

# EXTREMAL PROBLEMS ON GRAPHS, DIRECTED GRAPHS AND HYPERGRAPHS

by

TIMOTHY DUNCAN TOWNSEND

A thesis submitted to  
The University of Birmingham  
for the degree of  
DOCTOR OF PHILOSOPHY (PURE MATHEMATICS)

School of Mathematics  
The University of Birmingham  
August 2015

UNIVERSITY OF  
BIRMINGHAM

**University of Birmingham Research Archive**

**e-theses repository**

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

# ABSTRACT

This thesis is concerned with extremal problems on graphs and similar structures.

We first study degree conditions in uniform hypergraphs that force matchings of various sizes. Our main result in this area improves bounds of Markström and Ruciński on the minimum  $d$ -degree which forces a perfect matching in a  $k$ -uniform hypergraph on  $n$  vertices. We also extend bounds of Bollobás, Daykin and Erdős by asymptotically determining the minimum vertex degree which forces a matching of size  $t < n/2(k-1)$  in a  $k$ -uniform hypergraph on  $n$  vertices. Further asymptotically tight results on  $d$ -degrees which force large matchings are also obtained. Our approach is to prove fractional versions of the above results and then translate these into integer versions.

We then study connectivity conditions in tournaments that ensure the existence of partitions of the vertex set that satisfy various properties. In 1982 Thomassen asked whether there exists an integer  $f(k, t)$  such that every strongly  $f(k, t)$ -connected tournament  $T$  admits a partition of its vertex set into  $t$  vertex classes  $V_1, \dots, V_t$  such that for all  $i$  the subtournament  $T[V_i]$  induced on  $T$  by  $V_i$  is strongly  $k$ -connected. Our main result in this area implies an affirmative answer to this question. In particular we show that  $f(k, t) = O(k^7 t^4)$  suffices. As another application of our main result we give an affirmative answer to a question of Song as to whether, for any integer  $t$ , there exists an integer

$h(t)$  such that every strongly  $h(t)$ -connected tournament has a 1-factor consisting of  $t$  vertex-disjoint cycles of prescribed lengths. We show that  $h(t) = O(t^5)$  suffices.

Finally we investigate the typical structure of graphs and directed graphs with some forbidden subgraphs. Motivated by his work on the classification of countable homogeneous oriented graphs, Cherlin asked about the typical structure of triangle-free oriented graphs. We give an answer to this question (which is not quite the predicted one). Our approach is based on the ‘hypergraph containers’ method, developed independently by Saxton and Thomason as well as by Balogh, Morris and Samotij. Moreover, our results generalise to forbidden transitive tournaments and forbidden oriented cycles of any order, and also apply to digraphs. Along the way we prove several stability results for weighted extremal digraph problems, which we believe are of independent interest.

We also determine, for all  $k \geq 6$ , the typical structure of graphs that do not contain an induced  $2k$ -cycle. This verifies a conjecture of Balogh and Butterfield. Surprisingly, the typical structure of such graphs is richer than that encountered in related results. The approach we take also yields an approximate result on the typical structure of graphs without an induced 8-cycle or without an induced 10-cycle.

# ACKNOWLEDGEMENTS

I wish to thank my supervisors, Daniela Kühn and Deryk Osthus, for their constant support and guidance.

# CONTENTS

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Extremal graph theory . . . . .	1
1.2	Matchings in graphs and hypergraphs . . . . .	2
1.3	Partitions in graphs and tournaments . . . . .	4
1.4	The typical structure of graphs and digraphs with a given forbidden subgraph	6
<b>2</b>	<b>Fractional and integer matchings in uniform hypergraphs</b>	<b>8</b>
2.1	Chapter introduction . . . . .	8
2.1.1	Large matchings in hypergraphs with large degrees . . . . .	8
2.1.2	Large matchings in hypergraphs with many edges . . . . .	11
2.1.3	Large fractional matchings . . . . .	12
2.2	Notation, tools and preliminary results . . . . .	14
2.2.1	Notation . . . . .	14
2.2.2	Tools and preliminary results . . . . .	14

2.3	Minimum edge-density conditions for fractional matchings . . . . .	17
2.4	Minimum degree conditions for fractional matchings . . . . .	23
2.5	Constructing integer matchings from fractional ones . . . . .	26
2.6	Variants of Theorems 2.1.4 and 2.1.6 . . . . .	37
<b>3</b>	<b>Proof of a tournament partition conjecture and an application to 1-factors with prescribed cycle lengths</b>	<b>46</b>
3.1	Chapter introduction . . . . .	46
3.1.1	Partitioning tournaments into highly connected subtournaments . .	46
3.1.2	Partitioning tournaments into vertex-disjoint cycles . . . . .	49
3.2	Notation, tools and preliminary results . . . . .	51
3.3	Proof of Theorem 3.1.5 . . . . .	56
3.4	Partitioning tournaments into vertex-disjoint cycles . . . . .	69
<b>4</b>	<b>On the structure of oriented graphs and digraphs with forbidden tournaments or cycles</b>	<b>72</b>
4.1	Chapter introduction . . . . .	72
4.1.1	Oriented graphs and digraphs with forbidden subdigraphs . . . . .	72
4.1.2	Oriented graphs and digraphs with forbidden tournaments or cycles	73
4.1.3	Overview of the chapter . . . . .	76
4.2	Notation and tools . . . . .	78
4.3	Digraph containers . . . . .	80

4.4	Rough structure of typical $T_{k+1}$ -free oriented graphs and digraphs . . . . .	84
4.5	Exact structure of typical $T_{k+1}$ -free oriented graphs and digraphs . . . . .	90
4.6	Rough structure of typical $C_k$ -free oriented graphs and digraphs . . . . .	104
4.7	Typical $C_k$ -free oriented graphs and digraphs are not acyclic . . . . .	126
<b>5</b>	<b>Forbidding induced even cycles in a graph: typical structure and counting</b>	<b>131</b>
5.1	Chapter introduction . . . . .	131
5.1.1	Graphs with forbidden induced graphs . . . . .	131
5.1.2	Graphs with forbidden induced cycles . . . . .	132
5.1.3	Overview of the chapter . . . . .	134
5.2	Notation and tools . . . . .	136
5.3	Approximate structure of typical induced- $C_{2k}$ -free graphs . . . . .	138
5.3.1	Tools: containers, stability and removal lemmas . . . . .	139
5.3.2	Stability and removal lemmas for even cycles . . . . .	141
5.3.3	Approximate structure of typical induced $C_{2k}$ -free graphs . . . . .	149
5.4	The number of $k$ -templates . . . . .	150
5.5	Properties of near- $k$ -templates . . . . .	159
5.6	Derivation of Theorem 5.1.1 from the main lemma . . . . .	165
5.7	Estimation of $ F_Q^1 $ . . . . .	169
5.8	Estimation of $ F_Q^2 $ . . . . .	177



5.9	Estimation of $ F_Q^3 $ . . . . .	188
5.10	Proof of Lemma 5.6.1 . . . . .	202
5.11	The Removal Lemma for complete 2-coloured multigraphs . . . . .	204
<b>List of references</b>		<b>217</b>

# CHAPTER 1

## INTRODUCTION

### 1.1 Extremal graph theory

Extremal graph theory is the study of graphs that are extremal with a certain property, where ‘extremal’ means maximal or minimal with respect to some graph parameter. It can be seen as the study of how local properties of a graph affect global structure in that graph, and vice versa. Indeed, in extremal graph theory many problems are concerned with finding sufficient conditions on a graph  $G$  that force  $G$  to have some particular structure, while many other problems are concerned with determining the structure of graphs that satisfy some hereditary property, such as not containing a given subgraph. Both of these ideas are encompassed in Turán’s theorem, which gives the maximum number of edges that a graph on  $n$  vertices can have if it does not contain a  $k$ -clique – that is, a subgraph on  $k$  vertices with all possible edges present. Turán’s theorem also gives an exact description of the unique edge-maximal graph not containing a  $k$ -clique, and forms the cornerstone of extremal graph theory.

Here we study three such areas in extremal graph theory. We first discuss conditions that force matchings of various sizes in graphs and hypergraphs. We then consider conditions on graphs and tournaments that force the existence of partitions of the vertex set that satisfy some property. Finally we discuss the typical structure of graphs, oriented graphs and directed graphs that do not contain a given forbidden (induced) subgraph.

## 1.2 Matchings in graphs and hypergraphs

A *matching* in a graph  $G$  is a set of edges of  $G$ , no two of which share a vertex. A *perfect matching* in  $G$  is a matching such that every vertex of  $G$  is contained in some edge in the matching. A graph is *bipartite* if its vertex set can be partitioned into two sets in such a way that no two vertices in the same edge are in the same partition class. If  $A$  is a set of vertices in a graph  $G$ , we let  $N(A)$  denote the set of vertices of  $G$  that share an edge with a vertex in  $A$ .

The problem of determining which graphs contain a perfect matching has long been well understood. In particular, a simple characterisation of all bipartite graphs that contain a perfect matching was proved by Hall as early as 1935.

**Theorem 1.2.1 (Hall’s theorem).** [38] *Let  $G$  be a bipartite graph with vertex partition classes  $X, Y$  such that  $|X| = |Y|$ .  $G$  contains a perfect matching if and only if  $|N(A)| \geq |A|$  for every  $A \subseteq X$ .*

In 1947 Tutte characterised all graphs that contain a perfect matching.

**Theorem 1.2.2 (Tutte’s theorem).** [81] *A graph  $G = (V, E)$  has a perfect matching if and only if, for every  $U \subseteq V$ , the graph  $G - U$  has at most  $|U|$  components with an odd number of vertices.*

The following corollary of a result of Dirac on Hamilton cycles from 1952, which provides an easy to check sufficient condition that guarantees a perfect matching in a graph, is also very useful. The *degree* of a vertex  $v$  in a graph is defined to be  $d(v) := |N(\{v\})|$ , and the *minimum degree* of a graph  $G = (V, E)$  is defined to be  $\delta(G) := \min_{v \in V} (d(v))$ .

**Theorem 1.2.3.** [26] *A graph  $G$  with an even number  $n \geq 4$  of vertices contains a perfect matching if  $\delta(G) \geq n/2$ .*

Note that this minimum degree condition is best possible, in the sense that there exists a graph  $G$  on an even number of vertices without a perfect matching, such that  $\delta(G) = n/2 - 1$ . This can be seen by considering the graph on  $n$  vertices consisting of two disjoint  $n/2$ -cliques, where  $n$  is equal to 2 mod 4. Note also that Theorem 1.2.3 can also be easily proved directly, or derived from Tutte's theorem.

A *k-uniform hypergraph* is a pair  $(V, E)$  where  $V$  is a set of vertices and  $E$  is a set of subsets of  $V$ , each of size exactly  $k$ , which we call edges. Note that 2-uniform hypergraphs are just graphs. Similarly to the graph case, a *matching* in a hypergraph is a set of edges of a hypergraph, no two of which share a vertex, and a *perfect matching* in a hypergraph is a matching such that every vertex is contained in some edge in the matching. While it can be proved that the problem of determining whether a graph contains a perfect matching is decidable in polynomial time, it has been shown that, for  $k \geq 3$ , the problem of determining whether a  $k$ -uniform hypergraph contains a perfect matching is NP-complete. As such, rather than attempt to completely characterise all  $k$ -uniform hypergraphs that contain a perfect matching, it makes sense to try to determine reasonable sufficient conditions on a  $k$ -uniform hypergraph that force it to contain a perfect matching. This is the main focus of Chapter 2. Our main result in Chapter 2 can be seen as a hypergraph analogue of the Dirac type Theorem 1.2.3 above, in the sense that we give a minimum degree type condition that forces the existence of a perfect matching in a

$k$ -uniform hypergraph. Though the minimum degree bounds that we give are the best known, they are unlikely to be best possible. We do however also determine some best possible minimum degree type conditions that force the existence of matchings of various other sizes in  $k$ -uniform hypergraphs (in particular for matchings that are at most half the size of a perfect matching).

All results in Chapter 2 are joint work with Kühn and Osthus, and a slightly abridged version of Chapter 2 has been published in the European Journal of Combinatorics (see [51]).

### 1.3 Partitions in graphs and tournaments

Much work has been done on problems relating to partitions of graphs into subgraphs that inherit some properties of the original graph. For instance, the following result was proved by Hajnal, and independently Thomassen, in 1983. For a graph  $G = (V, E)$  and a subset  $A \subset V$  we let  $G[A]$  denote the subgraph of  $G$  with vertex set  $A$  and edge set consisting of all edges in  $E$  that are contained in  $A$ .

**Theorem 1.3.1.** *[37, 77] For every  $\ell$  there exists  $k = k(\ell)$  such that the vertex set of every graph with minimum degree at least  $k$  can be partitioned into sets  $A$  and  $B$  in such a way that  $G[A]$  and  $G[B]$  both have minimum degree at least  $\ell$ .*

Later, it was shown by Stiebitz [75] that  $k = 2\ell + 1$  is sufficient, which can be seen to be best possible by considering the complete graph on  $2\ell + 1$  vertices. We say  $G$  is  $k$ -connected if  $|V| > k$  and for any set  $S \subseteq V$  with  $|S| < k$ ,  $G - S$  is connected. Hajnal, and independently Thomassen, also proved a similar result to Theorem 1.3.1, but with the notion of minimum degree replaced by that of connectivity.

**Theorem 1.3.2.** *[37, 77] For every  $\ell$  there exists  $k = k(\ell)$  such that the vertex set of*

every  $k$ -connected graph can be partitioned into sets  $A$  and  $B$  in such a way that  $G[A]$  and  $G[B]$  are both  $\ell$ -connected.

Another result similar to Theorem 1.3.1 was conjectured by El-Zahar, where this time the partition guaranteed by the minimum degree condition is a partition into vertex disjoint cycles of prescribed lengths.

**Conjecture 1.3.3.** [27] *Let  $G = (V, E)$  be a graph. Suppose  $|V| = n_1 + \cdots + n_k$  and  $\delta(G) \geq \lceil n_1/2 \rceil + \cdots + \lceil n_k/2 \rceil$  where  $n_i \geq 3$  for every  $i \in \{1, \dots, k\}$ . Then  $G$  contains  $k$  disjoint cycles of lengths  $n_1, \dots, n_k$ , respectively.*

Conjecture 1.3.3 has been proved for all sufficiently large values of  $|V|$  by Abbasi [1].

A *tournament* is a complete graph (that is, a graph with all possible edges present) with an orientation assigned to each edge. In Chapter 3 we investigate similar results to those mentioned here, but for tournaments rather than graphs. In particular we prove a corresponding result to Theorem 1.3.2 for tournaments, for some analogous notion of connectivity. This settles a problem set by Thomassen [77] in 1983. We are also able to use the methods employed to prove a corresponding result to Conjecture 1.3.3 for tournaments, where the minimum degree condition is replaced by a connectivity condition. This settles a question of Song [73].

All results in Chapter 3 are joint work with Kühn and Osthus, and have been published in *Combinatorica* (see [52]).

## 1.4 The typical structure of graphs and digraphs with a given forbidden subgraph

The enumeration and description of the typical structure of graphs with given side constraints has become a successful and popular area at the interface of probabilistic, enumerative, and extremal combinatorics (see e.g. [14] for a survey of such work). For example, the following classical result of Erdős, Kleitman and Rothschild from 1976 asymptotically determines for every  $k \geq 3$  the logarithm of the number of graphs on  $n$  vertices that do not contain a  $k$ -clique.

**Theorem 1.4.1.** [32] *For every  $k \geq 3$ , the number of graphs on  $n$  vertices that do not contain a  $k$ -clique is  $2^{\frac{n^2}{2}(1-\frac{1}{k-1})+o(n^2)}$ .*

This result was strengthened by Kolaitis, Prömel and Rothschild in 1987, who showed that ‘almost all’ graphs that do not contain a  $k$ -clique are  $(k-1)$ -partite, for every  $k \geq 3$  (the case  $k = 3$  of this was already proved in [32]). More formally, given a class of graphs  $\mathcal{A}$ , we let  $\mathcal{A}_n$  denote the set of all graphs in  $\mathcal{A}$  that have precisely  $n$  vertices, and we say that *almost all graphs in  $\mathcal{A}$  have property  $\mathcal{B}$*  if

$$\lim_{n \rightarrow \infty} \frac{|\{G \in \mathcal{A}_n : G \text{ has property } \mathcal{B}\}|}{|\mathcal{A}_n|} = 1.$$

Then the result of Kolaitis, Prömel and Rothschild is as follows.

**Theorem 1.4.2.** [46] *For every  $k \geq 3$ , almost all graphs that do not contain a  $k$ -clique are  $(k-1)$ -partite.*

There are now many precise results on the number and typical structure of  $H$ -free graphs (that is, graphs that do not contain a fixed graph  $H$  as a not necessarily induced subgraph).

However, the corresponding questions for digraphs and oriented graphs are almost all wide open. (Briefly, an oriented graph is a graph with an orientation assigned to each edge; a digraph is similar but also allows two edges to span a pair of vertices - one oriented in each direction.) In Chapter 4 we investigate such questions. In particular we determine the typical structure of oriented graphs that do not contain a transitive tournament of size  $k$ , and of oriented graphs that do not contain an oriented cycle of size  $k$ , as well as proving corresponding results for digraphs. This answers a question of Cherlin [21]. The corresponding asymptotic counting results follow immediately from these structural results.

Given a fixed graph  $H$ , a graph is called *induced- $H$ -free* if it does not contain  $H$  as an induced subgraph. Associated counting and structural questions are equally natural as in the case of  $H$ -free graphs, but seem harder to solve. Thus much less is known about the typical structure and number of induced- $H$ -free graphs than that of  $H$ -free graphs. In Chapter 5 we determine the typical structure of induced- $C_{2k}$ -free graphs (from which the corresponding asymptotic counting result again follows immediately). This verifies a conjecture of Balogh and Butterfield [10].

All results in Chapter 4 are joint work with Kühn, Osthus and Zhao, while all results in Chapter 5 are joint work with Kühn, Osthus and Kim.



# CHAPTER 2

## FRACTIONAL AND INTEGER MATCHINGS IN UNIFORM HYPERGRAPHS

### 2.1 Chapter introduction

#### 2.1.1 Large matchings in hypergraphs with large degrees

A *k-uniform hypergraph* is a pair  $G = (V, E)$  where  $V$  is a finite set of vertices and the edge set  $E$  consists of unordered  $k$ -tuples of elements of  $V$ . A *matching* (or *integer matching*)  $M$  in  $G$  is a set of disjoint edges of  $G$ . The *size* of  $M$  is the number of edges in  $M$ . We say  $M$  is *perfect* if it has size  $|V|/k$ . Given  $S \in \binom{V}{d}$ , where  $0 \leq d \leq k-1$ , let  $\deg_G(S) = |\{e \in E : S \subseteq e\}|$  be the *degree* of  $S$  in  $G$ . Let  $\delta_d(G) = \min_{S \in \binom{V}{d}} \{\deg_G(S)\}$  be the *minimum d-degree* of  $G$ . When  $d = 1$ , we refer to  $\delta_1(G)$  as the *minimum vertex degree* of  $G$ . Note that  $\delta_0(G) = |E|$ .

For integers  $n, k, d, s$  satisfying  $0 \leq d \leq k-1$  and  $0 \leq s \leq n/k$ , we let  $m_d^s(k, n)$  denote the

minimum integer  $m$  such that every  $k$ -uniform hypergraph  $G$  on  $n$  vertices with  $\delta_d(G) \geq m$  has a matching of size  $s$ . We write  $o(1)$  to denote some function that tends to 0 as  $n$  tends to infinity. The following degree condition for forcing perfect matchings has been conjectured in [39, 50] and has received much attention recently.

**Conjecture 2.1.1.** *Let  $n$  and  $1 \leq d \leq k - 1$  be such that  $n, d, k, n/k \in \mathbb{N}$ . Then*

$$m_d^{n/k}(k, n) = \left( \max \left\{ \frac{1}{2}, 1 - \left( \frac{k-1}{k} \right)^{k-d} \right\} + o(1) \right) \binom{n-d}{k-d}.$$

The first term in the lower bound here is given by the following parity-based construction from [49]. For any integers  $n, k$ , let  $H'$  be a  $k$ -uniform hypergraph on  $n$  vertices with vertex partition  $A \cup B = V(H')$ , such that  $||A| - |B|| \leq 2$  and  $|A|$  and  $n/k$  have different parity. Let  $H'$  have edge set consisting of all  $k$ -element subsets of  $V(H')$  that intersect  $A$  in an odd number of vertices. Observe that  $H'$  has no perfect matching, and that for every  $1 \leq d \leq k - 1$  we have that  $\delta_d(H') = (1/2 + o(1)) \binom{n-d}{k-d}$ . The second term in the lower bound is given by the hypergraph  $H(n/k)$  defined as follows. Let  $H(s)$  be the  $k$ -uniform hypergraph on  $n$  vertices with edge set consisting of all  $k$ -element subsets of  $V(H(s))$  intersecting a given (fixed) subset of  $V(H(s))$  of size  $s - 1$ , that is  $H(s) = K_n^{(k)} - K_{n-s+1}^{(k)}$ .

For  $d = k - 1$ ,  $m_{k-1}^{n/k}(k, n)$  was determined exactly for large  $n$  by Rödl, Ruciński and Szemerédi [69]. This was generalized by Treglown and Zhao [78, 79], who determined the extremal families for all  $d \geq k/2$ . The extremal constructions are similar to the parity based one of  $H'$  above. This improves asymptotic bounds in [58, 68, 69]. Recently, Keevash, Knox and Mycroft [41] investigated the structure of hypergraphs whose minimum  $(k - 1)$ -degree lies below the threshold and which have no perfect matching.

For  $d < k/2$  less is known. In [5] Conjecture 2.1.1 was proved for  $k - 4 \leq d \leq k - 1$ , by reducing it to a probabilistic conjecture of Samuels. In particular, this implies Con-

jecture 2.1.1 for  $k \leq 5$ . Khan [43], and independently Kühn, Osthus and Treglown [53], determined  $m_1^{n/k}(k, n)$  exactly for  $k = 3$ . Khan [44] also determined  $m_1^{n/k}(k, n)$  exactly for  $k = 4$ . It was shown by Hàn, Person and Schacht [39] that for  $k \geq 3$ ,  $1 \leq d < k/2$  we have  $m_d^{n/k}(k, n) \leq ((k-d)/k + o(1)) \binom{n-d}{k-d}$ . (The case  $d = 1$  of this is already due to Daykin and Häggkvist [24].) These bounds were improved by Markström and Ruciński [55], using similar techniques, to

$$m_d^{n/k}(k, n) \leq \left( \frac{k-d}{k} - \frac{1}{k^{k-d}} + o(1) \right) \binom{n-d}{k-d}.$$

The main result in this chapter improves on this bound, using quite different techniques.

**Theorem 2.1.2.** *Let  $d, k \in \mathbb{N}$  with  $1 \leq d < k/2$  be fixed. Then for all  $n$  such that  $n, n/k \in \mathbb{N}$ ,*

$$m_d^{n/k}(k, n) \leq \left( \frac{k-d}{k} - \frac{k-d-1}{k^{k-d}} + o(1) \right) \binom{n-d}{k-d}.$$

We also consider degree conditions that force smaller matchings. As a consequence of the results of Kühn, Osthus and Treglown as well as those of Khan mentioned above,  $m_1^s(k, n)$  is determined exactly whenever  $s \leq n/k$  and  $k \leq 4$  (for details see the concluding remarks in [53]). More generally, we propose the following version of Conjecture 2.1.1 for non-perfect matchings.

**Conjecture 2.1.3.** *For all  $\varepsilon > 0$  and all integers  $n, d, k, s$  with  $1 \leq d \leq k-1$  and  $0 \leq s \leq (1-\varepsilon)n/k$  we have*

$$m_d^s(k, n) = \left( 1 - \left( 1 - \frac{s}{n} \right)^{k-d} + o(1) \right) \binom{n-d}{k-d}.$$

In fact it may be that the bound holds for all  $s \leq n/k - C$ , for some  $C$  depending only on  $d$  and  $k$ . The lower bound here is given by  $H(s)$ . The case  $d = k-1$  of

Conjecture 2.1.3 follows easily from the determination of  $m_{k-1}^s(k, n)$  for  $s$  close to  $n/k$  in [69]. Bollobás, Daykin and Erdős [15] determined  $m_1^s(k, n)$  for small  $s$ , i.e. whenever  $s < n/2k^3$ . For  $1 \leq d \leq k-2$  we are able to determine  $m_d^s(k, n)$  asymptotically for non-perfect matchings of any size at most  $n/2(k-d)$ . Note that this proves Conjecture 2.1.3 in the case  $k/2 \leq d \leq k-2$ , say.

**Theorem 2.1.4.** *Let  $\varepsilon > 0$ , fix integers  $k, d$  with  $1 \leq d \leq k-2$ , and fix  $0 \leq a < \min\{1/2(k-d), (1-\varepsilon)/k\}$ . Then for all  $n$  such that  $n, an \in \mathbb{N}$ ,*

$$m_d^{an}(k, n) = (1 - (1-a)^{k-d} + o(1)) \binom{n-d}{k-d}.$$

## 2.1.2 Large matchings in hypergraphs with many edges

In proving Theorem 2.1.4 it will be useful for us to consider the following related problem. A classical theorem of Erdős and Gallai [31] determines the number of edges in a graph which forces a matching of a given size. In 1965, Erdős [28] made a conjecture which would generalize this to  $k$ -uniform hypergraphs.

**Conjecture 2.1.5.** *Let  $n, k \geq 2$  and  $1 \leq s \leq n/k$  be integers. Then*

$$m_0^s(k, n) = \max \left\{ \binom{ks-1}{k}, \binom{n}{k} - \binom{n-s+1}{k} \right\} + 1.$$

For  $k=3$  this conjecture was verified by Frankl [35]. For the case  $k=4$ , Conjecture 2.1.5 was verified asymptotically by Alon, Frankl, Huang, Rödl, Ruciński and Sudakov [5]. Recently, Frankl confirmed the conjecture exactly for  $s \leq n/2k$ , i.e. when the aim is to cover at most half of the vertices of the hypergraph.

**Theorem 2.1.6.** [34] *Let  $n, k, s \in \mathbb{N}$  be such that  $n, k \geq 2$  and  $n \geq (2s-1)k - s + 1$ .*

Then

$$m_0^s(k, n) = \binom{n}{k} - \binom{n-s+1}{k} + 1.$$

It is possible to prove a variant of Theorem 2.1.6 that, for small values of  $k$ , yields the result for a larger range of  $s$  (see Theorem 2.6.1).

### 2.1.3 Large fractional matchings

Our approach to proving our results uses the concepts of fractional matchings and fractional vertex covers. A *fractional matching* in a  $k$ -uniform hypergraph  $G = (V, E)$  is a function  $w : E \rightarrow [0, 1]$  of weights of edges, such that for each  $v \in V$  we have  $\sum_{e \in E: v \in e} w(e) \leq 1$ . The *size* of  $w$  is  $\sum_{e \in E} w(e)$ . We say  $w$  is *perfect* if it has size  $|V|/k$ . A *fractional vertex cover* in  $G$  is a function  $w : V \rightarrow [0, 1]$  of weights of vertices, such that for each  $e \in E$  we have  $\sum_{v \in e} w(v) \geq 1$ . The *size* of  $w$  is  $\sum_{v \in V} w(v)$ .

A key idea (already used e.g. in [5, 68]) is that we can switch between considering the largest fractional matching and the smallest fractional vertex cover of a hypergraph. The determination of these quantities are dual linear programming problems, and hence by the Duality Theorem they have the same size.

For  $s \in \mathbb{R}$  we let  $f_d^s(k, n)$  denote the minimum integer  $m$  such that every  $k$ -uniform hypergraph  $G$  on  $n$  vertices with  $\delta_d(G) \geq m$  has a fractional matching of size  $s$ . It was shown in [68] that  $f_{k-1}^{n/k}(k, n) = \lceil n/k \rceil$ . Recently, Treglown and Zhao have extended their work in [78, 79] to determine  $m_d^{n/k}(k, n)$  also in the case when  $d < k/2$  and  $f_d^{n/k}(k, n)$  is significantly less than  $\binom{n}{k-d}/2$  (see [80]).

To prove Theorem 2.1.4, we use Theorem 2.1.6, along with methods similar to those de-

veloped in [5], to convert the edge-density conditions for the existence of matchings into corresponding minimum degree conditions for the existence of fractional matchings (see Proposition 2.4.1). We then use the Weak Hypergraph Regularity Lemma to prove Theorem 2.1.4 by converting our fractional matchings into integer ones. Note that applying this method with Theorem 2.6.1 rather than Theorem 2.1.6 yields a variant of Theorem 2.1.4 that, for small values of  $k$ , holds for a larger range of  $a$  (see Theorem 2.6.2). Our argument also gives the following theorem which, for  $1 \leq d \leq k-2$ , asymptotically determines  $f_d^s(k, n)$  for fractional matchings of any size up to  $n/2(k-d)$ . Note that this determines  $f_d^s(k, n)$  asymptotically for all  $s \in (0, n/k)$  whenever  $d \geq k/2$ .

**Theorem 2.1.7.** *Let  $n, k \geq 3$ , and  $1 \leq d \leq k-2$  be integers and let  $0 \leq a \leq \min\{1/2(k-d), 1/k\}$ . Then*

$$f_d^{an}(k, n) = (1 - (1-a)^{k-d} + o(1)) \binom{n-d}{k-d}.$$

We prove Theorem 2.1.2 in a similar fashion, via the following two theorems.

**Theorem 2.1.8.** *Let  $n, k \geq 2$ ,  $d \geq 1$  be integers. Then*

$$f_0^{n/(k+d)}(k, n) \leq \left( \frac{k}{k+d} - \frac{k-1}{(k+d)^k} + o(1) \right) \binom{n}{k}.$$

**Theorem 2.1.9.** *Let  $n, k \geq 3$ ,  $1 \leq d \leq k-2$  be integers. Then*

$$f_d^{n/k}(k, n) \leq \left( \frac{k-d}{k} - \frac{k-d-1}{k^{k-d}} + o(1) \right) \binom{n-d}{k-d}.$$

The rest of the chapter is organised as follows. In Section 2.2 we lay out some notation, set out some useful tools, and prove some preliminary results. Section 2.3 is the heart of the chapter, in which we prove Theorem 2.1.8. In Section 2.4 we derive Theorems 2.1.7 and 2.1.9, and in Section 2.5 we derive Theorems 2.1.2 and 2.1.4. We conclude with Section 2.6, where we prove variants of Theorems 2.1.4 and 2.1.6, as mentioned above.

## 2.2 Notation, tools and preliminary results

### 2.2.1 Notation

Since in many of the proofs in this chapter we often consider vertex degrees, when  $S = \{v\}$  is a set containing only one vertex we write  $d_G(v)$  to denote  $\deg_G(S)$  and we refer to  $d_G(v)$  as the degree of  $v$  (in  $G$ ). We let  $e(G)$  denote the number of edges in a hypergraph  $G$ , and let  $|G|$  denote the number of its vertices. For a set  $V$  and a positive integer  $k$  we let  $\binom{V}{k}$  denote the set of all  $k$ -element subsets of  $V$ . For  $m \in \mathbb{N}$  we let  $[m]$  denote the set  $\{1, \dots, m\}$ . Whenever we refer to a  $k$ -tuple, we assume that it is unordered. Given a hypergraph  $G = (V, E)$  and a set  $S \subseteq V$ , we refer to the pair  $(V \setminus S, \{e \subseteq V : S \cap e = \emptyset, e \cup S \in E\})$  as the *neighbourhood hypergraph* of  $S$  (in  $G$ ). If  $S = \{v\}$  has just one element then we may refer to this pair as the neighbourhood hypergraph of  $v$ . For  $U \subseteq V$  we denote by  $G[U]$  the hypergraph induced by  $U$  on  $G$ , that is the hypergraph with vertex set  $U$  and edge set  $\{e \in E : e \subseteq U\}$ .

### 2.2.2 Tools and preliminary results

In proving some of our results we will use the lower bound given by the earlier construction  $H(s)$ , for all integers  $n, d, k, s$  with  $k \geq 2$  and  $0 \leq d \leq k - 1$  and  $0 \leq s \leq n/k$ :

$$m_d^s(k, n) \geq f_d^s(k, n) \geq (1 - (1 - s/n)^{k-d} + o(1)) \binom{n-d}{k-d}. \quad (2.2.1)$$

Now, as mentioned in Section 2.1, a key tool in this chapter is that the determination of the size of the largest fractional matching of a  $k$ -uniform hypergraph is a linear programming

problem, and its dual problem is to determine the size of the smallest fractional vertex cover of the hypergraph. The following proposition, which follows by the Duality Theorem, will be very useful to us.

**Proposition 2.2.2.** *Let  $k \geq 2$  and let  $G$  be a  $k$ -uniform hypergraph. The size of the largest fractional matching of  $G$  is equal to the size of the smallest fractional vertex cover of  $G$ .*

In the rest of this section we collect some preliminary results.

**Proposition 2.2.3.** *Let  $G = (V, E)$  be a hypergraph,  $E' \subseteq E$ ,  $S \subseteq V$ , and let  $w$  be a fractional vertex cover of  $G$ . Then*

$$e(G) \leq \sum_{e \in E} \sum_{v \in e \setminus S} w(v) + \sum_{e \in E'} \sum_{v \in e \cap S} w(v) + |E \setminus E'|.$$

*Proof.* As  $w$  is a fractional vertex cover of  $G$ ,

$$\begin{aligned} e(G) &= |E'| + |E \setminus E'| \leq \sum_{e \in E'} \sum_{v \in e} w(v) + |E \setminus E'| \\ &\leq \sum_{e \in E} \sum_{v \in e \setminus S} w(v) + \sum_{e \in E'} \sum_{v \in e \cap S} w(v) + |E \setminus E'|. \end{aligned}$$

□

The following crude bound will sometimes be useful.

**Proposition 2.2.4.** *Suppose that  $k \geq 2$  and  $0 < a, c < 1$  are fixed. Then for every  $\varepsilon > 0$  there exists  $n_0 = n_0(k, \varepsilon)$  such that if  $n \geq n_0$  and  $f_0^{an}(k, n) \leq c \binom{n}{k}$  then  $f_0^{an+1}(k, n) \leq (c + \varepsilon) \binom{n}{k}$ .*

*Proof.* Suppose  $G$  is a  $k$ -uniform hypergraph on  $n$  vertices with  $e(G) \geq (c + \varepsilon) \binom{n}{k}$ . Choose an arbitrary hyperedge  $e \in E(G)$ , and delete all edges incident to any  $v \in e$ , to form the



new hypergraph  $G'$ . Then

$$e(G') \geq e(G) - k \binom{n-1}{k-1} \geq (c + \varepsilon) \binom{n}{k} - k \binom{n-1}{k-1} \geq c \binom{n}{k},$$

where the last inequality holds as  $n_0$  is sufficiently large. So by assumption,  $G'$  has a fractional matching  $M$  of size  $an$ . Note then that  $M \cup \{e\}$  is a fractional matching of  $G$ . So indeed  $G$  has a fractional matching of size  $an + 1$ , as required.  $\square$

In the next section we will prove Theorem 2.1.8 by induction. For this we will need Theorem 2.2.5, which will establish the base case of this induction. Theorem 2.2.5 is an easy consequence of the Erdős-Gallai Theorem from [31].

**Theorem 2.2.5.** *For  $k = 2$  and  $x \leq 1/3$  we have*

$$f_0^{xn}(k, n) = (1 - (1 - x)^k + o(1)) \binom{n}{k}.$$

The next proposition will also be needed in the proof of Theorem 2.1.8. To prove this proposition we will need a well-known theorem of Baranyai [13] from 1975.

**Theorem 2.2.6 (Baranyai's Theorem).** *If  $n \in \ell\mathbb{N}$  then the complete  $\ell$ -uniform hypergraph on  $n$  vertices decomposes into edge-disjoint perfect matchings.*

**Proposition 2.2.7.** *Let  $n, k, \ell$  be integers with  $k \geq 2$  and  $1 \leq \ell \leq k$ , and let  $\eta \in [0, 1]$ . Let  $V$  be a set of size  $n$ . Suppose  $S \subseteq V$ , with  $|S| \in \ell\mathbb{N}$ . Then there exists  $\tilde{E} \subseteq \{e \in \binom{V}{k} : |e \cap S| = \ell\}$  such that for every  $v \in S$ ,*

$$|\{e \in \tilde{E} : v \in e\}| = \left\lfloor \eta \binom{|S|}{\ell-1} \binom{n-|S|}{k-\ell} \right\rfloor. \quad (2.2.8)$$

*Proof.* The cases where  $\ell = 1$  or  $\eta = 0$  are trivial. So suppose that  $2 \leq \ell \leq k$  and

$\eta \in (0, 1)$ . Apply Theorem 2.2.6 to find a decomposition of the complete  $\ell$ -uniform hypergraph on  $S$  into edge-disjoint perfect matchings  $M_1, \dots, M_{\binom{|S|-1}{\ell-1}}$ .

We now construct  $\tilde{E}$  by adding  $k$ -tuples from  $\{e \in \binom{V}{k} : |e \cap S| = \ell\}$  greedily, under the following constraints:

(i) for all  $i \in \{1, \dots, \binom{|S|-1}{\ell-1}\}$ , we do not add any  $k$ -tuples in  $\{e \in \binom{V}{k} : e \cap S \in M_{i+1}\}$  unless we have already added all  $k$ -tuples in  $\{e \in \binom{V}{k} : e \cap S \in M_i\}$ ;

(ii) for every  $v \in S$ ,

$$|\{e \in \tilde{E} : v \in e\}| \leq \eta \binom{|S|}{\ell-1} \binom{n-|S|}{k-\ell}.$$

It is clear that (i) and (ii) ensure that the set  $\tilde{E}$  obtained in this way satisfies (2.2.8) for every  $v \in S$ .  $\square$

## 2.3 Minimum edge-density conditions for fractional matchings

We will use the following lemma to prove Theorem 2.1.8 inductively.

**Lemma 2.3.1.** *Let  $k \geq 3$  be fixed. Suppose that  $a \in (0, 1/(k+1)]$ ,  $c \in (0, 1)$  and that there exists  $n_0 \in \mathbb{N}$  such that for all  $n \geq n_0$  we have*

$$f_0^{an/(1-a)}(k-1, n) \leq c \binom{n}{k-1}. \quad (2.3.2)$$

*Then for all  $\varepsilon > 0$  there exists  $n_1 \in \mathbb{N}$  such that for all  $n \geq n_1$  any  $k$ -uniform hypergraph*

$G$  on  $n$  vertices with at least  $an$  vertices of degree at least

$$D := (c(1-a)^{k-1} + (1 - (1-a)^{k-1}) + \varepsilon) \binom{n-1}{k-1}$$

has a fractional matching of size  $an$ .

*Proof.* Let  $\varepsilon > 0$  and choose  $n_1$  sufficiently large. Consider a  $k$ -uniform hypergraph  $G = (V, E)$  on  $n$  vertices with at least  $an$  vertices of degree at least  $D$ . Let  $Y \subseteq V$  be the set of  $\lceil an \rceil$  vertices of highest degree. Let  $w$  be a fractional vertex cover of  $G$  of least size. Consider the vertex  $v_0 \in Y$  with the lowest weight  $w(v_0)$ . Let  $H$  be the neighbourhood hypergraph of  $v_0$  in  $G$ . So

$$e(H) = d_G(v_0) \geq D = (c(1-a)^{k-1} + (1 - (1-a)^{k-1}) + \varepsilon) \binom{n-1}{k-1}.$$

Let  $H' := H[V \setminus Y]$ . Since the number of edges in  $H$  with at least one vertex in  $Y$  is at most  $(1 - (1-a)^{k-1} + o(1)) \binom{n-1}{k-1}$ , it follows that

$$\begin{aligned} e(H') &\geq e(H) - (1 - (1-a)^{k-1} + o(1)) \binom{n-1}{k-1} \geq (c(1-a)^{k-1} + \varepsilon/2) \binom{n-1}{k-1} \\ &\geq (c + \varepsilon/3) \binom{|H'|}{k-1}, \end{aligned}$$

where in the last two inequalities we use that  $n_1$  was chosen sufficiently large. Note that  $|H'| \geq n/2$ , so we may assume that  $|H'| \geq n_0$ . Now, (2.3.2) and Proposition 2.2.4 together imply that  $H'$  has a fractional matching of size

$$a|H'|/(1-a) + 1 = a(n - \lceil an \rceil)/(1-a) + 1 \geq an.$$

So let  $M$  be a fractional matching of  $H'$  of size  $an$ . Note that for all  $v \in V \setminus Y$ ,

$$\sum_{e \in E(H') : v \in e} M(e) \leq 1.$$

So we have that

$$\sum_{v \in V} w(v) \geq \sum_{v \in Y} w(v) + \sum_{e \in E(H')} \sum_{v \in e} M(e)w(v).$$

By the minimality of  $w(v_0)$ , this implies that

$$\begin{aligned} \sum_{v \in V} w(v) &\geq anw(v_0) + \sum_{e \in E(H')} \sum_{v \in e} M(e)w(v) = \sum_{e \in E(H')} M(e)w(v_0) + \sum_{e \in E(H')} \sum_{v \in e} M(e)w(v) \\ &= \sum_{e \in E(H')} M(e) \left( w(v_0) + \sum_{v \in e} w(v) \right) \geq \sum_{e \in E(H')} M(e) = an. \end{aligned}$$

The last inequality holds because by definition of  $H'$  we have  $e \cup \{v_0\} \in E$  for all  $e \in E(H')$ , and so  $w(v_0) + \sum_{v \in e} w(v) \geq 1$ .

Hence the size of  $w$  is at least  $an$ , so by Proposition 2.2.2 the largest fractional matching in  $G$  has size at least  $an$ .  $\square$

The proof of Theorem 2.1.8 proceeds as follows. Suppose  $G$  has no fractional matching of size  $n/(k+d)$ . Then we use Lemma 2.3.1 and induction to show that  $G$  contains few vertices of high degree. Moreover, by duality we show that  $G$  has a small fractional vertex cover. We combine these two facts to show that the number of edges of  $G$  does not exceed the expression stated in Theorem 2.1.8.

**Proof of Theorem 2.1.8.** The proof will proceed by induction on  $k$ . The base step,  $k = 2$ , follows by Theorem 2.2.5, setting  $x := 1/(2+d)$ .

Now consider some  $k > 2$  and suppose that the theorem holds for all smaller values of  $k$ . Fix  $d \geq 1$ . Let  $\varepsilon > 0$  and let  $n_0 \in \mathbb{N}$  be sufficiently large compared to  $1/\varepsilon$ ,  $k$  and  $d$ . For convenience let us define

$$\xi := \left( \frac{k-1}{k+d-1} - \frac{k-2}{(k+d-1)^{k-1}} \right) \left( \frac{k+d-1}{k+d} \right)^{k-1} + \left( 1 - \left( \frac{k+d-1}{k+d} \right)^{k-1} \right) < 1.$$

Consider any  $k$ -uniform hypergraph  $G = (V, E)$  on  $n \geq n_0$  vertices, and suppose that the largest fractional matching of  $G$  is of size less than  $n/(k+d)$ . Then by Proposition 2.2.2 there exists a fractional vertex cover, *w* say, of  $G$  with size less than  $n/(k+d)$ . Let  $a := 1/(k+d)$ . So  $a/(1-a) = 1/(k+d-1)$ . Let

$$c := \frac{k-1}{k+d-1} - \frac{k-2}{(k+d-1)^{k-1}} + \varepsilon/4.$$

Then by induction,

$$f_0^{n'/(k+d-1)}(k-1, n) \leq c \binom{n'}{k-1},$$

for all sufficiently large  $n'$ . Thus, as  $n_0$  is sufficiently large, Lemma 2.3.1 implies that there are less than  $n/(k+d)$  vertices of  $G$  with degree at least  $(\xi + \varepsilon/2) \binom{n-1}{k-1}$ .

Let  $S$  be the set of  $|S|$  vertices of  $G$  with highest degree, where  $|S| \in k!\mathbb{N}$  is minimal such that  $|S| \geq n/(k+d)$ . So  $d_G(v) < (\xi + \varepsilon/2) \binom{n-1}{k-1}$  for all  $v \in V \setminus S$ . For every  $i \in \{0, \dots, k\}$  let  $S_i := \{e \in \binom{V}{k} : |e \cap S| = i\}$ . Given  $X \subseteq \binom{V}{k}$ , for all  $v \in V$  let  $t_X(v) := |\{e \in X : v \in e\}|$ . Note that for all  $v \in S$  the value of  $t_{S_i}(v)$  is the same and  $t_{S_0}(v) = 0$ . Let  $\ell \in \{0, \dots, k\}$  be maximal such that for any  $v \in S$  we have  $\sum_{i=0}^{\ell-1} t_{S_i}(v) \leq \xi \binom{n-1}{k-1}$ . Let  $E''' := \binom{V}{k} \setminus S_k$ . Then for each  $v \in S$ ,

$$t_{E'''}(v) = \left( 1 - \frac{1}{(k+d)^{k-1}} + o(1) \right) \binom{n-1}{k-1} > \xi \binom{n-1}{k-1}. \quad (2.3.3)$$

The final inequality holds here for sufficiently large  $n_0$ , as it rearranges to  $d(k+d-1)^{k-2} + (k-2) + o(1) > 1$ . This shows that  $\ell \leq k-1$ . Let

$$\eta := \left( \xi \binom{n-1}{k-1} - \sum_{i=1}^{\ell-1} t_{S_i}(v) \right) / \binom{|S|}{\ell-1} \binom{n-|S|}{k-\ell}.$$

So  $\eta \in [0, 1)$ . Apply Proposition 2.2.7 with parameters  $n, k, \ell, \eta$  to obtain a set  $\tilde{E} \subseteq S_\ell$  such that for every  $v \in S$ ,

$$t_{\tilde{E}}(v) = \left\lfloor \eta \binom{|S|}{\ell-1} \binom{n-|S|}{k-\ell} \right\rfloor.$$

Let  $E'' := \bigcup_{i=0}^{\ell-1} S_i \cup \tilde{E}$ . Then each  $v \in S$  satisfies

$$t_{E''}(v) = \left\lfloor \xi \binom{n-1}{k-1} \right\rfloor. \quad (2.3.4)$$

We can now give a lower bound on the size of  $E''$  as follows: for each vertex  $v \in S$  we count the number of  $k$ -tuples in  $E''$  that contain  $v$ , and then adjust for the  $k$ -tuples that contain several vertices of  $S$  and were thus counted several times as a result. Since  $S_0 \subseteq E''$  this yields

$$|E''| \stackrel{(2.3.4)}{\geq} \left\lfloor \xi \binom{n-1}{k-1} \right\rfloor \frac{n}{k+d} + |S_0| - \sum_{j=1}^{k-1} (j-1) |S_j|.$$

Note that since  $E'' \subseteq E'''$  we only need to consider values of  $j$  up to  $k-1$  in the summation, rather than  $k$ . Now, note that

$$\begin{aligned} |S_0| - \sum_{j=1}^k (j-1) |S_j| &= \binom{n}{k} - \sum_{v \in S} \binom{n-1}{k-1} = \binom{n}{k} - \left( \frac{n}{k+d} + o(1) \right) \binom{n-1}{k-1} \\ &= \left( 1 - \frac{k}{k+d} + o(1) \right) \binom{n}{k}. \end{aligned}$$

Hence, as  $(k-1)|S_k| = ((k-1)/(k+d)^k + o(1))\binom{n}{k}$ ,

$$|E''| \geq (\xi + o(1))\binom{n-1}{k-1}\frac{n}{k+d} + \left(1 - \frac{k}{k+d} + \frac{k-1}{(k+d)^k} + o(1)\right)\binom{n}{k}. \quad (2.3.5)$$

Now, let  $E' := E \cap E''$ . Also, note that by Proposition 2.2.3,

$$e(G) \leq \sum_{e \in E} \sum_{v \in e \setminus S} w(v) + \sum_{e \in E'} \sum_{v \in e \cap S} w(v) + |E \setminus E'|.$$

Recall that  $d_G(v) < (\xi + \varepsilon/2)\binom{n-1}{k-1}$  for all  $v \in V \setminus S$  and that by (2.3.4) the number of edges in  $E'$  incident to  $v$  is at most  $\xi\binom{n-1}{k-1}$  for all  $v \in S$ . So

$$e(G) \leq \sum_{v \in V} (\xi + \varepsilon/2)\binom{n-1}{k-1}w(v) + |E \setminus E'|.$$

Now note that  $|E \setminus E'| \leq |\binom{V}{k} \setminus E''| = \binom{n}{k} - |E''|$  and recall that the size of  $w$  is less than  $n/(k+d)$ . So

$$\begin{aligned} e(G) &< (\xi + \varepsilon/2)\binom{n-1}{k-1}\frac{n}{k+d} + \binom{n}{k} - |E''| \\ &\stackrel{(2.3.5)}{\leq} (\xi + \varepsilon/2)\binom{n-1}{k-1}\frac{n}{k+d} - (\xi + o(1))\binom{n-1}{k-1}\frac{n}{k+d} \\ &\quad + \left(\frac{k}{k+d} - \frac{k-1}{(k+d)^k} + o(1)\right)\binom{n}{k} \\ &\leq \left(\frac{k}{k+d} - \frac{k-1}{(k+d)^k} + \varepsilon\right)\binom{n}{k}. \end{aligned}$$

The final inequality holds since  $n_0$  is sufficiently large. By definition, this shows that

$$f_0^{n/(k+d)}(k, n) \leq \left(\frac{k}{k+d} - \frac{k-1}{(k+d)^k} + o(1)\right)\binom{n}{k}.$$

This completes the inductive step and hence the proof.  $\square$

## 2.4 Minimum degree conditions for fractional matchings

The following proposition generalises Proposition 1.1 in [5], with a similar proof idea. It allows us to transform bounds involving edge densities into bounds involving  $d$ -degrees.

**Proposition 2.4.1.** *Let  $\varepsilon \geq 0$ , let  $k, d, n$  be integers with  $n \geq k \geq 3$ ,  $1 \leq d \leq k - 2$ , and  $d < (1 - \varepsilon^{1/d})n$ . Let  $a \in [0, (1 - \varepsilon^{1/d})/k]$ . Suppose  $H$  is a  $k$ -uniform hypergraph on  $n$  vertices, such that for at least  $(1 - \varepsilon) \binom{n}{d}$   $d$ -tuples of vertices  $L \in \binom{V(H)}{d}$  we have*

$$\deg_H(L) \geq f_0^{an}(k - d, n - d).$$

*Then  $H$  has a fractional matching of size  $an$ .*

*Proof.* The outline of the proof goes as follows. We will assume that there is no fractional matching of size  $an$  in a  $k$ -uniform hypergraph  $H = (V, E)$  on  $n$  vertices and then show that for more than  $\varepsilon \binom{n}{d}$   $d$ -tuples of vertices  $L \in \binom{V}{d}$ , the neighbourhood hypergraph  $H(L)$  of  $L$  in  $H$  has no fractional matching of size  $an$ . This will imply that for more than  $\varepsilon \binom{n}{d}$   $d$ -tuples of vertices  $L$ ,  $\deg_H(L) = e(H(L)) < f_0^{an}(k - d, n - d)$ . This will prove the result in contrapositive.

So suppose  $H = (V, E)$  is an  $n$ -vertex  $k$ -uniform hypergraph, with no fractional matching of size  $an$ . Then by Proposition 2.2.2,  $H$  has a fractional vertex cover,  $w$  say, of size less than  $an$ . Let

$$E_w := \left\{ e \in \binom{V}{k} : \sum_{v \in e} w(v) \geq 1 \right\},$$

and let  $H_w := (V, E_w)$ . Since  $H \subseteq H_w$  we can, without loss of generality, replace  $H$  with  $H_w$ . Let  $U \subseteq V$  be the set of  $\lfloor \varepsilon^{1/d} n \rfloor + d$  vertices of smallest weights. Let  $\mathcal{L} := \binom{U}{d}$ . Note



that

$$|\mathcal{L}| = \binom{\lfloor \varepsilon^{1/d} n \rfloor + d}{d} > \frac{(\varepsilon^{1/d} n)^d}{d!} = \varepsilon \frac{n^d}{d!} \geq \varepsilon \binom{n}{d}.$$

Consider any  $L \in \mathcal{L}$ . Let  $H_w(L)$  be the neighbourhood hypergraph of  $L$  in  $H_w$ . We will show that  $H_w(L)$  has no fractional matching of size  $an$ . Without loss of generality we may assume that the elements of  $L$  all have equal weights,  $w(L)$  say. (If not, we could replace these weights by their average, which would alter neither  $\sum_{v \in V} w(v)$  nor  $\sum_{v \in e} w(v)$  for any  $e \supseteq L$ . These are the only two quantities involving weights that we will consider in what follows.) Observe that  $w(L) < 1/k$ , else the size of  $w$  would be at least

$$\frac{n(1 - \varepsilon^{1/d})}{k} \geq an.$$

We now define a new weight function  $w'(v)$  on the vertices in  $V$ :

$$w'(v) := \min \{ \max \{ 0, w^*(v) \}, 1 \}, \quad \text{where} \quad w^*(v) := \frac{w(v) - w(L)}{1 - kw(L)}.$$

Note that only for vertices  $u \in U \setminus L$  can it be that  $w^*(u) < 0$ . Note also that since  $w(v) \geq 0$  for all  $v \in V$ , we have that  $w^*(u) \geq -w(L)/(1 - kw(L))$  for such vertices  $u$ . Hence,

$$\begin{aligned} \sum_{v \in V} w'(v) &\leq \left( \sum_{v \in V} w^*(v) \right) + |U \setminus L| \frac{w(L)}{1 - kw(L)} < \frac{an - nw(L) + \varepsilon^{1/d} nw(L)}{1 - kw(L)} \\ &= an \frac{1 - (1/a)(1 - \varepsilon^{1/d})w(L)}{1 - kw(L)} \leq an, \end{aligned}$$

and for any given  $e \in \{e' \in E_w : e' \supseteq L\}$  we have that

$$\sum_{v \in e} w'(v) \geq \min \left\{ \frac{\sum_{v \in e} w(v) - kw(L)}{1 - kw(L)}, 1 \right\} \geq \min \left\{ \frac{1 - kw(L)}{1 - kw(L)}, 1 \right\} = 1.$$

Moreover,  $\sum_{v \in L} w'(v) = 0$ . It follows that the function  $w'$  restricted to  $V \setminus L$  is a fractional

vertex cover of  $H_w(L)$  of size less than  $an$ , and so by Proposition 2.2.2,  $H_w(L)$  has no fractional matching of size  $an$ , which completes the proof.  $\square$

We can now derive Theorems 2.1.7 and 2.1.9.

**Proof of Theorem 2.1.7.** Let  $k' := k - d$  and  $n' := n - d$ . Note that Theorem 2.1.6 implies that

$$m_0^{an}(k, n) = (1 - (1 - a)^k + o(1)) \binom{n}{k} \quad (2.4.2)$$

for all  $a \leq 1/2k$ . Now Proposition 2.2.4 implies that for all  $0 \leq a \leq \min\{1/2(k-d), 1/k\}$ ,

$$\begin{aligned} f_0^{an}(k-d, n-d) &= f_0^{a(n'+d)}(k', n') \leq f_0^{an'+1}(k', n') \leq m_0^{an'}(k', n') + o(1) \binom{n'}{k'} \\ &\stackrel{(2.4.2)}{=} \left(1 - (1 - a)^{k'} + o(1)\right) \binom{n'}{k'} = (1 - (1 - a)^{k-d} + o(1)) \binom{n-d}{k-d}. \end{aligned}$$

The upper bound in Theorem 2.1.7 follows now from Proposition 2.4.1 applied with  $\varepsilon = 0$ .

The lower bound follows from (2.2.1).  $\square$

**Proof of Theorem 2.1.9.** Let  $k' := k - d$  and  $n' := n - d$ . Then Theorem 2.1.8 and Proposition 2.2.4 together imply that

$$\begin{aligned} f_0^{n/k}(k-d, n-d) &= f_0^{(n'+d)/(k'+d)}(k', n') \leq f_0^{n'/(k'+d)+1}(k', n') \\ &\leq \left(\frac{k'}{k'+d} - \frac{k'-1}{(k'+d)^{k'}} + o(1)\right) \binom{n'}{k'} \\ &= \left(\frac{k-d}{k} - \frac{k-d-1}{k^{k-d}} + o(1)\right) \binom{n-d}{k-d}. \end{aligned}$$

So Theorem 2.1.9 follows now from Proposition 2.4.1 applied with  $\varepsilon = 0$ .  $\square$

The case  $\varepsilon > 0$  of Proposition 2.4.1 will be used in the next section.

## 2.5 Constructing integer matchings from fractional ones

We will construct integer matchings from fractional ones using the Weak Hypergraph Regularity Lemma. Before stating this we will need the following definitions.

Given a  $k$ -tuple  $(V_1, \dots, V_k)$  of disjoint subsets of the vertices of a  $k$ -uniform hypergraph  $G = (V, E)$ , we define  $(V_1, \dots, V_k)_G$  to be the  $k$ -partite subhypergraph with vertex classes  $V_1, \dots, V_k$  induced on  $G$ . We let

$$d_G(V_1, \dots, V_k) = \frac{e((V_1, \dots, V_k)_G)}{\prod_{i \in \{1, \dots, k\}} |V_i|}$$

denote the *density* of  $(V_1, \dots, V_k)_G$ .

**Definition 2.5.1 ( $\varepsilon$ -regularity).** *Let  $\varepsilon > 0$ , let  $G = (V, E)$  be a  $k$ -uniform hypergraph, and let  $V_1, \dots, V_k \subseteq V$  be disjoint. We say that  $(V_1, \dots, V_k)_G$  is  $\varepsilon$ -regular if for every subhypergraph  $(V'_1, \dots, V'_k)_G$  with  $V'_i \subseteq V_i$  and  $|V'_i| \geq \varepsilon |V_i|$  for each  $i \in \{1, \dots, k\}$ , we have that*

$$|d_G(V'_1, \dots, V'_k) - d_G(V_1, \dots, V_k)| < \varepsilon.$$

The following result was proved by Chung [23]. The proof follows the lines of that of the original Regularity Lemma for graphs [76].

**Lemma 2.5.2 (Weak Hypergraph Regularity Lemma).** *For all integers  $k \geq 2$ ,  $L_0 \geq 1$ , and every  $\varepsilon > 0$  there exists  $N = N(\varepsilon, L_0, k)$  such that if  $G = (V, E)$  is a  $k$ -uniform hypergraph on  $n \geq N$  vertices, then  $V$  has a partition  $V_0, \dots, V_L$  such that the following properties hold:*

- (i)  $L_0 \leq L \leq N$  and  $|V_0| \leq \varepsilon n$ ,
- (ii)  $|V_1| = \dots = |V_L|$ ,
- (iii) for all but at most  $\varepsilon \binom{L}{k}$   $k$ -tuples  $\{i_1, \dots, i_k\} \in \binom{[L]}{k}$ , we have that  $(V_{i_1}, \dots, V_{i_k})_G$  is  $\varepsilon$ -regular.

We call the partition classes  $V_1, \dots, V_L$  *clusters*, and  $V_0$  the *exceptional set*. For our purposes we will in fact use the degree form of the Weak Hypergraph Regularity Lemma.

**Lemma 2.5.3 (Degree Form of the Weak Hypergraph Regularity Lemma).** *For all integers  $k \geq 2$ ,  $L_0 \geq 1$  and every  $\varepsilon > 0$ , there is an  $N = N(\varepsilon, L_0, k)$  such that for every  $d \in [0, 1)$  and for every hypergraph  $G = (V, E)$  on  $n \geq N$  vertices there exists a partition of  $V$  into  $V_0, V_1, \dots, V_L$  and a spanning subhypergraph  $G'$  of  $G$  such that the following properties hold:*

- (i)  $L_0 \leq L \leq N$  and  $|V_0| \leq \varepsilon n$ ,
- (ii)  $|V_1| = \dots = |V_L| =: m$ ,
- (iii)  $d_{G'}(v) > d_G(v) - (d + \varepsilon)n^{k-1}$  for all  $v \in V$ ,
- (iv) every edge of  $G'$  with more than one vertex in a single cluster  $V_i$ , for some  $i \in \{1, \dots, L\}$ , has at least one vertex in  $V_0$ ,
- (v) for all  $k$ -tuples  $\{i_1, \dots, i_k\} \in \binom{[L]}{k}$ , we have that  $(V_{i_1}, \dots, V_{i_k})_{G'}$  is  $\varepsilon$ -regular and has density either 0 or greater than  $d$ .

The proof is very similar to that of the degree form of the Regularity Lemma for graphs, and will use the following easy proposition.

**Proposition 2.5.4.** *Let  $\varepsilon' \leq \rho \leq 1/2$ . Suppose  $G$  is a  $k$ -uniform hypergraph and  $(V'_1, \dots, V'_k)_G$  is a  $\varepsilon'$ -regular subhypergraph of  $G$  with vertex classes  $V'_1, \dots, V'_k$ , all of size  $m$ . Suppose that there are vertex sets  $V_i \subseteq V'_i$  with  $|V_i| \geq \rho m$  for all  $i \in \{1, \dots, k\}$ . Then  $(V_1, \dots, V_k)_G$  is  $(\varepsilon'/\rho)$ -regular.*

*Proof.* Consider any  $V_i^* \subseteq V_i$  with  $|V_i^*| \geq \varepsilon'|V_i|/\rho$  for all  $i \in \{1, \dots, k\}$ . Then  $|V_i^*| \geq (\varepsilon'/\rho)\rho m = \varepsilon'm$ . So as  $(V'_1, \dots, V'_k)_G$  is an  $\varepsilon'$ -regular  $k$ -tuple, it follows that

$$|d_G(V_1^*, \dots, V_k^*) - d_G(V'_1, \dots, V'_k)| \leq \varepsilon'.$$

Similarly, as  $\rho \geq \varepsilon'$ , we have that

$$|d_G(V_1, \dots, V_k) - d_G(V'_1, \dots, V'_k)| \leq \varepsilon'.$$

Hence,

$$|d_G(V_1^*, \dots, V_k^*) - d_G(V_1, \dots, V_k)| \leq 2\varepsilon' \leq \frac{\varepsilon'}{\rho}.$$

So by definition,  $(V_1, \dots, V_k)_G$  is an  $(\varepsilon'/\rho)$ -regular  $k$ -tuple. □

We will use the notation  $a \ll b$  to mean that we can find an increasing function  $f$  for which all of the conditions in the proof are satisfied whenever  $a \leq f(b)$ .

**Proof of Theorem 2.5.3.** Let  $\varepsilon > 0$ ,  $k \geq 2$ ,  $L_0 \in \mathbb{N}$  and  $d \in [0, 1)$ . We may assume that  $\varepsilon \leq 1$ . We choose further positive constants  $\varepsilon', L'_0$  satisfying

$$\frac{1}{L'_0}, \varepsilon' \ll \varepsilon, d, \frac{1}{L_0}, \frac{1}{k}$$

By the Weak Hypergraph Regularity Lemma, there exists  $N' = N'(\varepsilon', L'_0, k)$  such that if

we let  $N := 4N'/\varepsilon \geq N'$  and  $G = (V, E)$  is a  $k$ -uniform hypergraph on  $n \geq N$  vertices,  $G$  has a partition of its vertices into  $V'_0, \dots, V'_{L'}$  such that:

- (a)  $L'_0 \leq L' \leq N'$  and  $|V'_0| \leq \varepsilon' n$ ,
- (b)  $|V'_1| = \dots = |V'_{L'}| =: m'$ ,
- (c) for all but at most  $\varepsilon' \binom{L'}{k} \leq \varepsilon' L'^k$   $k$ -tuples  $\{i_1, \dots, i_k\} \in \binom{[L']}{k}$ , we have that  $(V'_{i_1}, \dots, V'_{i_k})_G$  is  $\varepsilon'$ -regular.

We will remove some edges from  $G$  to obtain a graph  $G'$  and a partition  $V_0, V_1, \dots, V_L$  of its vertices, satisfying properties (i)–(v), by carrying out the following steps:

- (i) For each  $k$ -tuple  $\{i_1, \dots, i_k\} \in \binom{[L']}{k}$ , if  $(V'_{i_1}, \dots, V'_{i_k})_G$  is not  $\varepsilon'$ -regular, then colour all edges in  $(V'_{i_1}, \dots, V'_{i_k})_G$  red. For any  $v \in V$ , if there are at least  $\varepsilon n^{k-1}/10$  red edges incident to  $v$ , then we move  $v$  to  $V'_0$ . Then delete all red edges that do not have a vertex in  $V'_0$ .

After deleting these edges, we observe that the degree of any vertex  $v \in V$  is greater than  $d_G(v) - \varepsilon n^{k-1}/10$ .

We have at most  $\varepsilon' L'^k m'^k \leq \varepsilon' n^k$  red edges by (c), and so the number of vertices we have moved to  $V'_0$  is at most

$$\frac{k\varepsilon' n^k}{\varepsilon n^{k-1}/10} = \frac{10k\varepsilon' n}{\varepsilon} \leq \frac{\varepsilon n}{4}.$$

- (ii) Next, consider each  $k$ -tuple  $\{i_1, \dots, i_k\} \in \binom{[L']}{k}$  such that  $(V'_{i_1}, \dots, V'_{i_k})_G$  is  $\varepsilon'$ -regular and has density  $d_G(V'_{i_1}, \dots, V'_{i_k}) \leq d + \varepsilon'$ . Colour all edges in these  $k$ -tuples blue.

For each  $v \in V'_{i_1}$  such that there are more than  $(d + 2\varepsilon')m'^{k-1}$  edges in  $(V'_{i_1}, \dots, V'_{i_k})_G$  incident to  $v$ , mark all but  $(d + 2\varepsilon')m'^{k-1}$  of these edges. Proceed similarly for each  $v \in V'_{i_2}, \dots, V'_{i_k}$ .

Let  $X$  be the set of vertices in  $V'_{i_1}$  having more than  $(d + 2\varepsilon')m'^{k-1}$  incident edges in  $(V'_{i_1}, \dots, V'_{i_k})_G$ . Note that

$$d_G(X, V'_{i_2}, \dots, V'_{i_k}) > \frac{(d + 2\varepsilon')m'^{k-1}|X|}{m'^{k-1}|X|} = d + 2\varepsilon'.$$

So since  $(V'_{i_1}, \dots, V'_{i_k})_G$  is  $\varepsilon'$ -regular, we have that  $|X| < \varepsilon'm'$ . Similarly for  $V'_{i_2}, \dots, V'_{i_k}$ . So we mark at most  $k\varepsilon'm'^k$  edges in  $(V'_{i_1}, \dots, V'_{i_k})_G$ .

We carry out this process for all  $\varepsilon'$ -regular  $k$ -tuples of clusters with density at most  $d + \varepsilon'$ . There are at most  $\binom{L'}{k}$  such  $k$ -tuples, so the total number of edges marked is at most

$$\binom{L'}{k} k\varepsilon'm'^k \leq \varepsilon'n^k.$$

- (iii) For every vertex  $v \in V$ , if there are at least  $\varepsilon n^{k-1}/10$  marked edges incident to  $v$ , then move  $v$  to  $V'_0$  and delete all blue edges that do not have a vertex in  $V'_0$ .

For every  $v \in V$  we delete fewer than  $(d + 2\varepsilon')m'^{k-1}\binom{L'-1}{k-1} + \varepsilon n^{k-1}/10$  edges incident to  $v$  in this step. We marked at most  $\varepsilon'n^k$  edges, so the number of vertices we move to  $V'_0$  is at most

$$\frac{k\varepsilon'n^k}{\varepsilon n^{k-1}/10} = \frac{10k\varepsilon'n}{\varepsilon} \leq \frac{\varepsilon n}{4}.$$

- (iv) Delete all those edges which for some  $i \in \{1, \dots, L'\}$  have more than one vertex in  $V'_i$  and which have no vertices in  $V'_0$ .

For every  $v \in V$  the number of edges incident to  $v$  that we delete in this step is fewer than

$$km' \binom{n-2}{k-2} \leq km'n^{k-2} \leq kn^{k-1}/L' \leq kn^{k-1}/L'_0 \leq \varepsilon n^{k-1}/4.$$

- (v) Finally, we ensure that all clusters have the same size by splitting each cluster into smaller subclusters of size  $\lceil \varepsilon n/(4L') \rceil$ . Move the vertices that are left over in each cluster after this process into the exceptional set  $V'_0$ . Call this new exceptional set  $V_0$  and the other new clusters  $V_1, \dots, V_L$ .

We now check that the graph,  $G'$  thus obtained, together with the vertex partition  $V_0, V_1, \dots, V_L$ , satisfies properties (i)–(v); (ii) and (iv) are clear. Let us consider property (i). We have that

$$L_0 \leq L'_0 \leq L' \leq L,$$

and also

$$L \leq \frac{m'}{\lceil (\varepsilon n)/(4L') \rceil} L' \leq \frac{4L'}{\varepsilon} \leq \frac{4N'}{\varepsilon} = N.$$

So we see that  $L_0 \leq L \leq N$ . Using (a) and that we have added at most  $\varepsilon n/4$  vertices to the exceptional set in each of steps (1), (3) and (5), we have that

$$|V_0| \leq \varepsilon' n + \frac{3\varepsilon n}{4} \leq \varepsilon n.$$

So property (i) is satisfied.

For property (iii) we combine our previous observations to see that for every vertex  $v \in V$ ,



the number of edges incident to  $v$  that we have removed is fewer than

$$\begin{aligned} & \varepsilon n^{k-1}/10 + \left( (d + 2\varepsilon')m'^{k-1} \binom{L' - 1}{k - 1} + \varepsilon n^{k-1}/10 \right) + \varepsilon n^{k-1}/4 \\ & \leq (d + 2\varepsilon' + 9\varepsilon/20)n^{k-1} \leq (d + \varepsilon)n^{k-1}. \end{aligned}$$

Hence, for every  $v \in V$  we have that:

$$d_{G'}(v) > d_G(v) - (d + \varepsilon)n^{k-1}.$$

Finally we check that property (v) is satisfied. So consider any  $k$  clusters  $V_{i_1}, \dots, V_{i_k}$ , (where  $\{i_1, \dots, i_k\} \in \binom{[L]}{k}$ ). Then either  $(V_{i_1}, \dots, V_{i_k})_{G'}$  has density 0 or it is an induced subhypergraph of an  $\varepsilon'$ -regular subhypergraph  $(V_{j_1}, \dots, V_{j_k})_G$ , of density greater than  $d + \varepsilon'$ , (for some  $\{j_1, \dots, j_k\} \in \binom{[L']}{k}$ ). Let us assume that  $d_{G'}(V_{i_1}, \dots, V_{i_k}) \neq 0$ . Since  $|V_{i_1}| = \dots = |V_{i_k}| \geq \varepsilon m'/4$ , we can apply Proposition 2.5.4 to see that  $(V_{i_1}, \dots, V_{i_k})_{G'}$  is  $\varepsilon'/(\varepsilon/4)$ -regular with density greater than  $d + \varepsilon' - \varepsilon' = d$ . Together with our choice of  $\varepsilon'$  this implies that  $(V_{i_1}, \dots, V_{i_k})_{G'}$  is  $\varepsilon$ -regular and has density greater than  $d$ , as required.  $\square$

We now define a type of hypergraph that will be essential in our application of the Weak Hypergraph Regularity Lemma.

**Definition 2.5.5 (Reduced Hypergraph).** *Let  $G = (V, E)$  be a  $k$ -uniform hypergraph. Given parameters  $\varepsilon > 0$ ,  $d \in [0, 1)$  and  $L_0 \geq 1$  we define the reduced hypergraph  $R = R(\varepsilon, d, L_0)$  of  $G$  as follows. Apply the degree form of the Weak Hypergraph Regularity Lemma to  $G$ , with parameters  $\varepsilon$ ,  $d$ ,  $L_0$  to obtain a spanning subhypergraph  $G'$  and a partition  $V_0, \dots, V_L$  of  $V$ , with exceptional set  $V_0$  and clusters  $V_1, \dots, V_L$ . Then  $R$  has vertices  $V_1, \dots, V_L$ , and there exists an edge between  $V_{i_1}, \dots, V_{i_k}$  precisely when*

$(V_{i_1}, \dots, V_{i_k})_{G'}$  is  $\varepsilon$ -regular with density greater than  $d$ .

The following lemma tells us that this reduced hypergraph (almost) inherits the minimum degree properties of the original hypergraph. The proof is similar to that of the well known version for graphs, but we include it here for completeness.

**Lemma 2.5.6.** *Suppose  $c > 0$ ,  $k \geq 2$ ,  $1 \leq \ell \leq k - 1$ ,  $L_0 \geq 1$ , and  $0 < \varepsilon \leq d \leq c^3/64$ . Let  $G$  be a  $k$ -uniform hypergraph with  $\delta_\ell(G) \geq c|G|^{k-\ell}$ . Let  $R = R(\varepsilon, d, L_0)$  be the reduced hypergraph of  $G$ . Then at least  $\binom{|R|}{\ell} - d^{1/3}(2k)^\ell |R|^\ell$  of the  $\ell$ -tuples of vertices of  $R$  have degree at least  $(c - 4d^{1/3})|R|^{k-\ell}$ .*

*Proof.* Let  $G'$  be the spanning subhypergraph of  $G$  obtained by applying the degree form of the Weak Hypergraph Regularity Lemma to  $G$  with parameters  $\varepsilon, d, L_0$ ; let  $V_1, \dots, V_L$  denote the vertices of  $R$ , and let  $m$  denote the size of these clusters.

First recall that given any vertex  $x \in V(G')$  we know that  $d_{G'}(x) > d_G(x) - (d + \varepsilon)|G|^{k-1}$ . Note that since  $|V_0| \leq \varepsilon|G|$ , we have that the number of edges incident to  $x$  that contain a vertex in  $V_0$  is at most  $\varepsilon|G|\binom{|G|}{k-2} \leq \varepsilon|G|^{k-1}$ . Hence for all  $v \in V(G' - V_0)$ , we have that

$$d_{G'-V_0}(v) > d_G(v) - (d + 2\varepsilon)|G|^{k-1} \geq d_G(v) - 3d|G|^{k-1}.$$

We call an  $\ell$ -tuple  $A$  of vertices of  $G' - V_0$  *bad* if  $\deg_{G'-V_0}(A) \leq \deg_G(A) - 3d^{1/3}|G|^{k-\ell}$ . So for each  $v \in V(G' - V_0)$  there are at most  $\binom{k-1}{\ell-1}d^{2/3}|G|^{\ell-1}$  bad  $\ell$ -tuples  $A$  with  $v \in A$ . (This follows by double-counting the number of pairs  $(A, e)$  where  $A$  is a bad  $\ell$ -tuple with  $v \in A$  and  $e \in E(G) \setminus E(G' - V_0)$  is an edge containing  $A$ .) This in turn implies that in total at most  $\binom{k-1}{\ell-1}d^{2/3}|G|^\ell$  of the  $\ell$ -tuples  $A$  are bad. Given  $1 \leq s \leq k$  and an  $s$ -tuple  $(V_{i_1}, \dots, V_{i_s})$  of clusters of  $R$ , we say that an  $s$ -tuple  $A$  of vertices of  $G' - V_0$  *lies in*  $(V_{i_1}, \dots, V_{i_s})$  if  $|A \cap V_{i_\alpha}| = 1$  for all  $\alpha \in \{1, \dots, s\}$ . We call an  $\ell$ -tuple  $(V_{i_1}, \dots, V_{i_\ell})$  of clusters of  $R$  *nice* if there are less than  $d^{1/3}m^\ell$  bad  $\ell$ -tuples  $A$  of vertices of  $G' - V_0$

which lie in  $(V_{i_1}, \dots, V_{i_\ell})$ . So less than  $\binom{k-1}{\ell-1} d^{1/3} |G|^\ell / m^\ell \leq d^{1/3} (2k)^\ell |R|^\ell$  of the  $\ell$ -tuples of clusters of  $R$  are not nice. Hence it suffices to show that any nice  $\ell$ -tuple of clusters of  $R$  has degree at least  $(c - 4d^{1/3})|R|^{k-\ell}$  in  $R$ .

Consider any nice  $\ell$ -tuple of clusters of  $R$ , say  $(V_{i_1}, \dots, V_{i_\ell})$ . Let  $\mathcal{A}$  denote the set of all  $\ell$ -tuples  $A$  of vertices of  $G' - V_0$  which lie in  $(V_{i_1}, \dots, V_{i_\ell})$  and are not bad. So  $|\mathcal{A}| \geq (1 - d^{1/3})m^\ell$ . Moreover, the number of edges  $e$  of  $G' - V_0$  with  $|e \cap V_{i_\alpha}| = 1$  for all  $\alpha \in \{1, \dots, \ell\}$  is at least

$$\sum_{A \in \mathcal{A}} \deg_{G' - V_0}(A) \geq |\mathcal{A}| (c - 3d^{1/3}) |G|^{k-\ell} \geq (c - 4d^{1/3}) |G|^{k-\ell} m^\ell. \quad (2.5.7)$$

Now suppose that the degree in  $R$  of  $(V_{i_1}, \dots, V_{i_\ell})$  is less than  $(c - 4d^{1/3})|R|^{k-\ell}$ . Then the number of  $(k - \ell)$ -tuples  $\{j_1, \dots, j_{k-\ell}\} \in \binom{[L]}{k-\ell}$  for which  $(V_{i_1}, \dots, V_{i_\ell}, V_{j_1}, \dots, V_{j_{k-\ell}})_{G'}$  is  $\varepsilon$ -regular with density greater than  $d$  is less than  $(c - 4d^{1/3})|R|^{k-\ell}$ . Note that at most  $m^k$  edges of  $G' - V_0$  lie in such a subhypergraph. So the number of edges  $e$  of  $G' - V_0$  with  $|e \cap V_{i_\alpha}| = 1$  for all  $\alpha \in \{1, \dots, \ell\}$  is less than

$$(c - 4d^{1/3})|R|^{k-\ell} m^k \leq (c - 4d^{1/3})|G|^{k-\ell} m^\ell,$$

contradicting (2.5.7). This completes the proof.  $\square$

The following lemma uses all of the previous results of this section to allow us to convert our fractional matchings into integer ones.

**Lemma 2.5.8.** *Let  $k \geq 2$  and  $1 \leq \ell \leq k - 1$  be integers, and let  $\varepsilon > 0$ . Suppose that for some  $b, c \in (0, 1)$  and some integer  $n_0$ , any  $k$ -uniform hypergraph on  $n \geq n_0$  vertices with at least  $(1 - \varepsilon)\binom{n}{\ell}$   $\ell$ -tuples of vertices of degree at least  $cn^{k-\ell}$  has a fractional matching of size  $(b + \varepsilon)n$ . Then there exists an integer  $n'_0$  such that any  $k$ -uniform hypergraph  $G$  on*

$n \geq n'_0$  vertices with  $\delta_\ell(G) \geq (c + \varepsilon)n^{k-\ell}$  has an (integer) matching of size at least  $bn$ .

*Proof.* Define  $n'_0 \in \mathbb{N}$  and new constants  $\varepsilon'$  and  $d$  such that  $0 < 1/n'_0 \ll \varepsilon' \ll d \ll \varepsilon, c, 1/k, 1/n_0$ . Let  $G$  be a  $k$ -uniform hypergraph on  $n \geq n'_0$  vertices, with  $\delta_\ell(G) \geq (c + \varepsilon)n^{k-1}$ . Let  $G'$  be the spanning subhypergraph of  $G$  obtained by applying the degree form of the Weak Hypergraph Regularity Lemma to  $G$  with parameters  $\varepsilon', d, n_0$ . Let  $R := R(\varepsilon', d, n_0)$  be the corresponding reduced hypergraph, and let  $L := |R|$ . By Lemma 2.5.6 at least  $(1 - \varepsilon)\binom{L}{\ell}$   $\ell$ -tuples of vertices of  $R$  have degree at least

$$(c + \varepsilon - 4d^{1/3})L^{k-\ell} \geq cL^{k-\ell}.$$

So by the assumption in the statement of the lemma,  $R$  has a fractional matching,  $F$  say, of size  $(b + \varepsilon)L$ .

For each  $e \in E(R)$ , let  $K_e := \lceil (1 - 2\varepsilon')F(e)m \rceil$ , where  $m$  is the size of each of the clusters of  $R$ . Now construct an integer matching,  $M$  say, in  $G$  by greedily adding to  $M$  edges of  $G'$  until, for each  $e = \{V_{j_1}, \dots, V_{j_k}\} \in E(R)$ ,  $M$  contains precisely  $K_e$  edges of  $(V_{j_1}, \dots, V_{j_k})_{G'}$ . Note that at each stage of this process the number of vertices in each  $V_i \in V(R)$  that would be covered by  $M$  is at most

$$\sum_{e: V_i \in e} K_e \leq \sum_{e: V_i \in e} ((1 - 2\varepsilon')F(e)m + 1) \leq (1 - 2\varepsilon')m + \binom{L-1}{k-1} \leq (1 - \varepsilon')m.$$

Note also that for every edge  $e = \{V_{j_1}, \dots, V_{j_k}\} \in E(R)$ , we have that  $(V_{j_1}, \dots, V_{j_k})_{G'}$  is  $\varepsilon'$ -regular with density  $d > \varepsilon'$ . So indeed, by the definition of  $\varepsilon'$ -regularity, it is possible to successively add edges to  $M$  in order to obtain a matching  $M$  as desired.

Note that the size of  $M$  is

$$\sum_{e \in E(R)} K_e \geq \sum_{e \in E(R)} (1 - 2\varepsilon') F(e) m = (1 - 2\varepsilon') m(b + \varepsilon) L \geq (1 - 2\varepsilon')(b + \varepsilon)(1 - \varepsilon') n \geq bn.$$

So indeed  $G$  has an (integer) matching of size at least  $bn$ .  $\square$

We are now in a position to prove Theorem 2.1.4.

**Proof of Theorem 2.1.4.** Let  $\varepsilon' > 0$  and let  $0 < \varepsilon'' \ll \varepsilon', \varepsilon, 1/k, 1/2(k - d) - a$ . Let  $n_0 \in \mathbb{N}$  be sufficiently large and suppose that  $n \geq n_0$ . Let  $k' := k - d$  and  $n' := n - d$ . Then

$$\begin{aligned} f_0^{(a+\varepsilon'')n}(k-d, n-d) &= f_0^{(a+\varepsilon'')(n'+d)}(k', n') \leq f_0^{(a+2\varepsilon'')n'}(k', n') \leq m_0^{(a+2\varepsilon'')n'}(k', n') \\ &\stackrel{(2.4.2)}{\leq} \left(1 - (1 - a - 2\varepsilon'')^{k'} + \frac{\varepsilon'}{4}\right) \binom{n'}{k'} \\ &\leq \left(1 - (1 - a)^{k-d} + \frac{\varepsilon'}{2}\right) \binom{n-d}{k-d}. \end{aligned}$$

So by Proposition 2.4.1, if  $H$  is a  $k$ -uniform hypergraph on  $n$  vertices such that for at least  $(1 - \varepsilon'') \binom{n}{d}$   $d$ -tuples of vertices  $L \in \binom{V(H)}{d}$  we have

$$\deg_H(L) \geq \frac{1 - (1 - a)^{k-d} + \varepsilon'/2}{(k-d)!} n^{k-d} \geq \left(1 - (1 - a)^{k-d} + \frac{\varepsilon'}{2}\right) \binom{n-d}{k-d},$$

then  $H$  has a fractional matching of size  $(a + \varepsilon'')n$ . So by Lemma 2.5.8, any  $k$ -uniform hypergraph  $G$  on  $n \geq n'_0$  vertices (where  $n'_0$  is sufficiently large) with

$$\delta_d(G) \geq (1 - (1 - a)^{k-d} + \varepsilon') \binom{n-d}{k-d} \geq \left(\frac{1 - (1 - a)^{k-d} + \varepsilon'/2}{(k-d)!} + \varepsilon''\right) n^{k-d}$$

has an (integer) matching of size at least  $an$ . This gives the upper bound in Theorem 2.1.4. The lower bound follows from (2.2.1).  $\square$

We can prove Theorem 2.1.2 in a similar way, but to do so we will need to use the absorbing technique as introduced by Rödl, Ruciński and Szemerédi [69]. More precisely, we use the existence of a small and powerful matching  $M_{abs}$  in  $G$  which, by ‘absorbing’ vertices, can transform any almost perfect matching into a perfect matching.  $M_{abs}$  has the property that whenever  $X$  is a sufficiently small set of vertices of  $G$  not covered by  $M_{abs}$  (and  $|X| \in k\mathbb{N}$ ) there exists a matching in  $G$  which covers precisely the vertices in  $X \cup V(M_{abs})$ . Since this part of the proof of Theorem 2.1.2 is very similar to the corresponding part of the proof of Theorem 1.1 in [5], we only sketch it.

**Proof of Theorem 2.1.2 (sketch).** Let  $\varepsilon > 0$  and suppose that  $G$  is a  $k$ -uniform hypergraph on  $n$  vertices with minimum  $d$ -degree at least

$$\left( \frac{k-d}{k} - \frac{k-d-1}{k^{k-1}} + \varepsilon \right) \binom{n-d}{k-d} > \left( \frac{1}{2} + \varepsilon \right) \binom{n-d}{k-d}. \quad (2.5.9)$$

(2.5.9) implies that we can use the Strong Absorbing Lemma from [39] to find an absorbing matching  $M_{abs}$  in  $G$ , and set  $G' := G \setminus V(M_{abs})$ . Using the degree condition, Theorem 2.1.9 gives us a perfect fractional matching in  $G'$  for sufficiently large  $n$ . Lemma 2.5.8(ii) then transforms this into an almost perfect integer matching  $M_{alm}$  in  $G'$ . We then extend  $M_{alm} \cup M_{abs}$  to a perfect matching of  $G$  by using the absorbing property of  $M_{abs}$ .  $\square$

## 2.6 Variants of Theorems 2.1.4 and 2.1.6

Using a method similar to that employed in proving Theorem 2.1.2 it is possible to prove a variant of Theorem 2.1.6 that verifies Conjecture 2.1.5 asymptotically for all  $k, s \in \mathbb{N}$  satisfying  $k \geq 4$  and  $s/n < a_k$ , where  $a_k$  is the unique solution in  $(0, 1/(k+1))$  to

$g_k(x) = 1$ , where

$$g_k(x) := \frac{1 - (1 - 2x)^{k-1}}{(1 - x)^{k-1}},$$

(see Theorem 2.6.1). For small values of  $k$  this allows us to verify Conjecture 2.1.5 asymptotically for some values of  $s$  not covered by Theorem 2.1.6. For example for  $k = 4$  this allows  $s$  to range up to  $0.567n/k$ . This approach also yields a slight improvement, for small values of  $k$ , to the range of matching sizes allowed in Theorem 2.1.4 (see Theorem 2.6.2). But for large  $k$  Theorem 2.1.6 gives the better bounds on the matching sizes allowed (as  $a_k$  is close to  $0.48/k$  in this case). The purpose of this section is to prove Theorems 2.6.1 and 2.6.2, below.

**Theorem 2.6.1.** *Let  $n, k \geq 4$  and  $0 \leq a < a_k$  be such that  $n, k, an \in \mathbb{N}$ . The minimum number of edges in a  $k$ -uniform hypergraph on  $n$  vertices which forces a matching of size  $an$  is*

$$(1 - (1 - a)^k + o(1)) \binom{n}{k}.$$

**Theorem 2.6.2.** *Let  $\varepsilon > 0$  and let  $n, k, d$  be integers with  $1 \leq d \leq k - 4$ , and let  $0 \leq a < \min\{a_{k-d}, (1 - \varepsilon)/k\}$  be such that  $an \in \mathbb{N}$ . Then*

$$m_d^{an}(k, n) = (1 - (1 - a)^{k-d} + o(1)) \binom{n-d}{k-d}.$$

We first show that the equation  $g_k(x) = 1$  does indeed have a unique solution in  $(0, 1/(k + 1))$ , as claimed.

**Proposition 2.6.3.** *Let  $k \in \mathbb{N}$  with  $k \geq 3$ . The equation  $g_k(x) = 1$  has a unique solution in  $(0, 1/(k + 1))$ .*

*Proof.* Note that  $g_k(x)$  is strictly increasing for  $x \in (0, 1/(k + 1))$  and that  $g_k(0) = 0$ .

Also note that

$$\begin{aligned} g_k(x) &= \frac{1 - (1 - 2x)^{k-1}}{(1 - x)^{k-1}} \frac{(1 + x)^{k-1}}{(1 + x)^{k-1}} = \frac{(1 + x)^{k-1} - (1 - x - 2x^2)^{k-1}}{(1 - x^2)^{k-1}} \\ &> \frac{(1 + x)^{k-1} - (1 - x)^{k-1}}{(1 - x^2)^{k-1}} > (1 + x)^{k-1} - (1 - x)^{k-1} > 2(k - 1)x, \end{aligned}$$

for  $x \in (0, 1/2)$ , so  $g_k(1/(k + 1)) > 2(k - 1)/(k + 1) \geq 1$  for  $k \geq 3$ . So indeed  $g_k(x) = 1$  has a unique solution in  $(0, 1/(k + 1))$ .  $\square$

Propositions 2.6.4 and 2.6.5 below will allow us to use induction in the subsequent arguments, by establishing the base case and the validity of the inductive step respectively. For a proof of Proposition 2.6.4 see Corollary 2.1 in [5].

**Proposition 2.6.4.** *For  $k = 4$ ,  $x \leq 1/5$ ,*

$$f_0^{xn}(k, n) = (1 - (1 - x)^k + o(1)) \binom{n}{k}.$$

**Proposition 2.6.5.** *For all  $k \in \mathbb{N}$  with  $k \geq 4$  we have  $a_k/(1 - a_k) < a_{k-1}$ .*

*Proof.* Recall from the proof of Proposition 2.6.3 that  $g_{k-1}(x)$  is strictly increasing for  $x \in (0, 1/k)$ . Since  $a_k/(1 - a_k) < 1/k$ , it suffices to show that  $g_{k-1}(a_k/(1 - a_k)) < g_{k-1}(a_{k-1})$ . As by definition  $g_{k-1}(a_{k-1}) = 1 = g_k(a_k)$ , it suffices to show  $g_{k-1}(a_k/(1 - a_k)) < g_k(a_k)$ . For clarity, we denote  $a_k$  simply by  $a$  for the remainder of this proof. So it suffices to show that

$$\frac{(1 - a)^{k-2} - (1 - 3a)^{k-2}}{(1 - 2a)^{k-2}} = \frac{1 - (1 - \frac{2a}{1-a})^{k-2}}{(1 - \frac{a}{1-a})^{k-2}} < \frac{1 - (1 - 2a)^{k-1}}{(1 - a)^{k-1}}. \quad (2.6.6)$$



This in turn is equivalent to  $s(a) + t(a) < 0$ , where

$$\begin{aligned} s(a) &:= (1-a) \left( (1-a)^2 \right)^{k-2} - (1-2a)^{k-2}, \\ t(a) &:= (1-2a) \left( (1-2a)^2 \right)^{k-2} - (1-a) \left( (1-a)(1-3a) \right)^{k-2}. \end{aligned}$$

Now,

$$\begin{aligned} s(a) &= (1-2a+a^2)^{k-2} - (1-2a)^{k-2} - a \left( (1-a)^2 \right)^{k-2} \\ &= \left( \sum_{i=1}^{k-2} \binom{k-2}{i} a^{2i} (1-2a)^{k-2-i} \right) - a \left( (1-a)^2 \right)^{k-2} \\ &\leq \left( \sum_{i=1}^{k-2} (k-2) \binom{k-3}{i-1} a^{2i} (1-2a)^{k-2-i} \right) - a \left( (1-a)^2 \right)^{k-2} \\ &= (k-2)a^2 \left( \sum_{i=0}^{k-3} \binom{k-3}{i} a^{2i} (1-2a)^{k-3-i} \right) - a \left( (1-a)^2 \right)^{k-2} \\ &= (k-2)a^2 (1-2a+a^2)^{k-3} - a \left( (1-a)^2 \right)^{k-2} \\ &= a \left( (1-a)^2 \right)^{k-3} (a(k-2) - (1-a)^2). \end{aligned}$$

Note that  $a((1-a)^2)^{k-3} > 0$ , that  $q_1(a) := a(k-2) - (1-a)^2$  is increasing in  $a$  for  $a \in (0, 1/(k+1))$  and that  $q_1(1/(k+1)) < 0$ . Hence,  $s(a) < 0$ .

Similarly,

$$\begin{aligned} t(a) &= (1-a) (1-4a+3a^2+a^2)^{k-2} - (1-a) (1-4a+3a^2)^{k-2} - a \left( (1-2a)^2 \right)^{k-2} \\ &\leq (k-2)(1-a)a^2 \left( (1-2a)^2 \right)^{k-3} - a \left( (1-2a)^2 \right)^{k-2} \\ &= a \left( (1-2a)^2 \right)^{k-3} (a(k-2)(1-a) - (1-2a)^2). \end{aligned}$$

Note that  $a((1-2a)^2)^{k-3} > 0$ , that  $q_2(a) := a(k-2)(1-a) - (1-2a)^2$  is increasing in  $a$  for  $a \in (0, 1/(k+1))$  and that  $q_2(1/(k+1)) < 0$ . Hence,  $t(a) < 0$ .

So  $s(a) + t(a) < 0$  and so (2.6.6) holds. □

Theorem 2.6.7, below, is an analogous result to Theorem 2.1.8 for matchings of size at most  $a_k n$ , and is proved using a similar inductive argument.

**Theorem 2.6.7.** *Let  $n, k \geq 4$  be integers and let  $0 \leq a < a_k$ . Then*

$$f_0^{an}(k, n) = (1 - (1 - a)^k + o(1)) \binom{n}{k}.$$

*Proof.* The proof will proceed by induction on  $k$ . The base step,  $k = 4$ , follows by Theorem 2.6.4, (as  $a_4 < 1/5$  by definition).

Now consider some  $k > 4$  and suppose that the theorem holds for all smaller values of  $k$ . Let  $\varepsilon > 0$  and let  $n_0 \in \mathbb{N}$  be sufficiently large compared to  $1/\varepsilon$  and  $k$ . Fix any  $a$  with  $0 < a < a_k$ . For convenience let us define

$$\xi(a) := \left(1 - \left(1 - \frac{a}{1 - a}\right)^{k-1}\right) (1 - a)^{k-1} + 1 - (1 - a)^{k-1} = 1 - (1 - 2a)^{k-1}.$$

Consider any  $k$ -uniform hypergraph  $G = (V, E)$  on  $n \geq n_0$  vertices, and suppose that the largest fractional matching of  $G$  is of size less than  $an$ . Then by Proposition 2.2.2 there exists a fractional vertex cover,  $w$  say, of  $G$  with size less than  $an$ . By Proposition 2.6.5,

$$\frac{a}{1 - a} < \frac{a_k}{1 - a_k} < a_{k-1}. \tag{2.6.8}$$

Let  $c := 1 - (1 - a/(1 - a))^{k-1} + \varepsilon/2$ . Then by induction and (2.6.8),

$$f_0^{an'/(1-a)}(k - 1, n') \leq c \binom{n'}{k - 1},$$

for all sufficiently large  $n'$ . So, as  $n_0$  is sufficiently large, Lemma 2.3.1 implies that there are less than  $an$  vertices of  $G$  with degree at least  $(\xi(a) + \varepsilon)\binom{n-1}{k-1}$ .

Let  $S$  be the set of  $\lceil an \rceil - 1$  vertices of  $G$  with highest degree. Note then that  $d_G(v) < (\xi(a) + \varepsilon)\binom{n-1}{k-1}$  for all  $v \in V \setminus S$ . Given  $X \subseteq V \setminus S$ , for all  $s \in S$  let  $t_X(s)$  denote the number of  $k$ -tuples of vertices of  $G$  consisting of  $s$  and  $k-1$  vertices from  $V \setminus S$  such that at least one of these  $k-1$  vertices lies in  $X$ .

We claim that  $X \subseteq V \setminus S$  can be chosen such that  $t_X(s) \geq \xi(a)\binom{n-1}{k-1}$  for all  $s \in S$ . Indeed, if we take  $X$  to be  $V \setminus S$  then for each  $s \in S$  we have that

$$t_X(s) = \binom{|X|}{k-1} = ((1-a)^{k-1} + o(1)) \binom{n-1}{k-1} \geq \xi(a) \binom{n-1}{k-1}.$$

To see that the final inequality holds for sufficiently large  $n_0$ , note that

$$\frac{\xi(a)}{(1-a)^{k-1}} = g_k(a) < g_k(a_k) = 1, \quad (2.6.9)$$

by definition of  $a$  and  $a_k$ , and the fact that  $g_k(x)$  is strictly increasing for  $x \in (0, 1/(k+1))$ .

Choose  $X \subseteq V \setminus S$  of minimal size with the property that  $t_X(s) \geq \xi(a)\binom{n-1}{k-1}$  for all  $s \in S$ . Note that  $t_X(s) = t_X(s')$  for all  $s, s' \in S$  and  $t_X(s) \leq (\xi(a) + \varepsilon)\binom{n-1}{k-1}$ . (The latter holds since we may assume that  $n_0$  is sufficiently large.) Also let

$$E' := \{e \in E : e \cap S = \emptyset\} \cup \{e \in E : |e \cap S| = 1, |e \cap X| \geq 1\}.$$

So

$$|E \setminus E'| \leq \binom{n}{k} - ((1-a)^k + o(1)) \binom{n}{k} - \xi(a) \binom{n-1}{k-1} (\lceil an \rceil - 1). \quad (2.6.10)$$

Now, note that by Proposition 2.2.3,

$$e(G) \leq \sum_{e \in E} \sum_{v \in e \setminus S} w(v) + \sum_{e \in E'} \sum_{v \in e \cap S} w(v) + |E \setminus E'|.$$

Together with the facts that  $d_G(v) < (\xi(a) + \varepsilon) \binom{n-1}{k-1}$  for all  $v \in V \setminus S$  and that the number of edges in  $E'$  incident to  $s$  is at most  $t_X(s) \leq (\xi(a) + \varepsilon) \binom{n-1}{k-1}$  for all  $s \in S$ , this implies that

$$e(G) \leq \sum_{v \in V} (\xi(a) + \varepsilon) \binom{n-1}{k-1} w(v) + |E \setminus E'|.$$

Now, recalling that the size of  $w$  is less than  $an$  and that  $a < a_k < 1/(k+1)$  gives

$$e(G) < (\xi(a) + \varepsilon) \binom{n-1}{k-1} an + |E \setminus E'| \stackrel{(2.6.10)}{\leq} (1 - (1-a)^k + \varepsilon) \binom{n}{k}.$$

By definition, this shows that  $f_0^{an}(k, n) \leq (1 - (1-a)^k + o(1)) \binom{n}{k}$ . This, along with the lower bound (2.2.1), completes the inductive step and hence the proof.  $\square$

Note that the main constraint which restricts the range of  $a$  here is given by (2.6.9).

The following lemma tells us that the reduced hypergraph defined in Section 2.5 (almost) inherits the edge density properties of the original hypergraph. It can be seen as an analogue of Lemma 2.5.6, with the notion of edge density replacing that of minimum degree.

**Lemma 2.6.11.** *Suppose  $c > 0$ ,  $k \geq 2$ ,  $L_0 \geq 1$  and  $0 < \varepsilon \leq d \leq c/4$ . Let  $G$  be a  $k$ -uniform hypergraph with  $e(G) \geq c|G|^k$ . Let  $R = R(\varepsilon, d, L_0)$  be the reduced hypergraph of  $G$ . Then  $e(R) \geq (c - 4d)|R|^k$ .*

*Proof.* Let  $G'$  be the spanning subhypergraph of  $G$  obtained by applying the degree form of the Weak Hypergraph Regularity Lemma to  $G$  with parameters  $\varepsilon, d, L_0$ ; let  $V_1, \dots, V_L$  denote the vertices of  $R$ , and let  $m$  denote the size of these clusters.

To prove (i), suppose that  $e(R) < (c - 4d)|R|^k$ . So there are less than  $(c - 4d)|R|^k$   $k$ -tuples  $\{i_1, \dots, i_k\} \in \binom{[L]}{k}$  such that  $(V_{i_1}, \dots, V_{i_k})_{G'}$  has non-zero density. Note that at most  $m^k$  edges lie in such a  $k$ -tuple. So

$$e(G' - V_0) < (c - 4d)|R|^k m^k \leq (c - 4d)|G|^k.$$

However, given any vertex  $x \in V(G')$  we know that  $d_{G'}(x) > d_G(x) - (d + \varepsilon)|G|^{k-1}$ . Note that since  $|V_0| \leq \varepsilon|G|$ , we have that the number of edges incident to  $x$  that contain a vertex in  $V_0$  is at most  $\varepsilon|G|\binom{|G|}{k-2} \leq \varepsilon|G|^{k-1}$ . Hence for all  $v \in V(G' - V_0)$ , we have that  $d_{G'-V_0}(v) > d_G(v) - (d + 2\varepsilon)|G|^{k-1}$ . So

$$e(G' - V_0) = \frac{\sum_{v \in V(G'-V_0)} d_{G'-V_0}(v)}{k} > \frac{\left(\sum_{v \in V(G)} d_G(v)\right) - \left(\sum_{v \in V_0} d_G(v)\right) - (d + 2\varepsilon)|G|^k}{k}.$$

Since  $|V_0| \leq \varepsilon|G|$ , this implies that

$$e(G' - V_0) > e(G) - \frac{(d + 3\varepsilon)|G|^k}{k} > (c - 4d)|G|^k,$$

a contradiction. This proves (i). □

The next lemma allows us to convert our fractional matchings into integer ones, similarly to Lemma 2.5.8, but again replacing the notion of minimum degree by that of edge density. The proof is almost identical to that of Lemma 2.5.8, just using Lemma 2.6.11 instead of Lemma 2.5.6, and so is omitted here.

**Lemma 2.6.12.** *Let  $k \in \mathbb{N}$  with  $k \geq 2$ , and let  $\varepsilon > 0$ . Suppose that for some  $b, c \in (0, 1)$  and some integer  $n_0$ , any  $k$ -uniform hypergraph  $G^*$  on  $n \geq n_0$  vertices with  $e(G^*) \geq cn^k$  has a fractional matching of size  $(b + \varepsilon)n$ . Then there exists an integer  $n'_0$  such that any  $k$ -uniform hypergraph  $G$  on  $n \geq n'_0$  vertices with  $e(G) \geq (c + \varepsilon)n^k$  has an (integer)*

*matching of size at least  $bn$ .*

We are now in a position to prove Theorems 2.6.1 and 2.6.2. The proof of Theorem 2.6.2 is almost identical to that of Theorem 2.1.4, just using Theorem 2.6.7 instead of Theorem 2.1.6, and so the details are omitted here. Theorem 2.6.1 follows immediately from Lemma 2.6.12, Theorem 2.6.7 and the lower bound (2.2.1).

## CHAPTER 3

# PROOF OF A TOURNAMENT PARTITION CONJECTURE AND AN APPLICATION TO 1-FACTORS WITH PRESCRIBED CYCLE LENGTHS

### 3.1 Chapter introduction

#### 3.1.1 Partitioning tournaments into highly connected subtournaments

As discussed in Chapter 1, there is a rich literature of results and questions relating to partitions of (di)graphs into subgraphs which inherit some properties of the original (di)graph. Hajnal [37] and Thomassen [77] proved that for every  $k$  there exists an integer

$f(k)$  such that every  $f(k)$ -connected graph has a vertex partition into sets  $S$  and  $T$  so that both  $S$  and  $T$  induce  $k$ -connected graphs. In this chapter we investigate a corresponding question for tournaments.

A tournament is an orientation of a complete graph. A tournament is *strongly connected* if for every pair of vertices  $u, v$  there exists a directed path from  $u$  to  $v$  and a directed path from  $v$  to  $u$ . For any integer  $k$  we call a tournament  $T$  *strongly  $k$ -connected* if  $|V(T)| > k$  and the removal of any set of fewer than  $k$  vertices results in a strongly connected tournament. We denote the subtournament induced on a tournament  $T$  by a set  $U \subseteq V(T)$  by  $T[U]$ .

The following problem was posed by Thomassen (see [65]).

**Problem 3.1.1.** *Let  $k_1, \dots, k_t$  be positive integers. Does there exist an integer  $f(k_1, \dots, k_t)$  such that every strongly  $f(k_1, \dots, k_t)$ -connected tournament  $T$  admits a partition of its vertex set into vertex classes  $V_1, \dots, V_t$  such that for all  $i \in \{1, \dots, t\}$  the subtournament  $T[V_i]$  is strongly  $k_i$ -connected?*

If  $k_i = 1$  for all  $i \in \{2, \dots, t\}$  then  $f(k_1, \dots, k_t)$  exists and is at most  $k_1 + 3t - 3$ . This follows by an easy induction on  $t$ , taking  $V_t$  to be a set inducing a directed 3-cycle. Chen, Gould and Li [20] showed that every strongly  $t$ -connected tournament with at least  $8t$  vertices admits a partition into  $t$  strongly connected subtournaments. This gives the best possible connectivity bound in the case  $k_1 = \dots = k_t = 1$  and  $|V(T)| \geq 8t$ . Until now even the existence of  $f(2, 2)$  was open. The main result in this chapter answers all cases of the above problem of Thomassen in the affirmative.

**Theorem 3.1.2.** *Let  $T$  be a tournament on  $n$  vertices and let  $k, t \in \mathbb{N}$  with  $t \geq 2$ . If  $T$  is strongly  $10^7 k^6 t^3 \log(kt^2)$ -connected then there exists a partition of  $V(T)$  into  $t$  vertex*



classes  $V_1, \dots, V_t$  such that for all  $i \in \{1, \dots, t\}$  the subtournament  $T[V_i]$  is strongly  $k$ -connected.

The above bound is unlikely to be best possible. It would be interesting to establish the correct order of magnitude of  $f(k_1, \dots, k_t)$  for all fixed  $k_i$  and  $t$ . In fact, we believe a linear bound may suffice.

**Conjecture 3.1.3.** *There exists a constant  $c$  such that the following holds. Let  $T$  be a tournament on  $n$  vertices and let  $k, t \in \mathbb{N}$ . If  $T$  is strongly  $ckt$ -connected then there exists a partition of  $V(T)$  into  $t$  vertex classes  $V_1, \dots, V_t$  such that for all  $i \in \{1, \dots, t\}$  the subtournament  $T[V_i]$  is strongly  $k$ -connected.*

It would also be interesting to know whether Theorem 3.1.2 can be generalised to digraphs.

**Question 3.1.4.** *Does there exist, for all  $k, t \in \mathbb{N}$ , a function  $\widehat{f}(k, t)$  such that for every strongly  $\widehat{f}(k, t)$ -connected digraph  $D$  there exists a partition of  $V(D)$  into  $t$  vertex classes  $V_1, \dots, V_t$  such that for all  $i \in \{1, \dots, t\}$  the subdigraph  $D[V_i]$  is strongly  $k$ -connected?*

Recently Kim, Kühn and Osthus [45] proved a stronger version of the case  $t = 2$  of Theorem 3.1.2, which ensures that the bipartite digraph  $T[V_1, V_2]$  is also strongly  $k$ -connected.

Instead of proving Theorem 3.1.2 directly, we first prove the following somewhat stronger result. It establishes the existence of small but powerful ‘linkage structures’ in tournaments, and Theorem 3.1.2 follows from it as an immediate corollary. These linkage structures are partly based on ideas of Kühn, Lapinskas, Osthus and Patel [48], who proved a conjecture of Thomassen by showing that for every  $k$  there exists an integer  $\tilde{f}(k)$  such that every strongly  $\tilde{f}(k)$ -connected tournament contains  $k$  edge-disjoint Hamilton cycles.

**Theorem 3.1.5.** *Let  $T$  be a tournament on  $n$  vertices, let  $k, m, t \in \mathbb{N}$  with  $m \geq t \geq 2$ . If  $T$  is strongly  $10^7 k^6 t^2 m \log(ktm)$ -connected then  $V(T)$  contains  $t$  disjoint vertex sets  $V_1, \dots, V_t$  such that for every  $j \in \{1, \dots, t\}$  the following hold:*

- (i)  $|V_j| \leq n/m$ ,
- (ii) *for any set  $R \subseteq V(T) \setminus \bigcup_{i=1}^t V_i$  such that  $|V_j \cup R| > k$  the subtournament  $T[V_j \cup R]$  is strongly  $k$ -connected.*

Recently, Pokrovskiy [59] has also used similar linkage structure ideas to prove that every strongly  $452k$ -connected tournament is  $k$  linked. This is an improvement on a result of Kühn, Lapinskas, Osthus and Patel [48] that we use here (Theorem 3.2.3), and would in fact allow us to slightly improve the connectivity bounds in Theorems 3.1.2, 3.1.5, and 3.1.7. However, we present these Theorems here in their original form to remain consistent with the published version of this chapter.

### 3.1.2 Partitioning tournaments into vertex-disjoint cycles

Theorem 3.1.5 also has an application to another problem on tournaments, this time concerning partitioning the vertices of a tournament into vertex-disjoint cycles of prescribed lengths.

Reid [64] proved that any strongly 2-connected tournament on  $n \geq 6$  vertices admits a partition of its vertices into two vertex-disjoint cycles (unless the tournament is isomorphic to the tournament on 7 vertices which contains no transitive tournament on 4 vertices). Chen, Gould and Li [20] showed that every strongly  $t$ -connected tournament with at least  $8t$  vertices admits a partition into  $t$  vertex-disjoint cycles. This answered a question of

Bollobás (see [64]), namely what is the least integer  $g(t)$  such that all but a finite number of strongly  $g(t)$ -connected tournaments admit a partition into  $t$  vertex-disjoint cycles? Song proved the following strengthening of Reid's result.

**Theorem 3.1.6.** [73] *Let  $T$  be a tournament on  $n \geq 6$  vertices and let  $3 \leq L \leq n - 3$ . If  $T$  is strongly 2-connected then  $T$  contains two vertex-disjoint cycles of lengths  $L$  and  $n - L$  (unless  $T$  is isomorphic to the tournament on 7 vertices which contains no transitive tournament on 4 vertices).*

Song [73] also posed a question that generalises the question of Bollobás. Namely, for any integer  $t$ , what is the least integer  $h(t)$  such that all but a finite number of strongly  $h(t)$ -connected tournaments admit a partition into  $t$  vertex-disjoint cycles of prescribed lengths? Until now, for  $t \geq 3$ , even the existence of  $h(t)$  remained open. The following consequence of Theorem 3.1.5 settles this question in the affirmative.

**Theorem 3.1.7.** *Let  $T$  be a tournament on  $n$  vertices, let  $t \in \mathbb{N}$  with  $t \geq 2$  and let  $L_1, \dots, L_t \in \mathbb{N}$  with  $L_1, \dots, L_t \geq 3$  and  $\sum_{j=1}^t L_j = n$ . If  $T$  is strongly  $10^{10}t^4 \log t$ -connected then  $T$  contains  $t$  vertex-disjoint cycles of lengths  $L_1, \dots, L_t$ .*

Camion's theorem (see [19]) states that every strongly connected tournament contains a Hamilton cycle. So certainly  $g(1) = h(1) = 1$ . Note that Song [73] showed that  $g(2) = h(2) = 2$ . Clearly  $g(k) \leq h(k)$  for all  $k$ . Song [73] conjectured that  $g(k) = h(k)$  for all  $k$ . Showing that  $h(k)$  is linear would already be a very interesting step towards this.

Theorem 3.1.7 has a similar flavour to the El-Zahar conjecture. This determines the minimum degree which guarantees a partition of a graph into vertex-disjoint cycles of prescribed lengths and was proved for all large  $n$  by Abbasi [1]. A related result to Theorem 3.1.7 for oriented graphs (where the assumption of connectivity is replaced by that of high minimum semidegree) was proved by Keevash and Sudakov [42].

The rest of the chapter is organised as follows. In Section 3.2 we lay out some notation, set out some useful tools, and prove some preliminary results. Section 3.3 is the heart of the chapter in which we prove Theorem 3.1.5. In Section 3.4 we deduce Theorem 3.1.7.

## 3.2 Notation, tools and preliminary results

We write  $|T|$  for the number of vertices in a tournament  $T$ . We denote the in-degree of a vertex  $v$  in a tournament  $T$  by  $d_T^-(v)$ , and we denote the out-degree of  $v$  in  $T$  by  $d_T^+(v)$ . We say that a set  $A \subseteq V(T)$  *in-dominates* a set  $B \subseteq V(T)$  if for every vertex  $b \in B$  there exists a vertex  $a \in A$  such that there is an edge in  $T$  directed from  $b$  to  $a$ . Similarly, we say that a set  $A \subseteq V(T)$  *out-dominates* a set  $B \subseteq V(T)$  if for every vertex  $b \in B$  there exists a vertex  $a \in A$  such that there is an edge in  $T$  directed from  $a$  to  $b$ . We denote the *minimum semidegree* of  $T$  (that is, the minimum of the minimum in-degree of  $T$  and the minimum out-degree of  $T$ ) by  $\delta^0(T)$ . We say that a tournament  $T$  is *transitive* if we may enumerate its vertices  $v_1, \dots, v_m$  such that there is an edge in  $T$  directed from  $v_i$  to  $v_j$  if and only if  $i < j$ . In this case we call  $v_1$  the *source* of  $T$  and  $v_m$  the *sink* of  $T$ . The *length* of a path is the number of edges in the path. If  $P = x_1 \dots x_\ell$  is a path directed from  $x_1$  to  $x_\ell$  then we denote the set  $\{x_1, \dots, x_\ell\} \setminus \{x_1, x_\ell\}$  of interior vertices of  $P$  by  $\text{Int}(P)$ , and if  $1 \leq i < j \leq \ell$  we say that  $x_i$  is an *ancestor* of  $x_j$  in  $P$  and that  $x_j$  is a *descendant* of  $x_i$  in  $P$ . We say that an ordered pair of vertices  $(x, y)$  is *k-connected* in a tournament  $T$  if the removal of any set  $S \subseteq V(T) \setminus \{x, y\}$  of fewer than  $k$  vertices from  $T$  results in a tournament containing a directed path from  $x$  to  $y$ . A tournament  $T$  is called *k-linked* if  $|T| \geq 2k$  and whenever  $x_1, \dots, x_k, y_1, \dots, y_k$  are  $2k$  distinct vertices in  $V(G)$  there exist vertex-disjoint paths  $P_1, \dots, P_k$  such that  $P_i$  is a directed path from  $x_i$  to  $y_i$  for each  $i \in \{1, \dots, k\}$ . For clarity we may sometimes refer to a strongly connected tournament as a strongly 1-connected tournament. Throughout the chapter we write  $\log x$  to mean

$\log_2 x$ .

We now collect some preliminary results that will prove useful to us. The following proposition follows straightforwardly from the definition of linkedness.

**Proposition 3.2.1.** *Let  $k \in \mathbb{N}$ . Then a tournament  $T$  is  $k$ -linked if and only if  $|T| \geq 2k$  and whenever  $(x_1, y_1), \dots, (x_k, y_k)$  are ordered pairs of (not necessarily distinct) vertices of  $T$ , there exist distinct internally vertex-disjoint paths  $P_1, \dots, P_k$  such that for all  $i \in \{1, \dots, k\}$  we have that  $P_i$  is a directed path from  $x_i$  to  $y_i$  and that  $\{x_1, \dots, x_k, y_1, \dots, y_k\} \cap V(P_i) = \{x_i, y_i\}$ .*

**Proposition 3.2.2.** *Let  $k, s \in \mathbb{N}$  and let  $T$  be a  $ks$ -linked tournament. Let  $(x_1, y_1), \dots, (x_k, y_k)$  be ordered pairs of (not necessarily distinct) vertices of  $T$ . Then there exist distinct internally vertex-disjoint paths  $P_1, \dots, P_k$  such that for all  $i \in \{1, \dots, k\}$  we have that  $P_i$  is a directed path from  $x_i$  to  $y_i$  with  $\{x_1, \dots, x_k, y_1, \dots, y_k\} \cap V(P_i) = \{x_i, y_i\}$  and such that  $|\text{Int}(P_1) \cup \dots \cup \text{Int}(P_k)| \leq |T|/s$ .*

*Proof.* By Proposition 3.2.1  $T$  contains  $ks$  distinct internally vertex-disjoint paths  $P_1^1, \dots, P_k^s$  such that for all  $i \in \{1, \dots, k\}$  and  $j \in \{1, \dots, s\}$  we have that  $P_i^j$  is a directed path from  $x_i$  to  $y_i$  and that  $\{x_1, \dots, x_k, y_1, \dots, y_k\} \cap V(P_i^j) = \{x_i, y_i\}$ . The disjointness of the paths implies that there is a  $j \in \{1, \dots, s\}$  with  $|\text{Int}(P_1^j) \cup \dots \cup \text{Int}(P_k^j)| \leq |T|/s$ . So the result follows by setting  $P_i := P_i^j$  for all  $i \in \{1, \dots, k\}$ .  $\square$

We will also use the following theorem from [48] in proving Theorem 3.1.5.

**Theorem 3.2.3.** [48] *For all  $k \in \mathbb{N}$  with  $k \geq 2$  every strongly  $10^4 k \log k$ -connected tournament is  $k$ -linked.*

The following lemma, which we will also use in proving Theorem 3.1.5, is very similar to Lemma 8.3 in [48]. The proof proceeds by greedily choosing vertices  $v_1 = v, v_2, \dots, v_i$

such that the size of their common in-neighbourhood is minimised at each step. We omit the proof since it is almost identical to the one in [48].

**Lemma 3.2.4.** *Let  $T$  be a tournament, let  $v \in V(T)$  and suppose  $c \in \mathbb{N}$ . Then there exist disjoint sets  $A, E \subseteq V(T)$  such that the following properties hold:*

- (i)  $1 \leq |A| \leq c$  and  $T[A]$  is a transitive tournament with sink  $v$ ,
- (ii) either  $E = \emptyset$  or  $E$  is the common in-neighbourhood of all vertices in  $A$ ,
- (iii)  $A$  out-dominates  $V(T) \setminus (A \cup E)$ ,
- (iv)  $|E| \leq (1/2)^{c-1} d_T^-(v)$ .

The next lemma follows immediately from Lemma 3.2.4 by reversing the orientations of all edges.

**Lemma 3.2.5.** *Let  $T$  be a tournament, let  $v \in V(T)$  and suppose  $c \in \mathbb{N}$ . Then there exist disjoint sets  $B, E \subseteq V(T)$  such that the following properties hold:*

- (i)  $1 \leq |B| \leq c$  and  $T[B]$  is a transitive tournament with source  $v$ ,
- (ii) either  $E = \emptyset$  or  $E$  is the common out-neighbourhood of all vertices in  $B$ ,
- (iii)  $B$  in-dominates  $V(T) \setminus (B \cup E)$ ,
- (iv)  $|E| \leq (1/2)^{c-1} d_T^+(v)$ .

The following well-known observation will be useful in proving the subsequent technical lemma, which is essential to the proof of Theorem 3.1.5.

**Proposition 3.2.6.** *Let  $k \in \mathbb{N}$  and let  $T$  be a tournament. Then  $T$  contains less than  $2k$  vertices of out-degree less than  $k$ , and  $T$  contains less than  $2k$  vertices of in-degree less than  $k$ .*

We call a non-empty tournament  $Q$  a *backwards-transitive path* if we may enumerate the vertices of  $Q$  as  $q_1, \dots, q_{|Q|}$  such that there is an edge in  $Q$  from  $q_i$  to  $q_j$  if and only if either  $j = i + 1$  or  $i \geq j + 2$ . The following lemma shows that if a tournament  $T$  can be split into vertex-disjoint backwards transitive paths then there exist small (not necessarily disjoint) sets  $U$  and  $W$  which are ‘quickly reachable in a robust way’.

**Lemma 3.2.7.** *Let  $k, \ell \in \mathbb{N}$  and let  $T$  be a tournament on vertex set  $V = Q_1 \dot{\cup} \dots \dot{\cup} Q_\ell$ , with  $|Q_j| \geq k + 1$  for all  $j \in \{1, \dots, \ell\}$ . Suppose that, for each  $j \in \{1, \dots, \ell\}$ ,  $T[Q_j]$  is a backwards-transitive path. Then there exist sets  $U, W, U', W'$  satisfying the following properties:*

- $U \subseteq U' \subseteq V(T)$  and  $W \subseteq W' \subseteq V(T)$ ,
- $|U|, |W| \leq 2k(k + 1)$  and  $|U'|, |W'| = \ell(k + 1)$ ,
- for any set  $S \subseteq V(T)$  of size at most  $k - 1$ , and for every vertex  $v$  in  $V(T) \setminus S$ , there exists a directed path (possibly of length 0) in  $T[(U' \cup \{v\}) \setminus S]$  from  $v$  to a vertex in  $U$  and a directed path in  $T[(W' \cup \{v\}) \setminus S]$  from a vertex in  $W$  to  $v$ .

*Proof.* We prove only the existence of  $U, U'$ ; the existence of  $W, W'$  follows by a symmetric argument. Let the backwards-transitive paths  $T[Q_j]$  have vertices enumerated  $q_j^1, \dots, q_j^{|Q_j|}$  such that there is an edge in  $T[Q_j]$  from  $q_j^a$  to  $q_j^b$  if and only if either  $b = a + 1$  or  $a \geq b + 2$ . For  $i \in \{1, \dots, k + 1\}$  let  $T_i := T[\{q_1^i, \dots, q_\ell^i\}]$ . Thus  $|T_i| = \ell$ . Let  $U_i \subseteq V(T_i)$  be a set of  $\min\{2k, \ell\}$  vertices of lowest out-degree in  $T_i$ , let  $U' := V(T_1) \cup \dots \cup V(T_{k+1})$ , and let  $U := U_1 \cup \dots \cup U_{k+1}$ . Then clearly  $|U| \leq 2k(k + 1)$  and  $|U'| = \ell(k + 1)$ . Now suppose

$S \subseteq V(T)$  is of size at most  $k - 1$  and  $v \in V(T) \setminus S$ . We need to show that there exists a directed path (possibly of length 0) in  $T[(U' \cup \{v\}) \setminus S]$  from  $v$  to a vertex in  $U$ . We consider four cases:

- (i) If  $v \in U$  then we are clearly done.
- (ii) If  $v \in V(T_i) \setminus U$  for some  $i \in \{1, \dots, k+1\}$  and  $V(T_i) \cap S = \emptyset$ , then let  $u \in U \cap V(T_i) = U_i$ . Since the vertices of each  $U_i$  were picked to have minimal out-degree in  $T_i$ , we have that  $d_{T_i}^+(u) \leq d_{T_i}^+(v)$ , so there is an edge in  $T$  from either  $v$  or one of its out-neighbours in  $T_i$  to  $u$ . So there is a directed path in  $T_i$  of length at most two from  $v$  to  $u$  and we are done.
- (iii) If  $v \in V(T_i) \setminus U$  for some  $i \in \{1, \dots, k+1\}$  and  $V(T_i) \cap S \neq \emptyset$ , then first note that since  $v \in V(T_i) \setminus U$ , it must be that  $\ell = |T_i| > 2k$ . Note then that by Proposition 3.2.6 and our choice of  $U$  we have that  $d_{T_i}^+(v) \geq k$ . Hence, since  $|S| \leq k - 1$ , there is at least one  $j \in \{1, \dots, \ell\}$  such that  $q_j^i$  is an out-neighbour of  $v$  and such that  $Q_j \cap S = \emptyset$ . Also since  $|S| \leq k - 1$ , there is some  $i' \in \{1, \dots, k+1\}$  such that  $V(T_{i'}) \cap S = \emptyset$ . Since  $T[Q_j]$  is a backwards-transitive path, there is a directed path in  $T[Q_j \cap U']$  from  $q_j^i$  to  $q_j^{i'}$ , and by (i), (ii) there is a directed path (possibly of length 0) in  $T_{i'}$  from  $q_j^{i'}$  to a vertex in  $U$ . So piecing these paths together gives us a directed path  $P$  in  $T[U' \setminus S]$  from  $v$  to  $U$  as required. (Indeed, note that  $P$  avoids  $S$  since both  $Q_j$  and  $T_{i'}$  avoid  $S$ .)
- (iv) If  $v \in V(T) \setminus U'$  then note that  $v = q_j^i$  for some  $j \in \{1, \dots, \ell\}$  and some  $i > k+1$ . Now since  $T[Q_j]$  is a backwards-transitive path, there are edges in  $T$  directed from  $v$  to each of the vertices  $q_j^1, \dots, q_j^k$ . Since  $|S| \leq k - 1$ , there is some  $i \in \{1, \dots, k\}$  such that  $q_j^i \notin S$ . By (i)–(iii) there is a directed path in  $T[U' \setminus S]$  from  $q_j^i$  to a vertex in  $U$ . So this path together with the edge directed from  $v$  to  $q_j^i$  is the directed path required.



This covers all cases and we are done. □

### 3.3 Proof of Theorem 3.1.5

The purpose of this section is to prove Theorem 3.1.5. Very briefly, the proof strategy is as follows: suppose for simplicity that  $k = t = m = 2$ . We aim to construct small disjoint out-dominating sets  $A_1, \dots, A_4$  (i.e. for every vertex  $v \in V(T)$  there is an edge from each  $A_i$  to  $v$ ) so that each  $A_i$  induces a transitive subtournament of  $T$ . Similarly, we aim to construct small disjoint in-dominating sets  $B_i$ . Then for each  $i$  we find a short path  $P_i$  joining the sink of  $B_i$  to the source of  $A_i$ , using the assumption of high connectivity. Let  $V_1 := D_1 \cup D_2$  and  $V_2 := D_3 \cup D_4$ , where  $D_i := A_i \cup V(P_i) \cup B_i$  for  $i = 1, \dots, 4$ .

Now it is easy to check that Theorem 3.1.5(ii) holds: consider  $R$  as in (ii) and delete an arbitrary vertex  $s$  from  $V_1 \cup R$  to obtain a set  $W$ . To prove (ii) we have to show that for any  $x, y \in W$  there is a path from  $x$  to  $y$  in  $T[W]$ . To see this note that, without loss of generality,  $W$  still contains all of  $D_1$  (otherwise we consider  $D_2$  instead). Since  $B_1$  is in-dominating, there is an edge from  $x$  to some  $b \in B_1$ . Similarly, there is an edge from some  $a \in A_1$  to  $y$ . Since  $A_1$  and  $B_1$  induce transitive tournaments, we can now find a path from  $b$  to  $a$  in  $T[D_1]$  by utilizing  $P_1$  (see Claim 1).

The main problem with this approach is that one cannot quite achieve the above domination property: for every  $A_i$  there is a small exceptional set which is not out-dominated by  $A_i$  (and similarly for  $B_i$ ). We overcome this obstacle by using the notion of ‘safe’ vertices introduced before Claim 2. With this notion, we can still find a short path from an exceptional vertex  $x$  to  $B_i$  (rather than a single edge).

**Proof of Theorem 3.1.5.** Let  $x_1, \dots, x_{kt}$  be  $kt$  vertices of lowest in-degree in  $T$ . Let

$y_1, \dots, y_{kt}$  be  $kt$  vertices in  $V(T) \setminus \{x_1, \dots, x_{kt}\}$  whose out-degree in  $T$  is as small as possible. Define

$$\widehat{\delta}^-(T) := \min_{v \in V(T) \setminus \{x_1, \dots, x_{kt}\}} d_T^-(v) \quad \text{and} \quad \widehat{\delta}^+(T) := \min_{v \in V(T) \setminus \{y_1, \dots, y_{kt}\}} d_T^+(v).$$

Let  $c := \lceil \log(32k^2tm) \rceil$ . We may repeatedly apply Lemmas 3.2.4 and 3.2.5 with parameter  $c$  (removing the dominating sets each time) to obtain disjoint sets of vertices  $A_1, \dots, A_{kt}, B_1, \dots, B_{kt}$  and sets of vertices  $E_{A_1}, \dots, E_{A_{kt}}, E_{B_1}, \dots, E_{B_{kt}}$  satisfying the following properties for all  $i \in \{1, \dots, kt\}$ , where we write  $D := \bigcup_{i=1}^{kt} (A_i \cup B_i)$ .

- (i)  $1 \leq |A_i| \leq c$  and  $T[A_i]$  is a transitive tournament with sink  $x_i$ ,
- (ii)  $1 \leq |B_i| \leq c$  and  $T[B_i]$  is a transitive tournament with source  $y_i$ ,
- (iii) either  $E_{A_i} = \emptyset$  or  $E_{A_i}$  lies in the common in-neighbourhood of all vertices in  $A_i$ ,
- (iv) either  $E_{B_i} = \emptyset$  or  $E_{B_i}$  lies in the common out-neighbourhood of all vertices in  $B_i$ ,
- (v)  $T[A_i]$  out-dominates  $V(T) \setminus (D \cup E_{A_i})$ ,
- (vi)  $T[B_i]$  in-dominates  $V(T) \setminus (D \cup E_{B_i})$ ,
- (vii)  $|E_{A_i}| \leq (1/2)^{c-1} \widehat{\delta}^-(T)$ ,
- (viii)  $|E_{B_i}| \leq (1/2)^{c-1} \widehat{\delta}^+(T)$ .

For  $j \in \{1, \dots, t\}$  define  $j^* := \{(j-1)k+1, \dots, (j-1)k+k\}$ , define  $A_j^* := \bigcup_{i \in j^*} A_i$ , and similarly define  $B_j^* := \bigcup_{i \in j^*} B_i$ . Define  $E_A := E_{A_1} \cup \dots \cup E_{A_{kt}}$  and  $E_B := E_{B_1} \cup \dots \cup E_{B_{kt}}$ . Finally define  $E := E_A \cup E_B$ . Note that

$$|E_A| \leq kt \left( \frac{1}{2} \right)^{c-1} \widehat{\delta}^-(T) \leq \frac{1}{16km} \widehat{\delta}^-(T), \quad (3.3.1)$$

by our choice of  $c$ . Similarly,  $|E_B| \leq \widehat{\delta}^+(T)/(16km)$ .

For the remainder of the proof we will assume that  $|E_A| \leq |E_B|$ . The case  $|E_A| > |E_B|$  follows by a symmetric argument. Note then that

$$|E| \leq |E_A| + |E_B| \leq 2|E_B| \leq \widehat{\delta}^+(T)/(8km). \quad (3.3.2)$$

Our aim is to use the dominating sets  $A_i, B_i$  to construct the sets  $V_i$  required. Roughly speaking, for each  $i \in \{1, \dots, kt\}$  our aim is to use the high connectivity of  $T$  in order to find vertex-disjoint paths  $P_i$  in  $T - D$  directed from the sink of  $B_i$  to the source of  $A_i$ . We will then form disjoint vertex sets  $V_1, \dots, V_t$  with

$$A_j^* \cup B_j^* \cup \bigcup_{i \in j^*} V(P_i) \subseteq V_j. \quad (3.3.3)$$

**Claim 1:** *Suppose that  $j \in \{1, \dots, t\}$  and that  $V_j \subset V(T)$  satisfies (3.3.3). Then for any pair of vertices  $x \in V(T) \setminus (D \cup E_B)$  and  $y \in V(T) \setminus (D \cup E_A)$ , the ordered pair  $(x, y)$  is  $k$ -connected in  $T[V_j \cup \{x, y\}]$ .*

Indeed, if we delete an arbitrary set  $S \subset V_j \setminus \{x, y\}$  of at most  $k - 1$  vertices then there is some  $i \in j^*$  such that  $S \cap (A_i \cup B_i \cup V(P_i)) = \emptyset$ . So there is an edge from  $x$  to some vertex  $b \in B_i$  (since  $B_i$  is in-dominating and  $x \notin D \cup E_{B_i}$ ) and an edge from  $b$  to the sink of  $B_i$  (if  $b$  is not the sink of  $B_i$ ); and similarly there is an edge from some vertex  $a \in A_i$  to  $y$  and an edge from the source of  $A_i$  to  $a$  (if  $a$  is not the source of  $A_i$ ). Then these at most four edges together with  $P_i$  form a directed walk from  $x$  to  $y$  in  $T[(V_j \setminus S) \cup \{x, y\}]$ , which we can shorten if necessary to find a directed path from  $x$  to  $y$  in  $T[(V_j \setminus S) \cup \{x, y\}]$ , as required.

Claim 1 is a step towards constructing sets  $V_j$  as required in Theorem 3.1.5. However note that this construction so far ignores the problem of finding paths to or from the (relatively few) vertices in  $D \cup E$  (in order to satisfy Theorem 3.1.5(ii)), and the problem of controlling the sizes of the vertex sets  $V_1, \dots, V_t$  (in order to satisfy Theorem 3.1.5(i)). To address the former problem we will introduce the notion of ‘safe’ vertices and will construct the sets  $V_1, \dots, V_t$  (which will eventually satisfy (3.3.3)) in several steps.

We will colour some vertices of  $V(T)$  with colours in  $\{1, \dots, t\}$ , and at each step  $V_j$  will consist of all vertices of colour  $j$ . At each step we will call a vertex  $v$  in  $V_j$  *forwards-safe* if for any set  $S \not\ni v$  of at most  $k-1$  vertices, there is a directed path (possibly of length 0) in  $T[V_j \setminus S]$  from  $v$  to  $V_j \setminus (D \cup E_B \cup S)$ . Similarly we will call a vertex  $v$  in  $V_j$  *backwards-safe* if for any set  $S \not\ni v$  of at most  $k-1$  vertices, there is a directed path (possibly of length 0) in  $T[V_j \setminus S]$  to  $v$  from  $V_j \setminus (D \cup E_A \cup S)$ . We call a vertex *safe* if it is both forwards-safe and backwards-safe. We also call any vertex in  $V(T) \setminus (V' \cup E)$  safe, where  $V' := \bigcup_{j=1}^t V_j$ . Note that the following properties are satisfied at every step:

- all vertices outside  $D \cup E$  are safe,
- all vertices in  $V' \setminus (D \cup E_B)$  are forwards-safe and all vertices in  $V' \setminus (D \cup E_A)$  are backwards-safe,
- if  $v \in V_j$  has at least  $k$  forwards-safe out-neighbours in  $V_j$  then  $v$  itself is forwards-safe; the analogue holds if  $v$  has at least  $k$  backwards-safe in-neighbours in  $V_j$ ,
- if  $v \in V_j$  is safe and in the next step we enlarge  $V_j$  by colouring some more (previously uncoloured) vertices with colour  $j$  then  $v$  is still safe.

Our aim is to first colour the vertices in  $D$  as well as some additional vertices in such a way as to make all coloured vertices safe (see Claim 3). We will then choose the paths

$P_i$  and colour the vertices on these paths, as well as some additional vertices, in such a way as to make all coloured vertices safe (see Claim 4). Finally we will colour all those vertices in  $E$  which are not coloured yet, as well as some additional vertices, in such a way as to make all coloured vertices safe (see Claim 5). The sets  $V_1, \dots, V_t$  thus obtained will satisfy (3.3.3) and all vertices of  $T$  will be safe. So the next claim will then imply that the sets  $V_1, \dots, V_t$  satisfy Theorem 3.1.5(ii). In order to ensure that Theorem 3.1.5(i) holds as well, we will ensure that in each step we do not colour too many vertices.

**Claim 2:** *Suppose that  $V_1, \dots, V_t$  satisfy (3.3.3) and that  $j \in \{1, \dots, t\}$ . Then for any pair of vertices  $x, y \in V_j \cup (V(T) \setminus V')$  that are both safe, the ordered pair  $(x, y)$  is  $k$ -connected in  $T[V_j \cup \{x, y\}]$ .*

This is immediate from the definitions and Claim 1.

So our goal is to modify our construction so as to ensure that  $V_1, \dots, V_t$  satisfy (3.3.3) and that every vertex in  $V(T)$  is safe. We start with no vertices of  $T$  coloured, and we now begin to colour them. We first colour the vertices in  $D = \bigcup_{j=1}^t (A_j^* \cup B_j^*)$  by giving every vertex in  $A_j^* \cup B_j^*$  colour  $j$ . We now wish to ensure that every vertex in  $D$  is safe.

**Claim 3:** *We can colour some additional vertices of  $T$  in such a way that every coloured vertex is safe, and at most*

$$(k+1)^2(2k^2t + 4k^2t) \tag{3.3.4}$$

*vertices are coloured in total.*

To prove Claim 3 first note that, since  $T$  is by assumption strongly  $10^7 k^6 t^2 m \log(ktm)$ -

connected, it certainly holds that

$$\delta^0(T) \geq 10^7 k^6 t^2 m \log(ktm). \quad (3.3.5)$$

Hence

$$\widehat{\delta}^-(T) - |E_A| \stackrel{(3.3.1)}{\geq} \widehat{\delta}^-(T)/2 \geq \delta^0(T)/2 \stackrel{(3.3.5)}{\geq} 10^6 k^6 t^2 m \log(ktm), \quad (3.3.6)$$

and similarly

$$\widehat{\delta}^+(T) - |E| \stackrel{(3.3.2)}{\geq} \widehat{\delta}^+(T)/2 \geq \delta^0(T)/2 \stackrel{(3.3.5)}{\geq} 10^6 k^6 t^2 m \log(ktm). \quad (3.3.7)$$

Since  $|D| \leq 2kct$ , (3.3.5) implies that for each  $v \in \{x_1, \dots, x_{kt}, y_1, \dots, y_{kt}\}$  in turn we may greedily choose  $k$  uncoloured in-neighbours and  $k$  uncoloured out-neighbours, all distinct from each other, and colour them the same colour as  $v$ . Now the number of coloured vertices is at most  $2kct + 4k^2t$ . So we may greedily choose, for each coloured vertex  $v$  not in  $\{x_1, \dots, x_{kt}, y_1, \dots, y_{kt}\}$  in turn,  $k$  distinct uncoloured in-neighbours not in  $E_A$ , and colour them the same colour as  $v$ . Indeed, this is possible since by (3.3.6) the number of in-neighbours of  $v$  outside  $E_A$  is at least  $(k+1)(2kct + 4k^2t)$ . Now the number of coloured vertices is at most  $(k+1)(2kct + 4k^2t)$ , so by (3.3.7) we may greedily choose, for each coloured vertex  $v$  not in  $\{x_1, \dots, x_{kt}, y_1, \dots, y_{kt}\}$  in turn,  $k$  distinct uncoloured out-neighbours not in  $E$ , and colour them the same colour as  $v$ . Note that the number of coloured vertices is now at most  $(k+1)^2(2kct + 4k^2t)$  and that every coloured vertex is safe, by construction.

We now wish to find the paths  $P_i$  discussed earlier and colour the vertices on these paths appropriately. For  $i \in \{1, \dots, kt\}$  we define an  $i$ -path to be a directed path from the sink of  $B_i$  to the source of  $A_i$ .

**Claim 4:** *For every  $j \in \{1, \dots, t\}$  and every  $i \in j^*$  there exists an  $i$ -path  $P_i$  in  $T$  with previously uncoloured internal vertices, such that all such paths are vertex-disjoint from each other. Moreover we can colour the internal vertices of  $P_i$  with colour  $j$  as well as colouring some additional (previously uncoloured) vertices of  $T$  in such a way that every coloured vertex is safe, and at most*

$$67k^4t^2 \log m + n/(2m) \quad (3.3.8)$$

*vertices are coloured in total.*

We will prove Claim 4 in a series of subclaims. The paths  $P_i$  that we construct for Claim 4 will be either ‘short’ or ‘long’; we deal with these two cases separately. Firstly, for every  $j \in \{1, \dots, t\}$  and every  $i \in j^*$  in turn we choose, if possible, an  $i$ -path of length at most  $k + 1$  with uncoloured internal vertices, vertex-disjoint from all previously chosen paths. For each  $i \in \{1, \dots, kt\}$  for which we find such a path, let  $P_i$  be that path. Let  $\mathcal{P}_{short}$  be the set of paths  $P_i$  of length at most  $k + 1$  found in this way, let  $\mathcal{I}_{short} := \{i \in \{1, \dots, kt\} : \mathcal{P}_{short} \text{ contains an } i\text{-path}\}$ , and let  $\mathcal{I}_{long} := \{1, \dots, kt\} \setminus \mathcal{I}_{short}$ . We colour the internal vertices of each  $i$ -path in  $\mathcal{P}_{short}$  with colour  $j$  (where  $j$  is such that  $i \in j^*$ ). Note that since some of these vertices may be in  $E$ , it is important that we ensure that they are safe.

**Claim 4.1:** *We may colour some (previously uncoloured) vertices of  $T$  in such a way that all coloured vertices are safe, and at most*

$$54k^4t^2 \log m \quad (3.3.9)$$

*vertices are coloured in total. In particular we can ensure that the internal vertices of all paths in  $\mathcal{P}_{short}$  are safe.*

We do this (similarly to before) as follows. By (3.3.4) the number of coloured vertices after colouring the short paths is at most  $(k+1)^2(2kct + 4k^2t) + k^2t$ , so by (3.3.6) we may greedily choose, for every path in  $\mathcal{P}_{short}$  and every internal vertex  $v$  on that path in turn,  $k$  distinct uncoloured in-neighbours not in  $E_A$ , and colour them the same colour as  $v$ . (Note that  $v \notin \{x_1, \dots, x_t, y_1, \dots, y_t\}$  since all the paths in  $\mathcal{P}_{short}$  had uncoloured internal vertices when we chose them.) Now the number of coloured vertices is at most  $(k+1)^2(2kct + 4k^2t) + (k+1)k^2t$ , so by (3.3.7) we may greedily choose, for every path in  $\mathcal{P}_{short}$  and every internal vertex  $v$  on that path, as well as the  $k$  in-neighbours of  $v$  just chosen, in turn,  $k$  distinct uncoloured out-neighbours not in  $E$ , and colour them the same colour as  $v$ . Note that the number of coloured vertices is now at most

$$(k+1)^2(2kct + 4k^2t) + (k+1)k^2t \leq 54k^4t^2 \log m$$

and that every coloured vertex is safe, by construction.

Now we must find  $i$ -paths  $P_i$  for all  $i \in \mathcal{I}_{long}$ ; note that they will all be of length at least  $k+2$ . Initially, for every  $j \in \{1, \dots, t\}$  and every  $i \in j^* \cap \mathcal{I}_{long}$  we will in fact seek  $13k^4t$  distinct internally vertex-disjoint  $i$ -paths with uncoloured internal vertices, such that for every  $i' \in \mathcal{I}_{long} \setminus \{i\}$ , all  $i$ -paths are vertex-disjoint from all  $i'$ -paths. We seek so many such paths because complications later in the proof may require us to colour some vertices in some of the  $i$ -paths with  $i \in j^* \cap \mathcal{I}_{long}$  a colour other than  $j$ , so some spare paths are necessary. It is also important that we control the sizes of these paths so that we are able to control the sizes of the vertex sets  $V_1, \dots, V_t$ .

**Claim 4.2:** *For every  $i \in \mathcal{I}_{long}$  we can find a set  $\mathcal{P}_{i,long}$  of  $13k^4t$  distinct internally vertex-disjoint  $i$ -paths with uncoloured internal vertices, such that for every  $i' \in \mathcal{I}_{long} \setminus \{i\}$ , all paths in  $\mathcal{P}_{i,long}$  are vertex-disjoint from all paths in  $\mathcal{P}_{i',long}$ . Moreover, we may choose the*



sets  $\mathcal{P}_{i, \text{long}}$  such that the total number of internal vertices on the paths in  $\bigcup_{i \in \mathcal{I}_{\text{long}}} \mathcal{P}_{i, \text{long}}$  is at most  $n/(2m)$ .

Indeed, consider the tournament  $T'$  induced on  $T$  by the uncoloured vertices as well as the sinks of  $B_i$  and the sources of  $A_i$ , for every  $i \in \mathcal{I}_{\text{long}}$ . By assumption  $T$  is strongly  $10^7 k^6 t^2 m \log(ktm)$ -connected, so by (3.3.9)  $T'$  is certainly strongly  $2.6 \times 10^5 k^5 t^2 m \log(26k^5 t^2 m)$ -connected. So by Theorem 3.2.3  $T'$  is  $26k^5 t^2 m$ -linked. So since  $|\mathcal{I}_{\text{long}}| \leq kt$ , Proposition 3.2.2 implies that we may find, for each  $i \in \mathcal{I}_{\text{long}}$ , the  $13k^4 t$   $i$ -paths required, and we may do so in such a way that the total number of internal vertices on these paths is at most  $|V(T')|/(2m) \leq n/(2m)$ , as required.

For each  $i \in \mathcal{I}_{\text{long}}$ , we obtain from each of the paths in  $\mathcal{P}_{i, \text{long}}$  a possibly shorter path by deleting from the path any vertex  $v$  such that there is an edge in  $T$  directed from an ancestor of  $v$  in the path to a descendant of  $v$  in the path. We replace each of the paths in  $\mathcal{P}_{i, \text{long}}$  by the corresponding shorter path obtained. Note that this ensures that each of the paths in  $\mathcal{P}_{i, \text{long}}$  is now a backwards-transitive path of length at least  $k + 2$ . As before, it is important that we now ensure that the internal vertices on these paths are coloured in such a way as to be safe, while also colouring them in accordance with the requirements of Claim 4; we do this as follows.

**Claim 4.3:** *For every  $j \in \{1, \dots, t\}$  and every  $i \in j^* \cap \mathcal{I}_{\text{long}}$  we may colour the internal vertices of all paths in  $\mathcal{P}_{i, \text{long}}$  as well as some additional (previously uncoloured) vertices of  $T$  in such a way that every coloured vertex is safe and at least one path  $P_i$  in  $\mathcal{P}_{i, \text{long}}$  has all vertices coloured with colour  $j$ . Moreover, we can do this so that at most*

$$67k^4 t^2 \log m + n/(2m) \tag{3.3.10}$$

vertices are coloured in total.

Indeed, for each  $j \in \{1, \dots, t\}$  consider the tournament induced on  $T$  by the set of all interior vertices of all paths in  $\mathcal{P}_{i, \text{long}}$  for all  $i \in j^* \cap \mathcal{I}_{\text{long}}$ . Note that this tournament satisfies the assumptions of Lemma 3.2.7 (with  $13k^4t \cdot |j^* \cap \mathcal{I}_{\text{long}}|$  playing the role of  $\ell$ ) since each of the paths in each of the sets  $\mathcal{P}_{i, \text{long}}$  is a backwards-transitive path of length at least  $k+2$ . So consider the sets  $U, W$  each of size at most  $2k(k+1)$  and the sets  $U', W'$  each of size at most  $13k^5t(k+1)$  given by Lemma 3.2.7. Let us call them  $U_j, W_j, U'_j, W'_j$  respectively. By the properties of  $U_j, W_j, U'_j, W'_j$  and the definitions of forwards-safe and backwards-safe, it is clear that if every vertex in  $U'_j$  is coloured  $j$  and every vertex in  $U_j$  is forwards-safe, and every vertex in  $W'_j$  is coloured  $j$  and every vertex in  $W_j$  is backwards-safe, then for all  $i \in j^* \cap \mathcal{I}_{\text{long}}$  every vertex on paths in  $\mathcal{P}_{i, \text{long}}$  that is coloured  $j$  will be safe. So for each  $j \in \{1, \dots, t\}$  we colour all vertices in  $U'_j \cup W'_j$  with colour  $j$ , and we now aim to make every vertex in  $U_j$  forwards-safe and every vertex in  $W_j$  backwards-safe; we accomplish this (similarly to the way we have made vertices safe before) as follows. By (3.3.9) the number of coloured vertices is at most  $54k^4t^2 \log m + 26k^5t^2(k+1)$ , so by (3.3.6) we may greedily choose, for every  $j \in \{1, \dots, t\}$  and for each vertex in  $W_j$  in turn,  $k$  distinct uncoloured in-neighbours not in  $E_A$ , and colour them  $j$ . Now, the number of coloured vertices is at most  $54k^4t^2 \log m + 26k^5t^2(k+1) + 2k^2(k+1)t$ , so by (3.3.7) we may greedily choose, for every  $j \in \{1, \dots, t\}$  and for each vertex in  $U_j$  and each of the  $k$  in-neighbours of each of the vertices in  $W_j$  just chosen in turn,  $k$  distinct uncoloured out-neighbours not in  $E$ , and colour them  $j$ . Let  $Z$  be the set of all those vertices that we have just coloured to make all vertices in each  $U_j$  forwards-safe and all vertices in each  $W_j$  backwards-safe. Note that  $|Z| \leq 2k^2(k+1)t + k(2k(k+1)t + 2k^2(k+1)t) < 13k^4t$ .

Note also that some of the vertices in  $Z$  may be contained in some of the paths in  $\mathcal{P}_{i, \text{long}}$  for some  $i \in \mathcal{I}_{\text{long}}$ ; this is the reason for which we found spare paths. For each  $i \in \mathcal{I}_{\text{long}}$ ,

since  $|\mathcal{P}_{i, \text{long}}| = 13k^4t$ , there is at least one path in  $\mathcal{P}_{i, \text{long}}$  that contains no vertices in  $Z$ ; let  $P_i$  be one such path. Colour any uncoloured vertices remaining in paths in the sets  $\mathcal{P}_{i, \text{long}}$  with colour  $j$ , where  $j$  is such that  $i \in j^*$ . In particular the vertices of  $P_i$  all have colour  $j$ . So we have now found our paths  $P_i$  for all  $i \in \mathcal{I}_{\text{long}}$ , and every coloured vertex is safe by construction. Also note that the number of coloured vertices is now at most

$$54k^4t^2 \log m + 13k^4t + n/(2m) \leq 67k^4t^2 \log m + n/(2m),$$

as required for Claim 4.3.

This completes the proof of Claim 4.

Now that we have built all of the structure required, it remains for us to colour the uncoloured vertices in  $E$  in such a way as to ensure that they are safe. This is essential as, recalling the definition, uncoloured vertices in  $E$  are not safe.

**Claim 5:** *We can colour the uncoloured vertices in  $E$  as well as some additional (previously uncoloured) vertices of  $T$  in such a way that every coloured vertex is safe, and at most  $n/m$  vertices are coloured in total.*

In order to prove Claim 5 we colour all the uncoloured vertices  $v \in E$  by distinguishing three cases. We first colour all uncoloured vertices  $v \in E$  which satisfy the assumptions of Case 1, then we colour all uncoloured vertices  $v \in E$  which satisfy the assumptions of Case 2, and then we colour all uncoloured vertices  $v \in E$  which satisfy the assumptions of Case 3.

**Case 1:** *There exist (not necessarily distinct)  $j_1, j_2 \in \{1, \dots, t\}$  such that  $|\{i \in j_1^* : v \in$*

$$|E_{A_i}| \leq |\{i \in j_1^* : v \in E_{B_i}\}| \text{ and } |\{i \in j_2^* : v \in E_{A_i}\}| \geq |\{i \in j_2^* : v \in E_{B_i}\}|.$$

Note that by (3.3.2) it certainly holds that  $|E| \leq n/(8km)$ . So by (3.3.8) the number of uncoloured vertices not in  $E$  is at least

$$n \left( 1 - \frac{1}{2m} - \frac{1}{8km} \right) - 67k^4 t^2 \log m \geq n - \frac{3n}{4m}. \quad (3.3.11)$$

Either there are  $k$  such vertices that are all out-neighbours of  $v$ , or there are not, in which case there must be  $k$  such vertices that are all in-neighbours of  $v$ .

**Case 1.1:** If  $v$  has  $k$  uncoloured out-neighbours not in  $E$ , we colour them and  $v$  with colour  $j_1$ . This ensures that  $v$  is forwards-safe. To see that  $v$  is backwards-safe too, note that if  $v \notin E_{A_i}$  then there is an edge in  $T$  directed to  $v$  from a (safe) vertex in  $A_i$ , but similarly that if  $v \in E_{B_i}$  then there is an edge in  $T$  directed to  $v$  from a (safe) vertex in  $B_i$ . Together with our assumption that  $|\{i \in j_1^* : v \in E_{A_i}\}| \leq |\{i \in j_1^* : v \in E_{B_i}\}|$  this ensures that  $v$  has  $k$  safe in-neighbours of its colour. So  $v$  is backwards-safe.

**Case 1.2:** If  $v$  does not have  $k$  uncoloured out-neighbours outside  $E$  then  $v$  must have  $k$  uncoloured in-neighbours not in  $E$ ; we colour them and  $v$  with colour  $j_2$ . This ensures that  $v$  is backwards-safe. To see that  $v$  is forwards-safe too, note that if  $v \notin E_{B_i}$  then there is an edge in  $T$  directed from  $v$  to a (safe) vertex in  $B_i$ , but similarly that if  $v \in E_{A_i}$  then there is an edge in  $T$  directed from  $v$  to a (safe) vertex in  $A_i$ . Together with our assumption that  $|\{i \in j_2^* : v \in E_{A_i}\}| \geq |\{i \in j_2^* : v \in E_{B_i}\}|$  this ensures that  $v$  has  $k$  safe out-neighbours of its colour. So  $v$  is forwards-safe.

By (3.3.11) we can repeat this process greedily for all vertices  $v \in E$  which satisfy the assumptions of Case 1. Note that after this step all coloured vertices are safe.

**Case 2:** For all  $j \in \{1, \dots, t\}$  it holds that  $|\{i \in j^* : v \in E_{A_i}\}| < |\{i \in j^* : v \in E_{B_i}\}|$ .

We consider two sub-cases:

**Case 2.1:** If  $v$  has  $k$  uncoloured out-neighbours not in  $E$  then colour them and  $v$  with colour 1.

**Case 2.2:** Otherwise, since (3.3.7) implies that  $\widehat{\delta}^+(T) \geq kt + k + |E|$ , an averaging argument shows that there is some  $j \in \{1, \dots, t\}$  such that  $v$  has  $k$  out-neighbours of colour  $j$  (recall that all currently coloured vertices are safe), in which case we colour  $v$  with colour  $j$ .

In either case it is clear that  $v$  is now forwards-safe. A similar argument as in Case 1.1 shows that  $v$  is backwards-safe too.

**Case 3:** For all  $j \in \{1, \dots, t\}$  it holds that  $|\{i \in j^* : v \in E_{A_i}\}| > |\{i \in j^* : v \in E_{B_i}\}|$ .

We consider two sub-cases:

**Case 3.1:** If  $v$  has  $k$  uncoloured in-neighbours not in  $E_A$  then colour them and  $v$  with colour 1. (Note that none of these in-neighbours  $w$  can lie in  $E_B$ . Indeed, if  $w \in E_B$  then  $w$  satisfies the assumptions of one of the first two cases (as  $w \notin E_A$  implies  $|\{i \in j^* : v \in E_{A_i}\}| = 0$ ) and so  $w$  would have already been coloured.)

**Case 3.2:** Otherwise, since (3.3.6) implies that  $\widehat{\delta}^-(T) \geq kt + k + |E_A|$ , an averaging argument shows that there is some  $j \in \{1, \dots, t\}$  such that  $v$  has  $k$  in-neighbours of colour  $j$  (recall that all currently coloured vertices are safe), in which case we colour  $v$  with colour  $j$ .

In either case it is clear that  $v$  is now backwards-safe. Again, a similar argument as in Case 1.2 shows that  $v$  is forwards-safe too.

This covers all cases, so we have now coloured all vertices in  $E$  in such a way that all coloured vertices are safe. Note that for each of the at most  $|E| \leq n/(8mk)$  vertices in  $E$  that were uncoloured at the start of the proof of Claim 5 we have coloured at most  $k$  (previously uncoloured) vertices not in  $E$  in this step. So by (3.3.11) the total number of coloured vertices is at most  $3n/(4m) + (k+1)|E| \leq n/m$ , as required.

Now the only uncoloured vertices remaining are not in  $E$  and so they are safe. So all vertices in  $T$  are now safe. This completes the construction of the vertex sets required, where the colour classes of colours  $1, \dots, t$  correspond to the vertex sets  $V_1, \dots, V_t$  respectively. Since the number of coloured vertices is at most  $n/m$ , the size of each  $V_j$  is certainly at most  $n/m$ . And since we have ensured that every vertex in  $T$  is safe, Claim 2 implies that the  $V_j$  satisfy the requirements of Theorem 3.1.5.  $\square$

### 3.4 Partitioning tournaments into vertex-disjoint cycles

The purpose of this section is to derive Theorem 3.1.7 from Theorem 3.1.5.

**Proof of Theorem 3.1.7.** Note that by averaging there is at least one value  $j \in \{1, \dots, t\}$  for which  $L_j \geq n/t$ . Without loss of generality let  $L_1 \geq n/t$ . Let  $\tilde{J} := \{j \in \{1, \dots, t\} : L_j < n/(2t^2)\}$ . For  $j \in \tilde{J}$  let  $L'_j := \lceil n/t^2 \rceil$ . For  $j \in \{2, \dots, t\} \setminus \tilde{J}$  let  $L'_j := L_j$ . Let  $L'_1 := L_1 - \sum_{j=2}^t (L'_j - L_j)$ . Note that  $L'_1 \geq n/t^2$  and that  $\sum_{j=1}^t L'_j = n$ .

Since  $10^{10}t^4 \log t \geq 10^7 2^6 t^2 (2t^2) \log(2t(2t^2))$ , we have by Theorem 3.1.5 that  $V(T)$  contains  $t$  disjoint sets of vertices,  $V_1, \dots, V_t$ , such that for every  $j \in \{1, \dots, t\}$  the following hold:

(i)  $|V_j| \leq n/(2t^2)$ ,

(ii) for any set  $R \subseteq V(T) \setminus \bigcup_{i=1}^t V_i$  the subtournament  $T[V_j \cup R]$  is strongly 2-connected.

Construct a partition  $V'_1, \dots, V'_t$  of the vertices of  $T$ , such that for every  $j \in \{1, \dots, t\}$  it holds that  $V_j \subseteq V'_j$  and that  $|V'_j| = L'_j$ . This is possible, since for every  $j \in \{1, \dots, t\}$  we have  $L'_j \geq n/(2t^2) \geq |V_j|$ . Note that, for every  $j \in \{1, \dots, t\}$ ,  $T[V'_j]$  is strongly 2-connected.

Now, since  $n/t^2 > 7$ , we have by Theorem 3.1.6 that for each  $j \in \tilde{J}$ ,  $T[V'_j]$  contains two vertex-disjoint cycles of lengths  $L_j$  and  $L'_j - L_j$ . The cycle of length  $L_j$  we call  $C_j$  and the cycle of length  $L'_j - L_j$  we call  $C'_j$ . Since for every  $j \in \tilde{J}$  we have that  $|C'_j| = L'_j - L_j > n/2t^2 \geq |V_j|$ , there is at least one vertex in  $V(C'_j) \cap (V'_j \setminus V_j)$ . Call one such vertex  $v_j$ . Let  $R$  be the set of all vertices  $v_j$  for  $j \in \tilde{J}$ .

Now let  $V''_1 := V'_1 \cup \bigcup_{j \in \tilde{J}} V(C'_j)$ . Note that  $|V''_1| = L_1$ . Note also that (ii) implies that  $T[V'_1 \cup R]$  is strongly 2-connected; so certainly it is strongly 1-connected. We now claim that  $T[V''_1]$  is strongly 1-connected. Indeed, suppose  $x, y \in V''_1$ , and we wish to find a path directed from  $x$  to  $y$  in  $T[V''_1]$ . First note that if  $x \notin V'_1$  then  $x \in V(C'_j)$  for some  $j \in \tilde{J}$ , so there is a path  $Q_j$  in  $T[V(C'_j)]$ , possibly of length 0, from  $x$  to  $v_j \in R$ . Similarly note that if  $y \notin V'_1$  then  $y \in V(C'_i)$  for some  $i \in \tilde{J}$ , so there is a path  $Q'_i$  in  $T[V(C'_i)]$ , possibly of length 0, to  $y$  from  $v_i \in R$ . Since  $T[V'_1 \cup R]$  is strongly 1-connected there exists a path  $P$  in  $T[V'_1 \cup R]$  directed from  $v_j$  to  $v_i$ . So  $Q_j P Q'_i$  is a walk in  $T[V''_1]$  directed from  $x$  to  $y$ . So indeed  $T[V''_1]$  is strongly 1-connected.

Note also that for every  $j \in \{2, \dots, t\} \setminus \tilde{J}$  we have that  $T[V'_j]$  is strongly 2-connected, so certainly strongly 1-connected. So by Camion's theorem  $T[V''_1]$  contains a Hamilton cycle,  $C_1$  say, and for every  $j \in \{2, \dots, t\} \setminus \tilde{J}$  we have that  $T[V'_j]$  contains a Hamilton cycle,  $C_j$

say.

Now the cycles  $C_1, \dots, C_t$  are vertex-disjoint and are of lengths  $L_1, \dots, L_t$  respectively, so this completes the proof.  $\square$



# CHAPTER 4

## ON THE STRUCTURE OF ORIENTED GRAPHS AND DIGRAPHS WITH FORBIDDEN TOURNAMENTS OR CYCLES

### 4.1 Chapter introduction

#### 4.1.1 Oriented graphs and digraphs with forbidden subdigraphs

Recall from Chapter 1 that, while much is known about the number and typical structure of graphs that do not contain some forbidden subgraph, the corresponding questions for digraphs and oriented graphs are almost all wide open. Until now the only results of this type for oriented graphs were proved by Balogh, Bollobás and Morris [8, 9] who classified the possible ‘growth speeds’ of oriented graphs with a given property. Moreover Robinson [66, 67], and independently Stanley [74], counted the number of acyclic digraphs.

In this chapter we determine the typical structure of oriented graphs that do not contain a transitive tournament of size  $k$ , and of oriented graphs that do not contain an oriented cycle of size  $k$ , as well as proving corresponding results for digraphs.

### 4.1.2 Oriented graphs and digraphs with forbidden tournaments or cycles

A *digraph* is a pair  $(V, E)$  where  $V$  is a set of vertices and  $E$  is a set of ordered pairs of distinct vertices in  $V$  (note that this means that in a digraph we do not allow loops or multiple edges in the same direction). An *oriented graph* is a digraph with at most one edge between two vertices, so may be considered as an orientation of a simple undirected graph. In this chapter we denote a transitive tournament on  $k$  vertices by  $T_k$ , and a directed cycle on  $k$  vertices by  $C_k$ . We only consider labelled graphs and digraphs.

Clearly any transitive tournament is  $C_k$ -free for any  $k$ , and any bipartite digraph is  $T_3$ -free. In 1998 Cherlin [21] gave a classification of countable homogeneous oriented graphs. He remarked that ‘the striking work of [46] does not appear to go over to the directed case’ and made the following conjectures.<sup>1</sup>

**Conjecture 4.1.1 (Cherlin).**

- (i) *Almost all  $T_3$ -free oriented graphs are tripartite.*
- (ii) *Almost all  $C_3$ -free oriented graphs are acyclic, i.e. they are subgraphs of transitive tournaments.*

The first main result of this chapter not only verifies part (i) of this conjecture, but shows

---

<sup>1</sup>Note that oriented graphs are referred to as digraphs in [21].

for all  $k \geq 2$  that almost all  $T_{k+1}$ -free oriented graphs are  $k$ -partite. Note that in particular this shows that in fact almost all  $T_3$ -free oriented graphs are actually even bipartite. We also prove the analogous result for digraphs.

**Theorem 4.1.2.** *Let  $k \in \mathbb{N}$  with  $k \geq 2$ . Then the following hold.*

- (i) *Almost all  $T_{k+1}$ -free oriented graphs are  $k$ -partite.*
- (ii) *Almost all  $T_{k+1}$ -free digraphs are  $k$ -partite.*

Theorem 4.1.2 can be viewed as a directed version of Theorem 1.4.2. Note also that (i) means that, unlike in related results, the typical structure of a  $T_3$ -free oriented graph is *not* close to that of the extremal  $T_3$ -free oriented graph: it is easy to see that the latter is the blow up of a directed triangle (this fact was probably the motivation for Conjecture 4.1.1(i)).

The next main result of this chapter shows in particular that part (ii) of Conjecture 4.1.1 is in fact false. We actually show something stronger, namely that for all  $k \geq 3$  and for almost all  $C_k$ -free oriented graphs on  $n$  vertices, the number of edges we must change in order to get an acyclic oriented graph is  $\Omega(n/\log n)$ . We also prove an analogous version of this result for digraphs. However, Conjecture 4.1.1(ii) is not too far from being true, as we prove also that almost all  $C_k$ -free oriented graphs are close to acyclic, in the sense that we only need to change sub-quadratically many edges in order to obtain an acyclic oriented graph. In the case when  $k$  is even we prove an analogous version of this result for digraphs too. We also obtain a (less restrictive) structural result for odd  $k$ .

In order to state the theorem precisely we need to introduce a little terminology. Given a labelled digraph or oriented graph  $G$  with vertex labels  $1, \dots, n$  and an ordering  $\sigma : [n] \rightarrow [n]$ , a *backwards edge* in  $G$  with respect to this ordering is any edge directed from a

vertex labelled  $i$  to a vertex labelled  $j$ , where  $\sigma(i) > \sigma(j)$ . A *transitive-optimal ordering* of  $G$  is any ordering of  $V(G)$  that minimises the number of backwards edges in  $G$  with respect to the ordering. We say that a directed graph is a *transitive-bipartite blow up* if it can be obtained from a transitive tournament by replacing some of its vertices by complete balanced bipartite digraphs. More formally, a directed graph  $G = (V, E)$  is a transitive-bipartite blow up if  $V$  admits a partition  $A_1, \dots, A_t$  such that for all  $i, j \in [t]$  with  $i < j$ , the graph induced on  $G$  by  $A_i$  is either a single vertex or a complete balanced bipartite digraph (with edges in both directions) and the edges in  $E$  between  $A_i$  and  $A_j$  are precisely those edges directed from  $A_i$  to  $A_j$  (and no others).

**Theorem 4.1.3.** *Let  $k, n \in \mathbb{N}$  with  $k \geq 3$ . There exists  $c > 0$  such that for every  $\alpha > 0$  the following hold.*

- (i) *Almost all  $C_k$ -free oriented graphs on  $n$  vertices have between  $cn/\log n$  and  $\alpha n^2$  backwards edges in a transitive-optimal ordering.*
- (ii) *Almost all  $C_k$ -free digraphs on  $n$  vertices have at least  $cn/\log n$  backwards edges in a transitive-optimal ordering. Moreover,*
  - (a) *if  $k$  is even then almost all  $C_k$ -free digraphs on  $n$  vertices have at most  $\alpha n^2$  backwards edges in a transitive-optimal ordering,*
  - (b) *if  $k$  is odd then almost all  $C_k$ -free digraphs on  $n$  vertices can be made into a subgraph of a transitive-bipartite blow up by changing at most  $\alpha n^2$  edges.*

We believe that in fact almost all  $C_k$ -free oriented graphs have linearly many backwards edges, and that an analogous result holds for  $C_k$ -free digraphs in the case when  $k$  is even.

**Conjecture 4.1.4.** *Let  $k, n \in \mathbb{N}$  with  $k \geq 3$ . Then the following hold.*

- (i) *Almost all  $C_k$ -free oriented graphs on  $n$  vertices have  $\Theta(n)$  backwards edges in a transitive-optimal ordering.*
- (ii) *If  $k$  is even then almost all  $C_k$ -free digraphs on  $n$  vertices have  $\Theta(n)$  backwards edges in a transitive-optimal ordering.*

It is not clear to us what to expect in the case when  $k$  is odd.

**Question 4.1.5.** *Suppose that  $k$  is odd and  $\alpha > 0$ . Do almost all  $C_k$ -free digraphs on  $n$  vertices have at most  $\alpha n^2$  backwards edges in a transitive-optimal ordering?*

An undirected version of Theorem 4.1.3 for forbidden odd cycles was proved by Lamken and Rothschild [54], who showed that for odd  $k$ , almost all  $C_k$ -free graphs are bipartite. (So the situation for oriented graphs is very different from the undirected one.) For even  $k$  the undirected problem is far more difficult. Despite major recent progress by Morris and Saxton [56] the problem of counting the number of  $C_k$ -free graphs is still open for even  $k$ .

We remark that in Theorem 4.1.2 we actually get exponential bounds on the proportion of  $T_{k+1}$ -free oriented graphs and digraphs that are not  $k$ -partite. We also get similar exponential bounds in Theorem 4.1.3.

### 4.1.3 Overview of the chapter

A key tool in our proofs is a recent and very powerful result, due to Saxton and Thomason [70], and independently Balogh, Morris and Samotij [11], which introduces the notion of hypergraph containers to give an upper bound on the number of independent sets in

hypergraphs. Briefly, the result states that under suitable conditions on a uniform hypergraph  $G$ , there is a small collection  $\mathcal{C}$  of small subsets (known as containers) of  $V(G)$  such that every independent set of vertices in  $G$  is a subset of some element of  $\mathcal{C}$ . The precise statement of this result (Theorem 4.3.2) is deferred until Section 4.3. It should be noted that the method of hypergraph containers is much more general than this, and Theorem 4.3.2 is just one of many useful applications of it; in particular other applications are explored in [11] and [70]. Saxton and Thomason also gave a short proof of a somewhat weaker version of the result (which would still have been strong enough for our purposes) in [71]. Roughly speaking, the use of hypergraph containers allows us to reduce an asymptotic counting problem to an extremal problem.

Our approach to proving the main results of this chapter is as follows. Firstly, in Section 4.3 we use the main result of [70] to derive a container result which is applicable to digraphs. Then in Section 4.4 we apply this digraph containers result in a relatively standard way to show that almost all  $T_{k+1}$ -free oriented graphs, and almost all  $T_{k+1}$ -free digraphs, are close to  $k$ -partite (see Lemma 4.4.9). In Section 4.5 we combine Lemma 4.4.9 with an inductive argument to prove the results on the exact structure of typical  $T_{k+1}$ -free oriented graphs and digraphs given by Theorem 4.1.2.

In Section 4.6 we use the digraph containers result to show that almost all  $C_k$ -free oriented graphs are close to acyclic, and that the analogous result for digraphs holds in the case when  $k$  is even (see Lemma 4.6.21(i), (ii)). For odd  $k$  we show that almost all  $C_k$ -free digraphs are close to a subgraph of a transitive-bipartite blow up (see Lemma 4.6.21(iii)). Finally, in Section 4.7 we complete the proof of Theorem 4.1.3 by giving a lower bound on the number of backwards edges in  $C_k$ -free oriented graphs and digraphs, the upper bounds in Theorem 4.1.3 being given by Lemma 4.6.21.

As part of the proofs in Sections 4.4 and 4.6 we prove several stability results on digraphs

which are potentially of independent interest:

- (i) Suppose  $k \in \mathbb{N}$  and  $G$  is a  $T_{k+1}$ -free digraph on  $n$  vertices with  $e(G) \geq \text{ex}_{di}(n, T_{k+1}) - o(n^2)$ . Then  $G$  is close to a complete balanced  $k$ -partite digraph. (See Lemma 4.4.3.)
- (ii) Suppose  $k \in \mathbb{N}$  with  $k \geq 4$  and  $k$  even, and suppose  $G$  is a  $C_k$ -free digraph on  $n$  vertices with  $e(G) \geq \text{ex}_{di}(n, C_k) - o(n^2)$ . Then  $G$  is close to a transitive tournament. (See Lemma 4.6.13.)
- (iii) Suppose  $k \in \mathbb{N}$  with  $k \geq 3$  and  $k$  odd, and suppose  $G$  is a  $C_k$ -free digraph on  $n$  vertices with  $e(G) \geq \text{ex}_{di}(n, C_k) - o(n^2)$ . Then  $G$  is close to a transitive-bipartite blow up. (See Lemma 4.6.17.)

Here  $\text{ex}_{di}(n, H)$  denotes the maximum number of edges among all  $H$ -free digraphs on  $n$  vertices. The corresponding Turán type results which determine  $\text{ex}_{di}(n, H)$  for  $H = T_k$  and  $H = C_k$  were proved by Brown and Harary [17] and Häggkvist and Thomassen [36] respectively. These stability results are used in the proofs of Theorems 4.1.2(ii) and 4.1.3(ii). We actually prove ‘weighted’ generalisations of (i) and (ii) which can be used to prove the assertions about oriented graphs in Theorems 4.1.2(i) and 4.1.3(i).

Before starting on any of this however, we lay out some notation and set out some useful tools in Section 4.2, below.

## 4.2 Notation and tools

For a set  $X$  we let  $X^{(r)}$  denote the set of all (unordered) subsets of  $X$  of size  $r$ . Recall that an  $r$ -uniform hypergraph, or  $r$ -graph, is a pair  $(V, E)$  where  $V$  is a set of vertices and  $E \subseteq V^{(r)}$ . If  $G = (V, E)$  is a graph, digraph, oriented graph or  $r$ -graph, we let

$V(G) := V$ ,  $E(G) := E$ ,  $v(G) := |V(G)|$ , and  $e(G) := |E(G)|$ . For a digraph  $G = (V, E)$  define  $\Delta^0(G)$  as the maximum of  $d^+(v)$  and  $d^-(v)$  among all  $v \in V$ . We write  $uv$  for the edge directed from  $u$  to  $v$ . For a vertex  $v \in V$ , define the out-neighbourhood of  $v$  in  $G$  to be  $N_G^+(v) := \{u \in V : vu \in E\}$ , and similarly define the in-neighbourhood of  $v$  in  $G$  to be  $N_G^-(v) := \{u \in V : uv \in E\}$ . Given a set  $U \subseteq V$ , we sometimes also write  $N_U^+(v) := N_G^+(v) \cap U$  and define  $N_U^-(v)$  similarly. For disjoint subsets  $U, U' \subseteq V$  we let  $G[U, U']$  denote the subdigraph of  $G$  with vertex set  $U \cup U'$  whose edge set consists of all edges between  $U$  and  $U'$  in  $G$  (in both directions). We let  $e(U, U') := e(G[U, U'])$ . If  $Q$  is a  $k$ -partition of  $[n]$  with partition classes  $V_1 \dots, V_k$ , and  $G$  is a (di)graph or oriented graph with vertex set  $[n]$ , we say that  $Q$  is a *k-partition of  $G$*  if for every  $i \in [k]$  we have that  $E(G)$  contains no edges  $uv$  with  $u, v \in V_i$ . We assume  $k$ -partitions to be unordered unless otherwise stated. For two digraphs  $G$  and  $G'$  on vertex set  $[n]$ , we write  $G = G' \pm \varepsilon n^2$  if  $G$  can be obtained from  $G'$  by changing (i.e. adding, deleting, or changing the orientation of) at most  $\varepsilon n^2$  edges. Given an  $r$ -graph  $H = (V, E)$  and  $\sigma \in V^{(d)}$ , where  $0 \leq d \leq r - 1$ , let  $d_H(\sigma) := |\{e \in E : \sigma \subseteq e\}|$  be the *degree* of  $\sigma$  in  $H$ . We may simply write  $d(\sigma)$  for  $d_H(\sigma)$  when it is obvious which  $r$ -graph  $H$  we are working with. The *average vertex degree* of  $H$  is defined to be  $(1/|V|) \sum_{v \in V} d_H(\{v\})$ . In this chapter, given an oriented graph  $H$  and a digraph  $H'$ ,

- let  $f(n, H)$  denote the number of labelled  $H$ -free oriented graphs on  $n$  vertices,
- let  $T(n, k)$  denote the number of labelled  $k$ -partite oriented graphs on  $n$  vertices,
- let  $f^*(n, H')$  denote the number of labelled  $H'$ -free digraphs on  $n$  vertices,
- let  $T^*(n, k)$  denote the number of labelled  $k$ -partite digraphs on  $n$  vertices.

We will make use of the following conventions throughout this chapter and Chapter 5. Given  $a, b \in \mathbb{R}$  with  $0 < a, b < 1$ , we will use the notation  $a \ll b$  to mean that we can find



an increasing function  $g$  for which all of the conditions in the proof are satisfied whenever  $a \leq g(b)$ . We write  $\log x$  to mean  $\log_2 x$ , and we assume all graphs, oriented graphs, and digraphs to be labelled unless otherwise stated. We also assume all large numbers to be integers, so that we may sometimes omit floors and ceilings for the sake of clarity.

We define  $H(p) := -p \log p - (1-p) \log(1-p)$ , the binary entropy function. The following bound will prove useful to us. For  $n \geq 1$  and  $0 < p < 1/2$ ,

$$\binom{n}{\leq pn} := \sum_{i=0}^{\lfloor pn \rfloor} \binom{n}{i} \leq 2^{H(p)n}. \quad (4.2.1)$$

In a number of our proofs we shall also use the following Chernoff bound.

**Theorem 4.2.2 (Chernoff bound).** *Let  $X$  have binomial distribution and let  $a > 0$ . Then*

$$P(X < \mathbb{E}[X] - a) < \exp\left(-\frac{a^2}{2\mathbb{E}[X]}\right).$$

### 4.3 Digraph containers

Our main tool in this chapter is Theorem 4.3.2 from [70]. Given a hypergraph  $G$  satisfying certain degree conditions, it gives a small set of almost independent sets in  $G$  (containers) which together contain all independent sets of  $G$ . In our applications the vertex set of  $G$  will be the edge set of the complete digraph, and the hyperedges will correspond to copies of the forbidden subdigraph. To formulate the degree conditions we need the following definition.

**Definition 4.3.1.** *Let  $G$  be an  $r$ -graph on  $n$  vertices with average vertex degree  $d$ . Let*

$\tau > 0$ . Given  $v \in V(G)$  and  $1 \leq j \leq r, n$ , let

$$d^{(j)}(v) := \max\{d(\sigma) : v \in \sigma \subseteq V(G), |\sigma| = j\}.$$

If  $d > 0$  we define  $\delta_j = \delta_j(\tau)$  by the equation

$$\delta_j \tau^{j-1} n d = \sum_{v \in V(G)} d^{(j)}(v).$$

Then the co-degree function  $\delta(G, \tau)$  is defined by

$$\delta(G, \tau) := 2^{\binom{r}{2}-1} \sum_{j=2}^r 2^{-\binom{j-1}{2}} \delta_j.$$

If  $d = 0$  we define  $\delta(G, \tau) := 0$ .

**Theorem 4.3.2.** [70, Corollary 2.7] Suppose that  $0 < \varepsilon < \frac{1}{2}$  and  $\tau \leq \frac{1}{144r^2!r}$ . Let  $G$  be an  $r$ -graph with vertex set  $[n]$  satisfying  $\delta(G, \tau) \leq \frac{\varepsilon}{12r!}$ . Then there exists a constant  $c = c(r)$  and a collection  $\mathcal{C}$  of subsets of  $[n]$  with the following properties.

- (a) For every independent set  $I$  of  $G$  there exists  $C \in \mathcal{C}$  such that  $I \subseteq C$ .
- (b)  $e(G[C]) \leq \varepsilon e(G)$  for all  $C \in \mathcal{C}$ .
- (c)  $\log |\mathcal{C}| \leq c \log\left(\frac{1}{\varepsilon}\right) n \tau \log\left(\frac{1}{\tau}\right)$ .

We will apply Theorem 4.3.2 to prove Theorem 4.3.3 below, which is a digraph analogue to [70, Theorem 1.3]. To state this we need the following definitions. Given a digraph  $G = (V, E)$ , let  $f_1(G)$  be the number of pairs  $u, v \in V$  such that exactly one of  $uv$  and  $vu$  is an edge of  $G$ , and let  $f_2(G)$  be the number of pairs  $u, v \in V$  such that both  $uv$  and  $vu$  are edges of  $G$ . The following definition of the *weighted size* of  $G$  will be crucial in this

chapter. For  $a \in \mathbb{R}$  with  $a \geq 1$  we define

$$e_a(G) := a \cdot f_2(G) + f_1(G).$$

This definition allows for a unified approach to extremal problems on oriented graphs and digraphs. We will be mainly interested in the cases  $a = 2$  and  $a = \log 3$ . The former is useful because each digraph  $G$  contains  $4^{f_2(G)} 2^{f_1(G)} = 2^{e_2(G)}$  (labelled) subdigraphs, and the latter is useful because each digraph  $G$  contains  $3^{f_2(G)} 2^{f_1(G)} = 2^{e_{\log 3}(G)}$  (labelled) oriented subgraphs. Given a digraph  $H$ , define the *weighted Turán number*  $\text{ex}_a(n, H)$  as the maximum  $e_a(G)$  among all  $H$ -free digraphs  $G$  on  $n$  vertices. (So  $\text{ex}_2(n, H)$  equals  $\text{ex}_{di}(n, H)$  which was defined in Section 4.1.3.) For  $A, B \subseteq V$  we will sometimes write  $e_a(A, B)$  to denote  $e_a(G[A, B])$ .

Given an oriented graph  $H$  with  $e(H) \geq 2$ , we let

$$m(H) = \max_{H' \subseteq H, e(H') > 1} \frac{e(H') - 1}{v(H') - 2}.$$

**Theorem 4.3.3.** *Let  $H$  be an oriented graph with  $h := v(H)$  and  $e(H) \geq 2$ , and let  $a \in \mathbb{R}$  with  $a \geq 1$ . For every  $\varepsilon > 0$ , there exists  $c > 0$  such that for all sufficiently large  $N$ , there exists a collection  $\mathcal{C}$  of digraphs on vertex set  $[N]$  with the following properties.*

- (a) *For every  $H$ -free digraph  $I$  on  $[N]$  there exists  $G \in \mathcal{C}$  such that  $I \subseteq G$ .*
- (b) *Every digraph  $G \in \mathcal{C}$  contains at most  $\varepsilon N^h$  copies of  $H$ , and  $e_a(G) \leq \text{ex}_a(N, H) + \varepsilon N^2$ .*
- (c)  $\log |\mathcal{C}| \leq c N^{2-1/m(H)} \log N$ .

Note that in (a), since  $I, G$  are labelled digraphs,  $I \subseteq G$  means that  $I$  is contained in  $G$  in the labelled sense, i.e. the copy of  $I$  in  $G$  has the same vertex labels as  $I$ .

The following corollary is a straightforward consequence of Theorem 4.3.3. We will not use it elsewhere in the chapter, but we include it to illustrate what one can achieve even just with a ‘direct’ application of Theorem 4.3.3. It would be interesting to obtain a version of Corollary 4.3.4 for general forbidden digraphs  $H$ .

**Corollary 4.3.4.** *For every oriented graph  $H$  with  $e(H) \geq 2$ , we have that  $f(n, H) = 2^{\text{ex}_{\log 3}(n, H) + o(n^2)}$  and  $f^*(n, H) = 2^{\text{ex}_2(n, H) + o(n^2)}$ .*

*Proof.* We only prove the first part here; the proof of the second part is almost identical. Clearly  $f(n, H) \geq 2^{\text{ex}_{\log 3}(n, H)}$ . By Theorem 4.3.3, for every  $\varepsilon > 0$  there is a collection  $\mathcal{C}$  of digraphs on  $[n]$  satisfying properties (a)–(c). We know that every digraph  $G \in \mathcal{C}$  contains  $2^{e_{\log 3}(G)}$  oriented subgraphs. Since each  $H$ -free oriented graph is contained in some  $G \in \mathcal{C}$ , and  $|\mathcal{C}| \leq 2^{cn^{2-1/m(H)} \log n}$ ,

$$f(n, H) \leq \sum_{G \in \mathcal{C}} 2^{e_{\log 3}(G)} \leq 2^{\text{ex}_{\log 3}(n, H) + \varepsilon n^2 + o(n^2)}.$$

We are done by letting  $\varepsilon \rightarrow 0$ . □

The proof of Theorem 4.3.3 is similar to that of [70, Theorem 1.3]. We first define the hypergraph  $D(N, H)$ , which will play the role of  $G$  in Theorem 4.3.2.

**Definition 4.3.5.** *Let  $H$  be an oriented graph, let  $r := e(H)$  and let  $N \in \mathbb{N}$ . The  $r$ -graph  $D(N, H)$  has vertex set  $U = ([N] \times [N]) \setminus \{(i, i) : i \in [N]\}$ , where  $B \in U^{(r)}$  is an edge whenever  $B$ , considered as a digraph with vertices in  $[N]$ , is isomorphic to  $H$ .*

We wish to apply Theorem 4.3.2 to  $D(N, H)$ . To do this we require an upper bound on  $\delta(D(N, H), \tau)$  for some suitable value of  $\tau$ . We give one in the following lemma, the proof

of which is identical to that of [70, Lemma 9.2] and is therefore omitted here.

**Lemma 4.3.6.** *Let  $H$  be an oriented graph with  $r := e(H) \geq 2$ , and let  $\gamma \leq 1$ . For  $N$  sufficiently large,  $\delta(D(N, H), \gamma^{-1}N^{-1/m(H)}) \leq r2^{r^2}v(H)!^2\gamma$ .*

We now state a supersaturation result, which we will use to bound the number of edges in containers. It is the digraph analogue of the well-known supersaturation result of Erdős and Simonovits [33]. Its proof is almost the same, and is omitted here.

**Lemma 4.3.7 (Supersaturation).** *Let  $H$  be a digraph on  $h$  vertices, and let  $a \in \mathbb{R}$  with  $a \geq 1$ . For any  $\varepsilon > 0$ , there exists  $\delta > 0$  such that the following holds for all sufficiently large  $n$ . For any digraph  $G$  on  $n$  vertices, if  $G$  contains at most  $\delta n^h$  copies of  $H$ , then  $e_a(G) \leq \text{ex}_a(n, H) + \varepsilon n^2$ .*

We may now apply Theorem 4.3.2 to  $D(N, H)$ , using Lemmas 4.3.6 and 4.3.7, to obtain Theorem 4.3.3. The details of this are identical to the proof of Theorem 1.3 in [70] and are omitted here.

## 4.4 Rough structure of typical $T_{k+1}$ -free oriented graphs and digraphs

In this section we prove a stability result for  $T_{k+1}$ -free digraphs. We apply this (together with Theorem 4.3.3) at the end of this section to determine the ‘rough’ structure of typical  $T_{k+1}$ -free oriented graphs and digraphs.

The *Turán graph*  $\text{Tu}_k(n)$  is the largest complete  $k$ -partite graph on  $n$  vertices (thus each vertex class has  $\lfloor n/k \rfloor$  or  $\lceil n/k \rceil$  vertices). Let  $t_k(n) := e(\text{Tu}_k(n))$ . Let  $DT_k(n)$  be the

digraph obtained from  $Tu_k(n)$  by replacing each edge of  $Tu_k(n)$  by two edges of opposite directions. Obviously, for all  $k \in \mathbb{N}$ ,  $DT_k(n)$  is  $T_{k+1}$ -free so  $\text{ex}_a(n, T_{k+1}) \geq e_a(DT_k(n)) = a \cdot t_k(n)$ . In Lemma 4.4.1 below we show that  $DT_k(n)$  is the unique extremal digraph for  $T_{k+1}$ . This result is not needed for any of our proofs, but we believe it is of independent interest, in addition to being useful for illustrating the general proof method of Lemma 4.4.3. Note that the case  $a = 2$  of Lemma 4.4.1 is already due to Brown and Harary [17].

**Lemma 4.4.1.** *Let  $a \in \mathbb{R}$  with  $3/2 < a \leq 2$  and let  $k, n \in \mathbb{N}$ . Then  $\text{ex}_a(n, T_{k+1}) = a \cdot t_k(n)$ , and  $DT_k(n)$  is the unique extremal  $T_{k+1}$ -free digraph on  $n$  vertices.*

*Proof.* Note that  $DT_k(n)$  is the unique  $k$ -partite digraph  $D$  on  $n$  vertices which maximises  $e_a(D)$ . Moreover,  $e_a(DT_k(n)) = a \cdot t_k(n)$ . Thus, it suffices to show that for every non- $k$ -partite  $T_{k+1}$ -free digraph  $G$  on  $n$  vertices, there exists a  $k$ -partite digraph  $H$  on the same vertex set such that  $e_a(G) < e_a(H)$ .

We prove this by induction on  $k$ . This is trivial for the base case  $k = 1$ , as the only  $T_2$ -free digraph is the empty graph. Suppose that  $k > 1$  and that the claim holds for  $k - 1$ . Let  $G = (V, E)$  be a non- $k$ -partite  $T_{k+1}$ -free digraph on  $n$  vertices. Without loss of generality, suppose that  $d^+(x) = \Delta^0(G)$  for some vertex  $x \in V$ . Let  $S := N^+(x)$  and  $T := V \setminus S$ . Since  $G$  is  $T_{k+1}$ -free we have that  $G[S]$  is  $T_k$ -free. By induction hypothesis, either (i) there is a  $(k - 1)$ -partite digraph  $H'$  on  $S$  such that  $e_a(G[S]) < e_a(H')$ , or (ii)  $G[S]$  is  $(k - 1)$ -partite and hence there is trivially a  $(k - 1)$ -partite digraph  $H'$  on  $S$  such that  $e_a(G[S]) = e_a(H')$ . Next we want to replace all the edges inside  $T$  with edges between  $S$  and  $T$  as follows. Suppose  $y, z \in T$  with  $yz \in E$ . Then there are  $y', z' \in S$  with  $yy' \notin E$  and  $z'z \notin E$  otherwise  $d^+(y) \geq |S| + 1$  or  $d^-(z) \geq |S| + 1$ , contradicting the assumption  $\Delta^0(G) = d^+(x) = |S|$ . We now replace  $yz$  with  $yy'$  and  $z'z$ . By the definition of  $e_a(\cdot)$ , and since  $a \leq 2$ , the gain of adding the edges  $yy'$  and  $z'z$  is at least  $2(a - 1)$ , while the loss of removing  $yz$  is at most one. Thus, since  $a > 3/2$ , we have that  $e_a(G)$  increases by  $2a - 3 > 0$ . Note that this procedure does not change the in-degree or out-degree of

any vertex in  $T$ . We repeat this for every edge inside  $T$ . At the end we obtain a digraph  $G'$  with no edge inside  $T$ . We replace  $G'[S] = G[S]$  with the  $(k-1)$ -partite digraph  $H'$  obtained before to obtain a  $k$ -partite digraph  $H$ . We now consider the two cases (i) and (ii) discussed previously. If  $e_a(G[S]) < e_a(H')$  then it is clear that  $e_a(G) < e_a(H)$ , and we are done. Otherwise,  $G[S]$  is  $(k-1)$ -partite. In this case, since  $G$  is not  $k$ -partite, there must be an edge inside  $T$ , so that there exists  $y, z$  as above. Hence  $e_a(G) < e_a(G')$ , and so clearly  $e_a(G) < e_a(H)$  in this case too, as required.  $\square$

We now prove a stability version of Lemma 4.4.1. The proof idea builds on that of Lemma 4.4.1. The proof will also make use of the following proposition, which can be proved by a simple but tedious calculation, which we omit here.

**Proposition 4.4.2.** *Let  $k, n \in \mathbb{N}$  with  $n \geq k \geq 2$  and let  $s > 0$ . Suppose  $G$  is a  $k$ -partite graph on  $n$  vertices in which some vertex class  $A$  satisfies  $|A - n/k| \geq s$ . Then*

$$e(G) \leq t_k(n) - s \left( \frac{s}{2} - k \right).$$

**Lemma 4.4.3 (Stability).** *Let  $a \in \mathbb{R}$  with  $3/2 < a \leq 2$  and let  $k \in \mathbb{N}$ . For any  $\beta > 0$  there exists  $\gamma > 0$  such that the following holds for all sufficiently large  $n$ . If a digraph  $G$  on  $n$  vertices is  $T_{k+1}$ -free, and  $e_a(G) \geq \text{ex}_a(n, T_{k+1}) - \gamma n^2$ , then  $G = DT_k(n) \pm \beta n^2$ .*

*Proof.* Choose  $\gamma$  and  $n_0$  such that  $1/n_0 \ll \gamma \ll \beta$ , and consider any  $n \geq n_0$ . We follow the proof of Lemma 4.4.1, in which we fix a vertex  $x \in V$  with  $d^+(x) = \Delta^0(G)$ , and let  $S := N^+(x)$  and  $T := V \setminus S$ , and proceed by induction on  $k$ . Again, the base case  $k = 1$  is trivial as the only  $T_2$ -free digraph is the empty graph. Let  $m_1$  be the number of edges of  $G[T]$ , and let  $m_2$  be the number of non-edges between  $T$  and  $S$  (in  $G$ ).

Let  $G'$  be the digraph obtained from  $G$  by replacing each edge inside  $T$  with two edges

between  $T$  and  $S$  as in the proof of Lemma 4.4.1. Then

$$e_a(G) \leq e_a(G') - (2a - 3)m_1. \quad (4.4.4)$$

Since there are  $m_2 - 2m_1$  non-edges between  $T$  and  $S$  in  $G'$ , and adding any one of them would increase  $e_a(G')$  by at least  $a - 1$  (since  $a \leq 2$ ), we have that

$$e_a(G') \leq |T||S|a + e_a(G[S]) - (m_2 - 2m_1)(a - 1). \quad (4.4.5)$$

Let

$$m_3 := \text{ex}_a(|S|, T_k) - e_a(G[S]). \quad (4.4.6)$$

Then  $m_3 \geq 0$  because  $G[S]$  is  $T_k$ -free.

Let  $H$  be the  $k$ -partite digraph obtained from  $DT_{k-1}(|S|)$  (on  $S$ ) by adding the vertex set  $T$  together with all the edges (in both directions) between  $S$  and  $T$ . Then

$$e_a(H) = |T||S|a + \text{ex}_a(|S|, T_k). \quad (4.4.7)$$

Altogether, this gives that

$$\begin{aligned} e_a(G) &\stackrel{(4.4.4)}{\leq} e_a(G') - (2a - 3)m_1 \\ &\stackrel{(4.4.5)}{\leq} |T||S|a + e_a(G[S]) - (m_2 - 2m_1)(a - 1) - (2a - 3)m_1 \\ &\stackrel{(4.4.6)}{=} |T||S|a + \text{ex}_a(|S|, T_k) - m_3 - (m_2 - 2m_1)(a - 1) - (2a - 3)m_1 \\ &\stackrel{(4.4.7)}{=} e_a(H) - m_3 - (m_2 - 2m_1)(a - 1) - (2a - 3)m_1. \end{aligned}$$

Let  $s := ||T| - \frac{n}{k}|$ . By Proposition 4.4.2 we have that if  $s \geq 4k$  then  $\text{ex}_a(n, T_{k+1}) \geq$



$a \cdot t_k(n) \geq e_a(H) + as^2/4$ . So if  $s \geq 4k$  then

$$e_a(G) \leq \text{ex}_a(n, T_{k+1}) - \frac{as^2}{4} - m_3 - (m_2 - 2m_1)(a - 1) - (2a - 3)m_1,$$

and if  $s < 4k$  then

$$e_a(G) \leq \text{ex}_a(n, T_{k+1}) - m_3 - (m_2 - 2m_1)(a - 1) - (2a - 3)m_1.$$

In either case, since  $\text{ex}_a(n, T_{k+1}) - \gamma n^2 \leq e_a(G)$  by assumption, we have that  $m_1 \leq \frac{\gamma}{2a-3}n^2$ ,  $m_2 \leq (\frac{\gamma}{a-1} + \frac{2\gamma}{2a-3})n^2$ ,  $m_3 \leq \gamma n^2$ , and  $s^2 \leq 4\gamma n^2/a$ .

Recall that  $e_a(G[S]) = \text{ex}_a(|S|, T_k) - m_3$ . Hence we have by induction hypothesis that, since  $\gamma \ll \beta$  and  $|S| = \Delta^0(G)$  is sufficiently large,  $G[S] = DT_{k-1}(|S|) \pm (\beta/2)|S|^2$ . Note that we can obtain the digraph  $DT_k(n)$  from  $G$  by removing  $m_1$  edges inside  $T$ , adding  $m_2$  edges between  $T$  and  $S$ , changing at most  $(\beta/2)n^2$  edges inside  $S$ , and changing the adjacency of at most  $s$  vertices. Thus

$$G = DT_k(n) \pm (m_1 + m_2 + (\beta/2)n^2 + 2sn),$$

and so we have that  $G = DT_k(n) \pm \beta n^2$ , as required.  $\square$

We also need the Digraph Removal Lemma of Alon and Shapira [6].

**Lemma 4.4.8 (Removal Lemma).** *For any fixed digraph  $H$  on  $h$  vertices, and any  $\gamma > 0$  there exists  $\varepsilon' > 0$  such that the following holds for all sufficiently large  $n$ . If a digraph  $G$  on  $n$  vertices contains at most  $\varepsilon' n^h$  copies of  $H$ , then  $G$  can be made  $H$ -free by deleting at most  $\gamma n^2$  edges.*

We are now ready to combine Theorem 4.3.3 with Lemma 4.4.3 to show that almost all

$T_{k+1}$ -free oriented graphs and almost all  $T_{k+1}$ -free digraphs are almost  $k$ -partite.

**Lemma 4.4.9.** *For every  $k \in \mathbb{N}$  with  $k \geq 2$  and any  $\alpha > 0$  there exists  $\varepsilon > 0$  such that the following holds for all sufficiently large  $n$ .*

- (i) *All but at most  $f(n, T_{k+1})2^{-\varepsilon n^2}$   $T_{k+1}$ -free oriented graphs on  $n$  vertices can be made  $k$ -partite by changing at most  $\alpha n^2$  edges.*
- (ii) *All but at most  $f^*(n, T_{k+1})2^{-\varepsilon n^2}$   $T_{k+1}$ -free digraphs on  $n$  vertices can be made  $k$ -partite by changing at most  $\alpha n^2$  edges.*

*Proof.* We only prove (i) here; the proof of (ii) is almost identical. Let  $a := \log 3$ . Choose  $n_0 \in \mathbb{N}$  and  $\varepsilon, \gamma, \beta > 0$  such that  $1/n_0 \ll \varepsilon \ll \gamma \ll \beta \ll \alpha, 1/k$ . Let  $\varepsilon' := 2\varepsilon$  and  $n \geq n_0$ . By Theorem 4.3.3 (with  $T_{k+1}, n$  and  $\varepsilon'$  taking the roles of  $H, N$  and  $\varepsilon$  respectively) there is a collection  $\mathcal{C}$  of digraphs on vertex set  $[n]$  satisfying properties (a)–(c). In particular, by (a), every  $T_{k+1}$ -free oriented graph on vertex set  $[n]$  is contained in some digraph  $G \in \mathcal{C}$ . Let  $\mathcal{C}_1$  be the family of all those  $G \in \mathcal{C}$  for which  $e_{\log 3}(G) \geq \text{ex}_{\log 3}(n, T_{k+1}) - \varepsilon' n^2$ . Then the number of (labelled)  $T_{k+1}$ -free oriented graphs not contained in some  $G \in \mathcal{C}_1$  is at most

$$|\mathcal{C}| 2^{\text{ex}_{\log 3}(n, T_{k+1}) - \varepsilon' n^2} \leq 2^{-\varepsilon n^2} f(n, T_{k+1}),$$

because  $|\mathcal{C}| \leq 2^{n^2 - \varepsilon'}$ , by (c), and  $f(n, T_{k+1}) \geq 2^{\text{ex}_{\log 3}(n, T_{k+1})}$ . Thus it suffices to show that every digraph  $G \in \mathcal{C}_1$  satisfies  $G = DT_k(n) \pm \alpha n^2$ . By (b), each  $G \in \mathcal{C}_1$  contains at most  $\varepsilon' n^{k+1}$  copies of  $T_{k+1}$ . Thus by Lemma 4.4.8 we obtain a  $T_{k+1}$ -free digraph  $G'$  after deleting at most  $\gamma n^2$  edges from  $G$ . Then  $e_{\log 3}(G') \geq \text{ex}_{\log 3}(n, T_{k+1}) - (\varepsilon' + \gamma)n^2$ . We next apply Lemma 4.4.3 to  $G'$  and derive that  $G' = DT_k(n) \pm \beta n^2$ . As a result, the original digraph  $G$  satisfies  $G = DT_k(n) \pm (\beta + \gamma)n^2$ , and hence  $G = DT_k(n) \pm \alpha n^2$  as required.  $\square$

## 4.5 Exact structure of typical $T_{k+1}$ -free oriented graphs and digraphs

From Section 4.4 we know that a typical  $T_{k+1}$ -free oriented graph is almost  $k$ -partite (and similarly for digraphs). In this section we use this information to show inductively that we can omit the ‘almost’ in this statement (see Lemma 4.5.7 and the proof of Theorem 4.1.2 at the end of this section). Lemma 4.5.7 relies on several simple observations about the typical structure of almost  $k$ -partite oriented graphs and digraphs (see Lemmas 4.5.3, 4.5.5 and 4.5.6).

Recall that  $t_k(n)$  denotes the maximum number of edges in a  $k$ -partite (undirected) graph on  $n$  vertices, i.e. the number of edges in the  $k$ -partite Turán graph on  $n$  vertices. We say that a  $k$ -partition of vertices is *balanced* if the sizes of any two partition classes differ by at most one. Given a  $k$ -partition  $Q$  of  $[n]$  with partition classes  $V_1, \dots, V_k$ , and a graph, oriented graph or digraph  $G = (V, E)$  on vertex set  $[n]$ , and an edge  $e = uv \in E$  with  $u \in V_i$  and  $v \in V_j$ , we call  $e$  a *crossing edge* if  $i \neq j$ . In Lemma 4.5.1 below we give upper and lower bounds on  $T(n, k)$  and  $T^*(n, k)$ , in terms of  $t_k(n)$  (recall that  $T(n, k)$  and  $T^*(n, k)$  were defined in Section 4.2). Lemma 4.5.1 is used in the proof of Theorem 4.1.2.

**Lemma 4.5.1.** *Let  $k \geq 2$ . For sufficiently large  $n$  we have the following:*

- (i)  $\frac{k^n 3^{t_k(n)}}{2k!n^{k-1}} \leq \frac{1}{2k!} \left( \lfloor \frac{n}{k} \rfloor, \dots, \lfloor \frac{n+k-1}{k} \rfloor \right) 3^{t_k(n)} < T(n, k) < k^n 3^{t_k(n)}.$
- (ii)  $\frac{k^n 4^{t_k(n)}}{2k!n^{k-1}} < T^*(n, k) < k^n 4^{t_k(n)}.$

*Proof.* We only prove (i) here; the proof of (ii) is similar. For the upper bound note that  $k^n$  counts the number of ordered  $k$ -partitions of  $[n]$ , and that for each such  $k$ -partition  $Q$  the number of oriented graphs for which every edge is a crossing edge with respect to  $Q$

is at most  $3^{t_k(n)}$ .

For the lower bound we will count the number of (unordered) balanced  $k$ -partitions. Each such  $k$ -partition gives rise to  $3^{t_k(n)}$   $k$ -partite oriented graphs. Since the vertex classes of a balanced  $k$ -partition of  $[n]$  have sizes  $\lfloor \frac{n}{k} \rfloor, \dots, \lfloor \frac{n+k-1}{k} \rfloor$ , the number of such  $k$ -partitions is

$$\frac{1}{k!} \binom{n}{\lfloor \frac{n}{k} \rfloor, \dots, \lfloor \frac{n+k-1}{k} \rfloor}.$$

We now show that for any given balanced  $k$ -partition  $Q$ , almost all  $k$ -partite oriented graphs for which  $Q$  is a  $k$ -partition have no other possible  $k$ -partitions. Given a balanced  $k$ -partition  $Q$  of  $[n]$  with partition classes  $A_1, \dots, A_k$ , consider a random oriented graph where for each potential crossing edge we choose the edge to be either directed one way, directed the other way, or not present, each with probability  $1/3$ , independently. So each  $k$ -partite oriented graph for which  $Q$  is a  $k$ -partition is equally likely to be generated. Given a set of vertices  $A$  in a digraph  $G$ , we define their common out-neighbourhood  $N^+(A) := \bigcap_{v \in A} N_G^+(v)$ . By Theorem 4.2.2 we have that almost all graphs in the probability space satisfy the following:

( $\alpha$ ) whenever  $\ell \leq k$  and  $i \in [k]$  and  $v_1, \dots, v_\ell \in V(G) \setminus A_i$ , we have that

$$|N^+(\{v_1, \dots, v_\ell\}) \cap A_i| \geq (n/k)(1/3)^{\ell+1}.$$

We now claim that if a  $k$ -partite oriented graph  $G$  has  $k$ -partition  $Q$  and satisfies ( $\alpha$ ) then  $Q$  is the unique  $k$ -partition of  $G$ . Indeed, suppose that  $Q'$  is a  $k$ -partition of  $G$  with vertex classes  $A'_1, \dots, A'_k$ . We will show that  $Q' = Q$ . Consider any  $k$  vertices  $v_1, \dots, v_k$  that are such that  $G[\{v_1, \dots, v_k\}]$  is a transitive tournament. Such a set of  $k$  vertices exists by ( $\alpha$ ). Clearly no two of these vertices can be in the same vertex class of  $Q$  or  $Q'$ . Without loss of generality let us assume that  $v_i \in A_i$  and  $v_i \in A'_i$  for every  $i \in [k]$ .

Define  $N_i := N^+(\{v_1, \dots, v_k\} \setminus \{v_i\})$ . Since  $N_i$  is the common out-neighbourhood of  $\{v_1, \dots, v_k\} \setminus \{v_i\}$  it must be that  $N_i$  is a subset of  $A_i$  and a subset of  $A'_i$ . Note that  $Q$  and  $Q'$  agree on all vertices so far assigned to a partition class of  $Q'$ . Now consider any vertex  $w$  not yet assigned to a partition class of  $Q'$ , and suppose  $w \in A_j$  for some  $j \in [k]$ . For every  $i \in [k]$  with  $i \neq j$  we have by  $(\alpha)$  that

$$|N^+(w) \cap N_i| = |N^+(\{w, v_1, \dots, v_k\} \setminus \{v_i\}) \cap A_i| \geq (n/k)(1/3)^{k+1} \geq 1.$$

This together with the previous observation that  $N_i \subseteq A'_i$  implies that  $w \notin A'_i$ . So  $w \in A'_j$ . Since  $w$  was an arbitrary unassigned vertex we have that  $A_i = A'_i$  for every  $i \in [k]$ , and so  $Q = Q'$ , which implies the claim. This completes the proof of the middle inequality in Lemma 4.5.1.

To prove the first inequality note that if  $a_1 + \dots + a_k = n$  then  $\binom{n}{a_1, \dots, a_k}$  is maximised by taking  $a_j := \lfloor \frac{n+j-1}{k} \rfloor$  for every  $j \in [k]$ . This implies that

$$k^n = \sum_{a_1 + \dots + a_k = n} \binom{n}{a_1, \dots, a_k} \leq n^{k-1} \binom{n}{\lfloor \frac{n}{k} \rfloor, \dots, \lfloor \frac{n+k-1}{k} \rfloor},$$

which in turn implies the first inequality in Lemma 4.5.1, and hence completes the proof.  $\square$

For a given oriented graph or digraph  $G$  on vertex set  $[n]$  we call a  $k$ -partition  $Q$  of  $[n]$  *optimal* if the number of non-crossing edges in  $G$  with respect to  $Q$  is at most the number of non-crossing edges in  $G$  with respect to  $Q'$  for every other  $k$ -partition  $Q'$  of  $[n]$ .

Given  $k \geq 2$  and  $\eta > 0$  we define  $F(n, T_{k+1}, \eta)$  to be the set of all labelled  $T_{k+1}$ -free oriented graphs on  $n$  vertices that have at most  $\eta n^2$  non-crossing edges in an optimal  $k$ -partition. We define  $F_Q(n, T_{k+1}, \eta) \subseteq F(n, T_{k+1}, \eta)$  to be the set of all such oriented

graphs for which  $Q$  is an optimal  $k$ -partition. Similarly, we define  $F^*(n, T_{k+1}, \eta)$  to be the set of all labelled  $T_{k+1}$ -free digraphs on  $n$  vertices that have at most  $\eta n^2$  non-crossing edges in an optimal  $k$ -partition, and we define  $F_Q^*(n, T_{k+1}, \eta) \subseteq F^*(n, T_{k+1}, \eta)$  to be the set of all such digraphs for which  $Q$  is an optimal  $k$ -partition. Define

$$f(n, T_{k+1}, \eta) := |F(n, T_{k+1}, \eta)| \quad \text{and} \quad f_Q(n, T_{k+1}, \eta) := |F_Q(n, T_{k+1}, \eta)|,$$

and similarly

$$f^*(n, T_{k+1}, \eta) := |F^*(n, T_{k+1}, \eta)| \quad \text{and} \quad f_Q^*(n, T_{k+1}, \eta) := |F_Q^*(n, T_{k+1}, \eta)|.$$

Then Lemma 4.4.9 implies that for every  $\eta > 0$  there exists  $\varepsilon' > 0$  such that

$$f(n, T_{k+1}) \leq f(n, T_{k+1}, \eta)(1 + 2^{-\varepsilon' n^2}) \quad \text{and} \quad f_Q^*(n, T_{k+1}) \leq f_Q^*(n, T_{k+1}, \eta)(1 + 2^{-\varepsilon' n^2}) \quad (4.5.2)$$

for all sufficiently large  $n$ . (So  $\varepsilon' = 2\varepsilon$ , where  $\varepsilon$  is as given by Lemma 4.4.9.)

Given an oriented graph or digraph  $G$  on vertex set  $V$  and disjoint subsets  $U, U' \subseteq V$  we let  $\vec{e}_G(U, U')$  denote the number of edges in  $E(G)$  directed from vertices in  $U$  to vertices in  $U'$ . For convenience we will sometimes write  $\vec{e}(U, U')$  for  $\vec{e}_G(U, U')$  if this creates no ambiguity. Given  $k \in \mathbb{N}$ ,  $\eta, \mu > 0$ , and a  $k$ -partition  $Q$  of  $[n]$  with vertex classes  $A_1, \dots, A_k$ , we define  $F_Q(n, \eta, \mu)$  (respectively  $F_Q^*(n, \eta, \mu)$ ) to be the set of all labelled oriented (respectively directed) graphs on  $n$  vertices for which  $Q$  is an optimal  $k$ -partition and that satisfy the following:

(F1) the number of non-crossing edges with respect to  $Q$  is at most  $\eta n^2$ ,

(F2) whenever  $U_i \subseteq A_i$  and  $U_j \subseteq A_j$  with  $|U_i|, |U_j| \geq \mu n$  for distinct  $i, j \in [k]$ , we have that  $\vec{e}(U_i, U_j), \vec{e}(U_j, U_i) \geq |U_i||U_j|/6$ ,

(F3)  $||A_i| - n/k| \leq \mu n$  for every  $i \in [k]$ .

Note that property (F2) is similar to the property that the bipartite graph on vertex classes  $A_i, A_j$  whose edges are directed from  $A_i$  to  $A_j$  is  $\mu$ -regular of density at least  $1/6$  (and similarly for edges directed from  $A_j$  to  $A_i$ ) and that the ‘reduced graph’  $R$  that has vertex set  $\{A_1, \dots, A_k\}$  and edges between pairs that are  $\mu$ -regular of density at least  $1/6$  is a complete digraph.

Define  $F_Q(n, T_{k+1}, \eta, \mu)$  to be the set of all oriented graphs in  $F_Q(n, \eta, \mu)$  that are  $T_{k+1}$ -free. Similarly define  $F_Q^*(n, T_{k+1}, \eta, \mu)$  to be the set of all digraphs in  $F_Q^*(n, \eta, \mu)$  that are  $T_{k+1}$ -free. Note that  $F_Q(n, T_{k+1}, \eta, \mu) \subseteq F_Q^*(n, T_{k+1}, \eta, \mu)$ . Define  $f_Q(n, T_{k+1}, \eta, \mu) := |F_Q(n, T_{k+1}, \eta, \mu)|$  and  $f_Q^*(n, T_{k+1}, \eta, \mu) := |F_Q^*(n, T_{k+1}, \eta, \mu)|$ .

The next lemma shows that  $f_Q(n, T_{k+1}, \eta)$  and  $f_Q(n, T_{k+1}, \eta, \mu)$  are asymptotically equal for any  $k$ -partition  $Q$  and suitable parameter values, (and similarly for  $f_Q^*(n, T_{k+1}, \eta)$  and  $f_Q^*(n, T_{k+1}, \eta, \mu)$ ).

**Lemma 4.5.3.** *Let  $k \geq 2$  and let  $0 < \eta, \mu < 1$  be such that  $\mu^2 \geq 24H(\eta)$ . There exists an integer  $n_0 = n_0(\mu, k)$  such that the following hold for all  $n \geq n_0$  and for every  $k$ -partition  $Q$  of  $[n]$ :*

$$(i) \quad f_Q(n, T_{k+1}, \eta) - f_Q(n, T_{k+1}, \eta, \mu) \leq 3^{t_k(n) - \frac{\mu^2 n^2}{100}}.$$

$$(ii) \quad f_Q^*(n, T_{k+1}, \eta) - f_Q^*(n, T_{k+1}, \eta, \mu) \leq 4^{t_k(n) - \frac{\mu^2 n^2}{100}}.$$

*Proof.* We only prove (i) here; the proof of (ii) is similar. We choose  $n_0$  such that  $1/n_0 \ll \mu, 1/k$ . We wish to count the number of  $G \in F_Q(n, T_{k+1}, \eta) \setminus F_Q(n, T_{k+1}, \eta, \mu)$ . Let  $Q$  have vertex classes  $A_1, \dots, A_k$ . The number of ways that at most  $\eta n^2$  non-crossing edges can be placed is at most

$$\binom{n^2}{\leq \eta n^2} \stackrel{(4.2.1)}{\leq} 2^{H(\eta)n^2}.$$

If  $||A_i| - n/k| > \mu n$  for some  $i \in [k]$  then by Proposition 4.4.2 the number of possible crossing edges is at most

$$t_k(n) - \mu n \left( \frac{\mu n}{2} - k \right) \leq t_k(n) - \frac{\mu^2 n^2}{3}.$$

We can conclude that the number of  $G \in F_Q(n, T_{k+1}, \eta) \setminus F_Q(n, T_{k+1}, \eta, \mu)$  that fail to satisfy (F3) is at most

$$2^{H(\eta)n^2} 3^{t_k(n) - \mu^2 n^2 / 3}.$$

Every  $G \in F_Q(n, T_{k+1}, \eta) \setminus F_Q(n, T_{k+1}, \eta, \mu)$  that satisfies property (F3) must fail to satisfy property (F2). For a given choice of at most  $\eta n^2$  non-crossing edges, consider the random oriented graph  $H$  where for each possible crossing edge with respect to  $Q$  we choose the edge to be either directed in one direction, directed in the other direction, or not present, each with probability  $1/3$ , independently. Note that the total number of ways to choose the crossing edges is at most  $3^{t_k(n)}$ , and each possible configuration of crossing edges is equally likely. So an upper bound on the number of  $G \in F_Q(n, T_{k+1}, \eta) \setminus F_Q(n, T_{k+1}, \eta, \mu)$  that fail to satisfy property (F2) is

$$2^{H(\eta)n^2} 3^{t_k(n)} \mathbb{P}(H \text{ on } A_1, \dots, A_k \text{ fails to satisfy (F2)}).$$

Note that the number of choices for  $U_i \subseteq A_i$  and  $U_j \subseteq A_j$  as in property (F2) is at most  $(2^n)^2$  and that  $\mathbb{E}(\vec{e}_H(U_i, U_j)) = |U_i||U_j|/3 \geq \mu^2 n^2 / 3$ . Hence by Theorem 4.2.2 we get that

$$\begin{aligned} \mathbb{P}(H \text{ on } A_1, \dots, A_k \text{ fails to satisfy (F2)}) &\leq (2^n)^2 \exp \left( -\frac{\mathbb{E}(\vec{e}_H(U_i, U_j))}{8} \right) \\ &\leq 2^{2n} \exp \left( -\frac{\mu^2 n^2}{24} \right). \end{aligned}$$



So summing these upper bounds gives us that

$$\begin{aligned} f_Q(n, T_{k+1}, \eta) - f_Q(n, T_{k+1}, \eta, \mu) &\leq 2^{H(\eta)n^2} 3^{t_k(n)} \left( 3^{-\mu^2 n^2/3} + 2^{2n} e^{-\frac{\mu^2 n^2}{24}} \right) \\ &\leq 3^{t_k(n)} 3^{-\frac{\mu^2 n^2}{24} (\log_3 e - \log_3 2)} 2^{2n+1} \leq 3^{t_k(n) - \frac{\mu^2 n^2}{100}}, \end{aligned}$$

where we use that  $\mu^2 \geq 24H(\eta)$  and that  $1/n_0 \ll \mu$ . □

The following proposition allows us to find many disjoint copies of  $T_k$  in any graph in  $F_Q^*(n, \eta, \mu)$ . It will be useful in proving Lemmas 4.5.5 and 4.5.7. We omit the proof, since it amounts to embedding a small oriented subgraph into the  $\mu$ -regular blow-up of a complete digraph which can be done greedily (see e.g. [25, Lemma 7.5.2] for the ‘undirected’ argument).

**Proposition 4.5.4.** *Let  $n, k \in \mathbb{N}$ , let  $\eta, \mu > 0$ , let  $Q$  be a  $k$ -partition of  $[n]$  with vertex classes  $A_1, \dots, A_k$ , and suppose  $G \in F_Q^*(n, \eta, \mu)$ . For every  $i \in [k]$  let  $B_i \subseteq A_i$  with  $|B_i| \geq 12^{k-2} \mu n$ . Let  $\sigma$  be a permutation of  $[k]$ . Then  $G$  contains a copy of  $T_k$  on vertices  $v_1, \dots, v_k$  where for all distinct  $i, j \in [k]$  we have that  $v_i \in B_i$  and that there is an edge from  $v_i$  to  $v_j$  if and only if  $\sigma(i) < \sigma(j)$ .*

We now show that in an optimal partition each vertex is contained in only a small number of non-crossing edges.

**Lemma 4.5.5.** *Let  $n, k \geq 2$ , let  $\eta, \mu > 0$ , let  $Q$  be a  $k$ -partition of  $[n]$  with vertex classes  $A_1, \dots, A_k$ , and suppose  $G \in F_Q^*(n, T_{k+1}, \eta, \mu)$ . Then for every  $i \in [k]$  and every  $x \in A_i$  we have that*

$$|N_{A_i}^+(x)| + |N_{A_i}^-(x)| \leq 12^{k-2} 2\mu n.$$

*Proof.* Suppose not, so that there exists  $x \in A_i$ , for some  $i \in [k]$ , such that  $|N_{A_i}^+(x)| +$

$|N_{A_i}^-(x)| > 12^{k-2}2\mu n$ . Since  $Q$  is an optimal  $k$ -partition of  $G$ , it must be that

$$|N_{A_j}^+(x)| + |N_{A_j}^-(x)| \geq |N_{A_i}^+(x)| + |N_{A_i}^-(x)| > 12^{k-2}2\mu n$$

for every  $j \in [k]$ .

For every  $j \in [k]$  define  $B_j$  to be  $N_{A_j}^+(x)$  if  $|N_{A_j}^+(x)| \geq |N_{A_j}^-(x)|$ , and  $N_{A_j}^-(x)$  otherwise. So  $|B_j| \geq 12^{k-2}2\mu n$ . Let  $J^+$  be the set of all  $j \in [k]$  such that  $B_j = N_{A_j}^+(x)$ , and let  $J^- := [k] \setminus J^+$ . Fix a permutation  $\sigma$  of  $[k]$  with the property that  $\sigma(i) < \sigma(j)$  whenever  $i \in J^-$  and  $j \in J^+$ . Now Proposition 4.5.4 implies that  $G$  contains a copy of  $T_k$  on vertices  $v_1, \dots, v_k$  where for all distinct  $i, j \in [k]$  we have that  $v_i \in B_i$  and that the edge between  $v_i$  and  $v_j$  is directed towards  $v_j$  if and only if  $\sigma(i) < \sigma(j)$ . By the definition of  $\sigma$ ,  $x$  together with this copy of  $T_k$  forms a copy of  $T_{k+1}$ . This is a contradiction, since  $G \in F_Q^*(n, T_{k+1}, \eta, \mu)$ , and so this completes the proof.  $\square$

The following result shows that an optimal partition of a graph does not change too much upon the removal of just two vertices from the graph.

**Lemma 4.5.6.** *Let  $k \geq 2$ , and let  $0 < \mu < 1/(3k^2)^{12}$  and  $0 < \eta < \mu^2/3$ . There exists an integer  $n_0 = n_0(\mu, k)$  such that the following holds for all  $n \geq n_0$ . Let  $Q$  be a partition of  $[n]$  with vertex classes  $A_1, \dots, A_k$  and let  $x, y$  be distinct elements of  $A_1$ . Then there is a set  $\mathcal{P}$  of  $k$ -partitions of  $[n] \setminus \{x, y\}$ , with  $|\mathcal{P}| \leq e^{\mu^{2/3}n}$ , such that, for every  $G \in F_Q^*(n, T_{k+1}, \eta, \mu)$ , every optimal  $k$ -partition of  $G - \{x, y\}$  is an element of  $\mathcal{P}$ .*

*Proof.* First note that, for any  $G \in F_Q^*(n, T_{k+1}, \eta, \mu)$ , we have by definition that the number of non-crossing edges in  $G$  with respect to  $Q$  is at most  $\eta n^2$ . So certainly the number of non-crossing edges in  $G - \{x, y\}$  with respect to the partition  $A_1 \setminus \{x, y\}, A_2, \dots, A_k$  is at most  $\eta n^2$ .

Consider an arbitrary  $k$ -partition  $B_1, \dots, B_k$  of  $[n] \setminus \{x, y\}$ . We claim that if there exists  $i \in [k]$  and distinct  $j, j' \in [k]$  such that  $|A_j \cap B_i|, |A_{j'} \cap B_i| \geq \mu n$ , then for any  $G \in F_Q^*(n, T_{k+1}, \eta, \mu)$  the number of non-crossing edges in  $G - \{x, y\}$  with respect to the partition  $B_1, \dots, B_k$  is larger than  $\eta n^2$  (and hence  $B_1, \dots, B_k$  cannot be an optimal  $k$ -partition of  $G - \{x, y\}$ ). Indeed, if we find such  $i, j, j'$  then by (F2) we have that the number of non-crossing edges in  $G - \{x, y\}$  with respect to the partition  $B_1, \dots, B_k$  is at least

$$e_G(B_i) \geq e_G(A_j \cap B_i, A_{j'} \cap B_i) \geq 2 \cdot \frac{1}{6}(\mu n)^2 > \eta n^2,$$

which proves the claim.

We let  $\mathcal{P}$  be the set of all  $k$ -partitions of  $[n] \setminus \{x, y\}$  for which no such  $i, j, j'$  exist. So by the above claim we have that for every  $G \in F_Q^*(n, T_{k+1}, \eta, \mu)$ , every optimal  $k$ -partition of  $G - \{x, y\}$  is an element of  $\mathcal{P}$ . So it remains to show that  $|\mathcal{P}| \leq e^{\mu^{2/3}n}$ . Consider an element of  $\mathcal{P}$  with partition classes  $B_1, \dots, B_k$ . For every  $i \in [k]$ , let  $S_i := \{j : |A_j \cap B_i| \geq \mu n\}$ . Note that for every  $i \in [k]$  we have that  $|S_i| \leq 1$ , by definition of  $\mathcal{P}$ . Note also that  $|A_j| \geq n/k - \mu n > k\mu n$  for every  $j \in [k]$ , and thus for every  $i \in [k]$  we have that  $|S_i| = 1$ . Let  $A'_1 := A_1 \setminus \{x, y\}$  and let  $A'_i := A_i$  for every  $i \in \{2, \dots, k\}$ . So every element of  $\mathcal{P}$  can be obtained by starting with the  $k$ -partition  $A'_1, \dots, A'_k$ , applying a permutation of  $[k]$  to the partition class labels, and then for every ordered pair of partition classes moving at most  $\mu n$  elements from the first partition class to the second. Hence, since  $|A_j| \leq n/k + \mu n \leq 2n/k$ , we have that

$$|\mathcal{P}| \leq k! \left( \binom{2n/k}{\leq \mu n} \right)^{k-1} \leq k! \left( \mu n \left( \frac{2en/k}{\mu n} \right)^{\mu n} \right)^{k(k-1)} \leq k! (\mu n)^{k^2} \left( \frac{1}{\mu^2} \right)^{\mu k^2 n} \leq e^{\mu^{2/3}n},$$

as required. □

Define  $F'_Q(n, T_{k+1}, \eta)$  to be the set of all oriented graphs in  $F_Q(n, T_{k+1}, \eta)$  that have at least

one non-crossing edge with respect to  $Q$ . Define  $f'_Q(n, T_{k+1}, \eta) := |F'_Q(n, T_{k+1}, \eta)|$ . Similarly define  $F_Q^*(n, T_{k+1}, \eta)$  to be the set of all digraphs in  $F_Q^*(n, T_{k+1}, \eta)$  that have at least one non-crossing edge with respect to  $Q$ , and define  $f_Q^{*'}(n, T_{k+1}, \eta) := |F_Q^{*'}(n, T_{k+1}, \eta)|$ . In the following result we use Lemmas 4.4.9, 4.5.3, 4.5.5 and 4.5.6 to give upper bounds on  $f'_Q(n, T_{k+1}, \eta)$  and  $f_Q^{*'}(n, T_{k+1}, \eta)$  for any  $k$ -partition  $Q$  and suitable parameter values.

**Lemma 4.5.7.** *For all  $k \geq 2$  there exist  $\eta > 0$  and  $C \in \mathbb{N}$  such that for all  $n \in \mathbb{N}$  and all  $k$ -partitions  $Q$  of  $[n]$  the following hold.*

$$(i) \quad f'_Q(n, T_{k+1}, \eta) \leq 3^{t_k(n)} C 2^{-\eta n}.$$

$$(ii) \quad f_Q^{*'}(n, T_{k+1}, \eta) \leq 4^{t_k(n)} C 2^{-\eta n}.$$

*Proof.* We only prove (i) here; the proof of (ii) is similar. Choose  $C, n_0 \in \mathbb{N}$  and  $\varepsilon, \eta, \mu > 0$  such that

$$1/C \ll 1/n_0 \ll \varepsilon \ll \eta \ll \mu \ll 1/k.$$

Define  $F_Q(n, T_{k+1})$  to be the set of all  $T_{k+1}$ -free oriented graphs on  $n$  vertices for which  $Q$  is an optimal  $k$ -partition, and define  $f_Q(n, T_{k+1}) = |F_Q(n, T_{k+1})|$ .

The proof proceeds by induction on  $n$ . In fact, in addition to (i) we will inductively show that

$$f_Q(n, T_{k+1}) \leq 3^{t_k(n)} (1 + C 2^{-\eta n}). \quad (4.5.8)$$

The result holds trivially for  $n < n_0$  since  $1/C \ll 1/n_0$ . So let  $n \geq n_0$  and let us assume that for every  $k$ -partition  $Q'$  of  $[n-2]$  we have that

$$f_{Q'}(n-2, T_{k+1}) \leq 3^{t_k(n-2)} (1 + C 2^{-\eta(n-2)}). \quad (4.5.9)$$

Let  $Q$  have partition classes  $A_1, \dots, A_k$ . Define  $F'_Q(n, T_{k+1}, \eta, \mu)$  to be the set of all

graphs in  $F_Q(n, T_{k+1}, \eta, \mu)$  that have at least one non-crossing edge with respect to  $Q$ . Define  $f'_Q(n, T_{k+1}, \eta, \mu) := |F'_Q(n, T_{k+1}, \eta, \mu)|$ . We will first find an upper bound for  $f'_Q(n, T_{k+1}, \eta, \mu)$ . We will find this bound in four steps. Note that (F3) implies that  $f'_Q(n, T_{k+1}, \eta, \mu) = 0$  unless  $||A_i| - n/k| \leq \mu n$ , so we may assume that this inequality holds.

**Step 1:** Let  $B_1$  be the number of ways to choose a single non-crossing edge  $xy$  with respect to  $Q$ . Then  $B_1 \leq n^2$ . Let  $A_i$  be the partition class of  $Q$  containing  $x$  and  $y$ .

**Step 2:** Let  $B_2$  be the number of ways to choose the edges that do not have an endpoint in  $\{x, y\}$ . By Lemma 4.5.6 there is a set  $\mathcal{P}$  of  $k$ -partitions of  $[n] \setminus \{x, y\}$ , with  $|\mathcal{P}| \leq e^{\mu^{2/3}n}$ , such that, for every  $G \in F_Q(n, T_{k+1}, \eta, \mu)$ , every optimal  $k$ -partition of  $G - \{x, y\}$  is an element of  $\mathcal{P}$ . So we have by our inductive hypothesis that

$$B_2 \leq \sum_{Q' \in \mathcal{P}} f_{Q'}(n-2, T_{k+1}) \stackrel{(4.5.9)}{\leq} e^{\mu^{2/3}n} 3^{t_k(n-2)} (1 + C2^{-\eta(n-2)}) \leq 3^{t_k(n-2)} C e^{\mu^{1/2}n}.$$

**Step 3:** Let  $B_3$  be the number of possible ways to construct the edges between  $x, y$  and the vertices outside  $A_i$ . Let  $U$  be the set of edges chosen in Step 2. One can view  $U$  as a subset of the edge set of a graph  $G$  in  $F'_Q(n, T_{k+1}, \eta, \mu)$ . Let  $U'$  be the subset of  $U$  consisting of all those edges in  $U$  that do not have an endpoint in  $A_i$ . So  $U'$  can be viewed as the edge set of a subgraph  $G'$  of  $G$  with  $G' \in F_{\tilde{Q}}(n - |A_i|, 5\eta, 3\mu)$ , where the set of partition classes of  $\tilde{Q}$  is  $\{A_1, \dots, A_k\} \setminus \{A_i\}$ . By repeatedly applying Proposition 4.5.4 to  $G'$  we can find at least  $n/k - \mu n - 12^{k-3}3\mu n$  vertex-disjoint copies of  $T_{k-1}$  in  $G'$ , each with precisely one vertex in each of the  $A_j$  for  $j \neq i$ . Consider the  $2(k-1)$  potential edges between  $x, y$  and the vertices of any such  $T_{k-1}$ . If we wish for our graph to remain  $T_{k+1}$ -free then not all of the possible  $3^{2(k-1)}$  sets of such edges are allowed. So since the number of vertices outside

$A_i$  not contained in one of these  $T_{k-1}$  is at most  $(k-1)(2\mu n + 12^{k-3}3\mu n) \leq \mu^{1/2}n \log_3 e/2$ , we have that

$$B_3 \leq (3^{2(k-1)} - 1)^{n/k} 3^{2(\mu^{1/2}n \log_3 e/2)} < (3^{2(k-1)} (1 - 3^{-2k}))^{n/k} e^{\mu^{1/2}n} \leq 3^{2\frac{k-1}{k}n} e^{-\frac{n}{9^k k}} e^{\mu^{1/2}n}.$$

**Step 4:** Let  $B_4$  be the number of possible ways to construct the edges between  $x, y$  and the other vertices in  $A_i$ . Note that by Lemma 4.5.5,  $x$  and  $y$  each have at most  $12^{k-2}2\mu n$  neighbours inside  $A_i$ , and for each of these the edge between them may be oriented in either direction. So since  $|A_i| \leq n$ , we have that

$$\begin{aligned} B_4 &\leq \left( \binom{n}{\leq 12^{k-2}2\mu n} \right)^2 (2^{12^{k-2}2\mu n})^2 \leq \left( \binom{n}{12^{k-2}2\mu n} \right)^2 (2^{12^{k-2}2\mu n})^2 (2^{12^{k-2}2\mu n})^2 \\ &\leq \left( \frac{2e}{12^{k-2}\mu} \right)^{12^{k-2}4\mu n} \leq e^{\mu^{1/2}n}. \end{aligned}$$

In Steps 1–4 we have considered all possible edges, and so  $f'_Q(n, T_{k+1}, \eta, \mu) \leq B_1 \cdot B_2 \cdot B_3 \cdot B_4$ . Together with the fact that  $t_k(n) \geq t_k(n-2) + 2((k-1)/k)(n-2)$  this implies that

$$\begin{aligned} f'_Q(n, T_{k+1}, \eta, \mu) &\leq n^2 3^{t_k(n-2)} C 3^{2((k-1)/k)n} e^{3\mu^{1/2}n} e^{-n/(k9^k)} \\ &\leq 3^{t_k(n)} C e^{-n/(2k9^k)} \leq 3^{t_k(n)} C 2^{-3\eta n}. \end{aligned} \tag{4.5.10}$$

Now, note that since  $F_Q(n, T_{k+1}, \eta, \mu), F'_Q(n, T_{k+1}, \eta) \subseteq F_Q(n, T_{k+1}, \eta)$  we have that

$$\begin{aligned} f'_Q(n, T_{k+1}, \eta) &= |F'_Q(n, T_{k+1}, \eta) \cap F_Q(n, T_{k+1}, \eta, \mu)| + |F'_Q(n, T_{k+1}, \eta) \setminus F_Q(n, T_{k+1}, \eta, \mu)| \\ &= |F'_Q(n, T_{k+1}, \eta, \mu)| + |F'_Q(n, T_{k+1}, \eta) \setminus F_Q(n, T_{k+1}, \eta, \mu)| \\ &\leq |F'_Q(n, T_{k+1}, \eta, \mu)| + |F_Q(n, T_{k+1}, \eta) \setminus F_Q(n, T_{k+1}, \eta, \mu)| \\ &= f'_Q(n, T_{k+1}, \eta, \mu) + (f_Q(n, T_{k+1}, \eta) - f_Q(n, T_{k+1}, \eta, \mu)). \end{aligned}$$

This together with Lemma 4.5.3(i) gives us that

$$f'_Q(n, T_{k+1}, \eta) \leq f'_Q(n, T_{k+1}, \eta, \mu) + 3^{t_k(n) - \frac{\mu^2 n^2}{100}} \stackrel{(4.5.10)}{\leq} 3^{t_k(n)} C 2^{-\eta n},$$

which proves (i). So it remains to prove (4.5.8).

Note that the number of graphs in  $F_Q(n, T_{k+1}, \eta, \mu)$  for which every edge is a crossing edge with respect to  $Q$  is at most  $3^{t_k(n)}$ . Since  $f_Q(n, T_{k+1}, \eta, \mu) - f'_Q(n, T_{k+1}, \eta, \mu)$  is precisely the number of such graphs, we have that

$$f_Q(n, T_{k+1}, \eta, \mu) - f'_Q(n, T_{k+1}, \eta, \mu) \leq 3^{t_k(n)}.$$

This together with (4.5.10) implies that

$$f_Q(n, T_{k+1}, \eta, \mu) \leq 3^{t_k(n)} (1 + C 2^{-3\eta n}).$$

Together with Lemma 4.5.3(i) this implies that

$$f_Q(n, T_{k+1}, \eta) \leq f_Q(n, T_{k+1}, \eta, \mu) + 3^{t_k(n) - \frac{\mu^2 n^2}{100}} \leq 3^{t_k(n)} (1 + C 2^{-2\eta n}). \quad (4.5.11)$$

On the other hand, Lemma 4.4.9(i) implies that

$$\begin{aligned} f(n, T_{k+1}) - f(n, T_{k+1}, \eta) &\leq f(n, T_{k+1}) 2^{-\varepsilon n^2} \stackrel{(4.5.2)}{\leq} 2f(n, T_{k+1}, \eta) 2^{-\varepsilon n^2} \\ &\stackrel{(4.5.11)}{\leq} 2k^n 3^{t_k(n)} (1 + C 2^{-2\eta n}) 2^{-\varepsilon n^2} \leq 3^{t_k(n)} C 2^{-2\eta n}. \end{aligned}$$

Now this together with (4.5.11) implies that

$$f_Q(n, T_{k+1}) \leq f_Q(n, T_{k+1}, \eta) + (f(n, T_{k+1}) - f(n, T_{k+1}, \eta)) \leq 3^{t_k(n)} (1 + C 2^{-\eta n}).$$

This completes the proof of (4.5.8). □

We can now finally prove Theorem 4.1.2 using Lemmas 4.5.1 and 4.5.7 together with the bounds in (4.5.2).

**Proof of Theorem 4.1.2.** We only prove (i) here; the proof of (ii) is similar. Let  $\eta, C$  be given by Lemma 4.5.7 and choose  $n_0$  and  $\varepsilon$  such that  $1/n_0 \ll \varepsilon \ll \eta, 1/k$ . Consider any  $n \geq n_0$  and let  $\mathcal{Q}$  be the set of all  $k$ -partitions of  $[n]$ . Note that

$$f'(n, T_{k+1}, \eta) \leq \sum_{Q \in \mathcal{Q}} f'_Q(n, T_{k+1}, \eta).$$

So by Lemma 4.5.7(i) and the fact that  $|\mathcal{Q}| \leq k^n$  we have that

$$f'(n, T_{k+1}, \eta) \leq k^n 3^{t_k(n)} C 2^{-\eta n}.$$

Recall from (4.5.2) that

$$f(n, T_{k+1}) \leq f(n, T_{k+1}, \eta)(1 + 2^{-\varepsilon n^2}).$$

Together with the fact that  $f(n, T_{k+1}, \eta) = f'(n, T_{k+1}, \eta) + T(n, k)$  and the upper bound in Lemma 4.5.1(i), this implies that

$$\begin{aligned} f(n, T_{k+1}) - T(n, k) &\leq f'(n, T_{k+1}, \eta) + f(n, T_{k+1}, \eta) 2^{-\varepsilon n^2} \\ &= f'(n, T_{k+1}, \eta) + (f'(n, T_{k+1}, \eta) + T(n, k)) 2^{-\varepsilon n^2} \\ &\leq k^n 3^{t_k(n)} C 2^{-\eta n} + k^n 3^{t_k(n)} (1 + C 2^{-\eta n}) 2^{-\varepsilon n^2}. \end{aligned}$$

Now the lower bound in Lemma 4.5.1(i) gives us that  $f(n, T_{k+1}) - T(n, k) = o(T(n, k))$ . So  $f(n, T_{k+1}) = T(n, k)(1 + o(1))$ , as required. □



## 4.6 Rough structure of typical $C_k$ -free oriented graphs and digraphs

In this section we prove several stability results for  $C_k$ -free digraphs (Lemmas 4.6.5, 4.6.13 and 4.6.17). These are used (together with Theorem 4.3.3) at the end of the section to determine the ‘rough’ structure of typical  $C_k$ -free oriented graphs and digraphs.

We will make use of the following definitions. For disjoint sets of vertices  $A, A'$ , we define  $\vec{K}(A, A')$  to be the oriented graph on vertex set  $A \cup A'$  with edge set consisting of all the  $|A||A'|$  edges that are directed from  $A$  to  $A'$ . Given a digraph  $G$ ,  $A \subseteq V(G)$  and  $x \in V(G) \setminus A$ , we say that  $G[A, \{x\}]$  is an *in-star* if  $G[A, \{x\}] = \vec{K}(A, \{x\})$ , and we say that  $G[A, \{x\}]$  is an *out-star* if  $G[A, \{x\}] = \vec{K}(\{x\}, A)$ . The following proposition will prove useful to us many times in this section.

**Proposition 4.6.1.** *Let  $a \in \mathbb{R}$  with  $1 \leq a \leq 2$ , let  $k \in \mathbb{N}$  with  $k \geq 2$  and let  $G$  be a  $C_{k+1}$ -free digraph. Suppose  $G$  contains a copy  $C$  of  $C_k$  with vertex set  $A \subseteq V(G)$ , and let  $x \in V(G) \setminus A$ . Then the following hold:*

- (i)  $e_a(A, \{x\}) \leq k$ ,
- (ii) if  $G[A, \{x\}]$  is not an in-star or an out-star, then  $e_a(A, \{x\}) \leq k - 2 + a$ ,
- (iii) if  $G[A, \{x\}]$  is not an in-star or an out-star, and contains no double edges, then  $e_a(A, \{x\}) \leq k - 1$ .

Suppose moreover that for some  $\ell \in \{k - 1, k\}$ ,  $G$  contains a copy  $C'$  of  $C_\ell$  with vertex set  $A' \subseteq V(G)$ , where  $A \cap A' = \emptyset$ . Then the following hold:

(iv)  $e_a(A, A') \leq k\ell$ ,

(v) if  $G[A, A'] \notin \{\vec{K}(A, A'), \vec{K}(A', A)\}$ , then  $e_a(A, A') \leq k\ell - 2 + a$ ,

(vi) if  $G[A, A'] \notin \{\vec{K}(A, A'), \vec{K}(A', A)\}$ , and moreover  $G[A, A']$  contains no double edges, then  $e_a(A, A') \leq k\ell - 1$ .

*Proof.* Write  $C = v_1v_2 \dots v_k$ . For  $i \in [k]$  let  $Q_i := \{v_ix, xv_{i+1}\}$ , where  $v_{k+1} := v_1$ . Since  $G$  is  $C_{k+1}$ -free we have that  $|E(G) \cap Q_i| \leq 1$  for every  $i \in [k]$ . Hence  $e(A, \{x\}) \leq k$ . We can now prove (i)–(vi).

(i) This follows since  $e_a(A, \{x\}) \leq e(A, \{x\}) \leq k$ .

(ii) Suppose that  $G[A, \{x\}]$  is not an in-star or an out-star. Note that if for some  $j \in [k]$  we have that  $E(G) \cap Q_j = \emptyset$  then, since  $|E(G) \cap Q_i| \leq 1$  for all  $i \in [k]$ ,  $e_a(A, \{x\}) \leq e(A, \{x\}) \leq k - 1 \leq k - 2 + a$  as required. So we may assume that  $|E(G) \cap Q_i| = 1$  for every  $i \in [k]$ . Since  $G[A, \{x\}]$  is not an in-star or an out-star, there exists some  $j \in [k]$  such that  $E(G) \cap Q_j = \{xv_{j+1}\}$  and  $E(G) \cap Q_{j+1} = \{v_{j+1}x\}$ ; that is, there exists a double edge in  $G[A, \{x\}]$ . So since  $e(G[A, \{x\}]) \leq k$  we have that  $e_a(A, \{x\}) \leq k - 2 + a$ , as required.

(iii) Suppose that  $G[A, \{x\}]$  is not an in-star or an out-star, and contains no double edges. Just as in the proof of (ii), we have that if  $|E(G) \cap Q_i| = 1$  for every  $i \in [k]$  then there exists a double edge in  $G[A, \{x\}]$ . So we may assume that for some  $j \in [k]$  we have that  $E(G) \cap Q_j = \emptyset$ . This implies that  $e_a(A, \{x\}) \leq e(A, \{x\}) \leq k - 1$ , as required.

(iv) This is immediate from (i).

(v) Suppose that  $G[A, A'] \notin \{\vec{K}(A, A'), \vec{K}(A', A)\}$ . We claim that there exists  $x \in A'$  such that  $G[A, \{x\}]$  is not an in-star or an out-star. Indeed, suppose not. Then since

$G[A, A'] \notin \{\vec{K}(A, A'), \vec{K}(A', A)\}$ , there must exist distinct vertices  $y', z' \in A'$  such that  $G[A, \{y'\}]$  is an in-star and  $G[A, \{z'\}]$  is an out-star. Let  $P'$  be the subpath of  $C'$  from  $y'$  to  $z'$ . Thus  $P'$  has length  $s$  for some  $1 \leq s \leq k-1$ . Let  $y, z \in A$  be not necessarily distinct vertices such that the subpath  $P$  of  $C$  from  $y$  to  $z$  has length  $k-s-1$ . Then  $yPzy'P'z'y$  is a copy of  $C_{k+1}$  in  $G$ , which contradicts the assumption that  $G$  is  $C_{k+1}$ -free. Hence there does exist  $x \in A$  such that  $G[A, \{x\}]$  is not an in-star or an out-star. So by (i) and (ii) we have that  $e_a(A, A') \leq k\ell - 2 + a$  as required.

(vi) This proof is almost identical to that of (v), just using (iii) instead of (ii), and so is omitted.

□

For  $k \in \mathbb{N}$  define  $T_{n,k}^+$  (up to isomorphism) to be the digraph on vertex set  $[n]$  with all edges  $ij$  where  $i < j$  and all edges  $ji$  where  $i < j$  and  $\lfloor (i-1)/k \rfloor = \lfloor (j-1)/k \rfloor$ . So if  $n = sk$  for some  $s \in \mathbb{N}$  then  $T_{n,k}^+$  is obtained from  $T_s$  by blowing up each vertex to a copy of the complete digraph  $DK_k$ . Note that  $T_{n,k}^+$  is  $C_{k+1}$ -free, and for all  $a \in \mathbb{R}$  with  $1 \leq a \leq 2$  we have that

$$e_a(T_{n,k}^+) = \binom{n}{2} + \left\lfloor \frac{n}{k} \right\rfloor \binom{k}{2} (a-1) + \binom{n-k \lfloor \frac{n}{k} \rfloor}{2} (a-1). \quad (4.6.2)$$

We will first show that  $T_{n,k}^+$  is an extremal digraph for  $C_{k+1}$ . The resulting formula for  $\text{ex}_a(n, C_{k+1})$  will be used in the proofs of Lemmas 4.6.5 and 4.6.13, but we will not refer to  $T_{n,k}^+$  itself again. Note that the case  $a = 2$  of Lemma 4.6.3 corresponds to finding the digraph Turán number of  $C_{k+1}$ , and is already due to Häggkvist and Thomassen [36].

**Lemma 4.6.3.** *Let  $a \in \mathbb{R}$  with  $1 \leq a \leq 2$  and let  $k \in \mathbb{N}$ . Then*

$$\text{ex}_a(n, C_{k+1}) = e_a(T_{n,k}^+).$$

*Proof.* The proof proceeds by induction on  $k$ . The case  $k = 1$  is trivial. So suppose  $G$  is a  $C_{k+1}$ -free digraph on  $n$  vertices for some  $k > 1$ . The proof now proceeds by induction on  $n$ . The cases  $n = 1, \dots, k$  are trivial. So suppose  $n > k$ . Note that if  $G$  is also  $C_k$ -free then by our inductive hypothesis on  $k$ ,

$$e_a(G) \leq e_a(T_{n,k-1}^+) \leq e_a(T_{n,k}^+).$$

Otherwise,  $G$  contains a copy of  $C_k$ , say on vertex set  $A \subseteq V(G)$ . So by Proposition 4.6.1(i) we have that  $e_a(A, \{x\}) \leq k$  for every  $x \in V(G) \setminus A$ . Hence  $e_a(A, V(G) \setminus A) \leq k(n-k)$ . Note that by our inductive hypothesis on  $n$ ,  $e_a(G[V(G) \setminus A]) \leq e_a(T_{n-k,k}^+)$ . Hence,

$$\begin{aligned} e_a(G) &= e_a(G[V(G) \setminus A]) + e_a(A, V(G) \setminus A) + e_a(G[A]) \\ &\stackrel{(4.6.2)}{\leq} \binom{n-k}{2} + \left( \left\lfloor \frac{n-k}{k} \right\rfloor \binom{k}{2} + \binom{(n-k) - k \left\lfloor \frac{n-k}{k} \right\rfloor}{2} \right) (a-1) + k(n-k) + a \binom{k}{2} \\ &= \binom{n}{2} + \left\lfloor \frac{n}{k} \right\rfloor \binom{k}{2} (a-1) + \binom{n - k \left\lfloor \frac{n}{k} \right\rfloor}{2} (a-1) = e_a(T_{n,k}^+). \end{aligned}$$

So indeed  $\text{ex}_a(n, C_{k+1}) = e_a(T_{n,k}^+)$ , as required.  $\square$

We will next prove three stability results for  $C_{k+1}$ -free digraphs. The first will cover the case  $1 \leq a < 2$  (where  $a$ , as usual, is the parameter in the definition of the weighted size of a digraph) and will be used to prove a structural result on  $C_{k+1}$ -free oriented graphs. The second covers the case  $a = 2$  and  $k$  odd, and will be used to prove an analogous structural result on  $C_{k+1}$ -free digraphs for odd  $k$ . The third covers the case  $a = 2$  and  $k$  even, and will be used to prove a (less restrictive) structural result on  $C_{k+1}$ -free digraphs

for even  $k$ . The proofs of the first two of these stability results will make use of a result of Chudnovsky, Seymour and Sullivan [22], which we state below. To do so we need to introduce the following notation. Let  $\beta(G)$  denote the number of backwards edges in  $G$  with respect to a transitive-optimal ordering of  $G$ . Let  $\gamma(G)$  denote the number of unordered non-adjacent pairs of vertices in  $G$ ; that is unordered pairs  $u, v$  of vertices such that  $uv \notin E(G)$  and  $vu \notin E(G)$ .

**Lemma 4.6.4.** [22] *Let  $G$  be a  $\{C_2, C_3\}$ -free digraph. Then  $\beta(G) \leq \gamma(G)$ .*

It is conjectured in [22] that in fact  $\beta(G) \leq \gamma(G)/2$  for all  $\{C_2, C_3\}$ -free digraphs  $G$ . If true, this would be best possible.

**Lemma 4.6.5 (Stability when  $a < 2$ ).** *Let  $a \in \mathbb{R}$  with  $1 \leq a < 2$  and let  $k \in \mathbb{N}$  with  $k \geq 2$ . Then for all  $\varepsilon > 0$  there exist  $\delta > 0$  and  $n_0 \in \mathbb{N}$  such that every  $C_{k+1}$ -free digraph  $G$  on  $n \geq n_0$  vertices with*

$$e_a(G) \geq \binom{n}{2} - \delta n^2$$

*satisfies  $G = T_n \pm \varepsilon n^2$ .*

*Proof.* We prove the lemma via the following claim.

**Claim:** *Let  $k \in \mathbb{N}$  with  $k \geq 2$  and let  $\varepsilon > 0$ . Suppose that there exist  $\delta' > 0$  and  $n'_0 \in \mathbb{N}$  such that every  $\{C_k, C_{k+1}\}$ -free digraph  $G$  on  $n' \geq n'_0$  vertices with*

$$e_a(G) \geq \binom{n'}{2} - \delta' n'^2$$

*satisfies  $G = T_{n'} \pm \varepsilon n'^2/(2k^2)$ . Then there exist  $\delta > 0$  and  $n_0 \in \mathbb{N}$  such that every  $C_{k+1}$ -free digraph  $G$  on  $n \geq n_0$  vertices with*

$$e_a(G) \geq \binom{n}{2} - \delta n^2$$

satisfies  $G = T_n \pm \varepsilon n^2$ .

In order to check that the claim implies the lemma, we proceed by induction on  $k$ . For the base case  $k = 2$  the assumption of the claim is satisfied, since if  $\delta' := \varepsilon/(4k^2)$  and if  $G$  is a  $\{C_2, C_3\}$ -free digraph on  $n'$  vertices with  $e_a(G) \geq \binom{n'}{2} - \delta' n'^2$  then  $\gamma(G) \leq \delta' n'^2$ , and so applying Lemma 4.6.4 yields the assumption of the claim. So the conclusion of the claim holds, which is precisely the statement of the lemma for  $k = 2$ . For  $k > 2$  the assumption of the claim is satisfied by the inductive hypothesis (since any  $\{C_k, C_{k+1}\}$ -free digraph is certainly a  $C_k$ -free digraph) and so the conclusion of the claim holds, which is precisely the statement of the lemma for  $k$ . So by induction the lemma holds and we are done.

Thus it remains to prove the claim. (Note that, apart from in the base case  $k = 2$ , in the above argument it would suffice for the assumption in the statement of the claim to refer to  $C_k$ -free digraphs, rather than  $\{C_k, C_{k+1}\}$ -free digraphs. As such, this claim is stronger than strictly necessary for our purposes, since the assumption is weaker. However, this approach allows us to prove the base case at the same time as the inductive step, and so yields a shorter proof.)

**Proof of claim:** Choose  $\delta$  and  $n_0$  such that  $1/n_0 \ll \delta \ll 1/k, 2-a, \delta'$  and  $1/n_0 \ll 1/n'_0, \varepsilon$ . Let  $G$  be a  $C_{k+1}$ -free digraph on  $n \geq n_0$  vertices with

$$e_a(G) \geq \binom{n}{2} - \delta n^2. \quad (4.6.6)$$

Let  $t \geq 0$  denote the maximum number of vertex-disjoint copies of  $C_k$  in  $G$ . Let  $\mathcal{C} = \{C^1, \dots, C^t\}$  be a set of  $t$  vertex-disjoint copies of  $C_k$  in  $G$ . Let  $V_1 := V(C^1) \cup \dots \cup V(C^t)$  and  $V_2 := V(G) \setminus V_1$ . Let  $n_1 := |V_1|$  and  $n_2 := |V_2|$ . Note that  $G[V_2]$  is  $C_k$ -free.

Note that Proposition 4.6.1(i) implies that  $e_a(V_1, V_2) \leq n_1 n_2$ , since  $G$  is  $C_{k+1}$ -free. Also, for  $i = 1, 2$ , since  $G[V_i]$  is  $C_{k+1}$ -free, Lemma 4.6.3 and (4.6.2) together imply that

$$e_a(G[V_i]) \leq \text{ex}_a(n_i, C_{k+1}) = \binom{n_i}{2} + \left\lfloor \frac{n_i}{k} \right\rfloor \binom{k}{2} (a-1) + \binom{n_i - k \left\lfloor \frac{n_i}{k} \right\rfloor}{2} (a-1) \leq \binom{n_i}{2} + \delta n^2.$$

(The last inequality holds since  $1/n_0 \ll \delta \ll 1/k$ .) Together with (4.6.6) this implies that

$$\begin{aligned} e_a(G[V_1]) &= e_a(G) - e_a(V_1, V_2) - e_a(G[V_2]) \\ &\geq \left( \binom{n}{2} - \delta n^2 \right) - n_1 n_2 - \left( \binom{n_2}{2} + \delta n^2 \right) = \binom{n_1}{2} - 2\delta n^2, \end{aligned} \quad (4.6.7)$$

and similarly that

$$e_a(G[V_2]) = e_a(G) - e_a(V_1, V_2) - e_a(G[V_1]) \geq \binom{n_2}{2} - 2\delta n^2, \quad (4.6.8)$$

and that

$$e_a(V_1, V_2) = e_a(G) - e_a(G[V_1]) - e_a(G[V_2]) \geq n_1 n_2 - 3\delta n^2. \quad (4.6.9)$$

We now consider the digraph  $G'$  defined on vertex set  $[t] \cup V_2$  as follows. Firstly,  $G'[V_2] := G[V_2]$ . For vertices  $i, j \in [t]$  we have that  $ij \in E(G')$  if and only if  $G[V(C^i), V(C^j)] = \vec{K}(V(C^i), V(C^j))$ . For a vertex  $x \in V_2$  and an element  $i \in [t]$  we have that  $ix \in E(G')$  if and only if  $G[V(C^i), \{x\}]$  is an in-star and that  $xi \in E(G')$  if and only if  $G[V(C^i), \{x\}]$  is an out-star.

Note that by Proposition 4.6.1(iv),  $e_a(G[V(C^i), V(C^j)]) \leq k^2$  (for all  $i \neq j$ ). Moreover, Proposition 4.6.1(v) implies that if  $i, j \in [t]$  and  $ij, ji \notin E(G'[[t]])$  then  $e_a(G[V(C^i), V(C^j)])$

$\leq k^2 - 2 + a$ . Let  $s := \binom{t}{2} - e_a(G'[[t]]) = \binom{t}{2} - e(G'[[t]])$ . Then

$$\begin{aligned} \binom{n_1}{2} - 2\delta n^2 &\stackrel{(4.6.7)}{\leq} e_a(G[V_1]) = \sum_{\substack{i,j \in [t] \\ i < j}} e_a(G[V(C^i), V(C^j)]) + \sum_{i \in [t]} e_a(G[V(C^i)]) \\ &\leq \binom{n_1}{2} - s(2 - a) + t \binom{k}{2} (a - 1). \end{aligned} \quad (4.6.10)$$

Thus  $s(2 - a) \leq 3\delta n^2$ , i.e.

$$e_a(G'[[t]]) \geq \binom{t}{2} - \frac{3\delta n^2}{2 - a}.$$

Similarly, Proposition 4.6.1(ii) implies that if  $i \in [t]$  and  $x \in V_2$  and  $ix, xi \notin E(G'[[t]])$  then  $e_a(G[V(C^i), \{x\}]) \leq k - 2 + a$ . So we have that

$$e_a(G'[[t], V_2]) \stackrel{(4.6.9)}{\geq} tn_2 - \frac{3\delta n^2}{2 - a}.$$

So recalling that  $e_a(G'[V_2]) = e_a(G[V_2])$  we have that

$$e_a(G') = e_a(G'[[t]]) + e_a(G'[V_2]) + e_a(G'[[t], V_2]) \stackrel{(4.6.8)}{\geq} \binom{t + n_2}{2} - \frac{8\delta n^2}{2 - a}. \quad (4.6.11)$$

Since  $t + n_2 \geq n/k$  we have that  $t + n_2 \geq n'_0$  and that  $8\delta n^2/(2 - a) \leq \delta'(t + n_2)^2$ , and hence by (4.6.11) that

$$e_a(G') \geq \binom{t + n_2}{2} - \delta'(t + n_2)^2. \quad (4.6.12)$$

We now claim that  $G'$  must be  $\{C_k, C_{k+1}\}$ -free. Indeed, suppose not. If  $G'$  contains a copy of  $C_{k+1}$  then it is clear that  $G$  also contains a copy of  $C_{k+1}$ , contradicting our assumption. So we may assume that  $G'$  contains a copy of  $C_k$ . Since  $G'[V_2] = G[V_2]$  is  $C_k$ -free by construction, the vertex set of any copy of  $C_k$  in  $G'$  must contain some  $i \in [t]$ . But then  $G$  would clearly contain a copy of  $C_{k+1}$  using two of the vertices in  $V(C^i)$ , again contradicting our assumption that  $G$  is  $C_{k+1}$ -free. So  $G'$  is  $\{C_k, C_{k+1}\}$ -free, as claimed.



Thus by (4.6.12) and the assumption in the statement of the claim we have that  $G' = T_{t+n_2} \pm \varepsilon(t+n_2)^2/(2k^2)$ . Together with the definition of  $G'$  this implies that  $G = T_n \pm \varepsilon n^2$ , as required. This completes the proof of the claim, and hence completes the proof of the lemma.  $\square$

The rough strategy of the next proof is similar to that of Lemma 4.6.5.

**Lemma 4.6.13 (Stability when  $a = 2$  and  $k$  is odd).** *Let  $k \in \mathbb{N}$  with  $k \geq 3$  and  $k$  odd. Then for all  $\varepsilon > 0$  there exist  $\delta > 0$  and  $n_0 \in \mathbb{N}$  such that every  $C_{k+1}$ -free digraph  $G$  on  $n \geq n_0$  vertices with*

$$e(G) \geq \binom{n}{2} - \delta n^2$$

*satisfies  $G = T_n \pm \varepsilon n^2$ .*

*Proof.* We prove the lemma via the following claim.

**Claim:** *Let  $k \in \mathbb{N}$  with  $k \geq 3$  and  $k$  odd, and let  $\varepsilon > 0$ . Suppose that there exist  $\delta' > 0$  and  $n'_0 \in \mathbb{N}$  such that every  $\{C_{k-1}, C_k\}$ -free digraph  $G$  on  $n' \geq n'_0$  vertices with*

$$e(G) \geq \binom{n'}{2} - \delta' n'^2$$

*satisfies  $G = T_{n'} \pm \varepsilon n'^2/(2k^2)$ . Then there exist  $\delta > 0$  and  $n_0 \in \mathbb{N}$  such that every  $C_{k+1}$ -free digraph  $G$  on  $n \geq n_0$  vertices with*

$$e(G) \geq \binom{n}{2} - \delta n^2$$

*satisfies  $G = T_n \pm \varepsilon n^2$ .*

In order to check that the claim implies the lemma, we proceed by induction on  $\ell := (k+1)/2$ . The argument is similar to that in the proof of Lemma 4.6.5. (As before,

Lemma 4.6.4 implies that in the base case  $\ell = 2$  of the induction, the assumption of the claim holds.)

**Proof of claim:** Choose  $\delta$  and  $n_0$  such that  $1/n_0 \ll \delta \ll 1/k, \delta'$  and  $1/n_0 \ll 1/n'_0, \varepsilon$ . Let  $G$  be a  $C_{k+1}$ -free digraph on  $n \geq n_0$  vertices with

$$e(G) \geq \binom{n}{2} - \delta n^2. \quad (4.6.14)$$

Let  $t \geq 0$  denote the maximum number of vertex-disjoint copies of  $C_k$  in  $G$ . Let  $\mathcal{C} = \{C^1, \dots, C^t\}$  be a set of  $t$  vertex-disjoint copies of  $C_k$  in  $G$ . Let  $V_1 := V(C^1) \cup \dots \cup V(C^t)$  and  $n_1 := |V_1|$ . Now let  $t^* \geq 0$  denote the maximum number of vertex-disjoint copies of  $C_{k-1}$  in  $G[V \setminus V_1]$ . Let  $\mathcal{C}^* = \{C_*^1, \dots, C_*^{t^*}\}$  be a set of  $t^*$  vertex-disjoint copies of  $C_{k-1}$  in  $G[V \setminus V_1]$ . Let  $V_2 := V(C_*^1) \cup \dots \cup V(C_*^{t^*})$  and  $n_2 := |V_2|$ . Let  $V_3 := V(G) \setminus (V_1 \cup V_2)$  and  $n_3 := |V_3|$ . Note that  $G[V_2 \cup V_3]$  is  $C_k$ -free and that  $G[V_3]$  is  $\{C_{k-1}, C_k\}$ -free.

Proposition 4.6.1(i) implies that  $e(V_1, V_2) \leq n_1 n_2$  and that  $e(V_1, V_3) \leq n_1 n_3$ , since  $G$  is  $C_{k+1}$ -free, and that  $e(V_2, V_3) \leq n_2 n_3$ , since  $G[V_2 \cup V_3]$  is  $C_k$ -free. Also, similarly to the proof of Lemma 4.6.5 we use Lemma 4.6.3 to get that  $e(G[V_i]) \leq \binom{n_i}{2} + \delta n^2$  for  $i = 1, 2, 3$ . Together with (4.6.14) this implies that

$$\begin{aligned} e(G[V_1]) &= e(G) - e(V_1, V_2) - e(V_1, V_3) - e(V_2, V_3) - e(G[V_2]) - e(G[V_3]) \\ &\geq \left( \binom{n}{2} - \delta n^2 \right) - n_1 n_2 - n_1 n_3 - n_2 n_3 - \left( \binom{n_2}{2} + \delta n^2 \right) - \left( \binom{n_3}{2} + \delta n^2 \right) \\ &= \binom{n_1}{2} - 3\delta n^2, \end{aligned}$$

and similarly that  $e(G[V_2]) \geq \binom{n_2}{2} - 3\delta n^2$ , and that  $e(G[V_3]) \geq \binom{n_3}{2} - 3\delta n^2$ , and that  $e(V_i, V_j) \geq n_i n_j - 4\delta n^2$  for all  $i, j \in \{1, 2, 3\}$ ,  $i \neq j$ .

We now consider the digraph  $G'$  defined on vertex set  $([t] \times \{0\}) \cup ([t^*] \times \{1\}) \cup V_3$  as follows. Firstly, for every vertex  $v \in V(G')$  define  $f(v)$  to be  $\{v\}$  if  $v \in V_3$ , to be  $V(C^i)$  if  $v = (i, 0) \in [t] \times \{0\}$ , and to be  $V(C_*^i)$  if  $v = (i, 1) \in [t^*] \times \{1\}$ . Now let  $G'[V_3] := G[V_3]$  and for vertices  $u, v \in V(G')$  with  $|\{u, v\} \cap V_3| \leq 1$  define  $uv \in E(G')$  if and only if  $G[f(u), f(v)] = \vec{K}(f(u), f(v))$ .

Note that by the Kővári-Sós-Turán theorem,  $G$  contains at most  $\delta n^2$  double edges, since  $G$  is  $C_{k+1}$ -free and  $k+1$  is even, and  $1/n_0 \ll \delta \ll 1/k$ . Note also that by Proposition 4.6.1(iv), if  $u, v \in [t] \times \{0\}$  then  $e(G[f(u), f(v)]) \leq k^2$ . If, moreover,  $uv, vu \notin E(G'[[t] \times \{0\}])$  and  $G[f(u), f(v)]$  contains no double edge, then by Proposition 4.6.1(vi) we have that  $e(G[f(u), f(v)]) \leq k^2 - 1$ . Using that  $e(G[V_1]) \geq \binom{n_1}{2} - 3\delta n^2$  one can now argue similarly as in (4.6.10) to see that

$$e(G'[[t] \times \{0\}]) \geq \binom{t}{2} - 5\delta n^2.$$

Also, Proposition 4.6.1(iv) implies that if  $u \in [t] \times \{0\}$  and  $v \in [t^*] \times \{1\}$  then  $e(G[f(u), f(v)]) \leq k(k-1)$ . If, moreover,  $uv, vu \notin E(G'[[t] \times \{0\}, [t^*] \times \{1\}])$  and  $G[f(u), f(v)]$  contains no double edge, then by Proposition 4.6.1(vi) we have that  $e(G[f(u), f(v)]) \leq k(k-1) - 1$ . Using that  $e(V_1, V_2) \geq n_1 n_2 - 4\delta n^2$  one can again argue similarly as in (4.6.10) to see that

$$e(G'[[t] \times \{0\}, [t^*] \times \{1\}]) \geq tt^* - 5\delta n^2.$$

Furthermore, Proposition 4.6.1(i) implies that if  $u \in [t] \times \{0\}$  and  $v \in V_3$  then  $e(G[f(u), f(v)]) \leq k$ . If, moreover,  $uv, vu \notin E(G'[[t] \times \{0\}, V_3])$  and  $G[f(u), f(v)]$  contains no double edge, then by Proposition 4.6.1(iii) we have that  $e(G[f(u), f(v)]) \leq k - 1$ . Using that  $e(V_1, V_3) \geq n_1 n_3 - 4\delta n^2$  one can again argue similarly as in (4.6.10) to see that

$$e(G'[[t] \times \{0\}, V_3]) \geq tn_3 - 5\delta n^2.$$

Using that  $G[V_2 \cup V_3]$  is  $C_k$ -free, in a similar way we get that  $e(G'[[t^*] \times \{1\}]) \geq \binom{t^*}{2} - 5\delta n^2$  and that  $e(G'[[t^*] \times \{1\}, V_3]) \geq t^*n_3 - 5\delta n^2$ . So recalling that  $e(G'[V_3]) = e(G[V_3]) \geq \binom{n_3}{2} - 3\delta n^2$  we have that

$$\begin{aligned} e(G') &= e(G'[[t] \times \{0\}]) + e(G'[[t^*] \times \{1\}]) + e(G'[V_3]) \\ &\quad + e(G'[[t] \times \{0\}, [t^*] \times \{1\}]) + e(G'[[t] \times \{0\}, V_3]) + e(G'[[t^*] \times \{1\}], V_3) \\ &\geq \binom{t + t^* + n_3}{2} - 28\delta n^2. \end{aligned} \tag{4.6.15}$$

Since  $t + t^* + n_3 \geq n/k$ , we have that  $t + t^* + n_3 \geq n'_0$  and that  $28\delta n^2 \leq \delta'(t + t^* + n_3)^2$ , and hence by (4.6.15) that

$$e(G') \geq \binom{t + t^* + n_3}{2} - \delta'(t + t^* + n_3)^2. \tag{4.6.16}$$

We now claim that  $G'$  must be  $\{C_{k-1}, C_k\}$ -free. Indeed, suppose not. If  $G'$  contains a copy of  $C_k$  then since  $G'[V_3] = G[V_3]$  is  $C_k$ -free by construction, the vertex set of such a copy of  $C_k$  in  $G'$  must contain some  $u \in ([t] \times \{0\}) \cup ([t^*] \times \{1\})$ . But then  $G$  would clearly contain a copy of  $C_{k+1}$  using two of the vertices in  $f(u)$ , contradicting our assumption that  $G$  is  $C_{k+1}$ -free. Similarly, if  $G'$  contains a copy of  $C_{k-1}$  then since  $G[V_3]$  is  $C_{k-1}$ -free by construction, the vertex set of such a copy of  $C_{k-1}$  in  $G'$  must contain some  $u \in ([t] \times \{0\}) \cup ([t^*] \times \{1\})$ . If there exists such a  $u \in [t] \times \{0\}$  then  $G$  would clearly contain a copy of  $C_{k+1}$  using three of the vertices in  $f(u)$ , since  $|f(u)| = k \geq 3$ . Otherwise, there exists  $u \in [t^*] \times \{1\}$  and a copy of  $C_{k-1}$  in  $G'$  that uses  $u$  but no vertices in  $[t] \times \{0\}$ . But then  $G[V_2 \cup V_3]$  would clearly contain a copy of  $C_k$  using two of the vertices in  $f(u)$ , contradicting our previous observation that  $G[V_2 \cup V_3]$  is  $C_k$ -free.

So  $G'$  is  $\{C_{k-1}, C_k\}$ -free, as claimed. Thus by (4.6.16) and the assumption in the statement

of the claim we have that  $G' = T_{t+t^*+n_3} \pm \varepsilon(t+t^*+n_3)^2/(2k^2)$ . Together with the definition of  $G'$  this implies that  $G = T_n \pm \varepsilon n^2$ , as required. This completes the proof of the claim, and hence completes the proof of the lemma.  $\square$

We now prove a digraph stability result for forbidden odd cycles. Here both the Turán and the stability results allow for a richer structure than in the previous two lemmas: For even  $k$ , a near extremal graph can be obtained from a transitive tournament by blowing up some of its vertices into complete bipartite graphs of arbitrary size (see Section 4.1 for the precise definition). This makes the proof more difficult than the previous two.

**Lemma 4.6.17 (Stability when  $a = 2$  and  $k$  is even).** *Let  $k \in \mathbb{N}$  be even. Then for all  $\varepsilon > 0$  there exist  $\delta > 0$  and  $n_0 \in \mathbb{N}$  such that every  $C_{k+1}$ -free digraph  $G$  on  $n \geq n_0$  vertices with*

$$e(G) \geq \binom{n}{2} - \delta n^2 \tag{4.6.18}$$

*can be made into a transitive-bipartite blow up by changing at most  $\varepsilon n^2$  edges.*

To give an idea of the proof, consider the triangle-free case  $k = 2$ . In this case, we first consider a maximal collection  $\mathcal{A}$  of disjoint double edges in  $G$ . It is easy to see that for almost all pairs of (double) edges  $u_1u_2, v_1v_2 \in \mathcal{A}$ , either (i)  $G[\{u_1, u_2, v_1, v_2\}]$  is a complete balanced bipartite digraph or (ii)  $G$  contains all four possible edges from  $\{u_1, u_2\}$  to  $\{v_1, v_2\}$  (or vice versa). We consider the following auxiliary ‘semi-oriented graph’  $G'$  whose vertex set is  $\mathcal{A}$ . In case (i), we include an (undirected) red edge between  $u_1u_2$  and  $v_1v_2$  in  $G'$ . In case (ii), we include a blue edge directed from  $u_1u_2$  to  $v_1v_2$  in  $G'$  (or vice versa). One can now show that the red edges induce a set of disjoint almost complete graphs  $R$  in  $G'$ . We then contract each such red almost complete graph  $R$  into a vertex  $v_R$  to obtain an oriented graph  $J$  (vertices of  $G'$  which are not involved in any of these  $R$  are also retained in  $J$ ). So all edges of  $J$  are blue. Crucially, it turns out that

$J$  is close to a transitive tournament. Moreover, in  $G$  each  $v_R$  corresponds to an almost complete bipartite digraph, so altogether this shows that the subgraph of  $G$  induced by the edges in  $\mathcal{A}$  is close to a transitive-bipartite blow up. One can generalize this argument to incorporate the vertices of  $G$  not covered by edges in  $\mathcal{A}$  (these will only be incident to blue edges in  $G'$  and  $J$  and not to any red ones).

To formalize the above argument, we make use of the following definitions. A *semi-oriented graph* is obtained from an undirected graph by first colouring each of the edges either red or blue and then giving an orientation to each of the blue edges. So a semi-oriented graph is a pair  $G = (V, E)$ , where  $V$  is a set of vertices and  $E$  is a set of coloured edges, some of which are red and undirected and the rest of which are blue and directed. We define basic notions such as induced subgraphs of  $G$  in the obvious way. Given a vertex  $v \in V$  we denote the set of all vertices  $x \in V$  for which there is a blue directed edge  $vx \in E$  by  $N_G^+(v)$ . We call the vertices in  $N_G^+(v)$  *blue out-neighbours* of  $v$ . We define the sets  $N_G^-(v)$  of *blue in-neighbours* of  $v$  and  $N_G^{\text{red}}(v)$  of *red neighbours* of  $v$  in a similar way. If  $x \in N_G^+(v) \cup N_G^-(v)$  we say that  $x$  is a *blue neighbour* of  $v$ .

We denote the complete bipartite digraph (with edges in both directions) with vertex classes of sizes  $a$  and  $b$  by  $DK_{a,b}$ . The following simple proposition will also be used in the proof of Lemma 4.6.17.

**Proposition 4.6.19.** *Let  $k \in \mathbb{N}$  be even and let  $G$  be a  $C_{k+1}$ -free digraph. Suppose  $G$  contains a copy of  $DK_{k/2, k/2}$  with vertex classes  $A, B$ . Then the following hold.*

- (i) *Suppose  $x \in V(G) \setminus (A \cup B)$ . Then  $e(G[\{x\}, A \cup B]) \leq k$ , with equality only if  $G[\{x\}, A \cup B] = \overrightarrow{K}(\{x\}, A \cup B)$  or  $G[\{x\}, A \cup B] = \overrightarrow{K}(A \cup B, \{x\})$  or  $G[\{x\} \cup A \cup B] = DK_{k/2+1, k/2}$ .*
- (ii) *Suppose  $G$  contains another copy of  $DK_{k/2, k/2}$  with vertex classes  $C, D$  such that*

$(C \cup D) \cap (A \cup B) = \emptyset$ . Then  $e(G[C \cup D, A \cup B]) \leq k^2$ , with equality only if  $G[C \cup D, A \cup B] = \vec{K}(C \cup D, A \cup B)$  or  $G[C \cup D, A \cup B] = \vec{K}(A \cup B, C \cup D)$  or  $G[C \cup D \cup A \cup B] = DK_{k,k}$ .

*Proof.* (i) Note that if there is an edge in  $G[\{x\}, A \cup B]$  from a vertex in  $A$  to  $x$  then there cannot be an edge in  $G[\{x\}, A \cup B]$  directed from  $x$  to a vertex in  $B$ , or else  $G[\{x\}, A \cup B]$  would clearly contain a copy of  $C_{k+1}$  that uses these two edges, contradicting the fact that  $G$  is  $C_{k+1}$ -free. Similarly, if there is an edge in  $G[\{x\}, A \cup B]$  directed from a vertex in  $B$  to  $x$  then there cannot be an edge in  $G[\{x\}, A \cup B]$  from  $x$  to a vertex in  $A$ , and the result now follows.

(ii) The fact that  $e(G[C \cup D, A \cup B]) \leq k^2$  follows immediately from (i). Now let us suppose that  $e(G[C \cup D, A \cup B]) = k^2$ . So (i) implies that, for every  $x \in C \cup D$ , either  $G[\{x\}, A \cup B] = \vec{K}(\{x\}, A \cup B)$  or  $G[\{x\}, A \cup B] = \vec{K}(A \cup B, \{x\})$  or  $G[\{x\} \cup A \cup B] = DK_{k/2+1, k/2}$ . We consider cases.

**Case 1:** *There exists  $x \in C \cup D$  with  $G[\{x\}, A \cup B] = \vec{K}(\{x\}, A \cup B)$ .* In this case, consider a vertex  $y \in A \cup B$ . Without loss of generality we may assume that  $x \in C$  and  $y \in A$ . By (i) (applied with  $C \cup D$  playing the role of  $A \cup B$  and  $y$  playing the role of  $x$ ) we have that either  $G[\{y\}, C \cup D] = \vec{K}(\{y\}, C \cup D)$  or  $G[\{y\}, C \cup D] = \vec{K}(C \cup D, \{y\})$  or  $G[\{y\} \cup C \cup D] = DK_{k/2+1, k/2}$ . But if  $G[\{y\}, C \cup D] = \vec{K}(\{y\}, C \cup D)$  then there is an edge directed from  $y$  to a vertex in  $D$  and an edge directed from  $x \in C$  to  $y$ , and as in the proof of (i) this implies that  $G$  contains a copy of  $C_{k+1}$ , which is a contradiction. If  $G[\{y\} \cup C \cup D] = DK_{k/2+1, k/2}$  then either  $G[\{y\}, C] = DK_{1, k/2}$ , in which case there is an edge directed from  $x$  to a vertex in  $B$  and an edge directed from  $y \in A$  to  $x$ , or else  $G[\{y\}, D] = DK_{1, k/2}$ , in which case there is an edge directed from  $y$  to a vertex in  $D$  and an edge directed from  $x \in C$  to  $y$ . In either case there must be a copy of  $C_{k+1}$  in  $G$ , as in the proof of (i), which is a contradiction. Hence it must be that  $G[\{y\}, C \cup D] = \vec{K}(C \cup D, \{y\})$ .

Since  $y \in A \cup B$  was arbitrary,  $G[\{y'\}, C \cup D] = \vec{K}(C \cup D, \{y'\})$  for every  $y' \in A \cup B$ , which implies that  $G[C \cup D, A \cup B] = \vec{K}(C \cup D, A \cup B)$ .

**Case 2:** *There exists  $x \in C \cup D$  with  $G[\{x\}, A \cup B] = \vec{K}(A \cup B, \{x\})$ .* In this case it follows by a symmetric argument to that of Case 1 that  $G[C \cup D, A \cup B] = \vec{K}(A \cup B, C \cup D)$ .

**Case 3:**  $G[\{x\} \cup A \cup B] = DK_{k/2+1, k/2}$  for all  $x \in C \cup D$ . In this case, suppose that there exist  $y, z \in C$  with  $G[\{y\}, A] = DK_{1, k/2}$  and  $G[\{z\}, B] = DK_{1, k/2}$ . (i) implies that  $y \neq z$ . There is a path  $P$  in  $G[C \cup D]$  of length  $k-2$  from  $y$  to  $z$ , so for any  $a \in A$  and  $b \in B$  we have that  $yPzbay$  is a copy of  $C_{k+1}$  in  $G$ , which is a contradiction. So we may assume that  $G[\{y\}, A] = DK_{1, k/2}$  for all  $y \in C$ . It now easily follows that  $G[\{z\}, B] = DK_{1, k/2}$  for all  $z \in D$  and thus  $G[C \cup D \cup A \cup B] = DK_{k, k}$ .

We have now considered all cases, and so this completes the proof of (ii).

□

**Proof of Lemma 4.6.17.** Choose  $n_0, \delta, \varepsilon_1, \varepsilon_2$  such that  $1/n_0 \ll \delta \ll \varepsilon_1 \ll \varepsilon_2 \ll 1/k, \varepsilon$ . Let  $G$  be a  $C_{k+1}$ -free digraph on  $n \geq n_0$  vertices which satisfies (4.6.18).

Let  $t \geq 0$  denote the maximum number of vertex-disjoint copies of  $DK_{k/2, k/2}$  in  $G$ . Let  $\mathcal{A} = \{A^1, \dots, A^t\}$  be a collection of vertex sets of  $t$  vertex-disjoint copies of  $DK_{k/2, k/2}$  in  $G$ . Let  $V_1 := A^1 \cup \dots \cup A^t$  and let  $V_2 := V(G) \setminus V_1$ . Note that  $G[V_2]$  is  $DK_{k/2, k/2}$ -free, and hence by the Kővári-Sós-Turán theorem  $G[V_2]$  contains at most  $\delta n^2$  double edges.

**Claim 1:** *For each  $i \in [t]$  there are at most  $6\delta^{1/2}n$  vertices  $x \in V_2$  for which  $G[A^i \cup \{x\}] = DK_{k/2, k/2+1}$ .*



Indeed, suppose that there exists a set  $X$  of more than  $6\delta^{1/2}n$  such vertices. Proposition 4.6.19(ii) and the fact that  $1/n_0 \ll \delta, 1/k$  together imply that  $e(G[V_1]) \leq \binom{|V_1|}{2} + \delta n^2$ . Moreover,  $e(V_1, V_2) \leq |V_1||V_2|$  by Proposition 4.6.19(i). Together with (4.6.18) this implies that  $e(G[V_2]) \geq \binom{|V_2|}{2} - 2\delta n^2$ . This together with our previous observation that  $G[V_2]$  contains at most  $\delta n^2$  double edges implies that  $G[V_2]$  contains at most  $3\delta n^2$  pairs of vertices with no edge between them. Hence  $e(G[X]) \geq \binom{|X|}{2} - 3\delta n^2$ . But this means that there are  $x, y \in X$  such that  $xy \in E(G)$  and such that both  $x$  and  $y$  are joined with double edges to the same vertex class of  $G[A^i] = DK_{k/2, k/2}$ , which contradicts the fact that  $G$  is  $C_{k+1}$ -free.

Let  $G^*$  be the digraph obtained from  $G$  by deleting the at most  $\delta n^2$  double edges in  $G[V_2]$  and deleting the double edges between  $A^i$  and all the vertices  $x \in V_2$  for which  $G[A^i \cup \{x\}] = DK_{k/2, k/2+1}$  (for each  $i \in [t]$ ). By Claim 1, for each  $i \in [t]$ , the number of the latter double edges is at most  $6\delta^{1/2}n \cdot k/2 = 3k\delta^{1/2}n$ . Thus

$$e(G^*) \geq e(G) - 2 \left( \delta n^2 + \frac{n}{k} \cdot 3k\delta^{1/2}n \right) \stackrel{(4.6.18)}{\geq} \binom{n}{2} - 7\delta^{1/2}n^2. \quad (4.6.20)$$

Consider the semi-oriented graph  $G' = (V', E')$  where  $V' := \mathcal{A} \cup V_2$  and the edge set  $E'$  is defined as follows. Firstly, for every vertex  $v \in V'$  define  $f(v)$  to be  $v$  if  $v \in \mathcal{A}$ , and to be  $\{v\}$  if  $v \in V_2$ . If  $u, v \in V'$  then there is a blue edge in  $E'$  directed from  $u$  to  $v$  if  $G[f(u), f(v)] = \vec{K}(f(u), f(v))$ . If  $u, v \in \mathcal{A}$  then there is a red edge in  $E'$  between  $u$  and  $v$  if  $G[f(u) \cup f(v)] = DK_{k,k}$ . So  $G'[V_2] = G^*[V_2]$ .

Note that, since  $G$  is  $C_{k+1}$ -free,  $G'$  cannot contain any copy of  $C_{k+1}$  which contains at least one blue edge and in which all the blue edges are oriented consistently (as any such copy of  $C_{k+1}$  in  $G'$  would correspond to a  $C_{k+1}$  in  $G$ ).

Let  $V'_0$  denote the set of all vertices  $v \in V'$  for which there are at least  $\delta^{1/4}n$  vertices  $u \in V'$  such that  $G'$  does not contain an edge between  $v$  and  $u$ . Note that Proposition 4.6.19 together with (4.6.20) implies that  $|E'| \geq \binom{|V'|}{2} - 7\delta^{1/2}n^2$ . Hence  $|V'_0| \leq 15\delta^{1/4}n$ . Let  $G'' := G' - V'_0$ ,  $V'' := V' \setminus V'_0$  and  $E'' := E(G'')$ .

**Claim 2:**

- (a) *For every vertex  $v \in V''$  there are at most  $\delta^{1/4}n$  vertices  $u \in V''$  such that  $G''$  does not contain an edge between  $v$  and  $u$ .*
- (b)  *$G''$  does not contain a triangle  $uvw$  such that both  $uv$  and  $vw$  are red edges and  $wu$  is a blue edge.*
- (c)  *$G''$  does not contain a triangle  $uvw$  such that  $uv$  is a red edge and both  $vw$  and  $wu$  are (directed) blue edges.*

Indeed, (a) is clear from the definition of  $G''$ , while (b) and (c) follow easily from the fact that  $G$  is  $C_{k+1}$ -free.

Given  $q, q' \in \mathbb{N}$ , we say that  $U \subseteq V''$  is a *red  $(q, q')$ -clique* if  $|U| \geq q'$  and  $|U \setminus N_{G''}^{\text{red}}(u)| \leq q$  for all  $u \in U$ .

**Claim 3:** *Suppose that  $R$  is a red  $(\delta^{1/4}n, \varepsilon_1 n)$ -clique. Then the following hold.*

- (a)  *$G''[R]$  does not contain a blue edge.*
- (b) *No vertex  $v \in V'' \setminus R$  has both a red and a blue neighbour in  $R$ .*
- (c) *No vertex  $v \in V'' \setminus R$  has both a blue in-neighbour and a blue out-neighbour in  $R$ .*

Suppose that  $R'$  is another red  $(\delta^{1/4}n, \varepsilon_1 n)$ -clique such that  $R \cap R' = \emptyset$ . Then the following hold.

- (d)  $G''$  cannot contain both a red edge and a blue edge between  $R$  and  $R'$ .
- (e)  $G''$  cannot contain both a directed blue edge from some vertex in  $R$  to some vertex in  $R'$  and a directed blue edge from some vertex in  $R'$  to some vertex in  $R$ .

First note that (a) follows immediately from Claim 2(b) and the definition of a red  $(\delta^{1/4}n, \varepsilon_1 n)$ -clique. To prove (b), suppose that some vertex  $v \in V'' \setminus R$  has both a red and a blue neighbour in  $R$ . Claim 2(a) and the fact that  $|R| \geq \varepsilon_1 n$  imply that either  $v$  has at least  $\varepsilon_1 n/3$  red neighbours in  $R$  or at least  $\varepsilon_1 n/3$  blue neighbours in  $R$  (or both). Suppose that the former holds (the argument for the latter is similar). Let  $u \in R$  be a blue neighbour of  $v$ . Since  $R$  is a red  $(\delta^{1/4}n, \varepsilon_1 n)$ -clique, all but at most  $\delta^{1/4}n < \varepsilon_1 n/3$  vertices of  $R$  are red neighbours of  $u$ . So there exists a red neighbour  $u' \in R$  of  $u$  which is also a red neighbour of  $v$ . Then the triangle  $uu'v$  contradicts Claim 2(b). This proves (b). The argument for (c) is similar. (d) follows from (b) and Claim 2(a), while (e) follows from (b), (c) and Claim 2(a).

**Claim 4:** *There exists a collection  $\mathcal{R}$  of pairwise disjoint red  $(\delta^{1/4}n, \varepsilon_1 n)$ -cliques such that, writing  $V_{\mathcal{R}}$  for the set of all those vertices in  $V''$  covered by these red  $(\delta^{1/4}n, \varepsilon_1 n)$ -cliques, the following holds:*

- (a) *For every  $R \in \mathcal{R}$  and every  $v \in R$  all red neighbours of  $v$  lie in  $R$ .*
- (b) *Every  $v \in V'' \setminus V_{\mathcal{R}}$  has less than  $\varepsilon_1 n$  red neighbours (and all of these lie in  $V'' \setminus V_{\mathcal{R}}$ ).*

To prove Claim 4, let  $\mathcal{R}$  be a collection of pairwise disjoint red  $(\delta^{1/4}n, \varepsilon_1 n)$ -cliques such

that the set  $V_{\mathcal{R}}$  of all those vertices in  $V''$  covered by these red  $(\delta^{1/4}n, \varepsilon_1 n)$ -cliques is maximal and, subject to this condition, such that  $|\mathcal{R}|$  is minimal. We will show that  $\mathcal{R}$  is as required in Claim 4.

To prove that Claim 4(a) holds, suppose first that there is some vertex  $x \in V'' \setminus V_{\mathcal{R}}$  that has a red neighbour in some  $R \in \mathcal{R}$ . Then Claims 2(a) and 3(b) together imply that  $|(R \cup \{x\}) \setminus N_{G''}^{\text{red}}(x)| \leq \delta^{1/4}n$ . Moreover, by Claims 2(a) and 3(a),(b) we have that  $|(R \cup \{x\}) \setminus N_{G''}^{\text{red}}(v)| \leq \delta^{1/4}n$  for every  $v \in R$ . So  $R \cup \{x\}$  is a red  $(\delta^{1/4}n, \varepsilon_1 n)$ -clique, contradicting our choice of  $\mathcal{R}$ .

Suppose next that there are distinct  $R, R' \in \mathcal{R}$  such that  $G''$  contains a red edge between  $R$  and  $R'$ . Then Claims 3(a),(d) imply that  $G''[R \cup R']$  does not contain a blue edge. Together with Claim 2(a) this implies that  $R \cup R'$  is a red  $(\delta^{1/4}n, \varepsilon_1 n)$ -clique, again contradicting our choice of  $\mathcal{R}$ . Altogether this proves Claim 4(a).

To check Claim 4(b), suppose that some  $v \in V'' \setminus V_{\mathcal{R}}$  has at least  $\varepsilon_1 n$  red neighbours. Claim 2(b) implies that  $G''[N_{G''}^{\text{red}}(v)]$  cannot contain a blue edge. Together with Claim 2(a) this implies that  $G''[N_{G''}^{\text{red}}(v)]$  is a red  $(\delta^{1/4}n, \varepsilon_1 n)$ -clique. But Claim 4(a) implies that  $N_{G''}^{\text{red}}(v) \subseteq V'' \setminus V_{\mathcal{R}}$ , contradicting our choice of  $\mathcal{R}$ . This completes the proof of Claim 4.

Let  $G'''$  be the semi-oriented graph obtained from  $G''$  by deleting all the red edges which are not covered by some  $R \in \mathcal{R}$ . Note that by Claim 4(b) at most  $\varepsilon_1 n^2$  red edges are deleted. Let  $J$  be the oriented graph obtained from  $G'''$  by contracting each  $R \in \mathcal{R}$  into a single vertex  $v_R$ . So  $V(J)$  consists of all these vertices  $v_R$  as well as all the vertices in  $V'' \setminus V_{\mathcal{R}}$ . Let  $J_2 := J[V'' \setminus V_{\mathcal{R}}] = G'''[V'' \setminus V_{\mathcal{R}}]$  and let  $J_1 := J - V(J_2)$ . Claims 3(c),(e) and Claim 4(a) together imply that  $J$  is indeed an oriented graph. Moreover, by Claim 2(a)  $J_1$  is a tournament and  $J[V(J_1), V(J_2)]$  is a bipartite tournament (i.e. for all  $v_R \in V(J_1)$  and  $v \in V(J_2)$  either  $v_R v$  or  $vv_R$  is a directed edge of  $J$ ).

**Claim 5:**

- (a)  $J$  does not contain a copy of  $C_3$  having at least one vertex in  $V(J_1)$ .
- (b)  $J_1$  is a transitive tournament.
- (c)  $J$  can be made into a transitive tournament by changing at most  $\varepsilon_2 n^2$  edges in  $E(J_2)$ .

Suppose that (a) does not hold and let  $xyv_R$  be a copy of  $C_3$  in  $J$ . We only consider the case when  $x \in V(J_1)$  and  $y \in V(J_2)$ ; the other cases are similar. So let  $R' \in \mathcal{R}$  be such that  $x = v_{R'}$ . Claim 2(a) and the definition of  $J$  together imply that  $R'$  contains a blue in-neighbour  $x'$  of  $y$  (in  $G''$ ). Moreover, Claims 2(a) and 3(c),(e) imply that  $|R \setminus N_{G''}^-(x')| \leq \delta^{1/4}n$  and  $|R \setminus N_{G''}^+(y)| \leq \delta^{1/4}n$ . Together with the fact that  $R$  is a red  $(\delta^{1/4}n, \varepsilon_1 n)$ -clique this implies that  $R$  contains a path  $P = u \dots v$  of length  $k - 2$  where  $u \in N_{G''}^+(y)$  and  $v \in N_{G''}^-(x')$ . So  $Px'y$  is a  $C_{k+1}$  in  $G''$  in which all the blue edges are oriented consistently. Using the fact that the edge  $vx'$  is blue, it is now easy to see that  $Px'y$  corresponds to a  $C_{k+1}$  in  $G$ , a contradiction. This proves (a). (b) follows from (a) and our previous observation that  $J_1$  is a tournament.

It remains to prove (c). Note that  $e(J_2) \geq \binom{|J_2|}{2} - 2\varepsilon_1 n^2$  by Claim 2(a) and the definition of  $J$  (and of  $G'''$ ). Moreover,  $J_2 = G'''[V'' \setminus V_{\mathcal{R}}]$  is a  $C_{k+1}$ -free oriented graph. So Lemma 4.6.5 implies that  $J_2 = T_{|J_2|} \pm \varepsilon_2 n^2$ . Let  $\sigma_2 : V(J_2) \rightarrow [|J_2|]$  be a transitive-optimal ordering of the vertices of  $J_2$ . Let  $r := |J_1| = |\mathcal{R}|$  and let  $v_{R_1}, \dots, v_{R_r}$  be the unique transitive ordering of the vertices of  $J_1$ . We claim that for every vertex  $x \in V(J_2)$  there exists an index  $i_x \in [r]$  such that all the  $v_{R_i}$  with  $i \leq i_x$  are in-neighbours of  $x$  in  $J$  while all the  $v_{R_i}$  with  $i > i_x$  are out-neighbours of  $x$  in  $J$ . (Indeed, suppose not. Since  $J[V(J_1), V(J_2)]$  is a bipartite tournament this implies that there are indices  $i < j$  such that  $v_{R_i}$  is an out-neighbour of  $x$  in  $J$  and  $v_{R_j}$  is an in-neighbour of  $x$  in  $J$ . But then  $xv_{R_i}v_{R_j}$  is a copy

of  $C_3$  contradicting (a).) For each  $i \in [r]$  let  $X_i := \{x \in V(J_2) : i_x = i\}$ . Note that there are no indices  $i < j$  such that  $J$  contains a directed edge from some vertex  $x \in X_j$  to some vertex  $x' \in X_i$  (otherwise  $xx'v_{R_{i+1}}$  would be a copy of  $C_3$  contradicting (a)). Consider the vertex ordering  $\sigma$  obtained from  $v_{R_1}, \dots, v_{R_r}$  by including all the vertices in  $X_i$  between  $v_{R_i}$  and  $v_{R_{i+1}}$  in the ordering induced by  $\sigma_2$  (for each  $i \in [r]$ ). This vertex ordering shows that (c) holds.

Recall that for each  $R \in \mathcal{R}$  the set  $\bigcup R$  is a subset of  $V(G)$  of size  $k|R|$ .

**Claim 6:** *Each red  $(\delta^{1/4}n, \varepsilon_1 n)$ -clique  $R \in \mathcal{R}$  satisfies  $G[\bigcup R] = DK_{|R|/2, |R|/2} \pm \delta^{1/5}n^2$ .*

To prove Claim 6, pick  $v \in R$  and write  $N_{G''}^{\text{red}}(v) \cap R = \{v_1, \dots, v_s\}$ . Recall that  $v$  corresponds to a copy of  $DK_{k/2, k/2}$  in  $G$ , and let  $A$  and  $B$  denote the vertex classes of this copy. Similarly, each  $v_i$  corresponds to a copy of  $DK_{k/2, k/2}$  in  $G$ . Let  $A_i$  and  $B_i$  denote its vertex classes. Recall from the definition of  $G''$  that  $G[A \cup A_i \cup B \cup B_i] = DK_{k, k}$ . By swapping  $A_i$  and  $B_i$  if necessary, we may assume that the vertex classes of this copy of  $DK_{k, k}$  are  $A \cup A_i$  and  $B \cup B_i$ . Since  $G$  is  $C_{k+1}$ -free, neither  $G[A_1 \cup \dots \cup A_s]$  nor  $G[B_1 \cup \dots \cup B_s]$  contains an edge. Thus whenever  $v_i v_j$  is a red edge in  $G''$  then  $G[A_i \cup A_j \cup B_i \cup B_j]$  is a copy of  $DK_{k, k}$  with vertex classes  $A_i \cup A_j$  and  $B_i \cup B_j$ . But since  $R$  is a red  $(\delta^{1/4}n, \varepsilon_1 n)$ -clique, for each  $i \in [s]$  all but at most  $\delta^{1/4}n$  vertices in  $\{v_1, \dots, v_s\}$  are red neighbours of  $v_i$  and  $|R \setminus \{v_1, \dots, v_s\}| \leq \delta^{1/4}n$ . Thus  $G[\bigcup R] = DK_{|R|/2, |R|/2} \pm \delta^{1/5}n^2$ , as required.

Using Claims 5(c) and 6 it is now straightforward to check that  $G$  can be made into a transitive-bipartite blow up by changing at most  $\varepsilon n^2$  edges.  $\square$

We now have all the tools we need to show that almost all  $C_k$ -free oriented graphs are close to acyclic, and that for all even  $k$  almost all  $C_k$ -free digraphs are close to acyclic,

and that for all odd  $k$  almost all  $C_k$ -free digraphs are close to a transitive-bipartite blow up. The proof of Lemma 4.6.21 is almost identical to that of Lemma 4.4.9, using Lemmas 4.6.5, 4.6.13 and 4.6.17 instead of Lemma 4.4.3, and so is omitted here.

**Lemma 4.6.21.** *For every  $k \in \mathbb{N}$  with  $k \geq 3$  and any  $\alpha > 0$  there exists  $\varepsilon > 0$  such that the following holds for all sufficiently large  $n$ .*

- (i) *All but at most  $f(n, C_k)2^{-\varepsilon n^2}$   $C_k$ -free oriented graphs on  $n$  vertices can be made into subgraphs of  $T_n$  by changing at most  $\alpha n^2$  edges.*
- (ii) *If  $k$  is even then all but at most  $f^*(n, C_k)2^{-\varepsilon n^2}$   $C_k$ -free digraphs on  $n$  vertices can be made into subgraphs of  $T_n$  by changing at most  $\alpha n^2$  edges.*
- (iii) *If  $k$  is odd then all but at most  $f^*(n, C_k)2^{-\varepsilon n^2}$   $C_k$ -free digraphs on  $n$  vertices can be made into a subgraph of a transitive-bipartite blow up by changing at most  $\alpha n^2$  edges.*

## 4.7 Typical $C_k$ -free oriented graphs and digraphs are not acyclic

Let  $\mathcal{O}_{n,k}$  be the set of all labelled  $C_k$ -free oriented graphs on  $n$  vertices and let  $\mathcal{O}_{n,k}^*$  be the set of all labelled  $C_k$ -free digraphs on  $n$  vertices. We show that almost all graphs in  $\mathcal{O}_{n,k}$  and almost all graphs in  $\mathcal{O}_{n,k}^*$  have at least  $cn/\log n$  backwards edges in a transitive-optimal ordering, for some constant  $c > 0$ . Let  $\mathcal{O}_{n,k,r}$  be the set of all labelled  $C_k$ -free oriented graphs on  $n$  vertices with exactly  $r$  backwards edges in a transitive-optimal ordering. Let  $\mathcal{O}_{n,k,\leq r} := \bigcup_{i \in \{0,1,\dots,[r]\}} \mathcal{O}_{n,k,i}$ , and define the digraph analogues  $\mathcal{O}_{n,k,r}^*$  and  $\mathcal{O}_{n,k,\leq r}^*$  in a similar way.

**Lemma 4.7.1.** *Let  $k \geq 3$  and let  $n \in \mathbb{N}$  be sufficiently large. Then*

$$(i) \quad |\mathcal{O}_{n,k,n/2^{13}}| \geq 2^{n/2^{14}} |\mathcal{O}_{n,k,\leq n/(2^{14} \log n)}|,$$

$$(ii) \quad |\mathcal{O}_{n,k,n/2^{13}}^*| \geq 2^{n/2^{14}} |\mathcal{O}_{n,k,\leq n/(2^{14} \log n)}^*|.$$

Note that Lemma 4.7.1 together with Lemma 4.6.21 immediately yields Theorem 4.1.3.

**Proof of Lemma 4.7.1.** We only prove the case  $k = 3$  of (i) here; the proofs for (ii) and the case  $k > 3$  are very similar. Let  $m_2 := \lfloor n/2^{13} \rfloor$ . Fix  $m_1 \in \mathbb{Z}$  with  $0 \leq m_1 \leq m_2/(2 \log n)$ . For every oriented graph  $G$  fix some transitive-optimal ordering  $\sigma_G : V(G) \rightarrow [n]$ .

Consider an auxiliary bipartite graph  $H$  with vertex classes  $\mathcal{O}_{n,3,m_1}$  and  $\mathcal{O}_{n,3,0}$  whose edge set is defined as follows. Let there be an edge in  $H$  between  $A \in \mathcal{O}_{n,3,m_1}$  and  $B \in \mathcal{O}_{n,3,0}$  if the graph  $B$  can be obtained from the graph  $A$  by deleting the  $m_1$  backwards edges with respect to  $\sigma_A$ . Note that every graph generated in this way from a graph  $A \in \mathcal{O}_{n,3,m_1}$  belongs to  $\mathcal{O}_{n,3,0}$ , so  $A$  certainly has at least one neighbour in  $\mathcal{O}_{n,3,0}$ .

We claim that, in  $H$ , a graph  $B \in \mathcal{O}_{n,3,0}$  has at most  $\binom{n^2/2}{m_1} 2^{m_1}$  neighbours in  $\mathcal{O}_{n,3,m_1}$ . Indeed, any graph in  $\mathcal{O}_{n,3,m_1}$  that can generate  $B$  in the described way can be obtained from  $B$  by choosing exactly  $m_1$  of the at most  $n^2/2$  pairs of vertices that have no edge between them in  $B$ , and then adding edges between them with some orientations (for which there are  $2^{m_1}$  possibilities).



Together with our previous observation that, in  $H$ , every graph  $A \in \mathcal{O}_{n,3,m_1}$  has at least one neighbour in  $\mathcal{O}_{n,3,0}$ , this implies that

$$|\mathcal{O}_{n,3,m_1}| \leq \sum_{A \in \mathcal{O}_{n,3,m_1}} d_{\mathcal{O}_{n,3,0}}(A) = \sum_{B \in \mathcal{O}_{n,3,0}} d_{\mathcal{O}_{n,3,m_1}}(B) \leq |\mathcal{O}_{n,3,0}| \binom{n^2/2}{m_1} 2^{m_1}. \quad (4.7.2)$$

For a graph  $G \in \mathcal{O}_{n,3,0}$  we define a *flippable 4-set* in  $G$  to be any set of 4 vertices, with labels  $w, x, y, z$  say, satisfying the following:

- the vertices  $w, x, y, z$  are consecutive in the ordering  $\sigma_G$ ; that is  $\sigma_G(w) + 3 = \sigma_G(x) + 2 = \sigma_G(y) + 1 = \sigma_G(z)$ ,
- $\sigma_G(w) - 1$  is divisible by 4.

Note that every graph in  $\mathcal{O}_{n,3,0}$  has  $\lfloor n/4 \rfloor$  flippable 4-sets.

Now consider an auxiliary bipartite graph  $H'$  with vertex classes  $\mathcal{O}_{n,3,0}$  and  $\mathcal{O}_{n,3,m_2}$  whose edge set is defined as follows. Let there be an edge in  $H'$  between  $B \in \mathcal{O}_{n,3,0}$  and  $C \in \mathcal{O}_{n,3,m_2}$  if the graph  $C$  can be obtained from the graph  $B$  by choosing exactly  $m_2$  flippable 4-sets in  $B$  with respect to  $\sigma_B$  and, for each flippable 4-set  $w, x, y, z$  chosen, deleting all edges between the vertices  $w, x, y, z$  and then adding the edges of a 4-cycle  $wxyz$ . Note that every graph generated in this way from a graph  $B \in \mathcal{O}_{n,3,0}$  belongs to  $\mathcal{O}_{n,3,m_2}$ .

We claim that, in  $H'$ , a graph  $B \in \mathcal{O}_{n,3,0}$  has exactly  $\binom{\lfloor n/4 \rfloor}{m_2}$  neighbours in  $\mathcal{O}_{n,3,m_2}$ . Indeed, the neighbours of  $B$  are precisely those graphs generated by choosing exactly  $m_2$  of the exactly  $\lfloor n/4 \rfloor$  flippable 4-sets in  $B$  with respect to  $\sigma_B$ , and then changing the edges between pairs of vertices in these flippable 4-sets in the described way. Each choice of  $m_2$  flippable 4-sets generates a different graph. So the claim holds.

We claim also that, in  $H'$ , a graph  $C \in \mathcal{O}_{n,3,m_2}$  has at most  $2^{8m_2}$  neighbours in  $\mathcal{O}_{n,3,0}$ . Indeed, first note that any graph in  $\mathcal{O}_{n,3,m_2}$  with at least one neighbour in  $\mathcal{O}_{n,3,0}$  contains exactly  $m_2$  induced 4-cycles. Any graph in  $\mathcal{O}_{n,3,0}$  that can generate  $C$  in the described way can be obtained from  $C$  by choosing for each of the  $m_2$  induced 4-cycles an ordering of the 4 vertices respecting the order of the 4-cycle (of which there are 4), and then changing the edges between pairs of vertices in these 4-cycles to some transitive configuration with respect to the chosen ordering (for which there are  $2^6$  possibilities). So indeed the claim holds.

So using these degree bounds gives us that

$$|\mathcal{O}_{n,3,0}| \binom{\lfloor n/4 \rfloor}{m_2} = \sum_{B \in \mathcal{O}_{n,3,0}} d_{\mathcal{O}_{n,3,m_2}}(B) = \sum_{C \in \mathcal{O}_{n,3,m_2}} d_{\mathcal{O}_{n,3,0}}(C) \leq |\mathcal{O}_{n,3,m_2}| 2^{8m_2}. \quad (4.7.3)$$

Now (4.7.2) and (4.7.3) together imply that

$$\frac{|\mathcal{O}_{n,3,m_2}|}{|\mathcal{O}_{n,3,m_1}|} \geq \frac{\binom{\lfloor n/4 \rfloor}{m_2}}{\binom{n^{2/2}}{m_1} 2^{m_1} 2^{8m_2}}. \quad (4.7.4)$$

Since  $n$  is sufficiently large we have that  $\binom{\lfloor n/4 \rfloor}{m_2} \geq \left(\frac{n}{8m_2}\right)^{m_2}$  and  $\binom{n^{2/2}}{m_1} 2^{m_1} \leq n^{2m_1}$ . Hence the right hand side of (4.7.4) is at least

$$\left(\frac{n}{2^{11}m_2}\right)^{m_2} n^{-2m_1} \geq \left(\frac{n}{2^{11}m_2}\right)^{m_2} n^{-\frac{m_2}{\log n}} = 2^{m_2 \log\left(\frac{n}{2^{11}m_2}\right) - \frac{m_2}{\log n} \log n} \geq 2^{m_2}.$$

So this together with (4.7.4) gives us that  $|\mathcal{O}_{n,3,m_2}| \geq 2^{m_2} |\mathcal{O}_{n,3,m_1}|$  for any integer  $0 \leq m_1 \leq m_2/(2 \log n)$ . So since  $n$  is sufficiently large,

$$|\mathcal{O}_{n,3,m_2}| \geq \frac{2^{m_2}}{n} |\mathcal{O}_{n,3,\leq m_2/(2 \log n)}| \geq 2^{n/2^{14}} |\mathcal{O}_{n,3,\leq m_2/(2 \log n)}|,$$

as required.



# CHAPTER 5

## FORBIDDING INDUCED EVEN CYCLES IN A GRAPH: TYPICAL STRUCTURE AND COUNTING

### 5.1 Chapter introduction

#### 5.1.1 Graphs with forbidden induced graphs

Recall that, given a fixed graph  $H$ , a graph is called *induced- $H$ -free* if it does not contain  $H$  as an induced subgraph. As mentioned in Chapter 1, much less is known about the typical structure and number of induced- $H$ -free graphs than that of  $H$ -free graphs, though considerable work has been done in this area (see, e.g. [3, 10, 40, 61, 62, 63]). In particular, Prömel and Steger [63] obtained an asymptotic counting result for the number of induced- $H$ -free graphs on  $n$  vertices, showing that the logarithm of this number is essentially determined by the so-called colouring number of  $H$ . This was generalised

to arbitrary hereditary properties independently by Alekseev [2] as well as Bollobás and Thomason [16]. The recent exciting developments on Hypergraph Containers in [11, 70] that we utilised in Chapter 4 have also opened up the opportunity to replace counting results on induced- $H$ -free graphs by more precise results which identify the typical asymptotic structure.

In this chapter we determine the typical structure of induced- $C_{2k}$ -free graphs (from which the corresponding asymptotic counting result follows immediately). The key difficulty we encounter is that the typical structure turns out to be more complex than encountered in previous results on forbidden induced subgraphs. This requires new ideas and a more intricate analysis when ‘excluding’ classes of graphs which might be candidates for typical induced- $C_{2k}$ -free graphs.

### 5.1.2 Graphs with forbidden induced cycles

Given graphs  $H_1, \dots, H_m$ , we say  $G$  can be covered by  $H_1, \dots, H_m$  if  $V(G)$  admits a partition  $A_1 \cup \dots \cup A_m = V(G)$  such that  $G[A_i]$  is isomorphic to  $H_i$  for every  $i \in \{1, \dots, m\}$ .

Prömel and Steger proved in [61] that almost all induced- $C_4$ -free graphs can be covered by a clique and an independent set, and in [60] characterised the structure of almost all induced- $C_5$ -free graphs too. More recently, Balogh and Butterfield [10] determined the typical structure of induced- $H$ -free graphs for a wide class of graphs  $H$ . In particular they proved that almost all induced- $C_7$ -free graphs can be covered by either three cliques or two cliques and an independent set, and that for  $k \geq 4$  almost all induced- $C_{2k+1}$ -free graphs can be covered by  $k$  cliques. They also conjectured that for  $k \geq 6$  almost all induced- $C_{2k}$ -free graphs can be covered by  $k-2$  cliques and a graph whose complement is a disjoint union of stars and triangles. The main result of this chapter completely verifies

this conjecture.

**Theorem 5.1.1.** *For  $k \geq 6$ , almost all induced- $C_{2k}$ -free graphs can be covered by  $k - 2$  cliques and a graph whose complement is a disjoint union of stars and triangles.*

Theorem 5.1.1 together with the discussed results in [10, 32, 60, 61] implies that the typical structure of induced- $C_k$ -free graphs is determined for every  $k \in \mathbb{N}$  apart from  $k \in \{6, 8, 10\}$ . For the cases  $k = 8$  and  $k = 10$  the methods we use to prove Theorem 5.1.1 allow us to also prove an approximate result on the typical structure of induced- $C_k$ -free graphs. In order to state this result we require the following definitions.

Given  $\eta > 0$  and graphs  $G$  and  $G'$  on the same vertex set, we say  $G'$  is  $\eta$ -close to  $G$  if  $G'$  can be made into  $G$  by changing (i.e. adding or deleting) at most  $\eta|G|^2$  edges. We say a graph  $G$  is a *sun* if either  $G$  consists of a single vertex or  $V(G)$  can be partitioned into sets  $A, B$  such that  $E(G) = \{uv : |\{u, v\} \cap B| \leq 1\}$ . We call  $A$  the *body* of the sun and  $B$  the *side* of the sun. Note that all stars and cliques (including triangles) are suns, and that we consider a single vertex to be both a star of order one and a clique of order one.

**Theorem 5.1.2.**

- (i) *For every  $\eta > 0$ , almost all induced  $C_{10}$ -free graphs are  $\eta$ -close to graphs that can be covered by three cliques and a graph whose complement is a disjoint union of cliques.*
- (ii) *For every  $\eta > 0$ , almost all induced  $C_8$ -free graphs are  $\eta$ -close to graphs that can be covered by two cliques and a graph whose complement is a disjoint union of suns.*

We remark that in Theorems 5.1.1 and 5.1.2 we get exponential bounds on the proportion of induced- $C_{2k}$ -free graphs that do not satisfy the relevant structural description. Our proofs also show that the  $k - 2$  cliques in the covering have size close to  $n/(k - 1)$  in

Theorem 5.1.1, with analogous bounds in Theorem 5.1.2. Theorem 5.1.1 also strengthens a result by Kang, McDiarmid, Reed and Scott [40] showing that almost all induced- $C_{2k}$ -free graphs have a linear sized homogeneous set. (Their results were motivated by the Erdős-Hajnal conjecture, and actually apply to a large class of forbidden graphs  $H$ .)

It would of course be interesting to determine the typical structure of induced- $C_6$ -free graphs.

**Question 5.1.3.** *What is the typical structure of induced- $C_6$ -free graphs?*

It seems likely that almost all induced- $C_6$ -free graphs can be covered by one clique and one cograph, where a cograph is a graph not containing an induced copy of  $P_4$ . Another natural question is that of the typical structure of induced- $H$ -free graphs of a given density. In particular, an intriguing question is whether their typical structure exhibits a non-trivial ‘phase transition’ as found for triangle-free graphs [57] and more generally  $K_r$ -free graphs [12].

### 5.1.3 Overview of the chapter

A key tool in our proofs is the hypergraph container method of Balogh, Morris and Samotij [11], and independently Saxton and Thomason [70], that we utilised in Chapter 4. The precise statement of the application used here is deferred until Section 5.3.

Given a graph  $G$  and a set  $A \subseteq V(G)$ , we denote by  $G[A]$  the graph induced on  $G$  by  $A$ , and we denote the complement of  $G$  by  $\overline{G}$ . For  $k \in \mathbb{N}$  and a set  $V$  of vertices we define an *ordered  $k$ -partition* of  $V$  to be a  $k$ -partition of  $V$  such that one partition class is labelled and the rest are unlabelled. If  $Q$  is an ordered  $k$ -partition with labelled class  $Q_0$  and unlabelled classes  $Q_1, \dots, Q_{k-1}$  then we write  $Q = (Q_0, \{Q_1, \dots, Q_{k-1}\})$ .

For  $k \geq 4$ , we say that a graph  $G$  is a  $k$ -template if  $V(G)$  has an ordered  $(k-1)$ -partition  $Q = (Q_0, \{Q_1, \dots, Q_{k-2}\})$  such that  $G[Q_i]$  is a clique for all  $i \in [k-2]$  and one of the following holds.

- $k = 4$  and  $\overline{G}[Q_0]$  is a disjoint union of suns.
- $k = 5$  and  $\overline{G}[Q_0]$  is a disjoint union of stars and cliques.
- $k \geq 6$  and  $\overline{G}[Q_0]$  is a disjoint union of stars and triangles.

Clearly every  $k$ -template is induced- $C_{2k}$ -free. If  $V(G)$  has such an ordered  $(k-1)$ -partition  $Q$ , we say that  $G$  is a  $k$ -template on  $Q$ , or  $G$  has ordered  $(k-1)$ -partition  $Q$ . If  $Q'$  is the (unordered)  $(k-1)$ -partition with the same partition classes as  $Q$ , we may also say that  $G$  is a  $k$ -template on  $Q'$ . Thus Theorem 5.1.1 can be reformulated as:

*‘For  $k \geq 6$ , almost all induced  $C_{2k}$ -free graphs are  $k$ -templates.’*

Theorem 5.1.2 can be similarly reformulated in terms of 4- and 5-templates. As mentioned earlier, the main difficulty in proving Theorem 5.1.1 (compared to related results) is that typically  $G[Q_0]$  is close to, but not quite, a complete graph. This makes it very difficult to rule out other similar classes of graphs as typical structures. To overcome this we use tools such as Ramsey’s theorem to classify the graphs according to the neighbourhoods of certain vertices.

More precisely, our approach to proving the main result of this chapter is as follows. Firstly, in Section 5.3 we use the hypergraph containers result discussed above to show that almost all induced- $C_{2k}$ -free graphs are close to being a  $k$ -template, for every  $k \geq 4$  (see Lemma 5.3.1). Note that Lemma 5.3.1 immediately implies Theorem 5.1.2.



In Section 5.4 we prove upper and lower bounds on the number of  $k$ -templates on  $n$  vertices (see Lemmas 5.4.4 and 5.4.7). In Section 5.5 we prove some preliminary results about graphs that are close to being a  $k$ -template.

In Section 5.6 we state a key result which is a version of Theorem 5.1.1 with respect to a given ordered  $(k - 1)$ -partition (see Lemma 5.6.1) and use it together with Lemma 5.4.7 to derive Theorem 5.1.1. The remainder of the chapter is devoted to proving Lemma 5.6.1 via an inductive argument, which we introduce at the end of Section 5.6. This argument involves partitioning the class of graphs considered in Lemma 5.6.1 into three ‘bad’ classes of graphs, and in each of Sections 5.7, 5.8 and 5.9 we use Lemma 5.4.4 and the results in Section 5.5 to prove an upper bound on the number of graphs in a different one of these classes (see Lemmas 5.7.3, 5.8.9 and 5.9.15). In particular, Lemmas 5.7.3 and 5.8.9 already show that almost all induced- $C_{2k}$ -free graphs are ‘extremely close’ to being  $k$ -templates (see Proposition 5.9.1). In Section 5.10 we use Lemmas 5.3.1, 5.7.3, 5.8.9 and 5.9.15 to complete the inductive argument set up in Section 5.6 and so prove Lemma 5.6.1. The final section of the chapter, Section 5.11, consists of the proof of a specialised version of the Removal Lemma that we state and use in Section 5.3. Before starting on any of this however, we lay out some notation and set out some useful tools in Section 5.2, below.

## 5.2 Notation and tools

Given a graph  $G$ , a vertex  $x$ , and an ordered  $(k - 1)$ -partition  $Q = (Q_0, \{Q_1, \dots, Q_{k-2}\})$  of  $V(G)$ , we let  $N(x), \overline{N}(x)$  denote the set of neighbours and non-neighbours of  $x$  in  $G$ , respectively. We also let  $N_{Q_i}(x), \overline{N}_{Q_i}(x)$  denote the set of neighbours of  $x$  in  $Q_i$  and non-neighbours of  $x$  in  $Q_i$ , respectively. We sometimes use the notation  $d_{G,Q}^i(x) = |N_{Q_i}(x)|$  and  $\overline{d}_{G,Q}^i(x) = |\overline{N}_{Q_i}(x)|$  when we want to emphasise which graph we are working with.

For a set  $A$  of vertices in  $G$ , we define

$$N(A) := \bigcap_{v \in A} N(v), \quad \overline{N}(A) := \bigcap_{v \in A} \overline{N}(v),$$

$$N_{Q_i}(A) := \bigcap_{v \in A} N_{Q_i}(v), \quad \overline{N}_{Q_i}(A) := \bigcap_{v \in A} \overline{N}_{Q_i}(v).$$

If it does not generate any ambiguity, we may write  $N_i(x), \overline{N}_i(x), N_i(A)$  and  $\overline{N}_i(A)$  for  $N_{Q_i}(x), \overline{N}_{Q_i}(x), N_{Q_i}(A)$ , and  $\overline{N}_{Q_i}(A)$  respectively. Given  $A, B \subseteq V(G)$ , we define

$$N^*(A, B) := N(A) \cap \overline{N}(B), \quad N_i^*(A, B) := N_i(A) \cap \overline{N}_i(B).$$

In the case when  $A$  and  $B$  both have size one, containing vertices  $a, b$  respectively, we may write  $N^*(a, b)$  for  $N^*(A, B)$  and  $N_i^*(a, b)$  for  $N_i^*(A, B)$ .

Given a  $(k-1)$ -partition  $Q$  of  $[n]$  with partition classes  $Q_0, \dots, Q_{k-2}$ , and a graph  $G = (V, E)$  on vertex set  $[n]$ , and an edge or non-edge  $e = uv$  with  $u \in Q_i$  and  $v \in Q_j$ , we call  $e$  *internal* if  $i = j$ .

We denote a path on  $m$  vertices by  $P_m$ . Given a path  $P = p_1 \dots p_m$  and a sequence  $A_1, \dots, A_m$  of sets of vertices, we say that  $P$  *has type*  $A_1, \dots, A_m$  if  $p_\ell \in A_\ell$  for every  $\ell \in [m]$ . We call a graph a *linear forest* if it is a forest such that all components are paths or isolated vertices.

Given  $\ell, t \in \mathbb{N}$  we let  $R_\ell(t)$  denote the  $\ell$ -colour Ramsey number for monochromatic  $t$ -cliques, i.e.  $R_\ell(t)$  is the smallest  $N \in \mathbb{N}$  such that every  $\ell$ -colouring of the edges of  $K_N$  yields a monochromatic copy of  $K_t$ .

We define

$$n_k := \left\lceil \frac{n}{k-1} \right\rceil.$$

In a number of the proofs in this chapter we shall use the following formulation of Chernoff bounds.

**Lemma 5.2.1 (Chernoff bound).** *Let  $X$  have binomial distribution and let  $0 < a \leq \mathbb{E}[X]$ . Then*

$$(i) \ P(X > \mathbb{E}[X] + a) \leq \exp\left(-\frac{a^2}{4\mathbb{E}[X]}\right).$$

$$(ii) \ P(X < \mathbb{E}[X] - a) \leq \exp\left(-\frac{a^2}{2\mathbb{E}[X]}\right).$$

We define  $\xi(p) := -3p(\log p)/2$ . The following bounds will prove useful to us. For  $n \geq 1$  and  $3 \log n/n \leq p \leq 10^{-11}$ ,

$$\binom{n}{\leq pn} := \sum_{i=0}^{\lfloor pn \rfloor} \binom{n}{i} \leq pn \left(\frac{en}{pn}\right)^{pn} \leq 2^{\xi(p)n}, \quad (5.2.2)$$

and

$$\xi(p) \leq \frac{3}{2}p \left(\frac{1}{p}\right)^{1/8} \leq p^{3/4}. \quad (5.2.3)$$

### 5.3 Approximate structure of typical induced- $C_{2k}$ -free graphs

The main result of this section is Lemma 5.3.1, which approximately determines the typical structure of induced- $C_{2k}$ -free graphs. As mentioned earlier, we make use of a

‘containers theorem’ which reduces the proof of Lemma 5.3.1 to an extremal problem involving induced- $C_{2k}$ -free graphs. More precisely, the argument is structured as follows.

We first introduce a number of tools (see Subsection 5.3.1): a ‘containers theorem’ (Theorem 5.3.2), a Stability theorem (Theorem 5.3.3), and two Removal Lemmas (Theorem 5.3.4, Lemma 5.3.5). In Subsection 5.3.2 we use Theorem 5.3.3 to derive a Stability result involving induced- $C_{2k}$ -free graphs (Lemma 5.3.7). Similarly we use Theorem 5.3.4 to derive another specialised version of the Removal Lemma (Lemma 5.3.9). In Subsection 5.3.3 we use Theorem 5.3.2 together with Lemmas 5.3.5, 5.3.7 and 5.3.9 to determine the approximate structure of typical induced- $C_{2k}$ -free graphs.

We denote the number of (labelled) induced- $C_{2k}$ -free graphs on  $n$  vertices by  $F(n, k)$ .

**Lemma 5.3.1.** *Let  $k \geq 4$ . For every  $\eta > 0$  there exists  $\varepsilon > 0$  such that the following holds for all sufficiently large  $n$ . All but at most  $F(n, k)2^{-\varepsilon n^2}$  induced- $C_{2k}$ -free graphs on  $n$  vertices can be made into a  $k$ -template by changing at most  $\eta n^2$  edges.*

Note that Lemma 5.3.1 immediately implies Theorem 5.1.2.

### 5.3.1 Tools: containers, stability and removal lemmas

The key tool in this section is Theorem 5.3.2, which is an application of the more general theory of Hypergraph Containers developed in [11, 70]. We use the formulation of Theorem 1.5 in [70]. We require the following definitions in order to state it.

A 2-coloured multigraph  $G$  on vertex set  $[N]$  is a pair of edge sets  $G_R, G_B \subseteq [N]^{(2)}$ , which we call the red and blue edge sets respectively. If  $H$  is a fixed graph on vertex set  $[h]$ , a copy of  $H$  in  $G$  is an injection  $f : [h] \rightarrow [N]$  such that for every edge  $uv$  of  $H$ ,

$f(u)f(v) \in G_R$ , and for every non-edge  $u'v'$  of  $H$ ,  $f(u')f(v') \in G_B$ . We write  $H \subseteq G$  if  $G$  contains a copy of  $H$ , and we say that  $G$  is  $H$ -free if there are no copies of  $H$  in  $G$ . We say that  $G$  is *complete* if  $G_R \cup G_B = [N]^{(2)}$ . We denote by  $G^B$  the graph on vertex set  $[N]$  and edge set  $G_B$ .

**Theorem 5.3.2.** *Let  $H$  be a fixed graph with  $h := |V(H)|$ . For every  $\varepsilon > 0$ , there exists  $c > 0$  such that for all sufficiently large  $N$ , there exists a collection  $\mathcal{C}$  of complete 2-coloured multigraphs on vertex set  $[N]$  with the following properties.*

- (a) *For every graph  $I$  on  $[N]$  that contains no induced copy of  $H$ , there exists  $G \in \mathcal{C}$  such that  $I \subseteq G$ .*
- (b) *Every  $G \in \mathcal{C}$  contains at most  $\varepsilon N^h$  copies of  $H$ .*
- (c)  $\log |\mathcal{C}| \leq c N^{2-(h-2)/(\binom{h}{2}-1)} \log N$ .

Another tool that we will use is the following classical Stability theorem of Erdős and Simonovits (see e.g. [29, 30, 72]). By  $\text{Tu}_k(n)$  we denote the *Turán graph*, the largest complete  $k$ -partite graph on  $n$  vertices, and we define  $t_k(n) := e(\text{Tu}_k(n))$ . Given a family  $\mathcal{H}$  of fixed graphs, we say a graph  $G$  is  $\mathcal{H}$ -free if  $G$  does not contain any  $H \in \mathcal{H}$  as a (not necessarily induced) subgraph, and we say  $G$  is *induced- $\mathcal{H}$ -free* if  $G$  does not contain any  $H \in \mathcal{H}$  as an induced subgraph.

**Theorem 5.3.3.** *Let  $\mathcal{H} = \{H_1, \dots, H_\ell\}$  be a family of fixed graphs, and define  $k := \min_{1 \leq i \leq \ell} \chi(H_i)$ . For every  $\delta > 0$  there exists  $\varepsilon > 0$  such that the following holds for all sufficiently large  $n$ . If a graph  $G$  on  $n$  vertices is  $\mathcal{H}$ -free and  $e(G) \geq t_{k-1}(n) - \varepsilon n^2$ , then  $G$  can be obtained from  $\text{Tu}_{k-1}(n)$  by changing at most  $\delta n^2$  edges.*

The final tools that we introduce in this subsection are the following two Removal Lemmas. The first is an extension of the Induced Removal Lemma to families of forbidden graphs,

and is due to Alon and Shapira [7]. The original statement of this theorem also applies to infinite families of forbidden graphs, but the version for finite families is sufficient for our purposes. The second is a version of the Removal Lemma applicable to complete 2-coloured multigraphs. The proof of this is similar to that of the standard Removal Lemma, and quite long, so we defer it until the end of the chapter (see Section 5.11). For two sets  $A, B$ , we denote their symmetric difference by  $A \triangle B$ . For 2-coloured multigraphs  $G, G'$  on the same vertex set we define their distance by  $\text{dist}(G, G') := |G_R \triangle G'_R| + |G_B \triangle G'_B|$ .

**Theorem 5.3.4.** [7] *For every finite family of fixed graphs  $\mathcal{H}$  and every  $\varepsilon > 0$ , there exists  $\delta > 0$  such that the following holds for all sufficiently large  $n$ . If a graph  $G$  on  $n$  vertices contains at most  $\delta n^h$  induced copies of each graph  $H \in \mathcal{H}$  on  $h$  vertices, then  $G$  can be made induced- $\mathcal{H}$ -free by adding or deleting at most  $\varepsilon n^2$  edges.*

**Theorem 5.3.5.** *For every fixed graph  $H$  on  $\ell$  vertices, and every  $\varepsilon > 0$ , there exists  $\delta > 0$  such that the following holds for all sufficiently large  $n$ . If a complete 2-coloured multigraph  $G$  on  $n$  vertices contains at most  $\delta n^\ell$  copies of  $H$ , then there exists a complete 2-coloured multigraph  $\hat{G}$  on vertex set  $V(G)$  such that  $\hat{G}$  is  $H$ -free and  $\text{dist}(G, \hat{G}) \leq \varepsilon n^2$ .*

### 5.3.2 Stability and removal lemmas for even cycles

Suppose  $H$  is a complete 2-coloured multigraph on  $m$  vertices with  $H_R \cap H_B = \emptyset$ . If  $m = 3$  and  $|H_R| \leq 1$  we call  $H$  a *mostly blue triangle*. For  $k \in \{4, 5, 6\}$ , if  $m = 4$  and  $|H_R| \geq 6 - k$  and  $H^B$  contains a copy of  $P_4$  then we call  $H$  a *k-good tetrahedron*. The following technical proposition will be useful in proving Lemmas 5.3.7 and 5.3.9.

**Proposition 5.3.6.** *Let  $k \geq 4$  and let  $G$  be a complete 2-coloured multigraph on  $2k$  vertices. If  $G$  satisfies one of the following properties then  $G$  contains a copy of  $C_{2k}$ . Below,  $r_i$  always denotes a red edge.*

- (E1)  $G_R \triangle G_B$  is a set of at most  $k$  disjoint (red or blue) edges.
- (E2)  $G_R \triangle G_B$  is the edge set of two vertex disjoint copies of a blue  $K_k$ .
- (E3)  $G_R \triangle G_B$  is the edge set of a union of disjoint graphs  $K_3^1, K_3^2, r_1, \dots, r_{k-3}$ , where each  $K_3^i$  is a mostly blue triangle.
- (E4)  $G_R \triangle G_B$  is the edge set of a union of disjoint graphs  $K_4^1, r_1, \dots, r_{k-2}$ , where  $K_4^1$  is a 4-good tetrahedron.
- (E5)  $k \geq 5$  and  $G_R \triangle G_B$  is the edge set of a union of disjoint graphs  $K_4^1, r_1, \dots, r_{k-2}$ , where  $K_4^1$  is a 5-good tetrahedron.
- (E6)  $k \geq 6$  and  $G_R \triangle G_B$  is the edge set of a union of disjoint graphs  $K_4^1, r_1, \dots, r_{k-2}$ , where  $K_4^1$  is a 6-good tetrahedron.

*Proof.* Let  $V(G) = \{v_1, \dots, v_{2k}\}$ . Let  $C = c_1 \dots c_{2k}$  be a  $2k$ -cycle. Note that if there exists a permutation  $\sigma$  of  $[2k]$  such that for every edge  $c_i c_j \in E(C)$  we have  $v_{\sigma(i)} v_{\sigma(j)} \in G_R$  and such that for every non-edge  $c_{i'} c_{j'} \notin E(C)$  we have  $v_{\sigma(i')} v_{\sigma(j')} \in G_B$ , then  $v_{\sigma(1)} \dots v_{\sigma(2k)}$  is a copy of  $C_{2k}$  in  $G$ . We call such a permutation  $\sigma$  a *covering permutation from  $C$  to  $G$* . For ease of reading, we will write a permutation  $\sigma$  on  $[2k]$  using the notation  $\sigma = (\sigma(1), \dots, \sigma(2k))$ . If  $\sigma$  restricted to  $\{m, m+1, \dots, 2k\}$  is the identity permutation, we may simply write  $\sigma = \{\sigma(1), \dots, \sigma(m-1)\}$  instead. So for example if  $\sigma = (1, 3, 4, 2)$  is a covering permutation from  $C$  to  $G$ , then  $v_1 v_3 v_4 v_2 v_5 \dots v_{2k}$  is a copy of  $C_{2k}$  in  $G$ .

We now show that each of the properties (E1), ..., (E6) imply that there exists a covering permutation from  $C$  to  $G$ , and hence that  $G$  contains a copy of  $C_{2k}$ .

- (E1) There exists  $b, r \in \mathbb{N} \cup \{0\}$  with  $b + r \leq k$  such that, by relabelling vertices if necessary,  $G_B \setminus G_R = \{v_1 v_2, \dots, v_{2b-1} v_{2b}\}$  and  $G_R \setminus G_B = \{v_{2b+1} v_{2b+2}, \dots, v_{2(b+r)-1} v_{2(b+r)}\}$ .

Depending on the value of  $b$  we find the following covering permutations  $\sigma$  from  $C$  to  $G$ , as required.

- If  $b = 0$  then  $\sigma$  is the identity permutation.
- If  $b = 1$  then  $\sigma = (1, 3, 4, 2)$ .
- If  $b \geq 2$  then  $\sigma = (1, 3, \dots, 2b - 1, 2, 4, \dots, 2b)$ .

(E2) Let  $\{v_1, \dots, v_k\}, \{v_{k+1}, \dots, v_{2k}\}$  be the respective vertex sets of the two copies of a blue  $K_k$  in  $G_R \triangle G_B$ . Then  $\sigma = (1, k+1, 2, k+2, \dots, k, 2k)$  is a covering permutation from  $C$  to  $G$ , as required.

(E3) Let  $V(K_3^1) = \{v_1, v_2, v_3\}, V(K_3^2) = \{v_4, v_5, v_6\}$  and  $V(r_i) = \{v_{2i+5}, v_{2i+6}\}$  for every  $i \in [k-3]$ . Depending on the colour of the edges in  $K_3^1, K_3^2$  we find the following covering permutations  $\sigma$  from  $C$  to  $G$ , as required.

- If  $K_3^1, K_3^2$  both contain no red edges, then  $\sigma = (1, 4, 2, 5, 3, 6)$ .
- If  $K_3^1$  contains exactly one red edge  $v_1v_2$  and  $K_3^2$  contains no red edges, then  $\sigma = (4, 1, 2, 5, 3, 6)$ .
- If  $K_3^1$  contains exactly one red edge  $v_1v_2$  and  $K_3^2$  contains exactly one red edge  $v_5v_6$ , then  $\sigma = (1, 2, 4, 3, 5, 6)$ .

(E4) Let  $V(K_4^1) = \{v_1, v_2, v_3, v_4\}$  and  $V(r_i) = \{v_{2i+3}, v_{2i+4}\}$  for every  $i \in [k-2]$ . Depending on the configuration of red edges in  $K_4^1$  we find the following covering permutations  $\sigma$  from  $C$  to  $G$ , as required.

- If  $K_4^1$  contains exactly three red edges  $v_1v_2, v_2v_3, v_3v_4$ , then  $\sigma$  is the identity permutation.
- If  $K_4^1$  contains exactly two red edges  $v_1v_2, v_2v_3$ , then  $\sigma = (1, 2, 3, 5, 6, 4)$ .
- If  $K_4^1$  contains exactly two red edges  $v_1v_2, v_3v_4$ , then  $\sigma = (1, 2, 5, 6, 3, 4)$ .



- (E5) We may assume that  $K_4^1$  contains exactly one red edge, since that is the only case not covered by (E4). Let  $V(K_4^1) = \{v_1, v_2, v_3, v_4\}$  and  $V(r_i) = \{v_{2i+3}, v_{2i+4}\}$  for every  $i \in [k-2]$ , and let  $v_1v_2$  be the red edge in  $K_4^1$ . Then  $\sigma = (1, 2, 5, 6, 3, 7, 8, 4)$  is a covering permutation from  $C$  to  $G$ , as required.
- (E6) We may assume that  $K_4^1$  contains no red edges, since that is the only case not covered by (E5). Let  $V(K_4^1) = \{v_1, v_2, v_3, v_4\}$  and  $V(r_i) = \{v_{2i+3}, v_{2i+4}\}$  for every  $i \in [k-2]$ . Then  $\sigma = (1, 5, 6, 2, 7, 8, 3, 9, 10, 4)$  is a covering permutation from  $C$  to  $G$ , as required.

□

We now use Theorem 5.3.3 and Proposition 5.3.6 to prove the following more specialised Stability result involving  $C_{2k}$ -free 2-coloured multigraphs.

**Lemma 5.3.7.** *Let  $k \geq 4$ . For every  $\delta > 0$  there exists  $\varepsilon > 0$  such that the following holds for all sufficiently large  $n$ . If a complete 2-coloured multigraph  $G$  on vertex set  $[n]$  is  $C_{2k}$ -free and  $|G_R \cap G_B| \geq t_{k-1}(n) - \varepsilon n^2$ , then the graph  $([n], G_R \cap G_B)$  can be obtained from  $Tu_{k-1}(n)$  by changing at most  $\delta n^2$  edges.*

*Proof.* Choose  $n_0 \in \mathbb{N}$  and  $\varepsilon > 0$  such that  $1/n_0 \ll \varepsilon \ll \delta$ . Let  $n \geq n_0$ . Since  $G$  is  $C_{2k}$ -free, we know by Proposition 5.3.6 that no  $2k$  vertices of  $G$  induce on  $G$  a 2-coloured multigraph  $G'$  that satisfies (E1). So, since  $G$  is complete, the graph  $([n], G_R \cap G_B)$  must be  $Tu_k(2k)$ -free. Note that  $\chi(Tu_k(2k)) = k$ . By Theorem 5.3.3, this together with the fact that  $|G_R \cap G_B| \geq t_{k-1}(n) - \varepsilon n^2$  implies that the graph  $([n], G_R \cap G_B)$  can be obtained from  $Tu_{k-1}(n)$  by changing at most  $\delta n^2$  edges. □

The following proposition characterises the structure of graphs without  $k$ -good tetrahedrons. It will be useful in proving Lemma 5.3.9. The proof is fairly straightforward so we

give only a sketch of it here.

**Proposition 5.3.8.** *Let  $G$  be a 2-coloured multigraph with  $G_R \cap G_B = \emptyset$ .*

- (i) *If  $G$  does not contain a 6-good tetrahedron then  $G^B$  is a disjoint union of stars and triangles.*
- (ii) *If  $G$  does not contain a 5-good tetrahedron then  $G^B$  is a disjoint union of stars and cliques.*
- (iii) *If  $G$  does not contain a 4-good tetrahedron then  $G^B$  is a disjoint union of suns.*

*Proof.* (i) follows immediately from the fact that if  $G$  is 6-good tetrahedron-free then  $G^B$  does not contain a  $P_4$ .

To see (ii), note that if  $G$  is 5-good tetrahedron-free and  $P$  is a copy of  $P_4$  in  $G^B$ , then  $G^B[V(P)] = K_4$ . So every component  $H$  of  $G^B$  is either a star or a triangle or contains a  $K_4$ . But in the latter case it is easy to check that  $H$  is actually a clique.

It remains to prove (iii). If  $G$  is 4-good tetrahedron-free and  $P$  is a copy of  $P_4$  in  $G^B$ , then  $G^B[V(P)]$  is either a  $K_4$  or a copy of the graph  $K_4^-$  obtained from  $K_4$  by deleting one edge. So every component  $H$  of  $G^B$  is either a star or a clique or contains an induced copy of  $K_4^-$ . Using induction on  $|H|$ , it is not hard to show that in the latter case  $H$  must be a sun. □

We now use Theorem 5.3.4 together with Propositions 5.3.6 and 5.3.8 to prove the following more specialised Removal Lemma involving even cycles.

**Lemma 5.3.9.** *For every  $k \geq 4$  and every  $\delta > 0$  there exists  $\varepsilon > 0$  such that the following holds for all sufficiently large  $n$ . Suppose  $G$  is a complete 2-coloured multigraph on  $n$*

vertices such that  $G_R \cap G_B = E(Tu_{k-1}(n))$ . Let  $Q$  be the unique  $(k-1)$ -partition of the vertices of  $G$  such that no partition class induces an edge in  $G_R \cap G_B$ . Suppose further that  $G$  contains at most  $\varepsilon n^{2k}$  copies of  $C_{2k}$ . Then there exists a  $k$ -template  $T = (V(G), E^T)$  on  $Q$  such that  $|G_R \triangle E^T| \leq \delta n^2$ .

*Proof.* We first prove the lemma in the case  $k \geq 6$ . Choose  $n_0 \in \mathbb{N}$  and  $\varepsilon, \gamma > 0$  such that  $1/n_0 \ll \varepsilon \ll \gamma \ll \delta, 1/k$ . Let  $n \geq n_0$  and let  $Q = (Q_1, \dots, Q_{k-1})$ . Let  $c := \varepsilon^{1/3}$ .

We claim that for no two distinct  $i, j \in [k-1]$  do  $G[Q_i]$  and  $G[Q_j]$  both contain at least  $cn^k$  copies of a blue  $K_k$ . Indeed, if they do then there are at least  $c^2 n^{2k} > \varepsilon n^{2k}$  sets of  $2k$  vertices that each induce on  $G$  a 2-coloured multigraph  $G'$  that satisfies (E2). By Proposition 5.3.6 each such  $G'$  contains a copy of  $C_{2k}$ . This contradicts the assumption that  $G$  contains at most  $\varepsilon n^{2k}$  copies of  $C_{2k}$ , which proves the claim.

Thus there exists  $J \subseteq [k-1]$  with  $|J| \leq 1$  such that for all  $i \in [k-1]$  with  $i \notin J$ ,  $G[Q_i]$  contains fewer than  $cn^k$  copies of a blue  $K_k$ . Together with Theorem 5.3.4 (applied to  $G^B[Q_i]$ ) this implies that  $G[Q_i]$  can be made free of blue cliques of size  $k$  by changing the colour of at most  $\gamma n^2$  edges inside  $Q_i$ . So by Turán's Theorem, for all  $i \in [k-1]$  with  $i \notin J$ ,  $G[Q_i]$  must have at least

$$(k-1) \binom{n/(k-1)^2}{2} - 2\gamma n^2 \geq \frac{n^2}{4(k-1)^3}$$

red edges.

**Claim 1:** *There is at most one index  $i \in [k-1]$  such that  $G[Q_i]$  contains at least  $cn^3$  mostly blue triangles. Moreover, if there is such an index  $i$  then  $J \subseteq \{i\}$ , and if there is no such index then  $J = \emptyset$ .*

Indeed, suppose for a contradiction that there exist distinct  $i, j \in [k-1]$  such that  $Q_i, Q_j$

both contain at least  $cn^3$  mostly blue triangles. Note that any class that contains at least  $cn^k$  copies of a blue  $K_k$  must contain at least  $cn^3$  mostly blue triangles. So we may assume that  $J \subseteq \{i, j\}$ . Thus for every index  $\ell \neq i, j$ ,  $G[Q_\ell]$  contains at least  $n^2/(4(k-1)^3)$  red edges. Thus there are at least  $2\varepsilon n^{2k}$  sets of  $2k$  vertices that each induce on  $G$  a 2-coloured multigraph  $G'$  that satisfies (E3). (To see this, note that to choose such a set of  $2k$  vertices we may choose, for both indices  $i, j$ , the vertices of any one of the at least  $cn^3$  mostly blue triangles in  $G[Q_i], G[Q_j]$  respectively, and then choose, for each index  $\ell \neq i, j$ , any one of the at least  $n^2/(4(k-1)^3)$  red edges in  $Q_\ell$ .) By Proposition 5.3.6 each such  $G'$  contains a copy of  $C_{2k}$ . This contradicts the assumption that  $G$  contains at most  $\varepsilon n^{2k}$  copies of  $C_{2k}$ , which proves the claim.

Let  $J'$  consist of the index  $j_0 \in [k-1]$  such that  $G[Q_{j_0}]$  contains at least  $cn^3$  mostly blue triangles, if such an index exists. Otherwise let  $J' := \emptyset$ . Thus  $J \subseteq J'$ . For all  $i \in [k-1]$  with  $i \notin J'$ , Claim 1 together with Theorem 5.3.4 (applied to  $G^B[Q_i]$ ) implies that  $G[Q_i]$  can be made free of mostly blue triangles by changing the colour of at most  $\gamma n^2$  edges inside  $Q_i$ . This implies that the blue edges inside  $Q_i$  after such a change form a matching. Hence  $G[Q_i]$  contains at most  $2\gamma n^2$  blue edges.

If  $J' = \emptyset$  then  $G[Q_i]$  contains at most  $2\gamma n^2$  blue edges for all  $i \in [k-1]$ , and hence  $|G_B \setminus G_R| \leq \delta n^2$  (since  $\gamma \ll \delta, 1/k$ ). In this case we are done by setting  $T$  to be  $K_n$ . Otherwise,  $J' = \{j_0\}$  and it suffices to show that the blue edges in  $G[Q_{j_0}]$  can be made into the edge set of a disjoint collection of stars and triangles by changing the colour of at most  $\gamma n^2$  edges inside  $Q_{j_0}$ , since then we are done by setting  $T$  to be  $K_n$  minus this disjoint collection of stars and triangles.

**Claim 2(a):**  $G[Q_{j_0}]$  contains fewer than  $cn^4$  6-good tetrahedrons.

Indeed, otherwise there are at least  $\varepsilon^{1/2}n^{2k}$  sets of  $2k$  vertices that each induce on  $G$  a 2-coloured multigraph  $G'$  that satisfies (E6). (To see this, note that to choose such a set of  $2k$  vertices we may first choose the vertices of any one of the at least  $cn^4$  6-good tetrahedrons, and then choose, for each other class  $Q_i$ , any one of the at least  $n^2/(4(k-1)^3)$  red edges in  $Q_i$ .) By Proposition 5.3.6 each such  $G'$  contains a copy of  $C_{2k}$ . This contradicts the assumption that  $G$  contains at most  $\varepsilon n^{2k}$  copies of  $C_{2k}$ , which proves the claim.

Claim 2(a) together with Theorem 5.3.4 (applied to  $G^B[Q_{j_0}]$ ) implies that  $G[Q_{j_0}]$  can be made free of 6-good tetrahedrons by changing the colour of at most  $\gamma n^2$  edges inside  $Q_{j_0}$ . Proposition 5.3.8(i) implies that after such a change, all blue edges inside  $Q_{j_0}$  form a disjoint collection of stars and triangles, as required. This completes the proof in the case  $k \geq 6$ .

For the case  $k = 5$ , the proof is almost identical to the case  $k \geq 6$ , except that instead of Claim 2(a) we prove the following weaker claim, which follows in a similar way.

**Claim 2(b):**  $G[Q_{j_0}]$  contains fewer than  $cn^4$  5-good tetrahedrons.

Claim 2(b) together with Theorem 5.3.4 (applied to  $G^B[Q_{j_0}]$ ) implies that  $G[Q_{j_0}]$  can be made free of 5-good tetrahedrons by changing the colour of at most  $\gamma n^2$  edges inside  $Q_{j_0}$ . Proposition 5.3.8(ii) implies that after such a change, all blue edges inside  $Q_{j_0}$  form a disjoint collection of stars and cliques. We are now done by setting  $T$  to be  $K_n$  minus this disjoint collection of stars and cliques.

For the case  $k = 4$ , the proof is again almost identical to the case  $k \geq 6$ , except that instead of Claim 2(a) we prove the following even weaker claim, which follows in a similar way.

**Claim 2(c):**  $G[Q_{j_0}]$  contains fewer than  $cn^4$  4-good tetrahedrons.

Claim 2(c) together with Theorem 5.3.4 (applied to  $G^B[Q_{j_0}]$ ) implies that  $G[Q_{j_0}]$  can be made free of 4-good tetrahedrons by changing the colour of at most  $\gamma n^2$  edges inside  $Q_{j_0}$ . Proposition 5.3.8(iii) implies that after such a change, all blue edges inside  $Q_{j_0}$  form a disjoint collection of suns. We are now done by setting  $T$  to be  $K_n$  minus this disjoint collection of suns.  $\square$

### 5.3.3 Approximate structure of typical induced $C_{2k}$ -free graphs

We are now in a position to prove the main result of this section.

**Proof of Lemma 5.3.1.** Choose  $n_0 \in \mathbb{N}$  and  $\varepsilon, \delta, \gamma, \beta > 0$  such that  $1/n_0 \ll \varepsilon \ll \delta \ll \gamma \ll \beta \ll \eta, 1/k$ . Let  $\varepsilon' := 2\varepsilon$  and  $n \geq n_0$ . First we claim that  $F(n, k) \geq 2^{t_{k-1}(n)}$ . To see this, first note that any graph  $G$  that contains  $\overline{\text{Tu}_{k-1}(n)}$  is induced- $C_{2k}$ -free (since for any set of  $2k$  vertices on  $G$ , 3 of them must form a triangle). Moreover, there are precisely  $2^{t_{k-1}(n)}$  such graphs for any given labelling of the vertices, which proves the claim.

By Theorem 5.3.2 (with  $C_{2k}, n$  and  $\varepsilon'$  taking the roles of  $H, N$  and  $\varepsilon$  respectively) there is a collection  $\mathcal{C}$  of complete 2-coloured multigraphs on vertex set  $[n]$  satisfying properties (a)–(c). In particular, by (a), every induced- $C_{2k}$ -free graph on vertex set  $[n]$  is contained in some  $G \in \mathcal{C}$ . Let  $\mathcal{C}_1$  be the family of all those  $G \in \mathcal{C}$  for which  $|G_R \cap G_B| \geq t_{k-1}(n) - \varepsilon' n^2$ . Then the number of (labelled) induced- $C_{2k}$ -free graphs not contained in some  $G \in \mathcal{C}_1$  is at most

$$|\mathcal{C}| 2^{t_{k-1}(n) - \varepsilon' n^2} \leq 2^{-\varepsilon n^2} F(n, k),$$

because  $|\mathcal{C}| \leq 2^{n^2 - \varepsilon'}$ , by (c), and  $F(n, k) \geq 2^{t_{k-1}(n)}$ . We claim that for every  $G \in \mathcal{C}_1$

there exists a complete 2-coloured multigraph  $\tilde{G}$  and a  $k$ -template  $T$  on partition  $Q = \{Q_0, Q_1, \dots, Q_{k-2}\}$  such that

$$\tilde{G}_R \cap Q_i^{(2)} = E(T[Q_i]) \quad \text{and} \quad \tilde{G}_R \cap \tilde{G}_B \cap Q_i^{(2)} = \emptyset$$

for every  $i \in \{0, 1, \dots, k-2\}$ , and  $\text{dist}(G, \tilde{G}) \leq \eta n^2$ . (Note that this claim implies that every induced- $C_{2k}$ -free graph contained in  $G$  can be made into a  $k$ -template by changing a total of at most  $\eta n^2$  edges within the vertex classes  $Q_i$ .) Indeed, by (b), each  $G \in \mathcal{C}_1$  contains at most  $\varepsilon' n^{2k}$  copies of  $C_{2k}$ . Thus by Theorem 5.3.5 there exists a complete 2-coloured multigraph  $G'$  on the same vertex set that is  $C_{2k}$ -free, such that  $\text{dist}(G, G') \leq \delta n^2$ . Then  $|G'_R \cap G'_B| \geq t_{k-1}(n) - (\varepsilon' + \delta)n^2$ . Thus by Lemma 5.3.7 there exists a complete 2-coloured multigraph  $G''$  on the same vertex set, with  $G''_R \cap G''_B = E(\text{Tu}_{k-1}(n))$  and such that  $\text{dist}(G', G'') \leq \gamma n^2$ . Note that  $G''$  can contain at most  $\gamma n^{2k}$  copies of  $C_{2k}$ , since  $G'$  is  $C_{2k}$ -free. Let  $Q = \{Q_0, Q_1, \dots, Q_{k-2}\}$  be the unique  $(k-1)$ -partition of  $V(G'')$  such that no partition class induces an edge in  $G''_R \cap G''_B$ . Thus by Lemma 5.3.9, there exists a  $k$ -template  $T = (V(G), E^T)$  on  $Q$  such that  $|G''_R \triangle E^T| \leq \beta n^2$ . Define  $\tilde{G}$  to be the 2-coloured multigraph with  $\tilde{G}_R \cap \tilde{G}_B = G''_R \cap G''_B$  and  $\tilde{G}_R \cap Q_i = E(T[Q_i])$  for every  $i \in \{0, 1, \dots, k-2\}$ . Then  $\text{dist}(G, \tilde{G}) \leq (\delta + \gamma + \beta)n^2 \leq \eta n^2$ , and  $\tilde{G}$  satisfies the required properties. This proves the claim and thus the lemma.  $\square$

## 5.4 The number of $k$ -templates

For  $k \geq 4$  we denote the set of all  $k$ -templates on  $n$  vertices by  $T(n, k)$ . Let  $T_Q(n, k)$  denote the set of all  $k$ -templates on  $n$  vertices for which  $Q$  is an ordered  $(k-1)$ -partition. The aim of this section is to estimate  $|T_Q(n, k)|$  and  $|T(n, k)|$  (see Lemmas 5.4.4 and 5.4.7 respectively). Before we start with this we need to introduce some more notation. A

$k$ -sun is defined as follows.

- If  $k = 4$ , a  $k$ -sun is any sun (as defined in Section 5.3.3).
- If  $k = 5$ , a  $k$ -sun is a star or a clique.
- If  $k \geq 6$ , a  $k$ -sun is a star or a triangle.

Note that the results of this section are only needed for Theorem 5.1.1 (and not Theorem 5.1.2) and so we would only need to consider the case  $k \geq 6$ . However, including the cases  $k = 4, 5$  makes little difference to the proofs, and are also interesting in their own right, so we work with all  $k \geq 4$  throughout this section.

Let  $F_k(n)$  denote the set of all  $n$ -vertex graphs whose complement is a disjoint union of  $k$ -suns. Define  $f_k(n) := |F_k(n)|$ . A pair of vertices  $x, y$  is called a *twin pair* if  $N(x) \setminus \{y\} = N(y) \setminus \{x\}$ .

The following two lemmas give some estimates of the value of  $f_k(n)$ . Note that we do not make use of the upper bound in Lemma 5.4.1 anywhere, but we include it for its intrinsic interest. It would not be difficult to obtain more accurate bounds, though an asymptotic formula would probably require more work.

**Lemma 5.4.1.** *For all  $n \in \mathbb{N}$  and  $k \geq 4$ ,*

$$2^{n \log n - en \log \log n} \leq f_k(n) \leq 2^{n \log n - n \log \log n}.$$

*Proof.* Let  $P(n)$  denote the number of partitions of an  $n$  element set. It is well known (see e.g. [18]) that

$$2^{n \log n - en \log \log n} \leq P(n) \leq 2^{n \log n - n \log \log n}.$$



We will count the number  $f_k(n)$  of graphs  $G \in F_k(n)$ . Note that  $f_k(n) \geq P(n)$  follows by considering each partition class as the vertex set of a star in  $\overline{G}$ . This then immediately yields the lower bound in Lemma 5.4.1. Now note that if we choose a partition of  $[n]$  into the vertex sets of disjoint suns in  $\overline{G}$  (for which there are at most  $2^{n \log n - n \log \log n}$  choices), and then for every vertex choose whether the vertex will be in the body of its sun or side of its sun (for which there are a total of  $2^n$  choices), we can generate every possible graph  $G \in F_k(n)$  (note that some such graphs can be generated by multiple different choices). This yields the upper bound in Lemma 5.4.1.  $\square$

**Lemma 5.4.2.** *For  $k \geq 4$  and  $n > s \geq 10^7$ ,*

$$s^{s/2} \leq \frac{f_k(n)}{f_k(n-s)} \quad \text{and} \quad \frac{f_k(n)}{f_k(n-1)} \leq n^2.$$

*Proof.* By Lemma 5.4.1,

$$f_k(n) \geq f_k(s)f_k(n-s) \geq 2^{s \log s - es \log \log s} f_k(n-s) \geq 2^{s \log s/2} f_k(n-s),$$

which gives us the lower bound in the statement of the lemma.

For the upper bound, note that every graph in  $F_k(n)$  has a twin pair. For any twin pair  $i, j \in [n]$  the number of graphs in  $F_k(n)$  for which  $i, j$  are twins is at most  $2f_k(n-1)$ , since every such graph can be obtained from a graph in  $F_k(n-1)$  on vertex set  $[n] \setminus \{i\}$  by adding the vertex  $i$  and choosing whether to add the edge  $ij$  (note that all other edges incident to  $i$  are prescribed, since  $i, j$  are twins). Thus

$$f_k(n) \leq \sum_{0 < i \leq n-1} \sum_{i < j \leq n} 2f_k(n-1) \leq n^2 f_k(n-1),$$

as required.  $\square$

The following proposition can be proved by a simple but tedious calculation, which we omit here. Note that (i) was already stated in Chapter 2 (Proposition 4.4.2), but we state it again here for convenience.

**Proposition 5.4.3.** *Let  $k, n \in \mathbb{N}$  with  $n \geq k \geq 2$  and let  $0 < s < n$ .*

- (i) *Suppose  $G$  is a  $k$ -partite graph on  $n$  vertices in which some vertex class  $A$  satisfies  $|A - n/k| \geq s$ . Then*

$$e(G) \leq t_k(n) - s \left( \frac{s}{2} - k \right).$$

- (ii)  $t_{k-1}(n) \geq t_{k-1}(n - s) + sn(k - 2)/(k - 1) - s(k - 2) - t_{k-1}(s)$ .

**Lemma 5.4.4.** *Let  $k \geq 4$ . There exists  $n_0 \in \mathbb{N}$  such that for every  $n \geq n_0$  and every ordered  $(k - 1)$ -partition  $Q$  of  $[n]$ , the number of  $k$ -templates on  $Q$  satisfies*

$$|T_Q(n, k)| \leq 2^{6(\log n)^2} 2^{t_{k-1}(n)} f_k(n_k),$$

where we recall that  $n_k := \lceil n/(k - 1) \rceil$ .

*Proof.* Denote the classes of  $Q$  by  $Q_0, Q_1, \dots, Q_{k-2}$  and let  $b := ||Q_0| - \lceil \frac{n}{k-1} \rceil|$ . Then by Proposition 5.4.3(i) the number of  $k$ -templates on this partition is at most

$$f_k(|Q_0|) 2^{\sum_{0 \leq i < j \leq k-2} |Q_i||Q_j|} \leq f_k(n_k + b) 2^{t_{k-1}(n) - b(b/2 - (k-1))}.$$

Let  $h(b) := f_k(n_k + b) 2^{t_{k-1}(n) - b(b/2 - (k-1))}$ . Then by Lemma 5.4.2,

$$\frac{h(b+1)}{h(b)} \leq \left( \frac{n}{k-1} + b + 2 \right)^2 2^{-((2b+1)/2 - (k-1))}.$$

Thus  $h(b)$  is a decreasing function for  $b \geq 3 \log n$ . This together with Lemma 5.4.2 gives

us that the number of  $k$ -templates on  $Q$  is at most

$$\begin{aligned} h(b) &\leq f_k(n_k + 3 \log n) 2^{t_{k-1}(n)} \leq (n^2)^{3 \log n} 2^{t_{k-1}(n)} f_k(n_k) \\ &= 2^{6(\log n)^2} 2^{t_{k-1}(n)} f_k(n_k), \end{aligned}$$

as required. □

We call a component of a graph *non-trivial* if it contains at least 2 vertices. The proof of Lemma 5.4.7 will make use of the following proposition.

**Proposition 5.4.5.** *Let  $k \geq 4$ . There exists  $n_0 \in \mathbb{N}$  such that the following holds for every  $n \geq n_0$ . Let  $Q$  be a balanced ordered  $(k-1)$ -partition of  $[n]$ . The proportion of  $k$ -templates  $G$  on  $Q$  that are such that  $\overline{G}[Q_0]$  has at most one non-trivial component is at most  $2^{-n}$ .*

*Proof.* Since  $Q$  is balanced, the number of  $k$ -templates on  $Q$  is at least  $2^{t_{k-1}(n)} f_k(\lfloor \frac{n}{k-1} \rfloor)$ .

We can generate all possible edge sets for  $G[Q_0]$  such that  $\overline{G}[Q_0]$  has at most one non-trivial component in the following way. Note that for every such  $G[Q_0]$ ,  $\overline{G}[Q_0]$  contains at most one disjoint sun  $S$  of order at least two. For every vertex in  $Q_0$  we choose whether it will belong to the body of  $S$ , the side of  $S$ , or neither (for which there are a total of at most  $3^n$  choices). Hence the number of  $k$ -templates  $G$  on  $Q$  that are such that  $\overline{G}[Q_0]$  has at most one non-trivial component is at most  $3^n 2^{t_{k-1}(n)}$ .

Since we have by Lemma 5.4.1 that  $f_k(m) \geq 2^{m \log m - em \log \log m}$  for all  $m \in \mathbb{N}$ , the result follows (with some room to spare). □

The following trivial observation will be useful in the proof of Lemma 5.4.7.

If a graph  $G$  is a disjoint union of suns then  $G$  contains no induced 4-cycles. (5.4.6)

**Lemma 5.4.7.** *For every  $k \geq 4$  there exists  $n_0 \in \mathbb{N}$  such that the following holds for all  $n \geq n_0$ , where we recall that  $n_k := \lceil n/(k-1) \rceil$ . The number of  $k$ -templates on vertex set  $[n]$  satisfies*

$$|T(n, k)| \geq \frac{(k-1)^n}{2(k-2)!n^k} 2^{t_{k-1}(n)} f_k(n_k).$$

*Proof.* Choose  $n_0$  such that  $1/n_0 \ll 1/k$ , and let  $n \geq n_0$ . Given a  $k$ -template  $G$  on vertex set  $[n]$  and an ordered  $(k-1)$ -partition  $Q = (Q_0, \{Q_1, \dots, Q_{k-2}\})$  of  $[n]$ , we say that  $G$  is  $Q$ -compatible if  $G$  is a  $k$ -template on  $Q$  and the following hold:

( $\alpha$ ) Whenever  $\ell \leq 2k$  and  $0 \leq i \leq k-2$  and  $v_1, v_2, \dots, v_\ell \in V(G) \setminus Q_i$ , we have that

$$|\overline{N}_{Q_i}(\{v_1, v_2, \dots, v_\ell\})| \geq \frac{n}{2^{\ell+1}(k-1)}.$$

( $\beta$ )  $\overline{G}[Q_0]$  has at least 2 non-trivial components.

**Claim 1:** *Given a balanced ordered  $(k-1)$ -partition  $Q = (Q_0, \{Q_1, \dots, Q_{k-2}\})$  of  $[n]$ , the number of  $Q$ -compatible  $k$ -templates  $G$  on vertex set  $[n]$  is at least  $2^{t_{k-1}(n)-1} f_k(n_k)/n^2$ .*

Indeed, consider a random graph  $G$  where for each potential crossing edge with respect to  $Q$  we choose the edge to be present or not, each with probability  $1/2$ , independently; we let  $G[Q_0]$  be one of the  $f_k(|Q_0|)$  graphs in  $F_k(|Q_0|)$ , chosen uniformly at random; and we choose all edges to be present inside  $Q_i$  for every  $i > 0$ . So each  $k$ -template on  $Q$  is equally likely to be generated. Note that the number of potential crossing edges with respect to  $Q$  is  $2^{t_{k-1}(n)}$ . This together with Lemma 5.4.2 implies that the number of graphs in the

probability space is at least  $2^{t_{k-1}(n)} f_k(n_k)/n^2$ . By Lemma 5.2.1(ii) and Proposition 5.4.5 respectively, we have that at least half of all graphs  $G$  in the probability space satisfy  $(\alpha)$  and  $(\beta)$ , which proves the claim.

**Claim 2:** *Given two balanced ordered  $(k-1)$ -partitions  $Q = (Q_0, \{Q_1, \dots, Q_{k-2}\})$  and  $Q' = (Q'_0, \{Q'_1, \dots, Q'_{k-2}\})$  of  $[n]$ , and a  $k$ -template  $G$  on  $[n]$  that is  $Q$ -compatible and  $Q'$ -compatible, there exist  $k$  vertices  $u_0, v_0, v_1, \dots, v_{k-2} \in [n]$  such that  $G[\{u_0, v_0, v_1, \dots, v_{k-2}\}]$  contains exactly one edge  $u_0 v_0$  and  $u_0 \in Q_0 \cap Q'_0$  and  $v_i \in Q_i \cap Q'_i$  for all  $i \geq 0$ .*

To show this, we first choose a set  $U = \{u_{0,1}, w_{0,1}, u_{0,2}, w_{0,2}, u_1, w_1, \dots, u_{k-2}, w_{k-2}\}$  of  $2k$  vertices such that  $u_{0,1}, w_{0,1}, u_{0,2}, w_{0,2} \in Q_0$  and  $u_i, w_i \in Q_i$  for every  $i > 0$  and

$$E(G[U]) = \{u_{0,1}u_{0,2}, u_{0,2}w_{0,1}, w_{0,1}w_{0,2}, w_{0,2}u_{0,1}, u_1w_1, \dots, u_{k-2}w_{k-2}\}.$$

This is possible since  $G$  satisfies  $(\alpha), (\beta)$  with respect to  $Q$ . Now if there exist distinct  $i, j > 0$  such that  $u_i, w_i, u_j, w_j \in Q'_0$  then  $\overline{G}[Q'_0]$  contains the induced 4-cycle  $u_i u_j w_i w_j$ , which by (5.4.6) contradicts the fact that  $G$  is a  $k$ -template on  $Q'$ . So, by relabelling vertices if necessary, we may assume that  $u_2, \dots, u_{k-2} \notin Q'_0$ . If  $u_{0,1}, w_{0,1} \notin Q'_0$  then by the pigeon-hole principle there must exist  $i > 0$  such that  $Q'_i$  contains at least 2 elements of  $\{u_{0,1}, w_{0,1}, u_2, \dots, u_{k-2}\}$ , contradicting the assumption that  $G[Q'_i]$  is a clique. So, by relabelling vertices if necessary, we may assume that  $u_{0,1} \in Q'_0$ , and similarly that  $u_{0,2} \in Q'_0$ . Now if  $u_1, w_1 \in Q'_0$  then  $\overline{G}[Q'_0]$  contains the induced 4-cycle  $u_{0,1} u_1 u_{0,2} w_1$ , which by (5.4.6) contradicts the fact that  $G$  is a  $k$ -template on  $Q'$ . So, by relabelling vertices if necessary, we may assume that  $u_1 \notin Q'_0$ , and thus  $u_1, \dots, u_{k-2} \notin Q'_0$ . Recall that for all  $i > 0$ ,  $G[Q'_i]$  is a clique, so  $Q'_i$  can contain at most one vertex in  $\{u_1, \dots, u_{k-2}\}$ . Thus we may assume, by relabelling indices if necessary, that  $u_{0,1}, u_{0,2} \in Q_0 \cap Q'_0$  and  $u_i \in Q_i \cap Q'_i$  for every  $i > 0$ . So setting  $u_0 := u_{0,1}, v_0 := u_{0,2}$  and  $v_i := u_i$  for all  $i > 0$  yields the

required set of vertices.

**Claim 3:** *If there exist balanced ordered  $(k-1)$ -partitions  $Q = (Q_0, \{Q_1, \dots, Q_{k-2}\})$  and  $Q' = (Q'_0, \{Q'_1, \dots, Q'_{k-2}\})$  of  $[n]$ , and a  $k$ -template  $G$  on  $[n]$  that is both  $Q$ -compatible and  $Q'$ -compatible, then  $Q = Q'$ .*

Consider any  $k$  vertices  $u_0, v_0, \dots, v_{k-2} \in V(G)$  that are such that  $G[\{u_0, v_0, v_1, \dots, v_{k-2}\}]$  contains exactly one edge  $u_0v_0$  and  $u_0 \in Q_0 \cap Q'_0$  and  $v_i \in Q_i \cap Q'_i$  for all  $i \geq 0$ . Such vertices exist by Claim 2. For  $i > 0$  define

$$\overline{N}_i := \overline{N}_{Q_i}(\{u_0, v_0, \dots, v_{k-2}\} \setminus \{v_i\}).$$

$$\overline{N}'_i := \overline{N}_{Q'_i}(\{u_0, v_0, \dots, v_{k-2}\} \setminus \{v_i\}).$$

Since both  $\overline{N}_i$  and  $\overline{N}'_i$  are subsets of the common non-neighbourhood of  $\{u_0, v_0, v_1, \dots, v_{k-2}\} \setminus \{v_i\}$ , neither can intersect  $Q_j$  or  $Q'_j$  for  $j \notin \{0, i\}$ . Note that all vertices in  $\overline{N}_i$  are adjacent. Thus  $|\overline{N}_i \cap Q'_0| \leq 1$ , since otherwise  $\overline{G}[Q'_0]$  contains an induced 4-cycle on  $u_0, v_0$  together with 2 vertices from  $\overline{N}_i$ , which by (5.4.6) contradicts the fact that  $G$  is a  $k$ -template on  $Q'$ . Similarly,  $|\overline{N}'_i \cap Q_0| \leq 1$ . Define

$$\overline{N}_i^\dagger := (\overline{N}_i \cup \overline{N}'_i) \setminus (Q_0 \cup Q'_0).$$

Then  $\overline{N}_i^\dagger \subseteq Q_i \cap Q'_i$ .

Now we consider any vertex  $w \in Q_0$ . Since  $G$  satisfies  $(\alpha)$  with respect to  $Q$ , we have

that for every  $i > 0$ ,

$$\begin{aligned} |\overline{N}_{Q'_i}(w)| &\geq |\overline{N}(w) \cap \overline{N}_i^\dagger| \geq |\overline{N}(w) \cap \overline{N}_i| - 1 \\ &= |\overline{N}_{Q_i}(\{u_0, v_0, \dots, v_{k-2}, w\} \setminus \{v_i\})| - 1 \geq \frac{n}{2^{k+1}(k-1)} - 1 \geq 1. \end{aligned} \quad (5.4.8)$$

Thus  $w$  must belong to  $Q'_0$ , since  $G[Q'_i]$  is a clique for every  $i > 0$ . Hence  $Q_0 \subseteq Q'_0$ . In the same way we can show that  $Q'_0 \subseteq Q_0$ . Thus  $Q_0 = Q'_0$ .

Now we consider any vertex  $w \in Q_j$ , for  $j > 0$ . Since  $G$  satisfies  $(\alpha)$  with respect to  $Q$ , we have (similarly to (5.4.8)) that for every  $i \neq j$  with  $i > 0$ ,

$$|\overline{N}_{Q'_i}(w)| \geq |\overline{N}(w) \cap \overline{N}_i^\dagger| \geq 1.$$

Thus  $w \in Q'_0 \cup Q'_j$ . Together with the fact that  $Q_0 = Q'_0$  this implies that  $w \in Q'_j$ . Thus  $Q_j \subseteq Q'_j$  for all  $j > 0$ .

Hence  $Q = Q'$ , which proves the claim.

We now count the number of balanced ordered  $(k-1)$ -partitions. Since the vertex classes of a balanced ordered  $(k-1)$ -partition of  $[n]$  have sizes  $\lceil \frac{n}{k-1} \rceil, \lceil \frac{n-1}{k-1} \rceil, \dots, \lceil \frac{n-k+2}{k-1} \rceil$ , the number of such  $(k-1)$ -partitions is

$$\frac{1}{(k-2)!} \binom{n}{\lceil \frac{n}{k-1} \rceil, \lceil \frac{n-1}{k-1} \rceil, \dots, \lceil \frac{n-k+2}{k-1} \rceil}.$$

This together with Claims 1 and 3 implies that

$$|T(n, k)| \geq \frac{1}{2(k-2)!n^2} \binom{n}{\lceil \frac{n}{k-1} \rceil, \lceil \frac{n-1}{k-1} \rceil, \dots, \lceil \frac{n-k+2}{k-1} \rceil} 2^{t_{k-1}(n)} f_k(n_k). \quad (5.4.9)$$

Now note that if  $a_1 + \dots + a_{k-1} = n$ , then  $\binom{n}{a_1, a_2, \dots, a_{k-1}}$  is maximized by taking  $a_j := \lceil \frac{n-j+1}{k-1} \rceil$  for every  $j$ . This implies that

$$(k-1)^n = \sum_{a_1 + \dots + a_{k-1} = n} \binom{n}{a_1, a_2, \dots, a_{k-1}} \leq n^{k-2} \binom{n}{\lceil \frac{n}{k-1} \rceil, \lceil \frac{n-1}{k-1} \rceil, \dots, \lceil \frac{n-k+2}{k-1} \rceil},$$

which together with (5.4.9) implies the result.  $\square$

## 5.5 Properties of near- $k$ -templates

In this section we collect some properties of graphs which are close to being  $k$ -templates. In particular, when  $k \geq 6$ , this means we consider graphs  $G$  which have a vertex partition such that each vertex class induces on  $G$  an almost complete graph. (As in the previous section, we will need the results of this section for the main results of this chapter only for the case  $k \geq 6$ , but we prove the results for all  $k \geq 4$  since it makes little difference to the proofs.) More formally, given  $k \geq 4$ , a graph  $G$  on vertex set  $[n]$ , and an ordered  $(k-1)$ -partition  $Q$  of  $[n]$  we define

$$h(Q, G) := \sum_{i=0}^{k-2} |E(\overline{G}[Q_i])|.$$

We say  $Q$  is an *optimal ordered  $(k-1)$ -partition of  $G$*  if  $h(Q, G)$  is the minimum value  $h(Q', G)$  takes over all partitions  $Q'$  of  $[n]$ . Note that if  $h(Q, G) = 0$  then  $G$  is a  $k$ -template on  $Q$ , and that the following also holds.

$$\text{If } k \geq 6 \text{ then every } k\text{-template } G' \text{ on } Q \text{ satisfies } h(Q, G') \leq n. \quad (5.5.1)$$

Note that (5.5.1) does not hold for  $k \in \{4, 5\}$ . We will require the following definitions in what follows.



- Recall that  $F(n, k)$  denotes the set of all labelled induced- $C_{2k}$ -free graphs on vertex set  $[n]$ .
- Given  $n \in \mathbb{N}$ ,  $k \geq 4$ , and  $\eta > 0$ , we define  $F(n, k, \eta) \subseteq F(n, k)$  to be the set of all graphs in  $F(n, k)$  such that  $h(Q, G) \leq \eta n^2$  for some optimal ordered  $(k-1)$ -partition  $Q$  of  $G$ .
- Given further an ordered  $(k-1)$ -partition  $Q = (Q_0, \{Q_1, \dots, Q_{k-2}\})$  of  $[n]$  we define  $F_Q(n, k) \subseteq F(n, k)$  to be the set of all graphs in  $F(n, k)$  for which  $Q$  is an optimal ordered  $(k-1)$ -partition
- Similarly we define  $F_Q(n, k, \eta) \subseteq F(n, k, \eta)$  to be the set of all graphs in  $F(n, k, \eta)$  for which  $Q$  is an optimal ordered  $(k-1)$ -partition.

Recall that, given a graph  $G$  on vertex set  $[n]$  and an index  $i \in \{0, 1, \dots, k-2\}$ , we let  $d_{G,Q}^i(x), \bar{d}_{G,Q}^i(x)$  denote the number of neighbours and non-neighbours of  $x$  in  $Q_i$ , respectively. The following proposition follows immediately from the definition of optimality.

**Proposition 5.5.2.** *Let  $k \geq 4$ , let  $\eta > 0$ , let  $Q = (Q_0, \{Q_1, \dots, Q_{k-2}\})$  be an ordered  $(k-1)$ -partition of  $[n]$ , and let  $G \in F_Q(n, k, \eta)$ . For any two distinct indices  $i, j \in \{0, 1, \dots, k-2\}$  every vertex  $x \in Q_i$  satisfies  $\bar{d}_{G,Q}^j(x) \geq \bar{d}_{G,Q}^i(x)$ .*

Next we show that for most graphs which are close to being  $k$ -templates, the bipartite graphs between the partition classes are quasirandom.

Given  $k \geq 4$ ,  $\nu = \nu(n) > 0$  and an ordered  $(k-1)$ -partition  $Q = (Q_0, \{Q_1, \dots, Q_{k-2}\})$  of  $[n]$ , we define the following properties that a graph on vertex set  $[n]$  may satisfy with respect to  $Q$ .

(F1) $_{\nu}$  If  $U_i \subseteq Q_i$  and  $U_j \subseteq Q_j$  with  $|U_i||U_j| \geq \nu^2 n^2$  for distinct  $0 \leq i, j \leq k-2$ , then

$$\frac{1}{4} \leq \frac{|e(U_i, U_j)|}{|U_i||U_j|} \leq \frac{3}{4}.$$

(F2) $_{\nu}$   $||Q_i| - \frac{n}{k-1}| \leq \nu n$  for every  $0 \leq i \leq k-2$ .

Given  $\eta, \mu > 0$  we define  $F_Q(n, k, \eta, \mu)$  to be the set of all graphs in  $F_Q(n, k, \eta)$  that satisfy (F1) $_{\mu}$  and (F2) $_{\mu}$  with respect to  $Q$ .

**Lemma 5.5.3.** *Let  $n \geq k \geq 4$ , let  $0 < \eta < 1$ , let  $6k/n \leq \nu = \nu(n) \leq 1$ , let  $6 \log n \leq m = m(n) \leq 10^{-11}n^2$ , and let  $Q = (Q_0, \{Q_1, \dots, Q_{k-2}\})$  be an ordered  $(k-1)$ -partition of  $[n]$ . Then the following hold (recall that  $\xi(p)$  was defined at the end of Section 5.2).*

- (i) *The number of graphs  $G$  in  $F_Q(n, k, \eta)$  that fail to satisfy (F1) $_{\nu}$  with respect to  $Q$  and have at most  $m$  internal non-edges is at most  $2^{t_{k-1}(n) + \xi(m/n^2)n^2} 2^{2n+1} \exp(-\nu^2 n^2/32)$ .*
- (ii) *The number of graphs  $G$  in  $F_Q(n, k, \eta)$  that fail to satisfy (F2) $_{\nu}$  with respect to  $Q$  and have at most  $m$  internal non-edges is at most  $2^{t_{k-1}(n) + \xi(m/n^2)n^2} \exp(-\nu^2 n^2/6)$ .*

*Proof.* For both (i) and (ii) we consider constructing such a graph  $G$ . By (5.2.2) there are at most  $\binom{n^2}{\leq m} \leq 2^{\xi(m/n^2)n^2}$  choices for the internal edges of  $G$ .

We first prove (i). For a given choice of internal edges, consider the random graph  $H$  where for each possible crossing edge with respect to  $Q$  we choose the edge to be present or not, with probability  $1/2$ , independently. Note that the total number of ways to choose the crossing edges is at most  $2^{t_{k-1}(n)}$ , and each possible configuration of crossing edges is equally likely. So an upper bound on the number of graphs  $G \in F_Q(n, k, \eta)$  that fail to satisfy property (F1) $_{\nu}$  with respect to  $Q$  and that have at most  $m$  internal non-edges is

$$2^{t_{k-1}(n) + \xi(m/n^2)n^2} \mathbb{P}(H \text{ fails to satisfy (F1)}_{\nu} \text{ with respect to } Q). \quad (5.5.4)$$

Note that the number of choices for  $U_i \subseteq Q_i, U_j \subseteq Q_j$  with  $|U_i||U_j| \geq \nu^2 n^2$  is at most  $2^{2n}$  and that  $\mathbb{E}(e(U_i, U_j)) = |U_i||U_j|/2 \geq \nu^2 n^2/2$ . Hence by Lemma 5.2.1,

$$\mathbb{P}(H \text{ fails to satisfy (F1)}_\nu \text{ with respect to } Q) \leq 2^{2n+1} \exp\left(-\frac{\nu^2 n^2}{32}\right).$$

This together with (5.5.4) yields the result.

We now prove (ii). If  $||Q_i| - \frac{n}{k-1}| > \nu n$  for some  $0 \leq i \leq k-2$ , then by Proposition 5.4.3(i) the number of crossing edges in  $G$  is at most

$$t_{k-1}(n) - \frac{\nu^2 n^2}{3}.$$

We can conclude that the number of  $G \in F_Q(n, k, \eta)$  that fail to satisfy (F2) $_\nu$  with respect to  $Q$  and that have at most  $m$  internal non-edges is at most

$$2^{\xi(m/n^2)n^2} 2^{t_{k-1}(n) - \frac{\nu^2 n^2}{3}} \leq 2^{t_{k-1}(n) + \xi(m/n^2)n^2} \exp\left(-\frac{\nu^2 n^2}{6}\right),$$

as required. □

We will apply the following special case of Lemma 5.5.3 in Section 5.10 in the proof of Lemma 5.6.1.

**Corollary 5.5.5.** *Let  $k \geq 4$  and let  $0 < \eta, \mu < 10^{-11}$  be such that  $\mu^2 > 24\xi(\eta)$ . There exists an integer  $n_0 = n_0(\mu, k)$  such that for all  $n \geq n_0$  and every ordered  $(k-1)$ -partition  $Q$  of  $[n]$ ,*

$$|F_Q(n, k, \eta) \setminus F_Q(n, k, \eta, \mu)| \leq 2^{t_{k-1}(n) - \frac{\mu^2 n^2}{100}}.$$

*Proof.* We choose  $n_0$  such that  $1/n_0 \ll \eta, \mu, 1/k$ . Applying Lemma 5.5.3 with  $\mu, \eta n^2$

playing the roles of  $\nu, m$  respectively yields that

$$|F_Q(n, k, \eta) \setminus F_Q(n, k, \eta, \mu)| \leq 2^{t_{k-1}(n) + \xi(\eta)n^2} 2^{2n+1} \left( e^{-\frac{\mu^2 n^2}{6}} + e^{-\frac{\mu^2 n^2}{32}} \right) \leq 2^{t_{k-1}(n) - \frac{\mu^2 n^2}{100}},$$

as required.  $\square$

The next proposition follows immediately from [10, Lemma 2.22]. We will use it to find induced copies of  $C_{2k}$ . (Usually  $T$  will be a suitable induced subgraph of  $C_{2k}$  and the  $A_i, B_i$  will be the intersection of (non-)neighbourhoods of vertices that we have already embedded.)

**Proposition 5.5.6.** *Let  $n_0, k \in \mathbb{N}$  and  $\eta, \mu > 0$  be chosen such that  $k \geq 4$  and  $1/n_0 \ll \eta \ll \mu \ll 1/k$ . Then the following holds for all  $n \in \mathbb{N}$  with  $n \geq n_0$ . Let  $Q = (Q_0, \{Q_1, \dots, Q_{k-2}\})$  be an ordered  $(k-1)$ -partition of  $[n]$  and suppose  $G \in F_Q(n, k, \eta, \mu)$ . Let  $I \subseteq \{0, 1, \dots, k-2\}$ . For every  $i \in I$  let  $A_i, B_i \subseteq Q_i$  be disjoint with  $|A_i|, |B_i| \geq \mu^{1/2}n$ . Let  $T$  be a  $2|I|$ -vertex graph with a perfect matching whose edges are  $v_i u_i$  for every  $i \in I$ . Then there exists an injection  $f : V(T) \rightarrow V(G)$  such that  $f(v_i) \in A_i, f(u_i) \in B_i$  for every  $i \in I$ , and  $f(V(T))$  induces on  $G$  a copy of  $T$ .*

Finally we show that if  $G$  is close to being a  $k$ -template then removing a small number of vertices from  $G$  does not alter its optimal ordered  $(k-1)$ -partition very much.

Given  $m, n \in \mathbb{N}$  and an ordered  $(k-1)$ -partition  $Q$  of  $[n]$ , we define  $\mathcal{P}(Q, m)$  to be the collection of all ordered  $(k-1)$ -partitions of  $[n]$  that can be obtained from  $Q$  by moving at most  $m$  vertices between partition classes, and possibly choosing a different partition class to be the labelled one. Then it is easy to see that

$$|\mathcal{P}(Q, m)| \leq k \binom{n}{m} k^m \leq k \left( \frac{ekn}{m} \right)^m \leq 2^{m \log(ek^2 n/m)}. \quad (5.5.7)$$

Given an ordered  $(k-1)$ -partition  $Q$  of  $[n]$  and a set  $S \subseteq [n]$ , let  $Q - S$  denote the ordered  $(k-1)$ -partition (possibly with some empty classes) obtained from  $Q$  by deleting all elements of  $S$  from their partition classes.

**Lemma 5.5.8.** *Let  $k \geq 4$ , let  $0 < \eta, \mu \leq 1/k^3$ , let  $0 < \nu = \nu(n) \leq 1/k^3$ , and let  $0 \leq m = m(n) \leq n^2$  with  $\nu^2 > 4m/n^2$  for all  $n \in \mathbb{N}$ . There exists  $n_0 \in \mathbb{N}$  such that the following holds for all  $n \geq n_0$ . Let  $Q = (Q_0, \{Q_1, \dots, Q_{k-2}\})$  be an ordered  $(k-1)$ -partition of  $[n]$  and let  $S \subseteq [n]$  with  $|S| \leq n/k^2$ . Then for every  $G \in F_Q(n, k, \eta, \mu)$  that satisfies  $(F1)_\nu$  with respect to  $Q$  and that has at most  $m$  internal non-edges, every optimal ordered  $(k-1)$ -partition of  $G - S$  is an element of  $\mathcal{P}(Q - S, k^4\nu^2n)$ .*

*Proof.* Let  $G \in F_Q(n, k, \eta, \mu)$  have at most  $m$  internal non-edges and satisfy  $(F1)_\nu$  with respect to  $Q$ , and let  $Q' = (Q'_0, \{Q'_1, \dots, Q'_{k-2}\})$  be an optimal ordered  $(k-1)$ -partition of  $G - S$ . By optimality of  $Q'$  it must be that  $G - S$  has at most  $m$  internal non-edges with respect to  $Q'$ .

For every  $i \in \{0, 1, \dots, k-2\}$ , since  $|Q_i - S| \geq n/(k-1) - \mu n - n/k^2 \geq n/k$ , the pigeon-hole principle implies that there exists  $j \in \{0, 1, \dots, k-2\}$  such that  $|Q_i \cap Q'_j| \geq n/k^2$ . We define a function  $\sigma$  by setting  $\sigma(i)$  to be an index in  $\{0, 1, \dots, k-2\}$  that satisfies  $|Q_i \cap Q'_{\sigma(i)}| \geq n/k^2$ , for every  $i \in \{0, 1, \dots, k-2\}$ . Suppose for a contradiction that there exists  $i' \in \{0, 1, \dots, k-2\}$  with  $i \neq i'$  such that  $|Q_{i'} \cap Q'_{\sigma(i)}| \geq k^2\nu^2n$ . Then since  $G$  satisfies  $(F1)_\nu$  with respect to  $Q$  we have that the number of internal non-edges in  $G - S$  with respect to  $Q'$  is at least  $|Q_i \cap Q'_{\sigma(i)}||Q_{i'} \cap Q'_{\sigma(i)}|/4 \geq \nu^2n^2/4 > m$ . This contradicts our previous observation that  $G - S$  has at most  $m$  internal non-edges with respect to  $Q'$ . Hence  $\sigma$  is a permutation on  $\{0, 1, \dots, k-2\}$ . Moreover  $|Q_i \cap Q'_j| < k^2\nu^2n$  for all  $j \in \{0, 1, \dots, k-2\}$  with  $j \neq \sigma(i)$ .

Let  $\mathcal{P}$  be the set of all ordered  $(k-1)$ -partitions of  $[n] \setminus S$  for which such a permutation exists. So by the above we have that for every  $G \in F_Q(n, k, \eta)$  that satisfies  $(F1)_\nu$  with

respect to  $Q$  and that has at most  $m$  internal non-edges, every optimal ordered  $(k-1)$ -partition of  $G-S$  is an element of  $\mathcal{P}$ . So it remains to show that  $\mathcal{P} \subseteq \mathcal{P}(Q-S, k^4\nu^2n)$ . This follows from the observation that every element of  $\mathcal{P}$  can be obtained by starting with the (labelled)  $(k-1)$ -partition  $Q_0 \setminus S, Q_1 \setminus S, \dots, Q_{k-2} \setminus S$ , applying a permutation of  $\{0, 1, \dots, k-2\}$  to the partition class labels, then for every ordered pair of partition classes moving at most  $k^2\nu^2n$  elements from the first partition class to the second, and finally unlabelling all but one of the resulting partition classes.  $\square$

The following is an immediate corollary of Lemma 5.5.8, applied with  $\mu, \eta n^2$  playing the roles of  $\nu, m$ , respectively.

**Corollary 5.5.9.** *Let  $k \geq 4$  and  $0 < \eta, \mu < 1/k^3$  with  $\mu^2 > 4\eta$ . There exists  $n_0 \in \mathbb{N}$  such that the following holds for all  $n \geq n_0$ . Let  $Q = (Q_0, \{Q_1, \dots, Q_{k-2}\})$  be an ordered  $(k-1)$ -partition of  $[n]$  and let  $S \subseteq [n]$  with  $|S| \leq n/k^2$ . Then for every  $G \in F_Q(n, k, \eta, \mu)$ , every optimal ordered  $(k-1)$ -partition of  $G-S$  is an element of  $\mathcal{P}(Q-S, k^4\mu^2n)$ .*

## 5.6 Derivation of Theorem 5.1.1 from the main lemma

The following lemma is the key result in our proof of Theorem 5.1.1. Together with Lemma 5.4.4 it implies that, for  $k \geq 6$ , almost all induced- $C_{2k}$ -free graphs  $G$  with a given optimal ordered  $(k-1)$ -partition are  $k$ -templates. Recall that  $n_k := \lceil n/(k-1) \rceil$ , that  $f_k(n)$  and  $T_Q(n, k)$  were defined at the beginning of Section 5.4, and that  $F_Q(n, k)$  was defined at the beginning of Section 5.5.

**Lemma 5.6.1.** *For every  $n, k \in \mathbb{N}$  with  $k \geq 6$  there exists  $C \in \mathbb{N}$  such that the following holds. For every ordered  $(k-1)$ -partition  $Q$  of  $[n]$ ,*

$$|F_Q(n, k)| \leq |T_Q(n, k)| + 5C2^{-n^{\frac{1}{2k^2}}/3} f_k(n_k) 2^{t_{k-1}(n)}.$$

Lemma 5.6.1 will be proved in the remaining sections of this chapter. We will now use it to derive Theorem 5.1.1.

**Proof of Theorem 5.1.1.** Let  $n_0 \in \mathbb{N}$  be as in Lemma 5.4.7, let  $C \in \mathbb{N}$  be as in Lemma 5.6.1, let  $n_1 \in \mathbb{N}$  satisfy  $1/n_1 \ll 1/k$ , let  $n \in \mathbb{N}$  with  $n \geq \max\{n_0, n_1\}$ , and let  $\mathcal{Q}$  be the set of all ordered  $(k-1)$ -partitions of  $[n]$ . Since  $T(n, k) \subseteq F(n, k)$  and  $T_Q(n, k) \subseteq F_Q(n, k)$  for every  $Q \in \mathcal{Q}$ , Lemma 5.6.1 implies that

$$\begin{aligned} |F(n, k)| - |T(n, k)| &= |F(n, k) \setminus T(n, k)| \leq \sum_{Q \in \mathcal{Q}} |F_Q(n, k) \setminus T_Q(n, k)| \\ &= \sum_{Q \in \mathcal{Q}} (|F_Q(n, k)| - |T_Q(n, k)|) \leq 5C(k-1)^n 2^{-n^{\frac{1}{2k^2}}/3} f_k(n_k) 2^{t_{k-1}(n)} \\ &\leq C 2^{-n^{\frac{1}{2k^2}}/4} \frac{(k-1)^n}{2(k-2)!n^k} f_k(n_k) 2^{t_{k-1}(n)}. \end{aligned}$$

This together with Lemma 5.4.7 implies that

$$|F(n, k)| - |T(n, k)| \leq C 2^{-n^{\frac{1}{2k^2}}/4} |T(n, k)| = o(|T(n, k)|),$$

where we use the little  $o$  notation with respect to  $n$ . So  $|F(n, k)| = (1 + o(1))|T(n, k)|$ , as required.  $\square$

Sections 5.7–5.10 are devoted to proving Lemma 5.6.1 by an inductive argument. Throughout Sections 5.7–5.10 we fix constants  $C, k, n_0 \in \mathbb{N}$  with  $k \geq 6$  and  $\varepsilon, \eta, \mu, \gamma, \beta, \alpha > 0$  such that

$$\frac{1}{C} \ll \frac{1}{n_0} \ll \varepsilon \ll \eta \ll \mu \ll \gamma \ll \beta \ll \alpha \ll \frac{1}{k}. \quad (5.6.2)$$

We also set  $M := R_{2k-2}(\lceil \frac{1}{\gamma} \rceil) + 1$ , fix an arbitrary integer  $n \geq n_0$ , and fix an arbitrary ordered  $(k-1)$ -partition  $Q = (Q_0, \{Q_1, \dots, Q_{k-2}\})$  of  $[n]$ .

We make the following inductive assumption in Sections 5.7, 5.8 and 5.9: for every  $n' \leq n - 1$ , and every ordered  $(k - 1)$ -partition  $Q' = (Q'_0, \{Q'_1, \dots, Q'_{k-2}\})$  of  $[n']$ ,

$$|F_{Q'}(n', k) \setminus T_Q(n', k)| \leq 5C2^{-(n')^{\frac{1}{2k^2}}/3} f_k(n'_k) 2^{t_{k-1}(n')}.$$

Note that this together with Lemma 5.4.4 implies that

$$|F_{Q'}(n', k)| \leq 6C2^{6(\log n')^2} f_k(n'_k) 2^{t_{k-1}(n')}. \quad (5.6.3)$$

We now give a number of definitions that will be used in Sections 5.7–5.10. Given an index  $i \in \{0, 1, \dots, k - 2\}$ , we call a vertex  $x$  of a graph  $G$  *i-light* if at least one of the following holds.

$$(A1) \quad d_{G,Q}^i(x) \leq \alpha n.$$

$$(A2) \quad \bar{d}_{G,Q}^i(x) \leq \alpha n.$$

$$(A3) \quad \text{There exists } z \in V(G) \text{ such that } |N_i^*(x, z)| + |N_i^*(z, x)| \leq \alpha n.$$

(Intuitively, the neighbourhood in  $Q_i$  of an *i-light* vertex is ‘atypical’, and this is unlikely to happen.)

Given  $\psi > 0$  and an index  $i \in \{0, 1, \dots, k - 2\}$ , we call  $\{x, y_1, y_2, y_3\} \subseteq V(G)$  a  $(k, x, i, \psi)$ -*configuration* if it satisfies the following.

$$(C1) \quad G[\{x, y_1, y_2, y_3\}] \text{ is a linear forest.}$$

$$(C2) \quad \bar{d}_{G,Q}^j(x) \geq 13 \cdot 6^k \psi n \text{ for all } j \in \{0, 1, \dots, k - 2\} \setminus \{i\}.$$



(C3) There exists  $i' \neq i$  such that  $d_{G,Q}^j(x) \geq 13 \cdot 6^k \psi n$  for all  $j \in \{0, 1, \dots, k-2\} \setminus \{i, i'\}$ .

(C4)  $\min\{d_{G,Q}^i(y_j), \bar{d}_{G,Q}^i(y_j)\} \leq \psi^2 n$  for all  $j \in [3]$ .

(Intuitively, (C1)–(C3) of the definition of  $(k, x, i, \psi)$ -configurations are useful for ‘building’ induced copies of  $C_{2k}$ , so the existence of a  $(k, x, i, \psi)$ -configuration in an induced- $C_{2k}$ -free graph  $G$  severely constrains the choices for the remaining edge set of  $G$ . The bounds arising from this are still not sufficiently strong though; we also need (C4), which gives further constraints on the choices for the remaining edge set of  $G$ .)

We partition  $F_Q(n, k, \eta, \mu)$  into the sets  $T_Q, F_Q^1, F_Q^2, F_Q^3$  defined as follows.

(F0)  $T_Q := T_Q(n, k) \cap F_Q(n, k, \eta, \mu)$ .

(F1)  $F_Q^1 \subseteq F_Q(n, k, \eta, \mu) \setminus T_Q$  is the set of all remaining graphs  $G$  which satisfy one of the following.

(i)  $G$  contains a  $(k, x, i, \psi)$ -configuration for some  $i \in \{0, 1, \dots, k-2\}$ , some  $x \in V(G)$  and some  $\psi \in \{\beta^{1/2}, \beta^2\}$ .

(ii)  $G$  contains a vertex  $x$  which is both  $i$ -light and  $j$ -light for some distinct indices  $i, j \in \{0, 1, \dots, k-2\}$ .

(F2)  $F_Q^2 \subseteq F_Q(n, k, \eta, \mu) \setminus (T_Q \cup F_Q^1)$  is the set of all remaining graphs that for some  $i \in \{0, 1, \dots, k-2\}$  contain a vertex  $x \in Q_i$  that satisfies  $\bar{d}_{G,Q}^i(x), d_{G,Q}^i(x) \geq \beta n$ .

(F3)  $F_Q^3 := F_Q(n, k, \eta, \mu) \setminus (T_Q \cup F_Q^1 \cup F_Q^2)$  is the set of all remaining graphs.

Sections 5.7, 5.8 and 5.9 are devoted to proving upper bounds on  $|F_Q^1|, |F_Q^2|$  and  $|F_Q^3|$  respectively. As mentioned earlier, it turns out that  $F_Q^3$  is the class of induced- $C_{2k}$ -

free graphs which are ‘extremely close’ to being  $k$ -templates (see Proposition 5.9.1). In Section 5.10 we will use these bounds to complete the proof of Lemma 5.6.1.

## 5.7 Estimation of $|F_Q^1|$

To estimate  $|F_Q^1|$  we will bound the number of graphs satisfying (F1)(i) and (F1)(ii) separately. The main difficulty is in estimating those satisfying (F1)(i), i.e. the ones containing a  $(k, x, i, \psi)$ -configuration. The idea here is that a  $(k, x, i, \psi)$ -configuration has many potential extensions into an induced copy of  $C_{2k}$ . More precisely, given a  $(k, x, i, \psi)$ -configuration  $H$  we can find many disjoint ‘skeleton’ graphs  $L$  with the same number of components as  $H$  such that  $H \cup L$  is a linear forest on  $2k$  vertices (i.e.  $H \cup L$  has a potential extension into an induced  $C_{2k}$ ). Thus each skeleton induces a restriction on further edges that can be added. Since the skeletons are disjoint we obtain many edge restrictions in total, and thus a good bound on the number of graphs containing a  $(k, x, i, \psi)$ -configuration. The next two propositions are used to formalise the notion of extendibility into an induced  $C_{2k}$ . (Roughly, in these propositions one can consider  $L_1$  as a  $(k, x, i, \psi)$ -configuration and  $L_2$  as an associated skeleton.)

**Proposition 5.7.1.** *Let  $c \geq 1$  and let  $L_1, L_2$  be disjoint linear forests, each with exactly  $c$  components, such that  $|V(L_1)| + |V(L_2)| = 2k$ . Then there exists a set  $E'$  of edges between  $V(L_1)$  and  $V(L_2)$  such that the graph  $(V(L_1) \cup V(L_2), E' \cup E(L_1) \cup E(L_2))$  is isomorphic to  $C_{2k}$ .*

The proof of Proposition 5.7.1 is trivial, and is omitted. Proposition 5.7.2 follows from an easy application of Proposition 5.7.1, and we give only a brief sketch of the proof.

**Proposition 5.7.2.** *Let  $c \geq 1$  and let  $L_1, L_2$  be linear forests that satisfy the following.*

- $V(L_1) \cap V(L_2) = \{x\}$ .
- $|V(L_1)|, |V(L_2)| > 1$ .
- $d_{L_1}(x) + d_{L_2}(x) = 2$ .
- $L_1$  and  $L_2 - \{x\}$  both have exactly  $c$  components.
- $|V(L_1) \cup V(L_2)| = 2k$ .

Then there exists a set  $E'$  of edges between  $V(L_1) \setminus \{x\}$  and  $V(L_2) \setminus \{x\}$  such that the graph  $(V(L_1) \cup V(L_2), E' \cup E(L_1) \cup E(L_2))$  is isomorphic to  $C_{2k}$ .

*Proof.* If  $d_{L_1}(x) = 0$  we apply Proposition 5.7.1 to  $L_1 - x, L_2$ ; if  $d_{L_2}(x) = 0$  we apply Proposition 5.7.1 to  $L_1, L_2 - x$ . If  $d_{L_1}(x) = d_{L_2}(x) = 1$  one can easily find  $E'$  directly. □

**Lemma 5.7.3.**  $|F_Q^1| \leq C 2^{-\frac{\beta^2 n}{14k}} f_k(n_k) 2^{t_{k-1}(n)}$ .

*Proof.* Let  $F_{Q,(i)}^1$  denote the set of all graphs in  $F_Q^1$  that satisfy (F1)(i). Similarly let  $F_{Q,(ii)}^1$  denote the set of all graphs in  $F_Q^1$  that satisfy (F1)(ii). Clearly,

$$|F_Q^1| \leq |F_{Q,(i)}^1| + |F_{Q,(ii)}^1|. \quad (5.7.4)$$

We will first estimate the number of graphs in  $F_{Q,(i)}^1$ . Any graph  $G \in F_{Q,(i)}^1$  can be constructed as follows. We first choose  $\psi \in \{\beta^2, \beta^{1/2}\}$ , and then perform the following steps.

- We choose an index  $i \in \{0, 1, \dots, k-2\}$ , a set of three (labelled) vertices  $Y = \{y_1, y_2, y_3\}$  in  $[n]$ , a vertex  $x \in [n] \setminus Y$ , and a set  $E$  of edges between these four vertices such that  $Y \cup \{x\}$  spans a linear forest. Let  $b_1$  denote the number of

such choices. The choices in the next steps will be made such that  $Y \cup \{x\}$  is a  $(k, x, i, \psi)$ -configuration in  $G$ .

- Next we choose the graph  $G'$  on vertex set  $[n] \setminus Y$  such that  $G[[n] \setminus Y] = G'$ . Let  $b_2$  denote the number of possibilities for  $G'$ .
- Next we choose the set  $E'$  of edges in  $G$  between  $Y$  and  $Q_i \setminus (Y \cup \{x\})$  such that  $E'$  is compatible with our previous choices. Let  $b_3$  denote the number of possibilities for  $E'$ .
- Finally we choose the set  $E''$  of edges in  $G$  between  $Y$  and  $[n] \setminus (Q_i \cup Y \cup \{x\})$  such that  $E''$  is compatible with our previous choices. Let  $b_4$  denote the number of possibilities for  $E''$ .

Hence,

$$|F_{Q,(i)}^1| \leq 2 \max_{\psi \in \{\beta^2, \beta^{1/2}\}} \{b_1 \cdot b_2 \cdot b_3 \cdot b_4\}. \quad (5.7.5)$$

We then estimate the number of graphs in  $F_{Q,(ii)}^1$ . Any graph  $G \in F_{Q,(ii)}^1$  can be constructed as follows.

- We first choose a single vertex  $x$  from  $[n]$  and distinct indices  $i, j \in \{0, 1, \dots, k-2\}$ . Let  $c_1$  denote the number of such choices. The choices in the next steps will be made such that  $x$  is both  $i$ -light and  $j$ -light in  $G$ .
- Next we choose the graph  $G'$  on vertex set  $[n] \setminus \{x\}$  such that  $G[[n] \setminus \{x\}] = G'$ . Let  $c_2$  denote the number of possibilities for  $G'$ .
- Next we choose the set  $E$  of edges in  $G$  between  $\{x\}$  and  $(Q_i \cup Q_j) \setminus \{x\}$  such that  $E$  is compatible with our previous choices. Let  $c_3$  denote the number of possibilities for  $E$ .

- Finally we choose the set  $E'$  of edges in  $G$  between  $\{x\}$  and  $[n] \setminus (Q_i \cup Q_j \cup \{x\})$ .

Let  $c_4$  denote the number of possibilities for  $E'$ .

Hence,

$$|F_{Q,(ii)}^1| \leq c_1 \cdot c_2 \cdot c_3 \cdot c_4. \quad (5.7.6)$$

The following series of claims will give upper bounds for the quantities  $b_1, \dots, b_4, c_1, \dots, c_4$ . Claims 1 and 5 are trivial, while the proof of Claim 6 is almost identical to that of Claim 2; we give proofs of Claims 2,3,4,7 and 8.

**Claim 1:**  $b_1 \leq 2^6 k n^4$ .

**Claim 2:**  $b_2 \leq C 2^{\mu^{1/2}n} f_k(n_k) 2^{t_{k-1}(n-3)}$ .

Indeed, note that for every graph  $\tilde{G} \in F_{Q,(i)}^1$ , Corollary 5.5.9 together with (5.5.7) implies that every optimal ordered  $(k-1)$ -partition of  $\tilde{G}[[n] \setminus Y]$  is contained in some set  $\mathcal{P}$  of size at most  $2^{\mu n}$ . Since  $G[[n] \setminus Y]$  is clearly induced- $C_{2k}$ -free, this together with (5.6.3) implies that

$$\begin{aligned} b_2 &\leq \sum_{Q' \in \mathcal{P}} |F_{Q'}(n-3, k)| \leq 6C 2^{\mu n} 2^{6(\log n)^2} f_k(\lceil (n-3)/(k-1) \rceil) 2^{t_{k-1}(n-3)} \\ &\leq C 2^{\mu^{1/2}n} f_k(n_k) 2^{t_{k-1}(n-3)}, \end{aligned}$$

as required.

**Claim 3:**  $b_3 \leq 2^{4\psi^{3/2}n}$ .

Indeed, for every graph  $\tilde{G} \in F_{Q,(i)}^1$  for which  $\{x, y_1, y_2, y_3\}$  is a  $(k, x, i, \psi)$ -configuration we

have that  $\min\{d_{\tilde{G},Q}^i(y_j), \bar{d}_{\tilde{G},Q}^i(y_j)\} \leq \psi^2 n$  for all  $j \in [3]$ . So  $b_3 \leq \prod_{j=1}^3 h(j)$  where  $h(j)$  denotes the number of possibilities for a set of edges between  $\{y_j\}$  and  $Q_i \setminus (Y \cup \{x\})$  such that either  $d_{G,Q}^i(y_j) \leq \psi^2 n$  or  $\bar{d}_{G,Q}^i(y_j) \leq \psi^2 n$ . Note that by (5.2.2),  $h(j) \leq 2 \binom{n}{\leq \psi^2 n} \leq 2^{\xi(\psi^2)n+1}$ . Hence,

$$b_3 \leq \prod_{j=1}^3 h(j) \leq (2^{\xi(\psi^2)n+1})^3 \stackrel{(5.2.3)}{\leq} 2^{4\psi^{3/2}n},$$

as required.

**Claim 4:**  $b_4 \leq 2^{3(k-2)n/(k-1)} 2^{\mu^{1/2}n} 2^{-\psi n/11^k}$ .

Indeed, first define  $L$  to be the graph on vertex set  $Y \cup \{x\}$  that satisfies  $E(L) = E$ . We say an induced subgraph  $H$  of  $G' - x$  is an *L-compatible skeleton* if it satisfies the following.

- $|V(H)| = 2k - 4$ .
- $G'[V(H) \cup \{x\}]$  is a linear forest.
- In  $G'$ ,  $x$  has  $2 - d_L(x)$  neighbours in  $V(H)$ .
- $L$  and  $H$  have the same number of components.

Given an  $L$ -compatible skeleton  $H$ , note that Proposition 5.7.2, applied with  $L, G'[V(H) \cup \{x\}]$  playing the roles of  $L_1, L_2$  respectively, implies that there exists a set  $E_{L,H}$  of possible edges between  $Y$  and  $V(H)$  such that  $(Y \cup \{x\} \cup V(H), E \cup E(H) \cup E_{L,H})$  is isomorphic to  $C_{2k}$ .

We will show that there exist a large number of disjoint  $L$ -compatible skeletons in  $G' - x$ . Since there is a limited number of ways to choose edges between  $Y$  and each of these

$L$ -compatible skeletons so as not to create an induced copy of  $C_{2k}$ , this will imply the claim.

For every index  $j \neq i$ , let  $N_j^1(x), N_j^2(x) \subseteq N_{Q_j}(x)$  be disjoint with  $|N_j^1(x)|, |N_j^2(x)| \geq \lfloor \frac{1}{2}|N_{Q_j}(x)| \rfloor$ . Similarly, let  $\overline{N}_j^1(x), \overline{N}_j^2(x) \subseteq \overline{N}_{Q_j}(x)$  be disjoint with  $|\overline{N}_j^1(x)|, |\overline{N}_j^2(x)| \geq \lfloor \frac{1}{2}|\overline{N}_{Q_j}(x)| \rfloor$ .

Note that we may assume that there exists an index  $i' \in \{0, 1, \dots, k-2\} \setminus \{i\}$  such that in  $G'$ ,  $|\overline{N}_{Q_j}(x)| \geq 12 \cdot 6^k \psi n$  for all  $j \in \{0, 1, \dots, k-2\} \setminus \{i\}$  and  $|N_{Q_j}(x)| \geq 12 \cdot 6^k \psi n$  for all  $j \in \{0, 1, \dots, k-2\} \setminus \{i, i'\}$ , since otherwise  $\{x, y_1, y_2, y_3\}$  cannot be a  $(k, x, i, \psi)$ -configuration. Define  $\ell_1, \dots, \ell_{k-2}$  such that  $\{\ell_1, \dots, \ell_{k-2}\} = \{0, 1, \dots, k-2\} \setminus \{i\}$  and  $\ell_{k-2} = i'$ . Thus the following hold.

- (a)  $|N_{\ell_j}^1(x)|, |N_{\ell_j}^2(x)|, |\overline{N}_{\ell_j}^1(x)|, |\overline{N}_{\ell_j}^2(x)| \geq 6 \cdot 6^k \psi n$  for all  $j \in \{1, \dots, k-3\}$ .
- (b)  $|\overline{N}_{\ell_{k-2}}^1(x)|, |\overline{N}_{\ell_{k-2}}^2(x)| \geq 6 \cdot 6^k \psi n$ .

We now show that  $G' - x$  contains at least  $5 \cdot 6^k \psi n$  disjoint  $L$ -compatible skeletons. Define  $t$  to be the number of components of  $L$ , and define  $s := d_L(x)$ . Then  $1 \leq t \leq 4$  and  $0 \leq s \leq 2$ . Note that  $t + s \geq 2$ , since a 4-vertex linear forest with one component contains no isolated vertices. We consider two cases. In each case we will describe the length and type of  $t$  path components,  $P^1, \dots, P^t$ , each with an even number of vertices. Proposition 5.5.6 (applied repeatedly) together with (a),(b) will then imply that  $G' - x$  contains at least  $5 \cdot 6^k \psi n$  disjoint  $L$ -compatible skeletons, each consisting exactly of  $t$  components isomorphic to  $P^1, \dots, P^t$ . (We can apply Proposition 5.5.6 here since in each case  $P^1 \cup \dots \cup P^t$  will contain a perfect matching.)

**Case 1:**  $s = 2$ .

- For  $1 \leq r \leq t-1$ ,  $P^r$  is a  $K_2$  of type  $\overline{N}_{\ell_r}^1(x), \overline{N}_{\ell_r}^2(x)$ .
- $P^t$  is a  $P_{2k-2t-2}$  of type

$$\overline{N}_{\ell_t}^1(x), \overline{N}_{\ell_t}^2(x), \overline{N}_{\ell_{t+1}}^1(x), \overline{N}_{\ell_{t+1}}^2(x), \dots, \overline{N}_{\ell_{k-2}}^1(x), \overline{N}_{\ell_{k-2}}^2(x).$$

**Case 2:** *Either  $s = 1$  or  $s = 0, t > 1$ .*

- For  $1 \leq r \leq 1-s$ ,  $P^r$  is a  $K_2$  of type  $N_{\ell_r}^1(x), \overline{N}_{\ell_r}^1(x)$ .
- $P^{2-s}$  is a  $P_{2k-2t-2}$  of type

$$N_{\ell_{2-s}}^1(x), \overline{N}_{\ell_{2-s}}^2(x), \overline{N}_{\ell_{3-s}}^1(x), \overline{N}_{\ell_{3-s}}^2(x), \dots, \overline{N}_{\ell_{k-t-s}}^1(x), \overline{N}_{\ell_{k-t-s}}^2(x).$$

- For  $k-t-s+1 \leq r \leq k-2$ ,  $P^r$  is a  $K_2$  of type  $\overline{N}_{\ell_r}^1(x), \overline{N}_{\ell_r}^2(x)$ .

Since  $t+s \geq 2$ , this covers all cases. Now fix a set  $SK$  of  $5 \cdot 6^k \psi n$  disjoint  $L$ -compatible skeletons in  $G' - x$ , and let  $H \in SK$ . Let  $h_H$  denote the number of possibilities for a set  $E^*$  of edges between  $Y$  and  $V(H)$ . Note that such a set  $E^*$  cannot equal  $E_{L,H}$ , since  $G$  needs to be induced- $C_{2k}$ -free. Thus  $h_H \leq 2^{|Y||V(H)|} - 1 = 2^{6(k-2)} - 1$ . Note that by (F2) $_{\mu}$  the number of vertices outside  $Q_i$  that are not contained in some graph  $H \in SK$  is at most  $(k-2)n/(k-1) + \mu n - 10(k-2)6^k \psi n$ . Hence,

$$\begin{aligned} b_4 &\leq 2^{3(k-2)n/(k-1)-30(k-2)6^k \psi n+3\mu n} \prod_{H \in SK} h_H \\ &\leq 2^{3(k-2)n/(k-1)-30(k-2)6^k \psi n+3\mu n} \left(2^{6(k-2)} (1 - 2^{-6(k-2)})\right)^{5 \cdot 6^k \psi n} \\ &\leq 2^{3(k-2)n/(k-1)} 2^{3\mu n} e^{-5 \cdot 6^k \psi n/(2^{6(k-2)})} \leq 2^{3(k-2)n/(k-1)} 2^{\mu^{1/2} n} 2^{-\psi n/11^k}, \end{aligned}$$

as required.



**Claim 5:**  $c_1 \leq k^2 n$ .

**Claim 6:**  $c_2 \leq C 2^{\mu^{1/2} n} f(n_k) 2^{t_{k-1}(n-1)}$ .

**Claim 7:**  $c_3 \leq 2^{7\xi(\alpha)n}$ .

Indeed, for every graph  $\tilde{G} \in F_{Q, (ii)}^1$  for which  $x$  is both  $i$ -light and  $j$ -light, we have that, for every  $\ell \in \{i, j\}$ , either  $\min\{|N_{Q_\ell}(x)|, |\overline{N}_{Q_\ell}(x)|\} \leq \alpha n$  or else there exists a vertex  $z \neq x$  such that  $|N_\ell^*(x, z)| + |N_\ell^*(z, x)| \leq \alpha n$ .

For  $\ell \in \{i, j\}$ , let  $h(\ell, 1)$  denote the number possibilities for a set of edges in  $G$  between  $\{x\}$  and  $Q_\ell \setminus \{x\}$  such that  $\min\{|N_{Q_\ell}(x)|, |\overline{N}_{Q_\ell}(x)|\} \leq \alpha n$ . Then  $h(\ell, 1) \leq 2^{\binom{n}{\leq \alpha n}} \leq 2^{\xi(\alpha)n+1}$ . For  $\ell \in \{i, j\}$ , let  $h(\ell, 2)$  denote the number possibilities for a set of edges between  $\{x\}$  and  $Q_\ell \setminus \{x\}$  such that there exists a vertex  $z \neq x$  such that  $|N_\ell^*(x, z)| + |N_\ell^*(z, x)| \leq \alpha n$ . Then  $h(\ell, 2) \leq n^{\binom{|N_{Q_\ell}(z)|}{\leq \alpha n}} \binom{|\overline{N}_{Q_\ell}(z)|}{\leq \alpha n} \leq 2^{3\xi(\alpha)n}$ .

Hence

$$c_3 \leq (h(i, 1) + h(i, 2))(h(j, 1) + h(j, 2)) \leq (2^{\xi(\alpha)n+1} + 2^{3\xi(\alpha)n})^2 \leq 2^{7\xi(\alpha)n},$$

as required.

**Claim 8:**  $c_4 \leq 2^{(k-3)n/(k-1)} 2^{2\mu n}$ .

Indeed, since the number of possible edges between  $\{x\}$  and  $[n] \setminus (Q_i \cup Q_j \cup \{x\})$  is at most  $(k-3)n/(k-1) + 2\mu n$ , we have that  $c_4 \leq 2^{(k-3)n/(k-1)+2\mu n}$ , as required.

Now (5.7.5) together with Claims 1–4 and Proposition 5.4.3(ii) implies that

$$\begin{aligned}
& |F_{Q,(i)}^1| \\
& \leq 2 \max_{\psi \in \{\beta^2, \beta^{1/2}\}} \left\{ 2^6 k n^4 \cdot C 2^{\mu^{1/2}n} f_k(n_k) 2^{t_{k-1}(n-3)} \cdot 2^{4\psi^{3/2}n} \cdot 2^{\frac{3(k-2)n}{k-1}} 2^{\mu^{1/2}n} 2^{-\frac{\psi n}{11k}} \right\} \\
& \leq \max_{\psi \in \{\beta^2, \beta^{1/2}\}} \left\{ C f_k(n_k) 2^{t_{k-1}(n-3) + \frac{3(k-2)n}{k-1}} 2^{-\frac{\psi n}{12k}} \right\} \leq C f_k(n_k) 2^{t_{k-1}(n)} 2^{-\frac{\beta^2 n}{13k}}.
\end{aligned} \tag{5.7.7}$$

Similarly, (5.7.6) together with Claims 5–8 and Proposition 5.4.3(ii) implies that

$$\begin{aligned}
|F_{Q,(ii)}^1| & \leq k^2 n \cdot C 2^{\mu^{1/2}n} f_k(n_k) 2^{t_{k-1}(n-1)} \cdot 2^{7\xi(\alpha)n} \cdot 2^{\frac{(k-3)n}{k-1}} 2^{2\mu n} \\
& \leq C 2^{\mu^{1/3}n} f_k(n_k) 2^{t_{k-1}(n-1) + \frac{(k-2)n}{k-1}} 2^{-\frac{n}{k-1}} 2^{7\xi(\alpha)n} \leq C f_k(n_k) 2^{t_{k-1}(n)} 2^{-\frac{n}{k}}.
\end{aligned} \tag{5.7.8}$$

Now (5.7.4) together with (5.7.7) and (5.7.8) implies that

$$|F_Q^1| \leq C f_k(n_k) 2^{t_{k-1}(n)} \left( 2^{-\frac{\beta^2 n}{13k}} + 2^{-\frac{n}{k}} \right) \leq C f_k(n_k) 2^{t_{k-1}(n)} 2^{-\frac{\beta^2 n}{14k}},$$

as required.  $\square$

## 5.8 Estimation of $|F_Q^2|$

Given  $G \in F_Q^2 \cup F_Q^3$  and  $i \in \{0, 1, \dots, k-2\}$ , let  $A_G^i := \{x \in Q_i : \bar{d}_{G,Q}^i(x), d_{G,Q}^i(x) \geq \beta n\}$ . The key result of this section (Lemma 5.8.6) states that  $A_G^i$  has bounded size. To prepare for this, we will classify the pairs of vertices in  $A_G^i$  according to their (non-)neighbourhood intersection pattern. The fact that  $G \notin F_Q^1$  allows us to observe some restrictions on these patterns (see Propositions 5.8.4 and 5.8.5). In the proof of Lemma 5.8.6 we use a Ramsey argument to restrict our view to one abundant type of pattern. This quickly leads to a contradiction if  $|A_G^i|$  is large. Using the fact that  $G \notin F_Q^1$  we show that the remainder

of each class (i.e.  $G[Q_i \setminus A_G^i]$ ) induces a very simple structure (Proposition 5.8.2). We translate this structural information into a sufficiently strong bound on the number of graphs in  $F_G^2$ , in Lemma 5.8.9.

Let  $\mathcal{L}$  denote the collection of all 4-vertex linear forests. The following proposition is an analogue of Proposition 5.3.8(i) that can be applied to graphs rather than 2-coloured multigraphs. It follows immediately from Proposition 5.3.8(i).

**Proposition 5.8.1.** *Let  $G$  be a graph such that for every  $H \in \mathcal{L}$ ,  $G$  is induced  $H$ -free. Then  $\overline{G}$  is a disjoint union of stars and triangles.*

**Proposition 5.8.2.** *Let  $G \in F_Q^2 \cup F_Q^3$  and  $i \in \{0, 1, \dots, k-2\}$ . Then  $\overline{G}[Q_i \setminus A_G^i]$  is a disjoint union of stars and triangles.*

*Proof.* Suppose for a contradiction that  $\overline{G}[Q_i \setminus A_G^i]$  is not a disjoint union of stars and triangles. Then Proposition 5.8.1 implies that  $\overline{G}[Q_i \setminus A_G^i]$  contains an induced copy of a graph in  $\mathcal{L}$ , with vertex set  $\{x, y_1, y_2, y_3\}$  say. We will show that  $\{x, y_1, y_2, y_3\}$  is a  $(k, x, i, \beta^{1/2})$ -configuration, which contradicts the fact that  $G \notin F_Q^1$ . Note that  $G[\{x, y_1, y_2, y_3\}]$  is a linear forest, and so  $\{x, y_1, y_2, y_3\}$  satisfies (C1). By the definition of  $A_G^i$  we have that  $\min\{d_{G,Q}^i(y_j), \bar{d}_{G,Q}^i(y_j)\} \leq \beta n$  for all  $j \in [3]$ , and so  $\{x, y_1, y_2, y_3\}$  satisfies (C4). Since  $G \notin F_Q^1$ ,  $x$  is  $j$ -light for at most one index  $j \in \{0, 1, \dots, k-2\}$ . Since  $x \in Q_i \setminus A_G^i$ ,  $x$  is  $i$ -light. Thus for every  $j \in \{0, 1, \dots, k-2\}$  with  $i \neq j$  we have that  $x$  is not  $j$ -light, and hence  $d_{G,Q}^j(x), \bar{d}_{G,Q}^j(x) > \alpha n > 13 \cdot 6^k \cdot \beta^{1/2} n$ , and so  $\{x, y_1, y_2, y_3\}$  satisfies (C2) and (C3). Therefore  $\{x, y_1, y_2, y_3\}$  is a  $(k, x, i, \beta^{1/2})$ -configuration, as required.  $\square$

The following definitions will be useful in order to show that  $|A_G^i|$  is small. Suppose  $S$  is a star or triangle. If  $S$  is a star on at least three vertices, we call the unique vertex in  $S$  of degree greater than one the *centre* of  $S$ . Otherwise we call the vertex of  $S$  with the smallest label the centre of  $S$ .

Let  $G \in F_Q^2 \cup F_Q^3$  and  $i, j \in \{0, 1, \dots, k-2\}$  and let  $x, y \in A_G^i$ .

- We say  $x, y$  are *j-irregular* if  $|\overline{N}_j(\{x, y\})| \leq \gamma n$ .
- We say  $x, y$  are *j-asymmetric* if  $|N_j^*(x, y)| + |N_j^*(y, x)| > 3\gamma n$  and either  $|N_j^*(x, y)| \leq \gamma n$  or  $|N_j^*(y, x)| \leq \gamma n$ .
- We say  $x, y$  are *j-identical* if  $|N_j^*(x, y)| + |N_j^*(y, x)| \leq 3\gamma n$ .

Roughly speaking, if one of the above holds then the neighbourhoods of  $x, y$  do not behave in a ‘random’ like way (thus constraining the number of possibilities for choosing the neighbourhoods). The following statement follows immediately from the above definitions and the fact that  $\gamma \ll \alpha$ .

$$\text{If } x, y \text{ are } j\text{-identical then } x, y \text{ are both } j\text{-light.} \quad (5.8.3)$$

**Proposition 5.8.4.** *Let  $G \in F_Q^2 \cup F_Q^3$  and  $i \in \{0, 1, \dots, k-2\}$  and let  $x, y \in A_G^i$ . Then  $x, y$  are *j-identical* for at most one index  $j \in \{0, 1, \dots, k-2\}$ .*

*Proof.* Suppose  $x, y$  are *j-identical* for some  $j \in \{0, 1, \dots, k-2\}$  and suppose  $j' \in \{0, 1, \dots, k-2\}$  with  $j' \neq j$ . It suffices to show that  $x, y$  are not *j'-identical*. Note that  $x$  is *j-light* by (5.8.3). Since  $G \notin F_Q^1$ ,  $x$  is *j''-light* for at most one index  $j'' \in \{0, 1, \dots, k-2\}$ . Thus  $x$  is not *j'-light*, and hence by (5.8.3)  $x, y$  are not *j'-identical*, as required.  $\square$

**Proposition 5.8.5.** *Let  $G \in F_Q^2 \cup F_Q^3$  and  $i \in \{0, 1, \dots, k-2\}$  and let  $x, y \in A_G^i$ . Then there exists an index  $j \in \{0, 1, \dots, k-2\}$  such that  $x, y$  are *j-irregular* or *j-asymmetric* (or both).*

*Proof.* Suppose for a contradiction that for every index  $\ell \in \{0, 1, \dots, k-2\}$ ,  $x, y$  are neither *ℓ-irregular* nor *ℓ-asymmetric*. Since, by Proposition 5.8.4,  $x, y$  are *j-identical* for

at most one index  $j$ , and  $k \geq 6$ , we may assume without loss of generality that  $x, y$  are not  $\ell$ -identical for  $\ell \in \{1, 2, 3\}$ . We consider the following two cases.

**Case 1:**  $x, y$  are adjacent.

In this case we define sets  $A_\ell, B_\ell$  for  $\ell \in \{0, 1, \dots, k-2\}$  as follows. We will use these sets to extend  $x, y$  into an induced copy of  $C_{2k}$ .

- Let  $A_1 := N_1^*(x, y)$  and  $B_1 := \overline{N}_1(\{x, y\})$ .
- Let  $A_2 := N_2^*(y, x)$  and  $B_2 := \overline{N}_2(\{x, y\})$ .
- For every  $\ell \in \{0, 1, \dots, k-2\} \setminus \{1, 2\}$ , let  $A_\ell, B_\ell \subseteq \overline{N}_\ell(\{x, y\})$  be disjoint and satisfy  $|A_\ell|, |B_\ell| \geq \lfloor |\overline{N}_\ell(\{x, y\})|/2 \rfloor$ .

Since for every  $\ell \in \{0, 1, \dots, k-2\}$   $x, y$  are neither  $\ell$ -irregular nor  $\ell$ -asymmetric, and for every  $\ell \in \{1, 2\}$   $x, y$  are not  $\ell$ -identical, we have that  $|A_\ell|, |B_\ell| \geq \gamma n/3$  for every  $\ell \in \{0, 1, \dots, k-2\}$ . This together with Proposition 5.5.6 and the fact that  $\mu \ll \gamma$  implies that there exists in  $G$  an induced copy of  $P_{2k-2}$  of type

$$A_1, B_1, A_0, B_0, A_3, B_3, \dots, A_{k-2}, B_{k-2}, B_2, A_2.$$

By the definition of the sets  $A_\ell, B_\ell$ , the vertices of this  $P_{2k-2}$  together with  $x, y$  induce on  $G$  a copy of  $C_{2k}$ . This contradicts the fact that  $G \in F_Q(n, k)$ .

**Case 2:**  $x, y$  are not adjacent.

In this case we define sets  $A_\ell, B_\ell$  for  $\ell \in \{0, 1, \dots, k-2\}$  as follows. Similarly to the previous case, we will find an induced  $C_{2k}$  which contains  $x, y$  together with exactly one

vertex from each of these sets.

- Let  $A_1 := N_1^*(x, y)$  and  $B_1 := N_1^*(y, x)$ .
- Let  $A_2 := N_2^*(x, y)$  and  $B_2 := \overline{N}_2(\{x, y\})$ .
- Let  $A_3 := N_3^*(y, x)$  and  $B_3 := \overline{N}_3(\{x, y\})$ .
- For every  $\ell \in \{0, 1, \dots, k-2\} \setminus \{1, 2, 3\}$ , let  $A_\ell, B_\ell \subseteq \overline{N}_\ell(\{x, y\})$  be disjoint and satisfy  $|A_\ell|, |B_\ell| \geq \lfloor |\overline{N}_\ell(\{x, y\})|/2 \rfloor$ .

Since for every  $\ell \in \{0, 1, \dots, k-2\}$   $x, y$  are neither  $\ell$ -irregular nor  $\ell$ -asymmetric, and for every  $\ell \in \{1, 2, 3\}$   $x, y$  are not  $\ell$ -identical, we have that  $|A_\ell|, |B_\ell| \geq \gamma n/3$  for every  $\ell \in \{0, 1, \dots, k-2\}$ . As before, this together with Proposition 5.5.6 implies that there exists in  $G$  an induced copy of the graph  $H$  that consists of the following two components:

- One  $P_{2k-4}$  of type  $A_2, B_2, A_0, B_0, A_4, B_4, \dots, A_{k-2}, B_{k-2}, B_3, A_3$ .
- One  $K_2$  of type  $A_1, B_1$ .

By the definition of the sets  $A_\ell, B_\ell$ , the vertices of  $H$  together with  $x, y$  induce on  $G$  a copy of  $C_{2k}$ . This contradicts the fact that  $G \in F_Q(n, k)$ .

This covers all cases, and hence completes the proof. □

Recall from Section 5.6 that  $M := R_{2k-2}(\lceil \frac{1}{\gamma} \rceil) + 1$ .

**Lemma 5.8.6.** *Let  $G \in F_Q^2 \cup F_Q^3$  and  $i \in \{0, 1, \dots, k-2\}$ . Then  $|A_G^i| < M$ .*

*Proof.* Suppose for a contradiction that  $|A_G^i| \geq M$ . Consider an auxiliary complete graph  $H_i$  with  $V(H_i) = A_G^i$ . We define a  $(2k-2)$ -edge-colouring  $\mathcal{C}$  of  $H_i$  with colours  $\{a_0, b_0, a_1, b_1, \dots, a_{k-2}, b_{k-2}\}$  as follows.

- For every  $j \in \{0, 1, \dots, k-2\}$ , an edge  $xy \in E(H)$  is coloured  $a_j$  if  $x, y$  are  $j$ -irregular and for every  $j' \in \{0, 1, \dots, k-2\}$  with  $j' < j$ ,  $x, y$  are not  $j'$ -irregular.
- An edge  $xy \in E(H)$  that was not coloured in the previous step is coloured  $b_j$  if  $x, y$  are  $j$ -asymmetric, and for every  $j' \in \{0, 1, \dots, k-2\}$  with  $j' < j$ ,  $x, y$  are not  $j'$ -asymmetric.

Note that by Proposition 5.8.5, every edge is coloured by a unique colour in  $\mathcal{C}$ .

Now since  $M > R_{2k-2}(\lceil 1/\gamma \rceil)$ ,  $H_i$  contains a monochromatic clique of size at least  $1/\gamma$ . Let  $X = \{x_1, x_2, \dots, x_{\lceil 1/\gamma \rceil}\}$  be the vertex set of such a monochromatic clique. We consider the following two cases.

**Case 1:**  $X$  has colour  $a_j$  for some  $j \in \{0, 1, \dots, k-2\}$ .

In this case every pair of vertices in  $X$  is  $j$ -irregular, by definition of  $\mathcal{C}$ . Let  $X' := \{x_1, x_2, \dots, x_{\lceil \beta/2\gamma \rceil}\}$  and suppose  $z, z' \in X'$ . By the definition of  $j$ -irregularity,  $|\overline{N}_j(z) \cap \overline{N}_j(z')| \leq \gamma n$ . Note also that  $|\overline{N}_j(z)| \geq \beta n$  by Proposition 5.5.2 and the fact that  $z \in A_G^i$ . So by the inclusion-exclusion principle,

$$\begin{aligned} 2n/(k-1) &\geq n/(k-1) + \mu n \geq |Q_j| \geq \sum_{z \in X'} |\overline{N}_j(z)| - \sum_{\substack{z, z' \in X' \\ z \neq z'}} |\overline{N}_j(z) \cap \overline{N}_j(z')| \\ &\geq \beta \lceil \beta/2\gamma \rceil n - \lceil \beta^2/(4\gamma^2) \rceil \gamma n \geq \beta^2 n/5\gamma > 2n/(k-1), \end{aligned}$$

where the last inequality follows from the fact that  $\gamma \ll \beta$ . This is a contradiction.

**Case 2:**  $X$  has colour  $b_j$  for some  $j \in \{0, 1, \dots, k-2\}$ .

In this case every pair of vertices in  $X$  is  $j$ -asymmetric, by definition of  $\mathcal{C}$ . Suppose  $\ell, \ell' \in [\lceil 1/\gamma \rceil]$  are distinct. By the definition of  $j$ -asymmetry, exactly one of the following holds.

$$(a) \quad |N_j^*(x_\ell, x_{\ell'})| \leq \gamma n \text{ and } |N_j^*(x_{\ell'}, x_\ell)| > 2\gamma n.$$

$$(b) \quad |N_j^*(x_{\ell'}, x_\ell)| \leq \gamma n \text{ and } |N_j^*(x_\ell, x_{\ell'})| > 2\gamma n.$$

Consider the auxiliary tournament  $T$  with  $V(T) = X$  and  $E(T) = \{\overrightarrow{x_\ell x_{\ell'}} : \ell, \ell' \text{ satisfy (a)}\}$ . By Redei's theorem every tournament contains a directed Hamilton path. So, by relabelling the indices if necessary, we may assume that  $\overrightarrow{x_\ell x_{\ell+1}} \in E(T)$  for every  $\ell \in [\lceil 1/\gamma \rceil - 1]$ . Thus for every  $\ell \in [\lceil 1/\gamma \rceil - 1]$ ,

$$\begin{aligned} |\overline{N}_j(x_{\ell+1})| &= |(\overline{N}_j(x_\ell) \setminus N_j^*(x_{\ell+1}, x_\ell)) \cup N_j^*(x_\ell, x_{\ell+1})| \leq |\overline{N}_j(x_\ell)| - 2\gamma n + \gamma n \\ &\leq |\overline{N}_j(x_\ell)| - \gamma n. \end{aligned}$$

Hence,

$$|\overline{N}_j(x_{\lceil 1/\gamma \rceil})| \leq |\overline{N}_j(x_1)| - \left(\frac{1}{\gamma} - 1\right) \cdot \gamma n \leq |Q_j| - (1 - \gamma)n < 0,$$

which is a contradiction.

This covers all cases, and hence completes the proof. □

Suppose  $G \in F_Q^2$  and  $i \in \{0, 1, \dots, k-2\}$ . By Proposition 5.8.2,  $\overline{G}[Q_i \setminus A_G^i]$  is a disjoint



union of stars and triangles. Let  $\mathcal{S}$  be the set of components of  $\overline{G}[Q_i \setminus A_G^i]$  with the largest number of vertices. Let  $S^\circ$  be the component in  $\mathcal{S}$  whose centre  $c$  has the smallest label. Define  $Y_i = Y_i(G, Q)$  to be the set of all isolated vertices in  $\overline{G}[Q_i \setminus A_G^i]$  together with all vertices in  $V(S^\circ) \setminus \{c\}$ .

**Lemma 5.8.7.** *Let  $G \in F_Q^2$  and  $i \in \{0, 1, \dots, k-2\}$ . Then  $|Y_i| \geq 10n/\log n$ .*

*Proof.* Define  $s := \lceil 10n/\log n \rceil$ . Suppose for a contradiction that  $|Y_i| < s$ . Since  $G \in F_Q^2$ , there exists an index  $i' \in \{0, 1, \dots, k-2\}$  such that  $|A_G^{i'}| > 0$ . Let  $x \in A_G^{i'}$ . The definition of  $A_G^{i'}$  together with Proposition 5.5.2 implies that  $|\overline{N}_{Q_j}(x)| \geq \beta n$  for every  $j \in \{0, 1, \dots, k-2\}$ . This together with Lemma 5.8.6 implies that  $|\overline{N}_{Q_i}(x) \setminus A_G^i| \geq \beta n - M > 2s$ . Also, since  $|Y_i| < s$ , at most  $s$  components in  $\overline{G}[Q_i \setminus A_G^i]$  are isolated vertices and every component in  $\overline{G}[Q_i \setminus A_G^i]$  has order at most  $s$ . Thus there are at least two non-trivial components  $S, S'$  of  $\overline{G}[Q_i \setminus A_G^i]$  that each contain a non-neighbour of  $x$ .

Since  $S$  is a non-trivial component of  $\overline{G}[Q_i \setminus A_G^i]$  there exist vertices  $y, y' \in S$  such that  $xy, yy' \notin E(G[Q_i])$ . Let  $y'' \in S'$  be such that  $xy'' \notin E(G[Q_i])$ . Since  $y''$  belongs to a different component of  $\overline{G}[Q_i \setminus A_G^i]$  to  $y$  and  $y'$ , it follows that  $yy'', y'y'' \in E(G[Q_i])$ . Thus,

$$E(G[\{x, y, y', y''\}]) \in \{\{yy'', y'y''\}, \{xy', yy'', y'y''\}\}. \quad (5.8.8)$$

**Claim:**  $\{x, y, y', y''\}$  is a  $(k, x, i, \beta^2)$ -configuration.

Indeed, by (5.8.8),  $G[\{x, y, y', y''\}]$  is a linear forest and so  $\{x, y, y', y''\}$  satisfies (C1). As observed above,  $\overline{d}_{G,Q}^j(x) \geq \beta n > 13 \cdot 6^k \beta^2 n$  for every  $j \in \{0, 1, \dots, k-2\}$ , and so  $\{x, y, y', y''\}$  satisfies (C2). Since  $G \notin F_Q^1$ , there do not exist distinct  $j, j' \in \{0, 1, \dots, k-2\}$  such that  $x$  is both  $j$ -light and  $j'$ -light. So there exists  $j \in \{0, 1, \dots, k-2\}$  such that for every  $j' \in \{0, 1, \dots, k-2\}$  with  $j' \neq j$ ,  $d_{G,Q}^{j'}(x) > \alpha n > 13 \cdot 6^k \beta^2 n$ , and so  $\{x, y, y', y''\}$

satisfies (C3). Since  $S, S'$  each contain at most  $s$  vertices,  $y, y', y''$  each have at most  $s$  non-neighbours in  $G[Q_i \setminus A_G^i]$ . This together with Lemma 5.8.6 implies that  $y, y', y''$  each have at most  $s + M \leq \beta^4 n$  non-neighbours in  $G[Q_i]$ , and so  $\{x, y, y', y''\}$  satisfies (C4). Hence  $\{x, y, y', y''\}$  is a  $(k, x, i, \beta^2)$ -configuration, as required.

The above claim contradicts the fact that  $G \notin F_Q^1$ , and hence completes the proof.  $\square$

Lemma 5.8.7 guarantees a large set of vertices in each class  $Q_i$  (namely  $Y_i$ ) with an extremely restricted (non-)neighbourhood. This is the key idea in our estimation of  $|F_Q^2|$ .

**Lemma 5.8.9.**  $|F_Q^2| \leq C 2^{-n} f_k(n_k) 2^{t_{k-1}(n)}$ .

*Proof.* Define  $s := \lceil 10n / \log n \rceil$ . Since by Lemma 5.8.7  $|Y_i(G, Q)| \geq s$  for every graph  $G \in F_Q^2$ , any graph  $G \in F_Q^2$  can be constructed as follows.

- First we choose sets  $S_\ell \subseteq Q_\ell$  such that  $|S_\ell| = s$ , for every  $\ell \in \{0, 1, \dots, k-2\}$ . Let  $b_1$  denote the number of such choices.
- Next we choose the graph  $G'$  on  $[n] \setminus \bigcup_{\ell \in \{0, 1, \dots, k-2\}} S_\ell$  such that  $G'[[n] \setminus \bigcup_{\ell \in \{0, 1, \dots, k-2\}} S_\ell] = G'$ . Let  $b_2$  denote the number of possibilities for  $G'$ .
- Next we choose the set  $E'$  of internal edges of  $G$  that are incident to at least one vertex in  $\bigcup_{\ell \in \{0, 1, \dots, k-2\}} S_\ell$  in such a way that the resulting graph  $G$  will satisfy  $S_\ell \subseteq Y_\ell(G, Q)$  for every  $\ell \in \{0, 1, \dots, k-2\}$ . Let  $b_3$  denote the number of possibilities for  $E'$ .
- Finally we choose the set  $E''$  of crossing edges of  $G$  that are incident to at least one vertex in  $\bigcup_{\ell \in \{0, 1, \dots, k-2\}} S_\ell$ . Let  $b_4$  denote the number of possibilities for  $E''$ .

Hence,

$$|F_Q^2| \leq b_1 \cdot b_2 \cdot b_3 \cdot b_4. \quad (5.8.10)$$

The following series of claims will give upper bounds for the quantities  $b_1, \dots, b_4$ . The proof of Claim 2 is almost identical to that of Claim 2 in Lemma 5.7.3; we give proofs of the others.

**Claim 1:**  $b_1 \leq 2^n$ .

Indeed,

$$b_1 \leq \left( \binom{n}{\lceil \frac{10n}{\log n} \rceil} \right)^{k-1} \leq \left( \left( \frac{e \log n}{10} \right)^{\frac{10n}{\log n}} \right)^{k-1} \leq 2^n,$$

as required.

**Claim 2:**  $b_2 \leq C 2^{\mu^{1/2}n} f_k(\lceil n/(k-1) - s \rceil) 2^{t_{k-1}(n-(k-1)s)}$ .

**Claim 3:**  $b_3 \leq 2^n$ .

Indeed, for every graph  $G^* \in F_Q^2$  for which  $S_\ell \subseteq Y_\ell(G^*, Q)$  for every  $\ell \in \{0, 1, \dots, k-2\}$ , let  $G_{B,\ell}^* := \overline{G^*}[Q_\ell \setminus A_{G^*}^\ell]$ . Then each  $S_\ell$  consists of isolated vertices in  $G_{B,\ell}^*$  as well as non-centre vertices of a single component  $\tilde{C}$  of  $G_{B,\ell}^*$ . (Note that  $\tilde{C}$  is a star or triangle in  $G_{B,\ell}^*$ , with some centre  $u \in Q_\ell \setminus (A_{G^*}^\ell \cup S_\ell)$ .) By Lemma 5.8.6, we also have that  $|A_{G^*}^\ell| \leq M$ .

Hence  $b_3 \leq \prod_{\ell=0}^{k-2} \prod_{j=1}^3 h(\ell, j)$ , where the quantities  $h(\ell, j)$  are defined as follows. Let  $h(\ell, 1)$  denote the number of ways to choose a set  $\tilde{A}^\ell \subseteq Q_\ell \setminus S_\ell$  of size at most  $M$ . (In what follows  $\tilde{A}^\ell$  will play the role of  $A_G^i$ .) Then  $h(\ell, 1) \leq n^M$ . Given such a set  $\tilde{A}^\ell$ , let  $h(\ell, 2)$  denote the number of ways to choose  $\tilde{C}$ . Then  $h(\ell, 2) \leq n 2^{|S_\ell|} + n^3 \leq n 2^{s+1}$ . (Indeed, if  $\tilde{C}$  is a star we have at most  $n$  choices for the centre  $u$ , and for every vertex

$v \in S_\ell$  we can choose whether  $v$  is adjacent to  $u$  or not; if  $\tilde{C}$  is a triangle we have at most  $n^3$  choices for its vertices.) Given a set  $\tilde{A}^\ell$  as above, let  $h(\ell, 3)$  denote the number of possible sets of edges between  $S_\ell$  and  $\tilde{A}^\ell$ . Then  $h(\ell, 3) \leq 2^{|S_\ell||\tilde{A}^\ell|} \leq 2^{sM}$ . Hence

$$b_3 \leq \prod_{\ell=0}^{k-2} \prod_{j=1}^3 h(\ell, j) \leq (n^M \cdot n2^{s+1} \cdot 2^{sM})^{k-1} \leq 2^n,$$

as required.

**Claim 4:**  $b_4 \leq 2^{(k-2)sn - \binom{k-1}{2}s^2}$ .

Indeed, note that, for a fixed index  $\ell \in \{0, 1, \dots, k-2\}$ , the number  $h_\ell$  of possible crossing edges in  $G$  that are incident to a vertex in  $S_\ell$  is at most  $s(n - |Q_\ell|)$ . Also, the number of possible crossing edges in  $G$  that are incident to two vertices in  $\bigcup_{\ell \in \{0, 1, \dots, k-2\}} S_\ell$  is exactly  $\binom{k-1}{2}s^2$ . Hence,

$$b_4 \leq 2^{\sum_{\ell=0}^{k-2} h_\ell} 2^{-\binom{k-1}{2}s^2} \leq 2^{(k-2)sn - \binom{k-1}{2}s^2},$$

as required.

Note that  $t_{k-1}(s(k-1)) = \binom{k-1}{2}s^2$  and that by Lemma 5.4.2,  $f_k(n_k) \geq s^{s/2} f_k(\lceil n/(k-1) - s \rceil) \geq 2^{4n} f_k(\lceil n/(k-1) - s \rceil)$ . These observations together with (5.8.10), Claims 1–4 and Proposition 5.4.3(ii) imply that

$$\begin{aligned} |F_Q^2| &\leq 2^n \cdot C2^{\mu^{1/2}n} f_k(\lceil n/(k-1) - s \rceil) 2^{t_{k-1}(n-(k-1)s)} \cdot 2^n \cdot 2^{(k-2)sn - \binom{k-1}{2}s^2} \\ &\leq C2^{3n} 2^{-4n} f_k(n_k) 2^{t_{k-1}(n-(k-1)s) + (k-2)sn - s(k-1)(k-2) - t_{k-1}(s(k-1))} \\ &\leq C2^{-n} f_k(n_k) 2^{t_{k-1}(n)}, \end{aligned}$$

as required. □

## 5.9 Estimation of $|F_Q^3|$

The information we have gained so far allows us to easily deduce that every  $G \in F_Q^3$  is extremely close to being a  $k$ -template (see Proposition 5.9.1). One advantage of this is that it allows us to use more precise estimates when applying induction (see Corollary 5.9.3).

**Proposition 5.9.1.** *Let  $G \in F_Q^3$  and  $i \in \{0, 1, \dots, k-2\}$ . Then the following hold.*

- (i)  $\overline{G}[Q_i]$  is a disjoint union of stars and triangles.
- (ii)  $G$  contains at most  $n$  internal non-edges.
- (iii) Every vertex  $x \in Q_i$  satisfies  $\overline{d}_{G,Q}^i(x) < \beta n$ .

*Proof.* (i) Since  $G \notin F_Q^2$ , every vertex  $x \in Q_i$  satisfies  $\min\{d_{G,Q}^i(x), \overline{d}_{G,Q}^i(x)\} < \beta n$ .

Thus  $A_G^i = \emptyset$ , and so by Proposition 5.8.2,  $\overline{G}[Q_i]$  is a disjoint union of stars and triangles.

(ii) This follows immediately from (i).

(iii) Let  $x \in Q_i$ . Let us first show that  $d_{G,Q}^i(x) \geq \beta n$ . Suppose not. Then  $\overline{d}_{G,Q}^i(x) = |Q_i| - d_{G,Q}^i(x) - 1 > |Q_i| - \beta n - 1 \geq n/(k-1) - \mu n - \beta n - 1$ . Thus for every  $j \in \{0, 1, \dots, k-2\}$  with  $j \neq i$ , Proposition 5.5.2 implies that

$$\begin{aligned} d_{G,Q}^j(x) &= |Q_j| - \overline{d}_{G,Q}^j(x) \leq |Q_j| - \overline{d}_{G,Q}^i(x) \\ &< \left( \frac{n}{k-1} + \mu n \right) - \left( \frac{n}{k-1} - \mu n - \beta n - 1 \right) = \beta n + 2\mu n + 1 < \alpha n, \end{aligned}$$

where the last inequality follows from the fact that  $\mu, \beta \ll \alpha$ . Thus  $x$  is both  $i$ -light and  $j$ -light, which contradicts the fact that  $G \notin F_Q^1$ . Thus  $d_{G,Q}^i(x) \geq \beta n$ . This

together with the fact (observed in the proof of (i), above) that  $A_G^i = \emptyset$  implies that  $\vec{d}_{G,Q}^i(x) < \beta n$ , as required.

□

Recall the definition of property  $(F1)_\nu$  in Section 5.5. We define  $T_Q^*(n, k) \subseteq F_Q^3$  to be the set of all (labelled) graphs in  $F_Q^3$  that satisfy property  $(F1)_{(40n \log n)^{1/2}/n}$  with respect to  $Q$ . Proposition 5.9.1(ii) together with Lemma 5.5.3(i) applied with  $(40n \log n)^{1/2}/n, n$  playing the roles of  $\nu, m$  respectively implies that

$$|F_Q^3 \setminus T_Q^*(n, k)| \leq 2^{t_{k-1}(n) - n \log n/5}. \quad (5.9.2)$$

So (5.9.2) allows us to restrict our attention to the class  $T_Q^*(n, k)$ . In particular, this allows us to apply property  $(F1)_\nu$  to much smaller vertex sets than in the preceding sections. This in turn gives us a much better bound on the number of partitions that may arise after deleting a small number of vertices. More precisely, Lemma 5.5.8 applied with  $(40n \log n)^{1/2}/n, n$  playing the roles of  $\nu, m$  respectively implies the following result. Recall that  $\mathcal{P}(Q, s)$  was defined before (5.5.7).

**Corollary 5.9.3.** *Let  $S \subseteq [n]$  with  $|S| \leq n/k^2$ . Then for every  $G \in T_Q^*(n, k)$ , every optimal ordered  $(k-1)$ -partition of  $G - S$  is an element of  $\mathcal{P}(Q - S, 40k^4 \log n)$ .*

In order to estimate  $|T_Q^*(n, k)|$  (and thus  $|F_Q^3|$ ) we will further split  $T_Q^*(n, k)$  into four classes  $\mathcal{A}_1, \dots, \mathcal{A}_4$ . To define these classes we require some further notation. We say that  $G$  contains a  $(6, 3)$ -forest with respect to  $Q$  if there exist distinct indices  $i, j \in \{0, 1, \dots, k-2\}$  such that there exist six vertices in  $Q_i \cup Q_j$  that induce on  $G$  a linear forest with at most three components. A  $(6, 3)$ -forest has potential extensions into an

induced  $C_{2k}$ , so its existence in every  $G \in \mathcal{A}_3$  (see below) constrains the possible edge sets for  $G$  (and thus the number of choices for  $G$ ). To obtain a significant constraint on the possible edge sets however, we first need to exclude the situations that arise in the classes  $\mathcal{A}_1$  and  $\mathcal{A}_2$ , described below. These involve the structure of the stars of the complement graph inside the vertex classes, so to describe these classes of graphs recall that the centres of stars and triangles were defined before Proposition 5.8.4. Given a graph  $G \in F_Q^3$  and an index  $i \in \{0, 1, \dots, k-2\}$  we define the following sets.

- $C^i(G, Q)$  is the set of all centres of triangles and non-trivial stars in  $\overline{G}[Q_i]$ .
- $C_{high}^i(G, Q)$  is the set of all centres of stars in  $\overline{G}[Q_i]$  of order at least  $n^{1-\frac{1}{2k^2}}/200k^2$ .
- $B_{high}^i(G, Q)$  is the set of all vertices in  $Q_i$  which have a non-neighbour in  $C_{high}^i$ .
- $C_{low}^i(G, Q)$  is the set of all centres of triangles and non-trivial stars in  $\overline{G}[Q_i]$  of order less than  $n^{1-\frac{1}{2k^2}}/200k^2$ .
- $B_{low}^i(G, Q)$  is the set of all vertices in  $Q_i$  which have a non-neighbour in  $C_{low}^i$ .
- $C_0^i(G, Q)$  is the set of all isolated vertices in  $\overline{G}[Q_i]$ .

We may sometimes write  $C^i$  for  $C^i(G, Q)$  when the graph  $G$  and ordered  $(k-1)$ -partition  $Q$  we consider are clear from the context (and similarly for  $C_{high}^i, B_{high}^i, C_{low}^i, B_{low}^i, C_0^i$ ). Note that Proposition 5.9.1(i) implies that  $C_{high}^i, B_{high}^i, C_{low}^i, B_{low}^i, C_0^i$  form a partition of  $Q_i$ . Given a subset  $B \subseteq B_{low}^i$ , we denote by  $C(B)$  the set of all vertices in  $C_{low}^i$  that have a non-neighbour in  $B$ .

We partition  $T_Q^*(n, k)$  into the sets  $\mathcal{A}_1, \dots, \mathcal{A}_4$  defined as follows.

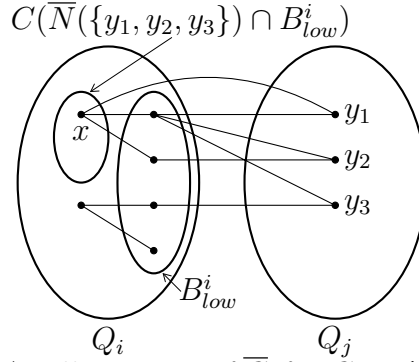


Figure 1: An illustration of  $\overline{G}$  for  $G \in \mathcal{A}_2$ . Note that (5.9.4) implies that at most one of  $xy_1, xy_2$  is an edge in  $\overline{G}$ .

- $\mathcal{A}_1$  is the set of all graphs  $G \in T_Q^*(n, k)$  for which there exist distinct indices  $i, j \in \{0, 1, \dots, k-2\}$  such that  $|B_{low}^i| \geq n/2k^2$  and there exist distinct vertices  $y_1, y_2, y_3 \in Q_j$  that satisfy  $|\overline{N}(\{y_1, y_2, y_3\}) \cap B_{low}^i| \leq n/200k^2$ .
- $\mathcal{A}_2$  is the set of all graphs  $G \in T_Q^*(n, k) \setminus \mathcal{A}_1$  for which there exist distinct indices  $i, j \in \{0, 1, \dots, k-2\}$  such that  $|B_{low}^i| \geq n/2k^2$  and there exist distinct vertices  $y_1, y_2, y_3 \in Q_j$  with  $y_1, y_2 \notin C^j(G, Q)$  that satisfy

$$C(\overline{N}(\{y_1, y_2, y_3\}) \cap B_{low}^i) \cap \overline{N}(\{y_1, y_2\}) = \emptyset. \quad (5.9.4)$$

(See Figure 1.)

- $\mathcal{A}_3$  is the set of all graphs  $G \in T_Q^*(n, k) \setminus (\mathcal{A}_1 \cup \mathcal{A}_2)$  such that  $G$  contains a  $(6, 3)$ -forest with respect to  $Q$ .
- $\mathcal{A}_4 := T_Q^*(n, k) \setminus (\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3)$  is the set of all remaining graphs.

We will estimate the sizes of  $\mathcal{A}_1, \dots, \mathcal{A}_4$  separately. Lemma 5.9.5 below gives a bound on  $|\mathcal{A}_1|$ . The idea of the proof of Lemma 5.9.5 is that in this case the neighbourhoods of  $y_1, y_2, y_3$  are ‘atypical’, and hence a Chernoff estimate (see Claim 4) shows that graphs in  $\mathcal{A}_1$  are rare.



**Lemma 5.9.5.**  $|\mathcal{A}_1| \leq C 2^{-n/150k^2} f_k(n_k) 2^{t_{k-1}(n)}.$

*Proof.* Any graph  $G \in \mathcal{A}_1$  can be constructed as follows.

- First we choose distinct indices  $i, j \in \{0, 1, \dots, k-2\}$ , distinct vertices  $y_1, y_2, y_3 \in Q_j$ , and a set  $E$  of edges between  $y_1, y_2, y_3$ . Let  $b_1$  denote the number of such choices. The choices in the next steps will be made such that  $G$  satisfies  $|B_{low}^i| \geq n/2k^2$  and  $|\overline{N}(\{y_1, y_2, y_3\}) \cap B_{low}^i| \leq n/200k^2$ .
- Next we choose the graph  $G'$  on vertex set  $[n] \setminus \{y_1, y_2, y_3\}$  such that  $G[[n] \setminus \{y_1, y_2, y_3\}] = G'$ . Let  $b_2$  denote the number of possibilities for  $G'$ .
- Next we choose the set  $E'$  of edges in  $G$  between  $\{y_1, y_2, y_3\}$  and  $Q_j \setminus \{y_1, y_2, y_3\}$ . Let  $b_3$  denote the number of possibilities for  $E'$ .
- Finally we choose the set  $E''$  of edges in  $G$  between  $\{y_1, y_2, y_3\}$  and  $[n] \setminus Q_j$  such that  $E''$  is compatible with our previous choices. Let  $b_4$  denote the number of possibilities for  $E''$ .

Hence,

$$|\mathcal{A}_1| \leq b_1 \cdot b_2 \cdot b_3 \cdot b_4. \quad (5.9.6)$$

The following series of claims will give upper bounds for the quantities  $b_1, \dots, b_4$ . Claim 1 is trivial; we give proofs of the others.

**Claim 1:**  $b_1 \leq 2^3 k^2 n^3$ .

**Claim 2:**  $b_2 \leq C 2^{2(\log n)^3} f_k(n_k) 2^{t_{k-1}(n-3)}.$

Indeed, note that for every graph  $\tilde{G} \in \mathcal{A}_1$ , Corollary 5.9.3 together with (5.5.7) implies

that every optimal ordered  $(k-1)$ -partition of  $\tilde{G}[[n] \setminus \{y_1, y_2, y_3\}]$  is contained in some set  $\mathcal{P}$  of size at most  $2^{(\log n)^3}$ . Since  $G[[n] \setminus \{y_1, y_2, y_3\}]$  is clearly induced- $C_{2k}$ -free, this together with (5.6.3) implies that

$$\begin{aligned} b_2 &\leq \sum_{Q' \in \mathcal{P}} |F_{Q'}(n-3, k)| \leq 6C 2^{(\log n)^3} 2^{6(\log n)^2} f_k(\lceil (n-3)/(k-1) \rceil) 2^{t_{k-1}(n-3)} \\ &\leq C 2^{2(\log n)^3} f_k(n_k) 2^{t_{k-1}(n-3)}, \end{aligned}$$

as required.

**Claim 3:**  $b_3 \leq 2^{3\xi(\beta)n}$ .

Indeed, for every graph  $\tilde{G} \in \mathcal{A}_1$  and every  $\ell \in [3]$ , Proposition 5.9.1(iii) implies that  $\bar{d}_{\tilde{G}, Q}^j(y_\ell) < \beta n$ . Thus

$$b_3 \leq \binom{n}{\leq \beta n}^3 \leq 2^{3\xi(\beta)n},$$

as required.

**Claim 4:**  $b_4 \leq 2^{3((k-2)n/(k-1)+\mu n)-n/128k^2}$ .

Consider the graph obtained by starting with the graph  $([n], E(G') \cup E')$  and adding edges between  $\{y_1, y_2, y_3\}$  and  $[n] \setminus Q_j$  randomly, independently, with probability  $1/2$ . Note that the number of graphs that this process can generate is at most  $2^{3((k-2)n/(k-1)+\mu n)}$ , with each such graph equally likely to be generated. So an upper bound on  $b_4$  is given by

$$b_4 \leq 2^{3((k-2)n/(k-1)+\mu n)} \mathbb{P} \left( |\bar{N}(\{y_1, y_2, y_3\}) \cap B_{low}^i| \leq \frac{n}{200k^2} \right).$$

Since  $G'$  was chosen such that  $|B_{low}^i| \geq n/2k^2$ , we have that  $\mathbb{E}(|\bar{N}(\{y_1, y_2, y_3\}) \cap B_{low}^i|) \geq$

$n/16k^2$ . So Lemma 5.2.1(ii) implies that

$$\mathbb{P}\left(|\overline{N}(\{y_1, y_2, y_3\}) \cap B_{low}^i| \leq \frac{n}{200k^2}\right) \leq \exp\left(-\frac{n}{128k^2}\right) \leq 2^{-\frac{n}{128k^2}}.$$

Hence  $b_4 \leq 2^{3((k-2)n/(k-1)+\mu n)-n/128k^2}$ , as required.

Now (5.9.6) together with Claims 1–4 and Proposition 5.4.3(ii) implies that

$$\begin{aligned} |\mathcal{A}_1| &\leq 2^3 k^2 n^3 \cdot C 2^{2(\log n)^3} f_k(n_k) 2^{t_{k-1}(n-3)} \cdot 2^{3\xi(\beta)n} \cdot 2^{3((k-2)n/(k-1)+\mu n)-n/128k^2} \\ &\leq C 2^{-n/150k^2} f_k(n_k) 2^{t_{k-1}(n-3)+3(k-2)n/(k-1)-3(k-2)-3} \\ &\leq C 2^{-n/150k^2} f_k(n_k) 2^{t_{k-1}(n)}, \end{aligned}$$

as required. □

**Lemma 5.9.7.**  $|\mathcal{A}_2| \leq C 2^{-n^{1/2k^2}/3} f_k(n_k) 2^{t_{k-1}(n)}.$

*Proof.* Note that for every  $G \in \mathcal{A}_2$  and every  $s \in \{0, 1, \dots, k-2\}$  the definition of  $C^s(G, Q)$  implies that  $|Q_s \setminus C^s(G, Q)| \geq |Q_s|/2$ . So any graph  $G \in \mathcal{A}_2$  can be constructed as follows. We first choose  $a \in \mathbb{N}$  such that  $n/2k^2 \leq a \leq n$ , and then perform the following steps.

- We choose distinct indices  $i, j \in \{0, 1, \dots, k-2\}$ , a set  $W = \{y_1, y_2\} \cup \{w_\ell^s : \ell \in [2], s \in \{0, 1, \dots, k-2\} \setminus \{j\}\}$  of vertices satisfying  $y_1, y_2 \in Q_j$  and  $w_1^s, w_2^s \in Q_s$  for every  $s \in \{0, 1, \dots, k-2\} \setminus \{j\}$ , a vertex  $y_3 \in Q_j \setminus W$ , and a set  $E$  of edges between the vertices in  $W \cup \{y_3\}$ . Let  $b_1$  denote the number of such choices. The choices in this step and the next steps will be made such that  $y_1, y_2 \notin C^j(G, Q)$  and  $w_1^s, w_2^s \notin C^s(G, Q)$  for every  $s \in \{0, 1, \dots, k-2\} \setminus \{j\}$ , and  $|B_{low}^i| = a$  and  $C(Y) \cap \overline{N}(\{y_1, y_2\}) = \emptyset$ , where  $Y := \overline{N}(\{y_1, y_2, y_3\}) \cap B_{low}^i(G, Q)$ .

- Next we choose the graph  $G'$  on vertex set  $[n] \setminus W$  such that  $G[[n] \setminus W] = G'$ . Let  $b_2$  denote the number of possibilities for  $G'$ .
- Next we choose the set  $E'$  of internal edges in  $G$  with exactly one endpoint in  $W$  such that  $E'$  is compatible with our previous choices. Let  $b_3$  denote the number of possibilities for  $E'$ .
- Next we choose the set  $E''$  of crossing edges in  $G$  between  $W$  and  $B_{low}^i \setminus W$  such that  $E''$  is compatible with our previous choices. Let  $b_4$  denote the number of possibilities for  $E''$ .
- Finally we choose the set  $E'''$  of crossing edges in  $G$  between  $W$  and  $[n] \setminus (W \cup B_{low}^i)$  such that  $E'''$  is compatible with our previous choices. Let  $b_5$  denote the number of possibilities for  $E'''$ .

Hence

$$|\mathcal{A}_2| \leq n \max_{n/2k^2 \leq a \leq n} \{b_1 \cdot b_2 \cdot b_3 \cdot b_4 \cdot b_5\}. \quad (5.9.8)$$

The main idea of the proof is that since  $Y$  is large for  $G \in \mathcal{A}_2$ , it follows that  $C(Y)$  is also large. So the assumption that every element of  $C(Y)$  has at least one neighbour in  $\{y_1, y_2\}$  places a significant restriction on the number of choices for  $G$ . The role of the  $w_\ell^s$  is to ‘balance out’ the vertex classes, i.e. in the proof of Claim 5 it will be useful that  $W$  contains two vertices from each vertex class.

The following series of claims will give upper bounds for the quantities  $b_1, \dots, b_5$ . Claims 1 and 4 are trivial, and the proof of Claim 2 proceeds in an almost identical way to that of Claim 2 in the proof of Lemma 5.9.5; we give proofs of Claims 3 and 5.

**Claim 1:**  $b_1 \leq k^2 n^{2k-1} 2^{\binom{2k-1}{2}}$ .

**Claim 2:**  $b_2 \leq C 2^{2(\log n)^3} f_k(n_k) 2^{t_{k-1}(n-(2k-2))}$ .

**Claim 3:**  $b_3 \leq n^{4(k-1)}$ .

Indeed, for every graph  $\tilde{G} \in \mathcal{A}_2$  such that  $y_1, y_2 \notin C^j(\tilde{G}, Q)$  and  $w_1^s, w_2^s \notin C^s(\tilde{G}, Q)$  for every  $s \in \{0, 1, \dots, k-2\} \setminus \{j\}$ , Proposition 5.9.1(i) implies that  $\bar{d}_{\tilde{G}, Q}^j(y_\ell), \bar{d}_{\tilde{G}, Q}^s(w_\ell^s) \leq 2$  for every  $\ell \in [2]$  and every  $s \in \{0, 1, \dots, k-2\} \setminus \{j\}$ . Thus

$$b_3 \leq n^{2|W|} \leq n^{4(k-1)},$$

as required.

**Claim 4:**  $b_4 \leq 2^{(2k-4)a}$ .

**Claim 5:**  $b_5 \leq 2^{(2k-4)(n-a)} 2^{-2n^{1/2k^2}/5}$ .

Indeed, suppose  $G$  satisfies  $C(Y) \cap \overline{N}(\{y_1, y_2\}) = \emptyset$ . Since we choose  $G$  such that  $|B_{low}^i| = a \geq n/2k^2$ , the fact that  $G \notin \mathcal{A}_1$  implies that  $|Y| > n/200k^2$ . Now the definitions of  $C_{low}^i, B_{low}^i$  imply that

$$|C(Y)| \geq \frac{200k^2|Y|}{n^{1-1/2k^2}} \geq n^{1/2k^2}.$$

So since in  $G$  every vertex in  $C(Y)$  must have at least one neighbour in  $\{y_1, y_2\}$ ,

$$\begin{aligned} b_5 &\leq 2^{2 \sum_{s \in \{0, 1, \dots, k-2\} \setminus \{j\}} |[n] \setminus (Q_s \cup B_{low}^i)|} 2^{2|[n] \setminus (Q_j \cup B_{low}^i \cup C(Y))|} 3^{|C(Y)|} \\ &\leq 2^{(2k-4)(n-a)} 2^{-2n^{1/2k^2}/5}, \end{aligned} \tag{5.9.9}$$

as required. The second inequality of (5.9.9) is where it is important that  $W$  contains two vertices from each vertex class.

Now (5.9.8) together with Claims 1–5 and Proposition 5.4.3(ii) implies that

$$\begin{aligned}
|\mathcal{A}_2| &\leq n \cdot k^2 n^{2k-1} 2^{\binom{2k-1}{2}} \cdot C 2^{2(\log n)^3} f_k(n_k) 2^{t_{k-1}(n-(2k-2))} \\
&\quad \cdot n^{4(k-1)} \cdot \max_{n/2k^2 \leq a \leq n} \left\{ 2^{(2k-4)a} \cdot 2^{(2k-4)(n-a)} 2^{-2n^{1/2k^2}/5} \right\} \\
&\leq C 2^{-n^{1/2k^2}/3} f_k(n_k) \cdot 2^{t_{k-1}(n-(2k-2)) + (2k-2)(k-2)n/(k-1) - (2k-2)(k-2) - t_{k-1}(2k-2)} \\
&\leq C 2^{-n^{1/2k^2}/3} f_k(n_k) 2^{t_{k-1}(n)},
\end{aligned}$$

as required. □

As mentioned earlier, a  $(6, 3)$ -forest (with edge set  $E$  say) is a useful building block for constructing many induced copies of  $C_{2k}$ . More precisely, in Lemma 5.9.10 we will show that there are many ‘ $E$ -compatible’ linear forests  $H$ , which play a similar role to that of the skeletons in the proof of Lemma 5.7.3. Each such  $E \cup E(H)$  gives us a non-trivial restriction on the remaining edge set, resulting in an adequate bound on  $|\mathcal{A}_3|$ .

**Lemma 5.9.10.**  $|\mathcal{A}_3| \leq C 2^{-\frac{n}{2^{14k}}} f_k(n_k) 2^{t_{k-1}(n)}$ .

*Proof.* Any graph  $G \in \mathcal{A}_3$  can be constructed as follows.

- First we choose distinct indices  $i, j \in \{0, 1, \dots, k-2\}$ , a set  $X \subseteq Q_i \cup Q_j$  of six vertices, and a set  $E$  of edges between vertices in  $X$  such that the graph  $(X, E)$  is a linear forest with at most three components (so  $E$  will be the edge set of a  $(6, 3)$ -forest in  $G$ ). Let  $b_1$  denote the number of such choices.
- Next we choose a graph  $G'$  on vertex set  $[n] \setminus X$  such that  $G'[[n] \setminus X] = G'$ . Let  $b_2$  denote the number of possibilities for  $G'$ .
- Next we choose the set  $E'$  of internal edges in  $G$  with exactly one endpoint in  $X$ . Let  $b_3$  denote the number of possibilities for  $E'$ .

- Finally we choose the set  $E''$  of crossing edges in  $G$  between  $X$  and  $[n] \setminus X$  such that  $E''$  is compatible with our previous choices. Let  $b_4$  denote the number of possibilities for  $E''$ .

Hence,

$$|\mathcal{A}_3| \leq b_1 \cdot b_2 \cdot b_3 \cdot b_4. \quad (5.9.11)$$

The following series of claims will give upper bounds for the quantities  $b_1, \dots, b_4$ . Claim 1 is trivial, and the proofs of Claims 2 and 3 follow in an almost identical way to those of Claims 2 and 3 in the proof of Lemma 5.9.5, so we give only a proof of Claim 4.

**Claim 1:**  $b_1 \leq 2^{15} k^2 n^6$ .

**Claim 2:**  $b_2 \leq C 2^{2(\log n)^3} f_k(n_k) 2^{t_{k-1}(n-6)}$ .

**Claim 3:**  $b_3 \leq 2^{6\xi(\beta)n}$ .

**Claim 4:**  $b_4 \leq 2^{\frac{6(k-2)n}{k-1}} 2^{\mu^{1/4}n} 2^{-\frac{n}{2^{13k}}}$ .

Indeed, we define an *E-compatible forest* to be a linear forest  $H$  on  $2k - 6$  vertices, with the same number of components as  $(X, E)$ , such that  $V(H) \cap Q_s$  induces a clique on two vertices for every  $s \in \{0, 1, \dots, k-2\} \setminus \{i, j\}$ . Note that an *E-compatible forest* exists since  $2k - 6 \geq 2 \cdot 3$  and  $(X, E)$  has at most three components. Moreover, an *E-compatible forest* contains a perfect matching, so Proposition 5.5.6 implies that for every graph  $\tilde{G} \in \mathcal{A}_3$ , the number of disjoint *E-compatible forests* in  $\tilde{G}$  is at least

$$\left\lfloor \frac{n/(k-1) - \mu n - 2\mu^{1/2}n}{2} \right\rfloor \geq \frac{n}{2(k-1)} - 3\mu^{1/2}n.$$

Hence  $G'$  contains at least  $n/2(k-1) - 3\mu^{1/2}n$  disjoint  $E$ -compatible forests. Now fix a set  $CF$  of  $n/2(k-1) - 3\mu^{1/2}n$  disjoint  $E$ -compatible forests in  $G'$ , and let  $H \in CF$ . Let  $h_H$  denote the number of possibilities for a set  $E^*$  of edges between  $X$  and  $V(H)$ . By Proposition 5.7.1 there exists at least one set  $\tilde{E}$  of edges between  $X$  and  $V(H)$  such that the graph  $(X \cup V(H), E \cup E(H) \cup \tilde{E})$  is isomorphic to  $C_{2k}$ . So since  $G$  must be induced- $C_{2k}$ -free, we must have that  $E^* \neq \tilde{E}$ , and hence  $h_H \leq 2^{|X||V(H)|} - 1 = 2^{12(k-3)} - 1$ . Note that the number of vertices outside  $Q_i \cup Q_j$  that are not contained in some graph  $H \in CF$  is at most  $(k-3)n/(k-1) + 2\mu n - (2k-6)(n/2(k-1) - 3\mu^{1/2}n) \leq 6k\mu^{1/2}n$ . Hence,

$$\begin{aligned} b_4 &\leq 2^{6 \cdot \max\{|Q_i|, |Q_j|\}} 2^{6(6k\mu^{1/2}n)} \prod_{H \in CF} h_H \\ &\leq 2^{6(n/(k-1) + \mu n)} 2^{6(6k\mu^{1/2}n)} \left(2^{12(k-3)} (1 - 2^{-12(k-3)})\right)^{n/(2(k-1)) - 3\mu^{1/2}n} \\ &\leq 2^{\frac{6(k-2)n}{k-1}} 2^{40k\mu^{1/2}n} e^{-\frac{n/(2(k-1))}{2^{12(k-3)}}} \leq 2^{\frac{6(k-2)n}{k-1}} 2^{\mu^{1/4}n} 2^{-\frac{n}{2^{13k}}}, \end{aligned}$$

as required.

Now (5.9.11) together with Claims 1–4 and Proposition 5.4.3(ii) implies that

$$\begin{aligned} |\mathcal{A}_3| &\leq 2^{15} k^2 n^6 \cdot C 2^{2(\log n)^3} f_k(n_k) 2^{t_{k-1}(n-6)} \cdot 2^{6\xi(\beta)n} \cdot 2^{\frac{6(k-2)n}{k-1}} 2^{\mu^{1/4}n} 2^{-\frac{n}{2^{13k}}} \\ &\leq C 2^{-\frac{n}{2^{14k}}} f_k(n_k) 2^{t_{k-1}(n-6) + 6(k-2)n/(k-1) - 6(k-2) - t_{k-1}(6)} \\ &\leq C 2^{-\frac{n}{2^{14k}}} f_k(n_k) 2^{t_{k-1}(n)}, \end{aligned}$$

as required. □

The next proposition shows that for every  $G \in \mathcal{A}_4$ , the small stars and triangles in  $\overline{G}[Q_0]$  do not cover too many vertices.



**Proposition 5.9.12.** *For every  $G \in \mathcal{A}_4$  and index  $i \in \{0, 1, \dots, k-2\}$ ,  $|B_{low}^i| < n/2k^2$ .*

*Proof.* Suppose for a contradiction that there exists a graph  $G \in \mathcal{A}_4$  such that  $|B_{low}^i| \geq n/2k^2$  for some index  $i \in \{0, 1, \dots, k-2\}$ . Since  $G \in \mathcal{A}_4 \subseteq F_Q^3$ ,  $G$  is not a  $k$ -template. This fact together with Proposition 5.9.1(i) implies that there exists an index  $j \in \{0, 1, \dots, k-2\} \setminus \{i\}$  and a non-edge  $y_1 y_3$  inside  $Q_j$ . At most one of  $y_1, y_3$  can be in  $C^j$  (by definition of  $C^j$ ), and so without loss of generality we assume that  $y_1 \notin C^j$ . So Proposition 5.9.1(i) implies that  $\bar{d}_{G,Q}^j(y_1) \leq 2$ . This together with the observation that  $|Q_j \setminus C^j| \geq |Q_j|/2$  (by definition of  $C^j$ ) implies that there exists a vertex  $y_2 \in Q_j \setminus C^j$  that is a neighbour of  $y_1$ .

Define  $Y := \bar{N}(\{y_1, y_2, y_3\}) \cap B_{low}^i$ . Since  $|B_{low}^i| \geq n/2k^2$  and  $G \notin \mathcal{A}_1$ ,  $|Y| > n/200k^2$ . Since  $|B_{low}^i| \geq n/2k^2$  and  $G \notin \mathcal{A}_2$ ,  $C(Y)$  contains a vertex  $x_3 \in \bar{N}(\{y_1, y_2\})$ . Since  $x_3 \in C(Y)$  there exists a vertex  $x_1 \in Y$  that is a non-neighbour of  $x_3$ . By Proposition 5.9.1(iii),  $\bar{d}_{G,Q}^i(x_1), \bar{d}_{G,Q}^i(x_3) \leq \beta n$ . So since  $|Y| > n/200k^2 \geq 2\beta n$ , there exists a vertex  $x_2 \in Y \cap N(\{x_1, x_3\})$ .

Then  $E(G[\{x_1, x_2, x_3, y_1, y_2, y_3\}]) = \{x_1 x_2, x_2 x_3, y_1 y_2\} \cup E'$  with  $E' \subseteq \{y_2 y_3, y_3 x_3\}$ . Thus the set  $\{x_1, x_2, x_3, y_1, y_2, y_3\} \subseteq Q_i \cup Q_j$  induces on  $G$  a linear forest with at most three components, and so  $G$  contains a  $(6, 3)$ -forest with respect to  $Q$ . This contradicts the fact that  $G \notin \mathcal{A}_3$ , and hence completes the proof.  $\square$

We now have sufficient information about the set  $\mathcal{A}_4$  of remaining graphs to count them directly (i.e.  $\mathcal{A}_4$  is the only class for which we do not use induction in our estimates). In particular, we now know that in  $\bar{G}$  every vertex class is the union of triangles and stars, where crucially the number of triangles and small stars is not too large (see Proposition 5.9.12). This allows us to show by a direct counting argument that  $|\mathcal{A}_4|$  is negligible.

**Lemma 5.9.13.**  $|\mathcal{A}_4| \leq 2^{-\frac{n \log n}{3k^2}} f_k(n_k) 2^{t_{k-1}(n)}$ .

*Proof.* Any graph  $G \in \mathcal{A}_4$  can be constructed as follows.

- First we choose a partition of  $Q_i$  into five sets,  $C_h^i, B_h^i, C_\ell^i, B_\ell^i, C_z^i$ , for every  $i \in \{0, 1, \dots, k-2\}$ . Let  $b_1$  denote the number of such choices.
- Next we choose the set  $E$  of crossing edges in  $G$  with respect to  $Q$ . Let  $b_2$  denote the number of possibilities for  $E$ .
- Finally we choose the set  $E'$  of internal edges in  $G$  with respect to  $Q$  such that  $G$  satisfies  $C_h^i = C_{high}^i$ ,  $B_h^i = B_{high}^i$ ,  $C_\ell^i = C_{low}^i$ ,  $B_\ell^i = B_{low}^i$ , and  $C_z^i = C_0^i$  for every  $i \in \{0, 1, \dots, k-2\}$ . Let  $b_3$  denote the number of possibilities for  $E'$ .

Hence

$$|\mathcal{A}_4| \leq b_1 \cdot b_2 \cdot b_3. \quad (5.9.14)$$

The following series of claims will give upper bounds for the quantities  $b_1, b_2, b_3$ . Claims 1 and 2 are trivial; we give only a proof of Claim 3.

**Claim 1:**  $b_1 \leq 5^n$ .

**Claim 2:**  $b_2 \leq 2^{t_{k-1}(n)}$ .

**Claim 3:**  $b_3 \leq 2^{\frac{(k-1/2)n \log n}{k^2}}$ .

For any given  $i \in \{0, 1, \dots, k-2\}$  and any vertex  $x \in B_{high}^i$ , the number of possibilities for the unique non-neighbour of  $x$  in  $C_{high}^i$  (namely the centre of the star in  $\overline{G}$  containing  $x$ ) is  $|C_{high}^i|$ . Now consider  $x \in B_{low}^i$ . Then  $x$  has a unique non-neighbour  $y$  in  $C_{low}^i$ , and has the possibility of either being part of a triangle in  $\overline{G}$  or a star in  $\overline{G}$ . Note also that

$|B_{low}^i| < n/2k^2$  by Proposition 5.9.12, and that by definition of  $C_{high}^i$ ,

$$|C_{high}^i| \leq \frac{200k^2n}{n^{1-1/2k^2}} \leq 200k^2n^{1/2k^2}.$$

Hence,

$$\begin{aligned} b_3 &\leq \prod_{i=0}^{k-2} (2|C_{low}^i|)^{|B_{low}^i|} |C_{high}^i|^{|B_{high}^i|} \leq \prod_{i=0}^{k-2} n^{\frac{n}{2k^2}} (200k^2)^n (n^{\frac{1}{2k^2}})^n = 2^{\frac{(k-1)n \log n}{k^2}} (200k^2)^{n(k-1)} \\ &\leq 2^{\frac{(k-1/2)n \log n}{k^2}}, \end{aligned}$$

as required.

Now (5.9.14) together with Claims 1–3 and Lemma 5.4.1 implies that

$$\begin{aligned} |\mathcal{A}_4| &\leq 5^n \cdot 2^{t_{k-1}(n)} \cdot 2^{\frac{(k-1/2)n \log n}{k^2}} \\ &\leq 5^n 2^{-\frac{n \log n}{2k^2}} 2^{n_k \log n_k - en_k \log \log n_k} 2^{en_k \log \log n_k} 2^{t_{k-1}(n)} \leq 2^{-\frac{n \log n}{3k^2}} f_k(n_k) 2^{t_{k-1}(n)}, \end{aligned}$$

as required. □

Recall that  $F_Q^3 = (F_Q^3 \setminus T_Q^*(n, k)) \cup \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \cup \mathcal{A}_4$ . The following bound on  $|F_Q^3|$  follows immediately from this observation together with (5.9.2) and Lemmas 5.9.5, 5.9.7, 5.9.10 and 5.9.13.

**Lemma 5.9.15.**  $|F_Q^3| \leq 2C 2^{-n^{\frac{1}{2k^2}}/3} f(n_k) 2^{t_{k-1}(n)}.$

## 5.10 Proof of Lemma 5.6.1

**Proof of Lemma 5.6.1.** Recall from Section 5.6 that we prove Lemma 5.6.1 by induction on  $n$  and that we choose constants satisfying (5.6.2). The fact that  $1/C \ll 1/n_0, 1/k$  implies that the statement of Lemma 5.6.1 holds for all  $n \leq n_0$ . So suppose that  $n > n_0$  and that the statement holds for all  $n' < n$ . Then we obtain the bounds in Lemmas 5.7.3, 5.8.9 and 5.9.15. These bounds together with the fact that  $F_Q(n, k, \eta, \mu) = T_Q \cup F_Q^1 \cup F_Q^2 \cup F_Q^3$  and  $T_Q \subseteq T_Q(n, k)$  imply that

$$\begin{aligned} |F_Q(n, k, \eta, \mu) \setminus T_Q(n, k)| &\leq C \left( 2^{-\beta^2 n / 14^k} + 2^{-n} + 2 \cdot 2^{-n \frac{1}{2k^2} / 3} \right) f_k(n_k) 2^{t_{k-1}(n)} \\ &\leq 3C 2^{-n \frac{1}{2k^2} / 3} f_k(n_k) 2^{t_{k-1}(n)}. \end{aligned}$$

This together with Corollary 5.5.5 implies that

$$\begin{aligned} |F_Q(n, k, \eta) \setminus T_Q(n, k)| &\leq |F_Q(n, k, \eta) \setminus F_Q(n, k, \eta, \mu)| + |F_Q(n, k, \eta, \mu) \setminus T_Q(n, k)| \\ &\leq \left( 2^{-\frac{\mu^2 n^2}{100}} + 3C 2^{-n \frac{1}{2k^2} / 3} f_k(n_k) \right) 2^{t_{k-1}(n)} \\ &\leq 4C 2^{-n \frac{1}{2k^2} / 3} f_k(n_k) 2^{t_{k-1}(n)}. \end{aligned} \tag{5.10.1}$$

Note that Lemma 5.3.1 (applied with  $\eta/2$  playing the role of  $\eta$ ) together with (5.5.1) implies that

$$|F(n, k) \setminus F(n, k, \eta)| \leq 2^{-\varepsilon n^2} |F(n, k, \eta)|. \tag{5.10.2}$$

Let  $\mathcal{Q}$  denote the set of all ordered  $(k-1)$ -partitions of  $[n]$ , and recall that our choice of  $Q \in \mathcal{Q}$  was arbitrary. Now (5.10.1) together with (5.10.2) and Lemma 5.4.4 implies that

$$\begin{aligned} |F(n, k) \setminus F(n, k, \eta)| &\leq 2^{-\varepsilon n^2} \sum_{Q' \in \mathcal{Q}} (|F_{Q'}(n, k, \eta) \setminus T_{Q'}(n, k)| + |T_{Q'}(n, k)|) \\ &\leq 2^{-\varepsilon n^2} (k-1)^n \left( 4C 2^{-n \frac{1}{2k^2}/3} + 2^{6(\log n)^2} \right) f_k(n_k) 2^{t_{k-1}(n)} \\ &\leq C 2^{-\varepsilon n^2/2} f_k(n_k) 2^{t_{k-1}(n)}. \end{aligned}$$

Now this together with (5.10.1) implies that

$$\begin{aligned} |F_Q(n, k)| &\leq |F_Q(n, k, \eta)| + |F(n, k) \setminus F(n, k, \eta)| \\ &\leq |T_Q(n, k)| + |F_Q(n, k, \eta) \setminus T_Q(n, k)| + |F(n, k) \setminus F(n, k, \eta)| \\ &\leq |T_Q(n, k)| + \left( 4 \cdot 2^{-n \frac{1}{2k^2}/3} + 2^{-\varepsilon n^2/2} \right) C f_k(n_k) 2^{t_{k-1}(n)} \\ &\leq |T_Q(n, k)| + 5C 2^{-n \frac{1}{2k^2}/3} f_k(n_k) 2^{t_{k-1}(n)}, \end{aligned}$$

which completes the inductive step, and hence the proof.  $\square$

## 5.11 The Removal Lemma for complete 2-coloured multigraphs

This section is devoted to proving Theorem 5.3.5. The proof is similar to that of the Induced Removal Lemma, and we follow its proof as given in [4]. We begin with the following definitions.

For every two nonempty disjoint vertex sets  $A, B$  of a graph  $G$  we define  $e(A, B)$  to be the number of edges of  $G$  between  $A$  and  $B$ . The *edge density* of the pair is defined by

$d^G(A, B) = e(A, B)/|A||B|$ . We say that the pair  $A, B$  is  $\gamma$ -regular in  $G$  if for any two subsets  $A' \subseteq A$  and  $B' \subseteq B$  satisfying  $|A'| \geq \gamma|A|$  and  $|B'| \geq \gamma|B|$ , their edge density satisfies  $|d^G(A', B') - d^G(A, B)| < \gamma$ .

If  $c$  is an edge-colouring of  $G$  with colours  $c_1, \dots, c_\ell$  we define  $e_{c_i}(A, B)$  to be the number of edges of colour  $c_i$  between  $A$  and  $B$ . Then we define the  $c_i$ -edge density of  $A, B$  by  $d_{c_i}^G(A, B) = e_{c_i}(A, B)/|A||B|$ . We say that the pair  $A, B$  is  $(\gamma, c)$ -regular if for any two subsets  $A' \subseteq A$  and  $B' \subseteq B$  satisfying  $|A'| \geq \gamma|A|$  and  $|B'| \geq \gamma|B|$ , their  $c_i$ -edge density satisfies  $|d_{c_i}^G(A', B') - d_{c_i}^G(A, B)| < \gamma$  for every  $0 < i \leq \ell$ .

If  $G'$  is a 2-coloured multigraph we define  $e_{red}(A, B), e_{blue}(A, B)$  to be the number of edges between  $A$  and  $B$  that are in  $G'_R, G'_B$  respectively. Then we define the *red-edge density* of  $A, B$  by  $d_{red}^{G'}(A, B) = e_{red}(A, B)/|A||B|$ , and we define the *blue-edge density*  $d_{blue}^{G'}(A, B)$  similarly. We say that the pair  $A, B$  is  $\gamma$ -regular in  $G'$  if for any two subsets  $A' \subseteq A$  and  $B' \subseteq B$  satisfying  $|A'| \geq \gamma|A|$  and  $|B'| \geq \gamma|B|$ , their red-edge density satisfies  $|d_{red}^{G'}(A', B') - d_{red}^{G'}(A, B)| < \gamma$  and their blue-edge density satisfies  $|d_{blue}^{G'}(A', B') - d_{blue}^{G'}(A, B)| < \gamma$ .

We will sometimes write simply  $d(A, B)$  for  $d^G(A, B)$  when it is clear what graph we are working with (and similarly for  $d_{c_i}^G(A, B), d_{red}^G(A, B), d_{blue}^G(A, B)$ ).

One simple but useful property of regularity is that it is somewhat preserved when passing to subsets, as the following trivial proposition shows.

**Proposition 5.11.1.** *Let  $\delta > 0$  and  $\varepsilon > \gamma > 0$ . If  $A, B$  is a  $\gamma$ -regular pair with density  $\delta$  and  $A' \subseteq A$  and  $B' \subseteq B$  satisfy  $|A'| \geq \varepsilon|A|$  and  $|B'| \geq \varepsilon|B|$ , then  $A', B'$  is a  $\gamma \max\{2, \varepsilon^{-1}\}$ -regular pair with edge density at least  $\delta - \gamma$  and at most  $\delta + \gamma$ .*

The following lemma shows how the existence of regular pairs implies the existence of

many induced copies of a given fixed subgraph. For a proof see e.g. [4, Lemma 3.2].

**Lemma 5.11.2.** *For every  $0 < \eta < 1$  and  $k \in \mathbb{N}$  there exist  $\delta, \gamma > 0$  such that the following holds. Suppose that  $H$  is a fixed graph with vertices  $v_1, \dots, v_k$  and that  $V_1, \dots, V_k$  are disjoint vertex sets of a graph  $G$  such that for every  $0 < i < i' \leq k$  the pair  $V_i, V_{i'}$  is  $\gamma$ -regular with edge density at least  $\eta$  if  $v_i v_{i'}$  is an edge of  $H$ , and with edge density at most  $1 - \eta$  if  $v_i v_{i'}$  is not an edge of  $H$ . Then at least  $\delta \prod_{i=1}^k |V_i|$  of the  $k$ -tuples  $w_1, \dots, w_k$  with  $w_1 \in V_1, \dots, w_k \in V_k$  span induced copies of  $H$  in  $G$ , where  $w_i$  plays the role of  $v_i$  for every  $0 < i \leq k$ .*

The next lemma is an analogue of Lemma 5.11.2 for complete 2-coloured multigraphs. Note that in order to keep track of, and easily refer to, the large amount of constants in this section we sometimes define constants with a subscript that denotes the number of the result with which they are associated, as seen below.

**Lemma 5.11.3.** *For every  $0 < \eta < 1$  and  $k \in \mathbb{N}$  there exist  $\delta = \delta_{(5.11.3)}(\eta, k)$  and  $\gamma = \gamma_{(5.11.3)}(\eta, k) > 0$  such that the following holds. Suppose that  $H$  is a fixed graph with vertices  $v_1, \dots, v_k$  and that  $V_1, \dots, V_k$  are disjoint vertex sets of a complete 2-coloured multigraph  $G$  such that for every  $0 < i < i' \leq k$  the pair  $V_i, V_{i'}$  is  $\gamma$ -regular with red-edge density at least  $\eta$  if  $v_i v_{i'}$  is an edge of  $H$ , and with blue-edge density at least  $\eta$  if  $v_i v_{i'}$  is not an edge of  $H$ . Then at least  $\delta \prod_{i=1}^k |V_i|$  of the  $k$ -tuples  $w_1, \dots, w_k$  with  $w_1 \in V_1, \dots, w_k \in V_k$  span copies of  $H$  in  $G$ , where  $w_i$  plays the role of  $v_i$  for every  $0 < i \leq k$ .*

*Proof.* We first define  $E$  to be the set obtained from  $G_R$  by removing every edge  $e \in G_R \cap G_B$  between sets  $V_i, V_{i'}$  that are such that  $v_i v_{i'}$  is not an edge of  $H$ . Now define the graph  $G' := (V(G), E)$ . Then the assumptions of Lemma 5.11.2 (with  $G'$  playing the role of  $G$ ) are satisfied. In particular note that for non-edges  $v_i v_{i'}$  of  $H$ , the pair  $V_i, V_{i'}$  is  $\gamma$ -regular in  $G'$  since for any two subsets  $A \subseteq V_i$  and  $B \subseteq V_{i'}$  satisfying  $|A'| \geq \gamma|V_i|$  and

$|B'| \geq \gamma|V_{i'}|$ , we have that

$$|d^{G'}(A, B) - d^{G'}(V_i, V_{i'})| = |\overline{d^{G'}}(A, B) - \overline{d^{G'}}(V_i, V_{i'})| = |d_{blue}^G(A, B) - d_{blue}^G(V_i, V_{i'})| < \gamma.$$

Thus at least  $\delta \prod_{i=1}^k |V_i|$  of the  $k$ -tuples  $w_1, \dots, w_k$  with  $w_1 \in V_1, \dots, w_k \in V_k$  span induced copies of  $H$  in  $G'$ , where  $w_i$  plays the role of  $v_i$  for every  $0 < i \leq k$ . Now the way  $G'$  was constructed from  $G$  implies that each of these at least  $\delta \prod_{i=1}^k |V_i|$   $k$ -tuples  $w_1, \dots, w_k$  span copies of  $H$  in  $G$ , where  $w_i$  plays the role of  $v_i$  for every  $0 < i \leq k$ . This completes the proof.  $\square$

We say that a partition  $\mathcal{A} = \{V_i : 1 \leq i \leq k\}$  of the vertex set of a graph is *balanced* if every pair of partition classes differ in size by at most one. A refinement of a balanced partition  $\mathcal{A}$  is a balanced partition of the form  $\mathcal{B} = \{V_{i,j} : 1 \leq i \leq k, 1 \leq j \leq \ell\}$  such that  $V_{i,j} \subseteq V_i$  for every  $1 \leq i \leq k$  and  $1 \leq j \leq \ell$ . The *index* of  $\mathcal{A}$  with respect to a graph  $G$  is defined by

$$\text{ind}^G(\mathcal{A}) = \frac{1}{k^2} \sum_{1 \leq i < i' \leq k} (d^G(V_i, V_{i'}))^2.$$

Note that since  $d^G(V_i, V_{i'}) = (\sum_{1 \leq j, j' \leq \ell} d^G(V_{i,j}, V_{i',j'}))/\ell^2$ , Jansen's inequality implies that

$$\text{ind}^G(\mathcal{B}) \geq \text{ind}^G(\mathcal{A}). \quad (5.11.4)$$

For an edge-colouring  $c$  of  $G$  with colours  $c_1, \dots, c_\ell$ , the *colour-index* of  $\mathcal{A}$  with respect to a graph  $G$  and a colouring  $c$  is defined by

$$\text{ind}_c^G(\mathcal{A}) = \frac{1}{k^2} \sum_{1 \leq j \leq \ell} \sum_{1 \leq i < i' \leq k} (d_{c_j}^G(V_i, V_{i'}))^2.$$

The following is a formulation of the ‘many colours’ version of Szemerédi’s regularity



lemma, one version of which is detailed in [47, Theorem 1.18]. It follows by an almost identical proof to that of the original regularity lemma, just replacing the notion of the index  $\text{ind}^G(\mathcal{A})$  with that of the colour-index  $\text{ind}_c^G(\mathcal{A})$ .

**Theorem 5.11.5.** *For every  $m, t \in \mathbb{N}$  and  $\varepsilon > 0$  there exists an integer  $T = T_{(5.11.5)}(m, \varepsilon, t)$  such that the following holds. If  $G$  is a graph with  $n \geq T$  vertices,  $c$  an edge-colouring of  $G$  with  $t$  colours, and  $\mathcal{A}$  a balanced partition of the vertex set of  $G$  with at most  $m$  vertex classes, then there exists a refinement  $\mathcal{B}$  of  $\mathcal{A}$  with  $k$  vertex classes, where  $m \leq k \leq T$ , for which all but at most  $\varepsilon \binom{k}{2}$  pairs of partition classes of  $\mathcal{B}$  are  $(\varepsilon, c)$ -regular.*

Note that the original formulation of the above lemma allows also for a set of up to  $\varepsilon n$  exceptional vertices outside of this balanced partition, but one can first apply the original formulation with a somewhat smaller parameter in place of  $\varepsilon$ , and then evenly distribute these exceptional vertices among the sets of the partition to obtain Theorem 5.11.5.

Note that the function  $T_{(5.11.5)}(m, \varepsilon, t)$  detailed in Theorem 5.11.5 may be assumed to be monotone increasing in  $m$  and  $t$  and monotone decreasing in  $\varepsilon$ . We also assume similar monotonicity properties for other functions appearing here, and assume that the number of vertices  $n$  of the graph is sufficiently large, even when we do not explicitly mention it.

We define a bijection  $f$  between complete 2-coloured multigraphs and complete graphs with a edge colouring using only colours  $R, B, P$ , as follows. Given a complete graph  $G$  and an edge colouring  $c$  of  $G$  using only colours  $R, B, P$ , we set  $f(G)$  to be the complete 2-coloured multigraph on vertex set  $V(G)$  with edge sets  $G'_R, G'_B$  defined as follows.

- $G'_R \setminus G'_B$  is the set of edges of  $G$  coloured  $R$  by  $c$ .
- $G'_B \setminus G'_R$  is the set of edges of  $G$  coloured  $B$  by  $c$ .

- $G'_R \cap G'_B$  is the set of edges of  $G$  coloured  $P$  by  $c$ .

The following proposition allows us to easily pass from regular pairs in such an edge-coloured graph  $G$  to regular pairs in the corresponding complete 2-coloured multigraph  $f(G)$ .

**Proposition 5.11.6.** *Let  $\gamma > 0$ , let  $G$  be a complete graph, let  $c$  be an edge colouring of  $G$  with colours  $R, B, P$ , let  $U, V \subseteq V(G)$  be disjoint and let  $U' \subseteq U$  and  $V' \subseteq V$ . If  $|d_\ell^G(U, V) - d_\ell^G(U', V')| < \gamma$  for every  $\ell \in \{R, B, P\}$  then  $|d_{red}^{f(G)}(U', V') - d_{red}^{f(G)}(U, V)| < 2\gamma$  and  $|d_{blue}^{f(G)}(U', V') - d_{blue}^{f(G)}(U, V)| < 2\gamma$ . In particular, if  $U, V$  is a  $(\gamma, c)$ -regular pair in  $G$  then  $U, V$  is a  $2\gamma$ -regular pair in  $f(G)$ .*

*Proof.* By the triangle inequality,

$$\begin{aligned} |d_{red}^{f(G)}(U', V') - d_{red}^{f(G)}(U, V)| &= |(d_R^G(U', V') + d_P^G(U', V')) - (d_R^G(U, V) + d_P^G(U, V))| \\ &\leq |d_R^G(U', V') - d_R^G(U, V)| + |d_P^G(U', V') - d_P^G(U, V)| < 2\gamma, \end{aligned}$$

and similarly for blue. □

The following theorem is another analogue of Szemerédi's regularity lemma, this time applicable to complete 2-coloured multigraphs.

**Theorem 5.11.7.** *For every  $m \in \mathbb{N}$  and  $\varepsilon > 0$  there exists an integer  $T$  such that the following holds. If  $G$  is a complete 2-coloured multigraph with  $n \geq T$  vertices and  $\mathcal{A}$  is a balanced partition of the vertex set of  $G$  with at most  $m$  vertex classes, then there exists a refinement  $\mathcal{B}$  of  $\mathcal{A}$  with  $k$  vertex classes, where  $m \leq k \leq T$ , for which all but at most  $\varepsilon \binom{k}{2}$  pairs of partition classes of  $\mathcal{B}$  are  $\varepsilon$ -regular.*

Theorem 5.11.7 follows immediately from Theorem 5.11.5 (applied with  $f^{-1}(G), \varepsilon/2$  playing the roles of  $G, \varepsilon$  respectively) and Proposition 5.11.6.

The following corollary to Theorem 5.11.7 is useful to what follows. We omit the proof as it consists of an almost identical Ramsey argument to that in the proof of the analogous result for graphs (see [4, Corollary 3.4]), simply using Theorem 5.11.7 instead of the usual regularity lemma.

**Corollary 5.11.8.** *For every  $\ell \in \mathbb{N}$  and  $\gamma > 0$  there exists  $\delta = \delta_{(5.11.8)}(\ell, \gamma)$  such that for every complete 2-coloured multigraph  $G$  with  $n \geq \delta^{-1}$  vertices there exist disjoint vertex sets  $W_1, \dots, W_\ell \subseteq V(G)$  satisfying the following.*

- $|W_i| \geq \delta n$  for every  $1 \leq i \leq \ell$ .
- $W_i, W_j$  is  $\gamma$ -regular for all  $1 \leq i < j \leq \ell$ .
- Either  $d_{\text{red}}(W_i, W_j) \geq 1/2$  for all  $1 \leq i < j \leq \ell$ , or else  $d_{\text{blue}}(W_i, W_j) \geq 1/2$  for all  $1 \leq i < j \leq \ell$ , (or both).

Returning briefly to (uncoloured) graphs, the following lemma shows that if the index of a balanced partition  $\mathcal{A}$  is not much smaller than the index of its refinement  $\mathcal{B}$  then most of the pairs of partition classes of  $\mathcal{B}$  have densities that are close to the densities of the corresponding pairs of  $\mathcal{A}$ . For a proof see [4, Lemma 3.7].

**Lemma 5.11.9.** *Let  $\varepsilon > 0$  and  $k, \ell \in \mathbb{N}$ . Suppose that a balanced partition  $\mathcal{A} = \{V_i : 1 \leq i \leq k\}$  of a graph  $G$  and its refinement  $\mathcal{B} = \{V_{i,j} : 1 \leq i \leq k, 1 \leq j \leq \ell\}$  satisfy  $\text{ind}^G(\mathcal{B}) - \text{ind}^G(\mathcal{A}) \leq \varepsilon^4/64$ , and that the number of vertices of  $G$  is  $n > 512\varepsilon^{-4}k\ell$ . Then all but at most  $\varepsilon \binom{k}{2}$  pairs  $1 \leq i < i' \leq k$  are such that all but at most  $\varepsilon \ell^2$  pairs  $1 \leq j, j' \leq \ell$  satisfy  $|d(V_{i,j}, V_{i',j'}) - d(V_i, V_{i'})| < \varepsilon$ .*

The following lemma is a ‘many colours’ analogue of Lemma 5.11.9.

**Lemma 5.11.10.** *Let  $\varepsilon > 0$  and  $k, \ell, t \in \mathbb{N}$ . Let  $G$  be a graph and let  $c$  be an edge-colouring of  $G$  with colours  $c_1, \dots, c_t$ . Suppose that a balanced partition  $\mathcal{A} = \{V_i : 1 \leq i \leq k\}$  of  $G$  and its refinement  $\mathcal{B} = \{V_{i,j} : 1 \leq i \leq k, 1 \leq j \leq \ell\}$  satisfy  $\text{ind}_c^G(\mathcal{B}) - \text{ind}_c^G(\mathcal{A}) \leq \varepsilon^4/64$ , and that the number of vertices of  $G$  is  $n > 512\varepsilon^{-4}k\ell$ . Then all but at most  $t\varepsilon \binom{k}{2}$  pairs  $1 \leq i < i' \leq k$  are such that all but at most  $t\varepsilon\ell^2$  pairs  $1 \leq j, j' \leq \ell$  satisfy  $|d_{c_q}(V_{i,j}, V_{i',j'}) - d_{c_q}(V_i, V_{i'})| < \varepsilon$  for every  $1 \leq q \leq t$ .*

*Proof.* For every  $1 \leq q \leq t$  and all pairs  $1 \leq i < i' \leq k$ , let  $J_q(i, i')$  be the set of all pairs  $1 \leq j, j' \leq \ell$  that satisfy  $|d_{c_q}^G(V_{i,j}, V_{i',j'}) - d_{c_q}^G(V_i, V_{i'})| \geq \varepsilon$ . Let  $I_q$  be the set of all pairs  $1 \leq i < i' \leq k$  for which  $|J_q(i, i')| > \varepsilon\ell^2$ . For every  $1 \leq q \leq t$ , define  $G_q$  to be the (uncoloured) graph on vertex set  $V(G)$  with edge set equal to the set of  $q$ -coloured edges of  $G$ . Note that  $|d^{G_q}(V_{i,j}, V_{i',j'}) - d^{G_q}(V_i, V_{i'})| = |d_{c_q}^G(V_{i,j}, V_{i',j'}) - d_{c_q}^G(V_i, V_{i'})|$  for all  $1 \leq i < i' \leq k$  and  $1 \leq j, j' \leq \ell$ . For every  $1 \leq q \leq t$ , (5.11.4) implies that  $\text{ind}^{G_q}(\mathcal{B}) - \text{ind}^{G_q}(\mathcal{A}) \leq \text{ind}_c^G(\mathcal{B}) - \text{ind}_c^G(\mathcal{A}) \leq \varepsilon^4/64$ , and so we can apply Lemma 5.11.9 to  $G_q$  to obtain that  $|I_q| \leq \varepsilon \binom{k}{2}$ .

For all pairs  $1 \leq i < i' \leq k$ , let  $J(i, i')$  be the set of all pairs  $1 \leq j, j' \leq \ell$  that satisfy  $|d_{c_q}^G(V_{i,j}, V_{i',j'}) - d_{c_q}^G(V_i, V_{i'})| \geq \varepsilon$  for some  $1 \leq q \leq t$ . Let  $I$  be the set of all pairs  $1 \leq i < i' \leq k$  for which  $|J(i, i')| > t\varepsilon\ell^2$ .

Now suppose for a contradiction that  $(i, i') \in I$  but that for every  $1 \leq q \leq t$ ,  $(i, i') \notin I_q$ . By definition of  $I_q$  this means that  $|J_q(i, i')| \leq \varepsilon\ell^2$  for every  $1 \leq q \leq t$ . So  $|J(i, i')| \leq \sum_{1 \leq q \leq t} |J_q(i, i')| \leq t\varepsilon\ell^2$ . But then  $(i, i') \notin I$  by definition; a contradiction. Hence  $I \subseteq \bigcup_{1 \leq q \leq t} I_q$ . Now the union bound yields that  $|I| \leq \sum_{1 \leq q \leq t} |I_q| \leq t\varepsilon \binom{k}{2}$ , as required.  $\square$

The following lemma is a variant of the many colours regularity lemma more suited to our purposes. We omit its proof since it is almost identical to that of the analogous result for graphs (see [4, Lemma 4.1]), just using Theorem 5.11.5 and Lemma 5.11.10 instead of

the corresponding graph results.

**Lemma 5.11.11.** *For every  $m \in \mathbb{N}$ , every  $t \in \mathbb{N}$ , and every function  $0 < \mathcal{E}(r) < 1$  there exists an integer  $S$  such that the following holds. If  $G$  is a graph with  $n \geq S$  vertices and  $c$  is an edge-colouring of  $G$  with colours  $c_1, \dots, c_t$  then for some  $k \geq m$  and  $\ell \leq S/k$  there exists a balanced partition  $\mathcal{A} = \{V_i : 1 \leq i \leq k\}$  of  $G$  and a refinement  $\mathcal{B} = \{V_{i,j} : 1 \leq i \leq k, 1 \leq j \leq \ell\}$  of  $\mathcal{A}$  that satisfy the following.*

- *For all but at most  $\mathcal{E}(0)\binom{k}{2}$  pairs  $1 \leq i < i' \leq k$ , the pair  $V_i, V_{i'}$  is  $(\mathcal{E}(0), c)$ -regular.*
- *All pairs  $1 \leq i < i' \leq k$  are such that for all but at most  $\mathcal{E}(k)\binom{\ell}{2}$  pairs  $1 \leq j, j' \leq \ell$ , the pair  $V_{i,j}, V_{i',j'}$  is  $(\mathcal{E}(k), c)$ -regular.*
- *All but at most  $t\mathcal{E}(0)\binom{k}{2}$  pairs  $1 \leq i < i' \leq k$  are such that all but at most  $t\mathcal{E}(0)\ell^2$  pairs  $1 \leq j, j' \leq \ell$  satisfy  $|d_{c_q}(V_{i,j}, V_{i',j'}) - d_{c_q}(V_i, V_{i'})| < \mathcal{E}(0)$  for every  $1 \leq q \leq t$ .*

In what follows we need the following corollary to Lemma 5.11.11. We omit its proof since it is almost identical to that of the analogous result for graphs (see [4, Lemma 4.2]), just using Lemma 5.11.11 (applied with  $\mathcal{E}(r)/t$  playing the role of  $\mathcal{E}(r)$ ) instead of the corresponding graph result.

**Corollary 5.11.12.** *For every  $m \in \mathbb{N}$ , every  $t \in \mathbb{N}$ , and every function  $0 < \mathcal{E}(r) < 1$  there exists an integer  $S$  and a real number  $\delta$  such that the following holds. If  $G$  is a graph with  $n \geq S$  vertices and  $c$  is an edge-colouring of  $G$  with colours  $c_1, \dots, c_t$  then for some  $m \leq k \leq S$  there exists a balanced partition  $\mathcal{A} = \{V_i : 1 \leq i \leq k\}$  of  $G$  and an induced subgraph  $G'$  of  $G$  with a balanced partition  $\mathcal{A}' = \{V'_i : 1 \leq i \leq k\}$  that satisfy the following.*

- *For all  $1 \leq i \leq k$ ,  $V'_i \subseteq V_i$  and  $|V'_i| \geq \delta n$ .*

- For all pairs  $1 \leq i < i' \leq k$  the pair  $V_i', V_{i'}'$  is  $(\mathcal{E}(k), c)$ -regular.
- All but at most  $\mathcal{E}(0) \binom{k}{2}$  pairs  $1 \leq i < i' \leq k$  satisfy  $|d_{c_q}(V_i, V_{i'}) - d_{c_q}(V_i', V_{i'}')| < \mathcal{E}(0)$  for every  $1 \leq q \leq t$ .

The following lemma is an analogue of Corollary 5.11.12 applicable to complete 2-coloured multigraphs.

**Lemma 5.11.13.** *For every integer  $m$  and real  $0 < \varepsilon < 1$  there exists an integer  $S = S_{(5.11.13)}(m, \varepsilon)$  and a real number  $\delta = \delta_{(5.11.13)}(m, \varepsilon)$  such that the following holds. If  $G$  is a complete 2-coloured multigraph with  $n \geq S$  vertices then for some  $m \leq k \leq S$  there exists a balanced partition  $\mathcal{A} = \{V_i : 1 \leq i \leq k\}$  of  $V(G)$  and a balanced partition  $\mathcal{A}' = \{V_i' : 1 \leq i \leq k\}$  of a subset of  $V(G)$  that induces on  $G$  a complete 2-coloured submultigraph  $G'$  that satisfy the following.*

- For all  $1 \leq i \leq k$ ,  $V_i' \subseteq V_i$  and  $|V_i'| \geq \delta n$ .
- For all pairs  $1 \leq i < i' \leq k$  the pair  $V_i', V_{i'}'$  is  $\varepsilon$ -regular.
- All but at most  $\varepsilon \binom{k}{2}$  pairs  $1 \leq i < i' \leq k$  satisfy  $|d_{red}(V_i, V_{i'}) - d_{red}(V_i', V_{i'}')| < \varepsilon$  and  $|d_{blue}(V_i, V_{i'}) - d_{blue}(V_i', V_{i'}')| < \varepsilon$ .

Lemma 5.11.13 follows immediately from Corollary 5.11.12 (applied with  $f^{-1}(G)$  and the constant function that takes value  $\varepsilon/2$  everywhere playing the roles of  $G$  and  $\mathcal{E}(r)$  respectively) and Proposition 5.11.6. The key point of Lemma 5.11.13 is that in the partition  $\mathcal{A}'$ , the pair  $V_i', V_{i'}'$  is  $\mathcal{E}(k)$ -regular for *all* pairs  $1 \leq i < i' \leq k$ , rather than just for most of them.

For two sets  $A, B$ , we denote their symmetric difference by  $A \triangle B$ . For 2-coloured multigraphs  $G, G'$  we define their distance by  $\text{dist}(G, G') := |G_R \triangle G'_R| + |G_B \triangle G'_B|$ .

We are now in a position to prove Theorem 5.3.5.

**Proof of Theorem 5.3.5.** We prove the contrapositive; that is we assume that every  $H$ -free complete 2-coloured multigraph  $\widehat{G}$  on vertex set  $V(G)$  satisfies  $\text{dist}(G, \widehat{G}) > \varepsilon n^2$ , and we show that this implies that  $G$  contains more than  $\delta n^\ell$  copies of  $H$ .

We assume  $\varepsilon < 1$  and define  $m := 25/\varepsilon$  and  $\delta_{(5.11.3)}^* := \delta_{(5.11.3)}(\varepsilon/24, \ell)$  and  $\gamma_{(5.11.3)}^* := \gamma_{(5.11.3)}(\varepsilon/24, \ell)$ . We set

$$\delta := \delta_{(5.11.3)}^* \cdot \left( \beta \delta_{(5.11.13)} \left( m, \min \left\{ \frac{\varepsilon}{24}, \alpha \right\} \right) \right)^\ell,$$

with  $\alpha := \min\{\varepsilon/24, \beta \gamma_{(5.11.3)}^*\}$  and  $\beta := \delta_{(5.11.8)}(\ell, \gamma_{(5.11.3)}^*)$ .

We apply Lemma 5.11.13 to  $G$  to find  $m \leq k \leq S_{(5.11.13)}(m, \min\{\varepsilon/24, \alpha\})$  and  $\mathcal{A} = \{V_i : 1 \leq i \leq k\}$ ,  $G'$  and  $\mathcal{A}' = \{V'_i : 1 \leq i \leq k\}$  that satisfy  $|V'_i| \geq \delta_{(5.11.13)}(m, \min\{\varepsilon/24, \alpha\})n$ , ensuring also that all pairs of partition classes of  $\mathcal{A}'$  are in particular  $\alpha$ -regular, and that the red-edge densities and blue-edge densities of all but at most  $\varepsilon \binom{k}{2}/24$  of them differ from those of the corresponding pairs of  $\mathcal{A}$  by at most  $\varepsilon/24$ .

Now for each  $1 \leq i \leq k$  we use Corollary 5.11.8 on  $G[V'_i]$  to obtain vertex sets  $W_{i,j}$  that satisfy the following.

(W1)  $|W_{i,j}| \geq \beta |V'_i|$  for every  $1 \leq j \leq \ell$ .

(W2)  $W_{i,j}, W_{i,j'}$  is  $\gamma_{(5.11.3)}^*$ -regular for all  $1 \leq j < j' \leq \ell$ .

(W3) Either  $d_{\text{red}}(W_{i,j}, W_{i,j'}) \geq 1/2$  for all  $1 \leq j < j' \leq \ell$ , or else  $d_{\text{blue}}(W_{i,j}, W_{i,j'}) \geq 1/2$  for all  $1 \leq j < j' \leq \ell$ , (or both).

Note that by Proposition 5.11.1, for all  $1 \leq i < i' \leq k$  and  $1 \leq j, j' \leq \ell$ , the pair  $W_{i,j}, W_{i',j'}$  is  $\gamma_{(5.11.3)}^*$ -regular, and its red-edge density and blue-edge density both differ from those of  $V_i', V_{i'}'$  by at most  $\varepsilon/24$ .

We define  $\tilde{G}$  to be the complete 2-coloured multigraph obtained from  $G$  by making the following changes to the sets  $G_R, G_B$ .

- (i) For all  $1 \leq i < i' \leq k$  for which either  $|d_{red}^G(V_i, V_{i'}) - d_{red}^G(V_i', V_{i'}')| \geq \varepsilon/24$  or  $|d_{blue}^G(V_i, V_{i'}) - d_{blue}^G(V_i', V_{i'}')| \geq \varepsilon/24$ , for all  $v \in V_i$  and  $v' \in V_{i'}$  we ensure that  $vv' \in \tilde{G}_B \setminus \tilde{G}_R$  if  $|d_{red}^G(V_i', V_{i'}')| \leq 1/2$ , and we ensure that  $vv' \in \tilde{G}_R \setminus \tilde{G}_B$  otherwise. Since there are at most  $\varepsilon \binom{k}{2}/24$  such pairs  $1 \leq i < i' \leq k$ , these changes can be made so as to add at most  $2\varepsilon \binom{n}{2}/24$  edges to each of  $G_R, G_B$  and remove at most  $2\varepsilon \binom{n}{2}/24$  edges from each of  $G_R, G_B$ , for sufficiently large  $n$ .
- (ii) For all  $1 \leq i < i' \leq k$  for which (i) does not apply and that are such that  $|d_{red}^G(V_i', V_{i'}')| < 2\varepsilon/24$ , for all  $v \in V_i$  and  $v' \in V_{i'}$  we ensure that  $vv' \in \tilde{G}_B \setminus \tilde{G}_R$ . Similarly, for all  $1 \leq i < i' \leq k$  for which (i) does not apply and that are such that  $|d_{blue}^G(V_i', V_{i'}')| < 2\varepsilon/24$ , for all  $v \in V_i$  and  $v' \in V_{i'}$  we ensure that  $vv' \in \tilde{G}_R \setminus \tilde{G}_B$ . These changes can be made so as to add at most  $3\varepsilon \binom{n}{2}/24$  edges to each of  $G_R, G_B$  and remove at most  $3\varepsilon \binom{n}{2}/24$  edges from each of  $G_R, G_B$ .
- (iii) For each fixed  $1 \leq i \leq k$ , if all red-edge densities of pairs from  $W_{i,1}, \dots, W_{i,\ell}$  are at least  $1/2$  then for all distinct  $v, v' \in V_i$  we ensure that  $vv' \in \tilde{G}_R \setminus \tilde{G}_B$ . Otherwise (W3) implies that all blue-edge densities of pairs from  $W_{i,1}, \dots, W_{i,\ell}$  are at least  $1/2$  in which case for all distinct  $v, v' \in V_i$  we ensure that  $vv' \in \tilde{G}_B \setminus \tilde{G}_R$ . By the choice of  $m$  these changes can be made so as to add at most  $\varepsilon \binom{n}{2}/24$  edges to each of  $G_R, G_B$  and remove at most  $\varepsilon \binom{n}{2}/24$  edges from each of  $G_R, G_B$ .

Thus  $\text{dist}(G, \tilde{G}) \leq 4(2\varepsilon \binom{n}{2} + 3\varepsilon \binom{n}{2} + \varepsilon \binom{n}{2})/24 \leq \varepsilon n^2$ . So by assumption  $\tilde{G}$  contains a copy



of  $H$ . Denote the vertices of this copy  $v_1, \dots, v_\ell$  and choose  $i_1, \dots, i_\ell$  such that  $v_j \in V_{i_j}$  for every  $1 \leq j \leq \ell$ . Now suppose that  $v_j v_{j'}$  corresponds to an edge in  $H$ . Then  $v_j v_{j'}$  is a red edge in  $\tilde{G}$ . We claim that the pair  $W_{i_j, j}, W_{i_{j'}, j'}$  has red-edge density at least  $\varepsilon/24$  in  $G$ . To show this we consider three cases.

**Case 1:**  $i_j = i_{j'}$ .

In this case (iii) together with the fact that  $v_j v_{j'}$  is a red edge in  $\tilde{G}$  implies that  $W_{i_j, j}, W_{i_{j'}, j'}$  has red-edge density at least  $1/2 \geq \varepsilon/24$  in  $G$ , as required.

**Case 2:**  $i_j \neq i_{j'}$  and either  $|d_{red}^G(V_{i_j}, V_{i_{j'}}) - d_{red}^G(V'_{i_j}, V'_{i_{j'}})| \geq \varepsilon/24$  or else  $|d_{blue}^G(V_{i_j}, V_{i_{j'}}) - d_{blue}^G(V'_{i_j}, V'_{i_{j'}})| \geq \varepsilon/24$ .

In this case (i) together with the fact that  $v_j v_{j'}$  is a red edge in  $\tilde{G}$  implies that  $d_{red}^G(V'_{i_j}, V'_{i_{j'}}) \geq 1/2$ , and hence  $W_{i_j, j}, W_{i_{j'}, j'}$  has red-edge density at least  $1/2 - \varepsilon/24 \geq \varepsilon/24$  in  $G$ , as required.

**Case 3:**  $i_j \neq i_{j'}$  and  $|d_{red}^G(V_{i_j}, V_{i_{j'}}) - d_{red}^G(V'_{i_j}, V'_{i_{j'}})| < \varepsilon/24$  and  $|d_{blue}^G(V_{i_j}, V_{i_{j'}}) - d_{blue}^G(V'_{i_j}, V'_{i_{j'}})| < \varepsilon/24$ .

In this case (ii) together with the fact that  $v_j v_{j'}$  is a red edge in  $\tilde{G}$  implies that  $d_{red}^G(V'_{i_j}, V'_{i_{j'}}) \geq 2\varepsilon/24$ , and hence  $W_{i_j, j}, W_{i_{j'}, j'}$  has red-edge density at least  $2\varepsilon/24 - \varepsilon/24 = \varepsilon/24$  in  $G$ , as required.

This covers all cases and so proves the claim.

Similarly we can show that if  $v_j v_{j'}$  corresponds to a non-edge in  $H$  then the pair  $W_{i_j, j}, W_{i_{j'}, j'}$

has blue-edge density at least  $\varepsilon/24$  in  $G$ . So recalling that we observed that for all  $1 \leq i, i' \leq k$  and  $1 \leq j < j' \leq \ell$ , the pair  $W_{i,j}, W_{i',j'}$  is  $\gamma_{(5.11.3)}^*$ -regular in  $G$ , we see that  $W_{i_1,1}, \dots, W_{i_\ell,\ell}$  satisfy the regularity and density conditions (over  $G$ ) required for Lemma 5.11.3 to ensure the existence of at least

$$\delta_{(5.11.3)}^* \prod_{1 \leq j \leq \ell} |W_{i_j,j}| \stackrel{(W1)}{\geq} \delta_{(5.11.3)}^* \cdot \left( \beta \delta_{(5.11.13)} \left( m, \min \left\{ \frac{\varepsilon}{24}, \alpha \right\} \right) n \right)^\ell = \delta n^\ell$$

copies of  $H$  in  $G$ . □

# LIST OF REFERENCES

- [1] S. Abbasi, The solution of the El-Zahar Problem, Ph.D. Thesis, Rutgers University 1998.
- [2] V.E. Alekseev, Range of values of entropy of hereditary classes of graphs, *Diskret. Mat.* 1992, 148–157.
- [3] N. Alon, J. Balogh, B. Bollobás, R. Morris, The structure of almost all graphs in a hereditary property, *J. Combin. Theory B* **101** (2011), 85–110.
- [4] N. Alon, E. Fischer, M. Krivelevich, M. Szegedy, Efficient testing of large graphs, *Combinatorica* **20** (2000), 451–476.
- [5] N. Alon, P. Frankl, H. Huang, V. Rödl, A. Ruciński, B. Sudakov, Large matchings in uniform hypergraphs and the conjectures of Erdős and Samuels, *J. Combin. Theory A* **119** (2012), 1200–1215.
- [6] N. Alon, A. Shapira, Testing subgraphs in directed graphs, *J. Comput. System Sci.* **69** (2004), 354–382.
- [7] N. Alon, A. Shapira, A characterisation of the (natural) graph properties testable with one-sided error, in Proc. of FOCS 2005, 429–438, *SIAM J. Comput.* (Special Issue on FOCS 05) **37** (2008), 1703–1727.
- [8] J. Balogh, B. Bollobás, R. Morris, Hereditary properties of combinatorial structures: posets and oriented graphs, *J. Graph Theory* **56** (2007), 311–332.
- [9] J. Balogh, B. Bollobás, R. Morris, Hereditary properties of tournaments, *Electronic J. Combin.* **14** (2007).
- [10] J. Balogh, J. Butterfield, Excluding induced subgraphs: Critical graphs, *Random Structures and Algorithms* **38** (2011), 100–120.
- [11] J. Balogh, R. Morris, W. Samotij, Independent sets in hypergraphs, *J. Amer. Math. Soc* **28** (2015), 669–709.

- [12] J. Balogh, R. Morris, W. Samotij, L. Warnke, The typical structure of sparse  $K_{r+1}$ -free graphs, *Trans. Amer. Math. Soc.*, to appear.
- [13] Z. Baranyai, On the factorization of the complete uniform hypergraph, in: A. Hajnal, R. Rado, V.T. Sós, Infinite and finite sets, *Proc. Coll. Keszthely, 1973*, Amsterdam, Netherlands: North-Holland, (1975), 91–108.
- [14] B. Bollobás, Hereditary and monotone properties of combinatorial structures, in A. Hilton, J. Talbot, *Surveys in Combinatorics 2007* (Cambridge University Press, Cambridge, 2007), 1–40.
- [15] B. Bollobás, D.E. Daykin, P. Erdős, Sets of independent edges of a hypergraph, *Quart. J. Math. Oxford* **27** (1976), 25–32.
- [16] B. Bollobás, A. Thomason, Hereditary and monotone properties of graphs, *The mathematics of Paul Erdős II* **14** (1997), 70–78.
- [17] W.G. Brown, F. Harary, Extremal digraphs, in *Combinatorial theory and its applications*, *Coll. Math. Soc. J. Bolyai* **4** (1970), 135–198.
- [18] N.G. de Bruijn, Asymptotic methods in analysis, (Dover, New York, NY, 1958).
- [19] P. Camion, Chemins et circuits hamiltoniens des graphes complets, *C. R. Acad. Sci. Paris* **249** (1959), 2151–2152.
- [20] G. Chen, R.J. Gould, H. Li, Partitioning vertices of a tournament into independent cycles, *J. Combin. Theory B* **83** (2001), 213–220.
- [21] G. Cherlin, The classification of countable homogeneous directed graphs and countable homogeneous  $n$ -tournaments, *AMS Memoir* **131** (1998), AMS, Rhode Island.
- [22] M. Chudnovsky, P. Seymour, B. Sullivan, Cycles in dense digraphs, *Combinatorica* **28** (2008), 1–18.
- [23] F.R.K. Chung, Regularity Lemmas for hypergraphs and quasi-randomness, *Random Structures and Algorithms* **2** (1991), 241–252.
- [24] D.E. Daykin, R. Häggkvist, Degrees giving independent edges in a hypergraph, *Bull. Austral. Math. Soc.* **23** (1981), 103–109.
- [25] R. Diestel, Graph Theory, *Graduate Texts in Mathematics* **173** 4th ed. (2010), Springer-Verlag, Heidelberg.
- [26] G.A. Dirac, Some theorems on abstract graphs, *Proc. London. Math. Soc.* **2** (1952), 69–81.
- [27] M. El-Zahar, On circuits in graphs, *Discrete Math.* **50** (1984), 227–230.

- [28] P. Erdős, A problem on independent  $r$ -tuples, *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.* **8** (1965), 93–95.
- [29] P. Erdős, On some new inequalities concerning extremal properties of graphs, in P. Erdős and G. Katona, eds., *Theory of Graphs* (Academic Press, New York, 1968), 77–81.
- [30] P. Erdős, Some recent results on extremal problems in graph theory, in P. Rosenstiehl, ed., *Theory of Graphs* (Gordon and Breach, New York; and Dunod, Paris, 1976), 117–123.
- [31] P. Erdős, T. Gallai, On maximal paths and circuits of graphs, *Acta Math. Acad. Sci. Hung.* **10** (1959), 337–356.
- [32] P. Erdős, D.J. Kleitman, B.L. Rothschild, Asymptotic enumeration of  $K_n$ -free graphs, in *Colloquio Internazionale sulle Teorie Combinatorie* (Rome, 1973), Vol. II, 19–27. *Atti dei Convegni Lincei* **17**, Accad. Naz. Lincei, Rome, 1976.
- [33] P. Erdős, M. Simonovits, Supersaturated graphs and hypergraphs, *Combinatorica* **3** (1983), 181–192.
- [34] P. Frankl, Improved bounds for Erdős’ Matching Conjecture, *J. Combin. Theory A* **120** (2013), 1068–1072.
- [35] P. Frankl, On the maximum number of edges in a hypergraph with given matching number, *arXiv:1205.6847v1*.
- [36] R. Häggkvist, C. Thomassen, On pancyclic digraphs, *J. Combin. Theory Series B* **20** (1976), 20–40.
- [37] P. Hajnal, Partition of graphs with condition on the connectivity and minimum degree, *Combinatorica* **3** (1983), 95–99.
- [38] P. Hall, On representatives of subsets, *J. London Math. Soc.* **10** (1935), 26–30.
- [39] H. Hàn, Y. Person, M. Schacht, On perfect matchings in uniform hypergraphs with large minimum vertex degree, *SIAM J. Discrete Math.* **23** (2009), 732–748.
- [40] R.J. Kang, C. McDiarmid, B. Reed, A. Scott, For most graphs  $H$ , most  $H$ -free graphs have a linear homogeneous set, *Random Structures and Algorithms* **45** (2014) 343–361.
- [41] P. Keevash, F. Knox and R. Mycroft, Polynomial-time perfect matchings in dense hypergraphs, *Advances in Mathematics* **269** (2014) 265–334.
- [42] P. Keevash, B. Sudakov, Triangle packings and 1-factors in oriented graphs, *J. Combin. Theory B* **99** (2009), 709–727.

- [43] I. Khan, Perfect matchings in 3-uniform hypergraphs with large vertex degree, *SIAM J. Discrete Math.* **27** (2013), 1021–1039.
- [44] I. Khan, Perfect matchings in 4-uniform hypergraphs, *J. Combin. Theory B* **116** (2016), 333–366.
- [45] J. Kim, D. Kühn, D. Osthus, Bipartitions of highly connected tournaments, *arXiv:1411.1533*.
- [46] Ph.G. Kolaitis, H.J. Prömel, B.L. Rothschild,  $K_{\ell+1}$ -free graphs: asymptotic structure and a 0 – 1 law, *Trans. Amer. Math. Soc.* **303** (1987), 637–671.
- [47] J. Komlós, M. Simonovits, Szemerédi’s regularity lemma and its applications in graph theory, Paul Erdős is eighty, *Proc. Coll. Bolyai Math. Soc.* (Keszthely, 1993), 295–352.
- [48] D. Kühn, J. Lapinskas, D. Osthus, V. Patel, Proof of a conjecture of Thomassen on Hamilton cycles in highly connected tournaments, *Proc. London Math Soc.* **109** (2014) 733–762.
- [49] D. Kühn, D. Osthus, Matchings in hypergraphs of large minimum degree, *J. Graph Theory* **51** (2006), 269–280.
- [50] D. Kühn, D. Osthus, Embedding large subgraphs into dense graphs, in: S. Huczynka, J. Mitchell, C. Roney-Dougal (Eds.), *Surveys in Combinatorics, London Math. Soc. Lecture Note Ser.* **365**, Cambridge University Press, 2009, 137–167.
- [51] D. Kühn, D. Osthus, T. Townsend, Fractional and integer matchings in uniform hypergraphs, *European J. Combin* **38** (2014), 83–96.
- [52] D. Kühn, D. Osthus, T. Townsend, Proof of a tournament partition conjecture and an application to 1-factors with prescribed cycle lengths, *Combinatorica*, Online first publication, doi: 10.1007/s00493-015-3186-8, (2015), 1–19.
- [53] D. Kühn, D. Osthus, A. Treglown, Matchings in 3-uniform hypergraphs, *J. Combin. Theory B* **103** (2013), 291–305.
- [54] E. Lamken, B.L. Rothschild, The number of odd-cycle-free graphs, in *Finite and Infinite Sets (A. Hajnal, L. Lovász, and V.T. Sós, eds.)* (1984), 547–553. *Coll. Math. Soc. J. Bolyai* **37**.
- [55] K. Markström, A. Ruciński, Perfect matchings and Hamiltonian cycles in hypergraphs with large degrees, *European J. Combin.* **32** (2011), 677–687.
- [56] R. Morris, D. Saxton, The number of  $C_{2\ell}$ -free graphs, *arXiv:1309.2927v2*.
- [57] D. Osthus, H.J. Prömel, A. Taraz, For which densities are random triangle-free graphs almost surely bipartite?, *Combinatorica* **23** (2003), 105–150.

- [58] O. Pikhurko, Perfect matchings and  $K_4^3$ -tilings in hypergraphs of large co-degree, *Graphs Combinatorics* **24** (2008), 391–404.
- [59] A. Pokrovskiy, Highly linked tournaments, *J. Combin. Theory B* **115** (2015), 339–347.
- [60] H.J. Prömel, A. Steger, Almost all Berge graphs are perfect, *Combin. Prob. Comput.* **1** (1992), 53–79.
- [61] H.J. Prömel, A. Steger, Excluding induced subgraphs: quadrilaterals, *Random Structures and Algorithms* **2** (1991), 55–71.
- [62] H.J. Prömel, A. Steger, Excluding induced subgraphs II: extremal graphs, *Discrete Appl. Math.* **44** (1993), 283–294.
- [63] H.J. Prömel, A. Steger, Excluding induced subgraphs III: a general asymptotic, *Random Structures and Algorithms* **3** (1992), 19–31.
- [64] K.B. Reid, Two complementary circuits in two-connected tournaments, *In: Cycles in Graphs, B.R. Alspach, C.D. Godsil, Eds., Ann. Discrete Math.* **27** (1985), 321–334.
- [65] K.B. Reid, Three problems on tournaments, *Graph Theory and its applications: East and West, Ann. New York Acad. Sci.* **576** (1989), 466–473.
- [66] R.W. Robinson, Enumeration of acyclic digraphs, in *R. C. Bose et al. (Eds.), Proc. Second Chapel Hill Conf. on Combinatorial Mathematics and its Applications* (Univ. North Carolina, Chapel Hill, N.C., 1970), 391–399.
- [67] R.W. Robinson, Counting labeled acyclic digraphs, in *F. Harary (Ed.), New Directions in the Theory of Graphs* (Academic Press, NY, 1973), 239–273.
- [68] V. Rödl, A. Ruciński, E. Szemerédi, Perfect matchings in uniform hypergraphs with large minimum degree, *European J. Combin.* **27** (2006), 1333–1349.
- [69] V. Rödl, A. Ruciński, E. Szemerédi, Perfect matchings in large uniform hypergraphs with large minimum collective degree, *J. Combin. Theory A* **116** (2009), 616–636.
- [70] D. Saxton, A. Thomason, Hypergraph containers, *Inventiones mathematicae*, to appear.
- [71] D. Saxton, A. Thomason, Simple containers for simple hypergraphs, *arXiv:1402.5400*.
- [72] M. Simonovits, A method for solving extremal problems in graph theory, in P. Erdős and G. Katona, eds., *Theory of Graphs* (Academic Press, New York, 1968), 279–310.
- [73] Z.M. Song, Complementary cycles of all lengths in tournaments, *J. Combin. Theory B* **57** (1993), 18–25.

- [74] R.P. Stanley, Acyclic orientations of graphs, *Discrete Math.* **5** (1973), 171–178.
- [75] M. Stiebitz, Decomposing graphs under degree constraints, *J. Graph Theory* **23** (1996), 321–324.
- [76] E. Szemerédi, Regular partitions of graphs, *Colloq. Internat. CNRS* **260** (1978), 399–401.
- [77] C. Thomassen, Graph decomposition with constraints on the connectivity and minimum degree, *J. Graph Theory* **7** (1983), 165–167.
- [78] A. Treglown, Y. Zhao, Exact minimum degree thresholds for perfect matchings in uniform hypergraphs, *J. Combin. Theory A* **119** (2012), 1500–1522.
- [79] A. Treglown, Y. Zhao, Exact minimum degree thresholds for perfect matchings in uniform hypergraphs II, *J. Combin. Theory A* **120** (2013), 1463–1482.
- [80] A. Treglown, Y. Zhao, A note on perfect matchings in uniform hypergraphs, *arXiv:1503.03357*.
- [81] W.T. Tutte, The factorization of linear graphs, *J. London Math. Soc.* **22** (1947), 107–111.