THE REGULARITY METHOD IN DIRECTED GRAPHS AND HYPERGRAPHS

by

RICHARD MYCROFT

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Mathematics
The University of Birmingham
May 2010

ABSTRACT

In recent years the regularity method has been used to tackle many embedding problems in extremal graph theory. This thesis demonstrates and develops three different techniques which can be used in conjunction with the regularity method to solve such problems. These methods enable us to prove an approximate version of the well-known Sumner's universal tournament conjecture, first posed in 1971, which states that any tournament G on $2n-2$ vertices contains a copy of any directed tree T on n vertices. An analysis of the extremal cases then proves that Sumner's universal tournament conjecture holds for any sufficiently large n.

Our methods are also applied to the problem of obtaining hypergraph analogues of Dirac's theorem. Indeed, we show that for any $k \geqslant 3$ and any $1 \leqslant \ell \leqslant k-1$ with $k-\ell \nmid k$, any k-uniform hypergraph on n vertices with minimum degree at least $\frac{n}{\lceil k/(k-\ell) \rceil (k-\ell)} + o(n)$ contains a Hamilton ℓ-cycle. This result confirms a conjecture of Hàn and Schacht, and is best possible up to the $o(n)$ error term. Together with results of Rödl, Ruciński and Szemerédi, this result asymptotically determines the minimum degree which forces an ℓ-cycle for any ℓ with $1 \leqslant \ell \leqslant k-1$.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Daniela Kühn and Deryk Osthus for their support and guidance in the writing of this thesis and throughout the course of my studies. In particular I acknowledge their coauthorship of the main results given in this thesis. Peter Keevash was also a coauthor of the work presented in Chapter 4, and Conjecture 1.7 was proposed in conjunction with Peter Allen and Olly Cooley. The work presented in Chapters 2, 3, 4 and 5 roughly corresponds to that presented in [29], [30], [24] and [28] respectively.

I would also like to thank my wonderful fiancée Zoe for her constant support and devotion.

CONTENTS

CHAPTER 1

INTRODUCTION

The problem of finding a fixed substructure within a given structure is a major area of extremal combinatorics. Early examples of such results include Dirac's Theorem [10], which states that any graph G on $n \geqslant 3$ vertices with minimum degree at least $n/2$ contains a Hamilton cycle, and Tutte's Theorem, which completely characterises those graphs G which contain a perfect matching (a set of disjoint edges of G containing every vertex of G). More recently, much progress has been made on similar problems through the use of *regularity*.

In this thesis we demonstrate and develop three different techniques by which the regularity method may be applied to solve embedding problems in directed graphs and hypergraphs. Namely, we demonstrate and develop the uses of the 'hypergraph blow-up lemma', an 'absorbing' structure, and a randomised embedding algorithm. Each method is used to obtain significant new results. Indeed, our application of a randomised embedding algorithm to embed directed graphs culminates in the proof for large n of the well-known Sumner's universal tournament conjecture, first posed in 1971, which states that any tournament on $2n-2$ vertices contains any directed tree on n vertices. Next we use the 'blow-up lemma' and an 'absorbing' structure to obtain analogues of Dirac's theorem for hypergraphs, giving asymptotically for any k and ℓ the minimum degree threshold required to guarantee that a k-uniform hypergraph contains a Hamilton ℓ-cycle (for any $1 \leqslant \ell \leqslant k-1$, an ℓ-cycle in a k-graph is a natural generalisation of a cycle in a graph).

The reader should note that whilst we apply the randomised embedding algorithm to directed graphs, and the 'blow-up lemma' and the 'absorbing' method to hypergraphs, these methods are not specific to these cases. Indeed, each method can be used to solve embedding problems in simple graphs, directed graphs or hypergraphs. Also, to keep this introduction relatively short and easy-going, we defer the definitions of many terms used to the relevant later chapters of this thesis.

1.1 Trees in tournaments

A *directed graph* G, or *digraph*, consists of a set of vertices $V(G)$ and a set of edges $E(G)$, where each edge of G is an ordered pair of vertices of G. So we think of an edge (u, v) as being *directed* from u to v. A *tournament* is a digraph formed by orienting each edge of a complete (undirected) graph. A tournament need not contain any directed cycles, so we cannot guarantee the presence of any digraph containing such a cycle in an arbitrary tournament, no matter how large. Instead, we might ask how large a tournament G must be to ensure that G contains a copy of a directed tree T on n vertices (where a *directed tree* is formed by orienting each edge of an undirected tree). The well-known Sumner's universal tournament conjecture, first posed in 1971 (see *e.g.* [38, 46]), states that any tournament on $2n-2$ vertices contains a copy of any directed tree T on n vertices. We prove that this conjecture holds for all large n.

Theorem 1.1 *There exists n_0 such that the following holds. Let T be a directed tree on $n \geqslant n_0$ vertices, and G a tournament on $2n-2$ vertices. Then G contains a copy of T.*

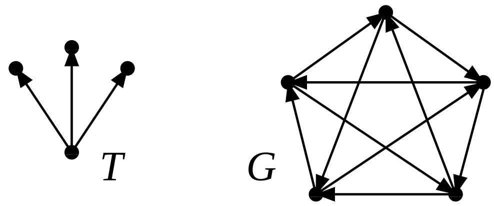

Figure 1.1: A tree T on four vertices and a tournament G on five vertices such that G does not contain T. Sumner's universal tournament conjecture for $n = 4$ states that any tournament on six vertices contains any tree on four vertices.

This bound is best possible for any n. To see this, let T be a star with all edges directed outwards, and let G be a regular tournament on $2n-3$ vertices. Then every vertex of G has $n-2$ inneighbours and $n-2$ outneighbours, and so G does not contain a copy of T, whose central vertex has $n-1$ outneighbours (see Figure 1.1). There are also 'near-extremal' examples whose structure is substantially different: let T be obtained from a directed path on $\ell \geqslant 1$ vertices by adding y outneighbours to the terminal vertex of the path and z inneighbours to the initial vertex of the path, where $y + z = n - \ell$. Let G consist of regular tournaments on sets Y and Z of size $2y-1$ and $2z-1$ respectively, together with an arbitrary tournament on a set X of $\ell - 1$ vertices so that all edges are oriented from Z to X, from X to Y and from Z to Y. Then $|T| = n$ and $|G| = 2n - \ell - 3$, but it is easy to see that G does not contain T. These examples play a significant role in the proof of Theorem 1.1.

Many previous results have been obtained towards Sumner's conjecture. Let $f(n)$ denote the smallest integer such that any tournament on $f(n)$ vertices contains any directed tree on n vertices. So Sumner's conjecture states that $f(n) = 2n - 2$. Chung (see [46]) observed that $f(n) \leqslant n^{1+o(1)}$, and Wormald [46] improved this to $f(n) \leqslant O(n \log n)$. The first linear bound on $f(n)$ was established by Häggkvist and Thomason [14], who showed that $f(n) \leqslant 12n$, and also that $f(n) \leqslant (4 + o(1))n$. Havet [17] then showed that $f(n) \leqslant 38n/5$, and later Havet and Thomassé [19] used their notion of median orders to improve this to $f(n) \leqslant 7n/2$. Finally El Sahili used the same notion to prove the best known bound for general n, namely that $f(n) = 3n - 3$. We make extensive use of this result (actually, any linear bound on $f(n)$ would suffice for our purposes; the factor of 3 is not essential).

Theorem 1.2 (El Sahili [11]) *Let T be a directed tree on n vertices, and let G be a tournament on $3n-3$ vertices. Then G contains a copy of T.*

Sumner's conjecture is also known to hold for some special classes of directed trees. In particular, Havet and Thomassé [19] proved it for 'outbranchings', again using median orders. Here an *outbranching* is a directed tree T in which we may choose a root vertex $t \in T$ so that for any vertex $t' \in T$, the path between t and t' in T is directed from t to t'. (Outbranchings are also known as arborescences.)

Theorem 1.3 (Havet and Thomassé [19]) *Let T be an outbranching on n vertices, and let G be a tournament on $2n-2$ vertices. Then G contains a copy of T.*

For many types of directed trees, Sumner's conjecture holds with room to spare. A classical result of this type is Redei's theorem.

Theorem 1.4 (Redei [37]) *Any tournament contains a spanning directed path.*

Havet and Thomassé [20] proved a much stronger result, namely that every tournament on $n \geqslant 8$ vertices contains every orientation of the path on n vertices (this was a conjecture of Rosenfeld). Reid and Wormald [38] also proved Sumner's conjecture for other (very restricted) classes of directed trees.

In Chapter 2, we use the regularity method, combined with a randomised embedding algorithm, to prove an approximate version of Sumner's universal tournament conjecture. This is (1) of the following theorem. In the process we also prove (2), a stronger result for directed trees of bounded maximum degree, which is of independent interest.

Theorem 1.5 *Let $\alpha > 0$. Then the following properties hold.*

(1) There exists n_0 such that for any $n \geqslant n_0$, any tournament G on at least $2(1+\alpha)n$ vertices contains any directed tree T on n vertices.

(2) Let Δ be any positive integer. Then there exists n_0 such that for any $n \geqslant n_0$, any tournament G on at least $(1+\alpha)n$ vertices contains any directed tree T on n vertices with $\Delta(T) \leqslant \Delta$.

In Chapter 3 we consider the extremal cases of the previous theorem. Indeed, we use both methods and results of the previous chapter to show that Sumner's universal tournament conjecture holds for large n whenever the directed tree T is not 'close' to one of the extremal examples to the conjecture given earlier. By a careful consideration of these extremal examples, we can show that the conjecture holds in these cases too, completing the proof of Theorem 1.1. In particular we use both parts of Theorem 1.5 extensively, and so Theorem 1.5(1) and its proof are not made redundant by the stronger Theorem 1.1.

There are several directions in which our results can be extended. For example, the error term in Theorem 1.5(2) cannot be completely omitted (to see this, consider the 'near-extremal' example discussed after the statement of Theorem 1.1 with $x \geqslant 2$ and $y = z = \Delta - 1$). It would be interesting to know whether the term αn can be reduced to a constant depending only on Δ.

Another class of directed trees where Sumner's conjecture can be strengthened is that of trees with few leaves. The first result in this direction was proved by Häggkvist and Thomason [14]. Havet and Thomassé (see [18]) then proposed the following generalisation of Sumner's conjecture.

Conjecture 1.6 (Havet and Thomassé [18]) *Let T be a directed tree on n vertices with k leaves. Then every tournament on $n+k-1$ vertices contains a copy of T.*

Céroi and Havet [7] proved that this conjecture holds for $k \leqslant 3$, from which they deduced that Sumner's conjecture holds for all directed trees with at most four leaves.

However, Conjecture 1.6 is still far from optimal for many directed trees T, such as a binary tree. Preliminary investigations have suggested the following stronger conjecture.

Conjecture 1.7 *Let T be a directed tree on n vertices which is not an instar or an outstar, and let $\Delta_p(T) := \max\{d^-(x) + d^+(y)\}$, where the maximum is over all $x, y \in T$ such that there is a directed path in T from x to y. Then any tournament G on $n + \Delta_p(T) - 2$ vertices contains a copy of T.*

The 'near-extremal' examples discussed after the statement of Theorem 1.1 show that this would be best possible, in the sense that the -2 cannot be replaced by -3. It would be interesting to know if the methods of this thesis can be extended to prove these conjectures (for large n). Finally, the following far-reaching generalisation of Sumner's conjecture was made by Burr [6].

Conjecture 1.8 (Burr [6]) *Any $(2n - 2)$-chromatic digraph contains any directed tree on n vertices.*

Burr verified the corresponding statement with $(n - 1)^2$ in place of $2n - 2$. More recently, El Sahili [12] proved that any n-chromatic digraph contains the path formed by $n - 1$ forward edges followed by one backwards edge, but to our knowledge, no further progress has been made.

1.2 Hamilton cycles in hypergraphs

Recall that Dirac's theorem [10] states that any graph G on $n \geqslant 3$ vertices with minimum degree at least $n/2$ contains a Hamilton cycle (*i.e.* there is a cyclic ordering of all n vertices of

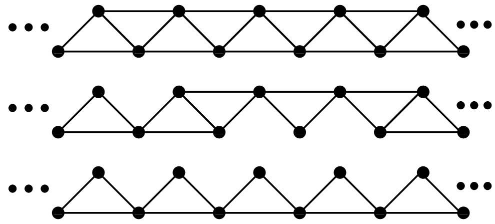

Figure 1.2: Segments of a 2-cycle (top), a cycle (middle) and a 1-cycle (bottom) in a 3-graph; the triangles represent edges.

G such that every pair of consecutive vertices forms an edge of G). Chapters 4 and 5 develop the regularity method to obtain analogues of this result for hypergraphs. For this, we first make the following standard definitions. A *k-uniform hypergraph*, or *k-graph* $\mathcal{H}$, consists of a set of *vertices* $V(\mathcal{H})$ and a set of *edges* $E(\mathcal{H})$, where every edge $e \in E(\mathcal{H})$ consists of precisely k vertices. Let $\mathcal{H}$ be a k-graph, and let A be a set of $k-1$ vertices of $\mathcal{H}$. Then the *degree* of A, denoted $d_{\mathcal{H}}(A)$, is the number of edges of $\mathcal{H}$ which contain A as a subset. The *minimum degree* $\delta(\mathcal{H})$ of $\mathcal{H}$ is then the minimum value of $d_{\mathcal{H}}(A)$ taken over all sets A of $k-1$ vertices of $\mathcal{H}$.

We say that a k-graph $\mathcal{C}$ on n vertices is a *cycle* if its vertices can be given a cyclic ordering $v_1, \dots, v_n$ so that every pair of consecutive vertices v_i, v_{i+1} lies in an edge of $\mathcal{C}$ and every edge of $\mathcal{C}$ consists of k consecutive vertices. $\mathcal{C}$ is an *ℓ-cycle* if every pair of consecutive edges intersect in precisely ℓ vertices. Figure 1.2 illustrates these definitions in the case $k = 3$. Note that if $\mathcal{C}$ is an ℓ-cycle of order n, then $(k-\ell)|n$, since every edge of $\mathcal{C}$ contains exactly $k-\ell$ vertices which were not contained in the previous edge. We say that a cycle is *loose* if every pair of consecutive edges intersects in a single vertex, with the possible exception of one pair of edges, which may intersect in more than one vertex. So if $(k-1)|n$ then a loose cycle of order n is a 1-cycle, but this final condition allows us to consider loose cycles whose order is not a multiple of $k-1$. A *Hamilton cycle* in a k-graph $\mathcal{H}$ is a sub-k-graph of $\mathcal{H}$ which is a cycle containing every vertex of $\mathcal{H}$.

In [39] and [40], Rödl, Ruciński and Szemerédi proved the following theorem for $\ell = k-1$;

the other cases follow, since if $(k-\ell)|n$ then any $(k-1)$-cycle of order n contains an ℓ-cycle on the same vertices.

Theorem 1.9 (Rödl, Ruciński and Szemerédi [39, 40]) *For all $k \geqslant 3$, $1 \leqslant \ell \leqslant k-1$ and any $\eta > 0$ there exists n_0 so that if $n \geqslant n_0$ and $(k-\ell)|n$ then any k-graph $\mathcal{H}$ on n vertices with $\delta(\mathcal{H}) \geqslant \left(\frac{1}{2}+\eta\right) n$ contains a Hamilton ℓ-cycle.*

This proved a conjecture of Katona and Kierstead [22]. A simple well-known construction (see Proposition 4.1) shows that Theorem 1.9 is best possible up to the error term ηn if $(k-\ell)|k$.

It is therefore natural to ask what minimum degree condition on $\mathcal{H}$ guarantees a Hamilton ℓ-cycle for the remaining values of k and ℓ. Kühn and Osthus [31] showed that any 3-graph $\mathcal{H}$ on n vertices with $\delta(\mathcal{H}) \geqslant (\frac{1}{4}+o(1))n$ contains a loose Hamilton cycle. In Chapter 4, we use the regularity method, combined with the recent 'hypergraph blow-up lemma' due to Keevash [23] to extend this result to any $k \geqslant 3$, proving Theorem 1.10. If $(k-\ell)|n$ then Theorem 1.10 is a special case of Theorem 1.11, which we prove later by other means. However, the proof of Theorem 1.10 has independent interest; we believe that this is the first application of the hypergraph blow-up lemma to find a spanning structure within a general hypergraph, a method with many further potential applications.

Theorem 1.10 *For all $k \geqslant 3$ and any $\eta > 0$ there exists n_0 so that if $n \geqslant n_0$ then any k-graph $\mathcal{H}$ on n vertices with $\delta(\mathcal{H}) \geqslant (\frac{1}{2(k-1)}+\eta)n$ contains a loose Hamilton cycle.*

Another simple well-known example (see Proposition 4.2) shows that this result is best possible up to the error term ηn. In fact, Proposition 4.2 actually tells us more than this, namely that up to the error term, this minimum degree condition is best possible to ensure the existence of any (not necessarily loose) Hamilton cycle in $\mathcal{H}$. So by combining Theorem 1.10 and Proposition 4.2 we obtain a hypergraph analogue of Dirac's theorem, namely that the minimum degree needed to find a Hamilton cycle in a k-graph of order n is asymptotically $\frac{n}{2(k-1)} + o(n)$.

Theorem 1.10 was also proved simultaneously and independently by Hàn and Schacht [16], who used an ‘absorbing’ method. In fact, they proved the more general result that if $1 \leqslant \ell < k/2$, then any k-graph $\mathcal{H}$ on n vertices with $(k-\ell)|n$ and $\delta(\mathcal{H}) \geqslant (\frac{1}{2(k-\ell)} + o(1))n$ contains a Hamilton ℓ-cycle. They raised the question of determining the correct minimum degree for those values of k and ℓ not covered by their result or by Theorem 1.9. In Chapter 5, we use the regularity method, combined with a development of the ‘absorbing method’ of Hàn and Schacht, to answer this question with the following theorem.

Theorem 1.11 *For all $k \geqslant 3$, $1 \leqslant \ell \leqslant k-1$ such that $(k-\ell) \nmid k$ and any $\eta > 0$ there exists n_0 so that if $n \geqslant n_0$ and $(k-\ell)|n$ then any k-graph $\mathcal{H}$ on n vertices with $\delta(\mathcal{H}) \geqslant \left(\frac{1}{\lceil \frac{k}{k-\ell} \rceil (k-\ell)} + \eta\right) n$ contains a Hamilton ℓ-cycle.*

Proposition 4.2 also shows that this result is best possible up to the error term ηn. Thus Theorem 1.9 and Theorem 1.11 together give asymptotically, for any k and ℓ, the minimum degree required to guarantee that a k-graph on n vertices contains a Hamilton ℓ-cycle.

The difference in the minimum degree threshold between the cases $(k-\ell) \mid k$ and $(k-\ell) \nmid k$ is perhaps surprising. For example, if $k = 9$ then the minimum degree threshold for an 8-cycle or a 6-cycle is asymptotically $n/2$, whereas for a 7-cycle it is instead $n/10$. This difference is essentially a consequence of the fact that in the $(k-\ell) \mid k$ case every Hamilton ℓ-cycle contains a perfect matching. The minimum degree threshold for the latter is known to be close to $n/2$ (see Proposition 4.1).

One natural extension of the work presented here would be to consider other notions of degree. Let $\mathcal{H}$ be a k-graph on n vertices; then the degree of a set A of vertices of $\mathcal{H}$, denoted $d(A)$ is defined to be the number of edges containing A. For $1 \leqslant r \leqslant k-1$, let $\delta_r(\mathcal{H})$ denote the minimum r-degree of $\mathcal{H}$, *i.e.* the minimum of $d(A)$ taken over all sets A of r vertices of $\mathcal{H}$. Then the notion of degree defined earlier is the $(k-1)$-degree. Until recently, very little was known regarding the minimum r-degree thresholds which force spanning subgraphs for $r < k-1$. However, recently Pikhurko [36] showed that for $k/2 \leqslant r \leqslant k-1$ the minimum r-degree threshold which forces a perfect matching in a k-graph $\mathcal{H}$ on n vertices is asymptotically $n/2$. Particular interest is devoted to the 1-degree, also known as the vertex degree. Hàn, Person and Schacht [15] recently showed that the 1-degree threshold needed to guarantee a perfect matching in a 3-graph is asymptotically $\frac{5}{9}\binom{n}{2}$. In light of the close connections between the problem of finding a perfect matching in a k-graph and the problem of finding a Hamilton cycle in a k-graph, it would be interesting to know whether the results and methods of Pikhurko and of Hàn, Person and Schacht can be used to prove similar results for Hamilton cycles in these cases.

Other (less restrictive) notions of hypergraph cycles have also been considered. For example, a *Berge-cycle* [4] is defined to consist of a sequence of vertices where each pair of consecutive vertices is contained in a common edge.

1.3 An introduction to the regularity method

In this section we give a brief outline of the use of the regularity method in tackling embedding problems. For simplicity we present the method as it applies to simple graphs, but for directed graphs and hypergraphs the ideas are similar.

Let G be a bipartite graph with vertex classes U and V of size m. Also let d be the density of G. Then we say that G is ε-regular if for any subsets $U' \subseteq U$ and $V' \subseteq V$ of size at least εm we have $d' = d \pm \varepsilon$, where d' is the density of $G(U', V')$. The notion of ε-regularity is very useful for embedding problems, as in many ways G behaves like a random bipartite graph on vertex classes U and V with dm^2 edges chosen uniformly at random, in the sense that many properties of G are shared by this random graph with high probability. In particular, if H is a

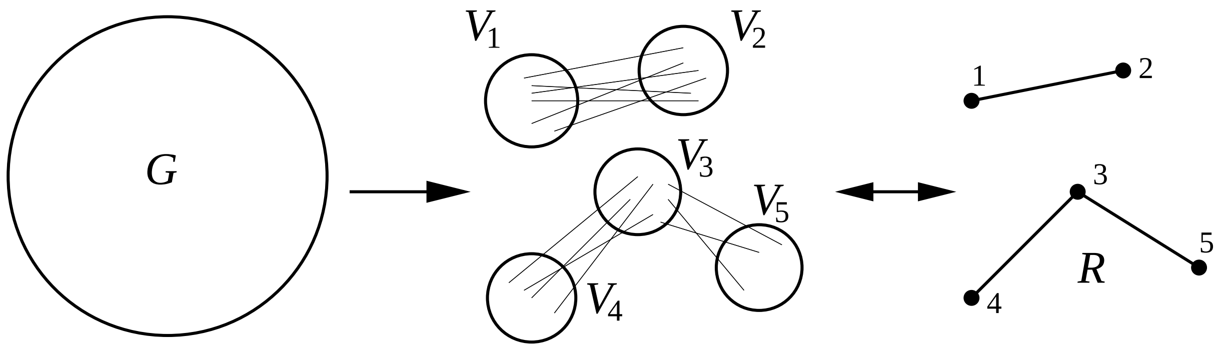

Figure 1.3: An illustration of steps (1) and (2) of the regularity method. Given a graph G, we apply the Szemerédi regularity lemma to partition the vertices of G into clusters V_i, so that the edges between most pairs of clusters are regular. We then define a reduced graph R, whose vertices correspond to the clusters of G, and with an edge in R whenever the bipartite graph induced by the corresponding clusters of G is regular and dense.

small graph, then the *counting lemma* tells us approximately how many copies of H there are in G. This result can be applied to graphs formed of multiple ε-regular subgraphs. For example, if G is a tripartite graph on vertex classes U, V and W of size m such that $G[U,V]$, $G[U,W]$ and $G[V,W]$ are each ε-regular with densities d_1, d_2 and d_3 respectively, then the number of triangles in G is approximately the number we would expect in the corresponding random graph on the same vertex classes, namely $d_1 d_2 d_3 m^3$.

To apply the notion of ε-regularity to an arbitrary graph G, we also need a version of the *Szemerédi regularity lemma* (see *e.g.* [45]). This states that the vertex set of any large graph G can be partitioned into a bounded number of clusters such that for almost all pairs of clusters U, V, the edges between U and V form an ε-regular bipartite subgraph $G(U,V)$ of G. We then typically consider a *reduced graph* R of G, whose vertices correspond to these clusters, and where two clusters U and V form an edge if $G(U,V)$ is ε-regular and has high density. See Figure 1.3 for an illustration of this process.

We can now describe a 'typical' application of the regularity method to find a copy of a subgraph H in G. This proceeds as follows.

1. Apply the Szemerédi regularity lemma to G to partition the vertices of G into clusters.

2. Form the reduced graph R as described above.

3. Find a 'useful' subgraph R' of R, for example a perfect matching or a Hamilton cycle.

4. Use the fact that every edge e of R' corresponds to an ε-regular and dense bipartite subgraph of G to find a copy of H in G.

At steps 3 and 4 there are a wide variety of possible approaches, three of which are illustrated in this thesis. However, in each case the advantage conferred by using the regularity method is that the subgraph R' of the reduced graph R is simpler (and hence easier to find) than the subgraph H which we wish to find in G.

For example, in Chapter 2, we wish to find a directed tree T of bounded degree in a tournament G. To do this, we first apply a version of the Szemerédi regularity lemma for directed graphs to partition G into clusters. We then show that (at least for certain well-behaved tournaments) the reduced directed graph R contains a directed Hamilton cycle R'. Next a randomised embedding algorithm allocates each vertex of T to a vertex of this cycle (which corresponds to a cluster of G). Finally, by analysing this algorithm we show that with high probability it is possible to use the regularity indicated by each edge of R' to embed each vertex of T within the cluster of G to which it was allocated so as to form a copy of T in G. So by the regularity lemma we have reduced the problem of finding a complex subgraph T in G to the problem of finding a simple subgraph R' in R plus the problem of using the existence of R' to embed T in G.

In Chapter 4 we instead wish to find a loose Hamilton cycle within a k-graph. This is also achieved using the regularity method as described above (using a k-graph version of the Szemerédi regularity lemma and defining a reduced k-graph R). Here Step 3 is to find a collection of disjoint copies of a small k-graph $\mathcal{A}_{k,1}$ in R whose union includes almost every vertex of

R, whilst for Step 4 we apply the blow-up lemma to find spanning paths within the clusters corresponding to each copy of $\mathcal{A}_{k,1}$ in R. This method is explained in more detail in Chapter 4. Finally in Chapter 5 we apply the regularity method in the proof of each of three lemmas used in the proof of Theorem 1.11.

1.4 A word on notation

Throughout Chapters 1, 2, 3 and 5 of this thesis we employ notational conventions consistently. For example, we refer to graphs with upper-case letters in italic font (*e.g.* G, H), whereas we use calligraphic characters (such as $\mathcal{G}$ and $\mathcal{H}$) to denote hypergraphs. However, in Chapter 4 the hypergraph blow-up lemma due to Keevash [23] and other results from the same paper play an extensive role. These results are very technical, and so employing different notational conventions in Chapter 4 to those used by Keevash [23] would cause the reader substantial difficulties. Instead we have opted for the lesser of two evils, namely to use the notational conventions used by Keevash [23] in Chapter 4. All differences in notation are appropriately introduced in Chapter 4.

The following notation is used consistently throughout this thesis. The set of integers from 1 to r is denoted by $[r]$. For a set A, we use $\binom{A}{k}$ to denote the collection of subsets of A of size k, and similarly $\binom{A}{\leqslant k}$ to denote the collection of non-empty subsets of A of size at most k. We write $x = y \pm z$ to denote that $y - z \leqslant x \leqslant y + z$, and use $o(1)$ to denote a function which tends to zero as n tends to infinity, holding all other variables involved constant. Also, we write *e.g.* $0 < a_1 \ll a_2 \ll a_3 \ll a_4 \leqslant 1$ to mean that we can choose the constants $a_1, \dots, a_4$ from right to left. More precisely, there are increasing functions f_1, f_2, f_3 such that, given a_4, whenever we choose some $a_3 \leqslant f_3(a_4)$, $a_2 \leqslant f_2(a_3)$ and $a_1 \leqslant f_1(a_2)$, all calculations needed in the following argument are valid. Hierarchies with more constants are defined similarly. Also, we sometimes write 'let $x \ll y$' when y has an already fixed positive value; by this we mean that there exists some $x_0 > 0$ such that for any $0 < x < x_0$ the subsequent statements hold. We omit floors and ceilings throughout this thesis whenever they do not affect the argument.

CHAPTER 2

EMBEDDING TREES IN TOURNAMENTS

In this chapter we use the regularity method combined with a randomised embedding algorithm to prove an approximate version of Sumner's universal tournament conjecture, as well as a stronger result for directed trees of bounded degree. This is Theorem 1.5, which is restated below.

Theorem 1.5 *Let $\alpha > 0$. Then the following properties hold.*

(1) There exists n_0 such that for any $n \geqslant n_0$, any tournament G on at least $2(1+\alpha)n$ vertices contains any directed tree T on n vertices.

(2) Let Δ be any positive integer. Then there exists n_0 such that for any $n \geqslant n_0$, any tournament G on at least $(1 + \alpha)n$ vertices contains any directed tree T on n vertices with $\Delta(T) \leqslant \Delta$.

2.1 Outline of the proof of Theorem 1.5

The notion of a robust outexpander (which was introduced for dense graphs in [33]) is crucial to the proof of Theorem 1.5. Informally, a digraph G is a robust outexpander if for any set $S \subseteq V(G)$ which is not too large or too small, the number of vertices with many inneighbours in S is substantially bigger than $|S|$. Kühn, Osthus and Treglown [33] showed that any robust outexpander G of linear minimum semidegree contains a Hamilton cycle. (Here the minimum semidegree is the minimum of the minimum indegree and the minimum outdegree.) By applying this result to the 'reduced digraph' obtained from the Szemerédi regularity lemma, we may split most of the vertices of G into sets $V_1, V_2, \dots, V_k$ so that the set of edges from V_i to V_{i+1} for each i (addition of the indices taken modulo k) forms a regular and dense bipartite graph. This structure is very useful for embedding directed trees. On the other hand, it is easy to show that if a tournament G is not a robust outexpander of linear minimum semidegree, then the vertices of G can be split into two parts so that almost all of the edges between the two parts are directed the same way (see Lemma 2.11). By iterating this split, we either obtain a part which is a robust outexpander of linear minimum semidegree, or find that G is close to being a transitive tournament, a case which we consider separately.

To begin, in Section 2.2 we define the concepts we use, and prove various useful lemmas. Then in Sections 2.3 and 2.4 we use the regularity method to show that Theorem 1.11 holds with the added condition that G is a robust outexpander of linear minimum semidegree. Indeed, in Section 2.3, we consider the case where the tournament G is a robust outexpander of linear minimum semidegree on $(1 + \alpha)n$ vertices, and T is a directed tree on n vertices of bounded maximum degree. Here we apply the Szemerédi regularity lemma to partition the vertices of G into clusters $V_1, V_2, \dots, V_k$, and form the 'reduced digraph' R as outlined in the introduction. As described above, we apply the result of Kühn, Osthus and Treglown [33] to find a Hamilton cycle R' in the reduced digraph R. This implies that we can relabel the clusters of G so that the set of edges from V_i to V_{i+1} is regular and dense for each i; it is this structure on G which we use to embed T in G. One attempt to do so would be to embed each vertex $t \in T$ in the cluster either preceding or succeeding the cluster containing the parent t' of t, according to the direction of the edge between t and t'. However, for many directed trees (such as an anti-directed path) this method fails to give an approximately uniform allocation of vertices of T to the clusters of G, which we require for the embedding to be successful. Instead, we modify this method so that each vertex is embedded as above with probability 1/2 and is embedded in the same cluster as its parent with probability 1/2. We show that with high probability this randomised algorithm indeed gives an approximately uniform allocation of vertices of T to the clusters of G, and so (using the regularity of edges directed from V_i to V_{i+1}) successfully embeds T in G.

In Section 2.4 we begin by strengthening the result from Section 2.3, showing that if T is a directed tree on n vertices of bounded maximum degree, and G is a tournament on $(1 + \alpha)n$ vertices whose reduced graph defined on the clusters $V_1, \dots, V_k$ contains a Hamilton cycle, then we can embed T in G so that the vertices of a chosen small set $H \subseteq V(T)$ are embedded within a specified set $U \subseteq V(G)$. To do this, we embed all vertices 'far' from H by the method described above, which ensures that the vertices of T are allocated approximately uniformly amongst the clusters of G. The remaining vertices of T are instead embedded to ensure that every vertex of H is embedded within U. This result allows us to embed directed trees T of unbounded maximum degree. Indeed, we define for any tree T a 'core tree' T_Δ, which has the properties that T_Δ has bounded maximum degree, but each component of $T - T_\Delta$ is small. Given a tournament G which is a robust outexpander of linear minimum semidegree on $(2+\alpha)n$ vertices, we again split most of the vertices of G into clusters $V_1, V_2, \dots, V_k$ as described above. We then choose subsets $V_i' \subseteq V_i$ at random so that $|\bigcup_i V_i'|$ is roughly equal to $|T_\Delta|$, and embed T_Δ into these subsets (actually we first extend T_Δ to an 'extended tree' T_{ext} and embed T_{ext} into these subsets), using the strengthened result for directed trees of bounded degree to restrict certain vertices of T_Δ to vertices of G with many inneighbours and outneighbours in $G - \bigcup_i V_i'$.

Since each component of $T - T_\Delta$ is small, this allows us to embed the components of $T - T_\Delta$ one by one among the unoccupied vertices of G to complete the embedding of T in G.

It is a simple exercise to demonstrate that any transitive tournament on n vertices contains any directed tree on n vertices. In Section 2.5, we prove an analogue of this for almost-transitive tournaments G. This means that the vertices of G can be ordered so that almost all of the edges of G are directed towards the endvertex which is greater in this order. We show that if G is an almost-transitive tournament on $(1+\alpha)n$ vertices and T is a directed tree on n vertices then G contains T.

Finally, in Section 2.6, we use the robust outexpander dichotomy to prove Theorem 1.5. Here we describe the proof of the first statement; the proof of the second is very similar. So let G be a tournament on $2(1+\alpha)n$ vertices and let T be a directed tree on n vertices. If G is a robust outexpander of linear minimum semidegree, then our results of Sections 2.3 and 2.4 show that G contains T, as desired. On the other hand, if G is not a robust outexpander of linear minimum semidegree then we may split G into two parts as described above. By iterating this split, we must either obtain a part which is a robust outexpander of linear minimum semidegree or find that G is almost-transitive. In the latter case we may embed T in G by the results of Section 2.5. So we may suppose that the splitting process yields a part which is a robust outexpander of linear minimum semidegree. Then we divide T into parts to be embedded amongst the parts of G, so that each part of G receives a part of T approximately proportional to its size. However, the robust outexpander part of G actually receives slightly more vertices of T than it would from a proportional split. The results from Sections 2.3 and 2.4 guarantee that this part of T can still be embedded into the corresponding part of G. Since the other parts of G then receive slightly fewer vertices of T than they would from a proportional split it is possible to embed the remainder of T.

2.2 Definitions and preliminary results

2.2.1 Graphs

For a (simple) graph G, we write $V(G)$ for the vertex set of G, and $|G|$ for the number of vertices of G. $E(G)$ denotes the set of edges of G, and $e(G) := |E(G)|$. Similarly for sets $X, Y \subseteq V(G)$, $e(X, Y)$ denotes the number of edges between X and Y. We sometimes write $v \in G$ to mean $v \in V(G)$, and often denote an edge between x and y by xy. The *degree* of a vertex $v \in G$, denoted $d(v)$, is the number of edges $e \in E(G)$ incident to v. We denote the minimum and maximum degree (taken over all vertices of G) by $\delta(G)$, and $\Delta(G)$ respectively. The *distance* $d(u, v)$ between vertices $u, v \in G$ is the length of the shortest path in G connecting u and v.

2.2.2 Directed graphs

Recall that a directed graph G, or digraph, consists of a vertex set $V(G)$ and an edge set $E(G)$, where each edge $e \in E$ is an ordered pair (u, v) of vertices of G. Again we write $|G|$ for $|V(G)|$, and $v \in G$ to mean $v \in V(G)$. For vertices $u, v \in V(G)$ we write $u \to v$ or $v \leftarrow u$ to denote that $(u, v) \in E(G)$. If $u \to v$ then we say that v is an *outneighbour* of u, that u is an *inneighbour* of v, and that the edge (u, v) is *directed from* u *to* v. Sometimes we use the term *neighbour* of v to mean a vertex which is either an inneighbour or an outneighbour of v. For any vertex $v \in G$, we denote the set of all outneighbours of v by $N^+_G(v)$, or simply $N^+(v)$ when G is clear from the context. Similarly we write $N^-_G(v)$ or $N^-(v)$ to denote the set of all inneighbours of v. Then the *outdegree* of v, denoted $d^+_G(v)$, is defined by $d^+_G(v) := |N^+_G(v)|$. Similarly the *indegree* of v, denoted $d^-_G(v)$, is defined by $d^-_G(v) := |N^-_G(v)|$. Again we may write $d^+(v)$ or $d^-(v)$ when G is clear from the context. We define the minimum outdegree of

G, denoted $\delta^+(G)$, to be the minimum of $d^+(v)$ taken over all vertices $v \in G$, and the minimum indegree, denoted $\delta^-(G)$, to be the minimum of $d^-(v)$ taken over all vertices $v \in G$. Then the minimum *semidegree* of G, denoted $\delta^0(G)$, is the minimum of $\delta^-(G)$ and $\delta^+(G)$. For disjoint sets $U, V \subseteq V(G)$ we write $G[U \to V]$ to denote the bipartite subgraph of G formed by edges directed from U to V.

We say that a directed graph G is an *oriented graph* if for any $u, v \in G$ at most one of $u \to v$ and $u \leftarrow v$ holds. So an oriented graph may be obtained by assigning a direction to each edge of an undirected graph. We call this undirected graph the *underlying graph*, and denote it by G_{under}. Recall that a tournament is an orientation of a complete graph, so the underlying graph of a tournament is a complete graph. Equivalently, G is a tournament if and only if for any distinct $u, v \in V(G)$ precisely one of $u \to v$ and $u \leftarrow v$ holds.

2.2.3 Trees

A *tree* is a connected graph which does not contain any cycles. We often use the fact that for any subtree T' of a tree T and any vertex $x \in T$ there is a unique vertex $y \in T'$ which minimises $d(x, y)$ over all $y \in T'$. Let T be a tree on n vertices; then T has $n - 1$ edges. For any vertex $x \in T$ and edge $e \in E(T)$ incident to x, the *weight of* e *from* x, denoted $w_e(x)$, is the number of vertices $y \neq x$ of T for which e is the first edge of the path from x to y. Equivalently, $w_e(x)$ is the order of the component of $T - e$ which does not contain x. Each vertex $y \neq x$ of T contributes to the weight from x of precisely one edge incident to x, so the sum of the weights from x over all edges incident to x is $n-1$. Also, if xy is an edge of T, then $w_e(x)+w_e(y) = n$. A vertex of a tree is a *leaf* if it has degree one.

A *rooted tree* is a tree with a specified vertex r as a *root*. In a rooted tree every vertex x other than the root has a *parent*; this is defined to be the unique neighbour y of x with $d(y, r) < d(x, r)$. If

y is the parent of x then we say that x is a *child* of y. An *ancestral ordering* of the vertices of a rooted tree is a linear order in which the root appears first and every other vertex appears after its parent.

Recall that a directed tree T is a digraph formed by orienting each edge of an (undirected) tree. Equivalently, the underlying graph T_{under} of T is a tree and at most one of $x \to y$ and $x \leftarrow y$ holds for any pair of vertices x and y of T. Given a specified vertex r as a root, we define parents and children of vertices of the directed tree T exactly as in the underlying tree T_{under}. Similarly $\Delta(T) = \Delta(T_{under})$, and the weight $w_e(x)$ of an edge e incident to a vertex x is defined as in T_{under}. We say that a vertex of a directed tree is a *sink* vertex if it has no outneighbours, and a *source* vertex if it has no inneighbours. Since a directed tree on n vertices has $n-1$ edges, any directed tree must contain at least one sink vertex and at least one source vertex.

Let T be a directed tree, and let x be a vertex of T. We say that a component of $T - x$ is an *incomponent of* x if the unique edge between x and this component is directed towards x, and an *outcomponent of* x if this edge is directed away from x. The *inweight of* x, denoted $w^-(x)$, is then the number of vertices in incomponents of x, and the *outweight of* x, denoted $w^+(x)$, is the number of vertices in outcomponents of x. Equivalently, the inweight of x is the sum of $w_e(x)$ taken over all edges e incident to x which are directed towards x, and the outweight can be defined similarly. In the same way we define incomponents and outcomponents for a subtree T_c of T. Indeed, for any component T' of $T - T_c$ there is precisely one edge between T' and T_c. If this edge is directed towards a vertex of T' then we say that T' is an *outcomponent of* T_c, whereas if this edge is directed towards T_c we say that T' is an *incomponent of* T_c. As when T_c is a single vertex we define the *inweight of* T_c, denoted $w^-(T_c)$, to be the number of vertices in incomponents of T_c, and the *outweight of* T_c, denoted $w^+(T_c)$, to be the number of vertices in outcomponents of T_c. Again these inweights and outweights can equivalently be defined as the sum of the weights of the appropriate edges of T.

We now prove three lemmas relating to trees. The first two of these enable us to split a tree into several pieces with properties that are useful for the analysis of the randomised embedding algorithm used in Section 2.3.

Lemma 2.1 *Let T be a tree on $n \geqslant 3$ vertices. Then there exist subtrees T' and T'' of T such that T' and T'' intersect in precisely one vertex of T, every edge of T lies in precisely one of T' and T'', and $e(T'), e(T'') \geqslant e(T)/3$.*

Proof. We begin by showing that T must contain a vertex v such that every edge e incident to v has $w_e(v) \leqslant n/2$. Recall that if $e = uv$, then $w_e(u) + w_e(v) = n$, and so at most one of $w_e(u) > n/2$ and $w_e(v) > n/2$ can hold. Since T contains n vertices and $n-1$ edges, by the pigeonhole principle T contains a vertex v so that no edge e incident to v has $w_e(v) > n/2$.

Now, choose such a vertex $v \in T$, and let $v_1, \dots, v_r$ be the neighbours of $v \in T$. For each i, let S_i be the set of vertices x of T such that v_i lies on the path from v to x. Then every vertex of T other than v lies in precisely one set S_i. Now, for each i, let T_i be the tree $T[S_i \cup \{v\}]$. Then each T_i is a subtree of T and every edge of T is contained in precisely one T_i. So $\{e(T_i) : i \in [r]\}$ is a set of positive integers, none greater than $2(n-1)/3$, which sum to $n-1$. Thus there exists $A \subseteq [r]$ such that the sum of elements of $\{e(T_i) : i \in A\}$ lies between $(n-1)/3$ and $2(n-1)/3$. Then if we take $T' = \bigcup_{i \in A} T_i$ and $T'' = \bigcup_{i \notin A} T_i$ then T' and T'' satisfy the conditions of the lemma (in particular, $V(T') \cap V(T'') = \{v\}$). $\square$

Lemma 2.2 *Suppose that $1/n \ll 1/\Delta, \varepsilon, 1/k$. Let T be a tree on n vertices satisfying $\Delta(T) \leqslant \Delta$ and rooted at t_1. Then there exist pairwise disjoint subsets $F_1, \dots, F_r$ of $V(T)$, and vertices $v_1, \dots, v_r$ (not necessarily distinct) of T such that:*

(1) $|\bigcup_{i \in [r]} F_i| \geqslant (1-\varepsilon)n$.

(2) $|F_i| \leqslant n^{2/3}$ *for each* i.

(3) For any $i \in [r]$, let $x \in \{t_1\} \cup \bigcup_{j<i} F_j$, and let $y \in F_i$. Then the path from x to y in T includes the vertex v_i.

(4) For any $y \in F_i$ we have $d_T(v_i, y) \geqslant k^3$.

Proof. We begin by splitting T into a family $\mathcal{F}$ of subtrees of T by repeated use of Lemma 2.1. So initially let $\mathcal{F} = \{T\}$, and then repeat the following step. Let T_{large} be the largest member of $\mathcal{F}$. Use Lemma 2.1 to split T_{large} into subtrees T' and T'' which intersect in a single vertex, partition the edges of T_{large}, and satisfy $e(T'), e(T'') \geqslant e(T_{large})/3$. Then remove T_{large} from $\mathcal{F}$, and replace it by the two smaller trees T' and T''. Since in each step we split the largest member of $\mathcal{F}$, at any stage we have $|T_1| \leqslant 3|T_2|$ for any $T_1, T_2 \in \mathcal{F}$. So after at most $3n^{1/3}$ steps we must have that $|T^*| \leqslant n^{2/3}$ for every $T^* \in \mathcal{F}$; at this point we terminate the process.

Observe that if $T', T'' \in \mathcal{F}$, then T' and T'' intersect in at most one vertex. Now, form a graph $\mathcal{G}_{\mathcal{F}}$ with vertex set $\mathcal{F}$ and with an edge between $T', T'' \in \mathcal{F}$ if and only if T' and T'' have a common vertex. Then $\mathcal{G}_{\mathcal{F}}$ is connected, and so contains a spanning tree $\mathcal{T}_{\mathcal{F}}$. Choose T_0 to be a member of $\mathcal{F}$ containing the root t_1 of T, and let $T_0, T_1, \ldots, T_r$ be an ancestral ordering of the members of $\mathcal{F}$ (thought of as vertices of the tree $\mathcal{T}_{\mathcal{F}}$). Now, for each $1 \leqslant i \leqslant r$ let v_i be the common vertex of T_i and its parent in $\mathcal{T}_{\mathcal{F}}$. Then define F_i for each $i \in [r]$ by

$$F_i = V(T_i) \setminus \{x \in T : d_T(v_i, x) < k^3\}.$$

It remains to show that $F_1, \ldots, F_r$ and $v_1, \ldots, v_r$ satisfy the required properties. (4) is immediate from the definition of F_i, and (2) holds since each T_i contained at most $n^{2/3}$ vertices. For (1), observe that every vertex of T was contained in at least one of the subtrees T_i, and that in forming the sets F_i, we deleted at most Δ^{k^3} vertices from each of the at most $3n^{1/3}$ sets $V(T_i)$, so in total at most $3n^{1/3}\Delta^{k^3} \leqslant \varepsilon n$ vertices of T are not contained in any of the sets F_i.

For condition (3), suppose that $T_1' T_2' T_3'$ is a path in $\mathcal{T}_{\mathcal{F}}$, and some vertex v lies in $T_1' \cap T_3'$, but

$v \notin T_2'$. Let $v' \in T_1' \cap T_2'$ and let $v'' \in T_2' \cap T_3'$. Then $v' \neq v''$, as otherwise T_1' and T_3' would have a common vertex other than v. So there is a path from v' to v'' in T which does not contain v, so T contains a cycle, giving a contradiction. Similarly it follows that for any path $T_{i_1} \dots T_{i_j}$ in $\mathcal{T}_{\mathcal{F}}$, if T_{i_1} and T_{i_j} have a common vertex v, then v lies in each of $T_{i_1}, \dots, T_{i_j}$, and so if $T_{i_{j-1}}$ is the parent of T_{i_j} in $\mathcal{T}_{\mathcal{F}}$ then $v = v_{i_j}$. Now, for any $i \in [r]$, if $x \in \{t_1\} \cup \bigcup_{j<i} F_j$ and $y \in F_i$, then $x \in T_j$ for some $0 \leqslant j < i$ and $y \in T_i$. Let $T_j T_1' \dots T_s' T_i$ be the path from T_j to T_i in $\mathcal{T}_{\mathcal{F}}$, then T_s' is the parent of T_i in $\mathcal{T}_{\mathcal{F}}$. This means that v_i is the common vertex of T_s' and T_i, so $T_j \cup T_1' \cup \dots \cup T_s'$ contains a path P_1 from x to v_i and T_i contains a path P_2 from v_i to y. But the property we have proved before implies that P_1 and P_2 only intersect in v_i. Thus $P_1 \cup P_2$ is the path in T from x to y, and $v_i \in P_1 \cup P_2$, as required. It also follows that the sets F_i are pairwise disjoint. $\square$

Recall that in Section 2.3, we describe a randomised algorithm for embedding the vertices of a directed tree T in a digraph G. Whenever this algorithm embeds a vertex t of T in G, it reserves a set of vertices of G in which to embed the children of t. No other vertices may be embedded in this set until all the children of t have been embedded. For this to work, we need to ensure that there are not too many of these reserved sets at any point. This motivates the following definition. If T is a rooted tree on n vertices, then we say that an ancestral ordering of the vertices of T is *tidy* if it has the property that for any initial segment $\mathcal{I}$ of the order, at most $\log_2 n$ vertices in $\mathcal{I}$ have a child not in $\mathcal{I}$. The following lemma shows that such an order exists for any tree T.

Lemma 2.3 *Let T be a tree on n vertices rooted at some $t_0 \in T$. Then there exists a tidy ancestral ordering of the vertices of T.*

Proof. We prove that for any r, the vertices of any rooted tree T on fewer than 2^r vertices can be given an ancestral ordering so that fewer than r vertices in any initial segment $\mathcal{I}$ have neighbours outside $\mathcal{I}$. Indeed, suppose that this statement is false, and let T rooted at t_0 be a counterexample of minimal order, say of order n. Let r be minimal such that $n < 2^r$. Then let $T_1, \ldots, T_s$ be the components of $T - t_0$, ordered in increasing size, and let t_i be the neighbour of t_0 in T_i. We think of t_i as the root of the tree T_i. Then $|T_i| < 2^{r-1}$ for $i \leqslant s-1$, and $T_s < 2^r$. So since T was a minimal counterexample, we can find an ancestral ordering of the vertices of each T_i so that for any $i \leqslant s-1$, any initial segment of the order of the vertices of T_i contains fewer than $r-1$ vertices with children outside the initial segment, and any initial segment of the order of the vertices of T_s contains fewer than r vertices with children outside the initial segment. Now, we order the vertices of T as follows. Begin with t_0, then add the vertices of T_1 in their order. Next, add the vertices of T_2 in their order, and continue in this fashion. Since the order of the vertices of each T_i was ancestral, this order is also ancestral. Also, any initial segment $\mathcal{I}$ of this order contains fewer than r vertices with children outside $\mathcal{I}$, contradicting the choice of T. $\square$

2.2.4 Probabilistic estimates

The next lemma, relating to binomial distributions, is used to show that in the randomised algorithm we use in Section 2.3, the cluster to which a vertex is allocated is almost independent of the cluster to which a vertex far away is allocated. We use $\mathcal{B}(n,p)$ to denote the binomial distribution with parameters n and p, *i.e.* the number of successes in n independent trials, each of which has probability p of success. So $\mathbb{E}(\mathcal{B}(n,p)) = np$.

Lemma 2.4 *Suppose that* $1/k \ll p, (1-p), \varepsilon$*, that* $n \geqslant k^3/6$*, and that* $X = \mathcal{B}(n,p)$*. Then for any* $r \in [k]$*,*

$$\mathbb{P}(X \equiv r \mod k) = \frac{(1 \pm \varepsilon)}{k}.$$

Proof. For each $x \in \{0, \ldots, n\}$ let p_x denote $\mathbb{P}(X = x)$, so $p_x = \binom{n}{x} p^x (1-p)^{n-x}$. Let $\mu = np$, so $\mathbb{E}(X) = \mu$, and let $p_\mu = \max\{p_{\lfloor \mu \rfloor}, p_{\lceil \mu \rceil}\}$, so $p_x \leqslant p_\mu$ for any x. Moreover, if $x \leqslant y \leqslant \mu$ or

$\mu \leqslant y \leqslant x$ then $p_x \leqslant p_y$. So for any $r, i \in [k]$,

$$\begin{aligned}
\mathbb{P}(X \equiv r \mod k) &= \sum_{\substack{0 \leqslant x \leqslant \mu-k \\ x \equiv r \mod k}} p_x + \sum_{\substack{\mu-k < x \leqslant \mu+k \\ x \equiv r \mod k}} p_x + \sum_{\substack{\mu+k < x \leqslant n \\ x \equiv r \mod k}} p_x \\
&\leqslant \sum_{\substack{0 \leqslant x \leqslant \mu-k \\ x \equiv r \mod k}} p_{x+i} + 2p_\mu + \sum_{\substack{\mu+k < x \leqslant n \\ x \equiv r \mod k}} p_{x-k+i} \\
&\leqslant \mathbb{P}(X \equiv r+i \mod k) + 2p_\mu.
\end{aligned}$$

So $\mathbb{P}(X \equiv r \mod k) = 1/k \pm 2p_\mu = (1 \pm \varepsilon)/k$ for any $r \in [k]$, using a standard result (*e.g.* [5], Section 1.2) on the binomial distribution which states that $p_\mu = O(n^{-1/2}) = O(k^{-3/2})$. □

The following two results give useful tail estimates for random variables. The first is an Azuma-type inequality which bounds the sum of many small and almost independent random variables. This is derived in [44] from a result in [34]. ([44] uses a random walk to embed trees in sparse undirected graphs.) The second gives standard Chernoff-type bounds for the binomial and hypergeometric distributions. The hypergeometric random variable X with parameters (n, m, k) is defined as follows. Let N be a set of size n, and fix a set $S \subseteq N$ of size $|S| = m$. Now choose a set $T \subseteq N$ of size $|T| = k$ uniformly at random. Then $X = |T \cap S|$. Note that $\mathbb{E}X = km/n$.

Lemma 2.5 ([44], Proposition 1.1) *Let $X_1, \ldots, X_n$ be random variables taking values in $[0,1]$ such that for each $k \in [n]$,*

$$\mathbb{E}[X_k \mid X_1, \ldots, X_{k-1}] \leqslant a_k.$$

Let $\mu \geqslant \sum_{i=1}^n a_i$. Then for any $0 < \delta < 1$,

$$\mathbb{P}[\sum_{i=1}^{n} X_i > (1+\delta)\mu] \leqslant e^{-\frac{\delta^2 \mu}{3}}.$$

Proposition 2.6 ([21], Corollary 2.3 and Theorem 2.10) *Suppose X has binomial or hypergeometric distribution and $0 < a < 3/2$. Then $\mathbb{P}(|X - \mathbb{E}X| \geqslant a\mathbb{E}X) \leqslant 2e^{-\frac{a^2}{3}\mathbb{E}X}$.*

2.2.5 Regularity and robust outexpanders

To prove Theorem 1.5 we make use of a directed version of Szemerédi's Regularity lemma. For this, we make the following definitions. If G is an undirected bipartite graph with vertex classes X and Y, then the density of G is defined as

$$d(X,Y) := \frac{e(X,Y)}{|X||Y|}.$$

Now, for any $\varepsilon > 0$, we say that G is ε-*regular* if for any $X' \subseteq X$ and $Y' \subseteq Y$ with $|X'| \geqslant \varepsilon|X|$ and $|Y'| \geqslant \varepsilon|Y|$ we have $|d(X',Y') - d(X,Y)| < \varepsilon$.

Recall that for disjoint sets X and Y of vertices of a digraph G, we use $G[X \to Y]$ to denote the edges of G directed from X to Y. We say $G[X \to Y]$ is ε-*regular* if the underlying bipartite graph of $G[X \to Y]$ is ε-regular. The density of $G[X \to Y]$ is similarly defined to be the density of the underlying bipartite graph. Next we state the degree form of the regularity lemma for digraphs. A regularity lemma for digraphs was proven by Alon and Shapira [2]. The degree form follows from this in the same way as the undirected version (see [32] for a sketch of the latter).

Lemma 2.7 (Regularity Lemma for digraphs) *For any ε, M' there exist M, n_0 such that if G is a digraph on $n \geqslant n_0$ vertices and $d \in [0,1]$, then there exists a partition of $V(G)$ into $V_0, \dots, V_k$ and a spanning subgraph G' of G such that*

(1) $M' \leqslant k \leqslant M$,

(2) $|V_0| \leqslant \varepsilon n$,

(3) $|V_1| = \cdots = |V_k|$*,*

(4) $d^+_{G'}(x) > d^+_G(x) - (d+\varepsilon)n$ *for all vertices* $x \in V(G)$*,*

(5) $d^-_{G'}(x) > d^-_G(x) - (d+\varepsilon)n$ *for all vertices* $x \in V(G)$*,*

(6) *for all* $i \in [k]$ *the digraph* $G'[V_i]$ *is empty,*

(7) *for all* $1 \leqslant i, j \leqslant k$ *with* $i \neq j$ *the pair* $G'[V_i \to V_j]$ *is* ε*-regular and either has density 0 or density at least* d*.*

We refer to $V_1, \ldots, V_k$ as *clusters*. Given a graph G on n vertices, we form the *reduced digraph* R *of* G *with parameters* ε, d *and* M' by applying the regularity lemma with these parameters to obtain $V_0, \ldots, V_k$. R is then the digraph on vertex set $\{1, \ldots, k\}$, with $i \to j$ an edge precisely when $G'[V_i \to V_j]$ is ε-regular with density at least d.

One particular regular structure appears frequently in Section 2.3 and Section 2.4. We say that a digraph G is an *ε-regular d-dense cycle of cluster tournaments* if $V(G) = V_1 \cup \cdots \cup V_k$, where the sets V_i are pairwise disjoint and of equal size, and for each i, $G[V_i]$ is a tournament and $G[V_i \to V_{i+1}]$ is ε-regular with density at least d (where addition and subtraction on the indices of clusters should be taken modulo k). We often refer to the sets V_i as clusters, as we obtain them by an application of the regularity lemma.

Now, let $V_1, \ldots, V_k$ be disjoint sets of m vertices, and let G be a digraph on vertex set $V_1 \cup \cdots \cup V_k$. Let S be a subset of some cluster V_i. Then we say that S is (c, γ)-*good* if for any $V'_{i-1} \subseteq V_{i-1}$ and $V'_{i+1} \subseteq V_{i+1}$ with $|V'_{i-1}| \geqslant cm$ and $|V'_{i+1}| \geqslant cm$, S contains at least $\gamma\sqrt{m}$ vertices which each have at least γm inneighbours in V'_{i-1} and at least γm outneighbours in V'_{i+1}. Our main tool in the use of regularity in this chapter is the next lemma, which states that if G is a regular and dense cycle of cluster tournaments, then any subset V'_i of any cluster V_i with $|V'_i| \geqslant \gamma m/2$ contains a (c, γ)-good subset S of size at most $\sqrt{m}$.

Lemma 2.8 *Suppose that $1/m \ll \varepsilon \ll \gamma \ll c, d$. Let G be an ε-regular d-dense cycle of cluster tournaments on clusters $V_1, \ldots, V_k$, each of size m. Then for any i and for any $V_i' \subseteq V_i$ of size $|V_i'| = \gamma m/2$, there exists a (c, γ)-good set $S \subseteq V_i'$ with $|S| \leqslant \sqrt{m}$.*

Proof. Given $V_i' \subseteq V_i$ of size $|V_i'| = \gamma m/2$, choose $S \subseteq V_i'$ at random by including each vertex of V_i' in S with probability $1/\gamma\sqrt{m}$, independently of the outcome for each other vertex. Then by Proposition 2.6, with probability $1 - o(1)$, $|S| \leqslant \sqrt{m}$.

Now, $G[V_{i-1} \to V_i']$ and $G[V_i' \to V_{i+1}]$ are each $(2\varepsilon/\gamma)$-regular with density at least $d/2$. So all but at most $2\varepsilon m/\gamma$ vertices $v_{i-1} \in V_{i-1}$ have at least $\gamma dm/5$ outneighbours in V_i'. Fix any such $v_{i-1} \in V_{i-1}$. Then $G[V_i' \cap N^+(v_{i-1}) \to V_{i+1}]$ is $(5\varepsilon/\gamma d)$-regular with density at least $d/2$. So all but at most $5\varepsilon m/\gamma d$ vertices $v_{i+1} \in V_{i+1}$ have at least $\gamma d^2 m/20$ inneighbours in $V_i' \cap N^+(v_{i-1})$. We therefore conclude that all but at most $7\varepsilon m^2/\gamma d$ pairs (v_{i-1}, v_{i+1}) with $v_{i-1} \in V_{i-1}$, $v_{i+1} \in V_{i+1}$ have at least $\gamma d^2 m/20$ common neighbours in V_i'.

By Proposition 2.6, for each such pair (v_{i-1}, v_{i+1}) the probability that (v_{i-1}, v_{i+1}) has fewer than $d^2\sqrt{m}/25$ common neighbours in S decreases exponentially with m, whilst the number of such pairs is quadratic in m. Thus with probability $1 - o(1)$ our randomly selected S has the property that all but at most $7\varepsilon m^2/\gamma d$ pairs (v_{i-1}, v_{i+1}) with $v_{i-1} \in V_{i-1}$, $v_{i+1} \in V_{i+1}$ have at least $d^2\sqrt{m}/25$ common neighbours in S. We may therefore fix an outcome of our random choice of S such that both of these events of probability $1 - o(1)$ occur.

So if $|V_{i-1}'| \geqslant cm$ and $|V_{i+1}'| \geqslant cm$, then we know that at least $c^2m^2/2$ pairs (v_{i-1}, v_{i+1}) with $v_{i-1} \in V_{i-1}'$, $v_{i+1} \in V_{i+1}'$ have at least $d^2\sqrt{m}/25$ common neighbours $s \in S$. Thus there are at least $c^2d^2m^{5/2}/50$ triples of such vertices (v_{i-1}, s, v_{i+1}), so at least $c^2d^2\sqrt{m}/100 \geqslant \gamma\sqrt{m}$ vertices in S must lie in the common neighbourhood of at least $c^2d^2m^2/100$ such pairs (v_{i-1}, v_{i+1}). Each of these vertices therefore has at least $c^2d^2m/100 \geqslant \gamma m$ neighbours in each of V_{i-1}' and V_{i+1}', as required. $\square$

We also make use of the following well known observation, which says that if G is a regular and dense cycle of cluster tournaments on clusters $V_1, \ldots, V_k$, and we select subsets $U_1 \subseteq V_1, \ldots, U_k \subseteq V_k$ uniformly at random, then with high probability the restriction of G to these subsets is also regular and dense. This follows from a lemma of Alon et al. [1] showing that ε-regularity is equivalent to almost all vertices having the expected degree and almost all pairs of vertices having the expected common neighbourhood size. We include the proof for completeness.

Lemma 2.9 *Suppose that $1/m \ll k \ll \varepsilon \ll \varepsilon' \ll d$ and that $m^{1/3} \leqslant m' \leqslant m$. Let G be an ε-regular d-dense cycle of cluster tournaments on clusters $V_1, \ldots, V_k$, each of size m. For each $i \in [k]$, choose $U_i \subseteq V_i$ of size m' uniformly at random, and independently of all other choices. Then with probability $1 - o(1)$, $G[U_1 \cup \cdots \cup U_k]$ is an ε'-regular $d/2$-dense cycle of cluster tournaments.*

Proof. We need to show that with high probability, $G[U_i \to U_{i+1}]$ is ε'-regular with density at least $d/2$ for each i. So fix some $i \in [k]$, and let $d_i \geqslant d$ be the density of $G[V_i \to V_{i+1}]$. Also, let B_i be the set of vertices $v \in V_i$ for which $|N^+(v) \cap V_{i+1}| \neq (d_i \pm \varepsilon)m$, and let D_i be the set of pairs $v_1 \neq v_2$ of vertices of V_i for which $|N^+(v_1) \cap N^+(v_2) \cap V_{i+1}| \neq (d_i^2 \pm 3\varepsilon)m$. Then since $G[V_i \to V_{i+1}]$ is ε-regular, $|B_i| \leqslant 2\varepsilon m$. Also, there are at most $2\varepsilon m^2$ pairs in D_i which contain a vertex of B_i, and each $v \in V_i \setminus B_i$ lies in at most $2\varepsilon m$ pairs in D_i, so $|D_i| \leqslant 4\varepsilon m^2$. So let $B'_i = B_i \cap U_i$ and similarly let D'_i consist of the pairs in D_i for which both vertices of the pair are in U_i. Then by Proposition 2.6, the probability that either $|B'_i| > 4\varepsilon m'$ or $|D'_i| > 8\varepsilon(m')^2$ declines exponentially with m.

Now, for each of the at most m' vertices $v \in U_i \setminus B_i$, by Proposition 2.6 the probability that $|N^+(v) \cap U_{i+1}| \neq (d_i \pm 2\varepsilon)m'$ decreases exponentially with m. Also, for each of the at most $\binom{m'}{2}$ pairs $v_1 \neq v_2$ with $v_1, v_2 \in U_i \setminus D_i$, the probability that $|N^+(v_1) \cap N^+(v_2) \cap U_{i+1}| \neq (d_i^2 \pm 4\varepsilon)m'$ decreases exponentially with m. So with probability $1 - o(1)$, for each i none of these events of exponentially declining probability hold.

Fix such an outcome of our random choices. Then for each i at least $(1-4\varepsilon)m'$ vertices $v \in U_i$ have $|N^+(v_1) \cap U_{i+1}| = (d_i \pm 2\varepsilon)m'$ and at least $\binom{m'}{2} - 8\varepsilon(m')^2$ pairs $v_1, v_2 \in U_i$ have $|N^+(v_1) \cap N^+(v_2) \cap U_{i+1}| = (d_i^2 \pm 4\varepsilon)m'$. It then immediately follows from Lemma 3.2 of [1] that for each i, $G[U_i \to U_{i+1}]$ is ε'-regular (and it is clear that this has density at least $d_i/2 \geqslant d/2$), as desired. $\square$

We now turn to the concept of a robust outexpander. Let $\mu > 0$, let G be a digraph on n vertices, and let $S \subseteq V(G)$. Then the *robust μ-outneighbourhood* of S, denoted $RN^+_\mu(S)$, is defined to be the set of vertices of G with at least μn inneighbours in S. For constants $0 < \mu \leqslant \nu < 1$, we say that a digraph G on n vertices is a *robust (μ, ν)-outexpander* if $|RN^+_\mu(S)| \geqslant |S| + \mu n$ for all $S \subseteq V(G)$ with $\nu n < |S| < (1-\nu)n$. A recent result from [33] (which in turn relies on results from [26, 25]) states that every robust outexpander with linear minimum semidegree contains a Hamilton cycle. We make use of this to prove the next lemma, which states that any robust outexpander tournament G contains a regular and dense cycle of cluster tournaments which includes almost all of the vertices of G. We use this structure when we embed a directed tree T in a tournament G which is a robust outexpander. Note that this lemma includes the first three steps of the regularity method as described in the introduction. Indeed, given a tournament G which is a robust outexpander of linear minimum semidegree, the effect of this lemma is to apply the Szemerédi regularity lemma to G, to form the reduced graph R of G, and to find a Hamilton cycle R' in R, which corresponds to a regular and dense cycle of cluster tournaments in G.

Lemma 2.10 *Suppose that $1/n \ll 1/M \ll 1/M' \ll \varepsilon \ll d \ll \mu \ll \nu \ll \eta$. Let G be a tournament on n vertices which is a robust (μ, ν)-outexpander with $\delta^0(G) \geqslant \eta n$. Then G contains an ε-regular d-dense cycle of cluster tournaments on clusters $V_1, \dots, V_k$, where $|\bigcup_{i=1}^k V_i| > (1-\varepsilon)n$, and $M' \leqslant k \leqslant M$.*

Proof. Let R be the reduced digraph of G with parameters ε, d and M' obtained by applying Lemma 2.7, and let $k = |R|$, so $M' \leqslant k \leqslant M$. Then by Lemma 12 of [33], $\delta^0(R) \geqslant \eta|R|/2$, and R is a robust $(\mu/2, 2\nu)$-outexpander. Then by Theorem 14 of [33], which states that any robust outexpander of linear minimum semidegree contains a Hamilton cycle, we know that R contains a Hamilton cycle. Let $V_1, \dots, V_k$ be the clusters of R in the order of the cycle. Then $|\bigcup_{i=1}^k V_i| > (1-\varepsilon)n$, $G[V_i]$ is a tournament for each i and, since $V_i \to V_{i+1}$ is an edge of R for each i, $G'[V_i \to V_{i+1}]$ is ε-regular with density at least d. (Here G' is the spanning subgraph of G obtained by Lemma 2.7.) $\square$

Of course, we sometimes need to embed a directed tree T in a tournament G which is not a robust outexpander. The next lemma is a useful tool in this situation; it states that if a tournament G is not a robust outexpander then $V(G)$ can be partitioned into two sets so that most edges between the two sets have the same direction.

Lemma 2.11 *Suppose that $1/n \ll \mu \ll \nu$, that G is a tournament on n vertices and that G is not a robust (μ, ν)-outexpander. Then we can partition $V(G)$ into sets S and S' such that $\nu n < |S|, |S'| < (1-\nu)n$ and $e(G[S \to S']) \leqslant 4\mu n^2$.*

Proof. Since G is not a robust (μ, ν)-outexpander there exists $S \subseteq V(G)$ such that $|RN^+_\mu(S)| < |S| + \mu n$ and $\nu n < |S| < (1-\nu)n$. Choose such an S, and let $S' = V(G) \setminus S$, so $\nu n < |S'| < (1-\nu)n$ also.

Since G is a tournament, at most $2\mu n + 1$ vertices $v \in S$ have $d^-_{G[S]}(v) < \mu n$, and so at most $2\mu n + 1$ vertices $v \in S$ have $v \notin RN^+_\mu(S)$. So $|RN^+_\mu(S) \setminus S| \leqslant 3\mu n + 1$, and so the number of edges from S to S' is at most $(3\mu n + 1)|S| + \mu n|S'| \leqslant 4\mu n^2$. $\square$

By repeated application of Lemma 2.11, we may obtain a decomposition of a tournament G into sets S_i which either induce robust outexpanders or are small, and where for all $i < j$, almost all edges between S_i and S_j are directed from S_i to S_j. In particular, if all the S_i are small then G is close to being a transitive tournament. We use this decomposition in Section 3.6 to prove Theorem 1.1.

Lemma 2.12 *Suppose that $1/n \ll \mu \ll \nu \ll \eta \ll \zeta \ll 1$. Let G be a tournament on n vertices. Then we may choose disjoint subsets $S_1, \ldots, S_r$ of $V(G)$ such that:*

(i) $|\bigcup_{i \in [r]} S_i| \geqslant (1-\zeta)n$,

(ii) for each $i \in [r]$, any vertex $v \in S_i$ has at most ζn inneighbours in $\bigcup_{j>i} S_j$ and at most ζn outneighbours in $\bigcup_{j<i} S_j$, and

(iii) for each $i \in [r]$, either $G[S_i]$ is a robust (μ, ν)-outexpander with $\delta^0(G[S_i]) \geqslant \eta n$ or $|S_i| < \zeta n$.

Proof. To prove the lemma, we use an algorithm which keeps track of an ordered family $\mathcal{S}^\tau$ of disjoint subsets of $V(G)$, and a set B^τ of bad edges of G, at each time τ. Initially, let $\mathcal{S}^1 := (V(G))$, and let $B^1 := \emptyset$. Then at time $\tau \geqslant 1$, we have $\mathcal{S}^\tau = (S_1^\tau, \ldots, S_\tau^\tau)$, and the algorithm proceeds as follows.

1. Let S_ℓ^τ be the largest member of $\mathcal{S}^\tau$ which is not a robust (μ,ν)-outexpander with $\delta^0(G[S_\ell^\tau]) \geqslant \eta n$. If there is no such member of $\mathcal{S}^\tau$, or if $|S_\ell^\tau| < \zeta n$, then terminate. If there is more than one largest such member, then choose one of these arbitrarily.

2. If some $v \in S_\ell^\tau$ has $d^+_{G[S_\ell^\tau]}(v) < \eta n$, then let

$$\mathcal{S}^{\tau+1} := (S_1^\tau, \ldots, S_{\ell-1}^\tau, S_\ell^\tau \setminus \{v\}, \{v\}, S_{\ell+1}^\tau, \ldots, S_\tau^\tau),$$

   let $B^{\tau+1} := B^\tau \cup E(G(\{v\} \to S_\ell^\tau \setminus \{v\}))$, and proceed to step (5).

3. Similarly, if some $v \in S_\ell^\tau$ has $d^-_{G[S_\ell^\tau]}(v) < \eta n$, then let

$$\mathcal{S}^{\tau+1} := (S_1^\tau, \dots, S_{\ell-1}^\tau, \{v\}, S_\ell^\tau \setminus \{v\}, S_{\ell+1}^\tau, \dots, S_\tau^\tau),$$

   let $B^{\tau+1} := B^\tau \cup E(G(S_\ell^\tau \setminus \{v\} \to \{v\}))$, and proceed to step (5).

4. If $G[S_\ell^\tau]$ is not a robust (μ, ν)-outexpander then apply Lemma 2.11 to partition the vertices of S_ℓ^τ into sets S' and S'' such that $\nu|S_\ell^\tau| \leqslant |S'|, |S''| \leqslant (1-\nu)|S_\ell^\tau|$ and at most $4\mu|S_\ell^\tau|^2$ edges of $G[S_\ell^\tau]$ are directed from S'' to S'. Then let

$$\mathcal{S}^{\tau+1} := (S_1^\tau, \dots, S_{\ell-1}^\tau, S', S'', S_{\ell+1}^\tau, \dots, S_\tau^\tau)$$

   and let $B^{\tau+1} := B^\tau \cup E(G(S'' \to S'))$.

5. Finally, for each $i \in [\tau+1]$, delete from $S_i^{\tau+1}$ any vertex v which lies in more than $\sqrt{\eta} n$ edges of $B^{\tau+1}$.

At any time τ, if the algorithm does not terminate at step (1) then S_ℓ^τ is split in precisely one of steps (2), (3) and (4). So at each time τ, either the algorithm terminates or $|\mathcal{S}^\tau|$ increases from τ to $\tau+1$ (in forming $\mathcal{S}^{\tau+1}$) by reducing the size of the largest member of S^τ. Therefore the algorithm must terminate at some time $\tau_{end} \leqslant n$. Take $r := \tau_{end}$, and $S_i := S_i^r$ for each i. Then since the algorithm terminated at step (1) of time r, (iii) must hold.

To see (i), observe that the split in step (4) occurs for at most $1/\zeta\nu$ times $\tau < \tau_{end}$. This is because any set obtained by a split in step (4) must have size at least $\zeta\nu n$ (since $|S_\ell^\tau| \geqslant \zeta n$, and the sets S', S'' obtained have $|S'|, |S''| \geqslant \nu|S_\ell^\tau|$). Also, at each time $\tau < \tau_{end}$, the number of edges added to form $B^{\tau+1}$ from B^τ is at most ηn if the algorithm carried out the split in step (2) or (3), and at most $4\mu n^2$ if the algorithm carried out the split in step (4). Since $\tau_{end} \leqslant n$, and the split in step (4) is carried out in at most $1/\zeta\nu$ steps, we must have

$$|B^{\tau_{end}}| \leqslant \eta n^2 + \frac{4\mu n^2}{\nu\zeta} \leqslant 2\eta n^2.$$

Since $B^1 \subseteq \cdots \subseteq B^{\tau_{end}}$, any vertex of G which was ever deleted in step (5) must lie in at least $\sqrt{\eta}n$ edges of $B^{\tau_{end}}$, and so at most $4\sqrt{\eta}n \leqslant \zeta n$ vertices of G can have been deleted in step (5) over the entire course of the algorithm. But any vertex which was not deleted lies in some S_i, and so (i) holds.

Finally, for (ii) fix any $i \in [r]$ and any $v \in S_i$. Observe that all edges directed from v to $\bigcup_{j<i} S_j$ and all edges directed from $\bigcup_{j>i} S_j$ to v are contained in B^r. This means that there are at most $\sqrt{\eta}n$ such edges, as otherwise v would have been deleted in step (5) at some point. Since i and v were arbitrary, (ii) must hold. $\square$

2.3 Embedding trees of bounded maximum degree in a robust outexpander

2.3.1 Introduction

Our aim in this section is the following lemma on embedding directed trees of bounded maximum degree in robust outexpander tournaments.

Lemma 2.13 *Suppose that $1/n \ll \mu \ll \nu \ll \eta \ll \alpha, 1/\Delta$, that G is a tournament on $(1+\alpha)n$ vertices which is a robust (μ, ν)-outexpander with $\delta^0(G) \geqslant \eta n$ and that T is a directed tree on n vertices with $\Delta(T) \leqslant \Delta$. Then G contains a copy of T.*

The proof of this lemma shows that we could actually put $1/\Delta$ lower down in the hierarchy, but this is how we apply this lemma later on. To prove this lemma, we begin by applying Lemma 2.10 to find a regular and dense cycle of cluster tournaments in G, containing almost all of the vertices of G. We then use Lemma 2.14 to find a copy of T within this structure. This lemma is stated separately, and in a stronger form than necessary, as we also make use of it in Section 2.4.

Lemma 2.14 *Suppose that $1/n \ll 1/k, 1/\Delta \ll \varepsilon \ll d \ll \alpha \leqslant 2$, and that $m = n/k$. Let G be an ε-regular d-dense cycle of cluster tournaments on clusters $V_1, \dots, V_k$ of equal size $(1+\alpha)m$. Let v^* be a vertex of V_1 with at least d^2m inneighbours in V_k and at least d^2m outneighbours in V_2. Finally, let T be a directed tree on n vertices, rooted at t_1 and with $\Delta(T) \leqslant \Delta$. Then G contains a copy of T, where the vertex t_1 of T corresponds to the vertex v^* of G.*

The main problem in achieving this is to allocate the vertices of T to the clusters V_i in such a way that we can then use the ε-regularity of each $G[V_i \to V_{i+1}]$ to embed the vertices of T in G. When we say we allocate v to V_i this means that v is to be embedded to a vertex of V_i, but this embedding has not been fixed yet. We wish to allocate each vertex of T to a cluster V_i so that for most edges $u \to v$ of T, if u is allocated to V_i then v is allocated to V_{i+1}. So if u is allocated to a cluster V_i and $u \to v$ then we say that the *canonical allocation* of v is to the cluster V_{i+1}, whereas if $u \leftarrow v$ then we say that the canonical allocation of v is to the cluster V_{i-1}. If we allocate v canonically, then we say that the edge between u and v has been allocated canonically. One way of allocating the vertices of T to the clusters V_i would be to begin by allocating the root t_1 to V_1, and then to allocate all remaining vertices canonically. However, to successfully embed the vertices of T within the clusters to which they are allocated we need the vertices of T to be approximately evenly distributed amongst the k clusters. This method often fails to achieve this, *e.g.* if T is an anti-directed path.

To obtain an 'even distribution' for any sufficiently large directed tree of bounded maximum degree, we modify the method so that some vertices (selected randomly) are allocated to the same cluster as their parent, rather than being allocated canonically. However, having large components of vertices which are allocated to the same cluster may prevent a successful embedding of these vertices within this cluster, and so we also require that such components are small. This is the motivation behind the Vertex Allocation Algorithm given in the next subsection, which we use to allocate the vertices of T.

2.3.2 Allocating the vertices of T

We use the following random process to allocate the vertices of T to the clusters V_i.

Vertex Allocation Algorithm:

Input: A directed tree T on n vertices, a root vertex $t_1 \in T$, and clusters $V_1, \dots, V_k$.

Initialisation: Choose an ancestral ordering $t_1, \dots, t_n$ of the vertices of T.

Procedure: At time $\tau = 1$, allocate t_1 to V_1. Then at time $\tau \geqslant 1$, we allocate t_τ. Let t_σ be the parent of t_τ, which must have appeared before t_τ in the ordering and has therefore already been allocated. Then:

- If $d_T(t_\tau, t_1)$ is odd, then allocate t_τ canonically.
- If $d_T(t_\tau, t_1)$ is even, then allocate t_τ to the same cluster as t_σ with probability $1/2$, and allocate t_τ canonically with probability $1/2$ (where these choices are made independently for each vertex).

Termination: Terminate when every vertex of T has been processed and therefore allocated to some cluster V_j.

Note that the cluster to which a vertex t is allocated by this algorithm depends only on the cluster to which its parent vertex was allocated and the outcome of the random choice when embedding t (if $d(t, t_1)$ is even). Since these choices were independent, the probability of any possible outcome does not depend on which ancestral order of the vertices was chosen in the initialisation step. Now, we say that an edge of T is *allocated within a cluster* if both of its endvertices are allocated to the same cluster. Then we say that an allocation of the vertices of a directed tree T to clusters $V_1, \ldots, V_k$ is *semi-canonical* if

(i) every edge of T is either allocated canonically or is allocated within a cluster,

(ii) every edge of T incident to t_1 is allocated canonically, and

(iii) every component of the subgraph of T formed by all edges allocated within a cluster contains at most $\Delta(T)$ vertices.

The next lemma shows that if we allocate the vertices of a directed tree T to clusters $V_1, \ldots, V_k$ by applying the Vertex Allocation Algorithm, then the allocation obtained is semi-canonical, and also that if vertices t and t' are far apart in T then the cluster to which t is allocated is almost independent of the cluster to which t' is allocated. As a consequence, if T is sufficiently large and has bounded maximum degree, each cluster has approximately equally many vertices of T allocated to it. These properties allow us to embed the vertices of such a T into a regular and dense cycle of cluster tournaments G in the next subsection.

Lemma 2.15 *Let T be a directed tree on n vertices rooted at t_1. Allocate the vertices of T to clusters $V_1, \ldots, V_k$ by applying the Vertex Allocation Algorithm. Then the following properties hold.*

(a) The allocation obtained is semi-canonical.

(b) *Suppose that $1/k \ll \delta$. Let u and v be vertices of T such that u lies on the path from t_1 to v, and $d_T(u,v) \geqslant k^3$. Then for any $i, j \in [k]$,*

$$\mathbb{P}(v \textit{ is allocated to } V_i \mid u \textit{ is allocated to } V_j) = \frac{1 \pm \delta/4}{k}.$$

(c) *Now suppose that $1/n \ll 1/\Delta, 1/k \ll \delta$, and that $\Delta(T) \leqslant \Delta$. Then with probability $1 - o(1)$, each of the k clusters V_i has at most $(1+\delta)m$ vertices of T allocated to it, where $m = n/k$.*

Proof. (a) The Vertex Allocation Algorithm allocates every vertex either canonically or to the same cluster as its parent, so every edge is allocated canonically or within a cluster. Furthermore, a vertex t can only be allocated to the same cluster as its parent if $d(t_1, t)$ is even, and so each edge incident to t_1 is allocated canonically. Finally, since edges allocated within a cluster can only be formed when we allocate t_i such that $d(t_i, t_1)$ is even, any such component is a star formed by some t_j and some of the children of t_j.

(b) Since the order in which the vertices are allocated is ancestral, at the stage in our algorithm when we have just allocated u, no other vertex on the path $P(u,v)$ in T from u to v has yet been allocated. So suppose that we have just allocated u to cluster V_j, let ℓ be the length of $P(u,v)$, so $\ell \geqslant k^3$, and let $u = v_0, v_1, \ldots, v_\ell = v$ be the vertices of $P(u,v)$. Then let $E = \{i \geqslant 1 \colon d(v_i, t_1) \text{ is even}\}$, so E indicates the vertices with a random element in their allocation, and let $O = [\ell] \setminus E$, so O indicates the vertices which are allocated deterministically.

We then split the edges of $P(u,v)$ into four classes:

$$\begin{aligned} F_{canon} =& \{v_{i-1} \to v_i : i \in O\} \\ B_{canon} =& \{v_{i-1} \leftarrow v_i : i \in O\} \\ F_{random} =& \{v_{i-1} \to v_i : i \in E\} \\ B_{random} =& \{v_{i-1} \leftarrow v_i : i \in E\}. \end{aligned}$$

Then every edge of $P(u,v)$ lies in one of these 4 sets, and so $|F_{canon}| + |B_{canon}| + |F_{random}| + |B_{random}| = \ell$. Furthermore, each edge in F_{canon} is allocated canonically, and hence from some V_i to V_{i+1}. Similarly, edges in B_{canon} are allocated from some V_i to V_{i-1}. Meanwhile, edges in F_{random} or B_{random} are allocated from some V_i to V_{i+1} or V_{i-1} respectively with probability $1/2$, and within some V_i with probability $1/2$. So let R be the sum of the number of edges from F_{random} which are allocated canonically and the number of edges from B_{random} which are *not* allocated canonically. Since the outcome of the random experiment for each edge is independent of the outcome for any other edge, R has distribution $\mathcal{B}(|E|, 1/2)$. Now, u was allocated to cluster V_j, and so v is allocated to cluster V_i, where

$$i \equiv j + |F_{canon}| - |B_{canon}| + R - |B_{random}| \mod k.$$

But since $|E| \geqslant \lfloor \ell/2 \rfloor \geqslant k^3/3$, Lemma 2.4 implies that for any $r \in [k]$, the probability that $i = r$ is $\frac{1 \pm \delta/4}{k}$, as desired.

(c) Use Lemma 2.2 to choose pairwise disjoint subsets $F_1, F_2, \ldots, F_r$ of $V(T)$ and vertices $v_1, \ldots, v_r \in V(T)$ such that $|\bigcup_{i \in [r]} F_i| \geqslant (1 - \delta/2k)n$ and $|F_i| \leqslant n^{2/3}$ for each i, also such that if $j < i$, then any path from t_1 or any vertex of F_j to any vertex of F_i passes through the vertex v_i, and finally such that $d(v_i, F_i) \geqslant k^3$ for each i. We prove that $(\dagger)$ with probability $1 - o(1)$, the total number of vertices from any of the sets F_i allocated to cluster V_j is at most $(1 + \delta/2)m$.

This proves the lemma, as the number of vertices of T not contained in any of the sets F_i is at most $\delta m/2$, and so in total at most $(1+\delta)m$ vertices of T are allocated to any cluster V_j.

To prove (†), define random variables X_i^j for each $i \in [r]$, $j \in [k]$ by

$$X_i^j = \frac{\text{\# of vertices of } F_i \text{ allocated to cluster } V_j}{n^{2/3}},$$

so that each X_i^j lies in the range $[0,1]$. Then since the cluster to which a vertex t of T is allocated is dependent only on the cluster to which the parent of t is allocated and on the outcome of the random choice made when allocating t, $\mathbb{E}[X_i^j \mid X_{i-1}^j, \dots, X_1^j, v_i \in V_s] = \mathbb{E}[X_i^j \mid \; v_i \in V_s]$ for all $s \in [k]$. Here we write $v_i \in V_s$ to denote the event that v_i is allocated to V_s. So for any i and j,

$$\begin{aligned}\mathbb{E}[X_i^j \mid X_{i-1}^j, \dots, X_1^j] &\leqslant \max_{s\in[k]} \mathbb{E}[X_i^j \mid X_{i-1}^j, \dots, X_1^j, v_i \in V_s] = \max_{s\in[k]} \mathbb{E}[X_i^j \mid v_i \in V_s] \\ &= \max_{s\in[k]} \frac{\sum_{x\in F_i} \mathbb{P}(x \in V_j \mid v_i \in V_s)}{n^{2/3}} \leqslant \frac{(1+\delta/4)|F_i|}{kn^{2/3}}.\end{aligned}$$

using (b). So, by Lemma 2.5, with probability $1-o(1)$ we have that for each j,

$$\sum_{i\in[r]} X_i^j \leqslant \frac{(1+\delta/2)m}{n^{2/3}}$$

and so for each j, the total number of vertices from any of the sets F_i allocated to cluster V_j is at most $(1+\delta/2)m$, proving (†). □

2.3.3 Embedding the vertices of T

Suppose that we have applied the Vertex Allocation Algorithm to find an approximately uniform allocation of the vertices of T to the clusters of G. We now wish to embed T in G so that each vertex is embedded in the cluster to which it is allocated. In principle we could use the blow–

up lemma for this. However numerous complications arise, for instance because we embed some edges within clusters and because we allow Δ to be comparatively large in Section 2.4. Instead, we embed the vertices of T as follows. Firstly, to deal with the problem of edges which are allocated within a cluster, we embed components of T formed by such edges at the same time, using Theorem 1.2 (it would also be easy to do this directly). To do this we make the following definition. Let T be a directed tree on n vertices with root t_1, and let the vertices of T be allocated to clusters $V_1, \ldots, V_k$ by a semi-canonical allocation. Then the *canonical tree* T_{canon} of T is formed by contracting to a single vertex each component of the subgraph of T formed of edges which are allocated within a cluster. Since the allocation is semi-canonical, each such component contains at most Δ vertices – we say that these vertices *correspond* to that contracted vertex in T_{canon}. Note also that no edge incident to t_1 is contracted; we let the root of T_{canon} be the vertex corresponding to t_1. We proceed through all of the vertices of T_{canon} in turn using a tidy ancestral order, and at time τ we embed all of the vertices of T which correspond to the vertex τ of T_{canon} in one step.

Secondly, we must ensure that at each time τ it is possible to carry out this embedding. To do this, each time we embed a vertex $t \in T$ to a vertex $v \in G$, we use Lemma 2.8 to select sets A_t^+ and A_t^- of outneighbours and inneighbours of v in the clusters succeeding and preceding that of v, each of size at most $2\sqrt{m}$, which are reserved until all of the children of t have been embedded. Indeed, while these sets are reserved, no vertices may be embedded in them other than

(i) the children of t, and

(ii) those vertices of T which correspond to the same vertex of T_{canon} as a child of t.

We refer to these vertices as the *canonical children* of t; observe that there are at most Δ^2 such vertices. Since we proceed through the vertices of T_{canon} in a tidy ancestral order, this means that at any time τ not too many such sets are reserved, and so only a small proportion of the vertices of any cluster are reserved. When we later come to embed a child t' of t for which the edge tt' was allocated canonically, we embed t' in A_t^+ (if $t \to t'$) or A_t^- (if $t \leftarrow t'$) in such a way that we can choose $A_{t'}^+$ and $A_{t'}^-$ as desired.

When reading the next algorithm, one should bear in mind that often it is not apparent that a choice can be made as required by the algorithm. Indeed, if such a choice is not possible then the algorithm terminates with failure. Lemma 2.16 shows that under certain conditions on G, it is always possible to make such choices, and so we can be sure that the algorithm succeeds.

Vertex Embedding Algorithm

Input:

- A directed tree T rooted at t_1.
- A constant α and a positive integer m.
- A digraph G on vertex set $V = V_1 \cup \cdots \cup V_k$, where each V_i has size $(1+\alpha)m$, and a semi-canonical allocation of the vertices of T to the clusters V_i, with t_1 allocated to V_1.
- Finally, a vertex $v^* \in V_1$ to which t_1 should be embedded, and constants c and γ.

Initialisation: Form the canonical tree T_{canon} of T as explained above, and choose a tidy ancestral ordering $1, 2, \ldots, n'$ of the vertices of T_{canon}. Let $t_1, \ldots, t_n$ be a corresponding order of the vertices of T (so if $t_i \in T$ corresponds to $i \in T_{canon}$ and $t_j \in T$ corresponds to $j \in T_{canon}$ then t_i appears before t_j if and only if $i < j$.)

Procedure: At time τ we embed the vertices $t_r, \ldots, t_{r+s-1}$ of T corresponding to vertex τ of T_{canon}. Each vertex t_i is embedded to a vertex v_i of G, where $v_1 = v^*$.

Then, for each t_i we reserve sets $A^+_{t_i}$ and $A^-_{t_i}$ of vertices of G for the canonical children of t_i. To do this, at each time τ with $1 \leqslant \tau \leqslant n'$, take the following steps.

(1) We say that a vertex t_i of T is *open* at time τ if t_i has been embedded but some child of t_i has not yet been embedded. Define the set B^τ of vertices of G unavailable for use at time τ to consist of the vertices already occupied and the sets reserved for the canonical children of open vertices, so

$$B^\tau = \{v_1, \ldots, v_{r-1}\} \cup \bigcup_{t_i\,:\,t_i \text{ is open}} (A^+_{t_i} \cup A^-_{t_i}).$$

For each cluster V_j, let $V_j^\tau = V_j \setminus B^\tau$, so V_j^τ is the set of available vertices of V_j.

(2) If $\tau = 1$ embed t_1 to v_1. Alternatively, if $\tau > 1$:

(2.1) Precisely one of the vertices $t_r, \ldots, t_{r+s-1}$ of T corresponding to vertex τ of T_{canon} has a parent already embedded; we may assume this vertex is t_r. Let t_p be the already-embedded parent (so $p < r$, and when t_p was embedded sets $A^+_{t_p}$ and $A^-_{t_p}$ were chosen). Let V_j be the cluster to which t_p is embedded.

(2.2) If $t_p \to t_r$, choose a set S of $3s$ vertices of $A^+_{t_p} \subseteq V_{j+1}$ such that for each $v \in S$

$$|N^+(v) \cap V^\tau_{j+2}| \geqslant \gamma m \text{ and } |N^-(v) \cap V^\tau_j| \geqslant \gamma m.$$

If $t_p \leftarrow t_r$, choose a set S of $3s$ vertices of $A^-_{t_p} \subseteq V_{j-1}$ so for each $v \in S$

$$|N^+(v) \cap V^\tau_j| \geqslant \gamma m \text{ and } |N^-(v) \cap V^\tau_{j-2}| \geqslant \gamma m.$$

(2.3) Then choose a copy of $T[t_r, \ldots, t_{r+s-1}]$ in $G[S]$, and embed each vertex t_i to the corresponding vertex v_i in this copy.

(3) In step (2), we embedded each of $t_r, \dots, t_{r+s-1}$ in the same cluster; let V_q be this cluster. For each $r \leqslant i \leqslant r+s-1$, choose sets

$$A_{t_i}^+ \subseteq N^+(v_i) \cap V_{q+1}^\tau \text{ and } A_{t_i}^- \subseteq N^-(v_i) \cap V_{q-1}^\tau$$

such that the sets $A_{t_i}^+$ and $A_{t_i}^-$ are all pairwise disjoint, each $A_{t_i}^+$ and each $A_{t_i}^-$ is (c, γ)-good, and $|A_{t_i}^+|, |A_{t_i}^-| \leqslant 2\sqrt{m}$ for each i.

Whenever there are several choices (for example if there are several possibilities for S in (2.2)), take the lexicographically first of these. This ensures that for each input, the output is uniquely defined (*i.e.* we can view the algorithm as being deterministic).

Termination: If at any point it is not possible to make the choice required, terminate with failure. Otherwise, terminate after every vertex of T_{canon} has been processed, at which point $\psi(t_i) = v_i$ for each $t_i \in T$ is an embedding ψ of T into G, by construction.

Lemma 2.16 *Suppose that* $1/n \ll 1/\Delta, 1/k \ll \varepsilon \ll \gamma \ll c \ll d \ll \alpha \leqslant 2$, *and let* $m = n/k$.

(1) Let T *be a directed tree on at most* n *vertices with root* t_1 *and* $\Delta(T) \leqslant \Delta$.

(2) Let G *be an* ε*-regular* d*-dense cycle of cluster tournaments on clusters* $V_1, \dots, V_k$, *each of size* $(1+\alpha)m$, *and let* $v^* \in V_1$ *have at least* γm *inneighbours in* V_k *and at least* γm *outneighbours in* V_2.

(3) Let the vertices of T *be allocated to the clusters* $V_1, \dots, V_k$ *so that at most* $(1+\alpha/2)m$ *vertices are allocated to any one cluster* V_i, *and so that the allocation is semi-canonical.*

Then the Vertex Embedding Algorithm applied to T *and* G *(with this allocation and constants* c *and* γ*) successfully embeds* T *into* G *with* t_1 *embedded to* v^*.

Proof. The Vertex Embedding Algorithm only fails if at some point it is not possible to make the required choice. So to demonstrate that the algorithm succeeds, it is enough to show that it is always possible to make the required choices.

In the initialisation we are required to choose a tidy ancestral ordering of the vertices of the rooted tree T_{canon}; the existence of such a choice is guaranteed by Lemma 2.3. Now, consider the set of unavailable vertices B^τ at some time τ. Since the Vertex Embedding Algorithm embeds each vertex in the cluster to which it was allocated, we know that at most $(1+\alpha/2)m$ vertices of each V_j are already occupied. Furthermore, suppose that vertex t_i of T is open at time τ. Then t_i must correspond to a vertex $\tau' < \tau$ of T_{canon}, such that τ' has a child $\tau'' \geqslant \tau$. Since we are processing the vertices of T_{canon} in a tidy order, there can be at most $\log_2 n' \leqslant \log_2 n$ such vertices of T_{canon}. As each vertex of T_{canon} corresponds to at most Δ vertices of T, at most $\Delta \log_2 n$ vertices of T are open at time τ. Therefore, at any time τ, the total number of vertices in reserved sets $A^+_{t_i}$ and $A^-_{t_i}$ is at most $4\Delta\sqrt{m}\log_2 n \leqslant \alpha m/4$. So for any cluster V_j, at any time τ at most $(1+\alpha/2)m + \alpha m/4$ vertices of V_j are unavailable, and so $|V^\tau_j| \geqslant \alpha m/4$.

We can now demonstrate that it is possible to make the other choices that the algorithm asks for. Indeed, in step (2.2), if $t_p \to t_r$ with t_p embedded into V_j, then the algorithm has to choose a set S of $3s \leqslant 3\Delta$ vertices of $A^+_{t_p}$ such that each $v \in S$ has $|N^+(v) \cap V^\tau_{j+2}| \geqslant \gamma m$ and $|N^-(v) \cap V^\tau_j| \geqslant \gamma m$. But when $A^+_{t_p}$ was chosen at an earlier time τ', it was chosen to be (c,γ)-good. Since the vertex v_p to which t_p was embedded is in cluster V_j, $A^+_{t_p} \subseteq V_{j+1}$. Moreover, since $|V^\tau_j| \geqslant \alpha m/4 \geqslant (1+\alpha)cm$ and $|V^\tau_{j+2}| \geqslant \alpha m/4 \geqslant (1+\alpha)cm$, $A^+_{t_p}$ must contain at least $\gamma\sqrt{m}$ vertices v such that $|N^+(v) \cap V^\tau_{j+2}| \geqslant \gamma m$ and $|N^-(v) \cap V^\tau_j| \geqslant \gamma m$. Furthermore, since t_r is a child of t_p, t_p has been open since its embedding, and so only canonical children of t_p (of which there are at most Δ^2) can have been embedded in $A^+_{t_p}$. So it is indeed possible to select such a set S of $3s$ vertices as required. The argument for the case when $t_p \leftarrow t_r$ is similar. As for (2.3), observe that $G[S]$ is a tournament on $3s$ vertices, and that $T[t_r, \ldots, t_{r+s-1}]$ is a directed tree on s vertices. So by Theorem 1.2, $G[S]$ contains a copy of $T[t_r, \dots, t_{r+s-1}]$, so we may choose such a copy.

Finally we come to step (3). In this step we have just embedded at most Δ vertices $t_r, \dots, t_{r+s-1}$ in some cluster V_q, and we wish to choose sets $A^+_{t_i}$ and $A^-_{t_i}$ for each such vertex t_i. When embedding these vertices we ensured that for each i the vertex v_i to which t_i was embedded satisfied $|N^+(v_i) \cap V^\tau_{q+1}| \geqslant \gamma m$ (for $\tau = 1$ this holds instead by the condition on the outneighbours of $v^* = v_1$). So suppose we have chosen $A^+_{t_r}, A^+_{t_{r+1}}, \dots, A^+_{t_{r+\ell-1}}$ and we now wish to choose $A^+_{t_\ell}$. Then the previously chosen $A^+_{t_i}$ contain at most $2\Delta\sqrt{m}$ vertices between them, and so at least $3\gamma m/4 \geqslant (1+\alpha)\gamma m/2$ vertices of $N^+(v_\ell) \cap V^\tau_{q+1}$ have not been used in these previous sets. So by Lemma 2.8, we may choose a (c, γ)-good set $A^+_{t_\ell} \subseteq N^+(v_\ell) \cap V^\tau_{q+1}$ of size at most $2\sqrt{m}$ which is disjoint from all of the previously chosen $A^+_{t_i}$. Do this for each vertex t_i in turn; the choice of the sets $A^-_{t_i}$ is similar. □

We can now give the proof of the main lemmas of this section, beginning with the proof of Lemma 2.14.

Proof of Lemma 2.14. Apply the Vertex Allocation Algorithm to allocate the vertices of T to the clusters $V_1, \dots, V_k$. Then by Lemma 2.15(a) this allocation is semi-canonical, and by Lemma 2.15(c) at most $(1+\alpha/2)m$ vertices are allocated to each of the k clusters V_i. Next, apply the Vertex Embedding Algorithm to T and G, giving this allocation as input. By Lemma 2.16, this successfully embeds T in G with t_1 embedded to v^*. □

Proof of Lemma 2.13. If $\alpha > 2$ then G contains a copy of T by Theorem 1.2. So we may assume that $\alpha \leqslant 2$. We begin by introducing new constants $1/n \ll 1/M \ll 1/M' \ll \varepsilon \ll \varepsilon' \ll d \ll \mu$. Then by Lemma 2.10, G contains an ε-regular d-dense cycle of cluster tournaments G' on clusters $V_1, \dots, V_k$, where $M' \leqslant k \leqslant M$, and $|V_1| = \dots = |V_k| \geqslant (1-\varepsilon)(1+\alpha)n/k \geqslant$

$(1+\alpha/2)n/k$. For each i choose $V_i' \subseteq V_i$ of size $|V_i'| = (1+\alpha/2)n/k$ uniformly at random. By Lemma 2.9 we may fix an outcome of these choices so that $G'' = G'[V_1' \cup \cdots \cup V_k']$ is a ε'-regular $d/2$-dense cycle of cluster tournaments. So by Lemma 2.14 G'' contains a copy of T, so G contains T also. □

We finish this section with an analogous result to Lemma 2.14 for small directed trees (*i.e.* the result does not demand that $|T|$ is large compared to $|G|$).

Lemma 2.17 *Suppose that $1/m \ll 1/k, 1/\Delta \ll \varepsilon \ll d \ll \alpha \leqslant 2$, and that $1/k \ll \delta$. Let G be an ε-regular d-dense cycle of cluster tournaments on clusters $V_1, \dots, V_k$, each of size $(1+\alpha)m$, and let $v^* \in V_1$ have at least d^2m inneighbours in V_k and at least d^2m outneighbours in V_2. Let T be a directed tree on at most m vertices, rooted at t_1 and with $\Delta(T) \leqslant \Delta$, and let T^{far} be the subgraph of T induced by the vertices $x \in T$ with $d(t_1, x) \geqslant k^3$. Let $\mathcal{G}_T$ denote the set of copies of T in G for which the vertex t_1 of T corresponds to vertex v^* of G. Then $\mathcal{G}_T$ is non-empty. Furthermore, there exists a probability distribution on $\mathcal{G}_T$ such that if a member of $\mathcal{G}_T$ is selected at random according to this distribution, then for each i,*

$$\mathbb{E}(\#\text{ vertices of } T^{far} \text{ embedded in } V_i) \leqslant \frac{(1+\delta)|T^{far}|}{k}.$$

The probability distribution is actually, for each member of $\mathcal{G}_T$, the probability that applying first the Vertex Allocation Algorithm and then the Vertex Embedding Algorithm gives this copy of T in G (recall that actually the Vertex Embedding Algorithm is purely deterministic).

Proof. Apply the Vertex Allocation Algorithm to allocate the vertices of T to the clusters V_i. Since $|T| \leqslant m$, at most m vertices can be allocated to any cluster, and the allocation obtained is semi-canonical by Lemma 2.15(a). Next, introduce constants $\varepsilon \ll \gamma \ll c \ll d$, and apply the Vertex Embedding Algorithm to embed T in G. By Lemma 2.16, this successfully embeds

T in G, with t_1 embedded to v^*, and every vertex of T embedded in the cluster to which it was allocated. So it remains only to show that for each i, the expected number of vertices of T^{far} allocated to V_i is at most $(1 + \delta)|T^{far}|/k$. But since for any $x \in T^{far}$ we have $d(x, t_1) \geqslant k^3$, by Lemma 2.15(b) applied with $u = t_1$,

$$\mathbb{P}(x \text{ is allocated to } V_i) = \frac{(1 \pm \delta)}{k}$$

for each i, and the result follows. □

2.4 Embedding trees of unbounded maximum degree in a robust outexpander

2.4.1 Introduction

Having proved the desired result for directed trees of bounded maximum degree, we now move onto proving a similar result for directed trees with no such bound, with a constant of 2 rather than 1 in the condition on the order of G. This is the following lemma.

Lemma 2.18 *Suppose that $1/n \ll \mu \ll \nu \ll \eta \ll \alpha$, that G is a tournament on $2(1 + \alpha)n$ vertices which is a robust (μ, ν)-outexpander with $\delta^0(G) \geqslant \eta n$ and that T is a directed tree on n vertices. Then G contains a copy of T.*

To prove this, in Section 2.4.2 we define the core tree $T_\triangle$ of a tree T. This is a subtree of T which has bounded maximum degree, and the property that all components of $T - T_\triangle$ are small. Then in Section 2.4.4 we show that $T_\triangle$ can be extended to an 'extended tree' T_{ext} which also has bounded maximum degree, and also has the property that few vertices of T_{ext}

have neighbours outside T_{ext}. We embed the extended tree T_{ext} by a similar method to that of the previous section. We need to do this so that the small number of vertices of T_{ext} with neighbours outside T_{ext} are embedded to vertices of G with large in- and outdegree in G. In Section 2.4.5 we use our results from Section 2.3 to prove Lemma 2.27 on embedding trees of bounded maximum degree. This is similar to Lemma 2.14, but allows us also to demand that a small subset $H \subseteq V(T)$ of the vertices of T, satisfying certain conditions, should be embedded in a small subset U of the vertices of G. This allows us to embed T_{ext} in G in the desired manner. Finally, in Section 2.4.6 we complete the proof of Lemma 2.18 by first using Lemma 2.27 to embed T_{ext} into G as described and then embedding each component of $T - T_{ext}$ in the unoccupied vertices of G.

2.4.2 The core tree

Let T be a tree on n vertices, and let $\Delta \geqslant 2$ be fixed. Then we say that a vertex x of T is Δ-*core* if every edge e incident to x has $w_e(x) \leqslant (1 - 1/\Delta)n$. We call the subgraph of T induced by Δ-core vertices of T the *core tree of* T *with parameter* Δ, and denote it by T_Δ. With this definition, for any tree T, the core tree T_Δ is the same as the Δ-heart of T considered by Häggkvist and Thomason in [14]. The following proposition states the most important properties of the core tree. These properties are also noted in Section 3 of [14], but we include the proof for completeness.

Proposition 2.19 *Let T be a tree on n vertices and let $\Delta \geqslant 2$. Then:*

(i) T_Δ is a tree containing at least one vertex.

(ii) $w_e(x) \geqslant n/\Delta$ if $e = xy$ is an edge of T_Δ.

(iii) $\Delta(T_\Delta) \leqslant \Delta$.

(iv) Every component subtree T' of $T - T_\Delta$ has $|T'| \leqslant n/\Delta$.

Proof. For (i), note that since $\Delta \geqslant 2$, for any edge $e = uv$ of T at most one of $w_e(u) > (1 - 1/\Delta)n$ and $w_e(v) > (1 - 1/\Delta)n$ holds. Since T has more vertices than edges, there must therefore be some vertex $v \in T$ such that $w_e(v) \leqslant (1 - 1/\Delta)n$ for every edge e incident to v, and so $v \in T_\Delta$. It remains to show that T_Δ is connected. Observe that if u, v, w are distinct vertices of T such that there is an edge between u and v and an edge between v and w, then $w_{uv}(u) > w_{vw}(v)$. Now, suppose $x, y \in T_\Delta$, and let $x = v_1, v_2, \ldots, v_r = y$ be the vertices of the path from x to y in T (in order). Suppose for a contradiction that some v_i is not in T_Δ. Then for some neighbour z of v_i, $w_{v_i z}(v_i) > (1 - 1/\Delta)n$. If $z \neq v_{i+1}$, then for each $i \leqslant j \leqslant r-1$ we have $w_{v_j v_{j+1}}(v_{j+1}) > (1 - 1/\Delta)n$, and so $y \notin T_\Delta$, giving a contradiction. On the other hand, if $z = v_{i+1}$, then for each $2 \leqslant j \leqslant i$, $w_{v_{j-1} v_j}(v_{j-1}) > (1 - 1/\Delta)n$, and so $x \notin T_\Delta$, again giving a contradiction.

Now, (ii) is immediate from the fact that if $e = xy$ is an edge of T then $w_e(x) + w_e(y) = n$. Then (iii) follows directly from (ii), as the sum of $w_e(v)$ over all edges incident to v is $n - 1$.

Finally, for (iv), observe that for any such T' there is $u \in T'$, $v \in T_\Delta$ with $e = uv$ an edge of T. Suppose that $|T'| > n/\Delta$. Then $w_e(v) \geqslant |T'| > n/\Delta$, and so $w_e(u) \leqslant (1 - 1/\Delta)n$. But since $w_{e'}(u) < w_e(v) \leqslant (1 - 1/\Delta)n$ for every other edge e' incident to u, this means that $u \in T_\Delta$, giving a contradiction. $\square$

Note that T_Δ is an undirected tree obtained from an undirected tree T. However we often refer to the core tree of a directed tree T; this means the directed tree formed by taking the core tree T_Δ of the underlying graph T_{under} (an undirected tree) and directing each edge of T_Δ as it is directed in T.

The idea behind this definition is that the core tree is a bounded degree tree. The general technique we use to work with a tree T of unbounded maximum degree (in both this and later sections) is to first consider the core tree T_Δ, and then consider separately each component of

$T - T_\Delta$, making use of the fact that each such component is small.

The following proposition is needed in the proof of Lemma 2.21, which we shall use in the next chapter. Essentially the latter states that if trees T^1 and T^2 almost partition a tree T, then the core tree T_Δ is not much larger than $T^1_\Delta \cup T^2_\Delta$.

Proposition 2.20 *Let T be a tree on n vertices, let x be a leaf of T, and let $\Delta \geqslant 2$. Then $|(T-x)_\Delta| \geqslant |T_\Delta| - 1$.*

Proof. Let y be a vertex of $T_\Delta - (T-x)_\Delta$, and let z be an arbitrary vertex of $(T-x)_\Delta$. Then for some edge e incident to y we have $w_e(y) > (1-1/\Delta)(n-1)$ in $T-x$. Since by Proposition 2.19(iv) the component of $(T-x)-(T-x)_\Delta$ containing y contains at most $(n-1)/\Delta$ vertices, this edge must in fact be the first edge of the path in T from y to z. If e is also the first edge of the path in T from y to x then we have $w_e(y) > (1-1/\Delta)(n-1)+1 \geqslant (1-1/\Delta)n$ in T, and so $y \notin T_\Delta$, giving a contradiction. So y must lie on the path in T from x to z. Since $y \in T_\Delta$ we must have $w_e(y) \leqslant (1-1/\Delta)n$ in T, and so in T we have

$$(1-\frac{1}{\Delta})n - 1 \leqslant (1-\frac{1}{\Delta})(n-1) < w_e(y) \leqslant (1-\frac{1}{\Delta})n.$$

Clearly this can hold for at most one vertex y on the path from x to z. So $|T_\Delta - (T-x)_\Delta| \leqslant 1$, as desired. □

Lemma 2.21 *Let T be a tree on n vertices, let $\Delta \geqslant 2$ and let $\gamma, \alpha > 0$. Also let T^1 and T^2 be subtrees of T such that $|T^1 \cup T^2| \geqslant (1-\gamma)n$. Suppose also that $|T^1_\Delta|, |T^2_\Delta| \leqslant \alpha n$. Then $|T_\Delta| \leqslant \gamma n + 2\alpha n + 2n/\Delta$.*

Proof. Arbitrarily choose vertices $x_1 \in T^1_\Delta$ and $x_2 \in T^2_\Delta$, and let P be the path from x_1 to x_2 (so P is also a subtree of T). Then let $T^* := T^1 \cup P \cup T^2$, so $|T^*| \geqslant (1-\gamma)n$. Furthermore, T^* can be formed from T by repeated leaf-deletions. So by Proposition 2.20 we must have

$|T| - |T^*| \geqslant |T_\Delta| - |T^*_\Delta|$, and so

$$|T_\Delta| \leqslant |T| - |T^*| + |T^*_\Delta| \leqslant \gamma n - |P - (T^1 \cup T^2)| + |T^*_\Delta|. \tag{2.22}$$

Let $T^*_c := T^1_\Delta \cup P \cup T^2_\Delta$. We claim that $T^*_\Delta \subseteq T^*_c$. Indeed, suppose for a contradiction that there exists a vertex $y \in T^*_\Delta - T^*_c$. Since T^*_c is a subtree of T, every vertex of T^*_c lies in the same component C of $T^* - y$. Note that $T^* - C$ is a tree. Now, T^1_Δ and T^2_Δ are subtrees of C, so by Proposition 2.19(iv) $T^* - C$ contains at most $|T^1|/\Delta$ vertices of T^1 and at most $|T^2|/\Delta$ vertices of T^2. Let e be the edge of T^* between y and C. Then since $y \in T^*_\Delta$, $w_e(y) \leqslant (1 - 1/\Delta)|T^*|$ in T^*. So at least $|T^*|/\Delta$ vertices of T^* lie in components of $T^* - y$ other than C. As every vertex of P lies in C, either at least $|T^1|/\Delta$ vertices of T^1 lie in components of $T^* - y$ other than C, or at least $|T^2|/\Delta$ vertices of T^2 lie in components of $T^* - y$ other than C. In the former case this implies that $T^* - C$ contains more that $|T^1|/\Delta$ vertices of T^1, and in the latter case this implies that $T^* - C$ contains more that $|T^2|/\Delta$ vertices of T^2. In either case this yields a contradiction.

Now, $|T^*_c| \leqslant 2\alpha n + |P - (T^1_\Delta \cup T^2_\Delta)|$. Since $(P \cap T^1) - T^1_\Delta$ is contained within a single component of $T^1 - T^1_\Delta$, $|(P \cap T^1) - T^1_\Delta| \leqslant |T^1|/\Delta$, by Proposition 2.19(iv). Similarly $|(P \cap T^2) - T^2_\Delta| \leqslant |T^2|/\Delta$. So

$$|T^*_\Delta| \leqslant |T^*_c| \leqslant 2\alpha n + \frac{|T^1| + |T^2|}{\Delta} + |P - (T^1 \cup T^2)|.$$

So by (2.22)

$$|T_\Delta| \leqslant \gamma n + \frac{|T^1| + |T^2|}{\Delta} + 2\alpha n \leqslant \gamma n + \frac{2n}{\Delta} + 2\alpha n.$$

□

2.4.3 Leading paths

Let T be a tree on n vertices, rooted at t_1, let $H \subseteq V(T)$, and let k be a positive integer. For any vertex $x \in T$, there is a unique path in T from x to t_1; let P_x denote the set of the first k vertices of this path, starting from x. Let $H^1 = \bigcup_{x \in H} P_x$, and then for each $i \geqslant 1$ let H^{i+1} be formed from H^i by adding the vertices of P_x for any $x \in H^i$ with at least two children in H^i. After at most n steps we must have $H^i = H^{i+1}$, when we terminate the process. We refer to this final H^i as H *with leading paths included*, denoted $\mathcal{P}_k(H)$. So $H \subseteq \mathcal{P}_k(H) \subseteq V(T)$. Note that $\mathcal{P}_k(H)$ depends on both the value of k and the root t_1 of T.

Next we prove two results which enable us to make use of this definition. The first shows that if H is small then $\mathcal{P}_k(H)$ is small, and the second shows that if H is small then it is possible to embed any component T' of $T[\mathcal{P}_k(H)]$ in a regular and dense cycle of cluster tournaments such that the vertices of $V(T') \cap H$ are embedded in the first cluster and the 'root' of H is embedded in a given cluster.

Proposition 2.23 *Let k be any positive integer, let T be a tree on n vertices, rooted at some $t_1 \in T$, and let $H \subseteq V(T)$. Then $|\mathcal{P}_k(H)| \leqslant 3k|H|$.*

Proof. Consider any component T' of $T[\mathcal{P}_k(H)]$, and let t_1' be the unique vertex of T' with minimal $d(t_1, t_1')$. Then every vertex of T' lies on the path from some vertex of H to t_1, and so T' is precisely the set of vertices in paths between t_1' and vertices of $H \cap V(T')$. Thus only t_1' and vertices of H can be leaves of T'. It follows that $T[\mathcal{P}_k(H)]$ has at most $2|H|$ leaves. Since $T[\mathcal{P}_k(H)]$ is a forest, it follows that the number of vertices of $T[\mathcal{P}_k(H)]$ with at least two children in $T[\mathcal{P}_k(H)]$ is also at most $2|H|$. Furthermore, any vertex $x \in T$ for which the vertices of P_x were added to $\mathcal{P}_k(H)$ at any stage is either a member of H or has at least two children in $\mathcal{P}_k(H)$. This is true for at most $3|H|$ vertices x, and for each such vertex at most k vertices were added. $\square$

Lemma 2.24 *Suppose that $1/m \ll 1/k \ll \varepsilon \ll d$. Let T be a directed tree rooted at some $t_1 \in T$. Let $H \subseteq V(T)$ be of size $|H| \leqslant m/10k$, let T' be a component of $T[\mathcal{P}_k(H)]$ which does not contain t_1, and let t_1' be the unique vertex of T' with minimal $d(t_1', t_1)$. Let G be an ε-regular d-dense cycle of cluster tournaments on clusters $V_1, \dots, V_k$, each of size m. Then for any $j \in [k]$, G contains a copy of T' with the vertex t_1' corresponding to some vertex of V_j, and every vertex in $V(T') \cap H$ corresponding to some vertex of V_1.*

Proof. Informally, from the perspective of t_1', T' begins with a path of length $k-1$ (from t_1' to t, say) before possibly branching out. So we find a copy of T' in G by first embedding the vertices of this path so that t_1' is embedded in V_j and t is embedded in V_1. We then embed all of the remaining vertices of T' in V_1.

More formally, note that for each $0 \leqslant s \leqslant k-1$ there is precisely one vertex x_s of T' with $d(t_1', x_s) = s$ (so $x_0 = t_1'$, and $x_i \notin H$ for any $i < k-1$). Let $F \subseteq [k-1]$ be the set of those s such that $x_{s-1} \rightarrow x_s$, and let $B \subseteq [k-1]$ be the set of s such that $x_{s-1} \leftarrow x_s$. Then $|F| + |B| = k-1$, so either $|F| > k-j$ or $|B| \geqslant j-1$. Suppose first that $|B| \geqslant j-1$. Then choose $B' \subseteq B$ of size $j-1$. We allocate the vertices of T' to the clusters $V_1, \dots, V_j$. Begin by allocating x_0 to V_j. Then for each $s \in [k-1]$ in turn, let V_i be the cluster to which x_{s-1} was allocated, and allocate x_s to V_i if $s \notin B'$, or to V_{i-1} if $s \in B'$. Then since $|B'| = j-1$, x_{k-1} is assigned to V_1. Finally, allocate all other vertices of T' to V_1. Then every edge of T' is allocated either canonically or within a cluster.

Next we embed T' in G so that every vertex is embedded within the cluster to which it is allocated. To begin, by a standard regularity argument we may choose for each i a set $V_i' \subseteq V_i$ so that $|V_i'| \geqslant 9m/10$ and every vertex $v \in V_i'$ has at least $dm/2$ outneighbours in V_{i+1}'. Let $G' = G[V_1' \cup \dots \cup V_k']$. Now, for each i, let S_i be the set of vertices of T' allocated to V_i. So $|S_2|, \dots, |S_k| \leqslant k-1$ and $|S_1| \leqslant |T'|$. Then by Proposition 2.23, $3|S_1| \leqslant 3|T'| \leqslant 9k|H| \leqslant |V_1'|$. So by Theorem 1.2 we may embed $T'[S_1]$ in $G[V_1']$. Now suppose that we have successfully embedded $T'[S_1 \cup \cdots \cup S_{i-1}]$ in $G[V'_1 \cup \cdots \cup V'_{i-1}]$ for some $i \leqslant j$. Then precisely one vertex $t \in S_i$ has a neighbour $t' \in S_{i-1}$, and t' has already been embedded to some $v' \in V'_{i-1}$. Now v' has at least $dm/2 \geqslant 3|S_i|$ outneighbours in V'_i, and so by Theorem 1.2 we may embed $T'[S_i]$ among these outneighbours. Let v be the vertex to which t is embedded; then since v is an outneighbour of v', we have extended our embedding to an embedding of $T'[S_1 \cup \cdots \cup S_i]$ in $G[V'_1 \cup \cdots \cup V'_i]$. Continuing in this manner we obtain an embedding of T' in G, with t'_1 embedded in V_j and $V(T') \setminus \{x_0, \dots, x_{k-2}\} \supseteq V(T') \cap H$ embedded into V_1, as desired. A similar argument achieves this if $|F| > k - j$. $\square$

2.4.4 The extended tree

The next lemma combines the ideas of the core tree and leading paths to give the structure within a tree T which we use to prove Lemma 2.18. It shows that given a tree T we may extend the core tree T_Δ of T to an 'extended tree' T_{ext} which, like T_Δ, has bounded maximum degree (although this bound is now much larger than Δ). T_{ext} also has the property that only a small subset H of the vertices of T_{ext} have neighbours outside T_{ext}, and that few vertices of T_{ext} are close to a vertex of $\mathcal{P}_k(H)$.

Lemma 2.25 *Suppose that $1/n, 1/\Delta^* \ll 1/\Delta, 1/k, \omega \ll 1$. Let T be a tree on n vertices, and choose any vertex $t_1 \in T_\Delta$ as the root of T. Then there exists a subtree T_{ext} of T and a subset $H \subseteq V(T_{ext})$ which satisfy the following properties.*

(i) $T_\Delta \subseteq T_{ext}$.

(ii) $\Delta(T_{ext}) \leqslant \Delta^*$.

(iii) For any edge e between $T - T_{ext}$ and T_{ext}, the endvertex of e in T_{ext} lies in H.

(iv) The number of vertices $v \in T_{ext}$ which satisfy $1 \leqslant d(v, \mathcal{P}_k(H)) \leqslant k^3$ is at most ωn.

(v) $|H| \leqslant \frac{n}{\Delta^{k^{1/\omega}}}$.

Proof. We consider the subgraph $T - E(T_\Delta)$ of T obtained by deleting the edges (but not the vertices) of T_Δ from T. Each vertex $v \in T_\Delta$ lies in a separate component of $T - E(T_\Delta)$; we denote the component containing v by T_v. Then the trees T_Δ and $\{T_v : v \in T_\Delta\}$ partition the edges of T, and the trees $\{T_v : v \in T_\Delta\}$ partition the vertices of T.

We say that a vertex $v \in T_\Delta$ is *i-heavy* if $|T_v| \geqslant \Delta_i := \Delta^{k^i}$. For any integer i, let H_i denote the set of i-heavy vertices in T_Δ. So $|H_i| \leqslant n/\Delta_i$, and so by Proposition 2.23 we have $|\mathcal{P}_k(H_i)| \leqslant 3kn/\Delta_i$ for each i. We wish to choose a large integer t so that few vertices of T lie in trees T_v for which v is not in H_t but is close to a member of $\mathcal{P}_k(H_t)$. The next claim shows that this is possible.

Claim. For some natural number $1/\omega \leqslant t \leqslant 3/\omega$ we have

$$\Big| \bigcup_{\substack{v \in V(T_\Delta) \setminus H_t \\ d(v, \mathcal{P}_k(H_t)) \leqslant k^3}} T_v \Big| \leqslant \omega n. \tag{2.26}$$

To prove the claim, observe that for each integer i with $1/\omega \leqslant i \leqslant 3/\omega$, if $v \in V(T_\Delta) \setminus H_{i-1}$ then $|T_v| < \Delta_{i-1}$, and so

$$\Big| \bigcup_{\substack{v \in V(T_\Delta) \setminus H_{i-1} \\ d(v, \mathcal{P}_k(H_i)) \leqslant k^3}} T_v \Big| < |\mathcal{P}_k(H_i)| \Delta^{k^3+1} \Delta_{i-1} \leqslant \frac{3k\Delta^{k^3+1}\Delta^{k^{i-1}} n}{\Delta^{k^i}} \leqslant \frac{3kn}{\Delta^{k^i/2}} \leqslant \frac{\omega n}{3}.$$

Now let

$$B_i := \bigcup_{\substack{v \in H_{i-1} \setminus H_i \\ d(v, \mathcal{P}_k(H_i)) \leqslant k^3}} T_v.$$

Then the sets B_i are pairwise disjoint subsets of $V(T)$. If the claim is false, then $|B_i| > 2\omega n/3$ for every integer i with $1/\omega \leqslant i \leqslant 3/\omega$, and so $|\bigcup_{1/\omega \leqslant i \leqslant 3/\omega} B_i| > n$, giving a contradiction. This completes the proof of the claim.

Fix such a value of t, and let $H = H_t$. We define the extended tree T_{ext} by $T_{ext} := T_\Delta \cup \bigcup_{v \in V(T_\Delta) \setminus H} T_v$. Then T_{ext} is a subtree of T with $T_\Delta \subseteq T_{ext}$, so (i) is satisfied. Since $H \subseteq V(T_\Delta)$, we have $H \subseteq V(T_{ext})$ as desired. Also (ii) holds since any vertex $u \in T_{ext}$ has at most Δ neighbours in T_Δ and at most Δ_t neighbours in the single tree T_v with $v \in T_\Delta$ which contains u. So $\Delta(T_{ext}) \leqslant \Delta + \Delta_t \leqslant \Delta + \Delta^{k^{3/\omega}} \leqslant \Delta^*$. For (iii), observe that if $u \notin T_{ext}$, then u must lie in some T_v with $v \in H$. But then if u has a neighbour in T_{ext} this neighbour must be v. For (iv), consider any $u \in T_{ext}$ satisfying $1 \leqslant d(u, \mathcal{P}_k(H)) \leqslant k^3$. Since $d(u, \mathcal{P}_k(H)) \geqslant 1$ we know that $u \notin H_t$, so if $u \in T_\Delta$ then u is counted in (2.26). If $u \notin T_\Delta$ then there exists v such that $u \in T_v$ and $v \in V(T_\Delta) \setminus H$. Note that $\mathcal{P}_k(H) \subseteq V(T_\Delta)$ (since $t_1 \in T_\Delta$). This in turn implies that $d(v, \mathcal{P}_k(H)) < d(u, \mathcal{P}_k(H)) \leqslant k^3$. So u is also counted in (2.26). Finally, for (v), recall that $|H| \leqslant n/\Delta_t \leqslant n/\Delta^{k^{1/\omega}}$. □

2.4.5 Embedding trees of bounded maximum degree with restrictions

In this section we prove the following lemma, which is similar to Lemma 2.14, but which allows us to restrict some vertices of T to a subset of $V(G)$.

Lemma 2.27 *Suppose that $1/n \ll 1/\Delta, 1/k \ll \varepsilon \ll d \ll \alpha, \lambda \leqslant 1/2$, that $m = n/k$, that $\lambda \leqslant \alpha/4$ and that $\delta = d\lambda/8k$. Let T be a directed tree on n vertices rooted at t_1 and with $\Delta(T) \leqslant \Delta$. Let $H \subseteq V(T)$ be such that $|H| \leqslant \delta n/7k$ and $|\{x \in T \colon 1 \leqslant d(x, \mathcal{P}_k(H)) \leqslant k^3\}| \leqslant \delta n$. Let G be an ε-regular d-dense cycle of cluster tournaments on clusters $V_1, \dots, V_k$, each of size $(1+\alpha)m$, and let $U \subseteq V_1 \cup \dots \cup V_k$ have size $|U| \geqslant \lambda n$. Then T can be embedded in G so that each vertex $t \in H$ is embedded to some $u \in U$.*

Proof. We may assume without loss of generality that $|U \cap V_1| \geqslant \lambda m$. If $t_1 \notin H$, then add t_1 to H, so now we have $|H| \leqslant \delta n/6k$. Moreover, the new $\mathcal{P}_k(H)$ is the union of the old $\mathcal{P}_k(H)$ and $\{t_1\}$. So now

$$|\{x \in T\colon 1 \leqslant d(x, \mathcal{P}_k(H)) \leqslant k^3\}| \leqslant \delta n + \Delta^{k^3+1} \leqslant \frac{3\delta n}{2}. \tag{2.28}$$

Also, introduce a new constant ε' with $\varepsilon \ll \varepsilon' \ll d$. To begin, for each i choose disjoint sets $X_i, Y_i \subseteq V_i$ such that

- $|X_i| = (1 + \alpha/2)m$ and $|Y_i| = 3\lambda m/4 \leqslant \alpha m/4$,
- every vertex of $X_i \cup Y_i$ has at least $d\lambda m/2$ inneighbours in Y_{i-1} and at least $d\lambda m/2$ outneighbours in Y_{i+1}, and
- $Y_1 \subseteq U \cap V_1$.

The existence of such sets can be shown by a standard regularity argument. Indeed, choose disjoint sets $X_i', Y_i' \subseteq V_i$ such that $|X_i'| = (1 + \alpha/2 + d^2)m$, $|Y_i'| = (3\lambda/4 + d^2)m$ and $Y_1' \subseteq U \cap V_1$. Then both $G[Y_{i-1}' \to X_i' \cup Y_i']$ and $G[X_i' \cup Y_i' \to Y_{i+1}']$ are $2\varepsilon/\lambda$-regular with density at least $3d/4$. So all but at most $9\varepsilon m/\lambda \leqslant d^2 m$ vertices in $X_i' \cup Y_i'$ have at least $9d\lambda m/16$ inneighbours in Y_{i-1}' and at least $9d\lambda m/16$ outneighbours in Y_{i+1}'. Delete $d^2 m$ vertices from X_i' and $d^2 m$ vertices from Y_i' including these $d^2 m$ vertices of small degree (for each $i \in [k]$). Then the sets X_i and Y_i thus obtained from X_i' and Y_i' are as desired.

Each vertex of $\mathcal{P}_k(H)$ and every child of any such vertex is to be embedded in the sets Y_i, whilst the remaining vertices of T are to be embedded in the sets X_i. Observe that by Proposition 2.23, $|\mathcal{P}_k(H)| \leqslant 3k|H| \leqslant \delta n/2$. Moreover, (2.28) implies that there are at most $3\delta n/2$ children of vertices of $\mathcal{P}_k(H)$ outside $\mathcal{P}_k(H)$. So at most $2\delta n = d\lambda m/4$ vertices are to be embedded in the sets Y_i.

Next, let $T_1, \ldots, T_r$ be the component subtrees of $T[\mathcal{P}_k(H)]$ and $T - \mathcal{P}_k(H)$. So each vertex of T lies in precisely one of the T_i. Let T^{con} be the directed tree obtained by contracting each T_i to a single vertex i. We may assume the T_i were labelled so that $t_1 \in T_1$ and $1, 2, \ldots, r$ is an ancestral order of the vertices of T^{con}. Then let

$$\begin{aligned}
J =& \{i : T_i \text{ is a component subtree of } T[\mathcal{P}_k(H)]\}, \\
L =& \{i : T_i \text{ is a component subtree of } T - \mathcal{P}_k(H) \text{ and } |T_i| \geqslant \sqrt{n}\}, \\
Q =& \{i : T_i \text{ is a component subtree of } T - \mathcal{P}_k(H) \text{ and } |T_i| < \sqrt{n}\}.
\end{aligned}$$

Note that each vertex of H lies in some T_i such that $i \in J$. For each $i > 1$, T_i contains precisely one vertex with a neighbour in some T_j with $j < i$. (Furthermore, if $i \in L \cup Q$ then this j must belong to J.) Let t_i be this vertex, then the children of vertices of $\mathcal{P}_k(H)$ which are not in $\mathcal{P}_k(H)$ are precisely the vertices t_i for $i \in L \cup Q$. For each i let T_i^{far} be the set of vertices $x \in T_i$ with $d(t_i, x) \geqslant k^3$. Then

$$\sum_{i \in L \cup Q} |V(T_i) \setminus T_i^{far}| \leqslant \frac{3\delta n}{2} \tag{2.29}$$

by (2.28). Finally, for each i let $T_i^{\leqslant} = T[V(T_1) \cup \cdots \cup V(T_i)]$, so $T_i^{\leqslant}$ is the graph formed from the union of $T_1, \ldots, T_i$ by also adding the edges between $T_1, \ldots, T_i$.

We use a randomised algorithm to embed the vertices of T in G. At each time τ this algorithm embeds the vertices of T_τ. Indeed, if $\tau \in J$, we use Lemma 2.24 to embed T_τ in the sets Y_i so that the vertices of $H \cap V(T_\tau)$ are embedded in $Y_1 \subseteq U$. If $\tau \in L$, we use Lemma 2.14 to embed T_τ in the sets X_i (except for the vertex t_τ, which is embedded in some Y_i) so that approximately equally many vertices of T_τ are embedded in each set X_i. Finally, if $\tau \in Q$ we use Lemma 2.17 to randomly embed T_τ in the sets X_i (again with the exception of the vertex t_τ, which is embedded in some Y_i) so that the expected number of vertices of T_τ^{far} embedded in each set X_i is approximately equal. Together the embeddings of each T_i in G form an embedding of T in G such that every vertex of H is embedded in U, as desired. At any time τ we can choose the desired embedding of T_τ unless there are insufficient vertices remaining unoccupied in one of the sets X_i. We show that this is unlikely to happen for any i, so with positive probability the algorithm finds a copy of T in G, proving the lemma.

Tree Embedding Algorithm.

At time $\tau = 1$, we wish to embed T_1 in G. Recall that we ensured that $t_1 \in H$, so $1 \in J$. We embed T_1 in Y_1. Indeed, $|Y_1| = 3\lambda m/4$, and $|T_1| \leqslant |\mathcal{P}_k(H)| \leqslant \delta n/2 = d\lambda m/16$, and so Y_1 contains a copy of T_1 by Theorem 1.2. Choose (deterministically) such a copy, and embed each vertex of T_1 to the corresponding vertex in this copy.

So after completing the first step, the algorithm has obtained an embedding of $T_1 = T_1^{\leqslant}$ in G such that any vertex of $H \cap V(T_1^{\leqslant})$ is embedded in Y_1, and only vertices of $\mathcal{P}_k(H)$ and their children have been embedded in the sets Y_i.

At a given time $\tau > 1$ we may therefore suppose that the algorithm has found an embedding of $T_{\tau-1}^{\leqslant}$ in G so that each vertex of $H \cap V(T_{\tau-1}^{\leqslant})$ is embedded in Y_1, and only vertices of $\mathcal{P}_k(H)$ and their children have been embedded in the sets Y_i. (Recall that this implies that at most $d\lambda m/4$ vertices are embedded in the sets Y_i.) We wish to extend this embedding to include T_τ, and we do this by the following steps.

- For each i let X_i^τ and Y_i^τ consist of the unoccupied vertices of X_i and Y_i respectively. If $|X_i^\tau| < |T_\tau|/k + \alpha m/4$ for some i, then terminate the algorithm with failure. So we may assume that $|X_i^\tau| \geqslant |T_\tau|/k + \alpha m/4$ for each i. Also, since at most $d\lambda m/4$ vertices have been embedded in the sets Y_i, every vertex of $X_i \cup Y_i$ must have at least $d\lambda m/4$ inneighbours in Y_{i-1}^τ and at least $d\lambda m/4$ outneighbours in Y_{i+1}^τ.

- By definition, t_τ is the unique vertex of T_τ with a neighbour which has already been embedded. Let t'_τ be this neighbour, and let v'_τ be the vertex to which t'_τ was embedded. Also let V_j be the cluster into which t_τ should be embedded so that the edge between t_τ and t'_τ is embedded canonically. Then v'_τ has at least $d\lambda m/4$ neighbours in Y_j^τ, and so by a standard regularity argument, we may choose some such neighbour $v_\tau \in Y_j^\tau$ which has at least $\alpha dm/8$ outneighbours in X_{j+1}^τ and at least $\alpha dm/8$ inneighbours in X_{j-1}^τ.

- Now, if $\tau \in L$, for each i consider a set $Z_i^\tau \subseteq X_i^\tau$ of size $(1+\alpha/8)|T_\tau|/k$ chosen uniformly at random and independently of all other choices. We can do this since $(1+\alpha/8)|T_\tau|/k \leqslant |T_\tau|/k + \alpha m/8 \leqslant |X_i^\tau|$ for each $i \in [k]$. Then since $G[X_1^\tau \cup \dots X_k^\tau]$ is an $(16\varepsilon/\alpha)$-regular $d/2$-dense cycle of cluster tournaments, by Lemma 2.9 $G[Z_1^\tau, \dots, Z_k^\tau]$ is an ε'-regular $d/4$-dense cycle of cluster tournaments with probability $1 - o(1)$. Also with probability $1 - o(1)$, v_τ has at least $\alpha d|T_\tau|/16k$ outneighbours in Z_{j+1}^τ and at least $\alpha d|T_\tau|/16k$ inneighbours in Z_{j-1}^τ. So we may choose (deterministically) sets Z_i^τ satisfying these two properties. Now delete a single vertex (chosen arbitrarily) from Z_j^τ, and replace it by v_τ, and let G^τ be the restriction of G to the new $Z_1^\tau, \dots, Z_k^\tau$. Then G^τ is a $(2\varepsilon')$-regular $(d/8)$-dense cycle of cluster tournaments with clusters of size $(1+\alpha/8)|T_\tau|/k$. So by Lemma 2.14 G^τ contains a copy of T_τ with at most $(1+\alpha/8)|T_\tau|/k$ vertices of T_τ embedded in each X_i, and with t_τ embedded to v_τ. Embed each vertex of T_τ to the corresponding vertex in this copy.

- If instead $\tau \in Q$, then arbitrarily choose $Z_j^\tau \subseteq X_j^\tau \cup \{v_\tau\}$ of size $\alpha m/8$ with $v_\tau \in Z_j^\tau$, and for each $i \neq j$ choose $Z_i^\tau \subseteq X_i^\tau$ of size $\alpha m/8$ uniformly at random and independently of all other choices. Then $G^\tau := G[Z_1^\tau, \dots, Z_k^\tau]$ is a $(16\varepsilon/\alpha)$-regular $d/2$-dense cycle of cluster tournaments. Also, with probability $1 - o(1)$, v_τ has at least $\alpha^2 dm/128$ outneighbours in Z_{j+1}^τ and at least $\alpha^2 dm/128$ inneighbours in Z_{j-1}^τ, so we may fix (deterministically) our choices of the Z_i^τ such that this event holds. Then by Lemma 2.17 the set of copies of T_τ in G^τ such that t_τ is embedded to v_τ is non-empty, and furthermore there exists a probability distribution on this set so that if a copy is chosen according to this distribution, then the expected number of vertices of T_τ^{far} embedded in each Z_i^τ is at most $(1+\sqrt{\varepsilon})|T_\tau^{far}|/k$. Choose (deterministically) such a distribution, and choose randomly such a copy according to this distribution. Embed each vertex of T_τ to the corresponding vertex in this copy.

- Finally, if $\tau \in J$, then since v'_τ has at least $d\lambda m/4$ neighbours in Y_j^τ, we may choose sets $Z_1^\tau \subseteq Y_1^\tau, \dots, Z_k^\tau \subseteq Y_k^\tau$, each of size $d\lambda m/4$, so that every vertex of Z_j^τ is a neighbour of v'_τ. Let G^τ be the restriction of G to the sets Z_i^τ; then G^τ is a $(8\varepsilon/d\lambda)$-regular $(d/2)$-dense cycle of cluster tournaments. Since $|H| \leqslant \delta n/6k = d\lambda m/48k$, by Lemma 2.24, G^τ contains a copy of T_τ, with vertex t_τ embedded in Y_j^τ, and with every vertex of $H \cap V(T_\tau)$ corresponding to a vertex of Y_1^τ. Embed each vertex of T_τ to the corresponding vertex in this copy.

- In either case, we have extended the embedding of $T_{\tau-1}^{\leqslant}$ in G to an embedding of $T_\tau^{\leqslant}$ in G, such that every vertex of $H \cap V(T_\tau^{\leqslant})$ is embedded in $Y_1 \subseteq U$, and only vertices of $\mathcal{P}_k(H)$ and their children have been embedded in the sets Y_i.

Since $T_r^{\leqslant} = T$, if the algorithm does not terminate with failure then at time r, after embedding T_r it has obtained an embedding of T in G so that every vertex of H is embedded in U, as desired. At this point the algorithm terminates with success.

It remains to show that with positive probability this algorithm does not terminate with failure before embedding T_r. Suppose first that $\sum_{j \in Q} |T_j| < \alpha m/8$. Then for any $i \in [k]$ and at any time τ, the number of vertices embedded in X_i is at most

$$\frac{1+\alpha/8}{k} \sum_{\substack{j \in L \\ j < \tau}} |T_j| + \sum_{\substack{j \in Q \\ j < \tau}} |T_j| \leqslant \frac{(1+\alpha/8)(n-|T_\tau|)}{k} + \frac{\alpha m}{8} < \left(1+\frac{\alpha}{4}\right) m - \frac{|T_\tau|}{k}$$

and so $|X_i^\tau| \geqslant |T_\tau|/k + \alpha m/4$. Therefore the algorithm cannot terminate with failure at any point. So we may assume that $\sum_{j \in Q} |T_j| \geqslant \alpha m/8$.

Let OUT be the set of all possible courses of the algorithm until termination. Since the only random choices made by the algorithm are the choices of where to embed the T_i for each $i \in Q$, any possible course of the algorithm $\mathcal{C} \in OUT$ can be uniquely described by the embeddings f_i of T_i into G for each $i \in Q$ such that the algorithm does not terminate before embedding T_i. So we may define a probability space with sample space OUT where for any $\mathcal{C} \in OUT$, $\mathbb{P}(\mathcal{C})$ is defined to be the probability that the algorithm takes course $\mathcal{C}$. So

$$\mathbb{P}(\mathcal{C}) = \prod_{j \in Q} \mathbb{P}(F_j \mid F_i \colon i < j, i \in Q).$$

where F_j denotes the event that f_j is the embedding of T_j into G, if T_j is embedded at some point during $\mathcal{C}$, and is taken to be true otherwise.

Now, we define the random variable W_j^i in this probability space as follows. For any $\mathcal{C} \in OUT, j \in Q$ and $i \in [k]$, let

$$W_j^i(\mathcal{C}) = \begin{cases} \frac{\text{\# of vertices from } T_j^{far} \text{ embedded in } X_i}{\sqrt{n}} & \text{if } T_j \text{ is embedded during } \mathcal{C}, \\ \frac{|T_j^{far}|}{k\sqrt{n}} & \text{otherwise.} \end{cases}$$

Since $|T_j^{far}| \leqslant |T_j| < \sqrt{n}$ for each $j \in Q$, W_j^i is a well-defined function from OUT to $[0,1]$, and so is a well-defined random variable in our probability space.

For any $j \in Q$ and $\mathcal{C}_a, \mathcal{C}_b \in OUT$, let $\mathcal{C}_a \sim_j \mathcal{C}_b$ if and only if $\mathcal{C}_a$ and $\mathcal{C}_b$ share the same course before time j (*i.e.* they embed $T_1, \ldots, T_{j-1}$ identically) or T_j is not embedded at any point in either $\mathcal{C}_a$ or $\mathcal{C}_b$. Then $\sim_j$ is an equivalence relation on OUT. For any equivalence class $\mathcal{C}^*$ of $\sim_j$ other than the class of $\mathcal{C}$ for which T_j is not embedded, every $\mathcal{C} \in \mathcal{C}^*$ shares the same course before time j. So for each $\mathcal{C} \in \mathcal{C}^*$, the same probability distribution on the set of copies of

T_j was chosen at time j, and a copy was then chosen according to this distribution. So further partition $\mathcal{C}^*$ into $\mathcal{C}_1^*, \dots, \mathcal{C}_a^*$ by this choice, so courses $\mathcal{C}, \mathcal{C}' \in \mathcal{C}^*$ are in the same $\mathcal{C}_s^*$ if and only if T_j is embedded identically in $\mathcal{C}$ and $\mathcal{C}'$. Now

$$\mathbb{E}(W_j^i \mid \mathcal{C}^*) = \sum_s \mathbb{E}(W_j^i \mid \mathcal{C}_s^*)\mathbb{P}(\mathcal{C}_s^* \mid \mathcal{C}^*),$$

but every member of $\mathcal{C}_s^*$ embeds T_j identically, so $\mathbb{E}(W_j^i \mid \mathcal{C}_s^*)$ is simply the number of vertices of T_j^{far} embedded in X_i in this common embedding, divided by $\sqrt{n}$. Also, $\mathbb{P}(\mathcal{C}_s^* \mid \mathcal{C}^*)$ is the probability that this embedding of T_j is chosen when the random choice of the embedding of T_j is made. So by our (deterministic) choice of the probability distribution on the copies of T_j in G,

$$\mathbb{E}(W_j^i \mid \mathcal{C}^*) \leqslant \frac{(1+\sqrt{\varepsilon})|T_j^{far}|}{k\sqrt{n}}. \tag{2.30}$$

If instead $\mathcal{C}^*$ is the class of all $\mathcal{C}$ such that T_j is not embedded in $\mathcal{C}$, then $\mathbb{E}(W_j^i \mid \mathcal{C}^*) = |T_j^{far}|/k\sqrt{n}$ by definition, and so (2.30) holds in this case also.

Now, for any equivalence class $\mathcal{C}^*$ other than the class in which T_j is not embedded, the embeddings of $T_1, \dots, T_{j-1}$ are identical amongst the members of $\mathcal{C}^*$, and so

$$\mathbb{E}(W_j^i \mid \mathcal{C}^*, W_s^i \colon s \in Q, s < j) = \mathbb{E}(W_j^i \mid \mathcal{C}^*).$$

Clearly this equality also holds for the class $\mathcal{C}^*$ in which T_j is not embedded, and so for any $i \in [k]$,

$$\mathbb{E}(W_j^i \mid W_s^i \colon s \in Q, s < j) \leqslant \max_{\mathcal{C}^*} \mathbb{E}(W_j^i \mid \mathcal{C}^*, W_s^i \colon s \in Q, s < j) \leqslant \frac{(1+\sqrt{\varepsilon})|T_j^{far}|}{k\sqrt{n}}.$$

Since $\sum_{j\in Q}|T_j| \geqslant \alpha m/8$, by Lemma 2.5, for any i the probability that

$$\sum_{j\in Q} W_j^i \leqslant \frac{(1+\alpha/8)\sum_{j\in Q}|T_j|}{k\sqrt{n}} \tag{2.31}$$

does not hold decreases exponentially with n. So with probability $1-o(1)$, (2.31) holds for each $i \in [k]$.

To finish the proof, we show that if (2.31) holds for each $i \in [k]$, then the algorithm cannot terminate with failure, and therefore successfully embeds T in G as desired. Indeed, the algorithm only terminates with failure if at some time τ we have $|X_i^\tau| < |T_\tau|/k + \alpha m/4$ for some i. But for any $i \in [k]$ and any time τ, only vertices from subtrees T_s such that $s \in L \cup Q$ and $s < \tau$ have been embedded in X_i before time τ. So the number of vertices embedded in X_i before time τ is at most

$$\begin{aligned}
&\frac{(1+\alpha/8)}{k}\sum_{s\in L\setminus\{\tau\}}|T_s| + \sum_{s\in Q\setminus\{\tau\}}|V(T_s)\setminus T_s^{far}| + \sqrt{n}\sum_{s\in Q\setminus\{\tau\}} W_s^i \\
\overset{(2.29)}{\leqslant}\, &\frac{(1+\alpha/8)}{k}\sum_{s\in L}|T_s| + \frac{3\delta n}{2} + \sqrt{n}\sum_{s\in Q} W_s^i + \frac{\delta n}{2} - \frac{|T_\tau|}{k} \\
\overset{(2.31)}{\leqslant}\, &\frac{(1+\alpha/8)}{k}\sum_{s\in L\cup Q}|T_s| + 2\delta n - \frac{|T_\tau|}{k} \leqslant (1+\frac{\alpha}{4})m - \frac{|T_\tau|}{k}.
\end{aligned}$$

To see that the second line holds, note that $|T_\tau|/k < \sqrt{n}/k < \delta n/2$ whenever $\tau \in Q$ and $|T_\tau|/k < |\mathcal{P}_k(H)| < \delta n/2$ whenever $\tau \in J$. So if (2.31) holds, then at any time τ and for any $i \in [k]$, $|X_i^\tau| \geqslant |T_\tau|/k + \alpha m/4$, and so the algorithm succeeds. This completes the proof of Lemma 2.27. $\square$

2.4.6 Proof of Lemma 2.18

To prove Lemma 2.18, we first prove a stronger result for directed trees T whose core tree T_Δ is large. This is the next lemma, which is also used in Chapter 3.

Lemma 2.32 *Suppose that* $1/n \ll 1/\Delta \ll \mu \ll \nu \ll \eta \ll \gamma \ll \beta \ll 1$. *Let* T *be a directed tree on* n *vertices such that* $|T_\Delta| \geqslant \beta n$*, and let* G *be a robust* (μ, ν)*-outexpander tournament on at least* $(2-\gamma)n$ *vertices, with* $\delta^0(G) \geqslant \eta|G|$. *Then* G *contains a copy of* T.

To prove Lemma 2.32, we apply Lemma 2.25 to find a subtree T_{ext} of T and a subset $H \subseteq V(T_{ext})$. Then we find a cluster cycle $\mathcal{C}$ in G such that $|\mathcal{C}|$ is slightly larger than $|T_{ext}|$. We then embed T_{ext} into $\mathcal{C}$ using Lemma 2.27, restricting H to a set U of vertices of $\mathcal{C}$ which have many inneighbours and outneighbours outside $\mathcal{C}$. Finally we use this property of U to embed the vertices of $T - T_{ext}$ in $V(G) \setminus V(\mathcal{C})$ and thereby complete the embedding.

Proof. We begin by introducing new constants $\Delta^*, M, M', \varepsilon, d$ and α which satisfy

$$\frac{1}{n} \ll \frac{1}{\Delta^*} \ll \frac{1}{M} \ll \frac{1}{M'}, \frac{1}{\Delta} \ll \varepsilon \ll d \ll \mu \ll \nu \ll \eta \ll \gamma \ll \alpha \ll \beta \ll 1.$$

If $|G| \geqslant 3n$, then G contains a copy of T by Theorem 1.2. So we may assume that $|G| < 3n$. Now, since G is a robust (μ, ν)-outexpander with $\delta^0(G) \geqslant \eta|G|$, Lemma 2.10 implies that G contains an ε-regular d-dense cycle of cluster tournaments on clusters $V_1, \dots, V_k$ each of equal size between $(1-\varepsilon)|G|/k \geqslant (1-\varepsilon)(2-\gamma)m \geqslant 2(1-\gamma)m$ and $|G|/k \leqslant 3m$, where $m := n/k$ and $M' \leqslant k \leqslant M$. So we may remove vertices from each V_i to obtain a 2ε-regular $(d/2)$-dense cycle of cluster tournaments G' on clusters $V_1', \dots, V_k'$ each of size $2(1-\gamma)m$. So $|G'| = 2(1-\gamma)n$. Let

$$\delta := \frac{d\alpha\beta}{160k}.$$

Choose any vertex $t_1 \in T_\Delta$ as the root of T. Then by Lemma 2.25, we may choose a subtree T_{ext} of T and a subset $H \subseteq V(T_{ext})$ satisfying the following properties.

(i) $T_\Delta \subseteq T_{ext}$.

(ii) $\Delta(T_{ext}) \leqslant \Delta^*$.

(iii) For any edge e between $T - T_{ext}$ and T_{ext}, the endvertex of e in T_{ext} lies in H.

(iv) The number of vertices $v \in T_{ext}$ which satisfy $1 \leqslant d(v, \mathcal{P}_k(H)) \leqslant k^3$ is at most $\delta\beta n$.

(v) $|H| \leqslant n/\Delta^{k^{1/\delta\beta}} \leqslant \delta\beta n/7k$.

Let $T_1^+, \dots, T_r^+$ be the outcomponents of T_{ext} in T and let $T_1^-, \dots, T_s^-$ be the incomponents of T_{ext} in T. Then for each i let v_i^+ be the vertex of T_i^+ with an inneighbour in T_{ext} and let v_i^- be the vertex of T_i^- with an outneighbour in T_{ext}. By (i) and Proposition 2.19(iv) each T_i^+ and each T_i^- contains at most n/Δ vertices. Let $x = |T_{ext}|$, let $y = |T_1^+ \cup \dots \cup T_r^+|$ and let $z = |T_1^- \cup \dots \cup T_s^-|$, so $x + y + z = n$.

Since $T_\Delta \subseteq T_{ext}$, we have $x \geqslant \beta n$. Also, all but at most $2y + x - \alpha n/2$ vertices of G have at least $y + x/2 - \alpha n/4$ outneighbours, and all but at most $2z + x - \alpha n/2$ vertices of G have at least $z + x/2 - \alpha n/4$ inneighbours. So at least $(2-\gamma)n - 2y - 2z - 2x + \alpha n \geqslant \alpha n/2$ vertices of G satisfy both of these conditions. Let U_0 be the set of these vertices, so $|U_0| \geqslant \alpha n/2$, and each $v \in U_0$ has at least $y + x/2 - \alpha n/4$ outneighbours and at least $z + x/2 - \alpha n/4$ inneighbours.

From each cluster V_i' of G' choose a set X_i of $(1+\alpha)x/k$ vertices uniformly at random, and let $X := X_1 \cup \dots \cup X_k$. Then $|X| = (1+\alpha)x$. For any single vertex $u \in G'$, the probability that u is included in X is $(1+\alpha)x/|G'| \geqslant x/2n$, so by Proposition 2.6, with probability at least $2/3$ the set $U := X \cap U_0$ satisfies $|U| \geqslant \alpha x/5 \geqslant \alpha\beta n/5$. Also, for any vertex $v \in U$, the expected number of outneighbours of v outside X is at least

$$\begin{aligned}\left(y+\frac{x}{2}-\frac{\alpha n}{4}\right)\left(1-\frac{(1+\alpha)x}{|G'|}\right) &\geqslant y-\frac{\alpha n}{4}+\frac{x}{2}-\frac{(1+\alpha)xy}{2(1-\gamma)n}-\frac{(1+\alpha)x^2}{4(1-\gamma)n}\\ &\geqslant y-\frac{\alpha n}{4}+\frac{2xn-2xy-x^2-2\gamma xn-2\alpha xy-\alpha x^2}{4(1-\gamma)n}\\ &\geqslant y+\frac{x^2}{4n}-2\alpha n \geqslant y+\frac{\beta^2 n}{4}-2\alpha n \geqslant y+2\alpha n,\end{aligned}$$

where in the first inequality of the third line we used the fact that $2n-2y-x \geqslant x$. A similar calculation shows that for each $v \in U$, the expected number of inneighbours of v outside X is at least $z+2\alpha n$. So by Proposition 2.6 we find that with probability at least $2/3$, every vertex $v \in U$ has at least $y+\alpha n$ outneighbours outside X and at least $z+\alpha n$ inneighbours outside X. Fix a choice of X such that both these events of probability at least $2/3$ occur.

Since every vertex of U has either at least $(|G|-|X|)/2 \geqslant y+z+\alpha n$ inneighbours outside X or at least $y+z+\alpha n$ outneighbours outside X, we may choose a set $U' \subseteq U$ of size $|U'| \geqslant |U|/2 \geqslant \alpha\beta n/10$ such that either

(α_1) every $v \in U'$ has at least $y+\alpha n$ outneighbours outside X and at least $y+z+\alpha n$ inneighbours outside X, or

(α_2) every $v \in U'$ has at least $y+z+\alpha n$ outneighbours outside X and at least $z+\alpha n$ inneighbours outside X.

So $G'[X]$ is a $(3\varepsilon/\beta)$-regular $(d/2)$-dense cycle of cluster tournaments on clusters $X_1,\dots,X_k$ of size $(1+\alpha)x/k$, and $U' \subseteq X_1 \cup \dots \cup X_k$ has size $|U'| \geqslant \alpha\beta x/10$. Also T_{ext} is a directed tree on x vertices rooted at t_1 and with $\Delta(T_{ext}) \leqslant \Delta^*$, and $H \subseteq V(T_{ext})$ has $|H| \leqslant \delta\beta n/7k \leqslant \delta x/7k$ and $|\{t \in T_{ext} : 1 \leqslant d(t,\mathcal{P}_k(H)) \leqslant k^3\}| \leqslant \delta\beta n \leqslant \delta x$. So by Lemma 2.27 (with $\alpha\beta/10$, Δ^* and $d/2$ in place of λ, Δ and d respectively), $G'[X]$ contains a copy of T_{ext} in which every vertex of H is embedded to a vertex of U'.

In either case, let V_{ext} be the set of vertices of G to which T_{ext} is embedded. We may now complete the embedding of T in G. If U' satisfies (a), then we first proceed through the trees T_i^+ in turn. For each T_i^+, let u_i^+ be the inneighbour of v_i^+ in T_{ext} (so $u_i^+ \in H$ by (iii)). Then u_i^+ has been embedded to some vertex $v \in U'$. This $v \in U'$ has at least $y + \alpha n$ outneighbours outside V_{ext}, of which at most y have been used for embedding the trees T_j^+ for $j < i$. So there are at least αn outneighbours of v outside V_{ext} available to embed T_i^+, and so since $|T_i^+| \leqslant n/\Delta \leqslant \alpha n/3$, by Theorem 1.2 we can embed T_i^+ among these vertices. In this way we may embed each of the T_i^+. We then proceed through the T_i^- similarly. For each T_i^- let u_i^- be the inneighbour of v_i^- in T_{ext} (so $u_i^- \in H$ by (iii)). Then u_i^- has been embedded to some vertex $v \in U'$. This $v \in U'$ has at least $y + z + \alpha n$ inneighbours outside V_{ext}, of which at most $y + z$ have been used for embedding the trees $T_1^+, \ldots, T_r^+$ and the trees T_j^- for $j < i$. So there are at least αn inneighbours of v outside V_{ext} available to embed T_i^-, and so since $|T_i^-| \leqslant n/\Delta \leqslant \alpha n/3$, again by Theorem 1.2 we can embed T_i^- among these vertices. If U' satisfies (b) we can embed T similarly, first embedding the T_i^-, and then the T_i^+. Either way we have completed the embedding of T in G. □

We can now give the proof of Lemma 2.18, which is restated below.

Lemma 2.18 *Suppose that $1/n \ll \mu \ll \nu \ll \eta \ll \alpha$, that G is a tournament on $2(1+\alpha)n$ vertices which is a robust (μ, ν)-outexpander with $\delta^0(G) \geqslant \eta n$ and that T is a directed tree on n vertices. Then G contains a copy of T.*

Proof. Introduce new constants Δ, γ and β with

$$\frac{1}{n} \ll \frac{1}{\Delta} \ll \mu \ll \nu \ll \eta \ll \gamma \ll \beta \ll \alpha.$$

If $|T_\Delta| \geqslant \beta n$, then G contains a copy of T by Lemma 2.32. So we may assume that $|T_\Delta| < \beta n$. As in the proof of the previous lemma, let $T_1^+, \dots, T_r^+$ be the outcomponents of T_Δ in T, and for each i let v_i^+ be the vertex of T_i^+ with an inneighbour in T_Δ. Similarly, let $T_1^-, \dots, T_s^-$ be the incomponents of T_Δ in T, and for each i let v_i^- be the vertex of T_i^- with an outneighbour in T_Δ. By (i) and Proposition 2.19(iv) each T_i^+ and each T_i^- contains at most n/Δ vertices. Let $x = |T_\Delta|$, let $y = |T_1^+ \cup \dots \cup T_r^+|$ and let $z = |T_1^- \cup \dots \cup T_s^-|$, so $x + y + z = n$.

Now, all but at most $2y + x + \alpha n/2$ vertices of G have at least $y + x/2 + \alpha n/4$ outneighbours in G, and all but at most $2z + x + \alpha n/2$ vertices of G have at least $z + x/2 + \alpha n/4$ inneighbours in G. So at least $2(1 + \alpha)n - 2y - 2z - 2x - \alpha n = \alpha n$ vertices of G satisfy both of these conditions. Choose any $\alpha n/8$ of these vertices to form U_0. Then $|U_0| = \alpha n/8$, and each $v \in U_0$ has at least $y + x/2 + \alpha n/8$ outneighbours outside U_0 and at least $z + x/2 + \alpha n/8$ inneighbours outside U_0. Since every vertex v of G has either $d^+(v) \geqslant (1 + \alpha)n - 1 \geqslant y + z + \alpha n$ or $d^-(v) \geqslant (1 + \alpha)n - 1 \geqslant y + z + \alpha n$, we can choose a set $U' \subseteq U_0$ of size $|U'| \geqslant \alpha n/16$ which satisfies either

(a) every $v \in U'$ has at least $y + \alpha n/8$ outneighbours outside U' and at least $y + z + \alpha n/8$ inneighbours outside U', or

(b) every $v \in U'$ has at least $y + z + \alpha n/8$ outneighbours outside U' and at least $z + \alpha n/8$ inneighbours outside U'.

Since $|T_\Delta| < \beta n \leqslant |U'|/3$, and $G[U']$ is a tournament, by Theorem 1.2 we may embed T_Δ in $G[U']$. We may then proceed to embed all the T_i^+ and T_j^- exactly as in the previous lemma, completing the embedding of T in G. $\square$

2.5 Embedding trees in an almost-transitive tournament

We say that a tournament is *transitive* if its vertices can be given a linear order so that every edge is directed towards the endvertex which is greater in this order. It is easy to show that any transitive tournament G on n vertices contains any directed tree T on n vertices, by first showing that the vertices of T can be given a linear order so that every edge is directed towards the endvertex which is greater in this order, and then embedding each vertex of T to the vertex of G in the corresponding position (in the order of vertices of G).

In this section, we prove an approximate version of this result, namely that if a tournament on $(1+\alpha)n$ vertices (for some small α) is sufficiently close to being transitive, then it contains any directed tree on n vertices. To state this lemma precisely, we say that a tournament G on n vertices is ε*-almost-transitive* if the vertices of G can be given an order $v_1, \ldots, v_n$ so that at most εn^2 edges are directed against the ordering of the vertices, *i.e.* they are directed from v_i to v_j where $i > j$.

The proof of this lemma is by a similar method to the proof of Theorem 1.5 in the next section. The approach is that if the lemma is false, then there is some $\alpha > 0$ for which the lemma does not hold, and so the infimum a_{inf} of all α for which the lemma does hold is greater than zero. We then choose α slightly less than a_{inf} and apply (to a smaller subtree) the fact that the lemma holds for any $\alpha' > a_{inf}$ to show that the lemma holds for α, giving a contradiction.

Lemma 2.33 *For all $\alpha > 0$ there exists $\varepsilon_0 > 0$ and $n_0 \in \mathbb{N}$ such that for any $\varepsilon \leqslant \varepsilon_0$ and any $n \geqslant n_0$, any ε-almost-transitive tournament G on at least $(1+\alpha)n$ vertices contains any directed tree T on n vertices.*

Proof. We consider the set A of all positive values of α such that the lemma holds. More precisely, A is the set of all positive values of α such that there exist $\varepsilon_0 > 0$ and $n_0 \in \mathbb{N}$ so that for any $n \geqslant n_0$ and $\varepsilon \leqslant \varepsilon_0$, any ε-almost-transitive tournament G on at least $(1+\alpha)n$ vertices contains a copy of any directed tree T on n vertices. So if $\alpha' \in A$ and $\alpha'' > \alpha'$ then $\alpha'' \in A$. Also $2 \in A$ by Theorem 1.2, and so we may define $a_{inf} = \inf A$, with $0 \leqslant a_{inf} \leqslant 2$. Then for any $\alpha' > a_{inf}$, $\alpha' \in A$. With this definition the lemma is equivalent to the statement that $a_{inf} = 0$, so suppose for a contradiction that $a_{inf} > 0$. Let

$$\gamma \ll \frac{1}{\Delta} \ll a_{inf} \qquad \text{and} \qquad \alpha = a_{inf} - \gamma,$$

so we may assume that $1/\Delta \ll \alpha$. Then $\alpha + 2\gamma > a_{inf}$, so $\alpha + 2\gamma \in A$, and so by definition of A there exist $\varepsilon'_0 > 0$ and $n'_0 \in \mathbb{N}$ such that for any $\varepsilon' \leqslant \varepsilon'_0$ and $n' \geqslant n'_0$, any ε'-almost-transitive tournament G on at least $(1 + \alpha + 2\gamma)n'$ vertices contains a copy of any directed tree T on n' vertices. Moreover, we may assume that $\varepsilon'_0 \ll \gamma$. Fix such an ε'_0 and n'_0, and let $1/n_0 \ll 1/n'_0, \gamma$ and $\varepsilon_0 \ll \varepsilon'_0$. We show that for any $n \geqslant n_0$ and $\varepsilon \leqslant \varepsilon_0$, any ε-almost-transitive tournament G on at least $(1+\alpha)n$ vertices contains a copy of any directed tree T on n vertices. It then follows that $\alpha \in A$, yielding a contradiction and therefore proving the lemma.

So let $\varepsilon \leqslant \varepsilon_0$ and $n \geqslant n_0$, let G be an ε-almost-transitive tournament on at least $(1+\alpha)n$ vertices and let T be a directed tree on n vertices. If $|G| \geqslant 3n$, then G contains a copy of T by Theorem 1.2, and so we may assume that $|G| < 3n$. Since G is ε-almost-transitive, we may order the vertices of G as $v_1, \ldots, v_{|G|}$ so that at most $\varepsilon|G|^2 \leqslant 9\varepsilon n^2$ edges are directed from v_j to v_i where $i < j$. Now, at most $18\sqrt{\varepsilon}n$ vertices of G are incident to more than $\sqrt{\varepsilon}n$ such edges; let G' be the subgraph of G obtained by deleting these vertices from G, and let $v'_1, v'_2, \ldots, v'_{|G'|}$ be the vertices of G' in the inherited order. Then G' is a tournament on at least $(1+\alpha-18\sqrt{\varepsilon})n$ vertices such that for any vertex v'_i there are at most $\sqrt{\varepsilon}n$ vertices v'_j for which the edge between v'_i and v'_j is directed towards $v'_{\min\{i,j\}}$.

We now consider three possibilities for the core tree T_Δ, in each case showing that T can be embedded in G'.

Case 1: Some vertex $t \in T_\Delta$ has $d^+_{T_\Delta}(t) \geqslant 2$. Then let F^- be the (possibly empty) forest consisting of each incomponent of t. Similarly let the outcomponents of t be partitioned into two forests, F_1^+ and F_2^+. Since $d^+_{T_\Delta}(t) \geqslant 2$, by Proposition 2.19(ii) at least two outcomponents of t each contain at least n/Δ vertices, and so we may choose F_1^+ and F_2^+ so that $|F_1^+|, |F_2^+| \geqslant n/\Delta$. Note that $|F^-| = w^-(t)$, and $|F_1^+| + |F_2^+| = w^+(t)$, so in particular $w^+(t) \geqslant 2n/\Delta$, and also recall that $w^+(t) + w^-(t) = n - 1$.

We first determine where to embed the vertex t. For this, let

$$p := \begin{cases} 3\gamma n + \sqrt{\varepsilon} n + 1 & \text{if } w^-(t) < \gamma n, \\ (1+\alpha+2\gamma)w^-(t) + \sqrt{\varepsilon} n + 1 & \text{if } w^-(t) \geqslant \gamma n. \end{cases}$$

and embed t to the vertex v'_p of G'. This can be done, as we show later that $p < |G'|$. We embed F^- in the vertices preceding v'_p and F_1^+, F_2^+ in the vertices succeeding v'_p in the vertex ordering of G'. Embedding F^- is possible because p is a little larger than one might expect, whereas embedding F_1^+ and F_2^+ can be done successively, which gives us enough room for both. Let $S^- = N^-(v'_p) \cap \{v'_1, \dots, v'_{p-1}\}$, and $S^+ = N^+(v'_p) \cap \{v'_{p+1}, \dots, v'_{|G'|}\}$. Then S^- and S^+ are disjoint, $|S^-| \geqslant p - \sqrt{\varepsilon} n - 1$ and $|S^+| \geqslant |G'| - p - \sqrt{\varepsilon} n$. Next we embed F^- in $G'[S^-]$. Indeed, if $w^-(t) < \gamma n$ then $|S^-| \geqslant 3\gamma n$, and so by Theorem 1.2 we can embed F^- in $G'[S^-]$. Alternatively, if $w^-(t) \geqslant \gamma n$, let $n' = w^-(t) \geqslant n'_0$ and $\varepsilon' = |G|^2\varepsilon/(n')^2 \leqslant \varepsilon'_0$, then F^- is a forest on n' vertices, and $G'[S^-]$ is an ε'-almost-transitive tournament on at least $(1+\alpha+2\gamma)n'$ vertices. So by the choice of ε'_0 and n'_0 we can embed F^- in $G'[S^-]$.

Finally we complete the embedding of T in G' by embedding F_1^+ and F_2^+ in $G'[S^+]$. Now,

$$\begin{aligned} |S^+| &\geqslant |G'| - p - \sqrt{\varepsilon} n \\ &\geqslant (1+\alpha-18\sqrt{\varepsilon})n - (3\gamma n + (1+\alpha+2\gamma)w^-(t) + \sqrt{\varepsilon} n + 1) - \sqrt{\varepsilon} n \\ &\geqslant (1+\alpha)w^+(t) - 5\gamma n - 20\sqrt{\varepsilon} n \geqslant (1+\alpha)w^+(t) - 6\gamma n. \end{aligned}$$

Let $n' = |F_1^+|$, so $n'_0 \leqslant n/\Delta \leqslant n'$ and $n' \leqslant w^+(t) - n/\Delta$, and again let $\varepsilon' = |G|^2\varepsilon/(n')^2$, so $\varepsilon' \leqslant \varepsilon'_0$. Then $G'[S^+]$ is an ε'-almost-transitive tournament on $|S^+| \geqslant (1+\alpha)(n' + n/\Delta) - 6\gamma n \geqslant (1+\alpha+1/\Delta)n' + (\alpha/\Delta - 6\gamma)n \geqslant (1+\alpha+2\gamma)n'$ vertices, and so by our choice of n'_0 and ε'_0, we may embed F_1^+ in $G'[S^+]$.

Now, let S^+_{rem} consist of the vertices of S^+ not occupied by the vertices of F_1^+. We embed F_2^+ in S^+_{rem} in a similar manner. Indeed, we now let $n' = |F_2^+|$, so again $n'_0 \leqslant n/\Delta \leqslant n'$, and again let $\varepsilon' = |G|^2\varepsilon/(n')^2 \leqslant \varepsilon'_0$. Then

$$\begin{aligned}|S^+_{rem}| = |S^+| - |F_1^+| &\geqslant (1+\alpha)w^+(t) - 6\gamma n - (w^+(t) - |F_2^+|)\\ &= (1+\alpha)n' + \alpha|F_1^+| - 6\gamma n \geqslant (1+\alpha+2\gamma)n',\end{aligned}$$

so $G'[S^+_{rem}]$ is an ε'-almost-transitive tournament on at least $(1+\alpha+2\gamma)n'$ vertices, and so by our choice of n'_0 and ε'_0, we may embed F_2^+ in $G'[S^+_{rem}]$.

Case 2: Some vertex $t \in T_\Delta$ has $d^-_{T_\Delta}(t) \geqslant 2$. Then we may embed T in G' by the same method as in Case 1, the main difference being that the roles of outdegrees and outneighbours are switched with those of indegrees and inneighbours.

Case 3: T_Δ is a directed path (possibly consisting of just a single vertex). Then let $w^+ = w^+(T_\Delta)$ and $w^- = w^-(T_\Delta)$ be as defined in Section 2.2, and partition the vertices of G' into three sets $S^- = \{v'_1, \dots, v'_{w^- + \alpha n/3}\}$, $S = \{v'_{w^-+\alpha n/3+1}, \dots, v'_{|G'|-w^+-\alpha n/3}\}$ and $S^+ = \{v'_{|G'|-w^+-\alpha n/3+1}, \dots, v'_{|G'|}\}$. Then since $w^+ + w^- + |T_\Delta| = n$, we know that $|S| = |G'| - w^- - w^+ - 2\alpha n/3 \geqslant |T_\Delta|$. Therefore by Theorem 1.4 we may embed T_Δ in $G'[S]$. Now, let $T_1^+, \dots, T_r^+$ be the outcomponents of T_Δ in T, and let $T_1^-, \dots T_s^-$ be the incomponents of T_Δ in T. For each i let $t_i^+ \in T_\Delta$ be the vertex of T_Δ which has a neighbour in T_i^+, and let v_i^+ be the vertex of G' to which t_i^+ was embedded. Similarly, for each i let $t_i^- \in T_\Delta$ be the vertex of T_Δ which has a neighbour in T_i^-, and let v_i^- be the vertex of G' to which t_i^- was embedded. Now, every vertex of T lies in T_Δ or one of the T_i^+ or T_i^-. Furthermore $|T_i^+|, |T_j^-| \leqslant n/\Delta$ for each i and j by Proposition 2.19(iv).

We complete the embedding of T in G' by greedily embedding each T_i^+ in $N^+(v_i^+) \cap S^+$, and each T_i^- in $N^-(v_i^-) \cap S^-$. Indeed, suppose we have already embedded $T_1^+, \dots, T_{i-1}^+$, and we now wish to embed T_i^+. Then

$$|N^+(v_i^+) \cap S^+| \geqslant |S^+| - \sqrt{\varepsilon}n \geqslant w^+ + \frac{\alpha n}{3} - \sqrt{\varepsilon}n \geqslant w^+ + \frac{\alpha n}{4}.$$

At most w^+ of these vertices have already been occupied by vertices of $T_1^+, \dots, T_{i-1}^+$, and so there remain at least $\alpha n/4$ available vertices in which to embed T_i^+. Since $|T_i^+| \leqslant n/\Delta \leqslant \alpha n/12$, we may embed T_i^+ in these available vertices by Theorem 1.2. Continuing in this way we may embed all of the T_i^+, and the T_i^- may be embedded similarly, to give us a copy of T in G'.

Any directed tree in which every vertex has at most one outneighbour and at most one inneighbour is a directed path. So T_Δ must fall into at least one of the three cases, and so we can find a copy of T in G', and hence in G, contradicting our assumption that $a_{inf} > 0$. So we must have $a_{inf} = 0$, and so the lemma holds. $\square$

2.6 Proof of Theorem 1.5

Recall the statement of Theorem 1.5.

Theorem 1.5 *Let $\alpha > 0$. Then the following properties hold.*

(1) There exists n_0 such that for any $n \geqslant n_0$, any tournament G on at least $2(1+\alpha)n$ vertices

contains any directed tree T on n vertices.

(2) *Let Δ be any positive integer. Then there exists n_0 such that for any $n \geqslant n_0$, any tournament G on at least $(1+\alpha)n$ vertices contains any directed tree T on n vertices with $\Delta(T) \leqslant \Delta$.*

The proofs of each of the two statements of the theorem are very similar, so to avoid repetition we prove the first statement, explaining in footnotes where the proof of the second statement differs.

2.6.1 Partitioning the vertices of G and T

As in the last section, we consider the set A of all positive values of α such that the theorem holds. So $\alpha' \in A$ if and only if there exists n_0 such that for any $n \geqslant n_0$, any tournament on at least $2(1+\alpha')n$ vertices contains any directed tree on n vertices. So if $\alpha' \in A$ and $\alpha'' > \alpha'$ then $\alpha'' \in A$, and also $1/2 \in A$ by Theorem 1.2. Thus we may define $a_{inf} = \inf A$, and then the theorem is equivalent to the statement that $a_{inf} = 0$. So suppose $a_{inf} > 0$, and choose constants

$$\frac{1}{n_0} \ll \frac{1}{n_0'} \ll \mu \ll \nu \ll \eta \ll \zeta \ll \frac{1}{\Delta'} \ll \gamma \ll a_{inf}.$$

Let $\alpha = a_{inf} - \mu$, so $\alpha \leqslant 1/2$, and we may assume that $\gamma \ll \alpha$. Then $\alpha + 2\mu \in A$, and so for any $n' \geqslant n_0'$, any tournament on at least $2(1+\alpha+2\mu)n'$ vertices contains any directed tree on n' vertices. We prove that if $n \geqslant n_0$, any tournament G on at least $2(1+\alpha)n$ vertices contains any directed tree on n vertices. This proves that $\alpha \in A$, giving a contradiction to our assumption that $a_{inf} > 0$, and so proving the theorem.[1]

[1]For the bounded degree case, fix any value of Δ, and here A is defined by $\alpha' \in A$ if and only if there exists n_0 such that for any $n \geqslant n_0$, any tournament on at least $(1+\alpha')n$ vertices contains any directed tree T on n vertices with $\Delta(T) \leqslant \Delta$. So if $\alpha' \in A$ and $\alpha'' > \alpha'$ then $\alpha'' \in A$, and also $2 \in A$ by Theorem 1.2. Thus we may define $a_{inf} = \inf A$; then the theorem is equivalent to the statement that $a_{inf} = 0$. So suppose $a_{inf} > 0$, and choose

So let G be a tournament on at least $2(1+\alpha)n$ vertices[2], and let T be a directed tree on n vertices. If $|G| \geqslant 3n$ then G contains T by Theorem 1.2, so we may assume that $|G| < 3n$. By Lemma 2.12, we may choose disjoint subsets $S_1, \dots, S_r$ of $V(G)$ such that:

(i) $|\bigcup_{i\in[r]} S_i| \geqslant (1-\zeta)|G|$,

(ii) for each $i \in [r]$, any vertex $v \in S_i$ has at most $\zeta|G|$ inneighbours in $\bigcup_{j>i} S_j$ and at most $\zeta|G|$ outneighbours in $\bigcup_{j<i} S_j$, and

(iii) for each $i \in [r]$, either $G[S_i]$ is a robust (μ,ν)-outexpander with $\delta^0(G[S_i]) \geqslant \eta|G|$ or $|S_i| < \zeta|G|$.

Let $G' := G[\bigcup_{i=1}^r S_i]$. Then G' is a tournament, and by (i) we have[3]

$$|G'| \geqslant (1-\zeta)|G| \geqslant (2+2\alpha-3\zeta)n. \tag{2.35}$$

Suppose now that for each $i \in [r]$ we have $|S_i| < \gamma n$. Then G' is 2γ-almost-transitive. Indeed, order the vertices of G' as $v_1, v_2, \dots, v_{|G'|}$ beginning with all the vertices of S_1, then the vertices of S_2, and so forth. Say that an edge of G' is *bad* if it is directed from some v_i to some v_j, where $j < i$. Then by (ii), the number of bad edges in G' is at most

$$2\zeta|G||G'| + \sum_{i\in[r]} \binom{|S|}{2} \leqslant 18\zeta n^2 + \frac{3\gamma n^2}{2} \leqslant 2\gamma n^2.$$

constants $1/n_0 \ll 1/n_0' \ll \mu \ll \nu \ll \eta \ll \zeta \ll 1/\Delta' \ll \gamma \ll 1/\Delta, a_{inf}$. Let $\alpha = a_{inf} - \mu$, so $\alpha < 2$, and we may assume that $\gamma \ll \alpha$. Then $\alpha + 2\mu \in A$, so for any $n' \geqslant n_0'$, any tournament on at least $(1+\alpha+2\mu)n'$ vertices contains any directed tree T on n' vertices with $\Delta(T) \leqslant \Delta$. Using this, we prove that if $n \geqslant n_0$, any tournament G on at least $(1+\alpha)n$ vertices contains any directed tree T on n vertices with $\Delta(T) \leqslant \Delta$. This proves that $\alpha \in A$, giving a contradiction to our assumption that $a_{inf} > 0$, and so proving the theorem.

[2]For the bounded degree case, instead let G be a tournament on at least $(1+\alpha)n$ vertices.

[3]For the bounded degree case we have instead

$$|G'| \geqslant (1-\zeta)|G| \geqslant (1+\alpha-3\zeta)n. \tag{2.34}$$

Since $|G'| \geqslant n$ by 2.35, $|G'|$ is indeed 2γ-almost-transitive. So by Lemma 2.33 G' contains a copy of T, which is also a copy of T in G.

We may therefore assume that for some $i \in [r]$ we have $|S_i| \geqslant \gamma n$. Fix such an i; then by (iii), $G[S_i]$ is a (μ, ν)-robust outexpander with $\delta^0(G[S_i]) \geqslant \eta n$. For this i, let $S = S_i$, let $S^+ = \bigcup_{i<j\leqslant r} S_j$ and let $S^- = \bigcup_{1\leqslant j<i} S_j$. So $G' = G[S^+ \cup S^- \cup S]$. Also, by (ii) any vertex $u \in S^+$ has at most $3\zeta n$ outneighbours in $S \cup S^-$. Similarly each vertex of S has at most $3\zeta n$ outneighbours in S^- and at most $3\zeta n$ inneighbours in S^+, and each vertex of S^- has at most $3\zeta n$ inneighbours in $S^+ \cup S$. Define β, β^+, β^- by $|S| = \beta|G'|$, $|S^+| = \beta^+|G'|$, and $|S^-| = \beta^-|G'|$, so $\beta + \beta^+ + \beta^- = 1$ and $\beta \geqslant \gamma n/|G'| \geqslant \gamma/3$.

Suppose first that β^+ and β^- are both small. More precisely, suppose that $\beta^+, \beta^- \leqslant \alpha\beta^2/20$, and so $\beta \geqslant 1 - \alpha/10$. Then we find a copy of T in $G[S]$ (and therefore in G). Indeed, T is a directed tree on n vertices, and $G[S]$ is a (μ, ν)-robust outexpander with $\delta^0(G[S]) \geqslant \eta n$. Furthermore,

$$|S| = \beta|G'| \overset{(2.35)}{\geqslant} (2 + 2\alpha - 3\zeta)(1 - \frac{\alpha}{10})n \geqslant 2(1 + \frac{\alpha}{2})(1 - \frac{\alpha}{10})n \geqslant 2(1 + \frac{\alpha}{4})n$$

and so by Lemma 2.18 $G[S]$ (and therefore G) contains a copy of T.[4]

So we may assume that at least one of β^+ and β^- is greater than $\alpha\beta^2/20$, so in particular, $\beta \leqslant 1 - \alpha\beta^2/20$. We next partition the vertices of T according to the values of β^+ and β^-.

Case 1: β^- is large but β^+ is small. More precisely, $\beta^+ \leqslant \alpha\beta^2/20$ and $\beta^- > \alpha\beta^2/20$. Then we partition the vertex set of T into T^- and T^0, where every edge of T between T^- and T^0 is directed from T^- to T^0, and $|T^-| = \beta^-(1 - \alpha\beta)n$. We can form T^0 greedily by successively

[4]For the bounded degree case, $|S| = \beta|G'| \overset{(2.34)}{\geqslant} (1 + \alpha - 3\zeta)(1 - \frac{\alpha}{10})n \geqslant (1 + \frac{\alpha}{4})n$, and so $G[S]$ (and therefore G) contains a copy of T by Lemma 2.13.

removing a sink vertex from T and adding it to T^0. Since $\beta^+ + \beta + \beta^- = 1$,

$$|T^0| = n - |T^-| = \beta n(1 + \alpha - \alpha\beta) + (1 - \alpha\beta)\beta^+ n \leqslant \beta n(1 + \alpha - \alpha\beta) + \frac{\alpha\beta^2 n}{20}.$$

Case 2: β^+ is large but β^- is small. More precisely, $\beta^- \leqslant \alpha\beta^2/20$ and $\beta^+ > \alpha\beta^2/20$. Then we similarly partition the vertex set of T into T^0 and T^+, where every edge of T between T^0 and T^+ is directed from T^0 to T^+, and $|T^+| = \beta^+(1 - \alpha\beta)n$. Again $|T^0| = n - |T^+| \leqslant \beta n(1 + \alpha - \alpha\beta) + \alpha\beta^2 n/20$.

Case 3: β^+ and β^- are both large. More precisely, $\beta^+, \beta^- > \alpha\beta^2/20$. Then we partition the vertex set of T into pieces T^-, T^0 and T^+ such that all edges of T between T^- and T^0 are directed from T^- to T^0, all edges of T between T^0 and T^+ are directed from T^0 to T^+ and all edges of T between T^- and T^+ are directed from T^- to T^+. Also $|T^+| = \beta^+(1 - \alpha\beta)n$ and $|T^-| = \beta^-(1 - \alpha\beta)n$, so $|T^0| = \beta(1 + \alpha - \alpha\beta)n$.

Note that in each of the three cases T^0 satisfies

$$|T^0| \leqslant \beta(1 + \alpha - \alpha\beta)n + \frac{\alpha\beta^2 n}{20} \leqslant \beta(1 + \alpha)n - \frac{\alpha\beta^2 n}{2}. \tag{2.36}$$

2.6.2 Embedding T in G

Having partitioned the vertices of G' into three sets S, S^+ and S^-, and the directed tree T into three forests T^+, T^0, T^-, we now complete the proof by embedding T in G, with T^-, T^0 and T^+ embedded in $G[S^-], G[S]$ and $G[S^+]$ respectively. Indeed, the fact that $G[S]$ is a robust (μ, ν)-outexpander enables us to embed slightly more vertices in $G[S]$ than the βn that would be embedded in $G[S]$ if the vertices of T were distributed proportionately amongst $G[S]$, $G[S^+]$

and $G[S^-]$. This gives us some leeway for embedding T^+ and T^- in $G[S^+]$ and $G[S^-]$ respectively, which by our choice of α is sufficient to successfully complete these embeddings.

So let $T_1^-, \dots, T_x^-$ be the component subtrees of T^-, let $T_1^+, \dots, T_y^+$ be the component subtrees of T^+, and let $T_1, \dots, T_z$ be the component subtrees of T^0. Let the *contracted tree* T_{con} be formed from T by contracting each T_i^+, T_i^- and T_i to a single vertex.

To begin the embedding, we embed into $G[S]$ every T_i satisfying $|T_i| \geqslant n/\Delta'$. Note that there are at most Δ' such T_i. Also, the union of all such T_i is a forest on at most $|T^0|$ vertices, and the tournament $G[S]$ is a robust (μ, ν)-outexpander on

$$\beta|G'| \overset{(2.35)}{\geqslant} \beta(2 + 2\alpha - 3\zeta)n \overset{(2.36)}{\geqslant} 2\left(1 + \frac{\alpha\beta}{10}\right)|T^0| \geqslant 2(1+\gamma^2)|T^0|$$

vertices with $\delta^0(G[S]) \geqslant \eta n$, and hence $G[S]$ contains a copy of this forest by Lemma 2.18.[5]

Now, choose an order of the vertices of T_{con}, beginning with the at most Δ' vertices corresponding to the T_i which we have just embedded, and such that any vertex of T_{con} has at most Δ' neighbours preceding it in this order. (To do this, choose one of the Δ' vertices corresponding to the T_i which have already been embedded, and then choose any ancestral ordering of the vertices of T_{con}, beginning with the chosen vertex, so every vertex has at most one neighbour preceding it in this order. Now move the remaining $\Delta' - 1$ vertices corresponding to the T_i which have already been embedded to the front of this order; then every vertex gains at most $\Delta' - 1$ preceding neighbours.) We proceed through the remaining vertices of T_{con} in this order, at each step embedding the directed tree T_i, T_i^+ or T_i^- corresponding to the current vertex of T_{con} in the unoccupied vertices of the tournament $G[S], G[S^+]$ or $G[S^-]$ respectively.

So suppose first that the current vertex t^* of T_{con} corresponds to some T_i. Since T_i has not

[5]For the bounded degree case, $|S| \geqslant (1+\gamma^2)|T^0|$ by a similar calculation, and so $G[S]$ contains a copy of this forest by Lemma 2.13.

already been embedded, we know that $|T_i| < n/\Delta'$. Also, since t^* has at most Δ' neighbours preceding it in T_{con}, the vertices of T_i have at most Δ' neighbours outside T_i which have already been embedded. Since T_i is a component of T^0, each of these neighbours of vertices in T_i lies either in T^- (in which case it is an inneighbour) or in T^+ (in which case it is an outneighbour). So let $t_1^-, \dots, t_p^-$ be the vertices in T^- which are inneighbours of some vertex in T_i and which have previously been embedded, and let $v_1^-, \dots, v_p^-$ be the vertices of $G'[S^-]$ to which $t_1^-, \dots, t_p^-$ were embedded. Similarly, let $t_1^+, \dots, t_q^+$ be the vertices in T^+ which are outneighbours of some vertex in T_i and which have previously been embedded, and let $v_1^+, \dots, v_q^+$ be the vertices of $G'[S^+]$ to which $t_1^+, \dots, t_q^+$ were embedded. Finally let S^* be the set of unoccupied vertices in $S \cap N^+(v_1^-, \dots, v_p^-) \cap N^-(v_1^+, \dots, v_q^+)$. Then we wish to embed T_i in S^*. For this, note that

$$\begin{aligned} |S^*| &\geqslant |S| - 3(p+q)\zeta n - |T^0| \overset{(2.36)}{\geqslant} \beta|G'| - 3\Delta'\zeta n - (\beta(1+\alpha)n - \frac{\alpha\beta^2 n}{2}) \\ &\overset{(2.35)}{\geqslant} \beta(1+\alpha)n - (3+3\Delta')\zeta n - \beta(1+\alpha)n + \frac{\alpha\beta^2 n}{2} \geqslant \frac{\alpha\beta^2 n}{3} \geqslant \frac{3n}{\Delta'} \geqslant 3|T_i|. \end{aligned}$$

Note that this calculation is valid for both the bounded degree case and the unbounded degree case, with plenty of room to spare in the unbounded case. So by Theorem 1.2, $G[S^*]$ contains a copy of T_i, to which we embed T_i.

Alternatively, if the current vertex of T_{con} corresponds to some T_i^-, then similarly the vertices of T_i^- have at most Δ' neighbours outside T_i^- which have already been embedded, all of which are outneighbours. As before we let $v_1, \dots, v_r$ be the vertices of $G'[S \cup S^+]$ to which these vertices have been embedded, and let S^* be the set of unoccupied vertices of $S^- \cap N^-(v_1, \dots, v_r)$. Note that at most $|T^-| - |T_i^-|$ vertices of T^- have already been embedded. Since some T_i^- exists we have

$$
\begin{aligned}
|S^*| &\geqslant |S^-| - 3r\zeta n - (|T^-| - |T_i^-|) && (2.37) \\
&\overset{(2.35)}{\geqslant} \beta^-(2+2\alpha)n - (3+3\Delta')\zeta n - \beta^-(1-\alpha\beta)n + |T_i^-| \\
&\geqslant \beta^-(1+2\alpha+\frac{\alpha\beta}{2})n + |T_i^-|. && (2.38)
\end{aligned}
$$

In the final line we used the fact that $\beta^- \geqslant \alpha\beta^2/20$ and $\beta \geqslant \gamma/3$ (so $\zeta, 1/\Delta' \ll \gamma, \beta, \beta^-$). So $|S^*| \geqslant 2(1+\alpha+2\mu)|T_i^-|$. Therefore if $|T_i^-| \geqslant \beta^- n/2$, then $|T_i^-| \geqslant \alpha\beta^2 n/40 \geqslant \alpha\gamma^2 n/360 \geqslant n_0'$, and so we can embed T_i^- in $G[S^*]$ by our choice of n_0'. On the other hand, if $|T_i^-| < \beta^- n/2$ then $|S^*| \geqslant 3|T_i^-|$ by (2.37), and so we can embed T_i^- in $G[S^*]$ by Theorem 1.2.[6]

Finally, if the current vertex of T_{con} corresponds to some T_i^+, we embed T_i^+ in the unoccupied vertices of S^+ by a similar method to the method used to embed some T_i^- in the unoccupied vertices of $G[S^-]$. We continue in this manner until we have embedded the T_i, T_i^+ or T_i^- corresponding to each vertex of T_{con}, at which point we have obtained an embedding of T in G, completing the proof. At each stage in this proof we had 'room to spare' in our choices, and so the fact that the expressions for $|T_i|$, $|T_i^+|$ and $|T_i^-|$ and other such expressions may not be integers is not a problem. □

[6] For the bounded degree case

$$
\begin{aligned}
|S^*| &\geqslant |S^-| - 3r\zeta n - (|T^-| - |T_i^-|) \overset{(2.34)}{\geqslant} \beta^-(1+\alpha)n - (3+3\Delta')\zeta n - \beta^-(1-\alpha\beta)n + |T_i^-| \\
&\geqslant \beta^-(\alpha + \frac{\alpha\beta}{2})n + |T_i^-|.
\end{aligned}
$$

So $|S^*| \geqslant (1+\alpha+2\mu)|T_i^-|$. Therefore if $|T_i^-| \geqslant \beta^-\alpha n/2$, then $|T_i^-| \geqslant n_0'$, and so we can embed T_i^- in $G[S^*]$ by our choice of n_0'. On the other hand, if $|T_i^-| < \beta^-\alpha n/2$ then $|S^*| \geqslant 3|T_i^-|$, and so we can embed T_i^- in $G[S^*]$ by Theorem 1.2.

CHAPTER 3

A PROOF OF SUMNER'S UNIVERSAL TOURNAMENT CONJECTURE FOR LARGE n

In this chapter, we use results and methods from Chapter 2 to prove that Sumner's universal tournament conjecture holds for any sufficiently large n. This is Theorem 1.1, which is restated below.

Theorem 1.1 *There exists n_0 such that the following holds. Let T be a directed tree on $n \geqslant n_0$ vertices, and G a tournament on $2n-2$ vertices. Then G contains a copy of T.*

3.1 Outline of the proof of Theorem 1.1

In Section 3.2, we prove some preliminary results. First, we introduce the notion of an 'almost-regular' tournament G, which is a tournament in which every vertex has in- and outdegree approximately equal to $|G|/2$. Such tournaments play an important role in the proof of Theorem 1.1. Indeed, in Section 3.4 we prove that Sumner's universal tournament conjecture holds for any almost-regular tournament G, a result which we use repeatedly when considering general tournaments. Section 3.2 also contains three auxiliary lemmas for embedding a directed tree T in a tournament G. These results are derived from Theorems 1.2 and 1.5(1) and are used extensively in later sections:

- Lemma 3.2 is designed to embed a directed tree T which is similar to an outstar, in the sense that T contains a vertex t with no inneighbours such that every component of $T-t$ is small. In particular, this lemma is useful in the case where $|T_\Delta| = 1$.

- In Lemma 3.3, we consider a subtree T_c of T with the property that every component of $T - T_c$ is small, showing that a suitable embedding of T_c in G can be extended to an embedding of T in G. In particular, we often apply this lemma with $T_c = T_\Delta$.

- In Lemma 3.4 we consider the case where the vertices of G can be partitioned into disjoint sets Y and Z such that almost all edges between Y and Z can be directed the same way. Here we show that if the vertices of T are partitioned appropriately between forests F^- and F^+, then to be able to embed T in G it is sufficient to embed the largest component of F^+ within Y.

We begin the proof of Theorem 1.1 in Section 3.3, by proving the case where $|T_\Delta| = 1$ (Lemma 3.5). Note that this includes the extremal case where T is an in- or outstar. To do this, we first embed the single vertex of T_Δ to a vertex of G with appropriate in- and outdegree. We then use Lemma 3.2, Lemma 3.3 and Theorem 1.3 to embed the components of $T - T_\Delta$ appropriately among the remaining vertices of G to obtain a copy of T in G.

In Section 3.4 we use the regularity method to prove

- Lemma 3.14, which states that Theorem 1.1 holds in the case where G is almost-regular and T_Δ is small enough to be embedded within a single cluster of G.

To prove this, we first select an appropriate cluster or pair of clusters of G in which to embed T_Δ, and then use Lemma 3.3 to extend this embedding of T_Δ to an embedding of T in G. We also prove that if we additionally assume that $|T_\Delta| \geqslant 2$ then the result holds with room to spare, *i.e.* we can allow G to be of order $(2-\alpha)n$, where α is small. It is not hard to show that any almost-regular tournament G is also a robust outexpander, and so we may combine Lemma 3.14 with Lemma 2.32 to prove

- Lemma 3.19, which states that Theorem 1.1 holds whenever G is almost-regular, and indeed holds with room to spare if $|T_\Delta| \geqslant 2$.

The primary goal of Section 3.5 is

- Lemma 3.20, which states that Theorem 1.1 holds for all directed trees T for which T_Δ is small.

In particular, the 'near extremal' constructions described in the introduction are dealt with in this part of the proof. For this, let $x = |T_\Delta|$, let y be the number of vertices which lie in outcomponents of T_Δ, and let z be the number of vertices which lie in incomponents of T_Δ, so $x + y + z = n$. Our assumption in Section 3.5 is that x is small. If at least $6x$ vertices $v \in G$ satisfy both $d^+(v) \geqslant y + \varepsilon n$ and $d^-(v) \geqslant z + \varepsilon n$, then either at least $3x$ of these vertices $v \in G$ satisfy $d^+(v) \geqslant y + \varepsilon n$ and $d^-(v) \geqslant y + z + \varepsilon n$ or at least $3x$ of these vertices satisfy $d^+(v) \geqslant y + z + \varepsilon n$ and $d^-(v) \geqslant z + \varepsilon n$. We may then use Theorem 1.2 to embed T_Δ among these $3x$ vertices and then apply Lemma 3.3 twice to extend this embedding to an embedding of T in G. We may therefore assume that fewer that $6x$ vertices $v \in G$ satisfy these conditions, from which we deduce

- Lemma 3.21, which states that we may assume that the tournament G contains two almost-regular subtournaments on vertex sets Y and Z which between them contain almost all of the vertices of G.

This structure allows us to make extensive use of Lemma 3.19. Indeed, by using Lemma 3.19 to embed a suitable subforest of T into Y or Z, and then using Lemma 3.4 to embed the remainder of T, we show in Lemmas 3.23 and 3.24 that we may assume that T_Δ is a short directed path and that most of the remainder of T is attached to the endvertices of this path. We then consider the case $|T_\Delta| = 2$ separately, proving that Theorem 1.1 holds for such T in Lemma 3.25. This allows us to assume for the proof of Lemma 3.20 that $|T_\Delta| \geqslant 3$. Since T_Δ is a directed path, we can use Redei's theorem (Theorem 1.4) to embed T_Δ within a set W of $|T_\Delta|$ vertices which have high in- and outdegree, and then apply Lemmas 3.2 and 3.3 to complete the embedding.

Finally, in Section 3.6 we complete the proof of Theorem 1.1. By Lemma 3.20 we may assume for this that T_Δ is large. None of the extremal or near-extremal cases satisfy this condition, so we always have a little room to spare in our calculations in this part of the proof. We proceed by using Lemma 2.12 to split the tournament G into disjoint subtournaments, where each subtournament is either small or a robust outexpander of large minimum semidegree. If there is just one such subtournament then this subtournament contains a copy of T by Lemma 2.32. By using Lemma 3.4 we prove Lemma 3.29, which shows that if there are two such subtournaments then these must also together contain a copy of T. We may therefore assume in the proof of Theorem 1.1 that there are at least three such subtournaments of G. In this case we use Lemma 2.32, Theorem 1.5 and Theorem 1.2 to embed T into these subtournaments.

3.2 Preliminary results

3.2.1 Almost-regular tournaments

In a regular directed graph G, every vertex v has $d^+(v) = d^-(v) = e(G)/|G|$. We say that a directed graph G is γ-*almost-regular* if every vertex $v \in G$ has $d^+(v), d^-(v) \geqslant (1-\gamma)e(G)/|G|$. In particular, if G is a tournament then G is γ-almost-regular if and only if every vertex $v \in G$ has $d^+(v), d^-(v) \geqslant (1-\gamma)(|G|-1)/2$. The next proposition shows that for a large tournament G only one of these two bounds is needed to ensure that G contains an almost-spanning almost-regular tournament.

Proposition 3.1 *Suppose that $1/n \ll \alpha \ll \gamma \ll 1$. Let G be a tournament on n vertices in which at least one of the following holds:*

(i) $d^+(v) \geqslant (1-\alpha)(n-1)/2$ *for every* $v \in G$*,*

(ii) $d^-(v) \geqslant (1-\alpha)(n-1)/2$ *for every* $v \in G$*,*

(iii) $d^+(v) \leqslant (1+\alpha)(n-1)/2$ *for every* $v \in G$*,*

(iv) $d^-(v) \leqslant (1+\alpha)(n-1)/2$ *for every* $v \in G$*.*

Then G contains a γ-almost-regular subtournament G' on at least $(1-\gamma)n$ vertices.

Proof. We prove (i); then (ii), (iii) and (iv) follow immediately. Suppose that G has at least $\sqrt{\alpha}n$ vertices with $d^+(v) > (1+\sqrt{\alpha})(n-1)/2$. Then

$$\binom{n}{2} = e(G) = \sum_{v \in G} d^+(v) > (1-\alpha)\binom{n}{2} + \sqrt{\alpha}n \cdot \frac{\sqrt{\alpha}(n-1)}{2} = \binom{n}{2},$$

giving a contradiction. So there are at most $\sqrt{\alpha}n$ vertices of G with $d^+(v) > (1+\sqrt{\alpha})(n-1)/2$. Delete all of these vertices of G, and let G' be the obtained subtournament. Then $n - \sqrt{\alpha}n \leqslant |G'| \leqslant n$. Also, every vertex of G' has

$$d^+_{G'}(v) \geqslant \frac{(1-\alpha)(n-1)}{2} - \sqrt{\alpha}n \geqslant \frac{(1-\gamma)(|G'|-1)}{2}$$

and

$$d^-_{G'}(v) \geqslant n - 1 - \sqrt{\alpha}n - \frac{(1+\sqrt{\alpha})(n-1)}{2} \geqslant \frac{(1-\gamma)(|G'|-1)}{2}.$$

So G' is a γ-almost-regular tournament on at least $(1-\gamma)n$ vertices, as desired. $\square$

3.2.2 Some embedding results

The following three lemmas are the main tools we use to embed directed trees in tournaments. We use Theorem 1.2 in the proofs of all three lemmas, although the factor of 3 in Theorem 1.2 is not critical to our proof; any linear bound would suffice. For the proof of Lemma 3.4 we also require the use of Theorem 1.5(1).

Lemma 3.2 *Let T be a directed tree on n vertices, rooted at t, such that t has no inneighbours in T, and every component of $T - t$ contains at most d vertices. Let G be a tournament whose vertex set is partitioned into three sets, $\{v\}, N$ and X, where $|N| \geqslant n-1$, every vertex of N is an outneighbour of v, and at least $3d$ vertices of N each have at least $6d$ inneighbours in X and at least $6d$ outneighbours in X. Then T can be embedded in G in such a way that t is embedded to v and at most $4d$ vertices of X are occupied by this embedding.*

Proof. Let $N' \subseteq N$ consist of all vertices of N with at least $6d$ inneighbours in X and at least $6d$ outneighbours in X. Then $|N'| \geqslant 3d$. We begin by embedding t to the vertex v. Now let $T_1, \ldots, T_r$ be the components of $T - t$, in order of decreasing order. For each i, let t_i be the single vertex of T_i which is an outneighbour of t. Then we embed $T_1, \ldots, T_r$ in turn in $N \cup X$, with each t_i embedded in N and each T_i embedded in the vertices not occupied by the embeddings of $T_1, \ldots, T_{i-1}$. This gives an embedding of T in G. So suppose that we have embedded $T_1, \ldots, T_{i-1}$ in this manner, and we now wish to embed T_i. Then at most $n-1$ vertices of T have been embedded. At least one of these vertices (namely t) was not embedded in N, so at least one vertex of N must be unoccupied.

Suppose that N' contains at least one unoccupied vertex v_i, and also that fewer than $3d$ vertices of X have been occupied. Then v_i has at least $3d$ unoccupied inneighbours in X and at least $3d$ unoccupied outneighbours in X. Embed t_i to v_i. We then proceed through the outcomponents of t_i in T_i in turn. Suppose that when we come to embed an outcomponent of t_i we have previously embedded m vertices of T_i. Then the current outcomponent has order at most $d-m$. Also, v_i has at least $3d-m \geqslant 3(d-m)$ outneighbours in X which have not yet been occupied, so by Theorem 1.2 we may embed this outcomponent amongst the outneighbours of v_i in X. Similarly we may embed the incomponents of t_i in turn amongst the inneighbours of v_i in X, and so we obtain an embedding of T_i in the unoccupied vertices of G. Note that all vertices of T_i apart from t_i are embedded in X.

Now suppose instead that every vertex of N' has been occupied, but still that fewer than $3d$ vertices of X have been occupied. Then at least one of the T_j with $j < i$ must have had $|T_j| = 1$, and so T_i consists of one single vertex, namely t_i. We may therefore embed t_i to any unoccupied vertex of N (recall that there is at least one such vertex).

Finally, suppose that at least $3d$ vertices of X have been occupied. Then at least $3d+1$ vertices of T have been embedded outside N, and so N contains at least $n-1-(n-(3d+1)) = 3d$ unoccupied vertices. Since $|T_i| \leqslant d$, by Theorem 1.2 we may embed T_i among these unoccupied vertices.

By embedding each T_i in this fashion we obtain an embedding of T in G with t embedded to v. Furthermore, the only vertices embedded in X are those in some T_i such that when we came to embed T_i, N' contained at least one unoccupied vertex v_i, and fewer than $3d$ vertices of X had been occupied. The embedding of T_i occupied at most another d vertices of X, and so at most $4d$ vertices of X can have been occupied in total. □

Lemma 3.3 *(a) Let T be a directed tree, and let T_c be a subtree of T such that every component of $T - T_c$ contains at most d vertices. Let G be a tournament whose vertices are partitioned into two sets S and N such that for every vertex $v \in S$ we have*

(i) $|N^+(v) \cap N| \geqslant |T - T_c| + 2d$*, and*

(ii) $|N^-(v) \cap N| \geqslant |T - T_c| + 2d$.

Then any embedding of T_c in $G[S]$ can be extended to an embedding of T in G.

(b) Suppose that in addition to the above assumptions we choose a set $N' \subseteq N$ and an integer $r \leqslant |T - T_c|$, so that every vertex $v \in S$ satisfies

(iii) $|N^+(v) \cap N'| \geqslant r + 2d$*, and*

(iv) $|N^-(v) \cap N'| \geqslant r + 2d$.

Then any embedding of T_c in $G[S]$ can be extended to an embedding of T in G such that at least r vertices of T are embedded in N'.

(c) Suppose that no edges of T are directed from T_c to $T - T_c$. Then conditions (i) and (iii) may be dropped without affecting the validity of the above result. Likewise if no edges of T are directed from $T - T_c$ to T_c, then the above results hold even without conditions (ii) and (iv).

Proof. Let $n := |T|$. We prove (b) and (c); for (a), apply (b) with $r := |T - T_c|$ and $N' := N$. Let $T_1, \ldots, T_q$ be the components of $T - T_c$, so $|T_i| \leqslant d$ for each i. Suppose that we have successfully extended the embedding of T_c in $G[S]$ to an embedding of $T_c \cup T_1 \cup \cdots \cup T_{s-1}$ in G; we now demonstrate how to extend this embedding to an embedding of $T_c \cup T_1 \cup \cdots \cup T_s$ in G. Indeed, there is precisely one edge between T_c and T_s. Let $t \in T_c$ and $t_s \in T_s$ be the endvertices of this edge, and let v be the vertex in S to which t is embedded.

Suppose that t_s is an outneighbour of t. By (i), v has at least $|T - T_c| + 2d$ outneighbours in N. At most $|T_1| + \cdots + |T_{s-1}|$ of these outneighbours are occupied by the embedding of $T_c \cup T_1 \cup \cdots \cup T_{s-1}$, and so v has at least $|T_s| + 2d \geqslant 3|T_s|$ outneighbours in N which are not occupied by this embedding. Now, by (iii), v has at least $r + 2d$ outneighbours in N'. If at most $r - |T_s|$ of these outneighbours are occupied by the embedding of $T_c \cup T_1 \cup \cdots \cup T_{s-1}$, then by Theorem 1.2 we may embed T_s amongst the at least $2d + |T_s| \geqslant 3|T_s|$ unoccupied outneighbours of v in N'. If instead $r - k$ of these outneighbours are occupied, for some $1 \leqslant k \leqslant |T_s| - 1$, then by Theorem 1.2 we may embed T_s amongst the $2|T_s| + k$ unoccupied outneighbours in N' and some arbitrary $|T_s| - k$ outneighbours of v in $N \setminus N'$. Then at least k vertices of N' are occupied by this embedding of T_s. Finally, if at least r outneighbours of v in N' have been occupied by this embedding, then we may embed T_s within the at least $3|T_s|$ unoccupied outneighbours of v in N.

If instead t_s is an inneighbour of t, then we may extend the embedding similarly, using (ii) and (iv) rather than (i) and (iii). So we may extend the embedding of T_c in $G[S]$ to an embedding of T in G by proceeding through each T_i in this manner. Also conditions (i) and (iii) are only required if at least one edge of T is directed from T_c to $T - T_c$, and conditions (ii) and (iv) are only required if at least one edge of T is directed from $T - T_c$ to T_c. Finally, note that after each T_s is embedded, either every vertex of $T_1 \cup \cdots \cup T_s$ has been embedded in N', or at least r vertices of $T_1 \cup \cdots \cup T_s$ have been embedded in N'. Since $|T_1 \cup T_2 \cup \cdots \cup T_q| = |T - T_c| \geqslant r$, we can be sure that at least r vertices of N' are occupied by the embedding of T, as desired. $\square$

Lemma 3.4 *Suppose that $1/n \ll \gamma \ll \alpha \ll 1$. Let T be a directed tree on n vertices, and let* *forests F^- and F^+ be induced subgraphs of T such that $V(F^-)$ and $V(F^+)$ partition $V(T)$ and every edge between F^- and F^+ is directed from F^- to F^+. Let T_1^+ and T_2^+ be the largest and second largest components of F^+ respectively. Also, let Y and Z be disjoint sets such that*

$$|Y| \geqslant |F^+| + |T_2^+| + \alpha n \text{ and } |Z| \geqslant 2|F^-| + \alpha n.$$

Let G be a tournament on vertex set $Y \cup Z$ such that every vertex of Y has at most γn outneighbours in Z, and every vertex of Z has at most γn inneighbours in Y. Then any embedding of T_1^+ in $G[Y]$ can be extended to an embedding of T in G.

Proof. Let $T_1, \dots, T_r$ be the components of F^- and F^+, ordered so that $T_1 = T_1^+$ and so that for each $2 \leqslant i \leqslant r$ there is exactly one edge of T between T_i and $T_1 \cup \cdots \cup T_{i-1}$. Then we have an embedding of T_1 in $G[Y]$. We proceed through the trees T_i in turn, embedding each T_i in $G[Y]$ if T_i is a component of F^+, or in $G[Z]$ if T_i is a component of F^-. Each T_i is embedded so that the embeddings of $T_1, \dots, T_i$ form an embedding of the subtree of T induced by the vertices of $T_1, \dots, T_i$. Suppose that we have successfully embedded $T_1, \dots, T_{i-1}$ in this manner, and we wish to extend this embedding to include T_i. Note that there is precisely one edge e between T_i and $T_1 \cup \cdots \cup T_{i-1}$. Let t be the endvertex of e in $T_1 \cup \cdots \cup T_{i-1}$, and let v be the vertex to which t was embedded.

If T_i is a component of F^+, then $t \in F^-$, so $v \in Z$. In this case we embed T_i within the unoccupied outneighbours of v in Y. Since $v \in Z$, $|N^+(v) \cap Y| \geqslant |Y| - \gamma n \geqslant |F^+| + |T_2^+| + \alpha n/2$. At most $|F^+| - |T_i|$ of these vertices are occupied by the embeddings of $T_1, \dots, T_{i-1}$. Since $i \geqslant 2$, T_i is not the largest component of F^+, and so has order $|T_i| \leqslant |T_2^+|$. So at least $2|T_i| + \alpha n/2$ outneighbours of v in Y remain unoccupied. So if $|T_i| \geqslant \alpha n/2$ then by Theorem 1.5(1) we may embed T_i in these unoccupied vertices of $N^+(v) \cap Y$. On the other hand, if $|T_i| < \alpha n/2$ then by Theorem 1.2 we may embed T_i in these unoccupied vertices of $N^+(v) \cap Y$.

Now suppose instead that T_i is a component of F^-. Then $t \in F^+$, so $v \in Y$. Here we embed T_i within the unoccupied inneighbours of v in Z. Since $v \in Y$, $|N^-(v) \cap Z| \geqslant |Z| - \gamma n \geqslant 2|F^-| + \alpha n/2$, and at most $|F^-| - |T_i|$ of these vertices are occupied by the embeddings of $T_1, \ldots, T_{i-1}$. So at least $2|T_i| + \alpha n/2$ such vertices remain unoccupied. So as before, if $|T_i| \geqslant \alpha n/2$ then by Theorem 1.5(1) we may embed T_i in these unoccupied vertices of $N^-(v) \cap Z$, whereas if $|T_i| < \alpha n/2$ then by Theorem 1.2 we may embed T_i in these unoccupied vertices of $N^-(v) \cap Z$. By proceeding through all of the trees T_i in this manner we obtain an embedding of T in G.

$\square$

Observe that if in the statement of Lemma 3.4 we let T_1^- and T_2^- be the largest and second-largest components of F^- respectively, and replaced the conditions on the sizes of Z and Y by the conditions that $|Y| \geqslant 2|F^+| + \alpha n$ and $|Z| \geqslant |F^-| + |T_2^-| + \alpha n$, then we could conclude that any embedding of T_1^- in $G[Z]$ can be extended to an embedding of T in G. To see this, either note that the proof is still valid with appropriate changes (switching inneighbours and outneighbours and so forth) or observe that this is the effect of reversing the direction of every edge of T and every edge of G, in which case the embedding problem is the same. Sometimes when referring to Lemma 3.4 we implicitly mean this 'dual' of Lemma 3.4 instead.

3.3 Embedding trees whose core tree is a single vertex

In this section we verify that Sumner's universal tournament conjecture holds for large directed trees T whose core tree T_Δ contains only one vertex, *i.e.* trees which are 'star-shaped'. Such trees can be embedded by selecting an appropriate vertex to which to embed the single vertex of T_Δ, and then using Lemma 3.3 or Lemma 3.2 to extend this embedding to an embedding of T in G.

Lemma 3.5 *Suppose that $1/n \ll 1/\Delta \ll 1$. Let T be a directed tree on n vertices with*

$|T_\Delta| = 1$*, and let* G *be a tournament on* $2n-2$ *vertices. Then* G *contains a copy of* T*.*

Proof. Introduce constants α and γ with $1/\Delta \ll \alpha \ll \gamma \ll 1$. Let t be the single vertex of T_Δ, let y be the outweight of T_Δ, and let z be the inweight of T_Δ. Also, let T_1 be the subtree of T formed by t and all of its outcomponents, and let T_2 be the subtree of T formed by t and all of its incomponents. Then $y + z = n - 1$, $|T_1| = y + 1$ and $|T_2| = z + 1$. Now, suppose that G contains a vertex v such that

(i) either $d^+(v) \geqslant y + 2n/\Delta$ or $y = 0$, and

(ii) either $d^-(v) \geqslant z + 2n/\Delta$ or $z = 0$.

Then embed t to v. By Proposition 2.19(iv) each component of $T - t$ contains at most n/Δ vertices. So by Lemma 3.3(c) we may extend the embedding of t in $\{v\}$ to an embedding of T_1 in $\{v\} \cup N^+(v)$ (since if $y = 0$ then T_1 consists of the single vertex t). Also by Lemma 3.3(c), we may extend the embedding of t in $\{v\}$ to an embedding of T_2 in $\{v\} \cup N^-(v)$ (since if $z = 0$ then t is the only vertex of T_2). These two embeddings only overlap in the vertex v, and so combining these two embeddings gives an embedding of T in G.

So we may assume that every vertex $v \in G$ has either $d^+(v) < y + 2n/\Delta$ or $d^-(v) < z + 2n/\Delta$. Let $Y := \{v \in G : d^+(v) < y + 2n/\Delta\}$ and let $Z := \{v \in G : d^-(v) < z + 2n/\Delta\}$. Then every vertex of G lies in precisely one of Y and Z, so $|Y| + |Z| = 2n - 2$. Thus we must have either $|Y| \geqslant 2y$ or $|Z| \geqslant 2z$. Furthermore, if $y = 0$ and $|Y| \geqslant 1$ then each $v \in Y$ has $d^+(v) < 2n/\Delta$ and therefore $d^-(v) \geqslant z + 2n/\Delta$, and so satisfies (ii). We may therefore assume that if $y = 0$ then $|Y| = 0$ and similarly that if $z = 0$ then $|Z| = 0$. So without loss of generality we may assume that $|Y| \geqslant 2y$ and $y > 0$ (otherwise reverse the direction of every edge of T and every edge of G; then we would have $|Y| \geqslant 2y$ and $y > 0$ at this stage, and the embedding problem is the same). Observe that by definition of Y we must also have $|Y| \leqslant 2y + 4n/\Delta + 1$.

Now suppose that $y \geqslant \alpha n$. Since $y \in \mathbb{N}$ and $|Y| \geqslant 2y$, Y must contain a vertex v which satisfies $|N^+(v) \cap Y| \geqslant y$. Choose a subset $N' \subseteq N^+(v) \cap Y$ of size y. For any vertex $u \in Y$,

$$d^+_{G[Y]}(u) = |N^+(u) \cap Y| \leqslant d^+_G(u) < y + \frac{2n}{\Delta} \leqslant \frac{(1+\alpha)(|Y|-1)}{2}.$$

So by Proposition 3.1 $G[Y]$ contains a γ-almost-regular tournament on at least $2(1-\gamma)y$ vertices. So at most $|Y| - 2(1-\gamma)y \leqslant 3\gamma y$ vertices of Y have fewer than $(1-2\gamma)y$ inneighbours in Y or fewer than $(1-2\gamma)y$ outneighbours in Y. Since $|N'| = y$, at most $6\gamma y + 1$ vertices of N' have more than $(1-3\gamma)y$ inneighbours in N', and at most $6\gamma y + 1$ vertices of N' have more than $(1-3\gamma)y$ inneighbours in N'. So at least $(1-16\gamma)y$ vertices of N' have at least γy inneighbours in $Y \setminus N'$ and at least γy outneighbours in $Y \setminus N'$. Certainly therefore at least $3n/\Delta$ vertices of N' have at least $6n/\Delta$ inneighbours in $Y \setminus (\{v\} \cup N')$ and at least $6n/\Delta$ outneighbours in $Y \setminus (\{v\} \cup N')$. So by Lemma 3.2 we may embed T_1 in Y, with t embedded to v, and at most $4n/\Delta$ vertices embedded outside $N' \cup \{v\}$. Let V' be the set of vertices of G not occupied by this embedding of T_1. Since v has at least $|G| - 1 - (y + 2n/\Delta) \geqslant z + 6n/\Delta$ inneighbours in G, all outside $N' \cup \{v\}$, v must have at least $z + 2n/\Delta$ unoccupied inneighbours in V'. So by Lemma 3.3(c) we may extend the embedding of t in $\{v\}$ to an embedding of T_2 in $\{v\} \cup V'$. These two embeddings only overlap in the vertex v, and so combine to give an embedding of T in G.

So we may assume that $1 \leqslant y < \alpha n$. Then every vertex $v \in Y$ has

$$d^-(v) \geqslant |G| - 1 - y - \frac{2n}{\Delta} \geqslant n + \frac{2n}{\Delta}. \tag{3.6}$$

Let T_3 be the subtree of T formed by every vertex $t' \in T$ for which T contains a directed path from t to t'. Then $t \in T_3$, and (taking t as the root vertex) T_3 is an outbranching. Also $T_3 \subseteq T_1$, so $|T_3| \leqslant y + 1$, and so by Theorem 1.3, we may embed T_3 in $G[Y]$. Since $T_\Delta \subseteq T_3$, by Proposition 2.19(iv) each component of $T - T_3$ contains at most n/Δ vertices. So as every edge of T between $T - T_3$ and T_3 is directed from $T - T_3$ to T_3, and also since by (3.6) every vertex of Y has at least $|T - T_3| + 2n/\Delta$ inneighbours which were not occupied by the embedding of T_3, we may extend the embedding of T_3 in $G[Y]$ to an embedding of T in G by Lemma 3.3(c).

□

3.4 The regularity lemma and its applications to embedding trees

In this section we use the regularity method to prove Lemma 3.14, which states that Theorem 1.1 holds in the case where T_Δ is small and G is almost-regular. Specifically, we consider directed trees T for which T_Δ is substantially smaller than the size of a cluster obtained by applying the regularity lemma to a tournament G; our approach here is essentially to select an appropriate cluster or pair of clusters of G in which to embed T_Δ so that we may then embed the components of $T - T_\Delta$ in the remaining clusters of G. We finish the section by combining Lemmas 3.14 and 2.32 to obtain Lemma 3.19, which states that Theorem 1.1 holds whenever G is almost-regular.

To prove Lemma 3.14 we begin by applying the regularity lemma to partition the vertices of G into clusters $V_1, \dots, V_k$. We deduce from the fact that G is almost-regular that every cluster V_i has roughly the same number of edges entering it as leaving it. In Lemma 3.9 we consider the case where some cluster V_i has the property that for any other cluster V_j there are many edges from V_i to V_j and many edges from V_j to V_i. Here we show that may embed T in G by first embedding T_Δ within the cluster V_i, and then using the regularity of edges between pairs of clusters to embed the remaining vertices of T. In Lemma 3.13 we instead consider the case where there is no such cluster. Here we show that we may embed T in G by the same argument provided that T_Δ has large inweight and outweight. So it remains only to show that

Lemma 3.14 holds in the case where T_Δ has small outweight. In this case we select two clusters V_i and V_j for which the corresponding vertices in the 'reduced digraph' of G has small common outneighbourhood. We then embed the vertices of T_Δ within these two clusters, before using the regularity of edges between pairs of clusters to embed the remaining vertices of T in G. So in terms of the description of the regularity method given in the introduction, the subgraph we find in the reduced digraph is a single vertex or pair of vertices with given properties on the edges entering and leaving the corresponding cluster or clusters of G.

We say that an oriented graph G on clusters $V_1, \dots, V_k$ of equal size is an *ε-regular cluster tournament* if for any $i, j \in [k]$ with $i \neq j$ the subdigraph $G[V_i \to V_j]$ is ε-regular and for any $i \in [k]$ the subdigraph $G[V_i]$ is a tournament. If G is a cluster tournament on clusters $V_1, \dots, V_k$ then we denote the density of $G[V_i \to V_j]$ by d_{ij} for any $i, j \in [k]$ (the tournament G should be clear from the context). The following corollary of the digraph regularity lemma (Lemma 2.7) shows that any sufficiently large tournament G contains an almost-spanning ε-regular cluster tournament G^* such that all vertices have similar in- and outdegrees in both G and G^*.

Corollary 3.7 *Suppose that $1/n \ll 1/M \ll 1/M' \ll \varepsilon$. Let G be a tournament on n vertices. Then there exist disjoint subsets $V_1, \dots, V_k \subseteq V(G)$ of equal size and a subgraph $G^* \subseteq G$ on vertex set $V_1 \cup \dots \cup V_k$ such that:*

(i) $M' \leqslant k \leqslant M$,

(ii) G^* *is an ε-regular cluster tournament,*

(iii) $\bigcup_{i \in [k]} V_i \geqslant (1 - \varepsilon)n$,

(iv) $d^+_{G^*}(x) > d^+_G(x) - 2\varepsilon n$ *for all vertices* $x \in V(G)$, *and*

(v) $d^-_{G^*}(x) > d^-_G(x) - 2\varepsilon n$ *for all vertices* $x \in V(G)$.

Proof. Apply Lemma 2.7 with $d = 0$ to obtain a partition $V_0, \dots, V_k$ of $V(G)$ and a subgraph $G' \subseteq G$ which satisfy the conditions of Lemma 2.7. In particular (i) and (iii) are satisfied. Now form G^* from $G'[V_1 \cup \dots \cup V_k]$ by adding every edge of G for which both endvertices lie in the same cluster V_i. So $G^* \subseteq G$, and by (7) of Lemma 2.7 and the fact that $G^*[V_i]$ is a tournament for each $i \in [k]$ we have (ii). Finally note that using (4) of Lemma 2.7 we have

$$d^+_{G^*}(x) \geqslant d^+_{G'}(x) - |V_0| \geqslant d^+_G(x) - 2\varepsilon n.$$

Similarly $d^-_{G^*}(x) \geqslant d^-_G(x) - 2\varepsilon n$ using (5) of Lemma 2.7. □

It follows immediately from the definition of regularity that if U and V are sets of size m, and $G[U \to V]$ is ε-regular with density d, then all but at most $2\varepsilon m$ vertices of U have $(d \pm \varepsilon)m$ outneighbours in V. The next lemma is a generalisation of this fact, considering the number of outneighbours of vertices in one cluster within a cluster tournament.

Lemma 3.8 *Suppose that $1/m \ll 1/k \ll \varepsilon \ll \varepsilon' \ll 1$. Let G be an ε-regular cluster tournament on clusters $V_1, \dots, V_k$, each of size m. Let $V'_j \subseteq V_j$ for each $j \in [k]$ be fixed. Then for any i, all but at most $\varepsilon' m$ vertices of V_i have $\sum_{j \in [k] \setminus \{i\}} d_{ij}|V'_j| \pm \varepsilon' km$ outneighbours in $\bigcup_{j \in [k] \setminus \{i\}} V'_j$ and $\sum_{j \in [k] \setminus \{i\}} d_{ji}|V'_j| \pm \varepsilon' km$ inneighbours in $\bigcup_{j \in [k] \setminus \{i\}} V'_j$.*

Proof. Fix some $i \in [k]$. Then let L be the set of all $j \in [k] \setminus \{i\}$ such that $|V'_j| \geqslant \varepsilon m$ and $d_{ij} \geqslant \sqrt{\varepsilon}$. For each $j \in L$, let A_j denote the set of vertices of V_i which have fewer than $(1 - \sqrt{\varepsilon})d_{ij}|V'_j|$ outneighbours in V'_j. Then for each $j \in L$, the subdigraph of $G[V_i \to V_j]$ induced by A_j and V'_j has density less than $(1 - \sqrt{\varepsilon})d_{ij} \leqslant d_{ij} - \varepsilon$. Since $G[V_i \to V_j]$ is ε-regular with density d_{ij}, and $|V'_j| \geqslant \varepsilon m$, we must have $|A_j| < \varepsilon m$.

Now, fix a vertex $v \in V_i$. Suppose that v appears in at most $\sqrt{\varepsilon}|L|$ of the sets A_j with $j \in L$.

Then

$$\begin{aligned}
|N^+(v) \cap \bigcup_{j \in L} V'_j| &\geqslant \sum_{j \in L : v \notin A_j} (1-\sqrt{\varepsilon}) d_{ij} |V'_j| \\
&\geqslant \sum_{j \in [k] \setminus \{i\}} (1-\sqrt{\varepsilon}) d_{ij} |V'_j| - \sum_{j \in [k] \setminus (L \cup \{i\})} d_{ij} |V'_j| - \sum_{j \in L : v \in A_j} d_{ij} |V'_j| \\
&\geqslant \sum_{j \in [k] \setminus \{i\}} d_{ij} |V'_j| - \sqrt{\varepsilon} km - \sqrt{\varepsilon} km - \sqrt{\varepsilon} |L| m \\
&\geqslant \sum_{j \in [k] \setminus \{i\}} d_{ij} |V'_j| - 3\sqrt{\varepsilon} km.
\end{aligned}$$

Since at most $\sqrt{\varepsilon}m$ vertices $v \in V_i$ appear in more than $\sqrt{\varepsilon}|L|$ of the sets A_j with $j \in L$, we may conclude that there are at most $\sqrt{\varepsilon}m$ vertices $v \in V_i$ with fewer than $\sum_{j \in [k] \setminus \{i\}} d_{ij} |V'_j| - 3\sqrt{\varepsilon} km$ outneighbours in $\bigcup_{j \in [k] \setminus \{i\}} V'_j$. A similar argument shows that there are at most $\sqrt{\varepsilon}m$ vertices $v \in V_i$ with more than $\sum_{j \in [k] \setminus \{i\}} d_{ij} |V'_j| + 3\sqrt{\varepsilon} km$ outneighbours in $\bigcup_{j \in [k] \setminus \{i\}} V'_j$.

Now, let L' be the set of all $j \in [k]$ such that $|V'_j| \geqslant \varepsilon m$ and $d_{ji} \geqslant \sqrt{\varepsilon}$. Then the same argument applied to inneighbours rather than outneighbours shows that there are at most $\sqrt{\varepsilon}m$ vertices $v \in V_i$ with fewer than $\sum_{j \in [k] \setminus \{i\}} d_{ji} |V'_j| - 3\sqrt{\varepsilon} km$ inneighbours in $\bigcup_{j \in [k] \setminus \{i\}} V'_j$ and at most $\sqrt{\varepsilon}m$ vertices $v \in V_i$ with more than $\sum_{j \in [k] \setminus \{i\}} d_{ji} |V'_j| + 3\sqrt{\varepsilon} km$ inneighbours in $\bigcup_{j \in [k] \setminus \{i\}} V'_j$. Since $\varepsilon \ll \varepsilon'$, this completes the proof. □

The next two lemmas are used in the proof of Lemma 3.14; we state them separately as we also refer to them in Section 3.5. Both of these consider an ε-regular cluster tournament G on k clusters with the property that for some cluster V_i the density of edges leaving V_i and the density of edges entering V_i are each roughly $1/2$. Lemma 3.9 considers the case where for many clusters V_j the density of edges between V_i and V_j is large in both directions, showing that in this case G contains a copy of a directed tree T of the type considered. Lemma 3.13 considers the alternative, namely that for almost all clusters V_j the density of edges between V_i

and V_j is small in one direction, showing that in this case G contains a copy of T provided that T_Δ has large inweight and large outweight.

Lemma 3.9 *Suppose that* $1/n \ll 1/\Delta', \beta \ll 1/k \ll \varepsilon \ll \gamma \ll \alpha \ll 1/\Delta \ll 1$. *Let* T *be a directed tree on* n *vertices with* $|T_{\Delta'}| \leqslant \beta n$ *and* $|T_\Delta| \geqslant 2$, *and let* G *be an* ε*-regular cluster tournament on clusters* $V_1, \dots, V_k$, *each of size* $m \geqslant 2(1-\gamma)n/k$. *Suppose also that for some* $i \in [k]$ *we have*

$$\sum_{j \in [k]\setminus\{i\}} d_{ij} \geqslant \frac{(1-3\gamma)k}{2} \quad \textit{and} \sum_{j \in [k]\setminus\{i\}} d_{ji} \geqslant \frac{(1-3\gamma)k}{2},$$

and also that there are at least αk *values of* $j \in [k] \setminus \{i\}$ *such that* $d_{ij} \geqslant \alpha$ *and* $d_{ji} \geqslant \alpha$. *Then* G *contains a copy of* T.

Proof. Fix such a value of i, and introduce a new constant ε' with $\varepsilon \ll \varepsilon' \ll \gamma$. Since $\Delta \leqslant \Delta'$, we must have $T_\Delta \subseteq T_{\Delta'}$. Also, since $|T_\Delta| \geqslant 2$, we may choose an edge $t^- \to t^+$ of T_Δ, which therefore is also an edge of $T_{\Delta'}$. Let T^+ and T^- be the two components formed when this edge is deleted from T, labelled so that $t^+ \in T^+$ and $t^- \in T^-$. Similarly, let $T^+_{\Delta'}$ and $T^-_{\Delta'}$ be the two components formed by the deletion of the edge $t^- \to t^+$ from $T_{\Delta'}$, labelled with $t^+ \in T^+_{\Delta'}$ and $t^- \in T^-_{\Delta'}$. Then T^+ and T^- partition the vertices of T, and there is precisely one edge of T between T^+ and T^-, which is directed towards T^+. Furthermore, since $t^- \to t^+$ was an edge of T_Δ, by Proposition 2.19(ii) we have $|T^+|, |T^-| \geqslant n/\Delta$.

Let $J \subseteq [k] \setminus \{i\}$ satisfy $|J| \geqslant \alpha k$ and also that for any $j \in J$ we have $d_{ij} \geqslant \alpha$ and $d_{ji} \geqslant \alpha$. Then $\sum_{j \in J} d_{ij} \geqslant \alpha^2 k$ and $\sum_{j \in J} d_{ji} \geqslant \alpha^2 k$. By Lemma 3.8 (applied with $V'_j = \emptyset$ for each $j \notin J$) at most $\varepsilon' m$ vertices of V_i have fewer than

$$\sum_{j \in J} d_{ij} m - \varepsilon' k m \geqslant \alpha^2 k m - \varepsilon' k m \geqslant \frac{\alpha^2 k m}{2} \tag{3.10}$$

outneighbours in $\bigcup_{j\in J} V_j$ or fewer than $\sum_{j\in J} d_{ji}m-\varepsilon' km \geqslant \alpha^2 km/2$ inneighbours in $\bigcup_{j\in J} V_j$. Also by Lemma 3.8 at most $\varepsilon' m$ vertices of V_i have fewer than

$$\sum_{j\in[k]\setminus\{i\}} d_{ij}m - \varepsilon' km \geqslant \frac{(1-3\gamma-2\varepsilon')km}{2} \geqslant (1-5\gamma)n \tag{3.11}$$

outneighbours in $\bigcup_{j\in[k]\setminus\{i\}} V_j$ or fewer than $\sum_{j\in[k]\setminus\{i\}} d_{ji}m - \varepsilon' km \geqslant (1-5\gamma)n$ inneighbours in $\bigcup_{j\in[k]\setminus\{i\}} V_j$. Finally, at most $m/2+1$ vertices of V_i have fewer than $m/4$ inneighbours in V_i. So we may choose a set S^+ of $m/10$ vertices of V_i which do not fall into any of these categories. Since $|T^+_{\Delta'}| \leqslant |T_{\Delta'}| \leqslant \beta n \leqslant m/30$, by Theorem 1.2 we may embed $T^+_{\Delta'}$ in S^+. Let $S^+_{\Delta'}$ be the set of vertices of S^+ occupied by this embedding of $T^+_{\Delta'}$, and let v^+ be the vertex to which t^+ was embedded. Recall that $|T^-| \geqslant n/\Delta$, so

$$|T^+| = n - |T^-| \leqslant (1-\frac{1}{\Delta})n.$$

Furthermore, every component of $T^+ - T^+_{\Delta'}$ is a component of $T - T_{\Delta'}$ and thus has order at most n/Δ' by Proposition 2.19(iv). So by (3.10) and (3.11), and since $\gamma \ll 1/\Delta$, we may apply Lemma 3.3(b) to extend the embedding of $T^+_{\Delta'}$ in $S^+_{\Delta'}$ to an embedding of T^+ in $S^+_{\Delta'} \cup \bigcup_{j\in[k]\setminus\{i\}} V_j$ so that at least $\alpha^2 n/3$ vertices of $\bigcup_{j\in J} V_j$ are occupied by this embedding of T^+.

Now, at least $m/4 - m/10 = 3m/20$ vertices of $V_i \setminus S^+_{\Delta'}$ are inneighbours of v^+. For each $j \in [k]\setminus\{i\}$, let o_j denote the number of vertices of V_j which are occupied by our embedding of T^+, and let $V'_j \subseteq V_j$ consist of those vertices of V_j which are not occupied by this embedding. So $|V'_j| = m - o_j$ for each j. Note that since $d_{ij}+d_{ji} \leqslant 1$ we have $d_{ij} \leqslant 1-\alpha$ for each $j \in J$.

Then by Lemma 3.8, at most $\varepsilon' m$ vertices of V_i have fewer than

$$\begin{aligned}\sum_{j\in[k]\setminus\{i\}} d_{ij}(m-o_j) - \varepsilon' km &\geqslant \sum_{j\in[k]\setminus\{i\}} d_{ij}m - \varepsilon' km - \sum_{j\in J} d_{ij}o_j - \sum_{j\in[k]\setminus(\{i\}\cup J)} d_{ij}o_j \\ &\overset{(3.11)}{\geqslant} (1-5\gamma)n - (1-\alpha)\sum_{j\in J} o_j - \sum_{j\in[k]\setminus(\{i\}\cup J)} o_j \\ &\geqslant (1-5\gamma)n - \sum_{j\in[k]\setminus\{i\}} o_j + \alpha\sum_{j\in J} o_j \\ &\geqslant (1-5\gamma)n - \sum_{j\in[k]\setminus\{i\}} o_j + \alpha^3 n/3 \geqslant n - \sum_{j\in[k]\setminus\{i\}} o_j + \frac{2n}{\Delta'}\end{aligned} \tag{3.12}$$

outneighbours in $\bigcup_{j\in[k]\setminus\{i\}} V_j'$ or fewer than

$$\sum_{j\in[k]\setminus\{i\}} d_{ji}(m-o_j) - \varepsilon' km \geqslant n - \sum_{j\in[k]\setminus\{i\}} o_j + \frac{2n}{\Delta'},$$

inneighbours in $\bigcup_{j\in[k]\setminus\{i\}} V_j'$. So we may choose a set S^- of $m/10$ vertices of $V_i \setminus S^+_{\Delta'}$, none of which fall into these two categories, and all of which are inneighbours of v^+. Since $|T^-_{\Delta'}| \leqslant |T_{\Delta'}| \leqslant \beta n \leqslant m/30$, by Theorem 1.2 we may embed $T^-_{\Delta'}$ in S^-. Let $S^-_{\Delta'}$ be the set of vertices of S^- occupied by this embedding of $T^-_{\Delta'}$. Then since

$$|T^-| = n - |T^+| \leqslant n - \sum_{j\in[k]\setminus\{i\}} o_j,$$

the right hand side of (3.12) is at least $|T^-| + 2n/\Delta'$. Also every component of $T^- - T^-_{\Delta'}$ is a component of $T - T_{\Delta'}$ (and so has order at most n/Δ' by Proposition 2.19(iv)). So by Lemma 3.3(b) we may extend the embedding of $T^-_{\Delta'}$ in $S^-_{\Delta'}$ to an embedding of T^- in $S^-_{\Delta'} \cup \bigcup_{j\in[k]\setminus\{i\}} V_j'$. Then the embeddings of T^+ and T^- do not overlap, and so together these embeddings form an embedding of T in G. □

Given an ε-regular cluster tournament G on clusters $V_1, \dots, V_k$, we define the *reduced digraph of G with parameter d*, denoted $R_G(d)$, to be the directed graph on vertex set $[k]$ in which $i \to j$ if and only if $d_{ij} \geqslant d$. Observe that since $d_{ij} + d_{ji} \leqslant 1$ for any i and j, if $d > 1/2$ then $R_G(d)$ is an oriented graph.

Lemma 3.13 *Suppose that $1/n \ll 1/\Delta', \beta \ll 1/k \ll \varepsilon \ll \gamma \ll \alpha \ll 1$. Let T be a directed tree on n vertices with $|T_{\Delta'}| \leqslant \beta n$, and let y and z be the outweight and inweight of $T_{\Delta'}$ respectively. Let G be an ε-regular cluster tournament on clusters $V_1, \dots, V_k$, each of size $m \geqslant 2(1-\gamma)n/k$. Suppose that for some $i \in [k]$ we have*

$$\sum_{j \in [k] \setminus \{i\}} d_{ij} \geqslant \frac{(1-3\gamma)k}{2} \quad \text{and} \sum_{j \in [k] \setminus \{i\}} d_{ji} \geqslant \frac{(1-3\gamma)k}{2},$$

and also that there are at most αk values of $j \in [k] \setminus \{i\}$ such that $d_{ij} \geqslant \alpha$ and $d_{ji} \geqslant \alpha$. Then:

(i) There are at most $2\alpha k$ values of $j \in [k] \setminus \{i\}$ such that $d_{ij} < 1 - 2\alpha$ and $d_{ji} < 1 - 2\alpha$.

(ii) Let $R := R_G(1-2\alpha)$. Then $|N^+_R(i)|, |N^-_R(i)| \geqslant (1-10\alpha)k/2$.

(iii) If $y, z \geqslant 14\alpha n$, then G contains a copy of T.

Proof. Fix such an i, and introduce a new constant ε' with $\varepsilon \ll \varepsilon' \ll \gamma$. For (i), note that since $d_{ij} + d_{ji} \leqslant 1$ for any $j \in [k] \setminus \{i\}$, and

$$\sum_{j \in [k] \setminus \{i\}} (d_{ij} + d_{ji}) \geqslant (1-3\gamma)k,$$

there are at most $3\sqrt{\gamma}k \leqslant \alpha k$ values of $j \in [k] \setminus \{i\}$ for which $d_{ij} + d_{ji} < 1 - \sqrt{\gamma}$. So there are at most $2\alpha k$ values of $j \in [k] \setminus \{i\}$ for which $d_{ij} < 1 - \alpha - \sqrt{\gamma}$ and $d_{ji} < 1 - \alpha - \sqrt{\gamma}$, so (i) holds.

For (ii), observe that by (i) we have

$$\begin{aligned}\frac{(1-3\gamma)k}{2} \leqslant \sum_{j\in[k]\setminus\{i\}} d_{ij} &\leqslant \sum_{\substack{j\in[k]\setminus\{i\}\\ d_{ij}\geqslant 1-2\alpha}} d_{ij} + \sum_{\substack{j\in[k]\setminus\{i\}\\ d_{ij},d_{ji}<1-2\alpha}} d_{ij} + \sum_{\substack{j\in[k]\setminus\{i\}\\ d_{ij}\leqslant 2\alpha}} d_{ij}\\ &\leqslant |N_R^+(i)| + 2\alpha k + 2\alpha k,\end{aligned}$$

so $|N_R^+(i)| \geqslant (1-10\alpha)k/2$. A similar calculation shows that $|N_R^-(i)| \geqslant (1-10\alpha)k/2$.

For (iii), let N^+ and N^- denote $N_R^+(i)$ and $N_R^-(i)$ respectively, and let $V^+ := \bigcup_{j\in N^+} V_j$ and $V^- := \bigcup_{j\in N^-} V_j$, so V^+ and V^- are disjoint. By Lemma 3.8, V_i contains at most $\varepsilon' m$ vertices with fewer than

$$\begin{aligned}\sum_{j\in N^+} d_{ij}m - \varepsilon' km \geqslant |N_R^+(i)|(1-2\alpha)m - \varepsilon' km &\geqslant \frac{(1-10\alpha)(1-2\alpha)km}{2} - \varepsilon' km\\ &\geqslant \frac{(1-12\alpha-2\varepsilon')km}{2} \geqslant (1-13\alpha)n\end{aligned}$$

outneighbours in V^+ and at most $\varepsilon' m$ vertices with fewer than $\sum_{j\in N^-} d_{ji}m - \varepsilon' km \geqslant (1-13\alpha)n$ inneighbours in V^-. Choose a set S of $m/2$ vertices of V_i, not including any of these at most $2\varepsilon' m$ vertices. Since $|T_{\Delta'}| \leqslant \beta n \leqslant m/6$, by Theorem 1.2 we may embed $T_{\Delta'}$ in S. Let $S_{\Delta'}$ be the set of vertices of S occupied by this embedding of $T_{\Delta'}$. Also let T_1 be the tree formed by $T_{\Delta'}$ and all of its outcomponents, and let T_2 be the tree formed by $T_{\Delta'}$ and all of its incomponents. Note that all of these out- and incomponents have order at most $n/\Delta' \ll \alpha n$ by Proposition 2.19(iv). In addition $|T_1| = n - z \leqslant (1-14\alpha)n$ and $|T_2| = n - y \leqslant (1-14\alpha)n$. So by Lemma 3.3(c) we may extend the embedding of $T_{\Delta'}$ in $S_{\Delta'}$ to an embedding of T_1 in $S_{\Delta'} \cup V^+$. Similarly by Lemma 3.3(c) we may extend the embedding of $T_{\Delta'}$ in $S_{\Delta'}$ to an embedding of T_2 in $S_{\Delta'} \cup V^-$. Then these embeddings do not overlap outside $T_{\Delta'}$, so we may combine them to form an embedding of T in G. $\square$

The next lemma shows that Sumner's universal tournament conjecture holds whenever n is sufficiently large, G is an almost-regular tournament and the core tree T_Δ of T is small. Actually we prove a slightly stronger result in this case, considering a tournament on fewer than $2n-2$ vertices. Later on we make use of the fact that we have a little room to spare in the order of the tournament. Much of the work for this lemma is done by the two previous lemmas.

Lemma 3.14 *Suppose that* $1/n \ll 1/\Delta', \beta \ll \gamma \ll 1/\Delta \ll 1$. *Let* T *be a directed tree on* n *vertices such that* $|T_{\Delta'}| \leqslant \beta n$ *and* $|T_\Delta| \geqslant 2$. *Let* G *be a* γ*-almost-regular tournament on at least* $(2-\gamma)n$ *vertices. Then* G *contains a copy of* T.

Proof. Introduce new constants $\varepsilon, \varepsilon', \alpha, M$, and M' with

$$1/n \ll 1/\Delta', \beta \ll 1/M \ll 1/M' \ll \varepsilon \ll \varepsilon' \ll \gamma \ll \alpha \ll 1/\Delta \ll 1.$$

If $|G| \geqslant (2+\gamma)n$, then G contains a copy of T by Theorem 1.5(1). So we may assume that $|G| = (2 \pm \gamma)n$. Observe that $d^+(v), d^-(v) \geqslant (1-\gamma)(|G|-1)/2 \geqslant (1-2\gamma)n$ for all $v \in G$.

Since $\Delta \leqslant \Delta'$, we must have $T_\Delta \subseteq T_{\Delta'}$. Also, since $|T_\Delta| \geqslant 2$, we may choose an edge $t^- \to t^+$ of T_Δ, which must also lie in $T_{\Delta'}$. Let T^+ and T^- be the two components formed when this edge is deleted from T, labelled so that $t^+ \in T^+$ and $t^- \in T^-$. Similarly, let $T^+_{\Delta'}$ and $T^-_{\Delta'}$ be the two components formed by the deletion of the edge $t^- \to t^+$ from $T_{\Delta'}$, labelled with $t^+ \in T^+_{\Delta'}$ and $t^- \in T^-_{\Delta'}$. Then T^+ and T^- partition the vertices of T, and there is precisely one edge of T between T^+ and T^-, which is directed towards T^+. Furthermore, $|T^+|, |T^-| \geqslant n/\Delta$.

Let disjoint subsets $V_1, \ldots, V_k$ and a subgraph $G^* \subseteq G$ satisfy the conditions of Corollary 3.7. So $M' \leqslant k \leqslant M$, and G^* is an ε-regular cluster tournament on clusters $V_1, \ldots, V_k$ of equal size m, where

$$\frac{2(1-\gamma)n}{k} \leqslant \frac{(2-\gamma)n - 3\varepsilon n}{k} \leqslant m \leqslant \frac{(2+\gamma)n}{k}. \tag{3.15}$$

Also, for each $v \in G^*$ we have $d^+_{G^*}(v) \geqslant d^+_G(v) - 2\varepsilon|G| \geqslant d^+_G(v) - 5\varepsilon n$ and $d^-_{G^*}(v) \geqslant d^-_G(v) - 5\varepsilon n$. So for each $i \in [k]$ we have

$$\begin{aligned}\sum_{j\in[k]\setminus\{i\}} d_{ij} = \sum_{j\in[k]\setminus\{i\}} \frac{e_{G^*}(V_i \to V_j)}{m^2} &\geqslant \sum_{v\in V_i} \frac{d^+_{G^*}(v) - m}{m^2} \\ &\geqslant \sum_{v\in V_i} \frac{d^+_G(v) - 5\varepsilon n - m}{m^2} \geqslant \frac{(1-2\gamma)n - 5\varepsilon n - m}{m} \overset{(3.15)}{\geqslant} \frac{(1-3\gamma)k}{2},\end{aligned} \tag{3.16}$$

and similarly $\sum_{j\in[k]\setminus\{i\}} d_{ji} \geqslant (1-3\gamma)k/2$.

So if there exists some $i \in [k]$ for which there are at least αk values of $j \in [k] \setminus \{i\}$ such that $d_{ij} \geqslant \alpha$ and $d_{ji} \geqslant \alpha$, then by Lemma 3.9 we may embed T in G^*, and therefore in G. So we may assume that for each $i \in [k]$ fewer than αk values of $j \in [k] \setminus \{i\}$ satisfy $d_{ij} \geqslant \alpha$ and $d_{ji} \geqslant \alpha$. Then by Lemma 3.13(ii) we may assume that $R := R_G(1-2\alpha)$ has

$$\delta^0(R) \geqslant \frac{(1-10\alpha)k}{2}. \tag{3.17}$$

Let y be the number of vertices in outcomponents of $T_{\Delta'}$, and let z be the number of vertices in incomponents of $T_{\Delta'}$, so $y + z + |T_{\Delta'}| = n$. So if $y, z \geqslant 14\alpha n$ then G^* (and therefore G) contains a copy of T by Lemma 3.13(iii). We may therefore assume without loss of generality that $z < 14\alpha n$.

Now, since $|R| = k$ we may choose a vertex $i \in R$ with $d^+_R(i) \leqslant k/2$. Then we may choose a vertex $j \in N^+_R(i)$ with at most $d^+_R(i)/2$ outneighbours in $N^+_R(i)$. So $i \to j$ and $|N^+_R(i) \cap N^+_R(j)| \leqslant k/4$. For this choice of i and j, let

$$\begin{aligned} A &:= N^+_R(i) \cap N^+_R(j), \\ B &:= N^+_R(i) \setminus N^+_R(j), \\ C &:= N^+_R(j) \setminus N^+_R(i). \end{aligned}$$

Then A, B and C are disjoint, and $|B|, |C| \geqslant k/2 - 5\alpha k - |A| \geqslant k/4 - 5\alpha k$ by (3.17). Now, choose a set S^+ of $m/2$ vertices of V_j such that each vertex $v \in S^+$ has

(i) at least $m/2$ inneighbours in V_i,

(ii) at least $\sum_{\ell \in A} d_{j\ell} m - \varepsilon' km \geqslant (1-2\alpha)m|A| - \varepsilon' km$ outneighbours in $\bigcup_{\ell \in A} V_\ell$, and

(iii) at least $\sum_{\ell \in C} d_{j\ell} m - \varepsilon' km \geqslant (1-2\alpha)m|C| - \varepsilon' km$ outneighbours in $\bigcup_{\ell \in C} V_\ell$.

We can be sure that such a choice is possible, as by Lemma 3.8 there are at most $2\varepsilon' m$ vertices of V_j which fail either of (ii) and (iii), and since $G^*[V_i \to V_j]$ is ε-regular with density $d_{ij} \geqslant 1 - 2\alpha$ there are at most εm vertices of V_j which fail (i). Then since $|T^+_{\Delta'}| \leqslant \beta n \leqslant m/6$, by Theorem 1.2 we can embed $T^+_{\Delta'}$ in S^+. Let v^+ be the vertex to which t^+ is embedded. Then v^+ has at least $m/2$ inneighbours in V_i. Choose a set S^- of $m/3$ of these inneighbours so that every vertex $v \in S^-$ has at least

$$\sum_{\ell \in A \cup B = N^+_R(i)} d_{i\ell} m - \varepsilon' km \geqslant (1-2\alpha)m|N^+_R(i)| - \varepsilon' km \overset{(3.17)}{\geqslant} (1-13\alpha)n \tag{3.18}$$

outneighbours in $\bigcup_{\ell \in A \cup B} V_\ell$. Again we can be sure that such a choice is possible, since by Lemma 3.8 at most $\varepsilon' m$ vertices of V_i fail this condition. Then since $|T^-_{\Delta'}| \leqslant \beta n \leqslant m/9$, by Theorem 1.2 we can embed $T^-_{\Delta'}$ in S^-. Let $S^+_{\Delta'}$ and $S^-_{\Delta'}$ be the sets of vertices of G occupied by $T^+_{\Delta'}$ and $T^-_{\Delta'}$ respectively.

Let T_3 be the tree formed by $T_{\Delta'}$ and all of its incomponents. Let T_4 be the tree formed by $T^+_{\Delta'}$ and all of its outcomponents, and let T_5 be the tree formed by $T^-_{\Delta'}$ and all of its outcomponents in T^- (*i.e.* all of its outcomponents except T^+). Note that $T_3 \cup T_4 \cup T_5 = T$. Then $|T_3| = |T_{\Delta'}| + z < 15\alpha n$, $|T_4| \leqslant |T^+| \leqslant n - |T^-| \leqslant (1 - 1/\Delta)n$, and similarly $|T_5| \leqslant (1 - 1/\Delta)n$. Every vertex of G has at least $(1 - 2\gamma)n$ inneighbours in G, so by Lemma 3.3(c) we may extend the embedding of $T_{\Delta'}$ in $S^+_{\Delta'} \cup S^-_{\Delta'}$ to an embedding of T_3 in G. For each $\ell \in [k] \setminus \{i\}$, let $V'_\ell \subseteq V_\ell$ consist of the vertices of V_ℓ which are not occupied by this embedding.

By (ii) and (iii), every vertex of $S^+_{\Delta'}$ then has at least $(1-2\alpha)(|A|+|C|)m - 2\varepsilon' km - |T_3| \geqslant (1-28\alpha)n$ outneighbours in $\bigcup_{\ell \in A \cup C} V'_\ell$ (here we also use the fact that $|A|+|C| = |N^+_R(j)| \geqslant (1-10\alpha)k/2$ by (3.17)). Since also $1/\Delta' \ll \alpha \ll 1/\Delta$ and every component of $T_4 - T^+_{\Delta'}$ has order at most n/Δ', by Lemma 3.3(c) we may extend the embedding of $T^+_{\Delta'}$ in $S^+_{\Delta'}$ to an embedding of T_4 in $S^+_{\Delta'} \cup \bigcup_{\ell \in A \cup C} V'_\ell$. Furthermore, since every vertex of $S^+_{\Delta'}$ has at least $(1-2\alpha)|C|m - \varepsilon' km - |T_3| \geqslant n/\Delta$ outneighbours in $\bigcup_{\ell \in C} V'_\ell$, and $|T_4 - T^+_{\Delta'}| = |T^+ - T_3| \geqslant n/2\Delta$, by Lemma 3.3(b) we can ensure that this embedding of T_4 occupies at least $n/2\Delta$ vertices of $\bigcup_{\ell \in C} V'_\ell$. So crucially at most $|T_4| - n/2\Delta$ vertices of T_4 are embedded in $\bigcup_{\ell \in A \cup B} V_\ell$. For each $\ell \in A \cup B$, let $V''_\ell \subseteq V_\ell$ consist of those vertices which are not occupied by the embedding of T_3 and T_4.

Finally, by (3.18), every vertex of $S^-_{\Delta'}$ has at least

$$(1-13\alpha)n - (|T_4| - \frac{n}{2\Delta}) - |T_3| \geqslant n - |T_4| + \frac{n}{3\Delta}$$

outneighbours in $\bigcup_{\ell \in A \cup B} V''_\ell$. Since $|T_5 - T^-_{\Delta'}| \leqslant n - |T_4|$, by Lemma 3.3(c) we can extend the embedding of $T^-_{\Delta'}$ in $S^-_{\Delta'}$ to an embedding of T_5 in $S^-_{\Delta'} \cup \bigcup_{\ell \in A \cup B} V''_\ell$. Then the embeddings of T_3, T_4 and T_5 do not overlap outside $S^+_{\Delta'} \cup S^-_{\Delta'}$, and so together form an embedding of T in G. $\square$

The following corollary of Lemmas 3.14 and 2.32 is the main result of this section, stating that Sumner's universal tournament conjecture holds with a little room to spare in the case where G is an almost-regular tournament and the core tree T_Δ of T has order greater than one (we need this extra room in the proof of Lemmas 3.21 and 3.23). Indeed, we show that a large almost-regular tournament G is also a robust outexpander, and so if T_Δ is large, then we can embed T

in G by Lemma 2.32. On the other hand, if T_Δ is small but has more than one vertex, then we may embed T in G by Lemma 3.14. In particular, together with Lemma 3.5 (which deals with the case $|T_\Delta| = 1$), this means that at this stage, we have proved that Sumner's conjecture holds for all large almost-regular tournaments.

Lemma 3.19 *Suppose that $1/n \ll \gamma \ll 1/\Delta \ll 1$. Let T be a directed tree on n vertices with $|T_\Delta| > 1$. Then every γ-almost-regular tournament G on at least $(2-\gamma)n$ vertices contains a copy of T.*

Proof. Introduce constants $\mu, \nu, \eta, \Delta', \beta, \gamma'$ such that

$$\frac{1}{n} \ll \frac{1}{\Delta'} \ll \mu \ll \nu \ll \eta \ll \gamma \ll \beta \ll \gamma' \ll \frac{1}{\Delta} \ll 1.$$

Let G be a γ-almost-regular tournament on at least $(2-\gamma)n$ vertices. Then we show that G is a robust (μ,ν)-outexpander. Indeed, let $S \subseteq V(G)$ satisfy $\nu|G| \leqslant |S| \leqslant 2|G|/3$. Then at least $(1-\gamma)|S|(|G|-1)/2$ edges originate in S. At most $\binom{|S|}{2}$ of these have both endvertices in S, so at least $(1-\gamma)|S|(|G|-1)/2 - \binom{|S|}{2} \geqslant |S|((1-\gamma)(|G|-1)-|S|)/2 \geqslant \nu|G|^2/10$ edges leave S. So at least $\nu|G|/20 \geqslant 3\mu|G|$ vertices outside S have at least $\nu|G|/20 \geqslant 3\mu|G|$ inneighbours in S. At most $2\mu|G|$ vertices of S have fewer than $\mu|G|$ inneighbours in S, and so $|RN^+_\mu(S)| \geqslant |S| + \mu|G|$, as desired. On the other hand, if $S \subseteq V(G)$ satisfies $2|G|/3 < |S| \leqslant (1-\nu)|G|$, every vertex of G has at least $|G|/7 \geqslant \mu|G|$ inneighbours in S. So $|RN^+_\mu(S)| = |G| \geqslant |S| + \mu|G|$, as desired.

So G is indeed a robust (μ,ν)-outexpander. Clearly $\delta^0(G) \geqslant \eta|G|$. So if $|T_{\Delta'}| \geqslant \beta n$, then by Lemma 2.32, G contains a copy of T. So we may assume that $|T_{\Delta'}| \leqslant \beta n$. But G is also a γ'-almost-regular tournament on at least $(2-\gamma')n$ vertices, and so by Lemma 3.14, G contains a copy of T. □

3.5 Embedding trees whose core tree is small

We now turn our attention to the general case of the problem. As when considering almost-regular tournaments, we consider the problem of embedding directed trees whose core trees are small separately from the case when the core trees are large. In this section we consider directed trees with small core trees, proving the following lemma.

Lemma 3.20 *Suppose $1/n \ll \beta, 1/\Delta' \ll 1$. Let T be a directed tree on n vertices with $|T_{\Delta'}| \leqslant \beta n$, and let G be a tournament on $2n-2$ vertices. Then G contains a copy of T.*

We begin by showing that we may assume that the tournament G consists of two large disjoint almost-regular tournaments, with almost all of the edges between them directed the same way.

Lemma 3.21 *Suppose that $1/n \ll \beta, 1/\Delta \ll \gamma \ll \eta \ll 1$. Let T be a directed tree on n vertices with $|T_\Delta| \leqslant \beta n$, and let G be a tournament on $2n-2$ vertices. Let y be the outweight of T_Δ, and let z be the inweight of T_Δ. Then the following properties hold.*

(i) If $z < \eta n$ or $y < \eta n$ then G contains a copy of T.

(ii) Either G contains a copy of T, or we can find disjoint sets $Y, Z \subseteq V(G)$ such that $|Y| \geqslant (2-\gamma)y$ and $|Z| \geqslant (2-\gamma)z$, $G[Y]$ and $G[Z]$ are γ-almost-regular, any vertex of Y has at most $3\gamma n$ outneighbours in Z and any vertex of Z has at most $3\gamma n$ inneighbours in Y.

Proof. Introduce new constants $M, M', \varepsilon, \varepsilon', \alpha, \gamma^*$ and Δ^* such that

$$\frac{1}{n} \ll \beta, \frac{1}{\Delta} \ll \frac{1}{M} \ll \frac{1}{M'} \ll \varepsilon \ll \varepsilon' \ll \gamma \ll \alpha \ll \eta \ll \gamma^* \ll \frac{1}{\Delta^*} \ll 1.$$

Partition the vertex set of G into sets A, B, C, D, E such that:

$$A \subseteq \{v \in G : d^+(v) \leqslant y + \varepsilon n\},$$
$$B \subseteq \{v \in G : y + \varepsilon n < d^+(v) < n - \varepsilon n\},$$
$$C \subseteq \{v \in G : d^+(v), d^-(v) \geqslant n - \varepsilon n\},$$
$$D \subseteq \{v \in G : z + \varepsilon n < d^-(v) < n - \varepsilon n\},$$
$$E \subseteq \{v \in G : d^-(v) \leqslant z + \varepsilon n\}.$$

These subset relations may not all be equality, for example in the case where z is very small, when we have $y + \varepsilon n \geqslant n - \varepsilon n$. However, it is clear that each vertex $v \in G$ lies in at least one of these five sets, so we may choose such a partition of $V(G)$. Let $x := |T_\triangle|$, so $x + y + z = n$ and $x \leqslant \beta n$.

Suppose that $|B| \geqslant 3x$. Then by Theorem 1.2 we may embed $T_\triangle$ in $G[B]$. Let $S_\triangle \subseteq B$ be the set of vertices occupied by this embedding of $T_\triangle$. Then every vertex of $S_\triangle$ has at least $y+\varepsilon n-x \geqslant y+2n/\Delta$ outneighbours outside $S_\triangle$ and at least $|G|-x-(n-\varepsilon n) \geqslant y+z+2n/\Delta$ inneighbours outside $S_\triangle$. Let T_1 be the subtree of T formed by $T_\triangle$ and all outcomponents of $T_\triangle$, and let T_2 be the subtree of T formed by $T_\triangle$ and all incomponents of $T_\triangle$. Then $|T_1| = x+y$ and $|T_2| = x + z$. By Proposition 2.19(iv), all incomponents and outcomponents of $T_\triangle$ contain at most n/Δ vertices, so by Lemma 3.3(c) we may extend our embedding of $T_\triangle$ in $S_\triangle$ to an embedding of T_1 in G. Then each vertex of $S_\triangle$ still has at least $z + 2n/\Delta$ inneighbours outside $S_\triangle$ which are not occupied by this embedding of T_1, so by Lemma 3.3(c) we may also extend our embedding of $T_\triangle$ in $S_\triangle$ to an embedding of T_2 in G which avoids vertices occupied by the embedding of $T_1 - T_\triangle$. Then these embeddings of T_1 and T_2 do not overlap outside $T_\triangle$, and so together form an embedding of T in G. We may therefore assume that $|B| < 3x \leqslant 3\beta n$. By the same argument (embedding first T_2 and then T_1 in G) we may assume that $|D| < 3x \leqslant 3\beta n$.

If $|T_{\Delta^*}| = 1$, then G contains a copy of T by Lemma 3.5. So we may assume that $|T_{\Delta^*}| \geqslant 2$. Now, if $z < \eta n$, then every $v \in E$ satisfies $d^-(v) < (\eta + \varepsilon)n < 2\eta n$, so $|E| \leqslant 4\eta n + 1$, and so $|B \cup D \cup E| \leqslant 4\eta n + 1 + 6\beta n \leqslant 5\eta n$. Let $G' := G[A \cup C]$. Then $|G'| \geqslant 2n - 2 - 5\eta n$, and every vertex $v \in G'$ has $d^+(v) \leqslant n + \varepsilon n$. So by Proposition 3.1, G' contains a γ^*-almost-regular subtournament G' on at least $(2 - \gamma^*)n$ vertices. Since $|T_{\Delta^*}| \geqslant 2$, by Lemma 3.19 G' contains a copy of T, so G contains a copy of T also. If instead we have $y < \eta n$, then we may similarly embed T in $G[C \cup E]$. So if $z < \eta n$ or $y < \eta n$ then G contains a copy of T, completing the proof of (i). So for (ii), we may assume that $y, z \geqslant \eta n$.

Suppose now that $|C| \geqslant 5\varepsilon' n$. Let disjoint subsets $V_1, \dots, V_k$ and a subgraph $G^* \subseteq G$ satisfy the conditions of Corollary 3.7. So $M' \leqslant k \leqslant M$, and G^* is an ε-regular cluster tournament on clusters $V_1, \dots, V_k$ of equal size m, where

$$\frac{(1-\varepsilon)|G|}{k} \leqslant m \leqslant \frac{|G|}{k}.$$

We show that G^* has the property that for some $i \in [k]$ we have

$$\sum_{j \in ([k] \setminus \{i\})} d_{ij} \geqslant \frac{(1-3\varepsilon')k}{2} \quad \text{and} \sum_{j \in ([k] \setminus \{i\})} d_{ji} \geqslant \frac{(1-3\varepsilon')k}{2}. \tag{3.22}$$

Indeed, if for some $i \in [k]$ we have $\sum_{j \in ([k] \setminus \{i\})} d_{ij} < (1 - 3\varepsilon')k/2$, then by Lemma 3.8 all but at most $\varepsilon' m$ vertices of V_i have at most

$$\sum_{j \in ([k] \setminus \{i\})} d_{ij} m + \varepsilon' k m < \frac{(1-\varepsilon')km}{2} < n - 8\varepsilon n$$

outneighbours in $\bigcup_{j \in ([k] \setminus \{i\})} V_j$ (in the graph G^*), and hence at most $n - 8\varepsilon n + (|G| - |G^*|) + |V_i| + 2\varepsilon|G| < n - \varepsilon n$ outneighbours in G. So at most $\varepsilon' m$ vertices of V_i lie in C. Similarly if for some $i \in [k]$ we have $\sum_{j \in ([k] \setminus \{i\})} d_{ji} < (1 - 3\varepsilon')k/2$ then again at most $\varepsilon' m$ vertices of V_i lie in C. Since $|C| \geqslant 5\varepsilon' n > 2\varepsilon' mk + (|G| - |G^*|)$, there must be some $i \in [k]$ which satisfies

(3.22). Fix such an i. Then if at least αk values of $j \in [k] \setminus \{i\}$ have $d_{ij} \geqslant \alpha$ and $d_{ji} \geqslant \alpha$ then G^* contains a copy of T by Lemma 3.9 (applied with ε' in the place of γ). Alternatively, if at most αk values of $j \in [k] \setminus \{i\}$ have $d_{ij} \geqslant \alpha$ and $d_{ji} \geqslant \alpha$ then since $y, z \geqslant \eta n$, G^* contains a copy of T by Lemma 3.13(iii) (again applied with ε' in the place of γ). So in either case G contains a copy of T, and so we may assume that $|C| < 5\varepsilon' n$.

So to prove (ii), observe that we must therefore have $|B \cup C \cup D| \leqslant 5\varepsilon' n + 6\beta n \leqslant 6\varepsilon' n$. Trivially $|A| \leqslant 2y + 2\varepsilon n + 1$ and $|E| \leqslant 2z + 2\varepsilon n + 1$, and so we must have

$$|A| \geqslant 2n - 2 - 6\varepsilon' n - 2z - 2\varepsilon n - 1 \geqslant 2y - 7\varepsilon' n, \text{ and}$$
$$|E| \geqslant 2n - 2 - 6\varepsilon' n - 2y - 2\varepsilon n - 1 \geqslant 2z - 7\varepsilon' n.$$

So by Proposition 3.1, $G[A]$ contains a γ-almost-regular subtournament on at least $(2-\gamma)y$ vertices, and $G[E]$ contains a γ-almost-regular subtournament on at least $(2-\gamma)z$ vertices. Let Y and Z be the vertex sets of these subtournaments respectively. Then any vertex of Y has at least $(1-2\gamma)y$ outneighbours in Y, and so has at most $y + \varepsilon n - (1-2\gamma)y \leqslant 3\gamma n$ outneighbours in Z. Similarly any vertex of Z has at least $(1-2\gamma)z$ inneighbours in Z, and so has at most $3\gamma n$ inneighbours in Y. So Y and Z are as required for (ii). □

The next lemma builds on the previous lemma and is in turn used in the proof of Lemma 3.24.

Lemma 3.23 *Suppose that $1/n \ll \beta, 1/\Delta' \ll \alpha \ll 1/\Delta \ll 1$. Let T be a directed tree on n vertices with $|T_{\Delta'}| \leqslant \beta n$. Let y and z be the outweight and inweight of $T_{\Delta'}$ respectively. Suppose that forests F^- and F^+ are induced subgraphs of T which partition the vertices of T, such that $|F^+| \leqslant y + 2\alpha n$, $|F^-| \leqslant z - \alpha n$, and every edge of T between F^- and F^+ is directed from F^- to F^+. Suppose also that either*

(i) no component of F^+ has order greater than $y - \alpha n$, or

(ii) the largest component T_1 of F^+ has $|(T_1)_\Delta| \geqslant 2$.

Then any tournament G on $2n-2$ vertices contains a copy of T.

Proof. Let G be a tournament on $2n-2$ vertices, and let T_1 and T_2 be the largest and second largest components of F^+ respectively. Introduce new constants γ and η with

$$\frac{1}{n} \ll \beta, \frac{1}{\Delta'} \ll \gamma \ll \alpha \ll \frac{1}{\Delta} \ll \eta \ll 1.$$

Then by Lemma 3.21 we may assume that $y, z \geqslant \eta n$. Also by Lemma 3.21 we may find subsets $Y, Z \subseteq V(G)$ such that $|Y| \geqslant (2-\gamma)y$, $|Z| \geqslant (2-\gamma)z$, $G[Y]$ is γ-almost-regular, each vertex of Y has at most $3\gamma n$ outneighbours in Z, and each vertex of Z has at most $3\gamma n$ inneighbours in Y. Then $|Y| \geqslant 3|F^+|/2 + \alpha n \geqslant |F^+| + |T_2| + \alpha n$, and $|Z| \geqslant 2|F^-| + \alpha n$, and so by Lemma 3.4 any embedding of T_1 in $G[Y]$ may be extended to an embedding of T in G.

It therefore suffices to embed T_1 in $G[Y]$. If $|T_1| < y/2$ then we may do this by Theorem 1.2. If instead $|T_1| \geqslant y/2 \geqslant \eta n/2$ and we also have (i), then $|T_1| \leqslant y - \alpha n$. Since $|Y| \geqslant (2-\gamma)y \geqslant 2|T_1| + \alpha n$ we may embed T_1 in $G[Y]$ by Theorem 1.5(1). Finally, if $|T_1| \geqslant \eta n/2$ and we also have (ii), then $|T_1| \leqslant |F^+| \leqslant y + 2\alpha n$ and $|(T_1)_\Delta| \geqslant 2$. Since $\gamma \leqslant 9\alpha/\eta$, $G[Y]$ is a $9\alpha/\eta$-almost-regular tournament on at least $(2-\gamma)y \geqslant (2-9\alpha/\eta)|T_1|$ vertices, and so we may embed T_1 in $G[Y]$ by Lemma 3.19. So in any case we may embed T_1 in $G[Y]$, completing the proof. $\square$

Observe that as with Lemma 3.4 a 'dual' form of Lemma 3.23 can be proved similarly. For this we instead require that $|F^+| \leqslant y - \alpha n$ and $|F^-| \leqslant z + 2\alpha n$, and also either that no component of F^- has order greater than $z - \alpha n$ or that the largest component T_1 of F^- has $|(T_1)_\Delta| \geqslant 2$. If these conditions are met then we may conclude that G contains a copy of T. As with Lemma 3.4, we sometimes implicitly refer to this 'dual' when referring to Lemma 3.23.

In the next lemma we show that Lemma 3.20 holds for any directed tree T whose core tree T_Δ is not a directed path in which most of the outweight and inweight of T_Δ lies at the endvertices of T_Δ. We say that a vertex t of a directed tree T is an *outleaf* if t has one inneighbour and no outneighbours, or an *inleaf* if t has one outneighbour and no inneighbours.

Lemma 3.24 *Suppose that $1/n \ll \beta, 1/\Delta' \ll 1/\Delta \ll \sigma \ll 1$. Let T be a directed tree on n vertices with $|T_{\Delta'}| \leqslant \beta n$, and let y and z be the outweight and inweight of $T_{\Delta'}$ respectively. Let G be a tournament on $2n-2$ vertices. Then either G contains a copy of T, or T_Δ is a directed path whose outleaf has outweight at least $y - \sigma n$ and whose inleaf has inweight at least $z - \sigma n$.*

Proof. Introduce new constants α and η with

$$\frac{1}{n} \ll \beta, \frac{1}{\Delta'} \ll \alpha \ll \frac{1}{\Delta} \ll \sigma \ll \eta \ll 1.$$

Then by Lemma 3.21 we may assume that $y, z \geqslant \eta n$. Also, if $|T_\Delta| = 1$ then G contains a copy of T by Lemma 3.5, so we may assume that $|T_\Delta| \geqslant 2$.

Suppose that some vertex $t \in T$ has the property that $w^-(t) \leqslant z - \alpha n - 1$, and also that every outcomponent of t contains at most $w^+(t) - 3\alpha n = |V^+| - 3\alpha n$ vertices, where the set V^- consists of t and every vertex in an incomponent of t, and $V^+ := V(T) \setminus V^-$. Then $|V^-| \leqslant w^-(t) + 1 \leqslant z - \alpha n$, and every edge of T between V^- and V^+ is directed from V^- to V^+. Also, each component of $T[V^+]$ contains at most $w^+(t) - 3\alpha n$ vertices. Now, select a source vertex from the largest component of $T[V^+]$, delete this vertex from V^+, and add it to V^-. Repeat this step until we have $|V^+| \leqslant y + 2\alpha n$ and $|V^-| \leqslant z - \alpha n$. For these final V^+ and V^-, let $F^+ := T[V^+]$ and let $F^- := T[V^-]$. Then F^- and F^+ are forests which partition the vertices of T, with $|F^+| \leqslant y + 2\alpha n$ and $|F^-| \leqslant z - \alpha n$. Also, every edge of T between F^- and F^+ is directed from F^- to F^+. Finally, since we always deleted a vertex from the largest component of $T[V^+]$, no component of F^+ contains more than $|F^+| - 3\alpha n \leqslant y - \alpha n$ vertices.

So by Lemma 3.23(i) G contains a copy of T. So we may assume that

(†) *there is no vertex $t \in T$ such that $w^-(t) \leqslant z - \alpha n - 1$ and every outcomponent of t contains at most $w^+(t) - 3\alpha n$ vertices. In particular, this implies that for every inleaf t of T_Δ, at least $n/2\Delta$ vertices of T lie in incomponents of t.*

Indeed, if T_Δ contains some inleaf t such that fewer than $n/2\Delta \leqslant z - \alpha n - 1$ vertices of T lie in incomponents of t, then by the definition of T_Δ at least $n/2\Delta - 1$ vertices of T lie in outcomponents of t other than the outcomponent containing the remaining vertices of T_Δ. Moreover, the definition of T_Δ also implies that at least n/Δ vertices of T lie in the one component of $T - t$ containing $T_\Delta - t$. Altogether this shows that every outcomponent of t contains at most $w^+(t) - n/2\Delta + 1 \leqslant w^+(t) - 3\alpha n$ vertices, a contradiction. By the same argument with the roles of incomponents and outcomponents switched, we may assume that

(††) *there is no vertex $t \in T$ such that $w^+(t) \leqslant y - \alpha n - 1$ and every incomponent of t contains at most $w^-(t) - 3\alpha n$ vertices. It follows from this that for every outleaf t of T_Δ, at least $n/2\Delta$ vertices of T lie in outcomponents of t.*

Claim. *If T_Δ has at least two inleaves or at least two outleaves, then G contains a copy of T.*

To prove the claim, suppose that T_Δ has two outleaves t and t' (the proof for inleaves is similar). Then we form a set V^+ of size between $n - z + \alpha n$ and $y + 2\alpha n$ such that any edge of T between V^+ and $V^- := V(G) \setminus V^+$ is directed from V^- to V^+. We may do this by repeatedly selecting a sink vertex of T, adding it to V^+ and removing it from T. Now, by (††) at least $n/2\Delta$ vertices lie in outcomponents of t, and at least $n/2\Delta$ vertices lie in outcomponents of t'. Furthermore, if T' is an outcomponent of t, then any sink vertex in T' is a sink vertex in T, and the same is true if T' is instead an outcomponent of t'. So we may form V^+ and V^- as described above so that additionally V^+ contains at least $n/2\Delta$ vertices from outcomponents of t and at least $n/2\Delta$ vertices from outcomponents of t'. Fix such a choice of V^+ and V^-, and let $F^+ := T[V^+]$ and $F^- := T[V^-]$ be the induced forests. Then $|F^+| \leqslant y + 2\alpha n$ and $|F^-| = n - |F^+| \leqslant z - \alpha n$, and every edge of T between F^- and F^+ is directed from F^- to F^+. So if every component of F^+ contains at most $y - \alpha n$ vertices, then G contains a copy of T by Lemma 3.23(i). We may therefore assume that the largest component T^+ of F^+ contains more than $y - \alpha n \geqslant |F^+| - n/4\Delta$ vertices. Since F^+ includes at least $n/2\Delta$ vertices from outcomponents of t and at least $n/2\Delta$ vertices from outcomponents of t', it follows that T^+ contains at least $n/4\Delta$ vertices from outcomponents of t and at least $n/4\Delta$ vertices from outcomponents of t'. As a consequence T^+ must contain t and t'. Furthermore, we must have $t, t' \in (T^+)_{4\Delta}$, and so $|(T^+)_{4\Delta}| \geqslant 2$. So G contains a copy of T by Lemma 3.23(ii), which proves the claim.

We may therefore assume that T_Δ has at most one outleaf and at most one inleaf. So T_Δ is a path with one inleaf and one outleaf. Let $t_1, \ldots, t_x$ be the vertices of this path, labelled so that t_1 is the inleaf of T_Δ (so $t_1 \to t_2$), t_x is the outleaf of T_Δ (so $t_{x-1} \to t_x$), and for each $i \in [x-1]$ there is an edge of T_Δ between t_i and t_{i+1}.

Now suppose that the inweight of T_Δ is less than $z - 2\alpha n$. Let the set V^- consist of all vertices of T which lie in T_Δ or in incomponents of T_Δ. Then $|V^-| \leqslant z - 2\alpha n + |T_\Delta| \leqslant z - \alpha n$ (since $|T_\Delta| \leqslant |T_{\Delta'}| \leqslant \beta n$). Also, every edge of T between V^- and $V^+ := V(T) \setminus V^-$ is directed from V^- to V^+. Choose a source vertex of $T[V^+]$, delete it from V^+, and add it to V^-, and repeat this step until we have $|V^-| \leqslant z - \alpha n$ and $|V^+| \leqslant y + 2\alpha n$. For these final V^- and V^+, let $F^+ := T[V^+]$ and $F^- := T[V^-]$ be the induced forests. Then $|F^-| \leqslant z - \alpha n$, $|F^+| \leqslant y + 2\alpha n$, and every edge of T between F^- and F^+ is directed from F^- to F^+. Also, every component of F^+ is contained within a component of $T - T_\Delta$, and so has order at most $n/\Delta \leqslant y - \alpha n$ by

Proposition 2.19(iv). So G contains a copy of T by Lemma 3.23(i). We may therefore assume that the inweight of T_Δ is at least $z - 2\alpha n$, and by a similar argument we may also assume that the outweight of T_Δ is at least $y - 2\alpha n$. It follows that the outweight of T_Δ is at most $n - (z - 2\alpha n) \leqslant y + 3\alpha n$ and that the inweight of T_Δ is at most $n - (y - 2\alpha n) \leqslant z + 3\alpha n$.

We now suppose that fewer than $y - \sigma n$ vertices of T lie in outcomponents of t_x. Let T_1 be the subtree of T formed by T_Δ and all of its outcomponents. Initially let the set $V^+ := V(T_1)$, so $|V^+| \leqslant y + 4\alpha n$, and every edge of T between V^+ and $V^- := V(G) \setminus V^+$ is directed from V^- to V^+. Choose a sink vertex of $T[V^-]$, delete it from V^- and add it to V^+, and repeat this step until we have $|V^+| \leqslant y + 4\alpha n$ and $|V^-| \leqslant z - 2\alpha n$. Fix these final V^+ and V^- and let $F^- := T[V^-]$ and $F^+ := T[V^+]$ be the induced forests. So $|F^+| \leqslant y + 4\alpha n$, $|F^-| \leqslant z - 2\alpha n$, and every edge of T between F^- and F^+ is directed from F^- to F^+. Also $T_1 \subseteq F^+$, so T_1 is contained within a single component T^+ of F^+. Since at least $y - 2\alpha n$ vertices of T lie in outcomponents of T_Δ, at least $\sigma n/2$ vertices of T lie in outcomponents of T_Δ other than the outcomponents of t_x. Moreover, since t_x is an outleaf of T_Δ, by (††) at least $n/2\Delta$ vertices lie in outcomponents of t_x. So $t_{x-1} \in (T^+)_{2\Delta}$ and $t_x \in (T^+)_{2\Delta}$. So $|(T^+)_{2\Delta}| \geqslant 2$. But since the outweight of T_Δ is at least $y - 2\alpha n$ we have $|T^+| \geqslant |T_1| \geqslant y - 2\alpha n$, and so T^+ must be the largest component of F^+. So G contains a copy of T by Lemma 3.23(ii).

So we may assume that at least $y - \sigma n$ vertices of T lie in outcomponents of t_x, as desired. If fewer than $z - \sigma n$ vertices of T lie in incomponents of t_1, then we may similarly embed T in G, so we may also assume that at least $z - \sigma n$ vertices of T lie in incomponents of t_1. So at most $3\sigma n$ vertices of T do not lie in incomponents of t_1 or outcomponents of t_x. It remains only to show that T_Δ is a directed path. So suppose for a contradiction that T_Δ is not a directed path. Then there is some $i \in [x-1]$ such that $t_i \leftarrow t_{i+1}$. Choose the minimal such i (note $i > 1$ as t_1 is an inleaf of T_Δ). Then t_i has two inneighbours and no outneighbours in T_Δ. So at least two incomponents of t_i contain at least n/Δ vertices, and so no incomponent of t_i contains more than $w^-(t_i) - n/\Delta \leqslant w^-(t_i) - 3\alpha n$ vertices. Also, at most $3\sigma n \leqslant y - \alpha n - 1$ vertices of T lie in outcomponents of t_i, contradicting (††). $\square$

We can now prove that Sumner's universal tournament conjecture holds for any large directed tree T whose core tree T_Δ contains precisely two vertices.

Lemma 3.25 *Suppose that $1/n \ll 1/\Delta' \ll 1$. Let T be a directed tree on n vertices with $|T_{\Delta'}| = 2$, and let G be a tournament on $2n-2$ vertices. Then G contains a copy of T.*

Proof. Introduce new constants $\Delta, \varepsilon, \gamma$ and η with

$$\frac{1}{n} \ll \beta, \frac{1}{\Delta'} \ll \frac{1}{\Delta} \ll \varepsilon \ll \gamma \ll \eta \ll 1.$$

Then $|T_{\Delta'}| = 2 \leqslant \beta n$. Also, since $\Delta \leqslant \Delta'$ we have $T_\Delta \subseteq T_{\Delta'}$. If $|T_\Delta| = 1$, then by Lemma 3.5 G contains a copy of T. So we may assume that $T_\Delta = T_{\Delta'}$. Let t_2 and t_1 be the vertices of T_Δ, labelled so that $t_2 \to t_1$. Let y be the outweight of T_Δ, and let z be the inweight of T_Δ, so $y + z = n - 2$. Then by Lemma 3.24 (with ε in the place of σ), we may assume that t_2 has inweight at least $z - \varepsilon n$, and also that t_1 has outweight at least $y - \varepsilon n$. Let T_1 be the subtree of T consisting of all vertices which lie in T_Δ or in outcomponents of T_Δ, and let T_2 be the subtree of T consisting of all vertices which lie in T_Δ or in incomponents of T_Δ. So $|T_1| = y + 2$ and $|T_2| = z + 2$. By Lemma 3.21(i) we may assume that $y, z \geqslant \eta n$.

As in the proof of Lemma 3.21, we partition the vertices of G into sets A, B, C, D and E, where:

$$A := \{v \in G : d^+(v) \leqslant y + \varepsilon n\},$$
$$B := \{v \in G : y + \varepsilon n < d^+(v) < n - \varepsilon n\},$$
$$C := \{v \in G : d^+(v), d^-(v) \geqslant n - \varepsilon n\},$$
$$D := \{v \in G : z + \varepsilon n < d^-(v) < n - \varepsilon n\},$$
$$E := \{v \in G : d^-(v) \leqslant z + \varepsilon n\}.$$

Since $y, z \geqslant \eta n$ and $\varepsilon \ll \eta$ this is indeed a partition. Suppose first that $|B| \geqslant 2$. Then we may embed T_Δ in $G[B]$. Let $S_\Delta \subseteq B$ be the set of vertices occupied by T_Δ. Then every vertex of S_Δ has at least $y + \varepsilon n - 1 \geqslant y + 2n/\Delta$ outneighbours outside S_Δ and at least $|G| - 2 - (n - \varepsilon n) \geqslant y + z + 2n/\Delta$ inneighbours outside S_Δ. So by Lemma 3.3(c) we may extend the embedding of T_Δ in S_Δ to an embedding of T_1 in G. This embedding of T_1 occupies at most y vertices of G outside S_Δ, and so we may apply Lemma 3.3(c) again to extend the embedding of T_Δ in S_Δ to an embedding of T_2 in G so that the embeddings of T_1 and T_2 do not overlap outside T_Δ. Then together the embeddings of T_1 and T_2 form an embedding of T in G. So we may assume that $|B| \leqslant 1$. If $|D| \geqslant 2$ we may embed T in G in the same way by embedding T_Δ in D and then extending this embedding to embeddings of first T_2 and then T_1 in G which do not overlap outside T_Δ. So we may also assume that $|D| \leqslant 1$.

Now suppose that $|C| \geqslant 3$. Then we may choose vertices $v_2, v_1 \in C$ with $v_2 \to v_1$ and $|N^+(v_1) \cap N^+(v_2)| \geqslant \eta n \geqslant \eta n/2 + 2n/\Delta$. Embed t_1 to v_1 and t_2 to v_2. Then since $|N^+(v_1)|, |N^+(v_2)| \geqslant n - \varepsilon n \geqslant y + 2n/\Delta$, by Lemma 3.3(c) we may extend the embedding of T_Δ in $\{v_1, v_2\}$ to an embedding of T_1 in G so that at least $\eta n/2$ vertices of T_1 are embedded in $N^+(v_1) \cap N^+(v_2)$. Then at most $y + 2 - \eta n/2$ vertices of $N^-(v_1) \cup N^-(v_2)$ are occupied by this embedding, and so in each of $N^-(v_1)$ and $N^-(v_2)$ at least $n - \varepsilon n - (y + 2 - \eta n/2) \geqslant z + 2n/\Delta$ vertices remain unoccupied. So by Lemma 3.3(c) we may extend the embedding of T_Δ in

$\{v_1, v_2\}$ to an embedding of T_2 in G which does not overlap with the embedding of T_1 outside T_Δ. Then together these embeddings form an embedding of T in G. So we may assume that $|C| \leqslant 2$, and hence that $|A \cup E| \geqslant 2n - 6$.

Claim. *Either some vertex of A has at least y outneighbours in $A \cup B \cup D$ or some vertex of E has at least z inneighbours in $B \cup D \cup E$.*

Indeed, suppose for a contradiction that both of these statements are false. Then certainly every vertex of A has fewer than y outneighbours in A and every vertex of E has fewer than z inneighbours in E. So $|A| \leqslant 2y-1$ and $|E| \leqslant 2z-1$. Since $y+z = n-2$ and $|A \cup E| \geqslant 2n-6$, we must have $|A| = 2y - 1$ and $|E| = 2z - 1$, and also $|B| = 1, |D| = 1$ and $|C| = 2$. Then every vertex of A must have $y - 1$ outneighbours in A, and so no vertex of A can have an outneighbour in B or in D. Likewise, every vertex of E must have $z - 1$ inneighbours in E, and so no vertex of E can have an inneighbour in B or in D. But then if we let b be the vertex in B and d be the vertex in D we have $d^+(b) = d^+(d) \pm 3$, contradicting the definition of B and D. So either some vertex of A has at least y outneighbours in $A \cup B \cup D$ or some vertex of E has at least z inneighbours in $B \cup D \cup E$. This completes the proof of the claim.

If some $v \in A$ has at least y outneighbours in $A \cup B \cup D$, then we embed T_1 in $G[A]$ so that we may then embed the incomponents of t_2 and t_1 in the unoccupied vertices of E and A respectively. For this, note that $|E| \leqslant 2(z + \varepsilon n) + 1$, so $|A| \geqslant 2n - 2z - 2\varepsilon n - 7 \geqslant 2y - 3\varepsilon n$ (and similarly we have $|E| \geqslant 2z - 3\varepsilon n$). Since every $a \in A$ has at most $y + \varepsilon n$ outneighbours in A, by Proposition 3.1 $G[A]$ contains a γ-almost-regular subtournament on at least $(2 - \gamma)y$ vertices. Let Y be the vertex set of this subtournament. Now,

$$|(A \cup B \cup D) \setminus Y| \leqslant 2 + (2y + 2\varepsilon n + 1) - (2 - \gamma)y \leqslant 2\gamma y,$$

so v must have at least $(1-2\gamma)y$ outneighbours in Y. Also, since $v \in A$ we have

$$(1-2\gamma)y \leqslant |N^+(v) \cap Y| \leqslant y + \varepsilon n \leqslant (1+2\gamma)y.$$

So at most $10\gamma y$ vertices of $N^+(v) \cap Y$ have more than $(1-3\gamma)y$ outneighbours in $N^+(v) \cap Y$, and at most $10\gamma y$ vertices of $N^+(v) \cap Y$ have more than $(1-3\gamma)y$ inneighbours in $N^+(v) \cap Y$. Since every vertex of Y has at least $(1-2\gamma)y$ inneighbours in Y and at least $(1-2\gamma)y$ outneighbours in Y, this means that at least $|N^+(v) \cap Y| - 20\gamma y \geqslant 3n/\Delta$ vertices of $N^+(v) \cap Y$ have at least $\gamma y \geqslant 6n/\Delta$ outneighbours in $Y \setminus N^+(v)$ and at least $6n/\Delta$ inneighbours in $Y \setminus N^+(v)$. Let T^+ be the tree formed by t_1 and its outcomponents, so $|T^+| \leqslant y + 1$. Then every component of $T^+ - t_1$ is a component of $T - T_\Delta$ and so has order at most n/Δ by Proposition 2.19(iv). So by Lemma 3.2 (applied with $N := N^+(v) \cap (A \cup B \cup D)$ and $X := Y \setminus N^+(v)$), we may embed T^+ in $G[A \cup B \cup D]$ so that t_1 is embedded to v and at most $4n/\Delta$ vertices are embedded outside $N^+(v)$.

Since $v \in A$ we have $d^+(v) \leqslant y + \varepsilon n$, and so v has at least

$$|Y| - 1 - (y + \varepsilon n) - \frac{4n}{\Delta} \geqslant 7\varepsilon n \tag{3.26}$$

inneighbours in Y which are not occupied by the embedding of T^+. Let T^* be the tree formed by all vertices of T which do not lie in outcomponents of t_1 or incomponents of t_2. Then every edge incident to t_1 in T^* is directed towards t_1. Also, $|T^*| \leqslant n - (y - \varepsilon n) - (z - \varepsilon n) = 2\varepsilon n + 2$, so certainly every component of $T^* - t_1$ has order at most $2\varepsilon n + 1$. Together with (3.26) and Theorem 1.2 this shows that we may extend the embedding of t_1 in $\{v\}$ to an embedding of T^* in $\{v\} \cup (N^-(v) \cap Y)$ so that the embeddings of T^+ and T^* only overlap in the vertex t_1. Then in particular t_2 is embedded to some vertex $v_2 \in Y$.

To complete the embedding, observe that every vertex of Y has at least $(1-2\gamma)y$ outneighbours in Y, and therefore at most $3\gamma y$ outneighbours outside Y. So v_2 has at least $|E| - 3\gamma y \geqslant z + 2n/\Delta$ inneighbours in E, none of which have been occupied by the embeddings of T^+ and T^*. Let T^- be the subtree of T consisting of t_2 and all of its incomponents. Then $|T^-| \leqslant z+1$, and each component of $T^- - t_2$ is a component of $T - T_\Delta$ and so has order at most n/Δ by Proposition 2.19(iv). So by Lemma 3.3(c) we may extend the embedding of t_2 in $\{v_2\}$ to an embedding of T^- in $\{v_2\} \cup E$. These embeddings together form an embedding of T in G.

If instead some $v \in E$ has at least z inneighbours in $B \cup D \cup E$ then we may similarly embed T in G by choosing Z to be the vertex set of a γ-almost-regular subtournament of $G[E]$ on at least $(2-\gamma)z$ vertices and embedding T^- in $G[B \cup D \cup E]$, then embedding $T^* - t_2$ in the unoccupied vertices of Z, before finally embedding $T^+ - t_1$ in $G[A]$. $\square$

We can now give the proof of Lemma 3.20. It was necessary to prove Lemma 3.25 separately from this as the method of proof does not hold for $|T_\Delta| = 2$ (we cannot obtain the partition of $V(G)$ into Y^* and Z^* in this case).

Proof of Lemma 3.20. Introduce new constants γ, α, Δ and η with

$$\frac{1}{n} \ll \beta, \frac{1}{\Delta'} \ll \frac{1}{\Delta} \ll \gamma \ll \alpha \ll \eta \ll 1.$$

Let y' be the outweight of $T_{\Delta'}$ and let z' be the inweight of $T_{\Delta'}$. Then by Lemma 3.21 we may assume that $y', z' \geqslant \eta n$. Similarly let y and z be the outweight and inweight of T_Δ respectively. If $|T_\Delta| = 1$, then G contains a copy of T by Lemma 3.5. If instead $|T_\Delta| = 2$ then G contains a copy of T by Lemma 3.25. So we may assume that $\ell := |T_\Delta| \geqslant 3$, and by Lemma 3.24 we may assume that T_Δ is a directed path. Let $t_1, \dots, t_\ell$ be the vertices of T_Δ, labelled so that $t_i \to t_{i+1}$ for each $i \in [\ell - 1]$. Then by Lemma 3.24 we may also assume that the inweight of t_1 is at least $z' - \gamma n$ and that the outweight of t_ℓ is at least $y' - \gamma n$. This implies that $z \geqslant z' - \gamma n$ and

$y \geqslant y' - \gamma n$. Since $y' + z' + |T_{\Delta'}| = y + z + |T_\Delta| = n$ it follows that we must have

$$y = y' \pm 2\gamma n \text{ and } z = z' \pm 2\gamma n. \tag{3.27}$$

Finally, by Lemma 3.21 we may assume that there are disjoint sets $Y, Z \subseteq V(G)$ such that:

(a) $|Y| \geqslant (2-\gamma)y'$ and $|Z| \geqslant (2-\gamma)z'$,

(b) $G[Y]$ and $G[Z]$ are γ-almost-regular, and

(c) any vertex of Y has at most $3\gamma n$ outneighbours in Z and any vertex of Z has at most $3\gamma n$ inneighbours in Y.

Let $X := V(G) \setminus (Y \cup Z)$, so $|X| \leqslant 2\gamma n$. Let T^* be the subtree of T formed by deleting from T all vertices in outcomponents of t_ℓ or incomponents of t_1. So $|T^*| \leqslant n - (z' - \gamma n) - (y' - \gamma n) \leqslant 3\gamma n$. Let T^+ be the subtree of T formed by t_ℓ and its outcomponents, and let T^- be the subtree of T formed by t_1 and its incomponents. So $|T^+| \leqslant y + 1$ and $|T^-| \leqslant z + 1$. Also, each component of $T^+ - t_\ell$ and each component of $T^- - t_1$ is a component of $T - T_\Delta$ and so has order at most n/Δ by Proposition 2.19(iv).

Suppose that some vertex $v \in X$ has at least αn inneighbours in Y and at least αn outneighbours in Z. Since $\ell \geqslant 3$, we may choose i with $1 < i < \ell$. Embed t_i to v. Let T_a be the subtree of T^* consisting of t_i and all of its outcomponents, and let T_b be the subtree of T^* consisting of t_i and all of its incomponents. Then $|T_a|, |T_b| \leqslant |T^*| \leqslant 3\gamma n$. So by Lemma 3.3(c) we may extend the embedding of t_i in $\{v\}$ to an embedding of T_a in $Z \cup \{v\}$, and similarly we may extend the embedding of t_i in $\{v\}$ to an embedding of T_b in $Y \cup \{v\}$. Then in particular t_1 is embedded to some $v_1 \in Y$ and t_ℓ is embedded to some $v_\ell \in Z$. So v_1 has at least $|Z| - 3\gamma n \geqslant z + 3\gamma n + 2n/\Delta$ inneighbours in Z, at most $3\gamma n$ of which are occupied by the embedding of T_a. Similarly v_ℓ has at least $|Y| - 3\gamma n \geqslant y + 3\gamma n + 2n/\Delta$ outneighbours in Y, at most $3\gamma n$ of which are occupied by the embedding of T_b. So by Lemma 3.3(c) we may extend the embedding of t_1 in $\{v_1\}$ to an embedding of T^- in $\{v_1\} \cup Z$ and also extend the embedding of t_ℓ in $\{v_\ell\}$ to an embedding of T^+ in $\{v_\ell\} \cup Y$ so that these embeddings together form a copy of T in G.

So we may assume that no vertex of X has at least αn inneighbours in Y and at least αn outneighbours in Z. Let $X^+ \subseteq X$ consist of all vertices of X with fewer than αn inneighbours in Y, and let $X^- \subseteq X \setminus X^+$ consist of all vertices of $X \setminus X^+$ with fewer than αn outneighbours in Z. Let $Y^* := Y \cup X^-$ and let $Z^* := Z \cup X^+$, so Y^* and Z^* partition the vertices of G. Then any vertex of Y^* has at most αn outneighbours in Z, and thus at least $z + \alpha n$ inneighbours in Z^* (by (a), (3.27) and the fact that $z' \geqslant \eta n$). Similarly any vertex of Z^* has at most αn inneighbours in Y, and therefore at least $y + \alpha n$ outneighbours in Y^*. Let $W \subseteq V(G)$ consist of all vertices in Y^* with at least $y + \alpha n$ outneighbours in Y^* and all vertices in Z^* with at least $z + \alpha n$ inneighbours in Z^*.

Now suppose that $|W| \geqslant |T_\Delta|$. Since T_Δ is a directed path, by Theorem 1.4 we may embed T_Δ in $G[W]$. Let $S_\Delta \subseteq W$ be the set of vertices occupied by this embedding. Then $|S_\Delta| = |T_\Delta| \leqslant |T_{\Delta'}| \leqslant \beta n$. So every vertex of S_Δ has at least $y + \alpha n/2 \geqslant y + 2n/\Delta$ outneighbours in $Y^* \setminus S_\Delta$ and at least $z + \alpha n/2 \geqslant z + 2n/\Delta$ inneighbours in $Z^* \setminus S_\Delta$. Let T_1 be the subtree of T consisting of T_Δ and all of its outcomponents, and let T_2 be the subtree of T consisting of T_Δ and all of its incomponents. So $|T_1| = \ell + y$ and $|T_2| = \ell + z$. Also, each component of $T_1 - T_\Delta$ and each component of $T_2 - T_\Delta$ is a component of $T - T_\Delta$, and so has order at most n/Δ by Proposition 2.19(iv). So by Lemma 3.3(c) we may extend the embedding of T_Δ in S_Δ to an embedding of T_1 in $Y^* \cup S_\Delta$. Similarly by Lemma 3.3(c) we may extend the embedding of T_Δ in S_Δ to an embedding of T_2 in $Z^* \cup S_\Delta$. These embeddings of T_1 and T_2 do not overlap outside T_Δ, and so together form an embedding of T in G.

We may therefore assume that $|W| < |T_\Delta|$, and hence that $|G - W| \geqslant 2n - 1 - \ell$. Since $y+z = n-\ell$, we must have either $|Y^* \setminus W| \geqslant 2y$ or $|Z^* \setminus W| \geqslant 2z$. Suppose that $|Y^* \setminus W| \geqslant 2y$. Then $Y^* \setminus W$ contains a vertex v_ℓ with at least y outneighbours in Y^*. So we may choose a set $N \subseteq N^+(v_\ell) \cap Y^*$ with $|N| = y$. Then $|N \cap Y| \geqslant y - (|Y^*| - |Y|) \geqslant y - 2\gamma n$. Now, by (a), (b) and (3.27) every vertex of Y has at least $(1 - 2\sqrt{\gamma})y$ inneighbours in Y and at least $(1 - 2\sqrt{\gamma})y$ outneighbours in Y. Since $|N| = y$, at most $6\sqrt{\gamma}y$ vertices of $N \cap Y$ have more than $(1 - 3\sqrt{\gamma})y$ inneighbours in $N \cap Y$, and at most $6\sqrt{\gamma}y$ vertices of $N \cap Y$ have more than $(1 - 3\sqrt{\gamma})y$ outneighbours in $N \cap Y$. So at least $|N \cap Y| - 12\sqrt{\gamma}n \geqslant 3n/\Delta$ vertices of N have at least $6n/\Delta$ inneighbours in $Y^* \setminus (N \cup \{v_\ell\})$ and at least $6n/\Delta$ outneighbours in $Y^* \setminus (N \cup \{v_\ell\})$. This means that by Lemma 3.2 (applied with $Y^* \setminus (N \cup \{v_\ell\})$ playing the role of X) we may embed T^+ in Y^* with t_ℓ embedded to v_ℓ, and at most $4n/\Delta$ vertices of T^+ embedded outside N. Since $v_\ell \notin W$, v_ℓ has at most $y + \alpha n$ outneighbours in Y^*, and so v_ℓ has at least $|Y| - 1 - (y + \alpha n) - 4n/\Delta \geqslant 9\gamma n$ inneighbours in Y which are not occupied by the embedding of T^+. Since $|T^*| \leqslant 3\gamma n$, by Lemma 3.3(c) we may extend the embedding of t_ℓ in v_ℓ to an embedding of T^* in Y which only overlaps the embedding of T^+ in t_ℓ. The vertex t_1 of T is therefore embedded to some vertex $v_1 \in Y$. By (3), v_1 then has at least $|Z| - 3\gamma n \geqslant z + 2n/\Delta$ inneighbours in Z, none of which have been occupied by the embeddings of T^* and T^+ so far. So by Lemma 3.3(c) we may extend the embedding of t_1 in $\{v_1\}$ to an embedding of T^- in $Z \cup \{v_1\}$. Then the embeddings of T^+, T^- and T^* combine to form an embedding of T in G. If instead we have $|Z^* \setminus W| \geqslant 2z$, then we may embed T in G similarly, first embedding T^- in Z^*, then embedding T^* in the unoccupied vertices of Z, and finally embedding T^+ in Y. So in either case G contains a copy of T, completing the proof. $\square$

3.6 Proof of Theorem 1.1

Having proved that Sumner's conjecture holds for directed trees of small core, we now show that the same is true for directed trees of large core, which completes the proof of Theorem 1.1. We begin with an embedding result similar to Lemma 3.23.

Lemma 3.28 *Suppose that $1/n \ll 1/\Delta \ll \mu \ll \nu \ll \eta \ll \gamma \ll \alpha \ll \beta \ll 1$. Let T be a directed tree on n vertices, and let forests F^- and F^+ be induced subgraphs of T which partition the vertices of T such that $|F^+| \geqslant 6\alpha n$. Suppose also that every edge of T between F^- and F^+ is directed from F^- to F^+. Let Y and Z be disjoint sets with $|Y| \geqslant 2|F^+| - 2\alpha n$ and $|Z| \geqslant 2|F^-| + \alpha n$, and let G be a tournament on vertex set $Y \cup Z$ such that every vertex of Y has at most $\gamma|G|$ outneighbours in Z and every vertex of Z has at most $\gamma|G|$ inneighbours in Y. Finally, let T_1^+ be the largest component of F^+, and suppose that either*

(i) $|T_1^+| \leqslant |F^+| - 3\alpha n$,

(ii) $G[Y]$ *is a robust* (μ, ν)*-outexpander with* $\delta^0(G[Y]) \geqslant \eta|Y|$ *and* $|(T_1^+)_\Delta| \geqslant \beta n$*, or*

(iii) $\Delta(T_1^+) \leqslant \Delta$.

Then G contains a copy of T.

Proof. First observe that if $|G| \geqslant 3n$, then G contains a copy of T by Theorem 1.2. So we may assume that $|G| < 3n$, and hence that every vertex of Y has at most $3\gamma n$ outneighbours in Z and every vertex of Z has at most $3\gamma n$ inneighbours in Y. Let T_2^+ be the second largest component of F^+. Then $|F^+| - |T_2^+| \geqslant |F^+|/2 \geqslant 3\alpha n$, so $|Y| \geqslant |F^+| + |T_2^+| + \alpha n$. Since $|Z| \geqslant 2|F^-| + \alpha n$, by Lemma 3.4 any embedding of T_1^+ in $G[Y]$ may be extended to an embedding of T in G. So it is sufficient to embed T_1^+ in $G[Y]$.

Note that $|Y| \geqslant 10\alpha n$, so if $|T_1^+| < \alpha n$, then $G[Y]$ contains a copy of T_1^+ by Theorem 1.2. Alternatively, suppose that $|T_1^+| \geqslant \alpha n$. If (i) holds, then $|T_1^+| \leqslant |Y|/2 - 2\alpha n$, and so $|Y| \geqslant (2+\alpha)|T_1^+|$. So $G[Y]$ contains a copy of T_1^+ by Theorem 1.5(1). If instead (ii) holds then G contains a copy of T_1^+ by Lemma 2.32. Finally, if (iii) holds then G contains a copy of T_1^+ by Theorem 1.5(2), completing the proof. $\square$

Observe that as with Lemma 3.4 and Lemma 3.23, a 'dual' form of Lemma 3.28 can be proved similarly. For this we instead require that $|F^-| \geqslant 6\alpha n$, $|Y| \geqslant 2|F^+| + \alpha n$ and $|Z| \geqslant 2|F^-| - 2\alpha n$, and also either that the largest component $(T_1^-)_\Delta$ of F^- contains at most $|F^-| - 3\alpha n$ vertices, or that $G[Z]$ is a robust (μ, ν)-outexpander with $\delta^0(G[Z]) \geqslant \eta|Z|$ and $|(T_1^-)_\Delta| \geqslant \beta n$, or that $\Delta(T_1^-) \leqslant \Delta$. If these conditions are met we may conclude that G contains a copy of T. As with Lemma 3.4, we sometimes implicitly refer to this 'dual' when referring to Lemma 3.28.

The next lemma is the final result we need to prove Theorem 1.1. It states that if we can find disjoint subsets $Y, Z \subseteq V(G)$ containing almost all of the vertices of G, so that $G[Y]$ and $G[Z]$ are robust outexpanders of large minimum semidegree with almost all edges between Y and Z directed the same way, then G contains a copy of T.

Lemma 3.29 *Suppose that $1/n \ll 1/\Delta \ll \mu \ll \nu \ll \eta \ll \gamma \ll \alpha \ll \beta \ll 1$. Let T be a directed tree on n vertices with $|T_\Delta| \geqslant \beta n$. Let Y and Z be disjoint sets with $|Y \cup Z| \geqslant (2-\alpha)n$, and let G be a tournament on vertex set $Y \cup Z$ such that*

(i) $G[Y]$ is a robust (μ, ν)-outexpander with $\delta^0(G[Y]) \geqslant \eta|Y|$,

(ii) $G[Z]$ is a robust (μ, ν)-outexpander with $\delta^0(G[Z]) \geqslant \eta|Z|$, and

(iii) every vertex of Y has at most $\gamma|G|$ outneighbours in Z, and every vertex of Z has at most $\gamma|G|$ inneighbours in Y.

Then G contains a copy of T.

Proof. If $|Y \cup Z| \geqslant (2+\alpha)n$, then G contains a copy of T by Theorem 1.5(1). So we may assume that $|Y \cup Z| = (2 \pm \alpha)n$. Suppose first that $|Z| < 64\alpha n$. Then $|Y| \geqslant (2-65\alpha)n$, and hence $G[Y]$ contains a copy of T by (i) and Lemma 2.32. Similarly if $|Y| < 64\alpha n$, then by (ii) and Lemma 2.32 $G[Z]$ contains a copy of T. So we may assume that $|Y| \geqslant 64\alpha n$ and $|Z| \geqslant 64\alpha n$.

So we may form a forest F_1^+ of order between $|Y|/2 + 4\alpha n$ and $|Y|/2 + 5\alpha n$ by repeatedly choosing a sink vertex of T, deleting it from T and adding it to F_1^+. Let $F_1^- := T - F_1^+$, so that

$$\frac{|Z|}{2} - 6\alpha n \leqslant n - \frac{|Y|}{2} - 5\alpha n \leqslant |F_1^-| \leqslant n - \frac{|Y|}{2} - 4\alpha n \leqslant \frac{|Z|}{2} - 3\alpha n. \tag{3.30}$$

We therefore have $|Y| \geqslant 2|F_1^+| - 10\alpha n$ and $|Z| \geqslant 2|F_1^-| + 6\alpha n$. Note also that $|F_1^+| \geqslant 36\alpha n$. Let T' be the largest component of F_1^+. If $|T'| \leqslant |F_1^+| - 18\alpha n$ or $|T'_\Delta| \geqslant \beta n/3$ then G contains a copy of T by (i), (iii) and Lemma 3.28. So we may assume that $|T'| > |F_1^+| - 18\alpha n$, and that $|T'_\Delta| < \beta n/3$.

Next we form a forest F_2^- which is a subgraph of T and which contains F_1^-. To do this, take F_2^- initially to be F_1^-. Then select a source vertex of F_1^+, delete it from F_1^+ and add it to F_2^-, and repeat this step until $|Z|/2 + 4\alpha n \leqslant |F_2^-| \leqslant |Z|/2 + 5\alpha n$, and let $F_2^+ := T - F_2^-$. Then by (3.30) we have $|F_1^+ \cap F_2^-| = |F_2^-| - |F_1^-| \leqslant 11\alpha n$. Also $|F_2^+| \leqslant |Y|/2 - 3\alpha n$, and so we have both $|Z| \geqslant 2|F_2^-| - 10\alpha n$ and $|Y| \geqslant 2|F_2^+| + 6\alpha n$. Observe also that $|F_2^-| \geqslant 36\alpha n$. Let T'' be the largest component of F_2^-. Then if $|T''| \leqslant |F_2^-| - 18\alpha n$ then G contains a copy of T by (ii), (iii) and Lemma 3.28. So we may assume that $|T''| > |F_2^-| - 18\alpha n$. Clearly $|T' \cap T''| \leqslant |F_1^+ \cap F_2^-| \leqslant 11\alpha n$, and so $|T' \cup T''| \geqslant |T'| + |T''| - |T' \cap T''| > (1-47\alpha)n$. This implies that $|T''_\Delta| \geqslant \beta n/3$, as otherwise by Lemma 2.21 we would have $|T_\Delta| < \beta n$, a contradiction. Thus G contains a copy of T by (ii), (iii) and Lemma 3.28, as desired. $\square$

Proof of Theorem 1.1. Introduce new constants with

$$\frac{1}{n} \ll \frac{1}{\Delta} \ll \mu \ll \nu \ll \eta \ll \gamma \ll \alpha \ll \alpha' \ll \beta \ll 1.$$

If $|T_\Delta| < \beta n$ then G contains a copy of T by Lemma 3.20. So we may assume that $|T_\Delta| \geqslant \beta n$. Let $x := |T_\Delta|$, let y be the outweight of T_Δ, and let z be the inweight of T_Δ, so $x + y + z = n$. Also let T_1 be the subtree of T formed by T_Δ and all outcomponents of T_Δ, and let T_2 be the subtree of T formed by T_Δ and all incomponents of T_Δ, so $|T_1| = x + y$, and $|T_2| = x + z$.

By Lemma 2.12 we may choose disjoint subsets $S_1, \dots, S_r$ of $V(G)$ such that

(i) $|\bigcup_{i \in [r]} S_i| \geqslant (1 - \gamma)|G|$,

(ii) for each $i \in [r]$, any vertex $v \in S_i$ has at most $\gamma|G|$ inneighbours in $\bigcup_{j>i} S_j$ and at most $\gamma|G|$ outneighbours in $\bigcup_{j<i} S_j$, and

(iii) for each $i \in [r]$, either $G[S_i]$ is a robust (μ, ν)-outexpander with $\delta^0(G[S_i]) \geqslant \eta|G|$ or $|S_i| < \gamma|G|$.

Let i be maximal such that $|S_1 \cup \dots \cup S_{i-1}| < \max\{2(z - \alpha n), 4\alpha n\}$, and let j be minimal such that $|S_{j+1} \cup \dots \cup S_r| < \max\{2(y - \alpha n), 4\alpha n\}$. Since $y + z \leqslant n - \beta n$, by (i) we have $i \leqslant j$ (though equality is possible here). Let $Z := S_1 \cup \dots \cup S_i$, let $Y := S_j \cup \dots \cup S_r$ and let $X := S_{i+1} \cup \dots \cup S_{j-1}$. Then we have

$$|Z \setminus S_i| < \max\{2(z - \alpha n), 4\alpha n\} \text{ and } |Y \setminus S_j| < \max\{2(y - \alpha n), 4\alpha n\}. \tag{3.31}$$

Also, by the maximality of i and the minimality of j we have

$$|Z| \geqslant z + \alpha n \text{ and } |Y| \geqslant y + \alpha n. \tag{3.32}$$

Claim. *If* $|Z \setminus S_i| \geqslant 11\alpha n$ *or* $|Y \setminus S_j| \geqslant 11\alpha n$ *then* G *contains a copy of* T.

To prove the claim, suppose first that $|Y \setminus S_j| \geqslant 11\alpha n$. Let $X^- := Z \cup X \cup S_j$ and $X^+ := Y \setminus S_j$. By (3.31) we have $|X^+| < 2y - 2\alpha n$. Also, by (ii) every vertex in X^- has at most $\gamma|G|$ inneighbours in X^+ and every vertex in X^+ has at most $\gamma|G|$ outneighbours in X^-. Now, $T_1 - T_\Delta$ is a forest on $y > |X^+|/2 + \alpha n$ vertices in which each component has order at most n/Δ by Proposition 2.19(iv). So by repeatedly deleting a source vertex of $T_1 - T_\Delta$, we may obtain a subforest F^+ on between $|X^+|/2 + 2\alpha n/3$ and $|X^+|/2 + \alpha n$ vertices. So $|F^+| \geqslant 6\alpha n$, and each component of F^+ has order at most $n/\Delta \leqslant |F^+| - 3\alpha n$. Let $F^+ := T - F^-$, so every edge of T between F^- and F^+ is directed from F^- to F^+. Since $|X^+| + |X^-| \geqslant (1-\gamma)|G|$ by (i), we have

$$|F^-| = n - |F^+| \leqslant n - \frac{|X^+|}{2} - \frac{2\alpha n}{3} \leqslant \frac{|X^-|}{2} - \frac{\alpha n}{2}.$$

So $|X^-| \geqslant 2|F^-| + \alpha n$, and $|X^+| \geqslant 2|F^+| - 2\alpha n$, and so G contains a copy of T by Lemma 3.28(i). If instead $|Z \setminus S_i| \geqslant 11\alpha n$ then G contains a copy of T similarly. This proves the claim.

We may therefore assume that $|Z \setminus S_i| < 11\alpha n$ and $|Y \setminus S_j| < 11\alpha n$. Suppose first that $i = j$. Then $|S_i| \geqslant (1-\gamma)|G| - 22\alpha n \geqslant (2-\alpha')n$, so by (iii) $G[S_i]$ is a robust (μ,ν)-outexpander with $\delta^0(G[S_i]) \geqslant \eta|G| \geqslant \eta|S_i|$. Thus G contains a copy of T by Lemma 2.32. Now suppose instead that $i \neq j$, and also that $|X| < 12\alpha' n$. Then $|S_i \cup S_j| \geqslant (1-\gamma)|G| - |X| - 22\alpha n \geqslant (2-13\alpha')n$. Now if $|S_i| < \gamma|G|$, then we must have $|S_j| \geqslant (2-14\alpha')n$. Then by (iii) $G[S_j]$ must be a robust (μ,ν)-outexpander with $\delta^0(G[S_j]) \geqslant \eta|G| \geqslant \eta|S_j|$, so $G[S_j]$ contains a copy of T by Lemma 2.32. Alternatively, if $|S_j| < \gamma|G|$ then $G[S_i]$ contains a copy of T similarly. Finally, if $|S_i|, |S_j| \geqslant \gamma|G|$, then by (iii) $G[S_i]$ and $G[S_j]$ must both be robust (μ,ν)-outexpanders with $\delta^0(G[S_i]) \geqslant \eta|G| \geqslant \eta|S_i|$ and $\delta^0(G[S_j]) \geqslant \eta|S_j|$. Also, by (ii) every vertex of S_i has at most $\gamma|G|$ inneighbours in S_j, and every vertex of S_j has at most $\gamma|G|$ outneighbours in S_i. So $G[S_i \cup S_j]$ contains a copy of T by Lemma 3.29.

So we may assume that $i \neq j$, and also that $|X| \geqslant 12\alpha' n$. We next consider two cases for the size of X, in each case showing that T may be embedded in G.

Case 1: $|X| \geqslant (1+\alpha)x$**.**

Since by Proposition 2.19(iii) we have $\Delta(T_\Delta) \leqslant \Delta$, by Theorem 1.5(2) we may embed T_Δ in $G[X]$. Let $X' \subseteq X$ consist of the vertices occupied by this embedding. Now, by (ii) every vertex of X' has at most $\gamma|G|$ inneighbours in Y, and hence by (3.32) at least $y + \alpha n/2$ outneighbours in Y. Since by Proposition 2.19(iv) every component of $T_1 - T_\Delta$ has order at most n/Δ, by Lemma 3.3(c) we may extend the embedding of T_Δ in $G[X']$ to an embedding of T_1 in $G[X' \cup Y]$. Similarly by (ii) every vertex of X' has at most $\gamma|G|$ outneighbours in Z, and hence by (3.32) at least $z + \alpha n/2$ inneighbours in Z. Since by Proposition 2.19(iv) every component of $T_2 - T_\Delta$ has order at most n/Δ, by Lemma 3.3(c) we may extend the embedding of T_Δ in $G[X']$ to an embedding of T_2 in $G[X' \cup Z]$. Since these embeddings of T_1 and T_2 only overlap in T_Δ, they together form an embedding of T in G.

Case 2: $|X| < (1+\alpha)x$**.**

Observe that if $|Z| \leqslant 2z + \alpha n$ and $|Y| \leqslant 2y + \alpha n$, then by (i) and the fact that $x = |T_\Delta| \geqslant \beta n$ we have

$$|X| \geqslant (1-\gamma)|G| - |Z| - |Y| \geqslant 2n - 2z - 2y - 3\alpha n \geqslant 2x - 3\alpha n \geqslant (1+\alpha)x,$$

contradicting our assumption on X. So at least one of $|Z| > 2z + \alpha n$ and $|Y| > 2y + \alpha n$ must hold. This gives us three further cases, which we consider separately.

Case 2(a): $|Z| > 2z + \alpha n$**,** $|Y| \leqslant 2y + \alpha n$**.**

In this case it is sufficient to embed T_2 in $G[X \cup Z]$. Indeed, by (ii) every vertex of $X \cup Z$ has at most $\gamma|G|$ inneighbours in Y, and therefore by (3.32) at least $y + \alpha n/2$ outneighbours in Y. Since by Proposition 2.19(iv) every component of $T - T_2$ has order at most n/Δ, any embedding of T_2 in $G[X \cup Z]$ can be extended to an embedding of T in G by Lemma 3.3(c).

Now, if $|X \cup Z| \geqslant 2|T_2| + 2\alpha n$, then we may embed T_2 in $G[X \cup Z]$ by Theorem 1.5(1). So we may assume that $|X \cup Z| < 2|T_2| + 2\alpha n$. Also, by (i) we have

$$|X \cup Z| \geqslant (1-\gamma)|G| - |Y| \geqslant 2n - 2y - 2\alpha n = 2x + 2z - 2\alpha n = 2|T_2| - 2\alpha n.$$

So $|X \cup Z| = 2|T_2| \pm 2\alpha n$. In particular, since $|T_2| \geqslant |T_\Delta| \geqslant \beta n$, we have $|X \cup Z| \geqslant \beta n$. By repeatedly deleting a source vertex of T_Δ, we may form a forest F which is an induced subgraph of T_Δ (consisting of the undeleted vertices of T_Δ) so that every edge between $T_\Delta - F$ and F is directed from T_Δ to F, and also so that

$$\frac{|X|}{2} + \frac{2\alpha'|T_2|}{3} \leqslant |F| \leqslant \frac{|X|}{2} + \alpha'|T_2|.$$

Let $F^- := T_2 - F$. Then

$$|F^-| = |T_2| - |F| \leqslant |T_2| - \frac{|X|}{2} - \frac{2\alpha'|T_2|}{3} \leqslant \frac{|Z|}{2} - \frac{\alpha'|T_2|}{2}.$$

So $|X| \geqslant 2|F| - 2\alpha'|T_2|$ and $|Z| \geqslant 2|F^-| + \alpha'|T_2|$. Also, $|F| \geqslant |X|/2 \geqslant 6\alpha'|T_2|$, and since F is a subtree of T_Δ, by Proposition 2.19(iii) each component C of F has $\Delta(C) \leqslant \Delta$. Since by (ii) every vertex of X has at most $\gamma|G| \leqslant 2\gamma|X \cup Z|/\beta$ outneighbours in Z and every vertex of Z has at most $\gamma|G| \leqslant 2\gamma|X \cup Z|/\beta$ inneighbours in X, $G[X \cup Z]$ contains a copy of T_2 by Lemma 3.28, as required.

Case 2(b): $|Z| \leqslant 2z + \alpha n$**,** $|Y| > 2y + \alpha n$**.**

In this case T may be embedded in G by the same method as in the previous case, with the roles of inneighbours and outneighbours switched. So we begin by embedding T_1 in $G[X \cup Y]$, and then use Lemma 3.3(c) to extend this embedding to an embedding of T in G.

Case 2(c): $|Z| > 2z + \alpha n$**,** $|Y| > 2y + \alpha n$**.**

In this case, we partition T into three forests as follows. Initially take F^- to be the forest formed by all incomponents of T_Δ, and F^+ to be the forest formed by all outcomponents of T_Δ. Then select a source vertex of T_Δ, delete it from T_Δ and add it to F^-. Repeat this step until $2|F^-| + \alpha n \leqslant |Z| \leqslant 2|F^-| + 2\alpha n$. Next, select a sink vertex of T_Δ, delete it from T_Δ and add it to F^+. Repeat this step until $2|F^+| + \alpha n \leqslant |Y| \leqslant 2|F^+| + 2\alpha n$. Then let F consist of all vertices remaining in T_Δ. So F is a subgraph of T_Δ. Also, by (i)

$$|F| = n - |F^-| - |F^+| \leqslant n - \frac{|Y|}{2} - \frac{|Z|}{2} + 2\alpha n \leqslant \frac{|X|}{2} + 3\alpha n,$$

so (since $|X| \geqslant \alpha' n$) $|X| \geqslant |F| + \alpha n$. We embed the components of F^-, F and F^+ in turn amongst the vertices of Z, X and Y respectively. Indeed, the proof is similar to the proof of Lemma 3.4, but with three forests instead of two.

Let $C_1, \dots, C_s$ be the components of F^-, F and F^+, ordered so that C_1 is a component of F, and for each $i \in [s-1]$, C_{i+1} has precisely one neighbour in $C_1 \cup \dots \cup C_i$. We embed the C_i in turn, so that each component of F^- is embedded in $G[Z]$, each component of F is embedded in $G[X]$, and each component of F^+ is embedded in $G[Y]$. We also require that after each C_i is embedded, the embeddings of $C_1, \dots, C_i$ together form an embedding in G of the subtree of T induced by the vertices of $C_1, \dots, C_i$. So suppose that we have successfully embedded $C_1, \dots, C_{i-1}$ in this manner, and we now wish to extend this embedding to include C_i. Then if $i \geqslant 2$, there is precisely one edge of T between C_i and $C_1 \cup \dots \cup C_{i-1}$. Let t be the endvertex of this edge in $C_1 \cup \dots \cup C_{i-1}$, and let v be the vertex to which t was embedded. If C_i is a component of F^-, then $i \geqslant 2$, the edge between t and C_i is directed towards t and $v \in X \cup Y$. So we may let S consist of the inneighbours of v in Z. Then by (ii) we have $|S| \geqslant |Z| - \gamma|G|$. Let $S' \subseteq S$ consist of the unoccupied vertices of S. Since at most $|F^-| - |C_i|$ vertices of S are occupied by the embeddings of $C_1, \dots, C_{i-1}$,

$$|S'| \geqslant |Z| - \gamma|G| - |F^-| + |C_i| \geqslant 2|C_i| + \frac{\alpha n}{2}.$$

So if $|C_i| < \alpha n/2$ then $G[S']$ contains a copy of T by Theorem 1.2, and if $|C_i| \geqslant \alpha n/2$ then $G[S']$ contains a copy of T by Theorem 1.5(1). Alternatively, if C_i is a component of F^+, then $i \geqslant 2$, the edge between t and C_i is directed towards C_i and $v \in X \cup Z$. So we may let S consist of the outneighbours of v in Y, and let $S' \subseteq S$ consist of the unoccupied vertices of S. Then we may embed C_i in S' by the same argument as used when C_i is a component of F^-. Finally, suppose that C_i is a component of F. Then if $i \geqslant 2$ and $t \in F^+$, let S consist of the inneighbours of v in X. If instead $i \geqslant 2$ and $t \in F^-$, let S consist of the outneighbours of v in X. If $i = 1$ then let $S = X$. Then by (ii) we have $|S| \geqslant |X| - \gamma|G|$. Again let $S' \subseteq S$ consist of the unoccupied vertices of S. Then it suffices to embed C_i in $G[S']$. Since at most $|F| - |C_i|$ vertices have been embedded in S, we have $|S'| \geqslant |X| - \gamma|G| - |F| + |C_i| \geqslant |C_i| + \alpha n/2$. Now, C_i is a subtree of T_Δ, so $\Delta(C_i) \leqslant \Delta$ by Proposition 2.19(iii). So if $|C_i| \geqslant \alpha n/4$, then $G[S']$ contains a copy of C_i by Theorem 1.5(2). On the other hand, if $|C_i| < \alpha n/4$, then $G[S']$ contains a copy of C_i by Theorem 1.2. So in any case we may embed C_i as desired, completing the proof. $\square$

CHAPTER 4

LOOSE HAMILTON CYCLES IN UNIFORM HYPERGRAPHS

In this chapter we prove Theorem 1.10, which is restated below for ease of reference.

Theorem 1.10 *For all $k \geqslant 3$ and any $\eta > 0$ there exists n_0 so that if $n \geqslant n_0$ then any k-graph $\mathcal{H}$ on n vertices with $\delta(\mathcal{H}) \geqslant (\frac{1}{2(k-1)} + \eta)n$ contains a loose Hamilton cycle.*

4.1 Extremal examples

Before proceeding to the proof of Theorem 1.10, in this section we present the constructions which show that Theorems 1.9, 1.10 and 1.11 are each best possible up to the error term ηn. These constructions are well known, but we include them here for completeness.

The following proposition shows that, if $(k-\ell)|k$, then Theorem 1.9 is best possible up to the error term ηn. By a *perfect matching* in a k-graph $\mathcal{H}$, we mean a set of disjoint edges of $\mathcal{H}$ whose union contains every vertex of $\mathcal{H}$.

Proposition 4.1 *For all $k \geqslant 3$, $1 \leqslant \ell \leqslant k-1$ and every $n \geqslant 3k$ such that $(k-\ell)|k$ and $k|n$ there exists a k-graph $\mathcal{H}$ on n vertices with $\delta(\mathcal{H}) \geqslant \frac{n}{2} - k$ which does not contain a Hamilton ℓ-cycle.*

Proof. Choose $\frac{n}{2} - 1 \leqslant a \leqslant \frac{n}{2} + 1$ so that a is odd. Let V_1 and V_2 be disjoint sets of size a and $n-a$ respectively, and let $\mathcal{H}$ be the k-graph on vertex set $V = V_1 \cup V_2$ and with all those k-element subsets S of V such that $|S \cap V_1|$ is even as edges. Then $\delta(\mathcal{H}) \geqslant \min(a, n-a) - k + 1 \geqslant \frac{n}{2} - k$. Now, any Hamilton ℓ-cycle C in $\mathcal{H}$ would contain a perfect matching, consisting of every $\frac{k}{k-\ell}$th edge of C. Every edge in this matching would contain an even number of vertices from V_1, and so $|V_1|$ would be even. Since $|V_1| = a$ is odd, $\mathcal{H}$ cannot contain a Hamilton ℓ-cycle. $\square$

A recent construction of Markström and Ruciński ([35]) shows that Proposition 4.1 still holds if we drop the requirement that $k \mid n$. Part (i) of the next proposition gives a lower bound on the minimum degree required to guarantee that a k-graph on n vertices contains a Hamilton cycle (of any kind). In particular, this shows that Theorem 1.10 is best possible, up to the error term ηn. Likewise, part (ii) shows that Theorem 1.11 is best possible, again up to the error term ηn.

Proposition 4.2 *For all integers $k \geqslant 3$, $1 \leqslant \ell \leqslant k-1$ and every $n \geqslant 2k-1$,*

(i) there exists a k-graph $\mathcal{H}$ on n vertices with $\delta(\mathcal{H}) \geqslant \lceil \frac{n}{2k-2} \rceil - 1$ which does not contain a Hamilton cycle.

(ii) there exists a k-graph $\mathcal{H}'$ on n vertices with $\delta(\mathcal{H}') \geqslant \lceil \frac{n}{\lceil \frac{k}{k-\ell} \rceil (k-\ell)} \rceil - 1$ which does not contain a Hamilton ℓ-cycle.

Proof. For (i), let V_1 and V_2 be disjoint sets of size $\lceil \frac{n}{2k-2} \rceil - 1$ and $n - \lceil \frac{n}{2k-2} \rceil + 1$ respectively. Let $\mathcal{H}$ be the k-graph on vertex set $V = V_1 \cup V_2$ whose edges are all those k-sets of vertices which contain at least one vertex from V_1. Then $\mathcal{H}$ has minimum degree $\delta(\mathcal{H}) = \lceil \frac{n}{2k-2} \rceil - 1$. However, any cyclic ordering of the vertices of $\mathcal{H}$ must contain $2k-2$ consecutive vertices $v_1, \dots, v_{2k-2}$ from V_2, but then v_{k-1} and v_k cannot be contained in a common edge consisting of k consecutive vertices, and so $\mathcal{H}$ cannot contain a Hamilton cycle.

For (ii), let $a := \lceil \frac{n}{\lceil \frac{k}{k-\ell} \rceil (k-\ell)} \rceil - 1$ and form $\mathcal{H}'$ as in (i), with disjoint vertex sets V_1' and V_2' of sizes a and $n-a$ respectively. Then $\delta(\mathcal{H}') = a$. However, an ℓ-cycle on n vertices has $n/(k-\ell)$ edges and every vertex on such a cycle lies in at most $\lceil \frac{k}{k-\ell} \rceil$ edges. Since $\lceil \frac{k}{k-\ell} \rceil |V_1'| < n/(k-\ell)$, $\mathcal{H}'$ cannot contain a Hamilton ℓ-cycle. $\square$

4.2 Outline of the proof of Theorem 1.10

In our proof of Theorem 1.10 we construct the loose Hamilton cycle by finding several paths and joining them into a spanning cycle. Here a k-graph P is a *path* if its vertices can be given a linear ordering such that every edge of P consists of k consecutive vertices, and so that every pair of consecutive vertices of P lie in an edge of P. Similarly as for cycles, we say that a path P is *loose* if edges of P intersect in at most one vertex. The ordering of the vertices of P naturally gives an ordering of the edges of P. We say that any vertex of P which lies in the initial edge of P, but not the second edge of P, is an *initial vertex*. Similarly, any vertex of P which lies in the final edge of P but not the penultimate edge is a *final vertex*. Also, we refer to vertices of P which lie in more than one edge of P as *link vertices*. So a loose path P has $k-1$ initial vertices, $k-1$ final vertices, and one link vertex in each pair of consecutive edges.

In Section 4.3, we introduce various ideas we need in the proof of Theorem 1.10. In particular, we state an analogue of the Szemerédi regularity lemma for hypergraphs due to Rödl and Schacht [41] and Theorem 4.5 due to Keevash [23]. The latter provides a useful way of applying the hypergraph blow-up lemma. In Section 4.4, we prove various auxiliary results, including a result on finding loose paths in complete k-partite k-graphs, and an approximate minimum degree condition to guarantee a near-perfect packing of H with a particular k-graph $\mathcal{A}_k$. Finally, in Section 4.5 we use the regularity method to prove Theorem 1.10 as follows.

4.2.1 Imposing structure on H

We begin in Section 4.5.1 with the following steps, which correspond to those in the description of the regularity method in the introduction.

1. We apply the hypergraph regularity lemma to partition the vertex set of H into clusters.
2. Next, we define a suitable 'reduced k-graph' R of H, as discussed in the introduction.
3. We find copies of a suitable auxiliary k-graph $\mathcal{A}_k$ covering almost all vertices of R.

We use this structure of the reduced k-graph R to find a Hamilton cycle in H. In the remaining part of Section 4.5.1, we split the sub-k-graph of H corresponding to each copy of $\mathcal{A}_k$ in R into the same number of vertex-disjoint k-partite k-graphs H^i on vertex sets X^i. These are suitable for embedding almost spanning loose paths (the sizes of the vertex classes of each H^i are chosen to meet this condition). We also form an 'exceptional' loose path L_e which contains all of the vertices of H not contained in any of the X^i (actually, if $|V(H)|$ is not divisible by $k-1$, then L_e contains two consecutive edges which intersect in more than one vertex).

4.2.2 The linking strategy

To complete the proof, in Section 4.5.3 we use the structure imposed on H to find a Hamilton cycle in H by the following process.

(a) The k-graphs H^i are connected by means of a walk $W = e_1, \dots, e_\ell$ in the 'supplementary graph'. This graph (which we define in Section 4.5.2) has vertices $1, \dots, t'$ corresponding to the k-graphs H^i.

(b) Using Lemma 4.19, each edge e_j of W is used to create a short 'connecting' loose path L_j in H joining two different H^is.

(c) L_e and the paths L_j are extended to 'prepaths' (these can be thought of as a path minus an initial vertex and a final vertex) $L_e^* = I_0 L_e F_0$ and $L_j^* = I_j L_j F_j$, where I_0, F_0 and all I_j, F_j are sets of size $k-2$. These prepaths have the property that there are large sets I'_j and F'_j such that L_j^* can be extended to a loose path by adding any vertex of I'_j as an initial vertex and any vertex of F'_j as a final vertex. Similarly there are large sets $I'_{\ell+1}$ and F'_0 so that L_e^* can be extended to a path by adding any vertex of $I'_{\ell+1}$ as an initial vertex and any vertex of F'_0 as a final vertex. I'_{j+1} and F'_j both lie in the same H^i (for all $j = 0, \dots, \ell$).

(d) For each H^i and for all those pairs I'_{j+1}, F'_j which lie in H^i, we choose a loose path L'_{j+1} inside H^i from F'_j to I'_{j+1}. For each i, we use the hypergraph blow-up lemma (in the form of Theorem 4.5) to ensure that together all those L'_j which lie in H^i use all the remaining vertices of H^i.

(e) The loose Hamilton cycle is then the concatenation $L_e^* L'_1 L_1^* \dots L'_\ell L_\ell^* L'_{\ell+1}$.

4.2.3 Controlling divisibility

Note that the number of vertices of a loose path is 1 modulo $k-1$. So in order to apply Theorem 4.5 to obtain spanning loose paths in a subgraph of H^i, we need this subgraph to satisfy this condition. So we choose our paths sequentially to satisfy the following congruences modulo $k-1$.

(a) L_e is chosen with $|V(H) \setminus V(L_e)| \equiv -1$.

(b) Let $X^i(j-1)$ be the subset of X^i obtained by removing $V(L_1), \dots, V(L_{j-1})$. (All the X^i are disjoint from $V(L_e)$.) Let d_i be the number of times that W visits H^i. When choosing L_j, for every X^i it traverses (except the final one) we arrange to intersect $X^i(j-1)$ in a set of size $\equiv t_i(j) \equiv |X^i(j-1)| + d_i$ (the size modulo $k-1$ of the intersection of L_j with the final X^i it traverses is then determined by the sizes of the other intersections). The choice of L_e in (a) ensures that after all L_j have been picked, the remaining part $X^i(\ell)$ of X^i has size $\equiv -d_i$.

(c) Each L_j is extended to a prepath L_j^* by adding I_j and F_j. Similarly, L_e is extended into a prepath L_e^* by adding I_0 and F_0. Now the remaining part of X^i has size $\equiv d_i$.

(d) It remains to select d_i paths L_j' within each X^i: each uses $\equiv 1$ vertices, so the divisibility conditions are satisfied.

4.3 Regularity and the blow-up lemma

4.3.1 Graphs and complexes

A *hypergraph* H consists of a vertex set $V(H)$ and an edge set $E(H)$, where each edge e satisfies $e \subseteq V(H)$. In particular, a k-graph is a hypergraph in which every edge has size k. A hypergraph H is a k-*complex* if every edge has size at most k and H forms a simplicial complex, *i.e.* if $e_1 \in H$ and $e_2 \subseteq e_1$ then $e_2 \in H$. Throughout Chapter 4, we frequently identify a hypergraph H with its edge set, writing $e \in H$ if e is an edge of H and using $|H|$ to denote the number of edges of H.

We say that a hypergraph H is r-*partite* if its vertex set X is divided into r parts $X_1, \dots, X_r$ so that for any edge $e \in H$, $|e \cap X_i| \leqslant 1$ for each i. We call the X_i the *vertex classes* of H and say that the partition $X_1, \ldots, X_r$ of X is *equitable* if every X_i has the same size. A set $A \subseteq X$ is r*-partite* if $|A \cap X_i| \leqslant 1$ for each i. So every edge of an r-partite hypergraph is r-partite. In the same way we consider r-partite k-graphs and r-partite k-complexes. Given a k-graph H, we define a k-complex $H^{\leqslant} = \{e_1 \colon e_1 \subseteq e_2 \text{ and } e_2 \in H\}$ and a $(k-1)$-complex $H^{<} = \{e_1 \colon e_1 \subset e_2$ and $e_2 \in H\}$. Conversely, for a k-complex H we define the k-graph $H_=$ to be the 'top level' of H, *i.e.* $H_= = \{e \in H \colon |e| = k\}$. (Here $V(H) = V(H^{\leqslant}) = V(H^{<}) = V(H_=)$.)

Given a k-graph G and a set W of vertices of G, $G[W]$ denotes the sub-k-graph of G obtained by removing all vertices and edges not contained in W (in this case, we say G is *restricted to* W). For a k-graph G and a sub-k-graph $H \subseteq G$ write $G - H$ for $G[V(G) \setminus V(H)]$.

Now, let $X_1, \ldots, X_r$ be disjoint vertex sets, and let $X = X_1 \cup \cdots \cup X_r$. Given $A \in \binom{[r]}{\leqslant k}$, we write $K_A(X)$ for the complete $|A|$-partite $|A|$-graph whose vertex classes are all the X_i with $i \in A$. The *index* of an r-partite subset S of X is $i(S) = \{i \in [r] : S \cap X_i \neq \emptyset\}$. Furthermore, given any set $B \subseteq i(S)$, we write $S_B = S \cap \bigcup_{i \in B} X_i$. Similarly, given $A \in \binom{[r]}{\leqslant k}$ and an r-partite k-graph or k-complex H on vertex set X we write H_A for the collection of edges in H of index A and let $H_\emptyset = \{\emptyset\}$. In particular, if H is a k-complex then $H_{\{i\}}$ is the set of all those vertices in X_i which lie in an edge of H (and thus form a (singleton) edge of H). In general, we often view H_A as an $|A|$-partite $|A|$-graph with vertex set X_A. Also, given a k-complex H we similarly write $H_{A^{\leqslant}} = \bigcup_{B \subseteq A} H_B$ and $H_{A^{<}} = \bigcup_{B \subset A} H_B$. We write H^*_A for the $|A|$-graph whose edges are those r-partite sets $S \subseteq X$ of index A for which all proper subsets of S belong to H. (In other words, a set S with index A satisfies $S \in H^*_A$ if and only if for all $j < |A|$ the edges of H which have size j and are subsets of S form a complete j-graph on $|S|$ vertices.) Then the *relative density of H at index A* is $d_A(H) = |H_A|/|H^*_A|$. The *absolute density of H_A* is $d(H_A) = |H_A|/|K_A(X)|$. (Note that $|K_A(X)| = \prod_{i \in A} |X_i|$.) If H is a k-partite k-complex we may simply write $d(H)$ for $d(H_{[k]})$. Similarly, the *density* of a k-partite k-graph H on

$X = X_1 \cup \cdots \cup X_k$ is $d(H) = |H|/|K_{[k]}(X)|$.

Finally, for any vertex v of a hypergraph H, we define the *vertex degree* $d(v)$ of v to be the number of edges of H which contain v. Note that this is not the same as the degree defined earlier, which was for sets of $k-1$ vertices. The *maximum vertex degree* of a hypergraph H is the maximum of $d(v)$ taken over all vertices v of H. The *vertex neighbourhood* $VN(v)$ of v is the set of all vertices $u \in H$ for which some edge of H contains both u and v. For a k-partite k-complex H on vertex set $X_1 \cup \cdots \cup X_k$ we also define the *neighbourhood complex* $H(v)$ of a vertex $v \in X_i$ for some i to be the $(k-1)$-partite $(k-1)$-complex with vertex set $\bigcup_{j \neq i} X_j$ and edge set $\{e \in H : e \cup \{x\} \in H\}$.

4.3.2 Regular complexes

In this subsection we define the concept of regular complexes (which was first introduced in the k-uniform case by Rödl and Skokan [43]) in the form used by Rödl and Schacht [41]. This is a generalisation of the concept of regularity in graphs. Roughly speaking, we say that a k-complex G is regular if the restriction of G to any large subcomplex of lower rank has similar densities to G. More precisely, let G be an r-partite k-complex on the vertex set $X = X_1 \cup \cdots \cup X_r$. For any $A \in \binom{[r]}{\leqslant k}$, we say that G_A is *ε-regular* if for any $H \subseteq G_{A^<}$ with $|H^*_A| \geqslant \varepsilon |G^*_A|$ we have

$$\frac{|G_A \cap H^*_A|}{|H^*_A|} = d_A(G) \pm \varepsilon.$$

We say G is *ε-regular* if G_A is ε-regular for every $A \in \binom{[r]}{\leqslant k}$. To illustrate the definition for $k = 3$, suppose that $A = [3]$. Then for instance the top level of $G_{[2]}$ is the bipartite subgraph of G induced by X_1 and X_2 and G^*_A is the set of (graph) triangles in G. So roughly speaking, the regularity condition states that if we consider a subgraph of $G_{[2]} \cup G_{\{1,3\}} \cup G_{\{2,3\}}$ which spans a large number of triangles, then the proportion of these which also form an edge of G_A is close to $d_A(G)$, *i.e.* close to the proportion of (graph) triangles in G between X_1, X_2 and X_3

which form an edge of G.

Roughly speaking, the hypergraph regularity lemma states that an arbitrary k-graph can be split into pieces, each of which forms a regular k-complex. The version of the regularity lemma we use also involves the notion of a 'partition complex', which is a certain partition of the edges of a complete k-complex. As before, let $X = X_1 \cup \cdots \cup X_r$ be an r-partite vertex set. A *partition k-system P on X* consists of a partition P_A of the edges of $K_A(X)$ for each $A \in \binom{[r]}{\leqslant k}$. We refer to the partition classes of P_A as *cells*. So every edge of $K_A(X)$ is contained in precisely one cell of P_A. P is a *partition k-complex on X* if it also has the property that whenever $S, S' \in K_A(X)$ lie in the same cell of P_A, we have that S_B and S'_B lie in the same cell of P_B for any $B \subseteq A$. This property of S, S' forms an equivalence relation on the edges of $K_A(X)$, which we refer to as *strong equivalence*. To illustrate this, again suppose that $k = 3$ and $A = [3]$. Then if P is a partition k-complex, $P_{\{1\}}$, $P_{\{2\}}$ and $P_{\{3\}}$ together yield a vertex partition Q_1 refining X_1, X_2, X_3. Q_1 naturally induces a partition Q_2 of the 3 complete bipartite graphs induced by the pairs X_i, X_j. $P_{\{1,2\}}$, $P_{\{2,3\}}$ and $P_{\{1,3\}}$ also yield a partition Q'_2 of these complete bipartite graphs. The requirement of strong equivalence now implies that Q'_2 is a refinement of Q_2. At the next level, Q'_2 naturally induces a partition Q_3 of the set of triples induced by X_1, X_2 and X_3. As before, strong equivalence implies that the partition $P_{\{1,2,3\}}$ of these triples is a refinement of Q_3.

Let P be a partition k-complex on $X = X_1 \cup \cdots \cup X_r$. For $i \in [k]$, the cells of $P_{\{i\}}$ are called *clusters* (so each cluster is a subset of some X_i). We say that P is *vertex-equitable* if all clusters have the same size. P is *a-bounded* if $|P_A| \leqslant a$ for every A (*i.e.* if $K_A(X)$ is divided into at most a cells by the partition P_A). Also, for any r-partite set $Q \in \binom{X}{\leqslant k}$, we write C_Q for the set of all edges lying in the same cell of P as Q, and write $C_{Q^{\leqslant}}$ for the r-partite k-complex whose vertex set is X and whose edge set is $\bigcup_{Q' \subseteq Q} C_{Q'}$. (Since P is a partition k-complex, $C_{Q^{\leqslant}}$ is indeed a complex.) The partition k-complex P is *ε-regular* if $C_{Q^{\leqslant}}$ is ε-regular for every

r-partite $Q \in \binom{X}{\leqslant k}$.

Given a partition $(k-1)$-complex P on X and $A \in \binom{[r]}{k}$, we can define an equivalence relation on the edges of $K_A(X)$, namely that $S, S' \in K_A(X)$ are equivalent if and only if S_B and S'_B lie in the same cell of P for any strict subset $B \subset A$. We refer to this as *weak equivalence*. Note that if the partition complex P is a-bounded, then $K_A(X)$ is divided into at most a^k classes by weak equivalence. If we let G be an r-partite k-graph on X, then we can use weak equivalence to refine the partition $\{G_A, K_A(X) \setminus G_A\}$ of $K_A(X)$ (*i.e.* two edges of G_A are in the same cell if they are weakly equivalent and similarly for the edges not in G_A). Together with P, this yields a partition k-complex which we denote by $G[P]$. If $G[P]$ is ε-regular then we say that G is *perfectly ε-regular with respect to P*. Note that if $G[P]$ is ε-regular then P must be ε-regular too.

Finally, we say that r-partite k-graphs G and H on X are *ν-close* if $|G_A \triangle H_A| < \nu|K_A(X)|$ for every $A \in \binom{[r]}{k}$, *i.e.* if there are few edges contained in G but not in H and vice versa.

The version of the regularity lemma which we use to split our k-graph H into regular k-complexes actually states that there is some k-graph G which is close to H and which is regular with respect to some partition complex. This is sufficient for our purposes, as we avoid the use of any edges in $G \setminus H$, so every edge used lies in both G and H. There are various other forms of the regularity lemma for k-graphs which give information on H itself (the first of these were proved in [43, 13]); we use one such form in Chapter 5. However, these versions do not have the hierarchy of densities necessary for the application of the blow-up lemma (see [23] for a fuller discussion of this point). This version is due to Rödl and Schacht [41] (actually it is a very slight restatement of their result).

Theorem 4.3 (Theorem 14, [41]) *Suppose integers n, a, r, k and reals ε, ν satisfy $1/n \ll \varepsilon \ll 1/a \ll \nu, 1/r, 1/k$ and where $a!r$ divides n. Suppose also that H is an r-partite k-graph whose vertex classes $X_1, \ldots, X_r$ form an equitable partition of its vertex set X, where $|X| = n$. Then* *there is an a-bounded ε-regular vertex-equitable partition $(k-1)$-complex P on X and an r-partite k-graph G on X that is ν-close to H and perfectly ε-regular with respect to P.*

One important property of regular complexes is that they remain regular when restricted to a large subset of their vertex set. For regular k-partite k-complexes this property is formalised by the following lemma, a special case of Lemma 6.18 in [23]:

Lemma 4.4 (Restriction of regular complexes) *Suppose $0 < \varepsilon \ll \varepsilon' \ll d \ll c \ll 1/k$, and that G is an ε-regular k-partite k-complex on vertex set $X = X_1 \cup \cdots \cup X_k$ such that $G_{\{i\}} = X_i$ for each i and $d(G) > d$. Let W be a subset of X such that $|W \cap X_i| \geqslant c|X_i|$ for each i. Then the restriction $G[W]$ of G to W is ε'-regular, with $d(G[W]) > d(G)/2$ and $d_{[k]}(G[W]) > d_{[k]}(G)/2$.*

4.3.3 Robustly universal complexes

Apart from Theorem 4.3, the other main tool we use in the proof of Theorem 1.10 is the recent hypergraph blow-up lemma of Keevash. This result involves not only a k-complex G, but also a k-graph M of 'marked' edges on the same vertex set. If the pair (G, M) is 'super-regular', then this blow-up lemma can be applied to embed any spanning bounded-degree k-complex in $G \setminus M$, *i.e.* within G but avoiding any marked edges. We apply this with $M = G \setminus H$ where G is the k-graph given by Theorem 4.3. Super-regularity is a stronger notion than regularity. A result in [23] states that every ε-regular k-complex can be made super-regular by deleting a few of its vertices. Unfortunately, the notion of hypergraph super-regularity is very technical, but the following definition avoids many of these technicalities. Let J' be a k-partite k-complex. Roughly speaking, we say that J' is robustly D-universal if the following holds: even after deletion of many vertices of J', the resulting complex J has the property that one can find in J a copy of any k-partite k-complex L which has vertex degree at most D and whose vertex classes are the same as those of J. Condition (i) puts a natural restriction on the number of vertices we are allowed to delete from the neighbourhood complex of a vertex of G' and condition (iii) states that for a few vertices u of L we can even prescribe a 'target set' in G into which u is embedded.

Definition. (Robustly universal complexes) Suppose that J' is a k-partite k-complex on $V' = V'_1 \cup \cdots \cup V'_k$ with $J'_{\{i\}} = V'_i$ for each $i \in [k]$. We say that J' is (c, c_0)*-robustly* D*-universal* if whenever

(i) $V_j \subseteq V'_j$ are sets with $|V_j| \geqslant c|V'_j|$ for all $j \in [k]$, such that writing $V = \bigcup_{j\in[k]} V_j$ and $J = J'[V]$ we have $|J(v)_=| \geqslant c|J'(v)_=|$ for any $j \in [k]$ and $v \in V_j$,

(ii) L is a k-partite k-complex of maximum vertex degree at most D on some vertex set $U = U_1 \cup \cdots \cup U_k$ with $|U_j| = |V_j|$ for all $j \in [k]$,

(iii) $U_* \subseteq U$ satisfies $|U_* \cap U_j| \leqslant c_0|X_j|$ for every $j \in [k]$, and sets $Z_u \subseteq V_i(u)$ satisfy $|Z_u| \geqslant c|V_{i(u)}|$ for each $u \in U_*$,

then J contains a copy of L, in which for each $j \in [k]$ the vertices of U_j correspond to the vertices of V_j, and u corresponds to a vertex of Z_u for every $u \in U_*$.

So our use of the blow-up lemma is hidden through this definition. Of course, we need to obtain robustly universal complexes. This is the purpose of the next theorem, which states that given a regular k-partite k-complex G with sufficient density, and a k-partite k-graph M on the same vertex set which is small relative to G, we can delete a small number of vertices from their common vertex set so that $G \setminus M$ is robustly universal. It is a special case of Theorem 6.32 in [23].

Theorem 4.5 (Theorem 6.32, [23]) *Suppose that* $1/n \ll \varepsilon \ll c_0 \ll d^* \ll d_a \ll \theta \ll d, c, 1/k, 1/D, 1/C$*,* G *is a* k*-partite* k*-complex on* $V = V_1 \cup \cdots \cup V_k$ *with* $n \leqslant |G_{\{j\}}| =$

$|V_j| \leqslant Cn$ for every $j \in [k]$, G is ε-regular with $d_{[k]}(G) \geqslant d$ and $d(G_{[k]}) \geqslant d_a$, and $M \subseteq G_=$ with $|M| \leqslant \theta|G_=|$. Then we can delete at most $2\theta^{1/3}|V_j|$ vertices from each V_j to obtain $V' = V'_1, \ldots, V'_k$, $G' = G[V']$ and $M' = M[V']$ such that

(i) $d(G') > d^*$ and $|G'(v)_=| > d^*|G'_=|/|V'_i|$ for every $v \in V'_i$, and

(ii) $G' \setminus M'$ is (c, c_0)-robustly D-universal.

4.4 Preliminary results

In this section we collect the preliminary results we need to prove Theorem 1.10. In order to apply Theorem 4.5, we need to know under what conditions we can find particular loose paths in complete k-partite k-graphs, which is the topic of the next subsection.

4.4.1 Loose paths in complete graphs

The problem of when we can find particular loose paths in a complete k-partite k-graph can be reformulated in terms of the question of which strings satisfying certain adjacency conditions can be produced from a fixed character set; the following lemma is the result we need.

Lemma 4.6 *Let ℓ and $a_1, \ldots, a_k$ be integers such that $0 \leqslant a_i < \ell/2$ for all i, and $\ell = \sum_{i=1}^{k} a_i$. Then for any $a, b \in [k]$ there exists a string of length ℓ on alphabet $x_1, \ldots, x_k$ such that the following properties hold:*

(1) no two consecutive characters are equal,

(2) the first character is not x_a and the final character is not x_b,

(3) the number of occurrences of character x_i is a_i.

Proof. Note that the conditions on ℓ and the a_i imply that $\ell \geqslant 3$. We construct the required string by starting with an 'empty string' of ℓ blank positions, and for each i inserting precisely a_i copies of character x_i. This ensures that condition (3) is satisfied. We fill the empty positions in the following order: first the first position, then the third, and so on through the odd-numbered positions, until we reach either position ℓ or position $\ell - 1$ (dependent on whether ℓ is odd or even). We then fill the second position, then the fourth, and so on until all positions are filled. Note that if we proceed by inserting all copies of one character, then all the copies of another character, and so forth, then condition (1) must be satisfied. This is because to get two consecutive copies of x_i, we must have inserted a copy of x_i at some odd position p, then $p+2, p+4$, and so on until reaching ℓ or $\ell-1$, and then filled even positions $2, 4, 6, \ldots, p-1$. However, this would imply that we had inserted at least $\ell/2$ copies of character x_i, contradicting the fact that $a_i < \ell/2$.

We therefore only need to determine an order to insert the different characters so as to satisfy (2). We first consider the case $a \neq b$, say $a = 1$ and $b = 2$. In this case we insert x_2 first, x_1 last, and the remaining character blocks in any order in between. Clearly this prevents the first character from being x_1 and the last from being x_2, and so (2) is satisfied. Now we may assume $a = b$, say $a = b = 1$. Then if ℓ is odd, we insert the characters in the following order: $x_2, x_3, \ldots, x_k, x_1$. Then all the copies of x_1 must be in even positions (since $a_1 < \ell/2$), and so (2) is satisfied. Alternatively, if ℓ is even, we insert first x_i for some $i \neq 1$ with $a_i > 0$, then x_1, and then the remaining blocks of characters in any order. (Note that these include at least one character since $\ell \geqslant 3$ and $a_i < \ell/2$ imply that at least three i have $a_i \geqslant 1$.) So neither the first nor last character can be x_1, and so (2) is again satisfied. $\square$

The next lemma is the result we were aiming for in this section, giving information about which loose paths can be found in complete k-partite k-graphs. Note that the maximum vertex degree of a loose path is two, and so this lemma tells us when we can find a loose path in a robustly universal complex.

Lemma 4.7 *Let G be a complete k-partite k-graph on vertex set $V_1 \cup \cdots \cup V_k$. Let $b_1, \ldots, b_k$ be integers with $0 \leqslant b_i \leqslant |V_i|$ for each i. Suppose that*

- $n := \frac{1}{k-1}((\sum_{i=1}^{k} b_i) - 1)$ *is an integer,*
- $\frac{n}{2} + 1 \leqslant b_i \leqslant n$ *for all i.*

Then for any $a, b \in [k]$, there exists a loose path in G with an initial vertex in V_a, a final vertex in V_b, and containing b_i vertices from V_i for each $i \in [k]$.

Proof. Note first that n is the number of edges such a path must contain. Let $a_i = n - b_i$ for each i, so that $0 \leqslant a_i < (n-1)/2$. By Lemma 4.6 we can find a string S of length $n-1$ on the alphabet $V_1, V_2, \ldots, V_k$ such that V_i appears a_i times, no two consecutive characters are identical, the first character is not V_a and the final character is not V_b. Let S_i be the ith character of S. To construct a loose path P in G, first choose any vertex from V_a to be the initial vertex of P, and any vertex from V_b to be the final vertex of P. We also use S to choose the link vertices of P: choose the ith link vertex (*i.e.* the vertex lying in the intersection of the ith and $(i+1)$th edges of P) to be any member of S_i not yet chosen. We have now assigned two vertices to each edge of P. Finally, we complete P by assigning to each edge one as yet unchosen vertex from each of the $k-2$ classes not yet represented in that edge. This is possible since precisely a_i link vertices are from the class V_i and so the total number of vertices used from V_i is $n - a_i = b_i$. Since G is complete we know that each edge of P is an edge of G, and so P is a loose path satisfying all the conditions of the lemma. $\square$

4.4.2 Walks and connectedness in k-graphs

A *walk* W in a hypergraph H consists of a sequence of edges $e_1, \dots, e_\ell$ of H and a sequence $x_0, \dots, x_\ell$ of (not necessarily distinct) vertices of H, satisfying $x_{i-1} \neq x_i$ for all $i \in [\ell]$, and also $x_0 \in e_1$, $x_\ell \in e_\ell$ and $x_i \in e_i \cap e_{i+1}$ for all $i \in [\ell - 1]$. The *length of* W is the number of its edges. We say that x_0 is the *initial vertex of* W, x_ℓ is the *final vertex of* W, and that $x_1, \dots, x_{\ell-1}$ are the *link vertices of* W. By a *walk from* x *to* y we mean a walk with initial vertex x and final vertex y.

Note that the vertices of a hypergraph H can be partitioned using the equivalence relation $\sim$, where $x \sim y$ if and only if either $x = y$ or there exists a walk from x to y.We call the equivalence classes of this relation *components* of H. We say that H is *connected* if it has precisely one component. Observe that all vertices of an edge of H must lie in the same component.Finally, note that if H is a connected hypergraph of order n, then for any two vertices x, y of H we can find in a walk from x to y of length at most n in H.

4.4.3 Random splitting

In this section we obtain, with high probability, a lower bound on the density of a subgraph of a k-partite k-graph chosen uniformly at random. We use Azuma's inequality on the deviation of a martingale from its mean.

Lemma 4.8 (Azuma [3]) *Suppose $Z_0, \dots, Z_m$ is a martingale,* i.e. *a sequence of random variables satisfying $\mathbb{E}(Z_{i+1} \mid Z_0, \dots, Z_i) = Z_i$, and that $|Z_i - Z_{i-1}| \leqslant c_i$ for some constants c_i and all $i \in [m]$. Then for any $t \geqslant 0$,*

$$\mathbb{P}(|Z_m - Z_0| \geqslant t) \leqslant 2 \exp\left(-\frac{t^2}{2\sum_{i=1}^m c_i^2}\right).$$

Lemma 4.9 *Suppose $1/n \ll c, \beta, 1/k, 1/b < 1$, and that H is a k-partite k-graph on vertex set $X = X_1 \cup \cdots \cup X_k$, where $n \leqslant |X_i| \leqslant bn$ for each $i \in [k]$. Suppose also that H has density $d(H) \geqslant c$ and that for each i we have $\beta|X_i| \leqslant t_i \leqslant |X_i|$. If we choose a subset $W_i \subseteq X_i$ with $|W_i| = t_i$ uniformly at random and independently for each i, and let $W = W_1 \cup \cdots \cup W_k$, then the probability that $H[W]$ has density $d(H[W]) > c/2$ is at least $1 - 1/n^2$. Moreover, the same holds if we choose W_i by including each vertex of X_i independently with probability $t_i/|X_i|$.*

Proof. Let $m = |X|$. To prove the first assertion, we obtain our subsets $W_i \subseteq X_i$ through the following two-stage random process, independently for each i. First we assign the vertices of each X_i into sets X_i^1 and X_i^2 independently at random, with each vertex being assigned to X_i^1 with probability $t_i/|X_i|$, and assigned to X_i^2 otherwise. Then, in the (highly probable) event that we have $|X_i^1| \neq t_i$ we select uniformly at random a set of vertices to transfer between X_i^1 and X_i^2 to obtain from X_i^1 the set W_i with $|W_i| = t_i$. For each i, no subset $W_i \subseteq X_i$ of size t_i is more likely to result from this process than any other, so we have chosen each W_i uniformly at random. It remains to show that $H[W]$ is likely to have high density. We do this by noting that $H[X^1]$ is likely to have high density (where $X^1 = X_1^1 \cup \cdots \cup X_k^1$) and that with high probability we only need to transfer a small number of vertices to form $W = W_1 \cup \cdots \cup W_k$, which can have only a limited effect on the density.

More precisely, let $x_1, \ldots, x_m$ be an ordering of the vertices of X, and for each $i \in [m]$ let the random variable Y_i take the value 1 if $x_i \in X^1$, and 0 otherwise. Recall that we write $|H|$ to denote the number of edges of a k-graph H. For all $i = 0, \ldots, m$ we now define random variables Z_i by $Z_i = \mathbb{E}(|H[X^1]| \mid Y_1, \ldots, Y_i)$. Then as we formed each X_i^1 by assigning vertices of X_i independently at random into X_i^1 and X_i^2, the sequence $Z_0, \ldots, Z_m$ is a martingale, and we have $Z_m = |H[X^1]|$ and $Z_0 \geqslant c \prod_{i=1}^k t_i$. Also, for any vertex x_i, let $f(i)$ be such that $x_i \in X_{f(i)}$ (*i.e.* $f(i)$ is the index of x_i). Then $|Z_i - Z_{i-1}| \leqslant \prod_{j \neq f(i)} |X_j| \leqslant (bn)^{k-1}$ for all

$i \in [m]$. Thus we can apply Lemma 4.8 to obtain

$$\mathbb{P}\left(|Z_m - Z_0| \geqslant \frac{c\prod_{i=1}^k t_i}{4}\right) \leqslant 2\exp\left(-\frac{c^2\prod_{i=1}^k t_i^2}{32mb^{2k-2}n^{2k-2}}\right) \leqslant \frac{1}{n^3}.$$

Therefore the event that $d(H[X^1]) > 3c/4$ has probability at least $1-1/n^3$. Also, by a standard Chernoff bound, for each $i \in [k]$ the event that $|X_i^1| = t_i \pm |X_i|^{2/3}$ has probability at least $1-1/n^3$. Thus with probability at least $1-1/n^2$ all of these events happen. Now, if $|X_i^1| > t_i$, we choose a set of $|X_i^1| - t_i$ vertices of X_i^1 uniformly at random and move these vertices from X_i^1 to X_i^2. Similarly, if $|X_i^1| < t_i$, then we choose a set of $t_i - |X_i^1|$ vertices of X_i^2 uniformly at random and move these vertices to X_i^1. In either case, for any i this action can decrease $d(H[X^1])$ by at most $||X_i^1| - t_i|/|X_i^1| \ll c$. Thus if we let W be the set obtained from X^1 in this way, we have $d(H[W]) > c/2$, proving the first part of the lemma.

The proof the 'moreover part' is the same except that we can omit the 'transfer' step at the end of the proof. □

4.4.4 Decomposition of G into copies of $\mathcal{A}_k$

Let $U_0, U_1, U_2, \ldots, U_{2k-3}$ be $2k-2$ pairwise disjoint sets of size $k-1$. Then we denote by $\mathcal{A}_k$ the k-graph whose vertex set $V(\mathcal{A}_k)$ is the disjoint union of the sets U_i, for $0 \leqslant i \leqslant 2k-3$, and whose edges consist of all k-tuples of the form $U_i \cup \{x\}$, with $i > 0$ and $x \in U_0$ (see Figure 4.1). So $|V(\mathcal{A}_k)| = 2(k-1)^2$. An $\mathcal{A}_k$*-packing* in a k-graph G is a collection of pairwise vertex-disjoint copies of $\mathcal{A}_k$ in G.

Lemma 4.10 *Suppose* $0 < 1/m \ll \theta \ll \psi \ll 1/k$, *and that* G *is a* k*-graph on* $[m]$ *such that* $|N_G(S)| > (\frac{1}{2(k-1)} + \theta)m$ *for all but at most* θm^{k-1} *sets* $S \in \binom{[m]}{k-1}$. *Then* G *has an* $\mathcal{A}_k$*-packing which covers more than* $(1-\psi)m$ *vertices of* G.

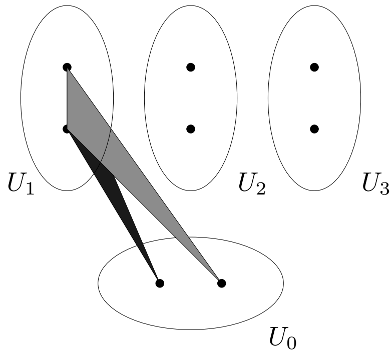

Figure 4.1: The 3-graph $\mathcal{A}_3$ (only edges involving U_1 are shown)

The k-graph $\mathcal{A}_k$ is identical to the k-graph $\mathcal{F}_{k,1}$ defined in Section 5.6, where a generalisation of Lemma 4.10 is proved (Lemma 5.17). Since the proof of Lemma 5.17 in no way depends on Lemma 4.10, we omit the proof of Lemma 4.10.

Corollary 4.11 *Lemma 4.10 still holds if we insist that the sub-k-graph of G induced by the vertices covered by the $\mathcal{A}_k$-packing must be connected.*

Proof. Apply Lemma 4.10 to obtain an $\mathcal{A}_k$-packing $A_1, \dots, A_\ell$ in G with $m_0 := |\bigcup_{i=1}^{\ell} V(A_i)| > (1-\psi/2)m$, and let A be the sub-k-graph of G induced by $\bigcup_{i=1}^{\ell} V(A_i)$. By hypothesis at most θm^{k-1} sets $S \in \binom{[m]}{k-1}$ have fewer than $m/(2(k-1))$ neighbours in G and so at most θm^{k-1} sets $T \in \binom{V(A)}{k-1}$ have no neighbours in $V(A)$. By the definition of a component, no edges of A contain vertices from different components of A. Therefore the largest component C of A must contain at least $(1-\psi)m$ vertices. Indeed, if not then there are $m_0^{k-2}(\psi m/2)/(k-1)! \gg \theta m^{k-1}$ sets $T \in \binom{V(A)}{k-1}$ which meet at least two components of A and thus have no neighbours in A, a contradiction (we can obtain such a set T by choosing $k-2$ vertices arbitrarily in $V(A)$ and then choosing the final vertex in a different component of A than the first vertex). Thus we may take the $\mathcal{A}_k$-packing consisting of all those copies A_i of $\mathcal{A}_k$ with $V(A_i) \subseteq V(C)$. $\square$

4.5 Proof of Theorem 1.10

Let H be as in Theorem 1.10. In our proof we use constants that satisfy the hierarchy

$$\frac{1}{n} \ll \varepsilon \ll d^* \ll d_a \ll \frac{1}{a} \ll \nu, \frac{1}{r} \ll \theta \ll d \ll c \ll \phi \ll \delta \ll \eta \ll \frac{1}{k}.$$

Furthermore, for any of these constants α, we use $\alpha \ll \alpha' \ll \alpha'' \ll \dots$ and assume that the above hierarchy also extends to the additional constants, *e.g.* $d'' \ll c \ll c'' \ll \phi$.

4.5.1 Imposing structure on H

Step 1. Applying the regularity lemma and forming the reduced graph

Let H_1 be the sub-k-graph obtained from H by removing up to $a!r$ vertices so that $|V(H_1)|$ is divisible by $a!r$. Let $T = T_1 \cup \dots \cup T_r$ be an equitable r-partition of the vertices of H_1, and let H_2 consist of all those edges of H_1 that are r-partite sets in T. Then H_2 is an r-partite k-graph with order divisible by $a!r$, and so we may apply the regularity lemma (Theorem 4.3), which yields an a-bounded ε-regular vertex-equitable partition $(k-1)$-complex P on T and an r-partite k-graph G on T that is ν-close to H_2 and perfectly ε-regular with respect to P.

Let $M = G \setminus H_2$. So any edge of $G \setminus M$ is also an edge of H. Let $V_1, \dots, V_m$ be the clusters of P. So $T = V_1 \cup \dots \cup V_m$ and G is m-partite with vertex classes $V_1 \cup \dots \cup V_m$. Note that $m \leqslant ar$ since P is a-bounded. Moreover, since P is vertex-equitable, each V_i has the same size. So let $n_1 = |V_i| = |T|/m$.

We next define a reduced k-graph R whose vertices correspond to the clusters V_i, whilst an edge e of R indicates that that within the cells of P corresponding to e we can find a subcomplex to which we can apply Theorem 4.5. For this we would like G to have high density in these cells, and M to have low density. Thus we define the *reduced k-graph* R on $[m]$ as follows: a k-tuple S of vertices of R corresponds to the k-partite union $S' = \bigcup_{i \in S} V_i$ of clusters. The edges of R are those $S \in \binom{[m]}{k}$ for which $G[S']$ has density at least c'' (*i.e.* $|G[S']| > c''|K_S(S')|$) and for which $M[S']$ has density at most $\nu^{1/2}$ (*i.e.* $|M[S']| < \nu^{1/2}|K_S(S')|$).

The edges in the reduced graph are useful in the following way. Given an edge $S \in R$, let $S' = \bigcup_{i \in S} V_i$ again. Using weak equivalence (defined in Section 4.3.2), the cells of P induce a partition $C^{S,1}, \ldots, C^{S,m_S}$ of the edges of $K_S(S')$. Recall that $m_S \leqslant a^k$. Therefore at most $c''|K_S(S')|/3$ edges of $K_S(S')$ can lie in sets $C^{S,i}$ with $|C^{S,i}| \leqslant c''|K_S(S')|/(3a^k)$. Furthermore, $|M[S']| < \nu^{1/2}|K_S(S')|$ (as $S \in R$) and so at most $\nu^{1/4}|K_S(S')|$ edges of $K_S(S')$ can lie in sets $C^{S,i}$ with $|M \cap C^{S,i}| \geqslant \nu^{1/4}|C^{S,i}|$. Together with the fact that $|G[S']| > c''|K_S(S')|$ this now implies that more than $c''|K_S(S')|/2$ edges of $G[S']$ lie in sets $C^{S,i}$ with $|C^{S,i}| > c''|K_S(S')|/(3a^k)$ and $|M \cap C^{S,i}| < \nu^{1/4}|C^{S,i}|$. Thus there must exist such a set $C^{S,i}$ that also satisfies $|G \cap C^{S,i}| > c''|C^{S,i}|/2$. Fix such a choice of $C^{S,i}$ and denote it by C^S. Let G^S be the k-partite k-complex on the vertex set S' consisting of $G \cap C^S$ and the cells of P that 'underlie' C^S, *i.e.* for any edge $Q \in G \cap C^S$ we have

$$G^S = (G[P])_{Q^\leqslant} = (G \cap C^S) \cup \bigcup_{Q' \subset Q} C_{Q'}. \tag{4.12}$$

(Recall that $(G[P])_{Q^\leqslant}$ and $C_{Q'}$ were defined in Section 4.3.2.) We also define the k-partite k-graph $M^S = G^S \cap M$ on vertex set S'. Then the following properties hold:

(A1) G^S is ε-regular.

(A2) G^S has k-th level relative density $d_{[k]}(G^S) \geqslant d'$.

(A3) G^S has absolute density $d(G^S) \geqslant d'_a$.

(A4) M^S satisfies $|M^S| < 2\nu^{1/4}|G^S|/c''$.

(A5) $(G^S)_{\{i\}} = V_i$ for any $i \in S$.

Indeed, (A1) follows from (4.12) since G is perfectly ε-regular with respect to P. To see (A2), note that $(G^S_{[k]})^* = C^S$ and so $d_{[k]}(G^S) = |G^S_{[k]}|/|(G^S_{[k]})^*| = |G^S \cap C^S|/|C^S| > c''/2$ by our choice of C^S. Similarly, (A3) follows from our choice of C^S since

$$d(G^S) = \frac{|G^S_{[k]}|}{|K_S(S')|} = \frac{|G^S \cap C^S|}{|C^S|} \cdot \frac{|C^S|}{|K_S(S')|} > \frac{(c'')^2}{6a^k} > d'_a.$$

(A4) holds since $|G^S| \geqslant |G \cap C^S| > c''|C^S|/2$ and $|M^S| \leqslant |M \cap C^S| < \nu^{1/4}|C^S|$. Finally, (A5) follows from (4.12) and the fact that $C_{\{v\}} = V_i$ for all $v \in V_i$.

Step 2. Choosing an $\mathcal{A}_k$-packing of R

The next step in our proof is to use Corollary 4.11 to find an $\mathcal{A}_k$-packing in the reduced k-graph R. For this we need an approximate minimum degree condition for R. Let

$$J = \left\{ I \in \binom{[m]}{k-1} : |N_R(I)| \leqslant \left(\frac{1}{2(k-1)} + \phi \right) m \right\}.$$

We next show that J is small, *i.e.* almost all $(k-1)$-tuples of vertices of R have degree at least $(1/(2(k-1)) + \phi)m$ in R. Consider how many edges of H do not belong to $G[S']$ for some edge $S \in R$. (Recall that $S' = \bigcup_{i \in S} V_i$.) There are three possible reasons why an edge $e \in H$ does not belong to such a restriction:

(i) e is not an edge of G. This could be because e lies in H but not H_1, in H_1 but not H_2, or in H_2 but not G. There are at most $a!rn^{k-1}$ edges of the first type, at most n^k/r of the second type, and at most νn^k of the third type.

(ii) $e \in G$ contains vertices from $V_{i_1}, \ldots, V_{i_k}$ such that the restriction of M to $S' = \bigcup_{i \in S} V_i$ satisfies $|M[S']| \geqslant \nu^{1/2}|K_S[S']|$, where $S = \{i_1, \ldots, i_k\}$. (Note that since G and thus M is m-partite, we must have $i_1 < \cdots < i_k$.) Since G and H_2 are ν-close and thus $|M| \leqslant \nu n^k$ there are at most $\nu^{1/2}n^k$ edges of this type.

(iii) $e \in G$ contains vertices from $V_{i_1}, \ldots, V_{i_k}$ such that the restriction of G to $\bigcup_{i \in S} V_i$ has density less than c''. There are at most $c''n^k$ edges of this type.

Therefore there are fewer than $2c''n^k$ edges of H that do not belong to the restriction of G to S' for some $S \in R$, and so we have

$$\begin{aligned}|J|n_1^{k-1} \cdot \left(\frac{1}{2(k-1)} + \eta\right) n &< \sum_{I \in J} \sum_{x_i \in V_i, i \in I} |N_H(\{x_i\}_{i \in I})| \\ &< 2c''n^k + \sum_{I \in J} |N_R(I)| n_1^k < 2c''n^k + |J| \left(\frac{1}{2(k-1)} + \phi\right) mn_1^k.\end{aligned}$$

Since $n - a!r \leqslant mn_1 \leqslant n$ we deduce $|J|n_1^{k-1}(\eta - \phi)n < 2c''n^k < 3c''(mn_1)^{k-1}n$, and so $|J| < \phi m^{k-1}$ (since $c'' \ll \phi \ll \eta$). This allows us to apply Corollary 4.11 (with $G = R$) to obtain an $\mathcal{A}_k$-packing $A_1, \ldots, A_t$ in R with $|\bigcup_{i=1}^t A_i| > (1-\delta)m$, such that the sub-$k$-graph of R induced by $\bigcup_{i=1}^t A_i$ is connected. For $i \in [t]$, let the vertex set of A_i be $U_0^i \cup U_1^i \cup \cdots \cup U_{2k-3}^i$, with each U_j^i of size $k-1$, so that the edge set is $\{U_j^i \cup \{x\} : j \in [2k-3], x \in U_0^i\}$.

Step 3. Forming the exceptional path

Given a sub-k-graph R' of R and a cluster V_i, we say that V_i *belongs to* R' if $i \in V(R')$. Let V_0' contain the at most $a!r$ vertices of H we removed at the start of the proof, and also the vertices in all those clusters not belonging to some copy of $\mathcal{A}_k$ in our packing (there are at most δn of the latter). We incorporate these vertices into a path L_e which is to form part of our loose Hamilton cycle. We also include in V_0' an arbitrary choice of δn_1 vertices from each V_y for which $y \in U_j^i$

for some $j \in [2k-3]$ and some $i \in [t]$ (we do not modify any of the V_y for which $y \in U_0^i$). We add up to $k-3$ more vertices from U_1^1 (say) to V_0' so that$|V_0'| \equiv 0 \mod k-2$. We delete all these vertices from the clusters they belonged to and still write V_y for the subcluster of a cluster V_y obtained in this way. This gives $|V_0'| \leqslant 5\delta n/2$.

We now construct a path L_e in H which contains all the vertices in V_0' and avoid all the clusters V_y with $y \in U_0^i$. Let $V_{>0} = \bigcup\{V_y : y \in U_j^i, j \in [2k-3], i \in [t]\}$. So we use only vertices from V_0' and $V_{>0}$ in forming L_e. Recall that if $|V(H)|$ is not a multiple of $k-1$, then a loose Hamilton cycle contains a single pair of edges which intersect in more than one vertex: we make allowance for this here. Choose $A, B \subseteq V_{>0}$ satisfying $|A| = |B| = k-1$, $|A \cap B| \equiv 1 - |V(H)| \mod k-1$ and $1 \leqslant |A \cap B| \leqslant k-1$. Now choose distinct $x_0, x_1 \in V_{>0} \setminus (A \cup B)$ such that $\{x_0\} \cup A \in H$ and $\{x_1\} \cup B \in H$ (we show in a moment that such x_0, x_1 exist). These edges form the first 2 edges of L_e. To complete L_e, let $Z_1, \dots, Z_s$ be any partition of the vertices of V_0' into sets of size $k-2$. For each $i = 1, \dots, s$ we proceed greedily in forming L_e: choose any $x_{i+1} \in V_{>0} \setminus (A \cup B)$ such that $Z_i \cup \{x_i, x_{i+1}\} \in H$ (where the x_i are all chosen to be distinct).

Let us now check that there is always such a vertex available. Indeed, every set in $\binom{V(H)}{k-1}$ has at least $(1/(2(k-1)) + \eta)n$ neighbours and we can choose any such neighbour which lies in $V_{>0}$ and has not already been used. But $|V(H) \setminus V_{>0}| \leqslant n/(2(k-1)) + |V_0'|$ and at most $|V_0'| + 2k \leqslant 3\delta n$ vertices have been used before. Thus (since $\delta \ll \eta$) for each choice of an x_i we have at least $\eta n/2$ vertices of $V_{>0}$ to choose from. Moreover, these vertices must be contained in at least $\eta n/(2n_1)$ different V_y such that $y \in U_j^{i'}$ ($j > 0$). Thus we can avoid choosing a vertex from any single V_y more than $6\delta n_1/\eta \leqslant \delta' n_1/2$ times. The path L_e thus formed has edges $\{x_0\} \cup A$, $B \cup \{x_1\}$ and $\{x_i, x_{i+1}\} \cup Z_i$ for all $i \in [s]$. So all the vertices of V_0' are included in L_e. For each cluster V_y, we still denote the subset of V_y lying in $V(H - L_e)$ by V_y. Then each V_y with $y \in U_0^i$ for some i still satisfies $|V_y| = n_1$, and each V_y with $y \in U_j^i$ for some

$j > 0$ satisfies

$$(1-\delta')n_1 \leqslant (1-\delta-\frac{\delta'}{2})n_1-(k-3) \leqslant |V_y| \leqslant (1-\delta)n_1. \tag{4.13}$$

In addition

$$|V(H)\setminus V(L_e)| \equiv |V(H)|-|A\cup B\cup\{x_0,x_1\}| \equiv -1 \mod k-1. \tag{4.14}$$

Note that L_e need not be a loose path, but that even if it is not it may still form part of a loose Hamilton cycle. Also observe that $|V(L_e)| \leqslant 6\delta n$.

Step 4. Splitting our copies of $\mathcal{A}_k$

The next step of the proof is to split the copies $A_1,\dots,A_t$ of $\mathcal{A}_k$ (more precisely the clusters belonging to the A_i) into sub-k-complexes of G which we later use to embed spanning loose paths. Consider any A_i. For convenient notation we identify each U^i_j in A_i with $[k-1]$ (but recall that they are disjoint sets). For each $y \in U^i_0 = [k-1]$ we have $|V_y| = n_1$, and so we can partition $V_y \setminus V(L_e)$ uniformly at random into $2k-3$ pairwise disjoint subsets $S^i_{y,1},\dots,S^i_{y,2k-3}$, each of size $\frac{n_1}{2k-3}$. Similarly, given $z \in U^i_j = [k-1]$ with $j \in [2k-3]$, (4.13) and the fact that $\delta' \ll \eta$ imply that we can partition V_z uniformly at random into $k-1$ pairwise disjoint subsets $T^i_{j,z}$ and $\{U^i_{j,z,w}\}_{w\in[k-1]\setminus\{z\}}$ so that $\frac{n_1}{2k-3} \leqslant |T^i_{j,z}| \leqslant \frac{(1-\eta)2n_1}{2k-3}$ and $|U^i_{j,z,w}| = \frac{(1-\eta)2n_1}{2k-3}$ for all $w \in [k-1]\setminus\{z\}$. Figure 4.2 shows how we do this in the case $k=3$.

We arrange these pieces into $(k-1)(2k-3)$ collections of k sets as follows: for each $y \in U^i_0$ and each $j \in [2k-3]$ we have a collection consisting of $S^i_{y,j}$, $T^i_{j,y}$ and $\{U^i_{j,z,y}\}_{z\neq y}$. (3 of these collections are illustrated in Figure 4.2.) For convenient notation we relabel these collections as

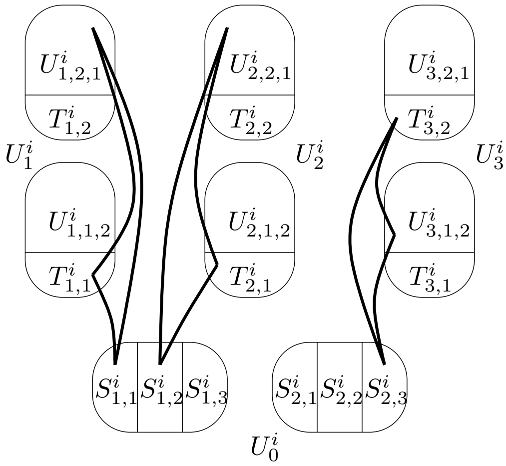

Figure 4.2: Splitting up A_i in the case $k = 3$.

$\{X_{i,1}, \dots, X_{i,k}\}$ with $1 \leqslant i \leqslant t' = (k-1)(2k-3)t$, where for all $i \in [t']$

$$|X_{i,1}| = \frac{n_1}{2k-3},\ \frac{n_1}{2k-3} \leqslant |X_{i,2}| \leqslant \frac{(1-\eta)2n_1}{2k-3} \text{ and } |X_{i,j}| = \frac{(1-\eta)2n_1}{2k-3} \text{ for } 3 \leqslant j \leqslant k, \tag{4.15}$$

$$(1-\delta')n_1 \leqslant \sum_{j=2}^{k} |X_{i,j}| \leqslant (1-\delta)n_1. \tag{4.16}$$

((4.16) follows from (4.13) using the fact that all the $U^{i'}_{j',z,w}$ have equal size.) Let $X_i = \bigcup_{j\in[k]} X_{i,j}$, so each X_i is a k-partite set, on which we now find a sub-k-complex G_i of G that is suitable for applying Theorem 4.5.

Consider any copy $A_{i'}$ in our $\mathcal{A}_k$-packing. Note that for each of the $(k-1)(2k-3)$ collections $\{X_{i,1}, \dots, X_{i,k}\}$ obtained by splitting up the clusters belonging to $A_{i'}$ there is an edge $S(i) \in A_{i'}$ such that each $X_{i,j}$ lies in a cluster belonging to $S(i)$ (and these clusters are distinct for each of $X_{i,1}, \dots, X_{i,k}$). Recall that $S'(i)$ denotes the union $\bigcup_{\ell\in S(i)} V_\ell$ of all the clusters belonging to $S(i)$. Let G_i denote the restriction of the k-partite k-complex $G^{S(i)}$ (which was defined in Section 4.5.1) to X_i, *i.e.* $G_i = G^{S(i)}[X_i]$. Let $M_i = M \cap G_i = M^{S(i)}[X_i]$. We claim that we may choose the above collections $\{X_{i,1}, \ldots, X_{i,k}\}$ such that

$$d(H[X_i]) \geqslant \frac{c''}{4} \text{ for all } i \in [t']. \tag{4.17}$$

Indeed, since $S(i) \in R$, $G[S'(i)]$ has absolute density at least c'' and $M[S'(i)]$ has density at most $\nu^{1/2}$. Since $G \setminus M \subseteq H$ and $\nu \ll c''$ this shows that $H[S'(i)]$ has density at least $c''/2$. Lemma 4.9 now implies that each $H[X_i]$ has density at least $c''/4$ with probability $1 - 1/n_1^2$, and so with non-zero probability this is true for all $i \in [t']$.

Lemma 4.4 and properties (A1)–(A3) and (A5) imply that G_i is an ε'-regular k-partite k-complex on vertex set X_i, with absolute density $d(G_i) \geqslant d(G^{S(i)})/2 \geqslant d_a$, relative density $d_{[k]}(G_i) \geqslant d$, and $(G_i)_{\{j\}} = X_{i,j}$ for each j. Moreover, using $\nu \ll \theta \ll c$, property (A4) and the fact that $d(G_i) \geqslant d(G^{S(i)})/2$ we see that

$$|M_i| \leqslant |M^{S(i)}| < \frac{2\nu^{1/4}|G^{S(i)}|}{c''} \leqslant \theta|G_i|.$$

So by Theorem 4.5, we can delete at most $\theta'|X_{i,j}|$ vertices from each $X_{i,j}$ so that if we let $X'_{i,j} \subseteq X_{i,j}$ consist of the undeleted vertices, and let $X'_i := \bigcup_{j=1}^{k} X'_{i,j}$, $G'_i := G_i[X'_i]$ and $M'_i := M_i[X'_i]$, then $G'_i \setminus M'_i$ is (c, ε'')-robustly 2^k-universal. Let X'' denote the set of vertices deleted from any $X_{i,j}$, so $|X''| \leqslant \theta' n$. We may assume that $|X''|$ is divisible by $k-2$. The latter helps us to extend L_e into a path which contains all the vertices in X''.

Step 5. Extending the exceptional path L_e

When extending L_e in order to incorporate X'', we have to remove some more vertices from some of the $X'_{i,j}$, and we wish to do this so that the remainder still satisfies (i) in the definition of robust universality. For this reason, we partition each $X'_{i,j}$ into two parts $AX'_{i,j}$ and $BX'_{i,j}$ as follows (where we write BX'_i for $\bigcup_{j\in[k]} BX'_{i,j}$):

(B1) For all i, j and every $v \in X'_{i,j}$ we have $|(G'_i(v)[BX'_i])_=| \geqslant 2c|G'_i(v)_=|$.

(B2) Every set of $k-1$ vertices of H has at least $n/(4k)$ neighbours in $\bigcup_{i,j} AX'_{i,j}$.

(Recall that for a $(k-1)$-complex F, $F_=$ denotes the '$(k-1)$th level' of F.) To see that such a partition exists, consider a partition obtained by assigning each vertex to a part with probability $1/2$ independently of all other vertices. (B2) is then satisfied with high probability by a standard Chernoff bound. Now consider (B1). The 'moreover' part of Lemma 4.9 implies that with high probability we have for all i and for all $v \in X'_{i,j}$ that $d((G'_i(v)[BX'_i])) \geqslant d(G'_i(v))/2$. Also, a standard Chernoff bound implies that with high probability $|BX'_{i,j'}| \geqslant |X'_{i,j'}|/3$ for all $j' \in [k]$. Thus

$$|(G'_i(v)[BX'_i])_=| = d((G'_i(v)[BX'_i])) \prod_{j' \neq j} |BX'_{i,j'}| \geqslant \frac{d(G'_i(v))}{2} \prod_{j' \neq j} \frac{|X'_{i,j'}|}{3} \geqslant 2c|G'_i(v)_=|.$$

Now, we extend our path L_e to include the vertices in X'', using only vertices from $\bigcup_{i,j} AX'_{i,j}$. We proceed similarly to when constructing L_e. So we split X'' into sets $Z_1, ..., Z_{s'}$ of size $k-2$ (so $s' \leqslant \theta' n$). Letting x_0 be a final vertex of L_e, for $i \in [s']$, we successively choose x_i to be a neighbour of the $(k-1)$-tuple $Z_i \cup \{x_{i-1}\}$ contained in some $AX'_{i',j'}$ and not already included in L_e, and extend L_e by the edge $Z_i \cup \{x_{i-1}, x_i\}$, continuing to denote the extended path by L_e. Recall that L_e originally contained at most $6\delta n$ vertices. Since $|X''| \leqslant \theta' n$, after each extension of L_e we have $|V(L_e)| < \eta n$. So (B2) implies that for each choice of x_i we have at least $n/(5k)$ suitable vertices and hence at least $t'/(5k)$ of the sets $AX'_{i'}$ contain such a suitable vertex. This shows that we can choose the x_i in such a way that at most $\theta'' n_1$ vertices are chosen from any single $AX'_{i'}$.

For each $i \in [t']$ let $X^i = X^i_1 \cup \cdots \cup X^i_k$ be the vertices remaining after the removal from X'_i of the at most $\theta'' n_1$ vertices used in extending L_e. Let $G^i := G'_i[X^i]$ and $M^i := M'_i[X^i]$. Also, let x_e be an initial vertex of L_e and let y_e be a final vertex of L_e. By adding 2 more edges to L_e if necessary (one at the start and one at the end), we may assume that x_e lies in at least $|H[X^i]|/(2|X^i|)$ edges of $H[X^i]$, where i is such that $x_e \in X^i$ and that the analogue holds for the final vertex y_e of L_e.

We claim that the above steps give us the following useful structure: a path L_e which is ready to form part of a loose Hamilton cycle, disjoint k-partite vertex sets $X^i = X^i_1 \cup \cdots \cup X^i_k$, and G^i and M^i for each $i \in [t']$ satisfying the following properties:

(C1) Every vertex of H lies in either the path L_e or precisely one of the k-partite sets X^i.

(C2) For each i, G^i is a k-partite sub-k-complex of G on the vertex set X^i. M^i is the k-partite k-graph $M \cap G^i$, and $G^i \setminus M^i \subseteq H$. Clearly these statements remain true after the deletion of up to εn_1 vertices of X^i.

(C3) Even after the deletion of up to εn_1 vertices of X^i, the following statement holds. Let L be a k-partite k-complex on vertex set $U = U_1 \cup \cdots \cup U_k$, where $|U_j| = |X^i_j|$ for each j, and let L have maximum vertex degree at most 2^k. Let $\ell \leqslant 2(t')^2$ and suppose we have $u_1, \ldots, u_\ell \in U$ and sets $Z_s \subseteq X^i_{i(u_s)}$ with $|Z_s| \geqslant c|X^i_{i(u_s)}|$ for $s \in [\ell]$. Then $G^i \setminus M^i$ contains a copy of L, in which the vertices of U_j correspond to the vertices of X^i_j, and u_s corresponds to a vertex in Z_s.

(C4) For each i, $H^i = H[X^i]$ has density at least c', even after the deletion of up to εn_1 vertices of X^i.

(C5) If we delete up to εn_1 vertices from any X^i, and let $t_j = |X^i_j|$ for each $j \in [k]$ after this removal, and let $n'_i = \frac{(\sum t_j)-1}{k-1}$, then $n'_i/2 + 1 \leqslant t_j \leqslant n'_i$ for all j.

(C6) The initial vertex x_e of L_e lies in at least $|H[X^i]|/(2|X^i|)$ edges of $H[X^i]$, where i is such that $x_e \in X^i$. The analogue holds for the final vertex y_e of L_e.

(When we talk of removing a vertex of X^i we implicitly mean that G^i, M^i and H^i are all restricted to the remaining vertices of X^i.) These properties hold for the following reasons. (C1) holds as every vertex deleted from an X_i has been added to L_e, whilst (C2) is clear as whenever we deleted vertices we simply restricted G and M to the remaining vertices. For (C3), recall that $G'_i \setminus M'_i$ was (c, ε'')-robustly 2^k-universal. Moreover, for all $i \in [t']$ and all $j \in [k]$ we have $|X^i_j| \geqslant |X'_{i,j}|/2 \geqslant c|X'_{i,j}|$, since we ensured that we only deleted $\theta'' n_1$ vertices from any single AX'_i (and none from BX'_i). Furthermore by (B1) we know that $|G^i(v)_=| \geqslant |(G'_i(v)[BX'_i])_=| \geqslant c|G'_i(v)_=|$ for any $v \in X^i$. (Also, even if we had arbitrarily deleted a further εn_1 vertices from X'_i when obtaining X^i, G^i and M^i, these bounds would still hold.) So $G^i \setminus M^i$ satisfies (i) in the definition of a robustly universal complex (with X^i_j in place of V_j). The sets Z_s satisfy (iii) in the definition and so we can find the required copy of L (even after the deletion of up to εn_1 more vertices of X^i). (C4) follows from (4.17) and the fact X^i was formed by deleting at most $(\theta' + \theta'')n_1 \ll c'|X_i|$ vertices from X_i. Similarly, for (C5) note that (even after up to εn_1 more deletions) we have deleted at most $2\theta'' n_1$ vertices from each X_i since we split the clusters to form the X_i. Therefore, from (4.15) and (4.16) and the fact that $\theta'' \ll \delta \ll \delta' \ll \delta'' \ll \eta$, we have (even after deletions) that

- $\frac{n_1}{2k-3} - 2\theta'' n_1 \leqslant |X^i_1| \leqslant \frac{n_1}{2k-3}$;
- $\frac{n_1}{2k-3} - 2\theta'' n_1 \leqslant |X^i_2| \leqslant \frac{(1-\eta)2n_1}{2k-3}$;
- $\frac{(1-\eta)2n_1}{2k-3} - 2\theta'' n_1 \leqslant |X^i_j| \leqslant \frac{(1-\eta)2n_1}{2k-3}$ for $3 \leqslant j \leqslant k$;
- $n'_i \geqslant \frac{1}{k-1}\left(n_1\left(1 - \delta' + \frac{1}{2k-3} - 2k\theta''\right) - 1\right) \geqslant \frac{(1-\eta)2n_1}{2k-3}$;
- $n'_i \leqslant \frac{n_1}{k-1}\left(1 - \delta + \frac{1}{2k-3}\right) \leqslant \frac{(2-\delta)n_1}{2k-3}$.

So property (C5) follows. Finally, (C6) follows by the choice of L_e.

4.5.2 The supplementary graph

Roughly speaking, our aim is to find a spanning loose path in each $G^i \setminus M^i$ (and thus in H^i) such that all these paths together with L_e form a loose Hamilton cycle of H. So we have to ensure that the complete k-partite k-graph on X^i contains a spanning loose path (for this, we need $|X^i| \equiv 1 \mod k-1$) and we need to join up all the loose paths we find in the H^i. The purpose of this section is to find the 'connecting loose paths' which join up the X^i in such a way that the divisibility problems are dealt with as well. To do this, we first define a supplementary hypergraph R^* whose vertices correspond to the X^i. We show that R^* is connected and that 'along' edges of R^* we can find our loose paths in H which join up all the X^i.

The vertex set of the *supplementary hypergraph* R^* is $[t']$. A subset $e \subseteq [t']$ of size at least 2 is an edge of R^* if there exists an edge $S_e \in R$ such that for all $j \in S_e$ there are $i_j \in e$ and $\ell_j \in [k]$ with $X^{i_j}_{\ell_j} \subseteq V_j$ and $e = \{i_j : j \in S_e\}$. (We fix one such edge S_e for every $e \in R^*$.) Then every edge of R^* has size at most k. We say that X^i *belongs* to an edge $e \in R^*$ if $i \in e$. Similarly, X^i *belongs to* some subhypergraph $R' \subseteq R^*$ if $i \in V(R')$.

Lemma 4.18 *The supplementary graph R^* is connected.*

Proof. Recall that we chose the copies A_ℓ of $\mathcal{A}_k$ in such a way that the sub-k-graph A of R induced by $\bigcup_{\ell=1}^t A_\ell$ is connected. Suppose that R^* is not connected. Let R^*_1 be a component of R^* and let $R^*_2 = R^* - R^*_1$. Let $R_1 = \{j \in [m] : X^i_s \subseteq V_j$ for some $i \in V(R^*_1), s \in [k]\}$. So R_1 corresponds to the set of all those clusters which meet some X^i belonging to R^*_1. Define R_2 similarly.Then $R_1 \cup R_2 = V(A)$ and thus A contains some edge S intersecting both R_1 and R_2. But then S corresponds to an edge of R^* intersecting both $V(R^*_1)$ and $V(R^*_2)$, a contradiction.
$\square$

The next lemma shows that within the X^i belonging to an edge of R^*, we can find a reasonably short loose path in H and we may choose (modulo $k-1$) how many vertices this path uses from each X^i. Using the connectedness of R^*, this allows us to find the connecting loose paths which join up the X^i, having control over the divisibility properties. We also insist that the path in Lemma 4.19 avoids a number of 'forbidden vertices', to enable us to ensure that our connecting loose paths are disjoint, and that the endvertices of these paths lie in many edges of the relevant H^i.

Lemma 4.19 *Suppose that $e \in R^*$ and that for every $i \in e$ there is an integer t_i such that $0 \leqslant t_i \leqslant k-1$ and $\sum_{i\in e} t_i \equiv 1 \mod k-1$. Let $i', i'' \in e$ be distinct. Moreover, suppose that Z is a set of at most $100(t')^2k^3$ 'forbidden' vertices of H. Then in the sub-k-graph of H induced by $\bigcup_{i\in e} X^i$ we can find a loose path L with the following properties.*

- *L contains at most $4k^3$ vertices.*
- *L has an initial vertex u in $X^{i'}$ and a final vertex v in $X^{i''}$.*
- *$|V(L) \cap X^i| \equiv t_i \mod k-1$ for each $i \in e$.*
- *L contains no forbidden vertices,* i.e. $V(L) \cap Z = \emptyset$.
- *u lies in at least $|H^{i'}|/(2|X^{i'}|)$ edges of $H^{i'}$, and v lies in at least $|H^{i''}|/(2|X^{i''}|)$ edges of $H^{i''}$.*

Proof. Recall that in Section 4.5.1 we assigned a k-partite k-complex G^S to every edge $S \in R$ such that (A1)–(A5) are satisfied. To simplify notation, we write S for the edge $S_e \in R$ corresponding to e and suppose that $S = [k]$. For each $j \in S = [k]$ choose $i_j \in e$ and $\ell_j \in [k]$ such that $X^{i_j}_{\ell_j} \subseteq V_j$ and such that $e = \{i_j : j \in S = [k]\}$. To simplify notation we write Y_j for $X^{i_j}_{\ell_j} \setminus Z$, $Y = \bigcup_{j\in[k]} Y_j$ and assume that $i' = i_1$ and $i'' = i_k$. For each $i \in e$ let J_i be the set of all $j \in S = [k]$ with $i_j = i$. So the sets J_i are disjoint and their union is $[k]$. Pick some

$j \in J_i$ and let $t'_j = t_i$ and $t'_s = 0$ for all $s \in J_i \setminus \{j\}$. Our path L is to consist of t'_j vertices from each Y_j (modulo $k-1$) and thus of t_i vertices from each X^i (modulo $k-1$).

Since G^S satisfies (A1)–(A3) and (A5), Lemma 4.4 implies that the restriction of $G^S[Y]$ is ε'-regular, with absolute density at least $d(G^S)/2 \geqslant d_a$, relative density at index $[k]$ at least d and $(G^S)_{\{j\}}[Y] = Y_j$. Furthermore, (A4) together with the fact that $d(G^S[Y]) \geqslant d(G^S)/2$ imply that

$$|M^S[Y]| < |M^S| < \frac{2\nu^{1/4}|G^S|}{c''} \leqslant \theta|G^S[Y]|.$$

Thus Theorem 4.5 implies that we can delete $\theta'|Y_j|$ vertices from each Y_j to obtain a subset Y'_j such that the complex $G^S[Y'] \setminus M^S[Y']$ is (c, ε'')-robustly 2^k-universal, where $Y' = \bigcup_{j\in[k]} Y'_j$.

Now, let $v_j = (k+2)(k-1) + t'_j$. Then $\sum v_j \equiv 1 \mod k-1$ and so $n' = ((\sum v_i) - 1)/(k-1)$ is an integer. Furthermore, $k(k+2) \leqslant n' \leqslant k(k+3)$, and so $n'/2 + 1 \leqslant v_j \leqslant n'$ for each j. Thus by Lemma 4.7 we can find a loose path in the complete k-partite k-graph on vertex set Y', beginning in Y'_1, finishing in Y'_k and using v_j vertices from each Y'_j. Since $G^S[Y'] \setminus M^S[Y']$ is robustly 2^k-universal, we can find such a loose path L in $G^S[Y'] \setminus M$ and hence in $H - Z$. (Indeed, we can do this by finding the complex $L^{\leqslant}$, which has maximum vertex degree at most 2^k. Note that we use the definition with $G = G'$ in (i)). Note that L contains at most $k(k-1)(k+3) \leqslant 4k^3$ vertices.

To see that we can insist on the final condition of the lemma, recall that $d(H^i) \geqslant c'$ by (C4). Thus for all $j \in [k]$ at least $c'|X^i_j|/2$ vertices of X^i_j lie in at least $|H^i|/(2|X^i_j|)$ edges of H^i, and so we may restrict the initial and final vertices of L to these sets of vertices (minus the vertices of Z) by (iii) in the definition of robust universality. $\square$

4.5.3 Constructing the loose Hamilton cycle

Our Hamilton cycle of H is to consist of L_e and paths in each H^i as well as paths connecting the X^i. However, we need to make sure that all these paths join up nicely, motivating the following definition. Suppose L is a path in some k-graph K with initial vertex x' and final vertex y'. Also, let $I, F \subseteq V(K) \setminus V(L)$ be disjoint sets of size $k-2$. Then $L^* = I \cup F \cup V(L)$ is a *prepath*. Note that L^* is not (the vertex set of) a k-graph, but that if we can find vertices $x, y \in V(K) \setminus L^*$ such that $\{x, x'\} \cup I, \{y, y'\} \cup F \in K$, then adding x and y to L^* gives another path. We refer to all such vertices $x \in V(K)$ as *possible initial vertices* of L^* and to all such vertices $y \in V(K)$ as *possible final vertices*. If L, L' and L'' are disjoint loose paths, I, F, x, y are as before, x is also the final vertex of L' and y is also the initial vertex of L'' then I and F together with L', L, L'' form a single loose path, illustrating how we join paths together.

We start by converting our exceptional path L_e into a prepath. Recall that $|V(L_e)| < \eta n$ and that the initial vertex x_e of L_e and its final vertex y_e satisfy (C6). Let $a \in [t']$ and $u_a \in [k]$ be such that $x_e \in X^a_{u_a}$. Pick any $u'_a \in [k]$ with $u_a \neq u'_a$. (C4) and (C6) together imply that there is a set $I_0 \subseteq X^a \setminus (X^a_{u_a} \cup X^a_{u'_a})$ for which $X^a_{u'_a}$ contains at least $c|X^a|$ vertices v which form an edge of H^a together with $I_0 \cup \{x_e\}$. Let $I'_0 \subseteq X^a_{u'_a}$ be such a set of vertices. Similarly, letting $b \in [t']$, $u_b \neq u'_b \in [k]$ be such that $y_e \in X^b_{u_b}$, there is a set $F_0 \subseteq X^b \setminus (X^b_{u_b} \cup X^b_{u'_b} \cup I_0)$ for which $X^b_{u'_b}$ contains at least $c|X^b|$ vertices v which form an edge of H^b together with $F_0 \cup \{y_e\}$. Let $F'_0 \subseteq X^b_{u'_b}$ be such a set of vertices. Let L^*_e be the prepath $I_0 \cup F_0 \cup V(L_e)$. Then I'_0 is a set of possible initial vertices of L^*_e and F'_0 is a set of possible final vertices. (We do not remove I_0 from X^a and F_0 from X^b at this stage.)

Since by Lemma 4.18 the supplementary graph R^* is connected, we can find a walk W from b to a in R^* such that every $i \in [t'] = V(R^*)$ appears as an initial, link or final vertex in W (these vertices were defined in Section 4.4.2) and such that W has length $\ell \leqslant (t')^2$. Let $e_1, \ldots, e_\ell$ be the edges of this walk, let $r_1 = b, r_{\ell+1} = a$, and let $r_2, \dots, r_\ell$ be the link vertices of the walk. For each $i \in [t']$, let $d_i = |\{j \in [\ell+1] : r_j = i\}|$, *i.e.* the number of times i appears as an initial, link or final vertex in W. So $d_i > 0$ for every i and $\sum d_i = \ell + 1$.

Next we apply Lemma 4.19 to each edge e_j to find a loose path L_j in H. We then extend L_j to a prepath L_j^* with many possible initial vertices in X^{r_j} and many possible final vertices in $X^{r_{j+1}}$. We do this for each $e_1, \dots, e_\ell$ in turn. So suppose that $s \in [\ell]$ and that for all $j = 1, \dots, s-1$ we have defined loose paths L_j in H as well as sets I_j, F_j extending L_j to a prepath L_j^* which satisfy the following properties:

(D1) L_j lies in the sub-k-graph of H induced by $\bigcup_{i \in e_j} X^i$ and contains at most $4k^3$ vertices.

(D2) The initial vertex x_j of L_j lies in X^{r_j} and its final vertex y_j lies in $X^{r_{j+1}}$.

(D3) $I_j \subseteq X^{r_j}$ and $F_j \subseteq X^{r_{j+1}}$.

(D4) There is a set $I'_j \subseteq X^{r_j}$ of at least $c|X^{r_j}|$ possible initial vertices for L_j^*. Similarly, there is a set $F'_j \subseteq X^{r_{j+1}}$ of at least $c|X^{r_{j+1}}|$ possible final vertices for L_j^*.

(D5) All the prepaths $L_e^*, L_1^*, \dots, L_{s-1}^*$ are disjoint.

(D6) For each $i \in [t']$ and all $j = 0, \dots, s-1$ let $X^i(j) = X^i \setminus (V(L_1) \cup \dots \cup V(L_j))$, where $X^i(0) = X^i$. For each $j \in [s-1]$ set $t_i(j) = |X^i(j-1)| + d_i$. Then for every $i \in e_j$ with $i \neq r_{j+1}$ we have $|V(L_j) \cap X^i| \equiv t_i(j) \mod k-1$. Moreover $|V(L_j) \cap X^{r_{j+1}}| \equiv 1 - \sum_{i \in e_j,\, i \neq r_{j+1}} t_i(j) \mod k-1$.

Let us now show how to find L_s, I_s and F_s. Apply Lemma 4.19 with $e = e_s$, $i' = r_s$, $i'' = r_{s+1}$ and with $Z = L_1^* \cup \dots L_{s-1}^* \cup I_0 \cup F_0$ to find a loose path L_s which satisfies (D1), (D2), (D6) and is disjoint from $L_e^*, L_1^*, \dots, L_{s-1}^*$. Moreover, the initial vertex x_s of L_s lies in at least $|H^{r_s}|/(2|X^{r_s}|)$ edges of H^{r_s}, and the final vertex y_s of L_s lies in at least $|H^{r_{s+1}}|/(2|X^{r_{s+1}}|)$

edges of $H^{r_{s+1}}$. We can now use the latter property to choose sets I_s and F_s which extend L_s to a prepath L^*_s satisfying (D3)–(D5). The argument for this is similar to that for the extension of L_e to L^*_e. Altogether this shows that we can find prepaths $L^*_1, \dots, L^*_\ell$ satisfying (D1)–(D6).

For each $i \in [t']$ we let j_i be the maximal integer such that $i \in e_{j_i}$. Thus $X^i(\ell) = X^i(j_i) = X^i(j_i - 1) \setminus V(L_{j_i})$ by (D1). But if $i \neq r_{\ell+1}$ then (D5) and (D6) together imply that

$$|V(L_{j_i}) \cap X^i(j_i - 1)| = |V(L_{j_i}) \cap X^i| \equiv t_i(j_i) \equiv |X^i(j_i - 1)| + d_i \mod k - 1$$

and so $|X^i(\ell)| \equiv -d_i \mod k-1$. We claim that this also holds if $i = r_{\ell+1}$. To see this, recall that since L_j is loose, we have $|V(L_j)| \equiv 1 \mod k-1$ for each $j \in [\ell]$. Hence

$$\begin{aligned} |X^{r_{\ell+1}}(\ell)| &= |V(H) \setminus V(L_e)| - \sum_{j \in [\ell]} |V(L_j)| - \sum_{i \in [t'],\ i \neq r_{\ell+1}} |X^i(\ell)| \\ &\overset{(4.14)}{\equiv} -1 - \ell + \sum_{i \in [t']\ i \neq r_{\ell+1}} d_i \equiv -d_{r_{\ell+1}} \mod k-1 \end{aligned}$$

as $\ell + 1 = \sum_{i \in [t']} d_i$. Let $Y^i = X^i \setminus (L^*_e \cup L^*_1 \cup \dots \cup L^*_\ell)$. Since by (D3) for each $i \in [t']$ there are exactly $2(k-2)d_i$ vertices of X^i which lie in $L^*_e, L^*_1, \dots, L^*_\ell$ but not in $L_e, L_1, \dots, L_\ell$, this in turn implies that

$$|Y^i| \equiv -d_i - 2(k-2)d_i \equiv d_i \mod k-1. \tag{4.20}$$

Let $x_{\ell+1} = x_e$, $y_0 = y_e$, $L^*_0 = L^*_e$, $I_{\ell+1} = I_0$ and $I'_{\ell+1} = I'_0$. In order to complete our prepaths $L^*_0, \dots, L^*_\ell$ to a Hamilton cycle we wish to choose d_i disjoint loose paths $L^i_1, \dots, L^i_{d_i}$ within each $H[Y^i]$ which together contain all the vertices in Y^i and which 'connect' successive prepaths L^*_j. We achieve this as follows. Let J_i be the set of all $j \in [\ell+1]$ with $r_j = i$. So J_i is the set of positions at which i occurs as an initial, final or link vertex in our walk W and $|J_i| = d_i$. Let $j_1 \leqslant \dots \leqslant j_{d_i}$ be the elements of J_i. Then we choose the L^i_s ($s \in [d_i]$) in such a way that the initial vertex of L^i_s lies in $F'_{j_s - 1}$ and its final vertex lies in I'_{j_s}, all the L^i_s are disjoint and together they cover all the vertices in Y^i. To see that this can be done, first note that $|X^i \setminus Y^i| \leqslant \ell(4k^3 + 2(k-2)) + 2(k-2) \ll \varepsilon n_1$. So using Lemma 4.7 together with (C5) and (4.20) it is easy to check that the complete k-partite k-graph on Y^i contains such paths (*e.g.* first choose $L^i_1, \dots, L^i_{d_i-1}$, each consisting of precisely 2 edges, and then apply (C5) and Lemma 4.7 to find a loose path L^i_s containing all the remaining vertices of Y^i). Now (C3) and (D4) together imply that $G^i[Y^i] \setminus M^i[Y^i]$ contains the k-complexes induced by these paths (*i.e.* it contains $(L^i_1)^{\leqslant}, \dots, (L^i_{d_i})^{\leqslant}$). But this means that we can find the required paths $L^i_1, \dots, L^i_{d_i}$ in each $H[Y^i]$.

Finally, for each $s \in [d_i]$ write L'_{j_s} for L^i_s and x'_{j_s} for its initial and y'_{j_s} for its final vertex (where j_s is as defined in the previous paragraph). To obtain our Hamilton cycle of H we first traverse $L_0 = L_e$, then we use the edge $F_0 \cup \{y_0, x'_1\}$ in order to move to the initial vertex x'_1 of L'_1. (This is possible since $x'_1 \in F'_0$.) Now we traverse L'_1 and use the edge $I_1 \cup \{y'_1, x_1\}$ to get to x_1. (Again, this is possible since $y'_1 \in I'_1$.) Next we traverse L_1 and use the edge $F_1 \cup \{y_1, x'_2\}$ to move to x'_2. We continue in this way until we have reached the initial vertex $x_{\ell+1} = x_e$ of $L_0 = L_e$ again. (So in the last step we traversed $L'_{\ell+1}$ and used the edge $I_{\ell+1} \cup \{y'_{\ell+1}, x_{\ell+1}\}$.) This completes the proof of Theorem 1.10. $\square$

CHAPTER 5

HAMILTON ℓ-CYCLES IN UNIFORM HYPERGRAPHS

In this chapter we prove Theorem 1.11, which is restated below for ease of reference.

Theorem 1.11 *For all $k \geqslant 3$, $1 \leqslant \ell \leqslant k-1$ such that $(k-\ell) \nmid k$ and any $\eta > 0$ there exists n_0 so that if $n \geqslant n_0$ and $(k-\ell)|n$ then any k-graph $\mathcal{H}$ on n vertices with $\delta(\mathcal{H}) \geqslant \left(\frac{1}{\lceil \frac{k}{k-\ell} \rceil (k-\ell)} + \eta\right) n$ contains a Hamilton ℓ-cycle.*

5.1 Outline of the proof of Theorem 1.11

In our proof of Theorem 1.11 we construct the Hamilton ℓ-cycle by finding several ℓ-paths and joining them into a spanning ℓ-cycle. Here a k-graph P is an *ℓ-path* if its vertices can be given a linear ordering such that every edge of P consists of k consecutive vertices, and so that every pair of consecutive edges of P (in the natural ordering induced on the edges) intersect in precisely ℓ vertices. We say that an enumeration $v_1, v_2, \dots, v_r$ of the vertices of P is a *vertex sequence* of P if the edges of P are $\{v_{s(k-\ell)+1}, \dots, v_{s(k-\ell)+k}\}$ for each $0 \leqslant s \leqslant (r-k)/(k-\ell)$.

We say that ordered sets A and B are *ordered ends* of P if $|A| = |B| = \ell$ and A and B are initial and final segments of a vertex sequence of P. This allows us to join up ℓ-paths in the following manner. Let P and Q be ℓ-paths, and let P^{beg} and P^{end} be ordered ends of P, and Q^{beg} and Q^{end} be ordered ends of Q. Suppose that $P^{end} = Q^{beg}$, and that $V(P) \cap V(Q) = P^{end}$. Then the k-graph with vertex set $V(P) \cup V(Q)$ and with all the edges of P and of Q is an ℓ-path with ordered ends P^{beg} and Q^{end}.

Our proof of Theorem 1.11 uses ideas of [16], which in turn were based on the 'absorbing path' method of [39] and [40]. Our proof contains further developments of the method, which may be of independent interest. We use the regularity method to prove three preliminary lemmas, which are as follows.

- The 'absorbing path lemma' states that if $(k-\ell) \nmid k$ then in any sufficiently large k-graph of large minimum degree there exists an ℓ-path P which can 'absorb' any small set X of vertices outside P. By this we mean that for any such small set X there is another ℓ-path Q with the same ordered ends as P and with $V(Q) = V(P) \cup X$. Then we can think of replacing P with Q as 'absorbing' the vertices of X into P.
- The 'path cover lemma' states that any sufficiently large k-graph satisfying the minimum degree condition of Theorem 1.11 can be almost covered by a bounded number of disjoint ℓ-paths.
- The 'diameter lemma' states that if $(k-\ell) \nmid k$ then any sufficiently large k-graph of large minimum degree has small diameter in the sense that we can find an ℓ-path from any ordered ℓ-set of vertices to any other ordered ℓ-set of vertices.

These three lemmas then combine to prove Theorem 1.11 as follows. Firstly, we find in $\mathcal{H}$ an absorbing ℓ-path, and then we almost cover the induced k-graph on the remaining vertices by disjoint ℓ-paths. Next we use the diameter lemma to connect up all of these ℓ-paths to form an

ℓ-cycle C which thus contains almost every vertex of $\mathcal{H}$. Finally, we absorb all vertices of $\mathcal{H}$ not contained in C into our absorbing path, thereby forming an ℓ-cycle containing every vertex of $\mathcal{H}$.

The combination of these three preliminary lemmas to prove Theorem 1.11 is very similar to the method of Hàn and Schacht in [16]. On the other hand, to prove these preliminary results substantial changes to the method of Hàn and Schacht were required. For example, for $1 \leqslant \ell < k/2$ the diameter lemma as described above is an immediate consequence of the minimum degree condition of $\mathcal{H}$. Indeed, if A and B are ordered ℓ-sets of vertices of $\mathcal{H}$, then we may add any $k - 1 - 2\ell$ vertices from outside A and B to $A \cup B$ to obtain a set S of size $k - 1$. Then we can apply the minimum degree condition of $\mathcal{H}$ to find a vertex $x \in \mathcal{H} - S$ such that $S \cup \{x\}$ is an edge of $\mathcal{H}$; then this single edge is the desired path in $\mathcal{H}$. However, if $\ell \geqslant k/2$ then things are more difficult. Indeed, in Section 5.4 we use the regularity method (with the notion of strong hypergraph regularity introduced in Section 5.3) to prove a 'diameter lemma' as stated above. A similar assertion for the case $\ell = k - 1$, (called the 'Connecting Lemma'), was proved in [40]. The proof is quite different from ours.

In a similar way, it is more difficult to prove an absorbing path lemma for $\ell \geqslant k/2$ than for $1 \leqslant \ell < k/2$. So whereas Hàn and Schacht were able to prove a similar result using only weak hypergraph regularity, we have found it necessary to use strong hypergraph regularity for the proof of our absorbing path lemma, which is given in Section 5.5. Actually, we cannot absorb arbitrary sets of vertices, but only 'good' ℓ-sets of vertices. We show that most ℓ-sets of vertices are good, which is sufficient for our purposes. This weaker notion of absorption may be useful for other problems.

In Section 5.6 we prove the path cover lemma. A similar result was already proved in [40]. The main difference is that they used weak regularity, whereas we have used strong regularity, but this is simply to avoid having to introduce multiple notions of regularity — weak regularity would have sufficed for this part of our proof.

Finally, in Section 5.7 we complete the proof as outlined earlier.

5.2 Definitions and a preliminary result

In this section we give various definitions which we use throughout the rest of this chapter. Due to the substantial differences between the system of notation used in this chapter and that used in the previous chapter (as discussed in the introduction) we redefine several terms here which were previously defined in Chapter 4.

Let $\mathcal{H}$ be a k-graph on vertex set V, with edge set E. Then the *order* of $\mathcal{H}$, denoted $|\mathcal{H}|$, is the number of vertices of $\mathcal{H}$ (so $|\mathcal{H}| = |V|$). For $A \subseteq V$, the *neighbourhood* of A is $N_{\mathcal{H}}(A) := \{B \subseteq V : A \cup B \in E, A \cap B = \emptyset\}$. The *degree* of A, denoted $d_{\mathcal{H}}(A)$, is the number of edges of $\mathcal{H}$ which contain A as a subset, so $d_{\mathcal{H}}(A) = |N_{\mathcal{H}}(A)|$. This is consistent with our previous definition of degree for sets of $k-1$ vertices. For any $V' \subseteq V$, the *restriction of $\mathcal{H}$ to V'*, denoted $\mathcal{H}[V']$, is the k-graph with vertex set V' and edges all those edges of $\mathcal{H}$ which are subsets of V'.

Given two ordered ℓ-sets of vertices of $\mathcal{H}$, say S and T, an *ℓ-path from S to T* in $\mathcal{H}$ is an ℓ-path in $\mathcal{H}$ which has a vertex sequence beginning with the ordered ℓ-set S and ending with the ordered ℓ-set T (*i.e.* an ℓ-path with ordered ends S and T). We say that a k-graph $\mathcal{H}$ is *s-partite* if its vertex set V can be partitioned into s vertex classes $V_1, \dots, V_s$ such that no edge of $\mathcal{H}$ contains more than one vertex from any vertex class V_i. We denote by $\mathcal{K}[V_1, \dots, V_s]$ the *complete s-partite k-graph* with vertex classes $V_1, \dots, V_s$, *i.e.* the k-graph with vertex set $V = V_1 \cup \dots \cup V_s$ and edges all sets $S \in \binom{V}{k}$ with $|S \cap V_i| \leqslant 1$ for all i.

The following proposition regarding the existence of ℓ-paths in complete k-partite k-graphs is required in the proof of both the diameter lemma and the absorbing path lemma.

Proposition 5.1 *Suppose that $k \geqslant 3$, and that $1 \leqslant \ell \leqslant k-1$ is such that $(k-\ell) \nmid k$. Let V be a set of vertices partitioned into k vertex classes $V_1, ..., V_k$, with $|V_i| = k\ell(k-\ell)+1$ for each i, and let P^{beg} and P^{end} be disjoint ordered sets of ℓ vertices from V such that $|P^{beg} \cap V_i| \leqslant 1$ and $|P^{end} \cap V_i| \leqslant 1$ for each $1 \leqslant i \leqslant k$. Then $\mathcal{K}[V_1, ..., V_k]$ contains an ℓ-path P from P^{beg} to P^{end} containing every vertex of V (so $|V(P)| = k^2\ell(k-\ell)+k$).*

Proof. To prove this result, we consider strings (finite sequences of characters) on character set $[k]$. We denote the ith character of a string S by S_i. By an ordering of $[k]$ we mean a string of length k which contains each character precisely once. Let A and B be orderings of $[k]$. We say that A and B are adjacent if we can obtain B from A by swapping a single pair of adjacent characters in A. So for example, 12345 is adjacent to 12435.

Suppose that A and B are adjacent orderings of $[k]$, and let i and $i+1$ be the positions in A of the characters swapped to obtain B from A (so $1 \leqslant i \leqslant k-1$). Since $(k-\ell) \nmid k$ we may choose $p \in \{1,2\}$ such that $(k-\ell) \nmid ((p-1)k+i)$. Then define the string $S(A,B)$ to consist of p consecutive copies of A followed by $(k-\ell+1)-p$ copiesof B. Then $S(A,B)$ has length $(k-\ell+1)k$ and the property that $S(A,B)$ starts with A and ends with B. Note that the only consecutive subsequence of S of length k which contains some character more than once is $S' = S(A,B)_{(p-1)k+i+1} \dots S(A,B)_{pk+i}$. In other words, S' contains the final $k-i$ characters of A and the first i characters of B, and the first and final character of S' is A_{i+1}. Therefore, as $(k-\ell) \nmid ((p-1)k+i)$, we know that no character appears twice in $S(A,B)_{r(k-\ell)+1}, \dots, S(A,B)_{r(k-\ell)+k}$ for any $0 \leqslant r \leqslant k$. Furthermore, $S(A,B)$ contains the same number of copies of each character.

Now, choose a string C to be any ordering of $[k]$ such that for $1 \leqslant i \leqslant \ell$, the ith vertex of the ordered set P^{beg} lies in vertex class V_{C_i}. Define a string D to be an ordering of $[k]$ such that for

$1 \leqslant i \leqslant \ell$, the ith vertex of the ordered set P^{end} lies in vertex class $V_{D_{i+k-\ell}}$, and the characters D_i for $1 \leqslant i \leqslant k-\ell$ appear in the same order as they do in C. Then we may transform C into D through at most $k\ell$ swaps of pairs of consecutive vertices. So we may choose $A^0, \dots, A^{k\ell}$ to be orderings of $[k]$ such that $A^0 = C$, $A^{k\ell} = D$, and for any $0 \leqslant i \leqslant k\ell - 1$, A^i and A^{i+1} are either adjacent or identical.

Then for each $0 \leqslant i \leqslant k\ell - 1$ we may choose a string S^i of length $k(k-\ell+1)$ such that S^i starts with A^i and ends with A^{i+1}, each character appears an equal number of times in S^i and for each $0 \leqslant r \leqslant k$ no character appears more than once in $S^i_{r(k-\ell)+1}, \dots, S^i_{r(k-\ell)+k}$. Indeed, if A^i and A^{i+1} are adjacent, take S^i to be $S(A^i, A^{i+1})$, and if $A^i = A^{i+1}$, take S^i to be the string consisting of $k-\ell+1$ consecutive copies of A^i. For each $0 \leqslant i \leqslant k\ell-2$, let T^i be the string obtained by deleting the final k characters of S^i, and let $T^{k\ell-1} = S^{k\ell-1}$. Let S be the string formed by concatenating $T^0, \dots, T^{k\ell-1}$. Then S starts with C and ends with D and has the property that no character appears twice in $S_{r(k-\ell)+1}, \dots, S_{r(k-\ell)+k}$ for any $0 \leqslant r \leqslant k^2\ell$. Also $|S| = k^2\ell(k-\ell)+k$, and so since S contains each character the same number of times, each character appears $k\ell(k-\ell)+1$ times in S.

We can now construct the vertex sequence of our desired ℓ-path P. To do so, let P have vertex sequence beginning with P^{beg} and ending with P^{end}. In between, let the ith vertex of P be chosen from V_{S_i}, and make these choices without choosing the same vertex twice. Then P contains all $k\ell(k-\ell)+1$ vertices from each vertex class and is an ℓ-path. Indeed, the edges of an ℓ-path P consist of the vertices in positions $r(k-\ell)+1, \dots, r(k-\ell)+k$ for $0 \leqslant r \leqslant |E(P)|-1$. So by construction these vertices are from different vertex classes, and so form an edge in $\mathcal{K}[V_1, \dots, V_k]$. □

Note that Proposition 5.1 would not hold if instead we had $(k-\ell) \mid k$, as it would not be possible to choose p as in the proof.

5.3 The regularity lemma for k-graphs

5.3.1 Regular complexes

In this section we give the definition of a regular complex, and results on regularity, which we use throughout the rest of Chapter 5. The reader should be alert to the fact that there are subtle differences in the definition of regularity used here compared to the definition used in Chapter 4, and that the results used (primarily the regularity lemma) are consequently also different.

A *hypergraph* $\mathcal{H}$ consists of a vertex set $V(\mathcal{H})$ and an edge set $E(\mathcal{H})$, where every edge $e \in E(\mathcal{H})$ is a non-empty subset of $V(\mathcal{H})$. So a k-graph as defined earlier is a hypergraph in which every edge has size k. A hypergraph $\mathcal{H}$ is a *complex* if whenever $e \in E(\mathcal{H})$ and e' is a non-empty subset of e we have that $e' \in E(\mathcal{H})$. All the complexes considered in this chapter have the property that every vertex forms an edge. A complex $\mathcal{H}$ is a *k-complex* if every edge of $\mathcal{H}$ consists of at most k vertices. The edges of size i are called *i-edges* of $\mathcal{H}$. We write $|\mathcal{H}| := |V(\mathcal{H})|$ for the *order* of $\mathcal{H}$. Given a k-complex $\mathcal{H}$, for each $i = 1, \ldots, k$ we denote by $\mathcal{H}_i$ the *underlying i-graph of $\mathcal{H}$*. So the vertices of $\mathcal{H}_i$ are those of $\mathcal{H}$ and the edges of $\mathcal{H}_i$ are the i-edges of $\mathcal{H}$.

Note that a k-graph $\mathcal{H}$ can be turned into a k-complex, which we denote by $\mathcal{H}^{\leqslant}$, by making every edge into a complete i-graph $K_k^{(i)}$, for each $1 \leqslant i \leqslant k$. (In a more general k-complex we may have i-edges which do not lie within an $(i+1)$-edge.) Given $k \leqslant s$, a *(k,s)-complex* $\mathcal{H}$ is an s-partite k-complex, by which we mean that the vertex set of $\mathcal{H}$ can be partitioned into sets $V_1, \ldots, V_s$ such that every edge of $\mathcal{H}$ meets each V_i in at most one vertex.

Given $i \geqslant 2$, an i-partite i-graph $\mathcal{H}_i$, and an i-partite $(i-1)$-graph $\mathcal{H}_{i-1}$ on the same vertex set, we write $\mathcal{K}_i(\mathcal{H}_{i-1})$ for the set of i-sets of vertices which form a copy of the complete

$(i-1)$-graph $K_i^{(i-1)}$ on i vertices in $\mathcal{H}_{i-1}$. We define the *density of* $\mathcal{H}_i$ *with respect to* $\mathcal{H}_{i-1}$ to be

$$d(\mathcal{H}_i|\mathcal{H}_{i-1}) := \frac{|\mathcal{K}_i(\mathcal{H}_{i-1}) \cap E(\mathcal{H}_i)|}{|\mathcal{K}_i(\mathcal{H}_{i-1})|}$$

if $|\mathcal{K}_i(\mathcal{H}_{i-1})| > 0$, and $d(\mathcal{H}_i|\mathcal{H}_{i-1}) := 0$ otherwise. More generally, if $\mathbf{Q} := (Q(1), Q(2), \ldots, Q(r))$ is a collection of r subhypergraphs of $\mathcal{H}_{i-1}$, we define $\mathcal{K}_i(\mathbf{Q}) := \bigcup_{j=1}^{r} \mathcal{K}_i(Q(j))$ and

$$d(\mathcal{H}_i|\mathbf{Q}) := \frac{|\mathcal{K}_i(\mathbf{Q}) \cap E(\mathcal{H}_i)|}{|\mathcal{K}_i(\mathbf{Q})|}$$

if $|\mathcal{K}_i(\mathbf{Q})| > 0$, and $d(\mathcal{H}_i|\mathbf{Q}) := 0$ otherwise.

We say that $\mathcal{H}_i$ is (d_i, δ, r)*-regular with respect to* $\mathcal{H}_{i-1}$ if every r-tuple $\mathbf{Q}$ with $|\mathcal{K}_i(\mathbf{Q})| > \delta|\mathcal{K}_i(\mathcal{H}_{i-1})|$ satisfies $d(\mathcal{H}_i|\mathbf{Q}) = d_i \pm \delta$. Instead of $(d_i, \delta, 1)$-regularity we sometimes refer to (d_i, δ)*-regularity*.

Given $3 \leqslant k \leqslant s$ and a (k, s)-complex $\mathcal{H}$, we say that $\mathcal{H}$ is $(d_k, \ldots, d_2, \delta_k, \delta, r)$*-regular* if the following conditions hold:

- For every $i = 2, \ldots, k-1$ and for every i-tuple K of vertex classes either $\mathcal{H}_i[K]$ is (d_i, δ)-regular with respect to $\mathcal{H}_{i-1}[K]$ or $d(\mathcal{H}_i[K]|\mathcal{H}_{i-1}[K]) = 0$.

- For every k-tuple K of vertex classes either $\mathcal{H}_k[K]$ is (d_k, δ_k, r)-regular with respect to $\mathcal{H}_{k-1}[K]$ or $d(\mathcal{H}_k[K]|\mathcal{H}_{k-1}[K]) = 0$.

Here we write $\mathcal{H}_i[K]$ for the restriction of $\mathcal{H}_i$ to the union of all vertex classes in K. We sometimes denote $(d_k, \ldots, d_2)$ by $\mathbf{d}$ and refer to $(\mathbf{d}, \delta_k, \delta, r)$-regularity.

We need the following lemma which states that the restriction of regular complexes to a sufficiently large set of vertices is still regular.

Lemma 5.2 *Let k, s, r, m be positive integers and $\alpha, d_2, \ldots, d_k, \delta, \delta_k$ be positive constants such that*

$$\frac{1}{m} \ll \frac{1}{r}, \delta \leqslant \min\{\delta_k, d_2, \ldots, d_{k-1}\} \leqslant \delta_k \ll \alpha \ll d_k, \frac{1}{s}.$$

Let $\mathcal{H}$ be a $(\mathbf{d}, \delta_k, \delta, r)$-regular (k, s)-complex with vertex classes $V_1, \ldots, V_s$ of size m. For each i let $V_i' \subseteq V_i$ be a set of size at least αm. Then the restriction $\mathcal{H}' = \mathcal{H}[V_1' \cup \cdots \cup V_s']$ of $\mathcal{H}$ to $V_1' \cup \cdots \cup V_s'$ is $(\mathbf{d}, \sqrt{\delta_k}, \sqrt{\delta}, r)$-regular.

Proof. Note that since $\mathcal{H}'$ is the restriction of $\mathcal{H}$ to a subset of its vertex set, for any $2 \leqslant i \leqslant k$, any i-tuple K of vertex classes and any subhypergraph $\mathbf{Q}$ of $\mathcal{H}'_{i-1}[K]$ we have that $d(\mathcal{H}'_i[K]|\mathbf{Q}) = d(\mathcal{H}_i[K]|\mathbf{Q})$. It is therefore sufficient to show that $\sqrt{\delta}|\mathcal{K}_i(\mathcal{H}'_{i-1})| \geqslant \delta|\mathcal{K}_i(\mathcal{H}_{i-1})|$ for each $2 \leqslant i \leqslant k-1$, and similarly that $\sqrt{\delta_k}|\mathcal{K}_k(\mathcal{H}'_{k-1})| \geqslant \delta_k|\mathcal{K}_k(\mathcal{H}_{k-1})|$. This follows by induction on i from the dense hypergraph counting lemma (Corollary 6.11 in [27]). Indeed, suppose that for some $2 \leqslant i \leqslant k$ we have $\sqrt{\delta}|\mathcal{K}_j(\mathcal{H}'_{j-1})| \geqslant \delta|\mathcal{K}_j(\mathcal{H}_{j-1})|$ for any $2 \leqslant j < i$. Then the $(i-1, s)$-complex with j-edges $\mathcal{H}'_j$ for each $2 \leqslant j < i$ is $(d_{i-1}, \ldots, d_2, \sqrt{\delta}, \sqrt{\delta}, 1)$-regular. Then since $\delta \ll \alpha$, by the dense hypergraph counting lemma, if $2 \leqslant i \leqslant k-1$ then we have $\sqrt{\delta}|\mathcal{K}_i(\mathcal{H}'_{i-1})| \geqslant \delta|\mathcal{K}_i(\mathcal{H}_{i-1})|$. Similarly if $i = k$ then since $\delta_k \ll \alpha$ we have $\sqrt{\delta_k}|\mathcal{K}_k(\mathcal{H}'_{k-1})| \geqslant \delta_k|\mathcal{K}_k(\mathcal{H}_{k-1})|$, completing the induction. $\square$

5.3.2 Statement of the regularity lemma

In this section we state the version of the regularity lemma for k-graphs due to Rödl and Schacht [41], which we use several times in proving Theorem 1.11. To prepare for this we first need some more notation. Suppose that V is a finite set of vertices and $\mathcal{P}^{(1)}$ is a partition of V into sets $V_1, \ldots, V_{a_1}$, which we call *clusters*. Given $k \geqslant 3$ and any $j \in [k]$, we denote by $\mathrm{Cross}_j = \mathrm{Cross}_j(\mathcal{P}^{(1)})$ the set of all those j-subsets of V that meet each V_i in at most 1 vertex. For every set $A \subseteq [a_1]$ with $2 \leqslant |A| \leqslant k-1$ we write Cross_A for all those $|A|$-subsets of V

that meet each V_i with $i \in A$. Let $\mathcal{P}_A$ be a partition of Cross_A. We refer to the partition classes of $\mathcal{P}_A$ as *cells*. For each $i = 2, \ldots, k-1$ let $\mathcal{P}^{(i)}$ be the union of all the $\mathcal{P}_A$ with $|A| = i$. So $\mathcal{P}^{(i)}$ is a partition of Cross_i.

$\mathcal{P}(k-1) = \{\mathcal{P}^{(1)}, \ldots, \mathcal{P}^{(k-1)}\}$ is a *family of partitions on* V if the following condition holds. Recall that a_1 denotes the number of clusters in $\mathcal{P}^{(1)}$. Consider any $B \subseteq A \subseteq [a_1]$ such that $2 \leqslant |B| < |A| \leqslant k-1$ and suppose that $S, T \in \mathrm{Cross}_A$ lie in the same cell of $\mathcal{P}_A$. Let $S_B := S \cap \bigcup_{i \in B} V_i$ and define T_B similarly. Then S_B and T_B lie in the same cell of $\mathcal{P}_B$.

To illustrate this condition, suppose that $k = 4$ and $A = [3]$. Then $\mathcal{P}_{\{1,2\}}$, $\mathcal{P}_{\{2,3\}}$ and $\mathcal{P}_{\{1,3\}}$ partition the edges of the 3 complete bipartite graphs induced by the pairs V_1V_2, V_2V_3 and V_1V_3. These partitions together naturally induce a partition $\mathcal{Q}$ of the set of triples induced by V_1, V_2 and V_3. The above condition says that $\mathcal{P}_{\{1,2,3\}}$ must be a refinement of $\mathcal{Q}$.

Given $1 \leqslant i \leqslant j \leqslant k$ with $i < k$, $J \in \mathrm{Cross}_j$ and an i-set $Q \subseteq J$, we write C_Q for the set of all those i-sets in Cross_i that lie in the same cell of $\mathcal{P}^{(i)}$ as Q. (In particular, if $i = 1$ then C_Q is the cluster containing the unique element in Q.) The *polyad* $\widehat{P}^{(i)}(J)$ *of* J is defined by $\widehat{P}^{(i)}(J) := \bigcup_Q C_Q$, where the union is over all i-subsets Q of J. So we can view $\widehat{P}^{(i)}(J)$ as an j-partite i-graph (whose vertex classes are the clusters intersecting J). We let $\widehat{\mathcal{P}}^{(j-1)}$ be the set consisting of all the $\widehat{P}^{(j-1)}(J)$ for all $J \in \mathrm{Cross}_j$. So for each $K \in \mathrm{Cross}_k$ we can view $\bigcup_{i=1}^{k-1} \widehat{P}^{(i)}(K)$ as a $(k-1,k)$-complex.

We say that $\mathcal{P} = \mathcal{P}(k-1)$ is (η, δ, t)-*equitable* if

- there exists $\mathbf{d} = (d_{k-1}, \ldots, d_2)$ such that $d_i \geqslant 1/t$ and $1/d_i \in \mathbb{N}$ for all $i = 2, \ldots, k-1$,
- $\mathcal{P}^{(1)}$ is a partition of V into a_1 clusters of equal size, where $1/\eta \leqslant a_1 \leqslant t$,
- for all $i = 2, \ldots, k-1$, $\mathcal{P}^{(i)}$ is a partition of Cross_i into at most t cells,

- for every $K \in \mathrm{Cross}_k$, the $(k-1,k)$-complex $\bigcup_{i=1}^{k-1} \widehat{P}^{(i)}(K)$ is $(\mathbf{d}, \delta, \delta, 1)$-regular.

Note that the final condition implies that for all $i = 2, \ldots, k-1$ the cells of $\mathcal{P}^{(i)}$ have almost equal size.

Let $\delta_k > 0$ and $r \in \mathbb{N}$. Suppose that $\mathcal{H}$ is a k-graph on V and $\mathcal{P} = \mathcal{P}(k-1)$ is a family of partitions on V. Given a polyad $\widehat{P}^{(k-1)} \in \widehat{\mathcal{P}}^{(k-1)}$, we say that $\mathcal{H}$ is (δ_k, r)*-regular with respect to* $\widehat{P}^{(k-1)}$ if $\mathcal{H}$ is (d, δ_k, r)-regular with respect to $\widehat{P}^{(k-1)}$ for some d. We say that $\mathcal{H}$ is (δ_k, r)*-regular with respect to* $\mathcal{P}$ if

$$\left|\bigcup\{\mathcal{K}_k(\widehat{P}^{(k-1)}) : \ \mathcal{H} \text{ is not } (\delta_k, r)\text{-regular with respect to } \widehat{P}^{(k-1)} \in \widehat{\mathcal{P}}^{(k-1)}\}\right| \leqslant \delta_k |V|^k.$$

This means that not much more than a δ_k-fraction of the k-subsets of V form a $K_k^{(k-1)}$ that lies within a polyad with respect to which $\mathcal{H}$ is not regular.

Now we are ready to state the regularity lemma.

Theorem 5.3 (Rödl and Schacht [41], Theorem 17) *Let $k \geqslant 3$ be a fixed integer. For all positive constants η and δ_k and all functions $r : \mathbb{N} \to \mathbb{N}$ and $\delta : \mathbb{N} \to (0, 1]$, there are integers t and n_0 such that the following holds for all $n \geqslant n_0$ which are divisible by $t!$. Suppose that $\mathcal{H}$ is a k-graph of order n. Then there exists a family of partitions $\mathcal{P} = \mathcal{P}(k-1)$ of the vertex set V of $\mathcal{H}$ such that*

1. *$\mathcal{P}$ is $(\eta, \delta(t), t)$-equitable and*
2. *$\mathcal{H}$ is $(\delta_k, r(t))$-regular with respect to $\mathcal{P}$.*

Similar results were proved earlier by Rödl and Skokan [43] and Gowers [13]. Note that the constants in Theorem 5.3 can be chosen so that they satisfy the following hierarchy:

$$\frac{1}{n_0} \ll \frac{1}{r} = \frac{1}{r(t)}, \delta = \delta(t) \ll \min\{\delta_k, \frac{1}{t}\} \ll \eta.$$

5.3.3 The reduced k-graph

To prove the absorbing lemma and the path cover lemma, we use the so-called reduced k-graph. Suppose that we have constants

$$\frac{1}{n_0} \ll \frac{1}{r}, \delta \ll \min\{\delta_k, \frac{1}{t}\} \leqslant \delta_k, \eta \ll d \ll \theta \ll \mu, \frac{1}{k}.$$

and a k-graph $\mathcal{H}$ on V of order $n \geqslant n_0$ with $\delta(\mathcal{H}) \geqslant (\mu + \theta)n$. We may apply the regularity lemma to $\mathcal{H}$ to obtain a family of partitions $\mathcal{P} = \{\mathcal{P}^{(1)}, \dots, \mathcal{P}^{(k-1)}\}$ of V. Then the *reduced k-graph* $\mathcal{R} = \mathcal{R}(\mathcal{H}, \mathcal{P})$ is the k-graph whose vertices are the clusters of $\mathcal{H}$, *i.e.* the parts of $\mathcal{P}^{(1)}$. A k-tuple of clusters forms an edge of $\mathcal{R}$ if there is some polyad $\widehat{P}^{(k-1)}$ induced on these k clusters such that $\mathcal{H}$ is (d', δ_k, r)-regular with respect to $\widehat{P}^{(k-1)}$ for some $d' \geqslant d$. To make use of the reduced k-graph, we need to show that it almost inherits the minimum degree condition from $\mathcal{H}$.

Lemma 5.4 *All but at most $\theta|\mathcal{R}|^{k-1}$ sets $S \in \binom{V(\mathcal{R})}{k-1}$ satisfy $d_{\mathcal{R}}(S) \geqslant \mu|\mathcal{R}|$.*

Similar results have been proved previously, but we include the short proof for completeness. We say that an edge e of $\mathcal{H}$ is *useful* if it lies in $\mathcal{K}_k(\widehat{P}^{(k-1)})$ for some $\widehat{P}^{(k-1)} \in \widehat{\mathcal{P}}^{(k-1)}$ such that $\mathcal{H}$ is (d', δ_k, r)-regular with respect to $\widehat{P}^{(k-1)}$ for some $d' \geqslant d$. Note that if e lies in $\mathcal{K}_k(\widehat{P}^{(k-1)})$ then $\widehat{P}^{(k-1)} = \widehat{P}^{(k-1)}(e)$ is the polyad of e. Moreover, if e is a useful edge of $\mathcal{H}$, and $V_{i_1}, \dots, V_{i_k}$ are the clusters containing the vertices of e, then these k clusters form an edge of $\mathcal{R}$.

Lemma 5.5 *At most $2dn^k$ edges of $\mathcal{H}$ are not useful.*

Proof. There are three reasons why an edge of $\mathcal{H}$ may not be useful. Firstly, it may lie in $\binom{V}{k} \setminus \mathrm{Cross}_k$. Since $\mathcal{P}^{(1)}$ partitions V into a_1 clusters of equal size, there are at most $\frac{n}{a_1}n^{k-1} \leqslant \eta n^k$ edges of this type. Secondly, the edge may lie in a polyad $\widehat{P}^{(k-1)} \in \widehat{\mathcal{P}}^{(k-1)}$ such that $|E(\mathcal{H}) \cap \mathcal{K}_k(\widehat{P}^{(k-1)})| \leqslant d|\mathcal{K}_k(\widehat{P}^{(k-1)})|$. There are at most dn^k edges of this type. Finally, the edge may lie in a polyad $\widehat{P}^{(k-1)} \in \widehat{\mathcal{P}}^{(k-1)}$ such that $\mathcal{H}$ is not (δ_k, r)-regular with respect to $\widehat{P}^{(k-1)}$. Since $\mathcal{H}$ is (δ_k, r)-regular with respect to $\mathcal{P}$, there are at most $\delta_k n^k$ edges of this type. So altogether, at most $(\delta_k + d + \eta)n^k \leqslant 2dn^k$ edges of $\mathcal{H}$ are not useful. $\square$

Proof of Lemma 5.4. Let $m = |V_1| = \cdots = |V_{a_1}|$ be the size of the clusters. We say that a $(k-1)$-tuple of clusters of $\mathcal{H}$ is *poor* if there are at least $\theta m^{k-1}n$ edges of $\mathcal{H}$ which intersect each of the $k-1$ clusters in precisely one vertex and which are not useful. Then it follows from Lemma 5.5 that at most $\theta|\mathcal{R}|^{k-1}$ such $(k-1)$-tuples are poor. So it remains to show that any $(k-1)$-tuple which is not poor has many neighbours in $\mathcal{R}$. But if $V_{i_1}, \ldots, V_{i_{k-1}}$ is a $(k-1)$-tuple which is not poor, then there are at least $m^{k-1}\delta(\mathcal{H}) - \theta m^{k-1}n \geqslant \mu m^{k-1}n$ useful edges of $\mathcal{H}$ which intersect each of $V_{i_1}, \ldots, V_{i_{k-1}}$ in precisely one vertex. For any other cluster V_j at most m^k edges of $\mathcal{H}$ intersect each of $V_{i_1}, \ldots, V_{i_{k-1}}, V_j$ in precisely one vertex, and so there are at least $\mu n/m = \mu|\mathcal{R}|$ choices of V_j such that there is at least one such useful edge. This useful edge indicates the existence of a polyad satisfying the conditions of an edge in the reduced k-graph $\mathcal{R}$. $\square$

5.3.4 The embedding and extension lemmas

To prove Theorem 1.11 we also use an embedding lemma, which guarantees the existence of a copy of a complex $\mathcal{G}$ of bounded maximum degree inside a suitable regular complex $\mathcal{H}$, where the order of $\mathcal{G}$ is allowed to be linear in the order of $\mathcal{H}$. In order to state this lemma, we need some more definitions.

The degree of a vertex x in a complex $\mathcal{G}$ is the number of edges containing x. The *maximum vertex degree* of $\mathcal{G}$ is the largest degree of a vertex of $\mathcal{G}$. Suppose that $\mathcal{H}$ is a (k, s)-complex with vertex classes $V_1, \ldots, V_s$, which all have size m. Suppose also that $\mathcal{G}$ is a (k, s)-complex with vertex classes $X_1, \ldots, X_s$ of size at most m. We say that $\mathcal{H}$ *respects the partition of* $\mathcal{G}$ if whenever $\mathcal{G}$ contains an i-edge with vertices in $X_{j_1}, \ldots, X_{j_i}$, then there is an i-edge of $\mathcal{H}$ with vertices in $V_{j_1}, \ldots, V_{j_i}$. On the other hand, we say that a labelled copy of $\mathcal{G}$ in $\mathcal{H}$ is *partition-respecting* if for each $i = 1, \ldots, s$ the vertices corresponding to those in X_i lie within V_i.

Lemma 5.6 (Embedding lemma, [8], Theorem 3) *Let Δ, k, s, r, m_0 be positive integers and let $c, d_2, \ldots, d_k, \delta, \delta_k$ be positive constants such that $1/d_i \in \mathbb{N}$ for all $i < k$,*

$$\frac{1}{m_0} \ll \frac{1}{r}, \delta \ll \min\{\delta_k, d_2, \ldots, d_{k-1}\} \leqslant \delta_k \ll d_k, \frac{1}{\Delta}, \frac{1}{s}$$

and

$$c \ll d_2, \ldots, d_k.$$

Then the following holds for all integers $m \geqslant m_0$. Suppose that $\mathcal{G}$ is a (k, s)-complex of maximum vertex degree at most Δ with vertex classes $X_1, \ldots, X_s$ such that $|X_i| \leqslant cm$ for all $i = 1, \ldots, s$. Suppose also that $\mathcal{H}$ is a $(\mathbf{d}, \delta_k, \delta, r)$-regular (k, s)-complex with vertex classes $V_1, \ldots, V_s$, all of size m, which respects the partition of $\mathcal{G}$. Then $\mathcal{H}$ contains a labelled partition-respecting copy of $\mathcal{G}$.

We also use the following weak version of a lemma from [8]. Roughly speaking, it states that if $\mathcal{G}$ is an induced subcomplex of $\mathcal{G}'$, and $\mathcal{H}$ is suitably regular, then almost all copies of $\mathcal{G}$ in $\mathcal{H}$ can be extended to a large number of copies of $\mathcal{G}'$ in $\mathcal{H}$. We write $|\mathcal{G}|_{\mathcal{H}}$ for the number of labelled partition-respecting copies of $\mathcal{G}$ in $\mathcal{H}$.

Lemma 5.7 (Extension lemma, [8], Lemma 5) *Let k, s, r, b', b'', m_0 be positive integers, where*

$b' < b''$, and let $c, \beta, d_2, \ldots, d_k, \delta, \delta_k$ be positive constants such that $1/d_i \in \mathbb{N}$ for all $i < k$ and

$$\frac{1}{m_0} \ll \frac{1}{r}, \delta \ll c \ll \min\{\delta_k, d_2, \ldots, d_{k-1}\} \leqslant \delta_k \ll \beta, d_k, \frac{1}{s}, \frac{1}{b''}.$$

Then the following holds for all integers $m \geqslant m_0$. Suppose that $\mathcal{G}'$ is a (k,s)-complex on b'' vertices with vertex classes $X_1, \ldots, X_s$ and let $\mathcal{G}$ be an induced subcomplex of $\mathcal{G}'$ on b' vertices. Suppose also that $\mathcal{H}$ is a $(\mathbf{d}, \delta_k, \delta, r)$-regular (k,s)-complex with vertex classes $V_1, \ldots, V_s$, all of size m, which respects the partition of $\mathcal{G}'$. Then all but at most $\beta|\mathcal{G}|_{\mathcal{H}}$ labelled partition-respecting copies of $\mathcal{G}$ in $\mathcal{H}$ are extendible to at least $cn^{b''-b'}$ labelled partition-respecting copies of $\mathcal{G}'$ in $\mathcal{H}$.

The proofs of Lemmas 5.6 and 5.7 rely on the hypergraph counting lemma (Theorem 9 in [42]). In particular, the extension lemma is a straightforward consequence of the counting lemma. Actually both the embedding lemma and the extension lemma involved the additional condition that $1/d_k \in \mathbb{N}$. However, this can easily be achieved by working with a subcomplex $\mathcal{H}'$ of $\mathcal{H}$ which is $(d'', d_{k-1}, \ldots, d_2, \delta_k, \delta, r)$-regular with respect to $\widehat{P}^{(k-1)}$ for some $d'' \gg \delta_k$ with $1/d'' \in \mathbb{N}$. The existence of such an $\mathcal{H}'$ follows immediately from the slicing lemma ([41], Proposition 22), which is proved using a simple application of a Chernoff bound.

Now suppose that we have applied the regularity lemma (Theorem 5.3) to a k-graph $\mathcal{H}$ to obtain a reduced k-graph $\mathcal{R}$. An edge e of $\mathcal{R}$ indicates that we can apply the embedding lemma or the extension lemma to the subcomplex of $\mathcal{H}$ whose vertex classes are the clusters $V_1, \ldots, V_k$ corresponding to the vertices of e. More precisely, since e is an edge of $\mathcal{R}$, there is some polyad $\widehat{P}^{(k-1)} = \widehat{P}^{(k-1)}(K)$ (where $K \in \mathrm{Cross}_k$) induced by $V_1, \ldots, V_k$ such that $\mathcal{H}$ is (d', δ_k, r)-regular with respect to $\widehat{P}^{(k-1)}$ for some $d' \geqslant d$. Let $\mathcal{H}^*$ be the (k,k)-complex obtained from the $(k-1,k)$-complex $\bigcup_{i=1}^{k-1} \widehat{P}^{(i)}(K)$ by adding $E(\mathcal{H}) \cap \mathcal{K}(\widehat{P}^{(k-1)})$ as the 'kth level'. Then $\mathcal{H}^*$ is a $(\mathbf{d}, \delta_k, \delta, r)$-regular subcomplex of $\mathcal{H}$, where $\mathbf{d} = (d', d_{k-1}, \ldots, d_2)$, and $(d_{k-1}, \ldots, d_2)$ is as in the definition of an (η, δ, t)-equitable family of partitions. Also $\mathcal{H}^*$ satisfies the conditions of the embedding (or extension) lemma. So in particular, the embedding lemma implies that if $m := |V_1|$ and $\mathcal{G}$ is a k-partite k-graph of bounded maximum vertex degree whose vertex classes have size at most cm, then $\mathcal{H}$ contains a copy of $\mathcal{G}$.

5.4 The diameter lemma

In this section, we prove a diameter lemma, which states that if $\mathcal{H}$ is a sufficiently large k-graph of large minimum degree, then for any ordered ℓ-sets A and B of vertices of $\mathcal{H}$ we can find an ℓ-path from A to B in $\mathcal{H}$. To prove this result, we first consider a k-graph $\mathcal{W}(k, \ell)$, for which a similar statement is easier to prove (Proposition 5.8). For $k/2 \leqslant \ell \leqslant k-2$, the k-graph $\mathcal{W}(k, \ell)$ has $4\ell - k + 2$ vertices in three disjoint sets X, Y and Z, where $X = \{x_1, \dots, x_\ell\}, Y = \{y_1, \dots, y_\ell\}$ and $Z = \{z_1, \dots, z_{2\ell-k+2}\}$. $\mathcal{W}(k, \ell)$ has $2\ell - k + 2$ edges, where for $1 \leqslant i \leqslant 2\ell - k + 2$ the ith edge of $\mathcal{W}(k, \ell)$ is $\{x_1, \dots, x_{\ell+1-i}\} \cup \{y_1, \dots, y_{k-2-\ell+i}\} \cup \{z_i\}$. So each edge of $\mathcal{W}(k, \ell)$ intersects the following edge in precisely $k-2$ vertices. We sometimes view $\mathcal{W}(k, \ell)$ as a $(4\ell - k + 2)$-partite k-graph with a single vertex in each vertex class, and consider the $(k, 4\ell - k + 2)$-complex $\mathcal{W}(k, \ell)^{\leqslant}$. We refer to the ordered sets X and Y as the ordered ends of $\mathcal{W}(k, \ell)$.

The next proposition states that for most pairs of sets S and T of ℓ vertices in a k-graph $\mathcal{H}$ of large minimum degree, $\mathcal{H}$ contains many copies of $\mathcal{W}(k, \ell)$ with S and T as ordered ends.

Proposition 5.8 *Suppose that $k \geqslant 3$, that $k/2 \leqslant \ell \leqslant k-2$ and that $1/n \ll \gamma \ll \beta \ll \mu, 1/k$. Let $\mathcal{H}$ be a k-graph on n vertices such that $d(S) \geqslant \mu n$ for all but at most γn^{k-1} sets $S \in \binom{V(\mathcal{H})}{k-1}$. Then for all but at most $\beta n^{2\ell}$ pairs S, T of ordered ℓ-sets of vertices of $\mathcal{H}$ there are at least $\beta n^{2\ell-k+2}$ copies of $\mathcal{W}(k, \ell)$ in $\mathcal{H}$ with ordered ends S and T.*

Proof. We refer to the at most γn^{k-1} sets S of $k-1$ vertices in $\mathcal{H}$ which do not satisfy $d(S) \geqslant \mu n$ as unfriendly $(k-1)$-sets. We say that a pair of ℓ-sets S and T is unfriendly if there exist $S' \subseteq S$, $T' \subseteq T$ such that $S' \cup T'$ is an unfriendly $(k-1)$-set. Then for any unfriendly $(k-1)$-set B, there are at most $2^{k-1}n^{2\ell-k+1}$ pairs of ℓ-sets S and T with $S' \cup T' = B$ for some $S' \subseteq S$ and $T' \subseteq T$, and so since there are at most γn^{k-1} unfriendly $(k-1)$-sets, and $\gamma \ll \beta \ll 1/k$, we know that there are at most $\beta n^{2\ell}$ unfriendly pairs of ℓ-sets.

To complete the proof, it is sufficient to show that if the pair S, T of ordered ℓ-sets is not unfriendly, then $\mathcal{H}$ contains at least $\beta n^{2\ell-k+2}$ copies of $\mathcal{W}(k,\ell)$ with ordered ends S and T. Let $S = \{x_1, \dots, x_\ell\}$, and let $T = \{y_1, \dots, y_\ell\}$. For each $1 \leqslant i \leqslant 2k-\ell+2$ we choose a vertex z_i such that $z_i \notin S \cup T$, $z_i \neq z_j$ for any $j < i$, and such that $\{x_1, \dots, x_{\ell+1-i}, y_1, \dots, y_{k-2-\ell+i}, z_i\}$ is an edge of $\mathcal{H}$. This is possible for each i as we know that S, T is not an unfriendly pair, and so $d(\{x_1, \dots, x_{\ell+1-i}, y_1, \dots, y_{k-2-\ell+i}\}) \geqslant \mu n$, and hence there are at least $\mu n - (4\ell - k + 2)$ vertices to choose from. Then S, T and the chosen vertices z_i together form a copy of $\mathcal{W}(k,\ell)$ in $\mathcal{H}$ with ordered ends S and T. Since $\beta \ll \mu$, by counting the choices we could have made for the z_i we find that $\mathcal{H}$ contains at least $\beta n^{2\ell-k+2}$ copies of $\mathcal{W}(k,\ell)$ with ordered ends S and T. $\square$

The following proposition relates the k-graph $\mathcal{W}(k,\ell)$ to a k-graph $\mathcal{P}(k,\ell)$ which consists of several ℓ-paths from one ordered ℓ-set to another. We say that ℓ-paths P and Q with ordered ends P^{beg}, P^{end}, Q^{beg} and Q^{end} are *internally disjoint* if P and Q do not intersect other than in these ordered ends. Note that the proof of this proposition uses Proposition 5.1. As a consequence this proposition and each of the remaining results of this section, including the diameter lemma, require that $(k-\ell) \nmid k$.

Proposition 5.9 *Suppose that $k \geqslant 3$ and that $k/2 \leqslant \ell \leqslant k-1$ is such that $(k-\ell) \nmid k$. Then there exists a $(4\ell-k+2)$-partite k-graph $\mathcal{P}(k,\ell)$ such that the following conditions hold.*

(1) $\mathcal{P}(k,\ell)$ is the union of $4\ell+1$ internally disjoint ℓ-paths, each containing between $k^2\ell$ and $2k^5$ vertices, with identical ordered ℓ-sets T_1 and T_2 as ordered ends (we refer to these as

the ordered ends of $\mathcal{P}(k,\ell)$). In particular, $\mathcal{P}(k,\ell)$ contains at most $10k^6$ vertices.

(2) The vertex classes of $\mathcal{P}(k,\ell)$ are disjoint sets V_w, one for each vertex w of $\mathcal{W}(k,\ell)$.

(3) Whenever $v_1 \in V_{w_1}, v_2 \in V_{w_2}, \ldots, v_k \in V_{w_k}$ are such that $\{v_1, v_2, \ldots, v_k\}$ is an edge of $\mathcal{P}(k,\ell)$, $\{w_1, \ldots, w_k\}$ is an edge of $\mathcal{W}(k,\ell)$. Furthermore, let X and Y be the ordered ends of $\mathcal{W}(k,\ell)$. Then the ordered ends of $\mathcal{P}(k,\ell)$ are contained in $\bigcup_{w\in X} V_w$ and $\bigcup_{w\in Y} V_w$ respectively.

Proof. For every vertex w of $\mathcal{W}(k,\ell)$, take a large vertex set V_w. Define $\mathcal{W}^*$ to have vertex set $V = \bigcup_{w\in\mathcal{W}(k,\ell)} V_w$, and edges precisely those k-sets of vertices which lie in sets corresponding to an edge of $\mathcal{W}(k,\ell)$. We construct $\mathcal{P}(k,\ell)$ to be a sub-k-graph of $\mathcal{W}^*$, with the ordered ends of $\mathcal{P}(k,\ell)$ in the sets V_w corresponding to the ordered ends of $\mathcal{W}(k,\ell)$. Then $\mathcal{P}(k,\ell)$ is a $(4\ell-k+2)$-partite k-graph which satisfies (2) and (3).

For each $1 \leqslant i \leqslant 2\ell-k+2$ let e_i be the ith edge of $\mathcal{W}(k,\ell)$ as in the definition of $\mathcal{W}(k,\ell)$. Then for each $1 \leqslant i \leqslant 2\ell-k+1$ we know that $|e_i \cap e_{i+1}| = k-2$, and so we may choose S_i to be an ordered set of ℓ vertices chosen from $\bigcup_{w\in e_i\cap e_{i+1}} V_w$. Also, let S_0 and $S_{2\ell-k+2}$ be ordered sets of ℓ vertices chosen from the V_w corresponding to the ordered ends of $\mathcal{W}(k,\ell)$. So S_0 and $S_{2\ell-k+2}$ are subsets of $\bigcup_{w\in e_1} V_w$, and $\bigcup_{w\in e_{2\ell-k+2}} V_w$ respectively. We choose these sets S_i to be disjoint and to contain at most one vertex from any one vertex class V_w. Then by Proposition 5.1, for each $1 \leqslant i \leqslant 2\ell-k+2$ we can find an ℓ-path from S_{i-1} to S_i in $\mathcal{K}[V_w : w \in e_i]$ which contains $k^2\ell(k-\ell)+k$ vertices. We do this so that the ℓ-paths chosen only intersect in the appropriate S_i. Then the union of all of these ℓ-paths is an ℓ-path P from S_0 to $S_{2\ell-k+2}$ with

$$k^2\ell \leqslant k^2\ell(k-\ell)+k \leqslant |P| \leqslant (2\ell-k+2)(k^2\ell(k-\ell)+k) \leqslant 2k^5.$$

In the same way we find another 4ℓ ℓ-paths from S_0 to $S_{2\ell-k+2}$, so that all $4\ell+1$ of the ℓ-paths obtained are internally disjoint. Then $\mathcal{P}(k,\ell)$ is the union of all of these ℓ-paths. $\square$

Fix any such $\mathcal{P}(k,\ell)$, which we henceforth refer to as $\mathcal{P}(k,\ell)$. Also, let S_1 and S_2 be the ordered ends of $\mathcal{P}(k,\ell)$, so that S_1 and S_2 are disjoint ordered ℓ-sets. Let $\mathcal{S}(k,\ell)$ be the complex with vertex set $S_1 \cup S_2$ and with edges being all subsets of S_1 and all subsets of S_2. Then since each of the ℓ-paths which form $\mathcal{P}(k,\ell)$ contain at least $k^2\ell$ vertices, the complex $\mathcal{S}(k,\ell)$ is an induced subcomplex of the complex $\mathcal{P}(k,\ell)^{\leqslant}$ corresponding to $\mathcal{P}(k,\ell)$, so under appropriate circumstances we may use the extension lemma (Lemma 5.7) to extend $\mathcal{S}(k,\ell)$ to $\mathcal{P}(k,\ell)$. This is the key to the following lemma, which states that for the values of k and ℓ considered, almost all pairs of ordered ℓ-sets of vertices of a sufficiently large k-graph of large minimum degree form the ordered ends of a copy of $\mathcal{P}(k,\ell)$.

Lemma 5.10 *Suppose that $k \geqslant 3$, that $k/2 \leqslant \ell \leqslant k-1$ is such that $(k-\ell) \nmid k$, and that $1/n \ll \beta \ll \mu, 1/k$. Let $\mathcal{H}$ be a k-graph of order n with $\delta(\mathcal{H}) \geqslant \mu n$. Then there are at most $\beta n^{2\ell}$ pairs of ordered ℓ-sets S_1 and S_2 of vertices of $\mathcal{H}$ for which $\mathcal{H}$ does not contain a copy of $\mathcal{P}(k,\ell)$ with ordered ends S_1 and S_2.*

Proof. To prove this, we use hypergraph regularity. So introduce new constants

$$\frac{1}{n} \ll \frac{1}{r}, \delta \ll c \ll \min\{\delta_k, \frac{1}{t}\} \leqslant \delta_k, \eta \ll d \ll \gamma \ll \beta \ll \mu.$$

We may assume that $t!$ divides $|\mathcal{H}|$, so apply the regularity lemma to $\mathcal{H}$, and let $V_1, \dots, V_{a_1}$ be the clusters of the partition obtained. As in Section 5.3.3, we say that an edge of $\mathcal{H}$ is useful if it lies in $\mathcal{K}_k(\widehat{P}^{(k-1)})$ such that $\mathcal{H}$ is (d', δ_k, r)-regular with respect to $\widehat{P}^{(k-1)}$ for some $d' \geqslant d$. Let $\mathcal{H}'$ be the subgraph of $\mathcal{H}$ consisting of all useful edges. Note that no edge of $\mathcal{H}'$ contains 2 vertices from the same cluster. Then by Lemma 5.5, at most $2dn^k$ edges of $\mathcal{H}$ are not useful, and so $d_{\mathcal{H}'}(S) \geqslant \mu n/2$ for all but at most γn^{k-1} of the $(k-1)$-sets S of vertices of $\mathcal{H}'$.

Let C_1 and C_2 be cells of the partition $\mathcal{P}^{(\ell)}$ obtained from the regularity lemma. We say that C_1 and C_2 are *connected* if $\mathcal{H}'$ contains a copy $\mathcal{W}$ of $\mathcal{W}(k,\ell)$ with ordered ends A and B such that

$A \in C_1$, $B \in C_2$, and such that no two vertices of $\mathcal{W}$ lie in the same cluster. We first show that there are at most $\beta n^{2\ell}/2$ pairs A and B of ordered ℓ-sets of vertices of $\mathcal{H}$ such that either

(i) at least one of A and B does not lie in a cell of $\mathcal{P}^{(\ell)}$, or

(ii) the cells C_A and C_B of $\mathcal{P}^{(\ell)}$ which contain A and B respectively are not connected.

Indeed, for (i) note that at most $\ell^2 \frac{n}{a_1} n^{\ell-1} \leqslant \ell^2 \eta n^\ell$ ordered ℓ-sets of vertices of $\mathcal{H}$ do not lie in Cross_ℓ, and so there are at most $\ell^2 \eta n^{2\ell}$ pairs A and B of ordered ℓ-sets of vertices of $\mathcal{H}$ such that at least one of A and B does not lie in a cell of $\mathcal{P}^{(\ell)}$. Similarly, for (ii) note that there are at most $\ell^2 \eta n^{2\ell}$ pairs A and B of ordered ℓ-sets such that the cells C_A and C_B of $\mathcal{P}^{(\ell)}$ which contain A and B respectively share at least one cluster. Finally, if the cells C_A and C_B of $\mathcal{P}^{(\ell)}$ which contain A and B respectively do not share any clusters, but are not connected, then $\mathcal{H}'$ must contain fewer than $\binom{4\ell-k+2}{2}\eta n^{2\ell-k+2}$ copies of $\mathcal{W}(k,\ell)$ with ordered ends A and B. So by Proposition 5.8, there are at most $\beta n^{2\ell}/3$ pairs A and B of ordered ℓ-sets of vertices of $\mathcal{H}$ which lie in such pairs of cells of $\mathcal{P}^{(\ell)}$.

To prove the lemma, it is therefore sufficient to show that there are at most $\beta n^{2\ell}/2$ pairs S_1, S_2 of ordered ℓ-sets of vertices of $\mathcal{H}$ such that C_{S_1} and C_{S_2} are connected cells of $\mathcal{P}^{(\ell)}$ but S_1 and S_2 do not form the ordered ends of a copy of $\mathcal{P}(k,\ell)$ in $\mathcal{H}$. So suppose cells C_1 and C_2 of $\mathcal{P}(k,\ell)$ are connected. Then there is a copy $\mathcal{W}$ of $\mathcal{W}(k,\ell)$ in $\mathcal{H}'$ with ordered ends A and B such that $A \in C_1$, $B \in C_2$, and such that no two vertices of $\mathcal{W}$ lie in the same cluster. Since every edge of $\mathcal{H}'$ is a useful edge, for each edge $e \in E(\mathcal{W})$ the polyad $\widehat{P}^{(k-1)}(e)$ of e is such that $\mathcal{H}$ is (d', δ_k, r)-regular with respect to $\widehat{P}^{(k-1)}(e)$ for some $d' \geqslant d$. Then these polyads 'fit together'. By this we mean that if edges e and e' of $\mathcal{W}$ intersect in q vertices, then

$$\left(\bigcup_{i=1}^{k-1} \widehat{P}^{(i)}(e)\right) \cap \left(\bigcup_{i=1}^{k-1} \widehat{P}^{(i)}(e')\right) = \bigcup_{i=1}^{q} \widehat{P}^{(i)}(e \cap e'),$$

i.e. the intersection of the $(k-1,k)$-complexes corresponding to e and e' is the (q,q)-complex corresponding to $e \cap e'$. Therefore we can define $\mathcal{H}^*$ to be the $(k, 4\ell - k + 2)$-complex obtained from the $(k-1, 4\ell - k + 2)$-complex $\bigcup_{e \in E(\mathcal{W})} \bigcup_{i=1}^{k-1} \widehat{P}^{(i)}(e)$ by adding $E(\mathcal{H}) \cap \bigcup_{e \in E(\mathcal{W})} \mathcal{K}(\widehat{P}^{(k-1)}(e))$ as the 'kth level'. Then $\mathcal{H}^*$ is a $(\mathbf{d}, \delta_k, \delta, r)$-regular $(k, 4\ell - k + 2)$-complex, where $\mathbf{d} = (d', d_{k-1}, \ldots, d_2)$ and $(d_{k-1}, \ldots, d_2)$ is as in the definition of an (η, δ, t)-equitable family of partitions. (Here we may assume a common density d' for the kth level by applying the slicing lemma ([41], Proposition 22) if necessary.) Furthermore, by construction $\mathcal{H}^*$ respects the partition of the complex $\mathcal{W}(k,\ell)^{\leqslant}$ corresponding to $\mathcal{W}(k,\ell)$, and so property (3) of Proposition 5.9 implies that $\mathcal{H}^*$ also respects the partition of $\mathcal{P}(k,\ell)^{\leqslant}$. Let S_1 and S_2 be disjoint ordered ℓ-sets lying in the cells C_1 and C_2 of $\mathcal{P}^{(\ell)}$ respectively. Then $S_1 \cup S_2$ is the vertex set of a labelled copy $\mathcal{S}$ of $\mathcal{S}(k,\ell)$ in $\mathcal{H}^*$. So by Lemma 5.7, for all but at most $\beta |C_1||C_2|/2$ choices of $S_1 \in C_1$ and $S_2 \in C_2$ we can extend the labelled complex $\mathcal{S}$ to at least one labelled partition-respecting copy of $\mathcal{P}(k,\ell)$ with ordered ends S_1 and S_2. Summing over all C_1 and C_2, we find that there are at most $\beta n^{2\ell}/2$ ordered ℓ-sets S_1 and S_2 of vertices of $\mathcal{H}$ which lie in connected cells of $\mathcal{P}^{(\ell)}$ and which cannot be extended to a labelled partition-respecting copy of $\mathcal{P}(k,\ell)$, completing the proof. $\square$

We can now prove the following corollary, the diameter lemma we were aiming for. The idea behind this is that if S and T are ordered ℓ-sets in a large k-graph $\mathcal{H}$ of large minimum degree, then there are many ordered ℓ-sets S' and T' such that $\mathcal{H}$ contains ℓ-paths from S to S' and T to T'. So by the previous lemma, at least one such pair S' and T' form the ordered ends of a copy of $\mathcal{P}(k,\ell)$. Combining these ℓ-paths gives an ℓ-path from S to T.

Corollary 5.11 (Diameter lemma) *Suppose that $k \geqslant 3$, that $1 \leqslant \ell \leqslant k-1$ is such that $(k-\ell) \nmid k$, and that $1/n \ll \mu, 1/k$. Let $\mathcal{H}$ be a k-graph of order n with $\delta(\mathcal{H}) \geqslant \mu n$. Then for any two disjoint ordered ℓ-sets S and T of vertices of $\mathcal{H}$, there exists an ℓ-path P in $\mathcal{H}$ from S to T such that P contains at most $8k^5$ vertices.*

Proof. Recall that if $\ell < k/2$ we can find such an ℓ-path consisting of just one single edge, so we may assume that $\ell \geqslant k/2$. Introduce constants β, β' such that $1/n \ll \beta' \ll \beta \ll \mu, 1/k$. Let A be an arbitrary ordered ℓ-set of vertices of $\mathcal{H}$, and let X be an arbitrary set of 3ℓ vertices which is disjoint from A. We begin by showing that there are many ordered ℓ-sets B such that $\mathcal{H}$ contains an ℓ-path P from A to B having at most 3ℓ vertices, none of which are from X. To show this, we demonstrate how a vertex sequence of P may be chosen, and then count the number of choices.

Since A is to be an ordered end of P, we begin the vertex sequence of P with the ordered ℓ-set A. We then arbitrarily choose any $k - \ell - 1$ vertices of $\mathcal{H}$ to add to the sequence. To finish the sequence, we repeatedly make use of the fact that $\delta(\mathcal{H}) \geqslant \mu n$. More precisely, we repeat the following step: let V be the set of the final $k-1$ vertices of the current vertex sequence. Then $d_{\mathcal{H}}(V) \geqslant \mu n$, and so there are at least $\mu n - 6\ell$ vertices which together with V form an edge of $\mathcal{H}$ and which are not in the vertex sequence constructed thus far or in X. Choose v to be one of these vertices, and append it to the vertex sequence. We stop as soon as the number r of vertices in the sequence satisfies $r > 2\ell$ and $r \equiv k$ (modulo $(k - \ell)$), so in particular $r \leqslant 3\ell$. Let B be the ordered set consisting of the last ℓ vertices of the sequence. Then $\mathcal{H}$ contains an ℓ-path P with this vertex sequence, and P is therefore an ℓ-path of order at most 3ℓ from A to B which does not contain any vertex of X. There are at least $(\mu n - 6\ell)^{r-\ell}$ vertex sequences we could have chosen, and hence there are at least $(\mu n - 6\ell)^{r-\ell}/n^{r-2\ell} > \beta n^{\ell}$ possibilities for an ordered ℓ-set B such that there is an ℓ-path from A to B in $\mathcal{H}$, not containing any vertex of X.

Now, let S and T be the two ordered ℓ-sets of vertices of $\mathcal{H}$ given in the statement of the corollary. Then there are at least βn^{ℓ} ordered ℓ-sets S' of vertices of $\mathcal{H}$ such that there exists an ℓ-path P_1 from S to S' in $\mathcal{H}$, which contains at most 3ℓ vertices and such that $V(P_1) \cap T = \emptyset$. Likewise for each such choice of S' and P_1, there are at least βn^{ℓ} ordered ℓ-sets T' of vertices of $\mathcal{H}$ such that there exists an ℓ-path P_2 from T to T' of order at most 3ℓ in $\mathcal{H}$ and such that

$V(P_2) \cap V(P_1) = \emptyset$. By Lemma 5.10, at most $\beta' n^{2\ell}$ of these pairs S', T' do not form ordered ends of a copy of $\mathcal{P}(k,\ell)$ in $\mathcal{H}$. Since $\beta' \ll \beta$ we may therefore choose such a pair S', T' such that S' and T' are ordered ends of a copy of $\mathcal{P}(k,\ell)$ in $\mathcal{H}$. Then there are at least $4\ell+1$ internally disjoint ℓ-paths of order at most $2k^5$ from S' to T' in $\mathcal{H}$. At most 4ℓ of these ℓ-paths contain any vertex from $V(P_1) \setminus S'$ or $V(P_2) \setminus T'$, and so we may choose an ℓ-path Q from S' to T' in $\mathcal{P}(k,\ell) \subseteq \mathcal{H}$ of order at most $2k^5$ which contains no vertex from $V(P_1) \setminus S'$ or $V(P_2) \setminus T'$. Then $P_1 Q P_2$ is the ℓ-path from S to T of order at most $2k^5 + 6\ell \leqslant 8k^5$ we seek. $\square$

5.5 The absorbing path lemma

Let $\mathcal{H}$ be a k-graph, and let S be a set of $k-\ell$ vertices of $\mathcal{H}$. Recall that an ℓ-path P in $\mathcal{H}$ with ordered ends P^{beg} and P^{end} is *absorbing* for S if P does not contain any vertex of S, and $\mathcal{H}$ contains an ℓ-path Q with the same ordered ends P^{beg} and P^{end}, where $V(Q) = V(P) \cup S$. This means that if P is a section of an ℓ-path P^* which does not contain any vertices of S, then we can 'absorb' the vertices of S into P^* by replacing P with Q. P^* is still an ℓ-path after this change as P and Q have the same ordered ends. Similarly, we say that an ℓ-path P in $\mathcal{H}$ with ordered ends P^{beg} and P^{end} can *absorb* a collection $S_1, \dots, S_r$ of $(k-\ell)$-sets of vertices of $\mathcal{H}$ if P does not contain any vertex of $\bigcup_{i=1}^r S_i$, and $\mathcal{H}$ contains an ℓ-path Q with the same ordered ends P^{beg} and P^{end}, where $V(Q) = V(P) \cup \bigcup_{i=1}^r S_i$. The reason we absorb $(k-\ell)$-sets is that the order of an ℓ-path must be congruent to k, modulo $k-\ell$. The next result describes the absorbing path as a k-graph, which we use to absorb a set S. Note that the proof of this proposition uses Proposition 5.1. As a consequence, this proposition and each of the remaining results of this section, including the absorbing path lemma, require that $(k-\ell) \nmid k$.

Proposition 5.12 *Suppose that $k \geqslant 3$, and that $1 \leqslant \ell \leqslant k-1$ is such that $(k-\ell) \nmid k$. Then there is a k-partite k-graph $\mathcal{AP}(k,\ell)$ with the following properties.*

(1) $|\mathcal{AP}(k,\ell)| \leqslant k^4$.

(2) The vertex set of $\mathcal{AP}(k,\ell)$ *consists of two disjoint sets* S *and* X *with* $|S| = k - \ell$.

(3) $\mathcal{AP}(k,\ell)$ *contains an* ℓ*-path* P *with vertex set* X *and ordered ends* P^{beg} *and* P^{end}.

(4) $\mathcal{AP}(k,\ell)$ *contains an* ℓ*-path* Q *with vertex set* $S \cup X$ *and ordered ends* P^{beg} *and* P^{end}.

(5) No edge of $\mathcal{AP}(k,\ell)$ *contains more than one vertex of* S.

(6) No vertex class of $\mathcal{AP}(k,\ell)$ *contains more than one vertex of* S.

Proof. Let $V_1, \dots, V_k$ be disjoint vertex sets of size $k\ell(k-\ell)+1$. Let S be an ordered $(k-\ell)$-set such that for each $1 \leqslant i \leqslant k-\ell$, the ith vertex of S lies in $V_{\ell+i}$. Let P be an ℓ-path in $\mathcal{K}[V_1, \dots, V_k]$ with ordered ends P^{beg} and P^{end} such that both P^{beg} and P^{end} contain at most one vertex from each V_i and such that $V(P) = (V_1 \cup \dots \cup V_k) \setminus S$. (One can easily choose such a P if for all $j = 1, \dots, |P|$ one chooses the jth vertex of P in the V_i for which $j \equiv i$ modulo k.) Then $V(P) \cup S = V_1 \cup \dots \cup V_k$. Thus we can apply Proposition 5.1 to obtain an ℓ-path Q from P^{beg} to P^{end} in $\mathcal{K}[V_1, \dots, V_k]$ such that $V(Q) = V(P) \cup S$. By swapping some vertices in S with some vertices in $V(Q) \setminus S$ (lying in the same V_i) if necessary we can ensure that the vertices in S are distributed in such a way that in some vertex sequence of Q they have distance at least k from each other. (This ensures (5).) We can now take $\mathcal{AP}(k,\ell) := P \cup Q$. $\square$

Fix an $\mathcal{AP}(k,\ell)$ satisfying Proposition 5.12, which we refer to simply as $\mathcal{AP}(k,\ell)$ for the rest of this section. Let $b(k,\ell) := |\mathcal{AP}(k,\ell)| - k + \ell$, so that $b(k,\ell)$ is the number of vertices of the ℓ-path P in the definition of $\mathcal{AP}(k,\ell)$.

Given a $(k-\ell)$-set S of vertices of $\mathcal{H}$, we think of S as a labelled (k,k)-complex with no i-edges for any $i \geqslant 2$. Then the extension lemma (Lemma 5.7) tells us that for most such $(k-\ell)$-sets S there are many labelled copies of $\mathcal{AP}(k,\ell)^{\leqslant}$ extending S in $\mathcal{H}$, and so $\mathcal{H}$ contains many absorbing paths for these sets S.

Suppose that $\mathcal{H}$ is a k-graph on n vertices, and that c is a positive constant. We say that a $(k-\ell)$-set S of vertices of $\mathcal{H}$ is *c-good* if $\mathcal{H}$ contains at least $cn^{b(k,\ell)}$ absorbing paths for S, each on $b(k,\ell)$ vertices. S is *c-bad* if it is not c-good. The next lemma states that for the values of k and ℓ we are interested in, and any small c, if $\mathcal{H}$ is sufficiently large and has large minimum degree, then almost all $(k-\ell)$-sets S of vertices of $\mathcal{H}$ are c-good.

Lemma 5.13 *Suppose that $k \geqslant 3$, that $1 \leqslant \ell \leqslant k-1$ is such that $(k-\ell) \nmid k$, and that $1/n \ll c \ll \gamma \ll \mu, 1/k$. Let $\mathcal{H}$ be a k-graph on n vertices such that $\delta(\mathcal{H}) \geqslant \mu n$. Then at most $\gamma n^{k-\ell}$ sets S of $k-\ell$ vertices of $\mathcal{H}$ are c-bad.*

Proof. Let $b = b(k,\ell)$, and introduce new constants

$$\frac{1}{n} \ll \frac{1}{r}, \delta \ll c \ll \min\{\delta_k, \frac{1}{t}\} \leqslant \delta_k, \eta \ll d \ll \gamma.$$

We may assume that $t!$ divides $|\mathcal{H}|$, so apply the regularity lemma to $\mathcal{H}$, and let $V_1, \dots, V_{a_1}$ be the clusters of the partition obtained. Let $m = n/a_1$ be the size of each of these clusters. Form the reduced k-graph $\mathcal{R}$ on these clusters as defined in Section 5.3.3.

We begin by showing that almost all sets of $k-\ell$ vertices of $\mathcal{H}$ are contained in clusters which lie in a common edge of $\mathcal{R}$. More precisely, for all but at most $\gamma n^{k-\ell}/2$ sets $\{v_1, \dots, v_{k-\ell}\}$ of $k-\ell$ vertices of $\mathcal{H}$ we can choose clusters $V_{i_1}, \dots, V_{i_k}$ such that $v_j \in V_{i_j}$ for each $1 \leqslant j \leqslant k-\ell$ and such that $\{V_{i_1}, \dots, V_{i_k}\}$ forms an edge of $\mathcal{R}$. Indeed, by Lemma 5.4, $d_{\mathcal{R}}(S) \geqslant 1$ for all but at most $\gamma a_1^{k-\ell}/3$ 'neighbourless' sets S of $k-\ell$ clusters. At most $\eta n^{k-\ell} \ll \gamma n^{k-\ell}$ sets T of $k-\ell$ vertices of $\mathcal{H}$ do not lie in $\mathrm{Cross}_{k-\ell}$. But if $T \in \mathrm{Cross}_{k-\ell}$, then unless the set S of clusters containing the vertices of T is one of the at most $\gamma a_1^{k-\ell}/3$ 'neighbourless' sets of $k-\ell$ clusters (which is the case for at most $\gamma n^{k-\ell}/3$ sets of $k-\ell$ vertices of $\mathcal{H}$), there is an edge of $\mathcal{R}$ containing S as required.

Now, suppose that $V_{i_1}, \dots, V_{i_k}$ are clusters which form an edge of $\mathcal{R}$. Note that there are $m^{k-\ell}$ sets $\{v_1, \dots, v_{k-\ell}\}$ such that $v_j \in V_{i_j}$ for each $1 \leqslant j \leqslant k-\ell$. Since $e = \{V_{i_1}, \dots, V_{i_k}\}$ is an edge of $\mathcal{R}$, we may define the complex $\mathcal{H}^*$ corresponding to e as in the paragraph after the statement of the extension lemma (Lemma 5.7). Then $\mathcal{H}^*$ satisfies the conditions of the extension lemma (with $\gamma/2$ and k playing the roles of β and s), and respects the partition of $\mathcal{AP}(k,\ell)$. Let S be an ordered set of size $k-\ell$, which we may view as a labelled (k,k)-complex with no j-edges for $j \geqslant 2$. Then by Lemma 5.7, all but at most $\gamma m^{k-\ell}/2$ ordered sets $S' = \{v_1, \dots, v_{k-\ell}\}$ such that $v_j \in V_{i_j}$ for each j (these are the labelled copies of S) are extendible to at least $cb(k,\ell)!n^{b(k,\ell)}$ labelled partition-respecting copies of $\mathcal{AP}(k,\ell)$ in $\mathcal{H}$. This is where we use property (5) of Proposition 5.12 – it ensures that the complex S is an induced subcomplex of $\mathcal{AP}(k,\ell)$. For each copy C of $\mathcal{AP}(k,\ell)$, $C - S'$ is an absorbing path for S' on $b(k,\ell)$ vertices, and so $\mathcal{H}^*$ (and therefore $\mathcal{H}$) contains at least $cn^{b(k,\ell)}$ absorbing paths on $b(k,\ell)$ vertices for S'. So at most $\gamma m^{k-\ell}/2$ such sets S' are c-bad.

Recall that the number of $(k-\ell)$-sets of vertices of $\mathcal{H}$ which do not lie in distinct clusters corresponding to an edge of $\mathcal{R}$ is at most $\gamma n^{k-\ell}/2$. Summing over all sets of $k-\ell$ clusters, we see that at most $\gamma n^{k-\ell}/2$ of the $(k-\ell)$-sets which do lie in distinct clusters corresponding to an edge of $\mathcal{R}$ are c-bad. Thus at most $\gamma n^{k-\ell}$ sets of $k-\ell$ vertices of $\mathcal{H}$ are c-bad, completing the proof. $\square$

We are now in a position to prove the main lemma of this section. It states that for any positive c, if $\mathcal{H}$ is a sufficiently large k-graph of large minimum degree, then we can find an ℓ-path in $\mathcal{H}$ which contains a small proportion of the vertices of $\mathcal{H}$, includes all vertices of $\mathcal{H}$ which lie in many c-bad $(k-\ell)$-sets and can absorb any small collection of c-good $(k-\ell)$-sets of vertices of $\mathcal{H}$.

Lemma 5.14 (Absorbing path lemma) *Suppose that $k \geqslant 3$, that $1 \leqslant \ell \leqslant k-1$ is such that $(k-\ell) \nmid k$, and that $1/n \ll \alpha \ll c \ll \gamma \ll \mu, 1/k$. Let $\mathcal{H}$ be a k-graph of order n with*

$\delta(\mathcal{H}) \geqslant \mu n$. *Then* $\mathcal{H}$ *contains an* ℓ*-path* P *on at most* μn *vertices such that the following properties hold:*

(1) Every vertex of $\mathcal{H} - V(P)$ *lies in at most* $\gamma n^{k-\ell-1}$ c*-bad* $(k-\ell)$*-sets.*

(2) P *can absorb any collection of at most* αn *disjoint* c*-good* $(k-\ell)$*-sets of vertices of* $\mathcal{H} - V(P)$.

Proof. Let $b := b(k, \ell)$, and choose a family $\mathcal{T}$ of ordered b-sets of vertices of $\mathcal{H}$ at random by including each ordered b-set T into $\mathcal{T}$ with probability $c^2 n^{1-b}$, independently of all other ordered b-sets. Now, for any c-good set S of $k - \ell$ vertices of $\mathcal{H}$, the expected number of $T \in \mathcal{T}$ for which $\mathcal{H}$ contains an absorbing path for S with T as a vertex sequence is at least $c^3 n$, by the definition of a c-good set. So by a standard Chernoff bound, with probability $1 - o(1)$, for every c-good $(k-\ell)$-set S of vertices of $\mathcal{H}$ the number of such ordered b-sets $T \in \mathcal{T}$ is at least $c^3 n/2$. Furthermore, with probability $1 - o(1)$ we have $|\mathcal{T}| \leqslant 2c^2 n$. The expected number of ordered pairs T, T' in $\mathcal{T}$ which intersect (*i.e.* for which the corresponding unordered sets intersect) is at most $(c^2 n^{1-b})^2 b^2 n^{2b-1} = c^4 b^2 n$. So with probability at least 1/2 the number of such pairs is at most $2c^4 b^2 n$. Thus we may fix an outcome of our random selection of $\mathcal{T}$ such that all of these events hold.

Delete from $\mathcal{T}$ every $T \in \mathcal{T}$ which intersects any other $T' \in \mathcal{T}$. Also delete from $\mathcal{T}$ every $T \in \mathcal{T}$ which is not a vertex sequence of an absorbing path for some c-good $(k-\ell)$-set S of vertices of $\mathcal{H}$. Let $T_1, \dots, T_q$ be the remaining members of $\mathcal{T}$. So $q \leqslant 2c^2 n$, and for each $1 \leqslant i \leqslant q$ we can choose an ℓ-path P_i in $\mathcal{H}$ with vertex sequence T_i which is absorbing for some such S. Then all the ℓ-paths P_i are disjoint, and for every c-good $(k-\ell)$-set S of vertices of $\mathcal{H}$ at least $c^3 n/2 - 2c^4 b^2 n \geqslant \alpha n$ of the ℓ-paths P_i are absorbing.

Let X be the set of vertices of $\mathcal{H}$ which are not contained in any P_i and which lie in more than $\gamma n^{k-\ell-1}$ c-bad $(k-\ell)$-sets. Then $|X| \leqslant \gamma n$, since by Lemma 5.13 there are at most $\gamma^2 n^{k-\ell}/(k-\ell)$ c-bad $(k-\ell)$-sets in total. We use the minimum degree condition on $\mathcal{H}$ to greedily construct an ℓ-path P_0 containing all vertices in X and not intersecting the previous paths P_i, $1 \leqslant i \leqslant q$. Then if we incorporate each of the P_i $(0 \leqslant i \leqslant q)$ into the ℓ-path P we are constructing, condition (1) of the lemma is satisfied. So let X' be a set of $k-1$ vertices of X. Then $d_{\mathcal{H}}(X') \geqslant \mu n$ by the minimum degree condition on $\mathcal{H}$. Since $|\bigcup_{i=1}^{q} P_i| < \mu n$, we may choose a vertex $x \in V(\mathcal{H}) \setminus \bigcup_{i=1}^{q} V(P_i)$ which together with X' forms an edge of $\mathcal{H}$. Then $X' \cup \{x\}$ is the first edge of P_0. We then greedily extend P_0 as follows. Let X'' be the set of the final ℓ vertices of the vertex sequence of P_0. Add to X'' any $k-1-\ell \geqslant 1$ vertices from X not yet contained in P_0. Then $d_{\mathcal{H}}(X'') \geqslant \mu n$, and so we may choose a vertex y of $\mathcal{H}$ which is not in $\bigcup_{i=1}^{q} P_i$ nor already contained in P_0. We then extend P_0 by the edge $X'' \cup \{y\}$. At the end of this process we obtain an ℓ-path P_0 which is disjoint from all the P_i $(i = 1, \dots q)$, which contains every vertex of X, and which satisfies $|V(P_0)| \leqslant 2\gamma n$. Let P_i^{beg} and P_i^{end} be ordered ends of P_i for each $0 \leqslant i \leqslant q$.

To complete the proof, we now use the diameter lemma (Corollary 5.11) to greedily join each ordered ℓ-set P_i^{end} to the ordered ℓ-set P_{i+1}^{beg} by an ℓ-path P_i', such that P_i' intersects P_i and P_{i+1} only in the sets P_{i+1}^{beg} and P_i^{end} and does not intersect any other P_j or any previously chosen P_j'. More precisely, suppose we have chosen such $P_0', \dots, P_{i-1}'$. Let $\mathcal{H}'$ be the k-graph obtained from $\mathcal{H}$ by removing all the vertices in $P_0, \dots, P_q$ and all the vertices in $P_0', \dots, P_{i-1}'$ and then adding back P_i^{end} and P_{i+1}^{beg}. Then $\delta(\mathcal{H}') \geqslant \mu n/2$, and so we may apply Corollary 5.11 to find an ℓ-path P_i' in $\mathcal{H}'$ from P_i^{end} to P_{i+1}^{beg} containing at most $8k^5$ vertices. Having found these ℓ-paths, the absorbing path P^* is the ℓ-path $P_0 P_0' P_1 P_1' P_2 \dots P_{q-1} P_{q-1}' P_q$.

To see that P^* satisfies condition (2) of the lemma, suppose we have a collection $S_1, \dots, S_r$ of at most αn disjoint c-good $(k-\ell)$-sets of vertices of $\mathcal{H}$. Each of these c-good sets S_i has at least

αn absorbing paths in the ℓ-path P^*. So for each $1 \leqslant i \leqslant r$ choose a unique absorbing path P_{j_i} for S_i in P^*. Then by the definition of an absorbing path we may absorb each set S_i into P_{j_i} to obtain an ℓ-path Q_{j_i} with vertex set $S_i \cup V(P_{j_i})$ and with the same ordered ends as P_{j_i}. Replacing each P_{j_i} by Q_{j_i} gives us an ℓ-path Q^* with vertex set $V(P^*) \cup \bigcup_{i=1}^{r} S_i$ and with the same ordered ends as P^*, as desired. □

5.6 The path cover lemma

In this section we prove the following lemma, which states that the vertices of a k-graph of large minimum degree can be almost covered by a bounded number of disjoint ℓ-paths.

Lemma 5.15 (Path cover lemma) *Suppose $k \geqslant 3$, that $1 \leqslant \ell \leqslant k-1$, and that $1/n \ll 1/D \ll \varepsilon \ll \mu, 1/k$. Let $\mathcal{H}$ be a k-graph of order n with $\delta(\mathcal{H}) \geqslant (\frac{1}{\lceil \frac{k}{k-\ell} \rceil (k-\ell)} + \mu)n$. Then $\mathcal{H}$ contains a set of at most D disjoint ℓ-paths covering all but at most εn vertices of $\mathcal{H}$.*

Note that the condition $(k-\ell) \nmid k$ is not needed for this lemma. Let

$$a := \left\lceil \frac{k}{k-\ell} \right\rceil (k-\ell) \tag{5.16}$$

and let $\mathcal{F}_{k,\ell}$ be the k-graph whose vertex set is the disjoint union of sets $A_1, \dots, A_{a-1}$ and B of size $k-1$ and whose edges are all the k-sets of the form $A_i \cup \{b\}$ (for all $i = 1, \dots, a-1$ and all $b \in B$). Recall that an $\mathcal{F}_{k,\ell}$*-packing* in a k-graph $\mathcal{R}$ is a collection of pairwise vertex-disjoint copies of $\mathcal{F}_{k,\ell}$ in $\mathcal{R}$.

The idea of the proof of the path cover lemma is to apply the regularity lemma to $\mathcal{H}$ in order to obtain a reduced k-graph $\mathcal{R}$. Recall that by Lemma 5.4 the minimum degree of $\mathcal{H}$ is almost inherited by $\mathcal{R}$. So we can use the following lemma (Lemma 5.17) to obtain an almost perfect

$\mathcal{F}_{k,\ell}$-packing in $\mathcal{R}$. Consider any copy $\mathcal{F}$ of $\mathcal{F}_{k,\ell}$ in this packing. We repeatedly apply the embedding lemma (Lemma 5.6) to the sub-k-graph $\mathcal{H}(\mathcal{F})$ of $\mathcal{H}$ corresponding to $\mathcal{F}$ to obtain a bounded number of ℓ-paths which cover almost all vertices of $\mathcal{H}(\mathcal{F})$. Doing this for all the copies of $\mathcal{F}_{k,\ell}$ in the $\mathcal{F}_{k,\ell}$-packing of $\mathcal{R}$ gives a set of ℓ-paths as required in Lemma 5.15.

Lemma 5.17 *Suppose that $k \geqslant 3$, that $1 \leqslant \ell \leqslant k-1$, and that $1/n \ll \theta \ll \varepsilon \ll 1/k$. Let $\mathcal{H}$ be a k-graph of order n such that $d(S) \geqslant n/a$ for all but at most θn^{k-1} sets $S \in \binom{V(\mathcal{H})}{k-1}$, where a is as defined in (5.16). Then $\mathcal{H}$ contains an $\mathcal{F}_{k,\ell}$-packing covering all but at most $(1-\varepsilon)n$ vertices.*

Note that $\mathcal{F}_{k,1}$ is the k-graph $\mathcal{A}_k$ considered in Chapter 4. So Lemma 4.10 is a special case of Lemma 5.17.

Proof. Let $\mathcal{F}^1, \dots, \mathcal{F}^t$ be a maximal $\mathcal{F}_{k,\ell}$-packing in $\mathcal{H}$. Let X be the set of vertices of $\mathcal{H}$ not covered by any of the $\mathcal{F}^i$, and suppose for a contradiction that $|X| > \varepsilon n$. For a $(k-1)$-set S of vertices of $\mathcal{H}$, we write $N(S)$ for $N_{\mathcal{H}}(S)$, $d(S)$ for $d_{\mathcal{H}}(S)$, $N_X(S)$ for $N(S) \cap X$ and $d_X(S)$ for $|N_X(S)|$. Since $\theta \ll \varepsilon$, we may greedily choose a collection $\mathcal{S} \subseteq \binom{X}{k-1}$ of size at least $2\theta n$ such that the members of $\mathcal{S}$ are pairwise disjoint and such that $d(S) \geqslant n/a$ for every $S \in \mathcal{S}$. We may therefore consider two cases, both leading to a contradiction.

Case 1. *There exists a collection $S_1, \dots, S_r$ of $r \geqslant \theta n$ disjoint $(k-1)$-subsets of X such that each S_i satisfies $d_X(S_i) \geqslant \varepsilon n/2a$.*

In this case, we count the pairs (i, B) for $1 \leqslant i \leqslant r$ and $B \in \binom{X}{k-1}$ such that $x \in N(S_i)$ for each $x \in B$. Then since each of the S_i satisfies $d_X(S_i) \geqslant \varepsilon n/2a$, the number of such pairs is at least $r\binom{\varepsilon n/2a}{k-1} \geqslant \theta n \binom{\varepsilon n/2a}{k-1}$. Since there are $\binom{|X|}{k-1}$ sets $B \in \binom{X}{k-1}$, at least one such B must lie in at least $(a-1)$ such pairs. But then the corresponding S_i together with this B form a copy of $\mathcal{F}_{k,\ell}$ contained in $\mathcal{H}[X]$, contradicting the maximality of our $\mathcal{F}_{k,\ell}$-packing.

Case 2. *There exists a collection $S_1, \dots, S_r$ of $r \geqslant \theta n$ disjoint $(k-1)$-subsets of X such that each S_i satisfies $d(S_i) \geqslant n/a$ and $d_X(S_i) < \varepsilon n/2a$.*

In this case, we say that $\mathcal{F}^i$ is *good* for S_j if $\mathcal{F}^i$ contains at least k vertices from $N(S_j)$. Note that $|F_{k,\ell}| = a(k-1)$. Each S_j has at least $\frac{n}{a} - \frac{\varepsilon n}{2a}$ neighbours in $\bigcup_{i=1}^{t} V(\mathcal{F}^i)$, and at most $(k-1)\frac{(1-\varepsilon)n}{a(k-1)}$ of these neighbours lie in copies $\mathcal{F}^i$ which are not good for S_j. Thus for each S_j at least $\frac{\varepsilon n}{2a^2(k-1)}$ of the $\mathcal{F}^i$ must be good for S_j. We next count the number of pairs $(j, \mathcal{T})$, where $1 \leqslant j \leqslant r$ and $\mathcal{T}$ is a collection of $(k-1)$ copies $\mathcal{F}^i$ which are good for S_j. This number must be at least $r\binom{\frac{\varepsilon n}{2a^2(k-1)}}{k-1} \geqslant \sqrt{n}\binom{t}{k-1}$, and so there must be some such $\mathcal{T}$ and some $R \subseteq [r]$ with $|R| \geqslant \sqrt{n}$ such that for every $j \in R$ and every $\mathcal{F}^i$ in $\mathcal{T}$, $\mathcal{F}^i$ is good for S_j. For each $j \in R$ and each $\mathcal{F}^i \in \mathcal{T}$, fix $K^{i,j}$ to be a subset of $N(S_j) \cap \mathcal{F}^i$ of size k. Then for some $R' \subseteq R$ of size $(a-1)k$ we have that for any fixed i, $K^{i,j}$ is the same set (denoted K^i), for every $j \in R'$. Arbitrarily partition R' into k sets $R'_1, \dots, R'_k$ of size $(a-1)$, and label the vertices of each K^i to be $\{v_i^1, v_i^2, \dots, v_i^k\}$. Then for each $1 \leqslant s \leqslant k$, the sets S_j for $j \in R'_s$ and the set $\{v_1^s, \dots, v_{k-1}^s\}$ form a copy of $\mathcal{F}_{k,\ell}$, and these copies are mutually disjoint and are contained in $X \cup \bigcup_{\mathcal{F}^i \in \mathcal{T}} V(\mathcal{F}^i)$, contradicting the maximality of our $\mathcal{F}_{k,\ell}$-packing. $\square$

Lemma 5.18 *Let P be an ℓ-path on n vertices and let a be as defined in (5.16). Then there is a k-colouring of P with colours $1, \dots, k$ such that colour k is used $n/a \pm 1$ times and the sizes of all other colour classes are as equal as possible.*

Proof. Let $x_1, \dots, x_n$ be a vertex sequence of P. Colour vertices $x_k, x_{k+a}, x_{k+2a}, \dots$ with colour k and remove these vertices from the sequence $x_1, \dots, x_n$. Colour the remaining vertices in turn with colours $1, \dots, k-1$ as follows. Colour the first vertex with colour 1. Suppose that we just coloured the ith vertex with some colour j. Then we colour the next vertex with colour $j+1$ if $j \leqslant k-2$ and with colour 1 if $j = k-1$. To show that this yields a proper colouring, it suffices to show that every edge of P contains some vertex of colour k. Clearly this holds for the first edge e_1 of P and for all edges intersecting e_1 (since x_k lies in all those edges). Note that the first vertex of the ith edge e_i of P is $x_{f(i)}$, where $f(i) = (i-1)(k-\ell)+1$. Also note that $i^* := \lceil \frac{k}{k-\ell} \rceil + 1$ is the smallest integer so that $f(i^*) > k$.In other words, the

i^*th edge e_{i^*} of P is the first edge which does not contain x_k. But the vertices of e_{i^*} are $x_{a+1}, \dots, x_{a+k}$. So e_{i^*} as well as all succeeding edges which intersect e_{i^*} contain a vertex of colour k (namely x_{a+k}). Continuing in this way gives the claim. Indeed, colour k is used $\lceil \frac{n-(k-1)}{a} \rceil$ times and $\lceil \frac{n-(k-1)}{a} \rceil \geqslant \lceil \frac{n}{a} - \frac{k-1}{\frac{k}{k-\ell}(k-\ell)} \rceil = \lceil \frac{n}{a} + \frac{1}{k} - 1 \rceil \geqslant \frac{n}{a} - 1$. Also $\lceil \frac{n-(k-1)}{a} \rceil \leqslant \lceil \frac{n}{a} \rceil \leqslant \frac{n}{a} + 1$. □

Proof of Lemma 5.15. Choose new constants such that

$$\frac{1}{n} \ll \frac{1}{D} \ll \frac{1}{r}, \delta, c \ll \min\{\delta_k, \frac{1}{t}\} \leqslant \delta_k, \eta \ll d \ll \theta \ll \varepsilon.$$

We may assume that $t!|n$, so apply Theorem 5.3 (the regularity lemma) to $\mathcal{H}$, and let $V_1, \dots, V_{a_1}$ be the clusters of the partition obtained. Let $m = n/a_1$ be the size of each of these clusters. Form the reduced k-graph $\mathcal{R}$ on these clusters as discussed in Section 5.3.3. Lemmas 5.4 and 5.17 together imply that $\mathcal{R}$ has an $\mathcal{F}_{k,\ell}$-packing $\mathcal{A}$ covering all but at most $\varepsilon n/2$ vertices of $\mathcal{R}$. Consider any copy $\mathcal{F}$ of $\mathcal{F}_{k,\ell}$ in this packing. Our aim is to cover almost all vertices in the clusters belonging to $\mathcal{F}$ by a bounded number of disjoint ℓ-paths.

So let $A_1, \dots, A_{a-1}$ and B be $(k-1)$-element subsets of $V(\mathcal{F})$ as in the definition of $\mathcal{F}_{k,\ell}$. So the edges of $\mathcal{F}_{k,\ell}$ are all the k-tuples of the form $A_i \cup \{b\}$ for all $i = 1, \dots, a-1$ and all $b \in B$. Pick $b \in B$ and consider the edge $A_1 \cup \{b\} =: e$. Let $\mathcal{V}$ be the set of all clusters corresponding to vertices in A_1 and let V_b be the cluster corresponding to b. Define the complex $\mathcal{H}^*$ corresponding to the edge e as in the paragraph after the statement of the extension lemma (Lemma 5.7). Then Lemma 5.18 and the embedding lemma (Lemma 5.6 applied to $\mathcal{H}^*$) together imply that the sub-k-graph of $\mathcal{H}$ spanned by the vertices in $V_b \cup \bigcup_{V \in \mathcal{V}} V$ contains an ℓ-path P_1 on $acm/(a-1)$ vertices which intersects each cluster from $\mathcal{V}$ in $cm/(k-1) \pm 1$ vertices and V_b in $cm/(a-1) \pm 1$ vertices.Lemma 5.2 implies that the subcomplex of $\mathcal{H}^*$ obtained by deleting the vertices of P_1 is still $(\mathbf{d}, \sqrt{\delta_k}, \sqrt{\delta}, r)$-regular. So we can find another ℓ-path P_2 which is disjoint from P_1 and intersects each cluster from $\mathcal{V}$ in $cm/(k-1) \pm 1$ vertices and V_b in $cm/(a-1) \pm 1$ vertices. We do this until we have used about $m/(k-1)$ vertices in each cluster from $\mathcal{V}$. So we have found $1/c$ disjoint ℓ-paths. Now we pick $b' \in B \setminus \{b\}$ and argue as before to get $1/c$ disjoint ℓ-paths, such that each of them intersects (the remainder of) each cluster from $\mathcal{V}$ in $cm/(k-1) \pm 1$ vertices and $V_{b'}$ in $cm/(a-1) \pm 1$ vertices. We do this for all the $k-1$ vertices in B. However, when considering the last vertex b'' of B, we stop as soon as one of the subclusters from $\mathcal{V}$ has size less than $\varepsilon m/4a$ (and thus all the other subclusters from $\mathcal{V}$ have size at most $\varepsilon m/2a$) since we need to ensure that the subcomplex of $\mathcal{H}^*$ restricted to the remaining subclusters is still $(\mathbf{d}, \sqrt{\delta_k}, \sqrt{\delta}, r)$-regular. So in total we have chosen close to $(k-1)/c$ disjoint ℓ-paths covering all but at most $\varepsilon m/2a$ vertices in each cluster from $\mathcal{V}$ and covering between $m/(a-1) - \varepsilon m/2a$ and $m/(a-1)$ vertices in each cluster V_b with $b \in B$. We now repeat this process for each of $A_2, \ldots, A_{a-1}$ in turn. When considering the final set A_{a-1}, we also stop choosing paths for some $b \in B$ if the subcluster V_b has size less than $\varepsilon m/4a$. Altogether this gives us a collection of close to $(k-1)(a-1)/c$ disjoint ℓ-paths covering all but at most $\varepsilon m/2$ vertices in the clusters belonging to $\mathcal{F}$. Doing this for all the copies of $\mathcal{F}_{k,\ell}$ in the $\mathcal{F}_{k,\ell}$-packing $\mathcal{A}$ of $\mathcal{R}$ we obtain a collection of at most $|\mathcal{A}|(k-1)(a-1)/c \ll D$ disjoint ℓ-paths covering all but at most $\varepsilon m/2$ vertices from each cluster, and hence all but at most $\varepsilon|\mathcal{H}|$ vertices of $\mathcal{H}$, as required. □

5.7 Proof of Theorem 1.11

We use the following two results in our proof of Theorem 1.11. The first says that if $1 \leqslant s \leqslant k-1$ and $\mathcal{H}$ is a large k-graph in which all sets of s vertices have a large neighbourhood, then if we choose $R \subseteq V(\mathcal{H})$ uniformly at random, with high probability all sets of s vertices have a large neighbourhood in R.

Lemma 5.19 (Reservoir Lemma) *Suppose that $k \geqslant 2$, that $1 \leqslant s \leqslant k-1$, and that $1/n \ll \alpha, \mu, 1/k$. Let $\mathcal{H}$ be a k-graph of order n with $d_{\mathcal{H}}(S) \geqslant \mu\binom{n}{k-s}$ for any set $S \in \binom{V(\mathcal{H})}{s}$, and* *let R be a subset of $V(\mathcal{H})$ of size αn chosen uniformly at random. Then the probability that $|N_{\mathcal{H}}(S) \cap \binom{R}{k-s}| \geqslant \mu\binom{\alpha n}{k-s} - n^{k-s-1/3}$ for every $S \in \binom{V(\mathcal{H})}{s}$ is $1 - o(1)$.*

The proof of Lemma 5.19 is a standard probabilistic proof, which proceeds by applying Chernoff bounds to the size of the neighbourhood of each set S, and summing the probabilities of failure over all S. We omit the details.

The second result is the following theorem of Daykin and Häggkvist [9], giving an upper bound on the vertex degree needed to guarantee a perfect matching in a k-graph $\mathcal{H}$.

Theorem 5.20 ([9], Theorem 3) *Suppose that $k \geqslant 2$ and $k|n$. Let $\mathcal{H}$ be a k-graph of order n with minimum vertex degree at least $\frac{k-1}{k}\left(\binom{n-1}{k-1} - 1\right)$. Then $\mathcal{H}$ contains a perfect matching.*

Proof of Theorem 1.11. In our proof we use constants that satisfy the hierarchy

$$\frac{1}{n} \ll \frac{1}{D} \ll \varepsilon \ll \alpha \ll c \ll \gamma \ll \gamma' \ll \eta \ll \eta' \ll \frac{1}{k}.$$

Apply Lemma 5.14 to find an absorbing ℓ-path P_0 in $\mathcal{H}$ which contains at most $\eta n/4$ vertices and which can absorb any set of at most αn c-good $(k-\ell)$-sets of vertices of $\mathcal{H}$. Define the $(k-\ell)$-graph $\mathcal{G}$ on the same vertex set as $\mathcal{H}$ to consist of all the $(k-\ell)$-sets of vertices of $\mathcal{H}$ which are c-good . Then by condition (1) of Lemma 5.14, $d_{\mathcal{G}}(v) \geqslant \binom{n-1}{k-\ell-1} - \gamma n^{k-\ell-1} \geqslant (1-\gamma')\binom{n}{k-\ell-1}$ for every vertex v in $V(\mathcal{G}) \setminus V(P_0)$.

Now, let R be a set of αn vertices of $\mathcal{H}$ chosen uniformly at random. Then by Lemma 5.19, with probability $1 - o(1)$ we have that $|N_{\mathcal{G}}(v) \cap \binom{R}{k-\ell-1}| \geqslant (1-2\gamma')\binom{\alpha n}{k-\ell-1}$ for every vertex v in $V(\mathcal{G}) \setminus V(P_0)$. Likewise, with probability $1 - o(1)$ we have that

$$|N_{\mathcal{H}}(S) \cap R| \geqslant \left(\frac{1}{\lceil\frac{k}{k-\ell}\rceil(k-\ell)} + \frac{\eta}{2}\right)\alpha n.$$

for any $(k-1)$-set S of vertices of $\mathcal{H}$. Finally, $\mathbb{E}[|R \cap V(P_0)|] = \alpha|P_0|$, and so with probability at least 1/2 we have that $|R \cap V(P_0)| \leqslant \alpha\eta n/2$. Thus we may fix a choice of R such that each of these three properties holds. Let $R' = R \setminus V(P_0)$, so $|R'| \geqslant (1-\eta/2)\alpha n$. Then $|N_{\mathcal{G}}(v) \cap \binom{R'}{k-\ell-1}| \geqslant (1-\eta')\binom{\alpha n}{k-\ell-1}$ for every vertex v in $V(\mathcal{G}) \setminus V(P_0)$, and $|N_{\mathcal{H}}(S) \cap R'| \geqslant \frac{\alpha n}{\lceil \frac{k}{k-\ell} \rceil (k-\ell)}$ for any $(k-1)$-set S of vertices of $\mathcal{H}$.

Let $V' = V(\mathcal{H}) \setminus (V(P_0) \cup R)$, and let $\mathcal{H}' = \mathcal{H}[V']$ be the restriction of $\mathcal{H}$ to V'. Then as $|V(P_0) \cup R| \leqslant \eta n/2$, we must have

$$\delta(\mathcal{H}') \geqslant \left(\frac{1}{\lceil \frac{k}{k-\ell} \rceil (k-\ell)} + \frac{\eta}{2} \right) n.$$

We may therefore apply Lemma 5.15 to $\mathcal{H}'$ to find a set of at most D disjoint ℓ-paths $P_1, \ldots, P_q$ in $\mathcal{H}'$ which include all but at most εn vertices of $\mathcal{H}'$. Let X be the set of vertices not included in any of these ℓ-paths, so $|X| \leqslant \varepsilon n$.

For each $0 \leqslant i \leqslant q$, let P_i^{beg} and P_i^{end} be ordered ends of P_i. Next we find disjoint ℓ-paths P_i' for each $0 \leqslant i \leqslant q$, so that P_i' is an ℓ-path from P_i^{end} to P_{i+1}^{beg} (subindices taken modulo $q+1$) which only contains vertices from $R' \cup P_i^{end} \cup P_{i+1}^{beg}$, and which contains at most $8k^5$ vertices in total. So suppose that we have found such ℓ-paths $P_0', \ldots, P_{i-1}'$. Let $R_i = (R' \cup P_i^{end} \cup P_{i+1}^{beg}) \setminus \bigcup_{j=0}^{i-1} V(P_j')$. Then $\delta(\mathcal{H}[R_i]) \geqslant \frac{\alpha n}{\lceil \frac{k}{k-\ell} \rceil (k-\ell)} - 8k^5 D \geqslant \alpha n/2k$, and so by Corollary 5.11 we can choose such an ℓ-path P_i' in $\mathcal{H}[R_i]$.

Then $C = P_0 P_0' P_1 P_1' \ldots P_q P_q'$ is an ℓ-cycle containing almost every vertex of $\mathcal{H}$. Indeed, C contains every vertex of $\mathcal{H}$ except for those in X and those in R' not contained in any P_i'. So let $R'' = V(\mathcal{H}) \setminus V(C)$. Then $(1-\eta)\alpha n \leqslant |R''| \leqslant (\alpha+\varepsilon)n$. Since $(k-\ell)|n$ and $(k-\ell)\big||C|$ (as C is an ℓ-cycle), we also have $(k-\ell)\big||R''|$. Furthermore, $N_{\mathcal{G}[R'']}(v) \geqslant (1-2\eta')\binom{\alpha n}{k-\ell-1}$ for every vertex $v \in R''$. Since $k-\ell \geqslant 2$, Theorem 5.20 tells us that $\mathcal{G}[R'']$ contains a perfect matching, and so we can partition R'' into at most αn c-good $(k-\ell)$-sets of vertices of $\mathcal{H}$. Since P_0 can absorb any collection of at most αn c-good $(k-\ell)$-sets, there exists an ℓ-path Q_0 with the same ordered ends as P_0 and such that $V(Q_0) = V(P_0) \cup R''$. Then $C' = Q_0 P_0' P_1 P_1' \dots P_q P_q'$ is a Hamilton ℓ-cycle in $\mathcal{H}$, completing the proof of Theorem 1.11. $\square$

LIST OF REFERENCES

[1] N. Alon, R.A. Duke, H. Lefmann, V. Rödl and R. Yuster, The algorithmic aspects of the regularity lemma, *Journal of Algorithms* **16** (1994), 80–109.

[2] N. Alon and A. Shapira, Testing subgraphs in directed graphs, *Journal of Computer and System Sciences* **69** (2004), 354–382.

[3] K. Azuma, Weighted sums of certain dependant random variables, *Tohoku Mathematical Journal* **19** (1967), 357–367.

[4] J.C. Bermond, A. Germa, M.C. Heydemann and D. Sotteau, Hypergraphes hamiltoniens, *Problèmes Combinatoires et Théorie des Graphes, Orsay* **260** (1976), 39–43.

[5] B. Bollobás, Random graphs, Cambridge University Press, 2001.

[6] S.A. Burr, Subtrees of directed graphs and hypergraphs, *Proceedings of the Eleventh Southeastern Conference on Combinatorics, Graph Theory and Computing* (1980), 227-239.

[7] S. Céroi and F. Havet, Trees with three leaves are $(n+1)$-unavoidable, *Discrete Applied Mathematics* **141** (2004), 19–39.

[8] O. Cooley, N. Fountoulakis, D. Kühn, D. Osthus, Embeddings and Ramsey numbers of sparse k-uniform hypergraphs, *Combinatorica* **29** (2009), 263–297.

[9] D.E. Daykin and R. Häggkvist, Degrees giving independent edges in a hypergraph, *Bulletin of the Australian Mathematical Society* **23** (1981), 103–109.

[10] G.A. Dirac, Some theorems on abstract graphs, *Proceedings of the London Mathematical Society* **2** (1952), 69–81.

[11] A. El Sahili, Trees in tournaments, *Journal of Combinatorial Theory, Series B* **92** (2004), 183–187.

[12] A. El Sahili, Paths with two blocks in k-chromatic digraphs, *Discrete Mathematics* **287** (2004), 151–153.

[13] W. T. Gowers, Hypergraph regularity and the multidimensional Szemerédi theorem, *Annals of Mathematics* **166** (2007), 897–946.

[14] R. Häggkvist and A.G. Thomason, Trees in tournaments, *Combinatorica* **11** (1991), 123–130.

[15] H. Hàn, Y. Person, and M. Schacht, On perfect matchings in uniform hypergraphs with large minimum vertex degree, *SIAM Journal on Discrete Mathematics* **23** (2009), 732–748.

[16] H. Hàn and M. Schacht, Dirac-type results for loose Hamilton cycles in uniform hypergraphs, *Journal of Combinatorial Theory, Series B* **100** (2010), 332–346.

[17] F. Havet, Trees in tournaments, *Discrete Mathematics* **243** (2002), 121–134.

[18] F. Havet, On unavoidability of trees with k leaves, *Graphs and Combinatorics* **19** (2003), 101–110.

[19] F. Havet and S. Thomassé, Median orders of tournaments: a tool for the second neighbourhood problem and Sumner's conjecture, *Journal of Graph Theory* **35** (2000), 244–256.

[20] F. Havet and S. Thomassé, Oriented Hamiltonian paths in tournaments: a proof of Rosenfeld's conjecture, *Journal of Combinatorial Theory, Series B* **78** (2000), 243–273.

[21] S. Janson, T. Łuczak and A. Ruciński, Random graphs, Wiley-Interscience, 2000.

[22] G.Y. Katona and H.A. Kierstead, Hamiltonian chains in hypergraphs, *Journal of Graph Theory* **30** (1999), 205–212.

[23] P. Keevash, A hypergraph blow-up lemma, preprint.

[24] P. Keevash, D. Kühn, R. Mycroft and D. Osthus, Loose Hamilton cycles in hypergraphs, preprint.

[25] P. Keevash, D. Kühn and D. Osthus, An exact minimum degree condition for Hamilton cycles in oriented graphs, *Journal of the London Mathematical Society* **79** (2009), 144–166.

[26] L. Kelly, D. Kühn and D. Osthus, A Dirac type result for oriented graphs, *Combinatorics, Probability and Computing* **17** (2008), 689–709.

[27] Y. Kohayakawa, V. Rödl and J. Skokan, Hypergraphs, quasi-randomness and conditions for regularity, *Journal of Combinatorial Theory, Series A* **97** (2002), 307–352.

[28] D. Kühn, R. Mycroft and D. Osthus, Hamilton ℓ-cycles in uniform hypergraphs, *Journal of Combinatorial Theory, Series A*, **117** (2010), 910–927.

[29] D. Kühn, R. Mycroft and D. Osthus, An approximate version of Sumner's universal tournament conjecture, *Journal of Combinatorial Theory, Series B*, to appear.

[30] D. Kühn, R. Mycroft and D. Osthus, A proof of Sumner's universal tournament conjecture for large tournaments, *Proceedings of the London Mathematical Society*, to appear.

[31] D. Kühn and D. Osthus, Loose Hamilton cycles in 3-uniform hypergraphs of high minimum degree, *Journal of Combinatorial Theory, Series B* **96** (2006), 767–821.

[32] D. Kühn and D. Osthus, Embedding large subgraphs into dense graphs, in Surveys in Combinatorics 2009, eds. S. Huczynska, J.D. Mitchell and C.M. Roney-Dougal, *London Mathematical Society Lecture Note* **365**, Cambridge University Press, 2009, 137–168.

[33] D. Kühn, D. Osthus and A. Treglown, Hamiltonian degree sequences in digraphs, *Journal of Combinatorial Theory, Series B* **100** (2010), 367–380.

[34] C. McDiarmid, Concentration, in Probabilistic Methods for Algorithmic Discrete Mathematics, eds. M. Habib, C. McDiarmid, J. Ramirez-Alfonsin and B. Reed, Springer, 1998, 195–248.

[35] K. Markström and A. Ruciński, Perfect matchings and Hamilton cycles in hypergraphs with large degrees, preprint.

[36] O. Pikhurko, Perfect matchings and K_4^3-tilings in hypergraphs of large codegree, *Graphs and Combinatorics* **24** (2008), 391–404.

[37] L. Redei, Ein kombinatorischer Satz, *Acta Litteraria Szeged* **7** (1934), 39–43.

[38] K.B. Reid and N.C. Wormald, Embedding oriented n-trees in tournaments, *Studia Scientiarum Mathematicarum Hungarica* **18** (1983) 377–387.

[39] V. Rödl, A. Ruciński and E. Szemerédi, A Dirac-type theorem for 3-uniform hypergraphs, *Combinatorics, Probability & Computing* **15** (2006), 229–251.

[40] V. Rödl, A. Ruciński and E. Szemerédi, An approximate Dirac-type theorem for k-uniform hypergraphs, *Combinatorica* **28** (2008), 229–260.

[41] V. Rödl and M. Schacht, Regular partitions of hypergraphs: regularity lemmas, *Combinatorics, Probability & Computing* **16** (2007), 833–885.

[42] V. Rödl and M. Schacht, Regular partitions of hypergraphs: counting lemmas, *Combinatorics, Probability & Computing* **16** (2007), 887–901.

[43] V. Rödl and J. Skokan, Regularity lemma for uniform hypergraphs, *Random Structures & Algorithms* **25** (2004), 1–42.

[44] B. Sudakov and J. Vondrák, A randomized embedding algorithm for trees, *Combinatorica*, to appear.

[45] E. Szemerédi, Regular partitions of graphs, *Problèmes Combinatoires et Théorie des Graphes, Orsay* (1976), 399–401.

[46] N.C. Wormald, Subtrees of large tournaments, *Combinatorial Mathematics X, Springer Lecture Notes in Mathematics* **1036** (1983) 417–419.