t]$, we have

$$\left(1 - \frac{4\varepsilon}{5}\right)\frac{n}{\log^5 n} \le |W_j| \le |S_j| \le \left(1 - \frac{3\varepsilon}{5}\right)\frac{n}{\log^5 n} + \frac{\varepsilon n}{\log^5 n} \le \left(1 + \frac{2\varepsilon}{5}\right)\frac{n}{\log^5 n}.$$

We now define the bucket sets that we use when constructing the desired family of Berge paths. For the first $t/2$ clusters of $\bigcup_{j\in[t]} S_j$ we use R_2 as a bucket set and for the other $t/2$ clusters we use A. We use the following general notation: For every $j \in [t-1] \setminus \{t/2\}$, we set

$$L_j = \begin{cases} R_2 & \text{if } j < t/2 \\ A & \text{otherwise.} \end{cases}$$

Moreover, we define the following set of hyperedges for every $j \in [t-1] \setminus \{t/2\}$:

$$\mathcal{E}_j := E_H(S_j, W_{j+1}^{(r-1)}) \cup E_H(W_j^{(r-1)}, S_{j+1}) \cup E_H(S_j, W_{j+1}, L_j^{(r-2)}) \cup E_H(W_j, S_{j+1}, L_j^{(r-2)}).$$

For all pairs $\{a,b\}$ of distinct vertices $a, b \in V(H)$ we define $\mathcal{E}_j(\{a,b\}) = \{e \in \mathcal{E}_j : \{a,b\} \subseteq e\}$. We now consider the bipartite graph $G_j := (S_j \cup S_{j+1}, E_j)$ with

$$E_j := \big\{\{a,b\} : a \in S_j,\ b \in S_{j+1} \text{ and } \mathcal{E}_j(\{a,b\}) > 0\big\}.$$

We claim that there exist Berge paths as required by the claim if for every $j \in [t-1] \setminus \{t/2\}$ there is a perfect matching in G_j. Indeed, assume first that we found such perfect matchings. Let G be the union of all graphs G_j, then the union of all these perfect matchings induces a collection $\{P'_1, \ldots, P'_T\}$ of $T \le 4n/\log^5 n$ pairwise edge-disjoint paths in G that cover the whole vertex set $\bigcup_{j\in[t]} S_j$. These paths naturally correspond to weak paths in H in the following way: For each path $P'_i = (v_1, e_1, v_2, \ldots, e_{r_i-1}, v_{r_i})$ in G, with $i \in [T]$ and $r_i \in \mathbb{N}$, we can fix a weak path $P_i = (v_1, f_1, v_2, \dots, f_{r_i-1}, v_{r_i})$ with $f_j \in \mathcal{E}_{j'}(e_j)$ where j' satisfies $v_j \in S_{j'}$. We show that the paths $\{P_i\}_{i\in[T]}$ fulfil all requirements asked by the claim.

First of all, P_i is a Berge path for every $i \in [T]$, which can be seen as follows: Let $P'_i = (v_1, e_1, v_2, \dots, e_{r_i-1}, v_{r_i})$ and $P_i = (v_1, f_1, v_2, \dots, f_{r_i-1}, v_{r_i})$ as above and assume for a contradiction that there exist distinct indices $j_1, j_2 \in [r_i - 1]$ with $f_{j_1} = f_{j_2}$. Let j'_1, j'_2 denote the indices with $f_{j_1} \in \mathcal{E}_{j'_1}(e_{j_1})$ and $f_{j_2} \in \mathcal{E}_{j'_2}(e_{j_2})$ and $e_{j_1} \neq e_{j_2}$. Notice that we must have $e_{j_1} \cup e_{j_2} \subseteq f_{j_1}$. We assume without loss of generality that $j'_1 < j'_2$.

Assume first that $f_{j_1} \in E_H(S_{j'_1}, W^{(r-1)}_{j'_1+1})$. This implies $\varnothing \neq e_{j_1} \cap S_{j'_1} = f_{j_1} \cap S_{j'_1} = f_{j_2} \cap S_{j'_1} = e_{j_2} \cap S_{j'_1}$ and $e_{j_1} \cap W_{j'_1+1}, e_{j_2} \cap W_{j'_1+1} \neq \varnothing$, which forces both e_{j_1} and e_{j_2} to be edges of the perfect matching of $G_{j'_1}$, a contradiction.

Similarly, we get a contradiction if we assume that $f_{j_1} \in E_H(S^{(r-1)}_{j'_1}, W_{j'_1+1})$ or $f_{j_2} \in E_H(S_{j'_2}, W^{(r-1)}_{j'_2+1})$ or $f_{j_2} \in E_H(S^{(r-1)}_{j'_2}, W_{j'_2+1})$.

So, let us assume that

$$f_{j_1} \in E_H\big(S_{j'_1}, W_{j'_1+1}, L^{(r-2)}_{j'_1}\big) \cup E_H\big(W_{j'_1}, S_{j'_1+1}, L^{(r-2)}_{j'_1}\big)$$

and

$$f_{j_2} \in E_H\big(S_{j'_2}, W_{j'_2+1}, L^{(r-2)}_{j'_2}\big) \cup E_H\big(W_{j'_2}, S_{j'_2+1}, L^{(r-2)}_{j'_2}\big).$$

If $j'_1, j'_2 < t/2$ or $j'_1, j'_2 > t/2$, we get a contradiction by a similar argument as before. Hence we assume that $j'_1 < t/2$ and $j'_2 > t/2$. Since $j'_1 < t/2$ we have $|f_{j_1} \cap R_1| \geq 1$. Since $j'_2 > t/2$ we have $L_{j'_2} = A$ and $|f_{j_2} \cap (R_2 \cup B)| = 2$ and hence in particular $f_{j_2} \cap R_1 = \varnothing$, a contradiction to the assumption $f_{j_1} = f_{j_2}$.

Next observe that $V(P_i) \cap Z = \varnothing$ for every $i \in [T]$ since $e \cap Z = \varnothing$ for every $e \in \mathcal{E}_j$ and $j \in [t-1] \setminus \{t/2\}$, which gives Property (3) of the claim.

Let $i, i' \in [T]$ be distinct indices. We have $V^*(P_i) \cap V^*(P_{i'}) = V(P'_i) \cap V(P'_{i'}) = \varnothing$ by the construction of the weak paths in H and the disjointness of the paths $\{P'_i\}_{i\in[T]}$ in G. Assume for a contradiction that there exists a hyperedge $e \in E(P_i) \cap E(P_{i'})$. Then, by the definition of the sets $\mathcal{E}_j$ we know that there needs to be an index $j \in [t]$ such that $|e \cap S_j| = 1$ and thus $S_j \cap V^*(P_i) = e \cap S_j = S_j \cap V^*(P_{i'})$, a contradiction to $V^*(P_i) \cap V^*(P_{i'}) = \varnothing$. Hence $E(P_i) \cap E(P_{i'}) = \varnothing$. Therefore, Property (1) of the claim is fulfilled.

For every $M \subseteq Z$ and $i \in [T]$ we have $V^*(P_i) \cap V^*(Q_M) = \varnothing$ since $V^*(P_i) \subseteq W \cup \big(Y \setminus V^*(Q)\big)$ and $V^*(Q_M) = V^*(Q) \cup M \subseteq V^*(Q) \cup Z$ and $V^*(Q) \subseteq Y$.

Finally, let $M \subseteq Z$ and $i \in [T]$. We aim to prove that $E(P_i) \cap E(Q_M) = \varnothing$. Let $e \in E(P)$ be any hyperedge. Assume first that $e \in \mathcal{E}_j$ with $j < t/2$. Then we have $|e \cap W| \geq r - 1$. Since for every hyperedge $e' \in E(Q_M)$ it holds that $|e' \cap W| \leq r - 2$ by Property (Q2), we have $e \notin E(Q_M)$.

Assume now that $e \in \mathcal{E}_j$ with $j > t/2$. Then we have $|e \cap W| \geq r - 1$ or $e \in E_H(S_j, W_{j+1}, A^{(r-2)}) \cup E_H(W_j, S_{j+1}, A^{(r-2)})$. In the first of the two cases, it holds that $e \notin E(Q_M)$ by the same argument as above. Thus suppose that

$$e \in E_H(S_j, R_2, A^{(r-2)}) \cup E_H(R_2, S_{j+1}, A^{(r-2)}). \tag{4.3}$$

This means in particular that

$$\begin{aligned} e \notin\, & E_H(A) \cup E_H(A^{(2)}, R_1^{(r-2)}) \cup E_H(M, A^{(r-1)}) \cup E_H(M, A, R_1^{(r-2)}) \\ & \cup E_H(B) \cup E_H(B^{(2)}, R_2^{(r-2)}) \cup E_H(A, B^{(r-1)}). \end{aligned}$$

Hence, by Property (Q2) it must hold that $e \in E_H(A, B, R_2^{(r-2)})$. Let $\{b\} = e \cap B$. Then we have $b \in S_j \cup S_{j+1}$ by (4.3). Observe that since $e \cap Z = \varnothing$, we have $e \in E(Q)$ and hence $b \in V^*(Q)$, a contradiction to $S_j \cup S_{j+1} \subseteq (W \cup Y) \setminus V^*(Q)$. Therefore $e \notin E(Q_M)$, which gives Property (P3).

Hence it remains to prove the existence of a perfect matching in G_j for every $j \in [t-1] \setminus \{t/2\}$. We show that for every $C \subseteq S_j$ of size $|C| \leq |S_j|/2$ we have $|N_{G_j}(C)| > |C|$. By symmetry and since $|S_{j'}| = |S_{j''}|$ for every $j, j'' \in [t]$, the same argument works if $C \subseteq S_{j+1}$. Using Hall's condition a perfect matching is then guaranteed to exist in the graph G_j.

So, let $C \subseteq S_j$ be a subset of size $|C| \leq |S_j|/2$ and recall that $(C \cup N_{G_j}(C)) \cap L_j = \varnothing$. Assume for a contradiction that $|N_{G_j}(C)| \leq |C|$. Then there exists a set $D \subseteq S_{j+1}$ with $|D| = |C|$ and $N_{G_j}(C) \subseteq D$. We distinguish two cases.

Case 1: Assume that $|C| \leq n/(10^r \log^5 n)$. Then

$$e_H\big(C, N_{G_j}(C), L_j^{(r-2)}\big) \geq e_H\big(C, W_{j+1}, L_j^{(r-2)}\big) \overset{(P3)}{\geq} |C| \left(\frac{1}{2^{r-1}} + \frac{\gamma}{2}\right) p|W_{j+1}| \binom{|L_j|}{r-2}.$$

On the other hand we get

$$e_H\big(C, N_{G_j}(C), L_j^{(r-2)}\big) \leq e_H(C, D, L_j^{(r-2)}) \overset{(H2)}{\leq} |C|^2 \binom{|L_j|}{r-2} p + \varepsilon' \frac{|C| n^{r-1} p}{\log^5 n},$$

which leads to a contradiction since $|C| \leq n/(10^r \log^5 n) \leq \frac{1}{2^{r-1}} \left(1 - \frac{4\varepsilon}{5}\right) \frac{n}{\log^5 n} \leq \frac{1}{2^{r-1}} |W_{j+1}|$ and by choice of ε'.

Case 2: Assume that $n/(10^r \log^5 n) < |C|$. Then

$$e_H(C, N_{G_j}(C)^{(r-1)}) \geq e_H(C, W_{j+1}^{(r-1)}) \overset{(P2)}{\geq} |C| \cdot \left(\frac{1}{2^{r-1}} + \frac{\gamma}{2}\right) p \binom{|W_{j+1}|}{r-1},$$

and

$$e_H(C, N_{G_j}(C)^{(r-1)}) \leq e_H(C, D^{(r-1)}) \overset{(H1)}{\leq} (1+\varepsilon') p |C| \binom{|D|}{r-1}.$$

This leads to a contradiction with $|C| = |D| \leq \left(1 + \frac{2\varepsilon}{5}\right) \frac{n}{2 \log^5 n} \leq \frac{1+2\varepsilon}{2} |W_{j+1}|$ and by the choices of ε and ε'. □

This finishes the proof of Theorem 4.7. □

Finally, we are in the position to present the proof of Theorem 4.1.

Proof of Theorem 4.1. Let $r \geq 3$ and $\gamma > 0$ as well as $p \geq \log^{8r} n / n^{r-1}$ be given. Let $\mathcal{H} = H^{(r)}(n, p)$. Set $\varepsilon' = (10^{-3r} \gamma^2)^r$. By Lemma 4.6 we know that $H^{(r)}(n, p)$ is a.a.s. (p, ε')-pseudorandom. We condition on this event assuming that $\mathcal{H}$ is (p, ε')-pseudorandom. Let $H \subseteq \mathcal{H}$ be any spanning subhypergraph with minimum vertex degree at least $(1/2^{r-1} + \gamma) p \binom{n}{r-1}$. Then, by Theorem 4.7 it holds that H contains a Hamilton Berge cycle. □

4.2 Weak and Berge Hamiltonicity in dense hypergraphs

In this subsection we investigate the local resilience of complete r-uniform hyergraphs with respect to containing weak Hamilton cycles and also with respect to Berge Hamiltonicity.

Setting $p = 1$ in Theorem 4.1 yields that for every $\gamma > 0$ and sufficiently large n, every spanning subhypergraph of $K_n^{(r)}$ with minimum vertex degree at least $\left(\frac{1}{2^{r-1}} + \gamma\right)\binom{n}{r-1}$ contains a Hamilton Berge cycle and hence also a weak Hamilton cycle. Considering the following cases suggests that the bound given by Theorem 4.1 might not be optimal for weak Hamilton cycles: For even n, the disjoint union of two copies of the complete r-uniform hypergraph $K_{n/2}^{(r)}$ on $n/2$ vertices has minimum vertex degree $\binom{n/2-1}{r-1}$ but is disconnected. For odd n, the hypergraph H on n vertices that is the composition of two copies of $K_{\lceil n/2\rceil}^{(r)}$ sharing one vertex satisfies $\delta_1(H) = \binom{\lceil n/2\rceil-1}{r-1}$ but does not contain a weak Hamilton cycle. In fact, the following proposition assures that a minimum degree strictly larger than $\binom{\lceil n/2\rceil-1}{r-1}$ already suffices to guarantee a weak Hamilton cycle in an r-uniform hypergraph on n vertices.

Proposition 4.23. *Let $r \geq 3$ and $n \geq r$ and let H be an r-uniform hypergraph on n vertices. If $\delta_1(H) > \binom{\lceil n/2\rceil-1}{r-1}$, then H contains a weak Hamilton cycle.*

Proof of Proposition 4.23. Let $r \geq 3$ and let $H = (V, E)$ be an r-uniform hypergraph on $n \geq r$ vertices with $\delta_1(H) > \binom{\lceil n/2\rceil-1}{r-1}$. Consider the underlying graph $G_H = (V, F_H)$ where

$$F_H := \big\{\{x, y\} : \exists e \in E \text{ such that } \{x, y\} \subseteq e\big\}.$$

Then, $\delta(G_H) \geq \lceil n/2\rceil$ as otherwise there would be a vertex $v \in V$ with vertex degree at most $\binom{\lceil n/2\rceil-1}{r-1}$ in H. Dirac's theorem (Theorem 1.1) implies that G_H contains a Hamilton cycle. By construction of G_H this Hamilton cycle corresponds to a weak Hamilton cycle in H. □

Observe that for $r \geq 3$ an r-uniform hypergraph on r vertices with one hyperedge is weak Hamiltonian. This is not true for Hamilton Berge cycles since a hypergraph on n vertices with a Berge Hamilton cycle must have at least n hyperedges. Hence, in order to guarantee a Berge Hamilton cycle, the lower bound on the minimum vertex degree must be larger than the one in Proposition 4.23. The following proposition states that the lower bound $\binom{\lceil n/2\rceil-1}{r-1} + n - 1$ is sufficient.

Proposition 4.24. *Let $r \geq 3$ and let H be an r-uniform hypergraph on $n > 2r - 2$ vertices. If $\delta_1(H) \geq \binom{\lceil n/2\rceil-1}{r-1} + n - 1$ then H contains a Hamilton Berge cycle.*

In the proof of Proposition 4.24 we follow the proof idea of Dirac's theorem [63].

Proof of Proposition 4.24. Let $r \geq 3$ and let $H = (V, E)$ be an r-uniform hypergraph on $n > 2r - 2$ vertices with $\delta_1(H) \geq \binom{\lceil n/2\rceil-1}{r-1} + n - 1$. Let $P = (v_1, e_1, v_2, \ldots, e_{k-1}, v_k)$ be a longest Berge path in H.

For every $v \in V$ we define $E'(v) = \big\{e \in E \setminus \{e_1, \ldots, e_k\} : v \in e\big\}$. The condition on the minimum vertex degree implies that we have $|E'(v_1)|, |E'(v_k)| \geq \binom{\lceil n/2\rceil-1}{r-1}$. Since P is a longest Berge path, it holds for every $e \in E'(v_1)$ that $e \subseteq V^*(P)$. The same is true for v_k.

We claim that there exist distinct hyperedges $e \in E'(v_1)$ and $e' \in E'(v_k)$ as well as an index $i \in [k]$ such that $v_{i+1} \in e \cap V^*(P)$ and $v_i \in e' \cap V^*(P)$. Assume for a contradiction that this is not true. Then there exists a subset $S \subseteq V^*(P)$ such that $|S| \leq \lfloor k/2\rfloor$ with

$f \subseteq S$ for every $f \in E'(v_1)$ or with $f' \subseteq S$ for every $f' \in E'(v_k)$. Suppose that $f \subseteq S$ for every $f \in E'(v_1)$. Then $\delta_1(v_1) \leq \binom{|S|-1}{r-1} + k - 1 \leq \binom{\lfloor n/2 \rfloor - 1}{r-1} + n - 1$, which is a contradiction. Hence there exist $e \in E'(v_1)$ and $e' \in E'(v_k)$ with the claimed property. Let $C = (v_1, e, v_{i+1}, e_{i+1}, v_{i+2}, \ldots, v_k, e', v_i, e_{i-1}, \ldots, e_1, v_1)$ be the Berge cycle that can be constructed from P using e and e'. If $k = n$, then C is a Hamilton Berge cycle and we are done. Otherwise we get a contradiction as follows. Let $(v'_1, e'_1, v'_2, \ldots, v'_k, e'_k, v'_1) := C$.

Suppose that $V \setminus V(C) \neq \varnothing$. Due to the large minimum vertex degree, it can be seen quickly that H is connected. Therefore there exist a vertex $v \in V \setminus V(C)$ and a hyperedge $e \in E$ such that $v \in e$ and $e \cap V(C) \neq \varnothing$. Let $e'_i \in \{e'_1, \ldots, e'_k\}$ such that $e \cap e'_i \neq \varnothing$. Let $v' \in e \cap e'_i$. We may assume without loss of generality that $v' \neq v'_{i+1}$. Then $P' := (v, e, v', e'_i, v'_{i+1}, e'_{i+1}, \ldots, e'_{k+1}, v'_1, e'_1, \ldots, e'_{i-2}, v'_{i-1})$ is a Berge path of length k, a contradiction to the assumption that P is a longest Berge path.

Suppose now that $V = V(C)$. Let $e \in E \setminus E(C)$ be any hyperedge such that e intersects $V \setminus V^*(C)$. Such a hyperedge exists due to the assumption on the minimum vertex degree. By choice of e and since $V = V(C)$, there exists an index $j \in [k]$ and distinct vertices x and $y \in e'_j$ such that $x \notin V^*(C)$ and $\{x, y\} \subseteq e$. We may assume without loss of generality that $y \neq v'_{j+1}$. Then $P' := (x, e, y, e'_j, v'_{j+1}, e'_{j+1}, \ldots, v'_k, e'_{k+1}, v'_1, \ldots, e'_{j-2}, v'_{j-1})$ is a Berge path of length k, again a contradiction. □

4.3 Positional games

In this section we discuss a relation between local resilience results of random r-uniform hypergraphs and biased Maker-Breaker games played on the edge set of complete r-uniform hypergraphs. This allows us to apply Theorem 4.1 in order to obtain a bound on the threshold bias for games, where Maker wins if he builds a Hamilton Berge cycle. Then we investigate strict and monotone biased Avoider-Enforcer games, where Avoider wins if by the end of the game his hypergraph is a Berge-acyclic hypergraph with at most one additional hyperedge.

4.3.1 Local resilience of hypergraphs and Maker-Breaker games

Although Maker-Breaker games played on the edge set of complete graphs have been extensively studied, to the best of our knowledge there are no known results on Maker-Breaker games played on the edge set of complete uniform hypergraphs. Following the proof of Theorem 1.10 by Ferber, Krivelevich, and Naves [77], one can show that local resilience results of random r-uniform hypergraphs imply bounds on the threshold biases of the corresponding Maker-Breaker games played on $E(K_n^{(r)})$. More precisely, the following analogue theorem holds.

Theorem 4.25. *For every integer $r \geq 3$, and real $0 < \varepsilon \leq 1/100$ the following holds if n is sufficiently large. Let $p = p(n) \in (0, 1)$ and let $\mathcal{P}$ be a monotone increasing graph property such that $H^{(r)}(n, p)$ has a.a.s. local resilience at least ε with respect to $\mathcal{P}$. Then Maker has a winning strategy in the $(1 : \lceil \frac{\varepsilon}{10rp} \rceil)$ Maker-Breaker game $\big(E(K_n^{(r)}), \mathcal{P}\big)$.*

The proof of Theorem 4.25 is almost identical to the proof of Theorem 1.10. The main difference lies in the additional game that Maker simulates. This game is a so called MinBox game, which is defined as follows. A MinBox(n', D, α, b') game is an $(1 : b')$ Maker-Breaker game played on a family of n' disjoint sets, called *boxes*, each having size at least D. Maker's goal in $\mathrm{MinBox}(n', D, \alpha, b')$ is to claim at least $\alpha|F|$ elements from every box F. In the proof of Theorem 4.25 the game $\mathrm{MinBox}\big(n, 4\delta_1(K_n^{(r)}), p/2, rb_r\big)$ is simulated instead of the game $\mathrm{MinBox}\big(n, 4\delta(K_n), p/2, 2b_2\big)$, which is simulated in the proof of Theorem 1.10, where b_r and b_2 are the biases of the Maker-Breaker games in the statements of Theorem 4.25 and Theorem 1.10, respectively. Since the rest of the proof works verbatim, we only explain Maker's strategy roughly. In particular we do not repeat any details and calculations of the proof but rather refer to the proof of Theorem 1.10 in [77].

The idea of the proof of Theorem 4.25 is to provide Maker with a random strategy that ensures him to win asymptotically almost surely. This implies in particular that Maker has a deterministic winning strategy. The random strategy of Maker is to generate a random hypergraph $H = H^{(r)}(n,p)$ by successively exposing a hyperedges of $K_n^{(r)}$ and assigning each exposed hyperedge to H with probability p. Whenever Maker exposes a hyperedge that was not yet claimed by Breaker, Maker inserts the hyperedge into his hypergraph. In the end, Maker's hypergraph is a subgraph of H that contains all hyperedges of H except for those that were taken by Breaker. Thus the goal is then to show that a.a.s. at most an ε-proportion of the incident hyperedges at every vertex of H were claimed by Breaker.

Let us discuss some more details on Maker's turns. Maker pretends to additionally play a $\mathrm{MinBox}\big(n, 4\delta_1(K_n^{(r)}), p/2, rb_r\big)$ game, where each box F_x corresponds to a vertex $x \in V(K_n^{(r)})$. Maker's strategy is divided into two stages, where it is shown that a.a.s. all hyperedges of $K_n^{(r)}$ are exposed before Maker reaches the second stage.

In each turn during the first stage, for each hyperedge $\{x_1, \dots, x_r\}$ that was claimed by Breaker in the previous move, Maker pretends that Breaker has claimed one free element from each of the boxes $F_{x_1}, \dots, F_{x_r}$. Then Maker chooses a vertex x for which F_x still contains free elements and is active (which means that Maker has claimed less than $p|F_x|/2$ elements from F_x so far) and such that the difference between the number of elements taken by Breaker from F_x and the number of elements taken by Maker from F_x multiplied by rb_r is largest among all choices of x.

Next Maker successively exposes hyperedges incident to x until either he exposes a hyperedge that is assigned to H or there are no unexposed hyperedges incident to x left. In the latter case, Maker declares the current turn as a failure of type I and claims $\lceil p|F_x|/2\rceil$ free elements from F_x. In the other case, if the last exposed hyperedge already belongs to Breaker, Maker claims one free element of F_x, skips his move in the original game, and declares the current turn as a failure of type II. If, on the other hand, this last exposed hyperedge, say $\{x_1, \dots, x_r\}$, is still free, Maker claims $\{x_1, \dots, x_r\}$ in the original game as well as an element from F_{x_i} for every $i \in [r]$.

The proof of Theorem 4.25 then comes down to proving an asymptotically almost sure upper bound on the number of failures of type II. As already mentioned above, for more details and calculations we refer to the proof of Theorem 1.10 by Ferber, Krivelevich, and Naves [77].

Using Theorem 4.1, we obtain as an immediate consequence of Theorem 4.25 the following bound on the threshold bias of the $(1:b)$ Maker-Breaker game played on $E(K_n^{(r)})$, where Breaker wins if his hypergraph contains a Hamilton Berge cycle.

Corollary 4.26. *For every $r \geq 3$ and sufficiently large n, Maker has a winning strategy in the $(1:b)$ Berge Hamiltonicity game played on $E(K_n^{(r)})$ if $b \leq n^{r-1}/(1000r \log^{8r} n)$.*

Proof. By Theorem 4.1 we know that $H^{(r)}(n,p)$ has a.a.s. local resilience at least $1/100$ with respect to containing a Hamilton Berge cycle whenever $p \geq \log^{8r} n/n^{r-1}$. Therefore, Theorem 4.25 assures that for sufficiently large n, Maker has a winning strategy for the $(1 : n^{r-1}/(1000r \log^{8r} n))$ Berge Hamiltonicity game played on the edge set of $K_n^{(r)}$. □

4.3.2 Avoiding Berge cycles in Avoider-Enforcer games

Turning to misère versions of the Maker-Breaker games studied in the previous subsection, we are interested in Avoider-Enforcer games played on the edge set of complete uniform hypergraphs, where the first player, which is now Avoider, has to avoid Berge cycles in his hypergraph rather than aiming to build a Hamilton Berge cycle. We restrict the problem to 3-uniform hypergraphs as the calculations in our proofs would be more involved for higher uniformities. However, the proof ideas should also work for higher uniformities.

The first theorem below gives a lower bound of $3000n^2 \log^2 n$ on the bias such that both in the monotone and in the strict $(1 : b)$ Avoider-Enforcer game, by the end of the game Avoider's hypergraph is a Berge-acyclic hypergraph with at most one additional hyperedge.

Theorem 4.27. *For n sufficiently large and $b \geq 3000n^2 \log^2 n$, Avoider can ensure that in the monotone as well as in the strict $(1 : b)$ Avoider-Enforcer game played on $E(K_n^{(3)})$ by the end of the game Avoider's hypergraph is a Berge-acyclic hypergraph with at most one additional hyperedge.*

The next theorem concerns the strict $(1 : b)$ Avoider Enforcer game played on the edge set of the complete 3-uniform hypergraph, where Avoider's task is to keep his graph Berge-acyclic. In this game he has a winning strategy for some bias b between $3000n^2 \log^2 n$ and $3001n^2 \log^2 n$.

Theorem 4.28. *For n sufficiently large, there is a bias $3000n^2 \log^2 n \leq b \leq 3001n^2 \log^2 n$ such that Avoider can ensure that in the strict $(1 : b)$ Avoider-Enforcer game played on $E(K_n^{(3)})$ Avoider's hypergraph is Berge-acyclic by the end of the game.*

Let us mention at this point that Theorems 4.27 and 4.28 are hypergraph analogues of the following two theorems, which were proved in joint work with Dennis Clemens, Yury Person, and Tuan Tran [55].

Theorem 4.29 ([55])**.** *For n sufficiently large and $b \geq 200n \log n$, Avoider can ensure that in the monotone as well as in the strict $(1 : b)$ Avoider-Enforcer game played on $E(K_n)$ by the end of the game Avoider's graph is a forest with at most one additional edge.*

The second theorem provides the existence of a bias between $200n \log n$ and $201n \log n$ for which Avoider can keep his graph acyclic.

Theorem 4.30 ([55])**.** *For n sufficiently large, there is a bias $200n \log n \leq b \leq 201n \log n$ such that Avoider can ensure that in the strict $(1 : b)$ Avoider-Enforcer game played on $E(K_n)$ Avoider's graph is a forest by the end of the game.*

We start with the proof of Theorem 4.27, which is an extension to hypergraphs of the proof of Theorem 4.29.

Proof of Theorem 4.27. Let n be large enough and let $b \geq 3000n^2 \log^2 n$. In the following we will provide Avoider with a strategy that ensures that by the end of the game Avoider's graph is Berge-acyclic plus at most one additional hyperedge.

Let t be the smallest integer with

$$n \left(\frac{t+1}{10 \log n} \right)^t < 3. \tag{4.4}$$

An easy calculation shows that $t = \Theta(\log n)$, in particular, we have for large n that

$$t < \log n/3. \tag{4.5}$$

To succeed, Avoider will play according to t stages in increasing order and each stage consists of several consecutive rounds where it is possible that a stage lasts zero rounds, i.e. that a stage does not occur at all. In the first $t-1$ stages, Avoider always claims exactly one hyperedge in each round, connecting three components of his hypergraph such that the sum of their sizes is minimal (whenever we talk about components, we mean the components of Avoider's hypergraph). In the last stage, which will be shown to last at most one round, Avoider will claim an arbitrary further hyperedge. We refer to hyperedges, neither taken by Avoider nor by Enforcer, as *unclaimed* hyperedges.

Starting with Stage 1, Avoider plays according to the following rules.

Stage k (for $k \in [t-1]$). If there exists an unclaimed hyperedge e between three components T_1, T_2, and T_3 with $\sum_{i\in[3]} |V(T_i)| = k+1$, Avoider claims such a hyperedge, thus creating a component on the vertex set $V(T_1) \cup V(T_2) \cup V(T_3)$. Then it is Enforcer's turn and the round is over.

Avoider is going to play according to Stage k in the next round as well. If there is no such hyperedge e to be claimed at Stage k, Avoider proceeds with Stage $k+1$. (As mentioned above it might happen that there is no hyperedge to be claimed at Stage k already when Avoider enters Stage k. In that case, this stage lasts zero rounds, and Avoider immediately proceeds with Stage $k+1$.)

Stage t. In every further round, Avoider claims exactly one arbitrary free hyperedge.

One can easily verify that Avoider can follow the strategy. Moreover, as long as Avoider plays according to the strategy of the first $t-1$ stages, his graph remains Berge-acyclic. Thus, in order to show that the above described strategy is indeed a winning strategy, it remains to show that the last stage lasts at most one round. As a first step we aim to bound the number of rounds a given stage lasts. Let n_k denote the number of rounds in Stage $k-1$. Observe that Avoider creates components of size exactly k only in this stage. Thus, the number of such components is always bounded from above by n_k.

Claim 4.31. *For every $k \leq t$,*

$$n_k \leq n \left(\frac{k}{10 \log n} \right)^{k-1}.$$

Proof. The claim is obviously true for $k=1$. So, let $k > 1$ and we proceed by induction. When Avoider enters Stage $k-1$ every existing component contains at most $k-1$ vertices and there are no unclaimed hyperedges between any three components T_1, T_2, and T_3 with

$\sum_{i\in[3]}|V(T_i)| \le k-1$. In particular, every unclaimed hyperedge is either between three components T_1, T_2, and T_3 with $\sum_{i\in[3]}|V(T_i)| \ge k$ or between two components T_1' and T_2' each of size at most $k-1$ or within a component, which has size at most $k-1$. The first case contributes at most

$$\sum_{\substack{1\le i\le j\le \ell\le k-1\\ i+j+\ell\ge k}} ij\ell n_i n_j n_\ell$$

unclaimed hyperedges since n_i is an upper bound on the number of components of size exactly i. The second and third cases yield at most $(k-1)n^2+(k-1)^2n$ unclaimed hyperedges by the following reason: Let n_i' denote the number of components of order i immediately after the end of Stage $k-2$. Then, after $k-2$ stages, the number of unclaimed hyperedges between pairs of components is at most

$$\begin{aligned}\sum_{i,j\in[k-1]} i\binom{j}{2}n_i'n_j' = \sum_{j\in[k-1]}\binom{j}{2}n_j'\sum_{i\in[k-1]} in_i' &= \sum_{j\in[k-1]}\binom{j}{2}n_j'n\\ &\le n(k-1)\sum_{j\in[k-1]} jn_j' = (k-1)n^2,\end{aligned}$$

since $\sum_{i\in[k-1]} in_i' = \sum_{j\in[k-1]} jn_j' = n$. Moreover, the number of unclaimed hyperedges within components after $k-1$ stages is at most $\sum_{i\in[k-1]}\binom{i}{3}n_i' \le (k-1)^2\sum_{i\in[k-1]} in_i' = (k-1)^2n$.

Therefore, at the beginning of Stage $k-1$, the number of unclaimed hyperedges is at most

$$\sum_{\substack{1\le i\le j\le \ell\le k-1\\ i+j+\ell\ge k}} ij\ell n_i n_j n_\ell + (k-1)n^2 + (k-1)^2n.$$

Since in each but possibly the last round at least $b+1$ hyperedges are claimed (1 by Avoider and at least b by Enforcer), we conclude

$$n_k \le \frac{1}{b+1}\left(\sum_{\substack{1\le i\le j\le \ell\le k-1\\ i+j+\ell\ge k}} ij\ell n_i n_j n_\ell + (k-1)n^2 + (k-1)^2n\right) + 1. \tag{4.6}$$

Using the induction hypothesis we find an upper bound on the sum $\sum_{\substack{1\le i\le j\le \ell\le k-1\\ i+j+\ell=s}} ij\ell n_i n_j n_\ell$ for $s=k,\ldots,3k-3$ as follows:

$$\sum_{\substack{1\le i\le j\le \ell\le k-1\\ i+j+\ell=s}} ij\ell n_i n_j n_\ell \le \frac{n^3}{(10\log n)^{s-3}}\sum_{\substack{1\le i\le j\le \ell\le s-1\\ i+j+\ell=s}} i^i j^j \ell^\ell.$$

For $s\le 6$, it is easy to check that

$$\sum_{\substack{1\le i\le j\le \ell\le s-1\\ i+j+\ell=s}} i^i j^j \ell^\ell < 5s^{s-1}.$$

On the other hand, for $s\ge 7$ observe that we have for every $2\le i\le s/2$

$$\left(\frac{i}{s}\right)^i \le \left(\frac{2}{s}\right)^2 \tag{4.7}$$

by an easy calculation for $i \le 3$ and since

$$\frac{i^i}{s^{i-2}} \le \frac{i^i}{(2i)^{i-2}} \le \frac{i^2}{2^{i-2}} \le 4$$

for every $i \ge 4$. Therefore, we also obtain for $s \ge 7$

$$\sum_{\substack{1\le i\le j\le \ell\le s-1\\ i+j+\ell=s}} i^i j^j \ell^\ell < \sum_{\substack{1\le j\le \ell\le s-1\\ j+\ell=s-1}} j^j \ell^\ell + \sum_{2\le i\le s/3} i^i s^{s-i} < s^{s-1} + \sum_{2\le j\le s/2} j^j s^{s-j} + \sum_{2\le i\le s/3} i^i s^{s-i}$$
$$= s^{s-1}\left(1 + 2s \sum_{2\le i\le s/2}\left(\frac{i}{s}\right)^i\right) \overset{(4.7)}{\le} s^{s-1}\left(1+2s\sum_{2\le i\le s/2}\left(\frac{2}{s}\right)^2\right) < 5s^{s-1}.$$

Observing that for every $s \ge k$ we have

$$\left(\frac{s}{10\log n}\right)^{s-1} = \left(\frac{k}{10\log n}\right)^{k-1}\prod_{i=1}^{s-k}\frac{k+i-1}{10\log n}\left(1+\frac{1}{k+i-1}\right)^{k+i-1}$$
$$\le \left(\frac{k}{10\log n}\right)^{k-1}\left(\frac{3ke}{10\log n}\right)^{s-k} \overset{(4.5)}{\le} \left(\frac{k}{10\log n}\right)^{k-1} 2^{k-s}, \quad (4.8)$$

we can simplify Equation (4.6) using $b \ge 3000n^2\log^2 n$ and using the above established upper bounds as follows

$$n_k \le \frac{1}{3000n^2\log^2 n}\left(\sum_{s=k}^{3k-3} 500n^3\log^2 n\left(\frac{s}{10\log n}\right)^{s-1} + (k-1)n^2 + (k-1)^2 n\right) + 1$$
$$\overset{(4.8)}{\le} \frac{n}{6}\left(\frac{k}{10\log n}\right)^{k-1}\sum_{s=k}^{3k-3} 2^{k-s} + \frac{k}{3000\log^2 n} + \frac{k^2}{3000n\log^2 n} + 1$$
$$\overset{(4.5)}{\le} \frac{n}{2}\left(\frac{k}{10\log n}\right)^{k-1} + \frac{1}{9000\log n} + \frac{1}{27000n} + 1 \le n\left(\frac{k}{10\log n}\right)^{k-1}.$$

This completes the proof of Claim 4.31. □

Now, analogously to the calculation of the proof of Claim 4.31 it follows that, when Avoider enters the last stage, Stage t, the number of remaining unclaimed hyperedges is bounded by

$$\sum_{\substack{1\le i\le j\le \ell\le t\\ i+j+\ell\ge t+1}} ij\ell n_i n_j n_\ell + tn^2 + t^2 n \le 500n^3\log^2 n\left(\frac{t+1}{10\log n}\right)^t \sum_{s=t+1}^{3t} 2^{t+1-s} + tn^2 + t^2 n$$
$$\overset{(4.4)}{<} 2999n^2\log^2 n + n^2\log n + n\log^2 n < 3000n^2\log^2 n$$

by the choice of t ($t < \log n/3$) and for n sufficiently large. Thus, this last stage lasts at most one round. □

Now we turn to the case of the strict rules, when Enforcer has to claim exactly b hyperedges during each round (except possibly for the last one). The proof is a modification of the proof of Theorem 4.30.

Proof of Theorem 4.28. We claim that for large enough n, there exists b with $3000n^2 \log^2 n \leq b \leq 3001n^2 \log^2 n$ and the remainder of $\binom{n}{3}$ divided by $b+1$ is at least $n^2 \log^2 n$.

Before proving this claim let us explain how the theorem follows then. Let b be given as above. Avoider now plays according to the same strategy as given in the proof of Theorem 4.27 until he reaches Stage t, where again t is the smallest integer with $n\left(\frac{t+1}{10\log n}\right)^t < 3$. At this point, Avoider's graph is still Berge-acyclic and its components are all of size at most t. Analogously to the proof of Theorem 4.27, there can be at most $t^2 n < n\log^2 n/9$ unclaimed hyperedges within components and at most $tn^2 < n^2 \log n/3$ unclaimed hyperedges between pairs of components. However, since the remainder of the division $\binom{n}{2}/(b+1)$ is at least $n^2\log^2 n$, there exist unclaimed hyperedges connecting three distinct components when Avoider enters Stage t (provided n is large enough). Now, Avoider just claims one such hyperedge arbitrarily. His graph remains Berge-acyclic and afterwards, Enforcer must take all remaining hyperedges. Observe that in the case when Avoider is the second player, he does not even claim a hyperedge in the last round.

So, it only remains to prove the above mentioned claim. Let $b_1 = \lceil 3000.5 \cdot n^2 \log^2 n\rceil$. Moreover, let

$$\binom{n}{3} = q_1(b_1+1) + r_1 \quad \text{with} \quad 0 \leq r_1 \leq b_1 \text{ and } q_1 \sim \frac{n}{c\log^2 n} \text{ with } c := 18003.$$

If $r_1 > n^2 \log^2 n$, we are done by setting $b = b_1$. Otherwise, let $b = b_1 - \lceil cn\log^4 n\rceil$. Then

$$\binom{n}{3} = q_1(b+1) + (r_1 + q_1\lceil cn\log^4 n\rceil).$$

Moreover, for large enough n, we obtain $r_1 + q_1\lceil cn\log^4 n\rceil < b$, and therefore the remainder of the division $\binom{n}{3}$ by $(b+1)$ is at least $r_1 + q_1\lceil cn\log^4 n\rceil > n^2\log^2 n$, while $3000n^2\log^2 n \leq b \leq 3001n^2\log^2 n$. □

4.4 Concluding remarks

It is still an open problem what the threshold for the appearance of a Berge Hamilton cycle in random r-uniform hypergraphs is. Theorem 4.1 implies an upper bound of $\log^{8r} n/n^{r-1}$ on this threshold, leaving a polylogarithmic gap to the best-known lower bound (see Theorem 1.9). It is worth mentioning that the proof of Theorem 4.1 works for slightly lower edge probabilities than $\log^{8r} n/n^{r-1}$. However, using our proof method, the exponent of $\log n$ needs to be strictly larger than a multiple of r. Since we do not believe that this is the correct answer, we did not intend to optimise the polylogarithmic factor for the sake of readability.

Having investigated the local resilience of random hypergraphs with respect to Berge Hamiltonicity, a natural problem to study next is the local resilience of random hypergraphs with respect to Hamilton cycles with a different notion of cycles, for instance with the notion of ℓ-cycles. For tight Hamilton cycles such a result seems to be hard to achieve. Indeed, determining the local resilience of the complete r-uniform hypergraph with respect to tight

Hamilton cycles turned out to be challenging. The best-known bound on the minimum vertex degree of an n-vertex 3-uniform hypergraph that implies the containment of a tight Hamilton cycle is $\big((5-\sqrt{5})/3+\varepsilon\big)\binom{n}{2}$, as shown by Rödl and Ruciński [143]. Non-trivial bounds for higher uniformities are not known yet. Rödl and Ruciński [142] posed the following conjecture, which would yield an asymptotically tight minimum vertex degree condition for tight Hamilton cycles.

Conjecture 4.32 (Rödl, Ruciński [142])**.** *For each integer $r \geq 3$ and real $\varepsilon > 0$ there is an integer n_0 such that the following holds. If H is an r-uniform hypergraph on $n \geq n_0$ vertices with*

$$\delta_1(H) \geq \left(1 - \left(1 - \frac{1}{r}\right)^{r-1} + \varepsilon\right)\binom{n}{r-1},$$

then H contains a tight Hamilton cycle.

We remark that while the minimum vertex degree condition for weak Hamilton cycles in Proposition 4.23 is tight, the one that we established for Hamilton Berge cycles in Proposition 4.24 can probably be improved.

Finally, using Theorem 4.1 and Theorem 4.25 we obtained a lower bound on the threshold bias of the $(1:b)$ Maker-Breaker game played on $E(K_n^{(r)})$, where Maker wins if his hypergraph contains a Hamilton Berge cycle. In Subsection 4.3.2 we determined upper bounds on the threshold biases of Avoider-Enforcer games played on $E(K_n^{(3)})$, where Avoider has to keep his hypergraph (almost) Berge-acyclic. We believe that for both games the bounds are not optimal and would be interested in knowing the threshold biases for the games that we studied.

5

Rainbow matchings in multigraphs

A conjecture by Aharoni and Berger suggests that every bipartite multigraph, the edges of which are coloured with n colours such that each colour class induces a matching of size $n+1$, contains a rainbow matching of size n. As elucidated in Subsection 1.2.2, this is a generalisation of famous open conjectures by Ryser and by Brualdi and Stein on Latin squares.

In this chapter we study general multigraphs, the edges of which are coloured with n colours. We prove that if each of the n colour classes covers $3n+o(n)$ vertices of the multigraph and induces a disjoint union of cliques, then there exists a rainbow matching of size n. In the setting above, this implies that matching sizes of $3n/2+o(n)$ suffice to guarantee a rainbow matching of size n. Thus our result marks a step towards the conjecture of Aharoni and Berger. Moreover, it solves an algebraic problem by Grinblat asymptotically.

In Section 5.1 we formulate our main result in terms of rainbow matchings in multigraphs, rainbow matchings in equivalence classes, and using the algebraic terminology of Grinblat's question. We prove the equivalence of these formulations in Section 5.2. The proof of the main result is presented in Section 5.3. Finally, we close this chapter with a discussion of some open problems in Section 5.4.

This chapter is based on joint work with Dennis Clemens and Alexey Pokrovskiy [56] and on joint work with Dennis Clemens [53].

5.1 Introduction

In this chapter we asymptotically affirm a question by Grinblat on sets not belonging to algebras. Before formulating this question, let us recall that $\mathfrak{v}=\mathfrak{v}(n)$ was defined by Grinblat [85] as the minimal cardinal number such that the following is true:

> "Let $\mathcal{A}_1,\dots,\mathcal{A}_n$ be algebras on a set X such that for each $i\in[n]$ there exist at least $\mathfrak{v}(n)$ pairwise disjoint sets in $\mathcal{P}(X)\setminus\mathcal{A}_i$. Then there exists a family $\{U_i^1,U_i^2\}_{i\in[n]}$ of $2n$ pairwise disjoint subsets of X such that, for each $i\in[n]$, if $Q\in\mathcal{P}(X)$ and Q contains one of the two sets U_i^1 and U_i^2 and its intersection with the other one is empty, then $Q\notin\mathcal{A}_i$."

As shown by Grinblat [85], a lower bound on $\mathfrak{v}(n)$ is $3n-2$ for every $n\in\mathbb{N}$. He asked whether this bound is tight for every $n\geq 4$:

Question 5.1 (Grinblat, [85])**.** *Is it true that $\mathfrak{v}(n) = 3n - 2$ for $n \geq 4$?*

By proving the following theorem, we give an asymptotic answer to Question 5.1.

Theorem 5.2. *For every $\delta > 0$ there exists $n_0 = n_0(\delta) = 144/\delta^2$ such that for every $n \geq n_0$ it holds that $\mathfrak{v}(n) \leq (3 + \delta)n$.*

Nivasch and Omri [134] proved the upper bound $\mathfrak{v}(n) \leq 16n/5 + \mathcal{O}(1)$, using the following equivalent definition of $\mathfrak{v}(n)$ in the context of equivalence relations. Let X be a finite set and let A be an equivalence relation on X. If $x, y \in X$ are equivalent under A, we write $x \sim_A y$. By

$$[x]_A := \{y \in X : x \sim_A y\}$$

we denote the *equivalence class* under A of an element $x \in X$, while the *kernel* of A is defined as

$$\ker(A) := \big\{x \in X : |[x]_A| \geq 2\big\}.$$

Following Nivasch and Omri [134], we let $\mathfrak{v}_1(n)$ be the minimal number such that if $A_1, \dots, A_n$ are equivalence relations on X with $|\ker(A_i)| \geq \mathfrak{v}_1(n)$ for each $i \in [n]$, then $A_1, \dots, A_n$ contain a *rainbow matching*, i.e. a set of $2n$ distinct elements $x_1, y_1, \dots, x_n, y_n \in X$ with $x_i \sim_{A_i} y_i$ for each $i \in [n]$. This identity is mainly based on the fact that there is a natural correspondence between algebras and equivalence relations. Using these definitions, it turns out that $\mathfrak{v}(n) = \mathfrak{v}_1(n)$ holds. Indeed, given an equivalence relation A on the set X, we can define the algebra $\mathcal{A} := \big\{\bigcup_{x \in S}[x]_A : \ S \subseteq X\big\}$. Conversely, given some algebra $\mathcal{A}$ on X, one can define the equivalence relation A on X the equivalence classes of which are the inclusion minimal sets in $\mathcal{A}$. A complete argument to show that $\mathfrak{v}(n) = \mathfrak{v}_1(n)$ is presented in Section 5.2.

Thus, using the terminology of Nivasch and Omri [134] Theorem 5.2 is equivalent to the following theorem.

Theorem 5.3. *For every $\delta > 0$ there exists $n_0 = n_0(\delta) = 144/\delta^2$ such that the following holds for every $n \geq n_0$. Let $A_1, \dots, A_n$ be n equivalence relations on a finite set X. If $|\ker(A_i)| \geq (3 + \delta)n$ for each $i \in [n]$, then $A_1, \dots, A_n$ contain a rainbow matching.*

Observe that it would suffice to prove Theorem 5.3 for the case that each equivalence class of $A_1, \dots, A_n$ has size 2 or 3. In the special case when all of these equivalence classes consist of 3 elements, the statement can be easily proved by a greedy argument even for $\delta = 0$.

Theorem 5.3 can be rephrased in the context of graphs. If $A_1, \dots, A_n$ are equivalence relations on a set X, let the vertices of an edge-coloured multigraph be the elements of X and, for each $i \in [n]$, let $\{x, y\} \in \binom{X}{2}$ be an edge of colour i if and only if $x \sim_{A_i} y$. This means that the equivalence relations are represented in this multigraph by colour classes, each of which is the disjoint union of non-trivial cliques, i.e. complete graphs with at least 2 vertices. A matching in an edge-coloured multigraph is called a *rainbow matching* if all its edges have distinct colours. Using this notion, we can reformulate Theorem 5.3 as follows.

Theorem 5.4. *For every $\delta > 0$ there exists $n_0 = n_0(\delta) = 144/\delta^2$ such that the following holds for every $n \geq n_0$. Let G be a multigraph, the edges of which are coloured with n colours. If each subgraph of G induced by a colour class has at least $(3+\delta)n$ vertices and is the disjoint union of non-trivial cliques, then G contains a rainbow matching of size n.*

As a direct consequence of Theorem 5.4 we obtain the following partial result towards the conjecture by Aharoni and Berger (Conjecture 1.13). The corollary constitutes the special case of Theorem 5.4 when the multigraph G is bipartite and thus each clique consists of two vertices:

Corollary 5.5. *For every $\varepsilon > 0$ there exists an integer $n_0 \geq 1$ such that for every $n \geq n_0$ the following holds. Let G be a bipartite multigraph whose edges are coloured with n colours and each colour class induces a matching of size at least $\left(\frac{3}{2} + \varepsilon\right) n$. Then G contains a rainbow matching of size n.*

For an independent direct proof of Corollary 5.5 we refer to the joint work [53] with Dennis Clemens.

In Section 5.3 we prove Theorem 5.4, which automatically provides a proof for Theorems 5.2 and 5.3. As already mentioned, the best-known lower bound on $\mathfrak{v}(n)$ is $3n - 2$ for each $n \geq 4$. Indeed, if all colour classes are identical and are the disjoint union of $n - 1$ triangles, then there is no rainbow matching of size n. Hence, Theorem 5.4 is asymptotically best possible. If $n = 3$, then $\mathfrak{v}(3) = 9 > 3n - 2$ as shown by Grinblat [85]. See Figure 5.1 for the lower bound $\mathfrak{v}(3) \geq 9$, which was also observed by Nivasch and Omri [134].

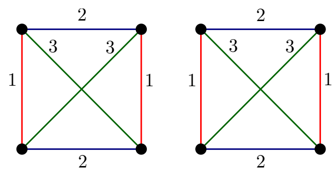

Figure 5.1: Example of a graph with 3 colour classes each of size 8 that has no rainbow matching of size 3.

5.2 Equivalence classes and algebras of sets

In this section we prove that $\mathfrak{v}(n) = \mathfrak{v}_1(n)$ for all $n \in \mathbb{N}$, where $\mathfrak{v}(n)$ and $\mathfrak{v}_1(n)$ are defined as in the previous section.

Proof. First we show that $\mathfrak{v}(n) \geq \mathfrak{v}_1(n)$ holds. Let $A_1, \dots, A_n$ be equivalence relations on a set X with $|\ker(A_i)| \geq \mathfrak{v}(n)$ for each $i \in [n]$. Let $\mathcal{A}_i := \left\{\bigcup_{x\in S}[x]_{A_i} :\ S \subseteq X\right\}$ for each $i \in [n]$. Recall that for a set X' a nonempty subset $\mathcal{A} \subseteq \mathcal{P}(X')$ is an *algebra* on X' if $\mathcal{A}$ is closed under complementation and under unions, i.e. if $M_1, M_2 \in \mathcal{A}$, then $X' \setminus M_1 \in \mathcal{A}$ and $M_1 \cup M_2 \in \mathcal{A}$. It can be easily seen that $\mathcal{A}_1, \dots, \mathcal{A}_n$ are algebras on X.

For each of the at least $\mathfrak{v}(n)$ elements $x \in \ker(A_i)$ it holds that $\{x\} \in \mathcal{P}(X) \setminus \mathcal{A}_i$. In particular, by the definition of $\mathfrak{v}(n)$, we find a family $\{U_i^1, U_i^2\}_{i\in[n]}$ such that if $Q \in \mathcal{P}(X)$ and Q contains one of the two sets U_i^1 and U_i^2 and its intersection with the other one is empty, then $Q \notin \mathcal{A}_i$. For every $i \in [n]$, we now choose $Q_i \in \mathcal{A}_i$ to be the inclusion minimal set satisfying $U_i^1 \subseteq Q_i$, and we note that $U_i^2 \cap Q_i \neq \varnothing$ is implied. By the minimality of Q_i, it turns out that every equivalence class of A_i that is contained in Q_i needs to intersect U_i^1, and thus, there is at least one such class $[z_i]_{A_i}$ intersecting both U_i^1 and U_i^2. Choosing arbitrary elements $x_i \in [z_i]_{A_i} \cap U_i^1$ and $y_i \in [z_i]_{A_i} \cap U_i^2$ for every $i \in [n]$ finally yields a rainbow matching as desired.

Let us now prove that $\mathfrak{v}(n) \leq \mathfrak{v}_1(n)$ holds. For this we need to argue that for every algebras $\mathcal{A}_1, \ldots, \mathcal{A}_n$ on a set X with at least $\mathfrak{v}_1(n)$ pairwise disjoint sets in $\mathcal{P}(X) \setminus \mathcal{A}_i$, for each $i \in [n]$, there is a family $\{U_i^1, U_i^2\}_{i \in [n]}$ as described earlier. To do so, for each $i \in [n]$, we define equivalence relations A_i on X the equivalence classes of which are the inclusion minimal sets in $\mathcal{A}_i$.

As, by the properties of an algebra, for every set $B \in \mathcal{P}(X) \setminus \mathcal{A}_i$ there is at least one element $b \in B$ with $\{b\} \notin \mathcal{A}_i$, we conclude $|\ker(A_i)| \geq \mathfrak{v}_1(n)$, for every $i \in [n]$. Thus, by definition of $\mathfrak{v}_1(n)$, we find a rainbow matching $x_1, y_1, \ldots, x_n, y_n$ as described above. Now, for every $i \in [n]$, let $U_i^1 := \{x_i\}$ and $U_i^2 := \{y_i\}$. Then, whenever $U_i^j \subseteq Q$ holds for some $Q \in \mathcal{P}(X)$ and $j \in \{1, 2\}$ we obtain $[y_i]_{A_i} = [x_i]_{A_i} \subseteq Q$, by definition of A_i, and thus $Q \cap U_i^{3-j} \neq \varnothing$. $\square$

5.3 Rainbow matchings

The aim of this section is to present the proof of Theorem 5.4. Our proof of this theorem will be by induction on n. Rather than proving the theorem directly we will first prove a technical lemma (Lemma 5.7), which is more amenable to induction. To state this lemma rigorously, we first need to introduce some definitions and notation.

5.3.1 Preliminaries

For any edge-coloured multigraph G, we denote by $c(e)$ the colour assigned to the edge $e \in E(G)$. For the sake of simplicity, we call an edge of colour c simply *c-edge*. Let F be a set or a sequence of edges, then we denote by $V(F) := \bigcup_{e \in F} e$ the *vertex set* of F.

Next we define switchings, which, given some rainbow matching M of size k, provide us a new rainbow matching of size k by replacing edges in M with edges in $E(G) \setminus M$. In Figure 5.2 we illustrate a switching of length 3.

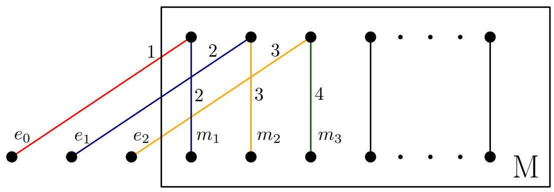

Figure 5.2: A $(1, 4)$-switching of length 3.

Definition 5.6 (Switching). *Let G be an edge-coloured multigraph and let M be a rainbow matching in G. We say that a sequence of edges $\sigma = (e_0, m_1, e_1, m_2, \ldots, e_{k-1}, m_k)$ is a* $(c(e_0), c(m_k))$-switching *of length k with respect to M if for each $i \neq j \in [k]$ we have*

(S1) $m_1, \ldots, m_k$ are distinct edges in M,

(S2) $e_{i-1} \in E_G[m_i, V \setminus V(M)]$,

(S3) $c(e_0) \neq c(m_i)$ *and* $c(e_i) = c(m_i)$*, and*

(S4) $e_{i-1} \cap e_{j-1} = \varnothing$.

Whenever it is clear from the context, we may omit writing explicitly with respect to which matching the considered switching is defined. The length of σ is denoted by $\ell(\sigma)$. Furthermore, we denote by $m(\sigma)$ the set of all edges of σ that are contained in the matching M and by $e(\sigma)$ the set of all other edges of σ. Observe that $\ell(\sigma) = |m(\sigma)| = |e(\sigma)|$.

For every colour c, we also define an *empty* (c, c)-switching σ_c^0. This switching has no edges, starts and ends at the colour c, has length zero, and satisfies $m(\sigma_c^0) = e(\sigma_c^0) = \varnothing$.

5.3.2 Switching lemma

In this subsection we state and prove Lemma 5.7. Roughly speaking, Lemma 5.7 says that given a rainbow matching M in a multigraph G, either there are few edges of colour c touching $V(G) \setminus V(M)$ for some colour c or there exists a larger rainbow matching.

Lemma 5.7 (Switching lemma)**.** *For each* $n \in \mathbb{N}$ *and* $\delta > 0$ *satisfying* $\delta\sqrt{n} \geq 12$*, the following holds. Let* $G = (V, E)$ *be a multigraph whose edges are coloured with* n *colours,* M *a rainbow matching of size* $n - 1$ *in* G*, and* c_0 *the colour that is missing in* M*.*

Suppose that for every colour c *in* G*, and every* (c_0, c)*-switching* σ *there are at least* $\big(\lceil (1+\delta)n \rceil - 4\ell(\sigma)\big)$ *disjoint* c*-edges between* $V \setminus \big(V(M) \cup V(\sigma)\big)$ *and* $V \setminus V(\sigma)$*. Then* G *has a rainbow matching of size* n*.*

An important special case of the condition in Lemma 5.7 is when $c = c_0$ and σ is the empty switching $\sigma_{c_0}^0$. In this case the condition says that there are at least $\lceil (1+\delta)n \rceil$ disjoint c_0-edges touching $V \setminus V(M)$.

Proof of Lemma 5.7. Let C be the set of colours of edges of G and $R := V \setminus V(M)$. We prove Lemma 5.7 by induction on n. For the initial case, we prove the theorem for all $n \leq 144$. Notice that if $n \leq 144$, then from $\delta\sqrt{n} \geq 12$, we obtain $\delta \geq 1$. This means in particular that there are $2n$ disjoint edges of colour c_0 in $E_G[R, V]$. However, there can be at most $|V(M)| = 2n-2$ disjoint c_0-edges in $E_G[R, V(M)]$. Hence, there exists a c_0-edge in $E_G[R, R]$, which can be added to M in order to obtain a rainbow matching of size n.

Now let $n > 144$ and assume that Lemma 5.7 holds for every $n' < n$. We may also assume that $\delta \leq 1$ since otherwise there is a rainbow matching of size n by the same argument as before. Let $G = (V, E)$ be a multigraph and M a rainbow matching of size $n - 1$ in G, which satisfies all the assumptions of the lemma. Suppose for the sake of contradiction that G does not have a rainbow matching of size n. The following claim produces a switching, a set of colours, and a set of edges that will later be used to reduce the problem to a smaller multigraph, to which we apply induction.

Claim 5.8. *There exist a colour* $c_2 \in C$*, a* (c_0, c_2)*-switching* $\sigma = (e_0, m_1, e_1, m_2)$ *and a subset* $C^* \subseteq C \setminus \{c_0, c_1, c_2\}$*, where* $c_1 := c(m_1)$*, with* $|C^*| = \lceil \delta n/6 \rceil$*, such that for each* $c \in C^*$ *there exists a* c*-edge* e_c *between* $V \setminus \big(V(M) \cup V(\sigma)\big)$ *and* $m_2 \setminus e_1$*.*

In Figure 5.3 we illustrate the switching, the set of colours and the edges that are guaranteed by Claim 5.8.

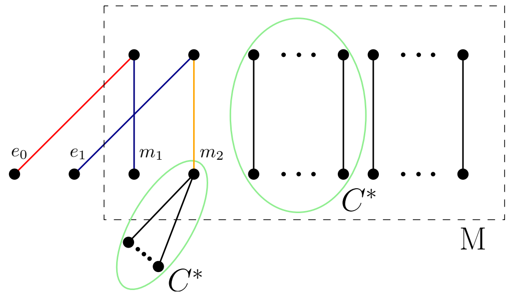

Figure 5.3: An illustration of the switching (e_0, m_1, e_1, m_2), the set $C^* \subseteq C \setminus \{c_0, c_1, c_2\}$ and the family of c-edges in $V \setminus \big(V(M) \cup V(\sigma)\big)$ with $c \in C^*$.

Proof of Claim 5.8. Let $C_1 := \big\{c \in C : \exists (c_0, c)\text{-switching of length } 1\big\}$. First we show that C_1 is big. By the assumption of the lemma, there exist $\lceil (1+\delta)n \rceil$ disjoint c_0-edges having an endvertex in R. If there exists a c_0-edge $e \in E_G[R, R]$, then $M \cup \{e\}$ is a rainbow matching of size n. Therefore we may assume that all c_0-edges from R end in $V(M)$, which implies that

$$|C_1| \geq \frac{\lceil (1+\delta)n \rceil}{2}. \tag{5.1}$$

For every $c \in C_1$, let $\sigma_c = (e_0^c, m_1^c)$ be an arbitrary but fixed (c_0, c)-switching of length 1. For a colour $c \in C_1$, we say that an edge $m \in M$ is *c-good* if there exist two disjoint c-edges in $E_G\big[R \setminus V(\sigma_c), m\big]$. Over the next few paragraphs we will find a large set of colours $C_2 \subseteq C_1$ and an edge m_2 such that m_2 is c-good for all $c \in C_2$.

By the assumption of the lemma, for every $c \in C_1$, there exist $\lceil (1+\delta)n \rceil - 4$ disjoint c-edges in $E_G\big[R \setminus V(\sigma_c), V \setminus V(\sigma_c)\big]$. If there exists a colour $c \in C_1$ and a c-edge $e \in E_G\big[R \setminus V(\sigma_c), R \setminus V(\sigma_c)\big]$, then there is a rainbow matching of size n, namely the union of the subset of M induced by the colours in $C \setminus \{c\}$, and the edges e and e_0^c. Therefore we may assume that for every $c \in C_1$, there exist $\lceil (1+\delta)n \rceil - 4$ disjoint c-edges in $E_G\big[R \setminus V(\sigma_c), V(M) \setminus V(\sigma_c)\big]$.

The maximum number of disjoint c-edges in $E_G\big[R \setminus V(\sigma_c), V(M) \setminus V(\sigma_c)\big]$ is less than twice the number of c-good edges $m \in M$ plus the number of edges $m \in M$ that are not c-good. Hence, for every $c \in C_1$, there exist at least $\lceil \delta n \rceil - 2$ edges in $M \setminus m(\sigma_c)$ that are c-good since otherwise there would be less than $2(\lceil \delta n \rceil - 3) + (n - \lceil \delta n \rceil + 2) = \lceil (1+\delta)n \rceil - 4$ edges of colour c in $E_G\big[R \setminus V(\sigma_c), V \setminus V(\sigma_c)\big]$, a contradiction to the assumption of the lemma. Next we find an edge m that is c-good for many colours $c \in C_1$. Let

$$\mu := \max_{m \in M} \{|C'| : C' \subseteq C_1 \setminus \{c(m)\} \text{ such that } m \text{ is } c\text{-good for each } c \in C'\}.$$

Double counting the pairs (c, m), where $c \in C_1 \setminus \{c(m)\}$ and m is a c-good edge, yields

$$\mu |M| \geq |C_1| \big(\lceil \delta n \rceil - 2\big)$$

and hence using (5.1) we obtain

$$\mu \geq \frac{(\delta n - 2)(1+\delta)}{2}.$$

This means that there exists an edge $m_2 = \{x, y\} \in M$ and a subset $C_2 \subseteq C_1 \setminus \{c(m_2)\}$ of size $\lceil (\delta n - 2)(1+\delta)/2 \rceil$ such that m_2 is c-good for every $c \in C_2$.

For every $c \in C_2$, let $x_c, y_c \in E_G\big[R \setminus V(\sigma_c), m_2\big]$ be disjoint edges of colour c such that $x_c \cap m_2 = \{x\}$ and $y_c \cap m_2 = \{y\}$ (such edges exist since m_2 is c-good). Let $X := \{x_c : c \in C_2\}$ and $Y := \{y_c : c \in C_2\}$. The remainder of the proof is split into the case that there exists a vertex in R that is incident to at least $1/3$ of the edges in X and the case that there does not exist such a vertex.

Case 1: Suppose that there exists a vertex $v \in R$ such that v is incident to at least $1/3$ of the edges in X. Using $\delta\sqrt{n} \geq 12$ and $\delta \leq 1$, notice that

$$\frac{|X|}{3} = \frac{|C_2|}{3} \geq \frac{(\delta n - 2)(1+\delta)}{6} \geq \left\lceil \frac{\delta n}{6} \right\rceil + 1.$$

Therefore, we can let X' be a subset of X consisting of $\lceil \delta n/6 \rceil + 1$ edges such that $v \in e$ for every $e \in X'$. Let e_1' be any edge in X'. Since $c(e_1') \in C_2$, there is an edge $e_1 \in Y$ with $c(e_1') = c(e_1)$. By the definition of X and Y, we also have $e_1 \cap e_1' = \varnothing$ and $V(\sigma_{c(e_1)}) \cap e_1' = \varnothing$, which imply $X' \subseteq E_G\big[R \setminus \big(V(\sigma_{c(e_1)}) \cup e_1\big), x\big]$. Set $c_1 := c(e_1)$, $c_2 := c(m_2)$, $e_0 := e_0^{c_1}$, and $m_1 := m_1^{c_1}$. We show that the set $C^* := \{c \in C_2 : x_c \in X' \setminus \{e_1'\}\}$, the sequence $\sigma := (e_0, m_1, e_1, m_2)$, and edges $e_c := x_c$ for each $c \in C^*$ are as desired in the claim.

First let us argue that σ is indeed a (c_0, c_2)-switching. Property (S1) is fulfilled since $m_1 \in M$ as (e_0, m_1) is a switching and since $m_2 \in M$ by the choice of m_2. Property (S2) holds since (e_0, m_1) is a switching and since $e_1 \cap m_2 = \{y\} \neq \varnothing$ and $e_1 \cap R \neq \varnothing$ by the definition of Y. As c_0 is not assigned to edges in M, we have $c(m_1) \neq c(e_0) \neq c(m_2)$. Moreover, we have $c(m_1) = c(e_1)$ by construction. This shows Property (S3). Finally, Property (S4) is satisfied since we have $e_1 \in E_G\big[R \setminus V(\sigma_{c_1}), m_2\big]$ by definition of $e_1 \in Y$, and hence $e_0 \cap e_1 = \varnothing$. Thus, σ is indeed a (c_0, c_2)-switching.

Observe that $C^* \subseteq C_2 \setminus \{c_1\} \subseteq C_1 \setminus \{c_1, c_2\} \subseteq C \setminus \{c_0, c_1, c_2\}$. Finally, for each $c \in C^* \subseteq C$, we have $e_c = x_c \in E_G\big[V \setminus \big(V(M) \cup V(\sigma)\big), x\big] = E_G\big[V \setminus \big(V(M) \cup V(\sigma)\big), m_2 \setminus e_1\big]$.

Case 2: Suppose that all vertices in R are incident to at most $1/3$ of the edges in X. Let e_1 be any edge in Y. Then at least $2/3$ of the edges in $X \subseteq E_G\big[R, x\big]$ are disjoint from $e_1 \in E_G\big[R, y\big]$ and at least $2/3$ of the edges in X are disjoint from $\sigma_{c(e_1)} = \big(e_0^{c(e_1)}, m_1^{c(e_1)}\big)$ since $e_0^{c(e_1)} \in E_G\big[R, m_1^{c(e_1)}\big]$ and $m_1^{c(e_1)} \cap \{x\} = \varnothing$. As a consequence, at least $1/3$ of the edges in X are disjoint from both e_1 and $\sigma_{c(e_1)}$. Since, as before, $|X|/3 \geq (\delta n - 2)(1+\delta)/6 \geq \lceil \delta n/6 \rceil + 1$ we can choose a subset $X^* \subseteq X$ of size $\lceil \delta n/6 \rceil$ such that for every $e \in X^*$ we have $c(e) \neq c(e_1)$ and $e \cap \big(e_1 \cup \sigma_{c(e_1)}\big) = \varnothing$. Set again $c_1 := c(e_1)$, $c_2 := c(m_2)$, $e_0 := e_0^{c_1}$, and $m_1 := m_1^{c_1}$. Analogously to the previous case, the set $C^* := \{c \in C_2 : x_c \in X^*\}$, the sequence $\sigma := (e_0, m_1, e_1, m_2)$, and the edges $e_c := x_c$ for each $c \in C^*$ are as desired in the claim. □

From now on we may assume the existence of the switching $\sigma = (e_0, m_1, e_1, m_2)$, the set $C^* \subseteq C$ and the edges e_c as in Claim 5.8. Next we define a subgraph G' of G, as well as a matching M' in G' to which we will apply induction. To this end we consider the following sets:

$$\begin{aligned} W &:= \{e \in M : \ c(e) \in C^*\}, \\ M' &:= M \setminus (m(\sigma) \cup W), \\ C' &:= C \setminus (C^* \cup \{c_0, c_1\}). \end{aligned}$$

Observe that C' is exactly the set of all colours assigned to the edges of M' plus colour c_2. Note further that $|W| = |C^*| = \lceil \delta n/6 \rceil$. Moreover, we set

$$\begin{aligned} n' &:= |C'| = \lfloor n(1-\delta/6) \rfloor - 2, \\ S &:= \bigcup\nolimits_{c \in C^*} e_c \cap R, \\ R' &:= R \setminus (V(\sigma) \cup S). \end{aligned}$$

See Figure 5.4 for an illustration of the sets $M', W \subseteq M$, the set $S \subseteq R$ and the switching (e_0, m_1, e_1, m_2).

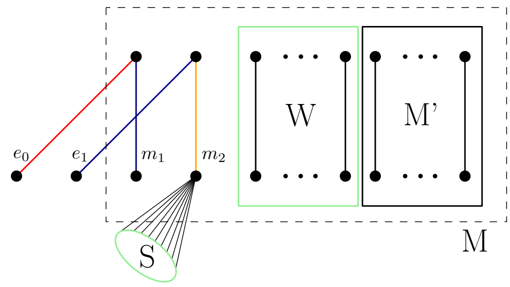

Figure 5.4: Sets $M', W \subseteq M$, set $S \subseteq R$ and switching (e_0, m_1, e_1, m_2).

To apply induction, we now consider the edge-coloured multigraph G' formed from G by deleting edges of colours from $C^* \cup \{c_0, c_1\}$ and vertices in $\big(V(\sigma) \cup S \cup V(W)\big)$. Formally, let $G' = (V', E')$ be the multigraph with vertex set

$$V' := V \setminus (V(\sigma) \cup S \cup V(W)) = R' \cup V(M')$$

and edge set

$$E' := \{e \in E_G[V'] : \ c(e) \in C'\}.$$

The edges of G' keep the colours they had in G.

With this notation in hand, we show that G', M', n', and c_2 satisfy the inductive assumption of the lemma. We prove the following claim.

Claim 5.9. *There is a constant $\delta' \geq 12/\sqrt{n'}$ such that for every colour $c \in C'$ and (c_2, c)-switching σ (in G', with respect to M') there are at least $\big(\lceil (1+\delta')n' \rceil - 4\ell(\sigma)\big)$ disjoint c-edges between $V' \setminus \big(V(M') \cup V(\sigma')\big)$ and $V' \setminus V(\sigma')$.*

Proof of Claim 5.9. Set $\delta' = (\lceil \delta n \rceil - 12)/n'$ and notice that the following holds:

$$\delta' n' \geq \delta n - 12 \geq 12(\sqrt{n} - 1) \geq 12\sqrt{n}(1 - \delta/12) \geq 12\sqrt{n}\sqrt{1-\delta/6} > 12\sqrt{n'}.$$

The second and third inequalities use $\delta \geq 12/\sqrt{n}$. Therefore $\delta' \geq 12/\sqrt{n'}$ holds.

Consider any colour $c \in C'$ and let σ' be some (c_2, c)-switching in G'. We need to show that there are at least $\big(\lceil (1+\delta')n' \rceil - 4\ell(\sigma')\big)$ disjoint c-edges in $E_{G'}\big[R' \setminus V(\sigma'), V' \setminus V(\sigma')\big]$. Recall that σ is the (c_0, c_2)-switching, given by Claim 5.8. Then the concatenation of σ and σ' gives a (c_0, c)-switching σ'' (in G w.r.t. M) of length $\ell(\sigma') + 2$. So, by the assumption of Lemma 5.7 on G, we can find at least $\lceil (1+\delta)n \rceil - 4\ell(\sigma'') = \lceil (1+\delta)n \rceil - 4(\ell(\sigma') + 2)$ disjoint

c-edges in $E_G\big[R \setminus V(\sigma''), V \setminus V(\sigma'')\big]$. As $|R \setminus V(\sigma'')| - |R' \setminus V(\sigma'')| = |S| \le \lceil \delta n/6 \rceil$, at least $\lceil (1+\delta)n \rceil - \lceil \delta n/6 \rceil - 4(\ell(\sigma')+2)$ of these disjoint edges belong to $E_G\big[R' \setminus V(\sigma''), V \setminus V(\sigma'')\big] \subseteq E_G\big[R' \setminus V(\sigma'), V \setminus V(\sigma')\big]$.

Assume first that there is no edge $e \in E_G\big[R' \setminus V(\sigma'), V(W) \cup S\big]$ with $c(e) = c$. Then, since at most 6 disjoint edges of colour c intersect $V(\sigma)$, the claim holds since the number of c-edges in $E_{G'}\big[R' \setminus V(\sigma'), V' \setminus V(\sigma')\big]$ is at least

$$\lceil (1+\delta)\, n \rceil - \lceil \tfrac{\delta n}{6} \rceil - 4(\ell(\sigma') + 2) - 6 = (1+\delta')n' - 4\ell(\sigma').$$

Assume then that there is an edge $e \in E_G\big[R' \setminus V(\sigma'), V(W) \cup S\big]$ with $c(e) = c$. If $e \cap S \neq \varnothing$, then $(M \setminus m(\sigma'')) \cup (\{e\} \cup e(\sigma''))$ is a rainbow matching of size n in G. Otherwise, if $e \cap V(W) \neq \varnothing$, then let $f \in W$ with $f \cap e \neq \varnothing$. By definition of W and S, and by Claim 5.8 we find an edge $g \in E_G(S, m_2 \setminus e_1)$ with $c(g) = c(f)$. Then $(M \setminus (m(\sigma'') \cup \{f\})) \cup (e(\sigma'') \cup \{e, g\})$ is a rainbow matching of size n in G, contradicting our assumption that M was maximum. □

Now we are able to finish the induction. By Claim 5.9, the multigraph G' satisfies the hypothesis of Lemma 5.7. Therefore, since $n' < n$, we can apply induction, which yields that G' contains a rainbow matching M'' of size n'. Now, $M'' \cup W \cup e(\sigma)$ forms a rainbow matching of size n, which is a contradiction to our assumption that there were no rainbow matchings in G of this size. □

5.3.3 Proof of the main result

With Lemma 5.7 in hand, we can now present the proof of Theorem 5.4, the idea of which is as follows. If a largest rainbow matching M in the given graph G has size less than n, we choose any colour c_0 that is not in M and consider a subgraph G' of G that is induced by the edges that are coloured by the colours of M or by c_0. The goal is then to show that G' contains a rainbow matching that uses all colours of G', which would lead to a contradiction to the maximality of M. Using the assumption on the number of vertices of the induced subgraph of each colour class allows us to show that the technical requirements of Lemma 5.7 are fulfilled. This is the place where we need to have for every colour $3n + o(n)$ rather than $3n - 2$ such vertices. Finally, Lemma 5.7 guarantees that there exists a rainbow matching using all colours in G'.

Proof of Theorem 5.4. Let $\delta > 0$, $n \ge 144/\delta^2$, and let G be as stated in the theorem. For the sake of contradiction, let us assume that a largest rainbow matching M in G has size smaller than n. Let C' be the set of colours in M plus one further colour c_0. Set $n' := |C'|$. In the following, we consider the multigraph $G' = (V, E')$ with $E' = \{e \in E : c(e) \in C'\}$. We now apply Lemma 5.7 to G' in order to find a rainbow matching of size n'. This is a contradiction since we assumed that M was a maximum matching.

Let $\delta' = \delta n/n'$ and observe that from $n \ge 144/\delta^2$, we have $\delta' \ge 12/\sqrt{n'}$. Let $c \in C'$ and let σ be any (c_0, c)-switching in G' with respect to M. By assumption on G, the number of vertices in $V \setminus \big(V(M) \cup V(\sigma)\big)$ that are incident to colour c is at least

$$\lceil (3+\delta)n \rceil - |V(M) \cup V(\sigma)| > \lceil (1+\delta)n \rceil - \ell(\sigma) \ge \lceil (1+\delta')n' \rceil - \ell(\sigma).$$

If in the subgraph induced by the colour class of c any two of these vertices are adjacent or have a common neighbour, then, since all the colour classes in G are unions of cliques, there is an edge, say e, of colour c between them, which leads to the rainbow matching $(M \setminus m(\sigma)) \cup (e(\sigma) \cup \{e\})$ of size n'. So, we may assume that there are at least $\big(\lceil(1+\delta')n'\rceil - \ell(\sigma)\big)$ disjoint edges of colour c in $E_{G'}\big[V \setminus \big(V(M) \cup V(\sigma)\big), V\big]$. Therefore and since there can be at most $3\ell(\sigma)$ disjoint c-edges in $E_{G'}\big[V \setminus \big(V(M) \cup V(\sigma)\big), V(\sigma)\big]$, there are at least $\big(\lceil(1+\delta')n'\rceil - 4\ell(\sigma)\big)$ disjoint c-edges in $E_{G'}\big[V \setminus \big(V(M) \cup V(\sigma)\big), V \setminus V(\sigma)\big]$. As c and σ were chosen arbitrarily, Lemma 5.7 now guarantees that G' contains a rainbow matching of size n'. □

5.4 Concluding remarks

We wonder how Theorem 5.3 changes if one adds the natural constraint that every pair of distinct elements belongs to at most one equivalence relation. More precisely, we are interested in the following problem.

Problem 5.10. *Determine the minimal number $\mathfrak{v}^*(n)$ such that if $A_1, \ldots, A_n$ are equivalence relations on a set X with $|\ker(A_i)| \geq \mathfrak{v}^*(n)$ and $A_i \cap A_j = \big\{(x,x) : x \in X\big\}$ for all distinct indices $i, j \in [n]$, then $A_1, \ldots, A_n$ contains a rainbow matching.*

Using the graph theoretic notion as before, the additional constraint means that the colour classes are pairwise disjoint. This can also be seen as restricting the problem to graphs instead of considering multigraphs. It is known that for every even n, there exists an edge-coloured bipartite graph whose colour classes induce matchings of size n and that does not contain a rainbow matching of size n. This follows from the fact that for every even n there exists a Latin square of order n without a transversal (see Subsection 1.2.2). For general n we thus obtain $\mathfrak{v}^*(n) > 2n - 2$. An upper bound on $\mathfrak{v}^*(n)$ follows directly from Theorem 5.4, i.e. $\mathfrak{v}^*(n) \leq \mathfrak{v}(n) = 3n + o(n)$.

Corollary 5.5 assures that a collection of n matchings of size $(3/2 + o(1))\, n$ in a bipartite multigraph guarantees a rainbow matching of size n. Aharoni, Kotlar, and Ziv [5] slightly improved the lower bound on the sizes of the matchings to $\lceil 3n/2 \rceil + 1$. For smaller matching sizes, it is unknown whether a rainbow matching of size $n-1$ exists. More generally, as suggested by Tibor Szabó (private communication), it would be interesting to determine upper bounds on the smallest integer $\mu(n,\ell)$ such that every family of n matchings of size $\mu(n,\ell)$ in a bipartite multigraph guarantees a rainbow matching of size $n - \ell$. One can verify that $\mu(n,l) \leq \frac{l+2}{l+1}n$. Moreover, it holds that $\mu(n,\sqrt{n}) \leq n$, which is a generalisation (see e.g. [4, 19]) of a result proved in the context of Latin squares by Woolbright [158], and independently by Brouwer, de Vries and Wieringa [45].

In order to approach Conjecture 1.13, one can also increase the number of matchings and fix their sizes to be equal to n instead of considering families of n matchings of sizes greater than n. Drisko [65] proved that a collection of $2n-1$ matchings of size n in a bipartite multigraph with partition classes of size n guarantees a rainbow matching of size n. This result is tight, as the factorisation of a cycle on $2n$ vertices with edges of multiplicity $n-1$ show. The problem was further investigated in the following two directions. Does the statement also hold if we omit the restriction on the sizes of the vertex classes? And how many matchings do we need to find a rainbow matching of size $n - \ell$ for every $\ell \geq 1$?

Aharoni and Berger [2] affirmed the first question by showing that for any two integers $s \leq t$, the maximal number of matchings of size t in a bipartite multigraph that do not contain a rainbow matching of size s is equal to $2(s-1)$.

The second question was studied recently by Barát, Gyárfás, and Sárközy in [19]. They proved that for every $\ell \geq 1$ any bipartite multigraph with $\left\lfloor \frac{\ell+2}{\ell+1} n \right\rfloor - (\ell+1)$ matchings of size n has a rainbow matching of size $n-\ell$. This result is best possible for $\ell = 0$ and $\lfloor n/2 \rfloor \leq \ell < n$.

Finally, if Conjecture 1.13 is true, it is of interest to see how tight it is. As shown by Barát and Wanless [20], one can find constructions of n matchings with $\left\lfloor \frac{n}{2} \right\rfloor - 1$ matchings of size $n+1$ and the remaining ones being of size n such that there is no rainbow matching of size n. We wonder whether the expression $\left\lfloor \frac{n}{2} \right\rfloor - 1$ above could also be replaced by $(1 - o(1))n$.

6 Enumerating spanning trees in series-parallel graphs

While in the preceding chapters we were examining questions concerning the *existence* of certain substructures, in this chapter we are interested in the *number* of spanning subgraphs. As motivated and summarised in Subsection 1.2.3, the enumeration of spanning trees and the analysis of series-parallel graphs (or SP graphs for short) have a rich history. By means of analytic techniques we show that the expected number of spanning trees in a connected labelled SP graph on n vertices chosen uniformly at random satisfies an estimate of the form

$$s\varrho^{-n}\big(1+o(1)\big),$$

where $s \approx 0.09063$ and $\varrho^{-1} \approx 2.08415$ can be computed explicitly (Theorem 6.1). We obtain analogue results for subfamilies of SP graphs including 2-connected SP graphs (Theorem 6.1), 2-trees (Theorem 6.8), and SP graphs with fixed excess (Theorem 6.12). Our proofs are based on analytic combinatorics, especially on the symbolic method and the singularity analysis of generating functions. The necessary analytic background was introduced in Section 2.3.

Since we focus on enumerative problems defined on SP graphs, let us quickly state the following alternative definition of SP graphs, which provides more insight into their structure and also justifies their name: let G be a graph and let s and t be two of its vertices. We say G is *series-parallel with terminals s and t* if G can be turned into the single edge $\{s,t\}$ by a sequence of the following operations: replacement of a pair of *parallel edges* (i.e. edges sharing two common endpoints) by a single edge, or replacement of a pair of *series edges* (i.e. non-parallel edges sharing a common endpoint of degree 2) by a single edge. A graph G is *2-terminal series-parallel* if there exist vertices s and t in G such that G is series-parallel with terminals s and t. Finally, a graph G is *series-parallel* if and only if each of its 2-connected components is a 2-terminal series-parallel graph (see e.g. [44]).

This chapter is structured in the following way. In Section 6.1 we prove Theorem 6.1 and analyse the behaviour of the growth constant of the expected number of spanning trees if we fix the edge density of a random 2-connected SP graph. We also comment on the variance of the number of spanning trees in a 2-connected SP graph chosen uniformly at random. Next, Section 6.2 is devoted to the analysis of the number of spanning trees in edge-maximal SP graphs (Theorem 6.8). The proof of Theorem 6.12, which deals with connected SP graphs with fixed excess, is presented in Section 6.3. We close the chapter with some concluding remarks and open questions in Section 6.4.

Throughout this chapter all graphs under study are labelled, unless stated otherwise. Furthermore, in contrast to the previous chapters, by a random object of a given family we mean an object chosen uniformly at random from all the elements of the same size, e.g. graphs on the same number of vertices. As already mentioned, this chapter is based on joint work with Juanjo Rué [75].

6.1 Spanning trees in connected and 2-connected SP graphs

The goal of this section is to analyse the expected value and the variance of spanning trees in random connected and 2-connected SP graphs as well as to elaborate on the growth constant of the expected number of spanning trees in random 2-connected SP graphs of a given edge density. Our main result in this respect is the following theorem. Its proof is presented in Subsection 6.1.1.

Theorem 6.1. *Let X_n and Z_n denote the number of spanning trees in a connected, respectively 2-connected labelled SP graph on n vertices chosen uniformly at random. Then*

$$\mathbb{E}[X_n]= s\varrho^{-n}\big(1+o(1)\big), \quad \text{where } s \approx 0.09063,\ \varrho^{-1} \approx 2.08415,$$
$$\mathbb{E}[Z_n]= p\varpi^{-n}\big(1+o(1)\big), \ \text{where } p \approx 0.25975,\ \varpi^{-1}\approx 2.25829.$$

Since the number of connected/2-connected SP graphs was already determined asymptotically by Bodirsky, Giménez, Kang, and Noy [33] (see Theorem 1.14), we may reduce the problem of estimating the number of spanning trees in random connected/2-connected SP graphs to the enumeration of connected/2-connected SP graphs carrying a distinguished spanning tree.

For this purpose, let $c_{n,m}$ and $b_{n,m}$ denote the number of connected and 2-connected SP graphs with a distinguished spanning tree, respectively. Let $C(x, y)$ and $B(x, y)$ be their associated counting formula, where x and y mark vertices and edges, respectively. Furthermore, let $\mathcal{D}$ denote the class of series-parallel networks carrying a distinguished spanning tree and let $D(x, y)$ be its associated generating function.

6.1.1 Expected number of spanning trees

The first step in our proof of Theorem 6.1 is the enumeration of SP networks that carry a distinguished spanning tree. For this purpose we need to introduce the following auxiliary class. Let $\bar{\mathcal{D}}$ denote the class of SP networks that carry a distinguished spanning forest with two components, each of which contains one of the poles. Let $\bar{D}(x, y)$ denote its associated (bivariate) exponential generating function (or EGF for short).

Recall that an SP network is either trivial, series, or parallel. By convention, we assume that networks with a root edge are parallel. Therefore, we define the following classes of networks.

Let $\mathcal{S}$ and $\bar{\mathcal{S}}$ denote the class of series networks that carry a distinguished spanning tree, respectively a distinguished spanning forest with two components, each of which contains one of the poles. We denote their associated EGFs by $S(x, y)$ and $\bar{S}(x, y)$, respectively.

Similarly, let $\mathcal{P}$ and $\bar{\mathcal{P}}$ denote the class of parallel networks that carry a distinguished spanning tree, respectively a distinguished spanning forest with two components each of which contains one of the poles. Observe that in both families the root edge might be present. Their associated EGFs are denoted by $P(x, y)$ and $\bar{P}(x, y)$, respectively. For the sake of readability we may omit the parameters whenever they are clear from the context.

We start with elaborating relations between D, $\bar{D}$, S, $\bar{S}$, P, and $\bar{P}$ in order to obtain a suitable system of equations. One can easily verify that

$$D(x, y) = y + S(x, y) + P(x, y), \tag{6.1}$$

and

$$\bar{D}(x, y) = y + \bar{S}(x, y) + \bar{P}(x, y). \tag{6.2}$$

Note that in Equation (6.1) the variable y on the right-hand side corresponds to a trivial network with a distinguished spanning tree, whereas in Equation (6.2) it corresponds to a trivial network with a distinguished spanning forest with two components of size 1.

Let us now analyse series networks. Observe that a series network N can be decomposed into at least two networks, where the 0-pole of the i-th network is identified with the ∞-pole of the $(i+1)$-th network. Equivalently, N can be decomposed into an ordered sequence formed by a network N' that is not series and an arbitrary network N'' that are joined by a series operation. If $N \in \mathcal{S}$, then each of these two networks contains a distinguished spanning tree (see Figure 6.1). Using the symbolic method we therefore have

$$S(x, y) = \big(D(x, y) - S(x, y)\big)xD(x, y) = \big(y + P(x, y)\big)xD(x, y). \tag{6.3}$$

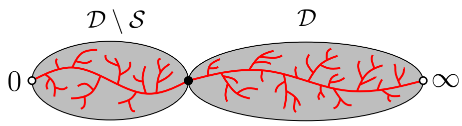

Figure 6.1: Decomposition of $N \in \mathcal{S}$.

If $N \in \bar{\mathcal{S}}$, then either $N' \in \mathcal{D}\backslash\mathcal{S}$ and $N'' \in \bar{\mathcal{D}}$, or $N' \in \bar{\mathcal{D}}\backslash\bar{\mathcal{S}}$ and $N'' \in \mathcal{D}$ (see Figure 6.2). This translates into the following equation:

$$\begin{aligned}\bar{S}(x, y) &= \big(D(x, y) - S(x, y)\big)x\bar{D}(x, y) + \big(\bar{D}(x, y) - \bar{S}(x, y)\big)xD(x, y)\\ &= \big(y + P(x, y)\big)x\bar{D}(x, y) + \big(y + \bar{P}(x, y)\big)xD(x, y).\end{aligned} \tag{6.4}$$

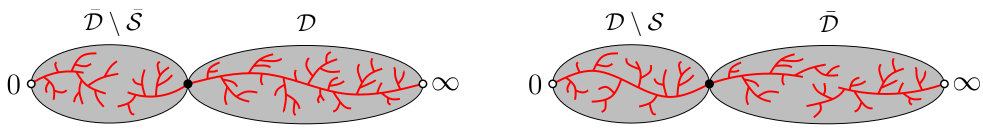

Figure 6.2: Decomposition of $N \in \bar{\mathcal{S}}$.

Finally, we turn to parallel networks. A parallel network can be described as a set of at least one series network if the root edge is present, or of at least two series networks, otherwise. If $N \in \mathcal{P}$, we need to distinguish between the case that the root edge is present and the case that it is not. In the second case, all series networks are in $\bar{\mathcal{S}}$ except for one, which is in $\mathcal{S}$. If in the first case the root edge is in the distinguished spanning tree of N, then all series

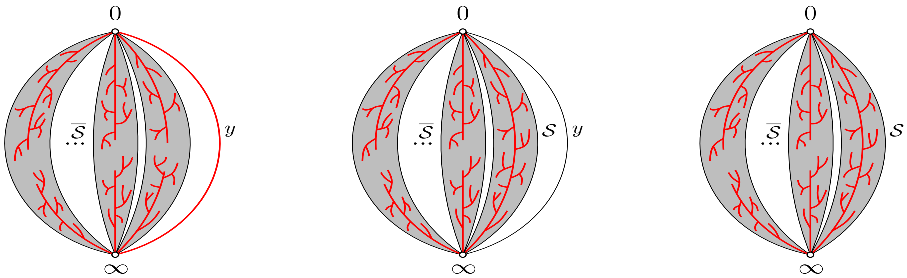

Figure 6.3: Decomposition of $N \in \mathcal{P}$.

networks are in $\bar{\mathcal{S}}$. If, on the other hand, the root edge is not in the spanning tree, then exactly one of the series networks is in $\mathcal{S}$ and all other networks are in $\bar{\mathcal{S}}$ (see Figure 6.3).

Thus, we get

$$P(x,y) = y\big(\exp(\bar{S}(x,y)) - 1\big) + y\big(S(x,y)\exp(\bar{S}(x,y))\big) + S\big(\exp(\bar{S}(x,y)) - 1\big). \tag{6.5}$$

If $N \in \bar{\mathcal{P}}$, then N can be decomposed into the root edge if present and into series networks in $\bar{\mathcal{S}}$ that are joined by a parallel operation. If the root edge is present, then there is at least one other network. In the other case, there must be at least two (see Figure 6.4). This gives rise to the following equation:

$$\bar{P}(x,y) = \big(\exp(\bar{S}(x,y)) - \bar{S}(x,y) - 1\big) + y\big(\exp(\bar{S}(x,y)) - 1\big). \tag{6.6}$$

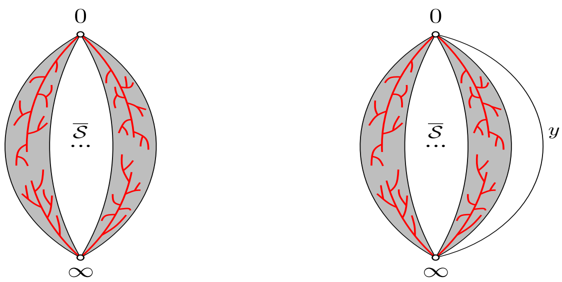

Figure 6.4: Decomposition of $N \in \bar{\mathcal{P}}$.

Using formal manipulations, we get that the system of Equations (6.1)–(6.6) defines the following implicit expression for $D(x,y)$:

$$D = \left(y + (1+y)\frac{xD^2}{1+xD}\right)\exp\left(-xD\frac{\big(y(1+xD) - (1+y)D\big)(2+xD)}{\big(y(1+xD) + (1+y)xD^2\big)(1+xD)^2}\right). \tag{6.7}$$

In order to study the singular behaviour of all the previous generating functions we could apply the Dromta-Lalley-Woods methodology for systems of functional equations (see e.g. [80]). However, as in this particular case we have an expression for $D(x,y)$ not depending on any other variables but x and y, we will analyse Equation (6.7) in order to get the singular behaviour of $D(x,y)$. The following theorem is reminiscent to [84, Lemma 3.3]:

Lemma 6.2. *Let $D(x,y)$ be the formal power series defined by $\Phi\big(x,y;D(x,y)\big)=0$, where*

$$\Phi(x,y;z) = z - \left(y + (1+y)\frac{xz^2}{1+xz}\right)\exp\left(-xz\frac{\big(y(1+xz)-(1+y)z\big)(2+xz)}{\big(y(1+xz)+(1+y)xz^2\big)(1+xz)^2}\right).$$

Then, for every $y>0$, it holds that $D(x,y)$ has a unique square-root singularity $R(y)$ such that $D(x,y)$ has a singular expansion of the following form in a domain dented at $x=R(y)$:

$$D(x,y) = D_0(y) + D_1(y)X(y) + D_2(y)X(y)^2 + D_3(y)X(y)^3 + \mathcal{O}\big(X(y)^4\big), \tag{6.8}$$

where $X(y)=\sqrt{1-x/R(y)}$.

For $y=1$ we have the numerical values $x=R(1)=R\approx 0.05668$, $D_0(1)\approx 1.82404$, $D_1(1)\approx -1.52769$, $D_2(1)\approx 1.34779$, and $D_3(1)\approx -1.25138$.

Proof. We fix $y>0$. A simple computation shows that $\Phi_z\big(0,y;D(0,y)\big)=1>0$ and $D(0,y)=y$. Hence, by the analytic implicit function theorem (see e.g. [80]) we know that $D(x,y)$ is analytic at $x=0$.

We continue with showing that $D(x,y)$ has a finite radius of convergence. Denote the singularity of the function $D(x,y)$ by $R(y)$. Observe that $[x^n]D_\varnothing(x,y)\le[x^n]D(x,y)$, where $D_\varnothing(x,y)$ is the generating function associated with SP networks without a distinguished spanning tree. As it is shown in [33], the radius of convergence $R_\varnothing(y)$ of $D_\varnothing(x,y)$ is finite. In particular, $0<R(y)\le R_\varnothing(y)<1<\infty$ and $D(x,y)$ ceases to be analytic at $x=R(y)$.

Observe that the only source of singularity for $D(x,y)$ is the condition

$$\Phi_z\big(R(y),y;D(R(y),y)\big)=0,$$

which means that the singularity arises from a branch point. Let us now justify that we have $\Phi_{zz}\big(R(y),y;D(R(y),y)\big)\neq 0$. This condition is enough in order to assure a square-root type singularity for each choice of y. For a contradiction, let us assume the opposite. Hence we have a solution (R_0,y_0,z_0) of the following system of equations:

$$\Phi(x,y;z)=0,\ \ \Phi_z(x,y;z)=0,\ \ \Phi_{zz}(x,y;z)=0.$$

Observe that $\Phi(x,y;z)=z-A(x,y;z)\exp\big(B(x,y;z)\big)$ with $A(x,y;z)$ and $B(x,y;z)$ being rational functions. Hence, $\Phi_z(x,y;z)=1-C(x,y;z)\exp(B(x,y;z))$ where again $C(x,y;z)$ is a rational function. Finally, $\Phi_{zz}(x,y;z)$ can be written in the form $E(x,y;z)\exp\big(B(x,y;z)\big)$ for a certain rational function $E(x,y;z)$.

In particular, combining the first two equations by eliminating the exponential term, we get the following system of rational equations:

$$zC(x,y;z)=A(x,y;z),\ \ E(x,y;z)=0.$$

After rearranging the denominators in both expressions, such a system can be transformed into a system of two polynomial equations $P_1(x,y;z)=0$, $P_2(x,y;z)=0$, from which we can get a new polynomial equation $Q(x,y)=0$ by eliminating the variable z. By carrying out the explained computations with `Maple`, we obtain

$$Q(x,y)=(-4y+yx-4)y(y+1)T(x,y),$$

where

$$T(x,y) = 100(1+y)^4 + 6917y(1+y)^3x + 1266y^2(1+y)^2x^2 - 1867y^3(1+y)x^3 + 280y^4x^4.$$

We now argue that $Q(x,y) = 0$ does not have a solution satisfying both $y > 0$ and $x < 1$. Observe that the first multiplicative term $-4y+yx-4$ gives the solution $(x,y) = (4+4/y, y)$. This means in particular that x is always greater than 1 if $y > 0$. It is also obvious that the multiplicative terms y and $y+1$ cannot contribute with the required solution. Therefore, we need to analyse the existence of solutions (x,y) of $T(x,y)$ with the condition $y > 0$ and $x < 1$. Using that $y(1+y)^3 > y^3(1+y)$ and $x > x^3$ for all $y > 0$ and $0 < x < 1$, we know that $6917y(1+y)^3x > 6917y^3(1+y)x^3 > 1867y^3(1+y)x^3$. Hence, $T(x,y) = 0$ does not have solutions with both $0 < x < 1$ and $y > 0$, which implies that the solution (x_0, y_0, z_0) of the equation $\Phi(x,y;z) = \Phi_z(x,y;z) = 0$ satisfies that $\Phi_{zz}(x_0,y_0;z_0) \neq 0$. Hence, the singularity of $D(x,y)$ is of a square-root type in a domain dented at $x = R(y)$. This proves the singular expansion in Equation (6.8).

In order to prove the special case of $y = 1$ in the statement of the lemma, we set $y = 1$, $R = R(1)$, and $X = X(1)$ (and consequently $x = R(1-X^2)$). By plugging the singular expansion of $D(x,1)$ in $\Phi(x,y;z) = 0$, taking the Taylor expansion in terms of X, and applying the method of indeterminate coefficients, we get the numerical values as claimed. Finally, observe that for each choice of $y > 0$, the generating function $D(x,y)$ is aperiodic, as for every n there exists a network on n vertices. Consequently, the singularity $R(y)$ is unique. □

Knowing that $D(x,y)$ admits a singular expansion of square root-type in a domain dented at $x = R(y)$, one can compute by means of indeterminate coefficients the exact expressions of $D_i(y)$ for $i \geq 1$ in terms of the function $D(R(y),y) = D_0(y)$, which satisfies the functional equation $\Phi\big(R(y), y, D_0(y)\big) = 0$. Although the expressions are long, we needed to compute the evaluations at $y = 1$ for enumerative purposes.

The following lemma gives the coefficients of the singular expansions (rounded up to 5 digits) of the EGFs $\bar{D}(x,1)$, $S(x,1)$, $\bar{S}(x,1)$, $P(x,1)$, and $\bar{P}(x,1)$. In order to get asymptotic estimates for these counting formulas we only need the multiplicative constant of the term $(1-x/R)^{1/2}$, in order to get the asymptotics in the 2-connected level we need expansions up to term $(1-x/R)^{3/2}$.

Lemma 6.3. *For each $y > 0$ the generating functions $\bar{D}$, S, $\bar{S}$, P, and $\bar{P}$ have a square-root singular expansion in a domain dented at $R(y)$, where $R(y)$ is the unique singularity of $D(x,y)$. Furthermore, for $y = 1$ the singular expansions of $\bar{D}$, S, $\bar{S}$, P, and $\bar{P}$ in a domain dented at $R := R(1) \approx 0.05668$ are*

$$\begin{aligned}
\bar{D}(x,1) &= \bar{D}_0(1)+\bar{D}_1(1)X+\bar{D}_2(1)X^2+\bar{D}_3(1)X^3+\mathcal{O}(X^4),\\
S(x,1) &= S_0(1)+S_1(1)X+S_2(1)X^2+S_3(1)X^3+\mathcal{O}(X^4),\\
\bar{S}(x,1) &= \bar{S}_0(1)+\bar{S}_1(1)X+\bar{S}_2(1)X^2+\bar{S}_3(1)X^3+\mathcal{O}(X^4),\\
P(x,1) &= P_0(1)+P_1(1)X+P_2(1)X^2+P_3(1)X^3+\mathcal{O}(X^4),\\
\bar{P}(x,1) &= \bar{P}_0(1)+\bar{P}_1(1)X+\bar{P}_2(1)X^2+\bar{P}_3(1)X^3+\mathcal{O}(X^4),
\end{aligned}$$

where $X = \sqrt{1-x/R}$ and the constants have the following approximate values:

	$i=0$	$i=1$	$i=2$	$i=3$
$\bar{D}_i(1)$	1.71871	−1.17120	1.17120	−0.59820
$S_i(1)$	0.17092	−0.27289	0.18433	−0.15440
$\bar{S}_i(1)$	0.30701	−0.43079	0.19616	−0.12220
$P_i(1)$	0.65312	−1.25480	1.16347	−1.09697
$\bar{P}_i(1)$	0.41170	−0.74041	0.58941	−0.47600

Proof. The first claim follows directly due to Equations (6.1)–(6.6), which are analytic and allow us to express $\bar{D}$, S, $\bar{S}$, P, and $\bar{P}$ in terms of D. In particular, all these generating functions have a unique singularity at $x = R(y)$. The second part follows by setting $y = 1$ and by plugging the singular expansion of $D(x, 1)$ into Equations (6.1)–(6.6). □

Now we turn to the analysis of $B(x, y)$, the EGF associated with the class of 2-connected SP graphs carrying a distinguished spanning tree. In Subsection 2.3.2 we have encountered in Equation 2.2 a relation between the EGF associated with 2-connected graphs (without a distinguished spanning tree) and the EGF associated with networks (without a distinguished spanning tree). In the context of 2-connected SP graphs with a distinguished spanning tree, Equation 2.2 translates to

$$2(1+y)B_y(x,y) = x^2\Big(1 + D(x,y) + \bar{D}(x,y) - y\big(\exp(\bar{S}(x,y)) - 1\big)\Big), \tag{6.9}$$

which means that when directing and possibly deleting an edge in a 2-connected SP graph with a distinguished spanning tree, the resulting object is a network, where labels are given to the poles, and which is either empty or of type $\mathcal{D}$ or of type $\bar{\mathcal{D}}$, but not a parallel network of type $\bar{\mathcal{P}}$ with an edge linking the poles (see Equation (6.6)).

A direct integration of Equation (6.9) is technically involved due to the relations between the generating functions associated with the different types of networks. However, we can get a simple expression of $B(x, y)$ in terms of the EGF associated with the networks just by combinatorial arguments using Tutte's decomposition and the dissymmetry theorem for tree-decomposable classes (Theorem 2.12). In the following lemma we provide such an equation.

Lemma 6.4. *The generating function $B(x, y)$ associated with the class of 2-connected SP graphs carrying a distinguished spanning tree can be expressed as*

$$B(x,y) = \frac{x^2}{2}y + B_R(x,y) + B_M(x,y) - B_{R-M}(x,y), \tag{6.10}$$

where

$$B_R(x,y) = \frac{x^2}{2}S(\bar{D} - \bar{S}), \tag{6.11}$$

$$B_M(x,y) = \frac{x^2}{2}\Big(S\big(\exp(\bar{S}) - \bar{S} - 1\big) + yS\big(\exp(\bar{S}) - 1\big) + y\big(\exp(\bar{S}) - \bar{S} - 1\big)\Big), \tag{6.12}$$

$$B_{R-M}(x,y) = \frac{x^2}{2}(S\bar{P} + \bar{S}P). \tag{6.13}$$

Proof. Applying Tutte's decomposition (see Subsection 2.3.2 for the combinatorial background) to 2-connected SP graphs bearing a distinguished spanning tree on at least 3 vertices only yields R-bricks (ring graphs) and M-bricks (multi-edge graphs), both carrying a distinguished spanning tree. In particular, there are no T-bricks since the family of h-networks is empty in our case. We obtain Expression (6.10) for $B(x, y)$ using Equation (2.4) to which we needed to add $x^2y/2$ since we also consider a single edge to be a 2-connected SP graph.

Let us study each term. Let R be a distinguished R-brick with a distinguished spanning tree. By definition, R is a cyclic chain of at least 3 networks that carries a spanning tree. In particular, exactly one of these networks is in $\bar{\mathcal{D}}$ while the other ones are in $\mathcal{D}$. This means that R can be decomposed into a non-series network in $\bar{\mathcal{D}}$ and a series network in $\mathcal{S}$ that are joined by a parallel operation and where the two poles are added to the graph, see also Figure 6.5. This gives Equation (6.11).

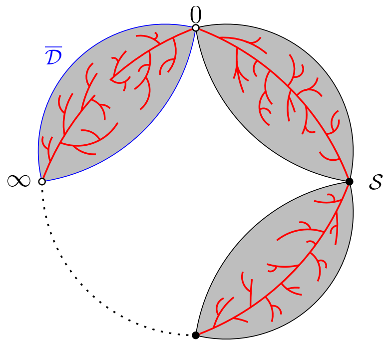

Figure 6.5: Decomposition of a distinguished R-brick in the RMT-tree.

We continue with M-bricks. Let M be a distinguished M-brick with a distinguished spanning tree. Then, M can be decomposed into at least three networks, all but possibly one of which are series and the possibly other one is a single edge. These networks are joined by a parallel operation and the two poles are again added to the graph. This situation is similar to the decomposition of parallel networks carrying a spanning tree that we considered for developing Equation (6.5). The main difference is that, by definition, M is decomposed into at least three and not into at least two networks.

We need to distinguish again between the two cases where there is a single edge component in M and where there is no such component. Observe that it is not possible that there are two such components in M since we are only considering simple graphs. In the former of the two cases, we note that if the edge of the single edge component is not contained in the distinguished spanning tree, then exactly one of the series networks is in $\mathcal{S}$ while all the others are in $\bar{\mathcal{S}}$. If, on the other hand, the edge is in the spanning tree, then all series networks must be in $\bar{\mathcal{S}}$. This gives rise to Equation (6.12).

Finally, we need to decompose 2-connected SP graphs with a distinguished spanning tree and with a distinguished $\{R, M\}$-edge in the RMT-tree. This means, that the distinguished edge corresponds to a virtual edge $\{x, y\}$ matching an R-brick and a M-brick. Hence the graphs can be decomposed into a series network and a parallel network by a parallel operation, where we need to add again the two poles to the graph. One of the two networks must be in $\mathcal{D}$ while the other one must be in $\bar{\mathcal{D}}$. As a consequence Equation (6.13) holds. See Figure 6.6 for an illustration of this situation. □

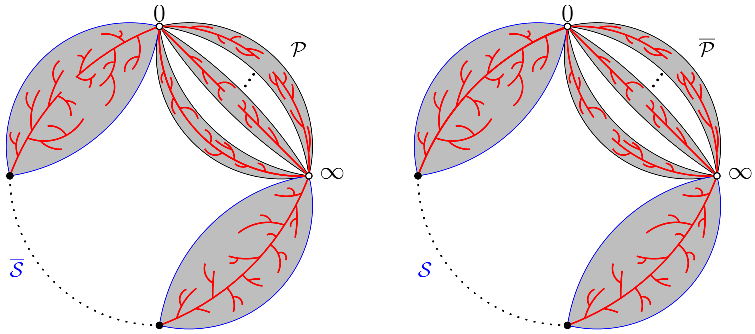

Figure 6.6: Decomposition of a distinguished $\{R, M\}$-edge in the RMT-tree.

With Lemmas 6.3 and 6.4 in hand, we can now turn to the analysis of the singular behaviour of $B(x, y)$.

Lemma 6.5. *Let $y > 0$. Then $B(x, y)$ has a unique square-root singularity, which is the unique singularity $R(y)$ of the function $D(x, y)$ from Lemma 6.2. Moreover, $B(x, y)$ has a singular expansion of the following form in a domain dented at $x = R(y)$:*

$$B(x, y) = B_0(y) + B_2(y)X(y)^2 + B_3(y)X(y)^3 + \mathcal{O}\big(X(y)^4\big), \tag{6.14}$$

where $X(y) = \sqrt{1 - x/R(y)}$.

For $y = 1$ we have $x = R(1) = R \approx 0.05668$, $B_0(1) \approx 0.00176$, $B_2(1) \approx -0.00394$, and $B_3(1) \approx 0.00062$.

Proof. Observe that the generating functions $B_R(x, y)$, $B_M(x, y)$ and $B_{R-M}(x, y)$ are analytic transformations of the generating functions for networks (namely, the EGFs that appear in Lemma 6.4). Hence, $B(x, y)$ has a unique dominant singularity, which is the same one as the coinciding singularity of the EGFs from Lemma 6.3, namely $R(y)$. Similarly, for each y we have that $B(x, y)$ admits a singular expansion in a domain dented at $R(y)$. In order to obtain it, we express the singular expansion of each of the network EGFs appearing in Equation (6.10) in terms of the singular expansions obtained in Lemma 6.3. Observe that Equation (6.9) implies that the singular expansion of $B(x, y)$ must start at $X(y)^3$, which gives in particular that $B_1(y) = 0$ (c.f. [80, Theorem VI.9]).

Finally, by setting $y = 1$ and by the same procedure as above using `Maple`, we obtain the approximation of $B_i(1)$ for $i \geq 0$ as stated in the lemma. In particular, the term $B_3(1)$ depends on all singular coefficients in Lemma 6.3. □

Next, we analyse the generating function $C(x, y)$ of connected SP graphs carrying a distinguished spanning tree. Since the singular expansion of $B(x, y)$ is of a square-root type with singular exponent $3/2$ as it is shown in Equation (6.14), we get the singular expansion of $C(x, y)$ immediately from [84, Proposition 3.10] (see also [71]).

Lemma 6.6. *The singularity of $C(x, y)$ is at*

$$\bar{\rho}(y) = \frac{\tau(y)}{\exp\big(B_x(\tau(y), y)\big)},$$

where $\tau(y)$ is the unique solution of the equation $\tau(y)B_{xx}(\tau(y), y) = 1$. The singular expansion of $C(x, y)$ in a domain dented at $\bar{\rho}(y)$ is

$$C(x, y) = C_0(y) + C_2(y)X(y)^2 + C_3(y)X(y)^3 + \mathcal{O}\big(X(y)^4\big),$$

where $X(y) = \sqrt{1 - x/\bar{\rho}(y)}$ and

$$\begin{aligned} C_0(y) &= \tau\big(1 + \log(\bar{\rho}(y)) - \log(\tau(y))\big) + B(\tau(y), y), \\ C_2(y) &= -\tau(y), \\ C_3(y) &= \frac{3}{2}\sqrt{\frac{2\bar{\rho}(y)\exp(B_x(\bar{\rho}(y), y))}{\tau B_{xxx}(\tau(y), y) - \tau B_{xx}(\tau(y), y)^2 + 2B_{xx}(\tau(y), y)}}\ . \end{aligned}$$

For $y = 1$ we have $\bar{\rho}(1) \approx 0.05288$, $C_0(1) \approx 0.05450$, $C_2(1) = -\tau \approx -0.05668$, and $C_3(1) \approx 0.00145$.

Proof. See [84, Proposition 3.10] for the general value of y. When $y = 1$, we use `Maple` to obtain the approximate values of the constants. The uniqueness of the singularity is assured by the aperiodicity of $C(x, y)$ with y being fixed (c.f. the proofs of [68, Lemma 7] and [68, Lemma 9]). □

Finally, we have all necessary ingredients to prove the main theorem of this section.

Proof of Theorem 6.1. We will prove the statement for X_n in detail. The result for Z_n is obtained *mutatis mutandis*.

Let us denote the class of all connected SP graphs on n vertices by $\mathcal{C}_n$, and the class of all connected SP graphs on n vertices carrying a distinguished spanning tree by $\mathcal{C}_n^s$. For a graph $G \in \mathcal{C}_n$ we write $s(G)$ for the number of spanning trees in G. Then, the expected value of X_n can be written as:

$$\mathbb{E}[X_n] = \sum_{G \in \mathcal{C}_n} s(G)\mathbb{P}[G] = \frac{\sum_{G \in \mathcal{C}_n} s(G)}{|\mathcal{C}_n|} = \frac{|\mathcal{C}_n^s|}{|\mathcal{C}_n|} = \frac{[x^n]C(x, 1)}{|\mathcal{C}_n|}. \tag{6.15}$$

It follows directly from Lemma 6.6 and the transfer theorem (Theorem 2.13) that the number of connected SP graphs on n vertices that carry a distinguished spanning tree is asymptotically equal to

$$\frac{C_3(1)}{\Gamma(-3/2)} n^{-5/2} \bar{\rho}(1)^{-n} n!\ .$$

From Theorem 1.14 we know that the number of connected SP graphs on n vertices is asymptotically equal to $c_s n^{-5/2} \varrho_s^{-n} n!$, where $c_s \approx 0.0067912$ and $\varrho_s \approx 0.11021$ are computable constants. Dividing the former by the latter as in Equation (6.15), we obtain that the expected value of X_n is asymptotically equal to $s\varrho^{-n}$, where $s \approx 0.09063$ and $\varrho^{-1} \approx 2.08415$.

The corresponding result for 2-connected SP graphs is obtained analogously by using [33, Theorem 2.6], which states that the number of 2-connected SP graphs on n vertices is asymptotically equal to $bn^{-5/2} r^{-n} n!$, where $b \approx 0.00101$ and $r \approx 0.12800$. □

6.1.2 Fixing the edge density

The previous results can be used to study random SP graphs with a fixed edge density, as well as limiting distributions for the number of edges. For the sake of brevity, we only discuss the problem for the family of 2-connected SP graphs, but the connected case is very similar.

The first main important observation is that the number of edges in a random 2-connected SP graph carrying a distinguished spanning tree follows a normal limiting distribution. Indeed, Lemma 6.5 shows that the singular behaviour of $B(x, y)$ is the same when choosing y in a real-valued neighbourhood of 1. Then, by Hwang's quasi-power theorem (see [96]) the distribution follows a normal limit law with linear expectation and linear variance. In particular, the number of edges is concentrated around its mean value. This behaviour is similar to the case of 2-connected SP graphs (without a spanning tree), where again the number of edges is normally distributed (see [33]).

Under these circumstances, our techniques provide a method to study the expected number of spanning trees in a random SP graph on n vertices of a given edge density μ. Following the arguments of [83, Theorem 3], for every $\mu > 0$ we can choose a value $y_0 > 0$ such that if we assign the weight y_0^k to each graph with k edges, then only the graphs with n vertices and with approximately μn edges (with a deviation of order $n^{1/2}$) have non-negligible weight. Such a technique is valid whenever Hwang's quasi-power theorem holds. Hence we can apply it in our context.

As a case example we plot the expected value of the random variable $Z_{n,\mu}$ that counts the number of spanning trees in a graph chosen uniformly at random from all 2-connected SP graphs with n vertices and edge density μ. Let $R(y)$ denote the radius of convergence of $B(x, y)$. Given an edge density μ, the right choice for y_0 is the unique positive solution of the following equation (see e.g. [83, Theorem 3]):

$$-y_0 \frac{R_y(y_0)}{R(y_0)} = \mu. \tag{6.16}$$

Observe that when μ tends to 1, the family of SP graphs under study are graphs with a small but positive number of cycles, whereas when μ tends to 2, the subfamily under study tends to the class of 2-trees. These cases correspond to the ones when y tends to 0 and infinity, respectively. Both cases will be analysed in full detail in Sections 6.2 and 6.3.

The precise computational method to obtain the exponential growth constant of the expected value of $Z_{n,\mu}$ as a function of the edge density is the following: For a given density μ we use (6.16) to obtain the corresponding y_0. Then we use the implicit expression of the singularity curve stated in [33, Theorem 2.2] in order to obtain the growth constant of the number of 2-connected SP graphs of edge density equal to μ.

To get the growth constant in the setting of 2-connected SP graphs carrying a spanning tree and with edge density μ, we perform the calculations explained in the proof of Lemma 6.2 for y_0. Finally, the exponential growth of the expected value of $Z_{n,\mu}$ is obtained by dividing these two numerical values as we did in the proof of Theorem 6.1. In Figure 6.7 we plot this exponential growth constant in terms of the edge density $\mu \in (1.07626, 1.97173)$.

We would like to mention that the non-plotted margins for μ correspond to values of y very close to 0 and when y tends to infinity. In both cases, the numerical method used to get the constant growth for the number of spanning trees fails because of indetermination of the operation to be carried out.

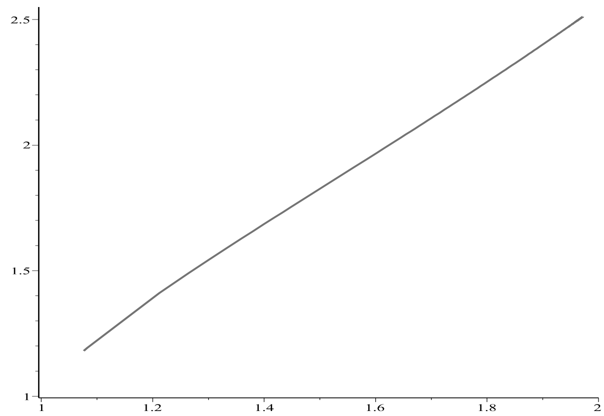

Figure 6.7: Exponential growth constant of the expected number of spanning trees in a random 2-connected SP graph (ordinate) as a function of its edge density (abscissa).

As already mentioned, the two cases when the edge density reaches its maximum and when it tends to its minimum will be analysed in Section 6.2 and in Section 6.3, respectively.

6.1.3 Variance of the number of spanning trees

Refining the combinatorics exploited in the proofs in Subsection 6.1.1, one has also access to the second moment of the random variables X_n and Z_n. In this subsection we develop this by determining the growth constant of the variance of Z_n. This will also show that Z_n is not concentrated around its expected value. Recall that Z_n was defined as the random variable that counts the number of spanning trees in a random 2-connected SP graph on n vertices.

In order to determine the growth constant of the second moment of Z_n, we will first study the asymptotic behaviour of the number of 2-connected SP graphs on n vertices carrying two distinguished spanning trees. As in Subsection 6.1.1, we start with the analysis of networks carrying spanning trees and spanning forests.

We define $\mathcal{D}^*$, $\mathcal{S}^*$, and $\mathcal{P}^*$ as the classes of SP, series, and parallel networks, respectively, each carrying two distinguished spanning trees. Let $D^*(x,y)$, $S^*(x,y)$, and $P^*(x,y)$ denote their EGFs.

In order to be able to analyse these generating functions, we need again some auxiliary classes. Let $\widetilde{\mathcal{D}}$ denote the class of all SP networks carrying a distinguished spanning tree and a distinguished spanning forest with (exactly) two components, each of which contains one pole. Furthermore, let $\hat{\mathcal{D}}$ denote the class of all SP networks carrying two distinguished spanning forests both with (exactly) two components, each of which contains one pole. Let $\widetilde{D}(x,y)$ and $\hat{D}(x,y)$ denote their EGFs. In the same way $\widetilde{\mathcal{S}}$, $\widetilde{\mathcal{P}}$, $\hat{\mathcal{S}}$, and $\hat{\mathcal{P}}$ as well as $\widetilde{S}(x,y)$, $\widetilde{P}(x,y)$, $\hat{S}(x,y)$, and $\hat{P}(x,y)$ are defined. We might again omit the parameters whenever they are clear from the context.

Following the proof of Theorem 6.1, we start with the following lemma that provides the growth constant of these generating functions.

Lemma 6.7. *For $y = 1$, the generating functions D^*, $\widetilde{D}$, $\hat{D}$, S^*, $\widetilde{S}$, $\hat{S}$, P^*, $\widetilde{P}$, and $\hat{P}$ have a square-root expansion in a domain dented at $R_2 \approx 0.02407$.*

Proof. We start again with elaborating relations between all given generating functions.

One can easily verify that the following relations hold:

$$\begin{aligned} D^*(x,y) &= y + S^*(x,y) + P^*(x,y), \\ \widetilde{D}(x,y) &= y + \widetilde{S}(x,y) + \widetilde{P}(x,y), \\ \hat{D}(x,y) &= y + \hat{S}(x,y) + \hat{P}(x,y). \end{aligned} \tag{6.17}$$

Recall that a series network can be decomposed into a network that is not series and an arbitrary SP network. Similarly to Equations (6.3) and (6.4), we get the following equations for the generating functions associated with the networks from the classes $\mathcal{S}$, $\widetilde{\mathcal{S}}$, and $\hat{\mathcal{S}}$.

$$\begin{aligned} S^*(x,y) &= \big(D^*(x,y) - S^*(x,y)\big)xD^*(x,y), \\ \widetilde{S}(x,y) &= \big(\widetilde{D}(x,y) - \widetilde{S}(x,y)\big)xD^*(x,y) + \big(D^*(x,y) - S^*(x,y)\big)x\widetilde{D}(x,y), \\ \hat{S}(x,y) &= \big(D^*(x,y) - S^*(x,y)\big)x\hat{D}(x,y) + \big(\hat{D}(x,y) - \hat{S}(x,y)\big)xD^*(x,y) \\ &\quad +2\big(\widetilde{D}(x,y) - \widetilde{S}(x,y)\big)x\widetilde{D}(x,y). \end{aligned} \tag{6.18}$$

Finally, a parallel network can be decomposed into a set of at least one series network if the root edge is present, or into at least two series networks, otherwise. In the first case we need to distinguish whether the root edge is in a distinguished spanning tree or not. By a careful case distinction, we get the following three equations:

$$\begin{aligned} P(x,y) &= S(x,y)\Big(\exp\big(\hat{S}(x,y)\big) - 1\Big) + (1+y)\widetilde{S}(x,y)^2 \exp\big(\hat{S}(x,y)\big) \\ &\quad + yS(x,y)\exp\big(\hat{S}(x,y)\big) + 2y\widetilde{S}(x,y)\exp\big(\hat{S}(x,y)\big) + y\Big(\exp\big(\hat{S}(x,y)\big) - 1\Big), \\ \widetilde{P}(x,y) &= \big(y + \widetilde{S}(x,y)\big)\Big(\exp\big(\hat{S}(x,y)\big) - 1\Big) + y\widetilde{S}(x,y)\exp\big(\hat{S}(x,y)\big), \\ \hat{P}(x,y) &= \Big(\exp\big(\hat{S}(x,y)\big) - \hat{S}(x,y) - 1\Big) + y\Big(\exp\big(\hat{S}(x,y)\big) - 1\Big). \end{aligned} \tag{6.19}$$

Equations (6.17)–(6.19) define a system of functional equations that can be analysed by means of Theorem 2.14 by setting $y = 1$. More precisely, each equation in this system is defined by an analytic function because the exponential function is an entire function. In addition, easy lower and upper bounds imply that the radius of convergence of $D(x,1)$, $\widetilde{D}(x,1)$, $\hat{D}(x,1)$ is in $(0,R]$, where $R \approx 0.05668$ is the constant from Lemma 6.2.

Solving the system of equations stated in Theorem 2.14 using `Maple` we get that $R_2 \approx 0.02407$. Furthermore, Theorem 2.14 assures that R_2 is the singularity of all generating functions appearing in Equations (6.17)–(6.19) and that they have a square-root expansion in a domain dented at R_2. As all generating functions written so far are aperiodic, the singularity is unique on the circle $|x| = R_2$. □

Let $B^*(x,y)$ denote the EGF associated with the class of all 2-connected SP graphs carrying two spanning trees. As in Lemmas 6.4 and 6.5 one can write $B^*(x,y)$ as an analytic combination of the generating functions from Lemma 6.7 by using Tutte's decomposition and

Equation (2.4). Therefore, the dominant singularity of $B^*(x, 1)$ is the same as the coinciding singularity of the EGFs appearing in Lemma 6.7, which is R_2, and moreover, $B^*(x, 1)$ has a square-root expansion in a domain dented at R_2. By Theorem 2.13 we get that the number of 2-connected SP graphs on n vertices carrying two spanning trees is asymptotically equal to $\Theta\left(n^{-5/2}R_2^{-n}n!\right)$.

Let $\mathcal{B}_n$ denote the number of 2-connected SP graphs on n vertices and $\mathcal{SB}_n^*$ the number of 2-connected SP graphs on n vertices carrying two spanning trees. The second moment of Z_n can be calculated in the following way:

$$\mathbb{E}[Z_n^2] = \sum_{G\in\mathcal{B}_n} s(G)^2\mathbb{P}[G] = \frac{|\mathcal{SB}_n^*|}{|\mathcal{B}_n|} = \frac{[x^n]B^*(x,1)}{|\mathcal{B}_n|},$$

where again $s(G)$ denotes the number of spanning trees in a graph G. By [33, Theorem 2.6] the number of 2-connected SP graphs on n vertices is asymptotically equal to $\Theta\left(n^{-5/2}r^{-n}n!\right)$, where $r \approx 0.12800$. This means that the second moment of Z_n is asymptotically equal to $\Theta\left(\varpi_2^{-n}\right)$, where $\varpi_2^{-1} \approx 5.31718$. The same holds for $\mathrm{Var}(Z_n) = \mathbb{E}[Z_n^2] - \mathbb{E}[Z_n]^2$ since $\mathbb{E}[Z_n]$ is approximately equal to $\Theta\left(\varrho^{-n}\right)$, where $\varrho^{-1} \approx 2.08415$ by Theorem 6.1. In particular, this implies that Z_n is not concentrated around its expected value.

6.2 Spanning trees in 2-trees

The previous analysis is done over *all* connected/2-connected SP graphs on a given number of vertices. Now we also address the study of extremal situations. First, we particularize the computation of the expected value in the case of a random 2-tree on n vertices, which maximizes the number of edges in an n-vertex SP graph. In this case, the expected value of the number of spanning trees is slightly bigger than the one in Theorem 6.1. In Subsection 6.1.2 we analysed the exponential growth constant of the expected number of spanning trees in a random 2-connected SP graph with respect to its edge density. Observe that if the edge density of SP graphs tends to its maximum, we reach the class of 2-trees. In this section we present an alternative, direct way to compute this growth constant in the setting of 2-trees. More precisely, we give a proof of the following theorem.

Theorem 6.8. *Let U_n denote the number of spanning trees in a labelled 2-tree on n vertices chosen uniformly at random. Then, the expected value of U_n is asymptotically equal to $s_2\varrho_2^{-n}$, where $s_2 \approx 0.14307$ and $\varrho_2^{-1} \approx 2.55561$.*

The proof of Theorem 6.8 is again based on the symbolic method, the extension of the dissymmetry theorem to tree-decomposable classes (Theorem 2.12), and the singularity analysis of generating functions. As in the previous section, we use x and y to mark vertices and edges, respectively. In this particular scenario the number of edges is determined once fixing the number of vertices. However, for pedagogical reasons, we will write all the equations keeping track of both parameters.

6.2.1 Enumerating 2-trees

In order to obtain the expected number of spanning trees in a random 2-tree, we aim to determine an asymptotic estimate of the number of 2-trees as well as of the number of 2-trees carrying a distinguished spanning tree. As a first step, Lemma 6.9 provides the singularity

ϱ_T and the singular expansion in a domain dented at ϱ_T of the generating function $T(x,y)$ associated with the class of 2-trees.

Lemma 6.9. *For $y = 1$ it holds that $T(x,y)$ has the unique square-root type singularity $\varrho_T = 1/(2e)$ and admits the following singular expansion in a domain dented at $x = \varrho_T$:*

$$T(x,1) = \frac{1}{12}e^{-3/2} - \frac{3}{16}e^{-3/2}X^2 + \frac{\sqrt{2}}{48}e^{-3/2}X^3 + \mathcal{O}(X^4),$$

where $X = \sqrt{1 - x/\varrho_T}$.

Proof. Let $\bar{\mathcal{T}}$ denote the class of labelled 2-trees rooted at an edge whose endpoints do not bear a label and let $\overline{T}(x,y)$ denote its associated generating function. By the rules of the symbolic method for pointing operations we have the following relation between $T(x,y)$ and $\overline{T}(x,y)$:

$$y\frac{\partial}{\partial y}T(x,y) = \frac{x^2}{2}\overline{T}(x,y).$$

Observe that we had to add the factor $x^2/2$ on the right-hand side because we had to add labels to the endpoints of the root edge. By integrating by substitution we get that

$$T(x,y) = \frac{x^2}{2}\int_0^y \frac{\overline{T}(x,z)}{z}dz = \frac{x^2}{2}\left(\overline{T}(x,y) - \frac{2}{3}x\overline{T}(x,y)^3\right).$$

In order to compute the radius of convergence of $T(x,y)$, it suffices to compute the radius of convergence of $\overline{T}(x,y)$ since their values coincide by the latter equation. A graph in $\bar{\mathcal{T}}$ can be reconstructed by merging at the root edge a set of pairs of graphs in $\bar{\mathcal{T}}$ that share a vertex, see Figure 6.8 for an illustration.

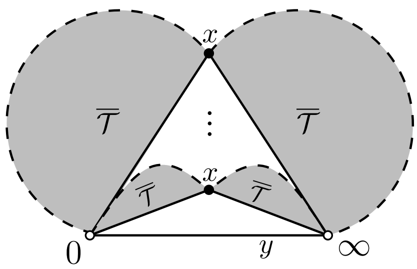

Figure 6.8: Decomposition of rooted 2-trees.

Using the symbolic method this gives rise to the following equation for $\overline{T}(x,y)$:

$$\overline{T}(x,y) = y\exp\left(x\overline{T}(x,y)^2\right).$$

By Lagrange's Theorem (see e.g. [80]) we get for every $n, m \geq 0$ that

$$\begin{aligned}[x^n][y^m]\overline{T}(x,y) &= [x^n]\tfrac{1}{m}[u^{m-1}]e^{mxu^2} \\ &= [x^n]\tfrac{1}{m\cdot(\frac{m-1}{2})!}m^{\frac{m-1}{2}}x^{\frac{m-1}{2}} = \begin{cases}\frac{1}{(\frac{m-1}{2})!}m^{\frac{m-3}{2}} & \text{if } n = \frac{m-1}{2}, \\ 0 & \text{otherwise.}\end{cases}\end{aligned}$$

We study the case $y = 1$ since the number of edges of a 2-tree is determined by its number of vertices. The inverse function of $\overline{T}(x,1)$ is given by $\psi(u) = \log(u)/u^2$. Let $\tau > 0$ such that $\psi'(\tau) = \big(1 - 2\log(\tau)\big)/\tau^3 = 0$, implying $\tau = \exp(1/2)$. Then we know from the analytic inverse function theorem that the singularity ϱ_T of $\overline{T}(x,1)$ satisfies $\varrho_T = \psi(\tau) = 1/(2e)$. Therefore, $\overline{T}(x,1)$ and hence also $T(x,1)$ have a square-root type singular expansion in a domain dented at ϱ_T. By the method of indeterminate coefficients we get that the singular expansion of $T(x,1)$ is of the form as stated in the lemma. Finally, by the aperiodicity of the generating functions under study, the dominant singularity of $T(x,1)$ is unique. □

6.2.2 Enumerating 2-trees carrying spanning trees

Now let us turn to 2-trees carrying a spanning tree. We denote this class by $\mathcal{T}^s$ and its associated generating function by $T^s(x,y)$. Similarly to Section 6.1, we need to analyse edge-maximal SP networks first to get access to the singular behaviour of $T^s(x,y)$. For this purpose, let $D^{\overline{r}}(x,y)$ denote the EGF associated with the class of all edge-maximal SP networks carrying a spanning tree that contains the root edge. Furthermore, let $D^r(x,y)$ denote the EGF associated with the class of all edge-maximal SP networks carrying a spanning tree that does not contain the root edge. Finally, let $D^\circ(x,y)$ denote the EGF associated with the class of all edge-maximal SP networks with a distinguished spanning forest that consists of exactly two components, each of which contains one of the poles. Observe that we have $D^{\overline{r}}(x,y) = D^\circ(x,y)$. Lemma 6.10 gives us the singular behaviours of $D^{\overline{r}} = D^\circ$ and D^r for $y = 1$.

Lemma 6.10. *We have $D^{\overline{r}} = D^\circ$ and for $y = 1$ the generating functions $D^{\overline{r}}$ and D^r have the same unique square-root singularity $R_T \approx 0.07197$. Furthermore, the singular expansions (with rounded coefficients) of $D^{\overline{r}}(x,1)$ and $D^r(x,1)$ in a domain dented at $x = R_T$ are:*

$$D^{\overline{r}}(x,1) = D^{\overline{r}}_0(1) + D^{\overline{r}}_1(1)X + D^{\overline{r}}_2(1)X^2 + D^{\overline{r}}_3(1)X^3 + \mathcal{O}(X^4),$$
$$D^r(x,1) = D^r_0(1) + D^r_1(1)X + D^r_2(1)X^2 + D^r_3(1)X^3 + \mathcal{O}(X^4),$$

where $X = \sqrt{1 - x/R_T}$ and the constants have the following approximate values:

	$i=0$	$i=1$	$i=2$	$i=3$
$D^{\overline{r}}_i(1)$	1.46516	−0.77028	0.53282	−0.40927
$D^r_i(1)$	0.34588	−0.77028	0.87870	−0.92279

Proof. Using the symbolic method (see also Figure 6.9) one can easily verify that the following system of equations hold:

$$\begin{aligned} D^{\overline{r}}(x,y) &= y \exp\Big(2\big(D^{\overline{r}}(x,y) + D^r(x,y)\big)xD^{\overline{r}}(x,y)\Big),\\ D^r(x,y) &= yx\big(D^{\overline{r}}(x,y) + D^r(x,y)\big)^2 \exp\Big(2\big(D^{\overline{r}}(x,y) + D^r(x,y)\big)xD^{\overline{r}}(x,y)\Big). \end{aligned} \tag{6.20}$$

We may set $y = 1$ since the number of edges of 2-trees is always given by the number of vertices. The two equations in (6.20) are defined by entire functions. As these structures carry spanning trees, the corresponding singularity is smaller than ϱ_T, and hence their singularity R_T is finite. We can apply Theorem 2.14 and know therefore that $D^{\overline{r}}(x,1)$ and $D^r(x,1)$ have the same singularity R_T and that they have a square-root singular expansion in a domain

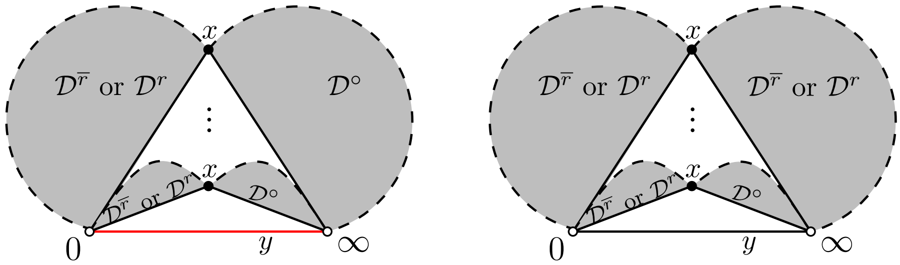

Figure 6.9: Possible decompositions of an edge-maximal SP network with a distinguished spanning tree that contains the root edge (depicted on the left hand side)/does not contain the root edge (depicted on the right hand side).

dented at this singularity. Solving the system of equations stated in Theorem 2.14 with `Maple` yields that $R_T \approx 0.07197$. By the aperiodicity of the counting formulas, this is the unique smallest dominant singularity. Using the method of indeterminate coefficients we get the exact coefficients of the singular expansions of $D^{\overline{r}}(x,1)$ and $D^r(x,1)$, which are the ones as stated in the lemma. □

Lemma 6.11. *For $y = 1$ it holds that $T^s(x,1)$ has a square-root singularity, which is the singularity R_T of $D^{\overline{r}}(x,1)$ and $D^r(x,1)$. Moreover, $T^s(x,1)$ has a singular expansion of the following form in a domain dented at $x = R_T \approx 0.07197$:*

$$T^s(x,1) = T_0^s + T_2^s X^2 + T_3^s X^3 + \mathcal{O}(X^4),$$

where $X = \sqrt{1 - x/R_T}$ and $T_0^s \approx 0.00290$, $T_2^s \approx -0.00669$, and $T_3^s \approx 0.00133$ are computable constants.

Proof. We define the Δe-*tree* $\overline{\tau}(G)$ of a graph G as the bipartite graph describing the incidences between edges and triangles of G. More precisely, the node set of $\overline{\tau}(G)$ is given by $E(G) \cup \big\{\{x_1,x_2,x_3\} : \{x_i,x_j\} \in E(G) \text{ for each } i \neq j \in [3]\big\}$ and two nodes e and Δ are neighbours if and only if $e \subsetneq \Delta$ in G.

Since we are dealing with graphs carrying a spanning tree we need to encode information about the distinguished spanning tree also in the associated Δe-tree. For this reason, we define five different types of nodes of $\overline{\tau}(G)$ if $G \in \mathcal{T}^s$. Let $V_{\overline{e}}$ and V_e denote the set of vertices of $\overline{\tau}(G)$ associated with edges of G that are/are not contained in the distinguished spanning tree of G. Moreover, for $i \in [3]$, let V_{Δ_i} denote the sets of vertices of $\overline{\tau}(G)$ associated with triangles of G that contain exactly $3 - i$ edges that are in the distinguished spanning tree.

Observe that for every $G \in \mathcal{T}^s$ the connected Δe-tree $\overline{\tau}(G)$ associated with it is uniquely defined. We claim that $\overline{\tau}(G)$ is a tree. Indeed, assume for a contradiction that there exists a cycle $C = \Delta^1 e^1 \Delta^2 \dots \Delta^k e^k \Delta^1$ in $\overline{\tau}(G)$. Let G' be the induced subgraph of G on the vertex set $\bigcup_{i \in [k]} \Delta^i$. In particular, being edge-maximal and K_4-minor free, G' is also a 2-tree. However, G' does not contain a vertex of degree 2, a contradiction.

Since $\overline{\tau}(G)$ is a tree, we know that $\mathcal{T}^s$ is a tree-decomposable class. Therefore we can apply Theorem 2.12. Because Δe-trees are bipartite, the equation in Theorem 2.12 simplifies to $\mathcal{T}^s \simeq \mathcal{T}^s_\circ - \mathcal{T}^s_{\circ-\circ}$. The class $\mathcal{T}^s_\circ$ is naturally partitioned into the five classes $\mathcal{T}^s_{\overline{e}}$, $\mathcal{T}^s_e$, $\mathcal{T}^s_{\Delta_1}$, $\mathcal{T}^s_{\Delta_2}$, and $\mathcal{T}^s_{\Delta_3}$ depending on whether the distinguished node is from the set $V_{\overline{e}}$, V_e, V_{Δ_1}, V_{Δ_2}, or

V_{Δ_3}. The class $\mathcal{T}^s_{\circ-\circ}$ is partitioned into the classes $\mathcal{T}^s_{\overline{e}-\Delta_1}$, $\mathcal{T}^s_{\overline{e}-\Delta_2}$, $\mathcal{T}^s_{e-\Delta_1}$, $\mathcal{T}^s_{e-\Delta_2}$, and $\mathcal{T}^s_{e-\Delta_3}$ by the structure of $\overline{\tau}(G)$. In terms of their associated generating functions these facts translate into the following equation

$$T^s = T^s_{\overline{e}} + T^s_e + T^s_{\Delta_1} + T^s_{\Delta_2} + T^s_{\Delta_3} - T^s_{\overline{e}-\Delta_1} - T^s_{\overline{e}-\Delta_2} - T^s_{e-\Delta_1} - T_{e-\Delta_2} - T^s_{e-\Delta_3}. \qquad (6.21)$$

Using the symbolic method, it is not difficult to check that the following equations are true. Recall that we have $D^{\overline{r}}(x,y) = D^{\circ}(x,y)$.

$$\begin{aligned}
&T^s_{\overline{e}}(x,y) = \frac{x^2}{2} D^{\overline{r}}(x,y), \qquad T^s_e = \frac{x^2}{2} D^r(x,y), \qquad T^s_{\Delta_1}(x,y) = \frac{x^3}{2}\left(D^{\overline{r}}(x,y)\right)^3, \\
&T^s_{\Delta_2}(x,y) = x^3 \left(D^{\overline{r}}(x,y)\right)^2 D^r(x,y), \qquad T^s_{\Delta_3}(x,y) = \frac{x^3}{2} D^{\overline{r}}(x,y)\left(D^r(x,y)\right)^2
\end{aligned}$$

and

$$\begin{aligned}
&T^s_{\overline{e}-\Delta_1}(x,y) = x^3\left(D^{\overline{r}}(x,y)\right)^3, && T^s_{\overline{e}-\Delta_2}(x,y) = x^3\left(D^{\overline{r}}(x,y)\right)^2 D^r(x,y), \\
&T^s_{e-\Delta_1}(x,y) = \frac{1}{2} x^3\left(D^{\overline{r}}(x,y)\right)^3, && T^s_{e-\Delta_2}(x,y) = 2x^3\left(D^{\overline{r}}(x,y)\right)^2 D^r(x,y), \\
&T^s_{e-\Delta_3}(x,y) = \frac{3}{2} x^3 D^{\overline{r}}(x,y)\left(D^r(x,y)\right)^2.
\end{aligned}$$

In particular, $T^s(x,y)$ can be expressed in terms of x, $D^{\overline{r}}(x,y)$, and $D^r(x,y)$ by plugging the previous equations in Equation (6.21). Hence one can see that the dominant singularity of $T^s(x,1)$ is the same as the coinciding one of $D^{\overline{r}}(x,1)$ and $D^r(x,y)$, namely R_T. Finally, we obtain the singular expansion of $T^s(x,1)$ in a domain dented at R_T by using Equation (6.21) and the singular expansions of $D^{\overline{r}}(x,1)$ and $D^r(x,1)$ from Lemma 6.10. □

6.2.3 Expected number of spanning trees

Knowing the singular expansions of the generating function $T(x,1)$ associated with the class of 2-trees (Lemma 6.9) as well as of the generating function $T^s(x,1)$ of the class of 2-trees carrying a distinguished spanning tree (Lemma 6.11), we have now all ingredients that we need to finally give a proof of Theorem 6.8.

Proof of Theorem 6.8. Let U_n denote the number of spanning trees in a random 2-tree on n vertices. Then, it holds that

$$\mathbb{E}[U_n] = \sum_{G\in\mathcal{T}_n} s(G)\mathbb{P}[G] = \frac{\sum_{G\in\mathcal{T}_n} s(G)}{|\mathcal{T}_n|} = \frac{|\mathcal{T}^s_n|}{|\mathcal{T}_n|} = \frac{[x^n]T^s(x,1)}{|\mathcal{T}_n|}, \qquad (6.22)$$

where $\mathcal{T}_n$ and $\mathcal{T}^s_n$ denote the class of 2-trees on n vertices and the class of 2-trees on n vertices carrying a distinguished spanning tree, respectively, and $s(G)$ denotes the number of spanning trees in a graph G. By Lemma 6.11 and Theorem 2.13 we get that the number of 2-trees on n vertices that carry a distinguished spanning tree is asymptotically equal to

$$\frac{T^s_3}{\Gamma(-3/2)} n^{-5/2} R_T^{-n} n!\ .$$

Furthermore, it follows from Lemma 6.9 and Theorem 2.13 that the number of 2-trees on n vertices is asymptotically equal to

$$\frac{\sqrt{2}e^{-3/2}}{48\cdot\Gamma(-3/2)}n^{-5/2}\varrho_T^{-n}n!\ .$$

Dividing the former by the latter as in Equation (6.22), we obtain that the expected value of U_n is asymptotically equal to $s_2\varrho_2^{-n}$, where $s_2\approx 0.14307$ and $\varrho_2^{-1}\approx 2.55561$. □

6.3 Spanning trees in series-parallel graphs with fixed excess

Having studied the number of spanning trees in edge-maximal SP graphs in the preceding subsection, we are now interested in the other extremal situation, namely when connected SP graphs have only few edges.

Recall that the *excess* of a graph is the number of its edges minus the number of its vertices. In this subsection we address the following question: given an integer k that does not depend on n, what is the expected number of spanning trees in a random connected SP graph with n vertices and excess equal to k if n is large? In order to study this question we exploit the structure of graphs with fixed excess introduced by Wright [159, 160]. The structure of graphs with fixed excess has been applied successfully in a wide variety of situations (see e.g. [29, 49, 135, 145]).

Our main result in this context is the following polynomial estimate (in n) of the expected number of spanning trees in a random connected SP graph on n vertices and with fixed excess equal to k.

Theorem 6.12. *Let $k>1$ be a fixed integer. Let $X_{n,k}$ denote the number of spanning trees in a connected labelled SP graph, on n vertices and with fixed excess equal to k, chosen uniformly at random. Then, when n is large enough,*

$$\mathbb{E}[X_{n,k}]=\tilde{c}(k)\frac{\Gamma(3k/2)}{\Gamma(2k+1/2)}\left(\frac{n}{2}\right)^{\frac{k+1}{2}}\big(1+o(1)\big),$$

where the function $\tilde{c}(k)$ satisfies the following equation if k is large enough:

$$\tilde{c}(k)=\tilde{c}\tilde{\gamma}^{-k}\big(1+o(1)\big), \tag{6.23}$$

with $\tilde{c}\approx 0.90959$ and $\tilde{\gamma}^{-1}\approx 2.60560$.

In order to deduce these expressions in Theorem 6.12 we analyse weighted cubic SP multigraphs on $2k$ vertices. These objects are reminiscent of the work [101] and are building on previous enumerative results on simple cubic planar graphs [34]. The asymptotic estimate stated in Theorem 6.12 arises when getting asymptotic estimates (in terms of k) for the number of such multigraphs.

The idea of the proof of Theorem 6.12 is as follows. We first provide in Subsection 6.3.1 asymptotically tight formulas for the number of connected SP graphs on n vertices with excess equal to k and for the number of such graphs that carry in addition a distinguished spanning tree. These formulas will be formulated in terms of the number of weighted cubic SP graphs without and with a spanning tree, respectively. In Subsection 6.3.2 we establish asymptotic estimates of these numbers. The proof of Theorem 6.12 then comes down to dividing the asymptotic estimate of the number of connected SP graphs on n vertices with excess k and a distinguished spanning tree by the asymptotic estimate of those graphs without a spanning tree. As in the previous sections, the proofs are based on the symbolic method and uses transfer theorems for the singularity analysis of generating functions.

6.3.1 Kernels of connected SP graphs with fixed excess

Let $\overline{C}_k(x)$ denote the EGF associated with connected SP graphs with excess equal to k carrying a distinguished spanning tree. Furthermore, we denote by $C_k(x)$ the EGF of connected SP graphs with excess equal to k.

The goal of this section is to determine asymptotically tight upper and lower bounds on $\overline{C}_k(x)$ and $C_k(x)$ in terms of the number of weighted cubic SP graphs without and with a spanning tree, respectively. To this end, we also need the EGF for rooted labelled trees, which we denote by $W(x)$. Using the symbolic method one can see that $W(x)$ satisfies the functional equation

$$W(x) = x \exp\big(W(x)\big).$$

Concerning the singular expansion, it is well known and not difficult to check that $W(x)$ has a unique square-root type singularity at $x = e^{-1}$, and in a domain dented at this point, $W(x)$ has an expansion of the form

$$W(x) = 1 - \sqrt{2}X + \mathcal{O}(X^2),$$

where $X = \sqrt{1-ex}$.

Let G be a graph with excess k. Starting from G, we build a multigraph that we call the *kernel* of G in the following way: first we delete recursively vertices of degree one and obtain thus the *core* of the graph (see for instance [135]). Then we continue by dissolving vertices of degree two, i.e. by replacing the two edges incident to a vertex of degree two by a single edge. The resulting connected multigraph $K(G)$ (the *kernel* of G) has minimum degree greater or equal than 3 and fixed excess equal to k. The vertices of the kernel of G can be labelled in the following way: the surviving $|V(K(G))|$ vertices in $K(G)$ carry labels from 1 to $|V(G)|$. The labels induce then an order of the vertices. The final labels in $[|V(K(G))|]$ of the kernel are the ones induced by the ordering of the original labels.

If G has a distinguished spanning tree, the construction of the kernel of G induces also a spanning tree in $K(G)$. Indeed, observe that the edges in G belonging to this distinguished structure induce a spanning tree in its core. Then, in the next step, two edges that are both in the distinguished spanning tree of the core and that are incident to the same vertex of degree 2 are replaced by a single edge that belongs to the distinguished structure of the kernel. If one of the two edges is not in the distinguished spanning tree of the core, then the new single edge does not belong to the distinguished structure of the kernel. Hence, the spanning tree of the core induces a spanning tree of the kernel. In particular, possible loops do not belong to the induced spanning tree of the kernel. Additionally, if there is a multiple edge, at most one of the copies belongs to the spanning tree. See an example of the construction of the kernel in Figure 6.10 together with the distinguished spanning structure.

It is straightforward how to reverse this construction: starting from a given kernel C one can build directly the initial graph by pasting over each vertex of C a rooted labelled tree and

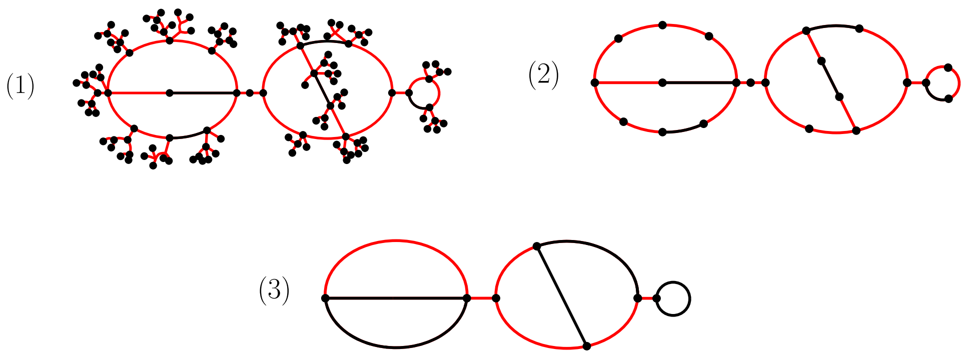

Figure 6.10: Construction of the kernel of an SP graph with a distinguished spanning tree.

by substituting each edge by a sequence of rooted labelled trees. In our approach, dealing with spanning trees, we need to consider a slight modification for edges not belonging to the spanning tree. More precisely, we need to substitute in this case each edge (neither a loop nor a multiple edge) by

$$1 + 2W(x) + 3W(x)^2 + \cdots = \sum_{r\geq 0}(r+1)W(x)^r = \frac{1}{(1-W(x))^2},$$

whose singular expansion in a domain dented at $x = e^{-1}$ is of the form $\frac{1}{2}X^{-2} + \mathcal{O}(X^{-1})$, where again $X = \sqrt{1-ex}$. The reason for this is that whenever we paste a sequence of rooted trees over the edge under consideration we must maintain the spanning structure acyclic and connected. If, in the sequence, we paste r different trees, the edge becomes subdivided into $r+1$ edges. We need to choose then which one of these $r+1$ edges is not in the spanning tree.

In the case of the loop the previous sequence must start at the term $2W(x)$ in order to obtain at the end a simple graph. Similarly, for multiple edges we need to assure that at the end we obtain a simple graph. This case-by-case analysis suggests that generating functions would be somehow involved (see the similar problem in general graphs in [159, 160]). However, we show that we are able to find closed formulas for the asymptotic estimates by means of a sequence of reductions.

In the problem that we study, our graph G is a connected SP graph. Hence, its kernel $K(G)$ is an SP multigraph, i.e. a K_4-minor free multigraph. Note that the number of possible kernels of excess k does not depend on the number of vertices of G. Moreover, observe that for a fixed value of k, a multigraph K with minimum degree at least 3 and of excess k maximizes its number of edges if and only if K is cubic. Hence, we can restrict ourselves to the study of connected SP graphs arising from a cubic kernel as these ones will provide the main contribution to the asymptotics.

For technical reasons we will deal with weighted cubic multigraphs. The *weight* (called the *compensation factor* in [97], see also [101]), has the following meaning. When we substitute edges of the kernel by sequences of rooted trees, a loop has two possible orientations that give the same simple graph. A double (triple) edge can be permuted in two (six) ways and still produce the same object. The enumerative study of weighted cubic multigraphs will be deferred to the end of this section. We denote by g_k the number of weighted cubic SP multigraphs with excess k. Additionally, we denote by $\overline{g}_k$ the number of weighted cubic SP multigraphs with excess k that carry a distinguished spanning tree. In the rest of this section, all cubic multigraphs are considered to be weighted.

Observe that a cubic multigraph of excess k has $2k$ vertices and $3k$ edges. Therefore, any spanning tree has $2k-1$ edges and hence $k+1$ edges are *not* used in the spanning structure. The first lemma gives a lower bound for $[x^n]\overline{C}_k(x)$:

Lemma 6.13. *For fixed $k > 1$, the following inequality holds:*

$$\overline{g}_k[x^n]\frac{\big(3-2W(x)\big)^{k+1}W(x)^{6k}}{\big(1-W(x)\big)^{4k+1}} < [x^n]\overline{C}_k(x). \tag{6.24}$$

Proof. The proof is reminiscent to the proof of Lemma 3 in [136]. The EGF on the left hand side of Equation (6.24) can be written in the following way:

$$W(x)^{2k}\frac{W(x)^{2k-1}}{\big(1-W(x)\big)^{2k-1}} \cdot \frac{\big(3-2W(x)\big)^{k+1}W(x)^{2k+2}}{\big(1-W(x)\big)^{2k+2}}.$$

This can be interpreted as follows: for a given cubic SP multigraph with a distinguished spanning tree we

(a) paste a rooted labelled tree over each of the $2k$ vertices,

(b) paste a sequence of at least one rooted labelled tree over each of the $2k-1$ edges belonging to the spanning tree of the kernel.

(c) paste a sequence of at least one rooted labelled tree over each of the $k+1$ edges not belonging to the spanning tree of the kernel, and then decide which of the new edges does not belong to the resulting spanning tree.

Points (a) and (b) contribute with terms $W(x)^{2k}$ and $(W(x)/(1-W(x)))^{2k-1}$, respectively, where the second term is associated with a sequence of at least one rooted labelled tree. For the computation of Point (c), recall that loops are not in spanning trees. Hence, to guarantee that the final graph is simple we need sequences of length at least two for edges that do not belong to the spanning tree (length one is enough for multiple edges, but length two is needed for loops). Hence, the computation in point (c) arise from the fact that

$$\sum_{r\geq 2}(r+1)W(x)^r = \frac{\big(3-2W(x)\big)W(x)^2}{\big(1-W(x)\big)^2}.$$

This construction is injective and does *not* give all possible connected SP graphs with excess k. Summing over all possible weighted cubic SP multigraphs we obtain the result as claimed. □

The next step is to get an upper bound for $[x^n]\overline{C}_k(x)$. This is provided by the following lemma:

Lemma 6.14. *For fixed $k > 1$, the following inequality holds:*

$$[x^n]\overline{C}_k(x) < \overline{g}_k\left([x^n]\frac{W(x)^{2k}}{\big(1-W(x)\big)^{4k+1}}\right)\big(1+o(1)\big). \tag{6.25}$$

Proof. The statement is proved by applying a similar argument to the one of Lemma 6.13. For a fixed cubic multigraph with a distinguished spanning tree, we now paste over each edge an arbitrary sequence of rooted trees (possibly empty), and take care of the special requirement on the edges of the kernel that do not belong to the initial spanning tree. In this way we generate *all* connected SP graphs with excess k that carry a spanning tree. Observe that it is possible to generate graphs that are not simple. Hence, this construction only gives an upper bound. Finally, the term $o(1)$ in Equation (6.25) arises from the set of kernels that are not cubic. □

We can now combine both lemmas to get an asymptotic estimate for $[x_n]\overline{C}_k(x)$.

Proposition 6.15. *For fixed $k > 1$, the following asymptotic estimate in n holds:*

$$[x^n]\overline{C}_k(x) = \overline{g}_k \frac{1}{2^{2k+1/2}} \frac{n^{2k-1/2}}{\Gamma(2k+1/2)} e^n \big(1 + o(1)\big). \tag{6.26}$$

Proof. We apply the transfer theorem for singularity analysis (Theorem 2.13) to Equations (6.24) and (6.25). In particular we get that

$$[x^n]\overline{C}_k(x) > \overline{g}_k [x^n] \frac{\big(3 - 2W(x)\big)^{k+1} W(x)^{6k}}{\big(1 - W(x)\big)^{4k+1}} = \overline{g}_k \frac{1}{2^{2k+1/2}} \frac{n^{2k-1/2}}{\Gamma(2k+1/2)} e^n \big(1 + o(1)\big),$$

and concerning the upper bound,

$$[x^n]\overline{C}_k(x) < \overline{g}_k \left([x^n] \frac{W(x)^{2k}}{\big(1 - W(x)\big)^{4k+1}} \right) \big(1 + o(1)\big) = \overline{g}_k \frac{1}{2^{2k+1/2}} \frac{n^{2k-1/2}}{\Gamma(2k+1/2)} e^n \big(1 + o(1)\big).$$

In both estimates we have exploited the fact that $W(e) = 1$. As these estimates have the same singular behaviour, we conclude the estimate in Equation (6.26). □

Before moving to the computation of $\overline{g}_k$ and g_k, let us mention that similar arguments give estimates for $[x^n]C_k(x)$. Indeed, using the argument in [136, Lemma 3] (or *mutatis mutandis* the previous arguments for $\overline{C}_k(x)$) one gets the following upper and lower bounds for $[x^n]C_k(x)$:

$$g_k [x^n] \frac{W(x)^{8k}}{\big(1 - W(x)\big)^{3k}} < [x^n]C_k(x) < g_k \left([x^n] \frac{W(x)^{2k}}{\big(1 - W(x)\big)^{3k}} \right) \big(1 + o(1)\big).$$

Again, by applying the transfer theorem for singularity analysis (Theorem 2.13) we get the estimate

$$[x^n]C_k(x) = g_k \frac{1}{2^{3k/2}} \frac{n^{3k/2-1}}{\Gamma(3k/2)} e^n \big(1 + o(1)\big). \tag{6.27}$$

6.3.2 Enumerating weighted cubic SP multigraphs with spanning trees

We complete the picture by getting asymptotic formulas (in k) for both g_k and $\overline{g}_k$. The main results we implicitly use are the transfer theorem (Theorem 2.13), joint with the fact that singular expansions can be integrated on dented domains (see e.g. [80, Theorem VI.9]).

Additionally, given a value k, one can obtain the corresponding values by getting the Taylor expansions of the generating functions that will be introduced in the next lines.

Let us start studying g_k. As mentioned before, g_k is the number of connected cubic SP multigraphs with weights and excess equal to k. The weights are defined rigorously in the following way: given a multigraph with l_1 loops, l_2 double edges and l_3 triple edges, its *weight* is $2^{-l_1-l_2}6^{-l_3}$. Weights are needed to encode edge symmetries of multigraphs. Let $G(u)$ be the EGF of connected weighted cubic SP multigraphs, where the variable u marks the excess. $G(u)$ satisfies the following system of functional equations:

$$\begin{aligned}
6u\frac{dG(u)}{du} &= d(u)+c(u),\\
b(u) &= \frac{u}{2}\big(d(u)+c(u)\big)+\frac{u}{2},\\
c(u) &= s(u)+p(u)+b(u),\\
d(u) &= \frac{b(u)^2}{u},\\
s(u) &= c(u)^2-c(u)s(u),\\
p(u) &= uc(u)+\frac{1}{2}uc(u)^2+\frac{u}{2}.
\end{aligned} \tag{6.28}$$

Full details concerning these equations can be found in [136, Section 3], building on results on [101] (see also [34]). The meaning of these EGF is the following: the term $6u\frac{dG(u)}{du}$ corresponds to the EGF of connected weighted cubic SP multigraphs where an edge (the *root edge*) is marked and oriented (remember that a cubic multigraph of excess k has $3k$ edges and each edge has 2 possible orientations). We have the following cases depending on the nature of the root edge: either the root edge is a loop (term $b(u)$) or a bridge (term $d(u)$) or when deleting it we get a series construction (term $s(u)$) or a parallel construction (term $p(u)$). The term $c(u)$ plays the role of an auxiliary EGF.

We proceed now with the analysis of this system.

Proposition 6.16. *The number $g_k=[u^k]G(u)$ of weighted connected cubic SP multigraphs with fixed excess equal to k has the following asymptotic estimate:*

$$g_k = ck^{-5/2}\gamma^{-k}\big(1+o(1)\big),$$

where $\gamma=\frac{4}{27}\sqrt{6\sqrt{3}-9}\approx 0.17481$, and $c\approx 0.06034$.

Proof. From the system of equations (6.28) we get a single equation for $c(u)$ by successive elimination. Computations give that $c(u)$ satisfies the following algebraic equation

$$\begin{aligned}
0 \;=\; & 8u+u^2+(-8+24u+6u^2)c(u)+(-4+24u+15u^2)c(u)^2+(8u+20u^2)c(u)^3\\
& +15u^2c(u)^4+6u^2c(u)^5+u^2c(u)^6.
\end{aligned}$$

We refer to [80, Section VII.8] for more details. From this equation we deduce that the dominant singularity of $c(u)$ is the smallest positive root of the equation

$$19683u^4+7776u^2-256=0,$$

whose value is equal to $\gamma = \frac{4}{27}\sqrt{6\sqrt{3}-9} \approx 0.17481$ ($-\gamma$ is also a root of this polynomial with the same absolute value, but it is easy to check that $c(u)$ is analytic at $u = -\gamma$). Using now Newton's Polygon Method (see [80, Section VII.7]) we get that $c(u)$ has a Puiseux's expansion of the following form in a domain dented at $u = \gamma$:

$$c(u) = c_0 + c_1 U + \mathcal{O}(U^2), \tag{6.29}$$

where $U = \sqrt{1 - u/\gamma}$, $c_0 \approx 0.61185$, and $c_1 \approx -1.08766$. Using Expansion (6.29) we deduce that $D(u)$ admits the following singular expansion in a domain dented at $u = \gamma$:

$$d(u) = d_0 + d_1 U + \mathcal{O}(U^2)$$

with $d_0 \approx 0.13306$ and $d_1 \approx -0.19574$. Finally, using that $6u\frac{dG(u)}{du} = d(u) + c(u)$ we conclude that the dominant singularity of $G(u)$ is located at $u = \gamma$. The proposition finally follows by integration of the Puiseux's series (by applying [80, Theorem VI.9]) in order to get the singular expansion of $G(u)$ and by the application of the transfer theorem (Theorem 2.13) $\square$

Let us now study $\overline{g}_k$. For this purpose, we refine the system of equations (6.28) in the following way: we denote by $\overline{G}(u)$ the EGF associated with connected weighted cubic SP multigraphs carrying a distinguished spanning tree, where, as before, u marks the excess. We study decompositions for $6u\frac{d\overline{G}(u)}{du}$, which correspond to the EGF of connected weighted cubic SP multigraphs carrying a distinguished spanning tree where an edge is distinguished and oriented.

In the following expressions we use the subindex 0 to denote that the distinguished and oriented edge belongs to the spanning tree, while we use the subindex 1 to denote the opposite. Using this convention, we write $d_0(u)$, $b_0(u)$, $s_0(u)$ and $p_0(u)$ for the cases where this distinguished edge is a bridge, a loop, defines a series construction or a parallel construction, respectively, in such a way that the distinguished edge belongs to the spanning tree of the initial structure. Analogue definitions are done for $d_1(u)$, $b_1(u)$, $s_1(u)$ and $p_1(u)$. The EGFs $c_0(u)$ and $c_1(u)$ are associated with auxiliary families.

Using the same arguments used to obtain the system of equations (6.28) yields the following, more involved system of equations:

$$\begin{aligned}
6u\frac{d\overline{G}(u)}{du} &= d_0(u) + c_0(u) + d_1(u) + c_1(u),\\
c_0(u) &= b_0(u) + s_0(u) + p_0(u),\\
c_1(u) &= b_1(u) + s_1(u) + p_1(u),\\
b_0(u) &= 0,\\
b_1(u) &= \frac{u}{2}\big(d_0(u) + c_0(u)\big) + u\big(d_1(u) + c_1(u)\big) + \frac{u}{2},\\
d_0(u) &= \frac{b_1(u)^2}{u},\\
d_1(u) &= 0,\\
s_0(u) &= \big(c_1(u) - s_1(u)\big)c_1(u) + \big(c_0(u) - s_0(u)\big)c_1(u) + \big(c_1(u) - s_1(u)\big)c_0(u),\\
s_1(u) &= \big(c_1(u) - s_1(u)\big)c_1(u),\\
p_0(u) &= \frac{u}{2} + uc_0(u) + 2uc_1(u) + \frac{1}{2}uc_0(u)^2,\\
p_1(u) &= u + uc_0(u) + 3uc_1(u) + uc_0(u)c_1(u) + 2uc_1(u)^2.
\end{aligned} \tag{6.30}$$

We now analyse this system of equations similarly to what we did when studying System (6.28):

Proposition 6.17. *The number $\overline{g}_k = [u^k]\overline{G}(u)$ of weighted connected cubic SP multigraphs carrying a spanning tree and with excess equal to k has the following asymptotic estimate:*

$$\overline{g}_k = \overline{c}k^{-5/2}\overline{\gamma}^{-k}\big(1+o(1)\big),$$

where $\overline{\gamma} \approx 0.06709$ and $\overline{c} \approx 0.06634$.

Proof. The arguments are exactly the same as in Proposition 6.16. From the system of equations (6.30) we get the following algebraic equation satisfied by $c_0(u)$:

$$\begin{aligned} 0 \;=\;& 121u^3 + 2304u + 7656u^2 \\ &+(51696u - 26620u^2 - 968u^3 - 4608)c_0(u) \\ &+(-384 + 34532u + 30888u^2 + 3388u^3)c_0(u)^2 \\ &+(256 - 1392u - 10516u^2 - 6776u^3)c_0(u)^3 \\ &+(32 - 1608u - 2376u^2 + 8470u^3)c_0(u)^4 \\ &+(-352u - 132u^2 - 6776u^3)c_0(u)^5 \\ &+(4u + 1144u^2 + 3388u^3)c_0(u)^6 \\ &+(-44u^2 - 968u^3)c_0(u)^7 + 121u^3c_0(u)^8. \end{aligned}$$

The singular point is located at $\overline{\gamma} \approx 0.06709$, and the Puiseux's expansion of $c_0(u)$ around $u = \overline{\gamma}$ is of the form

$$c_0(u) = c_{0,0} + c_{0,1}\overline{W} + \mathcal{O}(\overline{W}^2),$$

where $\overline{W} = \sqrt{1 - u/\overline{\gamma}}$, $c_{0,0} \approx 0.29896$, and $c_{0,1} \approx -0.47032$. We can then directly obtain the Puiseux's expansion of $c_1(u)$, $b_1(u)$ and $d_1(u)$ from the expansion of $c_0(u)$. By integrating the equation $6u\frac{d}{du}\overline{G}(u) = c_0(u)+c_1(u)+d_0(u)$ and applying the transfer theorem (Theorem 2.13) we get the estimate for $\overline{g}_k$ as it is claimed. □

6.3.3 Expected number of spanning trees

Joining the results from the previous two subsections, we can now complete the proof of Theorem 6.12.

Proof of Theorem 6.12. Let $\mathcal{C}_{n,k}$ denote the class of all connected SP graphs on n vertices with excess k and let $s(G)$ denote the number of spanning trees of a graph G. Due to Proposition 6.15 and Equation (6.27), the value $\mathbb{E}[X_{n,k}]$ is

$$\begin{aligned} \mathbb{E}[X_{n,k}] &= \sum_{G\in\mathcal{C}_{n,k}} s(G)\mathbb{P}[G] = \frac{\sum_{G\in\mathcal{C}_{n,k}} s(G)}{|\mathcal{C}_{n,k}|} = \frac{[x^n]\overline{C}_k(x)}{[x^n]C_k(x)} \\ &= \frac{\overline{g}_k \frac{1}{2^{2k+1/2}}\frac{n^{2k-1/2}}{\Gamma(2k+1/2)}e^n}{g_k\frac{1}{2^{3k/2}}\frac{n^{3k/2-1}}{\Gamma(3k/2)}e^n}(1+o(1)) = \frac{\overline{g}_k}{g_k}\frac{\Gamma(3k/2)}{\Gamma(2k+1/2)}\left(\frac{n}{2}\right)^{\frac{k+1}{2}}(1+o(1)), \end{aligned}$$

where, by using now the estimates obtained for g_k and $\overline{g}_k$ in Propositions 6.16 and 6.17, the function $\tilde{c}(k)$ satisfies the following equation if k is large enough:

$$\tilde{c}(k) = \tilde{c}\tilde{\gamma}^{-k}\big(1+o(1)\big), \tag{6.31}$$

with $\tilde{c} \approx 0.90959$ and $\tilde{\gamma}^{-1} \approx 2.60560$. □

6.4 Concluding remarks

In this chapter we have exploited the use of generating functions joint with analytic combinatorics to get exact expressions for the counting generating functions associated with connected SP graphs as well as with 2-connected SP graphs, 2-trees, and SP graphs with fixed excess in all cases with a distinguished spanning tree. As a consequence, we were able to get asymptotic estimates for the expected number of spanning trees in an object chosen uniformly at random from the family under study.

These techniques could be exploited in related families of graphs, as for instance the family of k-trees with $k \geq 3$. Even though the analytic techniques that we used allow access to higher moments, unfortunately, they do not provide a technique to determine the limit law of the number of spanning trees in SP graphs.

With the techniques used in this chapter one can also determine, with a bit more effort, the counting generating function of connected SP graphs carrying a distinguished spanning forest with a given number of components. From this, one can derive the expected number of components of a random spanning forest in a random connected SP graph. In particular, the main difficulty in this situation is that for encoding networks carrying a distinguished spanning forest, one needs more auxiliary classes than in the restricted case of spanning trees. Roughly speaking, one needs to define classes of SP networks that carry a spanning forest and such that either the two poles are contained in the same component or in different components.

The analysis of the generating function associated with networks as well as determining and analysing the generating functions associated with 2-connected and connected SP graphs carrying a spanning forest can be done similarly to the case of spanning trees. It is worth mentioning that Bousquet-Mélou and Courtiel [43] recently investigated the enumeration of regular planar maps carrying a distinguished spanning forest.

Finally, we would like to discuss similar results on planar graphs. Following the lines of [84], the tools developed in this chapter can be extended to graphs defined by 3-connected components. When the family under consideration is defined by a finite number of 3-connected graphs, the techniques used so far can be exploited to get analogue results. This would include, for instance, the family of graphs $\mathrm{Ex}(W_4)$ or $\mathrm{Ex}(W_5)$, where W_4 and W_5 are the wheel graphs with 4 and 5 external vertices, respectively (see [84]).

An intriguing problem is to extend the results of SP graphs to the random planar graph model. In this situation, the family of 3-connected components is infinite, and one requires additional results arising from map enumeration in order to control counting formulas for T-bricks. Although the number of rooted unlabelled maps carrying a spanning tree is a well-known fact (see e.g. [132]), to the best of our knowledge nothing is known when dealing with 3-connected planar graphs. The problem of getting the expected number of spanning trees in a planar graph chosen uniformly at random is an interesting problem to be considered in future investigations.

Acknowledgements

First of all, I would like to express my sincere gratitude to my supervisor Anusch Taraz for supporting me throughout the PhD study, for always taking the time to answer my questions even if his calender was crammed with appointments, for teaching me so many fascinating results and tools, for both the joint work and for the freedom to explore different fields, for being a wonderful, humorous, uncomplicated advisor, and for so much more concerning research and beyond.

My sincere thanks also go to my co-authors Peter Allen, Julia Böttcher, Dennis Clemens, Yury Person, Alexey Pokrovskiy, Juanjo Rué, Jakob Schnitzer, and Tuan Tran. Working with every single one of you was delightful. I am especially grateful for having been able to visit Peter Allen and Julia Böttcher at London School of Economics, Juanjo Rué at FU Berlin, and Yury Person at Goethe Universität Frankfurt. Thank you for sharing your immense knowledge, for support, inspiration, fruitful discussions, and enjoyable conversations.

I am especially grateful to Tibor Szabó for inviting me to his group workshop three years ago. That week was not only instructive, productive and fun but also an important cornerstone of my thesis as it was the starting point for various collaborations.

For reading parts of the thesis and valuable comments I would like to thank Dennis Clemens, Tina Janne Schmidt, Carl Georg Heise, Stefan König, and Damian Reding.

TopMath made it possible for me to encounter research at an early stage of my studies. I am truly thankful for the support I received within this program.

During the last years I was blessed with amazing office mates. Thank you, Jenna Thompson, for brightening up each day with your incredibly high spirits. Thank you, Dennis Clemens, for nice chats across our desks, our walks to the Campus Suite whenever there was need for coffee or chocolate, and for sharing ideas and insights.

For a familiar atmosphere I am grateful to all my colleagues at TU Hamburg and former colleagues at TU München. In particular I would like to thank Gisela Winterstein for interesting conversations about everything under the sun and for indispensable help when it comes to bureaucracy.

Moreover, I owe thanks to my parents, my brother, and my close relatives for supporting me in uncountably many ways, for listening to my problems, and encouraging me.

Finally, I am deeply grateful to Stefan König for encouraging and supporting me in every respect throughout this time. Thank you for always being there for me!

Bibliography

[1] S. Abbasi. How tight is the Bollobás-Komlós conjecture? *Graphs and Combinatorics*, 16:129–137, 2000.

[2] R. Aharoni and E. Berger. Rainbow matchings in r-partite r-graphs. *Electronic Journal of Combinatorics*, 16(1):R119, 2009.

[3] R. Aharoni, P. Charbit, and D. Howard. On a Generalization of the Ryser-Brualdi-Stein Conjecture. *Journal of Graph Theory*, 78(2):143–156, 2015.

[4] R. Aharoni, D. Kotlar, and R. Ziv. Rainbow sets in the intersection of two matroids. *Electronic Notes in Discrete Mathematics*, 43:39–42, 2013.

[5] R. Aharoni, D. Kotlar, and R. Ziv. Representation of large matchings in bipartite graphs. *arXiv:1601.00943*, 2016.

[6] P. Allen, J. Böttcher, J. Ehrenmüller, J. Schnitzer, and A. Taraz. A spanning bandwidth theorem in random graphs. Manuscript in preparation.

[7] P. Allen, J. Böttcher, J. Ehrenmüller, and A. Taraz. The bandwidth theorem in sparse graphs. Manuscript in preparation.

[8] P. Allen, J. Böttcher, J. Ehrenmüller, and A. Taraz. Local resilience of spanning subgrahs in sparse random graphs. *Electronic Notes in Discrete Mathematics*, 49:513–421, 2015.

[9] P. Allen, J. Böttcher, H. Hàn, Y. Kohayakawa, and Y. Person. Blow-up lemmas for sparse graphs. *Manuscript in preparation.*

[10] P. Allen, J. Böttcher, H. Hàn, Y. Kohayakawa, and Y. Person. Powers of Hamilton cycles in pseudorandom graphs. In A. Pardo and A. Viola, editors, *LATIN 2014: Theoretical Informatics*, pages 355–366. Springer, Berlin Heidelberg, 2014.

[11] P. Allen, J. Böttcher, Y. Kohayakawa, and Y. Person. Tight Hamilton cycles in random hypergraphs. *Random Structures & Algorithms*, 46(3):446–465, 2015.

[12] P. Allen, J. Böttcher, J. Skokan, and M. Stein. Regularity inheritance in pseudorandom graphs. Manuscript in preparation.

[13] N. Alon. The number of spanning trees in regular graphs. *Random Structures & Algorithms*, 1(2):175–181, 1990.

[14] N. Alon, M. Capalbo, Y. Kohayakawa, V. Rödl, A. Ruciński, and E. Szemerédi. Universality and tolerance. In *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on*, pages 14–21. 2000.

[15] N. Alon and J. H. Spencer. *The probabilistic method*. John Wiley & Sons, 3rd edition, 2008.

[16] S. Arnborg. Efficient algorithms for combinatorial problems on graphs with bounded decomposability — A survey. *BIT Numerical Mathematics*, 25(1):1–23, 1985.

[17] J. Balogh, B. Csaba, and W. Samotij. Local resilience of almost spanning trees in random graphs. *Random Structures & Algorithms*, 38(1-2):121–139, 2011.

[18] J. Balogh, C. Lee, and W. Samotij. Corrádi and Hajnal's theorem for sparse random graphs. *Combinatorics, Probability and Computing*, 21(1-2):23–55, 2012.

[19] J. Barát, A. Gyárfás, and G. N. Sárközy. Rainbow matchings in bipartite multigraphs. *arXiv:1505.01779*, 2015.

[20] J. Barát and I. M. Wanless. Rainbow matchings and transversals. *Australasian Journal of Combinatorics*, 59(1):211–217, 2014.

[21] G. Baxter. On fixed points of the composite of commuting functions. *Proceedings of the American Mathematical Society*, 15(6):851–855, 1964.

[22] J. Beck. Remarks on positional games. I. *Acta Math. Acad. Sci. Hungar.*, 40(1-2):65–71, 1982.

[23] J. Beck. *Combinatorial games: Tic-Tac-Toe theory*. Cambridge University Press, New York, 2008.

[24] C. Berge. *Graphs and hypergraphs*, volume 7. North-Holland, 1973.

[25] F. Bergeron, G. Labelle, and P. Leroux. *Combinatorial species and tree-like structures*, volume 67. Cambridge University Press, 1998.

[26] J. C. Bermond, A. Germa, M. C. Heydemann, and D. Sotteau. Hypergraphes hamiltoniens. In *Colloques Internationaux CNRS, Univ. Orsay*, volume 260, pages 39–43. 1976.

[27] O. Bernardi. Bijective counting of tree-rooted maps and shuffles of parenthesis systems. *Electronic Journal of Combinatorics*, 14:R9, 2007.

[28] O. Bernardi. A characterization of the Tutte polynomial via combinatorial embeddings. *Annals of Combinatorics*, 12(2):139–153, 2008.

[29] O. Bernardi and J. Rué. Enumerating simplicial decompositions of surfaces with boundaries. *European Journal of Combinatorics*, 33(3):302–325, 2012.

[30] N. Bernasconi, K. Panagiotou, and A. Steger. The degree sequence of random graphs from subcritical classes. *Combinatorics, Probability and Computing*, 18(5):647–681, 2009.

[31] N. Biggs. *Algebraic Graph Theory.* Cambridge University Press, 2nd edition, 1993.

[32] Y. Bilu and N. Linial. Lifts, discrepancy and nearly optimal spectral gap. *Combinatorica*, 26(5):495–519, 2006.

[33] M. Bodirsky, O. Giménez, M. Kang, and M. Noy. Enumeration and limit laws for series–parallel graphs. *European Journal of Combinatorics*, 28(8):2091–2105, 2007.

[34] M. Bodirsky, M. Kang, M. Löffler, and C. McDiarmid. Random cubic planar graphs. *Random Structures & Algorithms*, 30(1-2):78–94, 2007.

[35] B. Bollobás. Almost all Regular Graphs are Hamiltonian. *European Journal of Combinatorics*, 4(2):97–106, 1983.

[36] B. Bollobás. *Random graphs.* Springer, New York, 1998.

[37] B. Bollobás and A. G. Thomason. Threshold functions. *Combinatorica*, 7(1):35–38, 1987.

[38] M. Bóna (editor). *Handbook of Enumerative Combinatorics.* CRC Press, 2015.

[39] J. Böttcher, Y. Kohayakawa, and A. Taraz. Almost spanning subgraphs of random graphs after adversarial edge removal. *Combinatorics, Probability and Computing*, 22(05):639–683, 2013.

[40] J. Böttcher, K. P. Pruessmann, A. Taraz, and A. Würfl. Bandwidth, expansion, treewidth, separators and universality for bounded-degree graphs. *European Journal of Combinatorics*, 31(5):1217–1227, 2010.

[41] J. Böttcher, M. Schacht, and A. Taraz. Proof of the bandwidth conjecture of Bollobás and Komlós. *Mathematische Annalen*, 343(1):175–205, 2009.

[42] J. Böttcher, A. Taraz, and A. Würfl. Spanning embeddings of arrangeable graphs with sublinear bandwidth. *Random Structures & Algorithms*, 48(2):270–289, 2016.

[43] M. Bousquet-Mélou and J. Courtiel. Spanning forests in regular planar maps. *Journal of Combinatorial Theory, Series A*, 135:1–59, 2015.

[44] A. Brandstädt, V. B. Le, and J. P. Spinrad. *Graph classes: a survey.* Society for Industrial and Applied Mathematics, 1999.

[45] A. E. Brouwer, A. J. de Vries, and R. M. A. Wieringa. A lower bound for the length of partial transversals in a latin square. *Nieuw Archief Voor Wiskunde*, 26:330–332, 1978.

[46] R. A. Brualdi and H. J. Ryser. *Combinatorial matrix theory.* Cambridge University Press, 1991.

[47] E. Buß, H. Hàn, and M. Schacht. Minimum vertex degree conditions for loose Hamilton cycles in 3-uniform hypergraphs. *Journal of Combinatorial Theory, Series B*, 103(6):658–678, 2013.

[48] A. Cayley. A theorem on trees. *Quart. J. Math.*, 23:376–378, 1889.

[49] G. Chapuy. The structure of dominant unicellular maps, and a connnection between maps of positive genus and planar labelled trees. *Probability Theory and Related Fields*, 147(3):415–447, 2010.

[50] G. Chapuy, É. Fusy, M. Kang, and B. Shoilekova. A complete grammar for decomposing a family of graphs into 3-connected components. *Electronic Journal of Combinatorics*, 15(1):R148, 2008.

[51] F. R. K. Chung. Labelings of graphs. In *Selected Topics in Graph Theory*, volume 3, pages 151–168. Academic Press, San Diego, 1988.

[52] V. Chvátal and P. Erdős. Biased positional games. *Ann. Discrete Math.*, 2:221–229, 1978. Algorithmic aspects of combinatorics (Conf., Vancouver Island, B.C., 1976).

[53] D. Clemens and J. Ehrenmüller. An improved bound on the sizes of matchings guaranteeing a rainbow matching. *Electronic Journal of Combinatorics*, 23(2):P2.11, 2016.

[54] D. Clemens, J. Ehrenmüller, and Y. Person. A Dirac-type theorem for Berge cycles in random hypergraphs. Manuscript in preparation.

[55] D. Clemens, J. Ehrenmüller, Y. Person, and T. Tran. Keeping Avoider's graph almost acyclic. *Electronic Journal of Combinatorics*, 22(1):P1.60, 2015.

[56] D. Clemens, J. Ehrenmüller, and A. Pokrovskiy. On sets not belonging to algebras and rainbow matchings in graphs. *arXiv:1508.06437*.

[57] R. Cori, S. Dulucq, and G. Viennot. Shuffle of parenthesis systems and baxter permutations. *Journal of Combinatorial Theory, Series A*, 43(1):1 – 22, 1986.

[58] K. Corrádi and A. Hajnal. On the maximal number of independent circuits in a graph. *Acta Mathematica Hungarica*, 14(3):423–439, 1963.

[59] D. Dellamonica, Y. Kohayakawa, M. Marciniszyn, and A. Steger. On the resilience of long cycles in random graphs. *Electronic Journal of Combinatorics*, 15:R32, 2008.

[60] D. Dellamonica, Y. Kohayakawa, V. Rödl, and A. Ruciński. An improved upper bound on the density of universal random graphs. *Random Structures & Algorithms*, 46(2):274–299, 2014.

[61] J. Dénes and A. D. Keedwell. *Latin squares and their applications.* Academic Press, New York, 1974.

[62] R. Diestel. *Graph Theory.* Springer, 4th edition, 2010.

[63] G. A. Dirac. Some theorems on abstract graphs. *Proceedings of the London Mathematical Society*, 3(1):69–81, 1952.

[64] D. A. Drake. Maximal sets of Latin squares and partial transversals. *Journal of Statistical Planning and Inference*, 1(2):143–149, 1977.

[65] A. A. Drisko. Transversals in Row-Latin Rectangles. *Journal of Combinatorial Theory, Series A*, 84(2):181–195, 1998.

[66] M. Drmota. Systems of functional equations. *Random Structures & Algorithms*, 10(1-2):103–124, 1997.

[67] M. Drmota. *Random trees: an interplay between combinatorics and probability.* Springer, Wien, New York, 2009.

[68] M. Drmota, E. Fusy, M. Kang, V. Kraus, and J. Rué. Asymptotic study of subcritical graph classes. *SIAM Journal on Discrete Mathematics*, 25(4):1615–1651, 2011.

[69] M. Drmota, O. Giménez, and M. Noy. Vertices of given degree in series-parallel graphs. *Random Structures & Algorithms*, 36(3):273–314, 2010.

[70] M. Drmota, O. Giménez, and M. Noy. The maximum degree of series-parallel graphs. *Combinatorics, Probability and Computing*, 20(04):529–570, 2011.

[71] M. Drmota and M. Noy. Extremal parameters in sub-critical graph classes. In *ANALCO*, pages 1–7, 2013.

[72] A. Dudek and A. Frieze. Tight Hamilton cycles in random uniform hypergraphs. *Random Structures & Algorithms*, 42(3):374–385, 2013.

[73] A. Dudek, A. Frieze, P.-S. Loh, and S. Speiss. Optimal divisibility conditions for loose Hamilton cycles in random hypergraphs. *Electronic Journal of Combinatorics*, 19(4):P44, 2012.

[74] J. Edmonds. Paths, trees, and flowers. *Canadian Journal of Mathematics*, 17(3):449–467, 1965.

[75] J. Ehrenmüller and J. Rué. Spanning trees in random series-parallel graphs. *Advances in Applied Mathematics*, 75:18–55, 2016.

[76] P. Erdős and A. Rényi. On the evolution of random graphs. *Bull. Inst. Internat. Statist.*, 38(4):343–347, 1961.

[77] A. Ferber, M. Krivelevich, and H. Naves. Generating random graphs in biased Maker-Breaker games. *arXiv:1310.4096*, 2013.

[78] A. Ferber, R. Nenadov, U. Peter, A. Noever, and N. Škorić. Robust hamiltonicity of random directed graphs. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1752–1758. SIAM, 2015.

[79] P. Flajolet and A. Odlyzko. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, 3(2):216–240, 1990.

[80] P. Flajolet and R. Sedgewick. *Analytic combinatorics.* Cambridge University Press, 2009.

[81] C. M. Fortuin and P. W. Kasteleyn. On the random-cluster model. *Physica*, 57(4):536–564, 1972.

[82] H. Gebauer and T. Szabó. Asymptotic random graph intuition for the biased connectivity game. *Random Structures & Algorithms*, 35(4):431–443, 2009.

[83] O. Giménez and M. Noy. Asymptotic enumeration and limit laws of planar graphs. *Journal of the American Mathematical Society*, 22(2):309–329, 2009.

[84] O. Giménez, M. Noy, and J. Rué. Graph classes with given 3-connected components: asymptotic enumeration and random graphs. *Random Structures & Algorithms*, 42(4):438–479, 2013.

[85] L. Š. Grinblat. *Algebras of sets and combinatorics*, volume 214. American Mathematical Society, 2002.

[86] L. Š. Grinblat. Theorems on sets not belonging to algebras. *Electronic Research Announcements of the American Mathematical Society*, 10(6):51–57, 2004.

[87] L. Š. Grinblat. Families of sets not belonging to algebras and combinatorics of finite sets of ultrafilters. *Journal of Inequalities and Applications*, 2015(1):1–19, 2015.

[88] R. Häggkvist and A. Johansson. Orthogonal Latin Rectangles. *Combinatorics, Probability and Computing*, 17(04):519–536, 2008.

[89] J. Han and Y. Zhao. Minimum vertex degree threshold for loose Hamilton cycles in 3-uniform hypergraphs. *Journal of Combinatorial Theory, Series B*, 114:70–96, 2015.

[90] P. Hatami and P. W. Shor. A lower bound for the length of a partial transversal in a Latin square. *Journal of Combinatorial Theory, Series A*, 115(7):1103–1113, 2008.

[91] D. Hefetz, M. Krivelevich, M. Stojaković, and T. Szabó. Avoider-Enforcer: the rules of the game. *Journal of Combinatorial Theory, Series A*, 117(2):152–163, 2010.

[92] D. Hefetz, M. Krivelevich, M. Stojaković, and T. Szabó. *Positional games*. Oberwolfach Seminars 44. Springer Basel, 2014.

[93] D. Hefetz, M. Krivelevich, and T. Szabó. Avoider-Enforcer games. *Journal of Combinatorial Theory, Series A*, 114(5):840–853, 2007.

[94] D. Hefetz, M. Krivelevich, and T. Szabó. Sharp threshold for the appearance of certain spanning trees in random graphs. *Random Structures & Algorithms*, 41(4):391–412, 2012.

[95] H. Huang, C. Lee, and B. Sudakov. Bandwidth theorem for random graphs. *Journal of Combinatorial Theory, Series B*, 102(1):14–37, 2012.

[96] H.-K. Hwang. On convergence rates in the central limit theorems for combinatorial structures. *European Journal of Combinatorics*, 19(3):329 – 343, 1998.

[97] S. Janson, D. E. Knuth, T. Łuczak, and B. Pittel. The birth of the giant component. *Random Structures & Algorithms*, 4:233–358, 1993.

[98] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs.* John Wiley & Sons, New York, 2000.

[99] A. Johansson, J. Kahn, and V. Vu. Factors in random graphs. *Random Structures & Algorithms*, 33(1):1–28, 2008.

[100] J. Kahn, E. Lubetzky, and N. Wormald. The threshold for combs in random graphs. *Random Structures & Algorithms*, 2015.

[101] M. Kang and T. Łuczak. Two critical periods in the evolution of random planar graphs. *Transactions of the American Mathematical Society*, 364(8):4239–4265, 2012.

[102] R. M. Karp. Reducibility among combinatorial problems. In R. E. Miller and J. W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103, Boston, MA, 1972.

[103] I. Khan. Perfect matchings in 3-uniform hypergraphs with large vertex degree. *SIAM Journal on Discrete Mathematics*, 27(2):1021–1039, 2013.

[104] G. Kirchhoff. Ueber die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Vertheilung galvanischer Ströme geführt wird. *Annalen der Physik und Chemie*, 72(12):497–508, 1847.

[105] Y. Kohayakawa. Szemerédi's regularity lemma for sparse graphs. In *Foundations of Computational Mathematics*, pages 216–230. Springer Berlin Heidelberg, 1997.

[106] Y. Kohayakawa and V. Rödl. Szemerédi's regularity lemma and quasi-randomness. In *Recent Advances in Algorithms and Combinatorics*, pages 289–351. Springer New York, 2003.

[107] K. K. Koksma. A lower bound for the order of a partial transversal in a Latin square. *Journal of Combinatorial Theory*, 7(1):94–95, 1969.

[108] J. Komlós, G. N. Sárközy, and E. Szemerédi. On the square of a Hamiltonian cycle in dense graphs. *Random Structures & Algorithms*, 9(1-2):193–211, 1996.

[109] J. Komlós, G. N. Sárközy, and E. Szemerédi. Blow-up lemma. *Combinatorica*, 17(1):109–123, 1997.

[110] J. Komlós and E. Szemerédi. Limit distribution for the existence of Hamiltonian cycles in a random graph. *Discrete Mathematics*, 43(1):55–63, 1983.

[111] A. D. Korshunov. Solution of a problem of Erdős and Rényi on Hamilton cycles in non-oriented graphs. 17:760–764, 1976.

[112] A. V. Kostochka. The number of spanning trees in graphs with a given degree sequence. *Random Structures & Algorithms*, 6(2-3):269–274, 1995.

[113] D. Kotlar and R. Ziv. Large matchings in bipartite graphs have a rainbow matching. *European Journal of Combinatorics*, 38:97–101, 2014.

[114] M. Krivelevich. Embedding spanning trees in random graphs. *SIAM Journal on Discrete Mathematics*, 24(4):1495–1500, 2010.

[115] M. Krivelevich. The critical bias for the Hamiltonicity game is $(1+o(1))n/\ln n$. *Journal of the American Mathematical Society*, 24(1):125–131, 2011.

[116] M. Krivelevich, C. Lee, and B. Sudakov. Resilient pancyclicity of random and pseudo-random graphs. *SIAM Journal on Discrete Mathematics*, 24(1):1–16, 2010.

[117] M. Krivelevich and B. Sudakov. Sparse pseudo-random graphs are Hamiltonian. *Journal of Graph Theory*, 42(1):17–33, 2003.

[118] M. Krivelevich and B. Sudakov. Pseudo-random graphs. In *More sets, graphs and numbers*, pages 199–262. Springer Berlin Heidelberg, 2006.

[119] M. Krivelevich, B. Sudakov, and T. Szabó. Triangle factors in sparse pseudo-random graphs. *Combinatorica*, 24(3):403–426, 2004.

[120] D. Kühn and D. Osthus. Embedding large subgraphs into dense graphs. In S. Huczynka, J. Mitchel, and C. Roney-Dougal, editors, *Surveys in Combinatorics*, pages 137–167. London Math. Soc. Lecture Notes 365, 2009.

[121] D. Kühn and D. Osthus. On Pósa's conjecture for random graphs. *SIAM Journal on Discrete Mathematics*, 26(3):1440–1457, 2012.

[122] D. Kühn and D. Osthus. Hamilton cycles in graphs and hypergraphs: an extremal perspective. arXiv:1402.4268, 2014.

[123] D. Kühn, D. Osthus, and A. Treglown. Matchings in 3-uniform hypergraphs. *Journal of Combinatorial Theory, Series B*, 103(2):291–305, 2013.

[124] C. Lee and B. Sudakov. Dirac's theorem for random graphs. *Random Structures & Algorithms*, 41(3):293–305, 2012.

[125] A. Lehman. A solution of the Shannon switching game. *J. Soc. Indust. Appl. Math.*, 12:687–725, 1964.

[126] R. Lyons. Asymptotic enumeration of spanning trees. *Combinatorics, Probability and Computing*, 14:491–522, 2005.

[127] B. D. McKay. Spanning trees in regular graphs. *European Journal of Combinatorics*, 4(2):149–160, 1983.

[128] B. D. McKay, J. C. McLeod, and I. M. Wanless. The number of transversals in a Latin square. *Designs, Codes and Cryptography*, 40(3):269–284, 2006.

[129] R. Montgomery. Embedding bounded degree spanning trees in random graphs. *arXiv:1405.6559*, 2014.

[130] R. Montgomery. Sharp threshold for embedding combs and other spanning trees in random graphs. *arXiv:1405.6560*, 2014.

[131] J. W. Moon. Counting labelled trees. In *Canadian Mathematical Congress*, 1970.

[132] R. Mullin. On the enumeration of tree-rooted maps. *Canadian Journal of Mathematics*, 19:174–183, 1967.

[133] R. Nenadov and N. Škorić. Powers of cycles in random graphs and hypergraphs. *arXiv:1601.04034*, 2016.

[134] G. Nivasch and E. Omri. Rainbow matchings and algebras of sets. *arXiv:1503.03671*, 2015.

[135] M. Noy and L. Ramos. Random planar graphs with given minimum degree. arXiv:1403.5211, 2014.

[136] M. Noy, V. Ravelomanana, and J. Rué. On the probability of planarity of a random graph near the critical point. *Proceedings of the American Mathematical Society*, 143(3):925–936, 2015.

[137] K. Panagiotou, B. Stufler, and K. Weller. Scaling limits of random graphs from subcritical classes. arXiv:1411.1865, 2014.

[138] A. Pokrovskiy. Rainbow matchings and rainbow connectedness. *arXiv:1504.05373*, 2015.

[139] D. Poole. On weak Hamiltonicity of a random hypergraph. *arXiv:1410.7446*, 2014.

[140] L. Pósa. Hamiltonian circuits in random graphs. *Discrete Mathematics*, 14(4):359–364, 1976.

[141] O. Riordan. Spanning subgraphs of random graphs. *Combinatorics, Probability and Computing*, 9(2):125–148, 2000.

[142] V. Rödl and A. Ruciński. Dirac-type questions for hypergraphs — a survey (or more problems for Endre to solve). In *An Irregular Mind*, pages 561–590. Springer Berlin Heidelberg, 2010.

[143] V. Rödl and A. Ruciński. Families of triples with high minimum degree are Hamiltonian. *Discussiones Mathematicae Graph Theory*, 34(2):361–381, 2014.

[144] V. Rödl, A. Ruciński, and E. Szemerédi. A Dirac-type theorem for 3-uniform hypergraphs. *Combinatorics, Probability and Computing*, 15(1-2):229–251, 2006.

[145] J. Rué, I. Sau, and D. M. Thilikos. Asymptotic enumeration of non-crossing partitions on surfaces. *Discrete Mathematics*, 313(5):635–649, 2013.

[146] H. J. Ryser. Neuere Probleme der Kombinatorik. *Vorträge über Kombinatorik, Oberwolfach*, pages 69–91, 1967.

[147] A. Scott. Szemerédi's regularity lemma for matrices and sparse graphs. *Combinatorics, Probability and Computing*, 20(03):455–466, 2011.

[148] S. K. Stein. Transversals of Latin squares and their generalizations. *Pacific Journal of Mathematics*, 59(2):567–575, 1975.

[149] B. Sudakov and V. Vu. Local resilience of graphs. *Random Structures & Algorithms*, 33(4):409–433, 2008.

[150] A. Thomason. Pseudo-random graphs. *Annals of Discrete Mathematics*, 33:307–331, 1987.

[151] A. Thomason. Random graphs, strongly regular graphs and pseudorandom graphs. *Surveys in Combinatorics*, 123:173–195, 1987.

[152] B. A. Trakhtenbrot. Towards a theory of non-repeating contact schemes (in Russian). *Trudy Mathematichiskogo Instituta imeni V.A. Steklova*, 51:226–269, 1958.

[153] W. T. Tutte. *Connectivity in graphs.* University of Toronto Press, 1966.

[154] T. R. S. Walsh. Counting labelled three-connected and homeomorphically irreducible two-connected graphs. *Journal of Combinatorial Theory, Series B*, 32(1):1–11, 1982.

[155] I. M. Wanless. Transversals in Latin squares. *Quasigroups Related Systems*, 15:169–190, 2007.

[156] I. M. Wanless and B. S. Webb. The existence of latin squares without orthogonal mates. *Designs, Codes and Cryptography*, 40(1):131–135, 2006.

[157] D. B. West. *Introduction to graph theory.* Prentice Hall, 2001.

[158] D. E. Woolbright. An $n \times n$ Latin square has a transversal with at least $n - \sqrt{n}$ distinct symbols. *Journal of Combinatorial Theory, Series A*, 24(2):235–237, 1978.

[159] E. Wright. The number of connected sparsely edged graphs. *Journal of Graph Theory*, 1(4):317–330, 1977.

[160] E. Wright. The number of connected sparsely edged graphs. II. Smooth graphs and blocks. *Journal of Graph Theory*, 2(4):299–305, 1978.

[161] Y. Zhao. Recent advances on Dirac-type problems for hypergraphs. *arXiv:1508.06170*, 2015.

Index

Symbols

A

B