VILNIUS UNIVERSITY

VALENTAS KURAUSKAS

ON TWO MODELS OF RANDOM GRAPHS

Doctoral dissertation
Physical sciences, mathematics (01P)

Vilnius, 2013

The scientific work was carried out in 2009–2013 at Vilnius university.

Scientific supervisor:

prof. habil. dr. Mindaugas Bloznelis (Vilnius university, physical sciences, mathematics – 01P).

VILNIAUS UNIVERSITETAS

VALENTAS KURAUSKAS

DU ATSITIKTINIŲ GRAFŲ MODELIAI

Daktaro disertacija
Fiziniai mokslai, matematika (01P)

Vilnius, 2013 metai

Disertacija rengta 2009–2013 metais Vilniaus universitete.

Mokslinis vadovas:

prof. habil. dr. Mindaugas Bloznelis (Vilniaus universitetas, fiziniai mokslai, matematika – 01P).

Foreword

This thesis consists of two parts, each devoted to a different model of random graphs. Part I contains topics I have worked on under the supervision of Mindaugas Bloznelis in Vilnius University. Part II originated from the collaboration with my MSc supervisor in the University of Oxford, Colin McDiarmid. I am very grateful to both of my supervisors for all the help I received from them.

I would like to thank my parents, my brother and especially my wife Rosita for moral support during PhD. I would like to thank my colleagues Matas, Tomas and Paulius, who helped to keep the subject of combinatorics alive in Vilnius. I also thank Malcolm Strens for being a great supervisor in my previous company, for his help in combining my work and studies. My kind thanks also go to Jerzy Jaworski, Michał Karoński and Katarzyna Rybarczyk from Adam Mickiewicz University for making us feel welcome in Poznań.

Contents

Introduction

Formal introduction

This dissertation consists of two parts. The object of the first part of the dissertation is random intersection graphs and random intersection digraphs. The goal of the work was to determine certain asymptotic properties of such graphs (or digraphs). They include (a) the birth threshold for fixed-size complete subgraphs in the random intersection digraph; (b) the clique number of sparse random intersection graphs; (c) the chromatic number of random uniform intersection graphs. An additional goal was to better understand the connection of random intersection graphs and large real-world networks.

Random intersection graphs have been actively studied in the last decade. It has been shown that this model can produce instances with positive clustering coefficient and other commonly observed properties of real-world networks (such as the Internet, social and biological networks). The applications of such models include wireless networking, classification and epidemiology, see [21,22,46,53]. The work is also relevant from the computer science point of view: we consider classical NP-hard problems, but we restrict attention to a particular, rather general and practically important family of distributions of graphs.

The object of the second part of the dissertation is minor-closed classes of graphs without $k+1$ disjoint minors in $\mathcal{B}$, where a set $\mathcal{B}$ consists of 2-connected graphs. The problem of this part is to enumerate such classes asymptotically and prove properties of typical graphs in them. We study two general types of $\mathcal{B}$. As part of the work, we aim to answer a question of Bernardi, Noy and Welsh in this case.

The results in Part II build on the work of McDiarmid on addable minor-closed classes. The theory of graph minors has many applications in theoretical computer science and "has made a fundamental impact both outside the graph theory and within" [38]. Asymptotic enumeration of minor-closed classes was originally motivated by a particular case relevant both theoretically and practically, the planar graphs. Results of this kind can usually be directly applied to average-case complexity analysis of graph algorithms where the input is a uniformly random graph with some natural restrictions [12, 40, 48]. Another algorithmic application highlighted in the literature is as follows. The ability to count often gives knowledge how to construct large instances [52,85]. This can be used for system testing. The proofs in the second part are mostly based on combinatorial and probabilistic arguments (as opposed to the approach that uses mainly analysis of generating functions), and the results often hold with rather general conditions.

In the next two sections we specify the models and review the propositions that we prove in the thesis, this is done for each part separately.

Random intersection graphs

Let $S_1, S_2, \ldots, S_n$ be finite sets. The pairs uv where $u \neq v$ and $S_u \cap S_v \neq \emptyset$ define edges of a graph on the vertex set $[n] = \{1, \ldots, n\}$. This graph is called the *intersection graph* of $S_1, \ldots, S_n$, see Figure 1.

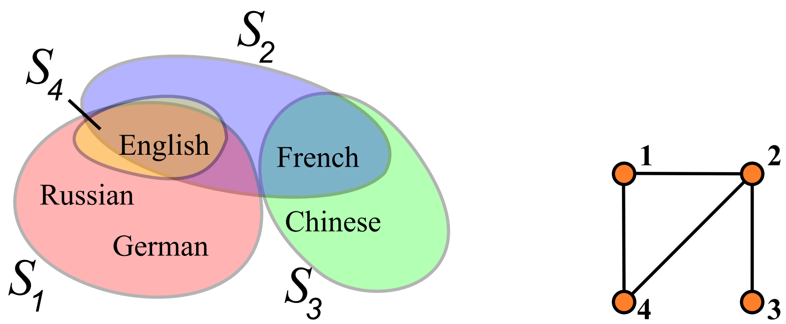

Figure 1: An intersection graph representing all communicating pairs, when, for example, S_v is the set of languages spoken by person v.

If the sets $S_1, \ldots, S_n$ are random subsets of some (finite) ground set W of m *attributes* (or keys), we obtain a random intersection graph. The first authors to consider such random graphs were Karoński, Scheinerman and Singer-Cohen (1999, [58]). They studied the *binomial random intersection graph* model $G(n, m, p)$, where an attribute w is added to the set S_v independently at random with probability p, for each pair (w, v), $w \in W$ and $v \in [n]$.

Godehardt and Jaworski [53] introduced a more general "active" random intersection graph $G(n, m, P)$, where each set S_v is generated independently at random in two stages: first the size X_v is drawn according to the probability distribution P, then a uniformly random subset S_v of size X_v is drawn (without replacement) from W. We give a more detailed description of this and related models in Section 1.1.

Studying random intersection graphs is motivated by a belief that they share some properties with large empirical networks. Large empirical networks are often observed to be sparse (the average number of neighbours of a vertex is bounded) and have a non-negligible clustering coefficient (which is the conditional probability that three randomly chosen vertices make up a triangle, given that the first two are neighbours of the third one). Unlike in many other models, the parameters of random intersection graphs can be chosen in a way that the resulting random instances have these two properties simultaneously.

For any collection $\{S_1, S_2, \ldots, S_n\}$ of subsets of W there is a unique *dual* collection $\{T_w\}_{w \in W}$ of subsets of $[n]$, where $T_w = \{v : w \in S_v\}$. In terms of intersection graphs, it can be understood as follows: each attribute $w \in W$ corresponds to a clique on the vertex set T_w; edges of all the cliques T_w define the set of edges of the intersection graph. We call cliques T_w *monochromatic*.

In their paper Karoński, Scheinerman and Singer-Cohen determined for which choices of p, the binomial random intersection graph $G(n, m, p)$ contains a clique on h vertices with high probability (when n and m are large). They solved the problem for any fixed h by showing that it is enough to consider a finite number of configurations of pairwise intersecting sets. Figure 2 shows two such configurations in the case $h = 4$.

Formalisms. Statements such as "$G(n, m, p)$ has a clique of size h with high probability" should be rigorously interpreted as follows. We consider a sequence of random graphs $\{G(n), n = 1, 2, \ldots\}$, where $G(n) = G(n, m, p)$ and $m = m(n)$, $p = p(n)$. For $n = 1, 2, \ldots$ we let $A = A(n)$ be the event that $G(n)$ has a clique of size h. Then a statement like "A holds with high probability" means that $\mathbb{P}(A(n)) \to 1$ as $n \to \infty$. Informal statements about $D(n, m, p_-, p_+)$ and $G(n, m, P)$ should be interpreted similarly. For example, when we talk about the parameter P, we actually have in mind a sequence of probability measures $\{P(n), n = 1, 2, ..\}$.

For a sequence $\{X_n, n = 1, 2, \ldots\}$ of random variables (for example X_n may be the size of a maximum clique in $G(n, m, p)$), we informally write that X_n is "asymptotically" $f(n)$ if for any $\epsilon > 0$ $\mathbb{P}(|X_n - f(n)| > \epsilon f(n)) \to 0$ as $n \to \infty$. A standard notation $X_n = f(n)(1 + o_P(1))$ will be used in the subsequent chapters.

Small subgraphs in random intersection digraphs

In Chapter 2 we ask a similar question for a related *binomial random intersection digraph* model $D(n, m, p_-, p_+)$. In this model, proposed by Bloznelis [16], each vertex $v \in [n]$ is assigned not one, but two random subsets, S_v^- and S_v^+. Each attribute $w \in W$ is included into S_v^- with probability p_- and into S_v^+ with probability p_+ independently. Then the random binomial intersection digraph is a directed graph on the vertex set $[n]$ with arcs $\{uv : S_u^- \cap S_v^+ \neq \emptyset\}$. Such a digraph makes sense if we interpret S_v^- and S_v^+ as sets of attributes (qualities) that v "likes" and "possesses" respectively.

We determine ranges of parameters for which $D(n, m, p_-, p_+)$ contains a copy of the complete directed graph on h vertices with a high probability. Depend-

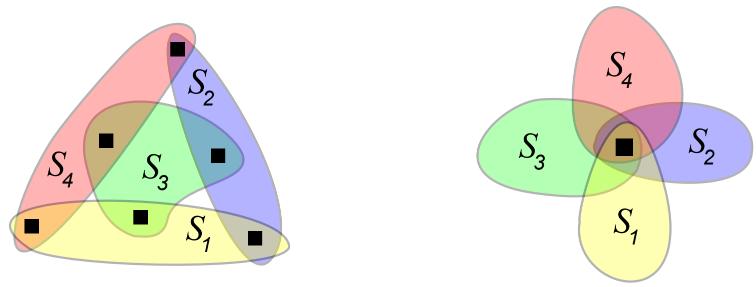

Figure 2: Two different configurations of intersecting sets that yield a clique on four vertices.

ing on the relationship between parameters p_-, p_+ and m (all of them can vary with n), four different patterns of intersecting sets can be most likely to realise the complete digraph on h vertices, and two of these patterns do not have an undirected counterpart.

Largest clique

But what can we say about the size of a maximum clique (*the clique number*) of a random intersection graph? In general, the problem is difficult due to two reasons. Firstly, the local clustering property causes a lot of dependence between the edges of the random graph. Secondly, the clique number can grow with n, and so it no longer suffices to consider only a finite number of patterns of sets. In Chapter 3 a solution for *sparse* uniform random intersection graphs $G(n, m, P)$ is presented. Here "sparse" means that m and P are such that the expected number of edges is linear in n.

Let $D(n)$ be the degree of a random vertex (or, equivalently, of vertex 1) in $G(n, m, P)$. We find that the clique number of $G(n, m, P)$ depends on the tail of the distribution of $D(n)$. If $D(n)$ is "asymptotically" power-law distributed with index $\alpha \in (1, 2)$ (for example, a Pareto distributed random variable X with $\mathbb{P}(X > t) = t^{-\alpha}$ is power-law with index α) then the largest clique is "asymptotically" of polynomial size. The order of the clique number in this case is the same as in a much simpler model without clustering studied by Janson, Łuczak and Norros (2010, [55]).

Meanwhile, if the degree variance is bounded ($\sup_n VarD(n) < \infty$), then the largest clique is with high probability "almost" monochromatic (generated by a single attribute, as in Figure 2, right) and its size is "asymptotically" logarithmic. This phenomenon is specific to random intersection graphs, and the clique number here is closely related to the maximum load problem: if N balls are thrown randomly to m bins, what is the maximum number of balls a bin receives? Both in the "power-law" and the "bounded variance" regimes our results are optimal up to the first-order asymptotic term. These two regimes cover most of the interesting choices of the parameters for sparse $G(n, m, P)$.

Furthermore, for each of the two main regimes there is a simple algorithm for finding large cliques. We prove that with high probability $G(n, m, P)$ is such, that the corresponding algorithm outputs a clique of asymptotically optimal size and terminates in polynomial time. These algorithms have a potential to be used and studied with large scale real-world graphs. A reader interested to see the simple pseudocode is welcome to jump directly to Section 3.4.

Chromatic index of random uniform hypergraphs

A *hypergraph* H is a pair (V, E), where V is a set and E is a collection of subsets of V called hyperedges, or simply *edges*. Intersection graphs are hypergraphs, where we put emphasis on pairwise intersections of edges. The *chromatic number* of a graph is the least number of colours needed to colour its vertices, so that no two neighbours receive the same colour. The *chromatic index* of a hypergraph is the least number of colours needed to colour its edges so that no pair of intersecting edges receives the same colour.

In Chapter 4 we study the chromatic index of $\mathbb{H}^{(k)}(m, n)$, the random hypergraph on the vertex set $[m]$ and with n edges drawn independently with replacement from all subsets of $[m]$ of size k. The problem is equivalent to the problem of determining the chromatic number of $G(n, m, k)$, the uniform random intersection graph with n vertices, m attributes and all subsets of size k. In the case when k is constant and n is much larger than m, a result by Pippenger and Spencer (1989, [82]) implies the answer. That result holds for arbitrary 'almost regular' hypergraphs, not just the random ones. For random hypergraphs we extend their result slightly and allow k to grow slowly with n. To do this, we exhibit a simple greedy algorithm (different from Pippenger and Spencer's one) and prove that it colours the edges with an (asymptotically) optimal number of colours.

Empirical aspects

In Section 1.2 we plot certain statistics for large real-world networks (such as, for example, the actor affiliation networks, where two actors are declared adjacent if they had a role in the same movie) and random intersection graphs with corresponding parameters. The statistics include assortativity and counts of pairs with given number of common neighbours.

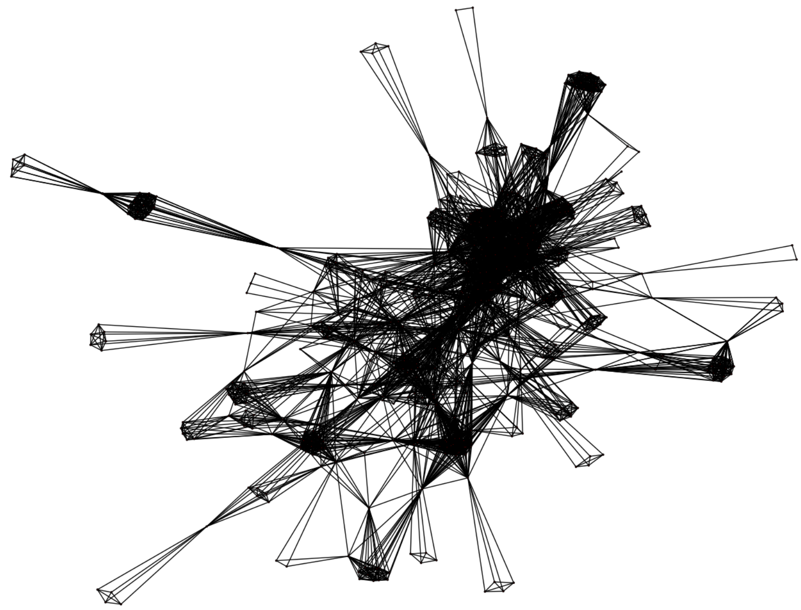

Figure 3: The Lithuanian actor affiliation network (data from IMDB: `http://www.imdb.com`). The 'union of cliques' structure, where each clique consists of actors participating in the same film, can clearly be seen here.

Minor-closed classes of graphs

The second part of this thesis is concerned with graphs that do not contain certain subgraphs.

Connected graphs that do not have any cycle are *trees*. Acyclic, but not necessarily connected graphs are called *forests*. Given a class of labelled graphs $\mathcal{A}$ (for example, the class of trees), we denote by $\mathcal{A}_n$ the restriction of $\mathcal{A}$ to graphs on the vertex set $[n] = \{1, .., n\}$. We study

(*) the asymptotic number of graphs in $\mathcal{A}_n$;

(**) the structure of a typical graphs in $\mathcal{A}$; more precisely, properties of a uniformly random graph from $\mathcal{A}_n$.

For example, the classic result of Cayley (1868) states that there are n^{n-2} trees on the vertex set $[n]$. Rényi (1959) proved that the number of forests on the same vertex set is $\sqrt{e}n^{n-2}(1 + o(1))$ as n tends to infinity.

Part II starts with investigation of the class of graphs that do not have $k + 1$ vertex-disjoint cycles. Erdős and Pósa (1965, [45]) showed that there is a constant c_k such that each graph that does not contain $k + 1$ disjoint cycles has a set of at most c_k vertices, whose removal results in an acyclic graph (a forest). It is known that the smallest possible c_k is of order $k \ln k$ [38]. (The important thing here is that no matter how large a graph is, if it has at most k disjoint cycles then we can "destroy" all of its cycles by removing just a few vertices.)

In Chapter 6 we present a proof that only k vertices are enough for typical graphs without $k+1$ disjoint cycles. A uniformly random such graph on the vertex set $[n]$ for large n is shown to be very close in distribution to the following simple construction a) pick a uniformly random set $S \subset [n]$ of size k; b) put a uniformly random forest on the remaining vertices $[n] \setminus S$; c) for each pair $\{x, y\} \subset [n]$ with at least one element in S, add the edge xy independently at random with probability $1/2$.

Given a graph G, the *contraction* of an edge $e = xy \in E(G)$ is the following operation: merge the endpoints x and y of e into a new vertex v_{xy}, so that v_{xy} becomes adjacent to all of the former neighbours of x and y. A graph H is called a *minor* of G if it can be obtained from G by applying a series of edge deletions, vertex deletions and edge contractions, see Figure 4.

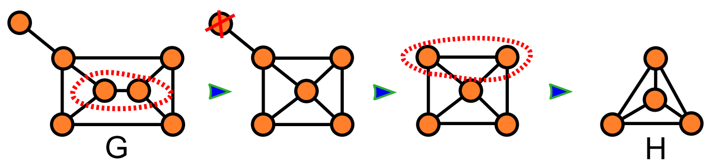

Figure 4: A sequence of vertex deletions and edge contractions showing that H is a minor of G.

A class of graphs $\mathcal{A}$ is *minor-closed* if for any $G \in \mathcal{A}$ every minor of G also belongs to $\mathcal{A}$. Minor-closed classes of graphs is the subject of the theory of graph minors developed by Robertson and Seymour in a series of more than twenty papers (1983-2004). One of the results is the following: each minor-closed class $\mathcal{A}$ can be characterised by a finite list $\mathcal{B}$ of minimal *excluded minors*. That is, to test whether a graph G is *not* in the class $\mathcal{A}$, it suffices to check whether any of the finitely many graphs in $\mathcal{B}$ is a minor of G. We denote this by $\mathcal{A} = \operatorname{Ex} \mathcal{B}$. For example, by an earlier work of Kuratowski (1930) and Wagner (1937) the class of planar graphs (graphs drawable on the plane so that edges can intersect only at their endpoints) can be characterised by two minimal non-planar excluded minors: $K_{3,3}$ and K_5 (here $K_{t,t}$ is the complete bipartite graph with both parts of size t and K_t is the complete graph on t vertices). Counting and studying properties of random planar graphs and other minor-closed classes of graphs has received a lot of attention in the last decade, we review the work most relevant to this thesis in Chapter 5. Some other examples of minor-closed classes are forests, series-parallel graphs, outerplanar graphs, graphs embeddable in a fixed surface (for example, the torus), graphs with a bounded treewidth, graphs knotlessly embeddable in

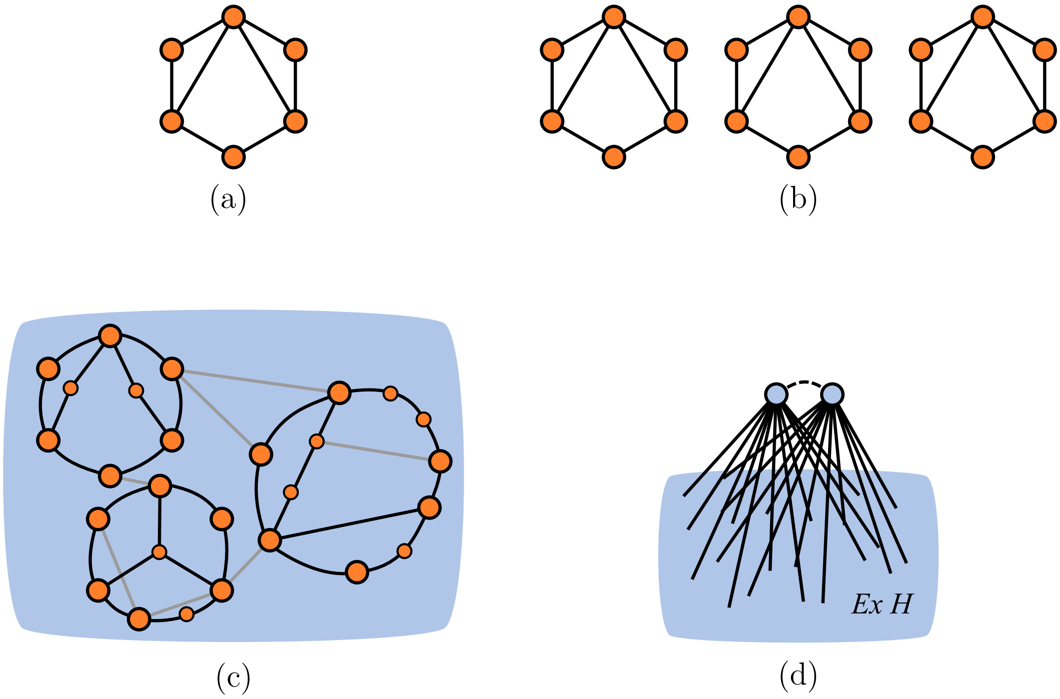

Figure 5: Illustration of a class of graphs handled in Chapter 7: (a) the graph H; (b) the forbidden minor: 3 disjoint copies of H; (c) a graph with three disjoint forbidden minors; (d) a typical graph without three disjoint minors H.

Euclidean 3-space, etc.

The class of graphs containing at most $k+1$ disjoint cycles is also minor-closed; the forbidden minor is $k+1$ disjoint copies of K_3. Chapter 7 generalizes results of Chapter 6 to classes with at most k disjoint excluded minors from a given fixed set $\mathcal{B}$ (with repetitions allowed). For the generalisation to work, the excluded minors in $\mathcal{B}$ have to necessarily satisfy a certain restriction: the class of graphs $\operatorname{Ex}\mathcal{B}$ must not contain arbitrarily large fans (a fan is a graph consisting of a path together with a vertex joined to each vertex on the path).

We postpone the formal statements of our theorems until Chapter 7; now we will just discuss one example, a straightforward application of our results with $k = 2$ and a set $\mathcal{B}_0 = \{H\}$, consisting of a particular graph H on six vertices shown in Figure 5 (a). Our result concerns the class $\mathcal{A}$ of graphs with the excluded minor shown in Figure 5 (b). Graphs that violate the requirement are, for instance, as in Figure 5 (c) (two of the subgraphs are *subdivisions* of H, the third can be seen to have H as a minor by contracting two edges incident to vertices marked with the smaller circles).

Our result implies that a uniformly random graph in $\mathcal{A}_n$ essentially consists of a random graph in $\operatorname{Ex}\mathcal{B}$ on $n-2$ vertices, two "apex vertices", and edges incident to each of the apex vertices which appear independently with probability 1/2.

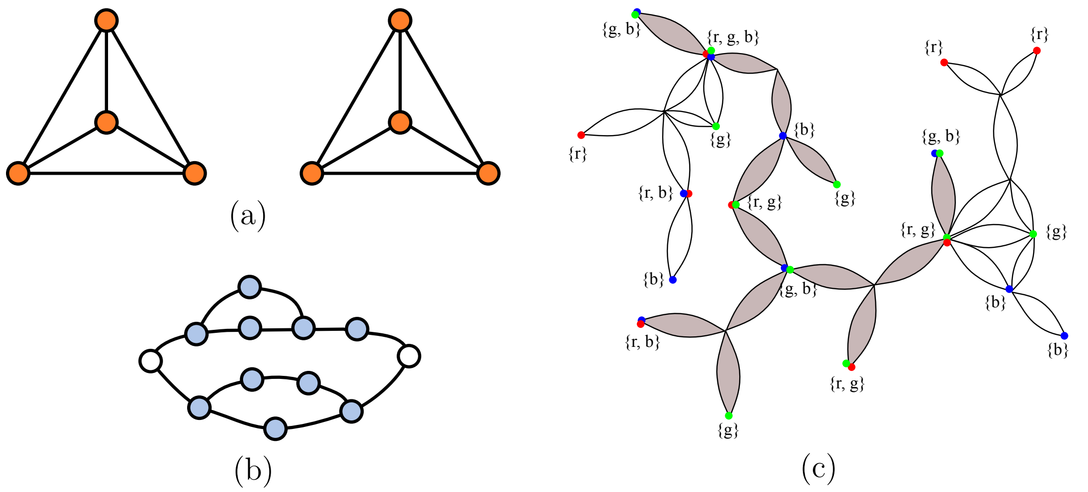

Figure 6: (a) A forbidden minor: two disjoint K_4; (b) a series-parallel network; (c) the "core" of a typical graph without two disjoint minors K_4. To complete the graph, add three new "apex" vertices x, y, z and join them to each node coloured r[ed], g[reen] and b[lue] respectively; replace leaf-like shapes by (non-series) series-parallel networks, and attach more series-parallel graphs at each vertex arbitrarily. Neither of x, y, z is allowed to create a K_4 minor alone.

With very high probability these two apex vertices are the only vertices that have linearly many neighbours, see Figure 5 (d).

The last two chapters are devoted to the next layer of disjoint forbidden minors. In Chapter 8 we prove results for general classes $\mathcal{B}$, such that $\mathrm{Ex}\,\mathcal{B}$ contains all fans, but $\mathcal{B}$ is "good enough". We show that a different general construction dictates the asymptotic number of graphs without $k+1$ disjoint excluded minors in $\mathcal{B}$. The motivating case behind quite general results of Chapter 8 was a particular set $\mathcal{B} = \{K_4\}$. The graphs without a minor K_4 are known as *series-parallel* graphs. This class is important in computer science, and has been used to study algorithms for problems that are hard for general graphs.

In Chapter 9 we obtain precise first-order asymptotics for the number of graphs without $k+1$ disjoint minors K_4. We prove that for large n, a typical such graph G on $\{1, \dots, n\}$ has a unique set S with the following properties:

(i) S is of size $2k+1$;

(ii) for any $x \in S$, the removal of $S \setminus \{x\}$ from G results in a series-parallel graph;

(iii) each vertex in S has linearly many neighbours.

A rather complete picture of typical graphs from this class is developed. Figure 6 illustrates some of the basic features.

Overview of the methods

Many different methods are used throughout the dissertation. The most important of them are discussed in more detail in the background chapters, Chapter 1 and Chapter 5.

Inequalities from probability theory and probabilistic method. We use various classic probabilistic methods such as the first and second moment method, Chebyshev's inequality quite intensively in Part I of the thesis. In addition, we use some less standard results from Ramsey theory and the theory of random graphs.

Concentration inequalities. In both parts an indispensable tool is Chernoff bounds for sums of independent random variables. Most of the applications require only that the bounds are exponential, the constant in the exponent is not important. In Chapter 4 we use more sophisticated concentration inequalities for martingales, from McDiarmid (1998) [72].

Differential equations method. This method is based on concentration inequalities for martingales. It was developed by Wormald [101] in the context of random graph processes, though Karp and Sipser had applied similar techniques in their earlier work [59]. The main idea is to show that a trajectory of a parameter of a random process is highly concentrated around its mean at all time steps. The curve for the mean is a solution of a system of differential equations. We apply this method for the random hypergraph edge colouring problem.

Theory of graph minors. Several major results in the theory of graph minors by Robertson and Seymour, see [89], are the starting point in the proofs of Part II. One of the key facts is that graphs with a planar excluded minor have a bounded tree-width.

Singularity analysis. Methods based on generating functions play an important role in Part II. While in Chapters 6, 7 and 8 we only make use of very simple results, such as the "exponential formula", Chapter 9 contains a full application of the singularity analysis method: we obtain decompositions of relevant classes, convert them to exponential generating functions and use either general theorems or our own observations in complex analysis to extract the asymptotic coefficients. Most of the tools of this kind can be found in the book "Analytic Combinatorics" by Flajolet and Sedgewick [48]. For counting tree-like structures we also find the work of Meir and Moon (1989, [79]) very general and useful.

Computational tools. We used computer to aid some of our work. Numerical estimates presented in Part II were carried out with the symbolic computation system `Maple`. Empirical analysis of real networks in Section 1.2 required larger scale computation, this was implemented using `Python` with its packages `numpy` and `matplotlib` and executed in the cluster of the Digital Science and Computing Centre of the Faculty of Mathematics and Informatics, Vilnius University. Some programming with `C++` and `Python` was used to explicitly construct all possible graphs related to classes with few disjoint minors K_4. Most of the illustrations in this thesis were created using `Xfig` and `Inkscape`. The dissertation itself was prepared and compiled with `XeLaTeX`.

The methods and proofs presented in this thesis are mathematically rigorous. The empirical parameters of large real networks were only evaluated for particular graphs and served mainly for illustratory purposes. Statistical inference or hypothesis testing using random intersection graph models can be seen as a potential future work in the area.

Content, originality and novelty

The content presented in this thesis has been created and prepared by the author of the thesis together with his co-authors.

The results obtained in the dissertation are original and all of them can be considered as new. Most of the problems of Part I had been considered by other authors with related but different models. In our work we propose several methods not used in this context before (applications of the balls and bins problem, extremal combinatorics and differential equations). The important phenomenon of largest clique being generated by a single attribute has been earlier also discovered by two other groups of researchers in related but more restricted models. Part II explores an entirely new type of minor-closed classes and similar results were unknown before. Some of the intermediate lemmas extend previously known ones.

Each of the results chapters is based on an article that has either been published or submitted for publication.

Section 1.2 – on paper [20] with M. Bloznelis and J. Jaworski and paper [24] with M. Bloznelis[1].

Chapter 2 – on paper [62];

[1]In these two works I essentially carried out only the empirical analysis.

Chapter 3 – on paper [25] with M. Bloznelis;

Chapter 4 – on paper [66] with K. Rybarczyk.

Chapter 6 – on paper [64] with C. McDiarmid;

Chapter 7 – on paper [65] with C. McDiarmid,

Chapter 8 and Chapter 9 – on paper [63].

The papers [24,62,64,65] have been published, the remaining papers have been submitted for publication.

Part I

Random intersection graphs

Chapter 1

Background

1.1 Random intersection graphs

In this section we briefly review some of the literature on random intersection graphs. For a more complete overview see upcoming survey papers [21, 22].

1.1.1 Models

Given positive integers n and m, and a probability measure P on $\{0, 1, \ldots, m\}$, the *random intersection graph* (also called the active random intersection graph) $G(n, m, P)$ on vertex set $V = \{1, 2, \ldots, n\}$ and with the set $W = \{w_1, w_2, \ldots, w_m\}$ of attributes is defined as follows. Let $S_1, S_2, \ldots, S_n$ be independent random subsets of W such that for any $v \in V$ and any $S \subseteq W$ we have $\mathbb{P}(S_v = S) = P(|S|)/\binom{m}{|S|}$. The edge set of $G(n, m, P)$ consists of those pairs $\{u, v\} \subseteq V$ for which $S_u \cap S_v \neq \emptyset$.

The model $G(n, m, P)$ is rather general. If the distribution P is the binomial distribution with parameters p and m, we obtain the binomial random intersection graph $G(n, p)$ studied in [58] and later by many other authors, including [37, 81, 92, 97]. If P is a degenerate distribution such that $P(k) = 1$ for a constant k, then we obtain the *uniform random intersection graph* $G(n, m, k)$ as, e.g., in [14] and many other papers, see [21]. Random uniform hypergraphs are closely related to the last model.

Several further natural extensions and modifications have been considered. An intersection threshold $s \geq 1$, may be required to create an edge [18] (i.e., vw is an edge if and only if $|S_v \cap S_w| \geq s$). A model dual to $G(n, m, P)$ and denoted $G^*(n, m, P)$ is called the *passive random intersection graph.* It has the vertex set W and for a pair $\{w', w''\}$ of distinct elements from W, $w'w''$ is an edge if and only if there is at least 1 (or, more generally, at least s) sets containing both w' and w''. More recent research, i.e., [19, 23, 27], focused on *inhomogeneous* random intersection graphs where both attributes and vertices are assigned random weights and an attribute w is added to the set S_v with probability proportional to the product of the weights of w and v.

The model of Chapter 2 is an adaptation of the binomial random intersection $G(n, m, p)$ that yields random directed graphs. The model of Chapter 3 is $G(n, m, P)$, and the model we will meet in Chapter 4 is $G(n, m, k)$.

1.1.2 Degrees, clustering and sparseness

A notable fact about these models is that they can be used to obtain sparse graphs with a power-law asymptotic degree distribution and a positive clustering coefficient.

Consider a sequence $\{G(n), n = 1, 2, \dots\}$ of random intersection graphs where $G(n) = G(n, m, P)$, $m = m(n)$ and $P = P(n)$. Let $X(n)$ (the size of a random set) be distributed according to $P(n)$, and write $Y(n) = n^{1/2}m^{-1/2}X(n)$. We will consider the following conditions

$$Y(n) \text{ converges weakly to a random variable } Y^* \text{ with mean } \mu; \tag{1.1}$$

$$\lim_{n\to\infty} \mathbb{E}\, Y(n) = \mathbb{E}\, Y^*. \tag{1.2}$$

The limiting degree distribution for active random intersection graphs was determined in increasing generality by Stark [97], Deijfen and Kets [37], Bloznelis [15, 16, 18] and Rybarczyk [93].

Theorem 1.1.1 (Bloznelis, [15, 18]) *Let $\{G(n), n = 1, 2, \dots\}$ be a sequence of random intersection graphs, such that (1.1) and (1.2) hold. Write $\mu = \mathbb{E}\, Y^*$.*

Then the degree $D(n)$ of vertex 1 in $G(n)$ converges weakly to a random variable D^*, which has a Poisson distribution with random intensity μY^*: $\mathbb{P}(D^* = k) = (k!)^{-1}\mathbb{E}\,(\mu Y^*)^k e^{-\mu Y^*}$.

Thus, if the random variable Y^* is power-law distributed and the conditions of the last theorem hold, then so is D^* [15]. If, additionally, $\mu = \mathbb{E}\, Y(n) = O(1)$, then the limiting degree is bounded (the graph is sparse). Let us remark that [18] considers more general intersection thresholds $s \geq 1$, and both the active and the passive random intersection graphs. Here and below for simplicity we present simplified results for the case $s = 1$ and the active random intersection graph only.

The clustering coefficient $\alpha(G)$ of a graph G with minimum degree at least two is defined as follows. Let (v_1^*, v_2^*, v_3^*) be a uniformly random triple of distinct vertices from G. Then

$$\alpha(G) = \mathbb{P}(v_1^* v_2^* \in E(G) \mid v_1^* v_3^*, v_2^* v_3^* \in E(G)).$$

The fact that $G(n, m, P)$ admits a clustering property (the clustering coefficient remains bounded away from zero as $n, m \to +\infty$) was observed by Deijfen and Kets [37]. The asymptotic properties of $\alpha(G)$ were later studied in [18].

Their results show that the parameters of $G(n, m, P)$ can be chosen so that its limiting clustering coefficient is an arbitrary positive number in the interval $[0, 1)$, and the limiting degree distribution is power-law. These two properties are reported to be present in many large scale real networks (complex networks). Other properties studied in random intersection graphs include formation of the giant component [5, 17, 27], connectivity [14] and diameter, see [22].

1.2 Relation to empirical networks

An interesting link between sparse[1] random intersection graphs and empirical networks was discovered by Bloznelis [18]. Given a graph G with $|V(G)| \geq 3$ and a random triple of vertices (v_1^*, v_2^*, v_3^*) as before, define the *conditional clustering coefficient*

$$\alpha^{[k]}(G) = \mathbb{P}(v_1^* v_2^* \in E(G) \mid v_1^* v_3^*, v_2^* v_3^* \in E(G), d(v_3^*) = k)$$

for each $k = 2, 3, \ldots$ such that $\mathbb{P}(d(v_3^*) = k) > 0$.

Bloznelis obtained a simple expression for $\alpha^{[k]}(G(n, m, P))$ and pointed out that for sparse random intersection graphs $G(n, m, P)$ with a power-law asymptotic degree distribution we have $\lim_{n\to\infty} \alpha^{[k]}(G(n, m, P)) \sim ck^{-1}$ as $k \to \infty$ where $c > 0$ is a constant. Remarkably, the shape of $\alpha^{[k]}$ obtained rigorously agrees with its shape for a wide variety of social networks observed earlier by other authors, see [18].

This result encouraged studying other measures commonly used in empirical networks. The correlation coefficient of the degrees of neighbouring nodes of a graph is called the *assortativity coefficient*. A very similar quantity, for a graph

[1] In this section *sparse* refers to sequences satisfying (1.1).

G and a uniformly random pair (v_1^*, v_2^*) of its vertices, is

$$r(G) = \frac{\mathbb{E}\, d(v_1^*)d(v_2^*) - (\mathbb{E}\, d(v_1^*))^2}{\mathbb{E}\, d(v_1^*)^2 - (\mathbb{E}\, d(v_1^*))^2}.$$

Theorem 1.2.1 ([20]) *Suppose $\{G(n), n = 1, 2, \dots\}$ is a sequence of random intersection graphs satisfying (1.1),*

$$\frac{m}{n} \to \beta \in (0, \infty) \text{ and} \tag{1.3}$$

$$\mathbb{E}\, Y(n)^3 \to \mathbb{E}\, (Y^*)^3. \tag{1.4}$$

Then

$$r(G(n)) = \frac{\mathbb{E}\, X}{\sqrt{\beta}(\mathbb{E}\, X \mathbb{E}\, X^3 - (\mathbb{E}\, X^2)^2) + \mathbb{E}\, X^2} + o(1).$$

For large n, $r(G(n))$ is positive, which tends to also be the case in social networks, see [20].

Another quantity, related to both $r(G)$ and $\alpha^{[k]}(G)$ is

$$b_k(G) = \mathbb{E}\, (d(v_1^*) \,|\, v_1^* v_2^* \in E(G), d(v_2^*) = k).$$

Let $\{G(n), n = 1, 2, \dots\}$ be as in Theorem 1.2.1. The asymptotics of $b_k(G(n))$ were determined in [20]. For a graph G and a pair $\{u, v\} \subseteq V(G)$, we let $d(u, v)$ denote the number of common neighbours (also known as codegree) of u and v. Define

$$d_k(G) = \mathbb{E}\, (d(v_1^*) - d(v_1^*, v_2^*) \,|\, v_1^* v_2^* \in E(G), d(v_2^*) = k).$$

Surprisingly, see [20], for any $k = 1, 2, \dots$, $d_k(G(n))$ is asymptotically a constant:

$$d_k(G(n)) = \mathbb{E}\, d(v_1^*) - \mathbb{E}\, d(v_1^*, v_2^*) + o(1) \qquad \text{as } n \to \infty.$$

That is, given that v_1^* is adjacent to v_2^*, the number of neighbours of v_1^* which are at distance 2 from v_2^* is (asymptotically) not affected by the degree of v_2^*. Thus there is only "one step" dependence between degrees of vertices in a random intersection graph. Meanwhile, there seems to be a longer range dependence in social networks, see Figure 1.1. The shape of the function in the simulated random intersection graph is close to a constant as in the theoretical estimate above. For empirical networks, however, d_k seems to increase with k, at least for moderately large k.

Yet another measure was introduced in [25]. Given a graph G and a uniformly

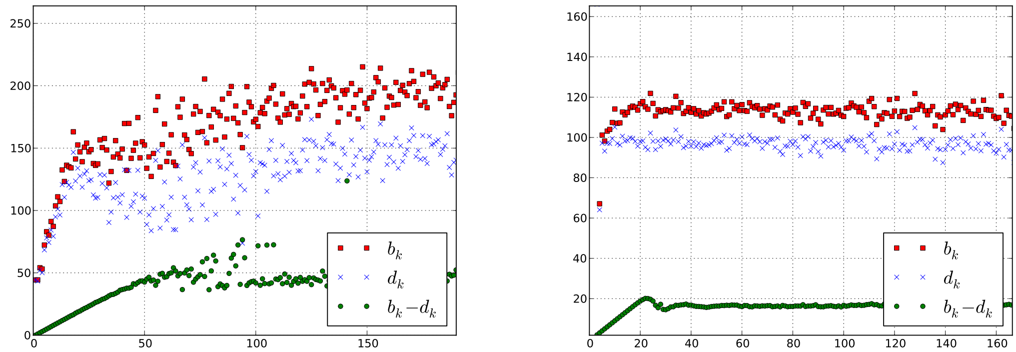

Figure 1.1: Decomposition of the conditional assortativity coefficient b_k. Left: IMDB actors affiliation network from all French language drama films ($n = 43204$ actors and $m = 5629$ films, data from `http://www.imdb.com`). Right: a random intersection graph with the same n and m and the same set sizes, where each actor reselects the subset of films of the same size as in the real data, but now independently and uniformly at random.

random pair (v_1^*, v_2^*) of its distinct vertices, the *clustering function* is defined as

$$r \mapsto cl_G(r) = \mathbb{P}(v_1^* v_2^* \in E(G) \,|\, d(v_1^*, v_2^*) = r),$$

for each r where $\mathbb{P}(d(v_1^*, v_2^*) = r) > 0$. That is, $cl_G(r)$ is the probability that two random vertices of a graph are adjacent, given that they have r common neighbours. Let $\{G(n), n = 1, 2, \dots\}$ be a sequence of random intersection graphs, such that (1.1), (1.3) hold and

$$\mathbb{E}\, Y(n)^2 \to \mathbb{E}\,(Y^*)^2.$$

Then, using Theorem 1.1.1, the degree distribution converges weakly to some random variable D^*. For a constant r, it was shown that we have convergence to a step function:

$$cl_{G(n)}(R) \to \begin{cases} 0, & \text{for } r = 0; \\ \frac{\alpha}{\alpha + (1-\alpha)e^{\Lambda}}, & \text{for } r = 1; \\ 1, & \text{for } r \geq 2. \end{cases}$$

Here α is the limit of the clustering coefficient of $G(n)$ and $\Lambda = (\mathbb{E}\, D^*/\beta)^{1/2}$.

[25] also looked at cl_G in several large empirical networks, including, IMDB actor affiliation networks and Facebook networks. In empirical data, this function was invariably found to be increasing with r, though the jump from 0 to 1 was less sharp. This was argued to be due to slow convergence of the error terms and the heavy tail of the degree distribution, see Figure 1.2.

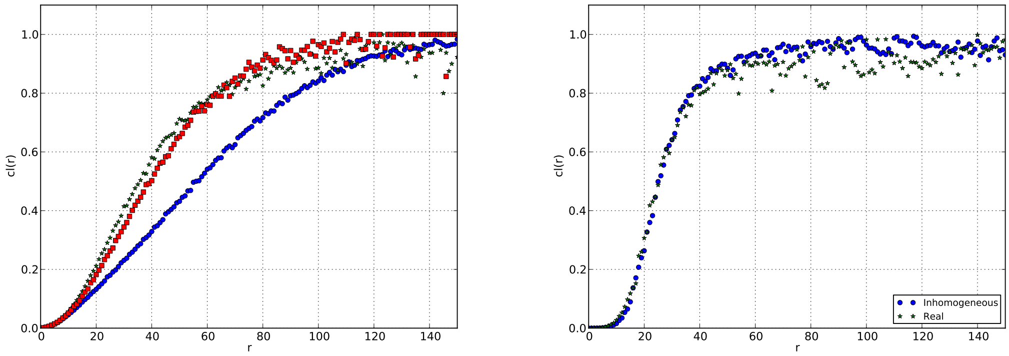

Figure 1.2: Clustering function cl_G for empirical graphs (illustrations from [25]). Left: three Facebook university networks. Right: The French drama actors affiliation network (green stars) and a corresponding random inhomogeneous intersection graph (blue circles).

1.3 Cliques and chromatic number

The clustering coefficient can be expressed as a ratio of the number of paths of length two and the number of triangles in the graph. Naturally, the clustering property of random intersection graphs propagates to larger local structures, such as cliques of size more than three.

Birth thresholds for cliques of size h, where h is fixed were determined already in the introductory paper of Karoński, Scheinerman and Singer-Cohen:

Theorem 1.3.1 ([58]) *Fix $\alpha > 0$ and consider a sequence of binomial random intersection graphs $\{G(n, m, p), n = 1, 2, \dots\}$ where $m = m(n) = \lfloor n^\alpha \rfloor$ and $p = p(n)$. Define*

$$\tau = \tau(K_h, n) = \begin{cases} n^{-1}m^{-1/h} & \text{for } \alpha < 2h/(h-1); \\ n^{-1/(h-1)}m^{-1/2} & \text{for } \alpha \geq 2h/(h-1). \end{cases}$$

$G(n, m, p)$ contains (respectively, does not contain) a copy of K_h whp if $p/\tau \to \infty$ (respectively, $p/\tau \to 0$).

It turns out, that in the first case the small cliques indeed are "born" at much smaller edge densities in $G(n, m, p)$ than, for example, in the Erdős-Rényi random graph $G(n, p)$.

Karoński, Scheinerman and Singer-Cohen also gave a more general result for arbitrary fixed size subgraphs H: they showed that one always needs to consider a finite number of "minimal" configurations of intersecting sets that "create" H, equivalently, the minimal covers of the edges of H by cliques. For $H = K_h$, one of the following two structures realises the birth threshold: (a) h sets that contain the same attribute or (b) h sets and $\binom{h}{2}$ attributes, where each of the attributes lies in one of the possible pairwise intersections.

Rybarczyk and Stark [92] used Stein's method and a nice idea to estimate and compare probabilities of different clique covers of K_h to show that the number of copies of K_h in $G(n,m,p)$ is approximately Poisson, when p is near the threshold $\tau(K_h, n)$.

The studies of clique number and the chromatic number of Erdős-Rényi random graph $G(n,p)$ have a long history and solid results. The work on these subjects continues from at least 1970s, when Bollobas, Erdős, Grimmett, Matula and McDiarmid made major early contributions. A remarkable technique developed to tackle these problems was the application of martingales in the theory of random graphs. In 1991 Łuczak settled the chromatic number problem for the remaining case of sparse random graphs. However, there still are some open questions even for $G(n,p)$. We refer to [34, 56] for the main results.

The clique number of the binomial random intersection graph $G(n,m,p)$, where $m = \lfloor n^\alpha \rfloor$, $0 < \alpha < 1$ and $mp^2 = O(1)$ was considered by Nikoletseas, Raptopoulos and Spirakis in [81]. They showed, that in this regime the largest clique is formed by a single attribute whp.

Motivated by a seemingly very different subject (the Erdős-Ko-Rado theorem), Balogh, Bohman and Mubayi [4] considered a random k-uniform hypergraph $G^k(n,p)$ where each of the $\binom{n}{k}$ hyperedges is included with probability p. They asked when $G^k(n,p)$ has the property that its maximum intersecting family is trivial, so that all sets in this family have a common element. They gave the answer for all p and a surprisingly wide range of k ($3 \le k \le n^{1/2-\epsilon}$ for any $\epsilon > 0$). Indeed, $G^k(n,p)$ was shown to have this property exactly, if $k = o(n^{1/4})$, and "approximately", if $k = o(n^{1/3})$; for larger k the property holds whp if and only if p is sufficiently large. This result is essentially equivalent to determining the clique number in the random uniform intersection graph $G(n,m,k)$ for a very wide range of parameters.

The chromatic number for $G(n,m,p)$ in the range of parameters yielding very sparse graphs was considered in [6]; some further bounds were obtained in [80]; see also Chapter 4 for more references.

Chapter 2

Small complete graphs in a random intersection digraph

2.1 Introduction

We consider the random binomial intersection digraph $D(n, m, p_-, p_+)$ defined in the Introduction. In [97] a network of co-authors of mathematical papers is mentioned as an illustration for random intersection graphs. One might alternatively define a *citation digraph* where V is a set of mathematicians and we draw an arc from u to v if and only if u has cited v. The underlying set W here would be the set of all mathematical papers; and $S^-(u)$ (respectively, $S^+(u)$) would correspond to the set of papers u has cited (respectively, co-authored).

We reviewed the practical importance of random intersection graphs in Chapter 1. In some applications considering directed intersection graphs makes sense and might lead to more precise/adequate models. In particular, one may obtain a digraph with power law indegree distribution and bounded outdegree distribution. In addition these digraphs have a clustering property when m is of order n [16].

In the problem of determining the birth threshold of small subgraphs one is interested in the question of how dense a graph should be to have a desired subgraph with certainty. There is a rich literature devoted to birth thresholds in random graphs with independent edges where each edge appears with the same probability, see, e.g., Chapter 3 of [56]. The threshold for a random (binomial) intersection graph to contain a fixed subgraph has been studied in [58].

Here we consider a similar problem for random intersection digraphs. Let $\overrightarrow{K}_h$ be the complete digraph on vertex set $[h] = \{1, \ldots, h\}$ containing arcs xy and yx for each pair of distinct vertices $x, y \in [h]$. We aim to determine critical values of the parameters for $D(n, m, p_-, p_+)$ to have with a high probability a subgraph

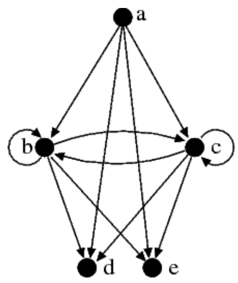

Figure 2.1: The diclique $(\{a, b, c\}; \{b, c, d, e\})$.

isomorphic to $\overrightarrow{K}_h$.

Given two finite sets C^- and C^+ we consider the ordered pair $C = (C^-; C^+)$ and the digraph $D(C)$ on the vertex set $C^- \cup C^+$ with the set of arcs $\{uv : u \in C^-, v \in C^+\}$. We call the pair C a *diclique*. We say that C is *proper* if C^-, C^+ are non-empty, otherwise say that it is *improper*. We remark that if the digraph $D(C)$ is non-empty then C must be proper and $D(C) \neq D(C')$ for $C \neq C'$. Therefore we will identify a proper diclique C with the corresponding digraph $D(C)$, see Figure 2.1.

To our knowledge, the diclique digraphs were first studied by Haralick [54], but in a different context.

In the random digraph D with vertex set V and attribute set W each attribute $w \in W$ defines a diclique $C(w) = (C^-(w); C^+(w))$ given by $C^-(w) = \{v \in V : w \in S^-(v)\}$ and $C^+(w) = \{v \in V : w \in S^+(v)\}$. It is convenient to interpret each attribute $w \in W$ as a distinct colour. Then all the attributes in W give rise to a family of dicliques of different colours which covers all arcs of D.

The paper is organised as follows. In the next section we present our main results. In Section 2.3 we give a general lemma for the birth threshold of a fixed directed graph H. In Section 2.4 we study a few special diclique covers of $\overrightarrow{K}_h$ and prove our main results Theorem 2.2.1 and Theorem 2.2.2.

We remind some standard notation used in the paper. For functions $f, g : \mathbb{N} \to \mathbb{R}_+$ we write $f \sim g$ if $\lim_{k\to\infty} f(k)/g(k) = 1$. We write $f = O(g)$ if $\limsup_{k\to\infty} f(k)/g(k) < \infty$, $f = \Omega(g)$ if $g = O(f)$ and $f = \Theta(g)$ if both $f = O(g)$ and $g = O(f)$. We write $f = o(g)$ if $f(k)/g(k) \to 0$.

Finally, thanks to an anonymous reviewer the author became aware of a related and very relevant result on the Poisson approximation of the number of cliques in sparse random intersection graphs by Rybarczyk and Stark [92].

2.2 Results

Before stating our main results we need to introduce some definitions related to dicliques. Without loss of generality we will assume that the set of vertices of the random digraph D is $V = [n]$.

For any diclique C, we call $V(C) = C^- \cup C^+$ the vertex set of C. Let $\mathcal{C} = \{C_1, C_2, \ldots, C_s\}$ be a family of dicliques, (we allow $\mathcal{C}$ to be a multiset and in this paper we consider only finite families $\mathcal{C}$). Let us denote by $V(\mathcal{C})$ the union of all vertices of the dicliques, $V(\mathcal{C}) = \bigcup V(C_i)$. We say that D *contains* $\mathcal{C}$ if there are distinct attributes $w_1, \ldots, w_s \in W$, such that $C_i \subseteq C(w_i)$ for each $i = 1, \ldots, s$ (the set operations for dicliques are defined componentwise). Also, let us call a diclique family *proper* if all its dicliques are proper.

Let $\mathcal{C}$ be any diclique family with $V(\mathcal{C}) = \{v_1, \ldots, v_r\} \subseteq [n]$ and assume that $v_1 < \cdots < v_r$. For any set $S = \{x_1, \ldots, x_r\} \subseteq [n]$ with $x_1 < \cdots < x_r$, let us denote by $M(\mathcal{C}, S)$ the diclique family which is an image of $\mathcal{C}$ obtained by renaming v_i to x_i for each $i = 1, \ldots, r$. We call $M(\mathcal{C}, S)$ a *copy* of $\mathcal{C}$.

Each family of dicliques $\mathcal{C}$ defines a digraph $H = H(\mathcal{C})$ with vertex set $V(\mathcal{C})$: an arc is present in H whenever it is present in some $D(C)$, $C \in \mathcal{C}$. We say that the family $\mathcal{C}$ is a *diclique cover* of H.

The digraph $\overrightarrow{K}_h$ can be covered by dicliques in many different ways. Consider the following important symmetric diclique covers of $\overrightarrow{K}_h$:

- $\mathcal{C}_M = \{([h]; [h])\}$, the monochromatic diclique cover;
- $\mathcal{C}_R = E(\overrightarrow{K}_h)$, the rainbow diclique cover, where $E(\overrightarrow{K}_h)$ is the set of arcs of $\overrightarrow{K}_h$ and we identify each arc uv with the diclique $(\{u\}; \{v\})$;
- $\mathcal{C}_{in} = \{([h] \setminus \{i\}; \{i\}) \,:\, i \in [h]\}$, the cover by in-stars;
- $\mathcal{C}_{out} = \{(\{i\}; [h] \setminus \{i\}) \,:\, i \in [h]\}$, the cover by out-stars.

The motivation for the names “monochromatic” and “rainbow” is that a single attribute (or colour) $w \in W$ may generate a copy of $\mathcal{C}_M$ in D, while $h(h-1)$ attributes are needed for a copy of $\mathcal{C}_R$.

We will consider a sequence of random digraphs $\{D(k), k = 1, 2, \ldots\}$ where $D(k) = D(n, m, p_-, p_+)$, $n = n(k)$ (we always assume that $n(k)$ is increasing), $m = m(k)$, $p_- = p_-(k)$ and $p_+ = p_+(k)$ all depend on k. If not stated otherwise all limits below are as $k \to \infty$.

Let us now define what a birth threshold function for $\overrightarrow{K}_h$ is. We would like to have a function that, for a sequence of random digraphs $\{D(k)\}$, indicates whether a copy of $\overrightarrow{K}_h$ is present whp. Since the sequence $\{D(k)\}$ depends on several parameters which are themselves sequences, such a function has to take into account all of them.

Let X be either a fixed digraph or a diclique family. Let $X \in D$ denote the event that the random digraph D contains a copy of X. Given a sequence of random digraphs $\{D(k)\}$ we call a function $\tau : \mathbb{N}^2 \times [0;1]^2 \to \mathbb{R}_+$ a *birth threshold function* for X if both of the following implications hold:

$$\tau(n, m, p_-, p_+) \to 0 \implies P(X \in D(k)) \to 0;$$
$$\tau(n, m, p_-, p_+) \to \infty \implies P(X \in D(k)) \to 1.$$

Given a sequence of random digraphs $\{D(k)\}$ and a birth threshold function τ for $\overrightarrow{K}_h$ we call a collection $\mathcal{L}$ of diclique covers of $\overrightarrow{K}_h$ the *leading set* if τ is a birth threshold function for each $\mathcal{C} \in \mathcal{L}$ and the following implications hold:

1) $\tau(n, m, p_-, p_+) = O(1) \implies P(\mathcal{C}' \in D(k)) \to 0$ for each proper diclique cover $\mathcal{C}'$ of $\overrightarrow{K}_h$ such that $\mathcal{C}' \notin \mathcal{L}$;

2) $\tau(n, m, p_-, p_+) = \Theta(1) \implies P(\mathcal{C} \in D(k)) = \Omega(1)$ for each $\mathcal{C} \in \mathcal{L}$.

In the case where $\mathcal{L}$ consists just of a single diclique family $\mathcal{C}$, we call $\mathcal{C}$ the *leading cover*.

We will consider the following birth threshold functions:

$$\tau_1 = nm^{1/h}p_-p_+; \qquad \tau_2 = n^{1/(h-1)}mp_-p_+;$$
$$\tau_3 = nmp_-^{h-1}p_+; \qquad \tau_4 = nmp_-p_+^{h-1}.$$

We are now ready to state our main result (see Figure 2.2 for an illustration).

Theorem 2.2.1 *Let $h \geq 3$ be a fixed integer. Write $\alpha_0 = 1 - \frac{1}{(h-1)^2}$. Let $\{D(k)\}$ be a sequence of random binomial intersection digraphs such that n is increasing, $m = \Theta(n^\alpha)$ for some $\alpha > 0$, $p_- \to 0$ and $p_+ \to 0$.*

(i) If $\alpha < \alpha_0$ and

(a) $m^{\frac{h-1}{h(h-2)}}p_- \to \infty$ then τ_3 is a birth threshold function for $\overrightarrow{K}_h$ with the leading cover $\mathcal{C}_{in}$;

(b) $m^{\frac{h-1}{h(h-2)}}p_- \to 0$ and $m^{\frac{h-1}{h(h-2)}}p_+ \to 0$ then τ_1 is a birth threshold function for $\overrightarrow{K}_h$ with the leading cover $\mathcal{C}_M$;

(c) $m^{\frac{h-1}{h(h-2)}}p_+ \to \infty$ then τ_4 is a birth threshold function for $\overrightarrow{K}_h$ with the leading cover $\mathcal{C}_{out}$.

(ii) If $\alpha \geq \alpha_0$ and

(a) $mp_+ \to 0$ then τ_3 is a birth threshold function for $\overrightarrow{K}_h$ with the leading cover $\mathcal{C}_{in}$;

(b) $\alpha \neq \alpha_0$, $mp_- \to \infty$ and $mp_+ \to \infty$ then τ_2 is a birth threshold function for $\overrightarrow{K}_h$ with the leading cover $\mathcal{C}_R$;

(c) $mp_- \to 0$ then τ_4 is a birth threshold function for $\overrightarrow{K}_h$ with the leading cover $\mathcal{C}_{out}$.

Let us introduce the following two collections of diclique covers. Let $\mathcal{S}_{in}$ be the collection of all proper diclique covers of $\overrightarrow{K}_h$ such that each $\mathcal{C} \in \mathcal{S}_{in}$ is a set of arc-disjoint in-stars, that is, for each $C \in \mathcal{C}$ we have $|C^+| = 1$ and for each $C_i, C_j \in \mathcal{C}$ with $i \neq j$ and $C_i^+ = C_j^+$ we have $C_i^- \cap C_j^- = \emptyset$. Similarly, let $\mathcal{S}_{out}$ be the collection of all proper diclique covers $\mathcal{C}$ of $\overrightarrow{K}_h$ such that for each $C \in \mathcal{C}$ we have $|C^-| = 1$ and for each $C_i, C_j \in \mathcal{C}$ with $i \neq j$ and $C_i^- = C_j^-$ we have $C_i^+ \cap C_j^+ = \emptyset$. Observe that $\mathcal{C}_R, \mathcal{C}_{in} \in \mathcal{S}_{in}$ and $\mathcal{C}_R, \mathcal{C}_{out} \in \mathcal{S}_{out}$.

For the "boundary" cases of the parameters in Theorem 2.2.1 we have:

Theorem 2.2.2 *Let $h, \{D(k)\}, n, \alpha_0, m$ be as in Theorem 2.2.1. Suppose $p_- \to 0$ and $p_+ \to 0$. If*

(a) $\alpha < \alpha_0$ and $m^{\frac{h-1}{h(h-2)}} p_- = \Theta(1)$ then τ_1, τ_3 are birth threshold functions for $\overrightarrow{K}_h$ with the leading set $\{\mathcal{C}_M, \mathcal{C}_{in}\}$.

(b) $\alpha < \alpha_0$ and $m^{\frac{h-1}{h(h-2)}} p_+ = \Theta(1)$ then τ_1, τ_4 are birth threshold functions for $\overrightarrow{K}_h$ with the leading set $\{\mathcal{C}_M, \mathcal{C}_{out}\}$.

(c) $\alpha = \alpha_0$, $mp_- \to \infty$ and $mp_+ \to \infty$ then τ_1, τ_2 are birth threshold functions for $\overrightarrow{K}_h$ with the leading set $\{\mathcal{C}_M, \mathcal{C}_R\}$.

(d) $\alpha > \alpha_0$ and $mp_+ = \Theta(1)$ then τ_2, τ_3 are birth threshold functions for $\overrightarrow{K}_h$ with the leading set $\mathcal{S}_{in}$.

(e) $\alpha > \alpha_0$ and $mp_- = \Theta(1)$ then τ_2, τ_4 are birth threshold functions for $\overrightarrow{K}_h$ with the leading set $\mathcal{S}_{out}$.

(f) $\alpha = \alpha_0$ and $mp_+ = \Theta(1)$ then τ_1, τ_2, τ_3 are birth threshold functions for $\overrightarrow{K}_h$ with the leading set $\{\mathcal{C}_M\} \cup \mathcal{S}_{in}$.

(g) $\alpha = \alpha_0$ and $mp_- = \Theta(1)$ then τ_1, τ_2, τ_4 are birth threshold functions for $\overrightarrow{K}_h$ with the leading set $\{\mathcal{C}_M\} \cup \mathcal{S}_{out}$.

We note that the argument used in the proof of Theorem 2.2.1 can also be extended to the case where p_- or p_+ are bounded away from zero. In this case the birth threshold function remains the same, but the leading sets have to be slightly modified.

Finally, for the case $h = 2$ we have

Remark 2.2.3 *Let $\{D(k)\}, n, m$ be as in Theorem 2.2.1. Suppose $h = 2, p_- \to 0$ and $p_+ \to 0$. Then τ_2 is a birth threshold function for $\overrightarrow{K}_2$ with the leading cover $\mathcal{C}_R$.*

2.3 Diclique covers: a general lemma

In this section we present some important estimates and Lemma 2.3.2 that relates the birth threshold of a fixed digraph with presence of its diclique covers. This is very similar to results for undirected random intersection graphs, see Theorem 3 of [58]. We postpone the proofs of the estimates (2.1) - (2.8) of this section till Section 2.5.

Let $\{D(k)\}$ be a sequence of random binomial intersection digraphs and write $D = D(k)$. Recall that we denote the vertex set of D by V and the attribute set by W. Suppose $m = \Theta(n^\alpha)$ for some $\alpha > 0$. In (2.1) - (2.5) we will assume that

$$p_- \to 0, \quad p_+ \to 0 \quad \text{and} \quad mp_-p_+ \to 0.$$

Note that in this case mp_-p_+ is asymptotically equivalent to the probability that a fixed directed edge exists.

Given a diclique $C = (C^-; C^+)$ and a set S the restriction of C to S is the diclique $C[S] = (C^- \cap S; C^+ \cap S)$. The restriction for a diclique family is defined by $\mathcal{C}[S] = \{C[S] : C \in \mathcal{C}, V(C) \cap S \neq \emptyset\}$.

Let $S \subseteq V$ and let C be a diclique with $V(C) \subseteq S$. We will say that a *monochromatic* C *occurs on* S if in the realization of D there is at least one attribute $w \in W$ such that $C = C(w)[S]$ (we say that w *generates* C on S). We denote the probability of the event that a monochromatic C occurs on S by $P(C)$.

We say that a diclique family $\mathcal{C} = \{C_1, \ldots, C_s\}$ is *induced* in D if there are distinct attributes $w_1, \ldots, w_s \in W$ such that $C_i = C(w_i)[V(\mathcal{C})]$, for each $i = 1, \ldots, s$ and for any attribute $w \in W \setminus \{w_1, \ldots, w_s\}$ the diclique $C(w)[V(\mathcal{C})]$ is improper. Thus if $\mathcal{C}$ is induced in D then D contains $\mathcal{C}$ (see Section 2.2). We denote the probability of the event that $\mathcal{C}$ is induced in D by $P(\mathcal{C})$. Let $S_1, S_2, \ldots, S_N$ be all the subsets of V of size $r = |V(\mathcal{C})|$, where $N = \binom{n}{r}$. Let

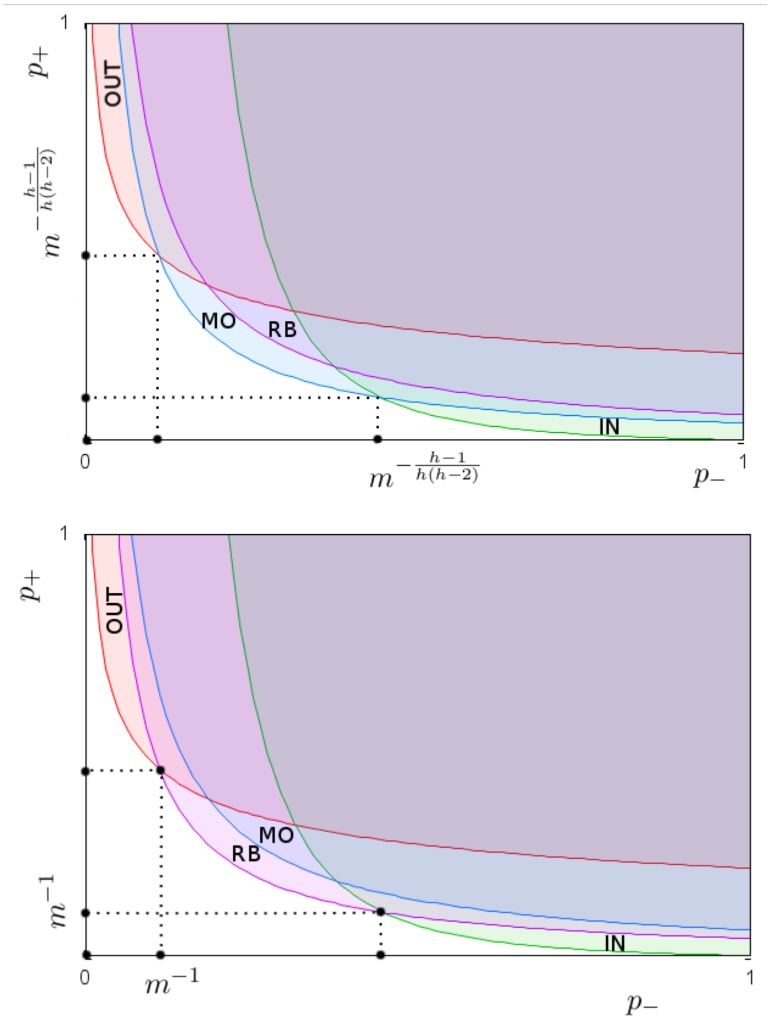

Figure 2.2: Schematic illustration of Theorems 2.2.1 and 2.2.2. Top: $\alpha < \alpha_0$, bottom: $\alpha > \alpha_0$. The coloured area is the region where D contains a copy of the special diclique cover whp (IN = "in-stars" cover $\mathcal{C}_{in}$, OUT="out-stars" cover $\mathcal{C}_{out}$, MO=monochromatic cover $\mathcal{C}_M$, RB = "rainbow" cover $\mathcal{C}_R$), the remaining area is where D does not contain that diclique cover whp. The white area is where D does not have a copy of $\overrightarrow{K}_h$ whp. On the contour between the white and the coloured area we know that D does not contain any other proper diclique cover whp (excluding the black points that correspond to the collections $\mathcal{S}_{in}$ and $\mathcal{S}_{out}$).

$\mathbb{I}_i = \mathbb{I}_{M(\mathcal{C},S_i)}$ be the indicator of the event that the copy $M(\mathcal{C}, S_i)$ is induced in D. Then the *number of induced copies* of $\mathcal{C}$ in D is defined by

$$X(\mathcal{C}) = \sum_{i=1}^{N} \mathbb{I}_i.$$

Let $S \subseteq V$ and suppose $C = (C^-; C^+)$ is a proper diclique such that $V(C) \subseteq S$. Then the probability that a monochromatic C occurs on S is

$$P(C) \sim \tilde{P}(C) := mp_-^{|C^-|} p_+^{|C^+|}. \tag{2.1}$$

Following [58], for a diclique family $\mathcal{C} = \{C_i\}$ we write $\sum \mathcal{C}^+ = \sum |C_i^+|$ and $\sum \mathcal{C}^- = \sum |C_i^-|$, and we denote by $|\mathcal{C}|$ the cardinality of $\mathcal{C}$.

For a diclique family $\mathcal{C}$ let $C_1, \ldots, C_t$ be all its *distinct* dicliques and let $a_1, \ldots, a_t$ be their multiplicities in $\mathcal{C}$. Let us denote $a_{\mathcal{C}} = (a_1!a_2! \ldots a_t!)^{-1}$.

Fix a diclique family $\mathcal{C}$ with $V(\mathcal{C}) \subseteq V$. If $\mathcal{C}$ is proper then the probability that $\mathcal{C}$ is induced in D is

$$P(\mathcal{C}) \sim \tilde{P}(\mathcal{C}) := a_{\mathcal{C}} \prod_{C \in \mathcal{C}} \tilde{P}(C) = a_{\mathcal{C}} m^{|\mathcal{C}|} p_-^{\sum \mathcal{C}^-} p_+^{\sum \mathcal{C}^+}. \tag{2.2}$$

Write

$$\mu(\mathcal{C}) = \mu(\mathcal{C}, n, m, p_-, p_+) := \frac{n^{|V(\mathcal{C})|}}{|V(\mathcal{C})|!} \tilde{P}(\mathcal{C}) = \frac{a_{\mathcal{C}}}{|V(\mathcal{C})|!} n^{|V(\mathcal{C})|} m^{|\mathcal{C}|} p_-^{\sum \mathcal{C}^-} p_+^{\sum \mathcal{C}^+}.$$

If $\mathcal{C}$ is proper then the number $X(\mathcal{C})$ of induced copies of $\mathcal{C}$ in D satisfies

$$\mathbb{E}\, X(\mathcal{C}) \sim \mu(\mathcal{C}). \tag{2.3}$$

More generally, suppose $\mathcal{C}$ is not necessarily proper. Suppose that, in addition, the following technical assumption is satisfied: for each $j = 0, \ldots, |V(\mathcal{C})|$

$$mp_-^j \to a_j \quad \text{and} \quad mp_+^j \to b_j \tag{2.4}$$

for some $a_j, b_j \in [0; \infty]$. Let $\hat{\mathcal{C}}$ be the diclique family obtained from $\mathcal{C}$ by taking only those dicliques $C \in \mathcal{C}$ that satisfy $mp_-^{|C^-|} p_+^{|C^+|} \to 0$ (for proper $\mathcal{C}$ we always have $\hat{\mathcal{C}} = \mathcal{C}$). Define

$$\tilde{\mu}(\mathcal{C}) = \tilde{\mu}(\mathcal{C}, n, m, p_-, p_+) := n^{|V(\mathcal{C})|} m^{|\hat{\mathcal{C}}|} p_-^{\sum \hat{\mathcal{C}}^-} p_+^{\sum \hat{\mathcal{C}}^+}.$$

Then

$$\mathbb{E}\, X(\mathcal{C}) = \Theta\left(\tilde{\mu}(\mathcal{C})\right). \tag{2.5}$$

Observe that the event "$\mathcal{C}$ is induced in D" allows any number of extra improper dicliques on $V(\mathcal{C})$.

For a proper diclique family $\mathcal{C}$ and any sequence of random graphs $\{D(k)\}$ (without any assumptions on p_-, p_+, m) we have

$$P(D(k) \text{ contains } \mathcal{C}) \leq \tilde{P}(\mathcal{C}); \tag{2.6}$$

$$P(\mathcal{C} \in D(k)) \leq \mu(\mathcal{C}); \tag{2.7}$$

(recall that the first event concerns a fixed diclique cover while the second one asks for any copy of $\mathcal{C}$ in $D(k)$) and, if (2.4) holds then for any diclique family $\mathcal{C}$

$$P(\mathcal{C} \in D(k)) \leq \tilde{\mu}(\mathcal{C}). \tag{2.8}$$

We will use the following simple technical lemma (known as the subsubsequence principle [56]) several times below.

Lemma 2.3.1 *For a positive integer t let $f_1, \ldots, f_t, g : \mathbb{N} \to \mathbb{R}$ be any functions. Suppose that for any increasing sequence of positive integers (n_k), $k = 1, 2 \ldots$ such that*

$$\lim_{k\to\infty} f_i(n_k) \text{ exists or is in } \{-\infty, \infty\} \text{ for each } i = 1, \ldots, t \tag{2.9}$$

we have $\lim_{k\to\infty} g(n_k) = b \in [-\infty; \infty]$. Then $\lim_{n\to\infty} g(n) = b$.

Proof Write $\bar{g} = \limsup_{n\to\infty} g(n)$. Then there is an increasing sequence of integers $(n_j), j = 1, 2, \ldots$ such that $g(n_j) \to \bar{g}$. By the Weierstrass theorem this sequence has a subsequence $(n_k), k = 1, 2, \ldots$ such that (2.9) holds. So $b = \lim_{k\to\infty} g(n_k) = \bar{g}$. Similarly $b = \liminf_{n\to\infty} g(n)$ and the claim follows. □

Let us call a diclique family $\mathcal{C}$ *simple* if it is proper and has no repetitive elements. The following result allows to find a birth threshold of a fixed digraph by considering just a constant number (which depends on h) of diclique covers.

Lemma 2.3.2 *Let $\{D(k)\}$ be a sequence of random binomial intersection digraphs such that n is increasing. Let $h \geq 2$ be an integer, and let H be a digraph with $V(H) = [h]$ and without isolated vertices. Suppose that $p_- \to 0$, $p_+ \to 0$ and $mp_-p_+ \to 0$. Then* *(a) if for each simple diclique cover* $\mathcal{C}$ *of* H *there is a non-empty set* $S \subseteq V(\mathcal{C})$ *such that* $\mathbb{E}\, X(\mathcal{C}[S]) \to 0$ *then whp* $D(k)$ *does not contain a copy of* H*;*

(b) if there is a simple diclique cover $\mathcal{C}$ *of* H *such that* $\mathbb{E}\, X(\mathcal{C}[S]) \to \infty$ *for each non-empty set* $S \subseteq V(\mathcal{C})$ *then whp* $D(k)$ *contains an induced copy of* $\mathcal{C}$ *(and therefore also a copy of* H *as an induced subgraph).*

Proof The part (a) is easy. Suppose the sequence $\{D(k)\}$ satisfies the conditions of (a), but $\limsup P(\overrightarrow{K}_h \in D) > 0$, where $D = D(k)$. Since the event $\overrightarrow{K}_h \in D$ implies $\mathcal{C} \in D$ for some simple diclique cover $\mathcal{C}$ of $\overrightarrow{K}_h$ and the number of such diclique covers is finite we have $\limsup P(\mathcal{C}_0 \in D) > 0$ for one of such covers $\mathcal{C}_0$. By Lemma 2.3.1 we may assume that (2.4) holds. Take a set $S \subseteq [h]$ such that $\mathbb{E}\,(X(\mathcal{C}_0[S])) \to 0$. By (2.5) and (2.8) we have

$$P(\mathcal{C}_0 \in D) \le P(\mathcal{C}_0[S] \in D) \le \tilde{\mu}(\mathcal{C}_0[S]) = \Theta(\mathbb{E}\, X(\mathcal{C}_0[S])) = o(1)$$

which is a contradiction.

Now let us prove (b). Let $\mathcal{C}$ be a simple diclique cover of H that satisfies the condition in (b). Note that $|V(\mathcal{C})| = h$ since H has no isolated vertices. Recall that the number of induced copies of $\mathcal{C}$ in D is $X = X(\mathcal{C}) = \sum_{i=1}^{N} \mathbb{I}_i$ where $\mathbb{I}_i = 1$ if and only if the diclique cover $M(\mathcal{C}, S_i)$ is induced in D and by (2.3)

$$\mathbb{E}\, X \sim \mu(\mathcal{C}) \sim \frac{n^h}{h!} \tilde{P}(\mathcal{C}) \to \infty.$$

So the claim will follow by the method of second moments if we show that $Var(X)/(\mathbb{E}\, X)^2 \to 0$. We have

$$\mathbb{E}\, X^2 = \mathbb{E}\, X + \sum_{i \ne j} \mathbb{I}_i \mathbb{I}_j.$$

If the sets S_i and S_j do not intersect, we have that $\mathbb{I}_i$ and $\mathbb{I}_j$ are independent and $\mathbb{E}\, \mathbb{I}_i \mathbb{I}_j = \mathbb{E}\, \mathbb{I}_i \mathbb{E}\, \mathbb{I}_j$. There are in total $\binom{n}{h}\binom{n-h}{h}$ (ordered) pairs of sets that do not intersect. So

$$\begin{aligned} Var(X) = \mathbb{E}\, X^2 - (\mathbb{E}\, X)^2 &\sim \mathbb{E}\, X + \frac{n^{2h}}{(h!)^2} (1 + o(1))\, P(\mathcal{C})^2 \\ &\quad - \frac{n^{2h}}{(h!)^2} (1 + o(1))\, P(\mathcal{C})^2 + \sum_{S_i \cap S_j \ne \emptyset} \mathbb{E}\, \mathbb{I}_i \mathbb{I}_j \\ &= o\left((\mathbb{E}\, X)^2\right) + \sum_{S_i \cap S_j \ne \emptyset} \mathbb{E}\, \mathbb{I}_i \mathbb{I}_j; \end{aligned}$$

and we have that

$$\frac{\sum_{S_i\cap S_j\neq\emptyset}\mathbb{E}\,\mathbb{I}_i\mathbb{I}_j}{(\mathbb{E}\,X)^2}\leq\frac{\sum_{s=1}^{h}\binom{n}{h}\binom{h}{s}\binom{n-h}{h-s}\max_{|S_i\cap S_j|=s}\mathbb{E}\,\mathbb{I}_i\mathbb{I}_j}{(\mathbb{E}\,X)^2}\leq C\sum_{s=1}^{h}T_s$$

for some constant C where

$$T_s=\frac{\max_{|S_i\cap S_j|=s}\mathbb{E}\,\mathbb{I}_i\mathbb{I}_j}{n^sP(\mathcal{C})^2}. \tag{2.10}$$

To show that $Var(X)/(\mathbb{E}\,X)^2\to 0$ it is enough to prove that $T_s\to 0$ for $s=1,\ldots,h$.

Fix a positive integer s, $s\leq h$ and two sets $S_i,S_j\subseteq V$ of size h such that $|S_i\cap S_j|=s$. Let $\mathcal{C}^i=M(\mathcal{C},S_i)=\{C_1^i,\ldots,C_t^i\}$ and $\mathcal{C}^j=M(\mathcal{C},S_j)=\{C_1^j,\ldots,C_t^j\}$ be two copies of the diclique family $\mathcal{C}$ (here $t=|\mathcal{C}|$). Let M be a matching (not necessarily perfect) in a bipartite graph with parts $X=\{x_1,\ldots,x_t\}$ and $Y=\{y_1,\ldots,y_t\}$. Let $L=X\setminus\{x_l:x_ly_r\in M\}$ and $R=Y\setminus\{y_r:x_ly_r\in M\}$. Define a diclique family[1]

$$\mathcal{C}^M=\{C_l^i\cup C_r^j:x_ly_r\in M\}\cup\{C_l^i:x_l\in L\}\cup\{C_r^j:y_r\in R\}.$$

Here the union of diclique families is 'multiset union' so that $\mathcal{C}^M$ contains exactly $|M|+|R|+|L|$ elements.

Let us call M *good* if $\mathcal{C}^M[S^i]\setminus\mathcal{C}^i$ and $\mathcal{C}^M[S^j]\setminus\mathcal{C}^j$ consist of improper dicliques only.

Proposition 2.3.3 *If both $\mathcal{C}^i$ and $\mathcal{C}^j$ are induced in D then there is a good matching M such that D contains $\mathcal{C}^M$.*

Proof By definition, if both $\mathcal{C}^i$ and $\mathcal{C}^j$ are induced in D then there are two lists $w^i=(w_1^i,\ldots,w_t^i)$ and $w^j=(w_1^j,\ldots,w_t^j)$ of attributes from W such that

$$C(w_s^i)[S^i]=C_s^i\quad\text{and}\quad C(w_s^j)[S^j]=C_s^j\quad\text{for } s=1,\ldots,t.$$

Also, there are no edges in $D[S^i]$ (respectively, $D[S^j]$) generated by attributes not in w^i (respectively, w^j). Notice also, that by definition the elements in each of the lists w^i,w^j must be distinct (although there can be some elements that belong to both lists). Thus all pairs x_ly_r where $w_l^i=w_r^j$ and $l,r\in\{1,\ldots,t\}$ define some matching M_0. Clearly for each diclique $C\in\mathcal{C}_{M_0}$ it is possible to assign a unique

[1]The notation should not be confused with the monochromatic cover $\mathcal{C}_M$.

element $w \in W$ such that $C \subseteq C(w)[S^i \cup S^j]$. The fact that M_0 is good follows by definition since both $\mathcal{C}^i$ and $\mathcal{C}^j$ are induced in D. □

Let $P^{**}(\mathcal{C})$ denote the probability of the event that D contains $\mathcal{C}$. It follows by Proposition 2.3.3 that

$$\mathbb{E}\,\mathbb{I}_i\mathbb{I}_j \leq \sum_M P^{**}(\mathcal{C}^M). \tag{2.11}$$

where the sum is over all good matchings M. The number of all good matchings M is at most $(t+1)^t$, which is constant. By (2.10) and (2.11) it suffices to show that

$$\frac{P^{**}(\mathcal{C}^M)}{n^s P(\mathcal{C})^2} \to 0$$

for each good matching M. Write $S = S^i \cap S^j$. We have $P^{**}(\mathcal{C}^M) \leq \tilde{P}(\mathcal{C}^M)$ by (2.6) and $P(\mathcal{C}_i) = P(\mathcal{C}_j) \sim \tilde{P}(\mathcal{C})$ by (2.2). For $x_l y_r \in M$ we have

$$\frac{\tilde{P}(C_l^i \cup C_r^j)}{\tilde{P}(C_l^i)\tilde{P}(C_r^j)} = \frac{1}{mp_-^{|C_l^{i-} \cap C_r^{j-}|} p_+^{|C_l^{i+} \cap C_r^{j+}|}} \leq \frac{1}{mp_-^{|S \cap C_r^{j-}|} p_+^{|S \cap C_r^{j+}|}}.$$

Therefore we get that there is a constant c such that:

$$\begin{aligned}
\frac{P^{**}(\mathcal{C}^M)}{n^s P(\mathcal{C})^2} &= \frac{P^{**}(\mathcal{C}^M)}{n^s P(\mathcal{C}^i) P(\mathcal{C}^j)} \\
&\leq \frac{c}{n^s} \prod_{x_l y_r \in M} \frac{\tilde{P}(C_l^i \cup C_r^j)}{\tilde{P}(C_l^i)\tilde{P}(C_r^j)} \prod_{x_l \in L} \frac{\tilde{P}(C_l^i)}{\tilde{P}(C_l^i)} \prod_{y_r \in R} \frac{\tilde{P}(C_r^j)}{\tilde{P}(C_r^j)} \\
&= \frac{c}{n^s} \prod_{x_l y_r \in M} \frac{1}{mp_-^{|C_l^{i-} \cap C_r^{j-}|} p_+^{|C_l^{i+} \cap C_r^{j+}|}} \\
&\leq \frac{c}{n^s} \prod_{y_r \notin R} \frac{1}{mp_-^{|S \cap C_r^{j-}|} p_+^{|S \cap C_r^{j+}|}} \\
&= O\left(\frac{1}{\mathbb{E}\, X(\mathcal{C}^j[S])}\right) = o(1).
\end{aligned}$$

Here we get the bound in the last line as follows. If all dicliques in $\mathcal{C}^j[S]$ are proper, we use (2.3). If $\mathcal{C}^j[S]$ has some improper dicliques, by Lemma 2.3.1 it is sufficient to consider the case where the assumption (2.4) holds. In this case we use (2.5) with the family $\hat{\mathcal{C}}^j[S]$. This completes the proof. □

The following fact follows easily from the estimates above.

Lemma 2.3.4 *Fix an integer $h \geq 2$. Let H be a digraph with $V(H) = [h]$ and at least one arc. Let $\{D(k)\}$ be a sequence of random binomial intersection digraphs such that $mp_-p_+ \to 0$. Let $\mathcal{S}$ be a collection of simple diclique covers of H.*

Suppose that $\mu(\mathcal{C}) = O(1)$ for each $\mathcal{C} \in \mathcal{S}$ and for each simple cover $\mathcal{C}$ of H, such that $\mathcal{C} \notin \mathcal{S}$ we have $\mu(\mathcal{C}) \to 0$. Let $\mathcal{C}$ be any proper diclique cover of H such that $\mathcal{C} \notin \mathcal{S}$. Then $P(\mathcal{C} \in D(k)) \to 0$.

Proof Suppose, the claim is false, i.e. there is a proper diclique cover $\mathcal{C}_0$ of H such that $\limsup P(\mathcal{C}_0 \in D(k)) > 0$ and $\mathcal{C}_0 \notin \mathcal{S}$. By assumption and (2.7) $\mathcal{C}_0$ cannot be simple. We may assume that $\mathcal{C}_0$ consists of some simple cover $\mathcal{C}_1$ of H and a proper diclique C. By (2.7) and the definition of μ: $P(\mathcal{C}_0 \in D(k)) \leq \mu(\mathcal{C}_0) \leq \mu(\mathcal{C}_1)\tilde{P}(C) \to 0$ which is a contradiction. $\square$

2.4 Main proofs

The four special (see Section 2.2) diclique covers of the digraph $\overrightarrow{K}_h$ have the following birth thresholds:

Lemma 2.4.1 *Fix an integer $h \geq 2$. Let $\{D(k)\}$ be a sequence of random binomial intersection digraphs such that n is increasing, $p_- \to 0$, $p_+ \to 0$ and $mp_-p_+ \to 0$. Then the diclique covers $\mathcal{C}_M$, $\mathcal{C}_R$, $\mathcal{C}_{in}$, $\mathcal{C}_{out}$ of $\overrightarrow{K}_h$ have birth threshold functions $\tau_1, \tau_2, \tau_3, \tau_4$, respectively.*

Proof Let $\mathcal{C}$ be one of the four special covers. To prove that τ is a birth threshold function for $\mathcal{C}$ we may use Lemma 2.3.2. By that lemma it is enough to show that whenever $p_- \to 0$, $p_+ \to 0$ and $mp_-p_+ \to 0$ we have

$$\tau \to 0 \implies \mathbb{E}\, X(\mathcal{C}) \to 0; \tag{2.12}$$

$$\tau \to \infty \implies \text{for each non-empty set } S \subseteq [h] : \mathbb{E}\, X(\mathcal{C}[S]) \to \infty. \tag{2.13}$$

Here $\tau = \tau(k) = \tau(n(k), m(k), p_-(k), p_+(k))$.

By Lemma 2.3.1 we may assume that (2.4) holds.

For the monochromatic cover $\mathcal{C}_M$ we see that for any non-empty set $S \subseteq V(\mathcal{C}_M)$ of size $s \leq h$, the restriction $\mathcal{C}_M[S]$ has the same form as the cover $\mathcal{C}_M$ defined for $h = s$ so:

$$\mu(\mathcal{C}_M[S]) = n^s m(p_-p_+)^s = m^{1-\frac{s}{h}}\tau_1^s.$$

So (2.12) and (2.13) follow by (2.3).

Consider now the diclique cover $\mathcal{C}_R$. Let $S \subseteq V(\mathcal{C}_R)$ be non-empty and write $s = |S|$. When $s < h$ the restriction $\mathcal{C} = \mathcal{C}_R[S]$ is not a proper diclique family as for each $v \in S$ it contains $(h-s)$ dicliques $(\{v\}; \emptyset)$ and $(h-s)$ dicliques $(\emptyset; \{v\})$. Therefore we need to use (2.5) with the family $\hat{\mathcal{C}}$.

If $mp_- \to 0$ and $mp_+ \to 0$ or $S = [h]$ then $\hat{\mathcal{C}} = \mathcal{C}$ and

$$\tilde{\mu}(\mathcal{C}) = n^s m^{s(s-1)+2s(h-s)}(p_- p_+)^{s(s-1)+s(h-s)} = m^{s(h-s)} \tau_2^{s(h-1)}$$

and (2.12) follows by (2.3). If $mp_- \to a_1 > 0$ and $mp_+ \to 0$ then $\hat{\mathcal{C}} = \mathcal{C} \setminus \{(\{v\}; \emptyset) : v \in S\}$.

$$\begin{aligned}\tilde{\mu}(\mathcal{C}) = n^s m^{s(s-1)+2s(h-s)}(p_- p_+)^{s(s-1)+s(h-s)}(mp_-)^{-s(h-s)} \\ = (p_-)^{-s(h-s)} \tau_2^{s(h-1)}.\end{aligned}$$

The expression $\tilde{\mu}(\mathcal{C})$ for the case $mp_+ \to b_1 > 0$ and $mp_- \to 0$ is similar (replace p_- with p_+).

If $mp_- \to a_1 > 0$ and $mp_+ \to b_1 > 0$ by (2.5) the family $\hat{\mathcal{C}}$ is exactly the cover $\mathcal{C}_R$ defined for $h = s$ and

$$\tilde{\mu}(\mathcal{C}) = n^{\frac{s(h-s)}{h-1}} \tau_2^{s(s-1)}.$$

In each case (2.13) holds by (2.5).

For the "in-stars" cover $\mathcal{C}_{in}$ we have

$$\mu(\mathcal{C}_{in}) = n^h m^h p_-^{h(h-1)} p_+^h = \tau_3^h$$

so the implication (2.12) holds by (2.3). Now let $S \subset V(\mathcal{C}_{in})$ be non-empty and write $s = |S|$. Suppose $s \geq 2$. Then the diclique family $\mathcal{C} = \mathcal{C}_{in}[S]$ is not proper as it contains $h - s$ copies of the diclique $(S; \emptyset)$. If $mp_-^s \to 0$, we use (2.5) with $\hat{\mathcal{C}} = \mathcal{C}$:

$$\tilde{\mu}(\mathcal{C}) = n^s m^{s+(h-s)} p_-^{s(s-1)+s(h-s)} p_+^s = m^{h-s} \tau_3^s.$$

If $mp_-^s \to a_s > 0$ then we have $\hat{\mathcal{C}} = \{(S \setminus \{v\}; \{v\}) : v \in S\}$ so

$$\tilde{\mu}(\mathcal{C}) = n^s m^{s+(h-s)} p_-^{s(s-1)+s(h-s)} p_+^s (mp_-^s)^{-(h-s)} = p_-^{-s(h-s)} \tau_3^s.$$

Now, if $s = 1$, in each of the cases $a_1 > 0$, $b_1 > 0$ and $a_1 = b_1 = 0$, see (2.4), by (2.5) we have that $\mathbb{E}\, X(\mathcal{C}) = \Omega(\tau_3)$. Therefore by (2.5) we see that

$$\mathbb{E}\, X(\mathcal{C}) \to \infty \tag{2.14}$$

when $\tau_3 \to \infty$ and $S \subseteq V(\mathcal{C}_{in})$ is non-empty and so (2.13) holds for $\mathcal{C}_{in}$. Finally, the case of $\mathcal{C}_{out}$ is symmetric to that of $\mathcal{C}_{in}$. □

Remark 2.4.2 *Let $\{D(k)\}$ be as in Lemma 2.4.1. Let $\mathcal{C}$ be one of the four special diclique covers of $\overrightarrow{K}_h$ and let τ be its birth threshold function given in Lemma 2.4.1. If $\tau = \Theta(1)$ then $P(a$ copy of $\mathcal{C}$ is induced in $D(k)) = \Omega(1)$.*

Proof From the proof of Lemma 2.4.1 we have that $\mu(\mathcal{C}) = \Theta(\tau^h)$ for $\mathcal{C} \in \{\mathcal{C}_{in}, \mathcal{C}_M, \mathcal{C}_{out}\}$ and $\mu(\mathcal{C}_R) = \Theta(\tau_2^{h(h-1)})$. So in each case, we have $\mu(\mathcal{C}) = \Theta(1)$.

By (2.3) the number $X = X(\mathcal{C})$ of induced copies of $\mathcal{C}$ satisfies

$$\mathbb{E}\, X = \Theta(\mu(\mathcal{C})) = \Theta(1).$$

The proof of Lemma 2.4.1 also shows that in each case $\mu(\mathcal{C}[S]) = \Omega(1)$ for each $S \subseteq V(C)$. Following the lines of the second part of the proof of Lemma 2.3.2 we see that

$$\frac{Var(X)}{(\mathbb{E}\, X)^2} = O(1).$$

Using the Cauchy-Schwartz inequality we get

$$P(X > 0) \geq \frac{(\mathbb{E}\, X)^2}{\mathbb{E}\, X^2} = \Omega(1).$$

□

The next lemma says that if p_- is sufficiently large then it is always 'better' to replace any diclique cover by a 'star' cover $\mathcal{C}'$:

Lemma 2.4.3 *Let $h \geq 3$ be an integer and let $\{D(k)\}$ be a sequence of random binomial intersection digraphs such that n is increasing, $p_- \to 0$ and $m^{\frac{h-1}{h(h-2)}} p_- = \Omega(1)$. Let $\mathcal{C}$ be a diclique family consisting of a single proper diclique $C = (C^-; C^+)$ with $V(C) \subseteq [h]$. Suppose $|C^+| \geq 2$ or $|C^- \cap C^+| = 1$. Let $\mathcal{C}'$ be a diclique family defined by*

$$\mathcal{C}' = \left\{ (C^- \setminus \{v\}; \{v\}) \,:\, v \in C^+ \right\}.$$

Then $\tilde{P}(\mathcal{C}') = \Omega(\tilde{P}(\mathcal{C}))$. More precisely,

(a) If $\mathcal{C} \neq \mathcal{C}_M$ or $m^{\frac{h-1}{h(h-2)}} p_- \to \infty$ then $\tilde{P}(\mathcal{C}) = o(\tilde{P}(\mathcal{C}'))$;

(b) If $\mathcal{C} = \mathcal{C}_M$ and $m^{\frac{h-1}{h(h-2)}} p_- = \Theta(1)$ then $\tilde{P}(\mathcal{C}') = \Theta(\tilde{P}(\mathcal{C}))$.

We note that $\mathcal{C}$ and $\mathcal{C}'$ are both diclique covers of the diclique C.

Proof If $|C^+| = 1$ and $|C^- \cap C^+| = 1$ then

$$\frac{\tilde{P}(\mathcal{C}')}{\tilde{P}(\mathcal{C})} = p_-^{-1} \to \infty$$

so we may assume that $|C^+| > 1$. Let $c_- = |C^-|$ and $c_+ = |C^+|$. Then

$$\frac{\tilde{P}(\mathcal{C}')}{\tilde{P}(\mathcal{C})} = \frac{m^{c_+} p_-^{\sum \mathcal{C}'^-} p_+^{c_+}}{m p_-^{c_-} p_+^{c_+}} = m^{c_+-1} p_-^{\sum \mathcal{C}'^- - c_-}$$

$$\geq m^{c_+-1} p_-^{(c_+-1)(h-1)-1} = \left(m p_-^{h-1-\frac{1}{c_+-1}}\right)^{c_+-1} \tag{2.15}$$

$$\geq \left(m p_-^{h-1-\frac{1}{h-1}}\right)^{c_+-1}. \tag{2.16}$$

Here the inequality in the second line follows from:

$$\sum \mathcal{C}'^- - c_- \leq (c_+ - 1)(h-1) - 1. \tag{2.17}$$

To see (2.17) consider three possible cases:

If $c_- \in \{1, 2, \ldots, h-2\}$ then

$$\sum \mathcal{C}'^- - c_- \leq c_- c_+ - c_- = c_-(c_+ - 1) \leq (h-1)(c_+ - 1) - (c_+ - 1).$$

If $c_- = h - 1$ then there is at most one diclique C' in $\mathcal{C}'$ with $C'^+ = \{v\}$ such that $v \in |C^+ \setminus C^-|$ so

$$\sum \mathcal{C}'^- - c_- \leq (c_- - 1)(c_+ - 1) + c_- - c_- = (h-1)(c_+ - 1) - (c_+ - 1),$$

and if $c_- = h$,

$$\sum \mathcal{C}'^- - c_- \leq c_+(c_- - 1) - c_- = (h-1)(c_+ - 1) - 1.$$

Note that in the inequality (2.16) the right hand side can be of the same order only for $c_+ = h$. But if $c_+ = h$ and $c_- < h$, we get

$$\sum \mathcal{C}'^- - c_- = (h-2)c_- < h(h-2) = c_+(h-1) - 1,$$

so in that case the right hand side of the inequality (2.15) of a smaller order than the left hand side. Now note that

$$m p_-^{h-1-\frac{1}{h-1}} = \Omega(1)$$

since $p_- = \Omega\left(m^{-\frac{h-1}{h(h-2)}}\right)$ and $m^{1-\frac{h-1}{h(h-2)}\left(h-1-\frac{1}{h-1}\right)} = 1$. Thus we have shown that $\tilde{P}(\mathcal{C}) = O(\tilde{P}(\mathcal{C}'))$ and the claim (a) holds. To complete the proof, note that for $\mathcal{C} = \mathcal{C}_M$ (2.17) and (2.16) become equalities. □

The next lemma shows that if we have a diclique family consisting of many "in-stars" centred at one vertex, we may merge all of them into a single diclique.

Lemma 2.4.4 *Let $h \geq 2$ be an integer and let H be a digraph obtained from the complete bipartite graph $K_{1,h-1}$ by orienting each edge towards the centre vertex v. Let $\mathcal{C}$ be any proper diclique cover of the digraph H of the form*

$$\mathcal{C} = \left\{(C_i^-; \{v\}),\, i = 1, \ldots, t\right\}$$

where $t \geq 2$ and $\bigcup_i C_i^- = V(H) \setminus \{v\}$.

Let $\mathcal{C}^* = \{(V(H) \setminus \{v\}; \{v\})\}$ be a cover of H and let $\{D(k)\}$ be a sequence of random binomial intersection digraphs such that n is increasing, $p_- \to 0$ and $mp_+ = O(1)$. Then

$$\tilde{P}(\mathcal{C}^*) = \Omega(\tilde{P}(\mathcal{C})).$$

More precisely,

(a) If $mp_+ = o(1)$ or $\sum \mathcal{C}^- > h-1$ then $\tilde{P}(\mathcal{C}) = o(\tilde{P}(\mathcal{C}^*))$;

(b) If $mp_+ = \Theta(1)$ and $\sum \mathcal{C}^- = h-1$ then $\tilde{P}(\mathcal{C}^*) = \Theta(\tilde{P}(\mathcal{C}))$.

Proof Using (2.2) we have

$$\frac{\tilde{P}(\mathcal{C})}{\tilde{P}(\mathcal{C}^*)} = \frac{a_{\mathcal{C}} m^t p_-^{\sum \mathcal{C}^-} p_+^t}{m p_-^{h-1} p_+} = a_{\mathcal{C}} (mp_+)^{t-1} p_-^{\sum \mathcal{C}^- - (h-1)} = O(1)$$

since $\sum \mathcal{C}^- \geq |\bigcup_i C_i| = h-1$. The claims (a) and (b) follow similarly. □

Let us remark that we use $\tilde{P}$ rather than P in Lemma 2.4.3 and Lemma 2.4.4 for convenience. By (2.2) we know that P can be replaced by $\tilde{P}$ as long as p_-, p_+ and mp_-p_+ all tend to 0. We are now ready to prove Theorem 2.2.1.

We split the proof of our main result into a few lemmas. In the four lemmas below we assume that $h, \{D(k)\}, n, \alpha_0, m$ are as in Theorem 2.2.1 $p_- \to 0, p_+ \to 0$ and $\mu(\mathcal{C}) = \mu(\mathcal{C}, n, m, p_-, p_+)$ is as defined in Section 2.3.

Lemma 2.4.5 *Suppose $\alpha < \alpha_0$ and $m^{\frac{h-1}{h(h-2)}} p_- = \Omega(1)$. Then τ_3 is a birth threshold function for $\overrightarrow{K}_h$. Furthermore, if $m^{\frac{h-1}{h(h-2)}} p_- \to \infty$ then the leading set is $\mathcal{L} = \{\mathcal{C}_{in}\}$ and if $m^{\frac{h-1}{h(h-2)}} p_- = \Theta(1)$ then the leading set is $\mathcal{L} = \{\mathcal{C}_{in}, \mathcal{C}_M\}$.*

In each of the cases above, if $\tau_3 = \Theta(1)$ then $\mu(\mathcal{C}) \to 0$ for any simple diclique cover $\mathcal{C} \notin \mathcal{L}$ and $\mu(\mathcal{C}) = \Theta(1)$ for $\mathcal{C} \in \mathcal{L}$.

Lemma 2.4.6 *Suppose $\alpha < \alpha_0$, $m^{\frac{h-1}{h(h-2)}} p_- \to 0$ and $m^{\frac{h-1}{h(h-2)}} p_+ \to 0$. Then τ_1 is a birth threshold function for $\overrightarrow{K}_h$ with the leading set $\mathcal{L} = \{\mathcal{C}_M\}$.*

If $\tau_1 = \Theta(1)$ then $\mu(\mathcal{C}) \to 0$ for any simple diclique cover $\mathcal{C} \notin \mathcal{L}$ and $\mu(\mathcal{C}) = \Theta(1)$ for $\mathcal{C} \in \mathcal{L}$.

Lemma 2.4.7 *Suppose $\alpha \geq \alpha_0$ and $mp_+ = O(1)$. Then τ_3 is a birth threshold function for $\overrightarrow{K}_h$. Furthermore, if $\alpha \geq \alpha_0$ and $mp_+ \to 0$ then the leading set is $\mathcal{L} = \{\mathcal{C}_{in}\}$; if $\alpha > \alpha_0$ and $mp_+ = \Theta(1)$ then the leading set is $\mathcal{L} = \mathcal{S}_{in}$; if $\alpha = \alpha_0$ and $mp_+ = \Theta(1)$ then the leading set is $\mathcal{L} = \{\mathcal{C}_M\} \cup \mathcal{S}_{in}$.*

In each of the cases above, if $\tau_3 = \Theta(1)$ then $\mu(\mathcal{C}) \to 0$ for any simple diclique cover $\mathcal{C} \notin \mathcal{L}$ and $\mu(\mathcal{C}) = \Theta(1)$ for $\mathcal{C} \in \mathcal{L}$.

Lemma 2.4.8 *Suppose $\alpha \geq \alpha_0$, $mp_- \to \infty$ and $mp_+ \to \infty$. Then τ_2 is a birth threshold function for $\overrightarrow{K}_h$. Furthermore, if $\alpha > \alpha_0$ then the leading set is $\mathcal{L} = \{\mathcal{C}_R\}$ and if $\alpha = \alpha_0$ then the leading set is $\mathcal{L} = \{\mathcal{C}_M, \mathcal{C}_R\}$.*

In each of the cases above, if $\tau_2 = \Theta(1)$ then $\mu(\mathcal{C}) \to 0$ for any simple diclique cover $\mathcal{C} \notin \mathcal{L}$ and $\mu(\mathcal{C}) = \Theta(1)$ for $\mathcal{C} \in \mathcal{L}$.

Proof of Theorem 2.2.1 Apply Lemmas 2.4.5-2.4.8 and notice that the cases (i)(c) and (ii)(c) follow by symmetry. ☐

Proof of Theorem 2.2.2 We note that if τ, τ' are birth threshold functions for the sequence $\{D(k)\}$ given by Theorem 2.2.2 then they are equivalent in the sense that $\log \tau = \Theta(\log \tau')$. The cases (a), (c), (d), (f) follow by Lemmas 2.4.5-2.4.8 and the remaining cases follow by symmetry. ☐

The idea of the proof of Lemmas 2.4.5-2.4.8 is to consider the birth threshold functions of the four special diclique covers and the boundaries determined by them, see Figure 2.2. We will use Lemma 2.4.3 and Lemma 2.4.4 to compare the probability of complicated diclique covers of $\overrightarrow{K}_h$ with the probability of appropriate special covers. In the proofs we write $D = D(k)$.

Proof of Lemma 2.4.5 By Lemma 2.3.1 we may assume that $mp_-p_+ \to a \in [0; \infty]$.

Suppose that $\tau_3 = \tau_3(k) = nmp_-^{h-1}p_+ \to \infty$ as $k \to \infty$. If $a = 0$ then by Lemma 2.4.1 the random digraph D contains a copy of $\mathcal{C}_{in}$ whp. If $a > 0$ consider a sequence of random digraphs $\{D'(k)\}$ where $D'(k) = D(n, m, p_-, p'_+)$, $p'_+ = (\omega m p_-)^{-1}$ and $\omega = \omega(n)$ grows slowly, say $\omega(n) = \ln n$. We have

$$\tau_3(n, m, p_-, p'_+) = nmp_-^{h-1}p'_+ = np_-^{h-2}\omega^{-1} = \Omega\left(n^{1-\frac{\alpha(h-1)}{h}}\omega^{-1}\right).$$

So $\tau_3(n, m, p_-, p'_+) \to \infty$, $mp_-p'_+ \to 0$ and $p'_+ = o(p_+)$. We have that $D'(k)$ contains a copy of $\mathcal{C}_{in}$ whp by Lemma 2.4.1 and therefore $D(k)$ contains a copy of $\mathcal{C}_{in}$ whp by monotonicity.

Now suppose that $\tau_3 = \Theta(1)$, so that $\mu(\mathcal{C}_{in}) = \tau_3^h = \Theta(1)$. Notice that since $\alpha < 1$ we have

$$mp_-p_+ = \frac{\tau_3}{np_-^{h-2}} = O\left(\frac{1}{nm^{-\frac{h-1}{h}}}\right) = O\left(n^{-1+\frac{\alpha(h-1)}{h}}\right) = o(1).$$

Let $\mathcal{C}$ be any simple diclique cover of $\overrightarrow{K}_h$. We will show that

$$\mu(\mathcal{C}) = O(1) \tag{2.18}$$

and furthermore we can replace $O()$ with $o()$ (respectively, $\Theta()$) if $\mathcal{C} \notin \mathcal{L}$ (respectively, $\mathcal{C} \in \mathcal{L}$).

Assuming we have proved (2.18), we obtain that $\tau_3 \to 0$ implies $\mu(\mathcal{C}) \to 0$, since both τ_3 and $\mu(\mathcal{C})$ are increasing multinomials in m, p_-, p_+. Now the fact that τ_3 is a birth threshold function for $\overrightarrow{K}_h$ follows by (2.6) and Markov's inequality since the number of simple diclique covers of $\overrightarrow{K}_h$ is finite. Remark 2.4.2 implies that in the case $\mu(\mathcal{C}_{in}) = \tau_3^h = \Theta(1)$ we have $P(\mathcal{C}_{in} \in D) = \Omega(1)$. If in addition $mp_-^{\frac{h-1}{h(h-2)}} = \Theta(1)$ then we have $\mu(\mathcal{C}_M) = \tau_1(n, m, p_-, p_+)^h = \Theta(1)$ and so $P(\mathcal{C}_M \in D) > 0$. So by Lemma 2.3.4 the set $\mathcal{L}$ is leading (for the birth threshold function τ_3 and $\overrightarrow{K}_h$).

So let us prove (2.18). Suppose $\mathcal{C} = \{C_1, C_2, \ldots, C_t\}$. Using the definition of μ and $\tilde{P}$, see Section 2.3, we have

$$\begin{aligned} \mu(\mathcal{C}) &= (h!)^{-1} n^h \tilde{P}(C_1)\tilde{P}(C_2)\ldots\tilde{P}(C_t) \\ &= O\left(n^h \tilde{P}(\mathcal{C}'_1)\tilde{P}(\mathcal{C}'_2)\ldots\tilde{P}(\mathcal{C}'_t)\right) && (2.19) \\ &= O(n^h \tilde{P}(\mathcal{C}_{in})) && (2.20) \\ &= O(\mu(\mathcal{C}_{in})) = O(1). && (2.21) \end{aligned}$$

Here in (2.19) we apply Lemma 2.4.3 so that for each $i = 1, \ldots, t$ the diclique family $\mathcal{C}'_i$ is obtained from the family $\mathcal{C}_i = \{C_i\}$ by splitting it into "in-stars". The resulting family $\mathcal{C}' = \mathcal{C}'_1 \cup \mathcal{C}'_2 \cup \cdots \cup \mathcal{C}'_t$ is a cover of $\overrightarrow{K}_h$ which consists of possibly duplicated or overlapping "in-stars". Next, in (2.20) we regroup the terms and apply Lemma 2.4.4.

More precisely, note that by Lemma 2.4.3 we can replace $O()$ with $o()$ in (2.19) in all cases except if

- $p_- m^{\frac{h-1}{h(h-2)}} = \Theta(1)$ and $\mathcal{C} = \mathcal{C}_M$; or
- for each $C \in \mathcal{C}$ we have $|C^+| = 1$ and $C^+ \cap C^- = \emptyset$.

We can complete the proof of (2.18) by noticing that since $\alpha < \alpha_0$ we have $p_+ = \Theta\left(\tau_3/(mnp_-^{h-1})\right) = o(m^{-1})$, so by Lemma 2.4.4 we can replace $O()$ with $o()$ in (2.20) for the second exceptional case whenever $\mathcal{C} \neq \mathcal{C}_{in}$. □

Proof of Lemma 2.4.6 First suppose that $\tau_1 = \tau_1(k) = nm^{1/h}p_-p_+ \to \infty$. By Lemma 2.4.1, some copy of $\mathcal{C}_M$ is induced in D whenever $mp_-p_+ \to 0$. Otherwise, using Lemma 2.3.1 we may assume that $mp_-p_+ \to a \in (0;\infty]$. Let $p'_- = (mp_+ \ln n)^{-1}$ and consider a sequence of random digraphs $\{D'(k)\}$ where $D'(k) = D(n, m, p'_-, p_+)$. Then $mp'_-p_+ \to 0$, $p'_- = o(p_-)$ and $\tau_1(n, m, p'_-, p_+) \to \infty$. So we have that $D'(k)$ contains an induced copy of $\mathcal{C}_M$ whp by Lemma 2.4.1 and by monotonicity $\mathcal{C}_M \in D$ whp.

Now suppose that $\tau_1 = \Theta(1)$. In this case we have $\mu(\mathcal{C}_M) = \tau_1^h = \Theta(1)$ and

$$mp_-p_+ = O\left(m\frac{1}{nm^{1/h}}\right) = O\left(\frac{m^{\frac{h-1}{h}}}{n}\right) = O\left(n^{-1+\alpha\frac{h-1}{h}}\right) \to 0.$$

Let us now show that for any simple diclique cover $\mathcal{C}$ of $\overrightarrow{K}_h$, such that $\mathcal{C} \neq \mathcal{C}_M$ we have $\mu(\mathcal{C}) \to 0$. Lemma 2.4.6 will then follow by monotonicity, Lemma 2.3.4, Lemma 2.4.1 and Remark 2.4.2.

Write $t = p_-p_+ = \Theta\left(n^{-1}m^{-1/h}\right)$. By the definition of μ, see (2.3), we have that for any simple diclique cover $\mathcal{C}$ of $\overrightarrow{K}_h$

$$\begin{aligned} h!\mu(\mathcal{C}) &= n^h m^{|\mathcal{C}|} p_+^{\sum \mathcal{C}^+} p_-^{\sum \mathcal{C}^-} \\ &= n^h m^{|\mathcal{C}|} p_+^{\sum \mathcal{C}^+ - \sum \mathcal{C}^-} t^{\sum \mathcal{C}^-} && (2.22) \\ &= n^h m^{|\mathcal{C}|} p_-^{\sum \mathcal{C}^- - \sum \mathcal{C}^+} t^{\sum \mathcal{C}^+}. && (2.23) \end{aligned}$$

Define $p_0 = m^{-\frac{h-1}{h(h-2)}}$. Clearly, if $\sum \mathcal{C}^- \geq \sum \mathcal{C}^+$, then by (2.23):

$$\mu(\mathcal{C}) = \mu(\mathcal{C}, n, m, p_-, p_+) = O\left(\mu(\mathcal{C}, n, m, p_0; t/p_0)\right).$$

Since $\tau_3(n, m, p_0, t/p_0) = O(1)$ by Lemma 2.4.6 we get $\mu' = \mu(\mathcal{C}, n, m, p_0, t/p_0) = O(1)$ and furthermore $\mu' \to 0$ if $\mathcal{C} \notin \{\mathcal{C}_M, \mathcal{C}_{in}\}$.

To see why $\mu(\mathcal{C}_{in}) \to 0$, note that $\sum \mathcal{C}_{in}^- > \sum \mathcal{C}_{in}^+$ and by (2.23) we have $\mu(\mathcal{C}_{in}, n, m, p_0; t/p_0) = o\left(\mu(\mathcal{C}_{in}, n, m, p_-, p_+)\right)$.

The case $\sum \mathcal{C}^+ > \sum \mathcal{C}^-$ is similar by symmetry: we use Lemma 2.4.5 with p_- and p_+ interchanged and replace τ_3 with τ_4, $\mathcal{C}_{in}$ with $\mathcal{C}_{out}$. □

Proof of Lemma 2.4.7 Suppose $\tau_3 \to \infty$. If $mp_-p_+ \to 0$ then by Lemma 2.4.1, D contains a copy of $\mathcal{C}_{in}$ whp. Otherwise let $\omega \to \infty$ not too fast so that $\omega = o\left(\tau_3^{1/(h-1)}\right)$ and let $p'_- = p_-/\omega$. Consider a sequence of random digraphs $\{D'(k)\}$ where $D'(k) = D(n, m, p'_-, p_+)$. By Lemma 2.4.1 and Lemma 2.3.1 $P(\mathcal{C}_{in} \in D'(k)) \to 1$ and so $P(\mathcal{C}_{in} \in D) \to 1$ by monotonicity.

Now suppose $\tau_3 = \Theta(1)$, so that $\mu(\mathcal{C}_{in}) = \tau_3^h = \Theta(1)$. In this case we have $mp_-p_+ = O(n^{-1/(h-1)}) \to 0$. Let $\mathcal{C}$ be a any simple diclique cover of $\overrightarrow{K}_h$. We will show that

$$\mu(\mathcal{C}) = O(1) \tag{2.24}$$

and furthermore we can replace $O()$ with $o()$ (respectively, $\Theta()$) if $\mathcal{C} \notin \mathcal{L}$ (respectively, $\mathcal{C} \in \mathcal{L}$).

Assuming (2.24) holds, since both τ_3 and $\mu(\mathcal{C})$ are monotone increasing multinomials in m, p_-, p_+ we have that $\tau_3 \to 0$ implies $\mu(\mathcal{C}) \to 0$. Therefore using Lemma 2.3.4, Lemma 2.4.1 and Remark 2.4.2 completes the proof of the lemma in the case where $mp_- \to 0$. For the boundary case $mp_- = \Theta(1)$, (2.24) and Lemma 2.4.9 below shows that each cover in $\mathcal{S}_{in}$ belongs to the leading set when $\alpha \geq \alpha_0$. If $\alpha = \alpha_0$ we use the fact that $\mu_1(\mathcal{C}_M) = \tau_1^h = \Theta(1)$ and Remark 2.4.2 to show that $\mathcal{C}_M$ also belongs to $\mathcal{L}$.

Let us check (2.24). We have

$$p_- = \left(\frac{\tau_3}{nmp_+}\right)^{1/(h-1)} = \Omega\left(m^{\frac{-1}{\alpha(h-1)}}\right) = \Omega\left(m^{-\frac{h-1}{h(h-2)}}\right)$$

and $m^{\frac{h-1}{h(h-2)}}p_- \to \infty$ when $\alpha > \alpha_0$. So we may apply Lemmas 2.4.3 and 2.4.4 as in the proof of Lemma 2.4.5 to get that

$$\mu(\mathcal{C}) = O(\mu(\mathcal{C}_{in})) = O(1) \tag{2.25}$$

Furthermore, Lemma 2.4.3 and Lemma 2.4.4 also give that we may replace $O()$ with $o()$ in (2.25) in all cases, except if

- $\mathcal{C} = \mathcal{C}_M$ and $m^{\frac{h-1}{h(h-2)}}p_- = \Theta(1)$ or
- $mp_+ = \Theta(1), |C^+| = 1$ for each $C \in \mathcal{C}$ and for each $j = 1, \ldots, h$ the diclique cover $\mathcal{C}_j$ obtained by taking all dicliques $C \in \mathcal{C}$ that have $C^+ = \{j\}$ satisfies $\sum \mathcal{C}_j^- = h - 1$.

The first exception occurs only if $\alpha = \alpha_0$ and $mp_+ = \Theta(1)$. The second exception represents all diclique covers $\mathcal{C} \in \mathcal{S}_{in}$. This completes the proof of (2.24). □

Proof of Lemma 2.4.8 Suppose $\tau_2 = n^{1/(h-1)} m p_- p_+ \to \infty$. If $m p_- p_+ \to 0$ the random digraph D contains a copy of $\mathcal{C}_R$ whp by Lemma 2.4.1. Otherwise, assume that $m p_- p_+ \to a \in (0; \infty]$. Let $p'_- = (m p_+ \ln n)^{-1} = o(p_-)$. Then the random digraph $D(n, m, p'_-, p_+)$ contains a copy of $\mathcal{C}_R$ whp by Lemma 2.4.1 since $m p'_- p_+ \to 0$ and $\tau_2 \to \infty$. Monotonicity and Lemma 2.3.1 imply that D contains a copy of $\mathcal{C}_R$ whp in all cases when $\tau_2 \to \infty$.

Now suppose that $\tau_2 = \Theta(1)$. Similarly as above, we have $\mu(\mathcal{C}_R) = \tau_2^{h-1} = \Theta(1)$ and $m p_- p_+ = O\left(n^{-1/(h-1)}\right) \to 0$. Let $\mathcal{C}$ be a simple diclique cover of $\overrightarrow{K}_h$. We will show that if $\mathcal{C} \not\mathcal{L}$ then $\mu(\mathcal{C}) \to 0$. As in the proof of Lemma 2.4.6 we will then be able to complete the proof using Lemma 2.4.1, Markov's inequality, Lemma 2.3.4 and Remark 2.4.2 (for the case $\alpha = \alpha_0$ notice that $\tau_1(n, m, p_-, p_+) = \Theta(1)$, so the birth threshold functions τ_1 and τ_2 become equivalent).

Suppose $\sum \mathcal{C}^- \geq \sum \mathcal{C}^+$. Using (2.22) we have

$$\mu(\mathcal{C}) = \mu(\mathcal{C}, n, m, p_-, p_+) = O\left(\mu(\mathcal{C}, n, m, t/p_0, p_0)\right)$$

where $t = p_- p_+$ and $p_0 = m^{-1}$. Also note that when $\mathcal{C}^- > \mathcal{C}^+$

$$\mu(\mathcal{C}) = o(\mu(\mathcal{C}, n, m, t/p_0, p_0)). \tag{2.26}$$

Since $\tau_3(n, m, t/p_0, p_0) = nm\left(n^{-1/(h-1)} m^{-1} p_0^{-1}\right)^{h-1} p_0 = \Theta(1)$ we can apply (2.25) from Lemma 2.4.7 for the sequence of random digraphs $\{D(n, m, t/p_0, p_0)\}$ to get that $\mu(\mathcal{C}, n, m, t/p_0, p_0) = O(1)$.

Lemma 2.4.7 also gives that $\mu(\mathcal{C}, n, m, t/p_0, p_0) \to 0$ for all simple diclique covers, except if $\mathcal{C} = \mathcal{C}_M$ in the case $\alpha = \alpha_0$ or if $\mathcal{C} \in \mathcal{S}_{in}$. It remains to check that $\mu(\mathcal{C}_0) \to 0$ for any diclique cover $\mathcal{C}_0 \in \mathcal{S}_{in}$ such that $\mathcal{C}_0 \neq \mathcal{C}_R$. But any $\mathcal{C}_0 \in \mathcal{S}_{in} \setminus \{\mathcal{C}_R\}$ has $\sum \mathcal{C}_0^- > \sum \mathcal{C}_0^+$, therefore $\mu(\mathcal{C}_0) \to 0$ by (2.26).

The case $\sum \mathcal{C}^- < \sum \mathcal{C}^+$ is similar because of symmetry. □

Lemma 2.4.9 *Let $h, \{D(k)\}, n, \alpha_0, m$ be as in Theorem 2.2.1. Suppose $p_- \to 0$ and $p_+ \to 0$, $\alpha \geq \alpha_0$ and $m p_+ = \Theta(1)$. Consider any diclique cover $\mathcal{C}_0 \in \mathcal{S}_{in}$. If $\tau_3 = \tau_3(n, m, p_-, p_+) = \Omega(1)$ then $P(\mathcal{C} \in D(k)) = \Omega(1)$ and if $\tau \to \infty$ then $D(k)$ contains $\mathcal{C}$ whp.*

Proof Write $D = D(k)$. By Lemma 2.4.7 we have $\mu(\mathcal{C}_0) = \Theta(1)$ whenever $\tau_3 = \Theta(1)$.

Assume first that $\tau_3 \to \infty$. We claim that for any non-empty set $S \subseteq [h]$:

$$\mu(\mathcal{C}_0[S]) = \Omega(\mu(\mathcal{C}_{in}[S])). \tag{2.27}$$

By (2.14) in the proof of Lemma 2.4.1 we have that $\mu(\mathcal{C}_{in}[S]) \to \infty$.

By Lemma 2.3.1 we may assume that (2.4) holds. Also, since $p_- \to 0$ we have $mp_-p_+ \to 0$.

First consider the case $S = \{v\}$ for some $v \in [h]$. By the definition of $\mathcal{S}_{in}$, $\mathcal{C}_0[S]$ consists of exactly $h-1$ dicliques $(\{v\}, \emptyset)$ and one or more dicliques $(\emptyset; \{v\})$. Let us apply (2.5). Since $mp_+ = \Theta(1)$, the set $\hat{\mathcal{C}}_0[S]$ is equal to $\hat{\mathcal{C}}_{in}[S]$, so (2.27) follows by (2.14) for the case $|S| = 1$.

Now suppose $|S| \geq 2$. Split a given diclique family $\mathcal{C}$ into the family of its proper dicliques $\mathcal{C}'$ and the family of its improper dicliques $\mathcal{C}''$ so that $\mathcal{C} = \mathcal{C}' \cup \mathcal{C}''$. Note that if $V(\mathcal{C}'') \subseteq V(\mathcal{C}')$ by (2.5) we have

$$\mathbb{E}\, X(\mathcal{C}) = \Theta(\mu(\mathcal{C}')\tilde{P}(\tilde{\mathcal{C}}'')) \tag{2.28}$$

where $\tilde{\mathcal{C}}''$ is the collection of dicliques C of $\mathcal{C}''$ that satisfy $\tilde{P}(C) \to 0$.

By Lemma 2.4.4 we have $\mu(\mathcal{C}_0[S]') = \Theta(\mu(\mathcal{C}_{in}[S]'))$. If $mp_-^s \to 0$ then $\tilde{\mathcal{C}}_{in}[S]''$ consists of $h - s$ dicliques $(S; \emptyset)$. Similarly, since $\mathcal{C}_0 \in \mathcal{S}_{in}$, the set $\tilde{\mathcal{C}}_0[S]''$ can be partitioned into $h - s$ families of improper dicliques $\mathcal{C}_1, \ldots, \mathcal{C}_{h-s}$ so that each $\mathcal{C}_i$ consists of pairwise disjoint dicliques, $\sum \mathcal{C}_i^- \leq s$ and $\sum \mathcal{C}_i^+ = 0$. We have for each $i = 1, \ldots, h - s$:

$$\prod_{C \in \mathcal{C}_i} \tilde{P}(C) = m^{|\mathcal{C}_i|} p_-^{\sum \mathcal{C}_i^-} = \Omega(\tilde{P}\,(S; \emptyset)).$$

If $mp_-^s \to a_s > 0$ then both $\tilde{\mathcal{C}}_0[S]''$ and $\tilde{\mathcal{C}}_{in}[S]''$ are empty. So in each case

$$\tilde{P}(\tilde{\mathcal{C}}_0[S]'') = \Theta\left(\prod_i \tilde{P}(\mathcal{C}_i)\right) = \Omega(\tilde{P}(\tilde{\mathcal{C}}_{in}[S]''))$$

and (2.27) follows by (2.28). Now (2.27) and Lemma 2.3.2 imply that D contains a copy of $\mathcal{C}_0$ whp when $\tau_3 \to \infty$.

Finally, we need to check that $P(\mathcal{C}_0 \in D) = \Omega(1)$ if $\tau_3 = \Theta(1)$. We note that in this case (2.27) still holds. Therefore the same argument as in the proof of Remark 2.4.2 shows that $P(\mathcal{C}_0 \in D) = \Omega(1)$. □

Proof of Remark 2.2.3 Write $D = D(k)$. Consider two simple diclique covers of $\overrightarrow{K}_2$, namely $\mathcal{C}_R = \mathcal{C}_{in} = \mathcal{C}_{out}$ and $\mathcal{C}_M$. Clearly, $\overrightarrow{K}_h \in D$ implies that $\mathcal{C}_R \in D$ or $\mathcal{C}_M \in D$.

Let us show that $\tau_2 \to \infty$ implies $P(\mathcal{C}_R \in D) \to 1$. If $mp_-p_+ \to 0$ this follows by Lemma 2.4.1. Otherwise, using Lemma 2.3.1 we may assume $mp_-p_+ \to c \in (0; \infty]$. Consider another sequence of random digraphs $\{D'(k)\}$ where $D'(k) =$

$D(n, m, (\omega m p_+)^{-1}, p_+)$ and $\omega = \ln n$. By Lemma 2.4.1 we have that $\mathcal{C}_R \in D'(k)$ whp, and by monotonicity $\mathcal{C}_R \in D$ whp.

Now suppose $\tau_2 = O(1)$. Then $\tau_1 = m^{-1/2}\tau_2 \to 0$ and $\mu(\mathcal{C}_M) = \tau_1^h \to 0$. This implies by (2.3) that any simple cover $\mathcal{C}$ of $\overrightarrow{K}_2$ such that $\mathcal{C} \neq \mathcal{C}_R$ must have $\mu(\mathcal{C}) \to 0$. By Lemma 2.4.1 and Lemma 2.3.4 for any proper diclique cover $\mathcal{C} \neq \mathcal{C}_R$ we have $P(\mathcal{C} \in D) \to 0$. Finally Remark 2.4.2 shows that $P(\mathcal{C}_R \in D) = \Omega(1)$ if $\tau_2 = \Theta(1)$ and so $\mathcal{C}_R$ is indeed the leading cover. □

2.5 Proofs of (2.1)-(2.8)

We will prove the estimates from Section 2.3.

Proof of (2.1) - (2.8) Similarly as in [58] we can represent D by two random n by m binary matrices R^- and R^+ where $R^-_{ij} = 1$ if and only if $j \in S^-(i)$ and $R^+_{ij} = 1$ if and only if $j \in S^+(i)$.

To prove (2.1) we will apply Lemma 2 from [58]. The probability of success p is the probability that a fixed key generates C on a fixed set $S \supseteq C^- \cup C^+$:

$$p = p_-^{|C^-|} p_+^{|C^+|} (1-p_-)^{|S|-|C^-|} (1-p_+)^{|S|-|C^+|}.$$

We have that $p \sim p_-^{|C^-|} p_+^{|C^+|}$ since $|S|$ is fixed and $p_-, p_+ \to 0$. Since C is proper and $mp_-p_+ \to 0$ we have that $mp \to 0$. By independence and the inclusion-exclusion principle (or by Lemma 2 of [58]) we have that $P(C) \sim mp \sim mp_-^{|C^-|} p_+^{|C^+|}$. Equation (2.2) follows from analogous reasoning as in the proof of Theorem 3 of [58]. For the random digraph D let N_C count the number of different attributes $w \in W$ that generate the diclique C (on the set $V(\mathcal{C})$). Suppose $\{C_1, C_2, \ldots, C_t\}$ are all distinct dicliques in $\mathcal{C}$ where C_i has multiplicity a_i, $i = 1 \ldots t$ and let $\{C_{t+1}, \ldots, C_M\}$ be the set of all proper dicliques on $V(\mathcal{C})$ that are not in $\mathcal{C}$. Then by Lemma 1 of [58]:

$$\begin{aligned} P(\mathcal{C}) &= P(N_{C_1} = a_1, \ldots, N_{C_t} = a_t, N_{C_{t+1}} = 0, \ldots, N_{C_M} = 0) \\ &\sim P(N_{C_1} \geq a_1, \ldots, N_{C_t} \geq a_t, N_{C_{t+1}} = 0, \ldots, N_{C_M} = 0) \\ &\sim \frac{P(C_1)^{a_1}}{a_1!} \frac{P(C_2)^{a_2}}{a_2!} \cdots \frac{P(C_t)^{a_t}}{a_t!} \sim a_{\mathcal{C}} m^{|\mathcal{C}|} p_-^{\sum \mathcal{C}^-} p_+^{\sum \mathcal{C}^+} \end{aligned}$$

since for $j > t$, $P(N_{C_j} = 0) = 1 - P(C_j) \to 1$. Now the equation (2.3) is immediate since $EX(\mathcal{C}) = \binom{n}{|V(\mathcal{C})|} P(\mathcal{C})$.

To see (2.5), recall that by Lemmas 1 and 2 of [58] we have for any diclique family $\mathcal{C}$:

$$P(\mathcal{C}) \sim \prod_{C\in\mathcal{C}} P(C) \sim K a_{\mathcal{C}} \prod_{C\in\hat{\mathcal{C}}} P(C)$$

where $K = \prod_{C\in\mathcal{C}\setminus\hat{\mathcal{C}}} P(C)$ is a constant.

The proof of the upper bounds (2.6) and (2.7) is much simpler: we sum the probability that D contains a fixed copy of $\mathcal{C}$ realised by a fixed tuple of attributes, $\prod_{C\in\mathcal{C}} \tilde{P}(C)$, over all $\binom{m}{a_1,\ldots,a_t}$ ways to pick a relevant tuple, and, for the bound (2.7), over all $\binom{n}{|V(\mathcal{C})|}$ sets of V of size $|V(C)|$. The estimate follows by the union bound. The estimate (2.8) follows similarly. □

Chapter 3

Large cliques in sparse random intersection graphs

3.1 Introduction

Bianconi and Marsili observed in 2006 [13] that "scale-free" real networks can have very large cliques; they gave an argument suggesting that the rate of divergence is polynomial if the degree variance is unbounded [13]. In a more precise analysis, Janson, Łuczak and Norros [55] showed exact asymptotics for the clique number in a power-law random graph model where edge probabilities are proportional to the product of weights of their endpoints.

Another feature of a real network that may affect formation of cliques is the clustering property: the probability of a link between two randomly chosen vertices increases dramatically after we learn about the presence of their common neighbour. An interesting question is whether and how the clustering property is related to the clique number.

With conditionally independent edges, the random graph of [55] does not have the clustering property and, therefore, can not explain such a relation.

Here we address this question by showing precise asymptotics for the clique number of a related random intersection graph model that admits a tunable clustering coefficient and power-law degree distribution. We find that the effect of clustering on the clique number only shows up for the degree sequences having a finite variance. We note that the finite variance is a necessary, but not sufficient condition for the clustering coefficient to attain a non-trivial value, see [18] and (3.5) below.

In the language of hypergraphs, we ask what is the largest intersecting family in a random hypergraph on the vertex set $[m]$, where n identically distributed and independent hyperedges have random sizes distributed according to P. A related problem for uniform hypergraphs was considered by Balogh, Bohman and Mubayi [4]. Although the motivation and the approach of [4] are different from ours, the result of [4] yields the clique number, for a particular class of random intersection graphs based on the subsets having the same (deterministic) number of elements.

We will consider a sequence $\{G(n)\} = \{G(n), n = 1, 2, \dots\}$ of random intersection graphs $G(n) = G(n, m, P)$, where $P = P(n)$ and $m = m(n) \to +\infty$ as $n \to +\infty$. Let $X(n)$ denote a random variable distributed according to $P(n)$ and define $Y(n) := \sqrt{\frac{n}{m}}X(n)$. If not explicitly stated otherwise, the limits below will be taken as $n \to \infty$. In this thesis we use the standard notation $o()$, $O()$, $\Omega()$, $\Theta()$, $o_P()$, $O_P()$, see, for example, [56]. For positive sequences $\{a_n\}$, $\{b_n\}$ we write $a_n \sim b_n$ if $a_n/b_n \to 1$ $a_n \ll b_n$ if $a_n/b_n \to 0$. For a sequence of events $\{\mathcal{A}_n\}$, we say that $\mathcal{A}_n$ occurs *whp*, if $\mathbb{P}(\mathcal{A}_n) \to 1$.

We will assume in what follows that

$$\mathbb{E}\, Y(n) = O(1). \tag{3.1}$$

This condition ensures that the expected number of edges in $G(n)$ is $O(n)$. Hence $G(n)$ is sparse. We remark, that if, in addition, $Y(n)$ converges in distribution to an integrable random variable, say Z, and $\mathbb{E}\, Y(n) \to \mathbb{E}\, Z$, then $G(n)$ has asymptotic degree distribution $Poiss(\lambda)$, where $\lambda = Z\mathbb{E}\, Z$, see, see Chapter 1 and [18]. In particular, if $Y(n)$ has asymptotic square integrable distribution, then $G(n)$ has asymptotic square integrable degree distribution too. Furthermore, if $Y(n)$ has a power-law asymptotic distribution, then $G(n)$ has asymptotic power-law degree distribution with the same exponent.

Our first result, Theorem 3.1.1, shows that in the latter case the clique number diverges polynomially. In fact, we do not require $Y(n)$ to have a limiting power-law distribution, but consider a condition that only involves the tail of $Y(n)$. Namely, we assume that for some $\alpha > 0$ and some slowly varying function L there is $0 < \epsilon_0 < 0.5$ such that for each sequence x_n with $n^{1/2-\epsilon_0} \le x_n \le n^{1/2+\epsilon_0}$ we have

$$\mathbb{P}\left(Y(n) \ge x_n\right) \sim L(x_n)x_n^{-\alpha}. \tag{3.2}$$

A function $L : \mathbb{R}_+ \to \mathbb{R}_+$ is called *slowly varying* if $\lim_{x\to\infty} L(tx)/L(x) = 1$ for any $t > 0$.

Theorem 3.1.1 *Let* $1 < \alpha < 2$. *Assume that* $\{G(n)\}$ *is a sequence of random* *intersection graphs satisfying (3.1), (3.2). Suppose that for some* $\beta > \max\{2 - \alpha, \alpha - 1\}$ *we have* $m = m(n) = \Omega(n^{\beta})$. *Then the clique number of* $G(n)$ *is*

$$\omega(G(n)) = (1 + o_P(1))\,(1 - \alpha/2)^{-\alpha/2}\,K(n) \tag{3.3}$$

where

$$K(n) = L\left((n \ln n)^{1/2}\right) n^{1-\alpha/2} (\ln n)^{-\alpha/2}.$$

We remark that adjacency relations of neighbouring vertices of a random intersection graph are statistically dependent events and this dependence is not negligible for $m = O(n)$. Theorem 3.1.1 says that in the case where the asymptotic degree distribution has infinite second moment ($\alpha < 2$), the asymptotic order (3.3) of a power-law random intersection graph is the same as that of the related model of [55] which has conditionally independent edges. Let us mention that the lower bound for the clique number $\omega(G(n))$ is obtained using a simple and elegant argument of [55], which is not sensitive to the statistical dependence of edges of $G(n)$. To show the matching upper bound we developed another approach based on a result of Alon, Jiang, Miller and Pritkin [1] in Ramsey theory.

In the case where the (asymptotic) degree distribution has a finite second moment we not only find the asymptotic order of $\omega(G(n))$, but also describe the structure of a maximal clique. To this aim, it is convenient to interpret attributes $w \in W$ as colours. The set of vertices $T(w) = \{v \in V : w \in S_v\}$ induces a clique in $G(n)$ which we denote (with some ambiguity of notation) $T(w)$. We say that every edge of $T(w)$ receives colour w and call this clique *monochromatic*. Note that $G(n)$ is covered by the union of monochromatic cliques $T(w)$, $w \in W$. We denote the size of the largest monochromatic clique by $\omega'(G(n))$. Clearly, $\omega(G(n)) \geq \omega'(G(n))$.

Denote $x \vee y = \max(x, y)$. The next theorem shows that the largest clique is a monochromatic clique (plus possibly a few extra vertices).

Theorem 3.1.2 *Assume that* $\{G(n)\}$ *is a sequence of random intersection graphs satisfying (3.1). Suppose that* $Var(Y(n)) = O(1)$. *Then*

$$\omega(G(n)) = \omega'(G(n)) + O_P(1).$$

If, in addition, for some positive sequence $\{\epsilon_n\}$ *converging to zero we have*

$$n\mathbb{P}(Y(n) > \epsilon_n n^{1/2}) \to 0 \tag{3.4}$$

then, for an absolute constant C,

$$\mathbb{P}\left(\omega(G(n)) \leq C \vee (\omega'(G(n)) + 3)\right) \to 1.$$

The condition (3.4) is not very restrictive. It is satisfied by uniformly square integrable sequences $\{Y(n)\}$. In particular, (3.4) holds if $\{Y(n)\}$ converges in distribution to a square integrable random variable, say Y_*, and $\mathbb{E}\, Y^2(n)$ converges to $\mathbb{E}\, Y_*^2$.

Next, we evaluate the size of the largest monochromatic clique. For this purpose we relate the random intersection graph to the balls into bins model. Let every vertex $v \in V$ throw $X_v := |S_v|$ balls into the bins $w_1, \ldots, w_m$ uniformly at random, subject to the condition that every bin receives at most one ball from each vertex. Then $\omega'(G(n))$ counts the maximum number of balls contained in a bin. Let $M(N, m)$ denote the maximum number of balls contained in any of m bins after N balls were thrown into m bins uniformly and independently at random. Our next result says that the probability distribution of $\omega'(G(n))$ can be approximated by that of $M(N, m)$, with $N \approx n\mathbb{E}\, X(n) = \mathbb{E}\,(X_1 + \cdots + X_n)$. The asymptotics of $M(N, m)$ are well known, see, e.g., Section 6 of Kolchin et al [60].

The *total variation distance* $d_{TV}(X, Y)$ between two random variables X and Y is the supremum over all (measurable) sets A of $|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|$.

Theorem 3.1.3 *Assume that $\{G(n)\}$ is a sequence of random intersection graphs satisfying $\mathbb{E}\, Y(n) = \Theta(1)$ and $VarY(n) = O(1)$. Then*

$$d_{TV}(\omega'(G(n)), M(\lfloor (mn)^{1/2} \mathbb{E}\, Y(n) \rfloor, m)) \to 0.$$

Remark 3.1.4 *For $n, m \to +\infty$ the relations $\mathbb{E}\, Y = \Theta(1), VarY = O(1)$ imply $n = O(m)$. In particular, the conditions of Theorem 3.1.3 rule out the case $m = o(n)$.*

Let us summarize our results about the clique number of a sparse random intersection graph $G(n)$ with a square integrable (asymptotic) degree distribution. We note that the conditional probability (called the clustering coefficient of $G(n)$)

$$\mathbb{P}(v_1 \sim v_2 | v_1 \sim v_3, v_2 \sim v_3) \approx (n/m)^{1/2} \mathbb{E}\, Y(n) / \mathbb{E}\, Y^2(n) \tag{3.5}$$

only attains a non-trivial value for $m = \Theta(n)$ and $\mathbb{E}\, Y^2(n) = \Theta(1)$. (Here $u \sim v$ is the event that u and v are adjacent in $G(n)$, i.e., $uv \in E(G(n))$.) In the latter case Theorems 3.1.2 and 3.1.3 together with the asymptotics for $M(N, m)$

(Theorem II.6.1 of [60]), imply that

$$\omega(G(n)) = \frac{\ln n}{\ln \ln n} \left(1 + o_P(1)\right).$$

In contrast, the clique number of a sparse Erdős-Rényi random graph $G(n, c/n)$ is at most 3, and in the model of [55], with square integrable asymptotic degree distribution, the largest clique has at most 4 vertices.

Each of our main results, Theorem 3.1.1 and Theorem 3.1.2, have corresponding simple polynomial algorithms that construct a clique of the optimal order whp. For a power-law graph with $\alpha \in (1; 2)$, it is the greedy algorithm of [55]: sort vertices in descending order according to their degree; traverse vertices in that order and 'grow' a clique, by adding a vertex if it is connected to each vertex in the current clique. For a graph with a finite degree variance we propose even simpler algorithm: for each pair of adjacent vertices, take any maximal clique formed by that pair and their common neighbours. Output the biggest maximal clique found in this way. More details and analysis of each of the algorithms are given in Section 3.4 below.

In practical situations a graph may be assumed to be distributed as a random intersection graph, but information about the subset size distribution may not be available. In such a case, instead of the condition (3.2) for the tail of the normalised subset size $Y(n)$, we may consider a similar condition for the tail of the degree $D_1(n)$ of the vertex $1 \in V$ in $G(n)$: there are constants $\alpha' > 1, \epsilon' > 0$ and a slowly varying function $L'(x)$ such that for any sequence t_n with $n^{1/2-\epsilon'} \le t_n \le n^{1/2+\epsilon'}$

$$\mathbb{P}(D_1(n) \ge t_n) \sim L'(t_n) t_n^{-\alpha'}. \tag{3.6}$$

The following lemma shows that, subject to an additional assumption, there is equivalence between conditions (3.2) and (3.6) .

Lemma 3.1.5 *Assume that $\{G(n)\}$ is a sequence of random intersection graphs such that for some $\epsilon > 0$ we have*

$$\mathbb{E}\, Y(n) \mathbf{I}_{Y(n) \ge n^{1/2-\epsilon}} \to 0. \tag{3.7}$$

Suppose that either $(\mathbb{E}\, Y(n))^2$ or $\mathbb{E}\, D_1(n)$ converges to a positive number, say, d.

Then both limits exist and are equal, $\lim \mathbb{E}\, D_1(n) = \lim(\mathbb{E}\, Y(n))^2 = d$. Furthermore, the condition (3.6) holds if and only if (3.2) holds. In that case, $\alpha' = \alpha$ and $L'(t) = d^{\alpha/2} L(t)$.

Thus, under a mild additional assumption (3.7), condition (3.2) of Theorem 3.1.1

can be replaced by (3.6). Similarly, the condition $VarY(n) = O(1)$ of Theorem 3.1.2 can be replaced by the condition $VarD_1(n) = O(1)$.

Lemma 3.1.6 *Assume that* $\{G(n)\}$ *is a sequence of random intersection graphs and for some positive sequence* $\{\epsilon_n\}$ *converging to zero we have*

$$\mathbb{E}\, Y^2(n)\mathbf{I}_{Y(n)>\epsilon_n n^{1/2}} \to 0. \tag{3.8}$$

Suppose that either $\mathbb{E}\, Y(n) = \Theta(1)$ *or* $\mathbb{E}\, D_1(n) = \Theta(1)$. *Then*

$$\mathbb{E}\, D_1(n) = (\mathbb{E}\, Y(n))^2 + o(1) \tag{3.9}$$

$$VarD_1(n) = (\mathbb{E}\, Y(n))^2(VarY(n) + 1) + o(1). \tag{3.10}$$

Cliques of random intersection graphs have been studied in [58], where edge density thresholds for emergence of small (constant-sized) cliques were determined, and in [92], where the Poisson approximation to the distribution of the number of small cliques was established. The clique number was studied in [81], see also [6], in the case, where $m \approx n^{\beta}$, for some $0 < \beta < 1$. We note that in the papers [58], [92], [81] a particular random intersection graph with the binomial distribution $P \sim Bin(p, m)$ was considered.

The rest of the Chapter is organized as follows. In Section 3.2 we study sparse random power-law intersection graphs with index $\alpha \in (1; 2)$, introduce the result on "rainbow" cliques in extremal combinatorics (Lemma 3.2.8) and prove Theorem 3.1.1. In Section 3.3 we relate our model to the balls and bins model and prove Theorem 3.1.2. In Section 3.4 we present and analyse algorithms for finding large cliques in $G(n, m, P)$. In Section 3.5 we prove Lemmas 3.1.5 and 3.1.6.

3.2 Power-law intersection graphs

3.2.1 Proof of Theorem 3.1.1

We start with introducing some notation. Given a family of subsets $\{S_v, v \in V'\}$ of an attribute set W', we denote $G(V', W')$ the *intersection graph* on the vertex set V' defined by this family: $u, v \in V'$ are adjacent (denoted $u \sim v$) whenever $S_u \cap S_v \neq \emptyset$. We say that an attribute $w \in W'$ *covers* the edge $u \sim v$ of $G(V', W')$ whenever $w \in S_u \cap S_v$. In this case we also say that the edge $u \sim v$ receives colour w. In particular, an attribute w covers all edges of the (monochromatic) clique subgraph T_w of $G(V', W')$ induced by the vertex set $T_w = \{v \in V' : w \in S_v\}$.

Given a graph H, we say that $G(V', W')$ *contains a rainbow* H if there is a subgraph $H' \subseteq G(V', W')$ isomorphic to H such that every edge of H' can be prescribed an attribute that covers this edge so that all prescribed attributes are different.

We denote by $e(G)$ the size of the set $E(G)$ of edges of a graph G. Given two graphs $G = (V(G), E(G))$ and $R = (V(R), E(R))$ we denote by $G \vee R$ the graph on vertices $V(G) \cup V(R)$ and with edges $E(G) \cup E(R)$. In what follows we assume that $V(G) = V(R)$ if not mentioned otherwise. Let t be a positive integer and let R be a non-random graph on the vertex set V'. Assuming that subsets S_v, $v \in V'$ are drawn at random, introduce the event $Rainbow(G(V', W'), R, t)$ that the graph $G(V', W') \vee R$ has a clique H of size $|V(H)| = t$ with the property that every edge of the set $E(H) \setminus E(R)$ can be prescribed an attribute that covers this edge so that all prescribed attributes are different.

In the case where every vertex v of the random intersection graph $G(n, m, P)$ includes attributes independently at random with probability $p = p(n)$, the size $X_v := |S_v|$ of the attribute set has binomial distribution $P \sim Binom(m, p)$. We denote such graph $G(n, m, p)$ and call it a *binomial* random intersection graph. We note that for $mp \to +\infty$ the sizes X_v of random sets are concentrated around their mean value $\mathbb{E}\, X_v = mp$. An application of Chernoff's bound (see, e.g., [72])

$$\mathbb{P}(|B - mp| > \epsilon mp) \leq 2e^{-\frac{1}{3}\epsilon^2 mp}, \tag{3.11}$$

where B is a binomial random variable $B \sim Binom(m, p)$ and $0 < \epsilon < 3/2$, implies

$$\mathbb{P}(\exists v \in [n] : |X_v - mp| > y) \leq n\mathbb{P}(|X_v - mp| > y) \to 0 \tag{3.12}$$

for any $y = y(n)$ such that $y/\sqrt{mp \ln n} \to \infty$ and $y/(mp) < 3/2$.

We write $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$.

Let us prove Theorem 3.1.1. For every member $G(n) = G(V, W)$ of a sequence $\{G(n)\}$ satisfying conditions of Theorem 3.1.1 and a number $\epsilon_1 \in (0, \epsilon_0)$ define the subgraphs $G_i \subseteq G(n)$, $i = 0, 1, 2$, induced by the vertex sets

$$V_0 = V_0(n) = \{v \in V(G(n)) : X_v < \theta_1\};$$
$$V_1 = V_1(n) = \{v \in V(G(n)) : \theta_1 \leq X_v \leq \theta_2\};$$
$$V_2 = V_2(n) = \{v \in V(G(n)) : \theta_2 < X_v\},$$

respectively. Here $X_v = |S_v|$ denotes the size of the attribute set prescribed to a vertex v and the numbers

$$\theta_1 = \theta_1(n) = m^{1/2} n^{-\epsilon_1}; \qquad \theta_2 = \theta_2(n) = \left((1-\alpha/2) m \ln n + m e_1\right)^{1/2},$$

with $e_1 = e_1(n) = \max(0, \ln L((n \ln n)^{1/2}))$. Note that $e_1 \equiv 0$ for $L(x) \equiv 1$. We have $V = V_0 \cup V_1 \cup V_2$ and $V_i \cap V_j = \emptyset$ for $i \neq j$. Theorem 3.1.1 follows from the three lemmas below. Let $K = K(n)$ be as in Theorem 3.1.1. The first lemma gives a lower bound for the clique number of $G(n)$.

Lemma 3.2.1 *For any $m = m(n)$*

$$\omega(G_2) = |V_2|(1 - o_P(1)) = (1 - o_P(1))\,(1-\alpha/2)^{-\alpha/2} K.$$

The next two lemmas provide an upper bound.

Lemma 3.2.2 *Suppose there is $\beta > \alpha - 1$ such that $m = \Omega(n^\beta)$. If $\epsilon_1 < \frac{\beta}{6}$ then there is $\delta > 0$ such that*

$$\mathbb{P}\left(\omega(G_0) \geq n^{1-\alpha/2-\delta}\right) \to 0.$$

Lemma 3.2.3 *Suppose there is $\beta > 2 - \alpha$ such that $m = \Omega(n^\beta)$. If $\epsilon_1 < \frac{\beta-2+\alpha}{24}$ then*

$$\omega(G_1) = o_P(K).$$

Proof of Theorem 3.1.1 We choose $0 < \epsilon_1 < \min\{(\alpha-1)/6, (\beta-2+\alpha)/24, \epsilon_0\}$. The theorem follows from the inequalities $\omega(G_2) \leq \omega(G) \leq \omega(G_0)+\omega(G_1)+\omega(G_2)$ and Lemmas 3.2.1, 3.2.2 and 3.2.3. □

3.2.2 Proof of Lemma 3.2.1

In this section we use ideas from [55] to give a lower bound on the clique number. We first note the following auxiliary facts.

Lemma 3.2.4 *Suppose $a = a_n, b = b_n$ are sequences of positive reals such that $0 < \ln 2b + 2a \to +\infty$. Let z_n be the positive root of*

$$a - \ln z - bz^2 = 0. \tag{3.13}$$

Then $z_n \sim \sqrt{\frac{2a+\ln(2b)}{2b}}$.

Proof Changing the variables $t = 2bz^2$ we get

$$t + \ln(t) = 2a + \ln(2b).$$

From the assumption it follows that $t + \ln t \sim t$ and therefore $z_n \sim \sqrt{\frac{2a+\ln(2b)}{2b}}$. □

Lemma 3.2.5 ([51]) *Let $x \to +\infty$. For any slowly varying function L and any $0 < t_1 < t_2 < +\infty$ the convergence $L(tx)/L(x) \to 1$ is uniform in $t \in [t_1, t_2]$. Furthermore, we have $\ln L(x) = o(\ln x)$.*

Proof of Lemma 3.2.1 Write $N = |V_2|$ and let

$$v^{(1)}, v^{(2)}, \dots, v^{(N)}$$

be the vertices of V_2 listed in an arbitrary order.

Consider a greedy algorithm for finding a clique in G proposed by Janson, Łuczak and Norros [55] (they use descending ordering by the set sizes, see also Section 3.4). Let $A^0 = \emptyset$. In the step $i = 1, 2, \dots, N$ let $A^i = A^{i-1} \cup \{v^{(i)}\}$ if $v^{(i)}$ is incident to each of the vertices $v^{(j)}$, $j = 1, \dots, i-1$. Otherwise, let $A^i = A^{i-1}$. This algorithm produces a clique H on the set of vertices A^N, and H demonstrates that $\omega(G_2) \geq |A^N|$.

Write $\theta = \theta_2$ and let $L_\theta = V_2 \setminus A^N$ be the set of vertices that failed to be added to A^N. We will show that

$$\frac{|L_\theta|}{N \vee 1} = o_P(1)$$

and

$$N = (1 - \alpha/2)^{-\alpha/2} L\left((n \ln n)^{1/2}\right) (\ln n)^{-\alpha/2} n^{1-\alpha/2} (1 - o_P(1)).$$

From (3.2) we obtain for $N \sim Binom(n, q)$ with $q = \mathbb{P}(X_n > \theta)$

$$\begin{aligned} \mathbb{E}\, N = nq &= n\mathbb{P}\left((m/n)^{1/2} Y_n > \theta\right) \\ &\sim L\left((n/m)^{1/2}\theta\right) n^{1-\alpha/2} m^{\alpha/2} \theta^{-\alpha} \\ &\sim (1 - \alpha/2)^{-\alpha/2} L(\sqrt{n \ln n}) (\ln n)^{-\alpha/2} n^{1-\alpha/2}. \end{aligned}$$

Here we used $L((n/m)^{1/2}\theta) \sim L(\sqrt{n \ln n})$ and $\ln L(\sqrt{n \ln n}) = o(\ln n)$, which follow using Lemma 3.2.5. Furthermore, by the concentration property of the binomial distribution, see, e.g., (3.11), we have $N = (1 - o_P(1))\mathbb{E}\, N$.

The remaining bound $|L_\theta|/(N \vee 1) \leq |L_\theta|/(N + 1) = o_P(1)$ follows from the bound $\mathbb{E}\left(L_\theta/(N + 1)\right) = o(1)$, which is shown below.

Let p_1 be the probability that two random independent subsets of $W = [m]$ of size $\lceil\theta\rceil$ do not intersect. The number of vertices in L_θ is at most the number of pairs in $x, y \in V_2$ where S_x and S_y do not intersect. Therefore by the first moment method

$$\mathbb{E}\,\frac{|L_\theta|}{N+1} = \mathbb{E}\,\mathbb{E}\left(\frac{|L_\theta|}{N+1}\Big|N\right) \le \mathbb{E}\,\mathbb{E}\left(\frac{\binom{N}{2}p_1}{N+1}\Big|N\right) \le \frac{p_1\mathbb{E}\,N}{2},$$

where

$$p_1 = \frac{\binom{m-\theta}{\theta}}{\binom{m}{\theta}} \le \left(1 - \frac{\theta}{m}\right)^\theta \le e^{-\theta^2/m}.$$

Now it is straightforward to check that for some constant c we have $p_1\mathbb{E}\,N \le c(\ln n)^{-\alpha/2} \to 0$. This completes the proof.

Let us briefly explain the intuition for the choice of θ. For simplicity assume $L(x) \equiv 1$ so that $e_1 = 0$. Could the same method yield a bigger clique if θ_2 is smaller? We remark that the product $p_1\mathbb{E}\,N$ as well as its upper bound $n^{1-\alpha/2}m^{\alpha/2}\theta^{-\alpha}e^{-\theta^2/m}$ (which we used above) are decreasing functions of θ. Hence, if we wanted this upper bound to be $o(1)$ then θ should be at least as large as the solution to the equation

$$n^{1-\alpha/2}m^{\alpha/2}\theta^{-\alpha}e^{-\theta^2/m} = 1$$

or, equivalently, to the equation

$$\alpha^{-1}\ln n + \frac{1}{2}\ln(m/n) - \ln\theta - \frac{\theta^2}{\alpha m} = 0. \tag{3.14}$$

Rewriting the latter relation in the form (3.13) where $a = \alpha^{-1}\ln n + (1/2)\ln(m/n)$ and $b = (\alpha m)^{-1}$ satisfy $be^{2a} = \alpha^{-1}n^{\frac{2}{\alpha}-1} \to +\infty$, we obtain from Lemma 3.2.4 that the solution θ of (3.14) satisfies

$$\theta \sim \sqrt{\frac{(2/\alpha)\ln n - \ln(n/m) + \ln(2/\alpha m)}{2/\alpha m}} \sim \sqrt{(1-\alpha/2)m\ln n}.$$

□

3.2.3 Proof of Lemma 3.2.2

Before proving Lemma 3.2.2 we collect some preliminary results.

Lemma 3.2.6 *Let h be a positive integer. Let $\{G(n)\}$ be a sequence of binomial random intersection graphs $G(n) = G(n, m, p)$, were $m = m(n)$ and $p = p(n)$*

satisfy $pn^{1/(h-1)}m^{1/2} \to a \in \{0,1\}$. *Then*

$$\mathbb{P}(G \textit{ contains a rainbow } K_h) \to a.$$

Proof The case $a = 1$ follows from Claim 2 of [58]. For the case $a = 0$ we have, by the first moment method,

$$\begin{aligned}\mathbb{P}(G \text{ contains a rainbow } K_h) &\le \binom{n}{h}(m)_{\binom{h}{2}} p^{2\binom{h}{2}} \\ &\le \left(n^{1/(h-1)}m^{1/2}p\right)^{h(h-1)} \to 0.\end{aligned}$$

□

Next is an upper bound for the size $\omega'(G)$ of the largest monochromatic clique.

Lemma 3.2.7 *Let* $1 < \alpha < 2$. *Assume that* $\{G(n)\}$ *is a sequence of random intersection graphs satisfying (3.1), (3.2). Suppose that for some* $\beta > \alpha - 1$ *we have* $m = \Omega(n^\beta)$. *Then there is a constant* $\delta > 0$ *such that* $\omega'(G(n)) \le n^{1-\alpha/2-\delta}$ *whp.*

Proof Let $X = X(n)$ and $Y = Y(n)$ be defined as in (3.1). Since for any $w \in W$ and $v \in V$

$$\mathbb{P}(w \in S_v) = \sum_{k=0}^{\infty} \frac{k}{m}\mathbb{P}(|S_v| = k) = \frac{\mathbb{E}\,X}{m} = \frac{\mathbb{E}\,Y}{\sqrt{mn}},$$

and the number of elements of the set $T_v = \{v : w \in S_v\}$ is binomially distributed

$$|T_w| \sim \mathit{Binom}\left(n, \frac{\mathbb{E}\,Y}{\sqrt{mn}}\right), \tag{3.15}$$

we have, for any positive integer k

$$\mathbb{P}(|T_w| \ge k) \le \binom{n}{k}\left(\frac{\mathbb{E}\,Y}{\sqrt{mn}}\right)^k \le \left(\frac{en}{k}\right)^k \left(\frac{\mathbb{E}\,Y}{\sqrt{mn}}\right)^k \le \left(\frac{c_1}{k}\sqrt{\frac{n}{m}}\right)^k$$

for $c_1 = e\sup_n \mathbb{E}\,Y$. Therefore, by the union bound,

$$\mathbb{P}\left(\omega'(G(n)) \ge k\right) \le m\left(\frac{c_1}{k}\sqrt{\frac{n}{m}}\right)^k.$$

Fix δ with $0 < \delta < \min((\beta - \alpha + 1)/4, 1 - \alpha/2, \beta/2)$. We have

$$\begin{aligned}\mathbb{P}\left(\omega'(G(n)) \geq n^{1-\alpha/2-\delta}\right) &\leq m\left(c_1 n^{\alpha/2-1/2+\delta} m^{-1/2}\right)^{\lceil n^{1-\alpha/2-\delta}\rceil}\\ &= m^{1-(\delta/\beta)\lceil n^{1-\alpha/2-\delta}\rceil}\left(c_1 n^{\alpha/2-1/2+\delta} m^{-1/2+\delta/\beta}\right)^{\lceil n^{1-\alpha/2-\delta}\rceil} \to 0\end{aligned}$$

since $m \to \infty$, $n^{1-\alpha/2-\delta} \to \infty$ and $m = \Omega(n^{\beta})$ implies

$$n^{\alpha/2-1/2+\delta} m^{-1/2+\delta/\beta} \to 0.$$

□

The last and the most important fact we need relates the maximum clique size with the maximum rainbow clique size in an intersection graph. An edge-colouring of a graph is called t-good if each colour appears at most t times at each vertex. We say that an edge-coloured graph contains a rainbow copy of H if it contains a subgraph isomorphic to H with all edges receiving different colours.

Lemma 3.2.8 ([1]) *There is a constant c such that every t-good coloured complete graph on more than $\frac{cth^3}{\ln h}$ vertices contains a rainbow copy of K_h.*

Proof of Lemma 3.2.2 Fix an integer $h > 1 + \frac{1}{\epsilon_1}$ and denote $t = n^{1-\alpha/2-\delta}$ and $k = \lceil \frac{cth^3}{\ln h} \rceil$, where positive constants δ and c are from Lemmas 3.2.7 and 3.2.8, respectively.

We first show that

$$\mathbb{P}(G_0 \text{ contains a rainbow } K_h) = o(1). \tag{3.16}$$

We note that for the binomial intersection graph $\tilde{G} = G(n, m, p)$ with $p = p(n) = m^{-1/2} n^{-\epsilon_1} + m^{-2/3}$ Lemma 3.2.6 implies

$$\mathbb{P}(\tilde{G} \text{ contains a rainbow } K_h) = o(1). \tag{3.17}$$

Let $\tilde{S}_v$ (respectively S_v), $v \in V$, denote the random subsets prescribed to vertices of $\tilde{G}$ (respectively $G(n)$). Given the set sizes $|S_v|, |\tilde{S}_v|$, $v \in V$, satisfying $|\tilde{S}_v| > \theta$, for each v, we couple the random sets of G_0 and $\tilde{G}$ so that $S_v \subseteq \tilde{S}_v$, for all $v \in V_0$. Now G_0 becomes a subgraph of $\tilde{G}$ and (3.16) follows from (3.17) and the fact that $\min_v |\tilde{S}_v| > \theta$ whp, see (3.12).

Next, we colour every edge $x \sim y$ of G_0 by an arbitrary element of $S_x \cap S_y$ and observe that the inequality $\omega'(G(n)) \leq t$ (which holds with probability $1 - o(1)$, by Lemma 3.2.7) implies that the colouring obtained is t-good. Furthermore, by Lemma 3.2.8, every k-clique of G_0 contains a rainbow clique; however the probability of the latter event is negligibly small by (3.16). We conclude that $\mathbb{P}(\omega(G_0) \geq k) = o(1)$ thus proving the lemma. □

3.2.4 Proof of Lemma 3.2.3

We start with a combinatorial lemma which is of independent interest.

Lemma 3.2.9 *Given positive integers $a_1, \ldots, a_k$, let $\{A_1, \ldots, A_k\}$ be a family of subsets of $[m]$ of sizes $|A_i| = a_i$. Let $d \geq k$ and let S be a random subset of $[m]$ of size d. Suppose that $a_1 + \cdots + a_k \leq m$. Then the probability*

$$\mathbb{P}\left(\{S \cap A_1, \ldots, S \cap A_k\} \text{ has a system of distinct representatives}\right) \tag{3.18}$$

is maximised when $\{A_i\}$ are mutually disjoint.

Proof Call any of $\binom{m}{d}$ possible outcomes c for S a configuration. Given $\mathcal{F} = \{A_1, \ldots, A_k\}$ let $\mathcal{C}_{DR}(\mathcal{F})$ be the set of configurations c such that $c \cap \mathcal{F} = \{c \cap A_1, \ldots, c \cap A_k\}$ has a system of distinct representatives. Write

$$p(\mathcal{F}) = \sum_{1 \leq i < j \leq k} |A_i \cap A_j|.$$

Suppose the claim is false. Out of all families that maximize (3.18) pick a family $\mathcal{F}$ with smallest $p(\mathcal{F})$. Then $p(\mathcal{F}) > 0$ and we can assume that there is an element $x \in [m]$ such that $x \in A_1 \cap A_2$. Since $\sum_{i=1}^{k} |A_i| \leq m$, there is an element y in the complement of $\bigcup_{A \in \mathcal{F}} A$.

Define $A_1' = (A_1 \setminus \{x\}) \cup \{y\}$ and consider the family $\mathcal{F}' = \{A_1', A_2, \ldots, A_k\}$. Observe that the family of configurations $\mathcal{C} = \mathcal{C}_{DR}(\mathcal{F}) \setminus \mathcal{C}_{DR}(\mathcal{F}')$ has the following property: for each $c \in \mathcal{C}$ we have $x \in c$ and it is not possible to find a set of distinct representatives for $c \cap \mathcal{F}$ where A_1 is matched with an element other than x (indeed such a set of distinct representatives, if existed, would imply $c \in \mathcal{C}_{DR}(\mathcal{F}')$). Consequently, there is a set of distinct representatives for sets $c \cap A_2, \ldots, c \cap A_k$ which does not use x. Since the latter set of distinct representatives together with y is a set of distinct representatives for $c \cap \mathcal{F}'$, we conclude that $c \notin \mathcal{C}_{DR}(\mathcal{F}')$ implies $y \notin c$.

Now, for $c \in \mathcal{C}$, let $c_{xy} = (c \cup \{y\}) \setminus \{x\}$ be the configuration with x and y swapped. Then $c_{xy} \notin \mathcal{C}_{DR}(\mathcal{F})$ and $c_{xy} \in \mathcal{C}_{DR}(\mathcal{F}')$, because $y \in c_{xy}$ and can be matched with A_1. Thus each configuration $c \in \mathcal{C}$ is assigned a unique configuration

$c_{xy} \in \mathcal{C}_{DR}(\mathcal{F}') \setminus \mathcal{C}_{DR}(\mathcal{F})$. This shows that $|\mathcal{C}_{DR}(\mathcal{F}')| \geq |\mathcal{C}_{DR}(\mathcal{F})|$. But $p(\mathcal{F}') \leq p(\mathcal{F}) - 1$, which contradicts our assumption about the minimality of $p(\mathcal{F})$. □

The next lemma is a version of a result of Erdős and Rényi about the maximum clique of the binomial random graph $G(n,p)$ (see, e.g., [56]).

Lemma 3.2.10 *Let $n \to +\infty$. Assume that probabilities $p_n \to 1$. Let $\{r_n\}$ be a positive sequence, satisfying $r_n = o(\tilde{K}^2)$, where $\tilde{K} = \frac{2\ln n}{1-p_n}$.*

There are positive sequences $\{\delta_n\}$ and $\{\epsilon_n\}$ converging to zero, such that $\delta_n \tilde{K} \to +\infty$ and for any sequence of non-random graphs $\{R_n\}$ with $V(R_n) = [n]$ and $e(R_n) \leq r_n$ the number X_n of cliques of size $\lfloor \tilde{K}(1+\delta_n) \rfloor$ in $G(n,p_n) \vee R_n$ satisfies

$$\mathbb{E}\, X_n \leq \epsilon_n.$$

Proof Write $p = p_n, r = r_n$ and $h = 1 - p$. Pick a positive sequence $\delta = \delta_n$ so that $\delta_n \to 0$ and $\ln^{-1} n + h + \frac{r}{\tilde{K}^2} = o(\delta)$. Let $a = \left\lfloor \tilde{K}(1+\delta) \right\rfloor$. We have

$$\mathbb{E}\, X_n \leq \binom{n}{a} p^{\binom{a}{2} - r} \leq \left(\frac{en}{a}\right)^a p^{\frac{a(a-1)}{2} - r} = e^{aB}, \tag{3.19}$$

where, by the inequality $\ln p \leq -h$, for n large enough,

$$\begin{aligned} B &\leq \ln(en/a) - \left(\frac{a-1}{2} - \frac{r}{a}\right) h \\ &\leq \ln n - \frac{ah}{2} + \frac{rh}{a} \leq (-1 + o(1))\delta \ln n \to -\infty. \end{aligned}$$

□

Lemma 3.2.11 *Let $\{G(n)\}$ be a sequence of binomial random intersection graphs, where $m = m_n \to +\infty$ and $p = p_n \to 0$ as $n \to +\infty$. Let $\{r_n\}$ be a sequence of positive integers. Denote $\bar{K} = 2e^{mp^2} \ln n$. Assume that $r_n \ll \bar{K}^2$ and*

$$mp^2 \to +\infty, \qquad \ln n \ll mp, \qquad \bar{K}p \to 0, \qquad \bar{K} \leq n/2. \tag{3.20}$$

There are positive sequences $\{\epsilon_n\}, \{\delta_n\}$ converging to zero such that $\delta_n \bar{K} \to +\infty$ and for any non-random graph sequence $\{R_n\}$ with $V(R_n) = V(G(n))$ and $e(R_n) \leq r_n$

$$\mathbb{P}\left(\mathit{Rainbow}(G(n), R_n, \bar{K}(1+\delta_n))\right) \leq \epsilon_n, \qquad n = 1, 2, \ldots \tag{3.21}$$

Here we choose $\{\delta_n\}$ such that $\bar{K}(1+\delta_n)$ were an integer.

Proof Let $\{x_n\}$ be a positive sequence such that

$$px_n \to 0, \quad x_n \ll mp \quad \text{and} \quad \sqrt{mp \ln n} \ll x_n$$

(one can take, e.g., $x_n = \varphi_n \sqrt{mp \ln n}$, with $\varphi_n \uparrow +\infty$ satisfying $\varphi_n^2 \bar{K} p \to 0$).

Given n, we truncate the random sets S_v, prescribed to vertices $v \in V$ of the graph $G = G(n, m, p)$, to the size $M = \lfloor mp + x_n \rfloor$. Denote

$$\bar{S}(v) = \begin{cases} S_v, \text{ if } |S_v| \le M, \\ M \text{ element random subset of } S_v, \text{ otherwise.} \end{cases}$$

We remark that for the event $B = \{S_v = \bar{S}_v, \forall v \in V\}$ Chernoff's bound implies

$$\mathbb{P}(B) = 1 - o(1). \tag{3.22}$$

Now, let $t \in [K; 2K]$ and let $T = \{u_1, \dots, u_t\}$ be a subset of V of size t. By R_T we denote the subgraph of R_n induced by the vertex set T. Given $i \in \{1, \dots, t\}$, let $T_i \subseteq \{u_1, \dots u_{i-1}\}$ denote the subset of vertices which are not adjacent to v_i in R_n. Let $A_T(i)$ denote the event that sets $\{\bar{S}_u \cap S_{u_i},\ u \in T_i\}$ have distinct representatives (in particular, none of the sets is empty). Furthermore, let A_T denote the event that all $A_T(i)$, $1 \le i \le t$ hold simultaneously

$$A_T = \bigcap_{i=1}^{t} A_T(i).$$

We shall prove below that whenever n is large enough

$$\mathbb{P}(A_T) \le \left(1 - (1-p)^M\right)^{\binom{t}{2} - e(R_T)}. \tag{3.23}$$

Next, proceeding as in Lemma 3.2.10 we find positive sequences $\{\delta'_n\}$, $\{\epsilon'_n\}$ converging to zero such that the number X'_n of subsets $T \subseteq V$ of size

$$a' = \left\lfloor \frac{2 \ln n}{(1-p)^M} (1 + \delta'_n) \right\rfloor$$

that satisfy the event A_T has expected value $\mathbb{E}\, X'_n \le \epsilon'_n$. For this purpose, we apply (3.19) to a' and $p' = 1 - (1-p)^M$, and use (3.23). We remark that $a' = \bar{K}(1 + \delta''_n)$, where $\{\delta''_n\}$ converges to zero and $\delta'' \bar{K} \to +\infty$. Indeed, we have $\delta'_n \ln n / (1-p)^M \to +\infty$, by Lemma 3.2.10, and we have $(1-p)^M = e^{-mp^2 - O(px + mp^3)}$ with $px + mp^3 = o(1)$. In particular, for large n, we have $a' \in [\bar{K}, 2\bar{K}]$.

The key observation of the proof is that events B and $Rainbow(G, R_n, a')$ imply $X'_n > 0$. Hence,

$$\mathbb{P}(Rainbow(G, R_n, a') \cap B) \leq \mathbb{P}(X'_n > 0) \leq \mathbb{E}\, X'_n \leq \epsilon'_n.$$

In the last step we used Markov's inequality. Finally, invoking (3.22) we obtain (3.21).

It remains to show (3.23). We write

$$\mathbb{P}(A_T) = \prod_{i=1}^{t} \mathbb{P}\left(A_T(i) | A_T(1), \ldots, A_T(i-1)\right)$$

and evaluate, for $1 \leq i \leq t$,

$$\mathbb{P}(A_T(i) | A_T(1), \ldots, A_T(i-1)) \leq \left(1 - (1-p)^M\right)^{|T_i|}. \tag{3.24}$$

Now (3.23) follows from the simple identity $\sum_{1 \leq i \leq t} |T_i| = \binom{t}{2} - e(R_T)$. Let us prove (3.24). For this purpose we apply Lemma 3.2.9. We first condition on $\{\bar{S}_u, u \in T_i\}$ and the size $|S_{v_i}|$ of S_{v_i}. By Lemma 3.2.9 the conditional probability

$$\mathbb{P}(A_T(i) \,\big|\, \bar{S}_u,\ u \in T_i,\ |S_{v_i}|)$$

is maximized when the sets $\bar{S}_u$, $u \in T_i$ are mutually disjoint (at this step we check the condition of Lemma 3.2.9 that $\sum_{u \in T_i} |\bar{S}_u| \leq tM < m$, for large n). Secondly, we drop the conditioning on $|S_{v_i}|$ and allow S_{v_i} to choose its element independently at random with probability p. In this way we obtain (3.24). □

Lemma 3.2.12 *Let $\{G(n)\}$ be a sequence of random binomial intersection graphs, where $m = m(n) \to +\infty$ and $p = p(n) \to 0$ as $n \to +\infty$. Assume that*

$$np = O(1), \qquad m(np)^3 \ll \bar{K}^2,$$

where $\bar{K} = 2e^{mp^2} \ln n$. Assume, in addition, that (3.20) holds.

Then there is a sequence $\{\delta_n\}$ converging to zero such that $\delta_n \bar{K} \to +\infty$ and

$$\mathbb{P}\left(\omega(G(n)) > \bar{K}(1 + \delta_n)\right) \to 0.$$

Proof Given n, let U be a random subset of $V = V(G(n))$ with binomial number of elements $|U| \sim Bin(n, p)$ and such that, for any $k = 0, 1, \ldots,$ conditionally, given the event $|U| = k$, the subset U is uniformly distributed over the class of subsets of V of size k. Recall that $T_w \subseteq V$ denotes the set of vertices that have chosen an attribute $w \in W$. We remark that T_w, $w \in W$ are iid random subsets having the same probability distribution as U.

We call an attribute w *big* if $|T_w| \geq 3$, otherwise w is *small*. Let W_B and W_S denote the sets of big and small attributes. Denote by G_B (respectively, G_S) the subgraph of $G = G(n)$ consisting of edges covered by big (respectively, small) attributes. We observe that, given G_B, the random sets T_z, $z \in W_S$, defining the edges of G_S are (conditionally) independent. We are going to replace them by bigger sets, denoted T'_z, by adding some more elements as follows. Given T_z, we first generate independent random variables $\mathbb{I}_z$ and $|\Delta_z|$, where $\mathbb{I}_z$ has Bernoulli distribution with success probability $p' = \mathbb{P}(|U| \leq 2)$ and where $\mathbb{P}(|\Delta_z| = k) = \mathbb{P}(|U| = k)/(1-p')$, $k = 3, 4, \ldots$. Secondly, for $\mathbb{I}_z = 1$ we put $T'_z = T_z$. Otherwise we put $T'_z = T_z \cup \Delta_z$, where Δ_z is a subset of $V \setminus T_z$ of size $|\Delta_z| - |T_z| \geq 1$ drawn uniformly at random. We note that given G_B, the random sets T'_z, $z \in W_S$ are (conditionally) independent and have the same probability distribution as U. Next we generate independent random subsets T'_w of V, for $w \in W_B$, so that they have the same distribution as U and were independent of G_S, G_B and T'_z, $z \in W_S$. Given G_B, the collection of random sets $\{T'_w, w \in W_B \cup W_S\}$ defines the binomial random intersection graph G' having the same distribution as $G(n, m, p)$.

We remark that $G_S \subseteq G'$ and every edge of G_S can be assigned a unique small attribute that covers this edge and the assigned attributes are all different. On the other hand, the graph G_B is relatively small. Indeed, since each w covers $\binom{|T_w|}{2}$ edges, the expected number of edges of G_B is at most

$$\mathbb{E} \sum_{w \in W} \binom{T_w}{2} \mathbb{I}_{\{|T_w| \geq 3\}} = m \mathbb{E} \binom{T_w}{2} \mathbb{I}_{\{|T_w| \geq 3\}} \leq m \sum_{k \geq 3} \binom{k}{2} \binom{n}{k} p^k.$$

Invoking the simple bound

$$\sum_{k \geq 3} \binom{k}{2} \binom{n}{k} p^k \leq (np)^2 (e^{np} - 1)/2 = O((np)^3)$$

we obtain $\mathbb{E}\, e(G_B) = O(m(np)^3)$.

Now we choose an integer sequence $\{r_n\}$ such that $m(np)^3 \ll r_n \ll \bar{K}^2$ and write, for an integer $K' > 0$,

$$\mathbb{P}\left(\omega(G) \geq K'\right) \leq \mathbb{E}\, \mathbb{P}\left(\omega(G) \geq K' | G_B\right) \mathbb{I}_{\{e(G_B) \leq r_n\}} + \mathbb{P}\left(e(G_B) \geq r_n\right). \quad (3.25)$$

Here, by Markov's inequality, $\mathbb{P}(e(G_B) \geq r_n) \leq r_n^{-1}\mathbb{E}\, e(G_B) = o(1)$. Furthermore, we observe that $\omega(G) \geq K'$ implies the event $Rainbow(G', G_B, K')$. Hence,

$$\mathbb{P}\left(\omega(G) \geq K'|G_B\right) \leq \mathbb{P}\left(Rainbow(G', G_B, K')|G_B\right).$$

We choose $K' = \bar{K}(1+\delta_n)$ and apply Lemma 3.2.11 to the conditional probability on the right. At this point we specify $\{\delta_n\}$ and find $\epsilon_n \downarrow 0$ such that $\mathbb{P}\left(Rainbow(G', G_B, K')|G_B\right) \leq \epsilon_n$ uniformly in G_B satisfying $e(G_B) \leq r_n$. Hence, (3.25) implies $\mathbb{P}\left(\omega(G) \geq \bar{K}(1+\delta_n)\right) \leq \epsilon_n + o(1) = o(1)$. □

Now we are ready to prove Lemma 3.2.3.

Proof of Lemma 3.2.3 Let

$$0 < \epsilon < 2^{-1}\min\{1,\, 1-2^{-1}\alpha,\, \beta - 2 + \alpha - 6\alpha\epsilon_1\} \tag{3.26}$$

and let $\bar{G}_1$ be the subgraph of G_1 induced by vertices $v \in V_1$ with $X_v \leq \theta$. Here $\theta^2 = (1-\varepsilon-2^{-1}\alpha)m\ln n$. Let $D = |V(G_1)\setminus V(\bar{G}_1)|$ denote the number of vertices of G_1 that do not belong to $\bar{G}_1$.

To prove the lemma we write $\omega(G_1) \leq D + \omega(\bar{G}_1)$ and show that each summand on the right is of order $o_P(K)$ for appropriately chosen $\epsilon = \epsilon(n) \to 0$.

Using (3.2) and Lemma 3.2.5 we estimate the expected value of D for $n \to +\infty$

$$\mathbb{E}\, D = n\left(\mathbb{P}(X_v \geq \theta) - \mathbb{P}(X_v \geq \theta_2)\right) \leq (h(\epsilon)+o(1))K. \tag{3.27}$$

Here $h(\epsilon) := (1-\epsilon-2^{-1}\alpha)^{-\alpha/2} - (1-2^{-1}\alpha)^{-\alpha/2} \to 0$ as $\varepsilon \to 0$. Letting $\epsilon \to 0$ we obtain from (3.27) that $D = o_P(K)$.

We complete the proof by showing that for any ε satisfying (3.26)

$$\mathbb{P}\left(\omega(\bar{G}_1) \geq 4n^{1-2^{-1}\epsilon-2^{-1}\alpha}\ln n\right) = o(1). \tag{3.28}$$

Note that $n^{1-2^{-1}\epsilon-2^{-1}\alpha}\ln n \ll K$.

Let $\bar{N}$ be a binomial random variable, $\bar{N} \sim Bin(n, \mathbb{P}(X_v > \theta_1))$, and let

$$\bar{n} = (1+\epsilon)n^{1-2^{-1}\alpha+\alpha\epsilon_1}L(n^{0.5-\epsilon_1}) \qquad \text{and} \qquad \bar{p}^2 = (1-2^{-1}\epsilon-2^{-1}\alpha)m^{-1}\ln n.$$

We couple $\bar{G}_1$ with the binomial random intersection graph $G' = G(\bar{n}, m, \bar{p})$ so that the event that $\bar{G}_1$ is isomorphic to a subgraph of G', denoted $\bar{G}_1 \subseteq G'$, has probability

$$\mathbb{P}(\bar{G}_1 \subseteq G') = 1 - o(1). \tag{3.29}$$

We argue that such a coupling is possible because the events $A = \{$every vertex of G' is prescribed at least θ attributes$\}$ and $B = \{|V(\bar{G}_1)| \le \bar{n}\}$ have very high probabilities. Indeed, the bound $\mathbb{P}(A) = 1-o(1)$ follows from Chernoff's inequality (3.12). To get the bound $\mathbb{P}(B) = 1-o(1)$ we first couple binomial random variables $|V(\bar{G}_1)| \sim Bin(n, \mathbb{P}(\theta_1 < X_v < \theta))$ and $\bar{N}$ so that $\mathbb{P}(|V(\bar{G}_1)| \le \bar{N}) = 1$ and then invoke the bound $\mathbb{P}(\bar{N} \le \bar{n}) = 1 - o(1)$, which follows from Chernoff's inequality.

Next we apply Lemma 3.2.12 to G' and obtain the bound

$$\mathbb{P}\left(\omega(G') > 4n^{1-2^{-1}\epsilon-2^{-1}\alpha} \ln \bar{n}\right) = o(1), \tag{3.30}$$

which together with (3.29) implies (3.28). □

3.3 Finite variance

In this section we prove Theorem 3.1.2. We note that the random power-law graph studied by Janson, Łuczak and Norros [55] whp does not contain K_4 as a subgraph if the degree distribution has a finite second moment. In our case a similar result holds for the rainbow K_4. Given a sequence of random intersection graphs $\{G(n)\}$, we show that the number of rainbow K_4 subgraphs of $G(n)$ is stochastically bounded as $n \to +\infty$ provided that the sequence of the second moments of the degree distributions is bounded. If, in addition, the sequence of degree distributions is uniformly square integrable, then $G(n)$ has no rainbow K_4 whp, see Lemma 3.3.3 below. We use these observations in the proof of Theorem 3.1.2.

3.3.1 Large cliques and rainbow K_4

Let U be a finite set and let $\mathcal{C} = \{C_1, \dots, C_r\}$ be a collection of (not necessarily distinct) subsets of U. We consider the complete graph K_U on the vertex set U and interpret subsets C_i as colours: an edge $x \sim y$ receives colour C_i (or just i) whenever $\{x, y\} \subseteq C_i$. We call $\mathcal{C}$ a *clique cover* if every edge of the clique K_U receives at least one colour. The edges spanned by the vertex set C_i form a subclique, which we call the *monochromatic clique* of colour i. We say that a vertex set $S \subseteq U$ is a *witness of a rainbow clique* if every edge of the clique K_S induced by S receives a non-empty collection of colours and it is possible to assign each edge one of its colours so that all edges of K_S were assigned different colours. For example, the collection $\mathcal{C} = \{A, B, C\}$, where $A = \{1, 2, 3\}$,

$B = \{1, 3, 4\}$ and $C = \{2, 4, 3\}$ is a clique cover of the set $\{1, 2, 3, 4\}$. It produces three monochromatic triangles and four rainbow triangles.

We start with a result that relates clique covers to rainbow clique subgraphs. For a clique cover $\mathcal{C} = \{C_1, \ldots, C_r\}$ denote by $p(\mathcal{C}) = \max_{i \neq j} |C_i \cap C_j|$ the size of maximum pairwise intersection.

Lemma 3.3.1 *Let k and p be positive integers. Let $h = h(k) > 0$ denote the smallest integer such that $\binom{h}{4} \geq k$. Let $\mathcal{C}$ be a clique cover of a finite set U and assume that* $\max_{C \in \mathcal{C}} |C| \geq |U| - h$ *and $p(\mathcal{C}) \leq p$.*

If, in addition, $|U| \geq t(k,p)$, where $t(k,p) = c\frac{h^3}{\ln h} p \left(\sqrt{2k} + 5 + 2p\right)$, then $\mathcal{C}$ produces at least k witnesses of rainbow K_4. Here c is the absolute constant of Lemma 3.2.8.

Proof Write $b = \max_i |C_i|$. We note that $\mathcal{C}$ has no rainbow K_h since otherwise there would be at least $\binom{h}{4} \geq k$ copies of rainbow K_4. Observe, that every monochromatic subclique of K_U has at most b vertices. Hence, each colour appears at most $b - 1$ times at each vertex of K_U. By Lemma 3.2.8, K_U has at most $c(b-1)h^3/\ln h$ vertices. That is, $b > a|U|$, where $a = \frac{\ln h}{ch^3}$ and c is an absolute constant. Fix $B \in \mathcal{C}$ with $|B| = b$ and a subset $S \subseteq U \setminus B$ of size h, say $S = \{x_1, \ldots, x_h\}$. Here we use the assumption $|U| \geq b + h$ telling that $U \setminus B$ has at least h elements, $|U \setminus B| = |U| - b \geq h$. We remark, that at least one pair of vertices of S, say $\{x_1, x_2\}$, receives at most 5 colours (it is covered by at most 5 sets from $\mathcal{C}$). Indeed, otherwise every edge of K_S received at least 6 distinct colours and, thus, each $S' \subseteq S$ of size $|S'| = 4$ induced a rainbow K_4. This contradicts to our assumption that there are fewer than $k \leq \binom{h}{4}$ rainbow copies of K_4.

We observe that the set of colours received by the pair $\{x_1, x_2\}$ is non-empty (since $\mathcal{C}$ is a clique cover) and fix one such colour, say $C_{x_1,x_2} \in \mathcal{C}$. Now, consider the set of pairs $\{\{x_1, y\},\ y \in B\}$ and pick a smallest family of sets from $\mathcal{C}$ such that each pair were covered by a member of the family (the smallest family means that any other family with fewer members would leave at least one uncovered pair). Since each member of the family intersects with B in at most p vertices (condition of the lemma) we conclude that such a family contains at least $\lceil b/p \rceil$ members. Furthermore, since the family is minimal, every member covers a pair $\{x_1, y\}$ which is not covered by other members. Hence, we can pick a set $B_1 \subseteq B$ of size $\lceil b/p \rceil$ so that every $\{x_1, y\}$, $y \in B_1$ is covered by a unique member, say $C_{x_1,y}$, of the family.

Next, remove from B_1 the elements y such that $x_2 \in C_{x_1,y}$ (there are at most 5 of them). Then remove those elements y which belong to the set C_{x_1,x_2} (there are at most p of them, since $|C_{x_1,x_2} \cap B| \leq p$). Call the newly formed set B'. Notice that

$$b' := |B'| \geq \frac{b}{p} - 5 - p > \frac{a|U|}{p} - 5 - p.$$

Let us consider the clique $\tilde{K}$ on the vertex set $B' \cup \{x_1, x_2\}$. For $y \in B'$, colour each edge $\{x_1, y\}$ of $\tilde{K}$ with the colour $C_{x_1,y}$. Colour the edge $\{x_1, x_2\}$ with C_{x_1,x_2} and for every edge $\{y_i, y_j\} \in B'$ use the colour B. Finally, for $y \in B'$, assign $\{x_2, y\}$ an arbitrary colour from the set of colours received by $\{x_2, y\}$ from the clique cover $\mathcal{C}$.

We claim that for any $y_1 \in B'$ and any $y_2 \in B' \setminus C_{x_2,y_1}$, the set $\{x_1, x_2, y_1, y_2\}$ witnesses a rainbow K_4. Indeed, by the construction, the colour C_{x_1,x_2} of the edge $\{x_1, x_2\}$ occurs only once, because $B' \cap C_{x_1,x_2} = \emptyset$. Similarly, for $x_1, x_2 \notin B$, the colour B of $\{y_1, y_2\}$ occurs only once. The colours of the two other edges incident to x_1 occur only once, since we removed all candidates y such that $x_2 \in C_{i_{x_1,y}}$, while constructing the set B'. Finally, we have $C_{x_2,y_1} \neq C_{x_2,y_2}$ since we chose y_2 outside C_{x_2,y_1}.

How many such witnesses can we form? For any y_1 we choose $|B'| - |B' \cap C_{x_2,y_1}| \geq |B'| - p$ suitable y_2. Repeating this for each y_1 we will produce every 4-set at most twice. Therefore $\tilde{K}$ contains at least

$$\frac{b'(b'-p)}{2} \geq \frac{1}{2}\left(\frac{a|U|}{p} - 5 - 2p\right)^2 \tag{3.31}$$

witnesses of rainbow K_4. But since the total number of witnesses of rainbow K_4 produced by $\mathcal{C}$ is less that k, the right-hand side of (3.31) is less than k. We obtain the inequality

$$|U| < \frac{p}{a}\left(\sqrt{2k} + 5 + 2p\right) = t(k, p),$$

which contradicts to the condition $|U| \geq t(k, p)$. □

In the remaining part of the subsection 3.3.1 we interpret attributes $w \in W$ as colours assigned to edges of a random intersection graph.

Lemma 3.3.2 *Let $G = G(k, m, P)$ be a random intersection graph and let $X_1, \ldots,$ X_k denote the sizes of random sets defining G. For any integers $x_1, \ldots, x_k$ such that the event $B = \{X_1 = x_1, \ldots, X_k = x_k\}$ has positive probability, we have*

$$\mathbb{P}(G \textit{ has a rainbow } K_k | B) \leq m^{-\frac{k(k-1)}{2}} (x_1 x_2 \ldots x_k)^{k-1}.$$

Proof Our intersection graph produces a rainbow clique on its k vertices whenever for some injective mapping, say f, from the set of pairs of vertices to the set of attributes, the event $A_f = \{\text{every pair } \{x,y\} \text{ is covered by } f(\{x,y\})\}$ occurs. By the independence, $\mathbb{P}(A_f|B) = \prod_i \frac{(x_i)_{k-1}}{(m)_{k-1}}$. Since there are $(m)_{\binom{k}{2}}$ possibilities to choose the map f, we obtain, by the union bound,

$$\mathbb{P}(G \text{ has a rainbow } K_k|B) \leq (m)_{\binom{k}{2}} \prod_i \frac{(x_i)_{k-1}}{(m)_{k-1}} \leq \frac{(x_1 x_2 \dots x_k)^{k-1}}{m^{k(k-1)/2}}.$$

□

Lemma 3.3.3 *Let $\{G(n)\}$ be a sequence of random intersection graphs such that $\mathbb{E}\, Y(n)^2 = O(1)$. Then the number $R = R(n)$ of 4-sets $S \subseteq V(G(n))$ that witness a rainbow K_4 in $G(n)$ satisfies as $n \to +\infty$*

$$\mathbb{E}\, R \leq \frac{(\mathbb{E}\, Y^2)^4}{4!} = O(1).$$

Furthermore, if for some positive sequence $\epsilon_n \to 0$ we have $n\mathbb{P}(Y(n) \geq \epsilon_n n^{1/2}) \to 0$ then $G(n)$ does not contain a rainbow K_4 whp.

Proof of Lemma 3.3.3 Denote $X_v = |S_v(n)|$ and $Y = Y(n)$. We write, using symmetry and the bound of Lemma 3.3.2,

$$\mathbb{E}\, R = \sum_{S \subseteq V, |S|=4} \mathbb{P}(S \text{ witnesses a rainbow } K_4) \leq \binom{n}{4} \mathbb{E}\left(\frac{(X_1 X_2 X_3 X_4)^3}{m^6} \wedge 1 \right).$$

Next, we apply the simple inequality $a^6 \wedge 1 \leq a^4$ and bound the right-hand side from above by $\frac{n^4}{4!} \frac{\mathbb{E}\,(X_1 X_2 X_3 X_4)^2}{m^4} = \frac{(\mathbb{E}\, Y^2)^4}{4!}$.

For the second part of the lemma, let $b = b(n) = \epsilon_n \sqrt{m}$ and let $A = A(n)$ be the event that $\max_{v \in V} X_v \leq b$. Let $\bar{A}$ denote the complement event. We write

$$\mathbb{P}(R \geq 1) \leq \mathbb{P}(R \geq 1, A) + \mathbb{P}(\bar{A}) \leq \mathbb{E}\, R\mathbf{I}_A + \mathbb{P}(\bar{A}). \tag{3.32}$$

By the union bound the second term is at most

$$n\mathbb{P}(X > b) = n\mathbb{P}(Y > \epsilon_n n^{1/2}) \to 0.$$

The first term by Lemma 3.3.2 satisfies

$$\mathbb{E}\, R\mathbf{I}_A \leq \binom{n}{4} m^{-6} \mathbb{E}\,(X_1 X_2 X_3 X_4)^3 \mathbf{I}_A \leq \frac{(\mathbb{E}\, X^2)^4 n^4 b^4}{4! m^6} = (\epsilon_n \mathbb{E}\, Y^2)^4 = o(1).$$

□

The next result shows that the structure of random intersection graphs with $\mathbb{E}\, Y(n)^2 = O(1)$ is relatively simple.

Lemma 3.3.4 *Let $\{G(n)\}$ be a sequence of random intersection graphs. Assume that $\mathbb{E}\, Y(n)^2 = O(1)$ and $m(n) \to \infty$ as $n \to +\infty$. Then whp each pair $\{w', w''\}$ of attributes is shared by at most two vertices of $G(n)$.*

The lemma says that the intersection of any two monochromatic cliques of $G(n)$ consists of at most one edge whp.

Proof For any pair of attributes w', w'' and a vertex v of $G(n)$, we have

$$\mathbb{P}(w', w'' \in S_v) = \sum_{k=0}^{m} \mathbb{P}(|S_v| = k) \frac{k(k-1)}{m(m-1)} = \frac{\mathbb{E}\, X^2 - \mathbb{E}\, X}{m(m-1)}$$
$$\leq \frac{\mathbb{E}\, Y^2}{n(m-1)} \leq \frac{c}{nm}.$$

Here $c > 0$ does not depend on m and n. By the union bound, the probability that there is a pair of attributes shared by k or more vertices is at most

$$\binom{m}{2}\binom{n}{k} \mathbb{P}(w', w'' \in S_v)^k \leq m^2 \left(\frac{en}{k}\right)^k \left(\frac{c}{nm}\right)^k \leq m^2 \left(\frac{ec}{km}\right)^k.$$

This probability tends to zero for any $k \geq 3$. □

Proof of Theorem 3.1.2 Let $R = R(n)$ denote the number of 4-sets $S \subseteq V(G(n))$ witnessing rainbow K_4 in $G(n)$. By Lemma 3.3.4, the intersection of any two monochromatic cliques has at most 2 vertices whp. In that case, by Lemma 3.3.1 (applied to the set of vertices U of the largest clique) either $\omega(G(n)) < t(R+1, 2)$ or $\omega(G) \leq \omega'(G) + h(R+1)$. Thus,

$$\omega(G(n)) \leq \omega'(G(n)) + Z(n)$$

where $Z(n) = t(R+1, 2) + h(R+1) = O_P(1)$, by Lemma 3.3.3.

If $n\mathbb{P}(Y(n) > \epsilon_n n^{1/2}) \to 0$ for some $\epsilon_n \to 0$ then by Lemma 3.3.3 $G(n)$ whp does not contain a rainbow K_4, so whp $\omega(G) \leq t(1, 2) \vee (\omega'(G) + 3)$. □

3.3.2 Monochromatic cliques and balls and bins

Here we prove Theorem 3.1.3. In the proof we use the fact that the maximum bin load $M(N, m)$ is a "smooth" function of the first argument N, see lemma below.

Lemma 3.3.5 *Let $\{N_n\}$ and $\{m_n\}$ be sequences of positive integers such that $N = N_n \to \infty$ and $m = m_n \to \infty$. Let $\{\delta_n\}, \{\epsilon_n\}$ be positive sequences converging to zero such that $\epsilon_n = o(\delta_n)$. For every n there is a coupling between random variables $M' = M'_n = M(\lfloor N(1+\epsilon_n)\rfloor, m)$ and $M = M_n = M(N, m)$ such that $M \le M'$ with probability one, and*

$$\mathbb{P}(M' - \delta_n \mathbb{E}\, M' \le M) \to 1. \tag{3.33}$$

If, additionally, $\mathbb{P}(M' > \delta_n^{-1}) \to 0$, then $M = M'$ whp.

Proof Given n, we label m bins by numbers $1, \dots, m$. Throw $\lfloor N(1+\epsilon_n)\rfloor$ balls into bins. This gives an instance of M'. Denote by L the label of the bin with the lowest index realising the maximum.

Now delete uniformly at random $\lfloor \epsilon_n N\rfloor$ balls. The configuration with the remaining N balls gives an instance of $M \le M'$. We remark that conditionally, given M', the number Δ of balls deleted from the bin L has a hypergeometric distribution with the mean value

$$\frac{M' \times \lfloor \epsilon_n N\rfloor}{\lfloor N(1+\epsilon_n)\rfloor} \le \epsilon_n M'.$$

Now the bin L contains $M' - \Delta \le M$ balls and, by Markov's inequality,

$$\mathbb{P}(M' - M \ge t) \le \mathbb{P}(\Delta \ge t) \le t^{-1}\mathbb{E}\,\Delta \le t^{-1}\epsilon_n \mathbb{E}\, M'.$$

Choosing $t = \delta_n \mathbb{E}\, M'$ yields (3.33). Similarly, if $\mathbb{P}(M' \ge \delta_n^{-1}) = o(1)$, then

$$\mathbb{P}(M' - M \ge 1) \le \mathbb{E}\,\Delta \mathbb{I}_{M' \le \delta_n^{-1}} + \mathbb{P}(M' > \delta_n^{-1}) \le \epsilon_n \delta_n^{-1} + o(1) \to 0.$$

□

Proof of Remark 3.1.4 Suppose $m = o(n)$, $\mathbb{E}\, Y = \Theta(1)$ and $\mathbb{E}\, Y^2 = O(1)$. Since $X = X(n)$ is a non-negative integer, we have $\mathbb{E}\, X^2 \ge \mathbb{E}\, X$. But $\mathbb{E}\, X^2 = O(m/n)$ and $\mathbb{E}\, X = \Theta((m/n)^{1/2})$, so $\mathbb{E}\, X^2 = o(\mathbb{E}\, X)$, a contradiction. □

Proof of Theorem 3.1.3 In view of Remark 3.1.4 it suffices to consider the case $m = \Omega(n)$. Denote $\epsilon_n = (2 + \ln^2 n)^{-1}$ so that $\epsilon_n \ln n = o(1)$ and $n\epsilon_n^2 \to +\infty$. Given n, write $\epsilon = \epsilon_n$ and denote $\bar{N} = n\mathbb{E}\, X_1 = \sqrt{mn}\,\mathbb{E}\, Y$ and

$$\bar{N}^- = \lfloor \bar{N}(1 - 4\epsilon)\rfloor, \qquad \bar{N}^+ = \lceil \bar{N}(1 + 4\epsilon)\rceil.$$

In order to generate an instance of $G(n)$ we draw a random sample $X_1, \dots, X_n$ from the distribution $P(n)$. Then choose random subsets $S_{v_i} \subseteq W$ of size X_i, $v_i \in V$, by throwing balls into m bins labelled $w_1, \dots, w_m$ (the j-th bin has label w_j and index j) as follows. Keep throwing balls labelled $i = 1$ until there are exactly X_i different bins containing a ball labelled i. Do the same for $i = 2, \dots, n$. Now, for each i, the bins containing balls labelled i make up the set S_{v_i}. In this way we obtain an instance of $G(n)$. Let X'_i denote the number of balls of label i thrown so far. Clearly, $X'_1, \dots, X'_n$ is a sequence of independent random variables and $X'_i \geq X_i$, for each i. We stop throwing balls if the number of balls $N' = \sum_i X'_i$ at least as large as $\bar{N}^+$. Otherwise we throw additional $\bar{N}^+ - N'$ unlabelled balls into bins.

Let us inspect the bins after j balls have been thrown. Let $\mathcal{M}(j)$ denote the set of balls contained in the bin with the largest number of balls and the smallest index. We note that the number $M(j) = |\mathcal{M}(j)|$ of balls in that bin has the same distribution as $M(j, m)$ (random variable defined before Theorem 3.1.3).

Denote, for short, $\omega' = \omega'(G(n))$ and $\bar{M} = M(\lfloor \bar{N} \rfloor)$. We observe that the event $\mathcal{A}_1 = \{\text{all balls of } \mathcal{M}(N') \text{ have different labels}\}$ implies $\omega'(G(n)) = M(N')$. Furthermore, if both events $\mathcal{A}_2 = \{M(\bar{N}^-) = M(\bar{N}^+)\}$ and $\mathcal{A}_3 = \{\bar{N}^- \leq N' \leq \bar{N}^+\}$ hold, then $\bar{M} = M(N')$. We shall show below that

$$\mathbb{P}(\mathcal{A}_r) = 1 - o(1), \qquad \text{for} \quad r = 1, 2, 3. \tag{3.34}$$

Now, (3.34) implies $\mathbb{P}(\omega' = \bar{M}) = 1 - o(1)$ and, since the distributions of $M(\lfloor \bar{N} \rfloor, m)$ and $\bar{M}$ coincide, we obtain

$$d_{TV}\big(\omega', M(\lfloor \bar{N} \rfloor, m)\big) = d_{TV}\big(\omega', \bar{M}\big) \leq \mathbb{P}\big(\omega' \neq \bar{M}\big) = o(1).$$

It remains to prove (3.34). Let us consider $\mathbb{P}(\mathcal{A}_3)$. We first replace X_i and X'_i by the truncated random variables

$$\tilde{X}_i = X_i \mathbb{I}_{\{X_i \leq \epsilon m\}} \qquad \text{and} \qquad \tilde{X}'_i = X'_i \mathbb{I}_{\{X_i \leq \epsilon m\}}, \quad 1 \leq i \leq n.$$

Denote $\tilde{N}' = \sum_i \tilde{X}'_i$ and introduce events $\tilde{\mathcal{A}}_3 = \{\bar{N}^- \leq \tilde{N}' \leq \bar{N}^+\}$ and $\mathcal{A}_4 = \{\max_{1 \leq i \leq n} X_i \leq \epsilon m\}$. Let $\bar{\mathcal{A}}_4$ denote the complement of $\mathcal{A}_4$. From the relation $\mathcal{A}_3 \cap \mathcal{A}_4 = \tilde{\mathcal{A}}_3 \cap \mathcal{A}_4$ we obtain

$$\mathbb{P}(\mathcal{A}_3) \geq \mathbb{P}(\mathcal{A}_3 \cap \mathcal{A}_4) = \mathbb{P}(\tilde{\mathcal{A}}_3 \cap \mathcal{A}_4) \geq \mathbb{P}(\tilde{\mathcal{A}}_3) - \mathbb{P}(\bar{\mathcal{A}}_4).$$

Furthermore, by the union bound and Markov's inequality, we have

$$\mathbb{P}(\bar{\mathcal{A}}_4) \le n\mathbb{P}(X_1 > \epsilon\, m) \le n\frac{\mathbb{E}\, X_1^2}{\epsilon^2 m^2} = \frac{\mathbb{E}\, Y^2}{\epsilon^2 m} = o(1),$$

since $m = \Omega(n)$ and $\epsilon^2 n \to +\infty$. Hence, $\mathbb{P}(\mathcal{A}_3) \ge \mathbb{P}(\tilde{\mathcal{A}}_3) - o(1)$. Secondly, we prove that $\mathbb{P}(\tilde{\mathcal{A}}_3) = 1 - o(1)$. For this purpose we show that, for large n,

$$\bar{N}(1-\epsilon) \le \mathbb{E}\, \tilde{N}' \le \bar{N}(1+2\epsilon) \quad \text{and} \quad \mathbb{P}\big(|\tilde{N}' - \mathbb{E}\, \tilde{N}'| \ge \epsilon\, \mathbb{E}\, \tilde{N}'\big) = o(1). \tag{3.35}$$

The proof of (3.35) is routine. Notice that conditionally, given $\tilde{X}_i = k$, we have $\tilde{X}_i' = \sum_{j=1}^{k} \xi_j$, where $\xi_1, \xi_2, \dots, \xi_k$ are independent geometric random variables with parameters

$$\frac{m}{m},\ \frac{m-1}{m},\ \dots,\ \frac{m-k+1}{m},$$

respectively. Since $\tilde{X}_i \le \epsilon\, m$, we only consider $k < \epsilon m$, so

$$\mathbb{E}\,(\tilde{X}_i'|\tilde{X}_i = k) = \frac{m}{m} + \frac{m}{m-1} + \dots + \frac{m}{m-k+1} \le \frac{k}{1-\epsilon} \le k(1+2\epsilon).$$

In the last step we used $\epsilon \le 1/2$. We conclude that

$$\tilde{X}_i \le \mathbb{E}\,(\tilde{X}_i'|\tilde{X}_i) \le \tilde{X}_i(1+2\epsilon). \tag{3.36}$$

From (3.36) we obtain

$$n\mathbb{E}\, \tilde{X}_1 \le \mathbb{E}\, \tilde{N}' \le (1+2\epsilon)n\mathbb{E}\, \tilde{X}_1. \tag{3.37}$$

Furthermore, invoking in (3.37) the inequalities $\mathbb{E}\, X_1 - s \le \mathbb{E}\, \tilde{X}_1 \le \mathbb{E}\, X_1$, where

$$s = \mathbb{E}\, X_1 \mathbb{I}_{\{X_1 > \epsilon m\}} \le (\epsilon\, m)^{-1} \mathbb{E}\, X_1^2 = (\epsilon\, n)^{-1} \mathbb{E}\, Y_1^2 = o(\epsilon),$$

we obtain the first part of (3.35). The second part of (3.35) follows from the inequalities $\tilde{N}' \ge N(1-\epsilon)$ and

$$Var \tilde{N}' \le 2n\mathbb{E}\, X_1^2 = 2m\mathbb{E}\, Y^2, \tag{3.38}$$

by Chebyshev's inequality. Let us show (3.38). Proceeding as in the proof of (3.36) we evaluate the conditional variance

$$Var(\tilde{X}_i'|\tilde{X}_i = k) = \sum_{j=1}^{k} Var(\xi_j) = \sum_{j=0}^{k-1} \frac{jm}{(m-j)^2} \le \frac{k^2}{2(1-\epsilon)^2 m} \le \frac{k^2}{m},$$

and obtain

$$\mathbb{E}\, Var(\tilde{X}_i'|\tilde{X}_i) \le \frac{\mathbb{E}\,\tilde{X}_i^2}{m}.$$

Furthermore, using (3.36) we write

$$Var(\mathbb{E}\,(\tilde{X}_i'|\tilde{X}_i)) \le \mathbb{E}\,(\mathbb{E}\,(\tilde{X}_i'|\tilde{X}_i))^2 \le \mathbb{E}\,\tilde{X}_i^2(1+2\epsilon)^2 \le \mathbb{E}\,\tilde{X}_i^2(1+8\epsilon).$$

Collecting these estimates we obtain an upper bound for the variance

$$Var(\tilde{X}_i') = \mathbb{E}\, Var(\tilde{X}_i'|\tilde{X}_i) + Var(\mathbb{E}\,(\tilde{X}_i'|\tilde{X}_i)) \le \mathbb{E}\,\tilde{X}_i^2(1+8\epsilon+m^{-1}) \le 2\mathbb{E}\,X_i^2.$$

This bound implies (3.38). We have shown (3.34) for $r=2$.

Let us prove (3.34) for $r=1$. We start with an auxiliary inequality. Given integers $x_1,\dots,x_n \ge 0$ consider a collection of $k = x_1+\cdots+x_n > 0$ labelled balls, containing x_i balls of label i, $1 \le i \le n$. The probability of the event that a random subset of r balls contains a pair of equally labelled balls is

$$\mathbb{P}(L \ge 1) \le \mathbb{E}\, L = \binom{r}{2}\binom{k}{2}^{-1}\sum_i \binom{x_i}{2} \le \left(\frac{r}{k}\right)^2 \sum_i x_i^2. \tag{3.39}$$

Here L counts pairs of equally labelled balls in the random subset.

We will show that $\mathbb{P}(\bar{\mathcal{A}}_1) = o(1)$. To this aim, we introduce events

$$\mathcal{A}_5 = \{M(\tilde{N}') \le \ln n\}, \qquad \mathcal{A}_6 = \{\sum_{1\le i\le n} (\tilde{X}_i')^2 \le m \ln n\},$$

estimate

$$\mathbb{P}(\bar{\mathcal{A}}_1) \le \mathbb{P}(\bar{\mathcal{A}}_1 \cap \mathcal{A}_3 \cap \mathcal{A}_4 \cap \mathcal{A}_5 \cap \mathcal{A}_6) + \mathbb{P}(\bar{\mathcal{A}}_3) + \mathbb{P}(\bar{\mathcal{A}}_4) + \mathbb{P}(\bar{\mathcal{A}}_5) + \mathbb{P}(\bar{\mathcal{A}}_6),$$

and show that each summand on the right is $o(1)$. For the first summand we estimate using (3.39)

$$\begin{aligned}
&\mathbb{P}(\bar{\mathcal{A}}_1 \cap \mathcal{A}_3 \cap \mathcal{A}_4 \cap \mathcal{A}_5 \cap \mathcal{A}_6) = \mathbb{E}\,\mathbb{P}(\bar{\mathcal{A}}_1|X_1,\dots,X_n)\mathbb{I}_{\mathcal{A}_3\cap\mathcal{A}_4\cap\mathcal{A}_5\cap\mathcal{A}_6} \\
&\le \mathbb{E}\left(\frac{M(\tilde{N}')^2}{(\tilde{N}')^2}\sum_i (\tilde{X}_i')^2 \mathbb{I}_{\mathcal{A}_3\cap\mathcal{A}_4\cap\mathcal{A}_5\cap\mathcal{A}_6}|X_1,\dots,X_n\right) \\
&\le \left(\frac{\ln n}{\bar{N}^+}\right)^2 m \ln n = O\left(\frac{\ln^3 n}{n}\right).
\end{aligned}$$

It remains to show $\mathbb{P}(\bar{\mathcal{A}}_r) = o(1)$, for $r = 5, 6$. We write $\mathbb{P}(\bar{\mathcal{A}}_5) = \mathbb{P}(\bar{\mathcal{A}}_5 \cap \mathcal{A}_3) + o(1)$ and estimate

$$\mathbb{P}(\bar{\mathcal{A}}_5 \cap \mathcal{A}_3) \leq \mathbb{P}(M(\bar{N}^+) > \ln n) = \mathbb{P}(\max_{j \in [m]} Z_j > \ln n) \leq m\mathbb{P}(Z_1 > \ln n) = o(1). \tag{3.40}$$

Here Z_j denotes the number of balls in the jth bin after $\bar{N}^+$ balls have been thrown. In the second inequality we applied the union bound and used the fact that $Z_1, \ldots, Z_m$ are identically distributed. To get the very last bound we write for binomially $Bin(\bar{N}^+, m^{-1})$ distributed Z_1 and $t = \lfloor \ln n \rfloor$,

$$\mathbb{P}(Z_1 \geq t) \leq \binom{\bar{N}^+}{t} m^{-t} \leq \left(\frac{e\bar{N}^+}{tm}\right)^t = o\left(m^{-1}\right).$$

To estimate $\mathbb{P}(\bar{\mathcal{A}}_6)$ we apply Markov's inequality,

$$\mathbb{P}(\bar{\mathcal{A}}_6) \leq (m \ln n)^{-1} n \mathbb{E}\,(\tilde{X}_1')^2 = \ln^{-1} n (Var(\tilde{X}_1') + (\mathbb{E}\,\tilde{X}_1')^2) = O(\ln^{-1} n).$$

Finally, we prove (3.34) for $r = 2$. Notice that the coupling between $M(\bar{N}^+)$ and $M(\bar{N}^-)$ is equivalent to the coupling provided by Lemma 3.3.5. Choose ϵ' solving $N^+ = (1 + \epsilon')N^-$ and note that $\epsilon' \sim 8\epsilon = O(\ln^{-2} n)$. The bound $\mathbb{P}(\mathcal{A}_2) = 1 - o(1)$ follows by Lemma 3.3.5 and the bound $\mathbb{P}(M(\bar{N}^+) > \ln n) = o(1)$, shown above. □

3.4 Algorithms for finding the largest clique

Random intersection graphs provide theoretical models for real networks, such as the affiliation (actor, scientific collaboration) networks. Although the model assumptions about the distribution of the family of random sets defining the intersection graph are rather stringent (independence and a particular form of the distribution), these models yield random graphs with clustering properties similar to those found in real networks, [18]. While observing a real network we may or may not have information about the sets of attributes prescribed to vertices. Therefore it is important to have algorithms suited to random intersection graphs that do not use any data related to attribute sets prescribed to vertices. In this section we consider two such algorithms that find cliques of order $(1 + o(1))\omega(G)$ in a random intersection graph G.

The Greedy-Clique algorithm of [55] finds a clique of the optimal order $(1 - o_P(1))\omega(G)$ in a random intersection graph, in the case where (asymptotic) degree distribution is a power-law with exponent $\alpha \in (1;2)$.

```
GREEDY-CLIQUE(G):
    Let v^(1), ..., v^(n) be V(G) sorted by their degrees, descending
    M ← ∅
    for i = 1 to n
        if v^(i) is adjacent to each vertex in M then
            M ← M ∪ {v^(i)}
    return M
```

Here we assume that graphs are represented by the adjacency list data structure. The implicit computational model behind our running time estimates in this section is random-access machine (RAM).

Proposition 3.4.1 *Assume that conditions of Theorem 3.1.1 hold. Suppose that* $\mathbb{E}Y = \Theta(1)$ *and that (3.7) holds for some* $\epsilon > 0$. *Then on input* $G = G(n)$ GREEDY-CLIQUE *outputs a clique of size* $\omega(G(n))(1 - o_P(1))$ *in time* $O(n^2)$.

By Lemma 3.1.5, the above result remains true if the conditions (3.2) and $\mathbb{E}\,Y(n) = \Theta(1)$ are replaced by the conditions (3.6) and $\mathbb{E}\,D_1 = \Theta(1)$. Proposition 3.4.1 is proved in a similar way as Lemma 3.2.1, but it does not follow from Lemma 3.2.1, since GREEDY-CLIQUE is not allowed to know the attribute subset sizes.

Proof The running time bound is obvious. We have to check that the algorithm returns a clique of the correct size. Fix any $\delta \in (0;1)$. Let $C = C(n)$ be the clique returned by the algorithm on input $G = G(n)$, and write $\omega = \omega(G(n))$. Let $A_\delta = A_\delta(n)$ be the event that $|C| < (1-\delta)\omega$. To prove the lemma, we have to show that $\mathbb{P}(A_\delta) \to 0$. Fix positive a, b, such that $a < 1/4$ and $b < (1 - 0.1\delta)^{-1/\alpha} - 1$. Let θ_2, e_1, K be as in Section 3.2.1. Set $\tilde{\theta} = \tilde{\theta}(n) = (1 + b)\theta_2$ and

$$\tau = \tau(n) = ((1 - \alpha/2)\mathbb{E}Y \ln n + e_1)n^{1/2}(1 + n^{-a}).$$

We will assume that n is large enough, so that $\lfloor \tilde{\theta} + 1 \rfloor \le m$. Let $V_2 = \{v \in [n] : X_v > \theta_2\}$ as before, and define random sets $Q = Q(n)$ and $R = R(n)$

$$Q = \{v \in [n] : D_v > \tau\} \quad \text{and} \quad R = \{v \in [n] : X_v > \tilde{\theta}\}.$$

Here D_v is the degree of v in $G(n)$. Let $B = B(n)$ be the event that $R \subseteq Q \subseteq V_2$. Write $\tilde{K} = \tilde{K}(n) = (1-\alpha/2)^{-\alpha}K$. Recall that in the proof of Lemma 3.2.1 it does not matter in which order the vertices of V_2 are considered when constructing the set A^N, in particular, the order may be random. We will assume that the vertices in $V_2 \cap Q$ are always considered first, in the order provided by the Greedy-Clique algorithm. Let L_θ be as in Lemma 3.2.1. We claim that

$$\mathbb{P}(A_\delta) \leq \mathbb{P}\left(|R| < (1-0.1\delta)\tilde{K}\right) + \mathbb{P}(\bar{B}) + \mathbb{P}\left(|V_2| > (1+0.1\delta)\tilde{K}\right) + \mathbb{P}\left(L_\theta > 0.1\delta|V_2|\right) + \mathbb{P}\left(\omega > (1+0.5\delta)\tilde{K}\right). \quad (3.41)$$

Indeed, suppose that $|R| \geq (1-0.1\delta)\tilde{K}$, $L_\theta \leq 0.1\delta|V_2|$, $|V_2| \leq (1+0.1\delta)\tilde{K}$, $\omega \leq (1+0.5\delta)\tilde{K}$ and B holds. Then A_δ does not hold, since

$$|C| \geq |Q| - L_\theta \geq |R| - L_\theta \geq (1-0.1\delta)\tilde{K} - 0.1\delta(1+0.1\delta)\tilde{K} \geq (1-0.5\delta)\tilde{K} \geq (1-\delta)\omega.$$

Now the last three terms on the right side of (3.41) tend to zero by the proof of Lemma 3.2.1 and Theorem 3.1.1. Furthermore, since $\mathbb{E}\,|R| = n\mathbb{P}(X > \tilde{\theta}) \sim (1+b)^{-\alpha}\tilde{K}$ and $(1+b)^{-\alpha} > 1-0.1\delta$, we get that $\mathbb{P}(|R| < (1-0.1\delta)\tilde{K}) \to 0$ by the concentration of $|R|$ (using, for example, (3.12)).

It remains to prove that B holds whp. Let us first show that $Q \subseteq V_2$ whp. By the union bound

$$\mathbb{P}(|Q \setminus V_2| > 0) \leq n\mathbb{P}(X_1 \leq \theta_2, D_1 > \tau) \leq n\mathbb{P}(D_1 > \tau | X_1 = \lfloor\theta_2\rfloor) \leq n\mathbb{P}\left(\sum_{v=2}^{n} \mathbb{I}_{S_v \cap S_1 \neq \emptyset} > \tau | X_1 = \lfloor\theta_2\rfloor\right) \leq n\mathbb{P}(Z_1 > \tau), \quad (3.42)$$

where $Z_1 = Z_1(n)$ is a random variable with distribution $Binom(n, p_1)$ and $p_1 = p_1(n) = \frac{\theta_2 \mathbb{E}\,X}{m}$. The last inequality follows by monotonicity, since the probability that S_2 intersects an independent uniformly random subset of $[m]$ of size $\lfloor\theta_2\rfloor$ is at most p_1. We have

$$\mathbb{E}\,Z_1 = np_1 = n^{1/2+o(1)} \quad \textit{and} \quad \tau = (1+n^{-a})\mathbb{E}\,Z_1.$$

Now by the Chernoff bound (3.11)

$$\mathbb{P}(Z_1 > \tau) = \mathbb{P}(Z_1 > (1+n^{-a})\mathbb{E}\,Z_1) \leq \exp\left(-n^{\frac{1}{2}-2a+o(1)}\right).$$

Putting this in (3.42) we get that $\mathbb{P}(|Q \setminus V_2| > 0) \to 0$.

Similarly we will show that $R \subseteq Q$ whp. We have

$$\mathbb{P}(|R \setminus Q| > 0) \leq n\mathbb{P}(D_1 \leq \tau, X_1 > \tilde{\theta}) \leq \mathbb{P}(D_1 \leq \tau | X_1 = \lfloor \tilde{\theta} + 1 \rfloor) = n\mathbb{P}(Z_2 \leq \tau),$$

where $Z_2 = Z_2(n) = Binom(n-1, p_2)$ and $p_2 = p_2(n)$ is the probability that S_2 intersects an independent uniformly random subset of $[m]$ of size $\lfloor \tilde{\theta} + 1 \rfloor$. Now (3.7), $\mathbb{E}\, Y = \Theta(1)$ and Lemma 3.2.5 imply that for $\tilde{a} = \tilde{a}(n) = n^{-\epsilon/2}$ we have

$$\mathbb{E}\, \tilde{\theta} X_2 \mathbb{I}_{\tilde{\theta} X_2 > \tilde{a} m} = \left(\frac{m}{n}\right)^{1/2} \tilde{\theta} \mathbb{E}\, Y \mathbb{I}_{Y > \tilde{a} n^{1/2}((1-\alpha/2)\ln n + e_1)^{-1}} = o(\tilde{\theta} \mathbb{E}\, X). \tag{3.43}$$

Next, observe that the bounds (3.55), (3.56) and (3.57) apply also when X_1 and X_2 are independent but with different distributions, in particular when $X_1 = \lfloor \tilde{\theta} + 1 \rfloor$ and X_2 has distribution $P(n)$; these inequalities together with (3.43) yield that

$$p_2 \sim \frac{\tilde{\theta} \mathbb{E}\, X}{m} \quad \text{and} \quad \mathbb{E}\, Z_2 \sim np_2 \sim (1+b)\tau.$$

Applying (3.11) again, we get that $\mathbb{P}(Z_2 \leq \tau) = \exp\left(-n^{1/2+o(1)}\right)$ and $\mathbb{P}(|R \setminus Q| > 0) \to 0$. □

For random intersection graphs with square integrable degree distribution we suggest the following simple algorithm.

```
MONO-CLIQUE(G):
    for uv ∈ E(G)
        D(uv) ← |Γ(u) ∩ Γ(v)|
    for uv ∈ E(G) in the decreasing order of D(uv)
        S ← Γ(u) ∩ Γ(v)
        if S is a clique then
            return S ∪ {u, v}
    return {1} ∩ V(G)
```

Here $\Gamma(v)$ denotes the set of neighbours of v.

Theorem 3.4.2 *Assume that $\{G(n)\}$ is a sequence of random intersection graphs such that $n = O(m)$ and $\mathbb{E}\, Y^2(n) = O(1)$. Let $C = C(n)$ be the clique constructed by* MONO-CLIQUE *on input $G(n)$. Then $\mathbb{E}\, (\omega(G(n)) - |C|)^2 = O(1)$. Furthermore, if there is a sequence $\{\omega_n\}$, such that $\omega_n \to \infty$ and $\omega(G(n)) \geq \omega_n$ whp, then $|C| = \omega(G(n))$ whp.*

Proof Given distinct vertices $v_1, v_2, v_3, v_4 \in [n]$, let $\mathcal{C}(v_1, v_2, v_3, v_4)$ be the event that $G(n)$ contains a cycle with edges $\{v_1v_2, v_2v_3, v_3v_4, v_1v_4\}$ and $S_{v_2} \cap S_{v_4} = \emptyset$. Let Z denote the number of tuples (v_1, v_2, v_3, v_4) of distinct vertices in $[n]$ such that $\mathcal{C}(v_1, v_2, v_3, v_4)$ hold. We will show below that

$$\mathbb{E}\, Z = O(1). \tag{3.44}$$

Let $S \subseteq [n]$ be the (lexicographically first) largest clique of $G(n)$. Denote $s = |S|$. If $s \leq 2$ or there is a pair $\{x, y\} \subseteq S$, $x \neq y$ such that $G(n)[\Gamma(x) \cap \Gamma(y)]$ is a clique, then the algorithm returns a clique of size s. Otherwise, for each such pair $\{x, y\}$ there are $x', y' \in \Gamma(x) \cap \Gamma(y)$, $x' \neq y'$ with $x'y' \notin E(G(n))$. That is, $\mathcal{C}(x, x', y, y')$ holds and $\binom{s}{2} \leq Z$. Thus, if $\binom{s}{2} > Z$, the algorithm returns a clique C of size s. Otherwise, the algorithm may fail and return a clique C of size 1. In any case we have that

$$s - |C| \leq \sqrt{2Z} + 1$$

and using (3.44)

$$\mathbb{E}\,(\omega(G(n)) - |C|)^2 \leq \mathbb{E}\,(\sqrt{2Z} + 1)^2 = O(1).$$

Also if $\omega(G(n)) \geq \omega_n$ whp, then by (3.44) and Markov's inequality

$$\mathbb{P}(|C| \neq \omega(G(n))) \leq \mathbb{P}(\omega(G(n)) < \omega_n) + \mathbb{P}\left(Z \geq \binom{\omega_n}{2}\right) \to 0.$$

It remains to show (3.44). What is the probability of the event $\mathcal{C}(1, 2, 3, 4)$? Clearly, $\mathcal{C}(1, 2, 3, 4)$ implies at least one of the following events:

- $\mathcal{A}_1$: there are distinct attributes $w_1, w_2, w_3, w_4 \in W$ such that $w_1 \in S_1 \cap S_2$, $w_2 \in S_2 \cap S_3$, $w_3 \in S_3 \cap S_4$ and $w_4 \in S_1 \cap S_4$;
- $\mathcal{A}_2$: there are distinct $w_1, w_2, w_3 \in W$, such that $w_1 \in S_1 \cap S_2 \cap S_3$, $w_2 \in S_3 \cap S_4$ and $w_3 \in S_1 \cap S_4$;
- $\mathcal{A}_3$: there are distinct $w_1, w_2, w_3 \in W$, such that $w_1 \in S_1 \cap S_2$, $w_2 \in S_2 \cap S_3$ and $w_3 \in S_1 \cap S_3 \cap S_4$;
- $\mathcal{A}_4$: there are distinct $w_1, w_2 \in W$, such that $w_1 \in S_1 \cap S_2 \cap S_3$ and $w_2 \in S_1 \cap S_3 \cap S_4$.

Conditioning on X_1, X_2, X_3, X_4 and using the union bound and independence we obtain, similarly as in Lemma 3.3.2

$$\begin{aligned}
\mathbb{P}(\mathcal{A}_1) &\le (m)_4 \mathbb{E}\,\frac{(X_1)_2(X_2)_2(X_3)_2(X_4)_2}{(m)_2^4} \le \frac{(\mathbb{E}\,Y^2)^4}{n^4};\\
\mathbb{P}(\mathcal{A}_2) = \mathbb{P}(\mathcal{A}_3) &\le (m)_3 \mathbb{E}\,\frac{(X_1)_2 X_2 (X_3)_2(X_4)_2}{(m)_2^3 m} \le \frac{(\mathbb{E}\,Y^2)^3(\mathbb{E}\,Y)}{m^{0.5}n^{3.5}};\\
\mathbb{P}(\mathcal{A}_4) &\le (m)_2 \mathbb{E}\,\frac{(X_1)_2 X_2 (X_3)_2 X_4}{(m)_2^2 m^2} \le \frac{(\mathbb{E}\,Y^2)^2(\mathbb{E}\,Y)^2}{mn^3}.
\end{aligned}$$

Furthermore, by symmetry,

$$\mathbb{E}\,X \le (n)_4\left(\mathbb{P}(\mathcal{A}_1) + \mathbb{P}(\mathcal{A}_2) + \mathbb{P}(\mathcal{A}_3) + \mathbb{P}(\mathcal{A}_4)\right) = O(1).$$

□

Proposition 3.4.3 *Consider a sequence of random intersection graphs $\{G(n)\}$ as in Theorem 3.1.3.* MONO-CLIQUE *can be implemented so that its expected running time on $G(n)$ is $O(n)$.*

Proof Let $\tilde{Z}$ denote the number of 4-cycles in $G(n)$, i.e., the number of tuples (v_1, v_2, v_3, v_4) of distinct vertices in $[n]$, such that $v_1v_2, v_2v_3, v_3v_4, v_1v_4 \in E(G(n))$. We will prove below that

$$\mathbb{E}\,\tilde{Z} = O(n). \tag{3.45}$$

Consider the running time of the first loop. We can assume that the elements in each list in the adjacency list structure are sorted in increasing order (recall that vertices are elements of $V = [n]$). Otherwise, given $G(n)$, they can be sorted using any standard sorting algorithm in time $O(n + \sum_{v\in[n]} D_v^2)$, where $D_v = d_{G(n)}(v)$ is the degree of v in $G(n)$. The intersection of two lists of lengths k_1 and k_2 can be found in $O(k_1 + k_2)$ time, so that expected total time for finding common neighbours is

$$O\left(n + \mathbb{E}\sum_{uv\in E(G(n))}(D_u + D_v)\right) = O\left(n + \mathbb{E}\sum_{v\in[n]} D_v^2\right) = O(n).$$

The last estimate follows by (3.66) in the proof of Lemma 3.1.6.

The second loop can be implemented so that the next edge uv with largest value of $D(uv)$ is found at each iteration (i.e., we do not sort the list of edges in advance). In this way picking the next edge requires at most $ce(G(n))$ steps c is a universal constant. We recall that the number of edges $uv \in E(G)$ with

$\Gamma(u,v) := \Gamma(u) \cap \Gamma(v) \neq \emptyset$ that fail to induce a clique is at most the number Z of cycles considered in the proof of Theorem 3.4.2 above. Therefore, the total number of steps used in picking $D(uv)$ in decreasing order is at most

$$Z\, e(G(n)) = \sum_{(i,j,k,l)} \mathbb{I}_{\mathcal{C}(i,j,k,l)} e(G(n)).$$

Now

$$e(G(n) = \sum_{s<t:\, \{s,t\}\cap\{i,j,k,l\}=\emptyset} \mathbb{I}_{\{s\sim t\}} + \sum_{s<t:\, \{s,t\}\cap\{i,j,k,l\}\neq\emptyset} \mathbb{I}_{\{s\sim t\}}.$$

Note, that the second sum on the right is at most $4n$. Also, if $\{s,t\}\cap\{i,j,k,l\} = \emptyset$, the events $s \sim t$ and $\mathcal{C}(i,j,k,l)$ are independent, therefore

$$\begin{aligned}\mathbb{E}\left(\mathbb{I}_{\mathcal{C}(i,j,k,l)} \sum_{s<t:\, \{s,t\}\cap\{i,j,k,l\}=\emptyset} \mathbb{I}_{\{s\sim t\}}\right) &= \mathbb{P}(\mathcal{C}(i,j,k,l)) \sum_{s<t:\, \{s,t\}\cap\{i,j,k,l\}=\emptyset} \mathbb{P}(s\sim t) \\ &\leq \mathbb{P}(\mathcal{C}(i,j,k,l))\mathbb{E}\, e(G(n)).\end{aligned}$$

Finally, invoking the simple bound $\mathbb{E}\, e(G(n)) = \binom{n}{2}\mathbb{P}(u \sim v) = O(n)$, and (3.44) we get

$$\mathbb{E}\, Z\, e(G(n)) \leq (\mathbb{E}\, e(G(n)) + 4n) \sum_{(i,j,k,l)} \mathbb{P}(\mathcal{C}(i,j,k,l)) = (\mathbb{E}\, e(G(n)) + 4n)\mathbb{E}\, Z = O(n).$$

Now let us estimate the time of the rest of the iteration of the second loop. The total expected time to find common neighbours is again $O(n)$, so we only consider the time spent for checking if $\Gamma(u,v)$ is a clique. This requires $c\, s_{uv}^2$ steps, where we denote $s_{uv} = |\Gamma(u,v)|$. Observe that u, v and $\Gamma(u,v)$ yield at least $s_{uv}(s_{uv}-1)$ 4-cycles in $G(n)$ of the form (u,x,v,y), $x, y \in \Gamma(u,v)$. Summing over all edges uv and noticing that each 4-tuple corresponding to 4-cycle in $G(n)$ can be obtained at most once, we get

$$\tilde{Z} \geq \sum_{uv\in E(G(n))} s_{uv}(s_{uv}-1) \geq \sum_{uv\in E(G(n))} (s_{uv}^2 - 1)/2.$$

So using (3.45) and the fact that $\mathbb{E}\, e(G(n)) = O(n)$ we obtain

$$\mathbb{E} \sum_{uv\in E(G(n))} s_{uv}^2 \leq 2\mathbb{E}\, \tilde{Z} + \mathbb{E}\, e(G(n)) = O(n).$$

Finally, let us bound $\mathbb{E}\, \tilde{Z}$. Let $\mathcal{A}_i$, $1 \leq i \leq 4$ be as in the proof of Theorem 3.4.2.

Let $\mathcal{A}_5$ be the event that there is $w \in W$ such that $w \in S_1 \cap S_2 \cap S_3 \cap S_4$. Using the union bound

$$\mathbb{P}(\mathcal{A}_5) \leq m\mathbb{E}\,\frac{X_1X_2X_3X_4}{m^4} = \frac{(\mathbb{E}\,Y)^4}{mn^2}.$$

Similarly as in the proof of Theorem 3.4.2 (we have to consider three other events similar to A_2 and A_4),

$$\mathbb{E}\,\tilde{Z} \leq (n)_4\left(\mathbb{P}(A_1) + 4\mathbb{P}(A_2) + 2\mathbb{P}(A_4) + \mathbb{P}(A_5)\right) = O(n).$$

□

Combining the next lemma with Theorem 3.1.3 we can show that MONO-CLIQUE whp finds a clique of size at least $\omega'(G(n))$.

Lemma 3.4.4 *Let $\{G(n)\}$ be as in Theorem 3.1.3 and let $M = M(G(n))$ be the monochromatic clique of size $\omega'(G(n))$ generated by the attribute with the smallest index. Then whp $G(n)$ has an edge uv such that $\{u, v\} \cup (\Gamma(u) \cap \Gamma(v)) = M$.*

Before we prove the lemma, we need several definitions. A *balls and bins configuration* $\mathcal{C}$ with parameters (N, m, n) is a pair $(bin_{\mathcal{C}}, label_{\mathcal{C}})$ of functions $bin_{\mathcal{C}} : [N] \to [m]$ and $label_{\mathcal{C}} : [N] \to [n]$. These functions determine the placement of N distinguishable balls with labels from $[n]$ into m bins. Recall that $f(S)$ and $f^{-1}(S)$ denote the image and preimage of a set S under a function f respectively. Given a balls and bins configuration $\mathcal{C}$, for each $i \in [n]$ we define $S_i(\mathcal{C}) = bin_{\mathcal{C}}(label_{\mathcal{C}}^{-1}(\{i\}))$, $X_i(\mathcal{C}) = |S_i(\mathcal{C})|$ and $X'_i(\mathcal{C}) = |label_{\mathcal{C}}^{-1}(\{i\})|$. Also define for $j \in [m]$ the set $T_j(\mathcal{C}) = bin_{\mathcal{C}}^{-1}(\{j\})$, write $Z_j(\mathcal{C}) = |T_j(\mathcal{C})|$ and interpret $T_j(\mathcal{C})$ as the set of balls in the bin j. We denote by $L(\mathcal{C})$ the (smallest) index $j \in [m]$ that maximizes $Z_j(\mathcal{C})$. Also, write $\omega'(\mathcal{C}) = Z_L(\mathcal{C})$, where $L = L(\mathcal{C})$.

The intersection graph $G(\mathcal{C})$ corresponding to $\mathcal{C}$ is the intersection graph on vertex set $[n]$ of the family of sets $\{S_i(\mathcal{C}) : i \in [n]\}$.

Given a function $f : [N] \to [m]$ and $x, y \in [N]$ let *exchange*(f, x, y) be an operation, the result of which is again a function $g : [N] \to [m]$, defined as follows.

$$g(w) = \begin{cases} f(x), & \text{if } w = y \\ f(y), & \text{if } w = x; \\ f(w), & \text{otherwise.} \end{cases}$$

Given a balls and bins configuration $\mathcal{C}$ with parameters (N, m, n), define random variables b_1, b_2, b'_1, b'_2 taking values[1] in the set

$$\{undefined\} \cup [N]$$

as follows. If $\omega'(\mathcal{C}) \leq 1$ let $b_1 = b_2 = undefined$. Otherwise, let (b_1, b_2) be chosen uniformly at random from all ordered pairs of distinct balls in the bin $L(\mathcal{C})$. We let $b'_1 = b'_2 = undefined$ if $N \leq 1$. Otherwise, we define (b'_1, b'_2) as a uniformly random pair from all $N(N-1)$ ordered pairs of distinct balls in $[N]$.

Now if $\omega'(\mathcal{C}) \leq 1$, define $bin_{\mathcal{C}'} = bin_{\mathcal{C}}$, otherwise define

$$bin_{\mathcal{C}'} = exchange(exchange(bin, b_1, b'_1), b_2, b'_2),$$

that is, we exchange the bins of b_1 and b'_1, then exchange the bins of b_2 and b'_2. Finally, let $T(\mathcal{C}) = T(\mathcal{C}, b_1, b_2, b'_1, b'_2)$ be a balls and bins configuration with parameters (N, m, n) and functions $(bin_{\mathcal{C}'}, label_{\mathcal{C}})$.

Lemma 3.4.5 *Let n, m be positive integers, let $(x_1, \ldots, x_n)$ be a sequence of nonnegative integers and let $N = x_1 + \cdots + x_n$. Let f be an arbitrary function $[N] \to [n]$, such that $|f^{-1}(\{j\})| = x_j$ for each $j \in [n]$. Let $\mathcal{C}$ be a random balls and bins configuration with parameters (N, m, n) defined as follows:*

- *$bin_{\mathcal{C}}$ is chosen uniformly at random from all m^N functions from $[N]$ to $[m]$ (i.e., each ball is thrown into a uniformly random bin);*
- *$label_{\mathcal{C}} = f$.*

Then $T(\mathcal{C})$ has the same distribution as $\mathcal{C}$.

Proof Let $z = (z_1, \ldots, z_m)$ be a vector of non-negative integers, such that $\sum_{i=1}^m z_i = \sum_{j=1}^n x_j = N$ and write $z_{max} = \max_{i \in [n]} z_i$. For a balls and bins configuration $\mathcal{C}'$, write $Z(\mathcal{C}') = (Z_1(\mathcal{C}'), \ldots, Z_m(\mathcal{C}'))$

Define a Markov chain M_z with state space $\mathcal{S}_z$ consisting of all balls and bins configurations $\mathcal{C}'$ with parameters (N, m, n), such that $label_{\mathcal{C}'} = f$ and $Z(\mathcal{C}') = z$, and with transitions given by the operation T.

Suppose $z_{max} \geq 2$. We claim, that the transition probabilities p_{st} of M_z satisfy

$$\sum_{s \in \mathcal{S}_z} p_{ts} = \sum_{s \in \mathcal{S}_z} p_{st} \text{ for any state } t \in \mathcal{S}_z. \tag{3.46}$$

(Note that the above sum must be equal to 1). This follows by a standard argument: let $H = H(V_1, V_2)$ be an edge-weighted bipartite graph, where $V_1 = \mathcal{S}_z$,

[1] We use a special value "undefined" to avoid the need of extra conditioning later.

$V_2 = \{s' : s \in \mathcal{S}_z\}$ is a disjoint copy of V_1 and for each $st' \in V_1 \times V_2$, let the weight $w(st')$ be the number of quadruples (b_1, b_2, b_1', b_2') leading from s to t. By symmetry, for any $s \in \mathcal{S}$,

$$\sum_{t \in V_2} w(st') = \sum_{t \in V_1} w(ts') = \frac{1}{|\mathcal{S}_z|} \sum_{s,t} w(st').$$

For any state in $\mathcal{S}_z$, any valid quadruple has the same probability $p = (z_{max}(z_{max} - 1)N_0(N_0 - 1))^{-1}$ to be picked by the random transition T. Therefore (3.46) follows by multiplying both sides of the last equality by p. (By considering "inverse" transitions one may further show that the transition matrix of this Markov chain is actually symmetric.)

Now by (3.46), the uniform distribution over $\mathcal{S}_z$ is a stationary distribution for M_z. Clearly, conditionally on $Z(\mathcal{C}) = z$, $\mathcal{C}$ is uniformly distributed over $\mathcal{S}_z$, and so $T(\mathcal{C})$ is also distributed uniformly over $\mathcal{S}_z$. The last statement is also true in the case where $z_{max} < 2$, since in this case T is the identity operation. We conclude that $T(\mathcal{C})$ has the same distribution as $\mathcal{C}$. □

Proof of Lemma 3.4.4 Using the iid random variables $X_1, \dots, X_n$ (distributed according to $P(n)$), we will construct two random balls and bins configurations $\mathcal{C}'$ and $\mathcal{C}''$, such that

(i) $\mathcal{C}'$ and $\mathcal{C}''$ have the same distribution;

(ii) $G(\mathcal{C}')$ and $G(\mathcal{C}'')$ are distributed as $G(n)$.

At the same time we will construct a pair of random vertices (u^*, v^*), such that

- whp $u^*, v^* \in T_L(\mathcal{C}'')$, where $L = L(\mathcal{C}'')$;
- $u^*, v^* \in \{0\} \cup [n]$;
- Given $X_1, \dots, X_n$, such that $N = X_1 + \cdots + X_n$, u^*, v^* are iid, and for any $x \in [n]$ $\mathbb{P}(u^* = x) = \frac{X_i}{N+1}$;
- Given $X_1, \dots, X_n$, u^*, v^* are conditionally independent of $\mathcal{C}'$.

Finally we will show that whp $\mathcal{C}'$ and $\mathcal{C}''$ differ only by the positions of at most four balls, so we will be able to use the simple distribution of (u^*, v^*) in $G(\mathcal{C}')$ to finish the proof of the lemma.

Construction of $\mathcal{C}'$ and $\mathcal{C}''$. We proceed in several steps. We start by drawing the sequence $X_1, \dots, X_n$. Define a random balls and bins configuration $\mathcal{C}_0$ with parameters (N, m, n), $N = \sum_{i=1}^{n} X_i$ and X_i balls labelled i as in Lemma 3.4.5. For $b \in [N]$ let $label_{\mathcal{C}_0}(b) = \min\{i \in [n] : X_1 + \cdots + X_i \geq b\}$ and notice that $label_{\mathcal{C}_0}$ depends only on $X_1, \ldots, X_n$. Now let $\mathcal{C}'_0 = \mathcal{C}_0$ and let $\mathcal{C}''_0 = T(\mathcal{C}_0)$. Let $b_1, b_2, b'_1, b'_2 \in \{undefined\} \cup [N]$ be the corresponding random balls used in the operation T as in Lemma 3.4.5.

In the *second phase* we complete the construction of $\mathcal{C}'$ and $\mathcal{C}''$ by adding more balls to $\mathcal{C}'_0$ and $\mathcal{C}''_0$ respectively so that $X_i(\mathcal{C}') = X_i(\mathcal{C}'') = X_i$ for each $i \in [n]$. We use pairing as much as we can. Specifically, assume that $\mathcal{C}'_t$ and $\mathcal{C}''_t$ are already defined. If there is no index i, such that $\max(X_i(\mathcal{C}'_t), X_i(\mathcal{C}''_t)) < X_i$, we stop. Otherwise, let i be smallest such index. Pick b^{**}_{t+1} independently and uniformly at random from $[m]$ and place $N + t + 1$ to the bin b^{**}_{t+1} in both $\mathcal{C}'_t$ and $\mathcal{C}''_t$, to obtain $\mathcal{C}'_{t+1}$ and $\mathcal{C}''_{t+1}$ respectively with parameters $(N + t + 1, m, n)$ (formally, $bin_{\mathcal{C}'_{t+1}}(N + t + 1) = b^{**}_{t+1}$, $label_{\mathcal{C}'_{t+1}}(N + t + 1) = i$, and $bin_{\mathcal{C}'_{t+1}}, label_{\mathcal{C}'_{t+1}}$ agree on $[N + t]$; $\mathcal{C}''_{t+1}$ is obtained from $\mathcal{C}''_t$ similarly).

Let t be the largest integer, such that $\mathcal{C}'_t$ and $\mathcal{C}''_t$ are defined. For $s = t, t+1, \ldots,$ while there is an index i such that $X_i(\mathcal{C}'_s) < X_i$ add a ball $N + s + 1$ labelled i to a uniformly random bin in $\mathcal{C}'_s$ to obtain $\mathcal{C}'_{s+1}$. Call the final configuration $\mathcal{C}'$. Similarly, but now independently, while there is an index i with $X_i(\mathcal{C}''_s) < X_i$, add a ball $N + s + 1$ labelled i to a uniformly random bin from $\mathcal{C}''_s$ to obtain $\mathcal{C}''_{s+1}$, and call the final configuration $\mathcal{C}''$.

Proof of (i) and (ii). By Lemma 3.4.5, $\mathcal{C}'_0$ and $\mathcal{C}''_0$ are identically distributed. The procedure to obtain $\mathcal{C}'$ from $\mathcal{C}'_0$ is the same as the procedure to obtain $\mathcal{C}''$ from $\mathcal{C}''_0$: we may ignore the coupling of the second phase since the distribution for each of the final configuration does not depend on the sequence of labels used in this phase. Therefore $\mathcal{C}'$ and $\mathcal{C}''$ are also identically distributed. Using a similar argument as in the proof of Theorem 3.1.3, we see that $G(\mathcal{C}')$ has the same distribution as $G(n)$.

Properties of $\mathcal{C}'$ and $\mathcal{C}''$. Let N' and N'' be the number of balls in $\mathcal{C}'$ and $\mathcal{C}''$ respectively. For $i \in \{1, 2\}$ let $l'_i = label_{\mathcal{C}_0}(b'_i)$ in the case where $b'_i \in [N]$, and let $l'_i = 0$ in the case where $b'_i = undefined$. Similarly let $l_i = label_{\mathcal{C}_0}(b_i)$ in the case where $b_i \in [N]$ and let $l_i = 0$ otherwise. Finally, define $w'_i = bin_{\mathcal{C}_0}(b'_i)$ if $b'_i \neq undefined$ and $w'_i = 0$ otherwise. Also define $L := L(\mathcal{C}'')$.

We will prove that all of the following events hold whp

- A : $N' = N''$, $label_{\mathcal{C}'} = label_{\mathcal{C}''}$, the functions $bin_{\mathcal{C}'}$, $bin_{\mathcal{C}''}$ agree on $[N'] \setminus \{b_1, b_2, b'_1, b'_2\}$ and for $i \in \{1, 2\}$ $bin_{\mathcal{C}'}(b'_i) = bin_{\mathcal{C}''}(b_i)$, $bin_{\mathcal{C}'}(b_i) = bin_{\mathcal{C}''}(b'_i)$;

- A' : for each $k \in [n] \setminus \{l'_1, l'_2\}$ there is $i \in \{1,2\}$ such that $S_k(\mathcal{C}') \cap S_{l'_i}(\mathcal{C}') = \emptyset$;
- B : for $i \in \{1,2\}$, $X_{l'_i}(\mathcal{C}_0) = X_{l'_i}$;
- C : b_1, b_2, b'_1, b'_2 are distinct and belong to $[N]$;
- D : $w'_1 \notin S_{\mathcal{C}_0}(l'_2)$ and $w'_2 \notin S_{\mathcal{C}_0}(l'_1)$;
- E : $L = L(\mathcal{C}_0)$, $T_L(\mathcal{C}''_0) = T_L(\mathcal{C}'')$ and all balls in $T_L(\mathcal{C}'')$ have distinct labels; also all balls in $T_L(\mathcal{C}')$ have distinct labels;
- F_1 : $L(\mathcal{C}_0) \notin S_{l'_1}(\mathcal{C}_0) \cup S_{l'_2}(\mathcal{C}_0)$;
- F_2 : for $j \in \{1,2\}$, b_j is the unique ball in $T_{w'_j}(\mathcal{C}'')$ with $label_{\mathcal{C}''}(b_j) = l_j$.

Assume that B, C, D, E, F_1 and F_2 occur, we will show that A occurs (this event is not essential but helps to make the proof clearer). Since B and F_1 hold, for $i \in \{1,2\}$, each ball in the set $label^{-1}_{\mathcal{C}_0}(\{l'_i\})$ is in a distinct bin of $\mathcal{C}'_0 = \mathcal{C}_0$, and not in the bin $L(\mathcal{C}_0)$. Using also C, each ball with label l'_i is in a distinct bin of $\mathcal{C}''_0$. Therefore no new balls with label l'_i are added in the random construction (of both $\mathcal{C}'$ and $\mathcal{C}''$), once $\mathcal{C}'_0$ and $\mathcal{C}''_0$ are defined and

$$S_{l'_i}(\mathcal{C}') \setminus \{w'_i\} = S_{l'_i}(\mathcal{C}'') \setminus \{L\}. \tag{3.47}$$

Now let $j \in \{1,2\}$. C, E, F_1 and F_2 imply that $L \in S_{l_j}(\mathcal{C}'_0)$ and $w'_j \notin S_{l_j}(\mathcal{C}'_0)$; meanwhile $L \notin S_{l_j}(\mathcal{C}''_0)$ and $w'_j \in S_{l_j}(\mathcal{C}''_0)$. By E, no new ball labelled l_j is added to the bin L in the second phase of the construction. By F_2, no new ball labelled l_j is added to the bin w'_j in the second phase. This means that

$$S_{l_j}(\mathcal{C}') \setminus \{L\} = S_{l_j}(\mathcal{C}'') \setminus \{w'_j\}. \tag{3.48}$$

Also, notice that the operation T does not modify the balls with labels $k \in [n] \setminus \{l_1, l_2, l'_1, l'_2\}$ therefore by the pairing $S_k(\mathcal{C}') = S_k(\mathcal{C}'')$ and A holds.

Now let A'' be the event that l'_1 and l'_2 belong to the set $M = label(T_L(\mathcal{C}''))$, $l'_1 \neq l'_2$, $|M| = \omega'(G(\mathcal{C}''))$ and $\Gamma_{G(\mathcal{C}'')}(l'_1) \cap \Gamma_{G(\mathcal{C}'')}(l'_2) = M \setminus \{l'_1, l'_2\}$. To prove the lemma, it suffices to show that A'' holds whp.

Suppose A', B, C, D, E, F_1, F_2 hold, but A'' does not. Then there is $k \in [n] \setminus \{l'_1, l'_2\}$, such that for both $i \in \{1,2\}$, $(S_{l'_i}(\mathcal{C}'') \setminus \{L\}) \cap S_k(\mathcal{C}'') \neq \emptyset$. Suppose $k \notin \{l_1, l_2\}$. We have $S_k(\mathcal{C}') = S_k(\mathcal{C}'')$, so if $S_k(\mathcal{C}'')$ intersects $S_{l'_i}(\mathcal{C}'') \setminus \{L\}$ then $S_k(\mathcal{C}')$ intersects $S_{l'_i}(\mathcal{C}')$ and A' does not hold, a contradiction. Therefore we can assume that $k \in \{l_1, l_2\}$.

Since $w_1', w_2' \notin S_{l_i'}(\mathcal{C}'')$ (by B, b_i' is the unique ball b in $T_{w_i'}(\mathcal{C}_0)$ with $label_{\mathcal{C}_0}(b) = l_1'$, and so by C, D and F_1, $w_1', w_2' \notin S_{l_i'}(\mathcal{C}_0'') = S_{l_i'}(\mathcal{C}''))$ we have that for $i \in \{1, 2\}$,

$$(S_{l_i'}(\mathcal{C}'') \setminus \{L, w_1', w_2'\}) \cap (S_k(\mathcal{C}'') \setminus \{L, w_1', w_2'\}) \neq \emptyset.$$

But using (3.47) and (3.48), $(S_x(\mathcal{C}') \setminus \{L, w_1', w_2'\}) = (S_x(\mathcal{C}'') \setminus \{L, w_1', w_2'\})$ for $x \in \{l_1, l_2, l_1', l_2'\}$, therefore A' does not occur. This is a contradiction. We conclude that indeed the events $A', \ldots, F_2$ imply A''.

To complete the proof of the lemma, it suffices to show that A', B, C, D, E, F_1, F_2 all occur whp.

The event B. For a random variable Y and an event A we will write $\mathbb{E}_* Y = \mathbb{E}(Y|X_1, \ldots, X_n)$ and $\mathbb{P}_*(A) = \mathbb{E}_* \mathbb{I}_A$.

Let $i \in \{1, 2\}$. We have $X_i(\mathcal{C}_0) = X_i$ if and only if each ball with label l_i' falls into a distinct bin of $\mathcal{C}_e$. Given $X_1, \ldots, X_n$ and l_i', the probability of this is

$$\frac{m-1}{m} \times \cdots \times \frac{m - X_{l_i'} + 1}{m} \geq 1 - \frac{1 + \cdots + (X_{l_i'} - 1)}{m} \geq 1 - \frac{X_{l_i'}^2}{2m}.$$

Therefore

$$\begin{aligned}
\mathbb{P}(\bar{B}) &= \mathbb{E}\,\mathbb{E}\left(\mathbb{I}_{\bar{B}}|X_1, \ldots, X_n, l_i'\right) \leq \mathbb{E}\,\mathbb{E}_*\left(1 \wedge \frac{X_{l_i'}^2}{2m}\right) \\
&\leq \mathbb{E}\left(\sum_{k=1}^{n}\left(1 \wedge \frac{X_k^2}{2m}\right)\frac{2X_k}{\mathbb{E}\,N}\right) + \mathbb{P}\left(N < \frac{\mathbb{E}\,N}{2}\right). \\
&= \frac{2n}{m^{1/2} n^{1/2} \mathbb{E}\,Y} \mathbb{E}\left(1 \wedge \frac{X_1^2}{2m}\right) X_1 + \mathbb{P}\left(N < \frac{\mathbb{E}\,N}{2}\right). \qquad (3.49)
\end{aligned}$$

We can bound the second term using Chebyshev's inequality:

$$\mathbb{P}\left(N < \frac{\mathbb{E}\,N}{2}\right) \leq \frac{4Var(N)}{(\mathbb{E}\,N)^2} \leq \frac{4\mathbb{E}\,Y^2}{n(\mathbb{E}\,Y)^2} = O(n^{-1}). \qquad (3.50)$$

We have

$$\begin{aligned}
&\mathbb{E}\left(1 \wedge \frac{X_1^2}{2m}\right) X_1 = \mathbb{E}\left(1 \wedge \frac{X_1^2}{2m}\right) X_1 \mathbb{I}_{X_1 \leq m^{1/2}} + \mathbb{E}\left(1 \wedge \frac{X_1^2}{2m}\right) X_1 \mathbb{I}_{X_1 > m^{1/2}} \\
&\leq \frac{m^{1/2}\mathbb{E}\,X_1^2}{2m} + \mathbb{E}\,X_1 \mathbb{I}_{X_1 > m^{1/2}} \leq \frac{m^{1/2}\mathbb{E}\,Y^2}{2n} + m^{-1/2}\mathbb{E}\,X_1^2.
\end{aligned}$$

Putting the last two bounds in (3.49) we get that

$$\mathbb{P}(\bar{B}) \le \frac{\mathbb{E}\,Y^2}{n^{1/2}\mathbb{E}\,Y} + \frac{2\mathbb{E}\,Y^2}{n^{1/2}\mathbb{E}\,Y} + o(n^{-1}) = O(n^{-1/2})$$

The event C**.** First, using (3.50), we have that

$$\mathbb{P}(\bar{C}) = \mathbb{P}(\bar{C}, N \ge 0.5\mathbb{E}\,N) + O(n^{-1}).$$

The balls b_1, b_2 are "undefined" if and only if all N balls fall into different bins. For $N > m$ this cannot happen. For $N \le m$, the probability of this is

$$\left(1 - \frac{1}{m}\right) \dots \left(1 - \frac{N-1}{m}\right) \le e^{-\frac{N(N-1)}{m}}.$$

So

$$\mathbb{P}(\{b_1, b_2\} \cap \{undefined\} \ne \emptyset, N \ge 0.5\mathbb{E}\,N) \le e^{-\frac{0.5\mathbb{E}\,N(0.5\mathbb{E}\,N-1)}{m}} = e^{-\Omega(n)} \to 0.$$

Notice that $b_1, b_2 \in [N]$ implies that $b_1', b_2' \in [N]$. Now recall that when b_1, b_2, b_1', $b_2' \in [N]$ (see Lemma 3.4.5), $b_1 \ne b_2$ and $b_1' \ne b_2'$. Finally for $i, j \in \{1, 2\}$ using the fact that, conditionally on N, b_i' is uniformly distributed in $[N]$ and independent of b_j,

$$\mathbb{P}(b_i' = b_j, N \ge 0.5\mathbb{E}\,N) \le \frac{2}{\mathbb{E}\,N} \to 0;$$

and it follows by the union bound and the above estimates that $\mathbb{P}(\bar{C}) = o(1)$.

The event D**.** We will define random vertices $u^*, v^* \in \{0\} \cup [n]$. We can assume that b_1' and b_2' are generated, given N, as follows. Let b_1^*, b_2^* be drawn independently and uniformly at random from $\{undefined\} \cup [N]$. If $b_1^* \ne b_2^*$ or $N \ge 2$ and $b_1, b_2 \notin \{undefined\}$, we set $(b_1', b_2') = (b_1^*, b_2^*)$. Otherwise, we let $(b_1', b_2') = (b_1'', b_2'')$, where (b_1'', b_2'') is a new uniformly random pair of distinct balls from $[N]$, independent from (b_1^*, b_2^*). For $i \in \{1, 2\}$ set $l_i^* = label_{\mathcal{C}_0}(b_i^*)$ if $b_i^* \in [N]$ and $l_i^* = 0$ otherwise and define $(u^*, v^*) = (l_1^*, l_2^*)$.

Using (3.50) we get that $(u^*, v^*) = (l_1', l_2')$ with probability $O(n^{-1})$.

Conditionally on $X_1, \dots, X_n$, u^* and v^* are independent (and also independent of $G(\mathcal{C}'))$, and

$$\mathbb{P}_*(u^* = i) = \mathbb{P}_*(v^* = i) = \frac{X_i}{N+1} \qquad \text{for } i = 1, 2, \dots, n.$$

Now we have

$$\mathbb{P}(\bar{D}) \leq \mathbb{P}(S_{l'_1}(\mathcal{C}') \cap S_{l'_2}(\mathcal{C}') \neq \emptyset) = \mathbb{P}(u^*v^* \in E(G(\mathcal{C}'))) + O(n^{-1}) \tag{3.51}$$
$$\leq \mathbb{E}\,\mathbb{E}_*\left(\mathbb{P}_*(u^*v^* \in E(G(\mathcal{C}')))\mathbb{I}_{N \geq 0.5\mathbb{E}\,N}\right) + \mathbb{P}(N < 0.5\mathbb{E}\,N) + O(n^{-1}).$$

The second term in the last line is $O(n^{-1})$ by (3.50). The first term is at most

$$\mathbb{E}\,\mathbb{E}_*\left(\sum_{i \neq j} \frac{X_i X_j}{(N+1)^2} \mathbb{I}_{N \geq 0.5\mathbb{E}\,N} \mathbb{P}_*(ij \in E(G(\mathcal{C}')))\right)$$
$$\leq \frac{4n^2}{(\mathbb{E}\,N+2)^2} \mathbb{E}\,X_1 X_2 \mathbb{P}_*(12 \in E(G(\mathcal{C}'))) \leq \frac{4n^2 \mathbb{E}\,X_1^2 X_2^2}{(\mathbb{E}\,N+2)^2 m} = O(n^{-1}).$$

This yields that $\mathbb{P}(\bar{D}) = O(n^{-1})$.

The event E**.** As discussed above, $\mathcal{C}'$, $\mathcal{C}''$ and the configuration used in the proof of Theorem 3.1.3 all have the same distribution. So $\mathbb{P}(E) \to 1$ by the proof of Theorem 3.1.3.

The event F_1**.** Fix $i \in \{1, 2\}$. Let F_{1i} be the event that the bin L in $\mathcal{C}_0$ does not contain a ball with label l_i^*. Since $l'_i = l_i^*$ whp, it is enough to show that $\mathbb{P}(\bar{F}_{1i}) = o(1)$. Let $\tilde{F}$ be the event that $N \geq 0.5\mathbb{E}\,N > 2$, $\sum_i X_i = N$, $\sum_i X_i^2 \leq m \ln n$ and $\omega'(\mathcal{C}_0) \leq (\ln n)^2$. Then

$$\mathbb{P}(\bar{F}_{1i}|\tilde{F}) \leq \mathbb{E} \sum_i \frac{X_i \omega'(\mathcal{C}_0)}{N} \frac{X_i}{N+1} \leq \frac{(\ln n)^3}{\mathbb{E}\,Y n} = o(1).$$

By the inequality (3.40) shown in the proof of Theorem 3.1.3

$$\mathbb{P}(\omega'(\mathcal{C}_0) > (\ln n)^2) \to 0. \tag{3.52}$$

By (3.38) and Markov's inequality, $\mathbb{P}(\sum_i X_i^2 > m \ln n) = o(1)$. Therefore by the union bound

$$\mathbb{P}(\tilde{F}) \geq 1 - \mathbb{P}(\omega'(\mathcal{C}_0) > (\ln n)^2) - \mathbb{P}(\sum_i X_i^2 > m \ln n) - \mathbb{P}(N < 0.5\mathbb{E}\,N) = 1 - o(1).$$

So $\mathbb{P}(\bar{F}_{1i}) \leq \mathbb{P}(\bar{F}_{1i}|\tilde{F}) + o(1) \to 0$.

The event F_2**.** Fix $j \in \{1, 2\}$. The event C implies that $b_j \in T_{w'_j}(\mathcal{C}''_0)$. The event E implies that $l_1 \neq l_2$. It suffices to prove that the following two events occur whp: $\tilde{A}$: in the configuration $\mathcal{C}_0$, the bin w'_j does not contain a ball with label l_j; and $\tilde{B}$: in the second phase no ball with label l_j is added to the bin w'_j. Denote $\tilde{S} = S_{l_j}(\mathcal{C}_0)$ (in the case where $b_j = undefined$, this set is empty). Since $\mathbb{P}(b'_j \neq b^*_j) = O(n^{-1})$ we may replace b'_j with b^*_j. Define $X_0 = X_0(n) = 0$ and recall that $l_j = 0$ when b_j is undefined. Using the conditional independence of b^*_j and $\mathcal{C}_0$,

$$\mathbb{P}_*(b^*_j \in bin^{-1}_{\mathcal{C}_0}(\tilde{S})) \leq (N+1)^{-1}\mathbb{E}_*|bin^{-1}_{\mathcal{C}_0}(\tilde{S})|$$

Given $X_1, \ldots, X_n$, l_j and $\mathcal{C}_0$, such that $\omega'(\mathcal{C}_0) \leq (\ln n)^2$ we have

$$bin^{-1}_{\mathcal{C}_0}(\tilde{S}) \leq X_{l_j}(\ln n)^2.$$

By symmetry, given $X_1, \ldots, X_n$ and $\mathcal{C}_0$ such that $\omega'(\mathcal{C}_0) \geq 2$, the ball b_j is uniformly distributed in $[N]$. We have

$$\mathbb{E}_* X_{l_j} \leq \frac{X_1^2 + \cdots + X_n^2 + 1}{N+1}.$$

Therefore, for n large enough, using (3.50), (3.52) and the above estimates

$$\begin{aligned}\mathbb{P}(\bar{\tilde{A}}) \leq & \mathbb{E}\,\mathbb{I}_{\omega(\mathcal{C}_0) \leq (lnn)^2}\mathbb{I}_{N \geq 0.5N}\mathbb{P}_*(b^*_j \in bin^{-1}_{\mathcal{C}_0}(\tilde{S})) \\ & + \mathbb{P}(b^*_j \neq b'_j) + \mathbb{P}(\omega(\mathcal{C}_0) > (\ln n)^2) + \mathbb{P}(N < 0.5\mathbb{E}\,N) \\ \leq & \frac{4n(\ln n)^2 \mathbb{E}\,X_1^2}{(\mathbb{E}\,N+2)^2} + o(1) = o(1).\end{aligned}$$

Now consider the event $\tilde{B}$. We have shown in the proof of Theorem 3.1.3 that there is $\delta_n \to 0$, such that

$$\mathbb{P}\left(N, N', N'' \in ((1-\delta_n)\mathbb{E}\,N, (1+\delta_n)\mathbb{E}\,N)\right) \to 1.$$

The number of balls added in the second phase depends on $\mathcal{C}_0$ and b_1, b_2, b'_1, b'_2, but suppose $T = 10\delta_n \mathbb{E}\,N$ balls are generated (independently) in advance together with their bin numbers (chosen independently from $[m]$). Then whp, these T balls are enough to complete the construction of $\mathcal{C}'$ and $\mathcal{C}''$, and, since $m = \Omega(n)$, the probability that at least one of these balls falls into the bin w'_j is at most

$$\frac{T}{m} \leq \frac{10\delta_n n^{1/2}}{m^{1/2}} \to 0.$$

So

$$\mathbb{P}(\bar{\tilde{B}}) \leq \frac{T}{m} + \mathbb{P}((N'-N) + (N''-N) > 4\delta_n \mathbb{E}\,N) \to 0.$$

The event A'. Denote by B_k the event that $S_k(\mathcal{C}') \cap S_{u^*}(\mathcal{C}') \neq \emptyset$ and $S_k(\mathcal{C}') \cap S_{v^*}(\mathcal{C}') \neq \emptyset$. Denote by B_{k1} the event that $S_k(\mathcal{C}') \cap S_{u^*}(\mathcal{C}') \cap S_{v^*}(\mathcal{C}') \neq \emptyset$, and by B_{k2} the event $B_k \setminus B_{k1}$. Since $(l_1', l_2') = (u^*, v^*)$ whp, it suffices to show that

$$\mathbb{P}(\exists k \notin \{u^*, v^*\} : B_k) \leq \mathbb{P}(\exists k \notin \{u^*, v^*\} : B_{k1}) + \mathbb{P}(\exists k \notin \{u^*, v^*\} : B_{k2}) \to 0.$$

In the proof we only need to work with the random intersection graph $G(\mathcal{C}')$. Using (3.51) above, we have

$$\begin{aligned}
&\mathbb{P}(\exists k \notin \{u^*, v^*\} : B_{k1}) \leq \mathbb{P}(u^* v^* \in E(G(\mathcal{C}'))) \to 0. \\
&\mathbb{P}(\exists k \notin \{u^*, v^*\} : B_{k2}) \leq \mathbb{P}(N < 0.5 \mathbb{E}\, N) + \mathbb{P}(u^* = v^*) \\
&\quad + \mathbb{E}\, \mathbb{I}_{N \geq 0.5 \mathbb{E}\, N} \mathbb{I}_{u^* \neq v^*} \sum_{k \in [n] \setminus \{u^*, v^*\}} \mathbb{P}_*(B_{k2})
\end{aligned}$$

The first and second terms on the right of the last inequality are $O(n^{-1})$ using (3.50). The last term is at most

$$\begin{aligned}
&\mathbb{E}\left(\sum_{i \neq j} \mathbb{I}_{N \geq 0.5 \mathbb{E}\, N} \frac{X_i}{(N+1)} \frac{X_j}{(N+1)} \sum_{k \in [n] \setminus \{i,j\}} \mathbb{P}_*(B_{k2} | u^* = i, v^* = j) \right) \\
&\leq \frac{4}{m^2 (\mathbb{E}\, N + 2)^2} \sum_{i,j,k \in [n], i \neq j, i \neq k, j \neq k} \mathbb{E}\, X_i^2 X_j^2 X_k^2 \leq \frac{4 (\mathbb{E}\, Y^2)^3}{n (\mathbb{E}\, Y)^2} = O(n^{-1})
\end{aligned}$$

Here in the last line we used

$$\mathbb{P}_*(B_{k2} | u^* = i, v^* = j) \leq \frac{X_i X_j (X_k)_2}{m^2},$$

which follows using the union bound and the observation that if $S_i(\mathcal{C}')$ and $S_j(\mathcal{C}')$ both intersect $S_k(\mathcal{C}')$, but the intersection of all three is empty, then there must be distinct elements w_1, w_2, such that $w_1 \in S_i(\mathcal{C}') \cap S_k(\mathcal{C}')$ and $w_2 \in S_j(\mathcal{C}') \cap S_k(\mathcal{C}')$. By the union bound we get $\mathbb{P}(\exists k \notin \{u^*, v^*\} : B_k) \to 0$.

□

3.5 Equivalence between set size and degree parameters

Here we prove Lemmas 3.1.5 and 3.1.6. In the proof we write $X = X(n)$, $Y = Y(n)$, and $D_1 = D_1(n)$. We denote $X_1, X_2, \dots$ the sizes of subsets $S_1, S_2, \dots \subseteq W$ prescribed to the vertices $1, 2, \dots \in V = [n]$ of $G(n)$.

Proof of Lemma 3.1.5 We start by showing that if either $\mathbb{E}\,Y$ or $\mathbb{E}\,D_1$ converges and for some positive sequence $\{a_n\}$ converging to zero (we write $a = a_n$ for short),

$$\mathbb{E}\,Y\mathbf{I}_{\{Y>(an)^{1/2}\}} \to 0 \tag{3.53}$$

then

$$\mathbb{E}\,Y = (\mathbb{E}\,D_1)^{1/2} + o(1). \tag{3.54}$$

We note that $\mathbb{E}\,D_1 = (n-1)\mathbb{P}(S_1 \cap S_2 \neq \emptyset)$. We estimate this probability using the inequalities, see Lemma 6 in [18],

$$\frac{X_1X_2}{m} \geq \mathbb{P}(S_1 \cap S_2 \neq \emptyset | X_1, X_2) \geq \max\left\{0, \left(\frac{X_1X_2}{m} - \frac{X_1^2X_2^2}{m^2}\right)\right\} =: Z. \tag{3.55}$$

Notice that $\mathbb{E}\,Y = \Omega(1)$. This is clear if $\mathbb{E}\,Y \to y \in (0;\infty)$. Otherwise, we have $\mathbb{E}\,D_1 \to d \in (0;\infty)$ and, by the first inequality of (3.55),

$$(n-1)\frac{(\mathbb{E}\,Y)^2}{n} \geq (n-1)\mathbb{P}(S_1 \cap S_2 \neq \emptyset) = \mathbb{E}\,D_1.$$

Furthermore, from $\mathbb{E}\,Y = \Omega(1)$ and (3.53) we conclude that $\mathbb{E}\,X\mathbf{I}_{\{X \geq (am)^{1/2}\}} = o(\mathbb{E}\,X)$. Using this bound we estimate $\mathbb{E}\,Z$ from below

$$\begin{aligned}\mathbb{E}\,Z &\geq \mathbb{E}\,Z\mathbf{I}_{X_1X_2 \leq am} \geq (1-a)m^{-1}\mathbb{E}\,X_1X_2\mathbf{I}_{X_1X_2 \leq am}\\ &\geq (1-a)m^{-1}\mathbb{E}\,X_1\mathbb{E}\,X_2 - m^{-1}\mathbb{E}\,X_1X_2\mathbf{I}_{X_1X_2>am},\end{aligned} \tag{3.56}$$

where

$$\begin{aligned}\mathbb{E}\,X_1X_2\mathbf{I}_{X_1X_2>am} &\leq \mathbb{E}\,X_1X_2\left(\mathbf{I}_{X_1>(am)^{1/2}} + \mathbf{I}_{X_2>(am)^{1/2}}\right)\\ &\leq 2\mathbb{E}\,X\mathbb{E}\,X\mathbf{I}_{X>(am)^{1/2}}\\ &= o((\mathbb{E}\,X)^2).\end{aligned} \tag{3.57}$$

Hence, $\mathbb{E}\,Z \geq (1-o(1))(\mathbb{E}\,X)^2$. Combining this inequality with (3.55) we obtain

$$\mathbb{P}(S_1 \cap S_2 \neq \emptyset) \sim m^{-1}(\mathbb{E}\,X)^2,$$

thus proving (3.54).

It remains to prove that (3.2)⇔(3.6). Since both implications are shown in much the same way, we only prove (3.2)⇒(3.6). For this purpose we fix $0 < \tilde{\epsilon} < \min\{\epsilon, \epsilon_0\}$ and show that for each $0 < \delta < 1$ and each sequence $\{t_n\}$ with

$n^{1/2-\tilde{\epsilon}} \leq t_n \leq n^{1/2+\tilde{\epsilon}}$

$$\liminf_n \big(\mathbb{P}(Y_1(n) \geq t_n)/\mathbb{P}(D_1(n) \geq t_n)\big) \geq (d^{1/2}(1+\delta))^{-\alpha}, \qquad (3.58)$$
$$\limsup_n \big(\mathbb{P}(Y_1(n) \geq t_n)/\mathbb{P}(D_1(n) \geq t_n)\big) \leq (d^{1/2}(1-\delta))^{-\alpha}. \qquad (3.59)$$

Here the random variable $Y_1(n) := (n/m)^{1/2} X_1(n)$ has the same distribution as $Y(n)$. We prove (3.58) and (3.59) by contradiction.

Proof of (3.58). Suppose there is an increasing sequence $\{n_k\}$ of positive integers and a sequence $\{b_k\}$ with $n_k^{1/2-\tilde{\epsilon}} \leq b_k \leq n_k^{1/2+\tilde{\epsilon}}$ such that, for some $0 < \delta < 1$,

$$\mathbb{P}(Y_1(n_k) \geq b_k) < (d^{1/2}(1+\delta))^{-\alpha}\mathbb{P}(D_1(n_k) \geq b_k), \quad k = 1, 2, \ldots. \qquad (3.60)$$

Define $\{l_k\}$ by the relation $b_k = d^{1/2}(1+\delta/2)l_k$, $k \geq 1$. Introduce events $\mathcal{A}_k = \{D_1(n_k) \geq b_k\}$, $\mathcal{B}_k = \{Y_1(n_k) \geq l_k\}$ and write

$$\mathbb{P}(\mathcal{A}_k) = \mathbb{P}(\mathcal{A}_k \cap \mathcal{B}_k) + \mathbb{P}(\mathcal{A}_k \cap \bar{\mathcal{B}}_k). \qquad (3.61)$$

In what follows we drop the subscript k and write b, l, n, m instead of b_k, l_k, n_k, m_k. We note that (3.2) together with (3.60) imply

$$\mathbb{P}(\mathcal{A} \cap \mathcal{B}) \leq \mathbb{P}(\mathcal{B}) \sim d^{\alpha/2}(1+\delta/2)^{\alpha}\mathbb{P}(Y_1(n) \geq b) \leq c_1 \mathbb{P}(\mathcal{A}),$$

where the constant $c_1 = ((1+\delta/2)/(1+\delta))^{\alpha} < 1$. Next we show that $\mathbb{P}(\mathcal{A} \cap \bar{\mathcal{B}}) = O(n^{-10})$ thus obtaining a contradiction to (3.60), (3.61).

Denote $x = \lfloor (m/n)^{1/2} l \rfloor$. Conditionally, given the event $\mathcal{C} = \{X_1(n) = x\}$, the random variable $D_1(n)$ has binomial distribution $Bin(n-1, p)$ with success probability $p = \mathbb{P}(S_1 \cap S_2 \neq \emptyset \big| |S_1| = x)$ satisfying $p \sim d^{1/2} l/n$. Indeed, the first inequality of (3.55) implies

$$p \leq \frac{x \mathbb{E}\, X_2}{m} = \frac{x (m/n)^{1/2} \mathbb{E}\, Y}{m} \sim d^{1/2} \frac{l}{n}.$$

Here we used $\mathbb{E}\, Y \to d^{1/2} > 0$. The second inequality of (3.55) implies, see (3.56),

$$p \geq \frac{1-a}{m} x \mathbb{E}\, X_2 \mathbb{I}_{\{x X_2 < am\}} = \frac{1-a}{m} x (\mathbb{E}\, X_2 - r) \sim \frac{x \mathbb{E}\, X_2}{m}.$$

Here $r = \mathbb{E}\, X_2 \mathbb{I}_{\{x X_2 \geq am\}} = o(\mathbb{E}\, X_2)$, for $a = a(n_k) = \ln^{-1} n_k$, cf. (3.57).

Next, since $b \sim (1+\delta/2)np$ and $np \sim d^{1/2} l = \Omega(n^{1/2-\tilde{\epsilon}})$ we obtain, by Chernoff's inequality, $\mathbb{P}(\mathcal{A}|\mathcal{C}) = O(n^{-10})$. Now, using the inequality $\mathbb{P}(\mathcal{A}|Y_1(n) = y) \leq \mathbb{P}(\mathcal{A}|\mathcal{C})$, for $y \leq l$, we obtain

$$\mathbb{P}(\mathcal{A} \cap \bar{\mathcal{B}}) = \mathbb{E}\,\mathbb{P}(\mathcal{A}|Y_1(n))\mathbb{I}_{\{Y(n)\leq l\}} \leq \mathbb{P}(\mathcal{A}|\mathcal{C}) = O(n^{-10}). \tag{3.62}$$

Proof of (3.59). Suppose there is an increasing sequence $\{n_k\}$ of positive integers and a sequence $\{b_k\}$ with $n_k^{1/2-\tilde{\epsilon}} \leq b_k \leq n_k^{1/2+\tilde{\epsilon}}$ such that, for some $0 < \delta < 1$,

$$\mathbb{P}(Y_1(n_k) \geq b_k) > (d^{1/2}(1-\delta))^{-\alpha}\mathbb{P}(D_1(n_k) \geq b_k), \quad k = 1, 2, \ldots. \tag{3.63}$$

Define $\{l_k\}$ by the relation $b_k = d^{1/2}(1-\delta/2)l_k$, $k \geq 1$. We write

$$\mathbb{P}(D_1(n_k) \geq b_k) = \mathbb{P}(Y_1(n_k) \geq l_k)\mathbb{P}(D_1(n_k) \geq b_k|Y_1(n_k) \geq l_k). \tag{3.64}$$

We note that, by (3.2) and (3.63), the first term on the right is at least $(c_2 + o(1))$ $\mathbb{P}(D_1(n_k) \geq b_k)$ where the constant $c_2 = ((1-\delta/2)/(1-\delta))^{\alpha} > 1$. Finally, we obtain a contradiction, by showing that the second term of (3.64) is $1 - O(n^{-10})$. Here we proceed as in (3.62) above. We write

$$\mathbb{P}(D_1(n_k) < b_k|Y_1(n_k) \geq l_k) \leq \mathbb{P}(D_1(n_k) < b_k|\mathcal{C})$$

and show that binomial probability on the right-hand side is $O(n^{-10})$ using Chernoff's inequality.

□

Proof of Lemma 3.1.6 The identity (3.9) follows from (3.54) since

$$\mathbb{E}\,Y\mathbf{I}_{Y>\epsilon_n n^{1/2}} \leq (\mathbb{E}\,Y^2\mathbf{I}_{Y>\epsilon_n n^{1/2}})^{1/2} \to 0.$$

Let us show (3.10). Denote N the number of 2-stars in $G = G(n)$ centered at vertex $1 \in V = [n]$. Introduce the events $\mathcal{A}_{ij} = \{i \sim j\}$, $i, j \in V$. Write, for short, $\mathcal{A} = \mathcal{A}_{12} \cap \mathcal{A}_{13}$. Let $\tilde{\mathbb{P}}$ denote the conditional probability given the sizes X_1, X_2, X_3 of the random subsets prescribed to vertices $1, 2, 3 \in V$. We remark that (3.10) follows from (3.9) combined with the simple identities

$$\mathbb{E}\,D_1(D_1 - 1) = 2\mathbb{E}\,N = (n-1)(n-2)\mathbb{P}(\mathcal{A}),$$

and the inequalities

$$(\mathbb{E}\,Y)^2\mathbb{E}\,Y^2 \ge n^2\mathbb{P}(\mathcal{A}) \ge (1-o(1))(\mathbb{E}\,Y)^2\mathbb{E}\,Y^2. \tag{3.65}$$

Let us prove (3.65). For this purpose we write (using the conditional independence of events $\mathcal{A}_{12}$ and $\mathcal{A}_{13}$, given X_1, X_2, X_3)

$$\mathbb{P}(\mathcal{A}) = \mathbb{E}\,\tilde{\mathbb{P}}(\mathcal{A}) = \mathbb{E}\,\tilde{\mathbb{P}}(\mathcal{A}_{12})\tilde{\mathbb{P}}(\mathcal{A}_{13})$$

and evaluate conditional probabilities $\tilde{\mathbb{P}}(\mathcal{A}_{ij})$ using (3.55). From the first inequality of (3.55) we obtain the first inequality of (3.65)

$$\mathbb{P}(\mathcal{A}) = \mathbb{E}\,\tilde{\mathbb{P}}(\mathcal{A}_{12})\tilde{\mathbb{P}}(\mathcal{A}_{13}) \le \mathbb{E}\,(X_1^2X_2X_3)/m^2 = (\mathbb{E}\,Y)^2\mathbb{E}\,Y^2/n^2.$$

Thus, even without the assumption (3.8) (we use this fact this in the proof of Proposition 3.4.3), we have

$$\mathbb{E}\,D_1 \le \mathbb{E}\,Y \quad \text{and} \quad \mathbb{E}\,D_1(D_1-1) \le \mathbb{E}\,Y^2\mathbb{E}\,Y. \tag{3.66}$$

To show the second inequality of (3.65) we apply the second inequality of (3.55) and use truncation. We denote $\mathbb{I}_i = \mathbb{I}_{\{X_i \le \epsilon_n m^{1/2}\}}$, $\bar{\mathbb{I}}_i = 1 - \mathbb{I}_i$ and write, cf. (3.56),

$$\begin{aligned} \mathbb{P}(A) &\ge \mathbb{E}\,\tilde{\mathbb{P}}(\mathcal{A})\mathbb{I}_1\mathbb{I}_2\mathbb{I}_3 \ge (1-\epsilon_n^2)^2\mathbb{E}\,(X_1^2X_2X_3/m^2)\mathbb{I}_1\mathbb{I}_2\mathbb{I}_3 \\ &\ge (1-\epsilon_n^2)^2\mathbb{E}\,(X_1^2X_2X_3/m^2)(1-\bar{\mathbb{I}}_1-\bar{\mathbb{I}}_2-\bar{\mathbb{I}}_3) \\ &= (1-o(1))(\mathbb{E}\,Y)^2\mathbb{E}\,Y^2/n^2. \end{aligned}$$

In the last step we used the fact that $\mathbb{E}\,Y^2 \ge (\mathbb{E}\,Y)^2 = \Omega(1)$ and the bounds

$$\begin{aligned} &\mathbb{E}\,X_1^2\bar{\mathbb{I}}_1 = (m/n)\,\mathbb{E}\,Y^2\mathbb{I}_{\{Y>\epsilon_n n^{1/2}\}} = o(\mathbb{E}\,X^2), \\ &\mathbb{E}\,X_j\bar{\mathbb{I}}_j = (m/n)^{1/2}\mathbb{E}\,Y\mathbb{I}_{\{Y>\epsilon_n n^{1/2}\}} = o(\mathbb{E}\,X), \qquad j=2,3. \end{aligned}$$

□

3.6 Concluding remarks

In this work we determined the order of the clique number in $G(n, m, P)$ for a wide range of $m = m(n)$ and $P = P(n)$. We saw that in sparse power-law random intersection graphs with unbounded degree variance, the clustering property of $G(n, m, P)$ has little influence in the formation of the maximum clique. This suggests that simpler models, such as the one in [55], may be preferable in the case of very heavy degree tails. However, when the degree variance is bounded, most random graph models, including the Erdős-Rényi graph and the model of [55] have only bounded size cliques whp. In contrast, we showed that in random intersection graphs the clique number can still diverge slowly.

We have a kind of "phase transition" as the tail index α for the random subset size (degree) varies, see (3.2). Assume, for example that $m = \Theta(n)$. When $\alpha < 2$, the random graph $G(n, m, P)$ whp contains cliques of only logarithmic size. When $\alpha > 2$, it whp contains a 'giant' clique of polynomial size. But what happens when (3.2) is satisfied with $\alpha = 2$ but the degree variance is unbounded?

We proposed a surprisingly simple algorithm for finding (almost) the largest clique in sparse random intersection graphs with finite degree variance. The performance of both Greedy-Clique and Mono-Clique algorithms can be of further interest, since these algorithms do not use the possibly hidden random subset structure. How well would they perform on arbitrary sparse empirical networks? Can we suspect a hidden intersecting sets structure for networks where the Mono-Clique algorithm performs well?

Another direction of possible future research would be to determine the asymptotic clique number in dense random intersection graphs (alternatively, the order of the largest intersecting set in dense random hypergraphs). For example, even in the random uniform hypergraph case where $m = \Theta(n)$ and the random subset size $X(n) = \Omega(n^{1/2})$ is deterministic, exact asymptotics of the clique number remain open.

Chapter 4

On the chromatic index of random uniform hypergraphs

4.1 Introduction

A hypergraph is a pair $H = (V, E)$, where $V = V(H)$ is a set of vertices and $E = E(H)$ is a family of subsets of V, called *edges* (we will allow multiple edges). H is *k–uniform* if all of its edges are of size k. The *chromatic index* of H, denoted $\chi'(H)$ is the smallest number of colours needed to colour its edges so that no two intersecting edges share the same colour. We consider the chromatic index of the random hypergraph $\mathbb{H}^{(k)}(n, N)$.

This problem is related to the generalisation of Vizing's theorem for hypergraphs. Some related results and conjectures may be found in [2] and [10]. Given a hypergraph H, let $deg_H(x)$ denote the *degree* of a vertex x, that is $deg_H(x) = |\{e : e \in E(H), x \in e\}|$. Let

$$D(H) = \max_{x \in V(H)} deg_H(x) \quad \text{and} \quad d(H) = \min_{x \in V(H)} deg_H(x).$$

Also let $C(H) = \max_{x \neq y} |\{e \in E(G) : x, y \in e\}|$. Vizing's theorem states that for any 2-uniform hypergraph without loops and multiple edges, $\chi'(H)$ is either $D(H)$ or $D(H) + 1$. Obviously, for any hypergraph

$$d(H) \leq D(H) \leq \chi'(H).$$

In 1989 Pippenger and Spencer [82] proved that Vizing's theorem may be extended in a certain sense to uniform hypergraphs in which degrees are concentrated around one value and $C(H)$ is small comparing to $D(H)$.

Theorem 4.1.1 (Pippenger and Spencer [82]) *For every k and every $\epsilon > 0$ there exist $\delta > 0$ and n_0 such that if H is a k-uniform hypergraph on $n \geq n_0$ vertices satisfying $d(H) \geq (1-\delta')D(H)$ and $C(H) \leq \delta' D(H)$ then*

$$\chi'(H) \leq (1+\epsilon)D(H).$$

In this thesis we are interested in asymptotic results concerning algorithmic version of the theorem for random hypergraphs. By $\binom{[n]}{k}$ we will denote the family of all k–element subsets of n. It will also be convenient for us to write $a = b \pm c$ for $a \in [b-c; b+c]$.

Theorem 4.1.1 implies that $\chi'(\mathbb{H}^{(k)}(n,N)) = \bar{d}(1 \pm \epsilon)$ when the degrees of $\mathbb{H}^{(k)}(n,N)$ are large and close to their mean $\bar{d} = \frac{kN}{n}$. This happens whp (with probability tending to 1 as $n \to \infty$) when $\ln n = o(\bar{d})$, however, k is assumed to be fixed. Motivated by the problem of determining the chromatic number of random intersection graphs (see below), we ask whether a similar result holds for random uniform hypergraphs even when the set sizes are unbounded. Moreover, we focus on the algorithmic aspect of the problem. The main result of the chapter is the following theorem.

Theorem 4.1.2 *For any $\epsilon > 0$ there is a constant $c_\epsilon > 0$, such that the following holds. Let $\mathbb{H}^{(k)}(n,N)$, where $k, n, N \geq 2$, be a random hypergraph on vertex set $[n]$ with N edges of size k drawn independently with replacement from the set $\binom{[n]}{k}$. Write $\bar{d} = \frac{kN}{n}$. Suppose*

$$k \leq c_\epsilon \ln\left(\frac{n}{\ln \bar{d}}\right) \quad \text{and} \quad k \leq c_\epsilon \ln\left(\frac{\bar{d}}{\ln n}\right). \tag{4.1}$$

Then there is an algorithm which colours all edges of $\mathbb{H}^{(k)}(n,N)$ with at most $\lceil \bar{d}(1+\epsilon) \rceil$ colours with probability at least $1 - \frac{2}{n} - \frac{2}{\bar{d}}$.

The algorithm of Theorem 4.1.2 is a simple polynomial algorithm which for each edge just selects a random available colour. Its description is given in Section 4.2. Our proof is an application of the differential equations method, which differs from the method used in [82]. It should be pointed out that by a simple coupling argument analogous theorems follow for random hypergraphs with independent edges and N edges chosen without replacement.

As mentioned above, the motivation for our research was studies on the chromatic number of uniform random intersection graphs.

Let N, n and k be positive integers. Moreover, let $V = \{v_1, \ldots, v_N\}$ and $\mathcal{W} = [n]$ be disjoint sets. Recall that by a *uniform random intersection* graph

$\mathcal{G}(N, n, k)$ we mean a graph on the vertex set V in which each vertex $v \in V$ chooses a set S_v uniformly at random from all k–element subsets of the set $\mathcal{W}$. Two vertices $v, v' \in V$ are connected by an edge in $\mathcal{G}(N, n, k)$ if $S_v \cap S_{v'} \neq \emptyset$. Naturally, $\mathcal{G}(N, n, k)$ is an edge graph of the hypergraph $\mathbb{H}^{(k)}(n, N)$. Therefore $\chi'(\mathbb{H}^{(k)}(n, N)) = \chi(\mathcal{G}(N, n, k))$, and Theorem 4.1.2 immediately applies. Results concerning the chromatic number of other models of random intersection graphs might be found in [6] and [80].

The chromatic number of any graph G on N vertices is related to its independence (stability) number $\alpha(G)$ and to the size of the largest clique $\omega(G)$ by simple inequalities

$$\chi(G) \geq \frac{N}{\alpha(G)} \quad \text{and} \quad \chi(G) \geq \omega(G). \tag{4.2}$$

In a series of papers [28, 29, 49, 69, 71] it was shown that for $G = G(N, p)$, the Erdős–Rényi random graph with independent edges, the first inequality of (4.2) is nearly an equality whp when $Np \to \infty$. The independence number of $\mathcal{G}(N, n, k)$ was studied in [94], where it was shown that whp the greedy algorithm usually constructs an independent set of the optimal size $\frac{n}{k}(1 \pm \epsilon)$ when the second inequality of (4.1) holds. Also, the main result of [4] implies that whp $\omega(\mathcal{G}(N, n, k)) = \frac{n}{k}(1 \pm \epsilon)$ whenever $k = o(n^{1/3})$. Our result shows that subject to the assumptions of Theorem 4.1.2, both inequalities of (4.2) are nearly equalities whp when $G = \mathcal{G}(N, n, k)$.

Naturally, it would be interesting to determine $\chi(\mathcal{G}(N, n, k))$ or, equivalently, $\chi'(\mathbb{H}^{(k)}(n, N))$ for other (all) ranges of parameters. In the light of [4, 94] (see also [25]), it seems that the first constraint of (4.1) might not be necessary. Meanwhile, the results of [69] imply that the second constraint should be not far from best possible.

Finally, let us make the following observation. The greedy colouring algorithms for $G(N, p)$ usually use about twice the chromatic number of colours (see, i.e., [50] or Section 7.2 of [56]). Our randomised greedy algorithm (see also [6] and [80]) with high probability colours $\mathcal{G}(N, n, k)$ with an asymptotically optimal number of colours. One could ask whether such an algorithm also exists for other ranges of k and N, perhaps in all cases where the clique number of $\mathcal{G}(N, n, k)$ is whp "approximately" equal to its chromatic number.

4.2 The algorithm

The edges of $\mathbb{H}^{(k)}(n, N)$ can be represented as a sequence $e_1, \ldots, e_N$ of iid sets, where each set is selected uniformly at random from $\binom{[n]}{k}$. We imagine $e_1, \ldots, e_N$

being added to the hypergraph one by one and coloured by a random valid colour. More precisely, we fix a positive integer q, the number of possible colours. The set e_1 is coloured with a uniformly random colour from $[q]$.

The random hypergraph obtained by adding and colouring the first i edges is denoted by $H(i)$. We always consider $H(i)$ together with the (random) colouring of its edges $C(i)$. For any set $S \subseteq [n]$ let $\mathcal{M}_S(i)$ denote the set of all colours $c \in [q]$ such that colouring S with c and adding it to $H(i)$ gives a properly edge-coloured hypergraph (a colouring is *proper* if each pair of intersecting edges has different colours). Let $M_S(i) = |\mathcal{M}_S(i)|$. Thus, $M_S(i)$ is the number of “available” colours for the set S after the step i.

For $i \geq 1$ the edge $e = e_{i+1}$ is coloured with a uniformly random colour $c(e)$ from the subset $\mathcal{M}_e(i)$ (conditionally on $\mathcal{M}_e(i)$, the colour $c(e)$ is independent of $H(i)$ and $C(i)$). If the set $\mathcal{M}_e(i)$ is empty for the random edge $e = e_{i+1}$, then the colour of e remains unassigned.

For a colour $c \in [q]$, let $L_c(i)$ be the number of vertices from $[n]$ which do not belong to an edge coloured c in the hypergraph $H(i)$. In the beginning we have $M_e(0) = q$ for each k-element subset of $[n]$ and $L_c(0) = n$ for each $c \in [q]$. We will prove Theorem 4.1.2 by showing that for any $\epsilon > 0$ the above algorithm with a large enough probability succeeds to colour every edge e_i for $i \leq N(1-\epsilon)$ when $q = \lceil \frac{kN}{n} \rceil$. From this it will follow, that the same algorithm with the claimed probability succeeds to colour all N edges when we start with $q = \lceil (1+2\epsilon)\bar{d} \rceil$.

4.3 Proofs

In the following sections we assume that the integers $N \geq 1, k \geq 2, n \geq 3$ are fixed.

4.3.1 One-step differences

We are going to use the differential equations method, see [30, 101]. We will analyse the randomized colouring algorithm with $q = \lceil \frac{kN}{n} \rceil$ colours. The final result of one run is a random object in the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\Omega = \binom{[n]}{k}^N \times \{0, \ldots, q\}^N$, $\mathcal{F}$ is the σ-field generated by the outcomes of all N edges and their colours, and $\mathbb{P}$ is as described in Section 4.2. The associated natural filter is

$$\{\emptyset, \Omega\} = \mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_N = \mathcal{F},$$

where $\mathcal{F}_i$ is the σ-field generated by the random edges $e_1, \ldots, e_i$ and their colours $c(e_1), \ldots, c(e_i)$. Corresponding to each $\mathcal{F}_i$ is a natural partition of Ω into blocks that generate $\mathcal{F}_i$, that is, a *block* of $\mathcal{F}_i$ is the set of all $\omega \in \Omega$ corresponding to a particular sequence of the first i edges and their colours.

For $i \in \{0, \ldots, N-1\}$, some nonnegative $\mu_M(i), \mu_L(i)$ and small nonnegative $k_1(i), k_2(i)$, we will consider the events that for each k-set $e \in \binom{[n]}{k}$

$$M_e(i) = \mu_M(i) \pm k_1(i), \tag{4.3}$$

and for each colour $c \in [q]$

$$L_c(i) = \mu_L(i) \pm k_2(i). \tag{4.4}$$

The numbers $\mu_M(i)$ and $\mu_L(i)$ will be defined later, but one can think for now that they are expected values of $M_e(i)$ and $L_c(i)$ respectively, while $k_1(i)$ and $k_2(i)$ are the corresponding "errors". So we will assume that $\mu_L(i) \leq n$.

Observe that the difference $M_e(i+1) - M_e(i)$ is either -1 or 0; and $L_c(i+1) - L_c(i)$ is either $-k$ or 0. If $M_e(i+1) - M_e(i) < 0$ or $L_c(i+1) - L_c(i) < 0$ we will say that M_e or, respectively, L_c, *decreases (at step $i+1$)*.

Lemma 4.3.1 *There is a constant $D \geq 1$ such that the following holds. Suppose $i \in \{0, 1, \ldots, N-1\}$ and positive numbers $\mu_M(i)$, $\mu_L(i)$, $k_1(i)$, $k_2(i)$ satisfy*

$$k^2 < n/2, \quad k_1(i) < \mu_M(i)/2 \quad \text{and} \quad k^2 + kk_2(i) < \mu_L(i) \leq n. \tag{4.5}$$

Let B_i be a block of the partition generating $\mathcal{F}_i$ (an atom of $\mathcal{F}_i$). Suppose (4.3) and (4.4) are satisfied conditionally on B_i. Then for each $e \in \binom{[n]}{k}$

$$\begin{aligned} &\mathbb{P}\left(M_e(i+1) - M_e(i) = -1 | B_i\right) \\ &= \frac{k^2 \mu_L(i)^{k-1}}{n^k} \left(1 \pm D\left(\frac{k_1(i)}{\mu_M(i)} + \frac{k^2 + kk_2(i)}{\mu_L(i)}\right)\right) \end{aligned}$$

and for each $c \in [q]$

$$\begin{aligned} &\mathbb{P}\left(L_c(i+1) - L_c(i) = -k | B_i\right) \\ &= \frac{\mu_L(i)^k}{\mu_M(i) n^k} \left(1 \pm D\left(\frac{k_1(i)}{\mu_M(i)} + \frac{k^2 + kk_2(i)}{\mu_L(i)}\right)\right). \end{aligned}$$

Proof Since i is fixed, we write $\mu_L = \mu_L(i)$, $\mu_M = \mu_M(i)$, $k_j = k_j(i)$, $M_S =$

$M_S(i)$, $L_c = L_c(i)$ and $\mathcal{M}_S = \mathcal{M}_S(i)$. Conditionally on B_i, all random variables that are $\mathcal{F}_i$-measurable are constant. This includes M_S for any $S \subseteq [n]$ and L_c for any $c \in [q]$, etc. Note that (4.3 - 4.5) imply that $M_e > 0$ for any $e \in \binom{[n]}{k}$, so each edge of $H(i)$ is assigned some colour on B_i.

In order for M_e to decrease at step $i+1$, e_{i+1} has to be incident to e. So

$$\mathbb{P}(M_e \text{ decreases } |B_i) = \sum_{f:f\sim e} \mathbb{P}(e_{i+1} = f)\mathbb{P}(M_e \text{ decreases}|e_{i+1} = f, B_i)$$

$$= \binom{n}{k}^{-1} \sum_{f:f\sim e} \mathbb{P}(M_e \text{ decreases}|e_{i+1} = f, B_i)$$

Here f ranges over all k–element subsets of $[n]$ intersecting e (we write $e \sim f$ if edges f and e share at least one vertex). Since the algorithm picks a colour $c = c(e_{i+1})$ uniformly at random from all available ones, and since M_e decreases only in the case if in the hypergraph $H(i)$ both e and e_{i+1} could be coloured c, we get

$$\mathbb{P}\left(M_e \text{ decreases } | \, e_{i+1} = f, B_i\right) = \frac{M_{f\cup e}}{M_f}.$$

We can approximate the denominator

$$M_f^{-1} = (\mu_M \pm k_1)^{-1} = \mu_M^{-1}\left(1 \pm \frac{2k_1}{\mu_M}\right) \tag{4.6}$$

to get

$$\mathbb{P}\left(M_e \text{ decreases } |B_i\right) = \binom{n}{k}^{-1} \mu_M^{-1}\left(1 \pm \frac{2k_1}{\mu_M}\right) \sum_{f:f\sim e} M_{f\cup e}. \tag{4.7}$$

For any set $S \subseteq [n]$ and any $c \in [q]$ let $A_{c,S} = A_{c,S}(i)$ be the indicator of the event that adding S coloured c to $H(i)$ gives a properly edge-coloured hypergraph. Then

$$M_S = \sum_{c\in[q]} A_{c,S}.$$

Let $\mathcal{L}_c = \mathcal{L}_c(i)$ be the set of vertices which are not contained in an edge coloured c in $H(i)$. Recall that $|\mathcal{L}_c| = L_c$. Note that $A_{c,e\cup f} = A_{c,e}A_{c,f\setminus e} = 1$ if both e and $f \setminus e$ are subsets of $\mathcal{L}_c$. If $e \subseteq \mathcal{L}_c$ then the number of k-subsets f of $\mathcal{L}_c$ intersecting e is at least $k\binom{L_c-k}{k-1}$ and at most $k\binom{L_c-1}{k-1}$. Using (4.4) and (4.5) we have

$$\binom{L_c - k}{k-1} \geq \frac{\mu_L^{k-1}}{(k-1)!}\left(1 - \frac{kk_2 + 1.5k^2}{\mu_L}\right).$$

Also, by (4.5) we have $x = k_2/\mu_L \leq 1/k$, so using simple inequalities $(1+x)^{k-1} \leq$

$1 + kx(1+x)^{k-1}$ and $(1+x)^{k-1} \le e$

$$\binom{L_c - 1}{k-1} \le \frac{\mu_L^{k-1}}{(k-1)!}\left(1 + \frac{ekk_2}{\mu_L}\right).$$

Thus if $A_{c,e} = 1$ then

$$\sum_{f:f\sim e} A_{c,f\setminus e} = \frac{k\mu_L^{k-1}}{(k-1)!}\left(1 \pm \frac{3(k^2+kk_2)}{\mu_L}\right).$$

Therefore

$$\begin{aligned}
\sum_{f:f\sim e} M_{f\cup e} &= \sum_{f:f\sim e}\sum_{c\in[q]} A_{c,f\cup e} \\
&= \sum_{c\in[q]}\sum_{f:f\sim e} A_{c,f\cup e} = \sum_{c\in[q]} A_{c,e}\frac{k\mu_L^{k-1}}{(k-1)!}\left(1 \pm \frac{3(k^2+kk_2)}{\mu_L}\right) \\
&= \frac{k\mu_L^{k-1}\mu_M}{(k-1)!}\left(1 \pm \frac{3(k^2+kk_2)}{\mu_L}\right)\left(1 \pm \frac{k_1}{\mu_M}\right).
\end{aligned}$$

In the last step we used $\sum_{c\in[q]} A_{c,e} = M_e = \mu_M \pm k_1$. Putting the last estimate into (4.7) we obtain

$$\begin{aligned}
&\mathbb{P}\left(M_e \text{ decreases } | B_i\right) \\
&= \frac{k^2\mu_L^{k-1}}{n^k}\left(1 \pm \frac{2k_1}{\mu_M}\right)\left(1 \pm \frac{k^2}{n}\right)\left(1 \pm \frac{3(k^2+kk_2)}{\mu_L}\right)\left(1 \pm \frac{k_1}{\mu_M}\right) \\
&= \frac{k^2\mu_L^{k-1}}{n^k}\left(1 \pm D'\left(\frac{k_1}{\mu_M} + \frac{k^2+kk_2}{\mu_L}\right)\right)
\end{aligned}$$

for some constant D'.

Now denote by $\mathcal{P}_c = \mathcal{P}_c(i)$ the set of k–element subsets of vertices in $H(i)$ which do not touch an edge of colour c. We have

$$\begin{aligned}
&\mathbb{P}(L_c \text{ decreases } | B_i) = \sum_{e\in\mathcal{P}_c} \mathbb{P}(L_c \text{ decreases } | e_{i+1} = e, B_i)\mathbb{P}(e_{i+1} = e) \\
&= \sum_{e\in\mathcal{P}_c} \frac{1}{M_e}\frac{1}{\binom{n}{k}} = \mu_M^{-1}\left(1 \pm \frac{2k_1}{\mu_M}\right)\binom{L_c}{k}\binom{n}{k}^{-1} \\
&= \frac{\mu_L^k}{\mu_M n^k}\left(1 \pm \frac{2k_1}{\mu_M}\right)\left(1 \pm \frac{3(k^2+kk_2)}{\mu_L}\right) \\
&= \frac{\mu_L^k}{\mu_M n^k}\left(1 \pm D''\left(\frac{k_1}{\mu_M} + \frac{k^2+kk_2}{\mu_L}\right)\right)
\end{aligned}$$

for some constant D''. Here we used (4.4), (4.5), (4.6) and

$$\frac{\mu_L^k}{n^k}\left(1-\frac{kk_2+k^2}{\mu_L}\right) \le \frac{(L_c)_k}{(n)_k} \le \frac{L_c^k}{n^k} \le \frac{\mu_L^k}{n^k}\left(1+\frac{3kk_2}{\mu_L}\right).$$

Setting $D = \max\{D', D''\}$ completes the proof (it is easy to check that we can take, i.e., $D = 30$). □

4.3.2 Differential equations

If M_e and L_c were always concentrated about the functions μ_M and μ_L, respectively, then by Lemma 4.3.2 we would have, informally,

$$\mu_M(i+1) - \mu_M(i) \approx -\frac{k^2 \mu_L(i)^{k-1}}{n^k}$$

and

$$\mu_L(i+1) - \mu_L(i) \approx -\frac{k\mu_L(i)^k}{n^k \mu_M(i)}.$$

We can rescale the "time" i and the random variables. Define new functions f and g by $t = i/N$, $\mu_M(i) = qf(t)$ and $\mu_L(i) = ng(t)$, so that

$$q(f(t+1/N) - f(t)) \approx -\frac{k^2 g(t)^{k-1}}{n}$$
$$n(g(t+1/N) - g(t)) \approx -\frac{kg(t)^k}{qf(t)}$$

or

$$N(f(t+1/N) - f(t)) \approx -kg(t)$$
$$N(g(t+1/N) - g(t)) \approx -\frac{g(t)^k}{f(t)}.$$

The left-hand sides above can be approximated by $f'(t)$ and $g'(t)$ respectively (assuming that f and g are differentiable).

This gives a system of differential equations:

$$\begin{cases} f'(t) = -kg(t)^{k-1} \\ g'(t) = -\frac{g(t)^k}{f(t)} \end{cases}$$

with the initial condition $f(0) = 1$ and $g(0) = 1$.

The solution is $f(t) = (1-t)^k$ and $g(t) = (1-t)$. This argument is not formal, but indicates the possible expressions for μ_M and μ_L. It will be formalized below.

4.3.3 Martingales

For the rest of this chapter we will use $t = t(i) = i/N$ and

$$\mu_M(i) = q(1-t)^k \qquad \mu_L(i) = n(1-t), \tag{4.8}$$

for $i = 0, 1, \ldots, N$. Using these functions yields a simpler estimate for the error of our conditional expectations.

Lemma 4.3.2 *There is a constant $D \geq 1$ such that the following is true. Suppose $i \in \{0, 1, \ldots, N-1\}$, (4.5) holds and let B_i be as in Lemma 4.3.1.*

$$\begin{aligned} &\mathbb{P}\left(M_e(i+1) - M_e(i) = -1 | B_i\right) \\ &= \frac{k^2(1-t)^{k-1}}{n} \pm D\left(\frac{(k^4 + k^3 k_2(i))(1-t)^{k-2}}{n^2} + \frac{k k_1(i)}{N(1-t)}\right) \end{aligned} \tag{4.9}$$

and

$$\mathbb{P}\left(L_c(i+1) - L_c(i) = -k | B_i\right) = \frac{1}{q} \pm D\left(\frac{k_1(i)}{q^2(1-t)^k} + \frac{k + k_2(i)}{N(1-t)}\right).$$

Proof Simply use (4.8) in Lemma 4.3.1. □

Let ω, C, K, ϵ be positive reals, $0 < \epsilon < 1$. Write $N_1 = \lfloor (1-\epsilon)N \rfloor$ and define for $i \in \{0, \ldots, N_1\}$

$$E_1(i) = \frac{kq}{\omega} e^{Ct} \quad \text{and} \quad E_2(i) = \frac{n}{\omega} \frac{e^{Ct}}{(1-t)^k}.$$

For any $i \in [N_1]$ and $e \in \binom{[n]}{k}$ set

$$M_e^-(i) = M_e(i) - \mu_M(i) - E_1(i); \tag{4.10}$$

$$M_e^+(i) = M_e(i) - \mu_M(i) + E_1(i); \tag{4.11}$$

and for each colour $c \in [q]$:

$$L_c^-(i) = L_c(i) - \mu_L(i) - E_2(i); \tag{4.12}$$

$$L_c^+(i) = L_c(i) - \mu_L(i) + E_2(i). \tag{4.13}$$

We will consider a stopping time T_S given by

$$T_S = \sup_i \Big\{ M_e^-(i) \le Kk\sqrt{q \ln n}, M_e^+(i) \ge -Kk\sqrt{q\ln n}$$
$$L_c^-(i) \le K\sqrt{kn\ln q}, L_c^+(i) \ge -K\sqrt{kn\ln q}$$
$$\text{for all } e \in \binom{[n]}{k}, c \in [q] \Big\}.$$

Now for each $e \in \binom{[n]}{k}$ define processes $\{\tilde{M}_e^-\} = \{\tilde{M}_e^-(i), i = 0, \ldots, N_1\}$ and $\{\tilde{M}_e^+\} = \{\tilde{M}_e^+(i), i = 0, \ldots, N_1\}$ where $\tilde{M}_e^{\pm}(0) = 0$ and for $i > 0$

$$\tilde{M}_e^-(i) = \sum_{j=1}^{i} \left(M_e^-(j) - M_e^-(j-1) \right) 1_{j \le T_S};$$
$$\tilde{M}_e^+(i) = \sum_{j=1}^{i} \left(M_e^+(j) - M_e^+(j-1) \right) 1_{j \le T_S}.$$

Also, for each $c \in [q]$ define processes $\{\tilde{L}_c^-\} = \{\tilde{L}_c^-(i), i = 0, \ldots, N_1\}$ and $\{\tilde{L}_c^+\} = \{\tilde{L}_c^+(i), i = 0, \ldots, N_1\}$ where $\tilde{L}_c^{\pm}(0) = 0$ and for $i > 0$

$$\tilde{L}_c^-(i) = \sum_{j=1}^{i} \left(L_c^-(j) - L_c^-(j-1) \right) 1_{j \le T_S};$$
$$\tilde{L}_c^+(i) = \sum_{j=1}^{i} \left(L_c^+(j) - L_c^+(j-1) \right) 1_{j \le T_S}.$$

Lemma 4.3.3 *Suppose that $2k^2/n \le \epsilon < 1$, $K \ge 1$, $(kN)/n \ge 3$, $C = 8Dk\epsilon^{-2}$, where D is the constant as in Lemma 4.3.2, and ω is such that*

$$4ke^C\epsilon^{-(k+1)} \le \omega \le K^{-1} \min\left(\frac{q}{\ln n}, \frac{n}{k \ln q} \right)^{1/2}.$$

Then for any $e \in \binom{[n]}{k}$ and any $c \in [q]$ the processes $\{\tilde{M}_e^-\}, \{\tilde{L}_c^-\}$ are supermartingales and $\{\tilde{M}_e^+\}, \{\tilde{L}_c^+\}$ are submartingales.

Proof Suppose $i \le \min(T_S, N_1)$. Then

$$M_e(i) = \mu_M(i) \pm k_1(i) \quad \forall e \in \binom{[n]}{2},$$
$$L_c(i) = \mu_L(i) \pm k_2(i) \quad \forall c \in [q],$$

where

$$k_1(i) = E_1(i) + Kk\sqrt{q \ln n},$$
$$k_2(i) = E_2(i) + K\sqrt{kn \ln q}.$$

Write $f(\tau) = e^{C\tau}$ and $g(\tau) = e^{C\tau}/(1-\tau)^k$. We have $k \geq 2$, so both f and g are increasing for $\tau \in [0, 1)$. Since $\omega \leq K^{-1}\sqrt{q/\ln n}$ and $\omega \leq K^{-1}\sqrt{n/(k \ln q)}$ we have

$$k_1(i) \leq 2E_1(i); \tag{4.14}$$
$$k_2(i) \leq 2E_2(i). \tag{4.15}$$

We claim that (4.5) is satisfied for each $i \in [0, N(1-\epsilon)]$. Indeed, since $\epsilon \geq 2k^2/n$, $\omega \geq 4ke^C\epsilon^{-(k+1)}$ and $g(\tau), f(\tau)$ are increasing

$$\frac{k_1(i)}{\mu_M(i)} \leq \frac{2E_1(i)}{q(1-t)^k} = \frac{2kg(t)}{\omega} \leq \frac{2ke^C}{\omega\epsilon^k} \leq \frac{1}{2};$$
$$\frac{kk_2(i)}{\mu_L(i)} \leq \frac{2kE_2(i)}{n(1-t)} \leq \frac{2kg(t)}{\omega(1-t)} \leq \frac{2ke^C}{\omega\epsilon^{k+1}} \leq \frac{1}{2};$$
$$\frac{k^2}{\mu_L(i)} = \frac{k^2}{n(1-t)} \leq \frac{k^2}{\epsilon n} \leq \frac{1}{2}.$$

So $k_1(i) \leq \mu_M(i)/2$ and $k^2 + kk_2(i) \leq \mu_L(i)$ as required.

We will first show that $\{\tilde{M}_e^+\}$ a submartingale and $\{\tilde{M}_e^-\}$ is a supermartingale. Since the increments are zero for $i > T_S$, it suffices to prove that

$$\mathbb{E}\left(\tilde{M}_e^-(i+1) - \tilde{M}_e^-(i)|B_i\right) \leq 0, \qquad \mathbb{E}\left(\tilde{M}_e^+(i+1) - \tilde{M}_e^+(i)|B_i\right) \geq 0$$

for each block (or atom) B_i of the partition corresponding to $\mathcal{F}_i$, where $T_S \geq i$. On such B_i, (4.3) and (4.4) hold by the definition of T_S.

Write

$$R_1(i) = D\left(\frac{(k^4 + k^3k_2(i))(1-t)^{k-2}}{n^2} + \frac{kk_1(i)}{N(1-t)}\right).$$

By Lemma 4.3.2

$$\mathbb{E}\left(\tilde{M}_e^+(i+1) - \tilde{M}_e^+(i)|B_i\right)$$
$$= \left(-\frac{k^2(1-t)^{k-1}}{n} \pm R_1(i)\right) + (\mu_M(i) - \mu_M(i+1)) + (E_1(i+1) - E_1(i)).$$

Some of the terms cancel out:

$$\begin{aligned}
&-\frac{k^2(1-t)^{k-1}}{n}+\mu_M(i)-\mu_M(i+1)\\
&=-\frac{k^2}{n}\left(1-\frac{i}{N}\right)^{k-1}+q\left(\left(1-\frac{i}{N}\right)^k-\left(1-\frac{i+1}{N}\right)^k\right)\\
&\geq-\frac{k^2\left(1-\frac{i}{N}\right)^{k-1}}{n}+\frac{kN}{n}\left(1-\frac{i}{N}\right)^k\left(1-\left(1-\frac{1}{N-i}\right)^k\right)\\
&\geq-\frac{k^2(1-\frac{i}{N})^{k-1}}{n}+\frac{kN}{n}\left(1-\frac{i}{N}\right)^k\left(\frac{k}{N-i}-\frac{k^2}{2(N-i)^2}\right)\\
&\geq-\frac{k^3}{Nn}\geq-\frac{k}{N}. \qquad (4.16)
\end{aligned}$$

Similarly, since $q=\lceil\frac{kN}{n}\rceil\leq\frac{kN}{n}+1$

$$-\frac{k^2\mu_L(i)^{k-1}}{n^k}+\mu_M(i)-\mu_M(i+1)\leq\frac{k}{N}.$$

We will also need later that since $\frac{k}{N}\leq\frac{k^2}{3n}<1$ we have

$$\mu_M(i)-\mu_M(i+1)<\frac{2k^2}{n}. \qquad (4.17)$$

Therefore

$$\begin{aligned}
&\mathbb{E}\left(M_e^-(i+1)-M_e^-(i)|B_i\right)\leq-(E_1(i+1)-E_1(i))+\frac{k}{N}+R_1(i);\\
&\mathbb{E}\left(M_e^+(i+1)-M_e^+(i)|B_i\right)\geq(E_1(i+1)-E_1(i))-\frac{k}{N}-R_1(i).
\end{aligned}$$

It remains to verify that

$$\frac{k}{N}+R_1(i)\leq E_1(i+1)-E_1(i). \qquad (4.18)$$

Since $f''(t)>0$ we have

$$\begin{aligned}
E_1(i+1)-E_1(i)&=\frac{kq}{\omega}\left(f\left(t+\frac{1}{N}\right)-f(t)\right) \qquad (4.19)\\
&\geq\frac{qkf'(t)}{\omega N}=\frac{Cqkf(t)}{\omega N}\geq\frac{Ck^2f(t)}{\omega n}.
\end{aligned}$$

Now, firstly,

$$\frac{k}{N}\leq\frac{1}{4}(E_1(i+1)-E_1(i)).$$

This is because by (4.19) the ratio of the left side and the right is at most

$$\frac{k}{N}\frac{4\omega N}{Cqkf(t)} \leq \frac{4\omega}{Cqf(0)} \leq \frac{4}{q^{1/2}C} \leq 1.$$

Here we used that $\omega \leq q^{1/2}$ and $C > 8$ from the assumption of the lemma. Secondly, for the first term of $R_1(i)$ we have

$$\frac{Dk^4(1-t)^{k-2}}{n^2} \leq \frac{1}{4}(E_1(i+1) - E_1(i)).$$

To see this, note that by (4.19), the ratio of the left and the right side expressions is at most

$$\frac{4Dk^2(1-t)^{k-2}\omega}{nCe^{Ct}} \leq \frac{4Dk^2\omega}{nC} \leq \frac{k\omega}{2n} \leq \frac{1}{2K}\sqrt{\frac{k}{n\ln q}} \leq 1.$$

Here we used the facts that $e^{C\tau}/(1-\tau)^{k-2}$ is increasing for $\tau \in [0,1)$ and the earlier assumptions about n, q, k, C and ω.

Thirdly,

$$\frac{Dk^3k_2(i)(1-t)^{k-2}}{n^2} \leq \frac{1}{4}(E_1(i+1) - E_1(i)).$$

Indeed, by (4.19), the ratio of the left and the right side is at most

$$\frac{4Dkk_2(i)\omega(1-t)^{k-2}}{nCe^{Ct}} \leq \frac{8DkE_2(i)\omega(1-t)^{k-2}}{nCe^{Ct}} \leq \frac{8Dk}{C(1-t)^2} \leq 1.$$

Here in the second inequality we used (4.15). Finally, for the last term of $R_1(i)$

$$\frac{Dkk_1(i)}{N(1-t)} \leq \frac{1}{4}(E_1(i+1) - E_1(i)).$$

This is because the ratio of the left side and the right by (4.19) and (4.14) is at most

$$\frac{4Dk_1(i)\omega}{Cqe^{Ct}(1-t)} \leq \frac{8Dkqe^{Ct}}{Cqe^{Ct}(1-t)} \leq \epsilon \leq 1.$$

Now (4.18) follows by combining bounds for each of the five terms of $2k/N + R_1(i)$.

Let us now show that $\{\tilde{L}_c^-\}$ is a supermartingale. Again it suffices to consider only the blocks B_i for which $i \le \min(T_S, N_1)$. By Lemma 4.3.2

$$\begin{aligned}
&\mathbb{E}\left(\tilde{L}_c^-(i+1) - \tilde{L}_c^-(i)|B_i\right) \\
&= \mathbb{E}\left(L_c(i+1) - L_c(i)|B_i\right) + (\mu_L(i) - \mu_L(i+1)) + (E_2(i) - E_2(i+1)) \\
&= -\frac{k\mu_L(i)^k}{\mu_M(i)n^k} \pm R_2(i) + (\mu_L(i) - \mu_L(i+1)) + (E_2(i) - E_2(i+1)),
\end{aligned}$$

where

$$R_2(i) = D\left(\frac{kk_1(i)}{q^2(1-t)^k} + \frac{k^2 + kk_2(i)}{N(1-t)}\right)$$

Now

$$\begin{aligned}
&-\frac{k\mu_L(i)^k}{\mu_M(i)n^k} + \mu_L(i) - \mu_L(i+1) \\
&= -\frac{k}{q} + n\left(\left(1 - \frac{i}{N}\right) - \left(1 - \frac{i+1}{N}\right)\right) \\
&= -\frac{k}{q} + \frac{n}{N} \in \left[0, \frac{2k}{q^2}\right],
\end{aligned} \tag{4.20}$$

since $\frac{kN}{n} \ge 3$ and $\frac{1}{q} \le \frac{n}{kN} \le \frac{1}{q}\left(1 + \frac{2}{q}\right)$. Therefore

$$\mathbb{E}\left(\tilde{L}_c^-(i+1) - \tilde{L}_c^-(i)|B_i\right) = E_2(i) - E_2(i+1) \pm \left(R_2(i) + \frac{2k}{q^2}\right).$$

We need to show that the above quantity is non-positive. Since $g''(t) > 0$ for $t \in [0,1)$, we have similarly as in (4.19)

$$E_2(i+1) - E_2(i) \ge \frac{g'(t)n}{N\omega} \ge \frac{g'(t)k}{q\omega} \ge \frac{Cg(t)k}{q\omega}. \tag{4.21}$$

We have

$$\frac{Dkk_1(i)}{q^2(1-t)^k} \le \frac{1}{4}(E_2(i+1) - E_2(i)),$$

since by (4.14) and (4.21) the ratio of the left and the right side is at most

$$\frac{4Dk}{q^2(1-t)^k}\frac{2kqf(t)}{\omega}\frac{q\omega}{Cg(t)k} = \epsilon^2 \le 1.$$

Let us now check that

$$\frac{k^2 D}{N(1-t)} \le \frac{1}{4}(E_2(i+1) - E_2(i)).$$

Indeed, the ratio of the left and the right side is by (4.21) at most

$$\frac{4k^2 D}{N(1-t)} \frac{q\omega}{Cg(t)k} \leq \frac{q\omega}{2N} \leq \frac{1}{K}\sqrt{\frac{k}{n \ln q}} \leq 1.$$

The first inequality follows, among others, because $(1-\tau)g(\tau)$ is increasing for $\tau \in [0,1)$. Next,

$$\frac{Dkk_2(i)}{N(1-t)} \leq \frac{1}{4}(E_2(i+1) - E_2(i)).$$

Indeed, by (4.15) and (4.21) the ratio of the left and the right side is at most

$$\frac{4Dk}{N(1-t)} \frac{2ng(t)}{\omega} \frac{q\omega}{Cg(t)k} \leq \epsilon \leq 1.$$

Finally,

$$\frac{2k}{q^2} \leq \frac{1}{4}\left(E_2(i+1) - E(i)\right).$$

Here, again, the ratio of the two sides is by (4.21) at most

$$\frac{2k}{q^2} \frac{4q\omega}{Cg(t)k} \leq \frac{\omega}{qDk} \leq 1$$

since $\omega \leq q^{1/2}$.

We have shown that

$$R_2(i) + \frac{2k}{q^2} \leq E_2(i+1) - E_2(i) \tag{4.22}$$

and

$$\mathbb{E}\left(\tilde{L}_c^-(i+1) - \tilde{L}_c^-(i)|B_i\right) \leq 0$$

as required, so $\{\tilde{L}_c^-\}$ is a supermartingale. The proof that $\{\tilde{L}_c^+\}$ is a submartingale is similar. □

4.3.4 Applying concentration results

Let us prove the main lemma of this paper.

Lemma 4.3.4 *Suppose $\epsilon \in (0,1)$, D is the constant from Lemma 4.3.2, $K = 8D^{1/2}$, $C = 8Dk\epsilon^{-2}$ and*

$$2k^2 e^C \epsilon^{-(k+2)} < \omega < K^{-1} \min\left(\frac{q}{\ln n}, \frac{n}{k \ln q}\right)^{1/2}.$$

Then

$$\mathbb{P}(T_S < \lfloor(1-\epsilon)N\rfloor) \le \frac{2}{q} + \frac{2}{n}. \tag{4.23}$$

Proof The condition of the lemma implies that $2k^2/n < \epsilon^{k+2}n^{-1/2} < \epsilon$ and $q > k^2 \ge 4$. Therefore by Lemma 4.3.3 $\{M_e^-\}, \{L_c^-\}$ are supermartingales and $\{M_e^+\}, \{L_c^+\}$ are submartingales for every $e \in \binom{[n]}{k}$ and $c \in [q]$.

Recall that $N_1 = \lfloor(1-\epsilon)N\rfloor$. We shall apply a martingale concentration result, Theorem 3.15 of [72] to show that for any $e \in \binom{[n]}{k}$ and $c \in [q]$:

$$\begin{aligned}
&\mathbb{P}(\tilde{M}_e^-(N_1) \ge Kk\sqrt{q\ln n}) \le \frac{1}{n^{k+1}};\\
&\mathbb{P}(\tilde{M}_e^+(N_1) \le -Kk\sqrt{q\ln n}) \le \frac{1}{n^{k+1}};\\
&\mathbb{P}(\tilde{L}_c^-(N_1) \ge K\sqrt{kn\ln q}) \le \frac{1}{q^2};\\
&\mathbb{P}(\tilde{L}_c^+(N_1) \le -K\sqrt{kn\ln q}) \le \frac{1}{q^2}.
\end{aligned}$$

Notice that if $T_S < N_1$ then at least one of the above events occurs (for some e or for some c), since all the processes remain "frozen" after the time T_S. Since there are $\binom{n}{k} \le n^k$ sets $e \in \binom{[n]}{k}$ and q colours c, the above inequalities and the union bound imply (4.23).

We will modify the submartingales and supermartingales slightly to turn them into martingales.

Set $\beta = 8D + 1$. Then

$$\omega > 2k^2e^C\epsilon^{-(k+2)} > 2kCe^C\epsilon^{-k}\beta^{-1}.$$

Notice that the bounds on ω in the assumption imply

$$\left(\frac{\ln n}{q}\right)^{1/2} \le \frac{1}{k^2K} \quad \text{and} \quad \left(\frac{k\ln q}{n}\right)^{1/2} \le \frac{1}{k^2K}. \tag{4.24}$$

For $i = 0, \ldots, N_1 - 1$ define random variables

$$Z_M^-(i+1) = -\mathbb{E}\,(\tilde{M}_e^-(i+1) - \tilde{M}_e^-(i)|\mathcal{F}_i).$$

Since $\{\tilde{M}_e^-\}$ is a supermartingale, $Z_M^-(i) \ge 0$. Also, since $\{\tilde{M}_e^+\}$ is a submartingale by Lemma 4.3.3

$$\mathbb{E}\,(\tilde{M}_e^-(i+1) - \tilde{M}_e^-(i)|\mathcal{F}_i) + 2(E_1(i+1) - E_1(i)) = \mathbb{E}\,(\tilde{M}_e^+(i+1) - \tilde{M}_e^+(i)|\mathcal{F}_i) \ge 0.$$

Therefore

$$0 \leq Z_M^-(i+1) \leq 2(E_1(i+1) - E_1(i)). \tag{4.25}$$

The sequence $\{\hat{M}_e^-\} = \{\hat{M}_e^-(i), i = 0, \ldots, N_1\}$ where

$$\hat{M}_e^-(i) = \tilde{M}_e^-(i) + \sum_{j=1}^{i} Z_M^-(j)$$

is a martingale (here and below we define the sum when j ranges from 1 to 0 to be equal to 0) and

$$\mathbb{P}\left(\tilde{M}_e^-(N_1) \geq Kk\sqrt{q \ln n}\right) \leq \mathbb{P}\left(\hat{M}_e^-(N_1) \geq Kk\sqrt{q \ln n}\right).$$

Let us estimate the maximum positive increment and the maximum conditional variance of $\{\hat{M}_e^-\}$ (see definitions before Theorem 3.15 in [72]).

Notice that since $(kN)/n \geq 3$, we have $n/(kN) \leq 3/(2q)$. Let $f(\tau), g(\tau)$ be as in the proof of Lemma 4.3.3. Since $f'(\tau), g'(\tau)$ are increasing, the following inequalities hold for $i = 0, \ldots, N_1 - 1$:

$$E_1(i+1) - E_1(i) \leq \frac{f'(1-\epsilon)kq}{\omega N} \leq \frac{k^2}{n}\frac{Ce^C}{\omega} \leq \frac{\beta k}{n} \leq \frac{\beta k^2}{n}; \tag{4.26}$$

$$E_2(i+1) - E_2(i) \leq \frac{g'(1-\epsilon)n}{\omega N} \leq \frac{3k}{2q\omega}\left(\frac{Ce^C}{\epsilon^k} + \frac{ke^C}{\epsilon^{k+1}}\right) \leq \frac{\beta}{q}. \tag{4.27}$$

We have $\hat{M}_e^-(i+1) - \hat{M}_e^-(i) = 0$ for $i > T_S$. If $i \leq T_S$, using (4.17), (4.25) and (4.26) the positive increment of $\{\hat{M}_e^-\}$ is

$$\begin{aligned} &M_e(i+1) - M_e(i) + \mu_M(i) - \mu_M(i+1) \\ &+ E_1(i) - E_1(i+1) + Z_M^-(i+1) \\ &\leq \frac{2k^2}{n} + \frac{\beta k^2}{n} \leq \frac{k^2(2+\beta)}{n}. \end{aligned}$$

Let R_1, R_2 be as in the proof of Lemma 4.3.3. For $i \in \{0, \ldots, N_1 - 1\}$, the conditional variance of the differences of $\{\hat{M}_e^-\}$ is

$$\begin{aligned} Var(\hat{M}_e^-(i+1) - \hat{M}_e^-(i)|\mathcal{F}_i) &\leq \max_{B_i} Var(M_e(i+1) - M_e(i)|B_i) \\ &\leq \max_{B_i} \mathbb{P}(M_e(i+1) - M_e(i) = -1|B_i) \\ &\leq \frac{k^2(1-t)^{k-1}}{n} + R_1(i) \leq \frac{k^2(1+\beta)}{n}. \end{aligned}$$

Here B_i ranges over all blocks (atoms) of the partition corresponding to $\mathcal{F}_i$; to get the third line we used Lemma 4.3.2 and to get the fourth line we used (4.18) and (4.26). Thus the maximum sum of conditional variances $\hat{v}$ for M satisfies

$$\hat{v} \leq \frac{Nk^2(1+\beta)}{n} \leq kq(1+\beta)$$

and the maximum conditional positive deviation b satisfies

$$b \leq \frac{k^2(2+\beta)}{n} \leq 2+\beta.$$

Theorem 3.15 of [72] states that for any $t \geq 0$

$$\mathbb{P}(\hat{M}_e^-(N_1) > t) \leq \exp\left(-\frac{t^2}{2\hat{v}(1+\frac{bt}{3\hat{v}})}\right).$$

Therefore

$$\begin{aligned}
\mathbb{P}(\hat{M}_e^-(N_1) \geq Kk\sqrt{q\ln n}) &\leq \exp\left(-\frac{k^2K^2 q\ln n}{2kq(1+\beta)+\frac{2}{3}(2+\beta)Kk\sqrt{q\ln n}}\right)\\
&= \exp\left(-\frac{kK^2\ln n}{2(1+\beta)+\frac{2}{3}(2+\beta)k^{-2}}\right)\\
&\leq e^{-\frac{kK^2\ln n}{3(1+\beta)}} \leq e^{-(k+1)\ln n}
\end{aligned}$$

by (4.24) since $k \geq 2$, $D \geq 1$ and $K = 8D^{1/2} > (9+36D)^{1/2}$.

Now consider the submartingale $\{\tilde{M}_e^+\}$ and define for $i = 0, \ldots, N_1 - 1$

$$Z_M^+(i+1) = -\mathbb{E}\,(\tilde{M}_e^+(i+1) - \tilde{M}_e^+(i)|\mathcal{F}_i).$$

Since $\tilde{M}_e^+$ is a submartingale and $\tilde{M}_e^-$ is a supermartingale we have

$$\mathbb{E}\,(\tilde{M}_e^+(i+1) - \tilde{M}_e^+(i)|\mathcal{F}_i) - 2(E_1(i+1) - E_1(i)) \leq 0$$

so

$$2(E_1(i) - E_1(i+1)) \leq Z_M^+(i+1) \leq 0. \tag{4.28}$$

The sequence

$$\hat{M}_e^+ = \{\tilde{M}_e^+(i) + \sum_{j=1}^{i} Z_M^+(j), \quad i = 0, 1, \ldots, N_1\}$$

is a martingale and

$$\mathbb{P}\left(\tilde{M}_e^+(N_1) \leq -Kk\sqrt{q \ln n}\right) \leq \mathbb{P}\left(-\hat{M}_e^+(N_1) \geq Kk\sqrt{q \ln n}\right).$$

Using (4.17), (4.26) and (4.28) the difference $-\hat{M}_e^-(i+1) - (-\hat{M}_e^-(i))$ is

$$\begin{aligned}
&M_e(i) - M_e(i+1) + \mu_M(i+1) - \mu_M(i) + E_1(i) - E_1(i+1) - Z_M(i+1) \\
&\leq 1 + E_1(i+1) - E_1(i) \leq 1 + \frac{k^2(1+\beta)}{n} \leq 2 + \beta.
\end{aligned}$$

Furthermore, the conditional variance of $-\hat{M}_e^+(i+1) + \hat{M}_e^+(i)$ is the same as the conditional variance of $M_e^-(i+1) - M_e^-(i)$ so $\hat{v} \leq kq(1+\beta)$. Now Theorem 3.15 of [72] yields

$$\mathbb{P}\left(-\hat{M}_e^+(N_1) \leq Kk\sqrt{q \ln n}\right) \leq e^{-(k+1)\ln n}$$

by the same calculation as in the corresponding bound for $\hat{M}_e^+$.

Now consider the supermartingale $\{\tilde{L}_c^-\}$. Similarly as above, define a martingale $\{\hat{L}_c^-\} = \{\hat{L}_c^-(i), i = 0, \ldots, N_1\}$, where

$$\hat{L}_c^-(i) = \tilde{L}_c^-(i) + \sum_{j=1}^{i} Z_L^-(j)$$

and $Z_L^-(i+1) = -E(\tilde{L}_c^-(i+1) - \tilde{L}_c^-(i)|\mathcal{F}_i)$.

Using the fact that $L_c^+(i) = L_c^-(i) + 2E_2(i)$ we obtain similarly as above

$$0 \leq Z_L^-(i) \leq 2(E_2(i+1) - E_2(i)).$$

Therefore using (4.20) and (4.27) the difference $\hat{L}_c^-(i+1) - \hat{L}_c^-(i)$ is

$$\begin{aligned}
&L_e(i+1) - L_e(i) + \mu_L(i) - \mu_L(i+1) + E_2(i+1) - E_2(i) \\
&\leq \frac{k}{q} + \frac{2k}{q^2} + \frac{\beta}{q} \leq \frac{k(1+\beta)}{q}.
\end{aligned}$$

Here we used $2k/q \leq (k-1)\beta$, since $k \geq 2, q \geq 3$ and $\beta > 8$. Now using Lemma 4.3.2 and the bounds (4.22), (4.27) for $i \in \{0, \ldots, N_1 - 1\}$ we get

$$
\begin{aligned}
&Var(\hat{L}_c^-(i+1) - \hat{L}_c^-(i)|\mathcal{F}_i) \leq \max_{B_i} Var(L_c(i+1) - L_c(i)|B_i) \\
&\leq k^2 \cdot \max_{B_i} \mathbb{P}(L_c \text{ decreases at step } i+1|B_i) \\
&\leq \frac{1}{q} + k^2 R_2(i) \leq \frac{k^2(1+\beta)}{q} \leq k(1+\beta).
\end{aligned}
$$

Hence the maximum sum of conditional variances $\hat{v}$ satisfies

$$
\hat{v} \leq \frac{k^2(1+\beta)N}{q} \leq (1+\beta)kn. \tag{4.29}
$$

So by Theorem 3.15 of [72], (4.24) and our choice of K

$$
\begin{aligned}
&\mathbb{P}\left(\tilde{L}_c^-(N_1) \geq K\sqrt{kn\ln q}\right) \leq \mathbb{P}\left(\hat{L}_c^-(N_1) \geq K\sqrt{kn\ln q}\right) \\
&\leq \exp\left(-\frac{K^2 kn \ln q}{2(1+\beta)kn + \frac{2}{3}k(1+\beta)K\sqrt{kn\ln q}}\right) \\
&\leq \exp\left(-\frac{K^2 \ln q}{2(1+\beta) + \frac{2}{3}(1+\beta)k^{-1}}\right) \leq \exp\left(-\frac{K^2 \ln q}{3(1+\beta)}\right) \leq e^{-2\ln q}.
\end{aligned}
$$

Finally, let us bound in exactly the same way the probability that $\{\tilde{L}_c^+\}$ ever attains a large negative value. Define for a martingale $\{\hat{L}_c+\} = \{\hat{L}_c^+(i), i = 0, \ldots, N_1\}$, where

$$
\hat{L}_c^+(i) = \tilde{L}_c^+(i) + \sum_{j=1}^{i} Z_L^+(j)
$$

and $Z_L^+(i+1) = -E(\tilde{L}_c^+(i+1) - \tilde{L}_c^+(i)|\mathcal{F}_i)$.

As before, by Lemma 4.3.3:

$$
2(E_2(i) - E_2(i+1)) \leq Z_L^+(i+1) \leq 0.
$$

Using (4.27) we get that the difference $(-\hat{L}_c^+(i+1) - (-\hat{L}_c^+(i))$ is at most

$$
\begin{aligned}
&L_c(i) - L_c(i+1) - \mu_L(i) + \mu_L(i+1) - E_2(i) + E_2(i+1) - Z_L^+(i+1) \\
&\leq k + \frac{\beta}{q} \leq k(1+\beta).
\end{aligned}
$$

The conditional variance of $-\hat{L}_c^+(i+1) + \hat{L}_c^+(i)$ is the same as the conditional variance of $\hat{L}_c^-(i+1) - \hat{L}_c^-(i)$, so the estimate (4.29) still holds.

Once again applying Theorem 3.15 of [72]:

$$\mathbb{P}(\tilde{L}_c^+(N_1) \leq -K\sqrt{kn\ln q}) \leq \mathbb{P}(-\hat{L}_c^+(N_1) \geq K\sqrt{kn\ln q}) \leq e^{-2\ln q},$$

as in the corresponding bound for $\{\hat{L}_c^-\}$. □

Proof of Theorem 4.1.2 We may assume that $\epsilon < 1$. Let ϵ' be such that $\frac{1}{1-\epsilon'} = 1 + \frac{\epsilon}{2}$, let $D \geq 1$ be the constant from Lemma 4.3.2 and define

$$C(\epsilon, k) = 16D^{1/2}k^3 e^{8Dk/\epsilon'^2}\epsilon'^{-(k+2)}.$$

Let $c_\epsilon > 0$ be a constant such that for all $k \geq 2$ we have $C(\epsilon, k)^2 < \frac{1}{2}e^{k/c_\epsilon}$ and observe that this implies $c_\epsilon \leq \epsilon'/8$. Suppose (4.1) hold. Then $n \geq 3$, $\bar{d} \geq 4$ and $\frac{k}{n} \leq \frac{\epsilon'}{2} \leq \frac{\epsilon}{2}$. Let

$$N' = \left\lceil \frac{N}{1-\epsilon'} \right\rceil \quad \text{and} \quad q = \left\lceil \frac{kN'}{n} \right\rceil.$$

Then,

$$\bar{d} \leq q \leq \lceil \bar{d}(1+\epsilon/2) + k/n \rceil \leq \lceil (1+\epsilon)\bar{d} \rceil \leq 3\bar{d} \leq \bar{d}^2,$$

and using the definition of c_ϵ and (4.1)

$$C(\epsilon, k)^2 < \frac{1}{2}e^{k/c_\epsilon} \leq \frac{1}{2}\min\left(\frac{n}{\ln \bar{d}}, \frac{\bar{d}}{\ln n}\right) \leq \min\left(\frac{n}{\ln q}, \frac{q}{\ln n}\right).$$

Therefore Lemma 4.3.4 applies for the random colouring process described in Section 4.3.3 with N' random hyperedges, q colours and n vertices. By that lemma, the probability that the process hits the stopping time until step $i = N - 1 \leq \lfloor N'(1-\epsilon') \rfloor$ is at most $\frac{2}{q} + \frac{2}{n}$. If $T_S \geq N - 1$, then, by the estimates from the proof of Lemma 4.3.3, and the definition of the process

$$M_e(N-1) \geq \mu_M(N-1) - (E_1(N-1) + Kk\sqrt{q\ln n}) \geq \frac{1}{2}\mu_M(N-1) > 0,$$

for any edge $e \in \binom{[n]}{k}$. So the N-th edge can be coloured successfully (which means all of the previous edges have been coloured successfully as well). The claim follows by a trivial observation, that the random hypergraph obtained after adding the first N out of N' edges has distribution exactly $\mathbb{H}^{(k)}(n, N)$. □

Part II

Random graphs with few disjoint excluded minors

Chapter 5

Background

5.1 Notation and main definitions

In this part calligraphic letters such as $\mathcal{A}, \mathcal{B}, \ldots$ will denote classes of objects, mostly classes of labelled graphs or graphs where some vertices are distinguished and/or unlabelled; these classes will always be closed under isomorphism. We denote by $\mathcal{A}_n, \mathcal{B}_n, \ldots$ the respective classes restricted to objects (graphs) of size n with labels $[n] = \{1, \ldots, n\}$. The exponential generating function for $\mathcal{A}$ is defined as

$$A(x) = \sum_{n=0}^{\infty} \frac{|\mathcal{A}_n|}{n!} x^n,$$

and we will use the notation $B(x)$ for the exponential generating function of $\mathcal{B}$, etc. If $A(x)$ is an infinite power series, then $[x^n]A(x)$ denotes its coefficient at x^n.

Given a set of graphs $\mathcal{B}$ and a graph G, a set $Q \subseteq V(G)$ is called a $\mathcal{B}$-blocker (or a $\mathcal{B}$-minor-blocker) if $G - Q \in \operatorname{Ex} \mathcal{B}$, i.e., $G - Q$ has no minor in the set $\mathcal{B}$. For a positive integer k, we denote by $k\mathcal{B}$ the class of graphs consisting of k vertex disjoint copies of graphs in $\mathcal{B}$ (with repetitions allowed). Thus $\operatorname{Ex}(k+1)\mathcal{B}$ is the class of graphs that do not have $k+1$ vertex disjoint subgraphs $H_1, \ldots, H_{k+1}$, each with a minor in $\mathcal{B}$. For a graph H, we will abbreviate $\operatorname{Ex}\{H\}$ to $\operatorname{Ex} H$, "$\{H\}$-blocker" to "H-blocker", etc.

Given a class of graphs $\mathcal{A}$, $\operatorname{apex}^k \mathcal{A}$ denotes the class of all graphs such that by deleting at most k vertices we may obtain a graph in $\mathcal{A}$; we use $\operatorname{apex} \mathcal{A}$ for $\operatorname{apex}^1 \mathcal{A}$.

Also, given a positive integer s, call a graph G an *s-fan* if G is a union of a complete bipartite graph with parts A and B, so that $|A| = s$, and a path P with $V(P) = B$. We call 1-fans simply *fans*.

A class of graphs is called *proper*, if it is not the class of all graphs.

5.2 Graph minors

5.2.1 Excluded minors and treewidth

We provide some additional definitions and results for graph minors. Many of the definitions are from Diestel's book "Graph Theory" [38].

Let G be a graph, T a tree, and let $\mathcal{V} = (V_t)_{t\in T}$ be a family of vertex sets $V_t \subseteq V(G)$ indexed by the vertices t of T. The pair $(T, \mathcal{V})$ is called a *tree-decomposition* of G if it satisfies the following three conditions:

(1) $V(G) = \cup_{t\in T} V_t$;

(2) for every edge $e \in E(G)$ there exists a $t \in V(T)$ such that both ends of e lie in V_t;

(3) $V_{t_1} \cap V_{t_3} \subseteq V_{t_2}$ whenever t_2 is on the path from t_1 to t_3 in T.

The *width* of $(T, \mathcal{V})$ is defined as $\max\{|V_t| - 1 : t \in V(T)\}$. The *treewidth* $\mathrm{tw}(G)$ of G is the least width of any decomposition of G. Treewidth measures how much G is "tree-like".

In the introduction we mentioned the classical result of Robertson and Seymour: each minor-closed class $\mathcal{A}$ can be characterised as $\mathcal{A} = \mathrm{Ex}\,\mathcal{B}$ where $\mathcal{B}$ is a finite set of *minimal excluded minors* for $\mathcal{A}$. We state two other important results from the theory of graph minors.

Theorem 5.2.1 *(Robertson and Seymour, 1986) Given a graph H, the graphs without H as a minor have bounded treewidth if and only if H is planar.*

It is well known that the class of graphs with treewidth at most 1 is the class of forests and the class of graphs with treewidth at most 2 is $\mathrm{Ex}\,K_4$. The other result is the generalisation of Erdős and Pósa theorem. (Our version of the proof is provided in Section 7.3.2.)

Theorem 5.2.2 *(Robertson and Seymour, 1986, [88]) Let $\mathcal{B}$ be a set of connected graphs and let k be a positive integer. The following statements are equivalent.*

- *There is a constant $c = c(k, \mathcal{B})$, such that each graph $G \in \mathrm{Ex}\,(k+1)\mathcal{B}$ has a $\mathcal{B}$-blocker of size at most c.*
- *$\mathcal{B}$ contains a planar graph.*

Classes of graphs with bounded treewidth play a prominent role in computer science. For many classical NP-complete graph problems, instances with constant treewidth can be solved in polynomial time, typically using a dynamic programming algorithm. Such problems include, for example, graph colouring and finding the maximum independent set. Similar questions are part of the domain of the rich area of parameterized complexity, see, e.g., [42].

5.2.2 Normal trees

A theorem of Kloks stated in [33] gives an equivalent definition of treewidth, which we find very useful in Chapter 7 and Chapter 8.

Let G be a graph, and let T be a rooted tree on the same vertex set $V(G)$, with root vertex r. (We do not insist that T is a subgraph of G.) The tree T induces a *tree-ordering* $\leq_T$ on $V(G)$, where $u \leq_T v$ if and only if u is on the path from r to v in T. T is a *normal* tree for G if u and v are comparable for every edge uv of G. We think of the tree T as hanging down from its root. We will say that u is above v (and v is below u) in T if $u <_T v$. Think of the graph G as fixed. Given a normal tree T for G, for each vertex v of G we define its set $\mathrm{AA}_T(v)$ of *active ancestors* by

$$\mathrm{AA}_T(v) = \{u <_T v : \ \exists z \geq_T v \text{ with } uz \in E(G)\}.$$

For brevity we write $a_T(v) = |\mathrm{AA}_T(v)|$.

Theorem 5.2.3 (Kloks) *The treewidth $tw(G)$ of a graph G satisfies*

$$tw(G) = \min_{T \in \mathcal{T}} \max_{v \in V(G)} a_T(v) \tag{5.1}$$

where $\mathcal{T}$ is the set of all normal trees for G.

We refer to [38] for more facts on treewidth (though note that in that book a normal tree in a graph is required to be a subgraph). We prove the theorem in Section 7.3.1.

5.3 Enumerating graphs from minor-closed classes

5.3.1 Planar graphs

Like the fundamental work of Robertson and Seymour on graph minors had its roots in the works of Kuratowski and Wagner on planar graphs, the motivating results on enumeration of minor-closed classes concern the class of planar graphs.

A theory for enumerating maps (3-connected planar graphs) was developed by Tutte in the 1960s, see [52]. Bender, Gao and Wormald [9] applied these results for counting 2-connected planar graphs. In 2005, Giménez and Noy finally obtained the asymptotic formula for the class $\mathcal{P} = \mathrm{Ex}\{K_{3,3}, K_5\}$ of planar graphs

$$|\mathcal{P}_n| \sim g n^{-7/2} \gamma^n n!$$

where g and $\gamma \approx 27.23$ are analytic constants. Giménez and Noy also obtained limit laws for the number of edges, number of blocks and number of components for uniformly random planar graphs of size n, see the survey [52]. The same authors developed a more general framework how to prove similar results for other classes with 2-connected excluded minors in the case where the exponential generating functions for 3-connected graphs are known. This framework, for example, has been applied in [32] to count series-parallel graphs and outerplanar graphs. Most of the analysis mentioned above was done using the generating functions approach and the methods of [48]. More recent very impressive results obtained with this technique include asymptotic enumeration of graphs with fixed genus [8, 36], improving earlier results of [73].

Steger and Panagiotou [85] gave an interesting application of the Boltzmann sampling technique for studying asymptotic number of vertices of degree k in random planar graphs. (Boltzmann sampling uses exponential generating functions for relevant classes to sample graphs of size n uniformly at random.)

5.3.2 General minor-closed classes

A parallel approach for proving enumerative results uses probabilistic and combinatorial methods rather than generating functions and singularity analysis. These results are usually less precise (i.e. explicit constants are not computed) but hold more generally.

In 2006, Norine, Seymour, Thomas and Wollan [83] proved that each proper minor-closed class of graphs $\mathcal{A}$ is *small*, that is, the supremum as $n \to \infty$ of the sequence

$$\left(\frac{|\mathcal{A}_n|}{n!}\right)^{1/n} \tag{5.2}$$

is finite, see also [43]. If the above sequence converges to $\gamma \in [0; \infty)$, we say that $\mathcal{A}$ has a *growth constant* $\gamma = \gamma(\mathcal{A})$.

For any class of graphs $\mathcal{A}$ we denote the upper and lower limits of (5.2) by

$\overline{\gamma}(\mathcal{A})$ and $\underline{\gamma}(\mathcal{A})$ respectively. Also, let $\rho(\mathcal{A})$ denote the radius of convergence of the exponential generating function A of $\mathcal{A}$. Of course, $\overline{\gamma}(\mathcal{A}) = \rho(\mathcal{A})^{-1}$ (if we assume that $0^{-1} = \infty$).

Bernardi, Noy and Welsh asked in 2010 whether every proper minor-closed class of graphs has a growth constant. To date, this question has been answered positively for many specific minor-closed classes of graphs, and no example of a class without a growth constant is known. A table with numeric values of growth constants for some specific classes is available in [52].

The work that is closest in spirit to this thesis, and which we use repeatedly in our proofs are papers by McDiarmid, Steger and Welsh (2005) [77] and McDiarmid (2009), [74].

A minor-closed class of graphs is called *addable*, if each excluded minor is 2-connected. We say that $\mathcal{A}$ is *bridge-addable* if given any graph in $\mathcal{A}$ and vertices u and v in distinct components of G, the graph obtained from G by adding an edge joining u and v must be in $\mathcal{A}$.

A class $\mathcal{A}$ is *smooth* if there is a constant γ such that $|\mathcal{A}_n|/(n|\mathcal{A}_{n-1}|) \to \gamma$ as $n \to \infty$. It is not difficult to see that if $\mathcal{A}$ is smooth then γ is the growth constant of $\mathcal{A}$. It was proved in [77] that each addable class of graphs has a growth constant, this was strengthened by McDiarmid [74]: he showed that each addable class of graphs is smooth. The last result implies that the answer to the question of Bernardi, Noy and Welsh is positive for the class of planar graphs $\mathcal{P}$ as well as for any minor-closed class characterised by a set of 2-connected excluded minors.

Another surprisingly useful property of a random graph from an addable class $\mathcal{A}$ is that with a high probability it has linearly many "pendant" copies of any fixed connected graph in $\mathcal{A}$. We state this formally, as it will be used in multiple places in Part II.

Let H be a connected graph on the vertex set $\{1, \ldots, h\}$ which we consider to be rooted at vertex 1, and let G be a graph on the vertex set $\{1, \ldots, n\}$, where $n > h$. Then an induced subgraph $\tilde{H}$ of G is a *pendant appearance* of H if (a) the increasing bijection from $\{1, \ldots, h\}$ to $V(\tilde{H})$ gives an isomorphism between H and $\tilde{H}$; and (b) there is exactly one edge in G between $V(\tilde{H})$ and the rest of G, and this edge is incident with the vertex of $\tilde{H}$ with smallest label.

Lemma 5.3.1 *Let $\mathcal{A}$ be a proper addable minor-closed class of graphs, and let H be a connected graph in $\mathcal{A}$. There is a constant $a > 0$ such that the following holds. For $R_n \in_u \mathcal{A}$, with probability $1 - e^{-\Omega(n)}$ R_n has at least $a \cdot n$ disjoint pendant appearances of H.*

Also, the complete limit distribution of the graph that remains by subtracting the largest component is available by [74]. We present this in more detail in Chapter 7.

More recent development in this direction concerns distribution of random graphs from minor-closed classes, where a pair of parameters (weights) is used to control the edge density and the number of components (connectivity) [76]. The last paper also contains an extensive list of literature regarding enumeration of minor-closed classes.

5.4 Non-addable classes of graphs

While each addable class of graphs is proved in [77] to have growth constant, the classes that are not addable are much less well behaved, and the question of Bernardi, Noy and Welsh has been answered only in a few cases.

One of the strongest results here concerns the important class of graphs embeddable on a surface of genus g, where g is any fixed non-negative integer. In particular, the case $g = 0$ corresponds to planar graphs and the case $g = 1$ corresponds to graphs embeddable on torus. It is known, see [73], that these classes are not addable, for example the graph $2K_5$ is not embeddable on torus. This class of graphs (for any $g \geq 0$) was initially shown to have a growth constant [73] (using combinatorial methods), then complete asymptotics were obtained independently in [8, 36] (using singularity analysis).

The class $\operatorname{Ex} P$ of graphs, where P is a fixed path has growth constant 0 [11]. Forests of paths, the graphs without the "bow-tie" graph as a minor and several other minor-closed classes were considered in [31]: while the growth constant for such classes was shown to exist, other properties, such that the asymptotics of the probability of connectedness seem to vary depending on the excluded minor.

The work presented in this part demonstrates two different infinite families of non-addable minor-closed classes that have a growth constant. There was some other work on disjoint excluded minors, using the approach of [64]. McDiarmid and Kang proved an analogue of Theorem 6.1.1 for unlabelled graphs with at most k disjoint cycles [57]. The class of graphs with at most k disjoint t-star minors (a t-star is the complete bipartite graph $K_{1,t}$) was studied in [75]. We note that results in [76] on random graphs from a weighted minor-closed class allow certain well-behaved non-addable classes as well, such as, for example, the class of graphs with at most k disjoint cycles.

5.5 Generating functions and singularity analysis

In Chapter 9 we make a substantial use of generating function methods for asymptotic enumeration, including the singularity analysis method developed by Flajolet and Odlyzko building on the work of many other authors, see [48].

Let $g(x)$ and $h(x)$ be complex functions and let D be a domain. Let $a \in \mathbb{C}$. We write $f(x) = O(g(x))$ (respectively, $f(x) = o(g(x))$) as $x \to a$, $x \in D$ if $|f(x)/g(x)|$ is bounded by a constant (respectively, $|f(x)/g(x)| \to 0$) uniformly as $x \to a$, $x \in D$.

Given numbers ϕ, R and ρ with $R > \rho > 0$ and $0 < \phi < \pi/2$, define a complex domain

$$\Delta(\phi, \rho, R) = \{z : |z| < R, z \neq \rho, |arg(z-\rho)| > \phi\}.$$

A central result of this method is the so called Transfer theorem (we state a simplified specific version relevant to us, as presented in [52]).

Theorem 5.5.1 *(Transfer) Let α be a real number, not a non-positive integer and let $R > \rho > 0$. Assume that $f(x)$ is analytic in a domain $\Delta = \Delta(\phi, \rho, R)$. If, as $x \to \rho$ in Δ*

$$f(x) = (1 - x/\rho)^{-\alpha}(1 + o(1))$$

Then

$$[x^n]f(x) = \frac{n^{\alpha-1}}{\Gamma(\alpha)}\rho^{-n}(1 + o(1)).$$

As it is common, we will apply this theorem to the exponential generating function of a class $\mathcal{A}$. To obtain the exponential generating function, one can use intuitive high level language to describe the class $\mathcal{A}$ in terms of $\mathcal{A}$ itself and other classes and operations on them. For example, if $\mathcal{C}$ is a class of connected graphs, then its relation to the class $\mathcal{A}$ of all graphs with components in $\mathcal{C}$ is written down symbolically as $\mathcal{A} = \mathrm{SET}(\mathcal{C})$. An automatic translation rule to generating functions then yields

$$A(x) = e^{C(x)}.$$

The example above is a very simple one, there are general rules how to automatically derive exponential generating functions for a great variety of combinatorial classes, this is called the “symbolic method”. We refer for an excellent presentation of the technique to the book “Analytic Combinatorics” by Flajolet and Sedgewick [48], see also Section 8.2.1.

Another theory relevant to us is the enumeration of tree-like structures, where general and very useful results have been obtained by Meir and Moon [79] (using earlier work by Bender and Canfield).

Chapter 6

Graphs with few disjoint cycles

6.1 Introduction

Call a set B of vertices in a graph G a *blocker* if the graph $G - B$ obtained by deleting the vertices in B has no cycles.

We let $\mathcal{F}$ denote the class of forests; let $\mathrm{apex}^k\mathcal{F}$ denote the class of graphs with a blocker of size at most k; and let $\mathrm{Ex}\,(k+1)C$ denote the class of graphs which do not have $k+1$ disjoint cycles. With this notation the Erdős-Pósa theorem says that

$$\mathrm{Ex}\,(k+1)C \subseteq \mathrm{apex}^{f(k)}\mathcal{F}.$$

Now clearly

$$\mathrm{Ex}\,(k+1)C \supseteq \mathrm{apex}^k\mathcal{F}; \tag{6.1}$$

for if a graph has a blocker B then it can have at most $|B|$ disjoint cycles. How much bigger is the left side of (6.1) than the right? Our main theorem is that the difference is relatively small: amongst all graphs without $k+1$ disjoint cycles, all but a small proportion have a blocker of size k. For any class $\mathcal{A}$ of graphs we let $\mathcal{A}_n$ denote the set of graphs in $\mathcal{A}$ on the vertex set $\{1, \ldots, n\}$. (When we say a 'class' of graphs it is assumed to be closed under automorphism.)

Theorem 6.1.1 *For each fixed positive integer k, as $n \to \infty$*

$$|(\mathrm{Ex}\,(k+1)C)_n| = (1 + e^{-\Omega(n)})\ |(apex^k\mathcal{F})_n|. \tag{6.2}$$

Graphs in Ex $2C$ (that is, with no two disjoint cycles) have been well characterised (Dirac [39], Lovász [67]); and from this characterisation we can much refine the above result for the case $k = 1$ – see Section 6.6 below. It seems that no such characterisation is known for graphs in Ex $3C$.

The natural partner to Theorem 6.1.1 is an asymptotic estimate for $|(\text{apex}^{\,k}\mathcal{F})_n|$. Recall the result of Rényi (1959) [87] that

$$|\mathcal{F}_n| \sim e^{\frac{1}{2}} n^{n-2} \sim \left(\frac{e}{2\pi}\right)^{\frac{1}{2}} n^{-\frac{5}{2}} e^n n! \qquad \text{as } n \to \infty. \tag{6.3}$$

Theorem 6.1.2 *For each fixed positive integer k, as $n \to \infty$*

$$|(\mathit{apex}^{\,k}\mathcal{F})_n| \sim c_k 2^{kn} |\mathcal{F}_n| \tag{6.4}$$

where $c_k = \left(2^{\binom{k+1}{2}} e^k k!\right)^{-1}$.

A class $\mathcal{A}$ of graphs has *growth constant* γ if

$$(|\mathcal{A}_n|/n!)^{1/n} \to \gamma \quad \text{as } n \to \infty.$$

The above results (6.2), (6.3) and (6.4) show that both $\text{apex}^{\,k}\mathcal{F}$ and $\text{Ex}\,(k+1)C$ have growth constant $2^k e$.

In order to prove Theorem 6.1.1, we use a seemingly minor 'redundant blocker' extension, Theorem 6.1.3, of the Erdős-Pósa theorem. Call a blocker B in a graph G *k-redundant* if $B \setminus \{v\}$ is still a blocker for all but at most k vertices $v \in B$. Thus a set B of vertices in G is a k-redundant blocker if and only if there is a subset S of B of size at most k such that each cycle in $G - S$ has at least two vertices in $B \setminus S$ (clearly in this case B is also a blocker). Theorem 6.1.3 says that if G does not have $k+1$ disjoint cycles then it has a small k-redundant blocker. Let us now take $f(k)$ as the least value that works in the Erdős-Pósa theorem; and recall that $f(k)$ is $\Theta(k \ln k)$.

Theorem 6.1.3 *If G does not have $k+1$ disjoint cycles then it has a k-redundant blocker of size at most $f(k)+k$.*

The above results yield asymptotic properties of typical graphs without $k+1$ disjoint cycles. We state three theorems. First we note that with high probability k vertices really stand out – they each have degree about $n/2$ whereas each other vertex has much smaller degree – and they form the only minimal blocker of sublinear size. We write $R_n \in_u \mathcal{A}$ to mean that the random graph R_n is sampled uniformly from the graphs in $\mathcal{A}_n$.

Theorem 6.1.4 *There is a constant $\delta > 0$ such that the following holds. Let k be a positive integer. For $n = 1, 2, \ldots$ let $R_n \in_u \text{Ex}\,(k+1)C$, and let S_n be the set of vertices in R_n with degree $> n/\ln n$. Then with probability $1 - e^{-\Omega(n)}$ we have:*

(i) $|S_n| = k$ *and* S_n *is a blocker in* R_n*;*

(ii) each blocker in R_n *not containing* S_n *has size* $> \delta n$*;*

(iii) R_n *has* k *disjoint triangles; and*

(iv) for any constant $\epsilon > 0$*, each vertex in* S_n *has degree between* $(\frac{1}{2} - \epsilon)n$ *and* $(\frac{1}{2} + \epsilon)n$*.*

The second theorem on the random graph $R_n \in_u \mathrm{Ex}\,(k+1)C$ concerns connectivity. Recall that the exponential generating function for the class $\mathcal{T}$ of (labelled) trees is $T(z) = \sum_{n\geq 1} n^{n-2}z^n/n!$, and $T(\frac{1}{e}) = \frac{1}{2}$. Also note that by Rényi's result (6.3), for $R_n \in_u \mathcal{F}$ we have $\mathbb{P}(R_n \text{ is connected}) \to e^{-\frac{1}{2}}$ as $n \to \infty$.

Theorem 6.1.5 *Let* $k \geq 0$ *be an integer, and let* $p_k = e^{-T(\frac{1}{2^k e})}$*. Then for* $R_n \in_u \mathrm{Ex}\,(k+1)C$ *we have*

$$\mathbb{P}(R_n \textit{ is connected }) \to p_k \quad \textit{as } \; n \to \infty. \tag{6.5}$$

In particular, $p_0 = e^{-1/2} = 0.606531$ *(as we already noted),* $p_1 = 0.814600$*,* $p_2 = 0.907879$*,* $p_3 = 0.953998$ *and* $p_4 = 0.977005$ *(to 6 decimal places).*

The third and final theorem presented here on the random graph $R_n \in_u \mathrm{Ex}\,(k+1)C$ concerns the chromatic number $\chi(R_n)$ and the clique number $\omega(R_n)$. It shows for example in the case $k = 2$ (concerning graphs with no three disjoint cycles) that both $\mathbb{P}(\chi(R_n) = \omega(R_n) = 3)$ and $\mathbb{P}(\chi(R_n) = \omega(R_n) = 4)$ tend to $\frac{1}{2}$ as $n \to \infty$.

Theorem 6.1.6 *Let* k *be a positive integer, and let the random graph* R *be picked uniformly from the set of all graphs on* $\{1, \ldots, k\}$*. For each* n *let* $R_n \in_u \mathrm{Ex}\,(k+1)C$*. Then for each* $3 \leq i \leq j \leq k+2$*, as* $n \to \infty$

$$\mathbb{P}\left((\omega(R_n) = i) \wedge (\chi(R_n) = j)\right) \to \mathbb{P}((\omega(R) = i-2) \wedge (\chi(R) = j-2))$$

(and for other values of i, j *the left side tends to 0).*

The plan of the rest of the chapter is as follows. First we prove Theorem 6.1.2 concerning the number of 'apex forests'. Next we prove the 'redundant blocker' result, Theorem 6.1.3, which we then use in the proof of our main result, Theorem 6.1.1. After that, we prove the three theorems on the random graph $R_n \in_u \mathrm{Ex}\,(k+1)C$, namely Theorems 6.1.4, 6.1.5 and 6.1.6. The last main section concerns the case $k = 1$ on graphs with no two disjoint cycles; and finally we make some remarks concerning extensions of the results presented earlier.

In our proofs we use some results developed in the study of random planar graphs [77, 78] and graphs from minor-closed classes [74]. Most of those results no longer work when a class of graphs fails to be closed under taking a disjoint union (as is the case here). However, for example to prove the main theorem, our extension of the Erdős-Pósa theorem allows us to decompose a graph with few disjoint cycles into a small redundant blocker and a forest, and we may then use the results mentioned above for random forests.

Initial work on this chapter was done in 2008 while the author of the thesis was studying for an MSc at the University of Oxford, with the second author, Colin McDiarmid, as supervisor. It was mainly written while the second author was at the Mittag-Leffler Institute during April 2009; and the support of that Institute is gratefully acknowledged.

6.2 Counting apex forests

The following lemma will be useful in this section and in Section 6.4. Call a pair of adjacent vertices in a graph a *spike* if it consists of a leaf and a vertex of degree 2, which are not contained in a component of just three vertices forming a path. Observe that distinct spikes are disjoint.

Lemma 6.2.1 *There exist constants $a > 0$ and $b > 0$ such that, for n sufficiently large, the number of forests $F \in \mathcal{F}_n$ with less than an spikes is less than $e^{-bn}|\mathcal{F}_n|$.*

Proof Let H be the path of 3 vertices rooted at an end vertex. By (6.3) the class $\mathcal{F}$ of forests has a growth constant, namely e. Thus we may apply the 'appearances theorem' Theorem 5.1 of [78] to lower bound the number of pendant appearances of H in a random forest; and each such appearance yields a spike. (A pendant appearance may be defined as follows. Let H be a connected graph on the vertex set $\{1, \ldots, h\}$ and let G be a graph on the vertex set $\{1, \ldots, n\}$. Then an induced subgraph $\tilde{H}$ of G is a *pendant appearance* of H if (a) the increasing bijection from $\{1, \ldots, h\}$ to $V(\tilde{H})$ gives an isomorphism between H and $\tilde{H}$; and (b) there is exactly one edge in G between $V(\tilde{H})$ and the rest of G, and this edge is incident with the vertex of $\tilde{H}$ with smallest label.) □

Proof of Theorem 6.1.2 By (6.3) we have

$$|\mathcal{F}_n| \sim (n)_k \, e^k |\mathcal{F}_{n-k}|. \tag{6.6}$$

Let $n > k$, let $V = \{1, \dots, n\}$, and consider the following constructions of graphs on V:

(i) Choose a k-set $S \subset V$, and put any graph on S ($\binom{n}{k} 2^{\binom{k}{2}}$ choices).

(ii) Put any forest F on $V \setminus S$ ($|\mathcal{F}_{n-k}|$ choices).

(iii) Add the edges of any bipartite graph H with parts S and $V \setminus S$ ($2^{k(n-k)}$ choices).

Clearly each graph constructed is in $(\text{apex}^k \mathcal{F})_n$, and each graph in $(\text{apex}^k \mathcal{F})_n$ is constructed at least once. By (6.6) the number of constructions is

$$\binom{n}{k} 2^{\binom{k}{2}} 2^{k(n-k)} |\mathcal{F}_{n-k}| \sim c_k 2^{kn} |\mathcal{F}_n|$$

so $|(\text{apex}^k \mathcal{F})_n|$ is at most this number.

Let us bound $|(\text{apex}^k \mathcal{F})_n|$ from below by showing that almost all of the constructions yield distinct graphs. Observe that $G \in (\text{apex}^k \mathcal{F})_n$ appears just once if and only if G has a unique blocker of size k. Fix $S = S_0$ for some k-set $S_0 \subseteq V$.

Let us say that a graph $G \in (\text{apex}^k \mathcal{F})_n$ is *good* if (a) $G - S_0 \in \mathcal{F}$; and (b) for each vertex $s \in S_0$ the forest $G - S_0$ has $k+1$ spikes such that s is adjacent to both vertices in each of these spikes, and so forms a triangle with each. If G is good then S_0 must be the unique blocker of size k in G. For if S' is another blocker, and $s \in S_0 \setminus S'$, then S' must contain a vertex from each spike in $G - S_0$ which forms a triangle with s, and so $|S'| \geq k+1$.

By Lemma 6.2.1, there exist constants $a > 0$ and $b > 0$ such that (assuming n is sufficiently large) the number of forests $F \in \mathcal{F}_{n-k}$ with less than an spikes is less than $e^{-bn} |\mathcal{F}_{n-k}|$. But if F has at least an spikes then there are at most

$$2^{k(n-k)} k \, \mathbb{P} \left(\text{Bin} \left(\lceil an \rceil, 1/4 \right) \leq k \right)$$

ways to choose the bipartite graph H in step (iii) so that the resulting graph constructed is not good. Hence, by considering separately the cases when F has $< an$ spikes and when F has $\geq an$ spikes, we see that the number of ways to choose F and H so that the resulting graph will be constructed more than once is at most

$$2^{k(n-k)} |\mathcal{F}_{n-k}| \left(e^{-bn} + k \, \mathbb{P} \Big(\text{Bin}(\lceil an \rceil, \frac{1}{4}) \leq k \Big) \right).$$

Now if $X \sim \mathrm{Bin}(m,p)$ and $0 \le j \le m$ then $\mathbb{P}(X \le j) \le \binom{m}{j}(1-p)^{m-j}$, and so since k is constant

$$\mathbb{P}\left(\mathrm{Bin}(\lceil an \rceil, \frac{1}{4}) \le k\right) \le \binom{\lceil an \rceil}{k}(3/4)^{an-k} = O(n^k(3/4)^{an}).$$

Thus the number of constructions that fail to yield a unique graph when the choice at step (i) is fixed is at most

$$2^{kn}|\mathcal{F}_n|\ e^{-\Omega(n)}. \tag{6.7}$$

Since the number of all possible sets S and graphs on S is $\binom{n}{k}2^{\binom{k}{2}} = O(n^k)$, it follows that the total number of constructions that fail to yield a unique graph is also at most the quantity in (6.7), which completes the proof. □

6.3 Redundant blockers: proof of Theorem 6.1.3

We will deduce Theorem 6.1.3 easily from the following lemma.

Lemma 6.3.1 *Let $k \ge 0$, let $G \in \mathrm{Ex}\,(k+1)C$, and let Q be a blocker in G. Then there are sets $S \subseteq Q$ with $|S| \le k$ and $A \subseteq V(G) \setminus Q$ with $|A| \le k$ such that there is no cycle C in $G - S$ with $|V(C) \cap (Q \cup A)| \le 1$.*

Note that the conclusion of the lemma is equivalent to saying that the graph $G - ((Q \setminus \{x\}) \cup A)$ is acyclic for each vertex $x \in Q \setminus S$; that is, each vertex $x \in Q \setminus S$ has at most one edge to each tree in the forest $G - (Q \cup A)$.

Proof We use induction on k. Clearly the result holds for the case $k = 0$, as we may take $A = S = \emptyset$. Let $j \ge 1$ and suppose that the result holds for $k = j - 1$. Let $G \in \mathrm{Ex}\,(j+1)C$ and let Q be a blocker in G. We may assume that for some tree T in the forest $G - Q$, and some vertex $y \in Q$, the induced subgraph $G[V(T) \cup \{y\}]$ has a cycle (as otherwise we may again take $A = S = \emptyset$ and we are done).

Fix one such tree T, and fix a root vertex r in T. For each vertex v in T let T_v denote the subtree of T rooted at v. (Thus T_r is T.) Let

$$R = \{v \in V(T) :\ G[V(T_v) \cup \{x\}] \text{ has a cycle for some } x \in Q\}.$$

By our assumption $R \ne \emptyset$. In the tree T, choose a vertex $u \in R$ at maximum distance from the root r. Let $z \in Q$ be such that $G[V(T_u) \cup \{z\}]$ has a cycle.

Let $G' = G - (V(T_u) \cup \{z\})$ and let $Q' = Q \setminus \{z\}$. Then clearly $G' \in \mathrm{Ex}\,(jC)$ and Q' is a blocker in G'. Hence we can apply the induction hypothesis to G' and Q', and obtain sets of vertices $S' \subseteq Q'$ and $A' \subseteq V(G') \setminus Q'$ each of size at most $j-1$, such that there is no cycle C in $G' - S'$ with $|V(C) \cap (Q' \cup A')| \leq 1$.

Now set $S = S' \cup \{z\}$ and $A = A' \cup \{u\}$. Suppose that there is a cycle C in $G - S$ with $|V(C) \cap (Q \cup A)| \leq 1$. We want to find a contradiction, since that will establish the induction step, and thus complete the proof of the lemma.

Note that C must have a vertex in the blocker Q: so we may let $x \in Q$ be the unique vertex in $V(C) \cap (Q \cup A)$. It follows that $u \notin V(C)$. But $V(C) \cap V(T_u)$ cannot be empty: for then C would be a cycle in $G' - S'$, and by induction we would have $2 \leq |V(C) \cap (Q' \cup A')| \leq |V(C) \cap (Q \cup A)|$.

Hence the connected graph $C - \{x\}$ is a subgraph of T with a vertex in T_u but not containing u. Therefore $C - \{x\}$ must be contained in a proper subtree T_w of T_u; but this implies that $w \in R$, which contradicts our choice of u. □

Proof of Theorem 6.1.3 Let $k \geq 0$ and let $G \in \mathrm{Ex}\,(k+1)C$. Let Q be a blocker in G of size at most $f(k)$. By Lemma 6.3.1, there are sets $S \subseteq Q$ with $|S| \leq k$ and $A \subseteq V(G) \setminus Q$ with $|A| \leq k$ such that there is no cycle C in $G - S$ with $|V(C) \cap (Q \cup A)| \leq 1$.

Then the set $B = Q \cup A$ is as required. For, given $v \in B \setminus S$, there cannot be a cycle C in $G - (B \setminus \{v\})$, since C would be a cycle in $G - S$ with $|V(C) \cap B| \leq 1$; and thus $B \setminus \{v\}$ is a blocker. □

6.4 Proof of the main theorem, Theorem 6.1.1

Let $\kappa(G)$ denote the number of connected components of a graph G. If the random variable X has the Poisson distribution with mean 1, then for each positive integer t we have $\mathbb{P}[X \geq t] \leq 1/t!$. Hence by Theorem 2.1 of McDiarmid, Steger, Welsh (2006) [78] applied to the class $\mathcal{F}$ of forests we have:

Lemma 6.4.1 *For each positive integer t*

$$|\{F \in \mathcal{F}_n : \kappa(F) \geq t+1\}| \leq |\mathcal{F}_n|/t!.$$

The idea of the proof of Theorem 6.1.1 is to give constructions which yield every graph in $(\mathrm{Ex}\,(k+1)C)_n$ at least once (as well as other graphs); show that there are few 'unrealistic' constructions; and show that few 'realistic' constructions yield a graph in $(\mathrm{Ex}\,(k+1)C \setminus \mathrm{apex}^k\mathcal{F})_n$.

Proof of Theorem 6.1.1 Fix a positive integer k. By Theorem 6.1.3, there is an integer $r \leq f(k)+k$ such that the following holds. For each graph G in $\mathrm{Ex}\,(k{+}1)C$ with at least r vertices, there is a blocker R of size r and a subset S of R of size k such that $R \setminus v$ is still a blocker for each vertex $v \in R \setminus S$.

Let $n > r$. Then the following constructions will yield every graph in $(\mathrm{Ex}\,(k{+}1)C)_n$ at least once (as well as other graphs).

(i) Choose an r-subset $R \subseteq V$, put any graph on R, and choose a k-subset $S \subseteq R$ ($\binom{n}{r}2^{\binom{r}{2}}\binom{r}{k} = O(n^r)$ choices).

(ii) Add the edges of any bipartite graph with parts S and $V \setminus R$ ($2^{k(n-r)}$ choices).

(iii) Put any forest F on $V \setminus R$ ($|\mathcal{F}_{n-r}|$ choices).

(iv) Add the edges of any bipartite graph with parts $R \setminus S$ and $V \setminus R$, subject to the restriction that each $v \in R \setminus S$ has at most one edge to each component tree of the forest F on $V \setminus R$.

We want upper bounds on the number of constructions. By the restriction in (iv) above, for each vertex $v \in R \setminus S$, the number of edges between v and the vertices in $V \setminus R$ is at most $\kappa(F)$. Let $t = t(n) \sim n(\ln n)^{-\frac{1}{2}}$. Then by Lemma 6.4.1

$$|\{F \in \mathcal{F}_{n-r} : \kappa(F) \geq t\}| \leq |\mathcal{F}_{n-r}|/(t-1)! \leq |\mathcal{F}_n|\ e^{-\Omega(n(\ln n)^{\frac{1}{2}})}.$$

Call a construction *realistic* if there are at most t edges between each vertex $v \in R \setminus S$ and the vertices in $V \setminus R$; and *unrealistic* otherwise. Then the number of unrealistic constructions is at most

$$O(n^r)\ 2^{kn}\ |\mathcal{F}_n|\ 2^{(r-k)(n-r)}\ e^{-\Omega(n(\ln n)^{\frac{1}{2}})} = |\mathcal{F}_n|\ e^{-\Omega(n(\ln n)^{\frac{1}{2}})}.$$

Thus there are relatively few unrealistic constructions, and we see that we need to consider only realistic constructions. Further, since $t = o(n)$, for n sufficiently large

$$\sum_{i=0}^{t} \binom{n-r}{i} \leq 2\binom{n-r}{t} \leq 2\left(\frac{ne}{t}\right)^t;$$

and so, in realistic constructions, the number of choices for step (iv) is

$$\left(\sum_{i=0}^{t} \binom{n-r}{i}\right)^{r-k} \leq 2^r \left(\frac{ne}{t}\right)^{tr} = (1+o(1))^n.$$

Let us bound the number of realistic constructions yielding a graph G in $(\mathrm{Ex}\,(k{+}1)C \setminus \mathrm{apex}^k\mathcal{F})_n$. Each such construction has a cycle contained in $V \setminus S$; and such a cycle C can touch at most $2(r-k)$ spikes, since as we travel around C we must visit $R \setminus S$ at least once between any three visits to distinct spikes.

Now suppose that each vertex in S is adjacent to both vertices of at least $2r-k$ spikes. Then the k vertices in S would each form triangles with at least $2r-k-2(r-k)=k$ spikes disjoint from C; and amongst these triangles we could find k disjoint ones (for example by picking the triangles greedily). But together with C there would now be at least $k+1$ disjoint cycles, contradicting the assumption that $G \in \mathrm{Ex}\,(k{+}1)C$. Hence, for at least one vertex v in S, v must be adjacent to both vertices of at most $2r-k-1 \le 2r$ spikes.

Therefore, given any choices at steps (i),(iii) and (iv), if F has z spikes then the number of choices at step (ii) to obtain a graph in $(\mathrm{Ex}\,(k{+}1)C \setminus \mathrm{apex}^k\mathcal{F})_n$ is at most

$$2^{k(n-r)}\; k\; \mathbb{P}[\mathrm{Bin}(z,1/4) \le 2r],$$

and arguing as before

$$\mathbb{P}[\mathrm{Bin}(z,1/4) \le 2r] \le \binom{z}{2r}(3/4)^{z-2r} = O(z^{2r}(3/4)^z).$$

By Lemma 6.2.1, there exist constants $a>0$ and $b>0$ such that (assuming n is sufficiently large) the number of graphs $F \in \mathcal{F}_{n-r}$ with less than an spikes is at most $e^{-bn}|\mathcal{F}_{n-r}|$. Hence, by considering separately the cases when F has $< an$ spikes and when F has $\ge an$ spikes, we see that the number of realistic constructions which yield a graph in $(\mathrm{Ex}\,(k{+}1)C \setminus \mathrm{apex}^k\mathcal{F})_n$ is at most

$$\begin{aligned} & O(n^r)\; 2^{k(n-r)}2^r \left(\frac{ne}{t}\right)^{tr} |\mathcal{F}_{n-r}| \left(e^{-bn} + O(z^{2r}(3/4)^z)\right) \\ = \;& e^{-\Omega(n)}2^{kn}\; |\mathcal{F}_n| \;=\; e^{-\Omega(n)}|(\mathrm{apex}^k\mathcal{F})_n| \end{aligned}$$

by Theorem 6.1.2. □

6.5 Proofs for random graphs R_n

In this section we prove Theorems 6.1.4, 6.1.5 and 6.1.6. The following lemma makes the task more straightforward. Recall that the total variation distance $d_{TV}(X,Y)$ between two random variables X and Y is the supremum over all events A of $|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|$.

Lemma 6.5.1 *Let k be a positive integer. Let $R_n \in_u \mathrm{Ex}\,(k+1)C$; let $R_n^a \in_u$ $\mathrm{apex}^k\mathcal{F}$; and let R_n^c denote the graph which is the result of a construction as in the proof of Theorem 6.1.2, where the steps are chosen uniformly at random. If X_n and Y_n are any two of these random variables, then the total variation distance between them satisfies*

$$d_{TV}(X_n, Y_n) = e^{-\Omega(n)}. \tag{6.8}$$

Proof Theorem 6.1.1 gives $d_{TV}(R_n, R_n^a) = e^{-\Omega(n)}$; and Theorem 6.1.2 and the inequality (6.7) give $d_{TV}(R_n^a, R_n^c) = e^{-\Omega(n)}$. □

Proof of Theorem 7.1.3 By Lemma 6.5.1, we may work with R_n^c rather than with R_n. Let $F_m \in_u \mathcal{F}_m$ for $m = 1, 2, \ldots$. If positive numbers $n_1, \ldots, n_j$ sum to at most m then $\prod_i n_i \le \left(\frac{m}{j}\right)^j$. Also, if vertex 1 has degree j in F_m and we delete this vertex then we obtain a forest with at least j components. Thus by considering the component sizes in F_{m-1}, and using Lemma 6.4.1

$$\begin{aligned}\mathbb{P}(\Delta(F_m) = j) &\le m \cdot \left(\frac{m}{j}\right)^j \frac{|\mathcal{F}_{m-1}|}{(j-1)!} \frac{1}{|\mathcal{F}_m|} \\ &\le j \left(\frac{m}{j}\right)^j \frac{1}{j!} \quad \text{since } m|\mathcal{F}_{m-1}| \le |\mathcal{F}_m| \\ &\le j \left(\frac{me}{j^2}\right)^j \quad \text{since } j! \ge (j/e)^j.\end{aligned}$$

Hence $\mathbb{P}(\Delta(F_m) \ge j) = e^{-\Omega(m)}$ if $j = \Omega(m/\ln m)$.

Recall that we let S denote the set used in the construction of R_n^c as in the proof of Theorem 6.1.2. The key observation now is that

$$\mathbb{P}(S_n \not\subseteq S) \le \mathbb{P}(\Delta(F_{n-k}) > n/\ln n - k),$$

and so by the above $\mathbb{P}(S_n \not\subseteq S) = e^{-\Omega(n)}$. But the number of constructions with S_n a proper subset of S is at most $2^{(k-1)n+o(n)}|\mathcal{F}_{n-k}|$, which is $2^{-n+o(n)}$ times the number of constructions, and hence $\mathbb{P}(S_n = S) = 1 - e^{-\Omega(n)}$.

We have now dealt with statement (i) in the theorem, so let us consider statements (ii) and (iii). By Lemma 6.2.1, there exists $\delta > 0$ such that the probability that F_{n-k} has a matching of size at least $5\delta n$ is $1 - e^{-\Omega(n)}$; and given such a matching, for each $j \in S$, the probability that j fails to be the central vertex of at least δn otherwise disjoint triangles (disjoint from $S \setminus \{j\}$) is at most

$$\mathbb{P}\left(\mathrm{Bin}(\lceil 5\delta n\rceil, \frac{1}{4}) < \delta n\right) \le e^{-\delta^2 n/8} = e^{-\Omega(n)}$$

by a standard Chernoff bound (namely that if $X \sim \mathrm{Bin}(m,p)$ and $a \geq 0$ then $\mathbb{P}(X \leq mp - a) \leq e^{-2a^2/m}$). But if vertex j is the central vertex of at least δn otherwise disjoint triangles, then any blocker not containing j must have size at least δn. Also in this case if n is sufficiently large that $\delta n \geq 2k$ we can pick k disjoint triangles in R_n^c one for each vertex in S. This deals with statements (ii) and (iii).

Finally, for statement (iv), a two-sided Chernoff bound shows that the number of constructions such that (iv) fails is $2^{kn-\Omega(n)}|\mathcal{F}_{n-k}|$; and it follows that (iv) holds with probability $1 - e^{-\Omega(n)}$. □

In order to prove Theorem 6.1.5 we shall use Lemma 4.3 of [74]. We need some definitions to present that lemma (in a simplified form).

Given a graph G on $\{1,\ldots,n\}$ let Big(G) denote the (lexicographically first) component of G with the most vertices, and let Frag(G) denote the graph induced on the vertices not in Big(G). Let $\mathcal{A}$ be a class of graphs. We say that $\mathcal{A}$ is *bridge-addable* if given any graph in $\mathcal{A}$ and vertices u and v in distinct components of G, the graph obtained from G by adding an edge joining u and v must be in $\mathcal{A}$. Given a graph H in $\mathcal{A}$, we say that H is *freely addable* to $\mathcal{A}$ if, given any graph G disjoint from H, the union of G and H is in $\mathcal{A}$ if and only if G is in $\mathcal{A}$. We say that the class $\mathcal{A}$ is *smooth* if $\mathcal{A}$ has growth constant γ and $\frac{|\mathcal{A}_n|}{n|\mathcal{A}_{n-1}|} \to \gamma$ as $n \to \infty$. Finally, note our standard convention that for the class $\mathcal{A}$ we will use $A(z)$ to denote its exponential generating function $\sum_{n\geq 0} |\mathcal{A}_n| z^n/n!$.

Lemma 6.5.2 *(McDiarmid [74]) Let the graph class $\mathcal{A}$ be bridge-addable; let $R_n \in_u \mathcal{A}_n$; let $\mathcal{B}$ denote the class of all graphs freely addable to $\mathcal{A}$; and suppose that $\mathbb{P}(\mathrm{Frag(R_n)} \in \mathcal{B}) \to 1$ as $n \to \infty$. Suppose further that $\mathcal{A}$ is smooth, with growth constant γ. Let $\mathcal{C}$ denote the class of connected graphs $\mathcal{B}$. Then $C(1/\gamma)$ is finite, and*

$$\mathbb{P}[R_n \textit{ is connected }] \to e^{-C(1/\gamma)} \quad \textit{as } n \to \infty.$$

Proof of Theorem 7.1.5 By Lemma 6.5.1, we may work with R_n^a, rather than with R_n. Let $\mathcal{A}$ denote apex$^k\mathcal{F}$: thus $R_n^a \in_u \mathcal{A}_n$. Clearly $\mathcal{A}$ is bridge-addable, and the class of graphs freely addable to $\mathcal{A}$ is $\mathcal{F}$. By Theorem 6.1.2, $\mathcal{A}$ is smooth, with growth constant $2^k e$. By Lemma 6.5.2 above, it now remains only to show that $\mathbb{P}(\mathrm{Frag(R_n^a)} \in \mathcal{F}) \to 1$ as $n \to \infty$. We may assume that $k \geq 1$.

By Theorem 6.1.2, the class apex$^{k-1}\mathcal{F}$ has growth constant $2^{k-1}e$, and so the class $\mathcal{D}$ of graphs with each component in apex$^{k-1}\mathcal{F}$ also has growth constant $2^{k-1}e$ (by the 'exponential formula'). If $G \in (\mathcal{A} \setminus \mathcal{D})_n$ and $\mathrm{Frag(G)} \notin \mathcal{F}$, then apart from a unique component of size at most $\lfloor n/2 \rfloor$ which is in $\mathcal{A} \setminus \text{apex}^{k-1}\mathcal{F}$ the rest of the graph is in $\mathcal{F}$; and the number of such graphs is at most

$$\sum_{t=0}^{\lfloor n/2 \rfloor} \binom{n}{t} |\mathcal{A}_t| \cdot |\mathcal{F}_{n-t}| = n!(e+o(1))^n 2^{kn/2} = 2^{-kn/2+o(n)} \cdot |\mathcal{A}_n|.$$

Thus $\mathbb{P}(\mathrm{Frag}(\mathrm{R}_n^a) \notin \mathcal{F}) = \mathrm{e}^{-\ (\mathrm{n})} = \mathrm{o}(1)$. □

Lemmas 4.3 and 4.4 in [74] may be used to yield further results on $\mathrm{Frag}(\mathrm{R_n})$.

Proof of Theorem 6.1.6 By Lemma 6.5.1 it is sufficient to consider R_n^c rather than R_n. Let H be the random graph on the set S of k vertices when constructing R_n^c as in the step (i) of the proof of Theorem 6.1.2. It is easy to see that, with probability $\to 1$ as $n \to \infty$, there are adjacent vertices in $V \setminus S$ which are adjacent to each vertex in S, and thus $\omega(R_n^c) = \omega(H) + 2$ and $\chi(R_n^c) = \chi(H) + 2$, which completes the proof. □

6.6 No two disjoint cycles

Let $\mathcal{D}^k$ denote the 'difference' class $\text{Ex}\,(k{+}1)C \setminus \text{apex}^{\,k}\mathcal{F}$, the class of graphs with no $k+1$ disjoint cycles but with no blocker of size at most k. Our main result, Theorem 6.1.1, shows that $\mathcal{D}^k$ is exponentially smaller than $\text{Ex}\,(k{+}1)C$. For the case $k = 1$ we can say much more about $\mathcal{D} = \mathcal{D}^1$, based on results from 1965 of Dirac [39] and Lovász [67], see also Lovász [68] problem 10.4.

We need some definitions and notation. The *2-core* or just *core* of a graph G is the unique maximal subgraph of minimum degree at least 2, and is denoted by core(G). Let $\tilde{\mathcal{K}}$ denote the class of graphs homeomorphic to K_5; let $\tilde{\mathcal{B}}$ denote the class of graphs homeomorphic to a multigraph $\tilde{K}_{3,t}$ formed from the complete bipartite graph $K_{3,t}$ for some $t \geq 0$ by possibly adding edges or multiple edges between vertices in the 'left part' of size 3 ($K_{3,0}$ has only a 'left part'); and let $\tilde{\text{cw}}$ denote the class of graphs homeomorphic to a multigraph formed from the t-vertex wheel W_t for some $t \geq 4$ by possibly adding parallel edges to some spokes. Let $\mathcal{K}$, $\mathcal{B}$, cw denote the classes of graphs G such that core(G) is in $\tilde{\mathcal{K}}$, $\tilde{\mathcal{B}}$, $\tilde{\text{cw}}$ respectively. Call the graphs in cw *generalised wheels*, and note that $\text{cw} \subseteq \mathcal{D}$.

Theorem 6.6.1 *(Dirac [39], Lovász [67])*

$$\text{Ex}\ 2C = (\mathit{apex}\ \mathcal{F}) \cup \mathit{cw} \cup \mathcal{B} \cup \mathcal{K}.$$

By Theorems 6.1.1 and 6.1.2, Ex $2C$ and apex $\mathcal{F}$ both have growth constant $2e$, and $\mathcal{D}$ = Ex $2C \setminus$ apex $\mathcal{F}$ is exponentially smaller. The next result shows that $\mathcal{D}$ is dominated by the class cw of generalised wheels, and gives an asymptotic formula for $|\mathcal{D}_n|$.

Theorem 6.6.2 *The classes $\mathcal{K}$ and $\mathcal{B}$ each have growth constant e, and cw has growth constant γ satisfying $e < \gamma < 2e$. Indeed $|cw_n| \sim c/n\ \gamma^n\ n!$, where the constants c and γ are given by equations (6.9) and (6.10). Thus $|\mathcal{D}_n| \sim c/n\ \gamma^n\ n!$ so that $\mathcal{D}$ has growth constant γ, and $\mathcal{D} \setminus cw$ has growth constant e. To 3 decimal places we have $c = 0.158$ and $\gamma = 4.346$.*

Proof Direct estimation shows easily that $\tilde{\mathcal{K}}$ has growth constant 1. Let $\mathcal{R}$ denote the class of rooted trees, so that $R(z) = \sum_{n\geq 1} n^{n-1}z^n/n!$. It is well known that the radius of convergence ρ_R of R equals $1/e$ and $R(1/e) = 1$. Since graphs in $\mathcal{K}$ are obtained from graphs in $\tilde{\mathcal{K}}$ by substituting rooted trees for vertices, we have $K(z) = \tilde{K}(R(z))$, and it follows that $\mathcal{K}$ has growth constant e. In a similar way we may see that $\mathcal{B}$ also has growth constant e.

Now let us consider cw. We need to see how graphs in cw are formed from simpler graphs. A 'hairy cycle' is a graph formed by attaching vertex-disjoint paths to a cycle. More precisely, a connected graph is a *hairy cycle* if its core is a cycle and each vertex not on the cycle has degree 1 or 2. A *coloured hairy cycle* is a hairy cycle in which each vertex on the cycle is coloured black or white. Let $\mathcal{H}^+$ be the class of coloured hairy cycles, and let $\mathcal{H}$ be the class of graphs in $\mathcal{H}^+$ such that at least 3 vertices on the cycle are either coloured black or have degree at least three. We shall see later that the difference between $\mathcal{H}^+$ and $\mathcal{H}$ is negligible.

Let $\mathcal{S}$ denote the class of homeomorphs of a star (sometimes called 'spiders'), rooted at the centre vertex, with the root coloured black or white. Thus the graphs in $\mathcal{S}$ correspond to a black or white root vertex and a set of oriented paths; and so $S(z) = 2ze^{z/(1-z)}$. Recall that the exponential generating function for cycles is $C(z) = -\frac{1}{2}\ln(1-z) - \frac{1}{2}z - \frac{1}{4}z^2$. Graphs in $\mathcal{H}^+$ are obtained from cycles by substituting a rooted graph from $\mathcal{S}$ for each vertex, so $H^+(z) = C(S(z))$.

Let $\tilde{\text{cw}}^+$ be the class of graphs G obtained by starting with a root vertex v and a graph $H \in \mathcal{H}^+$ not containing v; and joining v to each leaf of H and to each black vertex on the cycle in H, and then removing all colours. If the initial graph H is in $\mathcal{H}$ then $G \in \tilde{\text{cw}}$ (the rooting of v is irrelevant since the 'centre' vertex of a wheel is unique, so we may say $\tilde{\text{cw}} \subseteq \tilde{\text{cw}}^+$). Conversely, given a graph G in $\tilde{\text{cw}}^+$, with root vertex v, colour the vertices on the rim black if they are adjacent to v

and white otherwise, and then delete v. We obtain a graph H in $\mathcal{H}^+$, and if the initial graph G is in $\tilde{\text{cw}}$ then $H \in \mathcal{H}$. Hence $\tilde{W}(z) = zH(z)$ and $\tilde{W}^+(z) = zH^+(z)$.

Let cw^+ be the class of graphs formed by starting with a graph in $\tilde{\text{cw}}^+$ and substituting rooted trees for vertices. (Thus cw^+ is the class of graphs with 2-core in $\tilde{\text{cw}}^+$, except that we always treat the root as having degree at least 2.) Then, $\text{cw} \subseteq \text{cw}^+$, and arguing as earlier, $W(z) = \tilde{W}(R(z))$, and $W^+(z) = \tilde{W}^+(R(z)) = R(z)\ C(f(z))$ where $f(z) = S(R(z))$.

Observe that $S(\frac{1}{2}) = e > 1$, so there exists x with $0 < x < \frac{1}{2}$ such that $S(x) = 1$. Since $\rho_R = 1/e$ and $R(1/e) = 1$, there exists r with $0 < r < 1/e$ such that $R(r) = x$; and so

$$f(r) = S(R(r)) = 1. \tag{6.9}$$

We have a supercritical composition $C(f(z))$ (see [48] VI.9 page 411). It follows from standard results (see for example Theorem VI.3 and Example VI.11 in [48]) that

$$|\text{cw}_n^+| \sim c/n\ \gamma^n\ n! \quad \text{where}\ \ \gamma = 1/r\ \ \text{and}\ \ c = \frac{1}{2}R(r). \tag{6.10}$$

Finally, it is easy to see that $\mathcal{H}^+ \setminus \mathcal{H}$ has growth constant 1, and so $\text{cw}^+ \setminus \text{cw}$ has growth constant $e < \gamma$. Thus the asymptotic formula in (6.10) applies also to cw.

Numerical calculations yield c and γ as given in the theorem. Indeed $S(x) = 1$ for $x = 0.315411$ (to six decimal places); $c = x/2$; and $R(r) = x$ for $r = 0.230089$ (to six decimal places). □

6.7 Concluding Remarks

Our results are stated for a fixed number k of disjoint cycles, but they hold also when k is allowed to grow with n. Indeed it is straightforward to adapt the proofs to show that Theorem 6.1.2 holds as long as $k = o(n)$, and Theorem 6.1.1 holds for $k = o(\ln n/(\ln\ln n)^2)$ (in the proof take $t = \omega(n)\ n/\ln n$ where $\omega(n) \to \infty$ slowly as $n \to \infty$).

It would be interesting to know more about the difference class $\mathcal{D}^k = \text{Ex}\,(k+1)C \setminus \text{apex}^k\mathcal{F}$ for $k \geq 2$, ideally along the lines of the results on $\mathcal{D}^1$ in the last section. There are results for unlabelled graphs corresponding to the results given here for labelled graphs – see [57].

The Erdős-Pósa theorem was extended from disjoint cycles to suitable more general disjoint graph minors by Robertson and Seymour [88] in 1986. Our results can be extended in this direction, and we do so in the following chapters. For example, there is a result corresponding to Theorem 6.1.1 for ‘long’ cycles. Fix an integer $j \geq 3$, and call a cycle *long* if it has length at least j. Then amongst all graphs G on $\{1, \ldots, n\}$ which do not have $k+1$ disjoint long cycles, all but an exponentially small proportion have a set B of k vertices such that $G - B$ has no long cycles, indeed in Chapter 7more general minor-closed classes are considered and a condition for a similar result to hold is given.

There is also a version of the Erdős-Pósa theorem for directed graphs [86]: what can be said in this case?

As well as concerning a problem which is interesting in its own right, the results presented here are a step towards understanding the behaviour of random graphs from a minor-closed class where the excluded minors are not 2-connected, see the last section of [74].

Acknowledgement We would like to thank the referee of [64] for helpful detailed comments.

Chapter 7

Few disjoint minors in $\mathcal{B}$ when $\mathrm{Ex}\,\mathcal{B}$ excludes a fan

7.1 Introduction

The Erdős-Pósa theorem was generalised in 1986 by Robertson and Seymour [88]. Let $\mathcal{A}$ be a *minor-closed* class of graphs; that is, if $G \in \mathcal{A}$ and H is a minor of G then $H \in \mathcal{A}$. Then $\mathcal{A}$ may be written as $\mathrm{Ex}\,\mathcal{B}$, where $\mathcal{B}$ consists of the minor-minimal graphs not in $\mathcal{A}$, the *excluded minors of* $\mathcal{A}$; and $\mathcal{B}$ is finite by the fundamental result in 2004 by the same authors [89]. A *$\mathcal{B}$-minor-blocker* for a graph G is a set B of vertices such that $G - B$ is in $\mathrm{Ex}\,\mathcal{B}$.

The generalisation of the Erdős-Pósa theorem is as follows. Let $\mathcal{A}$ be any minor-closed class of graphs which does not include some planar graph, and let $\mathcal{B}$ be its set of excluded minors, so that $\mathcal{A}$ is $\mathrm{Ex}\,\mathcal{B}$. Then for each positive integer k there is a value $g(k)$ such that the following holds: each graph G which does not have as a minor a graph formed from $k+1$ vertex disjoint members of $\mathcal{B}$ contains a set B of at most $g(k)$ vertices such that $G - B$ is in $\mathcal{A}$ (that is, B is a $\mathcal{B}$-minor-blocker)[1]. In symbols we have

$$\mathrm{Ex}\,(k+1)\mathcal{B} \subseteq \mathrm{apex}^{g(k)}\,\mathcal{A}.$$

The assumption that some excluded minor is planar cannot be dropped [88].

Of course, there is an obvious containment result corresponding to (6.1), namely

$$\mathrm{Ex}\,(k+1)\mathcal{B} \supseteq \mathrm{apex}^{k}\,\mathcal{A}. \tag{7.1}$$

[1] Only the case of a single excluded minor is considered in [88] but the extension is straightforward, see also Proposition 7.3.6 below.

How much bigger is the left hand side than the right in this case? For which classes $\mathcal{B}$ is there an 'almost equality' result corresponding results of Chapter 6? Our main result provides a sufficient and essentially best possible condition for the class Ex $(k+1)\mathcal{B}$ to be 'almost apex'.

A class of graphs is *proper* if it is not the class of all graphs. Following [77] we call a minor-closed class $\mathcal{A}$ *addable* if each excluded minor is 2-connected. (Thus each such class contains all forests.) The *fan* F_j is the graph consisting of a path P_{j-1} of $j-1$ vertices together with a vertex joined to each vertex on this path. Observe that the addable class $\mathcal{F} = \text{Ex}\, C$ of forests does not contain the fan F_3: in contrast, the addable class Ex K_4 of series-parallel graphs contains all fans. The following theorem is our central result.

Theorem 7.1.1 *Let $\mathcal{A}$ be a proper minor-closed class of graphs, with set $\mathcal{B}$ of excluded minors. If $\mathcal{A}$ is addable and does not contain all fans, then for each positive integer k, as $n \to \infty$*

$$|(\text{Ex}\,(k+1)\mathcal{B})_n| = (1 + e^{-\Theta(n)})|(apex^k\, \mathcal{A})_n|. \tag{7.2}$$

On the other hand, if $\mathcal{A}$ contains all fans then this result fails; and indeed there is a constant c such that for all positive integers k and n

$$|(\text{Ex}\,(k+1)\mathcal{B})_n| \geq 2^{(k-c)n}\ |(apex^k\, \mathcal{A})_n|.$$

Let us consider a few examples illustrating this result. Recall that the number $f(k)$ in the Erdős-Pósa theorem [45] must be of order $k \ln k$. From (7.1) and Theorem 7.1.1 it follows that by removing just k vertices we can obtain:

- a forest from almost every graph with at most k disjoint cycles (Chapter 6);
- more generally, a graph without any cycles of length at least ℓ from almost every graph with at most k disjoint cycles of length at least ℓ (see also [26]);
- a collection of cacti (that is, a graph with each edge in at most one cycle) from almost every graph with at most k disjoint subdivisions of the diamond graph $D = K_4 - e$.

In contrast, by Remark 7.5.7 below, almost none of the graphs in Ex $2K_4$ can be turned into a series-parallel graph by removing one vertex.

A natural partner for this theorem is an asymptotic estimate for sizes of apex classes. Recall that a class $\mathcal{A}$ of graphs has *growth constant* $\gamma > 0$ if

$$|\mathcal{A}_n| = (\gamma + o(1))^n n! \quad \text{as } n \to \infty.$$

It is easy to see that if $\mathcal{A}$ has growth constant γ then $\text{apex}^j\, \mathcal{A}$ has growth constant $2^j\gamma$, see for example [11, 73], but we want a more precise result. Every proper addable minor-closed class of graphs has a growth constant $\gamma > 0$, see [74, 78]. For two sequences of reals (a_n) and (b_n) which are positive for n sufficiently large, we write $a_n \sim b_n$ if $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$. The next result extends Theorem 6.1.2 from Chapter 6 on forests.

Theorem 7.1.2 *Let $\mathcal{A}$ be a proper addable minor-closed class of graphs, with growth constant γ; and let k be a fixed positive integer. Then as $n \to \infty$*

$$|(\textit{apex}^k\, \mathcal{A})_n| \;\sim\; c_k 2^{kn}\; |\mathcal{A}_n|$$

where $c_k = \left(2^{\binom{k+1}{2}}\gamma^k k!\right)^{-1}$.

The above results yield asymptotic properties of typical graphs with at most k disjoint excluded minors. We state three theorems. First we note that with high probability k vertices really stand out – they each have degree about $n/2$ whereas each other vertex has much smaller degree – and they form the only minimal blocker of sublinear size. We write $R_n \in_u \mathcal{A}$ to mean that the random graph R_n is sampled uniformly from the graphs in $\mathcal{A}_n$. Thus for $R_n \in_u \text{Ex}\,(k+1)\mathcal{B}$, equation (7.2) in Theorem 7.1.1 says that R_n has a blocker of size k with probability $1 - e^{-\Theta(n)}$: the next theorem refines this result.

Theorem 7.1.3 *Let $\mathcal{A}$ be an addable minor-closed class of graphs which does not contain all fans, and let $\mathcal{B}$ be the set of excluded minors for $\mathcal{A}$. There is a constant $\delta > 0$ such that the following holds. Let k be a positive integer and let $0 < \epsilon < \frac{1}{2}$. For $n = 1, 2, \ldots$ let $R_n \in_u \text{Ex}\,(k+1)\mathcal{B}$, and let S_n be the set of vertices in R_n with degree $> \epsilon n$. Then with probability $1 - e^{-\Omega(n)}$ we have:*

(i) $|S_n| = k$ and S_n is a $\mathcal{B}$-minor-blocker in R_n;

(ii) each vertex in S_n has degree between $(\frac{1}{2} - \epsilon)n$ and $(\frac{1}{2} + \epsilon)n$; and

(iii) each $\mathcal{B}$-minor-blocker in R_n not containing S_n has size $> \delta n$.

Our second theorem on random graphs R_n concerns the clique number $\omega(R_n)$ and the chromatic number $\chi(R_n)$. Given a class $\mathcal{A}$ of graphs let

$$\omega(\mathcal{A}) = \sup\{\omega(G) : G \in \mathcal{A}\} \text{ and } \chi(\mathcal{A}) = \sup\{\chi(G) : G \in \mathcal{A}\}.$$

If $\mathcal{A}$ is a proper minor-closed class then these quantities are finite, since the average degree of the graphs in $\mathcal{A}$ is bounded, by a result of Mader [70], see also Theorem 7.22 and Corollary 5.23 in [38]. For example, if $\mathcal{A}$ is $\mathrm{Ex}\,C_4$ then $\omega(\mathcal{A}) = \chi(\mathcal{A}) = 3$ (since each block of each graph in $\mathcal{A}$ is an edge or a triangle). If also $\mathcal{A}$ is addable then we may use the 'pendant appearances theorem' of [77], restated as Lemma 7.2.1 below, to show that for $R_n \in_u \mathcal{A}$

$$\omega(R_n) = \omega(\mathcal{A}) \quad \text{and} \quad \chi(R_n) = \chi(\mathcal{A}) \quad \text{with probability } 1 - e^{-\Omega(n)}.$$

Recall that the total variation distance $d_{TV}(X, Y)$ between two random variables X and Y is the supremum over all (measurable) sets A of $|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|$. The next result shows that for $R_n \in_u \mathrm{Ex}\,(k+1)\mathcal{B}$ as defined below, the random pair consisting of $\omega(R_n)$ and $\chi(R_n)$ is very close in total variation distance to a certain simply defined pair of random variables.

Theorem 7.1.4 *Let $\mathcal{A}$ be an addable minor-closed class of graphs which does not contain all fans, and let $\mathcal{B}$ be the set of excluded minors for $\mathcal{A}$. Let k be a positive integer; let the random graph R be picked uniformly from the set of all graphs on $\{1, \ldots, k\}$; and let $X = \omega(R) + \omega(\mathcal{A})$ and $Y = \chi(R) + \chi(\mathcal{A})$. For each n let $R_n \in_u \mathrm{Ex}\,(k+1)\mathcal{B}$. Then*

$$d_{TV}\left((\omega(R_n), \chi(R_n)), (X, Y)\right) = e^{-\Omega(n)}.$$

For example, for $R_n \in_u \mathrm{Ex}\,3C_4$, both $\mathbb{P}(\chi(R_n) = 4)$ and $\mathbb{P}(\chi(R_n) = 5)$ are $\frac{1}{2} + e^{-\Omega(n)}$. Since there is only a finite range of relevant values, the result above is saying essentially that

$$\mathbb{P}\left(\omega(R_n) = i \text{ and } \chi(R_n) = j\right) = \mathbb{P}\left(X = i \text{ and } Y = j\right) + e^{-\Omega(n)}$$

for each $1 + \omega(\mathcal{A}) \le i \le j \le k + \chi(\mathcal{A})$, and the probability that the pair $(\omega(R_n), \chi(R_n))$ does not take values in this range is $e^{-\Omega(n)}$.

The third and final theorem on random graphs R_n presented here concerns connectivity. We let frag(G) denote $|V(G)|$ minus the maximum number of vertices in a component of G.

Theorem 7.1.5 *Let $\mathcal{A}$ be an addable minor-closed class of graphs which does not contain all fans; let $\mathcal{B}$ be the set of excluded minors for $\mathcal{A}$; and let $C(z) = \sum_{n\geq 1} |\mathcal{C}_n| z^n/n!$ be the exponential generating function for the class $\mathcal{C}$ of connected graphs in $\mathcal{A}$, with radius of convergence ρ. Given a positive integer k, for $R_n \in_u \operatorname{Ex}(k+1)\mathcal{B}$ we have as $n \to \infty$*

$$\mathbb{P}(R_n \text{ is connected}) \to e^{-C(\rho/2^k)}$$

and

$$\mathbb{E}\left[\mathrm{frag}(\mathrm{R_n})\right] \to (\rho/2^{\mathrm{k}})\ \mathrm{C}'(\rho/2^{\mathrm{k}}) < \infty.$$

See Chapter 6 for numerical values for these limiting probabilities in the case when $\mathcal{A}$ is the class of forests. We shall actually prove a detailed extension of this result, Theorem 7.6.1 below, concerning the limiting distribution of the unlabelled 'fragment' graph formed from the vertices not in the 'giant' component.

We now state two key intermediate results needed to prove our main theorem, Theorem 7.1.1. The first one extends a case of the Robertson-Seymour generalisation of the Erdős-Pósa theorem. Our extension asserts that, under suitable conditions, in graphs with few disjoint excluded minors there are small blockers with an additional 'redundancy' property. We write $R \cup v$ and $R \setminus v$ to denote $R \cup \{v\}$ and $R \setminus \{v\}$ respectively.

Lemma 7.1.6 *Let $\mathcal{B}$ be a set of 2-connected graphs containing at least one planar graph. Then for each integer $k \geq 1$ there is an integer $f(k)$ (depending on $\mathcal{B}$) such that the following holds. Each graph G in $\operatorname{Ex}(k+1)\mathcal{B}$ has a $\mathcal{B}$-minor-blocker R with $|R| \leq f(k)$ such that for all but at most k vertices v in R, the set $R \setminus v$ is still a $\mathcal{B}$-minor-blocker.*

The second result concerns the existence of vertex degrees of linear order. A class $\mathcal{A}$ of graphs has the *no-linear-degrees property* if, with $R_n \in_u \mathcal{A}$ as usual, for each $\delta > 0$ and each α we have $\mathbb{P}[\Delta(R_n) \geq \delta n] = O(e^{-\alpha n})$. (Here $\Delta(G)$ denotes the maximum vertex degree in G.) Observe that if $|\mathcal{A}_n| = O(\gamma^n n!)$ for some finite γ (as holds for every proper minor-closed class of graphs [43, 83]), and if $\mathcal{A}$ contains all fans, then $\mathcal{A}$ does not have this property: for there are $\frac{1}{2}(n-1)!$ fans on $[n]$ with vertex 1 as the centre vertex (for $n \geq 3$), and so

$$\mathbb{P}(\text{vertex 1 has degree } n-1 \text{ in } R_n) \geq \frac{\frac{1}{2}(n-1)!}{|\mathcal{A}_n|} = \Omega(n^{-1}\gamma^{-n}).$$

The next lemma shows that, as long as $\mathcal{A}$ is not too small, the absence of some fan yields the no-linear-degrees property.

Lemma 7.1.7 *Let the class $\mathcal{A}$ of graphs satisfy* $\liminf \left(\frac{|\mathcal{A}_n|}{n!}\right)^{1/n} > 0$*, and suppose that for some positive integer j, no graph in $\mathcal{A}$ contains the fan F_{j+2} as a minor. Then $\mathcal{A}$ has the no-linear-degrees property.*

The plan of this chapter is as follows. In Section 7.2 we count apex graphs and prove Theorem 7.1.2: this work needs no preliminaries.

Section 7.3 concerns redundant blockers. First we introduce a useful theorem of Kloks which relates normal trees and tree decompositions (and we provide a proof). Then we give two structural lemmas on normal trees and small 'splitting sets'. We use those lemmas to prove a result related to the Robertson-Seymour generalisation of the Erdős-Pósa theorem and then we prove Lemma 7.1.6.

In the next section, Section 7.4, we consider the no-linear-degrees property and prove Lemma 7.1.7. Following that, in Section 7.5, we complete the proof of Theorem 7.1.1. In Section 6.5 we use our main results to prove the theorems on properties of the random graph $R_n \in_u \text{Ex}\,(k+1)\mathcal{B}$. Finally, we make some concluding remarks.

7.2 Counting apex classes

We shall use the 'pendant appearances theorem', Theorem 4.1 from [77], several times, so for convenience we state here a suitable special case as a lemma.

Let H be a connected graph on the vertex set $\{1,\ldots,h\}$ which we consider to be rooted at vertex 1, and let G be a graph on the vertex set $\{1,\ldots,n\}$, where $n > h$. Then an induced subgraph $\tilde{H}$ of G is a *pendant appearance* of H if (a) the increasing bijection from $\{1,\ldots,h\}$ to $V(\tilde{H})$ gives an isomorphism between H and $\tilde{H}$; and (b) there is exactly one edge in G between $V(\tilde{H})$ and the rest of G, and this edge is incident with the vertex of $\tilde{H}$ with smallest label.

Lemma 7.2.1 *([77]) Let $\mathcal{A}$ be a proper addable minor-closed class of graphs, and let H be a connected graph in $\mathcal{A}$. There is a constant $a > 0$ such that the following holds. For $R_n \in_u \mathcal{A}$, with probability $1 - e^{-\Omega(n)}$ R_n has at least $a \cdot n$ disjoint pendant appearances of H.*

Proof of Theorem 7.1.2 Since $\mathcal{A}$ is proper minor-closed and addable, by Theorem 1.2 of [74] $\mathcal{A}$ is *smooth* with some growth constant $\gamma > 0$, that is

$|\mathcal{A}_n|/\left(n|\mathcal{A}_{n-1}|\right) \to \gamma$ as $n \to \infty$. Hence

$$|\mathcal{A}_n| \sim (n)_k\, \gamma^k\, |\mathcal{A}_{n-k}|.$$

Let $n > k$, let $V = \{1, \ldots, n\}$, and consider the following constructions of graphs on V:

(1) Choose a k-set $S \subseteq V$, and put any graph on S ($\binom{n}{k} 2^{\binom{k}{2}}$ choices).

(2) Put any graph $F \in \mathcal{A}$ on $V \setminus S$ ($|\mathcal{A}_{n-k}|$ choices).

(3) Add the edges of any bipartite graph B with parts S and $V \setminus S$ ($2^{k(n-k)}$ choices).

Clearly each graph constructed is in $(\text{apex}^k\, \mathcal{A})_n$, and each graph in $(\text{apex}^k\, \mathcal{A})_n$ is constructed at least once. The number of constructions is

$$\binom{n}{k} 2^{\binom{k}{2}} 2^{k(n-k)} |\mathcal{A}_{n-k}| \sim c_k 2^{kn} |\mathcal{A}_n|$$

so $|(\text{apex}^k\, \mathcal{A})_n|$ is at most this number.

Let us bound $|(\text{apex}^k\, \mathcal{A})_n|$ from below by showing that almost all of the constructions yield distinct graphs. Observe that $G \in (\text{apex}^k\, \mathcal{A})_n$ appears just once if and only if G has a unique k-set S of vertices such that $G - S$ is in $\mathcal{A}$.

Let $\mathcal{B}$ be the set of excluded minors for $\mathcal{A}$ and fix a graph $H \in \mathcal{B}$ (which must be 2-connected). Let $h = |V(H)|$. Fix a vertex v in H, and let H^- be the connected graph $H - v$. Let us say that a graph $G \in (\text{apex}^k\, \mathcal{A})_n$ is *good* if for some k-set S_0 it satisfies the following: (a) $G - S_0 \in \mathcal{A}$; and (b) for each vertex $s \in S_0$ there are $k+1$ pairwise disjoint sets $X_1(s), X_2(s), \ldots, X_{k+1}(s) \subseteq V(G) \setminus S_0$ such that each induced subgraph $G[X_i(s) \cup s]$ has a minor H. If G is good then S_0 must be the unique k-set S such that $G - S$ is in $\mathcal{A}$. For if S' is another set such that $G - S'$ is in $\mathcal{A}$, and $w \in S_0 \setminus S'$, then S' must contain a vertex from each of the sets $X_1(w), X_2(w), \ldots, X_{k+1}(w)$, and so $|S'| \geq k+1$.

Now by Lemma 7.2.1 (the 'pendant appearances theorem') there exist constants $a > 0$ and $b > 0$ such that the following holds for a random graph $R_n \in_u \mathcal{A}$: R_n contains at least $a \cdot n$ pairwise vertex-disjoint copies of H^- with probability at least $1 - e^{-bn}$ for n sufficiently large. If F has at least $a(n-k)$ such copies of H^- then there are at most

$$2^{k(n-k)} k\, \mathbb{P}\left(\text{Bin}(\lceil a(n-k) \rceil, 2^{-h+1}) \leq k\right)$$

ways to choose the bipartite graph B with parts S and $V \setminus S$ so that the resulting graph is not good. So the number of ways to choose the graphs F and B so that the resulting graph is constructed just once is at least

$$\begin{aligned} &2^{k(n-k)}|\mathcal{A}_{n-k}| \left(1 - e^{-b(n-k)} - k\mathbb{P}\left(\mathrm{Bin}(\lceil a(n-k)\rceil, 2^{-h+1}) \leq k\right)\right) \\ &= 2^{k(n-k)}|\mathcal{A}_{n-k}| \left(1 - e^{-\Omega(n)}\right), \end{aligned}$$

by a Chernoff bound. Summing over all sets S and all graphs on S we obtain

$$|\mathrm{apex}^k\,(\mathcal{A})_n| \geq c_k 2^{kn}|\mathcal{A}_n| \left(1 - e^{-\Omega(n)}\right),$$

as required. □

7.3 Redundant blockers

In this section, after some preliminary results we prove the 'redundant blockers lemma', Lemma 7.1.6.

7.3.1 Treewidth and normal trees

We prove Kloks'es theorem, Theorem 5.2.3 stated in Section 5.2.2.

Proof of Theorem 5.2.3 ($\geq$) Let G have treewidth k. We shall prove that for any given vertex $s_0 \in V(G)$ there is a normal tree T for G, rooted at s_0, such that

$$\max_{v \in V(G)} a_T(v) \leq k. \tag{7.3}$$

Let $(T_0, (V_t : t \in T_0))$ be a tree decomposition for G with $|V_t| = k+1$ for each node t in T_0, and with $|V_s \setminus V_t| = 1$ for each edge st in T_0 (it is easy to see that such a tree decomposition always exists, see for example [38]). We call V_t the *bag* for t.

For convenience we shall consider the following small modification of T_0. Pick a node u of T_0 with $s_0 \in V_u$. Suppose that $V_u = \{s_0, s_1, \ldots, s_k\}$. Let $u_0, u_1, \ldots, u_k$ be a path on $k+1$ new nodes, and identify u_k with u. Let $V_{u_j} = \{s_0, \ldots, s_j\}$ for each $j = 0, 1, \ldots, k$. Let T_1 be the tree we have formed from T_0 by adjoining the path, and note that $(T_1, (V_t : t \in T_1))$ is also a tree decomposition for G. Root T_1 at u_0.

The set V_{u_0} consists of the single vertex s_0: define $v(u_0) = s_0$. For each node t in T_1 other than u_0 denote its parent in T_1 by t'; and let $v(t)$ be the unique vertex in $V_t \setminus V_{t'}$. It is a property of tree decompositions that the nodes corresponding to the bags that contain $w \in V(G)$ form a subtree of T_1, which we call the tree for w. It follows that $v(s) \neq v(t)$ for distinct nodes s and t in T_1, since $v(s) = v(t)$ would imply that both V_s and V_t contain $v(s)$ but there is a node on the path from s to t in T_1 whose bag does not contain $v(s)$. Also, for each vertex v in G there is a node t in T_1 such that $v(t) = v$ (the node t with $v \in V_t$ which is nearest to the root). Thus the map $t \to v(t)$ gives a bijection between the nodes of T_1 and the vertices of G. Let T be the tree on $V(G)$ which corresponds to T_1 under this map; that is, let T be the tree on $V(G)$ with an edge $v(t)v(t')$ for each edge tt' in T_1, rooted at s_0. We claim that T is a normal tree for G and (7.3) holds.

To see that T is a normal tree for G, consider two vertices x and y which are incomparable in T. Let t_x and t_y be the nodes of T_1 with $v(t_x) = x$ and $v(t_y) = y$ respectively. Then t_x and t_y are incomparable in T_1, so the trees for x and y do not meet in T_1 and thus x and y are not adjacent in G.

It remains to prove (7.3). Fix a vertex x in G, and let t_x be the node of T_1 with $v(t_x) = x$. If $y \in \mathrm{AA}_T(x)$ then $y \in V_x \setminus x$; and so $a_T(x) = |\mathrm{AA}_T(x)| \leq |V_x| - 1 \leq k$, as required.

($\leq$) Let T be a normal tree for G, and let $\max_{v \in V} a_T(v) = k$. For each vertex x define the bag V_x as $\mathrm{AA}_T(x) \cup x$. We claim that $(T, (V_x : x \in V(T))$ is a tree decomposition of G, of width at most k. Certainly each bag V_x satisfies $|V_x| \leq k + 1$.

Let uv be an edge of G. Then u and v are comparable in T: without loss of generality suppose that $u <_T v$. So $u \in \mathrm{AA}_T(v)$, and thus both u and v are in the bag V_v. It remains to check that for each vertex w of G the nodes t of T whose bag V_t contains w form a subtree of T. But if $w \in \mathrm{AA}_T(v)$ then $w \in \mathrm{AA}_T(u)$ for each vertex u (other than w) on the path in T between w and v. This completes the proof. □

Note that the above proof yields a normal tree with the property (7.3) rooted at an arbitrary vertex in $V(G)$. Hence we can take the set $\mathcal{T}$ in (5.1) to be the set of all normal trees on $V(G)$ rooted at any chosen vertex $r \in V(G)$.

Note also that Theorem 5.2.3 fails if we additionally require the normal trees in $\mathcal{T}$ to be subgraphs of G. For example, consider a complete bipartite graph $K_{n,n}$ with $n \geq 3$. The treewidth of $K_{n,n}$ is n, but each normal tree T for $K_{n,n}$ which is its subgraph must be a path where the vertex w at distance $2n - 2$ from the root has $a_T(w) = 2(n - 1) > n$.

7.3.2 Disjoint subgraphs, splitting sets, treewidth and blockers

In this part we work with classes $\mathcal{H}$ of graphs which are closed under isomorphism but not necessarily minor-closed. Given a class $\mathcal{H}$ of graphs, denote by Forb $(\mathcal{H})$ the class $\mathcal{A}$ of all graphs G such that no subgraph of G is in $\mathcal{H}$. Also, call a set B of vertices such that $G - B \in$ Forb $(\mathcal{H})$ an *$\mathcal{H}$-subgraph-blocker*.

Let T be a normal tree and let H be a connected graph on a subset of $V(T)$. For a set A of vertices we say that A *splits* H in T if there is a vertex $v \in A$ such that either $v \in V(H)$ or H contains vertices above v and below v in T.

We state and prove two general lemmas regarding graphs without k disjoint subgraphs belonging to some class $\mathcal{H}$ of connected graphs. Their proofs have a similar structure and the proof of the former is a good warm-up for the proof of the latter. At the end of this section we also present an application of Lemma 7.3.1.

Lemma 7.3.1 *Let $\mathcal{H}$ be a non-empty class of connected graphs, let $k \geq 0$ be an integer, let $G \in$ Forb $(k+1)\mathcal{H}$, and let T be a normal tree for G. Then there is a set A of at most k vertices such that, for each subgraph $H \in \mathcal{H}$ of G, the set A splits H in T.*

Proof We will use induction on k. The statement is trivially true for the case $k = 0$, with $A = \emptyset$. Let $j \geq 1$ and suppose that the statement holds for $k = j - 1$. Let $G \in$ Forb $(j+1)\mathcal{H}$, and let T be a normal tree for G. Denote the subtree of T rooted at v by T_v. Let

$$B := \{v \in V(G) : G[V(T_v)] \text{ has a subgraph in } \mathcal{H}\}.$$

If $B = \emptyset$ then $G \in$ Forb $(\mathcal{H})$ so we may take $A = \emptyset$: thus we may assume that $B \neq \emptyset$. Consider a vertex $u \in B$ at maximum distance in T from the root r. If $u = r$ then $B = \{r\}$ so every subgraph of G in $\mathcal{H}$ must contain r. In this case we may take $A = \{r\}$: thus we may assume that $u \neq r$.

Let $G' = G - V(T_u)$ and let $T' = T - V(T_u)$. Since $G[V(T_u)]$ has a subgraph in $\mathcal{H}$, we have $G' \in$ Forb $j\mathcal{H}$. Clearly, T' is a normal tree for G'. Apply induction for G' and T' to obtain a set of at most $j-1$ vertices $A' \subseteq V(G')$ such that A' splits in T' each subgraph H of G' such that $H \in \mathcal{H}$.

Now let $A = A' \cup u$. We will show that A has the required property for G and T. Suppose H is a subgraph of G and $H \in \mathcal{H}$. If $V(H) \cap V(T_u)$ is empty then H is a subgraph of G', so there is a vertex $v \in A' \subseteq A$ such that H either contains

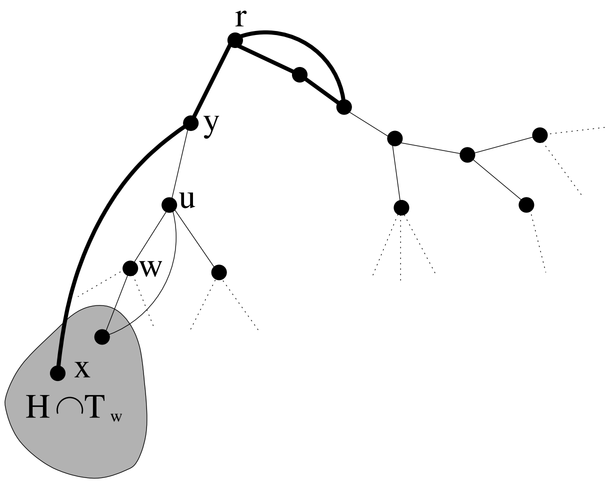

Figure 7.1: If $V(H) \cap [V(T_u) \setminus u] \neq \emptyset$ then H must have an edge between this set and some vertex $y \in AA_T(u)$. In this illustration H consists of a subgraph strictly below u together with the bold edges. In our terminology, $\{u\}$ splits H in T.

v or contains vertices both above and below v in T' and so in T. Thus it suffices to consider the case when $V(H) \cap V(T_u)$ is not empty.

We may assume that H does not contain u, as otherwise we are done. Let w be a child of u in T such that T_w contains a vertex in H. Then since $w \notin B$, T_w does not contain all of $V(H)$. Since H is connected, there must be an edge between a vertex x of H in T_w and a vertex y of H not in T_w; and since T is normal, y must be in $AA_T(u)$. But now x is below u and y is above u in T, and the proof is complete. □

We now assume that the class $\mathcal{H}$ consists of 2-connected graphs. The following lemma will be crucial in the proof of Lemma 7.3.8, and thus in the proof of Lemma 7.1.6. It asserts that any $\mathcal{H}$-subgraph-blocker can be transformed into one with a specific 'redundant' structure by adding a few extra vertices.

Lemma 7.3.2 *Let $\mathcal{H}$ be a non-empty set of 2-connected graphs. Let $k \geq 0$, let $G \in \text{Forb}\,(k+1)\mathcal{H}$, let $Q \subseteq V(G)$ be an $\mathcal{H}$-subgraph-blocker in G, and let T be a normal tree for $G-Q$. Then there are sets $S \subseteq Q$ with $|S| \leq k$ and $A \subseteq V(G-Q)$ with $|A| \leq k$, such that for each vertex $x \in Q \setminus S$ and each subgraph H of $G-(Q \setminus x)$ in $\mathcal{H}$ the set A splits $H - x$ in T.*

To read the last sentence, it may help to observe that, given $x \in Q \setminus S$ and a subgraph H of $G - (Q \setminus x)$ in $\mathcal{H}$, we must have $V(H) \cap Q = \{x\}$.

Proof The proof is similar to the proof of Lemma 7.3.1, except that in this case we use induction on k to obtain the set S together with the set A.

Clearly the case $k = 0$ holds, as we may take $A = S = \emptyset$. Let $j \geq 1$ and suppose that the result holds for the case $k = j - 1$. Let $G \in \text{Forb}\,(j+1)\mathcal{H}$, let $Q \subset V(G)$ be an $\mathcal{H}$-subgraph-blocker in G, and let T be a normal tree for $G - Q$. Let

$$B = \{v \in V(G) : \; G[V(T_v) \cup x] \notin \text{Forb}\,\mathcal{H} \text{ for some } x \in Q\}.$$

If B is empty then we are done (again take $A = S = \emptyset$, and note that there are no relevant graphs H); so assume that B is non-empty. Choose a vertex $u \in B$ with maximum distance from the root r in T.

Consider first the case $u = r$. Let $H \in \mathcal{H}$ be a subgraph of G with $V(H) \cap Q = \{x\}$. Since $u = r$, the vertices of the connected graph $H - x$ are not contained in $V(T_v)$ for any subtree T_v where v is a child of r. Also, since T is normal, there are no edges between subtrees T_v and $T_{v'}$ for distinct children of r. Hence H must contain r. Thus each subgraph in $\mathcal{H}$ of G which meets Q in just one vertex must contain r. Hence we may take $S = \emptyset$ and $A = \{r\}$.

We may now assume that $u \neq r$. Let $z \in Q$ be such that $G[V(T_u) \cup z] \notin \text{Forb}\,\mathcal{H}$. Let $G' = G - (V(T_u) \cup z)$, let $Q' = Q \setminus z$, and let $T' = T - V(T_u)$. Then clearly $G' \in \text{Forb}\, j\mathcal{H}$, Q' is an $\mathcal{H}$-subgraph-blocker in G', and T' is a normal tree for G'. Hence we can apply the induction hypothesis to G', Q' and T'. We obtain sets $S' \subseteq Q'$ with $|S'| \leq j - 1$ and $A' \subseteq V(G' - Q')$ with $|A'| \leq j - 1$, such that for each vertex $x \in Q' \setminus S'$ if $H \in \mathcal{H}$ is a subgraph of $G' - (Q' \setminus x))$ then A' splits $H - x$ in T'.

Now let $S = S' \cup z$ and $A = A' \cup u$. Let $x \in Q \setminus S$, and suppose that the subgraph H of $G - ((Q \setminus x) \cup A)$ is in $\mathcal{H}$. Note that $u \in A$ so $u \notin V(H)$. If $V(H) \cap V(T_u) = \emptyset$ then H is a subgraph of $G' - ((Q' \setminus x) \cup A')$: hence there is a vertex $v \in A' \subseteq A$ for which H has vertices above and below v in T' and so in T. This leaves the case that $V(H) \cap V(T_u) \neq \emptyset$. Suppose that H has no vertex above u in T: we want to find a contradiction.

Now $V(H - x)$ cannot be contained in $V(T_v)$ for any subtree T_v where v is a child of u, as this would imply that $v \in B$ which would contradict our choice of u. But as in the case $u = r$, since T is normal the connected graph $H - x$ cannot have vertices in subtrees T_v and $T_{v'}$ of T_u where v and v' are distinct children of u. Thus we have a contradiction, and the proof is complete. □

A variant of the following lemma was first proved by Robertson and Seymour [88] in order to prove the generalised Erdős-Pósa theorem. We state it in a more general form (proved by Thomassen [98]) and give a simple proof using Theorem 5.2.3 and Lemma 7.3.1.

Lemma 7.3.3 *Let $\mathcal{H}$ be a class of connected graphs and let k and w be nonnegative integers. If $G \in \text{Forb}\,(k+1)\mathcal{H}$ and $tw(G) \le w$ then G has an $\mathcal{H}$-*subgraph-blocker *of size at most $k(w+1)$.*

Proof of Lemma 7.3.3 Suppose $G \in \text{Forb}\,(k+1)\mathcal{H}$ and $tw(G) \le w$. By Theorem 5.2.3, there is a normal tree T for G such that $\max_{v \in V} a_T(v) \le w$. Let A be a set of at most k vertices as in Lemma 7.3.1, and let

$$B = A \cup (\cup_{v \in A} \text{AA}_T(v)).$$

Observe that $|B| \le k(w+1)$. We claim that B is an $\mathcal{H}$-subgraph-blocker in G. For suppose it is not, and let $H \in \mathcal{H}$ be a subgraph of $G - B$. By Lemma 7.3.1 there is a vertex $v \in A$ such that H contains vertices both above v and below v in T (note that $v \notin V(H)$ since $A \cap V(H) = \emptyset$). Since H is connected, it has an edge xy with x above v and y below v. But then $x \in \text{AA}_T(v) \subseteq B$, a contradiction. □

7.3.3 Treewidth and blockers: a more general case

Lemma 7.3.3 is what we need in this paper, to prove Lemma 7.1.6; but it does not apply to disconnected excluded subgraphs. We include for completeness a treatment of this case, and give a more general version of Lemma 7.3.3.

We give two preliminary lemmas. The proof of the first one uses induction much as in [88], but as in the proof of Lemma 7.3.1 we use normal trees. For any graph H we let $0H$ denote the graph with no vertices: thus for any graph H we have $G \cup 0H = G$.

Lemma 7.3.4 *For $t \ge 1$ let $H_1, H_2, \ldots, H_t$ be connected graphs. Let $k_1, k_2, \ldots, k_t$ be nonnegative integers, not all zero. If*

$$G \in \text{Ex}\ (k_1 H_1 \cup k_2 H_2 \cup \cdots \cup k_t H_t)$$

and T is a normal tree for G, then there is a set $A \subseteq V(G)$ and an integer j with $k_j \ge 1$, such that $|A| \le (\sum_i k_i) - 1$ and A splits in T each connected subgraph of G with a minor H_j.

Proof We use induction on $\sum_i k_i$. If $\sum_i k_i = 1$, then $G \in \text{Ex}\, H_j$ for some j: so we may take $A = \emptyset$ and we are done.

Let $s \ge 2$, suppose we have proved the hypothesis for each sequence $k'_1, k'_2, \ldots, k'_t$ with $\sum_i k'_i < s$, and let $\sum_i k_i = s$. We proceed as in the proof of Lemma 7.3.1.

Let B be the set of all vertices v of T such that $G[V(T_v)] \notin \mathrm{Ex}\,H_i$ for some i with $k_i \geq 1$, and let u be a vertex in B with maximum distance from the root.

Suppose first that u is the root of T. Then we may take $A = \{u\}$ as u must be in every subgraph of G with a minor in $\{H_1, \ldots, H_t\}$. So we may assume that u is not the root of T.

Let j be such that $k_j \geq 1$ and $G[V(T_u)] \notin \mathrm{Ex}\,H_j$. Write $T' = T - T_u$ and let $k'_i = k_i$, for $i \neq j$ and $k'_j = k_j - 1$. Since

$$G' := G - T_u \in \mathrm{Ex}\ (k'_1 H_1 \cup \cdots \cup k'_t H_t)\,,$$

T' is a normal tree for G' and $\sum_i k'_i = s - 1$, we may apply induction to find some l with $k'_l \geq 1$ and a set A' of at most $s-2$ vertices that splits in T' each connected subgraph H of G' with a minor H_l. We claim that $A = A' \cup u$ splits in T every connected subgraph of G with a minor H_l. As in the proof of Lemma 7.3.1, every connected subgraph H of G with a minor H_l such that H is not a subgraph of G' and $u \notin V(H)$ must have vertices both in $V(T_u)$ and $V(G')$ (or otherwise we get a contradiction to the choice of u). But then $\{u\}$ splits in T each such subgraph H. □

Lemma 7.3.5 *Let $\mathcal{A}$ be a minor-closed class of graphs (perhaps the class of all graphs), and let $\mathcal{H}$ consist of the graphs not in $\mathcal{A}$ together with an arbitrary class of connected graphs. Let s be the sum over the disconnected excluded minors M for $\mathcal{A}$ of the number $\kappa(M)$ of components (so $0 \leq s < \infty$). Let $k \geq 0$ be an integer, let $G \in \mathrm{Forb}\,(k+1)\mathcal{H}$, and let T be a normal tree for G. Then there is a set A of at most $s(k+1)+k$ vertices such that for each subgraph $H \in \mathcal{H}$ of G the set A splits in T some component of H.*

Proof Let $\tilde{\mathcal{H}}$ consist of the connected graphs in $\mathcal{H}$. By Lemma 7.3.1, there is a set $A_0 \subseteq V(G)$ with $|A_0| \leq k$ such that A_0 splits in T each subgraph $H \in \tilde{\mathcal{H}}$ of G.

Suppose that $\mathcal{A}$ has $j \geq 0$ disconnected excluded minors G_i, $i = 1, \ldots, j$. Let $i \in \{1, \ldots, j\}$. Since $G \in \mathrm{Ex}\,(k+1)G_i$, by Lemma 7.3.4, there is a set $A_i \subseteq V(G)$ with $|A_i| \leq \kappa(G_i)(k+1) - 1$ such that A_i splits in T some component of each subgraph G' of G which has a minor G_i, that is, of each G' in the complement $(\mathrm{Ex}\,G_i)^c$ of $\mathrm{Ex}\,G_i$.

Finally, observe that $\mathcal{H}$ is the union of $\tilde{\mathcal{H}}$ and $\cup_i(\mathrm{Ex}\,G_i)^c$, and so we may form a set A as required from the union of the $j+1$ sets A_i. □

Now from the above lemma and Theorem 5.2.3 following the lines of the proof of Lemma 7.3.3 we have:

Proposition 7.3.6 *Let $\mathcal{A}$ be a minor-closed class of graphs and let $\mathcal{H}$ consist of the graphs not in $\mathcal{A}$ together with an arbitrary class of connected graphs. Then for each pair of non-negative integers k and w there is an integer $f(k, w)$ such that if $G \in \mathrm{Forb}\,(k+1)\mathcal{H}$ and $tw(G) \leq w$ then G has an $\mathcal{H}$-subgraph-blocker of size at most $f(k, w)$.*

7.3.4 Proof of Lemma 7.1.6

We introduce the following fundamental result, Theorem (2.1) of Robertson and Seymour [88].

Lemma 7.3.7 *For every planar graph H, there is a number $\alpha(H)$ such that every graph with no minor H has treewidth at most $\alpha(H)$.*

We now prove the redundant blockers lemma, Lemma 7.1.6 (see Figure 2), quickly using the above results. We give a slightly more general version first.

Lemma 7.3.8 *Fix a 2-connected planar graph H_0. Then for each integer $k \geq 1$ there is an integer $f(k)$ such that the following holds.*

Let $\mathcal{H}$ be a set of 2-connected graphs such that $\mathcal{H}$ contains H_0 and all 2-connected graphs contractible to H_0. Then each graph G in $\mathrm{Forb}\,(k+1)\mathcal{H}$ has an $\mathcal{H}$-subgraph-blocker R with $|R| \leq f(k)$ such that for all but at most k vertices v in R, the set $R \setminus v$ is still an $\mathcal{H}$-subgraph-blocker.

Proof of Lemma 7.3.8 If a graph contains H_0 as a minor then it contains a 2-connected subgraph contractible to H_0. Thus $\mathrm{Forb}\,k\mathcal{H} \subseteq \mathrm{Ex}\,kH_0$ for each positive integer k.

Let $k \geq 1$ and let $G \in \mathrm{Forb}\,(k+1)\mathcal{H}$. Then $G \in \mathrm{Ex}\,(k+1)H_0$ and $(k+1)H_0$ is planar; so by Lemma 7.3.7, G has treewidth at most w for some constant $w = w(k, H_0)$. Therefore by Lemma 7.3.3 there is a positive integer $g(k)$ (depending only on k and H_0) such that each graph in $\mathrm{Forb}\,(k+1)\mathcal{H}$ has an $\mathcal{H}$-subgraph-blocker Q of size at most $g(k)$.

Now $G - Q \in \mathrm{Forb}\,\mathcal{H} \subseteq \mathrm{Ex}\,H_0$. Using Lemma 7.3.7 again, we see that $G - Q$ has treewidth at most α for some constant $\alpha = \alpha(H_0)$; and so it has a normal tree T with $a_T(v) \leq \alpha$ for each vertex $v \in V(T)$ by Theorem 5.2.3. Let A and S

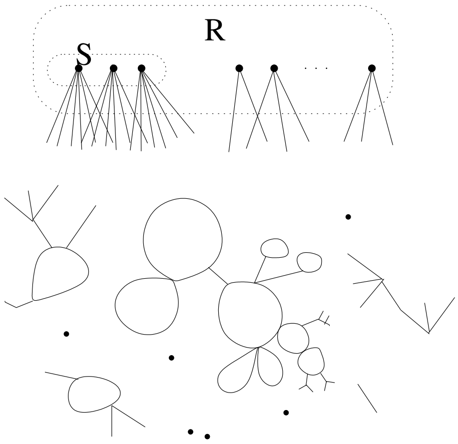

Figure 7.2: For example, it follows easily from Lemma 7.1.6 that we can always decompose a graph G in Ex $4D$ into a set R of a constant size and a collection of cacti, where each vertex in R, except at most 3 of them, can have at most 2 edges to each component of $G - R$. (Here D is the diamond graph $K_4 - e$.)

be sets obtained in Lemma 7.3.2 for k, G, Q and T. Define $R \subseteq V(G)$ by

$$R := Q \cup A \cup \left(\bigcup_{v \in A} \mathrm{AA}_T(v) \right).$$

Note that $|R| \leq f(k) := g(k) + (\alpha + 1)k$. We want to show that $R - u$ is still an $\mathcal{H}$-subgraph-blocker for G for each vertex $u \in R \setminus S$. This is clearly true for $u \in R \setminus Q$; so assume that $u \in Q \setminus S$, and some subgraph H of $G - (R \setminus u)$ is in $\mathcal{H}$. But then H is a subgraph of $G - ((Q \setminus u) \cup A)$, and so by Lemma 7.3.2, for some vertex $w \in A$, H must have vertices both above and below w in T. Hence the connected subgraph $H - u$ of $G - R$ has vertices both above and below w in T. But by the definition of normal tree, $AA_T(w)$ is a separating set for the vertices below w and the rest of T, so $H - u$ contains a vertex in $AA_T(w) \subset R$, a contradiction. □

Proof of Lemma 7.1.6 Let $\mathcal{H}$ be the class of all 2-connected graphs contractible to a graph in $\mathcal{B}$. Then Forb $\mathcal{H}$ = Ex $\mathcal{B}$, and more generally Forb $(k+1)\mathcal{H}$ = Ex $(k+1)\mathcal{B}$ for each $k \geq 0$. Also, a $\mathcal{B}$-blocker is the same as an $\mathcal{H}$-subgraph-blocker. Now the result follows directly from Lemma 7.3.8. □

7.4 Graph classes not containing all fans

In this section, after a preliminary lemma on coloured forests, we prove Lemma 7.1.7 on the no-linear-degrees property.

For graphs where each vertex is coloured black or white, we define the contraction operation as usual, but colour black each vertex resulting from contracting a connected graph which contains a black vertex. A forest is *rooted* if in each component tree a vertex is distinguished as the root: often we will think of the edges as being oriented away from the root. Let us say that a class $\mathcal{A}$ of graphs is *very small* (or $\mathcal{A}$ has growth constant 0) if

$$\left(\frac{|\mathcal{A}_n|}{n!}\right)^{\frac{1}{n}} \to 0 \quad \text{as } n \to \infty.$$

For example, the class of graphs with no path of length j is very small [11]: we shall use this result in the proof of the next lemma. Recall that the path P_j has j vertices.

Lemma 7.4.1 *Fix an integer $j \geq 2$ and $0 < \alpha \leq 1$. Let $\mathcal{F}'$ be the class of black/white coloured rooted forests such that no forest in $\mathcal{F}'$ has a black path P_j as a minor, and each forest in $\mathcal{F}'_n$ has at least αn black vertices. Then $\mathcal{F}'$ is very small.*

Proof Call a colouring of a forest as above *good*, and call a good colouring (black-) *maximal* if there is no vertex coloured white such that recolouring it black gives another good colouring. Let $\hat{\mathcal{F}}$ be the set of all possible (unrooted) forests F together with a specified maximal good colouring of F. Since $|\mathcal{F}'_n| \leq 4^n|\hat{\mathcal{F}}_n|$, it suffices for us to prove that $\hat{\mathcal{F}}$ is very small. (For an n-vertex forest, there are at most 2^n choices for the colouring and at most 2^n choices for the roots.)

Let $F \in \hat{\mathcal{F}}_n$. Observe that trimming off a white leaf yields another coloured forest in $\hat{\mathcal{F}}$. The *core* of F is the coloured forest obtained by repeatedly trimming off white leaves until none remain. Thus $core(F)$ is in $\hat{\mathcal{F}}_{n_1}$ for some $n_1 \geq \alpha n$.

The key observation is that in $core(F)$ each white vertex has degree 2. For suppose that vertex v is white and has degree at least 3. Let us change the colour of v to black. By the maximality of the colouring, $core(F)$ now has a black P_j minor. There must be a vertex w of the minor to which v is contracted; and if T_v denotes the tree in $core(F)$ containing v which is contracted to w, then with the original colour of v each vertex in T_v is white (for otherwise the colour change would not have mattered). Thus there are at least 3 edges between T_v and subtrees of the rest of $core(F)$, which each contain a black vertex. At least one of these subtrees must be entirely deleted when the minor is formed (since the minor has maximum degree 2): but if instead we added such a subtree to T_v and contracted all these vertices to form w then w would be black even without the colour change. Thus $core(F)$ has a black minor P_j with the original colouring, and this contradiction shows that indeed each white vertex in $core(F)$ has degree 2.

We have now seen that each coloured forest $F \in \hat{\mathcal{F}}$ may be obtained from an all black forest with no path P_j by subdividing edges using white vertices, and then attaching pendant trees using more white vertices (to reverse the initial operation of repeatedly trimming off white leaves). Thus the following constructions yield each member of $\hat{\mathcal{F}}_n$ at least once.

1. Choose a set $V_1 \subseteq V = [n]$ of $n_1 \geq \alpha n$ vertices; colour each of these vertices black; and choose a forest F' on V_1 which does not contain a path P_j.

2. Choose a set $V_2 \subseteq V \setminus V_1$ of $n_2 \geq 0$ vertices; colour each of these vertices white; and use them to subdivide edges in F'.

3. Take the remaining set $V_3 = V \setminus (V_1 \cup V_2)$ of vertices; colour each of these vertices white; and use them to form $n_1 + n_2$ trees T_x rooted at the vertices in $V_1 \cup V_2$.

4. The edges of the coloured forest F are those of the subdivided forest F' together with those of the trees T_x.

Let us bound the number of constructions. For step 1, let $\tilde{f}_{n_1}$ denote the number of forests on $[n_1]$ which do not contain a path P_j. Now consider step 2, after we have chosen the set V_2. List the $t \leq n_1 - 1$ edges of F' in some fixed order, with a fixed orientation: then we see that the number of ways to use the n_2 vertices in V_2 to subdivide the edges of F' is $(n_2+t-1)!/(t-1)! = (n_2+t-1)_{n_2} \leq n^{n_2}$. For step 3, recall that the number of forests on the vertex set V containing exactly $n_1 + n_2$ rooted trees with given roots is $(n_1+n_2)n^{n-n_1-n_2-1} \leq n^{n-n_1-n_2}$.

From the above we see that in total the number of constructions is at most

$$\begin{aligned} & \sum_{n_1=\lceil \alpha n \rceil}^{n} \binom{n}{n_1} \tilde{f}_{n_1} \sum_{n_2=0}^{n-n_1} \binom{n-n_1}{n_2} n^{n_2} n^{n-n_1-n_2} \\ = \; & (2n)^n \sum_{n_1=\lceil \alpha n \rceil}^{n} \binom{n}{n_1} \tilde{f}_{n_1} (2n)^{-n_1}. \end{aligned}$$

Now let $\epsilon > 0$, and suppose that $\epsilon \leq \frac{1}{e}$ so that $\sum_{m \geq 0} (e\epsilon/2)^m \leq 2$. From [11], there is a m_0 such that for each $m \geq m_0$ we have $\tilde{f}_m \leq (\epsilon m)^m$. Hence for all n

sufficiently large that $\alpha n \geq m_0$, the number of constructions is at most

$$
\begin{aligned}
& (2n)^n \sum_{n_1=\lceil \alpha n \rceil}^{n} \binom{n}{n_1} (\epsilon n_1)^{n_1} (2n)^{-n_1} \\
\leq \; & (2n)^n \sum_{n_1=\lceil \alpha n \rceil}^{n} \left(\frac{ne}{n_1} \cdot \frac{\epsilon n_1}{2n} \right)^{n_1} \\
\leq \; & (2n)^n \sum_{m \geq \lceil \alpha n \rceil} (e\epsilon/2)^m \leq 2 \cdot (2^{1-\alpha}(e\epsilon)^\alpha n)^n.
\end{aligned}
$$

This completes the proof, since we may make $2^{1-\alpha}(e\epsilon)^\alpha$ arbitrarily small by choice of ϵ. □

We have already seen that each proper minor-closed class of graphs that contains arbitrarily large fans fails to have the no-linear-degrees property. We now use Lemma 7.4.1 to prove Lemma 7.1.7, which says that excluding some fan ensures that a suitable class has the no-linear-degrees property.

Proof of Lemma 7.1.7 For a class $\mathcal{A}$ and a random graph $R_n \in_u \mathcal{A}$ the no-linear-degrees property is equivalent to the requirement that

$$
\mathbb{P}(d(v_0) \geq cn) = o(e^{-an}) \quad \text{for every } c, a > 0
$$

where v_0 is vertex 1 and $d(v_0)$ is its degree in the graph R_n.

Let j be a positive integer, and suppose that no graph in $\mathcal{A}$ contains as a minor the fan F_{j+2}. Fix $c \in (0, 1]$. We are going to bound the number of graphs in $\mathcal{A}_n$ such that $d(v_0) \geq cn$.

Consider a graph $G \in \mathcal{A}_n$. Let the least vertex in each component be the root vertex of that component (so v_0 is a root). Perform a depth-first search starting from each root vertex in G. Recall that the DFS procedure produces a spanning forest F of G where we consider each tree as rooted as above; also each tree is a normal subtree of its component.

Let T_r be a tree component of F, with root r; let v be any vertex in T_r; and let T_v be the subtree of T rooted at v. Observe that such a subtree T_v can have edges (in G) to at most j ancestors of v (from T_r), since otherwise we could form a minor F_{j+2} by contracting T_v into a single vertex and considering the path from v to the root of the tree. (Note also that by Theorem 5.2.3 this establishes that $\mathcal{A}$ has treewidth at most j, though we will not use this fact.)

Now colour black each vertex of F which is adjacent in G to v_0, and colour each other vertex (including v_0) white. We have described G as a forest of rooted trees on n vertices coloured black or white, with some additional edges (the 'back edges' from the DFS).

Consider the rooted forest $\bar{F} = F - v_0$ (set the nodes that were adjacent to v_0 in F as roots of the newly created trees). Note that $\bar{F}$ can have no black path P_{j+1} as a minor, and if $d(v_0) \geq cn$ then there are at least cn black vertices in $\bar{F}$. Thus by Lemma 7.4.1 the number of all possible rooted forests $\bar{F}$ is $o(\epsilon^n n^n)$ for every $\epsilon > 0$. But since there are at most 2^{n-1} ways to add the node v_0 back to the forest $\bar{F}$ to obtain a valid forest F, the number of different forests F we can get is also $o(\epsilon^n n^n)$ for all $\epsilon > 0$.

From the rooted forests F on $[n]$ we can construct the graphs in $\mathcal{A}_n$ by adding some DFS back edges. Let us show that there are not too many ways to do that. We shall see that, given $G \in \mathcal{A}_n$ and a DFS spanning forest F for G, we can record a small amount of information at the vertices and edges of F such that from F and this information we can reconstruct G. The total amount of information recorded is at most $(2j+1)n$ bits.

Assuming that we are able to do that, then the number of constructions on n vertices which yield a graph with $d(v_0) \geq cn$ is at most $|\mathcal{F}'_n| \cdot 2^n \cdot 2^{(2j+1)n} = o(\epsilon^n n!)$ for any $\epsilon > 0$. But $|\mathcal{A}_n| = \Omega(\gamma^n n!)$ for some $\gamma > 0$. Therefore, for any $\epsilon > 0$

$$\mathbb{P}(d(v_0) \geq cn) \leq \epsilon^n$$

for n sufficiently large. Thus the following lemma will complete the proof of Lemma 7.1.7. □

Lemma 7.4.2 *Let j be a positive integer and suppose that no graph in $\mathcal{A}$ contains as a minor the fan F_{j+2}. Given a graph $G \in \mathcal{A}$ and a DFS spanning forest F for G, we can uniquely describe G using F together with j bits for each vertex of F other than a root, and $j+1$ bits for each arc of F other than those leaving a root.*

Proof of Lemma 7.4.2 Consider a rooted tree T in F. For each vertex x other than the root, let L_x be the list of the ancestors of x (in F) other than the parent of x which have an edge (in G) to T_x, listed in order of increasing distance from the root. Thus L_x has length between 0 and j. Let B_x be the binary j-tuple $b_x(1), \ldots, b_x(j)$, where $b_x(i) = 1$ if the list L_x has length at least i and its ith member is adjacent to x, and otherwise $b_x(i) = 0$.

Also, for each edge xy of T oriented away from the root where x is not the root, let C_{xy} be the binary $(j+1)$-tuple $c_{xy}(1), \ldots, c_{xy}(j+1)$ defined as follows. Let L be the list L_x with the parent of x appended at the end; and for $i = 1, \ldots, j+1$

let $c_{xy}(i) = 1$ if the list L has length at least i and its ith member has a non-tree edge to T_y, and otherwise let $c_{xy}(i) = 0$.

Given G and F we can of course construct all the lists L_x and the tuples B_x and C_{xy}. But conversely, given F and the tuples C_{xy} we can construct the lists L_x and then we can use the tuples B_x to recover G.

We do this as follows. For each tree T in F, we construct the lists L_x by moving one step at a time away from the root. If x is a child of the root then L_x is empty. Now let xy be an edge in T oriented away from the root where x is not the root, and suppose that we know L_x. We form L_y as follows. Let L be L_x with the parent of x appended at the end. For $i = 1, \ldots, j+1$, if $c_{xy}(i) = 1$ we keep the ith member of L, otherwise we delete it, maintaining the same order: this gives L_y.

Thus we can determine each list L_x for x not the root; and now we can use F to determine the parent of x, and L_x and B_x to determine all its other ancestors to which it is adjacent. Thus we can determine G. □

7.5 Proof of Theorem 7.1.1

In the last two sections we proved Lemma 7.1.6 and Lemma 7.1.7. In this section, after a further few preliminary lemmas we use the earlier results to prove Theorem 7.1.1.

7.5.1 Minors, paths and pendant subgraphs

The following lemma is 'nearly obvious' but we spell out a proof.

Lemma 7.5.1 *Let the graph G have H as a minor. Then G has a subgraph $\tilde{H}$ which contracts to a graph isomorphic to H, and a family $(Q(xy) : xy \in E(H))$ of paths in $\tilde{H}$, which partition the edges of $\tilde{H}$ and have no internal vertices in common.*

Proof Since G has H as a minor, there is a family $(T^0_x : x \in V(H))$ of disjoint (that is, pairwise vertex-disjoint) subtrees of G, such that for each edge $xy \in E(H)$ there is at least one edge uv in G between the vertices of T^0_x and T^0_y. Form a set $D \subseteq E(G)$ by picking exactly one such edge uv for each edge $xy \in E(H)$. Call these the 'cross edges'.

Let $x \in V(H)$ and consider the tree T^0_x. Repeatedly remove leaves that are not incident with any cross edge, until no such leaves are left or the tree has just one vertex. The resulting tree T_x has the property that if it has at least two vertices then each leaf of T_x is incident with a cross edge. Do this for each vertex $x \in V(H)$, and let the subgraph $\tilde{H}$ of G be the graph consisting of all the trees T_x together with the edges in D. If we contract each of these trees T_x to a single vertex we obtain a copy of H. Thus it will suffice to show that there is an appropriate family of paths for $\tilde{H}$; and so the following claim will complete the proof.

Claim There is a family $(Q(uv) : uv \in D)$ of pairwise internally vertex-disjoint paths such that (a) for each edge uv in D, $Q(uv)$ contains the edge uv and if uv has one end vertex in T_x and one in T_y then so does $Q(uv)$; and (b) each edge in the trees T_x is contained in one of the paths.

We will prove the claim by induction on $|D|$. It is trivial if $|D| = 0$, so suppose that $|D| \geq 1$ and we know the result for smaller values. Let $uv \in D$ and suppose that u is in T_x and v is in T_y (and so $x \neq y$). We form the path $Q(uv)$ as follows.

If T_x consists just of u, or u is not a leaf of T_x, or u is incident with another edge in D, then let $P(u)$ be the trivial path consisting just of u. Otherwise u is a leaf of T_x and T_x has at least two vertices, and u is not incident with any edge in $D \setminus uv$: in this case we let the path $P(u)$ be the shortest path in T_x between u and a vertex u' such that either u' is incident with an edge in $D \setminus uv$ or u' has degree > 2 in T_x. Similarly we form a corresponding path $P(v)$ in T_y. Let $Q(uv)$ be the path formed by concatenating $P(u)$, uv and $P(v)$.

Now we remove uv from D, and from T_x and T_y we remove the edges and internal vertices of $Q(uv)$. It is easy to see that may use the induction hypothesis to obtain a family of paths for the new configuration, and then add the path $Q(uv)$ to complete the proof of the claim, and thus of the lemma. □

The next lemma follows quickly from the last one.

Lemma 7.5.2 *Let the graph G have H as a minor and let $W \subseteq V(G)$. Then G has a subgraph $\tilde{H}$ which contracts to a graph isomorphic to H, and a family $\mathcal{F}$ of paths in $\tilde{H}$ such that*

(a) the paths in $\mathcal{F}$ partition the edges of $\tilde{H}$,

(b) no path in $\mathcal{F}$ has an internal vertex in W, and

(c) $|\mathcal{F}| \leq |E(H)| + |W|$.

Proof By the last lemma, G has a subgraph $\tilde{H}$ which contracts to a graph isomorphic to H, and a family $(Q(xy) : xy \in E(H))$ of paths in $\tilde{H}$, which partition the edges of $\tilde{H}$ and have no internal vertices in common.

If a vertex $w \in W$ is internal for a path P in this family then P is the only such path for w, and we form two new paths P' and P'' by 'cutting' P at w (so that both of the new paths have w as an endpoint and not as an internal vertex). We do this for each vertex $w \in W$, and thus obtain a family $\mathcal{F}$ of paths as required. □

We call a connected subgraph H of G a *pendant subgraph* if there is exactly one edge of G between $V(H)$ and $V(G) \setminus V(H)$. From the last lemma we deduce:

Lemma 7.5.3 *Let $G = (V, E)$ be a graph, let $W \subset V$, let $G' = G - W$ and let $\mathcal{P}$ be a collection of pairwise vertex-disjoint pendant subgraphs of G'. Suppose the graph H has no isolated vertices, and let G have H as a minor. Then G has a subgraph $\tilde{H}$ which contracts to a graph isomorphic to H, and which has vertices in at most $2(|E(H)| + |W|)$ of the subgraphs in $\mathcal{P}$.*

Proof Invoke Lemma 7.5.2 to obtain a family $\mathcal{F}$ of at most $|E(H)|+|W|$ paths where no path has an internal vertex in W and $\bigcup_{P \in \mathcal{F}} P$ yields a graph which contracts to a graph isomorphic to H.

We claim that any path $P \in \mathcal{F}$ can touch at most 2 distinct pendant subgraphs in $\mathcal{P}$. Clearly the lemma will follow from this claim.

To establish the claim, assume for a contradiction that some path P in $\mathcal{F}$ shares vertices with each of 3 distinct subgraphs C_1, C_2 and C_3 in $\mathcal{P}$. Since P does not have internal vertices in W, we may assume it is entirely contained in G' (otherwise consider P less any vertices in W). Travel along P and without loss of generality suppose that C_1 is visited first and C_2 is visited second. But in order to visit C_2 after C_1 we must cross the bridge e connecting C_2 to the rest of G', and there is no path in $G' - e$ from $V(C_2)$ to $V(C_3)$, so P cannot reach C_3. □

7.5.2 Completing the proof of Theorem 7.1.1

Lemma 7.5.4 *Let $\mathcal{A}$ be an addable minor-closed class of graphs such that its family $\mathcal{B}$ of excluded minors includes at least one planar graph. Assume that $\mathcal{A}$ has the no-linear-degrees property. Then for each positive integer k, we have*

$$|(\mathrm{Ex}\,(k+1)\mathcal{B})_n| = (1 + e^{-\Omega(n)})|(apex^k\,\mathcal{A})_n|.$$

The idea of the proof is similar to that of the proof of Theorem 6.1.1 in Chapter 6. We first describe constructions which yield every graph in $(\mathrm{Ex}\,(k+1)\mathcal{B})_n$ at least once (as well as other graphs); we then show that there are few 'unrealistic' constructions; and finally we show that few 'realistic' constructions yield a graph not in $\mathrm{apex}^k\,\mathcal{A}$.

Proof of Lemma 7.5.4 Fix a positive integer k. By Lemma 7.1.6, there is a positive integer r such that the following holds. For each graph G in $\mathrm{Ex}\,(k+1)\mathcal{B}$ with at least r vertices, there is a $\mathcal{B}$-blocker R of size r and a subset S of R of size k such that $R \setminus v$ is still a $\mathcal{B}$-blocker for each vertex $v \in R \setminus S$.

Let $n > r$. By the above, the following constructions yield every graph in $(\mathrm{Ex}\,(k+1)\mathcal{B})_n$ at least once (as well as other graphs).

(i) Choose an r-subset $R \subseteq V$, put any graph on R, and choose a k-subset $S \subseteq R$ ($\binom{n}{r}2^{\binom{r}{2}}\binom{r}{k} = O(n^r)$ choices)

(ii) Add the edges of any bipartite graph $H(S, V \setminus R)$ with parts S and $V \setminus R$ ($2^{k(n-r)}$ choices)

(iii) Put any graph F in $\mathcal{A}$ on $V \setminus R$ ($|\mathcal{A}_{n-r}|$ choices)

(iv) Add the edges of any bipartite graph $H(R \setminus S, V \setminus R)$ with parts $R \setminus S$ and $V \setminus R$, subject to the restriction that for each $v \in R \setminus S$ the induced subgraph on $V \setminus (R \setminus v)$ is in $\mathcal{A}$.

By the graph minors theorem of Robertson and Seymour [89], $\mathcal{B}$ is a finite set of $j \geq 1$ graphs $H_1, \ldots, H_j$; and by assumption each H_i is 2-connected. Let m be the maximum number of edges in these graphs H_i.

Pick distinct vertices v_1 and r_1 in H_1, and consider the connected graph $H_1 - v_1$ in $\mathcal{A}$. Write $h_1 = |V(H_1)|$. From this graph, form the graph $\tilde{H} \in \mathcal{A}$ by attaching a path of length h_1 to r_1; let vertex r the other end of the path; and let $\tilde{H}$ be rooted at r. Call the part corresponding to $H_1 - v_1$ in a pendant appearance of $\tilde{H}$ a *spike* (following the terminology of Chapter 6) Our construction ensures that spikes must be disjoint.

By Lemma 7.2.1 (the 'pendant appearances' theorem of [77]), there exist constants $a > 0$ and $b > 0$ such that (assuming n is sufficiently large) the number of graphs $F \in \mathcal{A}_{n-r}$ with less than an spikes is at most $e^{-bn}|\mathcal{A}_{n-r}|$. We define further constants in terms of a and b. By a Chernoff bound, there is a constant $c > 0$ such that

$$\mathbb{P}\left(\mathrm{Bin}(\lceil an \rceil, 2^{-h_1+1}) < 2m + 2r\right) = O(e^{-cn}).$$

Let η, $0 < \eta < \frac{1}{2}$ be sufficiently small that

$$\left(\frac{e}{\eta}\right)^{\eta r} < e^{\min\{b,c\}}.$$

Let α be sufficiently large that $e^{\alpha} > 2^r$, and let $t = t(n) = \lceil \eta n \rceil$.

Call a construction *realistic* if there are at most t edges between each vertex $v \in R \setminus S$ and the vertices in $V \setminus R$; and *unrealistic* otherwise. Let $\mathcal{C}(n)$ denote the set of graphs in $\mathcal{A}_{n-r+1}$ such that vertex $n-r+1$ has degree $> t$. Since $\mathcal{A}$ has the no-linear-degrees property,

$$|\mathcal{C}(n)| = O(e^{-\alpha n})\ |\mathcal{A}_n|.$$

For a given choice of R and S, the number of graphs $\tilde{F}$ on $V \setminus S$ such that the induced subgraph $\tilde{F}[R \setminus S]$ is some fixed graph, and some vertex in $R \setminus S$ has $> t$ edges to the vertices in $V \setminus R$ is at most

$$(r-k)\ |\mathcal{C}(n)|\ 2^{(r-k-1)(n-r)}.$$

Hence the number of unrealistic constructions is at most

$$O(n^r) \cdot 2^{k(n-r)}\ 2^{(r-k-1)(n-r)} |\mathcal{C}(n)| \le O(n^r) \cdot 2^{rn} |\mathcal{A}_n|\ e^{-\alpha n} = |\mathcal{A}_n|\ e^{-\Omega(n)}.$$

Thus there are few unrealistic constructions so that we may ignore them further. Note that in realistic constructions, the number of choices for the bipartite graph $H(R \setminus S, V \setminus R)$ in step (iv) is

$$\left(\sum_{i=0}^{t} \binom{n-r}{i}\right)^{r-k} \le \left(n\binom{n}{t}\right)^{r} \le n^r \left(\frac{ne}{t}\right)^{tr} \le n^r \left(\frac{e}{\eta}\right)^{(\eta n+1)r}.$$

Let us bound the number of realistic constructions which yield a graph G in $\mathrm{Ex}\,(k+1)\,\mathcal{B} \backslash \mathrm{apex}^k\,\mathcal{A}$. For each such construction, the graph $G[V \backslash S]$ must contain a minimal subgraph K which contracts to an excluded minor H_i for some $i \in \{1, \ldots, j\}$; and by Lemma 7.5.3, such a subgraph K can touch at most $2(m+r-k)$ spikes.

Now suppose that each vertex v in S is adjacent to all $h_1 - 1$ vertices of each spike in a set A_v of at least $2m + 2r - k$ spikes. Since the graph K does not touch at least $2m + 2r - k - 2(m + r - k) = k$ spikes in A_v for each vertex $v \in S$, we can pick a spike in A_v (for example, greedily) for each $v \in S$ to form k disjoint subgraphs isomorphic to H_1 with none of them touching the graph K.

But now there are at least $k+1$ disjoint excluded minors in G, contradicting G being in $\mathrm{Ex}\,(k+1)\mathcal{B}$. Hence, for at least one vertex v in S, v must be adjacent to all $h_1 - 1$ vertices of at most $2m + 2r - k - 1 < 2m + 2r$ spikes.

Therefore, given any choices at steps (i),(iii) and (iv), if F has z spikes then the number of choices at step (ii) to obtain a graph in $\mathrm{Ex}\,(k+1)\mathcal{B} \setminus \mathrm{apex}^k \mathcal{A}$ is at most

$$2^{k(n-r)}\ k\ \mathbb{P}\left(\mathrm{Bin}(z, 2^{-h_1+1}) < 2m + 2r\right).$$

Hence, by considering separately the realistic constructions which yield a graph in $\mathrm{Ex}\,(k+1)\mathcal{B} \setminus \mathrm{apex}^k \mathcal{A}$ such that F has $< an$ spikes and those ones where F has $\geq an$ spikes, we see that the number of such constructions is at most

$$\begin{aligned} & O(n^{2r})\ 2^{kn} \left(\frac{e}{\eta}\right)^{\eta r n} |\mathcal{A}_{n-r}| \left(e^{-bn} + e^{-cn}\right) \\ = \ & e^{-\Omega(n)}\ 2^{kn}\ |\mathcal{A}_{n-k}| \ = \ e^{-\Omega(n)}\ |(\mathrm{apex}^k \mathcal{A})_n|. \end{aligned}$$

since $|(\mathrm{apex}^k \mathcal{A})_n| \geq 2^{k(n-k)} |\mathcal{A}_{n-k}|$. This completes the proof of the lemma. □

The following two simple facts will be useful.

Lemma 7.5.5 *Let $\mathcal{A}$ be a proper minor-closed class of graphs, with set $\mathcal{B}$ of excluded minors, such that no graph in $\mathcal{B}$ has a component which is a path. Let k be a positive integer. Then there is a positive integer t such that for all $n \geq t$*

$$(\mathrm{Ex}\,(k+1)\mathcal{B} \setminus \mathit{apex}^k \mathcal{A})_n \geq n!/(2 \cdot t!).$$

Proof We need only show that for some $t \geq 3$, the graph K_t is in $\mathrm{Ex}\,(k+1)\mathcal{B} \setminus \mathrm{apex}^k \mathcal{A}$, since then the graphs formed from K_t together with a disjoint path would also be in this class.

Let h be the least number of vertices in a graph in $\mathcal{B}$, so $h \geq 3$. Let $t = h + k$, and note that $t < (k+1)h$, since $(k+1)h - (h+k) = k(h-1) > 0$. Then K_t is not in $\mathrm{apex}^k \mathcal{A}$ because removing k vertices from K_t leaves a copy of K_h; and K_t cannot contain a minor in $(k+1)\mathcal{B}$ since $t < (k+1)h$. □

Lemma 7.5.6 *Let $\mathcal{A}$ be a proper minor-closed class of graphs with set $\mathcal{B}$ of excluded minors, and suppose that $\mathcal{A} \supseteq \mathit{apex}^l \mathcal{C}$ for some class $\mathcal{C}$ of graphs and some positive integer l. Then* $\mathrm{Ex}\,(k+1)\mathcal{B} \supseteq \mathit{apex}^{(k+1)(l+1)-1} \mathcal{C}$.

Proof Let $G \in \mathrm{apex}^{(k+1)(l+1)-1} \mathcal{C}$; and suppose for a contradiction that G has $k+1$ disjoint subgraphs $H_1, \ldots, H_{k+1}$ each with a minor in $\mathcal{B}$. Let $S \subseteq V(G)$ be a set of size at most $(k+1)(l+1)-1$ such that $G-S \in \mathcal{C}$. Since $\text{apex}^l\, \mathcal{C} \subseteq \text{Ex}\, \mathcal{B}$, each graph H_i must have at least $l+1$ vertices in S; and since the graphs H_i are pairwise disjoint we must have that $|S| \geq (k+1)(l+1)$, contradicting our choice of S. □

Now Theorem 7.1.1 will follow easily.

Proof of Theorem 7.1.1 For the first part, suppose that $\mathcal{A}$ is addable and does not contain all fans. Then directly from Lemmas 7.1.7 and 7.5.4, for each positive integer k, as $n \to \infty$

$$|(\text{Ex}\,(k+1)\mathcal{B})_n| = (1 + e^{-\Omega(n)})|(\text{apex}^k\, \mathcal{A})_n|.$$

But now Lemma 7.5.5 allows us to replace the $e^{-\Omega(n)}$ by $e^{-\Theta(n)}$.

Let us prove the second part of the theorem. Suppose that $\mathcal{A}$ contains all fans. Let $\mathcal{P}$ denote the class of all paths, so $\mathcal{A} \supseteq \text{apex}\, \mathcal{P}$. Then by Lemma 7.5.6 we have $\text{Ex}\,(k+1)\mathcal{B} \supseteq \text{apex}^{2k+1}\, \mathcal{P}$. So, by counting just the graphs where the first $2k+1$ vertices form an apex set,

$$\begin{aligned} |(\text{Ex}\,(k+1)\mathcal{B})_n| &\geq |(\text{apex}^{2k+1}\, \mathcal{P})_n| \\ &\geq 2^{(2k+1)(n-2k-1)} \frac{1}{2}(n-2k-1)! \\ &\geq n!\; 2^{(2k+1)n-o(n)}. \end{aligned}$$

But since $\mathcal{A}$ is proper minor-closed, it is small [83] (see [43] for another proof); that is, $|\mathcal{A}_n| = O(\gamma^n n!)$ for some γ. So

$$|(\text{apex}^k\, \mathcal{A})_n| \leq \binom{n}{k} 2^{kn} |\mathcal{A}_{n-k}| = O((2^k\gamma)^n n!).$$

Therefore

$$|(\text{Ex}\,(k+1)\mathcal{B})_n| / |(\text{apex}^k\, \mathcal{A})_n| \geq \left(2^{k+1-o(1)}/\gamma\right)^n$$

and we have completed the proof. □

Remark 7.5.7 *If the class $\mathcal{A}$ contains an apex class larger than apex$\,\mathcal{P}$ we can obtain a better lower bound than in the last part of the proof above. For example, the class* $\text{Ex}\,K_4$ *of series-parallel graphs contains apex$\,\mathcal{F}$, so for each fixed $k \geq 1$, by Lemma 7.5.6,* $|(\text{Ex}\,(k+1)K_4)_n| \geq 2^{(2k+1)n} e^{n-o(n)} n!$. *Also* $\text{Ex}\,K_4$ *has growth constant $\gamma \approx 9.07$ [11], and $2^{k+1}e \geq 4e \approx 10.87$ so $2^{k+1}e > \gamma$; and hence the graphs on $[n]$ in apex^k* $(\text{Ex}\,K_4)$ *form only an exponentially small proportion of those in* $\text{Ex}\,(k+1)K_4$.

7.6 Properties of the random graphs R_n

In this section we use the 'counting' results Theorems 7.1.1 and 7.1.2 to prove Theorems 7.1.3 and 7.1.4, as well as Theorem 7.6.1 which extends Theorem 7.1.5.

Let $\mathcal{M}^k$ be the multiset of graphs produced by the constructions in the proof of Theorem 7.1.2. For graphs $R_n \in_u \text{Ex}\,(k+1)\mathcal{B}$ and $R'_n \in_u \mathcal{M}^k$, that proof together with Theorem 7.1.1 gives that

$$d_{TV}(R_n, R'_n) = e^{-\Omega(n)} \tag{7.4}$$

where d_{TV} denotes total variation distance. Therefore it is enough to prove Theorems 7.1.3 and 7.1.4 with R_n replaced by R'_n (except for the easy lower bound in the second result). Notice that the graph R'_n can be generated by choosing the set S and the graphs $G[S]$, F and B in the steps in the proof of Theorem 7.1.2 uniformly at random.

Proof of Theorem 7.1.3 Suppose $S = S_0$ was chosen for step (1). Note that for each $v \in S_0$

$$\mathbb{P}\left(\left(\frac{1}{2} - \epsilon\right) n \le d(v) \le \left(\frac{1}{2} + \epsilon\right) n\right) = 1 - e^{-\Omega(n)}.$$

Indeed, since the graph $G[S]$ and the bipartite graph B are chosen uniformly at random, each vertex $v \in S_0$ has $d(v) \sim \text{Bin}\left(n-1, \frac{1}{2}\right)$.

To show that with probability $1 - e^{-\Omega(n)}$ no vertex $v \notin S_0$ may belong to S_n (that is, have degree $> \epsilon n$) we can apply Lemma 7.1.7 to the class $\mathcal{A}$ (as each proper addable minor-closed class of graphs has growth constant at least e). Thus for the graph F picked in step (2) uniformly from $\mathcal{A}_{n-k}$ we get

$$\mathbb{P}\left(\Delta(F) > \frac{\epsilon n}{2}\right) = e^{-\Omega(n)}.$$

Hence with probability $1 - e^{-\Omega(n)}$, given that $S = S_0$ was chosen in step (1), each vertex in $V(F)$ has degree not larger than $\frac{\epsilon n}{2} + k < \epsilon n$ in R'_n, for n sufficiently large. Considering now all k-subsets $S_0 \subset [n]$ completes the proof of (i) and (ii).

For (iii), let a, H, H^- and h be such as in the proof of Theorem 7.1.2. We say that a construction G obtained in that proof is *very good* if for each vertex $s \in S$ there are at least $m = 2^{-h}an$ pairwise disjoint subsets $X_1(s), X_2(s), \ldots, X_m(s) \subseteq V(F)$ such that each graph $G[X_i(s) \cup s]$ has H as a minor.

If G is very good and S' is another $\mathcal{B}$-blocker of G such that $S \not\subseteq S'$, then S' must have at least δn vertices where $\delta := 2^{-h}a$ (to see this, let $s \in S \setminus S'$ and note that S' has to contain a vertex from each of the disjoint sets $X_i(s)$). But given that $S = S_0$ and that F has at least $a(n-k)$ copies of H^-, the probability that a random construction R'_n is not very good is at most

$$k\,\mathbb{P}\left(\mathrm{Bin}\left(\lfloor a(n-k)\rfloor, 2^{-h+1}\right) < \lceil 2^{-h}an\rceil\right) = e^{-\Omega(n)}.$$

□

Proof of Theorem 7.1.4 Let $K, H \in \mathcal{A}$ be (fixed) connected graphs such that $\omega(K) = \omega(\mathcal{A})$ and $\chi(H) = \chi(\mathcal{A})$. Fix a k-set S_0 of vertices and a graph H_0 on S_0. By Lemma 7.2.1 (the pendant appearances theorem of [77]), there exists $a > 0$ such that, with probability $1 - e^{-\Omega(n)}$, the graph F chosen in step (2) of the construction has at least $2\lceil an\rceil$ disjoint sets of vertices such that the first $\lceil an\rceil$ induce copies of K, and the second $\lceil an\rceil$ induce copies of H.

Standard bounds for the binomial distribution now show that with probability $1 - e^{-\Omega(n)}$ there is a copy of K and a copy of H such that each possible edge between S_0 and these copies is present. Thus with probability $1 - e^{-\Omega(n)}$ we have $\omega(R'_n) \geq \omega(H_0) + \omega(\mathcal{A})$ and $\chi(R'_n) \geq \chi(H_0) + \chi(\mathcal{A})$. Hence, removing the conditioning on S and the graph on S, with probability $1 - e^{-\Omega(n)}$ we have $\omega(R'_n) \geq \omega(R') + \omega(\mathcal{A})$ and $\chi(R'_n) \geq \chi(R') + \chi(\mathcal{A})$, where R' denotes the induced subgraph $R'_n[S]$. Also $R' \sim R$; that is, R' and R have the same distribution.

Of course, the reverse inequalities, $\omega(R'_n) \leq \omega(R') + \omega(\mathcal{A})$ and $\chi(R'_n) \leq \chi(R') + \chi(\mathcal{A})$ always hold. We have now shown that

$$d_{TV}\left((\omega(R'_n), \chi(R'_n)), (X, Y)\right) = e^{-\Omega(n)}$$

and thus, by the discussion at the start of this section, that

$$d_{TV}\left((\omega(R_n), \chi(R_n)), (X, Y)\right) = e^{-\Omega(n)}.$$

To replace $e^{-\Omega(n)}$ by $e^{-\Theta(n)}$ note that $\mathbb{P}(R_n \in \mathcal{A}) = e^{-O(n)}$. □

We shall deduce Theorem 7.1.5 from a more general result, Theorem 7.6.1. For a graph G we let Big(G) denote its (lexicographically first) largest component, and let the *fragment* Frag(G) be G less Big(G). Let us use $v(G)$ to denote $|V(G)|$. Thus frag(G) = v(Frag(G)). We shall investigate the asymptotic behaviour of Frag($\mathrm{R_n}$), following the treatment in [74].

A class $\mathcal{A}$ of graphs is called *decomposable* if a graph is in $\mathcal{A}$ if and only if each component is. (It is easy to see that each addable minor-closed class is decomposable.) For any graph class $\mathcal{A}$ we let $\mathcal{UA}$ denote the corresponding unlabelled graph class, with members the equivalence classes of graphs in $\mathcal{A}$ under isomorphism.

Now let $\mathcal{A}$ be any decomposable class of (labelled) graphs, and let $A(x) = \sum_{n\geq 0} |\mathcal{A}_n| x^n/n!$ be its exponential generating function. Let $\mathcal{C}$ denote the class of connected graphs in $\mathcal{A}$, with exponential generating function $C(x)$. Recall the 'exponential formula', that $A(x) = e^{C(x)}$ (see for example [48]). (By convention the empty graph $\emptyset$ is in $\mathcal{A}$ and not in $\mathcal{C}$.) If $\rho > 0$ is such that $A(\rho)$ is finite, then we may obtain a natural 'Boltzmann Poisson distribution' on $\mathcal{UA}$, as follows. Let

$$\lambda(H) = \frac{\rho^{v(H)}}{\mathrm{aut}(H)} \quad \text{for each graph } H \in \mathcal{UA} \tag{7.5}$$

where aut(H) denotes the number of automorphisms of H. Then

$$\sum_{H \in \mathcal{UA}} \lambda(H) = A(\rho) = e^{C(\rho)}.$$

The *Boltzmann Poisson random graph* $R = R(\mathcal{A}, \rho)$ takes values in $\mathcal{UA}$, with

$$\mathbb{P}[R = H] = \frac{\lambda(H)}{A(\rho)} \quad \text{for each } H \in \mathcal{UA}. \tag{7.6}$$

It is shown in [74] that the number of components of R isomorphic to a given graph $H \in \mathcal{UC}$ has distribution Po$(\lambda(H))$, and numbers of components corresponding to distinct graphs in $\mathcal{UC}$ are independent; and thus the random number $\kappa(R)$ of components of R satisfies $\kappa(R) \sim \mathrm{Po}(C(\rho))$. Also, $v(R)$ is the sum of independent random variables $v(H)\mathrm{Po}(\lambda(H))$ for $H \in \mathcal{UC}$; and

$$\mathbb{P}[v(R) = n] = \frac{|\mathcal{A}_n| \rho^n/n!}{A(\rho)} \quad \text{for } n = 0, 1, 2, \ldots \tag{7.7}$$

We are interested in the limiting behaviour of the random graph Frag$(\mathrm{R_n})$. It is convenient to deal with the corresponding random unlabelled graph which we denote by $\mathcal{U}$Frag$(\mathrm{R_n})$.

Theorem 7.6.1 *Let $\mathcal{A}$ be an addable minor-closed class of graphs which does not contain all fans; let ρ be the radius of convergence of its exponential generating function $A(x)$; and let $\mathcal{B}$ be its set of excluded minors. Let $\mathcal{C}$ be the class of connected graphs in $\mathcal{A}$, with exponential generating function $C(x)$. Then $A(\rho) <$*

∞; and given a positive integer k, for $R_n \in_u \mathrm{Ex}\,(k+1)\mathcal{B}$ we have

$$d_{TV}(\mathcal{U}\mathrm{Frag(R_n), R}) \to 0 \quad \textit{as} \;\; \mathrm{n} \to \infty. \tag{7.8}$$

where $R = R(\mathcal{A}, \rho/2^k)$ is the Boltzmann Poisson random graph for $\mathcal{A}$ and $\rho/2^k$ as in (7.5) and (7.6) above. Further,

$$\mathbb{E}\,[\mathrm{frag(R_n)}] \to \mathbb{E}\,[\mathrm{v(R)}] = (\rho/2^{\mathrm{k}})\ \mathrm{C}'(\rho/2^{\mathrm{k}}) < \infty. \tag{7.9}$$

Since $\mathbb{P}(R = \emptyset) = e^{-C(\rho/2^k)}$, Theorem 7.1.5 follows as an immediate corollary. Also, $d_{TV}(\mathrm{frag(R_n), v(R)}) \to 0$ as $n \to \infty$, where the distribution of $v(R)$ is given by (7.7) with ρ replaced by $\rho/2^k$.

To prove (7.8) in Theorem 7.6.1 we use one preliminary lemma, followed by a lemma taken from [74]. After that, to prove (7.9) in Theorem 7.6.1 we use another preliminary lemma.

Lemma 7.6.2 *Let $\mathcal{A}$ be a minor-closed class of graphs, with set $\mathcal{B}$ of excluded minors. Let j be a positive integer, suppose that $\mathrm{Ex}\,j\mathcal{B}$ has a positive growth constant, and let $R_n \in_u \mathrm{Ex}\,j\mathcal{B}$. Then*

$$\mathbb{P}(\mathrm{Frag(R_n)} \in \mathcal{A}) = 1 - \mathrm{e}^{-\,(\mathrm{n})} \quad \textit{as}\ \mathrm{n} \to \infty.$$

Proof The case $j = 1$ is trivial, since $\mathcal{A} = \mathrm{Ex}\,\mathcal{B}$; so assume that $j \geq 2$. For $i = 1, 2, \ldots$ let $\mathcal{A}^i$ denote $\mathrm{Ex}\,i\mathcal{B}$, and let the exponential generating function $A^i(x)$ of the graphs in $\mathcal{A}^i$ have radius of convergence ρ_i. Then $\rho_{i+1} \leq \rho_i/2$, since from each graph G in $\mathcal{A}^i_n$ we may construct at least 2^n graphs in $\mathcal{A}^{i+1}_{n+1}$ by adding any set of edges between vertex $n+1$ and $V(G)$. Thus $\rho_j \leq 2^{-(j-1)}\rho_1 \leq \rho_1/2$.

Observe that if any component of R_n is in $\mathcal{A}^j \setminus \mathcal{A}^{j-1}$ then the remaining components must be in $\mathcal{A}$. Thus

$$\begin{aligned} \mathbb{P}(\mathrm{Frag(R_n)} \notin \mathcal{A}) \;&\leq\; \mathbb{P}(R_n \text{ has all components in } \mathcal{A}^{j-1}) \\ &+\; \mathbb{P}(\mathrm{Frag(R_n)\ has\ a\ component\ in\ } \mathcal{A}^{\mathrm{j}} \setminus \mathcal{A}^{\mathrm{j}-1}). \end{aligned} \tag{7.10}$$

The first term on the right side tends to 0, because the family $\mathcal{D}$ of graphs with all components from $\mathcal{A}^{j-1}$ has radius of convergence $\rho_{j-1} \geq 2\rho_j$. To see this, let $C(x)$ denote the exponential generating function of the connected graphs in $\mathcal{A}^{j-1}$: then, by the exponential formula, $\mathcal{D}$ has exponential generating function $e^{C(x)}$, and this converges for $0 < x < \rho_{j-1}$.

Consider the second term on the right side. Fix $\epsilon > 0$ sufficiently small that

$$\frac{(\rho_1^{-1}+\epsilon)(\rho_j^{-1}+\epsilon)}{(\rho_j^{-1}-\epsilon)^2} < 1.$$

There are constants $0 < a \le b$ such that for all non-negative integers n

$$a(\rho_j^{-1}-\epsilon)^n \le \frac{|\mathcal{A}_n^j|}{n!} \le b(\rho_j^{-1}+\epsilon)^n$$

(since $\mathcal{A}^j$ has growth constant ρ_j^{-1}) and

$$\frac{|\mathcal{A}_n|}{n!} \le b(\rho_1^{-1}+\epsilon)^n.$$

Since each component of $\mathrm{Frag}(\mathrm{R_n})$ has at most $n/2$ vertices, it now follows that the second term on the right side of (7.10) is at most

$$\begin{aligned} \frac{1}{|\mathcal{A}_n^j|}\sum_{t=0}^{\lfloor\frac{n}{2}\rfloor}\binom{n}{t}|\mathcal{A}_t^j||\mathcal{A}_{n-t}| &\le \frac{b^2 n!}{an!(\rho_j^{-1}-\epsilon)^n}\sum_{t=0}^{\lfloor\frac{n}{2}\rfloor}(\rho_j^{-1}+\epsilon)^t(\rho_1^{-1}+\epsilon)^{n-t} \\ &\le (b^2/a)n\left(\frac{(\rho_1^{-1}+\epsilon)(\rho_j^{-1}+\epsilon)}{(\rho_j^{-1}-\epsilon)^2}\right)^{n/2} = e^{-\Omega(n)}. \end{aligned}$$

□

We say that $\mathcal{A}$ is *bridge-addable* if given any graph G in $\mathcal{A}$ and vertices u and v in distinct components of G, the graph obtained from G by adding an edge joining u and v must be in $\mathcal{A}$. It is easy to see that each addable minor-closed class is bridge-addable. Given a graph H in $\mathcal{A}$, we say that H is *freely addable* to $\mathcal{A}$ if, given any graph G disjoint from H, the union of G and H is in $\mathcal{A}$ if and only if G is in $\mathcal{A}$. Recall that $\mathcal{A}$ is called smooth if $|\mathcal{A}_n|/\left(n|\mathcal{A}_{n-1}|\right)$ converges to some finite constant $\gamma > 0$ as $n \to \infty$. The following lemma is a combination of Lemmas 4.4 and 4.5 from McDiarmid [74].

Lemma 7.6.3 *Let the class $\mathcal{A}$ of graphs be minor-closed and bridge-addable; let $R_n \in_u \mathcal{A}$; let $\mathcal{B}$ denote the class of all graphs freely addable to $\mathcal{A}$; and suppose that $\mathbb{P}(\mathrm{Frag}(\mathrm{R_n}) \in \mathcal{B}) \to 1$ as $n \to \infty$. Suppose further that $\mathcal{A}$ is smooth, and its exponential generating function $A(x)$ has radius of convergence ρ, with $0 < \rho < \infty$.*

Then the exponential generating function $B(x)$ of $\mathcal{B}$ satisfies $0 < B(\rho) < \infty$; and $d_{TV}(\mathcal{U}\mathrm{Frag}(\mathrm{R_n}), \mathrm{R}) \to 0$ where $R = R(\mathcal{B}, \rho)$ is the Boltzmann Poisson random graph for $\mathcal{B}$ and ρ as defined in (7.5) and (7.6) above. Further, $\mathbb{E}\left[v(R)\right] = \rho C'(\rho) <$

∞, where $C(x)$ is the exponential generating function of the class $\mathcal{C}$ of connected graphs in $\mathcal{B}$.

Proof of (7.8) in Theorem 7.6.1 Let $\mathcal{A}^{k+1}$ denote $\text{Ex}\,(k+1)\mathcal{B}$. Since $\mathcal{A}$ is addable, it follows that $\mathcal{A}^{k+1}$ is bridge-addable and the class of graphs freely addable to $\mathcal{A}^{k+1}$ is $\mathcal{A}$. By Theorems 7.1.1 and 7.1.2, $\mathcal{A}^{k+1}$ is smooth and its exponential generating function has radius of convergence $\rho/2^k$. Thus by Lemma 7.6.2 we may use Lemma 7.6.3 to complete the proof. □

In order to prove (7.9) in Theorem 7.6.1 we need one more lemma. For a vertex v in a graph G, we let $\text{Comp}(v, G)$ denote the component containing v and let $\text{comp}(v, G)$ denote its number of vertices.

Lemma 7.6.4 *Let $\mathcal{A}$ be an addable minor-closed class of graphs which does not contain all fans; let $\mathcal{B}$ be its set of excluded minors; let k be a positive integer and let $R_n \in_u \text{Ex}\,(k+1)\mathcal{B}$. Then for each $\epsilon > 0$ there is a $K \geq 0$ such that*

$$\mathbb{E}\left[|\{v \in V(\mathrm{Frag(R_n)}) : \mathrm{comp(v, R_n) \geq K}\}|\right] < \epsilon.$$

Proof By (7.4), it suffices to prove this result with R_n replaced by R'_n. Recall that R'_n specifies a k-set S and $F_{n-k} \in \mathcal{A}$. Let A_n be the event that in R'_n, each vertex in S has an edge to $\mathrm{Big(F_{n-k})}$. We first show that

$$\mathbb{P}(A_n) = 1 - e^{-\Omega(n^{\frac{1}{2}})}. \tag{7.11}$$

Since $\mathcal{A}$ is bridge-addable, by Theorem 2.2 of [77], if we let $k = \lfloor n^{\frac{1}{2}} \rfloor$ then

$$\mathbb{P}(\kappa(F_{n-k}) \geq k+1) \leq 1/k! = e^{-\Omega(n^{\frac{1}{2}}\log n)}.$$

(Recall that $\kappa(G)$ denotes the number of components of G.) Thus

$$\mathbb{P}(|\mathrm{Big(F_{n-k})}| < \mathrm{n^{\frac{1}{2}}}) = \mathrm{e^{-\,(n^{\frac{1}{2}}\log n)}}.$$

Hence, if v_i is a vertex in S then

$$\mathbb{P}(v_i \text{ has no edge to } \mathrm{Big(F_{n-k})}) \leq 2^{-\mathrm{n^{\frac{1}{2}}}} + \mathbb{P}(|\mathrm{Big(F_{n-k})}| < \mathrm{n^{\frac{1}{2}}}) = \mathrm{e^{-\,(n^{\frac{1}{2}})}};$$

and now (7.11) follows.

Next observe that if A_n holds then S and $\mathrm{Big(F_{n-k})}$ are contained in $\mathrm{Big(R'_n)}$; and thus if $v \in V(\mathrm{Frag(R'_n)})$ then $v \in V(\mathrm{Frag(F_{n-k})})$, and there is no edge between S and Comp(v, F_{n-k}).

By (7.11) it suffices to upper bound $\mathbb{E}\left[|\{v \in V(\mathrm{Frag}(\mathrm{R}'_n)) : \mathrm{comp}(\mathrm{v}, \mathrm{R}'_n) \geq \mathrm{K}\}| \cdot 1_{\mathrm{A_n}}\right]$. For each $v \in V \setminus S$ let $B_n(v)$ be the event that there is no edge between S and Comp(v, F_{n-k}). By the above observations

$$\begin{aligned}
& \mathbb{E}\left[|\{v \in V(\mathrm{Frag}(\mathrm{R}'_n)) : \mathrm{comp}(\mathrm{v}, \mathrm{R}'_n) \geq \mathrm{K}\}| \cdot 1_{\mathrm{A_n}}\right] \\
\leq\ & \mathbb{E}\left[|\{v \in V \setminus S : v \in V(\mathrm{Frag}(\mathrm{F_{n-k}})), \mathrm{comp}(\mathrm{v}, \mathrm{F_{n-k}}) \geq \mathrm{K}, \mathrm{B_n(v)}\}|\right] \\
\leq\ & \sum_{v \in V \setminus S} \mathbb{P}\left(B_n(v) \mid \mathrm{comp}(v, F_{n-k}) \geq K, v \in V(\mathrm{Frag}(\mathrm{F_{n-k}}))\right) \cdot \mathbb{P}(v \in V(\mathrm{Frag}(\mathrm{F_{n-k}}))) \\
\leq\ & \sum_{v \in V \setminus S} 2^{-kK} \mathbb{P}(v \in V(\mathrm{Frag}(\mathrm{F_{n-k}}))) \\
\leq\ & 2^{-K} \mathbb{E}\left[\mathrm{frag}(\mathrm{F_{n-k}})\right] \ \leq\ \mathrm{c} \cdot 2^{-\mathrm{K}}.
\end{aligned}$$

In the last inequality here we used the result that $\mathbb{E}\left[\mathrm{frag}(\mathrm{F_{n-k}})\right] \leq \mathrm{c}$ for a constant c, see Lemma 2.6 of [74]. The lemma follows. □

Proof of (7.9) in Theorem 7.6.1 Let $\epsilon > 0$. By Lemma 7.6.3

$$\mathbb{E}\left[v(R)\right] = \sum_{H \in \mathcal{UC}} v(H)\lambda(H) = (\rho/2^k)\ C'(\rho/2^k) < \infty.$$

Thus for K sufficiently large

$$\mathbb{E}\left[v(R)\right] - \epsilon \leq \sum_{H \in \mathcal{UC}, v(H) \leq K} v(H)\lambda(H) \leq \mathbb{E}\left[v(R)\right].$$

By (7.8)

$$\mathbb{E}\left[|\{v \in V(\mathrm{Frag}(\mathrm{R_n})) : \mathrm{comp}(\mathrm{v}, \mathrm{R_n}) \leq \mathrm{K}\}|\right] \to \sum_{\mathrm{H} \in \mathcal{UC}, \mathrm{v(H)} \leq \mathrm{K}} \mathrm{v(H)}\lambda(\mathrm{H}),$$

and so

$$|\ \mathbb{E}\left[|\{v \in V(\mathrm{Frag}(\mathrm{R_n})) : \mathrm{comp}(\mathrm{v}, \mathrm{R_n}) \leq \mathrm{K}\}|\right] - \mathbb{E}\left[\mathrm{v(R)}\right]\ | < \epsilon + \mathrm{o}(1).$$

Hence by Lemma 7.6.4 with K sufficiently large

$$|\ \mathbb{E}\left[\mathrm{frag}(\mathrm{R_n})\right] - \mathbb{E}\left[\mathrm{v(R)}\right]\ | < 2\epsilon + \mathrm{o}(1)$$

and we are done. □

7.7 Concluding remarks

Consider an addable minor-closed class $\mathcal{A}$ of graphs, with set $\mathcal{B}$ of excluded minors. For such a class $\mathcal{A}$, we have learned much about the class of graphs with at most k disjoint excluded minors, and in particular about the relationship between $\mathrm{Ex}\,(k+1)\mathcal{B}$ and $\mathrm{apex}^k\mathcal{A}$. Here there are just two cases: if $\mathcal{A}$ does not contain all fans then the difference class $\mathrm{Ex}\,(k+1)\mathcal{B} \setminus (\mathrm{apex}^k\mathcal{A})$ forms an exponentially small proportion of $\mathrm{Ex}\,(k+1)\mathcal{B}$; and if $\mathcal{A}$ contains all fans then $\mathrm{Ex}\,(k+1)\mathcal{B}$ is exponentially larger than $\mathrm{apex}^k\mathcal{A}$ (at least for large k).

But what happens if the minors of $\mathcal{A}$ are not 2-connected? Consider for example the class $\mathrm{Ex}\, S_t$, where S_t denotes the star with t leaves (and thus with $t+1$ vertices). For each $t \geq 3$ almost all graphs in $\mathrm{Ex}\,(k+1)S_t$ are in $\mathrm{apex}^k\ \mathrm{Ex}\, S_t$, as in the case when $\mathcal{A}$ is addable and does not contain all fans; and for $t = 3$ the difference class $\mathrm{Ex}\,(k+1)S_t \setminus (\mathrm{apex}^k\ \mathrm{Ex}\, S_t)$ forms an exponentially small proportion of $\mathrm{Ex}\,(k+1)S_t$; but this is not the case for $t \geq 4$, where the proportion is $2^{-\Theta(n^{\frac{2t-5}{2t-4}})}$, see [75]. There is more to be learned about disjoint excluded minors in such classes of graphs.

A second natural question concerns the behaviour of $\mathrm{Ex}\,(k+1)\mathcal{B}$ when $\mathcal{A}$ is minor-closed and contains all fans. We have learned little about this case, other than the fact that $\mathrm{apex}^k\mathcal{A}$ is irrelevantly small in comparison (at least for large k).

A good starting point is to consider the class $\mathcal{A} = \mathrm{Ex}\, K_4$ of series-parallel graphs, see Remark 7.5.7. Clearly $\mathrm{Ex}\, 2K_4$ contains $\mathrm{apex}^3\mathcal{F}$, where $\mathcal{F}$ denotes the class of forests, and so if $\mathrm{Ex}\, 2K_4$ has a growth constant then it must be at least $8e$. It is not hard to see that this is not the right value, but it may give the right idea. We conjecture that almost all graphs G in $\mathrm{Ex}\, 2K_4$ contain a set S of three vertices such that any two form a K_4-minor-blocker, or equivalently every non-series-parallel subgraph of G has at least 2 vertices in S.

More generally, consider any (fixed) planar graph H. Perhaps there is a positive integer j_H such that the following is true for every positive integer k: almost all graphs G with at most k disjoint subgraphs contractible to H contain a set S of $(k+1)j_H - 1$ vertices such that each subgraph of G contractible to H contains at least j_H vertices from S. Observe that this is true for $H = jC_3$, with $j_H = j$. For if we let $\mathcal{B} = \{jC_3\}$ and $\mathcal{A} = \mathrm{Ex}\,\mathcal{B}$, then $\mathrm{Ex}\,(k+1)\mathcal{B} = \mathrm{Ex}\,(k+1)j\, C_3$ which is very close to $\mathrm{apex}^{(k+1)j-1}\,\mathcal{F}$ (where $\mathcal{F}$ is the class of forests); and for a graph G in this class, the set S consisting of the $(k+1)j-1$ apex vertices is such that each subgraph of G not in $\mathcal{A}$ contains at least j vertices from S.

Acknowledgement We would like to thank both referees of [65] for helpful comments which led to improvements in the presentation of the paper and this chapter.

Chapter 8

Few disjoint minors in $\mathcal{B}$ when $\mathrm{Ex}\,\mathcal{B}$ contains all fans

8.1 Introduction

In Chapter 7 we studied classes $\mathrm{Ex}\,(k+1)\mathcal{B}$ when the graphs in $\mathcal{B}$ are 2-connected and $\mathrm{Ex}\,\mathcal{B}$ does not contain all fans. What happens in the case when $\mathrm{Ex}\,\mathcal{B}$ contains all fans? We prove some general results in this chapter, and we consider specific forbidden minors, such as $2K_4$ in the next one.

We call a $\mathcal{B}$-blocker Q of a graph G *redundant* ("0-redundant", in the terminology of Chapter 7) if for each vertex $v \in Q$ the set $Q \setminus \{v\}$ is still a $\mathcal{B}$-blocker for G. We denote the class of graphs that have a redundant $\mathcal{B}$-blocker for size k by $\mathrm{rd}_k\,\mathcal{B}$. For a graph H, we will often abbreviate $\mathrm{Ex}\,\{H\}$ to $\mathrm{Ex}\,H$, $\mathrm{rd}_k\,\{H\}$ to $\mathrm{rd}_k\,H$, etc.

Given a positive integer s call a graph G an *s-fan* if G is a union of a complete bipartite graph with parts A and B, where $|A| = s$, and a path P with $V(P) = B$. We call 1-fans simply *fans*. Given a positive integer k and a set of graphs $\mathcal{B}$ we denote by $k\mathcal{B}$ the class of graphs consisting of k vertex disjoint copies of graphs in $\mathcal{B}$ (with repetitions allowed). Thus $\mathrm{Ex}\,(k+1)\mathcal{B}$ is the class of graphs that do not have $k+1$ vertex disjoint subgraphs $H_1, \ldots, H_{k+1}$, each with a minor in $\mathcal{B}$.

Recall that a classical result of Robertson and Seymour says that each minor-closed class can be characterised by a finite set $\mathcal{B}$ of *minimal excluded minors*, i.e. $\mathcal{A} = \mathrm{Ex}\,\mathcal{B}$ (see, e.g., [38]).

Let k be a positive integer and let $\mathcal{A}$ be a proper addable minor-closed class of graphs, with a set $\mathcal{B}$ of excluded minors. Recall that by Theorem 7.1.1 from Section 7.1

$$|(\mathrm{Ex}\,(k+1)\mathcal{B})_n| = (1 + e^{-\Theta(n)})|(\mathrm{apex}^k\,\mathcal{A})_n| \tag{8.1}$$

if $\mathcal{A}$ does not contain all fans. Suppose $\mathcal{A} = \operatorname{Ex}\mathcal{B}$ is addable but contains all fans. Then the class $\operatorname{apex}^k \mathcal{A} \subseteq \operatorname{Ex}(k+1)\mathcal{B}$ still seems a natural candidate to be the dominating subclass of $\operatorname{Ex}(k+1)\mathcal{B}$. However, we showed, see Remark 7.5.7, that for such $\mathcal{B}$ Theorem 7.1.1 fails, at least for large k. The first (and the main) theorem of this chapter shows that a very different subclass determines the convergence radius of $\operatorname{Ex}(k+1)\mathcal{B}$, namely, the class $\operatorname{rd}_{2k+1}\mathcal{B}$. Clearly, $\operatorname{rd}_{2k+1}\mathcal{B} \subseteq \operatorname{Ex}(k+1)\mathcal{B}$: if Q is a redundant blocker for G and $|Q| = 2k+1$ then each subgraph of G with a minor in $\mathcal{B}$ uses at least two vertices of Q, so we can find no more than k disjoint such subgraphs.

Theorem 8.1.1 *Let $\mathcal{A}$ be a proper addable minor-closed class of graphs, with a set $\mathcal{B}$ of minimal excluded minors and growth constant γ. Suppose $\mathcal{A}$ contains all fans, but not all 2-fans, nor all complete bipartite graphs $K_{3,t}$.*

Then there is a positive integer $k_0 = k_0(\mathcal{B})$ such that the following holds. Let k be a positive integer. If $k \geq k_0$,

$$\rho(\operatorname{Ex}(k+1)\mathcal{B}) = \rho(rd_{2k+1}\mathcal{B}) < \rho\left((\operatorname{Ex}(k+1)\mathcal{B}) \cap apex^{2k-1}\mathcal{A}\right).$$

If $k < k_0$, the class $\operatorname{Ex}(k+1)\mathcal{B}$ has a growth constant $2^k\gamma$. Furthermore, if $\rho(rd_{2k+1}\mathcal{B})^{-1} < 2^k\gamma$ then (8.1) holds.

For $t \geq 4$ we denote by W_t a wheel graph on t vertices. Some examples of classes $\mathcal{A}$ for which Theorem 8.1.1 applies are $\operatorname{Ex}K_4$, $\operatorname{Ex}K_{2,t}$ for $t \geq 3$, $\operatorname{Ex}W_5$, and $\operatorname{Ex}\{K_{3,t}, F_s\}$, where $t \geq 2$, $s \geq 5$ and F_s is a 2-fan on s vertices. The condition of Theorem 8.1.1 is not satisfied for, say, $\mathcal{A} = \operatorname{Ex}W_6$.

We believe that the condition of Theorem 8.1.1 can be weakened. For example, it may be possible to drop the requirement that $\mathcal{A}$ does not contain all graphs $K_{3,t}$, to get an assumption similar to the one of Theorem 7.1.1. We also believe that for $k \geq k_0$ the fraction of graphs in $(\operatorname{Ex}(k+1)\mathcal{B})_n$ that are not in $(\operatorname{rd}_{2k+1}\mathcal{B})_n$ is exponentially small.

Let $\mathcal{A}$ be an addable class of graphs. In [77], two properties are fundamental in the proof that $\mathcal{A}$ has a growth constant. First, the class $\mathcal{A}$ is *decomposable*, meaning that G is in $\mathcal{A}$ if and only if each component of G is. Second, $\mathcal{A}$ is *bridge-addable*, i.e., it is closed under adding bridges between distinct components. Classes $\operatorname{Ex}(k+1)\mathcal{B}$ are bridge-addable, but not decomposable. Theorem 8.1.1 reduces the problem of proving that a growth constant of $\operatorname{Ex}(k+1)\mathcal{B}$ exists, to the analogous problem for the class $\operatorname{rd}_{2k+1}\mathcal{B}$. Graphs in $\operatorname{rd}_{2k+1}\mathcal{B}$ with a fixed redundant blocker can be represented by a decomposable class of coloured graphs, see Section 8.4.

With stronger conditions on $\mathcal{A}$, this allows us to prove the following.

Theorem 8.1.2 *Let k be a positive integer and let $\mathcal{A}$ be a proper addable minor-closed class of graphs with a set $\mathcal{B}$ of minimal excluded minors. Suppose each graph in $\mathcal{B}$ is 3-connected, $\mathcal{A}$ does not contain all 2-fans, nor all complete bipartite graphs $K_{3,t}$, nor all wheels. Then* $\mathrm{Ex}\,(k+1)\mathcal{B}$ *has a growth constant.*

Classes $\mathcal{A}$ that satisfy the condition of Theorem 8.1.2 are, for instance, $\mathrm{Ex}\,K_4$ and $\mathrm{Ex}\,W_5$, but not $\mathrm{Ex}\,K_{2,3}$.

In Section 8.3 we prove our key structural lemmas and Theorem 8.1.1. Section 8.4 is similar, using a superadditivity argument as in [77], we prove Theorem 8.1.2 there.

8.2 Definitions

8.2.1 Definitions for coloured graphs

Let $t \geq 0$ be a fixed integer. We will consider $\{0,1\}^t$-coloured graphs G where each vertex $v \in V(G)$ is assigned a colour

$$\mathrm{col}(v) = \mathrm{col}_G(v) = (\mathrm{col}_1(v), \ldots, \mathrm{col}_t(v)) \in \{0,1\}^t.$$

We say that $v \in V(G)$ has colour i if $\mathrm{col}_i(v) = 1$. We denote by $\mathrm{Col}(v) = \mathrm{Col}_G(v) = \{k : c_k(v) \neq 0\}$ the set of all colours of v, similarly let $\mathrm{Col}(G)$ be the union of $\mathrm{Col}_G(v)$ for all $v \in V(G)$. Also, denote by $N(G)$ the uncoloured graph obtained by removing all colours from G. Whenever t is clear from the context or not important, we will call $\{0,1\}^t$-coloured graphs just *coloured graphs* or simply *graphs*. We call a vertex $v \in V(G)$ *coloured* if $\mathrm{Col}_G(v) \neq \emptyset$.

Let G be a $\{0,1\}^t$-coloured graph. Given a set $L = \{s_1, \ldots, s_t\}$ such that $s_1, \ldots, s_t \notin V(G)$ and $s_1 < \cdots < s_t$, we can obtain an (uncoloured) graph G^L on vertex set $V(G) \cup L$ by connecting s_i to each vertex $v \in V(G)$ that has colour i. We call G^L an *extension of* G. We denote by $\mathrm{Ext}(G)$ the set of all extensions G^L of G, and denote by $\mathrm{ext}(G)$ an arbitrary representative of $\mathrm{Ext}(G)$.

For a $\{0,1\}^t$-coloured graph G we define the contraction operation in the standard way (see, e.g., [38]) with the addition that the vertex w obtained from contracting an edge $uv \in E(G)$ has colours $\mathrm{Col}(w) = \mathrm{Col}(u) \cup \mathrm{Col}(v)$. A $\{0,1\}^t$-coloured graph H is a *subgraph* of G if H is a subgraph of G, if the colours are ignored, and for each $v \in V(H)$ we have $\mathrm{Col}_H(v) \subseteq \mathrm{Col}_G(v)$. H is a *coloured minor* of G if it can be obtained by contraction and subgraph operations from G.

When a (coloured) graph G has $V(G) \subseteq [n]$ for some positive integer n, we will usually assume that the new vertex w resulting from the contraction of an edge $e = xy$ has label $\min(x, y)$, so that $V(G/e) \subseteq [n]$. For a (coloured) graph G and $J \subseteq E(G)$ we will denote by G/J the graph resulting from the contraction of all of the edges in J. The operation G/J corresponds to a partition of $V(G)$ into a set of "bags" $\{Bag(v) : v \in V(G/J)\}$, where $Bag(v)$ is the set of vertices that contract to v. We call a subgraph H of G *stable* with respect to contraction of J in G if no pair of vertices of H is contracted into the same bag.

We say that two $\{0, 1\}^t$-coloured graphs G' and G'' are *isomorphic* if there is a bijection $f : V(G') \to V(G'')$ such that $xy \in E(G')$ if and only if $f(x)f(y) \in E(G'')$ and $\mathrm{Col}_{G'}(x) = \mathrm{Col}_{G''}(f(x))$ for each $x \in V(G')$.

For a $\{0, 1\}^t$-coloured graph G, we say that $S \subseteq V(G)$ has colour c if $c \in \mathrm{Col}_G(v)$ for some $v \in S$. We say that G has colour c if $V(G)$ does. For a vertex $v \in V(G)$ we let $\Gamma(v) = \Gamma_G(v)$ denote the set of neighbours of v in G. It will be convenient to call the colours 1, 2 and 3 *red, green and blue* respectively.

Let C be the set of cut points of G and let $\mathcal{B}$ be the set of its blocks. Fix $r \in V(G)$. Then the tree T_r with vertex set $C \cup \{r\} \cup \mathcal{B}$ and edges given by uB where $u \in C \cup \{r\}$, $B \in \mathcal{B}$ and $u \in V(B)$ will be called a *rooted block tree* of G, rooted at r. (This is a minor modification of the usual block tree, see [38].) We call graphs that are either 2-connected or isomorphic to K_2 *biconnected*.

For a graph G and a set S, we write $G \cap S = G[V(G) \cap S]$. For two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ we write $G_1 \cup G_2 = (V_1 \cup V_2, E_1 \cup E_2)$ and $G_1 \cap G_2 = (V_1 \cap V_2, E_1 \cap E_2)$.

8.2.2 Analytic combinatorics

We will apply the "symbolic method" of Flajolet and Sedgewick [48] to study the asymptotic number of graphs from various classes.

In this Chapter we follow the notational conventions used in [48]. The size of a graph G is the number of labelled vertices, while $V(G)$ refers to the set of all vertices of G, including the unlabelled (pointed) ones. The exponential generating function of $\mathcal{A}, \mathcal{B}, \dots$ is denoted $A(x), B(x), \dots$ respectively. For instance, $A(x) = \sum_{n=0}^{\infty} \frac{|\mathcal{A}_n|}{n!} x^n$. By $\mathcal{Z}$ we denote the class of graphs consisting of a single vertex with a label, such that $Z(x) = x$.

We use the notation $\mathcal{A}+\mathcal{B}$, $\mathcal{A}\times\mathcal{B}$, $\mathcal{A}(\mathcal{B})$ to denote the class of graphs obtained by the (disjoint) union, labelled product and composition operations respectively, see [48]. For a positive integer k, $\mathcal{A}^k$ denotes the class consisting of a sequence of k disjoint members from $\mathcal{A}$, and we define $\mathcal{A}^0$ to be the class with exponential generating function $A^0(x) = 1$. We also refer to [48] for the formal definition of the class SET($\mathcal{A}$) (obtained by taking arbitrary sets of elements of $\mathcal{A}$ and appropriately relabelling), the class SEQ($\mathcal{A}$) (obtained by taking any ordered sequence of elements of $\mathcal{A}$ and appropriately relabelling), and classes $\mathrm{SET}_{\geq k}(\mathcal{A})$ (sets of at least k elements) and $\mathrm{SEQ}_{\geq k}$ (sequences of at least k elements). Given a positive integer k, we will denote by $k \times \mathcal{A}$ a combinatorial class with the counting sequence $(k|\mathcal{A}_n|, n = 0, 1, \dots)$.

To denote dependence of a class $\mathcal{A}$ (or a generating function A) on a parameter l we will use either superscript $\mathcal{A}^{<l>}$, $\mathcal{A}^l$ or a subscript $\mathcal{A}_l$.[1]

If $\mathcal{A}$ and $\mathcal{B}$ have identical counting sequences $(|\mathcal{A}_n|, n = 0, 1, \dots), (|\mathcal{B}_n|, n = 0, 1, \dots)$ we call $\mathcal{A}$ and $\mathcal{B}$ *combinatorially isomorphic* and write $\mathcal{A} = \mathcal{B}$. We note that most of the decomposition results of Section 9.2 and onwards yield a stronger kind of isomorphism than just the combinatorial one: we prove unique decompositions of graphs from one class into unions of subgraphs with disjoint sets of labels from other classes. This is important since many of our proofs rely on the structure of graphs.

8.3 Structural results for $\mathrm{Ex}\,(k+1)\mathcal{B}$

8.3.1 The colour reduction lemma

The following simple lemma will be very useful in our structural proofs.

Lemma 8.3.1 *Let l be a non-negative integer. Let G be a $\{0,1\}^2$-coloured graph. Suppose G does not have $l+1$ disjoint connected subgraphs containing both colours. Then there is a set S of at most l vertices such that each component of $G - S$ has at most one colour.*

Proof For two new vertices $s, t \notin V(G)$ consider the extension $G' = G^{\{s,t\}}$. G' has $l + 1$ internally disjoint paths from s to t if and only if G has $l + 1$ disjoint connected subgraphs containing both colours. By Menger's theorem we may find a set S of at most l vertices in $V(G)$ such that S separates s from t in G', and hence each component of $G - S$ can have at most one colour. □

Given an integer s and a graph G, we define its *apex width of order s*, denoted $\mathrm{aw}_s\, G$ as the maximum number j such that G has a minor H on $j + s$ vertices

[1] When this coincides with the notation for the elements $\mathcal{A}_n$ in $\mathcal{A}$ with labels in $[n]$ or with a power of a class $\mathcal{A}^k$, the meaning should be determined from the context.

where H is a union of a tree T with $|V(T)| = j$ and a complete bipartite graph with parts $V(T)$ and $V(H) \setminus V(T)$. For a class of graphs $\mathcal{A}$ we define $\mathrm{aw}_s(\mathcal{A})$ to be the supremum of $\mathrm{aw}_s(G)$ over $G \in \mathcal{A}$. In this thesis we will only use the parameter aw_2. For example, it is easy to check that its value for classes $\mathrm{Ex}\,K_4$, $\mathrm{Ex}\,K_{2,t}$ and $\mathrm{Ex}\,K_5$ is 2, $t-1$ and ∞ respectively. In Section 8.3.4 below we give a condition to check if $\mathrm{aw}_j(\mathcal{A})$ is finite.

Given integers s and t, $1 \le s \le t$ and a $\{0,1\}^t$-coloured graph G, define its *coloured apex width of order s*, denoted $\mathrm{caw}_s(G)$, as the maximum number j, for which there are j pairwise disjoint connected subgraphs $H_1, \dots, H_j$ of G that have at least s common colours, i.e., $|\mathrm{Col}(H_1) \cap \cdots \cap \mathrm{Col}(H_j)| \ge s$. For a class of coloured graphs $\mathcal{A}$ we define $\mathrm{caw}_s(\mathcal{A})$ as the supremum of $\mathrm{caw}_s(G)$ over the graphs $G \in \mathcal{A}$.

We state one of our key structural lemmas next. We assume that all graphs here have vertices in $\mathbb{N}$.

Lemma 8.3.2 (Colour reduction lemma) *Let $t \ge 2$ be an integer and let G be a connected $\{0,1\}^t$-coloured graph. Suppose $\mathrm{caw}_2(G) \le j$, for some non-negative integer j.*

Then there is a connected $\{0,1\}^t$-coloured graph G' and a set $J \subseteq E(G')$ of size at most $(j+1)^{t-1} - 1$ such that a) each component of $G' - J$ has at most one colour, b) $G'/J = G$ and c) each component of $G' - J$ is stable with respect to contraction of J in G'.

Proof We use induction on t. For $t = 2$ by Lemma 8.3.1 there is a set B of at most j vertices in G such that each component of $G - B$ has at most one colour. Denote by V_{green} the set of vertices that belong to a component of $G - B$ that has the green colour.

Let $G_0 = N(G[B])$. For each vertex $v \in B$ take a new vertex $v' \notin V(G)$; let $B' = \{v' : v \in B\}$. Now define a matching $J = \{vv' : v \in B\}$ on $|B| \le j$ edges. Consider the $\{0,1\}^2$-coloured graph G_1 on the vertex set $B \cup B'$, with edges $E(G[B]) \cup J$ and colours

$$\mathrm{Col}_{G_1}(v) = \{red\} \cap \mathrm{Col}_G(v) \quad \text{and} \quad \mathrm{Col}_{G_1}(v') = \{green\} \cap \mathrm{Col}_G(v).$$

for each $v \in B$.

Now let G' be the union of G_1, $G-B$ and the set of edges $E_1 \cup E_2$ defined as follows:

$$E_1 = \{v'x : v \in B, x \in V_{green}, \text{ and } vx \in E(G)\};$$
$$E_2 = \{vx : v \in B, x \in V(G) \setminus (B \cup V_{green}), \text{ and } vx \in E(G)\}.$$

In words, G' is obtained from G by splitting each vertex $v \in B$, so that one of the new vertices inherits the green colour of v (if it had that colour) and all neighbours of v in V_{green} while the other vertex inherits the rest of the neighbours of v. Obviously, $G'/J = G$ if we make sure that the newly created vertices v' have larger labels than those in $V(G)$. By our construction, J separates B and B' in G_1 and each component of $G'-J$ containing green colour can have vertices only in $V_{green} \cup B'$, thus each component of $G'-J$ has at most one colour.

Consider a component C of $G'-J$. Since J is a matching and each edge of J is between different components of $G'-J$, the contraction of J in G' may not put two vertices of C into the same bag. This completes the proof for the case $t=2$.

Suppose $t>2$. Assuming that we have proved the claim for any $t'<t$, we now prove it with $t'=t$. Delete the colour t from G to get a $\{0,1\}^{t-1}$-coloured graph G_1. By induction, there is a $\{0,1\}^{t-1}$-coloured graph G_1' and a set of edges J_1, such that $|J_1| \le (j+1)^{t-2}-1$, each component of $G_1'-J_1$ has at most one colour, it is stable with respect to contraction of J_1 in G_1' and $G_1'/J_1 = G_1$. For a vertex v of G_1'/J_1, denote by $Bag(v)$ the set of vertices of G_1' that contract to v.

Now let us return the colour t back as follows. For each vertex v of G that has colour t in G pick one vertex $v' \in Bag(v) \subseteq V(G')$ and add the colour t to $\mathrm{Col}_{G'}(v')$: we obtain a $\{0,1\}^t$-coloured graph G_2 such that $G_2/J_1 = G$ and each component of G_2-J_1 can have at most two colours.

Now since each component C of G_2-J_1 is stable with respect to contraction of J_1 in G_2, we have $\mathrm{caw}_2(C) \le \mathrm{caw}_2(G) \le j$. Thus, by symmetry, we can apply the already proved case $t=2$ of the lemma to each such C to obtain a $\{0,1\}^t$-coloured graph C' and a set $J_C \subseteq E(C')$ of at most j edges, such that $\mathrm{Col}(C) = \mathrm{Col}(C')$, every component of $C'-J_C$ has at most one colour, is stable with respect to contraction of J_C in C' and $C'/J_C = C$. We assume that the labels for the new vertices are chosen so that they are larger than any label of G_2 and $V(C_1')$ and $V(C_2')$ remain disjoint for distinct components C_1 and C_2 of G_2.

For any $v \in V(C'/J_C)$ denote by $Bag_C(v)$ the set of all vertices of C' that contract to v. Now for any edge $e=xy \in J_1$, let C_x and C_y be the components of G_2-J_1 containing x and y respectively, and define $e'=x'y'$ where x' and y' are any vertices in $Bag_{C_x}(x)$ and $Bag_{C_y}(y)$ respectively. Set $J = \{e' : e \in J_1\} \cup \bigcup_C J_C$, where the union is over the components of $G_2 - J_1$. Finally, let G' be the graph obtained by adding J to the union of the disjoint graphs C', for each component C of $G_2 - J_1$.

Clearly, each component of $G' - J$ has at most one colour. Now consider the operation G'/J in two stages: in the first stage contract all edges $\bigcup_C J_C$, in the second stage contract the edges in J_1. Then at the first step we obtain the graph G_2, and in the second step, we obtain the graph G. Furthermore, if $\tilde{C}$ is a component of $G' - J$, then it is a component of $C' - J_C$ for some C. $\tilde{C}$ is stable with respect to contraction of J_C in C' and C is stable with respect to contraction of J_1 in G_2, therefore $\tilde{C}$ is stable with respect to contraction of J in G'.

Also, since G is connected G_2 has at most $|J| + 1$ components. Therefore $|J| \leq |J_1| + (|J_1| + 1)j \leq (j + 1)^{t-1} - 1$. □

8.3.2 Redundant blockers

The main result of this section is the following lemma (cf. Lemma 7.1.6 of the previous chapter).

Lemma 8.3.3 *Let k be a positive integer, and let $\mathcal{A}$ be a proper addable minor-closed class with a set $\mathcal{B}$ of minimal excluded minors. Suppose that* $\operatorname{aw}_2(\mathcal{A})$ *is finite.*

Then there is a constant $c = c(\mathcal{B}, k)$ such that any graph in $\operatorname{Ex}(k+1)\mathcal{B}$ *has a $\mathcal{B}$-blocker Q of size at most c and a set $S \subseteq Q$ of size at most $2k$ such that any subgraph H of G with $H \notin \mathcal{A}$ that meets Q in at most two points, also meets the set S.*

The proof will follow from a slightly more general result, Lemma 8.3.4 below; we first need a few definitions. Given a graph G, a set of graphs $\mathcal{B}$ and a set $Q \subseteq V(G)$, we say that Q is a $(j, \mathcal{B})$-*blocker* of G if G contains no subgraph H, such that $H \notin \operatorname{Ex}\mathcal{B}$ and $|V(H) \cap Q| \leq j$. We say that Q is a $(j, s, \mathcal{B})$-*blocker* of G if (a) Q is a $\mathcal{B}$-blocker for G and (b) G does not contain s pairwise disjoint subgraphs $H_1, \ldots, H_s \notin \operatorname{Ex}\mathcal{B}$, where each H_i, $i = 1, \ldots, s$ has at most j vertices in Q.

A graph H will be called $\mathcal{B}$-*critical* if $H \notin \operatorname{Ex}\mathcal{B}$ but $H' \in \operatorname{Ex}\mathcal{B}$ for any $H' \subset H$. Notice that if each graph in $\mathcal{B}$ is 2-connected, then so is each $\mathcal{B}$-critical graph.

As in Chapter 7 we will use normal trees for our proofs, see Section 7.3.1 for the definitions. Finally, denote by f^{*n} the n-th iteration of the function f, so that $f^{*0}(x) = x$ and $f^{*(n+1)}(x) = f(f^{*n}(x))$ for $n = 0, 1, \ldots$.

Lemma 8.3.4 *Let k be a positive integer, let $\mathcal{A}$ be a proper addable minor-closed class with a set $\mathcal{B}$ of minimal excluded minors. Suppose that* $\mathrm{aw}_2\,(\mathcal{A}) \leq j$*.*

There are positive constants $c_1 = c_1(\mathcal{B})$ and $w = w(\mathcal{B})$ such that the following holds. Define a function $f : \mathbb{N} \to \mathbb{R}_+$ by $f(q) = jc_1q^2 + jc_1wq$. Suppose Q is a non-empty $(2, k+1, \mathcal{B})$-blocker for a graph G. Then there are sets $S, Q' \subseteq V(G)$, such that $S, Q \subseteq Q'$, $|S| \leq 2k$, $|Q' \setminus S| \leq f^{*k}(|Q|)$ and $Q' \setminus S$ is a $(2, \mathcal{B})$-blocker for $G - S$.

Proof of Lemma 8.3.3 The assumption that $\mathrm{aw}_2\,(\mathcal{A}) < \infty$ implies that some planar graph, a 2-fan, is excluded from $\mathcal{A}$. By the theory of graph minors [88], see also [38] and Proposition 7.3.6 above, there is a constant $c' = c'(\mathcal{B}, k)$ such that every graph $G \in \mathrm{Ex}\,(k+1)\mathcal{B}$ has a $\mathcal{B}$-blocker Q_0 of size at most c' (we may assume Q_0 is non-empty). Such a set Q_0 is clearly also a $(2, k+1, \mathcal{B})$-blocker of G. Now Lemma 8.3.4 ensures that there is a $\mathcal{B}$-blocker Q of size at most $c = c(\mathcal{B}, k) = f^{*k}(c')$ and a set $S \subseteq Q$ of size at most $2k$ as required. □

Proof of Lemma 8.3.4 We use two results from the theory of graph minors of Robertson and Seymour [88]: since $\mathcal{A}$ excludes a planar graph (a 2-fan on $j+2$ vertices), the maximum treewidth over graphs in $\mathcal{A}$, denoted $w = w(\mathcal{B})$, is finite. Furthermore, the set $\mathcal{B}$ is finite.

Let $c_1 = c_1(\mathcal{B})$ be the maximum number of components that can be created by removing three vertices from a $\mathcal{B}$-critical graph. Since there is a finite number of graphs in $\mathcal{B}$, the number c_1 is finite (see Lemma 7.5.2 above). For example, we have $c_1(\{K_4\}) = 4$. Since $\mathcal{A}$ is addable, we have $j \geq 1$ and $c_1 \geq 1$. The case $V(G) = Q$ is also trivial (take $S = \emptyset$ and $Q' = Q$), so we will assume $Q \subset V(G)$.

We will prove the lemma by induction on k. The case $k = 0$ is trivial: we may take $Q' = Q$ and $S = \emptyset$. Assuming the lemma holds for $0 \leq k < k'$, where k' is a positive integer, we prove it for $k = k'$.

By (5.1), the graph $G - Q$ has a (rooted) normal tree T such that $\max_v a_T(v) \leq w$. Let r be the root of T. Form a set U of all such vertices $v \in V(G) \setminus Q$ for which there are some vertices $x, y \in Q \cup AA_T(v)$ that the subgraph of G induced on $V(T_v) \cup \{x, y\}$ has a minor in $\mathcal{B}$. Choose a vertex $u \in U$ with maximum distance from r in T. Let $R = Q \cup AA_T(u)$.

Let P be a set of minimum size such that $G[P \cup V(T_u)] \notin \mathrm{Ex}\,\mathcal{B}$. Then $|P| \in \{1, 2\}$. Fix a $\mathcal{B}$-critical graph H, such that $H \subseteq G[P \cup V(T_u)]$.

Consider the graph $G_1 = G[T_u] - u$. This graph consists of some connected components. Since $H \cap V(G_1) = H - (P \cup \{u\})$, the graph H has vertices in $t \leq c_1(\mathcal{B})$ such components. Call these components C_i, $i = 1, \ldots, t$. Fix a pair

$x, y \in R$ such that $\{x, y\} \cap P = \emptyset$. We claim, that for $i = 1, \ldots, t$ there is a set of at most j vertices $D_i(x, y)$, such that in $C_i - D_i(x, y)$ no component has edges to both x and y.

Let us show why it is true. In the component C_i colour vertices adjacent to x red, and those adjacent to y green to obtain a $\{0,1\}^2$-coloured graph C_i' (vertices adjacent to both x and y are coloured $\{red, green\}$, and the remaining vertices are coloured $\emptyset$). Suppose C_i' has $j + 1$ disjoint connected subgraphs containing both colours. Then since $\mathrm{aw}_2(\mathcal{A}) \leq j$ we would have that $G[\{x, y\} \cup V(C_i)] \notin \mathcal{A}$. But this contradicts to the choice of u: since T is normal, the vertices of the component C_i must be entirely contained in $V(T_{u'})$ for some $u' \in V(T_u - u)$, so $u' \in U$. Thus C_i' cannot have $j+1$ connected subgraphs containing both colours, so we may apply Lemma 8.3.1 to find a suitable set $D_i(x, y)$ of size at most j.

Now define sets S_0, Q_1 as follows. If $|P| = 1$, let $S_0 = P \cup \{u\}$, otherwise, let $S_0 = P$. Set

$$Q_1 = ((Q \cup AA(T_u) \cup \{u\}) \setminus S_0) \cup \bigcup_{i \in [t], x, y \in R \setminus P, x \neq y} D_i(x, y).$$

Writing $q = |Q|$ and considering the cases $|P| = 1$ and $|P| = 2$ separately we get that

$$|Q_1| \leq jc_1(q-1)(q-2+w) + q - 1 + w \leq f(q).$$

If Q_1 is a $(2, k, \mathcal{B})$-blocker for $G - S_0$, then we can use induction to find sets $S', Q' \subseteq V(G) \setminus S_0$ such that $S', Q_1 \subseteq Q'$, $|S'| \leq 2(k-1)$, $|Q' \setminus S'| \leq f^{*(k-1)}(|Q_1|) \leq f^{*k}(q)$ and $Q' \setminus S'$ is a $(2, \mathcal{B})$-blocker for $G - S_0$. Then the lemma follows with $S = S' \cup S_0$ and Q'.

It remains to show that Q_1 is a $(2, k, \mathcal{B})$-blocker for $G - S_0$. Assume this is not true. Let $\tilde{H} \subseteq G$ be a $k\mathcal{B}$-critical subgraph of $G - S_0$ showing that Q_1 is not a $(2, k, \mathcal{B})$-blocker: that is $\tilde{H} = H_1' \cup \cdots \cup H_k'$, where $H_i' \notin \mathrm{Ex}\,\mathcal{B}$, $i = 1, \ldots, k$ are 2-connected and pairwise disjoint; and for each $i \in [k]$ we have $|V(H_i') \cap Q_1| \leq 2$.

Now $\tilde{H}$ and H may not be disjoint: otherwise Q would not be a $(2, k+1, \mathcal{B})$ blocker for G. Let H' be a component of $\tilde{H}$ which shares at least one vertex with H. Then $V(H') \cap V(H) \subseteq V(H) \setminus S_0 \subseteq V(T_u)$.

Suppose first that $V(H') \cap Q_1$ consists of a single vertex v. Note, that we must have $v \in Q \setminus S_0$. The graph $H' - v$ cannot be entirely contained in $G[V(T_u)]$. To see why, observe first that by our construction in this case $u \in S_0$. Since $H' - v$ is connected it must be entirely contained in one of the proper subtrees of T_u, but this contradicts our choice of u.

Thus $H'-v$ must have a vertex a in $V(T_u-u)$ and a vertex b in $G-(S_0\cup Q_1\cup V(T_u))$. The set $AA_T(u)\subseteq Q_1\setminus\{v\}$ separates $V(T_u-u)$ from $G-(S_0\cup Q_1\cup V(T_u)$. On the other hand, there is a path from a to b in the connected graph $H'-v$: this is a contradiction.

Now suppose H' has exactly two vertices x, y in Q_1.

First consider the case where $x, y \in Q \cup AA_T(u)$. Let a be a vertex in $V(H)\cap V(H')$. It cannot be $a=u$ since in this case we have that $|V(H')\cap Q_1| \geq 3$. It follows that $a \in V(C')$ where C' is a component of $C_i - D_i(x,y)$ for some $i \in \{1,\ldots,t\}$. But C' cannot have edges to both x and y. This means that either x or y is a cut vertex in H': this contradicts the fact that $\tilde{H}$ is $k\mathcal{B}$-critical.

If $x \in Q$ and $y \in D_i(x',y')$ for some pair $\{x',y'\}$, then suppose that $H'-x$ is contained in $G[V(T_u)]$. By the choice of the set P this means that $u \in S_0$. But since T is normal, this contradicts the definition of u. Otherwise, suppose that $H'-x$ has a vertex in $G-(S_0\cup Q_1\cup V(T_u))$. Since $AA_T(u)\subseteq Q_1\setminus\{x,y\}$ we have that x must be a cut point in H': this is a contradiction to the $k\mathcal{B}$-criticality of $\tilde{H}$.

Finally, consider the case $x \in Q$ and $y = u$. Note, that the only case when $u \notin S_0$ by our construction is when there is no vertex z such that $G[V(T_u)\cup\{z\}]$ has a minor in $\mathcal{B}$. Thus H must contain at least one vertex in $G-(S_0\cup Q_1\cup V(T_u))$, and we saw earlier that it has a vertex in $(V(H)\setminus S_0)\subseteq V(T_u)$.

Again, each path in H' from $V(H')\cap V(T_u)$ to $V(H')\cap(V(G)\setminus(S_0\cup Q_1\cup T_u))$ must use x, since $AA_T(u)\subseteq Q_1\setminus\{x,y\}$ separates T_u from the rest of $G-Q$. So we obtain a contradiction to the fact that H' is 2-connected. In all of the cases we obtained a contradiction, so we conclude that Q_1 must be a $(2,k,\mathcal{B})$-blocker for $G-S_0$. ◻

8.3.3 Blockers of size $2k$

We will need another definition. We call a $\mathcal{B}$-blocker Q of a graph G a $(k,j,\mathcal{B})$-*double blocker* if there is a set $S\subseteq Q$ of size at most k, which is a redundant $\mathcal{B}$-blocker for $G-(Q\setminus S)$, and $Q\setminus S$ is a $(j,\mathcal{B})$-blocker for $G-S$. Such a set S is called a *special set* of Q.

Lemma 8.3.5 *Let k,l be positive integers and let $\mathcal{B}$ be the set of minimal excluded minors of a proper addable minor-closed class and assume that $\mathcal{B}$ contains a planar graph. Then there is a positive constant $w=w(\mathcal{B})$ such that the following holds.*

Suppose $G \in \mathrm{Ex}\,(k+1)\mathcal{B}$ has a $\mathcal{B}$-blocker Q of size at most q and a set $S\subseteq Q$ of size at most l such that $Q\setminus S$ is a $(2,\mathcal{B})$-blocker for $G-S$. Then G can be *represented as the union of two graphs, $G = G_1 \cup G_2$, where*

- *the graph G_1 has an $(l, 2, \mathcal{B})$-double blocker $Q_1 \supseteq Q$ of size at most $q+w+1$, such that S is the special set;*
- *$G_2 \in apex(\mathrm{Ex}\, k\mathcal{B})$ and $Q \subseteq V(G_1) \cap V(G_2) \subseteq Q_1$.*

Proof The set $\mathcal{B}$ contains a planar graph, so by the theory of graph minors, see [38, 88], the treewidth of $G - Q$ is bounded by a constant $w = w(\mathcal{B})$. Since the claim is trivial in the case $Q = V(G)$ (take $G_1 = G$, $Q_1 = Q$ and $G_2 = \bar{K}_Q$, where $\bar{K}_Q$ is the empty graph on Q), we will assume that $Q \subset V(G)$. By the Kloks theorem (5.1), $G - Q$ has a normal tree T such that the number of active ancestors satisfies $a_T(v) \leq w$ for each $v \in V(T)$. Denote by r the root of T.

Let U be the set of all vertices $v \in V(G-Q)$ such that $G[V(T_v) \cup \{x\}] \notin \mathrm{Ex}\,\mathcal{B}$ for some $x \in S$. If $U = \emptyset$ then Q is itself an $(l, 2, \mathcal{B})$-double blocker for G, so we may take $G_1 = G$ and $G_2 = \bar{K}_Q$. Now assume that U is non-empty. Let $u \in U$ be a vertex with maximum distance in T from the root r and let x_0 be a vertex in S showing that $u \in U$. Write $A = AA_T(u)$, let $G_1 = G[V(T_u) \cup Q \cup A]$, and let $G_2 = G - V(T_u)$. We claim that G_1 and G_2 are as required.

We have $V(G_1) \cap V(G_2) = Q \cup A$, so G_1 and G_2 share $|Q \cup A| \leq q+w$ vertices.

We will show that $Q_1 = Q \cup A \cup \{u\}$ is an $(l, 2, \mathcal{B})$-double blocker for G_1, and S is its special set. Indeed, using the assumption of the lemma, the set $Q_1 \setminus S = (Q \setminus S) \cup A \cup \{u\}$ is a $(2, \mathcal{B})$-blocker for $G_1 - S$. Now suppose that $G[V(T_u - u) \cup \{z\}] \notin \mathrm{Ex}\,\mathcal{B}$ for some $x \in S$. Let H be a $\mathcal{B}$-critical subgraph of $G[V(T_u - u) \cup \{x\}]$. Then, since T is normal, all vertices of the connected graph $H - x$ must be contained in $V(T_v)$ for some v strictly below u in T. This is a contradiction to the choice of u.

Finally observe that if $G_2 - x_0$ contains a minor in $k\mathcal{B}$ then since $V(T_u) \cup \{x_0\}$ and $V(G_2 - x_0)$ are disjoint, $G \notin \mathrm{Ex}\,(k+1)\mathcal{B}$. So $G_2 \in \mathrm{apex}\,(\mathrm{Ex}\, k\mathcal{B})$ is as claimed.

□

In the proof of the next lemma and in much of the remaining part of the Chapter, we will find it more convenient to represent graphs with small blockers as coloured graphs, where a colour class corresponds to a vertex in the blocker, and the set of colours of a vertex corresponds to the set of its neighbours in the blocker. What follows is an attempt to capture this formally.

Let r be a fixed integer. We call a graph G with r distinct distinguished vertices, or roots, an *r-rooted graph*. The roots will be labelled and ordered. We say that a class Π of r-rooted graphs is an r-property if Π is closed under isomorphism and under deleting edges between the roots. We say that an unrooted graph G has r-property Π if it is possible to root r of its vertices, so that the resulting r-rooted graph is in Π. We associate two classes with Π: the class $\mathcal{A}_\Pi$ of uncoloured, unrooted graphs that have r-property Π and the class $\tilde{\mathcal{A}}_\Pi$ of $\{0,1\}^r$-coloured graphs G, such that if $q_1 < \cdots < q_r$ are not elements of $V(G)$, then $G^{\{q_1,\dots,q_r\}}$ with roots $(q_1,\dots,q_r)$ belongs to Π.

The next proposition just spells out the well known fact that a class of rooted graphs has the same radius of convergence as the class of corresponding unrooted graphs.

Proposition 8.3.6 *Let Π be an r-property for some positive integer r. Then the sequence (5.2) for the class of rooted graphs Π, the class of unrooted graphs $\mathcal{A} = \mathcal{A}_\Pi$ and the class of $\{0,1\}^r$-coloured graphs $\tilde{\mathcal{A}} = \tilde{\mathcal{A}}_\Pi$ has the same set L of limit points, $L \subseteq [0;\infty]$.*

Proof The claim follows, since

$$|\tilde{\mathcal{A}}_n| \le |\mathcal{A}_{n+r}| \le |\Pi_{n+r}| \le 2^{\binom{r}{2}}(n+r)_r|\tilde{\mathcal{A}}_n|,$$

and if (5.2) has a limit for a subsequence $(n_k, k=1,\dots)$ for any of the classes, it has the same limit for the other two. □

The most important r-property for us will be the property of having a redundant blocker of size r. Formally, given a set $\mathcal{B}$ of graphs and a positive integer r, Π_0 is the set of all r-rooted graphs G, such that any subgraph of G containing just one of the roots of G is in $\mathrm{Ex}\,\mathcal{B}$.

Define $\mathcal{A}_{r,\mathcal{B}} = \tilde{\mathcal{A}}_{\Pi_0}$ and notice that $\mathcal{A}_{\Pi_0} = \mathrm{rd}_r\,\mathcal{B}$. To keep the notation simpler, below we will omit the subscript $\mathcal{B}$, since the class $\mathcal{B}$ will always be fixed. Proposition 8.3.6 implies that the $\rho(\mathcal{A}_r) = \rho(\mathrm{rd}_r\,\mathcal{B})$. We will also denote by $\mathcal{C}^r = \mathcal{C}^{r,\mathcal{B}}$ the class of connected graphs in $\mathcal{A}_r$.

Given a set of graphs $\mathcal{B}$ and a coloured graph G with $N(G) \in \mathrm{Ex}\,\mathcal{B}$, call a colour c *bad* for G (with respect to $\mathcal{B}$), if $N(G) \in \mathrm{Ex}\,(\mathcal{B})$, but adding to G a new vertex x_c connected to every vertex $v \in V(G)$ which has colour c we produce a graph with a minor in $\mathcal{B}$. Otherwise, call c good for G. Notice, that $G \in \mathcal{A}_k$ if and only if $N(G) \in \mathrm{Ex}\,\mathcal{B}$ and $\mathrm{Col}(G) \subseteq [k]$ and every colour is good for G.

Lemma 8.3.7 *Let k and r be positive integers, such that $k < r$, and let $\mathcal{B}$ be a finite set of graphs. Suppose $\mathrm{aw}_2\,(\mathrm{Ex}\,\mathcal{B})$ is finite. Let $\mathcal{A}$ be the class of graphs that have a $(k,2,\mathcal{B})$-double blocker of size r. Then $\rho(\mathcal{A}) = \rho(rd_{k+1}\,\mathcal{B})$.*

Proof Let Π be the r-property for "containing a $(k, 2, \mathcal{B})$-double blocker of size r", i.e., Π is the set of all graphs $G \in \mathcal{A}$ with r distinct roots $q_1, \ldots, q_r$ so that $\{q_1, \ldots, q_r\}$ is a $(k, 2, \mathcal{B})$-double blocker for G with a special set $\{q_1, \ldots, q_k\}$. Then $\mathcal{A} = \mathcal{A}_\Pi$ and $\tilde{\mathcal{A}} = \tilde{\mathcal{A}}_\Pi$ have the same radius of convergence by Proposition 8.3.6.

Let $\tilde{\mathcal{C}}$ be the class of connected graphs in $\tilde{\mathcal{A}}$. The exponential formula, see, i.e., [48], gives that for $n = 0, 1, 2, \ldots$

$$[x^n]\tilde{C}(x) \leq [x^n]\tilde{A}(x) \leq [x^n]e^{\tilde{C}(x)}$$

so $\rho(\tilde{\mathcal{C}}) = \rho(\tilde{\mathcal{A}})$. Similarly $\rho(\mathcal{C}^{k+1}) = \rho(\mathcal{A}_{k+1})$. Therefore by Proposition 8.3.6 it suffices to prove that

$$\rho(\tilde{\mathcal{C}}) = \rho(\mathcal{C}^k). \tag{8.2}$$

We have $\mathcal{C}_n^k \subseteq \tilde{\mathcal{C}}_n$, so $\rho(\mathcal{C}^k) \geq \rho(\tilde{\mathcal{C}})$. The difficult part is the opposite inequality. Our idea is to use the "Colour reduction lemma", Lemma 8.3.2, to represent each graph in $\tilde{\mathcal{C}}$ as a transformation of a finite set of disjoint graphs in $\mathcal{C}^k$.

Write $a = \mathrm{aw}_2\,(\mathrm{Ex}\,\mathcal{B})$. Consider a $\{0,1\}^r$-coloured graph $G \in \tilde{\mathcal{C}}$. Let G' be a $\{0,1\}^r$-coloured graph obtained by removing the colours $\{1, \ldots, k\}$ from G. Suppose $\mathrm{caw}_2\,(G') > a$. Then for any set $L = \{l_1, \ldots, l_r\}$ such that $l_1 < \cdots < l_r$ and $L \cap V(G) = \emptyset$, the graph $G^L - \{l_1, \ldots, l_k\}$ has a subgraph $H \notin \mathrm{Ex}\,\mathcal{B}$ such that H has at most two vertices in $\{l_{k+1}, \ldots, l_r\}$. But by the definition of $\tilde{\mathcal{A}}$ (and $\tilde{\mathcal{C}}$), $\{l_{k+1}, \ldots, l_r\}$ is a $(2, \mathcal{B})$-blocker for $G^L - \{l_1, \ldots, l_k\}$, a contradiction. Therefore $\mathrm{caw}_2\,(G') \leq a$.

By Lemma 8.3.2, there is a graph G_1 obtained from the union of $\kappa \leq N = (a+1)^{r-k-1}$ disjoint graphs, each with at most one colour in $\{k+1, \ldots, r\}$, and a set J of $m \leq N - 1$ edges between these graphs, such that $G_1/J = G'$ and each component of $G_1 - J$ is stable with respect to G_1/J.

Now return the colours $\{1, \ldots, k\}$ back: starting with the coloured graph G_1, for each $c \in \{1, \ldots, k\}$ and each $v \in V(G_1/J)$, add c to the set of colours for one of the vertices $v' \in Bag(v)$. Denote the newly obtained graph by G''. Then $G''/J = G$. Each component C of $G'' - J$ can have at most one colour $c \in \{k+1, \ldots, r\}$, and so at most $k+1$ colours in total. Crucially, if C contains a colour $c \in \{k+1, \ldots, r\}$, we can map the colour c to $k+1$, and this yields a graph C' in $\mathcal{C}^{k+1}$. Why? Since C is stable with respect to contraction of J in G_1 (and also in G''), C is isomorphic to a (coloured) subgraph of G. If there was a colour $j \in \mathrm{col}(C)$ which was bad for C, then it would be bad for G. But $G \in \tilde{\mathcal{A}}$, this gives a contradiction.

Recall that we assume that a contraction of an edge xy produces a new vertex with label $\min(x,y)$. Thus, each graph $G \in \tilde{\mathcal{C}}_n$ (as well as other graphs) can be obtained by choosing integers κ, l, m with $0 \le l, m, \kappa - 1 \le N-1$, a graph $G_0 \in \left(\mathcal{C}^{k+1}\right)^{\kappa}_{n+l}$, for each component of G_0, mapping the colour $k+1$ to an arbitrary colour in $\{k+1,\dots,r\}$, adding a set J of m edges to G_0 and finally contracting them. Therefore we have

$$\left[\frac{x^n}{n!}\right]\tilde{C}(x) \le \sum_{l=0}^{N-1}\sum_{\kappa=1}^{N}\sum_{m=0}^{N-1}(n+l)^{2m}\left[\frac{x^{n+l}}{(n+l)!}\right]\left((r-k)C^{k+1}\right)^{\kappa},$$

from which it follows that $\rho(\tilde{\mathcal{C}}) \ge \rho(\mathcal{C}^{k+1})$. □

Lemma 8.3.8 *Let $\mathcal{B}$ be any set of graphs. Let k and r be positive integers, $k \ge r$. Then*

$$\overline{\gamma}(rd_k\,\mathcal{B})^2 \le \overline{\gamma}(rd_{k+r}\,\mathcal{B})\overline{\gamma}(rd_{k-r}\,\mathcal{B}).$$

Proof By Proposition 8.3.6 it suffices to show that $\overline{\gamma}(\mathcal{A}_k)^2 \le \overline{\gamma}(\mathcal{A}_{k+r})\overline{\gamma}(\mathcal{A}_{k-r})$ (see above for the definition of $\mathcal{A}_k$). Fix positive integers l, n. Note that $\mathcal{A}_0 = \mathrm{Ex}\,\mathcal{B}$. We can partition the class $\mathcal{A}_{l,n}$ (of $\{0,1\}^l$-coloured graphs on vertex set $[n]$) into $|\mathcal{A}_{0,n}|$ disjoint subclasses according to the underlying uncoloured graph $N(G)$ of $G \in \mathcal{A}_{l,n}$.

Given an uncoloured graph $G \in \mathcal{A}_0$, let X_G be the number of ways to add colour 1 so that the resulting $\{0,1\}^1$-coloured graph is in $\mathcal{A}_1$. Since the constraint for redundant blockers involves only individual vertices, we can pick the sets of vertices coloured $1,2,\dots,l$ independently, in X_G^l ways. Therefore

$$|\mathcal{A}_{l,n}| = \sum_{G \in \mathcal{A}_{0,n}} X_G^l.$$

We see that the equality also holds when $l=0$. Choose G from $\mathcal{A}_{0,n}$ uniformly at random. Then $X = X_G$ is a random variable and $|\mathcal{A}_{l,n}| = |\mathcal{A}_{0,n}|\mathbb{E}\,X^l$. By the Cauchy-Schwarz inequality applied with random variables $X^{(l-r)/2}$ and $X^{(l+r)/2}$ we have

$$(\mathbb{E}\,X^l)^2 \le \mathbb{E}\,X^{l+r}\mathbb{E}\,X^{l-r},$$

or

$$|\mathcal{A}_{l,n}|^2 \le |\mathcal{A}_{l+r,n}||\mathcal{A}_{l-r,n}|.$$

Now the claim follows by dividing each side by $(n!)^2$, raising to $1/n$ and considering the subsequence that realizes the upper limit of the left side. □

Lemma 8.3.9 *Let $\mathcal{A}$ be a proper addable minor-closed class of graphs with a set $\mathcal{B}$ of minimal excluded minors. Suppose* $\mathrm{aw}_2\,(\mathrm{Ex}\,\mathcal{B})$ *is finite and there is a positive integer k_0 such that*

$$r = \frac{\overline{\gamma}(\mathrm{Ex}\,(k_0+1)\mathcal{B})}{\overline{\gamma}(\mathrm{Ex}\,k_0\mathcal{B})} > 2.$$

Then for any $k \geq k_0$

$$\overline{\gamma}(\mathrm{Ex}\,(k+1)\mathcal{B}) = \overline{\gamma}(rd_{2k+1}\,\mathcal{B}) \geq r\overline{\gamma}(\mathrm{Ex}\,k\mathcal{B}).$$

Proof In the proof we will need the following important consequence of the preceding lemmas.

Let t be a positive integer. By Lemma 8.3.3, every graph $G \in \mathrm{Ex}\,(t+1)\mathcal{B}$ has a $\mathcal{B}$-blocker Q, such that Q contains a set S of size at most $2t$ and $Q \setminus S$ is a $(2, \mathcal{B})$-blocker for $G - S$. Furthermore, the size of Q is bounded by a constant $c = c(\mathcal{B}, t)$.

For any integer $j \geq 0$ write $\overline{\gamma}_j = \overline{\gamma}(\mathrm{rd}_j\,\mathcal{B})$. Then

$$\overline{\gamma}(\mathrm{Ex}\,(t+1)\mathcal{B}) = \max(\overline{\gamma}_{2t+1}, \overline{\gamma}(\mathrm{apex}\,(\mathrm{Ex}\,t\mathcal{B}))). \tag{8.3}$$

Let us prove (8.3). Write $d = c + w$ where $w = w(\mathcal{B})$ is as in Lemma 8.3.5. We claim that for $n \geq d+1$ we have

$$[x^n]A(x) \geq [x^n]A_1(x)A_2(x), \tag{8.4}$$

where $A(x), A_1(x), A_2(x)$ are exponential generating functions of $\mathrm{Ex}\,(t+1)\mathcal{B}$, the class $\mathcal{A}_1$ of graphs that have a $(2t, 2, \mathcal{B})$-double blocker of size $d+1$ with d rooted vertices, and the class $\mathcal{A}_2$ of graphs in $\mathrm{apex}\,(\mathrm{Ex}\,t\mathcal{B})$ which have d pointed (i.e., unlabelled, but distinguishable) vertices respectively.

(8.4) can be seen as follows. Given graphs $G_1 \in \mathcal{A}_1$ and $G_2 \in \mathcal{A}_2$ with disjoint labels we can obtain a new graph by identifying the i-th distinguished vertex of G_1 with the i-th distinguished vertex of G_2 for $i = 1, \dots, d$ and removing repetitive edges. By Lemma 8.3.5, the set of all resulting graphs will contain all graphs in $\mathrm{Ex}\,(k+1)\mathcal{B}$ of size at least $d+1$. Note that if $G \in \mathrm{Ex}\,(k+1)\mathcal{B}$ has at least $d+1$ vertices, we may assume that G_1, G_2 given by Lemma 8.3.5 are such that G_1 has exactly $d+1$ vertices, $|Q_1| = d+1$ and $|V(G_2) \cap V(G_2)| = d$; otherwise G_1, G_2 can be extended by including extra isolated vertices, and Q_1 can be extended by adding more vertices from $G_1 - Q_1$.

Now rooting or pointing a fixed number of vertices does not change the convergence radius of a class (see, i.e., Proposition 8.3.6), therefore A_1 has the convergence radius $\rho_1 = \overline{\gamma}_{2t+1}^{-1}$ by Lemma 8.3.7, A_2 has convergence radius $\rho_2 = \overline{\gamma}(\mathrm{apex}\,(\mathrm{Ex}\,t\mathcal{B}))^{-1}$, and by the theory of generating functions, see i.e., [48], the convergence radius ρ of A is at least $\min(\rho_1, \rho_2)$. Finally, ρ is exactly this, since both $\mathrm{rd}_{2t+1}\,\mathcal{B}$ and $\mathrm{apex}\,(\mathrm{Ex}\,t\mathcal{B})$ are contained in $\mathrm{Ex}\,(t+1)\mathcal{B}$.

We will also use a simple bound $\overline{\gamma}(\mathrm{apex}\,(\mathcal{D})) \leq 2\overline{\gamma}(\mathcal{D})$, which is valid for any class of graphs $\mathcal{D}$, since $|\mathrm{apex}\,(\mathcal{D})_n| \leq n2^{n-1}|\mathcal{D}_{n-1}|$.

Let us now prove the lemma. We use induction on k. First consider the case $k = k_0$. We have $\overline{\gamma}(\mathrm{apex}\,(\mathrm{Ex}\,k_0\mathcal{B})) \leq 2\overline{\gamma}(\mathrm{Ex}\,k_0\mathcal{B}) < \overline{\gamma}(\mathrm{Ex}\,(k_0+1)\mathcal{B})$ and so by (8.3), only one candidate to realize $\overline{\gamma}(\mathrm{Ex}\,(k_0+1)\mathcal{B})$ remains:

$$\overline{\gamma}(\mathrm{Ex}\,(k_0+1)\mathcal{B}) = \overline{\gamma}_{2k_0+1}.$$

Now let $k' > k_0$ be an integer, assume we have proved the lemma with $k < k'$, let us now prove the case $k = k'$.

We have $\overline{\gamma}(\mathrm{Ex}\,k\mathcal{B}) = \overline{\gamma}_{2k-1}$ by induction, therefore

$$\overline{\gamma}(\mathrm{apex}\,(\mathrm{Ex}\,k\mathcal{B})) \leq 2\overline{\gamma}_{2k-1}.$$

Now Lemma 8.3.8 and induction yields

$$\overline{\gamma}_{2k+1} \geq \overline{\gamma}_{2k-1}\frac{\overline{\gamma}_{2k-1}}{\overline{\gamma}_{2k-3}} \geq r\overline{\gamma}_{2k-1}.$$

So finally

$$\overline{\gamma}_{2k+1} \geq r\overline{\gamma}_{2k-1} > 2\overline{\gamma}_{2k-1} \geq \overline{\gamma}(\mathrm{apex}\,(\mathrm{Ex}\,k\mathcal{B}))$$

and the claim follows by (8.3). □

We are now ready to prove our first main theorem.

Proof of Theorem 8.1.1 We use the notation from Lemma 8.3.9. By Lemma 8.3.10 below, $\mathrm{aw}_2\,(\mathcal{A}) < \infty$. As has been noticed in Chapter 7, since $\mathcal{A}$ contains all fans, $\mathrm{apex}^{2k+1}\,(\mathcal{P}) \subseteq \mathrm{Ex}\,(k+1)\mathcal{B}$, where $\mathcal{P}$ is the class of paths. So we have $\underline{\gamma}(\mathrm{Ex}\,(k+1)\mathcal{B}) \geq 2^{2k+1}$ for $k = 1, 2, \ldots$. Also by Theorem 7.1.2, $\gamma(\mathrm{apex}^k\,(\mathcal{A})) = 2^k\gamma$.

Let k_0 be the smallest positive integer, such that

$$\overline{\gamma}(\mathrm{Ex}\,(k_0+1)\mathcal{B}) > \gamma(\mathrm{apex}^{k_0}\,(\mathcal{A})).$$

Then for $1 \le j < k_0$, by applying (8.3) j times we have

$$\overline{\gamma}(\text{Ex}\,(j+1)\mathcal{B}) = 2^j\gamma.$$

Since $\text{apex}^j\,(\mathcal{A}) \subseteq \text{Ex}\,(j+1)\mathcal{B}$, it follows that $2^j\gamma$ is the growth constant of $\text{Ex}\,(j+1)\mathcal{B}$. Thus

$$\overline{\gamma}(\text{Ex}\,(k_0+1)\mathcal{B}) > 2^{k_0}\gamma = 2\gamma(\text{Ex}\,k_0\mathcal{B}),$$

therefore $\overline{\gamma}(\text{Ex}\,(k+1)\mathcal{B}) = \overline{\gamma}(\text{rd}_{\,2k+1}\,\mathcal{B})$ for all $k \ge k_0$ by Lemma 8.3.9.

Let us show that for $k \ge k_0$,

$$\overline{\gamma}\left((\text{Ex}\,(k+1)\mathcal{B}) \cap \text{apex}^{2k-1}\,(\mathcal{A})\right) < \overline{\gamma}_{2k+1}.$$

By Lemma 8.3.9 we have

$$\overline{\gamma}(\text{apex}\,(\text{Ex}\,k\mathcal{B})) \le 2\overline{\gamma}_{2k-1} < \overline{\gamma}_{2k+1}.$$

So, using Lemma 8.3.8

$$\overline{\gamma}_{2k} \le \sqrt{\overline{\gamma}_{2k+1}\overline{\gamma}_{2k-1}} \le 2^{-1/2}\overline{\gamma}_{2k+1}.$$

Now apply Lemma 8.3.5 (with $k, \mathcal{B}$ and $l = 2k-1$), Lemma 8.3.7 and an inequality analogous to (8.3), to get

$$\overline{\gamma}\left((\text{Ex}\,(k+1)\mathcal{B}) \cap \text{apex}^{2k-1}\,(\mathcal{A})\right) = \max(\overline{\gamma}_{2k}, \overline{\gamma}(\text{apex}\,(\text{Ex}\,k\mathcal{B}))) < \overline{\gamma}_{2k+1}.$$

Finally, let us show that (8.1) holds in the case $\overline{\gamma}_{2k+1} < 2^k\gamma$. Note that, in such case $1 \le k < k_0$. It cannot be that for some $j \in \{1, \dots, k-1\}$ we have $\overline{\gamma}_{2j+1} > 2^j\gamma$, since Lemma 8.3.9 would imply that $\overline{\gamma}_{2k+1} > 2^k\gamma$. Similarly, if $\overline{\gamma}_{2j+1} = 2^j\gamma$, we would have $\overline{\gamma}_{2k+1} \ge 2^{k-j}\overline{\gamma}_{2j+1} \ge 2^k\gamma$ by Lemma 8.3.8. Thus $\overline{\gamma}_{2j+1} < 2^j\gamma$ for all $j = 1, \dots, k-1$.

Trivially, $|(\text{Ex}\,(0+1)\mathcal{B})_n| = |\mathcal{A}_n| = |(\text{apex}^0\,(\mathcal{A}))_n|$. Using Lemma 8.4.11 (given in Section 8.4.2 below) and induction, we get that for $j \in \{1, \dots, k\}$

$$|(\text{Ex}\,(j+1)\mathcal{B})_n| = |(\text{apex}^j\,(\mathcal{A}))_n|(1 + e^{-\Theta(n)}).$$

□

8.3.4 When is apex width finite?

In Section 8.3.1 we introduced the apex width parameter, which is not standard. Here we present a characterisation of classes with bounded apex width in terms of excluded minors.

Lemma 8.3.10 *Let $\mathcal{A}$ be a minor-closed class. Then* $\mathrm{aw}_j(\mathcal{A}) < \infty$ *if and only if some j-fan and some bipartite graph $K_{j+1,t}$ does not belong to $\mathcal{A}$.*

For stating Theorem 8.1.2 in terms of minors, we will need another lemma.

Lemma 8.3.11 *Let $\mathcal{A}$ be a minor-closed class such that* $\mathrm{aw}_2(\mathcal{A}) < \infty$*. Then the following two statements are equivalent.*

(1) There is a constant c such that for each $G \in \mathcal{A}$ and each $v \in V(G)$, if $G - v$ is 2-connected then the degree of v in G is at most c.

(2) Some wheel does not belong to $\mathcal{A}$.

To prove the above lemmas, we need a few simple preliminary results. Recall that the *height* of a rooted tree is the number of edges in the longest path starting from the root. A *leaf* of a rooted tree is a vertex of degree 1, which is not the root. A straightforward fact is:

Lemma 8.3.12 *Let T be a rooted tree of size n, height h and with l leaves. Then $lh \geq n-1$.*

For positive integers j and s, denote by F^j_s the j-fan on s vertices. Also, let $K^*_{j,s}$ denote a graph obtained from the union of a $K_{j,s}$ and a $(j-1)$-star on the part of size j. Note, that for $j \leq s+1$, $\mathrm{aw}_j(K^*_{j+1,s}) = s+1$, and $K^*_{j+1,s}$ is isomorphic to a minor of $K_{j+1,s+j}$.

Proof of Lemma 8.3.10 If $\mathcal{A}$ contains all j-fans then $\mathrm{aw}_j(\mathcal{A}) = \infty$ by definition. If $\mathcal{A}$ contains all graphs $K_{j+1,t}$, then it contains all graphs $K^*_{j+1,t}$, where t is arbitrary and j is fixed, so again $\mathrm{aw}_j(\mathcal{A}) = \infty$.

Now suppose there are integers $t \geq 1$ and $s \geq 2$ such that $F^j_{j+s+1}, K_{j+1,t} \notin \mathcal{A}$. Let T be any tree on at least st vertices, let S be a set of j vertices, disjoint from T, and let H be the union of T and the complete bipartite graph with parts S and $V(T)$. If T has a path of length at least s, then H has a subgraph isomorphic to F^j_{j+s+1}. Otherwise we can root T at a vertex r, so that the resulting rooted tree T_r has height at most $s-1$. By Lemma 8.3.12, T_r has at least $\lceil \frac{st-1}{s-1} \rceil \geq t$ leaves. Therefore, contracting all internal vertices of T_r into a single vertex and using vertices in S, we obtain a minor of $K_{j+1,t}$. We have shown that $\mathrm{aw}_j(\mathcal{A}) < st$. □

It is well known that each 2-connected graph on at least 3 vertices has a contractible edge (so that contracting this edge yields again a biconnected graph), see, e.g., [38]. We need a simple refinement of this.

Lemma 8.3.13 *Let G be a 2-connected graph on at least 3 vertices, and let $x \in V(G)$. There is an edge $xy \in E(G)$, such that G/xy is biconnected.*

Proof Assume the claim is false: then for each neighbour u of x, $\{u, x\}$ must be a cut in G. Denote by $C(u)$ a component of $G - \{x, u\}$ of minimal size, and let u' minimize $|V(C(u))|$ over the neighbours u of x. Since G is 2-connected, Both x and u must have neighbours in $C(u')$. Let $z \in C(u')$ be a neighbour of x in G. Suppose $\{x, z\}$ is a cut in G. The graph $(G - C(u')) - x$ is connected (this follows using Menger's theorem), so $G - \{x, z\}$ must have a connected component which is strictly contained in $C(u')$. But this contradicts to the definition of u'. We conclude that $\{x, z\}$ is not a cut in G, so G/xz must be 2-connected. □

The following "simple fact" about the size of largest cycle (circumference) of a 2-connected graph is stated in [84]. For completeness, we include a proof.

Lemma 8.3.14 *Let $k \geq 3$ be an integer. There is a positive integer $N = N(k)$ such that each 2-connected graph with at least N vertices contains either a cycle of length k as a subgraph or the complete bipartite graph $K_{2,k}$ as a minor.*

Proof Write $N = k^{3k^2} + 1$, $\Delta = k^3 + 1$, and let G be a 2-connected graph of size at least N.

Let P be the longest path in G, and let x, y be its endpoints. Suppose P has length at least k^2. By Menger's theorem, G has a cycle C containing x and y. We can assume that $|V(C)| \leq k - 1$. The vertices in $V(C) \cap V(P)$ partition P into at most $k - 1$ subpaths, with internal vertices disjoint from C (and endpoints in C). One of these subpaths must have at least k vertices. This subpath, together with a part of the cycle C yields a cycle of length at least k in G.

Therefore we may assume that each path of G has length at most $k^2 - 1$. Let T_r be a rooted spanning tree of G. A rooted tree with maximum degree at most $\Delta - 1$ and height h can have at most

$$1 + (\Delta - 1) + \cdots + (\Delta - 1)^h \leq (\Delta - 1)^{h+1}$$

vertices. Since $|V(G)| \geq N$, T (and G) has a vertex v of degree at least Δ.

Consider the $\{0, 1\}^1$-coloured connected graph G' obtained from $G - v$ by setting $C_{G'}(u) = \{red\}$ for each neighbour u of v in G. Let T' be a minimal

(Steiner) subtree of G' containing all the red vertices. $|V(T')| \geq \Delta$ and T' has diameter at most k^2. By Lemma 8.3.12, T' has at least $\lceil(\Delta-1)/k^2\rceil \geq k$ leaves. By the minimality of T', each leaf has the red colour. Contract the internal vertices of T' to a single vertex; this vertex, the leaves of T' and the vertex x demonstrate that G has a minor $K_{2,k}$. □

Proof of Lemma 8.3.11 It is trivial to see that if $\mathcal{A}$ has arbitrarily large wheels, then (1) does not hold: the "hub" vertex x of a wheel W_t has $t-1$ neighbours, and $W_t - \{x\}$ is 2-connected. Suppose a wheel W_r is excluded from $\mathcal{A}$ and (1) does not hold: we will obtain a contradiction. Set $j = \mathrm{aw}_2\,(\mathcal{A})$. Take a graph $G \in \mathcal{A}$ and a vertex x such that $G-\{x\}$ is 2-connected, and x has $d \geq N(k)+3$ neighbours, where $k = \max(r-1, j+2)$ and $N(k)$ is as in Lemma 8.3.14. Let G' be the $\{0,1\}^1$-coloured graph obtained from $G-\{x\}$ by colouring the former neighbours of u $\{red\}$. By Lemma 8.3.13, the graph G' has an all-red 2-connected minor H of size d (repeatedly contract a contractible edge incident to an uncoloured vertex, until no uncoloured vertices remain). Now by Lemma 8.3.14, H either has a cycle C of length $k \geq r-1$, or $K_{2,k}$ as a minor. Recalling that v in G is incident to each red vertex of G', we get that $G \in \mathcal{A}$ has a minor W_{k+1} in the first case (contradiction to the fact that $W_r \notin \mathcal{A}$). In the second case, we see that G has a minor $K_{3,j+2}$, therefore also a minor $K^*_{3,j}$, so $\mathrm{aw}_2\,(\mathcal{A}) \geq j+1$, a contradiction. □

8.4 Growth constants for $\mathrm{Ex}\,(k+1)\mathcal{B}$

8.4.1 Proof of Theorem 8.1.2

In this section we prove Theorem 8.1.2. Similarly as McDiarmid, Steger and Welsh [77], we will make use of a version of Fekete's lemma.

Lemma 8.4.1 *Suppose $(f(n), n = 1,2,\dots)$ is a sequence of real numbers such that for any positive integers n, m*

$$f(n+m+1) \geq f(n) + f(m).$$

Then

$$\sup \frac{f(n)}{n+1} \leq \liminf \frac{f(n)}{n}; \quad \limsup \frac{f(n)}{n} \leq \sup \frac{f(n)}{n}.$$

Proof The second inequality is obvious. Fix any $d \in \mathbb{N}$. We will show that

$$\frac{f(d)}{d+1} \leq \liminf \frac{f(n)}{n}.$$

Define $f(0) := f(d+1) - f(d)$. For any $n \in \mathbb{N}$, let k and r be such integers that $n = (d+1)k + r$ and $0 \le r \le d$. Then using the assumption

$$\frac{f(n)}{n} \ge \frac{kf(d) + f(r)}{(d+1)k + r} \to \frac{f(d)}{d+1}, \quad \text{as } n \to \infty.$$

□

We will use the following lemma multiple times in the following sections (cf. Lemma 4.4.4 of [38]).

Lemma 8.4.2 *Let $\{v_1, v_2\}$ be a cut in a graph G. Suppose G_1 and G_2 are subgraphs of G with $V(G_1) \cap V(G_2) = \{v_1, v_2\}$ and $G_1 \cup G_2 = G$. Let H' be a subdivision of a 3-connected graph H, and suppose H' is not a subgraph of G_1 or G_2. Then either $G_1 \cap H'$ or $G_2 \cap H'$ is a path from v_1 to v_2.*

Proof For $i = 1, 2$ write $G'_i = G_i - \{v_1, v_2\}$. By the assumption, H' must have vertices both in G'_1 and in G'_2, and so $\{v_1, v_2\}$ is a cut in H' (both of these vertices must belong to H' since it is 2-connected).

Suppose first, that v_1 and v_2 are both on a path P' of H' which is a subdivided edge of H. Let P be the path connecting v_1 with v_2 in P'. If P has no internal vertices, then because H is 3-connected, the graph $H - \{v_1, v_2\}$ is connected, a contradiction. So P has at least one internal vertex, so that $H - \{v_1, v_2\}$ has exactly two parts, one of which is the path $P - \{v_1, v_2\}$. Since H' has vertices both in G'_1 and G'_2, one of these parts is contained in a component of G'_1 and another in a component of G'_2. The path P is a subgraph of P', so $H[V(P)] = P$, and in particular $v_1 v_2 \notin E(H)$. The claim follows.

Now suppose there is no path P in H' which is a subdivided edge of H and contains both v_1 and v_2. Using the fact that H is 3-connected we see easily that $H' - \{v_1, v_2\}$ must be connected, a contradiction. □

Let a positive integer l and a class $\mathcal{B}$ be fixed. As mentioned in the introduction, the class of $\{0,1\}^l$-coloured graphs $\mathcal{A}_l$ obtained from $\mathrm{rd}_l\,\mathcal{B}$ in Proposition 8.3.6 is decomposable. Unfortunately, this class is not bridge-addable: in the case $\mathcal{B} = \{K_4\}$, consider, for example, the coloured graph H obtained from K_3, by colouring each vertex with a distinct pair from $\{1, 2, 3\}$, see Figure 8.1. Then the graph $2H \in \mathcal{A}_3$, but no bridge can be added between the two components.

Let, as before, $\mathcal{C}^l$ be the class of connected graphs in $\mathcal{A}_l$. We call a graph $G \in \mathcal{C}^l$ *l-rootable* at a vertex $x \in V(G)$, if colouring x with $[l]$ yields a graph still in $\mathcal{C}^l$. G is l-rootable if it is l-rootable at some vertex $x \in V(G)$. For example, in the case $\mathcal{B} = \{K_4\}$, the coloured graph H from Figure 8.1 is not 3-rootable.

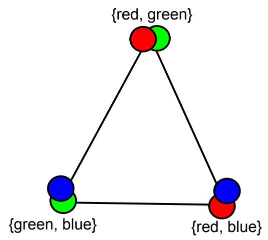

Figure 8.1: This graph is not 3-rootable with respect to $\mathcal{B} = \{K_4\}$.

Denote by $\mathcal{C}^{\bullet l}$ the class of all rooted graphs (with at least one vertex) that can be obtained by declaring an l-rootable vertex of a graph in $\mathcal{C}^l$ the root.

We will see next, that when $\mathcal{B}$ consists of 3-connected graphs, it is possible to partly restore the property of bridge-addability: if we add an edge xy connecting different components of $G \in \mathcal{A}_l$, we obtain a graph in $\mathcal{A}_l$, provided that x and y are rootable in their respective components.

Lemma 8.4.3 *Let l be a positive integer, and let $\mathcal{B}$ be as in Theorem 8.1.2. Then the class $\mathcal{U}'$ of graphs in $\mathcal{C}^l$ that are not l-rootable has $\overline{\gamma}(\mathcal{U}') \leq \overline{\gamma}(\mathcal{C}^{l-1})$.*

Proof Call $G \in \mathcal{C}^l$ *nice*, if there is $x \in V(G)$ such that $G - x$ has at least two components containing all l colours. We claim that in this case, G is rootable at x. Suppose the contrary.

Then there is a colour $i \in [l]$ such the graph G' obtained by adding a new vertex s to G connected to x and each vertex of G that has colour i, contains a $\mathcal{B}$-critical subgraph H.

Clearly $i \notin \mathrm{Col}_G(x)$ and $sx \in H$, otherwise we would have that $G \notin \mathcal{C}^l$. Suppose H shares vertices with more than one component of $G - x$. Let C be one such component. By Theorem 9.12 of [35] an expansion of a 3-connected graph at a vertex of degree at least 4 is 3-connected, and so H is a subdivision of a 3-connected graph. We may apply Lemma 8.4.2 with $G_1 = G'[V(C) \cup \{s,x\}]$, $G_2 = G' - V(C)$ and the cut $\{s,x\}$ to get that H meets either G_1 or G_2 just by a path from s to x: since $sx \in H$ this path must be sx, a contradiction.

Thus $V(H) \setminus \{s,x\}$ must be completely contained in $V(C)$ for a component C of $G - x$. But this means that $G \notin \mathcal{C}^l$: we may replace the edge sx by a path from x to s in $G' - C$, since $G - s$ has a component, disjoint from C with the colour i. We conclude that indeed G is rootable at x.

Denote by $\mathcal{U}$ the class of graphs in $\mathcal{C}^l$ that are not nice. Then $\mathcal{U}' \subseteq \mathcal{U}$, and we need to show that $\overline{\gamma}(\mathcal{U}) \leq \overline{\gamma}(\mathcal{C}^{l-1})$. Each graph $G \in \mathcal{U}$ contains a block B, such that for any cut vertex $y \in V(B)$, the components of $G - y$ disjoint from B can have at most $l - 1$ colours. (This can be seen as follows. Consider the rooted block tree T of G, and let r be its root. For any block B' let $r(B') \in V(G)$ be its parent in T, and denote by $G_{B'}$ the component of $G - r(B')$ containing $B - r(B')$. Define the set $\mathcal{S}$ of of all blocks B', such that $\mathrm{Col}(G_{B'}) = [l]$. We can assume it is non-empty, otherwise any block containing r has the required property. Pick a block $B \in \mathcal{S}$ with maximum distance from r in T. Then each component C of $G - V(B)$ contains at most $l-1$ colours: otherwise, if $C \subseteq G_B$ we would have contradiction to the choice of B, and if $C \subseteq G - G_{B'}$, then $r(B)$ would be a vertex showing that G is nice.)

By Lemma 8.3.11 for each colour $i \in [l]$ there can be at most a constant number c of vertices $x \in V(B)$ such that either x has colour i or the graph "attached" to B at x has colour i. Each graph in $\mathcal{C}^l$ with at most $l-1$ colours can be obtained from a pair (j, G_1), where $j \in [l]$ and $G_1 \in \mathcal{C}^{l-1}$ by mapping the colour j to l in G_1. Therefore, for $n \geq cl$ the coefficients $\left[\frac{x^n}{n!}\right] U(x)$ are bounded from above by

$$\left[\frac{x^n}{n!}\right] x^{cl} A^{(cl)}(x) \left(l(C^{l-1}(x))'\right)^{cl}$$

where $A(x)$ is the generating function of $\mathcal{A} = \mathrm{Ex}\,\mathcal{B}$ and $A^{(cl)}(x)$ is its cl-th derivative. The convergence radius of $U(x)$ is at least $\min(\rho(\mathcal{A}), \rho(\mathcal{C}^{l-1}))$. The class $\mathcal{C}^{l-1}$ contains all connected graphs in $\mathcal{A}$; by the exponential formula we have $\rho(\mathcal{A}) \geq \rho(\mathcal{C}^{l-1})$. So $\overline{\gamma}(\mathcal{U}) \leq \overline{\gamma}(\mathcal{C}^{l-1})$. □

We will need a simple technical lemma next.

Lemma 8.4.4 *Suppose classes of graphs $\mathcal{A}$ and $\mathcal{B}$ both have growth constants. Then $\gamma(\mathcal{A} \cup \mathcal{B})$ exists and is equal to* $\max(\gamma(\mathcal{A}), \gamma(\mathcal{B}))$.

Proof We may assume that $\gamma(\mathcal{A}) \geq \gamma(\mathcal{B})$. Clearly, $\underline{\gamma}(\mathcal{A} \cup \mathcal{B}) \geq \gamma(\mathcal{A})$. Suppose there is a subsequence $(n_k, k = 1, 2, \dots)$ such that

$$\left(\frac{|\mathcal{A}_{n_k} \cup \mathcal{B}_{n_k}|}{n_k!}\right)^{1/n_k} \to a > \gamma(\mathcal{A}).$$

If $\gamma(\mathcal{A}) > \gamma(\mathcal{B})$ then $|\mathcal{A}_n|/|\mathcal{B}_n| \to \infty$ and there is k_0 such that for $k \geq k_0$ we have $|\mathcal{A}_{n_k}| \geq |\mathcal{B}_{n_k}|$. If $\gamma(\mathcal{A}) = \gamma(\mathcal{B})$ then either $|\mathcal{A}_{n_k}| \geq |\mathcal{B}_{n_k}|$ or $|\mathcal{A}_{n_k}| \leq |\mathcal{B}_{n_k}|$ for infinitely many k. In this case, rename $\mathcal{A}$ and $\mathcal{B}$ if necessary, so that the former holds. We get that (n_k) contains a subsequence $(n'_l, l = 1, 2, \dots)$ such that $|\mathcal{A}_{n'_l} \cup \mathcal{B}_{n'_l}| \leq 2|\mathcal{A}_{n'_l}|$ and

$$\left(\frac{|\mathcal{A}_{n'_l} \cup \mathcal{B}_{n'_l}|}{n'_l!}\right)^{1/n'_l} \leq \left(\frac{2|\mathcal{A}_{n'_l}|}{n'_l!}\right)^{1/n'_l} \to \gamma(\mathcal{A}).$$

This is a contradiction. □

Lemma 8.4.5 *Let l be a positive integer and let $\mathcal{B}$ be as in Theorem 8.1.2. Suppose that $\mathcal{C}^{l-1}$ and $\mathcal{C}^{\bullet l}$ have growth constants. Then $\mathcal{C}^l$ has a growth constant*

$$\gamma(\mathcal{C}^l) = \max(\gamma(\mathcal{C}^{l-1}), \gamma(\mathcal{C}^{\bullet l})).$$

Proof Denote by $\mathcal{C}'$ the class of all l-rootable graphs in $\mathcal{C}^l$, and denote by $\mathcal{C}''$ the class of graphs $G \in \mathcal{C}^l$ that are either not l-rootable or have $\operatorname{Col}(G) \subseteq [l-1]$.

Then $\mathcal{C}'$ has a growth constant, since $\mathcal{C}^{\bullet l}$ does, and

$$|\mathcal{C}'_n| \leq |\mathcal{C}^{\bullet l}_n| \leq n|\mathcal{C}'_n|.$$

The class $\mathcal{C}''$ also has a growth constant: $\gamma(\mathcal{C}'') = \gamma(\mathcal{C}^{l-1})$. This is because by Lemma 8.4.3, $\overline{\gamma}(\mathcal{C}'') \leq \gamma(\mathcal{C}^{l-1})$, and $\mathcal{C}^{l-1} \subseteq \mathcal{C}''$.

Now, by Lemma 8.4.4, the class $\mathcal{C}^l = \mathcal{C}' \cup \mathcal{C}''$ is as claimed. □

The next lemma shows that for good enough $\mathcal{B}$, the class $\mathcal{C}^{\bullet l}$ is closed under joining smaller rooted graphs into "strings".

Lemma 8.4.6 *Let l be a positive integer and let $\mathcal{B}$ be a set of 3-connected graphs. Let G be a graph obtained from a non-empty set $\mathcal{S}$ of disjoint graphs in $\mathcal{C}^{\bullet l}$, and a path on the set $R(\mathcal{S})$ of the roots of the graphs in $\mathcal{S}$; and let G be rooted at $r \in R(\mathcal{S})$. Then $G \in \mathcal{C}^{\bullet l}$.*

Proof We first note that $\operatorname{Ex}\mathcal{B}$ contains all fans: indeed fans are series-parallel graphs, and if a fan F had a minor in $\mathcal{B}$, then since each 3-connected graph has K_4 as a minor (Lemma 3.2.1 of [38]), F would have K_4 as a minor, a contradiction.

Suppose the claim does not hold. Call $\mathcal{S}$ bad if the lemma fails for $\mathcal{S}$, a path P on $R(\mathcal{S})$ and a vertex $r \in V(P)$. Let $N(\mathcal{S})$ be the number of graphs in $\mathcal{S}$ that have size at least 2. Let ν be the size of the smallest bad set $\mathcal{S}$, and consider a set $\mathcal{S}'$ which minimizes $N(\mathcal{S})$ over bad sets $\mathcal{S}$ of size ν, let P' be a path on $R(\mathcal{S}')$ and let $r' \in V(P')$ be the root for which the lemma fails.

Then there is a colour $i \in [l]$ such that if we add a new vertex s to G, connect s to r' and every vertex of G that has colour i and remove all the colours, the resulting graph $G' \notin \operatorname{Ex}\mathcal{B}$. Let H be a $\mathcal{B}$-critical graph in G'.

Suppose all graphs in $\mathcal{S}$ are of size 1. Then G' is isomorphic to a minor of a fan, and $G' \in \operatorname{Ex}\mathcal{B}$, a contradiction.

Let G_1 be a graph in $\mathcal{S}'$ with root r_1 and size at least 2. If H has no vertices in $G_1 - r_1$, then we could replace G_1 with $G_1[\{r_1\}]$ and obtain a bad set $\mathcal{S}''$ with

$N(\mathcal{S}'') < N(\mathcal{S}')$. Thus, we may assume G_1 has at least one vertex coloured i, other than r_1. Since $G_1 \in \mathcal{C}^{\bullet l}$, $H - s$ also has a vertex not in G_1. Now $\{r_1, s\}$ is a cut in H, so by Lemma 8.4.2, either $H \cap (V(G_1) \cup \{s\})$ or $H \cap (V(G - G_1) \cup \{s, x\})$ is a path P_1 from r_1 to s. In the first case we may replace G_1 with the graph consisting of a single vertex r_1 coloured $\{i\}$ to obtain a set $\mathcal{S}''$ with $N(\mathcal{S}'') < N(\mathcal{S}')$. In the second case, we may replace P with the edge $r_1 s$ to show that $G_1 \notin \mathcal{C}^{\bullet l}$.

In each case we obtained a contradiction, so it must be that $G \in \mathcal{C}^{\bullet l}$. □

Lemma 8.4.7 *Let l be a positive integer and let $\mathcal{B}$ be as in Theorem 8.1.2. Then $\mathcal{C}^{\bullet l}$ has a growth constant.*

Proof Let n, m be positive integers. We claim that

$$|\mathcal{C}^{\bullet l}_{n+m+1}| \geq (n + m + 1)\binom{n+m}{n}|\mathcal{C}^{\bullet l}_n||\mathcal{C}^{\bullet l}_m|. \tag{8.5}$$

The above formula follows from the following construction for graphs on $n+m+1$ vertices. In the case $n \neq m$, pick a root vertex r, divide the remaining $n + m$ vertices into two parts of sizes n and m, and add a graph $G_1 \in \mathcal{C}^l$ of size n on the first part and a graph $G_2 \in \mathcal{C}^l$ of size m on the second part. Connect r to the roots r_1 and r_2 of G_1 and G_2 respectively, and declare r the root of the formed graph G. Each construction gives a unique graph, because given a graph G obtained in this way we may recover G_1 and G_2 uniquely by deleting the root of G and declaring the vertices adjacent to the two neighbours of G the roots of the respective components.

In the case $n = m$, we have to avoid obtaining each graph twice because of symmetry. So if $V(G_1)$ is lexicographically smaller than $V(G_2)$ output the graph G as above, otherwise output the graph $G - rr_1 + r_1r_2$. To finish the proof of (8.5), note the constructions G are always in the class $\mathcal{C}^{\bullet l}$ by Lemma 8.4.6.

Now let d, k be positive integers, $k \geq 2$ and set $n = kd$. Form a graph G by taking an arbitrary set $\mathcal{S}$ of k disjoint graphs in $\mathcal{C}^{\bullet l}$ of size d, adding a path, P rooted at one of the endpoints r and with $V(P)$ consisting of all roots of the graphs in $\mathcal{S}$. Declare r the root of G. By Lemma 8.4.6, $G \in \mathcal{C}^{\bullet l}$.

Note also, that we never construct a graph G twice: it is always possible to recover the path P and the set $\mathcal{S}$ uniquely from G. (Start with the root r of G. There can be only one edge $rx \in G$, such that $G - rx$ has a component C of size d: rx is the first edge of P. Delete rx and proceed in the same way with the component of $G - rx$ containing x, rooted at x.) Let $\mathcal{P}$ be the set of all graphs constructed in this way. Then

$$|\mathcal{P}_n| = \frac{(dk)!}{d!^k}|\mathcal{C}_d^{\bullet l}|^k$$

Next we observe, that since Ex $\mathcal{B}$ contains all apex paths, the class $\mathcal{R}$ of rooted uncoloured cycles is contained in $\mathcal{C}^{\bullet l}$. This class is clearly disjoint from $\mathcal{P}$, in which every graph has a bridge.

For $t = 1, 2, \ldots$ let $f(t) = \ln(|\mathcal{C}_t^{\bullet l}|/t!)$. We have

$$\frac{f(n)}{n} \geq \frac{1}{n} \ln\left(\frac{|\mathcal{P}_n| + |\mathcal{R}_n|}{n!}\right) > \frac{f(d)}{d}. \tag{8.6}$$

From (8.5) it follows that

$$f(n+m+1) \geq f(n) + f(m).$$

Thus by the modification of Fekete's lemma, Lemma 8.4.1

$$\sup \frac{f(n)}{n+1} \leq \liminf \frac{f(n)}{n}; \quad \limsup \frac{f(n)}{n} \leq \sup \frac{f(n)}{n}.$$

By (8.6), for $n = kd$ and any integer $k = 2, 3, \ldots$

$$\frac{f(n)}{n+1} - \frac{f(d)}{d+1} > \frac{d}{d+1}\left(\frac{f(n)}{n} - \frac{f(d)}{d}\right) > 0.$$

Therefore

$$\sup \frac{f(n)}{n+1} = \limsup \frac{f(n)}{n+1} = \limsup \frac{f(n)}{n},$$

and

$$\frac{f(n)}{n} \to \limsup \frac{f(n)}{n} \in [0; \infty].$$

Since Ex $\mathcal{B}$ is small by [43], we conclude that $\mathcal{C}^{\bullet l}$ has a growth constant.

Also, because Ex $\mathcal{B}$ includes all graphs without a 3-connected minor, by Lemma 3.2.1 of [38], $\gamma(\mathcal{C}^{\bullet l}) \in [\gamma(\text{Ex}\, K_4), \infty)$, where $\gamma(\text{Ex}\, K_4) = 9.073..$, see Section 9.2. □

Lemma 8.4.8 *Let l be a non-negative integer and let $\mathcal{B}$ be as in Theorem 8.1.2. Then the class $\mathcal{C}^l$ has a growth constant.*

Proof We use induction on l. The class $\mathcal{C}^0$ is the class of connected graphs in Ex $\mathcal{B}$, this class has a growth constant by [74, 77]. Suppose now that $l > 0$ and assume that we have proved the claim for each class Ex $(l'+1)\mathcal{B}$, where $l' < l$, we now prove it for $l' = l$.

The class $\mathcal{C}^{\bullet l}$ has a growth constant by Lemma 8.4.7. The class $\mathcal{C}^{(l-1)}$ has a growth constant by induction. So $\mathcal{C}^{l}$ has a growth constant by Lemma 8.4.5. □

We can now combine the lemmas of this section to finish the proof of Theorem 8.1.2.

Proof of Theorem 8.1.2 The class $\mathcal{C}^{2k+1}$ has a growth constant γ by Lemma 8.4.8. Since

$$[x^n]C^{2k+1}(x) \leq [x^n]A_{2k+1}(x) \leq [x^n]e^{C^{2k+1}(x)},$$

we get, see i.e. [48], that $\mathcal{A}_{2k+1}$ also has growth constant γ. Using Proposition 8.3.6, we see that $\mathrm{rd}_{2k+1}\mathcal{B}$ has growth constant γ as well. By the assumption of the theorem, there must be a constant c, such that $\mathcal{B}$ does not contain a wheel W_{c+1} as a minor (which is a planar graph). Now Theorem 8.1.1 completes the proof. □

8.4.2 Small blockers and small redundant blockers

In this section we collect several auxiliary lemmas. For $\mathcal{B}$ and k_0 as in Theorem 8.1.1, we can often conclude that $R_n \in \mathrm{Ex}\,(k+1)\mathcal{B}$ either has a blocker of size k (if $k < k_0$) with probability $1 - e^{-\Omega(n)}$ or (if $k \geq k_0$) it has a constant size $(2k, 2, \mathcal{B})$-double blocker with probability $1 - e^{-\Omega(n)}$. Using results of this section, it can be shown that this happens, for example, when $\gamma(\mathrm{rd}_{2k+1}\mathcal{B}) \neq 2^k\gamma(\mathrm{Ex}\,\mathcal{B})$ exists for all k.

Lemma 8.4.9 *Let k be a positive integer and let $\mathcal{B}$ be the set of minimal excluded minors for a proper addable minor-closed class of graphs. Suppose* $\mathrm{aw}_2(\mathrm{Ex}\,\mathcal{B})$ *is finite.*

If $\underline{\gamma}(\mathrm{Ex}\,(k+1)\mathcal{B}) > 2\overline{\gamma}(\mathrm{Ex}\,k\mathcal{B})$ *then there is a constant $r = r(k, \mathcal{B})$ such that all but at most $e^{-\Omega(n)}$ fraction of graphs in* $(\mathrm{Ex}\,(k+1)\mathcal{B})_n$ *have a* $(2k, 2, \mathcal{B})$*-double blocker of size r.*

Proof By Lemma 8.3.3 and Lemma 8.3.5 there is a constant $r = r(k, \mathcal{B}) > 2k$ such that every graph in $G \in \mathrm{Ex}\,(k+1)\mathcal{B}$ is a union of two graphs G_1 and G_2, where G_1 has a $(2k, 2, \mathcal{B})$-double blocker Q of size at most r with a special set S, $S \subseteq V(G_1) \cap V(G_2) \subseteq Q$, $G_2 \in \mathrm{apex}\,(\mathrm{Ex}\,k\mathcal{B})$ and Q is a $\mathcal{B}$-blocker for G.

We may assign each $G \in \mathrm{Ex}\,(k+1)\mathcal{B}$ a unique tuple $t(G) = (G_1, G_2, Q, S)$ as above. Call G *complex*, if $G_2 - (Q \setminus S)$ contains a subgraph $H \notin \mathrm{Ex}\,\mathcal{B}$ which has only one vertex $z \in S$. Observe, that if G is not complex, then Q is a $(2k, 2, \mathcal{B})$-double blocker for G, and S is its special set.

Suppose G is complex, and let H be a subgraph of $G_2 - (Q \setminus S)$ such that $V(H) \cap S = \{z\}$ for $z \in S$. Then $G_1 - z$ is disjoint from H, so $G_1 - z \in \mathrm{Ex}\, k\mathcal{B}$ and $G_1 \in \mathrm{apex}\,(\mathrm{Ex}\, k\mathcal{B})$. In this case, if G has at least r vertices, it can be obtained from a graph $\tilde{G}_1$ in $\mathrm{apex}\,(\mathrm{Ex}\, k\mathcal{B})$ which has $s = |V(G_1) \cap V(G_2)|$ distinguished vertices (roots) and another graph $\tilde{G}_2$ in $\mathrm{apex}\,(\mathrm{Ex}\, k\mathcal{B})$ which has s pointed vertices, by identifying the i-th rooted vertex with the i-th pointed vertex and merging edges between the distinguished vertices.

Thus the n-th coefficient, of the exponential generating function for the complex graphs is bounded by

$$[x^n] \sum_{s=0}^{r} x^s \left(A^{(s)}(x)\right)^2,$$

where A is the exponential generating function of $\mathrm{apex}\,(\mathrm{Ex}\, k\mathcal{B})$ and $A^{(s)}$ is the s-th derivative of A. This shows that the inverse radius of convergence for the class of complex graphs is at most

$$\overline{\gamma}(\mathrm{apex}\,(\mathrm{Ex}\, k\mathcal{B})) \leq 2\overline{\gamma}(\mathrm{Ex}\, k\mathcal{B}) < \underline{\gamma}(\mathrm{Ex}\,(k+1)\mathcal{B}),$$

see Proposition 8.3.6 and the proof of Theorem 8.1.1. Hence, all but at most $e^{-\Omega(n)}$ fraction of graphs in $|(\mathrm{Ex}\,(k+1)\mathcal{B})_n|$ are not complex, and therefore have a $(2k, 2, \mathcal{B})$-double blocker. □

We call a connected subgraph H of G a *pendant subgraph*, if there is exactly one edge in G between $V(H)$ and $V(G) \setminus V(H)$.

Lemma 8.4.10 *Let $\mathcal{A}$ be a proper addable minor-closed class of graphs. Let $H \in \mathcal{A}$ be a connected graph and let k be a non-negative integer.*

There is a constant $c > 0$, such that the random graph $R_n \in_u \mathrm{apex}^k\,(\mathcal{A})$ with probability $1 - e^{-\Omega(n)}$ has a set S of k vertices, such that $R_n - S$ contains a family $\mathcal{H}$ of at least cn pairwise disjoint pendant subgraphs isomorphic to H, and each vertex of S is incident to all vertices of every graph $\tilde{H} \in \mathcal{H}$.

Proof This fact is proved in the proof of Theorem 7.1.2. □

Lemma 8.4.11 *Let k be a positive integer and let $\mathcal{B}$ be the set of minimal excluded minors for a proper addable minor-closed class of graphs. Suppose* $\mathrm{Ex}\,\mathcal{B}$ *contains all fans,* $\mathrm{aw}_2\,(\mathrm{Ex}\,\mathcal{B})$ *is finite,*

$$|(\mathrm{Ex}\, k\mathcal{B})_n| \leq |(apex^{k-1}\,(\mathrm{Ex}\,\mathcal{B}))_n| \left(1 + e^{-\Omega(n)}\right)$$

and $\gamma_2 > \gamma_1$*, where* $\gamma_2 = \overline{\gamma}(apex\,(\mathrm{Ex}\,k\mathcal{B}))$*,* $\gamma_1 = \overline{\gamma}(rd_{2k+1}\,\mathcal{B})$*. Then*

$$|(\mathrm{Ex}\,(k+1)\mathcal{B})_n| = |(apex^k\,(\mathrm{Ex}\,\mathcal{B}))_n|\left(1 + e^{-\Theta(n)}\right).$$

Proof By Lemma 8.3.3 and Lemma 8.3.5 there is a constant $r = r(k, \mathcal{B}) > 2k$ such that every graph in $G \in \mathrm{Ex}\,(k+1)\mathcal{B}$ with at least r vertices can be generated as follows.

1) Pick $n_2 \in \{0, \dots, n\}$.

2) Pick a set $V_2 \subseteq [n]$ of size n_2.

3) Pick $q \in \{0, \dots, n_2 \wedge r\}$.

4) Pick a set $Q \subseteq V_2$ of size q.

5) Put any graph $G_2 \in \mathcal{A}$ on V_2. Here $\mathcal{A} = \mathrm{apex}\,(\mathrm{Ex}\,k\mathcal{B})$.

6) Add edges of any graph $G_1 \in \mathcal{D}$ on $V_1 = ([n] \setminus V_2) \cup Q$ (merge repetitive edges, if necessary). Here $\mathcal{D}$ is the class of graphs with a $(2k, 2, \mathcal{B})$-double blocker of size at most r.

Let u_n be the total number of constructions, i.e., the total number of different tuples $(n_2, q, V_2, Q, G_1, G_2)$ that can be generated by the above procedure. Denote by $\mathcal{U}$ be the combinatorial class with the counting sequence $(u_n, n = 0, 1, \dots)$. Also, let R_n be the graph obtained by taking the tuple $(n_2, q, V_2, Q, G_1, G_2)$ uniformly at random from all u_n possible tuples. (In the rest of the proof n_2, q, V_2, Q, G_1 and G_2 will be random variables.)

By Lemma 8.3.7, $\overline{\gamma}(\mathcal{D}) = \gamma_1$. Similarly as in the proof of (8.3)

$$u_n \le [x^n/n!] \sum_{q=0}^{r} x^q A^{(q)}(x) D^{(q)}(x),$$

so, see [48], $\overline{\gamma}(\mathcal{U}) \le \gamma_2$. Fix $\epsilon \in (0, 0.5)$ and $\delta > 0$ such that

$$(\gamma_1 + \delta)^{\epsilon}(\gamma_2 + \delta)^{1-\epsilon} < \gamma_2.$$

Let H be a graph of minimal size that can be obtained by removing one vertex from a graph in $\mathcal{B}$. Let c be a constant as in Lemma 8.4.10 applied with Ex $\mathcal{B}$ and H. We say that a set $S \subseteq V_2$ is *good* if $|S| = k$ and there is a family $\mathcal{H}_S$ of at least $cn/4$ disjoint pendant subgraphs $\tilde{H}$ in $G_2 - (S \cup Q)$ such that $\tilde{H}$ is isomorphic to H and every vertex $v \in S$ is incident to every vertex of $V(\tilde{H})$.

Define the following events:

$$A = \{G_2 \text{ has at least } (1-\epsilon)n \text{ vertices}\};$$
$$B = \{G_2 \in \mathrm{apex}^k\,(\mathrm{Ex}\,\mathcal{B})\};$$
$$C = \{G_2 \text{ has a good set } S\}.$$

We will show that

$$\mathbb{P}(\bar{A}) \leq e^{-\Omega(n)}; \quad \mathbb{P}(\bar{B}) \leq e^{-\Omega(n)}; \quad \mathbb{P}(\bar{C}) \leq e^{-\Omega(n)}. \tag{8.7}$$

and

$$\gamma(\mathcal{U}) = \gamma(\mathrm{apex}^k\,(\mathrm{Ex}\,\mathcal{B})) = \gamma_2. \tag{8.8}$$

For n large enough, A, B and C imply that either $R_n \in \mathrm{apex}^k\,(\mathrm{Ex}\,\mathcal{B})$ or R_n has $k+1$ disjoint subgraphs not in $\mathrm{Ex}\,\mathcal{B}$. Indeed, by Lemma 5.3 of [65], if S is a good set, then for a $\mathcal{B}$-critical subgraph H_1 of $G-S$, there is at most a constant number $N_{k,\mathcal{B}}$ of subgraphs $\tilde{H} \in \mathcal{H}_S$, which are not disjoint from H_1. For n large enough, $k < cn/4 - N_{k,\mathcal{B}}$, so we can construct k disjoint subgraphs not in $\mathrm{Ex}\,\mathcal{B}$, each containing one vertex from S, and each disjoint from H_1, producing $k+1$ disjoint forbidden subgraphs in total.

Denote $a_n' := |\mathrm{apex}^k\,(\mathrm{Ex}\,\mathcal{B})_n|$. Assuming (8.7) and (8.8) hold, we have for n large enough

$$\begin{aligned}&|(\mathrm{Ex}\,(k+1)\mathcal{B})_n \setminus (\mathrm{apex}^k\,(\mathrm{Ex}\,\mathcal{B}))_n| \leq u_n(\mathbb{P}(\bar{A}) + \mathbb{P}(\bar{B}) + \mathbb{P}(\bar{C}))\\ &= n!\gamma_2^{n-\Omega(n)} = e^{-\Omega(n)}a_n'.\end{aligned} \tag{8.9}$$

Let us show (8.7) and (8.8). Recall that by Theorem 7.1.2, $\gamma(\mathrm{apex}^l(\mathrm{Ex}\,\mathcal{B})) = 2^l\gamma(\mathrm{Ex}\,\mathcal{B})$ for any $l = 0, 1, 2, \ldots$. By the definition of apex classes and the assumption of the lemma

$$\begin{aligned}&|\mathrm{apex}\,(\mathrm{Ex}\,k\mathcal{B})_n \setminus (\mathrm{apex}^k\,(\mathrm{Ex}\,\mathcal{B}))_n| \leq n2^{n-1}|(\mathrm{Ex}\,k\mathcal{B})_{n-1} \setminus (\mathrm{apex}^{k-1}(\mathrm{Ex}\,\mathcal{B}))_{n-1}|\\ &\leq n2^{n-1}e^{-\Omega(n)}|(\mathrm{apex}^{k-1}\,(\mathrm{Ex}\,\mathcal{B}))_{n-1}|\end{aligned}$$

The last line is $e^{-\Omega(n)}a_n'$ again by Theorem 7.1.2. So

$$a_n := |\mathrm{apex}\,(\mathrm{Ex}\,k\mathcal{B})_n| \leq a_n'(1 + e^{-\Omega(n)}) \tag{8.10}$$

and $\gamma_2 = 2^k\gamma(\mathrm{Ex}\,\mathcal{B})$ is the growth constant of $\mathrm{apex}\,(\mathrm{Ex}\,k\mathcal{B})$. Since $u_n \geq a'_n$, we have $\underline{\gamma}(\mathcal{U}) \geq \gamma_2$ and (8.8) follows.

Let $d_n = |\mathcal{D}_n|$. There is a constant C, such that for any $n = 1, 2, \ldots$

$$a_n \leq Cn!(\gamma_2 + \delta)^n; \quad d_n \leq Cn!(\gamma_1 + \delta)^n.$$

Using (8.8), there is a constant $C' > 0$, such that the number of constructions with $n_2 < (1-\epsilon)n$ is at most

$$\begin{aligned}
&\sum_{n_2=0}^{\lfloor(1-\epsilon)n\rfloor} \sum_{q=0}^{r\wedge n_2} \binom{n}{n_2}\binom{n_2}{q} a_{n_2} d_{n-n_2+q} \\
&\leq C' n^{2q+1} n! \max_{n_2<(1-\epsilon)n} (\gamma_1+\delta)^{n-n_2}(\gamma_2+\delta)^{n_2} \\
&\leq C' n^{2q+1} n! \left((\gamma_2+\delta)^{1-\epsilon}(\gamma_1+\delta)^{\epsilon}\right)^n \\
&\leq e^{-\Omega(n)} u_n,
\end{aligned}$$

and the first bound of (8.7) follows.

The second bound of (8.7) follows by the first one and (8.10) since

$$\mathbb{P}(\bar{B}) \leq \mathbb{P}(\bar{A}) + \mathbb{P}(\bar{B}|A) = e^{-\Omega(n)}.$$

Fix an integer t, $(1-\epsilon)n \leq t \leq n$, and a subset $V_2 = \tilde{V}$ of size t. Conditionally on $V_2 = \tilde{V}$ and the event B, the random graph G_2 is a uniformly random graph on $\tilde{V}$ from $\mathrm{apex}^k\,(\mathrm{Ex}\,\mathcal{B})$.

By Lemma 8.4.10 there is a constant $c_1 > 0$ such that for all large enough n, conditionally on $V_2 = \tilde{V}$ and B, the graph G_2 with probability at least $1-e^{-c_1(1-\epsilon)n}$ has a set S, where every vertex in S is incident to every vertex of at least $c(1-\epsilon)n \geq cn/2$ disjoint pendant subgraphs of G_2 isomorphic to H. At most q such subgraphs can have vertices in Q, so if n is large enough then $cn/2 - q > cn/4$ and S is good.

Now

$$\mathbb{P}(\bar{C}) \leq \mathbb{P}(\bar{A}) + \mathbb{P}(\bar{B}) + \mathbb{P}(\bar{C}|A, B).$$

For large enough n, by the above argument and symmetry the last term on the right side is

$$\begin{aligned}
&\frac{1}{\mathbb{P}(A,B)} \sum_{\tilde{V}\subseteq[n], |\tilde{V}|\geq(1-\epsilon)n} \mathbb{P}(\bar{C}|V_2=\tilde{V}, B)\mathbb{P}(V_2=\tilde{V}, B) \\
&\leq e^{-c_1(1-\epsilon)n}\frac{\mathbb{P}(A,B)}{\mathbb{P}(A,B)} = e^{-\Omega(n)},
\end{aligned}$$

and the last bound of (8.7) follows.

Finally, the fact that $|(\mathrm{Ex}\,(k+1)\mathcal{B})_n \setminus (\mathrm{apex}^k\,(\mathrm{Ex}\,\mathcal{B}))_n| \geq e^{-\Theta(n)} a'_n$ follows by Lemma 7.5.5 from the previous chapter.

□

For $\mathcal{B}$ as in Theorem 8.1.2 and sufficiently large k we have, using Theorem 8.1.1, that $R_n \in_u \mathrm{rd}_{2k+1}\,\mathcal{B}$ belongs to $\mathrm{apex}^{2k-1}\,\mathcal{B} \supseteq \mathrm{apex}^k\,\mathcal{B}$ with probability $e^{-\Omega(n)}$. The next lemma shows that the two candidates for the main subclass of $\mathrm{Ex}\,(k+1)\mathcal{B}$ studied so far essentially do not overlap.

Lemma 8.4.12 *Let $\mathcal{A}$ be a proper addable minor-closed class. Let $\mathcal{B}$ be its set of minimal excluded minors. There is a constant $c > 0$ such that with probability $1 - e^{-\Omega(n)}$ every redundant blocker $R_n \in_u \mathrm{apex}^k\,(\mathrm{Ex}\,\mathcal{B})$ is of size at least cn.*

Proof Fix a graph $H \in \mathcal{B}$. Let $H_0 = H - v$, where $v \in V(H)$ is any vertex. Fix $\epsilon \in (0,1)$, and let $A_\epsilon = A_\epsilon(n)$ be the event that the random graph R_n has a unique blocker S of size k and the graph $R_n - S$ has at least ϵn pendant appearances $\tilde{H}$ of the graph H_0, such that every vertex of $\tilde{H}$ is connected to every vertex of S (call such pendant appearances *good*), and any $\mathcal{B}$-blocker not containing S has at least ϵn vertices. By Lemma 8.4.10 and Theorem 7.1.3 we can choose ϵ so, that A_ϵ occurs with probability $1 - e^{-\Omega(n)}$.

Let $c = \epsilon/2$. Let n be sufficiently large, so that $(\epsilon - c)n > 2$. Suppose A_ϵ occurs and R_n has a redundant blocker Q of size at most cn. Then Q must contain S and there must be at least one good appearance $\tilde{H}$ disjoint from Q. Now any vertex $x \in S$ together with $\tilde{H}$ induces a graph containing H in $R_n - (Q \setminus \{x\})$, thus Q is not a redundant blocker. So the probability that R_n has a redundant blocker of size at most cn is at most $\mathbb{P}(\bar{A}_\epsilon) = e^{-\Omega(n)}$. □

Chapter 9

Few disjoint minors K_4

9.1 Introduction

Recall that series-parallel graphs are exactly the class $\mathrm{Ex}\, K_4$. Asymptotic counting formulas and other properties of series-parallel and outerplanar graphs were obtained by Bodirsky, Giménez, Kang and Noy [32]; the degree distribution was studied by Bernasconi, Panagiotou and Steger [12] and by Drmota, Giménez and Noy [41].

The main result of this chapter concerns the number of graphs, not containing a minor isomorphic to $k+1$ disjoint copies of K_4 (i.e., $\mathcal{B} = \{K_4\}$). At the expense of delving deeply into the structure of series parallel graphs (coloured according to certain rules), and analysing specific generating functions, we obtain much sharper conclusions than in the previous chapter, as well as explicit numerical approximations for constants.

Theorem 9.1.1 *Let k be a positive integer. We have*

$$|(\mathrm{Ex}\,(k+1)K_4)_n| = (1 + e^{-\Theta(n)})|(rd_{2k+1}\, K_4)_n|.$$

There are constants $c_k > 0$ and $\gamma_k > 0$, such that

$$|(rd_{2k+1}\, K_4)_n| = c_k n^{-5/2} n! \gamma_k^n (1 + o(1)).$$

Furthermore, $\gamma_1 = 23.5241...$

The proof of the above theorem yields the following facts about the structure of typical graphs without a minor isomorphic to $(k+1)K_4$. Recall that given a class of graphs $\mathcal{A}$, we write $R_n \in_u \mathcal{A}$ to mean that R_n is a uniformly random graph drawn from $\mathcal{A}_n$.

Theorem 9.1.2 *Let k be a positive integer and let $R_n \in_u \mathrm{Ex}\,(k+1)K_4$.*

(a) There is a constant $a_k > 0$, such that with probability $1 - e^{-\Theta(n)}$, the graph R_n has a unique redundant K_4-blocker Q of size $2k+1$, each vertex in Q has degree at least $a_k n$, and any K_4-blocker Q' with $|Q \setminus Q'| > 1$, has at least $a_k n$ vertices.

(b) The probability that R_n is connected converges to $p_k = A(\gamma_k^{-1})$ where A is the exponential generating function of $\mathrm{Ex}\, K_4$ and γ_k is as in Theorem 9.1.1.

Let us point out that the complete asymptotic distribution of the 'fragment' graph of R_n (R_n minus its largest component) and the asymptotic Poisson distribution of the number of components in R_n can be easily obtained (in terms of A and γ_k) using results from [74]. Furthermore, the expected number of the vertices not in the largest component of R_n is $O(1)$, this holds more generally for random graphs from any bridge-addable class [74]. We provide an approach to evaluate γ_k, $k = 1, 2, \ldots$ numerically to arbitrary precision. Then p_k can be numerically evaluated using results of [32].

The proof of Theorem 9.1.1 is much more complicated than the proof in the case $\mathcal{B} = \{K_3\}$ in Chapter 6 or the more general Theorem 7.1.1. When $\mathrm{Ex}\,\mathcal{B}$ does not contain all fans, a random graph from $(\mathrm{Ex}\,(k+1)\mathcal{B})_n$ essentially consists of a random graph in $\mathrm{Ex}\,\mathcal{B}$ on $n-k$ vertices and k apex vertices with their neighbours chosen independently at random, each with probability 1/2 (see Chapter 7). Meanwhile, if Q is a redundant blocker for $G \in \mathrm{Ex}\,(k+1)K_4$, then the possible neighbours of a vertex $v \in Q$ depend on the series-parallel graph $G - Q$. To solve this, we obtain decompositions of the dominating subclass of $\mathrm{rd}_{2k+1}\,\mathcal{B}$ into tree-like structures and analyse the corresponding generating functions.

In our last result we look at classes $\mathrm{Ex}\,(k+1)\{K_{2,3}, K_4\}$. For $k = 0$, this corresponds to outerplanar graphs.

Theorem 9.1.3 *Let $\mathcal{B} = \{K_{2,3}, K_4\}$. The class $\mathrm{Ex}\,(k+1)\mathcal{B}$ has a growth constant γ'_k for each $k = 1, 2, \ldots$. We have*

$$\gamma'_1 = \gamma(\mathit{apex}\,(\mathrm{Ex}\,\mathcal{B})) = 2\gamma(\mathrm{Ex}\,\mathcal{B}) > \gamma(\mathit{rd}_3\,\mathcal{B})$$

and for a positive constant c

$$|(\mathrm{Ex}\,2\mathcal{B})_n| = cn^{-3/2}\gamma_1'^n n!(1 + o(1)).$$

However, for any $k \geq 2$

$$\gamma'_k = \gamma(rd_{2k+1}\,\mathcal{B}) > \gamma(apex^k\,(\mathrm{Ex}\,\mathcal{B})) \quad and \quad |(\mathrm{Ex}\,(k+1)\mathcal{B})_n| = e^{\Omega(n^{1/2})}\gamma_k'^n n!$$

The first few values are $\gamma'_1 = 14.642..$, $\gamma'_2 = 34.099..$, $\gamma'_3 = 130.023..$, and for $k \geq 2$ γ'_k admits a closed-form expression.

The last theorem shows that Theorem 8.1.1 does not hold in general with $k_0 = 1$. The unusual subexponential factor for $k \geq 2$ shows up because the underlying structure of typical graphs in $\mathrm{rd}_{2k+1}\,\{K_{2,3}, K_4\}$ is "path-like", whereas it is "tree-like" for graphs in $\mathrm{rd}_{2k+1}\,K_4$, see Figure 6 (c).

The structure of this Chapter is as follows. In Section 9.2, we explore the rich structure of the classes $\mathrm{rd}_{2k+1}\,K_4$, which we then translate into generating functions and apply analytic combinatorics to get the growth constant when $k = 1$. In Section 9.3 we count graphs obtained from unrooted Cayley trees, where edges, internal vertices and leaves are replaced with graphs from different classes. Then, in Section 9.4 we complete the proof of Theorem 9.1.1. In Section 9.5 we prove Theorem 9.1.3. Finally, in Section 9.6 we discuss open questions that arise from this work and give some concluding remarks.

9.2 Analytic combinatorics for $\mathrm{Ex}\,2K_4$

In this section we focus on the case $\mathcal{B} = \{K_4\}$ and the class $\mathcal{C}^3$. Recall that $\mathcal{A}_l$ denotes the set of $\{0,1\}^l$-coloured graphs G such that if $G \in \mathcal{A}_{l,n}$ then $\{n+1, \ldots, n+l\}$ is an l-redundant blocker for $G^{\{n+1,\ldots,n+l\}}$ and $\mathcal{C}^l$ is the class of connected graphs in $\mathcal{A}_l$. The main result of this section is the following.

Lemma 9.2.1 *Let $\mathcal{B} = \{K_4\}$. The class $\mathcal{C}^3$ has growth constant $\gamma(\mathcal{C}^3) = 23.5241..$.*

This shows that in Theorem 8.1.1 and Theorem 8.1.2 we have $k_0 = 1$ for $\mathcal{B} = \{K_4\}$:

Corollary 9.2.2 *Let $\mathcal{B} = \{K_4\}$. For any $k = 1, 2, \ldots$*

$$\gamma(\mathrm{Ex}\,(k+1)K_4) = \gamma(rd_{2k+1}\,K_4) = \gamma(\mathcal{A}_{2k+1}).$$

Proof Bodirsky, Giménez, Kang and Noy [32] showed that $\gamma(\mathrm{Ex}\,K_4) = 9.073\ldots$. By the exponential formula

$$C^3(x) \leq A_3(x) \leq e^{C^3(x)},$$

so by Proposition 8.3.6, Lemma 9.2.1 and Lemma 9.2.1

$$\gamma(\mathrm{rd}_3 K_4) = \gamma(\mathcal{A}_3) = \gamma(\mathcal{C}^3) > 2\gamma(\mathrm{Ex}\, K_4).$$

Now the claim follows by Lemma 8.3.9. □

9.2.1 Series-parallel networks

Recall that a graph G is called series-parallel if $G \in \mathrm{Ex}\, K_4$. A series-parallel graph G with an ordered pair of distinguished vertices s and t is called an *SP-network* if G is connected and adding an edge st to G, the resulting multigraph is 2-connected and series-parallel (so a network isomorphic to K_2 is also an SP-network). s and t are called the *poles* of G; s is the *source* of G and t is the *sink* of G. The poles have no label and do not contribute to the size of G. A vertex $v \in V(G)$ that is not a pole, is called an *internal vertex*.

Denote by $\mathcal{D}$ the class of all SP-networks and by $\mathcal{E}_2$ the class of SP-networks consisting of a single edge between the source and the sink. Also, denote by $\mathcal{E}_1$ the class of degenerate networks with source and sink represented by the same vertex and zero internal vertices. The corresponding exponential generating functions are $E_2(x) = E_1(x) = 1$.

Lemma 9.2.3 (Trakhtenbrot 1958, [99], see also [100]) *We have*

$$\mathcal{D} = \mathcal{E}_2 + \mathcal{S} + \mathcal{P}.$$

Here $\mathcal{S}$ and $\mathcal{P}$ are defined by $|\mathcal{S}_0| = |\mathcal{P}_0| = 0$ and

$$\mathcal{S} = (\mathcal{P} + \mathcal{E}_2) \times \mathrm{SEQ}_{\geq 1}(\mathcal{Z} \times (\mathcal{P} + \mathcal{E}_2));$$
$$\mathcal{P} = \mathcal{E}_2 \times \mathrm{SET}_{\geq 1}(\mathcal{S}) + \mathrm{SET}_{\geq 2}(\mathcal{S}).$$

Furthermore, the classes $\mathcal{S}$ and $\mathcal{P}$ correspond to disjoint classes of networks and the above relation corresponds to a unique decomposition of a graph $G \in \mathcal{S}$ (respectively, $G \in \mathcal{P}$) into subgraphs in $\mathcal{P} \cup \mathcal{E}_2$ (respectively, $\mathcal{S} \cup \mathcal{E}_2$) with pairwise disjoint sets of labels.

The last statement of the lemma asserts that there is a stronger kind of isomorphism than just combinatorial one. More precisely, the classes $\mathcal{S}$, $\mathcal{P}$ and $\mathcal{E}_2$ can (and will) be considered as classes of graphs, which naturally partition the class of all SP-networks, see Figure 9.1. A network $G \in \mathcal{S}$ is called a *series SP-network*. G can be decomposed uniquely into $k \geq 2$ networks $H_1, \ldots, H_k \in \mathcal{P} + \mathcal{E}_2$, where

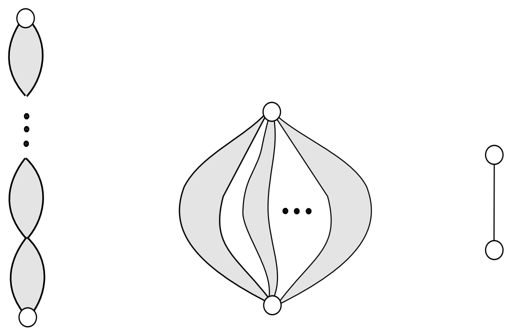

Figure 9.1: The structure of the class $\mathcal{D}$ provided by Lemma 9.2.3.

the sink of H_i is the source of H_{i+1} for $i = 1, \ldots, k-1$, the source of G is the source of H_1, the sink of G is the sink of H_k, and the sets of internal vertices are disjoint for H_i, H_j, $i \neq j$. We say that G is obtained from $H_1, \ldots, H_k$ by *series composition.*

A network $G \in \mathcal{P}$ with source s and sink t is called a *parallel SP-network.* G can be decomposed uniquely into $k \geq 2$ networks $S_1, \ldots, S_k \in \mathcal{S} \cup \mathcal{E}_2$, where at most one network is in $\mathcal{E}_2$. In such a decomposition, the sets of internal vertices of $S_1, \ldots, S_k$ are pairwise disjoint, the source of $S_1, \ldots, S_k$ is s, and the sink of $S_1, \ldots, S_k$ is t. We say that G is obtained from $S_1, \ldots, S_k$ by *parallel composition.* The above decomposition also implies that for any internal vertex v of $G \in \mathcal{P}$, we may represent G as a parallel composition of a network $S \in \mathcal{S}$ (where $S = S_j$ with $v \in V(S_j)$) and a network $D \in \mathcal{D}$ (where $D = \cup_{i \neq j} S_i$).

It has been shown, see [32] and [12], that the exponential generating functions of $\mathcal{D}$ and $\mathcal{P}$ satisfy

$$\frac{xD(x)^2}{1 + xD(x)} = \ln\left(\frac{1 + D(x)}{2}\right); \tag{9.1}$$

$$P(x) + 1 = \frac{D(x)}{1 + xD(x)}. \tag{9.2}$$

To keep formulas shorter, for exponential generating functions $A(x)$ we will often skip "(x)"; $x \in \mathbb{C}$ will usually be fixed, and its value should be clear from the context. Identities where the range of x is not explicitly stated, will hold for some $\delta > 0$ and any $x \in \mathbb{C}$ with $|x| < \delta$.

The following simple facts were used already in [99].

Proposition 9.2.4 *Let G be an SP-network with poles s and t. Then for each internal vertex v of G there is a path from s to t containing v.*

Proof If G is a parallel SP-network, then since G is 2-connected, there are internally disjoint paths, a path from v to s and a path from v to t. Connecting them, we obtain a path from s to t.

Suppose G is a series network. Let $H_1, \ldots, H_k$ be the decomposition of G into graphs in $\mathcal{P} + \mathcal{E}_2$ as in Lemma 9.2.3. Then for $i = 1, \ldots, k$, the network H_i contains a path P_i connecting its poles, and, if v is an internal vertex of H_i, also containing v. Connecting each of the paths P_i yields a path from s to t that contains v. □

Proposition 9.2.5 *Let G be a parallel SP-network with source s and sink t. Then for each internal vertex v of G there are two internally disjoint paths from s to t such that one of the paths contains v.*

Proof By Lemma 9.2.3 the graph G can be obtained by a parallel composition of two networks $S \in \mathcal{S}$ and $D \in \mathcal{D}$ with disjoint sets of vertices where v is an internal vertex of $\mathcal{S}$. By Proposition 9.2.4, there is a path P from s to t that contains v. Now D contains another path from s to t internally disjoint from P. □

Proposition 9.2.6 *Let $G \in \mathcal{D}$. The network G' obtained by adding a new vertex w connected to both poles of G satisfies $G' \in \mathcal{P}$.*

Proof This is an immediate consequence of Lemma 9.2.3, see the comment after it. □

9.2.2 Rooted graphs of multiple types

Let $F(x)$ and $B(x)$ denote the exponential generating functions of rooted connected series-parallel graphs and biconnected series-parallel graphs respectively. Then (see, e.g., [32, 52])

$$F(x) = xe^{B'(F(x))}. \tag{9.3}$$

An analogous formula works for any addable class of graphs. However, it fails for classes $\mathcal{C}^k$: we have to consider several types of rooted graphs instead.

Let G be a coloured graph with one pointed uncoloured vertex r, called the *root* of G. Let C be the set of all colours of G. We call a colour c *good* for G, if the graph obtained from G by adding a new vertex w connected to r and each vertex of G coloured c contains no K_4 as a minor. Otherwise we call c *bad* for G. We call G a *C-tree*, if the following three conditions are satisfied:

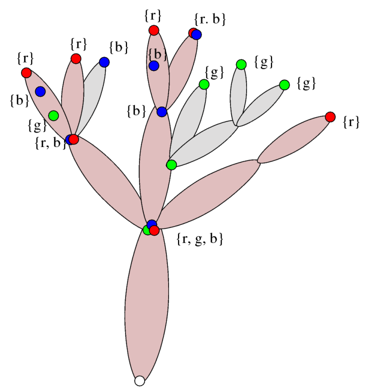

Figure 9.2: A $\{red, green, blue\}$-tree G. The elliptic shapes represent blocks, the white node is the root r. For each colour c the blocks on the path from the root vertex to a vertex coloured c form a sequence of SP-networks joined at their poles and only the "joints" of these networks can have colour c. All such blocks for any given colour c form a $\{c\}$-tree which is a "subtree" of G, here the subtree for $c = red$ is highlighted.

(a) each colour $c \in C$ is good for G,

(b) G is connected and it has no cut vertex x such that $G - x$ has a component without colours and without r, and

(c) r is not coloured, not a cut vertex of G and not the only vertex of G.

For a positive integer k and $C \subseteq [k]$, denote by $\mathcal{A}_C$ the family of all C-trees, see Figure 9.2. We define $\mathcal{A}_\emptyset = \emptyset$. If $C \neq \emptyset$, then for $n = 0, 1, 2$, $|\mathcal{A}_{C,n}|$ is equal to $0, 1$ and $2^{|C|}$ respectively. We will now study the exponential generating functions of $\mathcal{A}_C$; in the end of this section we will use the results to obtain the growth constant of $\mathcal{C}^k$.

Let G be a C-tree with root r, for some $C \subseteq [k]$. Then G has a unique rooted block tree T with root r. Consider any block B of G. Denote by $r(B)$ the vertex of B closest to r in G. For $v \in V(G)$, denote by G_v the subgraph of G induced on v and the vertices of all blocks of G that are ancestors of v in T (if v is not a cut vertex and $v \neq r$, then it has no ancestors), with the label and colour from v removed. We call the set of colours $\mathrm{Col}(v) \cup \mathrm{Col}(G_v)$ the *type* of v in G, and denote it by $type_G(v)$. For any block B of G and any colour c, let $X_c(B)$ denote the set of vertices $v \in V(B)$ such that $c \in type_G(v)$.

Proposition 9.2.7 *If G is a C-tree then for any block B of G and any $c \in C$ we have $X_c(B) \leq 1$.*

Proof Let G_B be the subgraph of G induced on the vertices of B and the vertices of all blocks that are ancestors of B in the rooted block tree of G. It is easy to see that G_B is a C'-tree for some $C' \subseteq C$. Therefore it suffices to prove the claim in the case where $r(B)$ is the root r of G.

Suppose $X_c(B) \geq 2$. Then G has a coloured minor isomorphic to a vertex-pointed 2-connected graph H obtained from $N(B)$ by setting $\mathrm{Col}(x) = \mathrm{Col}(y) = \{c\}$ for two distinct vertices $x, y \in V(B) \setminus \{r\}$. Now c is bad for H, since adding a new vertex with at least three neighbours to a 2-connected graph in $\mathrm{Ex}\, K_4$ yields K_4 as a minor. It follows that c is bad for G. □

For $k = 1, 2, \ldots$, denote by $\mathcal{B}_k$ the family of biconnected graphs in $\mathcal{A}_{[k]}$, such that each vertex has at most one colour. Again, if $B \in \mathcal{B}_k$ and $c \in [k]$, then there is exactly one vertex in B coloured c. For $n = 0, 1, 2$, $|\mathcal{B}_{1,n}|$ is equal to $0, 1, 2$ respectively; also $\mathcal{B}_{k,j} = 0$ if $j < k$.

For a set C of positive integers, denote by $\hat{\mathcal{A}}_C$ the set of all vertex-pointed graphs G, such that $\mathrm{Col}(G) = C$ and which further satisfy the conditions (a) and (b) of the definition of a C-tree. It is easy to check using the definition that $|\hat{\mathcal{A}}_{C,n}|$ is equal to 1 and $4^{|C|} - 2^{|C|}$ for $n = 0$ and 1 respectively. Each non-empty graph in $\hat{\mathcal{A}}_C$ can be decomposed uniquely into a (coloured and pointed) root vertex r and a set of graphs $G_1, \ldots G_t$ where for $i = 1, \ldots, t$, the graph G_i is a C_i-tree for some $C_i \subseteq C$, and $V(G_i) \cap V(G_j) = \{r\}$ for $i \neq j$.

Proposition 9.2.8 *Let C_1, C_2 be finite non-empty sets of positive integers. Suppose $G_1 \in \hat{\mathcal{A}}_{C_1}$ and $G_2 \in \hat{\mathcal{A}}_{C_2}$ have only their root vertex in common. Then $G_1 \cup G_2 \in \hat{\mathcal{A}}_{C_1 \cup C_2}$.*

Proof It is easy to see that the condition (b) holds for G. Suppose (a) does not hold, i.e., $c \in C_1 \cup C_2$ is bad for G. Consider the graph G^+ obtained by adding a new vertex w, connected to the root of G and each vertex coloured c. For $i = 1, 2$ let $G'_i = G^+[V(G_i) \cup \{w\}]$. By Lemma 8.4.2 for some $i = 1, 2$ we have $G'_i + rw$ has a subdivision of K_4, thus $G_i \notin \hat{\mathcal{A}}_{C_i}$. This is a contradiction. □

Lemma 9.2.9 *Let C be a finite non-empty set of positive integers. The exponential generating functions of $\hat{\mathcal{A}}_C$ and $\mathcal{A}_C$ are related by*

$$\hat{A}_C = \sum_{S \subseteq C} (-1)^{|C|-|S|} 2^{|S|} \exp\left(\sum_{S' \subseteq S} A_{S'} \right).$$

Notice, that by definition $A_\emptyset(x) = 0$ and $\hat{A}_\emptyset(x) = 1$.

Proof For any set $S \subseteq C$, define

$$\tilde{\mathcal{A}}_S = \cup_{T \subseteq S} \hat{\mathcal{A}}_T,$$

and note that $|\tilde{\mathcal{A}}_{S,0}| = 2^{|S|}$. Since $\mathcal{A}_S \cap \mathcal{A}_{S'} = \emptyset$ for $S \neq S'$, for any $C' \subseteq C$

$$\tilde{A}_{C'} = 2^{|C'|} \exp\left(\sum_{S \subseteq C'} A_S\right).$$

Fix a non-negative integer n, and for $S \subseteq C$ let $b_S = b_{S,n}$ be the number of graphs G in $\tilde{\mathcal{A}}_{C,n}$, such that $\mathrm{Col}(G) \cap S = \emptyset$. Then the number of graphs in $\tilde{\mathcal{A}}_{C,n}$ where some colour from C is missing, by the inclusion-exclusion principle is

$$\sum_{S \subseteq C, S \neq \emptyset} (-1)^{|S|-1} b_S.$$

Now $b_S = \left|\tilde{\mathcal{A}}_{C \setminus S,n}\right|$, therefore, summing over all $n \geq 0$

$$\tilde{A}_C - \hat{A}_C = \sum_{S \subseteq C, S \neq \emptyset} (-1)^{|S|-1} \tilde{A}_{C \setminus S} = \sum_{S \subset C} (-1)^{|C|-|S|-1} \tilde{A}_S$$

and

$$\hat{A}_C = \sum_{S \subseteq C} (-1)^{|C|-|S|} \tilde{A}_S = \sum_{S \subseteq C} (-1)^{|C|-|S|} 2^{|S|} \exp\left(\sum_{S' \subseteq S} A_{S'}\right).$$

□

Proposition 9.2.10 *Let k be a positive integer, and let C be a set of positive integers, $|C| \geq k$. Let $B \in \mathcal{B}_k$. For $i = 1, \ldots, k$, denote by v_i be the vertex of B coloured $\{i\}$. Let P be a partition of C into k non-empty sets $S_1, \ldots, S_k$ (listed in the lexicographic order), and let $G_1, \ldots, G_k$ be pairwise disjoint graphs, all disjoint from B, with $G_i \in \hat{\mathcal{A}}_{S_i}$, $i = 1, \ldots, k$.*

Then the graph G obtained by identifying the root r_i of G_i with v_i and setting $\mathrm{Col}_G(v_i) = \mathrm{Col}_{G_i}(r_i)$ for each i, is a C-tree.

Proof We have to show the conditions (a)-(c) of the definition of the C-tree are satisfied. It is trivial to check (b) and (c), so we will check just (a). Suppose it does not hold, i.e., $c \in C$ is bad for G. Consider the graph G^+ formed by adding to G a new vertex $w \notin V(G)$ and connecting w to the root r of G and each vertex coloured c. Let S_i be the set containing c. Let $G_1' = G[V(G_i) \cup w]$

and $G_2' = G^+ - (G_i - v_i)$. We have $G_1' \cup G_2' = G^+$ and $V(G_1') \cap V(G_2') = \{v_i, w\}$. Let K' be a subdivision of K_4 in G^+.

By Lemma 8.4.2, either K' is contained in G_1' or G_2', or the intersection of K' with G_j' is a path from v_i to w for some $j \in \{1, 2\}$. If K' is a subgraph of G_1', then c is bad for G_1; if K' is a subgraph of G_2', then c is bad for B. If $G_2' \cap K'$ is a path from w to v_i, then $G_1' + wv_i$ contains a subdivision of K_4, so c is bad for G_i. If $G_1' \cap K'$ is a path from w to v_i, then $B + wv_i$ contains a subdivision of K_4, and so c is bad for B. In each case we get a contradiction. □

We can now use the above observations and the decomposition into blocks, similarly as in (9.3) to give the exponential generating function for $\mathcal{A}_C$. Given a set $C \subseteq [k]$ for some positive integer k, let $\mathcal{P}(C)$ be the set of all set partitions of C, so that $|\mathcal{P}([j])|$ is the j-th Bell number.

Lemma 9.2.11 *Let k be a positive integer. For any non-empty set $C \subseteq [k]$, the exponential generating function of $\mathcal{A}_C$ is*

$$A_C(x) = \sum_{P \in \mathcal{P}(C)} B_{|P|}(x) \prod_{S \in P} \hat{A}_S(x).$$

Proof Each C-tree G may be decomposed into the (uncoloured) block B containing its root r, and a set of graphs G_v, such that $v \in X = \cup_{c \in C} X_c(B)$. Since the graph $\hat{G}_v$ obtained from G_v with vertex v coloured $\mathrm{Col}_G(v)$ and its label removed, is isomorphic to a coloured minor of G, we have that $\hat{G}_v \in \hat{\mathcal{A}}_{C'}$ where $C' = type_G(v)$. Let $v^{(1)}, \ldots, v^{(t)}$ be the vertices of X sorted according to their type in the lexicographic order, and for $v \in X$ let $ind(v)$ be the position of v in this list. The graph $\tilde{B}$ obtained from B by setting $\mathrm{Col}_{\tilde{B}}(v) = \{ind(v)\}$ for each $v \in X$ satisfies $\tilde{B} \in \mathcal{B}_t$. Now using Proposition 9.2.10 we see that each graph in $\mathcal{A}_C$ can be represented uniquely by and constructed from

- a root block $B \in \mathcal{B}_t$, for some $t \in [|C|]$,
- a partition $P = \{S_1, \ldots, S_t\}$ of C into non-empty sets (indexed in the lexicographic order), and
- pairwise disjoint graphs $G_1, \ldots, G_t$, all disjoint from B, where $G_i \in \hat{\mathcal{A}}_{S_i}$ for $i = 1, \ldots t$

by identifying the root r_i of G_i with the vertex v of B coloured $\{i\}$, and colouring that vertex $\mathrm{Col}_{G_i}(r_i)$. Now the exponential generating function is obtained in a standard way, see [48]. □

We write for shortness $\mathcal{A}_R = \mathcal{A}_{\{1\}}, \mathcal{A}_{RG} = \mathcal{A}_{\{1,2\}}$ and $\mathcal{A}_{RGB} = \mathcal{A}_{\{1,2,3\}}$.

Lemma 9.2.12 *The exponential generating functions of $\mathcal{A}_R, \mathcal{A}_{RG}$ and $\mathcal{A}_{RGB}$ satisfy*

$$\begin{aligned} A_R &= B_1\hat{A}_R; \\ A_{RG} &= B_1\hat{A}_{RG} + B_2\hat{A}_R^2; \\ A_{RGB} &= B_1\hat{A}_{RGB} + 3B_2\hat{A}_R\hat{A}_{RG} + B_3\hat{A}_R^3. \end{aligned}$$

Here

$$\begin{aligned} \hat{A}_R &= 2e^{A_R} - 1; \\ \hat{A}_{RG} &= 4e^{A_{RG}+2A_R} - 4e^{A_R} + 1; \\ \hat{A}_{RGB} &= 8e^{A_{RGB}+3A_{RG}+3A_R} - 12e^{A_{RG}+2A_R} + 6e^{A_R} - 1. \end{aligned}$$

Proof Notice, that by symmetry we have $A_C = A_{C'}$ whenever $|C'| = |C|$. The lemma follows by Lemma 9.2.9 and Lemma 9.2.11. □

9.2.3 Blocks of coloured trees (two colours)

In this section we present a decomposition for coloured graphs in the class $\mathcal{B}_2$. For a network G with two poles, we denote by G^+ the network obtained by connecting the poles with an edge. We will say that a colour c is *bad* for a (coloured) network G, if adding a new vertex w to G connected to the source of G and each vertex coloured c, we obtain a graph not in $\mathrm{Ex}\,\mathcal{B}$. Recall that in this section $\mathcal{B} = \{K_4\}$.

Lemma 9.2.13 *Let $\mathcal{S}_1$ be the class of $\{0,1\}^1$-coloured series SP-networks G where exactly one internal vertex is coloured $\{red\}$, and the colour red is good for G^+.*

Each graph in $\mathcal{S}_1$ admits a unique decomposition into two graphs in $\mathcal{D}$ or a graph $\mathcal{B}_2$ and a graph in $\mathcal{D}$. The exponential generating function of $\mathcal{S}_1$ is

$$S_1(x) = D(x)(xD(x) + B_2(x)).$$

We will use the following simple observation.

Lemma 9.2.14 *Let k be a positive integer and let G be a $\{0,1\}^k$-coloured graph with one pointed vertex r and exactly k coloured vertices, so that for each $i \in [k]$*

there is a vertex $v_i \in V(G) \setminus \{r\}$ *coloured* $\{i\}$*. Denote by* G_c *the network with source* r *and sink* u *obtained from* G *by removing the colour and the label from the vertex* u *coloured* $\{c\}$*.*

$G \in \mathcal{B}_k$ *if and only if for each* $c \in [k]$ *we have* $N(G_c) \in \mathcal{P} + \mathcal{E}_2$*.*

Proof If either $G \in \mathcal{B}_k$ or $N(G_c) \in \mathcal{P} + \mathcal{E}_2$, then G_c is biconnected, so $G_c \notin \mathcal{S}$.

($\Rightarrow$) Suppose $N(G_c) \notin \mathcal{P} + \mathcal{E}_2$ and let u be the vertex coloured $\{c\}$. Then $G_c + ru$ contains K_4 as a minor. We may replace ru by the path rwu where $w \notin V(G)$ to see that the colour c is bad for G, a contradiction.

($\Leftarrow$) Suppose we have $N(G_c) \in \mathcal{P} + \mathcal{E}_2$, but c is not good for G. Let u be the vertex coloured c. The assumption implies that with a new vertex $w \notin V(G)$, the graph $G' = N(G + rwu)$ contains K_4 as a minor. Then so does $G_c + ru$ and $N(G_c) \notin \mathcal{D}$, a contradiction. □

Proof of Lemma 9.2.13 Let $G \in \mathcal{S}_1$, let s and t be its source and sink respectively, and let v be the vertex coloured red. Then by Lemma 9.2.3, G may be decomposed into a sequence of (pairwise internally disjoint) networks $H_1, H_2, \ldots, H_k \in (\mathcal{P} + e)$ with $k \geq 2$ and vertices $x_1, \ldots, x_{k-1}$, where x_i is both the sink of H_i and the source of H_{i+1}.

Suppose v is an internal vertex of some H_j, $2 \leq j \leq k$. For $j = 1, \ldots, k$ denote by s_j and t_j the source and the sink of H_j respectively (we have $s_1 = s$ and $t_k = t$). By Proposition 9.2.5, H_j contains a cycle C with vertices v, s_j and t_j. By Lemma 9.2.3 and Proposition 9.2.4, there is a path P_1 from s to s_j in $H_1 \cup \cdots \cup H_{j-1}$, a path P_2 from t_j to t in $H_{j+1}, \ldots, H_k$ (which is trivial if $j = k$). Now the graph obtained from the union of C, P_1, P_2 and rt demonstrates that the colour red is bad for G^+. Therefore v cannot be an internal vertex of H_j, $j \geq 2$.

So v can have one of the following positions (and the cases are non-overlapping):

(a) $v = x_j$ for some $j \in [k-1]$;

(b) v is an internal vertex of H_1.

Suppose first that (a) holds. Denote by D_1 the SP-network with source s and sink v obtained from the union (series composition) of $H_1, \ldots, H_j$ and $x_1, \ldots, x_{j-1}$. Denote by D_2 the network with source v and sink t obtained from the union (series composition) of $H_{j+1}, \ldots, H_k$ and the vertices $x_{j+1}, \ldots, x_{k-1}$. By Lemma 9.2.3 and the comment thereafter, $D_1, D_2 \in \mathcal{D}$.

Now let $D_1', D_2' \in \mathcal{D}$ be arbitrary, and let G' be a network obtained by series composition of D_1' and D_2' by colouring the common pole $\{red\}$ and giving it an arbitrary label. Let s be the source of D_1' and G', and let t be the sink of D_2' and G'.

Lemma 9.2.3 and a comment after it, the decomposition of a graph $\mathcal{S}$ to a graph in $(\mathcal{P}+\mathcal{E}_2)$ and a sequence of graphs in $\mathcal{Z}\times(\mathcal{P}+e)$ is unique. Therefore, if $G'\in\mathcal{S}_1$ applying the decomposition of $G=G'$ into graphs $H_1,\dots,H_k$ as above, we recover $D_1=D_1'$ and $D_2=D_2'$.

Let us check that $G'\in\mathcal{S}_1$. Consider the network $\tilde{G}$ obtained from G'^{+} by making v a sink and t an internal vertex. $N(\tilde{G})$ is a parallel SP-network, since it is obtained by a series composition of the network st and the SP-network $\overleftarrow{D}_2'$ and a parallel composition of the resulting network with the SP-network D_1' (here $\overleftarrow{D}_2'$ denotes D_2' with its source and sink swapped. This change of orientation does not change the type of the network). By Proposition 9.2.6, the red colour is good for $\tilde{G}$, and so it is good for G'^{+}, and $G'\in\mathcal{S}_1$.

Now consider the case (b). Let $\tilde{H}_1$ be a $\{0,1\}^2$-coloured graph obtained from H_1 by colouring its sink green and assigning the label x_1. Since $\tilde{H}_1$ is a subgraph of G^{+}, if the red colour is bad for $\tilde{H}_1$, then it is also bad for G. If the green colour is bad for $\tilde{H}_1$, then the path P from s to x_1 in G^{+}, where P consists of the edge st and a path from t to x_1 in $D=H_2\cup\{x_2\}\cup\cdots\cup\{x_{k-1}\}\cup H_k\in\mathcal{D}$ shows that G^{+} contains K_4 as a minor, which is a contradiction. It follows that $\tilde{H}_1\in\mathcal{B}_2$ and $D\in\mathcal{D}$.

Now take an arbitrary graph $H'\in\mathcal{B}_2$ with root r, an arbitrary network $D'\in\mathcal{D}$ with source s' and sink t', and identify the green vertex of H' with the source of D' (call this vertex x) to obtain a network G' with source r and sink t'. Denote the vertex of G' coloured $\{red\}$ by v.

It is easy to see using the decomposition given by Lemma 9.2.3 that if $G'\in\mathcal{S}_1$, then the procedure described above applied with $G=G'$ recovers H' and D' of G' as $\tilde{H}_1$ and D respectively. It remains to show that $G'\in\mathcal{S}_1$.

Consider the graph $G''=G'^{+(w)}$ obtained by adding a new vertex w to G'^{+}, such that $\Gamma(w)=\{r,v\}$. Assume that red is bad for G'. Then G'' contains a subdivision K' of K_4. Since K' is 2-connected, it must contain both vertices v and r. Clearly, $\{r,x\}$ is a cut in G. Apply Lemma 8.4.2 to G'', and its subgraphs $\tilde{H}'=G''[V(H')\cup\{w\}]$ and $R=G''-(\tilde{H}'-\{r,x\})$. We consider three possible cases.

Case 1. K' is entirely contained in $\tilde{H}'$. Then $H'\notin\mathcal{B}_2$, which is a contradiction.

Case 2. $K'\cap\tilde{H}'$ is a path from r to x. Let G_1 be the network with poles r and x obtained by a series composition of rt' and $\overleftarrow{D'}$. Then the network $G_1+rx\in\mathcal{P}$ contains K_4 as a minor, this is a contradiction to Lemma 9.2.3.

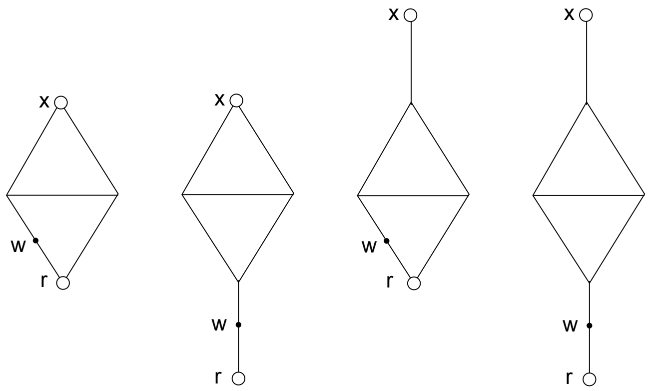

Figure 9.3: The part K_1' of the subdivision of K_4 contained in $\tilde{H}_1$ is a subdivision of one of these four types of graphs. Here the unlabelled vertices can be any vertices of $\tilde{H}_1$.

Case 3. $K' \cap R$ is a path P from r to x. Consider $K_1' = K' \cap \tilde{H}'$. Then K_1' is a subdivision of K_4 with a part of subdivided edge (i.e., the internal vertices of P) removed.

It must be that $w \in V(K_1')$, otherwise $N(G^+) \notin \mathcal{D}$, again contradicting Lemma 9.2.3. Since K' is 2-connected, w must have degree 2 in K_1'. We have that K_1' is a subdivision of one of the graphs shown in Figure 9.3, with the restriction that rw cannot be subdivided. Importantly, in all cases, the graph $K'' = K_1' - \{w, r, x\}$ is connected. The SP-network H' is parallel (it contains an internal vertex v), so by Lemma 9.2.3 it can be obtained in a unique way by parallel composition of some $l \geq 2$ networks $S_1, \ldots, S_l \in \mathcal{S} \cup e$. The connected graph H'' belongs to exactly one of these networks; change the indices if necessary, so that this network is S_1. Now, since $V(K') \cap V(S_2) \subseteq \{r, x\}$ we may use a path in S_2 from r to x to replace the path P and show that $\tilde{H}'$ also contains a subdivision of K_4. This demonstrates that $H' \notin \mathcal{B}_2$, which is a contradiction.

Combining the decompositions in each of the cases (a) and (b) yields the bijection

$$\mathcal{S}_1 = \mathcal{Z} \times \mathcal{D}^2 + \mathcal{D} \times \mathcal{B}_2,$$

which translates into the generating function, see [48], as claimed. □

Lemma 9.2.15 *Each graph in $\mathcal{B}_2$ admits a unique decomposition into a network in $\mathcal{D}$ and a network in $\mathcal{S}_1$. The exponential generating function of $\mathcal{B}_2$ satisfies*

$$B_2(x) = xD(x)S_1(x)$$

Proof Let $G \in \mathcal{B}_2$, let r be the root of G and let u and v be the vertices coloured $\{green\}$ and $\{red\}$ respectively. Consider the $\{0,1\}^1$-coloured network $\tilde{G}$ with

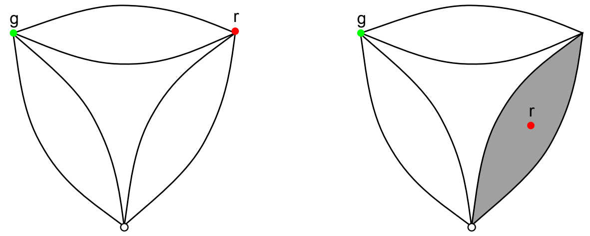

Figure 9.4: Graphs in the class $\mathcal{B}_2$ can be decomposed into one of the following constructions. Here the white shapes represent networks in $\mathcal{D}$, the grey shape represents a graph in $\mathcal{B}_2$ where the green vertex is converted into the sink. r and g mark vertices coloured red and green respectively.

poles r and u obtained with $c = green$ as in Proposition 9.2.14. $N(\tilde{G}) \in \mathcal{P}$ (it has an internal vertex v), so it can be decomposed uniquely using Lemma 9.2.3 into a network $D \in \mathcal{D}$ and a series graph S that contains v, both with poles s and u. Since the graph S^+ is isomorphic to a minor of $\tilde{G}$, if red is bad for S^+, then it is bad for $\tilde{G}$. Thus $S \in \mathcal{S}_1$.

Now take arbitrary networks $D' \in \mathcal{D}$ and $S' \in \mathcal{S}_1$ with disjoint sets of internal vertices and join them in parallel. Label the sink vertex u and colour it green. We claim that the resulting graph $G' \in \mathcal{B}_2$. Suppose, not. The colour green cannot be bad for G' by Proposition 9.2.6. Suppose red is bad for G'. The root r of G' is its unique pointed vertex (the common sink of D' and S'). Using Lemma 8.4.2, since $D' \in \mathcal{D}$, we get that red is bad for $S' + ru$, which contradicts to the definition of $\mathcal{S}_1$. So we have

$$\mathcal{B}_2 = \mathcal{Z} \times \mathcal{D} \times \mathcal{S}_2,$$

and applying the standard conversion to generating functions [48] completes the proof. □

We may combine the results of this section to obtain the full picture of graphs in the class $\mathcal{B}_2$, see Figure 9.4.

Corollary 9.2.16 *Each graph in $\mathcal{B}_2$ admits a unique decomposition into three graphs in $\mathcal{D}$ or two graphs in $\mathcal{D}$ and a graph in $\mathcal{B}_2$:*

$$\mathcal{B}_2 = \mathcal{Z}^2 \times \mathcal{D}^3 + \mathcal{Z} \times \mathcal{D}^2 \times \mathcal{B}_2.$$

Proof Combine Lemma 9.2.13 and Lemma 9.2.15. □

9.2.4 Blocks of coloured trees (general case)

In this section we give a nice characterisation of the class of coloured blocks $\mathcal{B}_k$ for arbitrary k. It turns out that each graph in $\mathcal{B}_k$ can be formed by substituting an SP-network for each edge of an "apex tree".

Let k be a positive integer. Let $\mathcal{T}'_k$ be the class of $\{0,1\}^k$-coloured Cayley trees T containing exactly k coloured vertices $v_1, \dots, v_k$, where vertex v_i has colour $\{i\}$ for each $i = 1, \dots, k$ with the following restriction: if a vertex $u \in V(T)$ has no colour, then it must have degree at least 3. Since every leaf is coloured by a unique colour, each $T \in \mathcal{T}'_k$ has only one automorphism. Therefore, we have $|\mathcal{T}'_{k,n}| = n!|(\mathcal{UT}'_k)_n|$, where $\mathcal{UT}'_k$ is the class of unlabelled $\{0,1\}^k$-coloured trees that can be obtained from the trees in $\mathcal{T}'_k$.

For $k = 1, 2, \dots$ the number of elements in $\mathcal{UT}'_k$ is

$$1, 1, 4, 31, 367, \dots$$

For example, all trees in the class $\mathcal{UT}'_3$ are shown in Figure 9.5. It is interesting to note, that the above sequence does not yet appear in the Sloane's Encyclopedia of Integer Sequences [95].

Now let $\mathcal{F}'_k$ be the class of all vertex-pointed graphs that can be obtained by taking a coloured tree $T \in \mathcal{T}'_k$, subdividing its edges arbitrarily (by inserting new uncoloured labelled vertices) to get a tree T', and finally adding a pointed root vertex r connected to each leaf of T' and each uncoloured vertex of degree 2 (edges rv where v is coloured or has at least 3 neighbours may be included or not included).

Let $\mathcal{F}'_k(\cdot, \mathcal{D})$ denote the class of graphs that can be obtained from graphs in $\mathcal{F}'_k$ by replacing their edges by arbitrary networks in $\mathcal{D}$. Since each network has an orientation (it starts with its source and ends with its sink), in order for such replacement to be well defined for a given $G \in \mathcal{F}'_k$ and $\{D_e \in \mathcal{D} : e \in E(H)\}$, the edges of G have to be oriented. We can assume that each edge in the tree $G - r$ points towards the red vertex, and each edge of G adjacent to r points away from r.

Theorem 9.2.17 *Let k be a positive integer. Each graph in $\mathcal{B}_k$ can be obtained in a unique way by substituting an SP-network for each edge of a graph in $\mathcal{F}'_k$:*

$$\mathcal{B}_k = \begin{cases} \mathcal{F}'_1(\cdot, \mathcal{P} + \mathcal{E}_2) = \mathcal{Z} \times (\mathcal{P} + \mathcal{E}_2) & \textit{for } k = 1 \\ \mathcal{F}'_k(\cdot, \mathcal{D}) & \textit{otherwise.} \end{cases}$$

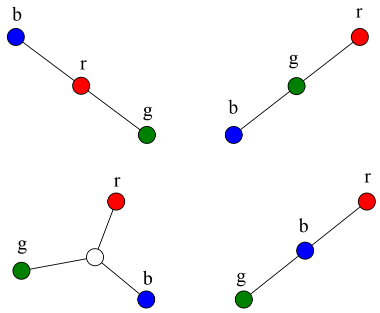

Figure 9.5: The isomorphism groups for the class of coloured trees $\mathcal{T}_3'$.

We will prove Theorem 9.2.17 using the next two lemmas.

Lemma 9.2.18 *Let $k \geq 2$ be an integer. Each pair $(H, \mathcal{D}_H)$, where $H \in \mathcal{F}_k'$ and $\mathcal{D}_H = \{D_e \in \mathcal{D} : e \in E(H)\}$ yields a unique graph $G = G(H, \mathcal{D}_H) \in \mathcal{B}_k$, where G is obtained from H by replacing e with D_e for each edge $e \in E(H)$.*

Proof Fix a pair $(H, \mathcal{D}_H)$ and let r be the root of $G = G(H, \mathcal{D}_H)$. Let us first prove that $G \in \mathcal{B}_k$. To this aim, by Lemma 9.2.14 it suffices to show that for each colour $c \in [k]$, if v_c is the vertex coloured $\{c\}$, then the network G_c with source r and sink v_c obtained from G as in Lemma 9.2.14 satisfies $N(G_c) \in \mathcal{P}$.

The graph G_c is 2-connected, so $N(G_c) \notin \mathcal{S} \cup \mathcal{E}_2$. Suppose $G_c + rv_c$ contains a subdivision K' of K_4. Let B be the block of $G_c - r$ containing the 2-core of $K' - r$ (the 2-core of a graph is the unique graph obtained by repeatedly deleting vertices of degree at most 1, until no such vertices remain). Since each vertex $v \in V(H) \setminus \{r\}$ is a cut point of $G_c - r$, we have that B is isomorphic to a subgraph of D_e for some $e = xy \in H$. Furthermore, G_c contains a path from the source to the sink of D_e which does not use any internal vertex of D_e. It follows by Lemma 8.4.2 that D_e is not an SP-network, a contradiction.

Let us now prove that given $G = G(H, \mathcal{D}_H)$ where $H \in \mathcal{F}_k'$ for any $k \geq 2$ we can always recover H and $\mathcal{D}_H$. We prove this claim by induction on $|V(H)|$.

Suppose $|V(H)| = 3$. Then the unique decomposition is provided by Corollary 9.2.16, and the corresponding tree in $\mathcal{F}_k'$ is the unique tree with two nodes u and v, which are the vertices of G coloured $\{green\}$ and $\{red\}$ respectively.

Now let $h \geq 4$ be an integer. Suppose that for any $\tilde{H} \in \mathcal{F}_k'$ on at most $h-1$ vertices, we can always recover $\tilde{H}$ and $\mathcal{D}_{\tilde{H}}$ given just the graph $G = G(\tilde{H}, \mathcal{D}_{\tilde{H}})$. Let $H \in \mathcal{F}_k'$ be a graph on h vertices with root r and let $\mathcal{D}_H = \{D_e \in \mathcal{D} : e \in E(H)\}$ be arbitrary. Denote the tree $H - r$ by T (it is a subdivision of a graph in $\mathcal{T}_k'$). Let u be a coloured vertex in $G - r$, such that there are no two coloured components in $G - \{r, u\}$. If there is more than one candidate for u, let u be such that $\mathrm{Col}(u) = \{j\}$ has the largest j. At least one such vertex exists since for any leaf x of the tree T, all coloured vertices of $T - u$ (and also in $G - \{r, u\}$) are in a single component. Furthermore, for any vertex u' of T that is an internal vertex of T, there are at least 2 coloured components in $T - u'$ (and also in $G - \{r, u'\}$). So u is the leaf vertex of T with the largest colour index.

Let C be the component of $G - \{r, u\}$ containing all coloured vertices. Then the network D'_{ru} with poles r and u obtained from $G - C$ is the network D_{ru}.

Now consider the graph $\tilde{G} = G - (V(D'_{ru}) \setminus \{r, u\})$. For each cut vertex v' of $\tilde{G}$, let $C(v', u)$ denote the component of $\tilde{G} - v'$ containing u. Let S be the set of cut vertices v' such that either v' is coloured or $G - C(v', u)$ is 2-connected. Finally, call a vertex $v' \in S$ a *candidate* if $G - C(v', u)$ does not contain any vertex from S. It is not difficult to see that there is exactly one candidate: the neighbour v of u in T.

Let D'_{uv} be the network obtained from $\tilde{G}[V(C(v, u)) \cup \{v\}]$ by making u the source and v the sink. We can see that $D'_{uv} = D_{uv}$ (the orientation is correct, since by definition $j > 1$).

Now consider the graph $\tilde{G} - V(D'_{uv} - v)$. If v is not coloured, colour it $\{j\}$, and add a dummy path vwr, where w is a vertex not in $\tilde{G}$. Denote the resulting graph by $\tilde{G}_2$. At the same time consider the graph obtained from $H - u$ by adding an edge rv, if it is not already there, and colouring v with $\{j\}$, if it is not coloured. Let P_{rwv} be the network with poles r and v obtained from the path rwv. If $rv \notin E(H)$, let $\tilde{D}_{rv} = P_{rwv}$ otherwise define $\tilde{D}_{rv} = D_{rv} \cup \tilde{P}$. The graph $\tilde{G}_2$ can be obtained from H' by replacing each edge by a network D'_e, where $D'_e = D_e$, if $e \in E(H') \setminus \{rv\}$ and $D'_{rv} = \tilde{D}_{rv}$.

By induction (rename colours, if necessary) we may recover H' and $\mathcal{D}'$ uniquely. Now we see, that connecting u to v and r in H', returning the original colour to v, and removing rv if $D'_{rv} = P_{rwv}$, recovers the graph H. For $e \in E(H) \setminus \{ru, rv, uv\}$, the network D_e is the network D'_e by induction, for $e \in \{rv, uv\}$, we have shown above that $D_e = D'_e$. Finally, if $rv \in H'$, we obtain D_{rv} as $D'_{rv} - w$. □

Lemma 9.2.19 *For any integer $k \geq 2$ we have $\mathcal{B}_k \subseteq \mathcal{F}'_k(\cdot, \mathcal{D})$.*

Proof We use induction on k. For $k = 2$, fix $G \in \mathcal{B}_2$. By Corollary 9.2.16, G admits either a representation by three SP-networks (we say that G is of the first type) or by two SP-networks and a (smaller) network $G' \in \mathcal{B}_2$ (we say that G is of the second type), see Figure 9.4. Let $G_0 = G$. For $i = 0, \ldots,$ if the graph G_i is of the second type, define $G_{i+1} = G'_i$. Let j be the index of the last G_i that has been defined.

To prove the lemma for the case $k = 2$ we apply induction on j. When $j = 0$, we have that G is of the first type, so by Corollary 9.2.16, it is a triangle $H \in \mathcal{F}'_{2,2}$ with each edge replaced by a network in $\mathcal{D}$. Now let $j' \geq 1$, assume that the claim holds for $k = 2$ and all $j \in \{0, \dots, j' - 1\}$, and suppose $j = j'$. Then by Corollary 9.2.16, the graph $G = G_0$ is obtained from G_1 by taking a series composition D of two graphs $D_1, D_2 \in \mathcal{D}$ with the common pole coloured $\{green\}$, identifying the sink of D with the green vertex u of G_1 and removing the colour from u. By induction, G_1 can be obtained from a graph $H' \in \mathcal{F}'_2$ by replacing each edge $e \in E(H')$ with a network $D'_e \in \mathcal{D}$.

Let H be a graph obtained from H' by inserting the vertex $u \notin V(H')$, so that u is connected to the green vertex u' of H' and the root, colouring u $\{green\}$ and removing the colour from u'. Clearly, $H \in \mathcal{F}'_2$. Also, for $e \in E(H')$ define $D_e = D'_e$, and let $D_{ru} = D_1$ and $D_{uu'} = D_2$. Thus G_0 can be obtained from the graph $H \in \mathcal{F}'_2$, by replacing each edge $e \in E(H)$ with D_e. This completes the proof for the case $k = 2$.

Assume now that we have proved the lemma for $\mathcal{B}_l$ with $l \in \{2, \dots, k-1\}$, and suppose $G \in \mathcal{B}_k$, where $k \geq 3$. Let u be the vertex of G coloured $\{k\}$. Remove the colour from u to obtain a graph $G' \in \mathcal{B}_{k-1}$. Use induction to find a graph $H' \in \mathcal{F}'_{k-1}$ and a set of networks $\mathcal{D}_{H'} = \{D'_e : e \in E(H')\}$ such that G' is the graph obtained by replacing each edge e of H' by D'_e. Let r be the root vertex of H', write $T' = H' - r$, and recall that T' is a subdivision of a tree in $\mathcal{T}'_{k-1}$.

The vertex u may have one of the following positions:

(a) $u \in V(T')$.

(b) u is an internal vertex of D'_e for some $e = xy \in E(T')$.

(c) u is an internal vertex of D'_{rv} for some $v \in V(T')$.

The case (a) is easy: we let H be the graph obtained from H' by colouring u with $\{k\}$, and let $\mathcal{D}_H = \mathcal{D}_{H'}$.

Consider the case (b). Suppose, u is not a cut vertex of D'_e. By Lemma 9.2.3 and Proposition 9.2.5, D_e contains a minor M isomorphic to the triangle K_3, such that x, y and u all belong to different bags. Now, since each component of $T' - xy$ contains at least one leaf of T', there are paths P_1 and P_2 in $G' - (D'_e - \{x, y\})$ from r to x and y respectively. Now M, P_1 and P_2 demonstrate that the colour k is bad for G. Thus, u must be a cut vertex of D'_e. Let H be the graph obtained from H'

by subdividing the edge xy with the vertex u. Let D_{ux} and D_{yu} be the networks with the common pole u (the orientation may be reversed, if necessary), such that D'_e results from the series composition of D_{ux} and D_{yu}. For $e \in E(H') \setminus \{e\}$, let $D_e = D'_e$, and define $\mathcal{D}_H = \{D_e : e \in E(H)\}$. Then we have $G = G(H, \mathcal{D}_H)$.

Now consider the case (c). Let G_1 be the graph obtained from $G[V(D'_{rv})]$, by colouring the vertex u $\{green\}$ and the vertex v $\{red\}$ and adding the edge rv, if $rv \notin E(G)$. For a $\{0,1\}^k$-coloured graph H and $c \in [k]$, let $(H)_c = H_c$ be as in Lemma 9.2.14. By Lemma 9.2.3, G_1 is 2-connected and $(G_1)_{red} \in \mathcal{P}$. By Lemma 9.2.14, $(G)_k \in \mathcal{P}$, and since G contains a path from r to v internally disjoint from D'_{rv} which can be contracted to an edge rv, we get that $(G_1)_{green} \in \mathcal{P}$. Therefore, applying Lemma 9.2.14 second time, we see that $G_1 \in \mathcal{B}_2$.

We have $N(G_1) \in \mathcal{P}$ and $G_1 \in \mathcal{B}_2$ by Lemma 9.2.14.

Using the already proved case $k = 2$, G_1 can be obtained from a graph $H_1 \in \mathcal{F}'_2$ by replacing each edge $e \in E(H_1)$ with an SP-network D''_e. Let $\tilde{H}_1$ be obtained from H_1 by setting $\mathrm{Col}_{\tilde{H}_1}(u) = \{k\}$ and $\mathrm{Col}_{\tilde{H}_1}(v) = \mathrm{Col}_G(v)$.

Now, if $D''_{rv} \in \mathcal{E}_2$ and $rv \notin E(G)$, let $H = H' \cup (\tilde{H}_1 - rv)$, otherwise, let $H = H' \cup \tilde{H}_1$. We can see that the tree $T = H - r$ is obtained from T' by attaching at the vertex v the graph $P = \tilde{H}_1 - r$ (which is a path from v to u). The vertex u is coloured $\{k\}$ in P and for each vertex $x \in V(P) \setminus \{v\}$ there is an edge $rx \in E(H)$ as required by the definition of $\mathcal{F}'_k$. Since $v \in V(T')$ and $T' \in \mathcal{F}'_{k-1}$, if v is not coloured it has degree at least 2 in T', and degree at least 3 in T. Hence $H \in \mathcal{F}'_k$. For $e \in E(H') \setminus \{rv\}$, define $D_e = D'_e$; for $e \in E(H_1) \setminus \{rv\}$, let $D_e = D''_e$. Finally, if $rv \in E(G)$, set $D_{rv} = D''_{rv} - rv$. Let $\mathcal{D}_H = \{D_e : e \in E(H)\}$: we have proved that $G = G(H, \mathcal{D}_H)$, as required. □

Proof of Theorem 9.2.17 The case $k = 1$ follows by Lemma 9.2.14. For $k \geq 2$, we combine Lemma 9.2.18 and Lemma 9.2.19. □

Given a class $\mathcal{A}$ of graphs and a parameter $X : \mathcal{A} \to \mathbb{Z}_{\geq 0}$, let $\mathcal{A}_{n,k} = \mathcal{A}^X_{n,k}$ denote the family of graphs $G \in \mathcal{A}_n$ with $X(G) = k$. We call

$$A(x, y) = \sum_{n \geq 0, k \geq 0} \frac{|\mathcal{A}_{n,k}|}{n!} x^n y^k$$

the *bivariate generating function* of $\mathcal{A}$ (where y "counts" X). Below, wherever X is not specified, y counts the number of edges, i.e. $X(G) = |E(G)|$.

It is not difficult to get the bivariate generating function F'_k for $\mathcal{F}'_k$, when k is small.

Lemma 9.2.20 *The bivariate generating functions of $\mathcal{F}_2'$ and $\mathcal{F}_3'$ are*

$$F_2(x,y) = \frac{x^2y^3}{1-xy^2}; \qquad F_3(x,y) = \frac{x^3y^4(3-2xy^2)(1+y)}{(1-xy^2)^3}.$$

Proof Denote by $\tilde{\mathcal{T}}_k'$ the class of trees obtained by subdividing edges of trees in $\mathcal{T}_k'$ arbitrarily. The graphs in $\tilde{\mathcal{T}}_2'$ are paths with coloured endpoints (the colours provide a unique orientation) and the univariate exponential generating function

$$\tilde{T}_2'(x) = \sum_{n=2}^{\infty} \frac{n!}{n!} x^n = \frac{x^2}{1-x}.$$

Each $T \in \tilde{\mathcal{T}}_2'$ on n vertices yields a unique fan $F \in \mathcal{F}_2'$ with $2n-1$ edges, so

$$F_2'(x,y) = \frac{\tilde{T}_2'(xy^2)}{y} = \frac{x^2y^3}{1-xy^2}.$$

Now consider $k = 3$. There are $3n!(n-2)$ trees $T \in \tilde{\mathcal{T}}_{3,n}'$ such that $N(T)$ is isomorphic to a path, and $n!\binom{n-2}{2}$ trees $T \in \tilde{\mathcal{T}}_{3,n}'$ which are subdivided 3-stars, see Figure 9.5. Therefore the exponential generating function of $\tilde{\mathcal{T}}_3'$ is

$$\tilde{T}_3'(x) = \sum_{n=3}^{\infty} 3(n-2)x^n + \sum_{n=4}^{\infty} \binom{n-2}{2} x^n = \frac{3x^3}{(1-x)^2} + \frac{x^4}{(1-x)^3}.$$

From each tree in $\tilde{\mathcal{T}}_3'$ on n vertices we can obtain two fans $F_1, F_2 \in \mathcal{F}_{3,n}'$ with $2n-1$ and $2n-2$ edges respectively (this is because the middle coloured vertex in the "path" case, and the centre of the star, in the "star" case may or may not be connected to the root). This yields the exponential generating function

$$F_3'(x) = \tilde{T}_3'(xy^2)(y^{-1} + y^{-2}) = \frac{x^3y^4(3-2xy^2)(1+y)}{(1-xy^2)^3},$$

as claimed. □

9.2.5 Growth of the class $\mathcal{A}_R$

We will use below the following fact about the class of SP-networks $\mathcal{D}$.

Lemma 9.2.21 *(Lemma 2.3 of [32]) Let $\mathcal{B}$ be the class of biconnected series-parallel graphs. Then*

$$\rho(\mathcal{B}) = \rho(\mathcal{D}) = \rho(\mathcal{P}) = \frac{(1+t_0)(t_0-1)^2}{t_0^3} = 0.1280.., \text{ and}$$
$$D(\rho(\mathcal{D})) = \frac{t_0^2}{1-t_0^2} = 1.8678..,$$

where $t_0 = 0.8070..$ is the unique positive solution of

$$(1-t^2)^{-1}\exp(-t^2/(1+t)) = 2.$$

By Lemma 9.2.12 and Theorem 9.2.17

$$A_R = B_1(2e^{A_R} - 1) = x(P+1)(2e^{A_R} - 1). \tag{9.4}$$

Notice, that if we add a new vertex w to any $\{red\}$-tree G, and connect it to the root and every vertex coloured red, we obtain a 2-connected series-parallel graph G'. This follows directly from the definition of a C-tree: if we delete any vertex $x \in V(G') \setminus \{w\}$, w has a neighbour in each of the components of $V(G) - \{x, w\}$, so $G' - x$ is connected. If we delete w we obtain the connected graph G. Thus each $\{red\}$-tree of size n gives a unique 2-connected series-parallel graph of size $n+2$ (we may label the root and the new vertex $n+1$ and $n+2$ respectively). Thus, if $\mathcal{B}$ is the class of biconnected series-parallel graphs, we have $\rho(\mathcal{B}) = \rho(\mathcal{D}) \leq \rho(\mathcal{A}_R)$. On the other hand, either looking at (9.4) or recalling that each SP-network yields a unique $\{red\}$-tree, we see that $\rho(\mathcal{A}_R) \leq \rho(\mathcal{D})$. We conclude that $\rho(\mathcal{A}_R) = \rho(\mathcal{D})$.

Proposition 9.2.22 *For any positive integer k, we have $\rho(\mathcal{B}_k) = \rho(\mathcal{D})$.*

Proof By Lemma 9.2.14

$$|\mathcal{B}_{k,n}| \leq n^k |(\mathcal{P} + \mathcal{E}_2)_{n-1}|,$$

so $\rho(\mathcal{B}_k) \geq \rho(\mathcal{P})$. By Theorem 9.2.17, each graph, obtained from a non-series SP-network G and a coloured path $P \in \tilde{\mathcal{T}}'_k$, by identifying the first endpoint of P with the sink of G and adding an edge between the source of G and second endpoint of P (if it is not already there), is in $\mathcal{B}_k$. So

$$|\mathcal{B}_{k,n}| \geq (n)_k |(\mathcal{P} + \mathcal{E}_2)_{n-k}|,$$

and $\rho(\mathcal{B}_k) \leq \rho(\mathcal{P})$. So $\rho(\mathcal{B}_k) = \rho(\mathcal{P}) = \rho(\mathcal{D})$ by Lemma 9.2.21. □

We remind a definition from [48]. Given two numbers ϕ, R with $R > 1$ and $0 < \phi < \pi/2$, define

$$\Delta(\phi, R) = \{z \in \mathbb{C} : |z| < R, z \neq 1, |arg(z-1)| > \phi\}.$$

A domain is a Δ-*domain at 1* if it is $\Delta(\phi, R)$ for some R and ϕ. For a complex number $\zeta \neq 0$, a Δ-domain at ζ is the image by the mapping $z \to \zeta z$ of a Δ-domain at 1.

For complex functions f, g we write $f(z) = O(g(z))$ as $z \to z_0$ if $|f(z)/g(z)|$ is bounded as $z \to z_0$.

The following fact is well known.

Lemma 9.2.23 *The exponential generating function* $R(x) = \sum_{n\geq 1} \frac{n^{n-1}x^n}{n!}$ *of rooted Cayley trees has a unique dominant singularity* e^{-1}*.* $R(x)$ *can be extended analytically to a* Δ*-domain* Δ *at* e^{-1}*, such that for all* $x \in \Delta$ *we have* $R(x) = xe^{R(x)}$ *and for* $x \to e^{-1}$*,* $x \in \Delta$ *we have*

$$R(x) = 1 - \sqrt{2}(1-ex)^{1/2} + O(1-ex). \tag{9.5}$$

Furthermore, $R(x)$ *is the unique solution* $y(x)$ *of* $y = xe^y$*, which is analytic at 0 and satisfies* $R(0) = 0$*.*

Proof See, e.g., Theorem VII.3 of [48] or Theorem 2.19 of [40]. For extension to a Δ-domain see, e.g., proof of Theorem 2.19 of [40]. The identity $R(x) = xe^{R(x)}$ for $|x| < e^{-1}$ is shown, i.e., in [48]. The identity then extends to the whole Δ domain by the Identity principle (see, e.g., Theorem 8.12 of [3]). $\square$

Recall that when we omit "(x)" in identities involving exponential generating functions and do not mention otherwise, we mean that they hold for some $\delta > 0$ and any $x \in \mathbb{C}$ with $|x| < \delta$. (If each side is an exponential generating function of a combinatorial class, this means that the counting sequences of both classes are identical.)

Lemma 9.2.24 *We have*

$$A_R = R(2B_1e^{-B_1}) - B_1.$$

Proof We may rewrite (9.4) as

$$\tilde{A}_R = Ee^{\tilde{A}_R}$$

where $\tilde{A}_R = A_R + B_1$ and $E = 2B_1 e^{-B_1}$. By Proposition 9.2.22, $\rho(\mathcal{B}_1) = \rho(\mathcal{D}) > 0$, so E is analytic at zero. Since $E'(0) = 2|\mathcal{B}_{1,1}| = 2 > 0$, E has an analytic inverse ψ_E at zero. Thus there is $\delta > 0$ such that for all $u \in \mathbb{C}$ with $|u| < \delta$ we have

$$f(u) = ue^{f(u)},$$

where $f(u) = \tilde{A}_R(\psi_E(u))$. Since $f(0) = E(0) = \psi_E(0) = 0$, we conclude (using Lemma 9.2.23) that $f(u) = R(u)$ for all u with $|u| < \delta$. This implies that there is $\epsilon > 0$, such that for all $x \in \mathbb{C}$ with $|x| < \epsilon$ we have $\tilde{A}_R(x) = f(E(x)) = R(E(x))$, or

$$A_R(x) = R(E(x)) - B_1(x).$$

Since the two analytic functions are identical on an open disc, they are identical for all x with $|x| < \rho(\mathcal{D})$. □

9.2.6 Growth of the class $\mathcal{A}_{RG}$

In contrast to $\mathcal{A}_R$, the exponential generating function of $\mathcal{A}_{RG}$ has a dominant singularity smaller than $\rho(\mathcal{D})$.

Lemma 9.2.25 *For $|x| < \rho(\mathcal{D})$ define a function $E = E(x)$ by*

$$E = 4B_1 \exp\left(2A_R - (4e^{A_R} - 1)B_1 + (2e^{A_R} - 1)^2 B_2\right).$$

The equation $E(x) = e^{-1}$ has only one solution $x_0 = 0.086468..$ in the interval $(0, \rho(\mathcal{D}))$ and $\rho(\mathcal{A}_{RG}) = x_0$.

Proof Combining Lemma 9.2.12, Lemma 9.2.20 and Theorem 9.2.17 we get

$$A_{RG} = B_1\left(4e^{A_{RG}+2A_R} - 4e^{A_R} + 1\right) + B_2(2e^{A_R} - 1)^2; \qquad (9.6)$$

$$B_1 = x(P + 1); \qquad B_2 = \frac{x^2 D^3}{1 - xD^2}.$$

Denote

$$B = (4e^{A_R} - 1)B_1 - (2e^{A_R} - 1)^2 B_2.$$

and

$$\tilde{A}_{RG} = A_{RG} + B.$$

We may rewrite (9.6) as

$$\tilde{A}_{RG} = Ee^{\tilde{A}_{RG}}.$$

Since all of the functions P, A_R, D, B_1, B_2 have convergence radius $\rho(\mathcal{D})$, the function $E(x)$ is analytic in the open disc $|x| < \rho(\mathcal{D})$. Furthermore, $E(x)$ is increasing for $x \in (0, \rho(\mathcal{D}))$. To see this, notice that the Taylor coefficients of $E_1 = E_1(x)$ given by

$$E_1 = 4\exp\left((2e^{A_R} - 1)^2 B_2\right)$$

are non-negative, so $E_1(x)$ is continuously increasing for $x \in (0, \rho(\mathcal{D}))$. Also, by (9.4) we have

$$2A_R - (4e^{A_R} - 1)B_1 = -B_1.$$

So

$$E_2 = B_1 \exp\left(2A_R - (4e^{A_R} - 1)B_1\right) = B_1 e^{-B_1}.$$

The function $B_1(x)$ continuously increases as $x \in (0, \rho(\mathcal{D}))$, since B_1 has non-negative Taylor coefficients, not all zero. Furthermore, $B_1(0) = 0$. By Lemma 9.2.21, (9.2) and numeric evaluation we get $B_1(\rho(\mathcal{D})) = 0.1929.. < 1$. Since the function $y(t) = te^{-t}$ continuously increases for $t \in (0,1)$ we conclude that both $y(B_1(x))$ and $E(x) = E_1(x)E_2(x)$ continuously increase for $x \in (0, \rho(\mathcal{D}))$.

We now claim that

$$\tilde{A}_{RG}(x) = R(E(x)), \tag{9.7}$$

where R is the exponential generating function for rooted Cayley trees. To see why, first note that

$$E'(0) = 4(P(0) + 1)e^0 = 4$$

and so, since E is analytic at 0 and $E(0) = 0$, $E(x)$ has an analytic inverse $\psi_E(u)$ for $|u| < \delta$, with some positive δ, such that $\psi_E(u) = 0$. For such u we have

$$f(u) = ue^{f(u)} \tag{9.8}$$

and we conclude as in the proof of Lemma 9.2.24 that for all $x \in \mathbb{C}$, $|x| < \rho$ where ρ is the radius of convergence of $R(E(x))$

$$\tilde{A}_{RG}(x) = f(E(x)) = R(E(x)).$$

Returning to A_{RG} we have

$$A_{RG}(x) = R(E(x)) - B(x). \tag{9.9}$$

Since A_{RG} has non-negative Taylor coefficients, by Pringsheim's theorem (see, e.g., [48]), it has a dominant singularity in $[0; \infty]$. The function E is continuously increasing for $x \in (0, 0.12] \subset (0, \rho(\mathcal{A}_R))$ and $E(0.12) = 0.6436.. > e^{-1}$, therefore there is exactly one solution of $E(x) = 1/e$ in $(0, \rho(\mathcal{D}))$; we call this solution x_0. Here we used (9.1), Lemma 9.2.24, (9.9) in Maple, to get the numeric evaluation of $E(0.12)$ and solve $E(x) = 1/e$. (Let us note here that D and R have explicit functional inverses, see [32, 48], therefore D, R, A_R, A_{RG} can be evaluated numerically at any point inside their disc of convergence).

The function A_{RG} is analytic for all $x < x_0$, and there is $\epsilon > 0$ such that B and E are analytic for all x with $|x| < x_0 + \epsilon$. Using the fact that E has an analytic inverse at x_0 (since $E'(x_0) > 0$) we conclude that x_0 must be a singularity of A_{RG}. □

9.2.7 Growth of the class $\mathcal{A}_{RGB}$.

For $\mathcal{A}_{RGB}$ we will apply a very similar analysis as in the previous section, we only have to work with slightly longer formulas.

Lemma 9.2.26 *For $|x| < \rho(\mathcal{A}_{RG})$ define $E(x) = E_1(x)E_2(x)$ where*

$$\begin{aligned}
E_1 &= 4\exp\{3B_2(2e^{A_R} - 1)^2(4e^{A_{RG}+2A_R} - 4e^{A_R} + 1) + B_3(2e^{A_R} - 1)^3\};\\
E_2 &= 2B_1\exp\{3A_{RG} + 3A_R + B_1(6e^{A_R} - 12e^{A_{RG}+2A_R} - 1)\}.
\end{aligned}$$

The equation $E(x) = e^{-1}$ has only one solution $x_1 = 0.044495..$ in the interval $(0, \rho(\mathcal{A}_{RG}))$ and $\rho(\mathcal{A}_{RGB}) = x_1$.

Proof By Lemma 9.2.12 we have

$$A_{RGB} = 8B_1e^{A_{RGB}+3A_{RG}+3A_R} - B,$$

where

$$\begin{aligned}
B = {}& B_1(12e^{A_{RG}+2A_R} - 6e^{A_R} + 1)\\
&- 3B_2(2e^{A_R} - 1)^2(4e^{A_{RG}+2A_R} - 4e^{A_R} + 1) - B_3(2e^{A_R} - 1)^3.
\end{aligned}$$

Setting $\tilde{A}_{RGB} = A_{RGB} + B$, we may rewrite this as

$$\tilde{A}_{RGB} = Ee^{\tilde{A}_{RGB}}.$$

Notice, that by Lemma 9.2.25, and Section 9.2.5, B_k (for $k \geq 1$), A_R and A_{RG} all convergence radius at least $\rho(\mathcal{A}_{RG})$. Therefore E is analytic (and continuous) at any point $x \in (0, \rho(\mathcal{A}_{RG}))$, We claim that $E(x)$ is increasing for $x \in (0, \rho(\mathcal{A}_{RG}))$. To see why, recall the notation of Lemma 9.2.12, and note that

$$E_1 = 4e^{3B_2\hat{A}_R^2\hat{A}_{RG}+B_3\hat{A}_R^3}$$

is an exponential generating function for a class of combinatorial objects, so its coefficients are non-negative, not all zero. Hence $E_1(x)$ is increasing for $x \in (0, \rho(\mathcal{A}_{RG}))$. Now, by (9.4) and (9.6), the exponent in E_2 is

$$\begin{aligned}
&3A_{RG} + 3A_R + B_1\left(6e^{A_R} - 12e^{A_{RG}+2A_R} - 1\right)\\
&= 3B_1\left(4e^{A_{RG}+2A_R} - 4e^{A_R} + 1\right) + 3B_2\left(2e^{A_R} - 1\right)^2\\
&+ 3B_1\left(2e^{A_R} - 1\right) + B_1\left(6e^{A_R} - 12e^{A_{RG}+2A_R} - 1\right)\\
&= -B_1 + 3B_2\hat{A}_R^2.
\end{aligned}$$

Thus

$$E_2 = 2B_1e^{-B_1}e^{3B_2\hat{A}_R^2}$$

is increasing for $x \in (0, \rho(\mathcal{A}_RG))$ using a similar argument as in the proof of Lemma 9.2.25. Since $E = E_1E_2$, E has the same property. Now since $E(0) = 0$, $E'(0) = 8|\mathcal{B}_{1,1}| = 8$, we have similarly as in the proof of Lemma 9.2.24

$$\tilde{A}_{RGB}(x) = R(E(x))$$

for all x in the disc of convergence of $\tilde{A}_{RGB}$, or, equivalently

$$A_{RGB}(x) = R(E(x)) - B(x).$$

Now using a numeric evaluation with $0.08 < \rho(\mathcal{A}_{RG})$ yields $E(0.08) = 0.855.. > 1/e$, and $B(x)$ is analytic for $x < \rho(\mathcal{A}_{RG})$. It follows similarly as in Lemma 9.2.25 that the smallest positive number x_1 such that $E(x_1) = 1/e$ is a dominant singularity of $R(E(x))$. Numerically solving with Maple yields $x_1 = 0.044495...$ (Here the numeric evaluation of $E(x)$ for $x \in (0, \mathcal{A}_{RG})$ can be easily carried out using the inverse functions of R and D, Lemma 9.2.24 and (9.9).) Since the convergence radius of $B(x)$ is at least $\rho(\mathcal{A}_{RG})$, it follows that x_1 is the convergence radius of $\mathcal{A}_{RGB}$. □

9.2.8 Completing the proofs

Proof of Lemma 9.2.1 Let $\mathcal{F}$ be the class of rooted series-parallel graphs. Bodirsky, Giménez, Kang and Noy [32] showed that the functional inverse ψ_F of F satisfies

$$\psi_F(u) = ue^{-B'(u)}$$

where

$$B(x) = \frac{1}{2}\ln(1 + xD(x)) - \frac{xD(x)(x^2D(x)^2 + xD(x) + 2 - 2x)}{4(1 + xD(x))}$$

is the exponential generating function of biconnected series-parallel graphs. Using Theorem 3.4 of [32], $\psi_F(u)$ is continuously increasing for $u \in [0, u_0)$, where $u_0 = F(\rho(\mathcal{F})) = 0.127969..$ (denoted $\tau(1)$ in [32]) and $\psi_F(u_0) = \rho(\mathcal{F}) = 0.11021...$

Recall that $\mathcal{A} = \mathcal{C}^{\bullet 3}$ is the class of all 3-rootable graphs rooted at a 3-rootable vertex. Then (cf. (9.3) and the proof of Lemma 9.2.9)

$$\mathcal{A} = 2^3 \times \mathcal{F} \times \prod_{S \subseteq [3]} \mathrm{SET}(\mathcal{A}_S(\mathcal{F})).$$

Therefore

$$A(x) = 8e^{A_{RGB}(F(x))}F(x)e^{\sum_{S\subset[3]} A_S(F(x))}. \tag{9.10}$$

By Lemma 9.2.26 $\rho(\mathcal{A}_{RGB}) = 0.044.. < F(\rho(\mathcal{F})) = u_0 = 0.1279...$ Let ρ be the unique solution in $(0, u_0)$ of

$$F(u) = \rho(\mathcal{A}_{RGB}),$$

so that $\rho = \psi_F(\rho(\mathcal{A}_{RGB})) = 0.042509..$

Since F and A_{RGB} have non-negative coefficients and $F'(\rho) \neq 0$, ρ is a singularity of $e^{A_{RGB}(F(x))}$. Moreover, since $\rho(\mathcal{A}_S) > \rho(\mathcal{A}_{RGB})$ for any $S \subset [3]$ by Lemma 9.2.25 and Lemma 9.2.26, we have that

$$g(x) = F(x)e^{\sum_{S\subset[3]} A_S(F(x))}$$

is analytic at ρ and $g(\rho) \neq 0$. It follows, see [48], that the radius of convergence of $A(x)$ is ρ, and so $\overline{\gamma}(\mathcal{A}) = \rho^{-1} = 23.524122...$ By Lemma 8.4.7, ρ^{-1} is the growth constant of $\mathcal{A} = \mathcal{C}^{\bullet 3}$.

Now

$$\overline{\gamma}(\mathcal{C}^2) \leq \gamma(\mathrm{apex}\,(\mathrm{Ex}\,K_4)) \leq 2\gamma(\mathrm{Ex}\,K_4) < \rho^{-1},$$

since $\gamma(\mathrm{Ex}\, K_4) = 9.07..$ by [32]. Also $\gamma(\mathcal{C}^2)$ exists and is equal to $\overline{\gamma}(\mathcal{C}^2)$ by Lemma 8.4.8. Applying Lemma 8.4.5 completes the proof. □

9.3 Counting tree-like graphs

9.3.1 Substituting edges, internal vertices and leaves of Cayley trees

Let $\mathcal{T}'$ be a class of trees. Let $\mathcal{D}, \mathcal{I}, \mathcal{L}$ be arbitrary non-empty classes of labelled objects. As before, we assume that all classes are closed under isomorphism of the labels. Although we use the same symbol to denote the class of series-parallel networks, in this section $\mathcal{D}$ will be an arbitrary class. We will consider the class $\mathcal{T}'(\mathcal{D}, \mathcal{I}, \mathcal{L})$ obtained from trees in $\mathcal{T}'$ by attaching to leaves, internal vertices and edges objects from $\mathcal{L}$, $\mathcal{I}$, $\mathcal{D}$ respectively. More precisely, denote by $L(T)$ and $I(T)$ the sets of all labelled leaves and labelled internal nodes of a tree T respectively (in this section, a node of T is called a leaf, if its degree is at most one; otherwise it is called an internal node). Then $\mathcal{T}'(\mathcal{D}, \mathcal{I}, \mathcal{L})$ is the class of all tuples $(T, \mathcal{D}', \mathcal{I}', \mathcal{L}')$ where $T \in \mathcal{T}'$, $\mathcal{D}' = \{D_e : e \in E(T)\}$, $\mathcal{I}' = \{I_v : v \in I(T)\}$ and $\mathcal{L}' = \{L_v : v \in L(T)\}$ are families of objects from $\mathcal{D}, \mathcal{I}$ and $\mathcal{L}$ respectively, and the sets of labels of each object in $\{T\} \cup \mathcal{L}_T \cup \mathcal{I}_T \cup \mathcal{D}_T$ are pairwise disjoint.

Suppose $\mathcal{I}, \mathcal{L}$ are classes of vertex-pointed graphs, $\mathcal{D}$ is a class of networks, and graphs in $\mathcal{T}'$ have at most one pointed (unlabelled) vertex. Then each object $\alpha = (T, \mathcal{D}_T, \mathcal{I}_T, \mathcal{L}_T) \in \mathcal{T}'(\mathcal{D}, \mathcal{I}, \mathcal{L})$ corresponds naturally to a graph $G(\alpha)$ defined as follows, see Figure 9.6. Starting with T, identify the pointed vertex of L_v with the node v of T for each $v \in L(T)$, identify the pointed vertex of I_u with the node u of T for each $u \in I(T)$ and replace each edge $uv \in E(T)$ by the network D_{uv}. To carry out the last substitution, fix a rule for orientation of edges of T. For instance, identify the source and the sink of D_{uv} with the smaller and the larger of $\{u, v\}$ respectively; a pointed vertex can be assumed to be smaller than any labelled vertex.

Let $\mathcal{T}$ be the class of (unrooted) Cayley trees. For example, if $\mathcal{Z}_C$ is the class of graphs consisting of a single pointed vertex coloured C and $\mathcal{E}_2$ is the class of trivial networks of size 0 containing a single edge, then the class $\mathcal{T}(\mathcal{E}_2, \mathcal{Z}_\emptyset, \mathcal{Z}_{\{red\}} \cup \mathcal{Z}_{\{green\}})$ is isomorphic to the class of all unrooted Cayley trees where the leaves are coloured either red or green.

For $\alpha \in \mathcal{T}'(\mathcal{D}, \mathcal{I}, \mathcal{L})$, let $T(\alpha)$ denote the underlying tree T. Our aim in this section is to enumerate general "supercritical" classes $\mathcal{T}(\mathcal{D}, \mathcal{I}, \mathcal{L})$ and obtain

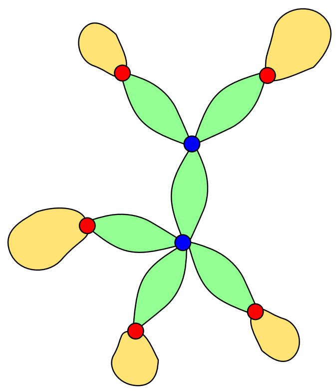

Figure 9.6: Theorem 9.3.1 characterises the growth of general "supercritical" class of graphs $\mathcal{T}(\mathcal{D},\mathcal{I},\mathcal{L})$ obtained by replacing edges, internal nodes and leaves of Cayley trees by objects from classes $\mathcal{D},\mathcal{I}$ and $\mathcal{L}$ respectively.

results on the underlying tree size: an application of this will be one of the key elements in the proof of Theorem 9.1.1.

Theorem 9.3.1 *Let $\mathcal{D},\mathcal{I},\mathcal{L}$ be non-empty classes of labelled objects. Let $\mathcal{A} = \mathcal{T}(\mathcal{D},\mathcal{I},\mathcal{L})$ and $\mathcal{A}' = \mathcal{R}(\mathcal{D},\mathcal{I},\mathcal{L})$. Suppose $\rho = \rho(\mathcal{A}) < \min(\rho(\mathcal{D}),\rho(\mathcal{I}),\rho(\mathcal{L}))$ and assume there are positive integers i,j,k with $gcd(k-i,j-i) = 1$ such that $\mathcal{A}$ contains an object of each of the sizes i,j and k.*

There are constants $a > 0$ and $c > 0$ such that the following holds. Let $R_n \in_u \mathcal{A}'$ or $R_n \in_u \mathcal{A}$ and let $Y_n = |V(T(R_n))|$.

1) For any $\epsilon > 0$, $\mathbb{P}\left(\left|\frac{Y_n}{n} - a\right| > \epsilon\right) = e^{-\Omega(n)}$;

2) $|\mathcal{A}_n| = (1+o(1))(an)^{-1}|\mathcal{A}'_n| = ca^{-1}n^{-5/2}n!\rho^{-n}(1+o(1))$;

3) A and A' converge at ρ.

To prove the theorem, we will need some preliminary results and a technical lemma. Let $\mathcal{D},\mathcal{I},\mathcal{L}$ be as in Theorem 9.3.1. Let $\mathcal{T}_1$ be the class of Cayley trees pointed at a leaf and containing at least two vertices. Consider the class $\mathcal{A}_1 = \mathcal{T}_1(\mathcal{D},\mathcal{I},\mathcal{L})$ with the bivariate generating function $A_1(x,s)$ where the second variable s counts the size of the underlying tree (which is the number of its nodes minus one). Then, writing $D = D(x)$, $I = I(x)$ and $L = L(x)$,

$$A_1(x,s) = sxDL + sxDI\left(e^{A_1(x,s)} - 1\right).$$

Consider additionally the class $\mathcal{A}_2$ with specification

$$\mathcal{A}_2 = \mathcal{A}_1 - \mathcal{Z}\times\mathcal{D}\times\mathcal{L} + \mathcal{Z}\times\mathcal{D}\times\mathcal{I},$$

Alternatively, $\mathcal{A}_2$ is the class $\mathcal{T}_1'(\mathcal{D}, \mathcal{I}, \mathcal{L})$, where $\mathcal{T}_1'$ is the same as $\mathcal{T}_1$, except that when we have a tree of size one (i.e., isomorphic to K_2), then its (unique) labelled vertex u is treated as an internal vertex and an object from $\mathcal{I}$, rather than from $\mathcal{L}$ is attached to it.

The bivariate generating function $A_2(x, s)$ of $\mathcal{A}_2$ satisfies

$$A_2(x, s) = sxDIe^{sxD(L-I)}e^{A_2(x,s)}. \tag{9.11}$$

Call a class $\mathcal{A}$ *aperiodic*, if there are positive integers i, j, k such that $i < j < k$, $|\mathcal{A}_i|, |\mathcal{A}_j|, |\mathcal{A}_k| > 0$ and $gcd(k - i, j - i) = 1$.

Lemma 9.3.2 *Let $\mathcal{D}, \mathcal{I}, \mathcal{L}$ be non-empty classes of labelled objects. If any of the classes $\mathcal{A} = \mathcal{T}(\mathcal{D}, \mathcal{I}, \mathcal{L}), \mathcal{A}_1, \mathcal{A}_2$ is aperiodic, then all of them are.*

Proof Given $\alpha \in \mathcal{T}'(\mathcal{D}, \mathcal{I}, \mathcal{L})$, and $x \in V^*(T(\alpha)) \cup E(T(\alpha))$ we denote by $Obj_x(\alpha)$ the object in $\mathcal{D} \cup \mathcal{I} \cup \mathcal{D}$ associated with x. Here $V^*(T) \subseteq V(T)$ denotes the set of labelled vertices of T.

1) Proof of $\mathcal{A}$ aperiodic $\implies$ $\mathcal{A}_2$ aperiodic. Consider an object α obtained from the path $P_2 = uvw$ where u and v are poles, with associated objects $D_{uv}, D_{vw} \in \mathcal{D}$, $I_v \in \mathcal{I}$ such that the label sets of D_{uv}, D_{vw}, I_v are pairwise disjoint, and disjoint from u. Let $\alpha_1, \alpha_2, \alpha_3 \in \mathcal{A}$ be objects of sizes i_1, i_2, i_3 respectively, such that $\gcd(i_3 - i_1, i_2 - i_1) = 1$. Let $l \in \{1, 2, 3\}$; we can assume that the label set of α_l is disjoint from the label set of α. Construct a new object α_l' from α_l and α as follows. If $|V(T(\alpha_l))| \neq 2$, then let x be a vertex of T with $d_T(v) \neq 1$. Merge α and α_l by identifying the vertex u of $T(\alpha)$ with x. If $V(T(\alpha_l))$ has two vertices, say a and b, then let α_l' be an object with underlying tree on edges $\{wv, va, vb\}$ by identifying u and a, so that $Obj_{vb}(\alpha_l') = Obj_{ab}(\alpha)$ and other associations are inherited from α and α_l, see Figure 9.7. We have $|\alpha_l'| = |\alpha_l| + |\alpha|$, and $\alpha_l \in \mathcal{A}_2$. Thus $\mathcal{A}_2$ contains objects of sizes $i_l' = i_l + |\alpha|$ for $l = 1, 2, 3$ and $\gcd(i_3' - i_1', i_2' - i_1') = \gcd(i_3 - i_2, i_2 - i_1) = 1$.

2) Proof of $\mathcal{A}_2$ aperiodic $\implies$ $\mathcal{A}_1$ aperiodic. By definition, the only objects in the symmetric difference of $\mathcal{A}_1$ and $\mathcal{A}_2$ are those, where the underlying tree has only one edge.

Let $\alpha \in \mathcal{A}_1$ be such that $|V(T(\alpha))| = 2$. $T(\alpha)$ contains one labelled and one pointed vertex. Let $\alpha_1, \alpha_2, \alpha_3 \in \mathcal{A}_2$ be objects of sizes i_1, i_2, i_3 respectively, such that $\gcd(i_3 - i_1, i_2 - i_1) = 1$. We may assume that the set of labels of α is disjoint from the set of labels of α_l for $l = 1, 2, 3$. From α_l we may obtain a new object as follows. Let u be an internal vertex of $T(\alpha_l)$, if $|V(T(\alpha_l))| \geq 3$, otherwise, let u

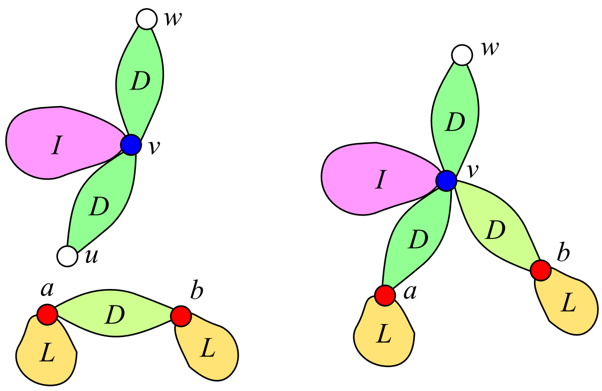

Figure 9.7: Top left: the object α, bottom left: the object α_l in the case $|T(\alpha_l)| = 2$, right: the object in $\mathcal{A}_1$ of size $|\alpha| + |\alpha_l|$ obtained by merging α_l and α.

be the unique labelled vertex of $T(\alpha_l)$. Merge the objects α and α_l by identifying the pointed vertex of $T(\alpha)$ with u, so that all the associated objects are inherited from the relevant tree. In particular, associate with u the object $Obj_u(\alpha_l) \in \mathcal{I}$. Call the resulting structure α_l', and note that $|\alpha_l'| = |\alpha_l| + |\alpha|$ and $\alpha_l' \in \mathcal{A}_1$ since u is an internal vertex of $T(\alpha_l')$. Similarly as above, it follows that $\mathcal{A}_1$ is aperiodic.

3) Proof of $\mathcal{A}_1$ aperiodic $\implies$ $\mathcal{A}$ aperiodic. From any object $\alpha \in \mathcal{A}_1$ we may obtain an object in $\mathcal{A}$ by labelling the pointed vertex u of α and associating with u an object in $\mathcal{L}$ of some fixed size. Now the claim follows similarly as in the previous cases. □

Lemma 9.3.3 *Let $\mathcal{D}, \mathcal{I}, \mathcal{L}$ be non-empty classes of labelled objects. For $\mathcal{A} = \mathcal{T}(\mathcal{D}, \mathcal{I}, \mathcal{L})$, $\rho(\mathcal{A}) = \rho(\mathcal{A}_1) = \rho(\mathcal{A}_2)$.*

Proof Constructions as in Lemma 9.3.2 show that for

$$(\mathcal{C}', \mathcal{C}'') \in \{(\mathcal{A}, \mathcal{A}_2), (\mathcal{A}_2, \mathcal{A}_1), (\mathcal{A}_1, \mathcal{A})\},$$

there is a positive integer s, such that from any object in $\mathcal{C}'_n$ we can construct a unique object in $\mathcal{C}''_{n+s}$. So $|\mathcal{C}''_{n+s}| \geq |\mathcal{C}'_n|$, $\rho(\mathcal{C}') \leq \rho(\mathcal{C}'')$ and the claim follows. □

For the function $f = f(x, s)$ below we denote by f_x and f_s its partial derivatives with respect to x and s.

Lemma 9.3.4 *Let $\mathcal{D}, \mathcal{I}, \mathcal{L}$ be non-empty classes of labelled objects. Let $\mathcal{A}_2$ be the class with the bivariate generating function $A_2(x, s)$ given in (9.11). Suppose that $\rho(\mathcal{A}_2) < m = \min(\rho(\mathcal{D}), \rho(\mathcal{I}), \rho(\mathcal{L}))$ and $\mathcal{A}_2$ is aperiodic.*

There is $\delta > 0$ such that the following holds. For any fixed $s \in [1 - \delta; 1 + \delta]$ we have

$$A_2(x, s) = R(f(x, s)), \tag{9.12}$$

where $f(x,s) = sxD(x)I(x)e^{sxD(x)(L(x)-I(x))}$ *and* R *is the Cayley tree function. The function* $A_2(x,s)$ *has a dominant singularity at* $\rho(s)$*, which is the smallest number in* $(0,m)$ *such that*

$$f(\rho(s),s) = e^{-1}.$$

Let $\rho = \rho(1)$*. We have* $f_s(\rho,1) > 0$*,* $f_x(\rho,1) > 0$ *and* $0 < \rho D(\rho)I(\rho) < 1$*,* $\rho(t)$ *is continuously differentiable for* $t \in [1-\delta, 1+\delta]$ *and* $\rho'(1) = -f_s(\rho,1)/f_x(\rho,1)$*. Furthermore* $A_2(x,s)$ *is analytic in a* Δ*-domain* Δ' *at* $\rho(s)$ *and for* $x \to \rho$*,* $x \in \Delta'$ *we have*

$$A_2(x,s) = 1 - c(s)(1 - x/\rho(s))^{1/2} + O((1 - x/\rho(s))) \qquad (9.13)$$

where $c(s) = (2e\rho(s)f_x(\rho(s),s))^{1/2}$ *is positive.*

It is not difficult to modify the proof and show that $O()$ holds uniformly for some $\delta > 0$ and $s \in [1-\delta, 1+\delta]$.

Proof We will write, for shortness, $D = D(x)$, $I = I(x)$ and $L = L(x)$.

Fix $s > 0$. We have $f(0,s) = 0$ and since $m > 0$, $f(x,s)$ is analytic at 0. Define $F^{[s]}(z,w) = f(z,s)e^w - w$. Then, see (9.11), the points $(x, A_2(x,s))$ are solutions of $F^{[s]}(x,y) = 0$. We have $R(f(0,s)) = 0$ and for x in a neighbourhood of 0, see Section 9.2.5,

$$R(f(x,s)) = f(x,s)e^{R(f(x,s))},$$

so $(x, R(f(x,s)))$ are also solutions of $F^{[s]}(x,y) = 0$. Since the derivative of $F^{[s]}$ with respect to w satisfies $F^{[s]}_w(0,0) = -1 \neq 0$ and $F^{[s]}(0,0) = 0$, (9.12) and the fact that $\rho(\mathcal{A}_2) > 0$ follow by the Analytic Implicit Function Theorem (Theorem B.4 of [48]) and the Identity principle.

To prove the rest of the lemma we will apply the "smooth implicit function schema" and a theorem of Meir and Moon [48,79]. The function $f(x,s)$ (and $F^{[s]}$) can have negative coefficients, therefore we will work with the function $A_1(x,s)$, which satisfies

$$A_1(x,s) = G^{[s]}(x, A_1(x,s)) \quad \text{where} \quad G^{[s]}(x,w) = sxDL + sxDI(e^w - 1).$$

(Alternatively, one could apply Theorem 2 of [79] directly to $F^{[s]}$.)

Suppose $f(x,1) < e^{-1}$ for all $x \in (0,m)$. Then for any $x \in (0,m)$, A_2 is analytic at x. Since by the Pringsheim's theorem (see [48]), $A_2(x,1)$ has a dominant singularity in $(0,\infty)$, we conclude that $\rho(\mathcal{A}_2) \geq m$, a contradiction. By continuity of $f(x,1)$, there exists a smallest positive $\rho \in (0,m)$, such that $f(\rho,1) = e^{-1}$.

An important observation is that $\rho D(\rho)I(\rho) < 1$. Suppose, this is false. $xD(x)I(x)$ continuously increases for $x \in (0, m)$ (it counts a non-empty combinatorial class), so there is a unique positive $x_0 \in (0, \rho]$ such that $x_0 D(x_0)I(x_0) = 1$. Since the function $h(z) = ze^{-z}$ is increasing for $z \in [0, 1)$ and $e^{xL(x)I(x)}$ is increasing for $x \in (0, m)$, we have that $f(x, 1) = h(xD(x)I(x))e^{xL(x)I(x)}$ is increasing for $x \in (0, x_0)$. But

$$f(x_0, 1) = e^{-1}e^{x_0 D(x_0)L(x_0)} = e^{-1}e^{L(x_0)/I(x_0)} > e^{-1},$$

so $\rho < x_0$, a contradiction. So $\rho D(\rho)I(\rho) < 1$, $\rho < m$, and we can further conclude that $f(x, 1)$ is continuously increasing for $x \in (0; \rho + \epsilon_1)$ for some $\epsilon_1 > 0$. This implies that $f_x(\rho, 1) > 0$. Furthermore, for $|x| < m$

$$f_s(x, s) = xDIe^{sxD(L-I)}(sxDL + 1 - sxDI),$$

and so $f_s(\rho, 1) > 0$.

Now consider a function $\tilde{F}(x, s) = f(x, s) - e^{-1}$, as a real function. Since $\tilde{F}(\rho, 1) = 0$, $\tilde{F}_x(\rho, 1) = f_x(\rho, 1) > 0$ and $\tilde{F}_s(\rho, 1) = f_s(\rho, 1) > 0$, by the Implicit Function Theorem (see, e.g., [90], Theorem 9.28), there is $\delta_1 > 0$ and a function $\rho(t) : \mathbb{R} \to \mathbb{R}$, such that for $t \in [1 - \delta_1, 1 + \delta_1]$, $\rho(1) = \rho$, $\rho(t)$ is continuously differentiable, $\tilde{F}(\rho(t), t) = 0$, $\rho'(1) = -\frac{\tilde{F}_s(\rho,1)}{\tilde{F}_x(\rho,1)} = -\frac{f_s(\rho,1)}{f_x(\rho,1)}$, and $\{(t, \rho(t)) : t \in (1 - \delta_1, 1 + \delta_1)\}$ contains all the solutions of $\tilde{F}(x, t) = 0$ in the region $[\rho - \delta_1, \rho + \delta_1] \times [1 - \delta_1, 1 + \delta_1]$.

Since $\rho < m$, $\rho I(\rho)D(\rho) < 1$, $\rho(t)$ is continuous at $t = 1$ and $xI(x)D(x)$ is analytic at $x = \rho$, we see that there is $\delta \in (0, \delta_1)$, such that for $t \in [1 - \delta, 1 + \delta]$, $s\rho(t)I(\rho(t))D(\rho(t)) < 1$ and $\rho(t) < m$. Now, since it is a product of two continuously increasing functions, $f(x, t) = h(txD(x)I(x))e^{txD(x)L(x)}$ increases for $x \in (0, \rho(t))$, so $\rho(t)$ is the smallest positive solution of $f(x, t) = e^{-1}$.

Assume $s \in [1 - \delta, 1 + \delta]$. Define $\tau(s) = 1 + s\rho(s)D(\rho(s))\left(L(\rho(s)) - I(\rho(s))\right)$. We claim that the function $G^{[s]}$ satisfies the "smooth implicit function schema" (Definition VII.4 of [48]). Indeed, $\rho(s) < m$, $\tau(s) < \infty$, the condition (I_1) is satisfied, since G is (bivariate) analytic for $|x| < m$ and $|w| < \infty$. The condition (I_2) follows since $G^{[s]}(0, 0) = G^{[s]}_w(0, 0) = 0$, and for any positive integer m, $[w^m]G^{[s]}(x, w) = (m!)^{-1}sxD(x)I(x)$ has non-negative coefficients, not all zero, since $\mathcal{D}, \mathcal{I}$ are non-empty. It remains to check the condition (I_3).

$$G^{[s]}(\rho(s), \tau(s)) = \tau(s) \quad \text{and} \quad G^{[s]}_w(\rho(s), \tau(s)) = 1.$$

Both identities follow after a simple calculation using the definition of $\tau(s)$ and the fact that $f(\rho(s), s) = e^{-1}$.

Furthermore, by Lemma 9.3.2, there are positive integers i_1, i_2, i_3, such that $|\mathcal{A}_{1,i_l}| > 0$, $\gcd(i_3 - i_1, i_2 - i_1) = 1$ and, denoting $\mathcal{A}_{1,n,m}$ the set of objects $\alpha \in \mathcal{A}_1$ with $|\alpha| = n$ and $|V(\alpha)| = m$,

$$[x^{i_l}]A_1(x, s) = [x^{i_l}](n!)^{-1} \sum_{m \geq 1} |\mathcal{A}_{1,i_l,m}| s^m > 0$$

for $l = 1, 2, 3$. Hence we can apply Theorem VII.3 of [48], which yields that $\rho(s)$ is the unique dominant singularity of $A_1(x, s)$ and there is a Δ-domain Δ at $\rho(s)$, such that for $x \to \rho_s$, $x \in \Delta$

$$A_1(x, s) = \tau(s) - c(s)(1 - z/\rho(s))^{1/2} + O((1 - x/\rho(s)))$$

where

$$c(s) = \left(\frac{2\rho(s) G_x^{[s]}(\rho(s), \tau(s))}{G_{ww}^{[s]}(\rho(s), \tau(s))} \right)^{1/2} = (2e\rho(s) f_x(\rho(s), s))^{1/2}.$$

The last equality follows using $f(\rho(s), s) = e^{-1}$, $G_{ww}^{[s]}(\rho(s), \tau(s)) = 1$ and comparing $G_x^{[s]}(\rho(s), \tau(s))$ and $f_x(\rho(s), s)$ term by term.

Finally, let us show that there is a Δ-domain at $\rho(s)$, such that $y(x) = A_1(x, s)$ is analytic at each $x \in \Delta'$. By Lemma VII.3 of [48], y is analytic in a region $D_0 = \{z \in \mathbb{C} : |z - \rho(s)| < r, |\arg(z) - \rho(s)| > \theta\}$ for some $r > 0$ and $0 < \theta < \frac{\pi}{2}$. By Note VII.17 of [48] y is analytic at any ξ with $|\xi| = \rho(s)$ and $\xi \neq \rho(s)$, i.e., there exists an open ball B_ξ centered at ξ and an analytic continuation of y in B_ξ. By compactness (see, e.g., proof of Theorem 2.19 of [40]), we may pick a finite number of points ξ, such that the respective balls cover all points $x \in D_0$ with $|x| = \rho(s)$, and the union of these finite balls together with $\{z : |z| < \rho(s)\}$ contains some Δ-domain Δ'.

□

Proof of Theorem 9.3.1 In Example IX.25 of [48], Flajolet and Sedgewick give a bivariate generating function $\tilde{R}(x, z)$ for rooted Cayley trees where the variable z counts leaves (the root is counted as a leaf only in the case $n = 1$)

$$\tilde{R}(x, z) = xz + x \left(e^{\tilde{R}(x,z)} - 1 \right) \tag{9.14}$$

which they show how to express using the usual (univariate) Cayley tree function

$$\tilde{R}(x,z) = x(z-1) + R(xe^{x(z-1)}).$$

Let $R(x,z)$ be the bivariate generating function for rooted Cayley trees, where z counts leaves, and the root is also counted as a leaf whenever its degree is at most one. Then, considering the cases when the root has 0, 1 or 2 children separately and using $xe^{R(x)} = R(x)$ we get

$$\begin{aligned} R(x,z) &= zx + zx\tilde{R}(x,z) + x(e^{\tilde{R}(x,z)} - \tilde{R}(x,z) - 1) \\ &= R\left(xe^{x(z-1)}\right)(x(z-1)+1) + x^2(z-1)^2 + x(z-1). \end{aligned} \tag{9.15}$$

Let us add a variable w that counts internal nodes:

$$R(x,w,z) = R(xw, z/w).$$

And another variable y, that counts edges:

$$R(x,y,w,z) = R(xyw, z/w)/y.$$

We get using (9.15)

$$R(x,y,w,z) = (xy(z-w)+1)R\left(xywe^{xy(z-w)}\right)y^{-1} + y(x(z-w))^2 + x(z-w).$$

Then the bivariate generating function $\mathcal{A}'$, where for $\alpha \in \mathcal{A}'$, the variable s counts the size of the underlying tree $T(\alpha)$ is

$$\begin{aligned} A'(x,s) &= R(sx, D, I, L) \\ &= (sxD(L-I)+1)\frac{R(f(x,s))}{D} + (sx(L-I))^2 D + sx(L-I). \end{aligned}$$

Here $D = D(x)$, $I = I(x)$, $L = L(x)$ are the exponential generating functions of $\mathcal{D}, \mathcal{I}$ and $\mathcal{L}$ respectively, and $f(x,s) = sxDIe^{sxD(L-I)}$ as before.

By Lemma 9.3.2, $\mathcal{A}_2$ is aperiodic. Denote $m = \min(\rho(\mathcal{D}), \rho(\mathcal{I}), \rho(\mathcal{L}))$. By Lemma 9.3.3, $\rho(\mathcal{A}_2) = \rho(\mathcal{A}) < m$. Therefore we can apply Lemma 9.3.4 to the class $\mathcal{A}_2$ and its bivariate generating function $A_2(x,s) = R(f(x,s))$.

Let $\rho = \rho(1) > 0$ be as in Lemma 9.3.4. Our constant a will be

$$a = -\frac{\rho'(1)}{\rho} = \frac{f_s(\rho,1)}{\rho f_x(\rho,1)}.$$

By Lemma 9.3.4, a is positive.

Fix a small $\epsilon \in (0, \min(0.5, 0.5a))$. Let $R'_n \in_u \mathcal{A}'$ be a uniformly random construction of size n and let $X'_n = |V(T(R'_n))|$. The key part of the proof will be to show that

$$\mathbb{P}(|X'_n - an| > \epsilon n) = e^{-\Omega(n)}. \tag{9.16}$$

Let δ be given by Lemma 9.3.4 applied with $\mathcal{D}, \mathcal{I}, \mathcal{L}$. We can assume that $\delta < \min\left(\frac{1}{2}, \frac{\epsilon}{16a}, \frac{\epsilon}{16a^2}\right)$. Fix $s \in \{1-\delta, 1, 1+\delta\}$. By Lemma 9.3.4, we may also assume that δ is small enough that

$$|\rho(s) - \rho - \rho'(1)| < \frac{\epsilon\rho\delta}{2} \quad \text{and} \quad 0 < \rho(s) < m. \tag{9.17}$$

We can write $A'(x,s) = E_1 D^{-1} A_2(x,s) + E_2$ where

$$\begin{aligned} E_1 &= E_1(x,s) = sxD(L-I) + 1; \\ E_2 &= E_2(x,s) = (sx(L-I))^2 D + sx(L-I). \end{aligned}$$

Using (9.17) $D(\rho(s)) > 0$, so $E_1 D^{-1}$ and E_2 are analytic at $x = \rho(s)$. So we have as $x \to \rho(s)$

$$E_1 D^{-1} = E_1(\rho(s)) D(\rho(s))^{-1} + O(x - \rho(s)), \quad E_2 = E_2(\rho(s)) + O(x - \rho(s)).$$

Using Lemma 9.3.4, $R(x) = xe^{R(x)}$, the fact that $m < \rho(s)$ and writing

$$A_2(x,s) = E_1 D^{-1} R(f(x,s)) + E_2 = E_1 sxIe^{sxD(L-I)+R(f(x,s))} + E_2$$

we get (see [48]) that $\rho(s)$ is the unique dominant singularity of $A'(x,s)$ and there is a Δ-domain Δ' at $\rho(s)$, such that for $x \to \rho(s)$, $x \in \Delta'$

$$A'(x,s) = c_0(s) - c_1(s)\left(1 - \frac{x}{\rho(s)}\right)^{1/2} + O\left(1 - \frac{x}{\rho(s)}\right),$$

with $c_1(s) = E_1(\rho(s), s)c(s)D(\rho(s))^{-1}$, $c_0(s) = c_1(s) + E_2(\rho(s), s)$.

Now by the "Transfer method" of Flajolet and Odlyzko (Theorem VI.1 and Theorem VI.3 of [48])

$$[x^n/n!]A'(x,s) = \frac{c_1(s)}{2\sqrt{\pi}} n^{-3/2} \rho(s)^{-n} \left(1 + O(n^{-1/2})\right),$$

the probability generating function of X'_n at s satisfies

$$\mathbb{E}\, s^{X'_n} = \frac{[x^n]A'(x,s)}{[x^n]A'(x,1)} = \left(\frac{\rho}{\rho(s)}\right)^n (1 + O(n^{-1/2})),$$

By Markov's inequality, for $s = 1 - \delta$

$$\begin{aligned}\mathbb{P}(X'_n \le (a-\epsilon)n) &= \mathbb{P}(s^{X'_n} \ge s^{(a-\epsilon)n}) \le \frac{\mathbb{E}\, s^{X'_n}}{s^{(a-\epsilon)n}} \\ &= \exp\left((p_1 - (a-\epsilon))n \ln s + o(1)\right) = e^{-\Omega(n)}.\end{aligned}$$

since by (9.17) and our choice of ϵ, δ

$$\begin{aligned}p_1 &= \frac{\ln \rho - \ln \rho(1-\delta)}{\ln(1-\delta)} \ge \frac{\ln(1 + a\delta - \delta\epsilon/2)}{-\ln(1-\delta)} \ge \frac{\ln(1 + a\delta - \delta\epsilon/2)}{\delta + \delta^2} \\ &\ge \frac{a\delta - \delta\epsilon/2 - (a\delta)^2}{\delta + \delta^2} \ge \left(a - \epsilon/2 - \delta a^2\right)(1-\delta) \\ &\ge a - \epsilon/2 - \delta a^2 - \delta a > a - \epsilon.\end{aligned}$$

Here we used simple inequalities $b - b^2 \le \ln(1+b) \le b$ and $1/(1+b) \ge 1 - b$, which are valid for any $b \in (-0.5, 0.5)$.

Now taking $s = 1 + \delta$, we get by Markov's inequality

$$\begin{aligned}\mathbb{P}(X'_n \ge (a+\epsilon)n) &\le \frac{\mathbb{E}\, s^{X'_n}}{s^{(a+\epsilon)n}} = \left(\frac{\rho}{\rho(s)s^{a+\epsilon}}\right)^n (1 + o(1)) \\ &= \exp\left((p_2 - (a+\epsilon))n \ln s + o(1)\right) = e^{-\Omega(n)}\end{aligned}$$

since similarly as above

$$\begin{aligned}p_2 &= \frac{\ln \rho - \ln \rho(s)}{\ln(1+\delta)} \le \frac{-\ln(1 - a\delta - \epsilon\delta/2)}{\ln(1+\delta)} \le \frac{a\delta + \delta\epsilon/2 + (a\delta + \delta\epsilon)^2}{\delta - \delta^2} \\ &\le (a + \epsilon/2 + 4a^2\delta)(1 + 2\delta) \le a + \epsilon/2 + 4a^2\delta + 4a\delta < a + \epsilon.\end{aligned}$$

This completes the proof of (9.16) and yields 1) with $Y_n = X'_n$.

Let us finish the proof of the theorem. Let $a_n(k,l)$ (respectively, $a'_n(k,l)$) be the number of objects α in $\mathcal{T}(\mathcal{D},\mathcal{I},\mathcal{L})_n$ (respectively, $\mathcal{R}(\mathcal{D},\mathcal{I},\mathcal{L})_n$) such that $|V(T(\alpha))| \in [k,l]$. Also let $A_n = a_n(1,n)$ and $A'_n = a'_n(1,n)$. Since each unrooted tree T corresponds to exactly $|V(T)|$ rooted trees

$$A_n \le A'_n \le nA_n \quad \text{and} \quad ka_n(k,l) \le a'_n(k,l) \le la_n(k,l). \tag{9.18}$$

By (9.16) we have

$$d_n = A'_n \mathbb{P}(|X'_n - an| > \epsilon n) = A'_n e^{-\Omega(n)}.$$

For $R_n \in_u \mathcal{A}$, let $X_n = X(R_n)$. The fact that 1) holds for $Y_n = X_n$ follows since

$$\mathbb{P}\left(|X_n - an| > \epsilon n\right) \leq \frac{nd_n}{A'_n} = e^{-\Omega(n)}.$$

Let us now show 2). Observe that since $X'_n \leq n$, by 9.16 it must be $a \in (0,1]$. Let $\epsilon' = \min(\epsilon, 1-a)$. By (9.18)

$$\frac{a'_n\left((a-\epsilon)n, (a+\epsilon')n\right)}{(a+\epsilon')n} \leq a_n\left((a-\epsilon)n, (a+\epsilon')n\right) \leq \frac{a'_n\left((a-\epsilon)n, (a+\epsilon')n\right)}{(a-\epsilon)n}.$$

So

$$a_n\left((a-\epsilon)n, (a+\epsilon')n\right) \in \left(\frac{(A'_n - d_n)(1-2\epsilon/a)}{an}, \frac{A'_n(1+2\epsilon/a)}{an}\right)$$

and also

$$A_n - a_n\left((a-\epsilon)n, (a+\epsilon')n\right) \leq d_n = e^{-\Omega(n)} A'_n.$$

So

$$\left|A_n - \frac{A'_n}{an}\right| \leq \frac{A'_n}{an}\left(\frac{2\epsilon}{a} + e^{-\Omega(n)}\right).$$

Letting ϵ go to zero shows that $A_n \sim \frac{A'_n}{an}$. Thus 2) follows with

$$c = \frac{c_1(1)}{2\sqrt{\pi}} = \frac{\rho D(\rho)(L(\rho) - I(\rho)) + 1}{D(\rho)} \left(\frac{e\rho f_x(\rho, 1)}{\pi}\right)^{1/2}.$$

Finally, since $R(e^{-1}) = 1$, we have that

$$A'(\rho) = A'(\rho, 1) = c_0(\rho) + c_1(\rho)$$

is finite. Using Lemma 9.3.3 we have $\rho(\mathcal{A}) = \rho$. By (9.18), the coefficients of $A(x)$ are dominated by the coefficients of $A'(x)$ and so $A(\rho) \leq A'(\rho) = c_0(\rho) + c_1(\rho)$. □

9.3.2 The case $\mathcal{B} = \{K_4\}$

In this section we will have $\mathcal{B} = \{K_4\}$ fixed and l a positive integer. Recall from the proof of Lemma 8.4.3, that $G \in \mathcal{C}^l$ is called nice if there is a vertex $x \in V(G)$ such that $G - x$ has at least two components containing all colours $[l]$. In this case, we call the vertex x *nice* in G. For $G \in \mathcal{A}_l$ we say that a vertex x is nice if it is nice in its connected component. Also recall that by $\mathcal{U} = \mathcal{U}^{<l>}$ we denote the class of graphs in $\mathcal{C}^l$ that are not nice. By Lemma 8.4.3, Lemma 8.4.7 and Lemma 8.4.5, we know that $\gamma(\mathcal{C}^l)$ exists and $\overline{\gamma}(\mathcal{U}) < \gamma(\mathcal{C}^l)$. Consider a graph $G \in \mathcal{C}^l$. Repeatedly "trim off" "pendant uncoloured subgraphs" from G (i.e. for $x \in V(G)$ such that $G - x$ has an uncoloured component H, delete $V(H)$ from G) until no such subgraphs remain. We call the remaining graph G' the *coloured core* of G.

Suppose G has a nice vertex r. Then since all coloured vertices remain in G', G' is also nice. Consider the rooted block tree T_r of G'. Recall that the nodes of T_r are $\{r\} \cup X \cup \mathcal{H}$, where X is the set of cut points of G' and $\mathcal{H}$ is the set of blocks of G'.

We call a block B of G' *simple* if there are at most two coloured components in $G' - E(B)$. If there are more than two coloured components in $G' - E(B)$, we call B *complex*. Suppose y is a nice vertex and $y \neq r$. Then, using Proposition 9.2.5 we see that every block node on the path P_{yr} from y to r in T_r must correspond to a simple block B. By Theorem 9.2.17, the edges of B form either a single edge or a parallel SP-network, where only the poles can be coloured. Furthermore, the poles s and t of B must be nice.

All paths $\{P_{yr} : y \text{ is nice}\}$ form an (unrooted) subtree T' of T_r, and the blocks corresponding to them form a connected subgraph of G'. Since a block node B of T' corresponds to a simple block, it has only two neighbours in T'. Therefore we may consider an unrooted tree T with $V(T) = \{y \in V(G') : y \text{ is nice}\}$ and $E(T) = \{xy : T' \text{ contains a path } xBy \text{ for some } B \in \mathcal{H}\}$. The trees T' and T do not depend on which nice vertex G is initially rooted at. We call T the *nice core tree* of G.

Fix $v \in V(T)$. Consider, the graph G'_v induced on v and the vertices of those components of $G' - v$ that do not contain any nice vertex, and pointed at the vertex v (by retaining the colour of v).

First suppose that v is a leaf node of T. By Proposition 9.2.8, G'_v admits a unique decomposition to a $\{0, 1\}^l$-coloured vertex (i.e. the root) and some graphs $H_1, \ldots, H_t$, where H_i is a C_i-tree with root v, such that $C_i \subseteq [l]$, $C_i \neq \emptyset$. The requirement that v is nice implies that at least one graph H_i must be an $[l]$-tree with the restriction that H_i does not have a subgraph H' rooted at a cut vertex $v' \neq v$, such that H' is an $[l]$-tree (otherwise v' would also be nice and v would not be a leaf vertex). The generating function counting the class $\bar{\mathcal{A}}$ of such graphs H_i is the same as the one given in Lemma 9.2.11, with the only difference that we do not allow an $[l]$-tree to be attached to the root block, so using Lemma 9.2.9 and

Lemma 9.2.11

$$\bar{A} = A_{[l]} - 2^l B_1(e^{A_{[l]}} - 1)\exp\left(\sum_{S\subset[l]} A_S\right)$$
$$= B_1\left(2^l \exp\left(\sum_{S\subset[l]} A_S\right) + \sum_{S\subset[l]} 2^{|S|}(-1)^{l-|S|}\exp\left(\sum_{S'\subseteq S} A_{S'}\right)\right)$$
$$+ \sum_{P\in\mathcal{P}([l]),|P|\geq 2} B_{|P|}\prod_{S\in P}\hat{A}_S.$$

If T has at least two vertices, then $G' - G'_v$ is non-empty and has a component that contains all colours $[l]$. Then the exponential generating function for the class $\mathcal{L}_2$ of all possible graphs G'_v is

$$L_2 = 2^l(e^{\bar{A}} - 1)\exp\left(\sum_{C\subset[l]} A_C\right).$$

If T has only one vertex, then there must be at least two graphs $H_i, H_j \in \bar{\mathcal{A}}$, so the exponential generating function for the class $\mathcal{L}_1$ of all possible graphs G'_v is

$$L_1 = 2^l(e^{\bar{A}} - \bar{A} - 1)\exp\left(\sum_{C\subset[l]} A_C\right).$$

Now suppose v is an internal node of T. Then $G'-G'_v$ has at least two components containing other nice vertices, and each such component contains all colours $[l]$. Therefore G'_v is in the class $\mathcal{I}$ with the exponential generating function

$$I = 2^l \exp\left(\bar{A} + \sum_{C\subset[l]} A_C\right).$$

Observe that our expressions for $\bar{A}, L_1, L_2, I$ all are given in terms of analytic functions of A_C, $C \subset [l]$ and B_i, $i \leq l$. Thus, by Lemma 9.2.1, Lemma 9.2.21 and Proposition 9.2.22 the convergence radii of each of these functions are at least $\rho(\mathcal{A}_{l-1}) > \rho(\mathcal{A}_l)$.

If G is a graph, v is a vertex of G and $\mathcal{A}$ is a class of vertex-pointed graphs, G' is obtained from G by *attaching* a graph $H \in \mathcal{A}$ at v if $G' = G\cup H$, $V(G')\cap V(H) = \{v\}$ and we assume that v inherits the label of G.

Recall that $\mathcal{F}$ denotes the class of all rooted series-parallel graphs. Let $\mathcal{F}_\circ$ denote the class of all vertex-pointed series-parallel graphs, so that $F_\circ(x) = F(x)/x$.

For two pointed graphs G_1 and G_2 with disjoint sets of labels, let $G_1 \times G_2$ be the pointed graph obtained by identifying their roots. Call classes $\mathcal{D}_1, \mathcal{D}_2$ of pointed graphs *uniquely mergeable*, if $G_1 \times G_2 \neq G_1' \times G_2'$ for all $G_1, G_1' \in \mathcal{D}_1$, $G_2, G_2' \in \mathcal{D}_2$, where $G_1 \times G_2$ and $G_1' \times G_2'$ are defined. If $\mathcal{D}_1$ and $\mathcal{D}_2$ are uniquely mergeable, we will identify with the combinatorial class $\mathcal{D}_1 \times \mathcal{D}_2$, the class of all graphs $G_1 \times G_2$, where $G_1 \in \mathcal{D}_1$, $G_2 \in \mathcal{D}_2$, and $G_1 \times G_2$ is defined.

Obviously, the classes $\mathcal{F}_\circ$ and $\mathcal{A}$ are uniquely mergeable, when $\mathcal{A}$ is $\mathcal{L}_1$, $\mathcal{L}_2$ or $\mathcal{I}$: the vertices of $G \in \mathcal{F}_\circ \times \mathcal{A}$ that belong to the graph $G_1 \in \mathcal{F}_\circ$ are exactly the root r of G and those vertices that are in the uncoloured components of $G - r$.

Given a class of graphs $\mathcal{A}$ and a class of rooted graphs $\mathcal{C}$, we denote by $\mathcal{A}(\mathcal{C})$ the class obtained from graphs in $\mathcal{A}$ by replacing each vertex by a graph in $\mathcal{B}$. Let $\mathcal{P}_+ = \mathcal{P} \cup \mathcal{E}_2$, be the class of non-series SP-networks. The above observations imply that each graph $G \in \mathcal{U}$ can be constructed as follows.

- Take a tree T of size at least one from the set of all unrooted trees $\mathcal{T}$ (i.e. a nice core).
- Replace each edge e of T by a network $D_e \in \mathcal{P}_+(\mathcal{F})$ (to fix the orientation, we may assume that edges of T are oriented away from the node with the smallest label in T).
- Attach at each leaf node of T a graph in $\mathcal{F}_\circ \times \mathcal{L}_1(\mathcal{F})$ (respectively in $\mathcal{F}_\circ \times \mathcal{L}_2(\mathcal{F})$) if T has one node (respectively, at least two nodes).
- Attach at each internal node of T a graph in $\mathcal{F}_\circ \times \mathcal{I}(\mathcal{F})$.

It is easy to see (for example, by fixing a root and comparing this construction with the construction of an $[l]$-tree) that the above decomposition is unique and the construction always yields a graph in $\mathcal{C}^l \setminus \mathcal{U}$.

Lemma 9.3.5 *Consider $\mathcal{B} = \{K_4\}$. Let $l \geq 2$ an integer, let $R_n \in_u \mathcal{C}^l$. Let Y_n denote the number of nice vertices in R_n. There is a positive constant a_l, such that*

$$\mathbb{P}(|Y_n - a_l n| > \epsilon n) = e^{-\Omega(n)}.$$

Proof Combining Corollary 9.2.2 with Lemma 8.3.9 we get that $\rho(\mathcal{C}^l) < \rho(\mathcal{C}^{l-1})$. Write

$$\tilde{\mathcal{T}} = \mathcal{T}(\mathcal{P}_+(\mathcal{F}), \mathcal{F}_\circ \times \mathcal{I}(\mathcal{F}), \mathcal{F}_\circ \times \mathcal{L}_2(\mathcal{F})),$$

and notice that $\tilde{\mathcal{T}}$ is aperiodic, since it contains, for example, all Cayley trees, where each node has colour $[l]$. The construction given in this section above yields the following identity

$$\mathcal{C}^l + \mathcal{Z} \times (\mathcal{F}_\circ \times \mathcal{L}_2(\mathcal{F})) = \tilde{\mathcal{T}} + \mathcal{Z} \times (\mathcal{F}_\circ \times \mathcal{L}_1(\mathcal{F})) + \mathcal{U}. \tag{9.19}$$

We will prove that the convergence radii of the exponential generating functions of $\mathcal{U}$, $\mathcal{P}_+(\mathcal{F})$, $\mathcal{F}_\circ \times \mathcal{I}(\mathcal{F})$, $\mathcal{F}_\circ \times \mathcal{L}_1(\mathcal{F})$ and $\mathcal{F}_\circ \times \mathcal{L}_2(\mathcal{F})$ are all at least $\rho(\mathcal{C}^{l-1})$. This implies that $|\mathcal{C}^l_n| = |\tilde{\mathcal{T}}_n|(1 + e^{-\Omega(n)})$ and the claim follows by Theorem 9.3.1.

Consider the class $\bar{\mathcal{C}}$ of rooted graphs obtained from graphs in $\mathcal{A}_{[l-1]}$ by replacing each labelled vertex by a graph in $\mathcal{F}$ and labelling the root. Then $\bar{C}(x) = xA_{[l-1]}(F(x))$ and $\bar{\mathcal{C}} \subseteq \mathcal{C}^{\bullet(l-1)}$ (defined in Section 8.4.1). Using Lemma 8.4.5, $\bar{\rho} := \rho(\bar{\mathcal{C}}) \geq \rho(\mathcal{C}^{\bullet(l-1)}) = \rho(\mathcal{C}^{l-1})$. By [32] the functional inverse $\psi_F(x)$ of F is increasing for $x \in (0, x_0)$ and $F(x_0) > \rho(\mathcal{D})$, where $x_0 = F(\rho(\mathcal{F})) = 0.1279..$ (denoted $\tau(1)$ in [32]). By Proposition 9.2.22 $\rho(\mathcal{D}) = \rho(\mathcal{B}_{l-1}) \geq \rho(\mathcal{A}_{[l-1]})$, so $x_0 \geq \rho(\mathcal{A}_{[l-1]})$. We see (using e.g., Section VI.9 of [48]) that $\bar{\rho} = \psi_F(\rho(\mathcal{A}_{l-1}))$.

By our construction above, for $i \in \{1,2\}$, $\rho(\mathcal{L}_i) \geq \rho(\mathcal{A}_{[l-1]})$, therefore $\rho(\mathcal{Z} \times (\mathcal{F}_0 \times \mathcal{L}_i(\mathcal{F}))) \geq \psi_F(\rho(\mathcal{A}_{[l-1]})) = \bar{\rho} \geq \rho(\mathcal{C}^{l-1})$. Since each graph in $\mathcal{P}_{+,n}$ yields a unique graph in $\mathcal{F}_{n+2}$, $\rho(\mathcal{P}_+(\mathcal{F})) \geq \rho(\mathcal{F}) \geq \rho(\mathcal{C}^{l-1})$. Finally, $\rho(\mathcal{U}) \geq \rho(\mathcal{U}') \geq \rho(\mathcal{C}^{l-1})$ by Lemma 8.4.3. □

9.4 Structure of random graphs in $\mathrm{Ex}\,(k+1)K_4$

9.4.1 Proof of Theorem 9.1.1 and Theorem 9.1.2

Let H be a fixed connected coloured graph on vertices $\{1,\ldots,h\}$. Following [77], we say that H *appears* in G at $W \subseteq V(G)$ if (a) the increasing bijection from $\{1,\ldots,h\}$ to W gives an isomorphism between H and $G[W]$ and (b) there is exactly one edge in G between W and the rest of G, and it is incident with the smallest element of W. We let $f_H(G)$ denote the number of sets W such that H appears at W in G.

Let $\mathcal{A}$ be a class of (coloured) graphs and let H be a connected graph, rooted at $r \in V(H)$. Let $G \in \mathcal{A}$, and let $S \subseteq V(G)$. Suppose G and S have the following property: if we take any number of pairwise disjoint copies of H, all disjoint from G, and add an edge between the root of each copy and a vertex in S then the resulting graph is still in $\mathcal{A}$. The set S is called an *H-attachable* subset of G (with respect to $\mathcal{A}$).

The next lemma and its proof is just an adaptation of Theorem 4.1 of [77] for graphs where not necessarily all of the vertices form an H-attachable set.

Lemma 9.4.1 *Let $\mathcal{C}$ be a non-empty class of (coloured) graphs, and suppose $\gamma(\mathcal{C}) = c \in [e^{-1}; \infty)$. Let H be a connected (coloured) graph on the vertex set $\{1, \ldots, h\}$ rooted at 1. Suppose there are constants $a \in (0, 1)$, $N_0 > 0$ and $d > 0$ such that the probability that $R_n \in_u \mathcal{C}$ has an H-attachable subset (with respect to $\mathcal{C}$) of size at least an is at least $1 - e^{-dn}$ all $n \geq N_0$. Fix α, such that $\alpha < d$ and $\alpha \leq a/(9e^2 c^h (h + 2)h!)$. Then there exists n_0 such that*

$$\mathbb{P}(f_H(R_n) \leq \alpha n) \leq e^{-\alpha n} \text{ for all } n \geq n_0.$$

Proof The proof is a simple modification of the proof of Theorem 4.1 of [77]. We skip some of the details and refer the reader for them to [77]. Write $\beta = a^{-1}e^2c^h(h+2)h!$ and let $\epsilon \in (0, 1/3)$ be such that $(\alpha\beta)^\alpha = 1-3\epsilon$. Let $f(n)$ denote the number of graphs in $\mathcal{C}_n$. Then since $\gamma(\mathcal{C}) = c$ there is $n_1 \geq N_0$ such that for each $n \geq n_1$ we have $e^{-\alpha n} \geq 2e^{-dn}$ and

$$2(1 - \epsilon)^n n! c^n \leq f(n) \leq (1 + \epsilon)^n n! c^n. \tag{9.20}$$

Assume that for infinitely many $n \geq n_1$ the claim of the lemma does not hold: that is, at least $e^{-\alpha n}$ fraction of graphs in $G \in \mathcal{C}_n$ "have few pendant appearances", i.e., $f_H(G) \leq \alpha n$. Let $\tilde{\mathcal{C}_n} \subseteq \mathcal{C}_n$ consist of those graphs in $\mathcal{C}_n$ that have few pendant appearances and an H-attachable subset with at least an vertices. Then

$$|\tilde{\mathcal{C}_n}| \geq f(n)(e^{-\alpha n} - e^{-dn}) \geq e^{-\alpha n}(1 - \epsilon)^n n! c^n.$$

Let $\delta \in (0, 1)$ be given by $\delta = \alpha h$. We can construct a graph G on $\lceil n(1 + \delta)\rceil$ vertices by putting a graph G_0 isomorphic to a graph in $\tilde{\mathcal{C}_n}$ on some n of these vertices, and adding $\lfloor \alpha n \rfloor$ disjoint copies of H on the remaining $\lceil \delta n \rceil$ vertices, so that for each added copy H' of H there is an edge between the least vertex of H' and some $y \in S_0$, where S_0 is the largest H-attachable subset of G_0. The number of such constructions $b_{\lceil (1+\delta)n \rceil}$ satisfies

$$\begin{aligned} b_{\lceil (1+\delta)n \rceil} &\geq \binom{\lceil (1 + \delta)n \rceil}{n} |\tilde{\mathcal{C}_n}| \binom{\lceil \delta n \rceil}{h, \ldots, h} \frac{(an)^{\lfloor \alpha n \rfloor}}{\lfloor \alpha n \rfloor !} \\ &\geq \lceil (1 + \delta)n \rceil ! e^{-\alpha n}(1 - \epsilon)^n c^n \frac{a^{\lfloor \alpha n \rfloor}}{h!(h!\alpha)^{\lfloor \alpha n \rfloor}}. \end{aligned}$$

Now [77] show that each graph in this way is constructed at most $\binom{\lfloor (h+2)\alpha n \rfloor}{\lfloor \alpha n \rfloor} \leq$

$e((h+2)e)^{\lfloor \alpha n \rfloor}$ times. This, after a similar calculation as in [77] yields that

$$f(\lceil (1+\delta)n \rceil) \geq \frac{b_{\lceil (1+\delta)n \rceil}}{e((h+2)e)^{\lfloor \alpha n \rfloor}} \geq c' f(\lceil (1+\delta)n \rceil) \left(\frac{1-\epsilon}{(1-3\epsilon)(1+\epsilon)^2} \right)^n,$$

for some constant $c' > 0$ which does not depend on n. Since $\frac{1-\epsilon}{(1-3\epsilon)(1+\epsilon)^2} > 1$, our assumption cannot hold for infinitely many n, a contradiction. □

Corollary 9.4.2 *Let $\mathcal{B} = \{K_4\}$, let $h \geq 1$ and $l \geq 2$ be integers and suppose $H \in \mathcal{C}_h^l$ is rootable at 1. There is a constant $a = a(l,h) > 0$ such that the random graph $R_n \in_u \mathcal{A}_l$ satisfies*

$$\mathbb{P}(f_H(R_n) \geq an) \geq 1 - e^{-\Omega(n)}.$$

Proof By Proposition 9.2.8 and the decomposition of Section 9.3.2 the set of nice vertices of a graph $G \in \mathcal{A}_l$ is H-attachable. The claim follows by Lemma 9.3.5 and Lemma 9.4.1. □

Fix positive integers r and l, $r > l$. Recall the class $\tilde{\mathcal{A}} = \tilde{\mathcal{A}}^{<\mathcal{B},l,r>}$ defined in the proof of Lemma 8.3.7: $\tilde{\mathcal{A}}$ is the class of $\{0,1\}^r$-coloured graphs corresponding to the class of graphs that have an $(l,2,\mathcal{B})$-double blocker of size r, and $\tilde{\mathcal{C}}$ is the class of connected such graphs. (In this section $\mathcal{B} = \{K_4\}$ is fixed.)

Suppose H is an induced subgraph of a coloured graph G. Similarly as in the previous chapters we will call H a *spike* of G if all of the following hold:

- H is a path $v_1 \dots v_{l+1}$;
- there is only one edge between $V(H)$ and $V(G-H)$, and this edge is uv_1 where $u \in V(G-H)$;
- $\mathrm{Col}_H(v_1) = \cdots = \mathrm{Col}_H(v_{l+1}) = \{1,\dots,l,x\}$ where $x \in \{l+1,\dots,r\}$;
- $u < v$ for each $v \in V(H)$.

It is easy to see that two different spikes must be pairwise disjoint.

Lemma 9.4.3 *Let $\mathcal{B} = \{K_4\}$, let r and l be positive integers, $r > l$ and consider the random graph $R_n \in_u \tilde{\mathcal{C}}$. There is a constant $a' = a'(r,l)$ such that*

$$\mathbb{P}(R_n \textit{ has less than } a'n \textit{ spikes}) \leq e^{-\Omega(n)}.$$

Proof Let H be a $\{0,1\}^{l+1}$-coloured path on the vertex set $[l+2]$ such that one of its endpoints is 1 and for $v \in \{2,\dots,l+2\}$, we have $\mathrm{Col}_H(v) = [l+1]$. By

Corollary 9.4.2, there are positive constants a, c and C, such that the number of graphs in $\mathcal{C}_n^{l+1}$ with at most an pendant appearances H is at most $Ce^{-cn}|\mathcal{C}_n^{l+1}|$ for every n.

Let $N = (1 + \mathrm{aw}_2(\mathrm{Ex}\, K_4))^{l-1} = 3^{l-1}$. In the proof of Lemma 8.3.7 we have shown that each graph in $\tilde{\mathcal{C}}_n$, as well as some other graphs, can be obtained as follows.

- Pick $\kappa \in [N]$ and $j, m \in \{0, \dots, N-1\}$;
- choose a partition $\mathcal{S}$ of $[n+j]$ into κ sets $V_1, \dots, V_\kappa$;
- for each $i = 1, \dots, \kappa$ put an arbitrary graph $H_i \in \mathcal{C}^{l+1}$ on V_i;
- for each $i = 1, \dots, \kappa$ choose $q_i \in \{l+1, \dots, r\}^\kappa$ and map the colour $l+1$ in H_i to q_i;
- choose a set J of m edges between the components $H_1, \dots, H_\kappa$ and add them to the resulting graph;
- finally, contract all edges J, so that the vertex resulting from a contraction of an edge $e = xy$ receives $max(x, y)$ as a label.

Consider the set $\mathcal{M}(n)$ of all possible constructions that yield a graph on the vertex set $[n]$ (the (multi-)set of the resulting graphs contains $\tilde{\mathcal{C}}_n$).

By Lemma 8.4.8 the class $\mathcal{C}^{l+1}$ has a growth constant γ. By Lemma 9.2.1 and the proof of Lemma 8.3.7

$$|\mathcal{M}(n)| = n!\gamma^{n(1+o(1))} \quad \text{and} \quad |\tilde{\mathcal{C}}_n| = n!\gamma^{n(1+o(1))}.$$

Fix $\kappa = \kappa_0$, $m = m_0$ $j = j_0$, $\mathcal{S} = \mathcal{S}_0$, $q = q_0$ and $J = J_0$ such that there is at least one construction in $\mathcal{M}(n)$ with these parameters. Then every choice of the graphs in $\{H_i\}$ yields a construction in $\mathcal{M}(n)$. In particular, writing $n_i = |V_i|$, there are in total $t_0 = \prod_{i=1}^{\kappa_0} |\mathcal{C}_{n_i}^{l+1}|$ constructions in $\mathcal{M}(n)$ with these parameters.

Note that the largest set V_j in $\mathcal{S}_0$ always contains $n' \geq n/\kappa_0 \geq n/N$ elements. It is easy to see that each pendant appearance of H in H_j yields a spike in H_j. There are at most $Ce^{-cn'}|\mathcal{C}_{n_i}^{l+1}|$ ways to choose the graph H_j so, that H_j has less than an' spikes. If H_j has more than an' spikes, then the graph G resulting from the construction has at least $an' - 2|J_0| \geq an' - 2N$ spikes, since the spikes are disjoint and each edge in J can touch at most two spikes.

Therefore there are at most

$$Ce^{-cn'}t_0 \leq Ce^{-(c/N)n}t_0$$

ways to finish the construction by choosing $H_1, \ldots, H_{\kappa_0}$, so that the resulting graph G has less than $(a/N)n - 2N$ spikes.

Since this bound holds for every $\kappa_0, m_0, j_0, \mathcal{S}_0, q_0$ and J_0, we get by the law of total probability that the number of constructions in $\mathcal{M}(n)$ that yield a graph with at most $(a/N)n - N$ spikes is at most

$$Ce^{-(c/N)n}|\mathcal{M}(n)| \leq n!e^{-(c/N)n}\gamma^{n+o(n)}.$$

So for any $a' < a/N$ and n large enough

$$\mathbb{P}(R_n \text{ has less than } a'n \text{ spikes}) \leq Ce^{-(c/N)n+o(n)} = e^{-\Omega(n)}.$$

□

Lemma 9.4.4 *Let $\mathcal{B} = \{K_4\}$, let l, r and K be positive integers, $r > l$. Then for $R_n \in_u \tilde{\mathcal{A}}$ we have*

$$\mathbb{P}(R_n \textit{ has at most } K \textit{ spikes}) \leq e^{-\Omega(n)}.$$

Proof Let $\tilde{\mathcal{A}}_1$ be the class of graphs in $\mathcal{A}$ that have at most K spikes, and let $\tilde{\mathcal{C}}_1$ be the class of graphs in $\tilde{\mathcal{C}}$ that have at most K spikes. Then $\tilde{\mathcal{A}}_1 \subseteq \mathrm{SET}(\tilde{\mathcal{C}}_1)$ and

$$\tilde{A}_1(x) \leq e^{\tilde{C}_1(x)}.$$

Using Lemma 9.4.3,

$$\overline{\gamma}(\tilde{\mathcal{A}}_1) \leq \overline{\gamma}(\tilde{\mathcal{C}}_1) < \gamma(\tilde{\mathcal{C}}) = \gamma(\tilde{\mathcal{A}}).$$

The lemma follows by (5.2). □

For an r-coloured graph G, $S_1, S_2 \subseteq [r]$, and sets Q_1, Q_2 disjoint from $V(H)$, such that $|S_i| = |Q_i|$ for $i = 1, 2$, we denote by $G^{S_1 \to Q_1, S_2 \to Q_2}$ the graph obtained by adding to G new vertices $Q_1 \cup Q_2$, and for $i = 1, 2$ adding an edge between $q_i^{(j)}$ and each vertex coloured $s_i^{(j)}$. Here $q_i^{(j)}$ and $s_i^{(j)}$ is the j-th smallest element in S_i and Q_i respectively.

Lemma 9.4.5 *Let k be a positive integer. Then*

$$|(\mathrm{Ex}\,(k+1)K_4)_n| = (1 + e^{-\Omega(n)})|(rd_{2k+1}\,K_4)_n|. \tag{9.21}$$

Proof By Lemma 8.3.9, Lemma 9.2.1, Lemma 8.4.9 and Theorem 8.1.2, there is a constant $r = r(k) > 2k$ such that all but an exponentially small fraction of graphs from $\mathrm{Ex}\,(k+1)K_4$, have a $(2k, 2, K_4)$-double blocker of size r. Here we will show that all but an exponentially small fraction of graphs in the latter class have a redundant blocker of size $2k+1$. Since each graph in $\mathrm{rd}_{2k+1} K_4$ is in the class $\mathrm{Ex}\,(k+1)K_4$, the claim will follow. We will use the idea of the proof of the main result of Chapter 7.

Fix $n \geq r$. All graphs in $(\mathrm{Ex}\,(k+1)K_4)_n$ that have a $(2k, 2, K_4)$-double blocker (and some other graphs) can be constructed as follows.

- Choose $Q \subseteq [n]$ of size r and $S \subset Q$ of size $2k$.
- Put an arbitrary graph in $\tilde{G} \in \tilde{\mathcal{A}} = \tilde{\mathcal{A}}^{<\{K_4\}, 2k, r>}$ on $[n] \setminus Q$.
- Put an arbitrary graph H on Q.
- Let $G = \tilde{G}^{\{1,\dots,2k\} \to S, \{2k+1,\dots,r\} \to Q \setminus S}$. This finishes the construction of a graph G.

Suppose $\tilde{G}$ has more than $K = r(k+2r+12)$ spikes. Then there is a (smallest) colour $q \in \{2k+1, \dots, r\}$ such that there are at least $k+2r+12$ spikes coloured $\{1, \dots, 2k, q\}$. Let x be the vertex in $Q \setminus S$ whose neighbours in G are the vertices coloured q in $\tilde{G}$. Denote $S' = S \cup \{x\}$.

Suppose G contains a $\mathcal{B}$-critical subgraph H' (that is, a subdivision of K_4) which has at most one vertex in S'. By Lemma 7.5.3, H' can touch at most $2(r + |E(K_4)|) = 2r + 12$ spikes in $\tilde{G}$. Thus there are at least k spikes that are disjoint from H. Form an arbitrary maximal matching in the set $S' \setminus V(H)$: the matching has exactly k pairs. For $y, z \in S'$ and a spike P in $\tilde{G}$, we have that $G[V(P) \cup \{y, z\}] \notin \mathrm{Ex}\, K_4$. Thus, we can produce k disjoint minors in $\mathcal{B}$ for each pair in the matching. The graph H yields $(k+1)$-st disjoint minor in $\mathcal{B}$.

Thus, whenever $\tilde{G}$ has at least K spikes and $G \in \mathrm{Ex}\,(k+1)\mathcal{B}$, we have that S' is a redundant blocker for G. So each construction G such that $G \in \mathrm{Ex}\,(k+1)K_4 \setminus \mathrm{rd}_{2k+1} K_4$ is formed by taking a graph $\tilde{G}$ with at most K spikes in the second step.

By Lemma 9.4.4, the number of choices for $\tilde{G}$, such that $\tilde{G}$ has less than K spikes is at most

$$e^{-\Omega(n)} |\mathcal{A}_{n-r}|.$$

Therefore if $\mathcal{D} = \mathrm{Ex}\,(k+1)K_4 \setminus \mathrm{rd}_{2k+1} K_4$, we have for $n \geq r$, $n \to \infty$

$$|\mathcal{D}_n| \leq \binom{n}{r} \binom{r}{2k} 2^{\binom{2k}{2}} e^{-\Omega(n)} |\tilde{\mathcal{A}}_{n-r}|$$

and $\overline{\gamma}(\mathcal{D}) < \gamma(\tilde{\mathcal{A}}) = \gamma(\mathrm{Ex}\,(k+1)K_4)$. □

Lemma 9.4.6 *Let $\mathcal{B} = \{K_4\}$, and let $l \geq 2$ be an integer. We have*

$$|(rd_l\,\mathcal{B})_{n+l}| = a_n(1 - e^{-\Omega(n)}) \quad where \quad a_n = 2^{\binom{l}{2}}\binom{n+l}{l}|\mathcal{A}_{l,n}|.$$

Proof Consider the following constructions of graphs on $[n+l]$: first pick a set $Q \subseteq [n+l]$ of size l, next take a graph $G_0 \in \mathcal{A}_l$ with $V(G_0) = [n+l] \setminus Q$ and an arbitrary graph H with $V(H) = Q$. Let $G = G_0^Q \cup H$. Each graph in $(\mathrm{rd}_l\,\mathcal{B})_{n+l}$ can be obtained in this way, so $|(\mathrm{rd}_l\,\mathcal{B})_{n+l}| \leq a_n$. We aim to bound the number of constructions that can be obtained twice, i.e., the ones which have two or more different redundant K_4-blockers of size l.

If G_0 has at least $l+1$ spikes then Q is a unique redundant blocker of size l. Indeed, if Q' is another such blocker, $Q' \neq Q$, then take a vertex $z \in Q \setminus Q'$ and any $x \in Q \setminus \{z\}$.

Now x, z and the vertices of any spike S induce a minor of K_4. Therefore, since $Q' \setminus \{x\}$ must still be a $\mathcal{B}$-blocker for G, Q' must contain a vertex from each of the $l+1$ spikes, and so $|Q'| > l$, a contradiction.

Thus every construction where Q is not the unique redundant blocker is obtained when G_0 has at most l spikes. By Corollary 9.4.2, there are at most

$$2^{\binom{l}{2}}\binom{n+l}{l}e^{-\Omega(n)}|\mathcal{A}_{l,n}| = a_n e^{-\Omega(n)}$$

such constructions, so the number of graphs in $(\mathrm{rd}_l\,\mathcal{B})_n$ that have a unique redundant blocker is at least $a_n(1 - e^{-\Omega(n)})$. □

Lemma 9.4.7 *Let $l \geq 2$ be an integer and let $\mathcal{A}_l$ be the class of coloured graphs defined for $\mathcal{B} = \{K_4\}$ in Section 8.3.3. Let $\rho_l = \rho(\mathcal{A}_l)$. Then $A_l(\rho_l) < \infty$ and there is a constant $a_l > 0$ such that*

$$|\mathcal{A}_{l,n}| = a_l n^{-5/2} n! \rho_l^{-n} (1 + o(1)).$$

Proof By the exponential formula we have $\rho(\mathcal{C}^l) = \rho_l$. We will show that

(*) C^l converges to some positive constant at ρ_l.

(**) $|\mathcal{C}^l_n| > 0$ for all $n \geq 1$.

(***) $\mathcal{C}^l$ is smooth, i.e. $|\mathcal{C}^l_{n+1}|/n|\mathcal{C}^l_n| \to \rho_l^{-1}$.

(****) For any $w = w(n) \to \infty$ we have

$$S(n,w) = \sum_{k=w}^{n-w} \binom{n}{k} |\mathcal{C}_k^l||\mathcal{C}_{n-k}^l| = o(|\mathcal{C}_n^l|).$$

Then $A_l(\rho_l) \le e^{C^l(\rho_l)} < \infty$ and Theorem 2 of Bell, Bender, Cameron and Richmond [7] yields

$$|\mathcal{A}_{l,n}| = \frac{1}{A_l(\rho_l)} |\mathcal{C}_n^l| \left(1 + o(1)\right).$$

We will use the identity (9.19) and the notation from the proof of Lemma 9.3.5. There we have shown that the class $\tilde{\mathcal{T}}$ has $\rho(\tilde{\mathcal{T}}) = \rho_l$ and by Lemma 9.3.1 $\tilde{T}(\rho_l) < \infty$. Since $\rho(\mathcal{F}_\circ)$, $\rho(\mathcal{L}_1(\mathcal{F}))$, $\rho(\mathcal{L}_2(\mathcal{F}))$, and $\rho(\mathcal{U})$ are all strictly larger than ρ_l, we have by (9.19) that $C^l(\rho_l) < \infty$. Since the coefficients of C^l are non-negative, (*) follows. The condition (**) is obvious ($\mathcal{C}^l$ includes, i.e., every uncoloured path on n vertices). Furthermore, by (9.19) and Theorem 9.3.1

$$|\mathcal{C}_n^l| = cn^{-5/2} n! \rho_l^{-n} (1 + o(1)) \tag{9.22}$$

for some constant $c > 0$, so (***) follows.

Finally, let us prove (****). Let $w = w(n) \to \infty$. We may assume $2 \le w(n) \le n/2$ for all n. By (9.22) for any $\epsilon > 0$, for all sufficiently large j we have $|\mathcal{C}_j^l| \le (c+\epsilon) j^{-5/2} j! \rho_l^{-j}$. So for n sufficiently large

$$S(n) = \sum_{k=w}^{n-w} \binom{n}{k} |\mathcal{C}_k^l||\mathcal{C}_{n-k}^l| \le (c+\epsilon)^2 \rho_l^{-n} n! \sum_{k=w}^{n-w} k^{-5/2} (n-k)^{-5/2}.$$

Now symmetry and a standard approximation of a sum by an integral gives for $w' = w - 1$

$$f(n) = \sum_{k=w}^{n-w} k^{-5/2} (n-k)^{-5/2} \le 2n^{-4} \int_{x=\frac{1}{2}}^{1-\frac{w'}{n}} x^{-5/2} (1-x)^{-5/2} dx.$$

Since for $t \in (1/2, 1)$

$$\int_{1/2}^{t} x^{-5/2} (1-x)^{-5/2} dx = -\frac{2(1 + 6t - 24t^2 + 16t^3)}{3t^{3/2}(1-t)^{3/2}}$$

we have

$$f(n) \le \frac{4(n^3 + 6w'n^2 - 24w'^2 n + 16w'^3)}{3n^4 (n-w')^{3/2} w'^{3/2}} = O\left(n^{-5/2} w^{-3/2}\right).$$

Thus $S(n) = O(|\mathcal{C}_n^l| w^{-3/2}) = o(|\mathcal{C}_n^l|)$. This completes the proof. □

We are now ready to prove Theorem 9.1.1.

Proof of Theorem 9.1.1 To replace Ω by Θ in the result of Lemma 9.4.5, note that by Lemma 8.4.12 and Theorem 7.1.2, $(\mathrm{Ex}\,(k+1)K_4)_n$ contains at least

$$|(\mathrm{apex}^k\, K_4)_n \setminus (\mathrm{rd}_{2k+1}\, K_4)_n| = n!(2^k \gamma(\mathrm{Ex}\, K_4))^{n+o(n)}$$

graphs that do not have a redundant K_4-blocker of size $2k+1$.

The theorem follows by Lemmas 9.2.1, 9.4.5, 9.4.6 and 9.4.7. □

Proof of Theorem 9.1.2 Let R'_n be a random construction as in the proof of Lemma 9.4.6, where we pick the set Q of size $2k+1$, the graph $G_0 \in \mathcal{A}_{l,n}$ and the graph H on Q uniformly at random. Then Theorem 9.1.1 and the proof of Lemma 9.4.6 imply that the total variation distance between R_n and R'_n satisfies

$$d_{TV}(R_n, R'_n) = e^{-\Theta(n)}.$$

Therefore it is enough to prove the theorem for the random graph R'_n. By Corollary 9.4.2, there is a constant $a_k > 0$, such that the graph G_0 has at least $a_k n$ spikes (and so, each vertex in Q has degree at least $a_k n$) with probability $1 - e^{-\Omega(n)}$.

Suppose G_0 has at least $a_k n$ spikes and there is a blocker Q' of R'_n and at least two distinct vertices $x, y \in Q' \setminus Q$. Then any spike and $\{x, y\}$ induces a minor K_4. Thus every such blocker must have at least $a_k n$ vertices with probability at least $1 - e^{-\Omega(n)}$. Similarly Q is with probability $1 - e^{-\Omega(n)}$ a unique redundant K_4-blocker for R_n of size $2k+1$. This finishes the proof of *(a)*.

Finally, *(b)* follows by a result of [74] (restated in a slightly more convenient form in Chapter 7). More precisely, by Lemma 7.6.2 the graph $Frag(R_n)$ obtained from R_n by removing its (lexicographically) largest component is in the class $\mathrm{Ex}\, K_4$ with probability $1 - e^{-\Omega(n)}$. The class $\mathrm{Ex}\,(k+1)K_4$ is bridge-addable and by Theorem 9.1.1 it is smooth and $\rho = \gamma(\mathrm{Ex}\,(k+1)K_4)^{-1} > 0$. Therefore by [74], see Lemma 7.6.3, $Frag(R_n)$ converges in total variation to the "Boltzmann-Poisson" random graph with parameters $\mathcal{B}$ and ρ, in particular $\mathbb{P}(|V(Frag(R_n))| = 0) \to p_k = A(\rho)^{-1}$. □

9.5 The class $\mathrm{Ex}\,(k+1)\{K_{2,3}, K_4\}$

The class $\mathrm{Ex}\,(k+1)\{K_{2,3}, K_4\}$ is a subclass of $\mathrm{Ex}\,(k+1)K_4$. Even though it does not satisfy the condition of Theorem 8.1.2, we can still adapt most of the techniques from the preceding sections (the proofs are simpler for this case).

9.5.1 Coloured cores and paths

In this section we fix $\mathcal{B} = \{K_{2,3}, K_4\}$. To avoid an additional index, we will accordingly write $\mathcal{C}^l = \mathcal{C}^{l,\mathcal{B}}$, $\mathcal{A}_l = \mathcal{A}_{l,\mathcal{B}}$, and assume that the definitions such as "good colour", "nice vertex" and "nice graph" are with respect to $\mathcal{B} = \{K_{2,3}, K_4\}$.

Lemma 9.5.1 *Let $\mathcal{B} = \{K_{2,3}, K_4\}$, let l be a positive integer and let $G \in \mathcal{C}^l$ be nice. The nice core tree of G is a path.*

Proof In Section 9.3.2 we showed that the nice core tree T of $G \in \mathcal{C}^{l,\mathcal{B}} \subset \mathcal{C}^{l,\{K_4\}}$ is a tree. If T has at least three leaves, we can produce a minor $K_{2,3}$ by adding a new vertex connected to each of these leaves, thus every colour is bad for G. □

Let $\tilde{\mathcal{D}}$ be the class of all biconnected outerplanar networks G such that adding a new vertex connected to each pole of G gives an outerplanar graph. Let $G \in \tilde{\mathcal{D}}$ be 2-connected. Then G has a unique Hamilton cycle H (see, e.g., [12]). The poles s and t of G must be neighbours in H, otherwise adding a new vertex connected to each pole yields a minor $K_{2,3}$. Conversely, if we take an arbitrary 2-connected outerplanar graph G and pick an oriented edge st from its Hamilton cycle H, the graph with source s and sink t obtained from G is from the class $\tilde{\mathcal{D}}$: this follows by the relation of 2-connected outerplanar graphs and polygon dissections [12]. Therefore from each 2-connected rooted outerplanar graph we can obtain two networks in $\tilde{\mathcal{D}}$ (the root becomes the source and either the left or the right neighbour of the root on the Hamilton cycle becomes the sink) with $n-2$ vertices. It follows that

$$\tilde{D}(x) = \frac{2B(x)}{x^2} - 1,$$

where $B(x)$ is the exponential generating function of rooted biconnected outerplanar graphs (which contains K_2). Bernasconi, Panagiotou and Steger [12] show that

$$B(x) = \frac{1}{2}(D(x) + x^2)$$

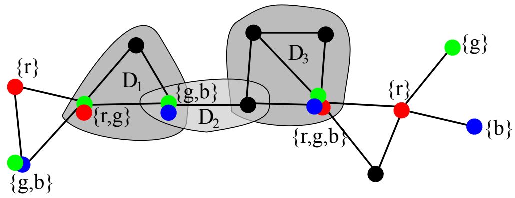

Figure 9.8: Coloured core of a graph in $\mathcal{C}^3$ in the case $\mathcal{B} = \{K_{2,3}, K_4\}$. Replacing the three grey networks (which belong to the class $\tilde{\mathcal{D}}$) with edges we obtain its nice core tree (a path of length three). The graphs attached to the endpoints of this path belong to the class $\mathcal{L}$.

where $D(x)$ is the exponential generating function for polygon dissections. Thus $\tilde{D}(x) = D(x)/x^2$ and we get by (4.1) of [12]

$$\tilde{D}(x) = \frac{1}{4x}\left(1 + x - \sqrt{x^2 - 6x + 1}\right). \tag{9.23}$$

Solving quadratic equations and using the "first principle" from [48] we get that

$$\rho(\tilde{\mathcal{D}}) = 3 - 2\sqrt{2} \quad \text{and} \quad \rho(\tilde{D})\tilde{D}(\rho(\tilde{\mathcal{D}})) = 1 - \frac{\sqrt{2}}{2} = 0.292\ldots \tag{9.24}$$

Lemma 9.5.2 *Let $\mathcal{B} = \{K_{2,3}, K_4\}$, let $l \geq 2$ be an integer, and let $\tilde{\mathcal{C}}^l$ be the class of coloured cores of nice graphs in $\mathcal{C}^l$. Then the class $\tilde{\mathcal{C}}^l$ has exponential generating function*

$$\tilde{C}^l(x) = \frac{2^{l-1}xL(x)^2}{1 - 2^l x\tilde{D}(x)}, \tag{9.25}$$

where L is the exponential generating function of a class $\mathcal{L} = \mathcal{L}^{<l>}$ with $\rho(\mathcal{L}) \geq \rho(\tilde{\mathcal{C}}^{l-1})$.

Proof Let $\mathcal{L} = \mathcal{L}^{<l>}$ be the class of all pointed $\{0,1\}^l$-coloured connected graphs G satisfying the following conditions: a) each colour is good for G; b) $G - r$ is connected, $\mathrm{Col}(G - r) = [l]$ and $\mathrm{Col}(r) = \emptyset$, where $r = r(G)$ is the root of G; c) for any $x \in V(G)$, each component of $G - x$ contains at least one colour or r, and at most one component contains all colours $[l]$ or r. The class $\mathcal{L}$ corresponds to the class $\mathcal{L}_2$ from Section 9.3.2.

We will now prove that $\rho(\mathcal{L}) \geq \rho(\tilde{\mathcal{C}}^{l-1})$. We first claim that for each positive integer i and each $n = 0, 1, \ldots,$

$$|\tilde{\mathcal{C}}_n^i| \geq |\tilde{\mathcal{C}}_n^{i-1}|. \tag{9.26}$$

Indeed, for any graph in $G \in \tilde{\mathcal{C}}_n^i$ we can assign a graph $G' \in \tilde{\mathcal{C}}_n^{i+1}$ by picking the lexicographically minimal pair of vertices coloured red and adding colour $i+1$ to their colour sets. Such a pair always exists since G is nice, and since the colour red is good for G, the colour $i+1$ must be good for G'.

Let $\mathcal{C}$ be the class of all pointed graphs H, such that

1. $\text{Col}(H) \subset [l]$;
2. if we add to H a new vertex w coloured $\text{Col}(H)$ and connected to the root of H and label the root of H arbitrarily, we obtain a graph in $\tilde{\mathcal{C}}^l$.

The coefficients of $C(x)$ are dominated by those of $\sum_{j\in[l-1]} \binom{l}{j} x^2 \tilde{C}^j(x)'$, therefore $\rho(\mathcal{C}) \geq \min_{j\in[l-1]} \rho(\tilde{\mathcal{C}}^j) \geq \rho(\tilde{\mathcal{C}}^{l-1})$. The last inequality follows using (9.26). Recall also that we denote by $\mathcal{Z}_\emptyset$ the class of graphs consisting of a single pointed uncoloured vertex only.

Now let $G \in \mathcal{L}$. G can be constructed as follows: take a pointed biconnected outerplanar graph B (B is the block of G containing the root with its colours removed), colour $t \leq 2l$ of its (non-root) vertices $v_1, \ldots, v_t$ with some subsets of $[l]$ and identify v_i with the root of a pointed graph $G_i \in \mathcal{C} \cup \mathcal{Z}_\emptyset$ (so that labels of G, $G_1, \ldots, G_t$ are disjoint). To verify the last statement, note first that the number of cut vertices and coloured vertices in B is bounded by $2l$, since each component of $G - E(B)$ can contain at most two components containing colour c for each $c \in [l]$ (see Section 9.2.2). Secondly, suppose u is a cut vertex of G, and let G_u the graph obtained from the component of $G - E(B)$ containing u by pointing the vertex u. The graph H obtained from G by contracting all vertices in $G - G_u$ into a single vertex w, setting $\text{Col}_H(w) = \text{Col}_G(G_u - u) \subset [l]$ and $\text{Col}_H(u) = \emptyset$ is a coloured minor of G: this shows that indeed $G_u \in \mathcal{C}$.

Let $\mathcal{B}_\circ$ denote the class of pointed biconnected outerplanar graphs. By the properties of biconnected outerplanar graphs discussed above in this section we see that $\rho(\mathcal{B}_\circ) = \rho(\tilde{\mathcal{D}}) \geq \rho(\tilde{\mathcal{C}}^1)$. (The last inequality follows since from each graph in $\tilde{\mathcal{D}}$ we may obtain a graph in $\tilde{\mathcal{C}}^1$ by labelling its poles and colouring them $\{red\}$.)

The above observations imply that the coefficients of $L(x)$ are bounded by the coefficients of

$$\sum_{t=0}^{2l} x^t B_\circ(x)^{(t+1)} (2^l(1 + C(x)))^t.$$

Therefore $\rho(\mathcal{L}) \geq \min(\rho(\mathcal{C}), \rho(\mathcal{B}_\circ)) \geq \rho(\tilde{\mathcal{C}}^{l-1})$.

Now let $\tilde{\mathcal{C}}^{l,1}$ denote the subclass of graphs in $\tilde{\mathcal{C}}^l$ with the nice tree of size 1, and let $\tilde{\mathcal{C}}^{l,2} = \tilde{\mathcal{C}}^l \setminus \tilde{\mathcal{C}}^{l,1}$. Using Lemma 9.5.1 and Section 9.3.2, each graph in $\tilde{\mathcal{C}}^{l,2}$ can be obtained as a series composition of $r \geq 1$ outerplanar networks $D_1, \ldots, D_r$, where only the poles of these networks can be coloured (with arbitrary colours in $2^{[l]}$), by attaching two rooted graphs L', L'' at the source of D_1 and the sink of

D_r respectively (attaching pendant coloured graphs to vertices corresponding to internal vertices of the nice core path is forbidden, otherwise some colour would be bad for the resulting graph), see Figure 9.8. Let $c \in [l]$. Since each pole of a network D_i is nice, there are two disjoint paths from each of the poles ending with a vertex coloured c. This shows that $D_i \in \tilde{\mathcal{D}}$.

It is easy to see that L', L'' must always be from the class $\mathcal{L}$. On the other hand, every such construction gives a valid graph in $\tilde{\mathcal{C}}^{l,2}$, and every graph in $\tilde{\mathcal{C}}^{l,2}$ is obtained exactly twice (reversing each network and their order in the sequence and swapping L' with L'' gives the same graph).

Therefore

$$2 \times \tilde{\mathcal{C}}^{l,2} = (2^l \times \mathcal{Z} \times \mathcal{L})^2 \times \tilde{\mathcal{D}} \times \mathrm{SEQ}((2^l \times \mathcal{Z}) \times \tilde{\mathcal{D}}).$$

Similarly

$$\tilde{\mathcal{C}}^{l,1} = 2^l \times \mathcal{Z} \times \mathrm{SET}_2(\mathcal{L}).$$

Finally, the identity $\tilde{\mathcal{C}}^l = \tilde{\mathcal{C}}^{l,1} + \tilde{\mathcal{C}}^{l,2}$ and a standard conversion to exponential generating functions [48] yields (9.25). □

Lemma 9.5.3 *Let $l \geq 2$ be an integer and let $\mathcal{L}^{<L>}, \tilde{\mathcal{C}}^l$ be as in Lemma 9.5.2. We have $\rho(\mathcal{L}^{<l>}) > \rho(\tilde{\mathcal{C}}^l) = r_l$, where*

$$r_l = \frac{1}{2^l}\left(1 - \frac{1}{2^l - 1}\right).$$

There are constants $R_1 = R_1(l) > r_l$ and $c_l' > 0$ and a function $g_1(x) = g_{1,l}(x)$ analytic for $x \in \mathbb{C}$ with $|x| < R_1$ such that

$$\tilde{C}^l(x) = \frac{c_l'}{1 - x/r_l} + g_1(x). \tag{9.27}$$

Proof A simple calculation shows that the unique solution of $2^l x\tilde{D}(x) = 1$ is r_l. Notice also that since $\tilde{\mathcal{D}}$ contains a network isomorphic to K_2 and a network isomorphic to an arbitrary cycle, we have $[x^n]2^l x\tilde{D}(x) > 0$ for $n = 1, 2, \ldots$ (that is, $2^l x\tilde{D}(x)$ is strongly aperiodic, see [48]).

We will also use that $r_2 = \frac{1}{6}$ and $r_{j+1} < r_j$ for any integer $j = 2, 3, \ldots$. Define

$$c_l' = \frac{r_l L(r_l)^2}{2(\tilde{D}(r_l) + r_l\tilde{D}'(r_l))}.$$

We prove the claim by induction on l. First consider the case $l = 2$. By Lemma 9.5.2 $\rho(\mathcal{L}^{<2>}) \geq \rho(\tilde{\mathcal{C}}^1) = \rho(\tilde{\mathcal{D}}) = 3 - 2\sqrt{2} > \frac{1}{6}$. We can write $\tilde{C}^2(x) = h(x)f(x)$ where $f(x) = (1 - 2^2 x\tilde{D}(x))^{-1}$ and $h(x) = 2^2 x L^{<l>}(x)^2$. By (9.24), $f(x)$ corresponds to a supercritical sequence schema and $h(x)$ is analytic in $\Delta = \{x \in \mathbb{C} : |x| < R_1\}$ for some $R_1 > r_2$, $h(r_2) > 0$.

We get using Theorem V.1 of [48], its proof and properties of meromorphic functions that $\tilde{C}^2$ is meromorphic and has only one pole r_2 (which is simple) in Δ, where it satisfies (9.27).

The proof of the general case $l \geq 3$ follows similarly, since using Lemma 9.5.2 and induction we have $\rho(\mathcal{L}^{<l>}) \geq \rho(\tilde{\mathcal{C}}^{l-1}) = r_{l-1} > r_l$, $2^l \rho(\tilde{\mathcal{D}})\tilde{D}(\rho(\tilde{\mathcal{D}})) > 1$ and the convergence radius of $2^l x\tilde{D}(x)$ is $\rho(\tilde{\mathcal{D}}) > r_2 > r_l$. □

9.5.2 Proof of Theorem 9.1.3

It remains to collect and combine the analytic results for classes related with Ex $\mathcal{B}$. Denote by $\mathcal{F}$ the class of connected rooted outerplanar networks (we reuse the symbol from the previous section). [32] show that the functional inverse of its exponential generating function F is

$$\psi_F(u) = ue^{\frac{1}{8}\left(\sqrt{1-6u+u^2}-5u-1\right)}. \tag{9.28}$$

Lemma 9.5.4 *Let $l \geq 2$ and let $\mathcal{B} = \{K_{2,3}, K_4\}$. Define $\sigma_l = \psi_F(r_l)$, where r_l is as in Lemma 9.5.3. There are constants $c_l > 0$, $R > \sigma_l$ and a function $g(x)$ analytic in $\{z \in \mathbb{C} : |x| < R\}$ such that*

$$C^l(x) = \frac{c_l}{1 - x/\sigma_l} + g(x).$$

Proof The class $\bar{\mathcal{C}} = \bar{\mathcal{C}}^{<l>}$ of nice graphs in $\mathcal{C}^l$ has exponential generating function

$$\bar{C}(x) = \tilde{C}(F(x)).$$

By [32], $\tau = F(\rho(\mathcal{F}))$ is the smallest positive solution of $3u^4 - 28u^3 + 70u^2 - 58u + 8 = 0$, and a numeric evaluation yields that $\tau = 0.170\cdots > 1/6 \geq r_l$. Furthermore, clearly $|\mathcal{F}_n| > 0$ for $n = 1, 2, \ldots$. Thus by 9.28 $\sigma_l = \psi_F(r_l)$ is the smallest positive solution of $F(x) = r_l$ and the unique dominant singularity of $\bar{C}(x)$.

Using Lemma 9.5.3 we get

$$\bar{C}(x) = \frac{c_l'}{1 - F(x)/r_l} + g_1(F(x))$$

and $g_1(x)$ is analytic for $|x| < R_1$ where $R_1 > r_l$. Since F has convergence radius larger than σ_l, there is $\epsilon > 0$, such that $F(x) < R_1$ for $x \in (0, \sigma_l + \epsilon)$. By the triangle inequality, for any $t \in \mathbb{C}$, $|t| < \sigma_1 + \epsilon$ we have $|F(t)| \leq F(|t|) < R_1$ so $g_1(F(x))$ is analytic at $x = t$.

Now applying the supercritical composition schema (Theorem V.1 of [48]) to the function $c_l'(1-F(x)/r_l)^{-1}$, we see that there is $R_2 \in (\sigma_l, \sigma_l+\epsilon)$ such that $\bar{C}(x)$ satisfies for $x \in \Delta := \{z \in \mathbb{C}, |z| < R_2\}$

$$\bar{C}(x) = \frac{c_l}{1 - x/\sigma_l} + g_2(x), \quad c_l = \frac{r_l c_l'}{\sigma_l F'(\sigma_l)}$$

for some function $g_2(x)$ which is analytic in Δ.

Finally, to obtain $C^l(x)$ we have to add to $\bar{C}(x)$ the exponential generating function $U^{<l>}(x)$ of graphs in $\mathcal{C}^l$ that are not nice. An argument analogous to the one presented in the proof of Lemma 8.4.3 shows that $\rho(\mathcal{U}^{<l>}) \geq \rho(\mathcal{C}^{l-1})$. Furthermore, [32] showed that the convergence radius of the class of outerplanar graphs $\rho(\text{Ex}\,\mathcal{B}) = 0.1365...$ Now in the case $l = 2$ the lemma follows, since $\rho(\mathcal{U}^2) \geq \rho(\mathcal{C}^1) \geq \rho(\text{Ex}\,\mathcal{B}) > \sigma_2 = 0.1353..$, so $U^{<2>}(x)$ is analytic at any t with $|t| < R := \min(\rho(\text{Ex}\,\mathcal{B}), R_2)$. Since $(r_l, l = 1, 2, \dots)$ is strictly decreasing and ψ_F is increasing for $x \in (0, \rho(\mathcal{F}))$, we have that $(\sigma_l, l = 1, 2, \dots)$ is strictly decreasing. Therefore have that $\rho(\mathcal{U}^{<l>}) \geq \rho(\mathcal{C}^{l-1}) = \sigma_{l-1} > \sigma_l$ by induction, and the lemma follows similarly as in the case $l = 2$. □

Lemma 9.5.5 *Let $l \geq 2$ be an integer and let $\mathcal{B} = \{K_{2,3}, K_4\}$. Let c_l, σ_l and g be as in Lemma 9.5.4. Then*

$$|\mathcal{C}_n^l| = c_l n! \sigma_l^{-n}(1 + o(1))$$

and

$$|\mathcal{A}_{l,n}| = b_l n^{-3/4} e^{2(c_l n)^{1/2}} n! \sigma_l^{-n}(1 + o(1))$$

where

$$b_l = \frac{c_l^{1/4} e^{c_l/2 + g(\sigma_l)}}{2\pi^{1/2}}.$$

Proof The lemma follows by Lemma 9.5.4 and Proposition 23 of [31]. □

Proof of Theorem 9.1.3 By [32], there are computable constants h and γ, with $\gamma^{-1} = 0.1365..$ such that the number of outerplanar graphs on vertex set $[n]$ is

$$hn^{-3/2}\gamma^n n!(1 + o(1)).$$

Using Lemma 9.5.4 and Theorem 7.1.2

$$\gamma(\mathrm{rd}_3\,\mathcal{B}) = \sigma_3^{-1} = 10.482.. < \gamma(\mathrm{apex}\,(\mathrm{Ex}\,\mathcal{B})) = 2\gamma = 14.642..$$

Therefore by Theorem 8.1.1 we have

$$|(\mathrm{Ex}\,2\mathcal{B})_n| = |(\mathrm{apex}\,(\mathrm{Ex}\,\mathcal{B}))_n|(1 + e^{-\Theta(n)}),$$

and using Theorem 7.1.2

$$|(\mathrm{Ex}\,2\mathcal{B})_n| = \frac{h}{2\gamma} n^{-3/2} n!(2\gamma)^n(1 + o(1)).$$

Now

$$\gamma(\mathrm{rd}_5\,\mathcal{B}) = \sigma_5^{-1} = 34.099..$$

and $\overline{\gamma}(\mathrm{apex}\,(\mathrm{Ex}\,2\mathcal{B})) \leq 4\gamma = 29.2.. < \gamma(\mathrm{rd}_5\,\mathcal{B})$. By Lemma 8.3.9 and Lemma 9.5.5

$$\gamma'_k = \gamma(\mathrm{Ex}\,(k+1)\mathcal{B}) = \gamma(\mathrm{rd}_{2k+1}\,\mathcal{B}) < \gamma(\mathrm{apex}^k\,(\mathrm{Ex}\,\mathcal{B}))$$

for each $k = 2, 3, \ldots$.

Furthermore, using Lemma 9.5.5 there are constants $b_{2k+1}, c_{2k+1} > 0$ such that

$$\begin{aligned}
|(\mathrm{Ex}\,(k+1)\mathcal{B})_{n+2k+1}| &\geq 2^{\binom{2k+1}{2}}|\mathcal{A}_{2k+1,n}| \\
&= 2^{\binom{2k+1}{2}} b_{2k+1} n^{-3/4} \exp\left(2(c_{2k+1}n)^{1/2}\right) n!(\gamma'_k)^n(1 + o(1)) \\
&= e^{\Omega(\sqrt{n+2k+1})}(n + 2k + 1)!(\gamma'_k)^{n+2k+1}.
\end{aligned}$$

Finally, the values of $\gamma'_k = \sigma_{2k+1}^{-1}$ for $k = 2, 3, \ldots$ can be obtained using the closed-form expression $\sigma_l = \psi_F(r_l)$. □

9.6 Concluding remarks

As the length of Chapters 8 and 9 indicates, analysis of classes without $k + 1$ disjoint minors in $\mathcal{B}$ becomes more involved as the excluded minors get more complicated. In the last two chapters we concentrated on families of sets $\mathcal{B}$, not covered by the results of Chapter 7: we found that indeed the highest-level structure of typical graphs in such cases may obey a different pattern.

There are a few possible directions of further research. One can conjecture that for $\mathcal{B}$ as in Theorem 8.1.1, and perhaps for more general $\mathcal{B}$, all but an exponentially small proportion of graphs in $(\mathrm{Ex}\,(k+1)\mathcal{B})_n$ belong to one of the classes $\mathrm{rd}_{2k+1}\,\mathcal{B}$ or $\mathrm{apex}^k\,(\mathrm{Ex}\,\mathcal{B})$. For certain $\mathcal{B}$ our results imply part of this conjecture, and we gave specific examples where this conjecture holds. To advance it further, one would need to develop a general way of comparing growth constants for two or more candidate subclasses. It is not clear whether this can be done without knowing the specific structure and generating functions for $\mathrm{rd}_{2k+1}\,\mathcal{B}$.

It seems plausible, that for classes $\mathcal{A}$ with $\mathrm{aw}_2\,(\mathcal{A}) \leq j$ an analogue of Theorem 1.2 is true with a more general kind of redundant blockers. One can go even further and formulate conjectures as in Section 7.7 about classes $\mathrm{Ex}\,(k+1)\mathcal{B}$ in the case when $\mathrm{Ex}\,\mathcal{B}$ contains all j-fans, $j \geq 2$, but not all $(j+1)$-fans. Yet another level of complexity would be to obtain any results in the case when $\mathcal{B}$ does not contain a planar graph.

In Section 9.2 we proved decompositions for the class $\mathrm{rd}_l\,K_4$ for general $l = 1, 2, \ldots$. In the following table we present growth constants for l up to 5 obtained automatically with the help of Maple (and a simple program to enumerate graphs in $\mathcal{T}_l'$). We explicitly proved validity of the numerical estimates up to $l = 3$ in this paper.

l	$\gamma(\mathrm{rd}_l\,K_4)$	Comment	$\gamma(\mathrm{rd}_l\,K_4)/(2^l e)$
1	9.073311..	$= \gamma(\mathrm{Ex}\,K_4)$, [32]	1.67..
2	12.677273..		1.17..
3	23.524122..	$= \gamma(\mathrm{Ex}\,2K_4)$	1.08..
4	45.5488..		1.05..
5	89.5511..	$= \gamma(\mathrm{Ex}\,3K_4)$	1.03..

The last column shows the ratio of the growth constant of $\mathrm{rd}_l\,K_4$ and the growth constant of the class $\mathrm{apex}^l\,(\mathrm{Ex}\,K_3)$ (see Chapter 7), where $\mathrm{Ex}\,K_3$ is the class of forests of labelled trees. Not surprisingly, the numerical estimates indicate that this ratio approaches 1 as l increases. A similar situation can be observed with the ratio $\gamma(\mathrm{rd}_l\,\{K_{2,3}, K_4\})/2^l$. This prompts the following questions: is it possible that for some $k = k(n) \to \infty$, a typical graph from $(\mathrm{Ex}\,(k+1)K_4)_n$ consists of a forest and $2k+1$ apex vertices with probability $1 - o(1)$? Can this be generalised?

Conclusion

Results of Part I

1. In Chapter 2 we considered sequences of random intersection graphs $\{D(n)\}$ where $D(n) = D(n, m, p_-, p_+)$ and $m = m(n)$, $p_- = p_-(n)$ and $p_+ = p_+(n)$. We defined the birth threshold function τ such that $\tau(n, m, p_-, p_+) \to \infty$ (respectively, 0) implies that $D(n)$ contains (respectively, does not contain) a copy of the complete directed graph $\overrightarrow{K}_h$ whp. Next, we introduced the notion of a diclique cover of a digraph. We showed that there are several possible cases of the relationship of the parameters m, p_- and p_+, and to each case corresponds one or more simple diclique covers. The "in-star" and the "out-star" covers that realise the birth threshold when p_- is much larger than p_+ (respectively, p_+ much larger than p_-), were not possible in the undirected case.

2. In Chapter 3 we considered sequences $\{G(n)\}$ of sparse 'active' random intersection graphs $G(n) = G(n, m, P)$, where $m = m(n)$, $P = P(n)$. We introduced a power-law tail condition (3.6) for the normalized random subset size $Y(n) = \sqrt{\frac{n}{m}}X(n)$, where $X(n)$ is distributed according to $P(n)$. We determined the asymptotic clique number in $G(n)$ (it is polynomial in n) when $Y(n)$ satisfies this condition with index $\alpha \in (1, 2)$ for a wide range of sequences $m = m(n)$, including the case $m = \Theta(n)$ that yields a non-vanishing clustering coefficient. Secondly, we considered the case where $G(n)$ is sparse and $\mathbb{E}\,Y = \Theta(1)$ and $VarY = \Theta(1)$. In this case we showed that the largest clique in $G(n)$ is monochromatic (plus possibly a stochastically bounded number extra vertices) and we proved that the total variation distance of $\omega'(G(n))$ and the size of the maximum bin when $(mn)^{1/2}\mathbb{E}\,Y(n)$ balls are thrown into m bins tends to zero. Thirdly, we described algorithms to find a clique of asymptotically optimal size in each of the above cases, and showed their correctness and efficiency. Finally, we proved a technical result on the relation of $Y(n)$ and the degree of a random vertex of $G(n)$.

3. In Chapter 4 we introduced a randomized greedy algorithm for colouring edges of the random uniform hypergraph $\mathbb{H}^{(k)}(n, m)$ and proved that there is a constant c_ϵ such that if $k \geq 2$, $k \leq c_\epsilon \ln\left(\frac{n}{\ln \bar{d}}\right)$ and $k \leq c_\epsilon \ln\left(\frac{\bar{d}}{\ln n}\right)$ then the algorithm properly colours the edges of $\mathbb{H}^{(k)}(n, m)$ with $\lceil \bar{d}(1+\epsilon) \rceil$ colours and probability at least $1 - \frac{2}{n} - \frac{2}{\bar{d}}$. For a sequence $\{H(n), n = n_0, n_0+1, \dots\}$ where $k = k(n)$, $m = m(n)$ and $H(n) = \mathbb{H}^{(k)}(n, m)$ satisfies the above condition this yields $\chi'(H(n)) = \bar{d}(1 + o_P(1))$.

4. In Section 1.2 we presented plots illustrating that parameters in real-world networks can, with interesting exceptions, be matched closely with those in random intersection graphs. This direction requires further research.

To sum up, we determined the behaviour of several important parameters in random intersection graphs. Notably, we made progress in the most practically relevant regime of sparse graphs with positive clustering.

Results of Part II

1. In Chapter 6 we proved that $|(\mathrm{Ex}\,(k+1)K_3)_n| = (1-e^{-\Omega(n)})|(\mathrm{apex}^k\,\mathcal{F})_n|$. Using this, we obtained precise asymptotic counting formula for graphs without $k+1$ disjoint cycles. We showed that with probability $1-e^{-\Omega(n)}$ a uniformly random graph R_n from $(\mathrm{Ex}\,(k+1)K_3)_n$ contains a unique vertex feedback set (blocker) of size k, we determined the asymptotic probability that R_n is connected and investigated the asymptotic distribution of the number of components, chromatic and clique numbers of R_n.

2. In Chapter 7 we generalized results of Chapter 6 and proved that $|(\mathrm{Ex}\,(k+1)\mathcal{B})_n| = (1-e^{-\Theta(n)})|(\mathrm{apex}^k\,\mathcal{A})_n|$, as long as the class $\mathcal{A} = \mathrm{Ex}\,\mathcal{B}$ is addable and does not contain all fans. We showed that this implies that such a class $\mathrm{Ex}\,(k+1)\mathcal{B}$ has a growth constant $2^k\gamma(\mathcal{A})$, i.e., the answer to the question of Bernardi, Noy and Welsh in this case is positive. We expressed asymptotics of $|(\mathrm{Ex}\,(k+1)\mathcal{B})_n|$ in terms of $|\mathcal{A}_n|$. Next, we showed that with probability $1-e^{-\Omega(n)}$ a random graph $R_n \in_u \mathrm{Ex}\,(k+1)\mathcal{B}$ contains a unique $\mathcal{B}$-blocker S of size k, such that each vertex in the blocker has a linear degree. We also generalized proofs of other asymptotic properties (connectivity, components, clique and chromatic number, etc.).

3. In Chapter 8 we considered addable classes $\mathcal{A} = \mathrm{Ex}\,\mathcal{B}$ such that $\mathcal{A}$ contains all fans, but not all 2-fans, nor all bipartite graphs $K_{3,t}$. We showed that there is a constant k_0, such that for $k \ge k_0$, $\overline{\gamma}(\mathrm{Ex}\,(k+1)\mathcal{B}) = \overline{\gamma}(\mathrm{rd}_{2k+1}\,\mathcal{B})$ and for a subsequence n_l realising this upper limit, a random graph $R_{n_l} \in_u \mathrm{Ex}\,(k+1)\mathcal{B}$ has no $\mathcal{B}$-blocker of size smaller than $2k$ with probability $1-e^{-\Omega(n_l)}$ as $l \to \infty$. We also proved that if we add a further condition that the minimal excluded minors are 3-connected and $\mathcal{A}$ does not contain all wheels, then $\mathrm{Ex}\,(k+1)\mathcal{B}$ has a growth constant. To obtain these results we proved two non-trivial graph-theoretical lemmas.

4. In Chapter 9 we proved that for $k = 1, 2, \ldots$ there are constants c_k, γ_k such that $|(\mathrm{Ex}\,(k+1)K_4)_n| = (1 - e^{-\Omega(n)})|(\mathrm{rd}_{\,2k+1}\,K_4)_n| \sim c_k n^{-3/2}\gamma_k^n n!$. We proved that $R_n \in_u \mathrm{Ex}\,(k+1)K_4$ whp has a unique redundant blocker of size $2k+1$, and each vertex in this blocker has a linear degree. Along the way we obtained decompositions for classes related to $\mathrm{rd}_{\,2k+1}\,K_4$ and proved a lemma for enumerating trees where leaves, internal vertices and edges are replaced with objects of different type. Lastly, we considered class $\mathrm{Ex}\,(k+1)\{K_{2,3}, K_4\}$ and showed that it behaves very differently.

The work explores a new subarea of asymptotic enumeration, i.e., counting graphs with few disjoint excluded minors. We saw that such classes are combinatorially tractable and have an interesting structure. We made progress on two rather general families of disjoint excluded minors, though an infinite number of unresolved important cases (i.e., no two disjoint minors K_5) remain.

Bibliography

[1] N. Alon, T. Jiang, Z. Miller and D. Pritkin, Properly colored subgraphs and rainbow subgraphs in edge-colorings with local constraints, *Random Struct. Algorithms* **23** (2003), 409–433.

[2] N. Alon and J. H. Kim, On the degree, size, and chromatic index of a uniform hypergraph, *J. Combin. Theory Ser. A* **77** (1997), 165–170.

[3] M. Beck, G. Marchesi, D. Pixton and L. Sabalka, A First Course in Complex Analysis, 2012, online lecture notes, available at http://math.sfsu.edu/beck/complex.html

[4] J. Balogh, T. Bohman and D. Mubayi. Erdős - Ko - Rado in random hypergraphs. *Combinatorics, Probability and Computing* **18** (2009), 629–646.

[5] M. Behrisch, Component evolution in random intersection graphs, *The Electron. J. Combin.* **14** (2007).

[6] M. Behrisch, A. Taraz and M. Ueckerdt, Colouring random intersection graphs and complex networks. *SIAM J. Discrete Math.* **23** (2009), 288–299.

[7] J. P. Bell, E. A. Bender, P. J. Cameron and L. B. Richmond. Asymptotics for the probability of connectedness and the distribution of number of components, *Electron. J. Combin.* **7** (2000) R33.

[8] E. A. Bender and Z. Gao. Asymptotic enumeration of labelled graphs with a given genus. *Electron. J. Combin.* **18** (2011) #P13.

[9] E. A. Bender, Z. Gao, N. C. Wormald, The number of 2-connected labelled planar graphs, *Electron. J. Combin.* **9** (2002), R43.

[10] C. Berge, On two conjectures to generalize Vizing's Theorem, *Le Matematiche* **45** (1990).

[11] O. Bernardi, M. Noy, and D. Welsh, Growth constants of minor-closed classes of graphs, *J. Combin. Theory B* **100** (2010) 468–484.

[12] N. Bernasconi, K. Panagiotou, A. Steger: On the Degree Sequences of Random Outerplanar and Series-Parallel Graphs, APPROX-RANDOM (2008), 303–316.

[13] G. Bianconi and M. Marsili, Emergence of large cliques in random scale-free networks. *Europhys. Lett.* **74** (2006), 740–746.

[14] S. R. Blackburn and S. Gerke, Connectivity of the uniform random intersection graph, *Discr. Math.* **309** (2009) 5130–5140.

[15] M. Bloznelis, Degree distribution of a typical vertex in a general random intersection graph, *Lith. Math. J.* **48** (2008) 38–45.

[16] M. Bloznelis, A random intersection digraph: Indegree and outdegree distributions, *Discr. Math.* **310** (2010) 2560–2566.

[17] M. Bloznelis, Component evolution in general random intersection graphs, *SIAM J. Discr. Math.* **24** (2010) 639–654.

[18] M. Bloznelis, Degree and clustering coefficient in sparse random intersection graphs, *Ann. Appl. Probab.* **23** (2013) 1254–1289.

[19] M. Bloznelis and J. Damarackas, Degree distribution of an inhomogeneous random intersection graph, arXiv:1212.6402 [math.PR] (2012).

[20] M. Bloznelis, J. Jaworski and V. Kurauskas, Assortativity and clustering in sparse random intersection graphs. *Electron. J. Probab.* **18** (2013), 1–24.

[21] M. Bloznelis, J. Jaworski, E. Godehardt, V. Kurauskas and K. Rybarczyk, Recent progress in network models – the models, 2013, in preparation.

[22] M. Bloznelis, J. Jaworski, E. Godehardt, V. Kurauskas and K. Rybarczyk, Recent progress in network models – the results, 2013, in preparation.

[23] M. Bloznelis and M. Karoński, Random intersection graph process, arXiv:1301.5579 [math.PR] (2012).

[24] M. Bloznelis and V. Kurauskas, Clustering function: a measure of social influence, arXiv:1207.4941 [math.PR] (2012), submitted.

[25] M. Bloznelis and V. Kurauskas, Large cliques in sparse random intersection graphs, arXiv:1302.4627 [math.CO] (2013), submitted.

[26] E. Birmele, J.A. Bondy and B.A. Reed, The Erdős-Pósa property for long circuits, *Combinatorica* **27** (2) (2007) 135–145.

[27] M. Bradonjic, A. A. Hagberg, N. W. Hengartner, N. Lemons, A. G. Percus, The phase transition in inhomogeneous random intersection graphs, CoRR abs/1301.7320 (2013).

[28] B. Bollobás, The chromatic number of random graphs. Combinatorica 8 (1988), 49–56.

[29] B. Bollobás and P. Erdős, Cliques in random graphs. Math. Proc. Camb. Phil. Soc. 80 (1976), 419–427.

[30] T. Bohman, The triangle-free process, *Adv. Math.* **221** (2009), 1653–1677.

[31] M. Bousquet-Mélou and K. Weller, Asymptotic properties of some minor-closed classes of graphs, arXiv:1303.3836 (2013).

[32] M. Bodirsky, O. Giménez, M. Kang and M. Noy, Enumeration and limit laws of series-parallel graphs, European J. Combin. **28** (2007) 2091–2105.

[33] H. L. Bodlaender, A Partial k-Arboretum of Graphs with Bounded Treewidth, *Theor. Comput. Sci.* **209** (1-2) (1998) 1–45.

[34] B. Bollobas, Random graphs, Cambridge university press, 2001.

[35] J. A. Bondy and U. S. R. Murty. *Graph Theory*, Springer Verlag, New York, 2008.

[36] G. Chapuy, E. Fusy, O. Giménez, B. Mohar and M. Noy. Asymptotic enumeration and limit laws for graphs of fixed genus. *J. Combinatorial Theory Ser A* **118** (2011) 748–777.

[37] M. Deijfen and W. Kets, Random intersection graphs with tunable degree distribution and clustering, *Probab. Eng. Inf. Sci.* **23** (2009) 661–674.

[38] R. Diestel, *Graph Theory*, third edition, Springer, 2005.

[39] G. Dirac, Some results concerning the structure of graphs, *Canad. Math. Bull.* **8** (1965) 459–463.

[40] M. Drmota, Random trees: an interplay between combinatorics and probability, Springer, 2009.

[41] M. Drmota, O. Giménez and M. Noy, Vertices of given degree in series-parallel graphs, *Random Struct. Algorithms*, **36** (2010), 273 – 314.

[42] R.G Downey and M. R. Fellows. Fundamentals of Parameterized complexity. *Undegraduate Texts in Computer Science*, Springer-Verlag, 2012.

[43] Z. Dvořak and S. Norine, Small graph classes and bounded expansion, *J. Combin. Theory B* **100** (2010) 171–175.

[44] P. Erdős, A.W. Goodman and L. Pósa, The representation of a graph by set intersections. *Canad. J. Math.* **18** (1966) 106–112.

[45] P. Erdős and L. Pósa, On independent circuits in a graph, *Canad. J. Math.* **17** (1965) 347–352.

[46] L. Eschenauer and V.D. Gligor, A key-management scheme for distributed sensor networks, *Proceedings of the 9th ACM Conference on Computer and Communications Security* (2002) 41–47.

[47] J. Fill , E. Scheinerman , K. Singer-Cohen, Random intersection graphs when $m = \omega(n)$: an equivalence theorem relating the evolution of the $G(n,m,p)$ and $G(n,p)$ models, *Random Struct. Algorithms*, **16**, 156–176, 2000.

[48] P. Flajolet and R. Sedgewick, *Analytic Combinatorics,* Cambridge University Press, 2009.

[49] A.M. Frieze, On the independence number of random graphs. Discrete Math. 81 (1990), 171–175.

[50] A.M. Frieze and C. McDiarmid, Algorithmic theory of random graphs. Random Structures Algorithms **10** (1997), 5–42.

[51] J. Galambos and E. Seneta, Regularly varying sequences, *Proc. Amer. Math. Soc.* **41**(1973) 110–116.

[52] O. Giménez, M. Noy, Counting planar graphs and related families of graphs, *Surveys in Combinatorics 2009*, Cambridge University Press, Cambridge (2009), 169–210.

[53] E. Godehardt and J. Jaworski, Two models of random intersection graphs for classification. In: O. Optiz and M. Schwaiger, Editors, *Studies in Classification, Data Analysis and Knowledge Organization* **22**, Springer, Berlin, 2003, 67–82.

[54] R. M. Haralick, The diclique representation and decomposition of binary relations, *J. ACM* **21** (1974) 356–366.

[55] S. Janson, T. Łuczak and I. Norros, Large cliques in a power-law random graph, *J. Appl. Probab.* **47** (2010), 1124–1135.

[56] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley-Interscience, 2000.

[57] M. Kang and C. McDiarmid, Random unlabelled graphs containing few disjoint cycles, *Random Struct. Algorithms* **38** (2011) 174–204.

[58] M. Karoński, E. R. Scheinerman and K. B. Singer-Cohen, On Random Intersection Graphs: The Subgraph Problem, *Combin. Probab. Comput.* **8** (1999) 131–159.

[59] R. M. Karp and M. Sipser, Maximum matchings in sparse random graphs, *IEEE Conference on Computing* (1981), 364–375.

[60] V. Kolchin, B. Sevstyanov and V.Chistyakov, Random Allocations, *V.H. Winston and Sons*, 1978, Washington D.C.

[61] V. Kurauskas, Graphs without k disjoint cycles, MSc thesis, University of Oxford, 2008.

[62] V. Kurauskas, On small subgraphs in a random intersection digraph. *Discrete Mathematics* **313** (2013) 872–885.

[63] V. Kurauskas, On graphs containing few disjoint excluded minors. Asymptotic number and structure of graphs containing few disjoint minors K4, 2013, submitted.

[64] V. Kurauskas and C. McDiarmid, Random graphs with few disjoint cycles, *Combin. Probab. Comput.*, **20** (2011) 763–775.

[65] V. Kurauskas and C. McDiarmid, Random graphs containing few disjoint excluded minors, *Random Struct. Algorithms*, doi: 10.1002/rsa.20447.

[66] V. Kurauskas and K. Rybarczyk, On the chromatic number in random uniform hypergraphs, 2013, submitted.

[67] L. Lovász, On graphs not containing independent circuits (Hungarian), *Mat. Lapok* **16** (1965), 289–299.

[68] L. Lovász, *Combinatorial Problems and Exercises*, second edition, North-Holland, 1993.

[69] T. Łuczak, The chromatic number of random graphs. Combinatorica **11** (1991), 45–54.

[70] W. Mader, Homomorphiesätze für Graphen, *Math. Ann.* **178** (1968) 154–168.

[71] D. Matula, The largest clique size in a random graph. Tech. Rep., Dept. Comp. Sci., Southern Methodist Univerity, Dallas, Texas (1976).

[72] C. McDiarmid, Concentration, in *Probabilistic Methods for Algorithmic Discrete Mathematics*, M. Habib, C. McDiarmid, J. Ramirez-Alfonsin and B. Reed Eds., Springer, New York (1998) 195–248.

[73] C. McDiarmid. Random graphs on surfaces. *J. Combin. Theory Ser. B* **98** (2008) 778–797.

[74] C. McDiarmid, Random graphs from a minor-closed class, *Combin. Probab. Comput.*, **18** (4) (2009) 583–599.

[75] C. McDiarmid. On graphs with few disjoint t-star minors. *European J. Comb.* **32** (2011) 1394–1406.

[76] C. McDiarmid: Random Graphs from a Weighted Minor-Closed Class. *Electron. J. Comb.* **20** (2013).

[77] C. McDiarmid, A. Steger and D. Welsh, Random planar graphs, *J. Combin. Theory B* **93** (2005) 187–205.

[78] C. McDiarmid, A. Steger and D. Welsh, Random graphs from planar and other addable classes, *Topics in Discrete Mathematics* (M. Klazar, J. Kratochvil, M. Loebl, J. Matousek, R. Thomas, P. Valtr eds), Algorithms and Combinatorics **26**, Springer, 2006, 231–246.

[79] A. Meir and J.W. Moon, On an asymptotic method in enumeration. *J. Combin. Theory Ser. A* **51** (1989), 77–89. Erratum: *J. Combin. Theory Ser. A* **52** (1989), 163.

[80] S. E. Nikoletseas, C. Raptopoulos and P. G. Spirakis, Colouring non-sparse random intersection graphs. *In: Mathematical Foundations of Computer Science 2009*, Springer Berlin Heidelberg, 2009, 600–611.

[81] S. Nikoletseas, C. Raptopoulos and P. G. Spirakis, Maximum cliques in graphs with small intersection number and random intersection graphs, *Mathematical Foundations of Computer Science 2012*, Springer Berlin Heidelberg, 2012. 728–739.

[82] N. Pippenger and J. Spencer, Asymptotic behavior of the chromatic index for hypergraphs, *J. Combin. Theory Ser. A* **51** (1989), 24–42.

[83] S. Norine, P. Seymour, R. Thomas, P. Wollan, Proper minor-closed families are small, *J. Combin. Theory B*, **96** (2006) 754–757.

[84] B. Oporowski, J. Oxley, and R. Thomas, Typical subgraphs of 3-and 4-connected graphs. *J. Combin. Theory B*, **57** (1993) 239–257.

[85] K. Panagiotou and A. Steger, On the degree sequence of random planar graphs, *Proceedings of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '11)*, 2011, 1198-1210.

[86] B.A. Reed, N. Robertson, P.D. Seymour and R. Thomas, Packing directed circuits, *Combinatorica* **16** (1996) 535 –554.

[87] A. Rényi, Some remarks on the theory of trees, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* **4** (1959) 73–85.

[88] N. Robertson and P. Seymour, Graph minors V. Excluding a planar graph, *J. Combin. Theory B* **41** (1986) 92–114.

[89] N. Robertson and P. Seymour, Graph minors. XX. Wagner's Conjecture, *J. Combin. Theory B* **92** (2004) 325–357.

[90] W. Rudin, *Principles of mathematical analysis*, McGraw-Hill, New York, 1964.

[91] K. Rybarczyk, Equivalence of the random intersection graph and G(n,p), *Random Struct. Algorithms*, **38** (2011), 205–234.

[92] K. Rybarczyk and D. Stark, Poisson approximation of the number of cliques in random intersection graphs, *J. Appl. Probab.*, **47** (2010), 826–840.

[93] K. Rybarczyk, The degree distribution in random intersection graphs, *In: Challenges at the Interface of Data Analysis, Computer Science, and Optimization*, 2013, 291–299. Springer, Berlin.

[94] K. Rybarczyk, Constructions of independent sets in random intersection graphs, manuscript.

[95] OEIS Foundation Inc. (2011), The On-Line Encyclopedia of Integer Sequences, `http://oeis.org`.

[96] K. Singer, Random Intersection Graphs, Ph.D. thesis, The Johns Hopkins University, 1995.

[97] D. Stark, The vertex degree distribution of random intersection graphs, *Random Struct. Algorithms* **24** (2004) 249–258.

[98] C. Thomassen, On the presence of disjoint subgraphs of a specified type, *J. Graph Theory* **12** (1988) 101–111.

[99] B. A. Trakhtenbrot, Towards a theory of non-repeating contact schemes (in Russian), *Trudi Mat. Inst. Akad. Nauk SSSR* **51** (1958), 226–269.

[100] T. R. S. Walsh. Counting labelled three-connected and homeomorphically irreducable two-connected graphs, *J. Combin. Theory Ser. B* 32 (1982), 1-11.

[101] N. C. Wormald, Differential equations for random processes and random graphs, *Ann. Appl. Probab.***5** (1995): 1217–1235.