Embedding Problems in Graphs and Hypergraphs

Eng Keat Hng

A thesis submitted for the degree of
Doctor of Philosophy

Department of Mathematics

The London School of Economics
and Political Science

April 2022

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that this thesis consists of 64121 words.

Statement of Co-authored Work

I confirm that Section 1.3 and Chapter 3 are based on [16], which is jointly authored with Oliver Cooley, Frederik Garbe, Mihyun Kang, Nicolás Sanhueza-Matamala and Julian Zalla.

I confirm that Section 1.4 and Chapter 4 are based on [3], which is jointly authored with Peter Allen, Julia Böttcher, Ewan Davies and Jozef Skokan.

Abstract

In this thesis, we explore several mathematical questions about substructures in graphs and hypergraphs, focusing on algorithmic methods and notions of regularity for graphs and hypergraphs.

We investigate conditions for a graph to contain powers of paths and cycles of arbitrary specified linear lengths. Using the well-established graph regularity method, we determine precise minimum degree thresholds for sufficiently large graphs and show that the extremal behaviour is governed by a family of explicitly given extremal graphs. This extends an analogous result of Allen, Böttcher and Hladký for squares of paths and cycles of arbitrary specified linear lengths and confirms a conjecture of theirs.

Given positive integers k and j with $j < k$, we study the length of the longest j-tight path in the binomial random k-uniform hypergraph $H^k(n,p)$. We show that this length undergoes a phase transition from logarithmic to linear and determine the critical threshold for this phase transition. We also prove upper and lower bounds on the length in the subcritical and supercritical ranges. In particular, for the supercritical case we introduce the `Pathfinder` algorithm, a depth-first search algorithm which discovers j-tight paths in a k-uniform hypergraph. We prove that, in the supercritical case, with high probability this algorithm finds a long j-tight path.

Finally, we investigate the embedding of bounded degree hypergraphs into large sparse hypergraphs. The blow-up lemma is a powerful tool for embedding bounded degree spanning subgraphs with wide-ranging applications in extremal graph theory. We prove a sparse hypergraph analogue of the blow-up lemma, showing that large sparse partite complexes with sufficiently regular small subcomplex counts and no atypical vertices behave as if they were complete for the purpose of embedding complexes with bounded degree and bounded partite structure.

Acknowledgements

I would like to thank my supervisors Julia Böttcher and Peter Allen for their guidance, support and constant encouragement. It has been my pleasure and privilege to work with you. Thank you for the fascinating mathematical problems, for all the discussions and for your time and patience. I am grateful to Jozef Skokan, my third supervisor in all but name, for the valuable wisdom, motivation and insights.

I would like to thank everyone in the Department of Mathematics at LSE. In particular, I am grateful to Kate Barker, Enfale Farooq, Sarah Massey, Edward Perrin and Rebecca Lumb for the efficient administrative support. I am thankful to my PhD colleagues for the mathematical discussions and fun mathematical puzzles, for making the PhD office a great working environment and for all the enjoyable times we had at LSE.

I would like thank all my friends, who made challenging times all the more bearable. Finally, I would like to thank my parents and my brothers for their love and constant support.

Contents

Chapter 1

Introduction

How do local properties, global properties and combinatorial parameters interact in graphs and hypergraphs? In this thesis we explore problems which investigate features that enable the presence of certain substructures in graphs and hypergraphs. Such problems are termed *embedding problems* because they may be framed as quests to embed a given structure into a combinatorial object with the relevant attributes. As a major theme in graph and hypergraph theory, the study of embedding problems has provoked many important conjectures and produced a multitude of elegant and influential results in discrete mathematics.

Embedding problems are ubiquitous in the study of graphs and hypergraphs. In extremal graph theory, one is primarily concerned with optimising the value of some parameter over a certain collection of graphs. A classical result in this direction is Turán's theorem [57], which investigates the maximum number of edges in a graph without a copy of the complete graph K_ℓ on ℓ vertices. By recasting this question as asking how many edges are required to guarantee that K_ℓ may be embedded, Turán's theorem may be viewed as a quintessential example of an embedding result in extremal graph theory. More generally, the study of conditions for the appearance of interesting subgraphs in a so-called host graph lies at the heart of extremal graph theory and has led to many prominent results such as Dirac's theorem [20] and the Erdős–Stone theorem [24]. In Chapter 2 we study conditions for the appearance of subgraphs known as powers of paths and cycles.

Embedding problems also feature prominently in the theory of random graphs. In its modern form, the celebrated phase transition result of Erdős and Rényi [23] for random graphs states that the order of the largest connected component in the binomial random graph on n vertices undergoes a strikingly abrupt change when p is in the region of $1/n$: with high probability all components are of at most logarithmic order when p is slightly smaller than $1/n$, but very soon after p exceeds $1/n$, a positive fraction of this 'sea' of components of logarithmic order swiftly coalesces to form a unique 'giant' component of linear order. While the result deals with component behaviour rather more comprehensively, we may reformulate the specific question about the order of the largest component as the embedding problem for a tree of linear size and here $1/n$ represents the so-called *threshold* for this embedding problem. More generally, the study of the threshold for the emergence of certain subgraphs is a major theme in the theory of random graphs, with a wide array of results for a variety of interesting subgraphs. In Chapter 3 we study the emergence of long paths in the binomial random k-uniform hypergraph.

Any story about embedding problems would be incomplete without the mention of systematic approaches to their study, exemplified by the development of the regularity method. This is a broad systematic framework for embedding combinatorial structures that stems from the celebrated regularity lemma of Szemerédi [56]. Originally motivated by a question about arithmetic progressions, Szemerédi's regularity lemma has been instrumental in the resolution of many long-standing open problems in extremal graph theory and has inspired breakthroughs in many areas of combinatorics. A classical implementation of the regularity method involves the joint application of Szemerédi's regularity lemma and the blow-up lemma of Komlós, Sárközy and Szemerédi [38], which is a tool for embedding bounded degree spanning subgraphs. In Chapter 4 we extend the regularity method by developing a blow-up lemma for sparse hypergraphs.

We outline how the remainder of this chapter is organised. In Section 1.1 we collect some common notation and terminology that we use throughout this thesis. In Section 1.2 we introduce the relevant background for the study of minimum degree conditions for powers of paths and cycles and state the main result, Theorem 1.5, that we prove in Chapter 2. In Section 1.3 we discuss the study of paths in random graphs and introduce notions of paths in hypergraphs to motivate our investigation of the emergence of long paths in random hypergraphs in Chapter 3. In Section 1.4 we motivate the development of a sparse hypergraph blow-up lemma in Chapter 4.

1.1 Notation

In this section we introduce some notation. Write $\mathbb{N}$ for the set of positive integers and $\mathbb{N}_0$ for the set $\mathbb{N} \cup \{0\}$. For $a \in \mathbb{N}_0$ write $[a]$ for the set $\{1, \ldots, a\}$ and $[a]_0$ for $[a] \cup \{0\}$. For a set S and $m \in \mathbb{N}_0$ let $\binom{S}{m}$ denote the set of subsets of S of size m. For positive real numbers x and y, we write $x \pm y$ to mean an appropriate real number z satisfying $x - y \leq z \leq x + y$. For example, for positive real numbers w, x, y and a the equation $w = (x \pm y)a$ would mean $(x - y)a \leq w \leq (x + y)a$. For an event $\mathcal{E}$ we write $\mathbf{1}_{\mathcal{E}}$ to denote its indicator function.

Let G be a graph. Write $V(G)$ and $E(G)$ for the vertex set and edge set of G respectively. Let $v(G) := |V(G)|$ and $e(G) := |E(G)|$. For sets $X, Y \subseteq V(G)$, let $E(X, Y) := \{xy \in E(G) : x \in X, y \in Y\}$ and $e(X, Y) := |E(X, Y)|$. Let $G[X]$ denote the subgraph of G *induced* by X. For a vertex $v \in V(G)$ and a subset $A \subseteq V(G)$, the *neighbourhood* $\Gamma_G(v; A)$ in A of v in G is the set of *neighbours* in A of v, that is, the vertices $u \in A$ such that $uv \in E(G)$. We will sometimes write $N_G(v; A)$ instead of $\Gamma_G(v; A)$. The *degree* $\deg_G(v; A)$ in A of v is $|\Gamma_G(v; A)|$. Given a subset $X \subseteq V(G)$ let $\Gamma_G(X; A) := \bigcap_{v \in X} \Gamma_G(v; A)$ denote the *common neighbourhood* in A of the vertices of X in G and write $\deg_G(X; A)$ for its cardinality $|\Gamma_G(X; A)|$. Given a subset $B \subseteq V(G)$ we write $N_G(B; A) := \bigcup_{v \in B} N_G(v; A)$ for the *joint neighbourhood* in A of the vertices of B in G. We will omit the set brackets in $\Gamma_G(\{v_1, \ldots, v_\ell\}; A)$ and $\deg_G(\{v_1, \ldots, v_\ell\}; A)$, and write $\Gamma_G(v_1, \ldots, v_\ell; A)$ and $\deg_G(v_1, \ldots, v_\ell; A)$ respectively instead. We will often drop the graph G in the subscripts when it is clear from context and omit the set A when we intend $A = V(G)$. The *minimum degree* of G is $\delta(G) := \min\{\deg(v) : v \in V(G)\}$ and the *maximum degree* of G is $\Delta(G) := \max\{\deg(v) : v \in V(G)\}$. We write $v_1 \cdots v_\ell$ to denote a clique in G with vertices $v_1, \ldots, v_\ell$. We write C_ℓ (resp. P_ℓ) for a cycle (resp. path) of *length* ℓ, that is, a cycle (resp. path) on ℓ vertices.

1.2 Minimum Degree Conditions for Powers of Cycles

The study of conditions on vertex degrees in a host graph G for the appearance of a target graph H is a major theme in extremal graph theory. A classical result in this area is the following theorem of Dirac about the existence of a Hamiltonian cycle.

Theorem 1.1 (Dirac [20]). *Every graph G on $n \geq 3$ vertices with minimum degree $\delta(G) \geq \frac{n}{2}$ contains a Hamiltonian cycle.*

The kth power of a graph G, denoted by G^k, is obtained from G by joining every pair of vertices at distance at most k. In 1962, Pósa conjectured an analogue of Dirac's theorem for the containment of the square of a Hamiltonian cycle. This conjecture was extended in 1974 by Seymour to general powers of a Hamiltonian cycle.

Conjecture 1 (Pósa–Seymour Conjecture [54])**.** *Let $k \in \mathbb{N}$. Every graph G on $n \geq 3$ vertices with minimum degree $\delta(G) \geq \frac{kn}{k+1}$ contains the kth power of a Hamiltonian cycle.*

Fan and Kierstead made significant progress, proving an approximate version of this conjecture for squares of paths and squares of cycles in sufficiently large graphs [25] and determining the best-possible minimum degree condition for the square of a Hamiltonian path [26]. Komlós, Sárközy and Szemerédi confirmed the Pósa–Seymour Conjecture for sufficiently large graphs.

Theorem 1.2 (Komlós, Sárközy and Szemerédi [39])**.** *Given $k \in \mathbb{N}$, there exists $n_0 \in \mathbb{N}$ such that for all integers $n \geq n_0$, any graph G on n vertices with minimum degree $\delta(G) \geq \frac{kn}{k+1}$ contains the kth power of a Hamiltonian cycle.*

In fact, their proof asserts a stronger result, guaranteeing kth powers of cycles of all lengths divisible by $k+1$ between $k+1$ and n, in addition to the kth power of a Hamiltonian cycle. The divisibility condition is necessary as balanced complete $(k+1)$-partite graphs contain kth powers of cycles of no other length.

Theorem 1.3 (Komlós, Sárközy and Szemerédi [39])**.** *Given $k \in \mathbb{N}$, there exists $n_0 \in \mathbb{N}$ such that for all integers $n \geq n_0$, any graph G on n vertices with minimum degree $\delta(G) \geq \frac{kn}{k+1}$ contains the kth power of a cycle $C^k_{(k+1)\ell}$ for any $1 \leq \ell \leq \frac{n}{k+1}$.*

There has been interest in generalising the Pósa-Seymour Conjecture. Allen, Böttcher and Hladký [5] determined exact minimum degree thresholds for sufficiently large graphs to contain squares of paths and cycles of arbitary given linear lengths. Staden and Treglown [55] proved a degree sequence analogue for the square of a Hamiltonian cycle. There has also been a lot of related work in the hypergraph setting for many notions of degree and cycle. For example, Rödl, Ruciński and Szemerédi [49] established the minimum codegree threshold for a tight Hamiltonian cycle in k-uniform hypergraphs. For a survey of related work in the hypergraph setting see [59].

In Chapter 2 we study exact minimum degree thresholds for the appearance of kth powers of paths and cycles of arbitrary given linear lengths. One possible guess as to what minimum degree $\delta = \delta(G)$ will guarantee which length $\ell = \ell(n, \delta)$ of kth power of a path (or longest kth power of a cycle) is the following. Since the minimum degree threshold for the kth power of a *Hamiltonian* cycle (or path) is roughly the same as that for a spanning K_{k+1}-factor, perhaps this remains true for smaller ℓ. If this were true, it would mean that one could expect $\ell(n, \delta)$ to be roughly $(k+1)(k\delta - (k-1)n)$. This is characterised by $(k+1)$-partite extremal examples, which are exemplified by the $k = 3$ example in Figure 1.2a.

However, Allen, Böttcher and Hladký [5] showed that this does not give the correct answer. For the case $k = 2$, they determined sharp thresholds attained by a family of extremal graphs which exhibit not a linear dependence between the length of the longest square of a path and the minimum degree, but rather piecewise linear dependence with jumps at certain points. In order to state the result of [5] as well as our result, we first introduce the following functions. Given $k, n, \delta \in \mathbb{N}$ with $\delta \in \left(\frac{(k-1)n}{k}, n-1\right]$, we define

$$\begin{aligned} r_p(k,n,\delta) &:= \max\left\{r \in \mathbb{N} : \left\lfloor \tfrac{(k-1)\delta-(k-2)n}{r} \right\rfloor > k\delta - (k-1)n\right\} \quad \text{and} \\ r_c(k,n,\delta) &:= \max\left\{r \in \mathbb{N} : \left\lceil \tfrac{(k-1)\delta-(k-2)n}{r} \right\rceil > k\delta - (k-1)n\right\}. \end{aligned} \tag{1.1}$$

Setting $s_p(k,n,\delta) := \left\lceil \frac{(k-1)\delta-(k-2)n}{r_p(k,n,\delta)} \right\rceil$ and $s_c(k,n,\delta) := \left\lceil \frac{(k-1)\delta-(k-2)n}{r_c(k,n,\delta)} \right\rceil$, we define

$$\begin{aligned} \mathrm{pp}_k(n,\delta) &:= \min\left\{(k-1)\left(\left\lfloor \tfrac{s_p(k,n,\delta)}{2} \right\rfloor + 1\right) + s_p(k,n,\delta), n\right\} \quad \text{and} \\ \mathrm{pc}_k(n,\delta) &:= \min\left\{(k-1)\left\lfloor \tfrac{s_c(k,n,\delta)}{2} \right\rfloor + s_c(k,n,\delta), n\right\}. \end{aligned} \tag{1.2}$$

Note that $r_p(k,n,\delta)$ and $r_c(k,n,\delta)$ are almost always the same, differing only for a very small number of values of δ. Note that the functions $\mathrm{pc}_k(n,\delta)$ and $\mathrm{pp}_k(n,\delta)$ satisfy $\mathrm{pc}_k(n,\delta) \le \mathrm{pp}_k(n,\delta)$. They also behave very similarly and differ only by a constant (dependent only on k) when r_p and r_c are equal. The behaviour of $\mathrm{pp}_3(n,\delta)$ is illustrated in Figure 1.1.

Before we discuss the result of Allen, Böttcher and Hladký [5] and our result, we shall define two closely related families of graphs which will serve as examples of extremal graphs. We obtain the n-vertex graph $G_p(k,n,\delta)$ by starting with the disjoint union of $k-1$ independent sets $I_1, \dots, I_{k-1}$ and $r := r_p(k,n,\delta)$ cliques $X_1, \dots, X_r$ with $|I_1| = \cdots = |I_{k-1}| = n - \delta$ and $|X_1| \ge \cdots \ge |X_r| \ge |X_1| - 1$. Then, insert all edges between X_i and I_j for each $(i,j) \in [r] \times [k-1]$ and all edges between I_i and I_j for each

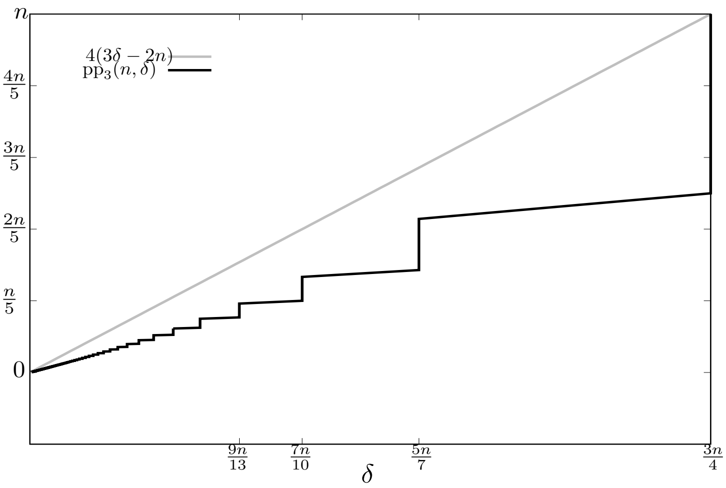
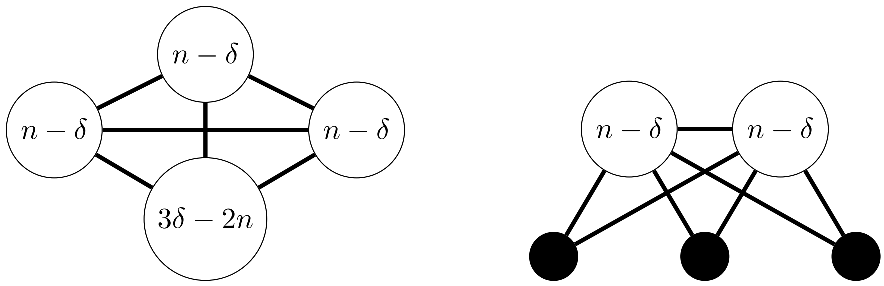


Figure 1.1: The behaviour of $\mathrm{pp}_3(n, \delta)$



(a) $K_{n-\delta,n-\delta,n-\delta,3\delta-2n}$

(b) $G_p(3, n, \delta)$

Figure 1.2: Graphs for $k = 3$

$(i, j) \in \binom{[k-1]}{2}$. This is a natural generalisation of the construction in [5]. Figure 1.2b shows an example with $k = 3$. Construct the graph $G_c(k, n, \delta)$ in the same way as $G_p(k, n, \delta)$ but with $r := r_c(k, n, \delta)$ and with additionally the arbitrary selection of a vertex $v \in X_1$ and the insertion of all edges between v and X_i for each $i \in [r]$ such that $|X_i| \neq |X_1|$.

Let us now discuss kth powers of paths and cycles in $G_p(k, n, \delta)$ and $G_c(k, n, \delta)$ respectively. We focus on the path case as the discussion for the cycle case is analogous. Consider an arbitrary kth power of a path $P_\ell^k \subseteq G_p(k, n, \delta)$ with its vertices in a natural order. Any $k + 1$ consecutive vertices form a clique, so any $k + 1$ consecutive vertices contain vertices from at most one clique X_i. Therefore, P_ℓ^k contains vertices from at most one clique X_i. Since P_ℓ^k has independence number $\left\lceil \frac{\ell}{k+1} \right\rceil$ and I_i is independent for each $i \in [k-1]$, we have $\ell - (k-1)\left\lceil \frac{\ell}{k+1} \right\rceil \leq \left\lceil \frac{(k-1)\delta-(k-2)n}{r_p(k,n,\delta)} \right\rceil$ and thus we deduce $\ell \leq \mathrm{pp}_k(n,\delta)$. Finally, observe that we can construct a copy of $P^k_{\mathrm{pp}_k(n,\delta)}$ in $G_p(k,n,\delta)$ as follows. Repeatedly take an unused vertex from I_i for each $i \in [k-1]$ and two unused vertices from X_1 in turn, until all vertices of X_1 are used and skipping I_i for each $i \in [k-1]$ if they become entirely used before X_1 does. Hence, $\mathrm{pp}_k(n,\delta)$ is the maximal length of the kth power of a path in $G_p(k,n,\delta)$. Analogously, $\mathrm{pc}_k(n,\delta)$ is the maximal length of the kth power of a cycle in $G_c(k,n,\delta)$.

The following is the result of Allen, Böttcher and Hladký [5] for the case $k=2$. It states that $\mathrm{pp}_2(n,\delta)$ and $\mathrm{pc}_2(n,\delta)$ are the maximal lengths of squares of paths and squares of cycles, respectively, guaranteed in an n-vertex graph G with minimum degree δ. Furthermore, G contains any shorter square of a cycle with length divisible by 3. These results are tight with $G_p(2,n,\delta)$ and $G_c(2,n,\delta)$ serving as extremal examples. In fact, both graphs contain squares of cycles C^2_ℓ for all lengths $3 \leq \ell \leq \mathrm{pc}_2(n,\delta)$ such that $\chi(C^2_\ell) \leq 4$. If G does not contain any one of these squares of cycles with chromatic number 4, then (ii) of Theorem 1.4 guarantees even longer squares of cycles C^2_ℓ in G, where ℓ is divisible by 3.

Theorem 1.4 (Allen, Böttcher and Hladký [5])**.** *For any $\nu > 0$ there exists $n_0 \in \mathbb{N}$ such that for all integers $n \geq n_0$ and $\delta \in [(\frac{1}{2}+\nu)n, \frac{2n-1}{3}]$ the following hold for all graphs G on n vertices with minimum degree $\delta(G) \geq \delta$.*

- *(i) $P^2_{\mathrm{pp}_2(n,\delta)} \subseteq G$ and $C^2_\ell \subseteq G$ for every $\ell \in \mathbb{N}$ with $\ell \in [3, \mathrm{pc}_2(n,\delta)]$ such that 3 divides ℓ.*
- *(ii) Either $C^2_\ell \subseteq G$ for every $\ell \in \mathbb{N}$ with $\ell \in [3, \mathrm{pc}_2(n,\delta)]$ and $\chi(C^2_\ell) \leq 4$, or $C^2_\ell \subseteq G$ for every $\ell \in \mathbb{N}$ with $\ell \in [3, 6\delta - 3n - \nu n]$ such that 3 divides ℓ.*

It was conjectured by Allen, Böttcher, and Hladký [5, Conjecture 24] that their result can be naturally generalised to higher powers. Our result states that their conjecture is indeed true. Note that $\chi(C^k_\ell) \leq k+2$ holds for all $\ell \geq k^2+k$, so this condition excludes only a number of lengths which is a function of k.

Theorem 1.5 (Hng [31])**.** *Given an integer $k \geq 3$ and $0 < \nu < 1$ there exists $n_0 \in \mathbb{N}$ such that for all integers $n \geq n_0$ and $\delta \in \left[\left(\frac{k-1}{k}+\nu\right)n, \frac{kn}{k+1}\right)$ the following hold for all graphs G on n vertices with minimum degree $\delta(G) \geq \delta$.*

- *(i) $P^k_{\mathrm{pp}_k(n,\delta)} \subseteq G$ and $C^k_\ell \subseteq G$ for every $\ell \in \mathbb{N}$ with $\ell \in [k+1, \mathrm{pc}_k(n,\delta)]$ such that $k+1$ divides ℓ.*

(ii) *Either* $C_\ell^k \subseteq G$ *for every* $\ell \in \mathbb{N}$ *with* $\ell \in [k+1, \mathrm{pc}_k(n,\delta)]$ *such that* $\chi(C_\ell^k) \leq k+2$, *or* $C_\ell^k \subseteq G$ *for every* $\ell \in \mathbb{N}$ *with* $\ell \in [k+1, (k+1)(k\delta - (k-1)n) - \nu n]$ *such that* $k+1$ *divides* ℓ.

As with the result of Allen, Böttcher and Hladký [5], our result is also tight with $G_p(k,n,\delta)$ and $G_c(k,n,\delta)$ serving as extremal examples. Note that (ii) implies the kth powers of cycles case of (i), as the latter is precisely the common part of the two cases in (ii). Hence, it will be sufficient to prove (ii) and the first part of (i). In Chapter 2 we shall prove Theorem 1.5, a result establishing exact minimum degree thresholds for the appearance of kth powers of paths and cycles of arbitrary given linear lengths.

1.3 Paths in Random Graphs and Hypergraphs

1.3.1 Paths in Random Graphs

The study of conditions for the emergence of interesting substructures is a major theme in random graph theory and has led to a number of highly influential results, including the prominent phase transition result of Erdős and Rényi [23] regarding the component profile of binomial random graphs.

While by definition any two vertices in a component are connected by a path, there is not necessarily a correlation between the order of the component and the lengths of such paths. Of course, if a component is small, then it can only contain short paths, but if a component is large, this does not guarantee the existence of a long path. Nevertheless, Ajtai, Komlós and Szemerédi [1] showed that if p is larger than $1/n$, then with high probability the binomial random graph does indeed contain a path of linear length.

We write whp as the shortened form of "with high probability", which means with probability tending to 1 as n tends to infinity. We write $G(n,p)$ for the random graph on vertex set $[n]$, in which each pair of vertices is connected by an edge with probability p independently. Incorporating various extensions of the results of Erdős and Rényi [23] and of Ajtai, Komlós and Szemerédi [1] by Pittel [48], by Łuczak [45], and by Kemkes and Wormald [34], gives the following.

Theorem 1.6. *Let* L *denote the length of the longest path in* $G(n,p)$.

(i) *If* $0 < \varepsilon < 1$ *is a constant and* $p = \frac{1-\varepsilon}{n}$, *then for any* $\omega = \omega(n)$ *such that*

$\omega \xrightarrow{n\to\infty} \infty$, *whp*

$$\frac{\ln n - \omega}{-\ln(1-\varepsilon)} \le L \le \frac{\ln n + \omega}{-\ln(1-\varepsilon)}.$$

(ii) *If* $0 < \varepsilon = \varepsilon(n) = o(1)$ *satisfies* $\varepsilon^5 n \to \infty$ *and* $p = \frac{1+\varepsilon}{n}$, *then whp*

$$\left(\frac{4}{3} + o(1)\right)\varepsilon^2 n \le L \le (1.7395 + o(1))\varepsilon^2 n.$$

Let us also note that Anastos and Frieze [10] determined L asymptotically in the range when $p = c/n$ for a sufficiently large constant c (in particular, c is much larger than 1).

1.3.2 Paths in Random Hypergraphs

Given $k \in \mathbb{N}$, a *k-uniform hypergraph* consists of a vertex set and an edge set, where each edge consists of precisely k distinct vertices. Let $H^k(n,p)$ denote the binomial random k-uniform hypergraph on vertex set $[n]$ in which each set of k distinct vertices forms an edge with probability p independently. Hence, a 2-uniform hypergraph is simply a graph and we have $H^2(n,p) = G(n,p)$.

There are several different ways of generalising the concept of paths in k-uniform hypergraphs. One important concept leads to a whole family of different types of paths which have been extensively studied. Each path type is defined by a parameter $j \in [k-1]$, which is a measure of how tightly connected the path is. Formally, we have the following definition.

Definition 1.7. Let $k, j \in \mathbb{N}$ satisfy $j \le k-1$ and let $\ell \in \mathbb{N}$. A *j-tight path of length ℓ* in a k-uniform hypergraph consists of a sequence of distinct vertices $v_1, \dots, v_{\ell(k-j)+j}$ and a sequence of edges $e_1, \dots, e_\ell$, where for $i = 1, \dots, \ell$ we have $e_i = \{v_{(i-1)(k-j)+1}, \dots, v_{(i-1)(k-j)+k}\}$, see Figure 1.3.

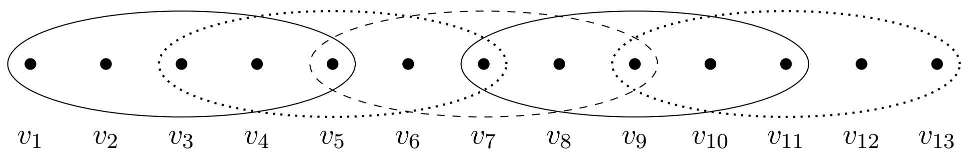


Figure 1.3: A 3-tight path of length 5 in a 5-uniform hypergraph

Note that the case $k = 2$ and $j = 1$ simply defines a path in a graph. For $k \geq 3$, the case $j = 1$ is often called a *loose path*, while the case $j = k - 1$ is often called a *tight path*.

In Chapter 3 we generalise Theorem 1.6 for j-tight paths in random *hypergraphs*. This is based on joint work in [16] with Oliver Cooley, Frederik Garbe, Mihyun Kang, Nicolás Sanhueza-Matamala and Julian Zalla. Our main result, Theorem 3.1, is a phase transition result for j-tight paths in the binomial random k-uniform hypergraph $H^k(n, p)$ which is very similar to Theorem 1.6. We determine for every pair of positive integers k and j with $j < k$ the critical threshold at which the length of the longest j-tight path transitions from being logarithmic to being linear. Furthermore, we can pin down what these lengths are both above and below the threshold. This is up to a tiny relative error below the threshold and up to a small constant factor for the majority of cases above the threshold, which is very similar to the conclusions of Theorem 1.6.

1.4 A Sparse Hypergraph Blow-up Lemma

The blow-up lemma is a powerful tool developed by Komlós, Sárközy and Szemerédi [38] for embedding spanning subgraphs of bounded degree. Roughly speaking, it says that sufficiently regular graphs with no atypical vertices behave almost like complete partite graphs for the purpose of embedding bounded degree graphs. The regularity method of combining Szemerédi's regularity lemma [56] with the blow-up lemma has led to many significant breakthroughs in extremal graph theory.

There are at least two natural directions in which to generalise the above results: to sparse host graphs and to hypergraphs. Sparse analogues of the regularity lemma were independently established by Kohayakawa [36] and Rödl, while Allen, Böttcher, Hàn, Kohayakawa and Person [4] proved blow-up lemmas for sparse pseudorandom graphs and random graphs. In the direction of hypergraphs, there are various generalisations of the regularity lemma to hypergraphs (e.g. [27], [28], [50], [51]) and Keevash [33] proved a hypergraph analogue of the blow-up lemma.

There is significant interest in combining both directions. In the study of sparse graphs, it is somewhat standard to consider suitably well-behaved graphs satisfying a pseudorandomness condition. A well-known and extensively studied pseudorandomness condition is the one known as jumbledness, which is a somewhat strong condition requiring significant control over edges between very small sets of vertices. For a variety of number theoretic applications however, it is desirable to obtain results with a weaker notion of pseudorandomness given in terms of small subgraph counts. This is sometimes referred to as a linear forms condition. Conlon, Fox and Zhao [13] proved a sparse weak regularity lemma for hypergraphs with a linear forms condition, together with a counting lemma. This was extended by Allen, Davies and Skokan [8] to a sparse regularity lemma for hypergraphs with corresponding counting and embedding lemmas.

In Chapter 4 we continue this line of research by developing a blow-up lemma for sparse hypergraphs. This is based on joint work with Peter Allen, Julia Böttcher, Ewan Davies and Jozef Skokan. Our main result, Theorem 4.5, is a sparse hypergraph blow-up lemma which roughly states that we may embed into any k-complex $\mathcal{G}$ with a balanced vertex partition any bounded degree partite k-complex H which is suitably compatible with the partition of $\mathcal{G}$ where we have typical counts and rooted counts of small partite complexes. To demonstrate how our result may be used, we apply Theorem 4.5 to prove Theorems 4.6 and 4.7, which are a result about biased Maker-Breaker games for bounded degree hypergraphs and a result on size Ramsey numbers of bounded degree hypergraphs respectively.

Chapter 2

Minimum Degree Conditions for Powers of Paths and Cycles

In this chapter we shall prove Theorem 1.5, a result establishing exact minimum degree thresholds for the appearance of kth powers of paths and cycles of arbitrary given linear lengths. We remark that while our proof of Theorem 1.5 uses the same basic strategy as used in [5] for the proof of Theorem 1.4 (that is, combining the regularity method and the stability method), our proof is not merely a generalisation of the proof of Theorem 1.4. In particular, the proof of our Stability Lemma turns out to be much more complex than in [5] and the analysis requires new insights.

This chapter is organised as follows. In Section 2.1 we introduce our tools. In Section 2.2 we outline our proof strategy and state the key lemmas in our proof. Then, we provide a proof of Theorem 1.5 which applies these lemmas. The main difficulty in our proof is proving Stability Lemma (that is, Lemma 2.13). In Section 2.3 we provide proofs for two special cases of our Stability Lemma and introduce a family of configurations which enables analysis of the general case. In Section 2.4 we analyse the aforementioned family of configurations and develop greedy-type methods, which we subsequently use in Section 2.5 in the proof of the general case of our Stability Lemma. In Sections 2.6 and 2.7 we provide proofs of our Extremal Lemma and Embedding Lemma respectively; these are applications of standard methods. Finally, we conclude this chapter with some remarks about our result and the extremal graph constructions.

2.1 Tools

In this section we provide various tools we will need and establish some useful properties of the functions introduced in (1.1) and (1.2). We begin with the following simple observations about matchings in graphs with given minimum degree.

Lemma 2.1.

(i) A graph G contains a matching with $\min\{\delta(G), \left\lfloor \frac{|V(G)|}{2} \right\rfloor\}$ *edges.*

(ii) Let $G = (U \cup V, E)$ be a bipartite graph with vertex classes U and V such that every vertex in U has degree at least u and every vertex in V has degree at least v. Then G contains a matching with $\min\{u + v, |U|, |V|\}$ *edges.*

Proof. For (i), let M be a maximum matching in G. We are done unless there are two vertices $u, v \in V(G)$ not contained in M. M is maximal so all neighbours of u and v are contained in M. There cannot be an edge $u'v'$ in M with $uv', vu' \in E(G)$ by the maximality of M, since then $uv'u'v$ would be an M-augmenting path. But this means that $\deg(u; e) + \deg(v; e) \leq 2$ for each $e \in M$, which implies that

$$\delta(G) + \delta(G) \leq \deg(u) + \deg(v) = \sum_{e \in M} \deg(u; e) + \deg(v; e) \leq 2|M|$$

and hence $|M| \geq \delta(G)$.

For (ii), let M be a maximum matching in G. We are done unless there are vertices $u \in U$ and $v \in V$ not contained in M. There cannot be an edge $u'v'$ in M with $uu', vv' \in E$ by the maximality of M, since then $uu'v'v$ would be an M-augmenting path. Now M is maximal so all neighbours of u and v are contained in M. This means that $\deg(u; e) + \deg(v; e) \leq 1$ for each $e \in M$, which implies that

$$u + v \leq \deg(u) + \deg(v) = \sum_{e \in M} \deg(u; e) + \deg(v; e) \leq |M|$$

and hence $|M| \geq u + v$. □

It will be useful to have the following simple observations about the sizes of common neighbourhoods and maximal cliques.

Lemma 2.2. *Let $k \in \mathbb{N}$, $u_1, \ldots, u_k$ be vertices of a graph G and $U \subseteq V(G)$. Then we have* $\deg(u_1, \ldots, u_k; U) \geq \sum_{i=1}^{k} \deg(u_i; U) - (k-1)|U|$. *In particular, if $\delta(G) \geq \delta$ then* $\deg(u_1, \ldots, u_k) \geq k\delta - (k-1)n$.

Proof. Let $X := \{u_i \mid i \in [k]\}$. Count $\rho := \sum_{i\in[k],v\in U} \mathbf{1}_{\{vu_i \in E(G)\}}$ in two ways. On the one hand, $\rho = \sum_{i\in[k]}\sum_{v\in U} \mathbf{1}_{\{vu_i\in E(G)\}} = \sum_{i\in[k]} \deg(u_i;U)$. On the other hand, noting that vertices in $U \setminus \Gamma(X)$ have at most $k-1$ neighbours in X, we obtain

$$\begin{aligned}\rho &= \sum_{v\in U}\sum_{i\in[k]} \mathbf{1}_{\{vu_i\in E(G)\}} = \sum_{v\in U} \deg(v;X)\\ &= \sum_{v\in\Gamma(X;U)} \deg(v;X) + \sum_{v\in U\setminus\Gamma(X)} \deg(v;X)\\ &\le k\deg(X;U) + (k-1)(|U| - \deg(X;U)) = \deg(X;U) + (k-1)|U|.\end{aligned}$$

It follows that $\deg(X;U) \ge \sum_{i=1}^{k} \deg(u_i;U) - (k-1)|U|$. Furthermore, if $\delta(G) \ge \delta$ then $\deg(u_i) \ge \delta$ for each $i \in [k]$, so it follows immediately that $\deg(X) \ge k\delta - (k-1)n$. □

Lemma 2.3. *Let $j,k,\ell \in \mathbb{N}$ satisfy $j \le \ell \le k+1$ and G be a graph on n vertices with minimum degree $\delta(G) > \frac{(k-1)n}{k}$. Then every copy of K_j in G can be extended to a copy of K_ℓ in G.*

Proof. Fix $k, \ell \in \mathbb{N}$ satisfying $\ell \le k+1$ and proceed by backwards induction on j. The case $j = \ell$ is trivial. For $j < \ell$, note that by Lemma 2.2 a copy of K_j has common neighbourhood of size at least $j\delta - (j-1)n > 0$. Therefore, we can extend it to a copy of K_{j+1} by adding to it a vertex in its common neighbourhood. The resultant copy of K_{j+1} can be extended to a copy of K_ℓ by the induction hypothesis. □

The following is a classical result of Hajnal and Szemerédi [30].

Theorem 2.4 (Hajnal and Szemerédi [30]). *For any graph G on n vertices with maximum degree $\Delta(G)$ and any integer $r \ge \Delta(G)+1$, there is a partition of $V(G)$ into r independent sets which are each of size $\lceil \frac{n}{r} \rceil$ or $\lfloor \frac{n}{r} \rfloor$.*

For our purposes we will need the following corollary of Theorem 2.4.

Corollary 2.5. *Let $k \in \mathbb{N}$. Let G be a graph on $n \ge k(k+1)$ vertices with $\delta := \delta(G) \ge \frac{(k-1)n}{k}$. Then G contains $\min\left\{k\delta - (k-1)n, \left\lfloor \frac{n}{k+1} \right\rfloor\right\}$ vertex-disjoint copies of K_{k+1}.*

Proof of Corollary 2.5. Let G and δ satisfy the corollary hypothesis. First consider $\delta \in \left(\frac{(k-1)n}{k}, \frac{kn}{k+1}\right)$. Apply Theorem 2.4 to $\overline{G}$ with $r := \Delta(\overline{G}) + 1 = n - \delta \in \left(\frac{n}{k+1}, \frac{n}{k}\right)$. Each part in the resultant partition has size $\lceil \frac{n}{r} \rceil = k+1$ or $\lfloor \frac{n}{r} \rfloor = k$, so there are $n - rk = k\delta - (k-1)n$ pairwise disjoint independent sets of size $k+1$. These correspond to $k\delta - (k-1)n$ vertex-disjoint copies of K_{k+1} in G.

Now consider $\delta \geq \frac{kn}{k+1}$. Apply Theorem 2.4 to $\overline{G}$ with $r := \left\lfloor \frac{n}{k+1} \right\rfloor > \Delta(\overline{G})$. Each part in the resultant partition has size $\lceil \frac{n}{r} \rceil \geq k+1$ or $\lfloor \frac{n}{r} \rfloor \geq k+1$, so there are $r = \left\lfloor \frac{n}{k+1} \right\rfloor$ pairwise disjoint independent sets of size at least $k+1$ in $\overline{G}$. These correspond to $\left\lfloor \frac{n}{k+1} \right\rfloor$ vertex-disjoint copies of K_{k+1} in G. □

For our purposes the following corollary of Theorem 1.2 will be useful.

Corollary 2.6. *Given $k \in \mathbb{N}$, there exists $n_0 \in \mathbb{N}$ such that for all integers $n \geq n_0$, any graph G on n vertices with minimum degree $\delta(G) \geq \frac{kn-1}{k+1}$ contains the kth power of a Hamiltonian path.*

Proof. Given $k \in \mathbb{N}$, Theorem 1.2 produces $n_0 \in \mathbb{N}$. Let G be a graph on $n \geq n_0$ vertices with minimum degree $\delta(G) \geq \frac{kn-1}{k+1}$. Obtain a new graph G^* by adding to G a vertex adjacent to all other vertices. Note that $\delta(G^*) \geq \frac{k(n+1)}{k+1}$, so we can appeal to Theorem 1.2 to find a copy of C^k_{n+1} in G^*. Deleting the additional vertex from this copy of C^k_{n+1} in G^* yields the desired copy of P^k_n in G. □

The following theorem of Andrásfai, Erdős and Sós gives a sufficient condition for a K_k-free graph to be in fact $(k-1)$-partite.

Theorem 2.7 (Andrásfai, Erdős and Sós [11])**.** *Let $k \geq 3$ be an integer. A K_k-free graph G on n vertices with minimum degree $\delta(G) > \frac{3k-7}{3k-4}n$ is $(k-1)$-partite.*

A connected graph G on n vertices is *panconnected* if for each pair $u, v \in V(G)$ of vertices and each $\mathrm{dist}_G(u, v) < \ell \leq n$ there is a path in G with ℓ vertices which has u and v as endpoints. The following theorem of Williamson gives a sufficient minimum degree condition for a graph to be panconnected.

Theorem 2.8 (Williamson [58])**.** *Every graph G on $n \geq 4$ vertices with minimum degree $\delta(G) \geq \frac{n}{2} + 1$ is panconnected.*

The following theorem of Erdős and Stone gives a sufficient condition for a graph to contain $K_{t,t,t}$, the complete tripartite graph on three sets of vertices of size t.

Theorem 2.9 (Erdős and Stone [24])**.** *Given $t \in \mathbb{N}$ and $\rho > 0$, there exists $n_0 \in \mathbb{N}$ such that every graph on $n \geq n_0$ vertices with at least $\left(\frac{1}{2} + \rho\right)\binom{n}{2}$ edges contains a copy of $K_{t,t,t}$.*

2.1.1 Properties of Some Functions

In this subsection, we collect some analytical data about the functions r_p, r_c, pp_k and pc_k. Note that for fixed $k, n \in \mathbb{N}$ the functions $r_p(k,n,\cdot)$ and $r_c(k,n,\cdot)$ are monotone decreasing while the functions $\mathrm{pp}_k(n,\cdot)$ and $\mathrm{pc}_k(n,\cdot)$ are monotone increasing. Note that the definition of $r := r_p(k,n,\delta)$ in (1.1) is equivalent to $r = \left\lfloor \frac{(k-1)\delta-(k-2)n}{k\delta-(k-1)n+1} \right\rfloor$, from which we obtain

$$\frac{n-\delta-1}{k\delta-(k-1)n+1} < r \le \frac{(k-1)\delta-(k-2)n}{k\delta-(k-1)n+1} \quad \text{and} \tag{2.1}$$

$$\frac{[(k-1)r+1]n-(r+1)}{kr+1} < \delta \le \frac{[(k-1)(r-1)+1]n-r}{k(r-1)+1}. \tag{2.2}$$

The following lemma gives bounds on the functions r_p, pp_k and pc_k.

Lemma 2.10. *Given an integer $k \ge 3$ and $\mu > 0$, there exists $\eta_0 > 0$ such that for every $0 < \eta < \eta_0$ there exists $n_2 \in \mathbb{N}$ such that the following hold for all $n \ge n_2$. Let $r_0 \in \mathbb{N}$ satisfy $r_p(n,\gamma) \le r_0$ for all $\gamma \ge \left(\frac{k-1}{k}+\mu\right)n$. For $\delta \in \left[\left(\frac{k-1}{k}+\mu\right)n, \left(\frac{k}{k+1}-2\eta\right)n\right]$ we have*

$$r_p(k,n,\delta+\eta n) \ge 2, \tag{2.3}$$

$$\mathrm{pp}_k(n,\delta) \le \left(1-\tfrac{\eta}{10r_0}\right)\mathrm{pp}_k(n,\delta+\tfrac{\eta}{2}n), \tag{2.4}$$

$$\mathrm{pp}_k(n,\delta+\eta n) \le \frac{k+1}{2}\left(\frac{(k-1)(\delta+3\eta n)-(k-2)n}{r_p(k,n,\delta+\eta n)}-2\right), \tag{2.5}$$

$$\delta - \frac{(k-1)(\delta+3\eta n)-(k-2)n}{r_p(k,n,\delta+\eta n)} > \frac{3k-4}{3}(n-\delta) \quad \textit{and} \tag{2.6}$$

$$\begin{aligned}\mathrm{pp}_k(n,\delta+\eta n) &\le \frac{19}{20}(k+1)(k\delta-(k-1)n)-2\\ &\le (k+1)(k\delta-(k-1)n)-10k^2\eta n.\end{aligned} \tag{2.7}$$

For $\delta' \in \left[\left(\frac{k-1}{k}+\mu\right)n, \left(\frac{2k-1}{2k+1}-2\eta\right)n\right] \cup \left[\left(\frac{2k-1}{2k+1}+\eta\right)n, \left(\frac{k}{k+1}-2\eta\right)n\right] =: A$ we have

$$\mathrm{pp}_k(n,\delta'+\eta n) \le \frac{3}{4}(k+1)(k\delta'-(k-1)n). \tag{2.8}$$

For $\delta'' \in \left[\left(\frac{k-1}{k}+\mu\right)n, \left(\frac{3k-2}{3k+1}-2\eta\right)n\right] \cup \left[\left(\frac{3k-2}{3k+1}+\eta\right)n, \left(\frac{2k-1}{2k+1}-2\eta\right)n\right] =: B$ we have

$$\mathrm{pp}_k(n,\delta''+\eta n) \le \frac{2}{3}(k+1)(k\delta''-(k-1)n). \tag{2.9}$$

For $\delta''' \ge \left(\frac{k}{k+1}-2\eta\right)n$ we have

$$r_p(k,n,\delta'''+\eta n) \le 2. \tag{2.10}$$

For $\delta_1 \in \left[\left(\frac{k-1}{k} + \mu\right) n, \frac{kn-1}{k+1}\right)$ *we have*

$$\mathrm{pp}_k(n, \delta_1) \leq \min\left\{(k+1)(k\delta_1 - (k-1)n) - 10k^2\eta n - (k+1), \tfrac{11n}{20}\right\}. \tag{2.11}$$

For $\delta_2 \in \left[\left(\frac{k-1}{k} + \mu\right) n, \frac{kn-1}{k+1}\right]$ *we have*

$$\mathrm{pc}_k(n, \delta_2) \leq \tfrac{11n}{20}. \tag{2.12}$$

Proof. Let $k \geq 3$ be an integer and $\mu > 0$. Pick $\eta_0 = \frac{\mu}{200k^2}$. For $0 < \eta < \eta_0$, pick $n_2 = \max\left\{\frac{10r_0}{\eta}, 100k\right\}$. Let $n \geq n_2$ be an integer.

Let $\delta \in \left[\left(\frac{k-1}{k} + \mu\right) n, \left(\frac{k}{k+1} - 2\eta\right) n\right]$. Set $\delta_+ := \delta + \eta n$, $r := r_p(n, \delta)$ and $r' := r_p(n, \delta_+)$. If $r' = 1$, then by (2.2) we have $\delta_+ \geq \frac{kn-1}{k+1} > \left(\frac{k}{k+1} - \eta\right) n$. This gives a contradiction, so we have $r' \geq 2$, i.e. (2.3). By (1.2) we have

$$\begin{aligned}
\mathrm{pp}_k(n, \delta) &\leq \frac{k+1}{2}\left(\frac{(k-1)\delta - (k-2)n}{r}\right) + \frac{3k-1}{2} \\
&\leq \left(1 - \frac{\eta}{10r_0}\right)\left(\frac{k+1}{2}\left(\frac{(k-1)(\delta + \frac{\eta n}{2}) - (k-2)n}{r'}\right) + \frac{k-1}{2}\right) \\
&\leq \left(1 - \frac{\eta}{10r_0}\right)\mathrm{pp}_k\left(n, \delta + \frac{\eta n}{2}\right),
\end{aligned}$$

so we have (2.4). By (1.2) we have

$$\begin{aligned}
\mathrm{pp}_k(n, \delta_+) &\leq \frac{k+1}{2}\left(\frac{(k-1)\delta_+ - (k-2)n}{r'}\right) + \frac{3k-1}{2} \\
&\leq \frac{k+1}{2}\left(\frac{(k-1)(\delta + 3\eta n) - (k-2)n}{r'} - 2\right),
\end{aligned}$$

so we have (2.5). By (1.1), for some $q \in [r', r'+1]$ we have

$$(k-1)\delta_+ - (k-2)n = q(k\delta_+ - (k-1)n + 1),$$

so we have

$$\frac{(k-1)(\delta + 3\eta n) - (k-2)n}{r'} = (k\delta_+ - (k-1)n + 1)\frac{q}{r'} + \frac{(k-1)\eta n}{r'}.$$

Since $n - \delta - 1 \geq (q-1)(k\delta - (k-1)n + 1)$ and $r' \geq 2$, we have (2.6).

By (1.1) we have

$$\frac{(k-1)\delta_+ - (k-2)n}{r'+1} < k\delta_+ - (k-1)n + 1. \tag{2.13}$$

Hence, by (1.2) we have $\mathrm{pp}_k(n, \delta_+) \leq \frac{k+1}{2}(k\delta_+ - (k-1)n + 1)\frac{r'+1}{r'} + \frac{3k-1}{2}$. Since $r' \geq 2$, we have (2.7) because

$$\begin{aligned}\mathrm{pp}_k(n, \delta_+) &\leq \frac{3(k+1)}{4}(k\delta_+ - (k-1)n + 1) + \frac{3k-1}{2} \\ &\leq \frac{19}{20}(k+1)(k\delta - (k-1)n) - 2 \\ &\leq (k+1)(k\delta - (k-1)n) - 10k^2\eta n.\end{aligned}$$

Let $\delta' \in A$. By (2.2) we have $r' \geq 2$. If $r' \geq 3$, then we have

$$\begin{aligned}\mathrm{pp}_k(n, \delta'_+) &\leq \frac{2(k+1)}{3}(k\delta'_+ - (k-1)n + 1) + \frac{3k-1}{2} \\ &\leq \frac{3}{4}(k+1)(k\delta' - (k-1)n),\end{aligned}$$

which gives (2.8). If $r' = 2$, we have $\delta' \in \left[\left(\frac{2k-1}{2k+1} + \eta\right)n, \left(\frac{k}{k+1} - 2\eta\right)n\right] =: A'$. By considering (2.13) for $\delta_+ = \frac{(2k-1)n}{2k+1}$ and the coefficient of δ_+ on both sides of the inequality, for $\delta' \in A'$ we can strengthen the inequality to

$$\frac{(k-1)\delta'_+ - (k-2)n}{3} < k\delta' - (k-1)n - 4.$$

From this, we obtain $\mathrm{pp}_k(n, \delta' + \eta n) \leq \frac{3}{4}(k+1)(k\delta' - (k-1)n)$ by an argument analogous to that for (2.7). For $\delta'' \in B$ we obtain (2.9) by an argument analogous to that for $\delta' \in A$. Let $\delta''' \geq \left(\frac{k}{k+1} - 2\eta\right)n$. If $r' \geq 3$, then by (2.2) we have $\delta'''_+ \leq \frac{(2k-1)n-3}{2k+1} < \left(\frac{k}{k+1} - \eta\right)n$. This gives a contradiction, so we have $r' \leq 2$.

Let $\delta_1 \in \left[\left(\frac{k-1}{k} + \mu\right)n, \frac{kn-1}{k+1}\right)$. Since $\mathrm{pp}_k(n, \cdot)$ is monotone increasing, we have $\mathrm{pp}_k(n, \delta_1) \leq \mathrm{pp}_k(n, \frac{kn-2}{k+1}) \leq \frac{n}{2} + k \leq \frac{11n}{20}$. By (2.2) we have $r \geq 2$ and by (1.1) we have $\frac{(k-1)\delta_1 - (k-2)n}{r+1} < k\delta_1 - (k-1)n + 1$. Then, by (1.2) we have

$$\begin{aligned}\mathrm{pp}_k(n, \delta_1) &\leq \frac{3(k+1)}{4}(k\delta_1 - (k-1)n + 1) + \frac{3k-1}{2} \\ &\leq (k+1)(k\delta_1 - (k-1)n) - 10k^2\eta n - (k+1),\end{aligned}$$

so we obtain (2.11). Let $\delta_2 \in \left[\left(\frac{k-1}{k} + \mu\right)n, \frac{kn-1}{k+1}\right]$. Since $\mathrm{pc}_k(n, \cdot)$ is monotone increasing, we have $\mathrm{pc}_k(n, \delta_2) \leq \mathrm{pc}_k(n, \frac{kn-1}{k+1}) \leq \frac{n}{2} + k \leq \frac{11n}{20}$, which gives (2.12). □

2.2 Main Lemmas and Proof of Theorem 1.5

Our proof of Theorem 1.5 uses the well-established technique combining the regularity method, which involves the joint application of Szemerédi's regularity lemma [56] and a blow-up lemma, and the stability method. However, this provides only a loose framework for proofs of this kind. In particular, the proof of our Stability Lemma is significantly more involved than in [5] and is our main contribution. For our application we will define the concept of a connected K_{k+1}-component of a graph, which generalises the concept of a connected triangle component of a graph introduced by Allen, Böttcher and Hladký [5].

In this section we explain how we utilise connected K_{k+1}-components, the regularity method and the stability method. We first introduce the necessary definitions and formulate our main lemmas. Then, we detail how these lemmas imply Theorem 1.5 at the end of this section.

2.2.1 Connected K_{k+1}-components and K_{k+1}-factors

Fix $k \in \mathbb{N}$ and let G be a graph. A K_{k+1}*-walk* is a sequence $t_1, \dots, t_p$ of copies of K_k in G such that for every $i \in [p-1]$ there is a copy c_i of K_{k+1} in G such that $t_i, t_{i+1} \subseteq c_i$. We say that t_1 and t_p are K_{k+1}*-connected* in G. A K_{k+1}*-component* of G is a maximal set C of copies of K_k in G such that every pair of copies of K_k in C is K_{k+1}-connected. Observe that this induces an equivalence relation on the copies of K_k of G whose equivalence classes are the K_{k+1}-components of G. The *vertices of a* K_{k+1}*-component* C are all vertices v of G such that v is a vertex of a copy of K_k in C. The *size of a* K_{k+1}*-component* C is the number of vertices of C, which we denote by $|C|$.

A K_{k+1}*-factor* F in G is a collection of vertex-disjoint copies of K_{k+1} in G. F is a *connected* K_{k+1}*-factor* if all copies of K_k in F are contained in the same K_{k+1}-component of G. The *size* of a K_{k+1}-factor F is the number of vertices covered by F. Let $\mathrm{CKF}_{k+1}(G)$ denote the maximum size of a connected K_{k+1}-factor in G. For $\ell \in [k]$ the *copies of* K_ℓ *of a* K_{k+1}*-component* C are all copies of K_ℓ of G which can be extended to a copy of K_k in C. For $\ell > k$ the *copies of* K_ℓ *of a* K_{k+1}*-component* C are all copies of K_ℓ of G to which a copy of K_k in C can be extended. In a (slight) abuse of notation, we shall write $K_\ell \subseteq C$ to mean that there is such a copy of K_ℓ.

2.2.2 Regularity Method

In our proof we use a combination of Szemerédi's regularity lemma [56] and a blow-up lemma. Generally, the regularity lemma produces a partition of a dense graph that is suitable for an application of the blow-up lemma, which enables us to embed a target graph in a large host graph. We first introduce some terminology to formulate the versions of these two lemmas we will use.

Let G be a graph and $d, \varepsilon \in (0,1]$. Let $U, W \subseteq V(G)$ be a pair of disjoint nonempty subsets. The *density* of the pair (U,W) is $d(U,W) := \frac{e(U,W)}{|U||W|}$. The pair (U,W) is *ε-regular* if for all $U' \subseteq U$ and $W' \subseteq W$ with $|U'| \geq \varepsilon|U|$ and $|W'| \geq \varepsilon|W|$ we have $|d(U',W') - d(U,W)| \leq \varepsilon$. The pair (U,W) is *(ε,d)-regular* if it is ε-regular and has density at least d. An *ε-regular partition* of G is a partition $\{V_i\}_{i\in[\ell]_0}$ of $V(G)$ with $|V_0| \leq \varepsilon|V(G)|$, $|V_i| = |V_j|$ for all $i,j \in [\ell]$ and such that for all but at most $\varepsilon\ell^2$ pairs $(i,j) \in [\ell]^2$ the pair (V_i, V_j) is ε-regular. An *(ε,d)-regular partition* of G is an ε-regular partition $\{V_i\}_{i\in[\ell]_0}$ of G such that each vertex $v \in V(G) \setminus V_0$ is incident to at most $(d+\varepsilon)n$ edges which are not contained in the (ε,d)-regular pairs of the partition. The *reduced graph R* of an (ε,d)-regular partition $\{V_i\}_{i\in[\ell]_0}$ of G is the graph on $V(R) = \{V_1, \ldots, V_\ell\}$ with edges $V_iV_j \in E(R)$ whenever (V_i,V_j) is an (ε,d)-regular pair. We say that G has *(ε,d)-reduced graph R* and call the partition classes V_i with $i \in [\ell]$ *clusters* of G.

The classical regularity lemma by Szemerédi [56] states that every large graph has an ε-regular partition with a bounded number of parts. Here we state the so-called minimum degree form of Szemerédi's regularity lemma (see, e.g., Lemma 7 in conjunction with Proposition 9 in [43]).

Lemma 2.11 (Regularity Lemma, minimum degree form)**.** *Given $\varepsilon \in (0,1)$ and $m_0 \in \mathbb{N}$, there exists $m_1 \in \mathbb{N}$ such that the following holds for all $d, \gamma \in (0,1)$ such that $\gamma \geq 2d + 4\varepsilon$. Every graph G on $n \geq m_1$ vertices with $\delta(G) \geq \gamma n$ has an (ε,d)-reduced graph R on m vertices with $m_0 \leq m \leq m_1$ and $\delta(R) \geq (\gamma - d - 2\varepsilon)m$.*

This lemma asserts the existence of a reduced graph R of G which 'inherits' the high minimum degree of G. We shall use this property to reduce our original problem of finding the kth power of a path (or cycle) in a graph on n vertices with minimum degree γn to the problem of finding an *arbitrary* connected K_{k+1}-factor of a desired size in a graph R on m vertices with minimum degree $(\gamma - d - 2\varepsilon)m$ (see Lemma 2.12). The new problem seeks a much less specific subgraph (connected K_{k+1}-factor) than the original problem and is therefore easier to tackle.

This kind of problem reduction is possible due to the blow-up lemma, which enables the embedding of a large bounded degree target graph H into a graph G with reduced graph R if there is a homomorphism from H to a subgraph T of R which does not 'overuse' any cluster of T. For our purposes we apply this lemma with $T = K_{k+1}$ and obtain for each copy of K_{k+1} in a connected K_{k+1}-factor F the kth power of a path which almost fills up the corresponding clusters of G. The K_{k+1}-connectedness of F then enables us to link up these kth powers of paths and obtain kth powers of paths and cycles of the desired lengths (see Lemma 2.12 (i) and (ii)). In addition, the blow-up lemma allows for some control over the start-vertices and end-vertices of kth powers of paths constructed in this manner (see Lemma 2.12 (iii)).

The following lemma sums up what we obtain from this embedding strategy. This is an application of standard methods and we provide its proof in Section 2.7.

Lemma 2.12 (Embedding Lemma)**.** *For any integer $k \geq 2$ and any $d > 0$ there exists $\varepsilon_{EL} > 0$ with the following property. Given $0 < \varepsilon < \varepsilon_{EL}$, for every $m_{EL} \in \mathbb{N}$ there exists $n_{EL} \in \mathbb{N}$ such that the following hold for any graph G on $n \geq n_{EL}$ vertices with (ε, d)-reduced graph R of G on $m \leq m_{EL}$ vertices.*

(i) $C^k_{(k+1)\ell} \subseteq G$ for every $\ell \in \mathbb{N}$ with $(k+1)\ell \leq (1-d)\,\mathrm{CKF}_{k+1}(R)\frac{n}{m}$.

(ii) If $K_{k+2} \subseteq C$ for each K_{k+1}-component C of R, then $C^k_\ell \subseteq G$ for every $\ell \in \mathbb{N}$ with $k+1 \leq \ell \leq (1-d)\,\mathrm{CKF}_{k+1}(R)\frac{n}{m}$ and $\chi(C^k_\ell) \leq k+2$.

Furthermore, let $\mathcal{T}'$ be a connected K_{k+1}-factor in a K_{k+1}-component C of R which contains a copy of K_{k+2}, let $u_{1,1} \dots u_{1,k}$ and $u_{2,1} \dots u_{2,k}$ be vertex-disjoint copies of K_k in G, and suppose that $X_{1,1} \dots X_{1,k}$ and $X_{2,1} \dots X_{2,k}$ are (not necessarily disjoint) copies of K_k in C in R such that $u_{i,j} \dots u_{i,k}$ has at least $\frac{2dn}{m}$ common neighbours in the cluster $X_{i,j}$ for each $(i,j) \in [2] \times [k]$. Let A be a set of at most $\frac{\varepsilon n}{m}$ vertices of G disjoint from $\{u_{1,1}, \dots, u_{1,k}, u_{2,1}, \dots, u_{2,k}\}$. Then

(iii) $P^k_\ell \subseteq G$ for every $\ell \in \mathbb{N}$ with $3m^{k+1} \leq \ell \leq (1-d)|\mathcal{T}'|\frac{n}{m}$, such that P^k_ℓ starts in $u_{1,1} \dots u_{1,k}$ and ends in $u_{2,k} \dots u_{2,1}$ (in those orders), P^k_ℓ contains no element of A, and at most $\frac{\varepsilon n}{m}$ vertices of P^k_ℓ are not in $\bigcup \mathcal{T}'$.

Note that the copies of K_{k+2} required in (ii) are essential for the embedding of kth powers of cycles which not $(k+1)$-chromatic.

2.2.3 Stability Method

The regularity method as just described leaves us with the task of finding a sufficiently large connected K_{k+1}-factor F in a reduced graph R of G. However, this is insufficient on its own. The Embedding Lemma (Lemma 2.12) gives the kth power of a path which covers a proportion of G roughly the same as the proportion λ of R covered by F. Furthermore, the extremal graphs for kth powers of paths and connected K_{k+1}-factors are the same, but the relative minimum degree $\gamma_R = \delta(R)/|V(R)|$ of R is in general slightly smaller than $\gamma_G = \delta(G)/|V(G)|$. Consequently, we cannot expect that λ is larger than the proportion $\mathrm{pp}_k\left(v(R), \gamma_R v(R)\right)/v(R)$ a maximum kth power of a path covers in a graph with relative minimum degree γ_R; in particular, λ is smaller than the proportion $\mathrm{pp}_k\left(v(G), \gamma_G v(G)\right)/v(G)$ we would like to cover for relative minimum degree γ_G. This is where our stability approach comes into the picture.

Roughly speaking, we will be more ambitious and aim for a connected K_{k+1}-factor in R larger than guaranteed by the relative minimum degree (see Lemma 2.13 (C1) and (C2)). We prove that we fail to find this larger connected K_{k+1}-factor only if R (and hence G) is 'very close' to the extremal graph $G_p(k, v(R), \delta(R))$, in which case we will say that R is *near-extremal* (see Lemma 2.13 (C3)). The following lemma, which we call Stability Lemma and prove in Section 2.3, does precisely this. Note that this lemma guarantees the copies of K_{k+2} required in Embedding Lemma (Lemma 2.12). We remark that the proof of Stability Lemma is our main contribution as it is significantly more involved than in [5] and the analysis requires new insights.

Lemma 2.13 (Stability Lemma)**.** *Given an integer $k \geq 3$ and $\mu > 0$, for any sufficiently small $\eta > 0$ there exists $m_0 \in \mathbb{N}$ such that if $\delta \in \left[(\frac{k-1}{k} + \mu)n, \frac{kn}{k+1}\right]$ and G is a graph on $n \geq m_0$ vertices with minimum degree $\delta(G) \geq \delta$, then either*

(C1) $\mathrm{CKF}_{k+1}(G) \geq (k+1)(k\delta - (k-1)n)$*, or*

(C2) $\mathrm{CKF}_{k+1}(G) \geq \mathrm{pp}_k(n, \delta + \eta n)$*, or*

(C3) G has $k-1$ vertex-disjoint independent sets of combined size at least $(k-1)(n-\delta) - 5k\eta n$ whose removal disconnects G into components which are each of size at most $\frac{19}{10}(k\delta - (k-1)n)$ and for each such component X all copies of K_k in G containing at least one vertex of X are K_{k+1}-connected in G.

Moreover, in (C2) and (C3) each K_{k+1}-component of G contains a copy of K_{k+2}.

Note that (C1) gives a connected K_{k+1}-factor whose size is significantly larger than $\mathrm{pp}_k(n, \delta)$, which is the maximum size we can guarantee in general (see Figure 1.1 for an illustration in the case $k = 3$). We also remark that since $\mathrm{pp}_k(n, \delta)$ is a function with relatively large jumps at certain points, around these points (C2) gives a connected K_{k+1}-factor whose size is significantly larger than $\mathrm{pp}_k(n,\delta)$. Furthermore, we clarify that the 'components' in (C3) refer to the usual connected components of a graph rather than K_{k+1}-components. It is notable that the two functions $\mathrm{pp}_k(n,\delta)$ and $\mathrm{pc}_k(n,\delta)$ are sufficiently similar that Stability Lemma handles both. We need to distinguish between kth powers of paths and kth powers of cycles only when we consider near-extremal graphs.

It remains to handle graphs with near-extremal reduced graphs. We have a great deal of structural information about these graphs, which we use to directly find the desired kth powers of paths and cycles. The following lemma, which we call Extremal Lemma, handles the near-extremal case. We provide a proof of this lemma in Section 2.6. Note that in this proof we make use of our Embedding Lemma (Lemma 2.12); accordingly, Lemma 2.14 inherits the upper bound m_{EL} on the number of clusters from Lemma 2.12.

Lemma 2.14 (Extremal Lemma)**.** *Given an integer $k \geq 3$, $0 < \nu < 1$ and $0 < \eta, d \leq \frac{\nu^4}{(k+1)^{13}10^8}$, there exists $\varepsilon_0 > 0$ such that for every $0 < \varepsilon < \varepsilon_0$ and every $m_{EL} \in \mathbb{N}$, there exists $N \in \mathbb{N}$ such that the following holds. Let G be a graph on $n \geq N$ vertices with $\delta(G) \geq \delta \in \left[\left(\frac{k-1}{k} + \nu\right)n, \frac{kn}{k+1}\right)$ and R be an (ε, d)-reduced graph of G on $m \leq m_{EL}$ vertices with a partition $\left(\bigsqcup_{i=1}^{k-1} I_i\right) \sqcup \left(\bigsqcup_{j=1}^{\ell} B_j\right)$ of $V(R)$ with $\ell \geq 2$. Suppose that*

- *(i) each K_{k+1}-component of R contains a copy of K_{k+2},*
- *(ii) $I_1, \ldots, I_{k-1}$ are independent sets in R with $\left|\bigcup_{i=1}^{k-1} I_i\right| \geq ((k-1)(n-\delta) - 5k\eta n)\frac{m}{n}$,*
- *(iii) for each $i \in [\ell]$ we have $0 < |B_i| \leq \frac{19m}{10n}(k\delta - (k-1)n)$, all copies of K_k in R containing at least one vertex of B_i are K_{k+1}-connected in R, and for $j \in [\ell] \setminus \{i\}$ there are no edges between B_i and B_j in R.*

Then G contains $P^k_{\mathrm{pp}_k(n,\delta)}$ and C^k_ℓ for each $\ell \in [k+1, \mathrm{pc}_k(n,\delta)]$ such that $\chi(C^k_\ell) \leq k+2$.

We now have all the ingredients for the proof of our main theorem. The regularity lemma (Lemma 2.11) gives a regular partition of the graph G with reduced graph R and Stability Lemma (Lemma 2.13) tells us that R either contains a large connected K_{k+1}-factor or is near-extremal. We find kth powers of paths and cycles in G by applying Embedding Lemma (Lemma 2.12) in the first case and Extremal Lemma (Lemma 2.14) in the second case.

Proof of Theorem 1.5. We first set up the necessary constants. Fix an integer $k \geq 3$ and $0 < \nu < 1$. With k as input, Corollary 2.6 produces n_{HP}. Set $\mu := \left(1 - \frac{2}{2k^2+2k+1}\right)\nu$ and choose $\eta > 0$ to be small enough for Lemmas 2.10, 2.13 and 2.14. In particular, $\eta \leq \frac{\nu^4}{(k+1)^{13}10^8}$. Given k, μ, η from above as input, Lemmas 2.10 and 2.13 produce positive integers m_0' and m_2 respectively. Let $r_0 \in \mathbb{N}$ satisfy $r_p(n, \gamma) \leq r_0$ for all $\gamma \geq \left(\frac{k-1}{k} + \mu\right) n$. Set $d := \frac{\eta}{5r_0}$ and $m_0 := \max\{m_0', m_2, d^{-1}\}$. With k and d as input, Lemma 2.12 then produces $\varepsilon_{EL} > 0$. For ν, η and d, Lemma 2.14 produces $\varepsilon_0 > 0$. Set $\varepsilon := \frac{1}{2}\min\{\varepsilon_{EL}, \varepsilon_0, \frac{\eta}{5r_0}\}$ and choose $m_{EL} \in \mathbb{N}$ such that Lemma 2.11 guarantees the existence of an (ε, d)-regular partition with at least m_0 and at most m_{EL} parts. With the constants $\nu, \eta, d, \varepsilon, m_{EL}$ from above as input, Lemma 2.12 and Lemma 2.14 produce n_{EL} and N respectively. Finally, set $n_0 := \max\{n_{HP}, m_{EL}, n_{EL}, N\}$.

Let $\delta \in \left[\left(\frac{k-1}{k} + \nu\right) n, \frac{kn}{k+1}\right)$ and let G be a graph on $n \geq n_0$ vertices with minimum degree $\delta(G) \geq \delta$. As remarked after Theorem 1.5, observe that it suffices to show that $P^k_{\mathrm{pp}_k(n,\delta)} \subseteq G$ and that (ii) of Theorem 1.5 holds. Furthermore, we need to treat the case $\delta = \frac{kn-1}{k+1}$ separately from the rest because in this case $P^k_{\mathrm{pp}_k(n,\delta)}$ is actually Hamiltonian.

Let us consider the case when $\delta \in \left[\left(\frac{k-1}{k} + \nu\right) n, \frac{kn-1}{k+1}\right)$. We first apply Lemma 2.11 to G to obtain an (ε, d)-reduced graph R on $m_0 \leq m \leq m_{EL}$ vertices with $\delta(R) \geq \delta' := \left(\frac{\delta}{n} - d - 2\varepsilon\right) m$. Note that $\delta' \in \left[\left(\frac{k-1}{k} + \mu\right) m, \frac{km}{k+1}\right]$. Then we apply Lemma 2.13 to R. According to this lemma, there are three possibilities.

Firstly, we could have that $\mathrm{CKF}_{k+1}(R) \geq (k+1)(k\delta' - (k-1)m)$. Now Lemma 2.12(i) guarantees that G contains C^k_ℓ for each positive integer $\ell \leq (1-d)\,\mathrm{CKF}_{k+1}(R)\frac{n}{m}$ divisible by $k+1$; by the choice of d and ε we have $(1-d)(k+1)(k\delta' - (k-1)m)\frac{n}{m} \geq (k+1)(k\delta - (k-1)n) - 10k^2\eta n$, so G contains C^k_ℓ for each positive integer $\ell \leq (k+1)(k\delta - (k-1)n) - \nu n$ divisible by $k+1$, i.e. the second case of Theorem 1.5(ii) holds. Since $P^k_\ell \subseteq C^k_\ell$ and we have (2.11), it follows that G contains $P^k_{\mathrm{pp}_k(n,\delta)}$.

Secondly, we could have that $\mathrm{CKF}_{k+1}(R) \geq \mathrm{pp}_k(m, \delta' + \eta m)$ and every K_{k+1}-component of R contains a copy of K_{k+2}. By Lemma 2.12(ii) G contains C^k_ℓ for each integer $k+1 \leq \ell \leq (1-d)\,\mathrm{CKF}_{k+1}(R)\frac{n}{m}$ such that $\chi(C^k_\ell) \leq k+2$; by (2.4) and the choice of η, d and ε we have $(1-d)\,\mathrm{CKF}_{k+1}(R)\frac{n}{m} \geq (1-d)\,\mathrm{pp}_k(n, \delta + \frac{\eta n}{2}) \geq \mathrm{pp}_k(n,\delta) \geq \mathrm{pc}_k(n,\delta)$, so G contains C^k_ℓ for each integer $k+1 \leq \ell \leq \mathrm{pc}_k(n,\delta)$ such that $\chi(C^k_\ell) \leq k+2$, i.e. the first case of Theorem 1.5(ii) holds. Since $P^k_\ell \subseteq C^k_\ell$, it also follows that G contains $P^k_{\mathrm{pp}_k(n,\delta)}$.

Thirdly, we could have that R is near-extremal. In this case every K_{k+1}-component of R contains a copy of K_{k+2} and R contains $k-1$ vertex-disjoint independent sets of combined size at least $(k-1)(m-\delta')-5k\eta m \geq ((k-1)(n-\delta)-5k\eta n)\frac{m}{n}$ whose removal disconnects R into components which are each of size at most $\frac{19}{10}(k\delta' - (k-1)m) \leq \frac{19m}{10n}(k\delta - (k-1)n)$ and for each such component X all copies of K_k in R containing at least one vertex of X are K_{k+1}-connected in R. But now G and R satisfy the conditions of Lemma 2.14, so it follows that G contains $P^k_{\mathrm{pp}_k(n,\delta)}$ and C^k_ℓ for each integer $k+1 \leq \ell \leq \mathrm{pc}_k(n,\delta)$ such that $\chi(C^k_\ell) \leq k+2$, i.e. the first case of Theorem 1.5(ii) holds.

It remains to deal with the special case $\delta = \frac{kn-1}{k+1}$. As with the main case, we apply Lemma 2.11 to G to obtain a reduced graph R, apply Lemma 2.13 to R with three possible outcomes and then apply Lemma 2.12(i), Lemma 2.12(ii) and Lemma 2.14 in the first, second and third cases respectively to obtain kth powers of cycles of the appropriate lengths. Finally, by Corollary 2.6 G contains a copy of $P^k_n = P^k_{\mathrm{pp}_k(n,\delta)}$. $\square$

2.3 Proving the Stability Lemma

In this section we provide a proof of our Stability Lemma for connected K_{k+1}-factors, Lemma 2.13. We divide the proof of Lemma 2.13 into three lemmas, which correspond to the following three cases.

(1) G has just one K_{k+1}-component (see Lemma 2.15),

(2) G has a K_{k+1}-component C that does not contain a copy of K_{k+2} (see Lemma 2.16),

(3) G has at least two K_{k+1}-components and each K_{k+1}-component contains a copy of K_{k+2} (see Lemma 2.17).

In the first case, the result follows from an application of a classical result of Hajnal and Szemerédi [30] in the form of Lemma 2.15. In the second case, the result follows from an inductive argument in the form of Lemma 2.16. Finally, we handle the third case in the form of Lemma 2.17. This turns out to be the main work and we will provide a sketch of its proof at the end of this section.

We now state Lemmas 2.15, 2.16 and 2.17, and provide a proof of Lemma 2.13 applying these lemmas. We provide the proofs of Lemmas 2.15 and 2.16 right after our proof of Lemma 2.13. Then, we introduce a family of configurations in Section 2.3.1 to prepare for the substantially more involved proof of Lemma 2.17. We analyse this family of configurations and develop greedy-type methods for the construction of connected K_{k+1}-factors in Section 2.4. These are applied in the proof of Lemma 2.17, which is provided in Section 2.5.

Lemma 2.15. *Let $k \in \mathbb{N}$ and $\delta \in \left[\frac{(k-1)n}{k}, \frac{kn}{k+1}\right]$. Let G be a graph on $n \geq k(k+1)$ vertices with minimum degree $\delta(G) \geq \delta$ and exactly one K_{k+1}-component. Then* $\mathrm{CKF}_{k+1}(G) \geq (k+1)(k\delta - (k-1)n)$.

Lemma 2.16. *Let $k \in \mathbb{N}$. Let G be a graph on n vertices with minimum degree $\delta(G) \geq \delta \geq \frac{(k-1)n}{k}$. Suppose that G has a K_{k+1}-component C which does not contain a copy of K_{k+2}. Then there is a set of $k\delta - (k-1)n$ vertex-disjoint copies of K_{k+1} which are all in C.*

Lemma 2.17. *Given an integer $k \geq 3$ and $\mu > 0$, for any sufficiently small $\eta > 0$ there exists $m_1 \in \mathbb{N}$ such that if $\delta \geq (\frac{k-1}{k} + \mu)n$ and G is a graph on $n \geq m_1$ vertices with minimum degree $\delta(G) \geq \delta$ such that G has at least two K_{k+1}-components and every K_{k+1}-component of G contains a copy of K_{k+2}, then either*

(D1) $\mathrm{CKF}_{k+1}(G) \geq \mathrm{pp}_k(n, \delta + \eta n)$, *or*

(D2) G has $k-1$ vertex-disjoint independent sets of combined size at least $(k-1)(n - \delta) - 5k\eta n$ whose removal disconnects G into components which are each of size at most $\frac{19}{10}(k\delta - (k-1)n)$ and for each component X all copies of K_k in G containing at least one vertex of X are K_{k+1}-connected in G.

Proof of Lemma 2.13. Given an integer $k \geq 3$, $\mu > 0$ and any $\eta > 0$ sufficiently small for the application of Lemma 2.17, Lemma 2.17 produces $m_1 \in \mathbb{N}$. Set $m_0 := \max\{m_1, k(k+1)\}$. Let $\delta \in \left[(\frac{k-1}{k} + \mu)n, \frac{kn}{k+1}\right]$ and let G be a graph on $n \geq m_0$ vertices with minimum degree $\delta(G) \geq \delta$.

If G has only one K_{k+1}-component then Lemma 2.15 implies $\mathrm{CKF}_{k+1}(G) \geq (k+1)(k\delta - (k-1)n)$. If G has a K_{k+1}-component C which does not contain a copy of K_{k+2} then Lemma 2.16 implies $\mathrm{CKF}_{k+1}(G) \geq (k+1)(k\delta - (k-1)n)$. In both cases we are in (C1). If G has at least two K_{k+1}-components and every K_{k+1}-component of G contains a copy of K_{k+2}, then Lemma 2.17 implies that we are in (C2) or (C3). □

Next, we provide the proofs of Lemmas 2.15 and 2.16.

Proof of Lemma 2.15. Fix $k \in \mathbb{N}$ and $\delta \in \left[\frac{(k-1)n}{k}, \frac{kn}{k+1}\right]$. Let G be a graph on $n \geq k(k+1)$ vertices with minimum degree $\delta(G) \geq \delta$ and exactly one K_{k+1}-component. Corollary 2.5 implies $\mathrm{CKF}_{k+1}(G) \geq (k+1)(k\delta - (k-1)n)$. $\square$

Proof of Lemma 2.16. We proceed by induction on k. For $k = 1$, let x be a vertex of C and define $U := \Gamma(x) \subseteq C$. Note that a K_2-component is a connected component, so in particular vertices with a neighbour in C are also in C. C contains no triangle, so U is an independent set. Pick a set S of δ vertices from U. Choose greedily for each $u \in S$ a distinct vertex $v \in V(G)$ such that uv is an edge. Since $S \subseteq U$ is an independent set, all these vertices are not elements of S. Since $\deg(u) \geq \delta$, we can find a distinct vertex for each $u \in S$. This yields a set M of δ vertex-disjoint edges all in C.

Now suppose $k \geq 2$. Let x be a vertex of C. Define $H := G[\Gamma(x)]$ and $C_1 := \{x_1 \dots x_{k-1} : x_1 \dots x_{k-1}x \in C\}$. Note that H is a graph on $m := |\Gamma(x)| \geq \delta$ vertices with minimum degree $\delta(H) \geq \delta - n + m \geq \frac{k-2}{k-1}m$ and C_1 is a nonempty union of some K_k-components of H. Since C does not contain a copy of K_{k+2}, any K_k-component $C' \subseteq C_1$ of H does not contain a copy of K_{k+1}. Let C' be such a K_k-component. Applying the induction hypothesis with H and C', we obtain a set F'' of $(k-1)\delta(H) - (k-2)m \geq (k-1)(\delta - n + m) - (k-2)m \geq k\delta - (k-1)n$ vertex-disjoint copies of K_k which are all in $C' \subseteq C_1$; since these copies of K_k lie in $\Gamma(x)$, they are also in C. Let $F' \subseteq F''$ be a subset of F'' containing $k\delta - (k-1)n$ vertex-disjoint copies of K_k. Choose greedily for each $f \in F'$ a distinct vertex $v \in V(G)$ such that fv is a copy of K_{k+1}. Since C' is K_{k+1}-free, all these vertices are not neighbours of x and in particular are not vertices of elements of F'. Since $|\Gamma(f)| \geq k\delta - (k-1)n$ by Lemma 2.2, we can find a distinct vertex for each $f \in F'$. This yields a set F of $k\delta - (k-1)n$ vertex-disjoint copies of K_{k+1} which are all in C. $\square$

2.3.1 Configurations

To prepare for the proof of Lemma 2.17, we first introduce some definitions useful for the analysis of the graph structure. Let G be a graph with K_{k+1}-components $C_1, \dots, C_r$. The K_{k+1}-*interior* $\mathrm{int}_k(G)$ of G is the set of vertices of G which are in more than one of the K_{k+1}-components. For a K_{k+1}-component C_i, the *interior* $\mathrm{int}(C_i)$ of C_i is the set of vertices of C_i which are in $\mathrm{int}_k(G)$. The *exterior* $\mathrm{ext}(C_i)$ of C_i is the set of vertices of C_i which are in no other K_{k+1}-component of G. To give an example, by definition the graph $G_p(k, n, \delta)$ has $r_p(k, n, \delta)$ K_{k+1}-components; its K_{k+1}-interior is the disjoint union of the $k-1$ independent sets $I_1, \dots, I_{k-1}$ (using notation from the definition of $G_p(k,n,\delta)$ on page 13 in Section 1.2) and its component exteriors are the cliques $X_1, \dots, X_{r_p(k,n,\delta)}$. Note that $\mathrm{int}_k(G_p(k,n,\delta))$ induces a complete $(k-1)$-partite graph and in particular contains no copy of K_k.

A key part of our proof of Lemma 2.17 involves the analysis of the case in which $\mathrm{int}_k(G)$ contains a copy of K_k, which corresponds to the case in which $\mathrm{int}_2(G)$ contains an edge in [5]. However, unlike in the case in [5], in our case the graph structure is not immediately amenable to the construction of connected K_{k+1}-factors. To overcome this, in this subsection we introduce a family of configurations which give graph structures that facilitate the construction of connected K_{k+1}-factors.

Definition 2.18 (Configurations)**.** Let $j, k, \ell \in \mathbb{N}$ satisfy $j < \ell \le k$. We say that a graph G *contains* the configuration $\dagger_k(\ell, j)$ if there is a (multi)set

$$\{u_i : i \in [k]\} \sqcup \{v_i : j < i \le \ell\} \sqcup \{w_{i,h} : j < i \le \ell, h \in [\ell - 1]\}$$

of (not necessarily distinct) vertices in $V(G)$ such that

(CG1) $u_1 \dots u_k$ is a copy of K_k in a K_{k+1}-component C of G,

(CG2) $u_1 \dots u_j v_{j+1} \dots v_\ell u_{\ell+1} \dots u_k$ is a copy of K_k of G not in C, and

(CG3) $u_{\ell+1} \dots u_k u_p w_{p,1} \dots w_{p,\ell-1}$ is a copy of K_k of G not in C for every $j < p \le \ell$.

We say that G *does not contain* the configuration $\dagger_k(\ell, j)$ if there is no such (multi)set of vertices in $V(G)$.

One may regard the configuration $\dagger_k(\ell, j)$ as a collection of copies of K_k which satisfies the following.

(i) There are $k - \ell$ vertices common to all the copies of K_k.

(ii) There is a 'central' copy of K_k in some K_{k+1}-component C and all the other copies of K_k do not belong to C.

(iii) After deleting the $k - \ell$ common vertices from the copies of K_k, we obtain a collection of copies of K_ℓ. The 'central' copy of K_ℓ shares j vertices with one other copy of K_ℓ and each of its remaining vertices has one copy of K_ℓ 'dangling' off it.

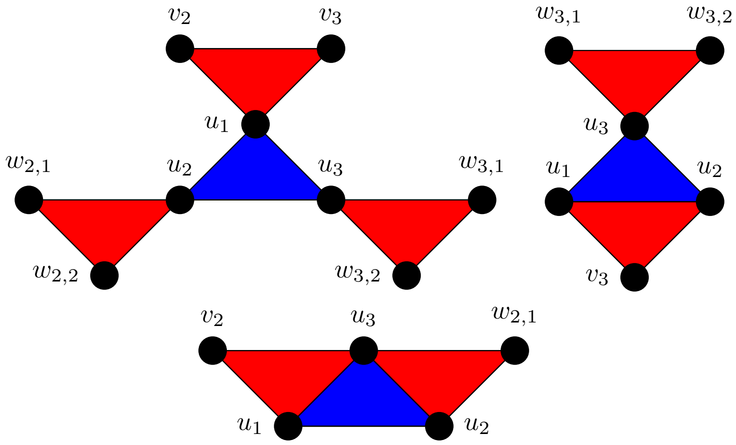


Figure 2.1: $\dagger_3(3,1)$, $\dagger_3(3,2)$ and $\dagger_3(2,1)$

Note that these configurations are by no means distinct, since 'non-central' copies of K_k and vertices not on the same copy of K_k need not be distinct. For example, a graph that contains $\dagger_3(3,2)$ also contains $\dagger_3(3,1)$ – set $v_2 := u_2, w_{2,1} := u_1, w_{2,2} := v_3$. The family of configurations for $k = 3$ can be found in Figure 2.1.

Now let us sketch the proof of Lemma 2.17. We distinguish two cases as follows.

(i) $\mathrm{int}_k(G)$ contains a copy of K_k,

(ii) $\mathrm{int}_k(G)$ does not contain a copy of K_k.

Case (i) is equivalent to G containing $\dagger_k(k,1)$, so G contains a member of our family of configurations. By Lemma 2.19, G in fact contains a configuration of the form $\dagger_k(\ell+1,\ell)$. Consider the configuration of this form contained in G with minimal ℓ. We distinguish two cases: when $\ell = 1$ and when $\ell > 1$. In the first case, Lemma 2.20 tells us that common neighbourhoods of a certain form are independent sets, which enables us to apply Lemma 2.22 to obtain the desired large connected K_{k+1}-factor. In the second case, we know that G does not contain $\dagger_k(2,1)$. Lemma 2.21(i) tells us that common neighbourhoods of a certain form are independent sets and we are able to apply Lemma 2.23 to obtain the desired large connected K_{k+1}-factor. We remark that the argument presented above for the second case is inadequate when δ is close to $\frac{(2k-1)n}{2k+1}$. We use an essentially similar but more tailored approach in the form of Lemmas 2.21(ii) and 2.24.

In Case (ii), G resembles our extremal graphs and has enough structure for the application of our construction methods to obtain the desired large connected K_{k+1}-factor. This approach works for most values of δ below $\frac{(2k-1)n}{2k+1}$. For $\delta \geq \frac{(2k-1)n}{2k+1}$ however, we find that our greedy-type methods are insufficient. To overcome this, we employ a Hall-type argument in the form of Lemma 2.32.

2.4 Structure and Methods

In this section we develop useful techniques for our proof of Lemma 2.17. These include structural results pertaining to the family of configurations defined in Section 2.3.1 and procedures for constructing connected K_{k+1}-factors.

2.4.1 Configurations and Structure

In this subsection we prove structural facts about our family of configurations for our proof of Lemma 2.17.

A key argument in our proof of Lemma 2.17 is that a graph without a sufficiently large connected K_{k+1}-factor in fact contains no member of the family of configurations defined previously in Section 2.3.1. The following lemma establishes an inductive-like relationship between the members of our family of configurations.

Lemma 2.19. *Let $j, k, \ell \in \mathbb{N}$ satisfy $j + 2 \leq \ell \leq k$ and G be a graph on n vertices with minimum degree $\delta = \delta(G) > \frac{(k-1)n}{k}$ and at least two K_{k+1}-components. Suppose that G does not contain $\dagger_k(\ell, j+1)$, $\dagger_k(\ell, \ell-1)$ or $\dagger_k(\ell - j, 1)$. Then G does not contain $\dagger_k(\ell, j)$.*

Proof. Suppose that G contains $\dagger_k(\ell, j)$. By Definition 2.18, there is a (multi)set

$$\{u_i : i \in [k]\} \sqcup \{v_i : j < i \leq \ell\} \sqcup \{w_{i,h} : j < i \leq \ell, h \in [\ell - 1]\}$$

of vertices in $V(G)$ such that (CG1)–(CG3) hold. The vertices $u_1, \ldots, u_k, v_{j+1}, \ldots, v_\ell$ are all distinct: if $u_a = v_b$ for some $j < a, b \leq \ell$, the copy of K_k containing v_b would share at least $k - \ell + j + 1$ vertices with the 'central' copy of K_k, thereby yielding $\dagger_k(\ell, j+1)$.

For each $j < i \leq \ell$ define $S_i := \Gamma(v_i, u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_k) \setminus \{u_i\}$ and $f_i := u_1 \ldots u_{i-1} u_{i+1} \ldots u_k$. Set $f' := u_1 \ldots u_j u_{\ell+1} \ldots u_k$. Let $j < i \leq \ell$. Observe that v_i has at least two non-neighbours in $\{u_{j+1}, \ldots, u_\ell\}$. Indeed, without loss of generality, suppose that v_i is adjacent to the vertices $u_{j+1}, \ldots, u_{\ell-1}$. Here $B := f' u_{j+1} \ldots u_{\ell-1} v_i$ is a copy of K_k sharing at least $k-1$ vertices with $u_1 \ldots u_k$ and at least $k - \ell + j + 1$ vertices with $f' v_{j+1} \ldots v_\ell$. If $B \in C$, then by taking B as the 'central' copy of K_k we have $\dagger_k(\ell, j+1)$. If $B \notin C$, then with B replacing $f' v_{j+1} \ldots v_\ell$ we have $\dagger_k(\ell, \ell-1)$.

Now by applying Lemma 2.2 with $U = V(G) \setminus \{u_1, \ldots, u_k, v_i\}$, we have

$$|S_i| \geq (k-1)(\delta - k) + (\delta - k + 2) - (k-1)(n-k-1) = k\delta - (k-1)n + 1 > 0.$$

Pick $s_i \in S_i$ and complete $f' s_i v_i$ to a copy $Z_i := f' s_i v_i z_{i,1} \ldots z_{i,\ell-j-2}$ of K_k by Lemma 2.3. Observe that $f_i s_i \in C$: if not, then $u_i w_{i,1} \ldots w_{i,k-1} \notin C$, $f_i u_i \in C$ and $f_i s_i \notin C$ would yield $\dagger_k(\ell, \ell-1)$ with $f_i u_i$ as the 'central' copy of K_k. Furthermore, we have $Z_i \in C$: if not, then $u_p w_{p,1} \ldots w_{p,k-1} \notin C$ for $j < p \leq \ell$, $f_i s_i \in C$ and $Z_i \notin C$ would yield $\dagger_k(\ell - j, 1)$ with $f_i s_i$ as the 'central' copy of K_k. But now $Z_i \in C$ for $j < i \leq \ell$ with $f' v_{j+1} \ldots v_\ell \notin C$ as the 'central' copy of K_k yields $\dagger_k(\ell - j, 1)$, which is a contradiction. □

The following lemma collects structural properties useful for the construction of connected K_{k+1}-factors in graphs which contain $\dagger_k(2,1)$.

Lemma 2.20. *Let $k \geq 2$ be an integer. Let $f = u_1 \ldots u_{k-1}$ be a copy of K_{k-1} in a graph G which lies in distinct K_{k+1}-components C_1 and C_2 of G. Let uv and wu_k be edges of G such that $fu, fv \in C_1$ and $fw, fu_k \in C_2$. Then $\Gamma(u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_k, w, u, v)$ is an independent set for each $i \in [k-1]$.*

Proof. Fix $i \in [k-1]$. Suppose that $X := \Gamma(u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_k, w, u, v)$ contains an edge $u'v'$. Let $U := u_1 \ldots u_{i-1} u_{i+1} \ldots u_{k-1}$. Note that $U u_i u \in C_1$ and $U u_i uv$ is a copy of K_{k+1} in G so $Uuv \in C_1$; also $Uuvu'v'$ is a copy of K_{k+2} in G so $Uu'v' \in C_1$. On the other hand, $U u_k w \in C_2$ and $U u_k w u' v'$ is a copy of K_{k+2} in G so $Uu'v' \in C_2$. Since no copy of K_k is in more than one K_{k+1}-component, this is a contradiction. Hence, X contains no edge and is therefore an independent set. □

The following lemma provides structural properties useful for the construction of connected K_{k+1}-factors in graphs containing $\dagger_k(\ell, \ell-1)$ for some $3 \leq \ell \leq k$ but not $\dagger_k(2,1)$.

Lemma 2.21. *Let $k \geq 2$ and $i \in [k-1]$ be integers. Let G be a graph which does not contain $\dagger_k(2,1)$ and $f = u_1 \ldots u_{k-1}$ be a copy of K_{k-1} in G which lies in distinct*

K_{k+1}-components C_1 and C_2 of G. Let uv be an edge of G such that $fu, fv \in C_1$ and w be a vertex of G such that $fw \in C_2$. Then

(i) $\Gamma(u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_{k-1}, w, u, v)$ *is an independent set.*

(ii) $\Gamma(x_1, \ldots, x_{i-1}, u_{i+1}, \ldots, u_{k-1}, w, u, v)$ *is an independent set for any copy $g := x_1 \ldots x_{i-1}$ of K_{i-1} in G where $x_j \in \Gamma(u_{j+1}, \ldots, u_{k-1}, w, u, v)$ for all $j < i$.*

Proof. We obtain (i) from (ii) by noting that $u_1 \ldots u_{i-1}$ is a copy of K_{i-1} in G such that for each $j < i$ we have $u_j \in \Gamma(u_{j+1}, \ldots, u_{k-1}, w, u, v)$. Hence, it remains to prove (ii).

Fix a copy $g := x_1 \ldots x_{i-1}$ of K_{i-1} such that for each $j < i$ we have $x_j \in \Gamma(u_{j+1}, \ldots, u_{k-1}, w, u, v)$. For each $j \in [i]$ set $U'_j := u_1 \ldots u_{j-1}$, $g_j := x_1 \ldots x_{j-1}$, $U_j := u_{j+1} \ldots u_{k-1}$ and $f_j := U'_j U_j$. Set $U_0 := f$. We shall prove by induction that $g_j U_j uv \in C_1$ for each $j \in [i]$. For $j = 1$, note that $fu \in C_1$ and fuv is a copy of K_{k+1} in G so $f_1 uv = g_1 U_1 uv \in C_1$. For $j \geq 2$, note that $g_{j-1} U_{j-1} uv \in C_1$ by the induction hypothesis and $g_j U_{j-1} uv$ is a copy of K_{k+1} in G so $g_j U_j uv \in C_1$, completing the proof by induction. In particular, we have $gU_i uv \in C_1$.

Set $X := \Gamma(x_1, \ldots, x_{i-1}, u_{i+1}, \ldots, u_{k-1}, w, u, v)$ and suppose that there is an edge $u'v'$ with $u', v' \in X$. Now by the definitions of g and X we have that $gU_i uvu'v'$ is a copy of K_{k+2} in G, so we have $gU_i u'v' \in C_1$. Furthermore, since $g_j U_j uv \in C_1$ and $g_j U_{j-1} uv$ is a copy of K_{k+1} in G for each $j \in [i]$, we have $g_j U_{j-1} u \in C_1$ for each $j \in [i]$.

Now we shall prove that $g_j U_{j-1} w \in C_2$ for each $j \in [i]$. For $j = 1$, we have $fw = g_1 U_0 w \in C_2$. For $j \geq 2$, observe that the induction hypothesis implies that $g_j U_{j-1} w \in C_2$: if not, then $g_{j-1} U_{j-2} u \in C_1$ from before, $g_{j-1} U_{j-2} w \in C_2$ by the induction hypothesis and $g_j U_{j-1} w \notin C_2$ would yield $\dagger_k(2,1)$ with $g_{j-1} U_{j-2} w \in C_2$ as the 'central' copy of K_k and $g_{j-1} U_{j-1}$ as the common vertices. This completes the proof by induction. In particular, we have $gU_{i-1} w \in C_2$. Now observe that $gU_i wu' \in C_2$: if not, then $gU_{i-1} u \in C_1$ from before, $gU_{i-1} w \in C_2$ and $gU_i wu' \notin C_2$ would yield $\dagger_k(2,1)$ with $gU_{i-1} w \in C_2$ as the 'central' copy of K_k and gU_i as the common vertices. Finally, $gU_i wu'v'$ is a copy of K_{k+1} in G so $gU_i u'v' \in C_2$, which contradicts our earlier deduction that $gU_i u'v' \in C_1$. Hence, X contains no edge and is therefore an independent set. $\square$

2.4.2 Constructing Connected K_{k+1}-factors

In this subsection we develop greedy-type procedures for constructing connected K_{k+1}-factors which exploit certain structures in graphs of interest, including those proved in

Section 2.4.1. Lemmas 2.22, 2.23, and 2.24 serve to formalise the achievable outcomes of these procedures.

Lemma 2.22 represents a greedy-type procedure for constructing connected K_{k+1}-factors in a graph using two parallel processes following two closely related partitions of the vertex set. The purpose of this lemma is to obtain sufficiently large connected K_{k+1}-factors in graphs containing $\dagger_k(2,1)$. The sets A and A' in Lemma 2.22 contain the vertices avoided by the two parallel processes. Note that the larger A and A' are, the smaller s_1 and t_1 are. Since the sizes of s_1 and t_1 are the key determinants of the attainable size of a connected K_{k+1}-factor, we think of A and A' as 'bad' sets. We remark that while we formally allow the quantities s_1, s_2, t_1 and t_2 to be negative to reduce the overall proof complexity, in practice they will always be non-negative.

Lemma 2.22. *Let $2 \le b \le c \le k$ be integers. Let G be a graph on n vertices with minimum degree $\delta = \delta(G) > \frac{(k-1)n}{k}$. Suppose there are two partitions of $V(G)$, one with vertex classes $U_1, U_2, X_1, \ldots, X_{k-1}, A$ and another with vertex classes V_1, V_2, $X_1, \ldots, X_{k-2}, X', A'$, such that*

(a) $U_1 \cap V_1 = \varnothing$,

(b) there are no edges between U_1 and U_2 and between V_1 and V_2,

(c) all copies of K_k in G with at least two vertices from U_1 and all other vertices from $\bigcup_{i=1}^{k-2} X_i$, or at least two vertices from V_1 and all other vertices from $\bigcup_{i=1}^{k-2} X_i$, are K_{k+1}-connected,

(d) $|X_i| \le n - \delta$ for $i \in [k-1]$ and $|X'| \le n - \delta$, and

(e) $X_i \cap \Gamma(g)$ is an independent set for each (i, g) where $i \in [k-2]$ and g is a clique of order at least i with at least two vertices from U_1 and all other vertices from $\bigcup_{j=1}^{i-1} X_j$, or at least two vertices from V_1 and all other vertices from $\bigcup_{j=1}^{i-1} X_j$.

Let F^U be a collection of vertex-disjoint copies of K_b in U_1 and F^V be a collection of vertex-disjoint copies of K_c in V_1. Let $d_1, d_2 \ge 0$ satisfy $|V_2| \ge 2d_2 + d_1$. Set

$$s_1 := \tfrac{k\delta-(k-1)n+(b-1)|U_2|-|A|}{2b-1}, \qquad t_1 := \tfrac{k\delta-(k-1)n+(c-1)|V_2|-|A'|-|U_1|(c-1)/b}{2c-1},$$
$$s_2 := \tfrac{k\delta-(k-1)n+(b-1)|U_2|-|V_1|}{2b-1} \qquad \text{and } t_2 := \tfrac{k\delta-(k-1)n+(c-1)|V_2|-|U_1|-|U_1|(c-1)/b}{2c-1}.$$

Then G contains a connected K_{k+1}-factor of size at least

$$(k+1)\min\left\{|F^U|, \left\lfloor\frac{|U_2|}{2}\right\rfloor, d_1, s_1, s_2\right\} + (k+1)\min\left\{|F^V|, d_2, t_1, t_2\right\}.$$

Moreover, if F^V is empty then G contains a connected K_{k+1}-factor of size at least

$$(k+1)\min\left\{|F^U|, \left\lfloor\frac{|U_2|}{2}\right\rfloor, d_1, s_1, s_2\right\}.$$

The proof of this lemma proceeds as follows. We first describe the greedy-type procedure used to construct a connected K_{k+1}-factor. Then, we prove that our procedure does indeed produce a connected K_{k+1}-factor of the desired size. This turns out to be an inductive argument where we need to justify that we can make 'good' choices at each step and the quantities s_1, s_2, t_1, t_2 are chosen to ensure success. For example, a copy of K_k extending a copy of K_b in F^U has at least $k\delta - (k-1)n + (b-1)|U_2| - |A|$ common neighbours not in some 'bad' set A; on the other hand, each copy of K_b in F^U may render up to $2b-1$ of these common neighbours unavailable, so $|F^U| \leq s_1$ ensures that there is still an available common neighbour.

Proof. Let $F_b^U \subseteq F^U$ and $F_c^V \subseteq F^V$ satisfy $|F_c^V| = \max\left\{0, \min\left\{|F^V|, d_2, t_1, t_2\right\}\right\}$ and $|F_b^U| = \max\left\{0, \min\left\{|F^U|, \left\lfloor\frac{|U_2|}{2}\right\rfloor, d_1, s_1, s_2\right\}\right\}$. In what follows, we will extend each clique in F_b^U to a copy of K_k using vertices in $U_1, X_1, \dots, X_{k-2}$ and each clique in F_c^V to a copy of K_k using vertices in $V_1, X_1, \dots, X_{k-2}$. These copies of K_k will then be extended to copies of K_{k+1} using vertices outside of $U_1, V_1, X_1, \dots, X_{k-2}$. Note that the resultant copies of K_{k+1} will be K_{k+1}-connected by (c).

We build up our desired connected K_{k+1}-factor by running two parallel processes, one starting from F_b^U in U_1 and the other starting from F_c^V in V_1. Each process is a two-stage step-by-step process performing steps in tandem with the other process. Set $\overline{F}_{b-1}^U, \overline{F}_{c-1}^V := \varnothing$. Stage one has steps $j = 1, \dots, k-b+1$. In step $j \in [c-b]$ of stage one, we extend copies of K_{b+j-1} in F_{b+j-1}^U to vertex-disjoint copies of K_{b+j} where possible. For each copy of K_{b+j-1} in F_{b+j-1}^U in turn we pick greedily, where possible, a common neighbour in U_1 which is not covered by $\overline{F}_{b-1}^U, \dots, \overline{F}_{b+j-2}^U, F_{b+j-1}^U$ or previously chosen common neighbours. The vertices selected lie in U_1, F_c^V is contained in V_1 and $U_1 \cap V_1 = \varnothing$ by (a), so no vertex of F_c^V is selected. Let $\overline{F}_{b+j-1}^U$ be the collection of copies of K_{b+j-1} in F_{b+j-1}^U which could not be extended and let F_{b+j}^U be the collection of vertex-disjoint copies of K_{b+j} which result from extending copies of K_{b+j-1} in F_{b+j-1}^U. In step $j \in [k-b+1] \setminus [c-b]$ of stage one, we extend copies of K_{b+j-1} in $F^U_{b+j-1} \cup F^V_{b+j-1}$ to vertex-disjoint copies of K_{b+j} where possible. For each copy of K_{b+j-1} in F^U_{b+j-1} in turn we pick greedily, where possible, a common neighbour in U_1 which is not covered by $\overline{F}^U_{b-1}, \dots, \overline{F}^U_{b+j-2}, F^U_{b+j-1}$ or previously chosen common neighbours. Let $\overline{F}^U_{b+j-1}$ be the collection of copies of K_{b+j-1} in F^U_{b+j-1} which could not be extended and let F^U_{b+j} be the collection of vertex-disjoint copies of K_{b+j} which result from extending copies of K_{b+j-1} in F^U_{b+j-1}. We do the same with F^V_{b+j-1} within V_1. We end stage one after step $k-b+1$. Set $\overline{F}^U_{k+1} := F^U_{k+1}$ and $\overline{F}^V_{k+1} := F^V_{k+1}$. At this point, we have collections $\overline{F}^U_i$ and $\overline{F}^V_j$ of vertex-disjoint copies of K_i and K_j respectively, for each $i = b, \dots, k+1$ and $j = c, \dots, k+1$, some of which may be empty. Let $\overline{F}^U = \bigcup_{i=b}^{k+1} \overline{F}^U_i$ and $\overline{F}^V = \bigcup_{i=c}^{k+1} \overline{F}^V_i$. Since $\overline{F}^U_i$ is the collection of copies of K_i which result from the extension of elements of F^U_b in stage one, we have $|\overline{F}^U| = |F^U_b|$; similarly, we have $|\overline{F}^V| = |F^V_c|$. Order the elements of $\overline{F}^U \cup \overline{F}^V$ with those in $\overline{F}^U$ coming before those in $\overline{F}^V$, those in each of $\overline{F}^U$ and $\overline{F}^V$ in increasing size order, and those in each of $\overline{F}^U$ and $\overline{F}^V$ of the same size in some arbitrary order. We will use this ordering when attempting to extend cliques in stage two.

We begin stage two with $\widetilde{F}^U_0 := \overline{F}^U$ and $\widetilde{F}^V_0 := \overline{F}^V$. Stage two has steps $j = 1, \dots, k-1$. In step $j \in [k-2]$ we attempt to extend each clique in $\widetilde{F}^U_{j-1} \cup \widetilde{F}^V_{j-1}$ of order at most k by one vertex using X_j. We extend cliques in the order mentioned previously. For each clique of order at most k in $\widetilde{F}^U_{j-1} \cup \widetilde{F}^V_{j-1}$ in turn we pick greedily, where possible, a common neighbour in X_j which is outside the previously chosen common neighbours. Let $\widetilde{F}^U_j$ and $\widetilde{F}^V_j$ be the collections of both cliques in $\widetilde{F}^U_{j-1}$ and $\widetilde{F}^V_{j-1}$ respectively of order $k+1$ and cliques resulting from the attempts to extend each clique of order at most k in $\widetilde{F}^U_{j-1}$ and $\widetilde{F}^V_{j-1}$ respectively by one vertex, no matter whether they were successful or not. In step $k-1$ we attempt to extend each clique of order at most k in $\widetilde{F}^U_{k-2} \cup \widetilde{F}^V_{k-2}$ by one vertex using vertices of G outside of $U_1 \cup V_1 \cup \left(\bigcup_{i=1}^{k-2} X_i\right)$ in a manner similar to that in earlier steps of stage two. We end stage two after step $k-1$ with collections $\widetilde{F}^U_{k-1}$ and $\widetilde{F}^V_{k-1}$ of $|F^U_b|$ and $|F^V_c|$ vertex-disjoint cliques in G respectively.

We shall prove that $\widetilde{F}^U_{k-1}$ and $\widetilde{F}^V_{k-1}$ are collections of $|F^U_b|$ and $|F^V_c|$ vertex-disjoint copies of K_{k+1} respectively. In fact, we shall prove that $\widetilde{F}^U_j$ and $\widetilde{F}^V_j$ are collections of $|F^U_b|$ and $|F^V_c|$ vertex-disjoint cliques of order at least $j+2$ respectively for each $j = b-2, \dots, k-1$. We shall first consider $\widetilde{F}^U_j$. We proceed by induction on j. The $j = b-2$ case is trivial. Consider $\widetilde{F}^U_j$ for $j \geq b-1$. By the induction hypothesis, $\widetilde{F}^U_{j-1}$ is a collection of $|F^U_b|$ vertex-disjoint cliques of order at least $j+1$. Hence, it suffices to show that the copies of K_{j+1} in $\widetilde{F}^U_{j-1}$ are all extended to copies of K_{j+2} in step j to prove our claim. Observe that this holds trivially when $|F^U_b| = 0$, so in what follows it is enough to consider when $|F^U_b| = \min\left\{|F^U|, \left\lfloor\frac{|U_2|}{2}\right\rfloor, d_1, s_1, s_2\right\}$.

Let f be a copy of K_{j+1} in $\widetilde{F}^U_{j-1}$ with $\ell \geq b$ vertices in U_1 and $\overline{f} = f \cap U_1$ be its corresponding clique in $\overline{F}^U$. Note that f has vertices from only $X_1, \dots, X_{j-1}, U_1$ and at most one vertex from each X_i. Define $I := \{i : |f \cap X_i| = 1\}$. Let $\overline{v}_i$ be the vertex of f in X_i for each $i \in I$.

First consider the case $j \leq k-2$. Every vertex v of U_1 has at least $\delta - |A| - \deg(v; U_1) - \sum_{h\neq j} \deg(v; X_h)$ neighbours in X_j and for each $i \in I$ the vertex $\overline{v}_i$ has at least $\delta - |A| - |U_2| - \deg(\overline{v}_i; U_1) - \sum_{h\neq j} \deg(\overline{v}_i; X_h)$ neighbours in X_j. By application of Lemma 2.2 and noting that $|X_j| = n - |A| - |U_2| - |U_1| - \sum_{h\neq j} |X_h|$, the number of common neighbours of f in X_j is at least

$$\begin{aligned} a_j :=& \sum_{v\in\overline{f}} \left(\delta - |A| - |U_2| - \deg(v; U_1) - \sum_{h\neq j} \deg(v; X_h)\right) \\ &+ \sum_{i\in I} \left(\delta - |A| - |U_2| - \deg(\overline{v}_i; U_1) - \sum_{h\neq j} \deg(\overline{v}_i; X_h)\right) \\ &- j\left(n - |A| - |U_2| - |U_1| - \sum_{h\neq j} |X_h|\right). \end{aligned} \tag{2.14}$$

Grouping terms together, we obtain

$$\begin{aligned} a_j =& (j+1)\delta - jn + (\ell-1)|U_2| - \left(\sum_{v\in f} \deg(v; U_1) - j|U_1|\right) \\ &- \sum_{h\neq j} \left(\sum_{v\in f} \deg(v; X_h) - j|X_h|\right) - |A|. \end{aligned} \tag{2.15}$$

Now we seek appropriate estimations of the terms in our expression. Since $\overline{f}$ could not be extended in step $\ell - b + 1$ of stage one, $\Gamma(\overline{f}; U_1)$ contains only vertices from elements of $\overline{F}^U_b, \dots, \overline{F}^U_\ell$ and $F^U_{\ell+1}$. For each $b \leq h \leq \ell$ the elements of $\overline{F}^U_h$ contain h vertices each while the elements of $F^U_{\ell+1}$ contain $\ell+1$ vertices each. Furthermore, we have $|F^U_b| = |\overline{F}^U_b| + \dots + |\overline{F}^U_\ell| + |F^U_{\ell+1}|$ by the definitions of $\overline{F}^U_b, \dots, \overline{F}^U_\ell$ and $F^U_{\ell+1}$. Hence, by applying Lemma 2.2 to U_1 and f, we obtain

$$\sum_{v\in f} \deg(v; U_1) - j|U_1| \leq \deg(f; U_1) \leq \deg(\overline{f}; U_1) \leq \ell|F^U_b| + |F^U_{\ell+1}|. \tag{2.16}$$

For $h \in [k-1]$, by applying Lemma 2.2 to X_h and f, we get

$$\sum_{v \in f} \deg(v; X_h) - j|X_h| \le \deg(f; X_h). \tag{2.17}$$

For $0 \le i < j$ let $f_i \in \widetilde{F}_i^U$ be the clique corresponding to f right before step $i+1$ of stage two, so $f_i = \overline{f} \cup \{\overline{v}_h : h \in I \cap [i]\}$. Let $h \in I$. By the induction hypothesis f_{h-1} is a clique of order at least $h+1$ with at least two vertices from U_1 and all other vertices from $\bigcup_{j=1}^{h-1} X_j$, so $\overline{v}_h$ has no neighbour in $\Gamma(f_{h-1}; X_h)$ by (e) applied with $(i, g) = (h, f_{h-1})$. Hence, we have $\deg(f_h; X_h) = 0$ for all $h \in I$. Together with (2.17) and the fact that $\deg(f; X_h) \le \deg(f_h; X_h)$ for all $h \in I$, we obtain

$$\sum_{h \in I} \left(\sum_{v \in f} \deg(v; X_h) - j|X_h| \right) \le \sum_{h \in I} \deg(f_h; X_h) = 0. \tag{2.18}$$

Given $i \notin I, i < j$, the clique f_i was not extended in step i of stage two. It follows that its common neighbourhood in X_i contains only vertices used to extend cliques that came before it in the size ordering, of which there were fewer than $m := |F_b^U| - |F_{\ell+1}^U|$. Noting further that $[j-1] \setminus I$ contains $\ell-2$ elements, by (2.17) and that $\deg(f; X_h) \le \deg(f_h; X_h)$ for all $h \in [j-1] \setminus I$, we get

$$\sum_{h \in [j-1] \setminus I} \left(\sum_{v \in f} \deg(v; X_h) - j|X_h| \right) \le \sum_{h \in [j-1] \setminus I} \deg(f_h; X_h) \le (\ell - 2)(m-1). \tag{2.19}$$

By (d) $|X_h| \le n - \delta$ for $h \in [k-1]$, so by (2.17) we have

$$\sum_{h=j+1}^{k-1} \left(\sum_{v \in f} \deg(v; X_h) - j|X_h| \right) \le \sum_{h=j+1}^{k-1} |X_h| \le (k-j-1)(n-\delta). \tag{2.20}$$

By (2.15), (2.16), (2.18), (2.19), (2.20) and that $m = |F_b^U| - |F_{\ell+1}^U|$, we obtain

$$\begin{aligned} a_j &\ge (j+1)\delta - jn - |A| + (\ell-1)|U_2| - \ell|F_b^U| - |F_{\ell+1}^U| - (\ell-2)m \\ &\quad - (k-j-1)(n-\delta) \\ &\ge k\delta - (k-1)n - |A| + (\ell-1)|U_2| - (\ell+1)|F_b^U| - (\ell-3)m. \end{aligned}$$

Since $\ell \ge b$ and by the definition of F_b^U and m we have $|U_2| \ge 2|F_b^U| \ge |F_b^U| + m$ and

$|F_b^U| \geq m$, we obtain

$$\begin{aligned}
a_j &\geq k\delta - (k-1)n - |A| + (\ell-1)|U_2| - (\ell+1)|F_b^U| - (\ell-3)m \\
&= k\delta - (k-1)n - |A| + (\ell-2)(|U_2| - |F_b^U| - m) + |U_2| - 3|F_b^U| + m \\
&\geq k\delta - (k-1)n - |A| + (b-2)(|U_2| - |F_b^U| - m) + |U_2| - 3|F_b^U| + m \\
&\geq k\delta - (k-1)n - |A| + (b-1)|U_2| - (2b-1)|F_b^U| + m.
\end{aligned}$$

Now by the definition of s_1 we have

$$\begin{aligned}
a_j &\geq k\delta - (k-1)n - |A| + (b-1)|U_2| - (2b-1)|F_b^U| + m \\
&\geq (2b-1)(s_1 - |F_b^U|) + m \geq m,
\end{aligned}$$

so we are indeed able to pick a vertex in X_j to extend f.

For the case $j = k-1$, an analysis analogous to (2.14) and (2.15) implies that the number of common neighbours of f outside of $U_1 \cup U_2 \cup V_1 \cup \left(\bigcup_{i=1}^{k-2} X_i\right)$ is at least

$$\begin{aligned}
a_{k-1} := k\delta - (k-1)n + (\ell-1)|U_2| - \left(\sum_{v\in f} \deg(v; U_1) - (k-1)|U_1|\right) \\
- \sum_{h=1}^{k-2}\left(\sum_{v\in f} \deg(v; X_h) - (k-1)|X_h|\right) - |V_1|.
\end{aligned}$$

By (2.16), (2.18), (2.19) and that $m = |F_b^U| - |F_{\ell+1}^U|$, we obtain

$$a_{k-1} \geq k\delta - (k-1)n - |V_1| + (\ell-1)|U_2| - (\ell+1)|F_b^U| - (\ell-3)m.$$

Since $\ell \geq b$ and by the definition of F_b^U and m we have $|U_2| \geq 2|F_b^U| \geq |F_b^U| + m$ and $|F_b^U| \geq m$, we obtain

$$\begin{aligned}
a_{k-1} &\geq k\delta - (k-1)n - |V_1| + (\ell-1)|U_2| - (\ell+1)|F_b^U| - (\ell-3)m \\
&= k\delta - (k-1)n - |V_1| + (\ell-2)(|U_2| - |F_b^U| - m) + |U_2| - 3|F_b^U| + m \\
&\geq k\delta - (k-1)n - |V_1| + (b-2)(|U_2| - |F_b^U| - m) + |U_2| - 3|F_b^U| + m \\
&\geq k\delta - (k-1)n - |V_1| + (b-1)|U_2| - (2b-1)|F_b^U| + m.
\end{aligned}$$

Now by the definition of s_2 we have

$$\begin{aligned}
a_{k-1} &\geq k\delta - (k-1)n - |V_1| + (b-1)|U_2| - (2b-1)|F_b^U| + m \\
&\geq (2b-1)(s_2 - |F_b^U|) + m \geq m,
\end{aligned}$$

so we are indeed able to pick a vertex outside of $U_1 \cup U_2 \cup V_1 \cup \left(\bigcup_{i=1}^{k-2} X_i\right)$ to extend f. This proves that copies of K_{j+1} in $\widetilde{F}^U_{j-1}$ are all extended to copies of K_{j+2} in step j and so by induction $\widetilde{F}^U_j$ is a collection of $|F^U_b|$ vertex-disjoint cliques of order at least $j+2$ for each $j = b-2, \dots, k-1$. In particular, $\widetilde{F}^U_{k-1}$ is a collection of $|F^U_b|$ vertex-disjoint copies of K_{k+1}.

The proof for the $\widetilde{F}^V_j$ case is very similar to that for the $\widetilde{F}^U_j$ case. We also proceed by induction on j and here the $j = c-2$ case is trivial. As in the $\widetilde{F}^U_j$ case, the desired outcome holds trivially when $|F^V_c| = 0$, so in what follows it is enough to consider when $|F^V_c| = \min\left\{|F^V|, d_2, t_1, t_2\right\}$. Let f be a copy of K_{j+1} in $\widetilde{F}^V_{j-1}$ with $\ell \geq c$ vertices in V_1 and $\overline{f} = f \cap V_1$ be its corresponding clique in $\overline{F}^V$. In particular, $\overline{f}$ has ℓ vertices. Define $I := \{i : |f \cap X_i| = 1\}$ and let $\overline{v}_i$ be the vertex of f in X_i for each $i \in I$. Note that f has vertices from only $X_1, \dots, X_{j-1}, V_1$ and at most one vertex from each X_i.

Consider $c-1 \leq j \leq k-1$. Let $m' := |F^V_c| - |F^V_{\ell+1}|$. An analysis analogous to (2.14) and (2.15) implies that the number of common neighbours of f in X_j is at least

$$\begin{aligned} b_j := (j+1)\delta - jn + (\ell-1)|V_2| - \left(\sum_{v \in f} \deg(v; V_1) - j|V_1|\right) \\ - \sum_{h \neq j} \left(\sum_{v \in f} \deg(v; X_h) - j|X_h|\right) - |A'|. \end{aligned} \tag{2.21}$$

By analyses similar to those for (2.16), (2.18), (2.19) and (2.20), we also have

$$\sum_{v \in f} \deg(v; V_1) - j|V_1| \leq \ell |F^V_c| + |F^V_{\ell+1}|, \tag{2.22}$$

$$\sum_{h \in I} \left(\sum_{v \in f} \deg(v; X_h) - j|X_h|\right) \leq 0, \tag{2.23}$$

$$\sum_{h \in [j-1] \setminus I} \left(\sum_{v \in f} \deg(v; X_h) - j|X_h|\right) \leq (\ell-2)(m' + |F^U_b| - 1) \quad \text{and} \tag{2.24}$$

$$\sum_{h=j+1}^{k-1} \left(\sum_{v \in f} \deg(v; X_h) - j|X_h|\right) \leq (k-j-1)(n-\delta), \tag{2.25}$$

respectively. The 'additional' term of $|F^U_b|$ in (2.24) (cf. (2.19)) arises because in each step of stage two we extend the cliques corresponding to F^U_b before those corresponding to F^V_c. Now by (2.21), (2.22), (2.23), (2.24) and (2.25) and that $m' = |F^V_c| - |F^V_{\ell+1}|$, we obtain

$$\begin{aligned}
b_j &\geq (j+1)\delta - jn - |A'| + (\ell-1)|V_2| - \ell|F_c^V| - |F_{\ell+1}^V| - (\ell-2)(m' + |F_b^U|) \\
&\qquad - (k-j-1)(n-\delta) \\
&\geq k\delta - (k-1)n - |A'| + (\ell-1)|V_2| - (\ell+1)|F_c^V| - (\ell-2)|F_b^U| - (\ell-3)m'.
\end{aligned}$$

Since $\ell \geq c$ and by the definition of F_b^U, F_c^V, d_1, d_2 and m we have $|F_c^V| \geq m'$ and $|V_2| \geq 2d_2 + d_1 \geq 2|F_c^V| + |F_b^U| \geq |F_c^V| + |F_b^U| + m'$, we obtain

$$\begin{aligned}
b_j &\geq k\delta - (k-1)n - |A'| + (\ell-1)|V_2| - (\ell+1)|F_c^V| - (\ell-2)|F_b^U| - (\ell-3)m' \\
&= k\delta - (k-1)n - |A'| + (\ell-2)(|V_2| - |F_c^V| - |F_b^U| - m') + |V_2| - 3|F_c^V| + m' \\
&\geq k\delta - (k-1)n - |A'| + (c-2)(|V_2| - |F_c^V| - |F_b^U| - m') + |V_2| - 3|F_c^V| + m' \\
&\geq k\delta - (k-1)n - |A'| + (c-1)|V_2| - (2c-1)|F_c^V| - (c-2)|F_b^U| + m'.
\end{aligned}$$

Now by the definition of t_1 we have

$$\begin{aligned}
b_j &\geq k\delta - (k-1)n - |A'| + (c-1)|V_2| - (2c-1)|F_c^V| - (c-2)|F_b^U| + m' \\
&\geq (2c-1)(t_1 - |F_c^V|) + m' + |F_b^U| \geq m' + |F_b^U|,
\end{aligned}$$

so we are indeed able to pick a vertex in X_j to extend f.

For the case $j = k-1$, an analogous analysis implies that the number of common neighbours of f outside of $U_1 \cup V_1 \cup V_2 \cup \left(\bigcup_{i=1}^{k-2} X_i\right)$ is at least

$$\begin{aligned}
b_{k-1} := k\delta - (k-1)n + (\ell-1)|V_2| - \left(\sum_{v\in f} \deg(v; V_1) - (k-1)|V_1|\right) \\
- \sum_{h=1}^{k-2}\left(\sum_{v\in f} \deg(v; X_h) - (k-1)|X_h|\right) - |U_1|.
\end{aligned}$$

By (2.22), (2.23) and (2.24) and that $m' = |F_c^V| - |F_{\ell+1}^V|$, we obtain

$$\begin{aligned}
b_{k-1} &\geq k\delta - (k-1)n - |U_1| + (\ell-1)|V_2| - \ell|F_c^V| - |F_{\ell+1}^V| - (\ell-2)(m' + |F_b^U|) \\
&\geq k\delta - (k-1)n - |U_1| + (\ell-1)|V_2| - (\ell+1)|F_c^V| - (\ell-2)|F_b^U| - (\ell-3)m'.
\end{aligned}$$

Since $\ell \geq c$ and by the definition of F_b^U, F_c^V, d_1, d_2 and m we have $|F_c^V| \geq m'$ and

$|V_2| \geq 2d_2 + d_1 \geq 2|F_c^V| + |F_b^U| \geq |F_c^V| + |F_b^U| + m'$, we obtain

$$\begin{aligned}
b_{k-1} &\geq k\delta - (k-1)n - |U_1| + (\ell-1)|V_2| - (\ell+1)|F_c^V| - (\ell-2)|F_b^U| - (\ell-3)m' \\
&= k\delta - (k-1)n - |U_1| + (\ell-2)(|V_2| - |F_c^V| - |F_b^U| - m') + |V_2| - 3|F_c^V| + m' \\
&\geq k\delta - (k-1)n - |U_1| + (c-2)(|V_2| - |F_c^V| - |F_b^U| - m') + |V_2| - 3|F_c^V| + m' \\
&\geq k\delta - (k-1)n - |U_1| + (c-1)|V_2| - (2c-1)|F_c^V| - (c-2)|F_b^U| + m'.
\end{aligned}$$

Now by the definition of t_2 we have

$$\begin{aligned}
b_{k-1} &\geq k\delta - (k-1)n - |U_1| + (c-1)|V_2| - (2c-1)|F_c^V| - (c-2)|F_b^U| + m' \\
&\geq (2c-1)(t_2 - |F_c^V|) + m' + |F_b^U| \geq m' + |F_b^U|,
\end{aligned}$$

so we are indeed able to pick a vertex outside of $U_1 \cup V_1 \cup V_2 \cup \left(\bigcup_{i=1}^{k-2} X_i\right)$ to extend f. This proves that copies of K_{j+1} in $\widetilde{F}_{j-1}^V$ are all extended to copies of K_{j+2} in step j and so by induction $\widetilde{F}_j^V$ is a collection of $|F_c^V|$ vertex-disjoint cliques of order at least $j+2$ for each $j = c-2, \ldots, k-1$. In particular, $\widetilde{F}_{k-1}^V$ is a collection of $|F_c^V|$ vertex-disjoint copies of K_{k+1}.

It remains to show that $\widetilde{F}_{k-1}^U \cup \widetilde{F}_{k-1}^V$ is a connected K_{k+1}-factor. Now $\widetilde{F}_{k-1}^U \cup \widetilde{F}_{k-1}^V$ consists of copies of K_k in G with either at least two vertices from U_1 and all other vertices from $\bigcup_{i=1}^{k-2} X_i$, or at least two vertices from V_1 and all other vertices from $\bigcup_{i=1}^{k-2} X_i$, so by (c) the copies of K_k in $\widetilde{F}_{k-1}^U \cup \widetilde{F}_{k-1}^V$ are pairwise K_{k+1}-connected. Hence, $\widetilde{F}_{k-1}^U \cup \widetilde{F}_{k-1}^V$ is a connected K_{k+1}-factor of size at least $(k+1)(|F_b^U| + |F_c^V|)$. If F^V is empty then we have $|F_c^V| = 0$, so $\widetilde{F}_{k-1}^U$ is a connected K_{k+1}-factor of size at least $(k+1)|F_b^U|$. $\square$

Lemma 2.23 is both the single partition analogue of and a straightforward consequence of Lemma 2.22. We will use it to find large connected K_{k+1}-factors when $\mathrm{int}_k(G)$ contains a copy of K_k, specifically in Lemmas 2.28 and 2.29.

Lemma 2.23. *Let $2 \leq b \leq k$ be integers. Let G be a graph on n vertices with minimum degree $\delta = \delta(G) > \frac{(k-1)n}{k}$. Suppose there is a partition of $V(G)$ into vertex classes $U_1, U_2, X_1, \ldots, X_{k-1}, A$ such that*

(a) there are no edges between U_1 and U_2,

(b) all copies of K_k in G with at least two vertices from U_1 and all other vertices from $\bigcup_{i=1}^{k-2} X_i$ are K_{k+1}-connected,

(c) $|X_i| \leq n - \delta$ *for* $i \in [k-1]$*, and*

(d) $X_i \cap \Gamma(g)$ *is an independent set for each* (i, g) *where* $i \in [k-2]$ *and* g *is a clique of order at least* i *with at least two vertices from* U_1 *and all other vertices from* $\bigcup_{j=1}^{i-1} X_j$*.*

Set $s_1 := \frac{k\delta-(k-1)n+(b-1)|U_2|-|A|}{2b-1}$*. Let* F *be a collection of vertex-disjoint copies of* K_b *in* U_1*. Then* G *contains a connected* K_{k+1}*-factor of size at least*

$$(k+1)\min\left\{|F|, \left\lfloor\frac{|U_2|}{2}\right\rfloor, s_1\right\}.$$

Proof. Fix a graph G and a partition of $V(G)$ with vertex classes $U_1, U_2, X_1, \ldots, X_{k-1}, A$ satisfying the lemma hypothesis. Define $V_1 = X' = A' = F^V := \varnothing$, $F^U := F$ and $V_2 := V(G) \setminus \left(\bigcup_{i=1}^{k-2} X_i\right)$. Set $d_1 := |V_2|$, $d_2 := 0$ and $c := b$. Then the result follows by application of Lemma 2.22, noting that $|V_2| \geq |U_2|$ and $|V_1| = 0$. □

We will find that Lemma 2.23 is sometimes inadequate, especially when $\mathrm{int}_k(G)$ does not contains a copy of K_k. This is partly due to the strength of conditions (b) and (d) forcing a large 'bad' set A. The conditions are necessary when $b > 2$, but we can weaken these conditions and sometimes do better when $b = 2$. We present this as Lemma 2.24. In this case, we require a smaller set of copies of K_k in G to be K_{k+1}-connected and $X_i \cap \Gamma(g)$ to be an independent set for a smaller set of copies g of K_{i+1} with $g \subseteq U_1 \cup \left(\bigcup_{j=1}^{i-1} X_j\right)$.

Lemma 2.24. *Let* $k \geq 2$ *be an integer. Let* G *be a graph on* n *vertices with minimum degree* $\delta = \delta(G) > \frac{(k-1)n}{k}$*. Suppose there is a partition of* $V(G)$ *into vertex classes* $U_1, U_2, X_1, \ldots, X_{k-1}, A$ *such that*

(a) *there are no edges between* U_1 *and* U_2*,*

(b) *all copies of* K_k *in* G *comprising an edge of* $G[U_1]$ *and a vertex from each of* $X_1, \ldots, X_{k-2}$ *are* K_{k+1}*-connected,*

(c) $|X_i| \leq n - \delta$ *for* $i \in [k-1]$*, and*

(d) $X_i \cap \Gamma(g)$ *is an independent set for each* (i, g) *where* $i \in [k-2]$ *and* g *is a copy of* K_{i+1} *comprising an edge of* $G[U_1]$ *and a vertex from each of* $X_1, \ldots, X_{i-1}$*.*

Let F *be a matching in* U_1*. Set* $q := k\delta - (k-1)n + |U_2| - |U_1| - |A|$*. Then* G *contains a connected* K_{k+1}*-factor of size at least* $(k+1)\min\{|F|, q\}$*.*

The proof approach is similar to that of Lemma 2.22; however, in this case we skip stage one and it turns out that we never fail to extend in stage two. Note that a copy of K_k extending an edge from F has at least q common neighbours outside of both U_1 (which contains F) and a 'bad' set A.

Proof. Let $\overline{F} \subseteq F$ satisfy $|\overline{F}| = \max\{0, \min\{|F|, q\}\}$. We will eventually extend each edge of $\overline{F}$ to a copy of K_{k+1} using vertices in $X_1, \ldots, X_{k-1}$. Note that the resultant copies of K_{k+1} will be K_{k+1}-connected by (b).

We build up our desired connected K_{k+1}-factor step-by-step, starting with the aforementioned matching $\widetilde{F}_0 := \overline{F}$ in U_1. We have steps $j = 1, \ldots, k-1$. In step j we extend each copy of K_{j+1} in $\widetilde{F}_{j-1}$ to a copy of K_{j+2} using X_j. For each copy of K_{j+1} in $\widetilde{F}_{j-1}$ in turn we pick greedily a common neighbour in X_j which is outside the previously chosen common neighbours to obtain a collection $\widetilde{F}_j$ of $|\overline{F}|$ vertex-disjoint copies of K_{j+2}. We claim that this is always possible for all $j \in [k-1]$. Observe that this holds trivially when $|\overline{F}| = 0$, so in what follows it is enough to consider when $|\overline{F}| = \min\{|F|, q\}$.

Let f be a copy of K_{j+1} in $\widetilde{F}_{j-1}$. Note that f has exactly one vertex in each X_i for $i < j$, exactly two vertices in U_1 and none elsewhere. Let v_1 and v_2 be the vertices of f in U_1, and let $\overline{v}_i$ be the vertex of f in X_i for each $i < j$. Every vertex v of U_1 has at least $\delta - |A| - |U_1| - \sum_{h\neq j} \deg(v; X_h)$ neighbours in X_j, and for each $i < j$ the vertex $\overline{v}_i$ has at least $\delta - |A| - |U_2| - |U_1| - \sum_{h\neq j} \deg(\overline{v}_i; X_h)$ neighbours in X_j. By application of Lemma 2.2 and noting that $|X_j| = n - |U_2| - |U_1| - |A| - \sum_{h\neq j} |X_h|$, the number of common neighbours of f in X_j is at least

$$\begin{aligned}
a_j := \sum_{i=1}^{2} &\left(\delta - |U_1| - |A| - \sum_{h\neq j} \deg(v_i; X_h)\right) \\
&+ \sum_{i=1}^{j-1} \left(\delta - |U_2| - |U_1| - |A| - \sum_{h\neq j} \deg(\overline{v}_i; X_h)\right) \\
&- j\left(n - |U_2| - |U_1| - |A| - \sum_{h\neq j} |X_h|\right).
\end{aligned}$$

Grouping terms together, we obtain

$$\begin{aligned} a_j = (j+1)\delta - jn + |U_2| - |U_1| - \sum_{h=1}^{j-1}\left(\sum_{v\in f}\deg(v;X_h) - j|X_h|\right) \\ - \sum_{h=j+1}^{k-1}\left(\sum_{v\in f}\deg(v;X_h) - j|X_h|\right) - |A|. \end{aligned} \tag{2.26}$$

For $h \in [k-1]$, by applying Lemma 2.2 to X_h and f, we get

$$\sum_{v\in f}\deg(v;X_h) - j|X_h| \leq \deg(f;X_h). \tag{2.27}$$

For $0 \leq i < j$ let $f_i \in \widetilde{F}_i$ be the clique corresponding to f right before step $i+1$, so $f_i = \{v_1, v_2\} \cup \{\overline{v}_h : h \in [i]\}$. Let $1 \leq h < j$. Now f_{h-1} is a clique of order $h+1$ comprising two vertices from U_1 and a vertex from each of $X_1, \ldots, X_{h-1}$, so $\overline{v}_h$ has no neighbour in $\Gamma(f_{h-1}; X_h)$ by (d) applied with $(i, g) = (h, f_{h-1})$. Hence, we have $\deg(f_h; X_h) = 0$ for all $h \in I$. Together with (2.27) and the fact that $\deg(f; X_h) \leq \deg(f_h; X_h)$ for all $h \in [j-1]$, we obtain

$$\sum_{h=1}^{j-1}\left(\sum_{v\in f}\deg(v;X_h) - j|X_h|\right) \leq \sum_{h=1}^{j-1}\deg(f_h;X_h) = 0. \tag{2.28}$$

By (c) $|X_h| \leq n - \delta$ for $h \in [k-1]$, so by (2.27) we have

$$\sum_{h=j+1}^{k-1}\left(\sum_{v\in f}\deg(v;X_h) - j|X_h|\right) \leq \sum_{h=j+1}^{k-1}|X_h| \leq (k-j-1)(n-\delta). \tag{2.29}$$

Putting together (2.26), (2.28) and (2.29), we get $a_j \geq q \geq |\overline{F}|$, so we are indeed able to pick a vertex in X_j to extend f. This proves that copies of K_{j+1} in $\widetilde{F}_{j-1}$ are all extended to copies of K_{j+2} in step j. Therefore, we terminate after step $k-1$ with a collection $\widetilde{F}_{k-1}$ of $|\overline{F}|$ vertex-disjoint copies of K_{k+1} in G. All copies of K_k in G comprising an edge of $G[U_1]$ and a vertex from each of $X_1, \ldots, X_{k-2}$ are K_{k+1}-connected by (b), so $\widetilde{F}_{k-1}$ is in fact a connected K_{k+1}-factor in G of size at least $(k+1)\min\{|F|, q\}$. □

2.5 The Proof of Lemma 2.17

In this section we provide a proof of Lemma 2.17, our stability result for graphs with at least two K_{k+1}-components where each K_{k+1}-component contains a copy of K_{k+2}.

We start with a couple of preparatory lemmas which collect some observations about K_{k+1}-components. Recall that the exterior $\mathrm{ext}(C)$ of a K_{k+1}-component C of a graph G is the set of vertices of C which are in no other K_{k+1}-component of G.

The first lemma states that K_{k+1}-components cannot be too small, that there are no edges between the exteriors of different components and that certain spots in a K_{k+1}-component induce a graph with minimum degree $k\delta - (k-1)n$.

Lemma 2.25. *Let $k \in \mathbb{N}$ and let G be a graph on n vertices with minimum degree $\delta(G) > \frac{(k-1)n}{k}$. Then*

(i) each K_{k+1}-component C satisfies $|C| > \delta$,

(ii) for distinct K_{k+1}-components C and C' there are no edges between $\mathrm{ext}(C)$ *and* $\mathrm{ext}(C')$,

(iii) for each K_{k+1}-component C, each copy $u_1 \dots u_{k-1}$ of K_{k-1} of C, and $U = \{v : u_1 \dots u_{k-1}v \in C\}$, we have $\delta(G[U]) \geq k\delta - (k-1)n$ and $|U| \geq k\delta - (k-1)n + 1$.

Proof. For (i) let M be a maximal clique in C. Note that $|M| \geq k+1$. Count $\rho := \sum_{m \in M, u \in V(G)} \mathbf{1}_{\{mu \in E(G)\}}$ in two ways. On the one hand,

$$\rho = \sum_{m \in M} \sum_{u \in V(G)} \mathbf{1}_{\{mu \in E(G)\}} = \sum_{m \in M} \deg(m) \geq |M|\delta.$$

On the other hand, noting that each vertex of G which is not a vertex of C is adjacent to at most $k-1$ vertices of M, while each vertex of C is adjacent to at most $|M|-1$ vertices of M, we obtain

$$\begin{aligned} \rho &= \sum_{u \in V(G)} \sum_{m \in M} \mathbf{1}_{\{mu \in E(G)\}} = \sum_{u \in V(G)} \deg(u; M) \\ &\leq \sum_{u \in C} |M| - 1 + \sum_{u \notin C} k - 1 = |C|(|M| - k) + (k-1)n \end{aligned}$$

and so $|M|\delta - (k-1)n \leq |C|(|M| - k)$. Since $(k-1)n < k\delta$ we conclude that $|C| > \delta$.

For (ii) suppose that u is a vertex in $\mathrm{ext}(C)$, v is a vertex in $\mathrm{ext}(C')$ and uv is an edge in G. Apply Lemma 2.3 to complete uv to a copy of K_k in G. Since this copy of K_k contains a vertex from each of $\mathrm{ext}(C)$ and $\mathrm{ext}(C')$, it is in both C and C', which is a contradiction.

For (iii) note that U is non-empty as $u_1 \dots u_{k-1}$ is a copy of K_{k-1} of C. Let $u_k \in U$, so by definition $u_1 \dots u_k \in C$. Since $\Gamma(u_1, \dots, u_k) \subseteq U$, by Lemma 2.2 we have $\deg(u_k;U) = \deg(u_1,\dots,u_k) \geq k\delta-(k-1)n$. Now $\{u_k\}\cup\Gamma(u_k;U)\subseteq U$ so $|U|\geq k\delta-(k-1)n+1$. $\square$

The next lemma says that graphs with more than one K_{k+1}-component have a non-empty K_{k+1}-interior and gives a lower bound on the size of said K_{k+1}-interior. This is an easy consequence of Lemma 2.25(i).

Lemma 2.26. *Let $k \geq 2$ be an integer. Let G be a graph on n vertices with minimum degree $\delta(G)=\delta>\frac{(k-1)n}{k}$ and more than one K_{k+1}-component. Then*

(i) $|\operatorname{int}_k(G)| \geq 2\delta-n+2>0$, and

(ii) for each K_{k+1}-component C of G we have $|\operatorname{ext}(C)|\leq n-\delta-1$.

Proof. For (i), let C and C' be distinct K_{k+1}-components of G. Lemma 2.25(i) tells us that $|C|,|C'|>\delta$. $\operatorname{int}_k(G)$ contains all vertices which are vertices of both C and C' so $|\operatorname{int}_k(G)|\geq|C_1|+|C_2|-n\geq 2\delta-n+2>0$.

For (ii), let C' be a K_{k+1}-component of G distinct from C. Now $\operatorname{ext}(C)$ contains no vertex of C' and by Lemma 2.25(i) we have $|C'|>\delta$, so it follows that $|\operatorname{ext}(C)|\leq n-\delta-1$. $\square$

Central to our proof of Lemma 2.17 is the construction of sufficiently large connected K_{k+1}-factors. Lemma 2.25(iii) enables us to find spots in a K_{k+1}-component which induce a graph with minimum degree $k\delta-(k-1)n$. In our proof of Lemma 2.17, we will often use this to find a large matching in such spots (this is possible due to Lemma 2.1(i)). The family of configurations introduced in Section 2.3.1, the structural analysis in Section 2.4.1 and our construction procedures in Section 2.4.2 will then enable us to extend such a matching to a connected K_{k+1}-factor.

As mentioned in Section 2.3.1, our proof of Lemma 2.17 considers two cases – when $\operatorname{int}_k(G)$ contains a copy of K_k and when $\operatorname{int}_k(G)$ does not contains a copy of K_k. In the first case, we prove that if $\operatorname{int}_k(G)$ contains a copy of K_k then $\mathrm{CKF}_{k+1}(G)\geq \mathrm{pp}_k(n,\delta+\eta n)$. In fact, we prove the contrapositive statement in Lemma 2.27, which involves proving that if $\mathrm{CKF}_{k+1}(G)<\mathrm{pp}_k(n,\delta+\eta n)$, then G does not contain the configurations $\dagger_k(\ell,j)$ for all $1\leq j<\ell\leq k$: it follows immediately from the definition of $\operatorname{int}_k(G)$ that any copy of K_k in $\operatorname{int}_k(G)$ acts as the 'central' copy of K_k in an instance of the configuration $\dagger_k(k,1)$. We use structural properties of these configurations proved in Section 2.4.1 and clique factor construction procedures from Section 2.4.2 to do so.

Lemma 2.27. *Let $k \geq 3$ be an integer and $\mu > 0$. Let $\eta > 0$ and $n \in \mathbb{N}$ satisfy $\frac{1}{n} \ll \eta \ll \mu, \frac{1}{k}$. Let G be a graph on n vertices with minimum degree $\delta(G) \geq \delta \geq \left(\frac{k-1}{k} + \mu\right) n$ and at least two K_{k+1}-components. Suppose that $\mathrm{CKF}_{k+1}(G) < \mathrm{pp}_k(n, \delta + \eta n)$. Then G does not contain the configuration $\dagger_k(\ell, j)$ for all $j, \ell \in \mathbb{N}$ such that $j < \ell \leq k$. In particular, $\mathrm{int}_k(G)$ is K_k-free.*

We prove Lemma 2.27 by induction on some function $f(j, \ell)$. While the specific choices of proof method and induction function are motivated by technical considerations, let us discuss the underlying ideas of our proof. We work in the context where $\mathrm{int}_k(G)$ contains a copy of K_k, which is equivalent to G containing $\dagger_k(k, 1)$ by definition, and we want to construct a sufficiently large connected K_{k+1}-factor. To this end, we seek increasingly structured graph configurations; this is achieved as a consequence of Lemma 2.19. Roughly speaking, the larger the interfaces between copies of K_k in different K_{k+1}-components, the more highly structured the configuration. Eventually, we arrive at a configuration of the form $\dagger_k(\ell, \ell - 1)$, which represents the 'pinnacle of evolution' with copies of K_k in different K_{k+1}-components that share a copy of K_{k-1}. These possess sufficient structure for the construction of a sufficiently large connected K_{k+1}-factor; we handle them in Lemmas 2.28 and 2.29. For technical reasons, we need treat $\dagger_k(2, 1)$ separately. We first consider the $j + 1 = \ell = 2$ case.

Lemma 2.28. *Let $k \geq 3$ be an integer. Let $\mu, \eta > 0$ and $n \in \mathbb{N}$ satisfy $\frac{1}{n} \ll \eta \ll \mu, \frac{1}{k}$. Let G be a graph on n vertices with minimum degree $\delta(G) \geq \delta \geq \left(\frac{k-1}{k} + \mu\right) n$ and at least two K_{k+1}-components. Suppose that $\mathrm{CKF}_{k+1}(G) < \mathrm{pp}_k(n, \delta + \eta n)$. Then G does not contain the configuration $\dagger_k(2, 1)$.*

Proof. Let $0 < \eta < \min\{\frac{1}{1000k^2}, \eta_0(k, \mu)\}$ and $n_1 := \max\{n_2(k, \mu, \eta), \frac{2}{\eta}\}$ with $\eta_0(k, \mu)$ and $n_2(k, \mu, \eta)$ given by Lemma 2.10. Suppose that G contains the configuration $\dagger_k(2, 1)$, so by Definition 2.18 there are vertices $u_1, \ldots, u_k, v_2, w_{2,1}$ in $V(G)$ such that (CG1)–(CG3) hold. Say $f := u_2 \ldots u_k$ lies in distinct K_{k+1}-components $C_1, \ldots, C_p$ and $f' := u_1 u_3 \ldots u_k$ lies in distinct K_{k+1}-components $C'_1, \ldots, C'_q$ with $p, q \geq 2$ and $f_0 := u_1 \ldots u_k \in C_1 = C'_1$. Define

$$U_i = \{y : fy \in C_i\} \text{ for } i \in [p] \text{ and } V_j = \{y : f'y \in C'_i\} \text{ for } j \in [q],$$

so $\{U_i\}_{i \in [p]}$ and $\{V_j\}_{j \in [q]}$ partition $\Gamma(f)$ and $\Gamma(f')$ respectively. Since any vertex $x \in U_i \cap V_j$ satisfies $x \in \Gamma(f_0) \subseteq U_1 \cap V_1$, we have

$$U_i \cap V_j = \varnothing \quad \text{for all } (i, j) \in ([p] \times [q]) \setminus \{(1, 1)\}. \tag{2.30}$$

By Lemma 2.25(iii) we have

$$|U_i|, |V_j| \geq k\delta - (k-1)n + 1 \tag{2.31}$$

for all $i \in [p], j \in [q]$. Since we have $\deg(u_1, \ldots, u_k) \geq k\delta - (k-1)n > 0$ by Lemma 2.2, we can pick a vertex $w \in \Gamma(f_0) \subseteq U_1 \cap V_1$. Now w has no neighbours in $\left(\bigcup_{1<i\leq p} U_i\right) \cup \left(\bigcup_{1<j\leq q} V_j\right)$, so by (2.31) we have

$$\delta \leq \deg(w) < n - \sum_{1<i\leq p} |U_i| - \sum_{1<j\leq q} |V_j| \leq n - 2(k\delta - (k-1)n + 1) \tag{2.32}$$

and we obtain $\delta \leq \frac{(2k-1)n-3}{2k+1} < \left(\frac{k}{k+1} - 2\eta\right) n$. By Lemma 2.2 we have

$$\begin{aligned} |\Gamma(f)| &= \sum_{i\in[p]} |U_i| \geq (k-1)\delta - (k-2)n \quad \text{and} \\ |\Gamma(f')| &= \sum_{j\in[q]} |V_j| \geq (k-1)\delta - (k-2)n, \end{aligned} \tag{2.33}$$

so we obtain

$$\begin{aligned} |U_1| = |\Gamma(f)| - \sum_{1<i\leq p} |U_i| &\overset{(2.32)}{\geq} |\Gamma(f)| - (n - \delta - 1) + \sum_{1<j\leq q} |V_j| \\ &\overset{(2.33)}{\geq} k\delta - (k-1)n + 1 + \sum_{1<j\leq q} |V_j| \overset{(2.31)}{\geq} 2(k\delta - (k-1)n + 1). \end{aligned} \tag{2.34}$$

By symmetry we have $|V_1| \geq 2(k\delta - (k-1)n + 1)$. We have $p, q \geq 2$, so $U_2, V_2 \neq \varnothing$ and we can pick $u \in U_2$ and $v \in V_2$. Now u and v have no neighbours in U_1 and V_1 respectively, so we conclude that

$$|U_1|, |V_1| < n - \delta. \tag{2.35}$$

We now define

$$\begin{aligned}
&X_i, Y_i := \Gamma(u_{i+2}, \ldots, u_k) \setminus \Gamma(u_{i+1}) \text{ for } i \in [k-1] \setminus \{1\}, \\
&X_1, X_1' := \Gamma(u_3, \ldots, u_k) \setminus \Gamma(u_2), \quad Y_1, Y_1' := \Gamma(u_3, \ldots, u_k) \setminus \Gamma(u_1), \\
&X_i' := \Gamma(u_2, \ldots, u_i, u_{i+2}, \ldots, u_k) \setminus \Gamma(u_{i+1}) \text{ for } i \in [k-1] \setminus \{1\}, \\
&Y_i' := \Gamma(u_1, u_3, \ldots, u_i, u_{i+2}, \ldots, u_k) \setminus \Gamma(u_{i+1}) \text{ for } i \in [k-1] \setminus \{1\}, \\
&Z_i' := X_{i+1}' \cap Y_{i+1}', \quad Z_i'' := Z_i' \cap \Gamma(w) \text{ for } i \in [k-2], \\
&A := \bigcup_{i=1}^{k-1} (X_i \setminus X_i') = \bigcup_{i=2}^{k-1} (X_i \setminus X_i'), \quad A' := \bigcup_{i=1}^{k-1} (Y_i \setminus Y_i') = \bigcup_{i=2}^{k-1} (Y_i \setminus Y_i'), \\
&A'' := \left(\bigcup_{i=1}^{k-2} Z_i'\right) \setminus \Gamma(w), \quad B := A \cup A' \cup A''.
\end{aligned}$$

Note that A is the set of vertices in G with at least two non-neighbours in f. Count $\rho := \sum_{v\in V(G),u\in f} \mathbf{1}_{\{vu\notin E(G)\}}$ in two ways. On the one hand,

$$\rho = \sum_{u\in f}\left(\sum_{v\in V(G)} \mathbf{1}_{\{vu\notin E(G)\}}\right) = \sum_{u\in f} |V(G)\setminus \Gamma(u)| \leq (k-1)(n-\delta).$$

On the other hand, we have

$$\rho = \sum_{v\in V(G)}\left(\sum_{u\in f} \mathbf{1}_{\{vu\notin E(G)\}}\right) = \sum_{v\in V(G)} |f\setminus \Gamma(v)| \geq n - |\Gamma(f)| + |A|.$$

Hence, by (2.33) we obtain $|A| \leq \sum_{i\in[p]} |U_i| - n + (k-1)(n-\delta)$. Similarly, A' is the set of vertices in G with at least two non-neighbours in f'. Hence, $|A'| \leq \sum_{j\in[q]} |V_j| - n + (k-1)(n-\delta)$. No vertex in $\left(\bigcup_{1<i\leq p} U_i\right) \cup \left(\bigcup_{1<j\leq q} V_j\right) \cup A''$ is adjacent to w, so $|A''| \leq n-\delta-1-\sum_{1<i\leq p}|U_i| - \sum_{1<j\leq q} |V_j|$. Therefore, we conclude that

$$|B| \leq |U_1| + |V_1| - 2[k\delta - (k-1)n + 1] - (n-\delta-1). \tag{2.36}$$

Let $1 < h \leq p$. By Lemma 2.25(iii) we have $\delta(G[U_h]) \geq k\delta - (k-1)n$, so we have a matching M in U_h with $|M| = \min\{k\delta - (k-1)n, \left\lfloor \frac{|U_h|}{2} \right\rfloor\}$ by Lemma 2.1(i). We shall check the conditions to apply Lemma 2.23 for $b = 2$ with U_h, $\bigcup_{i\neq h} U_i$, $Z''_1, \dots, Z''_{k-2}$, X_1, B and M as U_1, U_2, $X_1, \dots, X_{k-1}$, A and F respectively. By definition U_h and $\bigcup_{i\neq h} U_i$ partition $\Gamma(f)$. For each $i \in [k-2]$ the set Z''_i consists of the neighbours of w whose only non-neighbour in f_0 is u_{i+2}. The set X_1 consists of the vertices whose only non-neighbour in f is u_2. The set B consists of the non-neighbours of w whose only non-neighbour in f_0 is u_{i+2} for some $i \in [k-2]$ and the vertices with at least two non-neighbours in f or at least two non-neighbours in f'. Hence, U_h, $\bigcup_{i\neq h} U_i$, $Z''_1, \dots, Z''_{k-2}$, X_1, B form a partition of $V(G)$ such that there are no edges between U_h and $\bigcup_{i\neq h} U_i$. Note that $Z''_i \subseteq V(G)\setminus\Gamma(u_{i+2})$ for $i \in [k-2]$ and $X_1 \subseteq V(G)\setminus \Gamma(u_2)$, so $|Z''_i| \leq n-\delta$ for each $i\in[k-2]$ and $|X_1| \leq n-\delta$. For each $(e,i) \in E(G[U_h]) \times [k-2]$, by applying Lemma 2.20 for $i+1$ with $u_1, \dots, u_k, w$ as themselves, C_h as C_1, C_1 as C_2 and e as uv, we have that $Z''_i \cap \Gamma(e)$ is an independent set. Furthermore, all copies of K_k in G with at least two vertices from U_h and all other vertices from $\left(\bigcup_{i=1}^{k-2} Z''_i\right)$ are K_{k+1}-connected: we can construct a K_{k+1}-walk from such a copy g of K_k to f_0 by a step-by-step vertex replacement of the vertices of g with the vertices of f_0.

Now since the requisite conditions are satisfied, we apply Lemma 2.23 for $b = 2$ with U_h, $\bigcup_{i\neq h} U_i$, $Z''_1, \dots, Z''_{k-2}$, X_1, B and M as U_1, U_2, $X_1, \dots, X_{k-1}$, A and F

respectively; since $\sum_{i\neq h}|U_i| \geq |U_1|$ and by noting (2.34), (2.35) and (2.36), we obtain that $\mathrm{CKF}_{k+1}(G)$ is at least

$$(k+1)\min\left\{k\delta-(k-1)n, \left\lfloor\frac{|U_h|}{2}\right\rfloor, \left\lfloor\frac{\sum_{i\neq h}|U_i|}{2}\right\rfloor, \frac{k\delta-(k-1)n+\sum_{i\neq h}|U_i|-|B|}{3}\right\}$$
$$\geq (k+1)\min\left\{k\delta-(k-1)n, \left\lfloor\frac{|U_h|}{2}\right\rfloor\right\}.$$

Since (2.5) and (2.7) hold and $\mathrm{CKF}_{k+1}(G) < \mathrm{pp}_k(n, \delta+\eta n)$, we deduce that

$$|U_h| < 2(k\delta-(k-1)n) \text{ and } |U_h| < \frac{(k-1)(\delta+3\eta n)-(k-2)n}{r'}, \tag{2.37}$$

where $r' := r_p(n, \delta+\eta n)$. If furthermore $\delta \in \left[\left(\frac{k-1}{k}+\mu\right)n, \left(\frac{2k-1}{2k+1}-2\eta\right)n\right]$, by (2.8) we also deduce that

$$|U_h| < \frac{3}{2}(k\delta-(k-1)n)+1. \tag{2.38}$$

By symmetry we also have that for all $1 < j \leq q$,

$$|V_j| < 2(k\delta-(k-1)n) \text{ and } |V_j| < \frac{(k-1)(\delta+3\eta n)-(k-2)n}{r'}, \tag{2.39}$$

and if furthermore $\delta \in \left[\left(\frac{k-1}{k}+\mu\right)n, \left(\frac{2k-1}{2k+1}-2\eta\right)n\right]$ then

$$|V_j| < \frac{3}{2}(k\delta-(k-1)n)+1. \tag{2.40}$$

Let $1 < i \leq p$ and $1 < j \leq q$. By Lemma 2.1(i) and Lemma 2.25(iii) and noting (2.37) and (2.39), there are matchings M_u and M_v in U_i and V_j respectively with $|M_u| = \min\{\left\lfloor\frac{|U_i|}{2}\right\rfloor, k\delta-(k-1)n\} = \left\lfloor\frac{|U_i|}{2}\right\rfloor$ and $|M_v| = \min\{\left\lfloor\frac{|V_j|}{2}\right\rfloor, k\delta-(k-1)n\} = \left\lfloor\frac{|V_j|}{2}\right\rfloor$. Without loss of generality, suppose $|V_j| \geq |U_i|$. Let uv be an edge in U_i. Note that u and v each has at most $n-\delta-1-\sum_{h\neq i}|U_h| \overset{(2.33)}{\leq} |U_i|-(k\delta-(k-1)n+1)$ non-neighbours outside of $\left(\bigcup_{h\neq i}U_h\right)\cup\{u,v\}$. Hence, $\Gamma(u,v)$ has at most $2[|U_i|-(k\delta-(k-1)n+1)]$ non-neighbours outside of $\left(\bigcup_{h\neq i}U_h\right)\cup\{u,v\}$. Since we have $|V_j| \geq |U_i|$ and by (2.30) V_j is disjoint from $\left(\bigcup_{h\neq i}U_h\right)\cup\{u,v\}$, by (2.37) we obtain

$$|\Gamma(u,v;V_j)| \geq |V_j| - 2[|U_i|-(k\delta-(k-1)n+1)] > 0. \tag{2.41}$$

Hence, we may pick $x \in \Gamma(u,v;V_j)$. Suppose $\Gamma(u,v;V_j)$ is an independent set. Now x has no neighbour in $\left(\bigcup_{h\neq j}V_h\right)\cup\Gamma(u,v;V_j)$, so we have $n-\delta \geq \sum_{h\neq j}|V_h|+|\Gamma(u,v;V_j)| \geq |\Gamma(f')| - 2\left[|U_i|-(k\delta-(k-1)n+1)\right]$ by (2.33) and (2.41). Hence, we obtain

$$|V_j| \geq |U_i| \geq k\delta-(k-1)n+1+\frac{|\Gamma(f')|-(n-\delta)}{2} \overset{(2.33)}{\geq} \frac{3}{2}(k\delta-(k-1)n)+1. \tag{2.42}$$

Note that $U_i \cap V_j = \varnothing$ and w has no neighbours in $U_i \cup V_j$, so $n - \delta > |U_i \cup V_j| \geq 3[k\delta - (k-1)n + 1]$, which implies $\delta \leq \frac{(3k-2)n-4}{3k+1} < \left(\frac{2k-1}{2k+1} - 2\eta\right)n$. However, this means that (2.42) contradicts (2.38). Therefore, there is an edge $u'v'$ in $\Gamma(u, v; V_j)$. Then $uvu'v'u_3 \ldots u_k$ is a copy of K_{k+2} with $uvu_3 \ldots u_k \in C_i$ and $u'v'u_3 \ldots u_k \in C'_j$ so $C_i = C'_j$. Noting that $i \in [p] \setminus \{1\}$ and $j \in [q] \setminus \{1\}$ are arbitrary, we deduce that in fact $p = q = 2$.

We shall check the conditions to apply Lemma 2.22 for $b = c = 2$ with U_2, U_1, V_2, V_1, $Z''_1, \ldots, Z''_{k-2}$, X_1, Y_1, B, B, M_u and M_v as the inputs U_1, U_2, V_1, V_2, $X_1, \ldots, X_{k-1}$, X', A, A', F^U and F^V respectively. We know from earlier that U_2, U_1, $Z''_1, \ldots, Z''_{k-2}$, X_1, B form a partition of $V(G)$ such that there are no edges between U_2 and U_1, that $|Z''_i| \leq n - \delta$ for each $i \in [k-2]$ and $|X_1| \leq n - \delta$, that $Z''_i \cap \Gamma(e)$ is an independent set for each $(e, i) \in E(G[U_2]) \times [k-2]$ and that all copies of K_k in G with at least two vertices from U_2 and all other vertices from $\left(\bigcup_{i=1}^{k-2} Z''_i\right)$ are K_{k+1}-connected. By swapping the roles of u_1 and u_2, we also have that V_2, V_1, $Z''_1, \ldots, Z''_{k-2}$, Y_1, B form a second partition of $V(G)$ such that there are no edges between V_1 and V_2, that $|Y_1| \leq n - \delta$, that $Z''_i \cap \Gamma(e)$ is an independent set for each $(e, i) \in E(G[V_2]) \times [k-2]$ and that all copies of K_k in G with at least two vertices from V_2 and all other vertices from $\left(\bigcup_{i=1}^{k-2} Z''_i\right)$ are K_{k+1}-connected.

Since the conditions are satisfied, we apply Lemma 2.22 with the given inputs and $d_1 = d_2 = \left\lfloor \frac{|V_1|}{3} \right\rfloor$ to obtain that $\mathrm{CKF}_{k+1}(G)$ is at least

$$(k+1)\min\left\{\left\lfloor \tfrac{|U_2|}{2} \right\rfloor, \left\lfloor \tfrac{|U_1|}{2} \right\rfloor, \left\lfloor \tfrac{|V_1|}{3} \right\rfloor, \tfrac{k\delta-(k-1)n-|B|+|U_1|}{3}, \tfrac{k\delta-(k-1)n+|U_1|-|V_2|}{3}\right\}$$
$$+ (k+1)\min\left\{\left\lfloor \tfrac{|V_2|}{2} \right\rfloor, \left\lfloor \tfrac{|V_1|}{3} \right\rfloor, \tfrac{k\delta-(k-1)n-|B|+|V_1|-|U_2|/2}{3}, \tfrac{k\delta-(k-1)n+|V_1|-3|U_2|/2}{3}\right\}.$$

By (2.32), (2.33), (2.34), (2.35), (2.36) and (2.37), this is at least

$$(k+1)\left(\min\left\{\left\lfloor \tfrac{|U_2|}{2} \right\rfloor, \tfrac{2(k\delta-(k-1)n)}{3}\right\} + \min\left\{\left\lfloor \tfrac{|V_2|}{2} \right\rfloor, \tfrac{2(k\delta-(k-1)n)}{3} - \tfrac{|U_2|}{6}\right\}\right).$$

Now by Lemma 2.25(iii) we have $\left\lfloor \frac{|U_2|}{2} \right\rfloor, \left\lfloor \frac{|V_2|}{2} \right\rfloor \geq \frac{(k\delta-(k-1)n}{2}$ and we have (2.37), so in fact it is at least $(k+1)(k\delta - (k-1)n) \geq \mathrm{pp}_k(n, \delta + \eta n)$ by (2.7). However, this is a contradiction so G does not contain $\dagger_k(2, 1)$. $\square$

Next, we consider the $3 \leq j + 1 = \ell \leq k$ case.

Lemma 2.29. *Let $k, \ell \in \mathbb{N}$ satisfy $3 \leq \ell \leq k$. Let $\mu, \eta > 0$ and $n \in \mathbb{N}$ satisfy $\frac{1}{n} \ll \eta \ll \mu, \frac{1}{k}$. Let G be a graph on n vertices with minimum degree $\delta(G) \geq \delta \geq \left(\frac{k-1}{k} + \mu\right)n$*

and at least two K_{k+1}-components. Suppose that $\mathrm{CKF}_{k+1}(G) < \mathrm{pp}_k(n, \delta + \eta n)$ *and G does not contain $\dagger_k(2,1)$, $\dagger_k(\ell-1, \ell-2)$ or $\dagger_k(\ell-1,1)$. Then G does not contain the configuration $\dagger_k(\ell, \ell-1)$.*

Proof. Let $0 < \eta < \min\{\frac{1}{1000k^2}, \eta_0(k,\mu)\}$ and $n_1 := \max\{n_2(k,\mu,\eta), \frac{2}{\eta}\}$ with $\eta_0(k,\mu)$ and $n_2(k,\mu,\eta)$ given by Lemma 2.10. Suppose that G contains the configuration $\dagger_k(\ell, \ell-1)$, so by Definition 2.18 there are vertices $u_1, \ldots, u_k, v_\ell, w_{\ell,1}, \ldots, w_{\ell,\ell-1}$ in $V(G)$ such that (CG1)–(CG3) hold. For each $i \in [\ell-1]$ set $f_i := u_1 \ldots u_{i-1}u_{i+1} \ldots u_{\ell-1}u_{\ell+1} \ldots u_k$. Let $f := u_1 \ldots u_{\ell-1}u_{\ell+1} \ldots u_k$. Observe that u_ℓ and v_ℓ are distinct vertices: if not, $fu_\ell = fv_\ell$ would be a copy of K_k in two different K_{k+1}-components, giving a contradiction. Furthermore, $v_\ell u_\ell$ is not an edge: if not, $fu_\ell v_\ell$ would be a copy of K_{k+1} in G where fu_ℓ and fv_ℓ belong to different K_{k+1}-components, giving a contradiction. Hence, $u_1, \ldots, u_k, v_\ell, w_{\ell,1}, \ldots, w_{\ell,\ell-1}$ are all distinct vertices. Set $X_i := \Gamma(f_i) \setminus \{u_i\}$ for $i \in [\ell-1]$, $X := \bigcup_{i=1}^{\ell-1} X_i$ and $Y_j := \Gamma(u_\ell, v_\ell, w_{\ell,j}; X)$ for $j \in [\ell-1]$.

We claim that $Y_j = \varnothing$ for some $j \in [\ell-1]$. Indeed, suppose that $Y_j \neq \varnothing$ for all $j \in [\ell-1]$. Pick $y_j \in Y_j$ for each $j \in [\ell-1]$. Fix a function $\phi\colon [\ell-1] \to [\ell-1]$ such that $y_j \in X_{\phi(j)}$. Observe that $y_j f_{\phi(j)} u_\ell \in C$ for each $j \in [\ell-1]$: if not, then $fu_\ell \in C$, $fv_\ell \notin C$ and $y_j f_{\phi(j)} u_\ell \notin C$ would yield $\dagger_k(2,1)$ with fu_ℓ as the 'central' copy of K_k and $f_{\phi(j)}$ as the common vertices. Similarly, we have $y_j f_{\phi(j)} v_\ell \notin C$ for each $j \in [\ell-1]$. Now for each $j \in [\ell-1]$ apply Lemma 2.3 to complete $u_\ell \ldots u_k w_{\ell,j} y_j$ to a copy $D_j := u_\ell \ldots u_k w_{\ell,j} y_j y_{j,1} \ldots y_{j,\ell-3}$ of K_k. Observe that $D_j \in C$ for each $j \in [\ell-1]$: if not, then $y_j f_{\phi(j)} v_\ell \notin C$, $y_j f_{\phi(j)} u_\ell \in C$ and $D_j \notin C$ would yield $\dagger_k(\ell-1, \ell-2)$ with $y_j f_{\phi(j)} u_\ell$ as the 'central' copy of K_k, $y_j u_{\ell+1} \ldots u_k$ as the common vertices and D_j 'dangling off' u_ℓ. But now $D_j \in C$ for $j \in [\ell-1]$ with $u_\ell \ldots u_k w_{\ell,1} \ldots w_{\ell,\ell-1} \notin C$ as the 'central' copy of K_k yields $\dagger_k(\ell-1,1)$ with $u_\ell \ldots u_k$ as the common vertices, giving a contradiction. Hence, Y_j is empty for some $j \in [\ell-1]$.

Pick $j \in [\ell-1]$ such that $Y_j = \varnothing$, which exists by the claim above. Apply Lemma 2.2 with $U = V(G) \setminus \{u_1, \ldots, u_{\ell-1}, u_{\ell+1}, \ldots, u_k\}$ to obtain

$$|X_i| \geq (k-2)(\delta - k + 2) - (k-3)(n-k+1) = (k-2)\delta - (k-3)n - 1 \qquad (2.43)$$

for each $i \in [\ell-1]$. Since $X_h \cap X_i = \Gamma(f)$ for all $\{h, i\} \in \binom{[\ell-1]}{2}$, we have

$$|X| = \sum_{i=1}^{\ell-1} |X_i| - (\ell-2)|\Gamma(f)|. \qquad (2.44)$$

We claim that $w_{\ell,j} \notin X$. Indeed, suppose that $w_{\ell,j} \in X$. Without loss of generality, $w_{\ell,j} \in X_1$. Observe that $w_{\ell,j} f_1 u_\ell \in C$: if not, then $fu_\ell \in C$, $fv_\ell \notin C$ and $w_{\ell,j} f_1 u_\ell \notin C$ would yield $\dagger_k(2,1)$ with fu_ℓ as the 'central' copy of K_k and f_1 as the common vertices. Similarly, we have $w_{\ell,j} f_1 v_\ell \notin C$. But now $w_{\ell,j} f_1 v_\ell \notin C$, $w_{\ell,j} f_1 u_\ell \in C$ and $u_\ell \dots u_k w_{\ell,1} \dots w_{\ell,\ell-1} \notin C$ yields $\dagger_k(\ell-1, \ell-2)$ with $w_{\ell,j} f_1 u_\ell$ as the 'central' copy of K_k, $u_{\ell+1} \dots u_k w_{\ell,j}$ as the common vertices and $u_\ell \dots u_k w_{\ell,1} \dots w_{\ell,\ell-1}$ 'dangling off' u_ℓ, giving a contradiction. Now apply Lemma 2.2 with $U = X \setminus \{u_\ell, v_\ell\}$ to obtain

$$\begin{aligned} |Y_j| &\geq 2(\delta - n + |X|) + (\delta - n + |X| - 1) - 2(|X| - 2) \\ &= |X| - 3(n - \delta - 1). \end{aligned} \tag{2.45}$$

Denote by C' the K_{k+1}-component of G which contains fv_ℓ. Define

$$\begin{aligned} &W_1 := \{u \in \Gamma(f) : uf \in C\}, \quad W_2 := \{u \in \Gamma(f) : uf \in C'\} \quad \text{and} \\ &W_3 := \{u \in \Gamma(f) : uf \notin C, C'\}. \end{aligned}$$

Since $u_\ell \in W_1$ and $v_\ell \in W_2$, we have $W_1, W_2 \neq \varnothing$. Furthermore, $\Gamma(u_1, \dots, u_k) \subseteq W_1$ and $\Gamma(u_1, \dots, u_{\ell-1}, u_{\ell+1}, \dots, u_k, v_\ell) \subseteq W_2$. Let $w_1 \in W_1$ and $w_2 \in W_2$. Note that w_1 has no neighbour in $W_2 \cup W_3$ and w_2 has no neighbour in $W_1 \cup W_3$, so

$$|W_1 \cup W_3|, |W_2 \cup W_3| \leq n - \delta - 1. \tag{2.46}$$

Since Y_j is empty and (2.43), (2.44) and (2.45) hold, we obtain

$$0 = |Y_j| \geq (\ell - 1)((k-2)\delta - (k-3)n - 1) - (\ell - 2)|\Gamma(f)| - 3(n - \delta - 1).$$

By rearrangement, we obtain

$$|\Gamma(f)| = \sum_{i \in [3]} |W_i| \geq (k-2)\delta - (k-3)n - 1 + \tfrac{(k+1)\delta - kn + 2}{\ell - 2}. \tag{2.47}$$

By (2.46) and (2.47), we have

$$(k-1)\delta - (k-2)n + \tfrac{(k+1)\delta - kn + 2}{\ell - 2} \leq |W_1|, |W_2| \leq n - \delta - 1.$$

Hence, $\delta \leq \frac{[(\ell-1)(k-1)+1]n - \ell}{(\ell-1)k+1} \leq \frac{(2k-1)n-3}{2k+1} < \left(\frac{k}{k+1} - 2\eta\right) n$. By multiplying both sides of the first upper bound on δ by $\ell - 3$ and rearranging, we obtain

$$n - \delta - 1 + \tfrac{(k+1)\delta - kn + 2}{\ell - 2} \geq (\ell - 2)(k\delta - (k-1)n + 1).$$

Recalling (2.47) and $\ell \geq 3$, we obtain

$$\begin{aligned} |\Gamma(f)| &\geq (k-1)\delta - (k-2)n + (\ell-2)(k\delta - (k-1)n + 1) \\ &\geq (k-1)\delta - (k-2)n + k\delta - (k-1)n + 1 \end{aligned} \tag{2.48}$$

and

$$|W_1|, |W_2| \geq (\ell-1)[k\delta - (k-1)n + 1] \geq 2[k\delta - (k-1)n + 1]. \tag{2.49}$$

Now pick a vertex $w \in W_2$ and define

$$\begin{aligned} &Z_i := \Gamma(u_{i+2}, \ldots, u_k) \setminus \Gamma(u_{i+1}) \text{ for } \ell \leq i \leq k-1, \\ &Z_i := \Gamma(u_{i+1}, \ldots, u_{\ell-1}, u_{\ell+1} \ldots, u_k) \setminus \Gamma(u_i) \text{ for } i \in [\ell-1], \\ &Z_i' := \Gamma(u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_{\ell-1}, u_{\ell+1}, \ldots, u_k) \setminus \Gamma(u_i) \text{ for } i \in [\ell-1], \\ &Z_i' := \Gamma(u_1, \ldots, u_{\ell-1}, u_{\ell+1} \ldots, u_i, u_{i+2}, \ldots, u_k) \setminus \Gamma(u_{i+1}) \text{ for } \ell \leq i \leq k-1, \\ &Z_i'' := Z_i' \cap \Gamma(w) \text{ for } i \in [k-1], \\ &A_1 := \bigcup_{i=1}^{k-1} (Z_i \setminus Z_i'), \quad A_2 := \left(\bigcup_{i=1}^{k-1} Z_i' \right) \setminus \Gamma(w), \quad A := A_1 \cup A_2. \end{aligned}$$

Note that $|A_1|$ is the number of vertices in G with at least two non-neighbours in f. Count $\rho := \sum_{v \in V(G), u \in f} \mathbf{1}_{\{vu \notin E(G)\}}$ in two ways. On the one hand,

$$\rho = \sum_{u \in f} \left(\sum_{v \in V(G)} \mathbf{1}_{\{vu \notin E(G)\}} \right) = \sum_{u \in f} |V(G) \setminus \Gamma(u)| \leq (k-1)(n-\delta).$$

On the other hand,

$$\rho = \sum_{v \in V(G)} \left(\sum_{u \in f} \mathbf{1}_{\{vu \notin E(G)\}} \right) = \sum_{v \in V(G)} |f \setminus \Gamma(v)| \geq n - |\Gamma(f)| + |A_1|.$$

Hence, we have $|A_1| \leq |\Gamma(f)| - n + (k-1)(n-\delta)$. No vertex in $W_1 \cup W_3 \cup A_2$ is adjacent to w, so $|A_2| \leq n - \delta - 1 - |W_1| - |W_3|$. Therefore, noting that $|\Gamma(f)| = \sum_{i \in [3]} |W_i|$, we obtain

$$|A| \leq |W_2| - [k\delta - (k-1)n + 1]. \tag{2.50}$$

Lemma 2.25(iii) tells us that $\delta(G[W_1]) \geq k\delta - (k-1)n$, so by Lemma 2.1(i) and (2.49) we have a matching M of size $|M| = k\delta - (k-1)n$ in W_1. We shall check the conditions to apply Lemma 2.23 for $b = 2$ with W_1, $W_2 \cup W_3$, $Z_1'', \ldots, Z_{k-1}''$, A and M as U_1, U_2, $X_1, \ldots, X_{k-1}$, A and F respectively. By definition W_1 and $W_2 \cup W_3$ partition $\Gamma(f)$. For each $i \in [k-1]$ the set Z''_i consists of the neighbours of w whose only non-neighbour in f is u_i if $i < \ell$ and u_{i+1} if $i \geq \ell$. The set A consists of the non-neighbours of w with exactly one non-neighbour in f and the vertices with at least two non-neighbours in f. Hence, W_1, $W_2 \cup W_3$, $Z''_1, \dots, Z''_{k-1}$, A form a partition of $V(G)$ such that there are no edges between W_1, W_2 and W_3. Given $i \in [k-1]$ there exists $j \in [k]$ such that $Z''_i \subseteq V(G) \setminus \Gamma(u_j)$ so $|Z''_i| \leq n - \delta$. For each $(e, i) \in E(G[W_1]) \times [k-1]$, by applying Lemma 2.21(i) for i with $u_1, \dots, u_{\ell-1}, w$ as themselves, u_{a+1} as u_a for $\ell \leq a < k$, C as C_1, C' as C_2 and e as uv, we have that $Z''_i \cap \Gamma(e)$ is an independent set. Furthermore, all copies of K_k in G with at least two vertices from W_1 and all other vertices from $\left(\bigcup_{i=1}^{k-2} Z''_i\right)$ are K_{k+1}-connected: we can construct a K_{k+1}-walk from such a copy g of K_k to fu_ℓ by a step-by-step vertex replacement of the vertices of g with the vertices of fu_ℓ.

Since the requisite conditions are satisfied, we apply Lemma 2.23 for $b = 2$ with W_1, $W_2 \cup W_3$, $Z''_1, \dots, Z''_{k-1}$, A and M as U_1, U_2, $X_1, \dots, X_{k-1}$, A and F respectively; noting that (2.49) and (2.50) hold, we obtain that $\mathrm{CKF}_{k+1}(G)$ is at least

$$\begin{aligned}
&(k+1)\min\left\{k\delta - (k-1)n, \left\lfloor \tfrac{|W_2 \cup W_3|}{2} \right\rfloor, \tfrac{2[k\delta-(k-1)n]+|W_3|+1}{3}\right\} \\
&\geq (k+1)\min\left\{k\delta - (k-1)n, \tfrac{2[k\delta-(k-1)n]+|W_3|+1}{3}\right\}.
\end{aligned}$$

First suppose there is a vertex $u \in W_3$. Since $\Gamma(u, f) \subseteq W_3$, by Lemma 2.2 we have $|W_3| \geq |\Gamma(u, f)| \geq k\delta - (k-1)n$. This implies $\mathrm{pp}_k(n, \delta + \eta n) > \mathrm{CKF}_{k+1}(G) \geq (k+1)(k\delta - (k-1)n)$, which contradicts (2.7). Hence, we have $W_3 = \varnothing$. We distinguish three cases.

Case 1: $\delta \in \left[\left(\frac{k-1}{k} + \mu\right)n, \left(\frac{3k-2}{3k+1} - 2\eta\right)n\right] \cup \left[\left(\frac{3k-2}{3k+1} + \eta\right)n, \left(\frac{2k-1}{2k+1} - 2\eta\right)n\right]$. In this case, we have $\mathrm{pp}_k(n, \delta + \eta n) > \mathrm{CKF}_{k+1}(G) \geq \frac{2(k+1)(k\delta-(k-1)n)}{3}$, which contradicts (2.9).

Case 2: $\delta \in \left[\left(\frac{3k-2}{3k+1} - 2\eta\right)n, \left(\frac{3k-2}{3k+1} + \eta\right)n\right]$. Without loss of generality, we have $|W_1| \geq |W_2|$. By the upper bound on δ, we have $n - \delta - 1 \geq 3(k\delta - (k-1)n + 1) - (3k+1)\eta n - 2$. Now together with (2.48) we obtain $|W_1| \geq \frac{|\Gamma(f)|}{2} \geq \frac{9}{4}(k\delta - (k-1)n) + 3$ so $\left\lfloor \frac{|W_1|}{3} \right\rfloor \geq \frac{3}{4}(k\delta - (k-1)n)$. Note that $\delta(G[W_1]) \geq |W_1| - (n - \delta - |W_2|)$. By Corollary 2.5 applied to $G[W_1]$ with $k = 2$, (2.48) and (2.49), the number of vertex-disjoint triangles in $G[W_1]$ is at least

$$\begin{aligned}
&\min\left\{|\Gamma(f)| + |W_2| - 2(n-\delta), \left\lfloor \tfrac{|W_1|}{3} \right\rfloor\right\} \\
&\geq \min\left\{4(k\delta - (k-1)n) - (n-\delta), \left\lfloor \tfrac{|W_1|}{3} \right\rfloor\right\} \geq \tfrac{3}{4}(k\delta - (k-1)n).
\end{aligned}$$

Let T be a collection of $\frac{3}{4}(k\delta - (k-1)n)$ vertex-disjoint triangles in $G[W_1]$. We apply Lemma 2.23 for $b = 3$ with W_1, W_2, $Z''_1, \ldots, Z''_{k-1}$, A and T as U_1, U_2, $X_1, \ldots, X_{k-1}$, A and F respectively; the requisite conditions have been shown to be satisfied in the preceding application of Lemma 2.23. Noting (2.49), we obtain that $\mathrm{CKF}_{k+1}(G)$ is at least

$$(k+1)\min\left\{\tfrac{3}{4}(k\delta-(k-1)n), \left\lfloor\tfrac{|W_2|}{2}\right\rfloor, \tfrac{2[k\delta-(k-1)n]+|W_2|+1}{5}\right\} \geq \tfrac{3}{4}(k+1)(k\delta-(k-1)n),$$

so $\mathrm{pp}_k(n, \delta+\eta n) > \frac{3}{4}(k+1)(k\delta-(k-1)n)$, which contradicts (2.8).

Case 3: $\delta \in \left[\left(\frac{2k-1}{2k+1} - 2\eta\right)n, \frac{(2k-1)n-3}{2k+1}\right]$. Define

$$\widetilde{Z}_i = Z_i \cap \Gamma(w) \text{ for } i \in [k-1] \text{ and } \widetilde{A} := \left(\bigcup_{i=1}^{k-1} Z_i\right) \setminus \Gamma(w).$$

No vertex in $W_1 \cup \widetilde{A}$ is adjacent to w, so

$$|\widetilde{A}| \leq n - \delta - 1 - |W_1|. \tag{2.51}$$

By Lemma 2.25(iii) we have $\delta(G[W_1]) \geq k\delta - (k-1)n$, so there is a matching M of size $|M| = k\delta - (k-1)n$ in W_1 by Lemma 2.1(i) and (2.49). We shall check the conditions to apply Lemma 2.24 with W_1, W_2, $\widetilde{Z}_1, \ldots, \widetilde{Z}_{k-1}$, $\widetilde{A}$ and M as U_1, U_2, $X_1, \ldots, X_{k-1}$, A and F respectively. W_1 and W_2 partition $\Gamma(f)$ by definition. For each $i \in [k-1]$ the set $\widetilde{Z}_i$ consists of the neighbours v of w such that $\max\{j \in [k] \setminus \{\ell\} : vu_j \notin E(G)\}$ is well-defined and equal to i if $i < \ell$ and to $i+1$ if $i \geq \ell$. The set $\widetilde{A}$ consists of the non-neighbours of w with at least one non-neighbour in f. Hence, W_1, W_2, $\widetilde{Z}_1, \ldots, \widetilde{Z}_{k-1}$, $\widetilde{A}$ form a partition of $V(G)$ such that there are no edges between W_1 and W_2. Given $i \in [k-1]$ there is $j \in [k]$ such that $\widetilde{Z}_i \subseteq V(G) \setminus \Gamma(u_j)$ so $|\widetilde{Z}_i| \leq n - \delta$. Let $i \in [k-1]$ and let g be a copy of K_{i+1} comprising an edge e of $G[W_1]$ and a copy g' of K_{i-1} with a vertex from each of $\widetilde{Z}_1, \ldots, \widetilde{Z}_{i-1}$. By applying Lemma 2.21 (ii) for i with $u_1, \ldots, u_{\ell-1}, w$ as themselves, u_{a+1} as u_a for $\ell \leq a < k$, C as C_1, C' as C_2, e as uv and g' as g, we have that $\widetilde{Z}_i \cap \Gamma(g)$ is an independent set. Furthermore, all copies of K_k in G comprising an edge of $G[W_1]$ and a vertex from each of $\widetilde{Z}_1, \ldots, \widetilde{Z}_{k-2}$ are K_{k+1}-connected: we can construct a K_{k+1}-walk from such a copy g of K_k to fu_ℓ by a step-by-step vertex replacement of the vertices of g with the vertices of fu_ℓ.

Since the requisite conditions are satisfied, we apply Lemma 2.24 with W_1, W_2, $\widetilde{Z}_1, \ldots, \widetilde{Z}_{k-1}$, $\widetilde{A}$ and M as U_1, U_2, $X_1, \ldots, X_{k-1}$, A and F respectively; noting that in this case we have $\delta \geq \left(\frac{2k-1}{2k+1} - 2\eta\right) n$, (2.49) and (2.51), we obtain that $\mathrm{CKF}_{k+1}(G)$ is at least

$$\begin{aligned}
&(k+1)\min\left\{k\delta - (k-1)n, k\delta - (k-1)n + |W_2| - |W_1| - |\widetilde{A}|\right\} \\
&\geq (k+1)\min\left\{k\delta - (k-1)n, k\delta - (k-1)n + |W_2| - (n-\delta-1)\right\} \\
&\geq (k+1)\min\left\{k\delta - (k-1)n, k\delta - (k-1)n + (2k+1)\delta - (2k-1)n + 3\right\} \\
&\geq (k+1)(k\delta - (k-1)n - 2(2k+1)\eta n),
\end{aligned}$$

so $\mathrm{pp}_k(n, \delta + \eta n) > (k+1)(k\delta - (k-1)n - 2(2k+1)\eta n)$, contradicting (2.7). $\square$

Now we prove Lemma 2.27.

Proof of Lemma 2.27. Let $S = \{(j, \ell) \in \mathbb{N}^2 \mid j < \ell \leq k\}$. Note that $f : S \to \left[\frac{k(k-1)}{2}\right]$ given by $f(j, \ell) = \frac{\ell(\ell-1)}{2} - j + 1$ is bijective and $f(j, \ell) < f(j', \ell') \iff \ell < \ell'$ or $(\ell = \ell', j' < j)$. We proceed by induction on $f(j, \ell)$. By Lemma 2.28, G does not contain $\dagger_k(2, 1)$; this corresponds to the base case $f(j, \ell) = 1$. For $f(j, \ell) > 1$, there are two cases to consider: $j + 1 = \ell \leq k$ and $j + 1 < \ell \leq k$.

Consider the first case $j + 1 = \ell \leq k$. By the inductive hypothesis, G does not contain $\dagger_k(2, 1)$, $\dagger_k(\ell-1, \ell-2)$ or $\dagger_k(\ell-1, 1)$. Hence, by Lemma 2.29 G does not contain $\dagger_k(\ell, \ell-1)$. Consider the second case $j + 1 < \ell \leq k$. By the inductive hypothesis, G does not contain $\dagger_k(\ell, j+1)$, $\dagger_k(\ell, \ell-1)$ or $\dagger_k(\ell - j, 1)$. Hence, by Lemma 2.19 G does not contain $\dagger_k(\ell, j)$. This completes the proof by induction.

Finally, G does not contain $\dagger_k(k, 1)$ so $\mathrm{int}_k(G)$ is K_k-free. $\square$

It remains to handle the case where $\mathrm{int}_k(G)$ contains no copy of K_k. The following lemma represents an application of Lemma 2.24 for this case.

Lemma 2.30. *Let $k \geq 3$ be an integer. Let G be a graph on n vertices with minimum degree $\delta(G) \geq \delta > \frac{(k-1)n}{k}$, at least two K_{k+1}-components and $\mathrm{int}_k(G)$ K_k-free. Let $C_1, \ldots, C_p$ be the K_{k+1}-components of G. Set $q' := k\delta - (k-1)n + \sum_{j \neq 1} |\mathrm{ext}(C_j)| - |\mathrm{ext}(C_1)|$. Then*

$$\mathrm{CKF}_{k+1}(G) \geq (k+1)\min\left\{k\delta - (k-1)n, \left\lfloor \frac{|\mathrm{ext}(C_1)|}{2} \right\rfloor, q'\right\}.$$

Proof. By Lemma 2.26(i) we have $|\mathrm{int}_k(G)| \geq 2\delta - n + 2 > 0$. Pick $u_{k-1} \in \mathrm{int}_k(G)$ and recursively pick $u_i \in \Gamma(u_{k-1}, \ldots, u_{i+1}; \mathrm{int}_k(G))$ for $i \in [k-2]$. By Lemma 2.2 we have

$$\begin{aligned}
|\Gamma(u_{k-1}, \ldots, u_{i+1}; \mathrm{int}_k(G))| &\geq |\mathrm{int}_k(G)| - (k-i-1)(n-\delta) \\
&\geq (k-i+1)\delta - (k-i)n + 2 > 0
\end{aligned} \tag{2.52}$$

for each $i \in [k-1]$, so this is well-defined. For $i \in [k-1]$ define

$$L_i = \Gamma(u_{k-1}, \dots, u_{i+1}; \mathrm{int}_k(G)) \setminus \Gamma(u_i).$$

We want to apply Lemma 2.24 with $\mathrm{ext}(C_1), \bigcup_{j\neq 1} \mathrm{ext}(C_j), L_1, \dots, L_{k-1}$ and $\varnothing$ as $U_1, U_2, X_1, \dots, X_{k-1}$ and A respectively. We claim that $L_1, \dots, L_{k-1}$ give a partition of $\mathrm{int}_k(G)$. Indeed, each set L_i is nonempty by (2.52) and the fact that each u_i has at most $n-\delta$ non-neighbours. Furthermore, for each $v \in \mathrm{int}_k(G)$ we have $v \in L_h$ if and only if $h = \max\{a \in [k-1] : v \notin \Gamma(u_a)\}$; this quantity is well-defined because $\mathrm{int}_k(G)$ is K_k-free. Hence, $L_1, \dots, L_{k-1}, \mathrm{ext}(C_1), \bigcup_{j\neq 1} \mathrm{ext}(C_j)$ gives a partition of $V(G)$. No vertex of L_i is adjacent to u_i so $|L_i| \leq n-\delta$ for each $i \in [k-1]$ and $|\mathrm{int}_k(G)| \leq (k-1)(n-\delta)$. By Lemma 2.25(ii) there are no edges between $\bigcup_{j\neq 1} \mathrm{ext}(C_j)$ and $\mathrm{ext}(C_1)$. This means that vertices in $\mathrm{ext}(C_1)$ have neighbours in only $\mathrm{ext}(C_1)$ and $\mathrm{int}_k(G)$, so $\delta(\mathrm{ext}(C_1)) \geq \delta - |\mathrm{int}_k(G)| \geq k\delta - (k-1)n$. Hence, we have a matching M in $\mathrm{ext}(C_1)$ with $|M| = \min\left\{k\delta - (k-1)n, \left\lfloor \frac{|\mathrm{ext}(C_1)|}{2} \right\rfloor\right\}$ by Lemma 2.1(i). All copies of K_k in G containing an edge of $G[\mathrm{ext}(C_1)]$ belong to C_1, so they are all K_{k+1}-connected. Let $i \in [k-2]$ and let f be a copy of K_{i+1} comprising an edge of $G[\mathrm{ext}(C_1)]$ and a vertex from each of $L_1, \dots, L_{i-1}$. Since we have $L_1 \cup \dots \cup L_i \subseteq \Gamma(u_{k-1}, \dots, u_{i+1}; \mathrm{int}_k(G))$, an edge in $L_i \cap \Gamma(f)$ would form a copy of K_k in $\mathrm{int}_k(G)$ together with $u_{k-1}, \dots, u_{i+1}$ and the vertices of f in $L_1 \cup \dots \cup L_{i-1}$. This contradicts the assumption that $\mathrm{int}_k(G)$ is K_k-free, so $L_i \cap \Gamma(f)$ is an independent set.

Since the requisite conditions are satisfied, we apply Lemma 2.24 with $\mathrm{ext}(C_1)$, $\bigcup_{j\neq 1} \mathrm{ext}(C_j)$, $L_1, \dots, L_{k-1}$, $\varnothing$ and M as $U_1, U_2, X_1, \dots, X_{k-1}, A$ and F respectively to obtain that $\mathrm{CKF}_{k+1}(G)$ is at least $(k+1)\min\left\{k\delta - (k-1)n, \left\lfloor \frac{|\mathrm{ext}(C_1)|}{2} \right\rfloor, q'\right\}$. □

Now we aim to prove that if $\mathrm{CKF}_{k+1}(G) < \mathrm{pp}_k(n, \delta + \eta n)$ and $\mathrm{int}_k(G)$ contains no copy of K_k, then $\mathrm{int}_k(G)$ is in fact $(k-1)$-partite and its copies of K_{k-1} lie in at least $r_p(n, \delta+\eta n)$ K_{k+1}-components.

Lemma 2.31. *Let $k \geq 3$ be an integer. Let $\mu, \eta > 0$ and $n \in \mathbb{N}$ satisfy $\frac{1}{n} \ll \eta \ll \mu, \frac{1}{k}$. Let G be a graph on n vertices with at least two K_{k+1}-components and minimum degree $\delta(G) \geq \delta \geq (\frac{k-1}{k} + \mu)n$. Suppose $\mathrm{CKF}_{k+1}(G) < \mathrm{pp}_k(n, \delta + \eta n)$ and $\mathrm{int}_k(G)$ is K_k-free. Then $\mathrm{int}_k(G)$ is $(k-1)$-partite and all copies of K_{k-1} in $\mathrm{int}_k(G)$ are contained in at least $r_p(n, \delta + \eta n)$ K_{k+1}-components of G.*

Proof. Let $0 < \eta < \min\{\frac{1}{1000k^2}, \eta_0(k, \mu)\}$ and $n_1 := \max\{n_2(k, \mu, \eta), \frac{2}{\eta}\}$ with $\eta_0(k, \mu)$

and $n_2(k, \mu, \eta)$ given by Lemma 2.10. Set $r' := r_p(n, \delta + \eta n)$. Let $f := u_1 \dots u_{k-1}$ be a copy of K_{k-1} in $\mathrm{int}_k(G)$ and let $C_1, \dots, C_p$ be the K_{k+1}-components of G.

We claim that f is a copy of K_{k-1} of every K_{k+1}-component of G. Indeed, suppose f is not a copy of K_{k-1} of C_i for some $i \in [p]$. Since $|\Gamma(f)| \geq (k-1)\delta - (k-2)n$ by Lemma 2.2 and $|C_i| > \delta$ by Lemma 2.25(i), there is a vertex $w \in \Gamma(f)$ which is also a vertex of C_i. Now since $fw \notin C_i$, we have $fw \in C_j$ for some $j \neq i$ and hence w is a vertex of C_j. Since w is a vertex of both C_i and C_j, we have $w \in \mathrm{int}_k(G)$, which in turn implies that fw is a copy of K_k in $\mathrm{int}_k(G)$, contradicting our lemma hypothesis.

For $\delta \geq \left(\frac{k}{k+1} - 2\eta\right) n$, note that by Lemma 2.26(i) we have $|\mathrm{int}_k(G)| \geq 2\delta - n + 2 > \frac{3k-4}{3}(n-\delta)$, so $\delta(G[\mathrm{int}_k(G)]) \geq \delta - n + |\mathrm{int}_k(G)| > \frac{3k-7}{3k-4}|\mathrm{int}_k(G)|$. Then, Theorem 2.7 implies that $\mathrm{int}_k(G)$ is $(k-1)$-partite. Furthermore, by (2.10) and since G has at least two K_{k+1}-components, we have that all copies of K_{k-1} in $\mathrm{int}_k(G)$ are contained in at least $r' \leq 2$ K_{k+1}-components. Therefore, it remains to consider the case $\delta < \left(\frac{k}{k+1} - 2\eta\right) n$; by (2.3) we have $r' \geq 2$. For each $i \in [p]$, let U_i be the set of common neighbours v of f such that $fv \in C_i$. Since $\mathrm{int}_k(G)$ is K_k-free, we have $U_i \subseteq \mathrm{ext}(C_i)$ for each $i \in [p]$. Without loss of generality, let $\mathrm{ext}(C_1)$ be a largest K_{k+1}-component exterior of G.

Let $i \neq 1$. Applying Lemma 2.30 and noting that $|\mathrm{ext}(C_1)| \geq |\mathrm{ext}(C_i)|$, we have that $\mathrm{CKF}_{k+1}(G)$ is at least

$$(k+1)\min\left\{ \left\lfloor \frac{|\mathrm{ext}(C_i)|}{2} \right\rfloor, k\delta - (k-1)n \right\}.$$

Since $(k+1)(k\delta - (k-1)n) \geq \mathrm{pp}_k(n, \delta + \eta n)$ by (2.7), we deduce that $(k+1)\left\lfloor \frac{|\mathrm{ext}(C_i)|}{2} \right\rfloor < \mathrm{pp}_k(n, \delta + \eta n) \leq \frac{k+1}{2}\left(\frac{(k-1)(\delta+3\eta n)-(k-2)n}{r'} - 2\right)$ by (2.5). Hence, we have

$$|U_i| \leq |\mathrm{ext}(C_i)| < \frac{(k-1)(\delta + 3\eta n) - (k-2)n}{r'}. \tag{2.53}$$

By Lemma 2.25(i) we have $|\mathrm{ext}(C_i)| + |\mathrm{int}_k(G)| \geq |C_i| > \delta$, so by (2.6) we have

$$|\mathrm{int}_k(G)| > \delta - \frac{(k-1)(\delta + 3\eta n) - (k-2)n}{r'} > \frac{3k-4}{3}(n - \delta). \tag{2.54}$$

It follows that $\delta(G[\mathrm{int}_k(G)]) \geq \delta - n + |\mathrm{int}_k(G)| > \frac{3k-7}{3k-4}|\mathrm{int}_k(G)|$, so Theorem 2.7 implies that $\mathrm{int}_k(G)$ is $(k-1)$-partite. Let $I_1, \dots, I_{k-1}$ be the parts of $\mathrm{int}_k(G)$. For each $j \in [k-1]$ we have that I_j is an independent set, so $|I_j| \leq n - \delta$. Hence, we have $|I_j| = |\mathrm{int}_k(G)| - \sum_{h \neq j} |I_h| > (k-1)\delta - (k-2)n - \frac{(k-1)(\delta+3\eta n)-(k-2)n}{r'}$. Furthermore, each vertex in I_j is adjacent to all but at most $n - \delta - |I_j|$ vertices outside I_j.

It remains to show $p \geq r'$, so suppose $p < r'$. In particular, this implies $r' \geq 3$. Since (2.53) and $\sum_{i\in[p]} |\operatorname{ext}(C_i)| \geq \sum_{i\in[p]} |U_i| = |\Gamma(f)| \geq (k-1)\delta - (k-2)n$ hold, we obtain $|\operatorname{ext}(C_1)| \geq |\Gamma(f)| - \sum_{i\neq 1} |\operatorname{ext}(C_i)| > \frac{2[(k-1)\delta-(k-2)n]-3(k-1)(r'-2)\eta n}{r'}$. By Lemma 2.25(ii), there are no edges between $\operatorname{ext}(C_1)$ and $\bigcup_{i\neq 1} \operatorname{ext}(C_i)$, so every vertex in $\operatorname{ext}(C_1)$ has neighbours in $\operatorname{ext}(C_1)$ and $\operatorname{int}_k(G)$ only. Hence, we have $\delta(\operatorname{ext}(C_1)) \geq \delta - |\operatorname{int}_k(G)|$. By Lemma 2.1(i), there is a matching F_0 in $\operatorname{ext}(C_1)$ with

$$|F_0| = \min\left\{\delta - |\operatorname{int}_k(G)|, \frac{[(k-1)\delta - (k-2)n] - 3(k-1)(r'-1)\eta n}{r'}\right\}.$$

We now build up our desired connected K_{k+1}-factor step-by-step, starting from the aforementioned matching F_0 in $\operatorname{ext}(C_1)$. We have steps $j = 1, \dots, k-1$. In step j, we extend the K_{j+1}-factor F_{j-1} to a K_{j+2}-factor F_j using I_j. We greedily match vertices of I_j with distinct copies of K_{j+1} of F_{j-1} to form copies of K_{j+2}. We find that $|I_j| > |F_0| \geq |F_{j-1}|$, so we stop only when we encounter a vertex $x \in I_j$ which is not a common neighbour of any remaining copy of K_{j+1} of F_{j-1}. Since at most $n - \delta - |I_j|$ copies of K_{j+1} in F_{j-1} do not have x as a common neighbour, we obtain a K_{j+2}-factor F_j with at least $|F_{j-1}| - (n-\delta) + |I_j|$ copies of K_{j+2}.

We terminate after step $k-1$ with a collection F_{k-1} of at least $|F_0| - (k-1)(n-\delta) + |\operatorname{int}_k(G)|$ vertex-disjoint copies of K_{k+1} in G. Since each copy of K_{k+1} in F_{k-1} uses an edge of $F_0 \subseteq G[\operatorname{ext}(C_1)]$ and (2.54) holds, we deduce that F_{k-1} is in fact a connected K_{k+1}-factor of size at least $(k+1)(k\delta - (k-1)n - 3(k-1)\eta n)$. By (2.7), this means that $\mathrm{CKF}_{k+1}(G) \geq \mathrm{pp}_k(n, \delta + \eta n)$, which is a contradiction. This completes the proof. □

We prove in the following lemma that a graph which has very high minimum degree and is not near-extremal in fact contains a large connected K_{k+1}-factor. We handle this case separately as it turns out that our greedy-type methods in Section 2.4.2 are inadequate. To overcome this, we employ a Hall-type argument (see Lemma 2.1(ii)) to extend our large matching to a sufficiently large connected K_{k+1}-factor.

Lemma 2.32. *Let $k \geq 2$ be an integer. Let $\mu, \eta > 0$ and $n \in \mathbb{N}$ satisfy $\frac{1}{n} \ll \eta \ll \mu, \frac{1}{k}$. Let G be a graph on n vertices with minimum degree $\delta = \delta(G) \geq \left(\frac{2k-1}{2k+1} - 2\eta\right)n$, exactly two K_{k+1}-components, $\operatorname{int}_k(G)$ $(k-1)$-partite and either $|\operatorname{int}_k(G)| < (k-1)(n-\delta) - 5k\eta n$ or the larger K_{k+1}-component exterior X satisfies $|X| > \frac{19}{10}(k\delta - (k-1)n)$. Then $\mathrm{CKF}_{k+1}(G) \geq \mathrm{pp}_k(n, \delta + \eta n)$.*

Proof. Let $0 < \eta < \min\{\frac{1}{1000k^2}, \eta_0(k,\mu), \frac{k\mu^2}{k+1}\}$ and $n_1 := \max\{n_2(k,\mu,\eta), \frac{2}{\eta}\}$ with $\eta_0(k,\mu)$ and $n_2(k,\mu,\eta)$ given by Lemma 2.10. Let C_1 and C_2 be the two K_{k+1}-components of G. There is a partition of $V(G)$ into three vertex classes $\mathrm{int}_k(G), \mathrm{ext}(C_1)$ and $\mathrm{ext}(C_2)$; $\mathrm{int}_k(G)$ is further partitioned into $k-1$ independent sets $I_1, \dots, I_{k-2}$ and I_{k-1}. Without loss of generality, suppose $|\mathrm{ext}(C_1)| \geq |\mathrm{ext}(C_2)|$. Since I_i is an independent set, we have

$$|I_i| \leq n - \delta \quad \text{for each } i \in [k-1]. \tag{2.55}$$

If $\delta \geq \left(\frac{k}{k+1} - 2\eta\right) n$, then by Lemma 2.26(i) we have $|\mathrm{int}_k(G)| \geq 2\delta - n + 2 \geq (k-1)(n-\delta) - 2(k+1)\eta n$ and by Lemma 2.26(ii) we have $|\mathrm{ext}(C_1)| \leq n - \delta - 1 \leq k\delta - (k-1)n + 2(k+1)\eta n \leq \frac{19}{10}(k\delta - (k-1)n)$, which contradicts the lemma hypothesis. Therefore, we have $\delta < \left(\frac{k}{k+1} - 2\eta\right) n$. In particular, this means that $r' := r_p^{(k)}(n, \delta + \eta n) \geq 2$.

By (2.55) we have $|\mathrm{int}_k(G)| \leq (k-1)(n-\delta)$. By Lemma 2.25(i) we have $|C_1| > \delta$, so $|\mathrm{ext}(C_1)| > \delta - (k-1)(n-\delta) = k\delta - (k-1)n \geq 0$. By Lemma 2.25(ii), there are no edges between $\mathrm{ext}(C_1)$ and $\mathrm{ext}(C_2)$, so every vertex in $\mathrm{ext}(C_1)$ has neighbours in $\mathrm{ext}(C_1)$ and $\mathrm{int}_k(G)$ only. Hence, we have $\delta(\mathrm{ext}(C_1)) \geq \delta - |\mathrm{int}_k(G)| \geq \delta - (k-1)(n-\delta) = k\delta - (k-1)n$. Therefore, we can conclude by Lemma 2.1(i) that there is matching F_0 in $\mathrm{ext}(C_1)$ of size $|F_0| = \min\{k\delta - (k-1)n, \left\lfloor \frac{|\mathrm{ext}(C_1)|}{2} \right\rfloor\}$.

We build up the desired connected K_{k+1}-factor step-by-step, starting from the aforementioned matching F_0. We have steps $j = 1, \dots, k-1$. In step j we extend the K_{j+1}-factor F_{j-1} to a K_{j+2}-factor F_j using I_j. By Lemma 2.25(ii), there are no edges between $\mathrm{ext}(C_1)$ and $\mathrm{ext}(C_2)$, so every vertex in $\mathrm{ext}(C_1)$ has at least $\delta - |\mathrm{ext}(C_1)| - \sum_{h \neq j} |I_h|$ neighbours in I_j. For each $i \in [j-1]$, since I_i is an independent set, every vertex of I_i has at least $\delta - (n - |I_j| - |I_i|)$ neighbours in I_j. Therefore, by Lemma 2.2 every copy of K_{j+1} in F_{j-1} has at least

$$\begin{aligned} a_j &:= 2\left(\delta - |\mathrm{ext}(C_1)| - \sum_{h \neq j} |I_h|\right) + \sum_{i=1}^{j-1} (\delta - n + |I_j| + |I_i|) - j|I_j| \\ &= (j+1)\delta - (j-1)n - 2|\mathrm{ext}(C_1)| - \sum_{i=1}^{k-1} |I_i| - \sum_{i=j+1}^{k-1} |I_i| \end{aligned}$$

common neighbours in I_j. At the same time, since I_j is an independent set, every vertex of I_j has at least $\delta - (n - |\mathrm{ext}(C_1)| - |I_1| - \cdots - |I_j|)$ neighbours in $\mathrm{ext}(C_1) \cup I_1 \cup \cdots \cup I_j$, of which all but at most $|\operatorname{ext}(C_1)| + |I_1| + \cdots + |I_j| - (j+1)|F_{j-1}|$ are in F_{j-1}. Hence, every vertex in I_j has at least

$$\begin{aligned} b_j &:= \delta - (n - |\operatorname{ext}(C_1)| + |I_1| + \cdots + |I_j|) \\ &\qquad - (|\operatorname{ext}(C_1)| + |I_1| + \cdots + |I_{j-1}| - (j+1)|F_{j-1}|) - j|F_{j-1}| \\ &= \delta - n + |I_j| + |F_{j-1}| \end{aligned}$$

copies of K_{j+1} of F_{j-1} in its neighbourhood. Form an auxiliary bipartite graph with vertex set $F_{j-1} \cup I_j$, where $f \in F_{j-1}$ is adjacent to $u \in I_j$ if and only if fu is a copy of K_{j+2} in G. By Lemma 2.1(ii), there is a matching in the auxiliary bipartite graph with at least $\min\{a_j + b_j, |F_{j-1}|, |I_j|\}$ edges, which corresponds to a collection F_j of

$$|F_j| = \min\{a_j + b_j, |F_{j-1}|, |I_j|\} \tag{2.56}$$

vertex-disjoint copies of K_{j+2} in G. Lemma 2.26(i) tells us $|\operatorname{int}_k(G)| \geq 2\delta - n + 2$, so by (2.55) we have

$$|I_j| = |\operatorname{int}_k(G)| - \sum_{h \neq j} |I_h| > k\delta - (k-1)n \geq |F_0| \geq |F_{j-1}|. \tag{2.57}$$

Observe that by (2.55) we have

$$\begin{aligned} a_j + b_j &= (j+2)\delta - jn - 2|\operatorname{ext}(C_1)| - \sum_{i=j+1}^{k-1} |I_i| - \sum_{i \neq j} |I_i| + |F_{j-1}| \\ &\geq (2k-1)\delta - (2k-3)n - 2|\operatorname{ext}(C_1)| + |F_{j-1}|. \end{aligned} \tag{2.58}$$

Since by Lemma 2.26(ii) we have $|\operatorname{ext}(C_1)| \leq n - \delta - 1$ and recalling our assumption that $\delta \geq \left(\frac{2k-1}{2k+1} - 2\eta\right) n$, by (2.58) we have

$$a_j + b_j \geq |F_{j-1}| - 2(2k+1)\eta n. \tag{2.59}$$

Furthermore, by (2.58) and $\delta \geq \left(\frac{2k-1}{2k+1} - 2\eta\right) n$ we obtain

$$\text{if } |\operatorname{ext}(C_1)| \leq \left(\tfrac{2}{2k+1} - (2k-1)\eta\right) n, \text{ then } a_j + b_j \geq |F_{j-1}| \text{ for all } j. \tag{2.60}$$

All copies of K_k in G containing an edge of $G[\operatorname{ext}(C_1)]$ belong to C_1, so they are all K_{k+1}-connected. Therefore, F_{k-1} is a connected K_{k+1}-factor.

It remains to check that $(k+1)|F_{k-1}| \geq \operatorname{pp}_k(n, \delta + \eta n)$. We first consider when $|F_0| = k\delta - (k-1)n$. In this case, noting (2.56), (2.57) and (2.59) we have that

$|F_j| \geq |F_{j-1}| - 2(2k+1)\eta n$ for each $j \in [k-1]$, so F_{k-1} is a connected K_{k+1}-factor in G of size at least $(k+1)(k\delta - (k-1)n - 2(k-1)(2k+1)\eta n) \geq \mathrm{pp}_k(n, \delta + \eta n)$ by (2.7). Now consider when $|F_0| = \left\lfloor \frac{|\operatorname{ext}(C_1)|}{2} \right\rfloor$. We distinguish two cases.

Case 1: $a_j + b_j \geq |F_{j-1}|$ for each $j \in [k-1]$. In this case, F_{k-1} is a connected K_{k+1}-factor in G of size $(k+1)|F_0| = (k+1)\left\lfloor \frac{|\operatorname{ext}(C_1)|}{2} \right\rfloor$. Suppose that this is less than $\mathrm{pp}_k(n, \delta + \eta n)$. By (2.5) and $\delta \geq \left(\frac{2k-1}{2k+1} - 2\eta\right)n$, we have $|\operatorname{ext}(C_1)| < \frac{(k-1)(\delta+3\eta n)-(k-2)n}{2} \leq \frac{19}{10}(k\delta - (k-1)n)$. Furthermore, $|\operatorname{int}_k(G)| \geq n - 2|\operatorname{ext}(C_1)| > (k-1)(n-\delta) - 3(k-1)\eta n$. This contradicts the lemma hypothesis.

Case 2: $a_j + b_j < |F_{j-1}|$ for some $j \in [k-1]$. By (2.60), this means that $|\operatorname{ext}(C_1)| > \left(\frac{2}{2k+1} - (2k-1)\eta\right)n \geq 2(k\delta - (k-1)n) - (2k-1)\eta n$. By (2.56), (2.57) and (2.59) we have $|F_j| \geq |F_{j-1}| - 2(2k+1)\eta n$ for each $j \in [k-1]$, so F_{k-1} is a connected K_{k+1}-factor in G of size at least

$$(k+1)(|F_0| - 2(k-1)(2k+1)\eta n) \geq (k+1)(k\delta - (k-1)n - 6k^2\eta n).$$

By (2.7) this is at least $\mathrm{pp}_k(n, \delta + \eta n)$. □

Finally, we prove Lemma 2.17.

Proof of Lemma 2.17. Given an integer $k \geq 3$, $\mu > 0$ and any

$$0 < \eta < \min\{\frac{1}{1000k^2}, \eta_0(k,\mu), \frac{k\mu^2}{k+1}\},$$

let $m_1 := \max\{n_2(k,\mu,\eta), \frac{2}{\eta}, k(k+1)\}$ with $\eta_0(k,\mu)$ and $n_2(k,\mu,\eta)$ given by Lemma 2.10. Let $\delta \geq (\frac{k-1}{k} + \mu)n$. Let G be a graph on $n \geq m_1$ vertices with minimum degree $\delta(G) \geq \delta$ and at least two K_{k+1}-components, each of which contains a copy of K_{k+2}. Let $C_1, \dots, C_\ell$ be the K_{k+1}-components of G. Set $\alpha := |\operatorname{int}_k(G)|$.

Lemma 2.26(i) tells us $\operatorname{int}_k(G) \neq \varnothing$ and $|\operatorname{int}_k(G)| > 2\delta - n > (k-2)(n-\delta)$, so $\delta(G[\operatorname{int}_k(G)]) \geq \delta - n + |\operatorname{int}_k(G)| > \frac{k-3}{k-2}|\operatorname{int}_k(G)|$. Hence, any vertex in $\operatorname{int}_k(G)$ can be extended to a copy of K_{k-1} in $\operatorname{int}_k(G)$ by Lemma 2.3. In particular, $\operatorname{int}_k(G)$ contains a copy of K_{k-1}.

Suppose that (D1) does not hold. Lemma 2.27 tells us that $G[\operatorname{int}_k(G)]$ is K_k-free, so Lemma 2.31 implies that $\operatorname{int}_k(G)$ is $(k-1)$-partite and all copies of K_{k-1} in $G[\operatorname{int}_k(G)]$ (of which there is at least one) are contained in at least $r' := r_p(n, \delta + \eta n)$ K_{k+1}-components. Hence, G has at least r' K_{k+1}-components. Since $\operatorname{int}_k(G)$ is $(k-1)$-partite, we have $\alpha \leq (k-1)(n-\delta)$. Lemma 2.25(i) tells us that $|C_i| > \delta$, so

$$|\operatorname{ext}(C_i)| \geq \delta - \alpha + 1 \geq k\delta - (k-1)n + 1. \tag{2.61}$$

for each $i \in [\ell]$. In particular, every K_{k+1}-component has a non-empty exterior. Pick $x \in \mathrm{ext}(C_2)$. It has at least δ neighbours, none of which are in $\mathrm{ext}(C_1) \cup \mathrm{ext}(C_3) \cup \cdots \cup \mathrm{ext}(C_\ell)$ by Lemma 2.25(ii). Observe that

$$n = |\,\mathrm{int}_k(G)| + |\,\mathrm{ext}(C_1)| + \cdots + |\,\mathrm{ext}(C_\ell)| \quad \text{and} \tag{2.62}$$

$$n \geq 1 + \delta + |\,\mathrm{ext}(C_1)| + |\,\mathrm{ext}(C_3)| + \cdots + |\,\mathrm{ext}(C_\ell)|. \tag{2.63}$$

Without loss of generality, suppose $\mathrm{ext}(C_1)$ is a largest K_{k+1}-component exterior. By Lemma 2.25(ii), there are no edges between any pair of K_{k+1}-component exteriors. Note that for any K_{k+1}-component C, all copies of K_k in G containing at least one vertex of $\mathrm{ext}(C)$ are in C and are therefore K_{k+1}-connected in G. Hence, it is enough to prove that

$$\alpha \geq (k-1)(n-\delta) - 5k\eta n \text{ and } |\,\mathrm{ext}(C_1)| \leq \frac{19}{10}(k\delta - (k-1)n),$$

as this would imply that (D2) holds. Suppose this is not the case.

Claim 2.33. *G has exactly r' K_{k+1}-components.*

Proof. Suppose that $\ell \geq r'+1$. By (2.62) we have $(r'+1)(\delta - \alpha) + \alpha < n$. We consider two cases.

Case 1: $\alpha < (k-1)(n-\delta) - 5k\eta n$. Then we have

$$\begin{aligned}(r'+1)\delta &< n + r'\alpha < n + r'((k-1)(n-\delta) - 5k\eta n)\\ &= [(k-1)r'+1]n - (k-1)r'\delta - (kr'+1)\eta n - (4kr'-1)\eta n,\end{aligned}$$

which we rearrange to obtain

$$\delta + \eta n < \frac{[(k-1)r'+1]n - (4kr'-1)\eta n}{kr'+1}.$$

Comparing this with (2.2) applied to $r' := r_p(n, \delta + \eta n)$, we deduce $r' > (4kr'-1)\eta n \geq 4kr' - 1 \geq r'$, which is a contradiction.

Case 2: $|\,\mathrm{ext}(C_1)| > \frac{19}{10}(k\delta - (k-1)n)$. By (2.61) and (2.63), we have

$$1 + \delta + \frac{19}{10}(k\delta - (k-1)n) + (r'-1)[k\delta - (k-1)n + 1] \leq n,$$

which we simplify to

$$\frac{9}{10}(k\delta - (k-1)n) + r'[k\delta - (k-1)n] < n - \delta.$$

Since by (2.1) we have $r' \geq \frac{n-\delta-\eta n}{k\delta-(k-1)n+k\eta n+1}$, we deduce that

$$\frac{9}{10}(k\delta-(k-1)n)+\frac{n-\delta-\eta n}{k\delta-(k-1)n+k\eta n+1}[k\delta-(k-1)n] < n-\delta.$$

Since $\eta < \frac{k\mu^2}{k+1}$ and $k\delta-(k-1)n \geq k\mu n$, we have

$$\begin{aligned}(k\delta-(k-1)n+k\eta n+1)(1-\mu) &< k\delta-(k-1)n+(k+1)\eta n-\mu(k\delta-(k-1)n)\\ &\leq k\delta-(k-1)n+(k+1)\eta n-k\mu^2 n\\ &< k\delta-(k-1)n,\end{aligned}$$

so applying this to the previous inequality, we obtain

$$\frac{9}{10}k\mu n+(n-\delta-\eta n)(1-\mu) < n-\delta.$$

However, since $\eta < \mu$ and $n-\delta < \frac{n}{k}$, this is a contradiction. Therefore, G has exactly r' K_{k+1}-components. ■

In particular, this means that $r' \geq 2$. For $r' = 2$, Lemma 2.32 gives a contradiction, so it remains to consider the case $r' \geq 3$. First suppose $|\operatorname{ext}(C_1)| \leq \sum_{h\neq 1}|\operatorname{ext}(C_h)|$. By Lemma 2.30, we have

$$\mathrm{CKF}_{k+1}(G) \geq (k+1)\min\left\{\left\lfloor\frac{|\operatorname{ext}(C_1)|}{2}\right\rfloor, k\delta-(k-1)n\right\}.$$

We have $\mathrm{CKF}_{k+1}(G) < \mathrm{pp}_k(n,\delta+\eta n)$ by assumption, so by (2.7) we have

$$(k+1)\left\lfloor\frac{|\operatorname{ext}(C_1)|}{2}\right\rfloor < \mathrm{pp}_k(n,\delta+\eta n).$$

Hence, by (2.5) and (2.7) we obtain $|\operatorname{ext}(C_1)| < \frac{(k-1)(\delta+3\eta n)-(k-2)n}{r'}$ and $|\operatorname{ext}(C_1)| \leq \frac{19}{10}(k\delta-(k-1)n)$. Then, by (2.62) we have $|\operatorname{int}_k(G)| = n-\sum_{i\in[r']}|\operatorname{ext}(C_i)| > (k-1)(n-\delta)-3(k-1)\eta n$. This is contradicts our earlier supposition, so we have that

$$|\operatorname{ext}(C_1)| > \sum_{h\neq 1}|\operatorname{ext}(C_h)|. \tag{2.64}$$

Set $r := r_p(n,\delta)$. By (2.61) and (2.63), we have that

$$1+\delta+(r'-1)(k\delta-(k-1)n+1)+(r'-2)(k\delta-(k-1)n+1) \leq n.$$

Rearranging and applying (2.1), we obtain

$$2r'-3 \leq \frac{n-\delta-1}{k\delta-(k-1)n+1} < r \leq r'+1,$$

which gives $r = r' + 1 = 4$. In particular, by (2.2) we have $\delta \geq \left(\frac{3k-2}{3k+1} - 2\eta\right) n$. By (2.61) and (2.64), we have $|\operatorname{ext}(C_1)| > |\operatorname{ext}(C_2)| + |\operatorname{ext}(C_3)| \geq 2(k\delta - (k-1)n + 1)$. By (2.63) and the fact that $\delta \geq \left(\frac{3k-2}{3k+1} - 2\eta\right) n$, we have $|\operatorname{ext}(C_1)| < n - \delta - |\operatorname{ext}(C_3)| \leq 2[k\delta - (k-1)n] + 2(3k+1)\eta n$. Finally, by Lemma 2.30 and (2.7), we have

$$\mathrm{CKF}_{k+1}(G) \geq (k+1)\left(k\delta - (k-1)n - 2(3k+1)\eta n\right) \geq \mathrm{pp}_k(n, \delta + \eta n),$$

which is a contradiction. This completes the proof of Lemma 2.17. $\square$

2.6 Near-extremal Graphs

In this section we will provide our proof of Lemma 2.14. To this end, we start with two useful lemmas. The first lemma will be used to construct kth powers of paths and cycles from simple paths and cycles through repeated application.

Lemma 2.34. *Given $h \in \mathbb{N}$, $h' \in [h+1]$ and a graph G, let $T = t_1 \dots t_{(h+1)\ell+h'-1}$ be the hth power of a path in G and let W be a set of vertices disjoint from T. Let $Q_1 := t_1 \dots t_{h+1}$, $Q_i := t_{(h+1)(i-2)+1} \dots t_{(h+1)i}$ for each $1 < i \leq \ell$ and $Q_{\ell+1} := t_{(h+1)\ell-h} \dots t_{(h+1)\ell+h'-1}$. Suppose that there exists a permutation σ of $[\ell+1]$ such that for each $i \in [\ell+1]$ the vertices of $Q_{\sigma(i)}$ have at least i common neighbours in W. Then there is the $(h+1)$st power of a path*

$$(q_1 t_1 \dots t_{h+1}) \dots (q_\ell t_{(h+1)\ell-h} \dots t_{(h+1)\ell})(q_{\ell+1} t_{(h+1)\ell+1} \dots t_{(h+1)\ell+h'-1})$$

in G, with $q_i \in W$ for each $i \in [\ell+1]$, using every vertex of T. If T is a cycle on $(h+1)\ell$ vertices we let instead $Q_1 := t_{(h+1)\ell-h} \dots t_{(h+1)\ell} t_1 \dots t_{h+1}$, $Q_i := t_{(h+1)(i-2)+1} \dots t_{(h+1)i}$ for each $1 < i \leq \ell$ and σ be a permutation on $[\ell]$. Then, under the same conditions, we have the $(h+1)$st power of a cycle $C^{h+1}_{(h+2)\ell}$.

Proof. Choose for each i in succession $q_{\sigma(i)}$ to be any so far unused common neighbour of $Q_{\sigma(i)}$; the lemma hypothesis ensures that this is always possible. $\square$

The second lemma allows us to construct paths and cycles of desired lengths which keep certain 'bad' vertices far apart. We apply Theorem 2.8 in its proof.

Lemma 2.35. *Let H be a graph on $h \geq 10$ vertices and $B \subseteq V(H)$ be of size at most $\frac{h}{12}$. Suppose that every vertex in B has at least $3|B|+1$ neighbours in H, and every* *vertex outside B has at least $\frac{h}{2}+2|B|+2$ neighbours in H. Then for any $3 \leq \ell \leq h$ we can find a cycle C_ℓ of length ℓ in H on which no four consecutive vertices contain more than one vertex of B. Furthermore, if x and y are any two vertices not in B and $5 \leq \ell \leq h$, we can find an ℓ-vertex path P_ℓ whose end-vertices are x and y and on which no four consecutive vertices contain more than one vertex of $B \cup \{x,y\}$.*

Proof. If we seek a path in H from x to y and $xy \notin E(H)$, add xy as a 'dummy' edge. If we seek a cycle, let xy be any edge of H such that $x, y \notin B$. Hence, it suffices to show for each $3 \leq \ell \leq h$ and each edge $xy \in V(H)$ with $x, y \notin B$ that we can find a cycle C_ℓ of length ℓ with xy as an edge, on which no four consecutive vertices contain more than one vertex of B and on which any four consecutive vertices including a vertex of B contain neither x nor y.

Let $H_1 := H[V(H) \setminus B]$. Since H_1 is a graph on $h-|B| \geq 4$ vertices with minimum degree $\delta(H_1) \geq \frac{h}{2}+|B|+2 \geq \frac{h-|B|}{2}+1$, by Theorem 2.8 H_1 is panconnected. Hence, H_1 has paths between x and y of every number of vertices from 3 to $h-|B|$. By adding the edge xy to these paths, we obtain cycles of every length from 3 to $h-|B|$ with the desired properties.

To find the required cycles of length greater than $h-|B|$, we first construct a path P in H covering B with x as an end-vertex and xy as an edge. Let $B = \{b_1, \ldots, b_{|B|}\}$ and set $B' := B \cup \{x, y\}$. For each $i \in [|B|]$ choose distinct vertices $u_{i+1}, v_i \in V(H) \setminus B'$ adjacent to b_i. Every vertex in B has at least $3|B|+1$ neighbours in H, so we may pick these vertices to be distinct for all $i \in [|B|]$. Choose a different vertex $u_1 \in V(H) \setminus B'$ adjacent to y. We can do so as y has at least $\frac{h}{2}+2|B|+2$ neighbours in H and $h \geq 12|B|$. Let $i \in [|B|]$. Both u_i and v_i have $\frac{h}{2}+2|B|+2$ neighbours in H, so they have at least $4|B|+4$ common neighbours. At most $3|B|+3$ of these are in $B \cup \{x, y, u_1, \ldots, u_{|B|+1}, v_1, \ldots, v_{|B|}\}$, so we can find a thus far unused vertex w_i adjacent to u_i and v_i. We may pick the vertices $w_1, \ldots, w_{|B|}$ greedily as we require only $|B|$ vertices. Hence, we obtain a path

$$P = xyu_1w_1v_1b_1u_2w_2v_2b_2 \ldots v_{|B|}b_{|B|}u_{|B|+1}$$

on $4|B|+3$ vertices. Observe that any cycle containing P of length at least $4|B|+5$ has the desired properties.

Let $H_2 := H[V(H) \setminus (V(P) \setminus \{x, u_{|B|+1}\})]$. Since H_2 is a graph on $h-4|B|-1 \geq 4$ vertices with minimum degree $\delta(H_2) \geq \frac{h}{2}-2|B|+1 \geq \frac{h-4|B|-1}{2}+1$, by Theorem 2.8

H_2 is panconnected. Hence, H_2 has paths between x and $u_{|B|+1}$ of every number of vertices from 4 to $h - 4|B| - 1$. By adding the path P to these paths, we obtain cycles in H of every length from $4|B| + 5$ to h with the desired properties. □

Before providing the proof of Lemma 2.14 we first give an outline of our method. Recall that the Lemma is given a Szemerédi partition with a 'near-extremal' structure. We shall show that the underlying graph either also has a 'near-extremal' structure, or possesses features which lead to longer kth powers of paths and cycles than required for the conclusion of the Lemma. The complication we encounter is the insensitivity of the Szemerédi partition to the misassignment of sublinearly many vertices and the editing of subquadratically many edges.

Recall that the sets I_i are subsets of $V(R)$ and the elements of each set I_i correspond to clusters in $V(G)$. We denote by $\bigcup I_i$ the union of the elements of the set I_i as clusters in $V(G)$. We begin by collecting in a set W_i those vertices with 'few' neighbours in $\bigcup I_i$. We then show that if there are two vertex-disjoint edges in W_i, then the sets $\bigcup B_1$ and $\bigcup B_2$ 'belong' to the same K_{k+1}-component of G. We shall show that this enables us to construct very long kth powers of paths and cycles by applying Lemma 2.12.

It remains to consider when each W_i does not contain two vertex-disjoint edges – here each W_i is almost independent with 'near-extremal' size. The set $W = \bigcup_{i=1}^{k-1} W_i$ now resembles a 'near-extremal' interior and the minimum degree condition on G guarantees that almost every edge from W to $V(G) \setminus W$ is present. At this point, we would like to say that we can find a long path outside W with sufficiently nice properties (which we need because the bipartite graph $G[W, V(G) \setminus W]$ is unfortunately not actually complete) so that we can repeatedly apply Lemma 2.34 to extend it to the kth power of a path (and similarly for powers of cycles) using vertices from W. The purpose of Lemma 2.35 is precisely to provide paths and cycles with such nice properties. The rest of the proof then focuses on establishing the right conditions for the application of Lemma 2.35 and working out the details of the various applications of Lemma 2.34.

Proof of Lemma 2.14. Given an integer $k \geq 3$ and $0 < \nu < 1$ let $\eta > 0$ and $d > 0$ satisfy

$$\eta \leq \frac{\nu^4}{(k+1)^{13}10^8} \quad \text{and} \quad d \leq \frac{\nu^4}{(k+1)^{13}10^8}. \tag{2.65}$$

Given $k \geq 3$ and $d > 0$, Lemma 2.12 returns a constant $\varepsilon_{EL} > 0$. Set

$$\varepsilon_0 := \min\left\{\varepsilon_{EL}, \frac{\nu^4}{(k+1)^{13}10^8}\right\}. \tag{2.66}$$

Given $m_{EL} \in \mathbb{N}$ and $0 < \varepsilon < \varepsilon_0$, Lemma 2.12 returns a constant $n_{EL} \in \mathbb{N}$. Given $t = k$ and $\rho = \varepsilon^{1/2}$, Theorem 2.9 returns a constant $n_{ES} \in \mathbb{N}$. Set

$$N := \max\left\{n_{EL}, \nu^{-1}n_{ES}, 100m_{EL}^{k+2}, 100(k+1)\eta^{-1}\nu^{-1}\right\}. \tag{2.67}$$

Let $\delta \in \left[\left(\frac{k-1}{k}+\nu\right)n, \frac{kn}{k+1}\right)$ and let the graphs G and R and the partition $V(R) = \left(\bigcup_{i=1}^{k-1} I_i\right) \cup \left(\bigcup_{j=1}^{\ell} B_j\right)$ satisfy conditions (i)–(iii) of the lemma.

Note that Corollary 2.6 settles the specific case of finding P_n^k in G when $\delta \geq \frac{kn-1}{k+1}$. Therefore, by (2.11) and (2.12) in what follows it is sufficient to find

$$k\text{th powers of cycles and paths of all lengths up to } \frac{11n}{20}. \tag{2.68}$$

R is an (ε, d)-reduced graph of G, so

$$\delta(R) \geq \delta' := \left(\frac{\delta}{n} - d - 2\varepsilon\right)m. \tag{2.69}$$

Moreover, by (ii) for each $i \in [k-1]$ clusters in I_i have δ' neighbours outside I_i in R, so

$$|I_i| \leq m - \delta' = \left(1 - \frac{\delta}{n} + d + 2\varepsilon\right)m. \tag{2.70}$$

Set $I_J := \bigcup_{i\in J} I_i$ for each $J \subseteq [k-1]$. By (iii) each cluster $C \in B_j$ has neighbours only in $B_j \cup I_{[k-1]}$ in R, so by (2.70) we have $\delta' \leq \deg(C) = \deg\left(C, B_j \cup I_{[k-1]}\right) \leq \deg(C, B_j) + \left|I_{[k-1]}\right| \leq \deg(C, B_j) + (k-1)(m-\delta')$. Then, by (2.69) we have

$$|B_j| > \deg(C, B_j) \geq k\delta' - (k-1)m \geq \frac{m}{n}(k\delta - (k-1)n - k(d+2\varepsilon)n).$$

Since $\delta \geq \left(\frac{k-1}{k}+\nu\right)n$, we have $k\delta - (k-1)n \geq k\nu n$; then by (2.65) and (2.66) we obtain

$$|B_j| \geq \frac{38(k\delta - (k-1)n)m}{39n} \geq \frac{38k\nu m}{39}. \tag{2.71}$$

Set $\xi := \sqrt[4]{d + \varepsilon + 6k\eta}$. By (2.65) and (2.66), we have

$$\xi \leq \frac{\nu}{50(k+1)^3}. \tag{2.72}$$

For each $i \in [k-1]$ define W_i to be the set of vertices of G with no more than ξn neighbours in $\bigcup I_i$. Since $\xi > d + \varepsilon$, the independence of I_i and the definition of an

(ε, d)-regular partition imply that $\bigcup I_i \subseteq W_i$. Set $W_J := \bigcup_{i\in J} W_i$ and $I_J^* := \bigcup_{i\in J} (\bigcup I_i)$ for each $J \subseteq [k-1]$. Note that by (2.70) and condition (ii) we have

$$\begin{aligned} |I_J| &\geq \left|I_{[k-1]}\right| - (k-1-|J|)(m-\delta') \\ &\geq \frac{m}{n}|J|(n-\delta) - 5k\eta m - (k-1-|J|)(d+2\varepsilon)m \end{aligned} \tag{2.73}$$

for each $J \subseteq [k-1]$. Hence, we have

$$|I_J^*| \geq \frac{(1-\varepsilon)n}{m}|I_J| \geq |J|(n-\delta) - 5k\eta n - (k-|J|-1)(d+2\varepsilon)n - \varepsilon n \tag{2.74}$$

for each $J \subseteq [k-1]$.

The claim below states that if there are two vertex-disjoint edges in some W_i, then we have two vertex-disjoint copies of K_k on $W_{[k-1]}$.

Claim 2.36. *Suppose that for some $i \in [k-1]$ there are two vertex-disjoint edges in W_i. Then there are two vertex-disjoint copies of K_k in $W_{[k-1]}$ each comprising two vertices of W_i and a vertex of W_h for each $h \in [k-1] \setminus \{i\}$.*

Proof. We consider the $i = 1$ case and note that an analogous argument applies for each $i \neq 1$. We prove the following statement for all $2 \leq j \leq k$ by backwards induction on j. If there are two vertex-disjoint copies of K_j on $W_{[j-1]}$ each comprising two vertices of W_1 and a vertex of W_h for each $1 < h < j$, then there are two vertex-disjoint copies of K_k on $W_{[k-1]}$ each comprising two vertices of W_1 and a vertex of W_h for each $1 < h < k$. Setting $j = 2$ then gives our desired statement for the $i = 1$ case.

The statement is trivially true for $j = k$. Consider $2 \leq j < k$. Let $u_1 \dots u_j$ and $u_1' \dots u_j'$ be two vertex-disjoint copies of K_j on $W_{[j-1]}$ with $u_1, u_1' \in W_1$ and $u_{i+1}, u_{i+1}' \in W_i$ for each $i \in [j-1]$. By definition, u_1 and u_1' each has at most ξn neighbours in $I_{\{1\}}^*$ and u_{i+1} and u_{i+1}' each has at most ξn neighbours in $I_{\{i\}}^*$ for each $i \in [j-1]$. Then, by (2.65), (2.66), (2.72) and (2.74) we have

$$\begin{aligned} &\deg(u_1, \dots, u_j; W_j) \\ &\geq \sum_{i\in[j-1]} \left(\delta - n + |W_j| + \left|I_{\{i\}}^*\right| - \xi n\right) + \left(\delta - n + |W_j| + \left|I_{\{1\}}^*\right| - \xi n\right) - (j-1)|W_j| \\ &\geq -j(n-\delta) + |W_j| + \left|I_{[j-1]}^*\right| + \left|I_{\{1\}}^*\right| - j\xi n \\ &\geq -j(n-\delta) + \left|I_{[j]}^*\right| + \left|I_{\{1\}}^*\right| - j\xi n \\ &\geq n - \delta - 10k\eta n - 2(k-1)(d+2\varepsilon)n - j\xi n > 1. \end{aligned}$$

An analogous argument gives

$$\deg(u'_1, \dots, u'_j; W_j) \geq n - \delta - 10k\eta n - 2(k-1)(d+2\varepsilon)n - j\xi n > 1.$$

Hence, there are distinct vertices $u_{j+1} \in \Gamma(u_1, \dots, u_j; W_j)$ and $u'_{j+1} \in \Gamma(u'_1, \dots, u'_j; W_j)$. Notice that $u_1 \dots u_{j+1}$ and $u'_1 \dots u'_{j+1}$ are two vertex-disjoint copies of K_{j+1} on $W_{[j]}$ each comprising two vertices of W_1 and a vertex of W_h for each $1 < h \leq j$, so by the inductive hypothesis there are two vertex-disjoint copies of K_k on $W_{[k-1]}$ each comprising two vertices of W_1 and a vertex of W_h for each $1 < h < k$, completing the proof. ∎

Now suppose that for some $i \in [k-1]$ we have a copy $u_1 \dots u_k$ of K_k on $W_{[k-1]}$ with two vertices of W_i and a vertex of W_h for each $h \in [k-1] \setminus \{i\}$. We shall consider the $i = 1$ case and note that for each $i \neq 1$ an analogous version of the following argument applies. Without loss of generality, let $u_1 \in W_1$ and $u_{i+1} \in W_i$ for $i \in [k-1]$. We shall count the common neighbours of $u_1 \dots u_k$ outside $I^*_{[k-1]}$. By definition u_1 has at most ξn neighbours in $I^*_{\{1\}}$ and u_{i+1} has at most ξn neighbours in $I^*_{\{i\}}$ for each $i \in [k-1]$. Then, (2.65), (2.66), (2.72), (2.74) and the fact that $k\delta - (k-1)n \geq k\nu n$ imply that $u_1 \dots u_k$ has at least

$$\begin{aligned} \sum_{i \in [k-1]} & \left(\delta - \left|I^*_{[k-1]}\right| + \left|I^*_{\{i\}}\right| - \xi n\right) + \left(\delta - \left|I^*_{[k-1]}\right| + \left|I^*_{\{1\}}\right| - \xi n\right) \\ & - (k-1)\left(n - \left|I^*_{[k-1]}\right|\right) \\ & \geq (k-1)\delta - (k-2)n - \frac{k\delta - (k-1)n}{48} \end{aligned} \tag{2.75}$$

common neighbours outside $I^*_{[k-1]}$. Now the following claim tells us that we are done if we can find two vertex-disjoint copies of K_k which satisfy (2.75).

Claim 2.37. *Suppose that $u_1 \dots u_k$ and $u'_1 \dots u'_k$ are vertex-disjoint copies of K_k in G such that each of them has at least $(k-1)\delta - (k-2)n - \frac{k\delta-(k-1)n}{48}$ common neighbours outside $I^*_{[k-1]}$. Then G contains $P^k_{\mathrm{pp}_k(n,\delta)}$ and C^k_ℓ for each $\ell \in [k+1, \mathrm{pc}_k(n,\delta)]$ such that $\chi(C^k_\ell) \leq k+2$.*

Proof. Let D' be the set of clusters $C \in V(R) \setminus I_{[k-1]}$ such that $u_1 \dots u_k$ has at most $\frac{2dn}{m}$ common neighbours in C. By the hypothesis, $u_1 \dots u_k$ has at least $(k-1)\delta - (k-2)n - \frac{k\delta-(k-1)n}{48}$ common neighbours outside $I^*_{[k-1]}$. Of these, at most εn are in the exceptional set V_0 of the regular partition, and at most $\frac{2dn|D'|}{m}$ are in $\bigcup D'$. The remaining common neighbours all lie in $\bigcup\Big(V(R)\setminus(I_{[k-1]}\cup D')\Big)$, so by (ii) we have the inequality

$$\begin{aligned}&(k-1)\delta-(k-2)n-\frac{k\delta-(k-1)n}{48}-\varepsilon n-\frac{2dn|D'|}{m}\\&\le(m-\left|I_{[k-1]}\right|-|D'|)\frac{n}{m}\le n-(k-1)(n-\delta)+5k\eta n-|D'|\frac{n}{m}.\end{aligned}$$

Simplifying this, we obtain

$$(1-2d)\frac{n}{m}|D'|\le\varepsilon n+5k\eta n+\frac{k\delta-(k-1)n}{48},$$

so by (2.65) and (2.66) we have $|D'|\le\frac{(k\delta-(k-1)n)m}{40n}$.

Now let D be the set of clusters $C\in V(R)\setminus I_{[k-1]}$ such that either $u_1\dots u_k$ or $u_1'\dots u_k'$ has at most $\frac{2dn}{m}$ common neighbours in C. Since the same analysis holds for $u_1'\dots u_k'$, we obtain

$$|D|\le\frac{(k\delta-(k-1)n)m}{20n}. \tag{2.76}$$

We now show that there is a copy $X_1\dots X_{k-2}$ of K_{k-2} in R such that $X_j\in I_j\setminus D$ for each $j\in[k-2]$. In fact, we prove the following statement for all $i\in[k-2]$ by backwards induction on i: there is a copy $X_1\dots X_i$ of K_i in R such that $X_j\in I_j\setminus D$ for each $j\in[i]$. Setting $i=k-2$ then gives the desired statement.

Consider $i=1$. From (2.73) and (2.76) we conclude that

$$\begin{aligned}|I_1\setminus D|&\ge\frac{m}{n}\left(n-\delta-5k\eta n-(k-2)(d+2\varepsilon)n-\frac{k\delta-(k-1)n}{20}\right)\\&\ge\frac{m}{n}\left(n-\delta-\frac{k\delta-(k-1)n}{10}\right)>0,\end{aligned}$$

so we may choose $X_1\in I_1\setminus D$. Now consider $1<i\le k-2$. By the induction hypothesis, there is a copy $X_1\dots X_{i-1}$ of K_{i-1} such that $X_j\in I_j\setminus D$ for each $j\in[i-1]$. By (ii) I_j is an independent set for each $j\in[i-1]$, so $\Gamma(X_1,\dots,X_{i-1})\cap I_{[i-1]}=\varnothing$. Then applying Lemma 2.2, (2.65), (2.66), (2.69), (2.73) and (2.76), we obtain

$$\begin{aligned}\deg(X_1,\dots,X_{i-1};I_i)&\ge\deg(X_1,\dots,X_{i-1})-m+\left|I_{[i]}\right|\\&\ge\left|I_{[i]}\right|-(i-1)(m-\delta')\\&\ge\frac{m}{n}((k-1)(n-\delta)-5k\eta n)-(k-2)(m-\delta')\\&=\frac{m}{n}(n-\delta-(k-2)(d+2\varepsilon)n-5k\eta n)\\&\ge\frac{(k\delta-(k-1)n)m}{2n}>|D|,\end{aligned}$$

so we may pick $X_i \in \Gamma(X_1, \ldots, X_{i-1}) \cap (I_i \setminus D)$. Then, $X_1 \ldots X_i$ is a copy of K_i such that $X_j \in I_j \setminus D$ for each $j \in [i]$, concluding our inductive proof.

Hence, there is a copy $X_1 \ldots X_{k-2}$ of K_{k-2} such that $X_j \in I_j \setminus D$ for each $j \in [k-2]$. By (ii) I_j is an independent set for each $j \in [k-1]$, so $\Gamma(X_1, \ldots, X_{k-2}) \cap I_{[k-2]} = \varnothing$. Now by Lemma 2.2, (2.65), (2.66), (2.69), (2.71), (2.73) and (2.76), we have

$$\begin{aligned}
\deg(X_1, \ldots, X_{k-2}; B_1) &\geq \deg(X_1, \ldots, X_{k-2}) - m + |B_1| + \left|I_{[k-2]}\right| \\
&\geq |B_1| + \left|I_{[k-2]}\right| - (k-2)(m - \delta') \\
&\geq |B_1| + \frac{m}{n}((k-1)(n-\delta) - 5k\eta n) - (k-1)(m - \delta') \\
&= |B_1| - (5k\eta + (k-1)(d + 2\varepsilon))m \\
&\geq \frac{(k\delta - (k-1)n)m}{2n} > |D|,
\end{aligned}$$

so we may pick $X \in \Gamma(X_1, \ldots, X_{k-2}) \cap (B_1 \setminus D)$. By Lemma 2.2, (iii), (2.65), (2.66), (2.69), (2.70) and (2.76) we have

$$\begin{aligned}
\deg(X_1, \ldots, X_{k-2}, X; B_1) &\geq \deg(X_1, \ldots, X_{k-2}) - |I_{k-1}| \\
&\geq k\delta' - (k-1)m \\
&\geq \frac{(k\delta - (k-1)n)m}{2n} > |D|,
\end{aligned}$$

so we may pick $Y \in \Gamma(X_1, \ldots, X_{k-2}, X) \cap (B_1 \setminus D)$. By analogous argument we may pick $X' \in \Gamma(X_1, \ldots, X_{k-2}) \cap (B_2 \setminus D)$ and $Y' \in \Gamma(X_1, \ldots, X_{k-2}, X') \cap (B_2 \setminus D)$. Therefore, we have copies $X_1 \ldots X_{k-2}XY$ and $X_1 \ldots X_{k-2}X'Y'$ of K_k such that $X_j \in I_j \setminus D$ for each $j \in [k-2]$, $X, Y \in B_1 \setminus D$ and $X', Y' \in B_2 \setminus D$.

Since $\delta_R(B_1), \delta_R(B_2) \geq \delta' - |I|$ and $|B_i| > \delta_R(B_i)$ for all $i \in [2]$, by Lemma 2.1(i) we can find a matching $F_2 := M$ in $R[B_1 \cup B_2]$ with $\delta' - |I|$ edges. Using a step-by-step process with steps $1, \ldots, k-1$, we will extend the edges of F_2 to copies of K_{k+1} each consisting of an edge of F_2 and exactly one vertex from each I_i. The final collection of copies of K_{k+1} will have size at least $k\delta' - (k-1)m - 5k\eta m$. Let $i \in [k-1]$ and let F_{i+1} be the set of at least $i\delta' - (i-1)m - \sum_{h=i}^{k-1} |I_h| - 5k\eta m$ vertex-disjoint copies of K_{i+1} which we have immediately before step i. Every cluster in I_i has at most $m - |I_i| - \delta'$ non-neighbours outside I_i. Hence, every cluster in $|I_i|$ forms a copy of K_{i+2} with at least $|F_{i+1}| - (m - |I_i| - \delta') \geq (i+1)\delta' - im - \sum_{h=i+1}^{k-1} |I_h| - 5k\eta m$ copies of K_{i+1} of

F_{i+1}. Therefore, we may choose greedily clusters in I_i to obtain a set F_{i+2} of at least

$$\min\left\{(i+1)\delta' - im - \sum_{h=i+1}^{k-1} |I_h| - 5k\eta m, |I_i|\right\} \geq (i+1)\delta' - im - \sum_{h=i+1}^{k-1} |I_h| - 5k\eta m$$

vertex-disjoint copies of K_{i+2} formed from copies of K_{i+1} of F_{i+1} and clusters of I_i. After step $k-1$, we have a set $T := F_{k+1}$ of at least $k\delta' - (k-1)m - 5k\eta m$ vertex-disjoint copies of K_{k+1} each comprising an edge of M and a vertex from I_i for each $i \in [k-1]$. Let T_1 be the collection of the copies of K_{k+1} of T contained in $B_1 \cup I_{[k-1]}$ and T_2 the collection of those contained in $B_2 \cup I_{[k-1]}$. By (iii), all the copies of K_{k+1} in T_1 are in the same K_{k+1}-component as $X_1 \dots X_{k-2}XY$ and all the copies of K_{k+1} in T_2 are in the same K_{k+1}-component as $X_1 \dots X_{k-2}X'Y'$.

Apply Lemma 2.12 with $X_{i,j} = X_j$ for $(i,j) \in [2] \times [k-2]$ and $X_{i,k-1} = X, X_{i,k} = Y$ for $i \in [2]$ to find the kth power of a path starting with $u_1 \dots, u_k$ and ending with $u'_1 \dots u'_k$ using the copies of K_{k+1} in T_1. Similarly, apply Lemma 2.12 with $X_{i,j} = X_j$ for $(i,j) \in [2] \times [k-2]$, $X_{i,k-1} = X', X_{i,k} = Y'$ for $i \in [2]$ and A as the set of vertices of the kth power of a path we have above which are not in $\bigcup T_1$, to find the kth power of a path starting with $u'_1 \dots, u'_k$ and ending with $u_1 \dots u_k$ using the copies of K_{k+1} in T_2, intersecting the first only at $u_1, \dots, u_k$ and $u'_1, \dots, u'_k$. Choosing appropriate lengths for these kth power of paths and concatenating them yields the kth power of a cycle C^k_ℓ for any $6(k+1)m^{k+1} \leq \ell \leq (k+1)(1-d)\left(k\delta' - (k-1)m - 5k\eta m\right)\frac{n}{m}$. Applying Lemma 2.12 to a copy of K_{k+2} in a K_{k+1}-component directly yields C^k_ℓ for each $k+1 \leq \ell \leq (k+1)(1-d)\frac{n}{m}$ such that $\chi(C^k_\ell) \leq k+2$. By (2.65) and (2.67) we have $(k+1)(1-d)\frac{n}{m} \geq 6(k+1)m^{k+1}$, and by (2.65), (2.66) we have $(k+1)(1-d)\left(k\delta' - (k-1)m - 5k\eta m\right)\frac{n}{m} \geq \mathrm{pc}_k(n,\delta)$. It follows that G contains C^k_ℓ for each $\ell \in [k+1, \mathrm{pc}_k(n,\delta)]$ such that $\chi(C^k_\ell) \leq k+2$. For $\delta \in \left[\left(\frac{k-1}{k} + \nu\right)n, \frac{kn-1}{k+1}\right)$, note that by (2.65), (2.66) we have $(k+1)(1-d)\left(k\delta' - (k-1)m - 5k\eta m\right)\frac{n}{m} \geq \mathrm{pp}_k(n,\delta)$, so G contains $P^k_{\mathrm{pp}_k(n,\delta)}$. ∎

By Claim 2.36 and (2.75), if we can find two vertex-disjoint edges in some W_i, then we are done by Claim 2.37. Hence, we assume in the following that W_i does not contain two vertex-disjoint edges for each $i \in [k-1]$. This means that for each $i \in [k-1]$ there are two vertices in W_i which meet every edge in W_i. For each $i \in [k-1]$ let W'_i be W_i without these two vertices. Since neither of these two vertices has more than ξn neighbours in $I^*_{\{i\}} \subseteq W_i$, while $|I_i| \geq \frac{m}{k+1} - 5k\eta m - (k-2)(d+2\varepsilon)m$ by (2.73) and because $\delta < \frac{kn}{k+1}$, there is a vertex in W_i adjacent to no vertex of W_i. By (2.74) we conclude that

$$|J|(n-\delta) - 5k\eta n - (k-1-|J|)(d+2\varepsilon)n - \varepsilon n \le |I_J^*| \le |W_J| \le |J|(n-\delta) \quad (2.77)$$

for each $J \subseteq [k-1]$. Set $W := W_{[k-1]}$. For each $i \in [k-1]$ the total number of non-edges between W_i and $V(G) \setminus W_i$ is at most

$$\begin{aligned} |W_i||V(G) \setminus W_i| - |W_i|(\delta - 2) &= |W_i|(n - \delta + 2 - |W_i|) \\ &\le |W_i|((k-1)(n-\delta) + 2 - |W|). \end{aligned}$$

Hence, by (2.77) the total number of non-edges between W and $V(G) \setminus W$ is at most

$$|W|\big((k-1)(n-\delta) + 2 - |W|\big) \le |W|(5k\eta n + \varepsilon n + 2) \le 5k\eta n^2 + \varepsilon n^2 + 2n.$$

In particular, by the definition of ξ and (2.67), we have

$$\left|\left\{v \in V(G) \setminus W : \deg(v; W) < |W| - \xi^2 n\right\}\right| \le \xi^2 n. \quad (2.78)$$

Recall that the sets B_i are subsets of $V(R)$ and the elements of each set B_i correspond to clusters in $V(G)$. We shall denote by $\bigcup B_i$ the union of the elements of the set B_i as clusters in $V(G)$. By (iii) we have $|B_i| \le \frac{19m(k\delta-(k-1)n)}{10n}$, which together with $\delta < \frac{kn}{k+1}$, (2.65), (2.66) and (2.72) implies

$$\begin{aligned} \left|\bigcup B_i\right| &\le \frac{19}{10}(k\delta - (k-1)n) \le \frac{19}{20}((k-1)\delta - (k-2)n) \\ &< (k-1)\delta - (k-2)n - \xi n - (d+\varepsilon)n. \end{aligned} \quad (2.79)$$

By (iii) and the definition of an (ε, d)-regular partition, vertices in $\bigcup B_i$ have at most $(d+\varepsilon)n$ neighbours outside of $(\bigcup B_i) \cup I^*_{[k-1]}$; hence, by $\delta(G) \ge \delta$, (2.77) and (2.79) they have more than ξn neighbours in $\bigcup I_h$ for all $h \in [k-1]$. Now the definition of W_h implies $\bigcup B_i \cap W_h = \varnothing$ for all $(i,h) \in [\ell] \times [k-1]$, so in fact

$$\bigcup B_i \cap W = \varnothing \text{ for all } i \in [\ell]. \quad (2.80)$$

Furthermore, (2.65), (2.66), (2.72), (2.77) and (2.79) imply that $v \in \bigcup B_i$ has at least

$$\delta - |W| - (d+\varepsilon)n \ge k\delta - (k-1)n - (d+\varepsilon)n > \left|\bigcup B_i\right|/2 + 50\xi^2 n \quad (2.81)$$

neighbours in $\bigcup B_i$.

Now for each $i \in [\ell]$ let A_i be the set of vertices in $\bigcup B_i$ which are adjacent to at least $|W| - \xi^2 n$ vertices of W. By (2.78) we have

$$\left| \bigcup_{i \in [\ell]} \left(\bigcup B_i \right) \setminus A_i \right| \leq \xi^2 n. \tag{2.82}$$

Vertices which are neither in W nor in any of the sets A_i must either be in the exceptional set V_0 or in $(\bigcup B_i) \setminus A_i$ for some i, so we have

$$\left| V_0 \cup \bigcup_{i \in [\ell]} \left(\bigcup B_i \right) \setminus A_i \right| \leq \varepsilon n + \xi^2 n < 2\xi^2 n. \tag{2.83}$$

As such, (2.81) implies that

$$\delta(G[A_i]) \geq |A_i|/2 + 48\xi^2 n \tag{2.84}$$

and since $|B_i| > \delta' - \left| I_{[k-1]} \right| \geq k\delta' - (k-1)m$, we have

$$|A_i| \geq \left| \bigcup B_i \right| - \xi^2 n \geq (1-\varepsilon)\frac{n}{m}|B_i| - \xi^2 n \geq k\delta - (k-1)n - 2\xi^2 n \tag{2.85}$$

for each $i \in [\ell]$, where we have used (2.65), (2.66), (2.69) and the definition of ξ.

The following claim uses A_1 to obtain powers of cycles of all lengths up to near-extremal.

Claim 2.38. *$C_\ell^k \subseteq G$ for each $\ell \in [k+1, \frac{k+1}{2}|A_1|]$ such that $\chi(C_\ell^k) \leq k+2$.*

Proof. By Lemma 2.35 (with $B = \varnothing$) we find in A_1 a copy of $C_{2h'}$ for each $2h' \in [4, \min\{|A_1|, \frac{2n}{k+2}\}]$. We shall construct a copy of $C_{(k+1)h'}^k$ from this cycle by repeated application of Lemma 2.34. We have steps $j = 1, \dots, k-1$. In step j we start with a copy of $C_{(j+1)h'}^j$

$$T_j = q_{1,j-1} \dots q_{1,1} t_1 t_2 \dots q_{h',j-1} \dots q_{h',1} t_{2h'-1} t_{2h'}$$

in G, with $t_i \in A_1$ for $i \in [2h']$, $q_{f,g} \in W_g$ for $(f,g) \in [h'] \times [j-1]$, such that each vertex is adjacent to the immediately preceding j vertices in cyclic order.

Any $2(j+1)$-tuple of consecutive vertices on T_j comprises four vertices from A_1 and two vertices from W_i for each $i \in [j-1]$. Each vertex in A_1 has at least $|W_j| - \xi^2 n$ neighbours in W_j, while for each $i \in [j-1]$ a vertex in W_i has at least

$|W_j| - (n-\delta) + \left|I^*_{\{i\}}\right| - \xi n$ neighbours in W_j. Applying Lemma 2.2 and (2.74), we find that every $2(j+1)$-tuple of consecutive vertices on T_j has at least

$$\begin{aligned}&|W_j| - 4\xi^2 n - 2(j-1)\xi n - 2(j-1)(n-\delta) + 2\left|I^*_{[j-1]}\right| \\ &\geq |W_j| - 4\xi^2 n - 2(j-1)\xi n - 10k\eta n - 2(k-j)(d+2\varepsilon)n - 2\varepsilon n\end{aligned}$$

common neighbours in W_j. Since $\delta < \frac{kn}{k+1}$ and by (2.65), (2.66), (2.72) and (2.77), we have

$$|W_j| - 4\xi^2 n - 2(j-1)\xi n - 10k\eta n - 2(k-j)(d+2\varepsilon)n - 2\varepsilon n \geq \frac{n}{k+2}.$$

This means that we can apply Lemma 2.34 with G and W_j to obtain a copy of $C^{j+1}_{(j+2)h'}$

$$T_{j+1} = q_{1,j}\dots q_{1,1}t_1t_2\dots q_{h',j}\dots q_{h',1}t_{2h'-1}t_{2h'}$$

in G, with $t_i \in A_1$ for $i \in [2h']$, $q_{f,g} \in W_g$ for $(f,g) \in [h'] \times [j]$, such that each vertex is adjacent to the preceding j vertices in cyclic order. Terminating after step $k-1$ gives us a copy of $C^k_{(k+1)h'}$. Hence, we are able to find copies of C^k_h for $h \in [k+1, \frac{k+1}{2}\min\{|A_1|, \frac{2n}{k+2}\}]$ such that h is divisible by $k+1$.

To obtain a copy of C^k_h for h not divisible by $k+1$, we perform a so-called parity correction procedure. Fix $g \in [k]$. We seek a copy of $C^k_{(k+1)h'+g}$ with $h' \geq g$. Let $h'' := h' - g$. Pick (by Theorem 2.9) vertices $a_{i,j}$ for $(i,j) \in [g] \times [3]$ in A_1 such that $a_{i,1}a_{i,2}a_{i,3}$ is a triangle for each $i \in [g]$ and $a_{i,3}a_{i+1,1}$ is an edge for $i \in [g-1]$. Let $A = \{a_{i,j} \mid (i,j) \in [g] \times [3]\}$. Apply Lemma 2.35 to find a path $P'_1 = a_{1,1}p_{2h''}\dots p_1 a_{g,3}$ in $(A_1 \setminus A) \cup \{a_{1,1}, a_{g,3}\}$ on $2(h''+1)$ vertices whose end-vertices are $a_{1,1}$ and $a_{g,3}$. For each $a \in A$, insert a dummy vertex a' into G with the same adjacencies as a. Define $P_1^{(i)} := a_{i+1,2}a_{i+1,1}a_{i,3}a'_{i,2}a'_{i,1}$ for $i \in [g-1]$ and $P_1 := a_{1,2}P'_1a'_{g,2}a'_{g,1}$.

We shall construct a copy of $C^k_{(k+1)h'+g}$ from these paths by repeatedly applying Lemma 2.34 and suitably truncating and concatenating the resultant kth powers of paths. We have steps $j = 1, \dots, k-1$. In step 1 we start with the paths $P_1, P_1^{(1)}, \dots, P_1^{(g-1)}$. We seek to apply Lemma 2.34 with W_1 to each path. For P_1 take $Q_1 = a_{1,2}a_{1,1}$, $Q_2 = a_{1,2}a_{1,1}p_{2h''}p_{2h''-1}$, $Q_i = p_{2(h''-i+3)}p_{2(h''-i+3)-1}p_{2(h''-i+2)}p_{2(h''-i+2)-1}$ for $3 \leq i \leq h''+1$ and $Q_{h''+2} = p_2p_1a_{g,3}a'_{g,2}a'_{g,1}$, and apply Lemma 2.34 with W_1 to obtain the squared path

$$qa_{1,2}a_{1,1}q_{h'',1}p_{2h''}p_{2h''-1}\dots q_{1,1}p_2p_1q_1^{(g)}a_{g,3}a'_{g,2}a'_{g,1},$$

with $q_1^{(g)}, q_{x,1} \in W_1$ for each $x \in [h'']$, such that each vertex is adjacent to the preceding 2 vertices in cyclic order and $q_1^{(g)}$ adjacent to $a'_{g,1}$. Let P_2 be the result of replacing q in the above squared path with $a'_{1,3}$. For $P_1^{(i)}$ with $i \in [g-1]$, take $Q_1 = a_{i+1,2}a_{i+1,1}$, $Q_2 = a_{i+1,2}a_{i+1,1}a_{i,3}a'_{i,2}a'_{i,1}$, and apply Lemma 2.34 with W_1 to obtain the squared path

$$qa_{i+1,2}a_{i+1,1}q_1^{(i)}a_{i,3}a'_{i,2}a'_{i,1},$$

such that $q_1^{(i)} \in W_1$ adjacent to $a'_{g,1}$ and each vertex is adjacent to the preceding 2 vertices in cyclic order. Let $P_2^{(i)}$ be the result of replacing q in the above squared path with $a'_{i+1,3}$.

In step $j \geq 2$ we start with jth powers of paths

$$\begin{aligned} P_j =& (q_{j-2}^{(1)})' \dots (q_1^{(1)})' a'_{1,3}a_{1,2}a_{1,1}q_{h'',j-1} \dots q_{h'',1}p_{2h''}p_{2h''-1} \\ & \dots q_{1,j-1} \dots q_{1,1}p_2p_1q_{j-1}^{(g)} \dots q_1^{(g)}a_{g,3}a'_{g,2}a'_{g,1}, \\ P_j^{(i)} =& (q_{j-2}^{(i+1)})' \dots (q_1^{(i+1)})' a'_{i+1,3}a_{i+1,2}a_{i+1,1}q_{j-1}^{(i)} \dots q_1^{(i)}a_{i,3}a'_{i,2}a'_{i,1} \end{aligned}$$

for each $i \in [g-1]$. We seek to apply Lemma 2.34 with W_j to each of them. For P_j take $Q_1 = (q_{j-2}^{(1)})' \dots (q_1^{(1)})' a'_{1,3}a_{1,2}a_{1,1}$,

$$\begin{aligned} Q_2 &= (q_{j-2}^{(1)})' \dots (q_1^{(1)})' a'_{1,3}a_{1,2}a_{1,1}q_{h'',j-1} \dots q_{h'',1}p_{2h''}p_{2h''-1}, \\ Q_i &= q_{h''-i+3,j-1} \dots q_{h''-i+3,1}p_{2(h''-i+3)}p_{2(h''-i+3)-1} \\ & \quad q_{h''-i+2,j-1} \dots q_{h''-i+2,1}p_{2(h''-i+2)}p_{2(h''-i+2)-1} \end{aligned}$$

for each $3 \leq i \leq h''+1$, and

$$Q_{h''+2} = q_{1,j-1} \dots q_{1,1}p_2p_1q_{j-1}^{(g)} \dots q_1^{(g)}a_{g,3}a'_{g,2}a'_{g,1}.$$

Applying Lemma 2.34 with W_j yields the $(j+1)$st power of a path

$$\begin{aligned} & q(q_{j-2}^{(1)})' \dots (q_1^{(1)})' a'_{1,3}a_{1,2}a_{1,1}q_{h'',j} \dots q_{h'',1}p_{2h''}p_{2h''-1} \\ & \dots q_{1,j} \dots q_{1,1}p_2p_1q_j^{(g)} \dots q_1^{(g)}a_{g,3}a'_{g,2}a'_{g,1} \end{aligned}$$

with $q_j^{(g)}, q_{x,j} \in W_1$ for each $x \in [h'']$, such that each vertex is adjacent to the preceding $j+1$ vertices in cyclic order and $q_j^{(g)}$ adjacent to $a'_{g,1}$. Insert a dummy vertex $(q_{j-1}^{(1)})'$ into G with the same adjacencies as $q_{j-1}^{(1)}$. Define P_{j+1} to be the above $(j+1)$st power of a path with q replaced by $(q_{j-1}^{(1)})'$. For $P_j^{(i)}$ with $i \in [g-1]$, take

$$\begin{aligned} Q_1 &= (q_{j-2}^{(i+1)})' \dots (q_1^{(i+1)})' a'_{i+1,3}a_{i+1,2}a_{i+1,1} \text{ and} \\ Q_2 &= (q_{j-2}^{(i+1)})' \dots (q_1^{(i+1)})' a_{i+1,3}a_{i+1,2}a_{i+1,1}q_{j-1}^{(i)} \dots q_1^{(i)}a_{i,3}a'_{i,2}a'_{i,1}. \end{aligned}$$

Applying Lemma 2.34 with W_j yields the $(j+1)$st power of a path

$$q(q_{j-2}^{(i+1)})' \dots (q_1^{(i+1)})' a'_{i+1,3} a_{i+1,2} a_{i+1,1} q_j^{(i)} \dots q_1^{(i)} a_{i,3} a'_{i,2} a'_{i,1}$$

such that $q_j^{(i)} \in W_j$ adjacent to $a'_{g,1}$ and each vertex is adjacent to the preceding 2 vertices in cyclic order. Insert a dummy vertex $(q_{j-1}^{(i+1)})'$ into G with the same adjacencies as $q_{j-1}^{(i+1)}$. Define $P_{j+1}^{(i)}$ to be the above $(j+1)$st power of a path with q replaced by $(q_{j-1}^{(i+1)})'$.

After step $k-1$, we have kth powers of paths $P_{k-1}, P_{k-1}^{(1)}, \dots, P_{k-1}^{(g-1)}$. We delete the cloned vertices from each of them and concatenate the resultant kth powers of paths to obtain the kth power of a cycle on $(k+1)h' + g$ vertices. Therefore, we can obtain $C_{\ell'}^k$ for every $\ell' \in [k+1, \frac{k+1}{2}\min\{|A_1|, \frac{2n}{k+2}\}]$ such that $\chi(C_{\ell'}^k) \le k+2$. Since $\mathrm{pc}_k(n,\delta) \le \frac{(k+1)n}{k+2}$ by (2.68), we obtain the desired result. ∎

It remains to show that we have $C_{\ell'}^k \subseteq G$ for every $\frac{k+1}{2}|A_1| \le \ell' \le \mathrm{pc}_k(n,\delta)$ and that in the case $\delta \in \left[\left(\frac{k-1}{k} + \nu\right)n, \frac{kn-1}{k+1}\right)$ we have $P_{\mathrm{pp}_k(n,\delta)}^k \subseteq G$. To do so, we need to incorporate vertices which are not 'nice' enough to be included in the sets A_i. Define X_i as A_i together with all vertices in $V(G) \setminus W$ with at least $30\xi^2 n$ neighbours in A_i. Every vertex of $V(G) \setminus W$ has at least $\delta - |W|$ neighbours outside W, so by (2.77) every vertex of $V(G) \setminus W$ is in X_i for at least one i. Let $i, j \in [\ell]$ satisfy $i \neq j$. Since $A_h \subseteq \bigcup B_h$, we have $A_i \cap A_j = \varnothing$. By the definition of an (ε, d)-regular partition and (iii), vertices in A_i have at most $(d+\varepsilon)n$ neighbours outside of $(\bigcup B_i) \cup I_{[k-1]}^*$; by (2.80) vertices in A_i have at most $(d+\varepsilon)n < 30\xi^2 n$ neighbours in A_j. Hence, we have

$$A_i \cap X_j = \varnothing. \tag{2.86}$$

Then, it follows from (2.83) that

$$|X_i| < |A_i| + 2\xi^2 n. \tag{2.87}$$

We shall now show the desired outcome by considering three cases based on the values of $|X_i \cap X_j|$. The following claim deals with the case when $|X_i \cap X_j| \ge 2$ for some $i \neq j$.

Claim 2.39. *Suppose that $|X_i \cap X_j| \ge 2$ for some $i \neq j$. Then we have $C_\ell^k \subseteq G$ for every $\frac{k+1}{2}|A_1| \le \ell \le \mathrm{pc}_k(n,\delta)$ and if further $\delta \in \left[\left(\frac{k-1}{k} + \nu\right)n, \frac{kn-1}{k+1}\right)$ we also have $P_{\mathrm{pp}_k(n,\delta)}^k \subseteq G$.*

Proof. Let $i \neq j$ such that $|X_i \cap X_j| \geq 2$. Let u_1 and u_2 be distinct vertices of $X_i \cap X_j$. Let v_1 and v_2 be distinct neighbours in A_i of u_1 and u_2 respectively, and similarly w_1 and w_2 in A_j. Applying Lemma 2.35 in A_i, we can find a path from v_1 to v_2 of length α for any $4 \leq \alpha \leq |A_i| - 1$. We can find a similar path in A_j from w_1 to w_2. Concatenating these paths with u_1 and u_2, we can find a cycle $S_{2h'}$ of length $2h'$ in $X_i \cup X_j$ for any $12 \leq 2h' \leq \min\{|A_i| + |A_j| + 2, \frac{2n}{k+2}\}$. We shall construct the desired copy of $C^k_{(k+1)h'}$ from this cycle by repeated application of Lemma 2.34. We have steps $j = 1, \ldots, k-1$. In step j we start with a copy of $C^j_{(j+1)h'}$

$$T_j = q_{1,j-1} \ldots q_{1,1} t_1 t_2 \ldots q_{h',j-1} \ldots q_{h',1} t_{2h'-1} t_{2h'}$$

in G, with $t_p \in A_i \cup A_j \cup \{u_1, u_2\}$ for $p \in [2h']$, $q_{f,g} \in W'_g$ for each $f \in [h'], g \in [j-1]$, such that each vertex is adjacent to the immediately preceding j vertices in cyclic order and no $2(j+1)$-tuple of consecutive vertices on T_j uses both u_1 and u_2.

Any $2(j+1)$-tuple of consecutive vertices on T_j comprises four vertices from $A_i \cup A_j \cup \{u_1, u_2\}$ and two vertices from W'_h for each $h \in [j-1]$. Each vertex in $A_i \cup A_j$ has at least $|W'_j| - \xi^2 n$ neighbours in W'_j, u_1 and u_2 each has at least $\xi n - 2$ neighbours in W'_j, and for each $i \in [j-1]$ a vertex in W'_i has at least $|W'_j| - (n - \delta) + \left|I^*_{\{i\}}\right| - 2$ neighbours in W'_j. Hence, the four $2(j+1)$-tuples which use either u_1 or u_2 each has at least

$$\begin{aligned} &\xi n - 2 - 3\xi^2 n - 2(j-1)(n-\delta) + 2\left|I^*_{[j-1]}\right| \\ &\geq \xi n - 2 - 3\xi^2 n - 10k\eta n - 2(k-j+1)(d+\varepsilon)n > 100\ell \end{aligned}$$

common neighbours in W'_j, with the first inequality following from (2.77) and the second inequality following from (2.67), (2.72) and from

$$\ell \leq \nu^{-1}. \tag{2.88}$$

Every other $2(j+1)$-tuple of consecutive vertices on T_j has at least

$$\begin{aligned} &|W'_j| - 4\xi^2 n - 2(j-1)(n-\delta) + 2\left|I^*_{[j-1]}\right| \\ &\geq |W'_j| - 4\xi^2 n - 10k\eta n - 2(k-j+1)(d+2\varepsilon)n \end{aligned}$$

common neighbours in W'_j. By the definition of ξ, (2.65), (2.66) and (2.77), we have

$$|W'_j| - 4\xi^2 n - 10k\eta n - 2(k-j+1)(d+2\varepsilon)n \geq \frac{n}{k+2}.$$

This means that we can apply Lemma 2.34, with G, W'_j, and an ordering σ of the relevant $2(j+1)$-tuples which has all the $2(j+1)$-tuples containing u_1 or u_2 coming first, to obtain a copy of $C^{j+1}_{(j+2)h'}$

$$T_{j+1} = q_{1,j} \dots q_{1,1} t_1 t_2 \dots q_{h',j} \dots q_{h',1} t_{2h'-1} t_{2h'}$$

in G, with $t_p \in A_i \cup A_j \cup \{u_1, u_2\}$ for $p \in [2h']$, $q_{f,g} \in W'_g$ for each $f \in [h'], g \in [j]$, such that each vertex is adjacent to the immediately preceding j vertices in cyclic order and no $2(j+2)$-tuple of consecutive vertices on T_{j+1} uses both u_1 and u_2. Terminating after step $k-1$ gives us a copy of $C^k_{(k+1)h'}$. Hence, we are able to find copies of C^k_h for $h \in [k+1, \frac{k+1}{2} \min\{|A_i| + |A_j| + 2, \frac{2n}{k+2}\}]$ such that h is divisible by $k+1$.

To obtain a copy of C^k_h for h not divisible by $k+1$, we perform a parity correction procedure. Fix $g \in [k]$. We seek a copy of $C^k_{(k+1)h'+g}$ with $h' \geq g+7$. Let $h'' := h'-g \geq 7$. Let $u_1, u_2, v_1, v_2, w_1, w_2$ be the vertices previously picked. For the purpose of parity correction, pick (by Theorem 2.9) vertices $a_{x,y}$ for $(x,y) \in [g] \times [3]$ in A_i such that $a_{x,1}a_{x,2}a_{x,3}$ is a triangle for each $x \in [g]$ and $a_{x,3}a_{x+1,1}$ is an edge for $x \in [g-1]$. Let $A' = \{a_{x,y} \mid (x,y) \in [g] \times [3]\}$. Pick a common neighbour v of v_1 and $a_{1,1}$ in A_i which is not in $A' \cup \{v_2\}$. Applying Lemma 2.34 suitably, we can find a path in $A_i \setminus (A \cup \{v, v_1\})$ from $a_{g,3}$ to v_2 of length h for any $4 \leq h \leq |A_i| - 3g - 2$ and a path in A_j from w_1 to w_2 of length h for any $4 \leq h \leq |A_j| - 1$. Concatenating these paths with $u_1, u_2, v_1, v, a_{1,1}, a_{g,3}$, we can find a path of length $2h''+1$ in $A_i \cup A_j \cup \{u_1, u_2\}$ for any $15 \leq 2h''+1 \leq \min\{|A_i| + |A_j| - 3g + 3, \frac{2n}{k+2}\}$. This allows us to construct a copy of $C^k_{(k+1)h'+g}$ whenever $h' \geq g+7$ by applying the method used previously. Therefore, we can obtain $C^k_{\ell'} \subseteq G$ for every $\ell' \in [k^2+9k+7, \frac{k+1}{2} \min\{|A_i| + |A_j| - 3k, \frac{2n}{k+2}\}]$ such that $\chi(C^k_{\ell'}) \leq k+2$. By (2.85), (2.11) and (2.12) we have $\mathrm{pc}_k(n,\delta) \leq \frac{k+1}{2} \min\{|A_i| + |A_j| - 3k, \frac{2n}{k+2}\}$, so G contains $C^k_{\ell'}$ for every $\frac{k+1}{2}|A_1| \leq \ell' \leq \mathrm{pc}_k(n,\delta)$. For the case $\delta \in \left[\left(\frac{k-1}{k} + \nu\right) n, \frac{kn-1}{k+1}\right)$ we note that $P^k_\ell \subseteq C^k_\ell$ and by (2.85), (2.11) and (2.68) we have $\mathrm{pp}_k(n,\delta) \leq \frac{k+1}{2} \min\{|A_i|+|A_j|-3k, \frac{2n}{k+2}\}$, so G contains $P^k_{\mathrm{pp}_k(n,\delta)}$. This completes the proof. ∎

The following claim deals with the case when there exists $i \in [\ell]$ such that every vertex of A_i is adjacent to some vertex outside $X_i \cup W$.

Claim 2.40. *Suppose that there exists $i \in [\ell]$ such that every vertex of A_i is adjacent to some vertex outside $X_i \cup W$. Then we have $C^k_\ell \subseteq G$ for every $\frac{k+1}{2}|A_1| \leq \ell \leq \mathrm{pc}_k(n,\delta)$ and if further $\delta \in \left[\left(\frac{k-1}{k} + \nu\right) n, \frac{kn-1}{k+1}\right)$ we also have $P^k_{\mathrm{pp}_k(n,\delta)} \subseteq G$.*

Proof. Since we have

$$|A_i| \overset{(2.85)}{\geq} \left|\bigcup B_i\right| - \xi^2 n \overset{(2.71)}{\geq} \frac{38}{39}\nu(1-\varepsilon)n - \xi^2 n \overset{(2.72)}{\geq} 25\xi n \overset{(2.72),(2.88)}{>} 50\ell\xi^2 n,$$

there exists $j \neq i$ such that there are $50\xi^2 n$ vertices in A_i all adjacent to vertices of $X_j \setminus X_i$. No vertex of $X_j \setminus X_i$ is adjacent to $30\xi^2 n$ vertices of A_i (by definition of X_i), so there are two disjoint edges u_1v_1 and u_2v_2 from $u_1, u_2 \in A_i$ to $v_1, v_2 \in X_j$. Then, choosing distinct neighbours w_1 of v_1 and w_2 of v_2 in A_j and applying the same reasoning as in Claim 2.39 completes the proof. ∎

Now we deal with the remainder case. The following claim deals with finding the kth power of a path of the desired length in this case.

Claim 2.41. *Suppose that $\delta \in \left[\left(\frac{k-1}{k}+\nu\right)n, \frac{kn-1}{k+1}\right)$, for each $i \neq j$ we have $|X_i \cap X_j| \leq 1$ and for each i there is a vertex of A_i adjacent only to vertices in $X_i \cup W$. Then we have $P^k_{\mathrm{pp}_k(n,\delta)} \subseteq G$.*

Proof. In this case we have $|X_i| \geq \delta - |W| + 1$ for each $i \in [\ell]$. We first focus on finding the kth power of a path on $\mathrm{pp}_k(n,\delta)$ vertices when $\delta \in \left[\left(\frac{k-1}{k}+\nu\right)n, \frac{kn-1}{k+1}\right)$. Note that if $|X_i \cap X_j| = 1$ for some $i \neq j$, then we obtain the kth power of a path of the desired length as in Claim 2.39. We required two vertices in $|X_i \cap X_j|$ previously for a cycle to cross from X_i to X_j and back to X_i, whereas here we only need one vertex for a path to cross from X_i to X_j.

Hence, assume that the sets X_i are all disjoint. This implies that $\ell \leq \frac{n-|W|}{\delta-|W|+1}$. Note that $|W| \leq (k-1)(n-\delta)$ by (2.77), so we have

$$\ell \leq \frac{n-(k-1)(n-\delta)}{\delta-(k-1)(n-\delta)+1} = \frac{(k-1)\delta-(k-2)n}{k\delta-(k-1)n+1}.$$

Now if $\ell \geq r_p(n,\delta)+1$, we would have $r_p(n,\delta)+1 \leq \ell \leq \frac{(k-1)\delta-(k-2)n}{k\delta-(k-1)n+1}$, and so $r_p(n,\delta) \leq \frac{n-\delta-1}{k\delta-(k-1)n+1}$, but by (2.1) we have $r_p(n,\delta) \geq \frac{n-\delta}{k\delta-(k-1)n+1}$, so we have $\ell \leq r_p(n,\delta)$. Therefore, the largest of the sets X_i, say X_1, has at least

$$|X_1| \geq \frac{n-|W|}{\ell} \overset{(2.77)}{\geq} \frac{(k-1)\delta-(k-2)n}{\ell} \geq \frac{(k-1)\delta-(k-2)n}{r_p(n,\delta)} \tag{2.89}$$

vertices.

We wish to apply Lemma 2.35 with $H = G[X_1]$ and 'bad' vertices $B = X_1 \setminus A_1$. Note that by (2.87) B contains at most $2\xi^2 n$ vertices, so we have

$$|B| \overset{(2.87)}{\leq} 2\xi^2 n \overset{(2.72)}{\leq} \frac{\nu[(k-1)\delta-(k-2)n]}{100} \overset{(2.88)}{\leq} \frac{(k-1)\delta-(k-2)n}{100\ell} \overset{(2.89)}{\leq} \frac{|H|}{100}.$$

Moreover, we have $\delta(H) \geq \delta(G[X_1]) \geq 30\xi^2 n$ by definition of X_1, so every vertex of B has at least $30\xi^2 n \geq 9 \cdot 2\xi^2 n \geq 9|B|$ neighbours in H. For $v \in X_1 \setminus B = A_1$, we have

$$\deg(v; X_1) \overset{(2.84)}{\geq} \frac{|A_1|}{2} + 48\xi^2 n \overset{(2.87)}{>} \frac{|X_1|}{2} + 47\xi^2 n = \frac{|H|}{2} + 47\xi^2 n \overset{(2.67)}{\geq} \frac{|H|}{2} + 9|B| + 10.$$

Hence, we may apply Lemma 2.35 to obtain a path P in X_1 with $\alpha := \min\left\{|X_1|, \frac{2n}{k+2}\right\}$ vertices, on which no four consecutive vertices contain more than one vertex of B. Define $h' := \lfloor \frac{\alpha}{2} \rfloor$ and $\beta := \alpha - 2h' \in \{0, 1\}$. We shall construct the desired copy of $P^k_{\mathrm{pp}_k(n,\delta)}$ from P by repeated application of Lemma 2.34. We have steps $j = 1, \ldots, k-1$. In step j we start with a copy of $P^j_{(j+1)h'+j-1+\beta}$

$$T_j = q_{1,j-1} \ldots q_{1,1} t_1 t_2 \ldots q_{h',j-1} \ldots q_{h',1} t_{2h'-1} t_{2h'}$$
$$q_{h'+1,j-1} \ldots q_{h'+1,1} t_{2h'+1} \ldots t_{2h'+\beta}$$

in G, with $t_p \in X_1$ for $p \in [\alpha]$, $q_{f,g} \in W'_g$ for each $f \in [h'+1], g \in [j-1]$, such that each vertex is adjacent to the preceding j vertices and no $2(j+1)$-tuple of consecutive vertices on T_j contains more than one vertex of B.

There are at most $2|B| \leq 4\xi^2 n$ $2(j+1)$-tuples containing vertices of B. Any $2(j+1)$-tuple of consecutive vertices on T_j comprises four vertices from X_1 and two vertices from W'_i for each $i \in [j-1]$. Each vertex in A_1 has at least $|W'_j| - \xi^2 n$ neighbours in W'_j, each vertex in B has at least $\xi n - 2$ neighbours in W'_j, and for each $i \in [j-1]$ a vertex in W'_i has at least $|W'_j| - (n-\delta) + \left|I^*_{\{i\}}\right| - 2$ neighbours in W'_j. Hence, the $2(j+1)$-tuples which contain a vertex of B each has at least

$$\begin{aligned}
&\xi n - 2 - 3\xi^2 n - 2(j-1)(n-\delta) + 2\left|I^*_{[j-1]}\right| \\
&\geq \xi n - 2 - 3\xi^2 n - 10k\eta n - 2(k-j+1)(d+2\varepsilon)n > 100\ell
\end{aligned}$$

common neighbours in W'_j, with the first inequality following from (2.77) and the second inequality following from (2.67), (2.72) and (2.88). Every other $2(j+1)$-tuple of consecutive vertices on T_j has at least

$$\begin{aligned}
&|W'_j| - 4\xi^2 n - 2(j-1)(n-\delta) + 2\left|I^*_{[j-1]}\right| \\
&\geq |W'_j| - 4\xi^2 n - 10k\eta n - 2(k-j+1)(d+2\varepsilon)n
\end{aligned}$$

common neighbours in W'_j. By the definition of ξ, (2.65), (2.66) and (2.77), we have

$$|W'_j| - 4\xi^2 n - 10k\eta n - 2(k-j+1)(d+2\varepsilon)n \geq \frac{n}{k+2}.$$

This means that we can apply Lemma 2.34, with an ordering σ of the relevant $2(j+1)$-tuples which has all the $2(j+1)$-tuples containing vertices of B coming first, to obtain a copy of $P^{j+1}_{(j+2)h'+j+\beta}$

$$T_{j+1} = q_{1,j}\dots q_{1,1}t_1t_2\dots q_{h',j}\dots q_{h',1}t_{2h'-1}t_{2h'}$$
$$q_{h'+1,j}\dots q_{h'+1,1}t_{2h'+1}\dots t_{2h'+\beta}$$

in G, with $t_p \in X_1$ for $p \in [\alpha]$, $q_{f,g} \in W'_g$ for each $f \in [h'+1], g \in [j]$, such that each vertex is adjacent to the preceding j vertices and no $2(j+2)$-tuple of consecutive vertices on T_j contains more than one vertex of B. Terminating after step $k-1$ gives the kth power of a path on at least $(k+1)h'+k-1+\beta$ vertices. We consider two cases. First consider when $\alpha = \frac{2n}{k+2}$. In this case, we have the kth power of a path on at least

$$(k+1)\left(\frac{n}{k+2} - \frac{k+1}{k+2}\right) + k - 1 \geq \frac{(k+1)n}{k+2} - 2 \geq \mathrm{pp}_k(n,\delta)$$

vertices, with the inequality following from (2.68). Otherwise, we have $\alpha = |X_1|$. Define $h'' := \left\lfloor \frac{|X_1|}{2} \right\rfloor$ and $\beta' := |X_1| - 2h'' \in \{0,1\}$. In this case, we have the kth power of a path on at least

$$(k+1)h'' + k - 1 + \beta' = (k-1)(h''+1) + |X_1| \geq \mathrm{pp}_k(n,\delta)$$

vertices, with the inequality following from (2.89) and the definition of $\mathrm{pp}_k(n,\delta)$. ∎

Finally, the following claim deals with finding kth powers of cycles of the desired lengths in the remainder case.

Claim 2.42. *Suppose that for each $i \neq j$ we have $|X_i \cap X_j| \leq 1$ and for each i there is a vertex of A_i adjacent only to vertices in $X_i \cup W$. Then we have $C^k_\ell \subseteq G$ for every $\frac{k+1}{2}|A_1| \leq \ell \leq \mathrm{pc}_k(n,\delta)$.*

Proof. First consider when there is a cycle of sets (relabelling the indices if necessary) $X_1,\dots,X_s$ for some $3 \leq s \leq \ell$ such that $X_i \cap X_{i+1} = \{u_i\}$ for each i and the u_i are all distinct. In this case for each i we may choose neighbours $v_i \in A_i$ and $w_i \in A_{i+1}$ of u_i, and we may insist that these $3s$ vertices are distinct. Similarly as before, we may apply Lemma 2.35 to each $G[A_i]$ in turn and concatenate the resulting paths, in order to find a cycle $T_{2h'}$ for every $6s \leq 2h' \leq \min\{|A_i| + |A_j|, \frac{2n}{k+2}\}$ on which there are no quadruples using more than one vertex outside $\bigcup_{i\in[s]} A_i$. Arguing in a manner similar to Claim 2.39, we may repeatedly apply Lemma 2.34 to obtain a copy of $C^k_{(k+1)h'}$. Hence, we are able to find copies of C^k_h for $h \in [3s(k+1), \frac{k+1}{2}\min\{|A_i|+|A_j|, \frac{2n}{k+2}\}]$ such that h is divisible by $k+1$. To obtain a copy of C^k_h for h not divisible by $k+1$, we use a parity correction procedure analogous to that in Claim 2.39. Therefore, we can find copies of C^k_h for $h \in [k^2+3(s+1)k+(3s+1), \frac{k+1}{2}\min\{|A_i|+|A_j|-3k, \frac{2n}{k+2}\}]$. Hence, we have $C^k_{\ell'} \subseteq G$ for every $\ell' \in [k+1, \frac{k+1}{2}\min\{|A_i|+|A_j|-3k, \frac{2n}{k+2}\}]$ such that $\chi(C^k_{\ell'}) \leq k+2$.

Otherwise, no such cycle of sets exists. In this case, we have $\sum_{i=1}^{\ell}|X_i| \leq n-|W|+\ell-1$. Note that $|X_i| \geq \delta - |W| + 1$ for each $i \in [\ell]$, so this implies that $\ell \leq \frac{n-|W|-1}{\delta-|W|}$. Note that $|W| \leq (k-1)(n-\delta)$ by (2.77), so we have

$$\ell \leq \frac{n-(k-1)(n-\delta)-1}{\delta-(k-1)(n-\delta)} = \frac{(k-1)\delta-(k-2)n-1}{k\delta-(k-1)n}.$$

Now if $\ell \geq r_c(n,\delta)+1$, we would have $r_c(n,\delta)+1 \leq \ell \leq \frac{(k-1)\delta-(k-2)n}{k\delta-(k-1)n}$, and so $r_c(n,\delta) \leq \frac{n-\delta-1}{k\delta-(k-1)n}$, but we have $r_c(n,\delta) \geq \frac{n-\delta}{k\delta-(k-1)n}$, so we have $\ell \leq r_c(n,\delta)$. Therefore, the largest of the sets X_i, say X_1, has at least

$$|X_1| \geq \frac{n-|W|}{\ell} \geq \frac{(k-1)\delta-(k-2)n}{\ell} \geq \frac{(k-1)\delta-(k-2)n}{r_c(n,\delta)}$$

vertices.

As before, by Lemma 2.35 we find in X_1 for each $2h' \in [4, \min\{|X_1|, \frac{2n}{k+2}\}]$ a copy of $C_{2h'}$ on which no four consecutive vertices contain more than one vertex of B, and by repeated application of Lemma 2.34 we obtain the kth power of a cycle $C^k_{(k+1)h'}$ for each $(k+1)h' \in [2(k+1), \mathrm{pc}_k(n,\delta)]$. As before, we may apply a parity correction procedure for copies of C^k_h where h is not divisible by $k+1$. Therefore, we have copies of C^k_h for $h \in [k+1, \mathrm{pc}_k(n,\delta)\}]$ such that $\chi(C^k_{\ell'}) \leq k+2$. ■

Claims 2.39, 2.40, 2.41 and 2.42 collectively yield the desired outcome. □

2.7 Embedding Lemma

In this section we provide our proof of Lemma 2.12. To do so, we shall apply a version of a graph blow-up lemma by Allen, Böttcher, Hàn, Kohayakawa and Person [4]. We remark that the blow-up lemma of Komlós, Sárközy and Szemerédi [38] is perfectly adequate for this proof; our choice of blow-up lemma is not driven by necessity, but rather a desire to reduce the technical complexity of our proof.

We introduce some terminology in order to formulate this version of the blow-up lemma. We will use the definition of (ε, d)-regular as given in Section 2.2.2; this involves both an upper bound and a lower bound on densities. We note that the corresponding graph blow-up lemma in [4] applies to a more general class of graphs; in particular, the regularity condition in [4] is weaker and involves only a lower bound.

Let G and H be graphs with partitions $\mathcal{V} = \{V_i\}_{i\in[r]}$ and $\mathcal{X} = \{X_i\}_{i\in[r]}$ of their respective vertex sets. Let $\kappa \geq 1$. Let R be a graph on r vertices.

- $\mathcal{V}$ and $\mathcal{X}$ are *size-compatible* if $|V_i| = |X_i|$ for all $i \in [r]$.
- $\mathcal{V}$ is *κ-balanced* if there exists $m \in \mathbb{N}$ such that $m \leq |V_i| \leq \kappa m$ for all $i \in [r]$.
- $\mathcal{X}$ is an *R-partition* of H if each part of $\mathcal{X}$ is nonempty, and whenever there are edges of H between X_i and X_j, the pair ij is an edge of R,
- $\mathcal{V}$ is an *(ε, d)-regular R-partition* of G if for each edge $ij \in E(R)$ the pair (V_i, V_j) is (ε, d)-regular in G.

We say that R is an *(ε, d)-full-reduced graph* of the partition $\mathcal{V}$. Since each partition is subordinate to some graph in these definitions, for the sake of brevity we shall often refer to the graph–partition pair. For example, we shall say that $(H, \mathcal{X})$ is an R-partition and $(G, \mathcal{V})$ is an (ε, d)-regular R-partition.

We remark that the notion of an (ε, d)-regular R-partition as defined above is distinct from an (ε, d)-regular partition as defined in Section 2.2.2: in an (ε, d)-regular R-partition, the partition $\mathcal{V}$ does not have an exceptional set and we allow each vertex of G to be incident to possibly many edges which are not in (ε, d)-regular pairs. An (ε, d)-full-reduced graph is correspondingly distinct from an (ε, d)-reduced graph.

Definition 2.43 (Buffer sets)**.** Let R be a graph on r vertices, $(H, \mathcal{X})$ be an R-partition and $(G, \mathcal{V})$ be a size-compatible (ε, d)-regular R-partition. Let $\alpha > 0$. A family $\bar{\mathcal{X}} = \{\bar{X}_i\}_{i\in[r]}$ of subsets $\bar{X}_i \subseteq X_i$ is an *α-buffer* for H if

(i) the elements of $\bar{X}_i$ are isolated vertices in H,
(ii) $|\bar{X}_i| \geq \alpha |X_i|$ for all $i \in [r]$.

We remark that this corresponds to the notion of an (α, R')-buffer for H in [4] with R' as the empty spanning subgraph of R.

Definition 2.44 (Image restrictions)**.** Let R be a graph on r vertices, $(H, \mathcal{X})$ be an R-partition and $(G, \mathcal{V})$ be a size-compatible (ε, d)-regular R-partition with $G \subseteq K_n$.

Let $\mathcal{I} = \{I_x\}_{x \in V(H)}$ be a collection of subsets of $V(G)$, called *image restrictions*, and $\mathcal{J} = \{J_x\}_{x \in V(H)}$ be a collection of subsets of $V(K_n) \setminus V(G)$, called *restricting vertices*. We say that $\mathcal{I}$ and $\mathcal{J}$ are a $(\rho, \zeta, \Delta, \Delta_J)$-*restriction pair* if the following properties hold for each $i \in [r]$ and $x \in X_i$.

(a) The set $X_i^* \subseteq X_i$ of *image restricted* vertices in X_i, that is, vertices such that $I_x \neq V_i$, has size $|X_i^*| \leq \rho |X_i|$.
(b) If $x \in X_i^*$, then $I_x \subseteq V_i$ is of size at least $\zeta d^{|J_x|} |V_i|$.
(c) If $x \in X_i^*$, then $|J_x| + |\Gamma_H(x)| \leq \Delta$, and if $x \notin X_i^*$, then $J_x = \varnothing$.
(d) Each vertex of K_n appears in at most Δ_J of the sets of $\mathcal{J}$.
(e) If $x \in X_i^*$, then for each $xy \in E(H)$ with $y \in X_j$ the pair (V_i, V_j) is (ε, d)-regular in G.

Lemma 2.45 (Allen, Böttcher, Hàn, Kohayakawa and Person [4]). *For all $\Delta \geq 2, \Delta_J, \alpha, \zeta, d > 0$, $\kappa > 1$ there exists $\varepsilon, \rho > 0$ such that for all r_1 there exists $n_{BL} \in \mathbb{N}$ such that for all $n \geq n_{BL}$ the following holds. Let R be a graph on $r \leq r_1$ vertices. Let H and G be n-vertex graphs with κ-balanced size-compatible vertex partitions $\mathcal{X} = \{X_i\}_{i \in [r]}$ and $\mathcal{V} = \{V_i\}_{i \in [r]}$, respectively, which have parts of size at least $m \geq n/(\kappa r_1)$. Let $\bar{\mathcal{X}} = \{\bar{X}_i\}_{i \in [r]}$ be a family of subsets of $V(H)$, $\mathcal{I} = \{I_x\}_{x \in V(H)}$ be a family of image restrictions, and $\mathcal{J} = \{J_x\}_{x \in V(H)}$ be a family of restricting vertices. Suppose that*

(i) $\Delta(H) \leq \Delta$, $(H, \mathcal{X})$ is an R-partition, and $\bar{\mathcal{X}}$ is an α-buffer for H,

(ii) $(G, \mathcal{V})$ is an (ε, d)-regular R-partition,

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

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

Proof of Lemma 2.12. We proceed by checking the conditions for a suitable application of Lemma 2.45 to embed a relevant graph H into G. We first prove (i) and (ii). Fix $k \geq 2$, $d > 0$ and set $\Delta = 2k, \Delta_J = k, \kappa = 2, \alpha = \frac{d}{2}, \zeta = 1$. Now Lemma 2.45 outputs $\varepsilon_0, \rho_0 > 0$. We choose

$$\varepsilon_{EL} = \min \left\{ \frac{\varepsilon_0}{k+1}, \frac{d^2}{8(k+1)} \right\}.$$

Given $0 < \varepsilon < \varepsilon_{EL}, r_{EL} \in \mathbb{N}$, Lemma 2.45 outputs $n_{BL} \in \mathbb{N}$. We choose

$$n_{EL} = \max \left\{ n_{BL}, \frac{6r_{EL}^{k+2}}{\varepsilon}, \frac{4r_{EL}}{\rho_0} \right\}.$$

Let $n \geq n_{EL}$, let G be a graph on n vertices and let R be an (ε, d)-reduced graph of G on $r \leq r_{EL}$ vertices. Let $V_0, V_1, \ldots, V_r$ be the vertex classes of the (ε, d)-regular partition of G which underpins R. Fix a connected K_{k+1}-factor $\mathcal{F}$ in R which contains $c := \frac{\mathrm{CKF}_{k+1}(R)}{k+1}$ copies of K_{k+1}. Let $T_1, \ldots, T_c$ be the copies of K_{k+1} in $\mathcal{F}$. Let $\mathcal{V} := \{V_1, \ldots, V_r\}$. Let R' be the empty spanning subgraph of R. Let G^* be the subgraph of G induced on $\mathcal{V}$ and set $n^* := |V(G^*)|$. Note that $(G^*, \mathcal{V})$ is an (ε, d)-regular R-partition. Note that $|V_i| \geq (1-\varepsilon)\frac{n}{r} \geq \frac{n}{2r_{EL}}$ for all $i \in [r]$, so $\mathcal{V}$ is 2-balanced.

Let H be a copy of C_ℓ^k together with additional isolated vertices so that it has n^* vertices. Let $v_1, \ldots, v_{n^*}$ be its vertices, with $v_1, \ldots, v_\ell$ being the vertices of the copy of C_ℓ^k in a cyclic order. Let $C := \{v_i : i \in [\ell]\}$. Suppose that we have a vertex partition $\mathcal{X} := \{X_i\}_{i\in[r]}$ of H and a family $\bar{\mathcal{X}} := \{\bar{X}_i\}_{i\in[r]}$ of subsets of $V(H)$ such that $\mathcal{X}$ is size-compatible with $\mathcal{V}$, $(H, \mathcal{X})$ is an R-partition, and $\bar{\mathcal{X}}$ is an α-buffer for H. Define $\mathcal{I} := \{I_x\}_{x\in V(H)}$ and $\mathcal{J} := \{J_x\}_{x\in V(H)}$ by $I_x = V_i$ for $x \in X_i$ and $J_x = \varnothing$ for $x \in V(H)$. Note that $\mathcal{I}$ and $\mathcal{J}$ form a $(\rho_0, \zeta, \Delta, \Delta_J)$-restriction pair. Then, by Lemma 2.45 we will have an embedding $\phi : V(H) \to V(G^*)$, which will then complete our proof of (i) and (ii). Therefore, for suitable values of ℓ it remains to find a vertex partition $\mathcal{X}$ of H and a family $\bar{\mathcal{X}}$ of subsets of $V(H)$ such that $\mathcal{X}$ is size-compatible with $\mathcal{V}$, $(H, \mathcal{X})$ is an R-partition and $\bar{\mathcal{X}}$ is an α-buffer for H.

We start with (i) and we will consider $\ell \leq \frac{(1-d)(k+1)cn}{r}$ divisible by $k+1$. We first consider the case $\ell \leq \frac{(k+1)(1-d)n}{r}$ divisible by $k+1$, that is, when $c = 1$. Let $Y_1, \ldots, Y_{k+1}$ be the vertices of T_1. Define $\phi : V(H) \to V(R)$ as follows. For $i \leq \ell$, set $\phi(v_i) = Y_j$ with $j \equiv i \mod k+1$. For $i > \ell$, given $\phi(v_1), \ldots, \phi(v_{i-1})$, set $\phi(v_i) = V_j$ with $j = \min\{h : |\{b < i : \phi(v_b) = V_h\}| < |V_h|\}$. Set $X_i := \phi^{-1}(V_i)$, $\bar{X}_i := \phi^{-1}(V_i) \setminus C$ for $i \in [r]$. Define $\mathcal{X} := \{X_i\}_{i\in[r]}$ and $\bar{\mathcal{X}} := \{\bar{X}_i\}_{i\in[r]}$. Since all edges in H have pairs of vertices in C at most k apart in the cyclic order as endpoints and any $k+1$ consecutive vertices in the cyclic order are mapped to a copy of K_{k+1} in R, it follows that $(H, \mathcal{X})$ is an R-partition. Furthermore, for each $i \in [r]$ at most $\frac{\ell}{k+1} \leq (1-d)\frac{n}{r} \leq |V_i|$ vertices in C are mapped to V_i, so $\mathcal{X}$ is a vertex partition of H which is size-compatible with $\mathcal{V}$. Finally, $\bar{X}_i$ is a set of isolated vertices in H by definition and

$$|\bar{X}_i| = |X_i| - |C \cap X_i| \geq \left(1 - \frac{1-d}{1-\varepsilon}\right)|X_i| \geq \alpha|X_i|$$

for each $i \in [r]$, so $\bar{\mathcal{X}}$ is an α-buffer for H. This completes the proof in this case. We are done if $c = 1$, so we can assume $c \geq 2$ for the remainder of (i).

Next, we consider the case $\ell \in \left(\frac{(1-d)(k+1)n}{r}, \frac{(1-d)(k+1)cn}{r}\right]$ divisible by $k+1$. For each $i \in [c-1]$, fix a K_{k+1}-walk W_i whose first copy of K_k is in T_i and whose last is in T_{i+1}, which is of minimal length. We have $|W_i| \le \binom{r}{k}$ for each $i \in [c-1]$. Let W' be the K_{k+1}-walk obtained by concatenating $W_1, \dots, W_{c-1}$.

We describe how to construct the sequence $Q(W, \overrightarrow{U_{11} \dots U_{1k}})$ for any K_{k+1}-walk $W = (E_1, E_2, \dots)$ in R and any orientation $\overrightarrow{U_{11} \dots U_{1k}}$ of E_1, its first copy of K_k. We construct $Q(W, \overrightarrow{U_{11} \dots U_{1k}})$ iteratively as follows. Let $Q_1 = (U_{11}, \dots, U_{1k})$. Now for $2 \le i \le |W|$ successively, we define Q_i as follows. The last k vertices $U_{(i-1)1}, \dots, U_{(i-1)k}$ of Q_{i-1} are an orientation of E_{i-1}. We have $E_i = U_{(i-1)1} \dots U_{(i-1)(j-1)} U_{(i-1)(j+1) \dots U_{(i-1)k}} U_{ik}$ for some $j \in [k]$. Create Q_i by appending $(U_{ik}, U_{(i-1)1}, \dots U_{(i-1)(j-1)})$ to Q_{i-1}. At each step the last k vertices of Q_i are an orientation of E_i and every vertex of Q_i is adjacent in R to the k vertices preceding it in Q_i. Finally we let $Q(W, \overrightarrow{U_{11} \dots U_{1k}}) := Q_{|W|}$.

It is easy to check by induction that for any K_{k+1}-walk W whose first edge is $U_{11} \dots U_{1k}$, we have

$$|Q(W, \overrightarrow{U_{11} \dots U_{1k}})| + |Q(W, \overrightarrow{U_{1k} \dots U_{11}})| \equiv -2 \mod k+1. \tag{2.90}$$

Now consider the concatenation W' of the walks W_i. Let $U_{11} \dots U_{1k}$ be the first copy of K_k of W_1. If we construct $Q(W', \overrightarrow{U_{11} \dots U_{1k}})$ then the first copy of K_k $U_{i1} \dots U_{ik}$ and the last copy of K_k $U'_{i1} \dots U'_{ik}$ of each W_i obtain orientations, say $\overrightarrow{U_{i1} \dots U_{ik}}$ and $\overrightarrow{U'_{i1} \dots U'_{ik}}$. Clearly, there are sequences $\bar{Q}_i$ of vertices in T_i for $1 < i < c$, such that $Q(W', \overrightarrow{U_{11} \dots U_{1k}})$ is the concatenation of

$$Q(W_1, \overrightarrow{U_{11} \dots U_{1k}}), \bar{Q}_2, Q(W_2, \overrightarrow{U_{21} \dots U_{2k}}), \dots, \bar{Q}_{c-1}, Q(W_{c-1}, \overrightarrow{U_{(c-1)1} \dots U_{(c-1)k}}).$$

Let $\bar{Q}_1 := T_1 - U_{11} \dots U_{1k}$ and $\bar{Q}_c := T_c - U'_{c1} \dots U'_{ck}$. Define $f_i \equiv |\bar{Q}_i| \mod k+1$ for $i \in [c]$. Together with (2.90), we obtain

$$|Q(W', \overrightarrow{U_{1k} \dots U_{11}})| + \sum_{i \in [c-1]} (|Q(W_i, \overrightarrow{U_{i1} \dots U_{ik}})| + f_i) + f_c \equiv 0 \mod k+1. \tag{2.91}$$

Let Q' denote $Q(W', \overrightarrow{U_{1k} \dots U_{11}})$ and let Q^*_i denote $Q(W_i, \overrightarrow{U_{i1} \dots U_{ik}})$ for each $i \in [c-1]$. Define $q' := |Q'|$ and $q_i := |Q^*_i|$ for each $i \in [c-1]$. For a sequence Q of vertices of R, let $(Q)_h$ denote the hth term of Q.

Let $U_{11} \dots U_{1k}$ be the first copy of K_k in W_1. Orient it as $\overrightarrow{U_{11} \dots U_{1k}}$. Construct Q^*_i for $i \in [c-1]$ and Q' as described before, and define q_i, f_i for $i \in [c-1]$ and q', f_c as before. Let $T_i = Y_{i1} \dots Y_{i(k+1)}$ for $i \in [c]$ be such that $\overrightarrow{Y_{i2} \dots Y_{i(k+1)}}$ is the oriented last copy of K_k of W_{i-1} in Q^*_{i-1} for $2 \le i \le c$ and $\overrightarrow{Y_{1(k+1)} \dots Y_{12}}$ is the oriented first copy of K_k of W' in Q'. Define the following. Let $\alpha := \sum_{i=1}^{c-1}(q_i + f_i) + f_c + q'$.

$$p_0 := \max\left\{p \in \mathbb{Z} \mid \ell \ge p(1-d)(k+1)\frac{n}{r} + \alpha\right\},$$

$$t_i = \begin{cases} (1-d)\frac{n}{r} & \text{if } i \in [p_0] \\ \frac{\ell-\alpha}{k+1} - p_0(1-d)\frac{n}{r} & \text{if } i = p_0 + 1 \\ 0 & \text{if } i > p_0 + 1, \end{cases}$$

$$L_0 = 0, L_j = \sum_{i=1}^{j}[t_i(k+1) + q_i + f_i] \text{ for } j \in [c-1],$$

$$M_j = L_{j-1} + t_j(k+1) + f_j \text{ for } j \in [c].$$

Define $\phi : V(H) \to V(R)$ as follows. For $i \le \ell$, set

$$\phi(v_i) = \begin{cases} Y_{jh} & \text{if } L_{j-1} < i \le M_j, \text{ with } h \equiv i - L_{j-1} \mod k+1 \\ (Q^*_j)_{i-M_j} & \text{if } M_j < i \le L_j \\ (Q')_{M_c+q'+1-i} & \text{if } M_c < i \le \ell. \end{cases}$$

For $i > \ell$, given $\phi(v_1), \dots, \phi(v_{i-1})$, set $\phi(v_i) = V_j$ with $j = \min\{h : |\{b < i : \phi(v_b) = V_h\}| < |V_h|\}$.

Set $X_i := \phi^{-1}(V_i)$, $\bar{X}_i := \phi^{-1}(V_i) \setminus C$ for $i \in [r]$. Define $\mathcal{X} := \{X_i\}_{i\in[r]}$ and $\bar{\mathcal{X}} := \{\bar{X}_i\}_{i\in[r]}$. Since all edges in H have pairs of vertices in C at most k apart in the cyclic order as endpoints and any $k+1$ consecutive vertices in the cyclic order are mapped to a copy of K_{k+1} in R, it follows that $(H, \mathcal{X})$ is an R-partition. Furthermore, for each $i \in [r]$ at most $(1-d)\frac{n}{r} + \frac{2r\binom{r}{k}}{k+1} \le (1-d+\varepsilon)\frac{n}{r} \le (1-\varepsilon)\frac{n}{r} \le |V_i|$ vertices in C are mapped to V_i, so $\mathcal{X}$ is a vertex partition of H which is size-compatible with $\mathcal{V}$. Finally, $\bar{X}_i$ is a set of isolated vertices in H by definition and

$$|\bar{X}_i| = |X_i| - |C \cap X_i| \ge \left(1 - \frac{1-d+\varepsilon}{1-\varepsilon}\right)|X_i| \ge \alpha|X_i|$$

for each $i \in [r]$, so $\bar{\mathcal{X}}$ is an α-buffer for H. This completes the proof in this case and for (i).

We continue with (ii) and we will consider $\ell \le \frac{(1-d)(k+1)cn}{r}$ satisfying $\chi(C^k_\ell) \le k+2$. Pick $y \in [k] \cup \{0\}$ such that $\ell \equiv y \mod k+1$. In particular, we have $\ell \ge y(k+2)$. Let S be a copy of K_{k+2} in the same K_{k+1}-component as the copies of K_{k+1} in $\mathcal{F}$ and let

$Z_1, \dots, Z_{k+2}$ be the vertices of S. We first consider the case $\ell \le \frac{(k+1)(1-d)n}{r}$ satisfying $\chi(C_\ell^k) \le k+2$. Define $\phi : V(H) \to V(R)$ as follows. For $i \le \ell$, set

$$\phi(v_i) = \begin{cases} Z_j & \text{if } i \le \ell - y(k+2), \text{ with } j \equiv i \mod k+1, \\ Z_j & \text{if } \ell - y(k+2) < i \le \ell, \text{ with } j \equiv i \mod k+2. \end{cases}$$

For $i > \ell$, given $\phi(v_1), \dots, \phi(v_{i-1})$, set $\phi(v_i) = V_j$ with $j = \min\{h : |\{b < i : \phi(v_b) = V_h\}| < |V_h|\}$. Set $X_i := \phi^{-1}(V_i)$, $\bar{X}_i := \phi^{-1}(V_i) \setminus C$ for $i \in [r]$ and define $\mathcal{X} := \{X_i\}_{i \in [r]}$ and $\bar{\mathcal{X}} := \{\bar{X}_i\}_{i \in [r]}$. Since all edges in H have pairs of vertices in C at most k apart in the cyclic order as endpoints and any $k+1$ consecutive vertices in the cyclic order are mapped to a copy of K_{k+1} in R, it follows that $(H, \mathcal{X})$ is an R-partition. Furthermore, for each $i \in [r]$ at most $\frac{\ell}{k+1} \le (1-d)\frac{n}{r} \le |V_i|$ vertices in C are mapped to V_i, so $\mathcal{X}$ is a vertex partition of H which is size-compatible with $\mathcal{V}$. Finally, $\bar{X}_i$ is a set of isolated vertices in H by definition and

$$|\bar{X}_i| = |X_i| - |C \cap X_i| \ge \left(1 - \frac{1-d}{1-\varepsilon}\right)|X_i| \ge \alpha |X_i|$$

for each $i \in [r]$, so $\bar{\mathcal{X}}$ is an α-buffer for H. This completes the proof in this case. We are done if $c = 1$, so we can assume $c \ge 2$ for the remainder of (ii).

Next, we consider $\ell \in \left(\frac{(1-d)(k+1)n}{r}, \frac{(1-d)(k+1)cn}{r}\right]$ satisfying $\chi(C_\ell^k) \le k+2$. For each $i \in [c-1]$, fix a K_{k+1}-walk W_i whose first copy of K_k is in T_i and whose last is in T_{i+1}, which is of minimal length. We have $|W_i| \le \binom{r}{k}$ for each $i \in [c-1]$. Let W' be the K_{k+1}-walk obtained by concatenating $W_1, \dots, W_{c-1}$. Fix a K_{k+1}-walk W'' whose first copy of K_k is that of W_1, whose last is that of W_{c-1}, which includes a copy of K_k from S and is one of minimal length satisfying these conditions. We have $|W''| \le 2\binom{r}{k}$.

We construct the sequence $Q(W, \overrightarrow{U_{11} \dots U_{1k}})$ for any K_{k+1}-walk $W = (E_1, E_2, \dots)$ in R and any orientation $\overrightarrow{U_{11} \dots U_{1k}}$ of E_1, its first copy of K_k, identically to that in (i). Let $U_{11} \dots U_{1k}$ be the first copy of K_k in W_1 and orient it as $\overrightarrow{U_{11} \dots U_{1k}}$. Construct $Q(W', \overrightarrow{U_{11} \dots U_{1k}})$. Then, the first copy of K_k $U_{i1} \dots U_{ik}$ and the last copy of K_k $U'_{i1} \dots U'_{ik}$ of each W_i obtain orientations, say $\overrightarrow{U_{i1} \dots U_{ik}}$ and $\overrightarrow{U'_{i1} \dots U'_{ik}}$. Construct $Q(W_i, \overrightarrow{U_{i1} \dots U_{ik}})$ for $i \in [c]$. Clearly, there are sequences $\bar{Q}_i$ of vertices in T_i for $1 < i < c$, such that $Q(W', \overrightarrow{U_{11} \dots U_{1k}})$ is the concatenation of

$$Q(W_1, \overrightarrow{U_{11} \dots U_{1k}}), \bar{Q}_2, Q(W_2, \overrightarrow{U_{21} \dots U_{2k}}), \dots, \bar{Q}_{c-1}, Q(W_{c-1}, \overrightarrow{U_{(c-1)1} \dots U_{(c-1)k}}).$$

Let $\bar{Q}_1 := T_1 - U_{11} \dots U_{1k}$ and $\bar{Q}_c := T_c - U_{c1} \dots U_{ck}$. Define $f_i \equiv |\bar{Q}_i| \mod k+1$ for $i \in [c]$. Let Q_i^* denote $Q(W_i, \overrightarrow{U_{i1} \dots U_{ik}})$ for $i \in [c-1]$ and let Q'' denote $Q(W'', \overrightarrow{U_{1k} \dots U_{11}})$.

Define $q_i := |Q_i^*|$ for $i \in [c-1]$ and $q'' := |Q''|$. For a sequence Q of vertices of R, let $(Q)_h$ denote the hth term of Q. Let $T_i = Y_{i1} \dots Y_{i(k+1)}$ for all $i \in [c]$ be such that $\overrightarrow{Y_{i2} \dots Y_{i(k+1)}}$ is the oriented last copy of K_k of W_{i-1} in Q_{i-1} for $2 \le i \le c$ and $\overrightarrow{Y_{1(k+1)} \dots Y_{12}}$ is the oriented first copy of K_k of W'' in Q''.

Let $\alpha := \sum_{i=1}^{c-1}(q_i + f_i) + f_c + q''$. Pick $x \in [k] \cup \{0\}$ such that $\ell - \alpha \equiv x \mod k+1$. Let $Z_{k+2}, \dots, Z_3$ be the last k consecutive terms of Q'' which correspond to a copy of K_k in S. Define Q''' as the result of inserting x copies of $Z_{k+2}, \dots, Z_1$ into Q'' right before the last occurrence of $Z_{k+2}, \dots, Z_3$ in Q''. Let $q''' := |Q'''|$. Define $\alpha_x := \sum_{i=1}^{c-1}(q_i + f_i) + f_c + q''' = \alpha + x(k+2)$. Define the following.

$$p_0 := \max\left\{p \in \mathbb{Z} \mid \ell \ge p(1-d)(k+1)\frac{n}{r} + \alpha_x\right\},$$

$$t_i = \begin{cases} (1-d)\frac{n}{r} & \text{if } i \in [p_0] \\ \frac{\ell-\alpha_x}{k+1} - p_0(1-d)\frac{n}{r} & \text{if } i = p_0+1 \\ 0 & \text{if } i > p_0+1, \end{cases}$$

$$L_0 = 0, \quad L_j = \sum_{i=1}^{j}[t_i(k+1) + q_i + f_i] \text{ for } j \in [c-1],$$
$$M_j = L_{j-1} + t_j(k+1) + f_j \text{ for } j \in [c].$$

Define $\phi : V(H) \to V(R)$ as follows. For $i \le \ell$, set

$$\phi(v_i) = \begin{cases} Y_{jh} & \text{if } L_{j-1} < i \le M_j, \text{ with } h \equiv i - L_{j-1} \mod k+1 \\ (Q_j^*)_{i-M_j} & \text{if } M_j < i \le L_j \\ (Q''')_{M_c+q'''+1-i} & \text{if } M_c < i \le \ell. \end{cases}$$

For $i > \ell$, given $\phi(v_1), \dots, \phi(v_{i-1})$, set $\phi(v_i) = V_j$ with $j = \min\{h : |\{b < i : \phi(v_b) = V_h\}| < |V_h|\}$.

Set $X_i := \phi^{-1}(V_i)$, $\bar{X}_i := \phi^{-1}(V_i) \setminus C$ for $i \in [r]$. Define $\mathcal{X} := \{X_i\}_{i\in[r]}$ and $\bar{\mathcal{X}} := \{\bar{X}_i\}_{i\in[r]}$. Since all edges in H have pairs of vertices in C at most k apart in the cyclic order as endpoints and any $k+1$ consecutive vertices in the cyclic order are mapped to a copy of K_{k+1} in R, it follows that $(H, \mathcal{X})$ is an R-partition. Furthermore, for each $i \in [r]$ at most $(1-d)\frac{n}{r} + \frac{3r\binom{r}{k}+k(k+2)}{k+1} \le (1-d+\varepsilon)\frac{n}{r} \le (1-\varepsilon)\frac{n}{r} \le |V_i|$ vertices in C are mapped to V_i, so $\mathcal{X}$ is a vertex partition of H which is size-compatible with $\mathcal{V}$. Finally, $\bar{X}_i$ is a set of isolated vertices in H by definition and

$$|\bar{X}_i| = |X_i| - |C \cap X_i| \ge \left(1 - \frac{1-d+\varepsilon}{1-\varepsilon}\right)|X_i| \ge \alpha|X_i|$$

for each $i \in [r]$, so $\bar{\mathcal{X}}$ is an α-buffer for H. This completes the proof in this case and for (ii).

Now we prove (iii). Fix $k \geq 3$, $d > 0$ and let $\Delta = 2k, \Delta_J = k, \kappa = 2, \alpha = \frac{d}{2}, \zeta = 1$. Now Lemma 2.45 outputs $\varepsilon_0, \rho_0 > 0$. We choose

$$\varepsilon_{EL} = \min\left\{\frac{\varepsilon_0}{k+3}, \frac{d^2}{8(k+1)}\right\}.$$

Given $0 < \varepsilon < \varepsilon_{EL}, r_{EL} \in \mathbb{N}$, Lemma 2.45 outputs $n_{BL} \in \mathbb{N}$. We choose

$$n_{EL} = \max\left\{n_{BL}, \frac{6r_{EL}^{k+2}}{\varepsilon}, \frac{4r_{EL}}{\rho_0}\right\}.$$

Let $n \geq n_{EL}$, let G be a graph on n vertices and let R^* be an (ε, d)-reduced graph of G on $r \leq r_{EL}$ vertices. Let $V_0', V_1', \ldots, V_r'$ be the vertex classes of the (ε, d)-regular partition of G which gives rise to R^*. Let $\mathcal{T}'$ be the given connected K_{k+1}-factor in R^* with $t := |\mathcal{T}'|$ copies of K_{k+1}. Let $T_1', \ldots, T_t'$ be the copies of K_{k+1} of $\mathcal{T}'$. Let $A' := \{u_{i,j} \mid (i,j) \in [2] \times [k]\}$.

Consider $T_i' = X_{i,1}' \ldots X_{i,(k+1)}'$ for $i \in [t]$. Let $j \in [k+1]$. Remove the vertices of $A \cup A'$ from $X_{i,j}'$ to obtain $X_{i,j}$. We have $|X_{i,j}| \geq \varepsilon |X_{i,j}'|$ and $|X_{i,h}| \geq \varepsilon |X_{i,h}'|$, so the (ε, d)-regularity of $(X_{i,j}', X_{i,h}')$ implies that $(X_{i,j}, X_{i,h})$ is $(2\varepsilon, d - \varepsilon)$-regular.

Let $\{V_0, \ldots, V_r\}$ be the new vertex partition obtained by replacing each $X_{i,j}'$ with $X_{i,j}$ and let $\mathcal{V} := \{V_1, \ldots, V_r\}$. Let R be the $(2\varepsilon, d-\varepsilon)$-full-reduced graph of the partition $\mathcal{V}$. Every edge of R^* carries over to R, and let V_i be the vertex of R corresponding to V_i' in R^*. Let $\mathcal{T}$ be the connected K_{k+1}-factor in R corresponding to $\mathcal{T}'$. Let $T_1, \ldots, T_t$ be the copies of K_{k+1} in $\mathcal{T}$, with T_i corresponding to T_i' for all $i \in [t]$. Let G^* be the subgraph of G induced on $\mathcal{V}$. Let $n^* := |V(G^*)|$. Here $(G^*, \mathcal{V})$ is a $(2\varepsilon, d - \varepsilon)$-regular R-partition. Note that $|V_i| \geq (1 - 3\varepsilon)\frac{n}{r} \geq \frac{n}{2r_{EL}}$ for all $i \in [r]$, so $\mathcal{V}$ is 2-balanced.

Let $\ell' = \ell - 2k$. Let H be a copy of $P_{\ell'}^k$ together with additional isolated vertices so that it has n^* vertices. Let $w_1, \ldots, w_{n^*}$ be its vertices, with $w_1, \ldots, w_{\ell'}$ being the vertices of the copy of $P_{\ell'}^k$ in a path order. Let $P := \{w_i : i \in [\ell']\}$. Suppose that we have a vertex partition $\mathcal{X} := \{X_i\}_{i \in [r]}$ of H and a family $\bar{\mathcal{X}} := \{\bar{X}_i\}_{i \in [r]}$ of subsets of $V(H)$ such that $\mathcal{X}$ is size-compatible with $\mathcal{V}$, $(H, \mathcal{X})$ is an R-partition, and $\bar{\mathcal{X}}$ is an α-buffer for H. Suppose further that for each $j \in [k]$ we have $X_{1,j} = V_i$ and $X_{2,j} = V_h$ with i, h such that $w_j \in X_i$ and $w_{\ell'-j+1} \in X_h$. Define $\mathcal{I} := \{I_x\}_{x \in V(H)}$ and $\mathcal{J} := \{J_x\}_{x \in V(H)}$

as follows.

$$I_{w_j} = \begin{cases} V_i \cap \Gamma(u_{1,j}, \dots, u_{1,k}) & \text{for } j \in [k], \text{ with } w_j \in X_i \\ V_i & \text{for } k < j \leq \ell' - k, \text{ with } w_j \in X_i \\ V_i \cap \Gamma(u_{2,k}, \dots, u_{2,(\ell'-j+1)}) & \text{for } \ell' - k < j \leq \ell', \text{ with } w_j \in X_i, \end{cases}$$

$$J_{w_j} = \begin{cases} \{u_{1,j}, \dots, u_{1,k}\} & \text{for } j \in [k] \\ \varnothing & \text{for } k < j \leq \ell' - k \\ \{u_{2,k}, \dots, u_{2,(\ell'-j+1)}\} & \text{for } \ell' - k < j \leq \ell'. \end{cases}$$

Since $|\Gamma(u_{i,j}, \dots, u_{i,k}) \cap X_{i,j}| \geq \frac{2dn}{r} - \frac{2\varepsilon n}{r} \geq \frac{3dn}{2r}$ for each pair $(i,j) \in [2] \times [k]$ and $|V_i| \geq (1-2\varepsilon)\frac{n}{r} \geq \frac{2}{\rho_0}$, $\mathcal{I}$ and $\mathcal{J}$ form a $(\rho_0, \zeta, \Delta, \Delta_J)$-restriction pair.

Then, by Lemma 2.45 we will have an embedding $\phi : V(H) \to V(G^*)$ such that w_j is adjacent to $u_{1,j}, \dots, u_{1,k}$ for $j \in [k]$ and w_j is adjacent to $u_{2,k}, \dots, u_{2,(\ell-j+1)}$ for $\ell' - k < j \leq \ell'$. Together with $u_1, \dots, u_k, v_1, \dots, v_k$, this will yield a copy of P_ℓ^k which starts in $u_1, \dots, u_k$ and ends in $v_1, \dots, v_k$ (in those orders), contains no element of A and has at most $(d+\varepsilon)n$ vertices not in $\bigcup \mathcal{T}'$, which will then complete our proof of (iii). Therefore, for suitable values of ℓ' it remains to find a vertex partition $\mathcal{X}$ of H and a family $\bar{\mathcal{X}}$ of subsets of $V(H)$ such that $\mathcal{X}$ is size-compatible with $\mathcal{V}$, $(H, \mathcal{X})$ is an R-partition, $\bar{\mathcal{X}}$ is an α-buffer for H and for each $j \in [k]$ we have $X_{1,j} = V_i$ and $X_{2,j} = V_h$ with i, h such that $w_j \in X_i$ and $w_{\ell'-j+1} \in X_h$.

We consider $\ell \in \left(3r^{k+1}, \frac{(1-d)(k+1)tn}{r}\right]$. Let S be a copy of K_{k+2} in the same K_{k+1}-component of R as the copies of K_{k+1} in $\mathcal{T}$ and let $Z_1, \dots, Z_{k+2}$ be the vertices of S. Fix a K_{k+1}-walk W_0 whose first copy of K_k is $X_{1,1} \dots X_{1,k}$ and whose last is in T_1, which is of minimal length. For each $i \in [t-1]$, fix a K_{k+1}-walk W_i whose first copy of K_k is in T_i and whose last is in T_{i+1}, which is of minimal length. Fix a K_{k+1}-walk W_t whose first copy of K_k is in T_t, whose last is $X_{2,1} \dots X_{2,k}$, which includes a copy of K_k from S and is one of minimal length satisfying these conditions. We have $|W_i| \leq \binom{r}{k}$ for $i \in [t-1] \cup \{0\}$ and $|W_t| \leq 2\binom{r}{k}$. Let W' be the K_{k+1}-walk obtained by concatenating $W_0, \dots, W_t$.

We construct the sequence $Q(W, \overrightarrow{U_{11} \dots U_{1k}})$ for any K_{k+1}-walk $W = (E_1, E_2, \dots)$ in R and any orientation $\overrightarrow{U_{11} \dots U_{1k}}$ of E_1, its first copy of K_k, identically to that in (i). Orient $X_{1,1} \dots X_{1,k}$ as $\overrightarrow{X_{1,1} \dots X_{1,k}}$. Construct $Q(W', \overrightarrow{X_{1,1} \dots X_{1,k}})$. Then, the first copy of K_k $U_{i1} \dots U_{ik}$ and the last copy of K_k $U'_{i1} \dots U'_{ik}$ of each W_i obtain orientations, say $\overrightarrow{U_{i1}\dots U_{ik}}$ and $\overrightarrow{U'_{i1}\dots U'_{ik}}$. Construct $Q(W_i,\overrightarrow{U_{i1}\dots U_{ik}})$ for $i\in[t]\cup\{0\}$. Clearly, there are sequences $\bar{Q}_i$ of vertices in T_i for $i\in[t]$, such that $Q(W',\overrightarrow{X_{1,1}\dots X_{1,k}})$ is the concatenation of

$$Q(W_0,\overrightarrow{X_{1,1}\dots X_{1,k}}),\bar{Q}_1,Q(W_1,\overrightarrow{U_{11}\dots U_{1k}}),\dots,\bar{Q}_t,Q(W_t,\overrightarrow{U_{t1}\dots U_{tk}}).$$

Define $f_i\equiv|\bar{Q}_i| \mod k+1$ for $i\in[t]$. Let Q^*_0 denote $Q(W_0,\overrightarrow{X_{1,1}\dots X_{1,k}})$ and let Q^*_i denote $Q(W_i,\overrightarrow{U_{i1}\dots U_{ik}})$ for $i\in[t]$. Define $q_i:=|Q^*_i|$ for $i\in[t]\cup\{0\}$. For a sequence Q of vertices of R, let $(Q)_h$ denote the hth term of Q. For each $i\in[t]$ let $T_i=Y_{i1}\dots Y_{i(k+1)}$ be such that $\overrightarrow{Y_{i2}\dots Y_{i(k+1)}}$ is the oriented last copy of K_k of W_{i-1} in Q^*_{i-1}.

Let $\alpha:=q_0+\sum_{i=1}^t(q_i+f_i)$. Pick $x\in[k]\cup\{0\}$ such that $\ell'-\alpha\equiv x \mod k+1$. Let $Z_3,\dots,Z_{k+2}$ be the first k consecutive terms of Q^*_t which correspond to a copy of K_k in S. Define Q' as the result of inserting x copies of $Z_1,\dots,Z_{k+2}$ into Q^*_t right after the first occurrence of $Z_3,\dots,Z_{k+2}$ in Q^*_t. Let $q':=|Q'|$. Define $\alpha_x:=q_0+\sum_{i=1}^{t-1}(q_i+f_i)+f_t+q'=\alpha+x(k+2)$. Define the following.

$$p_0:=\max\left\{p\in\mathbb{Z}\mid \ell'\geq p(1-d)(k+1)\frac{n}{r}+\alpha_x\right\},$$

$$t_i=\begin{cases}(1-d)\frac{n}{r} & \text{if } i\in[p_0]\\ \frac{\ell'-\alpha_x}{k+1}-p_0(1-d)\frac{n}{r} & \text{if } i=p_0+1\\ 0 & \text{if } i>p_0+1,\end{cases}$$

$$L_0=q_0, L_j=q_0+\sum_{i=1}^{j}[t_i(k+1)+q_i+f_i] \text{ for } j\in[t-1],$$

$$M_0=0, M_j=L_{j-1}+t_j(k+1)+f_j \text{ for } j\in[t].$$

Define $\phi:V(H)\to V(R)$ as follows. For $i\leq\ell'$, set

$$\phi(v_i)=\begin{cases}Y_{jh} & \text{if } L_{j-1}<i\leq M_j, \text{ with } h\equiv i-L_{j-1} \mod k+1\\ (Q^*_j)_{i-M_j} & \text{if } M_j<i\leq L_j\\ (Q')_{i-M_t} & \text{if } M_t<i\leq\ell'.\end{cases}$$

For $i>\ell'$, given $\phi(v_1),\dots,\phi(v_{i-1})$, set $\phi(v_i)=V_j$ with $j=\min\{h:|\{b<i:\phi(v_b)=V_h\}|<|V_h|\}$.

Set $X_i:=\phi^{-1}(V_i)$, $\bar{X}_i:=\phi^{-1}(V_i)\setminus P$ for $i\in[r]$. Define $\mathcal{X}:=\{X_i\}_{i\in[r]}$ and $\bar{\mathcal{X}}:=\{\bar{X}_i\}_{i\in[r]}$. Since all edges in H have pairs of vertices in P at most k apart in the path order as endpoints and any $k+1$ consecutive vertices in the cyclic order are mapped to a copy of K_{k+1} in R, it follows that $(H, \mathcal{X})$ is an R-partition. Furthermore, for each $i \in [r]$ at most $(1-d)\frac{n}{r} + \frac{3r\binom{r}{k}+k(k+2)}{k+1} \le (1-d+\varepsilon)\frac{n}{r} \le (1-2\varepsilon)\frac{n}{r} \le |V_i|$ vertices in P are mapped to V_i, so $\mathcal{X}$ is a vertex partition of H which is size-compatible with $\mathcal{V}$. Finally, $\bar{X}_i$ is a set of isolated vertices in H by definition and

$$|\bar{X}_i| = |X_i| - |P \cap X_i| \ge \left(1 - \frac{1-d+\varepsilon}{1-3\varepsilon}\right)|X_i| \ge \alpha|X_i|$$

for each $i \in [r]$, so $\bar{\mathcal{X}}$ is an α-buffer for H. This completes the proof for (iii). $\square$

2.8 Concluding Remarks

Extremal graphs and minimum degree Our proofs provide a template for checking that $G_p(k,n,\delta)$ and $G_c(k,n,\delta)$ are the only extremal graphs up to some trivial modifications. We believe that the graph $G_p(k,n,\delta)$ remains extremal for kth powers of paths for all $\delta > \frac{(k-1)n}{k}$. However, the same is generally not true for $G_c(k,n,\delta)$ and kth powers of cycles: Allen, Böttcher and Hladký [5] sketched a construction, for infinitely many values of n, of graphs G on n vertices with $\delta(G) \ge \frac{n}{2} + \frac{\sqrt{n}}{5}$ which do not contain a copy of C_6^2. Their construction can be generalised to one for general powers of cycles.

Long kth powers of cycles Theorem 1.5(ii) states that if G does not contain any of various kth powers of cycles of lengths not divisible by $k+1$, then G must contain kth powers of cycles of every length divisible by $k+1$ up to $(k+1)(k\delta-(k-1)n)-\nu n$. We believe that the error term of νn can be removed, but it would involve significantly more technical work. This includes a new version of the Stability Lemma with more extremal cases and new extremal results corresponding to these additional extremal cases.

Chapter 3

Longest Paths in Random Hypergraphs

3.1 Introduction

In this chapter we prove a result for j-tight paths in random hypergraphs which is closely related to Theorem 1.6.

3.1.1 Main Theorem

The main result of this chapter is a phase transition result for j-tight paths in $H^k(n,p)$ that is very similar to Theorem 1.6. We write $x \ll y$ to mean that $x \leq y/C$ for some sufficiently large constant C and similarly $x \gg y$ to mean that $x \geq Cy$ for some sufficiently large constant C.

Theorem 3.1 (Cooley, Garbe, Hng, Kang, Sanhueza-Matamala and Zalla [16])**.** *Let $k, j \in \mathbb{N}$ satisfy $j < k$. Let $a \in \mathbb{N}$ be the unique integer satisfying $1 \leq a \leq k-j$ and $a \equiv k \bmod (k-j)$. Let $\varepsilon = \varepsilon(n) \ll 1$ satisfy $\varepsilon^3 n \xrightarrow{n\to\infty} \infty$ and let*

$$p_0 = p_0(n; k, j) := \frac{1}{\binom{k-j}{a}\binom{n-j}{k-j}}.$$

Let L be the length of the longest j-tight path in $H^k(n,p)$.

(i) *If $p = (1-\varepsilon)p_0$, then whp*

$$\frac{j\ln n - \omega + 3\ln\varepsilon}{-\ln(1-\varepsilon)} \leq L \leq \frac{j\ln n + \omega}{-\ln(1-\varepsilon)},$$

for any $\omega = \omega(n)$ such that $\omega \xrightarrow{n\to\infty} \infty$.

(ii) *If $p = (1+\varepsilon)p_0$ and $j \geq 2$, then for any δ satisfying $\delta \gg \max\{\varepsilon, \frac{\ln n}{\varepsilon^2 n}\}$, whp*

$$(1-\delta)\frac{\varepsilon n}{(k-j)^2} \leq L \leq (1+\delta)\frac{2\varepsilon n}{(k-j)^2}.$$

(iii) *If $p = (1+\varepsilon)p_0$ and $j = 1$, then for all $\delta \gg \varepsilon$ satisfying $\delta^2\varepsilon^3 n \xrightarrow{n\to\infty} \infty$, whp*

$$(1-\delta)\frac{\varepsilon^2 n}{4(k-1)^2} \leq L \leq (1+\delta)\frac{2\varepsilon n}{(k-1)^2}.$$

In other words, we have a phase transition at threshold p_0.

We will prove the upper bounds in all three cases using the first moment method. The lower bound in the subcritical case, i.e. in (i), will be proved using the second moment method—while the strategy is standard, there are significant technical complications to be overcome. However, the second moment method is not strong enough in the *supercritical* cases, and therefore we will prove the lower bounds in (ii) and (iii) by introducing the `Pathfinder` search algorithm which explores j-tight paths in k-uniform hypergraphs, and which is the main contribution of this chapter. The algorithm is based on a depth-first search process, but it is a rather delicate task to design it in such a way that it both correctly constructs j-tight paths and also admits reasonable probabilistic analysis. We will analyse the likely evolution of this algorithm and prove that whp it discovers a j-tight path of the appropriate length.

To help interpret Theorem 3.1, let us first observe that the results become stronger for smaller δ, so δ may be thought of as an error term. Furthermore, in all cases of the theorem we may choose δ to be no larger than an arbitrarily small constant, while in some cases we may even have $\delta \to 0$. In the subcritical regime (Theorem 3.1(i)), note that $-\ln(1-\varepsilon) = \varepsilon + O(\varepsilon^2)$ and that the term $3\ln\varepsilon$ in the lower bound becomes negligible (and in particular could be incorporated into ω) if ε is constant. For smaller ε, however, it represents a gap between the lower and upper bounds. In the supercritical case for $j \geq 2$ (Theorem 3.1(ii)), the length L is certainly of order $\Theta(\varepsilon n)$, but the lower and upper bounds differ by approximately a multiplicative factor of 2. In the supercritical case for $j = 1$ (Theorem 3.1(iii)), the lower and upper bounds differ by a multiplicative factor of $\Theta(\varepsilon)$. This has subsequently been improved by Cooley, Kang and Zalla [19], who lowered the upper bound to within a constant of the lower bound by analysing a structure similar to the 2-core in random hypergraphs. We will discuss all of these bounds and how they might be improved in more detail in Section 3.10.

Remark 3.2. In fact, the statement of Theorem 3.1 has been slightly weakened compared to what we actually prove in order to improve the clarity. More precisely, the full strength of the assumption on δ in (iii) is only required for the lower bound; the upper bound would in fact hold for any $\delta \gg \max\{\varepsilon, \frac{\ln n}{\varepsilon^2 n}\}$ as in (ii) (c.f. Lemma 3.35). Furthermore, the assumption that $\delta \gg \frac{\ln n}{\varepsilon^2 n}$ in (ii) is only needed for the upper bound; the lower bound holds with just the assumption that $\delta \gg \varepsilon$ (c.f. Lemma 3.30).

3.1.2 Related Work

The study of j-tight paths (and the corresponding notion of j-tight cycles) has been a central theme in hypergraph theory, with many generalisations of classical graph results, including Dirac-type and Ramsey-type (see [42, 46, 59] for surveys), as well as Erdős-Gallai-type results [2, 29].

There has also been some work on j-tight cycles in random hypergraphs. Dudek and Frieze [21, 22] determined the thresholds for the appearance of both loose and tight Hamilton cycles in $H^k(n, p)$, as well as determining the threshold for a j-tight Hamilton cycle up to a multiplicative constant. Recently, Narayanan and Schacht [47] pinpointed the precise value of the sharp threshold for the appearance of j-tight Hamilton cycles in k-uniform hypergraphs, provided that $k > j > 1$.

Theorem 3.1 addresses a range when p is significantly smaller than the threshold for a j-tight Hamilton cycle, and consequently the longest j-tight paths are far shorter. Recently Cooley [14] has extended the lower bound in Theorem 3.1(ii) to the range when $p = cp_0$ for some constant $c > 1$, and shown that with a much more difficult version of the common "sprinkling" argument, one can also find a j-tight *cycle* of approximately the same length.

Recall that for random graphs, the phase transition thresholds for the length of the longest path and the order of the largest component are both $1/n$. It is therefore natural to wonder whether something similar holds for j-tight paths in random hypergraphs, since for each $1 \leq j \leq k-1$, there is a notion of connectedness that is closely related to j-tight paths: two j-tuples J_1, J_2 of vertices are j-tuple-connected if there is a sequence of edges $e_1, \ldots, e_\ell$ such that $J_1 \subset e_1$ and $J_2 \subset e_\ell$, and furthermore any two consecutive edges e_i, e_{i+1} intersect in at least j vertices. A j-tuple component is a maximal collection of pairwise j-tuple-connected j-sets.

The threshold for the emergence of the giant j-tuple component in $H^k(n, p)$ is known to be

$$p_g = p_g(n; k, j) = \frac{1}{\left(\binom{k}{j} - 1\right)\binom{n-j}{k-j}}.$$

The case $k = 2$ and $j = 1$ is the classical graph result of Erdős and Rényi. The case $j = 1$ for general k was first proved by Schmidt-Pruzan and Shamir [53]. The case of general k and j was first proved by Cooley, Kang, and Person [18].

One might expect the threshold for the emergence of a j-tight path of linear length to have the same threshold. However, it turns out that this is only true in the case when $j = 1$. More precisely, in the case $j = 1$, the probability threshold of $\frac{1}{(k-1)\binom{n-j}{k-j}}$ given by Theorem 3.1 matches the threshold for the emergence of the giant (vertex-)component. However, for $j \geq 2$, the two thresholds do not match. A heuristic explanation for this is that when exploring a j-tuple component via a (breadth-first or depth-first) search process, each time we find an edge we may continue exploring a *j-tuple component* from any of the $\binom{k}{j} - 1$ new j-sets within this edge (all are new except the j-set from which we first found the edge). However, when exploring a *j-tight path*, the restrictions on the structure mean that not all j-sets within the edge may form the last j vertices of the path. For a as defined in Theorem 3.1, it will turn out that we only have $\binom{k-j}{a}$ choices for the j-set from which to continue the path (this will be explained in more detail in Section 3.4.2).

3.1.3 Overview

The remainder of the chapter is arranged as follows. In Section 3.2, we will analyse the structure of j-tight paths and prove some preliminary results concerning the number of automorphisms, which will be needed later. We also collect some standard probabilistic results which we will use. Subsequently, Section 3.3 will be devoted to a second moment calculation, which will be used to prove the lower bound on L in the subcritical case of Theorem 3.1. This is in essence a very standard method, although this particular application presents considerable technical challenges.

The second moment method breaks down when the paths become too long, and in particular it is too weak to prove the lower bounds in the supercritical case. Therefore the main contribution of this chapter is an alternative strategy, inspired by previous proofs of phase transition results regarding the order of the giant component. These proofs, due to Krivelevich and Sudakov [41] as well as Cooley, Kang, and Person [18] and Cooley, Kang, and Koch [17], are based on an analysis of search processes which explore components.

We therefore introduce the `Pathfinder` algorithm, which is in essence a depth-first search process for paths, in Section 3.4. In Section 3.5, we observe some basic facts about the `Pathfinder` algorithm, which we subsequently use in Section 3.6 ($j = 1$) and Section 3.7 ($j \geq 2$) to prove that whp the `Pathfinder` algorithm finds a j-tight path of the appropriate length, proving the lower bounds on L in the supercritical case of Theorem 3.1.

We collect together all of the previous results to complete the proof of Theorem 3.1 in Section 3.9. Finally in Section 3.10 we discuss some open problems, including possible strengthenings of Theorem 3.1.

3.2 Preliminaries

We first gather some notation and terminology which we will use throughout the chapter. Throughout this chapter, k and j are fixed integers with $1 \leq j \leq k-1$. All asymptotics are with respect to n, and we use the standard Landau notations $o(\cdot)$, $O(\cdot), \Theta(\cdot), \Omega(\cdot)$ with respect to these asymptotics. In particular, any value which is bounded by a function of k and j is $O(1)$. For $m, i \in \mathbb{N}$, we use $(m)_i := m(m-1)\dots(m-i+1)$ to denote the i-th falling factorial.

Recall that for $\ell \in \mathbb{N}$, a j-tight path of length ℓ in a k-uniform hypergraph contains ℓ edges and $(k-j)\ell + j$ vertices. Throughout the chapter, whenever j, k, ℓ are clear from the context, we will denote by

$$v = v_{j,k}(\ell) := (k-j)\ell + j \tag{3.1}$$

the number of vertices in such a path. Furthermore, for the rest of the chapter we fix a as in Theorem 3.1, i.e. a is the unique integer such that

$$1 \leq a \leq k-j \qquad \text{and} \qquad a \equiv k \pmod{k-j} \tag{3.2}$$

and we set

$$b := k - j - a. \tag{3.3}$$

Throughout the chapter we ignore floors and ceilings whenever these do not significantly affect the argument.

3.2.1 Structure of j-tight Paths

For $\ell \in \mathbb{N}$, let $\mathcal{P}_\ell$ be the set of all j-tight paths of length ℓ in the complete k-uniform hypergraph on $[n]$, denoted by $K_n^{(k)}$. Thus $\mathcal{P}_\ell$ is the set of *potential* j-tight paths of length ℓ in $H^k(n,p)$.

It is important to observe that, depending on the values of k and j, the presence of one j-tight path $P \in \mathcal{P}_\ell$ in $H^k(n,p)$ may instantly imply the presence of many more with exactly the same edge set. In the graph case, there are only two paths with exactly the same edge set (we obtain the second by reversing the orientation), but for general k and j there may be more.

Let us demonstrate this with the following example for the case $k = 5$ and $j = 2$ (see Figure 3.1).

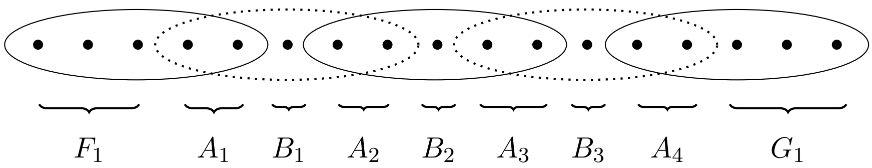


Figure 3.1: A 2-tight path of length 5 in a 5-uniform hypergraph, with a natural partition of vertices.

Observe that we have partitioned the vertices into sets $(F_1, A_1, \ldots)$ according to which edges they are in—each set of the partition is maximal with the property that every vertex in that set is in exactly the same edges of the j-tight path. Therefore we can re-order the vertices arbitrarily within any of these sets and obtain another j-tight path with the same edge set, and therefore also the same length. Similarly as for graphs, we can also reverse the orientation of the vertices (and also the edges) to obtain another j-tight path with the same edge set.

It will often be convenient to consider such paths as being the same, even though the order of vertices is different. Therefore we define an equivalence relation $\sim_\ell$ on $\mathcal{P}_\ell$ as follows. For any $A, B \in \mathcal{P}_\ell$, we say that $A \sim_\ell B$ if they have exactly the same edges.

We will be interested in the equivalence classes of this relation. Let $z_\ell = z_\ell(k,j)$ denote the size of each equivalence class of $\sim_\ell$ (note that, by symmetry, each equivalence class has the same size and so z_ℓ is well-defined). Further, let $\hat{\mathcal{P}}_\ell$ be the set of equivalence classes of $\sim_\ell$. Observe that if some $P \in \mathcal{P}_\ell$ is in $H^k(n,p)$, then so is every path in its equivalence class $\hat{P} \in \hat{\mathcal{P}}_\ell$. We abuse terminology slightly by saying that the equivalence class $\hat{P}$ lies in $H^k(n,p)$, and write $\hat{P} \subset H^k(n,p)$. We define $\hat{X}_\ell$ to be the number of equivalence classes for which this is the case. Then

$$\mathbb{E}(\hat{X}_\ell) = \sum_{\hat{P} \in \hat{\mathcal{P}}_\ell} \mathbb{P}\left(\hat{P} \subset H^k(n,p)\right) = |\hat{\mathcal{P}}_\ell| p^\ell = \frac{(n)_v}{z_\ell} p^\ell, \tag{3.4}$$

where $v = (k-j)\ell + j$ is the number of vertices in a j-tight path with ℓ edges (as defined in (3.1)).

We therefore need to estimate z_ℓ. To do so, we will analyse the structure of j-tight paths, inspired by the example in Figure 3.1. This analysis leads to the following lemma.

Lemma 3.3. *Let $s = s(j,k) := \left\lceil \frac{k}{k-j} \right\rceil - 1$. Then*

$$z_\ell = \begin{cases} \Theta(1) & \text{if } \ell \le s+1; \\ \frac{2}{b!}(a!b!)^{\ell-s}((k-j)!)^{2s} & \text{if } \ell \ge s+2. \end{cases}$$

In particular,

$$z_\ell = \Theta\left((a!b!)^\ell\right). \tag{3.5}$$

Proof. Let us first observe that if $\ell \le s+1$, then a j-tight path with ℓ edges has v vertices, where

$$v = (k-j)\ell + j \le k(\ell+1) \le k(s+2) = O(1),$$

and therefore $1 \le z_\ell \le v! = O(1)$, and the statement of the lemma follows for this case. We therefore assume that $\ell \ge s+2$.

We aim to determine the natural partition of the vertices of a j-tight path according to which edges they are in, as we did in the example in Figure 3.1.

Denote the edges of the j-tight path $P \in \mathcal{P}_\ell$ by $(e_1, \ldots, e_\ell)$, in the natural order. Recall that $s = \lceil \frac{k}{k-j} \rceil - 1$, and observe that s is the largest integer such that $(k-j)s < k$, and therefore the largest integer such that $e_i \cap e_{i+s} \neq \emptyset$. We define

$$\begin{aligned} F_i &:= e_i \setminus e_{i+1} && \text{for } 1 \le i \le s; \\ G_i &:= e_{\ell-s+i} \setminus e_{\ell-s+i-1} && \text{for } 1 \le i \le s. \end{aligned}$$

We also define

$$\begin{aligned} A_i &:= e_i \cap e_{i+s} && \text{for } 1 \le i \le \ell - s, \\ B_i &:= e_{i+s} \backslash (e_{i+s+1} \cup e_i) && \text{for } 1 \le i \le \ell - s - 1. \end{aligned}$$

Observe that $A_i \cup B_i = e_{i+s} \setminus e_{i+s+1}$. Furthermore, since s is the largest integer such that e_{i+s+1} intersects e_{i+1}, we have that $(e_{i+s} \backslash e_{i+s+1}) \subset e_{i+1}$ and that $A_{i+1} \subseteq (e_{i+1} \backslash e_i)$, and therefore $A_{i+1} \cup B_i = e_{i+1} \setminus e_i$. Since we also have $A_i \cap B_i = A_{i+1} \cap B_i = \emptyset$, the vertices of the path P are now partitioned into parts

$$(F_1, \dots, F_s, A_1, B_1, A_2, B_2, \dots, A_{\ell-s-1}, B_{\ell-s-1}, A_{\ell-s}, G_1, \dots, G_s)$$

(in the natural order along P). Observe further that the parts are of maximal size such that the vertices within each part are in exactly the same edges. We refer to $\bigcup_{i=1}^s F_i = e_1 \setminus e_{s+1}$ as the *head* of the path P and to $\bigcup_{i=1}^s G_i = e_\ell \setminus e_{\ell-s}$ as the *tail*. Note that the vertices within each part can be rearranged to obtain a new j-tight path with exactly the same edges. We can also change the orientation of the path (i.e. reverse the order of the edges) to obtain a new path with the same edge set. (If $\ell = 0, 1$, this reorientation would already have been counted, but recall that we have assumed that $\ell \ge s + 2$.) Thus we have

$$z_\ell = 2 \left(\prod_{i=1}^{s} |F_i|!|G_i|! \right) \left(\prod_{i=1}^{\ell-s} |A_i|! \right) \left(\prod_{i=1}^{\ell-s-1} |B_i|! \right). \tag{3.6}$$

It therefore remains to determine the sizes of the F_i, G_i, A_i, B_i.

Claim 3.4.

$$\begin{aligned} |F_i| = |G_i| &= k - j && \textit{for } 1 \le i \le s; \\ |A_i| &= a && \textit{for } 1 \le i \le \ell - s; \\ |B_i| &= b && \textit{for } 1 \le i \le \ell - s - 1. \end{aligned}$$

Proof. We certainly have

$$|F_i| = |e_i \setminus e_{i+1}| = |e_i| - |e_i \cap e_{i+1}| = k - j,$$

and similarly $|G_i| = k - j$. Furthermore,

$$|A_i| = |e_i \cap e_{i+s}| = k - s(k-j) = k - \left(\left\lceil \frac{k}{k-j} \right\rceil - 1 \right) (k-j),$$

so we have $1 \le |A_i| \le k-j$ and $|A_i| \equiv k \mod k-j$, which recall from (3.2) was precisely the definition of a, so $|A_i| = a$. Finally, observe that $A_i \cup B_i = e_{i+s} \setminus e_{i+s-1}$, and so

$$|B_i| = k - j - |A_i| = k - j - a \stackrel{(3.3)}{=} b,$$

as required. ■

Substituting the values from Claim 3.4 into (3.6), we obtain precisely the statement of Lemma 3.3. Thus the proof is complete. □

Equation (3.4) and Lemma 3.3 together give the following immediate corollary.

Corollary 3.5.

$$\mathbb{E}(\hat{X}_\ell) = \Theta(1)\frac{(n)_v}{(a!b!)^\ell}p^\ell.$$

3.2.2 Large Deviation Bounds

In this section we collect some standard results which will be needed later.

We will use the following Chernoff bound, (see e.g. [32, Theorem 2.1]). We use $\mathrm{Bin}(N,p)$ to denote the binomial distribution with parameters $N \in \mathbb{N}$ and $p \in [0,1]$.

Lemma 3.6. *If $X \sim \mathrm{Bin}(N,p)$, then for any $\xi \ge 0$*

$$\mathbb{P}(X \ge Np + \xi) \le \exp\left(-\frac{\xi^2}{2(Np + \frac{\xi}{3})}\right), \tag{3.7}$$

and

$$\mathbb{P}(X \le Np - \xi) \le \exp\left(-\frac{\xi^2}{2Np}\right).$$

It will often be more convenient to use the following one-sided form, which follows directly from Lemma 3.6.

Lemma 3.7. *Let $X \sim \mathrm{Bin}(N,p)$ and let $\alpha > 0$ be some arbitrarily small constant. Then with probability at least $1 - \exp(-\Theta(n^\alpha))$ we have $X \le 2Np + n^\alpha$.*

Proof. We distinguish two cases.

Case 1: $Np > n^\alpha$ By applying (3.7) with $\xi = Np$, we obtain

$$\mathbb{P}(X \ge 2Np + n^\alpha) \le \mathbb{P}(X \ge 2Np) \stackrel{(3.7)}{\le} \exp\left(-\frac{(Np)^2}{\frac{8}{3}Np}\right) \le \exp(-\Theta(n^\alpha)),$$

as required.

Case 2: $Np \le n^\alpha$ By applying (3.7) with $\xi = n^\alpha$, we obtain

$$\mathbb{P}(X \ge 2Np + n^\alpha) \le \mathbb{P}(X \ge Np + n^\alpha) \overset{(3.7)}{\le} \exp\left(-\frac{n^{2\alpha}}{2(n^\alpha + n^\alpha/3)}\right) = \exp(-\Theta(n^\alpha)),$$

which proves the assertion in this case. $\square$

3.3 Second Moment Method: Lower Bound

In this section we prove the lower bound in statement (i) of Theorem 3.1. The general basis of the argument is a completely standard second moment method — however, applying the method to this particular problem is rather tricky and so the argument is lengthy. For technical reasons that will become apparent during the proof, we need to handle the case when $2 \le j = k-1$ slightly differently. We therefore distinguish two cases:

- Case 1: Either $j \le k-2$ or $j = k-1 = 1$.
- Case 2: $2 \le j = k-1$.

Correspondingly, we split the lower bound we aim to prove into two lemmas. In Case 1, we need to prove the following.

Lemma 3.8. *Let $k, j \in \mathbb{N}$ satisfy $1 \le j \le k-1$, and additionally either $j \le k-2$ or $j = k-1 = 1$. Let $a \in \mathbb{N}$ be the unique integer satisfying $1 \le a \le k-j$ and $a \equiv k \bmod (k-j)$. Let $\varepsilon = \varepsilon(n) \ll 1$ satisfy $\varepsilon^3 n \xrightarrow{n\to\infty} \infty$ and let*

$$p = \frac{1-\varepsilon}{\binom{k-j}{a}\binom{n-j}{k-j}}.$$

Let L be the length of the longest j-tight path in $H^k(n,p)$. Then whp

$$L \ge \frac{j \ln n - \omega + 3\ln\varepsilon}{-\ln(1-\varepsilon)},$$

for any $\omega = \omega(n)$ such that $\omega \xrightarrow{n\to\infty} \infty$.

On the other hand, in Case 2 we have $k - j = 1$, and therefore the parameter a from Theorem 3.1 is simply 1. Thus also $\binom{k-j}{a} = 1$ and $p_0 = \frac{1}{n-k+1}$, and so the lower bound in Theorem 3.1 (i) simplifies to the following.

Lemma 3.9. *Let $k, j \in \mathbb{N}$ satisfy $2 \leq j = k-1$. Let $\varepsilon = \varepsilon(n) \ll 1$ satisfy $\varepsilon^3 n \xrightarrow{n\to\infty} \infty$ and let*

$$p = \frac{1-\varepsilon}{n-k+1}.$$

Let L be the length of the longest j-tight path in $H^k(n,p)$. Then whp

$$L \geq \frac{j \ln n - \omega + 3\ln\varepsilon}{-\ln(1-\varepsilon)},$$

for any $\omega = \omega(n)$ such that $\omega \xrightarrow{n\to\infty} \infty$.

Since the main ideas in the proofs of these two lemmas are essentially identical, we will treat Case 1 (i.e. Lemma 3.8) more carefully and address Case 2 (i.e. Lemma 3.9) in Section 3.3.2 by illustrating how to adapt the proof of Case 1.

3.3.1 Case 1: either $j \leq k-2$ or $j = k-1 = 1$

We will prove Lemma 3.8 with the help of various auxiliary results. Since these results are rather technical in nature, we defer their proofs to the end of this subsection.

Let us set $p = \frac{1-\varepsilon}{\binom{k-j}{a}\binom{n-j}{k-j}}$ and $\ell = \frac{j\ln n - \omega + 3\ln\varepsilon}{-\ln(1-\varepsilon)}$. Recall that $\mathcal{P}_\ell$ is the set of all j-tight paths of length ℓ in $K_n^{(k)}$ and therefore

$$\mathbb{E}(X_\ell^2) = \sum_{A,B\in\mathcal{P}_\ell} \mathbb{P}(A, B \subset H^k(n,p)).$$

The probability term in the sum is fundamentally dependent on how many edges the paths A and B share, so we will need to calculate the number of pairs of possible paths with given intersections.

For any $A, B \in \mathcal{P}_\ell$, let $Q(A,B)$ be the set of common edges of A and B and define $q(A,B) := |Q(A,B)|$. Observe that there is a natural partition of $Q(A,B)$ into *intervals*, where each interval is a maximal set of edges in $Q(A,B)$ which are consecutive along both A and B. Let $r(A,B)$ be the number of intervals in this natural partition of $Q(A,B)$. Set $\mathbf{c}(A,B) := (c_1,\dots,c_r)$, where $c_1 \geq \cdots \geq c_r \geq 1$, to be the lengths (i.e. the number of edges) of these intervals. Given non-negative integers q, r and an r-tuple $\mathbf{c} = (c_1,\dots,c_r)$ such that $c_1 \geq \cdots \geq c_r \geq 1$ and $c_1 + \cdots + c_r = q$, define

$$\mathcal{P}_\ell^2(q,r,\mathbf{c}) := \{(A,B) \in \mathcal{P}_\ell^2 : \ q(A,B) = q, \ r(A,B) = r, \ \mathbf{c}(A,B) = \mathbf{c}\}$$

For any $q, r, \mathbf{c}$ not satisfying these conditions, $\mathcal{P}_\ell^2(q,r,\mathbf{c})$ is empty. Recall from (3.1) that $v = (k-j)\ell + j$ is the number of vertices in a j-tight path of length ℓ.

Claim 3.10.

$$\mathbb{E}(X_\ell^2) \le ((n)_v)^2\, p^{2\ell} + \sum_{q\ge 1}\sum_{r\ge 1}\sum_{\mathbf{c}} |\mathcal{P}_\ell^2(q,r,\mathbf{c})| p^{2\ell-q}. \tag{3.8}$$

Thus we need to estimate $|\mathcal{P}_\ell^2(q,r,\mathbf{c})|$ for $q,r\ge 1$. Given $q,r\ge 1$, we define the parameter

$$T(r) = T_q(r) := (k-j)q + j + (r-1)\min\{j,k-j\}.$$

This slightly arbitrary-looking expression is in fact a lower bound on the number of vertices in $Q(A,B)$, as will become clear in the proof. We obtain the following.

Proposition 3.11. *There exists a constant $C>0$ such that for any $q\ge 1$ we have*

$$|\mathcal{P}_\ell^2(q,r,\mathbf{c})| \le ((n)_v)^2\, \frac{(\ell-q+1)^2 \ell^{2(r-1)}(a!b!)^q C^r}{(n-v)^{T(r)}}.$$

Proposition 3.11 together with (3.8) gives the following immediate corollary.

Corollary 3.12. *There exists a constant $C>0$ such that*

$$\mathbb{E}(X_\ell^2) \le ((n)_v)^2\, p^{2\ell}\left(1+\sum_{q=1}^{\ell}\sum_{r=1}^{q}\sum_{\substack{c_1+\cdots+c_r=q\\ c_1\ge\cdots\ge c_r\ge 1}} \frac{(\ell-q+1)^2\ell^{2(r-1)}(a!b!)^q C^r}{p^q (n-v)^{T(r)}}\right). \tag{3.9}$$

We bound the triple-sum using the following two results.

Proposition 3.13.

$$\sum_{r=1}^{q}\sum_{\substack{c_1+\cdots+c_r=q\\ c_1\ge\cdots\ge c_r\ge 1}} \frac{(\ell-q+1)^2\ell^{2(r-1)}(a!b!)^q C^r}{p^q (n-v)^{T(r)}} = O\left(n^{-j}\right)\frac{(\ell-q+1)^2}{(1-\varepsilon)^q}. \tag{3.10}$$

Claim 3.14.

$$\sum_{q=1}^{\ell} \frac{(\ell-q+1)^2}{(1-\varepsilon)^q} = \frac{2(1-\varepsilon)^{-\ell}}{\varepsilon^3}. \tag{3.11}$$

Substituting (3.10) and (3.11) into (3.9), using the fact that $\ell = \frac{j\ln n - \omega + 3\ln\varepsilon}{-\ln(1-\varepsilon)}$ and performing some elementary approximations leads to the following.

Claim 3.15. $\mathbb{E}(X_\ell^2) = ((n)_v)^2\, p^{2\ell}(1+o(1))$.

We can now use these auxiliary results to prove our lower bound.

Proof of Lemma 3.8. Recalling that $\mathcal{P}_\ell$ is the set of all possible j-tight paths of length ℓ in $H^k(n,p)$, clearly $\mathbb{E}(X_\ell) = |\mathcal{P}_\ell|p^\ell = (n)_v p^\ell$. Therefore by Claim 3.15, we have

$$\mathbb{E}(X_\ell^2) = \mathbb{E}(X_\ell)^2(1+o(1)),$$

and a standard application of Chebyshev's inequality shows that whp $X_\ell \geq 1$, i.e. whp

$$L(G(n,p)) \geq \ell = \frac{j\ln n - \omega + 3\ln\varepsilon}{-\ln(1-\varepsilon)}$$

as claimed. □

It would be tempting to try to generalise this proof to also prove a lower bound in the supercritical case. However, this strategy fails because as the paths A and B become longer, there are many more ways in which they can intersect each other, and therefore the terms which, in the subcritical case, were negligible lower order terms (i.e. $q \geq 1$) become more significant. We will therefore use an entirely different strategy for the supercritical case.

Now we prove the auxiliary results required for the proof of Lemma 3.8, i.e. the second moment method for the case when $j \leq k-2$ or $j = k-1 = 1$.

Proof of Claim 3.10. Observe that

$$\mathbb{E}(X_\ell^2) = \sum_{(A,B)\in\mathcal{P}_\ell^2} \mathbb{P}(A, B \subset H^k(n,p)) = \sum_{q,r,\mathbf{c}} |\mathcal{P}_\ell^2(q,r,\mathbf{c})|p^{2\ell-q}.$$

Furthermore, observe that in the case $q = 0$, we must have $r = 0$ and $\mathbf{c} = ()$ an empty sequence. In this case, we have

$$|\mathcal{P}_\ell^2(0,0,())| \leq (n)_v^2,$$

while for $q \geq 1$ clearly $\mathcal{P}_\ell^2(q,r,\mathbf{c})$ is empty unless also $r \geq 1$, and the result follows. □

Proof of Proposition 3.11. To estimate $|\mathcal{P}_\ell^2(q,r,\mathbf{c})|$, we will regard A and B as j-tight paths of length ℓ which must be *embedded* into $K_n^{(k)}$ subject to certain restrictions (so that the parameters $q, r, \mathbf{c}$ are correct), and estimate the number of ways of performing this embedding appropriately. We will denote the edges of A by $(e_1, \ldots, e_\ell)$ and the edges of B by $(f_1, \ldots, f_\ell)$, each in the natural order.

First we embed the path A; there are $(n)_v$ ways of choosing its vertices in order. Then we embed the path B subject to certain restrictions, since we must obtain the parameters $q, r, \mathbf{c}$. We first choose which of the edges f_i on B will lie in $Q(A,B)$—recall that the i-th interval must be of length c_i, and therefore must have the form $(f_{t_i}, \ldots, f_{t_i+c_i-1})$, for some $1 \leq t_i \leq \ell - c_i + 1$. Thus the i-th interval is determined by the choice of its first edge f_{t_i}. Having already chosen intervals of length $c_1, \ldots, c_{i-1}$, there are only $\ell - c_1 - c_2 - \cdots - c_{i-1}$ edges of B left, of which certainly the last $c_i - 1$ cannot be chosen for f_{t_i}, since then either the interval would intersect with another previously chosen interval, or it would extend beyond the end of B. Thus there are at most $\ell - c_1 - \cdots - c_i + 1$ possible choices for f_{t_i}. Subsequently, we choose which edges of A to embed this interval onto. The corresponding interval in A must have the form either

$$(e_{s_i}, \ldots, e_{s_i+c_i-1})$$

or

$$(e_{s_i}, \ldots, e_{s_i-c_i+1}),$$

depending on whether the orientation is with or against the direction on A. There are 2 choices for the orientation, and subsequently (arguing as for B) at most $\ell - c_1 - \cdots - c_i + 1$ choices for e_{s_i}.

Thus the number of ways of choosing where to embed the edges of $Q(A,B)$ is at most

$$\prod_{i=1}^{r} (\ell - c_1 - \cdots - c_i + 1) 2(\ell - c_1 - \cdots - c_i + 1) \leq 2^r (\ell - q + 1)^2 \ell^{2(r-1)}, \qquad (3.12)$$

where we have used the fact that $c_1 + \cdots + c_r = q$. Observe here that we may well have overcounted: for an interval of length 1, the factor of 2 is superfluous, since orientation makes no difference; furthermore, if $r > 1$, then having embedded the first interval is often more restrictive with respect to where the second may be embedded than we accounted for. However, this expression is certainly an upper bound.

Note also that for $i \leq r - 1$ we use the crude bound $\ell - c_1 - \cdots - c_i + 1 \leq \ell$, whereas we are more careful about c_r. The reason is that in the case $r = 1$ we will have to bound terms rather precisely, whereas for $r \geq 2$ we will have plenty of room to spare in the calculations.

We have now fixed how the edges of intervals in B are embedded onto intervals in A, but we also need to account for different ways of ordering the vertices in these intervals. Since the i-th interval forms a j-tight path of length c_i, there are z_{c_i} possible ways of re-ordering the vertices of B along it, but still embedding into A in a way consistent with the edge-assignment. This is true regardless of where the interval lies on A or B, even if it includes some of the head or tail of A or B.

One difficulty is that two different intervals may share vertices, and therefore not every re-ordering is admissible. However, we may certainly use z_{c_i} as an upper bound for each i. Thus by (3.5), the number of ways of choosing where to embed the vertices of B within edges of $Q(A, B)$ is at most

$$\prod_{i=1}^{r} z_{c_i} = \prod_{i=1}^{r} \Theta\left((a!b!)^{c_i}\right) = (a!b!)^q \Theta(1)^r. \tag{3.13}$$

We now need to bound the number of ways of embedding the remaining vertices of B, for which we need a *lower* bound on the number of vertices in edges of $Q(A, B)$, i.e. vertices of B which have already been embedded into A. Let us first consider a simple *upper* bound: the i-th interval contains $(k-j)c_i + j$ vertices, and so we have already embedded at most

$$\sum_{i=1}^{r}((k-j)c_i + j) = (k-j)q + rj \tag{3.14}$$

vertices, with equality if and only if no two intervals share a vertex. We find a *lower* bound by considering when the intervals share as many vertices as possible.

Let us first consider the intervals in their natural order along B. Then the number of vertices lying in two consecutive intervals is at most the size of the intersection of two *non-consecutive* edges $|e_i \cap e_{i+2}| = \max\{k - 2(k-j), 0\}$. Therefore the total number of vertices lying in more than one interval is at most

$$(r-1) \cdot \max\{k - 2(k-j), 0\} = (r-1)(j - \min\{j, k-j\}).$$

Thus using (3.14), the number of vertices already embedded is at least

$$T(r) = (k-j)q + j + (r-1)\min\{j, k-j\}.$$

Therefore we have at most $v - T(r)$ vertices of B still left to embed, for which there are at most

$$(n)_{v-T(r)} \leq \frac{(n)_v}{(n-v)^{T(r)}} \tag{3.15}$$

choices.

Combining (3.12), (3.13) and (3.15) with the fact that there were $(n)_v$ ways of embedding A, for $r \geq 1$ we obtain

$$\begin{aligned}|\mathcal{P}_\ell^2(q,r,\mathbf{c})| &\leq (n)_v 2^r (\ell-q+1)^2 \ell^{2(r-1)} (a!b!)^q \Theta(1)^r \frac{(n)_v}{(n-v)^{T(r)}} \\ &= (n)_v^2 \frac{(\ell-q+1)^2 \ell^{2(r-1)} (a!b!)^q \Theta(1)^r}{(n-v)^{T(r)}}\end{aligned}$$

as claimed. □

Proof of Proposition 3.13. It will turn out that for each q, the $r=1$ term is the most significant, so we will treat this case separately. We define the following functions for each positive integer q and $r \in [q]$.

$$y_q(r) := \sum_{\substack{c_1+\cdots+c_r=q \\ c_1\geq\cdots\geq c_r\geq 1}} 1 \quad \text{and} \quad x_q(r) := \frac{\ell^{2(r-1)} C^r}{(n-v)^{T(r)}} y_q(r).$$

We obtain

$$\sum_{r=1}^{q} \sum_{\substack{c_1+\cdots+c_r=q \\ c_1\geq\cdots\geq c_r\geq 1}} \frac{(\ell-q+1)^2 \ell^{2(r-1)} (a!b!)^q C^r}{p^q (n-v)^{T(r)}} = \frac{(a!b!)^q}{p^q} (\ell-q+1)^2 \sum_{r=1}^{q} x_q(r). \tag{3.16}$$

Observe that

$$y_q(r+1) \leq \sum_{c_{r+1}=1}^{q} y_{q-c_{r+1}}(r) \leq \sum_{c_{r+1}=1}^{q} y_q(r) \leq q \cdot y_q(r), \tag{3.17}$$

and

$$T(r+1) - T(r) = \min\{j, k-j\}, \tag{3.18}$$

so for $r \in [q]$ we have

$$\begin{aligned}\frac{x_q(r+1)}{x_q(r)} &= \frac{\ell^2 C}{(n-v)^{T(r+1)-T(r)}} \frac{y_q(r+1)}{y_q(r)} \\ &\leq \frac{\ell^2 C}{(n-v)^{\min\{j,k-j\}}} q = O\left(\frac{(\ln(\varepsilon^3 n^j))^3}{\varepsilon^3 n^{\min\{j,k-j\}}}\right),\end{aligned}$$

where we have used the fact that $q \leq \ell = O\left(\frac{\ln(\varepsilon^3 n^j)}{\varepsilon}\right)$. Now let us observe that in the case $j=1$, setting $\lambda := \varepsilon^3 n$ which tends to infinity by assumption, we have

$$\frac{x_q(r+1)}{x_q(r)} = O\left(\frac{(\ln\lambda)^3}{\lambda}\right) = O\left(\lambda^{-1/2}\right).$$

On the other hand, if $j \geq 2$, then (since we are in Case 1) we also have $j \leq k-2$, and so

$$\frac{x_q(r+1)}{x_q(r)} = O\left(\frac{(\ln n)^3}{\varepsilon^3 n^2}\right) = O\left(n^{-1/2}\right).$$

Setting

$$w := \begin{cases} \lambda^{1/2} & \text{if } j = 1, \\ n^{1/2} & \text{if } 2 \leq j \leq k-2, \end{cases}$$

we have $w \to \infty$ and $\frac{x_q(r+1)}{x_q(r)} = O(1/w)$ in all cases.[1] Therefore, we obtain

$$\sum_{r=1}^{q} x_q(r) = x_q(1)\left(1 + \sum_{i=1}^{q-1} O\left(\frac{1}{w}\right)^i\right) = \frac{Cy_q(1)}{(n-v)^{T(1)}} \cdot (1+o(1)) \leq \frac{2C}{n^{(k-j)q+j}}.$$

Substituting this upper bound into (3.16) gives

$$\begin{aligned}
\sum_{r=1}^{q} \sum_{\substack{c_1+\cdots+c_r=q \\ c_1 \geq \cdots \geq c_r \geq 1}} \frac{(\ell-q+1)^2 \ell^{2(r-1)} (a!b!)^q C^r}{p^q (n-v)^{T(r)}} &\leq \frac{(a!b!)^q}{p^q} (\ell-q+1)^2 \frac{2C}{n^{(k-j)q+j}} \\
&= O\left(n^{-j}\right)(\ell-q+1)^2 \left(\frac{a!b!}{pn^{k-j}}\right)^q \\
&= O\left(n^{-j}\right)\frac{(\ell-q+1)^2}{(1-\varepsilon)^q},
\end{aligned}$$

as required. $\square$

Proof of Claim 3.14. By a change of index $i = \ell - q$, we get

$$\begin{aligned}
\sum_{q=1}^{\ell} \frac{(\ell-q+1)^2}{(1-\varepsilon)^q} = \sum_{i=0}^{\ell-1} \frac{(i+1)^2}{(1-\varepsilon)^{\ell-i}} &\leq (1-\varepsilon)^{-\ell} \sum_{i=-2}^{\infty} (i+1)(i+2)(1-\varepsilon)^i \\
&\leq (1-\varepsilon)^{-\ell} \frac{\mathrm{d}^2}{\mathrm{d}\varepsilon^2}\left(\sum_{i=-2}^{\infty} (1-\varepsilon)^{i+2}\right) \\
&= (1-\varepsilon)^{-\ell} \frac{\mathrm{d}^2}{\mathrm{d}\varepsilon^2}\left(\frac{1}{\varepsilon}\right) \\
&= \frac{2(1-\varepsilon)^{-\ell}}{\varepsilon^3}
\end{aligned}$$

[1] Note that it is here that the argument fails for $2 \leq j = k-1$, since we would only obtain the bound

$$\frac{x_q(r+1)}{x_q(r)} = O\left(\frac{(\ln n)^3}{\varepsilon^3 n}\right),$$

and if ε is very small (i.e. $\varepsilon^3 n \to \infty$ very slowly), this may not tend to zero. If we were to assume the slightly stronger condition of $\frac{\varepsilon^3 n}{(\ln n)^3} \to \infty$ in Theorem 3.1, then this would not be an issue and we would not need to handle the case $2 \leq j = k-1$ separately.

as claimed. □

Proof of Claim 3.15. Using Proposition 3.13 and Claim 3.14, together with the fact that $\ell = \frac{j\ln n - \omega + 3\ln\varepsilon}{-\ln(1-\varepsilon)}$, we have

$$\begin{aligned}
&\sum_{q=1}^{\ell}\sum_{r=1}^{q}\sum_{\substack{c_1+\cdots+c_r=q\\ c_1\geq\cdots\geq c_r\geq 1}} \frac{(\ell-q+1)^2\ell^{2(r-1)}(a!b!)^q C^r}{p^q(n-v)^{T(r)}}\\
&\stackrel{(3.10)}{=} O(n^{-j})\sum_{q=1}^{\ell}\frac{(\ell-q+1)^2}{(1-\varepsilon)^q}\\
&\stackrel{(3.11)}{=} O(n^{-j})\frac{(1-\varepsilon)^{-\ell}}{\varepsilon^3}\\
&= O(1)\exp(-j\ln n - 3\ln\varepsilon - \ell\ln(1-\varepsilon))\\
&= O(1)\exp(-\omega) = o(1).
\end{aligned}$$

Substituting this into (3.9), we obtain $\mathbb{E}(X_\ell^2) = (n)_v^2 p^{2\ell}(1+o(1))$, as claimed. □

3.3.2 Case 2: $2 \leq j = k-1$

In this subsection we prove Lemma 3.9, i.e. the second moment method for the case when $2 \leq j = k-1$.

Since much of the proof is identical to the proof of Lemma 3.8, rather than repeating the argument, we will show how to adapt the previous proof to the special case when $2 \leq j = k-1$. Recall from Footnote 1 that the reason the proof did not go through for this case was that in (3.18) we have $T(r+1) - T(r) = \min\{j, k-j\} = 1$, and we obtain a single factor of n in the denominator of $\frac{x_q(r+1)}{x_q(r)}$, which is not quite enough to dominate the $\ell^2 q \leq \ell^3$ term in the numerator.

However, recall that $T(r) = T_q(r)$ represents a lower bound on the number of vertices of B already embedded in $Q(A, B)$ if this set splits into r intervals (for given q). To help illustrate the main idea in the adaptation of the previous proof, let us compare $T_q(2)$ with $T_q(1)$. We have $T_q(1) = (k-j)q + k - 1 = T_q(2) - 1$, but the only way of having two intervals which partition q edges and which together contain exactly $(k-j)q + k$ vertices is for the two intervals to have exactly one edge separating them, i.e. for the intervals to be of the form $f_{t_1}, \ldots, f_{t_2}$ and $f_{t_2+2}, \cdots, f_{t_3}$.[2] We call

[2] Observe that it is indeed possible to have two such intervals without the edge f_{t_2+1} between them also being shared, since the order of vertices either side of the separating edge may be different on A and B.

such a pair of intervals *adjacent.* Heuristically, if this is to happen then we have only one choice for where to place the second interval, rather than the factor of ℓ that we obtained previously (in the arguments leading to (3.12)). On the other hand, we must choose *which* of the intervals will be adjacent.

We therefore introduce a new parameter $r_1 = r_1(A, B)$, which is the number of pairs of intervals which are adjacent on B (and therefore also on A), and let $\mathcal{P}_\ell^2(q, r, r_1, \mathbf{c})$ be the subset of $\mathcal{P}_\ell^2$ with the appropriate parameters. For convenience, define $r_2 := r - r_1$. Instead of $T(r) = T_q(r)$ as in the previous case, we now define

$$\begin{aligned} T(r_1, r_2) = T_q(r_1, r_2) &= q + r(k-1) - r_1(k-2) - (r_2 - 1)(k-3) \\ &= q + r_1 + 2r_2 + k - 3. \end{aligned}$$

For convenience, we also define $r_1' := \max\{r_1, 1\}$ (so $r_1' = r_1$ unless $r_1 = 0$). The analogue of Proposition 3.11 is the following.

Proposition 3.16. *For $q, r_1 \geq 1$, there exists a constant C such that*

$$|\mathcal{P}_\ell^2(q, r, r_1, \mathbf{c})| \leq (n)_v^2 \frac{(\ell - q + 1)^2 \ell^{2(r_2 - 1)} C^r}{(n - v)^{T(r_1, r_2)}} \left(\frac{r^2}{r_1'}\right)^{r_1}.$$

Proof. In contrast to Case 1, when choosing where to place the intervals of $Q(A, B)$ on B, we first choose which pairs of intervals will be adjacent, and in which order such a pair appears along B. This is equivalent to choosing an auxiliary *adjacency graph* G, an oriented graph whose vertices are the intervals of $Q(A, B)$, and where an edge oriented from I_1 to I_2 in G indicates that these intervals will be adjacent and that I_1 will be the first of these to appear in the natural order along B. The number of ways of choosing r_1 such directed edges from among the r intervals is at most

$$\binom{\binom{r}{2}}{r_1} 2^{r_1} \leq \left(\frac{e(r^2/2)}{r_1'}\right)^{r_1} 2^{r_1} \leq \left(\frac{er^2}{r_1'}\right)^{r_1}. \tag{3.19}$$

Note that not every such choice is possible because in fact G must have maximum indegree and maximum outdegree at most 1, and furthermore must be acyclic. However, this expression certainly gives an upper bound.

We now observe that the components of G are simply directed paths (including isolated vertices, which are paths of length 0). Furthermore, for every directed path in the adjacency graph, choosing where on B to place the first edge of the first interval fixes the positions of all remaining edges of every interval on the path. We therefore consider the intervals corresponding to a component of G to be one *super-interval* (including the isolated vertices of G, which correspond to a single interval). The length of a super-interval consisting of $I_{i_1}, \ldots, I_{i_t}$ is

$$c_{i_1} + \cdots + c_{i_t} + t - 1 \geq c_{i_1} + \cdots + c_{i_t},$$

since the edge between two adjacent intervals also belongs to the super-interval. The number of super-intervals is $r - r_1 = r_2$.

Now we choose where to place the super-intervals on B, and as before we have at most ℓ choices for each, but for the last of the super-intervals we use the stronger bound $\ell - q + 1$, similarly to Case 1. Thus the number of ways of choosing the super-intervals on B is at most

$$\ell^{r_2 - 1}(\ell - q + 1). \tag{3.20}$$

We then need to choose where to place the super-intervals on A. (Note that while the edge between two adjacent intervals is not the same in A and B, which edge of A this is will naturally be fixed by the choice of where the adjacent intervals, which must lie either side of it, have been placed on A.) As before, for each super-interval we first choose an orientation along A, and subsequently there are at most ℓ choices for where to place the super-interval, or $\ell - q + 1$ for the last super-interval. Thus the number of ways of choosing where the super-intervals lie in A is at most

$$2^{r_2}\ell^{r_2 - 1}(\ell - q + 1). \tag{3.21}$$

Furthermore, by (3.5) the number of ways of ordering the vertices within the i-th interval in a way that is consistent with the choice of edges is at most

$$z_{c_i} = \Theta\left((a!b!)^{c_i}\right) = \Theta(1),$$

since $a = 1$ and $b = 0$. Since there are r intervals in total, the number of ways of re-ordering the vertices within $Q(A, B)$ is at most

$$\left(\frac{C}{2e}\right)^r \tag{3.22}$$

for some sufficiently large constant C. Thus combining the terms from (3.19), (3.20), (3.21) and (3.22), the number of ways of choosing where on A to embed the vertices within $Q(A, B)$ is at most

$$\left(\frac{er^2}{r_1'}\right)^{r_1} \ell^{2(r_2-1)}(\ell - q + 1)^2 2^{r_2} \left(\frac{C}{2e}\right)^r \leq \left(\frac{r^2}{r_1'}\right)^{r_1} \ell^{2(r_2-1)}(\ell - q + 1)^2 C^r.$$

This replaces the terms $(\ell-q+1)^2\ell^{2(r-1)}C^r$ from Proposition 3.11. All other terms remain the same as in Case 1, and observing that when $j=k-1$ we have $a=1$ and $b=0$, we obtain the statement of Proposition 3.16. $\square$

Now the analogue of Corollary 3.12 is the following

Corollary 3.17.

$$\mathbb{E}(X_\ell^2) \le (n)_v^2 p^{2\ell}\left(1+\sum_{q=1}^{\ell}\sum_{r=1}^{q}\sum_{r_1=0}^{r-1}\sum_{\substack{c_1+\cdots+c_r=q\\ c_1\ge\cdots\ge c_r\ge 1}}\frac{(\ell-q+1)^2\ell^{2(r_2-1)}C^r}{p^q(n-v)^{T(r_1,r_2)}}\left(\frac{r}{r_1'}\right)^{r_1}\right). \tag{3.23}$$

The following takes the place of Proposition 3.13.

Proposition 3.18.

$$\sum_{r=1}^{q}\sum_{r_1=0}^{r-1}\sum_{\substack{c_1+\cdots+c_r=q\\ c_1\ge\cdots\ge c_r\ge 1}}\frac{(\ell-q+1)^2\ell^{2(r_2-1)}C^r}{p^q(n-v)^{T(r_1,r_2)}}\left(\frac{r^2}{r_1'}\right)^{r_1} = O\left(n^{-j}\right)\frac{(\ell-q+1)^2}{(1-\varepsilon)^q}.$$

Proof. We first observe that

$$T(r_1,r_2) = q+r_1+2r_2+k-3 = q+2r-r_1+k-3 = O(\ell)$$

and therefore

$$\begin{aligned}(n-v)^{T(r_1,r_2)} = n^{q+2r-r_1+k-3}\left(1-O\left(\frac{\ell}{n}\right)\right)^{O(\ell)} &= n^{q+2r-r_1+k-3}\left(1-O\left(\frac{\ell^2}{n}\right)\right)\\ &= n^{q+2r-r_1+k-3}(1-o(1)).\end{aligned}$$

Since $p=\frac{1-\varepsilon}{n-k+1}$, we obtain

$$p^q(n-v)^{T(r_1,r_2)} = (1+o(1))(1-\varepsilon)^q n^{2r-r_1+k-3}. \tag{3.24}$$

As in Case 1, we define

$$y_q(r) := \sum_{\substack{c_1+\cdots+c_r=q\\ c_1\ge\cdots\ge c_r\ge 1}} 1,$$

but this time we define

$$x_q(r) := y_q(r)\sum_{r_1=0}^{r-1}\frac{\ell^{2r-2r_1}C^r}{n^{2r-r_1}}\left(\frac{r^2}{r_1'}\right)^{r_1} = y_q(r)\left(\frac{C\ell^2}{n^2}\right)^r\sum_{r_1=0}^{r-1}\left(\frac{nr^2}{\ell^2 r_1'}\right)^{r_1},$$

so that substituting these definitions into the triple-sum and using (3.24), we obtain

$$\begin{aligned}
&\sum_{r=1}^{q}\sum_{r_1=0}^{r-1}\sum_{\substack{c_1+\cdots+c_r=q\\ c_1\geq\cdots\geq c_r\geq 1}} \frac{(\ell-q+1)^2\ell^{2(r_2-1)}C^r}{p^q(n-v)^{T(r_1,r_2)}}\left(\frac{r^2}{r_1'}\right)^{r_1} \\
&= (1+o(1))\frac{(\ell-q+1)^2}{(1-\varepsilon)^q\ell^2 n^{k-3}}\sum_{r=1}^{q} x_q(r). \qquad (3.25)
\end{aligned}$$

Once again, the initial aim is to show that $\sum_{r=1}^{q} x_q(r) = (1+o(1))x_q(1)$. To achieve this, we define

$$z_{q,r}(r_1) := \left(\frac{nr^2}{\ell^2 r_1'}\right)^{r_1}.$$

Let us observe that, for $2 \leq r_1 \leq r-1$ we have

$$\begin{aligned}
\frac{z_{q,r}(r_1)}{z_{q,r}(r_1-1)} = \frac{nr^2}{\ell^2}\left(\frac{r_1^{r_1}}{(r_1-1)^{r_1-1}}\right)^{-1} &= \frac{nr^2}{\ell^2 r_1}\left(1+\frac{1}{r_1-1}\right)^{-(r_1-1)} \\
&\geq \frac{nr}{\ell^2}\cdot e^{-1} \\
&\geq n^{1/4},
\end{aligned}$$

since $\ell = O\left(\frac{\ln n}{\varepsilon}\right) = o\left(n^{1/3}\ln n\right)$. Meanwhile we also have $\frac{z_{q,r}(1)}{z_{q,r}(0)} = \frac{nr^2}{\ell^2} \geq n^{1/4}$, and so

$$z_{q,r}(r_1) \leq z_{q,r}(r-1)n^{-(r-1-r_1)/4}.$$

Therefore

$$\begin{aligned}
\sum_{r_1=0}^{r-1} z_{q,r}(r_1) &\leq z_{q,r}(r-1)\sum_{r_1=0}^{r-1} n^{-(r-1-r_1)/4} \\
&= \left(\frac{nr^2}{\ell^2\max\{r-1,1\}}\right)^{r-1}(1+o(1)) \\
&\leq \left(\frac{2nr}{\ell^2}\right)^{r-1}(1+o(1)),
\end{aligned}$$

which leads to

$$\begin{aligned}
x_q(r) &\leq y_q(r)\left(\frac{C\ell^2}{n^2}\right)^{r}\left(\frac{2nr}{\ell^2}\right)^{r-1}(1+o(1)) \\
&= (1+o(1))y_q(r)\frac{\ell^2}{2rn}\left(\frac{2Cr}{n}\right)^{r} \\
&= (1+o(1))x_q'(r), \qquad (3.26)
\end{aligned}$$

where we define

$$x'_q(r) := y_q(r)\frac{\ell^2}{2rn}\left(\frac{2Cr}{n}\right)^r.$$

Now observe that, since the definition of $y_q(r)$ is the same as in Case 1, (3.17) still holds, and so

$$\begin{aligned}\frac{x'_q(r+1)}{x'_q(r)} &= \frac{y_q(r+1)}{y_q(r)}\frac{r}{r+1}\frac{2C(r+1)}{n}\left(\frac{r+1}{r}\right)^r \\ &\leq q\cdot\frac{2Cr}{n}\cdot e = O\left(\frac{q^2}{n}\right) \leq n^{-1/4},\end{aligned}$$

since $q \leq \ell = o\left(n^{1/3}\ln n\right)$. We deduce that

$$\sum_{r=1}^{q} x'_q(r) \leq x'_q(1)\sum_{r=1}^{q} n^{-(r-1)/4} = (1+o(1))x'_q(1) \tag{3.27}$$

and therefore

$$\sum_{r=1}^{q} x_q(r) \overset{(3.26)}{\leq} (1+o(1))\sum_{r=1}^{q} x'_q(r) \overset{(3.27)}{=} (1+o(1))x'_q(1) = (1+o(1))\frac{C\ell^2}{n^2}.$$

Substituting this expression into (3.25), we obtain

$$\begin{aligned}\sum_{r=1}^{q}\sum_{r_1=0}^{r-1}\sum_{\substack{c_1+\dots+c_r=q\\ c_1\geq\dots\geq c_r\geq 1}} \frac{(\ell-q+1)^2\ell^{2(r_2-1)}C^r}{p^q(n-v)^{T(r_1,r_2)}}\left(\frac{r^2}{r'_1}\right)^{r_1} &= (1+o(1))\frac{C(\ell-q+1)^2}{(1-\varepsilon)^q n^{k-1}} \\ &= O\left(n^{-j}\right)\frac{(\ell-q+1)^2}{(1-\varepsilon)^q}\end{aligned}$$

since $j = k-1$. □

Finally observe that Claim 3.14 from Case 1 is still valid for this case. Thus as before we can combine the auxiliary results to prove the lower bound.

Proof of Lemma 3.9. By substituting the bound from Proposition 3.18 into (3.23), we obtain

$$\begin{aligned}\mathbb{E}(X_\ell^2) \leq\ & (n)_v^2 p^{2\ell}\left(1+O\left(n^{-j}\right)\sum_{q=1}^{\ell}\frac{(\ell-q+1)^2}{(1-\varepsilon)^q}\right) \\ \overset{\text{Cl.3.14}}{=} & (n)_v^2 p^{2\ell}\left(1+O\left(n^{-j}\right)\frac{2(1-\varepsilon)^\ell}{\varepsilon^3}\right),\end{aligned}$$

and exactly the same argument as in Case 1 shows that

$$O\left(n^{-j}\right)\frac{2(1-\varepsilon)^{\ell}}{\varepsilon^{3}} = o(1),$$

so

$$\mathbb{E}(X_\ell^2) \le (n)_v^2 p^{2\ell}(1+o(1)) = (1+o(1))\mathbb{E}(X_\ell)^2,$$

as required. □

3.4 The `Pathfinder` Algorithm

The proof strategy for the lower bound in the supercritical case is to define a depth-first search algorithm, which we call `Pathfinder` and which discovers j-tight paths in a k-uniform hypergraph, and to show that whp this algorithm, when applied to $H^k(n,p)$, will find a path of the appropriate length.

3.4.1 Special Case: Tight Paths in 3-uniform Hypergraphs

Before introducing the `Pathfinder` algorithm, we briefly describe the algorithm in the special case $k=3$ and $j=2$, in order to introduce some of the ideas required for the more complex general version.

In the special case, given a 3-uniform hypergraph H, the algorithm aims to construct a tight path in H starting at some *ordered pair* of vertices (v_1,v_2). It will maintain a partition of the (unordered) pairs into *neutral*, *active*, and *explored* pairs; initially only $\{v_1,v_2\}$ is active and all other pairs are neutral.

The algorithm now runs through the remaining $n-2$ vertices (apart from v_1,v_2) in turn, for each such vertex x making a *query* to reveal whether $\{v_1,v_2,x\}$ forms an edge of H. If we do not find such an edge, then the pair $\{v_1,v_2\}$ is labelled explored, and we choose a new ordered pair from which to begin (the corresponding unordered pair is then labelled active, and the corresponding vertices take the place of v_1,v_2). On the other hand, if we do find an edge $\{v_1,v_2,x\}$, then we set $v_3=x$, label the pair $\{v_2,v_3\}$ active and look for ways to extend the path from this pair.

More generally, at each step of the algorithm the current path will consist of vertices $v_1,v_2,\dots,v_{\ell+2}$, where ℓ is the length (i.e. number of edges of the path). The set of active pairs will consist of $\{v_i,v_{i+1}\}$ for $1\le i\le \ell+1$, and we will seek to extend the path from $\{v_{\ell+1}, v_{\ell+2}\}$. We therefore aim to query triples $\{v_{\ell+1}, v_{\ell+2}, x\}$, but we have some restrictions on when such a query can be made:

1. $\{v_{\ell+1}, v_{\ell+2}, x\}$ must not have been queried from $\{v_{\ell+1}, v_{\ell+2}\}$ before;

2. x may not lie in $\{v_1, \ldots, v_{\ell+2}\}$;

3. Neither $\{v_{\ell+1}, x\}$ nor $\{v_{\ell+2}, x\}$ may be explored.

The purpose of the first condition is clear: this ensures that we do not repeat previous queries and get stuck in a loop. The second condition forbids extensions which re-use a vertex which is already in the current path, which is also clearly necessary.

The third condition is perhaps the most interesting one. The algorithm would run correctly and find a tight path even without this condition, but it does ensure that no triple is ever queried more than once, which might otherwise occur as the triple $\{v_{\ell+1}, v_{\ell+2}, x\}$ might have been queried from, say, the explored pair $\{v_{\ell+1}, x\}$. While this would be permissible to create a new tight path, it would mean that the outcomes of some queries are dependent on each other, making the *analysis* of the algorithm far more difficult.

We therefore forbid such queries, which means that we may not find the longest path in the hypergraph, but if we still find a path of the required length, this is sufficient.

If we find an edge $\{v_{\ell+1}, v_{\ell+2}, x\}$ from the pair $\{v_{\ell+1}, v_{\ell+2}\}$, we set $v_{\ell+3} = x$, label $\{v_{\ell+2}, v_{\ell+3}\}$ active and continue exploring from this pair. If on the other hand we find no such edge from $\{v_{\ell+1}, v_{\ell+2}\}$, then we label $\{v_{\ell+1}, v_{\ell+2}\}$ explored, remove $v_{\ell+2}$ from the path and continue exploring from the previous active pair, i.e. $\{v_\ell, v_{\ell+1}\}$ (unless $\ell = 0$ in which case we have no further active pairs and we pick a new, previously neutral pair to start from, and order the vertices of this pair arbitrarily).

We now highlight a few ways in which the algorithm for general k and j differs from this special case, before introducing the algorithm more formally in Section 3.4.2.

Rather than the *pairs* of vertices, it will be the *j-sets* of vertices which are neutral, active or explored. We also begin our exploration process from a j-set rather than a pair.

In the special case, we also had an order of the vertices, and began with an ordered pair. In general, we will not necessarily have a *total* order of the vertices in the path, but we will have a *partial* order, or more precisely an *ordered partition* of each j-set into a set of size a and some sets of size $k - j$. This is connected to the fact that the last active j-set in the current path will contain the tail (see Section 3.2.1), and the ordered partition specifies which vertices belong to the sets $G_1, \dots, G_s$.

Related to this, depending on the values of k and j, when we discover an edge K from a j-set J, it may be that more than one j-set becomes active. More precisely, the tail will shift from $G_1, \dots, G_s$ to $G_2, \dots, G_{s+1}$, where $G_{s+1} = K \setminus J$, and any j-set containing the new tail and a vertices from G_1 is a valid place to continue extending the path, and therefore becomes active.

A consequence of this is that j-sets become active in *batches* of size $\binom{k-j}{a}$. Such a batch becomes active each time we discover an edge and from any j-set of the batch we can continue the path. Therefore we do not remove an edge (and decrease the length of the path) every time a j-set becomes explored—we only do this once *all* j-sets of the corresponding batch have become explored.

3.4.2 Hypergraph Exploration using DFS

In this section, we will describe the `Pathfinder` algorithm to find j-tight paths in k-uniform hypergraphs in full generality. We will use the following notation: if $\mathcal{F}$ is a family of sets and X is a set, we write $\mathcal{F} + X$ and $\mathcal{F} - X$ to mean $\mathcal{F} \cup \{X\}$ and $\mathcal{F} \setminus \{X\}$ respectively.

Recall from (3.2) that $a \in [k-j]$ is such that $a \equiv k \bmod (k-j)$, and from the statement of Lemma 3.3 that $s = \lceil \frac{k}{k-j} \rceil - 1 = \lceil \frac{j}{k-j} \rceil$. Let us define $r := s - 1 = \lceil \frac{j}{k-j} \rceil - 1$, so that $j = a + (k-j)r$.

Definition 3.19. Given a set J of j vertices, an *extendable partition of J* is an ordered partition $(C_0, C_1, \dots, C_r)$ of J such that $|C_0| = a$ and $|C_i| = k - j$ for all $i \in [r]$.

Note that if we have constructed a reasonably long j-tight path (i.e. of length at least s), the final j-vertices naturally come with an extendable partition $(C_0, C_1, \dots, C_r)$ according to which edges of the path they lie in, similar to the partition of all vertices of the path described in Section 3.2.1. The vertices within each part of the extendable partition could be re-ordered arbitrarily to obtain a new path with the same edge set. Therefore if we find a further edge from the final j-set to extend the path, there is more than one possibility for the final j-set of the extended path—it must contain $C_2, \dots, C_r$ and a further a vertices from C_1, which may be chosen arbitrarily. Thus an extendable partition provides a convenient way to describe the j-sets from which we might further extend the path.

Although for paths of length shorter than s the j-sets come only with a coarser (and therefore less restrictive) partition, it is convenient for a unified description of the algorithm for them to be given an extendable partition. In particular, we will start our search process from a j-set which we artificially endow with an extendable partition; this additional restriction is permissible for a lower bound on the longest path length.

We begin by giving an informal overview of the algorithm— the formal description follows.

At any given point, the algorithm will maintain a j-tight path P and a partition of the j-sets of $V(H)$ into *neutral*, *active* or *explored* sets. Initially, P is empty and every j-set is neutral. During the algorithm every j-set can change its status from neutral to active and from active to explored. The j-sets which are active or explored will be referred to as *discovered*.

The edges of P will be $e_1, \ldots, e_\ell$ (in this order), and every active j-set will be contained inside some edge of P. Whenever a new edge $e_{\ell+1}$ is added to the end of P, a batch $\mathcal{B}_{\ell+1}$ of neutral j-sets within that edge will become active: these are the j-sets from which we could potentially extend the current path. A j-set J becomes explored once all possibilities to extend P from J have been queried. Once all of the j-sets in the batch $\mathcal{B}_\ell$ corresponding to e_ℓ have been declared explored, e_ℓ will be removed from P.

The active sets will be stored in a "stack" structure (last in, first out). Each active j-set J will have an associated extendable partition $\mathcal{P}_J$ of J, and an index $i(J) \in \{0, \ldots, \ell\}$, where ℓ is the current length of P. The extendable partition will keep track of the ways in which we can extend P from J in a consistent manner, as described in Section 3.4.2. The index $i(J)$ will indicate that J belongs to the batch $\mathcal{B}_{i(J)}$ which was added when the edge $e_{i(J)}$ was added to P. Thus the algorithm will maintain a collection of batches $\mathcal{B}_0, \ldots, \mathcal{B}_\ell$, all of which consist of discovered j-sets which are inside $V(P)$. It will hold that $|\mathcal{B}_0| = 1$ and $|\mathcal{B}_i| = \binom{k-j}{a}$ for all $i \geq 1$, and all the batches will be disjoint.

All the j-sets from a single batch will change their status from neutral to active in a single step, and they will be added to the stack according to some fixed order which is chosen uniformly at random during the initialisation of the algorithm.

An iteration of the algorithm can be described as follows. Suppose J is the last active j-set in the stack. We will query k-sets K, to check whether K is an edge in H

Algorithm 1: `Pathfinder`

Input: Positive integers k and j such that $j < k$ and a k-uniform hypergraph H

Let $a \in [k-j]$ be such that $a \equiv k \bmod (k-j)$ and $r = \lceil \frac{j}{k-j} \rceil - 1$
For $i \in \{j, k\}$, let σ_i be a permutation of the i-sets of $V(H)$, chosen uniformly at random
$N \leftarrow \binom{V(H)}{j}$ `// neutral` j`-sets`
$A, E \leftarrow \emptyset$ `// active, explored` j`-sets`
$P \leftarrow \emptyset$ `// current` j`-tight path`
$\ell \leftarrow 0$ `// index tracking the current length of` P
$t \leftarrow 0$ `// "time", number of queries made so far`
while $N \neq \emptyset$ **do**
  Let J be the smallest j-set in N, according to σ_j `// "new start"`
  Choose an arbitrary extendable partition $\mathcal{P}_J$ of J
  $\mathcal{B}_0, A \leftarrow \{J\}$
  while $A \neq \emptyset$ **do**
    Let J be the last j-set in A
    Let $\mathcal{K}$ be the set of k-sets $K \subset V(H)$ such that $K \supset J$, K was not queried from J before, $K \setminus J$ is vertex-disjoint from P, and K does not contain any $J' \in E$
    if $\mathcal{K} \neq \emptyset$ **then**
      Let K be the first k-set in $\mathcal{K}$ according to σ_k
      $t \leftarrow t + 1$ `// a new query is made`
      if $K \in H$ **then** `// "query` K`"`
        $e_\ell \leftarrow K$
        $P \leftarrow P + e_\ell$ `//` P `is extended by adding` $K = e_\ell$
        $\ell \leftarrow \ell + 1$ `// length of` P `increases by one`
        Let $\mathcal{P}_J = (C_0, C_1, \ldots, C_r)$ be the extendable partition of J
        for *each* $Z \in \binom{C_1}{a}$ **do**
          $J_Z \leftarrow Z \cup C_2 \cup \cdots \cup C_r \cup (K \setminus J)$ `//` j`-set to be added`
          $\mathcal{P}_{J_Z} \leftarrow (Z, C_2, \ldots, C_r, K \setminus J)$ `// extendable partition`
          $i(J_Z) \leftarrow \ell$
          $A \leftarrow A + J_Z$ `//` j`-set becomes active`
        $\mathcal{B}_\ell \leftarrow \{J_Z : Z \in \binom{C_1}{a}\}$
      $(A_t, E_t, P_t) \leftarrow (A, E, P)$ `// update "snapshot" at time` t
    else if $\mathcal{K} = \emptyset$ **then** `// all extensions from` J `were queried`
      $A \leftarrow A - J$ `//` J `becomes explored`
      $E \leftarrow E + J$
      if $\mathcal{B}_\ell \subset E$ **then** `// the current batch is fully explored`
        $\mathcal{B}_\ell \leftarrow \emptyset$ `// empty this batch`
        $P \leftarrow P - e_\ell$ `// last edge of` P `is removed`
        $\ell \leftarrow \ell - 1$ `// length of` P `decreases by one`

or not. We only query a k-set K subject to the following conditions:

(Q1) K contains J;

(Q2) $K \setminus J$ is disjoint from the current path P;

(Q3) K was not queried from J before;

(Q4) K does not contain any explored j-set.

Condition (Q1) ensures that we only query k-sets with which we might sensibly continue the path in a j-tight manner. Condition (Q2) ensures that we do not re-use vertices that are already in P. Together, these two conditions guarantee that P will indeed always be a j-tight path. Moreover, Condition (Q3) ensures that we never query a k-set more than once from the same j-set, thus guaranteeing that the algorithm does not get stuck in an infinite loop. Finally Condition (Q4) ensures that we never query a k-set a second time from a *different* j-set (note that the possibility that K could have been queried from another *active* j-set is already excluded by Condition (Q2), since such an active j-set would lie within P). Note that, as described in Section 3.4.1, Condition (Q4) is not actually necessary for the correctness of the algorithm, but it does ensure independence of queries and is therefore necessary for our *analysis* of the algorithm.

If no such k-set K can be found in the graph H, then we declare J explored and move on to the previous active j-set in the stack. Moreover, if at this point all of the j-sets in the batch $\mathcal{B}_{i(J)}$ of J have been declared explored, the last edge e_ℓ of the current path is removed and ℓ is replaced by $\ell - 1$. If the set of active j-sets is now empty, we choose a new j-set J from which to start, declare J active and choose an extendable partition of J.

On the other hand, if we can find a suitable set K for J, we query K, and if it forms an edge, then according to the extendable partition of J, the set K will yield a new batch of j-sets (which previously were neutral and now become active). More precisely, if the extendable partition of J is $(C_0, C_1, \ldots, C_r)$, then the batch consists of all j-sets which contain $K \setminus J$ and $C_2, \ldots, C_r$, as well as a vertices of C_1. Thus the batch consists of $\binom{k-j}{a}$ many j-sets.

Finally, we keep track of a "time" parameter t, which counts the number of queries the algorithm has made. Initially, $t = 0$ and t increases by one each time we query a k-set.

During the analysis we will make reference to certain objects or families which are implicit in the algorithm at each time t even if the algorithm does not formally track them. These include the sets of neutral, active and discovered j-sets N_t, A_t, E_t and the current path P_t, which are simply the sets N, A, E and the path P at time t. We say that (A_t, E_t, P_t) is the *snapshot of H at time t*. We also refer to certain families of j-sets, including D_t (the discovered j-sets), R_t (the "new starts") and S_t (the "standard j-sets"), as well as families $F_t^{(1)}, F_t^{(2)}, F_t$ of $(k-j)$-sets (the "forbidden subsets"). The precise definitions of all of these families will be given when they become relevant.

3.4.3 Proof Strategy

Our aim is to analyse the `Pathfinder` algorithm and show that whp it finds a path of length at least $\frac{(1-\delta)\varepsilon n}{(k-j)^2}$, or at least $\frac{(1-\delta)\varepsilon^2 n}{4(k-j)^2}$ if $j = 1$. The overall strategy can be described rather simply: suppose that by some time t, which is reasonably large, we have not discovered a path of the appropriate length. Then whp (and disregarding some small error terms), the following holds:

(A) We have discovered at least $pt\binom{k-j}{a}$ many j-sets;

(B) Very few j-sets are active, therefore at least $pt\binom{k-j}{a}$ are explored;

(C) From each explored j-set, we queried at least $\binom{n'}{k-j}$ many k-sets, where $n' = \left(1 - \frac{(1-\delta)\varepsilon}{k-j}\right) n$.

(D) Thus the number of queries made is at least

$$pt\binom{k-j}{a}\binom{n'}{k-j} = t\frac{(1+\varepsilon)}{\binom{n}{k-j}}\binom{\left(1-\frac{(1-\delta)\varepsilon}{k-j}\right)n}{k-j}$$
$$\approx t(1+\varepsilon)(1-(1-\delta)\varepsilon) > t.$$

This yields a contradiction since the number of queries made is exactly t by definition.

The proof consists of making these four steps more precise. Three of these four steps are very easy to prove, once the appropriate error terms have been added:

Step (A) follows from a simple Chernoff bound applied to the number of edges discovered, along with the observation that for each edge, we discover $\binom{k-j}{a}$ many j-sets.

Step (B) follows from the observation that all active j-sets lie within some edge of the current path, and therefore there are at most $O(\varepsilon n)$ of them, which (for large enough t) is a negligible proportion of the number of discovered j-sets, and therefore almost all discovered j-sets must be explored.

Step (D) is a basic calculation arising from the bounds given by the previous three steps (though in the formal proof we do need to incorporate some error terms which we have omitted in this outline).

Thus the main difficulty is to prove Step (C). Recall that a k-set K containing J may not be queried for one of two reasons:

- $K \setminus J$ contains some vertex of P;
- K contains some explored j-set.

It is easy to bound the number of k-sets forbidden by the first condition, since we assumed that the path was never long— this is precisely what motivates the definition of n'. However, we also need to show that whp the number of k-sets forbidden by the second condition is negligible, which will be the heart of the proof.

3.5 Basic Properties of the Algorithm

Before analysing the likely evolution of the `Pathfinder` algorithm, we first collect some basic properties which will be useful later.

Note that there are two ways in which a j-set J can be discovered up to time t. First, it could have been included as a *new start* when the set of active j-sets was empty and we chose a j-set J from which to start exploring a new path (Line 10). Second, J could have been declared active if it was part of a batch of j-sets activated when we discovered an edge, which we refer to as a *standard activation* (Lines 20-30), and we refer to the j-sets which were discovered in this way as *standard* j-sets.

For any $t \geq 0$, let $\ell_t := |E(P_t)|$ be the length (i.e. number of edges) of the path found by the algorithm at time t.

Proposition 3.20. *At any time t, the number $|A_t|$ of active j-sets is at most*

$$|A_t| \leq 1 + \binom{k-j}{a} \ell_t. \tag{3.28}$$

Proof. Recall that by construction, every active j-set in A_t is contained in some edge of P_t. Moreover, every time an edge is added to the current path, exactly $\binom{k-j}{a}$ many j-sets are added via a standard activation. There is also exactly one further active j-set which was added as a new start, which gives the desired inequality. □

Note that equality does *not* necessarily hold, because some j-sets which once were active may already be explored.

For every t, let R_t be the set of all discovered j-sets at time t which were new starts, and let S_t be the discovered j-sets up to time t which are standard. Thus, for all t,

$$R_t \cup S_t = A_t \cup E_t.$$

Note that if the query at time t is answered positively, then $|S_t| = |S_{t-1}| + \binom{k-j}{a}$, and otherwise $|S_t| = |S_{t-1}|$. Thus, if $X_1, X_2, \dots$ are the indicator variables that track which queries are answered positively, i.e. X_i is 1 if the i-th k-tuple queried forms an edge and 0 otherwise, then we have

$$|S_t| = \binom{k-j}{a} \sum_{i=1}^{t} X_i. \tag{3.29}$$

Note that with input hypergraph $H = H^k(n,p)$, the $X_1, X_2, \dots$ are simply i.i.d. Bernoulli random variables with probability p. In particular, using Chernoff bounds, we can approximate $|S_t|$ when t is large.

Proposition 3.21. *Let $p = \frac{1+\varepsilon}{\binom{k-j}{a}\binom{n-j}{k-j}}$, let $t = t(n) \in \mathbb{N}$, and let $0 \le \gamma = \gamma(n) = O(1)$. Then when* `Pathfinder` *is run with input k, j and $H = H^k(n,p)$, with probability at least $1 - \exp(-\Theta(\gamma^2 pt))$ we have*

$$(1-\gamma)\frac{(1+\varepsilon)t}{\binom{n-j}{k-j}} \le |S_t| \le (1+\gamma)\frac{(1+\varepsilon)t}{\binom{n-j}{k-j}}.$$

In particular, if $\gamma^2 t n^{-(k-j)} \to \infty$, then these inequalities hold whp.

Proof. Using (3.29), the stated inequality is equivalent to

$$(1-\gamma)pt \le \sum_{i=1}^{t} X_i \le (1+\gamma)pt.$$

By the Chernoff bounds of Lemma 3.6, the probability that one of these inequalities fails is at most

$$\exp\left(-\frac{(\gamma pt)^2}{2pt}\right) + \exp\left(-\frac{(\gamma pt)^2}{2pt + \gamma pt}\right) = \exp\Big(-\Theta(\gamma^2 pt)\Big),$$

as required. □

Note that this proposition gives a lower bound on the number of discovered j-sets, but it does not immediately give an upper bound, since it says nothing about the number of new starts that have been made. (Later the number of new starts will be bounded by Proposition 3.33 in the case $j \geq 2$, ; we will not need such an upper bound in the case $j = 1$.)

How many queries are made from a given j-set J before it is declared explored? Clearly $\binom{n-j}{k-j}$ is an upper bound, since this is the number of k-sets that contain J, but some of these are excluded in the algorithm, and we will need a lower bound. In what follows, for convenience we slightly abuse terminology by referring to querying not a k-set $K \supset J$, but rather the $(k-j)$-set $K \setminus J$. (If J is already determined, this is clearly equivalent.)

There are two reasons why a $(k-j)$-set disjoint from the current j-set J may never be queried— either it contains a vertex of the current path, or it contains an explored j-set.

Definition 3.22. Consider an exploration of a k-uniform hypergraph H using the `Pathfinder` algorithm. Given t, let J be the last active set in the stack of A_t. We call a $(k-j)$-set $X \subset V(H) \setminus J$ *forbidden at time* t, if

(1) $X \cap V(P_t) \neq \emptyset$, or

(2) there exists an explored j-set $J' \in E_t$ such that $J' \subset (J \cup X)$.

If X satisfies (1) we say X is a *forbidden set of type 1*; if it satisfies (2) we say it is a *forbidden set of type 2*. Let $F^{(1)} = F_t^{(1)}$ and $F^{(2)} = F_t^{(2)}$ denote the corresponding sets of forbidden $(k-j)$-sets at time t, and let $F = F_t := F_t^{(1)} \cup F_t^{(2)}$ be the set of all forbidden $(k-j)$-sets at time t.

Observe that a $(k-j)$-set might be a forbidden set of both types, i.e. may lie in both $F^{(1)}$ and $F^{(2)}$. The following consequence of the definition of forbidden $(k-j)$-sets is crucial: if J is declared explored at time t and a $(k-j)$-set X disjoint from J is not in F_t, then X was queried from J by the algorithm (at some time $t' \leq t$). Thus, if the number of forbidden sets at time t is “small”, then a “large” number of queries were required to declare J explored.

Our aim is to bound the size of $F_t = F_t^{(1)} \cup F_t^{(2)}$. If the `Pathfinder` algorithm has not found a long path, then $F_t^{(1)}$ is small. More precisely, we obtain the following bound.

Proposition 3.23. *For all times $t \geq 0$,*

$$|F_t^{(1)}| \leq \ell_t \cdot (k-j)\binom{n-j-1}{k-j-1}.$$

Proof. Let J be the current active j-set in A_t. A $(k-j)$-set X is in $F_t^{(1)}$ if and only if $X \cap J = \emptyset$ and $X \cap V(P_t) \neq \emptyset$; thus $|F_t^{(1)}| \leq |V(P_t) \setminus J|\binom{n-j-1}{k-j-1}$. Since $J \subset V(P_t)$ and P_t has ℓ_t edges, we have $|V(P_t) \setminus J| = \ell_t \cdot (k-j)$, and the desired bound follows. $\square$

It remains to estimate the number of forbidden sets of type 2. To achieve this, in the next section we will give more precise estimates on the evolution of the algorithm run with input $H^k(n,p)$ (and in particular the evolution of discovered j-sets, which certainly includes all explored j-sets).

We will need to treat the case $j = 1$ separately from the case $j \geq 2$. We begin with the case $j = 1$, since this is significantly easier but introduces some of the ideas that will be used in the more complex case $j \geq 2$.

3.6 Algorithm Analysis: Loose Case ($j = 1$)

The case $j = 1$ is different from all other cases because the j-sets of the exploration process are simply vertices. This is important because there is a certain interplay between j-sets and vertices regarding where a path "lies"—in general, j-sets can only be blocked because they were previously explored, but vertices can be blocked because they are in the current path. Furthermore, for $j \geq 2$, we may revisit some vertices from a discarded branch of the depth-first search process, but for $j = 1$, since j-sets and vertices are the same, this is no longer possible.

This fundamental difference is reflected in the fact that the length of the longest path discovered by the `Pathfinder` algorithm in the supercritical case is significantly shorter for $j = 1$ (i.e. $\Theta(\varepsilon^2 n)$ rather than $\Theta(\varepsilon n)$). Indeed, it seems likely that this is in fact best possible up to a constant factor, i.e. that the longest loose path has length $\Theta(\varepsilon^2 n)$, rather than that either the algorithm or our analysis is far too weak. This is certainly the case for graphs, i.e. for $k = 2$; we will discuss this for general k in more detail in Section 3.10.

For convenience, we restate the result we are aiming to prove as a lemma.

Lemma 3.24. *Let $k \in \mathbb{N}$ and let $\varepsilon = \varepsilon(n)$ satisfy $\varepsilon^3 n \xrightarrow{n\to\infty} \infty$. Let*

$$p = (1+\varepsilon)p_0 = \frac{1+\varepsilon}{(k-1)\binom{n-1}{k-1}}$$

and let L be the length of the longest loose path in $H^k(n,p)$. Then for all $\delta \gg \varepsilon$ satisfying $\delta^2\varepsilon^3 n \xrightarrow{n\to\infty} \infty$, whp

$$L \geq (1-\delta)\frac{\varepsilon^2 n}{4(k-1)^2}.$$

We define

$$\ell_0 := \frac{(1-\delta)\varepsilon^2 n}{4(k-1)^2},$$

so our goal is to show that whp the `Pathfinder` algorithm discovers a path of length at least ℓ_0. We also define

$$T_0 := \frac{\varepsilon n\binom{n-1}{k-1}}{2(k-1)} = \frac{\varepsilon n}{2(k-1)^2 p_0}.$$

We will show that whp at some time $t \leq T_0$, we have $\ell_t \geq \ell_0$, as required. We begin with the following proposition, which is a simple application of Proposition 3.21.

Proposition 3.25. *At time $t = T_0$, whp we have*

$$|A_t \cup E_t| \geq (1 - o(\delta\varepsilon))(k-1)pt.$$

Proof. Since $|A_t \cup E_t| \geq |S_t|$ and $(k-1)pt = (1+\varepsilon)t/\binom{n-1}{k-1}$, we can apply Proposition 3.21: it is enough to find γ such that $\gamma = o(\delta\varepsilon)$ and $\gamma^2 pt \to \infty$. Recall that $\delta^2\varepsilon^3 n \to \infty$. Let $\omega = \delta^2\varepsilon^3 n$. Then $\gamma = \delta\varepsilon/\omega^{1/3}$ clearly satisfies $\gamma = o(\delta\varepsilon)$. On the other hand, by the choice of $t = T_0$, we have $pt = \Theta(\varepsilon n)$. Thus $\gamma^2 pt = \Theta(\omega^{1/3}) \to \infty$, as required. $\square$

Let T_1 denote the first time t at which

$$|A_t \cup E_t| = \left(1 - \frac{\delta\varepsilon}{3}\right)(k-1)pT_0 = \left(1 - \frac{\delta\varepsilon}{3}\right)(1+\varepsilon)\frac{\varepsilon n}{2(k-1)} \tag{3.30}$$

(recall that we ignore floors and ceilings). Then from Proposition 3.25, we immediately obtain the following.

Corollary 3.26. *Whp $T_1 \leq T_0$.*

We claim furthermore that this inequality implies that we must have a long loose path.

Proposition 3.27. *If $T_1 \le T_0$, then at time $t = T_1$ we have $\ell_t \ge \ell_0$.*

Proof. Suppose for a contradiction that $T_1 \le T_0$, but that at time $t = T_1$ we have $\ell_t < \ell_0$. Then by (3.30) and (3.28)

$$\begin{aligned}
|E_t| &= |A_t \cup E_t| - |A_t| \\
&\ge \left(1 - \frac{\delta\varepsilon}{3}\right)(k-1)pT_0 - ((k-1)\ell_0 + 1) \\
&= \left(1 + \varepsilon - \frac{\delta\varepsilon}{3} - O(\delta\varepsilon^2)\right)(k-1)p_0T_0 - \frac{(1-\delta)\varepsilon^2 n}{4(k-1)} - o(\delta\varepsilon^2 n) \\
&= \left(1 + \varepsilon - \frac{\delta\varepsilon}{3} - \frac{(1-\delta)\varepsilon}{2} - O(\delta\varepsilon^2) - o(\delta\varepsilon)\right)(k-1)p_0T_0 \\
&\ge \left(1 + \frac{\varepsilon}{2} + \frac{\delta\varepsilon}{7}\right)(k-1)p_0T_0,
\end{aligned}$$

where we have used the fact that $(k-1)p_0T_0 = \frac{\varepsilon n}{2(k-1)} = \Theta(\varepsilon n)$, and that $\delta\varepsilon^2 n \ge \varepsilon^3 n \to \infty$.

On the other hand, N_t, the set of neutral vertices, satisfies

$$\begin{aligned}
|N_t| = n - |A_t \cup E_t| &\overset{(3.30)}{=} n - (1 - o(\delta\varepsilon))(1+\varepsilon)\frac{\varepsilon n}{2(k-1)} \\
&= \left(1 - \frac{\varepsilon}{2(k-1)} + o(\delta\varepsilon) + O(\varepsilon^2)\right) n.
\end{aligned}$$

Note that no vertex of N_t can possibly have been forbidden at any time $t' \le t$. This implies, since the vertices of E_t are fully explored, that from each explored vertex we certainly queried any k-set containing the vertex and $k-1$ vertices of N_t. Thus the number of queries t that we have made so far certainly satisfies

$$\begin{aligned}
t &\ge |E_t| \binom{|N_t|}{k-1} \\
&\ge \left(1 + \frac{\varepsilon}{2} + \frac{\delta\varepsilon}{7}\right)(k-1)p_0T_0 \cdot \left(1 + O\left(\frac{1}{n}\right)\right) \frac{\left(1 - \frac{\varepsilon}{2(k-1)} + o(\delta\varepsilon) + O(\varepsilon^2)\right)^{k-1} n^{k-1}}{(k-1)!} \\
&= \left(1 + \frac{\varepsilon}{2} + \frac{\delta\varepsilon}{7}\right) T_0 \cdot \left(1 + O\left(\frac{1}{n}\right)\right)\left(1 - \frac{\varepsilon}{2} + o(\delta\varepsilon) + O(\varepsilon^2)\right) \\
&= \left(1 + \frac{\delta\varepsilon}{7} + o(\delta\varepsilon) + O(\varepsilon^2)\right) T_0 \\
&> T_0,
\end{aligned}$$

which gives the required contradiction since we assumed that $t = T_1 \le T_0$. □

Proof of Lemma 3.24. The statement of Lemma 3.24 follows directly from Corollary 3.26 and Proposition 3.27. □

Let us note that although we proved that whp $\ell_t \geq \ell_0$ at *some* time $t \leq T_0$, with a small amount of extra work we could actually prove that this even holds at exactly $t = T_0$: we would need a corresponding upper bound in Proposition 3.25, which follows from a Chernoff bound on the number of edges discovered so far and an upper bound on the number of new starts we have made by time T_0.

3.7 Algorithm Analysis: High-order Case ($j \geq 2$)

In the case $j \geq 2$, we will use the `Pathfinder` algorithm to study j-tight paths in $H^k(n,p)$ by running the algorithm up to a certain *stopping time* T_{stop}, i.e. until we have made T_{stop} queries. In order to define T_{stop}, we need some additional definitions.

Given some time $t \geq 0$ let D_t denote the set of all j-sets which are discovered by time t. With a slight abuse of notation, we will sometimes also use D_t to denote the j-uniform hypergraph on vertex set $[n]$ with edge set D_t. Note that a j-set J lies in D_t if and only if there exists $t' \leq t$ such that $J \in A_{t'}$, or in other words, every j-set which is discovered at time t was active at some time $t' \leq t$. Also, note that for every $t_1 \leq t_2$, $D_{t_1} \subseteq D_{t_2}$, i.e. the sequence of discovered j-sets is always increasing (although the sequence of active sets A_t is not).

Suppose that $0 \leq i \leq j$ and that I is an i-set. Then define $d(I) = d_t(I) = \deg_{D_t}(I)$ to be the number of j-sets of D_t that contain I.

Definition 3.28. Let $\varepsilon \ll \delta \leq 1$ be as in Theorem 3.1(ii),[3] and recall that $|R_t|$ is the number of new starts made by time t. Let

$$C_{k,j,j-1} \gg C_{k,j,j-2} \gg \cdots \geq C_{k,j,0} \gg 1$$

be some sufficiently large constants and let $0 < \beta \ll 1$ be a sufficiently small constant. Define

$$T_0 := \frac{n^{k-j+1}}{\varepsilon}.$$

We define T_{stop} to be the smallest time t such that one of the following stopping conditions hold:

[3] Recall from Remark 3.2 that we will not actually use the additional condition $\delta \gg \frac{\ln n}{\varepsilon^2 n}$ for the proof of the lower bound, c.f. Lemma 3.30.

(S1) `Pathfinder` found a path of length at least $(1-\delta)\frac{\varepsilon n}{(k-j)^2}$;

(S2) $t = T_0$;

(S3) $|R_t| \geq 2(k-j)!\sqrt{\frac{tn^\beta}{n^{k-j}}} + \frac{n^\beta}{2}$;

(S4) There exists some $0 \leq i \leq j-1$ and an i-set I with $d_t(I) \geq \frac{C_{k,j,i}t}{n^{k-j+i}} + n^\beta$.

We first observe that T_{stop} is well-defined.

Claim 3.29. *If* `Pathfinder` *is run on inputs k, j and any k-uniform hypergraph H on $[n]$, then one of the four stopping conditions is always applied.*

Proof. If none of the stopping conditions is applied, the algorithm will continue until all j-sets are explored (since a new start is always possible from any neutral j-set). If this occurs at time $t \geq T_0$, then **(S2)** would already have been applied (if none of the other stopping conditions were applied first). On the other hand, if this occurs at time $t \leq T_0$, then **(S4)** is certainly satisfied with $i = 0$ and $I = \emptyset$. $\square$

We will often use the fact that for $t \leq T_{\text{stop}}$, the (non-strict) inequalites in stopping conditions **(S1)**, **(S3)** and **(S4)** are reversed. For example, for $t \leq T_{\text{stop}}$ we have $|R_t| \leq 2(k-j)!\sqrt{\frac{tn^\beta}{n^{k-j}}} + n^\beta$. This is because

$$|R_t| \leq |R_{t-1}| + 1 < 2(k-j)!\sqrt{\frac{(t-1)n^\beta}{n^{k-j}}} + n^\beta + 1,$$

where the second inequality holds because we did not apply **(S3)** by time $t-1$ (and recall that we ignore floors and ceilings). In such a situation, we will slightly abuse terminology by saying that "by **(S3)**" we have $|R_t| \leq 2(k-j)!\sqrt{\frac{tn^\beta}{n^{k-j}}} + n^\beta$.

Our main goal is to show that whp it is **(S1)** which is applied first, i.e. the algorithm has indeed discovered a path of the appropriate length.

Lemma 3.30. *Let $k, j \in \mathbb{N}$ satisfy $2 \leq j \leq k-1$. Let $a \in \mathbb{N}$ be the unique integer satisfying $1 \leq a \leq k-j$ and $a \equiv k \bmod (k-j)$. Let $\varepsilon = \varepsilon(n) \ll 1$ satisfy $\varepsilon^3 n \xrightarrow{n\to\infty} \infty$ and let*

$$p_0 = p_0(n;k,j) := \frac{1}{\binom{k-j}{a}\binom{n-j}{k-j}}.$$

Let L be the length of the longest j-tight path in $H^k(n,p)$, and let $\delta \gg \varepsilon$.

Suppose `Pathfinder` *is run with input k, j and $H = H^k(n,p)$. Then whp* **(S1)** *is applied. In particular, whp*

$$L \geq \ell_{T_{\text{stop}}} = (1-\delta)\frac{\varepsilon n}{(k-j)^2}.$$

For the rest of this section, we will assume that all parameters are as defined in Lemma 3.30.

We first prove an auxiliary lemma which gives an upper bound on the number of forbidden $(k-j)$-sets up to time T_{stop}. Recall that $F_t^{(1)}$ and $F_t^{(2)}$ denote the sets of forbidden $(k-j)$-sets at time t of types 1 and 2, respectively. Let $f^{(i)} = f_t^{(i)} := |F_t^{(i)}|$ for $i = 1, 2$.

Lemma 3.31. *Let $t \leq T_{\text{stop}}$. Then*

$$f^{(1)} + f^{(2)} \leq (1-\delta/2)\varepsilon\binom{n-j}{k-j}.$$

In particular, from every explored j-set we made at least

$$(1-\varepsilon+\delta\varepsilon/2)\binom{n-j}{k-j}$$

queries.

Proof. Due to condition **(S1)**, the length ℓ_t of the path P_t at any time t is at most $\frac{(1-\delta)\varepsilon n}{(k-j)^2}$. Thus by Proposition 3.23 we have that

$$f^{(1)} \leq \frac{(1-\delta)\varepsilon n}{(k-j)} \cdot \binom{n-j-1}{k-j-1} \leq \left(1-\frac{2\delta}{3}\right)\varepsilon\binom{n-j}{k-j}. \tag{3.31}$$

By condition **(S2)**, we have $T_{\text{stop}} \leq \frac{n^{k-j+1}}{\varepsilon}$. Furthermore, by condition **(S4)**, for any $0 \leq i \leq j-1$ and any i-set I we have

$$d_t(I) \leq d_{T_{\text{stop}}}(I) \leq \frac{C_{k,j,i}T_{\text{stop}}}{n^{k-j+i}} + n^\beta \leq \frac{C_{k,j,i}}{\varepsilon n^{i-1}} + n^\beta.$$

Observe that if J is the current j-set, any forbidden $(k-j)$-set of type 2 can be identified by:

- choosing an integer $i = 0, \ldots, j-1$;
- choosing a proper subset $I \subset J$ of size i (there are $\binom{j}{i}$ possibilities);

- choosing an explored (and therefore discovered) j-set $J' \supset I$ such that $(J' \setminus I) \cap J = \emptyset$, (at most $d_t(I)$ possibilities);
- choosing a k-set K containing both J and J' (there are $\binom{n-2j+i}{k-2j+i}$ possibilities).

Then the forbidden $(k-j)$-set is $K \setminus J$. Note that if $j > k/2$, then $k - 2j + i$ may be negative for some values of i. In this case we interpret $\binom{n-2j+i}{k-2j+i}$ to be zero.

Therefore we obtain

$$\begin{aligned} f^{(2)} &\leq \sum_{i=0}^{j-1} \binom{j}{i} \cdot \left(\max_{|I|=i} d_t(I) \right) \cdot \binom{n-2j+i}{k-2j+i} \\ &\leq \sum_{i=0}^{j-1} 2^j \cdot \left(\frac{C_{k,j,i}}{\varepsilon n^{i-1}} + n^\beta \right) \cdot O(n^{-j+i}) \binom{n-j}{k-j} \\ &= O\left(\frac{1}{\delta \varepsilon^2 n^{j-1}} + \frac{n^\beta}{\delta \varepsilon n} \right) \delta\varepsilon \binom{n-j}{k-j}. \end{aligned}$$

Now recall that $\delta \gg \varepsilon$ and that we are considering the case $j \geq 2$, which means that $\delta\varepsilon^2 n^{j-1} \geq \varepsilon^3 n \to \infty$. Furthermore $\beta \ll 1$, which implies that $\delta\varepsilon n^{1-\beta} \geq \varepsilon^2 n^{2/3} \to \infty$, so we obtain

$$f^{(2)} = o(1)\delta\varepsilon \binom{n-j}{k-j}.$$

Together with (3.31), this leads to

$$f^{(1)} + f^{(2)} \leq \left(1 - \frac{2\delta}{3} + o(\delta)\right) \varepsilon \binom{n-j}{k-j} \leq (1 - \delta/2)\varepsilon \binom{n-j}{k-j}$$

as claimed. □

Our aim now is to prove Lemma 3.30, i.e. that whp stopping condition **(S1)** is applied. Our strategy is to show that whp each of the other three stopping conditions is *not* applied. The arguments for **(S2)** and **(S3)** are almost identical, so it is convenient to handle them together. We begin with the following proposition.

Proposition 3.32. *There exists an event $\mathcal{A}$ such that:*

(i) $\mathbb{P}(\mathcal{A}) = 1 - o(1)$;

(ii) *if $\mathcal{A}$ holds and either* **(S2)** *or* **(S3)** *is applied at time $t = T_{\mathrm{stop}}$, then*

$$|E_t| \geq \frac{(1 - 2\delta\varepsilon/5)(1+\varepsilon)t}{\binom{n-j}{k-j}}.$$

Proof. We first define the event $\mathcal{A}$ explicitly. For any time $t > 0$ we define

$$\gamma_t := \begin{cases} \sqrt{\frac{n^{k-j+\beta}}{t}} & \text{if } t < T_0, \\ \frac{\delta\varepsilon}{3} & \text{otherwise,} \end{cases}$$

and let

$$\mathcal{A}_t := \left\{ |S_t| \geq (1-\gamma_t)\, \frac{(1+\varepsilon)t}{\binom{n-j}{k-j}} \right\}.$$

Now we define

$$\mathcal{A} := \bigcap_{\frac{n^{k-j+\beta}}{4(k-j)!} \leq t \leq T_0} \mathcal{A}_t.$$

We now need to show that the two properties of the proposition are satisfied for this choice of $\mathcal{A}$. First observe that for $\frac{n^{k-j+\beta}}{4(k-j)!} \leq t < T_0$, Proposition 3.21 (applied with $\gamma = \gamma_t$) implies that

$$\mathbb{P}(\mathcal{A}_t) \geq 1 - \exp\left(-\Theta\left(\gamma_t^2 pt\right)\right) \geq 1 - \exp\left(-\Theta\left(\gamma_t^2 \frac{t}{n^{k-j}}\right)\right) \geq 1 - \exp\left(-\Theta\left(n^{\beta}\right)\right).$$

On the other hand, for $t = T_0$ again Proposition 3.21 implies that

$$\mathbb{P}(\mathcal{A}_{T_0}) \geq 1 - \exp\left(-\Theta\left(\gamma_t^2 pT_0\right)\right) = 1 - \exp\left(-\Theta\left(\delta^2 \varepsilon n\right)\right) = 1 - o(1),$$

where the convergence holds because $\delta^2\varepsilon n \geq \varepsilon^3 n \to \infty$. Therefore by applying a union bound,

$$\mathbb{P}(\mathcal{A}) \geq 1 - T_0 \exp\left(-\Theta\left(n^{\beta}\right)\right) - o(1) = 1 - o(1),$$

as required.

We now aim to prove the second statement, so let us assume that $\mathcal{A}$ holds, and we make a case distinction according to which of **(S3)** and **(S2)** is applied.

Case 1: (S3) is Applied

By applying Lemma 3.31 we can bound the number of queries made from each explored j-set at any time $t \leq T_{\text{stop}}$ from below by

$$(1 - \varepsilon + \delta\varepsilon/2)\binom{n-j}{k-j} \geq \frac{3n^{k-j}}{4(k-j)!}.$$

In particular, since **(S3)** is applied, we must have made at least $n^{\beta}/2$ new starts, and therefore at least $n^{\beta}/2 - 1 \geq n^{\beta}/3$ many j-sets are explored. Thus we have made at least $\frac{n^\beta}{3} \cdot \frac{3n^{k-j}}{4(k-j)!}$ queries, and therefore we may assume that $T_{\text{stop}} \geq \frac{n^{k-j+\beta}}{4(k-j)!}$. (Note that this in particular motivates why the definition of $\mathcal{A}$ did not include any $\mathcal{A}_t$ for $t < \frac{n^{k-j+\beta}}{4(k-j)!}$.)

Furthermore, since **(S2)** is *not* applied, we have $T_{\text{stop}} < T_0$. Therefore, the fact that $\mathcal{A}$ holds tells us that for $t = T_{\text{stop}}$,

$$|D_t| \geq |S_t| + |R_t| \geq (1-\gamma_t)(1+\varepsilon)\frac{t}{\binom{n-j}{k-j}} + |R_t|. \tag{3.32}$$

Since **(S3)** is applied at $t = T_{\text{stop}}$, we further have

$$|R_t| \geq 2(k-j)!\sqrt{\frac{tn^\beta}{n^{k-j}}} \geq \frac{3\gamma_t t}{2\binom{n-j}{k-j}}.$$

Substituting this inequality into (3.32), we obtain

$$|D_t| \geq (1-\gamma_t)(1+\varepsilon)\frac{t}{\binom{n-j}{k-j}} + \frac{3\gamma_t t}{2\binom{n-j}{k-j}} \geq (1+\varepsilon)\frac{t}{\binom{n-j}{k-j}}.$$

Furthermore, since **(S3)** is applied at $t = T_{\text{stop}}$, a new start must have been made at time t. This implies that the set of active sets at time A_t was empty, i.e. $|A_t| = 0$. This means that

$$|E_t| = |D_t| \geq (1+\varepsilon)\frac{t}{\binom{n-j}{k-j}} \geq \frac{(1-2\delta\varepsilon/5)(1+\varepsilon)t}{\binom{n-j}{k-j}},$$

as claimed.

Case 2: (S2) is Applied

We will use the trivial bound $|R_t| \geq 0$, and therefore $\mathcal{A}$ tells us that at time $t = T_0 = T_{\text{stop}}$ we have

$$|D_t| = |S_t| + |R_t| \geq \left(1 - \frac{\delta\varepsilon}{3}\right)(1+\varepsilon)\frac{t}{\binom{n-j}{k-j}}.$$

Furthermore, by **(S1)**,

$$\ell_t = O(\varepsilon n),$$

and therefore by (3.28)

$$|A_t| \leq 1 + \binom{k-j}{a}\ell_t = O(\varepsilon n) = O\left(\frac{\varepsilon^2 T_0}{n^{k-j}}\right).$$

Thus the number of explored sets at time T_0 satisfies

$$|E_{T_0}| = |D_{T_0}| - |A_{T_0}| \geq \frac{\left((1-\delta\varepsilon/3)(1+\varepsilon) - O\left(\varepsilon^2\right)\right) T_0}{\binom{n-j}{k-j}} \geq \frac{(1-2\delta\varepsilon/5)(1+\varepsilon)T_0}{\binom{n-j}{k-j}},$$

where in the last step we have used the fact that $\delta \gg \varepsilon$. □

The previous result enables us to prove the following.

Proposition 3.33. *Whp neither* **(S2)** *nor* **(S3)** *is applied.*

Proof. For any time $t \geq 0$, let us define the event

$$\mathcal{E}_t := \left\{ |E_t| \geq \frac{(1-2\delta\varepsilon/5)(1+\varepsilon)t}{\binom{n-j}{k-j}} \right\},$$

i.e. that the bound on $|E_t|$ from Proposition 3.32 holds. We will show that in fact it is not possible that $\mathcal{E}_t$ holds for any $t \leq T_{\text{stop}}$. Therefore, Proposition 3.32 implies that the probability that one of **(S3)** and **(S2)** is applied is at most $1 - \mathbb{P}(\mathcal{A}) = o(1)$. So suppose for a contradiction that $\mathcal{E}_t$ holds for some $t \leq T_{\text{stop}}$.

As in the proof of Proposition 3.32, an application of Lemma 3.31 implies that from each explored j-set at any time $t \leq T_{\text{stop}}$ we made at least

$$(1-\varepsilon+\delta\varepsilon/2)\binom{n-j}{k-j} \geq \frac{3n^{k-j}}{4(k-j)!}$$

queries. Therefore, by Proposition 3.32, the total number t of queries made satisfies

$$t \geq |E_t| \cdot (1-\varepsilon+\delta\varepsilon/2)\binom{n-j}{k-j} \geq (1-2\delta\varepsilon/5+\delta\varepsilon/2+O(\varepsilon^2))t > t,$$

yielding the desired contradiction. □

We next prove that whp **(S4)** is not applied. This may be seen as a form of *bounded degree lemma*. Both the result and the proof are inspired by similar results in [17, 18].

The intuition behind this stopping condition is that the *average* degree of an i-set should be of order $\frac{tp}{n^i} \sim \frac{t}{n^{k-j+i}}$, and **(S4)** guarantees that, for $t \leq T_{\text{stop}}$, no i-set exceeds this by more than a constant factor. The n^β-term can be interpreted as an error term which takes over when the average i-degree (i.e. the average degree over all i-sets) is too small to guarantee an appropriate concentration result.

Note, however, that due to the choice of T_0, the average i-degree is actually much smaller than n^β for any $i \geq 2$ (and possibly even for $i = 1$ if $\varepsilon = \Omega(n^{-\beta})$). Meanwhile, the statement for $i = 0$ is simply a statement about the number of discovered j-sets, which follows from a simple Chernoff bound on the number of edges discovered, together with **(S3)** to bound the number of new starts. Thus the strongest and most interesting case of the statement is when $i = 1$; nevertheless, our proof strategy is strong enough to cover all i and would even work for any $t > T_0$, provided **(S3)** has not yet been applied.

Lemma 3.34. *Whp* **(S4)** *is not applied.*

Proof. We will prove that the probability that **(S4)** is applied at a particular time $t \leq T_{\text{stop}}$, i.e. before any other stopping condition has been applied, is at most $\exp\left(-\Theta\left(n^{\beta/2}\right)\right) = o(n^{-k})$, and then a union bound over all possible t completes the argument.

We will prove the lemma by induction on i. For $i = 0$ the statement is just that the number of discovered j-sets is at most $C_{k,j,0}t/n^{k-j} + n^\beta$, which follows from Lemma 3.7 and **(S3)**. More precisely, using (3.29) and applying Lemma 3.7 with $\alpha = \beta/2$, we have that

$$\mathbb{P}\left(\frac{|S_t|}{\binom{k-j}{a}} \geq 2tp + n^{\beta/2}\right) \leq \exp\left(-\Theta\left(n^{\beta/2}\right)\right).$$

Furthermore, by **(S3)**, we have

$$\begin{aligned}
|R_t| &\leq 2(k-j)!\sqrt{\frac{tn^\beta}{n^{k-j}}} + \frac{n^\beta}{2} \\
&\leq \begin{cases} \frac{3n^\beta}{4} & \text{if } t \leq \frac{n^{k-j+\beta}}{64((k-j)!)^2}, \\ 16((k-j)!)^2 \frac{t}{n^{k-j}} + \frac{n^\beta}{2} & \text{if } t \geq \frac{n^{k-j+\beta}}{64((k-j)!)^2} \end{cases} \\
&\leq 16((k-j)!)^2 \frac{t}{n^{k-j}} + \frac{3n^\beta}{4}.
\end{aligned}$$

Thus with probability at least $1 - \exp\left(-\Theta\left(n^{\beta/2}\right)\right)$ we have

$$\begin{aligned}
|D_t| = |S_t| + |R_t| &\leq \binom{k-j}{a}\left(2tp + n^{\beta/2}\right) + 16((k-j)!)^2 \frac{t}{n^{k-j}} + \frac{3n^\beta}{4} \\
&\leq \left(3(k-j)! + 16((k-j)!)^2\right) \cdot \frac{t}{n^{k-j}} + n^\beta \\
&\leq \frac{20((k-j)!)^2 t}{n^{k-j}} + n^\beta,
\end{aligned}$$

and since we chose $C_{k,j,0} \gg 1$, and in particular $C_{k,j,0} > 20((k-j)!)^2$, this shows that whp **(S4)** is not applied because of $I = \emptyset$ (i.e. with $i = 0$). So we will assume that $i \geq 1$ and that **(S4)** is not applied for $0, 1, \ldots, i-1$.

Given $1 \leq i \leq j-1$ and an i-set I, let us consider the possible ways in which some j-sets containing I may become active.

- A *new start* at I occurs when there are no active j-sets and we make a new start at a j-set which happens to contain I. In this case $d(I)$ increases by 1;
- A *jump* to I occurs when we query a k-set containing I from a j-set not containing I and discover an edge. In this case $d(I)$ increases by at most $\binom{k-j}{a}$ (the number of new j-sets which become active in a batch, each of which may or may not contain I);
- A *pivot* at I occurs when we query a k-set from a j-set containing I and discover an edge. In this case $d(I)$ increases by at most $\binom{k-j}{a}$.

Each possibility makes a contribution to the degree of I according to how many j-sets containing I become active as a result of each type of event. We bound the three contributions separately.

New starts: Whenever we make a new start, we choose the starting j-set according to some (previously fixed) random ordering σ_j (recall that σ_j was a permutation of the j-sets chosen uniformly at random during the initialisation of the algorithm). By **(S3)**, at time $t \leq T_{\text{stop}}$ the number of new starts we have made is

$$|R_t| \leq 2(k-j)!\sqrt{\frac{tn^\beta}{n^{k-j}}} + \frac{n^\beta}{2}.$$

Observe that

$$\sqrt{\frac{tn^\beta}{n^{k-j}}} \leq \begin{cases} n^\beta & \text{if } t \leq n^{k-j+\beta}, \\ \frac{t}{n^{k-j}} & \text{if } t \geq n^{k-j+\beta}, \end{cases}$$

which means that the number of new starts satisfies

$$|R_t| \leq 2(k-j)!tn^{j-k} + 3(k-j)!n^\beta =: N^*.$$

Since the new starts are distributed randomly, the probability that a j-set chosen for a new start at time $t' \leq t$ contains I is precisely the proportion of neutral j-sets at time

t' which contain I. Since **(S4)** has not yet been applied, in particular with $i = 0$, the total number of non-neutral j-sets (which cannot be chosen for a new start) at time $t' \leq t$ is at most

$$d_{t'}(\emptyset) \leq d_t(\emptyset) \leq \frac{C_{k,j,0}t}{n^{k-j}} + n^\beta \leq \frac{C_{k,j,0}n}{\varepsilon} + n^\beta \leq n^{4/3} = o(n^j).$$

Thus the probability that the j-set chosen contains I is at most

$$\frac{\binom{n-i}{j-i}}{\binom{n}{j} - o(n^j)} \leq 2j!n^{-i}.$$

Therefore the number of new starts containing I is dominated by $\mathrm{Bin}(N^*, 2j!n^{-i})$, which has expectation at most $4k!tn^{j-k-i} + 1$ (since $n^{\beta-i} = o(1)$). By Lemma 3.7, with probability at least $1 - \exp(-\Theta(n^{\beta/2}))$ the number of new starts at I is at most

$$8k!tn^{j-k-i} + 2 + n^{\beta/2} \leq 8k!tn^{j-k-i} + n^{2\beta/3}.$$

Taking a union bound over all possible i-sets I, with probability at least

$$1 - \binom{n}{i}\exp(-\Theta(n^{\beta/2})) = 1 - \exp(-\Theta(n^{\beta/2})),$$

every i-set is contained in at most

$$8k!n^{j-k-i}t + n^{2\beta/3} \tag{3.33}$$

new starts.

Jumps: From each j-set J which became active in the search process, but which did not contain I, if we queried a k-set containing I and this k-set was an edge, then the degree of I may increase by up to $\binom{k-i}{j-i}$. To bound the number of such jumps, we distinguish according to the intersection $Z = J \cap I$, and denote $z := |Z|$. Observe that $0 \leq z \leq i-1$, and for each of the $\binom{i}{z}$ many z-sets $Z \subset I$, by the fact that **(S4)** has not been previously applied for this set Z, there are at most $d_t(Z) \leq \frac{C_{k,j,z}t}{n^{k-j+z}} + n^\beta$ many j-sets in D_t which intersect I in Z. For each such j-set J, there are at most $\binom{n}{k-j-i+z} \leq n^{k-j-i+z}$ many k-sets containing both J and I, i.e. which we might have queried from J and which would result in jumps to I.

Thus in total, the number of k-sets which we may have queried and which might have resulted in a jump to I is at most

$$\begin{aligned}\sum_{z=0}^{i-1}\binom{i}{z}\left(\frac{C_{k,j,z}t}{n^{k-j+z}}+n^{\beta}\right)n^{k-j-i+z} &= \sum_{z=0}^{i-1}\binom{i}{z}\left(\frac{C_{k,j,z}t}{n^{i}}+n^{k-j-i+z+\beta}\right)\\ &\le 2^{i}\left(\max_{0\le z\le i-1}C_{k,j,z}\frac{t}{n^{i}}+n^{k-j-1+\beta}\right)\\ &= 2^{i}\left(C_{k,j,i-1}\frac{t}{n^{i}}+n^{k-j-1+\beta}\right)=: N,\end{aligned}$$

since we chose $C_{k,j,j-1} \gg C_{k,j,j-2} \gg \ldots \gg C_{k,j,0}$. Then the number of edges that we discover which result in jumps to I is dominated by $\mathrm{Bin}(N,p)$. By Lemma 3.7, with probability at least $1-\exp(-\Theta(n^{\beta/2}))$ this random variable is at most

$$\begin{aligned}2Np+n^{\beta/2} &\le \frac{(k-j)!}{\binom{k-j}{a}}2^{i+2}C_{k,j,i-1}\frac{t}{n^{k-j+i}}+O\left(n^{-1+\beta}\right)+n^{\beta/2}\\ &\le \frac{(k-j)!}{\binom{k-j}{a}}2^{i+2}C_{k,j,i-1}\frac{t}{n^{k-j+i}}+2n^{\beta/2},\end{aligned}$$

and so the contribution to the degree of I made by jumps to I is at most

$$(k-j)!2^{i+2}C_{k,j,i-1}\frac{t}{n^{k-j+i}}+n^{2\beta/3}. \tag{3.34}$$

Pivots: Whenever we have a jump to I or a new start at I, some j-sets containing I become active. From these j-sets we may query further k-sets, potentially resulting in some more j-sets containing I becoming active. However, the number of such j-sets containing I that become active due to such a pivot is certainly at most $\binom{k-j}{a}$. Thus the number of further j-sets that become active due to pivots from some j-set J is at most $\binom{k-j}{a}\cdot\mathrm{Bin}\left(\binom{n-j}{k-j},p\right)$, which has expectation $1+\varepsilon$.

Furthermore, the number of such sequential pivots that we may make before leaving I in the j-tight path is $\lfloor\frac{k-i}{k-j}\rfloor \le k-i$. Thus the number of pivots arising from a single j-set containing I may be upper coupled with a branching process in which vertices in the first $(k-i)$ generations produce $\binom{k-j}{a}\cdot\mathrm{Bin}\left(\binom{n-j}{k-j},p\right)$ children, and thereafter no more children are produced.

We bound the total size of all such branching processes together. Suppose the contribution to the degree of I made by jumps and new starts is x. Then we have x vertices in total in the first generation, and by the arguments above, with probability $1-\exp(-\Omega(n^{\beta/2}))$ we have, by (3.33) and (3.34), that

$$x \le \left(8k!+(k-j)!2^{i+2}C_{k,j,i-1}\right)\frac{t}{n^{k-j+i}}+2n^{2\beta/3} \le 2^{i+3}k!C_{k,j,i-1}\frac{t}{n^{k-j+i}}+2n^{2\beta/3}.$$

For convenience, we will assume (for an upper bound) that in fact $x \geq n^\beta$. The number of children in the second generation is dominated by $\binom{k-j}{a} \cdot \mathrm{Bin}\left(x\binom{n-j}{k-j}, p\right)$, which has expectation $(1+\varepsilon)x$, and so by Lemma 3.7, with probability $1-\exp(-\Omega(n^{\beta/2}))$, the number of children is at most $2(1+\varepsilon)x + n^{\beta/2} \leq 4x$. Similarly, with probability $1-\exp(-\Omega(n^{\beta/2}))$, the number of vertices in the third generation is at most $16x$, and inductively the number of vertices in the m-th generation is at most $2^{2(m-1)}x$ for $1 \leq m \leq k-i+1$. Thus in total, with probability at least $1-\exp(-\Theta(n^{\beta/2}))$, the number of vertices in total in all these branching processes is at most

$$\sum_{m=1}^{k-i} 2^{2(m-1)}x \leq 2^{2k}x \leq 2^{3k+3}k!C_{k,j,i-1}\frac{t}{n^{k-j+i}} + n^\beta.$$

However, the vertices in the branching process exactly represent (an upper coupling on) the j-sets which can be discovered due to jumps to or new starts at I and the pivots arising from them, which are all of the j-sets containing I which we discover in the `Pathfinder` algorithm. Thus with probability at least $1-\exp(-\Theta(n^{\beta/2}))$, the number of j-sets containing I which became active is at most

$$2^{3k+3}k!C_{k,j,i-1}\frac{t}{n^{k-j+i}} + n^\beta \leq C_{k,j,i}\frac{t}{n^{k-j+i}} + n^\beta,$$

since we chose $C_{k,j,i} \gg C_{k,j,i-1}$. Taking a union bound over all $\binom{n}{i}$ many i-sets I, and observing that $\binom{n}{i}\exp(-\Theta(n^{\beta/2})) = o(1)$, the result follows. □

Proof of Lemma 3.30. The statement of Lemma 3.30 follows directly from Proposition 3.33 and Lemma 3.34. □

3.8 First Moment Method

In this section we prove the upper bounds in all three statements of Theorem 3.1. For convenience, we restate these bounds in the following lemma.

Lemma 3.35. *Let $k, j \in \mathbb{N}$ satisfy $1 \leq j \leq k-1$. Let $a \in \mathbb{N}$ be the unique integer satisfying $1 \leq a \leq k-j$ and $a \equiv k \bmod (k-j)$. Let $\varepsilon = \varepsilon(n) \ll 1$ satisfy $\varepsilon^3 n \xrightarrow{n\to\infty} \infty$ and let*

$$p_0 = p_0(n;k,j) := \frac{1}{\binom{k-j}{a}\binom{n-j}{k-j}}.$$

Let L be the length of the longest j-tight path in $H^k(n,p)$.

(i) *If* $p = \frac{1-\varepsilon}{\binom{k-j}{a}\binom{n-j}{k-j}}$*, then whp*

$$L \leq \frac{j \ln n + \omega}{-\ln(1-\varepsilon)},$$

for any $\omega = \omega(n) \xrightarrow{n\to\infty} \infty$.

(ii) *If* $p = \frac{1+\varepsilon}{\binom{k-j}{a}\binom{n-j}{k-j}}$*, then for any* δ *satisfying* $\delta \gg \max\{\varepsilon, \frac{\ln n}{\varepsilon^2 n}\}$*, whp*

$$L \leq (1+\delta)\frac{2\varepsilon n}{(k-j)^2}.$$

Note that the only difference between this statement and the upper bounds in Theorem 3.1 is that in Theorem 3.1 (iii) we assume $\delta^2\varepsilon^3 n \to \infty$ in place of $\delta \gg \frac{\ln n}{\varepsilon^2 n}$, but it is easy to see that the former condition implies the latter.

Proof. Since

$$\mathbb{P}(L \geq \ell) = \mathbb{P}(\hat{X}_\ell \geq 1) \leq \mathbb{E}(\hat{X}_\ell)$$

by Markov's inequality, it suffices to show that $\mathbb{E}(\hat{X}_\ell) \xrightarrow{n\to\infty} 0$ for the relevant values of ℓ and p.

We first prove the subcritical case (i.e. (i)), so we set $p = \frac{1-\varepsilon}{\binom{k-j}{a}\binom{n-j}{k-j}}$ and $\ell = \frac{j\ln n + \omega}{-\ln(1-\varepsilon)}$. It is convenient to assume that $\omega = o(\ln n)$, which is permissible since the statement becomes stronger for smaller ω. With this assumption we have $\ell = \Theta\left(\frac{\ln n}{\varepsilon}\right) = o(n)$. Then by Corollary 3.5,

$$\begin{aligned}
\mathbb{E}(\hat{X}_\ell) &= \Theta(1)\frac{(n)_v\,(1-\varepsilon)^\ell}{\left(a!b!\binom{k-j}{a}\binom{n-j}{k-j}\right)^\ell} \leq \Theta(1)\frac{n^v(1-\varepsilon)^\ell}{(n-k)^{\ell(k-j)}} \\
&\leq \Theta(1)\left(1+\frac{k}{n-k}\right)^{\ell(k-j)} n^j(1-\varepsilon)^\ell \\
&= \Theta(1)\left(1+O\left(\frac{\ell}{n}\right)\right)\exp(j\ln n + \ell\ln(1-\varepsilon)) \\
&= \Theta(1)(1+o(1))\exp(-\omega) \to 0,
\end{aligned}$$

which completes the proof of (i).

It remains to prove (ii), for which we set $p = \frac{1+\varepsilon}{\binom{k-j}{a}\binom{n-j}{k-j}}$ and $\ell = (1+\delta)\frac{2\varepsilon n}{(k-j)^2}$.

Observe that $v = (k-j)\ell + j = \Theta(\varepsilon n) \le \frac{n}{2}$. By applying Stirling's formula we obtain

$$\begin{aligned}
(n)_v &= \frac{n!}{(n-v)!} \\
&= (1+o(1))\sqrt{\frac{n}{n-v}}\frac{n^v}{e^v}\left(1+\frac{v}{n-v}\right)^{n-v} \\
&= O\left(\frac{n^v}{e^v}\exp\left((n-v)\left(\frac{v}{n-v}-\frac{v^2}{2(n-v)^2}+O\left(\frac{v^3}{(n-v)^3}\right)\right)\right)\right) \\
&= O\left(n^v\exp\left(\frac{-v^2}{2(n-v)}+O\left(\frac{v^3}{n^2}\right)\right)\right) \\
&= O\left(n^v\exp\left(-\frac{\ell^2(k-j)^2+O(\ell)}{2n(1+O(\varepsilon))}+O(\varepsilon^3 n)\right)\right) \\
&= O\left(n^v\exp\left(-\frac{\ell^2(k-j)^2}{2n}+O(\varepsilon^3 n)\right)\right),
\end{aligned}$$

where in the last line we have used the fact that $\ell/n = O(\varepsilon) = O(\varepsilon^3 n)$. Therefore by Corollary 3.5, we have

$$\begin{aligned}
\mathbb{E}(\hat{X}_\ell) &= \frac{O\left(n^v\exp\left(-\frac{\ell^2(k-j)^2}{2n}+O(\varepsilon^3 n)\right)\right)}{(a!b!)^\ell}\left(\frac{1+\varepsilon}{\binom{n-j}{k-j}\binom{k-j}{a}}\right)^\ell \\
&= O\left(n^j\exp\left(O(\varepsilon^3 n)\right)\left(\frac{n^{k-j}\exp\left(\frac{-\ell(k-j)^2}{2n}\right)(1+\varepsilon)}{\left(1+O\left(\frac{1}{n}\right)\right)n^{k-j}}\right)^\ell\right) \\
&= O\left(n^j\exp\left(O(\varepsilon^3 n)\right)\left(1+O\left(\frac{\ell}{n}\right)\right)\left(\exp\Big(-(1+\delta)\varepsilon\Big)(1+\varepsilon)\right)^\ell\right).
\end{aligned}$$

Now recall that $1+O(\ell/n) = 1+O(\varepsilon) = O(1)$, and furthermore

$$\begin{aligned}
\exp(-(1+\delta)\varepsilon)(1+\varepsilon) &= \exp\left(-(1+\delta)\varepsilon+\varepsilon+O\left(\varepsilon^2\right)\right) \\
&= \exp\left(-\delta\varepsilon+O\left(\varepsilon^2\right)\right) \le \exp\left(\frac{-\delta\varepsilon}{2}\right),
\end{aligned}$$

since $\delta \gg \varepsilon$. Therefore

$$\begin{aligned}
\mathbb{E}(\hat{X}_\ell) &= O\left(n^j\exp\left(O\left(\varepsilon^3 n\right)-\ell\delta\varepsilon/2\right)\right) \\
&= O\left(\exp\left(-\Theta\left(\varepsilon^2\delta n\right)+j\ln n\right)\right) \to 0,
\end{aligned}$$

by the fact that $\delta \gg \frac{\ln n}{\varepsilon^2 n}$. This completes case (ii). $\square$

3.9 Longest Paths: Proof of Theorem 3.1

The various statements contained in Theorem 3.1 have now all been proved.

- The upper bounds of statements (i), (ii) and (iii) of Theorem 3.1 follow from Lemma 3.35.
- The lower bound of statement (i) follows directly from Lemmas 3.8 and 3.9.
- The lower bound of statement (ii) is implied by Lemma 3.30, which is identical except that it omits the assumption that $\delta \gg \frac{\ln n}{\varepsilon^2 n}$.
- The lower bound of statement (iii) is precisely Lemma 3.24.

3.10 Concluding Remarks

Theorem 3.1 provides various bounds on the length L of the longest j-tight path, but these bounds may not be best possible. Let us examine each of the three cases in turn.

3.10.1 Subcritical Case

Here we proved the bounds

$$\frac{j \ln n - \omega + 3 \ln \varepsilon}{-\ln(1-\varepsilon)} \leq L \leq \frac{j \ln n + \omega}{-\ln(1-\varepsilon)}.$$

A more careful version of the first moment calculation implies that if $\ell = \frac{j \ln n + c}{-\ln(1-\varepsilon)}$ for some constant $c \in \mathbb{R}$, then the expected number of paths of length ℓ is asymptotically $d \cdot e^c$, where $d = \frac{(a!b!)^\ell}{z_\ell} = \frac{b!(a!b!)^s}{2((k-j)!)^{2s}}$. This suggests heuristically that in this range, the probability that $X_\ell = 0$, i.e. that there are no paths of length ℓ, is a constant bounded away from both 0 and 1, and that in fact the bounds on L are best possible up to the $3 \ln \varepsilon$ term in the lower bound. This term is negligible (and can be incorporated into ω) if ε is constant, but as ε decreases, it becomes more significant. The term arises because as ε decreases, the paths become longer, meaning that there are many more pairs of possible paths whose existences in $H^k(n,p)$ are heavily dependent on one another, and the second moment method breaks down. Thus to remove the $3 \ln \varepsilon$ term in the lower bound requires some new ideas.

3.10.2 Supercritical Case for $j \geq 2$

In this case, we had the bounds

$$(1-\delta)\frac{\varepsilon n}{(k-j)^2} \leq L \leq (1+\delta)\frac{2\varepsilon n}{(k-j)^2}.$$

Since in particular we may assume that $\delta \ll 1$, the upper bound (provided by the first moment method) and the lower bound (provided by the analysis of the `Pathfinder` algorithm) differ by approximately a factor of 2.

One possible explanation for this discrepancy comes from the fact that we do not query a k-set if it contains some explored j-set. As previously explained, this condition is not necessary to guarantee the correct running of the algorithm, but it is fundamentally necessary for our *analysis* of the algorithm, since it ensures that no k-set is queried twice and therefore each query is independent.

Removing this condition would allow us to try out many different paths with the same end (i.e. different ways of reaching the same destination), which could potentially lead to a longer final path since different sets of vertices are used in the current path and are therefore forbidden for the continuation.

It is not hard to prove that the length ℓ of the current path in the modified algorithm would very quickly reach almost $\frac{\varepsilon n}{(k-j)^2}$ (i.e. our lower bound). For each possible way of reaching this, it is extremely unlikely that the path can be extended significantly, and in particular to length $\frac{2\varepsilon n}{(k-j)^2}$. However, since there will be very many of these paths, it is plausible that at least one of them may go on to reach a larger size, and therefore our lower bound may not be best possible.

On the other hand, it could be that our upper bound is not best possible, i.e. that the first moment heuristic does not give the correct threshold path length. This could be because if there is one very long path, there are likely to be many more (which can be obtained by minor modifications), and so we may not have concentration around the expectation.

Therefore further study is required to determine the asymptotic value of L more precisely.

3.10.3 Supercritical Case for $j = 1$

For loose paths, we proved the bounds

$$(1-\delta)\frac{\varepsilon^2 n}{4(k-1)^2} \leq L \leq (1+\delta)\frac{2\varepsilon n}{(k-1)^2},$$

which differ by a factor of $\Theta(\varepsilon)$. In view of the supercritical case for $j \geq 2$, when the longest path is of length $\Theta(\varepsilon n)$ one might naively expect this to be the case for $j = 1$ as well, and that the lower bound is incorrect simply because the proof method is too weak for $j = 1$.

However, this is not the case for graphs, i.e. when $k = 2$ and $j = 1$, when the longest path is indeed of length $\Theta(\varepsilon^2 n)$. The analogous result for general k and $j = 1$ was recently achieved by Cooley, Kang and Zalla [19], who proved an upper bound of approximately $\frac{2\varepsilon^2 n}{(k-1)^2}$ by bounding the length of the longest loose *cycle* (via consideration of an appropriate 2-core-like structure) and using a sprinkling argument. Nevertheless, this leaves a multiplicative factor of 8 between the upper and lower bounds, which it would be interesting to close.

3.10.4 Critical Window

One might also ask what happens when ε is smaller than allowed here, i.e. when $\varepsilon^3 n \not\to \infty$. In the case $j = 1$, the lower bounds in the subcritical and supercritical case, of orders $\frac{\ln(\varepsilon^3 n)}{\varepsilon}$ and $\varepsilon^2 n$ respectively, would both be $\Theta(n^{1/3})$ when $\varepsilon^3 n = \Theta(1)$, which suggests that this may indeed be the correct critical window when $j = 1$. However, for $j \geq 2$, the bounds differ by approximately a factor of $n^{1/3}$ when $\varepsilon^3 n = \Theta(1)$. It would therefore be interesting to examine whether the statement of Theorem 3.1 remains true for $j \geq 2$ even for smaller ε.

Chapter 4

A Sparse Hypergraph Blow-up Lemma

4.1 Main Concepts and Theorem

In this chapter we prove a blow-up lemma for sparse hypergraphs after introducing the key concepts and preliminary definitions. A *hypergraph* H consists of a vertex set $V(H)$ and an edge set $E(H)$ containing subsets of $V(H)$. In particular, we allow singletons and $\varnothing$ to be edges in hypergraphs. This represents a departure from the norm, but this convention turns out to be convenient and its purpose becomes clear later. Write $v(H)$ for the number of vertices of H and $e(H)$ for the number of edges of H. A *subhypergraph* of a hypergraph H is a hypergraph F such that $V(F) \subseteq V(H)$ and $E(F) \subseteq E(H)$; we say that F is a *spanning* subhypergraph of H if $V(F) = V(H)$. The *induced subhypergraph* $H[I]$ of a hypergraph H on a subset $I \subseteq V(H)$ is the subhypergraph of H with vertex set I and edge set $\{e \in E(H) : e \subseteq I\}$. Given a hypergraph H and a subset $I \subseteq V(H)$, we write $H - I$ for the induced subhypergraph $H[V(H) \setminus I]$. If $I = \{x\}$ for some $x \in V(H)$, we drop the set brackets and write $H - x$. For $\ell \in \mathbb{N}$ we say that a hypergraph is *ℓ-uniform* if all its edges have size ℓ. For $\ell \in \mathbb{N}$ and a hypergraph H write $H^{(\ell)}$ for the edge-maximal ℓ-uniform spanning subhypergraph of H. An *embedding* of a hypergraph H into another hypergraph $\mathcal{G}$ is an injective function $\phi : V(H) \to V(\mathcal{G})$ such that $\phi(e) \in E(\mathcal{G})$ for all $e \in E(H)$.

A *complex* is a hypergraph whose edge set E is *down-closed*: if $e \in E$ and $f \subseteq e$ then $f \in E$. A *subcomplex* of a complex H is a subhypergraph of H which is a complex.

Given a complex H and a subset $I \subseteq V(H)$, the *induced subcomplex* of H on I is the induced subhypergraph of H on I. Given $k \in \mathbb{N}$, a *k-complex* is a complex whose edges are of size at most k. The *down-closure complex* of a hypergraph H is the edge-minimal complex with vertex set $V(H)$ of which H is a subhypergraph. The *down-closure complex* of a set S is the down-closure complex of the hypergraph with vertex set S and edge set $\{S\}$.

4.1.1 Weighted Hypergraphs and Homomorphisms

Our primary objective is the embedding of a k-complex H into another k-complex $\mathcal{G}$. It turns out to be convenient to consider $\mathcal{G}$ as a weighted hypergraph in our proof; for this reason we need definitions for a weighted setting. A *weighted hypergraph* consists of a vertex set V and a weight function from the power set of V to the non-negative real numbers. Given a weighted hypergraph $\mathcal{G}$ and a subset $U \subseteq V(\mathcal{G})$, the *weighted induced subhypergraph* $\mathcal{G}[U]$ of $\mathcal{G}$ on U is the weighted hypergraph with vertex set U whose weight function is equal to the weight function of $\mathcal{G}$ on the power set of U. Given a weighted hypergraph $\mathcal{G}$ and a subset $U \subseteq V(\mathcal{G})$, we write $\mathcal{G} - U$ for the induced subhypergraph $\mathcal{G}[V(\mathcal{G}) \setminus U]$. If $U = \{u\}$ for some $u \in V(\mathcal{G})$, we drop the set brackets and write $\mathcal{G} - u$.

A *weighted-k-graph* is a weighted hypergraph $\mathcal{G}$ with a weight function g such that $g(e) = 1$ for all $e \subseteq V(\mathcal{G})$ with $|e| > k$. The *weighted analogue* of a hypergraph $\mathcal{G}$ is the weighted hypergraph on $V(\mathcal{G})$ with weight function

$$g(e) = \begin{cases} 1 & \text{if } e \in E(\mathcal{G}), \\ 0 & \text{otherwise.} \end{cases}$$

In other words, the weight function is the indicator function for the edges. We will not explicitly distinguish between a hypergraph and its weighted analogue as it will be clear from context. We will use the calligraphic letters $\mathcal{D}$, $\mathcal{G}$ and $\mathcal{H}$ for weighted hypergraphs, and the corresponding lower case letters d, g and h for their weight functions.

A *homomorphism* from a hypergraph H to a weighted hypergraph $\mathcal{G}$ is a function $\phi : V(H) \to V(\mathcal{G})$ such that $|\phi(e)| = |e|$ for each $e \in E(H)$. This is represents a generalisation of the usual notion of homomorphism for unweighted graphs to weighted hypergraphs; in weighted hypergraphs, we are concerned about the size and weight of an edge rather than whether that edge is present. The *weight* of ϕ is

$$\mathcal{G}(\phi) := \prod_{e \in E(H)} g(\phi(e)).$$

We emphasise that this product does run over edges of size 0 or 1 in H. If $\mathcal{G}$ is the weighted analogue of an unweighted hypergraph $\mathcal{H}$, then the weight of ϕ is either 0 or 1, taking the latter value if and only if $\phi(e) \in E(\mathcal{H})$ for all $e \in E(H)$. In other words, $\mathcal{G}(\phi)$ acts as an indicator function for whether ϕ is a homomorphism in the unweighted sense. An *embedding* of a hypergraph H into a weighted hypergraph $\mathcal{G}$ is an injective homomorphism from H to $\mathcal{G}$. A *partial homomorphism* from a hypergraph H to a weighted hypergraph $\mathcal{G}$ is a homomorphism from an induced subhypergraph of H to $\mathcal{G}$.

We focus on a *partite setting* as follows. Let J be an index set. Let H be a hypergraph with a partition $\mathcal{X} = \{X_j\}_{j \in J}$ of $V(H)$ indexed by J and $\mathcal{G}$ be a weighted hypergraph with a partition $\mathcal{V} = \{V_j\}_{j \in J}$ of $V(\mathcal{G})$ indexed by J. We call the sets X_j and V_j the *parts* of H and $\mathcal{G}$ respectively. For emphasis we will often call the parts V_j of $\mathcal{G}$ *clusters*. We say that $\mathcal{V}$ and $\mathcal{X}$ are *size-compatible* if $|V_j| = |X_j|$ for all $j \in J$. For $\kappa \geq 1$ we say that $(\mathcal{G}, \mathcal{V})$ is *κ-balanced* if there exists $m \in \mathbb{N}$ such that $m \leq |V_j| \leq \kappa m$ for all $j \in J$. We say that a set of vertices in H (resp. $\mathcal{G}$) is *J-partite* if it contains at most one vertex from each part of H (resp. $\mathcal{G}$). We say that H is *J-partite* if all its edges are J-partite and say that $\mathcal{G}$ is J-partite if, writing g for the weight function of $\mathcal{G}$, we have $g(e) = 0$ for all non-J-partite $e \subseteq V(\mathcal{G})$. For $x \in X_j$ we write V_x to mean V_j. For a J-partite subset $S \subseteq V(H)$ we write $V_S = \prod_{x \in S} V_x$ for the collection of J-partite $|S|$-subsets of $V(\mathcal{G})$ with vertices in $\bigcup_{x \in S} V_x$. For J-partite $S \subseteq V(H)$ the *index* of S is $i(S) := \{j \in J : S \cap X_j \neq \varnothing\}$. For a collection $\mathcal{S}$ of J-partite subsets of $V(H)$ the *index* of $\mathcal{S}$ is the set $\iota(\mathcal{S}) := \{i(S) : S \in \mathcal{S}\}$. For a collection $\mathcal{S}$ of sets let $\bigcup \mathcal{S} := \bigcup_{S \in \mathcal{S}} S$.

Given an index set J, a hypergraph H with its vertex set partitioned into $\mathcal{X} = \{X_j\}_{j \in J}$ and a weighted hypergraph $\mathcal{G}$ with its vertex set partitioned into $\mathcal{V} = \{V_j\}_{j \in J}$, we say that a homomorphism from H to $\mathcal{G}$ is *J-partite* if it maps each X_j into V_j. Given $\ell \in \mathbb{N}_0$ and for each $i \in [\ell]$ a vertex $x_i \in V(H)$ and a subset $U_i \subseteq V_{x_i}$, define

$$\mathcal{G}(H; (U_i)_{i \in [\ell]}, (x_i)_{i \in [\ell]} \mid J; \mathcal{X}, \mathcal{V}) := \left(\prod_{i \in [\ell]} \frac{|V_{x_i}|}{|U_i|} \prod_{j \in J} |V_j|^{-|X_j|} \right) \sum_{\phi} \mathcal{G}(\phi),$$

where the sum is over all J-partite homomorphisms ϕ from H to $\mathcal{G}$ which map x_i into U_i for all $i \in [\ell]$. This is the expected weight of a uniformly random J-partite homomorphism from H to $\mathcal{G}$ which maps x_i into U_i for all $i \in [\ell]$. When $\ell = 0$ we write $\mathcal{G}(H \mid J; \mathcal{X}, \mathcal{V})$. When J, $\mathcal{X}$ and $\mathcal{V}$ are clear from context we often write $\mathcal{G}(H)$ and $\mathcal{G}\big(H; (U_i)_{i\in[\ell]}, (x_i)_{i\in[\ell]}\big)$ instead of $\mathcal{G}(H \mid J; \mathcal{X}, \mathcal{V})$ and $\mathcal{G}\big(H; (U_i)_{i\in[\ell]}, (x_i)_{i\in[\ell]} \mid J; \mathcal{X}, \mathcal{V}\big)$ respectively. We also often say partite homomorphism instead of J-partite homomorphism when J is clear from context. We often drop the tuple brackets when $\ell = 1$ and omit the set brackets for subsets of the form $U_i = \{v_i\}$.

4.1.2 Regularity Lemma

The regularity method typically involves the joint application of a regularity lemma and a blow-up lemma; here we shall state a sparse hypergraph regularity lemma by Allen, Davies and Skokan [8] that serves as a natural accompaniment to our main result — a sparse hypergraph blow-up lemma. To do so we provide some definitions. A partition of a set is *equitable* if each pair of parts differ in size by at most one. We say that a partition $\mathcal{P}$ of a set S *refines* another partition $\mathcal{Q}$ of S if every part of $\mathcal{P}$ is a subset of some part of $\mathcal{Q}$.

Given $k \geq 2$ and a set S, we say that the edges of a collection $\mathcal{S}$ of k-sets in S are *rainbow* for $(k-1)$-uniform hypergraphs $F_1, \ldots, F_k$ on S if the $(k-1)$-sets of any k-set in $\mathcal{S}$ can be labelled using each label in $[k]$ exactly once so that the subset labelled i is in F_i for all $i \in [k]$. A $(k-1)$-*family of partitions* $\mathcal{P}^*$ on S consists of the following. We have a partition $\mathcal{P}$ of S; we call this the *ground partition* of $\mathcal{P}^*$ and its parts 1-*cells*. We will sometimes refer to the 1-cells as *clusters*. Furthermore, for each $2 \leq j \leq k-1$ and each j-set J of 1-cells, we have a *supporting* partition of the j-sets of S with one vertex in each member of J into j-*cells*; we say that a partition is *supporting* if, for each j-cell of $\mathcal{P}^*$, there are j $(j-1)$ cells such that each edge of the given j-cell is rainbow for the chosen $(j-1)$-cells. We will talk about a j-*polyad* in such a family of partitions, by which we mean a choice of j 1-cells, $\binom{j}{2}$ 2-cells, and so on up to $\binom{j}{j-1}$ $(j-1)$-cells, which are supporting in the above sense.

Given a family of partitions $\mathcal{P}^*$ on a set S, we define the *density multicomplex* $\mathcal{D}^*$ on the vertex set $\mathcal{P}$ to have the weight function $d^* : \mathcal{P}^* \to [0, \infty)$ which for a cluster $P \in \mathcal{P}$ has $d^*(P) := 1$ and for each i-cell C with $i \geq 2$ we have

$$d^*(C) := |C|/|Q|,$$

where Q is the collection of all i-sets supported by the $(i-1)$-cells of $\mathcal{P}^*$ which support C. Given in addition a k-uniform hypergraph G on S, the *density multicomplex* of $(G, \mathcal{P}^*)$

is the multicomplex on $\mathcal{P}$ with weight function d^* extending the density multicomplex of $\mathcal{P}^*$ as follows. For each k-polyad Q of $\mathcal{P}^*$, let Q^* denote the set of k-edges supported by Q and let $d^*(Q) = \frac{1}{|Q^*|}\sum_{e\in Q^*} g(e)$. Observe that the density multicomplex keeps track of the (relative) density of each part of our family of partitions. In particular, if Q is a given k-polyad of $\mathcal{P}^*$ on clusters $V_1, \dots, V_k$, then we can write the number of edges of G supported by Q (or, if G is a weighted hypergraph, the sum of the weights of edges of G supported by Q) as

$$|V_1| \dots |V_k| \cdot \prod_{Q'} d^*(Q')\,,$$

where the product over Q' runs over Q and all its supporting j-cells for each $1 \le j \le k-1$. Here we say a $(k-2)$-cell is in the support of Q if it supports one of the $(k-1)$-cells supporting Q, and so on.

Given a k-complex F and a $(k-1)$-family of partitions $\mathcal{P}^*$ on S, we say $\phi : F \to \mathcal{P}^*$ is a *consistent embedding* if for each $e \in E(F)$ with $1 \le |e| \le k-1$, the part $\phi(e)$ is a $|e|$-cell of $\mathcal{P}^*$, if for each $2 \le |e| \le k-1$ the $(|e|-1)$-edges $\{e - \{x\} : x \in e\}$ are mapped bijectively to the $(|e|-1)$-cells which support $\phi(e)$, and if for each $|e| = k$ we have $\phi(e) = Q$, where Q is the unique k-polyad supported by the $(k-1)$-cells $\{\phi(e - \{x\}) : x \in e\}$. Given in addition a k-uniform hypergraph G on S, let $\mathcal{D}^*$ be the density multicomplex of $(G, \mathcal{P}^*)$. We write

$$\mathcal{D}^*(F, \phi) := \prod_{e\in E(F): e\neq\varnothing} d^*(\phi(e))\,.$$

Finally, we say a homomorphism $\psi : F \to G$ is *ϕ-agreeing* if for each $e \in E(F)$ with $1 \le |e| \le k-1$ we have $\psi(e) \in \phi(e)$. We write

$$G(F, \phi) := \Big(\prod_{x\in V(F)} \big|\phi(\{x\})\big|^{-1}\Big) \cdot \sum_{\substack{\psi: F\to G\\ \psi \text{ is } \phi\text{-agreeing}}} \prod_{\substack{e\in E(F)\\ |e|=k}} g(e)\,.$$

We now state a sparse hypergraph regularity lemma by Allen, Davies and Skokan [8]. Recall that $\Gamma(H)$ represents the homomorphism density of H in Γ.

Lemma 4.1 (Allen, Davies and Skokan [8, Lemma 25]). *Let $k \ge 3$ be an integer. For all $q, t_0, c, s \in \mathbb{N}$ and $\varepsilon > 0$, there exist $c^*, t_1, n_0 \in \mathbb{N}$ and $\eta > 0$ such that the following holds for all $n \ge n_0$. Let Γ be a k-uniform hypergraph, let $G_1, \dots, G_s$ be edge-disjoint k-uniform subhypergraphs of Γ and $\mathcal{Q}$ be an equitable partition of $V(\Gamma)$ into q parts.*

Suppose that there is some $p > 0$ such that for any k-uniform hypergraph H on at most c^* vertices we have $\Gamma(H) = (1 \pm \eta)p^{e(H)}$. Then there exists a $(k-1)$-family of partitions $\mathcal{P}^*$ on $V(\Gamma)$, and for each $i \in [s]$ a weighted k-uniform hypergraph G'_i with $g'_i(e) \le g_i(e)$ for each $e \in \binom{V(\Gamma)}{k}$, such that the following hold for each $i \in [s]$, where $\mathcal{D}^*_i$ is the density multicomplex of $(G'_i, \mathcal{P}^*)$.

(a) The ground partition $\mathcal{P}$ of $\mathcal{P}^*$ refines $\mathcal{Q}$ and we have $t_0 \le |\mathcal{P}| \le t_1$.

(b) We have $\sum_{e \in \binom{V(\Gamma)}{k}} (g_i(e) - g'_i(e)) \le \varepsilon p n^k$; if $e \in \binom{V(\Gamma)}{k}$ is such that $g_i(e) > g'_i(e)$, then we have $g'_i(e) = 0$ and for the k-polyad Q supporting e we have $d^*_i(Q) = 0$ and $g'_i(e') = 0$ for all e' supported by Q.

(c) For any k-complex F with at most c vertices and any consistent embedding $\phi : F \to \mathcal{P}^*$, we have $G'_i(F, \phi) = (1 \pm \varepsilon)\mathcal{D}^*_i(F, \phi)$.

(d) For each $j \in [k-1]$ there exists $d_j \in [1/t_1, 1]$ such that for each cell C of $\mathcal{P}^*$ we have $d^*_i(C) = (1 \pm \varepsilon)d_{|C|}$.

The reader may notice that the conclusions of Lemma 4.1 are somewhat different from a 'usual' regularity lemma; these have been adapted to suit the form of our main result and follow from a regularity lemma with a more conventional form [8, Lemma 24] by applying the counting machinery of [8]. Generally, when we apply Lemma 4.1, we will want to regularise one k-uniform unweighted hypergraph, in which case we would take $s = 1$ and g_1 is a function which takes values in $\{0, 1\}$. The hypergraph G'_1 is then a subgraph of G_1 in the usual sense. The reader familiar with hypergraph regularity will recognise that the family of partitions $\mathcal{P}^*$ is standard (and one really needs to consider a family of partitions for the result to be true). Conditions (a) and (d) say respectively that the ground partition refines the given one and is not too large; and that for each j the j-cells are roughly the same size and not too small. The latter is part of a condition sometimes called 'equitability'. Condition (b) says that G'_i is a subhypergraph of G_i: we obtain it by removing the few edges of G_i which are in 'sparse' or 'irregular' k-polyads. Finally condition (c) replaces the usual conditions that the family of partitions, and each G'_i with respect to the family of partitions, should have some regularity property. It is well known that a 2-cell being ε-regular (in the sense of Szemerédi) is implied by a counting condition in terms of the number of 2-edges and copies of the four-cycle C_4, and a similar statement holds for higher uniformities also; all these counting conditions, and more, are given by (c).

4.1.3 Main Theorem

The full version of our main theorem is technically complex and requires additional definitions to state. We shall first provide a simplified version of our sparse hypergraph blow-up lemma to remove some of this complexity and to motivate the extra definitions. It states that for any k-complex $\mathcal{G}$ with a balanced vertex partition $\{V_j\}_{j\in[r]}$ such that we have typical counts and rooted counts of small partite complexes, any compatible bounded degree partite complex H can be embedded into $\mathcal{G}$.

Theorem 4.2 (Allen, Böttcher, Davies, Hng and Skokan [3])**.** *Given $k, \Delta \geq 2$, $r \in \mathbb{N}$ and $\kappa \geq 1$, there exist $c \in \mathbb{N}$, $\eta > 0$ and $n_0 \in \mathbb{N}$ such that the following holds for all $n \geq n_0$. Let H and $\mathcal{G}$ be $[r]$-partite k-complexes on n vertices with κ-balanced size-compatible vertex partitions $\mathcal{X} = \{X_j\}_{j\in[r]}$ and $\mathcal{V} = \{V_j\}_{j\in[r]}$ respectively, such that $\Delta(H^{(2)}) \leq \Delta$, $\varnothing \in E(\mathcal{G})$ and $\{v\} \in E(\mathcal{G})$ for all $v \in V(\mathcal{G})$. Let $\mathcal{D}$ be a weighted hypergraph on $[r]$ with $d(\varnothing) = 1$, $d(\{j\}) = 1$ for all $j \in [r]$ and $d(e) > 0$ for all $e \subseteq [r]$ such that $|e| \leq k$. Let $\Delta_{\mathrm{aux}} := 2^{2^{64(\Delta^6+1)^2 r^2+1}+\Delta^2+1}(\Delta+1)\Delta$. Suppose that the following hold.*

(BLS1) For each $[r]$-partite k-complex F on at most $(\Delta_{\mathrm{aux}}+2)(\Delta+2)c$ vertices we have $\mathcal{G}(F) = (1 \pm \eta)\mathcal{D}(F)$.

(BLS2) For each $[r]$-partite k-complex F on at most $(\Delta^2+\Delta+2)c+1$ vertices with a partition $\mathcal{F} = \{F_j\}_{j\in[r]}$ of $V(F)$, each vertex $x \in V(F)$ and each vertex $v \in V_j$ with $j \in [r]$ such that $x \in F_j$, we have $\mathcal{G}(F; v, x) = (1 \pm \eta)\mathcal{D}(F)$.

Then there is an embedding ϕ of H into $\mathcal{G}$ such that $\phi(x) \in V_x$ for each $x \in V(H)$.

We will refer to the weighted hypergraph $\mathcal{D}$ which appears in the above theorem as the *density k-graph* of $\mathcal{G}$. This is a (small) abuse of notation, in that $\mathcal{D}$ is not defined uniquely by $\mathcal{G}$, but the weights are fixed up to a relative error $1 \pm \eta$. The density k-graph of $\mathcal{G}$ plays the same role as the density multicomplex of $(G, \mathcal{P}^*)$, where G is a k-uniform hypergraph: that is, it keeps track of the relative densities, and (BLS1) states that not only does it keep track of the number of k-edges between a given set of k clusters (obtained by taking F to be the down-closure of a k-edge with vertices in the given k vertices of J), but also counts of all other small complexes. We should stress that we really need to consider all small k-complexes F here, and not just those which are obtained by down-closure of some k-uniform hypergraph.

Theorem 4.2 broadly resembles the graph blow-up lemmas such as that of Komlós, Sárközy and Szemerédi [38]. One should think of (BLS1) as the equivalent of stating that various pairs of sets form graph regular pairs, and (BLS2) as dealing with 'super-regularity'. The latter is rather more complicated than the simple minimum degree condition of [38]; as seen in [4] something more is needed for a blow-up lemma already in sparse 2-graphs.

Before going on to state our full-strength main result, we should comment on the relation between Theorem 4.2 and Lemma 4.1. The latter gives us a family of partitions, but the former works with one single k-complex. The relation here is given by a *regular slice*. That is, given a family of partitions $\mathcal{P}^*$ on S, and a k-uniform hypergraph G, we can create a weighted k-graph $\mathcal{G}$ as follows. For each $x \in \mathcal{P}$ we let $Q_x = \{x\}$. For each $2 \leq i \leq k-1$ in succession, for each i-set f of clusters in $\mathcal{P}$, we pick one i-cell Q_f from $\mathcal{P}^*$ on the clusters f. We insist that Q_f is supported by the $(i-1)$-cells $\{Q_{f\setminus\{x\}} : x \in f\}$. Finally, we let $\mathcal{G}$ be the $\mathcal{P}$-partite k-graph on S whose weight function is defined as follows. We set $g(\varnothing) = 1$, for each $\mathcal{P}$-partite edge e with $1 \leq |e| \leq k-1$ we let $g(e)$ be equal to either zero or one, according to whether e is in one of the chosen Q_f, and finally we let $\mathcal{G}$ and G agree on edges of uniformity k. We refer to $\mathcal{G}$ as a *regular slice through* $(G, \mathcal{P}^*)$.

Observe that $\mathcal{G}$ inherits a density k-graph $\mathcal{D}$ on $\mathcal{P}$ from the density multicomplex $\mathcal{D}^*$ of $(G, \mathcal{P}^*)$ by setting the weight function of $\mathcal{D}$ to be given by

$$d(f) := \begin{cases} d^*(Q_f) & \text{if } 1 \leq |f| \leq k, \\ 1 & \text{otherwise.} \end{cases}$$

Furthermore, if $\phi : F \to \mathcal{P}^*$ is a consistent embedding such that $\phi(e)$ is one of the chosen cells for each $e \in E(F)$, then $G(F, \phi) = \mathcal{G}(F)$, where for the latter we view F as being $\mathcal{P}$-partite with the partition given by ϕ on the singleton edges of F. Thus Theorem 4.2 tells us that if we let $\mathcal{G}$ be induced by some r clusters of a regular slice $\mathcal{G}'$, if none of the k-polyads chosen on these r-clusters have density zero, and if in addition we are given (BLS2), then we can embed any appropriate spanning k-complex H into $\mathcal{G}$, and hence we can also embed the k-uniform hypergraph $H^{(k)}$ into G.

It may not be completely obvious where the consistency conditions (choosing a 3-cell supported on the chosen 2-cells, and so on) mentioned above come into Theorem 4.2. Observe however that if we chose for $\mathcal{G}$ a 3-cell which is *not* supported on the chosen 2-cells, then letting F be the down-closure of a 3-edge assigned to this 3-cell, we have

$\mathcal{G}(F) = 0$, and consequently $\mathcal{D}$ must take the value zero on this cell, which is not permitted. For a similar reason, in the above construction of $\mathcal{G}$ we did not bother to take only the edges of G which are supported by the chosen $(k-1)$-cells: those which are not supported cannot contribute to any $\mathcal{G}(F)$ anyway, because F is down-closed.

It might be surprising that in Theorem 4.2 we simply impose the condition $d(e) > 0$ for all e, without any lower bound that $d(e)$ cannot tend too fast to zero as n goes to infinity. It is however fairly easy to check that if F is the down-closure of a j-edge, and F' is the down-closure of a complete j-partite j-graph with parts of size c, both e-partite for some j-set $e \subset [r]$, then the two equations $\mathcal{G}(F) = (1 \pm \eta)\mathcal{D}(F)$ and $\mathcal{G}(F') = (1 \pm \eta)\mathcal{D}(F')$, both of which we assume to be true, imply that $d(e)$ cannot tend too fast to zero as n tends to infinity. The reason for this is that every e-partite j-edge in $\mathcal{G}$ gives us a homomorphism from F' to $\mathcal{G}$. The number of these homomorphisms is given by $|V_e|\mathcal{G}(F) = (1 \pm \eta)|V_e|\mathcal{D}(F)$, where $|V_e|$ is the total number of e-partite k-sets in $\mathcal{G}$. We therefore have $(1-\eta)|V_e|\mathcal{D}(F) \leq (1+\eta)|V_e|^c\mathcal{D}(F')$, but this equation fails if $d(e)$ tends to zero too fast as n tends to infinity.

How does our full-strength sparse hypergraph blow-up lemma differ from the simplified version? We do not require sufficiently regular and precise counts and rooted counts of small partite complexes in all parts of the partite host graph. Instead, we define the concept of a reduced complex, which encodes where we have counts of small partite complexes that are sufficiently regular and precise, and explain what it means for the complex H which we want to embed to be compatible with a reduced complex. Furthermore, we require only that the maximum degree of the reduced complex be bounded rather than the number of parts in the partition. The motivation for this is that in practical applications one often obtains a large regular structure, typically from a regularity lemma, which does not quite encompass the entire host graph; as such, it is imperative that we are able to avoid the parts of the host graph where the relevant notion of regularity fails and only embed complexes H that are compatible.

Definition 4.3 (Reduced complex)**.** Let R and R' be k-complexes on a set J and $\mathcal{D}$ be a weighted hypergraph on J. Let $c \in \mathbb{N}$ and $\eta > 0$.

(i) Given a J-partite complex H and a partition $\mathcal{X} = \{X_j\}_{j \in J}$ of $V(H)$, we say that $(H, \mathcal{X})$ is an *R-partition* if the set $I \subseteq J$ is an edge of R whenever there are edges in H with index I. We say that H is *R-partite* when we have a fixed partition $\mathcal{X}$ of $V(H)$ such that $(H, \mathcal{X})$ is an R-partition.

(ii) Given a J-partite weighted hypergraph $\mathcal{G}$ and a partition $\mathcal{V} = \{V_j\}_{j\in J}$ of $V(\mathcal{G})$, we say that $(\mathcal{G}, \mathcal{V})$ is an $(\eta, c, \mathcal{D})$-*typcount R-partition* if for each R-partite k-complex F on at most c vertices we have

$$\mathcal{G}(F) = (1 \pm \eta)\tfrac{g(\varnothing)}{d(\varnothing)}\mathcal{D}(F).$$

We call R a *reduced complex* for $(\mathcal{G}, \mathcal{V})$.

(iii) Given a J-partite weighted hypergraph $\mathcal{G}$ and a partition $\mathcal{V} = \{V_j\}_{j\in J}$ of $V(\mathcal{G})$, we say that $(\mathcal{G}, \mathcal{V})$ is $(\eta, c, \mathcal{D})$-*super-typcount on R'* if for each R'-partite k-complex F on at most $c+1$ vertices with its vertex set partitioned into $\mathcal{F} = \{F_j\}_{j\in J}$, each $x \in V(F)$ and each $v \in V_j$ with $j \in J$ such that $x \in F_j$, we have

$$\mathcal{G}(F; v, x) = (1 \pm \eta)\tfrac{g(\varnothing)}{d(\varnothing)}\mathcal{D}(F).$$

We also require the stronger rooted counts only in certain parts of our host graphs, which we encode using a spanning subcomplex R' of R. Here we define the concept of buffer sets in an R-partition $(H, \mathcal{X})$ in relation to a spanning subcomplex R' of R. The purpose of these buffer sets is that a subset of these vertices will be embedded last and we will need the edges of H in the vicinity of vertices in this subset to be associated with edges of R' (and not other edges of R) so that we have precisely estimated counts of small partite complexes rooted at these vertices.

Definition 4.4 (Buffer sets)**.** Let $\alpha > 0$ and $c \in \mathbb{N}$. Let R be a complex on a set J and R' be a spanning subcomplex of R. Let H be a J-partite complex with its vertex set partitioned into $\mathcal{X} = \{X_j\}_{j\in J}$ such that $(H, \mathcal{X})$ is an R-partition. We say that a family $\widetilde{\mathcal{X}} = \{\widetilde{X}_j\}_{j\in J}$ of subsets $\widetilde{X}_j \subseteq X_j$ is an (α, c, R')-*buffer* for $(H, \mathcal{X})$ if

(i) $|\widetilde{X}_j| \geq \alpha |X_j|$ for all $j \in J$, and

(ii) for each $j \in J$ and each $v_0 \in \widetilde{X}_j$, given any vertices $v_1, \ldots, v_c \in V(H)$ and any edges $e_1, \ldots, e_c \in E(H)$ such that $\{v_{i-1}, v_i\} \subseteq e_i$ for all $i \in [c]$, we have $i(e_i) \in E(R')$ for all $i \in [c]$.

We call the vertices in $\widetilde{\mathcal{X}}$ *potential buffer vertices.*

We now state the full version of our sparse hypergraph blow-up lemma. (BUL2) and (BUL3) correspond to (BLS1) and (BLS2) in Theorem 4.2, giving highly precise typical counts and super-typical counts respectively in the parts of $\mathcal{G}$ corresponding to R and R' respectively; (BUL1) requires that H be compatible with R, $R^{(2)}$ have bounded degree and edges of H in the vicinity of its potential buffer set be associated with edges of R'.

Theorem 4.5 (Allen, Böttcher, Davies, Hng and Skokan [3]). *Given $k, \Delta \geq 2$, $\Delta_R \in \mathbb{N}$, $\alpha \in (0,1]$ and $\kappa \geq 1$, there exist $c \in \mathbb{N}$ and $\eta > 0$ such that for every finite set J there exists $n_0 \in \mathbb{N}$ such that the following holds for all $n \geq n_0$. Let R be a k-complex on J and R' be a spanning subcomplex of R. Let H and $\mathcal{G}$ be J-partite k-complexes on n vertices with κ-balanced size-compatible vertex partitions $\mathcal{X} = \{X_j\}_{j \in J}$ and $\mathcal{V} = \{V_j\}_{j \in J}$ respectively, such that $\Delta(H^{(2)}) \leq \Delta$, $\varnothing \in E(\mathcal{G})$ and $\{v\} \in E(\mathcal{G})$ for all $v \in V(\mathcal{G})$. Let $\mathcal{D}$ be a weighted hypergraph on J with $d(\varnothing) = 1$, $d(\{j\}) = 1$ for all $j \in J$ and $d(e) > 0$ for all $e \in E(R)$. Let $\widetilde{\mathcal{X}} = \{\widetilde{X}_j\}_{j \in J}$ be a family of subsets of $V(H)$. Let $\Delta_{\mathrm{aux}} := 2^{2^{64(\Delta^6+1)^2\Delta_R^2+1}+\Delta^2+1}(\Delta+1)\Delta$. Suppose that*

(BUL1) $(H, \mathcal{X})$ *is an R-partition, $\widetilde{\mathcal{X}}$ is an $(\alpha, (\Delta^2+\Delta+2)c, R')$-buffer for $(H, \mathcal{X})$ and $\Delta(R^{(2)}) \leq \Delta_R$,*

(BUL2) $(\mathcal{G}, \mathcal{V})$ *is an $(\eta, (\Delta_{\mathrm{aux}}+2)(\Delta+2)c, \mathcal{D})$-typcount R-partition,*

(BUL3) $(\mathcal{G}, \mathcal{V})$ *is $(\eta, (\Delta^2+\Delta+2)c, \mathcal{D})$-super-typcount on R'.*

Then there is an embedding ϕ of H into $\mathcal{G}$ such that $\phi(x) \in V_x$ for each $x \in V(H)$.

Again, a comment on the relation between Theorem 4.5 and Lemma 4.1 is in order. If $\mathcal{G}'$ is a regular slice through $(G, \mathcal{P}^*)$ with density k-graph $\mathcal{D}'$, then Lemma 4.1 guarantees for us that we have (BUL2), and also that $d'(e) > 0$ for all edges e with $|e| \leq k-1$. However we generally will *not* be able to guarantee that $d'(e) > 0$ for all e with $|e| = k$; there will generally be some irregular and sparse k-polyads. What we then do is to let R be a k-complex whose k-edges e have $d'(e) > 0$, and we then define $\mathcal{G}$ and $\mathcal{D}$ by, for each e such that $d'(e) = 0$, setting $d(e) = 1$ and $g(f) = 1$ for each e-partite k-set f. It is easy to check that $\mathcal{G}$ also satisfies (BUL2), and an embedding of any R-partite H into $\mathcal{G}$ does not use the edges which we changed from $\mathcal{G}'$ and hence is also an embedding of H into $\mathcal{G}'$. Furthermore, at least in some situations it becomes easy to obtain (BUL3). For example, if H is such that an α-fraction of vertices $\widetilde{\mathcal{X}}$ of each part of H are isolated, then we can take R' to be the k-complex on J which contains all possible 1-edges and no edges of larger uniformity and we obtain an $(\alpha, (\Delta^2 + \Delta + 2)c, R')$-buffer. In this particular situation, (BUL3) asks for rooted counts of k-complexes F which have no edges of uniformity 2 or greater. These rooted counts are all equal to 1, so (BUL3) holds. This observation justifies that we can use Theorem 4.5 directly to embed almost-spanning graphs into regular partitions as produced by Lemma 4.1. We cannot (as one sees already for the graph blow-up lemma) expect a regularity lemma to give us something suitable for spanning embeddings of complexes without isolated vertices.

4.1.4 Applications

In this section we provide some relatively simple applications of our new blow-up lemma to demonstrate how it can be used. We start with a result concerning biased Maker-Breaker games on hypergraphs, and then turn to a result on size Ramsey numbers for bounded degree hypergraphs.

4.1.4.1 Maker-Breaker games

In a $(1 : b)$ biased Maker-Breaker game, we are given a finite ground set X and a collection $\mathcal{F} \subset \mathcal{P}(X)$ of *winning sets*. Alternately, Maker claims up to 1, and then Breaker up to b, of the elements of X, until no unclaimed elements of X remain. Maker wins if she has claimed all the elements of any winning set (and perhaps some further elements), and Breaker wins otherwise. Since this is a finite game of perfect information, it is determined: one of the two players has a winning strategy with best play. The *threshold bias* b^* of the game is defined to be the smallest natural number b such that Breaker wins the $(1 : b)$-game; assuming $\emptyset \notin \mathcal{F}$, this number is well-defined.

In particular, given k and n, if H is any k-uniform hypergraph, we can take X to be the edges of $K_n^{(k)}$ and $\mathcal{F}$ to be the edge sets of all isomorphic copies of H in $K_n^{(k)}$. Thus Maker wins this H*-game* if the edges of $K_n^{(k)}$ she eventually claims contain an isomorphic copy of H.

The H-game is fairly well understood when H is a fixed 2-graph and n is large. In particular, Bednarska and Łuczak [12] determined the order of magnitude of the threshold bias (though even for $H = K_3$, where the threshold bias is $\Theta(n^{1/2})$, we do not know the constant multiplying $n^{1/2}$), and their methods extend to give a lower bound on the threshold bias also for fixed k-uniform hypergraphs. However when H depends on n, much less is known. The threshold bias for the Hamiltonicity game in graphs was determined by Krivelevich [40], and recently Liebenau and Nenadov [44] found asymptotically the threshold bias for the K_r-factor (that is, $\frac{n}{r}$ vertex-disjoint copies of K_r). There is also a general lower bound on the threshold bias for any bounded-degree graph on up to n vertices, which is a consequence of the Sparse Blow-up Lemma, due to Allen, Böttcher, Kohayakawa, Naves and Person [7].

As an application of Theorem 4.5, we prove the following general lower bound on the threshold bias for the H-game, where H is any almost-spanning bounded-degree k-uniform hypergraph.

Theorem 4.6. *Given integers $\Delta, k \geq 2$ and $\gamma > 0$ there exists a constant $\nu > 0$ such that the following holds for all sufficiently large n. Let H be any $(1-\gamma)n$-vertex k-uniform hypergraph with $\Delta(H^{(2)}) \leq \Delta$, and let $b = n^{\nu}$. Then Maker wins the $(1:b)$ H-game on $K_n^{(k)}$.*

For convenience we write $H^{(2)}$ for the 2-level of the k-complex we obtain from H by down-closure. Thus the degree of v in $H^{(2)}$ is the number of vertices of $V(H) \setminus \{v\}$ which are in some k-edge with v.

The proof of this theorem is rather similar to the deduction of the $k = 2$ version of this result in [4]. Namely, we show that Maker has a randomised strategy that wins against any given Breaker strategy with positive probability. If Breaker had a winning strategy, then this would be impossible (Maker would always lose against Breaker's winning strategy) and hence Breaker does not have a winning strategy. Since the game is determined, it follows that Maker has a deterministic winning strategy. In our proof of Theorem 4.6 we shall use the celebrated theorem of Hajnal and Szemerédi [30] on graphs, Theorem 2.4.

Proof of Theorem 4.6. Given Δ and k, we let c, η be returned by Theorem 4.5 for input k, Δ, $\Delta_R = \Delta$, $\alpha = \frac{1}{2}\gamma$ and $\kappa = 2$. We set

$$\nu = \tfrac{1}{2}\binom{c+k}{k}^{-1}, \qquad s = \Delta + 1 = \Delta_R + 1\,, \quad \text{and} \quad J = [s]\,,$$

and let n_0 be given by Theorem 4.5 for this J (and the other constants specified before). Now let $\nu = \frac{1}{2}\binom{c+k}{k}^{-1}$, let $\varepsilon \ll \eta$ be sufficiently small and $n \geq n_0$ be sufficiently large for the following calculations and set

$$b = n^{\nu}\,, \qquad \ell = \frac{\varepsilon}{b}\binom{n}{k}\,.$$

Our goal is to argue that for *any* strategy Breaker uses there is a randomised Maker strategy which wins against this strategy with positive probability. A Breaker strategy by definition is a rule which, given the edges claimed by respectively Maker and Breaker in their previous turns, outputs the edges that Breaker should claim in the current turn; in particular, to define a Breaker strategy, we do not need to know H.[1] So, fix any such Breaker strategy.

For defining Maker's randomised strategy, fix any partition $V(K_n^{(k)}) = V_1 \cup \cdots \cup V_s$ with parts of sizes differing by at most one. Let Q denote the complete partite k-uniform hypergraph with parts $V_1, \ldots, V_s$. Maker shall only claim edges of Q and ignore all other edges – the reason for working in this partite setting is that it will give us, as we will show, a regular slice with the trivial family of partitions in which the ground partition is $\{V_1, \ldots, V_s\}$ and there is only one j-cell, for each $2 \leq j \leq k-1$, on any given j set of vertices and this j-cell contains all the partite j-edges. We shall later partition the target hypergraph H into s almost equally-sized independent sets $X_1, \ldots, X_s$ and embed it into the graph claimed by Maker, mapping each X_i to V_i.

Maker's strategy is now the following. She randomly orders the edges of Q, and in her ith turn tries to claim the ith edge in her list; if this edge was previously claimed by Breaker, she claims no edge in that turn.

It remains to argue that this strategy succeeds with positive probability. Indeed, let Γ be the hypergraph of the first ℓ edges in Maker's list, and G the subhypergraph of edges which Maker successfully claimed. Observe that Γ is distributed as the uniform random ℓ-edge subhypergraph of Q. We would like to apply our blow-up lemma to embed H into G, with partition $\mathcal{V} = \{V_j\}_{j\in[s]}$. For this, we have to show that G satisfies the assumptions of this lemma with positive probability.

We first claim that

$$e(G) \geq (1-3\varepsilon)\ell\,.$$

To see this, observe that in the ith turn, Maker chooses uniformly at random from the edges of Q which she has not previously chosen. Of these, at most $ib \leq \ell b = \varepsilon\binom{n}{k}$ were chosen by Breaker in previous rounds, and hence Maker's probability of picking a claimed edge is at most 2ε. The total number of edges Maker fails to claim is therefore

[1] When playing, the strategy S_H Breaker chooses will of course depend on H; but since Maker is successful against all strategies, Maker is in particular sucessful against S_H. This assumption that we do not know H when fixing the Breaker strategy will be important to argue, as we do after this proof, that we actually show something slightly stronger than Theorem 4.6.

stochastically dominated by $\mathrm{Bin}(\ell, 2\varepsilon)$, which with high probability by Chernoff's inequality is at most $3\varepsilon\ell$.

Hence G retains most edges of Γ. As we will show, this implies that we get good bounds on counts of partite k-complexes everywhere in our partition and hence can choose the complete k-complex on $[s]$ as our reduced k-complex R.

For applying suitable concentration inequalities in the following to establish these counts it is inconvenient that Γ is not distributed like a binomial random hypergraph, but rather has a fixed number ℓ of edges. To get around this, we shall now sandwich Γ between two binomially distributed hypergraphs Q_{p^-} and Q_{p^+}. For this purpose, set

$$p = \frac{\ell}{e(Q)}\,, \qquad p^+ = (1+\varepsilon)p\,, \quad \text{and} \quad p^- = (1-\varepsilon)p\,.$$

Note that $p = \Theta(n^{-\nu})$. With high probability, when we choose edges of Q independently with probability p^- to obtain Q_{p^-}, we obtain less than ℓ edges; when we choose with probability p^+ to obtain Q_{p^+}, we obtain more than ℓ edges. There is then a standard coupling $Q_{p^-} \subseteq \Gamma \subseteq Q_{p^+}$ which succeeds with high probability. Namely, choose $e(Q_{p^-})$ and $e(Q_{p^+})$ from the binomial distributions $\mathrm{Bin}(e(Q), p^-)$ and $\mathrm{Bin}(e(Q), p^+)$ respectively, and fail if we do not obtain $e(Q_{p^-}) \le \ell \le e(Q_{p^+})$. If we do not fail, then choose Q_{p^-} by selecting $e(Q_{p^-})$ edges uniformly at random, Γ by adding $\ell - e(Q_{p^-})$ further edges uniformly at random, and Q_{p^+} by adding a further $e(Q_{p^+}) - \ell$ uniform random edges.

We can now state more precisely in what setup we apply our blow-up lemma. We view each of the four partite k-uniform hypergraphs Q, Γ, Q_{p^-}, Q_{p^+} as k-complexes by adding all partite edges of uniformity less than k, and let the density k-graph $\mathcal{D}$ be the weighted-k-graph on $[s]$ in which all edges of uniformity less than k have weight 1, and all edges of uniformity k have weight p. As explained earlier, our reduced k-complex R is the the complete k-complex on $[s]$. Further, we choose the k-complex on $[s]$ whose edge set is $\{\emptyset, 1, 2, \dots, s\}$ as R'. That is, R' contains all edges of size 1 and smaller, and no larger edges.

We can now turn to verifying that the different assumptions of our blow-up lemma are satisfied. First observe that (BUL3) holds trivially. That is, the only k-complexes F which are R'-partite are 1-complexes. Since Q is identically equal to 1 on 1-edges, we see $Q(F) = 1$ for any 1-complex F, as required for (BUL3).

We shall next verify that (BUL2) holds. Let F be any k-complex as in (BUL2). By a minor modification of a theorem of Kim and Vu [35, Theorem 4.3.1], the number of embeddings of F into each of Q_{p^-} and Q_{p^+} is within a $(1 \pm \frac{1}{2}\eta)$-factor of their expectations. Specifically, the theorem as stated there refers to a random subgraph of $K_n^{(k)}$; however all of the expectations computed in the proof there are upper bounds for the corresponding expectations in our setting, and in addition the expected number of F-copies in Q_{p^-} and Q_{p^+} is of the same order of magnitude as in a p^+-random subgraph of $K_n^{(k)}$, so the proof applies in our setting also. Thus, by definition of $\mathcal{D}$, with high probability we have $Q_{p^-}(F), Q_{p^+}(F) = (1 \pm \frac{1}{2}\eta)\mathcal{D}(F)$ for all the k-complexes F of (BUL2), and so the same applies to Γ.

On the other hand, for any given such F, the number of embeddings of F using a given k-edge e in Q_{p^+} is stochastically dominated by the number of copies of F using e in the p^+-random subgraph of $K_n^{(k)}$. As proved by Kim and Vu [35, Theorem 4.2.4], with high probability for all edges e this quantity is at most $(p^+)^s n^t$, where F has $s+1$ k-edges and $t+k$ vertices.

Suppose that all the above mentioned likely events occur. Since G has at most $3\varepsilon\ell \le 3\varepsilon p^+ n^k$ edges fewer than Γ, it has at most $3\varepsilon(p^+)^{s+1}n^{t+k}$ fewer embeddings of F than Γ. We claim that the embeddings of F into G make up almost all of the homomorphic copies of F counted by $G(F)$: to see this, observe that the number of homomorphic copies of F is $\Theta(n^{v(F)}p^{e_k(F)})$, where $e_k(F) \le \binom{c}{k}$ counts the number of k-edges of F. By choice of ν, this is $\Omega(n^{v(F)-1/2})$, whereas trivially any homomorphic copy of F in G which is not an embedding uses at most $v(F)-1$ vertices of G, and so there are at most $n^{v(F)-1}$ such. By choice of ε, and since n is sufficiently large, we see that $G(F) = (1 \pm \eta)\mathcal{D}(F)$, verifying (BUL2).

It remains to verify (BUL1). To this end, let H be any k-uniform hypergraph with $\Delta(H^{(2)}) \le \Delta = \Delta_R$. and at most $(1-\gamma)n$ vertices. We view this as a k-complex by taking the down-closure. By the Hajnal–Szemerédi Theorem, Theorem 2.4, there is a partition of $V(H)$ into $(k-1)\Delta + 1$ parts $X_1, \dots, X_s$ which differ in size by at most one and such that all the edges of $H^{(2)}$ (and so all the edges of H) are partite. It follows that $\{X_i\}_{i\in[s]}$ is an R-partition. Since we chose $\alpha = \frac{1}{2}\gamma$, we have $|X_i| \le (1-\alpha)|V_i|$ for each i. We now add to each X_i exactly $|V_i| - |X_i|$ isolated vertices and let these form the buffer set $\tilde{X}_i$. Obviously, after adding these buffer sets we still have an R-partition. Since the buffer sets contain only independent vertices and $|\tilde{X}_i| = |V_i| - |X_i| \ge \alpha|V_i|$ we have that $\{\tilde{X}_i\}_{i\in[s]}$ is an (α, c', R')-buffer for any c', and hence in particular for $c' = (\Delta^2 + \Delta + 2)c$. Hence (BUL1) holds.

We conclude that we can apply Theorem 4.5 and hence H is a subgraph of G, as desired. $\square$

Note that this proof actually gives a slightly stronger conclusion than Theorem 4.6 claims: Maker actually ends up claiming a k-uniform hypergraph which contains not just any one H satisfying the conditions of the theorem, but all of them simultaneously (i.e. it is *universal*). To the best of our knowledge, previous to this result it was not even known that Maker has a winning strategy in the $(1 : b)$ H-game for any connected hypergraph H with $v(H) = \Theta(n)$ and any b growing with n (for constant b the result follows from Keevash [33]).

4.1.4.2 Size Ramsey numbers

The *ℓ-colour size Ramsey number* $\hat{r}_\ell(H)$ of a k-uniform hypergraph H is defined to be the minimum of $e(\Gamma)$ over k-uniform hypergraphs Γ with the following property: however the edges of Γ are ℓ-coloured, one of the colour classes contains a subgraph isomorphic to H. In this case, we also say that Γ is *ℓ-colour size Ramsey* for H.

We have the trivial bound $\hat{r}_\ell(H) \le \binom{r_\ell(H)}{k}$, where $r_\ell(H)$ is the usual ℓ-colour Ramsey number, since a complete graph on $r_\ell(H)$ vertices by definition has the desired property. It was proved by Cooley, Fountoulakis, Kühn and Osthus [15] that when H is an n-vertex k-uniform hypergraph with maximum degree at most Δ, there is a constant C depending on k, ℓ and Δ such that $r_\ell(H) \le Cn$, from which it follows $\hat{r}_\ell(H) = O(n^k)$.

For $k = 2$, i.e. graphs, Rödl and Szemerédi [52] proved that for some graphs H with $\Delta(H) = 3$ we have $\hat{r}_2(H) = \omega(n)$, and conjectured that there is $\varepsilon > 0$ such that for some H with n vertices and $\Delta(H) \le \Delta$ we have $\hat{r}_2(H) \ge n^{1+\varepsilon}$, and for all H with $\Delta(H) \le \Delta$ we have $\hat{r}_\ell(H) = O(n^{2-\varepsilon})$, where ε depends on Δ only. The former conjecture remains open, but the latter was proved by Kohayakawa, Rödl, Schacht and Szemerédi [37], who showed it holds with any $\varepsilon < \frac{1}{\Delta}$. This bound, which is generally believed to be rather far from optimal, nevertheless remains the state of the art.

For $k \ge 3$, to the best of our knowledge there was no result improving on the bound $\hat{r}_\ell(H) = O(n^k)$ mentioned above. Our blow-up lemma allows the following polynomial improvement.

Theorem 4.7. *For every k and Δ, there exists $\rho > 0$ such that the following holds for each constant ℓ and all sufficiently large n. For any n-vertex k-uniform hypergraph H*

with $\Delta(H^{(2)}) \le \Delta$*, we have* $\hat{r}_\ell(H) \le n^{k-\rho}$*.*

Proof. We shall show, by applying the regularity lemma, Lemma 4.1, and our blow-up lemma, Theorem 4.5, that the random k-uniform hypergraph $H^k(Cn, p)$ with $p = n^\rho$ for suitable C and ρ is ℓ-colour size-Ramsey for H with the stated properties asymptotically almost surely. We first need to fix the constants.

Given k and Δ, let $\eta > 0$ and c be returned by Theorem 4.5 for input k, Δ, $\Delta_R = \Delta$, $\alpha = \frac{1}{4}$ and $\kappa = 2$. Without loss of generality, we may assume η is small enough that

$$\binom{r_\ell(K^{(k)}_{\Delta+1})}{k} \cdot (1+\eta)^{2^{k+1}} \cdot 2k! \cdot \ell\eta \le \tfrac{1}{2}\,,$$

where we recall that $r_\ell(K^{(k)}_{\Delta+1})$ is the ℓ-colour Ramsey number of the complete k-uniform hypergraph on $\Delta+1$ vertices. Let c^*, t_1, n_0 and η^* be returned by Lemma 4.1 for input k, $q = 1$, $t_0 = r_\ell(K^{(k)}_{\Delta+1})$, c, $s = \ell$ and $\varepsilon = \eta$. We set $C = 2t_1$. We choose $\rho' > 0$ such that the following holds. For every k-uniform hypergraph F with at most c^* vertices, if $p = n^{-\rho'}$, asymptotically almost surely the number of labelled copies of F in the binomial random k-uniform hypergraph $H^k(Cn, p)$ is $(1 \pm \frac{1}{2}\eta^*)(Cn)^{v(F)}p^{e(F)}$. As in the proof of Theorem 4.6, this is possible by a theorem of Kim and Vu [35]. Without loss of generality we can assume $\rho' \le \frac{1}{2}$ is sufficiently small that $p^{\binom{c^*}{k}}n \ge n^{0.5}$. Set $\rho = \rho'/2$.

We further need the following property of $H^k(Cn, p)$, which will allow us later to argue that the number of so-called "bad" polyads is small. Suppose that $A_1, \dots, A_k$ are any edge-disjoint $(k-1)$-uniform hypergraphs on $[Cn]$ and that $Q = \mathrm{Rainbow}(A_1, \dots, A_k)$ is the collection of k-sets which contain one $(k-1)$-edge from each A_j. Then we require:

(Rain) For all choices of $Q = \mathrm{Rainbow}(A_1, \dots, A_k)$ with $|Q| \ge n^{k-0.25}$, the number of edges of $H^k(Cn, p)$ in Q is $(1 \pm \eta)p|Q|$.

The expected number of edges of $H^k(Cn, p)$ which lie in Q is $p|Q|$, and by Chernoff's inequality we see that with probability $1 - \exp(-\frac{1}{3}\eta^2 p|Q|)$ the actual number of edges is $(1 \pm \eta)p|Q|$. In particular, taking the union bound over the at most $2^{k(Cn)^{k-1}}$ choices of $A_1, \dots, A_k$, we see that (Rain) holds asymptotically almost surely.

Now, fix $\Gamma = H^k(Cn, p)$ and assume that the following three properties, which are true asymptotically almost surely, hold: Γ has at most $2p\binom{Cn}{k} \le (Cn)^{k-\rho'} \le n^{k-\rho}$ edges, property (Rain) holds, and we have:

(Count) For every k-uniform hypergraph F with at most c^* vertices, the number of labelled copies of F in Γ is $(1 \pm \frac{1}{2}\eta^*)(Cn)^{v(F)}p^{e(F)}$.

We claim that however $E(\Gamma)$ is ℓ-coloured, there is some colour class which contains any given n-vertex k-uniform hypergraph with the stated maximum degree. Since $|E(\Gamma)| \leq n^{k-\rho}$, this proves the theorem.

Indeed, let us start by checking that Γ satisfies the counting condition of Lemma 4.1. By (Count) and since there are at most $v(F)^2 \cdot (Cn)^{v(F)-1}$ homomorphisms which are not injective, which is by assumption much smaller than $(Cn)^{v(F)}p^{e(F)}$, we have $\Gamma(F) = (1 \pm \eta^*)p^{e(H)}$. This shows that we can apply Lemma 4.1. So, given any colouring of $E(\Gamma)$, we apply Lemma 4.1, with input as above, with G_i being the k-uniform hypergraph of edges of colour i, for each $1 \leq i \leq \ell$. We obtain a family of partitions $\mathcal{P}^*$ and weighted k-uniform hypergraphs G'_i satisfying the conclusions of that lemma. Let d^*_i be the corresponding weight function of the density multicomplex of $(G'_i, \mathcal{P}^*)$.

This family of partitions and hypergraphs G'_i, with their associated density multicomplexes is not yet suitable as an input for Theorem 4.5, which requires that we choose (a subhypergraph of) one of the colours G'_i as input and are only allowed to have a vertex partition, a density complex, and a reduced complex. How we choose an appropriate colour shall become clear later, but for getting from the multi-complex setting to the complex setting we select a regular slice (after restricting ourselves to $r_\ell(K^{(k)}_{\Delta+1})$ clusters) as follows. We select uniformly at random a collection of $r_\ell(K^{(k)}_{\Delta+1})$ clusters from the ground partition $\mathcal{P}$ of $\mathcal{P}^*$. We remark that by our choice of η the number of clusters in the ground partition $\mathcal{P}$ is usually much larger than the number $r_\ell(K^{(k)}_{\Delta+1})$ of clusters that we select in this way. The selected clusters form a vertex partition $\mathcal{V}' = \{V_j\}_{j \in J'}$. We shall use a subset of these clusters for Theorem 4.5. Between each pair of clusters from $\mathcal{V}'$, we then select one of the 2-cells of $\mathcal{P}^*$ supported by this pair of clusters uniformly at random; for each triple of selected clusters with their corresponding triple of selected 2-cells we then select one of the 3-cells of $\mathcal{P}^*$ supported by this triple of 2-cells from $\mathcal{P}^*$ uniformly at random; and so on up to $(k-1)$-cells. This selects in total $\binom{r_\ell(K^{(k)}_{\Delta+1})}{k}$ different k-polyads from $\mathcal{P}^*$. Call the collection of these selected polyads $\mathcal{S}$. By Lemma 4.1(d) the probability of having any given polyad Q in $\mathcal{S}$ is thus

$$\binom{r_\ell(K^{(k)}_{\Delta+1})}{k}\cdot(1\pm\eta)^{2^k}\binom{|\mathcal{P}|}{k}^{-1}\prod_{j=1}^{k-1}d_j^{\binom{k}{j}}\,.$$

Some polyads Q of our original partition $\mathcal{P}^*$ may be dense in none of our colours i, that is, $d_i^*(Q)$ is small for all i. These polyads are "bad" in the sense that we cannot use them for embedding H. More precisely, we mark a k-polyad Q of $\mathcal{P}^*$ as *bad* if there does not exist $1\le i\le\ell$ such that $d_i^*(Q)\ge\frac{p}{2\ell}$. We shall now show that very few polyads of $\mathcal{P}^*$ are bad and conclude that we can select the polyads in $\mathcal{S}$ so that none of them is bad by choice of our regularity parameter η. Indeed, let Q be any polyad of $\mathcal{P}^*$. By Lemma 4.1(c) applied with F being the down-closure of a $(k-1)$-uniform k-clique, the number of k-sets supported by Q is

$$(1\pm\eta)\Big(\frac{Cn}{|\mathcal{P}|}\Big)^k\prod_S d_1^*(S)\,,\tag{4.1}$$

where the product over S runs over all cells of each size from 1 to $k-1$ supporting Q, and we note that this quantity depends only on $\mathcal{P}^*$ and not on G_1 because we are only considering cells S of size up to $k-1$. Note that these supported k-sets are of the rainbow form (with the supporting graphs being the $(k-1)$-cells), and there are more than $n^{k-0.25}$ of them. Hence, by (Rain) the number of edges of Γ supported by Q is

$$(1\pm2\eta)p\Big(\frac{Cn}{|\mathcal{P}|}\Big)^k\prod_S d_1^*(S)\,.$$

Since the G_i partition Γ, we conclude that there is i such that the number of edges of G_i supported by Q is at least

$$(1-2\eta)\frac{p}{l}\Big(\frac{Cn}{|\mathcal{P}|}\Big)^k\prod_S d_1^*(S)\,.$$

Together with (4.1) this implies that the density of G_i relative to Q is at least $(1-10\eta)\frac{p}{\ell}$. Hence, if Q is bad then G_i' differs from G_i on Q and therefore $d_i^*(Q)=0$ by Lemma 4.1(b). Since the total number of edges $E(G_i)\setminus E(G_i')$ is at most $\eta p(Cn)^k$ by Lemma 4.1(b) for each $1\le i\le\ell$, we see that the number of bad polyads Q is at most

$$\frac{\ell\eta p(Cn)^k}{(1-2\eta)\cdot\frac{1}{\ell}\cdot p(\frac{Cn}{|\mathcal{P}|})^k\prod_S d_1^*(S)}\le(1+2\eta)^{2^k}\ell\eta|\mathcal{P}|^k\prod_{j=1}^{k-1}d_j^{-\binom{k}{j}}\,,$$

where the numbers $d_1, \dots, d_{k-1}$ are as in Lemma 4.1(d). By linearity of expectation, the expected number of bad polyads selected for $\mathcal{S}$ is at most

$$\binom{r_\ell(K^{(k)}_{\Delta+1})}{k} \cdot (1+\eta)^{2^k} \binom{|\mathcal{P}|}{k}^{-1} \prod_{j=1}^{k-1} d_j^{\binom{k}{j}} \cdot (1+2\eta)^{2^k} \ell\eta |\mathcal{P}|^k \prod_{j=1}^{k-1} d_j^{-\binom{k}{j}}$$
$$\leq \binom{r_\ell(K^{(k)}_{\Delta+1})}{k} \cdot (1+\eta)^{2^{k+1}} \cdot 2k! \cdot \ell\eta < 1\,,$$

where the final inequality is by choice of η. In particular, with positive probability none of our chosen k-polyads are bad. Fix such a choice $\mathcal{S}$.

We next want to determine the colour and the subset $\mathcal{V}$ of our selected clusters $\mathcal{V}'$ with which we want to apply Theorem 4.5. For this purpose, we draw an auxiliary ℓ-coloured complete k-uniform hypergraph with vertex set $\mathcal{V}'$ as follows. We put an edge of colour i on a given k-set of clusters from $\mathcal{V}'$ which supports the chosen k-polyad Q, where $i \in [\ell]$ is minimal such that $d_i^*(Q) \geq \frac{1}{2\ell}$. Since Q is not bad, such an index exists. By definition of $r_\ell(K^{(k)}_{\Delta+1})$, there are some $\Delta+1$ clusters, and a colour $\chi \in [\ell]$, such that all k-edges of our auxiliary hypergraph between these $\Delta+1$ colours are of colour χ. Let J index these $\Delta+1$ clusters. Suppose without loss of generality that $J = [\Delta+1]$, so the clusters of interest are $\mathcal{V} = \{V_i\}_{i\in[\Delta+1]}$. Let $\mathcal{G}$ be the J-partite k-complex on $V_1, \dots, V_{\Delta+1}$ obtained by taking all j-cells of polyads in $\mathcal{S}$ for each $2 \leq j \leq k-1$, and the k-edges of G_χ supported by the polyads in $\mathcal{S}$. Observe that by the choice of the colouring of our auxiliary hypergraph we have $d^*_\chi(Q) \geq \frac{1}{2\ell}$ for every polyad Q from $\mathcal{S}$ on J and hence, by Lemma 4.1(b), we conclude that G_χ and G'_χ are identical on $\mathcal{V}$. Moreover, by choice of t_0 and t_1 the complex $\mathcal{G}$ has at least $2n$ vertices. We let $\mathcal{D}$ be the density k-graph obtained from $\mathcal{D}^*_\chi$.

Now, given any k-uniform hypergraph H' with at most n vertices, we view it as a complex by down-closure and suppose $\Delta(H'^{(2)}) \leq \Delta$. We want to check the conditions of Theorem 4.5 so that we can use this theorem to find a copy of H' in G_χ. We start with (BUL1). For this we colour $V(H')$ equitably with at most $\Delta+1$ colours by applying Theorem 2.4 to $H'^{(2)}$, and let the colour classes be $X'_1, \dots, X'_{\Delta+1}$. This makes H' a J-partite k-complex. Letting our reduced complex R be the complete k-complex on J, we see that H' is R-partite. We enlarge H' to an R-partite complex H by adding to each part X'_i a set of $|V_i| - |X'_i|$ isolated vertices, and let the parts of H be $X_1, \dots, X_{\Delta+1}$. We let $\widetilde{X}_i$ be the set of isolated vertices in X_i, for each $i \in [\Delta+1]$, and let R' be the k-complex on J containing the empty set, all edges of size one, and no other edges. This gives us a $(\frac{1}{4}, (\Delta^2+\Delta+2)c, R')$-buffer for $(H, \widetilde{X})$, and so we have verified (BUL1).

By definition, if F is any J-partite k-complex with at most c vertices, then $\mathcal{G}(F) = G_\chi(F,\phi)$, where ϕ is the consistent embedding of F into $\mathcal{P}^*$ which maps each j-edge of F to the corresponding j-cell chosen in $\mathcal{S}$. By Lemma 4.1(c), we have

$$\mathcal{G}(F) = G_\chi(F,\phi) = (1\pm\eta)\mathcal{D}_i^*(F,\phi) = (1\pm\eta)\mathcal{D}(F)\,,$$

and this verifies (BUL2). Finally, as in the proof of Theorem 4.6, (BUL3) holds trivially. It follows that we can apply Theorem 4.5 to obtain an embedding of H into $\mathcal{G}$. Taking the induced embedding of H' into $\mathcal{G}$, we have in particular an embedding of the k-edges of H' into G_χ, as desired. □

Again, note that this proof (as with the proof of Kohayakawa, Rödl, Schacht and Szemerédi [37]) actually gives a stronger conclusion: the graph Γ has a colour class which contains simultaneously all n-vertex k-uniform hypergraphs H with $\Delta(H^{(2)}) \leq \Delta$. This property is called *partition universality.*

4.1.5 Link Graphs and Typically Hereditary Counting

A major theme of the graph regularity method is working with sufficiently regular host graphs so that we can design randomised vertex-by-vertex embedding procedures that can ensure regularity inheritance throughout the embedding procedure, even when embedding large target graphs. The encapsulation of this idea for hypergraphs with the regularity condition of having sufficiently regular and accurate counts of small complexes underpins the definition of typically hereditary counting (THC), a pseudorandomness condition which plays a central role in our proof of Theorem 4.5. This pseudorandomness condition was introduced by Allen, Davies and Skokan [8] and they proved a theorem which shows that THC follows from certain counting conditions which resemble (BUL2) in Theorem 4.5. THC is *defined* to have a hereditary property and so we may take typical links in THC graphs a very large number of times and still preserve THC.

To prepare for the definition of THC and the related theorem, we shall motivate and provide several definitions. Our proof of Theorem 4.5 relies on a vertex-by-vertex embedding procedure and requires us to keep track of valid choices for embedding the yet unembedded vertices of H; given that we are working with hypergraphs, in general we need to keep track of highly intricate and complicated structures and subsets, which turns out to be notationally highly inconvenient and untidy.

To circumvent this notational nightmare, we define the standard construction, which enables a reduction from the general partite setting to a setting in which partite homomorphisms send exactly one vertex of H to each part of the host graph. The idea is to obtain an object which retains all the essential structure and properties of $\mathcal{G}$ *for the purpose of embedding H* vertex by vertex while providing a structure (through duplication) suited to a simple updating procedure.

Definition 4.8 (Standard construction)**.** Let J be an index set. Let H be a J-partite hypergraph with a partition $\mathcal{X} = \{X_j\}_{j\in J}$ of $V(H)$ and $\mathcal{G}$ be a J-partite weighted hypergraph with a partition $\mathcal{V} = \{V_j\}_{j\in J}$ of $V(\mathcal{G})$. The *standard construction* of $(\mathcal{G}, \mathcal{V})$ with respect to $(H, \mathcal{X})$ is a $V(H)$-partite weighted hypergraph $\mathcal{G}'$ with vertex sets $\{V'_x\}_{x\in V(H)}$ and weight function g', where for each $x \in V(H)$ the set V'_x is a copy of the set V_j such that $x \in X_j$, and where for each $f \subseteq V(H)$ and each $e' \in V'_f$ we define

$$g'(e') = \begin{cases} g(e) & \text{if } f \in E(H), \\ 1 & \text{if } f \notin E(H), \end{cases}$$

where e is the natural projection of e' to $V(\mathcal{G})$. We will sometimes omit mention of the hypergraph H and the partitions $\mathcal{X}$ and $\mathcal{V}$ when they are clear from context.

In other words, the standard construction of $(\mathcal{G}, \mathcal{V})$ with respect to $(H, \mathcal{X})$ is constructed by duplicating the clusters and edges of $\mathcal{G}$ so that each vertex of H is assigned its own cluster and the weights of the edges of $\mathcal{G}$ associated with each edge of H are suitably preserved to reflect how they restrict the valid choices for the embedding of vertices *specifically in the context of embedding H*.

We also require a definition of the link graph of a vertex v in a weighted hypergraph $\mathcal{G}$; this provides a straightforward way to update the structures we track in the standard construction. Let J be an index set and let $\mathcal{G}$ be a weighted hypergraph with its vertex set partitioned into $\{V_j\}_{j\in J}$. For a vertex $v \in V_i$ with $i \in J$ we let $\mathcal{G}_v$ be the weighted hypergraph on the vertex sets $\{V_j\}_{j\in J\setminus\{i\}}$ with weight function g_v defined as follows. For $f \subseteq J \setminus \{i\}$ and $e \in V_f$, we set

$$g_v(e) = g(e)g(e \cup \{v\}).$$

We call $\mathcal{G}_v$ the *link graph* of v in $\mathcal{G}$. In the context of unweighted hypergraphs, this means that the edges of $\mathcal{G}_v$ are the edges of $\mathcal{G}$ whose extension by v is also an edge of $\mathcal{G}$; this is a natural generalisation in the context of an embedding procedure because the second term in the definition of $g_v(e)$ generalises the usual notion of link graph and the first term represents the sensible requirement that the edge be already present in $\mathcal{G}$.

Since we are working in the weighted setting, we need to work with the sum of the weights of the vertices in a set instead of the size of that set. In particular, it will be convenient to work with a normalised version of this notion. Let $\mathcal{G}$ be a weighted hypergraph with vertex sets $\{V_j\}_{j\in J}$ and weight function g. For a subset $U \subseteq V_j$ we write

$$\|U\|_{\mathcal{G}} := |V_j|^{-1} \sum_{u\in U} g(u).$$

We also introduce the notion of the order function of a linear order. Given a linear order τ on a finite set J, the *order function* of τ is the bijection $\tau : J \to [|J|]$ such that for each $j \in J$ we have $\tau(j) = i$ if and only if j is the ith element of J in the order according to τ. In practice, we will not distinguish between a linear order and its order function as it will be clear from context. We say that $I \subseteq J$ is an *initial segment* of τ if $\tau(I) = [|I|]$.

Now we provide a definition of a pseudorandomness condition introduced by Allen, Davies and Skokan [8]. In fact, in their work they define two very closely related notions: typically hereditary counting and *local* typically hereditary counting. We will work with the latter of these two notions and simply call it typically hereditary counting (THC) for the sake of brevity. We emphasise that the definition we give is recursive in nature; to provide a base for this recursion we shall include all weighted hypergraphs on $\varnothing$ in the definition. For a weighted hypergraph $\mathcal{H}$, a subset S of the powerset of $V(\mathcal{H})$ and a non-negative real number x we say that $\mathcal{H}$ *is identically x on S* if the weight function h of $\mathcal{H}$ satisfies $h(e) = x$ for all $e \in S$. We say that $\mathcal{H}$ *is identically x outside S* if the weight function h of $\mathcal{H}$ satisfies $h(e) = x$ for all $e \notin S$. In the definition of THC below, $\mathcal{H}$ and $\mathcal{D}$ are best thought of as standard constructions of some pair of weighted hypergraphs with respect to the complex H; in this setting, it is entirely expected that $\mathcal{H}$ and $\mathcal{D}$ be identically 1 in the places specified.

Definition 4.9 (Typically hereditary counting (THC))**.** Given $k \in \mathbb{N}$ and a finite set J endowed with a linear order τ, let H be a k-complex on J and $\mathcal{D}$ be a weighted hypergraph on J which is identically 1 outside $E(H)$. Let $\mathcal{H}$ be a J-partite weighted hypergraph with a partition $\mathcal{V} = \{V_j\}_{j\in J}$ of $V(\mathcal{H})$ which is identically 1 on any V_e such that $e \notin E(H)$. We say that $\mathcal{H}$ is an (η, c)-THC *graph with the linear order* τ *and density weighted hypergraph* $\mathcal{D}$ if either $J = \varnothing$ and both $\mathcal{H}$ and $\mathcal{D}$ are weighted hypergraphs on $\varnothing$, or the following three properties hold.

(THC1) For each J-partite k-complex F on at most c vertices, we have

$$\mathcal{H}(F) = (1 \pm v(F)\eta)\tfrac{h(\varnothing)}{d(\varnothing)}\mathcal{D}(F).$$

(THC2) If $|J| \geq 2$ and x is the first vertex of J according to τ, there is a set $V'_x \subseteq V_x$ with $\|V'_x\|_{\mathcal{H}} \geq (1-\eta)\|V_x\|_{\mathcal{H}}$ such that for each $v \in V'_x$ the link graph $\mathcal{H}_v$ is an (η, c)-THC graph on $J \setminus \{x\}$ with the linear order on $J \setminus \{x\}$ induced by τ and density weighted hypergraph $\mathcal{D}_x$ on $J \setminus \{x\}$.

(THC3) The set V'_x of (THC2) is computed by an algorithm whose input is $\mathcal{H}[\bigcup_{z\in I} V_z]$, where I is the set of vertices in J at distance at most $c+2$ from x in $H^{(2)}$. Furthermore, the algorithm decides whether $v \in V'_x$ using only the input $\mathcal{H}[\{v\} \cup (\bigcup_{z\in I\setminus\{x\}} V_z)]$. This last part is monotone in the following sense: if a vertex y is deleted from J, with $y \neq x$, and the corresponding cluster V_y is deleted from $\mathcal{H}$, then any $v \in V_x$ which is included in V'_x for the original J and $\mathcal{H}$ by the algorithm is still in V'_x for $J \setminus y$ and $\mathcal{H}[\bigcup_{z\in J\setminus\{y\}} V_z]$.

We say that $\mathcal{D}$ is a *density weighted hypergraph* of $\mathcal{H}$. We often omit mention of τ and $\mathcal{D}$ when they are clear from context.

The formula in (THC1) says that the density of copies of small complexes F in $\mathcal{H}$ can be estimated within a small relative error by a density weighted hypergraph $\mathcal{D}$, while (THC2) asserts that the link graph obtained from embedding the first vertex 'inherits' THC for all but a weighted η-fraction of possible choices. (THC3) guarantees the existence of an algorithm which produces a set satisfying (THC2) that depends only on the 'local' structure.

The following result of Allen, Davies and Skokan [8] tells us that weighted hypergraphs satisfying certain counting conditions are THC graphs. Note that the conclusion is valid for any linear order on the indexing set; this is convenient as it allows us to select a favourable linear order in applications.

Theorem 4.10 ([8]). *For all $k, \Delta \geq 2, c \geq \Delta + 2$ and $0 < \eta < 1/2$, there exists $\eta_0 > 0$ such that whenever $0 < \eta' < \eta_0$ the following holds. Let H be a k-complex with $\Delta(H^{(2)}) \leq \Delta$ and a linear order τ on $V(H)$. Suppose that $\mathcal{H}$ is a $V(H)$-partite weighted-k-graph with vertex sets $\{V_x\}_{x\in V(H)}$ which is identically 1 on any V_f such that $f \notin E(H)$, and $\mathcal{D}$ is a weighted hypergraph on $V(H)$ such that for all $V(H)$-partite k-complexes F on at most $(\Delta+2)c$ vertices we have*

$$\mathcal{H}(F) = (1 \pm \eta')\frac{h(\varnothing)}{d(\varnothing)}\mathcal{D}(F).$$

*Then $\mathcal{H}$ is an (η, c)-*THC *graph with the linear order τ and density weighted hypergraph $\mathcal{D}$.*

As mentioned previously, our main goal is the embedding into a suitable k-complex $\mathcal{G}$ of a compatible k-complex H. Our approach involves a vertex-by-vertex embedding procedure and for technical reasons we embed into the standard construction of $\mathcal{G}$ with respect to H; as such, we naturally seek the THC property for the standard construction of $\mathcal{G}$ with respect to H rather than $\mathcal{G}$ itself. In the context of Theorem 4.10, $\mathcal{H}$ can be thought of as the standard construction of some weighted hypergraph $\mathcal{G}$ with respect to the complex H; the theorem has a more general form to handle derivatives of standard constructions. In practice, we work with a whole array of auxiliary complexes derived from H and it would be rather cumbersome to spell out on every occasion that we work with the relevant standard construction; in view of this and to highlight the relevance of $\mathcal{G}$ in such situations, we shall define what it means for $\mathcal{G}$ to be THC for H.

Given $k \in \mathbb{N}$ and a finite set J, let $\mathcal{D}$ be a weighted hypergraph on J and H be a J-partite hypergraph with its vertex set partitioned into $\mathcal{X}$. Let τ be a linear order on $V(H)$ and set $\mathcal{J} := \{\{j\}\}_{j\in J}$. We say that a J-partite weighted hypergraph $\mathcal{G}$ with its vertex set partitioned into $\mathcal{V}$ is an (η, c)-THC *graph for* (H, τ) with density weighted hypergraph $\mathcal{D}$ if the standard construction of $(\mathcal{G}, \mathcal{V})$ with respect to $(H, \mathcal{X})$ is an (η, c)-THC graph with the standard construction of $(\mathcal{D}, \mathcal{J})$ with respect to $(H, \mathcal{X})$ as its density weighted hypergraph and τ as the linear order on $V(H)$.

4.1.6 Overview

The following is a high-level overview of the proof of Theorem 4.5. We want to embed a bounded degree k-complex H, given with a partition $\mathcal{X}$ of $V(H)$, bounded degree reduced complexes $R' \subseteq R$ and a system $\widetilde{\mathcal{X}}$ of potential buffer sets, into a

k-complex $\mathcal{G}$ with a compatible partition $\mathcal{V}$ of $V(\mathcal{G})$ and highly precise typical counts of small complexes on R and super-typical counts of small complexes rooted at a vertex on R'. Our embedding strategy resembles the one(s) used by Allen, Böttcher, Hàn, Kohayakawa and Person to prove blow-up lemmas for sparse graphs [4] — which in turn is heavily inspired by the proof of the original blow-up lemma [38] — but the details are very different and many new ideas are required. Furthermore, our pseudorandomness condition of having small complex counts cannot be qualitatively weakened any further: having a counting lemma, which is an essential prerequisite in these settings, would already give us these counts.

There are three stages in our embedding strategy: the preprocessing stage, the random greedy embedding stage and the buffer matching embedding stage. In the preprocessing stage, we prepare the complexes H and $\mathcal{G}$ by suitably subpartitioning their partition classes to obtain some additional properties. We first subpartition the partition classes X_j of H so that any pair of vertices in the same part of the new partition is at distance at least seven from each other; this uses a trick first utilised by Alon and Füredi [9] which applies the Hajnal–Szemerédi Theorem. By having only distant vertices in each part, we ensure that these vertices are sufficiently independent for the random greedy embedding stage. We randomly subpartition the clusters V_j of $\mathcal{G}$ to obtain a compatible new partition of $\mathcal{G}$.

Next, we pick a small linearly-sized set X_j^{buf} of *buffer vertices* with various extra properties from the set of potential buffer vertices $\widetilde{X}_j$ in X_j and a linear order τ on $V(H)$ satisfying certain good properties. The remaining vertices of X_j are placed in a main part X_j^{main}. We also randomly subpartition each cluster V_j of $\mathcal{G}$ into three parts V_j^{main}, V_j^{q} and V_j^{buf}, where the first part is large and the other two are much smaller. The random subpartitioning facilitates the retention of super-regularity of graph degrees for the subparts on our super-typicality reduced complex R' and reserves dedicated subparts for the different stages and aspects of our embedding procedure.

Now that H and $\mathcal{G}$ have been preprocessed, we describe our approach for the random greedy embedding stage. We use a *random greedy algorithm* to embed X_j into V_j; it proceeds vertex by vertex, embedding each vertex $x \in X_j^{\mathrm{main}}$ into its *candidate set* and avoiding certain *bad vertices.* Writing ϕ for the partial embedding of H into $\mathcal{G}'$ constructed thus far, the candidate set $\mathcal{C}(x) \subseteq V_j$ is the set of vertices which extends $\phi(e)$ to an edge of $\mathcal{G}$ for all embedded $e \subseteq V(H)$ such that $e \cup \{x\} \in E(H)$. Of course, we cannot reuse vertices in $\mathcal{G}$ as we want an embedding; taking this into account, we embed uniformly at random into the set of *available candidates* which are not bad.

As mentioned previously, it is convenient to work in the standard construction $\mathcal{G}'$ of $\mathcal{G}$ with respect to H; in this setting, the candidate set of a vertex x turns out to be the set of vertices in the cluster corresponding to x with weight 1. The success of our random greedy algorithm relies on a hereditary pseudorandomness condition called typically hereditary counting (THC). In our case, the counting conditions on $\mathcal{G}$ imply that $\mathcal{G}$ is an (η, c)-THC graph for H. While this may be the obvious choice of THC as we want to embed H into $\mathcal{G}$ and it is the right property for counting homomorphisms from H to $\mathcal{G}$ and for having sufficiently many (not necessarily available) candidates, it turns out to be insufficient to give the embedding (i.e. *injective* homomorphism) of H into $\mathcal{G}$ we want.

To circumvent this initial difficulty, we will construct an *auxiliary complex* H_+ which 'extends' the target graph H and enables us to obtain somewhat stronger THC properties from the given counting conditions. We also have to maintain certain *good* properties which serve to ensure local goodness each time we extend our partial embedding. Then, the bad vertices are those which would cause the failure of our THC property or the loss of local goodness; we know these to be few by Lemma 4.41. That we embed uniformly at random into a reasonably large set helps ensure that a partial partite homomorphism is unlikely at any point to cover (asymptotically) all of the set of 'good' available candidates of an unembedded vertex, since we do not try to embed the 'last few' vertices of any vertex part in this stage.

Unfortunately, we cannot guarantee that no unembedded vertex will have its set of 'good' available candidates become small at some point, but our random embedding approach ensures that such vertices are relatively rare (see Lemma 4.42). We put these vertices into a so-called *queue* Q_t and exclusively use the *queue reservoir* $V_j^{\mathrm{q}} \subseteq V_j$ that we reserved in advance to embed the queue. Since the queue is tiny, our small but significantly larger queue reservoir provides plenty of 'room' to embed the queue vertices; such an analysis is not viable with the main portion of vertices as we cannot obtain a sufficiently large reservoir for them. A similar logic applies to the advance reservation of an analogous buffer reservoir $V_j^{\mathrm{buf}} \subseteq V_j$ for the buffer vertices in X_j^{buf} with various extra properties which enable us to embed them in a separate buffer matching embedding stage: when embedding the final fraction of vertices in each part, there are simply too few vertices remaining to smooth out any atypical behaviour.

Now we want to show that each queue vertex $x \in Q_t \cap X_j$ maintains a significant positive fraction of its 'good' candidates as available candidates and other vertices $x' \in Q_t \cap X_j$ do not occupy almost all of $\mathcal{C}^{\mathrm{q}}(x)$. By a stochastic process inequality, this reduces to showing that the sum of the probabilities of embedding the vertices $x' \in Q_t \cap X_j$ to $U := \mathcal{C}^{\mathrm{q}}(x)$ is suitably bounded from above. We will approximate the aforementioned sum with a martingale-like stochastic process $R = \{R_b\}_{0 \leq b \leq T}$ whose value at time T is the desired sum of probabilities, whose value at each time is a sum of probabilities of randomly picking at a certain time a copy of the unembedded neighbourhood complex F of x' with a vertex in U from all copies of F and whose value at time zero can be precisely bound by applying THC for an auxiliary complex; we will seek to track these probabilities while embedding the first two neighbourhoods of x'. The main issues are the unpredictability of counts involving U since it is a very small set (sublinear v.s. linear errors) and the potential for undesirable behaviours to positively correlate. We use regularity arguments to show that the worst misbehaviours cannot be too significant in aggregate and utilise Cauchy–Schwarz arguments to isolate the 'unpredictable' U-related counts, incorporate distance-2 neighbours, perform stepwise updating and keep track of the error.

Finally, it remains to embed the buffer vertices X_j^{buf}. Since buffer vertices are far apart, their candidate sets will no longer change and it remains to simply find a system of distinct representatives from the available candidate sets by verifying Hall's condition: for each $Y \subseteq X_j^{\mathrm{buf}}$ the union U of the available candidate sets of $y \in Y$ satisfies $|Y| \leq |U|$. There are three natural cases: Y is a small fraction of X_j^{buf}, Y has an intermediate size and Y contains all but a small fraction of X_j^{buf}. The argument for the first case turns out to resemble that for the queue vertices, while the argument for the second case is a straightforward consequence of the uniform distribution of candidate sets.

To deal with the remaining case we require that the additional property that for every vertex $v \in V_j$ there are many buffer vertices $x \in X_j^{\mathrm{buf}}$ for which v is a candidate; this is where we utilise the super-typical counts on R' to anticipate the (auxiliary) future embedding of x to v. With this, we complete the proof of our blow-up lemma for sparse hypergraphs.

Before closing this section, let us now discuss how our proof differs from previous blow-up lemmas. We work in a sparse setting, with the densities much smaller than the regularity error parameter(s); in contrast, the original blow-up lemma [38] and the hypergraph blow-up lemma of Keevash [33] work in the dense setting where the reverse is typically true. Furthermore, the result of Keevash relies on a regularity condition arising from the *regular approximation lemma* [50]; this is inspired by a dense setting and has a noticeably different flavour from the octahedral-minimality concept employed in [8], which our work is based on.

Working in a sparse setting means that interesting sets often have sublinear size, rendering them largely invisible to summary statistics. To overcome this obstacle, the work on blow-up lemmas for sparse graphs in [6] achieves the precise control required for very small sets through strong pseudorandomness conditions such as bijumbledness on the underlying graph. In our case, working with counting conditions means that in general we do not have meaningful direct control over quantities involving sublinear sets. Instead, we rely on auxiliary constructions and averaging arguments to gain precious control over these quantities under specific circumstances and the 'global' nature of the hereditary property in THC to prevent loss of regularity precision.

4.2 Preliminaries and Tools

In this section we introduce some notation and collect some useful tools.

4.2.1 Probability

Here we collect the probabilistic inequalities we need. The following is a version of a Chernoff bound for hypergeometrically distributed random variables.

Theorem 4.11 ([32, Theorem 2.10])**.** *Let X be a hypergeometrically distributed random variable. Then for $\varepsilon \in (0, 3/2)$ we have*

$$\mathbb{P}(X > (1+\varepsilon)\mathbb{E}[X]) \leq e^{-\varepsilon^2\mathbb{E}[X]/3} \text{ and } \mathbb{P}(X < (1-\varepsilon)\mathbb{E}[X]) \leq e^{-\varepsilon^2\mathbb{E}[X]/3}.$$

The following is a version of Freedman's martingale inequality, for which a proof is provided by Allen, Böttcher, Hladký and Piguet [6].

Lemma 4.12 ([6, Lemma 5])**.** *Let Ω be a finite probability space and $(\mathcal{F}_i)_{i\in[n]_0}$ be a filtration. Suppose that we have $R > 0$, and for each $i \in [n]$ we have an $\mathcal{F}_i$-measurable non-negative random variable Y_i, non-negative real numbers λ and σ, and an event $\mathcal{E}$.*

(i) *Suppose that almost surely, either $\mathcal{E}$ does not occur, or* $\sum_{i\in[n]}\mathbb{E}[Y_i|\mathcal{F}_{i-1}] \leq \lambda$, $\sum_{i\in[n]}\mathrm{var}(Y_i|\mathcal{F}_{i-1}) \leq \sigma^2$ *and* $0 \leq Y_i \leq R$ *for each* $i \in [n]$. *Then for each* $\nu > 0$ *we have*

$$\mathbb{P}\left(\mathcal{E} \text{ and } \sum_{i\in[n]} Y_i > \lambda + \nu\right) \leq \exp\left(-\frac{\nu^2}{2\sigma^2 + 2R\nu}\right).$$

(ii) *Suppose that almost surely, either $\mathcal{E}$ does not occur, or* $\sum_{i\in[n]}\mathbb{E}[Y_i|\mathcal{F}_{i-1}] \geq \lambda$, $\sum_{i\in[n]}\mathrm{var}(Y_i|\mathcal{F}_{i-1}) \leq \sigma^2$ *and* $0 \leq Y_i \leq R$ *for each* $i \in [n]$. *Then for each* $\nu > 0$ *we have*

$$\mathbb{P}\left(\mathcal{E} \text{ and } \sum_{i\in[n]} Y_i < \lambda - \nu\right) \leq \exp\left(-\frac{\nu^2}{2\sigma^2 + 2R\nu}\right).$$

4.2.2 Notation for Copying in Complexes

Let H be a complex. For $\ell \in \mathbb{N}$ let $\vec{A} = (A_1, \dots, A_\ell)$ be an ℓ-tuple of pairwise disjoint subsets of $V(H)$ and $\vec{p} = (p_1, \dots, p_\ell)$ be an ℓ-tuple of positive integers. For $i \in [\ell]$ and $j \in [p_i]$ let $A_i^{(j)}$ be a set of cardinality $|A_i|$. Let $H(\vec{A}, \vec{p})$ be the edge-minimal complex with vertex set $\bigsqcup_{i\in[\ell]}\left(\bigsqcup_{j\in[p_i]} A_i^{(j)}\right)$ such that $H(\vec{A}, \vec{p})\left[\bigcup_{i\in[\ell]} A_i^{(j_i)}\right]$ is isomorphic to $H[\bigcup_{i\in[\ell]} A_i]$ under the natural isomorphism for all $(j_1, \dots . j_\ell) \in \prod_{i\in[\ell]}[p_i]$. We will drop the parentheses when $\ell = 1$ and simply write $H(A_1, p_1)$.

We think of $H(\vec{A}, \vec{p})$ as the complex obtained from H by deleting any vertex not in any A_i, creating p_i copies of each subset A_i and duplicating the edges of H in a 'partite' manner. The idea is to generalise the following example. Let H be the down-closure complex of a single edge xy and $\vec{A} := (\{x\}, \{y\})$. Then $H(\vec{A}, \vec{p}) = K_{p_1,p_2}$.

4.2.3 Notation for Ordered Complexes

Let H be a complex with a linear order τ on $V(H)$. We recursively define the sets $N^{-p}(x)$ and $N^{<p}(x)$ for $p \in \mathbb{N}$ and $x \in V(H)$ as follows. Set $N^{-1}(x), N^{<1}(x) := \{y \in N_{H^{(2)}}(x) : \tau(y) < \tau(x)\}$. For $p \geq 2$ set

$$N^{<p}(x) := N^{<(p-1)}(x) \cup \left(\bigcup_{y\in N^{-(p-1)}(x)} N^{-1}(y)\right)$$

and $N^{-p}(x) := N^{<p}(x) \setminus N^{<(p-1)}(x)$. Set $N^{>}(x) := \{y \in N_{H^{(2)}}(x) : \tau(y) > \tau(x)\}$. For $p \in \mathbb{N}$ set $N^{\leq p}(x) := N^{<p}(x) \cup \{x\}$, $H^{<p}(x) := H[N^{<p}(x)]$ and $H^{\leq p}(x) := H[N^{\leq p}(x)]$.

For the pth neighbourhood complex $H^{\leq p}(x)$ we will write $H^{\leq p}_{\times q}(x)$ to mean the complex $H(\vec{A},\vec{q})$ where $\vec{A}=(N^{<p}(x),\{x\})$ and $\vec{q}=(1,q)$. This is the complex obtained by making q copies of x and giving each of those copies the same adjacencies as the original x. We denote the q copies of x by $x^{(1)},\dots,x^{(q)}$.

4.2.4 Cauchy–Schwarz and Counts of Complexes

Here we state the Cauchy–Schwarz inequality and prove a useful lemma about counts of complexes.

Lemma 4.13 (Cauchy–Schwarz inequality)**.** *Let $\alpha_1,\dots,\alpha_t,\beta_1,\dots,\beta_t$ be real numbers. Then*

$$\left|\sum_{i\in[t]}\alpha_i\beta_i\right| \leq \sqrt{\sum_{i\in[t]}\alpha_i^2\sum_{i\in[t]}\beta_i^2}.$$

The following lemma applies the Cauchy–Schwarz inequality to establish relationships between counts of complexes.

Lemma 4.14. *Let H be a complex with a partition $\{A_1,B,A_2\}$ of $V(H)$ such that $e\cap A_1\neq\varnothing \implies e\cap A_2=\varnothing$ for all $e\in E(H)$. For $i\in[2]$ let $\vec{A_i}:=(A_i,B)$. Let $\vec{a}:=(2,1)$. Let $\mathcal{G}$ be a $V(H)$-partite weighted hypergraph with vertex sets $\{V_x\}_{x\in V(H)}$. Then*

$$\mathcal{G}(H)\leq\sqrt{\mathcal{G}(H(\vec{A_1},\vec{a}))\mathcal{G}(H(\vec{A_2},\vec{a}))}.$$

Proof. Set $\vec{b}:=(1,1)$. For $i\in[2]$ set $H_i:=H(\vec{A_i},\vec{b})$ and $H_{i,2}:=H(\vec{A_i},\vec{a})$. Set $H_3:=H[B]$. Let ψ be a partite homomorphism from H_3 to $\mathcal{G}$. Define

$$a_\psi:=\sum_{\phi:\phi|_{V(H_3)}=\psi}\ \prod_{e\in E(H)}g(\phi(e)),$$

where the sum is over all partite homomorphisms ϕ from H to $\mathcal{G}$ which are identical to ψ when restricted to H_3. For $i\in[2]$ define

$$a_{i,\psi}:=\sum_{\phi:\phi|_{V(H_3)}=\psi}\ \prod_{e\in E(H_i)}g(\phi(e)),$$

where the sum is over all partite homomorphisms ϕ from H_i to $\mathcal{G}$ which are identical to ψ when restricted to H_3. Since $E(H)=E(H_1)\cup E(H_2)$ and there is a natural correspondence between partite homomorphisms from H to $\mathcal{G}$ which are identical to ψ when restricted to H_3 and pairs (ϕ_1,ϕ_2) of partite homomorphisms where ϕ_1 and ϕ_2

are a partite homomorphism from H_1 to $\mathcal{G}$ and a partite homomorphism from H_2 to $\mathcal{G}$ respectively which are both identical to ψ when restricted to H_3, it follows that we have

$$a_\psi = a_{1,\psi} a_{2,\psi}. \tag{4.2}$$

Define

$$b_{i,\psi} := \sum_{\phi : \phi|_{V(H_3)} = \psi} \prod_{e \in E(H_{i,2})} g(\phi(e)),$$

where the sum is over all partite homomorphisms ϕ from $H_{i,2}$ to $\mathcal{G}$ which are identical to ψ when restricted to H_3. For each $i \in [2]$ we have

$$b_{i,\psi} = a_{i,\psi}^2 \tag{4.3}$$

by an argument analogous to that for (4.2).

Define $a := \sum_\phi \prod_{e \in E(H)} g(\phi(e))$, where the sum is over all partite homomorphisms ϕ from H to $\mathcal{G}$. For $i \in [2]$ define $b_i := \sum_\phi \prod_{e \in E(H_{i,2})} g(\phi(e))$, where the sum is over all partite homomorphisms ϕ from $H_{i,2}$ to $\mathcal{G}$. Since every partite homomorphism ϕ from H to $\mathcal{G}$ restricted to $V(H_3)$ is a partite homomorphism ψ from H_3 to $\mathcal{G}$, we find that $a = \sum_\psi a_\psi$, where the sum is over all partite homomorphisms ψ from H_3 to $\mathcal{G}$. By analogous arguments we also have $b_i = \sum_\psi b_{i,\psi}$, where the sums are over all partite homomorphisms ψ from H_3 to $\mathcal{G}$, for $i \in [2]$. Then applying (4.2), Lemma 4.13 and (4.3), we obtain

$$a = \sum_\psi a_{1,\psi} a_{2,\psi} \le \sqrt{\sum_\psi a_{1,\psi}^2 \sum_\psi a_{2,\psi}^2} = \sqrt{b_1 b_2},$$

where the sums are over all partite homomorphisms ψ from H_3 to $\mathcal{G}$. Hence, we have

$$\begin{aligned} \mathcal{G}(H) = a \prod_{x \in V(H)} |V_x|^{-1} &\le \left(b_1 \prod_{x \in V(H_{1,2})} |V_x|^{-1} \right)^{1/2} \left(b_2 \prod_{x \in V(H_{2,2})} |V_x|^{-1} \right)^{1/2} \\ &= \sqrt{\mathcal{G}(H_{1,2}) \mathcal{G}(H_{2,2})} \end{aligned}$$

as desired. □

4.2.5 Sparse Regularity

Sparse regularity turns out to be the right concept for analysing the behaviour of small subgraph counts and the local behaviour of our random greedy algorithm. Here we provide a definition and some useful lemmas.

Definition 4.15 (Sparse regularity)**.** Let $\ell \in \mathbb{N}$, $p > 0$ and G be an ℓ-uniform hypergraph. Let $U_1, \dots, U_\ell$ be pairwise disjoint nonempty subsets of $V(G)$. Write $e(U_1, \dots, U_\ell)$ for the number of edges in G with exactly one vertex in each of $U_1, \dots, U_\ell$. The *p-density* of $(U_1, \dots, U_\ell)$ is $d_p(U_1, \dots, U_\ell) := \frac{e(U_1,\dots,U_\ell)}{p|U_1|\dots|U_\ell|}$. We say that an ℓ-tuple $(U_1, \dots, U_\ell)$ is *(ε, p)-regular* if for all ℓ-tuples $(U_1', \dots, U_\ell')$ such that $U_i' \subseteq U_i$ and $|U_i'| \geq \varepsilon |U_i|$ for all $i \in [\ell]$, we have

$$|d_p(U_1', \dots, U_\ell') - d_p(U_1, \dots, U_\ell)| \leq \varepsilon.$$

We say that an (ε, p)-regular ℓ-tuple $(U_1, \dots, U_\ell)$ is *(ε)-regular* if it has p-density $d_p(U_1, \dots, U_\ell) = 1$. For an ℓ-partite ℓ-uniform hypergraph F with vertex sets $\{U_i\}_{i \in [\ell]}$, we say that F is (ε, p)-regular if the ℓ-tuple $(U_1, \dots, U_\ell)$ is (ε, p)-regular and that F is (ε)-regular if $(U_1, \dots, U_\ell)$ is (ε)-regular.

Lemma 4.16. *Let $\ell \geq 2$, $\varepsilon, p > 0$ and G be an ℓ-uniform hypergraph. Given an (ε, p) regular ℓ-tuple $(U_1, \dots, U_\ell)$ of pairwise disjoint nonempty subsets of $V(G)$ and an $(\ell - 1)$-tuple $(U_1', \dots, U_{\ell-1}')$ of subsets such that for each $i \in [\ell - 1]$ we have $U_i' \subseteq U_i$ and $|U_i'| \geq \varepsilon |U_i|$, the set*

$$U_\ell' := \{u \in U_\ell : d_p(U_1', \dots, U_{\ell-1}', \{u\}) < d_p(U_1, \dots, U_\ell) - \varepsilon\}$$

satisfies $|U_\ell'| < \varepsilon |U_\ell|$.

Proof. By the definitions of U_ℓ' and d_p we have

$$d_p(U_1', \dots, U_\ell') = \sum_{u \in U_\ell'} \frac{d_p(U_1', \dots, U_{\ell-1}', \{u\})}{|U_\ell'|} < d_p(U_1, \dots, U_\ell) - \varepsilon.$$

Since $(U_1, \dots, U_\ell)$ is (ε, p)-regular and $|U_i'| \geq \varepsilon |U_i|$ for all $i \in [\ell - 1]$, we conclude that $|U_\ell'| < \varepsilon |U_\ell|$. $\square$

The following is a defect version of the Cauchy–Schwarz inequality. This inequality and a proof can be found in [27, Fact B].

Lemma 4.17 ([27, Fact B])**.** *Let $m \in \mathbb{N}$ and $\delta, \alpha \geq 0$. Let $a_1, \dots, a_m, a$ be real numbers such that $\sum_{i \in [m]} a_i \geq am$. If we have*

$$\sum_{i \in [m]} a_i^2 < ma^2 \left(1 + \frac{\alpha \delta^2}{1 - \alpha}\right),$$

then for all $S \subseteq [m]$ such that $|S| \geq \alpha m$ we have $\sum_{i \in S} a_i = (1 \pm \delta) a |S|$.

The following lemma relates sparse regularity with small subgraph counts. We shall apply Lemma 4.17 to prove this.

Lemma 4.18. *Let $\ell \geq 2$, $q \in \mathbb{N}$ and $p > 0$. Let $\eta, \varepsilon \in (0,1]$ satisfy $2^{12}q\eta \leq \varepsilon^{2\ell+3}$. Let G be an ℓ-partite ℓ-uniform hypergraph with nonempty vertex sets $\{U_i\}_{i\in[\ell]}$. Let L be the unique $[\ell]$-partite ℓ-uniform hypergraph on $[\ell]$. Let $\vec{A} = (\{i\})_{i\in[\ell]}$. For $S \subseteq [2]$, let $\vec{p}_S$ be the ℓ-tuple where the ith entry is 2 if $i \in S$ and 1 otherwise, and set $L_S := L(\vec{A}, \vec{p}_S)$. Suppose that we have $G(L_S) = (1 \pm 4q\eta)p^{2^{|S|}}$ for all $S \subseteq [2]$. Then G is (ε)-regular.*

Proof. Let $(U'_1, \dots, U'_\ell)$ be any ℓ-tuple such that for all $i \in [\ell]$ we have $U'_i \subseteq U_i$ and $|U'_i| \geq \varepsilon|U_i|$. Let $\gamma_0 := 4q\eta$, $\gamma_1 := 4\left(\frac{q\eta}{\varepsilon}\right)^{1/2}$ and $\gamma_2 := 4\left(\frac{q\eta}{\varepsilon^{2\ell-1}}\right)^{1/4}$. For $h \in [2]$ set $\vec{W}_h := \prod_{i\in[\ell]\setminus\{h\}} U_i$. Set $b := \prod_{i\in[\ell]} |U_i|$. For each $\vec{u} = (u_i)_{i\in[\ell]\setminus\{2\}} \in \vec{W}_2$ set $a_{\vec{u}} := |U_2| G(L; \vec{u}, (i)_{i\in[\ell]\setminus\{2\}})$. We have

$$\sum_{\vec{u}\in\vec{W}_2} a_{\vec{u}} = G(L)b = (1 \pm \gamma_0)pb \text{ and } \sum_{\vec{u}\in\vec{W}_2} a_{\vec{u}}^2 = G(L_{\{2\}})b|U_2| = (1 \pm \gamma_0)p^2 b|U_2|.$$

Let $W' := U'_1 \times \left(\prod_{i=3}^{\ell} U_i\right)$. Applying Lemma 4.17 with $a = (1-\gamma_0)p|U_2|$, $\alpha = \varepsilon$ and $\delta = \gamma_1 - \gamma_0$, we obtain

$$G(L; U'_1, 1) = \frac{|U_1|}{b|U'_1|} \sum_{\vec{u}\in W'} a_{\vec{u}} = (1 \pm \gamma_1)p.$$

We also obtain $G(L_{\{1\}}; (U'_1, U'_1), (1, 1')) = (1 \pm \gamma_1)p^2$, where $1'$ represent the duplicate of 1 in $L_{\{1\}}$, by a similar argument with $L_{\{1\}}$ instead of L. Now for each $\vec{u} = (u_i)_{i\in[\ell]\setminus\{1\}} \in \vec{W}_1$ set

$$b_{\vec{u}} := |U'_1| G(L; (U'_1, \{u_2\}, \dots, \{u_\ell\}), (h)_{h\in[\ell]}).$$

We have

$$\sum_{\vec{u}\in\vec{W}_1} b_{\vec{u}} = G(L; U'_1, 1)b|U'_1|/|U_1| = (1 \pm \gamma_1)pb|U'_1|/|U_1|$$

and

$$\sum_{\vec{u}\in\vec{W}_1} b_{\vec{u}}^2 = G(L_{\{1\}}; (U'_1, U'_1), (1, 1'))b|U'_1|^2/|U_1| = (1 \pm \gamma_1)p^2 b|U'_1|^2/|U_1|.$$

Let $W'' := \prod_{i\in[\ell]\setminus\{1\}} U'_i$. Applying Lemma 4.17 with $a = (1-\gamma_1)p|U'_1|$, $\alpha = \varepsilon^{\ell-1}$ and $\delta = \gamma_2 - \gamma_1$, we obtain

$$G(L; (U'_i)_{i\in[\ell]}, (i)_{i\in[\ell]}) = \prod_{i\in[\ell]} |U'_i|^{-1} \sum_{\vec{u}\in W''} a_{\vec{u}} = (1 \pm \gamma_2)p.$$

Then, we obtain

$$d_{G(L)}(U'_1, \ldots, U'_\ell) = \frac{G(L; (U'_i)_{i\in[\ell]}, (i)_{i\in[\ell]})}{G(L)} = 1 \pm \varepsilon.$$

Hence, G is (ε)-regular. □

4.2.6 Equitable Partitions

To prepare the complex H for our embedding procedure, we require the following result on equitable partitioning related to the Hajnal–Szemerédi theorem (Theorem 2.4). The main difference is that the equitable partition we obtain needs to induce an equitable partition on a specified small subset of the vertices; in our applications this will be the set of buffer vertices.

Lemma 4.19. *Given a graph G and a subset U of $V(G)$ with $|U| \le \frac{|V(G)|}{2}$, setting $t = 6\Delta(G) + 1$, there is an equitable partition $\{V_i\}_{i\in[t]}$ of $V(G)$ such that each part is independent and $\{V_i \cap U\}_{i\in[t]}$ is an equitable partition of U.*

Proof. We first apply Theorem 2.4 to obtain an equitable partition $\{U_i\}_{i\in[t]}$ of U into independent sets in G. Then, fix a partition $\{V_i\}_{i\in[t]}$ of $V(G)$ into independent sets in G with $U_i \subseteq V_i$ for all $i \in [t]$ such that $\sum_{i,j\in[t]} ||V_i| - |V_j||$ is minimal. Observe that such a partition exists: since there are t parts and each vertex has at most $\Delta(G)$ neighbours, we can add the vertices of $V(G) \setminus U$ to the parts of $\{U_i\}_{i\in[t]}$ vertex-by-vertex while maintaining their independence. Without loss of generality, we may assume that V_1 is a smallest part. It follows that $t|V_1| \le |V(G)|$.

Suppose for a contradiction that $\{V_i\}_{i\in[t]}$ is not equitable, that is, there are distinct parts whose size differ by at least 2. In particular, there are parts V_i with $|V_i| \ge |V_1| + 2$. Note that for each such i every vertex in $V_i \setminus U$ has a neighbour in V_1, for otherwise we would be able to move into V_1 a vertex in some $V_i \setminus U$ with no neighbours in V_1 and obtain a new partition $\{W_i\}_{i\in[t]}$ of $V(G)$ into independent sets in G with $U_i \subseteq V_i$ for all $i \in [t]$ such that $\sum_{i,j\in[t]} ||W_i| - |W_j|| < \sum_{i,j\in[t]} ||V_i| - |V_j||$, thereby contradicting the minimality of $\{V_i\}_{i\in[t]}$. In particular, we must have $|V_1|, \Delta(G) \ge 1$.

Fix a set V_i with $|V_i| \ge |V_1| + 2$. Since $\{U_i\}_{i\in[t]}$ is equitable, we have $|U_i| \le |U_1| + 1 \le |V_1| + 1$ and therefore $|V_i \setminus U| \ge 1$. Fix a vertex $v \in V_i \setminus U$. By a previous argument, the vertices in $V(G) \setminus U$ which have no neighbour in V_1, of which there are at least $|V(G)| - |U| - \Delta(G)|V_1|$, are all contained in parts of size at most $|V_1| + 1$. Since

$\frac{|V(G)|-|U|-\Delta(G)|V_1|}{|V_1|+1} > \Delta(G)$, these vertices lie in at least $\Delta(G)+1$ distinct parts; hence, there is a part V_j which contains no neighbour of v and contains a vertex w, not in U, with no neighbour in V_1. Now by replacing V_1, V_i and V_j with $V_1 \cup \{w\}$, $V_i \setminus \{v\}$ and $V_j \cup \{v\} \setminus \{w\}$ respectively, we obtain a new partition $\{Z_i\}_{i\in[t]}$ of $V(G)$ into independent sets in G with $U_i \subseteq Z_i$ for all $i \in [t]$ such that $\sum_{i,j\in[t]} ||Z_i| - |Z_j|| < \sum_{i,j\in[t]} ||V_i| - |V_j||$, thereby contradicting the minimality of $\{V_i\}_{i\in[t]}$. $\square$

4.3 Standard Constructions and THC

In this section we introduce notation, terminology and useful technical results about standard constructions and THC.

4.3.1 Notation for Standard Constructions

It is useful to have the following notation for when we work with standard constructions. Let J be a finite set. Let H be a J-partite complex with its vertex set partitioned into $\mathcal{X} = \{X_j\}_{j\in J}$. Let $\mathcal{G}$ be a J-partite weighted hypergraph with its vertex set partitioned into $\mathcal{V} = \{V_j\}_{j\in J}$. Let $\mathcal{G}'$ with vertex sets $\{V'_x\}_{x\in V(H)}$ be the standard construction of $(\mathcal{G}, \mathcal{V})$ with respect to $(H, \mathcal{X})$. For $S \subseteq V(\mathcal{G}')$ write $S_{\to\mathcal{G}}$ to mean the natural projection of S into $V(\mathcal{G})$. For $v \in V(\mathcal{G}')$ write $v_{\to\mathcal{G}}$ to mean the natural projection of v into $V(\mathcal{G})$. Let $j \in J$. Let $x, y \in X_j$ be vertices. For $S \subseteq V'_x$ we write $S_{x\to y}$ to mean the natural projection of S into V'_y. For $v \in V'_x$ we write $v_{x\to y}$ to mean the natural projection of v into V'_y. For $S \subseteq V_j$ we write $S_{j\to x}$ to mean the natural projection of S into V'_x. For $v \in V_j$ we write $v_{j\to x}$ to mean the natural projection of v into V'_x. Let $\overline{H}$ be a J-partite complex with its vertex set partitioned into $\overline{\mathcal{X}} = \{\overline{X}_j\}_{j\in J}$ such that $X_j \subseteq \overline{X}_j$ for all $j \in J$. Let $\overline{\mathcal{G}}$ with vertex sets $\{\overline{V}_x\}_{x\in V(\overline{H})}$ be the standard construction of $(\mathcal{G}, \mathcal{V})$ with respect to $(\overline{H}, \overline{\mathcal{X}})$. For $S \subseteq V(\mathcal{G}')$ write $S_{\to\overline{\mathcal{G}}}$ to mean the natural projection of S into $V(\overline{\mathcal{G}})$. For $v \in V(\mathcal{G}')$ write $v_{\to\overline{\mathcal{G}}}$ to mean the natural projection of v into $V(\overline{\mathcal{G}})$. For $S \subseteq \bigcup_{x\in V(H)} \overline{V}_x$ write $S_{\to\mathcal{G}'}$ to mean the natural projection of S into $V(\mathcal{G}')$. For $v \in \bigcup_{x\in V(H)} \overline{V}_x$ write $v_{\to\mathcal{G}'}$ to mean the natural projection of v into $V(\mathcal{G}')$.

4.3.2 Technical Lemmas for Counts in Standard Constructions

Here we provide some technical results regarding counts in standard constructions. The following lemma formalises the correspondence between counts in a weighted hypergraph and counts in a standard construction of it.

Lemma 4.20. *Let J be a finite set. Let H be a J-partite complex with vertex partition $\mathcal{X} = \{X_j\}_{j\in J}$. Let $\mathcal{G}$ be a J-partite weighted hypergraph with vertex partition $\mathcal{V} = \{V_j\}_{j\in J}$. Let $\mathcal{H}$ be the standard construction of $(\mathcal{G}, \mathcal{V})$ with respect to $(H, \mathcal{X})$. Let F be a $V(H)$-partite complex with its vertex set partitioned into $\mathcal{F} = \{F_x\}_{x\in V(H)}$. For each $j \in J$ set $\overline{F}_j = \bigcup_{x\in X_j} F_x$. Let $\overline{F}$ be a J-partite spanning subcomplex of F with vertex partition $\overline{\mathcal{F}} = \{\overline{F}_j\}_{j\in J}$ for which the edge-maximal H-partite spanning subcomplex F' of F is a subcomplex. Then*

(i) $\mathcal{G}(\overline{F}) = \mathcal{H}(\overline{F}) = \mathcal{H}(F)$.

(ii) for each $x \in V(F)$, with $j \in J$ and $y \in X_j$ such that $x \in F_y$, and each $v \in V_j$, we have $\mathcal{G}(\overline{F}; v, x) = \mathcal{H}(\overline{F}; v_{j\to y}, x) = \mathcal{H}(F; v_{j\to y}, x)$.

Proof. We first consider (i). There is a correspondence between the $V(H)$-partite homomorphisms from $\overline{F}$ to $\mathcal{H}$ and the J-partite homomorphisms from $\overline{F}$ to $\mathcal{G}$ by the natural projection from $V(\mathcal{H})$ into $V(\mathcal{G})$, so we have $\mathcal{G}(\overline{F}) = \mathcal{H}(\overline{F})$. Since an edge of F which is not an edge of $\overline{F}$ is not H-partite, any $V(H)$-partite homomorphism from F to $\mathcal{H}$ must send any such edge into V'_f for some $f \notin E(H)$. The elements of V'_f for any $f \notin E(H)$ all have weight 1, so we have $\mathcal{H}(\overline{F}) = \mathcal{H}(F)$, completing the proof.

Now we consider (ii). The argument is analogous to that for (i), except that we need to account for the 'rooting' of x at v. Let $x \in V(F)$. Let $j \in J$ and $y \in X_j$ be such that $x \in F_y$. Let $v \in V_j$. There is a correspondence between the $V(H)$-partite homomorphisms from $\overline{F}$ to $\mathcal{H}$ which map x to $v_{j\to y}$ and the J-partite homomorphisms from $\overline{F}$ to $\mathcal{G}$ which map x to v by the natural projection from $V(\mathcal{H})$ into $V(\mathcal{G})$, so we have $\mathcal{G}(\overline{F}; v, x) = \mathcal{H}(\overline{F}; v_{j\to y}, x)$. Since an edge of F which is not an edge of $\overline{F}$ is not H-partite, any $V(H)$-partite homomorphism from F to $\mathcal{H}$ which maps x to $v_{j\to y}$ must send any such edge into V'_f for some $f \notin E(H)$. The elements of V'_f for any $f \notin E(H)$ all have weight 1, so we have $\mathcal{H}(\overline{F}; v_{j\to y}, x) = \mathcal{H}(F; v_{j\to y}, x)$, completing the proof. □

The following lemma formalises the correspondence between systems of counts in a weighted hypergraph and systems of counts in a standard construction of it.

Lemma 4.21. *Let $k, c \in \mathbb{N}$ and $\eta > 0$. Let R be a k-complex on a finite set J. Let H be a J-partite k-complex with a partition $\mathcal{X} = \{X_j\}_{j\in J}$ of $V(H)$ such that $(H, \mathcal{X})$ is an R-partition. Let $\mathcal{G}$ be a J-partite weighted hypergraph with vertex partition $\mathcal{V} = \{V_j\}_{j\in J}$,*

$\mathcal{D}$ be a weighted hypergraph on J and $\mathcal{J} = \{\{j\}\}_{j\in J}$. Let the standard constructions of $(\mathcal{G}, \mathcal{V})$ and $(\mathcal{D}, \mathcal{J})$ with respect to $(H, \mathcal{X})$ be $\mathcal{H}$ and $\mathcal{B}$ respectively. Suppose that $(\mathcal{G}, \mathcal{V})$ is an $(\eta, c, \mathcal{D})$-typcount R-partition. Then for each $V(H)$-partite k-complex F on at most c vertices we have $\mathcal{H}(F) = (1 \pm \eta)\frac{g(\varnothing)}{d(\varnothing)}\mathcal{B}(F)$.

Proof. Let F be a $V(H)$-partite k-complex on at most c vertices with vertex partition $\mathcal{F} = \{F_x\}_{x\in V(H)}$. Let $\overline{F}_j = \bigcup_{x\in X_j} F_x$ for each $j \in J$. Let $\overline{F}$ be the edge-maximal H-partite subcomplex of F with vertex partition $\overline{\mathcal{F}} = \{\overline{F}_j\}_{j\in J}$. By Lemma 4.20(i), we have $\mathcal{G}(\overline{F}) = \mathcal{H}(F)$ and $\mathcal{D}(\overline{F}) = \mathcal{B}(F)$. By definition we have $g(\varnothing) = h(\varnothing)$ and $d(\varnothing) = b(\varnothing)$. Hence, we deduce that

$$\mathcal{H}(F) = \mathcal{G}(\overline{F}) = (1 \pm \eta)\frac{g(\varnothing)}{d(\varnothing)}\mathcal{D}(F) = (1 \pm \eta)\frac{h(\varnothing)}{b(\varnothing)}\mathcal{B}(F),$$

completing the proof. □

4.3.3 Terminology for THC

Our proof of Theorem 4.5 involves the embedding of a k-complex H into another k-complex $\mathcal{G}$ by way of a vertex-by-vertex embedding procedure, with a central role played by the pseudorandomness condition THC and its hereditary property (THC2). Here we shall introduce terminology to enable description of the 'respecting' of this hereditary property in the context of embedding procedures.

In a typical setting, we have the standard construction $\mathcal{H}$ of a weighted hypergraph with respect to a complex H, where $\mathcal{H}$ is a THC graph with density weighted hypergraph $\mathcal{D}$, and we want to describe what it means to 'respect' THC in $\mathcal{H}$ as we embed H vertex-by-vertex. Let us now define this formally and more generally. Let H be a complex with a linear order τ on $V(H)$ and $\mathcal{D}$ be a weighted hypergraph on $V(H)$. Let $\mathcal{H}$ be a $V(H)$-partite weighted hypergraph with vertex sets $\{V_x\}_{x\in V(H)}$ which is an (η, c)-THC graph with the linear order τ and density weighted hypergraph $\mathcal{D}$ such that $g(\varnothing) = 1$. Let ϕ be a partial partite homomorphism from H to $\mathcal{H}$ such that $\mathrm{Dom}(\phi)$ is an initial segment of τ. Enumerate $\mathrm{Dom}(\phi)$ as $x_1, \dots, x_{|\mathrm{Dom}(\phi)|}$ according to the linear order τ. Set $\mathcal{H}_0 := \mathcal{H}$ and $\mathcal{D}_0 := \mathcal{D}$. For each $i \in [|\mathrm{Dom}(\phi)|]$, given $\mathcal{H}_{i-1}$ and $\mathcal{D}_{i-1}$, set $\mathcal{H}_i := (\mathcal{H}_{i-1})_{\phi(x_i)}$ and $\mathcal{D}_i := (\mathcal{D}_{i-1})_{x_i}$. We say that ϕ is (η, c)-THC-*respecting for* $(\mathcal{H}, H, \tau)$ if for all $i \in [|\mathrm{Dom}(\phi)|]$ the vertex $\phi(x_i)$ has weight 1 in $\mathcal{H}_{i-1}$ and belongs to the set V'_{x_i} of (THC2) returned by an algorithm whose existence is guaranteed by (THC3) for $\mathcal{H}_{i-1}$; in particular, for all $i \in [|\mathrm{Dom}(\phi)|]$ we have $g_i(\varnothing) = 1$ and $\mathcal{H}_i$ is an (η, c)-THC graph with the linear order on $V(H) \setminus \{x_h : h \in [i]\}$ induced by τ and density weighted hypergraph $\mathcal{D}_i$. Occasionally it may be more convenient to describe the partial partite homomorphism as a tuple of vertices. Let $\vec{v} = (v_i)_{i \in [\ell]}$ be a tuple of vertices of $\mathcal{H}$ with $v_i \in V_{x_i}$ for all $i \in [\ell]$. We say that $\vec{v}$ is (η, c)-THC-*respecting for* $(\mathcal{H}, H, \tau)$ if the function $\psi : \{x_i : i \in [\ell]\} \to V(\mathcal{H})$ given by $\phi(x_i) = v_i$ for $i \in [\ell]$ is a partial partite homomorphism from H to $\mathcal{H}$ which is (η, c)-THC-respecting for $(\mathcal{H}, H, \tau)$. The terminology introduced in this paragraph and the rest of this subsection is parameterised by η, c, $\mathcal{H}$, H and τ; for the sake of brevity, we will often omit some or all of the parameters which are clear from context. For example, we will say that ϕ is (η, c)-THC-respecting when $\mathcal{H}$, H and τ are clear from context and ϕ is THC-respecting for $(\mathcal{H}, H, \tau)$ when η and c are clear from context; if η, c, $\mathcal{H}$, H and τ are all clear from context we will simply say that ϕ is THC-respecting.

Often we wish to describe the embedding of a small interesting part of H within a partial partite homomorphism of H which 'respects' THC, but the vertices of interest are not embedded consecutively; examples include neighbourhoods of vertices. A partial partite homomorphism ψ from H to $\mathcal{H}$ is (η, c)-THC-*extendable for* $(\mathcal{H}, H, \tau)$ if there is a partial partite homomorphism ϕ from H to $\mathcal{H}$ which is (η, c)-THC-respecting for $(\mathcal{H}, H, \tau)$ such that $\mathrm{Dom}(\phi)$ is an initial segment of τ containing $\mathrm{Dom}(\psi)$ and $\phi|_{\mathrm{Dom}(\psi)} = \psi$. Let $\vec{v} = (v_x)_{x \in S}$ be a tuple of vertices of $\mathcal{H}$ with $S \subseteq V(H)$ and $v_x \in V_x$ for all $x \in S$. We say that $\vec{v}$ is (η, c)-THC-*extendable for* $(\mathcal{H}, H, \tau)$ if the function $\psi : S \to V(\mathcal{H})$ given by $\phi(x) = v_x$ for $x \in S$ is a partial partite homomorphism from H to $\mathcal{H}$ which is (η, c)-THC-extendable for $(\mathcal{H}, H, \tau)$.

While the main goal in our proof is to embed a complex H into the standard construction $\mathcal{H}$ of our host graph $\mathcal{G}$ with respect to H, our proof requires us to consider the embedding of H as a partial embedding of an auxiliary complex H_+ 'extending' H into the standard construction $\mathcal{H}^+$ of $\mathcal{G}$ with respect to H_+ and what it means to 'respect' THC in this extended setting. Let us now define this formally and more generally. Let H_+ be a k-complex with a linear order τ_+ on $V(H_+)$, $\mathcal{D}^+$ be a weighted hypergraph on $V(H_+)$ and $\mathcal{H}^+$ be a $V(H_+)$-partite weighted hypergraph with vertex sets $\{W_x\}_{x \in V(H_+)}$ such that $V(H) \subseteq V(H_+)$, $H_+[V(H)] = H$, $V(H)$ ordered according to τ forms an initial segment of τ_+, $\mathcal{D}^+[V(H)] = \mathcal{D}$, $W_x = V_x$ for $x \in V(H)$, $\mathcal{H}^+[\bigcup_{x \in V(H)} V_x] = \mathcal{H}$ and $\mathcal{H}^+$ is an (η, c)-THC graph with the linear order τ_+ and density weighted hypergraph $\mathcal{D}^+$. We say that a partial partite homomorphism ϕ

from H to $\mathcal{H}$ such that $\mathrm{Dom}(\phi)$ is an initial segment of τ is (η, c)-THC-*respecting for* $(\mathcal{H}^+, H_+, \tau_+)$ if the partial partite homomorphism ψ from H_+ to $\mathcal{H}^+$ given by $\psi(x) = \phi(x)_{\to\mathcal{H}^+}$ for $x \in \mathrm{Dom}(\phi)$ is (η, c)-THC-respecting for $(\mathcal{H}^+, H_+, \tau_+)$.

4.3.4 Technical Lemmas for THC

Here we provide two technical results about THC-respecting partial partite homomorphisms. The motivation for these is that we will work with THC for a variety of objects and we will need these technical results to show that THC is maintained concurrently for multiple objects of interest. The following lemma tells us that the property of THC-respecting for partial partite homomorphisms is naturally nested under suitable modification of the weighted hypergraphs; in particular, THC-respecting for partial partite homomorphisms is preserved under deletion of clusters in a suitable order-respecting manner.

Lemma 4.22. *Let $k, c \in \mathbb{N}$ and $\eta > 0$. Let G be a k-complex on a finite set J with a linear order τ and $\mathcal{G}$ be a J-partite weighted hypergraph with vertex sets $\mathcal{V} = \{V_j\}_{j\in J}$ which is identically* 1 *on any V_f such that $f \notin E(G)$ and is an (η, c)-*THC *graph with the linear order τ and density weighted hypergraph $\mathcal{D}$ on J. Let H be a k-complex on a finite set I with a linear order τ_I and $\mathcal{H}$ be an I-partite weighted hypergraph with vertex sets $\mathcal{U} = \{U_i\}_{i\in I}$ which is identically* 1 *on any V_f such that $f \notin E(H)$ and is an (η, c)-*THC *graph with the linear order τ_I and density weighted hypergraph $\mathcal{B}$ on I. Let ϕ be a* THC-*respecting partial partite homomorphism from G to $\mathcal{G}$ and let $I_\phi = \mathrm{Dom}(\phi) \cap I$. Let J' be the set of vertices in J at distance at most $c + 2$ from I_ϕ in $G^{(2)}$ and I' be the set of vertices in I at distance at most $c + 2$ from I_ϕ in $H^{(2)}$. Suppose that the following hold.*

(i) $\mathrm{Dom}(\phi)$ *is an initial segment of τ, I_ϕ is an initial segment of τ_I, $I' \subseteq J'$, for all edges $ab \in E(G[\mathrm{Dom}(\phi)])$ with $\tau(a) < \tau(b)$ we have $a \in I_\phi$ or $b \in \mathrm{Dom}(\phi) \setminus I$, and τ and τ_I induce the same order on I'.*

(ii) $H[I']$ with the linear order induced by τ_I is isomorphic to $G[I']$ with the linear order induced by τ, $\mathcal{H}\big[\bigcup_{i\in I'} U_i\big]$ is partite isomorphic to $\mathcal{G}\big[\bigcup_{j\in I'} V_j\big]$ and $\mathcal{B}[I']$ with the linear order induced by τ_I is isomorphic to $\mathcal{D}[I']$ with the linear order induced by τ.

Then the function $\phi' : \mathrm{Dom}(\phi) \cap I \to \bigcup_{i \in I} U_i$*, given by* $\phi'(x) = \phi(x)$ *for each* $x \in \mathrm{Dom}(\phi) \cap I$*, is a* THC*-respecting partial partite homomorphism from* H *to* $\mathcal{H}$*.*

Proof. Our proof proceeds by induction on $s = |\mathrm{Dom}(\phi) \cap I|$. For $s = 0$ the desired conclusion follows because $\mathcal{H}$ is an (η, c)-THC graph and $g(\varnothing) = h(\varnothing)$. Now consider $s \in [|I|]$. Let y be the last element of $\mathrm{Dom}(\phi) \cap I$ in the order according to τ_I and let ψ be the restriction of ϕ to $\{z \in \mathrm{Dom}(\phi) : \tau(z) < \tau(y)\}$. Now ψ is a THC-respecting partial partite homomorphism from G to $\mathcal{G}$ satisfying conditions (i) and (ii), so by the inductive hypothesis we conclude that the function $\psi' : \mathrm{Dom}(\psi) \cap I \to \bigcup_{i \in I} U_i$, given by $\psi'(x) = \psi(x)$ for each $x \in \mathrm{Dom}(\psi) \cap I$, is a THC-respecting partial partite homomorphism from H to $\mathcal{H}$.

By assumption ϕ is THC-respecting, so the vertex $\phi(y)$ has weight 1 in $\mathcal{G}_\psi$ and belongs to the set V'_y of (THC2) returned by an algorithm whose existence is guaranteed by (THC3) for $\mathcal{G}_\psi$. Furthermore, by (ii) we deduce that the input into the aforementioned algorithm for $\mathcal{H}_{\psi'}$ can be obtained by cluster deletion from the input into the aforementioned algorithm for $\mathcal{G}_{\psi'}$. Now by the the monotone property of aforementioned algorithm from (THC3) and since we have $\phi'(y) = \phi(y)$, it follows that ϕ' is a THC-respecting partial partite homomorphism from H to $\mathcal{H}$. □

We are also interested in how the property of THC-respecting behaves under cluster size adjustment, especially since we may work with subsets of sublinear sizes, which do not play well with THC-badness on their own. The following lemma tells us that the property of THC-respecting is well-behaved under cluster size adjustment.

Lemma 4.23. *Let* $k, c \in \mathbb{N}$*,* $\eta > 0$ *and* H *be a* k*-complex on a finite set* J *with a linear order* τ*. Let* $\mathcal{G}$ *be a* J*-partite weighted hypergraph with vertex sets* $\mathcal{V} = \{V_j\}_{j \in J}$ *which is identically* 1 *on any* V_f *such that* $f \notin E(H)$ *and is an* (η, c)-THC *graph with the linear order* τ *and density weighted hypergraph* $\mathcal{D}$ *on* J*. Let* j *be the first vertex of* J *according to* τ*,* $U_j \subseteq V_j$ *be a subset and* $\mathcal{H}$ *be the weighted hypergraph on* $V(\mathcal{G})$ *with weight function*

$$h(e) = \begin{cases} g(e) & \text{if } (e \cap V_j) \setminus U_j = \varnothing, \\ 0 & \text{otherwise.} \end{cases}$$

Suppose that $\mathcal{H}$ *is an* (η, c)-THC *graph with the linear order* τ *and density weighted hypergraph* $\mathcal{B}$ *on* J *such that we have* $d(e) = b(e)$ *for all* $e \neq \{j\}$*. Then any* THC-

respecting partial partite homomorphism ϕ from H to $\mathcal{H}$ is also THC-*respecting for* $(\mathcal{G}, H, \tau)$.

Proof. This follows immediately from the definition of THC-respecting, how $\mathcal{H}$ and $\mathcal{B}$ are related to $\mathcal{G}$ and $\mathcal{D}$ respectively, the fact that $\mathcal{G}$ and $\mathcal{H}$ are both (η, c)-THC graphs and that the link graphs after embedding the first vertex are entirely identical for both $\mathcal{G}$ and $\mathcal{H}$. □

4.4 Candidate Sets

In this section we provide definitions for some additional objects which are relevant to our embedding procedure.

4.4.1 Candidate Sets

Let J be a finite set. The *candidate set* of a J-partite hypergraph H in a J-partite weighted hypergraph $\mathcal{G}$ is

$$\mathcal{C}(H) := \left\{ \psi \text{ partite homomorphism from } H \text{ to } \mathcal{G} : \prod_{e \in E(H)} g(\psi(e)) = 1 \right\}.$$

For simplicity we will write $\mathcal{C}(e)$ to mean the candidate set of the down-closure complex of a single edge e. While the elements of $\mathcal{C}(H)$ are formally *functions*, it will be convenient to refer to their homomorphic images instead. Given vertices $x_1, \ldots, x_\ell \in V(H)$ and subsets $U_i \subseteq V_{x_i}$ for $i \in [\ell]$, set

$$\mathcal{C}(H; (U_i)_{i\in[\ell]}, (x_i)_{i\in[\ell]}) := \{\psi \in \mathcal{C}(H) : \psi(x_i) \in U_i \text{ for all } i \in [\ell]\}.$$

We will often drop the tuple brackets when $\ell = 1$ and omit the set brackets for subsets of the form $U_i = \{v_i\}$.

Let H be a hypergraph and $\mathcal{H}$ be a $V(H)$-partite weighted hypergraph with vertex sets $\{V_x\}_{x\in V(H)}$. Let ϕ be a partial partite homomorphism from H to $\mathcal{H}$. We call a vertex $x \in V(H)$ *embedded* if $x \in \mathrm{Dom}(\phi)$ and otherwise *unembedded*. Let the *candidate graph* $\mathcal{H}_\phi$ be the weighted hypergraph obtained from $\mathcal{H}$ by taking links of the vertices in $\mathrm{Im}(\phi)$ in some order. The *candidate set* of a $(V(H) \setminus \mathrm{Dom}(\phi))$-partite hypergraph F in $\mathcal{H}_\phi$ is

$$\mathcal{C}_\phi(F) := \left\{ \psi \text{ partite homomorphism from } F \text{ to } \mathcal{H} : \prod_{e \in E(F)} h(\psi(e)) = 1 \right\}.$$

For simplicity we will write $\mathcal{C}_\phi(e)$ to mean the candidate set of the down-closure complex of a single edge e. While the elements of $\mathcal{C}_\phi(F)$ are formally *functions*, it will be convenient to refer to their homomorphic images instead. Given vertices $x_1, \ldots, x_\ell \in V(F)$ and subsets $U_i \subseteq V_{x_i}$ for $i \in [\ell]$, set

$$\mathcal{C}_\phi(F; (U_i)_{i\in[\ell]}, (x_i)_{i\in[\ell]}) := \{\psi \in \mathcal{C}_\phi(F) : \psi(x_i) \in U_i \text{ for all } i \in [\ell]\}.$$

We will often drop the tuple brackets when $\ell = 1$ and omit the set brackets for subsets of the form $U_i = \{v_i\}$. Let $\vec{v}$ be a tuple of vertices in $\mathcal{H}$ such that for some partial partite homomorphism ϕ from H to $\mathcal{H}$ we have $\vec{v} = (\phi(x))_{x\in\mathrm{Dom}(\phi)}$. Write $\mathcal{H}_{\vec{v}}$ to mean the candidate graph $\mathcal{H}_\phi$ and for any $(V(H) \setminus \mathrm{Dom}(\phi))$-partite hypergraph F write $\mathcal{C}_{\vec{v}}(F)$ to mean the candidate set $\mathcal{C}_\phi(F)$.

Let H be a J-partite hypergraph with its vertex set partitioned into $\mathcal{X} = \{X_j\}_{j\in J}$ and $\mathcal{G}$ be a J-partite weighted hypergraph with its vertex set partitioned into $\mathcal{V} = \{V_j\}_{j\in J}$. Let $\mathcal{G}'$ be the standard construction of $(\mathcal{G}, \mathcal{V})$ with respect to $(H, \mathcal{X})$. Let ϕ be a partial partite homomorphism from H to $\mathcal{G}'$. For each $j \in J$ set $W_j^\phi := V_j \cap (\mathrm{Im}(\phi)_{\to\mathcal{G}})$ and for each $x \in X_j$ write W_x^ϕ to mean W_j^ϕ. For each $x \in V(H) \setminus \mathrm{Dom}(\phi)$ we define the *available candidate set*

$$A_\phi(x) := \{v \in \mathcal{C}_\phi(x) : v_{\to\mathcal{G}} \notin W_x^\phi\}.$$

In the remainder of this chapter, we will want to consider not just a single partial partite homomorphism, but rather a sequence $\phi_0, \phi_2, \ldots$ of them. We will want to refer to sets and quantities with reference to each of these partial partite homomorphisms. We will use the convention of attaching a subscript t to mean that it is with reference to ϕ_t. For example, $\mathcal{C}_t(x)$ would mean $\mathcal{C}(x)$ with reference to ϕ_t.

4.4.2 Binary Hypergraphs

We shall discuss the specific setting relevant to our situation. Many of the concepts we use are applicable to weighted hypergraphs in general, even though for us $\mathcal{G}$ will typically be the weighted analogue of a hypergraph; for the most part we will not need to concern ourselves with this. However, the setup for candidate sets is one of the few situations in which we will need to consider this particular aspect. We say that a weighted hypergraph is *binary* if it is $\{0, 1\}$-valued. This definition captures the specific property of weighted hypergraphs which arise as weighted analogues of hypergraphs.

The following lemma establishes that the sizes of candidate sets in binary weighted hypergraphs can be measured by suitable weighted homomorphic counts and that the property of being binary is preserved under the standard construction and taking links.

Lemma 4.24. *Let J be a finite set. Let $\mathcal{G}$ be a binary J-partite weighted hypergraph with its vertex set partitioned into $\mathcal{V} = \{V_j\}_{j\in J}$. Then the following hold.*

(i) $\mathcal{G}_v$ is binary for all $v \in V(\mathcal{G})$.

(ii) Let H be a J-partite hypergraph with a partition $\mathcal{X} = \{X_j\}_{j\in J}$ of $V(H)$. Then the standard construction of $(\mathcal{G}, \mathcal{V})$ with respect to $(H, \mathcal{X})$ is binary.

(iii) Given $\ell \in \mathbb{N}_0$, a J-partite hypergraph F, vertices $x_1, \ldots, x_\ell \in V(F)$ and subsets $U_i \subseteq V_{x_i}$ for $i \in [\ell]$, we have

$$\big|\mathcal{C}(F; (U_i)_{i\in[\ell]}, (x_i)_{i\in[\ell]})\big| = \mathcal{G}(F; (U_i)_{i\in[\ell]}, (x_i)_{i\in[\ell]}) \prod_{x\in V(F)} |V_x| \prod_{i\in[\ell]} \frac{|U_{x_i}|}{|V_{x_i}|}.$$

(iv) Let ϕ be a partial partite homomorphism from a hypergraph H on J to $\mathcal{G}$. Then $\mathcal{G}_\phi$ is binary and for all $(J \setminus \mathrm{Dom}(\phi))$-partite hypergraphs F we have $|\mathcal{C}_\phi(F)| = \mathcal{G}_\phi(F) \prod_{x\in V(F)} |V_x|$.

Proof. The statements (i), (ii) and (iii) follow from the definitions of $\mathcal{G}_v$, the standard construction of $(\mathcal{G}, \mathcal{V})$ with respect to $(H, \mathcal{X})$ and $\mathcal{C}(F; (U_i)_{i\in[\ell]}, (x_i)_{i\in[\ell]})$ respectively.

To prove (iv) we first prove the following by induction on $m = |\mathrm{Dom}(\phi)|$. For any partial partite homomorphism ϕ from a hypergraph H on J to $\mathcal{G}$, the weighted hypergraph $\mathcal{G}_\phi$ is binary. The case $m = 0$ follows from the lemma assumption. Now consider $m \in [|J|]$. Let x be an element of $\mathrm{Dom}(\phi)$ and ψ be the restriction of ϕ to $\mathrm{Dom}(\phi) \setminus \{x\}$. Since ψ is a partial partite homomorphism from H to $\mathcal{G}$ with $|\mathrm{Dom}(\psi)| = m - 1$, the inductive hypothesis tells us that $\mathcal{G}_\psi$ is binary. By applying (i) with $\mathcal{G}_\psi$ and $\phi(x)$ we deduce that $\mathcal{G}_\phi = (\mathcal{G}'_\psi)_{\phi(x)}$ is binary, completing our proof by induction.

Now let ϕ be a partial partite homomorphism from a hypergraph H on J to $\mathcal{G}$. Since $\mathcal{G}_\phi$ is binary, by (iii) it follows that for all $(J \setminus \mathrm{Dom}(\phi))$-partite hypergraphs F we have $|\mathcal{C}_\phi(F)| = \mathcal{G}_\phi(F) \prod_{x\in V(F)} |V_x|$. $\square$

The following lemma conveniently combines conclusions (ii) and (iv) of Lemma 4.24 for standard constructions, which we will often encounter.

Lemma 4.25. *Let J be a finite set, $\mathcal{G}$ be a binary J-partite weighted hypergraph with a partition $\mathcal{V} = \{V_j\}_{j\in J}$ of $V(\mathcal{G})$ and H be a J-partite hypergraph with a partition $\mathcal{X} = \{X_j\}_{j\in J}$ of $V(H)$. Let ϕ be a partial partite homomorphism from H to the standard construction $\mathcal{G}'$ of $(\mathcal{G}, \mathcal{V})$ with respect to $(H, \mathcal{X})$ on vertex sets $\{V'_x\}_{x\in V(H)}$. Then $\mathcal{G}'_\phi$ is binary and for all $(V(H)\setminus\mathrm{Dom}(\phi))$-partite hypergraphs F we have $|\mathcal{C}_\phi(F)| = \mathcal{G}'_\phi(F)\prod_{x\in V(F)}|V'_x|$.*

Proof. Since $\mathcal{G}'$ is binary by Lemma 4.24(ii), we can apply Lemma 4.24(iv) with $\mathcal{G}'$ in place of $\mathcal{G}$ to obtain the desired conclusions. □

The following lemma establishes a lower bound on certain complex-derived densities in $\mathcal{D}$ when $\mathcal{G}$ is binary and we have counting conditions which resemble (BUL2) and (THC1).

Lemma 4.26. *Let $k, \Delta, a \in \mathbb{N}$ and $\eta \in \left(0, \frac{1}{2}\right]$. Let R be a k-complex on a finite set J, $\mathcal{G}$ be a binary J-partite weighted hypergraph on n vertices with a partition $\mathcal{V} = \{V_j\}_{j\in J}$ of $V(\mathcal{G})$ and $\mathcal{D}$ be a weighted hypergraph on J such that $g(\varnothing) = d(\varnothing) = 1$, $d(\{j\}) = 1$ for all $j \in J$ and $(\mathcal{G}, \mathcal{V})$ is an $(\eta, (\Delta+1)a, \mathcal{D})$-typcount R-partition. Let F be a R-partite k-complex on at most $\Delta + 1$ vertices with a partition $\mathcal{F} = \{F_j\}_{j\in J}$ of $V(F)$ and a vertex $x \in V(F)$ such that $F_x = \{x\}$ and, setting $F_0 := F[V(F) \setminus \{x\}]$, we have $\mathcal{D}(F_0), \mathcal{D}(F) > 0$. Then we have*

$$\frac{\mathcal{D}(F)}{\mathcal{D}(F_0)} \geq \left(\frac{1-\eta}{(1+\eta)|V_{V(F)}|}\right)^{1/a}.$$

Proof. Let $\vec{a} := (a, a)$, $\vec{A} := (V(F) \setminus \{x\}, \{x\})$ and $F_1 := F(\vec{A}, \vec{a})$. Let $d_0 := \mathcal{D}(F_0)$ and $d_1 := \frac{\mathcal{D}(F)}{\mathcal{D}(F_0)}$. We have $\mathcal{G}(F_1) = (1 \pm \eta)\frac{\mathcal{D}(F_1)}{d(\varnothing)} = (1 \pm \eta)d_0^a d_1^{a^2}$ and $\mathcal{G}(F_0) \geq (1-\eta)d_0$. Now $\eta \in (0, 1)$ and $d_0, d_1 > 0$, so $\mathcal{G}(F_1)$ is strictly positive. Since $\mathcal{G}$ is a binary weighted hypergraph, in fact $\mathcal{G}(F_1) \geq |V_{V(F)}|^{-a}$ and $\mathcal{G}(F_0) \leq 1$. Now $d_1^{a^2} \geq \frac{\mathcal{G}(F_1)}{(1+\eta)d_0^a} \geq \left(\frac{1-\eta}{(1+\eta)|V_{V(F)}|}\right)^a$, so we have $d_1 \geq \left(\frac{1-\eta}{(1+\eta)|V_{V(F)}|}\right)^{1/a}$. □

4.4.3 Sparse Regularity and Candidate Sets

Here we prove some lemmas about sparse regularity for candidate sets. Let $k \in \mathbb{N}$, $\ell \geq 2$ and J be a finite set. Let $\mathcal{H}$ be a J-partite binary weighted hypergraph with vertex sets $\{V_j\}_{j\in J}$. Let $\mathcal{I} = \{I_i\}_{i\in[\ell]}$ be a collection of pairwise disjoint nonempty subsets of J and set $I := \bigcup_{i\in[\ell]} I_i$. Let F be an I-partite k-complex with vertex sets

$\{S_j\}_{j\in I}$. For each $i \in [\ell]$ set $T_i := \bigcup_{j\in I_i} S_j$ and write F_i for $F(T_i,(1))$. For $h \in \mathbb{N}_0$ let $x_1,\dots,x_h \in S_j \subseteq T_1$ be distinct vertices with $j \in I_1 \subseteq J$ and let $v_1,\dots,v_h \in V_j$. Write $\vec{x}$ and $\vec{v}$ for $(x_i)_{i\in[h]}$ and $(v_i)_{i\in[h]}$ respectively. Let $G^{\mathcal{H}}_{\mathcal{I},F,\vec{x},\vec{v}}$ denote the ℓ-partite ℓ-uniform hypergraph with vertex sets $\{\mathcal{C}(F_1;\vec{v},\vec{x})\} \cup \{\mathcal{C}(F_i)\}_{i\in[\ell]\setminus\{1\}}$, with $u_1\dots u_\ell$ an edge of $G^{\mathcal{H}}_{\mathcal{I},F,\vec{x},\vec{v}}$ if and only if $u_1\dots u_\ell \in \mathcal{C}(F;\vec{v},\vec{x})$. Write $G^{\mathcal{H}}_{\mathcal{I},F}$ for $G^{\mathcal{H}}_{\mathcal{I},F,(),()}$, that is, when $h=0$ and both $\vec{x}$ and $\vec{v}$ are empty tuples. The following lemma tells us that counts for a certain collection of complexes derived from F implies regularity in $G^{\mathcal{H}}_{\mathcal{I},F,\vec{x},\vec{v}}$.

Lemma 4.27. *Let $k \in \mathbb{N}$, $\ell \geq 2$ and J be a finite set. Let $\mathcal{H}$ be a J-partite binary weighted hypergraph with vertex sets $\{V_j\}_{j\in J}$. Let $\mathcal{I} = \{I_i\}_{i\in[\ell]}$ be a collection of pairwise disjoint nonempty subsets of J and set $I := \bigcup_{i\in[\ell]} I_i$. Let F be an I-partite k-complex with vertex sets $\{S_j\}_{j\in I}$. For each $i \in [\ell]$ set $T_i := \bigcup_{j\in I_i} S_j$ and write F_i for $F(T_i,(1))$. For $h \in \mathbb{N}_0$ let $x_1,\dots,x_h \in S_a \subseteq T_1$ be distinct vertices with $a \in I_1 \subseteq J$ and let $v_1,\dots,v_h \in V_a$. Write $\vec{x}$ and $\vec{v}$ for $(x_i)_{i\in[h]}$ and $(v_i)_{i\in[h]}$ respectively. Let*

$$\vec{T} := (T_1 \setminus \{x_i : i\in[h]\}, T_2,\dots,T_\ell,\{x_i : i \in [h]\}).$$

For $S \subseteq [2]$, let $\vec{q}_S$ be the $(\ell+1)$-tuple whose ith entry is 2 if $i \in S$ and 1 otherwise, and set $F_S := F(\vec{T},\vec{q}_S)$. Let $\mathcal{D}$ be a weighted hypergraph on J such that $\frac{\mathcal{D}(F_i)}{d(\varnothing)} > 0$ for all $i \in [\ell]$. Let $\eta,\varepsilon \in (0,1]$ satisfy $2^{12}(v(F)+2)\eta \leq \varepsilon^{2\ell+3}$. Suppose that $\mathcal{H}(F_S;\vec{v},\vec{x}) = (1 \pm v(F_S)\eta)\frac{\mathcal{D}(F_S)}{d(\varnothing)}$ for all $S \subseteq [2]$, $\mathcal{H}(F_i) = (1\pm v(F_i)\eta)\frac{\mathcal{D}(F_i)}{d(\varnothing)}$ for all $i \in [\ell]\setminus\{1\}$ and $\mathcal{H}(F_1;\vec{v},\vec{x}) = (1\pm v(F_1)\eta)\frac{\mathcal{D}(F_1)}{d(\varnothing)}$. Then $G^{\mathcal{H}}_{\mathcal{I},F,\vec{x},\vec{v}}$ is (ε)-regular.

Proof. Let L be the unique $[\ell]$-partite ℓ-uniform hypergraph on $[\ell]$. Set $\vec{A} := (\{i\})_{i\in[\ell]}$ and $p := \frac{\mathcal{D}(F)d(\varnothing)^{\ell-1}}{\prod_{i\in[\ell]}\mathcal{D}(F_i)} > 0$. For each $S \subseteq [2]$, set $\vec{p}_S$ to be the ℓ-tuple whose ith entry is 2 if $i \in S$ and 1 otherwise, and $L_S := L(\vec{A},\vec{p}_S)$. By Lemma 4.24(iii) applied in $\mathcal{H}$ with $\vec{x}$, $\vec{v}$, $F_1,\dots,F_\ell$ and F_S, we obtain

$$\begin{aligned} G^{\mathcal{H}}_{\mathcal{I},F,\vec{x},\vec{v}}(L_S) &= \frac{\mathcal{H}(F_S;\vec{v},\vec{x})}{\mathcal{H}(F_1;\vec{v},\vec{x})^{1+|S\cap\{1\}|}\prod_{i\in[\ell]\setminus\{1\}}\mathcal{H}(F_i)^{1+|S\cap\{i\}|}} \\ &= \frac{1\pm v(F_S)\eta}{\prod_{i\in[\ell]}(1\pm v(F_i)\eta)^{1+|S\cap\{i\}|}} \times \frac{\mathcal{D}(F_S)d(\varnothing)^{\ell+|S|-1}}{\prod_{i\in[\ell]}\mathcal{D}(F_i)^{1+|S\cap\{i\}|}} \\ &= (1\pm 4v(F_S)\eta)p^{2^{|S|}}. \end{aligned}$$

Then by Lemma 4.18 $G^{\mathcal{H}}_{\mathcal{I},F,x,v}$ is (ε)-regular. □

Let $k \in \mathbb{N}$, $\ell \geq 2$, H be a k-complex and $\mathcal{H}$ be a $V(H)$-partite binary weighted hypergraph with vertex sets $\{V_x\}_{x \in V(H)}$ which is identically 1 on any V_f such that $f \notin E(H)$. Let ϕ be a partial partite homomorphism from H to $\mathcal{H}$. Let $\mathcal{I} = \{I_i\}_{i \in [\ell]}$ be a collection of pairwise disjoint nonempty subsets of $V(H) \setminus \mathrm{Dom}(\phi)$ and set $I := \bigcup_{i \in [\ell]} I_i$. Let F be an I-partite k-complex with vertex sets $\{S_j\}_{j \in I}$. For each $i \in [\ell]$ set $T_i := \bigcup_{j \in I_i} S_j$ and write F_i for $F(T_i, (1))$. Let $G^{\phi}_{\mathcal{I},F}$ denote the ℓ-partite ℓ-uniform hypergraph with vertex sets $\{\mathcal{C}_\phi(F_i)\}_{i \in [\ell]}$, with $u_1 \dots u_\ell$ an edge of $G^{\phi}_{\mathcal{I},F}$ if $u_1 \dots u_\ell \in \mathcal{C}_\phi(F)$. Write $G^{\phi}_{\mathcal{I}}$ for $G^{\phi}_{\mathcal{I},H[I]}$. The following lemma tells us that the counting condition of THC implies regularity in $G^{\phi}_{\mathcal{I},F}$.

Lemma 4.28. *Let $k, c \in \mathbb{N}$, $\ell \geq 2$ and $0 < \eta \leq 2^{-12}c^{-1}$. Let H be a k-complex with a linear order τ on $V(H)$. Let $\mathcal{H}$ be a $V(H)$-partite binary weighted hypergraph with vertex sets $\{V_x\}_{x \in V(H)}$ which is identically 1 on any V_f such that $f \notin E(H)$ and $\mathcal{D}$ be a weighted hypergraph on $V(H)$ such that $\mathcal{H}$ is an (η, c)-THC graph with the linear order τ on $V(H)$ and density weighted hypergraph $\mathcal{D}$. Let ϕ be a THC-respecting partial partite homomorphism from H to $\mathcal{H}$. Let $\mathcal{I} = \{I_i\}_{i \in [\ell]}$ be a collection of pairwise disjoint nonempty subsets of $V(H) \setminus \mathrm{Dom}(\phi)$ and set $I := \bigcup_{i \in [\ell]} I_i$. Let F be an I-partite k-complex with vertex sets $\{S_x\}_{x \in I}$ such that $v(F) \leq c - 2$. Suppose that $\frac{\mathcal{D}_\phi(F_i)}{d_\phi(\varnothing)} > 0$ for all $i \in [\ell]$. Set $\varepsilon := (2^{12}(v(F)+2)\eta)^{1/(2\ell+3)}$. Then $G^{\phi}_{\mathcal{I},F}$ is (ε)-regular.*

Proof. Since $\mathcal{H}$ is an (η, c)-THC graph with the linear order τ on $V(H)$ and density weighted hypergraph $\mathcal{D}$ on $V(H)$ and ϕ is a THC-respecting partial partite homomorphism from H to $\mathcal{H}$, we have that $\mathcal{H}_\phi$ is an (η, c)-THC graph with the linear order on $V(H) \setminus \mathrm{Dom}(\phi)$ induced by τ and density weighted hypergraph $\mathcal{D}_\phi$ on $V(H) \setminus \mathrm{Dom}(\phi)$ such that $h_\phi(\varnothing) = 1$. For each $i \in [\ell]$ set $T_i := \bigcup_{x \in I_i} S_x$ and write F_i for $F(T_i, (1))$. Let $\vec{T} := (T_1, \dots, T_\ell)$. For $S \subseteq [2]$, let $\vec{p}_S$ be the ℓ-tuple whose ith entry is 2 if $i \in S$ and 1 otherwise, and set $F_S := F(\vec{T}, \vec{p}_S)$. By (THC1) we have $\mathcal{H}_\phi(F_S) = (1 \pm v(F_S)\eta)\frac{\mathcal{D}_\phi(F_S)}{d_\phi(\varnothing)}$ for all $S \subseteq [2]$ and $\mathcal{H}_\phi(F_i) = (1 \pm v(F_i)\eta)\frac{\mathcal{D}_\phi(F_i)}{d_\phi(\varnothing)}$ for all $i \in [\ell]$. Hence, $G^{\phi}_{\mathcal{I},F}$ is (ε)-regular by Lemma 4.27 with $h = 0$. $\square$

4.5 Good Partitions

The preprocessing stage of our embedding strategy involves refining the input partitions $\mathcal{X}^{\mathrm{LEM}}$ of $V(H)$ and $\mathcal{V}^{\mathrm{LEM}}$ of $V(\mathcal{G})$ to obtain new partitions $\mathcal{X}$ of $V(H)$ and $\mathcal{V}$ of $V(\mathcal{G})$ which have additional properties that we need for our proof. This in turn entails the replacement of various other input objects (for example, input reduced complex R_{LEM} by a new complex R) and the modification of constants. Since it is convenient to retain the notational choices of our blow-up lemma for the modified objects and constants, we use the suffix LEM to refer to the original input object.

We introduce the notion of a good vertex order for our random greedy algorithm. We need to analyse how the neighbours of buffer vertices are embedded during the course of our random greedy algorithm and putting them first in the order, with the neighbourhood of each buffer vertex forming an interval in the order, enables this.

Definition 4.29 (Good vertex order)**.** We say that a linear order τ on $V(H)$ is a *good vertex order* for $X \subseteq V(H)$ if the following conditions are satisfied.

(VO1) For all $x \in N_{H^{(2)}}(X)$ and $y \in V(H) \setminus N_{H^{(2)}}(X)$ we have $\tau(x) < \tau(y)$.

(VO2) For all $x \in X$ and $y \in V(H) \setminus X$ we have $\tau(y) < \tau(x)$.

(VO3) For all $x \in X$ we can enumerate $N_{H^{(2)}}(x)$ as $y_1, \dots, y_b$ such that $\tau(y_{h+1}) = \tau(y_h) + 1$ for all $h \in [b-1]$.

Now we define the properties we need from the refined partitions of H and $\mathcal{G}$ and show that we can obtain such refined partitions.

Definition 4.30 (Good H-partition)**.** A partition $\mathcal{X} = \{X_j\}_{j \in J}$ of $V(H)$ with a subpartition $\{X_{j,h}\}_{h \in [\ell_j]}$ of each part X_j is an $(\alpha, c, \Delta, \Delta_R, \kappa, \mu)$-*good H-partition* for reduced complexes $R' \subseteq R$ on a set J, with potential buffer $\widetilde{\mathcal{X}} = \{\widetilde{X}_j\}_{j \in J}$, buffer vertices $X^{\mathrm{buf}} \subseteq \bigcup_{j \in J} \widetilde{X}_j$ and a linear order τ on $V(H)$, if the following conditions are satisfied for each $j \in J$ and $h \in [\ell_j]$.

(H1) $(H, \mathcal{X})$ is an R-partition, $\widetilde{\mathcal{X}}$ is an $(\alpha, (\Delta^2 + \Delta + 2)c, R')$-buffer for $(H, \mathcal{X})$ and $\Delta(R^{(2)}) \leq \Delta_R$.

(H2) $\mathrm{dist}_{H^{(2)}}(x, y) \geq 7$ for each $x, y \in X_j$ with $x \neq y$.

(H3) $\mathrm{dist}_{H^{(2)}}(x, y) \geq c + 5$ for each $x, y \in X^{\mathrm{buf}}$ with $x \neq y$.

(H4) $|X_j^{\mathrm{buf}}| = 4\mu |X_j|$ and $|\{x \in X_j : x \in N_{H^{(2)}}(X^{\mathrm{buf}})\}| \leq 4\Delta_R \mu \kappa |X_j|$.

(H5) τ is a good vertex order for X^{buf}.

(H6) $H^{\leq 1}(x)$ is the same ordered complex F up to partite isomorphism for all $x \in X_j^{\mathrm{buf}}$. We then call X_j^{buf} an F-buffer.

(H7) $H^{\leq 3}(x)$ is the same ordered complex up to partite isomorphism for all $x \in X_{j,h}$.

(H8) $\ell_j \leq 2^{2^{\Delta_R^3+1}}$.

Definition 4.31 (Good $\mathcal{G}$-partition)**.** A $(c, \Delta, \Delta_R, \eta, \mu)$*-good $\mathcal{G}$-partition* is a partition $\mathcal{V} = \{V_j\}_{j \in J}$ of $V(\mathcal{G})$ with a subpartition $V_j = V_j^{\mathrm{main}} \sqcup V_j^{\mathrm{q}} \sqcup V_j^{\mathrm{buf}}$ for each $j \in J$, reduced complexes $R' \subseteq R$ and density weighted hypergraph $\mathcal{D}$, all on vertex set J, where the following conditions are satisfied. Let $\Delta_{\mathrm{aux}} := 2^{2^{\Delta_R^2+1}+\Delta^2+1}(\Delta+1)\Delta$.

(G1) $|V_j^{\mathrm{main}}| = (1-2\mu)|V_j|$ and $|V_j^{\mathrm{q}}| = |V_j^{\mathrm{buf}}| = \mu|V_j|$ for each $j \in J$.

(G2) $(\mathcal{G}, \mathcal{V})$ is an $(\eta, (\Delta_{\mathrm{aux}}+2)(\Delta+2)c, \mathcal{D})$-typcount R-partition.

(G3) $(\mathcal{G}, \mathcal{V})$ is $(\eta, (\Delta^2+\Delta+2)c, \mathcal{D})$-super-typcount on R'.

(G4) For each $ij \in E(R')$ and $v \in V_i$, we have

$$\deg_{\mathcal{G}^{(2)}}(v; V_j^{\mathrm{main}}) \geq (1-\eta)(1-2\mu)\deg_{\mathcal{G}^{(2)}}(v; V_j),$$
$$\deg_{\mathcal{G}^{(2)}}(v; V_j^{\mathrm{q}}) \geq (1-\eta)\mu\deg_{\mathcal{G}^{(2)}}(v; V_j).$$

The following lemma tells us that we can obtain good partitions of H and $\mathcal{G}$ from the partitions provided in our main theorem.

Lemma 4.32. *For all $k, \Delta, \Delta_R, c \in \mathbb{N}$, $\kappa \geq 2$, finite sets J, $\alpha, \eta \in (0,1]$ and sufficiently small $\mu > 0$, there exists $n_0 \in \mathbb{N}$ such that the following holds. Let $\beta := \frac{1}{8(\Delta^6+1)}$, $c_1 := \left(2^{2^{\Delta_R^2+1}+\Delta^2}(\Delta+1)\Delta+1\right)(\Delta+2)c$, $c_2 := (\Delta^2+\Delta+2)c$ and $\eta^{\mathrm{LEM}} := \frac{\eta^{4c_1+3}}{c_1 2^{4c_1+17}}$. Let R_{LEM} be a k-complex on a set J^{LEM} with $|J^{\mathrm{LEM}}| = \beta|J|$ and let R'_{LEM} be a spanning subcomplex of R_{LEM}. Let H and $\mathcal{G}$ be J^{LEM}-partite k-complexes on $n \geq n_0$ vertices with $\frac{\kappa}{2}$-balanced size-compatible vertex partitions $\mathcal{X}^{\mathrm{LEM}} = \{X_j^{\mathrm{LEM}}\}_{j \in J^{\mathrm{LEM}}}$ and $\mathcal{V}^{\mathrm{LEM}} = \{V_j^{\mathrm{LEM}}\}_{j \in J^{\mathrm{LEM}}}$ respectively which have parts of size at least $\frac{2n}{\kappa\beta|J|}$, such that $\Delta(H^{(2)}) \leq \Delta$, $\varnothing \in E(\mathcal{G})$ and $\{v\} \in E(\mathcal{G})$ for all $v \in V(\mathcal{G})$. Let $\mathcal{D}^{\mathrm{LEM}}$ be a weighted hypergraph on J^{LEM} with $d^{\mathrm{LEM}}(\varnothing) = 1$, $d^{\mathrm{LEM}}(\{j\}) = 1$ for all $j \in J^{\mathrm{LEM}}$ and $d^{\mathrm{LEM}}(e) > 0$ for all $e \in E(R_{\mathrm{LEM}})$. Let $\widetilde{\mathcal{X}}^{\mathrm{LEM}} = \{\widetilde{X}_j^{\mathrm{LEM}}\}_{j \in J^{\mathrm{LEM}}}$ be a family of subsets of $V(H)$. Suppose that*

(GP1) We have $\Delta(R^{(2)}_{\mathrm{LEM}}) \leq \beta\Delta_R$, $(H, \mathcal{X}^{\mathrm{LEM}})$ *is an* R_{LEM}*-partition and* $\widetilde{\mathcal{X}}^{\mathrm{LEM}}$ *is an* $(2\alpha, c_2, R'_{\mathrm{LEM}})$*-buffer for* $(H, \mathcal{X}^{\mathrm{LEM}})$,

(GP2) $(\mathcal{G}, \mathcal{V}^{\mathrm{LEM}})$ *is an* $(\eta^{\mathrm{LEM}}, 2c_1, \mathcal{D})$*-typcount* R_{LEM}*-partition.*

(GP3) $(\mathcal{G}, \mathcal{V}^{\mathrm{LEM}})$ *is* $(\eta, c_2, \mathcal{D})$*-super-typcount on* R'_{LEM}.

Then there is a k*-complex* R *on* J *and a spanning subcomplex* R' *of* R, κ*-balanced size-compatible vertex partitions* $\mathcal{X} = \{X_j\}_{j\in J}$ *and* $\mathcal{V} = \{V_j\}_{j\in J}$ *of* H *and* $\mathcal{G}$ *respectively whose parts are of size at least* $\frac{n}{\kappa|J|}$*, a weighted hypergraph* $\mathcal{D}$ *on* J *with* $d(\varnothing) = 1$, $d(\{j\}) = 1$ *for all* $j \in J$ *and* $d(e) > 0$ *for all* $e \in E(R)$*, a family* $\widetilde{\mathcal{X}} = \{\widetilde{X}_j\}_{j\in J}$ *of subsets of* $V(H)$*, subsets* $X_j^{\mathrm{buf}} \subseteq \widetilde{X}_j$ *for each* $j \in J$*, a partition* $\{X_{j,h}\}_{h\in[\ell_j]}$ *of* X_j *for each* $j \in J$*, and a partition* $V_j = V_j^{\mathrm{main}} \cup V_j^{\mathrm{q}} \cup V_j^{\mathrm{buf}}$ *for each* $j \in J$*, which give an* $(\alpha, c, \Delta, \Delta_R, \kappa, \mu)$*-good* H*-partition and a* $(c, \Delta, \Delta_R, \eta, \mu)$*-good* $\mathcal{G}$*-partition.*

Let us give a proof outline for this lemma. The proof of this lemma is rather similar to that of the good partitions lemma in [4]. The goal is to obtain good partitions of H and $\mathcal{G}$ from the partitions provided in our main theorem. We draw an auxiliary graph $G_{j'}$ on each part $X_{j'}^{\mathrm{LEM}}$ with edges between pairs of vertices at distance less than 7 and apply our variant of the Hajnal–Szemerédi theorem (Lemma 4.19) to each auxiliary graph together with $\widetilde{\mathcal{X}}_{j'}^{\mathrm{LEM}}$; we obtain refined partitions $\mathcal{X}$ and $\widetilde{\mathcal{X}}$ with new reduced complexes R and R' that satisfy the first two properties of a good H-partition. We greedily construct the sets X_j^{buf} and the good vertex order τ to obtain the buffer-related properties and construct the subpartitions $\{X_{j,h}\}_{h\in[\ell_j]}$ to obtain the remaining two properties. Then, we perform a (compatible) random refinement of $\mathcal{V}^{\mathrm{LEM}}$ to obtain a partition $\mathcal{V}$ and a further random subpartition of each cluster V_j into V_j^{main}, V_j^{q} and V_j^{buf}. To show that we do obtain the desired good $\mathcal{G}$-partition, we apply the concentration inequalities of Theorem 4.11 to establish size concentration for various subsets of the randomly selected clusters.

Proof. We require
$$\mu \leq \frac{\alpha}{2^{2^{\Delta_R+1}+5}\kappa\Delta^{c+4}\Delta_R^{c+4}}.$$
Set $n_0 = \frac{2^{8c_2+40}\kappa^4|J|^3}{\beta^2\mu^4\eta^8}$.

To obtain properties (H1) and (H2), we shall first refine the vertex partition $\mathcal{X}^{\mathrm{LEM}}$ of H. For each $j' \in J^{\mathrm{LEM}}$ let $G_{j'}$ be the graph with vertex set $X_{j'}^{\mathrm{LEM}}$ and edge set

$\{xy : x, y \in X_{j'}^{\mathrm{LEM}}, x \neq y, \mathrm{dist}_{H^{(2)}}(x, y) < 7\}$. Note that $\Delta(G_{j'}) \leq \Delta^6 + 1$. For each $j' \in J^{\mathrm{LEM}}$ we apply Lemma 4.19 with $G_{j'}$ and $\widetilde{\mathcal{X}}_{j'}^{\mathrm{LEM}}$ to obtain an equitable partition $\{X_j\}_{j \in J_{j'}}$ of X_i^{LEM} into β^{-1} parts which also partitions $\widetilde{\mathcal{X}}_{j'}^{\mathrm{LEM}}$ equitably and whose parts are independent sets in $G_{j'}$. Combining these partitions of $X_{j'}^{\mathrm{LEM}}$ and setting J to be the disjoint union of the sets $J_{j'}$, we obtain a partition $\mathcal{X} = \{X_j\}_{j \in J}$ of $V(H)$ into $|J|$ parts. Similarly, we also obtain a family $\widetilde{\mathcal{X}} = \{\widetilde{X}_j\}_{j \in J}$ of subsets of $V(H)$. We obtain R from R_{LEM} by replacing each $j' \in J^{\mathrm{LEM}}$ with an independent set on $J_{j'}$ and each ℓ-edge with a complete ℓ-partite ℓ-uniform hypergraph between the corresponding sets. We obtain R' from R'_{LEM} similarly. We obtain $\mathcal{D}$ from $\mathcal{D}^{\mathrm{LEM}}$ by replacing each $j' \in J^{\mathrm{LEM}}$ with an independent set on $J_{j'}$ and each ℓ-edge with a complete ℓ-partite ℓ-uniform hypergraph between the corresponding sets such that those edges have the same weight as the original ℓ-edge, and giving all other subsets weight zero. By construction of $\mathcal{D}$ and R, the weight function d of $\mathcal{D}$ satisfies $d(\varnothing) = 1$, $d(\{j\}) = 1$ for all $j \in J$ and $d(e) > 0$ for all $e \in E(R)$.

By construction (H2) is satisfied and $(H, \mathcal{X})$ is an R-partition. Furthermore, we have $\Delta(R) = \beta^{-1}\Delta(R_{\mathrm{LEM}}) \leq \Delta_R$. Since each part of $\mathcal{X}^{\mathrm{LEM}}$ has size at least $\frac{2n}{\kappa\beta|J|}$ and is equitably partitioned into β^{-1} parts, and n is sufficiently large, we have $|X_j| \geq \frac{n}{\kappa|J|}$ for each $j \in J$. Given $i, j \in J$ let $i', j' \in J^{\mathrm{LEM}}$ be such that $X_i \subseteq X_{i'}^{\mathrm{LEM}}$ and $X_j \subseteq X_{j'}^{\mathrm{LEM}}$. Then, noting that $|X_j| \geq \frac{n}{\kappa|J|}$ and that n is sufficiently large, we have

$$|X_i| \leq \beta|X_{i'}^{\mathrm{LEM}}| + 1 \leq \frac{\kappa\beta}{2}|X_{j'}^{\mathrm{LEM}}| + 1 \leq \frac{\kappa}{2}(|X_j| + 1) + 1 \leq \kappa|X_j|,$$

and

$$|\widetilde{X}_j| \geq \beta|\widetilde{X}_{j'}^{\mathrm{LEM}}| - 1 \geq 2\alpha\beta|X_{j'}^{\mathrm{LEM}}| - 1 \geq 2\alpha(|X_j| - 1) - 1 \geq \alpha|X_j|.$$

Hence, $\mathcal{X}$ is κ-balanced and $\widetilde{\mathcal{X}}$ is an (α, c_2, R')-buffer for $(H, \mathcal{X})$. Therefore, (H1) is satisfied.

To obtain properties (H3)–(H4), we shall choose the sets X_j^{buf} for each $j \in J$. We first partition $\widetilde{X}_j$ according to the partite isomorphism classes of the complexes $H[N_{H^{(2)}}(x) \cup \{x\}]$ for $x \in \widetilde{X}_j$. Note that there are at most $2^{2^{\Delta_R+1}}$ subsets in this partition. We take the largest subset S_j and choose X_j^{buf} greedily from within S_j, picking vertices one at a time and always at distance at least $c + 5$ from previously chosen buffer vertices. Any path in $H^{(2)}$ from x in some $\widetilde{X}_i$ to a vertex in $\widetilde{X}_j$ of length at most $c + 4$ corresponds to a path from i to j in $R^{(2)}$ of the same length obtained by taking indices, so vertices in X^{buf} at distance at most $c + 4$ from some vertex of

$\widetilde{X}_j$ must lie in the at most $\Delta_R^{c+4}+1$ subsets X_i^{buf} such that i is at distance at most $c+4$ from j in R. Furthermore, there are at most $\Delta^{c+4}+1$ vertices at distance at most $c+4$ from each vertex in $H^{(2)}$. Hence, there are at most $(\Delta^{c+4}+1)(\Delta_R^{c+4}+1)4\mu\kappa|X_j|$ vertices in $\widetilde{X}_j$ which are at distance at most $c+4$ from some vertex of X^{buf}. Since $|S_j| \geq 2^{-2^{\Delta_R+1}}\alpha|X_j|$, at each step we have at least

$$2^{-2^{\Delta_R+1}}\alpha|X_j| - (\Delta^{c+4}+1)(\Delta_R^{c+4}+1)4\kappa\mu|X_j| \geq 2^{-2^{\Delta_R+1}-1}\alpha|X_j|$$

vertices in S_j to choose from. Hence, we are able to choose the desired vertices of X_j^{buf}, so we have (H3).

Note that a vertex of X_i^{buf} has a neighbour in X_j only if $ij \in R^{(2)}$; since any pair of vertices in X_j are at distance at least seven, each such vertex has at most one neighbour in X_j. Hence, at most $4\Delta_R\mu\kappa|X_j|$ vertices of X_j are in $N_{H^{(2)}}(X^{\mathrm{buf}})$, so (H4) holds. Now we construct a good vertex order for X^{buf} on $V(H)$. For each $j \in J$ fix an ordering σ_j of the elements of the unique set equal to $i(N_{H^{(2)}}(x))$ for any $x \in X_j^{\mathrm{buf}}$. Enumerate the elements of X^{buf} as $x_1, \ldots, x_{|X^{\mathrm{buf}}|}$ and start with the empty order. For each i in succession, append the vertices of $N_{H^{(2)}}(x_i)$ in the order corresponding to σ_j where $x_i \in X_j$. Next, append the remaining vertices of $V(H) \setminus X^{\mathrm{buf}}$ in an arbitrary order. Finally, append the vertices of X^{buf} in an arbitrary order; let τ be the resultant linear order on $V(H)$. By construction, τ is a good vertex order for X^{buf} on $V(H)$ so (H5) is satisfied. Let $j \in J$. Since for any $x \in X_j^{\mathrm{buf}}$ the complexes $H[N_{H^{(2)}}(x) \cup \{x\}]$ are the same up to partite isomorphism and the vertices of $N_{H^{(2)}}(x)$ are ordered according to σ_j, we have that (H6) is satisfied.

Now we obtain properties (H7) and (H8). For each $j \in J$ partition X_j into $\{X_{j,h}\}_{h\in[\ell_j]}$ according to the partite isomorphism classes of the ordered complexes $H^{\leq 3}(x)$ for $x \in X_j$. By construction, for each $h \in [\ell_j]$ we have that $H^{\leq 3}(x)$ is the same ordered complex up to partite isomorphism for all $x \in X_{j,h}$, so (H7) is satisfied automatically. Let $j \in J$. By (H2), for any $x \in X_j$ the vertices of $H^{\leq 3}(x)$ have distinct indices, so the index $\iota(E(H^{\leq 3}(x)))$ uniquely identifies each partite isomorphism class. We have that $\iota(E(H^{\leq 3}(x)))$ is a subcomplex of $H^{\leq 3}(j)$ for all $x \in X_j$; since $\Delta(R^{(2)}) \leq \Delta_R$, we have $|N^{\leq 3}(j)| \leq \Delta_R^3 + 1$. Hence, there are at most $2^{2^{\Delta_R^3+1}}$ partite isomorphism classes and (H8) is satisfied. This shows that we have a $(\alpha, c, \Delta, \Delta_R, \kappa, \mu)$-good H-partition.

We shall now refine the partition $\mathcal{V}^{\mathrm{LEM}}$. For each $V_{j'}^{\mathrm{LEM}}$ choose, uniformly at random, an equitable partition into β^{-1} parts and assign these parts to parts of $X_{j'}^{\mathrm{LEM}}$

of the corresponding sizes, with parts of each size assigned uniformly at random; this is possible as $|X_{j'}^{\mathrm{LEM}}| = |V_{j'}^{\mathrm{LEM}}|$. For each $j' \in J^{\mathrm{LEM}}$ we obtain an equitable partition $\{X_j\}_{j\in J_{j'}}$ of $X_{j'}^{\mathrm{LEM}}$ such that $|V_j| = |X_j|$ for all $j \in J_{j'}$; combining these partitions of $X_{j'}^{\mathrm{LEM}}$ for $j \in J_{j'}$, we obtain a partition $\mathcal{V} = \{V_j\}_{j\in J}$ of $V(\mathcal{G})$. Then, for each $j \in J$ we choose, uniformly at random, a partition of V_j into one set V_j^{main} of size $(1-2\mu)|V_j|$ and two sets, V_j^{q} and V_j^{buf}, of size $\mu|V_j|$ each.

By construction $\mathcal{X}$ and $\mathcal{V}$ are size-compatible and (G1) holds. For $V_j \subseteq V_{j'}^{\mathrm{LEM}}$, note that each of the sets $V_j, V_j^{\mathrm{main}}, V_j^{\mathrm{q}}$ and V_j^{buf} is distributed as a uniformly random set of its size in $V_{j'}^{\mathrm{LEM}}$. Hence, given any set $U \subseteq V_{j'}^{\mathrm{LEM}}$, the size of each of the intersection sets $V_j \cap U, V_j^{\mathrm{main}} \cap U, V_j^{\mathrm{q}} \cap U$ and $V_j^{\mathrm{buf}} \cap U$ is hypergeometrically distributed. We shall work with the following choices of U: $N_{\mathcal{G}^{(2)}}(v; V_{j'}^{\mathrm{LEM}})$ for $v \in V_{i'}^{\mathrm{LEM}}$ and $j' \in J^{\mathrm{LEM}}$ with $i'j' \in E(R'_{\mathrm{LEM}})$.

Let $v \in V_{i'}^{\mathrm{LEM}}$ and $j' \in J^{\mathrm{LEM}}$ be such that $i'j' \in E(R'_{\mathrm{LEM}})$. Let $F_{i'j'}$ be the down-closure complex of $\{i', j'\}$. By the conditions on $\mathcal{D}^{\mathrm{LEM}}$ and applying (GP3) with $F_{i'j}$, v and i', we obtain

$$\deg_{\mathcal{G}^{(2)}}(v; V_{j'}^{\mathrm{LEM}}) \geq (1-\eta^{\mathrm{LEM}})d^{\mathrm{LEM}}(i'j')|V_{j'}^{\mathrm{LEM}}|.$$

Since $\mathcal{V}^{\mathrm{LEM}}$ is $\frac{\kappa}{2}$-balanced, we have $|V_{j'}^{\mathrm{LEM}}| \leq \frac{\kappa n}{2\beta|J|}$ for all $j' \in J^{\mathrm{LEM}}$. Hence, by Theorem 4.11 and Lemma 4.26 applied with c_1, $F_{i'j'}$ and i', the probability that the size of a given one of the intersection sets is not within a $\left(1 \pm \frac{\eta}{2^{c_2+1}}\right)$-factor of its expectation is at most

$$2\exp\left(-\frac{\eta^2\mu\beta\deg_{\mathcal{G}^{(2)}}(v; V_{j'}^{\mathrm{LEM}})}{3\cdot 2^{2c_2+4}}\right) \leq 2\exp\left(-\frac{\eta^2\mu\beta}{2^{2c_2+6}\kappa}\left(\frac{n}{\beta|J|}\right)^{1-2/c_1}\right).$$

There are at most $4|J|$ randomly selected sets and for each of these sets there are at most n sets U of interest. Hence, by taking a union bound, we find that asymptotically almost surely each of the intersection sets has size within a $\left(1 \pm \frac{\eta}{2^{c_2+1}}\right)$-factor of its expectation. At this point, we fix a partition $\mathcal{V}$ with this property and aim to show that this indeed gives a good $\mathcal{G}$-partition. (G4) follows directly from the construction of $\mathcal{V}$ and the guaranteed good properties of the relevant intersection sets.

Let F be a J-partite k-complex on at most $2c_1$ vertices with its vertex set partitioned into $\mathcal{F} = \{F_j\}_{j\in J}$. For each $j' \in J^{\mathrm{LEM}}$ set $F_{j'}^{\mathrm{LEM}} := \bigcup_{j\in J_{j'}} F_j$. We first consider when F is not J^{LEM}-partite with vertex set partition $\mathcal{F}^{\mathrm{LEM}} := \{F_{j'}^{\mathrm{LEM}}\}_{j'\in J^{\mathrm{LEM}}}$. Since edges of $\mathcal{G}$ and $\mathcal{D}$ of size at most k which are not J^{LEM}-partite have weight zero, we have

$\mathcal{G}(F \mid J; \mathcal{X}, \mathcal{V}) = \mathcal{D}(F \mid J; \mathcal{X}, \mathcal{V}) = 0$ and so $\mathcal{G}(F \mid J; \mathcal{X}, \mathcal{V}) = (1 \pm \eta)\mathcal{D}(F \mid J; \mathcal{X}, \mathcal{V})$ holds.

Now we consider when F is J^{LEM}-partite, and so R_{LEM}-partite, with its vertex set partitioned into $\mathcal{F}^{\mathrm{LEM}}$. Let $I := \{j' \in J^{\mathrm{LEM}} : F^{\mathrm{LEM}}_{j'} \neq \varnothing\}$. Note that if $|I| \leq 1$, then F contains no edges of size at least 2; in this case, we have $\mathcal{G}(F \mid J; \mathcal{X}, \mathcal{V}) = \mathcal{D}(F \mid J; \mathcal{X}, \mathcal{V}) = 1$ and so $\mathcal{G}(F \mid J; \mathcal{X}, \mathcal{V}) = (1 \pm \eta)\mathcal{D}(F \mid J; \mathcal{X}, \mathcal{V})$ holds. Now consider when $|I| \geq 2$. By the construction of $\mathcal{D}$ we have

$$\mathcal{D}(F \mid J; \mathcal{F}, \mathcal{V}) = \mathcal{D}^{\mathrm{LEM}}(F \mid J^{\mathrm{LEM}}; \mathcal{F}^{\mathrm{LEM}}, \mathcal{V}^{\mathrm{LEM}}).$$

Noting that we have the necessary counting conditions by (GP2), we apply Lemma 4.27 with $\mathcal{G}$ as a J^{LEM}-partite k-complex with the vertex partition $\mathcal{V}^{\mathrm{LEM}}$, the collection $\mathcal{I} = \{\{j'\}\}_{j' \in J^{\mathrm{LEM}}}$, the R_{LEM}-partite k-complex F, $h = 0$ and the weighted hypergraph $\mathcal{D}^{\mathrm{LEM}}$ on J^{LEM}, to deduce that $G^{\mathcal{G}}_{\mathcal{I},F}$ is $(\frac{\eta}{2})$-regular. Note that each edge of $G^{\mathcal{G}}_{\mathcal{I},F}$ corresponds to a J^{LEM}-partite copy of F in J^{LEM}-partite $\mathcal{G}$. Now the J-partite copies of F in J-partite $\mathcal{G}$, where each $x \in V(F)$ is mapped into F_x, correspond to the edges of $G^{\mathcal{G}}_{\mathcal{I},F}$ where for each $j' \in J^{\mathrm{LEM}}$ the tuple $(x)_{x \in F^{\mathrm{LEM}}_{j'}}$ is mapped into $\prod_{x \in F^{\mathrm{LEM}}_{j'}} F_x$. Hence, by the $(\frac{\eta}{2})$-regularity of $G^{\mathcal{G}}_{\mathcal{I},F}$ and (GP2) we obtain

$$\begin{aligned}
\mathcal{G}(F \mid J; \mathcal{F}, \mathcal{V}) &= \left(1 \pm \frac{\eta}{2}\right) \mathcal{G}(F \mid J^{\mathrm{LEM}}; \mathcal{F}^{\mathrm{LEM}}, \mathcal{V}^{\mathrm{LEM}}) \\
&= \left(1 \pm \frac{\eta}{2}\right) (1 \pm \eta^{\mathrm{LEM}}) \mathcal{D}(F \mid J^{\mathrm{LEM}}; \mathcal{F}^{\mathrm{LEM}}, \mathcal{V}^{\mathrm{LEM}}) \\
&= (1 \pm \eta)\mathcal{D}(F \mid J; \mathcal{F}, \mathcal{V}),
\end{aligned}$$

thus establishing (G2).

Let F be a R'-partite k-complex on at most $(\Delta^2 + \Delta + 2)c + 1$ vertices with its vertex set partitioned into $\mathcal{F} = \{F_j\}_{j \in J}$. For each $j' \in J^{\mathrm{LEM}}$ set $F^{\mathrm{LEM}}_{j'} := \bigcup_{j \in J_{j'}} F_j$. Let $x \in V(F)$. Let $b \in J^{\mathrm{LEM}}$ and $a \in J$ be such that $x \in F_a \subseteq F^{\mathrm{LEM}}_b$. Let $v \in V_a$. If F is not J^{LEM}-partite with vertex set partition $\mathcal{F}^{\mathrm{LEM}} := \{F^{\mathrm{LEM}}_{j'}\}_{j' \in J^{\mathrm{LEM}}}$, then as before we have $\mathcal{G}(F; v, x \mid J; \mathcal{X}, \mathcal{V}) = \mathcal{D}(F \mid J; \mathcal{X}, \mathcal{V}) = 0$ and so $\mathcal{G}(F; v, x \mid J; \mathcal{X}, \mathcal{V}) = (1 \pm \eta)\mathcal{D}(F \mid J; \mathcal{X}, \mathcal{V})$ holds.

Now we consider when F is J^{LEM}-partite, and so R'_{LEM}-partite, with its vertex set partitioned into $\mathcal{F}^{\mathrm{LEM}}$. By the construction of $\mathcal{D}$ we have

$$\mathcal{D}(F \mid J; \mathcal{F}, \mathcal{V}) = \mathcal{D}^{\mathrm{LEM}}(F \mid J^{\mathrm{LEM}}; \mathcal{F}^{\mathrm{LEM}}, \mathcal{V}^{\mathrm{LEM}}).$$

Let $I := \{j' \in J^{\mathrm{LEM}} : F^{\mathrm{LEM}}_{j'} \neq \varnothing\}$. If $|I| \leq 1$, then F contains no edges of size at least 2; in this case, we have $\mathcal{G}(F \mid J; \mathcal{X}, \mathcal{V}) = \mathcal{D}(F \mid J; \mathcal{X}, \mathcal{V}) = 1$ and so $\mathcal{G}(F \mid J; \mathcal{X}, \mathcal{V}) = (1 \pm \eta)\mathcal{D}(F \mid J; \mathcal{X}, \mathcal{V})$ holds. If $|I| = 2$ and $F^{\mathrm{LEM}}_b = \{x\}$, then F is a star graph; in this case, by our construction of $\mathcal{V}$ we obtain

$$\begin{aligned}\mathcal{G}(F; v, x \mid J; \mathcal{F}, \mathcal{V}) &= \left(1 \pm \frac{\eta}{2}\right)\mathcal{G}(F; v, x \mid J^{\mathrm{LEM}}; \mathcal{F}^{\mathrm{LEM}}, \mathcal{V}^{\mathrm{LEM}}) \\ &= \left(1 \pm \frac{\eta}{2}\right)(1 \pm \eta^{\mathrm{LEM}})\mathcal{D}(F \mid J^{\mathrm{LEM}}; \mathcal{F}^{\mathrm{LEM}}, \mathcal{V}^{\mathrm{LEM}}) \\ &= (1 \pm \eta)\mathcal{D}(F \mid J; \mathcal{F}, \mathcal{V}).\end{aligned}$$

Now consider when $|I| \geq 2$ and $F^{\mathrm{LEM}}_b \neq \{x\}$. Noting that we have the necessary counting conditions by (GP3), we apply Lemma 4.27 with $\mathcal{G}$ as a J^{LEM}-partite k-complex with the vertex partition $\mathcal{V}^{\mathrm{LEM}}$, the collection $\mathcal{I} = \{\{j'\}\}_{j' \in J^{\mathrm{LEM}}}$, the R'_{LEM}-partite k-complex F, the weighted hypergraph $\mathcal{D}^{\mathrm{LEM}}$ on J^{LEM} and x and v as mentioned previously, to deduce that $G^{\mathcal{G}}_{\mathcal{I},F,x,v}$ is $(\frac{\eta}{2})$-regular. Note that each edge of $G^{\mathcal{G}}_{\mathcal{I},F,x,v}$ corresponds to a J^{LEM}-partite copy of F in J^{LEM}-partite $\mathcal{G}$ with x mapped to v. Now the J-partite copies of F in J-partite $\mathcal{G}$, where each $y \in V(F) \setminus \{x\}$ is mapped into F_x and x mapped to v, correspond to the edges of $G^{\mathcal{G}}_{\mathcal{I},F,x,v}$ where for each $j' \in J^{\mathrm{LEM}}$ the tuple $(y)_{y \in F^{\mathrm{LEM}}_{j'}}$ is mapped into $\prod_{y \in F^{\mathrm{LEM}}_{j'}} F_y$ and x is mapped to v. Hence, by the $(\frac{\eta}{2})$-regularity of $G^{\mathcal{G}}_{\mathcal{I},F,x,v}$ and (GP3) we obtain

$$\begin{aligned}\mathcal{G}(F; v, x \mid J; \mathcal{F}, \mathcal{V}) &= \left(1 \pm \frac{\eta}{2}\right)\mathcal{G}(F; v, x \mid J^{\mathrm{LEM}}; \mathcal{F}^{\mathrm{LEM}}, \mathcal{V}^{\mathrm{LEM}}) \\ &= \left(1 \pm \frac{\eta}{2}\right)(1 \pm \eta^{\mathrm{LEM}})\mathcal{D}(F \mid J^{\mathrm{LEM}}; \mathcal{F}^{\mathrm{LEM}}, \mathcal{V}^{\mathrm{LEM}}) \\ &= (1 \pm \eta)\mathcal{D}(F \mid J; \mathcal{F}, \mathcal{V}),\end{aligned}$$

thus establishing (G3). Hence, we have a $(c, \Delta, \Delta_R, \eta, \mu)$-good $\mathcal{G}$-partition. $\square$

4.5.1 Good Partial Partite Homomorphisms

In a good $\mathcal{G}$-partition we have a partition $V_j = V^{\mathrm{main}}_j \cup V^{\mathrm{q}}_j \cup V^{\mathrm{buf}}_j$ of each cluster. For each $x \in X_j$ define the following.

$$\begin{aligned}\mathcal{C}^{\mathrm{main}}_\phi(x) &:= \mathcal{C}_\phi(x) \cap (V^{\mathrm{main}}_x)_{j \to x}, & A^{\mathrm{main}}_\phi(x) &:= A_\phi(x) \cap (V^{\mathrm{main}}_x)_{j \to x}, \\ \mathcal{C}^{\mathrm{q}}_\phi(x) &:= \mathcal{C}_\phi(x) \cap (V^{\mathrm{q}}_x)_{j \to x}, & A^{\mathrm{q}}_\phi(x) &:= A_\phi(x) \cap (V^{\mathrm{q}}_x)_{j \to x}, \\ \mathcal{C}^{\mathrm{buf}}_\phi(x) &:= \mathcal{C}_\phi(x) \cap (V^{\mathrm{buf}}_x)_{j \to x}, & A^{\mathrm{buf}}_\phi(x) &:= A_\phi(x) \cap (V^{\mathrm{buf}}_x)_{j \to x}.\end{aligned}$$

We define what it means for a partial partite homomorphism to be *good*. This aims to preserve desired local embedding properties that enable us to continue our embedding procedure for one more step; in particular, it ensures that we always have enough candidates in the queue and buffer reservoirs. For a partial partite homomorphism ϕ from a complex H to a weighted hypergraph $\mathcal{H}$ and a vertex $x \in V(H) \setminus \mathrm{Dom}(\phi)$, define $\pi_\phi(x) := |N_{H^{(2)}}(x) \cap \mathrm{Dom}(\phi)|$. Given a complex H, a linear order τ on $V(H)$ and $x \in V(H)$, let $\pi^\tau(x) := |\{y \in N_{H^{(2)}}(x) : \tau(y) < \tau(x)|\}$.

Definition 4.33 (Good partial partite homomorphism)**.** We call ϕ a *good partial partite homomorphism* if the following conditions hold.

(GPH1) For each $x \in \mathrm{Dom}(\phi)$ we have $\phi(x) \in V_x$.

(GPH2) For each $x \notin \mathrm{Dom}(\phi)$ we have

$$|\mathcal{C}^{\mathrm{main}}_\phi(x)| \geq (1-2\varepsilon)^{\pi_\phi(x)}(1-2\mu)|\mathcal{C}_\phi(x)|, \text{ and}$$
$$|\mathcal{C}^{\mathrm{q}}_\phi(x)|, |\mathcal{C}^{\mathrm{buf}}_\phi(x)| \geq (1-2\varepsilon)^{\pi_\phi(x)}\mu|\mathcal{C}_\phi(x)|.$$

4.6 The Setup and Pseudorandomness

In this section we discuss the extraction of useful THC information from our counting conditions on $\mathcal{G}$. Our embedding strategy involves the vertex-by-vertex construction of an embedding of H into $\mathcal{G}$ via a reduction to the standard construction setting and maintaining certain useful properties during the process; in particular, we maintain THC throughout the procedure. However, the straightforward choice of THC for H provides THC information strictly in accordance with the vertex order and only in *real-time*, while our queue and buffer treatments require THC information for a variety of neighbourhood-type restrictions on candidate sets given *in advance*. To circumvent this difficulty, we will construct auxiliary complexes (which can be seen as technical extensions of H) and use them to extract further THC information from the counting conditions.

4.6.1 Auxiliary Complex Constructions

Here we discuss the constructions of auxiliary complexes $\overline{H}$, H_x and H_+ in Lemmas 4.34, 4.35 and 4.36 which enable us to extract useful THC information from our counting conditions on $\mathcal{G}$. Let J be a finite set. Let H be a J-partite complex with $V(H)$ partitioned into $\mathcal{X} = \{X_j\}_{j\in J}$ and a linear order τ on $V(H)$. For $j \in J$ set

$$\mathcal{I}_j := \{\iota(E(H^{\leq 2}(x))) : x \in X_j\}$$

and for $x \in X_j$ write $\mathcal{I}_x$ to mean $\mathcal{I}_j$. Note that the elements of $\mathcal{I}_j$ are in fact sets (not just multisets) when (H2) holds for all $j \in J$. The following lemma constructs an auxiliary complex $\overline{H}$ which enables us to extract useful THC information about the induced subhypergraph corresponding to $H^{\leq 2}(y)$, with y restricted to a potential candidate set of another vertex x, from our counting conditions on $\mathcal{G}$; this gives us the neighbourhood THC properties needed to show that our random greedy algorithm completes successfully.

Lemma 4.34. *Let $k, \Delta, \Delta_R \in \mathbb{N}$. Let R be a k-complex on a finite set J with $\Delta(R^{(2)}) \leq \Delta_R$. Let H be a J-partite k-complex with $\Delta(H^{(2)}) \leq \Delta$ and its vertex set partitioned into $\mathcal{X} = \{X_j\}_{j\in J}$ such that $(H, \mathcal{X})$ is an R-partition and (H2) holds for all $j \in J$. Let τ be a linear order on $V(H)$. Then there exists a J-partite k-complex $\overline{H}$ with a partition $\overline{\mathcal{X}} = \{\overline{X}_j\}_{j\in J}$ of its vertex set and a linear order $\overline{\tau}$ on $V(\overline{H})$ such that the following hold.*

(AQ1) $X_j \subseteq \overline{X}_j$ *for all* $j \in J$, $\overline{H}[V(H)] = H$ *and* $(\overline{H}, \overline{\mathcal{X}})$ *is an R-partition.*

(AQ2) $V(H)$ *ordered according to* τ *forms an initial segment of* $\overline{\tau}$.

(AQ3) $\Delta\left(\overline{H}^{(2)}\right) \leq \left(2^{2^{\Delta_R^2+1}+\Delta^2} + 1\right)\Delta$.

(AQ4) *For each $j \in J$, $x \in X_j$ and $\mathcal{I} \in \mathcal{I}_j$, there are distinct vertices $y_1, \dots, y_{2^{\Delta^2}} \in \overline{X}_j$, $z_1, \dots, z_{|\bigcup \mathcal{I}|-1} \in V(\overline{H})$ such that the following hold. Set $Y = \{y_1, \dots, y_{2^{\Delta^2}}\}$ and $Z := \{z_1, \dots, z_{|\bigcup \mathcal{I}|-1}\}$.*

- *(i)* $(Y \cup Z) \cap V(H) = \varnothing$ *and* $N_{\overline{H}^{(2)}}(z) \cap V(H) = \varnothing$ *for all* $z \in Z$.
- *(ii)* $\iota(E(H[\{y\} \cup Z])) = \mathcal{I}$, $\iota(E(H^{\leq 2}(x))) = \iota(E(H[N^{<2}(x) \cup \{y\}]))$ *and* $N_{\overline{H}^{(2)}}(y) \cap V(H) = N^{-1}(x)$ *for all* $y \in Y$.

Proof. We shall explicitly construct $\overline{H}$ and $\overline{\tau}$. For each $(x, \mathcal{I}) \in X_j \times \mathcal{I}_j$, let $N_{x,\mathcal{I}}$ be the complex with vertex set $\{y_{x,\mathcal{I},i} : i \in [2^{\Delta^2}]\} \cup (\bigcup \mathcal{I} \setminus \{j\})$ and edge set $\{I \in \mathcal{I} : j \notin I\} \cup \{\{y_{x,\mathcal{I},i}\} \cup I \setminus \{j\} : I \in \mathcal{I}, j \in I, i \in [2^{\Delta^2}]\}$. We write $i_{x,\mathcal{I}}$ to refer to the copy of i

in $V(N_{x,\mathcal{I}})$. Start with the disjoint union

$$H \sqcup \left(\bigsqcup_{j \in J} \bigsqcup_{x \in X_j} \bigsqcup_{\mathcal{I} \in \mathcal{I}_j} N_{x,\mathcal{I}} \right).$$

For all $i \in [2^{\Delta^2}], j \in J, x \in X_j, \mathcal{I} \in \mathcal{I}_j$ and for all edges $e \in E(H)$ containing x, add $(e \setminus \{x\}) \cup \{y_{x,\mathcal{I},i}\}$ as an edge. Let $\overline{H}$ be the resultant k-complex. Pick an arbitrary linear order $\overline{\tau}$ on $V(\overline{H})$ with the elements of $V(H)$ ordered according to τ as an initial segment. For $j \in J$ set $\overline{X}_j := X_j \cup \{j_{x,\mathcal{I}} : x \in V(H), \mathcal{I} \in \mathcal{I}_x\} \cup \{y_{x,\mathcal{I},i} : x \in X_j, \mathcal{I} \in \mathcal{I}_j, i \in [2^{\Delta^2}]\}$. Set $\overline{\mathcal{X}} := \{\overline{X}_j\}_{j \in J}$.

We claim that $\overline{H}$ and $\overline{\tau}$ satisfy (AQ1)–(AQ4). Indeed, (AQ1) and (AQ2) hold by definition. Furthermore, the vertices of $N_{x,\mathcal{I}}$ fulfil the conditions of (AQ4) for each $j \in J$, $x \in X_j$ and $\mathcal{I} \in \mathcal{I}_j$. It remains to verify (AQ3). Let $j \in J$ and $x \in X_j$. Since $(H, \mathcal{X})$ is an R-partition, $\iota(E(H^{\leq 2}(x)))$ is a subset of the power set of $R^{\leq 2}(j)$. Now $|N^{\leq 2}(j)| \leq \Delta_R^2 + 1$, so we have $|\mathcal{I}_j| \leq 2^{2^{\Delta_R^2+1}}$. Note that every vertex in $V(\overline{H}) \setminus V(H)$ has degree at most 2^{Δ^2} in $H^{(2)}$. Let $y \in V(H)$ and consider $z \in N_{\overline{H}^{(2)}}(y)$. There are two possibilities: either $z \in N_{H^{(2)}}(y)$, or $z = y_{x,\mathcal{I},i}$ for some $x \in N^{>}(y), \mathcal{I} \in \mathcal{I}_x, i \in [2^{\Delta^2}]$. Hence, y has degree at most $\Delta + 2^{\Delta^2} \sum_{x \in N^{>}(y)} |\mathcal{I}_x| \leq (2^{2^{\Delta_R^2+1}+\Delta^2} + 1)\Delta$ in $\overline{H}^{(2)}$. Therefore, we have $\Delta\left(\overline{H}^{(2)}\right) \leq (2^{2^{\Delta_R^2+1}+\Delta^2} + 1)\Delta$. $\square$

The following lemma constructs an auxiliary complex H_+, which enables us to extract useful THC information from our counting conditions; this turns out to be a technical extension of THC for $\overline{H}$ which gives us the localised algorithmic THC properties needed to analyse the embedding of buffer vertices.

Lemma 4.35. *Let $k, \Delta, \Delta_R \in \mathbb{N}$. Let R be a k-complex on a finite set J with $\Delta(R^{(2)}) \leq \Delta_R$. Let H be a J-partite k-complex with $\Delta(H^{(2)}) \leq \Delta$ and its vertex set partitioned into $\mathcal{X} = \{X_j\}_{j \in J}$ such that $(H, \mathcal{X})$ is an R-partition and (H2) holds for all $j \in J$. For each $j \in J$ let $X_j^{\mathrm{buf}} \subseteq X_j$. Let τ be a good vertex order for X^{buf} on $V(H)$. Let $\overline{H}$ be a J-partite k-complex with a partition $\overline{\mathcal{X}} = \{\overline{X}_j\}_{j \in J}$ of its vertex set and a linear order $\overline{\tau}$ on $V(\overline{H})$ satisfying (AQ1)–(AQ4). Let $s := |N_{H^{(2)}}(X^{\mathrm{buf}})|$ and $h := |V(\overline{H})|$. Then there exists a J-partite k-complex H_+ with a partition $\mathcal{X}^+ = \{X_j^+\}_{j \in J}$ of its vertex set and a linear order τ_+ on $V(H_+)$ such that the following hold.*

(AM1) $\Delta\left(H_+^{(2)}\right) \leq \left(2^{2^{\Delta_R^2+1}+\Delta^2} + 1\right)(\Delta + 1)\Delta.$

(AM2) $(H_+, \mathcal{X}^+)$ *is an R-partition.*

(AM3) There are distinct vertices $z_1, \ldots, z_s$ and for each $i \in [\Delta]_0$ there are distinct vertices $w^{(i)}_{s+1}, \ldots, w^{(i)}_h$ such that the following hold. Let $Z = \{z_i : i \in [s]\}$ and $W_i = \{w^{(i)}_{s+1}, \ldots, w^{(i)}_h\}$ for each $i \in [\Delta]_0$.

(i) $z_1, \ldots, z_s$ is an initial segment of τ_+ in that order.

(ii) $H_+[Z \cup W_i]$ with the order induced by τ_+ is a copy of $(\overline{H}, \overline{\tau})$ for each $i \in [\Delta]_0$.

(iii) For each $i \in [\Delta]_0$ and each $w \in W_i$ we have $N_{H^{(2)}_+}(w) \subseteq Z \cup W_i$.

(iv) $V(H_+) = \bigcup_{i \in [\Delta]_0} Z \cup W_i$.

Proof. To construct H_+ and τ_+, we start with the disjoint union $\bigsqcup_{i=0}^{\Delta} \overline{H}_{(i)}$ of $\Delta+1$ copies of $\overline{H}$. For each $x \in V(\overline{H})$ and each $i \in [\Delta]_0$ write $x_{(i)}$ for the copy of x in $\overline{H}_{(i)}$. For each $t \in [s]$ in succession, with $z \in V(\overline{H})$ satisfying $\overline{\tau}(z) = t$, we identify the copies $z_{(i)}$ for $i \in [\Delta]_0$, that is, we delete those copies of z and add a vertex $\overline{z}$ which has their combined adjacencies. Let H_+ be the resultant k-complex and let $\mathcal{X}^+$ be the natural partition obtained by combining the partitions of the copies of $\overline{H}$ and suitably identifying vertices. Let τ_+ be the linear order on $V(H_+)$ where the identified vertices come first in the natural order induced by $\overline{\tau}$, followed by the remaining vertices of $\overline{H}_{(i)}$ in the natural order induced by $\overline{\tau}$ for each i in turn.

Each vertex in H_+ is the result of the identification of at most $\Delta+1$ vertices each with degree at most $\left(2^{2^{\Delta^2_R+1}+\Delta^2}+1\right)\Delta$, so (AM1) follows. The complex H_+ obtained from a disjoint union of copies of $\overline{H}$ by suitably identifying vertices, so (AM2) follows. We now verify (AM3). The identified vertices serve as $z_1, \ldots, z_s$ and for each $i \in [\Delta]_0$ the remaining vertices of $\overline{H}_{(i)}$ serve as the vertices in W_i. By construction, we obtain the properties (i)–(iv). $\square$

The following lemma constructs a useful auxiliary complex H_x which enables us to extract useful THC information from our rooted vertex counting conditions on $\mathcal{G}$.

Lemma 4.36. *Let $k, \Delta \in \mathbb{N}$. Let R be a k-complex on a finite set J. Let H be a J-partite k-complex with $\Delta(H^{(2)}) \leq \Delta$ and its vertex set partitioned into $\mathcal{X} = \{X_j\}_{j \in J}$ such that $(H, \mathcal{X})$ is an R-partition. For each $j \in J$ let $X^{\mathrm{buf}}_j \subseteq X_j$. Let τ be a good vertex order for X^{buf} on $V(H)$. Let $j \in J$ and $x \in X^{\mathrm{buf}}_j$. Let $s := \tau(\min(N_{H^{(2)}}(x))) - 1$, $b := |N_{H^{(2)}}(x)|$ and $h := |V(H)|$. Then there exists a J-partite k-complex H_x with a* *partition $\mathcal{X}^x = \{X^x_j\}_{j\in J}$ of its vertex set and a linear order τ_x on $V(H_x)$ such that the following hold.*

(AB1) $\Delta\left(H_x^{(2)}\right) \le (\Delta+1)\Delta$.

(AB2) $(H_x, \mathcal{X}^x)$ *is an R-partition.*

(AB3) $z_1, \dots, z_{s+b}$ *and* $w^{(i)}_{s+\min(i,b)+1}, \dots, w^{(i)}_h$ *for each* $i \in [\Delta]_0$ *are distinct vertices such that the following hold. Set* $Z_i = \{z_1, \dots, z_{s+i}\}$ *for each* $i \in [b]_0$ *and* $W_i = \{w^{(i)}_{s+\min(i,b)+1}, \dots, w^{(i)}_h\}$ *for each* $i \in [\Delta]_0$.

(i) $z_1, \dots, z_{s+b}$ *is an initial segment of* τ_x *in that order.*

(ii) $H_x[Z_{\min(i,b)} \cup W_i]$ *with the order induced by* τ_x *is a copy of* (H, τ) *for each* $i \in [\Delta]_0$.

(iii) $N_{H_x^{(2)}}(w) \subseteq Z_{\min(i,b)} \cup W_i$ *for each* $i \in [\Delta]_0$ *and each* $w \in W_i$.

(iv) $V(H_x) = \bigcup_{i\in[\Delta]_0} Z_{\min(i,b)} \cup W_i$.

Proof. To construct H_x and τ_x, we start with the disjoint union $\bigsqcup_{i=0}^{\Delta} H_{(i)}$ of $\Delta+1$ copies of H. For each $x \in V(H)$ and each $i \in [\Delta]_0$ write $x_{(i)}$ for the copy of x in $H_{(i)}$. For each $t \in [s]$ in succession, with $z \in V(H)$ satisfying $\tau(z) = t$, we identify the copies $z_{(i)}$ for $i \in [\Delta]_0$, that is, we delete those copies of z and add a vertex $\overline{z}$ which has their combined adjacencies. For each $\ell \in [b]$ in succession, with $z \in V(H)$ satisfying $\tau(z) = s+\ell$, we identify the copies $z_{(i)}$ for $i \in [\Delta] \setminus [\ell-1]$. Let H_x be the resultant k-complex and let $\mathcal{X}^x$ be the natural partition obtained by combining the partitions of the copies of H and suitably identifying vertices. Let τ_x be the linear order on $V(H_x)$ where the identified vertices come first in the natural order induced by τ, followed by the remaining vertices of $H_{(i)}$ in the natural order induced by τ for each i in turn.

Each vertex in H_x is the product of the identification of at most $\Delta+1$ vertices each with degree at most Δ, so (AB1) follows. The complex H_x obtained from a disjoint union of copies of H by suitably identifying vertices, so (AB2) follows. We now verify (AB3). The identified vertices serve as $z_1, \dots, z_{s+b}$ and for each $i \in [\Delta]_0$ the remaining vertices of $H_{(i)}$ serve as the vertices in W_i. By construction, we obtain the properties (i)–(iv). $\square$

4.6.2 The Setup

Here we shall collect the various conditions on the complexes H and $\mathcal{G}$ which we assume in our blow-up lemma and the extra conditions which we derive through good partitions. We will also describe the various constants which appear in our proof. Firstly, the following constants are chosen by the user.

k is the maximum size of an edge in H and $\mathcal{G}$.

Δ is the maximum degree of H.

The user also chooses Δ_R^{LEM}, α^{LEM}, κ^{LEM} and J^{LEM}; however, these are altered by our preprocessing through Lemma 4.32 to give the following.

Δ_R is the maximum degree of the reduced complex R, which captures the structure of H at the cluster level.

α is the required fraction of buffer vertices.

κ is the cluster size balancing factor.

J is the indexing set for the clusters.

Our blow-up lemma guarantee that the following constants exist.

c is the maximum size of the complexes whose counts we are required to control precisely.

η^{LEM} is the required precision for the counts we are required to control.

n_0 is the minimum size of the complexes H and $\mathcal{G}$ for which the blow-up lemma is valid.

Furthermore, we have the following auxiliary constants which play important roles in our proof.

μ is the fraction of each cluster of $\mathcal{G}$ contained in each small reservoir set of the subpartitions of $\mathcal{V}$.

ρ is both the fraction of vertices in each cluster which may enter the queue and the fraction of vertices in each cluster which may become exceptional.

η' is the initial precision of counts after preprocessing.

η is the precision of counts required for THC.

ε is regularity for pairs in THC graphs with precision η.

In order for our proofs to work we will require the constants to be in the following size order.

$$\begin{aligned} &0 < \eta' \ll \eta \ll \varepsilon \ll \rho \ll \mu \ll c \ll \alpha, \Delta^{-1}, \Delta_R^{-1}, \kappa^{-1} \text{ and} \\ &0 < n_0^{-1} \ll |J|^{-1}, \varepsilon, \end{aligned} \tag{4.4}$$

where by $x \ll y$ we mean that there is a non-decreasing function $f : (0,1] \to (0,1]$ such that our proof works if $0 < x \leq f(y)$.

We shall now define the setup, which encapsulates the choices of constants above and the enhanced setting in which we work. This encapsulation is convenient because we will use this setup in many of the lemmas to come.

Setup 4.37 (Setup)**.** *When we say that we* assume Setup 4.37 *we mean that we have made appropriate choices of constants in accordance with* (4.4) *and that we have k-complexes $R' \subseteq R$ on a finite set J with $\Delta(R^{(2)}) \leq \Delta_R$, J-partite k-complexes H and $\mathcal{G}$ on $n \geq n_0$ vertices with κ-balanced size-compatible vertex partitions $\mathcal{X} = \{X_j\}_{j \in J}$ and $\mathcal{V} = \{V_j\}_{j \in J}$ respectively whose parts are of size at least $\frac{n}{\kappa |J|}$ and where $\Delta(H^{(2)}) \leq \Delta$, $\varnothing \in E(\mathcal{G})$ and $\{v\} \in E(\mathcal{G})$ for all $v \in V(\mathcal{G})$, a weighted hypergraph $\mathcal{D}$ on J with $d(\varnothing) = 1$, $d(\{j\}) = 1$ for all $j \in J$ and $d(e) > 0$ for all $e \in E(R)$, a family $\widetilde{\mathcal{X}} = \{\widetilde{X}_j\}_{j \in J}$ of potential buffer vertices, subsets $X_j^{\mathrm{buf}} \subseteq \widetilde{X}_j$ for each $j \in J$, a good vertex order τ for X^{buf} on $V(H)$, a partition $\{X_{j,h}\}_{h \in [\ell_j]}$ of X_j for each $j \in J$, and a partition $V_j = V_j^{\mathrm{main}} \cup V_j^{\mathrm{q}} \cup V_j^{\mathrm{buf}}$ for each $j \in J$, which give an $(\alpha, c, \Delta, \Delta_R, \kappa, \mu)$-good H-partition and a $(c, \Delta, \Delta_R, \eta', \mu)$-good $\mathcal{G}$-partition.*

The following are also given. Let $\mathcal{J} := \{\{j\}\}_{j \in J}$. Let the standard constructions of $(\mathcal{G}, \mathcal{V})$ and $(\mathcal{D}, \mathcal{J})$ with respect to $(H, \mathcal{X})$ be $\mathcal{G}'$ and $\mathcal{D}'$ respectively. For each $j \in J$ write U_j for the unique set which is $V_{N^{-1}(x)}$ for any $x \in X_j^{\mathrm{buf}}$, F_j for the unique ordered partite complex, up to partite isomorphism, which is $H^{\leq 1}(x)$ for any $x \in X_j^{\mathrm{buf}}$, and a_j and b_j for the unique values of $\mathcal{D}(H^{-1}(x))$ and $\frac{\mathcal{D}(H^{\leq 1}(x))}{\mathcal{D}(H^{-1}(x))}$ respectively for any $x \in X_j^{\mathrm{buf}}$. For each $j \in J$, $h \in [\ell_j]$ and $p \in [3]$, write $n_{j,h}^{(p)}$ for the unique value of $|N^{<p}(y)|$ for $y \in X_{j,h}$.

4.6.3 Pseudorandomness from Counts

Here we establish several lemmas related to THC pseudorandomness. Firstly, we show that $\mathcal{G}$ is an (η, c)-THC graph for (H_+, τ_+), $(\overline{H}, \overline{\tau})$ and (H, τ); this is a key property for our random greedy algorithm to succeed and follows readily from Theorem 4.10 and the bounded degree of these complexes. Secondly, we establish that the weighted induced subhypergraph corresponding to the complex $H^{\leq 2}(y)$, with y restricted to a potential candidate set of a different vertex x, is an (η, c)-THC graph. This will be useful in showing that our random greedy algorithm completes successfully; we will interested in ensuring that potential candidate sets do not become overused and this will give us precise control over the conditional probability of embedding y into a potential candidate set of x for as long as we are able maintain THC. Proving this is mostly technical manipulation: $\overline{H}$ is constructed with the properties required for this to work.

Finally, we show that our host graph remains THC when we impose a neighbourhood restriction to reflect an intention to embed a buffer vertex $x \in X_j^{\mathrm{buf}}$ to a vertex $v \in V_j$. Our proof strategy requires us to understand how likely it is for a vertex $v \in V_j$ to be a candidate for a buffer vertex $x \in X_j^{\mathrm{buf}}$. As such, we will also need to understand how our weighted hypergraph behaves and evolves, under the assumption that we intend to embed x to v at the appropriate time; in particular, we want to show that the relevant weighted hypergraph is sufficiently well-behaved and THC-good. We will provide the technical constructions which allow us to establish this.

Let J be a finite set. Let H be a J-partite complex with a partition $\mathcal{X} = \{X_j\}_{j \in J}$ of $V(H)$ and a linear order τ on $V(H)$ such that (H2) holds for all $j \in J$. Let $\mathcal{G}$ be a J-partite weighted hypergraph with a partition $\mathcal{V} = \{V_j\}_{j \in J}$ of $V(\mathcal{G})$ and $\mathcal{D}$ be a weighted hypergraph on J. Let $\mathcal{J} := \{\{j\}\}_{j \in J}$. Let $\mathcal{G}'$ and $\mathcal{D}'$ be the standard constructions of $(\mathcal{G}, \mathcal{V})$ and $(\mathcal{D}, \mathcal{J})$ with respect to $(H, \mathcal{X})$. Let $j \in J$ and $x, y \in X_j$. Let ϕ be a partial partite homomorphism of $H^{-1}(x)$ into $\mathcal{G}'$ and set $\vec{v} = (\phi(z))_{z \in \mathrm{Dom}(\phi)}$. Define $\mathcal{G}^{x,\vec{v},y}$ and $\mathcal{D}^{x,\vec{v},y}$ to be the $N^{\leq 2}(y)$-partite weighted-k-graphs with vertex sets $\{V'_z\}_{z \in N^{\leq 2}(y)}$ and $\{\{z\}\}_{z \in N^{\leq 2}(y)}$ respectively and weight functions $g^{x,\vec{v},y}$ and $d^{x,\vec{v},y}$ respectively given as follows. For each set $f \subseteq N^{\leq 2}(y)$ and each edge $e \in V'_f$, we define

$$g^{x,\vec{v},y}(e) := \begin{cases} 0 & \text{if } f = \{y\} \text{ and } e \notin \mathcal{C}_{\vec{v}}(x), \\ g'(e) & \text{otherwise.} \end{cases}$$

For each set $e \subseteq N^{\leq 2}(y)$ we define

$$d^{x,\vec{v},y}(e) := \begin{cases} d'(y)\frac{\mathcal{D}(H[\mathrm{Dom}(\phi)\cup\{x\}])}{\mathcal{D}(H[\mathrm{Dom}(\phi)])} & \text{if } e = \{y\}, \\ d'(e) & \text{otherwise.} \end{cases}$$

Let $j \in J$, $v \in V_j$, $x \in X_j^{\mathrm{buf}}$ and $A \subseteq N_{H^{(2)}}(x)$. Define $\mathcal{G}^{v,x,A}$ and $\mathcal{D}^{v,x,A}$ to be the $V(H)$-partite weighted-k-graphs on $V(\mathcal{G}')$ and $V(\mathcal{D}')$ respectively with weight functions $g^{v,x,A}$ and $d^{v,x,A}$ respectively given as follows. For each set $f \subseteq V(H)$ and each edge $e \in V'_f$, writing v' for the copy of v in V'_x, we define

$$g^{v,x,A}(e) := \begin{cases} g'(e)g'(e \cup \{v'\}) & \text{if } f \subseteq A, \\ g'(e) & \text{otherwise.} \end{cases}$$

For each set $e \subseteq V(H)$ we define

$$d^{v,x,A}(e) := \begin{cases} d'(e)d'(e \cup \{x\}) & \text{if } e \subseteq A, \\ d'(e) & \text{otherwise.} \end{cases}$$

Let H_x be a J-partite k-complex with a partition $\mathcal{X}^x$ of its vertex set and a linear order τ_x on $V(H_x)$ satisfying (AB1)–(AB3). Let $\mathcal{G}^{x,*}$ and $\mathcal{D}^{x,*}$ be the standard constructions of $(\mathcal{G}, \mathcal{V})$ and $(\mathcal{D}, \mathcal{J})$ with respect to $(H_x, \mathcal{X}^x)$. Let x' denote the copy of x in W_b (from Lemma 4.36). Let $\mathcal{G}^{v,x,*}$ and $\mathcal{D}^{v,x,*}$ be the $V(H_x)$-partite weighted-k-graphs on $V(\mathcal{G}^{x,*})$ and $V(\mathcal{D}^{x,*})$ respectively with weight functions $g^{v,x,*}$ and $d^{v,x,*}$ respectively given as follows. For each set $f \subseteq V(H_x)$ and each edge $e \in V'_f$, writing v' for the copy of v in $V'_{x'}$, we define

$$g^{v,x,*}(e) := \begin{cases} g^{x,*}(e)g^{x,*}(e \cup \{v'\}) & \text{if } f \subseteq \{y_{s+1}, \dots, y_{s+b}\}, \\ g^{x,*}(e) & \text{otherwise.} \end{cases}$$

For each set $e \subseteq V(H_x)$ we define

$$d^{v,x,*}(e) := \begin{cases} d^{x,*}(e)d^{x,*}(e \cup \{x'\}) & \text{if } e \subseteq \{y_{s+1}, \dots, y_{s+b}\}, \\ d^{x,*}(e) & \text{otherwise.} \end{cases}$$

The following lemma is an amalgamation of the lemmas mentioned previously, which we unify into one to align the relevant constants. The proof of this lemma involves suitable applications of Theorem 4.10.

Lemma 4.38. *Let $k, \Delta \geq 2$, $\Delta_R \in \mathbb{N}$, $\Delta_{\mathrm{aux}} := \left(2^{2^{\Delta_R^2+1}+\Delta^2} + 1\right)(\Delta+1)\Delta$, $c \geq \Delta_{\mathrm{aux}}+2$ and $\eta \in (0, 1/2)$. Then there exists $\eta_0 > 0$ such that whenever $0 < \eta' < \eta_0$ the following holds. Let R be a k-complex on a finite set J with $\Delta(R^{(2)}) \leq \Delta_R$ and R' be a spanning subcomplex of R. Let H be a J-partite k-complex with $\Delta(H^{(2)}) \leq \Delta$ and a partition $\mathcal{X} = \{X_j\}_{j\in J}$ of $V(H)$ such that $(H, \mathcal{X})$ is an R-partition and (H2) holds for all $j \in J$. Let $\widetilde{\mathcal{X}} = \{\widetilde{X}_j\}_{j\in J}$ be an $(\alpha, (\Delta^2 + \Delta + 2)c, R')$-buffer for $(H, \mathcal{X})$. For $j \in J$ let X_j^{buf} be a subset of $\widetilde{X}_j$ such that (H3) holds. Let τ be a good vertex order for X^{buf} on $V(H)$. Let $\overline{H}$ be a J-partite k-complex with a partition $\overline{\mathcal{X}}$ of $V(\overline{H})$ and a linear order $\overline{\tau}$ on $V(\overline{H})$ satisfying (AQ1)–(AQ4). Let H_+ be a J-partite k-complex with a partition $\mathcal{X}^+$ of $V(H_+)$ and a linear order τ_+ on $V(H_+)$ satisfying (AM1)–(AM3). Let $\mathcal{G}$ be a J-partite weighted hypergraph with its vertex set partitioned into $\mathcal{V} = \{V_j\}_{j\in J}$. Let $\mathcal{D}$ be a weighted hypergraph on J. Suppose that for all R-partite k-complexes F on at most $(\Delta_{\mathrm{aux}} + 2)(\Delta + 2)c$ vertices we have*

$$\mathcal{G}(F) = (1 \pm \eta')\frac{g(\varnothing)}{d(\varnothing)}\mathcal{D}(F) \tag{4.5}$$

and for each R'-partite k-complex F on at most $(\Delta^2 + \Delta + 2)c + 1$ vertices with its vertex set partitioned into $\mathcal{F} = \{F_j\}_{j\in J}$, each $x \in V(F)$ and each $v \in V_j$ with $j \in J$ such that $x \in F_j$, we have

$$\mathcal{G}(F; v, x) = (1 \pm \eta')\frac{g(\varnothing)}{d(\varnothing)}\mathcal{D}(F). \tag{4.6}$$

Let $\mathcal{G}^+$, $\overline{\mathcal{G}}$ and $\mathcal{G}'$ be the standard constructions of $(\mathcal{G}, \mathcal{V})$ with respect to $(H_+, \mathcal{X}^+)$, $(\overline{H}, \overline{\mathcal{X}})$ and $(H, \mathcal{X})$ respectively. Let $\mathcal{J} := \{\{j\}\}_{j\in J}$. Let $\mathcal{D}^+$ be the standard construction $(\mathcal{D}, \mathcal{J})$ with respect to $(H_+, \mathcal{X}^+)$. The following statements hold.

- *(i) $\mathcal{G}$ is an (η, c)-*THC *graph for (H_+, τ_+), $(\overline{H}, \overline{\tau})$ and (H, τ) with density weighted hypergraph $\mathcal{D}$.*
- *(ii) For every $j \in J$, $x, y \in X_j$ and tuple $\vec{v} = (\phi(z))_{z\in \mathrm{Dom}(\phi)\cap N^{-1}(x)}$ of vertices in $\mathcal{G}'$ for some partial partite homomorphism ϕ from H to $\mathcal{G}'$ which is* THC*-respecting for $(\overline{\mathcal{G}}, \overline{H}, \overline{\tau})$, the weighted-$k$-graph $\mathcal{G}^{x,\vec{v},y}$ is an (η, c)-*THC *graph with the linear order on $N^{\leq 2}(y)$ induced by τ and density weighted hypergraph $\mathcal{D}^{x,\vec{v},y}$.*
- *(iii) Given $j \in J$, $v \in V_j$ and $x \in X_j^{\mathrm{buf}}$, for a J-partite k-complex H_x with a partition $\mathcal{X}^x$ of $V(H_x)$ and a linear order τ_x on $V(H_x)$ satisfying (AB1)–(AB3), we have*

that $\mathcal{G}^{v,x,*}$ is an (η, c)-THC graph with the linear order τ_x and density weighted hypergraph $\mathcal{D}^{v,x,*}$.

(iv) Given $j \in J$, $v \in V_j$ and $x \in X_j^{\mathrm{buf}}$, letting $y_1, \ldots, y_b$ be the elements of $N_{H^{(2)}}(x)$ in the order according to τ and $Y_i = \{y_h : h \in [i]\}$ for each $i \in [b]_0$, we have that $\mathcal{G}^{v,x,Y_i}$ is an (η, c)-THC graph with the linear order τ and density weighted hypergraph $\mathcal{D}^{v,x,Y_i}$ for each $i \in [b]_0$.

Proof. Given $k, \Delta \geq 2, c \geq \Delta + 2$ and $0 < \eta < 1/2$, Theorem 4.10 returns a constant $\eta_1 > 0$. Set $\eta^* := \min\left(\eta, \frac{\eta_1}{2(\Delta^2+\Delta+2)c+1}\right)$. Now given $k, \Delta_{\mathrm{aux}} \geq 2, (\Delta+2)c \geq \Delta_{\mathrm{aux}} + 2$ and $0 < \eta^* < 1/2$, Theorem 4.10 returns a constant $\eta_2 > 0$. Set $\eta_0 := \min(\eta^*, \eta_2)$ and let $0 < \eta' < \eta_0$. Let $\mathcal{D}'$ and $\overline{\mathcal{D}}$ be the standard constructions of $(\mathcal{D}, \mathcal{J})$ with respect to $(H, \mathcal{X})$ and $(\overline{H}, \overline{\mathcal{X}})$ respectively.

Let us first show (i). We start with H_+. Note that $\mathcal{G}^+$ is a $V(H_+)$-partite weighted-k-graph with vertex sets $\{V'_x\}_{x \in V(H_+)}$ which is identically 1 on any V'_f such that $f \notin E(H_+)$ and by Lemma 4.21 applied to (4.5) we have

$$\mathcal{G}^+(F) = (1 \pm \eta') \frac{g^+(\varnothing)}{d^+(\varnothing)} \mathcal{D}^+(F)$$

for all $V(H_+)$-partite k-complexes F on at most $(\Delta_{\mathrm{aux}} + 2)(\Delta + 2)c$ vertices. By (AM1) we have $\Delta\left(H_+^{(2)}\right) \leq \Delta_{\mathrm{aux}}$. Applying Theorem 4.10, we deduce that $\mathcal{G}^+$ is a $(\eta^*, (\Delta+2)c)$-THC graph with the linear order τ_+ and density weighted hypergraph $\mathcal{D}^+$. Therefore, by definition $\mathcal{G}$ is an (η, c)-THC graph for (H_+, τ_+) with density weighted hypergraph $\mathcal{D}$. Furthermore, observe that the arguments are entirely analogous for both $\overline{H}$ and H: these partite complexes satisfy the same maximum degree condition and the standard construction of $(\mathcal{G}, \mathcal{V})$ with respect to these partite complexes have the required form and satisfy the counting condition by Lemma 4.21.

Next we show (ii). Since $\overline{\mathcal{G}}$ is an $(\eta^*, (\Delta+2)c)$-THC graph, $\overline{\mathcal{G}}_\phi$ is an $(\eta^*, (\Delta+2)c)$-THC graph with the linear order induced by $\overline{\tau}$ and density weighted hypergraph $\overline{\mathcal{D}}_\phi$. Let $\mathcal{I} := \iota(E(H^{\leq 2}(y))) \in \mathcal{I}_j$. By the definition of $\overline{\mathcal{G}}$, there are distinct vertices $y' \in \overline{X}_j$ and $z_1, \ldots, z_{|N^{<2}(y)|} \in V(\overline{H})$ which satisfy (AQ4). Let $A := \{y', z_1, \ldots, z_{|N^{<2}(y)|}\}$. By (THC1) we have

$$\overline{\mathcal{G}}_\phi(F) = (1 \pm v(F)\eta^*)\overline{\mathcal{D}}_\phi(F)$$

for all A-partite k-complexes F on at most $(\Delta+2)c$ vertices. Hence, $\mathcal{H} := \overline{\mathcal{G}}_\phi[\bigcup_{z \in A} V'_z]$ is an (η, c)-THC graph by Theorem 4.10 applied with the complex $\overline{H}[A]$. By (AQ4) we have $\iota(E(H^{\leq 2}(y))) = \mathcal{I} = \iota(E(\overline{F}))$, $N_{\overline{H}^{(2)}}(z) \cap \mathrm{Dom}(\phi) = \varnothing$ for all $z \in A \setminus \{y'\}$ and $N_{\overline{H}^{(2)}}(y') \cap \mathrm{Dom}(\phi) = \mathrm{Dom}(\phi_t) \cap N^{-1}(x)$, so $\mathcal{H}$ is isomorphic to $\mathcal{G}^{x,\vec{v},y}$. Furthermore, $\overline{\mathcal{D}}_\phi[A]$ is isomorphic to $\mathcal{D}^{x,\vec{v},y}$. Therefore, $\mathcal{G}^{x,\vec{v},y}$ is an (η, c)-THC graph with density weighted hypergraph $\mathcal{D}^{x,\vec{v},y}$.

Now let us prove (iii). Let F be a $V(H_x)$-partite k-complex on at most $(\Delta^2+\Delta+2)c$ vertices with vertex partition $\mathcal{F} = \{F_y\}_{y \in V(H_x)}$. First consider the case where F_y is empty for every $y \in \{z_{s+1}, \dots, z_{s+b}\}$. In this case, by Lemma 4.21 and (4.5) we have

$$\mathcal{G}^{v,x,*}(F) = \mathcal{G}^{x,*}(F) = (1 \pm \eta') \frac{g^{x,*}(\varnothing)}{d^{x,*}(\varnothing)} \mathcal{D}^{x,*}(F) = (1 \pm \eta') \frac{g^{v,x,*}(\varnothing)}{d^{v,x,*}(\varnothing)} \mathcal{D}^{v,x,*}(F).$$

Otherwise, $\bigcup_{i \in [b]} F_{y_{s+i}}$ is nonempty. Set $F'_i = \bigcup_{y \in X_i^x} F_x$ for each $i \in J$ and set $\mathcal{F}' := \{F'_i\}_{i \in J}$. Let $\overline{F}$ be the complex with $V(\overline{F}) = V(F) \bigsqcup \{x'\}$ and $E(\overline{F}) = E(F) \cup \{e \cup \{x'\} : e \in F_f \cap E(F), f \cup \{x'\} \in E(H_x)\}$, writing x' for a copy of x. Set

$$\overline{F}_y := \begin{cases} F_y & \text{if } y \neq x', \\ F_{x'} \sqcup \{x'\} & \text{if } y = x', \end{cases} \quad \text{and } \overline{F}'_i := \begin{cases} F'_i & \text{if } i \neq j, \\ F'_j \sqcup \{x'\} & \text{if } i = j, \end{cases}$$

for $y \in V(H_x)$ and $i \in J$. Set $\overline{\mathcal{F}} := \{\overline{F}_y\}_{y \in V(H_x)}$ and $\overline{\mathcal{F}}' := \{\overline{F}'_i\}_{i \in J}$. Let $\widetilde{F}$ be the edge-maximal H_x-partite spanning subcomplex of $\overline{F}$.

Say $\widetilde{F}^{(2)}$ has components on vertex sets $X_1, \dots, X_\ell$ with $x' \in X_1$ and for each $i \in [\ell]$ let $\widetilde{F}_{(i)} := \widetilde{F}[X_i]$. Since $\widetilde{F}_{(1)}$ is contained within the $(\Delta^2+\Delta+2)c$-neighbourhood of x', x is in an $(\alpha, (\Delta^2+\Delta+2)c, R')$-buffer for $(H, \mathcal{X})$, and H_x is wholly derived from copies of $(H, \mathcal{X})$ by (AB3), we deduce that $\widetilde{F}_{(1)}$ is R'-partite. By (4.6) we have $\mathcal{G}(\widetilde{F}_{(1)}; v, x') = (1 \pm \eta') \frac{g(\varnothing)}{d(\varnothing)} \mathcal{D}(\widetilde{F}_{(1)})$. By (4.5) we have $\mathcal{G}(\widetilde{F}_{(i)}; v, x') = (1 \pm \eta') \frac{g(\varnothing)}{d(\varnothing)} \mathcal{D}(\widetilde{F}_{(i)})$ for each $i \in [\ell] \setminus \{1\}$. Now since $\widetilde{F}$ is a complex, all its edges must belong to some $F_{(i)}$. Hence, we obtain

$$\begin{aligned} \mathcal{G}(\widetilde{F}; v, x') &= g(\varnothing)^{1-\ell} \mathcal{G}(\widetilde{F}_{(1)}; v, x') \prod_{1 < h \leq \ell} \mathcal{G}(\widetilde{F}_{(i)}) \\ &= (1 \pm \eta')^\ell \frac{g(\varnothing)}{d(\varnothing)^\ell} \prod_{h \in [\ell]} \mathcal{D}(\widetilde{F}_{(i)}) = (1 \pm \eta')^\ell \frac{g(\varnothing)}{d(\varnothing)} \mathcal{D}(\widetilde{F}). \end{aligned}$$

By Lemma 4.20(ii), and the definitions of $\mathcal{G}^{v,x,*}$ and $\mathcal{D}^{v,x,*}$, we have $\mathcal{G}^{v,x,*}(F) = \mathcal{G}^{x,*}(\overline{F}; v_{j \to x'}, x') = \mathcal{G}(\widetilde{F}; v, x')$ and $\mathcal{D}^{v,x,*}(F) = \mathcal{D}^{x,*}(\overline{F}) = \mathcal{D}(\widetilde{F})$. Hence, we obtain $\mathcal{G}^{v,x,*}(F) = (1 \pm \eta')^\ell \frac{g^{v,x,*}(\varnothing)}{d^{v,x,*}(\varnothing)} \mathcal{D}^{v,x,*}(F)$. Therefore, we have

$$\mathcal{G}^{v,x,*}(F) = (1 \pm \eta')^\ell \frac{g^{v,x,*}(\varnothing)}{d^{v,x,*}(\varnothing)} \mathcal{D}^{v,x,*}(F)$$

for all $V(H_x)$-partite k-complexes F on at most $(\Delta^2+\Delta+2)c$ vertices and so by Theorem 4.10 $\mathcal{G}^{v,x,*}$ is an (η,c)-THC graph with the linear order τ_x and density weighted hypergraph $\mathcal{D}^{v,x,*}$.

Finally, let us prove (iv). Let $i \in [b]_0$ and let Z_i and W_i be as per (AB3). Since by (AB3)(ii) $H_x[Z_i \cup W_i]$ with the linear order induced by τ_x is a copy of (H,τ) and by construction $\mathcal{G}^{v,x,*}\left[\bigcup_{z\in Z_i\cup W_i} V'_z\right]$ is isomorphic to $\mathcal{G}^{v,x,Y_i}$, by our earlier deduction for (iii) we have

$$\mathcal{G}^{v,x,Y_i}(F) = (1\pm\eta')^{\ell}\frac{g^{v,x,Y_i}(\varnothing)}{d^{v,x,Y_i}(\varnothing)}\mathcal{D}^{v,x,Y_i}(F)$$

for all $V(H)$-partite k-complexes F on at most $(\Delta+2)c$ vertices and so by Theorem 4.10 $\mathcal{G}^{v,x,Y_i}$ is an (η,c)-THC graph with the linear order τ and density weighted hypergraph $\mathcal{D}^{v,x,Y_i}$. This completes the proof. $\square$

The following lemma enables us to obtain THC-respecting properties for $\mathcal{G}^{v,x,Y_i}$ and $\mathcal{G}'$ from $\mathcal{G}^{v,x,*}$.

Lemma 4.39. *Assume Setup 4.37. Let $c \in \mathbb{N}$ and $\eta > 0$. Let $j \in J$, $v \in V_j$ and $x \in X_j^{\mathrm{buf}}$. Enumerate $N_{H^{(2)}}(x)$ as $y_1,\dots,y_b$ in the order according to τ and let $Y_i = \{y_1,\dots,y_i\}$ for each $i \in [b]_0$. Let H_x be a J-partite k-complex with a partition $\mathcal{X}^x$ of its vertex set and a linear order τ_x on $V(H_x)$ satisfying (AB1)–(AB3). Suppose that $\mathcal{G}^{v,x,*}$ is an (η,c)-*THC *graph with the linear order τ_x and density weighted hypergraph $\mathcal{D}^{v,x,*}$ and $\mathcal{G}^{v,x,Y_i}$ is an (η,c)-*THC *graph with the linear order τ and density weighted hypergraph $\mathcal{D}^{v,x,Y_i}$ for each $i \in [b]_0$. Then for any* THC*-respecting partial partite homomorphism ϕ of H_x into $\mathcal{G}^{v,x,*}$ such that* $\mathrm{Dom}(\phi)$ *is an initial segment of τ_x of size at most $\tau(y_b)$, we have that for each $i \in [b]_0$ the function $\phi^{(i)} : \mathrm{Dom}(\phi) \to V(\mathcal{G}^{v,x,Y_i})$, given by $\phi^{(i)}(x) = \phi(x)_{\to\mathcal{G}^{v,x,Y_i}}$ for $x \in \mathrm{Dom}(\phi)$, is a* THC*-respecting partial partite homomorphism from H to $\mathcal{G}^{v,x,Y_i}$. In particular, the function $\phi' : \mathrm{Dom}(\phi) \to V(\mathcal{G}')$, given by $\phi'(x) = \phi(x)_{\to\mathcal{G}'}$ for each $x \in \mathrm{Dom}(\phi)$, is a* THC*-respecting partial partite homomorphism from H to $\mathcal{G}'$.*

Proof. Our proof inducts on $s = \max(|\mathrm{Dom}(\phi)| - i - \tau(y_1) + 1, 0)$. For $s = 0$, we obtain the desired conclusion by observing that $\mathcal{G}^{v,x,*}$ and $\mathcal{G}^{v,x,Y_i}$ with ϕ satisfy the conditions of Lemma 4.22 because $\mathcal{G}^{v,x,Y_i}$ is isomorphic to $\mathcal{G}^{v,x,*}[\bigcup_{z\in Z_i\cup W_i} V'_z]$ under the natural projection from $\mathcal{G}^{v,x,Y_i}$ into $\mathcal{G}^{v,x,*}$, $H_x[Z_i \cup W_i]$ with the order induced by τ_x is a copy of (H,τ), Z_i is an initial segment of τ_x and vertices in W_i have neighbours in only $Z_i \cup W_i$. Now consider $s \in [b]$. Let w be the last element of $\mathrm{Dom}(\phi)$ in the order according to τ and let ψ be the restriction of ϕ to $\mathrm{Dom}(\phi) \setminus \{w\}$. For $s = 1$ we obtain the desired conclusion by applying Lemma 4.23 with $\mathcal{G}_{\psi}^{v,x,Y_i}$ and $\mathcal{G}_{\psi}^{v,x,Y_{i+1}}$, observing that the restriction of ϕ to $\{w\}$ corresponds to a THC-respecting partial partite homomorphism from $H - \mathrm{Dom}(\psi)$ to $\mathcal{G}_{\psi}^{v,x,Y_{i+1}}$. For $s > 1$ we observe that ψ corresponds to a THC-respecting partial partite homomorphism from H to $\mathcal{G}^{v,x,Y_i}$ by the inductive hypothesis and this extends to ϕ because the link graph obtained by embedding the last vertex is identical to that for $\mathcal{G}^{v,x,Y_{i+1}}$. Finally, observe that $\mathcal{G}'$ is isomorphic to $\mathcal{G}^{v,x,Y_0}$. □

4.6.4 Bad Vertices

Our vertex-by-vertex embedding approach requires us to make embedding choices which enable the future continuation of our embedding procedure. We introduce the notion of bad vertices, which covers vertices that would cause the resultant partial partite homomorphism to lose desirable localised embedding properties or THC-related properties.

Definition 4.40 (Bad vertices, badness condition)**.** Let $k, c \in \mathbb{N}$, $\varepsilon > 0$ and J be a finite set. Let H be a J-partite k-complex with a partition $\mathcal{X}$ of $V(H)$ and a linear order τ on $V(H)$. Let H_+ be a J-partite k-complex with a partition $\mathcal{X}^+$ of $V(H_+)$ and a linear order τ_+ on $V(H_+)$ such that $V(H) \subseteq V(H_+)$, $H_+[V(H)] = H$ and $V(H)$ ordered according to τ forms an initial segment of τ_+. Let $\mathcal{G}$ be an (η, c)-THC graph for (H_+, τ_+) with density weighted hypergraph $\mathcal{D}$ on J and a partition $\mathcal{V}$ of $V(\mathcal{G})$. Let the standard construction of $(\mathcal{G}, \mathcal{V})$ with respect to $(H, \mathcal{X})$ and $(H_+, \mathcal{X}^+)$ be $\mathcal{G}'$ and $\mathcal{G}^+$ respectively. Let ϕ be a good partial partite homomorphism from H to $\mathcal{G}'$ which is THC-respecting for $(\mathcal{G}^+, H_+, \tau_+)$. Let $Q \subseteq V(H)$ be a set of vertices. Let $x \in V(H) \setminus \mathrm{Dom}(\phi)$ be the first unembedded vertex according to τ. We say that a vertex $v \in \mathcal{C}_\phi(x)$ is *bad for x with respect to H_+, ϕ and Q* if the extension $\phi \cup \{x \to v\}$ is not a good partial partite homomorphism from H to $\mathcal{G}'$ which is THC-respecting for $(\mathcal{G}^+, H_+, \tau_+)$, or there is a neighbour $y \in V(H) \setminus (\mathrm{Dom}(\phi) \cup Q)$ of x in $H^{(2)}$ such that

$$\deg_{\mathcal{G}^{(2)}}(v; A_\phi^{\mathrm{main}}(y)) < (1 - 2\varepsilon) d_\phi(i(x)i(y)) |A_\phi^{\mathrm{main}}(y)|. \tag{4.7}$$

Let $B_{H_+,\phi,Q}(x)$ be the set of vertices in $\mathcal{C}_\phi(x)$ which are bad for x with respect to ϕ, H_+ and Q. We will often omit mention of some or all of ϕ, H_+ and Q when they are clear from context; in particular, H_+ will always be clear in our applications. When we have a sequence $(\phi_t)_{t\in[\ell]_0}$ of good partial partite homomorphisms and a sequence $(Q_t)_{t\in[\ell]_0}$ of subsets of $V(H)$, with H_+ being clear from context, for each $t \in [\ell]_0$ we write $B_t(x)$ for the set of bad vertices for x with respect to ϕ_t, H_+ and Q_t.

The following lemma bounds the number of bad vertices with respect to a THC-respecting good partial partite homomorphism.

Lemma 4.41. *Let $k \in \mathbb{N}$, $c \geq 4$ and J be a finite set. Let $\mu, \varepsilon, \eta > 0$ satisfy $\mu \leq \frac{1}{4}$ and $4\eta^{1/7} \leq \varepsilon \leq \frac{\mu}{4\Delta c^{1/7}}$. Let H be a J-partite k-complex with a partition $\mathcal{X} = \{X_j\}_{j\in J}$ of $V(H)$ and a linear order τ on $V(H)$. Let H_+ be a J-partite k-complex with a partition $\mathcal{X}^+ = \{\overline{X}_j\}_{j\in J}$ of $V(H_+)$ and a linear order τ_+ on $V(H_+)$ such that $V(H) \subseteq V(H_+)$, $H_+[V(H)] = H$ and $V(H)$ ordered according to τ forms an initial segment of τ_+. Let $\mathcal{D}$ be a weighted hypergraph on J with $d(\varnothing) = 1$, $d(\{j\}) = 1$ for all $j \in J$ and $d(e) > 0$ for all $e \subseteq \{i(f) : f \in E(H)\}$. Let $\mathcal{G}$ be a binary (η, c)-*THC *graph for (H_+, τ_+) with density weighted hypergraph $\mathcal{D}$ and a partition $\mathcal{V} = \{V_j\}_{j\in J}$ of $V(\mathcal{G})$. Let the standard construction of $(\mathcal{G}, \mathcal{V})$ with respect to $(H, \mathcal{X})$ and $(H_+, \mathcal{X}^+)$ be $\mathcal{G}'$ and $\mathcal{G}^+$ respectively. Let ϕ be a good partial partite homomorphism from H to $\mathcal{G}'$ which is* THC*-respecting for $(\mathcal{G}^+, H_+, \tau_+)$. Let $Q \subseteq V(H)$ be a set of vertices such that for each $y \in V(H) \setminus (\mathrm{Dom}(\phi) \cup Q)$ we have*

$$|A^{\mathrm{main}}_\phi(y)| \geq (1 - 2\varepsilon)^{\pi_\phi(y)} \mu |\mathcal{C}^{\mathrm{main}}_\phi(y)|.$$

Then at most $5\Delta\varepsilon|\mathcal{C}_\phi(x)|$ vertices of $\mathcal{C}_\phi(x)$ are bad for the first vertex $x \in V(H)\setminus\mathrm{Dom}(\phi)$ according to τ with respect to ϕ, H_+ and Q.

Proof. We consider the various reasons for a vertex $v \in \mathcal{C}_\phi(x)$ to be bad for x with respect to ϕ, H_+ and Q. It could be that the extension $\phi \cup \{x \to v\}$ is not THC-respecting for $(\mathcal{G}^+, H_+, \tau_+)$. Since $\mathcal{G}'_\phi$ is binary by Lemma 4.25, ϕ is THC-respecting for $(\mathcal{G}^+, H_+, \tau_+)$ and (THC2) holds, this occurs for at most $\eta|\mathcal{C}_\phi(x)|$ vertices $v \in \mathcal{C}_\phi(x)$.

It could be that the extension $\phi \cup \{x \to v\}$ is not a good partial partite homomorphism. (GPH1) cannot fail because $v \in \mathcal{C}_\phi(x)$. Since ϕ is a good partial partite homomorphism, (GPH2) can only fail for neighbours $y \in V(H) \setminus \mathrm{Dom}(\phi)$ of x in $H^{(2)}$. Let $y \in V(H) \setminus \mathrm{Dom}(\phi)$ be a neighbour of x in $H^{(2)}$. By Lemma 4.28 with $\mathcal{G}'$ as $\mathcal{H}$ and the down-closure complex of xy as F we deduce that $G = G^\phi_{\{\{x\},\{y\}\}}$ is (ε)-regular. Since ϕ is a good partial partite homomorphism, we have $|\mathcal{C}^{\mathrm{q}}_\phi(y)|, |\mathcal{C}^{\mathrm{buf}}_\phi(y)| \geq (1-\varepsilon)^{\pi_\phi(y)}\mu|\mathcal{C}_\phi(y)| \geq \varepsilon|\mathcal{C}_\phi(y)|$ and $|\mathcal{C}^{\mathrm{main}}_\phi(y)| \geq (1-\varepsilon)^{\pi_\phi(y)}(1-2\mu)|\mathcal{C}_\phi(y)| \geq \varepsilon|\mathcal{C}_\phi(y)|$.

Let d be the density of G and define the following.

$$\begin{aligned}
U_\phi^{\mathrm{main}} &:= \{v \in \mathcal{C}_\phi(x) : \deg_G(v; \mathcal{C}_\phi^{\mathrm{main}}(y)) < (1-\varepsilon)d|\mathcal{C}_\phi^{\mathrm{main}}(y)|\},\\
U_\phi^{\mathrm{q}} &:= \{v \in \mathcal{C}_\phi(x) : \deg_G(v; \mathcal{C}_\phi^{\mathrm{q}}(y)) < (1-\varepsilon)d|\mathcal{C}_\phi^{\mathrm{q}}(y)|\},\\
U_\phi^{\mathrm{buf}} &:= \{v \in \mathcal{C}_\phi(x) : \deg_G(v; \mathcal{C}_\phi^{\mathrm{buf}}(y)) < (1-\varepsilon)d|\mathcal{C}_\phi^{\mathrm{buf}}(y)|\}.
\end{aligned}$$

By Lemma 4.16 we have $|U_\phi^{\mathrm{main}}|, |U_\phi^{\mathrm{q}}|, |U_\phi^{\mathrm{buf}}| < \varepsilon|\mathcal{C}_\phi(x)|$. Since x has at most Δ unembedded neighbours, there are at most $3\Delta\varepsilon|\mathcal{C}_\phi(x)|$ vertices $v \in \mathcal{C}_\phi(x)$ such that (GPH2) fails for $\phi \cup \{x \to v\}$.

It could be that there is some neighbour $y \in V(H) \setminus (Q \cup \mathrm{Dom}(\phi))$ of x in $H^{(2)}$ such that the badness condition (4.7) holds. As noted previously, the pair $(\mathcal{C}_\phi(x), \mathcal{C}_\phi(y))$ is (ε)-regular. By assumption and because (GPH2) holds for y with respect to ϕ we have

$$\begin{aligned}
|A_\phi^{\mathrm{main}}(y)| &\geq (1-2\varepsilon)^{\pi_\phi(y)}\mu|\mathcal{C}_\phi^{\mathrm{main}}(y)| \geq (1-2\varepsilon)^{2\pi_\phi(y)}(1-2\mu)\mu|\mathcal{C}_\phi(y)|\\
&\geq \varepsilon|\mathcal{C}_\phi(y)|,
\end{aligned}$$

so the badness condition (4.7) holds for at most $\varepsilon|\mathcal{C}_\phi(x)|$ vertices v in $\mathcal{C}_\phi(x)$. Since x has at most Δ neighbours, there are at most $\Delta\varepsilon|\mathcal{C}_\phi(x)|$ vertices $v \in \mathcal{C}_\phi(x)$ such that the badness condition (4.7) holds for some neighbour $y \in V(H) \setminus (Q \cup \mathrm{Dom}(\phi))$ of x.

Summing up, we conclude that there are at most $5\Delta\varepsilon|\mathcal{C}_\phi(x)|$ vertices of $\mathcal{C}_\phi(x)$ which are bad for x with respect to ϕ, H_+ and Q. $\square$

4.7 Random Greedy Embedding

In this section we prove a lemma which allows us to analyse the general behaviour of our random greedy algorithm. It tells us that if we embed uniformly at random into sets which are not too small, dense spots are unlikely to arise in the candidate sets of unembedded vertices. This roughly corresponds to the Main Lemma in [38] and has a similar proof approach. In our case, we obtain the necessary regularity property from THC small subgraph counts (see Lemma 4.28).

Lemma 4.42. *Assume Setup 4.37. Let $c \geq 4$ and $\mu, \rho, \varepsilon, \eta \in (0,1]$ satisfy*

$$4\eta^{1/7} \leq \varepsilon \leq \min\left(\tfrac{\rho}{2}, \tfrac{\mu 2^{-4/\rho}}{8\Delta c^{1/7}}\right).$$

Suppose that $\mathcal{G}$ is an (η, c)-THC graph for (H, τ) with density weighted hypergraph $\mathcal{D}$ and that for some integer T we have a sequence $\phi_0, \ldots, \phi_T$ of THC-respecting *partial partite homomorphisms from H to the standard construction $\mathcal{G}'$ of $(\mathcal{G},\mathcal{V})$ with respect to $(H,\mathcal{X})$, where ϕ_0 is the trivial partial partite homomorphism and each ϕ_t is obtained from ϕ_{t-1} by embedding the first vertex $x \in V(H)\setminus \mathrm{Dom}(\phi_{t-1})$ according to the linear order induced by τ to a uniform random vertex from a subset S of $\mathcal{C}_{t-1}(x)$ with $|S| \geq \frac{1}{8}\mu|\mathcal{C}_{t-1}(x)|$. Then the following holds with probability at least $1-|J|2^{-n/(\kappa|J|)}$. For every $j \in J$ and every set $W \subseteq V_j$ of size at least $\rho|V_j|$, the number of vertices $x \in X_j$ such that there exists $t = t(x)$ such that $x \notin \mathrm{Dom}(\phi_t)$ and we have*

$$|\mathcal{C}_t(x) \cap W_{j\to x}| < (1-2\varepsilon)^{\pi_t(x)}\frac{|\mathcal{C}_t(x)||W|}{|V_j|} \tag{4.8}$$

is at most $\rho|X_j|$.

Proof. Fix $j \in J$, a set $W \subseteq V_j$ of size at least $\rho|V_j|$ and a set $X \subseteq X_j$ of size $\rho|X_j|$. We aim to show that the probability of the following event is at most $2^{-4|V_j|}$. For each $x \in X$ there is a $t = t(x)$ when x is unembedded and satisfies (4.8). We then obtain the desired result by taking a union bound over all possible choices of j, W and X.

Let $x \in X$. Note that $\mathcal{C}_0(x) = V'_x$, so (4.8) is false for x at time zero. If there is a time t at which x is unembedded and (4.8) holds for x, then fix $t = t(x)$ as the least positive integer such that (4.8) holds for x. Since (4.8) holds at time t but not at time $t-1$, we must have $\mathcal{C}_t(x) \neq \mathcal{C}_{t-1}(x)$. The candidate set of x changes only when a neighbour of x in $H^{(2)}$ is embedded, so the vertex $y \in X_{j'}$ embedded at time t must be a neighbour of x in $H^{(2)}$ and hence we have $\pi_t(x) = \pi_{t-1}(x)+1$. Moreover, since (4.8) becomes true for x at time t, we have

$$\frac{|\mathcal{C}_t(x) \cap W_{j\to x}|}{|\mathcal{C}_t(x)|} < (1-2\varepsilon)^{\pi_t(x)}\frac{|W|}{|V_j|} \leq (1-2\varepsilon)\frac{|\mathcal{C}_{t-1}(x) \cap W_{j\to x}|}{|\mathcal{C}_{t-1}(x)|}. \tag{4.9}$$

By Lemma 4.28 with $\mathcal{G}'$ as $\mathcal{H}$ and the down-closure complex of xy as F we deduce that $G = G^{\phi_{t-1}}_{\{\{x\},\{y\}\}}$ is (ε)-regular. Let $d = \frac{|\mathcal{C}_{t-1}(xy)|}{|\mathcal{C}_{t-1}(y)||\mathcal{C}_{t-1}(x)|}$ be the density of G, $Z := W_{j\to x} \cap \mathcal{C}_{t-1}(x)$, $Z' := W_{j\to x} \cap \mathcal{C}_t(x)$ and

$$U := \{v \in \mathcal{C}_{t-1}(y) : \deg_G(v; Z) < (1-\varepsilon)d|Z|\}.$$

(4.8) does not hold at time $t-1$, so we have

$$|Z| \geq (1-2\varepsilon)^{\pi_{t-1}(x)}\frac{|\mathcal{C}_{t-1}(x)||W|}{|V_j|} \geq (1-2\varepsilon)^{\pi_{t-1}(x)}\rho|\mathcal{C}_{t-1}(x)| \geq \varepsilon|\mathcal{C}_{t-1}(x)|$$

and therefore by Lemma 4.16 we deduce that $|U| < \varepsilon|\mathcal{C}_{t-1}(y)|$.

Suppose to the contrary that y is not embedded into U. By the definition of U and applying (4.9), we have

$$|Z'| = \deg_G(w; Z) \geq (1-\varepsilon)d|Z| \geq \frac{(1-\varepsilon)|\mathcal{C}_{t-1}(xy)|}{(1-2\varepsilon)|\mathcal{C}_{t-1}(y)||\mathcal{C}_t(x)|}|Z'|. \tag{4.10}$$

Let F_x, F_y and F_{xy} be the down-closure complexes of the sets $\{x\}$, $\{y\}$ and $\{xy\}$ respectively. Since $\mathcal{G}$ is the weighted analogue of a complex, it is a binary weighted hypergraph. By applying Lemma 4.25 in $\mathcal{G}$ with F_y and F_{xy} paired with ϕ_{t-1} and F_x paired with ϕ_t, we obtain

$$\frac{|\mathcal{C}_{t-1}(xy)|}{|\mathcal{C}_{t-1}(y)||\mathcal{C}_t(x)|} = \frac{\mathcal{G}'_{t-1}(xy)}{\mathcal{G}'_{t-1}(y)\mathcal{G}'_t(x)}. \tag{4.11}$$

Since $\mathcal{G}'_{t-1}$ and $\mathcal{G}'_t$ are (η, c)-THC graphs, by applying (THC1) for F_y and F_{xy} in $\mathcal{G}'_{t-1}$ and F_x in $\mathcal{G}'_t$, we obtain

$$\frac{\mathcal{G}'_{t-1}(xy)}{\mathcal{G}'_{t-1}(y)\mathcal{G}'_t(x)} \geq \frac{(1-2\eta)\mathcal{D}'_{t-1}(xy)d'_t(\varnothing)}{(1+\eta)^2\mathcal{D}'_{t-1}(y)\mathcal{D}'_t(x)} = \frac{1-2\eta}{(1+\eta)^2}. \tag{4.12}$$

Putting together (4.10)-(4.12), we obtain

$$|Z'| \geq \frac{(1-\varepsilon)\mathcal{G}'_{t-1}(xy)}{(1-2\varepsilon)\mathcal{G}'_{t-1}(y)\mathcal{G}'_t(x)}|Z'| \geq \frac{(1-\varepsilon)(1-2\eta)}{(1-2\varepsilon)(1+\eta)^2}|Z'| > |Z'|,$$

which is a contradiction. Hence, y must have been embedded into U. Since ϕ_t is created by embedding y uniformly at random into a subset of $\mathcal{C}_{t-1}(y)$ of size at least $\frac{1}{8}\mu|\mathcal{C}_{t-1}(y)|$, the probability of embedding y into U, conditioning on any history up to, but not including, the embedding of y, is at most

$$\frac{\varepsilon|\mathcal{C}_{t-1}(y)|}{\frac{1}{8}\mu|\mathcal{C}_{t-1}(y)|} = 8\varepsilon\mu^{-1}. \tag{4.13}$$

Next, we argue that the probability that for each $x \in X$ there is a first time $t = t(x)$ at which x is unembedded and satisfies (4.8) is at most

$$\left(8\Delta\varepsilon\mu^{-1}\right)^{|X|} \leq 2^{-4|V_j|}.$$

Let us denote this event by $\mathcal{E}_X$. Note that $\mathcal{E}_X$ can be represented as the union of at most $\Delta^{|X|}$ events (since $\Delta(H) \leq \Delta$) $\mathcal{E}_{X,(y_x)}$ where each of these events involves specifying for each $x \in X$ a neighbour y_x of x whose embedding occurs at time $t(x)$ and causes x to satisfy (4.8). In other words, $\mathcal{E}_{X,(y_x)}$ is the intersection over $x \in X$ of events

$\mathcal{E}_{x,y_x}$ where each of the latter events involves a neighbour y_x of x being embedded at time $t(x)$ and causing x to satisfy (4.8).

By (4.13) the probability of $\mathcal{E}_{x,y_x}$, conditioning on any history up to, but not including, the embedding of y_x, is at most $8\varepsilon\mu^{-1}$. Since the vertices of X are at distance at least 3, the vertices y_x are distinct. Hence, the probability of each event $\mathcal{E}_{X,(y_x)}$ is the product of the aforementioned conditional probabilities for $\mathcal{E}_{x,y_x}$, giving

$$\mathbb{P}(\mathcal{E}_{X,(y_x)}) \le \left(8\varepsilon\mu^{-1}\right)^{|X|}.$$

Applying the union bound over the events $\mathcal{E}_{X,(y_x)}$, we conclude that

$$\mathbb{P}(\mathcal{E}_X) \le \left(8\Delta\varepsilon\mu^{-1}\right)^{|X|} \le 2^{-4|V_j|}.$$

Taking a union bound over the at most $2^{|V_j|}$ choices of W in V_j and the at most $2^{|X_j|} = 2^{|V_j|}$ choices of X' in X_j, we find that for any fixed $j \in J$ the probability that there exist subsets $W \subseteq V_j$ and $X \subseteq X_j$, of sizes $\rho|V_j|$ and $\rho|X_j|$ respectively, such that each vertex x of X satisfies (4.8) at some time t is at most $2^{2|X_j|} \cdot 2^{-4|X_j|} = 2^{-2|X_j|}$. Since $\mathcal{X}$ is κ-balanced, we have $|X_j| \ge \frac{n}{\kappa|J|}$. Hence, taking a union bound over the $|J|$ elements of J we conclude that, with probability at most $\sum_{j\in J} 2^{-2|X_j|} \le |J|2^{-n/(\kappa|J|)}$, there exists $j \in J$ and a subset $W \subseteq V_j$ such that there are $\rho|X_j|$ vertices x of X_j which satisfy (4.8) at some time t. $\square$

4.8 Queue Embedding

In this section we prove several lemmas — Lemmas 4.43, 4.44 and 4.48 — which enable us to show that our random greedy algorithm successfully completes. Let us now explain our approach. We want to show that queue candidate sets do not get overused, so queue vertices always have available queue candidates. Through a reduction by a stochastic process inequality, it suffices to bound the sum over the queue vertices of the probability of embedding a queue vertex into a potential queue candidate set, conditioned on the history up to, but not including, the embedding of that queue vertex.

Candidate sets are sublinearly small, so there is little hope of a direct estimate of these conditional probabilities. We use auxiliary constructions (see Lemmas 4.34 and 4.38(ii) in Sections 4.6.1 and 4.6.3 respectively) to extract counts of small complexes subject to certain neighbourhood-type restrictions. This turns out to be useful as we can obtain the count of the complex $H^{\leq 1}(y)$, with y going into a specified potential candidate set, at a time before any vertex of $H^{\leq 2}(y)$ is embedded. We then analyse the evolution of this count as we embed vertices in $N^{<2}(y)$ (updating the complex by removing vertices as they are embedded).

The treatment above gives us the count of y going to a potential candidate set, which is related to the conditional probability of interest; the hope is that the relevant sum over all queue vertices turns out to be reasonably well-behaved. Of course, how the count evolves is highly unpredictable for any single vertex y and we do not have advance sight of the queue. In particular, we cannot rule out the possibility that the misbehaviour of a vertex is highly correlated with that vertex entering the queue.

We show, through Lemmas 4.43 and 4.44, that only a small fraction of vertices have atypically high counts of y going to a specified potential candidate set; we also show, through Lemma 4.48, that the relevant sum of counts over these vertices turns out to be reasonably well-behaved, thereby establishing the required outcome. This means that misbehaving vertices, which we term *exceptional vertices*, constitute only a tiny fraction and occupy only a manageable fraction of potential candidate sets.

4.8.1 Exceptional Vertices

Here we introduce the notion of exceptional vertices in X_j in relation to a potential candidate set of a vertex in X_j. We show in Lemma 4.44 that these vertices are only a tiny fraction of X_j. The motivation for considering these vertices is that while we expect our random greedy algorithm to embed close to proportionally into most reasonably well-structured sets (even small ones), we cannot be sure that misbehaviour does not strongly correlate with entering the queue. As such, we want to understand the behaviour of misbehaving vertices and how much of a potential candidate set they may occupy, independent of the queue vertices.

We provide some definitions of useful objects and quantities. Assume Setup 4.37. For $a, b \in \mathbb{N}$ let $\delta_{a,b} := 2^a \varepsilon_2^{1-b}$. Let $j \in J$ and $x, y \in X_j$. Let $\vec{v} = (\psi(z))_{z \in \mathrm{Dom}(\psi)}$ be a tuple of vertices in $\mathcal{G}'$ for some partial partite homomorphism ψ from $H^{-1}(x)$ to $\mathcal{G}'$ and set $U := \mathcal{C}_{\vec{v}}(x)$. Here U represents a potential candidate set of a vertex x with its neighbours in $\mathrm{Dom}(\psi)$ embedded to the entries of $\vec{v}$. Let $T^{x,\vec{v},y}$ be the set of THC-respecting tuples in $V'_{N^{<2}(y)}$ for $(\mathcal{G}^{x,\vec{v},y}, H^{\leq 2}(y))$; this represents the 'good' set for the embedding of the vertices in $N^{<2}(y)$.

We need definitions and notation to describe the embedding of each vertex in $N^{<3}(y)$. Let $z \in N^{<3}(y)$ and $\delta_{y,z} := \delta_{a,b}$, where a and b are positive integers such that z is the ath vertex in $N^{<3}(y)$ and the bth vertex in $N^{<2}(y)$. Set $W^z := \{w \in N^{<2}(y) : \tau(w) \geq \tau(z)\}$, $W^z_- := W^z \setminus \{z\}$, $W^z_+ := W^z_- \cup \{z\}$, $H^z := H[W^z]$, $\widetilde{H}^z := H[W^z_-]$ and $\overline{H}^z := H[W^z_+]$; we use these sets of unembedded vertices and subcomplexes on H induced on them to describe the effect of embedding z. Set $d_{y,z} := \frac{|\mathcal{C}_\phi(\overline{H}^z)|}{|\mathcal{C}_\phi(z)||\mathcal{C}_\phi(\widetilde{H}^z)|}$. Let ϕ be a partial partite homomorphism from H to $\mathcal{G}'$ with $\mathrm{Dom}(\phi) = \{w \in V(H) : \tau(w) < \tau(z)\}$. Let $T^{x,\vec{v},y}_{\phi,z}$ be the set of tuples $\vec{u}$ in V'_{W^z} representing an element of $\mathcal{C}_\phi(H^z)$ where the concatenation of $(\phi(a))_{a \in \mathrm{Dom}(\phi) \cap N^{<2}(y)}$ and $\vec{u}$ produces a tuple in $T^{x,\vec{v},y}$; this represents the updated version of $T^{x,\vec{v},y}$ with the embedding of the vertices coming before z accounted for.

Now we shall define $W^{x,\vec{v},y}_{\phi,z}$, which contains the 'bad' vertices for the embedding of z; this comprises the vertices which would put 'at risk' the 'local' THC property of $\mathcal{G}^{x,\vec{v},y}$. For $z \in N^{-3}(y)$, set $W^{x,\vec{v},y}_{\phi,z} := \{w \in \mathcal{C}_\phi(z) : \deg_{G^\phi_{\{z\},W^z}}(w; T^{x,\vec{v},y}_{\phi,z}) < (1-\varepsilon_2) d_{y,z} |T^{x,\vec{v},y}_{\phi,z}|\}$. For $z \in N^{<2}(y)$, write $T^{x,\vec{v},y}_{\phi,z}(u)$ for the set of elements of $T^{x,\vec{v},y}_{\phi,z}$ with u as the z-entry and $F_{\phi,z}(u)$ for the set of elements of $\mathcal{C}_\phi(H^z)$ with u as the z-entry. Set $t^{x,\vec{v},y}_{\phi,z}(u) := |T^{x,\vec{v},y}_{\phi,z}(u)|$ and $f_{\phi,z}(u) := |F_{\phi,z}(u)|$. Set $W^{x,\vec{v},y}_{\phi,z} := \{w \in \mathcal{C}_\phi(z) : f_{\phi,z}(w) - t^{x,\vec{v},y}_{\phi,z}(w) > \delta^{1/2}_{y,z} d_{y,z} |\mathcal{C}_\phi(\widetilde{H}^z)|\}$.

Assume Setup 4.37 and suppose that for some integer T we have a sequence $\phi_0, \ldots, \phi_T$ of THC-respecting partial partite homomorphisms from H to $\mathcal{G}'$, where ϕ_0 is the trivial partial partite homomorphism and each ϕ_t is obtained from ϕ_{t-1} by embedding the first vertex $w \in V(H) \setminus \mathrm{Dom}(\phi_{t-1})$ to a vertex from $\mathcal{C}_{t-1}(w)$. Let $j \in J$ and $x, y \in X_j$. Let $\vec{v} = (\psi(z))_{z \in \mathrm{Dom}(\psi)}$ be a tuple of vertices in $\mathcal{G}'$ for some partial partite homomorphism ψ from $H^{-1}(x)$ to $\mathcal{G}'$. We say that $z \in N^{<3}(y) \cap \mathrm{Dom}(\phi_T)$ is a *trigger* for y under $(x, \vec{v})$ if no vertex embedded before z is a trigger for y under $(x, \vec{v})$ and z is embedded into $W^{x,\vec{v},y}_{\phi_{\tau(z)-1},z}$. In other words, z is a trigger for y if it puts 'at risk' the 'local' THC property of $\mathcal{G}^{x,\vec{v},y}$. We say that a trigger z for y under $(x, \vec{v})$ is an a-trigger for y under $(x, \vec{v})$ if z is the ath vertex in the ordering of $N^{<3}(y)$ induced by τ. We say that y is a-triggered under $(x, \vec{v})$ if the ath vertex in the ordering of $N^{<3}(y)$ induced by τ is a trigger for y under $(x, \vec{v})$.

We say that y is *exceptional for* $(x, \vec{v})$ if there is a trigger z for y under $(x, \vec{v})$. We say that y is (h, a)-*exceptional for* $(x, \vec{v})$ if $y \in X_{j,h}$ and y is a-triggered under $(x, \vec{v})$. Denote by $\mathrm{exc}^{x,\vec{v}}_{h,a}$ the set of (h, a)-exceptional vertices for $(x, \vec{v})$. For $b \in [n^{(2)}_{j,h}]$ denote by

a_b the unique integer such that the bth vertex in $N^{<2}(y)$ is the a_bth vertex in $N^{<3}(y)$, set $a_0 := 0$ and set

$$E_b^{x,\vec{v},h} := \bigcup_{a\in[a_b]} \mathrm{exc}_{h,a}^{x,\vec{v}} \text{ and } E_{b-}^{x,\vec{v},h} := \bigcup_{a\in[a_b-1]} \mathrm{exc}_{h,a}^{x,\vec{v}}.$$

For $A = E_b^{x,\vec{v},h}, E_{b-}^{x,\vec{v},h}$ we set $A(T) := \{y \in A : \tau(y) \leq T\}$. We sometimes omit b when we mean $b = n_{j,h}^{(2)}$ and omit h to refer to the objects obtained by taking a union over $h \in [\ell_j]$.

The following lemma tells us that for vertices which are not exceptional, certain quantities related to probabilities of embedding into potential candidate sets are typical in the sense that they are within a tiny relative error of their expected values. In particular, it means that all vertices which witness atypical values must be exceptional.

Lemma 4.43. *Assume Setup 4.37. Let $c \in \mathbb{N}$ and $0 < \eta < \frac{1}{10}$. Let $j \in J$ and $x, y \in X_j$. Let $\vec{v} = (\psi(z))_{z\in\mathrm{Dom}(\psi)\cap N^{-1}(x)}$ be a tuple of vertices in $\mathcal{G}'$ for some* THC*-respecting partial partite homomorphism ψ from H to $\mathcal{G}'$. Let ϕ be a* THC*-respecting partial partite homomorphism from H to $\mathcal{G}'$ with $N^{<2}(y) \subseteq \mathrm{Dom}(\phi)$ such that $(\phi(z))_{z\in N^{<2}(y)} \in T^{x,\vec{v},y}$. Suppose that $\mathcal{G}^{x,\vec{v},y}$ is an (η, c)-*THC *graph with the linear order on $N^{\leq 2}(y)$ induced by τ and density weighted hypergraph $\mathcal{D}^{x,\vec{v},y}$. Then*

$$|\mathcal{C}_\phi(y) \cap \mathcal{C}_{\vec{v}}(x)| = (1 \pm 4\eta)\frac{|\mathcal{C}_\phi(y)||\mathcal{C}_{\vec{v}}(x)|}{|V_j|}.$$

Proof. Since $(\phi(z))_{z\in N^{<2}(y)} \in T^{x,\vec{v},y}$, the restriction $\overline{\phi}$ of ϕ to $N^{<2}(y)$ gives a THC-respecting partial partite homomorphism from $H^{\leq 2}(y)$ to $\mathcal{G}^{x,\vec{v},y}$. Set $W := \mathrm{Dom}(\psi) \cap N^{-1}(x)$. Let F_x and F_y be the down-closure complexes of the sets $\{x\}$ and $\{y\}$ respectively. Since $\mathcal{G}$ is the weighted analogue of a complex, it is a binary weighted hypergraph. By applying Lemma 4.25 in $\mathcal{G}$ with F_x paired with ψ and F_y paired with ϕ, and Lemma 4.24(iv) in $\mathcal{G}^{x,\vec{v},y}$ with F_y paired with $\overline{\phi}$, we obtain

$$|\mathcal{C}_\phi(y)||\mathcal{C}_{\vec{v}}(x)| = \mathcal{G}'_\phi(y)\mathcal{G}'_{\vec{v}}(x)|V_j|^2 \text{ and } |\mathcal{C}_{\overline{\phi}}(y)| = \mathcal{G}^{x,\vec{v},y}_{\overline{\phi}}(y)|V_j|. \tag{4.14}$$

Since $\mathcal{G}'_{\vec{v}}$, $\mathcal{G}'_\phi$ and $\mathcal{G}^{x,\vec{v},y}_{\overline{\phi}}$ are (η, c)-THC graphs, by applying (THC1) for F_x in $\mathcal{G}'_{\vec{v}}$ and F_y in $\mathcal{G}'_\phi$ and $\mathcal{G}^{x,\vec{v},y}_{\overline{\phi}}$, we obtain

$$\frac{\mathcal{G}^{x,\vec{v},y}_{\overline{\phi}}(y)}{\mathcal{G}'_\phi(y)\mathcal{G}'_{\vec{v}}(x)} = \frac{(1 \pm \eta)d^{x,\vec{v},y}_{\overline{\phi}}(y)}{(1 \pm \eta)^2 d'_\phi(y)d'_{\vec{v}}(x)} = 1 \pm 4\eta. \tag{4.15}$$

Combining (4.14) and (4.15), we obtain

$$|\mathcal{C}_\phi(y) \cap \mathcal{C}_{\vec{v}}(x)| = |\mathcal{C}_{\overline{\phi}}(y)| = (1 \pm 4\eta)\frac{|\mathcal{C}_\phi(y)||\mathcal{C}_{\vec{v}}(x)|}{|V_j|}$$

as desired. $\square$

The following lemma tells us that exceptional vertices constitute only a small fraction of each X_j.

Lemma 4.44. *Assume Setup 4.37. Let $c \geq \Delta^2 + 3$ and $\mu, \rho, \varepsilon_2, \eta \in (0, 1]$ satisfy*

$$4(\Delta^2\eta)^{1/7} \leq \varepsilon_2 \leq \left(\frac{\mu\rho}{2^{\Delta^3+4}}\right)^{2^{\Delta^2}} c^{-1/7}.$$

Suppose that $\mathcal{G}$ is an (η, c)-THC graph for (H, τ) with density weighted hypergraph $\mathcal{D}$. Suppose that for some integer T we have a sequence $\phi_0, \dots, \phi_T$ of THC-respecting partial partite homomorphisms from H to $\mathcal{G}'$, where ϕ_0 is the trivial partial partite homomorphism and each ϕ_t is obtained from ϕ_{t-1} by embedding the first vertex $z \in V(H) \setminus \mathrm{Dom}(\phi_{t-1})$ according to the linear order induced by τ to a uniform random vertex from a subset S of $\mathcal{C}_{t-1}(z)$ with $|S| \geq \frac{1}{8}\mu|\mathcal{C}_{t-1}(z)|$. Then the following holds with probability at least $1 - 2^{2^{\Delta_R^3+1}}\Delta n^{\Delta+1}e^{-\frac{\rho n}{8\kappa|J|}}$. For any $j \in J$, $h \in [\ell_j]$, $a \in [n_{j,h}^{(3)}]$, $x \in X_j$ and tuple $\vec{v} = (\psi(z))_{z\in\mathrm{Dom}(\psi)}$ of vertices in $\mathcal{G}'$ for some partial partite homomorphism ψ from $H^{-1}(x)$ to $\mathcal{G}'$ such that for all $y \in X_{j,h}$ we have that $\mathcal{G}^{x,\vec{v},y}$ is an (η, c)-THC graph with the linear order on $N^{\leq 2}(y)$ induced by τ and density weighted hypergraph $\mathcal{D}^{x,\vec{v},y}$, the number of vertices which are (h, a)-exceptional for $(x, \vec{v})$ is at most $\rho|X_j|$.

The rest of this subsection builds up to the proof of Lemma 4.44. This is a consequence of the THC property of $\mathcal{G}^{x,\vec{v},y}$ proved in Lemma 4.38(ii). The complication here is that our embedding procedure does not actively seek to preserve the THC property for such structures; as such, we need to work a little harder to show that the pseudorandom structure means that we are very unlikely to make many bad choices. The following lemma tells us that $T_{\phi,z}^{x,\vec{v},y}$ contains all but a small linear fraction of $\mathcal{C}_\phi(H^z)$; we will apply this later to show that the trigger sets for exceptionality are small.

Lemma 4.45. *Assume Setup 4.37. Let $c \geq \Delta^2+1$ and $\varepsilon_2, \eta \in (0, 1]$ satisfy $4(\Delta^2\eta)^{1/7} \leq \varepsilon_2$. Suppose that $\mathcal{G}$ is an (η, c)-THC graph for (H, τ) with density weighted hypergraph $\mathcal{D}$. Let $j \in J$ and $x, y \in X_j$. Let $\vec{v} = (\psi(z))_{z\in\mathrm{Dom}(\psi)}$ be a tuple of vertices in $\mathcal{G}'$ for some* *partial partite homomorphism ψ from $H^{-1}(x)$ to $\mathcal{G}'$ such that $\mathcal{G}^{x,\vec{v},y}$ is an (η,c)-*THC *graph with the linear order on $N^{\leq 2}(y)$ induced by τ and density weighted hypergraph $\mathcal{D}^{x,\vec{v},y}$. Let $z \in N^{<3}(y)$. Let ϕ be a* THC*-respecting partial partite homomorphism from H to $\mathcal{G}'$ with* $\mathrm{Dom}(\phi) = \{w \in V(H) : \tau(w) < \tau(z)\}$*. Suppose that no vertex in* $\mathrm{Dom}(\phi)$ *is a trigger for y under $(x,\vec{v})$. Then $|T^{x,\vec{v},y}_{\phi,z}| \geq (1-\delta_{y,z})|\mathcal{C}_\phi(H^z)|$.*

Proof. Set $q := n^{(3)}_{j,h}$ and enumerate $N^{<3}(y)$ as $z_1,\dots,z_q$. We restate the desired outcome of Lemma 4.45 as follows and prove it by induction on p. Given $p \in [q]$ and any THC-respecting partial partite homomorphism ϕ from H to $\mathcal{G}'$ with $\mathrm{Dom}(\phi) = \{w \in V(H) : \tau(w) < \tau(z_p)\}$ such that no vertex in $\mathrm{Dom}(\phi)$ is a trigger for y under $(x,\vec{v})$, we have $|T^{x,\vec{v},y}_{\phi,z_p}| \geq (1-\delta_{y,z_p})|\mathcal{C}_\phi(H^{z_p})|$.

For $p = 1$, let ϕ be a THC-respecting partial partite homomorphism from H to $\mathcal{G}'$ with $\mathrm{Dom}(\phi) = \{w \in V(H) : \tau(w) < \tau(z_1)\}$. Observe that $T^{x,\vec{v},y}_{\phi,z_1} = T^{x,\vec{v},y}$ because $\mathrm{Dom}(\phi) \cap N^{<2}(y) = \varnothing$. Set $t := n^{(2)}_{j,h}$. Enumerate $N^{<2}(y)$ as $w_1,\dots,w_t$. For each $s \in [t]_0$ set $W_s := \{w_1,\dots,w_s\}$, $W_s^c := H^{<2}(y) \setminus W_s$ and $H_s := H^{<2}(y)[W_s]$. In other words, H_s is the subcomplex of $H^{<2}(y)$ induced on its first s vertices. We shall prove the following claim. Since (THC1) enables approximation of sizes of candidate sets by quantities from the appropriate density weighted hypergraph, we use such quantities in the expression in the claim statement because it simplifies the proof.

Claim 4.46. *Given $s \in [t]_0$, each $(u_i)_{i\in[s]}$ in V'_{W_s} which is a* THC*-respecting tuple for $(\mathcal{G}^{x,\vec{v},y}, H^{\leq 2}(y))$ has at least $(1-\eta)^{2(t-s)}|V'_{W_s^c}|\frac{\mathcal{D}(H^{<2}(y))}{\mathcal{D}(H_s)}$ extensions in $T^{x,\vec{v},y}$.*

Proof. Our proof proceeds by backwards induction on s. By definition of $T^{x,\vec{v},y}$, the statement with $s = t$ is trivially true. Now consider $0 \leq s < t$. Let $\vec{u} = (u_i)_{i\in[s]}$ be a THC-respecting tuple in V'_{W_s} for $(\mathcal{G}^{x,\vec{v},y}, H^{\leq 2}(y))$. $\mathcal{G}^{x,\vec{v},y}_{\vec{u}}$ is an (η,c)-THC graph, so by (THC1) and (THC2) applied in $\mathcal{G}^{x,\vec{v},y}_{\vec{u}}$ for $w_[s+1]$ there are at least $(1-\eta)|\mathcal{C}^{x,\vec{v},y}_{\vec{u}}(w_{s+1})| \geq (1-\eta)^2|V'_{w_{s+1}}|\frac{\mathcal{D}(H_{s+1})}{\mathcal{D}(H_s)}$ extensions of $\vec{u}$ to a THC-respecting tuple in $V'_{W_{s+1}}$ for $(\mathcal{G}^{x,\vec{v},y}, H^{\leq 2}(y))$. By the inductive hypothesis, any such extension $(u_i)_{i\in[s+1]}$ has at least $(1-\eta)^{2(t-s-1)}|V'_{W^c_{s+1}}|\frac{\mathcal{D}(H^{<2}(y))}{\mathcal{D}(H_{s+1})}$ extensions in $T^{x,\vec{v},y}$. Hence, $\vec{u}$ has at least $(1-\eta)^{2(t-s)}|V'_{W^c_s}|\frac{\mathcal{D}(H^{<2}(y))}{\mathcal{D}(H_s)}$ extensions in $T^{x,\vec{v},y}$ in total. ■

By assumption $\mathcal{G}^{x,\vec{v},y}$ is an (η,c)-THC graph, so the empty tuple () is THC-respecting for $(\mathcal{G}^{x,\vec{v},y}, H^{\leq 2}(y))$. Hence, we may apply Claim 4.46 to deduce that there are at least $(1-\eta)^{2t}\frac{\mathcal{D}(H^{<2}(y))}{d(\varnothing)}|V'_{N^{<2}(y)}|$ extensions of () in $T^{x,\vec{v},y}$. These correspond to distinct elements of $T^{x,\vec{v},y}$, so we have

$$\begin{aligned}|T^{x,\vec{v},y}_{\phi,z_1}| = |T^{x,\vec{v},y}| &\geq (1-\eta)^{2\Delta^2}\tfrac{\mathcal{D}(H^{<2}(y))}{d(\varnothing)}|V'_{N^{<2}(y)}| \\ &\geq (1-\delta_{y,z_1})|\mathcal{C}_{\tau(z_1)-1}(H^{z_1})|,\end{aligned}$$

where the final inequality is by (THC1) for $H^{<2}(y)$ in $\mathcal{G}'_\phi$.

Now consider $p > 1$. Let ϕ be a THC-respecting partial partite homomorphism from H to $\mathcal{G}'$ with $\mathrm{Dom}(\phi) = \{w \in V(H) : \tau(w) < \tau(z_p)\}$. Let ψ be the restriction of ϕ to $\{w \in V(H) : \tau(w) < \tau(z_{p-1})\}$. By the inductive hypothesis we have $|T^{x,\vec{v},y}_{\psi,z_{p-1}}| \geq (1-\delta_{y,z_{p-1}})|\mathcal{C}_\psi(H^{z_{p-1}})|$. Consider two cases. Firstly, we could have $z_{p-1} \in N^{-3}(y)$. Since z_{p-1} not is a trigger for y and $H^{z_{p-1}} = \widetilde{H}^{z_{p-1}}$, we have

$$\begin{aligned}\left|T^{x,\vec{v},y}_{\phi,z_p}\right| &= \deg_{G^\psi_{\{z_{p-1}\},W^{z_{p-1}}}}(\phi(z_{p-1}); T^{x,\vec{v},y}_{\psi,z_{p-1}}) \\ &\geq (1-\varepsilon_2)d_{y,z_{p-1}}|T^{x,\vec{v},y}_{\psi,z_{p-1}}| \\ &\geq (1-\delta_{y,z_{p-1}})(1-\varepsilon_2)d_{y,z_{p-1}}|\mathcal{C}_\psi(H^{z_{p-1}})|.\end{aligned}$$

Since ϕ and ψ are THC-respecting and $\widetilde{H}^{z_{p-1}} = H^{z_{p-1}}$, by (THC1) for $\overline{H}^{z_{p-1}}$ and z_{p-1} in $\mathcal{G}'_\psi$ and $\widetilde{H}^{z_p}$ in $\mathcal{G}'_\phi$ we have

$$d_{y,z_{p-1}}|\mathcal{C}_\psi(H^{z_{p-1}})| = \frac{|\mathcal{C}_\psi(\overline{H}^{z_{p-1}})|}{|\mathcal{C}_\psi(z_{p-1})|} \geq (1-3\Delta^2\eta)|\mathcal{C}_\phi(H^{z_p})|.$$

Putting together the previous two inequalities, we obtain

$$\begin{aligned}\left|T^{x,\vec{v},y}_{\phi,z_p}\right| &\geq (1-\delta_{y,z_{p-1}})(1-\varepsilon_2)(1-3\Delta^2\eta)|\mathcal{C}_\phi(H^{z_p})| \\ &\geq (1-\delta_{y,z_p})|\mathcal{C}_\phi(H^{z_p})|,\end{aligned}$$

completing the proof in this case.

Otherwise, we have $z_{p-1} \in N^{<2}(y)$. Since z_{p-1} not is a trigger for y, we have

$$\left|T^{x,\vec{v},y}_{\phi,z_p}\right| = t^{x,\vec{v},y}_{\psi,z_{p-1}}(\phi(z_{p-1})) \geq |\mathcal{C}_\phi(H^{z_p})| - \delta^{1/2}_{y,z_{p-1}}d_{y,z_{p-1}}|\mathcal{C}_\psi(\widetilde{H}^{z_{p-1}})|.$$

Since ϕ and ψ are THC-respecting, by (THC1) for $\overline{H}^{z_{p-1}}$ and z_{p-1} in $\mathcal{G}'_\psi$ and $\widetilde{H}^{z_{p-1}}$ in $\mathcal{G}'_\phi$ we have

$$d_{y,z_{p-1}}|\mathcal{C}_\psi(\widetilde{H}^{z_{p-1}})| = \frac{|\mathcal{C}_\psi(\overline{H}^{z_{p-1}})|}{|\mathcal{C}_\psi(z_{p-1})|} \leq (1+3\Delta^2\eta)|\mathcal{C}_\phi(H^{z_p})|.$$

Now putting together the previous two inequalities, we obtain

$$\begin{aligned}\left|T^{x,\vec{v},y}_{\phi,z_p}\right| &\geq |\mathcal{C}_\phi(H^{z_p})| - \delta^{1/2}_{y,z_{p-1}} d_{y,z_{p-1}} |\mathcal{C}_\psi(\widetilde{H}^{z_{p-1}})| \\ &\geq |\mathcal{C}_\phi(H^{z_p})| \left(1 - \delta^{1/2}_{y,z_{p-1}}(1+3\Delta^2\eta)\right) \\ &\geq |\mathcal{C}_\phi(H^{z_p})|(1-\delta_{y,z_p}),\end{aligned}$$

completing the proof in this case. □

The following lemma tells us that the trigger sets for exceptionality are always small.

Lemma 4.47. *Assume Setup 4.37. Let $c \geq \Delta^2+3$ and $\varepsilon_2, \eta \in (0,1]$ satisfy $4(\Delta^2\eta)^{1/7} \leq \varepsilon_2 \leq 2^{-2^{\Delta^2}(\Delta^3+1)}c^{-1/7}$. Suppose that $\mathcal{G}$ is an (η, c)-THC graph for (H,τ) with density weighted hypergraph $\mathcal{D}$. Let $j \in J$ and $x, y \in X_j$. Let $\vec{v} = (\psi(z))_{z\in \mathrm{Dom}(\psi)}$ be a tuple of vertices in $\mathcal{G}'$ for some partial partite homomorphism ψ from $H^{-1}(x)$ to $\mathcal{G}'$ such that $\mathcal{G}^{x,\vec{v},y}$ is an (η, c)-THC graph with the linear order on $N^{\leq 2}(y)$ induced by τ and density weighted hypergraph $\mathcal{D}^{x,\vec{v},y}$. Let $z \in N^{<3}(y)$. Let ϕ be a THC-respecting partial partite homomorphism from H to $\mathcal{G}'$ with $\mathrm{Dom}(\phi) = \{w \in V(H) : \tau(w) < \tau(z)\}$. Suppose that no vertex in $\mathrm{Dom}(\phi)$ is a trigger for y under $(x, \vec{v})$. Then*

$$|W^{x,\vec{v},y}_{\phi,z}| < \begin{cases} \varepsilon_2 |\mathcal{C}_\phi(z)| & \text{if } z \in N^{-3}(y), \\ \delta^{1/2}_{y,z}|\mathcal{C}_\phi(z)| & \text{if } z \in N^{<2}(y). \end{cases}$$

In particular, $|W^{x,\vec{v},y}_{\phi,z}| < 2^{\Delta^3}\varepsilon_2^{1/2^{\Delta^2}}|\mathcal{C}_\phi(z)|$.

Proof. Note that $G^{\phi}_{\{\{z\},\widetilde{W}^z\}}$ is (ε_2)-regular by Lemma 4.28. Consider two cases. Firstly, we could have $z \in N^{-3}(y)$. In this case, by Lemma 4.45 we have $|T^{x,\vec{v},y}_{\phi,z}| \geq (1-\delta_{y,z})|\mathcal{C}_\phi(H^z)| \geq \varepsilon_2|\mathcal{C}_\phi(H^z)|$, so by Lemma 4.16 we conclude that $|W^{x,\vec{v},y}_{\phi,z}| < \varepsilon_2|\mathcal{C}_\phi(z)|$. Otherwise, we have $z \in N^{<2}(y)$. Then, by Lemma 4.45 and the definitions of $f_{\phi,z}(w)$, $t^{x,\vec{v},y}_{\phi,z}(w)$ and $W^{x,\vec{v},y}_{\phi,z}$, we have

$$\begin{aligned}\delta_{y,z}|\mathcal{C}_\phi(H^z)| \geq |\mathcal{C}_\phi(H^z)| - |T^{x,\vec{v},y}_{\phi,z}| &= \sum_{w\in\mathcal{C}_\phi(z)} (f_{\phi,z}(w) - t^{x,\vec{v},y}_{\phi,z}(w)) \\ &> |W^{x,\vec{v},y}_{\phi,z}|\delta^{1/2}_{y,z}\frac{|\mathcal{C}_\phi(H^z)|}{|\mathcal{C}_\phi(z)|}.\end{aligned}$$

From this we obtain $|W^{x,\vec{v},y}_{\phi,z}| < \delta^{1/2}_{y,z}|\mathcal{C}_\phi(z)|$. □

Now we apply Lemma 4.47 to prove Lemma 4.44.

Proof of Lemma 4.44. For each $j \in J$, $h \in [\ell_j]$ and $x \in X_j$ let $\vec{V}_{j,h,x}$ denote the set of all tuples $\vec{v} = (\psi(z))_{z \in \mathrm{Dom}(\psi)}$ of vertices in $\mathcal{G}'$ for some partial partite homomorphism ψ from $H^{-1}(x)$ to $\mathcal{G}'$ such that for all $y \in X_{j,h}$ we have that $\mathcal{G}^{x,\vec{v},y}$ is an (η, c)-THC graph with the linear order on $N^{\leq 2}(y)$ induced by τ and density weighted hypergraph $\mathcal{D}^{x,\vec{v},y}$. Note that $|\vec{V}_{j,h,x}| \leq \Delta n^{\Delta}$.

Let $j \in J$, $h \in [\ell_j]$, $a \in [n^{(3)}_{j,h}]$, $x \in X_j$ and $\vec{v} \in \vec{V}_{j,h,x}$. Let $y \in X_{j,h}$ and let $z_{y,a}$ be the ath vertex in the ordering of $N^{<3}(y)$ according to τ. Note that if $\tau(z_{y,a}) > T$, then $z_{y,a}$ remains unembedded throughout and therefore cannot be an a-trigger for y. Hence, let us suppose that $\tau(z_{y,a}) \leq T$. Since $\phi_{\tau(z_{y,a})}$ is obtained from $\phi_{\tau(z_{y,a})-1}$ by embedding $z_{y,a}$ to a uniform random vertex from a subset of $\mathcal{C}_{\tau(z_{y,a})-1}(z_{y,a})$ of size at least $\frac{1}{8}\mu|\mathcal{C}_{\tau(z_{y,a})-1}(z_{y,a})|$, by Lemma 4.47 the probability of embedding $z_{y,a}$ such that it is an a-trigger for y, conditioning on the history up to but not including the embedding of $z_{y,a}$, is at most

$$\frac{\left|W^{x,\vec{v},y}_{\phi_{\tau(z_{y,a})-1}, z_{y,a}}\right|}{\frac{1}{8}\mu|\mathcal{C}_{\tau(z_{y,a})-1}(z_{y,a})|} \leq 2^{\Delta^3+3}\varepsilon_2^{1/2^{\Delta^2}}\mu^{-1} \leq \rho/2.$$

Let $x_{j,h} := |X_{j,h}|$. Enumerate $\{z_{y,a} : y \in X_{j,h}\}$ in the order according to τ as $u_{a,1}, \ldots, u_{a,x_{j,h}}$. For $i \in [x_{j,h}]_0$ set

$$\mathcal{F}^{(a)}_i := \begin{cases} \mathcal{F}_{\tau(u_{a,x_{j,h}})} & \text{if } i = x_{j,h}, \\ \mathcal{F}_{\tau(u_{a,i+1})-1} & \text{otherwise.} \end{cases}$$

For $i \in [x_{j,h}]$ set

$$Y_i := \begin{cases} 1 & \text{if } \phi_{\tau(u_{a,i})}(u_{a,i}) \in W^{x,\vec{v},y}_{\phi_{\tau(u_{a,i})-1}, u_{a,i}}, \\ 0 & \text{otherwise.} \end{cases}$$

Note that Y_i is $\mathcal{F}^{(a)}_i$-measurable and $0 \leq Y_i \leq 1$. By the previous paragraph, we have $\sum_{i \in [x_{j,h}]} \mathbb{E}\left[Y_i|\mathcal{F}^{(a)}_{i-1}\right] \leq \frac{\rho}{2}|X_j|$ and $\sum_{i \in [x_{j,h}]} \mathrm{var}\left(Y_i|\mathcal{F}^{(a)}_{i-1}\right) \leq \frac{\rho}{2}|X_j|$. Then, by applying Lemma 4.12, we deduce that the probability that more than $\rho|X_j|$ vertices y are (h,a)-exceptional for $(x, \vec{v})$ is at most $e^{-\rho|X_j|/8}$. This is at most $e^{-\frac{\rho n}{8\kappa|J|}}$ because we have $|X_j| \geq \frac{n}{\kappa|J|}$ as a consequence of $\mathcal{X}$ being κ-balanced.

Finally, by taking a union bound over all $j \in J$, $h \in [\ell_j]$, $x \in X_j$ and $\vec{v} \in \vec{V}_{j,h,x}$, we find that the desired outcome holds with probability at least

$$1 - 2^{2^{\Delta^3_R+1}}\Delta n^{\Delta+1}e^{-\frac{\rho n}{8\kappa|J|}}$$

as required. $\square$

4.8.2 Bounding the Contribution

We showed in Lemma 4.44 that only a small fraction of vertices have an atypically high probability of embedding into a given potential candidate set. We shall show that the sum of the probabilities of embedding these vertices into the potential candidate set is reasonably small; a martingale concentration argument then implies that these misbehaving vertices occupy only a small fraction of the potential candidate set. To this end, we shall define a suitable event $\mathcal{E}$ and prove Lemma 4.48, which states that $\mathcal{E}$ holds asymptotically almost surely.

The embedding behaviour of a vertex y is closely linked to its candidate set right before its embedding; to understand how this behaves, we shall study how the candidate set of the unembedded part of the neighbourhood complex $H^{\leq 1}(y)$ evolves as we embed $N^{<2}(y)$. Let us now provide definitions of useful objects and quantities. Firstly, we need definitions and notation to describe the evolution of the unembedded part of the neighbourhood complex. Assume Setup 4.37. Let $j \in J$ and $h \in [\ell_j]$. Set $q := n^{(2)}_{j,h}$. Let $y \in X_{j,h}$. Enumerate $N^{\leq 2}(y)$ as $u_{y,1}, \ldots, u_{y,q+1}$ in the order according to τ. Set $T_y := \tau(u_{y,q})$. For $b \in [q]_0$ set

$$U_{y,b,>} := N^{-1}(y) \setminus \{u_{y,i} : i \in [b]\} \quad \text{and} \quad F_{y,b} := H\left[U_{y,b,>} \cup \{y\}\right].$$

Note that $F_{y,0} = H^{\leq 1}(y)$ and $F_{y,q} = (\{y\}, \varnothing)$. We need notation to describe variants of the neighbourhood complex with attached second neighbours and multiple copies of y for our stepwise updating and Cauchy–Schwarz arguments. For $b \in [q]$ set

$$F'_{y,b} := H\left[\{u_{y,b}, y\} \cup U_{y,b-1,>}\right] \quad \text{and} \quad F''_{y,b} := H\left[\{u_{y,b}\} \cup U_{y,b-1,>}\right].$$

For $b \in [q]_0$ set $\vec{B}_{y,b} := (\{y\}, V(F_{y,b}) \setminus \{y\})$, $\vec{B}'_{y,b} := (\{y\}, V(F'_{y,b}) \setminus \{y\})$ and $\vec{B}''_{y,b} := (\{u_{y,b}\}, V(F''_{y,b}) \setminus \{u_{y,b}\})$. For $b \in [q]_0$ and $p \in \mathbb{N}$ set $\vec{p} := (p, 1)$,

$$F_{y,b,p} := F_{y,b}(\vec{B}_{y,b}, \vec{p}), \quad F'_{y,b,p} := F'_{y,b}(\vec{B}'_{y,b}, \vec{p}) \quad \text{and} \quad F''_{y,b,p} := F''_{y,b}(\vec{B}''_{y,b}, \vec{p}).$$

Let $Y_{y,b,p} := \{y_i : i \in [p]\}$ be the set of copies of y in $F_{y,b,p}$ and $F'_{y,b,p}$. Let $U_{y,b,p}$ be the set of copies of $u_{y,b}$ in $F''_{y,b,p}$.

To understand the probability of embedding a vertex y into a given potential candidate set U, we shall study the evolution of the probability of randomly picking a copy of $F_{y,b}$ with y going to a vertex in U from all copies of $F_{y,b}$; for this purpose we define the quantity $R_{x,\vec{v},y,b,p}$. In practice, we work with a different quantity $S_{x,\vec{v},y,b,p}$ obtained by replacing the denominator of $R_{x,\vec{v},y,b,p}$ with its deterministic estimate (within a small relative error by THC). Suppose that for some integer T we have a sequence $\phi_0, \dots, \phi_T$ of THC-respecting partial partite homomorphisms from H to $\mathcal{G}'$, where ϕ_0 is the trivial partial partite homomorphism and each ϕ_t is obtained from ϕ_{t-1} by embedding the first vertex $z \in V(H) \setminus \mathrm{Dom}(\phi_{t-1})$ to a uniform random vertex from a subset $S_{\phi_{t-1},z}$ of $\mathcal{C}_{t-1}(z)$. Set $\Xi_{j,h} := \{y \in X_{j,h} : T_y \le T\}$ and $\xi_{j,h} := |\Xi_{j,h}|$. Let $x \in X_j$ and $\vec{v} = (\psi(z))_{z \in \mathrm{Dom}(\psi)}$ be a tuple of vertices in $\mathcal{G}'$ for some partial partite homomorphism ψ from $H^{-1}(x)$ to $\mathcal{G}'$. Set $U := \mathcal{C}_{\vec{v}}(x)$. For $y \in \Xi_{j,h}$, $b \in [q]_0$ and $p \in \mathbb{N}$ define

$$\mathrm{num}(R_{x,\vec{v},y,b,p}), \mathrm{num}(S_{x,\vec{v},y,b,p}) := |\mathcal{C}_{\min(\tau(u_{y,b+1})-1,T)}(V(F_{y,b,p})) \cap (V_{U_{y,b,>}} \times U^p)|,$$

$$\mathrm{den}(R_{x,\vec{v},y,b,p}) := |\mathcal{C}_{\min(\tau(u_{y,b+1})-1,T)}(V(F_{y,b,p}))|,$$

$$\mathrm{den}(S_{x,\vec{v},y,b,p}) := \frac{\mathcal{D}_{\min(\tau(u_{y,b+1})-1,T)}(V(F_{y,b,p}))}{d_{\min(\tau(u_{y,b+1})-1,T)}(\varnothing)} |V_y|^p \prod_{u \in U_{y,b,>}} |V_u|,$$

$$R_{x,\vec{v},y,b,p} := \frac{\mathrm{num}(R_{x,\vec{v},y,b,p})}{\mathrm{den}(R_{x,\vec{v},y,b,p})} \quad \text{and} \quad S_{x,\vec{v},y,b,p} := \frac{\mathrm{num}(S_{x,\vec{v},y,b,p})}{\mathrm{den}(S_{x,\vec{v},y,b,p})}.$$

We need to describe the cumulative probabilities. For $b \in [q]_0$ and $p \in \mathbb{N}$ define

$$R_{x,\vec{v},j,h,b,p} := \sum_{y \in E_b^{x,\vec{v},h} \cap \Xi_{j,h}} R_{x,\vec{v},y,b,p} \quad \text{and} \quad S_{x,\vec{v},j,h,b,p} := \sum_{y \in E_b^{x,\vec{v},h} \cap \Xi_{j,h}} S_{x,\vec{v},y,b,p}.$$

We need notation to describe variants of these quantities for our stepwise updating and Cauchy–Schwarz arguments. For $y \in \Xi_{j,h}$, $b \in [q]$ and $p \in \mathbb{N}$ define

$$\mathrm{num}(S'_{x,\vec{v},y,b,p}) := \left| \mathcal{C}_{\tau(u_{y,b})-1}(V(F'_{y,b,p})) \cap (V_{U_{y,b,>}} \times S_{\phi_{\tau(u_{y,b})-1},u_{y,b}} \times U^p) \right|,$$

$$\mathrm{num}(S''_{x,\vec{v},y,b,p}) := \left| \mathcal{C}_{\tau(u_{y,b})-1}(V(F'_{y,b,p})) \cap (V_{U_{y,b,>}} \times W^{x,\vec{v},y}_{\phi_{\tau(u_{y,b})-1},u_{y,b}} \times U^p) \right|,$$

$$\mathrm{den}(S'_{x,\vec{v},y,b,p}), \mathrm{den}(S''_{x,\vec{v},y,b,p}) := \frac{\mathcal{D}_{\min(\tau(u_{y,b})-1,T)}(V(F'_{y,b,p}))}{d_{\min(\tau(u_{y,b})-1,T)}(\varnothing)} |V_y|^p \prod_{u \in \{u_{y,b}\} \cup U_{y,b-1,>}} |V_u|,$$

$$S'_{x,\vec{v},y,b,p} := \frac{\mathrm{num}(S'_{x,\vec{v},y,b,p})}{\mathrm{den}(S'_{x,\vec{v},y,b,p})} \quad \text{and} \quad S''_{x,\vec{v},y,b,p} := \frac{\mathrm{num}(S''_{x,\vec{v},y,b,p})}{\mathrm{den}(S''_{x,\vec{v},y,b,p})}.$$

For $b \in [q]$ and $p \in \mathbb{N}$ define

$$S'_{x,\vec{v},j,h,b,p} := \sum_{y \in E_b^{x,\vec{v},h} \cap \Xi_{j,h}} S'_{x,\vec{v},y,b,p} \quad \text{and} \quad S''_{x,\vec{v},j,h,b,p} := \sum_{y \in E_b^{x,\vec{v},h} \cap \Xi_{j,h}} S''_{x,\vec{v},y,b,p}.$$

Observe that $R_{x,\vec{v},j,h,q,1} = \sum_{y \in E_q^{x,\vec{v},h} \cap \Xi_{j,h}} \frac{|\mathcal{C}_{\tau(y)-1}(y) \cap U|}{|\mathcal{C}_{\tau(y)-1}(y)|}$ and for each $b \in [q]_0$ we have $R_{x,\vec{v},j,h,b,1} = (1 \pm (\Delta^2+1)\eta) S_{x,\vec{v},j,h,b,1}$.

Let $\mathcal{E}'$ be the event where for each $j \in J$, $h \in [\ell_j]$, $a \in [n^{(3)}_{j,h}]$, $x \in X_j$ and tuple $\vec{v} = (\psi(z))_{z \in \mathrm{Dom}(\psi)}$ of vertices in $\mathcal{G}'$ for some partial partite homomorphism ψ from $H^{-1}(x)$ to $\mathcal{G}'$ such that for all $y \in X_{j,h}$ we have that $\mathcal{G}^{x,\vec{v},y}$ is an (η, c)-THC graph with the linear order on $N^{\leq 2}(y)$ induced by τ and density weighted hypergraph $\mathcal{D}^{x,\vec{v},y}$, we have $\left|\mathrm{exc}^{x,\vec{v}}_{h,a}\right| \leq \rho |X_j|$. For $j \in J$, $h \in [\ell_j]$, $x \in X_j$, a tuple $\vec{v} = (\psi(z))_{z \in \mathrm{Dom}(\psi)}$ of vertices in $\mathcal{G}'$ for some partial partite homomorphism ψ from $H^{-1}(x)$ to $\mathcal{G}'$ such that for all $y \in \Xi_{j,h}$ we have that $\mathcal{G}^{x,\vec{v},y}$ is an (η, c)-THC graph with the linear order on $N^{\leq 2}(y)$ induced by τ and density weighted hypergraph $\mathcal{D}^{x,\vec{v},y}$, and $\ell \in [n^{(2)}_{j,h}]_0$, set $\mathcal{E}_{j,x,\vec{v},h,\ell}$ to be the event that $\mathcal{E}'$ holds, and that the following hold for all $b \in [\ell]_0$ and $p \in [\Delta^2 - b]_0$.

(EQ1) $S_{x,\vec{v},j,h,b,2^p} \leq 2^{4b+1} \frac{a_b \rho |\mathcal{C}_{\vec{v}}(x)|^{2^p}}{\mu^b |V_j|^{2^p-1}}$.

Let $\mathcal{E}$ be the event that $\mathcal{E}_{j,x,\vec{v},h,n^{(2)}_{j,h}}$ holds for each $j \in J$, $h \in [\ell_j]$, $x \in X_j$ and tuple $\vec{v} = (\psi(z))_{z \in \mathrm{Dom}(\psi)}$ of vertices in $\mathcal{G}'$ for some partial partite homomorphism ψ from $H^{-1}(x)$ to $\mathcal{G}'$ which is THC-extendable for $(\mathcal{G}', H, \tau)$ such that $\mathcal{G}'_{\vec{v}}(x) \geq \left(\frac{10\kappa|J|\mu\Delta \log n}{\rho n}\right)^{1/2^{\Delta^2-1}}$ and for all $y \in \Xi_{j,h}$ we have that $\mathcal{G}^{x,\vec{v},y}$ is an (η, c)-THC graph with the linear order on $N^{\leq 2}(y)$ induced by τ and density weighted hypergraph $\mathcal{D}^{x,\vec{v},y}$. The following lemma states that $\mathcal{E}$ holds asymptotically almost surely.

Lemma 4.48. *Assume Setup 4.37. Let $c \geq 2^{\Delta^2+2}$ and $\mu, \rho, \varepsilon_1, \varepsilon_2, \varepsilon_3, \eta \in (0,1]$ satisfy*

$$4(\Delta\eta)^{1/7} \leq \varepsilon_1, \quad 4(\Delta^2\eta)^{1/7} \leq \varepsilon_2 \leq \left(\frac{\mu\rho}{2^{\Delta^3+4}}\right)^{2^{\Delta^2}} c^{-1/7}, \quad 2^{\Delta^2/7+2}\eta^{1/7} \leq \varepsilon_3.$$

Suppose that $\mathcal{G}$ is an (η, c)-THC graph for (H, τ) with density weighted hypergraph $\mathcal{D}$ and that for some integer T we have a sequence $\phi_0, \dots, \phi_T$ of THC-respecting partial partite homomorphisms from H to $\mathcal{G}'$, where ϕ_0 is the trivial partial partite homomorphism and each ϕ_t is obtained from ϕ_{t-1} by embedding the first vertex $z \in V(H) \setminus \mathrm{Dom}(\phi_{t-1})$ according to the linear order induced by τ to a uniform random vertex from a subset $S_{\phi_{t-1},z}$ of $\mathcal{C}_{t-1}(z)$ with $|S_{\phi_{t-1},z}| \geq \frac{1}{8}\mu|\mathcal{C}_{t-1}(z)|$. Then

$$\mathbb{P}(\mathcal{E}) \geq 1 - 2^{2^{\Delta_R^3+1}} \left(\Delta n^{\Delta+1} e^{-\frac{\rho n}{8\kappa|J|}} + \Delta^5 n^{1-9\Delta}\right).$$

In particular, $\mathcal{E}$ holds asymptotically almost surely.

The rest of this section is devoted to proving Lemma 4.48. Our goal is to show that the event $\mathcal{E}$ holds asympototically almost surely, thereby establishing that relevant sums of quantities related to certain conditional probabilities are reasonably well-behaved. Our approach involves obtaining certain initial counts of complexes and analysing the evolution of these counts as we embed vertices. The following lemma enables us to control the one-step evolution of these counts for each vertex of interest.

Lemma 4.49. *Assume Setup 4.37. Suppose that for some integer T we have a sequence $\phi_0, \ldots, \phi_T$ of partial partite homomorphisms from H to $\mathcal{G}'$, where ϕ_0 is the trivial partial partite homomorphism and each ϕ_t is obtained from ϕ_{t-1} by embedding the first vertex $z \in V(H) \setminus \mathrm{Dom}(\phi_{t-1})$ according to the linear order induced by τ to a uniform random vertex from a subset $S_{\phi_{t-1},z}$ of $\mathcal{C}_{t-1}(z)$. Let $j \in J$, $h \in [\ell_j]$, $x \in X_j$, $\vec{v} = (\psi(z))_{z \in \mathrm{Dom}(\psi)}$ be a tuple of vertices in $\mathcal{G}'$ for some partial partite homomorphism ψ from $H^{-1}(x)$ to $\mathcal{G}'$, $b \in [n^{(2)}_{j,h}]$ and $y \in \Xi_{j,h}$. Then for each $p \in [\Delta^2 - b]_0$ we have*

$$\mathbb{E}\left[S_{x,\vec{v},y,b,2^p} | \mathcal{F}_{\tau(u_{y,b})-1}\right] \leq \frac{d'_{\tau(u_{y,b})-1}(u_{y,b})|V_{u_{y,b}}|S'_{x,\vec{v},y,b,2^p}}{\left|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}\right|}, \tag{4.16}$$

and if $y \in \Xi_{j,h} \setminus E^{x,\vec{v},h}_{b-}$ we also have

$$\mathbb{E}\left[I_{\left\{y \in \mathrm{exc}^{x,\vec{v}}_{h,a_b}\right\}} S_{x,\vec{v},y,b,2^p} | \mathcal{F}_{\tau(u_{y,b})-1}\right] \leq \frac{d'_{\tau(u_{y,b})-1}(u_{y,b})|V_{u_{y,b}}|S''_{x,\vec{v},y,b,2^p}}{\left|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}\right|}. \tag{4.17}$$

Proof. Let $U := \mathcal{C}_{\vec{v}}(x)$. For $u \in \mathcal{C}_{\tau(u_{y,b})-1}(u_{y,b})$ write $f_{x,\vec{v},y,b,2^p}(u)$ for the number of candidates for $F'_{y,b,2^p}$ at time $\tau(u_{y,b})-1$ such that the copies of y would be embedded to U and $u_{y,b}$ would be embedded to u. Since $\mathrm{num}(S'_{x,\vec{v},y,b,2^p})$ is the number of candidates for $F'_{y,b,2^p}$ at time $\tau(u_{y,b})-1$ such that the copies of y would be embedded to U and $u_{y,b}$ would be embedded into $S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}$ and $\mathrm{num}(S''_{x,\vec{v},y,b,2^p})$ is the number of candidates for $F'_{y,b,2^p}$ at time $\tau(u_{y,b})-1$ such that the copies of y would be embedded to U and $u_{y,b}$ would be embedded into $W^{x,\vec{v},y}_{\phi_{\tau(u_{y,b})-1},u_{y,b}}$, we have

$$\mathrm{num}(S'_{x,\vec{v},y,b,2^p}) = \sum_{u \in S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}} f_{x,\vec{v},y,b,2^p}(u), \text{ and}$$
$$\mathrm{num}(S''_{x,\vec{v},y,b,2^p}) = \sum_{u \in W^{x,\vec{v},y}_{\phi_{\tau(u_{y,b})-1},u_{y,b}}} f_{x,\vec{v},y,b,2^p}(u).$$

Since $\mathrm{num}(S_{x,\vec{v},y,b,2^p})$ is the number of candidates for $F_{y,b,2^p}$ at time $\tau(u_{y,b})$ such that the copies of y would be embedded to U and at time $\tau(u_{y,b})$ we embed $u_{y,b}$ to a uniform random vertex from a subset $S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}$ of $\mathcal{C}_{\tau(u_{y,b})-1}(u_{y,b})$, we obtain

$$\begin{aligned}\mathbb{E}[\mathrm{num}(S_{x,\vec{v},y,b,2^p})|\mathcal{F}_{\tau(u_{y,b})-1}] &\leq \frac{\sum_{u\in S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}} f_{x,\vec{v},y,b,2^p}(u)}{|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|} \\ &= \frac{\mathrm{num}(S'_{x,\vec{v},y,b,2^p})}{|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|}\end{aligned}$$

and if $y \in \Xi_{j,h} \setminus E^{x,\vec{v},h}_{b-}$ we also have

$$\begin{aligned}&\mathbb{E}\left[I_{\left\{y\in \mathrm{exc}^{x,\vec{v}}_{h,a_b}\right\}}\mathrm{num}(S_{x,\vec{v},y,b,2^p})|\mathcal{F}_{\tau(u_{y,b})-1}\right] \\ &\leq \frac{\sum_{u\in W^{x,\vec{v},y}_{\phi_{\tau(u_{y,b})-1},u_{y,b}}} f_{x,\vec{v},y,b,2^p}(u)}{|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|} = \frac{\mathrm{num}(S''_{x,\vec{v},y,b,2^p})}{|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|}.\end{aligned}$$

Since $\mathrm{den}(S'_{x,\vec{v},y,b,2^p}) = \mathrm{den}(S_{x,\vec{v},y,b,2^p})|V_{u_{y,b}}|d'_{\tau(u_{y,b})-1}(u_{y,b})$, we obtain

$$\begin{aligned}\mathbb{E}[S_{x,\vec{v},y,b,2^p}|\mathcal{F}_{\tau(u_{y,b})-1}] &\leq \frac{d'_{\tau(u_{y,b})-1}(u_{y,b})|V_{u_{y,b}}|\,\mathrm{num}(S'_{x,\vec{v},y,b,2^p})}{|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|\,\mathrm{den}(S'_{x,\vec{v},y,b,2^p})} \\ &\leq \frac{d'_{\tau(u_{y,b})-1}(u_{y,b})|V_{u_{y,b}}|S'_{x,\vec{v},y,b,2^p}}{|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|}\end{aligned}$$

and if $y \in \Xi_{j,h} \setminus E^{x,\vec{v},h}_{b-}$ we also have

$$\begin{aligned}&\mathbb{E}\left[I_{\left\{y\in \mathrm{exc}^{x,\vec{v}}_{h,a_b}\right\}}S_{x,\vec{v},y,b,2^p}|\mathcal{F}_{\tau(u_{y,b})-1}\right] \\ &\leq \frac{d'_{\tau(u_{y,b})-1}(u_{y,b})|V_{u_{y,b}}|\,\mathrm{num}(S''_{x,\vec{v},y,b,2^p})}{|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|\,\mathrm{den}(S''_{x,\vec{v},y,b,2^p})} \\ &\leq \frac{d'_{\tau(u_{y,b})-1}(u_{y,b})|V_{u_{y,b}}|S''_{x,\vec{v},y,b,2^p}}{|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|}\end{aligned}$$

as desired. □

The following lemma tells us that the relevant sum of conditional expectations which regulates the one-step evolution of the relevant counts remains well-behaved.

Lemma 4.50. *Assume Setup 4.37. Let $c \geq 2^{\Delta^2+2}$ and $\varepsilon_1, \varepsilon_2, \varepsilon_3, \eta \in (0,1]$ satisfy*

$$4(\Delta\eta)^{1/7} \leq \varepsilon_1, \quad 4(\Delta^2\eta)^{1/7} \leq \varepsilon_2 \leq 2^{-2^{\Delta^2}(\Delta^3+1)}c^{-1/7}, \quad 2^{\Delta^2/7+2}\eta^{1/7} \leq \varepsilon_3.$$

*Suppose that $\mathcal{G}$ is an (η, c)-*THC *graph for (H, τ) with density weighted hypergraph $\mathcal{D}$ and that for some integer T we have a sequence $\phi_0, \dots, \phi_T$ of* THC*-respecting partial partite homomorphisms from H to $\mathcal{G}'$, where ϕ_0 is the trivial partial partite homomorphism and each ϕ_t is obtained from ϕ_{t-1} by embedding the first vertex $z \in V(H) \setminus \mathrm{Dom}(\phi_{t-1})$ according to the linear order induced by τ to a uniform random vertex from a subset $S_{\phi_{t-1},z}$ of $\mathcal{C}_{t-1}(z)$ with $|S_{\phi_{t-1},z}| \geq \frac{1}{8}\mu|\mathcal{C}_{t-1}(z)|$. Let $j \in J$, $h \in [\ell_j]$, $x \in X_j$, $b \in [n^{(2)}_{j,h}]$ and $\vec{v} = (\psi(z))_{z\in\mathrm{Dom}(\psi)}$ be a tuple of vertices in $\mathcal{G}'$ for some partial partite homomorphism ψ from $H^{-1}(x)$ to $\mathcal{G}'$ which is* THC*-extendable for $(\mathcal{G}', H, \tau)$ such that for all $y \in \Xi_{j,h}$ we have that $\mathcal{G}^{x,\vec{v},y}$ is an (η, c)-*THC *graph with the linear order on $N^{\leq 2}(y)$ induced by τ and density weighted hypergraph $\mathcal{D}^{x,\vec{v},y}$. Suppose that $\mathcal{E}_{x,\vec{v},j,h,b-1}$ holds. Then for each $p \in [\Delta^2 - b]_0$ we have*

$$\sum_{y\in\Xi_{j,h}} \mathbb{E}\left[I_{\left\{y\in E_b^{x,\vec{v},h}\right\}} S_{x,\vec{v},y,b,2^p} | \mathcal{F}_{\tau(u_{y,b})-1}\right] \leq (1+\varepsilon_1)\frac{2^{4b}a_b\rho|\mathcal{C}_{\vec{v}}(x)|^{2^p}}{\mu^b|V_j|^{2^p-1}}.$$

Proof. Since $I_{\left\{y\in E_b^{x,\vec{v},h}\right\}} = I_{\left\{y\in E_{b-}^{x,\vec{v},h}\right\}} + \left(1 - I_{\left\{y\in E_{b-}^{x,\vec{v},h}\right\}}\right) I_{\left\{y\in \mathrm{exc}_{h,a_b}^{x,\vec{v}}\right\}}$ and $I_{\left\{y\in E_{b-}^{x,\vec{v},h}\right\}}$ is $\mathcal{F}_{\tau(u_{y,b})-1}$-measurable, we have

$$\begin{aligned} &\sum_{y\in\Xi_{j,h}} \mathbb{E}\left[I_{\left\{y\in E_b^{x,\vec{v},h}\right\}} S_{x,\vec{v},y,b,2^p} | \mathcal{F}_{\tau(u_{y,b})-1}\right] \\ &= \sum_{y\in\Xi_{j,h}\cap E_{b-}^{x,\vec{v},h}} \mathbb{E}\left[S_{x,\vec{v},y,b,2^p} | \mathcal{F}_{\tau(u_{y,b})-1}\right] \\ &\quad + \sum_{y\in\Xi_{j,h}\setminus E_{b-}^{x,\vec{v},h}} \mathbb{E}\left[I_{\left\{y\in \mathrm{exc}_{h,a_b}^{x,\vec{v}}\right\}} S_{x,\vec{v},y,b,2^p} | \mathcal{F}_{\tau(u_{y,b})-1}\right]. \end{aligned} \tag{4.18}$$

We shall first prove the following claim.

Claim 4.51. *Let $y \in \Xi_{j,h} \setminus E_{b-1}^{x,\vec{v},h}$ and z be the bth vertex in the ordering of $N^{<2}(y)$ according to τ. Then the restriction ψ of $\phi_{\tau(z)-1}$ to $N^{\leq 2}(y)$ is a* THC*-respecting partial partite homomorphism from $H^{\leq 2}(y)$ to $\mathcal{G}^{x,\vec{v},y}$.*

Proof. Since $y \in \Xi_{j,h} \setminus E_{b-1}^{x,\vec{v},h}$, no vertex in $\mathrm{Dom}(\psi)$ is a trigger for y under $(x, \vec{v})$. Then, by Lemma 4.45 we have $|T_{\psi,z}^{x,\vec{v},y}| \geq (1-\delta_{y,z})|\mathcal{C}_\psi(H^z)| > 0$. By the definition of

$T^{x,\vec{v},y}_{\psi,z}$, the concatenation of $(\psi(w))_{w\in \mathrm{Dom}(\psi)}$ with any $\vec{u} \in T^{x,\vec{v},y}_{\psi,z}(\psi(z))$ (which must exist) produces a tuple in $T^{x,\vec{v},y}$; by the definition of $T^{x,\vec{v},y}$, it follows that ψ is a THC-respecting partial partite homomorphism from $H^{\leq 2}(y)$ to $\mathcal{G}^{x,\vec{v},y}$. ■

Let $U := \mathcal{C}_{\vec{v}}(x)$. Suppose that $\mathcal{E}_{j,x,\vec{v},h,b-1}$ holds. We first consider when $u_{y,b} \in N^{-1}(y)$ for all $y \in \Xi_{j,h}$. In this case we have $F_{y,b-1,2^p} = F'_{y,b,2^p}$, so we have $S_{x,\vec{v},y,b-1,2^p} = S'_{x,\vec{v},y,b,2^p}$. For each $y \in \Xi_{j,h} \cap E^{x,\vec{v},h}_{b-1}$ we have

$$|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}| \geq \tfrac{1}{8}\mu|\mathcal{C}_{\tau(u_{y,b})-1}(u_{y,b})| \geq \tfrac{1-\eta}{8}\mu d'_{\tau(u_{y,b})-1}(u_{y,b})|V_{u_{y,b}}|.$$

Then, by the lower bound on $S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}$, Lemma 4.49, summing over $y \in \Xi_{j,h}\cap E^{x,\vec{v},h}_{b-1}$, and (EQ1), we obtain

$$\begin{aligned}\sum_{y\in\Xi_{j,h}\cap E^{x,\vec{v},h}_{b-1}} &\mathbb{E}[S_{x,\vec{v},y,b,2^p}|\mathcal{F}_{\tau(u_{y,b})-1}] \\ \leq \sum_{y\in\Xi_{j,h}\cap E^{x,\vec{v},h}_{b-1}} &\frac{d'_{\tau(u_{y,b})-1}(u_{y,b})|V_{u_{y,b}}|S'_{x,\vec{v},y,b,2^p}}{|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|} \\ \leq \frac{8S_{x,\vec{v},j,h,b-1,2^p}}{(1-\eta)\mu} &\leq \frac{2^{4b}a_{b-1}\rho|U|^{2^p}}{(1-\eta)\mu^b|V_j|^{2^p-1}}.\end{aligned} \tag{4.19}$$

Let $y \in \Xi_{j,h} \setminus E^{x,\vec{v},h}_{b-1}$. By Claim 4.51 the restriction ψ of $\phi_{\tau(u_{y,b})-1}$ to $N^{\leq 2}(y)$ is a THC-respecting partial partite homomorphism from $H^{\leq 2}(y)$ to $\mathcal{G}^{x,\vec{v},y}$, so by Lemma 4.28 and the lower bound on $S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}$ we have

$$\begin{aligned}\frac{S'_{x,\vec{v},y,b,2^p}}{|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|} &\leq (1+\varepsilon_3)\frac{S_{x,\vec{v},y,b-1,2^p}}{|\mathcal{C}_{\tau(u_{y,b})-1}(u_{y,b})|} \\ &\leq \frac{(1+\varepsilon_3)\left(1+2^{\Delta^2+2}\eta\right)|U|^{2^p}}{(1-\eta)d'_{\tau(u_{y,b})-1}(u_{y,b})|V_{u_{y,b}}||V_j|^{2^p}}.\end{aligned}$$

If $y \in \Xi_{j,h} \setminus E^{x,\vec{v},h}_{b-}$, then by the lower bound on $S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}$ and Lemmas 4.28 and 4.47 we also have

$$\begin{aligned}\frac{S''_{x,\vec{v},y,b,2^p}}{|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|} &\leq (1+\varepsilon_3)\frac{\rho S_{x,\vec{v},y,b-1,2^p}}{2|\mathcal{C}_{\tau(u_{y,b})-1}(u_{y,b})|} \\ &\leq \frac{(1+\varepsilon_3)\left(1+2^{\Delta^2+2}\eta\right)\rho|U|^{2^p}}{(1-\eta)2d'_{\tau(u_{y,b})-1}(u_{y,b})|V_{u_{y,b}}||V_j|^{2^p}}.\end{aligned}$$

Now by Lemma 4.49, summing over $y \in \Xi_{j,h} \cap E_{b-}^{x,\vec{v},h} \setminus E_{b-1}^{x,\vec{v},h}$ and noting that $|\operatorname{exc}_{h,a}^{x,\vec{v}}| \leq \rho|V_j|$ for all $a_{b-1} < a < a_b$, we obtain

$$\begin{aligned}
&\sum_{y \in \Xi_{j,h} \cap E_{b-}^{x,\vec{v},h} \setminus E_{b-1}^{x,\vec{v},h}} \mathbb{E}[S_{x,\vec{v},y,b,2^p} | \mathcal{F}_{\tau(u_{y,b})-1}] \\
&\leq \sum_{y \in \Xi_{j,h} \cap E_{b-}^{x,\vec{v},h} \setminus E_{b-1}^{x,\vec{v},h}} \frac{d'_{\tau(u_{y,b})-1}(u_{y,b}) |V_{u_{y,b}}| S'_{x,\vec{v},y,b,2^p}}{|S_{\phi_{\tau(u_{y,b})-1}, u_{y,b}}|} \\
&\leq \frac{(1+\varepsilon_3)\left(1+2^{\Delta^2+2}\eta\right)(a_b - a_{b-1} - 1)\rho|U|^{2^p}}{(1-\eta)|V_j|^{2^p-1}}.
\end{aligned} \tag{4.20}$$

By Lemma 4.49 and summing over $y \in \Xi_{j,h} \setminus E_{b-}^{x,\vec{v},h}$, we obtain

$$\begin{aligned}
&\sum_{y \in \Xi_{j,h} \setminus E_{b-}^{x,\vec{v},h}} \mathbb{E}\left[I_{\left\{y \in \operatorname{exc}_{h,a_b}^{x,\vec{v}}\right\}} S_{x,\vec{v},y,b,2^p} | \mathcal{F}_{\tau(u_{y,b})-1} \right] \\
&\leq \sum_{y \in \Xi_{j,h} \setminus E_{b-}^{x,\vec{v},h}} \frac{d'_{\tau(u_{y,b})-1}(u_{y,b}) |V_{u_{y,b}}| S''_{x,\vec{v},y,b,2^p}}{|S_{\phi_{\tau(u_{y,b})-1}, u_{y,b}}|} \\
&\leq \frac{(1+\varepsilon_3)\left(1+2^{\Delta^2+2}\eta\right)\rho|U|^{2^p}}{(1-\eta)2|V_j|^{2^p-1}}.
\end{aligned} \tag{4.21}$$

Putting together (4.18)–(4.21), we obtain

$$\sum_{y \in \Xi_{j,h}} \mathbb{E}\left[I_{\left\{y \in E_b^{x,\vec{v},h}\right\}} S_{x,\vec{v},y,b,2^p} | \mathcal{F}_{\tau(u_{y,b})-1} \right] \leq \frac{2^{4b} a_b \rho |U|^{2^p}}{(1-\eta)\mu^b |V_j|^{2^p-1}}.$$

Otherwise, we have $u_{y,b} \in N^{-2}(y)$ for all $y \in \Xi_{j,h}$. We first prove the following claim.

Claim 4.52. *For each $y \in \Xi_{j,h}$ we have*

$$S'_{x,\vec{v},y,b,2^p} \leq \frac{(1+\frac{\varepsilon_1}{2})|S_{\phi_{\tau(u_{y,b})-1}, u_{y,b}}|}{|\mathcal{C}_{\tau(u_{y,b})-1}(u_{y,b})|} \sqrt{S_{x,\vec{v},y,b-1,2^{p+1}}},$$

and for each $y \in \Xi_{j,h} \setminus E_{b-}^{x,\vec{v},h}$ we have

$$S''_{x,\vec{v},y,b,2^p} \leq \frac{(1+\frac{\varepsilon_1}{2})|W_{\phi_{\tau(u_{y,b})-1}, u_{y,b}}|}{|\mathcal{C}_{\tau(u_{y,b})-1}(u_{y,b})|} \sqrt{S_{x,\vec{v},y,b-1,2^{p+1}}}.$$

Proof. Apply Lemma 4.14 with the following. $H = F'_{y,b,2^p}$, $A_1 = \{u_{y,b}\}$, $B = U_{y,b-1,>}$, $A_2 = \{y\}$, $V_y = U$, $V_{u_{y,b}} = S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}$ and $V_u = V'_u$ for $u \in B$. Let $\vec{a} := (2,1)$, $\vec{A} := (A_1, B)$, $F = H(\vec{A}, \vec{a})$. We obtain

$$\mathrm{num}(S'_{x,\vec{v},y,b,2^p}) \leq \sqrt{\left|\mathcal{C}_{\tau(u_{y,b})-1}(F) \cap \left(S^2_{\phi_{\tau(u_{y,b})-1},u_{y,b}} \times V'_{U_{y,b-1,>}}\right)\right|} \times \sqrt{\mathrm{num}(S_{x,\vec{v},y,b-1,2^{p+1}})}.$$

If $y \in \Xi_{j,h} \setminus E^{x,\vec{v},h}_{b-}$, then an analogous application of Lemma 4.14, with $V_{u_{y,b}} = W_{\phi_{\tau(u_{y,b})-1},u_{y,b}}$ instead, gives us

$$\mathrm{num}(S''_{x,\vec{v},y,b,2^p}) \leq \sqrt{\left|\mathcal{C}_{\tau(u_{y,b})-1}(F) \cap \left(\left(W^{x,\vec{v},y}_{\phi_{\tau(u_{y,b})-1},u_{y,b}}\right)^2 \times V'_{U_{y,b-1,>}}\right)\right|} \times \sqrt{\mathrm{num}(S_{x,\vec{v},y,b-1,2^{p+1}})}.$$

Now $\phi_{\tau(u_{y,b})-1}$ is a THC-respecting partial partite homomorphism, so by applying Lemma 4.28 for F and (THC1) for $\mathcal{G}'_{\tau(u_{y,b})-1}$ we have

$$\begin{aligned}
&|\mathcal{C}_{\tau(u_{y,b})-1}(F) \cap (S^2_{\phi_{\tau(u_{y,b})-1},u_{y,b}} \times V'_{U_{y,b-1,>}})| \\
&\leq (1+\varepsilon_1)\frac{|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|^2|\mathcal{C}_{\tau(u_{y,b})-1}(F)|}{|\mathcal{C}_{\tau(u_{y,b})-1}(u_{y,b})|^2} \\
&\leq \frac{(1+\frac{\varepsilon_1}{2})^2|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|^2}{|\mathcal{C}_{\tau(u_{y,b})-1}(u_{y,b})|^2}\frac{\left(\mathrm{den}(S'_{x,\vec{v},y,b,2^p})\right)^2}{\mathrm{den}(S_{x,\vec{v},y,b-1,2^{p+1}})}.
\end{aligned}$$

If $y \in \Xi_{j,h} \setminus E^{x,\vec{v},h}_{b-}$, then an analogous argument gives us

$$\begin{aligned}
&|\mathcal{C}_{\tau(u_{y,b})-1}(F) \cap (W^2_{\phi_{\tau(u_{y,b})-1},u_{y,b}} \times V_{U_{y,b-1,>}})| \\
&\leq (1+\varepsilon_1)\frac{|W_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|^2|\mathcal{C}_{\tau(u_{y,b})-1}(F)|}{|\mathcal{C}_{\tau(u_{y,b})-1}(u_{y,b})|^2} \\
&\leq \frac{(1+\frac{\varepsilon_1}{2})^2|W_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|^2}{|\mathcal{C}_{\tau(u_{y,b})-1}(u_{y,b})|^2}\frac{\left(\mathrm{den}(S'_{x,\vec{v},y,b,2^p})\right)^2}{\mathrm{den}(S_{x,\vec{v},y,b-1,2^{p+1}})}.
\end{aligned}$$

Hence, we have

$$S'_{x,\vec{v},y,b,2^p} \leq \frac{(1+\frac{\varepsilon_1}{2})|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|}{|\mathcal{C}_{\tau(u_{y,b})-1}(u_{y,b})|}\sqrt{S_{x,\vec{v},y,b-1,2^{p+1}}},$$

and if $y \in \Xi_{j,h} \setminus E^{x,\vec{v},h}_{b-}$ we also have

$$S''_{x,\vec{v},y,b,2^p} \leq \frac{(1+\frac{\varepsilon_1}{2})|W_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|}{|\mathcal{C}_{\tau(u_{y,b})-1}(u_{y,b})|}\sqrt{S_{x,\vec{v},y,b-1,2^{p+1}}}$$

as required. ∎

Now by applying Lemma 4.49, Claim 4.52, the lower bound on $S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}$ and Lemma 4.47, we obtain

$$\begin{aligned}&\sum_{y\in\Xi_{j,h}\cap E^{x,\vec{v},h}_{b-}} \mathbb{E}\left[S_{x,\vec{v},y,b,2^p}|\mathcal{F}_{\tau(u_{y,b})-1}\right]\\ &\leq \frac{(1+\frac{\varepsilon_1}{2})}{1-\eta}\sum_{y\in\Xi_{j,h}\cap E^{x,\vec{v},h}_{b-}}\sqrt{S_{x,\vec{v},y,b-1,2^{p+1}}}\end{aligned}$$

and

$$\begin{aligned}&\sum_{y\in\Xi_{j,h}\setminus E^{x,\vec{v},h}_{b-}} \mathbb{E}\left[I_{\left\{y\in\mathrm{exc}^{x,\vec{v}}_{h,a_b}\right\}}S_{x,\vec{v},y,b,2^p}|\mathcal{F}_{\tau(u_{y,b})-1}\right]\\ &\leq \frac{(1+\frac{\varepsilon_1}{2})}{1-\eta}\sum_{y\in\Xi_{j,h}\setminus E^{x,\vec{v},h}_{b-}}\frac{\rho}{2}\sqrt{S_{x,\vec{v},y,b-1,2^{p+1}}}.\end{aligned}$$

By Lemma 4.13 applied with $\alpha_y = \sqrt{S_{x,\vec{v},y,b-1,2^{p+1}}}$ and $\beta_y = 1$ for $y \in E^{x,\vec{v},h}_{b-1}$ and (EQ1), we obtain

$$\sum_{y\in E^{x,\vec{v},h}_{b-1}}\sqrt{S_{x,\vec{v},y,b-1,2^{p+1}}} \leq \sqrt{a_{b-1}\rho|V_j|S_{x,\vec{v},j,h,b-1,2^{p+1}}} \leq \frac{2^{4b}a_{b-1}\rho|U|^{2^p}}{\mu^b|V_j|^{2^p-1}}.$$

Let $y \in \Xi_{j,h} \setminus E^{x,\vec{v},h}_{b-1}$. By Claim 4.51 the restriction ψ of $\phi_{\tau(u_{y,b})-1}$ to $N^{\leq 2}(y)$ is a THC-respecting partial partite homomorphism from $H^{\leq 2}(y)$ to $\mathcal{G}^{x,\vec{v},y}$. In particular, $\mathcal{G}^{x,\vec{v},y}_{\psi}$ is an (η, c)-THC graph with the linear order on $N^{\leq 2}(y) \setminus \mathrm{Dom}(\psi)$ induced by τ and density weighted hypergraph $\mathcal{D}^{x,\vec{v},y}_{\psi}$, so we have $\sqrt{S_{x,\vec{v},y,b-1,2^{p+1}}} \leq \left(1+2^{\Delta^2+2}\eta\right)\frac{|U|^{2^p}}{|V_j|^{2^p}}$. By summing over $y \in \Xi_{j,h} \cap E^{x,\vec{v},h}_{b-} \setminus E^{x,\vec{v},h}_{b-1}$ and noting that $|\mathrm{exc}^{x,\vec{v}}_{h,a}| \leq \rho|V_j|$ for all $a_{b-1} < a < a_b$, we obtain

$$\sum_{y\in\Xi_{j,h}\cap E^{x,\vec{v},h}_{b-}\setminus E^{x,\vec{v},h}_{b-1}}\sqrt{S_{x,\vec{v},y,b-1,2^{p+1}}} \leq \left(1+2^{\Delta^2+2}\eta\right)\frac{(a_b-a_{b-1}-1)\rho|U|^{2^p}}{|V_j|^{2^p-1}},$$

and by summing over $y \in \Xi_{j,h} \setminus E_{b-}^{x,\vec{v},h}$ we obtain

$$\sum_{y \in \Xi_{j,h} \setminus E_{b-}^{x,\vec{v},h}} \frac{\rho}{2}\sqrt{S_{x,\vec{v},y,b-1,2^{p+1}}} \leq \left(1 + 2^{\Delta^2+2}\eta\right) \frac{\rho|U|^{2^p}}{2|V_j|^{2^p-1}}.$$

Finally, putting together the previous five inequalities and (4.18), we obtain

$$\sum_{y \in \Xi_{j,h}} \mathbb{E}\left[I_{\left\{y \in E_b^{x,\vec{v},h}\right\}} S_{x,\vec{v},y,b,2^p} | \mathcal{F}_{\tau(u_{y,b})-1}\right] \leq \frac{(1+\frac{\varepsilon_1}{2})2^{4b} a_b \rho |U|^{2^p}}{(1-\eta)\mu^b |V_j|^{2^p-1}}$$

as desired. □

The following lemma establishes the probability of each subevent of our desired event.

Lemma 4.53. *Assume Setup 4.37. Let $c \geq 2^{\Delta^2+2}$ and $\varepsilon_1, \varepsilon_2, \varepsilon_3, \eta \in (0,1]$ satisfy*

$$4(\Delta\eta)^{1/7} \leq \varepsilon_1, \quad 4(\Delta^2\eta)^{1/7} \leq \varepsilon_2 \leq 2^{-2^{\Delta^2}(\Delta^3+1)} c^{-1/7}, \quad 2^{\Delta^2/7+2}\eta^{1/7} \leq \varepsilon_3.$$

*Suppose that $\mathcal{G}$ is an (η, c)-*THC *graph for (H, τ) with density weighted hypergraph $\mathcal{D}$ and that for some integer T we have a sequence $\phi_0, \ldots, \phi_T$ of* THC*-respecting partial partite homomorphisms from H to $\mathcal{G}'$, where ϕ_0 is the trivial partial partite homomorphism and each ϕ_t is obtained from ϕ_{t-1} by embedding the first vertex $z \in V(H) \setminus \mathrm{Dom}(\phi_{t-1})$ according to the linear order induced by τ to a uniform random vertex from a subset $S_{\phi_{t-1},z}$ of $\mathcal{C}_{t-1}(z)$ with $|S_{\phi_{t-1},z}| \geq \frac{1}{8}\mu|\mathcal{C}_{t-1}(z)|$. Let $j \in J$, $h \in [\ell_j]$, $x \in X_j$ and $\vec{v} = (\psi(z))_{z \in \mathrm{Dom}(\psi)}$ be a tuple of vertices in $\mathcal{G}'$ for some partial partite homomorphism ψ from $H^{-1}(x)$ to $\mathcal{G}'$ which is* THC*-extendable for $(\mathcal{G}', H, \tau)$ such that for all $y \in \Xi_{j,h}$ we have that $\mathcal{G}^{x,\vec{v},y}$ is an (η, c)-*THC *graph with the linear order on $N^{\leq 2}(y)$ induced by τ and density weighted hypergraph $\mathcal{D}^{x,\vec{v},y}$. Let $q := n_{j,h}^{(2)}$. Then for each $b \in [q]_0$ the event $\mathcal{E}_{j,x,\vec{v},h,b}$ holds with probability at least $\mathbb{P}(\mathcal{E}') - \sum_{\ell=1}^{b}\sum_{p=0}^{\Delta^2-\ell} \exp\left(-2^{4\ell-3}\frac{a_\ell \rho |\mathcal{C}_{\vec{v}}(x)|^{2^p}}{\mu^\ell |V_j|^{2^p-1}}\right)$. In particular, the event $\mathcal{E}_{j,x,\vec{v},h,q}$ holds with probability at least $\mathbb{P}(\mathcal{E}') - \Delta^4 \exp\left(-\frac{\rho|\mathcal{C}_{\vec{v}}(x)|^{2^{\Delta^2-1}}}{\mu|V_j|^{2^{\Delta^2-1}-1}}\right)$.*

Proof. Let $U := \mathcal{C}_{\vec{v}}(x)$. We proceed by induction on b. First consider $b = 0$. Note that $S_{x,\vec{v},j,h,0,2^p} = 0 \leq \frac{2a_0\rho|U|^{2^p}}{|V_j|^{2^p-1}}$ trivially holds for all $p \in [\Delta^2]_0$. Hence, we have $\mathbb{P}(\mathcal{E}_{j,x,\vec{v},h,0}) = \mathbb{P}(\mathcal{E}')$. Now consider $b \in [q]$.

Claim 4.54. *For each $p \in [\Delta^2 - b]_0$ we have*

$$\mathbb{P}\left(\mathcal{E}_{j,x,\vec{v},h,b-1} \text{ and } S_{x,\vec{v},j,h,b,2^p} > 2^{4b+1}\frac{a_b\rho|U|^{2^p}}{\mu^b|V_j|^{2^p-1}}\right)$$
$$\leq \exp\left(-2^{4b-3}\frac{a_b\rho|U|^{2^p}}{\mu^b|V_j|^{2^p-1}}\right).$$

Proof. Enumerate $\{u_{y,b} : y \in \Xi_{j,h}\}$ as $z_{b,1}, \ldots, z_{b,\xi_{j,h}}$ in the order according to τ. Set $\lambda := (1+\varepsilon_1)2^{4b}\frac{a_b\rho|U|^{2^p}}{\mu^b|V_j|^{2^p-1}}$. For $i \in [\xi_{j,h}]_0$ set $\mathcal{F}_i^{(b)}$ to be $\mathcal{F}_{\tau(z_{b,\xi_{j,h}})}$ if $i = \xi_{j,h}$ and $\mathcal{F}_{\tau(z_{b,i+1})-1}$ otherwise. For $i \in [\xi_{j,h}]$, with y satisfying $z_{b,i} = u_{y,b}$, set Y_i to be $S_{x,\vec{v},y,b,2^p}$ if $y \in \Xi_{j,h} \cap E_b^{x,\vec{v},h}$ and zero otherwise. Note that Y_i is $\mathcal{F}_i^{(b)}$-measurable and $0 \leq Y_i \leq R := 1 + (\Delta + 2^p)\eta$. Now when $\mathcal{E}_{j,x,\vec{v},h,b-1}$ holds, by Lemma 4.50 we have $\sum_{i\in[x_{j,h}]}\mathbb{E}\left[Y_i|\mathcal{F}_{i-1}^{(b)}\right] \leq \lambda$ and $\sum_{i\in[x_{j,h}]}\operatorname{var}\left(Y_i|\mathcal{F}_{i-1}^{(b)}\right) \leq R\lambda$. Then by Lemma 4.12 we have

$$\mathbb{P}\left(\mathcal{E}_{j,x,\vec{v},h,b-1} \text{ and } S_{x,\vec{v},j,h,b,2^p} > 2^{4b+1}\tfrac{a_b\rho|U|^{2^p}}{\mu^b|V_j|^{2^p-1}}\right) \leq \exp\left(-2^{4b-3}\tfrac{a_b\rho|U|^{2^p}}{\mu^b|V_j|^{2^p-1}}\right)$$

as desired. ∎

By the inductive hypothesis and Claim 4.54, we have

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{E}_{j,x,\vec{v},h,b}\right)\\
&= \mathbb{P}\left(\mathcal{E}_{j,x,\vec{v},h,b-1} \text{ and } S_{x,\vec{v},j,h,b,2^p} \leq 2^{4b+1}\tfrac{a_b\rho|U|^{2^p}}{\mu^b|V_j|^{2^p-1}} \text{ for all } p \in [\Delta^2-b]_0\right)\\
&\geq \mathbb{P}\left(\mathcal{E}_{j,x,\vec{v},h,b-1}\right) - \sum_{p=0}^{\Delta^2-b}\mathbb{P}\left(\mathcal{E}_{j,x,\vec{v},h,b-1} \text{ and } S_{x,\vec{v},j,h,b,2^p} > 2^{4b+1}\tfrac{a_b\rho|U|^{2^p}}{\mu^b|V_j|^{2^p-1}}\right)\\
&\geq \mathbb{P}(\mathcal{E}') - \sum_{\ell=1}^{b}\sum_{p=0}^{\Delta^2-\ell}\exp\left(-2^{4\ell-3}\tfrac{a_\ell\rho|U|^{2^p}}{\mu^\ell|V_j|^{2^p-1}}\right).
\end{aligned}$$

Finally, letting $b = q$ and noting that each of the at most Δ^4 summands in the sum is at most $\exp\left(-\frac{\rho|U|^{2^{\Delta^2-1}}}{\mu|V_j|^{2^{\Delta^2-1}-1}}\right)$, we obtain $\mathbb{P}\left(\mathcal{E}_{j,x,\vec{v},h,q}\right) \geq \mathbb{P}(\mathcal{E}') - \Delta^4\exp\left(-\frac{\rho|U|^{2^{\Delta^2-1}}}{\mu|V_j|^{2^{\Delta^2-1}-1}}\right)$. □

Finally, we provide a proof of Lemma 4.48.

Proof of Lemma 4.48. By Lemma 4.44 we find that

$$\mathbb{P}(\mathcal{E}') \geq 1 - 2^{2^{\Delta_R^3+1}}\Delta n^{\Delta+1}e^{-\frac{\rho n}{8\kappa|J|}}.$$

For each $j \in J$, $h \in [\ell_j]$ and $x \in X_j$ let $\vec{V}_{j,h,x}$ denote the set of all tuples $\vec{v} = (\psi(z))_{z \in \mathrm{Dom}(\psi)}$ of vertices in $\mathcal{G}'$ for some partial partite homomorphism ψ from $H^{-1}(x)$ to $\mathcal{G}'$ which is THC-extendable for $(\mathcal{G}', H, \tau)$ such that $\mathcal{G}'_{\vec{v}}(x) \geq \left(\frac{10\kappa|J|\mu\Delta\log n}{\rho n}\right)^{1/2^{\Delta^2-1}}$ and for all $y \in \Xi_{j,h}$ we have that $\mathcal{G}^{x,\vec{v},y}$ is an (η, c)-THC graph with the linear order on $N^{\leq 2}(y)$ induced by τ and density weighted hypergraph $\mathcal{D}^{x,\vec{v},y}$. Note that $|\vec{V}_{j,h,x}| \leq \Delta n^{\Delta}$. Let $j \in J$, $h \in [\ell_j]$, $x \in X_j$ and $\vec{v} \in \vec{V}_{j,h,x}$. By Lemma 4.53, the fact that $|V_j| \geq \frac{n}{\kappa|J|}$, and the condition on $\mathcal{G}'_{\vec{v}}(x)$, we have $\mathbb{P}(\mathcal{E}' \setminus \mathcal{E}_{j,x,\vec{v},h,n^{(2)}_{j,h}}) \leq \Delta^4 n^{-10\Delta}$. Hence, by summing over $j \in J$, $h \in [\ell_j]$, $x \in X_j$, $\vec{v} \in \vec{V}_{j,x}$, and applying (H8), we obtain

$$\begin{aligned}\mathbb{P}(\mathcal{E}) &\geq \mathbb{P}(\mathcal{E}') - \sum_{j \in J} \sum_{h \in [\ell_j]} \sum_{x \in X_j} \sum_{\vec{v} \in \vec{V}_{j,h,x}} \mathbb{P}(\mathcal{E}' \setminus \mathcal{E}_{j,x,\vec{v},h,n^{(2)}_{j,h}}) \\ &\geq 1 - 2^{2^{\Delta_R^3+1}} \left(\Delta n^{\Delta+1} e^{-\frac{\rho n}{8\kappa|J|}} + \Delta^5 n^{1-9\Delta}\right)\end{aligned}$$

as desired. Finally, the final expression tends to 1 as n tends to infinity so we obtain the desired conclusion. □

4.9 Buffer Embedding

Our random greedy algorithm will manage to embed the main part of each X_j into V_j; we now need to find a way to embed the carefully selected buffer vertices X_j^{buf}. Since the buffer vertices are pairwise far apart, their neighbours will have all been embedded and therefore their candidate sets will no longer change. As such, it will suffice to find a system of distinct representatives for the buffer vertices from their available candidate sets; we will do so by verifying Hall's condition in the auxiliary available candidate graph. Our analysis will consider three cases based on the size of the subset, two of which will be handled using methods from our previous analysis of the random greedy algorithm. The final case concerns almost spanning subsets of buffer vertices. Our method will require us to prove two key lemmas: firstly, that each v is a candidate for not too few buffer vertices, and secondly, that vertices in W are candidates for not too many buffer vertices for which v is a candidate. One difference in this analysis, compared to those from before, is that instead of considering candidates for a vertex x, we will consider vertices for which a vertex v is a candidate; this reversal of roles slightly complicates our analysis.

To show that each v is a candidate for not too few buffer vertices, we will establish a lower bound on the probability of each buffer vertex x having v as a candidate. This will involve showing that it is reasonably likely for the complex $H^{-1}(x)$ to be embedded into a suitable neighbourhood of v. However, a complication which arises from working with sparse structures here is that it is entirely possible for such a neighbourhood to become overly occupied by neighbours of buffer vertices and for there to be insufficient room left to obtain the desired outcome. We will show in Lemma 4.55 that these neighbourhoods do not become overly occupied and, through Lemma 4.61, that each buffer vertex is reasonably likely to have a vertex v as a candidate.

4.9.1 Bounding Occupancy by Initial Segment

We shall show that the sum of the probabilities of embedding neighbours of buffer vertices into a given vertex neighbourhood is reasonably small; a martingale concentration argument then implies that these vertices occupy only a small fraction of the vertex neighbourhood. To this end, we shall define a suitable event $\mathcal{E}^*$ and prove Lemma 4.55, which states that $\mathcal{E}^*$ holds asymptotically almost surely. Let us highlight that the argument presented in this subsection very much resembles the argument in Section 4.8.2.

The embedding behaviour of a vertex y is closely linked to its candidate set right before its embedding; to understand how this behaves, we shall study how the candidate set of the unembedded part of the neighbourhood complex $H^{\leq 1}(y)$ evolves as we embed $N^{<2}(y)$. We shall reuse notation introduced in Section 4.8.2 to describe the evolution of the unembedded part of the neighbourhood complex. Let us now provide definitions of useful objects and quantities. Assume Setup 4.37 and suppose that for some integer T we have a sequence $\phi_0, \dots, \phi_T$ of THC-respecting partial partite homomorphisms from H to $\mathcal{G}'$, where ϕ_0 is the trivial partial partite homomorphism and each ϕ_t is obtained from ϕ_{t-1} by embedding the first vertex $z \in V(H) \setminus \mathrm{Dom}(\phi_{t-1})$ to a uniform random vertex from a subset $S_{\phi_{t-1},z}$ of $\mathcal{C}_{t-1}(z)$. Let $j \in J$ and $h \in [\ell_j]$. Set $q := n^{(2)}_{j,h}$, $\Xi_{j,h} := N_{H^{(2)}}(X^{\mathrm{buf}}) \cap \{y \in X_{j,h} : T_y \leq T\}$, $\xi_{j,h} := |\Xi_{j,h}|$, $\Xi_j := \bigcup_{h\in[\ell_j]} \Xi_{j,h}$ and $\xi_j := |\Xi_j|$. Let $v \in V(H)$ and set $U := N_{\mathcal{G}^{(2)}}(v; V_j)$.

To understand the probability of embedding a vertex y into U, we shall study the evolution of the probability of randomly picking from all copies of $F_{y,b}$ a copy of $F_{y,b}$ with y going to a vertex in U; for this purpose we define the quantity $R_{v,y,b,p}$. In practice, we work with a different quantity $S_{v,y,b,p}$ obtained by replacing the denominator of

$R_{v,y,b,p}$ with its deterministic estimate (within a small relative error by THC). For $y \in \Xi_{j,h}$, $b \in [q]_0$ and $p \in \mathbb{N}$ define

$$\mathrm{num}(R_{v,y,b,p}), \mathrm{num}(S_{v,y,b,p}) := |\mathcal{C}_{\min(\tau(u_{y,b+1})-1,T)}(V(F_{y,b,p})) \cap (V_{U_{y,b,>}} \times U^p)|,$$

$$\mathrm{den}(R_{v,y,b,p}) := |\mathcal{C}_{\min(\tau(u_{y,b+1})-1,T)}(V(F_{y,b,p}))|,$$

$$\mathrm{den}(S_{v,y,b,p}) := \frac{\mathcal{D}_{\min(\tau(u_{y,b+1})-1,T)}(V(F_{y,b,p}))}{d_{\min(\tau(u_{y,b+1})-1,T)}(\varnothing)}|V_y|^p \prod_{u \in U_{y,b,>}} |V_u|,$$

$$R_{v,y,b,p} := \frac{\mathrm{num}(R_{v,y,b,p})}{\mathrm{den}(R_{v,y,b,p})} \quad \text{and} \quad S_{v,y,b,p} := \frac{\mathrm{num}(S_{v,y,b,p})}{\mathrm{den}(S_{v,y,b,p})}.$$

We need to describe the cumulative probabilities. For $b \in [q]_0$ and $p \in \mathbb{N}$ define

$$R_{v,j,h,b,p} := \sum_{y \in \Xi_{j,h}} R_{v,y,b,p} \quad \text{and} \quad S_{v,j,h,b,p} := \sum_{y \in \Xi_{j,h}} S_{v,y,b,p}.$$

We need notation to describe a related quantity for our stepwise updating argument. For $y \in \Xi_{j,h}$, $b \in [q]$ and $p \in \mathbb{N}$ define

$$\mathrm{num}(S'_{v,y,b,p}) := \left|\mathcal{C}_{\tau(u_{y,b})-1}(V(F'_{y,b,p})) \cap (V_{U_{y,b,>}} \times S_{\phi_{\tau(u_{y,b})-1}, u_{y,b}} \times U^p)\right|,$$

$$\mathrm{den}(S'_{v,y,b,p}) := \frac{\mathcal{D}_{\min(\tau(u_{y,b})-1,T)}(V(F'_{y,b,p}))}{d_{\min(\tau(u_{y,b})-1,T)}(\varnothing)}|V_y|^p \prod_{u \in \{u_{y,b}\} \cup U_{y,b-1,>}} |V_u|, \quad \text{and}$$

$$S'_{v,y,b,p} := \frac{\mathrm{num}(S'_{v,y,b,p})}{\mathrm{den}(S'_{v,y,b,p})}.$$

For $b \in [q]$ and $p \in \mathbb{N}$ define $S'_{v,j,h,b,p} := \sum_{y \in \Xi_{j,h}} S'_{v,y,b,p}$. Observe that we have $R_{v,j,h,q,1} = \sum_{y \in \Xi_{j,h}} \frac{|\mathcal{C}_{\tau(y)-1}(y) \cap U|}{|\mathcal{C}_{\tau(y)-1}(y)|}$ and we have $R_{v,j,h,b,1} = (1 \pm (\Delta^2+1)\eta) S_{v,j,h,b,1}$ for each $b \in [q]_0$.

For $j \in J$, $h \in [\ell_j]$, $v \in V(H)$ and $\ell \in [n^{(2)}_{j,h}]_0$, set $\mathcal{E}_{j,v,h,\ell}$ to be the event that the following hold for each $b \in [\ell]_0, p \in [\Delta^2 - b]_0$.

(EB1) $S_{v,j,h,b,2^p} \leq \frac{2^{2b+3}\kappa \Delta_R \mu (\deg_{\mathcal{G}^{(2)}}(v;V_j))^{2^p}}{|V_j|^{2^p-1}}$.

Let $\mathcal{E}^*$ be the event that $\mathcal{E}_{j,v,h,n^{(2)}_{j,h}}$ holds for each $i, j \in J$, $v \in V_i$ and $h \in [\ell_j]$ such that $ij \in E(R')$ and $\frac{\deg_{\mathcal{G}^{(2)}}(v;V_j)}{|V_j|} \geq \left(\frac{10|J|\Delta \log n}{\Delta_R \mu n}\right)^{1/2^{\Delta^2-1}}$. The following lemma states that $\mathcal{E}^*$ holds asymptotically almost surely.

Lemma 4.55. *Assume Setup 4.37. Let $c \geq \Delta + 4$ and $\varepsilon_1, \eta \in (0, 1]$ satisfy $4(\Delta\eta)^{1/7} \leq \varepsilon_1 \leq 2^{-\Delta^2}c^{-1/7}$. Suppose that $\mathcal{G}$ is an (η, c)-THC graph for (H, τ) with density weighted hypergraph $\mathcal{D}$ and that for some integer T we have a sequence $\phi_0, \dots, \phi_T$ of* THC-*respecting partial partite homomorphisms from H to $\mathcal{G}'$, where ϕ_0 is the trivial partial partite homomorphism and each ϕ_t is obtained from ϕ_{t-1} by embedding the first vertex $z \in V(H) \setminus \mathrm{Dom}(\phi_{t-1})$ according to the linear order induced by τ to a uniform random vertex from a subset $S_{\phi_{t-1},z}$ of $\mathcal{C}_{t-1}(z)$ with $|S_{\phi_{t-1},z}| \geq \frac{1}{2}|\mathcal{C}_{t-1}(z)|$. Suppose further that we have $\xi_j \leq 4\kappa\Delta_R\mu|X_j|$ for all $j \in J$ and $S_{v,y,0,2^p} \leq \frac{2(\deg_{\mathcal{G}^{(2)}}(v;V_j))^{2^p}}{|V_j|^{2^p}}$ for all $i, j \in J$ satisfying $ij \in E(R')$, $h \in [\ell_j]$, $v \in V_i$, $y \in \Xi_{j,h}$ and $p \in [\Delta^2]_0$. Then $\mathbb{P}(\mathcal{E}^*) \geq 1 - 2^{2^{\Delta_R^3+1}}\kappa\Delta_R\Delta^4 n^{1-10\Delta}$. In particular, $\mathcal{E}^*$ holds asymptotically almost surely.*

The rest of this subsection is devoted to proving Lemma 4.55. The proof is analogous to that of Lemma 4.48 and we begin with the following lemma which enables us to control the one-step evolution of the count for each individual vertex of interest.

Lemma 4.56. *Assume Setup 4.37. Suppose that for some integer T we have a sequence $\phi_0, \dots, \phi_T$ of partial partite homomorphisms from H to $\mathcal{G}'$, where ϕ_0 is the trivial partial partite homomorphism and each ϕ_t is obtained from ϕ_{t-1} by embedding the first vertex $z \in V(H) \setminus \mathrm{Dom}(\phi_{t-1})$ according to the linear order induced by τ to a uniform random vertex from a subset $S_{\phi_{t-1},z}$ of $\mathcal{C}_{t-1}(z)$. Let $j \in J$, $h \in [\ell_j]$, $v \in V(H)$, $b \in [n_{j,h}^{(2)}]$ and $y \in \Xi_{j,h}$. Then for each $p \in [\Delta^2 - b]_0$ we have*

$$\mathbb{E}\left[S_{v,y,b,2^p}|\mathcal{F}_{\tau(u_{y,b})-1}\right] \leq \frac{d'_{\tau(u_{y,b})-1}(u_{y,b})|V_{u_{y,b}}|S'_{v,y,b,2^p}}{\left|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}\right|}, \tag{4.22}$$

Proof. Let $U := N_{\mathcal{G}^{(2)}}(v; V_j)$. For $u \in \mathcal{C}_{\tau(u_{y,b})-1}(u_{y,b})$ write $f_{v,y,b,2^p}(u)$ for the number of candidates for $F'_{y,b,2^p}$ at time $\tau(u_{y,b}) - 1$ such that the copies of y would be embedded to U and $u_{y,b}$ would be embedded to u. Since $\mathrm{num}(S'_{v,y,b,2^p})$ is the number of candidates for $F'_{y,b,2^p}$ at time $\tau(u_{y,b}) - 1$ such that the copies of y would be embedded to U and $u_{y,b}$ would be embedded into $S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}$, we have

$$\mathrm{num}(S'_{v,y,b,2^p}) = \sum_{u \in S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}} f_{v,y,b,2^p}(u).$$

Since $\mathrm{num}(S_{v,y,b,2^p})$ is the number of candidates for $F_{y,b,2^p}$ at time $\tau(u_{y,b})$ such that the copies of y would be embedded to U and at time $\tau(u_{y,b})$ we embed $u_{y,b}$ to a uniform random vertex from a subset $S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}$ of $\mathcal{C}_{\tau(u_{y,b})-1}(u_{y,b})$, we obtain

$$\begin{aligned}\mathbb{E}[\mathrm{num}(S_{v,y,b,2^p})|\mathcal{F}_{\tau(u_{y,b})-1}] &\leq \frac{\sum_{u\in S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}} f_{v,y,b,2^p}(u)}{|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|}\\ &= \frac{\mathrm{num}(S'_{v,y,b,2^p})}{|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|}.\end{aligned}$$

Now since $\mathrm{den}(S'_{v,y,b,2^p}) = \mathrm{den}(S_{v,y,b,2^p})|V_{u_{y,b}}|d'_{\tau(u_{y,b})-1}(u_{y,b})$, we obtain

$$\begin{aligned}\mathbb{E}[S_{v,y,b,2^p}|\mathcal{F}_{\tau(u_{y,b})-1}] &\leq \frac{d'_{\tau(u_{y,b})-1}(u_{y,b})|V_{u_{y,b}}|\,\mathrm{num}(S'_{v,y,b,2^p})}{|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|\,\mathrm{den}(S'_{v,y,b,2^p})}\\ &\leq \frac{d'_{\tau(u_{y,b})-1}(u_{y,b})|V_{u_{y,b}}|S'_{v,y,b,2^p}}{|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|}\end{aligned}$$

as desired. □

The following lemma tells us that the relevant sum of conditional expectations which controls the one-step evolution of the counts of interest remains manageable.

Lemma 4.57. *Assume Setup 4.37. Let $c \geq \Delta + 4$ and $\varepsilon_1, \eta \in (0, 1]$ satisfy $4(\Delta\eta)^{1/7} \leq \varepsilon_1 \leq c^{-1/7}$. Suppose that $\mathcal{G}$ is an (η, c)-THC graph for (H, τ) with density weighted hypergraph $\mathcal{D}$ and that for some integer T we have a sequence $\phi_0, \ldots, \phi_T$ of THC-respecting partial partite homomorphisms from H to $\mathcal{G}'$, where ϕ_0 is the trivial partial partite homomorphism and each ϕ_t is obtained from ϕ_{t-1} by embedding the first vertex $z \in V(H) \setminus \mathrm{Dom}(\phi_{t-1})$ according to the linear order induced by τ to a uniform random vertex from a subset $S_{\phi_{t-1},z}$ of $\mathcal{C}_{t-1}(z)$ with $|S_{\phi_{t-1},z}| \geq \frac{1}{2}|\mathcal{C}_{t-1}(z)|$. Let $j \in J$, $h \in [\ell_j]$, $v \in V(H)$ and $b \in [n^{(2)}_{j,h}]$. Suppose further that we have $\xi_j \leq 4\kappa\Delta_R\mu|X_j|$ and that $\mathcal{E}_{j,v,h,b-1}$ holds. Then for each $p \in [\Delta^2 - b]_0$ we have*

$$\sum_{y\in\Xi_{j,h}} \mathbb{E}\left[S_{v,y,b,2^p}|\mathcal{F}_{\tau(u_{y,b})-1}\right] \leq (1+\varepsilon_1)\frac{2^{2b+2}\kappa\Delta_R\mu(\deg_{\mathcal{G}^{(2)}}(v; V_j^{\mathrm{main}}))^{2^p}}{|V_j|^{2^p-1}}.$$

Proof. Let $U := N_{\mathcal{G}^{(2)}}(v; V_j)$. Suppose that $\mathcal{E}_{j,v,h,b-1}$ holds. We first consider when $u_{y,b} \in N^{-1}(y)$ for all $y \in \Xi_{j,h}$. In this case we have $F_{y,b-1,2^p} = F'_{y,b,2^p}$, so we have $S_{v,y,b-1,2^p} = S'_{v,y,b,2^p}$. For each $y \in \Xi_{j,h}$ we have

$$|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}| \geq \tfrac{1}{2}|\mathcal{C}_{\tau(u_{y,b})-1}(u_{y,b})| \geq \tfrac{1-\eta}{2}d'_{\tau(u_{y,b})-1}(u_{y,b})|V_{u_{y,b}}|.$$

Then, by the lower bound on $S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}$, Lemma 4.56, summing over $y \in \Xi_{j,h}$ and (EB1) we obtain

$$\sum_{y\in\Xi_{j,h}} \mathbb{E}\left[S_{v,y,b,2^p}|\mathcal{F}_{\tau(u_{y,b})-1}\right] \leq \sum_{y\in\Xi_{j,h}} \frac{d'_{\tau(u_{y,b})-1}(u_{y,b})|V_{u_{y,b}}|S'_{v,y,b,2^p}}{\left|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}\right|} \leq \frac{2S_{v,j,h,b-1,2^p}}{1-\eta} \leq \frac{2^{2b+2}\kappa\Delta_R\mu|U|^{2^p}}{(1-\eta)|V_j|^{2^p-1}}.$$

Otherwise, we have $u_{y,b} \in N^{-2}(y)$ for all $y \in \Xi_{j,h}$. We first prove the following claim.

Claim 4.58. *For each $y \in \Xi_{j,h}$ we have*

$$S'_{v,y,b,2^p} \leq \frac{(1+\frac{\varepsilon_1}{2})|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|}{|\mathcal{C}_{\tau(u_{y,b})-1}(u_{y,b})|}\sqrt{S_{v,y,b-1,2^{p+1}}}.$$

Proof. Apply Lemma 4.14 with the following. $H = F'_{y,b,2^p}$, $A_1 = \{u_{y,b}\}$, $B = U_{y,b-1,>}$, $A_2 = \{y\}$, $V_y = U$, $V_{u_{y,b}} = S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}$ and $V_u = V'_u$ for $u \in B$. Let $\vec{a} := (2,1)$, $\vec{A} := (A_1, B)$, $F = H(\vec{A},\vec{a})$. We obtain

$$\mathrm{num}(S'_{v,y,b,2^p}) \leq \sqrt{\left|\mathcal{C}_{\tau(u_{y,b})-1}(F) \cap \left(S^2_{\phi_{\tau(u_{y,b})-1},u_{y,b}} \times V'_{U_{y,b-1,>}}\right)\right|} \times \sqrt{\mathrm{num}(S_{v,y,b-1,2^{p+1}})}.$$

Now $\phi_{\tau(u_{y,b})-1}$ is a THC-respecting partial partite homomorphism, so by Lemma 4.28 for F and (THC1) for $\mathcal{G}'_{\tau(u_{y,b})-1}$ we have

$$\begin{aligned}
&|\mathcal{C}_{\tau(u_{y,b})-1}(F) \cap (S^2_{\phi_{\tau(u_{y,b})-1},u_{y,b}} \times V'_{U_{y,b-1,>}})| \\
&\leq (1+\varepsilon_1)\frac{|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|^2|\mathcal{C}_{\tau(u_{y,b})-1}(F)|}{|\mathcal{C}_{\tau(u_{y,b})-1}(u_{y,b})|^2} \\
&\leq \frac{(1+\frac{\varepsilon_1}{2})^2|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|^2}{|\mathcal{C}_{\tau(u_{y,b})-1}(u_{y,b})|^2}\frac{\left(\mathrm{den}(S'_{v,y,b,2^p})\right)^2}{\mathrm{den}(S_{v,y,b-1,2^{p+1}})}.
\end{aligned}$$

Hence, we have

$$S'_{v,y,b,2^p} \leq \frac{(1+\frac{\varepsilon_1}{2})|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}|}{|\mathcal{C}_{\tau(u_{y,b})-1}(u_{y,b})|}\sqrt{S_{v,y,b-1,2^{p+1}}}$$

as required. ∎

By applying Lemma 4.56, Claim 4.58, summing over $y \in \Xi_{j,h}$, the lower bound on $S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}$, Lemma 4.13 applied with $\alpha_y = \sqrt{S_{v,y,b-1,2^{p+1}}}$ and $\beta_y = 1$ for $y \in \Xi_{j,h}$ and (EB1), we obtain

$$\begin{aligned}\sum_{y\in\Xi_{j,h}} \mathbb{E}\left[S_{v,y,b,2^p}|\mathcal{F}_{\tau(u_{y,b})-1}\right] &\leq \sum_{y\in\Xi_{j,h}} \frac{d'_{\tau(u_{y,b})-1}(u_{y,b})|V_{u_{y,b}}|S'_{v,y,b,2^p}}{\left|S_{\phi_{\tau(u_{y,b})-1},u_{y,b}}\right|} \\ &\leq \frac{1+\frac{\varepsilon_1}{2}}{1-\eta} \sum_{y\in\Xi_{j,h}} \sqrt{S_{v,y,b-1,2^{p+1}}} \\ &\leq \frac{1+\frac{\varepsilon_1}{2}}{1-\eta} \sqrt{\xi_{j,h} S_{v,j,h,b-1,2^{p+1}}} \\ &\leq \frac{1+\frac{\varepsilon_1}{2}}{1-\eta} \frac{2^{2b+2}\kappa\Delta_R\mu|U|^{2^p}}{|V_j|^{2^p-1}}\end{aligned}$$

as desired. $\square$

The following lemma shows that each subevent of our main event holds with very high probability.

Lemma 4.59. *Assume Setup 4.37. Let $c \geq \Delta + 4$ and $\varepsilon_1, \eta \in (0,1]$ satisfy $4(\Delta\eta)^{1/7} \leq \varepsilon_1 \leq 2^{-\Delta^2}c^{-1/7}$. Suppose that $\mathcal{G}$ is an (η, c)-*THC *graph for (H,τ) with density weighted hypergraph $\mathcal{D}$ and that for some integer T we have a sequence $\phi_0, \dots, \phi_T$ of* THC-*respecting partial partite homomorphisms from H to $\mathcal{G}'$, where ϕ_0 is the trivial partial partite homomorphism and each ϕ_t is obtained from ϕ_{t-1} by embedding the first vertex $z \in V(H) \setminus \mathrm{Dom}(\phi_{t-1})$ according to the linear order induced by τ to a uniform random vertex from a subset $S_{\phi_{t-1},z}$ of $\mathcal{C}_{t-1}(z)$ with $|S_{\phi_{t-1},z}| \geq \frac{1}{2}|\mathcal{C}_{t-1}(z)|$. Let $i,j \in J$ satisfy $ij \in E(R')$, $v \in V_i$ and $h \in [\ell_j]$. Let $q := n^{(2)}_{j,h}$ and $U := N_{\mathcal{G}^{(2)}}(v; V_j)$. Suppose further that we have $\xi_j \leq 4\kappa\Delta_R\mu|X_j|$ and $S_{v,y,0,2^p} \leq \frac{2|U|^{2^p}}{|V_j|^{2^p}}$ for all $p \in [\Delta^2]_0$ and $y \in \Xi_{j,h}$. Then for each $b \in [q]_0$ the event $\mathcal{E}_{j,v,h,b}$ holds with probability at least*

$$1 - \sum_{\ell=1}^{b} \sum_{p=0}^{\Delta^2-\ell} \exp\left(-\frac{2^{2\ell-1}\kappa\Delta_R\mu|U|^{2^p}}{|V_j|^{2^p-1}}\right).$$

In particular, the event $\mathcal{E}_{j,v,h,q}$ holds with probability at least

$$1 - \Delta^4 \exp\left(-\frac{\kappa\Delta_R\mu|U|^{2^{\Delta^2-1}}}{|V_j|^{2^{\Delta^2-1}-1}}\right).$$

Proof. We proceed by induction on b. First consider $b=0$. For each $p \in [\Delta^2]_0$, by assumption we have $S_{v,y,0,2^p} \leq \frac{2|U|^{2^p}}{|V_j|^{2^p}}$ for each $y \in \Xi_{j,h}$. Hence, we obtain $S_{v,j,h,0,2^p} = \sum_{y\in\Xi_{j,h}} S_{v,y,0,2^p} \leq \frac{8\kappa\Delta_R\mu|U|^{2^p}}{|V_j|^{2^p-1}}$. Therefore, we have $\mathbb{P}(\mathcal{E}_{j,v,h,0}) = 1$. Now consider $b \in [q]$.

Claim 4.60. *For each $p \in [\Delta^2 - b]_0$ we have*

$$\mathbb{P}\left(\mathcal{E}_{j,v,h,b-1} \text{ and } S_{v,j,h,b,2^p} > \frac{2^{2b+3}\kappa\Delta_R\mu|U|^{2^p}}{|V_j|^{2^p-1}}\right) \leq \exp\left(-\frac{2^{2b-1}\kappa\Delta_R\mu|U|^{2^p}}{|V_j|^{2^p-1}}\right).$$

Proof. Enumerate $\{u_{y,b} : y \in \Xi_{j,h}\}$ as $z_{b,1}, \ldots, z_{b,\xi_{j,h}}$ in the order according to τ. Set $\lambda := (1+\varepsilon_1)\frac{2^{2b+2}\kappa\Delta_R\mu|U|^{2^p}}{|V_j|^{2^p-1}}$. For $i \in [\xi_{j,h}]_0$ set $\mathcal{F}_i^{(b)}$ to be $\mathcal{F}_{\tau(z_{b,\xi_{j,h}})}$ if $i = \xi_{j,h}$ and $\mathcal{F}_{\tau(z_{b,i+1})-1}$ otherwise. For $i \in [\xi_{j,h}]$, with y satisfying $z_{b,i} = u_{y,b}$, set $Y_i := S_{x,\vec{v},y,b,2^p}$. Note that Y_i is $\mathcal{F}_i^{(b)}$-measurable and $0 \leq Y_i \leq R := 1 + (\Delta + 2^p)\eta$. Now when $\mathcal{E}_{j,v,h,b-1}$ holds, by Lemma 4.57 we have $\sum_{i\in[\xi_{j,h}]} \mathbb{E}\left[Y_i|\mathcal{F}_{i-1}^{(b)}\right] \leq \lambda$ and $\sum_{i\in[\xi_{j,h}]} \operatorname{var}\left(Y_i|\mathcal{F}_{i-1}^{(b)}\right) \leq R\lambda$. Then by Lemma 4.12 we have

$$\mathbb{P}\left(\mathcal{E}_{j,v,h,b-1} \text{ and } S_{v,j,h,b,2^p} > \frac{2^{2b+3}\kappa\Delta_R\mu|U|^{2^p}}{|V_j|^{2^p-1}}\right) \leq \exp\left(-\frac{2^{2b-1}\kappa\Delta_R\mu|U|^{2^p}}{|V_j|^{2^p-1}}\right)$$

as desired. ∎

By the inductive hypothesis and Claim 4.60, we have

$$\begin{aligned}
\mathbb{P}(\mathcal{E}_{j,v,h,b}) &= \mathbb{P}\left(\mathcal{E}_{j,v,h,b-1} \text{ and } S_{v,j,h,b,2^p} \leq \frac{2^{2b+3}\kappa\Delta_R\mu|U|^{2^p}}{|V_j|^{2^p-1}} \text{ for all } p \in [\Delta^2-b]_0\right) \\
&\geq \mathbb{P}(\mathcal{E}_{j,v,h,b-1}) - \sum_{p=0}^{\Delta^2-b} \mathbb{P}\left(\mathcal{E}_{j,v,h,b-1} \text{ and } S_{v,j,h,b,2^p} > \frac{2^{2b+3}\kappa\Delta_R\mu|U|^{2^p}}{|V_j|^{2^p-1}}\right) \\
&\geq 1 - \sum_{\ell=1}^{b}\sum_{p=0}^{\Delta^2-\ell} \exp\left(-\frac{2^{2\ell-1}\kappa\Delta_R\mu|U|^{2^p}}{|V_j|^{2^p-1}}\right).
\end{aligned}$$

Finally, letting $b = q$ and noting that each of the at most Δ^4 summands in the sum is at most $\exp\left(-\frac{\kappa\Delta_R\mu|U|^{2^{\Delta^2-1}}}{|V_j|^{2^{\Delta^2-1}-1}}\right)$, we obtain $\mathbb{P}(\mathcal{E}_{j,v,h,q}) \geq 1 - \Delta^4 \exp\left(-\frac{\kappa\Delta_R\mu|U|^{2^{\Delta^2-1}}}{|V_j|^{2^{\Delta^2-1}-1}}\right)$. □

Finally, we provide a proof of Lemma 4.55.

Proof of Lemma 4.55. For each $j \in J$ let $\vec{V}_j$ denote the set of all $v \in V_i$ such that $ij \in E(R')$ and $\frac{\deg_{\mathcal{G}^{(2)}}(v;V_j)}{|V_j|} \geq \left(\frac{10|J|\Delta\log n}{\Delta_R\mu n}\right)^{1/2^{\Delta^2-1}}$. Note that $|\vec{V}_j| \leq \frac{\kappa\Delta_R n}{|J|}$ for each $j \in J$. For $j \in J$ and $v \in \vec{V}_j$, by Lemma 4.59 and the fact that $|V_j| \geq \frac{n}{\kappa|J|}$, we have $\mathbb{P}\left(\mathcal{E}^c_{j,v,h,n^{(2)}_{j,h}}\right) \leq \Delta^4 n^{-10\Delta}$ for each $h \in [\ell_j]$. Now by summing over $j \in J$, $h \in [\ell_j]$ and $v \in \vec{V}_j$, and applying (H8), we obtain

$$\mathbb{P}(\mathcal{E}^*) \geq 1 - \sum_{j\in J}\sum_{v\in\vec{V}_j}\sum_{h\in[\ell_j]} \mathbb{P}\left(\mathcal{E}^c_{j,v,h,n^{(2)}_{j,h}}\right) \geq 1 - 2^{2^{\Delta_R^3+1}}\kappa\Delta_R\Delta^4 n^{1-10\Delta}$$

as desired. Finally, the final expression tends to 1 as n tends to infinity so we obtain the desired conclusion. □

4.9.2 Candidates for Many

Here we show that a buffer vertex x is reasonably likely to have any given vertex v in the corresponding cluster as a candidate, subject to the condition that each cluster does not become overfilled by vertices in a suitable initial segment. Broadly speaking, this follows from the THC property of $\mathcal{G}^{v,x,N^{-1}(x)}$. For technical reasons, we will in fact use the combined THC properties of $\mathcal{G}^{v,x,*}$ and $\mathcal{G}^{v,x,Y_i}$ for $i \in [b]_0$, which is stronger in general.

Lemma 4.61. *Assume Setup 4.37. Let $c \geq 4$ and $\mu, \varepsilon, \eta > 0$ satisfy $\mu \leq \frac{1}{4}$ and $4\eta^{1/7} \leq \varepsilon \leq \frac{\mu}{2^{\Delta+2}c^{1/7}}$. Let $j \in J$, $v \in V_j$ and $x \in X_j^{\mathrm{buf}}$. Let $y_1, \ldots, y_b$ be the neighbours of x in $H^{(2)}$ in the order according to τ. Let H_x be a J-partite k-complex with a partition $\mathcal{X}^x$ of its vertex set and a linear order τ_x on $V(H_x)$ satisfying (AB1)–(AB3). Let ϕ_0 be a good partial partite homomorphism from H to $\mathcal{G}'$ which is* THC*-respecting for $(\mathcal{G}^{v,x,*}, H_x, \tau_x)$ and whose domain contains no vertex at distance 2 or less from x. Let $Q_0 \subseteq X^{\mathrm{main}}$. Suppose $\deg_{\mathcal{G}^{(2)}}(v; V_{y_a}^{\mathrm{main}} \setminus \mathrm{Im}(\phi_0)) \geq \frac{1}{2}\deg_{\mathcal{G}^{(2)}}(v; V_{y_a})$ for each $a \in [b]$. Let $\phi_1, \ldots, \phi_b$ be good partial partite homomorphisms from H to $\mathcal{G}'$ and $Q_1, \ldots, Q_b$ be subsets of X^{main} where for each $t \in [b]$ the partial partite homomorphism ϕ_t is obtained from ϕ_{t-1} by embedding y_t to a uniform random vertex from $A_{t-1}^{\mathrm{main}}(y_t) \setminus B_{t-1}(y_t)$ and we have $Q_t = Q_{t-1} \cup \{z \in X^{\mathrm{main}} \setminus \mathrm{Dom}(\phi_t) : |A_t^{\mathrm{main}}(z)| < (1-2\varepsilon)^{\pi_t(z)}\mu|C_t^{\mathrm{main}}(z)|\}$. Then with probability at least $2^{-(b^2+5b)/2}b_j$ we have $\phi_b(H^{-1}(x)) \subseteq \mathcal{G}^{v,x,N^{-1}(x)}$.*

Proof. For $i \in [b]_0$ set $Y_i := \{y_1, \ldots, y_i\}$ and $H_i := H[Y_i]$. For $t \in [b]_0$ and $i \in [b]$, letting F_i be the down-closure complex of $\{x, y_i\}$, we shall write $\mathcal{C}_t^{v,x}(y_i)$ to mean $\mathcal{C}_{\phi_t}(F_i; v, x)$. By design ϕ_t is a good partial partite homomorphism from H to $\mathcal{G}'$ for all $t \in [b]_0$. We say that ϕ_t is a *buffer-friendly partial partite homomorphism* if it satisfies the following conditions.

(BFPH1) $\phi_t(H_t) \subseteq \mathcal{G}^{v,x,N^{-1}(x)}$.

(BFPH2) ϕ_t is THC-respecting for $(\mathcal{G}^{v,x,*}, H_x, \tau_x)$.

(BFPH3) $|A_t^{\mathrm{main}}(y_\ell) \cap \mathcal{C}_t^{v,x}(y_\ell)| \geq 2^{-t-1}|\mathcal{C}_t^{v,x}(y_\ell)|$ for $t < \ell \leq b$.

Let us first check that ϕ_0 is a buffer-friendly THC-respecting partial partite homomorphism from H to $\mathcal{G}'$. Indeed, (BFPH1) is vacuously satisfied, (BFPH2) holds by assumption and (BFPH3) holds by the assumption on $\deg_{\mathcal{G}^{(2)}}(v; V_i^{\mathrm{main}} \setminus \mathrm{Im}(\phi_0))$ and the fact that no vertex at distance 2 or less from x has been embedded. Since (BFPH2) holds, by Lemma 4.39 ϕ_0 is THC-respecting.

We shall now establish, for any $t \in [b]$, a lower bound on the probability that ϕ_t is a buffer-friendly THC-respecting good partial partite homomorphism from H to $\mathcal{G}'$, conditioned on ϕ_{t-1} being a buffer-friendly THC-respecting good partial partite homomorphism from H to $\mathcal{G}'$. Suppose that ϕ_{t-1} is a buffer-friendly THC-respecting good partial partite homomorphism from H to $\mathcal{G}'$. Observe that our embedding procedure automatically maintains (GPH1) and we maintain (BFPH1) if we embed y_t into $A^{\text{main}}_{t-1}(y_t) \cap \mathcal{C}^{v,x}_{t-1}(y_t)$. For (BFPH2) observe that since ϕ_{t-1} is THC-respecting for $(\mathcal{G}^{v,x,*}, H_x, \tau_x)$, $\mathcal{G}'_{t-1}$ is binary by Lemma 4.25 and (THC2) holds, we have that $\phi_t = \phi_{t-1} \cup \{y_t \to w\}$ is THC-respecting for $(\mathcal{G}^{v,x,*}, H_x, \tau_x)$, and therefore by Lemma 4.39 also THC-respecting for $(\mathcal{G}', H, \tau)$, for all but at most $\eta|\mathcal{C}^{v,x}_{t-1}(y_t)|$ vertices w in $A^{\text{main}}_{t-1}(y_t) \cap \mathcal{C}^{v,x}_{t-1}(y_t)$. Furthermore, by Lemma 4.39 ϕ_{t-1} is THC-respecting for $\mathcal{G}^{v,x,Y_t}$ and $\mathcal{G}^{v,x,Y_b}$.

Now we consider (GPH2) for unembedded neighbours z of y_t in $H^{(2)}$. Considering ϕ_{t-1} as a function into $\mathcal{G}^{v,x,Y_t}$, by Lemma 4.28 $G^{\phi_{t-1}}_{\{y_t\},\{z\}}$ is (ε)-regular. Since (GPH2) holds for ϕ_{t-1} we have $|\mathcal{C}^{\text{main}}_{t-1}(z)| \geq (1-2\varepsilon)^{\pi_{t-1}(z)}(1-2\mu)|\mathcal{C}_{t-1}(z)| \geq \varepsilon|\mathcal{C}_{t-1}(z)|$. By (THC1) in $\mathcal{G}^{v,x,Y_t}$ the density of $G^{\phi_{t-1}}_{\{y_t\},\{z\}}$ is $(1 \pm \frac{\varepsilon}{2})d_{t-1}(y_t z)$. Then, by Lemma 4.16 there are at most $\varepsilon|\mathcal{C}^{v,x}_{t-1}(y_t)|$ vertices w in $\mathcal{C}^{v,x}_{t-1}(y_t)$ such that

$$\deg_{\mathcal{G}^{(2)}}(w; \mathcal{C}^{\text{main}}_{t-1}(z)) < (1-2\varepsilon)d_{t-1}(y_t z)|\mathcal{C}^{\text{main}}_{t-1}(z)|.$$

We argue analogously for each of the other two conditions of (GPH2). Hence, since y_t has at most Δ unembedded neighbours in $H^{(2)}$, we find that (GPH2) fails to hold with respect to $\phi_{t-1} \cup \{y_t \to w\}$ for at most $3\Delta\varepsilon|\mathcal{C}^{v,x}_{t-1}(y_t)|$ vertices $w \in \mathcal{C}^{v,x}_{t-1}(y_t)$. For the badness condition (4.7) the argument is also analogous, noting that in this case we focus on unembedded neighbours $z \in V(H) \setminus Q_{t-1}$ of y_t in $H^{(2)}$ and for such vertices we have

$$|A^{\text{main}}_{t-1}(z)| \geq (1-2\varepsilon)^{\pi_{t-1}(z)}\mu|\mathcal{C}^{\text{main}}_{t-1}(z)| \geq \varepsilon|\mathcal{C}_{t-1}(z)|.$$

Hence, in total there are at most $\Delta\varepsilon|\mathcal{C}^{v,x}_{t-1}(y_t)|$ vertices of $\mathcal{C}^{v,x}_{t-1}(y_t)$ such that the badness condition holds for some unembedded neighbour of y_t.

It remains to consider (BFPH3). Let $t < \ell \leq b$ and consider ϕ_{t-1} as a map into $\mathcal{G}^{v,x,Y_b}$; by Lemma 4.28 $G^{\phi_{t-1}}_{\{y_t\},\{y_\ell\}}$ is (ε)-regular. Since (BFPH3) holds for ϕ_{t-1}, we have

$$|A^{\text{main}}_{t-1}(y_\ell) \cap \mathcal{C}^{v,x}_{t-1}(y_\ell)| \geq 2^{-t}|\mathcal{C}^{v,x}_{t-1}(y_\ell)| \geq \varepsilon|\mathcal{C}^{v,x}_{t-1}(y_\ell)|.$$

By (THC1) in $\mathcal{G}^{v,x,Y_b}$ the density of $G^{\phi_{t-1}}_{\{y_t\},\{y_\ell\}}$ is $(1 \pm \frac{\varepsilon}{2})d_{t-1}(y_t z)$. Then, by Lemma 4.16 there are at most $\varepsilon|\mathcal{C}^{v,x}_{t-1}(y_t)|$ vertices w in $\mathcal{C}^{v,x}_{t-1}(y_t)$ such that

$$\deg_{\mathcal{G}^{(2)}}(w; A^{\mathrm{main}}_{t-1}(y_\ell) \cap \mathcal{C}^{v,x}_{t-1}(y_\ell)) < (1-2\varepsilon)d_{t-1}(y_t z)|A^{\mathrm{main}}_{t-1}(y_\ell) \cap \mathcal{C}^{v,x}_{t-1}(y_\ell)|.$$

Now summing over $t < \ell \leq b$, we find that in total there are at most $\Delta\varepsilon|\mathcal{C}^{v,x}_{t-1}(y_t)|$ vertices $w \in \mathcal{C}^{v,x}_{t-1}(y_t)$ such that (BFPH3) fails to hold with respect to $\phi_{t-1} \cup \{y_t \to w\}$.

Therefore, given that ϕ_{t-1} is buffer-friendly THC-respecting good partial partite homomorphism, in total there are at most

$$8\Delta\varepsilon|\mathcal{C}^{v,x}_{t-1}(y_t)| \leq 2^{-t-1}|\mathcal{C}^{v,x}_{t-1}(y_t)| \leq \frac{1}{2}|A^{\mathrm{main}}_{t-1}(y_t) \cap \mathcal{C}^{v,x}_{t-1}(y_t)|$$

vertices $w \in A^{\mathrm{main}}_{t-1}(y_t) \cap \mathcal{C}^{v,x}_{t-1}(y_t)$ such that $\phi_{t-1} \cup \{y_t \to w\}$ is not a buffer-friendly THC-respecting good partial partite homomorphism. Observe that our embedding procedure embed y_t uniformly at random into a set of size at most $|\mathcal{C}_{t-1}(y_t)|$ vertices, of which at least

$$\frac{1}{2}|A^{\mathrm{main}}_{t-1}(y_t) \cap \mathcal{C}^{v,x}_{t-1}(y_t)| \geq 2^{-t-1}|\mathcal{C}^{v,x}_{t-1}(y_t)|$$

vertices w would make $\phi_t = \phi_{t-1} \cup \{y_t \to w\}$ a buffer friendly THC-respecting good partial partite homomorphism. Therefore, conditioning on the history and on ϕ_{t-1} being a buffer friendly THC-respecting good partial partite homomorphism, we find that the probability of ϕ_t being a buffer friendly THC-respecting good partial partite homomorphism is at least

$$\frac{2^{-t-1}|\mathcal{C}^{v,x}_{t-1}(y_t)|}{|\mathcal{C}_{t-1}(y_t)|} \geq 2^{-t-2} \prod_{A \subseteq Y_{t-1}} d(A \cup \{y_t, x\})$$

The conditional probabilities multiply, so we find that the probability that ϕ_b is a buffer-friendly THC-respecting good partial partite homomorphism from H to $\mathcal{G}'$, conditioned on ϕ_0 being a buffer-friendly THC-respecting good partial partite homomorphism from H to $\mathcal{G}'$, is at least

$$2^{-(b^2+5b)/2}\frac{\mathcal{D}(H^{\leq 1}(x))}{\mathcal{D}(H^{-1}(x))}.$$

This is a lower bound on the desired conditional probability, completing the proof. $\square$

4.10 Proof of the Blow-up Lemma

In this section we provide a proof of Theorem 4.5. We state the RGA Lemma (Lemma 4.62) and provide the actual proof of Theorem 4.5. Then, we prove Lemma 4.62 in Section 4.10.1.

The following is the RGA Lemma, which encapsulates the outcome of applying a suitable random greedy algorithm (RGA) to embed X^{main} into V^{main}. It guarantees the existence of a good partial partite homomorphism with certain desirable deterministic properties which the RGA produces with high probability.

Lemma 4.62 (RGA Lemma)**.** *Assume Setup 4.37. Then there is a* THC*-respecting good partial partite homomorphism* ϕ_{RGA} *from* H *to* $\mathcal{G}'$ *such that the following hold for each* $j \in J$*.*

(RGA1) $\mathrm{Dom}(\phi_{\mathrm{RGA}}) \cap X_j = X_j^{\mathrm{main}}$*, for every* $x, y \in X_j^{\mathrm{main}}$ *we have*

$$x = y \iff \phi_{\mathrm{RGA}}(x)_{\to\mathcal{G}} = \phi_{\mathrm{RGA}}(y)_{\to\mathcal{G}},$$

for each $x \in X_j^{\mathrm{main}}$ *we have* $\phi_{\mathrm{RGA}}(x)_{\to\mathcal{G}} \in V_j^{\mathrm{main}} \cup V_j^{\mathrm{q}}$ *and for each* $e \in E(H[\mathrm{Dom}(\phi_{\mathrm{RGA}})])$ *we have* $\phi_{\mathrm{RGA}}(e)_{\to\mathcal{G}} \in E(\mathcal{G})$*.*

(RGA2) $b_j \geq \frac{2^{4+(\Delta^2+5\Delta)/2}\log(\rho|V_j|)}{\mu|V_j|}$.

(RGA3) For every set $W \subseteq V_j$ *of size at least* $\rho|V_j|$*, there are at most* $\rho|X_j|$ *vertices* $x \in X_j^{\mathrm{buf}}$ *with fewer than* $\frac{b_j|W|}{2}$ *candidates in* $W_{j\to x}$*.*

(RGA4) For each $x \in X_j^{\mathrm{buf}}$*, letting*

$$Y_x = \left\{ y \in X_j^{\mathrm{buf}} : |\mathcal{C}^{\mathrm{buf}}(y)_{\to\mathcal{G}} \cap \mathcal{C}^{\mathrm{buf}}(x)_{\to\mathcal{G}}| > (1+4\eta)\frac{4|\mathcal{C}^{\mathrm{buf}}(y)||\mathcal{C}^{\mathrm{buf}}(x)|}{\mu^2|V_j|} \right\},$$

we have

$$\sum_{y \in Y_x} \frac{|\mathcal{C}^{\mathrm{buf}}(y)_{\to\mathcal{G}} \cap \mathcal{C}^{\mathrm{buf}}(x)_{\to\mathcal{G}}|}{|\mathcal{C}^{\mathrm{buf}}(y)|} \leq \frac{2^{2^{\Delta_R^3+1}+4\Delta^2+4}\Delta^3\rho|\mathcal{C}^{\mathrm{buf}}(x)|}{\mu^{\Delta^2+2}}.$$

(RGA5) For each $v \in V_j$ *we have*

$$|\{x \in X_j^{\mathrm{buf}} : v \in \mathcal{C}(x)_{\to\mathcal{G}}\}| \geq 2^{-(\Delta^2+5\Delta)/2+1} b_j \mu |X_j|.$$

(RGA6) For each $v \in V_j$ *and each set* $W \subseteq V_j$ *we have*

$$\sum_{w \in W} |\{x \in X_j^{\mathrm{buf}} : v, w \in \mathcal{C}(x)_{\to\mathcal{G}}\}| \leq 2^{\Delta+3} b_j^2 \mu |X_j| (|W| + 2\Delta^{1/7}\varepsilon|V_j|).$$

(RGA1) states that the vertices of X^{main} are embedded into $V^{\mathrm{main}} \cup V^{\mathrm{q}}$ by ϕ_{RGA}. The lower bound in (RGA2) ensures that candidate sets of buffer vertices are reasonably large, (RGA3) says that candidate sets of buffer vertices are distributed uniformly and (RGA4) asserts that atypically large intersections of candidate sets are highly unlikely. (RGA5) ensures that each vertex $v \in V_j$ is a candidate for reasonably many buffer vertices $x \in X_j^{\mathrm{buf}}$ and (RGA6) says that sets of buffer vertices with a specific candidate typically intersect as if they are random sets.

Now we give the full proof of Theorem 4.5.

Proof of Theorem 4.5. We first determine our choices of constants. Given integers $k, \Delta \geq 2$, $\Delta_R^{\mathrm{LEM}} \in \mathbb{N}$, and real numbers $\alpha^{\mathrm{LEM}} \in (0,1]$ and $\kappa^{\mathrm{LEM}} \geq 1$, let $\beta = \frac{1}{8(\Delta^6+1)}$, $\Delta_R = \beta^{-1}\Delta_R^{\mathrm{LEM}}$, $\alpha = \frac{1}{2}\alpha^{\mathrm{LEM}}$ and $\kappa = 2\kappa^{\mathrm{LEM}}$. We now choose $c \geq \Delta_{\mathrm{aux}} + 2$,

$$\mu = \frac{\alpha}{2^{2^{\Delta_R^3+1}+2\Delta^2+8}\kappa\Delta^{c+4}\Delta_R^{c+4}}, \qquad \rho \leq \frac{\mu^{\Delta^2+2}}{2^{2^{\Delta_R^3+1}+4\Delta^2+10}\Delta^3},$$

$$\varepsilon \leq \left(\frac{\mu 2^{-4/\rho}}{c\Delta 2^{\Delta^3+4}}\right)^{2^{\Delta^2}}, \text{ and} \qquad \eta \leq \frac{\varepsilon^7}{2^{14}}.$$

Given inputs k, Δ, Δ_R, c and η, Lemma 4.38 returns a constant $\eta_0 > 0$. Let $\eta' = \eta_0/2$, $c' = (\Delta_{\mathrm{aux}}+2)(\Delta+2)c$ and $\eta^{\mathrm{LEM}} = \frac{(\eta')^{2c'+3}}{c'2^{2c'+16}}$. Now Theorem 4.5 returns c and η^{LEM}. Given a finite set J^{LEM}, let J be a finite set of size $|J^{\mathrm{LEM}}|/\beta$. Given inputs k, Δ, Δ_R, α, κ, c, η', μ and J, Lemma 4.32 returns $n_1 \in \mathbb{N}$ and Lemma 4.62 returns $n_2 \in \mathbb{N}$. Let $n_0 = \max(n_1, n_2)$.

Let R_{LEM} be a k-complex on J^{LEM} and let R'_{LEM} be a spanning subcomplex of R_{LEM}. Let H and $\mathcal{G}$ be J^{LEM}-partite k-complexes on $n \geq n_0$ vertices with κ^{LEM}-balanced size-compatible vertex partitions $\mathcal{X}^{\mathrm{LEM}}$ and $\mathcal{V}^{\mathrm{LEM}}$ respectively such that $\Delta(H^{(2)}) \leq \Delta$, $\varnothing \in E(\mathcal{G})$ and $\{v\} \in E(\mathcal{G})$ for all $v \in V(\mathcal{G})$. Let $\mathcal{D}^{\mathrm{LEM}}$ be a weighted hypergraph on J^{LEM} with $d^{\mathrm{LEM}}(\varnothing) = 1$, $d^{\mathrm{LEM}}(\{j\}) = 1$ for all $j \in J^{\mathrm{LEM}}$ and $d^{\mathrm{LEM}}(e) > 0$ for all $e \in E(R_{\mathrm{LEM}})$. Let $\widetilde{\mathcal{X}}^{\mathrm{LEM}} = \{\widetilde{X}_j^{\mathrm{LEM}}\}_{j \in J^{\mathrm{LEM}}}$ be a family of potential buffer vertices. Suppose that the conditions (BUL1)–(BUL3) of Theorem 4.5 are satisfied. Then by Lemma 4.32 there is a k-complex R on J and a spanning subcomplex R' of R, κ-balanced size-compatible vertex partitions $\mathcal{X} = \{X_j\}_{j \in J}$ and $\mathcal{V} = \{V_j\}_{j \in J}$ of H and $\mathcal{G}$ respectively whose parts are of size at least $\frac{n}{\kappa|J|}$, a weighted hypergraph $\mathcal{D}$ on J with $d(\varnothing) = 1$, $d(\{j\}) = 1$ for all $j \in J$ and $d(e) > 0$ for all $e \in E(R)$, a family $\widetilde{\mathcal{X}} = \{\widetilde{X}_j\}_{j \in J}$ of potential buffer vertices, subsets $X_j^{\mathrm{buf}} \subseteq \widetilde{X}_j$ for each $j \in J$, a good vertex order τ for X^{buf} on $V(H)$, a partition $\{X_{j,h}\}_{h\in[\ell_j]}$ of X_j for each $j \in J$, and a partition $V_j = V_j^{\text{main}} \cup V_j^{\text{q}} \cup V_j^{\text{buf}}$ for each $j \in J$, which give an $(\alpha, c, \Delta, \Delta_R, \kappa, \mu)$-good H-partition and a $(c, \Delta, \Delta_R, \eta', \mu)$-good $\mathcal{G}$-partition. Hence, we may assume Setup 4.37 from now onwards.

By Lemma 4.62, there is a THC-respecting good partial partite homomorphism ϕ_{RGA} from H to $\mathcal{G}'$ with properties (RGA1)–(RGA6). Since by (H3) any two vertices in X^{buf} are at distance at least five in $H^{(2)}$, every vertex x in X^{buf} has all its neighbours already embedded and so the candidate set $\mathcal{C}(x)$ will no longer change. For each $j \in J$ let G_j be the bipartite graph with vertex sets X_j^{buf} and $\overline{V}_j := V_j \setminus \text{Im}(\phi_{\text{RGA}})_{\to\mathcal{G}}$ and edge set $\{xv : v \in \mathcal{C}(x)_{\to\mathcal{G}}\}$. We claim that we can find a system of matchings M_j in G_j for each $j \in J$.

Claim 4.63. *For each $j \in J$ the graph G_j contains a perfect matching.*

Suppose for now that Claim 4.63 holds. For each $j \in J$ fix a perfect matching M_j in G_j and for each $x \in X_j^{\text{buf}}$ let v_x be the unique vertex in $\overline{V}_j$ such that $xv_x \in M_j$. Let the function $\phi : V(H) \to V(\mathcal{G})$ be given as follows.

$$\phi(x) := \begin{cases} \phi_{\text{RGA}}(x)_{\to\mathcal{G}} & \text{if } x \in \text{Dom}(\phi_{\text{RGA}}) \\ v_x & \text{if } x \in X^{\text{buf}}. \end{cases}$$

We shall show that ϕ is an embedding of H into $\mathcal{G}$ such that $\phi(x) \in V_x$ for each $x \in V(H)$. By the definition of v_x and (RGA1), ϕ is injective, $\phi(x) \in V_x$ for each $x \in V(H)$ and $\phi(e) \in E(\mathcal{G})$ for all $e \in E(H[\text{Dom}(\phi_{\text{RGA}})])$. It remains to show that $\phi(e) \in E(\mathcal{G})$ for all $e \in E(H) \setminus E(H[\text{Dom}(\phi_{\text{RGA}})])$. Let $e \in E(H) \setminus E(H[\text{Dom}(\phi_{\text{RGA}})])$. Note that e contains at least one vertex from X^{buf}; in fact, e contains exactly one such vertex because by (H3) no two vertices in X^{buf} are adjacent to each other in $H^{(2)}$. Let that vertex be $x_e \in X_j$. Since $\phi(x_e) = v_{x_e} \in \mathcal{C}(x_e)_{\to\mathcal{G}}$, we have

$$1 = g'_{\phi_{\text{RGA}}}((v_{x_e})_{j\to x_e}) = \prod_{A\subseteq \text{Dom}(\phi_{\text{RGA}})} g'(\phi_{\text{RGA}}(A) \cup \{(v_{x_e})_{j\to x_e}\}).$$

Since $\mathcal{G}'$ is a binary weighted hypergraph, this means that $g'(\phi_{\text{RGA}}(A)\cup\{(v_{x_e})_{j\to x_e}\}) = 1$ for all $A \subseteq \text{Dom}(\phi_{\text{RGA}})$. In particular, by setting $A = e \setminus \{x_e\}$ and the definition of standard construction, we have $g(\phi(e)) = g'(\phi_{\text{RGA}}(A) \cup \{(v_{x_e})_{j\to x_e}\}) = 1$. Hence, we have $\phi(e) \in E(\mathcal{G})$, completing the proof that $\phi : V(H) \to V(\mathcal{G})$ is an embedding of H into $\mathcal{G}$ such that $\phi(x) \in V_x$ for each $x \in V(H)$. Now it remains to prove Claim 4.63.

Proof of Claim 4.63. Let $j \in J$. We shall find a perfect matching in G_j by verifying Hall's condition. Let $Y \subseteq X_j^{\mathrm{buf}}$ and set $U := \bigcup_{y \in Y} A(y)_{\to \mathcal{G}}$. We shall show that $|Y| \leq |U|$ by considering three cases based on the size of Y.

Consider when $0 \leq |Y| \leq \rho |X_j|$. Enumerate Y as $y_1, \ldots, y_{|Y|}$. For each $i = 1, \ldots, |Y|$ choose v_i uniformly at random from $\mathcal{C}^{\mathrm{buf}}(y_i)_{\to \mathcal{G}} \setminus \{v_1, \ldots, v_{i-1}\}$ if this is possible; if not, say that v_i does not exist. For each $\ell \in [|Y|]_0$ let $\mathcal{E}_\ell$ be the event that $|\mathcal{C}^{\mathrm{buf}}(y_i)_{\to \mathcal{G}} \setminus \{v_1, \ldots, v_{i-1}\}| \geq \frac{1}{2}|\mathcal{C}^{\mathrm{buf}}(y_i)|$ holds for each $i \in [\ell]$. Since ϕ_{RGA} is a THC-respecting good partial partite homomorphism, by (THC1) and (GPH2) for each $y \in Y$ we have $|\mathcal{C}^{\mathrm{buf}}(y)| \geq \frac{1}{2}\mu|\mathcal{C}(y)| \geq \frac{1-\eta}{2} b_j \mu |V_j|$. We claim that $\mathcal{E}_\ell$ holds with probability at least $1 - \sum_{i=1}^{\ell} \exp\left(-\frac{|\mathcal{C}^{\mathrm{buf}}(y_i)|}{16}\right)$ for all $\ell \in [|Y|]_0$; the statement for $\ell = |Y|$ would imply that asymptotically almost surely $|\mathcal{C}^{\mathrm{buf}}(y_i)_{\to \mathcal{G}} \setminus \{v_1, \ldots, v_{i-1}\}| \geq \frac{1}{2}|\mathcal{C}^{\mathrm{buf}}(y_i)|$ holds for each $i \in [|Y|]$. This would mean that asymptotically almost surely we could pick v_i for each $i \in [|Y|]$ by the procedure outlined above and in particular, there would be a valid selection of v_i for each $i \in [|Y|]$ such that they would all be distinct. Since $v_i \in U$ for each $i \in [|Y|]$, we could then conclude that $|Y| = |\{v_i : i \in [|Y|]\}| \leq |U|$.

We prove our claim by induction on ℓ. For $\ell = 0$, we have $\mathbb{P}(\mathcal{E}_0) = 1$ trivially. Consider $\ell \in [|Y|]$. For each $i \in [\ell - 1]$ define the random variable $Z_{\ell,i}$ as follows. Set $Z_{\ell,i} = 1$ if $v_i \in \mathcal{C}^{\mathrm{buf}}(y_\ell)_{\to \mathcal{G}}$ and $|\mathcal{C}^{\mathrm{buf}}(y_i)_{\to \mathcal{G}} \setminus \{v_1, \ldots, v_{i-1}\}| \geq \frac{1}{2}|\mathcal{C}^{\mathrm{buf}}(y_i)|$, and $Z_{\ell,i} = 0$ otherwise. $Z_{\ell,i}$ is an $\mathcal{F}_i$-measurable random variable which satisfies $0 \leq Z_{\ell,i} \leq 1$. When $\mathcal{E}_{\ell-1}$ holds we have

$$\sum_{i=1}^{\ell-1} \mathbb{E}[Z_{\ell,i} | \mathcal{F}_{i-1}] = \sum_{i=1}^{\ell-1} \mathbb{P}(Z_{\ell,i} = 1 | \mathcal{F}_{i-1}) \leq \sum_{i=1}^{\ell-1} \frac{|\mathcal{C}^{\mathrm{buf}}(y_i)_{\to \mathcal{G}} \cap \mathcal{C}^{\mathrm{buf}}(y_\ell)_{\to \mathcal{G}}|}{\frac{1}{2}|\mathcal{C}^{\mathrm{buf}}(y_i)|}.$$

Consider the summands in the final sum. If $y_i \notin Y_{y_\ell}$, the corresponding summand is at most $\frac{8|\mathcal{C}^{\mathrm{buf}}(y_\ell)|}{\mu^2 |V_j|}$; (RGA4) bounds the sum over the terms corresponding to $y_i \in Y_{y_\ell}$. Putting these together, we obtain

$$\begin{aligned}
\sum_{i=1}^{\ell-1} \mathrm{var}(Z_{\ell,i} | \mathcal{F}_{i-1}) &\leq \sum_{i=1}^{\ell-1} \mathbb{E}[Z_{\ell,i} | \mathcal{F}_{i-1}] \\
&\leq \sum_{i=1}^{\ell-1} \frac{|\mathcal{C}^{\mathrm{buf}}(y_i)_{\to \mathcal{G}} \cap \mathcal{C}^{\mathrm{buf}}(y_\ell)_{\to \mathcal{G}}|}{\frac{1}{2}|\mathcal{C}^{\mathrm{buf}}(y_i)|} \\
&\leq \frac{8(\ell-1)|\mathcal{C}^{\mathrm{buf}}(y_\ell)|}{\mu^2 |V_j|} + \frac{2^{2^{\Delta_R^3+1}+4\Delta^2+4} \Delta^3 \rho |\mathcal{C}^{\mathrm{buf}}(y_\ell)|}{\mu^{\Delta^2+2}} \\
&\leq \frac{|\mathcal{C}^{\mathrm{buf}}(y_\ell)|}{4}.
\end{aligned}$$

Then, by Lemma 4.12 we get

$$\mathbb{P}\left(\mathcal{E}_{\ell-1} \text{ and } \sum_{i=1}^{\ell-1} Z_{\ell,i} > \frac{|\mathcal{C}^{\mathrm{buf}}(y_\ell)|}{2}\right) \le e^{-\frac{|\mathcal{C}^{\mathrm{buf}}(y_\ell)|}{16}}.$$

Now applying the inductive hypothesis, we obtain

$$\begin{aligned}\mathbb{P}\left(\mathcal{E}_\ell\right) &= \mathbb{P}\left(\mathcal{E}_{\ell-1} \text{ and } \sum_{i=1}^{\ell-1} Z_{\ell,i} \le \frac{|\mathcal{C}^{\mathrm{buf}}(y_\ell)|}{2}\right)\\ &\ge \mathbb{P}\left(\mathcal{E}_{\ell-1}\right) - \mathbb{P}\left(\mathcal{E}_{\ell-1} \text{ and } \sum_{i=1}^{\ell-1} Z_{\ell,i} > \frac{|\mathcal{C}^{\mathrm{buf}}(y_\ell)|}{2}\right)\\ &\ge 1 - \sum_{i=1}^{\ell} \exp\left(-\frac{|\mathcal{C}^{\mathrm{buf}}(y_i)|}{16}\right),\end{aligned}$$

completing the proof in this case.

Consider when $\rho|X_j| < |Y| \le |X_j^{\mathrm{buf}}| - \rho|X_j| = |\overline{V}_j| - \rho|V_j|$. Suppose for a contradiction that $|Y| > |U|$, so $|\overline{V}_j \setminus U| > |\overline{V}_j| - |Y| \ge \rho|V_j|$. By (RGA3) there are at most $\rho|X_j|$ vertices of X_j^{buf} with fewer than $\frac{b_j}{2}|\overline{V}_j \setminus U|$ candidates in $\overline{V}_j \setminus U$. In particular, there is a vertex in Y with candidates in $\overline{V}_j \setminus U$, which contradicts the definition of U. Hence, we have $|U| \ge |Y|$.

Consider when $|Y| > |X_j^{\mathrm{buf}}| - \rho|X_j| = |\overline{V}_j| - \rho|V_j|$. For each $v \in \overline{V}_j$ set $\overline{C}(v) := \{x \in X_j^{\mathrm{buf}} : v \in \mathcal{C}(x)_{\to\mathcal{G}}\}$. Enumerate $\overline{V}_j \setminus U$ as $v_1, \ldots, v_{|\overline{V}_j \setminus U|}$. For each $i = 1, \ldots, |\overline{V}_j \setminus U|$ choose x_i uniformly at random from $\overline{C}(v_i) \setminus \{x_1, \ldots, x_{i-1}\}$ if this is possible; if not, say that x_i does not exist. For each $\ell \in [|\overline{V}_j \setminus U|]_0$ let $\overline{\mathcal{E}}_\ell$ be the event that $|\overline{C}(v_i) \setminus \{x_1, \ldots, x_{i-1}\}| \ge \frac{1}{2}|\overline{C}(v_i)|$ holds for each $i \in [\ell]$. We claim that $\overline{\mathcal{E}}_\ell$ holds with probability at least $1 - \sum_{i=1}^{\ell} \exp\left(-\frac{|\overline{C}(v_i)|}{16}\right)$ for all $\ell \in [|\overline{V}_j \setminus U|]_0$; by (RGA5) and since $\ell \le \rho|V_j|$, the statement for $\ell = |\overline{V}_j \setminus U|$ would imply that asymptotically almost surely $|\overline{C}(v_i) \setminus \{x_1, \ldots, x_{i-1}\}| \ge \frac{1}{2}|\overline{C}(v_i)|$ holds for each $i \in [|\overline{V}_j \setminus U|]$. This would mean that asymptotically almost surely we could pick x_i for each $i \in [|\overline{V}_j \setminus U|]$ by the procedure outlined above and in particular, there would be a valid selection of x_i for each $i \in [|\overline{V}_j \setminus U|]$ such that they would all be distinct. Since $x_i \in X_j^{\mathrm{buf}} \setminus Y$ for each $i \in [|\overline{V}_j \setminus U|]$, we could then conclude that $|\overline{V}_j \setminus U| = |\{x_i : i \in [|\overline{V}_j \setminus U|]\}| \le |X_j^{\mathrm{buf}} \setminus Y|$; since $|\overline{V}_j| = |X_j^{\mathrm{buf}}|$, this would imply $|Y| \le |U|$ as desired.

We prove our claim by induction on ℓ. For $\ell = 0$, we have $\mathbb{P}(\overline{\mathcal{E}}_0) = 1$ trivially. Consider $\ell \in [|\overline{V}_j \setminus U|]$. For each $i \in [\ell - 1]$ define the random variable $\overline{Z}_{\ell,i}$ as follows. Set $\overline{Z}_{\ell,i} = 1$ if $x_i \in \overline{C}(v_\ell)$ and $|\overline{C}(v_i) \setminus \{x_1, \ldots, x_{i-1}\}| \ge \frac{1}{2}|\overline{C}(v_i)|$, and $\overline{Z}_{\ell,i} = 0$

otherwise. $\overline{Z}_{\ell,i}$ is an $\overline{\mathcal{F}}_i$-measurable random variable which satisfies $0 \le \overline{Z}_{\ell,i} \le 1$. By applying (RGA5) and (RGA6) with $W = \{v_i : i \in [\ell-1]\}$, when $\overline{\mathcal{E}}_{\ell-1}$ holds we have

$$\begin{aligned}\sum_{i=1}^{\ell-1} \mathrm{var}(\overline{Z}_{\ell,i}|\overline{\mathcal{F}}_{i-1}) \le \sum_{i=1}^{\ell-1} \mathbb{E}[\overline{Z}_{\ell,i}|\overline{\mathcal{F}}_{i-1}] &\le \sum_{i=1}^{\ell-1} \frac{|\overline{C}(v_i) \cap \overline{C}(v_\ell)|}{\frac{1}{2}|\overline{C}(v_i)|} \\ &\le 2^{(\Delta^2+7\Delta)/2+3} b_j(\ell - 1 + 2\Delta^{1/7}\varepsilon|V_j|) \\ &\le \frac{|\overline{C}(v_\ell)|}{4}.\end{aligned}$$

Then, by Lemma 4.12 we get

$$\mathbb{P}\left(\overline{\mathcal{E}}_{\ell-1} \text{ and } \sum_{i=1}^{\ell-1} \overline{Z}_{\ell,i} > \frac{|\overline{C}(v_\ell)|}{2}\right) \le \exp\left(-\frac{|\overline{C}(v_\ell)|}{16}\right).$$

Applying the inductive hypothesis, we obtain

$$\begin{aligned}\mathbb{P}\left(\overline{\mathcal{E}}_\ell\right) &= \mathbb{P}\left(\overline{\mathcal{E}}_{\ell-1} \text{ and } \sum_{i=1}^{\ell-1} \overline{Z}_{\ell,i} \le \frac{|\overline{C}(v_\ell)|}{2}\right) \\ &\ge \mathbb{P}\left(\overline{\mathcal{E}}_{\ell-1}\right) - \mathbb{P}\left(\overline{\mathcal{E}}_{\ell-1} \text{ and } \sum_{i=1}^{\ell-1} \overline{Z}_{\ell,i} > \frac{|\overline{C}(v_\ell)|}{2}\right) \\ &\ge 1 - \sum_{i=1}^{\ell} \exp\left(-\frac{|\overline{C}(v_i)|}{16}\right),\end{aligned}$$

completing the proof in this case. This completes the verification of Hall's condition.

■

This completes the proof of Theorem 4.5. □

4.10.1 Proof of the RGA Lemma

Here we describe our random greedy algorithm, Algorithm RGA, and prove that with high probability it produces a partial partite homomorphism from H to $\mathcal{G}'$ consistent with Lemma 4.62. Algorithm RGA sequentially embeds vertices of H according to the good vertex order τ included in the provided good H-partition, thereby building up a sequence (ϕ_s) of good partial partite homomorphisms and a designated *queue* of vertices given as a sequence (Q_s) of subsets of $V(H)$ with Q_s being the queue at time s. Let $B_s(x)$ denote the set of bad vertices with respect to ϕ_s, H_+ and Q_s, with H_+ returned by Lemma 4.35. We add unembedded vertices y to the queue if the set $A_s^{\mathrm{main}}(y) \setminus B_s(y)$ becomes small and create ϕ_s from ϕ_{s-1} by embedding the sth vertex x

in the order according to τ uniformly at random into the set $A_s^{\mathrm{main}}(x) \setminus B_s(x)$ if x is not in the queue; if x is a queue vertex we embed uniformly at random into $A_s^{\mathrm{q}}(x) \setminus B_s(x)$ as long as this set is not small and halt with failure otherwise.

Algorithm 2: RGA

Input: H and $\mathcal{G}$ with partitions satisfying the Setup, a good vertex order τ for X^{buf} on $V(H)$, H_+ returned by Lemma 4.35

$t := 0$
$\phi_0 = \varnothing$ `// start with nothing embedded`
$Q_0 := \varnothing$ `// initial (lifetime) queue`
while $\mathrm{Dom}(\phi_t) \neq X^{\mathrm{main}}$ **do**
  Let $x \in V(H) \setminus \mathrm{Dom}(\phi_t)$ be the next vertex in the order τ
  if $x \in Q_t$ *and* $|A_t^{\mathrm{q}}(x) \setminus B_t(x)| < \frac{1}{8}\mu|\mathcal{C}_t(x)|$ **then**
    halt with failure
  Choose v uniformly at random in $\begin{cases} A_t^{\mathrm{main}}(x) \setminus B_t(x) \text{ if } x \notin Q_t \\ A_t^{\mathrm{q}}(x) \setminus B_t(x) \text{ if } x \in Q_t \end{cases}$
  $\phi_{t+1} := \phi_t \cup \{x \to v\}$
  $Q_{t+1} := Q_t$
  for $y \in V(H) \setminus \mathrm{Dom}\,\phi_{t+1}$ **do**
    if $|A_{t+1}^{\mathrm{main}}(y)| < (1-2\varepsilon)^{\pi_{t+1}(y)}\mu|\mathcal{C}_{t+1}^{\mathrm{main}}(y)|$ **then**
      $Q_{t+1} := Q_{t+1} \cup \{y\}$
  $t \leftarrow t+1$
$t_{\mathrm{RGA}} := t$

Proof of Lemma 4.62. Let $\Delta_{\mathrm{aux}} := 2^{2^{\Delta_R^2+1}+\Delta^2+1}(\Delta+1)\Delta$. We require

$$\mu \le \left(2^{2^{\Delta_R^3+1}+2\Delta^2+8}\kappa\Delta_R\right)^{-1}, \qquad \rho \le \frac{\mu^{\Delta^2+2}}{2^{2^{\Delta_R^3+1}+4\Delta^2+10}\Delta^3},$$

$$c \ge \Delta_{\mathrm{aux}}+2, \quad \varepsilon \le \left(\frac{\mu 2^{-4/\rho}}{c\Delta 2^{\Delta^3+4}}\right)^{2^{\Delta^2}} \quad \text{and} \qquad \eta \le \frac{\varepsilon^7}{2^{14}}.$$

We require η' to be small enough for Lemma 4.38 with inputs k, Δ, Δ_R, c and η. We require

$$n_0 \ge 2^{2^{\Delta_R^3+1}+2^{\Delta^2+2}+4\Delta^2+24\Delta+12}10^4\kappa^6\Delta^4|J|^2\rho^{-4}\mu^{-4}.$$

Let $\varepsilon_1 = \Delta^{1/7}\varepsilon$, $\varepsilon_2 = \Delta^{2/7}\varepsilon$ and $\varepsilon_3 = 2^{\Delta^2/7}\varepsilon$.

Assume Setup 4.37. We apply Lemmas 4.34 and 4.35 to obtain J-partite k-complexes $\overline{H}$ with a partition $\overline{\mathcal{X}}$ of $V(\overline{H})$ and a linear order $\overline{\tau}$ on $V(\overline{H})$ satisfying (AQ1)–(AQ4) and H_+ with a partition $\mathcal{X}^+$ of $V(H_+)$ and a linear order τ_+ on $V(H_+)$ satisfying (AM1)–(AM3) respectively. The conditions of Lemma 4.38 are satisfied, so the conclusions of Lemma 4.38 hold; in particular, by Lemma 4.38(i) $\mathcal{G}$ is an (η, c)-THC graph for $(\overline{H}, \overline{\tau})$, (H_+, τ_+) and (H, τ) with density weighted hypergraph $\mathcal{D}$. We run Algorithm RGA, thereby building up a sequence (ϕ_t) of good partial partite homomorphisms and a sequence (Q_t) of subsets of $V(H)$; let $B_t(x)$ denote the set of bad vertices with respect to ϕ_t, H_+ and Q_t. Let T be the time at which Algorithm RGA terminates. We first collect several facts about the running of Algorithm RGA.

Claim 4.64. *The following hold at each time t when Algorithm* RGA *is running.*

(INV1) *ϕ_t is a good partial partite homomorphism from H to $\mathcal{G}'$ which is* THC-*respecting for $(\mathcal{G}^+, H_+, \tau_+)$, $(\overline{\mathcal{G}}, \overline{H}, \overline{\tau})$ and $(\mathcal{G}, H, \tau)$.*

(INV2) *For each $x \in V(H) \setminus \mathrm{Dom}(\phi_t)$, either $x \in Q_t$ or we have*

$$|A_t^{\mathrm{main}}(x)| \geq (1 - 2\varepsilon)^{\pi_t(x)} \mu |\mathcal{C}_t^{\mathrm{main}}(x)|.$$

(INV3) *When we embed x to create ϕ_{t+1}, we do so uniformly at random into a set of size at least $\frac{1}{8}\mu|\mathcal{C}_t(x)|$.*

Proof. We require $5\Delta\varepsilon \leq \frac{\mu}{8}$. Algorithm RGA maintains (INV2) by definition. Since $\mathcal{G}$ is an (η, c)-THC graph for (H_+, τ_+) and by the definition of Algorithm RGA and bad vertices, ϕ_t is a good partial partite homomorphism from H to $\mathcal{G}'$ which is THC-respecting for $(\mathcal{G}^+, H_+, \tau_+)$. Now as a consequence of Lemma 4.22 for $(\mathcal{G}^+, H_+, \tau_+)$ compared with $(\overline{\mathcal{G}}, \overline{H}, \overline{\tau})$ and $(\mathcal{G}, H, \tau)$, we conclude that ϕ_t is THC-respecting for $(\overline{\mathcal{G}}, \overline{H}, \overline{\tau})$ and $(\mathcal{G}, H, \tau)$. Hence, Algorithm RGA maintains (INV1). When we embed x to create ϕ_{t+1}, either $x \in Q_t$ or not. In the former case, we have (INV3) because Algorithm RGA has not failed. In the latter case, by (INV2) and (GPH2) we have $|A_t^{\mathrm{main}}(x)| \geq (1 - 2\varepsilon)^{\pi_t(x)} \mu |\mathcal{C}_t^{\mathrm{main}}(x)| \geq (1 - 2\varepsilon)^{2\pi_t(x)}(1 - 2\mu)\mu|\mathcal{C}_t(x)|$. By Lemma 4.41 we have $|B_t(x)| \leq 5\Delta\varepsilon|\mathcal{C}_t(x)|$, so (INV3) follows. ∎

It turns out that we need a stronger version of (INV3) for neighbours of buffer vertices. We prove the following claim, which tells us that neighbours of buffer vertices have many available candidates and never enter the queue.

Claim 4.65. *Let $x \in X^{\mathrm{buf}}$ and let $y_1, \dots, y_b$ be the neighbours of x in $H^{(2)}$ in the order according to τ. Then for each $i \in [b]$ and $\tau(y_1) - 1 \le t < \tau(y_i)$ we have*

$$|A_t^{\mathrm{main}}(y_i)| \ge \tfrac{3}{4}(1-2\varepsilon)^{\pi_t(y_i)} d_t(y_i)|V_{y_i}| \tag{4.23}$$

and in particular y_i never enters the queue.

Proof. We require $\mu \le \frac{1}{8(1+2\kappa\Delta_R)}$. By (H3), buffer vertices are at distance at least five in $H^{(2)}$ and so neighbours of distinct buffer vertices are at distance at least three in $H^{(2)}$; hence, at time $\tau(y_1) - 1$ the available candidate set of each y_i is $A^{\mathrm{main}}_{\tau(y_1)-1}(y_i) = V^{\mathrm{main}}_{y_i} \setminus \mathrm{Im}(\phi_{\tau(y_1)-1})$. The size of $A^{\mathrm{main}}_{\tau(y_1)-1}(y_i)$ is by (H4), (G1) and the choice of μ at least

$$|V^{\mathrm{main}}_{y_i}| - 4\kappa\Delta_R\mu|V_{y_i}| = (1 - 2\mu - 4\kappa\Delta_R\mu)|V_{y_i}| \ge \tfrac{3}{4}|V_{y_i}|,$$

so y_i is not added to Q_t for any $t < \tau(y_1)$. By (H2) and because τ is a good vertex order for X^{buf} on $V(H)$, the vertices $y_1, \dots, y_b$ are embedded consecutively into clusters of $\mathcal{G}'$ which correspond to distinct clusters of $\mathcal{G}$. By Definition 4.40 of bad vertices with respect to ϕ and Q, for each time t with $\tau(y_1) - 1 \le t < \tau(y_i)$ we have

$$|A_t^{\mathrm{main}}(y_i)| \ge \tfrac{3}{4}(1-2\varepsilon)^{\pi_t(y_i)} d_t(y_i)|V_{y_i}|,$$

which gives (4.23) and that y_i never enters the queue. ∎

The following claim gives a stronger version of (INV3) for the neighbours of buffer vertices. Let $T_0 := |N_{H^{(2)}}(X^{\mathrm{buf}})|$.

Claim 4.66. *We have $Q_{T_0-1} = \varnothing$ and $T > T_0$. Moreover, for each time $s \le T_0$, when we embed z to create ϕ_{s+1} we do so uniformly at random into the set $A_s^{\mathrm{main}}(z) \setminus B_s(z)$ of size at least $\frac{1}{2}\max(|\mathcal{C}_s(z)|, d'_s(z)|V_z|)$.*

Proof. Since τ is a good vertex order for X^{buf} on $V(H)$, the first T_0 vertices to be embedded by Algorithm RGA are all in $N_{H^{(2)}}(X^{\mathrm{buf}})$. By Claim 4.65 none of these vertices enter the queue, so we have $Q_{T_0-1} = \varnothing$ and for each $s \le T_0$ the sth vertex z is embedded uniformly at random into $A_s^{\mathrm{main}}(z) \setminus B_s(z)$ to create ϕ_{s+1}. Algorithm RGA can halt with failure only when a queue vertex is being embedded, so in particular it can terminate only after time T_0. Now by putting together (4.23), Lemma 4.41 and (INV1) with (THC1), we find that, for $s \le T_0$ and the sth vertex z, we have

$$\begin{aligned}|A_s^{\mathrm{main}}(z) \setminus B_s(z)| &\ge \tfrac{3}{4}(1-2\varepsilon)^{\pi_s(z)} d'_s(z)|V_z| - 5\Delta\varepsilon|\mathcal{C}_s(z)| \\ &\ge \tfrac{1}{2}\max(|\mathcal{C}_s(z)|, d'_s(z)|V_z|)\end{aligned}$$

as desired. ∎

The following claim tells us that there are no dense spots in our embedding of the neighbours of any one buffer vertex.

Claim 4.67. *Let $j \in J$, $v \in V_j$ and $x \in X_j^{\mathrm{buf}}$. Let $y_1, \ldots, y_b$ be the elements of $N_{H^{(2)}}(x)$ in the order according to τ. Then, conditioning on the history up to the time right before the embedding of y_1, we embed $H^{-1}(x)$ to each element of $\mathcal{C}_0(H^{-1}(x))$ with probability at most $\frac{2^b}{a_j|U_j|}$.*

Proof. For each $i \in [b]_0$ set $H_i := H[\{y_1, \ldots, y_i\}]$. By Claim 4.66 each y_i is embedded uniformly at random into a subset of $\mathcal{C}_{\tau(y_i)-1}(y_i)$ of size at least $\frac{\mathcal{D}(H_i)}{2\mathcal{D}(H_{i-1})}|V_{y_i}|$. By multiplying the conditional probabilities, we find that we embed $H^{-1}(x)$ to each element of $\mathcal{C}_{\tau(y_1)-1}(H^{-1}(x)) = \mathcal{C}_0(H^{-1}(x))$ with probability at most $\frac{2^b}{a_j|U_j|}$, conditioning on the history up to time $\tau(y_1) - 1$. ∎

We establish a lower bound on certain relevant complex-derived densities in $\mathcal{D}$, which we will need for certain bad event probabilities to go to zero.

Claim 4.68. *Given a complex F on at most $\Delta+1$ vertices, a partition $\mathcal{F} = \{F_j\}_{j\in J}$ of $V(F)$ and a vertex $x \in V(F)$ such that $(F, \mathcal{F})$ is an R-partition and $F_x = \{x\}$, setting $F_0 := F[V(F) \setminus \{x\}]$, we have*

$$\frac{\mathcal{D}(F)}{\mathcal{D}(F_0)} \ge 2\left(\frac{10\kappa|J|\Delta\log n}{\rho n}\right)^{1/2^{\Delta^2+1}}.$$

Proof. We require $n \ge 2^{2^{\Delta^2+4}} 10^4 \kappa^6 |J|^2 \Delta^4 \rho^{-4}$. Apply Lemma 4.26 with $2^{\Delta^2}c$ to obtain

$$\frac{\mathcal{D}(F)}{\mathcal{D}(F_0)} \ge \left(\frac{1-\eta}{(1+\eta)|V_{V(F)}|}\right)^{\frac{1}{2^{\Delta^2}c}} \ge \left(\frac{2\kappa n}{|J|}\right)^{-\frac{\Delta+1}{2^{\Delta^2}c}} \ge 2\left(\frac{10\kappa|J|\Delta\log n}{\rho n}\right)^{1/2^{\Delta^2+1}}$$

as desired. ∎

In the following claim we show that candidate sets for individual vertices never become too small throughout the embedding procedure. This follows immediately from Claim 4.68.

Claim 4.69. *For each $x \in V(H)$ and at each time $t \le \tau(x) - 1$ and before the termination of Algorithm* RGA*, we have $\mathcal{G}'_t(x) \ge \left(\frac{10\kappa|J|\Delta \log n}{\rho n}\right)^{1/2^{\Delta^2+1}}$.*

Proof. Let $W := \mathrm{Dom}(\phi_t) \cap N^{-1}(x)$. Apply Claim 4.68 with $F := H[W \cup \{x\}]$, x and $F_0 := H[W]$ to obtain $\frac{\mathcal{D}(F)}{\mathcal{D}(F_0)} \ge 2\left(\frac{10\kappa|J|\Delta \log n}{\rho n}\right)^{1/2^{\Delta^2+1}}$. Since ϕ_t is a THC-respecting partial partite homomorphism from H to $\mathcal{G}'$, we have

$$\mathcal{G}'_t(x) \ge (1-\eta)\frac{\mathcal{D}(F)}{\mathcal{D}(F_0)} \ge \left(\frac{10\kappa|J|\Delta \log n}{\rho n}\right)^{1/2^{\Delta^2+1}}$$

as desired. ∎

To show that Algorithm RGA successfully runs to completion, we need to show that asymptotically almost surely the 'halt with failure' line is never reached; this failure condition is triggered when a queue vertex has too few available queue candidates. We shall show in Claim 4.71 that asymptotically almost surely there are always many available queue candidates. To do so, we first show in Claim 4.70 that the sum of the conditional probabilities of embedding vertices into potential queue candidate sets is reasonably close to its expected value and apply Lemma 4.12.

Observe that since Algorithm RGA preserves (INV3), the conditions of Lemmas 4.42 and 4.48 are met. By (H4), (G3) and Claim 4.66, the conditions of Lemma 4.55 are also met. Thus with probability at least

$$1 - |J|2^{-n/(\kappa|J|)} - 2^{2^{\Delta_R^3+1}}\left(\Delta n^{\Delta+1}e^{-\frac{\rho n}{8\kappa|J|}} + \Delta^5 n^{1-9\Delta} + \kappa\Delta_R\Delta^4 n^{1-10\Delta}\right),$$

the events $\mathcal{E}$, $\mathcal{E}^*$ and the following event $\mathcal{E}_{\mathrm{RGA}}$ hold. For every $j \in J$ and every subset $W \subseteq V_j$ with $|W| \ge \rho|V_j|$, the number of vertices $x \in X_j$ such that there exists a time $t = t(x)$ at which we have $|\mathcal{C}_t(x) \cap W_{j\to x}| < (1-2\varepsilon)^{\pi_t(x)}\frac{|\mathcal{C}_t(x)||W|}{|V_j|}$ and x is unembedded is at most $\rho|X_j|$.

Claim 4.70. *Suppose that the events $\mathcal{E}_{\mathrm{RGA}}$ and $\mathcal{E}$ hold. Then for any $j \in J$, $x \in X_j$ and tuple $\vec{v} = (\psi(z))_{z\in\mathrm{Dom}(\psi)}$ of vertices in $\mathcal{G}'$ for some partial partite homomorphism ψ from $H^{-1}(x)$ to $\mathcal{G}'$ which is* THC*-extendable for $(\mathcal{G}', H, \tau)$ such that $\mathcal{G}'_{\vec{v}}(x) \ge \left(\frac{10\kappa|J|\mu\Delta\log n}{\rho n}\right)^{1/2^{\Delta^2-1}}$ and for all $y \in \Xi_j$ we have that $\mathcal{G}^{x,\vec{v},y}$ is an (η, c)-*THC *graph with the linear order on $N^{\le 2}(y)$ induced by τ and density weighted hypergraph $\mathcal{D}^{x,\vec{v},y}$, we have*

$$\sum_{y\in Q_t\cap X_j:\tau(y)\le t}\frac{|\mathcal{C}_{\tau(y)-1}(y) \cap \mathcal{C}_{\vec{v}}(x)|}{|\mathcal{C}_{\tau(y)-1}(y)|} \le 2^{2^{\Delta_R^3+1}+4\Delta^2+3}\Delta^3\rho\mu^{-\Delta^2}|\mathcal{C}_{\vec{v}}(x)|$$

for any $t \le T$*.*

Proof. Let $j \in J$, $x \in X_j$ and $\vec{v}$ be a tuple of vertices in $\mathcal{G}'$ satisfying the conditions of the claim. Let $U := \mathcal{C}_{\vec{v}}(x)$. Since the sum is monotonically increasing in t and the upper bound is independent of t, it is enough to show that it holds at time T.

We first establish that $|Q_T \cap X_j| \le \rho|X_j|$. Set $W := V_j^{\mathrm{main}} \setminus (\mathrm{Im}(\phi_T))_{\to\mathcal{G}}$. We have $|X_j^{\mathrm{main}}| = (1-4\mu)|X_j|$ and $|V_j^{\mathrm{main}}| = (1-2\mu)|V_j|$ so $|W| \ge \mu|V_j| \ge \rho|V_j|$. Suppose that $x \in Q_T \cap X_j$. Then there is a first time t at which $x \in Q_t$. Since $A_t^{\mathrm{main}}(x) \supseteq \mathcal{C}_t(x) \cap W_{j\to x}$, by construction of Q_t in Algorithm RGA and (GPH2) we have

$$|\mathcal{C}_t(x) \cap W_{j\to x}| < (1-2\varepsilon)^{\pi_t(x)}\mu|\mathcal{C}_t(x)| \le (1-2\varepsilon)^{\pi_t(x)}\frac{|\mathcal{C}_t(x)||W|}{|V_j|}$$

so x satisfies (4.8) of Lemma 4.42. Since $|W| \ge \rho|V_j|$ and $\mathcal{E}_{\mathrm{RGA}}$ holds, we deduce that the number of vertices $x \in Q_T \cap X_j$ is at most $\rho|X_j|$.

Now since $\mathcal{E}_{j,x,\vec{v},h,n^{(2)}_{j,h}}$ holds for each $h \in [\ell_j]$, we have

$$\begin{aligned}\sum_{y\in E^{x,\vec{v}}\cap X_j:\tau(y)\le T}\frac{|\mathcal{C}_{\tau(y)-1}(y)\cap U|}{|\mathcal{C}_{\tau(y)-1}(y)|} &\le \sum_{h\in[\ell_j]} 2S_{x,\vec{v},j,h,n^{(2)}_{j,h},1} \\ &\le 2^{2^{\Delta_R^3+1}+4\Delta^2+2}\Delta^3\rho\mu^{-\Delta^2}|U|.\end{aligned} \tag{4.24}$$

For $y \in X_j \setminus E^{x,\vec{v}}$ such that $\tau(y) \le T$, the vertices in $N^{<2}(y)$ are embedded to a tuple in $T^{x,\vec{v},y}$, so by Lemma 4.43 we have $\frac{|\mathcal{C}_{\tau(y)-1}(y)\cap U|}{|\mathcal{C}_{\tau(y)-1}(y)|} = (1\pm 4\eta)\frac{|U|}{|V_j|}$. Combining this with (4.24) and the fact that $|Q_T \cap X_j| \le \rho|X_j|$, we obtain

$$\begin{aligned}&\sum_{y\in Q_T\cap X_j:\tau(y)\le T}\frac{|\mathcal{C}_{\tau(y)-1}(y)\cap U|}{|\mathcal{C}_{\tau(y)-1}(y)|} \\ &\le \sum_{y\in (Q_T\cap X_j)\setminus E^{x,\vec{v}}:\tau(y)\le T}\frac{|\mathcal{C}_{\tau(y)-1}(y)\cap U|}{|\mathcal{C}_{\tau(y)-1}(y)|} + \sum_{y\in E^{x,\vec{v}}\cap X_j:\tau(y)\le T}\frac{|\mathcal{C}_{\tau(y)-1}(y)\cap U|}{|\mathcal{C}_{\tau(y)-1}(y)|} \\ &\le 2\rho|U| + 2^{2^{\Delta_R^3+1}+4\Delta^2+2}\Delta^3\rho\mu^{-\Delta^2}|U| \le 2^{2^{\Delta_R^3+1}+4\Delta^2+3}\Delta^3\rho\mu^{-\Delta^2}|U|,\end{aligned}$$

completing the proof. ∎

Now we show that asymptotically almost surely there are many available queue candidates.

Claim 4.71. *Asymptotically almost surely for each* $x \in V(H)$ *and at each time* $t \le \min(\tau(x)-1, T)$*, we have* $|\mathcal{C}_t^{\mathrm{q}}(x) \cap \mathrm{Im}(\phi_t)| < \frac{1}{2}|\mathcal{C}_t^{\mathrm{q}}(x)|$*.*

Proof. We require $\rho \le \frac{\mu^{\Delta^2+2}}{2^{2^{\Delta_R^3+1}+4\Delta^2+10}\Delta^3}$. Suppose that the events $\mathcal{E}_{\mathrm{RGA}}$ and $\mathcal{E}$ hold. Let $j \in J$, $x \in X_j$ and $t \le \min(\tau(x)-1, T)$. By (INV1) ϕ_t is a partial partite homomorphism from H to $\mathcal{G}'$ which is THC-respecting for $(\overline{\mathcal{G}}, \overline{H}, \overline{\tau})$ and we have $\mathcal{C}^{\mathrm{q}}_t(x) \subseteq \mathcal{C}_t(x) = \mathcal{C}_{\vec{v}}(x)$ with $\vec{v} = (\phi_t(z))_{z \in \mathrm{Dom}(\phi_t) \cap N^{-1}(x)}$. By Lemma 4.38(ii) for all $y \in X_j$ we have that $\mathcal{G}^{x,\vec{v},y}$ is an (η, c)-THC graph with the linear order on $N^{\le 2}(y)$ induced by τ and density weighted hypergraph $\mathcal{D}^{x,\vec{v},y}$. By Claim 4.69 we have

$$\mathcal{G}'_{\vec{v}}(x) = \mathcal{G}'_t(x) \ge \left(\frac{10\kappa|J|\Delta\log n}{\rho n}\right)^{1/2^{\Delta^2+1}}.$$

Hence, by Claim 4.70 we obtain

$$\sum_{y \in Q_t \cap X_j^{\mathrm{main}} : \tau(y) \le t} \frac{|\mathcal{C}_{\tau(y)-1}(y) \cap \mathcal{C}_t(x)|}{|\mathcal{C}_{\tau(y)-1}(y)|} \le 2^{2^{\Delta_R^3+1}+4\Delta^2+3}\Delta^3\rho\mu^{-\Delta^2}|\mathcal{C}_t(x)|.$$

Since Algorithm RGA has not terminated, by the definition of Algorithm RGA we have $|A^{\mathrm{q}}_{\tau(y)-1}(y) \setminus B_{\tau(y)-1}(y)| \ge \frac{1}{8}\mu|\mathcal{C}_{\tau(y)-1}(y)|$, so we obtain

$$\sum_{y \in Q_t \cap X_j^{\mathrm{main}} : \tau(y) \le t} \frac{|\mathcal{C}_{\tau(y)-1}(y) \cap \mathcal{C}^{\mathrm{q}}_t(x)|}{|A^{\mathrm{q}}_{\tau(y)-1}(y) \setminus B_{\tau(y)-1}(y)|} \le \frac{2^{2^{\Delta_R^3+1}+4\Delta^2+6}\Delta^3\rho|\mathcal{C}_t(x)|}{\mu^{\Delta^2+1}}.$$

Note that the summand in the inequality above gives an upper bound for the probability of embedding $y \in Q_t$ to $\mathcal{C}^{\mathrm{q}}_t(x)$, conditioning on the history up to but not including the embedding of y. Furthermore, for $y \notin Q_t$ the probability of embedding to $\mathcal{C}^{\mathrm{q}}_t(x)$ is zero. Applying Lemma 4.12 with $R = 8\mu^{-1}$, we find that the probability that more than

$$\frac{2^{2^{\Delta_R^3+1}+4\Delta^2+7}\Delta^3\rho|\mathcal{C}_t(x)|}{\mu^{\Delta^2+1}} \le \frac{2^{2^{\Delta_R^3+1}+4\Delta^2+8}\Delta^3\rho|\mathcal{C}^{\mathrm{q}}_t(x)|}{\mu^{\Delta^2+2}}$$

vertices y from X_j^{main} are embedded into $\mathcal{C}^{\mathrm{q}}_t(x)$, and both $\mathcal{E}_{\mathrm{RGA}}$ and $\mathcal{E}$ hold, is at most

$$\exp\left(-\frac{2^{2^{\Delta_R^3+1}+4\Delta^2+1}\Delta^3\rho|V_j|}{\mu^{\Delta^2}}\left(\frac{10\kappa|J|\Delta\log n}{\rho n}\right)^{1/2^{\Delta^2+1}}\right).$$

If this bad event does not occur, then we have

$$|\mathcal{C}^{\mathrm{q}}_t(x) \cap \mathrm{Im}(\phi_t)| \le \frac{2^{2^{\Delta_R^3+1}+4\Delta^2+8}\Delta^3\rho|\mathcal{C}^{\mathrm{q}}_t(x)|}{\mu^{\Delta^2+2}} < \frac{1}{2}|\mathcal{C}^{\mathrm{q}}_t(x)|$$

as desired. Now the probability that such a bad event occurs for some $x \in X^{\mathrm{main}}$ and $t \le \tau(x) - 1$, or that at least one of $\mathcal{E}_{\mathrm{RGA}}$ and $\mathcal{E}$ does not hold, is at most

$$|J|2^{-n/(\kappa|J|)} + 2^{2^{\Delta_R^3+1}}\left(\Delta n^{\Delta+1}e^{-\frac{\rho n}{8\kappa|J|}} + \Delta^5 n^{1-9\Delta}\right)$$
$$+ n^2 \exp\left(-\frac{2^{2^{\Delta_R^3+1}+4\Delta^2+1}\Delta^3\rho|V_j|}{\mu^{\Delta^2}}\left(\frac{10\kappa|J|\Delta\log n}{\rho n}\right)^{1/2^{\Delta^2+1}}\right),$$

which tends to zero as n goes to infinity, completing the proof. ∎

We show that our random greedy algorithm completes successfully.

Claim 4.72. *Asymptotically almost surely Algorithm* RGA *does not halt with failure.*

Proof. Suppose that the good event of Claim 4.71 holds. Let $x \in X^{\mathrm{main}}$. By (INV1) $\phi_{\tau(x)-1}$ is a THC-respecting good partial partite homomorphism from H to $\mathcal{G}'$. By the good event of Claim 4.71, (GPH2) and Lemma 4.41 we have

$$\begin{aligned}|A^{\mathrm{q}}_{\tau(x)-1}(x) \setminus B_{\tau(x)-1}(x)| &\ge \tfrac{1}{2}(1-2\varepsilon)^{\pi^\tau(x)}\mu|\mathcal{C}_{\tau(x)-1}(x)| - 5\Delta\varepsilon|\mathcal{C}_{\tau(x)-1}(x)| \\ &\ge \tfrac{1}{8}\mu|\mathcal{C}_{\tau(x)-1}(x)|.\end{aligned}$$

Hence, we never reach the 'halt with failure' line of Algorithm RGA. ∎

Now suppose that $\mathcal{E}^*$ and the good event of Claim 4.72 holds. (RGA2) follows from Claim 4.68 by choice of n_0 and (RGA3) follows from $\mathcal{E}_{\mathrm{RGA}}$ because $X_j^{\mathrm{buf}} \subseteq X_j$ is an F_j-buffer. Since Algorithm RGA successfully completes, we have $T = |X^{\mathrm{main}}| = (1-4\mu)n$ and $\mathrm{Dom}(\phi_{\mathrm{RGA}}) \cap X_j = X_j^{\mathrm{main}}$ for each $j \in J$. Let $j \in J$. Since we embed each $x \in X_j^{\mathrm{main}}$ into a subset of $A^{\mathrm{main}}(x) \cup A^{\mathrm{q}}(x)$, it follows that for each $x \in X_j^{\mathrm{main}}$ we have $\phi_{\mathrm{RGA}}(x)_{\to\mathcal{G}} \in V_j^{\mathrm{main}} \cup V_j^{\mathrm{q}}$ and for every $x, y \in X_j^{\mathrm{main}}$ we have $\phi_{\mathrm{RGA}}(x)_{\to\mathcal{G}} = \phi_{\mathrm{RGA}}(y)_{\to\mathcal{G}}$ if and only if $x = y$. Let $e \in E(H[\mathrm{Dom}(\phi_{\mathrm{RGA}}]))$. If $e = \varnothing$, then $\phi_{\mathrm{RGA}}(e)_{\to\mathcal{G}} = \varnothing \in E(\mathcal{G})$ holds automatically. Now consider when $e \ne \varnothing$ and let x_e be the last element of e in the order according to τ. Since $\phi_{\mathrm{RGA}}(x_e) \in \mathcal{C}_{\tau(x_e)-1}(x_e)$, we have

$$1 = g'_{\tau(x_e)-1}(\phi_{\mathrm{RGA}}(x_e)) = \prod_{A \subseteq \mathrm{Dom}(\phi_{\tau(x_e)-1})} g'(\phi_{\tau(x_e)-1}(A) \cup \{\phi_{\mathrm{RGA}}(x_e)\}).$$

Since $\mathcal{G}'$ is a binary weighted hypergraph, we have $g'(\phi_{\tau(x_e)-1}(A) \cup \{\phi_{\mathrm{RGA}}(x_e)\}) = 1$ for all $A \subseteq \mathrm{Dom}(\phi_{\tau(x_e)-1})$. In particular, by setting $A = e \setminus \{x_e\}$ and the definition of standard construction, we have $g(\phi(e)) = g'(\phi_{\tau(x_e)-1}(A) \cup \{\phi_{\mathrm{RGA}}(x_e)\}) = 1$. Hence, we have $\phi_{\mathrm{RGA}}(e)_{\to\mathcal{G}} \in E(\mathcal{G})$, thereby establishing that (RGA1) holds.

Now we shall verify (RGA4). Let $j \in J$ and $x \in X_j^{\mathrm{buf}}$. By (INV1) ϕ_T is a partial partite homomorphism from H to $\mathcal{G}'$ which is THC-respecting for $(\overline{\mathcal{G}}, \overline{H}, \overline{\tau})$ and let $\vec{v} = (\phi_T(z))_{z \in N^{-1}(x)}$. By Claim 4.69 we have

$$\mathcal{G}'_{\vec{v}}(x) = \mathcal{G}'_T(x) \geq \left(\frac{10\kappa |J| \Delta \log n}{\rho n}\right)^{1/2^{\Delta^2+1}}.$$

By Lemma 4.38(ii) for all $y \in X_j$ we have that $\mathcal{G}^{x,\vec{v},y}$ is an (η, c)-THC graph with the linear order on $N^{\leq 2}(y)$ induced by τ and density weighted hypergraph $\mathcal{D}^{x,\vec{v},y}$. Now since $\mathcal{E}_{j,x,\vec{v},h,n_{j,h}^{(2)}}$ holds for each $h \in [\ell_j]$, we have

$$\begin{aligned}\sum_{y \in E^{x,\vec{v}} \cap X_j^{\mathrm{buf}}} \frac{|\mathcal{C}_T(y) \cap \mathcal{C}_T(x)|}{|\mathcal{C}_T(y)|} &\leq \sum_{h \in [\ell_j]} 2 S_{x,\vec{v},j,h,n_{j,h}^{(2)},1} \\ &\leq 2^{2^{\Delta_R^3+1}+4\Delta^2+2} \Delta^3 \rho \mu^{-\Delta^2} |\mathcal{C}_T(x)|.\end{aligned}$$

By (GPH2) for each $y \in X_j^{\mathrm{buf}}$ we have $|\mathcal{C}_T^{\mathrm{buf}}(y)| \geq \frac{\mu}{2} |\mathcal{C}_T(y)|$, so for each $y \in Y_x$ we have $|\mathcal{C}_T(y) \cap \mathcal{C}_T(x)| \geq |\mathcal{C}_T^{\mathrm{buf}}(y) \cap \mathcal{C}_T^{\mathrm{buf}}(x)| > (1+4\eta) \frac{4|\mathcal{C}_T^{\mathrm{buf}}(y)||\mathcal{C}_T^{\mathrm{buf}}(x)|}{\mu^2 |V_j|} \geq (1+4\eta) \frac{|\mathcal{C}_T(y)||\mathcal{C}_T(x)|}{|V_j|}$. Then, by Lemma 4.43 we have $y \in E^{x,\vec{v}} \cap X_j^{\mathrm{buf}}$ and therefore (RGA4) follows.

It remains to verify (RGA5) and (RGA6). We first show that neighbourhoods do not become overly occupied by neighbours of buffer vertices.

Claim 4.73. *Asymptotically almost surely for each $i, j \in J$ such that $ij \in E(R')$ and $v \in V_i$, we have* $\deg_{\mathcal{G}^{(2)}}\left(v; V_j^{\mathrm{main}} \setminus \mathrm{Im}\,(\phi_{T_0})\right) \geq \frac{1}{2} \deg_{\mathcal{G}^{(2)}}(v; V_j)$.

Proof. We require

$$\mu \leq \left(2^{2^{\Delta_R^3+1}+2\Delta^2+8} \kappa \Delta_R\right)^{-1}.$$

Suppose that the event $\mathcal{E}^*$ holds. Let $i, j \in J$ satisfy $ij \in E(R')$ and $v \in V_i$. Let $U := N_{\mathcal{G}^{(2)}}(v; V_j) \supseteq N_{\mathcal{G}^{(2)}}(v; V_j^{\mathrm{main}}) =: U'$. By (G3) and Claim 4.68 we have $|U| \geq \left(\frac{10\kappa|J|\Delta \log n}{\rho n}\right)^{1/2^{\Delta^2+1}} |V_j|$. Now $\mathcal{E}_{j,v,h,n_{j,h}^{(2)}}$ holds for all $h \in [\ell_j]$, so we have

$$\begin{aligned}\sum_{y \in N_{H^{(2)}}(X^{\mathrm{buf}}) \cap X_j} \frac{\left|\mathcal{C}_{\tau(y)-1}(y) \cap U\right|}{\left|\mathcal{C}_{\tau(y)-1}(y)\right|} &\leq 2 \sum_{h \in [\ell_j]} S_{v,j,h,n_{j,h}^{(2)},1} \\ &\leq 2^{2^{\Delta_R^3+1}+2\Delta^2+4} \kappa \Delta_R \mu |U|.\end{aligned}$$

By Claim 4.66 $\left|A^{\mathrm{main}}_{\tau(y)-1}(y) \setminus B_{\tau(y)-1}(y)\right| \geq \frac{1}{2}|\mathcal{C}_{\tau(y)-1}(y)|$ holds for all $y \in N_{H^{(2)}}(X^{\mathrm{buf}}) \cap X_j$, so we obtain

$$\sum_{y \in N_{H^{(2)}}(X^{\mathrm{buf}}) \cap X_j} \frac{\left|\mathcal{C}_{\tau(y)-1}(y) \cap U\right|}{\left|A^{\mathrm{main}}_{\tau(y)-1}(y) \setminus B_{\tau(y)-1}(y)\right|} \leq 2^{2^{\Delta_R^3+1}+2\Delta^2+5}\kappa\Delta_R\mu|U|.$$

Note that the summand in the inequality above gives an upper bound for the probability of embedding $y \in N_{H^{(2)}}(X^{\mathrm{buf}}) \cap X_j$ to U, conditioning on the history up to but not including the embedding of y. Furthermore, for $y \in N_{H^{(2)}}(X^{\mathrm{buf}}) \setminus X_j$ the probability of embedding to U is zero. Applying Lemma 4.12 with $R = 2$, we find that the probability that more than $2^{2^{\Delta_R^3+1}+2\Delta^2+6}\kappa\Delta_R\mu|U|$ vertices y from $N_{H^{(2)}}(X^{\mathrm{buf}}) \cap X_j$ are embedded into U is at most $\exp\left(-2^{2^{\Delta_R^3+1}+2\Delta^2+2}\kappa\Delta_R\mu|U|\right)$. Now noting that $|V_j| \geq \frac{n}{\kappa|J|}$, this probability is at most

$$\exp\left(-2^{2^{\Delta_R^3+1}+2\Delta^2+2}\frac{\Delta_R\mu n}{|J|}\left(\frac{10\kappa|J|\Delta\log n}{\rho n}\right)^{1/2^{\Delta^2+1}}\right).$$

If this bad event does not occur, then by (G4) we have

$$|U' \cap \mathrm{Im}(\phi_{T_0})| \leq |U \cap \mathrm{Im}(\phi_{T_0})| \leq 2^{2^{\Delta_R^3+1}+2\Delta^2+6}\kappa\Delta_R\mu|U| \leq \tfrac{1}{4}|U| \leq \tfrac{1}{2}|U'|,$$

which gives the desired outcome. Now the probability that a bad event occurs for some $ij \in E(R')$ and $v \in V_i$, and both $\mathcal{E}^*$ and the good event of Claim 4.72 hold, is at most

$$n^2\exp\left(-2^{2^{\Delta_R^3+1}+2\Delta^2+2}\frac{\Delta_R\mu n}{|J|}\left(\frac{10\kappa|J|\Delta\log n}{\rho n}\right)^{1/2^{\Delta^2+1}}\right),$$

which tends to zero as n goes to infinity, completing the proof. ∎

Now suppose that the good event of Claim 4.73 holds. Let $j \in J$ and $v \in V_j$. Enumerate X_j^{buf} as $x_1, \ldots, x_h$ in the order according to τ. Let $i \in [h]$. Enumerate $N_{H^{(2)}}(x_i)$ as $y_{i1}, \ldots, y_{ib}$ in the order according to τ. We apply Lemma 4.36 to obtain a J-partite k-complex H_{x_i} with a partition $\mathcal{X}^{x_i}$ of $V(H_{x_i})$ and a linear order τ_{x_i} on $V(H_{x_i})$ satisfying (AB1)–(AB3). By Lemma 4.38(iii) $\mathcal{G}^{v,x_i,*}$ is an (η, c)-THC graph with the linear order τ_{x_i} and density weighted hypergraph $\mathcal{D}^{v,x_i,*}$.

Enumerate the first $\tau(y_{i1}) - 1$ vertices in $V(H)$ as $z_1, \ldots, z_{\tau(y_{i1})-1}$ in the order according to τ. By (INV1) $\phi_{\tau(y_{i1})-1}$ is THC-respecting for $(\mathcal{G}^+, H_+, \tau_+)$, so for each $i \in [\tau(y_{i1}) - 1]$ the vertex $\phi(z_i)$ has weight 1 in $\mathcal{G}_i^+$ and belongs to the set V'_{x_i} of (THC2) returned by an algorithm whose existence is guaranteed by (THC3) for $\mathcal{G}^+_{i-1}$. Now since the input into the algorithm is the same for both $\mathcal{G}^+$ and $\mathcal{G}^{v,x_i,*}$ while we embed the vertices in $N_{H^{(2)}}(X^{\mathrm{buf}})$ and the weighted induced subhypergraphs of both $\mathcal{G}^+$ and $\mathcal{G}^{v,x_i,*}$ on the clusters associated with $N_{H^{(2)}}(X^{\mathrm{buf}})$ are identical, it follows that $\phi_{\tau(y_{i1})-1}$ is THC-respecting for $(\mathcal{G}^{v,x_i,*}, H_{x_i}, \tau_{x_i})$.

We have $\deg_{\mathcal{G}^{(2)}}\left(v; V^{\mathrm{main}}_{y_{i\ell}} \setminus \mathrm{Im}\left(\phi_{\tau(y_{i\ell})-1}\right)\right) \geq \frac{1}{2}\deg_{\mathcal{G}^{(2)}}(v; V_{y_{i\ell}})$ for each $\ell \in [b]$ by Claim 4.73 and because $x_i \in X^{\mathrm{buf}}_j \subseteq \widetilde{X}_j$. Since $\mathrm{Dom}(\phi_{\tau(y_{i1})-1}) \subseteq N_{H^{(2)}}(X^{\mathrm{buf}})$ and we have (H3), no vertex at distance $c+3$ or less from x_i in $H^{(2)}$ is embedded in $\phi_{\tau(y_{i1})-1}$. By construction and Claim 4.66, Algorithm RGA generates sequences $\phi_{\tau(y_{i1})}, \dots, \phi_{\tau(y_{ib})}$ and $Q_{\tau(y_{i1})-1}, \dots, Q_{\tau(y_{ib})}$ compatible with the requirements of Lemma 4.61. Hence, we conclude by Lemma 4.61 that, conditioned on the history up to time $\tau(y_{i1})-1$, the probability of embedding $H^{-1}(x_i)$ into $\mathcal{G}^{v,x_i,N^{-1}(x_i)}$ is at least $2^{-(b^2+5b)/2}b_j$. Furthermore, by Claim 4.67 we embed $H^{-1}(x_i)$ to each element of $\mathcal{C}(H^{-1}(x_i))$ with probability at most $\frac{2^b}{a_j|U_j|}$, conditioning on the history up to time $\tau(y_{i1})-1$, so by (G3) the probability of embedding $H^{-1}(x_i)$ into $\mathcal{G}^{v,x_i,N^{-1}(x_i)}$, conditioned on the history up to time $\tau(y_{i1})-1$, is at most $2^{b+1}b_j$.

Now for $i \in [h]$ define the Bernoulli random variable Y_i as follows. Set $Y_i = 1$ if either $H^{-1}(x_i)$ is embedded into $\mathcal{G}^{v,x_i,N^{-1}(x_i)}$ by $\phi_{\tau(y_{ib})}$ or the bad event of Claim 4.73 occurs by time $\tau(y_{i1})-1$. As argued previously, we have $Y_i = 1$ with probability of at least $2^{-(b^2+5b)/2}b_j$ and at most $2^{b+1}b_j$, conditioned on the history up to time $\tau(y_{i1})-1$, so we have $\sum_{i\in[h]} \mathbb{E}[Y_i|\mathcal{F}_{\tau(y_{i1})-1}] \geq 2^{2-(b^2+5b)/2}b_j\mu|X_j|$ and $\sum_{i\in[h]} \mathrm{var}(Y_i|\mathcal{F}_{\tau(y_{i1})-1}) \leq 2^{b+3}b_j\mu|X_j|$. Now we apply Lemma 4.12 to deduce that the probability that v is a candidate for fewer than $2^{1-(b^2+5b)/2}b_j\mu|X_j|$ vertices in X^{buf}_j and the good event of Claim 4.73 occurs is at most $\exp(-2^{-(b^2+6b+2)}b_j\mu|X_j|)$. By taking a union bound over $v \in V(\mathcal{G})$ and applying Claim 4.68, we find that the probability that the good event of Claim 4.73 occurs and that some bad event occurs is at most

$$n \exp\left(-2^{-(\Delta^2+6\Delta+2)}\frac{\mu n}{\kappa|J|}\left(\frac{10\kappa|J|\Delta\log n}{\rho n}\right)^{1/2^{\Delta^2+1}}\right).$$

This tends to zero as n tends to infinity, so the good event holds asymptotically almost surely; when the good event holds we have (RGA5).

It remains to establish (RGA6). We first provide definitions of objects and quantities we will use. Let $j \in J$, $v, w \in V_j$ and $x \in X^{\mathrm{buf}}_j$. Write $s(v,w)$ for the unique value of $\left|\mathcal{C}\left(H^{\leq 1}_{\times 2}(x); (v,w), (x^{(i)})_{i\in[2]}\right)\right|$ for all $x \in X^{\mathrm{buf}}_j$. Set $W_v := \{w \in V_j : s(v,w) >$

$(1+2\varepsilon_1)\,|U_j|a_j b_j^2\}$. The following claim tells us that W_v is not too large and that the total contribution of the vertices in W_v is small.

Claim 4.74. *For $j \in J$ and $v \in V_j$ we have $|W_v| < \varepsilon|V_j|$ and*

$$\sum_{w \in W_v} s(v,w) \le 2\varepsilon_1 |U_j||V_j|a_j b_j^2.$$

Proof. Let $x \in X_j^{\mathrm{buf}}$. Set $\mathcal{C} := \mathcal{C}(H^{\le 1}(x); v, x)$, $\mathcal{I} := \{\{x\}, N^{-1}(x)\}$ and $F := H^{\le 1}_{\times 2}(x)$. (G2) and (G3) give the necessary counting conditions, so by Lemma 4.27 $G := G^{\mathcal{G}'}_{\mathcal{I},F,x^{(2)},v}$ is (ε_1)-regular. By the definition of W_v we have $e_G(W_v, \mathcal{C}) > (1+2\varepsilon_1)|W_v||U_j|a_j b_j^2$ and by (G3) we have $|\mathcal{C}| = (1 \pm \eta)a_j b_j |U_j|$ and $e_G(V_j, \mathcal{C}) \le (1+\eta)|V_j||U_j|a_j b_j^2$, so we have $d_G(W_v, \mathcal{C}) > \frac{(1+2\varepsilon_1)b_j}{1+\eta} \ge (1+\varepsilon_1)d_G(V_j, \mathcal{C})$. Hence, by the (ε_1)-regularity of G we obtain $|W_v| < \varepsilon_1 |V_j|$. Now take a superset $W \supseteq W_v$ of size $|W| = \varepsilon_1|V_j|$. By the (ε_1)-regularity of G we have $d_G(W, \mathcal{C}) \le (1+\varepsilon_1)d_G(V_j, \mathcal{C})$, so we have

$$\sum_{w \in W_v} s(v,w) \le e_G(W, \mathcal{C}) \le (1+\varepsilon_1)\varepsilon_1 e_G(V_j, \mathcal{C}) \le 2\varepsilon_1 |U_j||V_j|a_j b_j^2$$

as desired. ∎

(RGA6) follows immediately from the good event of Claim 4.75.

Claim 4.75. *Asymptotically almost surely the following holds. For every $j \in J$, every $v \in V_j$ and every set $W \subseteq V_j$, we have*

$$\sum_{w \in W} |\{x \in X_j^{\mathrm{buf}} : v, w \in \mathcal{C}_T(x)\}| \le 2^{\Delta+3} b_j^2 \mu |X_j|(|W| + 2\varepsilon_1|V_j|).$$

Proof. Let $j \in J$. Enumerate X_j^{buf} as $x_1, \dots, x_a$. For each $i \in [a]$ consider the embedding of the neighbours $y_{i1}, \dots, y_{ib}$ of x_i. For $p \in [b]_0$ set $H_{ip} = H^{-1}(x_i)[\{y_{i1}, \dots, y_{ip}\}]$. By Claim 4.66 each y_{ip} is embedded uniformly at random into a subset of $\mathcal{C}_{\tau(y_{ip})-1}(y_{ip})$ of size at least $\frac{\mathcal{D}(H_{ih})}{2\mathcal{D}(H_{i(h-1)})}|V_{y_{ip}}|$. Let $v, w \in V_j$. By the discussion above, conditioning on the history up to the time right before the embedding of y_{i1}, we embed $N_{H^{(2)}}(x_i)$ to each element of $S(v,w)$ with probability at most $\frac{2^b}{a_j|U_j|}$. Note that

$$v, w \in \mathcal{C}_{\tau(y_{ib})}(x_i) = \mathcal{C}_T(x_i) \iff \phi_{\tau(y_{ib})}(H^{-1}(x_i)) \in S(v,w).$$

We first consider when $w \in V_j \setminus W_v$. In this case, the probability that $v, w \in \mathcal{C}_T(x_i)$, conditioning on the history up to the time right before the embedding of y_{i1}, is at most

$$\frac{s(v,w)2^b}{a_j|U_j|} \le (1+2\varepsilon_1)2^b b_j^2 \le (1+2\varepsilon_1)2^\Delta b_j^2.$$

Then, since the vertices of $N_{H^{(2)}}(x_i)$ are embedded consecutively, we apply Lemma 4.12 to find that the probability that

$$|\{x \in X_j^{\mathrm{buf}} : v, w \in \mathcal{C}_T(x)\}| > 2^{\Delta+3} b_j^2 \mu |X_j|$$

is at most $\exp\left(-2^{\Delta-3} b_j^2 \mu |X_j|\right)$.

Now consider W_v collectively. For each $i \in [a]$ set Y_i to be $|W_v \cap \mathcal{C}_T(x_i)|$ if $v \in \mathcal{C}_T(x_i)$ and zero otherwise. Applying Claim 4.74, we find that the expectation of Y_i, conditioning on the history up to the time right before the embedding of $\min(N_{H^{(2)}}(x_i))$, is at most

$$\sum_{w \in W_v} \frac{s(v,w)2^b}{a_j |U_j|} \leq 2^{b+1} \varepsilon_1 |V_j| b_j^2.$$

Then, since the vertices of $N_{H^{(2)}}(x_i)$ are embedded consecutively, we apply Lemma 4.12 to find that the probability that

$$\sum_{w \in W_v} |\{x \in X_j^{\mathrm{buf}} : v, w \in \mathcal{C}_T(x)\}| > 2^{\Delta+4} \varepsilon_1 |V_j| b_j^2 \mu |X_j|$$

is at most $\exp\left(-2^{\Delta-2} b_j^2 \mu |X_j|\right)$.

Let $\overline{\mathcal{E}}$ be the event that given $j \in J$ and $v \in V_j$ we have

$$|\{x \in X_j^{\mathrm{buf}} : v, w \in \mathcal{C}_T(x)\}| \leq 2^{\Delta+3} b_j^2 \mu |X_j|$$

for all $w \in V_j \setminus W_v$ and we have

$$\sum_{w \in W_v} |\{x \in X_j^{\mathrm{buf}} : v, w \in \mathcal{C}_T(x)\}| \leq 2^{\Delta+4} \varepsilon_1 |V_j| b_j^2 \mu |X_j|.$$

Putting the cases together, applying Claim 4.68 and taking a union bound over all choices of $j \in J$ and $v, w \in V_j$, we deduce that the probability of both the complement of $\overline{\mathcal{E}}$ and the good event thus far holding simulataneously is at most

$$n^2 \exp\left(-2^{\Delta-3} \tfrac{\mu n}{\kappa |J|} \left(\tfrac{10\kappa |J| \Delta \log n}{\rho n}\right)^{1/2^{\Delta^2}}\right),$$

which tends to one as n goes to infinity. To complete the proof, it remains to show that $\overline{\mathcal{E}}$ implies the desired outcome. Suppose $\overline{\mathcal{E}}$ holds. Let $j \in J$, $v \in V_j$ and $W \subseteq V_j$. Then we have

$$\begin{aligned}
&\sum_{w \in W} |\{x \in X_j^{\mathrm{buf}} : v, w \in \mathcal{C}_T(x)\}| \\
&\leq \sum_{w \in W \setminus W_v} |\{x \in X_j^{\mathrm{buf}} : v, w \in \mathcal{C}_T(x)\}| + \sum_{w \in W_v} |\{x \in X_j^{\mathrm{buf}} : v, w \in \mathcal{C}_T(x)\}| \\
&\leq 2^{\Delta+3} b_j^2 \mu |X_j| (|W| + 2\varepsilon_1 |V_j|)
\end{aligned}$$

as desired. ■

This completes the proof. □

Bibliography

[1] M. Ajtai, J. Komlós, and E. Szemerédi, *The longest path in a random graph*, Combinatorica **1** (1981), 1–12.

[2] P. Allen, J. Böttcher, O. Cooley, and R. Mycroft, *Tight cycles and regular slices in dense hypergraphs*, J. Combin. Theory Ser. A **149** (2017), 30–100.

[3] P. Allen, J. Böttcher, E. Davies, E. K. Hng, and J. Skokan, *A sparse hypergraph blow-up lemma*, In preparation.

[4] P. Allen, J. Böttcher, H. Hàn, Y. Kohayakawa, and Y. Person, *Blow-up lemmas for sparse graphs*, 2016, arXiv:1612.00622.

[5] P. Allen, J. Böttcher, and J. Hladký, *Filling the gap between Turán's theorem and Pósa's conjecture*, J. Lond. Math. Soc. (2) **84** (2011), no. 2, 269–302.

[6] P. Allen, J. Böttcher, J. Hladký, and D. Piguet, *Packing degenerate graphs*, Adv. Math. **354** (2019), 106739, 58.

[7] P. Allen, J. Böttcher, Y. Kohayakawa, H. Naves, and Y. Person, *Making spanning graphs*, 2017, arXiv:1711.05311.

[8] P. Allen, E. Davies, and J. Skokan, *Regularity inheritance in hypergraphs*, 2019, arXiv:1901.05955.

[9] N. Alon and Z. Füredi, *Spanning subgraphs of random graphs*, Graphs Combin. **8** (1992), no. 1, 91–94.

[10] M. Anastos and A. Frieze, *A scaling limit for the length of the longest cycle in a sparse random graph*, J. Combin. Theory Ser. B **148** (2021), 184–208.

[11] B. Andrásfai, P. Erdős, and V. T. Sós, *On the connection between chromatic number, maximal clique and minimal degree of a graph*, Discrete Math. **8** (1974), 205–218.

[12] M. Bednarska and T. Łuczak, *Biased positional games for which random strategies are nearly optimal*, Combinatorica **20** (2000), no. 4, 477–488.

[13] D. Conlon, J. Fox, and Y. Zhao, *A relative Szemerédi theorem*, Geom. Funct. Anal. **25** (2015), no. 3, 733–762.

[14] O. Cooley, *Paths, cycles and sprinkling in random hypergraphs*, 2021, arXiv:2103.16527.

[15] O. Cooley, N. Fountoulakis, D. Kühn, and D. Osthus, *Embeddings and Ramsey numbers of sparse k-uniform hypergraphs*, Combinatorica **29** (2009), no. 3, 263–297.

[16] O. Cooley, F. Garbe, E. K. Hng, M. Kang, N. Sanhueza-Matamala, and J. Zalla, *Longest paths in random hypergraphs*, SIAM J. Discrete Math. **35** (2021), no. 4, 2430–2458.

[17] O. Cooley, M. Kang, and C. Koch, *The size of the giant high-order component in random hypergraphs*, Random Structures Algorithms **53** (2018), no. 2, 238–288.

[18] O. Cooley, M. Kang, and Y. Person, *Largest components in random hypergraphs*, Combin. Probab. Comput. **27** (2018), no. 5, 741–762.

[19] O. Cooley, M. Kang, and J. Zalla, *Loose cores and cycles in random hypergraphs*, 2021, arXiv:2101.05008.

[20] G. A. Dirac, *Some theorems on abstract graphs*, Proc. London Math. Soc. (3) **2** (1952), 69–81.

[21] A. Dudek and A. Frieze, *Loose Hamilton cycles in random uniform hypergraphs*, Electron. J. Combin. **18** (2011), no. 1, Paper 48, 14.

[22] ———, *Tight Hamilton cycles in random uniform hypergraphs*, Random Structures Algorithms **42** (2013), no. 3, 374–385.

[23] P. Erdős and A. Rényi, *On the evolution of random graphs*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **5** (1960), 17–61.

[24] P. Erdös and A. H. Stone, *On the structure of linear graphs*, Bull. Amer. Math. Soc. **52** (1946), 1087–1091.

[25] G. Fan and H. A. Kierstead, *The square of paths and cycles*, J. Combin. Theory Ser. B **63** (1995), no. 1, 55–64.

[26] ______, *Hamiltonian square-paths*, J. Combin. Theory Ser. B **67** (1996), no. 2, 167–182.

[27] P. Frankl and V. Rödl, *Extremal problems on set systems*, Random Structures Algorithms **20** (2002), no. 2, 131–164.

[28] W. T. Gowers, *Hypergraph regularity and the multidimensional Szemerédi theorem*, Ann. of Math. (2) **166** (2007), no. 3, 897–946.

[29] E. Győri, G. Y. Katona, and N. Lemons, *Hypergraph extensions of the Erdős-Gallai theorem*, European J. Combin. **58** (2016), 238–246.

[30] A. Hajnal and E. Szemerédi, *Proof of a conjecture of P. Erdős*, Combinatorial theory and its applications, II (Proc. Colloq., Balatonfüred, 1969), North-Holland, Amsterdam, 1970, pp. 601–623.

[31] E. K. Hng, *Minimum degrees for powers of paths and cycles*, 2020, arXiv:2005.02210.

[32] S. Janson, T. Łuczak, and A. Rucinski, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.

[33] P. Keevash, *A hypergraph blow-up lemma*, Random Structures Algorithms **39** (2011), no. 3, 275–376.

[34] G. Kemkes and N. Wormald, *An improved upper bound on the length of the longest cycle of a supercritical random graph*, SIAM J. Discrete Math. **27** (2013), no. 1, 342–362.

[35] J. H. Kim and V. H. Vu, *Concentration of multivariate polynomials and its applications*, Combinatorica **20** (2000), no. 3, 417–434.

[36] Y. Kohayakawa, *Szemerédi's regularity lemma for sparse graphs*, Foundations of computational mathematics (Rio de Janeiro, 1997), Springer, Berlin, 1997, pp. 216–230.

[37] Y. Kohayakawa, V. Rödl, M. Schacht, and E. Szemerédi, *Sparse partition universal graphs for graphs of bounded degree*, Adv. Math. **226** (2011), no. 6, 5041–5065.

[38] J. Komlós, G. N. Sárközy, and E. Szemerédi, *Blow-up lemma*, Combinatorica **17** (1997), no. 1, 109–123.

[39] ______, *Proof of the Seymour conjecture for large graphs*, Ann. Comb. **2** (1998), no. 1, 43–60.

[40] M. Krivelevich, *The critical bias for the Hamiltonicity game is* $(1 + o(1))n/\ln n$, J. Amer. Math. Soc. **24** (2011), no. 1, 125–131.

[41] M. Krivelevich and B. Sudakov, *The phase transition in random graphs: A simple proof*, Random Structures Algorithms **43** (2013), no. 2, 131–138.

[42] D. Kühn and D. Osthus, *Hamilton cycles in graphs and hypergraphs: an extremal perspective*, Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. IV, Kyung Moon Sa, Seoul, 2014, pp. 381–406.

[43] D. Kühn, D. Osthus, and A. Taraz, *Large planar subgraphs in dense graphs*, J. Combin. Theory Ser. B **95** (2005), no. 2, 263–282.

[44] A. Liebenau and R. Nenadov, *The threshold bias of the clique-factor game*, J. Combin. Theory Ser. B **152** (2022), 221–247.

[45] T. Łuczak, *Cycles in a random graph near the critical point*, Random Structures Algorithms **2** (1991), no. 4, 421–439.

[46] D. Mubayi and A. Suk, *A survey of hypergraph Ramsey problems*, Discrete Mathematics and Applications, Springer Optim. Appl., vol. 165, Springer, 2020, pp. 405–428.

[47] B. Narayanan and M. Schacht, *Sharp thresholds for nonlinear Hamiltonian cycles in hypergraphs*, Random Structures Algorithms **57** (2020), no. 1, 244–255.

[48] B. Pittel, *A random graph with a subcritical number of edges*, Trans. Amer. Math. Soc. **309** (1988), no. 1, 51–75.

[49] V. Rödl, A. Ruciński, and E. Szemerédi, *An approximate Dirac-type theorem for k-uniform hypergraphs*, Combinatorica **28** (2008), no. 2, 229–260.

[50] V. Rödl and M. Schacht, *Regular partitions of hypergraphs: regularity lemmas*, Combin. Probab. Comput. **16** (2007), no. 6, 833–885.

[51] V. Rödl and J. Skokan, *Regularity lemma for k-uniform hypergraphs*, Random Structures Algorithms **25** (2004), no. 1, 1–42.

[52] V. Rödl and E. Szemerédi, *On size Ramsey numbers of graphs with bounded degree*, Combinatorica **20** (2000), no. 2, 257–262.

[53] J. Schmidt-Pruzan and E. Shamir, *Component structure in the evolution of random hypergraphs*, Combinatorica **5** (1985), no. 1, 81–94.

[54] P. Seymour, *Problem section*, Combinatorics: Proceedings of the British Combinatorial Conference 1973 (T. P. McDonough and V.C. Mavron, eds.), Cambridge University Press, 1974, pp. 201–202.

[55] K. Staden and A. Treglown, *On degree sequences forcing the square of a Hamilton cycle*, SIAM J. Discrete Math. **31** (2017), no. 1, 383–437.

[56] E. Szemerédi, *Regular partitions of graphs*, Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976), Colloq. Internat. CNRS, vol. 260, CNRS, Paris, 1978, pp. 399–401.

[57] P. Turán, *On an extremal problem in graph theory*, Mat. Fiz. Lapok **48** (1941), 436–452.

[58] J. E. Williamson, *Panconnected graphs. II*, Period. Math. Hungar. **8** (1977), no. 2, 105–116.

[59] Y. Zhao, *Recent advances on Dirac-type problems for hypergraphs*, Recent trends in combinatorics, IMA Vol. Math. Appl., vol. 159, Springer, 2016, pp. 145–165.