Colourings of Random Graphs

Annika Heckel

Balliol College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2016

Abstract

We study graph parameters arising from different types of colourings of random graphs, defined broadly as an assignment of colours to either the vertices or the edges of a graph.

The *chromatic number* of a graph is the minimum number of colours required for a vertex colouring where no two adjacent vertices are coloured the same. Determining the chromatic number is one of the classic challenges in random graph theory. In Chapter 3, we give new upper and lower bounds for the chromatic number of the dense random graph $\mathcal{G}(n,p)$ where $p \in (0,1)$ is constant. These bounds are the first to match up to an additive term of order $o(1)$ in the denominator, and in particular, they determine the average colour class size in an optimal colouring up to an additive term of order $o(1)$.

In Chapter 4, we study a related graph parameter called the *equitable chromatic number*. This is defined as the minimum number of colours needed for a vertex colouring where no two adjacent vertices are coloured the same and, additionally, all colour classes are as equal in size as possible. We prove one point concentration of the equitable chromatic number of the dense random graph $\mathcal{G}(n,m)$ with $m = \left\lfloor p\binom{n}{2}\right\rfloor$, $p < 1 - 1/e^2$ constant, on a subsequence of the integers. We also show that whp, the dense random graph $\mathcal{G}(n,p)$ allows an almost equitable colouring with a near optimal number of colours.

We call an edge colouring of a graph G a *rainbow colouring* if every pair of vertices is joined by a *rainbow path*, which is a path where no colour is repeated. The least number of colours where this is possible is called the *rainbow connection number* $\mathrm{rc}(G)$. For any graph G, $\mathrm{rc}(G) \geqslant \mathrm{diam}(G)$, where $\mathrm{diam}(G)$ denotes the diameter. In Chapter 5, we will see that in the random graph $G(n,p)$, rainbow connection number 2 is essentially equivalent to diameter 2. More specifically, we consider $G \sim \mathcal{G}(n,p)$ close to the diameter 2 threshold and show that whp $\mathrm{rc}(G) = \mathrm{diam}(G) \in \{2,3\}$. Furthermore, we show that in the random graph process, whp the hitting times of diameter 2 and of rainbow connection number 2 coincide.

In Chapter 6, we investigate sharp thresholds for the property $\mathrm{rc}(G) \leqslant r$ where r is a fixed integer. The results of Chapter 5 imply that for $r = 2$, the properties $\mathrm{rc}(G) \leqslant 2$ and $\mathrm{diam}(G) \leqslant 2$ share the same sharp threshold. For $r \geqslant 3$, the situation seems quite different. We propose an alternative threshold and prove that this is an upper bound for the sharp threshold for $\mathrm{rc}(G) \leqslant r$ where $r \geqslant 3$.

Acknowledgements

First and foremost, I would like to thank my supervisor Oliver Riordan for his guidance and for sharing his vast knowledge and insight.

The Mathematical Institute has provided me with a fantastic working environment and facilities. I was lucky to have been a member of the Combinatorics Group, where I have learned a lot and made many great friends.

Balliol College has been an invaluable source of support; I should single out my college adviser Frances Kirwan for providing exceptional practical guidance throughout the years.

My research has been supported by EPSRC, a Scatcherd European Scholarship, the Mathematical Prizes Fund and the Vice-Chancellors' Fund. I am grateful to have received support from the Esson Bequest Fund during my parental leave, as well as a Georgina Horlick Childcare Bursary from Balliol College.

I would like to thank my family for their love and encouragement.

Authorship

Chapters 5 and 6 are based on the papers [31, 32] which are joint work with Oliver Riordan.

To Albert and Fiona

Contents

1 Introduction **1**

1.1 Chromatic number 2

1.2 Rainbow connectivity 5

2 Preliminaries **9**

2.1 Asymptotic notation 9

2.2 Graph theoretic notions 9

2.3 Random graphs 10

2.3.1 Asymptotic equivalence 11

2.3.2 Threshold functions 11

2.4 Inequalities 12

2.4.1 Union of events 12

2.4.2 Chernoff bounds 13

2.4.3 Markov's Inequality and the First Moment Method 14

2.4.4 The Paley–Zygmund Inequality and the Second Moment Method 14

2.4.5 Bounded Differences Inequality 15

3 The chromatic number of dense random graphs **17**

3.1 Background and results 17

3.2 Outline 21

3.3 Preliminaries and notation 26

3.3.1 List of key facts and relations 28

3.3.2 On the solutions of (3.4) 29

3.4 Proof of the lower bound 33

3.5 Bounding the second moment 36

3.5.1 Outline 37

3.5.2 Typical overlap range 43

3.5.3 Pairs of partitions with many small overlap blocks 50

3.5.3.1 Preliminaries 50

3.5.3.2 Contribution from Cases I and II 54

3.5.3.3 Contribution from Case III 57

3.5.4 Very high overlap 65
3.6 Outlook 73

4 Equitable colourings 75
4.1 Background and results 75
4.2 Proof of Theorem 4.1 78
4.3 Outline of the proof of Theorem 4.2 79
4.4 The first moment 81
4.5 Choice of the subsequence 86
4.6 The second moment 89
4.6.1 Asymptotics of $S_{\mathbf{r}}$ 91
4.6.2 The typical overlap case 91
4.6.3 The intermediate overlap case 96
4.6.4 The high overlap case 97

5 The hitting time of rainbow connection number two 99
5.1 Background and results 99
5.2 Proofs 102
5.2.1 Overview 102
5.2.2 Definitions 102
5.2.3 Proofs of Theorems 5.1 and 5.4 103
5.2.4 Proof of Proposition 5.6 104
5.2.5 Proof of Proposition 5.7 113

6 On the threshold for rainbow connection number $r \geqslant 3$ 117
6.1 Background and results 117
6.2 Proof of Theorem 6.2 119
6.2.1 General observations 119
6.2.2 The main lemma 121
6.2.3 Completing the proof 130

Bibliography 132

List of Figures

3.1 The two cases if $p > 1 - 1/e^2$. 23
3.2 The three possible cases for the function $f(x)$. 31
3.3 Generating pairs of k-equipartitions with high overlap. 42

5.1 Two ways in which $\{x, w\}$ can be a fix for a dangerous pair other than $\{v, w\}$. 107
5.2 Possible ways in which the edge e could have been flagged before considering $\{v, w\}$. 115

6.1 The event $A_{v,w}$. 123
6.2 The tree of depth s obtained by the exploration of the rainbow s-neighbourhood of z. 124

Chapter 1

Introduction

A *graph* $G = (V, E)$, consisting of a set V of *vertices* with a set $E \subset \binom{V}{2}$ of *edges* between them, is a mathematical model of a network. In a random graph, there is typically a large number of vertices and the edges are drawn at random, representing a large disordered network. Random graph theory has a long and rich history which we will not attempt to give a full account of, instead we refer to the books by Bollobás [7] and by Janson, Łuczak and Ruciński [33].

All graphs that we consider are finite and simple (no loops or multiple edges), and usually the vertex set is taken to be $V = [n] = \{1, \ldots, n\}$ for a positive integer n and we let $N = \binom{n}{2}$ denote the maximum possible number of edges. Given n and $p \in [0, 1]$, the *binomial random graph* $\mathcal{G}(n, p)$ is the graph with vertex set $[n]$ where each of the N possible edges is present independently with probability p. This random graph model was originally proposed by Gilbert [26] in 1959.

A closely related model, the *uniform random graph* $\mathcal{G}(n, m)$, was introduced at the same time in a seminal paper by Erdős and Rényi [20]. Given positive integers n and $m \leqslant N$, $\mathcal{G}(n, m)$ is the random graph with vertex set $V = [n]$ and edge set E which is chosen uniformly at random from all $\binom{N}{m}$ possible edge sets of size exactly m.

We are usually interested in the *asymptotic behaviour* of random graphs: as the number n of vertices tends to infinity, can we find properties which the random graph possesses with probability approaching 1, or properties where this probability tends to 0? We say that a sequence of events $E = E(n)$ holds *with high probability (whp)* if $\lim_{n\to\infty} \mathbb{P}(E(n)) = 1$.

A colouring of a graph, in the broad sense, is an assignment of colours to either the vertices or the edges of a graph. We are going to study graph parameters which arise from different types of colourings and what they tell us about the global structure of the random graph in question.

In the remainder of this chapter, we will give a brief overview of the main results contained in this thesis. A longer introduction to each topic can be found at the beginning of each chapter. We assume familiarity with basic concepts from probability theory. We will also use standard graph theoretic notions and asymptotic notation which will be reviewed in Chapter 2.

1.1 Chromatic number

One of the central concepts in graph theory is that of a *proper colouring* where every vertex of a graph G is assigned a colour so that no two adjacent vertices are coloured the same. The minimum number of colours where this is possible is called the *chromatic number* $\chi(G)$. Typical applications are scheduling and resource allocation problems: suppose that a number of classes needs to be scheduled, but certain pairs of classes may not be scheduled at the same time, for example because some students are required to attend both. We can think of the classes as vertices in a graph and draw an edge between them whenever there is a conflict. The task is now to group the vertices into conflict free timeslots, which is exactly what is achieved by a proper colouring of the vertices. The chromatic number is the required minimum number of timeslots.

A classic challenge in random graph theory, which was raised in one of the earliest papers by Erdős and Rényi [21], is finding the likely value of the chromatic number of a random graph, and a wealth of results has been published over the years in this area. Most notably, in 1987 Bollobás [6] first determined the asymptotic value of the chromatic number of dense random graphs. In Chapter 3, which is based on the manuscript [30], we will give new upper and lower bounds for the chromatic number of these random graphs; these bounds determine, for the first time, the average colour class size in an optimal colouring up to an additive term of order $o(1)$. The main result of Chapter 3 is the following theorem.

Theorem 3.1. *Let $p \in (0,1)$ be constant, and consider the random graph $G \sim \mathcal{G}(n,p)$. Let $q = 1-p$, $b = \frac{1}{q}$, $\gamma = \gamma_p(n) = 2\log_b n - 2\log_b \log_b n - 2\log_b 2$ and $\Delta = \Delta_p(n) = \gamma - \lfloor \gamma \rfloor$. Then whp,*

$$\chi(G) = \frac{n}{\gamma - x_0 + o(1)},$$

where $x_0 \geqslant 0$ is the smallest nonnegative solution of

$$(1-\Delta+x)\log_b(1-\Delta+x) + \frac{(\Delta-x)(1-\Delta)}{2} \leqslant 0.$$

The lower bound in Theorem 3.1 is proved through a relatively easy first moment argument. Somewhat surprisingly, there are two different random variables whose first moments are the bottleneck for different constant values of p.

Proving good upper bounds for $\chi(\mathcal{G}(n,p))$ has historically been a greater challenge than finding good lower bounds, and the proof of Theorem 3.1 is no exception. The best previously known upper bounds for $\chi(\mathcal{G}(n,p))$ have been derived through refinements of the classic approach due to Bollobás [6]. A key ingredient in these proofs is to show that whp, every sufficiently large set of vertices contains a large independent set, which is a set of vertices without any edges between them. A colouring is then constructed by successively selecting independent sets and assigning a colour to each one until only a few vertices are left which can be coloured individually.

In contrast, the proof of the upper bound in Theorem 3.1 does not follow this strategy but is instead based on an elaborate application of the second moment method. Since this approach involves considering complete, global colourings, it yields a sharper bound than those that can be achieved through a step-by-step construction of a colouring.

In a nutshell, the proof involves considering all pairs of possible complete colourings and calculating the joint probability that they are both proper colourings of the random graph at the same time. Since this probability depends critically on how similar the two colourings are, the main technical difficulty lies in classifying how two colourings may overlap and splitting up the calculations into more manageable cases.

While Theorem 3.1 addresses the likely value of the chromatic number of dense random graphs, another important open problem is its *concentration*, i.e., the length of the shortest interval which contains the chromatic number whp (see for example [8]). In Chapter 4, we will harness the second moment calculations from the proof of Theorem 3.1 to show that a related graph parameter, the equitable chromatic number, is very sharply concentrated for the dense random graph $\mathcal{G}(n, m)$ on a subsequence of the integers.

An *equitable proper colouring* of a graph is a proper colouring where all colour classes are as equal in size as possible. Since the number of colours does not necessarily divide the number of vertices, this means that the colour class sizes may differ by 1. The minimum required number of colours is called the *equitable chromatic number* $\chi_=(G)$.

In view of possible applications, the notion of an equitable colouring is a natural one: in the scheduling problem described before, the colour class sizes are simply the numbers of classes which are scheduled to take place at the same time. To optimise the use of available resources such as the number of rooms in a building, it is often desirable to keep this number as equal as possible between different time slots.

For constant $p < 1 - 1/e^2$, we will consider the dense random graph $\mathcal{G}(n, m)$ with $m = m(n) = \lfloor pN \rfloor$. Refining the calculations from Chapter 3, we will prove that there is a subsequence $(n_j)_{j \geqslant 1}$ of the integers such that $\chi_= (\mathcal{G}(n_j, m(n_j)))$ is whp concentrated on one value. Moreover, we can pick the subsequence in such a way that whp, all colour classes in an optimal equitable colouring of $\mathcal{G}(n_j, m(n_j))$ have size exactly j.

Theorem 4.2. *Let $0 < p < 1 - 1/e^2$ be constant. There exists a strictly increasing sequence of integers $(n_j)_{j \geqslant 1}$ and $j_0 \geqslant 1$ such that*

a) for all $j \geqslant j_0$, $j | n_j$,

b) letting $b = \frac{1}{1-p}$ and $\gamma_j = 2\log_b n_j - 2\log_b \log_b n_j - 2\log_b 2$,

$$\gamma_j = j + o(1) \text{ as } j \to \infty, \text{ and}$$

c) letting $G \sim \mathcal{G}(n_j, m_j)$ with $m_j = \left\lfloor p\binom{n_j}{2} \right\rfloor$, with high probability as $j \to \infty$,

$$\chi_= (G) = \frac{n_j}{j}.$$

We will also show that there is an almost equitable proper colouring of $\mathcal{G}(n,p)$ with a near optimal number of colours.

Theorem 4.1. *Let $p \in (0,1)$ be constant, and consider the random graph $G \sim \mathcal{G}(n,p)$. Define b, γ and x_0 as in Theorem 3.1. Then whp, G has a colouring with*

$$\frac{n}{\gamma - x_0 + o(1)}$$

colours such that the sizes of all but $o\left(\frac{n}{\gamma}\right)$ colour classes differ by at most 1.

1.2 Rainbow connectivity

Another fundamental and global graph property is *connectivity*: starting from any vertex, can we reach any other vertex through a sequence of steps along edges of the graph? Moreover, if a given graph is connected, can we quantify how well connected it is?

The rainbow connection number $\mathrm{rc}(G)$ of a graph G was introduced in 2008 by Chartrand, Johns, McKeon and Zhang [12] as a novel way to measure the connectivity of G and has since attracted the attention of a great number of researchers (see for example the survey by Li and Sun [38]). Like the chromatic number, the rainbow connection number arises from a type of graph colouring: we call an assignment of colours to the *edges* of a graph G a *rainbow colouring* if every pair of vertices is joined by a *rainbow path*, which is a path where no colour is repeated. The *rainbow connection number* $\mathrm{rc}(G)$ is the least number of colours where this is possible. If $\mathrm{rc}(G)$ is low, this indicates that G is well connected.

We will investigate the following basic question: for a given $r \geqslant 2$, how many edges need to be added to n vertices at random in order to ensure that whp, the resulting graph has rainbow connection number r?

In Chapter 5, which is based on the paper [31], we obtain an essentially complete answer to this question for $r = 2$. The rainbow connection number of any graph is always at least as large as its diameter. We will see that rainbow connection number 2 and diameter 2 are essentially equivalent in random graphs.

It is well known (see [5]) that the 'critical window' for diameter 2 in random graphs $G \sim \mathcal{G}(n,p)$ occurs at $p(n) = \sqrt{\frac{2\log n + \omega(n)}{n}}$: if $\omega(n) \to \infty$, then whp $\mathrm{diam}(G) \leqslant 2$, and if $\omega(n) \to -\infty$, then whp $\mathrm{diam}(G) \geqslant 3$. In this range, the rainbow connection number is whp the same as the diameter of the random graph:

Theorem 5.1. *Let* $p = p(n) = \sqrt{\frac{2\log n + \omega(n)}{n}}$ *where* $\omega(n) = o(\log n)$ *and let* $G \sim \mathcal{G}(n,p)$*. Then whp* $\mathrm{rc}(G) = \mathrm{diam}(G) \in \{2, 3\}$.

In fact, we can prove an even stronger result. Consider the *random graph process*, which will be formally introduced in Chapter 2, where we start at time $t = 0$ with an empty graph on n vertices and successively add edges chosen uniformly at random from all edges not already present, until at time $t = N$ we have the complete graph on n vertices. Let $\tau_{\mathcal{R}}$ and $\tau_{\mathcal{D}}$ denote the *hitting times* of rainbow connection number 2 and diameter 2, respectively, i.e., the smallest t such that at time t the graph has rainbow connection number at most 2 and diameter at most 2, respectively. We always have $\tau_{\mathcal{R}} \geqslant \tau_{\mathcal{D}}$, but in fact, the two hitting times coincide whp.

Theorem 5.4. *In the random graph process* $(G_t)_{t=0}^{N}$*, whp* $\tau_{\mathcal{D}} = \tau_{\mathcal{R}}$.

In Chapter 6, which is based on the paper [32], we examine the case $r \geqslant 3$ where the arguments used to prove Theorems 5.1 and 5.4 do not apply.

It is not hard to show that for constant r, the graph property $\mathcal{R}_r$ of having rainbow connection number at most r has a *sharp threshold*. This is defined as a sequence $p^* = p^*(n) \in [0,1]$ such that for all constants $c < 1 < C$, if $p = p(n) \leqslant cp^*(n)$ for all sufficiently large n, then whp $\mathrm{rc}(\mathcal{G}(n,p)) > r$, and if $p = p(n) \geqslant Cp^*(n)$ for all sufficiently large n, then whp $\mathrm{rc}(\mathcal{G}(n,p)) \leqslant r$ (see also Section 2.3.2 in the next chapter).

The function $p(n) = \frac{(2\log n)^{1/r}}{n^{1-1/r}}$ is a sharp threshold for the graph property of having diameter at most r (see [5]), and the results of Chapter 5 imply that for $r = 2$, this is also a sharp threshold for $\mathcal{R}_2$. In Chapter 6, we propose an alternative sharp threshold for $\mathcal{R}_r$ where $r \geqslant 3$.

Conjecture 6.1. *Fix an integer $r \geqslant 3$, set $C = \frac{r^{r-2}}{(r-2)!}$, and let*

$$p(n) = \frac{(C\log n)^{1/r}}{n^{1-1/r}}.$$

Then $p(n)$ is a sharp threshold for the graph property $\mathcal{R}_r$.

The function $p(n)$ from Conjecture 6.1 is chosen so that if we colour the edges of the random graph independently and uniformly at random with r colours, the number of pairs of vertices which are joined by only a few rainbow paths is roughly the same as the number of edges in the graph. In one direction, we will give a heuristic argument why $p(n)$ could be a lower bound for the sharp threshold of $\mathcal{R}_r$. We will show that it is an upper bound in the following theorem.

Theorem 6.2. *Fix an integer $r \geqslant 3$ and $\varepsilon > 0$, and let $C = \frac{r^{r-2}}{(r-2)!}$. Set $p = p(n) = \frac{(C(1+\varepsilon)\log n)^{1/r}}{n^{1-1/r}}$, and let $G \sim \mathcal{G}(n,p)$. Then whp, $\mathrm{rc}(G) = r$.*

Chapter 2

Preliminaries

In this chapter we will formally introduce some of the basic concepts and results which will be used throughout the thesis.

2.1 Asymptotic notation

For two functions $f, g : \mathbb{N} \to \mathbb{R}$, we write

- $f = o(g)$ if $f(n)/g(n) \to 0$ as $n \to \infty$.
- $f = O(g)$ if there are constants C and n_0 such that $|f(n)| \leqslant Cg(n)$ for all $n \geqslant n_0$.
- $f = \Theta(g)$ if $f = O(g)$ and $g = O(f)$.
- $f = O^*(g)$ if there are constants C and n_0 such that $|f(n)| \leqslant (\log n)^C g(n)$ for all $n \geqslant n_0$, where $\log n$ denotes the natural logarithm.
- $f \sim g$ if $f(n) = (1 + o(1))g(n)$, i.e., $f(n)/g(n) \to 1$.
- $f \lesssim g$ if $f(n) \leqslant (1 + o(1))g(n)$.
- $f \gtrsim g$ if $f(n) \geqslant (1 + o(1))g(n)$.

2.2 Graph theoretic notions

A *graph* $G = (V, E)$ consists of a *vertex set* V and an *edge set*

$$E \subset \binom{V}{2} = \big\{\{u, v\} \mid u, v \in V, u \neq v\big\}.$$

All the graphs we will consider are finite, i.e., $|V| < \infty$. We will usually take $V = [n] = \{1, 2, \ldots, n\}$, and let $N = \binom{n}{2}$. Furthermore, all the graphs we study are *simple*, i.e., there are no multiple edges or loops from a vertex to itself. We denote an edge $e = \{u, v\}$ by uv or vu.

We say that two vertices $u, v \in V$ are *adjacent* if $uv \in E$, and we denote by

$$\Gamma(v) = \{u \in V \mid uv \in E\}$$

the *neighbourhood* of a vertex v, and by $d(v) = |\Gamma(v)|$ the *degree* of v.

A *path* of length l in G consists of a sequence v_0, v_1, $\ldots$, v_l of distinct vertices such that for all $0 \leqslant i \leqslant l-1$, $v_i v_{i+1} \in E$. We say that a path $v_0, \ldots, v_l$ *joins* the vertices v_0 and v_l. For two given vertices $u, v \in V$, we call the length of the shortest path in G which joins u and v the *graph distance* between u and v, denoted by $d(u, v)$. If there is no such path, we set $d(u, v) = \infty$.

We say that the graph G is *connected* if any two vertices are joined by a path. For a connected graph G, the *diameter* of G is defined as the maximum distance between any two vertices in V, i.e.,

$$\operatorname{diam}(G) = \max\{d(u, v) \mid u, v \in V\}.$$

For $n \in \mathbb{N}$, let Ω_n denote the *set of all graphs with vertex set* $[n]$. Any subset $\mathcal{Q} \subset \Omega_n$ is called a *graph property*. We say that a graph property $\mathcal{Q}$ is *monotone increasing* if it is preserved under the addition of edges, that is, if G_1 and G_2 are graphs in Ω_n with edge sets E_1 and E_2, respectively, such that $E_1 \subset E_2$, then $G_1 \in \mathcal{Q}$ implies $G_2 \in \mathcal{Q}$. We say that a graph property $\mathcal{Q}$ is *monotone decreasing* if it is preserved under the removal of edges, or, equivalently, if $\Omega_n \setminus \mathcal{Q}$ is monotone increasing.

2.3 Random graphs

Recall from the introduction that given $n \geqslant 1$ and $p \in [0, 1]$, we denote by $G \sim \mathcal{G}(n, p)$ the *binomial random graph* with vertex set $[n] = \{1, \ldots, n\}$ where each of the $N = \binom{n}{2}$ possible edges is present independently with probability p. Given $n, m \in \mathbb{N}$, we denote by $G \sim \mathcal{G}(n,m)$ the *uniform random graph* with vertex set $V = [n]$ and edge set E of size exactly m which is chosen uniformly at random from all $\binom{N}{m}$ possible edge sets of size m.

We will also study the *random graph process* $(G_t)_{t=0}^N$. At time $t = 0$, we start with the empty graph on n vertices, $G_0 = ([n], \emptyset)$. At time t, we choose one edge uniformly at random from all those not already present in the graph in G_t and add it to the graph to yield G_{t+1}, until at time N we have a complete graph $G_N = \left([n], \binom{[n]}{2}\right)$. For a monotone increasing graph property $\mathcal{Q}$, let $\tau_{\mathcal{Q}}$ denote the *hitting time* of $\mathcal{Q}$, i.e., the smallest t such that $G_t \in \mathcal{Q}$.

2.3.1 Asymptotic equivalence

If m is close to pN and n is large, then the random graphs $\mathcal{G}(n,m)$ and $\mathcal{G}(n,p)$ behave in a similar way. We will use the following asymptotic equivalence theorem (see Proposition 1.13 in [33]).

Theorem 2.1. *Let $\mathcal{Q} = \mathcal{Q}(n) \subset \Omega_n$ be a monotone graph property, and let $m = m(n)$ be a sequence of integers and $a \in [0,1]$. If for every sequence $p = p(n) \in [0,1]$ such that*

$$p = \frac{m}{N} + O\left(\sqrt{\frac{m(N-m)}{N^3}}\right),$$

it holds that $\mathbb{P}(\mathcal{G}(n,p) \in \mathcal{Q}) \to a$ as $n \to \infty$, then also $\mathbb{P}(\mathcal{G}(n,m) \in \mathcal{Q}) \to a$ as $n \to \infty$.

□

There is a corresponding result in the other direction which does not require monotonicity (see Proposition 1.12 in [33]).

2.3.2 Threshold functions

For $n \in \mathbb{N}$, consider a graph property $\mathcal{Q} = \mathcal{Q}(n) \subset \Omega_n$ and a sequence $p^* = p^*(n) \in [0,1]$. We call p^* a *threshold function* for $\mathcal{Q}$ if the following holds: for all sequences

$p = p(n) \in [0,1]$ such that $p = o(p^*)$, whp $\mathcal{G}(n,p) \notin \mathcal{Q}$, and for all sequences $p = p(n) \in [0,1]$ such that $p^* = o(p)$, whp $\mathcal{G}(n,p) \in \mathcal{Q}$. Furthermore, if there are constants $c, C > 0$ such that these properties hold for all sequences $p = p(n) \in [0,1]$ where for all large enough n, $p(n) \leqslant cp^*(n)$ and $p(n) \geqslant Cp^*(n)$, respectively, then we say that p^* is a *semi-sharp threshold function* for $\mathcal{Q}$. Finally, if these properties hold for any constants $c < 1$ and $C > 1$ and all sequences $p = p(n) \in [0,1]$ where for all large enough n, $p(n) \leqslant cp^*(n)$ and $p(n) \geqslant Cp^*(n)$, respectively, then p^* is called a *sharp threshold function.*

The notion of a semi-sharp threshold is non-standard, and lies in between that of a (weak) threshold and of a truly sharp threshold.

Bollobás and Thomason [10] showed that every monotone graph property has a threshold function, and Friedgut [23] gave conditions for the existence of a sharp threshold.

2.4 Inequalities

We will now review a number of inequalities and approximations.

2.4.1 Union of events

We start with a simple observation on the probability of a union of events (which also appeared in [32]).

Lemma 2.2. *Let A_i, $i = 1, \ldots, k$, be events in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then*

$$\sum_i \mathbb{P}(A_i) - \sum_i \sum_{j<i} \mathbb{P}(A_i \cap A_j) \leqslant \mathbb{P}\left(\bigcup_i A_i\right) \leqslant \sum_i \mathbb{P}(A_i).$$

Proof. The second inequality is the well known union bound. For the first inequality, which is a special case of the inclusion–exclusion principle, note that the events

$$A_i \setminus \bigcup_{j<i} (A_i \cap A_j)$$

are disjoint for different i. Hence, using the union bound again in the last step,

$$\begin{aligned}\mathbb{P}\left(\bigcup_i A_i\right) &= \mathbb{P}\left(\bigcup_i \left(A_i \setminus \bigcup_{j<i}(A_i \cap A_j)\right)\right) = \sum_i \mathbb{P}\left(A_i \setminus \bigcup_{j<i}(A_i \cap A_j)\right) \\ &= \sum_i \mathbb{P}(A_i) - \sum_i \mathbb{P}\left(\bigcup_{j<i}(A_i \cap A_j)\right) \geqslant \sum_i \mathbb{P}(A_i) - \sum_i \sum_{j<i} \mathbb{P}(A_i \cap A_j).\end{aligned}$$

□

2.4.2 Chernoff bounds

We will need a number of tail bounds for binomial distributions. Recall the well-known Chernoff bounds ([14], see also [33, p.26]).

Theorem 2.3 (Chernoff bounds)**.** *Let X be distributed binomially with parameters n and p, and let $0 < x < 1$.*

(i) If $x \geqslant p$, then $\mathbb{P}(X \geqslant nx) \leqslant \left[\left(\frac{p}{x}\right)^x \left(\frac{1-p}{1-x}\right)^{1-x}\right]^n$.

(ii) If $x \leqslant p$, then $\mathbb{P}(X \leqslant nx) \leqslant \left[\left(\frac{p}{x}\right)^x \left(\frac{1-p}{1-x}\right)^{1-x}\right]^n$.

□

The following corollary of Theorem 2.3 is given in Theorem 2.1 in [33].

Corollary 2.4. *Let X be distributed binomially with parameters n and p, and let $\varphi(x) = (1+x)\log(1+x) - x$ for $x \geqslant -1$ and $\varphi(x) = -\infty$ otherwise. Then for all $t \geqslant 0$,*

$$\mathbb{P}(X \leqslant np - t) \leqslant \exp\left(-np\varphi\left(-\frac{t}{np}\right)\right).$$

□

In many cases the following simpler bound is sufficient (see Corollary 2.3 in [33]).

Corollary 2.5. *Let X be distributed binomially with parameters n and p. If $0 < \varepsilon \leqslant \frac{3}{2}$, then*

$$\mathbb{P}\left(|X - np| \geqslant \varepsilon np\right) \leqslant 2e^{-\varepsilon^2 np/3}.$$

□

We will also need the following consequence of the Chernoff bounds (which appeared in [31] together with its proof).

Corollary 2.6. *Let $(n_i)_{i\in\mathbb{Z}}$ be a sequence of integers such that $n_i \to \infty$ as $i \to \infty$, and let $(p_i)_{i\in\mathbb{N}}$ be a sequence of probabilities. Let $X_i \sim \mathrm{Bin}(n_i, p_i)$, and let $k \in \mathbb{N}$ be constant. Suppose that $\mu_i := n_i p_i \to \infty$ as $i \to \infty$. Then*

$$\mathbb{P}(X_i \leqslant k) = O(\mu_i^k e^{-\mu_i}).$$

Proof. Applying Theorem 2.3 to X_i with $x_i = \frac{k}{n_i}$ gives

$$\mathbb{P}(X_i \leqslant k) = \mathbb{P}(X_i \leqslant n_i x_i) \leqslant \left(\frac{\mu_i}{k}\right)^k \left(\frac{1-p_i}{1-\frac{k}{n_i}}\right)^{n_i-k} = O\left(\mu_i^k \frac{e^{-\mu_i+p_i k}}{e^{-k}}\right) = O(e^{-\mu_i}\mu_i^k),$$

using the fact that $1-y \leqslant e^{-y}$ and that $\lim_{n\to\infty}(1-\frac{y}{n})^n = e^{-y}$ for every $y \in \mathbb{R}$. □

2.4.3 Markov's Inequality and the First Moment Method

A simple yet powerful tool in probabilistic combinatorics is the *first moment method*. It is based on Markov's Inequality, which says that for any integrable random variable $X \geqslant 0$ and constant $c > 0$, $\mathbb{P}(X \geqslant c) \leqslant \mathbb{E}[X]/c$.

If X is an integrable random variable which only takes nonnegative integer values, then Markov's Inequality gives

$$\mathbb{P}(X > 0) = \mathbb{P}(X \geqslant 1) \leqslant \mathbb{E}[X].$$

In typical application, we would like to find out whether or not a certain object exists whp, such as a colouring of a random graph with a certain number of colours. If $X = X(n)$ counts the number of these objects — the number of colourings of the random graph, say — and we can show that $\mathbb{E}[X] = o(1)$, then whp no such object exists. This approach is called the first moment method.

2.4.4 The Paley–Zygmund Inequality and the Second Moment Method

A proof of the whp *existence* of an object is slightly more complicated. In particular, showing that $\mathbb{E}[X] \to \infty$ is not sufficient, we also usually need a bound on the variance or the second moment of X.

This type of argument is called the *second moment method*, and there are several ways to phrase it. The variant we are going to use is based on the Paley–Zygmund Inequality ([48, 49], see also A.12 in [54]).

Theorem 2.7 (Paley–Zygmund Inequality)**.** *If $X \geqslant 0$ is an integrable random variable with finite variance and $c \in [0,1]$, then*

$$\mathbb{P}(X > c\,\mathbb{E}[X]) \geqslant (1-c)^2 \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}.$$

□

Setting $c = 0$ yields a lower bound for the probability of the event $\{X > 0\}$.

$$\mathbb{P}(X > 0) \geqslant \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}.$$

2.4.5 Bounded Differences Inequality

Another widely used tool in combinatorics is the method of bounded differences. The following inequality is also referred to as McDiarmid's Inequality and is similar to the Azuma–Hoeffding Inequality.

Theorem 2.8 (Bounded Differences Inequality [45])**.** *Let $X_1, \ldots, X_n$ be independent random variables where X_i takes values in the set Λ_i. Suppose that $f : \prod_{i=1}^n \Lambda_i \to \mathbb{R}$ is a measurable function and there are numbers c_i such that whenever $x, x' \in \prod_{i=1}^n \Lambda_i$ differ only in the i^{th} coordinate, then*

$$|f(x) - f(x')| \leqslant c_i.$$

Let $X = f(X_1, \ldots, X_n)$, then for all $t > 0$,

$$\mathbb{P}\left(|X - \mathbb{E}X| > t\right) \leqslant 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right).$$

The proof of this theorem is based on martingale inequalities, for full details see [45] or Chapter 2.4 in [33].

In the random graph setting $\mathcal{G}(n,p)$, the random variables X_i are usually defined in terms of the N independent indicator variables $(\mathbb{1}_e)_{e \in \binom{V}{2}}$ of the edges. Most importantly, given the vertex set $V = [n]$, we often take X_i to be the sequence of indicator variables $(\mathbb{1}_{ji})_{1 \leqslant j < i}$. These n random variables $X_1, \dots, X_n$ are independent. If we can show that a graph parameter does not change too much if the edges incident with any particular vertex are changed arbitrarily, then the Bounded Differences Inequality tells us that in the random graph $\mathcal{G}(n,p)$, this graph parameter is tightly concentrated about its mean.

Shamir and Spencer [53] famously gave this argument to show that for any function $p = p(n)$, the chromatic number of $\mathcal{G}(n,p)$ is whp concentrated on an interval of length $\sqrt{n}\omega(n)$ where $\omega(n)$ tends to infinity arbitrarily slowly.

Chapter 3

The chromatic number of dense random graphs

3.1 Background and results

The chromatic number, defined as the minimum number of colours needed for a vertex colouring where no two adjacent vertices are coloured the same, is one of the central topics in graph theory and has a wide range of applications including scheduling and resource allocation problems. The question of determining the chromatic number of random graphs was first asked in 1960 by Erdős and Rényi [21], and a plethora of results has since been published in this area (see for example the recent survey by Kang and McDiarmid [35]).

The order of magnitude of the chromatic number of the dense random graph $G \sim \mathcal{G}(n,p)$ where $p \in (0,1)$ is constant was first established in 1975 by Grimmett and McDiarmid, who showed that whp,

$$(1+o(1))\frac{n}{2\log_b n} \leqslant \chi(G) \leqslant (1+o(1))\frac{n}{\log_b n},$$

where $b = \frac{1}{1-p}$. They also conjectured that the asymptotic value of $\chi(G)$ lies near the lower bound. Establishing the asymptotic behaviour of $\chi(G)$ remained one of the major open problems in random graph theory until it was settled by a breakthrough result of Bollobás in 1987 [6], who used martingale inequalities to prove the matching upper bound: he showed that whp,

$$\chi(G) = (1+o(1))\frac{n}{2\log_b n}.$$

The same result was obtained independently by Matula and Kučera [44], using the expose-and-merge approach due to Matula [43].

Let

$$\gamma = \gamma_p(n) = 2\log_b n - 2\log_b \log_b n - 2\log_b 2. \tag{3.1}$$

Refining Bollobás' approach, more accurate bounds for $\chi(G)$ were given by McDiarmid [45, 46] who showed that whp,

$$\frac{n}{\gamma + \frac{2}{\log b} + o(1)} \leqslant \chi(G) \leqslant \frac{n}{\gamma + \frac{2}{\log b} - \frac{1}{2} - \frac{1}{1-\sqrt{1-p}} + o(1)}. \tag{3.2}$$

In particular, whp

$$\chi(G) = \frac{n}{2\log_b n - 2\log_b \log_b n + O(1)}.$$

The current best upper bound was obtained by Fountoulakis, Kang and McDiarmid [22] through a very accurate analysis of Bollobás' general approach, and the best known lower bound comes from a first moment argument due to Panagiotou and Steger [50]:

$$\frac{n}{\gamma + o(1)} \leqslant \chi(G) \leqslant \frac{n}{\gamma - 1 + o(1)}. \tag{3.3}$$

It should be noted that the proof of (3.2) does in fact yield the lower bound

$$\chi(G) \geqslant \frac{n}{\left\lfloor \gamma + \frac{2}{\log b} + o(1) \right\rfloor + o(1)}$$

whp, which is sharper than the lower bound in (3.3) for infinitely many n if $p > 1 - 1/e^2$.

As observed in [22], considering (3.3) in terms of the *colouring rate* $\bar{\alpha}(G) = n/\chi(G)$, defined as the average colour class size of a proper colouring with the minimum number of colours, these inequalities provide an explicit interval of length $1 + o(1)$ which contains $\bar{\alpha}(G)$ whp. In [35], Kang and McDiarmid remark that it is a natural problem to determine the value of $\bar{\alpha}(G)$ up to an error of size $o(1)$.

The following result settles this question, giving new upper and lower bounds for $\chi(G)$ which match up to the $o(1)$ term in the denominator.

Theorem 3.1. *Let $p \in (0,1)$ be constant, and consider the random graph $G \sim \mathcal{G}(n,p)$. Let $q = 1-p$, $b = \frac{1}{q}$, $\gamma = \gamma_p(n) = 2\log_b n - 2\log_b \log_b n - 2\log_b 2$ and $\Delta = \Delta_p(n) = \gamma - \lfloor \gamma \rfloor$. Then whp,*

$$\chi(G) = \frac{n}{\gamma - x_0 + o(1)},$$

where $x_0 \geqslant 0$ is the smallest nonnegative solution of

$$(1-\Delta+x)\log_b(1-\Delta+x) + \frac{(\Delta-x)(1-\Delta)}{2} \leqslant 0. \tag{3.4}$$

As $\Delta \geqslant 0$ is a solution of (3.4), x_0 is well-defined, and $0 \leqslant x_0 \leqslant \Delta$. We will see in Section 3.3.2 that for $p \leqslant 1 - 1/e^2$, the smallest nonnegative solution of (3.4) is $x_0 = 0$, while for $p > 1 - 1/e^2$, the solutions of (3.4) depend not only on p but also on n, and we have $0 \leqslant x_0 \leqslant 1 - \frac{2}{\log b}$ (in fact, the values of x_0 are dense in the interval $[0, 1 - \frac{2}{\log b}]$). Therefore, we can derive the following simpler bounds.

Corollary 3.2. *Let $p \in (0,1)$ be constant, $b = \frac{1}{1-p}$ and define γ as in (3.1). Consider the random graph $G \sim \mathcal{G}(n,p)$.*

a) If $p \leqslant 1 - 1/e^2$, then whp,

$$\chi(G) = \frac{n}{\gamma + o(1)}.$$

b) If $p > 1 - 1/e^2$, then whp,

$$\frac{n}{\gamma + o(1)} \leqslant \chi(G) \leqslant \frac{n}{\gamma - 1 + \frac{2}{\log b} + o(1)}.$$

For $p \leqslant 1 - 1/e^2$, the lower bound in Theorem 3.1 is simply the known lower bound (3.3) due to Panagiotou and Steger, which was obtained by estimating the first moment of the number of vertex partitions which induce proper colourings. The *first moment threshold* of this random variable, i.e., the point where the first moment changes from tending to 0 to tending to ∞, occurs at about $\frac{n}{\gamma + o(1)}$ colours.

For $p > 1 - 1/e^2$, we shall also employ the first moment method to establish our new lower bound, although a different first moment threshold will take precedence. The *independence number* $\alpha(G)$ is defined as the size of the largest *independent set* in G, i.e., the largest set of vertices without any edges between them. For $G \sim \mathcal{G}(n,p)$ with p constant, $\alpha(G)$ takes one of at most two explicitly known consecutive values whp (for more details see Section 3.2). In a proper colouring, each colour class forms an independent set, and so no colour class can be larger than $\alpha(G)$. It will turn out that for $p > 1 - 1/e^2$, γ is so close to the likely values of $\alpha(G)$ that the hardest part in colouring G is finding a sufficient number of disjoint independent sets of size $\lceil \gamma \rceil$ or larger. If we colour G with about $\frac{n}{\gamma - x}$ colours for some $x \geqslant 0$, then the average colour class size is about $\gamma - x$. If the independence number $\alpha(G)$ is not much larger than its likely values, then in every such colouring, there must be a certain minimum number of colour classes of size at least $\lceil \gamma \rceil$. Therefore, whp a partial colouring with this number of colour classes of size $\lceil \gamma \rceil$ exists if G is $\frac{n}{\gamma - x}$–colourable. Condition (3.4) describes the first moment threshold of the number of such partial colourings.

The upper bound in Theorem 3.1 is much harder to prove. In contrast to the best previous upper bounds, it will not be obtained through a variant of Bollobás' method but through the second moment method, and our approach will be outlined in Section 3.2. Analysing the second moment of the number of colourings of a random graph is a notoriously hard problem, as it involves examining the joint behaviour of all pairs of possible colourings, which varies considerably depending on how similar they are to each other. Therefore, we will distinguish three different ranges of "overlap" between different pairs of colourings; each range requires different tools and ideas which will be outlined in Section 3.5.1.

Related work

Recall from Section 2.4.5 that Shamir and Spencer [53] showed in 1987 that for any function $p = p(n)$, the chromatic number of $\mathcal{G}(n,p)$ is whp concentrated on an interval of length about $\sqrt{n}$. For constant p, this can be improved to an interval of length about $\sqrt{n}/\log n$ (this is an exercise in Chapter 7.3 of [4], see also [52]). For smaller functions $p = p(n)$, much more is known. Shamir and Spencer also showed that for $p = n^{-c}$ where $c > 1/2$ is constant, the chromatic number of $\mathcal{G}(n,p)$

is concentrated on an interval of *constant length*. Łuczak [40] improved this to an interval of length 2 for $c > 5/6$, and finally Alon and Krivelevich [3] showed two-point concentration for all $c > 1/2$.

None of these concentration results gives any information about the *location* of the intervals, however. Łuczak [39] determined the asymptotic value of the chromatic number of $G \sim \mathcal{G}(n,p)$ whenever $d_0/n \leqslant p = o(1)$ for some large enough constant d_0. For $p = d/n$, where d is constant, Achlioptas and Naor [2] gave two explicit values which the chromatic number of $\mathcal{G}(n,p)$ may take whp, and determined the chromatic number exactly for roughly half of all values d. Recently, Coja-Oghlan and Vilenchik [18] extended this result to almost all constant values d. For $p = n^{-c}$ where $3/4 < c \leqslant 1$, Coja-Oghlan, Panagiotou and Steger [16] gave three explicit values for the chromatic number.

In another direction, the search for sharp thresholds for k-colourability has received a lot of attention. Achlioptas and Friedgut [1] showed that for every fixed k there is a sharp threshold sequence $\frac{d_k(n)}{n}$ such that for any $\varepsilon > 0$, if $p = (d_k - \varepsilon)/n$, then whp $G \sim \mathcal{G}(n,p)$ is k-colourable, and if $p = (d_k + \varepsilon)/n$, then whp $G \sim \mathcal{G}(n,p)$ is not k-colourable. It is unknown whether the sequence $d_k(n)$ converges, but it follows from the results of Achlioptas and Naor [2] that if it does, it lies in an explicit interval of length of order $O(\log k)$. Recently, there has been considerable progress in this area by Coja-Oghlan and Vilenchik [17] and Coja-Oghlan [15], who used ideas inspired by statistical physics to narrow down the length of this interval to roughly 0.39.

3.2 Outline

From now on, let $p \in (0,1)$ be constant and $G \sim \mathcal{G}(n,p)$.

Independence number, first moment method and the lower bound

The chromatic number $\chi(G)$ is closely linked to the independence number $\alpha(G)$, and the behaviour of the independence number of random graphs is very well understood.

Recall that $b = 1/(1-p)$, and let

$$\alpha_0 = 2\log_b n - 2\log_b \log_b n + 2\log_b (e/2) + 1.$$

For p constant, Matula [41, 42] and independently Bollobás and Erdős [9] showed in the 1970s that whp,

$$\alpha(G) = \lfloor \alpha_0 + o(1) \rfloor = \left\lfloor \gamma + \frac{2}{\log b} + 1 + o(1) \right\rfloor, \tag{3.5}$$

pinning down $\alpha(G)$ to at most two values whp.

In a proper colouring each colour class forms an independent set, so for any graph G, $\chi(G) \geqslant n/\alpha(G)$. For a long time, the best known lower bound for the chromatic number of dense random graphs was obtained from this simple fact. McDiarmid [46] sharpened this to $n/(\alpha_0 - 1 + o(1))$ by considering the first moment of the number of independent sets of a certain size, and finally Panagiotou and Steger [50] used a first moment argument on the number of colourings to prove the lower bound $\frac{n}{\gamma+o(1)}$.

Recall from Section 2.4.3 that if $X = X(n) \geqslant 0$ is a sequence of nonnegative integer random variables such that $\mathbb{E}[X]$ tends to 0, then by Markov's inequality, $\mathbb{P}(X > 0) = \mathbb{P}(X \geqslant 1)$ tends to 0 as well. In [50], X is the number of all vertex partitions of G which induce valid colourings (i.e., unordered colourings) with $\frac{n}{\gamma+o(1)}$ colours. Since $\mathbb{E}[X] \to 0$ for an appropriate choice of the $o(1)$-term in the denominator, it follows that whp no proper colouring with this number of colours exists, and the lower bound (3.3) follows.

It turns out, however, that for $p > 1 - 1/e^2$, the chromatic number of $\mathcal{G}(n,p)$ can not in general be found near $\frac{n}{\gamma}$. This is because for colourings with about $\frac{n}{\gamma}$ colours, the average colour class size γ gets so close to $\alpha(G)$ that there are simply not enough disjoint independent sets of size at least $a := \lceil \gamma \rceil$.

Note that in this case $\alpha_0 - \gamma = \frac{2}{\log b} + 1 < 2$, so it follows from (3.5) that $\alpha(G) = a$ or $\alpha(G) = a + 1$ whp as shown in Figure 3.1. In particular, there are whp no independent sets larger than $a + 1$. Therefore, any colouring with average colour class size about γ must contain a certain proportion of colour classes of size at least a (and at most $a + 1$).

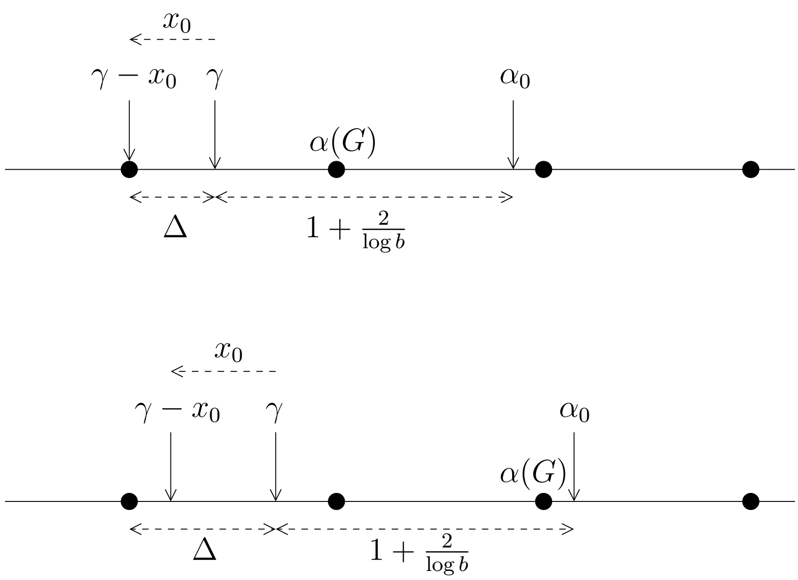

Figure 3.1: If $p > 1 - 1/e^2$, whp either $\alpha(G) = a = \lceil \gamma \rceil$ (top picture) or $\alpha(G) = a + 1$ (bottom picture). In the first case, there are only $o\,(n/\log n)$ independent sets of size $\lceil \gamma \rceil$, and so the colouring rate $n/\chi(G)$ drops back to the next smaller integer $\lfloor \gamma \rfloor$, i.e., $x_0 = \Delta$. In the second case, there are enough independent sets of size $\lceil \gamma \rceil$, but not necessarily enough *disjoint* ones, and we have to correct the colouring rate $n/\chi(G)$ by $x_0 \in [0, 1 - \frac{2}{\log b}]$ to reflect this.

In Section 3.4, we shall consider the number of such partial colourings with large colour classes (or rather, the number of sets of disjoint large independent sets inducing them) which are required for colourings with average colour class size of a little more than $\gamma - x_0$, where x_0 is the solution of (3.4). We will show that their expected number is $o(1)$, so whp no such partial colouring and hence no such complete colouring of G exists.

The second moment method and the upper bound

The upper bound in Theorem 3.1 will be proved using the second moment method. Fix an arbitrary $\varepsilon \in (0, 1)$. If we can show that whp,

$$\chi(G) \leqslant \frac{n}{\gamma - x_0 - 2\varepsilon}, \tag{3.6}$$

this suffices to establish the upper bound in Theorem 3.1. Letting

$$k = k(n) = \left\lceil \frac{n}{\gamma - x_0 - \varepsilon} \right\rceil, \tag{3.7}$$

we shall study k-colourings of G. We will only consider *equitable k-colourings* where the sizes of the colour classes differ by at most 1 (because the method fails if we allow general colourings).

We call a vertex partition into k parts a *k-equipartition* if the part sizes differ by at most 1. We call an ordered partition an *ordered k-equipartition* if the parts sizes differ by at most 1 and decrease in size (so the parts of size $\lceil \frac{n}{k} \rceil$ come first, followed by the parts of size $\lfloor \frac{n}{k} \rfloor$).

Denote by Z_k the *number of ordered k-equipartitions which induce proper colourings*, i.e., where all parts form independent sets. Then our goal will be to bound the second moment of Z_k in terms of $\mathbb{E}[Z_k]^2$. More specifically, our aim will be to show that for n large enough,

$$\frac{\mathbb{E}[Z_k^2]}{\mathbb{E}[Z_k]^2} \leqslant \exp\left(\frac{n}{\log^7 n}\right). \tag{3.8}$$

Let us briefly discuss why (3.8) suffices to prove (3.6). By the Paley–Zygmund Inequality, Theorem 2.7, (3.8) implies that

$$\mathbb{P}\left(Z_k > 0\right) \geqslant \frac{\mathbb{E}[Z_k]^2}{\mathbb{E}[Z_k^2]} \geqslant \exp\left(-\frac{n}{\log^7 n}\right) \tag{3.9}$$

for large enough n. Therefore, for n large enough,

$$\mathbb{P}(\chi(G) \leqslant k) \geqslant \mathbb{P}\left(Z_k > 0\right) \geqslant \exp\left(-\frac{n}{\log^7 n}\right). \tag{3.10}$$

The term on the right-hand side of course tends to 0, so it may at first seem that (3.10) is not particularly helpful in proving (3.6). However, as first noted by Frieze in [24], all is not lost in cases like these where we have a lower bound on a probability which tends to 0 sufficiently slowly. Using the Bounded Differences Inequality (Theorem 2.8, also often referred to as the Azuma–Hoeffding or McDiarmid Inequality), we will see that the chromatic number of random graphs is concentrated so tightly around its mean that by adding only a few additional colours, we can boost the lower bound (3.10) to a bound which tends to 1.

Indeed, consider $G \sim \mathcal{G}(n,p)$: arbitrarily modifying the edges incident with any particular vertex of G can change the value of $\chi(G)$ by at most 1. Therefore, by the

Bounded Differences Inequality, Theorem 2.8,

$$\mathbb{P}\Big(|\chi(G) - \mathbb{E}\left(\chi(G)\right)| \geqslant t \Big) \leqslant 2\exp\left(-\frac{2t^2}{n}\right). \tag{3.11}$$

This implies that $k \geqslant \mathbb{E}\left(\chi(G)\right) - \frac{n}{\log^3 n}$ for n large enough, because otherwise (3.11) with $t = \frac{n}{\log^3 n}$ would contradict (3.10). But then again by (3.11), if we let $\hat{k} = k + \frac{2n}{\log^3 n}$,

$$\mathbb{P}\left(\chi(G) > \hat{k}\right) \leqslant \mathbb{P}\left(\chi(G) > \mathbb{E}\left(\chi(G)\right) + \frac{n}{\log^3 n}\right) \to 0$$

as $n \to \infty$. Therefore, whp

$$\chi(G) \leqslant \hat{k} = k + \frac{2n}{\log^3 n} \leqslant \frac{n}{\gamma - x_0 - 2\varepsilon},$$

as required.

So to prove the upper bound in Theorem 3.1, it remains to show (3.8). Note that as

$$Z_k = \sum_{\pi \text{ an ordered } k\text{-equipartition}} \mathbb{1}_{\{\pi \text{ induces a proper colouring}\}},$$

by linearity of the expectation,

$$\mathbb{E}[Z_k^2] = \sum_{\pi_1, \pi_2 \text{ ordered } k\text{-equipartitions}} \mathbb{P}\left(\text{both } \pi_1 \text{ and } \pi_2 \text{ induce proper colourings}\right), \tag{3.12}$$

where the joint probability that both π_1 and π_2 induce proper colourings of course depends critically on how similar they are.

Classifying the amount of overlap between π_1 and π_2 and splitting up the calculation into manageable cases will be the main challenge of the proof. In Section 3.5, we will quantify the amount of overlap between two partitions, and in Sections 3.5.2–3.5.4, we will proceed to distinguish three different ranges of overlap and bound their respective contributions to (3.8). Each range will be tackled through a different approach. A more detailed overview of the different ideas and tools for each range is given in Section 3.5.1.

Remark

Like Bollobás' original proof of the asymptotic upper bound [6], our proof of the upper bound in Theorem 3.1 requires martingale concentration inequalities. This is necessary because for our choice of k, $\mathbb{E}[Z_k^2]/\mathbb{E}[Z_k]^2 \nrightarrow 1$, so the second moment method alone cannot yield the whp existence of a colouring.

However, it is possible to obtain an upper bound of the form $\frac{n}{\gamma - O(1)}$ using only the second moment method. For this, we need to work in $\mathcal{G}(n,m)$ with $m \approx p\binom{n}{2}$ instead of $\mathcal{G}(n,p)$ in order to control the variations in Z_k which can be attributed to variations in the number of edges. Furthermore, all colour classes need to be of exactly the same size $\gamma - t$ for some $t = O(1)$ which will be specified below. As $\gamma - t$ does not necessarily divide n, we may need to include up to $\gamma - t = O(\log n)$ extra vertices so that all colour classes are of exactly the same size — these vertices can then simply be removed afterwards once we have proved the whp existence of a colouring.

If n' denotes the number of vertices and Z denotes the number of equitable $\frac{n'}{\gamma - t}$-colourings of $G \sim \mathcal{G}(n', m)$, then we can show that $\mathbb{E}[Z^2]/E[Z]^2 \to 1$ for $t = \Delta + 1$ (yielding the upper bound $\frac{n+\gamma-\Delta-1}{\gamma-\Delta-1} = \frac{n}{\lfloor\gamma\rfloor-1} + 1$), and possibly also for $t = \Delta$. Choosing $t = \Delta + 1$ would also simplify the proof considerably, as much of the technical difficulty stems from pairs of partitions which overlap in subsets of parts of size at least $\lfloor\gamma\rfloor$, which is not possible if all parts of the partitions are of size $\gamma - \Delta - 1 = \lfloor\gamma\rfloor - 1$.

In the next chapter, we will prove a related result about the equitable chromatic number of dense random graphs. In Section 4.6, we will see how the arguments and conditions from this chapter can be adapted in this case to yield $\mathbb{E}[Z^2]/\mathbb{E}[Z]^2 \to 1$.

3.3 Preliminaries and notation

From now on, we will always assume that n is large enough so that various bounds and approximations hold, even when this is not stated explicitly. Recall that $\gamma =$

$2\log_b n - 2\log_b \log_b n - 2\log_b 2$ and that we fixed an arbitrary $\varepsilon \in (0,1)$, and let

$$k = \left\lceil \frac{n}{\gamma - x_0 - \varepsilon} \right\rceil .$$

Similarly, let

$$l = \left\lfloor \frac{n}{\gamma - x_0 + \varepsilon} \right\rfloor .$$

We are going to show that for any fixed $\varepsilon > 0$, whp $\chi(G) \geqslant l$, and that

$$\mathbb{P}\left(\chi(G) \leqslant k\right) \geqslant \exp\left(-\frac{n}{\log^7 n}\right).$$

Furthermore, let

$$\delta = \frac{n}{k} - \left\lfloor \frac{n}{k} \right\rfloor, \qquad k_1 = \delta k \quad \text{ and } \quad k_2 = (1-\delta)k.$$

If k does not divide n, then a *k-equipartition* consists of exactly k_1 parts of size $\left\lceil \frac{n}{k} \right\rceil$ and exactly k_2 parts of size $\left\lfloor \frac{n}{k} \right\rfloor$. If k divides n, then $k_1 = 0$ and $k_2 = k$, and a k-equipartition consists of exactly k_2 parts of size $\left\lfloor \frac{n}{k} \right\rfloor = \frac{n}{k}$. In an *ordered k-equipartition*, the first k_1 parts are of size $\left\lceil \frac{n}{k} \right\rceil$ and the remaining k_2 parts are of size $\left\lfloor \frac{n}{k} \right\rfloor$.

Let P denote the total number of ordered k-equipartitions of the n vertices, then

$$P = \frac{n!}{\left\lceil \frac{n}{k} \right\rceil !^{k_1} \left\lfloor \frac{n}{k} \right\rfloor !^{k_2}}. \tag{3.13}$$

Since by Stirling's approximation, $n! = \Theta\left(n^{n+1/2}e^{-n}\right)$,

$$P = k^n \exp(o(n)). \tag{3.14}$$

Given a k-equipartition, there are exactly

$$f = k_1 \cdot \binom{\lceil n/k \rceil}{2} + k_2 \cdot \binom{\lfloor n/k \rfloor}{2} = \frac{n\left(\frac{n}{k} - 1\right)}{2} + \frac{\delta(1-\delta)}{2} \cdot k \tag{3.15}$$

forbidden edges which are not present in G if the partition induces a proper colouring. Therefore, the probability that a given ordered k-equipartition induces a proper colouring is exactly q^f, so

$$\mu_k := \mathbb{E}[Z_k] = Pq^f. \tag{3.16}$$

Considering the known upper bound for $\chi(G)$ in (3.3), we may assume that

$$\frac{n}{k} = \gamma - x_0 - \varepsilon + o(1) \geqslant \gamma - 1,$$

since we can simply use (3.3) instead for all n where this is not the case. Recalling that

$$a = \lceil \gamma \rceil = \gamma + 1 - \Delta,$$

then by our assumption $\lceil \frac{n}{k} \rceil \in \{a-1, a\}$.

For easier notation, we will also assume that γ is not an integer, and that therefore $\lfloor \gamma \rfloor = a - 1$. We may replace γ by $\gamma - 1/n$, say, for all n where this is not the case (if there are any such n at all), and the statement of the theorem and its proof remain unchanged.

3.3.1 List of key facts and relations

Below is a list of some facts, bounds and approximations so that we can conveniently refer back to them later on.

(A) If $p > 1 - 1/e^2$, then whp $\alpha(G) \in \{a, a+1\}$, where $\alpha(G)$ denotes the independence number and $a = \lceil \gamma \rceil$ (as remarked in Section 3.2).

(B) By our assumption, $\lceil \frac{n}{k} \rceil \in \{a, a-1\}$, where $a = \lceil \gamma \rceil$. We may also assume that γ is not an integer.

(C) In a k-equipartition, there are $k_1 = \delta k$ parts of size $\lceil \frac{n}{k} \rceil$ and $k_2 = (1-\delta)k$ parts of size $\lfloor \frac{n}{k} \rfloor$, where $\delta = \frac{n}{k} - \lfloor \frac{n}{k} \rfloor$.

(D) $\gamma \sim a \sim \frac{n}{k} \sim \frac{n}{l} \sim 2\log_b n = \Theta(\log n)$

(E) $k \sim l \sim \frac{n}{\gamma} \sim \frac{n}{a} \sim \frac{n}{2\log_b n} = \Theta\left(\frac{n}{\log n}\right)$

(F) $q^{-\gamma/2} = b^{\gamma/2} = \frac{n}{2\log_b n} \sim k \sim l$

(G) $k^{\frac{1}{n/k}} = O(1)$ and $k^{\frac{1}{\log n}} = O(1)$.

(H) $f \sim n \log_b n$.

(J) For any integer function $\varphi = \varphi(n) = o(n)$, $\binom{n}{\varphi} \leqslant \exp(o(n))$.

(K) $\frac{\mu_k}{k_1!k_2!} \geqslant b^{\varepsilon n/4}$.

Proof. Note that $k_1!k_2! \leqslant k! \exp(o(n))$, since $\binom{k}{k_1} \leqslant 2^k$. Furthermore, by (E) we have $k \sim \frac{n}{2\log_b n}$, so with Stirling's formula, $k! = k^{k(1+o(1))} = n^{k(1+o(1))} = b^{n\left(\frac{1}{2}+o(1)\right)}$. From the definition (3.15) of f, we can see that $q^f = b^{-f} = b^{-\frac{n^2}{2k}+\frac{n}{2}} \exp(o(n))$. Therefore, from (3.14) and (3.16),

$$\frac{\mu_k}{k_1!k_2!} = \frac{Pq^f}{k_1!k_2!} = \frac{k^n q^f}{k!} \exp(o(n)) = \left(kb^{-\frac{n}{2k}}\right)^n \exp(o(n)).$$

Note that $\frac{n}{k} = \gamma - x_0 - \varepsilon + o(1) \leqslant \gamma - \varepsilon + o(1)$, so by (F),

$$kb^{-\frac{n}{2k}} \geqslant kb^{\frac{-\gamma+\varepsilon+o(1)}{2}} \sim b^{\varepsilon/2},$$

and therefore, for n large enough, $\frac{\mu_k}{k_1!k_2!} \geqslant b^{\varepsilon n/4}$ as required. □

3.3.2 On the solutions of (3.4)

In this section we will explore the solutions of the inequality (3.4) in Theorem 3.1, and prove some technical lemmas which will be needed later. All the proofs in this section are straightforward analytical arguments, so the reader might wish to skip the details.

Let

$$f(x) = f_n(x) = (1 - \Delta + x)\log_b(1 - \Delta + x) + (1 - \Delta)(\Delta - x)/2.$$

Then x_0 is defined in Theorem 3.1 as the smallest nonnegative solution of $f(x) \leqslant 0$. Since $f(\Delta) = 0$, x_0 is well-defined and $x_0 \in [0, \Delta]$. Note that

$$f'(x) = \log_b(1 - \Delta + x) + \frac{1}{\log b} - \frac{1 - \Delta}{2}$$
$$f''(x) = \frac{1}{(1 - \Delta + x)\log b} \geqslant 0.$$

Therefore, f is convex and there are three different possible cases for the location of x_0 as shown in Figure 3.2. In the first case, $f(0) \leqslant 0$ and therefore $x_0 = 0$. In the second and third case, $f(0) > 0$, so $x_0 > 0$. In the second case, x_0 lies strictly between 0 and Δ, and in the third case, $x_0 = \Delta$, which happens if and only if $f'(\Delta) \leqslant 0$, or equivalently $1 - \Delta \geqslant \frac{2}{\log b}$. This case corresponds to the upper picture in Figure 3.1.

We will first prove the two lemmas needed to obtain Corollary 3.2 from Theorem 3.1.

Lemma 3.3. *If $p \leqslant 1 - 1/e^2$, then $x_0 = 0$.*

Proof. Note that $p \leqslant 1 - 1/e^2$ is equivalent to $\log b \leqslant 2$, and therefore,

$$\log_b(1-\Delta) + \frac{\Delta}{2} \leqslant \frac{1}{2}\big(\log(1-\Delta) + \Delta\big) \leqslant 0,$$

since $\log(1-y) \leqslant -y$ for all $y \in [0,1)$. Hence, $f(0) \leqslant 0$. □

Lemma 3.4. *If $p > 1 - 1/e^2$, then $0 \leqslant x_0 \leqslant 1 - \frac{2}{\log b}$.*

Proof. Of course $x_0 \geqslant 0$ is true by definition, and if $\Delta \leqslant 1 - \frac{2}{\log b}$, then $x_0 \leqslant \Delta \leqslant 1 - \frac{2}{\log b}$. So suppose $\Delta > 1 - \frac{2}{\log b}$, then the claim follows if we can show $f\left(1 - \frac{2}{\log b}\right) \leqslant 0$. Note that $f\left(1 - \frac{2}{\log b}\right) \leqslant 0$ is equivalent to $g\left(2 - \frac{2}{\log b} - \Delta\right) \leqslant 0$, where

$$g(y) = y \log y + \frac{\log b}{2}\left(y + \frac{2}{\log b} - 1\right)(1-y).$$

Note that $g'(y) = \log y - y \log b + \log b$ and $g''(y) = \frac{1}{y} - \log b$.

The function g has no maximum in $\left(1 - \frac{2}{\log b}, 1\right)$: suppose that $y \in \left(1 - \frac{2}{\log b}, 1\right)$ with $g'(y) = 0$ and $g''(y) \leqslant 0$. It follows that $0 = \log y + (1-y)\log b \geqslant \log y + \frac{1-y}{y}$, but this is a contradiction since $\log z + \frac{1-z}{z} > 0$ for all $z \in (0,1)$.

In the boundary cases $y = 1 - \frac{2}{\log b}$ and $y = 1$, we have $g(y) \leqslant 0$. Since $1 - \frac{2}{\log b} < \Delta \leqslant 1$, it follows that $2 - \frac{2}{\log b} - \Delta \in \left[1 - \frac{2}{\log b}, 1\right)$, and therefore $g\left(2 - \frac{2}{\log b} - \Delta\right) \leqslant 0$ as required. □

We now proceed with the technical lemmas which we will need later on.

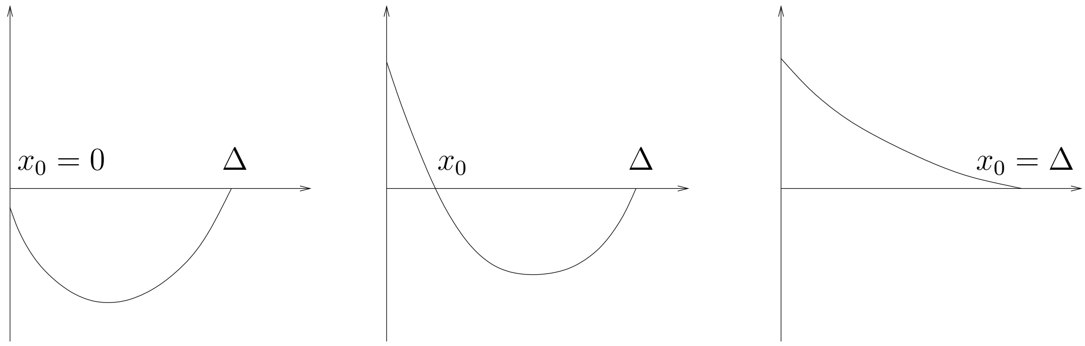

Figure 3.2: The three possible cases for the function $f(x)$. In the first case, $f(0) \leqslant 0$, so $x_0 = 0$. In the second case, $x_0 \in (0, \Delta)$. In the third case, Δ is the smallest nonnegative solution of $f(x) \leqslant 0$, so $x_0 = \Delta$.

Lemma 3.5. *Suppose $p > 1 - 1/e^2$, and fix $\varepsilon' > 0$. Then there is a constant $c_1 = c_1(\varepsilon') > 0$ such that if $x_0 > \varepsilon'$, then*

$$f(x_0 - \varepsilon') \geqslant c_1.$$

Proof. Since $x_0 > 0$, the definition of x_0 implies that $f(x) > 0$ for all $x \in [0, x_0)$, so by continuity $f(x_0) = 0$ and furthermore $f'(x_0) \leqslant 0$. As $f'' \geqslant \frac{1}{\log b}$ on $[0, \Delta]$ for all n, f'' is strongly convex on $[0, \Delta]$ with parameter at least $\frac{1}{\log b}$ for all n, and the claim follows. □

Lemma 3.6. *There is a constant $c_2 = c_2(\varepsilon) \in (0, 1)$ such that if $x_0 \leqslant \Delta - \varepsilon$, then*

$$1 - \Delta \leqslant \frac{2c_2}{\log b}.$$

Proof. As $f(\Delta) = 0$ and $f(x_0) \leqslant 0$, there cannot be an $x \in (x_0, \Delta)$ such that $f(x) > 0$, otherwise there would have to be a local maximum which is impossible since $f'' > 0$.

So $f(\Delta - \varepsilon) \leqslant 0$ and rearranging terms gives

$$1 - \Delta \leqslant -\frac{(1 - \varepsilon) \log(1 - \varepsilon)}{\varepsilon} \cdot \frac{2}{\log b},$$

so we can let $c_2 = -\frac{(1-\varepsilon)\log(1-\varepsilon)}{\varepsilon} \in (0, 1)$ as $\varepsilon \in (0, 1)$. □

Lemma 3.7. *Fix $\varepsilon' > 0$. There is a constant $c_3 = c_3(\varepsilon, \varepsilon') > 0$ such that if $\varepsilon' \leqslant y \leqslant \Delta - x_0 - \varepsilon$, then*

$$f(\Delta - y) \leqslant -c_3.$$

Proof. Let

$$c_3 = \min\left(\frac{\varepsilon^2}{4\log b}, \frac{\varepsilon'^2}{4\log b}\right) > 0.$$

Note that $x_0 + \varepsilon \leqslant \Delta - y \leqslant \Delta - \varepsilon'$. As $f'' > 0$, f has no internal maxima in $(x_0 + \varepsilon, \Delta - \varepsilon')$, so

$$f(\Delta - y) \leqslant \max\left(f(x_0 + \varepsilon), f(\Delta - \varepsilon')\right).$$

We distinguish two cases.

- **Case 1:** $f(x_0 + \varepsilon) \geqslant f(\Delta - \varepsilon')$

  Then as f' is increasing, $f'(x_0 + \varepsilon) = \log_b(1 - \Delta + x_0 + \varepsilon) + \frac{1}{\log b} - \frac{1-\Delta}{2} \leqslant 0$. For any $z_1, z_2 \geqslant 0$ with $z_1 + z_2 < 1$, we have $\log(z_1 + z_2) \geqslant \log(z_1) + z_2$, so

  $$f'\left(x_0 + \frac{\varepsilon}{2}\right) \leqslant f'(x_0 + \varepsilon) - \frac{\varepsilon}{2\log b} \leqslant -\frac{\varepsilon}{2\log b}.$$

  As f' is increasing, $f'(x) < 0$ for all $x \in [x_0, x_0 + \varepsilon]$, and since by definition $f(x_0) \leqslant 0$,

  $$\begin{aligned} f\left(x_0 + \varepsilon\right) &\leqslant \int_{x_0}^{x_0+\varepsilon} f'(x)\,\mathrm{d}x \leqslant \int_{x_0}^{x_0+\varepsilon/2} f'(x)\,\mathrm{d}x \leqslant \frac{\varepsilon}{2} f'\left(x_0 + \frac{\varepsilon}{2}\right) \\ &\leqslant -\frac{\varepsilon^2}{4\log b} \leqslant -c_3. \end{aligned}$$

- **Case 2:** $f(x_0 + \varepsilon) < f(\Delta - \varepsilon')$

  Then as f' is increasing, $f'(\Delta - \varepsilon') = \log_b(1 - \varepsilon') + \frac{1}{\log b} - \frac{1-\Delta}{2} \geqslant 0$. For any $z_1, z_2 \geqslant 0$ with $z_1 + z_2 < 1$, we have $\log(z_1 + z_2) \geqslant \log(z_1) + z_2$, so

  $$f'\left(\Delta - \frac{\varepsilon'}{2}\right) \geqslant f'(\Delta - \varepsilon') + \frac{\varepsilon'}{2\log b} \geqslant \frac{\varepsilon'}{2\log b}.$$

  As f' is increasing, $f'(x) \geqslant 0$ for all $x \geqslant \Delta - \varepsilon$, and since $f(\Delta) = 0$,

$$f(\Delta - \varepsilon') = -\int_{\Delta-\varepsilon'}^{\Delta} f'(x)\,\mathrm{d}x \leqslant -\int_{\Delta-\varepsilon'/2}^{\Delta} f'(x)\,\mathrm{d}x \leqslant -\frac{\varepsilon'}{2} f'\left(\Delta - \frac{\varepsilon'}{2}\right)$$
$$\leqslant -\frac{\varepsilon'^2}{4\log b} \leqslant -c_3.$$

□

Lemma 3.8. *Fix $\varepsilon' > 0$. There is a constant $c_4 = c_4(\varepsilon, \varepsilon') > 0$ such that if $\varepsilon' \leqslant \Delta - x_0 - \varepsilon \leqslant y \leqslant 1$, then*

$$(1-y)\log_b(1-y) + \frac{\Delta}{2}(1-y) - \frac{x_0+\varepsilon}{2} \leqslant -c_4.$$

Proof. Let

$$h(x) = (1-x)\log_b(1-x) + \frac{\Delta}{2}(1-x) - \frac{x_0+\varepsilon}{2}.$$

Note that $h(\Delta - x_0 - \varepsilon) = f(x_0+\varepsilon)$. Furthermore, $\lim_{x\to 1} h(x) = -\frac{x_0+\varepsilon}{2} \leqslant -\frac{\varepsilon}{2}$. Since $h''(x) = \frac{1}{(1-x)\log b} > 0$ for $x \in (0,1)$, h has no internal maxima in $(0,1)$, so since $\Delta - x_0 - \varepsilon \leqslant y \leqslant 1$,

$$h(y) \leqslant \max\left(h(\Delta - x_0 - \varepsilon), -\frac{\varepsilon}{2}\right) \leqslant \max\left(f(x_0+\varepsilon), -\frac{\varepsilon}{2}\right).$$

Applying Lemma 3.7 to $y' = \Delta - x_0 - \varepsilon$, we can see that $f(x_0+\varepsilon) \leqslant -c_3(\varepsilon, \varepsilon')$. Letting

$$c_4 = \min\left(c_3(\varepsilon, \varepsilon'), \frac{\varepsilon}{2}\right) > 0,$$

it follows that $h(y) \leqslant -c_4$ for all $\Delta - x_0 - \varepsilon \leqslant y \leqslant 1$. □

3.4 Proof of the lower bound

We may assume that $p > 1 - 1/e^2$, because otherwise $x_0 = 0$ and the lower bound in Theorem 3.1 is simply the known lower bound (3.3). Recall that we defined $l = \left\lfloor \frac{n}{\gamma_p(n) - x_0 + \varepsilon} \right\rfloor$ for an arbitrary fixed $\varepsilon \in (0,1)$. We may assume $x_0 - \varepsilon \geqslant 0$, because we can just use the known bound (3.3) instead for all n where this is not the case. We will show that any l-colouring must contain a certain proportion of large colour classes of size $a = \lceil \gamma \rceil$, and then prove that the expected number of unordered partial colourings with just these large colour classes tends to 0, which means that whp no such partial and therefore no complete l-colouring exists.

As $x_0 \leqslant \Delta$ (see Section 3.3.2),

$$\lceil\gamma\rceil \geqslant \lceil\gamma - x_0 + \varepsilon\rceil \geqslant \lceil\gamma\rceil,$$

so $\lceil\gamma - x_0 + \varepsilon\rceil = \lceil\gamma\rceil =: a$. This is very close to the independence number $\alpha(G)$: by (A) from Section 3.3.1, whp $\alpha(G) = a$ or $\alpha(G) = a+1$. In particular, whp there are no independent sets of size $a+2$ in G.

Recall that $\alpha_0 = \gamma + 1 + \frac{2}{\log b}$. Standard calculations show that for any $t = t(n) = O(1)$ such that $\alpha_0 - t$ is an integer, the expected number of independent sets of size $\alpha_0 - t$ in G is $n^{t+o(1)}$ (see also 3.c) in [45]). Therefore, as $a = \lceil\gamma\rceil = \gamma + 1 - \Delta = \alpha_0 - \frac{2}{\log b} - \Delta$,

$$\binom{n}{a} q^{\binom{a}{2}} = n^{\frac{2}{\log b} + \Delta + o(1)}$$
$$\binom{n}{a+1} q^{\binom{a+1}{2}} = n^{\frac{2}{\log b} + \Delta - 1 + o(1)}. \tag{3.17}$$

Note that as $\frac{2}{\log b} + \Delta - 1 \leqslant \frac{2}{\log b} < 1$ and by (E), $l = \Theta\left(\frac{n}{\log n}\right)$, it follows from Markov's inequality that whp only $o(l)$ independent sets of size $a+1$ exist in G.

We assume from now on that no independent sets of size $a+2$ and only $o(l)$ independent sets of size $a+1$ are present in G, both of which hold whp. A valid l-colouring consists of l disjoint independent sets of average size $\frac{n}{l} \geqslant \gamma - x_0 + \varepsilon$, and therefore a certain proportion of the l colour classes must be of size a in order to obtain an average colour class size of at least $\gamma - x_0 + \varepsilon$. More specifically, given an l-colouring, if we let $y \in [0,1]$ be the proportion of colour classes of size a in the colouring, and let $z = o(1)$ be such that there are exactly zl independent sets of size $a+1$ in G, then adding up the number of vertices in each colour class yields

$$n \leqslant ayl + (a+1)zl + (a-1)(1-y-z)l.$$

Therefore, since $n/l \geqslant \gamma - x_0 + \varepsilon$,

$$\gamma - x_0 + \varepsilon \leqslant y + a - 1 + 2z.$$

As $a = \gamma + 1 - \Delta$ and $z = o(1)$, it follows that

$$y \geqslant \Delta - x_0 + \varepsilon + o(1).$$

Hence, as $l \sim \frac{n}{2\log_b n}$ by (E), if a proper l-colouring exists and n is large enough, then in particular G must contain at least

$$s := \left\lceil \frac{(\Delta - x_0 + \varepsilon/2)\, n}{2\log_b n} \right\rceil$$

disjoint independent sets of size a. We shall call such an (unordered) collection of s disjoint independent sets of size a a *precolouring*, and denote by $\bar{Z}$ the number of precolourings in G.

Since for all $m \in \mathbb{N}$, $m^m/e^m \leqslant m! \leqslant m^{m+O(1)}/e^m$,

$$\begin{aligned}
\mathbb{E}[\bar{Z}] &= \frac{1}{s!}\binom{n}{a}\binom{n-a}{a}\cdots\binom{n-(s-1)a}{a} q^{s\binom{a}{2}} = \frac{n!\, q^{s\binom{a}{2}}}{s!\, a!^s (n-as)!} \\
&\leqslant \frac{e^{s-as} n^{n+O(1)} q^{s\binom{a}{2}}}{s^s a!^s (n-as)^{n-as}} = n^{O(1)} \left(\frac{n^a q^{\binom{a}{2}}}{e^{a-1} s a!}\right)^s \left(\frac{n}{n-as}\right)^{n-as}.
\end{aligned}$$

By (E), $a \sim 2\log_b n$, so it follows from (3.17) that $\frac{n^a}{a!} q^{\binom{a}{2}} \sim \binom{n}{a} q^{\binom{a}{2}} = n^{\frac{2}{\log b} + \Delta + o(1)}$. Furthermore, $e^{a-1} = n^{\frac{2}{\log b} + o(1)}$. Therefore, since $n^{1-o(1)} \leqslant s \leqslant n$,

$$\mathbb{E}[\bar{Z}] \leqslant n^{O(1)} \left(n^{\Delta + o(1)} s^{-1}\right)^s \left(1 - \frac{as}{n}\right)^{-n\left(1 - \frac{as}{n}\right)} \leqslant e^{o(n)} n^{-s(1-\Delta)} \left(1 - \frac{as}{n}\right)^{-n\left(1 - \frac{as}{n}\right)}.$$

As $as/n \sim \Delta - x_0 + \varepsilon/2$, this gives

$$\begin{aligned}
\mathbb{E}[\bar{Z}] &\leqslant e^{o(n)} n^{-s(1-\Delta)} \left(1 - \Delta + x_0 - \frac{\varepsilon}{2}\right)^{-\left(1 - \Delta + x_0 - \frac{\varepsilon}{2}\right) n} \\
&= b^{-\left(\left(1-\Delta+x_0-\frac{\varepsilon}{2}\right)\log_b\left(1-\Delta+x_0-\frac{\varepsilon}{2}\right) + (1-\Delta)\left(\Delta - x_0 + \frac{\varepsilon}{2}\right)/2 + o(1)\right) n}.
\end{aligned}$$

Note that with the exception of the $o(1)$ term, the expression in the exponent is now simply the left-hand side of condition (3.4) in Theorem 3.1 with $x = x_0 - \varepsilon/2$. As $\varepsilon/2 < \varepsilon \leqslant x_0$, we may apply Lemma 3.5 with $\varepsilon' = \varepsilon/2$ to conclude that

$$\mathbb{E}[\bar{Z}] \leqslant b^{-(c_1 + o(1))n} = o(1).$$

By Markov's inequality, whp no precolouring and consequently no proper l-colouring exists.

3.5 Bounding the second moment

We now proceed to the main part of the proof. Recall that to prove Theorem 3.1, it remains to show that for an arbitrary fixed $\varepsilon \in (0,1)$ and $k = k(n) = \left\lceil \frac{n}{\gamma - x_0 - \varepsilon} \right\rceil$, if n is large enough,

$$\mathbb{E}[Z_k^2]/\mu_k^2 \leqslant \exp\left(\frac{n}{\log^7 n}\right).$$

Because of the known upper bound (3.3), we may assume that $\frac{n}{k} \geqslant \gamma - 1$ and that therefore $x_0 + \varepsilon \leqslant 1$.

By (3.12), in order to bound $\mathbb{E}[Z_k^2]$, we need to study the joint probability that two partitions both induce proper colourings, a quantity which of course depends on how similar the two partitions are. To quantify the amount of overlap between two partitions, we define the *overlap sequence* $\mathbf{r}$. For $2 \leqslant i \leqslant a := \lceil \gamma \rceil$, given two ordered k-equipartitions π_1, π_2, denote by r_i the number of pairs of parts (the first being a part in π_1 and the second being a part in π_2) which intersect in exactly i vertices. Denote by

$$\mathbf{r} = (r_2, r_3, \ldots, r_a)$$

the *overlap sequence* of the two ordered k-equipartitions π_1, π_2. If the intersection of two parts contains at least two vertices, we call the intersection an *overlap block*. If there is only a single vertex in the intersection of two parts, we call that vertex a *singleton*. Note that since

$$\left\lceil \frac{n}{k} \right\rceil \leqslant \lceil \gamma \rceil = a,$$

no overlap block is larger than a.

Conversely, given an overlap sequence $\mathbf{r}$, denote by $P_{\mathbf{r}}$ the number of *ordered pairs* of ordered k-equipartitions with overlap sequence $\mathbf{r}$. Let

$$v = v(\mathbf{r}) = \sum_{i=2}^{a} i r_i$$

be the *number of vertices involved in the overlap*, and let

$$\rho = v/n \leqslant 1$$

denote the *proportion* of those vertices in the graph. Furthermore, denote by

$$d = d(\mathbf{r}) = \sum_{i=2}^{a} r_i \binom{i}{2} \tag{3.18}$$

the number of *common forbidden edges* that two ordered k-equipartitions π_1, π_2 with overlap sequence $\mathbf{r}$ share. Since the number of forbidden edges in one partition is exactly f, where f was defined in (3.15), both π_1 and π_2 induce proper colourings if and only if none of exactly $2f - d$ forbidden edges are present. Therefore, from (3.12) and (3.16),

$$\mathbb{E}[Z_k^2] = \sum_{\mathbf{r}} P_{\mathbf{r}} q^{2f-d} = \mu_k^2 \sum_{\mathbf{r}} \frac{P_{\mathbf{r}}}{P^2} b^d.$$

Let

$$Q_{\mathbf{r}} = \frac{P_{\mathbf{r}}}{P^2}, \tag{3.19}$$

then our goal is to show that for n large enough,

$$\frac{\mathbb{E}[Z_k^2]}{\mu_k^2} = \sum_{\mathbf{r}} Q_{\mathbf{r}} b^d \leqslant \exp\left(\frac{n}{\log^7 n}\right). \tag{3.20}$$

Since the summands in (3.20) vary considerably for different types of overlap sequences $\mathbf{r}$, we split up our calculations into three parts in Sections 3.5.2 – 3.5.4. The behaviour of the summands is rather different in each case, and so different methods and ideas will be required to bound them.

3.5.1 Outline

Typical case

In Section 3.5.2, we first discuss the typical type of overlap between pairs of partitions. If a partition is chosen uniformly at random from all possible ordered k-equipartitions, then the probability that two given vertices are in the same part is roughly $\frac{1}{k}$. Consequently, if two ordered k-equipartitions are sampled independently and uniformly at random, then the expected number d of forbidden edges they have in common, i.e., pairs of vertices which are in the same part in both partitions, is of order $\frac{n^2}{k^2} = O(\log^2 n)$. In particular, we do not expect the number v of vertices involved in the overlap to be much larger than $2d = O(\log^2 n)$.

Furthermore, the expected number of *triangles* which the two partitions have in common, i.e., triples of vertices that are in the same part in both partitions, is of order $\frac{n^3}{k^4} = O\left(\frac{\log^4 n}{n}\right) = o(1)$. Therefore, typically two partitions have no triangles or larger cliques in common, and overlap in about $O(\log^2 n)$ disjoint pairs of vertices.

In fact, we will cover a much larger range of overlap sequences $\mathbf{r}$ in Section 3.5.2, namely those $\mathbf{r}$ where at most a constant fraction of all vertices are involved in the overlap, i.e., where $v = v(\mathbf{r}) \leqslant cn$ for a constant c which will be defined in (3.21).

To bound the number of such pairs of partitions, we will count the number of corresponding *overlap matrices*. The overlap matrix between two partitions π_1 and π_2 is defined as the matrix $\mathcal{M} = (M_{xy})$, where M_{xy} denotes the number of vertices that are in part number x in π_1 and in part number y in π_2. If π_1 and π_2 overlap according to a given overlap sequence $\mathbf{r}$, then the entries of $\mathcal{M}$ are exactly r_i instances of the number i for all $2 \leqslant i \leqslant \left\lceil \frac{n}{k} \right\rceil$, as well as $n - v$ instances of the number 1, with the remaining entries 0. As π_1 and π_2 are ordered k-equipartitions, all rows and columns of $\mathcal{M}$ sum to $\left\lceil \frac{n}{k} \right\rceil$ or $\left\lfloor \frac{n}{k} \right\rfloor$.

Since there are typically few pairs and very few triangles or larger cliques in the overlap, one crucial idea is that we can count the number of overlap matrices by first placing any entries $2, 3, \ldots, \left\lceil \frac{n}{k} \right\rceil$ in the matrix separately, and then treating the rest of the matrix as a $0-1$ matrix with given row and column sums close to $\frac{n}{k}$. An important tool is Theorem 3.9, due to McKay, which gives an estimate for the number of $0-1$ matrices with prescribed row and column sums.

After some fairly accurate calculations, we will see that the contribution from each $\mathbf{r}$ in this case is bounded by an expression of the form $\prod_{i=2}^{a} \frac{T_i^{r_i}}{r_i!}$, where the terms T_i still depend on $\rho(\mathbf{r}) = v/n$. We will then show that if $\rho \leqslant c$, the terms T_i are small enough so that the overall contribution to (3.20) is bounded by $\exp\left(\frac{n}{\log^8 n}\right)$. The bound for the term $T_{\left\lceil \frac{n}{k} \right\rceil}$ will require condition (3.4) from Theorem 3.1 to hold.

Let us remark that if we work with $\mathcal{G}(n,m)$ instead of $\mathcal{G}(n,p)$ and conduct a much more detailed analysis, it is possible to show that if $p < 1 - 1/e$, the contribution from this range of $\mathbf{r}$ is in fact asymptotically equal to $e^{\Delta(1-\Delta)} \leqslant e^{1/4}$, where $\Delta = \gamma - \lfloor \gamma \rfloor$. The bulk of this contribution comes from overlap sequences of the form $\mathbf{r} = (r_2, 0, 0, \ldots, 0)$ with $r_2 = O(\log^2 n)$. However, only the coarser bound is needed for our result.

In Section 4.6.2 of the next chapter, we will perform this more accurate calculation in the special case where all colour classes have exactly the same size $\gamma + o(1)$ (and $\Delta = o(1)$).

Many small overlap blocks

An intermediate degree of overlap is examined in Section 3.5.3, where at least a constant fraction cn of vertices are involved in the overlap between the two partitions, but there are either still many small overlap blocks, or many vertices not involved in the overlap at all. More specifically, for an arbitrary constant $c' > 0$, we will consider all $\mathbf{r}$ with $\rho = v/n > c$ and

$$\sum_{2 \leqslant i \leqslant 0.6\gamma} i r_i \geqslant c'n \text{ or } \rho \leqslant 1 - c',$$

i.e., those $\mathbf{r}$ where there are either at least $c'n$ vertices not in the overlap, or at least $c'n$ vertices in overlap blocks of size at most 0.6γ.

Let us assume for the moment that $\frac{n}{k}$ is an integer in order to simplify notation. It will be useful to define a simple parameter β which measures how close the overlap of two ordered k-equipartitions is to consisting entirely of complete parts of size $\frac{n}{k}$ (with the remaining $n - v$ vertices being singletons not involved in the overlap).

Any overlap block contains at most $\frac{n}{k}$ vertices. Therefore, if we view the overlap blocks of two k-equipartitions as cliques making up a graph, then each vertex has degree at most $\frac{n}{k} - 1$ within this overlap graph. Hence, given the number v of vertices involved in the overlap and the number d of common forbidden edges, we know that

$2d \leqslant (\frac{n}{k}-1)v$, and we let

$$\beta = \frac{2d}{(\frac{n}{k}-1)v} \leqslant 1.$$

If β is close to 1, then the overlap consists almost entirely of very large overlap blocks which are almost entire parts.

In Section 3.5.3.2, we will first consider the case where β is not too close to 1 (so there are enough small overlap blocks). Thereafter, in Section 3.5.3.3, we will study the case where β is close to 1 (so the overlap of the pairs of partitions consists almost exclusively of very large overlap blocks), but there are still many vertices which are not involved in the overlap at all, i.e., $n-v$ is large enough.

In both cases, we will bound the number $P_{\mathbf{r}}$ of pairs of ordered k-equipartitions with overlap sequence $\mathbf{r}$ according to the same strategy. We fix the first ordered k-partition π_1 arbitrarily, and then we generate the second partition in the following way.

We first subdivide the parts of π_1 into overlap parameter blocks and singletons according to $\mathbf{r}$. In the first case, a fairly slack bound on the number of ways to do this will suffice (Lemma 3.13). In the second case, we need to be more careful, and so we will show that this can be done in subexponentially many ways (Lemma 3.15).

Thereafter, we sort the overlap blocks and singletons into k parts in order to form the new partition π_2. In the first case the number of ways to do this is simply bounded by k^{R+n-v}, where R denotes the number of overlap blocks. Bounding R in terms of β in (3.37) and (3.38), we will see that the overall contribution from the first case to the sum (3.20) is $o(1)$.

In the second case we again need a better bound for the number of ways to sort the overlap blocks into the k parts in order to form π_2. Note that in this case, almost the entire overlap consists of very large overlap blocks. If we sort these large overlap blocks into the k parts first, then they occupy their assigned parts almost completely. As there are $v=\rho n$ vertices in the overlap, this means that roughly ρk of the k parts are now filled or almost filled. The remaining smaller overlap blocks and singletons, of which there are roughly $(1-\rho)n$, do not have k parts to pick from. Instead, their choice is limited to about $(1-\rho)k$ parts. Therefore, in this case we get an additional factor of roughly $(1-\rho)^{(1-\rho)n}$.

As almost everything else turns out to be subexponential in the second case, this would be the end of the story if $\frac{n}{k}$ were indeed an integer: as long as $c < \rho < 1 - c'$, the overall contribution to the sum (3.20) would decrease exponentially, and in particular it would be $o(1)$.

Perhaps surprisingly, however, the fact that $\frac{n}{k}$ is not in general an integer is not purely a notational inconvenience. When we do distinguish between parts of size $\lceil \frac{n}{k} \rceil$ and $\lfloor \frac{n}{k} \rfloor$ (or rather, for technical reasons, between parts of size $a = \lceil \gamma \rceil$ and of size at most $a-1$) then there is an additional factor of size about $b^{v_1(1-\Delta)/2}$, where v_1 denotes the number of vertices in the overlap which are in parts of size a within the first partition π_1.

As $v_1 \leqslant v = \rho n$, this means that overall, in equation (3.45), we arrive at an expression which is roughly

$$(1-\rho)^{(1-\rho)n} b^{\rho(1-\Delta)n/2} = b^{n((1-\rho)\log_b(1-\rho)+\rho(1-\Delta)/2)}.$$

Noting in (3.46) that the proportion of vertices in sets of size a in a k-equipartition is roughly $\Delta - x_0 - \varepsilon$, it is now not very hard, but slightly tedious, to compare this last exponent to condition (3.4) in Theorem 3.1 in order to show that this expression is exponentially decreasing in n. We will need to consider several cases, and we will also use the technical Lemmas 3.7 and 3.8 from Section 3.3.2.

This suffices to show that the overall contribution from the second case to the sum (3.20) is $o(1)$.

High overlap

Finally, in Section 3.5.4 we will study those $\mathbf{r}$ where the corresponding pairs of partitions are very similar to each other. In this range, most of the overlap consists of almost entire parts which are merely permuted, with a few exceptional small overlap blocks and singletons.

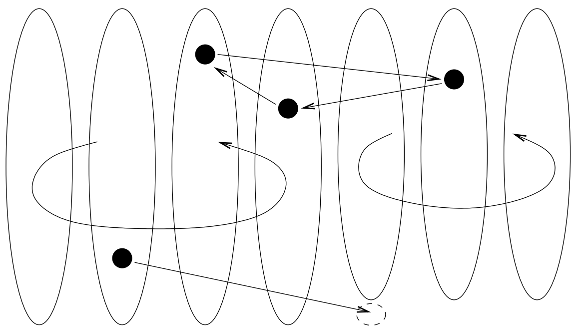

Figure 3.3: In the high overlap case, the second partition is largely generated by permuting the exceptional vertices and then permuting the parts of size $\lceil \frac{n}{k} \rceil$ (shown on the left) and of size $\lfloor \frac{n}{k} \rfloor$ (shown on the right). Exceptional vertices may also jump to smaller parts of size $\lfloor \frac{n}{k} \rfloor$.

We will show that this range of overlap contributes $O(\frac{k_1!k_2!}{\mu_k} \cdot 2^k)$ to the sum (3.20). Since we are sufficiently far above the first moment threshold for the number of colourings, this is $o(1)$, and summing up the contributions from each of the three cases yields (3.20) and thereby concludes the proof of Theorem 3.1.

It is helpful to first consider the *extreme case* of those pairs of partitions π_1, π_2 which are simply permutations of each other: as there are k_1 parts of size $\lceil \frac{n}{k} \rceil$ and k_2 parts of size $\lfloor \frac{n}{k} \rfloor$, there are exactly $Pk_1!k_2!$ such (ordered) pairs of partitions, where P is defined in (3.13) as the total number of k-equipartitions. The number of overlapping edges is maximal, so $d = f$. Therefore, from (3.16), the overall contribution to (3.20) is exactly

$$\frac{Pk_1!k_2!}{P^2} b^f = \frac{k_1!k_2!}{Pq^f} = \frac{k_1!k_2!}{\mu_k}.$$

More generally, we will consider pairs of partitions which are largely just permutations of each other, but where there are also a few *exceptional vertices* which are essentially permuted amongst themselves first, as shown in Figure 3.3. As the part sizes may vary by 1, however, the number of 'available slots' for exceptional vertices in each of the k parts may vary by 1. We will bound these variations with an additional factor 2^k.

From Section 3.5.3, we can assume that there are at most $2c'n$ exceptional vertices, where we can make the constant c' as small as we like. We will distinguish three different types of exceptional vertices. Starting with the first partition π_1, we will first select the exceptional vertices of each type, and bound the number of choices in Lemma 3.17. Then we generate π_2 and bound the number of ways to do this in Lemma 3.18. Finally, we examine how much each exceptional vertex subtracts from the maximum number f of shared forbidden edges between π_1 and π_2 in Lemma 3.19. Summing over the number of exceptional vertices, we will see that the overall contribution to (3.20) is of order $O(\frac{k_1!k_2!}{\mu_k} \cdot 2^k)$ if c' is chosen small enough.

3.5.2 Typical overlap range

We will first consider all those overlap sequences $\mathbf{r}$ where the proportion $\rho = v/n$ of the vertices which are involved in the overlap is at most

$$c = \frac{1 - c_2}{2} \in \left(0, \frac{1}{2}\right), \tag{3.21}$$

where c_2 is the constant from Lemma 3.6. So let

$$\mathcal{R}_1 = \{\mathbf{r} \mid \rho = \rho(\mathbf{r}) \leqslant c\}.$$

The vast majority of all pairs of partitions overlap in a parameter sequence $\mathbf{r} \in \mathcal{R}_1$, and this is also where the bulk of the sum (3.20) comes from. We will show that the contribution of the overlap sequences $\mathbf{r} \in \mathcal{R}_1$ to (3.20) is at most $\exp\left(\frac{n}{\log^8 n}\right)$. To do this, we will find a bound for the contribution from each $\mathbf{r}$ of the form $\prod_i \frac{T_i^{r_i}}{r_i!}$, and then bound the terms T_i.

Any pair of ordered k-equipartitions π_1 and π_2 defines a $k \times k$ *overlap matrix* $\mathcal{M} = (M_{xy})$, where M_{xy} is the number of vertices that are in part x in π_1 and in part y in π_2. Since π_1 and π_2 are ordered k-equipartitions, the first k_1 rows and columns of $\mathcal{M}$ sum to $\left\lceil \frac{n}{k} \right\rceil$ and the remaining k_2 rows and columns sum to $\left\lfloor \frac{n}{k} \right\rfloor$. If π_1 and π_2 overlap according to the overlap sequence $\mathbf{r}$, this means that for every $2 \leqslant i \leqslant a$, exactly r_i of the entries of the overlap matrix are i, exactly $n - v = n - v(\mathbf{r})$ entries are 1, and the remaining entries are 0.

Conversely, given such a matrix, the number of corresponding pairs of ordered k-equipartitions is given by the multinomial coefficient

$$\frac{n!}{\prod_{i=2}^{a} i!^{r_i}}.$$

This is because, given the matrix and n vertices, we must pick r_i sets of i vertices that correspond to the i-entries in the matrix for each i, as well as $n-v$ single vertices for each of the 1-entries, and then this exactly defines the two ordered k-equipartitions. Given $\mathbf{r}$, denote by $M_{\mathbf{r}}$ the number of corresponding matrices, then

$$P_{\mathbf{r}} = \frac{n!}{\prod_{i=2}^{a} i!^{r_i}} \cdot M_{\mathbf{r}}. \tag{3.22}$$

To bound $Q_{\mathbf{r}} = P_{\mathbf{r}}/P^2$, we will count the number $M_{\mathbf{r}}$ of corresponding overlap matrices in the following way. Take an empty $k \times k$-matrix, and write the number 2 in r_2 empty slots, write the number 3 in r_3 empty slots, and so on. There are at most

$$\binom{k^2}{r_2}\binom{k^2}{r_3}\cdots\binom{k^2}{r_a} \leqslant \frac{k^{2\sum_{i=2}^{a} r_i}}{\prod_{i=2}^{a} r_i!} \tag{3.23}$$

ways to do this. The rest of the matrix has entries 0 and 1, and the number of ways to fill in these entries is bounded by the total number of $k \times k$ 0-1 matrices where the row and column sums are given by $\lceil \frac{n}{k} \rceil$ or $\lfloor \frac{n}{k} \rfloor$ minus the values of the entries that are already written in these rows and columns. Note that we are of course overcounting $M_{\mathbf{r}}$, since not all placements of the numbers 2, 3, …, are valid, and not all 0-1 matrices are possible afterwards, but this will be insignificant.

To estimate the number of 0-1 matrices with prescribed row and column sums, we use the following result of McKay ([47], see also [27]).

Theorem 3.9. *Let $N(\mathbf{s}, \mathbf{t})$ be the number of $m \times n$ 0-1 matrices with row sums $\mathbf{s} = (s_1, \ldots, s_m)$ and column sums $\mathbf{t} = (t_1, \ldots, t_n)$. Let $S = \sum_{x=1}^{m} s_x$, $s = \max_x s_x$, $t = \max_y t_y$, $S_2 = \sum_{x=1}^{m} s_x(s_x - 1)$ and $T_2 = \sum_{y=1}^{n} t_y(t_y - 1)$.*

If $S \to \infty$ and $1 \leqslant \max\{s,t\}^2 < cS$ for some constant $c < \frac{1}{6}$, then

$$N(\mathbf{s}, \mathbf{t}) = \frac{S!}{\prod_{x=1}^{m} s_x! \prod_{y=1}^{n} t_y!} \exp\left(-\frac{S_2 T_2}{2S^2} + O\left(\frac{\max\{s,t\}^4}{S}\right)\right).$$

□

Having written the numbers $2, \dots, \lceil \frac{n}{k} \rceil$ in the matrix, the remaining 0-1 entries must be placed so that the rows sum to $\mathbf{s} = (s_1, \dots, s_k)$, and the columns sum to $\mathbf{t} = (t_1, \dots, t_k)$, where $s_x, t_y \leqslant \lceil \frac{n}{k} \rceil$ for all x, y. The exact values for s_x and t_y depend on the placement of the numbers $2, \dots, \lceil \frac{n}{k} \rceil$. In the terminology of Theorem 3.9, we have $S = n - v \geqslant (1-c)n \to \infty$ and $1 \leqslant \max\{s,t\}^2 \leqslant \lceil \frac{n}{k} \rceil^2 = O(\log^2 n) = o(S)$, so we can apply Theorem 3.9.

$$\begin{aligned} N(\mathbf{s},\mathbf{t}) &= \\ &= \frac{(n-v)!}{\prod_{x=1}^k s_x! \prod_{y=1}^k t_y!} \cdot \exp\left(-\frac{\sum_{x=1}^k s_x(s_x-1) \sum_{y=1}^k t_y(t_y-1)}{2(n-v)^2} + O\left(\frac{\log^4 n}{n} \right) \right) \\ &\lesssim \frac{(n-v)!}{\prod_{x=1}^k s_x! \prod_{y=1}^k t_y!}. \end{aligned} \tag{3.24}$$

Now the sequence $s_1, \dots, s_k$ can be obtained from the sequence $\lceil \frac{n}{k} \rceil$, $\lceil \frac{n}{k} \rceil$, $\dots$, $\lfloor \frac{n}{k} \rfloor$ (k_1 times $\lceil \frac{n}{k} \rceil$ and k_2 times $\lfloor \frac{n}{k} \rfloor$) by successively subtracting the number 2 from r_2 members of the sequence, the number 3 from r_3 members of the sequence, and so on.

The product $\prod_{x=1}^k s_x!$ can then be obtained from the product $\lceil \frac{n}{k} \rceil !^{k_1} \lfloor \frac{n}{k} \rfloor !^{k_2}$ by removing the corresponding $v = \sum_{i=2}^a i r_i$ factors of the factorials. If $\sum_{i=2}^a r_i \leqslant k_1$, the product of these factors is maximal if $s_x = \lceil \frac{n}{k} \rceil - i$ for exactly r_i values x for all $i \geqslant 2$. For all remaining values x, $s_x = \lceil \frac{n}{k} \rceil$ or $s_x = \lfloor \frac{n}{k} \rfloor$. Therefore, in this case

$$\prod_{x=2}^k s_x! \geqslant \left\lceil \frac{n}{k} \right\rceil !^{k_1} \left\lfloor \frac{n}{k} \right\rfloor !^{k_2} \cdot \prod_{i=2}^a \frac{\left(\lceil \frac{n}{k} \rceil - i \right)!^{r_i}}{\lceil \frac{n}{k} \rceil !^{r_i}}.$$

Note that the above remains valid if $\sum_{i=2}^a r_i > k_1$ — it is just not tight in this case. If $\lceil \frac{n}{k} \rceil < a = \lceil \gamma \rceil$ (so by our assumptions $\lceil \frac{n}{k} \rceil = a - 1$), there are no parts of size at least a in the partitions, so $r_a = 0$ and the above is still well-defined as there are no terms for $i = a$. The corresponding inequality of course also holds for $\prod_{y=2}^k t_y!$. Therefore,

$$N(\mathbf{s},\mathbf{t}) \lesssim \frac{(n-v)!}{\lceil \frac{n}{k} \rceil !^{2k_1} \lfloor \frac{n}{k} \rfloor !^{2k_2}} \prod_{i=2}^a \frac{\lceil \frac{n}{k} \rceil !^{2r_i}}{\left(\lceil \frac{n}{k} \rceil - i \right)!^{2r_i}}.$$

Using (3.22) and (3.23), this gives

$$P_{\mathbf{r}} \lesssim \frac{(n-v)!n!}{\lceil \frac{n}{k} \rceil !^{2k_1} \lfloor \frac{n}{k} \rfloor !^{2k_2}} \prod_{i=2}^{a} \frac{k^{2r_i} \lceil \frac{n}{k} \rceil !^{2r_i}}{i!^{r_i} r_i! \left(\lceil \frac{n}{k} \rceil - i \right)!^{2r_i}}. \tag{3.25}$$

Note that by Stirling's formula $n! \sim \sqrt{2\pi n} n^n / e^n$, and using $1 + x \leqslant e^x$,

$$\frac{(n-v)!}{n!} \lesssim \frac{(n-v)^{n-v} e^v}{n^n} = n^{-v} \left(1 - \frac{v}{n}\right)^{n-v} e^v \leqslant n^{-v} e^{v^2/n} = n^{-v} e^{\rho v}.$$

Together with (3.19), (3.25) and (3.13), and as $v = \sum_{i=2}^{a} i r_i$, this gives

$$Q_{\mathbf{r}} = \frac{P_{\mathbf{r}}}{P^2} \lesssim \prod_{i=2}^{a} \left(\frac{1}{r_i!} \left(\frac{e^{\rho i} k^2 \lceil \frac{n}{k} \rceil !^2}{n^i i! \left(\lceil \frac{n}{k} \rceil - i \right)!^2} \right)^{r_i} \right). \tag{3.26}$$

Recalling that $d = \sum_{i=2}^{a} \binom{i}{2} r_i$, and that by (B) from Section 3.3.1, $\lceil \frac{n}{k} \rceil \leqslant a$,

$$Q_{\mathbf{r}} b^d \lesssim \prod_{i=2}^{a} \left(\frac{1}{r_i!} \left(\frac{e^{\rho i} b^{\binom{i}{2}} k^2 \lceil \frac{n}{k} \rceil !^2}{n^i i! \left(\lceil \frac{n}{k} \rceil - i \right)!^2} \right)^{r_i} \right) \leqslant \prod_{i=2}^{a} \left(\frac{1}{r_i!} \left(\frac{e^{\rho i} b^{\binom{i}{2}} k^2 a!^2}{n^i i! \, (a-i)!^2} \right)^{r_i} \right)$$

as $\frac{\lceil \frac{n}{k} \rceil !}{\left(\lceil \frac{n}{k} \rceil - i \right)!} \leqslant \frac{a!}{(a-i)!}$ for all i. Therefore, letting

$$T_i := \frac{e^{\rho i} b^{\binom{i}{2}} k^2 a!^2}{n^i i! \, (a-i)!^2},$$

we have

$$Q_{\mathbf{r}} b^d \lesssim \prod_{i=2}^{a} \frac{T_i^{r_i}}{r_i!}.$$

By (B), either $\lceil \frac{n}{k} \rceil = a$ or $\lceil \frac{n}{k} \rceil = a - 1$. In the latter case, there are no parts of size a, so $r_a = 0$. Therefore,

$$Q_{\mathbf{r}} b^d \lesssim \prod_{i=2}^{\lceil \frac{n}{k} \rceil} \frac{T_i^{r_i}}{r_i!}. \tag{3.27}$$

Note that the terms T_i still depend on $\mathbf{r}$, but only through $\rho(\mathbf{r})$. The next lemma ensures that the terms T_i are small enough as long as $\rho \leqslant c$. Let

$$c_5 = \min \left\{ \frac{1}{10}, \frac{c}{2 \log b}, \frac{1-c}{2 \log b} \right\} \in (0, 1),$$

where c is defined in (3.21).

Lemma 3.10. *If* $\mathbf{r} \in \mathcal{R}_1$ *and* n *is large enough, then for all* $3 \leqslant i \leqslant \left\lceil \frac{n}{k} \right\rceil - 1$,

$$T_i \leqslant n^{-c_5},$$

and for $i \in \left\{2, \left\lceil \frac{n}{k} \right\rceil\right\}$,

$$T_i \leqslant n^{1-c_5}.$$

Proof. As usual, we assume throughout that n is large enough for our various bounds to hold. First, note that

$$\frac{T_{i+1}}{T_i} = \frac{e^{\rho} b^{\binom{i+1}{2}} (a-i)!^2}{b^{\binom{i}{2}} n(i+1)\,(a-i-1)!^2} = \frac{e^{\rho} b^i (a-i)^2}{n(i+1)}. \tag{3.28}$$

Now consider $i = 2$: since $a \sim \frac{n}{k} = O(\log n)$ by (D) in Section 3.3.1,

$$T_2 = \frac{e^{2\rho} b^{\binom{2}{2}} k^2 a!^2}{n^2 2!\,(a-2)!^2} \leqslant \frac{e^2 b k^2 a^4}{2n^2} = O\left(\log^2 n\right) \leqslant n^{1-c_5}.$$

By (3.28), for $i \leqslant 5$,

$$T_{i+1} = O\left(\frac{\log^2 n}{n}\right) \cdot T_i \leqslant n^{-1+o(1)} T_i,$$

so in particular for all $3 \leqslant i \leqslant 6$,

$$T_i \leqslant T_3 \leqslant n^{-1+o(1)} O(\log^2 n) \leqslant n^{-c_5}. \tag{3.29}$$

For $7 \leqslant i \leqslant 1.2 \log_b n$, note that as $a \leqslant 2 \log_b n$,

$$T_i \leqslant \frac{e^i b^{\frac{i^2}{2}} n^2 a^{2i}}{n^i} = n^2 \left(\frac{e b^{\frac{i}{2}} a^2}{n}\right)^i \leqslant n^2 \left(\frac{4e b^{0.6 \log_b n} \log_b^2 n}{n}\right)^i \leqslant n^{2-0.3i} \leqslant n^{-0.1} \leqslant n^{-c_5}.$$

Next, we take a look at the special case $i = a$. Since $a = \gamma + 1 - \Delta$, by (F),

$$b^{\binom{a}{2}} = b^{(\gamma-\Delta)(\gamma+1-\Delta)/2} = b^{\frac{\gamma}{2}(\gamma+1-2\Delta)} n^{o(1)} = ((1+o(1))k)^{\gamma+1-2\Delta} n^{o(1)} = k^{a-\Delta} n^{o(1)},$$

so by Stirling's formula,

$$T_a = \frac{e^{\rho a} b^{\binom{a}{2}} k^2 a!}{n^a} \sim \frac{e^{\rho a} k^{a-\Delta+2} n^{o(1)} \sqrt{2\pi a} a^a}{n^a e^a} \leqslant \left(\frac{ka}{n}\right)^a n^{2-\Delta} e^{-(1-\rho)a} n^{o(1)}. \tag{3.30}$$

Since by (D), $a \sim \frac{n}{k} \sim 2 \log_b n$ and as $\mathbf{r} \in \mathcal{R}_1$, this gives

$$T_a \leqslant n^{2-\Delta-(1-\rho)\frac{2}{\log b}+o(1)} \leqslant n^{2-\Delta-(1-c)\frac{2}{\log b}+o(1)}.$$

For $i \in \{a-1, a-2\}$, by (3.28) and (F) and since $a = \lceil \gamma \rceil$,

$$\frac{T_{i+1}}{T_i} \leqslant \frac{n^{o(1)} b^i}{n} = n^{1+o(1)}. \tag{3.31}$$

Therefore,

$$T_{a-1} = T_a n^{-1+o(1)} \leqslant n^{1-\Delta-(1-c)\frac{2}{\log b}+o(1)}.$$

To bound $T_{\lceil \frac{n}{k} \rceil}$, we need to distinguish between two cases. By (B), $\lceil \frac{n}{k} \rceil = a$ or $\lceil \frac{n}{k} \rceil = a - 1$, and γ is not an integer.

- **Case 1:** $\lceil \frac{n}{k} \rceil = a$.

  Since $\frac{n}{k} \leqslant \gamma - x_0 - \varepsilon$, $a = \lceil \gamma \rceil$ and $\Delta = \gamma - \lfloor \gamma \rfloor$, we have $x_0 + \varepsilon \leqslant \Delta$. By Lemma 3.6,

  $$1 - \Delta < \frac{2c_2}{\log b}.$$

  By the definition (3.21) of c, $c_2 = 1 - 2c$ and since $c_5 \leqslant \frac{c}{2 \log b}$,

  $$T_{\lceil \frac{n}{k} \rceil} = T_a \leqslant n^{2-\Delta-(1-c)\frac{2}{\log b}+o(1)} \leqslant n^{1-\frac{2c}{\log b}+o(1)} \leqslant n^{1-2c_5}.$$

- **Case 2:** $\lceil \frac{n}{k} \rceil = a - 1$.

  By the definition of $c_5 \leqslant \frac{1-c}{2 \log b}$,

  $$T_{\lceil \frac{n}{k} \rceil} = T_{a-1} \leqslant n^{1-(1-c)\frac{2}{\log b}+o(1)} \leqslant n^{1-2c_5}.$$

So in both cases we have $T_{\lceil \frac{n}{k} \rceil} \leqslant n^{1-2c_5} \leqslant n^{1-c_5}$. By (3.31), this gives

$$T_{\lfloor \frac{n}{k} \rfloor} \leqslant n^{-1+o(1)} T_{\lceil \frac{n}{k} \rceil} \leqslant n^{-c_5}.$$

Finally, for $i \geqslant 1.2 \log_b n$, by (3.28),

$$\frac{T_{i+1}}{T_i} = \frac{e^{\rho} b^i (a-i)^2}{n(i+1)} \geqslant n^{0.2+o(1)} \geqslant 1,$$

so for all $1.2 \log_b n \leqslant i \leqslant \lfloor \frac{n}{k} \rfloor$,

$$T_i \leqslant T_{\lfloor \frac{n}{k} \rfloor} \leqslant n^{-c_5}.$$

□

Let

$$R = \sum_{i=2}^{a} r_i$$

denote the *total number of overlap blocks*. The following lemma gives a bound for the quantity appearing in (3.27) in terms of R instead of the individual r_i's.

Lemma 3.11. *If n is large enough, then for all $\mathbf{r} \in \mathcal{R}_1$,*

$$Q_{\mathbf{r}} b^d \lesssim n^{-c_5 R/2} \exp\left(\frac{n}{\log^9 n}\right), \tag{3.32}$$

where $c_5 > 0$ is the constant from Lemma 3.10.

Proof. For $3 \leqslant i \leqslant \lfloor \frac{n}{k} \rfloor$, the previous lemma gives

$$\frac{1}{r_i!} T_i^{r_i} \leqslant n^{-c_5 r_i} \leqslant n^{-c_5 r_i/2}.$$

Now suppose $i \in \left\{2, \lceil \frac{n}{k} \rceil\right\}$. If $r_i \leqslant \frac{n}{\log^{11} n}$, then

$$\frac{1}{r_i!} T_i^{r_i} \leqslant n^{(1-c_5)r_i} \leqslant n^{-c_5 r_i/2} \exp\left(\frac{n}{2\log^9 n}\right).$$

Otherwise, if $r_i > \frac{n}{\log^{11} n}$, then since $r_i! \geqslant r_i^{r_i}/e^{r_i}$,

$$\frac{1}{r_i!} T_i^{r_i} \leqslant \left(\frac{e n^{1-c_5}}{r_i}\right)^{r_i} \leqslant \left(e n^{-c_5} \log^{11} n\right)^{r_i} \leqslant n^{-c_5 r_i/2}.$$

Together with (3.27), this gives the result. □

Now we are finally ready to sum (3.32) over all $\mathbf{r} \in \mathcal{R}_1$. For this, note that if n is large enough, then given R, there are at most $(2e \log_b n)^R$ ways to select $r_2, \ldots, r_a$ such that $\sum_{i=2}^{a} r_i = R$. This is because there are

$$\binom{R+a-2}{R} \leqslant \left(\frac{e\,(R+a-2)}{R}\right)^R \leqslant (e\,(1+a-2))^R \leqslant (2e \log_b n)^R \tag{3.33}$$

ways to write R as an ordered sum with $a-1$ nonnegative summands.

Using this and Lemma 3.11, if n is large enough, we can now simply take the sum over R.

$$\begin{aligned}\sum_{\mathbf{r}\in\mathcal{R}_1} Q_{\mathbf{r}} b^d &\lesssim \sum_{R=0}^{\infty} \left((2e \log_b n)^R\, n^{-c_5 R/2} \exp\left(\frac{n}{\log^9 n}\right)\right) \\ &= \exp\left(\frac{n}{\log^9 n}\right) \sum_{R=0}^{\infty} \left(\frac{2e \log_b n}{n^{c_5/2}}\right)^R \leqslant 2\exp\left(\frac{n}{\log^9 n}\right).\end{aligned}$$

Therefore, $\sum_{\mathbf{r}\in\mathcal{R}_1} Q_{\mathbf{r}} b^d \leqslant \exp\left(\frac{n}{\log^8 n}\right)$ for n large enough as required.

3.5.3 Pairs of partitions with many small overlap blocks

In this section, we will bound the contribution to the sum (3.20) from those overlap sequences $\mathbf{r}$ with $\rho = \rho(n) := v/n \geqslant c$, but where there are either still many singletons which are not involved in the overlap (so $n - v$ is large) or many vertices in 'small' overlap blocks of size at most 0.6γ. More specifically, fix a constant $0 < c' < 1$ and consider only those $\mathbf{r}$ with $\rho > c$ such that there are at least $c'n$ singletons or at least $c'n$ vertices in overlap blocks of size at most 0.6γ.

$$\mathcal{R}_2^{c'} = \left\{ \mathbf{r} \mid \rho > c \wedge \left(\sum_{2 \leqslant i \leqslant 0.6\gamma} ir_i \geqslant c'n \vee \rho \leqslant 1 - c' \right) \right\}. \tag{3.34}$$

We will prove that for any fixed $c' \in (0,1)$, the contribution to the sum (3.20) from these overlap sequences is negligible, i.e.,

$$\sum_{\mathbf{r} \in \mathcal{R}_2^{c'}} Q_{\mathbf{r}} b^d = o(1).$$

To do this, we will generate all pairs of partitions in this range by taking the first partition, grouping the vertices into subsets of its parts which will form the singletons and overlap blocks, and rearranging them into k sets to get the new partition. If we do this according to some $\mathbf{r} \in \mathcal{R}_2^{c'}$, then we can bound the number of ways to generate another partition and show that the number of overlapping edges d between the two partitions is small enough.

3.5.3.1 Preliminaries

We first need some notation and preliminary results. Since some of our bounds need to be extremely accurate, we distinguish between parts of size a and parts of size at most $a - 1$. Since by (B), $\lceil \frac{n}{k} \rceil \in \{a, a-1\}$, there may of course not be any parts of size a at all. Fix an *arbitrary ordered k-equipartition* π_1, and let

$$\mathcal{P}_2^{c'} = \left\{ \text{ordered } k\text{-equipartitions } \pi_2 \text{ such that } \mathbf{r}(\pi_1, \pi_2) \in \mathcal{R}_2^{c'} \right\}.$$

Given $\pi_2 \in \mathcal{P}_2^{c'}$, let

$V_1 =$ set of vertices in the overlap of π_1 and π_2 that are in parts of size a in π_1

$V_2 =$ set of vertices in the overlap of π_1 and π_2 that are in parts of size at most $a-1$ in π_1

$D_1 =$ set of overlapping forbidden edges between vertices in V_1

$D_2 =$ set of overlapping forbidden edges between vertices in V_2.

For $i \in \{1, 2\}$, let $v_i = |V_i|$ and $d_i = |D_i|$, so $v_1 + v_2 = v$ and $d_1 + d_2 = d$.

Given π_1 and π_2, we define the *overlap graph* of π_1 and π_2 as the union of all the vertices in overlap blocks together with all the joint forbidden edges. By definition, the overlap graph is a disjoint union of cliques, each containing between 2 and $\lceil \frac{n}{k} \rceil$ vertices. Note that the vertex set is exactly $V_1 \cup V_2$ and the edge set exactly $D_1 \cup D_2$. Denote by $\mathbf{g} = (g_j)_{j=1}^{v_1}$ the degree sequence in the overlap graph of the vertices in V_1. Then $g_j \leqslant a - 1$ for all j, so

$$2d_1 = \sum_{j=1}^{v_1} g_j \leqslant v_1 (a-1).$$

Similarly,

$$2d_2 \leqslant v_2 (a-2).$$

Let

$$\beta_1 = \frac{2d_1}{v_1 (a-1)} \leqslant 1$$
$$\beta_2 = \frac{2d_2}{v_2 (a-2)} \leqslant 1.$$

For $x \in (0, 1)$, denote by $w_{x,1}$ the proportion of vertices in V_1 which have degree at most $x(a-1)$ within the overlap graph, i.e.,

$$w_{x,1} = \frac{\#\, j \text{ with } g_j \leqslant x(a-1)}{v_1}.$$

Then, as $g_j \leqslant a-1$ for all j, for any $x \in (0, 1)$,

$$\beta_1 v_1 (a-1) = 2d_1 = \sum_{j=1}^{v_1} g_j \leqslant w_{x,1} v_1 x (a-1) + (1 - w_{x,1}) v_1 (a-1),$$

so

$$w_{x,1} \leqslant \frac{1-\beta_1}{1-x}. \tag{3.35}$$

Similarly, define $w_{x,2}$ as the proportion of vertices in V_2 that have degree in the overlap graph of at most $x\,(a-2)$. Analogously, we have

$$w_{x,2} \leqslant \frac{1-\beta_2}{1-x}.$$

Next, we need a bound for the total number of overlap blocks. As in the previous section, let

$$R = \sum_{i=2}^{a} r_i \tag{3.36}$$

denote the number of overlap blocks. Let R_1 and R_2 denote the number of overlap blocks in parts of size a and of size at most $a-1$ in π_1, respectively, so $R = R_1 + R_2$.

Note that

$$R_1 = \sum_{j=1}^{v_1} \frac{1}{g_j+1},$$

as every overlap block of s vertices contributes exactly s instances of the summand $\frac{1}{s}$. For any $0 < x < y < 1$, there are $w_{x,1}v_1$ values i such that $1 \leqslant g_i \leqslant x\,(a-1)$, at most $w_{y,1}v_1$ values i such that $x\,(a-1) < g_i \leqslant y\,(a-1)$, and for the remaining values i, $g_i > y\,(a-1)$. Therefore,

$$R_1 \leqslant \frac{w_{x,1}v_1}{2} + \frac{w_{y,1}v_1}{x\,(a-1)+1} + \frac{v_1}{y\,(a-1)+1} \leqslant \frac{w_{x,1}v_1}{2} + \frac{w_{y,1}v_1}{x\gamma} + \frac{v_1}{y\gamma}.$$

Using (3.35), it follows that

$$R_1 \leqslant \frac{(1-\beta_1)v_1}{2(1-x)} + \frac{(1-\beta_1)v_1}{x(1-y)\gamma-1} + \frac{v_1}{y\gamma-1}, \tag{3.37}$$

where in the last term $y\gamma$ was replaced by $y\gamma-1$ so that the corresponding expression holds for R_2 as well. Indeed, we can see that

$$R_2 \leqslant \frac{(1-\beta_2)v_2}{2(1-x)} + \frac{(1-\beta_2)v_2}{x(1-y)\gamma-1} + \frac{v_2}{y\gamma-1}. \tag{3.38}$$

We will now give some weaker conditions for π_2 which are more convenient to work with, and show that any $\pi_2 \in \mathcal{P}_2^{c'}$ meets these conditions.

Lemma 3.12. *If $\pi_2 \in \mathcal{P}_2^{c'}$ and n is large enough, then at least one of the following three conditions applies:*

I) $v_1 \geqslant \frac{n}{(\log\log n)^2}$ and $\beta_1 \leqslant 1 - \frac{(\log\log n)^4}{\log n}$.

II) $v_2 \geqslant \frac{n}{(\log\log n)^2}$ and $\beta_2 \leqslant 1 - \frac{(\log\log n)^4}{\log n}$.

III) Neither I nor II holds, and $c < \rho \leqslant 1 - c'$.

Proof. By the definition of $\mathcal{R}_2^{c'}$, it suffices to show that if $\sum_{2\leqslant i\leqslant 0.6\gamma} ir_i \geqslant c'n$, then I or II holds. So suppose that $\sum_{2\leqslant i\leqslant 0.6\gamma} ir_i \geqslant c'n$. Of those vertices that are in overlap blocks of size at most 0.6γ, either at least $c'n/2$ are in parts of size a or at least $c'n/2$ are in parts of size $a-1$ in π_1.

So say that at least $c'n/2$ of them are in parts of size a. In particular, $v_1 \geqslant c'n/2 \geqslant \frac{n}{(\log\log n)^2}$. Furthermore, if we denote by $\hat{r}_i$ the number of overlap blocks of size i in parts of size a in π_1, then

$$d_1 = \sum_{i=2}^{a} \binom{i}{2} \hat{r}_i \leqslant 0.3\gamma \sum_{2\leqslant i\leqslant 0.6\gamma} i\hat{r}_i + \frac{a-1}{2} \sum_{0.6\gamma < i\leqslant a} i\hat{r}_i.$$

Since $0.3\gamma \leqslant \frac{a-1}{2}$ and $\sum_{2\leqslant i\leqslant 0.6\gamma} i\hat{r}_i \geqslant c'n/2$ and $\gamma \leqslant \lceil\gamma\rceil = a$, this is at most

$$\frac{0.3\gamma c'n}{2} + \frac{a-1}{2}\left(v_1 - \frac{c'n}{2}\right) \leqslant \frac{a-1}{2}v_1 - 0.05ac'n.$$

Therefore,

$$\beta_1 = \frac{2d_1}{v_1\,(a-1)} \leqslant 1 - (0.1 + o(1))\frac{c'n}{v_1}.$$

As $v_1 \leqslant n$, this is at most

$$1 - 0.05c' < 1 - \frac{(\log\log n)^4}{\log n},$$

so I holds if n is large enough.

The second case is analogous and implies II. □

Still fixing the arbitrary ordered k-equipartition π_1, let

$$\mathcal{P}^{\mathrm{I}} = \Big\{\text{ordered } k\text{-equipartitions } \pi_2 \text{ such that } v_1 \geqslant \frac{n}{(\log\log n)^2} \text{ and } \beta_1 \leqslant 1 - \frac{(\log\log n)^4}{\log n}\Big\}$$

$$\mathcal{P}^{\mathrm{II}} = \Big\{\text{ordered } k\text{-equipartitions } \pi_2 \text{ such that } v_2 \geqslant \frac{n}{(\log\log n)^2} \text{ and } \beta_2 \leqslant 1 - \frac{(\log\log n)^4}{\log n}\Big\}$$

$$\mathcal{P}^{\mathrm{III}} = \Big\{\text{ordered } k\text{-equipartitions } \pi_2 \text{ such that } c < \rho \leqslant 1 - c'\Big\} \setminus \mathcal{P}^{\mathrm{I}} \setminus \mathcal{P}^{\mathrm{II}}, \quad (3.39)$$

where v_i, β_i and ρ refer to the overlap of π_1 and π_2. Then by Lemma 3.12 for n large enough,

$$\mathcal{P}_2^{c'} \subset \mathcal{P}^{\mathrm{I}} \cup \mathcal{P}^{\mathrm{II}} \cup \mathcal{P}^{\mathrm{III}}.$$

For an overlap sequence $\mathbf{r}$, denote by $P'_{\mathbf{r}}$ the number of ordered k-equipartitions with overlap $\mathbf{r}$ with π_1. Then by the definition (3.19) of $Q_{\mathbf{r}}$,

$$Q_{\mathbf{r}} = \frac{P_{\mathbf{r}}}{P^2} = \frac{P'_{\mathbf{r}}}{P}. \quad (3.40)$$

Using (3.14) in the last step, if n is large enough,

$$\sum_{\mathbf{r}\in\mathcal{R}_2^{c'}} Q_{\mathbf{r}} b^d = \sum_{\mathbf{r}\in\mathcal{R}_2^{c'}} \frac{P'_{\mathbf{r}}}{P} b^d = \sum_{\pi_2\in\mathcal{P}_2^{c'}} P^{-1} b^{d(\pi_1,\pi_2)} \leqslant \sum_{\pi_2\in\mathcal{P}^{\mathrm{I}}\cup\mathcal{P}^{\mathrm{II}}\cup\mathcal{P}^{\mathrm{III}}} P^{-1} b^{d(\pi_1,\pi_2)}$$
$$= \sum_{\pi_2\in\mathcal{P}^{\mathrm{I}}\cup\mathcal{P}^{\mathrm{II}}\cup\mathcal{P}^{\mathrm{III}}} k^{-n} b^{d(\pi_1,\pi_2)} \exp(o(n)), \quad (3.41)$$

where $d(\pi_1, \pi_2) := d(\mathbf{r})$ if $\mathbf{r}$ is the overlap sequence of π_1 and π_2.

We will now generate and count all $\pi_2 \in \mathcal{P}^{\mathrm{I}} \cup \mathcal{P}^{\mathrm{II}} \cup \mathcal{P}^{\mathrm{III}}$. Starting with π_1, we first subdivide the parts into overlap blocks and singletons. Then we arrange those overlap blocks and singletons into k new parts to generate π_2, and sum the resulting $b^{d(\pi_1,\pi_2)}$.

3.5.3.2 Contribution from Cases I and II

We start by generating the partitions in $\mathcal{P}^{\mathrm{I}} \cup \mathcal{P}^{\mathrm{II}}$. The strategy is as follows. We group the vertices into subsets of the parts of π_1 which form the overlap blocks and singletons for the overlap with π_2, and give a bound for the number of ways this can be done in Lemma 3.13. Then we sort the overlap blocks and singletons into the k parts of π_2. If there are R overlap blocks and $n-v$ singletons, then there are at most k^{n-v+R} choices for this. Considering (3.41), the term k^n cancels out with k^{-n}, leaving just $k^{-v+R}b^d$ multiplied by the bound from Lemma 3.13 as an upper bound for (3.41). If Cases I or II apply and we also use the bounds (3.37) and (3.38) for R, then d_i will be small enough in comparison to v_i for at least one $i \in \{1,2\}$ so that k^{-v+R} is much smaller than b^d, allowing us to bound the total contribution from Cases I and II to (3.41) and thereby to (3.20) by $o(1)$.

Lemma 3.13. *Denote by S the number of ways the n vertices can be grouped into subsets (of any size and number) of the parts of π_1. Then*

$$S \leqslant \exp\left(O\left(n \log\log n\right)\right).$$

Proof. If we sort the n vertices into a containers, this defines a subdivision of π_1 by letting all vertices be in the same set that are in the same part of π_1 and in the same container. Conversely, any possible subdivision of π_1 can be obtained in this way, since every part can only be partitioned into at most a non-empty sets. Therefore, as $a = O(\log n)$,

$$S \leqslant a^n = \exp\left(O\left(n \log\log n\right)\right).$$

□

We are now ready to show that the contribution to (3.41) from all ordered k-equipartitions in $\mathcal{P}^{\mathrm{I}} \cup \mathcal{P}^{\mathrm{II}}$ to (3.41) is $o(1)$.

Lemma 3.14.

$$\sum_{\pi_2 \in \mathcal{P}^{I} \cup \mathcal{P}^{II}} k^{-n} b^{d(\pi_1,\pi_2)} \exp(o(n)) = o(1).$$

Proof. Fix v_1, v_2, d_1 and d_2 so that I or II holds. Let

$$\mathcal{P}(v_1,v_2,d_1,d_2) = \left\{\pi_2 \in \mathcal{P}^{\mathrm{I}} \cup \mathcal{P}^{\mathrm{II}} \mid v_i(\pi_1,\pi_2) = v_i, d_i(\pi_1,\pi_2) = d_i, i = 1,2\right\}.$$

Arrange the n vertices into singletons and overlap blocks that are subsets of the parts of π_1 in accordance with v_1, v_2, d_1 and d_2. Now that we know the R overlap blocks and $n-v$ singletons, the number of ordered k-equipartitions π_2 with these overlap blocks with π_1 is at most k^{n-v+R}, since we need to sort $n-v$ singletons and R overlap blocks into k parts.

Therefore, if we let $x=\frac{1}{4}$ and $y=1-\frac{1}{\log\log n}$, then with (3.37), (3.38) and Lemma 3.13,

$$\sum_{\pi_2\in\mathcal{P}(r_1,r_2,d_1,d_2)} k^{-n}b^{d(\pi_1,\pi_2)} \leqslant Sk^{-n+n-v+\sum_{i=1}^{2}\left(\frac{2(1-\beta_i)v_i}{3}+\frac{4(1-\beta_i)v_i}{(1-y)\gamma-4}+\frac{v_i}{y\gamma-1}\right)}b^{d_1+d_2}\exp\left(o(n)\right)$$
$$\leqslant k^{\sum_{i=1}^{2}\left(-v_i+\frac{2(1-\beta_i)v_i}{3}+\frac{4(1-\beta_i)v_i}{(1-y)\gamma-4}+\frac{v_i}{y\gamma-1}\right)}b^{d_1+d_2}\exp\left(O(n\log\log n)\right). \quad (3.42)$$

Note that as by (F) from Section 3.3.1, $b^{\frac{\gamma}{2}}\sim k$, and since $\beta_i v_i\leqslant n$ for $i\in\{1,2\}$ and $a=\lceil\gamma\rceil\leqslant\gamma+1$,

$$b^{d_1+d_2}=b^{\beta_1 v_1\cdot\frac{a-1}{2}+\beta_2 v_2\cdot\frac{a-2}{2}}\leqslant b^{(\beta_1 v_1+\beta_2 v_2)\cdot\frac{\gamma}{2}}\leqslant k^{\beta_1 v_1+\beta_2 v_2}\exp(o(n)).$$

By (G) and since $v_i\leqslant n$, $k^{\frac{v_i}{y\gamma-1}}\leqslant\exp(O(n))$ for $i\in\{1,2\}$, and therefore (3.42) becomes

$$\sum_{\pi_2\in\mathcal{P}(r_1,r_2,d_1,d_2)} k^{-n}b^{d(\pi_1,\pi_2)}\leqslant k^{-\sum_{i=1}^{2}\left(v_i(1-\beta_i)\left(\frac{1}{3}-\frac{4}{(1-y)\gamma-4}\right)\right)}\exp\left(O(n\log\log n)\right).$$

Recall that $y=1-\frac{1}{\log\log n}$, so $(1-y)\gamma\to\infty$, and we have $\frac{1}{3}-\frac{4}{(1-y)\gamma-4}\geqslant\frac{1}{4}$ for n large enough. Since I or II holds, there is an $i\in\{1,2\}$ such that $(1-\beta_i)v_i\geqslant\frac{n(\log\log n)^2}{\log n}$, so by (G),

$$\sum_{\pi_2\in\mathcal{P}(r_1,r_2,d_1,d_2)} k^{-n}b^{d(\pi_1,\pi_2)}\leqslant k^{-\frac{n(\log\log n)^2}{4\log n}}\exp\left(O(n\log\log n)\right).$$

As $f=O(n\log n)$ by (H), and since $v_i\leqslant n$ and $d_i\leqslant f$ for $i\in\{1,2\}$, there are only at most $O(n^4\log^2 n)$ choices for the values of $v_i\leqslant n$ and d_i for $i\in\{1,2\}$. Hence,

$$\sum_{\pi_2\in\mathcal{P}^{\mathrm{I}}\cup\mathcal{P}^{\mathrm{II}}} k^{-n}b^{d(\pi_1,\pi_2)}\exp(o(n))\leqslant k^{-\frac{n(\log\log n)^2}{4\log n}}\exp\left(O(n\log\log n)\right)$$
$$=\exp\left(-\Theta\left(n(\log\log n)^2\right)\right)=o(1).$$

□

3.5.3.3 Contribution from Case III

We have to be a bit more careful in the case where neither I nor II holds. We will proceed similarly as in the proof of Lemma 3.14: we subdivide π_1 into subsets and then sort the singletons and overlap blocks into the k parts to form the new partition π_2. Since for both $i \in \{1, 2\}$, β_i is either close to 1 or v_i is negligibly small, most of the overlap blocks will be almost entire parts of π_1. If we place those large overlap blocks first, they occupy a constant fraction of about ρk of the k parts almost entirely, so the remaining roughly $(1-\rho)n$ vertices and smaller overlap blocks have fewer choices left, namely only about $(1-\rho)k$ choices each. This will give an additional factor of about $(1-\rho)^{(1-\rho)n}$. Almost everything else will turn out to be subexponential, except for a term which is about $b^{(1-\Delta)v/2}$. As $v = \rho n$, this will result in a total bound which is roughly of the form $b^{-((1-\rho)\log_b(1-\rho)-(1-\Delta)\rho/2)n}$. Comparing the exponent of this expression with condition (3.4) from Theorem 3.1 (using the technical lemmas we proved in Section 3.3.2), we will show that the sum is $o(1)$ for $c < \rho < 1 - c'$.

Instead of Lemma 3.13, which gave a fairly slack bound on the number of ways the vertices may be arranged into subsets of the parts of π_1, we now need a more accurate bound. The following lemma ensures that if Condition III applies, the number of ways to subdivide π_1 is subexponential.

Lemma 3.15. *Fix integers v_1, v_2, d_1, d_2 so that I and II do not hold as above. Denote by $S(v_1, v_2, d_1, d_2)$ the number of ways the vertices can be arranged into subsets of the parts of π_1 which form overlap blocks and singletons according to v_i and d_i, $i \in \{1, 2\}$. Then there is a function $S' = S'(n)$ which does not depend on v_i or d_i, $i = 1, 2$, such that*

$$S(v_1, v_2, d_1, d_2) \leqslant S' \leqslant \exp\left(o(n)\right).$$

Proof. We first split up the parts of size a. Since I does not hold, either $v_1 < \frac{n}{(\log\log n)^2}$ or $\beta_1 > 1 - \frac{(\log\log n)^4}{\log n}$.

In the first case, select the $v_1 < \frac{n}{(\log\log n)^2} = o(n)$ vertices which form the overlap blocks in parts of size a. Using (J) from Section 3.3.1, there are at most

$$\binom{n}{v_1} \leqslant \binom{n}{\left\lfloor \frac{n}{(\log\log n)^2} \right\rfloor} \leqslant \exp(o(n))$$

ways to do this. All the other vertices in parts of size a must be singletons. To find out how the v_1 vertices are arranged into overlap blocks, we can proceed as in the proof of Lemma 3.13: sort the v_1 vertices into a containers, and let those vertices be in the same overlap block that are in the same container and in the same part of π_1. There are

$$a^{v_1} \leqslant a^{\frac{n}{(\log\log n)^2}} = \exp\left(O\left(\frac{n}{\log\log n}\right)\right) \leqslant \exp(o(n))$$

possibilities for this, so altogether there are $\exp(o(n))$ ways to split up the parts of size a in the case $v_1 < \frac{n}{(\log\log n)^2}$.

In the second case, we have $\beta_1 > 1 - \frac{(\log\log n)^4}{\log n}$. Let $x = 1 - \frac{(\log\log n)^2}{(\log n)^{1/2}}$, then by (3.35), if π_2 overlaps with π_1 according to v_i and d_i, then

$$w_{x,1} \leqslant \frac{\frac{(\log\log n)^4}{\log n}}{\frac{(\log\log n)^2}{(\log n)^{1/2}}} = \frac{(\log\log n)^2}{(\log n)^{1/2}} =: \hat{w}_x \to 0.$$

This means that almost all of the v_1 vertices in the overlap must be arranged into large overlap blocks of size greater than $x\,(a-1)+1$. As $x \to 1$, we can assume $x > 1/2$. Therefore, any part of π_1 contains at most one such large overlap block, and we can group the vertices in parts of size a into overlap blocks and singletons in the following way.

- First we select the parts which contain large overlap blocks. There are at most

  $$2^k = \exp\left(O\left(n/\log n\right)\right) = \exp(o(n))$$

  choices.

- Next, given these $k' \leqslant k$ parts, we pick the vertices within the parts that are not in the large overlap blocks. Since $x \to 1$, there are at most

  $$k'(a - x(a-1) - 1) \leqslant (1-x)ak' = o\left(ak'\right) = o(n)$$

such vertices. Therefore, there are at most

$$\sum_{l \leqslant (1-x)ak'} \binom{ak'}{l} \leqslant ((1-x)ak'+1) \cdot \binom{ak'}{\lfloor (1-x)ak' \rfloor} \leqslant n \cdot \binom{n}{(1-x)n}$$
$$\leqslant \exp(o(n))$$

possibilities for this.

- Now we know all the large overlap blocks in V_1. From the remaining vertices, we choose those vertices that are not singletons, i.e., which are in overlap blocks of size at least 2, but not in big overlap blocks. There cannot be more than $\hat{w}_x v_1 \leqslant \hat{w}_x n = o(n)$ such vertices. Therefore, there are at most

$$\sum_{j \leqslant \hat{w}_x n} \binom{n}{j} \leqslant (\hat{w}_x n + 1) \binom{n}{\lfloor w_x n \rfloor} \leqslant \exp(o(n))$$

  choices.

- We have determined all of the large overlap blocks and which of the remaining vertices are singletons and which are in overlap blocks. It only remains to group the vertices that are in overlap blocks into subsets of the parts of π_1. As in the proof of Lemma 3.13, each such partition into subsets can be obtained by sorting the vertices into a containers, and since there are at most $\hat{w}_x v_1 \leqslant \hat{w}_x n$ vertices left, this can be done in at most

$$a^{\hat{w}_x n} = \exp\left[O\left(n \hat{w}_x \log \log n\right)\right] = \exp(o(n))$$

  ways.

Multiplying everything, and noting that none of the bounds depend on the specific choice of v_i and d_i, gives the bound $\exp(o(n))$ for the number of ways we can subdivide the parts of size a in the second case, and hence in both cases.

The bound $\exp(o(n))$ for subdividing the parts of size at most $a-1$ can be proved analogously. Multiplying those two bounds gives $S' = S'(n)$ such that

$$S(v_1, v_2, d_1, d_2) \leqslant S' \leqslant \exp(o(n)).$$

□

Now we are ready to show that the contribution to (3.41) from Case III is $o(1)$.

Lemma 3.16.

$$\sum_{\pi_2 \in \mathcal{P}^{III}} k^{-n} b^{d(\pi_1,\pi_2)} \exp(o(n)) = o(1).$$

Proof. Suppose we have fixed v_1, v_2, d_1 and d_2 in such a way that I and II do not hold but III does. Let

$$\mathcal{P}'(v_1, v_2, d_1, d_2) = \left\{ \pi_2 \in \mathcal{P}^{\mathrm{III}} \mid v_i(\pi_1, \pi_2) = v_i, d_i(\pi_1, \pi_2) = d_i, i = 1, 2 \right\}.$$

Let $u = 1 - \frac{(\log \log n)^5}{\log n} \to 1$. Recall that $\rho = v/n = (v_1 + v_2)/n$.

Claim. *For any $\pi_2 \in \mathcal{P}'(v_1, v_2, d_1, d_2)$, there are $(1 + o(1))\rho k$ 'large' overlap blocks of size at least $u(a - 2)$ in the overlap of π_1 and π_2.*

Proof. Of course there are asymptotically at most $\frac{v}{u(a-2)} \sim \rho k$ such blocks, so we only need to show that there are asymptotically at least ρk of them.

Note that if $v_i \geqslant \frac{n}{(\log \log n)^2}$ for $i \in \{1, 2\}$, then as I and II do not hold, $\beta_i > 1 - \frac{(\log \log n)^4}{\log n}$, and therefore,

$$\frac{1 - \beta_i}{1 - u} \leqslant \frac{1}{\log \log n} \to 0. \tag{3.43}$$

If $\pi_2 \in \mathcal{P}'(v_1, v_2, d_1, d_2)$, then by (3.35), there are at least

$$(1 - w_{u,1})v_1 + (1 - w_{u,2})v_2 \geqslant \sum_{i=1}^{2} \left(1 - \frac{1 - \beta_i}{1 - u}\right) v_i$$

vertices in large overlap blocks of size at least $u(a - 2)$. Since no overlap block contains more than a vertices, there are at least

$$\sum_{i=1}^{2} \left(1 - \frac{1 - \beta_i}{1 - u}\right) \frac{v_i}{a} \tag{3.44}$$

such large overlap blocks. As III holds, $v_1 + v_2 = v \geqslant cn$, so there can be at most one $i \in \{1, 2\}$ with $v_i < \frac{n}{(\log \log n)^2}$. If this is the case and j is the other element of $\{1, 2\}$, then $v_i \ll v_j \sim v \sim \rho n$, so together with (3.43), (3.44) is

$$o\left(\frac{n}{a}\right) + \left(1 - \frac{1 - \beta_j}{1 - u}\right) \frac{v_j}{a} = o(k) + (1 + o(1))\rho \frac{n}{a} \sim \rho k,$$

as $\frac{n}{a} \sim k$ by (E). Otherwise, if for both $i \in \{1,2\}$, $v_i \geqslant \frac{n}{(\log\log n)^2}$, (3.44) and (3.43) give

$$\sum_{i=1}^{2} \left(1 - \frac{1-\beta_i}{1-u}\right) \frac{v_i}{a} \geqslant \left(1 - \frac{1}{\log\log n}\right) \frac{v_1+v_2}{a} \sim \frac{v_1+v_2}{a} \sim \rho k.$$

So in both cases, there are asymptotically at least ρk large overlap blocks of size at least $u\,(a-2)$. ☐

Having subdivided the partition π_1 into overlap blocks and singletons according to v_1, v_2, d_1, d_2, we now generate all $\pi_2 \in \mathcal{P}'(v_1, v_2, d_1, d_2)$. Recall that R was defined in (3.36) as the total number of overlap blocks.

Claim. *There are at most*

$$(1-\rho)^{(1-\rho)n} k^{n-v+R} \exp(o(n))$$

other ordered k-equipartitions with the given overlap blocks with the original partition π_1.

Proof. We sort the overlap blocks and singletons into k parts to create a new ordered k-equipartition π_2, and start with the large sets of size at least $u\,(a-2)$. By the previous claim, there are $(1+o(1))\rho k$ of them, and each has at most k choices. As $u \to 1$, we can assume $u > 0.6$, so no two large overlap blocks can be assigned to the same part.

After we are finished with the large overlap blocks, the remaining vertices can either be sorted into the small remainder of the $(1+o(1))\rho k$ parts of π_2 which have been assigned a large block, or they can be sorted into the remaining $(1-\rho+o(1))k$ parts of π_2.

As $u \to 1$, we can fit at most $(1+o(1))\rho k \cdot (a - u\,(a-2)) = o(n)$ vertices into the remainder of the parts of π_2 with large overlap blocks. Therefore, by (J) there are at most

$$\binom{n}{o(n)} \leqslant \exp(o(n))$$

ways of picking these vertices, and for each there are at most k choices for which part of π_2 it is assigned to.

There are now at least $n - v - o(n) = (1 - \rho + o(1))n$ singletons and overlap blocks left to be assigned to the remaining $(1 - \rho + o(1))k$ parts. For each of these there are at most $(1 - \rho + o(1))k$ choices.

We have now sorted R overlap blocks and $n - v$ singletons into the k parts, and bounded the number of choices for each by at most k, and for $(1 - \rho + o(1))n$ of them by $(1 - \rho + o(1))k$. Therefore, in total there are at most

$$(1 - \rho + o(1))^{(1-\rho+o(1))n} k^{n-v+R} \leqslant (1 - \rho)^{(1-\rho)n} k^{n-v+R} \exp(o(n))$$

ways to build a new partition π_2 from the given overlap blocks and singletons. □

Now as in the previous part, let $x = \frac{1}{4}$ and $y = 1 - \frac{1}{\log \log n}$. Then as in (3.42), by Lemma 3.15, (3.37) and (3.38), and since $R = R_1 + R_2$,

$$\begin{aligned}
&\sum_{\pi_2 \in \mathcal{P}'(v_1,v_2,d_1,d_2)} k^{-n} b^{d(\pi_1,\pi_2)} \\
&\leqslant S(v_1, v_2, d_1, d_2)(1 - \rho)^{(1-\rho)n} k^{-n+n-v+\sum_{i=1}^{2}\left(\frac{2(1-\beta_i)v_i}{3} + \frac{4(1-\beta_i)v_i}{(1-y)\gamma-4} + \frac{v_i}{y\gamma-1}\right)} b^{d_1+d_2} \exp(o(n)) \\
&\leqslant (1 - \rho)^{(1-\rho)n} k^{\sum_{i=1}^{2}\left(-v_i + \frac{2(1-\beta_i)v_i}{3} + \frac{4(1-\beta_i)v_i}{(1-y)\gamma-4} + \frac{v_i}{y\gamma-1}\right)} \cdot b^{d_1+d_2} \exp\left(o(n)\right).
\end{aligned}$$

Note that as by (F), $b^{\frac{\gamma}{2}} \sim k$, and as $a = \lceil \gamma \rceil = \gamma + 1 - \Delta$,

$$\begin{aligned}
b^{d_1+d_2} &= b^{\beta_1 v_1 \cdot \frac{a-1}{2} + \beta_2 v_2 \cdot \frac{a-2}{2}} = b^{(\beta_1 v_1 + \beta_2 v_2)\cdot \frac{\gamma}{2} - \frac{1}{2}(\Delta \beta_1 v_1 + (1+\Delta)\beta_2 v_2)} \\
&\leqslant k^{\beta_1 v_1 + \beta_2 v_2} b^{-\frac{1}{2}(\Delta \beta_1 v_1 + (1+\Delta)\beta_2 v_2)} \exp(o(n)).
\end{aligned}$$

Since I and II do not hold, $v_i(1 - \beta_i) = o(n)$ for $i = 1, 2$, and therefore,

$$b^{d_1+d_2} \leqslant k^{\beta_1 v_1 + \beta_2 v_2} b^{-\frac{1}{2}(\Delta v_1 + (1+\Delta)v_2)} \exp(o(n)).$$

Hence,

$$\begin{aligned}
\sum_{\pi_2 \in \mathcal{P}'(v_1,v_2,d_1,d_2)} k^{-n} b^{d(\pi_1,\pi_2)} &\leqslant (1 - \rho)^{(1-\rho)n} k^{-\sum_{i=1}^{2}\left(v_i(1-\beta_i)\left(\frac{1}{3} - \frac{4}{(1-y)\gamma-4}\right)\right)} k^{\frac{v_1+v_2}{y\gamma-1}} \\
&\quad \cdot b^{-\frac{1}{2}(\Delta v_1 + (1+\Delta)v_2)} \exp\left(o(n)\right) \\
&\leqslant (1 - \rho)^{(1-\rho)n} k^{\frac{v_1+v_2}{y\gamma-1}} b^{-\frac{1}{2}(\Delta v_1 + (1+\Delta)v_2)} \exp\left(o(n)\right)
\end{aligned}$$

as $\frac{1}{3} - \frac{4}{(1-y)\gamma-4} > 0$ because $(1-y)\gamma \to \infty$. Since $\gamma \sim 2\log_b n$ and $y \to 1$,

$$k^{\frac{v_1+v_2}{y\gamma-1}} b^{-\frac{1}{2}(\Delta v_1+(1+\Delta)v_2)} \leqslant n^{\frac{v_1+v_2}{y\gamma-1}} b^{-\frac{1}{2}(\Delta v_1+(1+\Delta)v_2)} = b^{\frac{v_1+v_2}{2+o(1)}-\frac{1}{2}(\Delta v_1+(1+\Delta)v_2)}$$
$$\leqslant b^{\frac{1-\Delta}{2}\cdot v_1 - \frac{\Delta}{2}\cdot v_2} \exp(o(n)).$$

Hence,

$$\sum_{\pi_2 \in \mathcal{P}'(r_1,r_2,d_1,d_2)} k^{-n} b^{d(\pi_1,\pi_2)} \leqslant b^{n(1-\rho)\log_b(1-\rho)+\frac{v_1}{2}-\frac{\Delta v}{2}} \exp(o(n)). \tag{3.45}$$

We will show that this last expression is exponentially decreasing in n, and need to distinguish three cases, depending how large $\Delta - x_0 - \varepsilon$ is in comparison to ρ. Note that since γ is not an integer by (B),

$$\frac{n}{k} = \gamma - x_0 - \varepsilon + o(1) = \lfloor \gamma \rfloor + \Delta - x_0 - \varepsilon + o(1) = a - 1 + \Delta - x_0 - \varepsilon + o(1). \tag{3.46}$$

Roughly speaking, $\Delta - x_0 - \varepsilon$ is the proportion of parts of size a in a k-equipartition, and we need to distinguish between Case 1 where there are few (or no) such parts, Case 2 where there are more such parts but still not so many that all of the $v = \rho n$ vertices in the overlap can be in parts of size a, and finally Case 3 where there are enough parts of size a that the overlap blocks between π_1 and π_2 can all be in parts of size a in π_1. In the first case, we shall only need the condition that $c < \rho < 1 - c'$; the second and third cases are where condition (3.4) from Theorem 3.1 is crucial.

- **Case 1:** $\Delta - x_0 - \varepsilon < \Delta\rho$.

  Recall that by (C) in Section 3.3.1, $k_1 = \delta k$ where $\delta = \frac{n}{k} - \left\lfloor \frac{n}{k} \right\rfloor$. If $\frac{n}{k} \leqslant a - 1$, then there are no parts of size a in π_1. If $\frac{n}{k} > a - 1$, then by (B), $\left\lceil \frac{n}{k} \right\rceil = a$ and $\left\lfloor \frac{n}{k} \right\rfloor = a - 1$, so it follows from (3.46) that $\delta = \Delta - x_0 - \varepsilon + o(1) \leqslant \Delta\rho + o(1)$. Therefore, if $\frac{n}{k} > a - 1$, there are $k_1 = \delta k \leqslant \Delta\rho k + o(k)$ parts of size a in π_1, so $v_1 \leqslant k_1 a \leqslant \Delta\rho n + o(n)$. In both cases, from (3.45),

$$\sum_{\pi_2 \in \mathcal{P}'(r_1,r_2,d_1,d_2)} k^{-n} b^{d(\pi_1,\pi_2)} \leqslant b^{n(1-\rho)\log_b(1-\rho)+\frac{\Delta\rho n}{2}-\frac{\Delta}{2}\cdot\rho n} \exp(o(n))$$
$$= b^{n(1-\rho)\log_b(1-\rho)} \exp(o(n)) \leqslant b^{-c_6 n} \exp(o(n)), \tag{3.47}$$

  where $c_6 := \min\left(-(1-c)\log(1-c), -c'\log c'\right) > 0$, since $c < \rho \leqslant 1 - c'$.

- **Case 2:** $\Delta\rho \leqslant \Delta - x_0 - \varepsilon \leqslant \rho$.

  Note that $\Delta\rho \geqslant c\varepsilon$ since $\rho > c$, and $\Delta \geqslant \varepsilon + x_0 + \Delta\rho \geqslant \varepsilon$. Therefore, by (3.46) and as $\Delta \leqslant 1$,
  $$a - 1 + c\varepsilon \leqslant \frac{n}{k} + o(1) \leqslant a - \varepsilon + o(1),$$
  so in particular $\lfloor \frac{n}{k} \rfloor = a - 1$. By (C) and (3.46), π_1 has $k_1 = \delta k = (\Delta - x_0 - \varepsilon + o(1))k$ parts of size a. Therefore, v_1 can be at most $(\Delta - x_0 - \varepsilon + o(1))ka$, and as $ka \sim n$,
  $$\frac{v_1}{2} - \frac{\Delta v}{2} \leqslant \frac{\Delta - x_0 - \varepsilon}{2} \cdot n - \frac{\Delta}{2}\rho n + o(n) = n\left[\frac{\Delta}{2}(1-\rho) - \frac{x_0 + \varepsilon}{2}\right] + o(n).$$
  Hence, by (3.45),
  $$\sum_{\pi_2 \in \mathcal{P}'(r_1, r_2, d_1, d_2)} k^{-n} b^{d(\pi_1, \pi_2)} \leqslant b^{n\left[(1-\rho)\log_b(1-\rho) + \frac{\Delta}{2}(1-\rho) - \frac{x_0+\varepsilon}{2}\right]} \exp(o(n)).$$
  As remarked above, $\Delta\rho \geqslant c\varepsilon$, so $\Delta - \varepsilon - x_0 \geqslant c\varepsilon$. Therefore, we can apply Lemma 3.8 with $\varepsilon' = c\varepsilon$ and $c_4 = c_4(\varepsilon, c\varepsilon)$ to conclude that
  $$\sum_{\pi_2 \in \mathcal{P}'(r_1, r_2, d_1, d_2)} k^{-n} b^{d(\pi_1, \pi_2)} \leqslant b^{-c_4 n} \exp(o(n)). \tag{3.48}$$
  Note that the proof of Lemma 3.8 requires Lemma 3.7, which in turn uses condition (3.4).

- **Case 3:** $\Delta - x_0 - \varepsilon > \rho$.

  Noting that $v_1 + v_2 = v = \rho n$, we proceed from (3.45).
  $$\begin{aligned}\sum_{\pi_2 \in \mathcal{P}'(r_1, r_2, d_1, d_2)} k^{-n} b^{d(\pi_1, \pi_2)} &\leqslant b^{n(1-\rho)\log_b(1-\rho) + \frac{1-\Delta}{2}v} \exp(o(n)) \\ &= b^{n \cdot \left[(1-\rho)\log_b(1-\rho) + \frac{1-\Delta}{2}\rho\right]} \exp(o(n)).\end{aligned}$$
  Since $c \leqslant \rho \leqslant \Delta - x_0 - \varepsilon$, we can use Lemma 3.7 (the proof of which uses condition (3.4)) with $\varepsilon' = c$ to see that this expression is exponentially decreasing in n.
  $$\sum_{\pi_2 \in \mathcal{P}'(r_1, r_2, d_1, d_2)} k^{-n} b^{d(\pi_1, \pi_2)} \leqslant b^{-c_3 n} \exp(o(n)). \tag{3.49}$$

By (3.47), (3.48) and (3.49), if we let $c_7 = \min(c_3, c_4, c_6) > 0$, then

$$\sum_{\pi_2 \in \mathcal{P}'(r_1, r_2, d_1, d_2)} k^{-n} b^{d(\pi_1, \pi_2)} \leqslant b^{-c_7 n} \exp(o(n)).$$

Since there are only $O(n^4 \log^2 n)$ choices for the values of $v_i \leqslant n$ and $d_i \leqslant f = O(n \log n)$ for $i = 1, 2$, this implies

$$\sum_{\pi_2 \in \mathcal{P}^{\mathrm{III}}} k^{-n} b^{d(\pi_1, \pi_2)} \exp(o(\log n)) = o(1).$$

□

From Lemmas 3.14 and 3.16 together with (3.41), it follows that

$$\sum_{\mathbf{r} \in \mathcal{R}_2^{c'}} Q_{\mathbf{r}} b^d = o(1),$$

as required.

3.5.4 Very high overlap

We are left with those overlap sequences $\mathbf{r}$ where $\rho = v/n > 1 - c'$ and $\sum_{2 \leqslant i \leqslant 0.6\gamma} i r_i \leqslant c'n$ for any constant $c' \in (0, 1)$ of our choosing. This means that all but at most $c'n$ vertices are involved in the overlap, and of those vertices involved in the overlap, all but at most $c'n$ are in large overlap blocks of size at least 0.6γ. Roughly speaking, in this case the large overlap blocks are mostly just permuted amongst themselves, and there are a small number of exceptional vertices which need to be studied in more detail. Let

$$\mathcal{R}_3^{c'} = \left\{ \mathbf{r} \mid \rho > 1 - c', \sum_{2 \leqslant i \leqslant 0.6\gamma} i r_i \leqslant c'n \right\}. \tag{3.50}$$

We will show that if we pick $c' > 0$ small enough, then the contribution from $\mathcal{R}_3^{c'}$ to the sum (3.20) is $o(1)$. We will pick $c' > 0$ later in this section, and to ensure this is not circular, we will take care that none of the implicit constants in our O-notation depend on c'.

As in the previous section, let π_1 be an *arbitrary fixed ordered k-equipartition.* Recall that for an overlap sequence $\mathbf{r}$, we denote by $P'_{\mathbf{r}}$ the number of ordered k-equipartitions with overlap $\mathbf{r}$ with π_1, and that by (3.40), $Q_{\mathbf{r}} = \frac{P'_{\mathbf{r}}}{P}$. Let

$$\mathcal{P}_3 = \left\{ \text{ordered } k\text{-equipartitions } \pi_2 \text{ such that } \mathbf{r}(\pi_1, \pi_2) \in \mathcal{R}_3^{c'} \right\}, \tag{3.51}$$

and recall that $\mu_k = Pq^f$ by (3.16). Then

$$\begin{aligned} \sum_{\mathbf{r}\in\mathcal{R}_3^{c'}} Q_{\mathbf{r}} b^d &= \sum_{\mathbf{r}\in\mathcal{R}_3^{c'}} \frac{P'_{\mathbf{r}}}{P} b^d = \frac{b^f}{P} \sum_{\mathbf{r}\in\mathcal{R}_3^{c'}} P'_{\mathbf{r}} \cdot b^{d-f} = \frac{1}{\mu_k} \sum_{\mathbf{r}\in\mathcal{R}_3^{c'}} P'_{\mathbf{r}} \cdot b^{-(f-d)} \\ &= \frac{1}{\mu_k} \sum_{\pi_2 \in \mathcal{P}_3} b^{-(f-d(\pi_1,\pi_2))}, \end{aligned} \tag{3.52}$$

where $d(\pi_1, \pi_2) := d(\mathbf{r})$ if $\mathbf{r}$ is the overlap sequence of π_1 and π_2.

Starting with π_1, we will generate, and count the number of choices for, $\pi_2 \in \mathcal{P}_3$. Since $v = \rho n \geqslant (1-c')n$ and $\sum_{2\leqslant i\leqslant 0.6\gamma} ir_i \leqslant c'n$, most of the overlap between π_1 and π_2 consists of large overlap blocks which are merely permuted. More specifically, given $\pi_2 \in \mathcal{P}_3$, we call an overlap block *large* if it contains at least 0.53γ vertices, and let

$$L = \text{ set of large overlap blocks of size at least } 0.53\gamma.$$

No part of π_1 can contain more than one large overlap block, and some parts may not contain any large overlap block at all. It will be more important later to talk about the latter type of part, so given $\pi_2 \in \mathcal{P}_3$, let

$$T = \text{ set of parts of } \pi_1 \text{ containing no large overlap block.}$$

We call a vertex *exceptional* if it is either not in the overlap at all or not in a large overlap block. If $\pi_2 \in \mathcal{P}_3$, then by definition there are at most $2c'n$ exceptional vertices. We shall distinguish between *three types of exceptional vertices.*

Again given $\pi_2 \in \mathcal{P}_3$, let

$$\begin{aligned} S &= \text{set of exceptional vertices} \\ S_1 &= \text{set of exceptional vertices not in parts in } T\text{, i.e., in parts containing} \\ &\quad\ \text{a large overlap block} \\ S_2 &= \text{set of exceptional vertices in parts in } T \text{ which are either not in the} \\ &\quad\ \text{overlap at all or in overlap blocks of size at most 100} \\ S_3 &= \text{set of exceptional vertices in parts in } T \text{ which are in overlap blocks} \\ &\quad\ \text{of size greater than 100} \\ g &= \text{number of overlap blocks of vertices in } S_3. \end{aligned}$$

Let $s = |S|$, $s_i = |S_i|$, and $t = |T|$. Then, as the vertices in parts in T are exactly those in $S_2 \cup S_3$, and since by (B), $a - 2 \leqslant \lfloor \frac{n}{k} \rfloor \leqslant \lceil \frac{n}{k} \rceil \leqslant a$,

$$\frac{s_2 + s_3}{a} \leqslant t \leqslant \frac{s_2 + s_3}{a - 2}. \tag{3.53}$$

The vertices in S_3 are arranged in blocks of size between 100 and 0.53γ, so

$$\frac{s_3}{0.53\gamma} \leqslant g \leqslant \frac{s_3}{100}. \tag{3.54}$$

Fix $\mathbf{s} = (s_1, s_2, s_3)$, g, and t such that $s = s_1 + s_2 + s_3 \leqslant 2c'n$ and (3.53) and (3.54) hold, and let

$$\mathcal{P}(\mathbf{s}, t, g) = \big\{\pi_2 \in \mathcal{P}_3 \mid \mathbf{s}(\pi_1, \pi_2) = \mathbf{s},\ t(\pi_1, \pi_2) = t,\ g(\pi_1, \pi_2) = g\big\}.$$

Note that

$$\mathcal{P}_3 = \bigcup_{\mathbf{s}, t, g : s \leqslant 2c'n} \mathcal{P}(\mathbf{s}, t, g). \tag{3.55}$$

Starting with the fixed partition π_1 and given $\mathbf{s}$, g, t, we will generate all $\pi_2 \in \mathcal{P}(\mathbf{s}, t, g)$ and sum $b^{-(f - d(\pi_1, \pi_2))}$ to bound the contribution to (3.52). We will proceed in the following way: first, we choose all three sets of exceptional vertices, bounding the number of choices in Lemma 3.17. Next, we generate π_2 by permuting the exceptional vertices amongst themselves and then permuting all the parts, taking into account that part sizes may vary between $\lceil \frac{n}{k} \rceil$ and $\lfloor \frac{n}{k} \rfloor$. The number of ways to generate π_2 in this way is bounded in Lemma 3.18. Finally, in Lemma 3.19, we will examine how much each exceptional vertex of each type subtracts from the maximum possible number f of shared forbidden edges between π_1 and π_2, and we obtain a lower bound for $f - d(\pi_1, \pi_2)$ which will be used afterwards to bound $b^{-(f-d(\pi_1,\pi_2))}$ from above.

Lemma 3.17. *For fixed* $\mathbf{s} = (s_1, s_2, s_3)$*,* g *and* t *and the fixed partition* π_1*, there are at most*

$$\frac{n^{s_1} k^t 2^{s_3} t^g a^{s_3}}{s_1! g!}$$

ways to choose the sets S_1*,* S_2 *and* S_3 *and arrange the vertices in* S_3 *into* g *overlap blocks.*

Proof. We first choose the vertices in S_1 and the parts in T. For this there are at most

$$\binom{n}{s_1}\binom{k}{t} \leqslant \frac{n^{s_1} k^t}{s_1!}$$

possibilities. Next, we pick the vertices in S_3 from within the parts in T along with the g overlap blocks they make up. Since we do not know the exact sizes of these overlap blocks, we first write s_3 as an ordered sum of g positive summands, which can be done in $\binom{s_3-1}{g-1}$ ways. Next, we decide which of the parts in T each of the g blocks is in, for which there are at most t^g choices, and then we pick the vertices that belong to each of the g blocks. We know which part of size at most a each such vertex is in, and we choose s_3 vertices in total, so there are at most a^{s_3} possibilities for this. Finally, since we do not care about the order of the g overlap blocks, we can divide by $g!$. So overall, there are at most

$$\binom{s_3 - 1}{g - 1} t^g a^{s_3} \frac{1}{g!} \leqslant \frac{2^{s_3} t^g}{g!} a^{s_3}$$

ways of selecting the vertices in S_3 along with the g overlap blocks they are arranged in. The remaining vertices in the parts in T must be exactly those in S_2. □

Let

$$\tau = \max\left(1, \Gamma\left(\frac{s_2}{t}\right)^t\right), \tag{3.56}$$

where $\Gamma(\cdot)$ denotes the gamma function.

Lemma 3.18. *Given π_1, S_1, S_2, S_3 and the overlap blocks that the vertices in S_3 are arranged in, there are at most*

$$2^k \frac{(s_1+s_2+g)!}{\tau} k_1! k_2!$$

possibilities for π_2.

Proof. Note that each part in π_1 and π_2 contains at most one large overlap block from L, since one such block occupies more than half of a part. Therefore, since we know S_1, S_2 and S_3, we also know L. In each part of π_1, there are a certain number of 'slots' for exceptional vertices, with the rest of the part occupied by at most one block from L. The numbers of slots for exceptional vertices in parts of π_2 are essentially just a permutation of the numbers of slots in π_1, because the remainders of the parts in π_2 are again occupied by at most one block from L. However, as total part sizes vary between $\lceil \frac{n}{k} \rceil$ and $\lfloor \frac{n}{k} \rfloor$, the numbers of available slots in each part may also increase or decrease by 1.

Therefore, starting with π_1, we can generate every possible partition π_2 in the following way. Each of the k parts of π_1 contains a certain number of exceptional vertices, and for each part we decide whether or not to increase or decrease the number of available slots for exceptional vertices by 1, for which there are at most 2^k possibilities. We write the vertices in S_1 and S_2 along with the g blocks comprising the vertices in S_3 as a list and permute them, which can be done in

$$(s_1+s_2+g)!$$

ways. Now we divide up the list successively according to the number of available slots in each of the k parts (discarding the cases where this is not possible because one of the g blocks would have to be divided), and move the vertices from each division to the corresponding part. Finally, we permute all k_1 parts of (new) size $\lceil \frac{n}{k} \rceil$ and all k_2 parts of (new) size $\lfloor \frac{n}{k} \rfloor$, for which there are

$$k_1!k_2!$$

possibilities, and re-order the parts so that those of size $\lceil \frac{n}{k} \rceil$ come first, followed by those of size $\lfloor \frac{n}{k} \rfloor$, yielding the new ordered k-equipartition π_2.

However, we have overcounted the number of ways to generate π_2: each possible partition π_2 was counted at least τ times, where τ is defined in (3.56). To see this, suppose we have generated a partition π_2. Note that the number of available slots for exceptional vertices in the parts in T is at least $s_2 + s_3 - t$, since there were initially $s_2 + s_3$ exceptional vertices in the parts in T, and at most t slots can be 'lost'. So at least $s_2 + s_3 - t$ vertices were moved to the available slots in T, and of these, at most s_3 were in one of the g overlap blocks. Therefore, there were at least $s_2 - t$ vertices which were permuted and then moved to the parts in T as singletons. Denote the number of such singletons assigned to each of the parts in T by l_1, l_2, ..., l_t, where $\sum_{i=1}^{t} l_i \geqslant s_2 - t$. Then, since we do not care about the order of the vertices within the parts, we counted π_2 at least $\prod_{i=1}^{t} l_i!$ times.

Note that $l_i! = \Gamma(l_i + 1)$, where $\Gamma(\cdot)$ denotes the gamma function. By the Bohr–Mollerup Theorem (see for example §13.1.10 in [36]), $\log \Gamma(\cdot)$ is a convex function on the positive reals, so from Jensen's inequality,

$$\log \left(\prod_{i=1}^{t} l_i! \right) = \sum_{i=1}^{t} \log(\Gamma(l_i + 1)) \geqslant t \log \left(\Gamma \left(\frac{1}{t} \sum_{i=1}^{t} l_i + 1 \right) \right),$$

and therefore $\prod_{i=1}^{t} l_i! \geqslant \Gamma \left(\frac{s_2 - t}{t} + 1 \right)^t = \Gamma \left(\frac{s_2}{t} \right)^t$. Hence, we may divide our result by τ. □

Lemma 3.19. *If* $\pi_2 \in \mathcal{P}(\mathbf{s}, t, g)$*, then*

$$f - d(\pi_1, \pi_2) \geqslant 0.53\gamma s_1 + (\gamma/2 - 51)\, s_2 + 0.23\gamma s_3.$$

Proof. Note that the number $d(\pi_1, \pi_2)$ of shared forbidden edges is exactly the number of pairs of vertices which are in the same part in both π_1 and π_2, and f is the number of pairs of vertices which are in the same part of π_1. Therefore, if we let

$$E = \big\{ \{v, w\} \mid \ v \text{ and } w \text{ are in the same part of } \pi_1 \text{ but in different parts of } \pi_2 \big\},$$

then $f - d(\pi_1, \pi_2) = |E|$. Each exceptional vertex $v \in S$ contributes at least a certain amount to $|E|$ according to its type.

If v is in S_1, then v is in a part of π_1 which contains a large overlap block. Therefore, there are at least 0.53γ vertices $w \notin S$ such that $\{v, w\} \in E$. Therefore, the contribution from S_1 to $|E|$ is at least $0.53\gamma s_1$.

Since the vertices in S_2 are in overlap blocks of size at most 100, for each $v \in S_2$, there are at least $\lfloor \frac{n}{k} \rfloor - 100 \geqslant \gamma - 102$ vertices w such that $\{v, w\} \in E$. As the vertices in S_3 are exceptional and therefore in overlap blocks of size at most 0.53γ, for each $v \in S_3$, there are at least $\lfloor \frac{n}{k} \rfloor - 0.53\gamma \geqslant 0.46\gamma$ vertices w such that $\{v, w\} \in E$. However, we have counted each such pair $\{v, w\}$ twice, and must therefore divide the total number by 2. So the contribution from $S_2 \cup S_3$ to $|E|$ is at least $(\gamma/2 - 51)\, s_2 + 0.23\gamma s_3$. □

By (F) from Section 3.3.1, $b^{-\gamma} \sim k^{-2}$, so if $\pi_2 \in \mathcal{P}(\mathbf{s}, t, g)$, from Lemma 3.19,

$$b^{-(f-d(\pi_1,\pi_2))} \leqslant k^{-1.06s_1 - s_2 - 0.46s_3} \exp\left(O\left(s\right)\right) \leqslant n^{-1.05s_1} k^{-s_2 - 0.46s_3} \exp\left(O\left(s\right)\right).$$

Together with Lemmas 3.17 and 3.18, this gives

$$\begin{aligned}
&\sum_{\pi_2 \in \mathcal{P}(\mathbf{s},t,g)} b^{-(f-d(\pi_1,\pi_2))} \\
&\qquad \leqslant \frac{n^{s_1} k^t 2^{s_3} t^g a^{s_3}}{s_1! g!} 2^k \frac{(s_1 + s_2 + g)!}{\tau} k_1! k_2! n^{-1.05s_1} k^{-s_2 - 0.46s_3} \exp\left(O\left(s\right)\right) \\
&\qquad = k_1! k_2! 2^k n^{-0.05s_1} t^g a^{s_3} \frac{(s_1 + s_2 + g)!}{s_1! g! \tau} k^{t - s_2 - 0.46s_3} \exp\left(O\left(s\right)\right) \\
&\qquad = k_1! k_2! 2^k n^{-0.05s_1} t^g a^{s_3} \frac{(s_1 + s_2 + g)! s_2!}{s_1! s_2! g! \tau} k^{t - s_2 - 0.46s_3} \exp\left(O\left(s\right)\right).
\end{aligned}$$

Note that $\frac{(s_1+s_2+g)!}{s_1!s_2!g!} \leqslant 3^{s_1+s_2+g} = \exp\left(O\left(s\right)\right)$ and by (3.53), $s_2! \leqslant s_2^{s_2} \leqslant t^{s_2} a^{s_2}$, so together with (3.54),

$$\sum_{\pi_2 \in \mathcal{P}(\mathbf{s},t,g)} b^{-(f-d(\pi_1,\pi_2))} \leqslant k_1! k_2! 2^k n^{-0.05 s_1} t^{s_2+s_3/100} a^{s_2+s_3} \frac{1}{\tau} k^{t-s_2-0.46s_3} \exp\left(O\left(s\right)\right)$$
$$\leqslant k_1! k_2! 2^k n^{-0.05 s_1} t^{s_2+s_3/100} a^{s_2+s_3} \frac{1}{\tau} k^{-s_2-0.46s_3} \exp\left(O\left(s\right)\right), \tag{3.57}$$

as $k^t \leqslant k^{(s_2+s_3)/(a-2)} \leqslant \exp(O(s_2+s_3))$ by (G) from Section 3.3.1. Let

$$T(s_2,s_3) = t^{s_2+s_3/100} a^{s_2+s_3} \frac{1}{\tau} k^{-s_2-0.46s_3} \exp\left(C_2(s_2+s_3)\right),$$

where $C_2 > 0$ is the constant implicit in the term $O\left(s\right)$ above. We distinguish two cases.

- **Case 1:** $s_3 \geqslant 100 s_2$.

  By (3.53), $s_3 \leqslant s_2+s_3 \leqslant at$, so $s_2 \leqslant 0.01at$. Since again by (3.53), $s_2+0.46s_3 \geqslant 0.46(s_2+s_3) \geqslant 0.46(a-2)t$, and $t \leqslant k$ and $\tau \geqslant 1$,

  $$T(s_2,s_3) \leqslant t^{0.02at} a^{at} k^{-0.46(a-2)t} \exp(O(s_2+s_3)) \leqslant \left(\frac{k^{0.02}a}{k^{0.45}}\right)^{at} \exp(O(s_2+s_3))$$
  $$\leqslant n^{-0.4at} \exp(O(s_2+s_3)) \leqslant n^{-0.3(s_2+s_3)}$$

  if n is large enough.

- **Case 2:** $s_3 < 100 s_2$.

  Then by (3.53), $t \leqslant \frac{101 s_2}{a-2} \leqslant \frac{51 s_2}{\log_b n}$, so $\frac{s_2}{t} \geqslant \frac{\log_b n}{51}$. By the Stirling approximation of the Gamma function,

  $$\tau \geqslant \Gamma\left(s_2/t\right)^t \geqslant \left(\frac{\frac{s_2}{t}-1}{e}\right)^{s_2-t} \geqslant (\log n)^{s_2} \exp(O(s_2)).$$

  Furthermore, $t \leqslant \frac{s_2+s_3}{a-2} \leqslant \frac{s}{a-2} \leqslant \frac{2c'n}{a-2} \leqslant 3c'k$ if n is large enough, and therefore $\frac{t}{k} \leqslant 3c'$ for n large enough. So since $a \leqslant 2\log_b n = 2\log n/\log b$,

  $$T(s_2,s_3) \leqslant \left(\frac{ta}{k\log n}\right)^{s_2} \left(\frac{t^{0.01}a}{k^{0.46}}\right)^{s_3} \exp(O(s_2+s_3))$$
  $$\leqslant \left(\frac{6c'}{\log b}\right)^{s_2} \left(\frac{t^{0.01}a}{k^{0.46}}\right)^{s_3} \exp(O(s_2+s_3)) \leqslant \left(\frac{1}{2}\right)^{s_2} n^{-0.3s_3}$$

  for n large enough if $c' > 0$ is picked small enough. Pick the constant $c' > 0$ small enough for this.

Therefore in both cases, from (3.57) if n is large enough,

$$\sum_{\pi_2 \in \mathcal{P}(\mathbf{s},t,g)} b^{-(f-d(\pi_1,\pi_2))} \leqslant k_1!k_2!2^k n^{-0.05s_1} 2^{-s_2} n^{-0.3s_3} \exp(O(s_1))$$
$$\leqslant k_1!k_2!2^k n^{-0.04s_1} 2^{-s_2} n^{-0.3s_3}.$$

If we sum over s_1, s_2 and s_3 and recall that $g \leqslant s_3/100$ and $t \leqslant s_2 + s_3$, by (3.55) we get a bound for (3.52).

$$\sum_{\mathbf{r} \in \mathcal{R}_3^{c\prime}} Q_{\mathbf{r}} b^d \leqslant \frac{k_1!k_2!2^k}{\mu_k} \sum_{s_1,s_2,s_3} \left(0.01 s_3(s_2+s_3) n^{-0.04s_1} 2^{-s_2} n^{-0.3s_3}\right) = O\left(\frac{k_1!k_2!}{\mu_k} \cdot 2^k\right).$$

As $k = O\left(\frac{n}{\log n}\right)$, and since by (K) from Section 3.3.1, $\frac{\mu_k}{k_1!k_2!} \geqslant b^{\varepsilon n/4}$, we can see that this last expression is exponentially decreasing in n, and in particular it is $o(1)$ as required. This concludes the proof of Theorem 3.1.

3.6 Outlook

Recall that the chromatic number of random graphs is highly concentrated, and that in particular for constant p, there is an interval of length roughly $\sqrt{n}/\log n$ which contains $\chi(G)$ whp. The proof of this concentration result gives no clue about the location of this interval, however. While the new explicit bounds match up to the $o(1)$-term in the denominator, the absolute gap between them which is implicit in the proof is still at least $\frac{n \log\log n}{\log^3 n}$, which is of course larger than $\sqrt{n}/\log n$.

Therefore, a more detailed result about the size of the $o(1)$-term would be interesting. For $p \leqslant 1 - 1/e^2$, where the lower bound comes from the first moment threshold for the number of partitions which induce proper colourings [50], this gap is unavoidable as long as the upper bound is obtained through the study of partitions which induce equitable colourings, since the first moment thresholds of colourings and equitable colourings are separated by this distance.

Recall also that for small $p = p(n) \ll n^{-1/2}$, the chromatic number of $\mathcal{G}(n,p)$ is concentrated on two values whp [3], and this is generally the smallest possible interval one can hope for. In contrast, the question of the concentration of the chromatic number of dense random graphs is wide open. Even the most basic non-concentration results, such as showing that we do not in general have two-point concentration, would be of interest (see also [8]).

In the next chapter, we will show that for the dense random graph $\mathcal{G}(n, m)$, the *equitable* chromatic number is concentrated on one explicit value for a subsequence of the integers.

Chapter 4

Equitable colourings

4.1 Background and results

In the last chapter, we considered proper colourings of $G \sim \mathcal{G}(n,p)$ with p constant and determined the *average* colour class size in an optimal colouring up to a term of order $o(1)$. In many applications of proper graph colourings, it is desirable for the colour classes to be as equal in size as possible. For example, consider a scheduling problem where the vertices of a graph represent events and the edges signify pairs of events that may not be scheduled at the same time. While the number of colours in a proper colouring gives us the required number of time slots, the colour class sizes tell us how many events take place at the same time. It often makes sense to keep the number of parallel events as equal as possible in order to optimise the use of available resources, such as the number of available rooms in a building.

This is exactly what is achieved by an *equitable colouring* of the vertices of a graph G, i.e., a colouring where the colour class sizes differ by at most 1. Recall from the introduction that the *equitable chromatic number* $\chi_=(G)$ of a graph G is the minimum number colours needed for an equitable colouring. The *equitable chromatic threshold* $\chi_=^*(G)$ is defined as the least number k such that for all $l \geqslant k$, there is an equitable l-colouring of G. The Hajnal-Szemerédi Theorem states that if $\Delta(G)$ is the maximum degree of a graph G, then

$$\chi_=^*(G) \leqslant \Delta(G) + 1.$$

This was proved by Hajnal and Szemerédi in 1970 [28], confirming a conjecture by

Erdős. Chen, Li and Wu [13] conjectured that in fact $\chi^*_=(G) \leqslant \Delta(G)$ for every connected graph G which is not an odd cycle C_{2n+1}, the complete graph K_n, or the complete bipartite graph $K_{2n+1,2n+1}$ with equal odd part sizes.

For random graphs, Krivelevich and Patkós [37] proved, amongst other things, that for $G \sim \mathcal{G}(n,p)$ with $n^{-1/5+\varepsilon} \leqslant p \leqslant 0.99$, whp

$$\chi(G) \leqslant \chi_=(G) \leqslant (1+o(1))\chi(G),$$

and if $p < 0.99$ and $\log\log n \ll \log(np)$,

$$\chi(G) \leqslant \chi^*_=(G) \leqslant (2+o(1))\chi(G).$$

They also conjecture that there is a constant C such that if $C/n < p < 0.99$,

$$\chi^*_=(G) \leqslant (1+o(1))\chi(G).$$

Rombach and Scott give more results in their forthcoming paper [51].

Despite the fact that we used equitable colourings to prove the upper bound in Theorem 3.1, the proof itself does *not* imply the whp existence of an equitable colouring of $G \sim \mathcal{G}(n,p)$, p constant, with a near optimal number of colours. The only direct implication of the proof is that by (3.9), for k as in (3.7), the probability that there is an equitable k-colouring of G is at least $\exp\left(-\frac{n}{\log^7 n}\right)$.

We proceeded with bounded differences arguments to boost this lower bound to one that tends to 1 by adding only a few colours, but these arguments only apply to general colourings, not equitable ones.

We can, however, use a different bounded differences argument to show that there is an 'almost equitable' colouring with a near optimal number of colours.

Theorem 4.1. *Let $p \in (0,1)$ be constant, and consider the random graph $G \sim \mathcal{G}(n,p)$. Define b, γ and x_0 as in Theorem 3.1. Then whp, G has a colouring with*

$$\frac{n}{\gamma - x_0 + o(1)}$$

colours such that the sizes of all but $o\left(\frac{n}{\gamma}\right)$ colour classes differ by at most 1.

We will prove Theorem 4.1 in Section 4.2. While Theorem 4.1 implies the existence of an almost equitable colouring, the main focus of this chapter will be on completely equitable colourings.

Recall from the introduction that the question of the concentration of the chromatic number is an important open problem for dense random graphs – see for example [8]. While the bounds from the last chapter narrow down the smallest *explicit* interval known to contain the chromatic number whp, they do not improve the current best concentration result, namely that the chromatic number is whp contained in a (non-explicit) interval of size about $\sqrt{n}/\log n$.

We will adapt arguments from the proof of Theorem 3.1 to show that on a subsequence of the integers, the *equitable* chromatic number of the dense random graph $\mathcal{G}(n, m)$ is concentrated on one explicit value.

Theorem 4.2. *Let $0 < p < 1 - 1/e^2$ be constant. There exists a strictly increasing sequence of integers $(n_j)_{j \geqslant 1}$ and $j_0 \geqslant 1$ such that*

a) for all $j \geqslant j_0$, $j | n_j$,

b) letting $b = \frac{1}{1-p}$ and $\gamma_j = 2\log_b n_j - 2\log_b \log_b n_j - 2\log_b 2$,

$$\gamma_j = j + o(1) \text{ as } j \to \infty, \text{ and}$$

c) letting $G \sim \mathcal{G}(n_j, m_j)$ with $m_j = \left\lfloor p\binom{n_j}{2} \right\rfloor$, with high probability as $j \to \infty$,

$$\chi_{=}(G) = \frac{n_j}{j}.$$

In other words, we can pick a subsequence $(n_j)_{j \geqslant 1}$ of the integers so that whp as $j \to \infty$, the equitable chromatic number of $G \sim \mathcal{G}(n_j, m_j)$ with $m_j = \left\lfloor p\binom{n_j}{2} \right\rfloor$ is exactly $\frac{n_j}{j}$.

We will prove Theorem 4.2 in Sections 4.3–4.6. The proof relies on choosing the subsequence $(n_j)_{j \geqslant 1}$ in such a way that exactly as the expected number of equitable colourings starts tending to infinity, all the colour classes have exactly the same size. Working in $\mathcal{G}(n, m)$ instead of $\mathcal{G}(n, p)$, we can then show that the second moment of the number of equitable colourings is asymptotically equal to the square of the first moment, which will imply the whp existence of an equitable colouring.

4.2 Proof of Theorem 4.1

The proof is similar to an argument due to Frieze, see Fact 1 in [40]. In (3.9), we showed that for any fixed $\varepsilon > 0$ and

$$k = \left\lceil \frac{n}{\gamma - x_0 - \varepsilon} \right\rceil,$$

if n is large enough, then the probability that G has an equitable k-colouring is at least $\exp\left(-\frac{n}{\log^7 n}\right)$. Since $\varepsilon > 0$ was arbitrary, we can also pick a sequence $\varepsilon(n) \to 0$ so that this remains true for $k' = k'(n, \varepsilon(n))$.

Now let Y denote the maximum number of vertices which can be coloured using at most k' colours such that the colour classes are all of size exactly $\lceil \frac{n}{k'} \rceil$ or $\lfloor \frac{n}{k'} \rfloor$. Then

$$\mathbb{P}(Y = n) \geqslant \exp\left(-\frac{n}{\log^7 n}\right). \tag{4.1}$$

If we change all the edges of G incident with any particular vertex v, then Y changes by at most $2\log_b n$: a given partial colouring remains valid if we remove the colour class containing v (if v is in a colour class), which is of size at most $\lceil \frac{n}{k'} \rceil \leqslant 2\log_b n$, so Y cannot increase or decrease by more than $2\log_b n$. Therefore, the Bounded Differences Inequality, Theorem 2.8, implies that for any $t > 0$,

$$\mathbb{P}(|Y - \mathbb{E}Y| > t) \leqslant 2\exp\left(-t^2/(2n\log_b^2 n)\right). \tag{4.2}$$

Hence, $\mathbb{E}Y \geqslant n - \frac{n}{\log^{2.4} n}$; otherwise (4.1) would be a contradiction to (4.2) with $t = \frac{n}{\log^{2.4} n}$. But then again using (4.2), we get

$$\mathbb{P}\left(Y < n - \frac{2n}{\log^{2.4} n}\right) \leqslant \mathbb{P}\left(|Y - \mathbb{E}Y| > \frac{n}{\log^{2.4} n}\right) = o(1).$$

Therefore, whp there is a partial colouring with at most k' colours where all colour classes are of size exactly $\lceil \frac{n}{k'} \rceil$ or $\lfloor \frac{n}{k'} \rfloor$, and where at most $\frac{2n}{\log^{2.4} n} = o\left(\frac{n}{\log n}\right) = o\left(\frac{n}{\gamma}\right)$ vertices are left over. We can extend this to a complete colouring by creating an individual colour class for each leftover vertex, and denote by K the total number of colour classes. Then this colouring is of the desired form, and as $\gamma - x_0 = \Theta(\log n)$,

$$K \leqslant k' + \frac{2n}{\log^{2.4} n} = \frac{n}{\gamma - x_0 + o(1)} + O\left(\frac{n}{\log^{2.4} n}\right) = \frac{n}{\gamma - x_0 + o(1)}.$$

Furthermore, by Theorem 3.1, whp

$$K \geqslant \chi(G) = \frac{n}{\gamma - x_0 + o(1)}.$$

□

4.3 Outline of the proof of Theorem 4.2

From now on, fix $p < 1 - 1/e^2$. For $n \in \mathbb{N}$, let $m(n) = \left\lfloor p\binom{n}{2} \right\rfloor$ and $G \sim \mathcal{G}(n, m(n))$. For $u \in \mathbb{N}$, recall from the last chapter that an ordered partition of n vertices into u parts is called an *ordered u-equipartition* if all u parts have size $\left\lceil \frac{n}{u} \right\rceil$ or $\left\lfloor \frac{n}{u} \right\rfloor$ and decrease in size (so the parts of size $\left\lceil \frac{n}{u} \right\rceil$ come first, followed by the parts of size $\left\lfloor \frac{n}{u} \right\rfloor$). Denote by $X_{n,u}$ the number of ordered u-equipartitions of G which induce valid colourings.

We will start with a straightforward analysis of the first moment of $X_{n,u}$ in Section 4.4. Next, in Section 4.5, we will show that there is a strictly increasing sequence $(n_j)_{j \geqslant 1}$ which fulfils parts a) and b) of Theorem 4.2, and furthermore, letting $u_j = n_j/j$,

$$\mathbb{E}[X_{n_j,u_j}] \to \infty \text{ as } j \to \infty.$$

We will give a first moment argument to show that whp G has no equitable colouring with fewer than u_j colours, i.e., whp

$$\chi_{=}(G) \geqslant u_j = \frac{n_j}{j}.$$

The matching upper bound will be proved through the second moment method, adapting the arguments from the proof of Theorem 3.1. More specifically, in Section 4.6 we will show that

$$\mathbb{E}[X_{n_j,u_j}^2]/\mathbb{E}[X_{n_j,u_j}]^2 \to 1 \text{ as } j \to \infty.$$

Together with the Paley-Zygmund Inequality, Theorem 2.7, this implies that whp $X_{n_j,u_j} > 0$, so

$$\chi_=(G) \leqslant u_j = \frac{n_j}{j} \text{ whp as } j \to \infty,$$

concluding the proof of Theorem 4.2.

Remark

The fact that by our choice of n_j, all colour classes are of exactly the same size is crucial for $\mathbb{E}[X^2_{n_j,u_j}]/\mathbb{E}[X_{n_j,u_j}]^2 \to 1$. Roughly speaking, in order for this ratio to come out as 1, conditioning on the event E that one particular partition π_1 induces a colouring may not affect the expected number of other colourings very much, i.e., we must have $\mathbb{E}[X \mid E] \sim \mathbb{E}[X]$ where X denotes the number of partitions which induce valid colourings. But if the part sizes are not exactly the same, then after conditioning on E, we expect a lot of partitions which are very similar to π_1 to induce valid colourings as well. For example, suppose that $u \sim \frac{n}{\gamma(n)}$ and about half of all parts are of size $\lceil \frac{n}{u} \rceil$ and the other half of size $\lfloor \frac{n}{u} \rfloor$. Select one of the roughly $n/2$ vertices in the larger parts and move it to one of the roughly $u/2$ smaller parts. The probability that this vertex is not adjacent to any of the vertices in the new part is about $q^{\lfloor \frac{n}{u} \rfloor} = q^{\gamma(n)+O(1)} = \Theta\left(\frac{\log^2 n}{n^2}\right)$. So conditional on E inducing a valid colouring, we expect that roughly

$$\Theta\left(nu\frac{\log^2 n}{n^2}\right) = \Theta\left(\log n\right)$$

other partitions which can be obtained from π_1 by moving a single vertex from a larger to a smaller part induce valid colourings as well. Of course there are many more such partitions if we allow slightly larger variations. If $\mathbb{E}[X]$ tends to infinity relatively slowly, as it is the case for our sequence $(n_j)_{j\geqslant 1}$, then $\mathbb{E}[X \mid E] \gg \mathbb{E}[X]$ and $\mathbb{E}[X^2]/\mathbb{E}[X]^2 \nrightarrow 1$.

4.4 The first moment

Given integers n and u, $m = m(n)$ as before and $G \sim \mathcal{G}(n, m(n))$, we start by analysing the first moment of $X_{n,u}$. As in Chapter 3, let

$$\gamma = \gamma(n) = 2\log_b n - 2\log_b \log_b n - 2\log_b 2.$$

We will only examine the range

$$u = \frac{n}{\gamma + O(1)}.$$

Let $q = 1-p$, $N = \binom{n}{2}$, and $\delta = \delta_{n,u} = \frac{n}{u} - \left\lfloor \frac{n}{u} \right\rfloor$. If u does not divide n, then an ordered u-equipartition of n vertices consists of $u_{\mathrm{L}} = u_{\mathrm{L}}(n) = \delta_{n,u} u$ larger parts of size $\left\lceil \frac{n}{u} \right\rceil$, followed by $u_{\mathrm{S}} = u_{\mathrm{S}}(n) = (1-\delta_{n,u})u$ smaller parts of size $\left\lfloor \frac{n}{u} \right\rfloor$. If u divides n, then all $u = u_{\mathrm{S}}(n)$ parts are of size exactly $\frac{n}{u} = \left\lfloor \frac{n}{u} \right\rfloor$. As in (3.13), the total number of u-equipartitions is

$$P_{n,u} = \frac{n!}{\left\lceil \frac{n}{u} \right\rceil !^{u_{\mathrm{L}}} \left\lfloor \frac{n}{u} \right\rfloor !^{u_{\mathrm{S}}}}, \tag{4.3}$$

and as in (3.14),

$$P_{n,u} = u^n \exp(o(n)). \tag{4.4}$$

As in (3.15), denote by

$$f = f_{n,u} = u_{\mathrm{L}} \binom{\lceil n/u \rceil}{2} + u_{\mathrm{S}} \binom{\lfloor n/u \rfloor}{2} = \frac{n\left(\frac{n}{u} - 1\right)}{2} + \frac{\delta_{n,u}(1-\delta_{n,u})}{2} u \sim n \log_b n \tag{4.5}$$

the number of *forbidden edges* that may not be present in a u-equipartition for it to induce a valid u-colouring. Let

$$\varepsilon = Np - m \in [0,1].$$

Since $f = f_{n,u} \sim n\log_b n$ and by Stirling's formula $n! \sim \sqrt{2\pi n}\, n^n/e^n$, the probability that any such given partition induces a valid colouring of G is

$$\begin{aligned} \frac{\binom{N-f}{m}}{\binom{N}{m}} &= \frac{(N-f)!(qN+\varepsilon)!}{N!(qN-f+\varepsilon)!} \sim \frac{(N-f)^{N-f}(qN+\varepsilon)^{qN+\varepsilon}}{N^N(qN-f+\varepsilon)^{qN-f+\varepsilon}} \\ &= q^f \frac{(1-\frac{f}{N})^{N-f}(1+\frac{\varepsilon}{qN})^{qN+\varepsilon}}{(1-\frac{f}{qN}+\frac{\varepsilon}{qN})^{qN-f+\varepsilon}} \sim q^f \frac{(1-\frac{f}{N})^{N-f}(1+\frac{\varepsilon}{qN})^{qN}}{(1-\frac{f}{qN}+\frac{\varepsilon}{qN})^{qN-f}}. \end{aligned}$$

Using $\log(1+x) = x - \frac{x^2}{2} + O(x^3)$ for $x \to 0$ and that $\varepsilon \in (0,1)$ and $f = \Theta(n \log n)$,

$$\log \frac{\binom{N-f}{m}}{\binom{N}{m}} = f \log q - \frac{f^2 p}{2qN} + O\left(\frac{f^3}{N^2}\right).$$

Since $\frac{f^3}{N^2} = o(1)$,

$$\mu_{n,u} := \mathbb{E}[X_{n,u}] = P_{n,u} \frac{\binom{N-f_{n,u}}{m}}{\binom{N}{m}} \sim P_{n,u} q^{f_{n,u}} \exp\left(-\frac{f_{n,u}^2 p}{2qN}\right). \tag{4.6}$$

The expected number of *unordered* equitable partitions which induce valid colourings is

$$\bar{\mu}_{n,u} = \frac{\mu_{n,u}}{u_{\mathrm{L}}! u_{\mathrm{S}}!} \sim \frac{P_{n,u} q^{f_{n,u}}}{u_{\mathrm{L}}! u_{\mathrm{S}}!} \exp\left(-\frac{f_{n,u}^2 p}{2qN}\right). \tag{4.7}$$

The next lemma gives an approximation of $\bar{\mu}_{n,u}$ when u is not too close to n/γ.

Lemma 4.3. *Given n and u and $x = O(1)$ such that*

$$u = \frac{n}{\gamma + x},$$

then

$$\bar{\mu}_{n,u} = b^{-\frac{x}{2}n + o(n)}.$$

Proof. As in the proof of (K) in Section 3.3.1, since $1 \leqslant \binom{u}{u_{\mathrm{S}}} \leqslant 2^u$ and $u \sim \frac{n}{2 \log_b n}$, we can show that

$$u_{\mathrm{S}}! u_{\mathrm{L}}! = u! \exp(o(n)) = b^{n/2} \exp(o(n)). \tag{4.8}$$

Furthermore, by (4.5),

$$q^{f_{n,u}} = b^{-\frac{n(\frac{n}{u}-1)}{2}} \exp(o(n)) = b^{\frac{n}{2}(1-\gamma-x)} = b^{n\frac{1-x}{2}} \left(\frac{2 \log_b n}{n}\right)^n. \tag{4.9}$$

Finally, note that

$$\exp\left(-\frac{f_{n,u}^2 p}{2qN}\right) = \exp\left(O\left(\log^2 n\right)\right) = \exp(o(n)).$$

Together with (4.4), (4.7), (4.8), and (4.9), this gives

$$\bar{\mu}_{n,u} = \frac{u^n b^{n\frac{1-x}{2}}}{b^{n/2}} \left(\frac{2 \log_b n}{n}\right)^n \exp(o(n)) = b^{-\frac{x}{2}n} (1 + o(1))^n \exp(o(n)) = b^{-\frac{x}{2}n + o(n)}.$$

□

Later on, in Lemma 4.7, we will pick a sequence $(n_j)_{j\geqslant 1}$ such that $\bar{\mu}_{n_j,u_j}$ starts tending to infinity just as all parts of a u_j-equipartition are of size exactly j. For this, we will need the following lemmas which examine how much $\mu_{n,u}$ and $\bar{\mu}_{n,u}$ change if we increase u by 1.

It should be noted that $\bar{\mu}_{n,u}$ behaves in a slightly irregular way: while $\mu_{n,u}$ increases steadily if we increase u, as can be seen in Lemma 4.4, the increases in $\bar{\mu}_{n,u}$ are not as large as one might expect if n/u is close to an integer. This is because the product $u_{\mathrm{L}}!u_{\mathrm{S}}!$ is larger when n/u is close to an integer (i.e., if either u_{L} or u_{S} is close to u) than when n/u is sufficiently far away from any integers. However, in Lemma 4.5 we will show that even in the 'worst case scenario', $\bar{\mu}_{n,u}$ still increases by a sufficient amount.

Lemma 4.4. *Given n and u such that $u = \frac{n}{\gamma+O(1)}$,*

$$\frac{\mu_{n,u+1}}{\mu_{n,u}} \gtrsim b^{\frac{n^2}{2u(u+1)}-\frac{n}{2u}}.$$

Proof. First note that

$$P_{n,u+1} \geqslant P_{n,u}. \tag{4.10}$$

To see that this is true, note that the product $\left\lceil \frac{n}{u} \right\rceil !^{u_{\mathrm{L}}} \left\lfloor \frac{n}{u} \right\rfloor !^{u_{\mathrm{S}}}$ contains exactly n factors, and increasing u by 1 can only decrease those n factors.

Now let

$$x = \frac{n}{u} - \frac{n}{u+1} = \frac{n}{u(u+1)}.$$

If $\left\lfloor \frac{n}{u} \right\rfloor = \left\lfloor \frac{n}{u+1} \right\rfloor$, then $\delta_{n,u+1} = \delta_{n,u} - x$. Otherwise, $\left\lfloor \frac{n}{u} \right\rfloor = \left\lfloor \frac{n}{u+1} \right\rfloor + 1$ and $\delta_{n,u} + (1 - \delta_{n,u+1}) = x$. In both cases,

$$|\delta_{n,u+1}(1-\delta_{n,u+1}) - \delta_{n,u}(1-\delta_{n,u})| \leqslant x.$$

Therefore, by (4.5),

$$f_{n,u} - f_{n,u+1} \geqslant \frac{n^2}{2u(u+1)} - \frac{x}{2}(u+1) = \frac{n^2}{2u(u+1)} - \frac{n}{2u}. \tag{4.11}$$

Furthermore,

$$|f_{n,u} - f_{n,u+1}| = \frac{n^2}{2u(u+1)} + O\left(\frac{n}{u}\right) = O\left(\log^2 n\right),$$

so as $f_{n,u} \sim n \log_b n$,

$$\exp\left(-\frac{f_{n,u+1}^2 p}{2qN} + \frac{f_{n,u}^2 p}{2qN}\right) = \exp\left(o(1)\right) \sim 1.$$

Together with (4.6), (4.10), and (4.11), this completes the proof. □

Lemma 4.5. *Given n and u such that $u = \frac{n}{\gamma + O(1)}$,*

$$\frac{\bar{\mu}_{n,u+1}}{\bar{\mu}_{n,u}} \geqslant \exp\left(\Theta\left(\log n \log\log n\right)\right).$$

Proof. Let $u'_{\mathrm{L}} = \delta_{n,u+1}(u+1)$ and $u'_{\mathrm{S}} = (1-\delta_{n,u+1})(u+1)$, then

$$u'_{\mathrm{L}} + u'_{\mathrm{S}} = u + 1 = u_{\mathrm{L}} + u_{\mathrm{S}} + 1.$$

If $u_{\mathrm{L}} > \left\lfloor \frac{n}{u} \right\rfloor$, then given a u-equipartition of n vertices, we can form a $(u+1)$-equipartition by removing one vertex from $\left\lfloor \frac{n}{u} \right\rfloor$ parts of size $\left\lceil \frac{n}{u} \right\rceil$ and forming a new part of size $\left\lfloor \frac{n}{u} \right\rfloor$ from the removed vertices. In this case, $u'_{\mathrm{L}} = u_{\mathrm{L}} - \left\lfloor \frac{n}{u} \right\rfloor$ and $u'_{\mathrm{S}} = u_{\mathrm{S}} + \left\lfloor \frac{n}{u} \right\rfloor + 1$, and so

$$\frac{u'_{\mathrm{L}}!u'_{\mathrm{S}}!}{u_{\mathrm{L}}!u_{\mathrm{S}}!} = \frac{\prod_{t=1}^{\left\lfloor \frac{n}{u} \right\rfloor + 1} (u_{\mathrm{S}} + t)}{\prod_{t=0}^{\left\lfloor \frac{n}{u} \right\rfloor - 1} (u_{\mathrm{L}} - t)} \leqslant \frac{(u+1)^{\left\lfloor \frac{n}{u} \right\rfloor + 1}}{\left\lfloor \frac{n}{u} \right\rfloor !}. \tag{4.12}$$

Otherwise, if $u_{\mathrm{L}} \leqslant \left\lfloor \frac{n}{u} \right\rfloor$, then starting with a u-equipartition, we can form a $(u+1)$-equipartition by removing one vertex from each of the u_{L} parts of size $\left\lceil \frac{n}{u} \right\rceil$ and from $\left\lfloor \frac{n}{u} \right\rfloor - u_{\mathrm{L}}$ parts of size $\left\lfloor \frac{n}{u} \right\rfloor$, and forming a new part of size $\left\lfloor \frac{n}{u} \right\rfloor$ from the removed vertices. In this case, $u'_{\mathrm{S}} = \left\lfloor \frac{n}{u} \right\rfloor - u_{\mathrm{L}}$ and $u'_{\mathrm{L}} = u + 1 - \left\lfloor \frac{n}{u} \right\rfloor + u_{\mathrm{L}}$. Note that for all integers $1 \leqslant x_1 \leqslant x_2 \leqslant x_3 \leqslant x_4$ with $x_1 + x_4 = x_2 + x_3$, we have $x_1!x_4! \geqslant x_2!x_3!$. Therefore, if $u'_{\mathrm{S}} \geqslant u_{\mathrm{L}}$, then

$$u'_{\mathrm{L}}!u'_{\mathrm{S}}! \leqslant u'_{\mathrm{S}}!\,(u - u'_{\mathrm{S}})!(u+1) \leqslant u_{\mathrm{L}}!u_{\mathrm{S}}!(u+1). \tag{4.13}$$

Otherwise, if $u'_{\mathrm{S}} = \left\lfloor \frac{n}{u} \right\rfloor - u_{\mathrm{L}} < u_{\mathrm{L}} \leqslant \left\lfloor \frac{n}{u} \right\rfloor$, then

$$\frac{u'_{\mathrm{L}}!u'_{\mathrm{S}}!}{u_{\mathrm{L}}!u_{\mathrm{S}}!} = \frac{u'_{\mathrm{L}}!/u_{\mathrm{S}}!}{u_{\mathrm{L}}!/u'_{\mathrm{S}}!} \leqslant \frac{(u+1)^{u'_{\mathrm{L}} - u_{\mathrm{S}}}}{(u_{\mathrm{L}} - u'_{\mathrm{S}})!} = \frac{(u+1)^{1 + u_{\mathrm{L}} - u'_{\mathrm{S}}}}{(u_{\mathrm{L}} - u'_{\mathrm{S}})!} \leqslant \frac{(u+1)^{\left\lfloor \frac{n}{u} \right\rfloor + 1}}{\left\lfloor \frac{n}{u} \right\rfloor !}.$$

Comparing this to (4.12) and (4.13), we can see that in every case,

$$\frac{u'_{\mathrm{L}}!u'_{\mathrm{S}}!}{u_{\mathrm{L}}!u_{\mathrm{S}}!} \leqslant \frac{(u+1)^{\lfloor \frac{n}{u} \rfloor + 1}}{\lfloor \frac{n}{u} \rfloor !} \leqslant \frac{e^{\lfloor \frac{n}{u} \rfloor}(u+1)^{\lfloor \frac{n}{u} \rfloor + 1}}{\lfloor \frac{n}{u} \rfloor^{\lfloor \frac{n}{u} \rfloor}} \leqslant \left(\frac{e(u+1)}{\frac{n}{u} - 1} \right)^{\lfloor \frac{n}{u} \rfloor} (u+1)$$
$$\leqslant \left(\frac{e(u+1)u}{n-u} \right)^{n/u} (u+1).$$

Together with Lemma 4.4, and since $\frac{n}{u} = \gamma + O(1) = \Theta(\log n)$, this gives

$$\frac{\bar{\mu}_{n,u+1}}{\bar{\mu}_{n,u}} \gtrsim b^{\frac{n^2}{2u(u+1)} - \frac{n}{2u}} \left(\frac{n-u}{(u+1)u} \right)^{n/u} n^{O(1)} = b^{\frac{\gamma n}{2(u+1)}} \left(\frac{n}{u^2} \right)^{\gamma} n^{O(1)}$$
$$= \left(\frac{n}{2 \log_b n} \right)^{\frac{n}{u+1}} \left(\frac{n}{u^2} \right)^{\gamma} n^{O(1)} = \left(\frac{n^2}{u^2 \log_b n} \right)^{\gamma} n^{O(1)} = \left(\Theta \left(\log n \right) \right)^{\gamma} n^{O(1)}$$
$$= \exp \left(\Theta \left(\log n \log \log n \right) \right).$$

□

We will also need the following lemma which examines how much $\mu_{n,u}$ increases if n is increased by 1.

Lemma 4.6. *Given n and u such that $u = \frac{n}{\gamma + O(1)}$,*

$$\frac{\mu_{n+1,u}}{\mu_{n,u}} = \Theta \left(\frac{\log n}{n} \right).$$

Proof. Given a u-equipartition of n vertices, adding a vertex to a part of size $\lfloor \frac{n}{u} \rfloor$ yields a u-equipartition of $n+1$ vertices, so

$$f_{n+1,u} = f_{n,u} + \left\lfloor \frac{n}{u} \right\rfloor. \tag{4.14}$$

Therefore, since $\lfloor \frac{n}{u} \rfloor = \gamma(n) + O(1)$,

$$q^{f_{n+1,u} - f_{n,u}} = \Theta(1) \left(\frac{2 \log_b n}{n} \right)^2. \tag{4.15}$$

Furthermore, as $f_{n,u} = O(n \log n)$ and from (4.14),

$$-\frac{f_{n+1,u}^2 p}{2qN} + \frac{f_{n,u}^2 p}{2qN} = O \left(\frac{\log^2 n}{n} \right) = o(1). \tag{4.16}$$

Finally, note that by (4.3),

$$\frac{P_{n+1,u}}{P_{n,u}} \sim \frac{n}{2 \log_b n},$$

since the factorial is multiplied by $n+1 \sim n$ and the product in the denominator is multiplied by exactly one factor $\lfloor \frac{n}{u} \rfloor + 1 \sim \gamma \sim 2 \log_b n$ if n is increased to $n+1$.

Together with (4.6), (4.15) and (4.16), this completes the proof. □

4.5 Choice of the subsequence

In the next proposition, we choose an appropriate subsequence of the integers for Theorem 4.2.

Proposition 4.7. *There is a constant $j_0 \in \mathbb{N}$ and a strictly increasing sequence $(n_j)_{j\geqslant 1}$ such that for all $j \geqslant j_0$,*

a) $u_j := \frac{n_j}{j} \in \mathbb{Z}$.

b) $\gamma_j = j + o(1)$ *as* $j \to \infty$, *where* $\gamma_j = 2\log_b n_j - 2\log_b \log_b n_j - 2\log_b 2$.

c) $\bar{\mu}_{n_j,u_j} \to \infty$ *as* $j \to \infty$.

d) Let $G \sim \mathcal{G}(n_j, m_j)$ *with* $m_j = \left\lfloor p\binom{n_j}{2}\right\rfloor$, *then whp as* $j \to \infty$, *for all* $u \leqslant u_j - 1$, *G has no equitable u-colouring.*

Proof. It suffices to show that there is a strictly increasing sequence $(n_j)_{j\geqslant j_0}$ which fulfils a)–d) for some large enough j_0: given such a sequence, by b) n_j grows exponentially in j, so without loss of generality we can assume that $n_{j_0} \geqslant j_0$ and let $n_j = j$ for $1 \leqslant j < j_0$.

From Lemma 4.3, we can make the following two easy observations.

Observation 1. *Given integers n and u such that* $\left|\frac{n}{u} - \gamma(n)\right| \leqslant 10$, *if* $\bar{\mu}_{n,u} \geqslant 1$, *then*

$$u \geqslant \frac{n}{\gamma(n) + o_n(1)}.$$

Observation 2. *Given integers n and u such that* $\left|\frac{n}{u} - \gamma(n)\right| \leqslant 10$, *if* $\bar{\mu}_{n,u} \leqslant n$, *then*

$$u \leqslant \frac{n}{\gamma(n) + o_n(1)}.$$

Now for large enough $j \in \mathbb{N}$, let n_j be the smallest multiple of j such that, letting $u_j = n_j/j$,

(1) $\gamma(n_j) \in [j - 10, j + 10]$

(2) $\bar{\mu}_{n_j,u_j} \geqslant \log j$.

Claim 1. *If j is large enough, then n_j is well-defined.*

Proof. Fix j and consider the sequence $(g_t)_{t \geqslant j}$ with $g_t = \gamma(tj)$. If j is large enough and $t \geqslant j$, then

$$0 < g_{t+1} - g_t \leqslant 3\left(\log_b\left((t+1)j\right) - \log_b\left(tj\right)\right) = 3\log_b\left(1 + \frac{1}{t}\right) < \frac{1}{10}. \tag{4.17}$$

Furthermore, if j is large enough, then $g_j \leqslant j$ and $g_t \to \infty$ as $t \to \infty$. Therefore, if j is large enough, there is an integer $t_0 \geqslant j$ such that

$$g_{t_0} \in [j + 0.1, j + 0.2]\,.$$

By Lemma 4.3, letting $n' = t_0 j$,

$$\bar{\mu}_{n',t_0} \geqslant b^{(0.05+o(1))n'} \geqslant \log n' \geqslant \log j.$$

Therefore, n' is a multiple of j such that the two conditions (1) and (2) from the definition of n_j are fulfilled, so n_j is well-defined. □

The definition of n_j immediately implies a) and c). We now show that b) holds.

Claim 2. *As $j \to \infty$, $\gamma_j = j + o(1)$.*

Proof. By definition,

$$\bar{\mu}_{n_j,u_j} \geqslant \log j \geqslant 1,$$

so Observation 1 gives

$$u_j = \frac{n_j}{j} \geqslant \frac{n_j}{\gamma(n_j) + o(1)}.$$

In particular, $j \leqslant \gamma(n_j) + o(1) = \gamma_j + o(1) \leqslant 2\log_b n_j$ for j large enough (note that $n_j \geqslant j$ by definition). By (4.17), for j large enough,

$$\gamma(n_j) \geqslant \gamma(n_j - j) \geqslant \gamma(n_j) - \frac{1}{10} \geqslant j - \frac{1}{20},$$

so in particular

$$\gamma(n_j - j) \in [j - 10, j + 10]. \tag{4.18}$$

By the minimality of n_j in its definition, we must therefore have

$$\bar{\mu}_{n_j-j,u_j-1} < \log j \leqslant \log\left(2\log_b n_j\right) \leqslant n_j - j. \tag{4.19}$$

By Observation 2,

$$u_j - 1 = \frac{n_j - j}{j} \leqslant \frac{n_j - j}{\gamma(n_j - j) + o(1)},$$

so $j \geqslant \gamma(n_j - j) + o(1)$. By (4.17),

$$\gamma(n_j - j) = \gamma(n_j) + O\left(\log\left(1 + \frac{1}{(n_j - j)/j}\right)\right) = \gamma(n_j) + o(1).$$

Therefore, $\gamma(n_j) \leqslant j + o(1)$. □

Since $\gamma(n)$ is strictly increasing for large enough n, it follows that n_j is strictly increasing if j is large enough. It only remains to show d).

Claim 3. *If j is large enough, then*

$$\bar{\mu}_{n_j,u_j-1} \leqslant \frac{1}{j}.$$

Proof. By (4.19),

$$\bar{\mu}_{n_j-j,u_j-1} = \frac{\mu_{n_j-j,u_j-1}}{(u_j-1)!} < \log j. \tag{4.20}$$

Note that by Lemma 4.6,

$$\frac{\mu_{n_j,u_j-1}}{\mu_{n_j-j,u_j-1}} \leqslant \left(\Theta\left(\frac{\log n_j}{n_j}\right)\right)^j. \tag{4.21}$$

Since $u_j = \frac{n_j}{j}$, an equitable partition of n_j vertices into $u_j - 1$ parts consists of exactly j larger parts of size $j+1$ and $u_j - 1 - j$ smaller parts of size j. Hence,

$$\bar{\mu}_{n_j,u_j-1} = \frac{\mu_{n_j,u_j-1}}{j!(u_j-1-j)!}.$$

Together with (4.20) and (4.21) and the facts that $j! \geqslant j^j/e^j$, $j = \Theta(\log n_j)$ and $u_j = \Theta\left(n_j/\log n_j\right)$, this gives

$$\begin{aligned}\bar{\mu}_{n_j,u_j-1} &\leqslant \log j\left(\Theta\left(\frac{\log n_j}{n_j}\right)\right)^j \frac{(u_j-1)!}{j!(u_j-1-j)!} \leqslant \log j\left(\Theta\left(\frac{\log n_j}{n_j}\right)\right)^j \frac{(u_j-1)^j}{j!}\\ &\leqslant \log j\left(\Theta\left(\frac{u_j\log n_j}{jn_j}\right)\right)^j = \log j\left(\Theta\left(1/j\right)\right)^j < 1/j\end{aligned}$$

if j is large enough. □

To show d), note that whp, $\chi(G) = \frac{n}{\gamma(n)+o(1)}$ for $G \sim \mathcal{G}(n, m(n))$, as it is for $G \sim \mathcal{G}(n,p)$ with $p < 1-1/e^2$ by Theorem 3.1. This follows from Theorem 2.1 since for $p' = p+O(1/n)$, $\gamma_{p'}(n) = \gamma_p(n)+o(1)$; and because the properties of having and of not having a colouring with a certain given number of colours is monotone.

So whp $G \sim \mathcal{G}(n_j, m_j)$ has no general colouring and therefore no equitable colouring with less than $\frac{n_j}{\gamma_j+o(1)}$ colours. It only remains to prove that whp G has no equitable colouring with at least $\frac{n_j}{\gamma_j+o(1)}$ colours (where the $o(1)$-term is an arbitrary function that tends to 0) but at most $u_j - 1$ colours (of course $u_j - 1$ is of the form $\frac{n_j}{\gamma_j+o(1)}$ as well). By Lemma 4.5 and Claim 3, the expected number of (unordered) partitions which induce such a colouring is

$$\sum_{\frac{n}{\gamma_j+o(1)} \leqslant u \leqslant u_j-1} \bar{\mu}_{n_j,u} = O\left(\bar{\mu}_{n_j,u_j-1}\right) = O\left(1/j\right) = o(1),$$

so whp no equitable colouring with at most $u_j - 1$ colours exists. □

4.6 The second moment

For the proof of Theorem 4.2, it remains to show that for the sequence $(n_j)_{j\geqslant 1}$ from Proposition 4.7,

$$\mathbb{E}[X_{n_j,u_j}^2]/\mathbb{E}[X_{n_j,u_j}]^2 \to 1 \text{ as } j \to \infty.$$

We always have $\mathbb{E}[X_{n_j,u_j}^2] \geqslant \mathbb{E}[X_{n_j,u_j}]^2$ (this is true for any random variable), so it suffices to prove that

$$\mathbb{E}[X_{n_j,u_j}^2]/\mathbb{E}[X_{n_j,u_j}]^2 \leqslant 1 + o(1) \text{ as } j \to \infty. \tag{4.22}$$

The second moment calculations are similar to those in the proof of Theorem 3.1.

Let $j \geqslant j_0$, $N_j = \binom{n_j}{2}$, $m_j = \lfloor pN_j \rfloor$ and $f_j = f_{n_j,u_j}$ as in (4.5) and $P_j = P_{n_j,u_j}$ as in (4.3), and $G \sim \mathcal{G}(n_j, m_j)$. To simplify notation, for the rest of this section we will omit the indices of n_j, m_j, u_j and so on when the context is clear.

Note that since $j = \frac{n}{u}$ is an integer,

$$P = \frac{n!}{j!^u} \quad \text{and} \quad f = \frac{n(j-1)}{2}. \tag{4.23}$$

As in the last chapter in (3.12),

$$\mathbb{E}[X_{n,u}^2] = \sum_{\pi_1,\pi_2 \text{ ordered } u\text{-equipartitions}} \mathbb{P}\left(\text{both } \pi_1 \text{ and } \pi_2 \text{ induce proper colourings}\right).$$

We again quantify how similar two partitions π_1 and π_2 are to each other by considering their *overlap sequence* $\mathbf{r} = \mathbf{r}(\pi_1, \pi_2) = (r_i)_{i=2}^{j}$. As before, denote by r_i the number of pairs of parts which intersect in exactly i vertices, and call an intersection of size at least 2 between two parts an *overlap block*.

Again let $P_{\mathbf{r}}$ denote the number of ordered pairs π_1, π_2 with overlap sequence $\mathbf{r}$, and

$$\begin{aligned} v = v(\mathbf{r}) &= \sum_{i=2}^{j} i r_i \\ \rho &= v/n \\ d = d(\mathbf{r}) &= \sum_{i=2}^{j} r_i \binom{i}{2}. \end{aligned} \tag{4.24}$$

Two given ordered u-equipartitions π_1 and π_2 with overlap sequence $\mathbf{r}$ both induce valid colourings at the same time if and only if exactly $2f - d(\mathbf{r})$ given forbidden edges are not present in G, so by (4.6),

$$\mathbb{E}[X_{n,u}^2] = \sum_{\mathbf{r}} P_{\mathbf{r}} \cdot \frac{\binom{N-2f+d(\mathbf{r})}{m}}{\binom{N}{m}} = \mu_{n,u}^2 \sum_{\mathbf{r}} \frac{P_{\mathbf{r}}}{P^2} \cdot \frac{\binom{N-2f+d}{m}\binom{N}{m}}{\binom{N-f}{m}^2}.$$

Let

$$\begin{aligned} Q_{\mathbf{r}} &= \frac{P_{\mathbf{r}}}{P^2} \\ S_{\mathbf{r}} &= \frac{\binom{N-2f+d(\mathbf{r})}{m}\binom{N}{m}}{\binom{N-f}{m}^2}, \end{aligned} \tag{4.25}$$

then to prove (4.22), we need to show that as $j \to \infty$,

$$\sum_{\mathbf{r}} Q_{\mathbf{r}} S_{\mathbf{r}} \leqslant 1 + o(1). \tag{4.26}$$

In Section 4.6.1, we will determine the asymptotic value of $S_{\mathbf{r}}$. In Section 4.6.2, we will show that the contribution to the sum (4.26) from the typical overlap range is at most $1 + o(1)$, similarly to how this was done in Section 3.5.2 of the last chapter. In Sections 4.6.3 and 4.6.4, we will discuss how the remaining cases follow directly from simplifications of the arguments in Sections 3.5.3 and 3.5.4. The contribution from these cases to the sum (4.26) is $o(1)$, concluding the proof of Theorem 4.2.

4.6.1 Asymptotics of $S_{\mathbf{r}}$

Consider an overlap sequence $\mathbf{r}$. Again letting $\varepsilon = Np - m \in [0,1]$, and $d = d(\mathbf{r})$,

$$\begin{aligned} S_{\mathbf{r}} &= \frac{(N-2f+d)!\, N!\, (N-m-f)!^2}{(N-f)!^2\, (N-m)!\, (N-m-2f+d)!} \\ &= \frac{(N-2f+d)!\, N!\, (qN+\varepsilon-f)!^2}{(N-f)!^2\, (qN+\varepsilon)!\, (qN+\varepsilon-2f+d)!}. \end{aligned}$$

Since $d \leqslant f = O(n \log n) = o(N)$, applying Stirling's formula $n! \sim \sqrt{2\pi n}\, n^n/e^n$ gives

$$\begin{aligned} S_{\mathbf{r}} &\sim \frac{(N-2f+d)^{N-2f+d} N^N\, (qN+\varepsilon-f)^{2qN+2\varepsilon-2f}}{(N-f)^{2N-2f}(qN+\varepsilon)^{qN+\varepsilon}\, (qN+\varepsilon-2f+d)^{qN+\varepsilon-2f+d}} \\ &= q^{-d} \cdot \frac{\left(1-\frac{2f-d}{N}\right)^{N-2f+d} \left(1-\frac{f-\varepsilon}{qN}\right)^{2qN+2\varepsilon-2f}}{\left(1-\frac{f}{N}\right)^{2N-2f} (1+\frac{\varepsilon}{qN})^{qN+\varepsilon} \left(1-\frac{2f-d-\varepsilon}{qN}\right)^{qN+\varepsilon-2f+d}} \\ &\sim b^d \cdot \frac{\left(1-\frac{2f-d}{N}\right)^{N-2f+d} \left(1-\frac{f-\varepsilon}{qN}\right)^{2qN-2f}}{\left(1-\frac{f}{N}\right)^{2N-2f} (1+\frac{\varepsilon}{qN})^{qN} \left(1-\frac{2f-d-\varepsilon}{qN}\right)^{qN-2f+d}}. \end{aligned}$$

Using $\log(1+x) = x - \frac{x^2}{2} + O(x^3)$ for $x \to 0$, and as $d \leqslant f = O(n \log n)$, we get

$$\begin{aligned} \log S_{\mathbf{r}} &= d \log b + \frac{d^2}{2N} + \frac{f^2}{N} - \frac{d^2}{2qN} - \frac{f^2}{qN} - \frac{2df}{N} + \frac{2df}{qN} + O\left(\frac{f^3}{N^2}\right) \\ &= d \log b + \frac{-p(d^2+2f^2-4df)}{2qN} + o(1). \end{aligned}$$

Therefore,

$$S_{\mathbf{r}} \sim b^d \exp\left(-\frac{p(d^2+2f^2-4df)}{2qN}\right), \tag{4.27}$$

where $d = d(\mathbf{r})$ is given in (4.24).

4.6.2 The typical overlap case

Let

$$c = \frac{1}{2}\left(1 - \frac{\log b}{2}\right), \tag{4.28}$$

then $c \in (0,1)$ since $p < 1 - 1/e^2$. Similarly as in the proof of Theorem 3.1, let

$$\mathcal{R}_1 = \{\mathbf{r} \mid \rho = \rho(\mathbf{r}) \leqslant c\}.$$

In this section we will show how to amend the arguments from Section 3.5.2 to prove that

$$\sum_{\mathbf{r}\in\mathcal{R}_1} Q_{\mathbf{r}} S_{\mathbf{r}} \leqslant 1 + o(1). \tag{4.29}$$

We can proceed analogously to Section 3.5.2 to bound $Q_{\mathbf{r}}$ as in (3.26). However, we make one refinement to our argument: in (3.24), we simply bounded

$$\exp\left(-\frac{\sum_{x=1}^{u} s_x(s_x-1)\sum_{y=1}^{u} t_y(t_y-1)}{2(n-v)^2} + O\left(\frac{\log^4 n}{n}\right)\right) \lesssim 1.$$

Instead, as $\sum_{x=1}^{u} s_x = n - v$, applying Jensen's inequality with the convex function $x(x-1)$ gives

$$\sum_{x=1}^{u} s_x(s_x-1) \geqslant u \cdot \left(\frac{n-v}{u}\right)\left(\frac{n-v}{u}-1\right) = (n-v)\left(\frac{n-v}{u}-1\right),$$

and the corresponding inequality of course also holds for $\sum_{y=1}^{u} t_y(t_y-1)$. This yields the following refined version of (3.26), noting that $n/u = j$:

$$Q_{\mathbf{r}} \lesssim \exp\left(-\frac{1}{2}\left(\frac{n-v}{u}-1\right)^2\right)\prod_{i=2}^{j}\left(\frac{1}{r_i!}\left(\frac{e^{\rho i}u^2 j!^2}{n^i i!\,(j-i)!^2}\right)^{r_i}\right).$$

Recall that $d = \sum_{i=2}^{j}\binom{i}{2} r_i$. Together with (4.27), this gives

$$Q_{\mathbf{r}} S_{\mathbf{r}} \lesssim \exp\left(-\frac{1}{2}\left(\frac{n-v}{u}-1\right)^2 - \frac{p(d^2+2f^2-4df)}{2qN}\right) \cdot \prod_{i=2}^{j}\left(\frac{1}{r_i!}\left(\frac{e^{\rho i} b^{\binom{i}{2}} u^2 j!^2}{n^i i!\,(j-i)!^2}\right)^{r_i}\right).$$

Letting

$$\tilde{T}_i := \frac{e^{\rho i} b^{\binom{i}{2}} u^2 j!^2}{n^i i!\,(j-i)!^2},$$

we have

$$Q_{\mathbf{r}} S_{\mathbf{r}} \lesssim \exp\left(-\frac{1}{2}\left(\frac{n-v}{u}-1\right)^2 - \frac{p(d^2+2f^2-4df)}{2qN}\right)\prod_{i=2}^{j}\frac{\tilde{T}_i^{r_i}}{r_i!}. \tag{4.30}$$

Recalling that $p < 1 - 1/e^2$ and therefore $\log b < 2$, let

$$\tilde{c} = \min\left(\frac{1}{2}\left(\frac{1}{\log b}-\frac{1}{2}\right), \frac{1}{2}\right) \in (0,1).$$

Lemma 4.8. *If j (and thereby $n = n_j$) is large enough and $\rho \in \mathcal{R}_1$, then for all $3 \leqslant i \leqslant j$,*

$$\tilde{T}_i \leqslant n^{-\tilde{c}}.$$

Proof. The proof is similar to the proof of Lemma 3.10. We assume throughout that j (and $n = n_j$) is large enough for all bounds to hold. We first need to check that the assertion holds for $i = 3$ and $i = j$. For this, note that

$$\tilde{T}_3 \leqslant \frac{e^3 b^3 u^2 j^6}{n^3 3!} = n^{-1+o(1)} \leqslant n^{-\tilde{c}}. \tag{4.31}$$

Since $j = \gamma_j + o(1) = 2\log_b n - 2\log_b \log_b n - 2\log_b 2 + o(1)$,

$$b^{\binom{j}{2}} = \left(\frac{n}{2\log_b n}\right)^{j-1} n^{o(1)},$$

so with Stirling's formula and as $u = n/j = n^{1+o(1)}$ and $j \sim 2\log_b n$,

$$\tilde{T}_j = \frac{e^{\rho j} b^{\binom{j}{2}} u^2 j!}{n^j} = n^{o(1)} \frac{e^{\rho j} n j^j}{(2\log_b n)^{j-1} e^j} = n^{1+o(1)} e^{-j+\rho j}.$$

As $e^j = n^{\frac{2}{\log b}+o(1)}$, and $\rho \leqslant c = \frac{1}{2}\left(1 - \frac{\log b}{2}\right)$ since $\mathbf{r} \in \mathcal{R}_1$,

$$\tilde{T}_j \leqslant n^{1-(1-\rho)\frac{2}{\log b}+o(1)} = n^{\frac{1}{2}-\frac{1}{\log b}+o(1)} \leqslant n^{-\tilde{c}}. \tag{4.32}$$

Now as in (3.28),

$$\frac{\tilde{T}_{i+1}}{\tilde{T}_i} = \frac{e^{\rho} b^i (j-i)^2}{n(i+1)}.$$

In particular, for all $i \leqslant 0.8\log_b n$,

$$\frac{\tilde{T}_{i+1}}{\tilde{T}_i} \leqslant n^{-0.2+o(1)} \leqslant 1,$$

so by (4.31), for all $3 \leqslant i \leqslant 0.8\log_b n$,

$$\tilde{T}_i \leqslant \tilde{T}_3 \leqslant n^{-\tilde{c}}.$$

Furthermore, for $i \geqslant 1.2\log_b n$,

$$\frac{\tilde{T}_{i+1}}{\tilde{T}_i} \geqslant n^{0.2+o(1)} \geqslant 1,$$

so by (4.32), for all $1.2\log_b n \leqslant i \leqslant j$,

$$\tilde{T}_i \leqslant \tilde{T}_j \leqslant n^{-\tilde{c}}.$$

For the remaining case $0.8\log_b < i < 1.2\log_b n$, note that as $j \leqslant 2\log_b n$,

$$\tilde{T}_i \leqslant \frac{e^i b^{\frac{i^2}{2}} n^2 j^{2i}}{n^i} \leqslant n^{O(1)} \frac{n^{0.6i} j^{2i}}{n^i} \leqslant n^{O(1)-0.3i} = n^{-\Theta(\log n)} \leqslant n^{-\tilde{c}}.$$

□

Let

$$R_3 = \sum_{i=3}^{j} r_i,$$

then by Lemma 4.8, (4.30) and the definition (4.23) of f,

$$\begin{aligned}
Q_{\mathbf{r}} S_{\mathbf{r}} &\lesssim \exp\left(-\frac{1}{2}\left(\frac{n-v}{u}-1\right)^2 - \frac{p(d^2+2f^2-4df)}{2qN}\right)\frac{\tilde{T}_2^{r_2}}{r_2!} n^{-\tilde{c}R_3} \\
&\leqslant \exp\left(-\frac{1}{2}\left(\frac{n-v}{u}-1\right)^2 - \frac{p(f^2-2df)}{qN}\right)\frac{\tilde{T}_2^{r_2}}{r_2!} n^{-\tilde{c}R_3} \\
&\sim \exp\left(-\frac{1}{2}\left(\frac{n-v}{u}-1\right)^2 - \frac{2pf^2}{qn^2} + \frac{4pdf}{qn^2}\right)\frac{\tilde{T}_2^{r_2}}{r_2!} n^{-\tilde{c}R_3} \\
&\sim \exp\left(-\frac{1}{2}\left(\frac{n-v}{u}-1\right)^2 - \frac{p(j-1)^2}{2q} + \frac{2pd(j-1)}{qn}\right)\frac{\tilde{T}_2^{r_2}}{r_2!} n^{-\tilde{c}R_3}. \qquad (4.33)
\end{aligned}$$

We are almost ready to sum (4.33) over $\mathbf{r} \in \mathcal{R}_1$, but first we need to make a simple observation and then handle the cases where either v or d are large.

Lemma 4.9. *Given R_3, there are at most $(2e\log_b n)^{R_3}$ ways to select $r_3, \ldots, r_j$ such that $\sum_{i=3}^{j} r_i = R_3$.*

Proof. Since $j \leqslant 2\log_b n$, there are at most

$$\binom{R_3+j-3}{R_3} \leqslant \left(\frac{e\,(R_3+j-3)}{R_3}\right)^{R_3} \leqslant (e\,(1+j-3))^{R_3} \leqslant (2e\log_b n)^{R_3}$$

ways to write R_3 as an ordered sum of $j-2$ nonnegative summands. □

Lemma 4.10. *Let $\mathcal{R}_1^{\text{ex}}$ be the set of all $\mathbf{r} \in \mathcal{R}_1$ with $v = v(\mathbf{r}) \geqslant \frac{n}{\log^3 n}$ or $d = d(\mathbf{r}) \geqslant \frac{n}{\log^3 n}$, then*

$$\sum_{\mathbf{r}\in\mathcal{R}_1^{\text{ex}}} Q_{\mathbf{r}} S_{\mathbf{r}} = o(1).$$

Proof. Again we assume throughout that j is large enough for all estimates to be valid. Let $\mathbf{r} \in \mathcal{R}_1^{\text{ex}}$. We first note that as $d \leqslant f = O(n \log n)$, $j = \Theta(\log n)$ and $u = \Theta(n/\log n)$,

$$\exp\left(-\frac{1}{2}\left(\frac{n-v}{u}-1\right)^2 - \frac{p(j-1)^2}{2q} + \frac{2pd(j-1)}{qn}\right) = \exp\left(O(\log^2 n)\right), \quad (4.34)$$

and

$$\tilde{T}_2 \leqslant \frac{e^2 b^{\binom{2}{2}} u^2 j^4}{n^2 2!} = \Theta(\log^2 n). \quad (4.35)$$

Since $v = \sum_{i=2}^{j} i r_i \leqslant 2r_2 + 2R_3 \log_b n$ and $d = \sum_{i=2}^{j} \binom{i}{2} r_i \leqslant r_2 + 2R_3 \log_b^2 n$, if $\mathbf{r} \in \mathcal{R}_1^{\text{ex}}$, then either $r_2 \geqslant n/\log^6 n$ or $R_3 \geqslant n/\log^6 n$.

Case 1: $r_2 \geqslant n/\log^6 n$.

Then from (4.33), (4.34), and (4.35), and as $r_2! \geqslant {r_2}^{r_2}/e^{r_2}$,

$$\begin{aligned} Q_{\mathbf{r}} S_{\mathbf{r}} &\lesssim \exp\left(O(\log^2 n)\right) \frac{\left(\Theta\left(\log^2 n\right)\right)^{r_2}}{r_2!} n^{-\tilde{c}R_3} \\ &\leqslant \exp\left(O(\log^2 n)\right) \left(\frac{\Theta\left(\log^2 n\right)}{r_2}\right)^{r_2} n^{-\tilde{c}R_3} \\ &\lesssim \exp\left(O(\log^2 n)\right) \left(\frac{\log^9 n}{n}\right)^{r_2} n^{-\tilde{c}R_3}. \end{aligned}$$

With Lemma 4.9, summing over r_2 and R_3 gives

$$\begin{aligned} \sum_{\substack{\mathbf{r} \in \mathcal{R}_1^{\text{ex}} \\ r_2 \geqslant n/\log^6 n}} Q_{\mathbf{r}} S_{\mathbf{r}} &\leqslant \exp\left(O(\log^2 n)\right) \sum_{r_2 \geqslant n/\log^6 n,\ R_3} \left(\left(\frac{\log^9 n}{n}\right)^{r_2} \left(\frac{2e \log_b n}{n^{\tilde{c}}}\right)^{R_3}\right) \\ &= o(1). \end{aligned} \quad (4.36)$$

Case 2: $R_3 \geqslant n/\log^6 n$.

By Lemma 4.9, (4.34), and (4.35),

$$\begin{aligned} \sum_{\substack{\mathbf{r} \in \mathcal{R}_1^{\text{ex}} \\ R_3 \geqslant n/\log^6 n}} Q_{\mathbf{r}} S_{\mathbf{r}} &\leqslant \exp\left(O(\log^2 n)\right) \sum_{r_2,\ R_3 \geqslant n/\log^6 n} \left(\frac{\tilde{T}_2^{r_2}}{r_2!} \left(\frac{2e \log_b n}{n^{\tilde{c}}}\right)^{R_3}\right) \\ &= \exp\left(O(\log^2 n)\right) \left(\frac{2e \log_b n}{n^{\tilde{c}}}\right)^{\frac{n}{\log^6 n}} \sum_{t \geqslant 0} \left(\frac{2e \log_b n}{n^{\tilde{c}}}\right)^{t} = o(1). \end{aligned} \quad (4.37)$$

The claim now follows from (4.36) and (4.37). □

We will now sum (4.33) for all $\mathbf{r} \in \mathcal{R}_1 \setminus \mathcal{R}_1^{\text{ex}}$. If $v < \frac{n}{\log^3 n}$ and $d < \frac{n}{\log^3 n}$, then

$$\begin{aligned}
&\exp\left(-\frac{1}{2}\left(\frac{n-v}{u}-1\right)^2 - \frac{p(j-1)^2}{2q} + \frac{2pd(j-1)}{qn}\right) \\
&\sim \exp\left(-\frac{1}{2}\left(\frac{n}{u}-1\right)^2 - \frac{p(j-1)^2}{2q}\right) = \exp\left(-(j-1)^2\left(\frac{1}{2}+\frac{p}{2q}\right)\right) \\
&= \exp\left(-\frac{b}{2}(j-1)^2\right). \qquad (4.38)
\end{aligned}$$

Furthermore, note that

$$\tilde{T}_2 = \frac{e^{2\rho} b^{\binom{2}{2}} u^2 j!^2}{n^2 2!\,(j-2)!^2} = \frac{e^{2\rho} b u^2 j^2 (j-1)^2}{2n^2} = e^{2\rho}\frac{b}{2}(j-1)^2.$$

For all $\mathbf{r} \in \mathcal{R}_1 \setminus \mathcal{R}_1^{\text{ex}}$, $\rho = v/n < \frac{1}{\log^3 n}$, so

$$\tilde{T}_2 \leqslant e^{\frac{2}{\log^3 n}}\,\frac{b}{2}(j-1)^2 =: T.$$

Therefore, from (4.33) and together with (4.38) and Lemma 4.9,

$$\begin{aligned}
\sum_{\mathbf{r}\in\mathcal{R}_1\setminus\mathcal{R}_1^{\text{ex}}} Q_{\mathbf{r}} S_{\mathbf{r}} &\lesssim \exp\left(-\frac{b}{2}(j-1)^2\right) \sum_{r_2, R_3 \geqslant 0} \frac{T^{r_2}}{r_2!}\left(2en^{-\tilde{c}}\log_b n\right)^{R_3} \\
&\sim \exp\left(-\frac{b}{2}(j-1)^2\right) \sum_{r_2\geqslant 0}\frac{T^{r_2}}{r_2!} = \exp\left(-\frac{b}{2}(j-1)^2 + T\right) \\
&= \exp\left(-\frac{b}{2}(j-1)^2\left(1 - e^{\frac{2}{\log^3 n}}\right)\right) = \exp\left(-\frac{b}{2}(j-1)^2 O\left(\log^{-3} n\right)\right) \\
&= \exp\left(-O\left(\log^2 n\right) O\left(\log^{-3} n\right)\right) = \exp(o(1)) = 1 + o(1).
\end{aligned}$$

Together with Lemma 4.10, this gives (4.29).

4.6.3 The intermediate overlap case

The case where two partitions have an intermediate degree of overlap cases follows directly from a simplification of the arguments in Section 3.5.3.

Given c from (4.28) and an arbitrary constant $c' \in (0, 1-c)$, define $\mathcal{R}_2^{c'}$ as in (3.34). Note that by (4.27), for all $\mathbf{r}$,

$$Q_{\mathbf{r}} S_{\mathbf{r}} = Q_{\mathbf{r}} b^{d(\mathbf{r})} \exp\left(O(\log^2 n)\right) \leqslant Q_{\mathbf{r}} b^{d(\mathbf{r})} \exp\left(o(n)\right).$$

Therefore, we can proceed as in Section 3.5.3 to bound the contribution from $\mathcal{R}_2^{c'}$ to the sum (4.26). As from (3.41), Lemma 3.14 and (3.45), we can show that

$$\begin{aligned}\sum_{\mathbf{r}\in\mathcal{R}_2^{c'}} Q_{\mathbf{r}}S_{\mathbf{r}} &\leqslant o(1) + \sum_{cn\leqslant v\leqslant(1-c')n;\ v_1\leqslant v;\ d_1,d_2\leqslant f} b^{n(1-\rho)\log_b(1-\rho)+\frac{v_1}{2}-\frac{\Delta v}{2}}\exp(o(n)) \\ &= o(1) + \sum_{cn\leqslant v\leqslant(1-c')n;\ v_1\leqslant v} b^{n(1-\rho)\log_b(1-\rho)+\frac{v_1}{2}-\frac{\Delta v}{2}}\exp(o(n)),\end{aligned}$$

where we took the sum over v, the number of vertices in overlap blocks between two given partitions, and v_2, the number of such vertices in parts of size larger than j in the first partition. However, in our case there are no such parts since all parts have exactly the same size j, and furthermore $\Delta = \gamma - \lfloor\gamma\rfloor = o(1)$. Therefore, this can be sharpened to

$$\sum_{\mathbf{r}\in\mathcal{R}_2^{c'}} Q_{\mathbf{r}}S_{\mathbf{r}} \leqslant o(1) + \sum_{cn\leqslant v\leqslant(1-c')n} b^{n(1-\rho)\log_b(1-\rho)}\exp(o(n)) \leqslant o(1) + nb^{tn},$$

where $t = \min\left((1-c)\log(1-c), c'\log c'\right) < 0$ for any constant $0 < c' < 1-c$. Therefore,

$$\sum_{\mathbf{r}\in\mathcal{R}_2^{c'}} Q_{\mathbf{r}}S_{\mathbf{r}} = o(1).$$

4.6.4 The high overlap case

Fix an arbitrary ordered u-equipartition π_1 and a constant $c' \in (0, 1-c)$ which will be determined later. As in the last chapter, for an overlap sequence $\mathbf{r}$, we denote by $P'_{\mathbf{r}}$ the number of ordered k-equipartitions with overlap $\mathbf{r}$ with π_1. Define $\mathcal{R}_3^{c'}$ and $\mathcal{P}_3$ as in (3.50) and (3.51).

By the definition (4.25) of $Q_{\mathbf{r}}$, the asymptotics (4.27) of $S_{\mathbf{r}}$ and (4.6) of the first moment $\mu_{n,u}$, and since $d = d(\mathbf{r}) \leqslant f$ for all $\mathbf{r}$, we have

$$\begin{aligned}
\sum_{\mathbf{r}\in\mathcal{R}_3^{c'}} Q_{\mathbf{r}}S_{\mathbf{r}} &\sim \sum_{\mathbf{r}\in\mathcal{R}_3^{c'}} \left(\frac{P'_{\mathbf{r}}}{P} b^{d(\mathbf{r})} \exp\left(-\frac{p(d(\mathbf{r})^2 + 2f^2 - 4d(\mathbf{r})f)}{2qN}\right)\right) \\
&\sim \frac{1}{\mu_{n,u}} \sum_{\mathbf{r}\in\mathcal{R}_3^{c'}} P'_{\mathbf{r}} b^{d-f} \exp\left(-\frac{p(d^2+3f^2-4df)}{2qN}\right) \\
&= \frac{1}{\mu_{n,u}} \sum_{\mathbf{r}\in\mathcal{R}_3^{c'}} P'_{\mathbf{r}} b^{d-f} \exp\left(-\frac{p(f-d)(3f-d)}{2qN}\right) \\
&\leqslant \frac{1}{\mu_{n,u}} \sum_{\mathbf{r}\in\mathcal{R}_3^{c'}} P'_{\mathbf{r}} b^{d-f} = \frac{1}{\mu_{n,u}} \sum_{\pi_2\in\mathcal{P}_3} b^{-(f-d(\pi_1,\pi_2)))},
\end{aligned}$$

where $d(\pi_1, \pi_2) := d(\mathbf{r})$ if $\mathbf{r}$ is the overlap sequence of π_1 and π_2.

Comparing this to (3.52), we see that we can proceed as in Section 3.5.4 to bound the contribution to the sum (4.26) from $\mathcal{R}_3^{c'}$. Recall that we could choose the constant c' in such a way that this contribution is bounded by

$$O\left(\frac{k_1!k_2!}{\mu_k} 2^k\right),$$

where k was the number of parts, k_1, k_2 were the numbers of parts of size $\left\lceil \frac{n}{k} \right\rceil$ and $\left\lfloor \frac{n}{k} \right\rfloor$, respectively, and μ_k denoted the expected number of ordered k-equipartitions which induce valid colourings. The factor 2^k came from varying part sizes: in the proof of Lemma 3.18, we bounded the number of ways in which the part sizes may change between $\left\lfloor \frac{n}{k} \right\rfloor$ and $\left\lceil \frac{n}{k} \right\rceil$ by 2^k. Since we picked k large enough so that $\mu_k/(k_1!k_2!)$ grows exponentially in n, the overall product is still $o(1)$.

In our case, all parts have exactly the same size j, so there is no need to account for varying part sizes. Therefore, the arguments from Section 3.5.4 give

$$\sum_{\mathbf{r}\in\mathcal{R}_3^{c'}} Q_{\mathbf{r}}S_{\mathbf{r}} = O\left(\frac{u!}{\mu_{n,u}}\right) = O\left(\frac{1}{\bar{\mu}_{n,u}}\right).$$

This expression is $o(1)$ as soon as $\bar{\mu}_{n,u} \to \infty$, which is indeed the case by our choice of $(n_j)_{j\geqslant 1}$ — see c) of Proposition 4.7. This completes the proof of Theorem 4.2. □

Chapter 5

The hitting time of rainbow connection number two

5.1 Background and results

The rainbow connection number was introduced in 2008 by Chartrand, Johns, McKeon and Zhang [12] as a way of quantifying the connectivity of a graph. Given an edge-coloured connected graph G, we call a path a *rainbow path* if all of its edges have distinct colours, and we call the colouring a *rainbow colouring* if every pair of vertices is joined by at least one rainbow path. The least number of colours where this is possible is called the *rainbow connection number* (or *rainbow connectivity*) $\mathrm{rc}(G)$ of the graph G. Rainbow connectivity has received considerable attention since its introduction, being both of theoretical interest and highly applicable (see for example the survey [38]).

Generally speaking, a low rainbow connection number indicates that a graph is well connected. At the extreme end, the complete graph K_n has rainbow connection number 1, while any tree on n vertices has rainbow connection number $n-1$. It is easy to show that for any connected graph G on n vertices that is neither complete nor a tree,

$$1 < \mathrm{rc}(G) < n-1.$$

A trivial lower bound for the rainbow connection number of any graph is its diameter, as pointed out in [12]: in a rainbow colouring with r colours, every pair of vertices is joined by a path of length at most r. Kamčev, Krivelevich and Sudakov [34] also gave an upper bound for the rainbow connection number in terms of the diameters of pairs of spanning subgraphs.

In general, the rainbow connection number and the diameter of a graph can be far apart from each other — for instance, the complete bipartite graph $K_{1,n-1}$ has diameter 2 but rainbow connection number $n-1$. In random graphs, however, the rainbow connection number often does not stray far from the diameter. For example, Frieze and Tsourakakis [25] showed that for the random graph $\mathcal{G}(n,p)$ near the connectivity threshold, whp the rainbow connection number is asymptotically equal to the maximum of the diameter and the number of degree one vertices. Dudek, Frieze and Tsourakakis [19] proved that the rainbow connection number of the random r-regular graph with $r \geqslant 4$ fixed is of the same order as its diameter.

In this and the next chapter, we will study the occurrence of rainbow connection number r in random graphs where r is constant, particularly threshold functions. Recall the definitions of different types of thresholds from Section 2.3.2, especially the non-standard notion of a semi-sharp threshold.

We are interested in the graph property

$$\mathcal{R}_r = \{G : \mathrm{rc}(G) \leqslant r\} \tag{5.1}$$

of having rainbow connection number at most r. Caro, Lev, Roditty, Tuza and Yuster [11] showed that $\sqrt{\frac{\log n}{n}}$ is a semisharp threshold for $\mathcal{R}_2$, and He and Liang [29] proved that for general $r \geqslant 2$, $\frac{(\log n)^{1/r}}{n^{1-1/r}}$ is a semi-sharp threshold for $\mathcal{R}_r$. As observed by Friedgut [23], a coarse threshold can only occur near rational powers of n. More specifically, from Theorem 1.4 in [23], if the semi-sharp threshold for $\mathcal{R}_r$ were not sharp, there would be a sequence (n_k) and $p(n_k) = \Theta\left(\frac{(\log n_k)^{1/r}}{n_k^{1-1/r}}\right)$ such that $b_1 n_k^\alpha \leqslant p(n_k) \leqslant b_2 n_k^\alpha$ for some constants $b_1, b_2 \in \mathbb{R}$ and $\alpha \in \mathbb{Q}$, which is a contradiction. Therefore, a *sharp threshold* for the property $\mathcal{R}_r$ must exist.

Bollobás [5] showed that for any fixed $r \geqslant 2$, $\frac{(2\log n)^{1/r}}{n^{1-1/r}}$ is a sharp threshold for the graph property

$$\mathcal{D}_r = \{G : \mathrm{diam}(G) \leqslant r\} \tag{5.2}$$

of having diameter at most r. In particular, $\mathcal{R}_r$ and $\mathcal{D}_r$ have the same weak threshold. Since $\mathrm{rc}(G) \geqslant \mathrm{diam}(G)$ for any graph G, the function $\frac{(2\log n)^{1/r}}{n^{1-1/r}}$ is a lower bound for the sharp threshold of $\mathcal{R}_r$. It is a natural question whether the *sharp* thresholds for rainbow connection number r and diameter r coincide as well.

In this chapter, we will answer this question affirmatively for $r = 2$ in the strongest possible sense, showing that rainbow connection number 2 and diameter 2 occur essentially at the same time in random graphs and indeed even in the random graph process. We will examine the case $r \geqslant 3$ in Chapter 6, where the situation seems fundamentally different, and propose an alternative sharp threshold function in this case.

For the rest of this chapter we only consider the case $r = 2$, so to simplify notation we let

$$\mathcal{R} = \mathcal{R}_2 \text{ and } \mathcal{D} = \mathcal{D}_2.$$

Let us first consider $\mathcal{G}(n,p)$ close to the threshold for diameter 2.

Theorem 5.1. *Let* $p = p(n) = \sqrt{\frac{2\log n + \omega(n)}{n}}$ *where* $\omega(n) = o(\log n)$ *and let* $G \sim \mathcal{G}(n,p)$. *Then whp* $\mathrm{rc}(G) = \mathrm{diam}(G) \in \{2,3\}$.

From [5] (see also Theorem 10.10 and Corollary 10.11 in [7]), we immediately get the following corollaries.

Corollary 5.2. *Let* $p = \sqrt{\frac{2\log n + c}{n}}$ *where* $c \in \mathbb{R}$ *is a constant, and let* $G \sim \mathcal{G}(n,p)$. *Then* $\lim_{n\to\infty} \mathbb{P}(\mathrm{rc}(G) = 2) = e^{-e^{-c}/2}$ *and* $\lim_{n\to\infty} \mathbb{P}(\mathrm{rc}(G) = 3) = 1 - e^{-e^{-c}/2}$.

Corollary 5.3. *Let* $p = \sqrt{\frac{2\log n + \omega(n)}{n}}$ *where* $\omega(n) \to \infty$ *such that* $(1-p)n^2 \to \infty$, *and let* $G \sim \mathcal{G}(n,p)$. *Then* $\mathrm{rc}(G) = 2$ *whp.*

We will in fact prove something even stronger than Theorem 5.1. Consider the *random graph process* $(G_t)_{t=0}^N$, $N = \binom{n}{2}$, which was defined in Chapter 2. Recall also that for a monotone increasing graph property $\mathcal{Q}$, we let $\tau_\mathcal{Q}$ denote the *hitting time* of $\mathcal{Q}$, i.e., the smallest t such that G_t has property $\mathcal{Q}$.

As $\mathcal{D}$ and $\mathcal{R}$ are monotone increasing and $\mathcal{D}$ is clearly necessary for $\mathcal{R}$, we always have $\tau_{\mathcal{D}} \leqslant \tau_{\mathcal{R}}$ for any order in which the edges are added to the graph. But in fact, whp $\mathcal{D}$ and $\mathcal{R}$ occur at the same time.

Theorem 5.4. *In the random graph process $(G_t)_{t=0}^{N}$, whp $\tau_{\mathcal{D}} = \tau_{\mathcal{R}}$.*

5.2 Proofs

5.2.1 Overview

We will construct a rainbow colouring in three rounds. We first add some of the edges of the random graph and 2-colour them randomly, up to edge density $\sqrt{\frac{1.01 \log n}{n}}$. Then we add most of the remaining edges — up to edge density $\sqrt{\frac{1.99 \log n}{n}}$ — and colour them more intelligently by adding rainbow paths whenever possible to pairs of vertices which are joined by only a few rainbow paths already. At this point the resulting random graph has diameter 3 whp, and we will show that whp this graph together with the given edge colouring has a certain desirable monotone increasing property $\mathcal{M}$. Finally, we add all remaining edges, and prove that if the final random graph has diameter 2 and property $\mathcal{M}$, we can tweak the edge colouring 'by hand' so that it becomes a rainbow 2-colouring.

It should be noted that there is an alternative proof which uses Lemma 6.5 from the next chapter. It follows from this lemma that our second stage of colouring is not necessary: we could, in fact, randomly 2-colour the edges up to edge density $\sqrt{\frac{1.01 \log n}{n}}$, then add all of the remaining edges and tweak the edge colouring in a somewhat different way. Lemma 6.5 ensures this is possible whp. Both Lemma 6.5 and its proof are of a somewhat technical nature, so we present the simpler three colouring rounds version of the proof here.

5.2.2 Definitions

For the proofs of Theorems 5.1 and 5.4, we will need a number of definitions. In a graph G with a given edge 2-colouring, we call a pair of non-adjacent vertices *dangerous* if they are joined by at most $d = 66$ rainbow paths of length 2. Moreover, we call a pair of non-adjacent vertices *sparsely connected* if they are joined by at most $d = 66$ paths of length 2 (rainbow or otherwise) and *richly connected* otherwise.

Definition 5.5. *We say that a graph has* property $\mathcal{M}$ *if it has a spanning subgraph which has an edge 2-colouring such that*

(i) Every vertex is in at most 3 *dangerous pairs.*

(ii) Every vertex is joined by edges to both vertices of at most 15 *dangerous pairs.*

(iii) Every vertex is in at most one sparsely connected pair.

Note that $\mathcal{M}$ is a monotone increasing graph property because it is defined by the existence of a spanning subgraph with some property. The property of having a colouring satisfying conditions (i)–(iii) is not itself monotone increasing, since condition (ii) does not necessarily stay true if we add more edges.

The following two propositions will form the main part of our proof.

Proposition 5.6. *If* $p = \sqrt{\frac{1.99 \log n}{n}}$*, then whp the graph* $G \sim \mathcal{G}(n,p)$ *has property* $\mathcal{M}$*.*

Proposition 5.7. *If a graph has properties* $\mathcal{M}$ *and* $\mathcal{D}$*, it also has property* $\mathcal{R}$*.*

Before turning to the proofs of Propositions 5.6 and 5.7, we show how they can be used to prove Theorems 5.1 and 5.4.

5.2.3 Proofs of Theorems 5.1 and 5.4

Proof of Theorem 5.1. Let $p = p(n) = \sqrt{\frac{2\log n + \omega(n)}{n}}$ where $\omega(n) = o(\log n)$, and let $G \sim \mathcal{G}(n,p)$. Since p is well above the threshold $\frac{(\log n)^{1/3}}{n^{2/3}}$ for the property $\mathrm{rc}(G) \leqslant 3$ established by He and Liang [29], we certainly have $\mathrm{rc}(G) \leqslant 3$ whp. In fact, for this p, it is easy to check that a random 3-colouring is rainbow whp. Since $p^{\binom{n}{2}} = o(1)$, whp G is not complete, so whp $\mathrm{diam}(G) \geqslant 2$. Since $\mathrm{diam}(G) \leqslant \mathrm{rc}(G)$, it remains only to show that whp $\mathrm{diam}(G) = 2$ implies $\mathrm{rc}(G) = 2$.

For n large enough, $p \geqslant \sqrt{\frac{1.99 \log n}{n}}$. Since $\mathcal{M}$ is monotone increasing, it follows from Proposition 5.6 that whp G has property $\mathcal{M}$. By Proposition 5.7, if $\operatorname{diam}(G) = 2$, i.e., G has property $\mathcal{D}$, then G also has property $\mathcal{R}$, so its rainbow connection number is at most 2. □

For Theorem 5.4, we need to construct the random graph process so that we can couple it with $\mathcal{G}(n, p)$, $p \in [0, 1]$.

Proof of Theorem 5.4. Take a set V of vertices where $|V| = n$, and assign to each potential edge e a random variable X_e which is distributed uniformly on $[0, 1]$, independently. Order the potential edges in ascending order of the corresponding random variables X_e. Almost surely, no two of the X_e take the same value, and any order of the X_e is equally likely. Therefore, we can add the edges to the graph one-by-one in the ascending order of the corresponding X_e, yielding a random graph process $(G_t)_{t=0}^{N}$, $N = \binom{n}{2}$, with the required distribution.

Let $p = \sqrt{\frac{1.99 \log n}{n}}$ and let $G = (V, E)$ where $e \in E$ iff $X_e \leqslant p$. Then since the random variables X_e are i.i.d. and distributed uniformly on $[0, 1]$, $G \sim \mathcal{G}(n, p)$.

By Proposition 5.6, whp G has property $\mathcal{M}$. Since, as shown by Bollobás [5] (see Theorem 10.10 in [7]), $\sqrt{\frac{2 \log n}{n}}$ is a sharp threshold for the property $\mathcal{D}$, whp G does not have property $\mathcal{D}$.

Since in the random graph process we added the edges in ascending order of their corresponding random variables, there is a (random) time $0 \leqslant t \leqslant N$ such that $G = G_t$. Therefore, there is whp a time t such that G_t has property $\mathcal{M}$ but not property $\mathcal{D}$, so $\tau_{\mathcal{M}} < \tau_{\mathcal{D}}$ whp. From Proposition 5.7, we get $\tau_{\mathcal{R}} \leqslant \max\{\tau_{\mathcal{D}}, \tau_{\mathcal{M}}\} = \tau_{\mathcal{D}}$ whp, and together with the trivial observation $\tau_{\mathcal{D}} \leqslant \tau_{\mathcal{R}}$, this implies the result. □

5.2.4 Proof of Proposition 5.6

We will generate the graph and an edge 2-colouring together in two steps. First consider the random graph $G_1 \sim \mathcal{G}(n, p_1)$ where $p_1 = \sqrt{\frac{(1+\varepsilon) \log n}{n}}$ and $\varepsilon = 0.01$. We will colour the edges of this graph randomly.

Next, we will add more edges to generate $G_2 \sim \mathcal{G}(n,p)$ where $p = \sqrt{\frac{1.99 \log n}{n}}$. Each edge which is not already present will be added independently with probability p_2, where $1 - p = (1-p_1)(1-p_2)$. We will colour these new edges so they add a rainbow 2-path to a dangerous pair whenever possible. We will show in Lemma 5.11, Corollary 5.13 and Lemma 5.14 that whp this gives an edge colouring which fulfills conditions (i)–(iii) of property $\mathcal{M}$ (with G_2 itself as the spanning subgraph).

First step: a random colouring

Let $G_1 \sim \mathcal{G}(n, p_1)$ where $p_1 = \sqrt{\frac{(1+\varepsilon)\log n}{n}}$ and $\varepsilon = 0.01$. Colour the edges of G_1 using two colours independently and uniformly at random.

We will now gather some information about the structure of the random graph and of the dangerous pairs in G_1. Recall that we denote the neighbourhood of a vertex v by $\Gamma(v)$.

Lemma 5.8. *With probability $1 - o(n^{-2})$, for every vertex v in G_1,*

$$\sqrt{\left(1+\frac{\varepsilon}{2}\right) n \log n} \leqslant |\Gamma(v)| \leqslant \sqrt{(1+2\varepsilon) n \log n}.$$

Proof. For a given vertex v, the number of neighbours of v is binomially distributed with parameters $n-1$ and p_1 and has mean $(n-1)p_1 = \sqrt{\frac{(n-1)^2}{n}(1+\varepsilon)\log n} \sim \sqrt{(1+\varepsilon) n \log n}$. By Corollary 2.5, the event that v has more than $\sqrt{(1+2\varepsilon) n \log n}$ or fewer than $\sqrt{(1+\frac{\varepsilon}{2}) n \log n}$ neighbours has probability $o(n^{-3})$. Taking the union bound over all vertices gives the result. □

Lemma 5.9. *The probability that a given pair $\{v, w\}$ of vertices is dangerous in G_1 is $O(n^{-\frac{1}{2}(1+\frac{\varepsilon}{2})})$. Moreover, with probability $1 - o(n^{-2})$, every vertex in G_1 is in at most $n^{\frac{1}{2}(1-\frac{\varepsilon}{4})}$ dangerous pairs.*

Proof. Fix a vertex v and explore the graph in the following way. Test all edges incident with v and their colours. With probability $1 - o(n^{-3})$, $|\Gamma(v)| \geqslant \sqrt{(1+\frac{\varepsilon}{2}) n \log n}$ as in the proof of Lemma 5.8. Assume this is the case and condition on a choice for $\Gamma(v)$ of at least this size.

If $w \in \Gamma(v)$, then $\{v, w\}$ is not dangerous, so suppose $w \notin \Gamma(v) \cup \{v\}$. The number of edges between w and $\Gamma(v)$ which have the correct colour for a rainbow 2-path between w and v is distributed binomially with parameters $|\Gamma(v)|$ and $\frac{1}{2}p_1$, with mean at least $\frac{1}{2}\sqrt{(1+\varepsilon)(1+\frac{\varepsilon}{2})}\log n$ for every possible choice for $\Gamma(v)$. So the probability that w has at most d edges of the appropriate colour for a rainbow path to $\Gamma(v)$ is $O(n^{-\frac{1}{2}\sqrt{(1+\varepsilon)(1+\frac{\varepsilon}{2})}}(\log n)^d)$ by Corollary 2.6.

Therefore, conditional on a choice for $\Gamma(v)$ of size at least $\sqrt{(1+\frac{\varepsilon}{2})n\log n}$, the pair $\{v, w\}$ is dangerous with probability $O(n^{-\frac{1}{2}(1+\frac{\varepsilon}{2})})$, and this happens independently for different $w \notin \Gamma(v) \cup \{v\}$. So the number of dangerous pairs that v is in is dominated by a binomial random variable with parameters n and $O(n^{-\frac{1}{2}(1+\frac{\varepsilon}{2})})$, which has mean $O(n^{\frac{1}{2}(1-\frac{\varepsilon}{2})})$. By Corollary 2.5, with probability $1 - o(n^{-3})$, v is in at most $n^{\frac{1}{2}(1-\frac{\varepsilon}{4})}$ dangerous pairs. Taking the union bound over all v gives the result. □

We call a pair of non-adjacent vertices $\{x, y\}$ in G_1 a *fix* for a pair $\{v, w\}$ if adding an edge $e = xy$ of a certain colour would add a rainbow path of length 2 between v and w. We call a fix $\{x, y\}$ for a pair $\{v, w\}$ an *exclusive fix* if there is no other dangerous pair (other than possibly $\{v, w\}$ if $\{v, w\}$ is dangerous) that $\{x, y\}$ is a fix for.

We expect to have about $2np_1$ fixes for every pair $\{v, w\}$ (of the form $\{x, w\}$ where $x \in \Gamma(v)$ or $\{v, y\}$ where $y \in \Gamma(w)$). We will now show that in fact most of these fixes are exclusive.

Lemma 5.10. *Whp, every non-adjacent pair $\{v, w\}$ of vertices in G_1 has at least $2\sqrt{(1+\frac{\varepsilon}{4})n\log n}$ exclusive fixes.*

Proof. Take v out of the graph G_1 and just look at the remaining graph G_1'. Then by Lemma 5.8 and (a slight variant of) Lemma 5.9, with probability $1 - o(n^{-2})$, every vertex in G_1' has at most $\sqrt{(1+2\varepsilon)n\log n}$ neighbours and is in at most $n^{\frac{1}{2}(1-\frac{\varepsilon}{4})}$ dangerous pairs (dangerous within G_1').

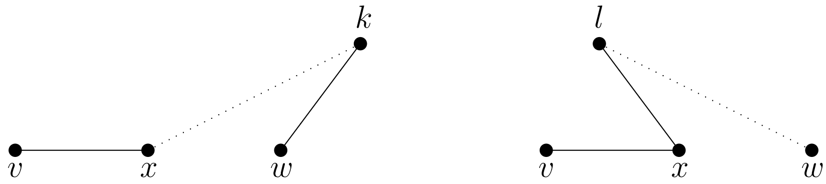

Figure 5.1: Two ways in which $\{x, w\}$ can be a fix for a dangerous pair other than $\{v, w\}$. The dotted lines show dangerous pairs.

In particular, if W_1' denotes the set of vertices x such that x is in a dangerous pair (within G_1') with a neighbour of w, and W_2' denotes the set of vertices x such that x is a neighbour of a vertex that is in a dangerous pair (within G_1') with w, then with probability $1 - o(n^{-2})$, $|W_1'| \leqslant n^{1-\frac{\varepsilon}{16}}$ and $|W_2'| \leqslant n^{1-\frac{\varepsilon}{16}}$.

In the whole graph G_1, let W_1 denote the set of all $x \in V \setminus \{v, w\}$ such that there is a neighbour $k \neq v$ of w such that $\{x, k\}$ is dangerous (in G_1). Let W_2 denote the set of all $x \in V \setminus \{v, w\}$ which have a neighbour $l \neq v$ such that $\{l, w\}$ is dangerous (in G_1). Any pair $\{s, t\} \subset V \setminus \{v\}$ which is dangerous in G_1 is also dangerous in G_1'. Therefore, $W_1 \subset W_1'$ and $W_2 \subset W_2'$.

A pair $\{x, w\}$ where $x \in \Gamma(v)$ can only fail to be an exclusive fix for $\{v, w\}$ in one of the following three ways. Either x and w are adjacent, or there is a $k \in \Gamma(w)$ such that $\{x, k\}$ is dangerous, or there is an $l \in \Gamma(x) \setminus \{v\}$ such that $\{l, w\}$ is dangerous (see Figure 5.1). If v and w are not adjacent, this can only happen if $x \in W_1 \cup W_2 \cup \Gamma'(w) \subset W_1' \cup W_2' \cup \Gamma'(w)$, where $\Gamma'(w)$ denotes the neighbourhood of w in G_1'.

Condition on G_1'. With probability $1 - o(n^{-2})$, $|W_1' \cup W_2' \cup \Gamma'(w)| \leqslant 3n^{1-\frac{\varepsilon}{16}}$. If this is the case, there are at least $n - 2 - 3n^{1-\frac{\varepsilon}{16}}$ potential neighbours x of v such that $\{x, w\}$ would be an exclusive fix for $\{v, w\}$; and each is actually adjacent to v with probability p_1 independently of each other and of G_1'.

Therefore, if v and w are not adjacent, the number of $x \in \Gamma(v)$ such that $\{x, w\}$ is an exclusive fix for $\{v, w\}$ is bounded from below by a binomial random variable with parameters $n - 2 - 3n^{1-\frac{\varepsilon}{16}}$ and $p_1 = \sqrt{\frac{(1+\varepsilon)\log n}{n}}$, which has mean greater than $\sqrt{(1 + \frac{\varepsilon}{2})n \log n}$ if n is large enough. By Corollary 2.5, it follows that with probability $1 - o(n^{-2})$, there are at least $\sqrt{(1+\frac{\varepsilon}{4})n\log n}$ exclusive fixes of the form $\{x, w\}$ where $x \in \Gamma(v)$. Analogously, with probability $1 - o(n^{-2})$, there are at least $\sqrt{(1+\frac{\varepsilon}{4})n\log n}$ exclusive fixes of the form $\{v, y\}$ where $y \in \Gamma(w)$, so overall with probability $1 - o(n^{-2})$, there are at least $2\sqrt{(1+\frac{\varepsilon}{4})n\log n}$ exclusive fixes for $\{v, w\}$. □

Second step: more edges with a more intelligent colouring

Now we are ready to introduce some additional edges which will be coloured more intelligently. Each edge which is not already present in the graph is now added independently with probability p_2, where p_2 is chosen so that $1-p = (1-p_1)(1-p_2)$. This ensures that after the second step, the probability that a particular edge is present is exactly $p = \sqrt{\frac{1.99\log n}{n}}$.

Note that $p_2 = p - p_1 + p_1 p_2 \geqslant p - p_1 = \frac{\sqrt{1.99\log n} - \sqrt{(1+\varepsilon)\log n}}{\sqrt{n}} \geqslant 0.4\sqrt{\frac{\log n}{n}}$ (recall that $\varepsilon = 0.01$).

Whenever a new edge is a fix for a dangerous pair, we give it the appropriate colour so that it adds a rainbow path of length 2 joining the dangerous pair. If there are several such dangerous pairs, we pick an arbitrary colour.

By Lemma 5.10, whp in G_1 there are at least $2\sqrt{(1+\frac{\varepsilon}{4})n\log n}$ exclusive fixes for every dangerous pair. Assume this from now on. These exclusive fixes will always get the correct colour for this pair if they are added. For a dangerous pair $\{v, w\}$ in G_1, let $N_{\{v,w\}}$ be the number of exclusive fixes of $\{v, w\}$ added in the second step. By definition, the sets of exclusive fixes are disjoint for different dangerous pairs. Therefore, conditional on G_1, the random variables $N_{\{v,w\}}$ are independent.

For a fixed dangerous pair $\{v, w\}$ in G_1, $N_{\{v,w\}}$ is bounded from below by a binomial random variable with parameters $2\sqrt{(1+\frac{\varepsilon}{4})n\log n}$ and p_2, which has mean at least $0.8\sqrt{1+\frac{\varepsilon}{4}}\log n$. Therefore, by Corollary 2.6,

$$\mathbb{P}(N_{\{v,w\}} \leqslant d) = O(n^{-0.8\cdot\sqrt{1+\frac{\varepsilon}{4}}}(\log n)^d) = O(n^{-0.8}). \tag{5.3}$$

Lemma 5.11. *In G_2 whp no vertex is in more than three dangerous pairs.*

Proof. Let L denote the event that every pair of vertices is either adjacent or has at least $2\sqrt{(1+\frac{\varepsilon}{4})n\log n}$ exclusive fixes in G_1, so L holds whp by Lemma 5.10. Let $v, w_1, \dots, w_4$ be distinct vertices, and let $D^v_{w_1,\dots,w_4}$ denote the event that the pairs $\{v, w_1\}, \dots, \{v, w_4\}$ are dangerous in G_2. Then

$$D^v_{w_1,\dots,w_4} \subset L^C \cup (L \cap D^v_{w_1,\dots,w_4}). \tag{5.4}$$

Let $\tilde{D}^v_{w_1,\dots,w_4}$ denote the event that $\{v, w_1\},\dots,\{v, w_4\}$ are dangerous in G_1. Then, since $D^v_{w_1,\dots,w_4} \subset \tilde{D}^v_{w_1,\dots,w_4}$,

$$\begin{aligned}
\mathbb{P}(L \cap D^v_{w_1,\dots,w_4}) &= \mathbb{P}(\tilde{D}^v_{w_1,\dots,w_4} \cap D^v_{w_1,\dots,w_4} \cap L) \\
&= \mathbb{P}(\tilde{D}^v_{w_1,\dots,w_4} \cap L)\mathbb{P}(D^v_{w_1,\dots,w_4} | \tilde{D}^v_{w_1,\dots,w_4} \cap L) \\
&\leqslant \mathbb{P}(\tilde{D}^v_{w_1,\dots,w_4})\mathbb{P}(D^v_{w_1,\dots,w_4} | \tilde{D}^v_{w_1,\dots,w_4} \cap L).
\end{aligned} \tag{5.5}$$

We first want to bound $\mathbb{P}(\tilde{D}^v_{w_1,\dots,w_4})$. If $z \in V \setminus \{v, w_1, \dots, w_4\}$, let E_z be the event that there is a rainbow path of length 2 from v to at least one w_i via z. The edge vz is present in G_1 with probability p_1, and if it is present, each edge zw_i is present in G_1 and has a different colour than vz with probability $\frac{p_1}{2}$, independently. Therefore, $q := \mathbb{P}(E_z) = p_1(1 - (1 - \frac{p_1}{2})^4) \sim 2p_1^2$, and the events E_z are independent for all $z \in V \setminus \{v, w_1, \dots, w_4\}$. Let K be the number of vertices z such that E_z holds. If $\{v, w_1\},\dots,\{v, w_4\}$ are all dangerous pairs, then $K \leqslant 4d$.

Since K is distributed binomially with parameters $n-5$ and q and with mean $(n-5)q \sim 2np_1^2 = 2(1+\varepsilon)\log n$, the probability that $K \leqslant 4d$ is $O(n^{-2(1+\frac{\varepsilon}{2})}(\log n)^{4d})$ by Corollary 2.6. Hence,

$$\mathbb{P}(\tilde{D}^v_{w_1,\dots,w_4}) = O(n^{-2}). \tag{5.6}$$

Conditional on G_1, if $\tilde{D}^v_{w_1,\dots,w_4}$ and L hold, the probability of the event $N_{\{v,w_i\}} \leqslant d$ that $\{v, w_i\}$ does not get at least $d+1$ of its exclusive fixes in the second round is $O(n^{-0.8})$ by (5.3), and these events are independent for different w_i. Therefore, by (5.5) and (5.6),

$$\mathbb{P}(L \cap D^v_{w_1,\dots,w_4}) = O(n^{-2}n^{-3.2}) = o(n^{-5}).$$

Hence, by (5.4),

$$\mathbb{P}\left(\bigcup_{v,w_1,\dots,w_4} D^v_{w_1,\dots,w_4}\right) \leqslant \mathbb{P}(L^C) + \mathbb{P}\left(\bigcup_{v,w_1,\dots,w_4} L \cap D^v_{w_1,\dots,w_4}\right)$$
$$\leqslant o(1) + n^5 o(n^{-5}) = o(1).$$

□

Lemma 5.12. *In G_2 whp no vertex is joined by edges to both vertices of more than 3 vertex disjoint dangerous pairs.*

Proof. Let v, u_i, w_i, $i = 1, \dots, 4$, be distinct vertices. Let A denote the event that v is adjacent in G_2 to all vertices u_i and w_i, $i = 1, \dots, 4$. Let D denote the event that all pairs $\{u_i, w_i\}$, $i = 1, \dots, 4$, are dangerous in G_2. Then we want to bound the probability of the event $A \cap D$.

For this, we will explore the edges of G_2 in several steps. First reveal the edges of the graph $G_1' = G_1 \setminus \{v\}$ and their colours. Let D' denote the event that all pairs $\{u_i, w_i\}$, $i = 1, \dots, 4$, are dangerous in G_1'. Then $D \subset D'$. By a variant of Lemma 5.9, a given pair $\{u_i, w_i\}$ is dangerous in G_1' with probability $O(n^{-\frac{1}{2}(1+\frac{\varepsilon}{2})})$, and it is easy to see that $\mathbb{P}(D') = O(n^{-2(1+\frac{\varepsilon}{4})})$. Indeed, for each $z \notin \{v, u_1, w_1, \dots, u_4, w_4\}$, the probability that z is the middle vertex of a rainbow path joining one of the pairs $\{u_i, w_i\}$, $i = 1, \dots, 4$, in G_1' is $2p_1^2(1 + o(1))$. These events are independent for different z, and at most $4d$ of these events can hold for D' to hold. Since $(n - 9)2p_1^2 \sim 2(1 + \varepsilon)\log n$, by Corollary 2.6, we have $\mathbb{P}(D') = O(n^{-2(1+\frac{\varepsilon}{2})}(\log n)^{4d}) = O(n^{-2(1+\frac{\varepsilon}{4})})$.

Next, reveal the edges of G_1 incident with v and their colours. They are independent from G_1'. For $k \in \{0, \dots, 8\}$, let A_k denote the event that v is adjacent in G_1 to exactly k of the vertices $\{u_1, w_1, \dots, u_4, w_4\}$. Then, since A_k and D' are independent,

$$\mathbb{P}(A_k \cap D') \leqslant \binom{8}{k} p_1^k O(n^{-2(1+\frac{\varepsilon}{4})}) = O(n^{-2-\frac{k}{2}}). \tag{5.7}$$

As before, let L denote the event that in G_1 all non-adjacent pairs of vertices have at least $2\sqrt{(1 + \frac{\varepsilon}{4})n \log n}$ exclusive fixes, which holds whp by Lemma 5.10. For every pair $\{u_i, w_i\}$, at most two exclusive fixes contain the vertex v (namely $\{v, u_i\}$ and $\{v, w_i\}$). So if L holds, then for n large enough, all pairs $\{u_i, w_i\}$, $i = 1, \ldots 4$, are either adjacent or have at least $2\sqrt{(1+\frac{\varepsilon}{8})n\log n}$ exclusive fixes which do not contain the vertex v. Call these fixes *v-free exclusive fixes*.

Now add the edges of G_2 not incident with v. Let D'' denote the event that every pair $\{u_i, w_i\}$, $i = 1, \ldots, 4$, not adjacent in G_1 now gets at most d of its v-free exclusive fixes. Note that $D \subset D''$. Conditional on G_1, if L holds and n is large enough, every non-adjacent pair $\{u_i, w_i\}$ has at least $2\sqrt{(1+\frac{\varepsilon}{8})n\log n}$ v-free exclusive fixes, and each one is added with probability $p_2 \geqslant 0.4\sqrt{\frac{\log n}{n}}$, independently. Hence, by Corollary 2.6, if L and D' hold,

$$\mathbb{P}(D'' \mid G_1) = \Big(O(n^{-0.8\sqrt{1+\frac{\varepsilon}{8}}}(\log n)^d)\Big)^4 = O(n^{-3.2}).$$

Finally, we add the remaining edges incident with v in G_2. Note that D'' depends on (G_1 and) the edges of G_2 not incident with v. Therefore, conditional on G_1, D'' and A are independent, so if $k \in \{0, \ldots, 8\}$, whenever L, D' and A_k hold in G_1, we have

$$\mathbb{P}(A \cap D \mid G_1) \leqslant \mathbb{P}(A \cap D'' \mid G_1) = \mathbb{P}(A \mid G_1)\mathbb{P}(D'' \mid G_1) = p_2^{8-k} O(n^{-3.2}).$$

This gives for $k \in \{0, \ldots, 8\}$,

$$\mathbb{P}(A \cap D \mid A_k \cap L \cap D') = O(n^{-3.2-\frac{8-k}{2}}(\log n)^{\frac{8-k}{2}}). \tag{5.8}$$

Since $A \cap D \subset \big(\bigcup_{k=0}^{8} A_k\big) \cap D'$, we have with (5.7) and (5.8),

$$\begin{aligned}
\mathbb{P}(A \cap D \cap L) &= \sum_{k=0}^{8} \mathbb{P}(A \cap D \cap L \cap A_k \cap D') \\
&= \sum_{k=0}^{8} \mathbb{P}(L \cap A_k \cap D')\mathbb{P}(A \cap D \mid L \cap A_k \cap D') \\
&\leqslant \sum_{k=0}^{8} \mathbb{P}(A_k \cap D')\mathbb{P}(A \cap D \mid L \cap A_k \cap D') \\
&\leqslant \sum_{k=0}^{8} O(n^{-2-\frac{k}{2}})O(n^{-3.2-\frac{8-k}{2}}(\log n)^{\frac{8-k}{2}}) \\
&= O(n^{-9.2}(\log n)^4) = o(n^{-9}).
\end{aligned} \tag{5.9}$$

Now, since we want to bound the probability that *there exist* vertices v, u_1, w_1, $\dots$, u_4, w_4 such that $A \cap D$ holds for them, we now add indices $A^{v,(u_i,w_i)_i}$, $D^{(u_i,w_i)_i}$ to our events A and D to make clear which vertices they refer to. The event L is a global event which is the same for all specific vertices v, u_1, w_1,…,u_4, w_4, so it does not require an index. Then using (5.9), the probability that there are vertices v, u_1, w_1,…,u_4, w_4 such that $A^{v,(u_i,w_i)_i} \cap D^{(u_i,w_i)_i}$ holds is at most

$$\mathbb{P}\left(\bigcup_{v,(u_i,w_i)_i} (A^{v,(u_i,w_i)_i} \cap D^{(u_i,w_i)_i})\right)$$
$$\leqslant \mathbb{P}(L^C) + \mathbb{P}\left(\bigcup_{v,(u_i,w_i)_i} (A^{v,(u_i,w_i)_i} \cap D^{(u_i,w_i)_i} \cap L)\right) = o(1) + n^9 o(n^{-9}) = o(1),$$

as required. □

Corollary 5.13. *In G_2 whp no vertex is joined by edges to both vertices of more than* 15 *dangerous pairs.*

Proof. By Lemma 5.12, whp no vertex is adjacent to both vertices of more than 3 vertex disjoint dangerous pairs, and by Lemma 5.11, whp every vertex is in at most 3 dangerous pairs. Assume this from now on.

Note that if a graph has maximum degree at most $\Delta \geqslant 1$ and more than $t(2\Delta-1)$ edges, where $t \in \mathbb{N}_0$, then it contains at least $t+1$ pairwise vertex-disjoint edges. This can be seen by induction on t — note that if one edge and its endpoints are removed from the graph, there are more than $t(2\Delta-1)-(2\Delta-1) = (t-1)(2\Delta-1)$ edges left.

Therefore, if some vertex v is joined to both vertices of more than $15 = 3\cdot(2\cdot 3-1)$ pairs, and every vertex is in at most 3 dangerous pairs, then v is joined to both vertices of at least $4 = 3+1$ pairwise disjoint dangerous pairs, which is not possible. □

Recall that we call a non-adjacent pair of vertices *sparsely connected* if they are joined by at most $d = 66$ paths of length 2 (rainbow or otherwise).

Lemma 5.14. *Whp every vertex in G_2 is in at most one sparsely connected pair.*

Proof. Consider some vertex v in G_2. Explore G_2 in the following way. Explore all edges incident with v. By Corollary 2.5, with probability $1-o(n^{-1})$, we have $|\Gamma(v)| \geqslant \sqrt{1.98n \log n}$. Condition on a choice for $\Gamma(v)$ where this is the case. Now for every vertex $w \notin \Gamma(v) \cup \{v\}$ (by definition sparse pairs are not adjacent), the probability that w has at most d edges to $\Gamma(v)$ is $O(e^{-\sqrt{1.98 \cdot 1.99} \log n}(\log n)^d) = O(n^{-1.98})$ by Corollary 2.6, and this is independent for different w. Therefore, for every possible choice for $\Gamma(v)$ of size at least $\sqrt{1.98n \log n}$, the probability that v is in two sparsely connected pairs is $O(n^2(n^{-1.98})^2) = O(n^{-1.96}) = o(n^{-1})$. Hence, the unconditioned probability that v is in two sparsely connected pairs is $o(n^{-1})$. Using the union bound, it follows that whp there is no such v. □

By Lemma 5.11, Corollary 5.13 and Lemma 5.14, the graph G_2 with the given edge colouring has property $\mathcal{M}$ whp (with G_2 itself as the spanning subgraph), which completes the proof of Proposition 5.6.

5.2.5 Proof of Proposition 5.7

To prove that $\mathcal{D}$ and $\mathcal{M}$ imply $\mathcal{R}$, we will take the edge 2-colouring given by property $\mathcal{M}$ and re-colour some edges to make a rainbow colouring. We will do this by first re-colouring paths joining sparsely connected dangerous pairs (this step only works if there are such paths at all, i.e., if we have diameter 2), and then doing the same for richly connected dangerous pairs.

So suppose properties $\mathcal{M}$ and $\mathcal{D}$ hold in some graph $G = (V, E)$. Take the spanning subgraph $G' = (V, E')$ and the edge 2-colouring of G' given by property $\mathcal{M}$. Do not assign colours to the edges in $E \setminus E'$ yet.

We will now assign some colours and change the colours of some edges in E' in order to make all dangerous pairs rainbow connected. We will *flag* all edges we (re-)assign a colour to as we go along so that they do not get reassigned another colour later on.

Call a pair of vertices *sparsely sub-connected* if it is sparsely connected in the subgraph G', and call it *richly sub-connected* otherwise. Call a pair *sub-dangerous* if it is dangerous in G'. Every sparsely connected pair in G is also sparsely sub-connected. Every dangerous pair in G is also sub-dangerous.

We start with the sparsely sub-connected sub-dangerous pairs. Take some arbitrary order of these pairs.

We will go through the sparsely sub-connected sub-dangerous pairs one by one in the given order, and each time ensure there is a rainbow path in E joining them, which is then flagged. Let $\{v, w\}$ be a pair we consider. Since $\mathcal{D}$ holds, either $vw \in E$, in which case we do not need to do anything, or v and w are joined by at least one path of length 2 in E. Let vzw be such a path.

It is not possible that both of the edges vz and zw are flagged already by the time we look at $\{v, w\}$: suppose that the edge $e = vz$ is flagged already. This can only have happened in one of the following two ways as shown in Figure 5.2. Either there is a vertex $w' \neq w$ such that $\{v, w'\}$ is sparsely sub-connected and sub-dangerous and the path vzw' was flagged for it, or there is a vertex z' such that $\{z, z'\}$ is sparsely sub-connected and sub-dangerous and the path zvz' was flagged for it. But the first case is impossible because by property $\mathcal{M}$, the vertex v is in at most one sparsely sub-connected pair (namely $\{v, w\}$). So the edge vz was flagged for a sparsely sub-connected sub-dangerous pair $\{z, z'\}$. Similarly, if zw is flagged already, this can only be because there is a vertex z'' such that $\{z, z''\}$ is sparsely sub-connected and sub-dangerous and zw was flagged for it. But then $z' \neq z''$, so z is in two sparsely sub-connected pairs, contradicting part (iii) of the definition of $\mathcal{M}$.

So take the path vzw. If necessary, adjust the colour of an un-flagged edge on it to make it a rainbow path, then flag both edges (if they are not flagged already).

Repeat this procedure until all sparsely sub-connected sub-dangerous pairs have rainbow paths. Now we will deal with richly sub-connected sub-dangerous pairs. Again take some arbitrary order of these pairs and consider them one by one.

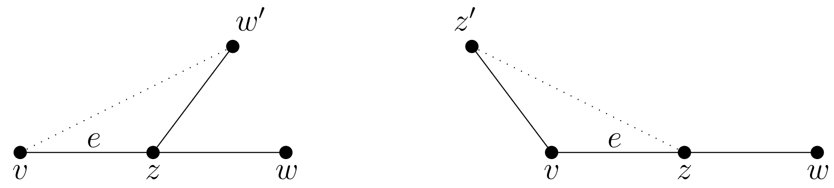

Figure 5.2: Possible ways in which the edge e could have been flagged before considering $\{v, w\}$. The dotted lines show sub-dangerous pairs other than $\{v, w\}$.

Let $\{v, w\}$ be the richly sub-connected sub-dangerous pair we consider. By definition, it is either adjacent in G, in which case we do not need to do anything, or joined by at least 67 paths of length 2 within E'. Let vzw be such a path. Then as before and as shown in Figure 5.2, the edge $e = vz$ can only be previously flagged for another (sparsely or richly sub-connected) sub-dangerous pair in one of the following two ways. Either there is a vertex $w' \neq w$ such that $\{v, w'\}$ is sub-dangerous and the path vzw' was flagged for it — since by property $\mathcal{M}$, v is in at most 3 sub-dangerous pairs in G', at most 3 edges in E' incident with v can be flagged this way (now or ever). Or there is a vertex z' such that $\{z, z'\}$ is sub-dangerous and the path zvz' was flagged for it — since by property $\mathcal{M}$, v is joined by edges to both vertices of at most 15 dangerous pairs in G', and for each such pair at most 2 edges incident with v are flagged, at most 30 edges in E' incident with v can be flagged this way (now or ever).

So at most 33 edges in E' incident with v can be flagged in this process. Analogously, at most 33 edges incident with w can be flagged. Since $\{v, w\}$ is joined by at least 67 paths of length 2 in G', there is at least one completely unflagged path at the time we look at $\{v, w\}$. Select one such path for $\{v, w\}$, adjust its colours if necessary to make it a rainbow path, then flag both of its edges and move on to the next richly sub-connected sub-dangerous pair.

Repeat this procedure until all richly sub-connected sub-dangerous pairs have rainbow paths. If there are any uncoloured edges left, assign them arbitrary colours. All sub-dangerous pairs are now joined by rainbow paths. It only remains to check that no non-sub-dangerous pairs have been rainbow disconnected in the process.

By the same argument as above (in the description of the procedure for richly sub-connected sub-dangerous pairs), for every vertex v at most 33 edges incident with v can be flagged and potentially re-coloured by the time we are done. If a pair $\{v, w\}$ is not sub-dangerous, it is either adjacent or is joined by at least 67 rainbow paths, of which at most 66 have been re-coloured. Therefore, every previously non-sub-dangerous pair still has at least one rainbow path left, so all pairs of vertices are joined by at least one rainbow path now. $\square$

Chapter 6

On the threshold for rainbow connection number $r \geqslant 3$

6.1 Background and results

In the last chapter, we saw that rainbow connection number 2 and diameter 2 happen essentially at the same time in the random graph $\mathcal{G}(n,p)$, and indeed in the random graph process. It is a natural question to ask whether this result may be extended to $r \geqslant 3$.

Recall the definitions (5.1) and (5.2) of the graph properties $\mathcal{R}_r$ and $\mathcal{D}_r$. As noted in the last chapter, we know that $\mathcal{R}_r$ has a sharp threshold, and that $\mathcal{R}_r$ and $\mathcal{D}_r$ share the same semisharp threshold. The results of the last chapter imply that $\mathcal{R}_2$ and $\mathcal{D}_2$ also share the same sharp threshold. However, the situation in the case $r \geqslant 3$ seems to be quite different from $r = 2$, and there are good reasons to believe that the following may be the true sharp threshold for $\mathcal{R}_r$ where $r \geqslant 3$.

Conjecture 6.1. *Fix an integer $r \geqslant 3$, set $C = \frac{r^{r-2}}{(r-2)!}$, and let*

$$p(n) = \frac{(C \log n)^{1/r}}{n^{1-1/r}}. \tag{6.1}$$

Then $p(n)$ is a sharp threshold for the graph property $\mathcal{R}_r$.

The constant $C = C(r)$ is chosen in such a way that if $G \sim \mathcal{G}(n, p(n))$ with $p(n)$ as in (6.1) and we r-colour the edges of G independently and uniformly at random, then the number of pairs of vertices which are joined by only a few rainbow paths is very roughly the same as the number of edges in G. For $r = 2$, this happens below sharp threshold for diameter 2 already, whereas for $r \geqslant 3$, this happens above the sharp threshold for diameter r.

In one direction, consider the following heuristic argument that (6.1) is a lower bound for the sharp threshold for $\mathcal{R}_r$. Let $\varepsilon > 0$ and $p = \frac{(C(1-\varepsilon)\log n)^{1/r}}{n^{1-1/r}}$, and colour the edges of $G \sim \mathcal{G}(n,p)$ with r colours independently and uniformly at random. For a given pair of vertices v, w, there are about n^{r-1} potential paths of length r joining them. Each is present with probability p^r and a rainbow path with probability $\frac{r!}{r^r}$. Therefore, the expected number of rainbow paths joining v and w is about $\frac{r!}{r^r}n^{r-1}p^r = (1-\varepsilon)\left(1-\frac{1}{r}\right)\log n$. If we assume that these potential rainbow paths behave roughly independently, then the distribution of the number of rainbow paths joining v and w can be approximated with a Poisson random variable with mean $(1-\varepsilon)\left(1-\frac{1}{r}\right)\log n$. The probability that v and w are not joined by any rainbow path is then about $n^{-(1-\varepsilon)\left(1-\frac{1}{r}\right)}$.

Hence, the expected number of pairs of vertices not joined by any rainbow path should be about $\Theta\left(n^{1+\frac{1}{r}+\varepsilon\left(1-\frac{1}{r}\right)}\right)$. Again assuming that these events behave roughly independently for different pairs of vertices, and approximating the number of such events that occur with a Poisson random variable with mean $\Theta\left(n^{1+\frac{1}{r}+\varepsilon\left(1-\frac{1}{r}\right)}\right)$, we would expect the overall probability of the random colouring being a rainbow colouring (i.e. that there is no pair of vertices not joined by any rainbow path) to be about $\exp\left(-\Theta\left(n^{1+\frac{1}{r}+\varepsilon\left(1-\frac{1}{r}\right)}\right)\right)$. Conditional on G having $O^*\left(n^{1+\frac{1}{r}}\right)$ edges, which holds with very high probability, there are $\exp\left(O^*\left(n^{1+\frac{1}{r}}\right)\right)$ possible edge colourings. The probability that there exists at least one rainbow colouring is then bounded by the total number of colourings multiplied by the probability that a random colouring is a rainbow colouring, which tends to 0.

The upper bound for the sharp threshold of the property $\mathcal{R}_r$ from the proof of the semi-sharp threshold in [29] is $\frac{(2^{20r}\log n)^{1/r}}{n^{1-1/r}}$. We will establish the other direction of Conjecture 6.1, i.e., we will prove the following result.

Theorem 6.2. *Fix an integer $r \geqslant 3$ and $\varepsilon > 0$, and let $C = \frac{r^{r-2}}{(r-2)!}$. Set $p = p(n) = \frac{(C(1+\varepsilon)\log n)^{1/r}}{n^{1-1/r}}$, and let $G \sim \mathcal{G}(n,p)$. Then whp,* $\operatorname{rc}(G) = r$.

6.2 Proof of Theorem 6.2

Let $G = (V, E) \sim \mathcal{G}(n, p)$. The basic idea of the proof is as follows. First we colour the edges of G independently and uniformly at random using r colours. We call a pair of vertices *dangerous* if it is joined by at most K rainbow paths of length r in this colouring, where K is a constant which will be defined later.

For each dangerous pair, we will select one path joining it and change the colours of the edges to make it a rainbow path, which will yield a rainbow colouring. To see that this is possible without any conflicts and that this does not rainbow-disconnect the pairs that previously had many rainbow paths, we need to study the structure of the graph and its dangerous pairs.

The rest of the chapter is organised as follows. Section 6.2.1 contains general observations on the distribution of edges of each colour and paths of length r in the randomly coloured graph. The heart of the proof is Section 6.2.2, where the key lemma is proved. This lemma ensures that when we later select a path of length r for every dangerous pair of vertices and recolour it to make it a rainbow path, it is possible to do so without using any edges from a path that was previously assigned to another dangerous pair of vertices. Finally, in Section 6.2.3, the recolouring procedure will be described in detail and we shall show that we can indeed find a rainbow r-colouring of the edges of G with this strategy.

6.2.1 General observations

For the rest of the chapter, define $p = p(n)$ as in Theorem 6.2, let $G \sim \mathcal{G}(n, p)$ and colour the edges of G independently and uniformly at random using r colours.

Lemma 6.3. *Let $\delta > 0$ be constant, let $W \subset V$ be a set of vertices with $|W| \sim n$, and let $v \in V$. Then for every colour, with probability $1 - o\left(\exp\left(-n^{1/r}\right)\right)$, there are at least $\frac{1-\delta}{r} np$ and at most $\frac{1+\delta}{r} np$ edges between v and W of the given colour.*

Proof. The number of such edges is distributed binomially with parameters $|W|$ (or $|W| - 1$ if $v \in W$) and p/r. Since $|W|p/r \sim ((1+\varepsilon)Cn \log n)^{1/r} / r$, the probability that there are fewer than $\frac{1-\delta}{r}np$ or more than $\frac{1+\delta}{r}np$ such edges is $o\left(\exp\left(-n^{1/r}\right)\right)$ by Corollary 2.5. $\square$ $\square$

For $k \in \mathbb{N}$, we call a path of length k in G a *k-path*, so a k-path is of the form $x_0x_1\dots x_k$ where the x_i are distinct vertices. We call a collection of paths in the graph *independent* if no two of them share any inner vertices.

Lemma 6.4. *There is a constant $c > 0$ such that whp every pair of vertices in G is joined by at least $c\log n$ independent r-paths.*

Proof. Fix two distinct vertices v and w. Step-by-step, we shall explore the $(r-1)$-neighbourhood of v, and apply Lemma 6.3 at each step with a suitable $\delta > 0$ to see that the sets we discover have the right size. Let $\delta > 0$ be such that $(1-\delta)(1+\varepsilon)^{1/r} = (1+\frac{\varepsilon}{2})^{1/r}$.

We start by considering all edges between v and $W_1 = V \setminus \{v, w\}$. By Lemma 6.3, with probability $1 - o\left(\exp\left(-n^{1/r}\right)\right)$ there are between $(1-\delta)np$ and $(1+\delta)np$ edges between v and W_1. Condition on this, and denote by N_1 the set of vertices in W_1 which are adjacent to v.

Next, let $W_2 = W_1 \setminus N_1$, and consider the edges between N_1 and W_2. Note that by our condition on the size of N_1, $|W_2| \sim n$. Furthermore, the edges between N_1 and W_2 are disjoint from and therefore independent of the edges we have considered so far. We go through the vertices z in N_1 one after the other, revealing the edges present. However, we disregard any edges to vertices which are adjacent to another vertex in N_1 which was considered earlier, so that the edges revealed form a tree. Applying Lemma 6.3 at each step, we can see that with probability $1 - |N_1| o\left(\exp\left(-n^{1/r}\right)\right) = 1 - o\left(\exp\left(-n^{1/2r}\right)\right)$, at each step there are between $(1-\delta)np$ and $(1+\delta)np$ edges.

Denote by N_2 the set of vertices in W_2 adjacent to a vertex in N_1, and let $W_3 = W_2 \setminus N_2$. We now proceed in the same way and explore the neighbours in W_3 of all vertices in N_2 disjointly, conditional on the neighbourhoods so far having the right sizes for all vertices according to Lemma 6.3.

We continue in this way until the entire $(r-1)$-neighbourhood N_{r-1} of the vertex v in W_1 is explored. Note that if all neighbourhoods have the right size, in total $O\left((np)^{r-1}\right) = o(n)$ vertices are revealed, so we can apply Lemma 6.3 at each step. With probability $1 - o(\exp\left(-n^{1/2r}\right))$, we now have a tree with at least $((1-\delta)np)^{r-1} = \left((1+\frac{\varepsilon}{2})Cn\log n\right)^{(r-1)/r}$ leaves. We can group the leaves together depending on which of the edges incident with v their path to v contains (i.e., which vertex in N_1 they originate from) — each group has size at least $\left((1+\frac{\varepsilon}{2})Cn\log n\right)^{(r-2)/r}$, and there are at least $\left(\left(1+\frac{\varepsilon}{2}\right)Cn\log n\right)^{1/r}$ groups. The edges between w and the leaves of the tree are independent from the edges that have been explored before. The probability that w has a neighbour in a given vertex group is therefore at least

$$\begin{aligned} 1-(1-p)^{\left((1+\frac{\varepsilon}{2})Cn\log n\right)^{(r-2)/r}} &\geqslant 1-\exp\left(-p\left(\left(1+\frac{\varepsilon}{2}\right)Cn\log n\right)^{(r-2)/r}\right) \\ &\geqslant 1-\exp\left(-\left(\left(1+\frac{\varepsilon}{2}\right)C\log n\right)^{(r-1)/r} n^{-1/r}\right) \\ &\sim \left(\left(1+\frac{\varepsilon}{2}\right)C\log n\right)^{(r-1)/r} n^{-1/r}, \end{aligned}$$

using the fact that $1-x \leqslant \exp(-x)$ for all $x \in \mathbb{R}$ and that $1-\exp(-x) \sim x$ as $x \to 0$. These events are independent for the different vertex groups, so the number of groups with at least one edge to w is distributed binomially. If we pick one edge from each such vertex group, this gives independent paths from v to w by construction. Since there are at least $\left((1+\frac{\varepsilon}{2})Cn\log n\right)^{1/r}$ vertex groups, the expected number of such paths is at least $(1+\frac{\varepsilon}{3})C\log n$ if n is large enough. Note that $\varphi(x) \to 1$ as $x \searrow -1$, where φ is the function defined in Corollary 2.4. Therefore, since $r \geqslant 3$ and $C = \frac{r^{r-2}}{(r-2)!} > 2$, if we pick $c > 0$ small enough, the probability that there are fewer than $c\log n$ independent r-paths joining v and w is $o(n^{-2})$ by Corollary 2.4. □

6.2.2 The main lemma

Let

$$L = \left\lceil \frac{17r}{\varepsilon(r-1)} \right\rceil, \ K = rL, \ \text{and} \ S = 2Lr^2 + 2.$$

Call a pair of vertices *dangerous* if it is joined by at most K independent rainbow r-paths in the random colouring. The following lemma will form the main part of the proof.

Lemma 6.5. *For a pair $\{v, w\}$ of vertices, denote by $A_{v,w}$ the event shown in Figure 6.1: there are L independent r-paths $P_1, \ldots, P_L$ joining v and w, and L r-paths $Q_1, \ldots, Q_L$ such that, writing $\{x_i, y_i\}$ for the end vertices of Q_i, the following conditions hold.*

i) For each i, P_i contains an edge e_i that is also on Q_i.

ii) The pairs $\{x_i, y_i\}$, $i = 1, \ldots, L$, and $\{v, w\}$ are distinct (but not necessarily disjoint).

iii) All pairs $\{x_i, y_i\}$, $i = 1, \ldots, L$, are dangerous.

Then whp $A_{v,w}$ does not hold for any pair $\{v, w\}$ of vertices.

The idea of the proof is the following. For one pair $\{x_i, y_i\}$ as in the lemma, the expected number of rainbow r-paths joining x_i and y_i is roughly $\frac{r!}{r^r} n^{r-1} p^r = \frac{r!}{r^r} C(1 + \varepsilon) \log n = \frac{r-1}{r}(1 + \varepsilon) \log n$. Since the rainbow paths behave roughly binomially, the probability that $\{x_i, y_i\}$ is dangerous is about $n^{-\frac{r-1}{r}(1+\varepsilon)} (\log n)^K$ by Corollary 2.6.

Therefore, given v and w, the probability that there is one path P_i containing an edge e_i which also lies on an r-path joining a dangerous pair $\{x_i, y_i\}$ is about

$$O^* \left(n^{2r-2} p^{2r-1} n^{-\frac{r-1}{r}(1+\varepsilon)} \right) = O^*(n^{-\varepsilon \frac{r-1}{r}}).$$

Therefore, if we can show that these events do not depend on each other too much for different P_i, then we would expect that the overall probability that there are L such paths is about $O^* \left(n^{-L\varepsilon \frac{r-1}{r}} \right)$. If L is chosen large enough, this will then be $o(n^{-2})$, completing the proof of the lemma.

A formal proof of this idea requires some care.

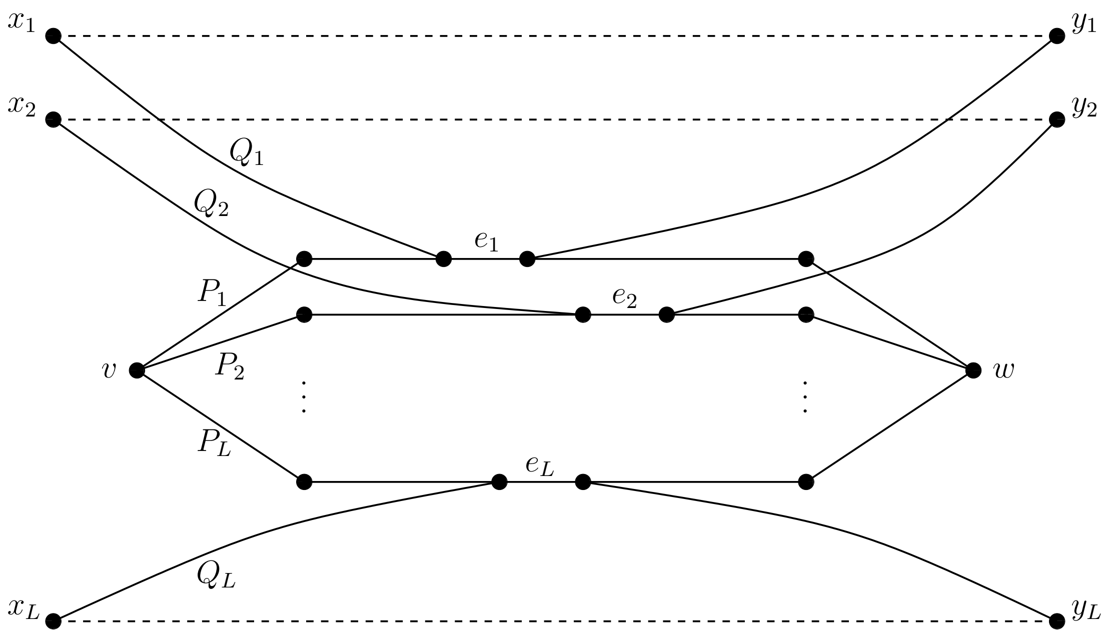

Figure 6.1: The event $A_{v,w}$. Dashed lines show dangerous pairs. The paths P_i only meet at v and w, while the paths Q_i may share vertices with each other and with the paths P_i. The pairs $\{x_i, y_i\}$ are distinct, but not necessarily disjoint.

Proof of Lemma 6.5. Fix distinct vertices v and w. Consider a possible configuration of vertices and edges for the paths P_i, Q_i, edges e_i and pairs $\{x_i, y_i\}$ as in conditions (i) and (ii) of $A_{v,w}$. Denote by k the number of vertices in the configuration other than v and w, and let l be the number of edges in the configuration. Then $k \leqslant 2(r-1)L$, as the configuration consists of L r-paths P_i with endpoints v and w, and L r-paths Q_i which each share at least two vertices with a path P_i.

Note that the configuration is connected and remains connected if we remove one edge on all but one path P_i, since there is still one v-w path left. Since a connected graph with m vertices has at least $m-1$ edges, it follows that $l-(L-1) \geqslant (k+2)-1$, so $l \geqslant k + L$. Therefore,

$$n^k p^l \leqslant (np)^k p^L \leqslant (np)^{2(r-1)L} p^L = n^{2(r-1)L} p^{(2r-1)L}. \tag{6.2}$$

Now condition on a specific such configuration being present in G. Let W denote the set of vertices involved in the configuration, including v and w, and let $V' = V \setminus W$. Then $|W| = k + 2 \leqslant 2Lr$, and $|V'| \sim n$. The edges between W and V' and

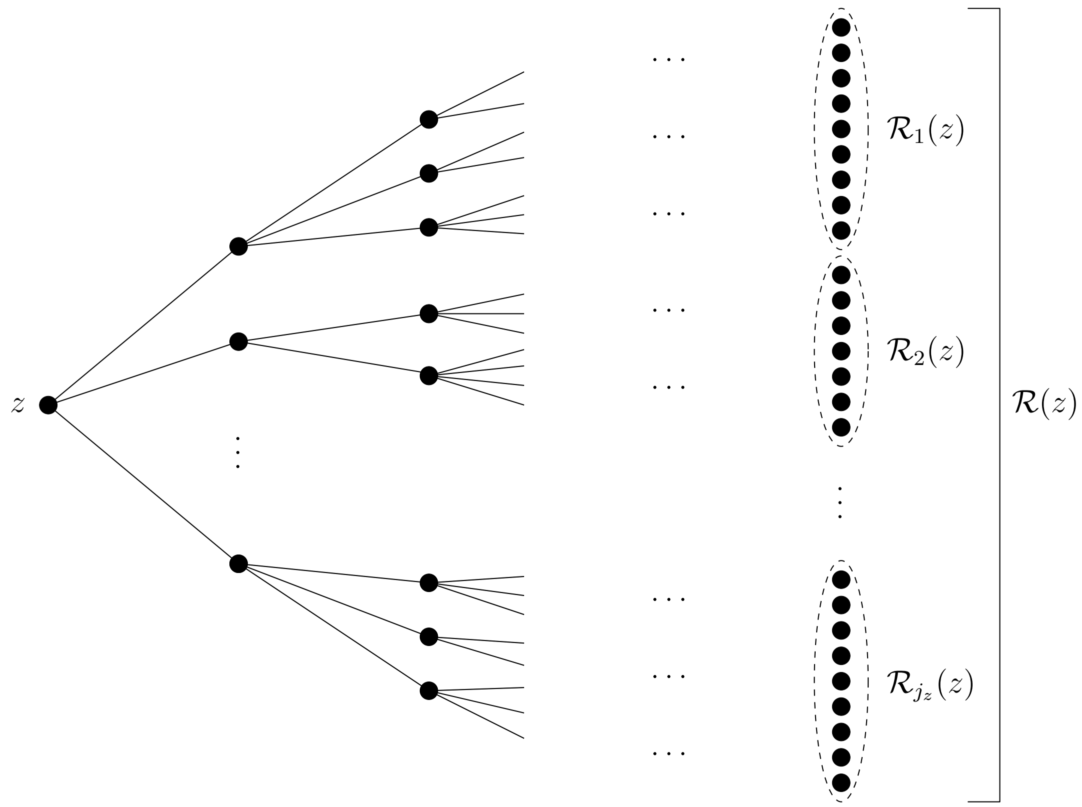

Figure 6.2: The tree of depth s obtained by the exploration of the rainbow s-neighbourhood of z. The paths from the root z to the leaves are rainbow paths. We have one such tree for each $z \in D$; these trees are disjoint.

within V' are disjoint from and therefore independent from the edges involved in the configuration.

Let $s = \left\lfloor \frac{r-1}{2} \right\rfloor$, and let D be the set of all vertices x_i and y_i from the configuration. We now explore the s-neighbourhoods $\Gamma^s(z)$ of the vertices $z \in D$ within V'. We want to find disjoint subsets of $\Gamma^s(z)$ such that all elements are joined to z by a rainbow s-path and all such paths are independent, except if they come from the same neighbour of z. We do this as in the proof of Lemma 6.4 — first explore the neighbours of $z_1 \in D$ in V', then explore their neighbours in V' and so on, then proceed with the next vertex $z_2 \in D$, and so on. As before, at each step, we disregard edges to vertices that have been explored already. Unlike in the proof of Lemma 6.4, at each step we only check for new neighbours joined by edges with colours not appearing on the path from z to the current vertex. We group together vertices that come from the same edge incident with some $z \in D$. This gives disjoint subsets $\mathcal{R}_j(z)$ of V' for every $z \in D$, $1 \leqslant j \leqslant j_z$, where the $\mathcal{R}_j(z)$ are the vertex groups which come from the same neighbour of z, as shown in Figure 6.2. Let $\mathcal{R}(z) = \bigcup_{1 \leqslant j \leqslant j_z} \mathcal{R}_j(z)$. Then by definition, the following properties hold.

i) For every $z_1, z_2 \in D$, $1 \leqslant j_1 \leqslant j_{z_1}$ and $1 \leqslant j_2 \leqslant j_{z_2}$, if $z_1 \neq z_2$ or if $j_1 \neq j_2$ then the sets $\mathcal{R}_{j_1}(z_1)$ and $\mathcal{R}_{j_2}(z_2)$ are disjoint.

ii) For every $z \in D$, every vertex $z' \in \mathcal{R}(z)$ is joined to z by a rainbow s-path $P_{z'}$ with all inner vertices in V'.

iii) For every z_1, z_2, j_1, j_2 and $z_1' \in \mathcal{R}_{j_1}(z_1)$, $z_2' \in \mathcal{R}_{j_2}(z_2)$, if $z_1 \neq z_2$ or if $j_1 \neq j_2$, then the paths $P_{z_1'}$ and $P_{z_2'}$ do not share any inner vertices.

Applying Lemma 6.3 at each step of our exploration with $\delta > 0$ such that $(1-\delta)(1+\varepsilon)^{1/r} \geqslant \left(1 + \frac{\varepsilon}{2}\right)^{1/r}$ and $(1+\delta)(1+\varepsilon)^{1/r} \leqslant (1+2\varepsilon)^{1/r}$, we see that with probability $1 - o(n^{-2Lr})$, the following additional properties hold.

iv) For all $z \in D$, $\left(C\left(1 + \frac{\varepsilon}{2}\right) n \log n\right)^{1/r} \leqslant j_z \leqslant \left(C\left(1 + 2\varepsilon\right) n \log n\right)^{1/r}$.

v) For all $z \in D$ and $1 \leqslant j \leqslant j_z$, $|\mathcal{R}_j(z)| = O^*\left(n^{\frac{s-1}{r}}\right)$.

vi) For every subset S of the r available colours such that $|S| = s$ and for all $z \in D$ there are at least

$$\frac{s!}{r^s}\left(C\left(1 + \frac{\varepsilon}{2}\right) n \log n\right)^{s/r}$$

and at most

$$\frac{s!}{r^s}\left(C\left(1 + 2\varepsilon\right) n \log n\right)^{s/r}$$

vertices $z' \in \mathcal{R}(z)$ such that the colours appearing on $P_{z'}$ are exactly the colours in S.

Assume (i) – (vi) from now on. Then $|\mathcal{R}| = O((n \log n)^{s/r}) = o(n)$, so $|V' \setminus \mathcal{R}| \sim n$.

Case 1: r is odd. In this case $s = \lfloor \frac{r-1}{2} \rfloor = \frac{r-1}{2}$. For every subset S of colours of size s, there are $\binom{r-s}{s} = s+1$ sets of colours of size s disjoint from S. Therefore, for every vertex u_1 in some set $\mathcal{R}(x_i)$ there are at least

$$(s+1)\frac{s!}{r^s}\left(C\left(1+\frac{\varepsilon}{2}\right)n\log n\right)^{s/r}$$

vertices u_2 in $\mathcal{R}(y_i)$ such that an edge u_1u_2 of the correct colour would complete a rainbow r-path from x_i to y_i.

Therefore, there are at least

$$\binom{r}{s}(s+1)\frac{(s!)^2}{r^{2s}}\left(C\left(1+\frac{\varepsilon}{2}\right)n\log n\right)^{2s/r} = \frac{r!}{r^{r-1}}\left(C\left(1+\frac{\varepsilon}{2}\right)n\log n\right)^{(r-1)/r}$$

potential edges between $\mathcal{R}(x_i)$ and $\mathcal{R}(y_i)$ such that each one would complete a rainbow path from x_i to y_i. Each of these edges is present and has the correct colour for a rainbow path with probability $\frac{1}{r}p$, independently from the edges that have been revealed so far. Therefore, the number of such edges is distributed binomially with mean at least

$$\frac{r!}{r^r}C\left(1+\frac{\varepsilon}{2}\right)\log n = \frac{r-1}{r}\left(1+\frac{\varepsilon}{2}\right)\log n.$$

If we denote by E_i the event that there are at most $2SK$ edges of the correct colour between $\mathcal{R}(x_i)$ and $\mathcal{R}(y_i)$ to complete a rainbow path between x_i and y_i, then by Corollary 2.6,

$$\mathbb{P}(E_i) = O^*\left(n^{-\left(1+\frac{\varepsilon}{2}\right)\frac{r-1}{r}}\right).$$

The events E_i are independent for different pairs $\{x_i, y_i\}$ since the pairs $\{x_i, y_i\}$ are distinct and all sets $\mathcal{R}(x_i)$, $\mathcal{R}(y_i)$ are disjoint. Hence,

$$\mathbb{P}\left(\bigcap_{1\leqslant i\leqslant L} E_i\right) = O^*\left(n^{-L\left(1+\frac{\varepsilon}{2}\right)\frac{r-1}{r}}\right) = O\left(n^{-L\left(1+\frac{\varepsilon}{4}\right)\frac{r-1}{r}}\right). \tag{6.3}$$

For every $z \in D$ and $1 \leqslant j \leqslant j_z$, let B_j^z denote the event that there are at least S edges between $\mathcal{R}_j(z)$ and $\mathcal{R} \setminus \mathcal{R}_j(z)$. Then

$$\mathbb{P}(B_j^z) \leqslant \left(|\mathcal{R}_j(z)||\mathcal{R}|p\right)^S = O^*\left(n^{S(s-1)/r}n^{Ss/r}p^S\right) = O^*\left(n^{-S/r}\right),$$

as $|\mathcal{R}_j(z)| = O^*\left(n^{(s-1)/r}\right)$ and $|\mathcal{R}| = O^*\left(n^{s/r}\right)$. Hence, letting $B = \bigcup_{(z,j):1\leqslant j\leqslant j_z} B_j^z$,

$$\mathbb{P}\left(B\right) = O^*\left(n^{1/r}n^{-S/r}\right) = o(n^{-2Lr}), \tag{6.4}$$

by choice of S.

If $B^{\mathrm{c}} \cap E_i^{\mathrm{c}}$ holds for some $1 \leqslant i \leqslant L$, then the pair $\{x_i, y_i\}$ is not dangerous. This is because we have at least $2SK$ edges of the correct colour between $\mathcal{R}(x_i)$ and $\mathcal{R}(y_i)$ to complete a rainbow path between x_i and y_i, but there are at most S such edges from each particular vertex group $\mathcal{R}_j(x_i)$ or $\mathcal{R}_j(y_i)$, so we can successively pick K such edges between pairwise distinct vertex groups $\mathcal{R}_j(x_i)$ or $\mathcal{R}_j(y_i)$, yielding K independent rainbow paths between x_i and y_i by property (iii).

Therefore, if all L pairs $\{x_i, y_i\}$ are dangerous, then $B \cup \bigcap_{1\leqslant i\leqslant L} E_i$ holds. Hence, by (6.3) and (6.4), the probability that all L pairs $\{x_i, y_i\}$ are dangerous is bounded by

$$O\left(n^{-L(r-1)\left(1+\frac{\varepsilon}{4}\right)/r}\right) + o(n^{-2Lr}).$$

Case 2: r is even.

In this case $s = \left\lfloor\frac{r-1}{2}\right\rfloor = \frac{r}{2} - 1$. Let $u \in V' \setminus \mathcal{R}$. Given a vertex u_1 in some $\mathcal{R}(x_i)$, there are at least

$$\binom{r-s}{s}\frac{s!}{r^s}\left(C\left(1+\frac{\varepsilon}{2}\right)n\log n\right)^{s/r}$$

and at most

$$\binom{r-s}{s}\frac{s!}{r^s}\left(C\left(1+2\varepsilon\right)n\log n\right)^{s/r}$$

vertices u_2 in $\mathcal{R}(y_i)$ such that adding edges u_1u and uu_2 of appropriate colours would complete a rainbow r-path from x_i to y_i via u_1, u and u_2. Therefore, there are at least

$$\binom{r}{s}\binom{r-s}{s}\frac{s!^2}{r^{2s}}\left(C\left(1+\frac{\varepsilon}{2}\right)n\log n\right)^{2s/r} = \frac{r!}{2r^{r-2}}\left(C\left(1+\frac{\varepsilon}{2}\right)n\log n\right)^{\frac{r-2}{r}} \tag{6.5}$$

and at most

$$\frac{r!}{2r^{r-2}}\left(C\left(1+2\varepsilon\right)n\log n\right)^{\frac{r-2}{r}} \tag{6.6}$$

pairs of vertices $u_1 \in \mathcal{R}(x_i)$, $u_2 \in \mathcal{R}(y_i)$ such that edges $u_1 u$ and uu_2 of appropriate colours would complete a rainbow path from x_i to y_i. For one such pair $\{u_1, u_2\}$ and $u \in V' \setminus \mathcal{R}$, denote by $M_u^{u_1,u_2}$ the event that the edges $u_1 u$ and uu_2 are present and have one of the two possible colour combinations. Then

$$\mathbb{P}(M_u^{u_1,u_2}) = \frac{2}{r^2} p^2. \tag{6.7}$$

Moreover, if the events $M_u^{u_1,u_2}$ and $M_u^{u_1',u_2'}$ hold for different pairs $\{u_1, u_2\}$ and $\{u_1', u_2'\}$, then u is adjacent to three or more distinct vertices from $\{u_1, u_2, u_1', u_2'\}$, so

$$\mathbb{P}\left(M_u^{u_1,u_2} \cap M_u^{u_1',u_2'}\right) = O(p^3). \tag{6.8}$$

For a vertex $u \in V' \setminus \mathcal{R}$, denote by F_u the event that u is the middle vertex of any path as above for any $1 \leqslant i \leqslant L$. Then by Lemma 2.2 and (6.5), (6.6), (6.7), (6.8),

$$\begin{aligned}
\mathbb{P}(F_u) &\geqslant L\frac{r!}{r^r}\left(C\left(1+\frac{\varepsilon}{2}\right)n\log n\right)^{\frac{r-2}{r}} p^2 - O^*\left(n^{\frac{2(r-2)}{r}}\right)O\left(p^3\right) \\
&\geqslant L\frac{r-1}{r}\left(1+\frac{\varepsilon}{2}\right)n^{-1}\log n - O^*\left(n^{-1-\frac{1}{r}}\right) \\
&\sim L\frac{r-1}{r}\left(1+\frac{\varepsilon}{2}\right)n^{-1}\log n.
\end{aligned}$$

The events F_u are independent for different $u \in V'\setminus\mathcal{R}$. Thus, since $|V'\setminus\mathcal{R}| \sim n$, the number of events F_u that hold is distributed binomially with mean asymptotically at least

$$L\frac{r-1}{r}\left(1+\frac{\varepsilon}{2}\right)\log n.$$

By Corollary 2.6, the probability of the event F that at most $2KLS$ of the events F_u hold is

$$\mathbb{P}(F) = O^*\left(n^{-(1+o(1))L\frac{r-1}{r}\left(1+\frac{\varepsilon}{2}\right)}\right) = O\left(n^{-L\frac{r-1}{r}\left(1+\frac{\varepsilon}{4}\right)}\right). \tag{6.9}$$

For every $z \in D$ and $1 \leqslant j \leqslant j_z$, denote by $\tilde{B}_j^z$ the event that there are at least S independent 2-paths from (not necessarily distinct) vertices in $\mathcal{R}_j(z)$ to (not necessarily distinct) vertices in $\mathcal{R}$ with middle vertices in $V' \setminus \mathcal{R}$. Then, since $|\mathcal{R}_j(z)| = O^*\left(n^{(s-1)/r}\right)$ and $|\mathcal{R}| = O^*\left(n^{s/r}\right)$,

$$\mathbb{P}(\tilde{B}_j^z) \leqslant \left(|\mathcal{R}_j(z)|\,|V|\,|\mathcal{R}|\,p^2\right)^S = O^*\left(\left(n^{\frac{s-1}{r}+1+\frac{s}{r}}p^2\right)^S\right) = O^*\left(n^{-S/r}\right).$$

Therefore, if we let $\tilde{B} = \bigcup_{(z,j):1\leqslant j\leqslant j_z} \tilde{B}_j^z$, then

$$\mathbb{P}(\tilde{B}) = O^*\left(n^{1/r} n^{-S/r}\right) = o(n^{-2Lr}), \tag{6.10}$$

by choice of S.

If neither $\tilde{B}$ nor F holds, then one of the pairs $\{x_i, y_i\}$ is not dangerous. This is because there are more than $2KLS$ vertices $u \in V' \setminus \mathcal{R}$ which are the middle vertices of rainbow paths joining pairs $\{x_i, y_i\}$, so there is an index $1 \leqslant i_0 \leqslant L$ such that there are more than $2KS$ vertices $u \in V' \setminus \mathcal{R}$ which are the middle vertices of rainbow r-paths joining the pair $\{x_{i_0}, y_{i_0}\}$ (we can just pick the index with the maximum number of such vertices u). If $\tilde{B}$ does not hold, at most S of those paths pass through any particular vertex group $\mathcal{R}_j(x_{i_0})$ or $\mathcal{R}_j(y_{i_0})$. Hence, we can successively select more than K rainbow r-paths joining $\{x_{i_0}, y_{i_0}\}$ which pass through pairwise distinct vertex groups. These rainbow paths are independent by property (iii), so $\{x_{i_0}, y_{i_0}\}$ is not dangerous.

Hence, if all pairs $\{x_i, y_i\}$ are dangerous, then $\tilde{B}$ or F holds, which by (6.9) and (6.10) has probability

$$O\left(n^{-L\left(1+\frac{\varepsilon}{4}\right)\frac{r-1}{r}}\right) + o\left(n^{-2Lr}\right).$$

So in each case, conditional on a configuration of paths P_i and Q_i, edges e_i and pairs $\{x_i, y_i\}$ as in conditions (i) and (ii) of the event $A_{v,w}$, the probability that all pairs $\{x_i, y_i\}$ are dangerous is at most

$$O\left(n^{-L\left(1+\frac{\varepsilon}{4}\right)\frac{r-1}{r}}\right) + o\left(n^{-2Lr}\right).$$

Using (6.2), it follows that the overall probability of $A_{v,w}$ is at most

$$\begin{aligned}
&O\left(n^{2(r-1)L} p^{(2r-1)L} n^{-L\left(1+\frac{\varepsilon}{4}\right)\frac{r-1}{r}}\right) + o(n^{-2L}) = \\
&= O\left(\left(n^{2r-2-\left(1+\frac{\varepsilon}{4}\right)\frac{r-1}{r}} p^{2r-1}\right)^L\right) + o(n^{-2}) = O\left(n^{-\frac{\varepsilon(r-1)}{8r}L}\right) + o(n^{-2}) \\
&= o(n^{-2}),
\end{aligned}$$

by choice of L. So whp, there is no such pair $\{v, w\}$. □

6.2.3 Completing the proof

To finish the proof, we want to construct a rainbow colouring of the edges of G from the given random colouring. By Lemmas 6.4 and 6.5, we can assume that every pair of vertices is joined by at least $c \log n$ independent r-paths for a constant $c > 0$, and that $A_{v,w}$ does not hold for any pair $\{v, w\}$ of vertices.

Recall that we call a pair of vertices dangerous if it is joined by at most K independent rainbow r-paths in the original random colouring. Take an arbitrary ordering of the dangerous pairs. We will go through them one by one, each time selecting an r-path joining the dangerous pair, changing its colours if necessary to make it a rainbow path, then *flagging* all edges on the path to ensure they do not get recoloured later on.

Let $\{v, w\}$ be the pair we consider. It is joined by at least $c \log n \geqslant rL$ independent r-paths if n is large enough. We want to find one such path where no edge is flagged yet.

So take a set $\mathcal{I}$ of rL independent r-paths joining v and w, and consider any such path P_1 in $\mathcal{I}$. Either none of its edges is flagged — in this case, we have found our path. Otherwise, it contains (at least) one edge which is also on an r-path joining a dangerous pair other than $\{v, w\}$. For this dangerous pair, one path of length r was flagged previously. Therefore, at most $r-1$ of the other paths in $\mathcal{I}$ can contain edges flagged for the same dangerous pair. Discard those paths and P_1. We are left with at least $r(L-1)$ paths joining v and w. Select any such path P_2 and proceed in the same way as with P_1: either P_2 is completely unflagged, or we remove P_2 and any other path with edges flagged for the same dangerous pair as P_2 from consideration, and are left with at least $r(L-2)$ paths. We repeat this procedure until we find a completely unflagged path. This happens at P_L at the latest. Otherwise, if P_L also contains an edge flagged for a new dangerous pair, then $A_{v,w}$ holds, a contradiction.

Therefore, there is a path joining $\{v, w\}$ where no edge is flagged at the time we consider $\{v, w\}$. Select this path, change its colours if necessary to make it a rainbow path, then flag all its edges and move on to the next dangerous pair. Repeat this procedure until all dangerous pairs have been assigned rainbow paths.

It only remains to check that during our recolouring procedure no previously non-dangerous pair has lost all of its rainbow paths. Let $\{v, w\}$ be a pair that was not dangerous before we started recolouring. Since it was originally joined by at least $K = rL$ rainbow paths, by the same argument as above for dangerous pairs, one of these paths must be completely unflagged, otherwise $A_{v,w}$ would hold. This path has retained its original colours and is therefore still a rainbow path. So all pairs of vertices are joined by rainbow paths now. $\square$

Bibliography

[1] D. Achlioptas and E. Friedgut. A sharp threshold for k-colorability. *Random Structures and Algorithms*, 14(1):63–70, 1999.

[2] D. Achlioptas and A. Naor. The two possible values of the chromatic number of a random graph. *Annals of Mathematics*, 162:1335–1351, 2005.

[3] N. Alon and M. Krivelevich. The concentration of the chromatic number of random graphs. *Combinatorica*, 17(3):303–313, 1997.

[4] N. Alon and J. Spencer. *The probabilistic method*. Wiley, 4th edition, 2015.

[5] B. Bollobás. The diameter of random graphs. *Transactions of the American Mathematical Society*, 267:41–52, 1981.

[6] B. Bollobás. The chromatic number of random graphs. *Combinatorica*, 8(1):49–55, 1988.

[7] B. Bollobás. *Random Graphs*. Cambridge University Press, second edition, 2001.

[8] B. Bollobás. How sharp is the concentration of the chromatic number? *Combinatorics, Probability and Computing*, 13(01):115–117, 2004.

[9] B. Bollobás and P. Erdős. Cliques in random graphs. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 80, pages 419–427. Cambridge University Press, 1976.

[10] B. Bollobás and A. Thomason. Threshold functions. *Combinatorica*, 7(1):35–38, 1987.

[11] Y. Caro, A. Lev, Y. Roditty, Z. Tuza, and R. Yuster. On rainbow connection. *Electronic Journal of Combinatorics*, 15:R57, 2008.

[12] G. Chartrand, G.L. Johns, K.A. McKeon, and P. Zhang. Rainbow connection in graphs. *Mathematica Bohemica*, 133:85–98, 2008.

[13] B. Chen, K. Lih, and P. Wu. Equitable coloring and the maximum degree. *European Journal of Combinatorics*, 15(5):443–447, 1994.

[14] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23:493–507, 1952.

[15] A. Coja-Oghlan. Upper-bounding the k-colorability threshold by counting covers. *The Electronic Journal of Combinatorics*, 20(3):P32, 2013.

[16] A. Coja-Oghlan, K. Panagiotou, and A. Steger. On the chromatic number of random graphs. *Journal of Combinatorial Theory B*, 98(5):980–993, 2008.

[17] A. Coja-Oghlan and D. Vilenchik. Chasing the k-colorability threshold. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 380–389, 2013.

[18] A. Coja-Oghlan and D. Vilenchik. The chromatic number of random graphs for most average degrees. *International Mathematics Research Notices*, page rnv333, 2015.

[19] A. Dudek, A. Frieze, and C. Tsourakakis. Rainbow connection of random regular graphs. *SIAM Journal on Discrete Mathematics*, 29(4):2255–2266, 2015.

[20] P. Erdős and A. Rényi. On random graphs, I. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

[21] P. Erdős and A. Rényi. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5:17–61, 1960.

[22] N. Fountoulakis, R. Kang, and C. McDiarmid. The t-stability number of a random graph. *The Electronic Journal of Combinatorics*, 17(1):R59, 2010.

[23] E. Friedgut. Sharp thresholds of graph properties, and the k-sat problem. *Journal of the Americal Mathematical Society*, 12(4):1017–1054, 1999.

[24] A. Frieze. On the independence number of random graphs. *Discrete Mathematics*, 81(2):171–175, 1990.

[25] A. Frieze and C. Tsourakakis. Rainbow connection of sparse random graphs. *Electronic Journal of Combinatorics*, 19:P5, 2012.

[26] E. Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

[27] C. Greenhill, B. McKay, and X. Wang. Asymptotic enumeration of sparse 0–1 matrices with irregular row and column sums. *Journal of Combinatorial Theory, Series A*, 113(2):291–324, 2006.

[28] A. Hajnal and E. Szemerédi. Proof of a conjecture of P. Erdős. *Combinatorial Theory and its Applications*, 2:601–623, 1970.

[29] J. He and H. Liang. On rainbow-k-connectivity of random graphs. *Information Processing Letters*, 112:406–410, 2012.

[30] A. Heckel. The chromatic number of dense random graphs. *Preprint available at* `arxiv.org/abs/1603.04836`.

[31] A. Heckel and O. Riordan. The hitting time of rainbow connection number two. *Electronic Journal of Combinatorics*, 19(4):P37, 2012.

[32] A. Heckel and O. Riordan. On the threshold for rainbow connection number r in random graphs. *Graphs and Combinatorics*, 32(1):161–174, 2016.

[33] S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs.* Wiley, New York, 2000.

[34] N. Kamčev, M. Krivelevich, and B. Sudakov. Some remarks on rainbow connectivity. *Journal of Graph Theory*, 2015.

[35] R. Kang and C. McDiarmid. Colouring random graphs. In *Topics in Chromatic Graph Theory*, volume 156 of *Encyclopedia of Mathematics and Its Applications*, pages 199–229. Cambridge University Press, 2015.

[36] S. Krantz. *Handbook of Complex Variables.* Birkhäuser, Boston, 1999.

[37] M. Krivelevich and B. Patkós. Equitable coloring of random graphs. *Random Structures and Algorithms*, 35(1):83–99, 2009.

[38] X. Li and Y. Sun. *Rainbow Connections of Graphs.* SpringerBriefs in Mathematics. Springer, New York, 2012.

[39] T. Łuczak. The chromatic number of random graphs. *Combinatorica*, 11(1):45–54, 1991.

[40] T. Łuczak. A note on the sharp concentration of the chromatic number of random graphs. *Combinatorica*, 11(3):295–297, 1991.

[41] D. Matula. On the complete subgraphs of a random graph. In *Proceedings of the 2nd Chapel Hill Conference on Combinatorial Mathematics and its Applications (Chapel Hill, NC, 1970)*, pages 356–369, 1970.

[42] D. Matula. The employee party problem. *Notices of the American Mathematical Society*, 19(2):A–382, 1972.

[43] D. Matula. Expose-and-merge exploration and the chromatic number of a random graph. *Combinatorica*, 7(3):275–284, 1987.

[44] D. Matula and L. Kučera. An expose-and-merge algorithm and the chromatic number of a random graph. In *Random Graphs '87 (Poznań 1987)*, pages 175–187. Wiley, 1990.

[45] C. McDiarmid. On the method of bounded differences. *Surveys in Combinatorics*, 141(1):148–188, 1989.

[46] C. McDiarmid. On the chromatic number of random graphs. *Random Structures & Algorithms*, 1(4):435–442, 1990.

[47] B. McKay. Asymptotics for 0-1 matrices with prescribed line sums. *Enumeration and Design*, pages 225–238, 1984.

[48] R. Paley and A. Zygmund. A note on analytic functions in the unit circle. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 28, pages 266–272. Cambridge University Press, 1932.

[49] R. Paley and A. Zygmund. On some series of functions, (3). In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 28, pages 190–205. Cambridge University Press, 1932.

[50] K. Panagiotou and A. Steger. A note on the chromatic number of a dense random graph. *Discrete Mathematics*, 309(10):3420–3423, 2009.

[51] P. Rombach and A. Scott. Equitable colourings of random graphs. *Preprint.*

[52] A. Scott. On the concentration of the chromatic number of random graphs. *Available at* `arxiv.org/abs/0806.0178`, 2008.

[53] E. Shamir and J. Spencer. Sharp concentration of the chromatic number on random graphs $G_{n,p}$. *Combinatorica*, 7(1):121–129, 1987.

[54] M. Talagrand. *Mean field models for spin glasses. Volume I: Basic examples*, volume 54 of *Ergebnisse der Mathematik und ihrer Grenzgebiete, 3. Folge.* Springer, 2010.