TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl Angewandte Geometrie und Diskrete Mathematik

Spanning subgraphs of growing degree
A generalised version of the Blow-up Lemma and its applications

Andreas Würfl

TECHNISCHE UNIVERSITÄT MÜNCHEN
Lehrstuhl Angewandte Geometrie und Diskrete Mathematik

Spanning subgraphs of growing degree
A generalised version of the Blow-up Lemma and its applications

Andreas Würfl

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende:		Univ.-Prof. Dr. Simone Warzel
Prüfer der Dissertation:	1.	Univ.-Prof. Dr. Anusch Taraz
	2.	Univ.-Prof. Mathias Schacht, Ph.D.
		Universität Hamburg
	3.	Prof. Dr. Deryk Osthus
		University of Birmingham / UK

Die Dissertation wurde am 29. November 2012 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 23. Januar 2013 angenommen.

Abstract

This thesis is concerned with embedding problems for spanning subgraphs of growing maximum degree into dense host graphs.

We generalise the well known Blow-up Lemma of Komlós, Sarközy, and Szemerédi by replacing the constant degree bound for the target graph with a bound on its arrangeability. Applications of the strengthened Blow-up Lemma include new embedding results for graphs of sublinear bandwidth and planar graphs. In addition, we determine the maximum size of a homogeneous set in typical graphs without an induced copy of C_5 or P_4 respectively. Our proofs are based on the regularity method.

Zusammenfassung

Die vorliegende Arbeit befasst sich mit der Einbettung von aufspannenden Subgraphen mit wachsendem Maximalgrad in dichte Trägergraphen.

Es wird eine Verallgemeinerung des bekannten Blow-up Lemmas von Komlós, Sarközy und Szemerédi bewiesen, die die konstante Gradschranke für Gastgraphen durch eine Schranke für die *arrangeability* ersetzt. Anwendungen des verallgemeinerten Blow-up Lemmas umfassen neue Einbettungsergebnisse für Graphen mit sublinearer Bandweite und planare Graphen. Desweiteren wird die maximale Größe einer homogenen Menge für typische Graphen ohne induzierte C_5-Kopie bzw. ohne induzierte P_4-Kopie bewiesen. Die Beweise basieren auf der Regularitätsmethode.

Contents

1 Introduction

> *"Extremal graph theory, in its strictest sense, is a branch of graph theory developed and loved by Hungarians."*
>
> – Béla Bollobás (1978)

Extremal combinatorics is a branch of discrete mathematics. The term "*extremal*" comes from the nature of the problems this field deals with: given a collection of finite objects (numbers, graphs, sets, etc.) that satisfies certain restrictions, how large or small can the collection be? The restriction imposed on the combinatorial objects often depends on a local property while extremality refers to a global one. The following problem illustrates typical features of an extremal combinatorial problem.

Given an integer n, what is the maximum number of elements in $\{1, \dots, n\}$ one can choose without choosing three distinct integers a_1, a_2, a_3 so that $a_3 - a_2 = a_2 - a_1$? The problem defines a restriction on the triples of elements (local) and asks for the maximal cardinality of a set satisfying this restriction (global). A generalised version of the preceding question has become known as the famous Erdős-Turán conjecture. An arithmetic progression of length k is a set of k integers $\{a_1, \dots, a_k\}$ with the property that all differences of the form $a_{i+1} - a_i$ are the same. In 1936, Erdős and Turán conjectured that any subset of the integers with positive density contains arbitrarily long arithmetic progressions. It took almost 40 years until Szemerédi proved this conjecture in 1975. The key element of Szemerédi's ground breaking proof – the *Regularity Lemma* – has become a central tool in graph theory and inspired a whole new series of important results.[1]

A classical result from the time before the Regularity Lemma, and one of the earliest results in extremal graph theory, is Turán's Theorem. Turán's Theorem determines the maximum number of edges in a graph that does not contain a copy of K_r. Any graph with more than this number of edges must contain a complete graph on r vertices.

Another classical result from extremal graph theory is the Theorem of Erdős and Stone. It considers the question how many edges are needed in a host graph G to ensure that a graph H can be embedded into G. Interestingly, neither the number of vertices nor the number of edges in H but its chromatic number determines the fraction of edges necessary to enforce an embedding into G.

The Erdős-Stone theorem applies to graphs H that are much smaller than the host graphs G into which they are embedded. It was again Szemerédi who, together with János Komlós and Gábor N. Sarközy, broke the ground with a first general embedding

[1] Endre Szemerédi has been awarded the 2012 Abel Prize because he "*has revolutionized discrete mathematics by introducing ingenious and novel techniques*". The Regularity Lemma is an outstanding example of those contributions.

result for graphs H that have the same number of vertices as their host graphs G. Their Blow-up Lemma paved the road for numerous important results on *global* structures (the spanning graph H) that arise from *local* properties (e.g. the minimum degree) of the host graph G.

Let us now give a short impression of a central concept in the Regularity Method. The Regularity Lemma not only was the key to an amazing result in number theory, it also became the foundation of a powerful technique in extremal graph theory. Roughly speaking, this lemma states that any sufficiently large graph decomposes into a relatively small number of parts that do have a very regular structure. More precisely, the vertex set of a graph G can be partitioned into a constant number of equally sized classes so that the following holds for most pairs (A, B) of such classes: The edges running between the classes A and B form a bipartite graph with random-like edge distribution. Later in this chapter we will explain what we mean by "random-like edge distribution". A partition into sets that have random-like edge distribution between them is also called a *regular partition* and the random-like bipartite graphs are the *regular pairs* of this partition.

The Blow-up Lemma states that – with respect to the embedding of graphs with constant maximum degree – regular pairs behave (almost) like complete bipartite graphs if they satisfy an additional minimum degree condition. Combining the Regularity Lemma and the Blow-up Lemma hence gives a powerful tool for the spanning embedding of graphs with constant maximum degree.

As a main result this thesis contains a generalisation of the Blow-up Lemma. Our version of the Blow-up Lemma does not require the constant degree bound for the target graph. Instead it relies on the more general concept of bounded arrangeability (which we will introduce later in this chapter) to derive a spanning embedding. This strengthening yields several applications: We obtain various embedding results for spanning subgraphs of potentially unbounded maximum degree. Among those we present a variant of the Bandwidth Conjecture of Bollobás and Komlós for arrangeable graphs. This in turn provides generalised embedding results, e.g. for F-factors, planar graphs or graphs embeddable into any (fixed) orientable surface.

As we will see in this introduction, many important results in extremal graph theory have indeed been developed by Hungarians. And we would never doubt the assessment of Béla Bollobás when it comes to graph theory. We will see in the following that Hungarians are no longer *the only ones* that develop and love extremal graph theory.

Before stating our results in Section 1.2 we will give a short account of the relevant previous work in the area in Section 1.1. In Section 1.3 we will then sketch and explain some techniques that are used in many parts of this thesis. Those mostly fall into the area of the Regularity Lemma or the Probabilistic Method.

1.1 History

In this section we provide some of the most important background material concerning extremal graph theory (see Section 1.1.1) and random graph theory (see Section 1.1.2). This will enable us to put our results in context later.

1.1.1 Extremal graph theory

We start this section with a short definition of the topic by Reinhard Diestel [33]: "Extremal graph theory is a branch of the mathematical field of graph theory. Extremal graph theory studies extremal (maximal or minimal) graphs which satisfy a certain property. Extremality can be taken with respect to different graph invariants, such as order, size or girth. More abstractly, it studies how global properties of a graph influence local substructures of the graph."

The subgraph containment problem The primordial problem in extremal graph theory has been solved by Paul Turán [92]. His by now classical result gives a tight lower bound on the number of edges that force a copy of a complete graph on r vertices. It states that K_r can be embedded into each n-vertex graph G with average degree strictly greater than $\frac{r-2}{r-1}n$.

Theorem 1.1 (Turán [92])
Every graph G on n vertices with average degree $d(G) > \frac{r-2}{r-1}n$ contains a complete graph K_r on r vertices as a subgraph.

This theorem was generalised to arbitrary fixed graphs H by Erdős, Stone, and Simonovits [45, 38]. Their result, sometimes also called the fundamental theorem of extremal graph theory, states the following: the average degree of a host graph needed to guarantee an embedding of a fixed graph H depends only on the *chromatic number* of H, i.e., the minimal number of colours necessary to colour the vertices of H in such a way that no two adjacent vertices receive the same colour.

Theorem 1.2 (Erdős, Stone [38])
For every constant $\gamma > 0$ and every fixed graph H with chromatic number $r \geq 2$ there is a constant $n_0 \in \mathbb{N}$ such that every graph G with $n \geq n_0$ vertices and average degree $d(G) \geq \left(\frac{r-2}{r-1} + \gamma\right) n$ contains a copy of H as a subgraph.

This result more or less settled the question of fixed subgraph containment and the focus shifted towards more complex structures, among them spanning subgraphs.

Spanning subgraphs This section translates Theorem 1.2 into a spanning setting, i.e., one where the graphs H and G have the same number of vertices. It is obvious that at least two changes are necessary here.

Since a spanning subgraph uses every single vertex of G, we need to control every single vertex of G. Hence the average degree condition must be replaced by one involving the minimum degree $\delta(G)$ of G. Moreover, there are simple examples of r-chromatic graphs H that show that the lower bound has to be raised to at least $\delta(G) \geq \frac{r-1}{r}n$: simply consider the case where G is the complete r-partite graph with partition classes almost but not exactly of the same size (thus G has minimum degree almost $\frac{r-1}{r}n$) and let H be the union of $\lfloor n/r \rfloor$ vertex disjoint r-cliques (see Figure 1.1).

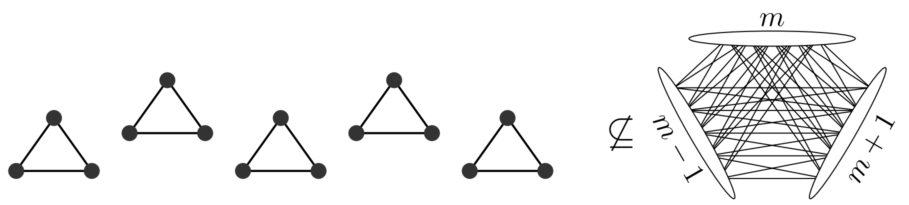

Figure 1.1: The complete 3-partite graph G on $3m$ vertices with partition classes of size $m-1, m, m+1$ does not contain m vertex disjoint triangles.

And indeed, the minimum degree $\frac{r-1}{r}n$ is sufficient to guarantee the existence of certain spanning subgraphs H. Maybe the most well-known example is Dirac's Theorem [34] which states that any graph G on n vertices contains a Hamilton cycle, i.e., a cycle on n vertices, if its minimum degree satisfies $\delta(G) \geq n/2$.

Another early result on large r-chromatic subgraphs of graphs with minimum degree $\frac{r-1}{r}n$ is due to Corrádi and Hajnal [30]. They showed that in a graph G with n vertices and minimum degree $\delta(G) \geq \frac{2}{3}n$ all but at most two vertices can be covered by vertex disjoint triangles.

Theorem 1.3 (Corrádi, Hajnal [30])
Let G be a graph on n vertices with minimum degree $\delta(G) \geq \frac{2}{3}n$. Then G contains $\lfloor n/3 \rfloor$ vertex disjoint triangles.

This was generalised by Hajnal and Szemerédi [49], who proved the K_r-analogon, i.e., that every graph G with $\delta(G) \geq \frac{r-1}{r}n$ must contain $\lfloor n/r \rfloor$ vertex disjoint cliques of order r.

Theorem 1.4 (Hajnal, Szemerédi [49])
Let G be a graph on n vertices with minimum degree $\delta(G) \geq \frac{r-1}{r}n$. Then G contains $\lfloor n/r \rfloor$ vertex disjoint copies of K_r.

If a graph G on n vertices contains $\lfloor n/|F| \rfloor$ vertex disjoint copies of a graph F we also say that G contains an *F-factor.*

The first non-trivial case of Theorem 1.4 is $r = 2$, the existence of a perfect matching which follows (for even n) from $\delta(G) \geq \frac{1}{2}n$. This is the same threshold as in Dirac's Theorem about the existence of a Hamilton cycle. Could we find a more connected structure in every graph G with high minimum degree such that the copies of K_r are part of this structure? Indeed, Pósa (see, e.g., [41]) suggested a further extension of Theorem 1.3. He conjectured that at the same degree threshold $\delta(G) \geq \frac{2}{3}n$ where the Theorem of Corrádi and Hajnal (Theorem 1.3) guarantees the existence of a spanning triangle factor, a graph must already contain the square of a Hamilton cycle. (Here the r-th power of a graph is obtained by inserting an edge between every two vertices with distance at most r in the original graph, and the square of a graph is its second power.) Note that the square of a cycle on at least $3t$ vertices contains t vertex disjoint triangles.

Conjecture 1.5 (Pósa [41])
An n-vertex graph G with minimum degree $\delta(G) \geq \frac{2}{3}n$ contains the square of a Hamilton cycle.

This conjecture was later generalised by Seymour [89] who suggested that one can exchange "square" and $\delta(G) \geq \frac{2}{3}n$ for the "r-th power" and $\delta(G) \geq \frac{r-1}{r}n$ (this is often called the Pósa-Seymour conjecture). Interestingly this again is the same threshold as given by Theorem 1.4 for K_r-factors. Komlós, Sarközy, Szemerédi [61] first proved Pósa's Conjecture (Conjecture 1.5), then an approximate version of the more general Pósa-Seymour conjecture [64] and finally [65] gave a full proof for fixed r and sufficiently large graphs G.

Theorem 1.6 (Komlós, Sarközy, Szemerédi [65])
For every integer $r \geq 1$ there is an integer n_0 such that every graph G on $n > n_0$ vertices with minimum degree $\delta(G) \geq \frac{r-1}{r}n$ contains the $(r-1)$-st power of a Hamilton cycle.

The same authors also pioneered embedding results for a more general class of graphs. Proving a conjecture of Bollobás [12, p. 437], Komlós, Sarközy, and Szemerédi [60] showed (for n is sufficiently large) that graphs of order n with minimum degree $(\frac{1}{2}+\gamma)n$ contain all trees on n vertices that respect a constant degree bound. They later generalised this to trees of maximum degree $O(n/\log n)$ which is optimal up to a constant factor. To our knowledge, this was the only result on spanning subgraphs that did not require a constant degree bound.

Theorem 1.7 (Komlós, Sarközy, Szemerédi [67])
For every $\gamma > 0$ there is an integer n_0 and a constant $c > 0$ such that the following holds. If T is a tree of order $n \geq n_0$ with $\Delta(T) \leq cn/\log n$, and G is a graph of order n with $\delta(G) \geq (\frac{1}{2}+\gamma)n$, then G contains T as a subgraph.

Several other results for bounded degree subgraphs have been obtained within the last decade. As examples we present the following two theorems due to Komlós, Sarközy, Szemerédi [66] on spanning F-factors (which solves a conjecture by Alon and Yuster [9]) and Kühn and Osthus [71] on spanning triangulations. For a detailed survey see [72] and the references therein.

Theorem 1.8 (Komlós, Sarközy, Szemerédi [66])
For every r-chromatic graph F there is a constant C such that every graph G of order n and minimum degree at least $\frac{r-1}{r}n + C$ contains an F-factor.

Theorem 1.9 (Kühn, Osthus [71])
There is n_0 such that every graph G with $n \geq n_0$ vertices and minimum degree at least $\frac{2}{3}n$ contains a spanning triangulation.

The results presented so far apply to very specific classes of target graphs (F-factors, powers of cycles, trees of bounded maximum degree). We conclude this section on extremal graph theory with a theorem that applies to a rather general class of (bounded degree) graphs – graphs of sublinear bandwidth. A graph G has bandwidth at most b, if there is a labelling of the vertices by numbers $1, \dots, n$, such that for every edge ij of the graph we have $|i-j| \leq b$. F-factors, powers of cycles, trees of bounded degree, and planar graphs of bounded degree all have sublinear bandwidth [17]. Hence the following general statement which was suggested by Bollobás and Komlós [58, Conjecture 16] and has been proved recently by Böttcher, Schacht, and Taraz [18] applies to each of these classes.

Theorem 1.10 (Böttcher, Schacht, and Taraz [18])
For all $r, \Delta \in \mathbb{N}$ and $\gamma > 0$, there are constants $\beta > 0$ and $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$, if H is an r-chromatic graph on n vertices with $\Delta(H) \leq \Delta$, and bandwidth at most βn and if G is a graph on n vertices with minimum degree $\delta(G) \geq \left(\frac{r-1}{r} + \gamma\right) n$, then G contains a copy of H.

1.1.2 Random graph theory

The field of random graph theory started around the middle of the 20^{th} century thanks to several papers of Erdős [39, 40, 42, 43]. He applied probabilistic arguments to show the existence of graphs with certain properties; two of these examples we shall present later in this section. Let us first define the notion of a random graph and shed some light on central concepts of random graph theory. A random graph is a probability space over a set of graphs. Random graph theory now revolves around the question: "What is the probability that a random graph graph does have a certain property."

There are two strong motivations in random graph theory. The one that initially inspired Erdős was the following. To prove the existence of a graph with a certain property, it suffices to show that a random graph does have this property with positive probability. There are numerous examples where a short probabilistic argument gives the existence of graphs with a prescribed property but where it is extremely difficult to find a deterministic construction. The approach via random graphs has become known as the *Probabilistic Method.*

The second motivation is the study of typical properties in a given class of graphs. This is indeed an interesting problem for many graph classes as the values of graph parameters often are highly concentrated even though a wide range of values is attained by the class. We demonstrate this using the example of random triangle-free graphs. It is well-known that triangle-free graphs can have arbitrarily large chromatic number. However, if a graph is drawn uniformly at random from the set of all triangle-free graphs it is bipartite with probability tending to 1. We also say that *almost all* triangle-free graphs are bipartite.

This thesis pursues problems of both types. We employ the Probabilistic Method to prove embedding results for spanning subgraphs. Later we give bounds on the size of homogeneous sets for almost all graphs in certain graph classes defined by forbidden induced subgraphs.

Almost all graphs The study of random graphs was started by Erdős and Rényi [43] in 1959. They introduced the $\mathcal{G}(n,p)$ random graph model where each edge is inserted into a graph of order n uniformly and independently with probability p where $p = p(n)$ is allowed to depend on n. A first question to ask is whether such a random graph typically is connected, or Hamiltonian, or contains a fixed subgraph H. Erdős and Rényi proved that the answer to these questions switches rapidly between "yes" and "no" when p crosses certain thresholds.

Theorem 1.11 (Erdős, Rényi [43])
Let $k \geq 3$. Then

$$\lim_{n\to\infty} \mathbb{P}[\mathcal{G}(n,p) \textit{ contains a } k\textit{-cycle}] = \begin{cases} 0 & \textit{if } pn \to 0, \\ 1 & \textit{if } pn \to \infty. \end{cases}$$

Interestingly, the threshold of Theorem 1.11 is independent of the cycle length k; cycles of all (constant) lengths appear as soon as p is significantly larger than $1/n$.

The $\mathcal{G}(n,p)$-model is not the only random graph model. Any class of graphs can be turned into a random graph model by choosing randomly among its members. A recent example is the following result by McDiarmid and Reed [77].

Theorem 1.12 (McDiarmid, Reed [77])
For every fixed surface S there are constants $0 < c < C$ such that

$$c \log n \leq \Delta(H) \leq C \log n$$

for almost all graphs H of order n which are embeddable into S.

Probabilistic versions of open problems is another popular scenario in random graph theory. One example is related to a famous conjecture by Erdős and Hajnal [37]. The conjecture suggests that for every graph H there is a positive constant $\varepsilon(H)$ such that all graphs G which do not contain H as an induced subgraph have a homogeneous set of size at least $|G|^{\varepsilon(H)}$. (A homogeneous set spans either a clique or a stable set in G.) This conjecture is open for most graphs H. Loebl, Reed, Scott, Thomason, Thomassé [75] however showed that for every graph H the statement is true for almost all graphs without induced H.

Theorem 1.13 (Loebl, Reed, Scott, Thomason, Thomassé [75])
For every H there is $\varepsilon(H) > 0$ such that almost all graphs G without an induced copy of H contain a clique or a stable set of size $|G|^{\varepsilon(H)}$.

While random graph theory is an interesting field in its own right, random graph models have proved a valuable tool for a variety of combinatorial problems. Even if we cannot prove that almost all graphs have a given property random graphs may yield valuable insights. Any positive probability certifies that at least one graph with the property in question exists. Deducing existence from positive expectation is the principle that lies at the core of the *Probabilistic Method.* The following paragraph gives two early and by now classical application of this concept.

The Probabilistic Method The basic *Probabilistic Method* can be described as follows: In order to prove the existence of a combinatorial structure with certain properties, we construct an appropriate probability space and show that a randomly chosen element in this space has the desired properties with positive probability. The following example was one of the first in which this idea has been applied successfully.

The *Ramsey number* $R(k)$ is the smallest integer n such that in any two-colouring of the edges of a complete graph on n vertices K_n by red and blue, there is a red K_k (i.e., a complete subgraph on k vertices all of whose edges are coloured red) or a blue one. Frank P. Ramsey [82] showed that $R(k)$ is finite for any integer k. His results imply an exponential upper bound of $R(k) \leq 2^{2k-3}$. It was Paul Erdős [39] who employed the Probabilistic Method to prove an exponential lower bound for $R(k)$.

Theorem 1.14 (Erdős [39])
Let $k \geq 3$. Then $R(k) > 2^{k/2}$.

The idea of the proof is as simple as intriguing. Consider a random two-colouring of the edges of K_n obtained by colouring each edge independently either red or blue, where each colour is equally likely. Then any fixed set of k vertices spans a monochromatic clique with probability at least $2^{1-\binom{k}{2}}$. Since there are $\binom{n}{k}$ subsets of $\{1, \ldots, n\}$ with k elements, the probability that at least one monochromatic subset occurs is at most $\binom{n}{k}2^{1-\binom{k}{2}}$. Thus, a colouring without a monochromatic K_k exists if $\binom{n}{k}2^{1-\binom{k}{2}} < 1$. This is the case for $n = \lfloor 2^{k/2} \rfloor$.

To our knowledge no explicit construction is known that achieves this lower bound. But a simple first moment calculation proves that a random edge-colouring of K_n with positive probability does not contain a monochromatic K_k. Hence, there is an edge-colouring of K_n that does not have a monochromatic K_k. Up to date this lower bound is essentially the best known.

Another founding example of the technique that had by then become known under the name *Probabilistic Method* is also due to Erdős. In 1959 he answered the question whether graphs without short cycles can have high chromatic number in the affirmative [40].

Theorem 1.15 (Erdős, [40])
For all positive integers k, ℓ there is a graph G on $\ell^{1+1/(2k)}$ vertices that does not contain any cycles of length up to k and has $\alpha(G) < \ell$. In particular, $\chi(G) \geq \ell^{1/k}$.

Half a century after its initial discovery the Probabilistic Method remains a versatile tool in combinatorics. The ideas behind it have been refined from simple first moment calculations to more and more sophisticated probabilistic arguments. Those also include randomised algorithms and their analysis – an important tool in our later proofs. The core argument of the Probabilistic Method remains unchanged: positive probability implies existence.

1.2 Results

In this section we present an overview of the results obtained throughout this thesis and put them in the context of the previous section. We will describe an extension of the Blow-up Lemma (Section 1.2.1) which is one of our core results. As a major application we provide a generalised version of the Conjecture of Bollobás and Komlós (Section 1.2.2). Moreover, we discuss planar subgraphs in dense graphs (Section 1.2.3). All results will be repeated with more details in the respective chapters that cover their proofs.

1.2.1 A Blow-up Lemma for growing degrees

As we have laid out in Section 1.1.1 the last decade has seen tremendous progress on various results for spanning subgraphs. The foundation of this success was laid in the Blow-up Lemma by Komlós, Sarközy, and Szemerédi [62]. In order to state this result we need some definitions.

Let $G = (V, E)$ be a graph and let $\varepsilon, \delta > 0$. A pair (A, B) with $A, B \subseteq V$, $A \cap B = \emptyset$ is called *ε-regular* if

$$\left| \frac{e(A, B)}{|A|\,|B|} - \frac{e(X, Y)}{|X|\,|Y|} \right| \leq \varepsilon$$

for all $X \subseteq A$, $Y \subseteq B$ with $|X| \geq \varepsilon|A|$, $|Y| \geq \varepsilon|B|$. A pair (A, B) is called *(ε, δ)-super-regular* if it is ε-regular and, in addition, $\deg(v, B) \geq \delta|B|$ and $\deg(w, A) \geq \delta|A|$ for all $v \in A$, $w \in B$.

Let G, H and R be graphs with vertex sets $V(G)$, $V(H)$, and $V(R) = \{1, \dots, r\}$. We say that H has an *R-partition* $V(H) = X_1 \mathbin{\dot\cup} \dots \mathbin{\dot\cup} X_r$, if for every edge $xy \in E(H)$ there are distinct $i, j \in [r]$ with $x \in X_i$, $y \in X_j$ and $ij \in E(R)$. G has a *corresponding (ε, δ)-super-regular R-partition* $V(G) = V_1 \mathbin{\dot\cup} \dots \mathbin{\dot\cup} V_r$, if $|V_i| = |X_i| =: n_i$ for all $i \in [r]$ and every pair (V_i, V_j) with $ij \in E(R)$ is (ε, δ)-super-regular. These partitions are *balanced* if $n_1 \leq n_2 \leq \dots \leq n_r \leq n_1 + 1$. We now state a basic version of the Blow-up Lemma.

Theorem 1.16 (Komlós, Sarközy, Szemerédi [62])
Given a graph R of order r and positive parameters δ, Δ there is a positive $\varepsilon = \varepsilon(\delta, \Delta, r)$ such that the following holds. Suppose that H and G are two graphs with the same number of vertices, where $\Delta(H) \leq \Delta$ and H has a balanced R-partition, and G has a corresponding (ε, δ)-super-regular R-partition. Then there is an embedding of H into G.

This result is one key ingredient in the proofs of Theorem 1.6, Theorem 1.8, Theorem 1.9, and Theorem 1.10 from Section 1.1. We want to draw the reader's attention to the constant degree bound for H in Theorem 1.16. This restriction is mirrored by the fact that all of the listed theorems (implicitly) contain a constant bound on the maximum degree of the subgraphs embedded.

Komlós [58] suggested to relax the constant degree bound in Theorem 1.16 by the following more generous restriction. A graph H is called *a-arrangeable* if its vertices can be ordered as $(x_1,\dots,x_n)$ in such a way that $\big|N\big(N(x_i,\{x_{i+1},\dots,x_n\}),\{x_1,\dots,x_i\}\big)\big|\le a$ for each $1\le i\le n$. This indeed generalises a constant degree bound as every graph H with $\Delta(H)\le a$ is (a^2-a+1)-arrangeable. At the same time, even 1-arrangeable graphs, e.g. stars, can have unbounded degree. We give examples for graph classes with bounded arrangeability in the following section. We prove the following theorem in Chapter 4.

Theorem (Theorem 4.3)
Given a graph R of order r and positive parameters δ, a there is a positive $\varepsilon=\varepsilon(\delta,a,r)$ such that the following holds. Suppose that H and G are two graphs with the same number of vertices, where H is a-arrangeable, $\Delta(H)\le\sqrt{n}/\log n$ and H has a balanced R-partition, and G has a corresponding (ε,δ)-super-regular R-partition. Then there is an embedding of H into G.

It remains to say that the restriction of $\Delta(H)\le\sqrt{n}/\log n$ is optimal up to the log-factor (see Proposition 4.35). Already in 1996 Komlós and Simonovits extended the degree bound of the Blow-up Lemma to $c\sqrt{n/\log n}$ for the case that H is a tree (see [69, Theorem 6.6]).

1.2.2 A generalisation of a conjecture by Bollobás and Komlós

The conjecture of Bollobás and Komlós (Theorem 1.10) presents an embedding result for the rather general class of graphs with sublinear bandwidth and constant maximum degree. In Section 1.1.1 we have mentioned that powers of cycles, trees of bounded degree, F-factors, and planar graphs of bounded degree all satisfy the requirements of Theorem 1.10. Eliminating the constant degree bound would add even more classes of graphs to this list. The degree bound is required in only two places in the proof of Theorem 1.10 – one being an application of the Blow-up Lemma (Theorem 1.16). Building on the generalised Blow-up Lemma presented in Section 1.2.1 we give a proof of the following generalisation of the conjecture of Bollobás and Komlós in Chapter 5.

Theorem (Theorem 4.9)
For all $r,a\in\mathbb{N}$ and $\gamma>0$, there are constants $\beta>0$ and $n_0\in\mathbb{N}$ such that for every $n\ge n_0$ the following holds.

If H is an r-chromatic, a-arrangeable graph on n vertices with $\Delta(H)\le\sqrt{n}/\log n$ and bandwidth at most βn and if G is a graph on n vertices with minimum degree $\delta(G)\ge\left(\frac{r-1}{r}+\gamma\right)n$, then G contains a copy of H.

The more generous degree bound of $\Delta(H)\le\sqrt{n}/\log n$ allows various applications in the spirit of Section 1.1.2. We give one example. So let S be an orientable surface and $\mathcal{H}_S(n)$ be the set of all graphs on n vertices which are embeddable into the surface S. We argue that the requirements of the above theorem are satisfied for almost all graphs in $\mathcal{H}_S(n)$.

A result by McDiarmid and Reed [77] (Theorem 1.12) states that there is a constant $C=C(S)$ such that almost all graphs H in $\mathcal{H}_S(n)$ have $\Delta(H)\le C\log n$. Moreover, almost all graphs in $\mathcal{H}_S(n)$ have sublinear bandwidth (Böttcher, Pruessmann, Taraz, and Würfl [17]), are $a = a(S)$-arrangeable (Rödl and Thomas [87]), and have chromatic number at most $r = \lfloor (7 + \sqrt{1 + 48g})/2 \rfloor$ (Heawood [51], Apple and Haken [10]). Combining those results with our theorem above yields the following application of the generalised conjecture of Bollobás and Komlós.

Theorem (Corollary 5.4)
Let $\gamma > 0$ and let S be an orientable surface of genus $g \in \mathbb{N}_0$. Set $r = \lfloor (7 + \sqrt{1 + 48g})/2 \rfloor$ and let G be any graph of order n with $\delta(G) \geq (\frac{r-1}{r} + \gamma)n$. If H is drawn uniformly at random from $\mathcal{H}_S(n)$, then G contains H almost surely.

In particular, we obtain the subsequent result on planar graphs.

Corollary (Corollary 5.4)
Let $\gamma > 0$ then $\delta(G) \geq (\frac{3}{4} + \gamma)n$ implies that G contains almost all planar graphs on n vertices as subgraphs.

1.2.3 Large planar subgraphs in dense graphs

Given a graph G we want to exhibit a large planar subgraph and therefore define $\mathrm{pl}(G)$ to be the number of edges in the largest planar subgraph of G. Moreover, we set $\mathrm{pl}(n, d) := \min\{\mathrm{pl}(G) : v(G) = n,\ \delta(G) \geq d\}$. Those definitions are due to Kühn, Osthus, and Taraz [74]. Using this notation we can write Theorem 1.9 as $\mathrm{pl}(n, \frac{2}{3}n) = 3n - 6$ for large n. Kühn, Osthus, and Taraz [74] prove that $\mathrm{pl}(n, \gamma n) \geq 2n - C(\gamma)$ and $\mathrm{pl}(n, (\frac{1}{2} + \gamma)n) \geq 3n - C(\gamma)$ for some constant C that only depends on γ.

In Chapter 6 we investigate the behaviour of $\mathrm{pl}(n, d)$ for values d slightly above $n/2$. We discover the following peculiar behaviour.

Theorem (Theorem 6.2, Theorem 6.3)

$$\mathrm{pl}(2m - 1, m) = (4.5 + o(1))m\ , \qquad \mathrm{pl}(2m - 1, m + 1) = (5 + o(1))m\ .$$

Note that the class of complete bipartite graphs shows that $\mathrm{pl}(2m, m) = 4m - 4$. Hence the parameter $\mathrm{pl}(n, d)$ has a "jump" at the threshold $d = n/2$. Contrary to that, $\mathrm{pl}(n, d)$ exhibits a very smooth behaviour for a minimum degree of the form $d = \gamma n$ with $0 < \gamma < 1/2$. In Chapter 7 we prove the following theorem.

Theorem (Theorem 7.5)
For every $\gamma \in (0, 1/2)$ there is n_γ such that $\mathrm{pl}(n, \gamma n) = 2n - 4k$ for every $n \geq n_\gamma$, where $k \in \mathbb{N}$ is the unique integer such that $k \leq 1/(2\gamma) < k + 1$.

1.2.4 Homogeneous sets

A famous conjecture by Erdős and Hajnal [37] suggests that all graphs that do not contain a fixed induced subgraph H have large homogeneous sets. To make this more precise, we define $\mathrm{hom}(G)$ as the size of the largest homogeneous set (i.e., clique or stable set) in G. Using this notation the conjecture of Erdős and Hajnal says the following.

Conjecture 1.17 (Erdős, Hajnal [37])
For every graph H there is $\varepsilon(H) > 0$ such that $\hom(G) \geq |G|^{\varepsilon(H)}$ if G does not contain H as an induced subgraph.

The conjecture is known to be true for a small number of graphs H, but open, among others, for $H = C_5$ (see [48]). Loebl, Reed, Scott, Thomason, and Thomassé [75] however showed that for every graph H the statement is true for almost all graphs without induced H. They ask for which graphs H it is true that almost all graphs without an induced copy of H do have a *linearly* sized homogeneous set.

Let $\mathcal{F}orb^*_n(H)$ denote the set of graphs on vertex set $[n]$ that do not contain an induced copy of H. For $\eta > 0$ we define $\mathcal{F}orb^*_n(H, c)$ to be the set of graphs in $\mathcal{F}orb^*_n(H)$ that have $c\binom{n}{2}$. We answer the question of Loebl et. al. in the affirmative. The statement even remains true if we restrict ourselves to graphs of a fixed density.

Theorem (Theorem 8.3)
For every $0 < c < 1$ there is $\eta > 0$ such that

$$\lim_{n\to\infty} \frac{|\{G \in \mathcal{F}orb^*_n(C_5, c) : \hom(G) \geq \eta n\}|}{|\mathcal{F}orb^*_n(C_5, c)|} = 1.$$

This leaves us in a peculiar situation. Even though it is known that almost every graph without an induced copy of C_5 has a homogeneous set of linear size it remains unknown whether all graphs G without an induced copy of C_5 do have a homogeneous set of size $|G|^{\varepsilon}$ for any fixed $\varepsilon > 0$.

Recently Kang, McDiarmid, Reed, and Scott [56] have shown that most graphs H are such that almost all graphs without an induced copy of H do have a linearly sized homogeneous set. One case that is not covered by their result is $H = P_4$ which gives the class of cographs. We show that almost all cographs do not have homogeneous sets of linear size.

Theorem (Theorem 9.2)
For every $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \frac{|\{G \in \mathcal{F}orb^*_n(P_4) : \hom(G) \geq n/\log^{1-\varepsilon} n\}|}{|\mathcal{F}orb^*_n(P_4)|} = 0\,.$$

1.3 Ideas and concepts

In this section we briefly describe the central ideas and concepts that are necessary to prove the theorems presented in the previous section. Virtually all of these are based on the Regularity Lemma and corresponding embedding results for regular partitions. Hence this section constitutes an introduction into the regularity method with a focus on the Blow-up Lemma. All definitions and results of this section will be repeated later.

First, in Section 1.3.1, we motivate the Blow-up Lemma. Section 1.3.2 and Section 1.3.3 introduce important ideas in the proof of our generalised Blow-up Lemma. We conclude with questions regarding the application of the (generalised) Blow-up Lemma in Section 1.3.4.

1.3.1 Regular partitions and the Blow-up Lemma

As mentioned earlier, the Regularity Lemma is a structural result that guarantees that every sufficiently large graph has a partition which we call regular. This partition contains regular pairs which we described as random-like bipartite graphs. We shall make this more precise in the following.

The *density* $d(A, B)$ of a bipartite graph (A, B) with partition classes A and B is the number of its edges divided by the number of possible edges $|A|\ |B|$. The pair (A, B) is called *ε-regular* if all subsets A' of A and all subsets B' of B with $|A'| \geq \varepsilon|A|$ and $|B'| \geq \varepsilon|B|$ have the property that their density $d(A', B')$ differs from the density $d(A, B)$ by at most ε. An *ε-regular partition* of a graph $G = (V, E)$ is a partition of its vertex set V into $V_1, \ldots, V_k$ such that (V_i, V_j) forms an ε-regular pair for all but at most an ε-fraction of all possible index pairs i, j. The sets V_i are also called the *clusters* of the partition. The Regularity Lemma then guarantees the existence of an ε-regular partition of G into k clusters where k only depends on ε but not on the size of G.

Note that an empty bipartite graph is ε-regular for any positive ε. However, an empty bipartite graph will not be useful in applications concerning the embedding of graphs. For these applications we need regular pairs to have many edges (*dense* pairs). Given an ε-regular partition of a graph we now define a so called *reduced graph* R of the partition. This graph contains one vertex for each cluster and an edge for every ε-regular pair with density at least δ for some constant δ (much bigger than ε).

In the first pages of this chapter we called the edge distribution in an ε-regular pair "random-like". We elaborate on this analogy now. To do so we compare properties of a random bipartite graph and a dense ε-regular pair. Let $G_p = (A \cup B, E_G)$ be a bipartite graph that contains every edge from $A \times B$ independently with probability p and let $H = (A \cup B, E_H)$ be an ε-regular pair with density p (p much larger then ε). In both cases the degree of most vertices in A is close to $p|B|$ (by symmetry the analogous statement is true for B). Subpairs (A', B') of reasonable size in (A, B) have density close to p in both graphs. And finally both graphs contain roughly the same number of copies of any fixed (bipartite) graph F. In particular, a dense ε-regular pair contains many copies of a fixed bipartite graph F.

We can even guarantee much bigger structures in dense regular pairs if we require the additional property that every vertex has a δ-fraction of all possible neighbours. Regular pairs that satisfy this minimum degree condition are called *(ε, δ)-super-regular*. The so-called Blow-up Lemma (see Theorem 1.16) implies that an (ε, δ)-super-regular pair (A, B) is (almost) as good as a complete bipartite graph when it comes to the embedding of bounded-degree subgraphs. More generally, the Blow-up Lemma allows for the embedding of spanning graphs into *systems* of super-regular pairs. So assume that a graph $G = (V, E)$ has an ε-regular partition with reduced graph R in which every dense pair is also (ε, δ)-super-regular. Further assume that H is another graph on the vertex set V and that all edges of H run between clusters that are adjacent in R. If, in addition, the maximum degree of H is bounded by a constant Δ, then the Blow-up Lemma guarantees a copy of H in G (if ε is sufficiently small).

Our generalised Blow-up Lemma (see Theorem 4.3) replaces the constant degree bound for H by the weaker constraint for its arrangeability. Besides that, the method remains unchanged: In order to find a spanning subgraph H of G we partition the host graph G into (super-)regular pairs and apply the Blow-up Lemma to a corresponding partition of the target graph H.

Let us close with the remark that the additional requirement of super-regularity is no severe restriction. Any regular pair can easily be transformed into a super-regular pair by omitting a few vertices (see Proposition 3.11).

1.3.2 A randomised embedding algorithm

A randomised embedding algorithm plays a central rôle in the proof of our generalised Blow-up Lemma. We present the approach using the example of a single (ε, δ)-super-regular pair. The ideas naturally carry over to systems of super-regular pairs. It is quite instructive to study a naive greedy approach for this embedding problem first. So assume that the graph $G = (V_1 \cup V_2, E_G)$ is an (ε, δ)-super-regular pair and $H = (X_1 \cup X_2, E_H)$ is a bipartite graph with $|V_i| = |X_i|$ for $i = 1, 2$. We fix an arbitrary order for the vertices of H and start (randomly) embedding them one by one into the pair (V_1, V_2). Doing so we have to respect the structure of H, i.e., if we want to embed a vertex x of H and a neighbour y of x has already been embedded into $w \in V_1$ then we have to embed x into the neighbourhood of w. This restricts our choice for the embedding of x in V_2. At this point a severe problem arises. If x has a large number of neighbours (say $y_1, \dots, y_k$) that have already been embedded, it could very well be that the images of the y_i do not have a common neighbour, thus leaving no choice for the embedding of x. We therefore need a more sophisticated embedding scheme.

The indispensable detail to overcome this obstacle lies in the structure of a-arrangeable graphs. As pointed out in Section 1.2.1, the generalised Blow-up Lemma requires the graph H to be a-arrangeable for some constant a (see Theorem 4.3). So let $(x_1, \dots, x_n)$ be an a-arrangeable ordering of the vertices in H. We again start a randomised embedding procedure. This time however we try to ensure that the embedding of one vertex does not block the embedding of its neighbours. To do so we simply forbid the embedding of x_i into vertices that would disproportionately restrict the embedding of a vertex x_j (with $j > i$). This is where the a-arrangeability comes into play. It guarantees that we indeed can find an embedding of x_i that does not disproportionately restrict the embedding of any vertex x_j with $j > i$.

Nevertheless the randomised algorithm is likely to get stuck towards the end of the embedding. By then most of the vertices of G are occupied which might again block all possible embeddings of a remaining vertex of H. Hence, we only embed most of H into the pair (V_1, V_2). We then try to find a simultaneous assignment of all remaining vertices of H to the unoccupied vertices in G. We address the question why such an assignment should exist in the subsequent section.

1.3.3 The auxiliary graph and weighted regularity

In the previous section we have sketched a randomised algorithm for the embedding of H into G. We now argue that in the setting of Section 1.3.2 this algorithm succeeds with positive probability, thereby proving that H is indeed a subgraphs of G. In order to analyse the embedding procedure we define a family of *auxiliary graphs* as follows. For $i = 1, 2$ and $t \in \{1, \dots, n\}$, the auxiliary graph $F_i(t)$ is the bipartite graph on vertex set $X_i \cup V_i$ that contains the following edges: xv is an edge in $F_i(t)$ if and only if after embedding vertices $x_1, \dots, x_{t-1}$ the vertex x can still be embedded into v. This is the case if and only if all neighbours of x that already got embedded got embedded into vertices that are adjacent to v in G. We want to illustrate this definition by the following example. Let H be a cycle on $2n$ vertices and let G be a balanced bipartite graph on the same number of vertices. We use the cyclical ordering of the vertices in H for the embedding. (Note that this ordering is 2-arrangeable.) The auxiliary graph $F_i(0)$ is a complete bipartite graph as every vertex of X_i could be embedded into any vertex of V_i at this point. This changes as we embed x_1 into some vertex $v_1 \in V_1$ thereby restricting the embedding of x_2 (and of x_n) to the neighbours of v_1 in V_2. Hence the neighbourhood of x_2 in $F_2(1)$ is just the neighbourhood of v_1 in V_2. While we embed vertex after vertex the auxiliary graphs keeps track of possible embeddings: At time t we check whether the neighbourhood of x_t in $F_i(t)$ contains vertices into which no vertex has been embedded yet. If this is the case we can embed x_t into those vertices. In particular, we find a spanning embedding of H into G if there are unoccupied vertices in the neighbourhood of x_t in $F_i(t)$ for all $t \in \{1, \dots, n\}$.

Indeed, the above event happens with positive probability. The argument relies on a concept that one could best describe as *weighted regularity*. The auxiliary graphs $F_i(t)$ are not ε-regular in terms of Szemerédi's classical definition as the degree of a vertex $x \in X_i$ strongly depend on the number of its neighbours that already got embedded. However, one can introduce a weight function on the vertices of X_i together with an corresponding definition of density that (potentially) makes those graphs weighted regular. In Section 4.3 we prove that the randomised embedding into super-regular pairs almost surely produces weighted regular auxiliary graphs. Moreover, all auxiliary graphs of a randomised embedding have linear minimum degree with positive probability. If this is the case we obtain perfect matchings in each auxiliary graph. Those perfect matchings at some point of time allow us to embed all remaining vertices of H simultaneously into G thereby guaranteeing the spanning embedding in Section 4.4.

1.3.4 A lemma for G and a lemma for H

With the generalised Blow-up Lemma at hand, we prove an extension of a Conjecture by Bollobás and Komlós [58] (see Theorem 4.9). For this purpose we need finely tuned tools for both the host graph and the target graph. Those we will call *lemma for G* and *lemma for H* respectively. The lemma for G will construct a (super-)regular partition of the graph G (building on the Regularity Lemma as described in Section 1.3.1) with a reduced graph of quite a specific structure. This lemma for G follows from a result by Böttcher, Schacht, and Taraz [18] from their proof of Theorem 1.10. The lemma for H

in turn provides a partition of H that is compatible with the partition of G in view of the embedding task. Here we prove a new version of a result by Böttcher et. al. that is adapted to the potentially unbounded degree of H.

After these preparations we can apply our generalised Blow-up Lemma (see Theorem 4.3) to embed H into G. This application differs from most previous ones in that we do not need to split up the graphs into parts spanned by a fixed number of clusters. Also the connecting via image restrictions becomes obsolete. Our Blow-up Lemma embeds the whole of H in one go.

1.4 Organisation

This thesis is structured as follows.

- Chapter 2 gives an overview of basic definitions and concepts which we use throughout the thesis.

- Chapter 3 introduces Szemerédi's Regularity Lemma in various forms as well as corresponding embedding results for regular partitions. We present the state of research and motivate upcoming results.

- Chapter 4 presents our core result, a Blow-up Lemma for graphs of growing degree. This is joint work with Julia Böttcher, Yoshiharu Kohayakawa, and Anusch Taraz [16].

- Chapter 5 contains an application of the extended Blow-up Lemma – the Conjecture of Bollobás and Komlós for arrangeable graphs. This is joint work with Julia Böttcher and Anusch Taraz [19].

- Chapter 6 and Chapter 7 investigate the size of planar subgraphs in dense graphs. These results are joint work with Oliver Cooley, Tomasz Łuczak, and Anusch Taraz [27] and Peter Allen and Jozef Skokan [3] respectively.

- Chapter 8 proves that almost all graphs without an induced copy of C_5 have a homogeneous set of linear size. This is joint work with Julia Böttcher and Anusch Taraz [21].

- Chapter 9 proves that for any $\varepsilon > 0$ almost all cographs (graphs without an induced copy of P_4) have no homogeneous set of size $n/\log^{1-\varepsilon} n$. This is joint work with Carlos Hoppen and Marc Noy [52].

- Chapter 10 contains some concluding remarks on the problems covered in preceding chapters.

2 Definitions

In this chapter we provide the definitions of most concepts that we shall use throughout the thesis. For all elementary graph theoretic concepts not defined in this chapter we refer the reader to, e.g., [33]. In addition, we defer most definitions that are not used in many chapters to later and only introduce them once we need them. In particular, all terminology concerning the regularity method is introduced in Chapter 3.

2.1 Basic notions

Graphs A *graph* is a tuple $G = (V, E)$ where V is a finite, non-empty set and $E \subseteq \binom{V}{2}$. In particular all graphs in this thesis are finite, simple, and undirected. We denote the vertices of G by $V(G) := V$ and the edges of G by $E(G) := E$. Their number is denoted by $v(G)$ and $e(G)$ respectively. Sometimes we also write $|G|$ for the number of vertices in G and call it the *order* of G. An *n-graph* is a graph of order n. We frequently use the shorthand uv for an edges $\{u, v\} \in E(G)$ and say that it *joins* the vertex u to the vertex v. We also say that u and v are *adjacent* vertices and that the vertex u is *incident* with the edge uv.

For a vertex $v \in V$ we write $N_G(v) := \{u \in V(G) : uv \in E(G)\}$ or simply $N(v)$ for the *neighbourhood* of v in G. Let $A, B \subseteq V$. Then $N(v, B) := N(v) \cap B$ denotes the set of neighbours of v in B. The union of all neighbourhoods of vertices in A is also denoted by $N_G(A) := \bigcup_{v \in A} N_G(v)$ or simply $N(A)$. The *common* or *joint* neighbourhood of two vertices v, w is defined as $N_G(v, w) := N_G(v) \cap N_G(w)$. The *degree* of a vertex $v \in V(G)$ is denoted by $\deg_G(v) := |N_G(v)|$ while $\deg_G(v, B) := |N_G(v) \cap B|$. We may omit the subscript G if there is no doubt about the graph in question.

Let $G = (V_G, E_G)$, $H = (V_H, E_H)$ be two graphs. A *homomorphism* from H to G is a function $\varphi : V_H \to V_G$ with $\{\varphi(v), \varphi(w)\} \in E_G$ for every $\{v, w\} \in E_H$. The graph H is called a *subgraph* of G if there is an injective homomorphism $\varphi : V_H \to V_G$. In this case we also say that G contains a *copy* of H or that φ is an *embedding* of H into G and write $H \subseteq G$. The graph H is an *induced subgraph* of G if there is an injective homomorphism $\varphi : V_H \to V_G$ with $\{\varphi(v), \varphi(w)\} \in E_G$ if and only if $\{v, w\} \in E_H$. We also say that G contains an *induced copy* of H. The set $A \subseteq V$ induces the subgraph $G[A] := (A, V(G) \cap \binom{A}{2})$. A subgraph $H \subseteq G$ is called *spanning* if $|H| = |G|$.

A graph $G = (V, E)$ is called *complete* if $E = \binom{V}{2}$; it is called a *stable set* if $E = \emptyset$. We say a set $W \subseteq V$ is *homogeneous* if it induces a complete graph or a stable set in G. A *partition* of a graph $G = (V, E)$ is a partition of its vertex set into disjoint[1] sets $V = V_1 \dot\cup \ldots \dot\cup V_k$. A partition $V = V_1 \dot\cup \ldots \dot\cup V_k$ is called *balanced* if

[1]The dot in $A \dot\cup B$ indicates that the union of A and B is a union between disjoint sets.

$|V_1| \leq |V_2| \leq \cdots \leq |V_k| \leq |V_1| + 1$. A graph $G = (V, E)$ is called *k-partite* if there is a partition $V = V_1 \cup \ldots \cup V_k$ of its vertices such that $G[V_i]$ is a stable set for all $i \in [k]$. A 2-partite graph is called *bipartite*.

For a graph $G = (V, E)$ and disjoint subsets $A, B \subseteq V$ we define $e(A, B)$ to be the number of edges between A and B. For disjoint nonempty vertex sets $A, B \subseteq V$ the *density* $d(A, B) := e(A, B)/(|A|\,|B|)$ of the pair (A, B) is the number of edges that run between A and B divided by the number of possible edges between A and B.

The graph $P_n = ([n], \{\{i, i+1\} : i \in [n-1]\})$ is called a *path* on n vertices, while $C_n = ([n], \{\{i, i+1\} : i \in [n-1]\} \cup \{\{1, n\}\})$ is called a *cycle* on n vertices. The complete graph on n vertices, which we also call a *clique*, is denoted by K_n, the complete bipartite graph with partition classes of size n and m is denoted by $K_{n,m}$. A *Hamilton cycle* is a spanning subgraph which is a cycle. Let H be a fixed graph. An *F-factor* is a spanning subgraph which consists of vertex disjoint copies of H. A *perfect matching* is a K_2-factor. The k-th *power* of a graph $G = (V, E)$ is denoted by G^k. It is a graph on the same vertex set V in which two vertices x and y are adjacent if and only if there is a path from x to y with length at most k in G.

A graph G is called *connected* if any two vertices are linked by a path in G. The vertex maximal connected subgraphs of G are called the *connected components* of G. A graph is called *acyclic* if it does not contain a subgraph which is a cycle. An acyclic, connected graph is a *tree* and the disjoint union of trees is a *forest*. Vertices of degree 1 in a tree are *leaves*. A tree in which all but at most one vertex are leaves is called a *star*.

A graph is called *planar* if it can be drawn into the plane in such a way that its edges do not cross. A *plane* graph is a graph embedded into the plane together with its embedding. The regions that arise from this embedding are called *faces*. A plane graph G is a *triangulation* if all its faces are bounded by triangles and a *quadrangulation* if all faces are bounded by 4-cycles. Note that a triangulation on n vertices has $3n - 6$ edges, while a quadrangulation has $2n - 4$ edges by *Euler's formula*. (Euler's formula for polyhedra states that any plane graph satisfies $v - e + f = 2$, where v, e, f denote the number of vertices, edges, faces in the plane embedding.)

A graph G is called *perfect* if $\omega(H) = \chi(H)$ for every induced subgraph H of G. The strong perfect graph theorem [23] asserts that perfect graphs are exactly those graphs that do not contain an induced subgraph which is a cycle C_{2k+1} or the complement of a cycle $\overline{C}_{2k+1}$ for some $k \geq 2$.

Graph parameters We use the following basic graph parameters: the *minimum degree* $\delta(G)$, the *maximum degree* $\Delta(G)$, and the *average degree* $d(G)$. The maximum size of a clique in G or the *clique number* of G is denoted by $\omega(G)$, the maximum size of a stable set in G or the *independence number* by $\alpha(G)$. The maximum size of a homogeneous set in G is denoted by $\hom(G) := \max\{\alpha(G), \omega(G)\}$. The *chromatic number* of G is the minimum integer k such that G is k-partite. It is denoted by $\chi(G)$.

The two subsequent parameters will play an important rôle throughout the thesis.

Definition (a-arrangeable)
Let a be an integer. A graph is called a-arrangeable *if its vertices can be ordered as*

$(x_1, \dots, x_n)$ *in such a way that* $\big|N\big(N(x_i, \mathrm{Right}_i), \mathrm{Left}_i\big)\big| \le a$ *for each* $1 \le i \le n$*, where* $\mathrm{Left}_i = \{x_1, x_2, \dots, x_i\}$ *and* $\mathrm{Right}_i = \{x_{i+1}, x_{i+2}, \dots, x_n\}$.

Definition (Bandwidth)
Let $G = (V, E)$ *be a graph on* n *vertices. The* bandwidth *of* G *is denoted by* $\mathrm{bw}(G)$ *and defined to be the minimum positive integer* b*, such that there exists a labelling of the vertices in* V *by numbers* $1, \dots, n$ *so that the labels of every pair of adjacent vertices differ by no more than* b*.*

The notion of arrangeability was introduced by Chen and Schelp in [22]. It generalises the concept of bounded degree: every graph with maximum degree Δ is $(\Delta(\Delta-1)+1)$-arrangeable while even 1-arrangeable graphs such as stars may have arbitrarily large degree. Chen and Schelp showed that planar graphs are 761-arrangeable which was later improved to 10-arrangeable by Kierstead and Trotter [57]. Graphs with bounded arrangeability exhibit several benign properties when it comes to embedding.

Key results of this thesis require graphs to have sublinear bandwidth. One can think of this restriction as some locality constraint: the neighbours of every vertex lie close to it in such a bandwidth ordering. Partitioning a graph along a bandwidth ordering will thus ensure that edges only run between consecutive partition classes. Böttcher, Pruessmann, Taraz, and Würfl [17] give several sufficient conditions for sublinear bandwidth.

Asymptotics For asymptotic notation we use the *Landau symbols* $\mathcal{O}(f(n))$, $\mathbf{o}(f(n))$, $\Omega(f(n))$ and $\Theta(f(n))$. The symbol $\ll$ is used exclusively for the relation among constants. For two positive real numbers $\varepsilon, \varepsilon'$ we write $\varepsilon' \ll \varepsilon$ to express that $\varepsilon' \le \varepsilon$ and that we can choose ε' smaller than any given positive ε.

Other We denote the set of positive integers by $\mathbb{N}$ and set $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. For any integer n we write $[n] := \{1, \dots, n\}$. For a real number x we denote by $\lfloor x \rfloor$ the greatest integer $\le x$, and by $\lceil x \rceil$ the least integer $\ge x$. Throughout this paper we omit floors and ceilings whenever this does not affect the argument. We write $x = a \pm b$ as a shorthand for $x \in [a-b, a+b]$. Unless we explicitly state a base all logarithms written as 'log' refer to the natural logarithm, i.e., are taken at base e.

2.2 Random variables and random graphs

The notation and terminology we use for probabilities, random variables, and related concepts is standard (in the theory of random graphs) and follows [54].

Some random variables that will play an important rôle throughout this thesis are *binomially distributed.* The properties of this distribution are very well understood. So let $\mathcal{X}$ be a binomially distributed random variable with parameters n and p. There is a variety of concentration results for $\mathcal{X}$ which often go under the name *Chernoff bound.*

We will mostly use the following inequality which is true for any $c \in [0, 3/2]$.

$$\mathbb{P}[|\mathcal{X} - pn| \geq c \cdot pn] \leq \exp\left(-\frac{c^2}{3}pn\right).$$

A *random graph* in the *Erdős-Renyí model* [43] denoted by $\mathcal{G}(n,p)$ is generated by including each of the $\binom{n}{2}$ possible edges on n vertices with probability p independently at random. By linearity of expectation, the expected number of edges incident to a vertex of $\mathcal{G}(n,p)$ equals $(n-1)p$. Therefore, the parameter p governs the average degree, or what we call the *density* of the graph.

A class of graphs that is closed under isomorphism is called a *graph property*. A *monotone* graph property is closed under taking subgraphs while a *hereditary* graph property is closed under taking induced subgraphs.

We say that a random graph $\mathcal{G}(n,p)$ has a property $\mathcal{P}$ *asymptotically almost surely* (abbreviated *a.a.s.*) if the probability that $\mathcal{G}(n,p) \in \mathcal{P}$ tends to 1 as n tends to infinity.

The Erdős-Renyí model has several convenient properties. If follows from Chernoff's inequality that $\mathcal{G}(n,p)$ for $0 < p < 1$ asymptotically almost surely has the following properties.

(i) every vertex has degree $(p + o(1))n$,

(ii) every k-tuple of vertices has $(p^k + o(1))n$ common neighbours,

(iii) every subset which is significantly larger than $\log n$ has density approximately p,

(iv) every fixed subgraph H is contained in $\mathcal{G}(n,p)$.

This may give an idea as to why the Erdős-Renyí model is so popular in random graph theory.

We would like to point out that Erdős and Renyí [42] initially proposed the random graph model $\mathcal{G}(n,m)$. A random graph in the $\mathcal{G}(n,m)$-model is drawn uniformly at random from the class of graphs with n vertices and m edges. The $\mathcal{G}(n,p)$- and the $\mathcal{G}(n,m)$-model for $m = p\binom{n}{2}$ are closely related as we shall see, e.g., in Section 8.4.

Other random graph models are defined by families of graphs. For an (infinite) family of graphs $\mathcal{F}$ let $\mathcal{F}_n$ be a graph drawn uniformly at random among all graphs of order n in $\mathcal{F}$. Much less than for $\mathcal{G}(n,p)$ is known about these graph models in general. As an example let us shortly discuss random planar graphs, i.e., the case where $\mathcal{F}$ is the set of all planar graphs. It has only recently been shown that the maximum degree of $\mathcal{F}_n$ asymptotically almost surely is of order $\Theta(\log n)$. At the same time the minimum degree a.a.s. is as low as 1. Many of the other facts known about $\mathcal{G}(n,p)$ are either not true or remain unknown.

3 The regularity method

We have laid out in the introduction how Szemerédi's Regularity Lemma [90] has paved the road for many important advances in extremal combinatorics. It originated in a proof of the existence of arithmetic progressions in sets with positive density, but has since then grown into a powerful framework for extremal problems of all kinds. In particular, it has been the key instrument for the solution of a number of long-standing open problems (such as, e.g., Theorem 1.6, Theorem 1.7, Theorem 1.8) in extremal graph theory (see the surveys [69, 72] for a more detailed account). Little surprising, virtually all results in this thesis build on the regularity method. Hence we now want to give a detailed introduction into the topic.

In this chapter we will state the Regularity Lemma (Section 3.1) and some immediate consequences which we group under the keywords regularity (Section 3.1) and super-regularity (Section 3.2). We continue with embedding results such as the Blow-up Lemma (Section 3.3) and conclude the chapter with some remarks on the concept of weighted regularity (Section 3.4).

3.1 Regular partitions of graphs

The Regularity Lemma talks about regular partitions and, in particular, about regular pairs. We define those concepts before we state the lemma. For basic graph theoretic definitions see Chapter 2. For the rest of this chapter, let $\varepsilon, d \in (0, 1]$ unless noted otherwise.

Definition 3.1 (ε-regular)
The pair (A, B) is said to be ε-regular *if, for all $X \subseteq A$, $Y \subseteq B$ with $|X| \geq \varepsilon|A|$, $|Y| \geq \varepsilon|B|$, one has $|d(X, Y) - d(A, B)| < \varepsilon$; otherwise (A, B) is said to be* ε-irregular. *An ε-regular pair (A, B) is called* (ε, d)-regular*, if it has density at least d.*

An ε-regular pair is also called *quasi-random* because regular pairs exhibiti properties similar to the ones of random graphs in the $\mathcal{G}(n, p)$-model stated in Section 2.2. There are sevxeral ways to define a regular partition. We present the original one by Szemerédi from [90].

Definition 3.2 (ε-regular partition)
Let $G = (V, E)$. An equitable partition *of V is a partition into pairwise disjoint sets $V_0, V_1, \ldots, V_k$ such that $|V_1| = \cdots = |V_k|$. The set V_0 is said to be* exceptional*, and one might have $V_0 = \emptyset$. Such an equitable partition of V is said to be an* ε-regular partition *if $|V_0| \leq \varepsilon|V|$ and not more than εk^2 of the pairs (V_i, V_j) are ε-irregular in G, where $1 \leq i < j \leq k$.*

With these definitions set Szemerédi's Regularity Lemma can be stated in two lines.

Lemma 3.3 (Regularity Lemma, Szemerédi [90])
For all $\varepsilon > 0$ and m there is M and N such that every graph $G = (V, E)$ on $n \geq N$ vertices has an ε-regular partition $V = V_0 \mathbin{\dot\cup} V_1 \mathbin{\dot\cup} \ldots \mathbin{\dot\cup} V_k$ with $m \leq k \leq M$.

The crucial point here is that the upper bound M does not depend on the order n but only on ε. However, $N \geq M$ and M grows rapidly with ε tending to 0. Proofs of the Regularity Lemma bound M by a tower of 2s with height proportional to ε^{-5} (and ε^{-5} cannot be replaced by anything better than ε^{-1} as was recently shown by Conlon and Fox [26]). As a consequence, results obtained with the help of this lemma typically talk about huge graphs only. And by *huge* graphs we mean graphs whose order easily exceeds the number of atoms in the universe. This is also the scale for most results of this thesis.

We will use the original version of the Regularity Lemma (Lemma 3.3) in Chapter 8. For other purposes it is convenient to have slightly different versions. Those might seem stronger at first glance, however all versions of the Regularity Lemma presented here directly imply each other. The following degree form of the Regularity Lemma can be found, e.g., in [69, Theorem 1.10]. It will be used in Chapter 6.

Lemma 3.4 (Regularity Lemma, degree form)
For every $\varepsilon, d > 0$ there is an n_0 such that the following is true for every graph $G = (V, E)$ on $n \geq n_0$ vertices. There is an equitable partition $V = V_0 \mathbin{\dot\cup} V_1 \mathbin{\dot\cup} \ldots \mathbin{\dot\cup} V_k$ and a subgraph $H \subseteq G$ such that

(i) $k \leq n_0$,

(ii) $|V_0| \leq \varepsilon |V|$,

(iii) $H[V_i]$ *is an empty graph for every* $i \in [k]$,

(iv) $H[V_i \mathbin{\dot\cup} V_j]$ *is* (ε, d)*-regular or empty for every* $i, j \in [k]$,

(v) $\deg_G(v) - \deg_H(v) \leq (d + \varepsilon)n$ *for all* $v \in V$.

Before we state a third form of the Regularity Lemma we introduce some more notation. The concept of a *reduced graph* captures the structure of a regular partition. One can think of it as a blueprint for the later embedding.

Definition 3.5 ((ε, d)-reduced graph)
*Let $G = (V, E)$ have the ε-regular partition $V = V_0 \mathbin{\dot\cup} V_1 \mathbin{\dot\cup} \ldots \mathbin{\dot\cup} V_k$. Then the (ε, d)-*reduced graph *R of G is defined as the graph on vertex set $[k]$ that has an edge ij if and only if (V_i, V_j) is an (ε, d)-regular pair. We also say that G is (ε, d)-regular* on *R.*

The sets V_i with $i \in [k]$ are also called the *clusters* of the partition; we may occasionally simplify notation and call a vertex i of the reduced graph a cluster and identify it with its corresponding set V_i. The reduced graphs R inherits some properties from G. The subsequent lemma states that minimum degree is one such property. This third form of the Regularity Lemma can be found, e.g., in [74, Proposition 9]. We will use it in Chapter 7.

Lemma 3.6 (Regularity Lemma, minimum degree form)
For all ε, d, γ with $0 < \varepsilon < d < \gamma < 1$ and for all m there is M such that every graph G on $n > M$ vertices with $\delta(G) \geq \gamma n$ has an (ε, d)-reduced graph R on k vertices with $m_0 \leq k \leq M$ and $\delta(R) \geq (\gamma - d - \varepsilon)k$.

At this point we want to emphasize that we can require the exceptional set V_0 in a regular partition to be empty. Since n is not necessarily divisible by the number of clusters k we then cannot expect the partition $V = V_1 \dot\cup \ldots \dot\cup V_k$ to be equitable any more. However, we can demand $|V_1| \leq |V_2| \leq \cdots \leq |V_k| \leq |V_1| + 1$. Recall that such a partition is a *balanced partition*. The following equi-partite version of Lemma 3.3 can be found in [68, Theorem 2].

The remainder of this section is dedicated to some convenient properties of regular pairs. The first of these can be found, e.g., as [7, Lemma 3.1] and says that subpairs of dense regular pairs are dense and regular.

Lemma 3.7 (Slicing lemma)
For every $0 < \varepsilon < d, \alpha$ the following is true. Let (A, B) be an (ε, d)-regular pair and let $A' \subseteq A$, $B' \subseteq B$ satisfy $|A'| \geq \alpha|A|$ and $|B'| \geq \alpha|B|$. Then (A', B') is an $(\varepsilon', d - \varepsilon)$-regular pair where $\varepsilon' = \max\{2\varepsilon, \varepsilon/\alpha\}$.

The second property says that an even degree distribution is to some point equivalent to ε-regularity. We define the *co-degree* of v and w in G to be $\deg_G(v, w) := |N_G(v) \cap N_G(w)|$. As we have pointed out in Section 1.3.1, most vertices in a regular pair have a degree close to the average degree. We make this more precise now. So let (A, B) be an (ε, d)-regular pair. Then all but an ε-fraction of the vertices in A have degree at least $(d-\varepsilon)|B|$ into B. Similarly, only a small fraction of the pairs $(v, w) \in A \times A$ have co-degree $\deg(v, w) < (d-\varepsilon)^2|B|$. In that sense ε-regularity forces an even distribution of degrees and co-degrees in the pair. Interestingly, the converse is also true. The subsequent statement (see, e.g., [35, Proposition 2.5]) says that an even degree and co-degree distribution implies a certain degree of regularity.

Lemma 3.8 (Degree/co-degree lemma)
Let $G = (A \dot\cup B, E)$ be a bipartite graph, $|A| = |B| = n$, and let at least $(1-5\varepsilon)n^2$ pairs of vertices $(v, w) \in A \times A$ satisfy

$$\deg_G(v), \deg_G(w) \geq (d - \varepsilon)n \ , \text{ and}$$
$$\deg_G(v, w) < (d + \varepsilon)^2 n \, ,$$

then G is $(16\varepsilon)^{1/5}$-regular.

This *degree/co-degree characterisation* of regular pairs was introduced by Alon, Duke, Lefmann, Rödl, and Yuster [6] in their proof of an algorithmic version of the Regularity Lemma. We will employ the degree/co-degree characterisation to prove regularity for the auxiliary graphs in Chapter 4.

3.2 Super-regularity

We have already stated the Blow-up Lemma (Theorem 1.16) of Komlós, Sarközy, and Szemerédi in the introduction. This result builds on the concept of *super-regular pairs* which we now define. Roughly speaking, a regular pair is super-regular if every vertex has a sufficiently large degree.

Definition 3.9 (super-regular pair)
An ε-regular pair (A, B) in a graph $G = (V, E)$ is (ε, d)-super-regular *if every vertex $v \in A$ has degree* $\deg(v, B) \geq d|B|$ *and every $v \in B$ has* $\deg(v, A) \geq d|A|$. *A partition $V = V_0 \cup V_1 \cup \ldots \cup V_k$ is (ε, d)-super-regular on a graph $R = ([k], E_R)$, if all pairs (V_i, V_j) with $ij \in E_R$ are (ε, d)-super-regular.*

An ε-regular pair may be far from (ε, δ)-super-regular in the sense that it might contain isolated vertices. However, there are not many vertices of small degree in a regular pair.

Proposition 3.10
Let (A, B) be an (ε, d)-regular pair and B' be a subset of B of size at least $\varepsilon|B|$. Then there are at most $\varepsilon|A|$ vertices v in A with $|N(v) \cap B'| < (d - \varepsilon)|B'|$.

Proof. Let $A' = \{v \in A : |N(v) \cap B'| < (d - \varepsilon)|B'|\}$. It follows that $d(A', B') < ((d - \varepsilon)|A'||B'|)/(|A'||B'|) = d - \varepsilon$. As (A, B) is (ε, d)-regular and $|B'| \geq \varepsilon|B|$ we conclude that $|A'| < \varepsilon|A|$. □

Applying Proposition 3.10 to all (ε, d)-regular pairs of a partition now gives a subgraphs which is super-regular on R. Note that we remove an ε-fraction of the vertices for each edge of R that is incident to the cluster. Hence the subgraph could be empty unless $\Delta(R) < 1/\varepsilon$. We will only consider reduced graphs of bounded degree for this reason. The following proposition can be found in [74, Proposition 8].

Proposition 3.11
Given $\varepsilon, d > 0$ and $\Delta \in \mathbb{N}$ set $\varepsilon' := 2\varepsilon\Delta/(1 - \varepsilon\Delta)$ and $d' := d - 2\varepsilon\Delta$. Let G have the balanced (ε, d)-regular partition $V_1 \cup \ldots \cup V_k$ with reduced graph R and let R' be a subgraph of R with $\Delta(R') \leq \Delta$. Then there are subsets $V_i' \subseteq V_i$ with $|V_i'| \geq (1 - \varepsilon\Delta)|V_i|$ for every $i \in [k]$ such that $G[V_i', V_j']$ is (ε', d')-super-regular for every $ij \in R'$.

The concept of super-regularity will be central in the following section. Unfortunately, even graphs with high minimum degree do not have a super-regular partition in general. However, a dense graph can be partitioned such that an almost spanning subgraph is super-regular on a bounded degree subgraph of the reduced graph. The few vertices that are not part of this almost spanning subgraph are then moved to the exceptional set V_0.

3.3 Embedding lemmas

We now turn to embedding lemmas for regular pairs and/or partitions. We start with two elementary result for the embedding of graphs with fixed number of vertices. Then we move on to spanning embeddings and the Blow-up Lemma (Section 3.3.1). Finally we present an almost-spanning embedding result that does not require a super-regular partition (Section 3.3.2).

The subsequent lemma is a standard embedding result for regular partitions and follows easily, e.g. from [68, Theorem 14].

Lemma 3.12 (Embedding lemma)
For every $d > 0$ and every integer k there exists $\varepsilon > 0$ with the following property. Let H be a graph on k vertices $v_1, \dots, v_k$. Let G be a graph. Let $V_1, \dots, V_k$ be clusters of an (ε, d)-regular partition of G with reduced graph $R = ([k], E_R)$. If there is a homomorphism from H to R, then G contains a copy of H.

In Chapter 8 we will be interested in the embedding of induced subgraphs. Dense regular pairs allow the embedding of induced subgraphs under the natural additional assumption that they are not too dense. This means that, in an ε-regular pair we can embed edges and "non-edges" as long as its density is bounded away from 0 and 1. The following lemma from [7] makes this more precise.

Lemma 3.13 (Induced embedding lemma)
For every $d > 0$ and every integer k there exists $\varepsilon > 0$ such that the following holds. Let H be a graph on k vertices $v_1, \dots, v_k$. Let G be a graph. Let $V_1, \dots, V_k$ be clusters of an ε-regular partition of G with complete reduced graph $R = ([k], \binom{[k]}{2})$. Moreover assume that $d(V_i, V_j) \geq d$ if $v_i v_j \in E(H)$ and $d(V_i, V_j) \leq 1 - d$ if $v_i v_j \notin E(H)$. Then G contains an induced induced copy of H.

3.3.1 The Blow-up Lemma

We have described in the introduction how the Blow-up Lemma is a key tool for many results on spanning subgraphs. We now state the full result by Komlós, Sárközy and Szemerédi [62] (for an alternative proof see [86]).

We first introduce some definitions. Let G, H and R be graphs with vertex sets $V(G)$, $V(H)$ and $V(R) = [r]$. We say that a graph H has an *R-partition* $V(H) = X_1 \dot\cup \dots \dot\cup X_r$, if for every edge $xy \in E(H)$ there are distinct $i, j \in [r]$ with $x \in X_i$, $y \in X_j$ and $ij \in E(R)$. The graph G has a *corresponding (ε, d)-super-regular R-partition* if $V(G) = V_1 \dot\cup \dots \dot\cup V_r$ is (ε, d)-super-regular on R and $|V_i| = |X_i|$ for all $i \in [r]$. These partitions are *balanced* if $n_1 \leq n_2 \leq \dots \leq n_r \leq n_1 + 1$.

Lemma 3.14 (Blow-up Lemma)
Given a graph R of order r and positive parameters d, Δ, c there exist positive constants ε, α such that the following holds. Suppose that H and G are two graphs with the same number of vertices, where $\Delta(H) \leq \Delta$ and H has a balanced R-partition $V(H) = X_1 \dot\cup \dots \dot\cup X_r$, and G has a corresponding (ε, d)-super-regular R-partition

$V(G) = V_1 \cup \ldots \cup V_r$. Further suppose that in each class X_i there is a set of at most αn_i special vertices y, each of which is equipped with a candidate set $C_y \subseteq V_i$ with $|C_y| \geq c\,n_i$. Then there is an embedding of H into G such that each special vertex is mapped to a vertex in its candidate set.

We also say that the special vertices y in Lemma 3.14 are *image restricted* to C_y. Thus Lemma 3.14 not only embeds a spanning subgraph H into a graph with a corresponding super-regular partition. It also allows us to control the embedding of a linear number of vertices.

3.3.2 Almost spanning subgraphs

If one is only interested in almost spanning subgraphs, i.e., in graphs that span a $(1-\mu)$-fraction of the host graph, the condition on the super-regularity can be dropped. For an almost spanning embedding it suffices that the pairs are (ε, d)-regular (as opposed to (ε, d)-super-regular for Lemma 3.14).

Lemma 3.15 (Almost spanning Blow-up Lemma)
Given a graph R of order r and positive parameters d, Δ, c, μ there exist constants $\varepsilon, \alpha > 0$ such that the following holds. Suppose that H and G are two graphs, where $\Delta(H) \leq \Delta$ and H has a balanced R-partition $V(H) = X_1 \cup \ldots \cup X_r$, and G has a balanced (ε, d)-super-regular R-partition $V(G) = V_1 \cup \ldots \cup V_r$. Further let $|X_i| = (1-\mu)n_i$ and $|V_i| = n_i$ for all $i \in [r]$ and suppose that in each class X_i there is a set of at most αn_i special vertices y, each of which is equipped with a candidate set $C_y \subseteq V_i$ with $|C_y| \geq c\,n_i$. Then there is an embedding of H into G such that each special vertex is mapped to a vertex in its candidate set.

Proof (sketch). Lemma 3.15 follows easily from Lemma 3.14 and Proposition 3.11. Given the graph R we choose ε small enough for Proposition 3.11 to guarantee the existence of a subgraph on a $(1-\mu)$-fraction of the vertices which is super-regular on R. We then apply the Blow-up Lemma (Lemma 3.14) to embed H into this subgraph.$\square$

3.4 Weighted regularity

We want to mention that the concept of ε-regularity can be extended to hypergraphs (see [81]). As in the graph case an algorithmic version of the Regularity Lemma for hypergraphs builds upon a degree/co-degree characterisation. Czygrinow and Rödl [32] establish such a degree/co-degree characterisation of regularity for hypergraphs. The hypergraph world, however, is more delicate. To address these difficulties they introduce the concept of *weighted regularity*. In the following we present a (simplified) version of their concept for the graph case. (For weighted regularity in graphs see also [31].)

So let $G = (V, \omega)$ be a graph with weight function $\omega : V \times V \to \mathbb{N}_0$, and let $K = 1 + \max\{\omega(v_1, v_2) : (v_1, v_2) \in V \times V\}$. For disjoint V_1, V_2 we define the *weighted* *density* as

$$d_\omega(V_1, V_2) = \frac{\sum \omega(v_1, v_2)}{K|V_1|\,|V_2|}\,,$$

where the sum is over all pairs $(v_1, v_2) \in V_1 \times V_2$. A pair (V_1, V_2) with $V_1 \cap V_2 = \emptyset$ is called (ε, ω)-*regular* if all subsets $W_i \subseteq V_i$, $i = 1, 2$ with $|W_i| \geq \varepsilon|V_i|$ satisfy

$$|d_\omega(V_1, V_2) - d_\omega(W_1, W_2)| < \varepsilon\,.$$

Weighted regular pairs are very similar to regular pairs in many ways. For one, most vertices in a weighted pair have a weighted degree close to the average weighted degree of the pair. Moreover, there also is a weighted version of Lemma 3.7.

Lemma 3.16 (Slicing lemma)
Let $d, \varepsilon > 0$ and let $G = (V_1 \,\dot\cup\, V_2, \omega)$ with $\omega : V_1 \times V_2 \to \mathbb{N}_0$ be an (ε, ω)-regular pair with weighted density d. Further let $W_1 \subseteq V_1$, $W_2 \subseteq V_2$ with $|W_i| \geq \gamma|V_i|$ for some $\gamma \geq \varepsilon$. Then $(W_1 \,\dot\cup\, W_2, \omega_{|W_1 \times W_2})$ is an $(\varepsilon', \omega_{|W_1 \times W_2})$-regular pair with weighted density d' where $\varepsilon' = \max\{2\varepsilon, \varepsilon/\gamma\}$ and $|d' - d| \leq \varepsilon$. Here $\omega_{|W_1 \times W_2}$ is the restriction of ω to $W_1 \times W_2$.

Proof. The definition of (ε, ω)-regularity implies $|d_\omega(W_1, W_2) - d_\omega(V_1, V_2)| \leq \varepsilon$ hence $(W_1 \,\dot\cup\, W_2, \omega_{|W_1 \times W_2})$ has the claimed density. Moreover, all subsets $W_1' \subseteq W_1$, $W_2' \subseteq W_2$ with $|W_i'| \geq (\varepsilon/\gamma)|W_i| \geq \varepsilon|V_i|$, $i = 1, 2$ satisfy $|d_\omega(W_1', W_2') - d_\omega(W_1, W_2)| \leq 2\varepsilon$ as

$$|d_\omega(W_1', W_2') - d_\omega(W_1, W_2)| \leq |d_\omega(W_1', W_2') - d_\omega(V_1, V_2)| + |d_\omega(V_1, V_2) - d_\omega(W_1, W_2)|$$

again by the (ε, ω)-regularity of $(V_1 \,\dot\cup\, V_2, \omega)$. □

Very much like standard regularity weighted regularity is closely linked to the distribution of degrees and co-degrees in the pair. These two parameters are defined in a straight forward fashion.

Definition 3.17 (Weighted degree and co-degree)
Let $G = (V_1 \,\dot\cup\, V_2, \omega)$ with $\omega : V_1 \times V_2 \to \mathbb{N}_0$ be a weighted bipartite graph. For $x, y \in V_1$ we define the weighted degree *of x as $\deg_\omega(x) := \sum_{z \in V_2} \omega(x, z)$ and the* weighted co-degree *of x and y as $\deg_\omega(x, y) := \sum_{z \in V_2} \omega(x, z)\omega(y, z)$.*

With this definition at hand we can state an analogon of Lemma 3.8 for weighted graphs; the following result is an easy corollary of [32, Lemma 4.2].

Lemma 3.18 (Weighted degree/co-degree lemma)
Let $\varepsilon > 0$ and $n \geq \varepsilon^{-6}$. Further let $G = (V_1 \,\dot\cup\, V_2, \omega)$ with $\omega : V_1 \times V_2 \to \mathbb{N}_0$ be a weighted bipartite graph with $|V_1| = |V_2| = n$. If

(i) $|\{x \in V_1 : |\deg_\omega(x) - d_\omega(V_1, V_2)n| > \varepsilon^{12}n\}| < \varepsilon^{12}n$ *and*

(ii) $|\{(x, y) \in V_1 \times V_1 : |\deg_\omega(x, y) - d_\omega(V_1, V_2)^2 n| \geq \varepsilon^6 n^2\}| \leq \varepsilon^6 n^2$

then (V_1, V_2) is a $(3\varepsilon, \omega)$-regular pair.

We will apply the concept of weighted regularity to auxiliary graphs that arise from a randomized embedding algorithm in Chapter 4. The nature of this algorithm is such that the degree of a vertex in the auxiliary graph drops significantly any time a neighbour of that vertex is embedded. The degrees in the auxiliary graphs differ widely in general for this reason. Hence it will be convenient to balance out this effect by a weight function that increases the weight of a vertex any time a predecessor of the vertex has been embedded.

4 A Blow-up Lemma for arrangeable graphs

4.1 Introduction

The last 15 years have witnessed an impressive series of results guaranteeing the presence of spanning subgraphs in dense graphs. In this area, the so-called *Blow-up Lemma* has become one of the key instruments. It emerged out of a series of papers by Komlós, Sárközy, and Szemerédi (see e.g. [60, 61, 62, 63, 64, 65, 66]) and asserts, roughly spoken, that we can find bounded degree spanning subgraphs in ε-regular pairs. It was used for determining, among others, sufficient degree conditions for the existence of F-factors, Hamilton paths and cycles and their powers, spanning trees and triangulations, and graphs of sublinear bandwidth in graphs, digraphs and hypergraphs (see the survey [72] for an excellent overview of these and related achievements). In this way, the Blow-up Lemma has reshaped extremal graph theory.

However, with very few exceptions, the embedded spanning subgraphs H considered so far came from classes of graphs with constant maximum degree, because the Blow-up Lemma requires the subgraph it embeds to have constant maximum degree. In fact, the Blow-up Lemma is usually the only reason why the proofs of the above mentioned results only work for such subgraphs.

The central purpose of this paper is to overcome this obstacle. We shall provide extensions of the Blow-up Lemma that can embed graphs whose degrees are allowed to grow with the number of vertices. These versions require that the subgraphs we embed are arrangeable.[1] We will formulate them in the following and subsequently present some applications.

Blow-up Lemmas. We first introduce some notation. Let G, H and R be graphs with vertex sets $V(G)$, $V(H)$, and $V(R) = \{1,\dots,r\} =: [r]$. For $v \in V(G)$ and $S, U \subseteq V(G)$ we define $N(v,S) := N(v) \cap S$ and $N(U,S) = \bigcup_{v\in U} N(v,S)$. Let $A, B \subset V(G)$ be non-empty and disjoint, and let $\varepsilon, \delta \in [0,1]$. The *density* of the pair (A,B) is defined to be $d(A,B) := e(A,B)/(|A||B|)$. The pair (A,B) is *ε-regular*, if $|d(A,B) - d(A',B')| \le \varepsilon$ for all $A' \subseteq A$ and $B' \subseteq B$ with $|A'| \ge \varepsilon|A|$ and $|B'| \ge \varepsilon|B|$. An ε-regular pair (A,B) is called *(ε,δ)-regular*, if $d(A,B) \ge \delta$ and *(ε,δ)-super-regular*, if $|N(v,B)| \ge \delta|B|$ for all $v \in A$ and $|N(v,A)| \ge \delta|A|$ for all $v \in B$. We say that H has an *R-partition* $V(H) = X_1 \,\dot\cup \dots \dot\cup\, X_r$, if for every edge $xy \in E(H)$ there are distinct

[1] We remark that it was already suggested in [58] to relax the maximum degree constraint to arrangeability.

$i, j \in [r]$ with $x \in X_i$, $y \in X_j$ and $ij \in E(R)$. G has a *corresponding* (ε, δ)-*super-regular R-partition* $V(G) = V_1 \dot\cup \ldots \dot\cup V_r$, if $|V_i| = |X_i| =: n_i$ for all $i \in [r]$ and every pair (V_i, V_j) with $ij \in E(R)$ is (ε, δ)-super-regular. In this case R is also called the *reduced graph* of the super-regular partition. Moreover, these partitions are *balanced* if $n_1 \le n_2 \le \cdots \le n_r \le n_1 + 1$. They are κ-*balanced* if $n_j \le \kappa n_i$ for all $i, j \in [r]$. The partition classes V_i are also called *clusters*.

With this notation, a simple version of the Blow-up Lemma of Komlós, Sárközy, and Szemerédi [62] can now be formulated as follows.

Theorem 4.1 (Blow-up Lemma [62])
Given a graph R of order r and positive parameters δ, Δ, there exists a positive $\varepsilon = \varepsilon(r, \delta, \Delta)$ such that the following holds. Suppose that H and G are two graphs with the same number of vertices, where $\Delta(H) \le \Delta$ and H has a balanced R-partition, and G has a corresponding (ε, δ)-super-regular R-partition. Then there exists an embedding of H into G.

We remark that Rödl and Ruciński [86] gave a different proof for this result. In addition, Komlós, Sárközy, and Szemerédi [63] gave an algorithmic proof.

Our first result replaces the restriction on the maximum degree of H in Theorem 4.1 by a restriction on its arrangeability. This concept was first introduced by Chen and Schelp in [22].

Definition 4.2 (a-arrangeable)
*Let a be an integer. A graph is called a-*arrangeable *if its vertices can be ordered as $(x_1, \ldots, x_n)$ in such a way that $\big|N\big(N(x_i, \mathrm{Right}_i), \mathrm{Left}_i\big)\big| \le a$ for each $1 \le i \le n$, where* $\mathrm{Left}_i = \{x_1, x_2, \ldots, x_i\}$ *and* $\mathrm{Right}_i = \{x_{i+1}, x_{i+2}, \ldots, x_n\}$.

Obviously, every graph H with $\Delta(H) \le a$ is $(a^2 - a + 1)$-arrangeable. Other examples for arrangeable graphs are planar graphs: Chen and Schelp showed that planar graphs are 761-arrangeable [22]; Kierstead and Trotter [57] improved this to 10-arrangeable. In addition, Rödl and Thomas [87] showed that graphs without K_s-subdivision are s^8-arrangeable. On the other hand, even 1-arrangeable graphs can have unbounded degree (e.g. stars).

Theorem 4.3 (Arrangeable Blow-up Lemma)
Given a graph R of order r, a positive real δ and a natural number a, there exists a positive real $\varepsilon = \varepsilon(r, \delta, a)$ such that the following holds. Suppose that H and G are two graphs with the same number of vertices, where H is a-arrangeable, $\Delta(H) \le \sqrt{n}/\log n$ and H has a balanced R-partition, and G has a corresponding (ε, δ)-super-regular R-partition. Then there exists an embedding of H into G.

Komlós, Sárközy, and Szemerédi proved that the Blow-up Lemma allows for the following strengthenings that are useful in applications. We allow the clusters to differ in size by a constant factor and we allow certain vertices of H to restrict their image in G to be taken from an a priori specified set of linear size. However, in contrast to the original Blow-up Lemma, we need to be somewhat more restrictive about the image restrictions: We still allow linearly many vertices in each cluster to have image restrictions, but now only a constant number of different image restrictions is permissible in each cluster (we shall show in Section 4.5 that this is best possible). In the following, we state an extended version of the Blow-up Lemma that makes this precise.

Theorem 4.4 (Arrangeable Blow-up Lemma, full version)
For all $C, a, \Delta_R, \kappa \in \mathbb{N}$ and for all $\delta, c > 0$ there exist $\varepsilon, \alpha > 0$ such that for every integer r there is n_0 such that the following is true for every $n \geq n_0$. Assume that we are given

(i) a graph R of order r with $\Delta(R) < \Delta_R$,

(ii) an a-arrangeable n-vertex graph H with maximum degree $\Delta(H) \leq \sqrt{n}/\log n$, together with a κ-balanced R-partition $V(H) = X_1 \dot\cup \ldots \dot\cup X_r$,

(iii) a graph G with a corresponding (ε, δ)-super-regular R-partition $V(G) = V_1 \dot\cup \ldots \dot\cup V_r$ with $|V_i| = |X_i| =: n_i$,

(iv) for every $i \in [r]$ a set $S_i \subseteq X_i$ of at most $|S_i| \leq \alpha n_i$ image restricted vertices, *such that $|N_H(S_i) \cap X_j| \leq \alpha n_j$ for all $ij \in E(R)$,*

(v) and for every $i \in [r]$ a family $\mathcal{I}_i = \{I_{i,1}, \ldots, I_{i,C}\} \subseteq 2^{V_i}$ of permissible image restrictions, *of size at least $|I_{i,j}| \geq cn_i$ each, together with a mapping $I\colon S_i \to \mathcal{I}_i$, which assigns a permissible image restriction to each image restricted vertex.*

Then there exists an embedding $\varphi\colon V(H) \to V(G)$ such that $\varphi(X_i) = V_i$ and $\varphi(x) \in I(x)$ for every $i \in [r]$ and every $x \in S_i$.

As we shall show, the upper bound on the maximum degree of H in Theorem 4.4 is optimal up to the log-factor (see Section 4.5). However, if we require additionally that every $(a+1)$-tuple of G has a big common neighbourhood then this degree bound can be relaxed to $o(n/\log n)$.

Theorem 4.5 (Arrangeable Blow-up Lemma, extended version)
Let $a, \Delta_R, \kappa \in \mathbb{N}$ and $\iota, \delta > 0$ be given. Then there exist $\varepsilon, \xi > 0$ such that for every r there is $n_0 \in \mathbb{N}$ such that the following holds for every $n \geq n_0$.

Assume that we are given a graph R of order r with $\Delta(R) < \Delta_R$, an a-arrangeable n-vertex graph with $\Delta(H) \leq \xi n/\log n$, together with a κ-balanced R-partition, and a graph G with a corresponding (ε, δ)-super-regular R-partition $V = V_1 \dot\cup \ldots \dot\cup V_r$. Assume that in addition for every $i \in [r]$ every tuple $(u_1, \ldots, u_{a+1}) \subseteq V \setminus V_i$ of vertices satisfies $|\bigcap_{j \in [a+1]} N_G(u_j) \cap V_i| \geq \iota |V_i|$. Then there exists an embedding of H into G.

Again, the degree bound of $\xi n/\log n$ for H in Theorem 4.5 is optimal up to the constant factor. The same degree bound can be obtained if we do require H only to be an almost spanning subgraph, even if the additional condition on $(a+1)$-tuples from Theorem 4.5 is dropped again.

Theorem 4.6 (Arrangeable Blow-up Lemma, almost spanning version)
Let $\mu > 0$ and assume that we have exactly the same setup as in Theorem 4.4, but with $\Delta(H) \le \xi n/\log n$ instead of the maximum degree bound given in (b)*, where ξ is sufficiently small compared to all other constants. Fix an a-arrangeable ordering of H, let X_i' be the first $(1-\mu)n_i$ vertices of X_i in this ordering, and set $H' := H[X_1' \cup \dots \cup X_r']$.*

Then there exists an embedding $\varphi\colon V(H') \to V(G)$ such that $\varphi(X_i') \subseteq V_i$ and $\varphi(x) \in I(x)$ for every $i \in [r]$ and every $x \in S_i \cap X_i'$.

Let us point out that one additional essential difference between these three versions of the Blow-up Lemma and Theorem 4.1 concerns the order of the quantifiers: the regularity ε that we require only depends on the maximum degree Δ_R of the reduced graph R, but *not* on the number of the vertices in R. Sometimes this is useful in applications. Clearly, we can reformulate our theorems to match the original order of quantifiers of Theorem 4.1; the lower bound on n_0 can be omitted in this case.

Applications. To demonstrate the usefulness of these extensions of the Blow-up Lemma, we consider two example applications that can now be derived in a relatively straightforward manner. At the end of this section we are going to mention a few further applications that are more difficult and will be proven in separate papers.

Our first application concerns F-factors in graphs of high minimum degree. This is a topic which is well investigated for graphs F of *constant* size. For a graph F on f vertices, an *F-factor* in a graph G is a collection of vertex disjoint copies of F in G such that all but at most $f-1$ vertices of G are covered by these copies of F.

A classical theorem by Hajnal and Szemerédi [49] states that each n-vertex graph G with minimum degree $\delta(G) \ge \frac{r-1}{r}n$ has a K_r-factor. Alon and Yuster [9] considered arbitrary graphs F and showed that, if r denotes the chromatic number of F, every sufficiently large graph G with minimum degree $\delta(G) \ge (\frac{r-1}{r} + \gamma)n$ contains an F-factor. This was improved upon by Komlós, Sárközy, and Szemerédi [66], who replaced the linear term γn in the degree bound by a constant $C = C(F)$; and by Kühn and Osthus [73], who, inspired by a result of Komlós [59], determined the precise minimum degree threshold for every constant size F up to a constant.

In contrast to the previous results we consider graphs F whose size may grow with the number of vertices n of the host graph G. More precisely, we allow graphs F of size linear in n. To prove this result, we use Theorem 4.4 (see Section 4.6) and hence we require that F is a-arrangeable and has maximum degree at most $\sqrt{n}/\log n$.

Theorem 4.7
For every a, r and $\gamma > 0$ there exist n_0 and $\xi > 0$ such that the following is true. Let G be any graph on $n \ge n_0$ vertices with $\delta(G) \ge (\frac{r-1}{r}+\gamma)n$ and let F be an a-arrangeable r-chromatic graph with at most ξn vertices and with maximum degree $\Delta(F) \le \sqrt{n}/\log n$. Then G contains an F-factor.

Our second application is a universality result for random graphs $\mathcal{G}(n,p)$ with constant p (that is, a graph on vertex set $[n]$ for which every $e \in \binom{[n]}{2}$ is inserted as an edge independently with probability p). A graph G is called *universal* for a family $\mathcal{H}$

of graphs if G contains a copy of each graph in $\mathcal{H}$ as a subgraph. For instance, graphs that are universal for the family of forests, of planar graphs and of bounded degree graphs have been investigated (see [5] and the references therein).

Here we consider the class

$$\mathcal{H}_{n,a,\xi} := \{H : |H| = n,\ H \text{ is } a\text{-arrangeable},\ \Delta(H) \leq \xi n/\log n\}$$

of arrangeable graphs whose maximum degree is allowed to grow with n. Using Theorem 4.5, we show that with high probability $\mathcal{G}(n,p)$ contains a copy of each graph in $\mathcal{H}_{n,a,\xi}$ (see Section 4.6). Universality problems for bounded degree graphs in (subgraphs of) random graphs with constant p were also considered in [53]. Another result for subgraphs of potentially growing degree and p tending to 0 can be found in [84]. Theorem 2.1 of [84] implies that any a-arrangeable graph of maximum degree $o(n^{1/4})$ can be embedded into $\mathcal{G}(n,p)$ with $p > 0$ constant with high probability.

Theorem 4.8
For all constants a, $p > 0$ there exists $\xi > 0$ such that $\mathcal{G}(n,p)$ is universal for $\mathcal{H}_{n,a,\xi}$ with high probability.

In addition, we use Theorem 4.4 in [19] to establish an analogue of the Bandwidth Theorem from [18] for arrangeable graphs. More precisely, we prove the following result.

Theorem 4.9 (Arrangeable Bandwidth Theorem [19])
For all $r, a \in \mathbb{N}$ and $\gamma > 0$, there exist constants $\beta > 0$ and $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$ the following holds. If H is an r-chromatic, a-arrangeable graph on n vertices with $\Delta(H) \leq \sqrt{n}/\log n$ and bandwidth at most βn and if G is a graph on n vertices with minimum degree $\delta(G) \geq \left(\frac{r-1}{r} + \gamma\right) n$, then there exists an embedding of H into G.

As we also show there, this implies for example that every graph G with minimum degree at least $(\frac{3}{4} + \gamma)n$ contains *almost every* planar graph H on n vertices, provided that $\gamma > 0$. In addition it implies that almost every planar graph H has Ramsey number $R(H) \leq 12|H|$.

Finally, another application of Theorem 4.4 appears in [3]. In that paper Allen, Skokan, and Würfl prove the following result, closing a gap left in the analysis of large planar subgraphs of dense graphs by Kühn, Osthus, and Taraz [74] and Kühn and Osthus [71].

Theorem 4.10 (Allen, Skokan, Würfl [3])
For every $\gamma \in (0, 1/2)$ there exists n_γ such that every graph on $n \geq n_\gamma$ vertices with minimum degree at least γn contains a planar subgraph with $2n - 4k$ edges, where k is the unique integer such that $k \leq 1/(2\gamma) < k + 1$.

Methods. To prove the full version of our Arrangeable Blow-up Lemma (Theorem 4.4), we proceed in two steps. Firstly, we use a random greedy algorithm to embed an almost spanning subgraph H' of the target graph H into the host graph G

(proving Theorem 4.6 along the way). Secondly, we complete the embedding by finding matchings in suitable auxiliary graphs which concern the remaining vertices in $V(H) \setminus V(H')$ and the unused vertices V^{Free} of G. The first step uses an approach similar to the one of Komlós, Sárközy, and Szemerédi [62]. The second step utilises ideas from Rödl and Ruciński's [86]. Let us briefly comment on the similarities and differences.

The use of a random greedy algorithm to prove the Blow-up Lemma appears in [62]. The idea is intuitive and simple: Order the vertices of the target graph H' arbitrarily and consecutively embed them into the host graph G, in each step choosing a random image vertex $\varphi(x)$ in the set $A(x)$ of those vertices which are still possible as images for the vertex x of H' we are currently embedding. If for some unembedded vertex x the set $A(x)$ gets too small, then call x critical and embed it immediately, but still randomly in $A(x)$. Our random greedy algorithm proceeds similarly, with one main difference. We cannot use an arbitrary order of the vertices of H', but have to use one which respects the arrangeability bound. Consequently, we also cannot embed critical vertices immediately – each vertex has to be embedded when it is its turn according to the given order. So we need a different strategy for dealing with critical vertices. We solve this problem by reserving a linear sized set of special vertices in G for the embedding of critical vertices, which are very few.

The second step is more intricate. Similarly to the approach in [86] we construct for each cluster V_i an auxiliary bipartite graph F_i with the classes $X_i \setminus V(H')$ and $V_i \cap V^{Free}$ and an edge between $x \in V(H)$ and $v \in V(G)$ whenever embedding x into v is a permissible extension of the partial embedding from the first step. Moreover, we guarantee that $V(H) \setminus V(H')$ is a stable set. Then, clearly, if each F_i has a perfect matching, there is an embedding of H into G. So the question remains how to show that the auxiliary graphs have perfect matchings. Rödl and Ruciński approach this by showing that their auxiliary graphs are super-regular. We would like to use a similar strategy, but there are two main difficulties. Firstly, because the degrees in our auxiliary graphs vary greatly, they cannot be super-regular. Hence we have to appropriately adjust this notion to our setting, which results in a property that we call weighted super-regular. Secondly, the proof that our auxiliary graphs are weighted super-regular now has to proceed quite differently, because we are dealing with the arrangeable graphs.

Structure. This paper is organised as follows. In Section 4.2 we provide notation and some tools. In Section 4.3 we show how to embed almost spanning arrangeable graphs, which will prove Theorem 4.6. In Section 4.4 we extend this to become a spanning embedding, proving Theorem 4.4. At the end of Section 4.4, we also outline how a similar argument gives Theorem 4.5. In Section 4.5 we explain why the degree bounds in the new versions of the Blow-up Lemma and the requirements for the image restrictions are essentially best possible. In Section 4.6, we give the proofs for our applications, Theorem 4.7 and Theorem 4.8.

4.2 Notation and preliminaries

All logarithms are to base e. For a graph G we write $V(G)$ for its vertex set, $E(G)$ for its edge set and denote the number of its vertices by $|G|$, its *maximum degree* by $\Delta(G)$ and its *minimum degree* by $\delta(G)$. Let $u, v \in V(G)$ and $U, W \subset V(G)$. The *neighbourhood* of u in G is denoted by $N_G(u)$, the neighbourhood of u in the set U by $N_G(u, U) := N_G(u) \cap U$. Similarly $N_G(U) = \bigcup_{x \in U} N_G(x)$ and $N_G(U, W) := N_G(U) \cap W$. The *co-degree* of u and v is $\deg_G(u, v) = |N_G(u) \cap N_G(v)|$. We often omit the subscript G.

For easier reading, we will often use x, y or z for vertices in the graph H that we are embedding, and u, v, w for vertices of the host graph G.

We shall also use the following version of the Hajnal-Szemerédi Theorem [49].

Theorem 4.11
Every graph G on n vertices and maximum degree $\Delta(G)$ can be partitioned into $\Delta(G)+1$ stable sets of size $\lfloor n/(\Delta(G)+1) \rfloor$ or $\lceil n/(\Delta(G)+1) \rceil$ each.

4.2.1 Arrangeability

Let H be a graph and $(x_1, x_2, \ldots, x_n)$ be an a-arrangeable ordering of its vertices. We write $x_i \prec x_j$ if and only if $i < j$ and say that x_i is *left* of x_j and x_j is *right* of x_i. We write $N^-(x) := \{y \in N_H(x) : y \prec x\}$ and $N^+(x) := \{y \in N_H(x) : x \prec y\}$ and call these the set of *predecessors* or the set of *successors* of x respectively. Predecessors and successors of vertex sets and in vertex sets are defined accordingly. Then $|N^+(x)| \le \Delta(H)$ for all $x \in V(H)$ and the definition of arrangeability says that $N^-\big(N^+(x_i)\big) \cap \{x_1, \ldots, x_i\}$ is of size at most a for each $i \in [n]$. Moreover, it follows that all $x \in V(H)$ satisfy $|N^-(x)| \le a$ and

$$e(H) = \sum_{x \in V(H)} |N^+(x)| = \sum_{x \in V(H)} |N^-(x)| \le an\,. \tag{4.1}$$

In the proof of our main theorem, it will turn out to be desirable to have a vertex ordering which is not only arrangeable, but also has the property that its final μn vertices form a stable set. More precisely we require the following properties.

Definition 4.12 (stable ending)
Let $\mu > 0$ and let $H = (X_1 \,\dot\cup \ldots \dot\cup\, X_r, E)$ be an r-partite, a-arrangeable graph with partition classes of order $|X_i| = n_i$ with $\sum_{i \in [r]} n_i = n$. Let $(v_1, \ldots, v_n)$ be an a-arrangeable ordering of H. We say that the ordering has a stable ending of order μn *if $W = \{v_{(1-\mu)n+1}, \ldots, v_n\}$ has the following properties*

(i) $|W \cap X_i| = \mu n_i$ *for every* $i \in [r]$,

(ii) $H[W]$ *is a stable set.*

The next lemma shows that an arrangeable order of a graph can be reordered to have a stable ending while only slightly increasing the arrangeability bound.

Lemma 4.13
Let a, Δ_R, κ be integers and let H be an a-arrangeable graph that has a κ-balanced R-partition with $\Delta(R) < \Delta_R$. Then H has a $(5a^2\kappa\Delta_R)$-arrangeable ordering with stable ending of order μn, where $\mu = 1/(10a(\kappa\Delta_R)^2)$.

Proof. Let $X = X_1 \dot\cup \ldots \dot\cup X_r$ be a κ-balanced R-partition of H with $|X_i| = n_i$. Further let $(x_1, \ldots, x_n)$ be any a-arrangeable ordering of H. In a first step we will find a stable set $W \subseteq X$ with $|W \cap X_i| = \mu n_i$ for $\mu = 1/(10a(\kappa\Delta_R)^2)$. Note that for every $i \in [r]$ a vertex $x \in X_i$ has only neighbours in sets X_j with $ij \in E(R)$. Further $H[X_i \cup \{X_j : ij \in E(R)\}]$ has at most $\kappa\Delta_R n_i$ vertices and is a-arrangeable. Therefore

$$\sum_{w \in X_i} \deg(w) \overset{(4.1)}{\leq} 2a\kappa\Delta_R n_i.$$

It follows that at least half the vertices $w \in X_i$ have $\deg(w) \leq 4a\kappa\Delta_R$. Let W'_i be the set of these vertices and m'_i be their number.

Now we greedily find a stable set $W \subseteq \bigcup_{i \in [r]} W'_i$ as follows. In the beginning we set $W = \emptyset$. Then we iteratively select an $i \in [r]$ with

$$|X_i \cap W|/n_i = \min_{j \in [r]} |X_j \cap W|/n_j\,, \tag{4.2}$$

choose an arbitrary vertex $x \in W'_i$, move it to W and delete x from W'_i and $N_H(x)$ from W'_j for all $j \in [r]$. We perform this operation until we have found a stable set W with $|W \cap X_i| = \mu n_i$ for all $i \in [r]$ or we attempt to choose a vertex from an empty set W'_{i^*}.

So assume that, at some point, we try to choose a vertex from an empty set W'_{i^*}. For each $i \in [r]$ let m_i be the number of vertices chosen from X_i (and moved to W) so far. Moreover, let $i \in [r]$ be such that $m_i < \mu n_i$ and consider the last step when a vertex from X_i was chosen. Before this step, $m_i - 1$ vertices of X_i and at most m_{i^*} vertices of X_{i^*} have been chosen. By (4.2) we thus have $(m_i - 1)/n_i \leq m_{i^*}/n_{i^*}$, which implies $m_i \leq \kappa m_{i^*} + 1$ because $n_i \leq \kappa n_{i^*}$. Hence, since W'_{i^*} became empty, we have

$$\begin{aligned} n_{i^*}/2 \leq m'_{i^*} &\leq m_{i^*} + \sum_{\{i^*, i\} \in E(R)} m_i 4a\kappa\Delta_R \\ &\leq m_{i^*} + (\Delta_R - 1)(\kappa m_{i^*} + 1)4a\kappa\Delta_R \leq m_{i^*}\, 5a(\kappa\Delta_R)^2\,. \end{aligned}$$

Thus $m_{i^*} \geq n_{i^*}/(10a(\kappa\Delta_R)^2)$. Since we then try to choose from W'_{i^*} we must have $m_{i^*}/n_{i^*} \leq m_i/n_i$ by (4.2), which implies $m_i \geq n_i/(10a(\kappa\Delta_R)^2) = \mu n_i$. Hence we indeed find a stable set W with $|W \cap X_i| = \mu n_i$ for all $i \in [r]$.

Given this stable set W we define a new ordering in which these vertices are moved to the end in order to form the stable ending. To make this more precise let $(x'_1, \ldots, x'_n)$ be the vertex ordering obtained from $(x_1, \ldots, x_n)$ by moving all vertices of W to the end (in any order). It remains to prove that $(x'_1, \ldots, x'_n)$ is $(5a^2\kappa\Delta_R)$-arrangeable. Let $L'_i = \{x'_1, \ldots, x'_i\}$ and $R'_i = \{x'_{i+1}, \ldots, x'_n\}$ be the vertices left and right of x'_i in the new ordering. We have to show that

$$\left|N\big(N(x'_i, R'_i), L'_i\big)\right| \leq 5a^2\kappa\Delta_R$$

for all $i \in [n]$. This is obvious for the vertices in W because they are now at the end and W is stable. For $x_i \notin W$ let $N_i' = N\big(N(x_i, R_i'), L_i'\big)$ be the set of predecessors of successors of x_i in the new ordering. N_i is defined analogously for the original ordering. Then all vertices in $N_i' \backslash N_i$ are neighbours of predecessors y of x_i in the original ordering with $y \in W$. There are at most a such left-neighbours of x_i and each of these has at most $4a\kappa\Delta_R$ neighbours by definition of W. Hence

$$|N_i'| \le |N_i| + a \cdot 4a\kappa\Delta_R \le a + 4a^2\kappa\Delta_R \le 5a^2\kappa\Delta_R\,. \qquad \square$$

4.2.2 Weighted regularity

In our proof we shall make use of a weighted version of ε-regularity. More precisely, we will have to deal with a bipartite graph whose vertices have very different degrees. The idea is then to give each vertex a weight antiproportional to its degree and then say that the graph is weighted regular if the following holds.

Definition 4.14 (Weighted regular pairs)
Let $\varepsilon > 0$ and consider a bipartite graph $G = (A \,\dot\cup\, B, E)$ with a weight function $\omega : A \to [0,1]$. For $A' \subseteq A$, $B' \subseteq B$ we define the weighted density

$$d_\omega(A', B') := \frac{\sum_{x \in A'} \omega(x)|N(x, B')|}{|A'| \cdot |B'|}\,.$$

We say that the pair (A, B) with weight function ω is weighted ε-regular *(with respect to ω) if for any $A' \subseteq A$ with $|A'| \ge \varepsilon|A|$ and any $B' \subseteq B$ with $|B'| \ge \varepsilon|B|$ we have*

$$|d_\omega(A, B) - d_\omega(A', B')| \le \varepsilon\,.$$

Many results for ε-regular pairs carry over to weighted ε-regular pairs. For one, subpairs of weighted regular pairs are weighted regular.

Proposition 4.15
Let $G = (A \,\dot\cup\, B, E)$ with weight function $\omega : A \to [0,1]$ be weighted ε-regular. Further let $A' \subseteq A$, $B' \subseteq B$ with $|A'| \ge \gamma|A|$ and $|B'| \ge \gamma|B|$ for some $\gamma \ge \varepsilon$ and set $\varepsilon' := \max\{2\varepsilon, \varepsilon/\gamma\}$. Then $(A' \,\dot\cup\, B', E \cap A' \times B')$ is a weighted ε'-regular pair with respect to the restricted weight function $\omega' : A' \to [0,1]$, $\omega'(x) = \omega(x)$.

Proof. Let $A' \subseteq A$ and $B' \subseteq B$ with $|A'| \ge \gamma|A|$, $|B'| \ge \gamma|B|$ be arbitrary. The definition of weighted ε-regularity implies that $|d_\omega(A, B) - d_\omega(A', B')| \le \varepsilon$. Moreover, $|d_\omega(A, B) - d_\omega(A^*, B^*)| \le \varepsilon$ for all $A^* \subseteq A'$ and $B^* \subseteq B'$ with $|A^*| \ge (\varepsilon/\gamma)|A'| \ge \varepsilon|A|$, $|B^*| \ge (\varepsilon/\gamma)|B'| \ge \varepsilon|B|$ for the same reason. It follows by triangle inequality that $|d_\omega(A', B') - d_\omega(A^*, B^*)| \le 2\varepsilon$. Hence $(A' \,\dot\cup\, B', E \cap A' \times B')$ with weight function $\omega' : A' \to [0,1]$ is a weighted ε'-regular pair where $\varepsilon' = \max\{2\varepsilon, \varepsilon/\gamma\}$. $\square$

If most vertices of a bipartite graph have the 'right' degree and most pairs have the 'right' co-degree then the graph is an ε-regular pair. This remains to be true for weighted regular pairs and weighted degrees and co-degrees.

Definition 4.16 (Weighted degree and co-degree)
Let $G = (A \mathbin{\dot\cup} B, E)$ be a bipartite graph and $\omega : A \to [0,1]$. For $x, y \in A$ we define the weighted degree *of x as $\deg_\omega(x) := \omega(x)|N(x,B)|$ and the* weighted co-degree *of x and y as $\deg_\omega(x,y) := \omega(x)\omega(y)|N(x,B) \cap N(y,B)|$.*

A proof of the following lemma can be found in the Appendix.

Lemma 4.17
Let $\varepsilon > 0$ and $n \geq \varepsilon^{-6}$. Further let $G = (A \mathbin{\dot\cup} B, E)$ be a bipartite graph with $|A| = |B| = n$ and let $\omega : A \to [\varepsilon, 1]$ be a weight function for G. If

(i) $|\{x \in A : |\deg_\omega(x) - d_\omega(A,B)n| > \varepsilon^{14}n\}| < \varepsilon^{12}n$ *and*

(ii) $|\{\{x,y\} \in \binom{A}{2} : |\deg_\omega(x,y) - d_\omega(A,B)^2 n| \geq \varepsilon^9 n\}| \leq \varepsilon^6\binom{n}{2}$

then (A,B) is a weighted 3ε-regular pair.

It is well known that a balanced (ε,δ)-super-regular pair has a perfect matching if $\delta > 2\varepsilon$ (see, e.g., [86]). Similarly, balanced weighted regular pairs with an appropriate minimum degree bound have perfect matchings (see the Appendix for a proof).

Lemma 4.18
Let $\varepsilon > 0$ and let $G = (A \mathbin{\dot\cup} B, E)$ with $|A| = |B| = n$ and weight function $\omega : A \to [\sqrt{\varepsilon}, 1]$ be a weighted ε-regular pair. If $\deg(x) > 2\sqrt{\varepsilon}n$ for all $x \in A \cup B$ then G contains a perfect matching.

4.2.3 Chernoff type bounds

Our proofs will heavily rely on the probabilistic method. In particular we will want to bound random variables that are close to being binomial. By close to we mean that the individual events are not necessarily independent but occur with certain probability even if condition on the outcome of other events. The following two variations on the classical bound by Chernoff make this more precise.

Lemma 4.19
Let $0 \leq p_1 \leq p_2 \leq 1$, $0 < c \leq 1$. Further let $\mathcal{A}_i$ for $i \in [n]$ be 0-1-random variables and set $\mathcal{A} := \sum_{i\in[n]} \mathcal{A}_i$. If

$$p_1 \leq \mathbb{P}\left[\mathcal{A}_i = 1 \;\middle|\; \begin{array}{l} \mathcal{A}_j = 1 \textit{ for all } j \in J \textit{ and} \\ \mathcal{A}_j = 0 \textit{ for all } j \in [i-1] \setminus J \end{array}\right] \leq p_2$$

for every $i \in [n]$ and every $J \subseteq [i-1]$ then

$$\mathbb{P}[\mathcal{A} \leq (1-c)p_1 n] \leq \exp\left(-\frac{c^2}{3}p_1 n\right)$$

and

$$\mathbb{P}[\mathcal{A} \geq (1+c)p_2 n] \leq \exp\left(-\frac{c^2}{3}p_2 n\right).$$

Similarly we can state a bound on the number of tuples of certain random variables.

Lemma 4.20
Let $0 < p$ and $a, m, n \in \mathbb{N}$. Further let $\mathcal{I} \subseteq \mathcal{P}([n]) \setminus \{\emptyset\}$ be a collection of m disjoint sets with at most a elements each. For every $i \in [n]$ let $\mathcal{A}_i$ be a 0-1-random variable. Further assume that for every $I \in \mathcal{I}$ and every $k \in I$ we have

$$\mathbb{P}\left[\mathcal{A}_k = 1 \;\middle|\; \begin{array}{l} \mathcal{A}_j = 1 \text{ for all } j \in J \text{ and} \\ \mathcal{A}_j = 0 \text{ for all } j \in [k-1] \setminus J \end{array}\right] \geq p$$

for every $J \subseteq [k-1]$ with $[k-1] \cap I \subseteq J$. Then

$$\mathbb{P}\left[\left|\{I \in \mathcal{I} : \mathcal{A}_i = 1 \text{ for all } i \in I\}\right| \geq \tfrac{1}{2}p^a m\right] \geq 1 - 2\exp\Big(-\frac{1}{12}p^a m\Big).$$

The proofs for both lemmas can be found in the Appendix. The first one is very close to the proof of the classical Chernoff bound while the second proof builds on the fact that the events $[\mathcal{A}_i = 1 \text{ for all } i \in I]$ have probability at least p^a for every $I \in \mathcal{I}$. In particular, in the special case $a = 1$, Lemma 4.19 implies Lemma 4.20.

4.3 An almost spanning version of the Blow-up Lemma

This section is dedicated to the proof of Theorem 4.6 which is a first step towards Theorem 4.4. We give a randomised algorithm for the embedding of an almost spanning subgraph H' into G and show that it is well defined and that it succeeds with positive probability.

This embedding of H' is later extended to the embedding of a spanning subgraph H in Section 4.4. Applying the randomised algorithm to H while only embedding H' provides the structural information necessary for the extension of the embedding. It is for this reason that we define a graph H while only embedding a subgraph $H' \subseteq H$ into G in this section.

Remark
In the following we shall always assume that each super-regular pair (V_i, V_j) appearing in the proof has density

$$d(V_i, V_j) = \delta$$

exactly*, and minimum degree*

$$\min\nolimits_{v \in V_i} \deg(v, V_j) \geq \tfrac{1}{2}\delta|V_j|, \quad \min\nolimits_{v \in V_j} \deg(v, V_i) \geq \tfrac{1}{2}\delta|V_i|, \tag{4.3}$$

since otherwise we can simply appropriately delete random edges to obtain this situation (while possibly increasing regularity to 2ε).

4.3.1 Constants, constants

Since there will be plenty of constants involved in the following proofs we give a short overview first.

Δ_R: the maximum degree of R is *strictly* smaller than Δ_R
r: the number of clusters
a: the arrangeability of H
s: the chromatic number of H
δ: the density of the pairs (V_i, V_j) in G
μ: the proportion of G that will be left after embedding H
ξ: some constant in the degree-bound of H
ε: the regularity of the pairs (V_i, V_j) in G
ε': the weighted regularity of the auxiliary graphs $F_i(t)$
κ: the maximum quotient between cluster sizes
γ: a threshold for moving a vertex into the critical set
λ: the fraction of vertices whose predecessors receive a special embedding
α: the fraction of vertices with image restrictions
c: the relative size of the image restrictions
C: the maximum number of image restrictions per cluster

Now let $C, a, \Delta_R, \kappa \in \mathbb{N}$ and $\delta, c, \mu > 0$ be given. We define the following constants.

$$\gamma = \frac{c}{2}\frac{\mu}{10}\delta^a \ , \tag{4.4}$$

$$\lambda = \frac{1}{25a}\delta\gamma \ , \tag{4.5}$$

$$\varepsilon' = \min\left\{\left(\frac{\lambda\delta^a}{6\cdot 2^{a^2+1}3^a}\right)^2, \left(\frac{7\gamma}{30}\right)^2\right\} \ , \tag{4.6}$$

$$\varepsilon = \min\left\{\frac{1}{\Delta_R(1+C)2^{a+1}}\varepsilon', \left(\frac{\varepsilon'}{3}\right)^{36}\right\} \ , \tag{4.7}$$

$$\alpha = \frac{\sqrt{\varepsilon}}{6} \ . \tag{4.8}$$

Furthermore, let r be given. Then we choose

$$\xi = \frac{8\varepsilon^2}{9\gamma^2\kappa r} \ . \tag{4.9}$$

Moreover, we ensure that n_0 is big enough to guarantee

$$\sqrt{n_0} \geq 48\frac{3^a 2^{a^2+1}a\kappa r}{\lambda\delta^a} \ , \quad n_0 \geq 60\frac{\kappa r}{\varepsilon^2\delta\mu}\log(12(n_0)^2), \ \text{ and } \quad \log n_0 \geq 36\frac{2^{a^2}a^2\kappa r}{\lambda} \ . \tag{4.10}$$

All logarithms are base e. In short, the constants used relate as

$$0 < \xi \ll \varepsilon \ll \alpha \ll \varepsilon' \ll \lambda \ll \gamma \ll \mu, \delta \leq 1\,.$$

Moreover, $\varepsilon \ll 1/\Delta_R$. Note that it follows from these definitions that $(1+\varepsilon/\delta)^a \leq 1+\sqrt{\varepsilon}/3$ and $(1-\varepsilon/\delta)^a \geq 1-\sqrt{\varepsilon}/3$ which implies

$$\frac{(\delta+\varepsilon)^a}{1+\sqrt{\varepsilon}/3} \leq \delta^a \leq \frac{(\delta-\varepsilon)^a}{1-\sqrt{\varepsilon}/3}, \quad \text{in particular} \quad (\delta-\varepsilon)^a \geq \frac{9}{10}\delta^a. \tag{4.11}$$

4.3.2 The randomised greedy algorithm

Let $V(H) = (x_1, \ldots, x_n)$ be an a-arrangeable ordering of H and let $H' \subseteq H$ be a subgraph induced by $\{x_1, \ldots, x_{(1-\mu)n}\}$. In this section we define a *randomised greedy algorithm* (RGA) for the embedding of $V(H')$ into $V(G)$. This algorithm processes the vertices of H vertex by vertex and thereby defines an embedding φ of H' into G. We say that vertex x_t gets embedded in time step t where t runs from 1 to $T = |H'|$. Accordingly $t(x) \in [n]$ is defined to be the time step in which vertex x will be embedded.

We explain the main ideas before giving an exact definition of the algorithm.

Preparing H: Recall that S_i is the set of *image restricted vertices* in X_i and set $S := \bigcup S_i$. We define L_i^* to be the last λn_i vertices in $X_i \setminus N(S)$ in the arrangeable ordering. Moreover, we define $X_i^* := N^-(L_i^*) \cup S_i$ and $X^* := \bigcup X_i^*$. Those vertices will be called the *important vertices*. The name indicates that they will play a major rôle for the spanning embedding. Important vertices shall be treated specially by the embedding algorithm. The a-arrangeability of H implies that

$$|X_i^*| \leq a\lambda n_i + \alpha n_i \tag{4.12}$$

for all $i \in [r]$.

Preparing G: Before we start embedding into G we randomly set aside $(\mu/10)n_i$ vertices in V_i for each $i \in [r]$. We denote these sets by V_i^{s} and call them the *special vertices*. All remaining vertices, i.e., $V_i^{\mathrm{o}} := V_i \setminus V_i^{\mathrm{s}}$ will be called *ordinary vertices*. As the name suggests our algorithm will try to embed most vertices of H' into the sets V_i^{o} and only if this fails resort to embedding into V_i^{s}. The idea is that the special vertices will be reserved for the important vertices and for those vertices in H' whose embedding turns out to be intricate. We define

$$V^{\mathrm{o}} := \bigcup_{i=1}^{r} V_i^{\mathrm{o}}, \quad V^{\mathrm{s}} := \bigcup_{i=1}^{r} V_i^{\mathrm{s}}.$$

Note that $V^{\mathrm{o}} \mathbin{\dot\cup} V^{\mathrm{s}}$ defines a partition of $V(G)$.

Candidate sets: While our embedding process is running, more and more vertices of G will be used up to accommodate vertices of H. For each time step $t \in [n]$ we denote by $V^{\mathit{Free}}(t) := V(G) \setminus \{v \in V(G) : \exists t' < t : \varphi(x_{t'}) = v\}$ the set of vertices where no vertex of H has been embedded yet. Obviously $\varphi(x_t) \in V^{\mathit{Free}}(t)$ for all t.

The algorithm will define sets $C_{t,x} \subseteq V(G)$ for $1 \leq t \leq T$, $x \in V(H)$, which we will call the *candidate set* for x at time t. Analogously

$$A_{t,x} := C_{t,x} \cap V^{Free}(t)$$

will be called the *available candidate set* for x at time t. Again we distinguish between the *ordinary candidate* set $C^{\mathrm{o}}_{t,x} := C_{t,x} \cap V^{\mathrm{o}}$ and the *special candidate set* $C^{\mathrm{s}}_{t,x} := C_{t,x} \cap V^{\mathrm{s}}$ or their respective available version $A^{\mathrm{o}}_{t,x} := A_{t,x} \cap V^{\mathrm{o}}$ and $A^{\mathrm{s}}_{t,x} := A_{t,x} \cap V^{\mathrm{s}}$.

Finally we define a set $Q(t) \subseteq V(H)$ and call it the *critical set* at time t. $Q(t)$ will contain the vertices whose available candidate set got too small at time t or earlier.

Algorithm RGA

Initialisation

Randomly select $V^{\mathrm{s}}_i \subseteq V_i$ with $|V^{\mathrm{s}}_i| = (\mu/10)|V_i|$ for each $i \in [r]$. For $x \in X_i \setminus S_i$ set $C_{1,x} = V_i$ and for $x \in S_i$ set $C_{1,x} = I(x)$. Set $Q(1) = \emptyset$.

Check that for every $i \in [r]$, $v \in V_i$, and every $j \in N_R(i)$ we have

$$\left| \frac{|N_G(v) \cap V^{\mathrm{s}}_j|}{|V^{\mathrm{s}}_j|} - \frac{|N_G(v) \cap V_j|}{|V_j|} \right| \leq \varepsilon\,. \tag{4.13}$$

Further check that every $x \in S_i$ has

$$|C^{\mathrm{s}}_{1,x}| = |I(x) \cap V^{\mathrm{s}}_i| \geq \tfrac{1}{20} c\mu\, n_i\,. \tag{4.14}$$

Halt with failure if any of these does not hold.

Embedding Stage

For $t \geq 1$, **repeat** the following steps.

Step 1 – Embedding x_t: Let $x = x_t$ be the vertex of H to be embedded at time t. Let $A'_{t,x}$ be the set of vertices $v \in A_{t,x}$ which satisfy (4.15) and (4.16) for all $y \in N^+(x)$:

$$(\delta - \varepsilon)|C^{\mathrm{o}}_{t,y}| \leq |N_G(v) \cap C^{\mathrm{o}}_{t,y}| \leq (\delta + \varepsilon)|C^{\mathrm{o}}_{t,y}|, \tag{4.15}$$
$$(\delta - \varepsilon)|C^{\mathrm{s}}_{t,y}| \leq |N_G(v) \cap C^{\mathrm{s}}_{t,y}| \leq (\delta + \varepsilon)|C^{\mathrm{s}}_{t,y}|. \tag{4.16}$$

Choose $\varphi(x)$ uniformly at random from

$$A(x) := \begin{cases} A^{\mathrm{o}}_{t,x} \cap A'_{t,x} & \text{if } x \notin X^* \text{ and } x \notin Q(t), \\ A^{\mathrm{s}}_{t,x} \cap A'_{t,x} & \text{else.} \end{cases} \tag{4.17}$$

Step 2 – Updating candidate sets: for each unembedded vertex $y \in V(H)$, set

$$C_{t+1,y} := \begin{cases} C_{t,y} \cap N_G(\varphi(x)) & \text{if } y \in N^+(x), \\ C_{t,y} & \text{otherwise.} \end{cases}$$

Step 3 – Updating critical vertices: We will call a vertex $y \in X_i$ *critical* if $y \notin X^*_i$ and

$$|A^{\mathrm{o}}_{t+1,y}| < \gamma n_i. \tag{4.18}$$

Obtain $Q(t+1)$ by adding to $Q(t)$ all critical vertices that have not been embedded yet. Set $Q_i(t+1) = Q(t+1) \cap X_i$.
Halt with failure if there is $i \in [r]$ with

$$|Q_i(t+1)| > \varepsilon' n_i \,. \tag{4.19}$$

Else, if there are no more unembedded vertices left in $V(H')$ **halt with success**, **otherwise** set $t \leftarrow t+1$ and go back to *Step 1*.

We have now defined our randomised greedy algorithm for the embedding of an almost spanning subgraph H' into G. The rest of this section is to prove that it succeeds with positive probability. This then implies Theorem 4.6.

In order to analyse the RGA we define auxiliary graphs which describe possible embeddings of vertices of H' into G. These auxiliary graphs inherit some kind of regularity from G with positive probability. We show that the algorithm terminates successfully whenever this happens.

In the subsequent Section 4.3.3 we show that conditions (4.13) and (4.14) hold with probability at least $5/6$. The INITIALISATION of the RGA succeeds whenever this happens. Moreover, we prove that the embedding of each vertex is randomly chosen from a set of linear size in Step 1 of the EMBEDDING STAGE.

In Section 4.3.4 we define auxiliary graphs and derive that all auxiliary graphs are weighted regular with probability at least $5/6$. We also show that condition (4.19) never holds if this is the case. Thus the EMBEDDING STAGE also terminates successfully with probability at least $5/6$.

We conclude that the whole RGA succeeds with probability at least $2/3$. This implies Theorem 4.6.

4.3.3 Initialisation and Step 1

This section is to prove that the INITIALISATION of the RGA succeeds with probability at least $5/6$ and that Step 1 of the EMBEDDING STAGE always chooses vertices from a non-empty set.

Lemma 4.21
The INITIALISATION *succeeds with probability at least* $5/6$*, i.e. both condition* (4.13) *and* (4.14) *hold for every* $i \in [r]$*,* $v \in V_i$*,* $j \in [r] \setminus \{i\}$*, and* $x \in S_i$ *with probability* $5/6$*.*

Proof of Lemma 4.21. Fix one $v \in V_i$, $j \in [r] \setminus \{i\}$. Since V_j^{s} is a randomly chosen subset of V_j we have

$$\mathbb{E}[|N_G(v) \cap V_j^{\mathrm{s}}|] = |N_G(v) \cap V_j| \frac{|V_j^{\mathrm{s}}|}{|V_j|} \geq \frac{\delta}{2} n_j \frac{\mu}{10} \,.$$

It follows from a Chernoff bound (see Theorem 4.40 in the Appendix) that

$$\mathbb{P}\left[|N_G(v) \cap V_j^{\mathrm{s}}| - |N_G(v) \cap V_j| \frac{|V_j^{\mathrm{s}}|}{|V_j|} > \varepsilon |V_j^{\mathrm{s}}| \right] \leq \exp\left(-\frac{\varepsilon^2}{6} \delta n_i \frac{\mu}{10} \right) \overset{(4.10)}{\leq} \frac{1}{12n^2} \,.$$

Similarly $c|V_i^{\mathrm{s}}| \geq c\frac{\mu}{10}n_i$ and

$$\mathbb{P}[c|V_i^{\mathrm{s}}| - |I(x) \cap V_i^{\mathrm{s}}| \geq \frac{c}{2}|V_i^{\mathrm{s}}|] \leq \exp\left(-\frac{1}{12}c\frac{\mu}{10}n_i\right) \leq \frac{1}{12n} .$$

A union bound over all $i \in [r]$, $v \in V_i$ and $j \in N_R(i)$ or over all $x \in S_i$ finishes the proof. □

Let us write $\pi(t,x)$ for the number of predecessors of x that already got embedded by time t:

$$\pi(t,x) := |\{t' < t : \{x, x_{t'}\} \in E(H)\}|.$$

Obviously $\pi(t,x) \leq a$ by the definition of arrangeability.

Lemma 4.22
Let $x \in X_i \setminus S_i$ and $t \leq T$ be arbitrary. Then

$$(1-\mu/10)(\delta-\varepsilon)^{\pi(t,x)}n_i \leq |C_{t,x}^{\mathrm{o}}| \leq (1-\mu/10)(\delta+\varepsilon)^{\pi(t,x)}n_i ,$$
$$(\mu/10)(\delta-\varepsilon)^{\pi(t,x)}n_i \leq |C_{t,x}^{\mathrm{s}}| \leq (\mu/10)(\delta+\varepsilon)^{\pi(t,x)}n_i .$$

If $x \in S_i$, $t \leq T$ then

$$\frac{9}{10}\gamma n_i \leq |C_{t,x}^{\mathrm{s}}| .$$

Proof. The Initialisation of the RGA defines the candidate sets such that $|C_{1,x}^{\mathrm{o}}| = (1-\mu/10)n_i$ and $|C_{1,x}^{\mathrm{s}}| = (\mu/10)n_i$ for every $x \in X_i \setminus S_i$. In the Embedding Stage conditions (4.15) and (4.16) guarantee that $C_{t,x}^{\mathrm{o}}$ and $C_{t,x}^{\mathrm{s}}$ respectively shrink by a factor of $(\delta \pm \varepsilon)$ whenever a vertex in $N^-(x)$ is embedded.

If $x \in S_i$ we still have $|C_{1,x}^{\mathrm{s}}| \geq (c\mu/20)n_i$ by (4.14). The statement follows as conditions (4.15) and (4.16) again guarantee that $C_{t,x}^{\mathrm{s}}$ shrinks at most by a factor of $(\delta-\varepsilon)^a$. Moreover, $\frac{1}{20}c\mu(\delta-\varepsilon)^a \geq \frac{9}{10}\gamma$ by (4.11) and the definition of γ. □

We now argue that $\varphi(x)$ is chosen from a non-empty set at the end of Step 1 in the Embedding Stage. In fact, we will show that $\varphi(x)$ is chosen from a set of size linear in n_i.

Lemma 4.23
For any vertex $x \in X_i$ that gets embedded in the Embedding Stage *$\varphi(x)$ is chosen randomly from a set $A(x)$ of size at least $(\gamma/2)n_i$.*
Moreover, if x gets embedded into V_i^{s}

$$|X_i^*| + |Q_i(t(x))| + |A_{t(x),x}^{\mathrm{s}} \setminus A(x)| \leq \frac{\delta}{18}|C_{t(x),x}^{\mathrm{s}}| .$$

If the RGA completes the Embedding Stage *successfully but $x \in X_i$ does not get embedded in the* Embedding Stage *we have*

$$|A_{T,x}^{\mathrm{s}}| \geq \frac{7\gamma}{10}n_i .$$

Proof. We claim that any $x \in X_i$ that gets embedded into V_i^σ during the EMBEDDING STAGE has

$$|A^\sigma_{t(x),x}| \geq \frac{7\gamma}{10} n_i \,. \tag{4.20}$$

We will establish equation (4.20) at the end of this proof.

In order to show the first statement of the lemma we now bound $|A^\sigma_{t(x),x} \setminus A(x)|$, i.e., we determine the number of vertices that potentially violate conditions (4.15) or (4.16). As H is a-arrangeable, the vertices $y \in N^+(x)$ share at most 2^a distinct ordinary candidate sets $C^{\mathrm{o}}_{t(x),y}$ in each V_j. The number of special candidate sets $C^{\mathrm{s}}_{t(x),y}$ in each V_j might be larger by a factor of C as they arise from the intersection with at most C sets $I_{j,k}$ (with $k \in [C]$) which are the image restrictions. Moreover, there are less than Δ_R many sets V_j with $j \in N_R(i)$ bounding the total number of candidate sets we have to care for by $\Delta_R(1+C)2^a$.

As we embed x into an ε-regular pair there are at most $2\varepsilon n_i$ vertices $v \in A^\sigma_{t(x),x}$ for each $C^{\mathrm{o}}_{t(x),y}$ that violate (4.15) (and the same number for each $C^{\mathrm{s}}_{t(x),y}$ that violate (4.16)) with $y \in N^+(x)$. Hence

$$|A^\sigma_{t(x),x} \setminus A(x)| \leq \Delta_R(1+C)2^{a+1}\varepsilon n_i \tag{4.21}$$

if x gets embedded into V_i^σ. Now $\Delta_R(1+C)2^{a+1}\varepsilon n_i \leq \gamma/5 n_i$ by (4.7) and

$$|A(x)| = |A^\sigma_{t(x),x}| - |A^\sigma_{t(x),x} \setminus A(x)| \geq (\gamma/2) n_i$$

follows.

Next we show the second statement of the lemma. If $x \in X_i$ gets embedded into V_i^{s} in the EMBEDDING STAGE we conclude

$$\begin{aligned} |X_i^*| + |Q_i(t(x))| + |A^{\mathrm{s}}_{t(x),x} \setminus A(x)| &\leq (a\lambda + \alpha)\, n_i + \varepsilon' n_i + \Delta_R(1+C)2^{a+1}\varepsilon n_i \\ &\overset{(4.5),(4.7)}{\leq} \left(\tfrac{1}{25}\delta\gamma + \alpha + 2\varepsilon'\right) n_i \overset{(4.8)}{\leq} \tfrac{1}{20}\delta\gamma\, n_i \\ &\leq \tfrac{\delta}{18} |C^{\mathrm{s}}_{t(x),x}| \end{aligned}$$

where the first inequality is due to (4.12), (4.19), and (4.21) and the last inequality is due to $|C^{\mathrm{s}}_{t(x),x}| \geq \frac{9}{10}\gamma n_i$ by Lemma 4.22.

We now return to Equation (4.20). In order to prove it we distinguish between the two cases of (4.17) in *Step 1* of the EMBEDDING STAGE. If $x \notin X^*$ has never entered the critical set, it is embedded into $A^{\mathrm{o}}_{t(x),x}$ and $|A^{\mathrm{o}}_{t(x),x}| \geq (7\gamma/10) n_i$ holds by condition (4.18). Else x gets embedded into $A^{\mathrm{s}}_{t(x),x}$. As only vertices from $Q_i(t(x))$ or X_i^* have been embedded into V_i^{s} so far, we can bound $|A^{\mathrm{s}}_{t(x),x}|$ by

$$\begin{aligned} |A^{\mathrm{s}}_{t(x),x}| &\geq |C^{\mathrm{s}}_{t(x),x}| - |Q_i(t(x))| - |X_i^*| \\ &\overset{(4.12)}{\geq} \frac{9\gamma}{10} n_i - \varepsilon' n_i - (a\lambda + \alpha) n_i \geq \frac{7\gamma}{10} n_i \end{aligned}$$

where the second inequality is due to Lemma 4.22 and the third inequality is due to our choice of constants. In any case we have $|A^{\sigma}_{t(x),x}| \geq \frac{7\gamma}{10} n_i$ if x gets embedded into V_i^{σ} (with $\sigma \in \{\mathrm{o}, \mathrm{s}\}$) in Step 1 of the Embedding Stage.

If the RGA completes the Embedding Stage successfully but $x \in X_i$ does not get embedded during the Embedding Stage the analogous argument gives

$$|A^{\mathrm{s}}_{T,x}| \geq |C^{\mathrm{s}}_{T,x}| - |Q_i(T)| - |X_i^*| \geq \frac{7\gamma}{10} n_i \,. \qquad \square$$

4.3.4 The auxiliary graph

We run the RGA as described above. In order to analyse it, we define *auxiliary graphs* $F_i(t)$ which monitor at every time step t whether a vertex $v \in V(G)$ is still contained in the candidate set of a vertex $x \in V(H)$. Let $F_i(t) := (X_i \,\dot\cup\, V_i, E(F_i(t)))$ where $xv \in E(F_i(t))$ if and only if $v \in C_{t,x}$. We stress that we use the candidate sets $C_{t,x}$ and not the set of available candidates $A_{t,x}$. This is well defined as $C_{t,x} \subseteq V_i$ for every $x \in X_i$ and every t. Note that $F_i(t)$ is a balanced bipartite graph. By Lemma 4.22 we have

$$(\delta - \varepsilon)^{\pi(t,x)} n_i \leq \deg_{F_i(t)}(x) \leq (\delta + \varepsilon)^{\pi(t,x)} n_i \tag{4.22}$$

for every $x \in X_i \setminus S_i$, i.e., the degree of x in $F_i(t)$ strongly depends on the number $\pi(t,x)$ of embedded predecessors. The main goal of this section is proving, however, that if we take this into account and weight the auxiliary graphs accordingly, then they are with high probability weighted regular (see Lemma 4.24). It will turn out that the RGA succeeds if this is the case (see Lemma 4.26).

More precisely, for $F_i(t)$ we shall use the weight function $\omega_t \colon X_i \to [0,1]$ with

$$\omega_t(x) := \delta^{a - \pi(t,x)} \,. \tag{4.23}$$

Observe that the weight function depends on t. For nicer notation, we write $\deg_{\omega,t}(x) := \deg_{\omega(t)}(x) = \omega_t(x)|N_{F_i(t)}(x)|$ for $x \in X_i$ and $d_{\omega,t}(X,Y) := d_{\omega(t)}(X,Y)$ for $X \subseteq X_i$ and $Y \subseteq V_i$. By (4.22) we have

$$\deg_{\omega,t}(x) \geq \delta^{a-\pi(t,x)}(\delta - \varepsilon)^{\pi(t,x)} n_i \overset{(4.11)}{\geq} (1 - \sqrt{\varepsilon}/3)\delta^a n_i \,, \tag{4.24}$$

$$\deg_{\omega,t}(x) \leq \delta^{a-\pi(t,x')}(\delta + \varepsilon)^{\pi(t,x')} n_i \overset{(4.11)}{\leq} (1 + \sqrt{\varepsilon}/3)\delta^a n_i \tag{4.25}$$

for every $x \in X_i \setminus S_i$ and t. Thus for every $i \in [r]$ and $t \leq T$ the auxiliary graph $F_i(t)$ satisfies

$$(1 - \sqrt{\varepsilon}/2)\delta^a \overset{(4.8)}{\leq} (1-\alpha)(1 - \sqrt{\varepsilon}/3)\delta^a \leq d_{\omega,t}(X_i, V_i) \leq (1 + \sqrt{\varepsilon}/2)\delta^a \,. \tag{4.26}$$

Let $\mathcal{R}_i(t)$ denote the event that $F_i(t)$ is weighted ε'-regular for ε' as in (4.6). Further let $\mathcal{R}_i$ be the event that $\mathcal{R}_i(t)$ for all $t \leq T$.

Lemma 4.24
We run the RGA in the setting of Theorem 4.6. Then $\mathcal{R}_i$ holds for all $i \in [r]$ with probability at least $5/6$.

We will use Lemma 4.17 and weighted degrees and co-degrees to prove Lemma 4.24.

Proof of Lemma 4.24. This proof checks the conditions of Lemma 4.17. Let

$$W_i^{(1)}(t) = \left\{x \in X_i : |\deg_{\omega,t}(x) - d_{\omega,t}(X_i, V_i)n_i| > \sqrt{\varepsilon}n_i\right\},$$
$$W_i^{(2)}(t) = \left\{\{x,y\} \in \binom{X_i}{2} : |\deg_{\omega,t}(x,y) - d_{\omega,t}(X_i, V_i)^2 n_i| \geq \sqrt[4]{\varepsilon}n_i\right\}$$

be the set of vertices and pairs which deviate from the expected (co-)degree. Let $W_i^{(1)} := \bigcup_{t\in[T]} W_i^{(1)}(t)$ and $W_i^{(2)} := \bigcup_{t\in[T]} W_i^{(2)}(t)$. We have $\varepsilon' \geq 3\varepsilon^{1/36}$ by (4.7), and by Lemma 4.17 all auxiliary graphs $F_i(t)$ with $t = 1, \dots, T$ are weighted ε'-regular if both

$$|W_i^{(1)}| < \sqrt{\varepsilon}n_i\,, \tag{4.27}$$
$$|W_i^{(2)}| \leq \sqrt[4]{\varepsilon}\binom{n_i}{2}\,. \tag{4.28}$$

Thus $\mathcal{R}_i$ occurs whenever equations (4.27) and (4.28) are satisfied. We will prove that this happens for a fixed $i \in [r]$ with probability at least $1 - n_i^{-1}$, which together with a union bound over $i \in [r]$ implies the statement of the lemma.

So fix $i \in [r]$. From (4.24), (4.25) and (4.26) we deduce that

$$|\deg_{\omega,t}(x) - d_{\omega,t}(X_i, V_i)n_i| \leq \sqrt{\varepsilon}n_i$$

for all $x \in X_i \setminus S_i$ and every $t \leq T$. But $|S_i| \leq \alpha n_i < \sqrt{\varepsilon}n_i$ by (4.8) and equation (4.27) is thus *always* satisfied.

It remains to consider (4.28). To this end let P_i be the set of all pairs $\{y,z\} \in \binom{X_i\setminus S_i}{2}$ with $N^-(y) \cap N^-(z) = \emptyset$. Observe that $|\binom{X_i}{2} \setminus \binom{X_i\setminus S_i}{2}| \leq \alpha n_i^2 \leq \frac{\sqrt{\varepsilon}}{6}n_i^2$ by (4.8) and

$$\left|\left\{\{y,z\} \in \binom{X_i}{2} : N^-(y) \cap N^-(z) \neq \emptyset\right\}\right| \leq a\Delta(H)n_i \leq a\frac{\xi n}{\log n}n_i$$
$$\leq \frac{2a\xi\kappa r}{\log n}\binom{n_i}{2} \overset{(4.9)}{\leq} \sqrt{\varepsilon}\binom{n_i}{2}\,.$$

Hence it suffices to show that

$$\mathbb{P}\left[|W_i^{(2)} \cap P_i| \leq \tfrac{1}{2}\sqrt[4]{\varepsilon}\binom{n_i}{2}\right] \geq \mathbb{P}\left[|W_i^{(2)} \cap P_i| \leq \tfrac{1}{2}\sqrt[4]{\varepsilon}|P_i|\right] > 1 - n_i^{-1}\,. \tag{4.29}$$

For this we first partition P_i into sets of mutually *predecessor disjoint* pairs, i.e., $P_i = K_1 \,\dot\cup \dots \dot\cup\, K_\ell$ such that for every $k \in [\ell]$, no vertex of X_i appears in two different pairs in K_k, and moreover no two pairs in K_k contain two vertices that have a common predecessor. Theorem 4.11 applied to the following graph asserts that there is such a partition with almost equally sized classes K_k: Let $\mathcal{P}$ be the graph on vertex set P_i with edges between exactly those pairs $\{y_1, y_2\}, \{y_1', y_2'\} \in P_i$ which have either $\{y_1, y_2\} \cap \{y_1', y_2'\} \neq \emptyset$ or $\big(N_H^-(y_1) \cup N_H^-(y_2)\big) \cap \big(N_H^-(y_1') \cup N_H^-(y_2')\big) \neq \emptyset$. This graph has maximum degree $\Delta(\mathcal{P}) < 2a\Delta(H)n_i \leq 2a(\xi n/\log n)n_i$. Hence Theorem 4.11 gives a partition $K_1 \mathbin{\dot\cup} \ldots \mathbin{\dot\cup} K_\ell$ of P_i into stable sets with $|K_k| \geq \lfloor |P_i|/\big(\Delta(\mathcal{P})+1\big)\rfloor \geq \log n/(8a\xi\kappa r)$ for all $k \in [\ell]$, where we used $|P_i| \geq n_i^2/4$.

Now fix $k \in [\ell]$ and consider the random variable $K_k' := K_k \cap W_i^{(2)}$. Our goal now is to show

$$\mathbb{P}\left[|K_k'| > \tfrac{1}{2}\sqrt[4]{\varepsilon}|K_k|\right] \leq n_i^{-3}\,, \tag{4.30}$$

as this together with another union bound over $k \in [\ell]$ with $\ell < n_i^2$ implies (4.29). We shall first bound the probability that some fixed pair $\{y, z\} \in K_k$ gets moved to $W_i^{(2)}(t)$ (and hence to K_k') at some time t.

For a pair $\{y, z\} \in K_k$ and $t \in [T]$ let $\mathcal{C}o_{t,y,z}$ denote the event that $|\deg_{\omega,t+1}(y,z) - \deg_{\omega,t}(y,z)| \leq \varepsilon n_i$. Why are we interested in these events? Obviously $\mathcal{C}o_{t,y,z}$ holds for all time steps t with $x_t \notin N^-(y) \mathbin{\dot\cup} N^-(z)$. This is because we have $\deg_{\omega,t+1}(y,z) = \deg_{\omega,t}(y,z)$ for such t. Moreover $|d_{\omega,t'}(X_i,V_i)^2 - \delta^{2a}| \leq 2\sqrt{\varepsilon}\delta^{2a}$ by (4.26). Thus the fact that $|N^-(y) \mathbin{\dot\cup} N^-(z)| \leq 2a$ and the definition of ω from (4.23) imply the following. If $\mathcal{C}o_{t,y,z}$ holds for all $t \leq T$, then

$$\begin{aligned}|\deg_{\omega,t'}(y,z) - d_{\omega,t'}(X_i,V_i)^2 n_i| &\leq |\deg_{\omega,t'}(y,z) - \delta^{2a}n_i| + |d_{\omega,t'}(X_i,V_i)^2 - \delta^{2a}|\, n_i \\ &\leq 2a\varepsilon n_i + 2\sqrt{\varepsilon}\delta^{2a}n_i \overset{(4.7)}{\leq} \sqrt[4]{\varepsilon}n_i\end{aligned}$$

for every $t' \leq T$. In other words, if $\mathcal{C}o_{t,y,z}$ holds for all $t \leq T$ then $\{y,z\} \notin K_k \cap W_i^{(2)}$. More precisely, we have the following.

Fact 4.25
For the smallest t with $\{y,z\} \in W_i^{(2)}(t)$ we have that $\mathcal{C}o_{t',y,z}$ holds for all $t' < t$ but not *for $t' = t$.*

Moreover, if $\mathcal{C}o_{t',y,z}$ holds for all $t' < t$ then

$$|\deg_{\omega,t}(y,z) - \deg_{\omega,0}(y,z)| \leq \big(\pi(t,y) + \pi(t,z)\big)\varepsilon n_i\,.$$

Recall that $\deg_{\omega,t}(y,z) = \delta^{a-\pi(t,y)}\delta^{a-\pi(t,z)}|C_{t,y} \cap C_{t,z}|$ and in particular (since $y, z \in X_i \setminus S_i$) $\deg_{\omega,0}(y,z) = \delta^{2a}n_i$. Hence

$$|C_{t,y} \cap C_{t,z}| \geq (\delta^{\pi(t,y)+\pi(t,z)} - \varepsilon\delta^{\pi(t,y)+\pi(t,z)-2a}(\pi(t,y) + \pi(t,z)))n_i \overset{(4.7)}{\geq} \varepsilon n_i\,. \tag{4.31}$$

We now claim that

$$\mathbb{P}[\mathcal{C}o_{t,y,z} \,|\, \mathcal{C}o_{t',y,z} \text{ for all } t' < t] \geq 1 - \frac{4\varepsilon}{\gamma}\,. \tag{4.32}$$

This is obvious if $x_t \notin N^-(y) \mathbin{\dot\cup} N^-(z)$. So assume we are about to embed an $x_t \in N^-(y) \mathbin{\dot\cup} N^-(z)$, which happens to be in X_j. Then $\varphi(x_t)$ is chosen randomly among at least $(\gamma/2)n_j$ vertices of V_j by Lemma 4.23. Out of those at most $2\varepsilon n_j$ vertices $v \in V_j$ have

$$\big|\deg(v, C_{t,y} \cap C_{t,z}) - d(V_i, V_j) \cdot |C_{t,y} \cap C_{t,z}|\big| > \varepsilon n_i$$

because $|C_{t,y} \cap C_{t,z}| \geq \varepsilon n_i$ by (4.31) and $G[V_i, V_j]$ is ε-regular. For every other choice of $\varphi(x_t) = v \in V_j$ we have

$$\begin{aligned}
&\big|\deg_{\omega,t+1}(y,z) - \deg_{\omega,t}(y,z)\big| \\
&\qquad = \big|\omega_{t+1}(y)\omega_{t+1}(z)\deg(v, C_{t,y} \cap C_{t,z}) - \omega_t(y)\omega_t(z) \cdot |C_{t,y} \cap C_{t,z}|\big| \\
&\qquad = \omega_{t+1}(y)\omega_{t+1}(z) \cdot \big|\deg(v, C_{t,y} \cap C_{t,z}) - \delta \cdot |C_{t,y} \cap C_{t,z}|\big| \\
&\qquad = \omega_{t+1}(y)\omega_{t+1}(z) \cdot \big|\deg(v, C_{t,y} \cap C_{t,z}) - d(V_i, V_j) \cdot |C_{t,y} \cap C_{t,z}|\big| \\
&\qquad \leq \omega_{t+1}(y)\omega_{t+1}(z) \cdot \varepsilon n_i \leq \varepsilon n_i\,.
\end{aligned}$$

Thus at most $2\varepsilon n_j$ out of $(\gamma/2)n_j$ choices for $\varphi(x_t)$ will result in $\overline{\mathcal{C}o_{t,y,z}}$, which implies (4.32), as claimed.

Finally, in order to show concentration, we will apply Lemma 4.19. For this purpose observe that by the construction of K_k for each time step $t \in [T]$ the embedding of x_t changes the co-degree of at most one pair in K_k, which we denote by $\{y_t, z_t\}$ if present. That is, $x_t \in N^-(y_t) \cup N^-(z_t)$. Now let $T_k \subseteq [T]$ be the set of time steps t with $\{y_t, z_t\}$ in K_k, i.e., let T_k be the set of time steps which actually change the co-degree of a pair in K_k. Since $|N^-(y) \cup N^-(z)| \leq 2a$ for every pair $\{y, z\} \in K_k$ we have $|T_k| \leq 2a|K_k|$. We define the following 0-1-variables $\mathcal{A}(t)$ for $t \in T_k$: Let $\mathcal{A}(t) = 1$ if and only if $\mathcal{C}o_{t',y_t,z_t}$ holds for all $t' \in [t-1] \cap T_k$ but not for $t' = t$. Fact 4.25 then implies $|K'_k| \leq \mathcal{A} := \sum_{t \in T_k} \mathcal{A}(t)$. Moreover, for any $t' < t$ with $\{y_{t'}, z_{t'}\} = \{y_t, z_t\}$ and $\mathcal{A}(t') = 1$ we have $\mathcal{A}(t) = 0$ by definition. Hence, for any $t \in T_k$ and $J \subseteq [t] \cap T_k$ we have

$$\mathbb{P}\left[\mathcal{A}(t) = 1 \,\middle|\, \begin{matrix} \mathcal{A}(t') = 1 \text{ for all } t' \in J \\ \mathcal{A}(t') = 0 \text{ for all } t' \in ([t] \cap T_k) \setminus J \end{matrix}\right] \leq 4\varepsilon/\gamma$$

by (4.32). Now either $|T_k| < 16a\varepsilon|K_k|/\gamma$ and thus $\mathcal{A} < 16a\varepsilon|K_k|/\gamma$ by definition. Or $|T_k| \geq 16a\varepsilon|K_k|/\gamma$ and

$$\mathbb{P}\left[\mathcal{A} \geq \frac{16a\varepsilon}{\gamma}|K_k|\right] \leq \mathbb{P}\left[\mathcal{A} \geq \frac{8\varepsilon}{\gamma}|T_k|\right] \leq \exp\left(-\frac{4\varepsilon}{3\gamma}|T_k|\right) \leq n_i^{-3},$$

by Lemma 4.19, where the last inequality follows from

$$\frac{4\varepsilon}{3\gamma}|T_k| \geq \frac{64a\varepsilon^2}{3\gamma^2}|K_k| \geq \frac{8\varepsilon^2 \log n}{3\gamma^2\xi\kappa r} \overset{(4.9)}{\geq} 3\log n_i\,.$$

Since $|K'_k| \leq \mathcal{A}$ and $16a\varepsilon/\gamma < \frac{1}{2}\sqrt[4]{\varepsilon}$ by (4.7) we obtain (4.30) as desired. □

We have now established that the auxiliary graph $F_i(t)$ for the embedding of X_i into V_i is weighted regular for all times $t \leq T$ with positive probability. The following lemma states that no critical set ever gets large in this case, i.e., if all auxiliary graphs remain weighted regular, then the RGA terminates successfully.

Lemma 4.26
For every $t \le T$ and $i \in [r]$ we have: $\mathcal{R}_i(t)$ implies that $|Q_i(t)| \le \varepsilon' n_i$. In particular, $\mathcal{R}_i$ for all $i \in [r]$ implies that the RGA completes the EMBEDDING STAGE *successfully.*

Proof. The idea of the proof is the following. Vertices only become critical because their available candidate set is significantly smaller than the average available candidate set. In other words, the weighted density between the set of critical vertices and $V_i^{\mathit{Free}}(t)$ deviates significantly from the weighted density of the auxiliary graph. Since the auxiliary graph is weighted regular it follows that there cannot be many critical vertices.

Indeed, assume for contradiction that there is $i \in [r]$ and $t \le T$ with $|Q_i(t)| > \varepsilon' n_i$ and such that $F_i(t)$ is weighted ε'-regular. Let $x \in Q_i(t)$ be an arbitrary critical vertex. Then x is an ordinary vertex and the available (ordinary) candidate set $A^{\mathrm{o}}_{t,x} = C_{t,x} \cap V_i^{\mathrm{o}} \cap V_i^{\mathit{Free}}(t)$ of x got small, that is,

$$|C_{t,x} \cap V_i^{\mathrm{o}} \cap V_i^{\mathit{Free}}(t)| \overset{(4.18)}{<} \gamma n_i \overset{(4.4)}{\le} \frac{\mu}{20}\delta^a n_i\,.$$

In the language of the auxiliary graph this means that

$$\deg_{\omega,t}(x, V_i^{\mathrm{o}} \cap V_i^{\mathit{Free}}(t)) = \omega_t(x)|C_{t,x} \cap V_i^{\mathrm{o}} \cap V_i^{\mathit{Free}}(t)| \le \frac{\mu}{20}\delta^a n_i.$$

Moreover $|V_i^{\mathrm{o}} \cap V_i^{\mathit{Free}}(t)| \ge |V_i^{\mathit{Free}}(t)| - |V_i^{\mathrm{s}}| \ge \frac{9}{10}\mu n_i \ge \varepsilon' n_i$. This implies

$$d_{\omega,t}(Q_i(t), V_i^{\mathrm{o}} \cap V_i^{\mathit{Free}}(t)) \le \frac{\mu/20\,\delta^a n_i}{9/10\,\mu n_i} = \frac{1}{18}\delta^a\,. \tag{4.33}$$

Since (4.26) and (4.33) imply that

$$d_{\omega,t}(X_i, V_i) - d_{\omega,t}(Q_i(t), V_i^{\mathrm{o}} \cap V_i^{\mathit{Free}}(t)) \ge \frac{1}{2}\delta^a - \frac{1}{18}\delta^a > \varepsilon',$$

but $F_i(t)$ is weighted ε'-regular we conclude that $|Q_i(t)| < \varepsilon' n_i$. □

Theorem 4.6 is now immediate from the following lemma.

Lemma 4.27
If we apply the RGA in the setting of Theorem 4.6, then with probability at least $2/3$ the event $\mathcal{R}_i$ holds for all $i \in [r]$ and the RGA finds an embedding of H' into G (obeying the R-partitions of H and G and the image restrictions).

Proof of Lemma 4.27. Let C, a, Δ_R, κ and δ, c, μ be given. Set the constants $\gamma, \varepsilon, \alpha$ as in (4.4)-(4.8). Let r be given and choose n_0, ξ as in (4.9)-(4.10). Further let R be a graph of order r with $\Delta(R) < \Delta_R$ and let G, H, H' have the required properties. Run the RGA with these settings. The INITIALISATION succeeds with probability at least 5/6 by Lemma 4.21. It follows from Lemma 4.24 that $\mathcal{R}_i$ occurs for all $i \in [r]$ with probability at least 5/6. This implies that no critical set Q_i ever violates the bound (4.19) by Lemma 4.26. Thus the EMBEDDING STAGE also succeeds with probability 5/6. We conclude that the RGA succeeds with probability at least 2/3. Thus an embedding φ of $H' = H[X_1' \,\dot\cup \ldots \dot\cup\, X_r']$ into G which maps X_i' into V_i exists. Moreover this embedding guarantees $\varphi(x) \in I(x)$ for all $x \in S_i \cap X_i'$ by definition of the algorithm. □

At the end of this section we want to point out that the minimum degree bound for H in Theorem 4.6 can be increased even further if we swap the order of the quantifiers. More precisely, for a fixed graph R we may choose ε such that almost spanning subgraphs of linear maximum degree can be embedded into a corresponding (ε, d)-regular R-partition.

Theorem 4.28
Given a graph R of order r and positive parameters a, κ, δ, μ there are $\varepsilon, \xi > 0$ such that the following holds. Assume that we are given

(i) a graph G with a κ-balanced (ε, δ)-regular R-partition $V(G) = V_1 \dot\cup \ldots \dot\cup V_r$ with $|V_i| =: n_i$ and

(ii) an a-arrangeable graph H with maximum degree $\Delta(H) \le \xi n$ (where $n = \sum n_i$), together with a corresponding R-partition $V(H) = X_1 \dot\cup \ldots \dot\cup X_r$ with $|X_i| \le (1-\mu)n_i$.

Then there is an embedding $\varphi\colon V(H) \to V(G)$ such that $\varphi(X_i) \subseteq V_i$.

Proof (sketch). Theorem 4.28 is deduced along the lines of the proof of Theorem 4.6.

Once more the randomised greedy algorithm from Section 4.3.2 is applied. It finds an embedding of H into G if all auxiliary graphs $F_i(t)$ remain weighted regular throughout the Embedding Stage. This in turn happens if each auxiliary graph $F_i(t)$ contains few pairs $\{x, y\} \in \binom{X_i}{2}$ whose weighted co-degree deviates from the expected value.

In the setting of Theorem 4.6 this is the case with positive probability as has been proven in Lemma 4.24: Inequality (4.29) states that the number of pairs with incorrect co-degree exceeds the bound of (4.28) with probability at most $1 - n_i^{-1}$. This particular argument is the only part of the proof of Theorem 4.6 that requires the degree bound of $\Delta(H) \le \xi n / \log n$. We then used (4.29) and a union bound over $i \in [r]$ to show that all auxiliary graphs $F_i(t)$ remain weighted regular throughout the Embedding Stage with probability at least 5/6. Since r can be large compared to all constants except n_0 we need the bound $1 - n_i^{-1}$ in (4.29).

In the setting of Theorem 4.28 however it suffices to replace this bound by a constant. More precisely, since we are allowed to choose ε depending on the order of R the proof of Lemma 4.24 becomes even simpler: Set ε small enough to ensure $3r\varepsilon^{3/4}\gamma \le 4a$. Note that inequality (4.32) implies that the expected number of pairs $\{y, z\} \in \binom{X_i}{2}$ with incorrect co-degree is bounded by $2a\frac{4\varepsilon}{\gamma}\binom{|X_i|}{2}$. It follows from Markov's inequality and the union bound over all $i \in [r]$ that all auxiliary graphs $F_i(t)$ remain weighted regular throughout the Embedding Stage with probability at least

$$1 - r\frac{\sqrt[4]{\varepsilon}\gamma}{8a\varepsilon} \ge \frac{5}{6}\,.$$

Choosing ε sufficiently small we can thus guarantee that the randomised greedy algorithm successfully embeds H into G with positive probability. □

Using the classical approach of Chvatal, Rödl, Szemerédi, and Trotter [25] Theorem 4.28 easily implies that all a-arrangeable graphs have linear Ramsey numbers. This result has first been proven (using the approach of [25]) by Chen and Schelp [22].

4.4 The spanning case

In this section we prove our main result, Theorem 4.4. We use the randomised greedy algorithm and its analysis from Section 4.3 to infer that the almost spanning embedding found in Theorem 4.6 can in fact be extended to a spanning embedding. We shortly describe our strategy in Section 4.4.1 and establish a minimum degree bound for the auxiliary graphs in Section 4.4.2 before we give the proof of Theorem 4.4 in Section 4.4.3. We conclude this section with a sketch of the proof of Theorem 4.5 in Section 4.4.4.

4.4.1 Outline of the proof

Let G, H satisfy the conditions of Theorem 4.4. We first use Lemma 4.13 to order the vertices of H such that the arrangeability of the resulting order is bounded and its last μn vertices form a stable set W. We then run the RGA to embed the almost spanning subgraph $H' = H[X \setminus W]$ into G. The RGA is successful and the resulting auxiliary graphs $F_i(T)$ are all weighted regular (that is, $\mathcal{R}_i$ holds) with probability $2/3$ by Lemma 4.27.

It remains to extend the embedding of H' to an embedding of H. Since W is stable it suffices to find for each $i \in [r]$ a bijection between

$$L_i := X_i \setminus W$$

and $V_i^{\mathit{Free}}(T)$ which respects the candidate sets, i.e., which maps x into $C_{T,x}$. Such a bijection is given by a perfect matching in $F_i^* := F_i(T)[L_i \cup V_i^{\mathit{Free}}(T)]$, which is the subgraph of $F_i(T)$ induced by the vertices left after the EMBEDDING PHASE of the RGA.

By Lemma 4.18 balanced weighted regular pairs with an appropriate minimum degree bound have perfect matchings. Now, $(L_i, V_i^{\mathit{Free}})$ is a subpair of a weighted regular pair and thus weighted regular itself by Proposition 4.15. Hence our main goal is to establish a minimum degree bound for $(L_i, V_i^{\mathit{Free}})$. More precisely we shall explain in Section 4.4.2 that it easily follows from the definition of the RGA that vertices in L_i have the appropriate minimum degree if $\mathcal{R}_i$ holds.

Proposition 4.29
Run the RGA in the setting of Theorem 4.4 and assume that $\mathcal{R}_j$ holds for all $j \in [r]$. Then every $x \in L_i$ has

$$\deg_{F_i(T)}(x, V_i^{\mathit{Free}}(T)) \geq 3\sqrt{\varepsilon'} n_i.$$

For vertices in V_i^{Free} on the other hand this is not necessarily true. But it holds with sufficiently high probability. This is also proved in Section 4.4.2.

Lemma 4.30
Run the RGA in the setting of Theorem 4.4 and assume that $\mathcal{R}_j$ holds for all $j \in [r]$. Then we have

$$\mathbb{P}\left[\forall i \in [r],\ \forall v \in V_i^{\mathit{Free}}(T) : \deg_{F_i(T)}(v, L_i) \geq 3\sqrt{\varepsilon'} n_i\right] \geq \frac{2}{3}\,.$$

4.4.2 Minimum degree bounds for the auxiliary graphs

In this section we prove Proposition 4.29 and Lemma 4.30. For the former we need to show that vertices $x \in L_i$ have an appropriate minimum degree in F_i^*, which is easy.

Proof of Proposition 4.29. Since $\mathcal{R}_j$ holds for all $j \in [r]$ the RGA completed the Embedding Stage successfully by Lemma 4.26. Note that all $x \in L_i$ did not get embedded yet. Thus

$$\deg_{F_i(T)}(x, V_i^{\mathit{Free}}(T)) = |A_{T,x}| \geq |A_{T,x}^{\mathrm{s}}| \geq \tfrac{7}{10}\gamma n_i \overset{(4.6)}{\geq} 3\sqrt{\varepsilon'} n_i \,,$$

for every $x \in L_i$ by Lemma 4.23. □

Lemma 4.30 claims that vertices in $V_i^{\mathit{Free}}(T)$ with positive probability also have a sufficiently large degree in F_i^*. We sketch the idea of the proof.

Let $x \in L_i$ and $v \in V_i^{\mathit{Free}}(T)$ for some $i \in [r]$. Recall that there is an edge $xv \in E(F_i(T))$ if and only if $\varphi(N^-(x)) \subseteq N_G(v)$. So we aim at lower-bounding the probability that $\varphi(N^-(x)) \subseteq N_G(v)$ for many vertices $x \in L_i$.

Now let $y \in N^-(x)$ be a predecessor of x. Recall that y is randomly embedded into $A(y)$, as defined in (4.17). Hence the probability that y is embedded into $N_G(v)$ is $|A(y) \cap N_G(v)|/|A(y)|$. Our goal will now be to show that these fractions are bounded from below by a constant for all predecessors of many vertices $x \in L_i$, which will then imply Lemma 4.30. To motivate this constant lower bound observe that a random subset A of X_j satisfies $|A \cap N_G(v)|/|A| = |N_G(v) \cap V_j|/|V_j|$ in expectation, and the right hand fraction is bounded from below by $\delta/2$ by (4.3). For this reason we call the vertex v *likely for* $y \in X_j$ and say that $\mathcal{A}_v(y)$ holds, if

$$\frac{|A(y) \cap N_G(v)|}{|A(y)|} \geq \frac{2}{3}\frac{|V_j \cap N_G(v)|}{|V_j|} \,.$$

Hence it will suffice to prove that for every $v \in V_i$ there are many $x \in L_i$ such that v is likely for all $y \in N^-(x)$.

We will focus on the last λn_i vertices x in $L_i \setminus N(S)$ (i.e., on vertices $x \in L_i^*$) as we have a good control over the embedding of their predecessors (who are in $X^* \setminus S$). Note that there indeed are λn_i vertices in $L_i \setminus N(S)$ as $\mu n_i - \alpha n_i \geq \lambda n_i$. For $i \in [r]$ and $v \in V_i$ we define

$$L_i^*(v) := \{x \in L_i^* : \mathcal{A}_v(y) \text{ holds for all } y \in N^-(x)\} \,.$$

Our goal is to show that a positive proportion of the vertices in L_i^* will be in $L_i^*(v)$. The following lemma makes this more precise.

Lemma 4.31
We run the RGA in the setting of Theorem 4.4 and assume that $\mathcal{R}_j$ holds for all $j \in [r]$. Then

$$\mathbb{P}\left[\forall i \in [r], \forall v \in V_i : |L_i^*(v)| \geq 2^{-a^2-1}|L_i^*|\right] \geq \frac{5}{6} \,.$$

Lemma 4.31 together with the subsequent lemma will imply Lemma 4.30.

Lemma 4.32
Run the RGA in the setting of Theorem 4.4 and assume that $\mathcal{R}_j$ holds for all $j \in [r]$ and that $|L_i^*(v)| \geq 2^{-a^2-1}|L_i^*|$. Then we have

$$\mathbb{P}\left[\forall i \in [r],\ \forall v \in V_i^{\mathit{Free}}(T) : \deg_{F_i(T)}(v, L_i) \geq 3\sqrt{\varepsilon'}n_i\right] \geq \frac{5}{6}\,.$$

Proof of Lemma 4.32. Let $i \in [r]$ and $v \in V_i$ be arbitrary and assume that the event of Lemma 4.31 occurs, this is, assume that we do have $|L_i^*(v)| \geq 2^{-a^2-1}|L_i^*|$. We claim that v almost surely has high degree in $F_i(T)$ in this case.

Claim *If $|L_i^*(v)| \geq 2^{-a^2-1}|L_i^*|$, then*

$$\mathbb{P}\left[\deg_{F_i(T)}(v, L_i) \geq 3\sqrt{\varepsilon'}n_i\right] \geq 1 - \frac{1}{n_i^2}\,.$$

This claim, together with a union bound over all $i \in [r]$ and $v \in V_i$, implies that

$$\mathbb{P}\left[\forall i \in [r], \forall v \in V_i : \deg_{F_i(T)}(v, L_i) \geq 3\sqrt{\varepsilon'}n_i\right] \geq \frac{5}{6}$$

if $|L_i^*(v)| \geq 2^{-a^2-1}|L_i^*|$ for all $i \in [r]$ and all $v \in V_i$. It remains to establish the claim.

Proof of Claim. Let $x \in L_i^*(v)$. Recall that $xv \in E(F_i)$ if and only if $\varphi(y) \in N_G(v)$ for all $y \in N^-(x)$. If the events $[\varphi(y) \in N_G(v)]$ were independent for all $y \in N^-(L_i^*(v))$ we could apply a Chernoff bound to infer that almost surely a linear number of the vertices $x \in L_i^*(v)$ is such that $[\varphi(y) \in N_G(v)]$ for all $y \in N^-(x)$. However, the events might be far from independent: just imagine two vertices x, x' sharing a predecessor y. We address this issue by partitioning the vertices into classes that do not share predecessors. We then apply Lemma 4.20 to those classes to finish the proof of the claim. Here come the details.

We partition $L_i^*(v)$ into predecessor disjoint sets. To do so we construct an auxiliary graph on vertex set $L_i^*(v)$ that has an edge xx' for exactly those vertices $x \neq x'$ that share at least one predecessor in H. As H is a-arrangeable, the maximum degree of this auxiliary graph is bounded by $a\Delta(H) - 1$. Hence we can apply Theorem 4.11 to partition the vertices of this auxiliary graph into stable sets $K_1 \,\dot\cup \ldots \dot\cup\, K_b$ with

$$|K_\ell| \geq \frac{|L_i^*(v)|}{a\Delta(H)} \geq \frac{2^{-a^2-1}\lambda\sqrt{n}}{a\,\kappa\, r}\log n \overset{(4.10)}{\geq} 48\left(\frac{3}{\delta}\right)^a \log n_i \tag{4.34}$$

for $\ell \in [b]$. Those sets are predecessor disjoint in H. We now want to apply Lemma 4.20. Let $\mathcal{I} = \{N^-(x) : x \in K_\ell\}$. The sets in $\mathcal{I}$ are pairwise disjoint and have at most a elements each. Name the elements of $\bigcup_{I\in\mathcal{I}} I = \{y_1, \ldots, y_s\}$ (with $s = |\bigcup_{I\in\mathcal{I}} I|$) in ascending order with respect to the arrangeable ordering. Furthermore, let $\mathcal{A}_k$ be a random variable which is 1 if and only if y_k gets embedded into $N_G(v)$. By the definition of $L_i^*(v)$, the event $\mathcal{A}_v(y_k)$ holds for each $k \in [s]$. It follows from the definition of $\mathcal{A}_v(y_k)$ that

$$\mathbb{P}[\mathcal{A}_k = 1] = \mathbb{P}[\varphi(y_k) \in N_G(v)] = \frac{|A(y_k) \cap N_G(v)|}{|A(y_k)|} \geq \frac{2}{3}\frac{|N_G(v) \cap V_j|}{|V_j|} \overset{(4.3)}{\geq} \frac{\delta}{3}\,.$$

This lower bound on the probability of $\mathcal{A}_k = 1$ remains true even if we condition on other events $\mathcal{A}_j = 1$ (or their complements $\mathcal{A}_j = 0$), because in this calculation the lower bound relies solely on $|A(y_k) \cap N_G(v)|/|A(y_k)|$, which is at least $\delta/3$ for all $k \in [m]$ regardless of the embedding of other y_j. Hence,

$$\mathbb{P}\left[\mathcal{A}_k = 1 \middle| \begin{matrix}\mathcal{A}_j = 1 \text{ for all } j \in J \\ \mathcal{A}_j = 0 \text{ for all } j \in [k-1] \setminus J\end{matrix}\right] \geq \frac{\delta}{3}$$

for every k and every $J \subseteq [k-1]$ (this is stronger than the condition required by Lemma 4.20). By Lemma 4.20, we have

$$\begin{aligned}
\mathbb{P}&\left[\left|\left\{x \in K_\ell : \varphi(N^-(x)) \subseteq N_G(v)\right\}\right| \geq \frac{1}{2}\left(\frac{\delta}{3}\right)^a |K_\ell|\right] \\
&\qquad = \mathbb{P}\left[\left|\left\{I \in \mathcal{I} : \mathcal{A}_i = 1 \text{ for all } i \in I\right\}\right| \geq \frac{1}{2}\left(\frac{\delta}{3}\right)^a |K_\ell|\right] \\
&\qquad \geq 1 - 2\exp\left(-\frac{1}{12}\left(\frac{\delta}{3}\right)^a |K_\ell|\right) \\
&\qquad \overset{(4.34)}{\geq} 1 - 2\exp(-4\log n_i) = 1 - 2 \cdot n_i^{-4}\,.
\end{aligned}$$

Applying a union bound over all $\ell \in [b]$ we conclude that

$$\begin{aligned}
\deg_{F_i(T)}(v, L_i) &\geq \left|\left\{x \in L_i^*(v) : \varphi(N^-(x)) \subseteq N_G(v)\right\}\right| \\
&\geq \sum_{\ell \in [b]} \frac{1}{2}\left(\frac{\delta}{3}\right)^a |K_\ell| = \frac{1}{2}\left(\frac{\delta}{3}\right)^a |L_i^*(v)| \geq \frac{\delta^a}{2 \cdot 2^{a^2+1} 3^a}\lambda n_i \overset{(4.6)}{\geq} 3\sqrt{\varepsilon'} n_i
\end{aligned}$$

with probability at least $1 - 2n_i \exp(-4\log n_i) \geq 1 - n_i^{-2}$. □

This concludes the proof of the lemma. □

The remainder of this section is dedicated to the proof of Lemma 4.31. This proof will use similar ideas as the proof of Lemma 4.32. This time, however, we are not only interested in the predecessors of $x \in L_i^*$ but in the predecessors of the predecessors. We call those *predecessors of second order* and say two vertices x, x' are *predecessor disjoint of second order* if $N^-(x) \cap N^-(x') = \emptyset$ and $N^-(N^-(x)) \cap N^-(N^-(x')) = \emptyset$.

To prove Lemma 4.31, we have to show that for any vertex $v \in V_i$ many vertices x in L_i^* are such that all their predecessors $y \in N^-(x)$ are likely for v. Note that $x \in L_i^*$ implies that $y \in N^-(x)$ gets embedded into the special candidate set $C^{\mathrm{s}}_{t(y),y}$.

It depends only on the embedding of the vertices in $N^-(y)$ whether a given vertex $v \in V_i$ is likely for y or not. Therefore, we formulate an event $\mathcal{B}_{v,x}(z)$, which, if satisfied for all $z \in N^-(y)$, will imply $\mathcal{A}_v(y)$ as we will show in the next proposition. Recall that $C^{\mathrm{s}}_{1,y} = V^{\mathrm{s}}_j$ for $y \in N^-(L^*_i) \subseteq X^*$ and $C^{\mathrm{s}}_{t(z)+1,y} = C^{\mathrm{s}}_{t(z),y} \cap N_G(\varphi(z))$. For $x \in L^*_i$ and $z \in N^-(N^-(x))$ let $\mathcal{B}_{v,x}(z)$ be the event that

$$\left| \frac{|C^{\mathrm{s}}_{t(z),y} \cap N_G(v)|}{|C^{\mathrm{s}}_{t(z),y}|} - \frac{|C^{\mathrm{s}}_{t(z)+1,y} \cap N_G(v)|}{|C^{\mathrm{s}}_{t(z)+1,y}|} \right| \leq \frac{2\varepsilon}{\delta - \varepsilon}$$

for all $y \in N^-(x)$.

Proposition 4.33
Let $i \in [r]$, $v \in V_i$, $x \in L^*_i$, and $z \in N^-(N^-(x))$, then

$$\mathbb{P}\left[\mathcal{B}_{v,x}(z) \middle| \mathcal{B}_{v,x}(z') \text{ for all } z' \in N^-(N^-(x)) \text{ with } t(z') < t(z)\right] \geq 1/2.$$

This remains true if we additionally condition on other events $\mathcal{B}_{v,\widetilde{x}}(\widetilde{z})$ (or their complements) with $\widetilde{z} \in N^-(N^-(\widetilde{x}))$ for $\widetilde{x} \in L^*_i$, as long as x and $\widetilde{x}$ are predecessor disjoint of second order.
Moreover, if $\mathcal{B}_{v,x}(z)$ occurs for all $z \in N^-(N^-(x))$, then $\mathcal{A}_v(y)$ occurs for all $y \in N^-(x)$.

Proof of Proposition 4.33. Let $x \in L^*_i$ and let $z \in N^-(N^-(x))$ lie in X_ℓ. Further assume that $\mathcal{B}_{v,x}(z')$ holds for all $z' \in N^-(N^-(x))$ with $t(z') < t(z)$. For $y \in N^-(x)$ let $j(y)$ be such that $y \in X_{j(y)}$. Then $\mathcal{B}_{v,x}(z')$ for all $z' \in N^-(y)$ with $t(z') < t(z)$ implies

$$\begin{aligned} \frac{|C^{\mathrm{s}}_{t(z),y} \cap N_G(v)|}{|C^{\mathrm{s}}_{t(z),y}|} &\geq \frac{|C^{\mathrm{s}}_{1,y} \cap N_G(v)|}{|C^{\mathrm{s}}_{1,y}|} - \frac{2\varepsilon \cdot a}{\delta - \varepsilon} \\ &= \frac{|V_{j(y)} \cap N_G(v)|}{|V_{j(y)}|} - \frac{2\varepsilon \cdot a}{\delta - \varepsilon} \overset{(4.3)}{\geq} \frac{\delta}{2} - \frac{2\varepsilon \cdot a}{\delta - \varepsilon} \end{aligned}$$

where the identity $C_{1,y} = V_{j(y)}$ is due to $y \notin S$. Hence $|C^{\mathrm{s}}_{t(z),y} \cap N_G(v)| \geq \varepsilon n_{j(\ell)}$ by (4.16) and our choice of constants. Now fix a $y \in N^-(x)$. As $(V_{j(y)}, V_\ell)$ is an ε-regular pair all but at most $4\varepsilon n_\ell$ vertices $w \in A_{t(z),z} \subseteq V_\ell$ simultaneously satisfy

$$\begin{aligned} \left| \frac{\left|N_G\big(w, C^{\mathrm{s}}_{t(z),y} \cap N_G(v)\big)\right|}{|C^{\mathrm{s}}_{t(z),y} \cap N_G(v)|} - d(V_{j(y)}, V_\ell) \right| &\leq \varepsilon\,, \text{ and} \\ \left| \frac{|N_G(w, C^{\mathrm{s}}_{t(z),y})|}{|C^{\mathrm{s}}_{t(z),y}|} - d(V_{j(y)}, V_\ell) \right| &\leq \varepsilon\,. \end{aligned}$$

Hence, all but at most $4\varepsilon a n_\ell$ vertices in V_ℓ satisfy the above inequalities for all $y \in N^-(x)$. If $\varphi(z) = w$ for a vertex w that satisfies the above inequalities for all $y \in N^-(x)$ we have

$$\left| \frac{|C^{\mathrm{s}}_{t(z)+1,y} \cap N_G(v)|}{|C^{\mathrm{s}}_{t(z),y} \cap N_G(v)|} - \frac{|C^{\mathrm{s}}_{t(z)+1,y}|}{|C^{\mathrm{s}}_{t(z),y}|} \right| \leq 2\varepsilon\,.$$

This implies $B_{v,x}(z)$ as

$$\left|\frac{|C^{\mathrm{s}}_{t(z)+1,y}\cap N_G(v)|}{|C^{\mathrm{s}}_{t(z)+1,y}|}-\frac{|C^{\mathrm{s}}_{t(z),y}\cap N_G(v)|}{|C^{\mathrm{s}}_{t(z),y}|}\right|\le 2\varepsilon\frac{|C^{\mathrm{s}}_{t(z),y}\cap N_G(v)|}{|C^{\mathrm{s}}_{t(z)+1,y}|}$$
$$\le 2\varepsilon\frac{|C^{\mathrm{s}}_{t(z),y}|}{|C^{\mathrm{s}}_{t(z)+1,y}|}\overset{(4.15)}{\le}\frac{2\varepsilon}{\delta-\varepsilon}\,.$$

Since $\varphi(z)$ is chosen randomly from $A(z)\subseteq A_{t(z),z}$ with $|A(z)|\ge(\gamma/2)n_\ell$ by Lemma 4.23, we obtain

$$\mathbb{P}\left[\mathcal{B}_{v,x}(z)\middle|\mathcal{B}_{v,x}(z')\text{ for all }z'\in N^-(N^-(x)),\ t(z')<t(z)\right]\ge 1-\frac{4\varepsilon a n_\ell}{(\gamma/2)n_\ell}\ge\frac{1}{2}\,.$$

Note that this probability follows alone from the ε-regularity of the pairs $(V_{j(y)},V_\ell)$ and the fact that $A(z)$ and $C^{\mathrm{s}}_{t(z),y}\cap N_G(v)$ are large. If x and $\widetilde{x}$ are predecessor disjoint of second order the outcome of the event $\mathcal{B}_{v,\widetilde{x}}(\widetilde{z})$ for $\widetilde{z}\in N^-(N^-(\widetilde{x}))$ does not influence those parameters. We can therefore condition on other events $\mathcal{B}_{v,\widetilde{x}}(\widetilde{z})$ as long as x and $\widetilde{x}$ are predecessor disjoint of second order.

It remains to show the second part of the proposition, that v is likely for all $y\in N^-(x)$ if $\mathcal{B}_{v,x}(z)$ holds for all $z\in N^-(N^-(x))$. Again let $x\in L_i^*$ and let $y\in N^-(x)$ lie in X_j. Recall that condition (4.13) in the definition of the RGA guarantees

$$\left|\frac{|V_j^{\mathrm{s}}\cap N_G(v)|}{|V_j^{\mathrm{s}}|}-\frac{|V_j\cap N_G(v)|}{|V_j|}\right|\le\varepsilon\,.$$

Moreover, $\mathcal{B}_{v,x}(z)$ for all $z\in N^-(y)$, $C^{\mathrm{s}}_{1,y}=V_j^{\mathrm{s}}$ (as $y\notin S$) and the fact that $|N^-(y)|\le a$ imply

$$\left|\frac{|C^{\mathrm{s}}_{t(y),y}\cap N_G(v)|}{|C^{\mathrm{s}}_{t(y),y}|}-\frac{|V_j^{\mathrm{s}}\cap N_G(v)|}{|V_j^{\mathrm{s}}|}\right|\le\frac{2\varepsilon\cdot a}{\delta-\varepsilon}\,.$$

As $2\varepsilon a/(\delta-\varepsilon)+\varepsilon\le\delta/36\le(\delta/18)|V_j\cap N_G(v)|/|V_j|$ we conclude that

$$\frac{|C^{\mathrm{s}}_{t(y),y}\cap N_G(v)|}{|C^{\mathrm{s}}_{t(y),y}|}\ge\frac{17}{18}\frac{|V_j\cap N_G(v)|}{|V_j|}\tag{4.35}$$

for all $y\in N^-(x)$ if $\mathcal{B}_{v,x}(z)$ for all $z\in N^-(N^-(x))$. Equation (4.35) in turn implies $\mathcal{A}_v(y)$ as only few vertices get embedded into V^{s} thus making $C^{\mathrm{s}}_{t,y}\approx A^{\mathrm{s}}_{t,y}$. More precisely, by Lemma 4.23 we have

$$\begin{aligned}\frac{|A(y)\cap N_G(v)|}{|A(y)|}&\ge\frac{|A^{\mathrm{s}}_{t(y),y}\cap N_G(v)|-|A^{\mathrm{s}}_{t(y),y}\setminus A(y)|}{|A^{\mathrm{s}}_{t(y),y}|}\\&\ge\frac{|C^{\mathrm{s}}_{t(y),y}\cap N_G(v)|-|X_j^*|-|Q_j(t(y))|-|A^{\mathrm{s}}_{t(y),y}\setminus A(y)|}{|C^{\mathrm{s}}_{t(y),y}|}\\&\ge\frac{|C^{\mathrm{s}}_{t(y),y}\cap N_G(v)|}{|C^{\mathrm{s}}_{t(y),y}|}-\frac{\delta}{18}\\&\overset{(4.35)}{\ge}\left(\frac{17}{18}-\frac{2}{18}\right)\frac{|V_j\cap N_G(v)|}{|V_j|}\,.\end{aligned}$$

□

We have seen that $x \in L_i^*$ also lies in $L_i^*(v)$ if $\mathcal{B}_{v,x}(z)$ holds for all $z \in N^-(N^-(x))$. To prove Lemma 4.31 it therefore suffices to show that an arbitrary vertex v has a linear number of vertices $x \in L_i^*$ with $\mathcal{B}_{v,x}(z)$ for all $z \in N^-(N^-(x))$.

Proof of Lemma 4.31. Let $i \in [r]$ and $v \in V_i$ be arbitrary. We partition L_i^* into classes of vertices that are predecessor disjoint of second order. Observe that for every $x \in L_i^*$ we have

$$\left|\left\{x' \in L_i^* : \begin{array}{l} N^-(x) \cap N^-(x') \neq \emptyset \text{ or} \\ N^-(N^-(x)) \cap N^-(N^-(x')) \neq \emptyset \end{array}\right\}\right| \leq (a\Delta(H))^2 \leq \frac{a^2 n}{\log^2 n} \overset{(4.10)}{<} \frac{\lambda n_i}{36 \cdot 2^{a^2} \log n_i}$$

as H is a-arrangeable. Recall that $|L_i^*| = \lambda n_i$. Therefore, Theorem 4.11 gives a partition $L_i^* = K_1 \,\dot\cup \ldots \dot\cup\, K_b$ with

$$|K_\ell| \geq 36 \cdot 2^{a^2} \log n_i \tag{4.36}$$

for all $\ell \in [b]$ such that the vertices in K_ℓ are predecessor disjoint of second order.

Next we want to apply Lemma 4.20. Let $\ell \in [b]$ be fixed. We define $\mathcal{I} = \{N^-(N^-(x)) : x \in K_\ell\}$. These sets are pairwise disjoint and have at most a^2 elements each. Name the elements of $\bigcup_{I \in \mathcal{I}} I = \{z_1, \ldots, z_{|\cup_{I\in\mathcal{I}} I|}\}$ in ascending order with respect to the arrangeable ordering. Then for every $I \in \mathcal{I}$ and every $z_k \in I$ we have

$$\mathbb{P}\left[\mathcal{B}_{v,x}(z_k) \middle| \begin{array}{l} \mathcal{B}_{v,x}(z_j) \text{ for all } z_j \in J, \\ \overline{\mathcal{B}_{v,x}}(z_j) \text{ for all } z_j \in \{z_1, \ldots, z_{k-1}\} \setminus J \end{array}\right] \geq \frac{1}{2}$$

for every $J \subseteq \{z_1, \ldots, z_{k-1}\}$ with $\{z_1, \ldots, z_{k-1}\} \cap I \subseteq J$ by Proposition 4.33. We set $K_\ell(v) := \{x \in K_\ell : \mathcal{B}_{v,x}(z) \text{ for all } z \in N^-(N^-(x))\}$ and apply Lemma 4.20 to derive

$$\mathbb{P}\left[|K_\ell(v)| \geq 2^{-a^2-1}|K_\ell|\right] \geq 1 - 2\exp\left(-\frac{1}{12}2^{-a^2}|K_\ell|\right) \overset{(4.36)}{\geq} 1 - 2\exp(-3\log n_i) = 1 - 2 \cdot n_i^{-3}.$$

Note that we have $\bigcup_{\ell\in[b]} K_\ell(v) \subseteq L_i^*(v)$ as the following is true for every $x \in K_\ell$ by Proposition 4.33: $\mathcal{B}_{v,x}(z)$ for all $z \in N^-(N^-(x))$ implies $\mathcal{A}_v(y)$ for all $y \in N^-(x)$. Taking a union bound over all $\ell \in [b]$ we thus obtain that

$$\mathbb{P}\left[|L_i^*(v)| \geq 2^{-a^2-1}|L_i^*|\right] \geq 1 - b \cdot 2n_i^{-3} \geq 1 - \frac{2}{n_i^2}.$$

One further union bound over all $i \in [r]$ and $v \in V_i$ finishes the proof. □

4.4.3 Proof of Theorem 4.4

Putting everything together, we conclude that the RGA gives a spanning embedding of H into G with probability at least 1/3. We now use Lemma 4.18, Proposition 4.29, and Lemma 4.30 to prove our main result.

Proof of Theorem 4.4. Let integers C, a, Δ_R, κ and $\delta, c > 0$ be given. Set $a' = 5a^2\kappa\Delta_R$ and $\mu = 1/(10a'(\kappa\Delta_R)^2)$. We invoke Theorem 4.6 with parameters C, a', Δ_R, κ and $\delta, c, \mu > 0$ to obtain $\varepsilon, \alpha > 0$. Let r be given and choose n_0 as in Theorem 4.6.

Now let R be a graph on r vertices with $\Delta(R) < \Delta_R$. And let G and H satisfy the conditions of Theorem 4.4, i.e., let G have the (ε, δ)-super-regular R-partition $V(G) = V_1 \cup\!\!\!\!\cdot\;\ldots \cup\!\!\!\!\cdot\; V_r$ and let H have a κ-balanced R-partition $V(H) = X_1 \cup\!\!\!\!\cdot\;\ldots \cup\!\!\!\!\cdot\; X_r$. Further let $\{x_1, \ldots, x_n\}$ be an a-arrangeable ordering of H. We apply Lemma 4.13 to find an a'-arrangeable ordering $\{x'_1, \ldots, x'_n\}$ of H with a stable ending of order μn. Let $H' = H[\{x'_1, \ldots, x'_{(1-\mu)n}\}]$ be the subgraph induced by the first $(1-\mu)n$ vertices of the new ordering. We take this ordering and run the RGA as described in Section 4.3.2 to embed H' into G. By Lemma 4.27 we have

$$\mathbb{P}\left[\text{RGA successful and } \mathcal{R}_i \text{ for all } i \in [r]\right] \geq \frac{2}{3}\,, \tag{4.37}$$

where $\mathcal{R}_i$ is the event that the auxiliary graph $F_i(t)$ is weighted ε'-regular for all $t \leq T$. Note that every image restricted vertex $x \in S_i \cap V(H')$ has been embedded into $I(x)$ by the definition of the RGA.

Now assume that $\mathcal{R}_i$ holds for all $i \in [r]$. It remains to embed the stable set $L_i = V(H) \setminus V(H')$. To this end we shall find in each $F_i^* := F_i(T)[L_i \cup\!\!\!\!\cdot\; V_i^{\mathit{Free}}(T)]$ a perfect matching, which defines a bijection from L_i to $V_i^{\mathit{Free}}(T)$ that maps every $x \in L_i$ to a vertex $v \in V_i^{\mathit{Free}}(T) \cap C_{T,x}$. Note that again $x \in S_i$ is embedded into $I(x)$ by construction.

Since $F_i(T)$ is weighted ε'-regular the subgraph F_i^* is weighted (ε'/μ)-regular by Proposition 4.15. Moreover

$$\mathbb{P}\left[\delta(F_i^*) \geq 3\sqrt{\varepsilon'}n_i \text{ for all } i \in [r]\right] \geq \frac{2}{3} \tag{4.38}$$

by Proposition 4.29 and Lemma 4.30. In other words, with probability at least $2/3$ all graphs $F_i^* = F_i(T)[L_i \cup\!\!\!\!\cdot\; V_i^{\mathit{Free}}]$ are balanced, bipartite graphs on $2\mu n_i$ vertices with $\deg(x) \geq 3\sqrt{\varepsilon'/\mu}(\mu n_i)$ for all $x \in L_i \cup V_i^{\mathit{Free}}$. Also note that $\omega(x) \geq \delta^a \geq \sqrt{\varepsilon'/\mu}$ for all $x \in L_i$ by definition of ε'. We conclude from Lemma 4.18 that F_i^* has a perfect matching if F_i^* has minimum degree at least $3\sqrt{\varepsilon'}n_i$. Hence, combining (4.37) and (4.38) we obtain that the RGA terminates successfully and all F_i^* have perfect matchings with probability at least $1/3$. Thus there is an almost spanning embedding of H' into G that can be extended to a spanning embedding of H into G. □

4.4.4 Proof of Theorem 4.5

We close this section by sketching the proof of Theorem 4.5, which is very similar to the proof of Theorem 4.4. We start by quickly summarising the latter. For two graphs G and H let the partitions $V = V_1 \cup\!\!\!\!\cdot\;\ldots \cup\!\!\!\!\cdot\; V_k$ and $X = X_1 \cup\!\!\!\!\cdot\;\ldots \cup\!\!\!\!\cdot\; X_k$ satisfy the requirements of Theorem 4.4. In order to find an embedding of H into G that maps the vertices of X_i onto V_i we proceeded in two steps. First we used a randomised greedy algorithm to embed an almost spanning part of H into G. This left us with sets $L_i \subseteq X_i$ and $V_i^{Free} \subseteq V_i$. We then found a bijection between the L_i and V_i^{Free} that completed the embedding of H.

More precisely, we did the following. We ran the randomised greedy algorithm from Section 4.3.2 and defined auxiliary graphs $F_i(t)$ on vertex sets $V_i \,\dot\cup\, X_i$ that kept track of all possible embeddings at time t of the embedding algorithm. We showed that the randomised greedy embedding succeeds for the almost spanning subgraph if all the auxiliary graphs remain weighted regular (Lemma 4.26). This in turn happens with probability at least 2/3 by Lemma 4.27. This finished stage one of the embedding (and also proved Theorem 4.6).

For the second stage of the embedding we assumed that stage one found an almost spanning embedding by time T and that all auxiliary graphs are weighted regular. We defined $F_i^*(T)$ to be the subgraph of $F_i(T)$ induced by $L_i \,\dot\cup\, V_i^{Free}$. This subgraph inherits (some) weighted regularity from $F_i(T)$. Moreover, we showed that all $F_i^*(T)$ have a minimum degree which is linear in n_i with probability at least 2/3 (see Proposition 4.29 and Lemma 4.30). Each $F_i^*(T)$ has a perfect matching in this case by Lemma 4.18. Those perfect matchings then gave the bijection of L_i onto V_i^{Free} that completed the embedding of H into G. We concluded that with probability at least 2/3 the almost spanning embedding found by the randomised greedy algorithm in stage one can be extended to a spanning embedding of H into G.

For the proof of Theorem 4.5 we proceed in exactly the same way. Note that Theorem 4.4 and Theorem 4.5 differ only in the following aspects. The first allows a maximum degree of $\sqrt{n}/\log n$ for H while the latter extends this to $\Delta(H) \leq \xi n/\log n$. This does not come free of charge. Theorem 4.5 not only requires the R-partition of G to be super-regular but also imposes what we call the *tuple condition*, that every tuple of $a+1$ vertices in $V \setminus V_i$ have a linearly sized joint neighbourhood in V_i. We now sketch how one has to change the proof of Theorem 4.4 to obtain Theorem 4.5.

Again we proceed in two stages. The first of those, which gives the almost spanning embedding, is identical to the previously described one: here the larger maximum degree is not an obstacle (see also Theorem 4.6). Again all auxiliary graphs are weighted regular by the end of the EMBEDDING PHASE with probability at least 2/3. Moreover, all vertices in L_i have linear degree in $F_i^*(T)$ by the same argument as before (see Proposition 4.29 and its proof).

It now remains to show to show that every vertex v in V_i^{Free} has a linear degree in the auxiliary graph $F_i^*(T)$. At this point we deviate from the proof of Theorem 4.4. Recall that L_i^* was defined as the last λn_i vertices of $X_i \setminus N(S)$ in the arrangeable ordering and $L_i^*(v)$ was defined as the set of vertices $x \in L_i^*$ with $\mathcal{A}_v(y)$ for all $y \in N^-(x)$. We still want to prove that $L_i^*(v)$ is large for every v as this again would imply the linear minimum degree for all $v \in V_i^{Free}$. However, the maximum degree $\Delta(H) \leq \xi n/\log n$ does not allow us to partition L_i into sets which are predecessor disjoint of second order any more. This, however, was crucial for our proof of $|L_i^*(v)| \geq 2^{-a^2-1}|L_i^*|$ (see the proof of Lemma 4.31).

We may, however, alter the definition of the event $\mathcal{A}_v(y)$ to overcome this obstacle. Instead of requiring that $|A(y) \cap N_G(v)|/|A(y)| \geq (2/3)|V_j \cap N_G(v)|/|V_j|$, we now define

$\mathcal{A}_v(y)$ in the proof of Theorem 4.5 to be the event that

$$\frac{|A(y) \cap N(v)|}{|A(y)|} \geq \frac{\iota}{2}\,.$$

We still denote by $L_i^*(v)$ the set of vertices $x \in L_i^*$ with $\mathcal{A}_v(y)$ for all $y \in N^-(x)$. Now the tuple condition guarantees that $|C_{t(y),y} \cap N_G(v)| \geq \iota n_j$ for any $y \in X_j$ and $v \in V \setminus V_j$. Since we chose $V_j^{\mathrm{s}} \subseteq V_j$ randomly we obtain $|C^{\mathrm{s}}_{t(y),y} \cap N_G(v)| \geq (\mu/20)\iota n_j$ for all $y \in X_j$ *almost surely*. The same arguments as in the proof of Proposition 4.33 imply that $A(y) \approx C^{\mathrm{s}}_{t(y),y}$ for all y that are predecessors of vertices $x \in L_i^*$. Hence,

$$\frac{|A(y) \cap N_G(v)|}{|A(y)|} \approx \frac{|C^{\mathrm{s}}_{t(y),y} \cap N_G(v)|}{|C^{\mathrm{s}}_{t(y),y}|} \approx \frac{(\mu/20)\iota n_j}{(\mu/10)\delta^a n_j} \geq \frac{\iota}{2}$$

for all $x \in L_i^*$ and all $y \in N^-(x)$ almost surely. If this is the case we have $|L_i^*(v)| = |L_i^*| = \lambda n_i$ and therefore the assertion of Lemma 4.31 holds also in this setting. It remains to show that the same is true for Lemma 4.32. Indeed, after some appropriate adjustments of the constants, the very same argument implies $\deg_{F_i(T)}(v, L_i) \geq 3\varepsilon' n_i$ for all $i \in [r]$ and $v \in V_i^{\mathit{Free}}$ if $|L_i^*(v)| = |L_i^*|$. More precisely, the change in the definition of $\mathcal{A}_v(y)$ will force smaller values of ε', that is, the constant in the bound of the joint neighbourhood of each $(a+1)$-tuple has to be large compared to the ε in the ε-regularity of the partition $V_1 \mathbin{\dot\cup} \ldots \mathbin{\dot\cup} V_k$. The constants then relate as

$$0 < \xi \ll \varepsilon \ll \varepsilon' \ll \lambda \ll \gamma \ll \mu, \delta, \iota \leq 1\,.$$

The remaining steps in the proof of Theorem 4.5 are identical to those in the proof of Theorem 4.4. For $i \in [r]$ the minimum degree in $F_i^*(T)$ together with the weighted regularity implies that $F_i^*(T)$ has a perfect matching. The perfect matching defines a bijection of L_i onto V_i^{Free} that in turn completes the embedding of H into G.

To wrap up, let us quickly comment on the different degree bounds for H in Theorem 4.4 and Theorem 4.5. The proof of Theorem 4.5 just sketched only requires $\Delta(H) = \xi n/\log n$. This is needed to partition L_i^* into *predecessor disjoint sets* in the last step in order to prove the minimum degree for the auxiliary graphs.

Contrary to that the proof of Theorem 4.4 partitions the vertices of L_i^* into sets which are *predecessor disjoint of second order*, i.e., which do have $N^-(N^-(x)) \cap N^-(N^-(x')) = \emptyset$ for all $x \neq x'$. This is necessary to ensure that there is a linear number of vertices x in L_i^* with $\mathcal{A}_v(y)$ for all $y \in N^-(x)$, i.e., whose predecessors all get embedded into $N_G(v)$ with probability $\delta/3$. More precisely we ensure that, all predecessors y of x have the following property. The predecessors $z_1, \ldots, z_k$ of y are embedded to a k-tuple $(\varphi(z_1), \ldots, \varphi(z_k))$ of vertices in G such that $\bigcap N(\varphi(z_i)) \cap N_G(v) \cap V_{j(y)}$ is large. This fact follows trivially from the tuple condition of Theorem 4.5 and hence we don't need a partition into predecessor disjoint sets of second order.

4.5 Optimality

The aim of this section is twofold. Firstly, we shall investigate why the degree bounds given in Theorem 4.4 and in Theorem 4.5 are best possible. Secondly, we shall why the conditions Theorem 4.4 imposes on image restrictions are so restrictive.

Optimality of Theorem 4.5. To argue that the requirement $\Delta(H) \leq n/\log n$ is optimal up to the constant factor we use a construction from [60] and the following proposition.

Proposition 4.34
For every $\varepsilon > 0$ the domination number of a graph $\mathcal{G}(n,p)$ with high probability is larger than $(1-\varepsilon)p\log n$.

Proof. The probability that a graph in $\mathcal{G}(n,p)$ has a dominating set of size r is bounded by

$$\binom{n}{r}\left(1-(1-p)^r\right)^{n-r} \leq \exp\left(r\log n - \exp(-rp)(n-r)\right) .$$

Setting $r = (1-\varepsilon)p\log n$ we obtain

$$\mathbb{P}\begin{bmatrix}\mathcal{G}(n,p) \text{ has a dominating} \\ \text{set of size } (1-\varepsilon)p\log n\end{bmatrix} \leq \exp\left((1-\varepsilon)p\log^2 n - \frac{n-(1-\varepsilon)p\log n}{n^{1-\varepsilon}}\right) \to 0$$

for every (fixed) positive ε. □

Let H be a tree with a root of degree $\frac{1}{2}\log n$, such that each neighbour of this root has $2n/\log n$ leaves as neighbours. This graph H almost surely is not a subgraph of $\mathcal{G}(n, 0.9)$ by Proposition 4.34 as the neighbours of the root form a dominating set.

Optimality of Theorem 4.4. The degree bound $\Delta(H) \leq \sqrt{n}/\log n$ is optimal up to the log-factor. More precisely, we can show the following.

Proposition 4.35
For every $\varepsilon > 0$ and $n_0 \in \mathbb{N}$ there are $n \geq n_0$, an $(\varepsilon, 1/2)$-super-regular pair (V_1, V_2) with $|V_1| = |V_2| = n$ and a tree $T \subseteq K_{n,n}$ with $\Delta(T) \leq \sqrt{n}+1$, such that (V_1, V_2) does not contain T.

Condition (e) of Theorem 4.4 allows only a constant number of permissible image restrictions per cluster. The following proposition shows that also this is best possible (up to the value of the constant).

Proposition 4.36
For every $\varepsilon > 0$, $n_0 \in \mathbb{N}$, and every $w\colon \mathbb{N} \to \mathbb{N}$ which goes to infinity arbitrarily slowly, there are $n \geq n_0$, an $(\varepsilon, 1/2)$-super-regular pair (V_1, V_2) with $|V_1| = |V_2| = n$ and a tree $T \subseteq K_{n,n}$ with $\Delta(T) \leq w(n)$ such that the following is true. The images of $w(n)$ vertices of T can be restricted to sets of size $n/2$ in $V_1 \cup V_2$ such that no embedding of T into (V_1, V_2) respects these image restrictions.

We remark that our construction for Proposition 4.36 does not require a spanning tree T, but only one on $w(n)+1$ vertices. Moreover, this proposition shows that the number of admissible image restrictions drops from linear (in the original Blow-up Lemma) to constant (in Theorem 4.4), if the maximum degree of the target graph H increases from constant to an increasing function.

We now give the constructions that prove these two propositions.

Proof of Proposition 4.35 (sketch). Let $\varepsilon > 0$ and n_0 be given, choose an integer k such that $1/k \ll \varepsilon$ and an integer n such that $k, n_0 \ll n$, and consider the following bipartite graph $G_k = (V_1 \mathbin{\dot\cup} V_2, E)$ with $|V_1| = |V_2| = n$. Let $W_1, \dots, W_k$ be a balanced partition of V_1. Now for each *odd* $i \in [k]$ we randomly and independently choose a subset $U_i \subseteq V_2$ of size $n/2$; and we set $U_{i+1} := V_2 \setminus U_i$. Then we insert exactly all those edges into E which have one vertex in W_i and the other in U_i, for $i \in [k]$. Clearly, every vertex in G has degree $n/2$. In addition, using the degree co-degree characterisation of ε-regularity it is not difficult to check that (V_1, V_2) almost surely is ε-regular.

Next, we construct the tree T as follows. We start with a tree T', which consists of a root of degree $\sqrt{n}-1$ and is such that each child of this root has exactly $\sqrt{n}$ leaves as children. For obtaining T, we then take two copies of T', call their roots x_1 and x_2, respectively, and add an edge between x_1 and x_2. Clearly, the two colour classes of T have size n and $\Delta(T) = \sqrt{n}+1$.

It remains to show that $T \not\subseteq G_k$. Assume for contradiction that there is an embedding φ of T into G_k such that $\varphi(x_1) \in W_1$. Note that $n-1$ vertices in T have distance 2 from x_1. Since G_k is bipartite φ has to map these $n-1$ vertices to V_1. In particular, one of them has to be embedded in W_2. However, the distance between W_1 and W_2 in G_k is greater than 2. □

The proof of Proposition 4.36 proceeds similarly.

Proof of Proposition 4.36 (sketch). Let ε, n_0 and w be given, choose n large enough so that $n_0 \le n$ and $1/w(n) \ll \varepsilon$, and set $k := w(n)$.

We reuse the graph $G_k = (V_1 \mathbin{\dot\cup} V_2, E)$ from the previous proof as ε-regular pair. Now consider any balanced tree T with a vertex x of degree $\Delta(T) = w(n) = k$. Let $\{y_1, \dots, y_k\}$ be the neighbours of x in T. For $i \in [k]$ we then restrict the image of y_i to $V_2 \setminus U_i$.

We claim that there is no embedding of T into G that respects these image restrictions. Indeed, clearly x has to be embedded into $W_j \subseteq V_1$ for some $j \in [k]$ because its neighbours are image restricted to subsets of V_2. However, by the definition of U_j this prevents y_j from being embedded into $V_2 \setminus U_j$. □

4.6 Applications

4.6.1 F-factors for growing degrees

This section is to prove Theorem 4.7. Our strategy will be to repeatedly embed a collection of copies of F into a super-regular r-tuple in G with the help of the Blow-up

Lemma version stated as Theorem 4.4. The following result by Böttcher, Schacht, and Taraz [18, Lemma 6] says that for $\gamma > 0$ any sufficiently large graph G with $\delta(G) \geq ((r-1)/r+\gamma)|G|$ has a regular partition with a reduced graph R that contains a K_r-factor. Moreover, all pairs of vertices in R that lie in the same K_r span super-regular pairs in G. Let $K_r^{(k)}$ denote the disjoint union of k complete graphs on r vertices each. For all $n, k, r \in \mathbb{N}$, we call an integer partition $(n_{i,j})_{i\in[k],j\in[r]}$ of $[n]$ (with $n_{i,j} \in \mathbb{N}$ for all $i \in [k]$ and $j \in [r]$) r-equitable, if $|n_{i,j} - n_{i,j'}| \leq 1$ for all $i \in [k]$ and $j, j' \in [r]$.

Lemma 4.37
For all $r \in \mathbb{N}$ and $\gamma > 0$ there exists $\delta > 0$ and $\varepsilon_0 > 0$ such that for every positive $\varepsilon \leq \varepsilon_0$ there exists K_0 and $\xi_0 > 0$ such that for all $n \geq K_0$ and for every graph G on vertex set $[n]$ with $\delta(G) \geq ((r-1)/r+\gamma)n$ there exists $k \in \mathbb{N}\setminus\{0\}$, and a graph $K_r^{(k)}$ on vertex set $[k]\times[r]$ with

(R1) $k \leq K_0$,

(R2) there is an r-equitable integer partition $(m_{i,j})_{i\in[k],j\in[r]}$ of $[n]$ with $(1+\varepsilon)n/(kr) \geq m_{i,j} \geq (1-\varepsilon)n/(kr)$ such that the following holds.[2]

For every partition $(n_{i,j})_{i\in[k],j\in[r]}$ of n with $m_{i,j} - \xi_0 n \leq n_{i,j} \leq m_{i,j} + \xi_0 n$ there exists a partition $(V_{i,j})_{i\in[k],j\in[r]}$ of V with

(V1) $|V_{i,j}| = n_{i,j}$,

(V2) $(V_{i,j})_{i\in[k],j\in[r]}$ *is (ε,δ)-super-regular on $K_r^{(k)}$.*

Using this partitioning result for G, Theorem 4.7 follows easily.

Proof of Theorem 4.7. We alternatingly choose constants as given by Theorem 4.4 and Lemma 4.37. So let $\delta, \varepsilon_0 > 0$ be the constants given by Lemma 4.37 for r and $\gamma > 0$. Further let $\varepsilon, \alpha > 0$ be the constants given by Theorem 4.4 for $C = 0$, a, $\Delta_R = r$, $\kappa = 2$, $c = 1$ and δ. We are setting $C = 0$ as we do not use any image restrictions in this proof. If necessary we decrease ε such that $\varepsilon \leq \varepsilon_0$ holds. Let K_0 and $\xi_0 > 0$ be as in Lemma 4.37 with ε as set before. For r let n_0 be given by Theorem 4.4. If necessary increase n_0 such that $n_0 \geq K_0$. Finally set $\xi = \xi_0$. In the following we assume that

(i) G is of order $n \geq n_0$ and has $\delta(G) \geq (\frac{r-1}{r}+\gamma)n$, and

(ii) H is an a-arrangeable, r-chromatic F-factor with $|F| \leq \xi n$, $\Delta(F) \leq \sqrt{n}/\log n$.

We need to show that $H \subseteq G$. For this purpose we partition H into subgraphs $H_1, \dots, H_k$, where H_i is to be embedded into $\cup_{j\in[r]} V_{i,j}$ later, as follows. Let $(m_{i,j})_{i\in[k],j\in[r]}$ be an r-equitable partition of $[n]$ with $m_{i,j} \geq (1-\varepsilon)n/(kr)$ as given by Lemma 4.37. For $i = 1, \dots, k-1$ we choose ℓ_i such that both

$$\big|m_{i,j} - \ell_i|F|\big| \leq |F|, \qquad \text{and} \qquad \left|\sum_{i'\leq i}(m_{i',j} - \ell_{i'}|F|)\right| \leq |F| \tag{4.39}$$

[2]The upper bound on $m_{i,j}$ is implicit in the proof of the lemma but not explicitly stated in [18].

for all $j \in [r]$. Let $n_{i,j} = \ell_i|F|$ for all $j \in [r]$ and set H_i to be $\ell_i r$ copies of F. Note that there exists an r-colouring of H_i in which each colour class $X_{i,j}$ has exactly $n_{i,j}$ vertices. Finally H_k is set to be $H \setminus (H_1 \cup \cdots \cup H_{k-1})$. Let $\chi : V(H_k) \to [r]$ be a colouring of H_k where the colour-classes have as equal sizes as possible and set $n_{k,j} := |\chi^{-1}(j)|$ and $X_{k,j} := \chi^{-1}(j)$ for $j \in [r]$. It follows from (4.39) that

$$|n_{i,j} - m_{i,j}| \leq |F| \leq \xi_0 n \tag{4.40}$$

for all $i \in [k]$, $j \in [r]$. Thus there exists a partition $(V_{i,j})_{i\in[k],j\in[r]}$ of $V(G)$ with properties (V1) and (V2) by Lemma 4.37.

We apply Theorem 4.4 to embed H_i into $G[V_{i,1} \cup \cdots \cup V_{i,r}]$ for every $i \in [k]$. Note that we have partitioned $V(H_i) = X_{i,1} \cup \cdots \cup X_{i,r}$ in such a way that $|X_{i,j}| = n_{i,j}$ and $vw \in E(H_i)$ implies $v \in X_{i,j}$, $w \in X_{i,j'}$ with $j \neq j'$. Now properties (V1), (V2) guarantee that $|V_{i,j}| = n_{i,j}$ and $(V_{i,j}, V_{i,j'})$ is an (ε, δ)-super-regular pair in G for all $i \in [k]$ and $j, j' \in [r]$ with $j \neq j'$. It follows that H_i is a subgraph of $G[V_{i,1} \cup \cdots \cup V_{i,r}]$ by Theorem 4.4. □

4.6.2 Random graphs and universality

Next we prove Theorem 4.8, which states that $G = \mathcal{G}(n, p)$ is universal for the class of a-arrangeable bounded degree graphs,

$$\mathcal{H}_{n,a,\xi} = \{H : |H| = n, H \ a\text{-arr.}, \Delta(H) \leq \xi n/\log n\}.$$

To prove this we will find a balanced partition of G and apply Theorem 4.5. For this purpose we also have to find a balanced partition of the graphs $H \in \mathcal{H}_{n,a,\xi}$. To this end we shall use the following result of Kostochka, Nakprasit, and Pemmaraju [70].

Theorem 4.38 (Theorem 4 from [70])
Every a-arrangeable[3] graph H with $\Delta(H) \leq n/15$ has a balanced k-colouring for each $k \geq 16a$.

A graph has a *balanced k-colouring* if the graph has a proper colouring with at most k colours such that the sizes of the colour classes differ by at most 1.

Proof of Theorem 4.8. Let a and p be given. Set $\Delta_R := 16a$, $\kappa = 1$, $\iota := \frac{1}{2}p^{a+1}$, $\delta := p/2$ and let R be a complete graph on $16a$ vertices. Set $r := 16a$ and let ε, ξ, n_0 as given by Theorem 4.5. Let $n \geq n_0$ and let $V = V_1 \dot\cup \ldots \dot\cup V_r$ be a balanced partition of $[n]$. Then we generate a random graph $G = \mathcal{G}(n, p)$ on vertex set $[n]$. Every pair (V_i, V_j) is $(\varepsilon, p/2)$-super-regular in G with high probability. Furthermore with high probability we have that every tuple $(u_1, \ldots, u_{a+1}) \subseteq V \setminus V_i$ satisfies $|\bigcap_{j\in[a+1]} N_G(u_j) \cap V_i| \geq \iota|V_i|$. So assume this is the case and let $H \in \mathcal{H}_{n,a,\xi}$. We partition H into $16a$ equally sized stable sets with the help of Theorem 4.38. Thus H satisfies the requirements of Theorem 4.5 and H embeds into G. □

[3] In fact [70] shows this result for the more general class of a-degenerate graphs.

4.7 Appendix

4.7.1 Weighted regularity

In this section we provide some background on *weighted regularity*. In particular, we supplement the proofs of Lemma 4.17 and Lemma 4.18. We start with a short introduction to the results on weighted regularity by Czygrinow and Rödl [32]. Their focus lies on hypergraphs. However, we only present the graph case here.

Czygrinow and Rödl define their weight function on the set of edges (whereas in our scenario we have a bipartite graph with weights on the vertices of one class). They consider weighted graphs $G = (V, \widetilde{\omega})$ where $\widetilde{\omega} : V \times V \to \mathbb{N}_{\geq 0}$. One can think of $\widetilde{\omega}(x, y)$ as the multiplicity of the edge (x, y). Their weighted degree and co-degree for $x, y \in V$ are then defined as

$$\deg^*_{\widetilde{\omega}}(x) := \sum_{y \in V} \widetilde{\omega}(x, y), \qquad \deg^*_{\widetilde{\omega}}(x, y) := \sum_{z \in V} \widetilde{\omega}(x, z)\widetilde{\omega}(y, z)\,.$$

For disjoint $A, B \subseteq V$ they define

$$d^*_{\widetilde{\omega}}(A, B) = \frac{\sum \widetilde{\omega}(x, y)}{K|A|\,|B|}\,,$$

where the sum is over all pairs $(x, y) \in A \times B$ and $K := 1 + \max\{\widetilde{\omega}(x, y) : (x, y) \in V \times V\}$.[4] A pair (A, B) in $G = (V, \widetilde{\omega})$ with $A \cap B = \emptyset$ is called $(\varepsilon, \widetilde{\omega})$-*regular* if

$$|d^*_{\widetilde{\omega}}(A, B) - d^*_{\widetilde{\omega}}(A', B')| < \varepsilon$$

for all $A' \subseteq A$ with $|A'| \geq \varepsilon|A|$ and all $B' \subseteq B$ with $|B'| \geq \varepsilon|B|$. As in the unweighted case, regular pairs can be characterised by the degree and co-degree distribution of their vertices. The following lemma (see [32, Lemma 4.2]) shows that a pair is weighted regular in the setting of Czygrinow and Rödl if most of the vertices have the correct weighted degree and most of the pairs have the correct weighted co-degree.

Lemma 4.39 (Czygrinow, Rödl [32])
Let $G = (A \,\dot\cup\, B, \widetilde{\omega})$ be a weighted graph with $|A| = |B| = n$ and let $\varepsilon, \xi \in (0, 1)$, $\xi^2 < \varepsilon$, $n \geq 1/\xi$. Assume that both of the following conditions are satisfied:

(i') $|\{x \in A : |\deg^*_{\widetilde{\omega}}(x) - K\, d^*_{\widetilde{\omega}}(A, B)n| > K\xi^2 n\}| < \xi^2 n$, *and*

(ii') $\left|\left\{\{x_i, x_j\} \in \binom{A}{2} : |\deg^*_{\widetilde{\omega}}(x_i, x_j) - K^2 d^*_{\widetilde{\omega}}(A, B)^2 n| \geq K^2\xi n\right\}\right| \leq \xi\binom{n}{2}$.

Then for every $A' \subseteq A$ with $|A'| \geq \varepsilon n$ and every $B' \subseteq B$ with $|B'| \geq \varepsilon n$ we have

$$|d^*_{\widetilde{\omega}}(A', B') - d^*_{\widetilde{\omega}}(A, B)| \leq 2\frac{\xi^2}{\varepsilon} + \frac{\sqrt{5\xi}}{\varepsilon^2 - \varepsilon\xi^2}\,.$$

[4] Czygrinow and Rödl require K to be strictly larger than the maximal weight for technical reasons.

The assertion of Lemma 4.39 implies that the pair (A, B) is $(\varepsilon', \widetilde{\omega})$-regular, where $\varepsilon' = \max\{\varepsilon, 2\xi^2/\varepsilon + \sqrt{5\xi}/(\varepsilon^2 - \varepsilon\xi^2)\}$, if the conditions of the lemma are satisfied.

Our goal is to translate this result into our setting of weighted regularity (see Section 4.2.2). We shortly recall our definition of weighted graphs and weighted regularity before we restate and prove Lemma 4.17.

Let $G = (A \cup B, E)$ be a bipartite graph and $\omega : A \to [0, 1]$ be our weight function for G. We define the weighted degree of a vertex $x \in A$ to be $\deg_\omega(x) = \omega(x)|N(x, B)|$ and the weighted co-degree of $x, y \in A$ as $\deg_\omega(x, y) = \omega(x)\omega(y)|N(x, B) \cap N(y, B)|$. Similarly, the weighted density of a pair (A', B') is defined as

$$d_\omega(A', B') := \sum_{x \in A'} \frac{\omega(x)|N(x, B')|}{|A'| \cdot |B'|} .$$

Again the pair (A, B) is called weighted ε-regular if

$$|d_\omega(A, B) - d_\omega(A', B')| \leq \varepsilon$$

for all $A' \subseteq A$ and $B' \subseteq B$ with $|A'| \geq \varepsilon|A|$ and $|B'| \geq \varepsilon|B|$. We now prove Lemma 4.17, which we restate here for the reader's convenience.

Lemma (Lemma 4.17)
Let $\varepsilon > 0$ and $n \geq \varepsilon^{-6}$. Further let $G = (A \cup B, E)$ be a bipartite graph with $|A| = |B| = n$ and let $\omega : A \to [\varepsilon, 1]$ be a weight function for G. If

(i) $|\{x \in A : |\deg_\omega(x) - d_\omega(A, B)n| > \varepsilon^{14}n\}| < \varepsilon^{12}n$ *and*

(ii) $|\{\{x, y\} \in \binom{A}{2} : |\deg_\omega(x, y) - d_\omega(A, B)^2 n| \geq \varepsilon^9 n\}| \leq \varepsilon^6 \binom{n}{2}$

then (A, B) is a weighted 3ε-regular pair.

Proof of Lemma 4.17. Let $\varepsilon > 0$, $G = (A \cup B, E)$ and $\omega : A \to [\varepsilon, 1]$ satisfy the requirements of the lemma. From this ω we define a weight function $\widetilde{\omega} : A \times B \to \mathbb{N}_{\geq 0}$ in the setting of Lemma 4.39. For $(x, y) \in A \times B$ we set

$$\widetilde{\omega}(x, y) := \begin{cases} \lceil C \cdot \omega(x) \rceil & \text{if } \{x, y\} \in E, \\ 0 & \text{otherwise,} \end{cases}$$

where $\varepsilon^{-13} - 1 \leq C \leq \varepsilon^{-14}$ is chosen such that $K = \max\{\widetilde{\omega}(x, y) + 1 : (x, y) \in A \times B\} \geq \varepsilon^{-13}$. (This is possible unless $E = \emptyset$.) Note that our choice of constants implies $K/C \leq 1 + 2\varepsilon^{13}$. Moreover, let $d^*_{\widetilde{\omega}}(A, B)$ be defined as above. The definition of $\widetilde{\omega}$ implies

$$C \deg_\omega(x) \leq \deg^*_{\widetilde{\omega}}(x) \leq C \deg_\omega(x) + |N(x, B)| \text{ and} \tag{4.41}$$
$$C^2 \deg_\omega(x, y) \leq \deg^*_{\widetilde{\omega}}(x, y) \leq C^2 \deg_\omega(x, y) + (2C + 1)|N(x, B) \cap N(y, B)| \tag{4.42}$$

for all $x, y \in A$. Here the second inequality follows from

$$\lceil C \cdot \omega(x) \rceil^2 \leq (C \cdot \omega(x) + 1)^2 \leq C^2 \big(\omega(x)\big)^2 + 2C + 1 .$$

Moreover,

$$C d_\omega(A', B') \le K d^*_{\widetilde{\omega}}(A', B') \le C d_\omega(A', B') + 1 \tag{4.43}$$

for all $A' \subseteq A$, $B' \subseteq B$ which in turn implies that

$$\big(C\, d_\omega(A', B')\big)^2 - \big(K\, d^*_{\widetilde{\omega}}(A', B')\big)^2 \le 1 \cdot (C + K) \tag{4.44}$$

for all $A' \subseteq A$, $B' \subseteq B$.

We now verify that conditions (i) and (ii) of Lemma 4.17 imply conditions (i') and (ii') of Lemma 4.39. Set $\xi := \varepsilon^6$ and let $x \in A$ be such that $|\deg_\omega(x) - d_\omega(A, B)n| \le \varepsilon^{14} n$. It follows from (4.41), (4.43) and the triangle inequality that

$$\begin{aligned}
|\deg^*_{\widetilde{\omega}}(x) - K\, d^*_{\widetilde{\omega}}(A, B)n| &\le |\deg^*_{\widetilde{\omega}}(x) - C \deg_\omega(x)| \\
&\quad + |C \deg_\omega(x) - C\, d_\omega(A, B)n| \\
&\quad + |C\, d_\omega(A, B)n - K\, d^*_{\widetilde{\omega}}(A, B)n| \\
&\le n + C\varepsilon^{14} n + n \le 3n \\
&\le K\xi^2 n\,.
\end{aligned}$$

Hence, condition (i) implies condition (i').

Now let $\{x, y\} \in \binom{A}{2}$ satisfy $|\deg_\omega(x, y) - d_\omega(A, B)^2 n| < \varepsilon^9 n$. It follows from (4.42), (4.43) and (4.44) that

$$\begin{aligned}
\left|\deg^*_{\widetilde{\omega}}(x, y) - K^2 d^*_{\widetilde{\omega}}(A, B)^2 n\right| &\le |\deg^*_{\widetilde{\omega}}(x, y) - C^2 \deg_\omega(x, y)| \\
&\quad + |C^2 \deg_\omega(x, y) - C^2 d_\omega(A, B)^2 n| \\
&\quad + \left|C^2 d_\omega(A, B)^2 n - K^2 d^*_{\widetilde{\omega}}(A, B)^2 n\right| \\
&< (2C + 1)n + C^2 \varepsilon^9 n + (C + K)n \\
&\le K^2 \xi n\,,
\end{aligned}$$

where the last inequality is due to $C \le K/\varepsilon$. Thus, condition (ii) implies condition (ii').

We conclude that $G = (A \,\dot\cup\, B, \widetilde{\omega})$ satisfies the requirements of Lemma 4.39. Hence every $A' \subseteq A$ with $|A'| \ge \varepsilon n$ and every $B' \subseteq B$ with $|B'| \ge \varepsilon n$ has

$$|d^*_{\widetilde{\omega}}(A', B') - d^*_{\widetilde{\omega}}(A, B)| \le 2\frac{\xi^2}{\varepsilon} + \frac{\sqrt{5\xi}}{\varepsilon^2 - \varepsilon\xi^2} \le \frac{5}{2}\varepsilon\,.$$

Together with (4.43) and the fact that $K/C \le 1 + 2\varepsilon^{13}$ this finishes the proof as we have

$$\begin{aligned}
|d_\omega(A', B') - d_\omega(A, B)| &\le |d_\omega(A', B') - \tfrac{K}{C} d^*_{\widetilde{\omega}}(A', B')| \\
&\quad + |\tfrac{K}{C} d^*_{\widetilde{\omega}}(A', B') - \tfrac{K}{C} d^*_{\widetilde{\omega}}(A, B)| \\
&\quad + |\tfrac{K}{C} d^*_{\widetilde{\omega}}(A, B) - d_\omega(A, B)| \\
&\le \tfrac{1}{C} + \tfrac{K}{C}\tfrac{5}{2}\varepsilon + \tfrac{1}{C} \\
&\le 3\varepsilon\,.
\end{aligned}$$

□

We want to point out that the requirement that ω is at least ε does not cause any problem when we apply Lemma 4.17 because one could simply increase the weight of all vertices x with $\omega(x) < \varepsilon$ to ε without changing the weighted densities in the subpairs by more than ε. Hence a graph with an arbitrary weight function is weighted 2ε-regular if the graph with the modified weight function is weighted ε-regular.

The remainder of this section is dedicated to the proof of Lemma 4.18 which we restate here.

Lemma (Lemma 4.18)
Let $\varepsilon > 0$ and let $G = (A \,\dot\cup\, B, E)$ with $|V_i| = n$ and weight function $\omega : A \to [\sqrt{\varepsilon}, 1]$ be a weighted ε-regular pair. If $\deg(x) > 2\sqrt{\varepsilon}n$ for all $x \in A \cup B$ then G contains a perfect matching.

Proof of Lemma 4.18. In order to prove that $G = (A \,\dot\cup\, B, E)$ has a perfect matching, we will verify the König–Hall criterion for G, i.e., we will show that $|N(S)| \geq |S|$ for every $S \subseteq A$. We distinguish three cases.

Case 1, $|S| < \varepsilon n$: The minimum degree $\deg(x) \geq 2\sqrt{\varepsilon}n$ implies $|N(S)| \geq 2\sqrt{\varepsilon}n \geq |S|$ for any non-empty set S.

Case 2, $\varepsilon n \leq |S| \leq (1-\varepsilon)n$: Note that $\deg(x) > 2\sqrt{\varepsilon}n$ and $\omega(x) \geq \sqrt{\varepsilon}$ for all $x \in A$ implies that $d_\omega(A, B) > 2\varepsilon$. We now set $T = B \setminus N(S)$. Since $d_\omega(S, T) = 0$ and (A, B) is a weighted-ε-regular pair with weighted density greater than 2ε we conclude that $|T| < \varepsilon n$.

Case 3, $|S| > (1-\varepsilon)n$: For every $y \in B$ we have $|S| + |N(y)| \geq (1 - \varepsilon + 2\sqrt{\varepsilon})n > n$ and thus $N(y) \cap S \neq \emptyset$. It follows that $N(S) = B$ if $|S| > (1-\varepsilon)n$. □

4.7.2 Chernoff type bounds

The analysis of our randomised greedy embedding (see Section 4.3.2) repeatedly uses concentration results for random variables. Those random variables are the sum of Bernoulli variables. If these are mutually independent we use a Chernoff bound (see, e.g., [54, Corollary 2.3]).

Theorem 4.40 (Chernoff bound)
Let $\mathcal{A} = \sum_{i\in[n]} \mathcal{A}_i$ be a binomially distributed random variable with $\mathbb{P}[\mathcal{A}_i] = p$ for all $i \in [n]$. Further let $c \in [0, 3/2]$. Then

$$\mathbb{P}[|\mathcal{A} - pn| \geq c \cdot pn] \leq \exp\left(-\frac{c^2}{3}pn\right).$$

However, we also consider scenarios where the Bernoulli variables are not independent.

Lemma (Lemma 4.19)
Let $0 \leq p_1 \leq p_2 \leq 1$, $0 < c \leq 1$. Further let $\mathcal{A}_i$ for $i \in [n]$ be a 0-1-random variable and set $\mathcal{A} := \sum_{i\in[n]} \mathcal{A}_i$. If

$$p_1 \leq \mathbb{P}\left[\mathcal{A}_i = 1 \,\middle|\, \begin{array}{l} \mathcal{A}_j = 1 \text{ for all } j \in J \text{ and} \\ \mathcal{A}_j = 0 \text{ for all } j \in [i-1] \setminus J \end{array}\right] \leq p_2 \tag{4.45}$$

for every $i \in [n]$ *and every* $J \subseteq [i-1]$ *then*

$$\mathbb{P}[\mathcal{A} \leq (1-c)p_1 n] \leq \exp\left(-\frac{c^2}{3}p_1 n\right)$$

and

$$\mathbb{P}[\mathcal{A} \geq (1+c)p_2 n] \leq \exp\left(-\frac{c^2}{3}p_2 n\right).$$

The somewhat technical conditioning in (4.45) allows us to bound the probability for the event $\mathcal{A}_i = 1$ even if we condition on any outcome of the events $\mathcal{A}_j$ with $j < i$.

The idea of the proof now is to relate the random variable $\mathcal{A}$ to a truly binomially distributed random variable and then use a Chernoff bound.

Proof of Lemma 4.19. For $k, \ell \in \mathbb{N}_0$ define $a_{\ell,k} = \mathbb{P}[\sum_{i\leq\ell}\mathcal{A}_i \leq k]$ and $b_{\ell,k} = \mathbb{P}[B_{\ell,p_1} \leq k]$ where B_{ℓ,p_1} is a binomially distributed random variable with parameters ℓ and p_1. So both $a_{\ell,k}$ and $b_{\ell,k}$ give a probability that a random variable (depending on ℓ and p_1) is below a certain value k. The following claim relates these two probabilities.

Claim *For every* $k \geq 0$, $\ell \geq 0$ *we have* $a_{\ell,k} \leq b_{\ell,k}$.

Proof. We will prove the claim by induction on ℓ. For $\ell = 0$ we trivially have $a_{0,k} = 1 = b_{0,k}$ for all $k \geq 0$. Now assume that the claim is true for $\ell - 1$ and every $k \geq 0$. Now

$$a_{\ell,0} \leq (1-p_1)a_{\ell-1,0} \leq (1-p_1)b_{\ell-1,0} = b_{\ell,0}\,.$$

As

$$\mathbb{P}\left[\mathcal{A}_\ell = 1 \;\middle|\; \begin{array}{l}\mathcal{A}_j = 1 \text{ for all } j \in J \text{ and}\\ \mathcal{A}_j = 0 \text{ for all } j \in [\ell-1]\setminus J\end{array}\right] \geq p_1$$

for every $J \subseteq [\ell-1]$ it follows that for $k \geq 1$

$$a_{\ell,k} \leq a_{\ell-1,k-1} + (a_{\ell-1,k} - a_{\ell-1,k-1})(1-p_1)\,. \tag{4.46}$$

This upper bound on $a_{\ell,k}$ implies that for every $k \geq 1$ we have

$$\begin{aligned}
a_{\ell,k} &\overset{(4.46)}{\leq} p_1 \cdot a_{\ell-1,k-1} + (1-p_1)\cdot a_{\ell-1,k}\\
&\leq p_1 \cdot b_{\ell-1,k-1} + (1-p_1)\cdot b_{\ell-1,k}\\
&= p_1\,\mathbb{P}[B_{\ell-1,p_1} \leq k-1] + (1-p_1)\,\mathbb{P}[B_{\ell-1,p_1} \leq k]\\
&= \mathbb{P}[B_{\ell-1,p_1} \leq k-1] + (1-p_1)\,\mathbb{P}[B_{\ell-1,p_1} = k]\\
&= \mathbb{P}[B_{\ell,p_1} \leq k] = b_{\ell,k}\,.
\end{aligned}$$

Here the second inequality is due to the induction hypothesis. This finishes the induction step and the proof of the claim. □

Now the first inequality of the lemma follows immediately. We set $\ell = n$, $k = (1-c)p_1 n$ and obtain

$$\mathbb{P}[\mathcal{A} \leq (1-c)p_1 n] = a_{n,(1-c)p_1 n} \leq b_{n,(1-c)p_1 n} = \mathbb{P}[B_{n,p_1} \leq (1-c)p_1 n] \leq \exp\left(-\frac{c^2}{3}p_1 n\right),$$

where the last inequality follows by Theorem 4.40.

The second assertion of the lemma follows by an analogous argument: set $a_{\ell,k} = \mathbb{P}[\sum_{i\le\ell}\mathcal{A}_i \ge k]$ and $b_{\ell,k} = \mathbb{P}[B_{\ell,p_2} \ge k]$ and obtain

$$a_{\ell,k} \le a_{\ell-1,k} + (a_{\ell-1,k-1} - a_{\ell-1,k})p_2\,. \tag{4.47}$$

It follows by induction on ℓ that

$$\begin{aligned} a_{\ell,k} &\overset{(4.47)}{\le} (1-p_2)\cdot a_{\ell-1,k} + p_2 \cdot a_{\ell-1,k-1} \\ &\le (1-p_2)\cdot b_{\ell-1,k} + p_2\cdot b_{\ell-1,k-1} \\ &= \mathbb{P}[B_{\ell-1,p_2} \ge k] + \mathbb{P}[B_{\ell-1,p_2} = k-1]p_2 \\ &= \mathbb{P}[B_{\ell,p_2} \ge k] = b_{\ell,k}\,. \end{aligned}$$

Once more the second inequality follows from the induction hypothesis. Setting $\ell = n$ and $k = (1+c)p_2 n$ and using Theorem 4.40 again then finishes the proof. □

In addition we need a similar result with a more complex setup.

Lemma (Lemma 4.20)
Let $0 < p$ and $a, m, n \in \mathbb{N}$. Further let $\mathcal{I} \subseteq \mathcal{P}([n]) \setminus \{\emptyset\}$ be a collection of m disjoint sets with at most a elements each. For every $i \in [n]$ let $\mathcal{A}_i$ be a 0-1-random variable. Further assume that for every $I \in \mathcal{I}$ and every $k \in I$ we have

$$\mathbb{P}\left[\mathcal{A}_k = 1 \;\middle|\; \begin{array}{l} \mathcal{A}_j = 1 \text{ for all } j \in J \text{ and} \\ \mathcal{A}_j = 0 \text{ for all } j \in [k-1]\setminus J \end{array}\right] \ge p$$

for every $J \subseteq [k-1]$ with $[k-1] \cap I \subseteq J$. Then

$$\mathbb{P}\Big[\big|\{I \in \mathcal{I} : \mathcal{A}_i = 1 \text{ for all } i \in I\}\big| \ge \tfrac{1}{2}p^a m\Big] \ge 1 - 2\exp\Big(-\frac{1}{12}p^a m\Big)\,.$$

Proof of Lemma 4.20. Let $p > 0$, $a, m, n \in \mathbb{N}$ and $\mathcal{I}$ be given. We order the elements of $\mathcal{I}$ as $\mathcal{I} = \{I_1, \ldots, I_m\}$ by their respective largest index. This means, the I_j are sorted such that $j' < j$ implies that there is an index $i_j \in I_j$ with $i < i_j$ for all $i \in I_{j'}$. For $i \in [m]$ we now define events $\mathcal{B}_i$ as

$$\mathcal{B}_i := \begin{cases} 1 & \text{if } \mathcal{A}_j = 1 \text{ for all } j \in I_i, \\ 0 & \text{otherwise.} \end{cases}$$

We claim that the events $\mathcal{B}_i$ satisfy equation (4.45) where the probability is bounded from below by p^a.

Claim *For every $i \in [m]$ and $J \subseteq [i-1]$ we have*

$$\mathbb{P}\left[\mathcal{B}_i = 1 \;\middle|\; \begin{array}{l} \mathcal{B}_j = 1 \text{ for all } j \in J \text{ and} \\ \mathcal{B}_j = 0 \text{ for all } j \in [i-1]\setminus J \end{array}\right] \ge p^a\,.$$

Proof. Let $i \in [m]$ and $J \subseteq [i-1]$ be given. We assume that $|I| = a$ for ease of notation. (The proof is just the same if $|I| < a$.) So let $I_i = \{i_1, \dots, i_a\}$ be in ascending order and define $i_0 := 0$. For $v \in \{0,1\}^{i_k - i_{k-1} - 1}$ let $H_k(v)$ be the 0-1-random variable with

$$H_k(v) = \begin{cases} 1 & \mathcal{A}_{i_{k-1}+\ell} = v_\ell \text{ for all } \ell \in [i_k - i_{k-1} - 1], \\ 0 & \text{otherwise.} \end{cases}$$

The rationale for this definition is the following. The outcome of $\mathcal{B}_i$ is determined by the outcome of the random variables $\mathcal{A}_{i_j}$. However, we cannot neglect the random variables $\mathcal{A}_\ell$ for $\ell \notin I_i$ as the $\mathcal{A}_\ell$ are not mutually independent. Instead we condition the probability of $\mathcal{A}_{i_j} = 1$ on possible outcomes of $\mathcal{A}_\ell$ with $\ell < i_j$. Now $H_k(v) = 1$ with $v \in \{0,1\}^{i_k - i_{k-1} - 1}$ represents one outcome for the $\mathcal{A}_\ell$ with $i_{k-1} < \ell < i_k$. We call the $v \in \{0,1\}^{i_k - i_{k-1} - 1}$ the *history* between $\mathcal{A}_{i_{k-1}}$ and $\mathcal{A}_{i_k}$. It follows from the requirements of Lemma 4.20 that for any tuple $(v_1, \dots, v_k) \in \{0,1\}^{i_1 - 1} \times \{0,1\}^{i_2 - i_1 - 1} \times \dots \times \{0,1\}^{i_k - i_{k-1} - 1}$ we have

$$\mathbb{P}\left[\mathcal{A}_{i_k} = 1 \,\middle|\, \begin{matrix} \mathcal{A}_{i_j} = 1 \text{ for all } j \in [k-1] \text{ and} \\ H_j(v_j) = 1 \text{ for all } j \in [k-1] \end{matrix}\right] \geq p\,. \tag{4.48}$$

However, we are not interested in every possible history $(v_1, \dots, v_k)$ as some of the histories cannot occur simultaneously with the event $\mathcal{B} = 1$ where

$$\mathcal{B} = 1 \text{ if and only if } \begin{bmatrix} \mathcal{B}_j = 1 \text{ for all } j \in J \text{ and} \\ \mathcal{B}_j = 0 \text{ for all } j \in [i-1] \setminus J \end{bmatrix}.$$

For ease of notation we define the following shortcuts

$$\mathcal{H}(v_1, \dots, v_k) = 1 \text{ if and only if } H_j(v_j) = 1 \text{ for all } j \in [k],$$
$$\mathcal{A}^{(k)} = 1 \text{ if and only if } \mathcal{A}_{i_j} = 1 \text{ for all } j \in [k].$$

Moreover, we define C_k to be the set of all tuples $(v_1, \dots, v_k) \in \{0,1\}^{i_1-1} \times \{0,1\}^{i_2 - i_1 - 1} \times \dots \times \{0,1\}^{i_k - i_{k-1} - 1}$ with

$$\mathbb{P}\left[\mathcal{H}(v_1, \dots, v_k) = 1 \text{ and } \mathcal{B} = 1\right] > 0\,.$$

In other words, the elements of C_k are those histories that are compatible with the event that we condition on in the claim. Note in particular that $\mathcal{B} = 1$ if and only if there is a $(v_1, \dots, v_a) \in C_a$ with $\mathcal{H}(v_1, \dots, v_a) = 1$. With these definitions we can rewrite the probability in the assertion of our claim as

$$\mathbb{P}\left[\mathcal{B}_i = 1 \mid \mathcal{B} = 1\right] = \sum_{(v_1, \dots, v_a) \in C_a} \mathbb{P}\left[\begin{matrix} \mathcal{A}^{(a)} = 1 \text{ and} \\ \mathcal{H}(v_1, \dots, v_a) = 1 \end{matrix} \,\middle|\, \mathcal{B} = 1\right].$$

We now prove by induction on k that

$$P_k := \sum_{(v_1, \dots, v_k) \in C_k} \mathbb{P}\left[\begin{matrix} \mathcal{A}^{(k)} = 1 \text{ and} \\ \mathcal{H}(v_1, \dots, v_k) = 1 \end{matrix} \,\middle|\, \mathcal{B} = 1\right] \geq p^k \tag{4.49}$$

for all $k \in [a]$. The induction base $k = 1$ is immediate from the requirements of the lemma as

$$P_1 = \sum_{v_1 \in C_1} \mathbb{P}\left[\begin{matrix}\mathcal{A}^{(1)} = 1 \text{ and} \\ \mathcal{H}(v_1) = 1\end{matrix} \,\middle|\, \mathcal{B} = 1\right] \overset{(4.48)}{\geq} p \sum_{v_1 \in C_1} \mathbb{P}\left[\mathcal{H}(v_1) = 1 \,\middle|\, \mathcal{B} = 1\right] = p\,.$$

The last equality above follows by total probability from the definition of C_1. So assume that the induction hypothesis holds for $k-1$. Then

$$\begin{aligned}
P_k &= \sum_{(v_1,\ldots,v_k) \in C_k} \mathbb{P}\left[\begin{matrix}\mathcal{A}^{(k)} = 1 \text{ and} \\ \mathcal{H}(v_1,\ldots,v_k) = 1\end{matrix} \,\middle|\, \mathcal{B} = 1\right] \\
&= \sum_{(v_1,\ldots,v_k) \in C_k} \mathbb{P}\left[\mathcal{A}_{i_k} = 1 \,\middle|\, \begin{matrix}\mathcal{B} = 1 \text{ and } \mathcal{A}^{(k-1)} = 1 \text{ and} \\ \mathcal{H}(v_1,\ldots,v_k) = 1\end{matrix}\right] \cdot \mathbb{P}\left[\begin{matrix}\mathcal{A}^{(k-1)} = 1 \text{ and} \\ \mathcal{H}(v_1,\ldots,v_k) = 1\end{matrix} \,\middle|\, \mathcal{B} = 1\right] \\
&\overset{(4.48)}{\geq} p \cdot \sum_{(v_1,\ldots,v_k) \in C_k} \mathbb{P}\left[\begin{matrix}\mathcal{A}^{(k-1)} = 1 \text{ and} \\ \mathcal{H}(v_1,\ldots,v_k) = 1\end{matrix} \,\middle|\, \mathcal{B} = 1\right] \\
&= p \cdot \sum_{(v_1,\ldots,v_{k-1}) \in C_{k-1}} \mathbb{P}\left[\begin{matrix}\mathcal{A}^{(k-1)} = 1 \text{ and} \\ \mathcal{H}(v_1,\ldots,v_{k-1}) = 1\end{matrix} \,\middle|\, \mathcal{B} = 1\right] \\
&= p \cdot P_{k-1} \geq p^k\,.
\end{aligned}$$

The claim now follows as

$$\mathbb{P}[\mathcal{B}_i = 1 \mid \mathcal{B} = 1] = \sum_{(v_1,\ldots,v_a) \in C_a} \mathbb{P}\left[\begin{matrix}\mathcal{A}^{(a)} = 1 \text{ and} \\ \mathcal{H}(v_1,\ldots,v_a) = 1\end{matrix} \,\middle|\, \mathcal{B} = 1\right] \geq p^a\,.$$

□

We have seen that the $\mathcal{B}_i$ are pseudo-independent and that they have probability at least p^a each. Thus we can apply Lemma 4.19 and derive

$$\mathbb{P}\left[|\{i \in [m] : \mathcal{B}_i = 1\}| \geq \frac{1}{2} p^a m\right] \geq 1 - 2\exp\left(-\frac{1}{12} p^a m\right)\,.$$

□

5 Spanning embeddings of arrangeable graphs with sublinear bandwidth

5.1 Introduction

The existence of spanning subgraphs in dense graphs has been investigated very successfully over the past decades. Its early stages can be traced back to results by Dirac [34] in 1952, who showed that a minimum degree of $n/2$ forces a Hamilton cycle in graphs of order n, and Corrádi and Hajnal [30] in 1963 as well as Hajnal and Szemerédi [49] in 1970, who proved that every graph G with $\delta(G) \geq \frac{r-1}{r}n$ must contain a family of $\lfloor n/r \rfloor$ vertex disjoint cliques, each of size r. The story gained new momentum when, in a series of papers in the 1990s, Komlós, Sarközy, and Szemerédi established a new methodology which, based on the Regularity Lemma and the Blow-up Lemma, paved the road to a series of results for spanning subgraphs with bounded maximum degree, such as powers of Hamilton cycles, trees, F-factors, and planar graphs (see the survey [72] for an excellent overview of these and related achievements).

During that period, Bollobás and Komlós [58] formulated a general conjecture which (approximately) included many of the results mentioned above. Böttcher, Schacht and Taraz proved this conjecture.

Theorem 5.1 (Böttcher, Schacht, Taraz [18])
For all $r, \Delta \in \mathbb{N}$ and $\gamma > 0$, there exist constants $\beta > 0$ and $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$ the following holds. If H is an r-chromatic graph on n vertices with $\Delta(H) \leq \Delta$ and bandwidth at most βn and if G is a graph on n vertices with minimum degree $\delta(G) \geq \left(\frac{r-1}{r} + \gamma\right) n$, then G contains a copy of H.

Here a graph H has *bandwidth* at most b if there exists a labelling of the vertices by numbers $1, \ldots, n$ such that for every edge $\{i, j\} \in E(H)$ we have $|i - j| \leq b$. It is well known that the restriction on the bandwidth in Theorem 5.1 cannot be omitted. On the other hand, powers of Hamilton cycles and F-factors have constant bandwidth. Moreover, bounded degree planar graphs and more generally any hereditary class of bounded degree graphs with small separators have bandwidth at most $O(n/\log n)$ (see [17]). Hence a rich class of graphs H is covered by Theorem 5.1.

However, a major constraint of this theorem is that it allows only H with constant maximum degree. In fact this is also true for most other results on spanning subgraphs mentioned above. There are only few exceptions, such as a result by Komlós, Sarközy, and Szemerédi [67], which shows that each sufficiently large graph with minimum degree at least $(\frac{1}{2}+\gamma)n$ contains all spanning trees of maximum degree $o(n/\log n)$.

One aim of this paper is to obtain a corresponding embedding result for a more general class of graphs with unbounded maximum degree. More precisely, we will generalise Theorem 5.1 to graphs with unbounded maximum degrees. We focus on arrangeable graphs.

Definition 5.2 (a-arrangeable)
Let a be an integer. A graph is called a-arrangeable if its vertices can be ordered as $(x_1,\dots,x_n)$ in such a way that $\big|N\big(N(x_i)\cap \mathrm{Right}_i)\big)\cap \mathrm{Left}_i\big)\big| \le a$ for each $1\le i\le n$, where $\mathrm{Left}_i=\{x_1,x_2,\dots,x_i\}$ and $\mathrm{Right}_i=\{x_{i+1},x_{i+2},\dots,x_n\}$.

Arrangeability was introduced by Chen and Schelp [22]. It generalises the concept of bounded maximum degree because graphs with maximum degree Δ are clearly $(\Delta^2-\Delta+1)$-arrangeable, and stars are 1-arrangeable. Moreover several important graph classes were shown to be constantly arrangeable: planar graphs are 10-arrangeable [57] (see also [22]) and graphs without a K_p-subdivision are p^8-arrangeable [87].

Our main result asserts that we can replace the constant maximum degree bound in Theorem 5.1 by a-arrangeability and $\Delta(H)\le\sqrt{n}/\log n$.

Theorem 5.3 (The bandwidth theorem for arrangeable graphs)
For all $r,a\in\mathbb{N}$ and $\gamma>0$, there exist constants $\beta>0$ and $n_0\in\mathbb{N}$ such that for every $n\ge n_0$ the following holds. If H is an r-chromatic, a-arrangeable graph on n vertices with $\Delta(H)\le\sqrt{n}/\log n$ and bandwidth at most βn and if G is a graph on n vertices with minimum degree $\delta(G)\ge\left(\frac{r-1}{r}+\gamma\right)n$, then G contains a copy of H.

The key ingredient for generalising Theorem 5.1 to Theorem 5.3 is a variant of the Blow-up Lemma for arrangeable graphs, obtained recently by Böttcher, Kohayakawa, Taraz, and Würfl in [16] (see Theorem 5.13).

Applications. We give one direct application of Theorem 5.3 (Corollary 5.4), and one application which uses the techniques needed in the proof of Theorem 5.3 (Theorem 5.6). Both applications concern graphs of fixed genus.

Let S be an orientable surface and denote by $g(S)$ the genus of S. Let $\mathcal{H}_S(n)$ be the family of n-vertex graphs embeddable on S and let $\mathcal{H}_S(n,\Delta)$ be the family of those graphs in $\mathcal{H}_S(n)$ with maximum degree at most Δ. The celebrated Four Colour Theorem [10, 85] and the affirmative solution of Heawood's Conjecture [51, 83] guarantee that each graph in $\mathcal{H}_S(n)$ can be coloured with

$$r(S) := \left\lfloor \frac{7+\sqrt{1+48g(S)}}{2} \right\rfloor \tag{5.1}$$

colours. Moreover, in [17] it was shown that graphs in $H\in\mathcal{H}_S(n,\Delta)$ have bandwidth at most

$$\mathrm{bw}(S,n,\Delta) := \frac{15n\log\Delta}{\log n-\log\min\big(1,g(S)\big)}\,. \tag{5.2}$$

Hence, as observed in [17], it is a direct consequence of Theorem 5.1 that large n-vertex graphs G with minimum degree at least $(\frac{r-1}{r}+\gamma)n$ contain all graphs from $\mathcal{H}_S(n,\Delta)$ as subgraphs, which extends results of Kühn, Osthus and Taraz [74] (see also [71]). With the help of Theorem 5.3 we are now able to say considerably more – namely, that in fact *almost all* graph from $\mathcal{H}_S(n)$ are contained in each such graph G.

Indeed, McDiarmid and Reed [77] proved that for each fixed S, if we draw a graph H uniformly at random from $\mathcal{H}_S(n)$ then asymptotically almost surely H has maximum degree of order

$$\Delta(S,n) := \Theta_S(\log n)\,. \tag{5.3}$$

Moreover, clearly $K_{r(S)+1}$ cannot be embedded in S and hence graphs from $\mathcal{H}_S(n)$ are $K_{r(S)+1}$-minor free. It thus follows from the result of Rödl and Thomas [87] mentioned above that the graphs in $\mathcal{H}_S(n)$ are $a(S)$-arrangeable with

$$a(S) := \big(r(S)+1\big)^8\,. \tag{5.4}$$

In conclusion, using (5.1), (5.2), (5.3) and (5.4) we immediately obtain the following corollary of Theorem 5.3.

Corollary 5.4
Let $\gamma>0$, let S be an orientable surface and let G be an n-vertex graph with $\delta(G)\ge \big(\frac{r(S)-1}{r(S)}+\gamma\big)n$. If H is drawn uniformly at random from $\mathcal{H}_S(n)$, then G contains H almost surely.

In particular, if $\delta(G)\ge(\frac{3}{4}+\gamma)n$ then G contains almost all planar graphs on n vertices.

Our second application concerns *Ramsey numbers* of graphs in $\mathcal{H}_S(n)$. For a graph H we denote by $R(H)$ the two-colour Ramsey number of H. Allen, Brightwell, and Skokan [2] proved that graphs with bounded maximum degree and small bandwidth have small Ramsey numbers.

Theorem 5.5 (Allen, Brightwell and Skokan [2])
For all $\Delta\in\mathbb{N}$, there exist constants $\beta>0$ and n_0 such that for every $n\ge n_0$ the following holds. If H is an n-vertex graph with maximum degree at most Δ and $\mathrm{bw}(H)\le\beta n$, then $R(H)\le(2\chi(H)+4)n$.

With the help of (5.1) and (5.2) this implies that for any fixed orientable surface S and any fixed Δ each graph $H\in\mathcal{H}_S(n,\Delta)$ satisfies $R(H)\le\big(2r(S)+4\big)n$ if n is sufficiently large. In particular, large planar graphs H with bounded maximum degree have Ramsey number $R(H)\le 12|H|$.

This together with the fact that planar graphs are known to have at most linear Ramsey number (see [22]) led Allen, Brightwell, and Skokan to conjecture that in fact *all* sufficiently large planar graphs H have Ramsey number at most $12|H|$. Combining their methods with ours we can now show that this is true for *almost all* planar graphs.

Theorem 5.6
Let S be an orientable surface. If H is drawn uniformly at random from $\mathcal{H}_S(n)$, then almost surely $R(H)\le(2r(S)+4)n$.

In particular, for almost every planar graph H we have $R(H)\le 12|H|$.

Organisation. In Section 5.2 we give an outline of our proof of Theorem 4.9. This proof builds on partitioning results for G and for H, which we present in Section 5.3, and on a variant of the Blow-up Lemma for arrangeable graphs, which we discuss in Section 5.4. We then present the actual proof of Theorem 4.9 in Section 5.5. We close with the proof of Theorem 5.6 in Section 5.6 and with some concluding remarks in Section 5.7.

5.2 Outline

Many of the results concerning the embedding of spanning, bounded degree graphs follow a general agenda which is nicely described in the survey paper [58] by Komlós. This agenda consists of five main steps: firstly preparing H, secondly preparing G, thirdly assigning parts of H to parts of G, fourthly connecting those parts, and fifthly embedding the parts of H separately, via the Blow-up Lemma.

In the proof of Theorem 4.9 we follow a similar agenda. The preparation for G uses, as is usual, Szemerédi's Regularity Lemma and some additional work to produce a suitable partition of G. For this step we can make use of a lemma from [18] (see Lemma 5.7).

The preparation of H (see Lemma 5.8) makes use of the bandwidth of H and produces a partition of H which is compatible to the partition of G (in this way we implicitly obtain an assignment of the parts of H to the parts of G). This step is also similar to the methods used in [18] to partition bounded degree graphs H. However, we need to strengthen this approach because we now deal with graphs H whose degrees are no longer bounded by a constant. In other words, we need a slightly different partitioning lemma for H in order to make this partition suitable for the Blow-up lemma that we will use in the next step.

In a final step we use the two partitions obtained to embed H into G. Our approach here is slightly different from the steps described by Komlós which are usually used (connecting the parts and embedding the parts of H separately). We use the Blow-up Lemma for arrangeable graphs, which was recently established in [16], to formulate an embedding result (see Theorem 5.14) which can handle super-regular and merely regular pairs simultaneously and make use of a spanning subgraph of the reduced graph of the partition for G. This enables us to embed H into G at once.

5.3 Lemmas for G and H

In this section we formulate a partitioning lemma for G, which asserts that G has a regular partition suitable for our purposes, and a corresponding partitioning lemma for H. Both these lemmas are tailored to the application of the version of the Blow-up Lemma that we will give in the next section.

We first introduce some notation. Let G, H and R be graphs with vertex sets $V(G)$, $V(H)$, and $V(R) = \{1, \dots, s\} =: [s]$. For $v \in V(G)$ and $S \subseteq V(G)$ we define $N(v, S) := N(v) \cap S$. Let $A, B \subseteq V(G)$ be non-empty and disjoint, and let $0 \le \varepsilon, \delta \le 1$.

The *density* of the pair (A, B) is $d(A, B) := e(A, B)/(|A||B|)$. The pair (A, B) is ε-*regular*, if $|d(A, B) - d(A', B')| \leq \varepsilon$ for all $A' \subseteq A$ and $B' \subseteq B$ with $|A'| \geq \varepsilon|A|$ and $|B'| \geq \varepsilon|B|$. An ε-regular pair (A, B) is called (ε, δ)-*regular* if $d(A, B) \geq \delta$, and (ε, δ)-*super-regular* if $|N(v, B)| \geq \delta|B|$ for all $v \in A$ and $|N(v, A)| \geq \delta|A|$ for all $v \in B$.

Let G have the partition $V(G) = V_1 \mathbin{\dot\cup} \ldots \mathbin{\dot\cup} V_s$ and H have the partition $V(H) = W_1 \mathbin{\dot\cup} \ldots \mathbin{\dot\cup} W_s$. We say that $(V_i)_{i\in[s]}$ is (ε, δ)-*(super-)regular on* R if (V_i, V_j) is an (ε, δ)-(super-)regular pair for every $ij \in E(R)$. In this case R is also called *reduced graph* of the (super-)regular partition. The partition classes V_i are also called *clusters*.

For all $n, k, r \in \mathbb{N}$, we call an integer partition $(n_{i,j})_{i\in[k],j\in[r]}$ of n r-*equitable*, if $|n_{i,j} - n_{i,j'}| \leq 1$ for all $i \in [k]$ and $j, j' \in [r]$. Let B_k^r be the kr-vertex graph obtained from a path on k vertices by replacing every vertex by a clique of size r and replacing every edge by a complete bipartite graph minus a perfect matching. More precisely, $V(B_k^r) = [k] \times [r]$ and

$$\{(i, j), (i', j')\} \in E(B_k^r) \quad \text{iff} \quad i = i' \text{ or } |i - i'| = 1 \wedge j \neq j'.$$

Let K_k^r be the graph on vertex set $[k] \times [r]$ that is formed by the disjoint union of k complete graphs on r vertices. Obviously, $K_k^r \subseteq B_k^r$.

Now we can formulate the partition lemma for G, which we take from [18, Lemma 6].

Lemma 5.7 (Lemma for G [18])
For all $r \in \mathbb{N}$ and $\gamma > 0$ there exists $d > 0$ and $\varepsilon_0 > 0$ such that for every positive $\varepsilon \leq \varepsilon_0$ there exist K_0 and $\xi_0 > 0$ such that for all $n \geq K_0$ and for every graph G on vertex set $[n]$ with $\delta(G) \geq (\frac{r-1}{r} + \gamma)n$ there exist $k \in [K_0]$ and a graph R_k^r on vertex set $[k] \times [r]$ with

(R1) $K_k^r \subseteq B_k^r \subseteq R_k^r$,

(R2) $\delta(R_k^r) \geq (\frac{r-1}{r} + \gamma/2)kr$, *and*

(R3) there is an r-equitable integer partition $(m_{i,j})_{i\in[k],j\in[r]}$ of n with $(1 + \varepsilon)n/(kr) \geq m_{i,j} \geq (1 - \varepsilon)n/(kr)$ such that the following holds.[1]

For every partition $(n_{i,j})_{i\in[k],j\in[r]}$ of n with $m_{i,j} - \xi_0 n \leq n_{i,j} \leq m_{i,j} + \xi_0 n$ there exists a partition $(V_{i,j})_{i\in[k],j\in[r]}$ of V with

(G1) $|V_{i,j}| = n_{i,j}$,

(G2) $(V_{i,j})_{i\in[k],j\in[r]}$ *is (ε, d)-regular on R_k^r, and*

(G3) $(V_{i,j})_{i\in[k],j\in[r]}$ *is (ε, d)-super-regular on K_k^r.*

The remainder of this section is dedicated to a corresponding partitioning lemma for H, which again will be similar to the Lemma for H in [18] (Lemma 8 in that paper). However, we need to strengthen the conclusion of this lemma. We shall point out the main differences below.

[1] The upper bound on $m_{i,j}$ is implicit in the proof of Lemma 7 in [18].

Again, we start with some definitions. Let H be a graph on n vertices and $\sigma : V(H) \to \{0,\dots,r\}$ be a proper $(r+1)$-colouring of H. A set $W \subseteq V(H)$ is called *zero free* if $\sigma^{-1}(0) \cap W = \emptyset$. Now assume that the vertices of H are labelled $1,\dots,n$ and that this labelling is a labelling of bandwidth at most βn for some $\beta > 0$. Given an integer ℓ, an $(r+1)$-colouring $\sigma : V(H) \to \{0,\dots,r\}$ of H is said to be (ℓ,β)-zero free with respect to such a labelling if any ℓ consecutive blocks contain at most one block with zeros. Here a block is a set of the form $B_t := \{(t-1)4r\beta n+1,\dots,t4r\beta n\}$, $t = 1,\dots,1/(4r\beta)$.

Lemma 5.8 (Lemma for H)
Let $r,k \geq 1$ be integers and let $\beta,\xi > 0$ satisfy $\beta \leq \xi^2/(1200r)$. Let H be a graph on n vertices and assume that H has a labelling of bandwidth at most βn and an $(r+1)$-colouring that is $(10/\xi,\beta)$-zero free with respect to this labelling. Let R^r_k be a graph with $V(R^r_k) = [k]\times[r]$ such that

(R1)* $K^r_k \subseteq B^r_k \subseteq R^r_k$, *and*

(R2)* *for every $i \in [k]$ there is a vertex $s_i \in ([k]\setminus\{i\})\times[r]$ with $\{s_i,(i,j)\} \in E(R^r_k)$ for every $j \in [r]$.*

Furthermore, suppose $(m_{i,j})_{i\in[k],j\in[r]}$ is an r-equitable integer partition of n with $m_{i,j} \geq 12\beta n$ for every $i \in [k]$ and $j \in [r]$. Then there exists a mapping $f : V(H) \to [k]\times[r]$ and a set of special vertices $X \subseteq V(H)$ with the following properties, where we set $W_{i,j} := f^{-1}(i,j)$.

(H1) $|X \cap W_{i,j}| \leq \xi n$ *and* $|N_H(X\cap W_{i,j}) \cap W_{i',j'}| \leq \xi n$ *for all* $i,i' \in [k]$, $j,j' \in [r]$,

(H2) $m_{i,j} - \xi n \leq |W_{i,j}| \leq m_{i,j} + \xi n$ *for every* $i \in [k]$ *and* $j \in [r]$,

(H3) *for every edge* $\{u,v\} \in E(H)$ *we have* $\{f(u),f(v)\} \in E(R^r_k)$, *and*

(H4) *if* $\{u,v\} \in E(H)\setminus E(H[X])$ *then* $\{f(u),f(v)\} \in E(K^r_k)$.

This lemma differs from Lemma 8 in [18] in that the conclusion (H4) is stronger. In order to obtain this stronger conclusion we had to strengthen the notion of zero-freeness as well. Nevertheless the proof of this modified Lemma for H closely follows the proof in [18]. We use the following propositions.

Proposition 5.9 (Proposition 20 in [18])
Let $c_1,\dots,c_r$ be such that $c_1 \leq c_2 \leq \dots \leq c_{r-1} \leq c_r \leq c_1 + x$ and $c'_1,\dots,c'_r$ be such that $c'_r \leq c'_{r-1} \leq \dots \leq c'_2 \leq c'_1 \leq c'_r + x$. If we set $c''_i := c_i + c'_i$ for all $i \in [r]$ then

$$\max_i\{c''_i\} \leq \min_i\{c''_i\} + x\,.$$

Proposition 5.10 (Proposition 22 in [18])
Assume that the vertices of H are labelled $1,\dots,n$ with bandwidth at most βn with respect to this labelling. Let $s \in [n]$ and suppose further that $\sigma\colon [n] \to \{0,\dots,r\}$ is a proper $(r+1)$-colouring of $V(H)$ such that $[s-2\beta n, s+2\beta n]$ is zero free.

Then for any two colours $l, l' \in [r]$ the mapping $\sigma'\colon [n] \to \{0, \dots, r\}$ defined by

$$\sigma'(v) := \begin{cases} l & \text{if } \sigma(v) = l', s < v \\ l' & \text{if } \sigma(v) = l, s + \beta n < v \\ 0 & \text{if } \sigma(v) = l, s - \beta n \le v \le s + \beta n \\ \sigma(v) & \text{otherwise} \end{cases}$$

is a proper *$(r+1)$-colouring of H.*

By repeatedly applying Proposition 5.10 we can transform a colouring of H into a balanced colouring by allowing some more vertices to be coloured with colour 0. This is a first step towards the proof of Lemma 5.8.

In order to make this precise we need the following definition. For $x \in \mathbb{N}$, a colouring $\sigma : [n] \to \{0, \dots, r\}$ is called *x-aligned*, if for each pair $a, b \in [n] \cup \{0\}$ and each $i \in [r]$, we have

$$\frac{b-a}{r} - x \le \left|\sigma^{-1}(i) \cap \{a+1, \dots, b\}\right| \le \frac{b-a}{r} + x$$

and $|\sigma^{-1}(0)| \le x$.

Proposition 5.11
Assume that the vertices of H are labelled $1, \dots, n$ with bandwidth at most βn and that H has an $(r+1)$-colouring that is $(2\ell, \beta)$-zero free with respect to this labelling. Let $\xi = 1/\ell$ and $\beta \le \xi^2/(12r)$. Then there exists a proper $(r+1)$-colouring $\sigma : V(H) \to \{0, \dots, r\}$ that is (ℓ, β)-zero free and $6\xi n$-aligned.

Proof. The idea of the proof is to split H into small parts and use Proposition 5.10 to switch colours in the parts. This allows us to even out differences in the sizes of the colour classes and obtain an aligned colouring.

Recall that the blocks $B_1, \dots, B_{1/4r\beta}$ of H are the vertex sets of the form $B_t = \{(t-1)4r\beta n + 1, \dots, t4r\beta n\}$.

We start by identifying so called *switching blocks.* They will be used to exchange the colours between parts of H. With the help of Proposition 5.10, which will colour some vertices with 0, we choose the switching blocks in such a way that every ℓ consecutive blocks contain at most one block which either had zeros in the original colouring or one switching block. As the ordering of H is $(2\ell, \beta)$-zero free this can be done so that every consecutive 3ℓ blocks contain at least one switching block. We next explain how to use the switching blocks.

Claim 5.12 *Let $\sigma : [n] \to \{0, \dots, r\}$ be a proper $(r+1)$-colouring of H, B_t a zero free block and π any permutation of $[r]$. Then there exists a proper $(r+1)$-colouring σ' of H with $\sigma'(v) = \sigma(v)$ for all $v \in \bigcup_{i<t} B_i$ and $\sigma'(v) = \pi(\sigma(v))$ for all $v \in \bigcup_{i>t} B_i$.*

Indeed, every permutation $[r]$ is the concatenation of at most r transpositions, i.e., permutations that exchange only two elements. We split the block B_t into r disjoint intervals of length $4\beta n$ and decompose π into at most r transpositions. The claim then follows from Proposition 5.10.

Let $\{s_1, s_2, \dots, s_p\}$ be the set of indices belonging to switching blocks. For ease of notation let $s_0 = 0$ and let $s_{p+1} = 1/(4r\beta)+1$. Further let $B^*(t) := \bigcup_{i\le t}\Big(\bigcup_{s_{i-1}\le j<s_i} B_j\Big)$, $c_i(t) := |\{v \in B^*(t) : \sigma(v) = i\}|$ and $\widetilde{c}_i(t) := |\{v \in B^*(t+1) \setminus B^*(t) : \sigma(v) = i\}|$ for $t \in [p]$. We inductively construct a proper $(r+1)$-colouring of H with

$$\max_i\{c_i(t)\} \le \min_i\{c_i(t)\} + \xi n \tag{5.5}$$

for every $t \in [p+1]$.

Note that any proper colouring of H satisfies (5.5) for $t = 1$ as $|B^*(1)| \le 3\ell 4r\beta n \le \xi n$ because $s_1 \le 3\ell$. So let σ be a proper $(r+1)$-colouring which satisfies (5.5) for all $t' \le t$. Without loss of generality we assume that $c_1(t) \le c_2(t) \le \dots \le c_r(t) \le c_1(t) + \xi n$. We define the switching for block t to be any permutation π which satisfies $\widetilde{c}_{\pi(r)}(t)+\xi n \ge \widetilde{c}_{\pi(1)}(t) \ge \widetilde{c}_{\pi(2)}(t) \ge \dots \ge \widetilde{c}_{\pi(r-1)}(t) \ge \widetilde{c}_{\pi(r)}(t)$. Such a permutation exists as $|B^*(t+1)\setminus B^*(t)| \le \xi n$. We apply Claim 5.12 to σ, the block B_t and the permutation π and obtain a new proper $(r+1)$-colouring σ'. Let $c'_i(t) := |\{v \in B^*(t) : \sigma'(v) = i\}|$. It follows from Proposition 5.9 that $c'_i(t+1) = c_i(t) + \widetilde{c}_{\pi(i)}(t)$ satisfies

$$\max_i\{c'_i(t+1)\} \le \min_i\{c'_i(t+1)\} + \xi n\,. \tag{5.6}$$

Therefore, the colouring σ' satisfies (5.5) for every $t' \le t+1$. Let σ^* be a colouring of H which satisfies (5.5) for every $t \le p+1$. Then σ^* is a proper $(r+1)$-colouring and (ℓ,β)-zero free by construction. It remains to show that σ^* is also $6\xi n$-aligned.

For this purpose consider any interval $[a,b] := \{a, a+1, \dots, b\} \subseteq [n]$ and let $a' = s_t$ for the smallest switching block index $s_t \ge a$. Similarly let b' be the biggest $s_{t'}$ with $s_{t'} \le b$. Fix a colour $i \in [r]$ and let $C_i := (\sigma^*)^{-1}(i)$. Clearly $C_i \cap [a,a']$ and $C_i \cap [b',b]$ are of size ξn at most. Moreover, if z denotes the number of vertices x in $[a',b']$ with $\sigma^*(x) = 0$, then

$$|C_i \cap [a',b']| = |C_i \cap [b']| - |C_i \cap [a']| \overset{(5.6)}{=} \frac{b'-a'-z}{r} \pm 2\xi n\,.$$

Because σ^* is (ℓ,β)-zero free we have $z \le \xi n$. Hence $\big((b-a)-(b'-a')-z\big)/r \le 2\xi n$ implies

$$|C_i \cap [a,b]| \le |C_i \cap [a',b']| \pm 2\xi n = \frac{b'-a'-z}{r} \pm 4\xi n = \frac{b-a}{r} \pm 6\xi n\,. \qquad \square$$

With the help of Proposition 5.11 and an appropriate method for "cutting up" a graph H with a balanced colouring we can now construct the homomorphism asserted by Lemma 5.8.

Proof of Lemma 5.8. Given r, k and β, let ξ, H and $R^r_k \supseteq B^r_k \supseteq K^r_k$ be as required. Assume without loss of generality that the vertices of R^r_k are labelled as induced by this copy of B^r_k. Assume moreover that the vertices of H are labelled $1, \dots, n$ with bandwidth at most βn and that H has a $(10/\xi,\beta)$-zero free $(r+1)$-colouring with respect to this labelling. Let $B_1, \dots, B_{1/(4r\beta)}$ be the corresponding blocks of H. Set

$\xi' = \xi/10$ and note that $\beta \le \xi^2/(1200r) = (\xi')^2/(12r)$. Therefore, by Proposition 5.11 with input β, $\ell = 1/\xi'$, and H, there is an (ℓ,β)-zero free and $6\xi' n$-aligned colouring $\sigma : V(H) \to \{0,\dots,r\}$ of H.

Given an r-equitable integer partition $(m_{i,j})_{i\in[k],j\in[r]}$ of n, set $M_i := \sum_{j\in[r]} m_{i,j}$ for $i \in [k]$. Now choose indices $0 = t_0 \le t_1 \le \cdots \le t_{k-1} \le t_k = 1/(4r\beta)$ such that B_{t_i} and B_{t_i+1} are zero free blocks and

$$\sum_{i'\le t_i} |B_{i'}| \le \sum_{i'\le i} M_{i'} < 12r\beta n + \sum_{i'\le t_i} |B_{i'}| \,. \tag{5.7}$$

Indeed, such t_i exist as σ is (ℓ,β)-zero free and, in particular, two out of every three consecutive blocks are zero free. Furthermore, the t_i are distinct because $m_{i,j} \ge 12\beta n$. The last βn vertices of the blocks B_{t_i} and the first βn vertices of the blocks B_{t_i+1} will be called *boundary vertices* of H. Observe that the choice of the t_i implies that boundary vertices are never assigned colour 0 by σ.

Using σ, we will now construct $f : V(H) \to [k] \times [r]$ and $X \subseteq V(H)$. For each $i \in [k]$, and each $v \in \bigcup_{t_{i-1}<i'\le t_i} B_{i'}$ we set

$$f(v) := \begin{cases} s_i & \text{if } \sigma(v) = 0, \\ (i, \sigma(v)) & \text{otherwise,} \end{cases}$$

where s_i is the vertex which exists by property (R2*). Further let

$$\begin{aligned} X_1 &:= \bigcup_{v\in\sigma^{-1}(0)} \big(\{v\} \cup N_H(v)\big), \\ X_2 &:= \big\{v \in V(H) \,:\, v \text{ is a boundary vertex}\big\}. \end{aligned}$$

It remains to show that f and $X := X_1 \cup X_2$ satisfy properties (H1)–(H4) of Lemma 5.8.

Recall that there are $1/(4r\beta)$ many blocks in the (ℓ,β)-zero free colouring σ. The bandwidth-ordering implies that all vertices from $X_1 \cup N(X_1)$ lie in blocks that either contain zeros or that are adjacent to blocks that contain zeros (because $|B_| \ge 4r\beta n$). Hence, at most $(3/\ell)/(4r\beta)+3$ out of $1/(4r\beta)$ blocks contain vertices from $X_1 \cup N(X_1)$. Furthermore, every $W_{i,j} = f^{-1}(i,j)$ contains at most βn boundary vertices and at most βn vertices adjacent to boundary vertices. Thus

$$\begin{aligned} |X \cap W_{i,j}| &\le |X_1| + |X_2 \cap W_{i,j}| \le \left(\frac{3(1/\ell)}{4r\beta} + 3\right) 4r\beta\, n + \beta n \\ &\le \frac{4}{4r\ell\beta} 4r\beta n + \beta n = \frac{4}{10}\xi n + \beta n \le \xi n \,, \end{aligned}$$

and

$$|N(X) \cap W_{i,j}| \le |N(X_1)| + |N(X_2) \cap W_{i,j}| \le \left(\frac{3(1/\ell)}{4r\beta} + 3\right) 4r\beta\, n + \beta n \le \xi n$$

and property (H1) holds.

It follows from (5.7) that $M_i - 12r\beta n \leq |\bigcup_{t_{i-1}<i'\leq t_i} B_{i'}| \leq M_i + 12r\beta n$. As σ is $6\xi'n$-aligned and $(m_{i,j})_{i\in[k],j\in[r]}$ is an r-equitable integer partition of n this implies

$$m_{i,j} - \xi n \leq \frac{M_i}{r} - 12\beta n - 6\xi' n \leq |f^{-1}(i,j)| \leq \frac{M_i}{r} + 12\beta n + 6\xi' n \leq m_{i,j} + \xi n$$

for every $j \in [r]$. Hence property (H2) is satisfied.

Let $\{u,v\} \in E(H) \setminus E(H[X])$ with $u \notin X$. Since vertices with colour 0 and their neighbours lie in X, we know that therefore $\sigma(u) \neq 0 \neq \sigma(v)$. Hence $f(u) = (i, \sigma(u))$ and $f(v) = (i', \sigma(v))$ for some $i, i' \in [r]$. If $i \neq i'$, u and v must both be boundary vertices, which contradicts $u \notin X$. Hence $i = i'$ and property (H4) follows.

Let $\{u,v\} \in E(H[X])$. As σ is a proper $(r+1)$-colouring, $\sigma(u) \neq \sigma(v)$. First assume that $\sigma(u) = 0$. Then there is an index $i \in [k]$ such that $f(u) = s_i$ and $f(v) = (i, \sigma(v))$. But $\{s_i, (i, \sigma(v))\} \in E(R_k^r)$ by condition (R2*) and so (H3) holds in this case. It remains to consider the case $\sigma(u) \neq 0 \neq \sigma(v)$. This implies that both u, v are boundary vertices of different colour. Since we started with an ordering of bandwidth at most βn we have $f(u) = (i, \sigma(u))$ and $f(v) = (i', \sigma(v))$ with $|i - i'| \leq 1$. Hence $\{f(u), f(v)\} \in E(B_k^r) \subseteq E(R_k^r)$ by condition (R1*) and so property (H3) also holds in this case. □

5.4 A Blow-up Lemma for arrangeable graphs

In this section we provide a Blow-up Lemma type result which we shall apply to prove Theorem 4.9 and Theorem 5.6. This results builds on the following Blow-up Lemma for arrangeable graphs from [16].

Theorem 5.13 (Arrangeable Blow-up Lemma, full version [16])
For all $C, a, \Delta_R, \kappa \in \mathbb{N}$ and for all $\delta', c > 0$ there exist $\varepsilon', \alpha' > 0$ such that for every integer s there is n_0 such that the following is true for every $n \geq n_0$. Assume that we are given

(*i*) *a graph R on vertex set $[s]$ with $\Delta(R) < \Delta_R$,*

(*ii*) *an a-arrangeable n-vertex graph H with maximum degree $\Delta(H) \leq \sqrt{n}/\log n$, together with a partition $V(H) = W_1 \,\dot\cup \ldots \dot\cup\, W_s$ such that $uv \in E(H)$ implies $u \in W_i$ and $v \in W_j$ with $ij \in E(R)$,*

(*iii*) *a graph G with a partition $V(G) = V_1 \,\dot\cup \ldots \dot\cup\, V_s$ that is (ε', δ')-super-regular on R and has $|W_i| \leq |V_i| =: n_i$ and $n_i \leq \kappa \cdot n_j$ for all $i, j \in [s]$,*

(*iv*) *for every $i \in [s]$ a set $S_i \subseteq W_i$ of at most $|S_i| \leq \alpha n_i$* image restricted vertices, *such that $|N_H(S_i) \cap W_j| \leq \alpha n_j$ for all $ij \in E(R)$,*

(*v*) *and for every $i \in [s]$ a family $\mathcal{I}_i = \{I_{i,1}, \ldots, I_{i,C}\} \subseteq 2^{V_i}$ of permissible* image restrictions, *of size at least $|I_{i,j}| \geq cn_i$ each, together with a mapping $I \colon S_i \to \mathcal{I}_i$, which assigns a permissible image restriction to each image restricted vertex.*

Then there exists an embedding $\varphi\colon V(H) \to V(G)$ such that $\varphi(W_i) = V_i$ and $\varphi(x) \in I(x)$ for every $i \in [s]$ and every $x \in S_i$.

This theorem requires super-regularity for all pairs used in the embedding. However, in applications this can usually not be guaranteed: Lemma 5.7 for example provides a partition of G where we know only for very few regular pairs that they are also super-regular.

The standard approach to deal with a situation like this is to apply the Blow-up Lemma only *locally* to small groups of clusters where super-regularity is guaranteed (such as the K_r-copies within K_k^r in Lemma 5.7) and to use image restrictions to connect these local embeddings into an embedding of the whole graph H.

Instead, here we combine Theorem 5.13 with a randomisation step in order to obtain the following version of the Blow-up Lemma for arrangeable graphs that can handle super-regular pairs and merely regular pairs at once.

This result will allow us to embed a spanning graph H at once by imposing the additional restriction that edges which are embedded into pairs that are regular but not necessarily super-regular are confined to a small subpair in this pair.

Theorem 5.14 (Arrangeable Blow-up Lemma, mixed version)
For all a, Δ_R, κ and for all $\delta > 0$ there exist $\varepsilon, \alpha > 0$ such that for every s there is n_0 such that the following is true for every $n_1, \dots, n_s$ with $n_0 \le n = \sum n_i$ and $n_i \le \kappa \cdot n_j$ for all $i, j \in [s]$. Assume that we are given graphs R, R^* with $V(R) = [s]$, $\Delta(R) < \Delta_R$ and $R^* \subseteq R$, and graphs G, H on $V(G) = V_1 \dot\cup \dots \dot\cup V_s$, $V(H) = W_1 \dot\cup \dots \dot\cup W_s$ with

(G1) $|V_i| = n_i$ for every $i \in [s]$,

(G2) $(V_i)_{i\in[s]}$ is (ε, δ)-regular on R, and

(G3) $(V_i)_{i\in[s]}$ is (ε, δ)-super-regular on R^*.

Further let H be a-arrangeable, $\Delta(H) \le \sqrt{n}/\log n$, and let there be a function $f : V(H) \to [s]$ and a set $X \subseteq V(H)$ with

(H1) $|X \cap W_i| \le \alpha n_i$ and $|N_H(X \cap W_i) \cap W_j| \le \alpha n_j$ for every $i \in [s]$ and every $ij \in E(R)$,

(H2) $|W_i| \le n_i$ for every $i \in [s]$,

(H3) for every edge $\{u, v\} \in E(H)$ we have $\{f(u), f(v)\} \in E(R)$,

(H4) for every edge $\{u, v\} \in E(H) \setminus E(H[X])$ we have $\{f(u), f(v)\} \in E(R^*)$.

Then $H \subseteq G$.

The idea of the proof is as follows. If $R = R^*$, that is, if all edges in R correspond to super-regular pairs in G, we are done by Theorem 5.13. In general of course this will not be the case. However, we will artificially create a situation like that: we carefully construct an auxiliary graph $G' \supseteq G$ which also has R as a reduced graph, but which has super-regular pairs for *all* edges in R. We then use Theorem 5.13 to embed H into G'. It will then remain to show that we constructed G' (and the image restrictions used in the application of Theorem 5.13) sufficiently carefully that this embedding in fact uses only edges from G.

Proof of Theorem 5.14. Let a, Δ_R, κ and $\delta > 0$ be given. Let $\varepsilon', \alpha' > 0$ as in Theorem 5.13 with $C := 1, a, \Delta_R, \kappa, \delta' := \delta/2$, and $c := 1/2$ and set $\varepsilon := \min\{\varepsilon'/2, 1/(2\Delta_R), \delta/2\}$, $\alpha := \alpha'$. Let s be given and choose n_0 as given by Theorem 5.13. Now let R, R^*, G, H have the required properties. In particular let $V(G) = V_1 \dot\cup \ldots \dot\cup V_s$, $V(H) = W_1 \dot\cup \ldots \dot\cup W_s$ be partitions such that $(V_i)_{i \in [s]}$ is (ε, δ)-regular on R and (ε, δ)-super-regular on R^*.

For $i \in [s]$ define U_i to be the set of all vertices $v \in V_i$ with $|N_G(v) \cap V_j| \geq (\delta - \varepsilon)n_j$ for all $j \in N_R(i)$. Since $\Delta(R) < \Delta_R$ and all pairs (V_i, V_j) with $j \in N_R(i)$ are (ε, δ)-regular we have

$$|U_i| \geq |V_i| - \Delta_R \varepsilon |V_i| \geq \tfrac{1}{2}|V_i| \,. \tag{5.8}$$

In the next step we construct a graph G' which is super-regular on all pairs (V_i, V_j) with $ij \in E(R)$. For every $ij \in E(R) \setminus E(R^*)$ we do the following. For every vertex $v \in V_i$ with $|N_G(v) \cap V_j| < (\delta - \varepsilon)n_j$ we add edges to δn_j randomly selected vertices in V_j thus ensuring the minimum degree for v in V_j. Let G' be the resulting graph. With positive probability, all pairs (V_i, V_j) with $ij \in E(R)$ are now $(2\varepsilon, \delta - \varepsilon)$-super-regular in G'. In particular, there exists at least one graph G' with (V_i, V_j) being an $(2\varepsilon, \delta - \varepsilon)$-super-regular pair in G' for every $ij \in E(R)$ and

$$G[V_i \cup V_j] = G'[V_i \cup V_j] \qquad \text{if } ij \in E(R^*), \tag{5.9}$$
$$G[U_i \cup U_j] = G'[U_i \cup U_j] \qquad \text{if } ij \in E(R). \tag{5.10}$$

As G' is (ε', δ')-super-regular for every $ij \in E(R)$ we have $H \subseteq G'$ by Theorem 5.13 even if, for every $i \in [s]$, we restrict the embedding of vertices in $S_i := W_i \cap X$ to $U_i \in \mathcal{I}_i := \{U_i\}$. This is possible by (5.8) and the fact that $|W_i \cap X| \leq \alpha n_i$ and $|N_H(W_i \cap X) \cap W_j| \leq \alpha n_j$ for all $i \in [s]$ and all $ij \in E(R)$.

Moreover, every $uv \in E(H) \cap W_i \times W_j$ with $ij \in E(R) \setminus E(R^*)$ has $u, v \in X$. Therefore, the embedding of H into G' also is an embedding of H into G by (5.9) and (5.10). $\square$

5.5 Proof of Theorem 5.3

Our strategy for this proof is as follows. We use the Lemma for G (Lemma 5.7) and the Lemma for H (Lemma 5.8) to get a partition of H and a matching regular partition of G which is (ε, δ)-(super-)regular wherever edges of H are to be embedded. Given these partitions, the Blow-up Lemma (Theorem 5.14) guarantees an embedding of H into G.

Proof of Theorem 5.3. We first set up the constants. Given r, a, $\gamma > 0$, let d, ε_0 be given by Lemma 5.7. Set $\Delta_R := 3r + 1/\gamma + 1$, $\kappa := 2$ and $\delta := d$ and let $\varepsilon_{T.5.14}$

and $0 < \alpha \leq 1$ be given by Theorem 5.14. Plug this $\varepsilon := \min\{\varepsilon_0, 1/4, \varepsilon_{T.5.14}\}$ into Lemma 5.7 and obtain K_0, ξ_0. If necessary decrease ξ_0 such that $\xi_0 \leq \alpha/(2rK_0)$. Choose β, ξ such that $\xi \leq \xi_0$ and $\beta \leq \xi^2/(1200r)$. Finally for every $s \leq r \cdot K_0$ let n_0 be sufficiently large for the application of Theorem 5.14.

Now let G be any graph on $n \geq n_0$ vertices with $\delta(G) \geq (\frac{r-1}{r}+\gamma)n$. Then Lemma 5.7 returns a $k \leq K_0$ and a graph $\widetilde{R}^r_k$ on vertex set $[k] \times [r]$ and an r-equitable integer partition $(m_{i,j})_{i\in[k],j\in[r]}$ with properties (R1)–(R3). In particular,

$$m_{i,j} \geq \frac{n}{2kr} \geq \frac{n}{2k}\frac{2K_0\xi_0}{\alpha} \geq \xi n \geq \sqrt{1200r\beta n} \geq 12\beta n$$

for all $i \in [k]$, $j \in [r]$.

With this integer partition we return to Lemma 5.8. Let H satisfy the conditions of Theorem 4.9, in particular H is r-chromatic and has bandwidth at most βn. Hence, clearly there is a labelling of bandwidth at most βn with a $(10/\xi, \beta)$-zero free $(r+1)$-colouring. Furthermore, we need to show that there is a graph R^r_k with $B^r_k \subseteq R^r_k \subseteq \widetilde{R}^r_k$ which satisfies conditions (R1*) and (R2*) of Lemma 5.8 and additionally has $\Delta(R^r_k) < \Delta_R$. Indeed, R^r_k can be obtained as follows. Recall that $\delta(\widetilde{R}^r_k) \geq \left(\frac{r-1}{r} + \gamma/2\right)kr$ by property (R2). Thus for every $i \in [k]$ there are at least $\frac{\gamma}{2}kr$ vertices $v \in ([k]\setminus\{i\}) \times [r]$ with $\{v, (i,j)\} \in E(\widetilde{R}^r_k)$ for all $j \in [r]$. We say that such a vertex v covers i. Now, consecutively choose for each $i = 1, \ldots, k$ a vertex $v_i \in [k] \times [r]$ among those vertices covering i which has been used as $v_{i'}$ as few times as possible for $i' < i$. Then the edges of R^r_k only consist of edges of B^r_k in $\widetilde{R}^r_k$ and all edges $\{v_i, (i,j) \in E(\widetilde{R}^r_k)$. Since $\Delta(B^r_k) \leq 3r$ we have by the choice of the v_i that $\Delta(R^r_k) \leq 3r + 2/\gamma < \Delta_R$. Hence R^r_k satisfies conditions (R1*) and (R2*) of Lemma 5.8.

As r, k, β, ξ and R^r_k and the r-equitable integer partition $(m_{i,j})_{i\in[k],j\in[r]}$ satisfy the requirements of Lemma 5.8, we obtain a mapping $f : V(H) \to [k] \times [r]$ and a set X which satisfy conditions (H1)–(H4). In the next step we will partition $V(G)$ into $(V_{i,j})_{i\in[k],j\in[r]}$. A vertex $x \in V(H)$ is then embedded into $V_{i,j} \subseteq V(G)$ if and only if $x \in f^{-1}(i,j)$.

Define $n_{i,j} := |f^{-1}(i,j)|$ and note that $m_{i,j} - \xi_0 n \leq n_{i,j} \leq m_{i,j} + \xi_0 n$ by property (H2). Thus there exists a partition of $V(G)$ into $(V_{i,j})_{i\in[k],j\in[r]}$ with properties (G1)–(G3) by Lemma 5.7. Moreover, $n_{i,j} \leq 2n_{i',j'}$ for all $i, i' \in [k]$ and $j, j' \in [r]$ by property (R3) and property (H2) as

$$n_{i,j} \leq m_{i,j} + \xi_0 n \leq (1+\varepsilon)\frac{n}{kr} + \xi_0 n \leq 2\left((1-\varepsilon)\frac{n}{kr} - \xi_0 n\right) \leq 2\left(m_{i',j'} - \xi_0 n\right) \leq 2n_{i',j'}\,.$$

Now all conditions of Theorem 5.14 are satisfied and thus $H \subseteq G$. □

5.6 Proof of Theorem 5.6

The proof of Theorem 5.6 closely follows the methods of Allen, Brightwell and Skokan [2]. The restriction on $\Delta(H)$ in their result (Theorem 5.5) originates from the embedding result they use (Theorem 24 in [2]). This embedding result in turn relies on the Blow-up Lemma and the Lemma for H in [18]. The following Lemma 5.15 is a consequence of our Lemma for H (Lemma 5.8). We shall use this lemma together with the Blow-up Lemma for arrangeable graphs (Theorem 5.13) to extend the result of Allen, Brightwell and Skokan to arrangeable graphs.

We denote by P_m^r the r-th power of a path P_m, that is, P_m^r has vertex set $[m]$ and edge set $\{uv : |u-v| \le r\}$. Analogously, C_m^r is the r-th power of the cycle C_m.

Lemma 5.15

For any $\xi > 0$ and for any natural numbers r', m_0 there exists $\beta > 0$ such that the following is true. Let H be a graph on n vertices that is r-colourable for $r \le r'$ and has $\mathrm{bw}(H) \le \beta n$. Then for any m with $2r \le m \le m_0$ there exists a homomorphism $f : H \to C_m^r$ with $|f^{-1}(i)| \le \frac{n}{m}(1+\xi)$ for every $i \in [m]$.

Proof. Let $\xi > 0$ and r', m_0 be given. We choose k' sufficiently large so that $m_0/k' \le \xi/3$ and so that $(k'+r'-1)/m$ is integer for each $m \in [m_0]$ and $r \in [r']$. We set

$$\xi' := \frac{\xi}{3k'r'} \qquad \text{and} \qquad \beta := \min\left(\frac{\xi'^2}{1200r'}, \frac{\xi}{6k'r'}\right).$$

Assume that H satisfies the requirements of the lemma. Observe that by the definition of β we can assume that the number of vertices n of H satisfies $n \ge 6k'r'/\xi \ge 6k'r/\xi$ and hence

$$1 + \xi' n = \frac{n}{k'r}\left(\frac{k'r}{n} + k'r\xi'\right) = \frac{n}{k'r}\left(\frac{k'r}{n} + \frac{\xi}{3}\right) \le \frac{n}{k'r} \cdot \frac{\xi}{2}. \tag{5.11}$$

Let m with $2r \le m \le m_0$ be given.

We would now like to start by applying Lemma 5.8 with parameters r, k' and β, ξ'. For this purpose let $R_{k'}^r$ be the graph obtained from $B_{k'}^r$ (defined in the beginning of Section 5.3) by adding all edges of the form $\{(i,j),(i+1,j)\}$ where $i \in [k'-1]$ and $i - j \equiv 0 \mod r$ (see Figure 5.1). These additional edges ensure that for every $i \in [k']$ there is a vertex $s_i = (i+1, i')$ or $s_i = (i-1, i')$ (where $i' \in [r]$ satisfies $i - i' \equiv 0 \mod r$) such that $\{s_i, (i,j)\} \in E(R_{k'}^r)$ for all $j \in [r]$. Hence the graph $R_{k'}^r$ satisfies conditions (R1*) and (R2*) of Lemma 5.8.

Furthermore let $\lfloor n/(k'r)\rfloor =: m_{1,1} \le m_{1,2} \le \cdots \le m_{k',r} := \lceil n/(k'r)\rceil$. Then Lemma 5.8 guarantees a mapping $f' : V(H) \to [k'] \times [r]$ and a set $X \subseteq V(H)$ with properties (H1)–(H4). In the following we call each set $f'^{-1}(i,j)$ with $i \in [k']$, $j \in [r]$ an *f'-class* and use these classes to define a homomorphism $f : V(H) \to C_m^r$ with the properties promised by Lemma 5.15.

We will construct f in two further steps. Recall that $V(R_{k'}^r) = [k'] \times [r]$ and consider the r-th power of a path $P_{k'+r-1}^r$ on vertex set $V(P_{k'+r-1}^r) = [k'+r-1]$. First we now define a mapping $f^* \colon [k'] \times [r] \to [k'+r-1]$ whose purpose is to group the f'-classes and which is a homomorphism from $R_{k'}^r$ to $P_{k'+r-1}^r$. Let $(i,j) \in [k'] \times [r]$. Observe that there are unique positive integers ℓ and x such that $x \in [r]$ and $i = -(r-j) + r \cdot \ell + x$. Then set $f^*(i,j) := r \cdot \ell + j$ (see also Figure 5.1). This guarantees for all $y \in [k'+r-1]$ that at most r pairs (i,j) are mapped to y, all of which have the same j-coordinate. In fact only the first and the last $r-1$ values y have less than r such pairs mapped to y, which we call the *exceptional preimages*. Moreover it is easy to verify that $|f^*(i,j) - f^*(i',j')| \le r$

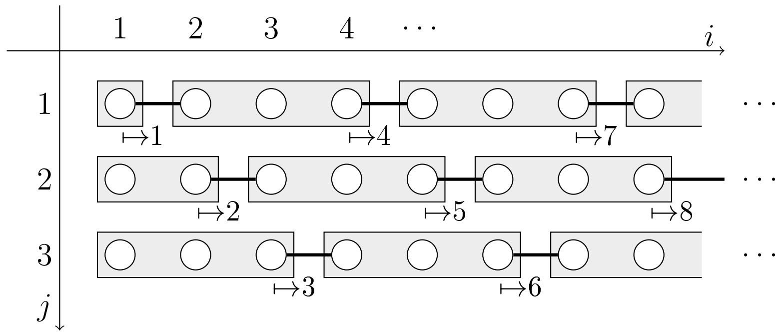

Figure 5.1: An illustration of f^* for $r = 3$. The white circle in column i and row j represents the set $f'^{-1}(i,j)$. The homomorphism f^* groups these sets as indicated. The thick vertical edges indicate the additional edges $\{(i,j), s_i\}$ of the reduced graph $R^r_{k'}$. For example $f^*(4,2) = 5$.

whenever $|i - i'| \leq 1$, that $f^*(i,j) = f^*(i',j')$ only if $j = j'$, and that $f^*(i,j) \neq f^*(s_i)$ for all $i, i' \in [k']$ and $j, j' \in [r]$. Hence f^* is a homomorphism from $R^r_{k'}$ to $P^r_{k'+r-1}$.

Our second step is to define the mapping $f^{**} \colon [k'+r-1] \to [m]$ by setting $f^{**}(y) := (y \mod m) + 1$ for all $y \in [k' + r - 1]$. Clearly f^{**} is a homomorphism from $P^r_{k'+r-1}$ to C^r_m. In conclusion, $f := f^{**} \circ f^* \circ f'$ is a homomorphism from H to C^r_m.

It remains to verify that also $|f^{-1}(i)| \leq \frac{n}{m}(1 + \xi)$ for every $i \in [m]$. Indeed, by (H2) of Lemma 5.8 we have $|(f')^{-1}(i,j)| = m_{i,j} \pm \xi' n$ for all $i \in [k'], j \in [r]$. Moreover, by construction the preimages of f^* are all of size at most r and only $2(r-1)$ of these preimages, the exceptional preimages, are smaller than r. The preimages of f^{**} are all of the same size and f^{**} maps at most one vertex with exceptional preimage under f^* to each vertex of C^r_m. Thus, because $f^{**} \circ f^*$ is a mapping from $[k'] \times [r]$ to $[m]$, the preimages of $f^{**} \circ f^*$ are all of size $\frac{k'r}{m} \pm r$. Hence, in total for each $i \in [m]$ we have

$$
\begin{aligned}
|f^{-1}(i)| &= (m_{i,j} \pm \xi' n) \cdot \left(\frac{k'r}{m} \pm r \right) = \left(\frac{n}{k'r} \pm 1 \pm \xi' n \right) \cdot \frac{k'r}{m} \left(1 \pm \frac{m}{k'} \right) \\
&\overset{(5.11)}{=} \frac{n}{k'r} \left(1 \pm \frac{\xi}{2} \right) \cdot \frac{k'r}{m} \left(1 \pm \frac{\xi}{3} \right) = \frac{n}{m} (1 \pm \xi),
\end{aligned}
$$

where we used $m_{i,j} = \frac{n}{k'r} \pm 1$ in the second equality and $\frac{m}{k'} \leq \frac{m_0}{k'} \leq \frac{1}{3}\xi$ in the third. □

For the proof of Theorem 5.6 we additionally need the following lemma, which is implicit in [2] in the proof of Theorem 5.5. Before we can state this lemma we need some further definitions.

Assume we are given a complete graph K_n whose edges are red/blue-coloured. Let A and B be disjoint vertex sets in K_n. Then (A, B) is a *coloured ε-regular* pair if (A, B) is an ε-regular pair in the subgraph of K_n formed by the red edges. It is easy to see that such a pair is also ε-regular in blue. A vertex partition $(V_i)_{i \in [s]}$ of $V(K_n)$ is called coloured ε-regular if all but at most $\varepsilon\binom{s}{2}$ of the pairs (V_i, V_j) with $\{i,j\} \in \binom{s}{2}$ are not coloured ε-regular. The *coloured reduced graph R* corresponding to this partition is the graph with vertex set $[s]$ and an edge for exactly each coloured ε-regular pair. Each edge ij of R is coloured in the majority-colour of the edges of (V_i, V_j). This clearly implies that if ij is a red edge of R, then the subgraph of (V_i, V_j) formed by the red edges is $(\varepsilon, \frac{1}{2})$-regular.

Lemma 5.16 (Implicit in [2])
For every $\varepsilon > 0$, r, and $\widetilde{m}$ there exists k_0 and n_0 such that the following is true for every $n \geq n_0$. Let the edges of K_n be red/blue-coloured.

(*a*) *The graph K_n has a coloured ε-regular partition $(V_i)_{i\in[k]}$ with $(2r+3)\widetilde{m} \leq k \leq k_0$ and $|V_1| \leq |V_2| \leq \ldots |V_k| \leq |V_1| + 1$.*

Let R be the coloured reduced graph corresponding to this partition and let m be any multiple of $r+1$ with $k \geq (2r+3)m$.

(*b*) *The graph R contains a monochromatic copy of C^r_m.*

We now apply Lemma 5.16, Lemma 5.15 and Theorem 5.13 to derive the following result, which in view of (5.1), (5.2), (5.3) and (5.4) directly implies Theorem 5.6.

Theorem 5.17
Given $a \geq 1$, there exists n_0 and $\beta > 0$ such that, whenever $n \geq n_0$ and H is an a-arrangeable n-vertex graph with maximum degree at most $\sqrt{n}/\log n$ and $\mathrm{bw}(H) \leq \beta n$, we have $R(H) \leq (2\chi(H)+4)n$.

Proof. Let a be given and set $r' := a+1$ (and observe that every a-arrangeable graph is r'-colourable). Set $\xi := 1/(100r')$. Choose ε as given by Theorem 5.13 with $C := 0$, a, $\Delta_R := 2r'+1$, $\kappa := 2$ and $\delta' := 1/4$, $c := 1$. If necessary decrease ε such that $\varepsilon \leq \xi/(4r')$. Further set $\widetilde{m} := 100r'^2$. Let n'_0 and k_0 be as returned by Lemma 5.16 for these $\varepsilon, r', \widetilde{m}$. Then continue the application of Theorem 5.13 with $s := k_0$ and obtain n''_0. Set $m_0 := k_0$ and

$$n_0 := \max\{n'_0\,,\, n''_0\,,\, 100m_0r'\}\,.$$

Let $\beta > 0$ be as given by Lemma 5.15 with parameters ξ, r', and m_0. Finally, let n and H be given, set $r := \chi(H)$, and assume we have a red/blue-colouring of the edges of $K_{(2r+4)n}$.

Lemma 5.16(*a*) asserts that there is a coloured ε-regular partition $(V'_i)_{i\in[k]}$ of $K_{(2r+4)n}$ with $(2r+3)\widetilde{m} \leq k \leq k_0$ whose clusters differ in size by at most 1. Let R' be the coloured reduced graph of the partition $(V'_i)_{i\in[k]}$. Let m be the multiple of $r+1$ which satisfies $(2r+3)m \leq k < (2r+3)(m+r+1)$. Observe that this and $k \geq (2r+3)\widetilde{m}$ implies $m \geq \widetilde{m} - r$ and thus

$$\frac{1}{2}m \geq \frac{1}{2}(\widetilde{m} - r) \geq 2r^2 + 5r + 3 \tag{5.12}$$

because $\widetilde{m} = 100r'^2 \geq 100r^2$. Further, $m \leq k \leq k_0 = m_0$ and so

$$\frac{m}{n} \leq \frac{m_0}{n_0} \leq \frac{1}{100r'} \leq \frac{1}{100r}\,. \tag{5.13}$$

We conclude that we have

$$|V_i'| \geq \frac{(2r+4)n}{k} - 1 \geq \frac{(2r+4)n}{(2r+3)(m+r+1)} - 1 \overset{(5.12)}{\geq} \frac{(2r+4)n}{(2r+3.5)m} - 1$$
$$= \left(1 + \frac{0.5}{2r+3.5} - \frac{m}{n}\right)\frac{n}{m} \overset{(5.13)}{\geq} \left(1 + \frac{1}{20r} - \frac{1}{100r}\right)\frac{n}{m} \geq (1+2\xi)\frac{n}{m}$$

because $\xi = 1/(100r') \leq 1/(100r)$. In addition, by Lemma 5.16(b) there is a monochromatic C_m^r in R', without loss of generality a red C_m^r. Let $U \subseteq V(K_{(2r+4)n})$ be the set of all vertices contained in clusters of this C_m^r.

Our next step is to apply Lemma 5.15 to the graph H with parameters ξ, r', m_0, β, r and m. This lemma guarantees a homomorphism $f\colon H \to C_m^r$ with $|f^{-1}(i)| \leq (1+\xi)\frac{n}{m}$ for every $i \in [m]$. By setting $W_i := f^{-1}(i)$ we obtain a partition $(W_i)_{i\in V(C_m^r)}$ of H.

We finish the proof with an application of Theorem 5.13. In this application we will not have image restricted vertices and we will use $R := C_m^r$. Observe that $\Delta(R) = 2r < \Delta_R$ and thus ((i)) of Theorem 5.13 is satisfied. The partition $(W_i)_{i\in V(C_m^r)}$ and the conditions on H guarantee that also condition ((ii)) of Theorem 5.13 is satisfied.

Now let G' be the subgraph of K_n with vertices U and all red edges of $K_{(2r+4)n}$ in U. In the following we consider this graph as an uncoloured graph. Clearly the partition $(V_i')_{i\in[k]}$ induces a partition $(V_i')_{i\in V(C_m^r)}$ of G' which is $(\varepsilon, \frac{1}{2})$-regular on C_m^r. Moreover, since C_m^r has maximum degree $2r$, by deleting from each of these clusters V_i' at most $2r\varepsilon|V_i'| \leq \frac{1}{2}\xi|V_i'|$ vertices we can obtain a partition $(V_i)_{i\in V(C_m^r)}$ of a subgraph G of G' which is $(\varepsilon, \frac{1}{4})$-super-regular on C_m^r and satisfies $|V_i| \geq (1+\xi)\frac{n}{m} \geq |W_i|$. Hence for G and $(V_i)_{i\in V(C_m^r)}$ also condition ((iii)) of Theorem 5.13 is satisfied.

Thus Theorem 5.13 implies that there is a copy of H in G. This copy corresponds to a red copy of H in the red/blue-coloured $K_{(2r+4)n}$. □

5.7 Concluding remarks

Optimality of Theorem 4.9. The degree bound $\Delta(H) \leq \sqrt{n}/\log n$ in Theorem 4.9 arises from our proof method: For the Blow-up Lemma, Theorem 5.13, such a degree bound is necessary (see [16, Proposition 35]). For trees H, however, the corresponding result of Komlós, Sarközy, and Szemerédi [67] requires only the weaker condition $\Delta(H) = o(n/\log n)$. It is thus well possible that our maximum degree condition is not best possible and could be improved to $o(n/\log n)$.

Blow-up Lemmas. In the original formulation of the Blow-up Lemma [62, 63, 86] the regularity ε required for the super-regular pairs depends on the number of clusters k' used in an application. Consequently, this lemma can never be used on the *whole cluster graph* obtained from an application of the Regularity Lemma: the number of clusters k the Regularity Lemma produces depends on the required regularity ε. Moreover, all pairs used in the embedding have to be super-regular.

The Blow-up Lemma for arrangeable graphs formulated in [16] overcomes the first difficulty: Here ε only depends on the maximum degree of the reduced graph of the super-regular partition that is used. (In fact, fairly straight-forward modifications of the original Blow-up Lemma proof from [62] would also allow for a corresponding result for bounded degree graphs.)

In Theorem 5.14 we also overcome the second difficulty: Pairs into which we only want to embed few edges are now allowed to be merely ε-regular. This allows us to avoid the occasionally tedious procedure of setting up suitable image restrictions and then applying the Blow-up Lemma several times. This might turn out could be useful for other applications as well.

Degeneracy. Though by now many important graph classes were shown to be a-arrangeable for some constant a, the notion of arrangeability has the disadvantage of seeming somewhat artificial at first sight. The notion of degeneracy is more natural (and more general): A graph H is d-degenerate if there is an ordering of its vertices such that each vertex has at most d neighbours to its left.

It would be very interesting to obtain an analogue of Theorem 4.9 for d-degenerate graphs. However, most likely this problem is very hard. Indeed, a version of the Blow-up Lemma for d-degenerate graphs would imply the difficult and long-standing Burr-Erdös conjecture [20], which states that degenerate graphs have linear Ramsey number.

6 Large planar subgraphs in dense graphs (I)

In this paper we study an extremal question of the following type. Given a monotone property $\mathcal{P}$ and a function $m(n)$, how large does the minimum degree of a graph G of order n have to be in order to guarantee a subgraph with at least $m(n)$ edges and propery $\mathcal{P}$?

This question has first been asked for $\mathcal{P}$ being the planar graphs by Kühn, Osthus, and Taraz [74]. They asymptotically determined the number of edges in planar subgraphs of graphs with given minimum degree d for many values of d. Among other results they proved that in the case of $m(n) = 3n - 6$ (which gives a spanning triangulation) a minimum degree of $(2/3 + \gamma)n$ with any constant $\gamma > 0$ suffices as long as n is large enough. This was later improved to the optimal degree bound of $(2/3)n$ by Kühn and Osthus [71].

We extend the results of Kühn, Osthus, and Taraz by giving optimal conditions on the minimum degree for $m(n) = (2.25 + o(1))n$ and $m(n) = (2.5 + o(1))n$. Doing so we discover an interesting threshold behaviour of $m(n)$ for a minimum degree d slightly above $n/2$.

Let $\mathrm{pl}(G)$ denote the number of edges in the largest planar subgraph of G, and set

$$\mathrm{pl}(n, d) := \min\{\mathrm{pl}(G) : v(G) = n,\ \delta(G) \geq d\}\,. \tag{6.1}$$

This definition is due to Kühn, Osthus, and Taraz [74] who proved asymptotic results on $\mathrm{pl}(n, d)$ for many values of d. In particular they showed that for every $\gamma > 0$ there is a constant $C(\gamma)$ such that

$$\begin{aligned} \mathrm{pl}(n, \gamma n) &\geq 2n - C(\gamma)\ ,\ \text{and} \\ \mathrm{pl}(n, (\tfrac{1}{2} + \gamma)n) &\geq 3n - C(\gamma)\ . \end{aligned}$$

The class of complete bipartite graphs shows that the first inequality is optimal up to the value of the constant for every $\gamma \leq 1/2$.[1] The same is true for the second inequality by Euler's formula. What remains unknown to this point is the tiny gap of $n/2 < d \leq n/2 + o(n)$. This paper investigates the curious behaviour of $\mathrm{pl}(n, d)$ when d is slightly above $n/2$. Let us state some upper bounds for $\mathrm{pl}(n, d)$.

Theorem 6.1
Let $m \geq 3$ then

$$\mathrm{pl}(2m - 1, m) \leq 4.5m - 3\,,\ \textit{and} \tag{6.2}$$

$$\mathrm{pl}(2m, m + 1) \leq 5m - 3\,. \tag{6.3}$$

[1] Optimal constants for this case have recently been shown by Allen, Skokan, and Würfl in [3].

Furthermore, for every $k \geq 3$ and m sufficiently large we have

$$\mathrm{pl}(2m, m + \tfrac{1}{4}m^{1/k}) \leq (5 + \tfrac{1}{k})m\,. \tag{6.4}$$

So if the upper bounds of Theorem 6.1 are tight we have the following situation. The parameter stays at $\mathrm{pl}(n,d) \leq 2n$ for all $d \leq n/2$, then there is a jump as d goes slightly above $n/2$, i.e., already $\mathrm{pl}(2m-1,m) = (4.5+o(1))m$. This is followed by a second jump to $\mathrm{pl}(2m, m+1) = (5+o(1))m$ but then again $\mathrm{pl}(n,d) = (2.5+o(1))n$ for $n/2+1 \leq d \leq n/2 + n^{o(1)}$.

The two main results of this paper are the following.

Theorem 6.2

$$\mathrm{pl}(2m-1, m) = (4.5 + o(1))m\,.$$

Theorem 6.3

$$\mathrm{pl}(2m, m+1) = (5 + o(1))m\,.$$

We start out with a partitioning result using the regularity method in Section 6.1. Subsequently we prove Theorems 6.2 and 6.3 in Sections 6.3 and 6.2 respectively. We conclude with some final remarks in Section 6.4. Before doing so we supplement the proof of Theorem 6.1.

Proof of Theorem 6.1. Note that for any graph $G = H_1 \cup H_2$ we have

$$\mathrm{pl}(G) \leq \mathrm{pl}(H_1) + \mathrm{pl}(H_2).$$

The inequality (6.2) now follows as we set $H_1 = (V_1 \mathbin{\dot\cup} V_2, E_1)$ to be a complete bipartite graph with $|V_2| = m$ and $H_2 = (V_2, E_2)$ to be a perfect matching (or a matching and one path of length 2 if m is odd). Analogously inequality (6.3) is obtained if $|V_2| = m+1$ and H_2 is a Hamiltonian cycle.

Erdős and Sachs [44] prove that there is a graph H on $(2d)^g$ vertices with girth g and minimum degree at least d. Euler's formula now implies that such a graph H has $\mathrm{pl}(H) \leq (2d)^g/(1-2/g)$. It follows that $\mathrm{pl}(m, \tfrac{1}{2}m^{1/k}) \leq m/(1-2/k)$ if m is sufficiently large compared to k. Hence, for every k there is m such that

$$\mathrm{pl}\left(m + \tfrac{1}{4}m^{1/k}, \tfrac{1}{2}m^{1/k}\right) \leq \frac{1}{1-2/k}\left(m + \tfrac{1}{4}m^{1/k}\right) \leq \left(1 + \frac{1}{k}\right)m\,.$$

We now derive (6.4) by choosing $H_1 = (V_1 \mathbin{\dot\cup} V_2, E_1)$ to be a complete bipartite graph with $|V_1| = m - \tfrac{1}{4}m^{1/k}$, $|V_2| = m + \tfrac{1}{4}m^{1/k}$ and $H_2 = (V_2, E_2)$ to be a graph that attains the bound above. Then $\delta(H_1 \mathbin{\dot\cup} H_2) \geq m + \tfrac{1}{4}m^{1/k}$ and

$$\mathrm{pl}(H_1 \cup H_2) \leq 4m - 4 + \left(1 + \frac{1}{k}\right)m \leq \left(5 + \frac{1}{k}\right)m\,.$$

□

This chapter is organised as follows. In Section 6.1 we provide a key tool for the previously stated results, a partitioning lemma for dense graphs without large planar subgraphs. Its proof relies on the regularity method which we also introduce here. We continue with the proofs of Theorem 6.3 in Section 6.2. Slight modifications yield the proof of Theorem 6.2 in Section 6.3. We conclude with some remarks in Section 6.4.

6.1 Preliminary observations

A plane graph G is a *triangulation* if all its faces are bounded by triangles, and a *quadrangulation* if all faces are bounded by 4-cycles. Note that a triangulation on n vertices has $3n-6$ edges while a quadrangulation has $2n-4$ edges by Euler's formula.

Throughout this paper we omit floors and ceilings whenever this does not affect the argument. We write $|G|$ for the order of a graph G, $\delta(G)$ for its minimum degree, $\Delta(G)$ for its maximum degree and $e(A,B)$ for the number of edges between A and B. The *density* of a bipartite graph $G=(A\mathbin{\dot\cup} B,E)$ is defined to be

$$d(A,B) := \frac{e(A,B)}{|A|\,|B|}\,.$$

A pair (A,B) is called *ε-regular* if $|d(X,Y)-d(A,B)|\le\varepsilon$ for every $X\subseteq A$, $Y\subseteq B$ with $|X|\ge\varepsilon|A|$, $|Y|\ge\varepsilon|B|$. Given $\delta\in[0,1]$, we say that a pair (A,B) is *(ε,δ)-regular*, if it is ε-regular and has density at least δ.

A partition $V=V_0\mathbin{\dot\cup} V_1\mathbin{\dot\cup}\dots\mathbin{\dot\cup} V_k$ is called a *balanced partition* if $|V_1|\le|V_2|\le\dots\le|V_k|\le|V_1|+1$. The following degree form of the Regularity Lemma can be found, e.g., in [69, Theorem 1.10].

Lemma 6.4 (Regularity Lemma, degree form)
For every $\varepsilon,d>0$ there is an $n_{L.6.4}$ such that the following is true for every graph $G=(V,E)$ on $n\ge n_{L.6.4}$ vertices. There is an balanced partition $V=V_0\mathbin{\dot\cup} V_1\mathbin{\dot\cup}\dots\mathbin{\dot\cup} V_k$ and a spanning subgraph $G'\subseteq G$ such that

(R1) $k\le n_{L.6.4}$,

(R2) $|V_0|\le\varepsilon|V|$,

(R3) $G'[V_i]$ *is an empty graph for every* $i\in[k]$,

(R4) $G'[V_i\mathbin{\dot\cup} V_j]$ *is (ε,d)-regular or empty for every* $i,j\in[k]$,

(R5) $\deg_G(v)-\deg_{G'}(v)\le(d+\varepsilon)n$ *for all* $v\in V$.

For a partition $V=V_0\mathbin{\dot\cup} V_1\mathbin{\dot\cup}\dots\mathbin{\dot\cup} V_k$ we define a *reduced graph* R on vertex set $[k]$. R has an edge ij if and only if $G'[V_i,V_j]$ is (ε,d)-regular.

The partition of Lemma 6.4 becomes a powerful tool when combined with an embedding result such as the Blow-up Lemma of Komlós, Sárközy, and Szemerédi [62].[2]

Theorem 6.5 (Blow-up Lemma, [62])
Given a graph R on $[k]$ and positive parameters d,Δ, there is a positive $\varepsilon=\varepsilon(\delta,\Delta,k)$ such that the following holds. Suppose that H and G are two graphs with $V(H)=X_1\mathbin{\dot\cup}\dots\mathbin{\dot\cup} X_k$, $V(G)=V_1\mathbin{\dot\cup}\dots\mathbin{\dot\cup} V_k$ and $|X_i|=|V_i|$ for all $i\in[k]$. Moreover, assume that

(R1) $G[V_i\mathbin{\dot\cup} V_j]$ *is (ε,d)-super-regular for every* $ij\in E_R$,

[2]For surveys on Szemerédi's Regularity Lemma and the Blow-up Lemma see [58, 69].

(R2) for every edge $xy \in E_H$ *there is* $ij \in E_R$ *with* $x \in X_i$ *and* $y \in X_j$*,*

(R3) $\Delta(H) \leq \Delta$.

Then there is an embedding of H *into* G.

Unfortunately, the properties guaranteed by the degree form of the Regularity Lemma are not strong enough to apply the Blow-up Lemma in its original form. Hence we combine it with the following lemma which can be found in [74, Proposition 8]. We include its proof for the sake of being self contained.

Lemma 6.6
Let $G = (V, E_G)$ *with* $V = V_1 \,\dot\cup \ldots \dot\cup\, V_k$ *and* $R = ([k], E_R)$ *be such that* $G[V_i \,\dot\cup\, V_j]$ *is an* (ε, d)*-regular pair whenever* $ij \in E_R$*. Then each* V_i *contains a subset* V_i' *of size* $(1 - \varepsilon k)|V_i|$ *such that for every* $ij \in E_R$ *the graph* $G[V_i' \,\dot\cup\, V_j']$ *is* $(\varepsilon/(1 - \varepsilon k), d - k\varepsilon)$*-super-regular.*

Proof. Consider a bipartit subgraph $G[V_i \,\dot\cup\, V_j]$ of G. By the definition of (ε, d)-regularity there are less than $\varepsilon|V_i|$ vertices in V_i which have less than $(d-\varepsilon)|V_j|$ neighbours in V_j. So for every set V_i we can choose a set $V_i' \subseteq V_i$ of size at least $(1-\varepsilon k)|V_i|$ such that for each $j \in N_R(i)$ all vertices $x \in V_i'$ have at least $(d - \varepsilon)|V_j|$ neighbours in V_j. It can be easily checked that for every edge ij of R the graph $G[V_i' \,\dot\cup\, V_j']$ is $(\varepsilon/(1 - \varepsilon k), d - k\varepsilon)$-super-regular. □

The following Lemma 6.7 will be the key to Theorem 6.2 and Theorem 6.3. We will need it only in the cases when either $a = 9/4$ or $a = 5/2$.

Lemma 6.7
Let $2 < a \leq 2.5$*. Then for every* $\eta > 0$ *there is an* n_0 *such that the set of vertices of every graph* $G = (V, E)$ *with* $|V| = n \geq n_0$*,* $\delta(G) \geq n/2$ *and* $\mathrm{pl}(G) < (a - 15\eta)n$ *can be partitioned into sets* $V = A \,\dot\cup\, B \,\dot\cup\, C \,\dot\cup\, D$ *such that*

(R1) every $v \in C$ *has* $\deg(v, B \cup D) \geq n/4$ *and*
every $v \in D$ *has* $\deg(v, A \cup C) \geq n/4$*,*

(R2) every $v \in A$ *has* $\deg(v, B \cup D) \geq (1/2 - 2\eta)n$ *and*
every $v \in B$ *has* $\deg(v, A \cup C) \geq (1/2 - 2\eta)n$*,*

(R3) $|A| + |B| \geq (2 - a/2 + 5\eta)n$*,*

(R4) $(1/2 - 2\eta)n \leq |A \cup C| \leq |B \cup D|(1/2 + 2\eta)n$ *.*

The lemma states that any sufficiently large graph without a large planar subgraph can be partitioned into two (roughly equal) sets $A \,\dot\cup\, C$ and $B \,\dot\cup\, D$, such that the bipartite graphs spanned by $A \cup (B \cup D)$, and $B \cup (A \cup C)$ are basically complete and sets A and B contain a significant proportion of all vertices. We call a partition $V = A \,\dot\cup\, B \,\dot\cup\, C \,\dot\cup\, D$ with the properties (1)-(4) a *working partition.*

As mentioned in the introduction the following theorem was proved by Kühn, Osthus, and Taraz.

Theorem 6.8 (Theorem 2 of [74])
For every $\gamma > 0$ there is $C = C(\gamma)$ such that every graph G of order n and minimum degree at least γn contains a planar subgraph with at least $2n - C$ edges.

We prove Lemma 6.7 using the Regularity Lemma, Theorem 6.8 and the following three propositions.

Proposition 6.9
For every $\alpha > 0$ there is $n_{P.6.9}$ such that for every $n \geq n_{P.6.9}$ the complete tripartite graph $G = (V_1 \cup V_2 \cup V_3, E)$, with $|V_1| = |V_2| = n$ and $\alpha n \leq |V_3| \leq (2-\alpha)n$, contains a spanning planar graph which is a triangulation and has maximum degree at most $4/\alpha + 4$.

Proof. We claim that for every integer $n \geq 2$ there is quadrangulation on $V_1 \,\dot\cup\, V_2$ with maximum degree four. We prove this claim by induction on n. Actually we prove the stronger statement that for every even n there is a quadrangulation H on $V_1 \,\dot\cup\, V_2$ with $|V_1| = |V_2| = n$ with the following properties: The maximum degree of H is at most four and there is a face which is bounded by vertices of degree at most three. A cycle on four vertices establishes the induction base ($n = 2$). Now let H be a quadrangulation on $2n$ vertices with maximum degree four and let v_1, v_2, v_3, v_4 be the boundary vertices of a face in H such that $\deg_H(v_i) \leq 3$. We embed a cycle on w_1, w_2, w_3, w_4 into said face and connect v_i to w_i. The graph obtained satisfies the induction hypothesis for $2(n+2)$.

If n is odd we construct a quadrangulation on $2n-2$ vertices as described before. We then insert the edge w_1w_2 into a face v_1, v_2, v_3, v_4 with $\deg_H(v_i) \leq 3$ and add the edges w_1v_1, w_2v_2, w_1v_3. The graph obtained is a quadrangulation H on $V_1 \,\dot\cup\, V_2$ with maximum degree four.

We set $n_{P.6.9} = \lceil 2/\alpha \rceil$ and assume that $n \geq n_{P.6.9}$ in the following. Let H be the planar graph on $V_1 \cup V_2$ constructed before and let it be embedded into the plane. The next step involves the dual graph of H. The dual graph H^* is obtained from H by replacing each face of H with a vertex and connecting two vertices in H^* if and only if the corresponding faces in H are bounded by a common edge.

Note that H has $|V_1| + |V_2| - 2 \geq |V_3|$ faces and that the dual graph H^* of H has a Hamilton cycle. We use this Hamilton cycle to construct a subgraph $H' \subseteq H$ that has $|V_3|$ faces with the additional property that no face is bounded by more than $4/\alpha + 4$ edges. To do so we partition the Hamilton cycle in H^* into $|V_3|$ paths P_i with almost equal lengths. Now H' is the subgraph obtained from H by removing all edges that lie between faces f_j and f_k whenever jk is an edge of a path P_i in H^*. The graph H' by construction has $|V_3|$ faces each with at most

$$2\left(\left\lceil \frac{2n-2}{|V_3|} \right\rceil + 1\right) \leq 2\left(\left\lceil \frac{2}{\alpha} \right\rceil + 1\right) \leq \frac{4}{\alpha} + 4$$

bounding edges. Finally we embed one vertex of V_3 in each face and add all edges from those vertices to the boundary vertices of the faces they are embedded in. The graph obtained is a triangulation with the properties of the proposition. $\square$

Proposition 6.10
For all positive α, δ, η there exist $\varepsilon_{P.6.10} > 0$ and $n_{P.6.10}$ such that for every $n \geq n_{P.6.10}$ and for every tripartite graph $G = (V_1 \cup V_2 \cup V_3, E)$ such that $|V_1| = |V_2| = n$, $\alpha \leq |V_3| \leq (2-\alpha)n$, and all three pairs are $(\varepsilon_{P.6.10}, \delta/2)$-regular, we have $\mathrm{pl}(G) \geq 3(1-\eta)|G|$.

Proof. Let α, δ, η be given and set $\Delta = 4/\alpha + 4$, $k = 3$, $\kappa = \max\{2, 1/\alpha\}$, and $d = \delta/4$. Now let ε be given by Theorem 6.5. Set $\varepsilon_{P.6.10} := \min\{\varepsilon/2, \eta/2\}$ and $n_{P.6.10} := n_{P.6.9}/(1-1.5\varepsilon)$ where $n_{P.6.9}$ is given by Proposition 6.9 with parameter α. Now assume that $G = (V_1 \uplus V_2 \uplus V_3, E)$ satisfies the conditions of the proposition. We apply Lemma 6.6 to obtain V_1', V_2', V_3' with $V_i' \subseteq V_i$ and $|V_i'| = (1-1.5\varepsilon)|V_i|$ such that $G[V_i' \uplus V_j']$ is (ε, d)-super-regular for every $i \neq j$. We use Proposition 6.9 to construct a planar triangulation H with $\Delta(H) \leq 4/\alpha + 4$ on the vertex set $X_1 \uplus X_2 \uplus X_3$ where $|X_i| = |V_i'|$. This H has $3(1-1.5\varepsilon)|G| - 4 \geq 3(1-\eta)|G|$ edges and is a subgraph of G by Theorem 6.5. □

The following lemma can be found, e.g., in [7, Lemma 3.1].

Proposition 6.11 (Slicing lemma)
For every $\varepsilon, \delta, \alpha > 0$ with $\alpha \geq \varepsilon$ the following is true. Let (V_1, V_2) be an (ε, δ)-regular pair and let $W_i \subseteq V_i$ with $|W_i| \geq \alpha |V_i|$. Then (W_1, W_2) is $(\varepsilon', \delta - \varepsilon)$-regular where $\varepsilon' = \max\{2\varepsilon, \varepsilon/\alpha\}$.

In order to proceed we need the definition of a *cactus*. A cactus is a graph H_C such that the edges of H_C can be decomposed into edge-disjoint triangles $T_1, \ldots, T_m$, where for $i = 2, 3, \ldots, m$,

$$\left| V(T_i) \cap \bigcup_{j=1}^{i-1} V(T_j) \right| \leq 1 .$$

A cactus H_C is called *κ-bounded* if every vertex of H_C lies in no more than κ triangles.

We have now assembled the tools to prove our partitioning result, Lemma 6.7.

Proof of Lemma 6.7. Let $2 < a \leq 2.5$ and $\eta > 0$ be given. W.l.o.g. we can assume that $\eta < 1/5$. Let $\alpha = \eta^3$, $\delta = \eta^2/8$ and let $\varepsilon_{P.6.10}$ and $n_{P.6.10}$ be as given by Proposition 6.10. Set $\varepsilon = \min\{\delta/2, \varepsilon_{P.6.10}\eta^3\}$ and let $n_{L.6.4}$ be given by Lemma 6.4. Finally set $n_0 = 2n_{L.6.4} \cdot n_{P.6.10}$.

Now take a graph G on $n \geq n_0$ vertices which has minimum degree $\delta(G) \geq n/2$ and $\mathrm{pl}(G) < (a - 15\eta)n$. Apply Lemma 6.4 to G with ε, δ and obtain an ε-regular partition $V = V_0 \uplus V_1 \uplus \ldots \uplus V_k$, a subgraph $G' \subseteq G$, and a reduced graph R of G. We claim that $\delta(R) \geq (1/2 - 2\delta)k$. Indeed, every vertex $v \in V_1 \uplus \ldots \uplus V_k$ has at least $\delta(G) - (\delta - 2\varepsilon)n$ neighbours in $G'[V_1 \uplus \ldots \uplus V_k]$. The edges towards these neighbours all lie in dense regular pairs of R. Hence $(n/k) \cdot \delta(R) \geq (1/2 - 2\delta)n$.

We will work on this subgraph G' and, with the help of several claims, establish that V has a partition into $A \uplus B \uplus C \uplus D$ with properties (1)-(4) unless $\mathrm{pl}(G) \geq \mathrm{pl}(G') \geq (a - 15\eta)n$. For better readability we postpone the proofs of these claims to the end of this argument.

We start with the vertices in $V_1 \mathbin{\dot\cup} \ldots \mathbin{\dot\cup} V_k$. The vertices in V_0 will be added at a later point of time. Now set $\kappa = 1/\eta^2$ and let R_C be a vertex-maximal κ-bounded cactus in R.

Claim 6.12 $|R_C| \leq (a - 2 - 11\eta)k$

Since R_C now covers less than $(1/2 - 2\delta)k \leq \delta(R)$ of the vertices of R, one can find two vertices w_1, w_2 in $V(R) \setminus V(R_C)$ which are adjacent. Recall that R_C is a maximal κ-bounded cactus. Therefore the neighbourhoods $W_1 := N(w_1)$ and $W_2 := N(w_2)$ only intersect in vertices that lie in κ many triangles of R_C. But there are at most $(\eta^2/2)k$ such vertices. Thus $W_1 \cup W_2$ covers at least $(1 - 4\delta)k - (\eta^2/2)k = (1 - \eta^2)k$ many vertices in R.

We now set

$$\begin{aligned} A' &:= W_1 \setminus V(R_C), & B' &:= W_2 \setminus V(R_C), \\ C' &:= (W_1 \cap V(R_C)) \setminus W_2, & D' &:= (W_2 \cap V(R_C)) \setminus W_1, \\ \widehat{C} &:= \{v \in C' : N_R(v, A' \cup C') \leq \eta^2 k\}, & \widehat{D} &:= \{v \in D' : N_R(v, B' \cup D') \leq \eta^2 k\}. \end{aligned}$$

We further define $X := (V(R) \setminus (W_1 \cup W_2)) \cup (W_1 \cap W_2)$ and finally set

$$\begin{aligned} A'' &:= A' \cup \widehat{C}, & B'' &:= B' \cup \widehat{D}, \\ C'' &:= C' \setminus \widehat{C}, & D'' &:= (D' \setminus \widehat{D}) \cup X\,. \end{aligned}$$

Claim 6.13 $|\widehat{C} \cup \widehat{D}| \geq \frac{1}{2}|C' \cup D'| - 3\eta^2 k/2$

Note that $|X| \leq \frac{3}{2}\eta^2 k$. It follows from $|C'| + |D'| \leq |R_C|$, Claim 6.12, and Claim 6.13 that

$$|C''| + |D''| \leq \tfrac{1}{2}|C' \cup D'| + 3\eta^2 k/2 + 3\eta^2 k/2 \leq \tfrac{1}{2}(a - 2 - 11\eta)k + 3\eta^2 k\,.$$

Hence $|A''| + |B''| \geq k - |C''| - |D''|$ implies

$$|A''| + |B''| \geq (2 - a/2 + 5\eta)k\,. \tag{6.5}$$

As $\delta(R) \geq (1/2 - 2\delta)k \geq (1/2 - \eta^2/2)k$ we also have

$$|A''| + |C''| \geq (1/2 - \eta^2)k\,, \tag{6.6}$$

$$|B''| + |D''| \geq (1/2 - \eta^2)k\,. \tag{6.7}$$

Our definitions already yield some valuable properties. One is that all vertices of R that lie in A'' have most of their neighbours in $B'' \cup D''$ and vertices in B'' have most of their neighbours in $A'' \cup C''$.

Claim 6.14 *All vertices $v \in A''$ and $w \in B''$ have*

$$\deg_R(v, A'' \cup C'') \leq \eta^2 k \quad \textit{and} \quad \deg_R(w, B'' \cup D'') \leq \eta^2 k\,.$$

Now let A, B, C^*, D^*, denote the set of vertices of G which correspond to A'', B'', C'', D'' in R respectively. We want to point out the following lower bounds.

$$\begin{aligned} |A \cup B| &\geq (2 - a/2 + 5\eta - \varepsilon)n \geq (2 - a/2 + 4\eta)n \ , \\ |A \cup C^*|, |B \cup D^*| &\geq (1/2 - \eta^2 - \varepsilon)n \geq (1/2 - 2\eta^2)n \ , \qquad (6.8) \\ |A|, |B| &\geq (1/4 + 4\eta)n \ . \qquad (6.9) \end{aligned}$$

The first and second bound are immediate from (6.5) and (6.6) and the fact that $(1 - \varepsilon)n \leq |V_1 \cupdot \ldots \cupdot V_k|$. The last bound follows as $|C^*| + |D^*| \leq (a/2 - 1 - 5\eta)n \leq (1/4 - 5\eta)n$. Thus the partition has property (3).

We now add the vertices of V_0 arbitrarily to either C^* or D^* (without changing the denotation).

Claim 6.15 *There are sets C, D with*

(a) $C^* \cupdot D^* = C \cupdot D$,

(b) $|C^* \setminus C|, |D^* \setminus D| \leq \eta n$, *and*

(c) every $v \in C$ has $\deg_G(v, B \cup D) \geq n/4$,
every $v \in D$ has $\deg_G(v, A \cup C) \geq n/4$.

This is to say, we can move a small fraction of the vertices from C^* to D^* and vice versa to ensure property (1). Moreover, inequality (6.8) together with Claim 6.15 imply

$$\begin{aligned} |A| + |C| &\geq |A| + |C^*| - \eta n \geq (1/2 - 2\eta)n \ , \\ |B| + |D| &\geq |B| + |D^*| - \eta n \geq (1/2 - 2\eta)n \end{aligned}$$

and hence the partition satisfies property (4). We claim that the partition also has property (2).

Claim 6.16 *Every vertex $v \in A$ satisfies* $\deg(v, B \cup D) \geq (1/2 - 2\eta)n$, *and every vertex $v \in B$ satisfies* $\deg(v, A \cup C) \geq (1/2 - 2\eta)n$.

Assuming all five claims hold, the partition $V = A \cupdot B \cupdot C \cupdot D$ satisfies the properties (1)-(4) of the lemma. Except for those five claims the proof of Lemma 6.7 is complete. □

We now supplement the missing claims.

Proof of Claim 6.12. Let $T_1, \ldots, T_m$ be the triangles of R_C and suppose that it contains exactly $(a - 2 - 11\eta)k$ many vertices. If R_C has more vertices we exchange R_C for a sub-cactus with the specified number of vertices.

We argue now that in this case $\mathrm{pl}(G) \geq (a - 15\eta)n$. Indeed, for every set of the cluster graph which is covered by R_C, let us split it randomly into κ parts, where the first $\kappa - 1$ parts each contain an η^3-fraction of all vertices. The last part contains a fraction of the vertices which is at least $1 - (\kappa - 1)\eta^3 \geq 1 - \eta$. Then, for each triangle T_i for $i = 1, \ldots, m$ of the cactus we choose three sets of this partition, where we select large parts of the corresponding sets of a cluster graph if these sets appear in the recursive construction for the first time, and small parts if the set already belonged to one of the triangles $T_1, \dots, T_{i-1}$. Thus, all but at most an η-fraction of the vertices which belonged to sets covered by R_C are distributed among vertex disjoint triples, in which either all three sets are of equal size, or two large parts are of equal size and one is (much) smaller. Let T_i be a triangle on (V_1, V_2, V_3) and let (V_1', V_2', V_3') be the subsets chosen for this triple. Then (V_1', V_2', V_3') is an $(\varepsilon_{P.6.10}, \delta/2)$-regular triple by Proposition 6.11. Moreover, it satisfies $n' := |V_1'| = |V_2'|$, $\alpha n' \le |V_3'| \le (2-\alpha)n'$ with $n' \ge n_{P.6.10}$. Thus $G[V_1' \cup V_2' \cup V_3']$ contains a planar subgraph with $3(1-\eta)|V_1' \cup V_2' \cup V_3'|$ many edges by Proposition 6.10. We conclude that all clusters covered by R_C together induce a subgraph H_1 with

$$\mathrm{pl}(H_1) \ge 3(1-\eta)^2|H_1| \,.$$

Now let $H_2 := G[V(G) \setminus V(H_1)]$. Note that $\delta(H_2) \ge \delta(G) - |H_1| \ge \eta|H_2|$. Thus

$$\mathrm{pl}(H_2) \ge 2|H_2| - C(\eta)$$

by Theorem 6.8. Together with $|H_1| \ge (a - 2 - 11\eta - \varepsilon)n$ we have

$$\begin{aligned} \mathrm{pl}(G) &\ge \mathrm{pl}(H_1) + \mathrm{pl}(H_2) \\ &\ge 2(|H_1| + |H_2|) + (1-6\eta)|H_1| - C(\eta) \\ &\ge (a - 15\eta)n \,. \end{aligned}$$

We deduce that R_C covers less than $(a-2-11\eta)k$ vertices of R as the above contradicts contradicts our assumption $\mathrm{pl}(G) < (a - 15\eta)n$. □

Proof of Claim 6.13. We call a vertex in R_C that lies in only one triangle a *leaf*. Let ℓ_i be the number of triangles in R_C that have exactly i leaves. We can assume that $\ell_0 = 0$ as we could simply successively erase triangles without leaves and obtain a cactus with the same number of vertices. Then $|R_C| = 2(\ell_1 + \ell_2 + \ell_3) + c$ where c is the number of connected components in R_C. We now distinguish four cases for a vertex $v \in C'$ which itself does not lie in κ many triangles; the same argument works for $v \in D'$.

Case 0: v is not a leaf. Then it is only adjacent to vertices $u \in C'$ that lie in κ many triangles or in the same connected component as v in the cactus. Otherwise we could extend the cactus by the triangle uvw_1.

Case 1: v is the only leaf in its triangle. Then it is only adjacent to vertices $u \in C'$ that lie in κ many triangles. Otherwise we would exchange the current triangle of v against the triangle uvw_1 and obtain a larger κ-bounded cactus.

Case 2: v is one of two leaves in its triangle. Let the other leaf be v'. Note that we can easily extend the cactus by removing the current triangle and adding $vv'w_1$ if $v' \in C'$. So let us assume that $v' \in D'$. Now either v or v' is only adjacent to vertices in its respective set that lie in κ many triangles. Otherwise we could replace the triangle containing v, v' with two triangles on vuw_1 and $v'u'w_2$. We want to point out that possibly $v' \notin C' \cup D'$. But this happens for at most $(\eta^2/2)k$ vertices v.

Case 3: v is one of three leaves in its triangle. If v has a neighbour $u \in C'$ which lies in less than κ many triangles we extend the cactus by the triangle vuw_1.

Note that at most $(\eta^2/2)k$ many vertices lie in κ many triangles. Indeed, R_C has at most $k/4$ many triangles while each triangle has at least one leaf.

We conclude that any vertex $v \in C'$ that is adjacent only to vertices $u \in A' \cup C'$ that lie in κ many triangles has $\deg(v, A' \cup C') \leq \eta^2 k$. (We also have $\deg(v, B' \cup D') \leq \eta k$ for $v \in D'$ if v is adjacent only to vertices $u \in B' \cup D'$ that lie in κ many triangles.) Therefore, at least $\ell_1 + \ell_2 + 3\ell_3 - (\eta^2/2)k$ many leaves v in $C' \cup D'$ have a high degree outside their class.

If $v \in C'$ is not a leaf, lies in less than κ many triangles, and is part of a component with at most $(\eta^2/2)k$ many vertices, then also $N(v, A' \cup C') \leq (\eta^2/2)k$. (The analogous statement holds for $v \in D'$.) But there less than $2/\eta^2$ components with more than $(\eta^2/2)k$ many vertices and at most $(\eta^2/2)k$ many vertices lie in κ many triangles. So at least $\max\{0, c - 2/\eta^2 - (\eta^2/2)k\}$ many of the non-leaf vertices satisfy $|N(v, A' \cup C')| \leq \eta^2 k$ or $|N(v, B' \cup D')| \leq \eta^2 k$ respectively.

Recall that $|C' \cup D'| \leq |R_C| = 2(\ell_1 + \ell_2 + \ell_3) + c$. Thus

$$\ell_1 + \ell_2 + 3\ell_3 - (\eta^2/2)k + \max\{0, c - 2/\eta^2 - (\eta^2/2)k\} \geq \tfrac{1}{2}|C' \cup D'| - 3\eta^2 k/2$$

vertices of $C' \cup D'$ lie in $\widehat{C} \cup \widehat{D}$ and the claim follows. □

Proof of Claim 6.14. By symmetry it suffices to establish the claim for vertices $v \in A''$. The case $v \in \widehat{C}$ follows by definition. So assume that $v \in A' = A'' \setminus \widehat{C}$. Since the cactus R_C was maximal, all neighbours $u \in A' \cup C'$ of v lie in κ many triangles. But again there are at most $(\eta^2/2)k$ many vertices that lie in κ many triangles. □

Note that Claim 6.14 implies that any vertex $x \in A$ and any $y \in B$ has

$$\deg_H(x, B \cup D^*) \geq (1/2 - 2\eta^2)n \,, \tag{6.10}$$
$$\deg_H(y, A \cup C^*) \geq (1/2 - 2\eta^2)n \,. \tag{6.11}$$

Indeed, $\deg_H(x) \geq (1/2 - \eta^2)n$ and x has at most $\eta^2 n$ neighbours in $A \cup C^*$ by Claim 6.14. The same argument works for $y \in B$.

Proof of Claim 6.15. It follows by double counting from (6.8), (6.9), and (6.11) that all but at most ηn many vertices $v \in C^*$ have $\deg_H(v, B) \geq (1 - 4\eta)|B| \geq n/4$. The analogous argument works for vertices in D^*. We one by one move vertices $v \in C^*$ with $\deg(v, B \cup D^*) < n/4$ from C^* to D^* and vertices $v \in D^*$ with $\deg(v, A \cup C^*) < n/4$ from D^* to C^*. Since all vertices have degree at least $n/2$, this process increases the density of edges between $A \cup C^*$ and $B \cup D^*$ in each step, and so must terminate. Furthermore we certainly move at most ηn vertices from each class. The new sets obtained are called C and D and satisfy the conditions of the claim. □

Proof of Claim 6.16. We argue that already the degrees in G' are large enough. To see this we once more turn to R. Let $v \in A$ be arbitrary and let $x \in A'' = A' \cup \widehat{C}$ be the cluster of R containing v. It follows from Claim 6.14 that $\deg_R(x, A'' \cup C'') \leq \eta^2 k$. Thus v has at most $\eta^2 n$ many neighbours in clusters of $A' \cup C' = A'' \cup C''$. It might have another $\frac{3}{2}\eta^2 n$ neighbours in clusters of X. All other neighbours must lie in clusters of $B' \cup D'$. As there are at least $(1/2 - \eta^2)n$ vertices in those clusters by (6.6) and $|(B \cup D^*) \setminus (B \cup D)| \leq \eta n$ we infer that

$$\deg_{G'}(v, B \cup D) \geq (1/2 - \eta^2)n - \eta n \geq (1/2 - 2\eta)n . \qquad \square$$

The following result on matchings in bipartite graphs will be convenient in the proof of both Theorem 6.2 and Theorem 6.3.

Proposition 6.17
Let $G = (A \,\dot\cup\, B, E)$ be a bipartite graph with $A = \{a_1, a_1', \ldots, a_m, a_m'\}$ and $|N(a_i)|$, $|N(a_i')| \geq (1 - \mu)m$, $N(a_i) \cap N(a_i') = \emptyset$ for all $i \in [m]$. Then there is a matching that covers all but $2\mu m$ vertices of A.

Proof. By deleting edges if necessary, we will assume that every vertex in A has degree exactly $(1-\mu)m$. Then it is easy to see that a minimal vertex cover in G must have at least $2(1 - \mu)m$ vertices. Indeed, G has $2(1 - \mu)m^2$ edges but the maximum degree is bounded by m. It follows from König's Theorem that G has a matching with $2(1-\mu)m$ edges. $\square$

6.2 Proof of Theorem 6.3

Before we go into details let us define a special class of graphs. A *jellyfish* is a connected graph with exactly one cycle. Equivalently, a jellyfish can be described as a tree plus one additional edge, or as a connected graph whose vertex number equals the edge number. It is easy to see that any graph G with $\delta(G) \geq 2$ has a partition into vertex-disjoint jellyfish.[3]

This section is dedicated to the proof of Theorem 6.3, i.e., we show that

$$\mathrm{pl}(2m, m + 1) \geq (5 + o(1))m.$$

We first give a rough outline of the main steps in the construction of the planar subgraph.

Partitioning G: We partition $V(G)$ into $A \,\dot\cup\, B \,\dot\cup\, C \,\dot\cup\, D$ with the help of Lemma 6.7 where $a = 2.5$ this time.

Finding a suitable planar subgraph F: Assume that $|B \cup D| > |A \cup C|$. Hence $\delta(G[B \cup D]) \geq 2$ and $G[B \cup D]$ has a planar subgraph F that is the disjoint union of jellyfish. In particular, $e(F) \geq n/2$.

Cutting up F: To simplify the handling of components we erase a small number of edges and obtain components which are either trees or jellyfish.

Pairing the leaves of F: We pair up the leaves of F in such a way that F plus the matching spanned by those pairs gives a planar graph. (Note that the pairs do not

[3] The authors wish to note that the correct collective noun for a group of jellyfish is a 'fluther'.

necessarily form edges in G, so the matching is imagined and will not be part of the planar subgraph of G which we construct.)

Attaching gadgets: We combine F with edges from the dense pairs $(A, B \cup D)$ and $(B, A \cup C)$ to obtain a planar graph H with sufficiently high average degree.

Taking care of left-over vertices: There may be vertices in $A \cup C$ that have not been used yet. We add those to H and connect each of them via two new edges.

The special case $|A \cup C| = |B \cup D|$: Finally we show how to adapt the proof when the two sides of the working partition have exactly the same size (and so we cannot assume $\delta(G[B \cup D]) \geq 2$).

We now give the details of the proof of Theorem 6.3.

Proof of Theorem 6.3. We will prove that for every $\eta > 0$ there is an integer n_0 such that any graph on n vertices with $n \geq n_0$ being an even integer that satisfies the requirements of Theorem 6.3 has a planar subgraph with at least $(2.5 - 3\eta^{1/12})n$ edges.

Let $\eta > 0$ be given. Set n_0 as given by Lemma 6.7 and let $G = (V, E)$ be a graph on $n \geq n_0$ vertices where n is an even integer. Assume that G does not have a planar subgraph with $(5 - 15\eta)n/2$ edges. Then G has a working partition $V = A \,\dot\cup\, B \,\dot\cup\, C \,\dot\cup\, D$ by Lemma 6.7. This partition might be such that $|B \cup D| = n/2 = |A \cup C|$. We will treat this case at the end of this section. For the moment we assume that $|B \cup D| > n/2$. Then $G[B \cup D]$ contains a spanning set of vertex-disjoint jellyfish F. For reasons we will explain later we will also admit some components into F which are actually trees. We say that F is *benign* if all its components can be classified as either Type I or Type II, where

(B1) every component of Type I is a tree which can be rooted such that the root z satisfies $\deg(z, A \cup C) \geq \frac{1}{2}|A \cup C|$ and every vertex $x \neq z$ in the tree satisfies $\deg(x, A \cup C) \geq (1 - 2\eta)|A \cup C|$,

(B2) every vertex y in a (jellyfish or tree) component of Type II satisfies $\deg(y, A) \geq (1 - 2\eta)|A|$, and

(B3) $\Delta(F) \leq |B \cup D|/2$.

In Lemma 6.21 we will prove that $G[B \cup D]$ with $\delta(G[B \cup D]) \geq 2$ contains a subgraph F which is a benign union of jellyfish and trees which has at least

So let us assume that F is a benign union of jellyfish and trees. Before we continue we set aside two vertices b_1, b_2 which have degree at most 2 in F and which lie in B (since $e(F) \leq |F|$ and $|B| \geq |F|/2$, such vertices exist). In doing so we lose at most 4 edges, which will not affect calculations significantly. Moreover, we reserve two vertices $a_1, a_2 \in A$ for later use. For technical reasons we do the following in each component of F which is a jellyfish. If the cycle of the component has less than two vertices of degree at least three, we select one vertex of degree two in the cycle and make it a *pseudo-leaf* (see Figure 6.1).

Moreover, we separate all jellyfish and trees with less than three leaves (where pseudo-leaves are counted as leaves). Those will be dealt with separately. Let $F' \subseteq F$ be the set of jellyfish and trees that have at least three leaves.

Now for a component S of F, let $L(S)$ denote the set of leaves. We aim to split the components of F into two classes which have an equal number of leaves. This will not always be possible, but we will choose an optimal partition. To this end, if $P \uplus Q$ is a partition of the components of F, we define $L(P) := \sum_{S \in P} L(S)$, similarly for Q and we define the *disparity* of the partition to be $|L(P) - L(Q)|$. We choose a partition $P \uplus Q$ with minimal disparity, and without loss of generality we may assume $L(P) \geq L(Q)$. We will consider three possible cases:

(R1) The disparity of $P \uplus Q$ is at most 1;

(R2) The disparity of $P \uplus Q$ is at least 2 and P has more than one component;

(R3) The disparity of $P \uplus Q$ is at least 2 and P has only one component.

In case (1) we simply pair up the leaves of F according to the partition. We choose an arbitrary order of the components $S_1, \ldots, S_p$ in P and an arbitrary order of the components $S'_1, \ldots, S'_q$ in Q. Then for some planar embedding of each component in which each leaf lies on a line in the outer face, we obtain an order of the leaves in that component by choosing the order of leaves along the line. We now have an ordering $\ell_1, \ldots, \ell_{L(P)}$ of all the leaves in P by taking all the leaves of S_1 in order, then all the leaves of S_2 in order and so on. Similarly we have an ordering $\ell'_1, \ldots, \ell'_{L(Q)}$ of all the leaves in Q. We now pair up ℓ_i with ℓ'_i for each $1 \leq i \leq L(Q)$. If the number of leaves is odd we remove the unpaired leaf from the forest. This only removes one edge which will not affect calculations significantly.

Case (2) is very similar, but now we alter the partition slightly to allow one component to appear in both classes, and its leaves to be divided between the two classes. Firstly, since $P \uplus Q$ was chosen to be a partition with minimal disparity, each component S of P satisfies $L(S) \geq L(P) - L(Q)$. Again we choose an ordering of the components of P and Q and of the leaves within each component, but now we take the last $d := L(P) - L(Q)$ leaves of S_p and pair them with the first d leaves of S_1 in the reverse order, i.e. we pair ℓ_i with $\ell_{\ell(P)+1-i}$ for $1 \leq i \leq d$. For the remaining leaves, we pair ℓ_{i+d} with ℓ'_i for each $1 \leq i \leq L(Q)$. We then relabel the leaves in such a way that the i-th pair contains ℓ_i and ℓ'_i.

Case (3) will be dealt with separately at the end of this section.

We now define the notion of a *socket*. We line up the components of F' in a way that the leaves ℓ_i and ℓ'_i lie opposite one another for all i.

Let ℓ_i^- be defined as ℓ_{i-1} if ℓ_i and ℓ_{i-1} lie in the same component, or the leaf with the highest index in the component containing ℓ_i otherwise. Define P_i to be the set of vertices on a path between ℓ_i and ℓ_i^- in F. If we have two choices for such a path (as will be the case in a jellyfish, since we may be able to go either way round the cycle) we choose the "inner path" i.e. the path which lies closer to the ℓ'_j if $\ell_i^- = \ell_{i-1}$, or the "outer path" otherwise. We similarly define $(\ell'_i)^-$ and P'_i. The i-th socket then consists of $P_i \cup P'_i$.

Here a leaf is called *open* for a_1 if it is a leaf ℓ_i with the highest index in its star or the leaf paired with this ℓ_i. It is called open for a_2 if it is a leaf ℓ'_i with the highest index

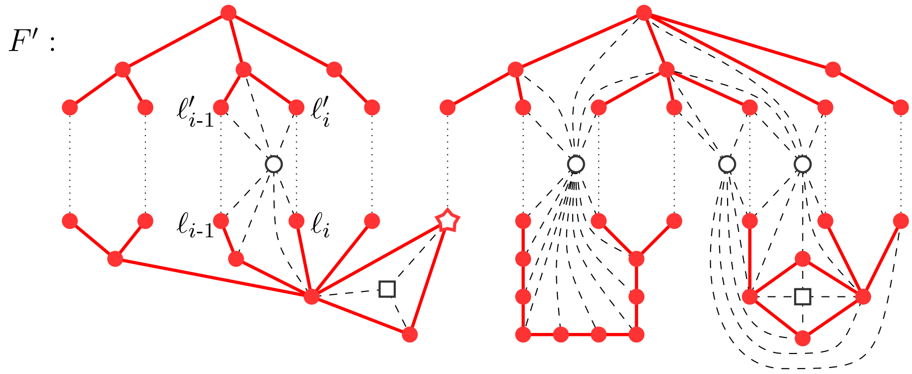

Figure 6.1: Pairing a partition into trees and jellyfish; the star-shaped vertex marks a pseudo-leaf; the black circles represent (examples of) sockets; the black squares represent hubs.

in its star or the leaf paired with this ℓ'_i. The only exception to this scheme appears when a pair of leaves is open for both a_1 and a_2. Instead of adding all four edges to this pair we only add three of them. The fourth edge (from a_1) instead goes to the leaf with the highest index in the star containing ℓ_{i+1}, where the leaves ℓ_i are ordered cyclically. Indeed, there is a planar embedding of the star forest with its gadgets and all edges from a_i to their respective open leaves. This gives another two edges per star (see Figure 6.7).

A *gadget* for a socket now consists of one or two vertices from $A \cup C$ that is/are adjacent to (most of) the vertices of the socket. Since a socket potentially has many vertices we only require the gadget to be adjacent to most vertices of a socket. More precisely we distinguish between the following two cases. An *ordinary gadget* is one vertex that is adjacent to at least $(1-\sqrt{\eta})|P_i \cup P'_i|$ of the vertices in $P_i \cup P'_i$. A *split gadget* consists of two vertices x, x' where x is adjacent to at least $(1-\sqrt{\eta})|P_i|$ vertices of P_i and x' is adjacent to $(1-\sqrt{\eta})|P'_i|$ vertices of P'_i. In addition we require that either x' is adjacent to $\{\ell_i^-, \ell_i\}$ or x is adjacent to $\{(\ell'_i)^-, \ell'_i\}$.

For each cycle in F (not only in F'!) we define a *hub gadget* to be a vertex from A which is adjacent to an $(1-\sqrt{\eta})$-fraction of all vertices of the cycle.

We will later see that the proof of Theorem 6.2 is a special case of this where all trees are stars, all sockets contain exactly six vertices, and every gadget is adjacent to six (or five and three respectively) of these vertices.

Next we prove that we can find gadgets and hubs for almost all sockets and cycles. So let g_o and g_s denote the number of ordinary and split gadgets respectively. By g_h we denote the number of hub gadgets.

A *potential half-gadget* for a side of the i-th pair is a vertex which is adjacent to $(1-\sqrt{\eta})|P'_i|$ vertices of P'_i or to $(1-\sqrt{\eta})|P_i|$ vertices of P_i respectively.

Claim 6.18 *Every side of a socket has at least $(1/4-\sqrt{\eta})n$ potential half-gadgets and every cycle has at least $(1/4-\sqrt{\eta})n$ potential hub gadgets.*

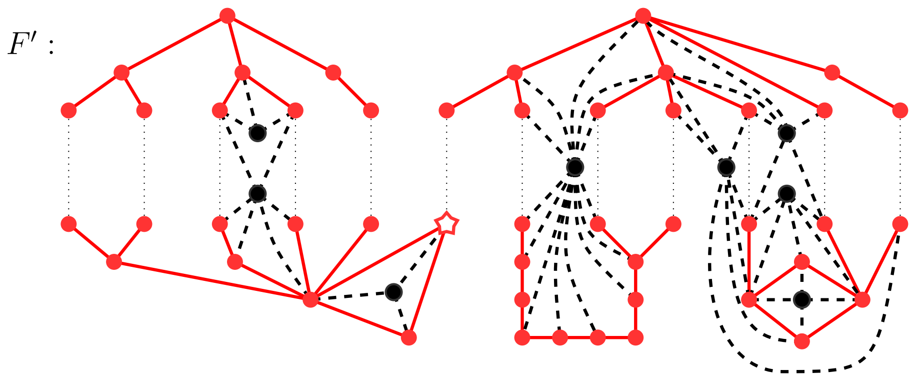

Figure 6.2: Some examples of ordinary, split and hub gadgets.

Proof. We will use properties (B1) and (B2) to derive this claim. Accordingly, we distinguish between half-gadgets for vertices in a Type I component (which have property (B1)) and half-gadgets (or hub gadgets) for vertices in a Type II component (which have property (B2)).

Case 1: P_i is part of a Type I component. Then P_i is part of a tree. Let z be the root of the tree and set $M_i := \{x \in A \cup C : |N(x) \cap (P_i \setminus \{z\})| \geq (1 - \sqrt{\eta})|P_i \setminus \{z\}|\}$. With (B1) we obtain

$$(1 - 2\eta)|A \cup C| \cdot |P_i \setminus \{z\}| \leq |M_i| \cdot |P_i \setminus \{z\}| + (|A \cup C| - |M_i|)(1 - \sqrt{\eta})|P_i \setminus \{z\}|$$

which is true if and only if

$$(1 - 2\sqrt{\eta})|A \cup C| \leq |M_i| \, .$$

This implies that at least an $(1-2\sqrt{\eta})$-fraction of the vertices in $A \cup C$ are adjacent to at least $(1-\sqrt{\eta})|P_i \setminus \{z\}|$ vertices in $P_i \setminus \{z\}$. This settles Case 1 as z is adjacent to at least $(1/2-2\sqrt{\eta})|A \cup C|$ of those vertices. Hence we have $(1/2-2\sqrt{\eta})|A \cup C| \geq (1/4-2\sqrt{\eta})n$ potential half-gadgets.

Case 2: P_i is part of a Type II component. We set $M_i = \{x \in A : |N(x) \cap P_i| \geq (1 - \sqrt{\eta})|P_i|\}$ and with the help of (B2) we obtain

$$(1 - 2\eta)|A| \cdot |P_i| \leq |M_i| \cdot |P_i| + (|A| - |M_i|)(1 - \sqrt{\eta})|P_i|$$

which is true if and only if

$$(1 - 2\sqrt{\eta})|A| \leq |M_i| \, .$$

This implies that at least $(1 - 2\sqrt{\eta})|A| \geq (1/4 - 2\sqrt{\eta})n$ vertices in A are potential half-gadgets for P_i.

The argument of Case 2 also shows that each cycle (may it be in F' or in $F \setminus F'$ has at least $(1 - 2\sqrt{\eta})|A| \geq (1/4 - 2\sqrt{\eta})n$ potential hub gadgets in A. $\square$

We now assign gadgets to sockets and cycles. We claim that we can do so in a way that gives many edges between the vertices of the gadgets and F'. Those edges (i.e., edges between $A \cup C$ and F') will be called *gadget edges*. Let H be the planar graph obtained by adding gadgets (with gadget edges) to F'. The hub gadgets assigned to cycles in $F \setminus F'$ will be of importance at a later point.

Claim 6.19 *There is an assignment of gadgets to sockets and cycles that uses g_o ordinary gadgets, g_s split gadgets, and g_h hub gadgets which contributes at least*

$$2e(F') + 2g_o + 4g_s - 34\sqrt{\eta}n \tag{6.12}$$

gadget-edges to H.

Proof. We want to apply Proposition 6.17 to derive that most pairs and cycles receive a gadget. If this is the case we obtain at least the number of edges given in (6.12).

Let s be the number of pairs and cycles in F' plus the number of cycles in $F \setminus F'$. Since every cycle has at least two non-leaf vertices, s is bounded by $|B \cup D|/2 \leq (1/4 + \eta)n$. It is easy to see that either $s \leq (1/4 - 3\sqrt{\eta})n$ or F has at least $4\sqrt{\eta}n$ pairs of leaves or cycles that correspond to sockets or to cycles with at most six vertices. In the latter case we will give up on $s - (1/4 - 3\sqrt{\eta})n \leq 4\sqrt{\eta}n$ of those pairs and cycles and try to find gadgets only for the other pairs and cycles. In doing so we lose at most $32\sqrt{\eta}n$ edges in gadgets we gave up.

Now assume that the number of pairs and cycles in F' plus the number of cycles in $F \setminus F'$ is at most $(1/4 - 3\sqrt{\eta})n$. Claim 6.18 states that each side of a socket and each cycle has at least $(1/4 - 2\sqrt{\eta})n$ potential half-gadgets or hub gadgets.

First we (greedily) assign one hub gadget to each cycle. Since there are at most $(1/2 + 2\eta)n/3$ cycles in F and we have $(1/4 - 2\sqrt{\eta})n$ potential hub gadgets for each cycle, this can be done. Denote by g_h the number of hub gadgets we have used in this step.

In our next step we greedily assign ordinary gadgets to pairs until no more ordinary gadgets can be found or all pairs have received gadgets. (Note that if one vertex is a potential half-gadget for both sides of a pair it also is an ordinary gadget for this pair.) Assume we have assigned g_o many ordinary gadgets.

Now we are left with at most $(1/4 - 3\sqrt{\eta})n - g_h - g_o$ many pairs that have not received a gadget yet. At least $(1/4 - 2\sqrt{\eta})n - g_h - g_o$ potential half-gadgets for each side of those pairs are still unused.

Recall that our definition of a split gadget requires a bit more than just two potential half-gadgets. Let x be a potential half-gadget for P_i and let x' be a potential half gadget for P_i'. If in addition x is adjacent to $\{(\ell_i')^-, \ell_i'\}$ or x' is adjacent to $\{\ell_i^-, \ell_i\}$ then x and x' form a split gadget for the i-th pair.

We claim that we can restrict the set of potential half-gadgets for each pair and each side in such a way that

(a) at least $(1/4 - 3\sqrt{\eta})n - g_h - g_o$ potential half-gadgets for each side of a socket remain, and

(b) any assignment of those potential half-gadgets to pairs is such that two potential half-gadgets of a pair form a split gadget.

To show this, fix any pair i which has not received a gadget yet. First assume that at least one side of pair i lies in a Type I component. Let P_i be the vertices of this side and let M be the set of potential half-gadgets for the opposite side P_i'. If follows from property (B1) that all but $2\eta|A \cup C|$ vertices in M are adjacent to ℓ_i (or to ℓ_i^- respectively). Hence, we can restrict the set of potential half-gadgets to $M' \subseteq M$ with $|M'| \geq (1/4 - 2\sqrt{\eta})n - g_h - g_o - 2\eta n \geq (1/4 - 3\sqrt{\eta})n - g_h - g_o$.

If both sides of pair i lie in Type II components the set M of potential half-gadgets for P_i' is a subset of A. It follows from property (B1) that all but $(1 - 2\eta)|A|$ of those vertices are adjacent to ℓ_i (or to ℓ_i^- respectively). Hence, we can restrict M to a subset $M' \subseteq M$ with $|M'| \geq (1/4 - 2\sqrt{\eta})n - g_h - g_o - 2\eta n \geq (1/4 - 3\sqrt{\eta})n - g_h - g_o$.

Obviously, the restricted sets of potential half-gadgets satisfy the second condition above.

It remains to find an assignment of the (restricted) potential half-gadgets to the pairs that covers all pairs. Such an assignment exists by Proposition 6.17 (with $\mu = 0$) as we are left with $(1/4 - 3\sqrt{\eta})n - g_h - g_o$ many pairs but each side has at least $(1/4 - 3\sqrt{\eta})n - g_h - g_o$ potential half-gadgets.

We have now seen that we can find gadgets for all but at most $4\sqrt{\eta}n$ sockets and cycles. Those sockets that we chose to disregard would contribute at most $32\sqrt{\eta}n$ edges (since each of those sockets would receive at most 8 edges from a split gadget).

Assume for the moment that we had an assignment of gadgets to sockets and cycles that would leave no side of a socket or cycle empty. Moreover, assume that all gadgets used were adjacent to all the vertices of their respective socket or cycle. We claim that such an assignment adds

$$\sum_{x \in V(F')} d_{F'}(x) + \ell(F') + 2g_s = 2e(F') + 2g_o + 4g_s$$

gadget-edges to H. Indeed, note that each vertex $x \in V(F')$ which is not a leaf lies in $d_{F'}(x)$ many sockets; each leaf lies in 2 sockets. Hence the gadgets add $d_{F'}(x) + \ell(F')$ many edges. Each split gadget adds another 2 edges (as we have 8 edges for a split gadget compared to 6 edges for an ordinary one).

However, we may not find gadgets for all sockets and the gadgets found may only be adjacent to an $(1 - \sqrt{\eta})$-fraction of their respective socket or cycle. Still we find an assignment of gadgets that adds at least

$$(1 - \sqrt{\eta})(2e(F') + 2g_o + 4g_s) - 32\sqrt{\eta}n \geq (2e(F') + 2g_o + 4g_s) - 34\sqrt{\eta}n$$

edges to H. □

Recall that a leaf ℓ_i in F' is called open for a_1 if it is the leaf with the highest index in its star or if it is the leaf ℓ_i' which is paired with this ℓ_i. It is called open for a_2 if it is a leaf ℓ_i' with the highest index in its star or the leaf paired with this ℓ_i' (unless this leaf ℓ_i is also open for a_1, in which case the leaf of the highest index in the component containing ℓ_{i+1} is also considered open for a_2).

We now connect a_1, a_2 (with $a_i \in A$) to their open leaves and to the vertices of components in $F \setminus F'$ in such a way that two vertices on each cycle and all vertices not on cycles are adjacent to both a_1 and a_2, and other vertices are adjacent to only one of these vertices. Moreover, we have already placed hub gadgets into cycles of $F \setminus F'$ (see Figure 6.3).

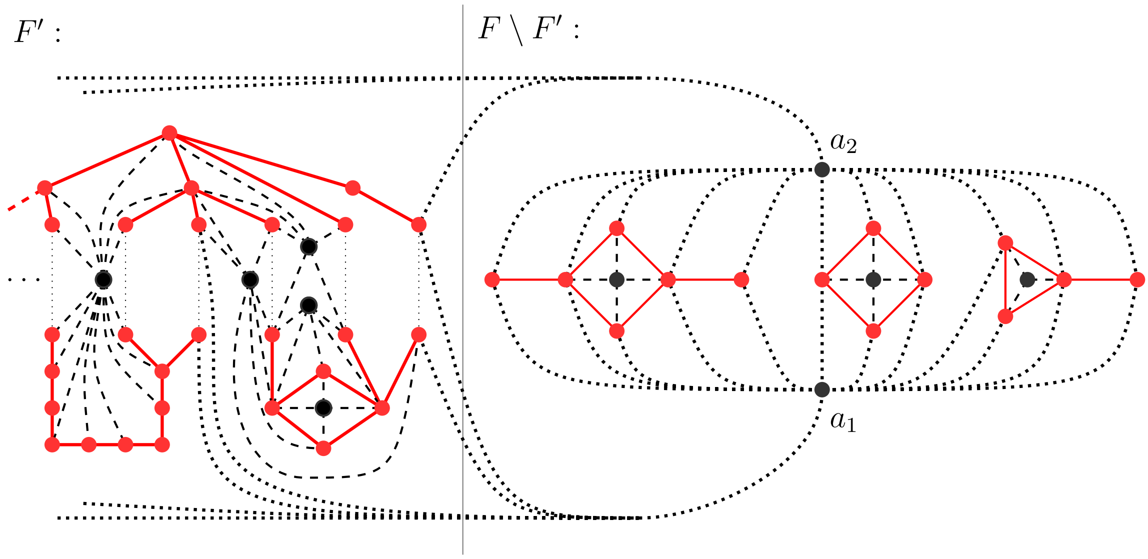

Figure 6.3: Connecting a_i to free leaves in F' and taking care of the components with at most two leaves in $F \setminus F'$.

Claim 6.20 *This adds another*

$$2|F \setminus F'| + 2c(F) - 8\eta n - \sqrt{\eta} n \tag{6.13}$$

edges to H, where $c(F)$ is the number of components in F.

Proof. Let us assume that in the graph G both a_i were adjacent to all vertices of F and that every cycle in $F \setminus F'$ obtained a hub which was adjacent to all vertices of the cycle. Then we obtain at least $2c(F')$ edges between the a_i and open leaves in F' as there are two open leaves for every component in F'. Moreover we obtain at least $2|F \setminus F'| + 2c(F \setminus F')$ edges between $F \setminus F'$ and $\{a_1, a_2\}$ and the hub gadgets: Indeed, two vertices of each component of $F \setminus F'$ are adjacent to a_1, a_2 and a hub gadget, while the rest are adjacent to exactly 2 of these (i.e., the rest are adjacent to either a_1 and a_2 or to one a_i and one hub gadget).

In total we get at least $2|F \setminus F'| + 2c(F') + 2c(F \setminus F')$ edges if our assumption holds. However, some edges between the a_i and F and some edges between the hub gadgets and $F \setminus F'$ are not present. Thus we lose at most $4\eta n$ edges from each a_i (as $\deg(a_i, B \cup D) \geq (1/2 - 2\eta)n \geq |B \cup D| - 4\eta n$) and at most $\sqrt{\eta} n$ edges from the hub gadgets (as each potential hub gadget is adjacent to a $(1-\sqrt{\eta})$-fraction of all vertices on its cycle). $\square$

After the assignment of gadgets there might be vertices left in $A \cup C$. Those will be treated with the help of b_1 and b_2. We simply connect each b_i to as many of the left-over vertices as possible. There are at least $(1/2 - 2\eta)n - g_o - 2g_s - g_h$ left-over vertices in $A \cup C$ and each b_i is adjacent to all but $2\eta n$ of them. Thus we have at least

$$2\left(n/2 - 2\eta n - g_o - 2g_s - g_h\right) - 4\eta n \tag{6.14}$$

edges incident to these vertices.

Our construction of the graph H is now complete. We bound the number of edges in H by summing up the contributions from (6.12)–(6.14).

$$e(H) \geq n + 2e(F) + 2c(F) - 2g_h - 16\eta n - 35\sqrt{\eta}n \geq n + 2e(F) - 51\sqrt{\eta}n \tag{6.15}$$

where the inequality is due to $g_h \leq c(F)$.

Recall that we have assumed that F was a benign jellyfish cover of $G[B \cup D]$. Such a jellyfish cover need not exist in general. However, Lemma 6.21 says that we find a jellyfish cover F of $G[B \cup D]$ and a subgraph $F^* \subseteq F$ with $e(F^*) \geq e(F) - 2\eta n$ such that F^* is benign. We apply our construction for H to this graph F^* and obtain a planar graph $H^* \subseteq G$ with

$$\mathrm{pl}(n, n/2+1) \geq e(H^*) \geq e(H) - 2\eta^{1/12}n \geq 2.5n - 3\eta^{1/12}n\,.$$

Here the last inequality follows from (6.15) and the fact that $e(F) \geq |B \cup D| \geq n/2$. $\square$

Except for the case $|B \cup D| = n/2$ and the proof of Lemma 6.21 the proof of Theorem 6.3 is complete.

Lemma 6.21
For every $\eta > 0$ there is n_0 such that the following is true for every graph G on $n \geq n_0$ vertices. Let $A \,\dot\cup\, B \,\dot\cup\, C \,\dot\cup\, D$ be a working partition of G and assume that $G[B \cup D]$ has minimum degree two. Then there is a subgraph $F \subseteq G[B \cup D]$, whose components are classified as Type I or Type II, with the following properties:

- *(a) the components of F are either trees or jellyfish,*
- *(b) $e(F) \geq |B \cup D| - 2\eta^{1/12}n$,*
- *(c) every Type I component in F is a tree which can be rooted such that the root z satisfies $\deg(z, A \cup C) \geq \frac{1}{2}|A \cup C|$ and every vertex $x \neq z$ in the tree satisfies $\deg(x, A \cup C) \geq (1 - \eta^{1/5})|A \cup C|$, and*
- *(d) every vertex y in a Type II component in F satisfies $\deg(y, A) \geq (1 - \eta^{1/3})|A|$,*

(e) $\Delta(F) \leq |B \cup D|/2$.

Proof. We first give a rough outline of the proof.

It follows from $\delta(G[B \cup D]) \geq 2$ that $B \cup D$ can be partitioned into sets with spanning jellyfish in $G[B \cup D]$. However, we will want to choose such a partition with certain useful properties. To this end, given a graph H we define the *degree sequence* of H to be the (unique) decreasing sequence of integers $(d_1, d_2, \ldots, d_{|H|})$ such that there is a bijection $f : V(H) \to [|H|]$ with $\deg_H(x) = d_{f(x)}$ for all $x \in V(H)$. Suppose we have a second graph H' on the same vertex set with degree sequence $d'_1, \ldots, d'_{|H|}$. We say that H is *lexicographically smaller* than H' if there is $j \in [|H|]$ with $d_i = d'_i$ for all $1 \leq i < j$ and $d_j < d'_j$. This gives us a partial order on partitions into jellyfish (two partitions are non-comparable if they have the same degree sequence). Now let F be a lexicographically minimal spanning fluther of jellyfish in $G[B \cup D]$.

We call a vertex of F *bad* if it has degree at least $\eta^{-1/4}$ in F. For a vertex z, we denote by $F(z)$ the tree induced by z and all vertices x of F which do not lie on a cycle and for which the unique path from x to a cycle of F passes through z. We partition $V(F) = X \cupdot Y \cupdot Z$ by letting Z be the set of bad vertices and setting $X := \bigcup_{z \in Z} F(z) \setminus Z$ and $Y := V(F) \setminus (X \cup Z)$.

We now prove that F, X, Y and Z satisfy certain properties.

Claim 6.22 *For any $0 < c < 1$, there are at most $2cn$ vertices of degree at least $1/c$ in F.*

Proof. Since $e(F) = |F| \leq (1/2 + 2\eta)n$, the number of vertices of degree at least $1/c$ in F is at most $\frac{2e(F)}{1/c} \leq 2c(1/2 + 2\eta)n \leq 2cn$. □

Note in particular that this implies that $|Z| \leq 2\eta^{1/4}n$.

Claim 6.23 $\Delta(F) \leq \lfloor |F|/2 \rfloor + 1$.

Proof. There can certainly not be two vertices whose degree sum is at least $2\lfloor |F|/2 \rfloor + 3$, since then $e(F) \geq 2\lfloor |F|/2 \rfloor + 3 - 1 > |F|$ which is a contradiction.

Now suppose there is one vertex x of degree at least $\lfloor |F|/2 \rfloor + 2$. Since F is a jellyfish, there is at most one edge between neighbours of x, and if all of the at least $\lfloor |F|/2 \rfloor$ remaining neighbours were not leaves, F would contain at least $2\lfloor |F|/2 \rfloor + 3 > |F|$ edges, a contradiction. Thus x has at least one neighbour y in F which is a leaf. But since $\delta(G[B \cup D]) \geq 2$, the vertex y has at least one other neighbour z in $V(F)$ whose degree in F is at most $\lfloor |F|/2 \rfloor$. Adding the edge yz and deleting the edge xy from F gives a new jellyfish F' in which the maximum degree is smaller than that of F, and which is therefore lexicographically smaller than F, contradicting the choice of F. □

Claim 6.24 *(I) For any $x \in X$ we have* $\deg_G(x, A \cup C) \geq (1 - \eta^{1/5})|A \cup C|$.

(II) For all but at most $\eta^{1/3}n$ vertices $y \in Y$ we have $\deg_G(y, A) \geq (1 - \eta^{1/3})|A|$.

Proof. Suppose (I) is false, so we have $x \in X$ with at most $(1 - \eta^{1/5})|A \cup C| \leq (1/2 - \eta^{1/5}/2)n$ neighbours in $|A \cup C|$. Then x has at least $\eta^{1/5}n/2$ neighbours in $B \cup D$. However, by Claim 6.22 there are at most $4\eta^{1/4}n$ vertices of degree at least $\eta^{-1/4}/2$ in F, and so x must be adjacent to a vertex y of degree less than $\eta^{-1/4}/2$ with $xy \notin E(F)$. However, x is also in some tree $F(z)$ for a bad vertex z. If we add the edge xy and delete an edge incident to z which is not in $F(z)$, then the degree of the bad vertex z has decreased, but no new vertex has become bad. Furthermore, we have a new jellyfish, or possibly two new jellyfish, in place of the old jellyfish, and so we have constructed a partition into jellyfish which is lexicographically smaller than F, contradicting the choice of F.

Now suppose (II) is false. Then we have

$$\begin{aligned} e(B \cup D, A) &\leq (|B \cup D| - \eta^{1/3}n)|A| + \eta^{1/3}n(1 - \eta^{1/3})|A| \\ &= |B \cup D||A| - \eta^{2/3}|A|n \\ &\leq (1/2 + 2\eta - \eta^{2/3})|A|n \\ &< (1/2 - 2\eta)|A|n, \end{aligned}$$

contradicting property (2) of Lemma 6.7. □

For each vertex in Z we now delete the incident edges which lie on a path between the bad vertex and the cycle of the jellyfish containing it. Note that for each bad vertex we delete only one edge if it does not lie on the cycle, or if it does lie on the cycle, we delete the two incident edges on the cycle. Thus we delete at most $4\eta^{1/4}n$ edges in total (by Claim 6.22).

The (additional) components we obtain by deleting edges satisfy condition (c) of the lemma: Every component has exactly one vertex from Z (with $\deg(z, A \cup C) \geq n/4 \geq \frac{1}{2}|A \cup C|$ by the properties of the working partition) and all other vertices x lie in X and thus have $\deg(x, A \cup C) \geq (1 - \eta^{1/5})|A \cup C|$ by property (I) of Claim 6.24.

Furthermore, for each vertex $y \in Y$ with $\deg_G(y, A) < (1 - \eta^{1/3})|A|$ we delete y from F together with all incident edges. The remaining vertices in Y span components in F that are either trees or jellyfish and that satisfy condition (d) by property (II) of Claim 6.24. Since we delete only vertices which are not bad, and since there are at most $\eta^{1/3}n$ of them, we are deleting at most $\eta^{-1/4}\eta^{1/3}n = \eta^{1/12}n$ edges in this step. In total we have deleted at most $4\eta^{1/4}n + \eta^{1/12}n \leq 2\eta^{1/12}n$ edges and all components obtained are either trees or jellyfish hence property (a) and (b) of the lemma hold.

Note that originally we may have had $\Delta(F) = \lfloor |B \cup D|/2 \rfloor + 1$, but a vertex of this degree would be bad and would have had at least one incident edge removed. Hence also condition (e) holds. □

6.2.1 The case of $|A \cup C| = |B \cup D|$

We have one final case left to consider, namely when $|A \cup C| = |B \cup D|$. We begin by splitting up the graph into more easily manageable parts. To this end, we say that a graph G on n vertices has property $\mathcal{P}_{\text{split}}$ with parameter η if there is a partition $V(G) = V_1 \cupdot V_2$ such that the following conditions hold:

(i) $|V_1| \geq \eta n$;

(ii) By adding at most $1/\eta^2$ edges to $G[V_2]$ we obtain the graph G' with $\delta(G') \geq |V_2|/2 + 1$;

(iii) There is a planar subgraph $H \subseteq G[V_1]$ with $e(H) \geq (2.5 - \eta)|V_1|$.

Essentially, property $\mathcal{P}_{\text{split}}$ says that we can split off a substantial proportion of the graph, V_1, and cover it with a planar subgraph of approximately the right size, while the remainder still (almost) satisfies the appropriate minimum degree condition.

The property $\mathcal{P}_{\text{split}}$ is the key to the case $|A \cup C| = |B \cup D|$. By the minimum degree condition, we know that both $G[A \cup C]$ and $G[B \cup D]$ have minimum degree at least one. We call a component in one of these graphs which has at least $1/\eta$ vertices a *large component*, and all others are *small components*. The union of large components in $G[B \cup D]$ is denoted L_1, and the small components by S_1. Similarly L_2 and S_2 denote the unions of large and small components in $G[A \cup C]$. Without loss of generality we assume that $|S_1| \leq |S_2|$.

Now let $S_2' \subseteq S_2$ be a subset of size $|S_1|$, chosen in such a way that all small components in S_2 except at most one lie either completely in S_2' or completely in $S_2 \setminus S_2'$.

The case $|A \cup C| = |B \cup D|$ is immediate from the following three propositions.

Proposition 6.25
If $|S_1 \cup S_2'| < \eta n$ *then* $\mathrm{pl}(G) \geq 2.5n - 3\eta^{1/12} n$.

Proposition 6.26
If $|S_1 \cup S_2'| \geq \eta n$ *then* G *has property* $\mathcal{P}_{\textit{split}}$ *with parameter* η.

Proposition 6.27
If G *has property* $\mathcal{P}_{\textit{split}}$ *with parameter* η *it has* $\mathrm{pl}(G) \geq 2.5n - 3\eta^{1/12} n$.

It remains to prove Proposition 6.25, Proposition 6.26, and Proposition 6.27.

Proof of Proposition 6.25. We conclude from $|S_1| \leq (\eta/2)n$ that $G[B \cup D \setminus V(S_1)]$ contains a collection of jellyfish and large trees with at least $(1-\eta)(1-\eta)n/2 \geq (1-2\eta)n/2$ edges. We can therefore add ηn (virtual) edges to $G[B \cup D]$ to increase its minimum degree to at least 2. We then apply the construction from the case $|A \cup C| < |B \cup D|$ to find a planar subgraph with at least $2.5n - 3\eta^{1/12} n$ edges. (This is indeed possible since the cases only differ in that $|A \cup C| = |B \cup D|$ does not guarantee $\delta(G[B \cup D]) \geq 2$.) Removing the virtual edges again gives a planar subgraph H of G with

$$e(H) \geq 2.5n - 2\eta^{1/12} n - 43\sqrt{\eta} n - 2\eta n \geq 2.5n + 3\eta^{1/12} n$$

by (6.15). □

Proof of Proposition 6.26. We define $V_1 = S_1 \cup S_2'$, $V_2 = V(G) \setminus V_1$ and claim that the partition $V(G) = V_1 \cup V_2$ is a certificate for the property $\mathcal{P}_{\text{split}}$. To prove this we verify conditions (1)-(3).

Obviously $|V_1| \geq \eta n$ and hence property (1) is satisfied.

To derive property (2) we set $G' = G[V_2]$. Let x be a vertex of L_1. Then x has no neighbours in S_1, and so

$$\begin{aligned} d_{G'}(x) \geq d_G(x) - |S_2'| &\geq n/2 + 1 - |S_2'| \\ &= (|G'| + |S_1| + |S_2'|)/2 + 1 - |S_2'| \\ &= |G'|/2 + 1. \end{aligned}$$

A similar argument shows that vertices of L_2 also satisfy this minimum degree condition in G', as do all vertices of $S_2 \setminus S_2'$ except possibily those in the one component of S_2 which was split between S_2' and $S_2 \setminus S_2'$. However, there are at most $1/\eta$ such vertices, each having at most $1/\eta$ neighbours in S_2', and therefore satisfying $d_{G'}(x) \geq |G'|/2+1-1/\eta$. By adding at most $1/\eta^2$ edges to G', we can ensure that these vertices also satisfy the minimum degree condition, as required by property (2).

It remains to verify condition (3), i.e., that $G[V_1]$ contains a planar subgraph H with $e(H) \geq (2.5 - \eta)|V_1|$. For this, we observe that a vertex in S_1 has at most $1/\eta$ neighbours in $B \cup D$, and therefore has at least $n/2 + 1 - 1/\eta$ neighbours in $A \cup C$ and in particular at least $|S_2'| - 1/\eta$ neighbours in S_2'. Similarly, a vertex in S_2' has at least $|S_1| - 1/\eta$ neighbours in S_1.

In order to construct a planar subgraph H of $G[V_1]$ with close to $2|V_1| + e(S_1 \cup S_2) \geq 2.5|V_1|$ edges we pair up the vertices of S_1 with the vertices of S_2' in the following fashion.

First we fix a planar embedding for each component. For each component we choose an arbitrary vertex and start a walk around the outer face of the component in clockwise direction. The vertices within the component are then ordered according to the order in which they appear on this walk. (One could also think of a depth first search where we always choose vertices from left to right.) The components of S_2' are ordered in the same way with the sole difference that the walk around a component has counter-clockwise direction. We then pair up the i-th vertex in S_1 with the i-th vertex in S_2'.

We now bring the components of S_1 into an arbitrary order. The components of S_2' are arranged such that a maximum number of components in S_1 are such that all vertices of the component C are adjacent to all vertices of components in S_2' that contain vertices paired with vertices of C. Since each component has less than $1/\eta$ vertices and every vertex of S_1 is adjacent to all but $1/\eta$ vertices of S_2' such a pairing differs in at most $1/\eta^3$ edges from one that contains all edges between components that share vertices of a pair. We assume the latter in the following and denote the i-th vertex of S_1 by v_i and the i-th vertex of S_2' by w_i.

We finally construct a planar subgraph H of $G[V_1]$ by adding all edges of the form $v_i w_i$ and $v_{i+1} w_i$. Moreover, we greedily add edges between S_1 and S_2' to all faces except for the outer face (see Figure 6.4). Finally we connect v_1 to all vertices of S_2' that lie on the outer face and we connect $w_{|S_2'|}$ to all vertices of S_1 that lie on the outer face. Note that v_1 and $w_{|S_2'|}$ might not be adjacent to all of those vertices but they are adjacent to all but $1/\eta$ of those vertices for sure. (In this step we might loose another $2/\eta$ edges that we already inserted between the components.)

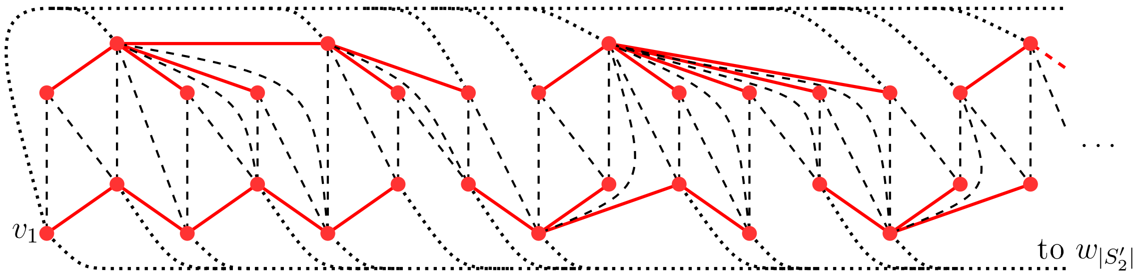

Figure 6.4: Small paired components and a maximal (bipartite) planar subgraph; the dotted edges are those toward v_1 and $w_{|S_2'|}$ respectively.

Summing up we obtain a planar subgraph $H \subseteq G[V_1]$ with at least

$$e(S_1) + e(S_2') + 2|V_1| - 4 - 1/\eta^3 - 4/\eta \geq (2.5 - \eta)|V_1|$$

edges thus showing property (3). Hence V_1 and V_2 satisfy all condidtions of property $\mathcal{P}_{\text{split}}$. $\square$

Proof of Proposition 6.27. Recall that G is a graph on $n \geq n_0/\eta$ vertices. As G satisfies property $\mathcal{P}_{\text{split}}$, we split it accordingly into V_1 and V_2, where V_1 can be covered by a planar subgraph with at least $(2.5 - \eta)|V_1|$ edges. Furthermore, we add at most $1/\eta^2$ edges to $G[V_2]$ to ensure that $\delta(G[V_2]) \geq |V_2|/2 + 1$ holds. We then continue to work with $G[V_2]$. It follows from Proposition 6.25 and Proposition 6.26 that either $\mathrm{pl}(G[V_2]) \geq (2.5 - 3\eta^{1/12})|V_2|$ or $G[V_2]$ again has property $\mathcal{P}_{\text{split}}$ with parameter η. If this is the case, we split off another subgraph on vertex set V_1', set $V_2 := V_2 \setminus V_1'$ and add at most $1/\eta^2$ edges to $G[V_2]$ to maintain a minimum degree of $|V_2|/2 + 1$. We continue this procedure until one of the following two cases holds.

Case 1: $|V_2| < n_0$: In this case, we simply take the planar subgraphs of the sets which were split off, which together contain at least $(2.5 - \eta)(n - n_0) = (2.5 - \eta)(1 - \eta)n \geq (2.5 - 4\eta)n$ edges.

Case 2: $\mathrm{pl}(G[V_2]) \geq (2.5 - 3\eta^{1/12})|V_2|$: In this case we take a planar subgraph H_2 of $G[V_2]$ with $e(H_2) \geq (2.5 - 3\eta^{1/12})|V_2|$, which together with the planar graphs on the sets which were split off, contains at least $(2.5n - 3\eta^{1/12})|V_2| + (2.5 - \eta)(n - |V_2|) \geq 2.5 - 3\eta^{1/12}n + 1/\eta^4$ edges.

We now need to delete those edges which we added in during the splitting steps. In each step we added at most $1/\eta^2$ edges. Furthermore, in each step, the size of the remaining graph decreased by a factor of at least $(1-\eta)$. We therefore required at most $\ln(\eta)/\ln(1-\eta) \leq 1/\eta^2$ steps, meaning we added at most $1/\eta^4$ edges. Thus deleting these edges again, we obtain a planar subgraph of G with at least $(2.5 - 3\eta^{1/12})n$ edges. $\square$

6.3 Proof of Theorem 6.2

Before we prove Theorem 6.2 we introduce one additional definition. A star forest F is a cycle-free graph with $\delta(F) \geq 1$ and $P_3 \not\subseteq F$. We will call a spanning star forest $F \subseteq G$ a *star cover* of G.

Proposition 6.28
Every graph G with $\delta(G) \geq 1$ has a star cover.

Proof. Let $\delta(G) \geq 1$ and let F be a vertex maximal star forest in G. Assume there is $v \in V(G) \setminus V(F)$. But then v must be adjacent to at least one leaf w in F and the star of w must have more than two vertices. We remove w from its current star and add the edge vw to F which gives a star forest with more vertices. □

Let G be a graph that satisfies the conditions of Theorem 6.2. In five steps we construct a planar subgraph H of G that has many edges.

Partitioning G: We partition $V(G)$ into $A \cup B \cup C \cup D$ with the help of Lemma 6.7 where $a = 2.25$.

Finding a suitable planar subgraph F: We have $|B \cup D| > |A \cup C|$.[4] Hence $\delta(G[B \cup D]) \geq 1$ and $G[B \cup D]$ has a star cover F. In particular, $e(F) \geq n/4$.

Pairing the leaves of F: We pair up the leaves of F in such a way that F plus the matching spanned by those pairs gives a planar graph. (Note that the pairs do not necessarily form edges in G, so the matching is imagined and will not be part of the planar subgraph of G which we construct.)

Attaching gadgets: We combine F with edges from the dense pairs $(A, B \cup D)$ and $(B, A \cup C)$ to obtain a planar graph H with reasonably high average degree.

Taking care of left-over vertices: There may be vertices in $A \cup C$ that have not been used yet. We put those into a planar graph with average degree 4.

We now give the details of the proof of Theorem 6.2, i.e., we show that $\mathrm{pl}(2m - 1, m) \geq (4.5 + o(1))m$.

Proof of Theorem 6.2. We will prove that for every $\eta > 0$ there is n_0 such that

$$\mathrm{pl}(n, \lceil n/2 \rceil) \geq (2.25 - 70\eta)n$$

for every odd integer $n \geq n_0$.

Let $\eta > 0$. Set n_0 as given by Lemma 6.7 and let $G = (V, E)$ be a graph on n vertices where $n \geq n_0$ is an odd integer. Assume that G does not have a planar subgraph with $(4.5 - 15\eta)n/2$ edges. Then G has a working partition $V = A \cup B \cup C \cup D$ by Lemma 6.7.

W.l.o.g. we have $|B \cup D| \geq n/2$ and $G[B \cup D]$ has a star cover F by Proposition 6.28. We call a star cover F *benign* if every star S in F satisfies at least one of the following conditions.

[4] In the odd case.

(b1) The center v of S has $\deg(v, A \cup C) \geq \frac{1}{2}|A \cup C|$ and every leaf w of S has $\deg(w, A \cup C)| \geq |A \cup C| - \eta n$, or

(b2) every vertex v in S has $\deg(v, A) \geq |A| - 2\eta n$.

We will later see that stars with at least $2/\eta$ many leaves satisfy property (b1) while stars with less than $2/\eta$ leaves (mostly) satisfy property (b2). We call the former *large stars* and the latter *small stars*. For the rest of the proof we will assume that $A \cup B \cup C \cup D$ is a working partition and that $G[B \cup D]$ contains a benign star cover F. Later, in Lemma 6.29, we will show that we indeed find a star cover F of $G[B \cup D]$ and an almost spanning star cover $F^* \subseteq F$ with $e(F^*) \geq e(F) - 2\eta n$ such that F^* satisfies properties (b1) and (b2).

We now extend F to a planar subgraph with many edges. First we set aside two leaves b_1, b_2 of F that lie in B. In doing so we lose two edges, which will not affect calculations significantly. Moreover, we separate all stars with less than three vertices, i.e., which are isolated edges. Those will be dealt with separately. Let $F' \subseteq F$ be the set of all stars with at least three vertices.

We now define the notion of a *socket*. The i-th socket or the socket for the pair (ℓ_i, ℓ_i') is the following set of vertices: $\{\ell_i^-, c(\ell_i), \ell_i, (\ell_i')^-, c(\ell_i'), \ell_i'\}$. Here $c(\ell)$ is the center of the star that contains the leaf ℓ. Moreover, ℓ_i^- is defined to be ℓ_{i-1} if ℓ_{i-1} and ℓ_i lie in the same star, or the leaf of highest index in the star containing ℓ_i otherwise. We define $(\ell_i')^-$ similarly. Now every pair of leaves has its unique socket (see Figure 6.5.

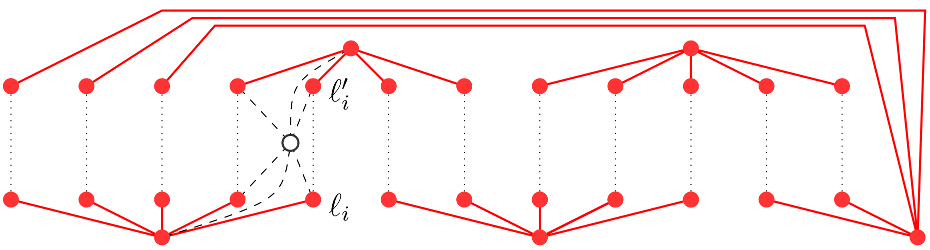

Figure 6.5: A pairing of an unbalanced partition; the dashed lines from the black circle end at vertices of the i-th socket.

A *gadget* for a socket consists of one or two vertices from $A \cup C$ that are adjacent to the vertices of the socket. We distinguish between an *ordinary gadget* which is one vertex that is adjacent to all six vertices of a socket and a *split gadget* which consists of two vertices which, in the case of the i-th socket, are adjacent to $\{\ell_i^-, c(\ell_i), \ell_i, (\ell_i')^-, \ell_i'\}$ and $\{(\ell_i')^-, c(\ell_i'), \ell_i'\}$ (or $\{\ell_i^-, \ell_i, (\ell_i')^-, c(\ell_i'), \ell_i'\}$ and $\{\ell_i^-, c(\ell_i), \ell_i\}$) respectively (see Figure 6.6). We denote the number of ordinary gadgets by g_o and the number of split gadgets by g_s and claim that we can find gadgets for almost all sockets. Let k denote the number of leaves in F', and note that $k \leq |A \cup C| \leq n/2 + 2\eta n$.

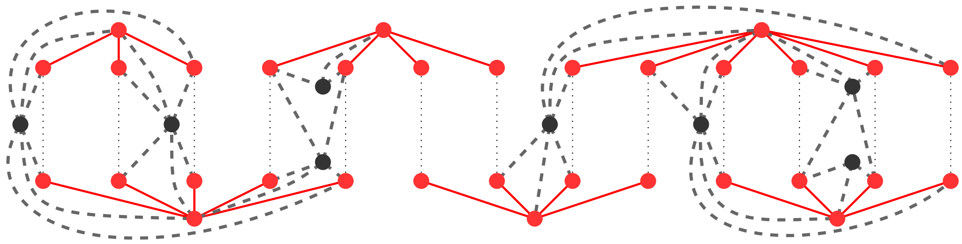

Figure 6.6: A pairing with (some) gadgets attached; dashed lines mark edges in the gadgets.

Claim *There is an assignment of gadgets to sockets that gives at least*

$$2k + 2g_o + 4g_s - 56\eta n \tag{6.16}$$

many edges towards F'.

Proof. We call a vertex a *potential half gadget* for a side of the i-th pair if it is adjacent to $P_i := \{\ell_i^-, c(\ell_i), \ell_i\}$ or to $P_i' := \{(\ell_i')^-, c(\ell_i'), \ell_i'\}$ respectively. It follows from property (b1) or (b2) that each side of the i-th pair has at least $(1/4 - 4\eta)n$ potential half gadgets. Moreover, we can restrict the choice of potential half gadgets for small stars to the set A and still have at least $(1/4 - 4\eta)n$ many of them. If a vertex is a potential half gadget for both sides of the i-th pair it is an ordinary gadget for the i-th pair. If x is adjacent to P_i and x' is adjacent to P_i' these two form a split gadget for the i-th pair if either x is adjacent to $(\ell_i')^-$ and ℓ_i' or x' is adjacent to ℓ_i^- and ℓ_i. This is the case for most potential half gadgets if the vertices on one side belong to a large star as leaves of large stars are adjacent to all but ηn vertices in $A \cup C$. If the leaves on both sides belong to small stars we choose the gadgets from A by definition. But then each leaf is adjacent to all but ηn many vertices in A. In both cases we find $(1/4 - 6\eta)n$ many potential half gadgets for each side that guarantee the existence of the two additional edges. We now find an assignment of gadgets to sockets that leaves few sockets empty.

In a first step we greedily fill sockets with ordinary gadgets. If no more ordinary gadgets can be found the sets of potential half gadgets for all remaining pairs are disjoint. We had $k/2$ many pairs in the beginning and that we have found g_o many ordinary gadgets. Thus $k/2 - g_o$ pairs (with $k/2 \le (1/4 + \eta)n$) remain and each leaf in those pairs has at least $(1/4 - 6\eta)n - g_o \ge (k/2 - g_o) - 7\eta n$ many potential half gadgets.

Thus we can apply Proposition 6.17 to an appropriate auxiliary graph to find an assignment of potential half gadgets to the sockets that leaves at most $14\eta n$ sockets (half) open. All other $g_s \ge k/2 - g_o - 14\eta n$ sockets receive split gadgets. This uses another $2g_s$ many vertices from $A \cup C$. All in all we have found

$$6g_o + 8g_s \ge 4\left(\tfrac{1}{2}k - 14\eta n\right) + 2g_o + 4g_s$$

edges from gadgets towards F'. □

In the next step we connect the vertices a_1, a_2 to *open leaves* of F. Here a leaf is called open for a_1 if it is a leaf ℓ_i with the highest index in its star or the leaf paired with this ℓ_i. It is called open for a_2 if it is a leaf ℓ'_i with the highest index in its star or the leaf paired with this ℓ'_i. The only exception to this scheme appears when a pair of leaves is open for both a_1 and a_2. Instead of adding all four edges to this pair we only add three of them. The fourth edge (from a_1) instead goes to the leaf with the highest index in the star containing ℓ_{i+1}, where the leaves ℓ_i are ordered cyclically. Indeed, there is a planar embedding of the star forest with its gadgets and all edges from a_i to their respective open leaves. This gives another two edges per star (see Figure 6.7).

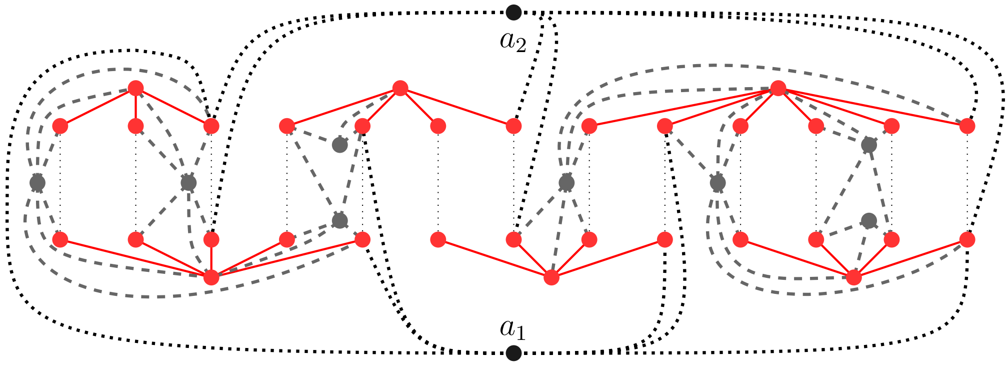

Figure 6.7: Connecting up two leaves per star to a_1 and/or a_2.

In case (2) this construction collides with the edges of the star that that has leaves on both sides. However, we only lose two edges in total because of this collision – one for each a_i.

Similarly we connect both a_i to all their neighbours among the vertices of stars with only two vertices (which we removed in the beginning). Together this adds another

$$2(|F \setminus F'| + (|F'| - k) - 4\eta n) \geq n - 2k - 8\eta n \tag{6.17}$$

edges to H as $\deg(a_i, B \cup D) \geq n/2 - 2\eta n \geq |B \cup D| - 4\eta n$.

After the assignment of gadgets there might be vertices left in $A \cup C$. Those will be treated with the help of b_1 and b_2. We simply connect each b_i to as many of the left over vertices as possible. There are at least $(1/2 - \eta)n - g_o - 2g_s$ left over vertices and the b_i have at least

$$n - 4\eta n - 2g_o - 4g_s \tag{6.18}$$

edges towards these vertices.

For the cases (1) and (2) the construction of H is complete at this point. Note that H is indeed planar. We sum up the terms (6.16), (6.17), and (6.18) and obtain

$$e(H) \geq e(F) + 2n - 68\eta n\,. \tag{6.19}$$

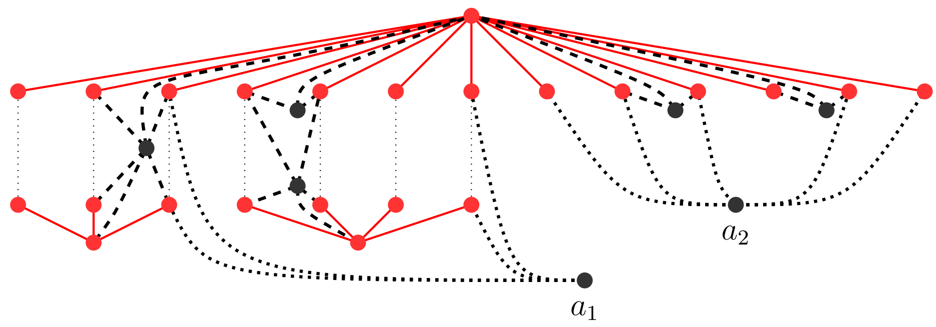

Figure 6.8: A pairing and gadgets for a giant star.

If we are in case (3) there is one giant star in F that has more than half of the total leaves in F. Let us assume that this star has even more leaves than its center has neighbours in $A \cup C$. Since F was a minimal star cover it follows that each leave did not have any other neighbour in $A \cup C$. Hence we have $|A \cup C| = |B \cup D| + 1$. We move the center of the giant star to $B \cup D$ and have $|B \cup D| > |A \cup D|$ afterwards. Hence we can exchange the rôles of $A \cup C$ and $B \cup D$ and again find a minimal star cover. It could be that this star cover again has a giant star that has more leaves than its center has neighbours in $A \cup C$. Then we once more move the center to $B \cup D$. It is easy to see that (after exchanging the rôles of $A \cup C$ and $B \cup D$ once more) there is no giant star with more leaves then center neighbours. We now pair up all other stars against the giant star. The resulting pairs are treated as in the previous cases, i.e., we find ordinary/split gadgets for those pairs. For the half-sockets of the giant star that did not receive a pairing we take a different approach, seeking only one half of a split gadget (see Figure 6.8). Finally we connect a vertex from A to all leaves that did not get paired up.

We now calculate how many edges this construction yields. The giant star has at most d half-sockets, where $d = d_{A \cup C}(x)$ and x is the center of the giant star, and at most $4\eta n$ of the neighbours of x in $A \cup C$ are not neighbours of some pair of leaves of the giant star. On the other hand, the stars in Q contain at most $|B \cup D|/2 \leq n/4 + \eta n$ half-sockets, each with at least $|A| - 6\eta n \geq n/4 - 6\eta n$ potential half-gadgets. Furthermore, at most $4\eta n$ pairs of potential half-gadgets for each (half-)socket do not admit the required additional two edges for a split gadget. Thus we can certainly find split gadgets (and half-gadgets for the unpaired leaves of the giant star) covering all but at most $15\eta n$ sockets and half sockets. Suppose that we use g_o ordinary gadgets, g_s split gadgets and g_h half-gadgets. Then the number of edges in gadgets is $6g_o + 8g_s + 3g_h$. Furthermore, from a_1 we obtain at least $2c(Q)$ edges, where $c(Q)$ is the number of stars in Q, and from a_2 we obtain at least g_h edges. Finally, we put the remaining $|A \cup C| - (g_o + 2g_s + g_h)$ vertices into a bipartite graph with b_1 and b_2 containing at least $2(|A \cup C| - (g_o + 2g_s + g_h)) - 4\eta n$ edges. Summing up all terms and including the edges from F we obtain a planar graph H with at least

$$\begin{aligned} e(F) + 2|A \cup C| + 4g_o + 4g_s + 2g_h + 2c(Q) - 4\eta n \\ &\geq e(F) + |A \cup C| + 2(\ell(F) - 30\eta n) + 2c(Q) - 4\eta n \\ &\geq e(F) + 2|A \cup C| + 2(|F| - 1) - 64\eta n \\ &\geq e(F) + 2n - 65\eta n \end{aligned}$$

edges. We conclude that also in case (3) we have

$$e(H) \geq e(F) + 2n - 65\eta n\,. \tag{6.20}$$

Recall that we assumed that F was a benign star cover of $G[B \cup D]$. Such a star cover need not exist in general. However Lemma 6.29 says that we find a star cover F of $G[B \cup D]$ and a subgraph $F^* \subseteq F$ with $e(F^*) \geq e(F) - 2\eta n$. We apply our construction for H to this graph F^* and obtain a planar graph $H^* \subseteq G$ with

$$\mathrm{pl}(n, \lceil n/2 \rceil) \geq e(H^*) \geq e(H) - 2\eta n \geq 2.25n - 70\eta n\,.$$

Here the last inequality follows from (6.19) and (6.20) and the fact that $e(F) \geq |B \cup D|/2 \geq n/4$. □

At this point the proof of Theorem 6.2 is complete except for the assumption that the star cover F is a benign star cover. The remainder of this section proves that we can indeed assume that (b1) or (b2) holds for a star cover F of $G[B \cup D]$. More precisely we show that $G[B \cup D]$ with $\delta(G[B \cup D]) \geq 1$ has a star cover F such that an almost spanning subgraph $F' \subseteq F$ satisfies (b1) or (b2) for each component.

Recall that a star is called large if it has at least $2/\eta$ many leaves. Else it is called *small*.

Lemma 6.29
Let $\eta > 0$ be given. Then there is an n_0 such that the set of vertices of every graph $G = (V, E)$ with $|V| = n \geq n_0$, $\delta(G) > n/2$ and $\mathrm{pl}(G) < (a - 15\eta)n$ can be partitioned into sets $V = A \cupdot B \cupdot C \cupdot D$ such that properties (1)–(4) of Lemma 6.7 (with $a = 2.25$) hold. Then there exist subsets $B' \subseteq B$ and $D' \subseteq D$ with the following properties.

$$|(B \cup D) \setminus (B' \cup D')| \leq 2\eta n$$

and $G[B' \cup D']$ has a star cover F^* such that

(b1) every center v of a large star in F^* has $\deg(v, A \cup C) \geq \frac{1}{2}|A \cup C|$ and every leaf w has $\deg(w, A \cup C) \geq |A \cup C| - \eta n$,

(b2) every vertex in a small star in F^* has $\deg(v, A) \geq |A| - \eta n$,

(b3) $e(F^*) \geq e(F) - 2\eta n$ for some star cover F of $G[B \cup D]$.

Proof of Lemma 6.29. Let $\eta > 0$ be given and choose n_0 as in Lemma 6.7 with parameters $\eta^3/4$ and $a = 2.25$. Now let G have the properties required by the lemma. In particular assume $\mathrm{pl}(G) \leq (2.25 - 15\eta)n \leq (2.25 - \frac{15}{4}\eta^3)n$. Then G has a working partition $V = A \mathbin{\dot\cup} B \mathbin{\dot\cup} C \mathbin{\dot\cup} D$ with parameters $a = 2.25$ and $\eta^3/4$ by Lemma 6.7. We proceed in three steps to find subsets $B' \subseteq B$ and $D' \subseteq D$ with and a star cover F^* of $B' \cup D'$ with the promised properties.

Finding a suitable star cover F of $G[B \cup D]$. If follows from the minimum degree of G and $|B \cup D| \geq n/2$ that $G[B \cup D]$ has minimum degree at least 1. Thus there is a star cover F in $G[B \cup D]$ by Proposition 6.28. However, we will want to choose such a star cover with certain useful properties. To this end, given a graph H we define the *degree sequence* of H to be the (unique, non-monotone-) decreasing sequence of integers $(d_1, d_2, \dots, d_{|H|})$ such that there is a bijection $f : V(H) \to [|H|]$ with $\deg_H(x) = d_{f(x)}$ for all $x \in V(H)$. Suppose we have a second graph H' on the same vertex set with degree sequence $d'_1, \dots, d'_{|H|}$. We say that H is *lexicographically smaller* than H' if there is $j \in [|H|]$ with $d_i = d'_i$ for all $1 \leq i < j$ and $d_j < d'_j$. This gives us a partial order on star covers (two partitions are non-comparable if they have the same degree sequence). Now let F be a lexicographically minimal star cover of $G[B \cup D]$.

Fixing large stars. The aim of this step is to make sure that each large star satisfies property (b1). First note that every center v of a large star in F satisfies $\deg(v, A \cup C) \geq n/4 \geq \frac{1}{2}|A \cup C|$ by condition (1) of Lemma 6.7. So assume that there is a large star in F that has a leaf w with $\deg(w, A \cup C) < |A \cup C| - \eta n$. Then w must have more than ηn neighbours in $B \cup D$. At least one of those neighbours must be a center of a star with less than $2/\eta - 1$ leaves or a leaf itself. So we remove w from its current star and attach it to the small star or to the leaf. In the latter case we also remove the leaf from its star thus forming an isolated edge. The star forest obtained is lexicographically smaller than F which contradicts our choice of F.

Fixing small stars. In this step we remove all vertices from $B \cup D$ which lie in small stars that violate condition (b2), i.e., that contain a vertex v with $\deg(v, A) < |A| - \eta n$. Let $B' \cup D'$ be the set of remaining vertices. Obviously, $F^* := F[B' \cup D']$ is a star forest in which every star satisfies either (b1) or (b2). It remains to show that $|(B \cup D) \setminus (B' \cup D')| \leq 2\eta n$. So let M be the set of vertices $v \in B \cup D$ with $\deg(v, A) < |A| - \eta n$. Since every vertex $w \in A$ has $\deg(w, B \cup D) \geq (1/2 - \eta^3/2)n \geq |B \cup D| - \eta^3 n$ it follows that $|W| \leq \eta^2 n$. Therefore, we erase at most $(2/\eta)\eta^2 n = 2\eta n$ vertices in small stars and $|(B \cup D) \setminus (B' \cup D')| \leq 2\eta n$ follows.

As $|(B \cup D) \setminus (B' \cup D')| \leq 2\eta n$ and F^* is a spanning star cover in $B' \cup D'$ we conclude that condition (b3) holds as

$$e(F^*) \geq e(F) - 2\eta n \geq \left(\frac{1}{4} - 2\eta\right) n\,.$$

□

Lemma 6.29 says that while $G[B \cup D]$ may not contain a star cover which is benign, we can always choose an almost spanning subset $B' \cup D' \subseteq B \cup D$ that has a benign star cover. We then use the construction from the proof of Theorem 6.2 on the (almost spanning) partition $A \cup B' \cup C \cup D'$ and obtain nearly the same number of edges.

6.4 Concluding remarks

It would be interesting to obtain tight versions of Theorem 6.2 and Theorem 6.3 for all values of m. We believe that the following should be true.

Conjecture 6.30
For every $m \in \mathbb{N}$ we have

$$\begin{aligned}\mathrm{pl}(2m-1, m) &\geq 4.5m - 4\,,\\ \mathrm{pl}(2m, m+1) &\geq 5m - 4\end{aligned}$$

Note that the difference of 1 to the upper bounds of Theorem 6.1 is due to parity issues. Another direction of research would be the case $d = o(n)$. If follows from Theorem 6.1 that $\mathrm{pl}(n, 1) = \lceil n/2 \rceil$ and $\mathrm{pl}(n, 2) = n$ while $\mathrm{pl}(n, n^{o(1)}) = (1 + o(1))n$. Kühn, Osthus, and Taraz showed the following.

Theorem 6.31 (Theorem 1 of [74])
For every $0 < \varepsilon < 1$ there is n_0 such that

$$\mathrm{pl}(n, 1500\sqrt{n}/\varepsilon^2) \geq (2 - \varepsilon)n$$

for every $n \geq n_0$.

Their proof relies on the fact that a minimum degree of order $\Theta(\sqrt{n})$ forces the existence of many copies of C_4. At the same time graphs with girth larger than four do not contain planar subgraph with more than $5/3\,n$ vertices. We conjecture that that the existence of cycles of a given length governs the evolution of planarity.

Conjecture 6.32
For every k there exist constants $c(k), C(k)$ such that

$$\mathrm{pl}(n, C(k)n^{c(k)}) = \left(\frac{2k+1}{2k-1} + o(1)\right) n\,.$$

This $c(k)$ is the infimum over all exponents which force every n-graph with minimum degree $n^{c(k)}$ to contain a cycle of length at most $2k$.

Another direction of research is to generalise the parameter $\mathrm{pl}(n, d)$ from planar subgraphs to subgraphs from other monotone properties. One such question could be: "What is the the maximum number of edges a subgraph can have without containing a fixed minor?"

7 Large planar subgraphs in dense graphs (II)

7.1 Introduction

A way to reformulate typical questions in extremal graph theory is the following. Given a property $\mathcal{P}$ and an edge density (or minimum vertex degree, etc.), what is the 'largest' member of $\mathcal{P}$ which must be contained in an n-vertex graph G with the given density? For many problems in extremal graph theory, the property $\mathcal{P}$ is somewhat trivial (for example, in Turán's theorem, $\mathcal{P}$ is the set of cliques). However this is not always the case: for example, in the Erdős-Stone [38] theorem, $\mathcal{P}$ is the set of complete r-partite graphs, and the problem of determining the 'largest' complete r-partite subgraph remains active, with most recently results and generalisations due to Nikiforov [78]. In 2005, Kühn, Osthus and Taraz [74] suggested the study of the property $\mathcal{P}$ consisting of all planar graphs, which, while well-studied in other parts of graph theory, have received relatively little attention from extremal graph theorists.

A plane graph is a drawing of a graph in the plane with no crossing edges. A graph is called planar if it has a plane graph drawing. The planarity of a graph G is defined as the maximum number of edges in a planar subgraph of G. We denote the planarity of G by $\mathrm{pl}(G)$. Kühn, Osthus and Taraz [74] investigated the connection between the minimum degree $\delta(G)$ and planarity $\mathrm{pl}(G)$ of a graph G by studying the parameter

$$\mathrm{pl}(n, d) := \min\{\mathrm{pl}(G) : |G| = n, \delta(G) \geq d\}.$$

Among other results they proved the following theorem.

Theorem 7.1
For each $\gamma > 0$ there exists n_γ such that $\mathrm{pl}(n, (2/3 + \gamma)n) = 3n - 6$ for every integer $n \geq n_\gamma$.

This was later improved by Kühn and Osthus [71] to the following result with the optimal bound on the minimum degree.

Theorem 7.2
There exists n_2 such that $\mathrm{pl}(n, (2/3)n) = 3n - 6$ for every integer $n \geq n_2$.

More recently, Cooley, Łuczak, Taraz and Würfl [27] showed the following threshold behaviour of $\mathrm{pl}(n, d)$ at minimum degree $d = n/2$.

Theorem 7.3
For every $\mu > 0$ there exists n_μ such that, for every $n \geq n_\mu$, we have that

$$\mathrm{pl}(n, \lceil n/2 \rceil) \geq (2.25 - \mu)n \qquad \textit{for } n \textit{ odd,}$$

and

$$\mathrm{pl}(n, n/2 + 1) \geq (2.5 - \mu)n \qquad \textit{for } n \textit{ even.}$$

This indeed constitutes a threshold behaviour since $\mathrm{pl}(n, \lfloor n/2 \rfloor) \leq 2n - 4$ for all integers n as one can see from the class of complete bipartite graphs. For smaller values of d one does not observe such rapid changes in the planarity. Indeed, Kühn, Osthus, and Taraz [74] showed that $\mathrm{pl}(n, d)$ varies only by a constant term for the whole range of $d = \gamma n$ with $\gamma \in (0, 1/2)$.

Theorem 7.4
For each $\gamma > 0$ there is $C = C(\gamma)$ such that $\mathrm{pl}(n, \gamma n) \geq 2n - C$ for every integer n.

For $\gamma < 1/2$ this is optimal up to the value of the constant C. For $\gamma \geq 1/2$ the above statement trivially holds: a Hamilton cycle with chords from one vertex on the inner face and from another vertex on the outer face proves that every n-vertex graph with minimum degree at least $n/2$ has a planar subgraph with $2n - 4$ edges. So it is natural to ask whether there are values $\gamma < 1/2$ such that $C(\gamma) = 4$. We answer this in the affirmative as we determine the optimal value of $C(\gamma)$ for all $0 < \gamma < 1/2$.

Theorem 7.5
For every $\gamma \in (0, 1/2)$ there exists n_γ such that $\mathrm{pl}(n, \gamma n) = 2n - 4k$ for every $n \geq n_\gamma$, where $k \in \mathbb{N}$ is the unique integer such that $k \leq 1/(2\gamma) < k + 1$. Hence, $C(\gamma) = 4\lfloor 1/(2\gamma) \rfloor$ for $n \geq n_\gamma$.

Note that the constants are best possible for the given minimum degree condition: the graph consisting of k disjoint copies of $K_{t,t}$ has $2kt$ vertices, is t-regular, and has no planar subgraph with more than $4kt - 4k$ edges because $K_{t,t}$ has no planar subgraph with more than $4t - 4$ edges.

7.2 Tools and lemmas

Our main tools in the proof are variants of the Regularity Lemma [90] and the Blow-up Lemma [62]. In order to formulate the versions that we will use, we first introduce some terminology.

Let $G = (V, E)$ be a graph and let $\varepsilon, d \in (0, 1]$. For disjoint nonempty sets $U, W \subset V$, we denote by $e(U, W)$ the number of edges between U and W, and define the *density* of the pair (U, W) as $d(U, W) := e(U, W)/|U||W|$. A pair (U, W) is *ε-regular* if

$$|d(U', W') - d(U, W)| \leq \varepsilon$$

for all $U' \subset U$ and $W' \subset W$ with $|U'| \geq \varepsilon|U|$ and $|W'| \geq \varepsilon|W|$. If the pair (U, W) is ε-regular and has density at least d, then we say that (U, W) is (ε, d)-*regular*.

An ε-*regular partition* of $G = (V, E)$ is a partition $V_0 \mathbin{\dot\cup} V_1 \mathbin{\dot\cup} \ldots \mathbin{\dot\cup} V_r$ of V with $|V_0| \leq \varepsilon|V|$, $|V_i| = |V_j|$ for all $i, j \in [r] := \{1, \ldots, r\}$, and such that, for all but at most εr^2 pairs $(i, j) \in [r]^2$, the pair (V_i, V_j) is ε-regular.

We say that an ε-regular partition $V_0 \mathbin{\dot\cup} V_1 \mathbin{\dot\cup} \ldots \mathbin{\dot\cup} V_r$ of a graph G is an (ε, d)-*regular partition* if the following is true. For every $i \in [r]$ and every vertex $v \in V_i$, there are at most $(\varepsilon + d)n$ edges incident to v which are not contained in (ε, d)-regular pairs of the partition.

Given an (ε, d)-regular partition $V_0 \mathbin{\dot\cup} V_1 \mathbin{\dot\cup} \ldots \mathbin{\dot\cup} V_r$ of a graph G, we define a graph R, called the *reduced graph* of the partition of G, where $R = (V(R), E(R))$ has $V(R) = \{V_1, \ldots, V_r\}$ and $V_iV_j \in E(R)$ whenever (V_i, V_j) is an (ε, d)-regular pair. We will usually omit the partition, and simply say that G has (ε, d)-*reduced graph* R. We call the partition classes V_i with $i \in [r]$ *clusters* of G. Observe that our definition of the reduced graph R implies that, for $T \subset V(R)$, we can, for example, refer to the set $\bigcup T$, which is a subset of $V(G)$.

In our proof, we require the minimum degree form of the Regularity Lemma.

Lemma 7.6 (Regularity Lemma, minimum degree form)
For all positive ε, d and γ with $0 < \varepsilon < d < \gamma < 1$ there is r_1 such that every graph G on $n > r_1$ vertices with minimum degree $\delta(G) \geq \gamma n$ has an (ε, d)-reduced graph R on r vertices such that $r \leq r_1$ and $\delta(R) \geq (\gamma - d - \varepsilon)r$.

Lemma 7.6 is an easy consequence of the original Regularity Lemma of Szemerédi [90]. Its proof can be found, for example, in [74, Proposition 9].

Now we outline our proof strategy for Theorem 7.5. First, we apply Lemma 7.6 to a given n-vertex graph G with minimum degree at least γn and obtain the reduced r-vertex graph R whose minimum degree is almost as large as γr. Then we need to distinguish two cases.

If any component of R has less than $2\delta(R)$ vertices, then it contains a triangle. Using this triangle we will find a small triangulation T (that is, a plane graph whose every face is a triangle) in G, and Theorem 7.4 will guarantee a subgraph S of the rest of the graph $G - V(T)$ such that the disjoint union of S and T has at least $2n$ edges.

It follows that each component of R has at least $2\delta(R) > r/(k + 1)$ vertices, and thus R has at most k components. These components correspond to k well-connected subgraphs of G and cover almost all vertices of G. In each subgraph we will find a quadrangulation (a plane graph whose every face has four edges) which has a certain 'accepting' property that allows the few remaining vertices to be inserted. We conclude that there is a collection of at most k vertex-disjoint quadrangulations covering all the vertices of G. Since every quadrangulation on m vertices has $2m-4$ edges, the theorem follows.

As one can see from the above outline, our argument divides into two cases, depending on whether the reduced graph R has a small component or not. In each case we shall need some embedding results, which we now describe in detail.

When the reduced graph R does have a small component, we will need the following embedding result, an easy case of the Counting Lemma (see, for example, Theorem 2.1 in [69]).

Lemma 7.7
For each $d > 0$ and $s \in \mathbb{N}$ there exist $\varepsilon > 0$ and m_0 such that whenever $m \geq m_0$ the following holds. Let U, V, W be three pairwise disjoint vertex sets each of size m. Suppose that each pair forms an (ε, d)-regular pair in a graph G. Then G contains every 3-partite triangulation on s vertices.

In the case that R has no small components, we will construct quadrangulations. For this we shall use a version of the Blow-up Lemma. In order to state this result, we need a further definition. A pair of disjoint sets of vertices U and W in a graph G is called (ε, δ)-*super-regular* if it is ε-regular, each vertex $u \in U$ has at least $\delta|W|$ neighbours in W, and each $w \in W$ has at least $\delta|U|$ neighbours in U.

The original version of the Blow-up Lemma, due to Komlós, Sárközy and Szemerédi [62], showed that, for the purposes of embedding graphs of bounded degree, super-regular pairs behave like complete bipartite graphs. In our proof, we will need to embed (planar) graphs with growing degrees, which is generally a very difficult problem. Fortunately for us, planar graphs are examples of arrangeable graphs, for which a suitable extension of the Blow-up Lemma [62] has recently been proven by Böttcher, Kohayakawa, Taraz and Würfl.

Definition 7.8 (a-arrangeable)
Let a be an integer. An n-vertex graph is called a-arrangeable if its vertices can be ordered as $(x_1, \ldots, x_n)$ in such a way that

$$\Big| N\Big(N(x_i) \cap \{x_{i+1}, x_{i+2}, \ldots, x_n\} \Big) \cap \{x_1, x_2, \ldots, x_i\} \Big| \leq a$$

for each $1 \leq i \leq n$.

Here, for a set of vertices S, we denote by $N(S)$ the set of those vertices not in S that are adjacent to some vertex in S.

Chen and Schelp showed that planar graphs are 761-arrangeable [22]; Kierstead and Trotter [57] improved this to 10-arrangeable. Thus, the following theorem of Böttcher, Kohayakawa, Taraz and Würfl [16] can be used to embed planar graphs whose maximum degree is not too large.

Theorem 7.9 (Arrangeable Blow-up Lemma)
For all $a, \Delta_R, \kappa \in \mathbb{N}$ and for all $\delta > 0$ there exists $\varepsilon > 0$ such that for every integer r there is n_0 such that the following is true for every $n_1, \ldots, n_r$ with $n_0 \leq n = \sum n_i$ and $n_i \leq \kappa \cdot n_j$ for all $i, j \in [r]$.

Let R be a graph of order r with $\Delta(R) < \Delta_R$. Assume that we are given a graph G with a partition $V(G) = V_1 \cup\!\!\!\cdot\, \ldots \cup\!\!\!\cdot\, V_r$ and a graph H with a partition $V(H) = X_1 \cup\!\!\!\cdot\, \ldots \cup\!\!\!\cdot\, X_r$ with $|V_i| = |X_i| = n_i$ such that (V_i, V_j) is an (ε, δ)-super-regular pair for every $ij \in E(R)$ and such that all edges of H run between sets X_i, X_j for which $ij \in E(R)$. Further assume that H is a-arrangeable and has $\Delta(H) \leq \sqrt{n}/\log n$. Then there exists an embedding $\varphi : V(H) \to V(G)$ of H to G such that $\varphi(X_i) = V_i$.

We will embed planar graphs into a large well connected subgraphs of G. These subgraphs will correspond to spanning trees of components of R. However, to be able to use Theorem 7.9, we shall need spanning trees whose maximum degree is bounded. This is the purpose of the following lemma.

Lemma 7.10
Given $k \in \mathbb{N}$, let R be a connected graph with minimum degree at least $v(R)/(2k)$. Then R has a spanning tree with maximum degree $8k$.

Proof. We define the *score* of a spanning tree T of R to be the sum of the squares of the degrees of vertices in T. Let T be a spanning tree of R with minimum score. Observe that T has less than $v(R)/(4k)$ vertices of degree $8k$, since the sum of the vertex degrees of T is $2v(R) - 2$.

Suppose that there is a vertex u of T whose degree in T exceeds $8k$. Observe that the removal of u from T disconnects T into more than $8k$ components, one of which, C, has less than $v(R)/(8k)$ vertices. Let v be the neighbour of u which is in C. Now v has at least $v(R)/(2k)$ neighbours in R, of which less than $v(R)/(8k)$ are in C and a further less than $v(R)/(4k)$ are of degree at least $8k$. It follows that v has a neighbour u' in R which is not in C and whose degree is less than $8k$. Let T' be obtained from T by deleting uv and inserting $u'v$. Then T' is still a spanning tree of v. Each vertex of T' has the same degree as in T except for u and u', which have respectively lost and gained one neighbour. It follows that the score of T' is smaller than that of T, which by contradiction completes the proof. $\square$

Finally, we need to specify which planar graphs we will embed into our spanning trees.

Let $H = (V, E)$ be a plane graph. We say that a k element subset $V' \subseteq V$ forms a *bag of order* k in H if there is $\{x_1, x_2\} \subseteq V \setminus V'$ such that $V' \cup \{x_1, x_2\}$ induces a copy of $K_{2,k}$ in H and all inner faces of $H[V' \cup \{x_1, x_2\}]$ are also faces of H. We call the vertices of V' which are not in the outer face the *interior vertices* of the bag.

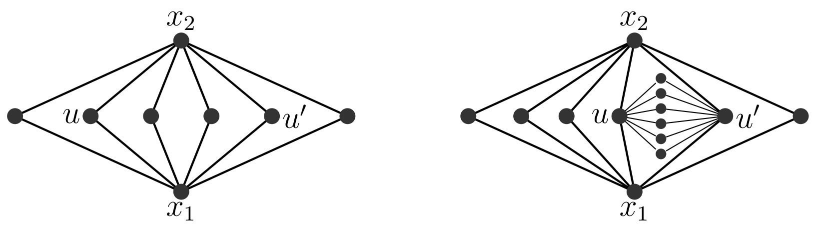

Figure 7.1: A bag of order $k = 6$; the same bag reordered and with an insertion of $\ell = 6$ vertices.

Observe that we can reorder the interior vertices of a bag without affecting the planarity of H. A bag is thus a very convenient structure into which one can put further vertices: if some vertex v (not in H) is adjacent in a supergraph G of H to any two interior vertices u, u' of a bag, then we can redraw H such that u and u' are consecutive in the bag, and insert v and $uv, u'v$ to obtain H'. If H is a quadrangulation, then H' is still a quadrangulation contained in G. Furthermore, if $v_1, \ldots, v_\ell \notin V(H)$ are all adjacent to u and u' in G, then we can insert all these vertices and edges to u and u', and still obtain a subgraph H'' of G which is a quadrangulation. Furthermore, $v_1, \ldots, v_\ell$ then form a bag of order ℓ in H''. This will be particularly useful in the proof of the following lemma.

Lemma 7.11
Let T be a tree of order $r \geq 2$, $n \geq (16r)^3$, and let G be an n-vertex graph with the parition $V_1 \uplus \ldots \uplus V_r$ of its vertex set such that $|V_i| \leq 2|V_j|$ for all $i \neq j$ and $G[V_i, V_j]$ is a complete bipartite graph whenever $ij \in E(T)$.

Then G contains a plane quadrangulation H with maximum degree $\Delta(H) \leq n^{1/3} + 2$ as a spanning subgraph. Furthermore, all but at most $9n^{2/3}$ vertices of H are contained in a collection of pairwise disjoint bags each of order in the interval $[n^{1/3}/2, n^{1/3}]$.

Proof. We first prove that G has a quadrangulation H with $\Delta(H) \leq n^{1/3} + 2$ by induction on r. So assume that $r = 2$ and G is a bipartite graph with partite sets V_1, V_2. We partition V_i into sets $W_{i,j}$ with sizes $|W_{i,j}| \leq n^{1/3}$ as equal as possible. The plane graph H is constructed as follows. Take $x_1, x_2 \in W_{1,1}$ and all of $W_{2,1}$ and embed the graph induced by these vertices into the plane. Let $y_1, y_2 \in W_{2,1}$ lie in the same face and embed $W_{1,1} \setminus \{x_1, x_2\}$ into this face connecting each vertex to y_1 and y_2. We continue greedily embedding sets $W_{i,j}$ into faces with two vertices of degree 2 from V_{3-i} and adding all edges in between. This process does not stop before all vertices of G have been embedded into the plane. The resulting graph H is a quadrangulation with $\Delta(H) \leq n^{1/3} + 2$.

Now assume that $r > 2$ and $1r \in E(T)$. Further assume that we have embedded $V_1 \uplus \ldots \uplus V_{r-1}$ this way and obtained a quadrangulation H' on $V \setminus V_r$. We extend H' to a quadrangulation H on V as follows. Again partition V_r into sets $W_{r,i}$ of size at most $n^{1/3}$. For each i, pick a pair of vertices u, u' in V_1 that have degree 2 and lie in the same face in H'. Embed all vertices from $W_{r,i}$ into this face and connect these to u, u'. Since $|V_1| \geq n/(2r-1) > 8n^{2/3}$, we do not run out of pairs in V_1.

It remains to show that most vertices lie in a collection of large disjoint bags. Recall that the planarity of H is preserved if we reorder the embedding in such a way that all vertices in $W_{i,j}$ of degree 2 in H form a bag. Since there are at most $2n^{2/3}$ many sets $W_{i,j}$, all but at most $4n^{2/3}$ many vertices lie in pairwise disjoint bags. Some of these bags might be small, i.e., they might have order less than $\frac{1}{2}n^{1/3}$. Assume that the bag in $W_{i,j}$ is small. Note that $|W_{i,j}| \geq \frac{9}{10}n^{1/3}$ by construction. Thus at least $\frac{1}{5}n^{1/3}$ pairs from $W_{i,j}$ have been used to embed other sets $W_{i',j'}$. But there are at most $2n^{2/3}$ many sets $W_{i,j}$. Hence, at most $10n^{1/3}$ bags are small. Consequently, all but at most $4n^{2/3} + 10n^{1/3} \cdot \frac{1}{2}n^{1/3} \leq 9n^{2/3}$ vertices lie in disjoint bags of size at least $\frac{1}{2}n^{1/3}$. □

7.3 Proof of Theorem 7.5

Given $\gamma > 0$, let $k \in \mathbb{N}$ be such that $k \leq 1/(2\gamma) < k+1$. We set $\beta = \gamma - 1/(2(k+1))$, $\delta = \beta/8$, $d = \beta/4$ and $s = C(\beta) + 6$, where $C(\beta)$ is the constant returned by Theorem 7.4. Next we choose ε such that 2ε is sufficiently small to apply Theorem 7.9 with $a = 10$, $\Delta_R = 8(k+1) + 1$ and $\kappa = 2$. We further insist that

$$\varepsilon \leq \min\left\{\frac{\beta}{8(k+1)+1}, \frac{1}{16(k+1)}, \frac{1}{10^5 k^4(8(k+1)+2)}\right\}.$$

Let r_1 be the parameter returned by Lemma 7.6 for d and ε as chosen. Let $n_0 \geq \max\left\{(16r_1)^3, 6(k+1)s\right\}$ be sufficiently large so that Theorem 7.9 applies with any $r \leq r_1$.

Suppose that $n \geq (4k) \cdot n_0$ and let G be an n-vertex graph with minimum degree $\delta(G) \geq \left(\frac{1}{2(k+1)} + \beta\right)n$. By Lemma 7.6 there is an (ε, d)-regular partition $V(G) = V_0 \mathbin{\dot\cup} \ldots \mathbin{\dot\cup} V_r$ with $r \leq r_1$ such that the corresponding reduced graph R satisfies $\delta(R) \geq \left(\frac{1}{2(k+1)} + \beta/2\right)r$. We distinguish two cases.

<u>Case 1:</u> R has a component with less than $2\delta(R)$ vertices. In this case R contains a triangle. It follows by Lemma 7.7 that G contains a triangulation T on s vertices, which has $3s-6$ edges. The graph $G - V(T)$ has minimum degree at least $\left(\frac{1}{2(k+1)} + \beta\right)n - s \geq \beta n$, where the last inequality is by our choice of n_0. Therefore, by Theorem 7.4, $G - V(T)$ contains a planar subgraph S with at least $2(n-s) - C(\beta)$ edges. Then G contains the disjoint union of S and T, which is planar and has at least $2n - 2s - C(\beta) + 3s - 6 \geq 2n$ edges (by choice of s) as required.

<u>Case 2:</u> Every component of R has at least $2\delta(R) > r/(k+1)$ vertices. It follows that R has $c \leq k$ components. We will show that we can cover G with c vertex-disjoint quadrangulations, which implies that G contains a planar subgraph with at least $2n - 4c \geq 2n - 4k$ edges as required.

Let C be a component of R, and T be its spanning tree with maximum degree $8(k+1)$ guaranteed by Lemma 7.10. Let V_i be any cluster of C and $ij \in T$. Observe that, by the (ε, d)-regularity of (V_i, V_j), at most $\varepsilon|V_i|$ vertices do *not* have at least $(d-\varepsilon)|V_j|$ neighbours in V_j. It follows that we can remove from each cluster V_i at most $8(k+1)\varepsilon|V_i|$ vertices and obtain a set V_i' whose every vertex has at least $(d-\varepsilon)|V_j|$ neighbours in each V_j such that $ij \in T$. Since $(8(k+1)+1)\varepsilon \leq \beta/8$ and $d - \delta = \beta/8$, if $ij \in T$, then each vertex in V_i' has at least $\delta|V_j'|$ neighbours in V_j'. Moreover, since $8(k+1)\varepsilon \leq 1/2$, for each i we have $|V_i'| \geq |V_i|/2$. Consequently, the pair (V_i', V_j') is 2ε-regular (see [69, Fact 1.5]) and, since $|V_i| = |V_j|$, we also have $|V_i'| \leq 2|V_j'|$ for each i, j. It follows that each edge ij of T corresponds to a $(2\varepsilon, \delta)$-super-regular pair (V_i', V_j'), and the cluster sizes are not too unbalanced, as required for Theorem 7.9.

Let G' be a graph whose vertex set is the union of the sets V_i', $i \in C$ and whose edges are all edges between V_i' and V_j' whenever $ij \in T$. Note that G' has n_C vertices, where n_C satisfies

$$n \geq n_C \geq 2\delta(R)\big(1 - 8(k+1)\varepsilon\big)\frac{(1-\varepsilon)n}{r} \geq \frac{n}{4k} \geq n_0.$$

By Lemma 7.11, G' contains a plane quadrangulation H in which the maximum degree is at most ${n_C}^{1/3} + 2 \leq n^{1/3} + 2$, and in which at most $9n_C^{2/3}$ vertices are not contained in bags of order between $\frac{1}{2}{n_C}^{1/3} \geq (n/32k)^{1/3}$ and $n_C^{1/3} \leq n^{1/3}$. By Theorem 7.9, H can be embeded into the subgraph of G induced on $\bigcup C$.

Repeating this for each component we obtain c vertex disjoint quadrangulations $H_1, \ldots, H_c$ in G, together with a collection $B_1, \ldots, B_\ell$ of pairwise disjoint bags of order at least $(n/32k)^{1/3}$ covering all but at most $9\sum_C n_C^{2/3} \leq 9n^{2/3}$ vertices of $\bigcup_{i\in[r]} V_i'$. In particular, we have that $\frac{1}{2}n^{2/3} < \ell \leq (32k)^{1/3}n^{2/3}$.

Let L be the set of vertices in none of the quadrangulations. Observe that every vertex in L is either in V_0 or in $V_i \setminus V_i'$ for some i; therefore, it follows that $|L| \leq (8(k+1)+1)\varepsilon n$. We say that a bag B_i is *good* for $u \in L$ if u has at least $n^{1/3}/(32k^2)$ neighbours in B_i. Since $(8(k+1)+1)\varepsilon n + 9n^{2/3} \leq \beta n$, each vertex $u \in L$ has at least $n/(2(k+1))$ neighbours contained in $B_1 \cup \cdots \cup B_\ell$. Of these at least $n/(2(k+1)) - \ell \cdot n^{1/3}/(32k^2) \geq n/(6k)$ lie in bags that are good for u. Hence, at least $n^{2/3}/(6k) \geq \ell/(24k^2)$ of the bags B_i are good for u. We now assign vertices L_i of L to each bag B_i sequentially as follows. From the collection of unassigned vertices of L for which B_i is good, we assign L_i to be any $n^{1/3}/(128k^2)$ of them if this is possible, and all of them if not. Suppose that after carrying out this procedure there is a vertex u of L which is not in any L_i. Then it must be the case that for each B_i good for u, we have $|L_i| = n^{1/3}/(128k^2)$. But there are at least $\ell/(24k^2)$ such B_i, and $\ell/(24k^2)\cdot n^{1/3}/(128k^2) > (8(k+1)+1)\varepsilon n \geq |L|$, (where the first inequality is by choice of ε) which is a contradiction.

We then work as follows. For each bag B_i, we reorder the interior vertices of B_i such that the first vertex of L_i is adjacent to the first and second interior vertices of B_i, the second vertex of L_i to the third and fourth, and so on. Because each vertex of L_i has at least $n^{1/3}/(32k^2)$ neighbours in B_i, and $|L_i| \leq n^{1/3}/(128k^2)$, this is possible. We now insert, for each j, the jth vertex of L_i to the interior face of B_i containing the $(2j-1)$st and $2j$th interior vertices, and add the edges to those two vertices. Let the plane graphs so constructed be $H_1', \ldots, H_c'$. By construction, these graphs are vertex disjoint and cover G, and since H_i was a quadrangulation, so H_i' is also a quadrangulation for each i. The disjoint union of $H_1', \ldots, H_c'$ is then a planar subgraph of G with $2n - 4c \geq 2n - 4k$ edges, as required.

7.4 Concluding remarks

There remain several open questions on planar graphs. In particular, it is possible that in Theorem 7.5 the constant n_γ can be taken to be an absolute constant provided $\gamma \gg n^{-1/2}$. Note that this is a natural lower bound since there are bipartite graphs without 4-cycles of minimum degree $\Theta(n^{1/2})$. Another possibility would be to investigate the behaviour of the planarity function $\mathrm{pl}(n, \gamma n)$ for $\gamma \in (1/2, 2/3]$ in more detail. Finally, one could ask these questions if the constraint imposed is that of edge density rather than minimum degree.

More generally, one could replace 'planar graphs' by some other property — topologically defined, or by forbidden minors, for example.

8 Homogeneous sets in graphs without an induced copy of C_5

8.1 Introduction and results

In this chapter we investigate classes of graphs that are defined by forbidding certain substructures. Let $\mathcal{H}$ be such a class. We focus on two related goals: to approximate the cardinality of $\mathcal{H}$ and to determine the structure of a typical graph in $\mathcal{H}$. In particular, we add the additional constraint that all graphs in $\mathcal{H}$ must have the same density c and would like to know how the answer to these questions depends on the parameter c.

The quantity $|\mathcal{H}_n|$ where $\mathcal{H}_n := \{G \in \mathcal{H} : V(G) = [n]\}$ is also called the *speed* of $\mathcal{H}$. Often exact formulas or good estimates for $|\mathcal{H}_n|$ are out of reach. In these cases, however, one might still ask for the asymptotic behaviour of the speed of $\mathcal{H}$. One prominent result in this direction was obtained by Erdős, Frankl and Rödl [36] who considered properties $\mathcal{F}orb(F)$ defined by a single forbidden (weak) subgraph F. They proved that for each graph F with $\chi(F) \geq 3$ the class $\mathcal{F}orb_n(F)$ of n-vertex graphs that do not contain F as a subgraph satisfies $|\mathcal{F}orb_n(F)| = 2^{\mathrm{ex}(F,n)+o(n^2)}$ where $\mathrm{ex}(F,n) := (\chi(F)-2)\binom{n}{2}/(\chi(F)-1)$. In other words, if $\chi(F) \geq 3$ then the speed of $\mathcal{F}orb(F)$ asymptotically only depends on the chromatic number of F.

Here we are interested in features of the picture at a more fine grained scale. More precisely, we fix a density $0 < c < 1$ and are interested in the number $|\mathcal{H}_n(c)|$ of graphs on n vertices with property $\mathcal{H}$ and density c. Let $\mathcal{F}orb_n(F,c) = \mathcal{F}orb_n(F) \cap \mathcal{G}_n(c)$ where $\mathcal{G}_n(c)$ is the set of all graphs on vertex set $[n]$ with $\lfloor c\binom{n}{2}\rceil$ edges. Here $\lfloor x \rceil$ denotes the nearest integer to x. For the sake of readability, we will always assume in the following that $c\binom{n}{2}$ is an integer since rounding issues would not affect our asymptotic considerations.

Straightforward modifications of the proof of the theorem of Erdős, Frankl and Rödl [36] yield the following bounds for $|\mathcal{F}orb_n(F,c)|$. Let F be a graph with $\chi(F) = r$. For all $c \in (0, \frac{r-2}{r-1})$ we have

$$\lim_{n\to\infty} \frac{\log_2 |\mathcal{F}orb_n(F,c)|}{\binom{n}{2}} = \tfrac{r-2}{r-1}\, H\left(\tfrac{r-1}{r-2}c\right), \tag{8.1}$$

where $H(x)$ is the *binary entropy* function, that is, for $x \in (0,1)$ we set $H(x) := -x\log_2 x - (1-x)\log_2(1-x)$. Here we denote by $\log_2$ the logarithm to base 2. Notice that $\lim_{n\to\infty} \log_2 |\mathcal{F}orb_n(F,c)|/\binom{n}{2} = 0$ for $c > \frac{r-2}{r-1}$ by the theorem of Erdős and Stone [38].

The analogous problem for a graph class $\mathcal{F}orb^*(F)$ (or $\mathcal{F}orb^*_n(F)$ respectively), characterised by a forbidden *induced* subgraph F, is more challenging and was first considered by Prömel and Steger [81]. They specified a graph parameter, the so-called colouring number $\chi^*(F)$ of F, that serves as a suitable replacement of the chromatic number in the theorem of Erdős, Frankl and Rödl. More precisely, they showed that $|\mathcal{F}orb_n^*(F)| = 2^{\mathrm{ex}^*(F,n)+o(n^2)}$ with $\mathrm{ex}^*(F,n) := \big(\chi^*(F)-2\big)\binom{n}{2}/\big(\chi^*(F)-1\big)$ where $\chi^*(F)$ is defined as follows. A *generalised r-colouring* of F with $r' \in [0,r]$ cliques is a partition of $V(F)$ into r' cliques and $r-r'$ independent sets. The *colouring number* $\chi^*(F)$ is the largest integer $r+1$ such that there is an $r' \in [r]$ for which F has no generalised r-colouring with r' cliques. For example, we have $\chi^*(C_5) = 3$ and $\chi^*(C_7) = 4$.

This naturally extends to hereditary graph properties, i.e., classes of graphs $\mathcal{H}$ which are closed under isomorphism and taking induced subgraphs (and may therefore be characterised by possibly infinitely many forbidden induced subgraphs). Let $\mathcal{F}(r,r')$ denote the family of all graphs that admit a generalised r-colouring with r' cliques. Then the colouring number of $\mathcal{H}$ is

$$\chi^*(\mathcal{H}) := \max\{r+1 \colon \mathcal{F}(r,r') \subset \mathcal{H} \text{ for some } r' \in [0,r]\}\ ,$$

and we set $\mathrm{ex}^*(\mathcal{H},n) := \big(\chi^*(\mathcal{H})-2\big)\binom{n}{2}/\big(\chi^*(\mathcal{H})-1\big)$. Observe that this definition implies $\chi^*\big(\mathcal{F}orb^*(F)\big) = \chi^*(F)$. And indeed Alekseev [1], and Bollobás and Thomason [13] generalised the result of Prömel and Steger to arbitrary hereditary graph properties $\mathcal{H}$ and showed that $|\,\mathbb{P}_n\,| = 2^{\mathrm{ex}^*(\mathcal{H},n)+o(n^2)}$.

More precise estimates for the speed were given for monotone properties $\mathcal{H}$ (properties that are closed under isomorphisms and taking subgraphs) by Balogh, Bollobás, and Simonovits [11] who showed that $2^{\mathrm{ex}^*(\mathcal{H},n)} \le |\mathcal{H}_n| \le 2^{\mathrm{ex}^*(\mathcal{H},n)+cn\log_2 n}$ for some constant c, and for hereditary properties $\mathcal{H}$ by Alon, Balogh, Bollobás, and Morris [4] who proved $2^{\mathrm{ex}^*(\mathcal{H},n)} \le |\mathcal{H}_n| \le 2^{\mathrm{ex}^*(\mathcal{H},n)+n^{2-\varepsilon}}$ for some $\varepsilon = \varepsilon(\mathcal{H}) > 0$ and n sufficiently large. Prömel and Steger [79, 80] gave even more precise results for the speed of $\mathcal{F}orb_n^*(C_4)$ and $\mathcal{F}orb_n^*(C_5)$ which they determined up to a factor of $2^{O(n)}$. In fact, they showed in [80] that almost all graphs in $\mathcal{F}orb_n^*(C_5)$ are generalised split graphs, that is, graphs of a rather simple structure which are defined as follows. We say that a graph $G = (V,E)$ admits a *generalised clique partition* if there is a partition $V = V_1 \,\dot\cup \ldots \dot\cup\, V_k$ of its vertex set such that $G[V_i]$ is a clique and for $i > j > 1$ we have $e(V_i,V_j) = e(V_j,V_i) = 0$. A graph G is a *generalised split graph* if G or its complement admit a generalised clique partition.

It is illustrative to compare this result to the celebrated strong perfect graph theorem [23]. A graph G is *perfect* if $\chi(G')$ equals the clique number $\omega(G')$ for all induced subgraphs G' of G. The strong perfect graph theorem asserts that *all* graphs without induced copies of odd cycles C_{2i+1}, $i > 1$ and without induced copies of their complements $\overline{C}_{2i+1}$ are perfect. Using this characterisation, it is easy to see that generalised split graphs are perfect. Consequently the result of Prömel and Steger implies that already *almost all* graphs without induced C_5 are perfect (observe that C_5 is self-complementary).

In this chapter, we consider induced C_5-free graphs of density c and provide bounds for their number. In the spirit of the result by Prömel and Steger we also relate this quantity to the number of n-vertex perfect graphs and generalised split graphs with density c.

Definition 8.1
We define the following graph classes:

$$\begin{aligned}\mathcal{C}(n,c) &:= \mathcal{F}orb_n^*(C_5,c) := \mathcal{F}orb_n^*(C_5) \cap \mathcal{G}_n(c)\,,\\ \mathcal{P}(n,c) &:= \big\{G \in \mathcal{G}_n(c) \colon G \text{ is perfect}\big\}\,,\\ \mathcal{S}(n,c) &:= \big\{G \in \mathcal{G}_n(c) \colon G \text{ is a generalised split graph}\big\}\,.\end{aligned}$$

Observe that for all n and $c \in [0,1]$ we have $\mathcal{S}(n,c) \subset \mathcal{P}(n,c) \subset \mathcal{C}(n,c)$. Our first main result now bounds the multiplicative error term between $|\mathcal{S}(n,c)|$ and $|\mathcal{C}(n,c)|$. In order to state this we define the following function. Let

$$h(c) := \begin{cases} H(2c)/2 & \text{if } 0 < c < \frac{1}{4}\,,\\ 1/2 & \text{if } \frac{1}{4} \le c \le \frac{3}{4}\,,\\ H(2c-1)/2 & \text{otherwise}\,. \end{cases} \tag{8.2}$$

Note that the classes of all generalised split graphs, all perfect graphs, and all graphs without induced C_5 are closed under taking complements. Hence, e.g., $|\mathcal{C}(n,c)| = |\mathcal{C}(n,1-c)|$ for all $c \in (0,1)$ and h is in fact symmetric in $(0,1)$. Further note that $H(|2c-1|)/2 = h(c)$ for $c < 1/4$ or $c > 3/4$.

Theorem 8.2
For all $c \in (0,1)$ we have

$$\lim_{n\to\infty} \frac{\log_2 |\mathcal{C}(n,c)|}{\binom{n}{2}} = \lim_{n\to\infty} \frac{\log_2 |\mathcal{P}(n,c)|}{\binom{n}{2}} = \lim_{n\to\infty} \frac{\log_2 |\mathcal{S}(n,c)|}{\binom{n}{2}} = h(c)\,.$$

The full proof of this theorem uses Szemerédi's Regularity Lemma and can be found in [21]. A sketch of the proof is given in Section 8.2.

We remark that Bollobás and Thomason [14] studied related questions of a more general type (see also the references in [14] for earlier results in this direction). They were interested in the probability $\mathbb{P}_{\mathcal{H}} := \mathbb{P}[\mathcal{G}(n,p) \in \mathcal{H}]$ of an arbitrary hereditary property $\mathcal{H}$ in the probability space $\mathcal{G}(n,p)$ and showed that for any $\mathcal{H}$ there are very simple properties $\mathcal{H}^*$ which closely approximate $\mathcal{H}$ in the probability space $\mathcal{G}(n,p)$. In this context, our Theorem 8.2 estimates the probability of $\mathcal{H} = \mathcal{F}orb_n^*(C_5)$ in the probability space $\mathcal{G}(n,m)$ with $m = c\binom{n}{2}$ and states that $\mathcal{H} = \mathcal{F}orb_n^*(C_5)$ is approximated by the property $\mathcal{H}^*$ of being a generalised split graph in $\mathcal{G}(n,m)$. The actual value of the probability $\mathbb{P}_{\mathcal{H}}$ was estimated by Marchant and Thomason in [76] for several properties $\mathcal{H}$, such as $\mathcal{H} = \mathcal{F}orb_n^*(C_5)$ (see [76, 91]). The probabilities $\mathbb{P}[\mathcal{G}(n,p) \in \mathcal{H}]$ and $\mathbb{P}[\mathcal{G}(n,m = p\binom{n}{2}) \in \mathcal{H}]$ are related (but not identical), and we discuss their relation in Section 8.4.

Let us now move from the question of approximating cardinalities to determining the structure of a typical element in $\mathcal{F}orb_n^*(C_5)$. A well-known conjecture by Erdős and Hajnal [37] states that any family of graphs that does not contain a certain fixed graph F as an induced subgraph must contain a *homogeneous set*, i.e. a clique or a stable set, which is of size at least some positive power of the number of vertices.

The conjecture is known to be true for certain graphs F, but open, among others, for $F = C_5$ (see [48]). However, Loebl, Reed, Scott, Thomason, and Thomassé [75] recently showed that for any graph F *almost all graphs* in $\mathcal{F}orb_n^*(F)$ have a homogeneous set of size at least some positive power of n. Moreover, they ask for which graphs F it is true that almost all graphs in $\mathcal{F}orb_n^*(F)$ do indeed have a *linearly* sized homogeneous set.

It may seem at first sight that our estimates derived in Theorem 8.2, carrying an $o(n^2)$ term in the exponent, are too rough to tell us something about the structure of almost all graphs in $\mathcal{F}orb_n^*(C_5)$ or $\mathcal{F}orb_n^*(C_5, c)$. However, we can combine them with the ideas of [75] to answer the question of Loebl, Reed, Scott, Thomason, and Thomassé in the affirmative for the case $F = C_5$. In fact, we can prove this assertion even in the case where we, again, restrict the class to graphs with a given density.

Theorem 8.3
For $\eta > 0$ denote by $\mathcal{F}orb_{n,\eta}^*(F,c)$ the set of graphs $G \in \mathcal{F}orb_n^*(F,c)$ with $\hom(G) := \max\{\alpha(G), \omega(G)\} < \eta n$. *Then for every $0 < c < 1$ there exists $\eta > 0$ such that*

$$\frac{|\mathcal{F}orb_{n,\eta}^*(C_5,c)|}{|\mathcal{F}orb_n^*(C_5,c)|} \to 0 \quad (n \to \infty).$$

We provide the proof of this theorem in Section 8.3.

Similar statements as in Theorem 8.2 and 8.3, for forbidden graphs F other than C_5, seem to require more work.

8.2 Concepts in the proof of Theorem 8.2

This section contains techniques and results from the proof of Theorem 8.2 that are relevant for the proof of Theorem 8.3. However, we do not present the full proof of Theorem 8.2 as those results have been obtained in [93]. The self-contained version of Section 8.2 can be found in [21].

This section is organised as follows. We start with a lemma on the lower bound of Theorem 8.2 in Section 8.2.1. For the upper bound we need some preparations: We shall apply Szemerédi's Regularity Lemma, which is introduced in Section 8.2.2. In Section 8.2.3 we prepare the tools that allow us to count graphs with a forbidden induced subgraph. In Section 8.2.4, finally, we prove the upper bound of Theorem 8.2.

Section 8.2.3 and Section 8.2.4 contain many concepts that are crucial for the for the proof of Theorem 8.3 which we present in Section 8.3. We omitted proofs whenever later arguments do not explicitly refer to them. Those proofs can be found in [21].

8.2.1 The lower bound of Theorem 8.2

The following lemma constitutes the lower bound of Theorem 8.2. The bound given can be obtained by explicitly constructing a class of generalised split graphs of sufficient size.

Lemma 8.4
For all $c, \gamma \in (0,1)$ there is n_0 such that for all $n \geq n_0$ we have

$$|\mathcal{S}(n,c)| \geq 2^{h(c)\binom{n}{2}-\gamma\binom{n}{2}}.$$

We will use the following bound for binomial coefficients (see, e.g., [55]). For every $\gamma > 0$ there exists n_0 such that for every integer $m \geq n_0$ and for every real $c \in (0,1)$ we have

$$2^{mH(c)-\gamma m} \leq \binom{m}{cm} \leq 2^{mH(c)}. \tag{8.3}$$

We call the term $-\gamma m$ in the first exponent the *error term of Equation* (8.3).

8.2.2 Regularity

In order to prove the upper bound from Theorem 8.2, i.e.,

$$|\mathcal{C}(n,c)| \leq 2^{h(c)\binom{n}{2}+\gamma\binom{n}{2}},$$

we will analyse the structure of graphs in $\mathcal{C}(n,c)$ by applying a variant of the Regularity Lemma suitable for our purposes.

Lemma 8.5 (Regularity Lemma, [90])
For all $\varepsilon > 0$ and k_0 there is k_1 such that every graph $G = (V,E)$ on $n \geq k_1$ vertices has an ε-regular partition $V = V_0 \mathbin{\dot\cup} V_1 \mathbin{\dot\cup} \ldots \mathbin{\dot\cup} V_k$ with $k_0 \leq k \leq k_1$.

The strength of this lemma becomes apparent when it is complemented with corresponding embedding lemmas, such as the following (see, e.g., [68]).

Lemma 8.6 (embedding lemma)
For every $d > 0$ and every integer k there exists $\varepsilon > 0$ with the following property. Let H be a graph on k vertices $v_1, \ldots, v_k$. Let G be a graph. Let $V_1, \ldots, V_k$ be clusters of an (ε, d)-regular partition of G with reduced graph $R = ([k], E_R)$. If there is a homomorphism from H to R, then H is a subgraph of G.

8.2.3 Embedding induced subgraphs

We remark that the concepts and ideas presented in this section are not new. They were used for various similar applications, e.g., by Bollobás and Thomason [14] or Loebl, Reed, Scott, Thomason, and Thomassé [75], as well as for different applications such as property testing, e.g., by Alon, Fischer, Krivelevich and Szegedy [7], or Alon and Shapira [8].

We start with an embedding lemma for induced subgraphs, which allows us to find an induced copy of a graph F in a graph G with reduced graph R if F is an induced subgraph of R (see, e.g., [7]).

Lemma 8.7 (injective embedding lemma for induced subgraphs)
For every $d > 0$ and every integer k there exists $\varepsilon > 0$ such that for all $f \leq k$ the following holds. Let $V_1, \dots, V_f$ be clusters of an ε-regular partition of a graph G such that for all $1 \leq i < j \leq f$ the pair (V_i, V_j) is ε-regular. Let $F = (V_F, E_F)$ be a graph on f vertices and let $g\colon V_F \to [f]$ be an injective mapping from F to the clusters of G such that for all $1 \leq i < j \leq f$ we have $d(V_i, V_j) \geq d$ if $\{g^{-1}(i), g^{-1}(j)\} \in E(F)$ and $d(V_i, V_j) \leq 1 - d$ if $\{g^{-1}(i), g^{-1}(j)\} \notin E(F)$. Then G contains an induced copy of F.

In contrast to Lemma 8.6, this lemma allows us only to embed one vertex per cluster of G. Our goal in the following will be to describe an embedding lemma for induced subgraphs which allows us to embed more than one vertex per cluster. Observe first, that for this purpose we must have some control over the existence of edges respectively non-edges *inside* clusters of a regular partition of G. This can be achieved by applying the following lemma to each of these clusters. It is not difficult to infer this lemma from the Regularity Lemma (Lemma 8.5) by applying Turán's theorem and Ramsey's theorem (see, e.g., [7]).

We use the following definition. An (μ, ε, k)-*subpartition* of a graph $G = (V, E)$ is a family of pairwise disjoint vertex sets $W_1, \dots, W_k \subseteq V$ with $|W_i| \geq \mu|V|$ for all $i \in [k]$ such that each pair (W_i, W_j) with $\{i, j\} \in \binom{k}{2}$ is ε-regular. A (μ, ε, k)-subpartition $W_1, \dots, W_k$ of G is *dense* if $d(W_i, W_j) \geq \frac{1}{2}$ for all $\{i, j\} \in \binom{k}{2}$, and *sparse* if $d(W_i, W_j) < \frac{1}{2}$ for all $\{i, j\} \in \binom{k}{2}$.

Lemma 8.8
For every k and ε there exists $\mu > 0$ such that every graph $G = (V, E)$ with $n \geq \mu^{-1}$ vertices either has a sparse or a dense (μ, ε, k)-subpartition.

The idea for the embedding lemma for induced subgraphs F of G now is as follows. We first find a regular partition of G. By Lemma 8.7, if a regular pair (V_i, V_j) in this partition is very dense then we can embed edges of F into (V_i, V_j), if it is very sparse then we can embed non-edges of F, and if its density is neither very small nor very big then we can embed both edges and non-edges of F. Moreover, Lemma 8.8 asserts that each cluster either has a sparse or a dense subpartition. In the first case we can embed non-edges inside this cluster, in the second case we can embed edges.

This motivates that we want to tag the reduced graphs with some additional information. For this purpose we colour an edge of the reduced graph white if the corresponding regular pair is sparse, grey if it is of medium density, and black if it is dense. Moreover we colour a cluster white if it has a sparse subpartition and black otherwise. We call a cluster graph that is coloured in this way a type. The following definitions make this precise.

Definition 8.9 (coloured graph, type)
A coloured graph *R is a triple (V_R, E_R, σ) such that (V_R, E_R) is a graph and $\sigma\colon V_R \cup E_R \to \{0, \frac{1}{2}, 1\}$ is a colouring of the vertices and the edges of this graph where $\sigma(V_R) \subset \{0, 1\}$. Vertices and edges with colour 0, $\frac{1}{2}$, and 1 are also called* white, grey, *and* black, *respectively.*

*Let $G = (V, E)$ be a graph and let $V = V_0 \mathbin{\dot\cup} V_1 \mathbin{\dot\cup} \ldots \mathbin{\dot\cup} V_k$ be an ε-regular partition of G with reduced graph $([k], E_R)$. The $(\varepsilon, \varepsilon', d, k')$-*type *$R$ corresponding to the partition $V_0 \mathbin{\dot\cup} V_1 \mathbin{\dot\cup} \ldots \mathbin{\dot\cup} V_k$ is the coloured graph $R = ([k], E_R, \sigma)$ with colouring*

$$\sigma(\{i,j\}) = \begin{cases} 0 & \text{if } d(V_i, V_j) < d\,, \\ 1 & \text{if } d(V_i, V_j) > 1 - d\,, \\ \frac{1}{2} & \text{otherwise}\,, \end{cases}$$

for all $\{i,j\} \in E_R$ and

$$\sigma(i) = \begin{cases} 0 & \text{if } G[V_i] \text{ has a sparse } (\mu, \varepsilon', k')\text{-subpartition}\,, \\ 1 & \text{if } G[V_i] \text{ has a dense } (\mu, \varepsilon', k')\text{-subpartition}\,, \end{cases}$$

for all $i \in [k]$, where μ is the constant from Lemma 8.8 for input k' and ε'. In this case we also simply say that G has $(\varepsilon, \varepsilon', d, k')$-type R.

By the discussion above a combination of the Regularity Lemma, Lemma 8.5, and Lemma 8.8 gives the following.

Lemma 8.10 (type lemma)
For every $\varepsilon, \varepsilon' \in (0, \frac{1}{2})$ and for all integers k', k_0 there are integers k_1 and n_0 such that for every $d > 0$ every graph G on at least n_0 vertices has an $(\varepsilon, \varepsilon', d, k')$-type $R = ([k], E_R, \sigma)$ with $k_0 \leq k \leq k_1$ and with at most εk^2 non-edges.

For formulating our embedding lemma we need one last preparation. We generalise the concept of a graph homomorphism to the setting of coloured graphs.

Definition 8.11 (coloured homomorphism)
Let $F = (V_F, E_F)$ be a graph and $R = (V_R, E_R, \sigma)$ be a coloured graph. A coloured homomorphism *from F to R is a mapping $h \colon V_F \to V_R$ with the following properties.*

(*a*) *If $u, v \in V_F$ and $h(u) \neq h(v)$ then $\{h(u), h(v)\} \in E_R$.*

(*b*) *If $\{u, v\} \in E_F$ then*
$h(u) = h(v)$ and $\sigma(h(u)) = 1$, or $h(u) \neq h(v)$ and $\sigma\Big(\{h(u), h(v)\}\Big) \in \{\frac{1}{2}, 1\}$.

(*c*) *If $\{u, v\} \notin E_F$ then*
$h(u) = h(v)$ and $\sigma(h(u)) = 0$, or $h(u) \neq h(v)$ and $\sigma\Big(\{h(u), h(v)\}\Big) \in \{0, \frac{1}{2}\}$.

If there is a coloured homomorphism from F to R we also write $F \xrightarrow{\sigma} R$.

The following embedding lemma states that a graph F is an induced subgraph of a graph G with type R if there is a coloured homomorphism from F to R. This lemma is, e.g, inherent in [8]. For completeness we provide its proof below.

Lemma 8.12 (embedding lemma for induced graphs)
For every pair of integers k, k' and for every $d \in (0,1)$ there are $\varepsilon, \varepsilon' > 0$ such that the following holds. Let $f \leq k'$ and G be a graph on n vertices with $(\varepsilon, \varepsilon', d, k')$-type R on k vertices. Let F be an f-vertex graph such that there is a coloured homomorphism from F to R. Then F is an induced subgraph of G.

8.2.4 The upper bound of Theorem 8.2

The following lemma constitutes the upper bound of Theorem 8.2.

Lemma 8.13
For all $c, \gamma \in (0,1)$ there is n_0 such that for all $n \geq n_0$ we have

$$|\mathcal{C}(n,c)| \leq 2^{h(c)\binom{n}{2}+\gamma\binom{n}{2}}.$$

The idea of the proof of Lemma 8.13 is as follows. We proceed in three steps. Firstly, we start by applying the Regularity Lemma to all graphs in $\mathcal{C}(n,c)$. For each of the regular partitions obtained in this way there is a corresponding type, and in total we only get a constant number K of different types. Secondly, we continue with a structural analysis of the possible types R for graphs from $\mathcal{C}(n,c)$ and infer from Lemma 8.12 that R cannot contain a triangle all of whose edges are grey (see Lemma 8.14). Thirdly, we prove that a coloured graph without such a grey triangle can only serve as a type for at most $\mathrm{UB}(n)$ graphs on n vertices (see Lemma 8.15). Multiplying $\mathrm{UB}(n)$ with K then gives the desired bound.

We start with the second step.

Lemma 8.14
For every integer $k' \geq 5$ and every $d > 0$ there exist $\varepsilon_{L8.14}, \varepsilon'_{L8.14} > 0$ such that for every $0 < \varepsilon \leq \varepsilon_{L8.14}$ and every $0 < \varepsilon' \leq \varepsilon'_{L8.14}$ the following is true. If G is a graph whose $(\varepsilon, \varepsilon', d, k')$-type R contains three grey edges forming a triangle then G has an induced C_5.

Next, we show an upper bound on the number of graphs on n vertices with a fixed type R, where R does not contain a triangle with three grey edges. We use the following definition.

$$\mathcal{R}(R, \varepsilon, \varepsilon', d, k', n, c) := \big\{G \in \mathcal{G}(n,c)\colon\ G \text{ has } (\varepsilon, \varepsilon', d, k')\text{-type } R\big\}. \tag{8.4}$$

We stress that $\mathcal{R}(R, \varepsilon, \varepsilon', d, k', n, c)$ and $\mathcal{R}(R', \varepsilon, \varepsilon', d, k', n, c)$ may have non-empty intersection for $R \neq R'$.

Lemma 8.15
For every c with $0 < c \leq \frac{1}{2}$, and every $\gamma > 0$ there exist $\varepsilon_{L8.15}, d_0 > 0$ and integers $n_{L8.15}$, k_0 such that for all positive $d \leq d_0$, $\varepsilon \leq \varepsilon_{L8.15}$, ε', and all integers $k \geq k_0$, k', $n \geq \max\{k, n_{L8.15}\}$ the following holds. If R is a coloured graph of order k which has at most εk^2 non-edges and does not contain a triangle with three grey edges, then

$$|\mathcal{R}(R, \varepsilon, \varepsilon', d, k', n, c)| \leq 2^{h(c)\binom{n}{2}+\gamma\binom{n}{2}}.$$

Proof. Let c, γ be given. Choose $\varepsilon_{L8.15}$, d_0, k_0 such that $\max\{4\varepsilon_{L8.15}, H(d_0), \frac{1}{k_0}\} \leq \frac{\gamma}{5}$. Let $n_{L8.15}$ be large enough to guarantee $\log_2(n_{L8.15}+1) \leq \frac{\gamma}{5}(n_{L8.15}-1)$. Let $\varepsilon \leq \varepsilon_{L8.15}$, ε', $d \leq d_0$, $k \geq k_0$, k', $n \geq \max\{n_{L8.15}, k\}$ be given.

Let $R = ([k], E_R, \sigma)$ be a coloured graph which has at most εk^2 non-edges and does not contain a triangle with three grey edges. We shall count the number of graphs in

$\mathcal{R}(R, \varepsilon, \varepsilon', d, k', n, c)$ by estimating the number of equipartitions $V_0 \uplus \ldots \uplus V_k$ of $[n]$, the number of choices for edges with one end in the exceptional set V_0 and edges in pairs (V_i, V_j) such that $\{i, j\} \notin E_R$, the number of choices for edges in clusters V_i such that i is white or black in R, and the number of choices for at most $c\binom{n}{2}$ edges in pairs (V_i, V_j) such that $\{i, j\}$ is a white, black, or grey edge of R.

The number of equipartitions $V_0 \uplus \ldots \uplus V_k$ of $[n]$ is bounded by

$$(k+1)^n = 2^{n \log_2(k+1)} \leq 2^{n \log_2(n+1)} \leq 2^{\frac{\gamma}{5}\binom{n}{2}}. \tag{8.5}$$

Let us now fix such an equipartition. There are at most εn^2 possible edges that have at least one end in V_0 and at most $\frac{\varepsilon}{2}n^2$ possible edges in pairs (V_i, V_j) such that $\{i, j\} \notin E_R$. Thus there are at most

$$2^{\frac{3}{2}\varepsilon n^2} \leq 2^{4\varepsilon\binom{n}{2}} \tag{8.6}$$

possible ways to distribute such edges. In addition, the number of ways to distribute edges in clusters V_i corresponding to white or black vertices of R is at most

$$2^{k\binom{n/k}{2}} \leq 2^{\frac{1}{k}\binom{n}{2}} \tag{8.7}$$

By definition, white edges of an $(\varepsilon, \varepsilon', d, k')$-type correspond to pairs with density at most d and black edges correspond to pairs with density at least $(1-d)$. Hence, by the symmetry of the binomial coefficient the number of ways to distribute edges in pairs (V_i, V_j) such that $\{i, j\}$ is a white or a black edge of R is at most

$$\binom{\left(\frac{n}{k}\right)^2}{d\left(\frac{n}{k}\right)^2}^{\binom{k}{2}} \leq 2^{\binom{k}{2}\left(\frac{n}{k}\right)^2 H(d)} \leq 2^{H(d)\binom{n}{2}}. \tag{8.8}$$

For later reference we now sum up the estimates obtained so far. The product of (8.5)–(8.8) gives less than

$$2^{\left(\frac{\gamma}{5}+4\varepsilon+H(d)+\frac{1}{k}\right)\binom{n}{2}} \leq 2^{\frac{4}{5}\gamma\binom{n}{2}} \tag{8.9}$$

choices for the partition $V_0 \uplus \ldots \uplus V_k$ and for the distribution of edges inside such a partition, except for the pairs (V_i, V_j) corresponding to grey edges of R.

It remains to take the grey edges E_g of R into account. By assumption E_g does not contain a triangle. Hence, by Turán's Theorem (see, e.g., [92]) we have $|E_g| \leq \frac{k^2}{4}$. It follows that there are at most $\frac{k^2}{4}\left(\frac{n}{k}\right)^2 = \frac{n^2}{4}$ possible places for edges in E_g-pairs (V_i, V_j), i.e., pairs such that $\{i, j\} \in E_g$. Hence the number N_g of possible ways to distribute at most $c\binom{n}{2}$ edges to E_g-pairs is at most $c\binom{n}{2}\binom{n^2/4}{c\binom{n}{2}}$. If $c < \frac{1}{4}$ then this gives

$$N_g \leq 2^{\frac{1}{2}\binom{n}{2}H(2c)+\gamma\binom{n}{2}}, \tag{8.10}$$

and if $\frac{1}{4} \leq c \leq \frac{1}{2}$ then

$$N_g \leq 2^{\frac{n^2}{4}} \leq 2^{\frac{1}{2}\binom{n}{2}+\frac{\gamma}{5}\binom{n}{2}}. \tag{8.11}$$

Combining (8.10) and (8.11) and recalling the definition of $h(c)$ in (8.2) gives

$$N_g \le 2^{h(c)\binom{n}{2}+\frac{\gamma}{5}\binom{n}{2}} . \tag{8.12}$$

Multiplying (8.9) and (8.12) gives the desired upper bound

$$|\mathcal{R}(R,\varepsilon,\varepsilon',d,k',n,c)| \le 2^{h(c)\binom{n}{2}+\frac{\gamma}{5}\binom{n}{2}+\frac{4}{5}\gamma\binom{n}{2}} = 2^{h(c)\binom{n}{2}+\gamma\binom{n}{2}} . \qquad \square$$

8.3 The proof of Theorem 8.3

Our proof of Theorem 8.3 consists of the following steps. We start, similarly as in the proof of Theorem 8.2, by constructing for each graph G in $\mathcal{F}orb^*_{n,\eta}(C_5,c)$ a type R of size independent of n with the help of the type lemma, Lemma 8.10. Next, we consider each cluster V_i of a partition of $V(G)$ corresponding to R separately. We shall show that the fact that G does not contain homogeneous sets of size ηn implies that $G[V_i]$ has many vertex disjoint induced copies of P_3, the path on three vertices, or many vertex disjoint induced copies of the anti-path $\overline{P}_3$, the complement of P_3 (see Lemma 8.16). Many induced copies of P_3 or $\overline{P}_3$ in two clusters V_i and V_j, however, limit the number of possibilities to insert edges between V_i and V_j without inducing a C_5 (see Lemma 8.17). Combining this with the proof strategy from Theorem 8.2 will give us an upper bound for the number of graphs from $\mathcal{F}orb^*_{n,\eta}(C_5,c)$ with type R (see Lemma 8.18). Finally, comparing this upper bound with the lower bound on $|\mathcal{F}orb^*_{n}(C_5,c)|$ from Theorem 8.2 will lead to the desired result.

We start by proving that graphs without big homogeneous sets contain many vertex disjoint induced P_3 or $\overline{P}_3$.

Lemma 8.16
Let G be a graph of order n with $\hom(G) \le n/6$. Then one of the following is true.

(i) G contains $n/6$ vertex disjoint induced copies of P_3, or

(ii) G contains $n/6$ vertex disjoint induced copies of $\overline{P}_3$.

Proof. Let G be an n-vertex graph with $\hom(G) \le n/6$. Select a maximal set of disjoint copies of P_3. If this set consists of less than $n/6$ paths then there is a subgraph $G' \subseteq G$ with $v(G') = n/2$ that has no induced P_3 and thus is a vertex disjoint union of cliques $Q_1,\dots,Q_\ell$. We claim that in G' we can find $n/6$ vertex disjoint induced $\overline{P}_3$, which proves the lemma.

Indeed, since $\hom(G) \le n/6$ we have $\ell \le n/6$ and for each $i \in [\ell]$ we have $q_i := |Q_i| \le n/6$. This implies $\sum_{i\in[\ell]} \lfloor q_i/2 \rfloor \ge \frac{1}{2}(\frac{n}{2} - \frac{n}{6}) = n/6$. It follows that we can find a set of $n/6$ vertex disjoint edges $E = \{e_1,\dots,e_{n/6}\}$ in these cliques in the following way. We first choose as many vertex disjoint edges in Q_1 as possible, then in Q_2, and so on, until we chose $n/6$ edges in total. Let Q_k be the last clique used in this process. Then for each clique Q_i with $i < k$ at most one vertex was unused in this process, and in Q_k

possibly several vertices were unused. Let X be the set of all these unused vertices together with all vertices from $\bigcup_{k<i\le\ell} Q_i$. Clearly $|X| = n/6$.

We consider the auxiliary bipartite graph $B = (X \cup E, E_B)$ with $\{x, e\} \in E_B$ for $x \in X$ and $e \in E$ iff x and e do not lie in the same clique of G'. We verify Hall's condition for B. So let $Y \subseteq X$. If $Y \not\subseteq Q_i$ for all i we have $|N(Y)| = |E| = |X| \ge |Y|$. Else if $Y \subseteq Q_i$ for some i then $|N(Y)| \ge |E| - (|Q_i| - |Y|)/2 \ge n/6 - n/12 + |Y|/2 \ge |Y|$ since $|Y| \le n/6$. It follows that B has a perfect matching, which means that there are $n/6$ vertex disjoint induced $\overline{P}_3$ in G' as claimed. $\square$

Now suppose we are given a graph G with vertex set $V_1 \mathbin{\dot\cup} V_2$ and no edges between V_1 and V_2. Let further H_1 and H_2 be such that for $i \in [2]$ the graph H_i induces a copy of P_3 or $\overline{P}_3$ in $G[V_i]$. Observe that, no matter which combination of P_3 or $\overline{P}_3$ we choose, we can create an induced C_5 in G by adding appropriate edges between H_1 and H_2. Since we are interested in graphs without induced C_5 this motivates why we call (H_1, H_2) a *dangerous pair* of (V_1, V_2).

Our next goal is to use these dangerous pairs in order to derive an upper bound on the number of possibilities to insert edges between V_1 and V_2 without creating an induced copy of C_5 if we know that (V_1, V_2) contains many dangerous pairs. In order to quantify this upper bound in Lemma 8.17 we use the following technical definition. We define $R(c) = c^4(1-c)^4$ and the function $r : (0,1) \to \mathbb{R}^+$ with

$$r(c) = \frac{1}{72}\begin{cases} R(2c) & \text{if } c < \frac{1}{4}, \\ (1/4)^4 & \text{if } c \in [\frac{1}{4}, \frac{3}{4}], \\ R(2c-1) & \text{otherwise.} \end{cases} \tag{8.13}$$

Recall in addition the definition of the function $h(c)$ from (8.2).

Lemma 8.17
For every $0 < c_0 \le \frac{1}{2}$ there is n_0 such that for all c with $c_0 \le 2c \le 1-c_0$ and $n \ge n_0$ the following holds. Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two n-vertex graphs, each of which contains $n/6$ vertex disjoint induced copies of P_3 or $n/6$ vertex disjoint induced copies of $\overline{P}_3$. Let $G = (V_1 \mathbin{\dot\cup} V_2, E)$ be the disjoint union of G_1 and G_2. Then there are at most

$$2^{2n^2\left(h(c)-r(c)\right)}$$

ways to add exactly $2cn^2$ edges to G that run between V_1 and V_2 without inducing a C_5 in G.

We remark that in the proof of this lemma we are going to make use of the following probabilistic principle: We can *count* the number of elements in a finite set X which have some property P, by determining the *probability* that an element which is chosen from X uniformly at random has property P.

Proof of Lemma 8.17. Given $c_0 \in (0, \frac{1}{2}]$ let n_0 be sufficiently large such that

$$n_0 e^{-2r(c_0/2)n_0^2} \le 2^{-2r(c_0/2)n_0^2}\,. \tag{8.14}$$

Now let c be such that $c_0 \leq 2c \leq 1 - c_0$. Observe first that it suffices to prove the lemma for $2c \leq \frac{1}{2}$, since induced C_5-free graphs are self-complementary and P_3 is the complement of $\overline{P}_3$. Hence, we assume from now on that $2c \leq \frac{1}{2}$. Observe moreover, that (8.14) remains valid if c_0 is replaced by c since $r(c)$ is monotone increasing in $[0, \frac{1}{2}]$. Let G_1, G_2, and G be as required.

Our first goal is to estimate the probability P^* of inducing no C_5 in G when choosing uniformly at random exactly $2cn^2$ edges between V_1 and V_2. Instead of dealing with P^* directly, we consider the following binomial random graph $\mathcal{G}(V_1, V_2, p)$ with $p = 2c$: we start with G and add each edge between V_1 and V_2 independently with probability p.

Now, let A be the event that $\mathcal{G}(V_1, V_2, p)$ contains exactly $2cn^2$ edges between V_1 and V_2, and let B be the event that $\mathcal{G}(V_1, V_2, p)$ contains no induced C_5. Observe that each graph with $2cn^2$ edges between V_1 and V_2 is equally likely in $\mathcal{G}(V_1, V_2, p)$ and thus

$$P^* = \mathbb{P}[B|A] \leq \frac{\mathbb{P}[B]}{\mathbb{P}[A]}\,. \tag{8.15}$$

Hence it suffices to estimate $\mathbb{P}[A]$ and $\mathbb{P}[B]$.

We first bound $\mathbb{P}[B]$. By assumption there are at least $n^2/36$ dangerous pairs in (V_1, V_2). Now fix such a dangerous pair (H_1, H_2). The probability that (H_1, H_2) induces a C_5 in $\mathcal{G}(V_1, V_2, p)$ is at least $p^2(1-p)^4$ unless H_1 and H_2 are both $\overline{P}_3$, and at least $p^4(1-p)^2$ unless H_1 and H_2 are both P_3. Thus (H_1, H_2) induces a copy of C_5 with probability at least

$$p^4(1-p)^4 = (2c)^4(1-2c)^4 \overset{(8.13)}{\geq} 72 \cdot r(c)\,.$$

Since we can upper bound the probability of B by the probability that none of the $n^2/36$ dangerous pairs in (V_1, V_2) induces a C_5 in $\mathcal{G}(V_1, V_2, p)$ we obtain

$$\mathbb{P}[B] \leq \big(1 - 72 \cdot r(c)\big)^{n^2/36} \leq e^{-2r(c)n^2}.$$

Note that the number of edges between V_1 and V_2 in $\mathcal{G}(V_1, V_2, p)$ is binomially distributed. Thus by Stirling's formula $\mathbb{P}[A] \geq 1/(\sqrt{2\pi p(1-p)}n) \geq 1/n$. By the choice of n_0, combining this with (8.15) gives

$$P^* \leq n\, e^{-2r(c)n^2} \overset{(8.14)}{\leq} 2^{-2r(c)n^2}. \tag{8.16}$$

It remains to estimate the number N of ways to choose exactly $2cn^2$ edges between V_1 and V_2. We have

$$N \;\leq \binom{n^2}{2cn^2} \overset{(8.3)}{\leq} 2^{2h(c)n^2}\,. \tag{8.17}$$

This implies that the number of ways to add exactly $2cn^2$ edges to G that run between V_1 and V_2 without inducing a C_5 is

$$P^* \cdot N \overset{(8.16),(8.17)}{\leq} 2^{-2r(c)n^2} \cdot 2^{2h(c)n^2}\,. \qquad \square$$

Next, we want to show that Lemma 8.17 allows us to derive an upper bound on the number of graphs G such that (a) G has no large homogeneous sets and (b) G has a fixed type R which does not contain a triangle with three grey edges. Our aim is to obtain an upper bound which is much smaller than the bound provided in Lemma 8.15 for the corresponding problem without restriction (a). Lemma 8.18 states that this is possible. Recall for this purpose the definition of $\mathcal{R}(R, \varepsilon, \varepsilon', d, k', n, c)$ from (8.4).

Lemma 8.18
For every c with $0 < c \leq \frac{1}{2}$, and every $\gamma > 0$ there exist $\varepsilon_0, d_0 > 0$ and an integer k_0 such that for all integers $k_1 \geq k_0$, k' there is an integer n_0 such that for all positive $d \leq d_0$, $\varepsilon \leq \varepsilon_0$, ε', and all integers $n \geq n_0$ and $k_0 \leq k \leq k_1$ the following holds. If R is a coloured graph of order k which has at most εk^2 non-edges and does not contain a triangle with three grey edges, and $\eta = 1/(6k_1)$, then

$$|\mathcal{R}(R, \varepsilon, \varepsilon', d, k', n, c) \cap \mathcal{F}orb^*_{n,\eta}(C_5, c)| \leq 2^{\left(h(c)-r(c)\right)\binom{n}{2}+\gamma\binom{n}{2}}.$$

In the proof of this lemma we combine the strategy of the proof of Lemma 8.15 with an application of Lemma 8.16 to all clusters of a partition corresponding to R, and an application of Lemma 8.17 to regular pairs of medium density. We shall make use of the following observation.

Using the definition of $r(c)$ from (8.13), it is easy to check that $f(c) := h(c) - r(c)$ is a concave function for $c \in (0, 1)$. Thus f enjoys the following property, which is a special form of Jensen's inequality (see, e.g., [50]).

Proposition 8.19 (Jensen's inequality)
Let f be a concave function, $0 < c < 1$, $0 < c_i < 1$ for $i \in [m]$ and let $\sum_{i=1}^{m} c_i = mc$. Then

$$\sum_{i=1}^{m} f(c_i) \leq m \cdot f(c).$$

Proof of Lemma 8.18. Let $c, \gamma > 0$ be given and choose ε_0, d_0, k_0 and n_0 as in the proof of Lemma 8.15. Let $d \leq d_0$ and k_1 be given and possibly increase n_0 so that $n_0 \geq 2k_1 n_{\text{L8.17}}$, where $n_{\text{L8.17}}$ is the constant from Lemma 8.17 with parameter d, and so that

$$\frac{3}{4}k_1^2 \log_2 n_0 + 2n_0 \leq \frac{\gamma}{10}\binom{n_0}{2}. \tag{8.18}$$

If necessary decrease ε_0 so that

$$3\varepsilon_0 \log_2 \frac{1}{c} \leq \frac{\gamma}{10}. \tag{8.19}$$

Let $\varepsilon \leq \varepsilon_0$, ε', $n \geq n_0$, and k with $k_0 \leq k \leq k_1$ be given.

Let $R = ([k], E_R, \sigma)$ be a coloured graph which has at most εk^2 non-edges and does not contain a triangle with three grey edges. In the proof of Lemma 8.15 we counted the number of graphs in $\mathcal{R}(R, \varepsilon, \varepsilon', d, k', n, c)$ by estimating the number of equipartitions $V_0 \mathbin{\dot\cup} \ldots \mathbin{\dot\cup} V_k$ of $[n]$, the number of choices for edges with one end in the exceptional set V_0 and edges in pairs (V_i, V_j) such that $\{i, j\} \notin E_R$, the number of choices for edges inside clusters V_i (such that i is white or black in R), and the number of choices for at most $c\binom{n}{2}$ edges in pairs (V_i, V_j) such that $\{i,j\}$ is a white, black, or grey edge of R. Now we are interested in the number of graphs in $\mathcal{R}(R,\varepsilon,\varepsilon',d,k',n,c) \cap \mathcal{F}orb^*_{n,\eta}(C_5,c)$. Clearly, we can use the same strategy, and it is easy to verify that the estimates in (8.5)–(8.8) and thus in (8.9) from the proof of Lemma 8.15 remain valid in this setting. From now on, as in the proof of Lemma 8.15, we fix a partition $V_0 \dot\cup \ldots \dot\cup V_k$ of $[n]$ and observe that also $m_g := |E_g| \leq \frac{k^2}{4}$ still holds for the grey edges E_g in R. However, we shall now use Lemma 8.17 to obtain an improved bound on the number of possible choices for edges in E_g-pairs (V_i, V_j), and use this to replace (8.12) by a smaller bound on the number N_g of possible ways to distribute at most $c\binom{n}{2}$ edges to E_g-pairs. Since in the following we do not rely on any interferences between different E_g-pairs, clearly N_g will be maximal if m_g is maximal, and hence we assume from now on that

$$m_g = \tfrac{k^2}{4}\,. \tag{8.20}$$

Let $s := |V_1| = \cdots = |V_k|$ and observe that

$$\frac{n}{k} \geq s \geq (1-\varepsilon)\frac{n}{k} \geq n_{\mathrm{L8.17}}\,. \tag{8.21}$$

By Lemma 8.16, for each cluster V_i of a partition P of a graph in $\mathcal{R}(R,\varepsilon,\varepsilon',d,k',n,c) \cap \mathcal{F}orb^*_{n,\eta}(C_5,c)$ such that P corresponds to R, we have that V_i contains either $s/6$ copies of P_3 or $s/6$ copies of $\overline{P}_3$. Hence we will assume from now on, that in our fixed partition the clusters V_i have this property.

We now upper bound N_g by multiplying the possible ways A to assign at most $c\binom{n}{2}$ edges to one E_g-pair each, and the maximum number B of ways to choose all these assigned edges in the corresponding pairs, without inducing a C_5. First observe that we have

$$A \leq \left(c\binom{n}{2}\right)^{m_g+1} \leq n^{3m_g} \overset{(8.20)}{\leq} 2^{\frac{3}{4}k^2\log_2 n}\,. \tag{8.22}$$

For estimating B, we now assume that we fixed an assignment in which each pair (V_i, V_j) with $\{i,j\} \in E_g$ is assigned $2c_{i,j}s^2$ edges. Let $\hat{c}$ be such that

$$\sum_{\{i,j\}\in E_g} 2c_{i,j}s^2 =: \hat{c}n^2 \leq c\binom{n}{2}. \tag{8.23}$$

Observe further that, since $\{i,j\} \in E_g$ is a grey edge of R, and we are interested in counting graphs with a partition corresponding to R we can assume that $d \leq 2c_{i,j} \leq (1-d)$. Hence, by (8.21) we can apply Lemma 8.17 with $c_0 = d$ to infer that for each $\{i,j\} \in E_g$ there are at most

$$B_{ij} \leq 2^{2s^2\left(h(c_{i,j}) - r(c_{i,j})\right)} \tag{8.24}$$

possible ways to choose the $2c_{i,j}s^2$ edges in (V_i, V_j) without inducing a C_5. Now let $2\widetilde{c} := \sum_{\{i,j\}\in E_g} 2c_{i,j}/m_g$ and observe that

$$\widetilde{c} \overset{(8.23)}{=} \hat{c}\frac{n^2}{2s^2m_g} \overset{(8.21)}{\leq} \hat{c}\frac{k^2}{2(1-\varepsilon)^2 m_g} \overset{(8.20)}{\leq} 2(1+3\varepsilon)\hat{c} \overset{(8.23)}{\leq} (1+3\varepsilon)c \leq \frac{3}{4}$$

Therefore, since $f(x) := h(x) - r(x)$ is a concave function for $x \in (0,1)$, which is moreover non-decreasing for $x \leq \frac{3}{4}$ we can infer from Lemma 8.19 that

$$\sum_{\{i,j\}\in E_g} f(c_{i,j}) \leq m_g \cdot f(\widetilde{c}) \leq m_g \cdot f\big(c(1+3\varepsilon)\big) \overset{(8.20)}{=} \frac{k^2}{4} f\big(c(1+3\varepsilon)\big)\,. \tag{8.25}$$

As $h(x)$ is a convex function with $h'(x) \leq \log_2(1/x)$ and $r(x)$ is non-decreasing for $x \leq 3/4$ we have

$$f\big(c(1+3\varepsilon)\big) \leq h(c+3\varepsilon) - r(c) \leq h(c) + 3\varepsilon h'(c) - r(c) \overset{(8.19)}{\leq} h(c) - r(c) + \frac{\gamma}{10}\,.$$

Together with (8.24) and (8.25) this implies

$$B = \prod_{\{i,j\}\in E_g} B_{ij} \leq 2^{2s^2(k^2/4)\left(h(c)-r(c)+\gamma/10\right)} \overset{(8.21)}{\leq} 2^{(n^2/2)\left(h(c)-r(c)+\gamma/10\right)}\,,$$

which in turn together with (8.22) gives

$$N_g \leq 2^{\frac{3}{4}k^2\log_2 n} \cdot 2^{(n^2/2)\left(h(c)-r(c)+\gamma/10\right)} \overset{(8.18)}{\leq} 2^{\left(h(c)-r(c)\right)\binom{n}{2}+(\gamma/5)\binom{n}{2}}\,.$$

By multiplying this with (8.9) from the proof of Lemma 8.15 we obtain

$$|\mathcal{R}(R,\varepsilon,\varepsilon',d,k',n,c) \cap \mathcal{F}orb^*_{n,1/6k}(C_5,c)| \leq 2^{\left(h(c)-r(c)\right)\binom{n}{2}+\gamma\binom{n}{2}}\,,$$

as claimed. $\square$

Lemma 8.18 together with the type lemma, Lemma 8.10, implies an upper bound on $|\mathcal{F}orb^*_{n,\eta}(C_5,c)|$. Now we can combine this with the lower bound on $|\mathcal{F}orb^*_n(C_5,c)|$ which follows from Lemma 8.4 in order to prove Theorem 8.3.

Proof of Theorem 8.3. Observe first that, since C_5 is self-complementary, it suffices to prove Theorem 8.3 for $c \leq 1/2$. Hence we assume $c \leq 1/2$ from now on.

We first need to set up some constants. Given $c \in (0, \frac{1}{2}]$, we choose $\gamma > 0$ such that $2\gamma < r(c)$. For input c and γ Lemma 8.4 supplies us with a constant $n_{\text{L8.4}}$. We apply Lemma 8.18 with input c and $\gamma/2$ to obtain $\varepsilon_{\text{L8.18}}$, d_0, and k_0. Next, we apply Lemma 8.14 with input d_0 and obtain constants $\varepsilon_{\text{L8.14}}$ and ε'. Let $\varepsilon := \min\{\varepsilon_{\text{L8.18}}, \varepsilon_{\text{L8.14}}\}$. For input ε, ε', and k_0 Lemma 8.10 returns constants k_1 and $n_{\text{L8.10}}$. With this parameter k_1 we continue the application of Lemma 8.18 and obtain $n_{\text{L8.18}}$. Choose $n_0 := \max\{n_{\text{L8.4}}, n_{\text{L8.10}}, n_{\text{L8.18}}, \frac{3}{\sqrt{\gamma}}k_1\}$, assume that $n \geq n_0$, and set $\eta := 1/(6k_1)$.

Now, for each graph $G \in \mathcal{F}orb^*_{n,\eta}(C_5,c)$ we apply the type lemma, Lemma 8.10, with parameters ε, ε', k_0, k' and d and obtain an $(\varepsilon,\varepsilon',d,k')$-type R of G on k vertices with $k_0 \leq k \leq k_1$ and with at most εk^2 non-edges. Let $\widetilde{\mathcal{R}}$ be the set of types obtained from these applications of Lemma 8.10. It follows that $|\widetilde{\mathcal{R}}| \leq 4^{\binom{k_1}{2}}2^{k_1} \leq 2^{k_1^2}$. By Lemma 8.14 applied with d_0, ε, and ε', no coloured graph in $\widetilde{\mathcal{R}}$ contains a triangle with three grey edges. Hence by Lemma 8.18 applied with c, $\gamma/2$, ε, ε', n, and k we have

$$\Big|\mathcal{R}(R,\varepsilon,\varepsilon',d,k',n,c) \cap \mathcal{F}orb^*_{n,\eta}(C_5,c)\Big| \leq 2^{\left(h(c)-r(c)\right)\binom{n}{2}+\frac{1}{2}\gamma\binom{n}{2}}\,.$$

Since, by Lemma 8.10,

$$\mathcal{F}orb^*_{n,\eta}(C_5,c) \subset \bigcup_{R\in\widetilde{\mathcal{R}}} \Big(\mathcal{R}(R,\varepsilon,\varepsilon',d,k',c,n) \cap \mathcal{F}orb^*_{n,\eta}(C_5,c)\Big)$$

we conclude from the choice of n_0 that

$$|\mathcal{F}orb^*_{n,\eta}(C_5,c)| \le 2^{k_1^2} \cdot 2^{\left(h(c)-r(c)\right)\binom{n}{2}+\frac{1}{2}\gamma\binom{n}{2}} \le 2^{\left(h(c)-r(c)\right)\binom{n}{2}+\gamma\binom{n}{2}}\,.$$

On the other hand, by Lemma 8.4 and the choice of n_0 we have

$$|\mathcal{F}orb^*_{n}(C_5,c)| \ge 2^{h(c)\binom{n}{2}-\gamma\binom{n}{2}}\,.$$

Since $2\gamma < r(c)$ by the choice of γ, this implies that almost all graphs in $\mathcal{F}orb^*_n(C_5,c)$ satisfy $\hom(G) \ge \eta n$. □

8.4 Concluding remarks

$\mathcal{G}(n,m)$ versus $\mathcal{G}(n,p)$

Our counting results can be interpreted as probabilities in the random graph model $\mathcal{G}(n,m=c\binom{n}{2})$. We can reformulate Theorem 8.2 as

$$\mathbb{P}\left[\mathcal{G}\left(n,m=c\tbinom{n}{2}\right) \in \mathcal{F}orb^*_n(C_5)\right] = \frac{|\mathcal{F}orb^*_n(C_5,c)|}{|\mathcal{G}(n,c)|} = 2^{(h(c)-H(c)+o(1))\binom{n}{2}}\,.$$

We now compare $\mathcal{G}(n,m=c\binom{n}{2})$ to the standard Erdős-Rényi model studied by Marchant and Thomason in [76]. They showed that

$$\mathbb{P}[\mathcal{G}(n,p) \in \mathcal{F}orb^*_n(C_5)] = 2^{c_p\binom{n}{2}}$$

where $c_p = \frac{1}{2}\max\{\log_2 p, \log_2(1-p)\}$. We can now derive the same estimate via Theorem 8.2: Obviously $\mathbb{P}[\mathcal{G}(n,p) \in \mathcal{F}orb^*_n(C_5)]$ equals

$$\max_c \mathbb{P}\left[e(\mathcal{G}(n,p)) = c\tbinom{n}{2}\right] \cdot \mathbb{P}\left[\mathcal{G}(n,m=c\tbinom{n}{2}) \in \mathcal{F}orb^*_n(C_5)\right] \cdot 2^{o(n^2)}\,.$$

Setting $g_p(c) = H(c) + c\log_2 p + (1-c)\log_2(1-p)$ we obtain

$$\mathbb{P}\left[e(\mathcal{G}(n,p)) = c\tbinom{n}{2}\right] = 2^{(g(c)+o(1))\binom{n}{2}}$$

and thus by Theorem 8.2

$$\mathbb{P}[\mathcal{G}(n,p) \in \mathcal{F}orb^*_n(C_5)] = 2^{(\max_c g_p(c)+h(c)-H(c)+o(1))\binom{n}{2}}\,.$$

The maximum is attained at $c = p/2$ for $p < 1/2$ and $c = (p+1)/2$ for $p > 1/2$. For $p = 1/2$ all values $c \in [1/4, 3/4]$ are optimal. Inserting the optimal value for c shows that the exponent is indeed equal to c_p as computed in [91]. This indicates that, e.g., a graph from $\mathcal{G}(n,1/4)$ that happens to be induced C_5-free will a.a.s. have $(1/8+o(1))\binom{n}{2}$ edges. Thus the $\mathcal{G}(n,m)$ model can also be used to derive that a typical element having a certain property might be far from a typical element in $\mathcal{G}(n,p)$.

Extensions

The final thing to be considered is whether more can be said about $\mathcal{F}orb_n^*(F, c)$ in the $\mathcal{G}(n, m)$ model. Indeed, there are at least three natural ways to enhance our results. First one might want to count $\mathcal{F}orb_n^*(F, c)$ for graphs F other than C_5. But already the case of a forbidden induced C_7 is more challenging, as tight upper bounds as in Lemma 8.15 are not so easy to derive. Furthermore, $h(c)$ does not seem to be unimodal for C_7.

Second it would be interesting to obtain even sharper asymptotic bounds for the speed of $\mathcal{F}orb_n^*(F, c)$. In [4], Alon et al. determine the speed of some hereditary properties up to a subquadratic term in the exponent. We believe that their techniques can be extended to the case of restricted density.

Finally one might want to prove a much better constant in the size of the linear homogeneous sets that can be found in almost all graphs in $\mathcal{F}orb_n^*(C_5, c)$. It easily follows from the fact that almost all graphs in $\mathcal{F}orb_n^*(C_5)$ are generalised split graphs (see [80]) that also almost all of them have a homogeneous set of size $(1/2 - o(1))n$. We believe that the same is true for the density restricted case and that this can be proven as in Theorem 8.3 combined with a stability type argument. We plan to return to this in the near future.

9 Homogeneous sets in graphs without an induced copy of P_4

9.1 Introduction

A well-known conjecture by Erdős and Hajnal [37] states that any family of graphs that does not contain a copy of a certain fixed graph F as an induced subgraph must contain a homogeneous set, i.e. a clique or a stable set, which is of size at least some positive power of the number of vertices. More formally, we define $\mathcal{F}orb_n^*(F)$ to be the set of all graphs on vertex set $[n]$ without *induced* subgraph F and we set $\hom(G) := \max\{\alpha(G), \omega(G)\}$, where, as usual, $\alpha(G)$ and $\omega(G)$ denote the size of a maximum independent set and of a maximum clique in a graph G, respectively.

Conjecture 9.1 (Erdős, Hajnal [37])
For every graph F there exists $\varepsilon(F) > 0$ such that $\hom(G) \geq n^{\varepsilon(F)}$ for all $G \in \mathcal{F}orb_n^*(F)$ and all $n \in \mathbb{N}$.

The conjecture is known to be true for some graphs F, such as all graphs on at most four vertices, all cliques [37] and the bull graph [24], but open in general (see [48]).[1] However, Loebl, Reed, Scott, Thomason, and Thomassé [75] showed that, for any fixed graph F, *almost all graphs* in $\mathcal{F}orb_n^*(F)$ have a homogeneous set of size at least some positive power of n. Moreover, they ask for the family of graphs F with the property that almost all graphs in $\mathcal{F}orb_n^*(F)$ have a *linearly* sized homogeneous set. In light of this, we say that a graph F has the *linear Erdős-Hajnal property* if there exists $c(F) > 0$ such that

$$\lim_{n\to\infty} \frac{|\{G \in \mathcal{F}orb_n^*(F) : \hom(G) \geq c(F)n\}|}{|\mathcal{F}orb_n^*(F)|} = 1\,.$$

The path P_3 on three vertices does not have this property, as the class $\mathcal{F}orb_n^*(P_3)$ is dominated by graphs which are the disjoint union of cliques of size $\log n$. Hence graphs in $\mathcal{F}orb_n^*(P_3)$ typically have $\hom(G) = \Theta(n/\log n)$. On the other hand, Böttcher, Taraz, and Würfl [21] showed that a cycle on five vertices satisfies the linear Erdős-Hajnal property, i.e., there is $c(C_5) > 0$ such that almost all graphs G in $\mathcal{F}orb_n^*(C_5)$ have $\hom(G) \geq c(C_5)n$.

Recently the question of Loebl et. al. was answered for almost all graphs F by Kang, McDiarmid, Reed, and Scott [56], who showed that almost all graphs F have the linear

[1] This seems to have changed recently: Sági [88] has announced a proof of the Erdős-Hajnal conjecture in November 2012.

Erdős-Hajnal property. The smallest graph for which the question has not been settled in their paper is the case $F = P_4$. This note proves that almost all graphs G in $\mathcal{F}orb^*_n(P_4)$ do have $\hom(G) = O(n/\log^{1-\varepsilon} n)$ where ε is an arbitrary positive constant, and hence P_4 does not satisfy the linear Erdős-Hajnal property. (The class of graphs $\mathcal{F}orb^*(P_4)$ is also the class of n-vertex *complement reducible graphs*, or *cographs* for short.)

Theorem 9.2
For every positive ε we have

$$\lim_{n\to\infty} \frac{|\{G \in Forb^*_n(P_4) : \hom(G) \le n/\log^{1-\varepsilon} n\}|}{|\mathcal{F}orb^*_n(P_4)|} = 1\,.$$

The proof of Theorem 9.2 relies on the one-to-one correspondence between cographs and a certain class of trees. The typical structure of these trees can be determined by means of generating functions, as we shall see in Section 9.2. This structure will help us find an upper bound on the size of homogeneous sets in typical cographs, which leads to the proof of Theorem 9.2 in Section 9.3.

9.2 Basic facts about cographs

There are several equivalent ways of defining the class of cographs, which we introduced as the class of graphs that do not contain an induced copy of P_4. Cographs can also be defined through graph operations usually known as the *(disjoint) union* and the *join*, or *(disjoint) sum*. Given a family of graphs $G_1 = (V_1, E_1), \dots, G_k = (V_k, E_k)$ whose vertex sets are mutually disjoint, the disjoint union $G_1 \cup \cdots \cup G_k$ is the graph with vertex set $V = V_1 \cup \cdots \cup V_k$ and edge set $E = E_1 \cup \cdots \cup E_k$. The disjoint sum $G_1 \times \cdots \times G_k$ is the graph obtained from the disjoint union by adding all edges of the form $\{x, y\}$ with $x \in V(G_i)$ and $y \in V(G_j)$ for $i \neq j$. The class of cographs may be defined recursively as follows:

(R1) The graph on a single vertex is a cograph.

(R2) If $G_1, \dots, G_k$ are cographs, so is the disjoint union $G_1 \cup \cdots \cup G_k$.

(R3) If $G_1, \dots, G_k$ are cographs, so is the join $G_1 \times \cdots \times G_k$.

We call the operation from (2) the *(disjoint) union* of graphs and the operation from (3) the *(disjoint) sum* of graphs.

Every cograph on vertex set $[n]$ can be represented as a rooted tree with n leaves labelled 1 through n and internal vertices that carry either the label "$\cup$" for disjoint union or "$\times$" for disjoint sum. Given such a tree T, we construct the corresponding cograph G_T by looking at T one level at a time, starting from the bottom and proceeding all the way to the root. A leaf of T corresponds to an isolated vertex in G_T that preserves its label in the tree, while an internal vertex generates, according to its label, the disjoint union or the disjoint sum of the cographs associated with the branches containing its children. This representation is unique if we require two things. First the labels of internal vertices must be alternating along any path in the tree and second all internal vertices (including the root, unless it is childless) must have at least two children. Trees with this property were named *cotrees* by Corneil, Lerchs, and Burlingham [28], who introduced this bijection.

We would like to point out that the cotree representation of a cograph G is very convenient if one wants to determine $\hom(G)$. We illustrate this with the following example. Let G be the disjoint sum of $G_1, \dots, G_k$. (This corresponds to a cotree whose root has label "$\times$" and with branches associated with $G_1, \dots, G_k$.) Then $\omega(G) = \sum_{i\in[k]} \omega(G_i)$ and $\alpha(G) = \max_{i\in[k]} \alpha(G_i)$. Hence we can use the cotree to recursively determine the size of the largest homogeneous set in G.[2]

We have seen how the cotree of a cograph allows us to easily determine graph parameters like α, ω, and hom. It is thus natural to study the structure of a typical cotree. The one-to-one correspondence between cographs and cotrees will then easily yield bounds on the size of homogeneous sets. As it turns out, the crucial parameter is the largest component under the root of a cotree. To be more precise, let $\mathcal{C}_n$ be the set of cotrees on n leaves labelled 1 through n whose root is labelled "$\times$" (hence $\mathcal{C}_n$ corresponds to the set of n-vertex connected cographs for all $n \geq 2$.). Let C_n be a cotree drawn uniformly at random in $\mathcal{C}_n$ and define $L(n)$ to be the number of leaves in the largest branch incident with the root of C_n. We may derive asymptotic information about the distribution of $L(n)$ using the general framework introduced by Gourdon [47], which leads to the following technical result.

Lemma 9.3
For any fixed integer k we have

$$\lim_{n\to\infty} \mathbb{P}[L(n) = n - k] = p_k$$

for some $p_k > 0$. In addition there exist constants $\kappa_1, \kappa_2 > 0$ such that

$$\mathbb{E}[L(n)] = n - \kappa_1\sqrt{n}(1 + o(1)) \quad \textit{and} \quad \mathrm{Var}[L(n)] = \kappa_2 n^{3/2}(1 + o(1))\,.$$

Proof of Lemma 9.3. The proof is based on a general framework introduced by Gourdon [47], who found the expected size of the largest component of some combinatorial structure $\mathcal{C}$ which may be constructed from some basic decomposition by means of combinatorial substitution. In our case, let

$$\mathcal{C} \stackrel{\text{def}}{=} \text{cotrees with at least two branches and a root labelled "}\times\text{"} \cup \{\varepsilon\},$$
$$\mathcal{P} \stackrel{\text{def}}{=} \text{labelled, connected cographs.}$$

There is a natural bijection between $\mathcal{C}$ and $\mathcal{P}$ that maps ε to the graph on a single vertex and a cotree with at least two branches (with a root labelled "$\times$") to the corresponding

[2]This property is the reason why largest cliques/stable sets in cographs can be found in linear time [29].

connected cograph. Moreover, the set $\mathcal{C}$ may be viewed as a set of objects formed by means of a function Φ that substitutes objects of $\mathcal{P}$ inside 'atoms', which corresponds to substituting a labelled cograph for each branch (atom) of the cotree and to partitioning the set of labels $\{1, \dots, n\}$ among the branches. We define

$$\begin{aligned} C(z) &\stackrel{\text{def}}{=} \text{ generating function for } \mathcal{C}, \\ P(z) &\stackrel{\text{def}}{=} \text{ generating function for } \mathcal{P}, \\ F(w) &\stackrel{\text{def}}{=} \text{ generating function for the combinatorial substitution,} \\ &\quad \text{ i.e., for the function } \Phi. \end{aligned}$$

In particular we have

$$C(z) = F(P(z)) \tag{9.1}$$

and, by definition,

$$C(z) = P(z) - z + 1 \quad \text{ and } \quad F(w) = e^w - w\,.$$

Hence (9.1) leads to

$$2P(z) - z + 1 = \exp(P(z))\,. \tag{9.2}$$

We wish to apply Theorem 1 of [47]. To do so, we have to show (among other things) that $P(z)$ is algebraic-logarithmic, i.e., that it can be written as

$$P(z) = c - \left(1 - \frac{z}{\rho}\right)^{\alpha} \left(\log \frac{1}{1 - z/\rho}\right)^{\beta} (d + o(1))$$

as z tends to the dominant singularity ρ. The dominant singularity of (9.2) is given by $\rho = 2\log 2 - 1$, and, using methods of singularity analysis (see Flajolet, Salvy, and Zimmermann [46]), we see that the local expansion of $P(z)$ equals

$$P(z) = \log 2 - \sqrt{\rho} \left(1 - \frac{z}{\rho}\right)^{1/2} + O(1 - z/\rho),$$

which is algebraic-logarithmic with $c = \log 2$, $d = \sqrt{\rho}$, $\alpha = 1/2$, and $\beta = 0$. Moreover,

$$[z^n]P(z) = \frac{\sqrt{3\sqrt{2} - 4}}{4\sqrt{\pi}} \frac{1}{n^{3/2}} \frac{1}{(3 - 2\sqrt{2})^n} \left(1 + o\left(\frac{1}{n}\right)\right)\,. \tag{9.3}$$

We apply Theorem 1 of [47] to derive that $L(n)$, which is the size of the largest $\mathcal{P}$-component in a random $\Phi(\mathcal{P})$ structure of size n (which is the number of leaves in a largest subtree under the root of the cotree), tends to the following discrete law:

$$\lim_{n \to \infty} \mathbb{P}[L(n) = n - k] = \frac{b_k \rho^k}{e^c - 1} =: p_k, \tag{9.4}$$

with ρ and c given above, and $b_k = [z^k]F'(P(z))$. Note that the b_k are easy to calculate as $F'(P(z)) = \exp(P(z)) - 1 = 2P(z) - z$ by (9.2). From (9.3) and (9.4) we deduce that

$$\mathbb{P}[L(n) = n - k] = O\left(k^{-3/2}\right) . \tag{9.5}$$

Moreover, by Corollary 1 of [47],

$$\begin{aligned}\mathbb{E}[L(n)] &= n - \kappa_1\sqrt{n}(1 + o(1)) \text{ and} \\ \mathrm{Var}[L(n)] &= \kappa_2 n^{3/2}(1 + o(1))\,,\end{aligned}$$

where the constants κ_i are given by

$$\begin{aligned}\kappa_1 &= \frac{dF''(c)}{\sqrt{\pi}F'(c)} = \sqrt{\rho/\pi}, \\ \kappa_2 &= \frac{dF''(c)}{\sqrt{\pi}F'(c)}\left(\frac{1}{\sqrt{\pi}} - \frac{\sqrt{\pi}}{4}\right) = \sqrt{\rho}\left(\frac{1}{\pi} - \frac{1}{4}\right).\end{aligned}$$ □

Lemma 9.3 states that most leaves of a typical cotree can be found in a single subtree. Using this structural property we now derive Theorem 9.2.

9.3 The proof of Theorem 9.2

We derive the statement by a first moment argument. Our goal will be to prove the following lemma.

Lemma 9.4
For every $\varepsilon > 0$, there exists a positive constant C such that

$$\mathbb{E}[\hom(G)] \le C\frac{n}{\log^{1-\varepsilon} n}$$

if G is drawn uniformly at random from the set of all cographs on vertex set $[n]$.

Lemma 9.4 together with Markov's inequality then easily implies Theorem 9.2 as for every $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \frac{|\{G \in \mathcal{F}orb^*_n(P_4) : \hom(G) \le n/\log^{1-2\varepsilon} n\}|}{|\mathcal{F}orb^*_n(P_4)|} \ge 1 - \lim_{n\to\infty}\frac{C}{\log^{\varepsilon} n} = 1\,.$$

Hence it remains to prove Lemma 9.4.

Proof of Lemma 9.4. We define $\omega(n)$ and $\alpha(n)$ to be the expected values of $\omega(C_n)$ and $\alpha(C_n)$, respectively, where C_n is drawn uniformly at random among all connected cographs on vertex set $[n]$. Similarly we define $\overline{\omega}(n)$ and $\overline{\alpha}(n)$ to be the expected values of $\omega(C_n)$ and $\alpha\left(C_n\right)$ when C_n is drawn uniformly at random among all disconnected cographs on vertex set $[n]$. Note that $\omega(C_n) = \alpha\left(\overline{C_n}\right)$, and the fact that cographs are closed under taking complements implies that $\omega(n) = \overline{\alpha}(n)$ and $\overline{\omega}(n) = \alpha(n)$.

Hence, $\mathbb{E}[\hom(G)] \leq \omega(n) + \alpha(n)$ if G is drawn uniformly at random from the set of all cographs on vertex set $[n]$. To show Lemma 9.4 it thus suffices to prove

$$\omega(n) = O\left(\frac{n}{\log^{1-\varepsilon} n}\right) \tag{9.6}$$

for every $\varepsilon > 0$.

It is easy to see that every subadditive function $\omega(n)$ satisfies one of the following two conditions.

(R1) For every $\varepsilon > 0$ there is C such that

$$\omega(n) \leq C\frac{n}{\log^{1-\varepsilon} n}, \text{ or}$$

(R2) the function $\omega(n)(\log n)/n$ is non-decreasing for n sufficiently large.

Indeed, if there is $\varepsilon > 0$ such that for every C there is n with

$$\omega(n) > C\frac{n}{\log^{1-\varepsilon} n}$$

the function $\omega(n)\log^{1-\varepsilon}(n)/n$ tends to infinity with $n \to \infty$. In particular, $\omega(n)(\log n)/n$ is non-decreasing for n sufficiently large.

We assume the latter for the rest of this proof.

Claim

$$\overline{\omega}(n) = \sum_{k\geq 1} \omega(n-k)p_k + O\left(\frac{\sqrt{n}}{\log^2 n}\right), \textit{ and} \tag{9.7}$$

$$\omega(n) \leq \sum_{k\geq 1} \left(\overline{\omega}(n-k) + \omega(k)\right) p_k \, . \tag{9.8}$$

Proof of Claim 9.4.1. The error term $O(\sqrt{n}/\log^2 n)$ in (9.7) is due to the (unlikely) case that the largest clique does not reside in the subgraph corresponding to the largest branch.

Inequality (9.8) is justified by the fact that $\overline{\omega}(k) \leq \omega(k)$ in the case where the cotree only has two branches (and the smaller branch thus corresponds to a disconnected cograph on k vertices). $\square$

Claim

$$\sum_{k\geq 1} \overline{\omega}(n-k)p_k \leq \frac{\omega(n)}{n}\left(\left(n - 2\kappa_1\sqrt{n}(1+o(1))\right) + O\left(\frac{\sqrt{n}}{\log n}\right)\right)$$

$$\sum_{k\geq 1} \omega(k)p_k \leq \omega(\kappa_1\sqrt{n})\,(1+o(1))$$

Proof of Claim 9.4.2. Recall that we have assume that $\omega(k)(\log k)/k$ is non-decreasing for all $k \geq \sqrt{n}/\log^2 n$. Moreover, we claim that $\omega(k)/k$ is non-increasing. To see this we show that, for every $k' \geq k$, we have

$$\omega(k) \geq \frac{k}{k'}\omega(k') \,. \tag{9.9}$$

So let $C_{k'}$ be drawn uniformly at random from the set of connected cographs on vertex set $[k']$ and let $M \subseteq [k']$ span a maximum clique in $C_{k'}$. Then the expected intersection of M with $[k]$ has $(k/k')|M|$ elements. Therefore

$$\mathbb{E}\left[\omega\left(C_{k'}[\{1,\dots,k\}]\right)\right] \geq \frac{k}{k'}\omega(k')$$

It follows from linearity of expectation and the fact that cographs are closed under taking induced subgraphs that

$$\omega(k) = \mathbb{E}\left[\omega(C_k)\right] \geq \mathbb{E}\left[\omega\big(C_{k'}[\{1,\dots,k\}]\big)\right]$$

where C_k is drawn uniformly at random from the set of all connected cographs on $[k]$. (We do not have equality here because $C_{k'}[\{1,\dots,k\}]$ is not necessarily connected.) This establishes (9.9).

We derive that

$$\frac{\omega(k)}{k} \leq \frac{\omega(\kappa_1\sqrt{n})}{\kappa_1\sqrt{n}}\left(1+\frac{\log(\kappa_1\sqrt{n}/k)}{\log k}\right) \leq \frac{\omega(\kappa_1\sqrt{n})}{\kappa_1\sqrt{n}}\left(1+\frac{5\log\log n}{\log n}\right)$$

for every $k \geq \sqrt{n}/\log^2 n$. Hence we can bound $\sum_{k\geq 1}\omega(k)p_k$ by

$$\begin{aligned}
\sum_{k\geq 1}\omega(k)p_k &\leq \sum_{k\geq 1}^{\sqrt{n}/\log^2 n} kp_k + \sum_{k>\sqrt{n}/\log^2 n}\omega(k)\,p_k \\
&\leq \frac{\sqrt{n}}{\log^2 n} + \frac{\omega(\kappa_1\sqrt{n})}{\kappa_1\sqrt{n}}\left(1+\frac{5\log\log n}{\log n}\right)\sum_{k>\sqrt{n}/\log^2 n} k\,p_k \\
&\leq \omega(\kappa_1\sqrt{n})\left(1+o(1)\right)\,.
\end{aligned}$$

It also follows from the monotonicity of $\omega(n)(\log n)/n$ that

$$\frac{\omega(n-k-j)}{n-k-j} \quad\leq\quad \frac{\omega(n)}{n}\frac{\log n}{\log(n-k-j)} \leq \frac{\omega(n)}{n}\left(1+\frac{(k+j)/(n-k-j)}{\log(n-k-j)}\right) \tag{9.10}$$

Moreover, Lemma 9.3 implies that

$$\sum_{j\geq 1}(n-k-j)p_j = (n-k) - \kappa_1\sqrt{n-k}(1+o(1)) \leq \frac{n-\kappa_1\sqrt{n}(1-o(1))}{n}(n-k) \tag{9.11}$$

and that

$$\sum_{k\geq 1}\left(\sum_{j\geq 1}\frac{k+j}{\log(n-k-j)}p_j\right)p_k \leq \frac{2}{\log n}\sum_{k\geq 1}\left(\sum_{j\geq 1}(k+j)p_j\right)p_k$$
$$= \frac{2}{\log n}2\kappa_1\sqrt{n}(1+o(1))\,.$$

Therefore we obviously have

$$\sum_{k\geq 1}\left(\sum_{j\geq 1}\frac{k+j}{\log(n-k-j)}p_j\right)p_k = O\left(\frac{\sqrt{n}}{\log n}\right)\,. \tag{9.12}$$

Putting everything together, we obtain

$$\begin{aligned}
\sum_{k\geq 1}\overline{\omega}(n-k)p_k &\overset{(9.7)}{=} \sum_{k\geq 1}\left(\sum_{j\geq 1}\omega(n-k-j)p_j\right)p_k + O\left(\frac{\sqrt{n}}{\log^2 n}\right)\\
&\overset{(9.10)}{\leq} \frac{\omega(n)}{n}\sum_{k\geq 1}\left(\sum_{j\geq 1}\left((n-k-j)+\frac{k+j}{\log(n-k-j)}\right)p_j\right)p_k + O\left(\frac{\sqrt{n}}{\log^2 n}\right)\\
&\overset{(9.12)}{\leq} \frac{\omega(n)}{n}\left(\sum_{k\geq 1}\left(\sum_{j\geq 1}(n-k-j)p_j\right)p_k + O\left(\frac{\sqrt{n}}{\log n}\right)\right)\\
&\overset{(9.11)}{\leq} \frac{\omega(n)}{n}\left(\frac{n-\kappa_1\sqrt{n}(1-o(1))}{n}\sum_{k\geq 1}(n-k)p_k + O\left(\frac{\sqrt{n}}{\log n}\right)\right)\\
&\leq \frac{\omega(n)}{n}\left(\frac{n^2-2\kappa_1\sqrt{n}n(1+o(1))}{n} + O\left(\frac{\sqrt{n}}{\log n}\right)\right)\,.
\end{aligned}$$

□

From Claim 9.4.2 and (9.8) we now derive that

$$\omega(n) \leq \frac{\omega(n)}{n}\left(\left(n-2\kappa_1\sqrt{n}(1+o(1))\right)+O\left(\frac{\sqrt{n}}{\log n}\right)\right)+\omega(\kappa_1\sqrt{n})\,(1+o(1))$$

and thus

$$\frac{2\kappa_1}{\sqrt{n}}\omega(n)(1+o(1)) \leq \omega(\kappa_1\sqrt{n})\,(1+o(1)) \tag{9.13}$$

It is easy to see that (9.13) does not hold for any function ω with

$$\omega(n) = \Omega\left(\frac{n}{\log^{1-\varepsilon} n}\right)$$

if $\varepsilon > 0$.

Moreover, $\alpha(n) = \overline{\omega}(n) \leq \omega(n) = \overline{\alpha}(n)$. In other words, the expected size of a maximum stable set in a connected cograph equals the expected size of a maximum clique in a disconnected cograph which is not larger than the expected size of a maximum clique in a connected cograph which in turn equals the expected size of a maximum stable set in a disconnected cograph. Summing up we conclude that for every $\varepsilon > 0$ there is C such that

$$\max\{\omega(n), \alpha(n)\} \leq C\frac{n}{\log^{1-\varepsilon} n}\,. \qquad \square$$

The methods employed above can also be used to prove the following lower bound on the expected value for the size of a largest homogeneous sets in a random cograph. For every $\varepsilon > 0$ there is a positive constant C such that

$$\mathbb{E}[\hom(G)] \geq C\frac{n}{\log^{1+\varepsilon} n}$$

if G is drawn uniformly at random from the set of all cographs on vertex set $[n]$.

10 Concluding remarks

The results presented in this thesis can be grouped into three different topics. The first containing the arrangeable Blow-up Lemma and its applications from Chapter 4 and Chapter 5 revolves around the embedding of spanning subgraphs with growing degrees into (super-)regular pairs. We have provided a variety of different embedding results that can be deduced from our generalised Blow-up Lemma for arrangeable graphs. Central tools in these proofs are related to the Regularity Lemma and the probabilistic method.

The second topic of this thesis is large planar subgraphs in dense graphs. This subject differs from the previous in two aspects. On the one hand we focus on a particular class of target graphs – planar graphs – and on the other hand we do not even stipulate a fixed target graph. Instead we try to maximize a graph parameter (the edge number) by embedding an a priory undetermined subgraph into the dense host graph. The arguments employed in the corresponding Chapter 6 and Chapter 7 are algorithmic and deterministic.

Thirdly, we study the typical structure of two classes of graphs. We determine the size of homogeneous sets in almost all graphs that do not contain an induced copy of C_5 or an induced copy of P_4 respectively. The setting of those problems is almost identical, however, we deploy completely different methods for the two cases. We apply the regularity method for C_5-free graphs in Chapter 8 while P_4-free graphs (cographs) are treated with the generating function approach in Chapter 9. In both cases we determine asymptotic bounds for the size of a largest homogeneous set in a typical element of the class.

In this final chapter of the thesis we give an outlook on future directions of research related to the three topics presented within the thesis.

Blow-up Lemma for arrangeable graphs Our Blow-up Lemma for arrangeable graphs applies to target graphs of maximum degree up to $\sqrt{n}/\log n$ where n is the order of the host graph. As we show in Section 4.5 this is optimal up to the log-factor. Also the constant number of permissible image restrictions per cluster is best possible. Hence it seems that not much more can be said in the standard graph setting.

Instead one might consider sparse versions of the (almost spanning) arrangeable Blow-up Lemma and of corresponding applications. Böttcher, Kohayakawa, and Taraz [15] have proved a bipartite version of their bandwidth theorem for dense subgraphs of the random graph $\mathcal{G}(n, p)$. Later Huang, Lee, and Sudakov [53] have given a sparse version of the bandwidth theorem for arbitrary chromatic numbers. A sparse, arrangeable version of the bandwidth theorem would be a canonical generalisation of the previously mentioned results.

One could also try to further weaken the assumptions of the Blow-up Lemma in the dense case. We did not find evidence as to why a degenerate version of the Blow-up Lemma should not be true. However, the properties of an arrangeable ordering are crucial for our proof of the arrangeable Blow-up Lemma. Thus our approach seems hopeless for the more general class of degenerate graphs. This is little surprising as any embedding result capable of embedding degenerate graphs of size linear in the size of the host graph would prove the Conjecture of Burr and Erdős [20]. Still it might be interesting to determine the maximum size of a subgraph embeddable with the methods of Chapter 4.

Planar subgraphs The parameter $\mathrm{pl}(n,d)$ is defined to be the largest integer k such that every graph with n vertices and minimum degree d contains a planar subgraph with at least k edges. The behaviour of $\mathrm{pl}(n,d)$ is well understood for values $d=\gamma n$ with $\gamma \in (0,1/2)$ or $\gamma \in (1/2,1]$. We have proved the existence of a jump in the evolution of $\mathrm{pl}(n,d)$ for d slightly above $n/2$. It would be interesting to have tight versions of Theorem 6.2 and Theorem 6.3. We believe that indeed

$$\mathrm{pl}(2m-1,m) \geq 4.5m-4, \qquad \mathrm{pl}(2m,m+1) \geq 5m-4\,.$$

However, those sharp results seem to require more diligence and/or different tools than our regularity based approach.

Another direction of research is to generalise the parameter $\mathrm{pl}(n,d)$ from planar subgraphs to subgraphs from other monotone properties. One such question could be: "What is the maximum number of edges a subgraph can have without containing a fixed minor?"

Homogeneous sets The maximum size of a homogeneous set in a typical graph without a forbidden induced subgraph F has been known for basically all graphs F. We have settled one of the remaining open cases $F=P_4$. The class of P_4-free graphs turns out to be one of the few examples where not almost all graphs without an induced copy of F do have linearly sized homogeneous sets. For $F=C_5$ we know that almost all F-free graphs have linearly sized homogeneous sets even if we restrict the class to graphs of a specific density. There does not seem to be an easy way to transfer this result to other graphs F. In the case of $F=C_7$, e.g., it already is a veritable task to determine the number of graphs with a given density. We wonder whether it is true for every $c \in (0,1)$ and every F that almost all F-free graphs of density c have a linearly sized homogeneous set as long as almost all F-free graphs have a linearly sized homogeneous set.

For the class of cographs ($F=P_4$) our methods give a lot more structural information about the typical elements of the class. Hence we believe that other graph parameters can be determined with similar arguments. It would also be interesting to determine the exact ξ such that almost all cographs on n vertices have $\hom(G)=(\xi+o(1))n/\log n$ if such a ξ does exist.

Acknowledgements

The past four years have been a huge, exciting endeavour. A number of people in Munich, São Paulo, London, and other places have shaped this thesis, my work and my life. I am indebted to them for support, inspiration, knowledge, and fun.

My appreciation goes to all members of M9 at TU München for providing a stimulating work environment. The discussions over lunch, coffee in the afternoon, and the occasional cake we shared have been very dear to me. In particular, I want to thank Jan (Honza) Hladky, Oliver Cooley, and Wei Huang who have shared an office with me at TU München. All three of them have been great office mates and a quick chat across our desks has always been a pleasure.

I want to thank the following institutions for a warm welcome, friendly support and a productive work environment: Instituto de Matématica e Estatístia at USP in São Paulo, Department of Mathematics at London School of Economics, and Department of Mathematics and Computer Science at Emory University in Atlanta.

My gratitude belongs to Michael Ritter, Christian Ludwig and Georg Wechslberger for technical support regarding, but not limited to, LaTeX and Linux/Unix.

I owe thanks to my coauthors Peter Allen, Julia Böttcher, Oliver Cooley, Carlos Hoppen, Yoshiharu Kohayakawa, Tomasz Łuczak, Marc Noy, Jozef Skokan, and Anusch Taraz for endless discussions, plenty of inspiration, and their priceless encouragement whenever I felt distressed about our joint projects. They have taught me everything I know about good scientific writing.

I am especially grateful to my family, for being supportive in every non-mathematical way I could imagine. It really helped me to know that they approved of what I was doing, even though they maybe did not understand why I was doing it. My family has provided an indispensable retreat whenever this thesis seemed to be hopelessly stuck.

I could not have done without the support of Veronika Bogner. Her affection and encouragement always helped me to put things into perspective.

Most thanks, however, go to Julia Böttcher and to my supervisor Anusch Taraz. The two of them have become what I would call my academic parents. With them I did my first steps in mathematical research as an undergraduate student starting with a study project in early 2007 and my Diploma thesis in 2008. They accompanied my journey over the four years of this thesis. Their guidance and support has made possible everything I have achieved.

Julia Böttcher and Anusch Taraz both have committed countless hours to help me find intuition, shape ideas, work out proofs, and settle details. Their optimism and composure have been essential whenever I got desperate. Their humorous way has made the whole endeavour as joyful as it has been.

Bibliography

[1] V. E. Alekseev. Range of values of entropy of hereditary classes of graphs. *Discrete Math. Appl.*, 3(2):191–199, 1993.

[2] P. Allen, G. Brightwell, and J. Skokan. Ramsey-goodness – and otherwise. http://arxiv.org/abs/1010.5079v1.

[3] P. Allen, J. Skokan, and A. Würfl. Maximum planar subgraphs in dense graphs. submitted, 2012.

[4] N. Alon, J. Balogh, B. Bollobás, and R. Morris. The structure of almost all graphs in a hereditary property. *Combin. Theory Ser. B*, 101(2):85–110, 2011.

[5] N. Alon and M. Capalbo. Sparse universal graphs for bounded-degree graphs. *Random Struct. Algorithms*, 31(2):123–133, 2007.

[6] N. Alon, R. A. Duke, H. Lefmann, V. Rödl, and R. Yuster. The algorithmic aspects of the regularity lemma. *J. Algorithms*, 16(1):80–109, 1994.

[7] N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000.

[8] N. Alon and A. Shapira. A characterization of the (natural) graph properties testable with one-sided error. *SIAM J. Comput.*, 37(6):1703–1727, 2008.

[9] N. Alon and R. Yuster. H-factors in dense graphs. *J. Combin. Theory Ser. B*, 66(2):269–282, 1996.

[10] K. Appel and W. Haken. The solution of the four-color-map problem. *Sci. Amer.*, 237(4):108–121, 152, 1977.

[11] J. Balogh, B. Bollobás, and M. Simonovits. The number of graphs without forbidden subgraphs. *J. Comb. Theory, Ser. B*, 91(1):1–24, 2004.

[12] B. Bollobás. *Extremal graph theory.* Dover Publications Inc., Mineola, NY, 2004. Reprint of the 1978 original.

[13] B. Bollobás and A. Thomason. Hereditary and monotone properties of graphs. In *The mathematics of Paul Erdős, II*, volume 14 of *Algorithms Combin.*, pages 70–78. Springer, Berlin, 1997.

[14] B. Bollobás and A. Thomason. The structure of hereditary properties and colourings of random graphs. *Combinatorica*, 20(2):173–202, 2000.

[15] J. Böttcher, Y. Kohayakawa, and A. Taraz. Almost spanning subgraphs of random graphs after adversarial edge removal. In *LAGOS'09—V Latin-American Algorithms, Graphs and Optimization Symposium*, volume 35 of *Electron. Notes Discrete Math.*, pages 335–340. Elsevier Sci. B. V., Amsterdam, 2009.

[16] J. Böttcher, Y. Kohayakawa, A. Taraz, and A. Würfl. A Blow-up Lemma for growing degrees. in preparation, 2012.

[17] J. Böttcher, K. P. Pruessmann, A. Taraz, and A. Würfl. Bandwidth, expansion, treewidth, separators and universality for bounded-degree graphs. *Eur. J. Comb.*, 31(5):1217–1227, 2010.

[18] J. Böttcher, M. Schacht, and A. Taraz. Proof of the bandwidth conjecture of Bollobás and Komlós. *Mathematische Annalen*, 343(1):175–205, 2009.

[19] J. Böttcher, A. Taraz, and A. Würfl. Spanning embeddings of arrangeable graphs with sublinear bandwidth. in preparation, 2012.

[20] S. A. Burr and P. Erdős. On the magnitude of generalized Ramsey numbers for graphs. In *Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. 1*, pages 215–240. Colloq. Math. Soc. János Bolyai, Vol. 10. North-Holland, Amsterdam, 1975.

[21] J. Böttcher, A. Taraz, and A. Würfl. Perfect graphs of fixed density: Counting and homogeneous sets. *Combinatorics, Probability and Computing*, 21(5):661–682, 2012.

[22] G. Chen and R. H. Schelp. Graphs with linearly bounded ramsey numbers. *J. Comb. Theory, Ser. B*, 57(1):138–149, 1993.

[23] M. Chudnovsky, N. Robertson, P. Seymour, and R. Thomas. The strong perfect graph theorem. *Annals of Mathematics*, 164:51–229, 2006.

[24] M. Chudnovsky and S. Safra. The Erdős-Hajnal conjecture for bull-free graphs. *J. Combin. Theory Ser. B*, 98(6):1301–1310, 2008.

[25] V. Chvátal, V. Rödl, E. Szemerédi, and W. T. Trotter, Jr. The Ramsey number of a graph with bounded maximum degree. *J. Combin. Theory Ser. B*, 34(3):239–243, 1983.

[26] D. Conlon and J. Fox. Bounds for graph regularity and removal lemmas, 2011. http://arxiv.org/abs/1107.4829.

[27] O. Cooley, T. Łuczak, A. Taraz, and A. Würfl. Large planar subgraphs in dense graphs. in preparation, 2012.

[28] D. G. Corneil, H. Lerchs, and L. S. Burlingham. Complement reducible graphs. *Discrete Appl. Math.*, 3(3):163–174, 1981.

[29] D. G. Corneil, Y. Perl, and L. K. Stewart. A linear recognition algorithm for cographs. *SIAM J. Comput.*, 14(4):926–934, 1985.

[30] K. Corrádi and A. Hajnal. On the maximal number of independent circuits in a graph. *Acta Math. Acad. Sci. Hungar.*, 14:423–439, 1963.

[31] B. Csaba and A. Pluhár. Weighted regularity lemma with applications, 2011. http://arxiv.org/abs/0907.0245.

[32] A. Czygrinow and V. Rödl. An algorithmic regularity lemma for hypergraphs. *SIAM J. Comput.*, 30(4):1041–1066, 2000.

[33] R. Diestel. *Graph Theory.* Springer, Berlin, 2005.

[34] G. A. Dirac. Some theorems on abstract graphs. *Proc. London Math. Soc. (3)*, 2:69–81, 1952.

[35] R. A. Duke, H. Lefmann, and V. Rödl. A fast approximation algorithm for computing the frequencies of subgraphs in a given graph. *SIAM J. Comput.*, 24(3):598–620, 1995.

[36] P. Erdős, P. Frankl, and V. Rödl. The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent. *Graphs Comb.*, 2:113–121, 1986.

[37] P. Erdős and A. Hajnal. Ramsey-type theorems. *Discrete Appl. Math.*, 25(1-2):37–52, 1989.

[38] P. Erdős and A. H. Stone. On the structure of linear graphs. *Bull. Amer. Math. Soc.*, 52:1087–1091, 1946.

[39] P. Erdős. Some remarks on the theory of graphs. *Bull. Amer. Math. Soc.*, 53:292–294, 1947.

[40] P. Erdős. Graph theory and probability. *Canad. J. Math.*, 11:34–38, 1959.

[41] P. Erdős. Problem 9. In M. Fiedler, editor, *In M. Fiedler, editor, Theory of graphs and its applications (Proceedings of the Symposium held in Smolenice in June 1963)*, page 234. Publishing House of the Czechoslovak Academy of Sciences, Prague, 1964.

[42] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[43] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.

[44] P. Erdős and H. Sachs. Reguläre Graphen gegebener Taillenweite mit minimaler Knotenzahl. *Wiss. Z. Uni. Halle (Math. Nat.)*, 12:251–257, 1963.

[45] P. Erdős and M. Simonovits. A limit theorem in graph theory. *Studia Sci. Math. Hungar*, 1:51–57, 1966.

[46] P. Flajolet, B. Salvy, and P. Zimmermann. Automatic average-case analysis of algorithms. *Theoret. Comput. Sci.*, 79(1, (Part A)):37–109, 1991. Algebraic and computing treatment of noncommutative power series (Lille, 1988).

[47] X. Gourdon. Largest component in random combinatorial structures. In *Proceedings of the 7th Conference on Formal Power Series and Algebraic Combinatorics (Noisy-le-Grand, 1995)*, volume 180, pages 185–209, 1998.

[48] A. Gyárfás. Reflections on a problem of Erdös and Hajnal. In R. L. Graham and J. Nešetřil, editors, *The Mathematics of Paul Erdös II*, volume 14 of *Algorithms and Combinatorics*, pages 93–98. Springer, Berlin, Germany, 1997.

[49] A. Hajnal and E. Szemerédi. Proof of a conjecture of P. Erdős. In *Combinatorial theory and its applications, II (Proc. Colloq., Balatonfüred, 1969)*, pages 601–623. North-Holland, Amsterdam, 1970.

[50] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities.* Cambridge Mathematical Library. Cambridge University Press, 1988. Reprint of the 1952 edition.

[51] P. J. Heawood. Map-colour theorem. *Quart. J. Math.*, 24:332–338, 1890.

[52] C. Hoppen, M. Noy, and A. Würfl. Homogeneous sets in cographs. in preparation.

[53] H. Huang, C. Lee, and B. Sudakov. Bandwidth theorem for random graphs. *J. Combin. Theory Ser. B*, 102(1):14–37, 2012.

[54] S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs.* John Wiley and Sons, 2000.

[55] S. Jukna. *Extremal Combinatorics.* Springer, June 2001.

[56] R. J. Kang, C. McDiarmid, B. Reed, and A. Scott. For most graphs H, most H-free graphs have a linear homogeneous set. submitted.

[57] H. A. Kierstead and W. T. Trotter. Planar graph coloring with an uncooperative partner. In *Planar graphs (New Brunswick, NJ, 1991)*, volume 9 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 85–93. Amer. Math. Soc., Providence, RI, 1993.

[58] J. Komlós. The Blow-up Lemma. *Combin. Probab. Comput.*, 8(1-2):161–176, 1999. Recent trends in combinatorics (Mátraháza, 1995).

[59] J. Komlós. Tiling Turán theorems. *Combinatorica*, 20(2):203–218, 2000.

[60] J. Komlós, G. N. Sárközy, and E. Szemerédi. Proof of a packing conjecture of Bollobás. *Combinatorics, Probability & Computing*, 4:241–255, 1995.

[61] J. Komlós, G. N. Sárközy, and E. Szemerédi. On the square of a Hamiltonian cycle in dense graphs. In *Proceedings of the Seventh International Conference on Random Structures and Algorithms (Atlanta, GA, 1995)*, volume 9, pages 193–211, 1996.

[62] J. Komlós, G. N. Sárközy, and E. Szemerédi. Blow-up Lemma. *Combinatorica*, 17(1):109–123, 1997.

[63] J. Komlós, G. N. Sárközy, and E. Szemerédi. An algorithmic version of the Blow-up Lemma. *Random Structures Algorithms*, 12(3):297–312, 1998.

[64] J. Komlós, G. N. Sárközy, and E. Szemerédi. On the Pósa-Seymour conjecture. *J. Graph Theory*, 29(3):167–176, 1998.

[65] J. Komlós, G. N. Sárközy, and E. Szemerédi. Proof of the Seymour conjecture for large graphs. *Ann. Comb.*, 2(1):43–60, 1998.

[66] J. Komlós, G. N. Sárközy, and E. Szemerédi. Proof of the Alon-Yuster conjecture. *Discrete Math.*, 235(1-3):255–269, 2001. Combinatorics (Prague, 1998).

[67] J. Komlós, G. N. Sárközy, and E. Szemerédi. Spanning trees in dense graphs. *Combin. Probab. Comput.*, 10(5):397–416, 2001.

[68] J. Komlós, A. Shokoufandeh, M. Simonovits, and E. Szemerédi. The regularity lemma and its applications in graph theory. In G. B. Khosrovshahi, A. Shokoufandeh, and M. A. Shokrollahi, editors, *Theoretical Aspects of Computer Science*, volume 2292 of *Lecture Notes in Computer Science*, pages 84–112, New York, NY, USA, 2000. Springer-Verlag New York, Inc.

[69] J. Komlós and M. Simonovits. Szemerédi's regularity lemma and its applications in graph theory. In *Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993)*, volume 2 of *Bolyai Soc. Math. Stud.*, pages 295–352. János Bolyai Math. Soc., Budapest, 1996.

[70] A. V. Kostochka, K. Nakprasit, and S. V. Pemmaraju. On equitable coloring of d-degenerate graphs. *SIAM J. Discrete Math.*, 19:83–95, 2005.

[71] D. Kühn and D. Osthus. Spanning triangulations in graphs. *J. Graph Theory*, 49(3):205–233, 2005.

[72] D. Kühn and D. Osthus. Embedding large subgraphs into dense graphs. In *Surveys in combinatorics 2009*, volume 365 of *London Math. Soc. Lecture Note Ser.*, pages 137–167. Cambridge Univ. Press, Cambridge, 2009.

[73] D. Kühn and D. Osthus. The minimum degree threshold for perfect graph packings. *Combinatorica*, 29:65–107, 2009.

[74] D. Kühn, D. Osthus, and A. Taraz. Large planar subgraphs in dense graphs. *J. Combin. Theory Ser. B*, 95(2):263–282, 2005.

[75] M. Loebl, B. Reed, A. Scott, A. Thomason, and S. Thomassé. Almost all F-free graphs have the Erdős-Hajnal property. In *An Irregular Mind*, volume 21 of *Bolyai Society Mathematical Studies*, pages 405–414. Springer, 2010.

[76] E. Marchant and A. Thomason. The structure of hereditary properties and 2-coloured multigraphs. *Combinatorica*, 31(1):85–93, 2011.

[77] C. McDiarmid and B. Reed. On the maximum degree of a random planar graph. *Combin. Probab. Comput.*, 17(4):591–601, 2008.

[78] V. Nikiforov. Some new results in extremal graph theory. `arXiv:1107.1121 [math.CO]`.

[79] H. J. Prömel and A. Steger. Excluding induced subgraphs: Quadrilaterals. *Random Structures Algorithms*, 2(1):55–72, 1991.

[80] H. J. Prömel and A. Steger. Almost all Berge graphs are perfect. *Combinatorics, Probability & Computing*, 1:53–79, 1992.

[81] H. J. Prömel and A. Steger. Excluding induced subgraphs. III. A general asymptotic. *Random Structures Algorithms*, 3(1):19–31, 1992.

[82] F. P. Ramsey. On a problem of formal logic. *Proceedings of the London Mathematical Society*, 30(1):361–376, 1930.

[83] G. Ringel and J. W. T. Youngs. Solution of the Heawood map-coloring problem. *Proc. Nat. Acad. Sci. U.S.A.*, 60:438–445, 1968.

[84] O. Riordan. Spanning subgraphs of random graphs. *Combin. Probab. Comput.*, 9(2):125–148, 2000.

[85] N. Robertson, D. Sanders, P. Seymour, and R. Thomas. The four-colour theorem. *J. Combin. Theory Ser. B*, 70(1):2–44, 1997.

[86] V. Rödl and A. Ruciński. Perfect matchings in ε-regular graphs and the blow-up lemma. *Combinatorica*, 19(3):437–452, 1999.

[87] V. Rödl and R. Thomas. Arrangeability and clique subdivisions. In *The mathematics of Paul Erdős, II*, volume 14 of *Algorithms Combin.*, pages 236–239. Springer, Berlin, 1997.

[88] G. Sági. On induced subgraphs of finite graphs not containing large empty and complete subgraphs, 2012. http://arxiv.org/abs/1211.3876.

[89] P. Seymour. Problem section. In *T. P. McDonough and V. C. Mavron, editors, Combinatorics (Proc. British Combinatorial Conf. 1973)*, number 13 in London Math. Soc. Lecture Note Ser., pages 201–204. Cambridge Univ. Press, London, 1974.

[90] E. Szemerédi. Regular partitions of graphs. In *Problèmes Combinatoires et Théorie des Graphes (Orsay, 1976)*, volume 260 of *Colloques Internationaux CNRS*, pages 399–401. CNRS, 1978.

[91] A. Thomason. Graphs, colours, weights and hereditary properties. In *Surveys in Combinatorics 2011*, volume 392 of *LMS Lecture Note Series*, pages 333–364. Cambridge University Press, 2011.

[92] P. Turán. Eine Extremalaufgabe aus der Graphentheorie. *Mat. Fiz. Lapok*, 48:436–452, 1941.

[93] A. Würfl. Über die Struktur von Graphen, die keinen induzierten C_5 enthalten (On the structure of graphs without an induced copy of C_5). Master Thesis, TU Munich, 2008.

Index

Symbols

A