Large Structures in Dense Directed Graphs

by

Tássio Naia dos Santos

A thesis submitted to
the University of Birmingham
for the degree of
Doctor of Philosophy

School of Mathematics
College of Engineering and Physical Sciences
University of Birmingham

July 2018

If you have an apple and I have an apple and we exchange apples, then you and I will still each have one apple. But if you have an idea and I have an idea and we exchange these ideas, then each of us will have two ideas.

George Bernard Shaw
(attributed)

Abstract

We answer questions in extremal combinatorics, for directed graphs. Specifically, we investigate which large tree-like directed graphs are contained in all dense directed graphs of large order. More precisely, let T be an oriented tree of order n; among others, we establish the following results.

(1) We obtain a sufficient condition which ensures every tournament of order n contains T, and show that almost every tree possesses this property.

(2) We prove that for all positive C, ε and sufficiently large n, every tournament of order $(1+\varepsilon)n$ contains T if $\Delta(T) \leq (\log n)^C$.

(3) We prove that for all positive Δ, ε and sufficiently large n, every directed graph G of order n and minimum semidegree $(1/2+\varepsilon)n$ contains T if $\Delta(T) \leq \Delta$.

(4) We obtain a sufficient condition which ensures that every directed graph G of order n with minimum semidegree at least $(1/2+\varepsilon)n$ contains T, and show that almost every tree possesses this property.

(5) We extend our method in (4) to a class of tree-like spanning graphs which includes all orientations of Hamilton cycles and large subdivisions of any graph.

Result (1) confirms a conjecture of Bender and Wormald and settles a conjecture of Havet and Thomassé for almost every tree; (2) strengthens a result of Kühn, Mycroft and Osthus; (3) is a directed graph analogue of a classical result of Komlós, Sárközy and Szemerédi and is implied by (4); and (5) is of independent interest.

Acknowledgements

If this thesis ever saw the light of day, it is because I could stand in the shoulders of giants—Big Friendly Giants with whom I have had the honour to share ideas, meals, dreams and worries, badminton games, laughs and so much more. I would like to thank you, whose support over the years has brought me so much joy and inspiration, and taught me more than I can hope to acknowledge.

First and foremost, I would like to thank my advisor, Richard, for his knowledge, understanding, kindness and support throughout the PhD, and for inspiring me to do more and better mathematics.

I would also like to thank Andrew and Julia for the careful reading of this thesis and for their many insightful suggestions.

My life has been enriched at and beyond work by many wonderful people I have met at Birmingham. My sincere thanks go to the Mathematics department, especially the combinatorics group—present and past—for being such a vibrant and warm community, and for many stimulating conversations.

My PhD would not have been nearly as fulfilling if it wasn't for some special friends. In a pseudo-random order (promise!) Frederik, Stefan, Fábio, Robert, Andrew, Stefano, Nicolás, Ryan and Ben.

This section would be incomplete without including André and Johanna (and family!) for your warm welcome and for being home away from home.

Rent can be expensive if you have a family but no salary. I am truly grateful for the financial support provided by CNPq (Proc. 201114/2004-3), however late it came.

I have been blessed with a wonderful family, who supports me well beyond reason. Special thanks go to my parents, Adonel and Célia, for giving me wings and roots. Vocês me deram carinho, educação, e sempre acreditaram em mim. Elisabeth and Kale, thank you for being there always, especially in times of need. Emmanuelle and Tamira, for your rooting and deep caring. I love you all.

And of course, Sophie and Ka'imi. Words can't do you justice, but all the people above know how much you mean to me. You are always there for me—in so many ways, especially when it is hard. Je vous aime à la folie, de montão. I am better, every day, because of you.

Contents

List of Theorems

Every **alg**orithm, **cl**aim, **conj**ecture, **cor**ollary, **f**act, **d**efinition, **l**emma, **p**roblem, **q**uestion, and **t**heorem in the thesis, with a brief description and references when obtained from other sources. Main results are starred. Definitions are indexed on page 140.

List of Figures

1 Introduction

In this thesis we study various conditions which ensure that certain (dense) directed graphs contain spanning copies of many (non-isomorphic) oriented trees. One of the most basic assertions in graph theory is that a graph G is connected if and only if it contains a spanning tree T. It is also simple to show that every graph G contains every tree of order $\delta(G)+1$, but the same is not true if the edges of G and T are given orientations (see Figure 1.1). To 'recover' the result, we may replace minimum degree $\delta(G)$ by minimum semidegree, where the minimum semidegree $\delta^0(G)$ of a digraph G is the minimum of all the in- and outdegrees of the vertices in G.

In the next subsections we introduce some problems of the same flavour, which arise from well known and studied properties of graphs when orientations are given to the edges. We also state there our main results. In Chapter 2 we state (sometimes with proof) auxiliary results, define concepts and introduce notation which is used in the proofs of our main results in Chapters 3, 4 and 5, as well as in the next section. Finally, in Chapter 6, we describe a few open problems and directions for further research. The results in this thesis are joint work with Richard Mycroft [66, 67].

1.1 Unavoidable trees

An oriented graph H on n vertices is *unavoidable* if every tournament on n vertices contains a copy of H; otherwise, we say that H is *avoidable* (see figures 1.1 and 1.2). In particular, if H contains a directed cycle then H must be avoidable, since a transitive tournament contains no directed cycles and hence no copy of H. It is therefore natural to ask which oriented trees are unavoidable. A classical

unavoidable

avoidable

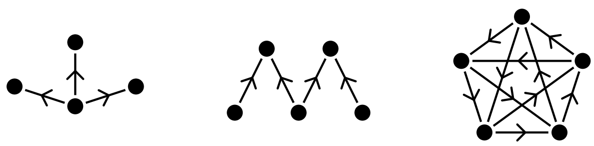


Figure 1.1: Two trees and a tournament (on the right) which avoids them.

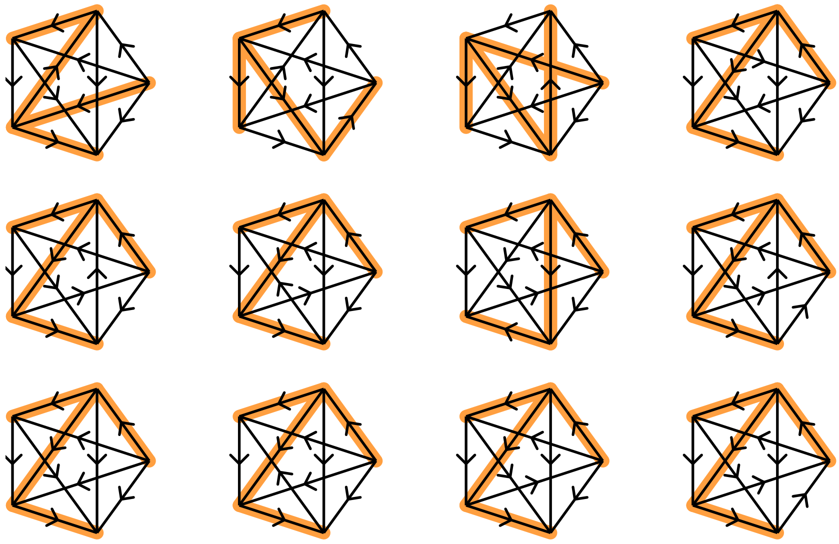


Figure 1.2: An unavoidable tree in all 12 nonisomorphic tournaments of order 5.

result of Rédei [68] states that every directed path is unavoidable (a directed path is a path $\bullet\!\!\to\bullet\to\cdots\to\!\!\bullet$ whose edges are all oriented towards the same leaf). Over 50 years later, Thomason [79] showed that every orientation of a cycle of order at least 2^{128} is unavoidable, except for those which yield directed cycles (see Figure 1.3). In particular this implies that all orientations of paths with order at least 2^{128} are unavoidable. Havet and Thomassé [41] then gave a complete answer for paths: with three exceptions, every orientation of a path is unavoidable (the exceptions are antidirected paths of order 3, 5 and 7, which are not contained in the directed cycle of length 3, the regular 5-vertex tournament and the Paley tournament on 7 vertices respectively). Significant attention has also been focused on the unavoidability of claws (a claw is an oriented graph formed by identifying the initial vertices of a collection of vertex-disjoint directed paths). Indeed Saks and Sós [70] conjectured that every claw on n vertices with maximum degree at most $n/2$ is unavoidable. Lu [60, 59] gave a counterexample to this conjecture, but in the other direction showed that every claw with maximum degree at most $3n/8$ is unavoidable. Lu, Wang and Wong [62] then extended these results by showing that every claw with maximum degree at most $19n/50$ is unavoidable, but that there exist claws with maximum degree approaching $11n/23$ which are avoidable. Finding the supremum of all $c > 0$ for which every claw with maximum degree at most cn is unavoidable remains an open problem.

Some oriented trees are far from being unavoidable. For example, the out-directed star S on n vertices (whose edges are oriented from the central vertex to each of the $n-1$ leaves) is not contained in a regular tournament on $2n-3$

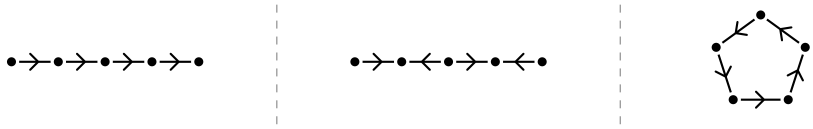


Figure 1.3: In a directed path (left) all edges are oriented consistently, 'towards' the same leaf; in an anti-directed path (centre) the orientation of the edges alternates; finally, in a directed cycle (right) the removal of any edge yields a directed path.

vertices, since each vertex of the latter has only $n-2$ outneighbours. That is, there exist tournaments with almost twice as many vertices as S which do not contain a copy of S. On the other hand, Bender and Wormald [8] proved that almost all oriented trees are 'almost unavoidable', in the sense that they are contained in almost all tournaments on the same number of vertices.

Theorem 1.1. [8, Theorem 4.4] Let $\mathcal{T}_n$ denote the set of all labelled oriented trees on n vertices. Then there is a subset $\mathcal{T}'_n \subseteq \mathcal{T}_n$ of size $(1-\mathrm{o}(1))|\mathcal{T}_n|$ such that a uniformly-random labelled tournament on n vertices asymptotically almost surely contains every tree in $\mathcal{T}'_n$.

In particular, it follows that if T is chosen uniformly at random from the set of all labelled oriented trees on n vertices, and G is chosen uniformly at random from the set of all labelled tournaments on n vertices, then asymptotically almost surely G contains a copy of T. In the same paper Bender and Wormald conjectured that this holds for every tournament G, or, in other words, that almost all labelled oriented trees are unavoidable. The main result of Chapter 3 is to prove this conjecture.

Theorem 1.2. Let T be chosen uniformly at random from the set of all labelled oriented trees on n vertices. Then asymptotically almost surely T is unavoidable.

The following definitions are crucial for the proof of Theorem 1.2. We say that a subtree T' of a tree T is *pendant* if $T-T'$ is connected. Next, we define 'nice' oriented trees, whose properties are useful for embedding in tournaments, as follows (see Figure 1.4). pendant

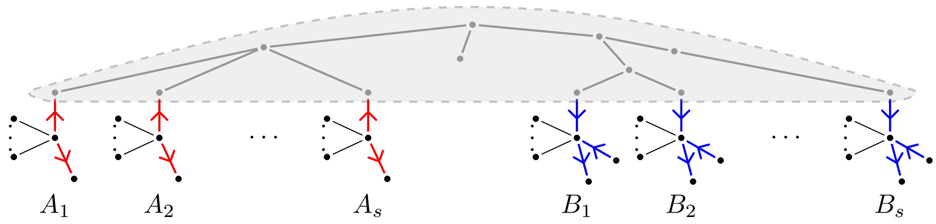


Figure 1.4: An α-nice tree T has $s = \lceil \alpha n \rceil$ pendant stars $A_1, \ldots A_s$ which contain an out-leaf of T such that the edge between $T - A_i$ and A_i is directed away from A_i, and also s pendant stars $B_1, \ldots, B_s$ which contain both an in-leaf of T and an out-leaf of T such that the edge between $T - B_i$ and B_i is directed towards B_i. In this illustration we only indicate the orientations of edges specified by this definition. The shaded area is the subtree $T - \bigcup_{i\in[s]} \big(V(A_i) \cup V(B_i)\big)$.

α-nice

Definition 1.3 (α-nice). For $\alpha > 0$ we say that an oriented tree T on n vertices is α-*nice* if, writing $s := \lceil \alpha n \rceil$, T contains $2s$ vertex-disjoint pendant oriented stars $A_1, \ldots, A_s$ and $B_1, \ldots, B_s$ such that for each $i \in [s]$

(i) A_i is a subtree of T which contains an out-leaf of T and the edge between A_i and $T - A_i$ is oriented away from A_i, and

(ii) B_i is a subtree of T which contains both an in-leaf of T and an out-leaf of T and the edge between B_i and $T - B_i$ is oriented towards B_i.

We note that the asymmetry in the definition above is a product of our proof; a similar definition (with the directions of the edges reversed) would also work.

Most of the work involved in proving Theorem 1.2 is in the proof of the following theorem, which states that large nice oriented trees with polylogarithmic maximum degree are unavoidable.

Theorem 1.4. For every $\alpha, C > 0$ there exists n_0 such that if T is an oriented tree on $n \geq n_0$ vertices such that

(i) $\Delta(T) \leq (\log n)^C$ and

(ii) T is α-nice,

then T is unavoidable.

Almost all labelled trees satisfy condition (i) of Theorem 1.4, as proved by Moon.

Theorem 1.5. [65, Corollaries 1 and 2] For every $\varepsilon > 0$, if T is chosen uniformly at random from the set of all labelled trees on n vertices, then asymptotically almost surely

$$(1-\varepsilon)\frac{\log n}{\log\log n} \leq \Delta(T) \leq (1+\varepsilon)\frac{\log n}{\log\log n}.$$

Since a uniformly-random orientation of a uniformly-random labelled tree yields a uniformly-random labelled oriented tree, Theorem 1.5 remains valid if we replace 'labelled tree' by 'labelled oriented tree'. We prove that almost all labelled oriented trees satisfy condition (ii) of Theorem 1.4.

Theorem 1.6. Let T be chosen uniformly at random from the set of all labelled oriented trees on n vertices. Then asymptotically almost surely T is $\frac{1}{250}$-nice.

Combining Theorems 1.4, 1.5 and 1.6 (with $C = \varepsilon = 1$ and $\alpha = \frac{1}{250}$) immediately proves Theorem 1.2.

Another natural question is to find, for a given oriented tree T, the smallest integer $t(T)$ such that every tournament on $t(T)$ vertices contains a copy of T. In particular, T is unavoidable if and only if $t(T) = |T|$, where $|T|$ denotes the order of T. Sumner conjectured the following. $t(T)$

Conjecture 1.7. [83] If T is an oriented tree, then $t(T) \leq 2|T| - 2$.

The example of an outdirected star described above demonstrates that this bound would be best possible. Kühn, Mycroft and Osthus [52, 51] used a randomised embedding algorithm to prove that Conjecture 1.7 holds for sufficiently large n; previous upper bounds on $t(T)$ had been established by Chung [19], Wormald [83], Häggkvist and Thomason [35], Havet [39], Havet and Thomassé [40] and El Sahili [29]. In particular, El Sahili proved that $t(T) \leq 3n - 3$ for every oriented tree T on n vertices, and this remains the best known upper bound on $t(T)$ for small n.

Alas, despite significant advances, it still not clear what parameter governs the 'embeddability' of an oriented tree. Havet and Thomassé [39] proposed investigating the parameter $h(\ell)$, defined as the smallest integer such that every tournament of order $n+h(\ell)$ contains every tree of order n which contains precisely ℓ leaves. Häggkvist and Thomason [35] have shown that $h(\ell) \leq 2^{512\ell^3}$. This was subsequently improved to $h(\ell) \leq \frac{3}{2}(\ell^2 - 3\ell) + 5$ by Havet [39], who jointly with Thomassé proposed the following conjecture. $h(\ell)$

Conjecture 1.8. If ℓ is an integer and $\ell \geq 2$, then $h(\ell) \leq \ell - 1$.

Conjecture 1.8 is known to hold for paths and it is not difficult to show that it holds for stars as well. We provide some support for this conjecture. Since every tree has at least 2 leaves, it follows that every unavoidable tree satisfies Havet and Thomassé's conjecture. Therefore, by Theorem 1.2, Conjecture 1.8 is correct for almost every tree.

Kühn, Mycroft and Osthus [51] also gave a stronger bound for $t(T)$ when T is a large oriented tree of bounded maximum degree. More precisely, they proved that for every $\alpha, \Delta > 0$, if n is sufficiently large then every oriented tree T on n vertices with $\Delta(T) \leq \Delta$ has $t(T) \leq (1+\alpha)n$. In other words, bounded degree oriented trees are close to being unavoidable, in that they are contained in every tournament of slightly larger order.

Our proof of Theorem 1.4 makes use of the aforementioned random embedding algorithm of Kühn, Mycroft and Osthus, using somewhat sharper estimates on certain quantities associated with the random embedding. In particular, using these stronger estimates we are able to establish the same bound on $t(T)$ for oriented trees whose maximum degree is at most polylogarithmic in n (rather than bounded by a constant as above). This is the following theorem, which we use repeatedly in the proof of Theorem 1.4, and which is of independent interest.

Theorem 1.9. For every $\alpha, C > 0$ there exists n_0 such that if T is an oriented tree on $n \geq n_0$ vertices with $\Delta(T) \leq (\log n)^C$ and G is a tournament on at least $(1+\alpha)n$ vertices, then G contains a copy of T.

Observe that, under the assumption that Theorem 1.4 holds, we can deduce Theorem 1.9 immediately by appending a linear number of pendant stars to T. However, since Theorem 1.9 plays a crucial role in the proof of Theorem 1.4, we cannot use this deduction.

We prove Theorems 1.2, 1.4, 1.6 and 1.9 in Chapter 3.

1.2 Spanning trees via high semidegree

We now turn away from tournaments and consider other dense digraphs. (All digraphs we consider are simple, see Figure 1.5.) There are many actively researched questions for digraphs with given minimum degree (for many distinct notions of degree). Before describing our main results, we mention a few of these.

A well-known open problem is whether the absence of short directed cycles forces a digraph to have small outdegrees; this is the following conjecture, raised 40 years ago by Caccetta and Häggkvist.

Conjecture 1.10. [17] Every simple digraph of order n with minimum outdegree at least r has a cycle with length at most $\lceil n/r \rceil$.

Caccetta and Häggkvist [17] have settled the case $r = 2$; it has also been confirmed for $r = 3$ by Hamidoune [37], for $r \in \{4, 5\}$ by Hoáng and Reed [42] and for $r \leq \sqrt{n/2}$ by Shen [73]; many other related results have been obtained (see, e.g., [7, 75]).

Rather than forcing a 'single' subgraph (such as a cycle), one can also ask for many disjoint copies of some fixed digraph. Indeed, significant attention has been given to the question of whether high (semi)degrees guarantee the existence of a perfect tiling. More precisely, an *H-tiling* of a graph G is a collection of vertex-disjoint copies of H in G; an H-tiling of G is *perfect* if it covers all vertices in G. We highlight three results in the area. Yuster [46] has shown that for every positive ε, h and every transitive tournament T, if H is the digraph obtained from T by replacing each vertex with an independent set of size $h \geq 1$, then there exists $N = N(\varepsilon, h, T)$ and $c = c(T) > 0$ such that if G is a graph of order $n \geq N$ and minimum degree cn, then every orientation of G admits an H-tiling covering all but at most εn vertices of G. Cuckler [22] and Yuster [84] asked whether every regular tournament of order $n \equiv 3 \bmod 6$ must contain a perfect directed triangle tiling (a directed triangle is a directed cycle of order 3); Keevash and Sudakov [47] have shown that every large oriented graph G of order n with minimum semidegree at least $(1/2 - \mathrm{o}(1))n$ admits a directed triangle tiling which covers all but at most 3 vertices of G; the question was finally settled by Li and Molla [58]. In a similar vein, Czygrinow, Kierstead and Molla [24] have shown that every digraph G with order n, where $n = ks$, and minimum total degree $\min_{v \in G} \deg^-(v) + \deg^+(v)$ at least $2k(s-1) - 1$ contains k vertex-disjoint transitive tournaments of order s (this is a generalisation, for digraphs, of a result of Hajnal and Szemerédi [36]). The number of results and conjectures in the area is large (see, e.g., [7, 23, 48, 54, 80, 81]), and the above are a small sample.

Let $\mathcal{F}$ be a family of trees of order n (for instance, the set of all oriented paths of order n). We investigate the following question: what is the smallest

H-tiling

perfect tiling

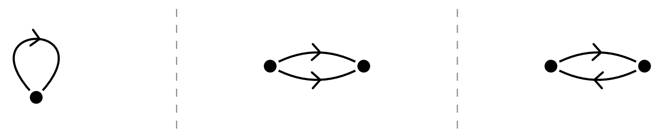


Figure 1.5: *Simple digraphs* contain no loops (left) or parallel arcs (centre); besides those, oriented graphs also do not contain 2-cycles (right).

simple digraph

integer $\delta_{n,\mathcal{F}}$ such that every digraph G of order n with semidegree $\delta^0(G) \geq \delta_{n,\mathcal{F}}$ contains every tree in $\mathcal{F}$? Note that if $\delta^0(G) = \lceil n/2 \rceil - 1$ then G may not even be connected (see Figure 1.6), so $\delta_{n,\mathcal{F}} \geq n/2$ whenever $\mathcal{F}$ is not empty. On the other hand, Ghouila-Houri [43] proved that if $\delta^0(G) \geq n/2$, then G contains a directed Hamilton cycle (i.e., a consistently oriented spanning cycle), and thus contains a spanning directed path. This has been extended recently by DeBiasio, Kühn, Molla, Osthus and Taylor [25], who proved that if $\delta^0(G) \geq n/2$ (and n is sufficiently large) then G contains every possible orientation of a Hamilton cycle, except perhaps for the antidirected one; the threshold for existence of antidirected Hamilton cycles is $\delta^0(G) \geq n/2 + 1$ (again, for sufficiently large n) and was established by DeBiasio and Molla [26] (see Figure 1.7).

The results described above are sharp in the sense that the value of $\delta_{n,\mathcal{F}}$ is determined precisely (see Figures 1.6 and 1.7). Hence if G has semidegree at least $n/2$ then G contains every orientation of a Hamilton path (this can be shown by changing the orientation of at most one edge in a Hamilton cycle of G to yield the desired Hamilton path), and as noted above this bound is best possible. To the best of our knowledge, not much else is known about spanning trees of general digraphs G with given semidegree if no other conditions are imposed on G. For instance, more is known about the presence of Hamilton cycles in G if G is an oriented graph [48] or if G is strongly connected [56].

The situation in the graph setting is quite different (see, e.g.: [53, 55]). For instance, a classical theorem by Komlós, Sárközy and Szemerédi [49] states that if $\delta(G) \geq \big(1/2 + o(1)\big)n$ then G contains every tree with bounded degree.

Theorem 1.11. [49] For every positive integer Δ, every real $0 < \alpha < 1/2$ and sufficiently large integer n, every graph of order n and minimum degree at least $\left(\frac{1}{2} + \alpha\right)n$ contains every spanning tree T with maximum degree at most Δ.

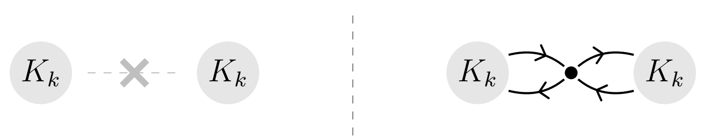


Figure 1.6: Digraphs with order n, and semidegree $\lceil n/2 \rceil - 1$ which do not contain any orientation of a Hamilton cycle. In the figure, K_k denotes a digraph of order k where each pair of vertices is connected by two edges, one in each direction. Left: two disjoint copies of K_k, so $n = 2k$. Right: two disjoint copies of K_k and a vertex connected (in both directions) to each vertex in each copy, so $n = 2k + 1$.

We remark the same authors have later improved the result above, replacing the constant bound Δ by $cn/\log n$, where c is some constant depending on α [50]. One of the main results we prove in Chapter 4 is a directed graph analogue of Theorem 1.11.

Theorem 1.12. For all positive real α, C, Δ there exists n_0 such that for all $n \geq n_0$ the following holds. If G is a directed graph of order n and minimum semidegree at least $(\frac{1}{2} + \alpha)n$, then G contains every (spanning) tree of order n such that $\Delta(T) \leq \Delta$.

We prove Theorem 1.12 by establishing a stronger result. Let $\mathcal{G}_{n,\alpha}$ be the set of all digraphs G of order n with $\delta^0(G) \geq (1/2 + \alpha)n$. Let $C > 0$ and T be an oriented tree of order n, where n is sufficiently large, and suppose that T has maximum underlying degree $\Delta(T) \leq (\log n)^C$. We describe a sufficient condition for T to be contained in every $G \in \mathcal{G}_{n,\alpha}$, and use this to prove that almost every oriented tree of order n is contained in every $G \in \mathcal{G}_{n,\alpha}$. Theorem 1.12 follows because every (sufficiently large) oriented tree of bounded degree satisfies this condition.

A path P of a digraph T is *bare* if every edge in $E(T) \setminus E(P)$ which contains a vertex $v \in P$ is incident to one of its endvertices. bare

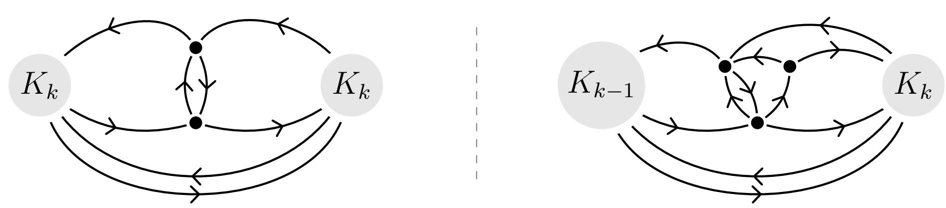


Figure 1.7: Two digraphs with order n and semidegree $n/2$ which do not contain an antidirected Hamilton cycle, due to Cai [18] (left) and to DeBiasio and Molla [26] (right). In the figure, K_p denotes a digraph of order p where each pair of vertices is connected by two edges, one in each direction, and arrowed-lines mean that all possible edges (in the corresponding direction) are present.

Theorem 1.13. Suppose that $1/n \ll 1/C$ and that $1/n \ll \lambda \ll \alpha$. Let G be a digraph of order n with $\delta^0(G) \geq (1/2+\alpha)n$, and let T be an oriented tree of order n such that $\Delta(T) \leq (\log n)^C$. If T contains either

(i) at least λn vertex-disjoint bare paths of order 7; or

(ii) at least λn vertex-disjoint edges incident to leaves,

then G contains a copy of T.

Not every tree satisfies one of the conditions above, as we now show. Let B_n be a rooted binary tree with $n/\log n$ leaves, where the root has degree 2 and every other vertex which is not a leaf has degree 3. Then B_n has $2n/\log n - 1$ vertices. Let T_n be the tree we obtain from B_n by appending $\log n - 2$ new leaf vertices to each leaf vertex of B_n so that T_n has has order $n-1$. Note that any bare-path in T_n is an edge, that $\Delta(T) \leq \log n$ and moreover that T contains at most $n/\log n$ distinct leaf-edges; this means that for all $\lambda > 0$ we can choose n sufficiently large so that T_n does not satisfy either condition of Theorem 1.13.

Since almost every tree has sublogarithmic maximum degree (by Theorem 1.5) and almost every tree satisfies (ii) (with $\lambda = \frac{1}{250}$, by Theorem 1.6) we immediately conclude the following.

Theorem 1.14. Let $\mathcal{T}_n$ be the set of all labelled oriented trees of order n. For all positive α, ε, there exists n_0 such that for all $n \geq n_0$ the following holds. There exists a family $\mathcal{F} \subseteq \mathcal{T}_n$ with $|\mathcal{F}| \geq (1-\varepsilon)|\mathcal{T}_n|$ such that if G is a digraph of order $n \geq n_0$ with $\delta^0(G) \geq (1/2+\alpha)n$ then G contains every tree $T \in \mathcal{F}$.

We conclude Chapter 4 by extending these methods to find spanning tree-like digraphs of bounded degree in all $G \in \mathcal{G}_{n,\alpha}$. Roughly speaking, a digraph H is tree-like if there exists a small subset $S \subseteq V(H)$ such that $H[S]$ has no edges and $H - S$ is a forest with at least one 'large' component. Examples of tree-like graphs include every orientation of a Hamilton cycle; indeed, every orientation of a large subdivision of every graph Q such that each edge has been subdivided a sufficient number of times.

Theorem 1.15. Suppose that $1/n \ll \alpha \ll C$. Let G be a digraph of order n with $\delta^0(G) \geq (1/2 + \alpha)n$. If Q is a subdivision of a graph Q_{under} such that

(i) $|Q_{\text{under}}| \leq (\log n)^C$;

(ii) each edge of Q_{under} has been subdivided at least $\log n$ times; and

(iii) $|Q| = n$;

then G contains every orientation of Q.

1.3 Trees via chromatic number

The *chromatic number* $\chi(G)$ of a graph G is the smallest integer k such that the vertices of G can be partitioned into k sets such that no edge of G has both endvertices in the same set. Throughout this thesis, the chromatic number of a digraph D is defined similarly, i.e.: as the chromatic number of the underlying (undirected) graph of D. chromatic number

A classical result in graph theory is that every graph G contains a subgraph H with $\delta(H) \geq \chi(G)-1$, and therefore contains a copy of every tree on $\chi(G)$ vertices. On the other hand, the following well-known result of Erdős (Theorem 1.16 below, see, e.g., [5, pp. 38–39]) implies that for every integer k, if F is contained in every graph G with $\chi(G) = k$ then F must be a forest. The *girth* of a graph is the smallest integer g such that g contains a cycle of order g. girth

Theorem 1.16 (Erdős, 1959). [30] For all positive k, ℓ there exists a graph G with $\chi(G) > k$ and $\text{girth}(G) > \ell$.

There are many open questions related to the existence of trees in undirected graphs of high chromatic number. For instance, Erdős and Sós [31] conjectured that if G is a graph and $|E(G)| > (k-2)|G|/2$ then G contains every tree of order k (a proof of this was announced in the early 1990's by Ajtai, Komlós, Simonovits and Szemerédi); and Gyárfás [34] and Sumner [76] conjectured that for any tree T and any integer n, if G is a graph of sufficiently high chromatic number, then either G contains an induced copy of T or G contains a copy of K_n (this conjecture remains open, but has been shown to be true for particular trees, such as paths [34] and subdivisions of stars [71], among others—see, e.g., [72]). Considering the relationship between chromatic number and orientations of graphs, Burr [16] (see also [83]) conjectured the following.

Conjecture 1.17. [16] If G is a graph with $\chi(G) \geq 2t-2$ and T is an oriented tree of order t, then every orientation of G contains a copy of T.

This is a far-reaching generalisation of Conjecture 1.7, since tournaments are orientations of complete graphs. Let $q(T)$ be the smallest integer such that every orientation of a $q(T)$-chromatic graph contains a copy of T; with this notation, Burr's conjecture is equivalent to the statement that if T is an oriented tree of order t then $q(T) \leq 2t-2$. Note that the example of an anti-directed star S (given in Section 1.1) implies that the value $2t-2$ cannot be replaced by a smaller function of t.

$q(T)$

Burr's conjecture is known to be true for some classes of trees. The next theorem states that every orientation of a graph G contains a directed path with at least as many vertices as the chromatic number of G.

Theorem 1.18. [32, 38, 69, 82] Every orientation of a graph G contains a directed path of order $\chi(G)$.

In other words, $q(T) = |T|$ for every directed path. This theorem was obtained independently and with distinct proofs (!) by Gallai, Hasse, Roy and Vitaver [32, 38, 69, 82]. Below is a (fifth) proof of this beautiful result, included for completeness (see, e.g.: [33]).

Proof. Fix an orientation D of G and let A be an edge-maximal acyclic subdigraph of D. We may assume that $V(A) = V(G)$. Let c be a colouring of $V(G)$ which assigns to each vertex v the number of vertices in a longest directed path of A ending in v. We will show that this is a proper colouring of G, and therefore some vertex receives label $\chi(G)$, completing the proof. First note that if uv is an edge of A, then $c(u) < c(v)$, because A is acyclic and thus the longest directed path P_u ending in u cannot contain v. Therefore c is a proper colouring of A and, in particular, c strictly increases over any directed path in A. Now consider any edge xy which is in D but not in A. Since A is edge-maximal, adding xy to A must create a directed cycle; it follows that there exists a directed path from y to x in A. Hence, $c(y) < c(x)$, and we conclude that c is a proper colouring of G. $\square$

Burr's conjecture has been settled for certain orientations of paths by El Sahili and Kouider [28], namely, for paths P of order at least 4 with at most 2 blocks (i.e: at most one 'change in direction') by proving that $q(P) = |P| + 1$ for such paths; Addario-Berry, Havet and Thomassé [2] have improved this to the best possible bound $q(P) = |P|$. Burr himself proved that $q(T) \leq (n-1)^2$ for every tree T of order n. This has been improved recently by Addario-Berry, Havet, Sales, Reed and Thomassé [1].

Theorem 1.19. [1] If T is an oriented tree T of order n, then

$$q(T) \leq \binom{n}{2} + 1.$$

In Chapter 5 we present some new preliminary results related to Burr's conjecture. Recall that if a digraph H is contained in every orientation of a k-chromatic graph G, then H cannot contain directed cycles. Moreover, H must be contained in every tournament of order k. The next theorem gives a bound on the chromatic number of acyclic subgraphs of tournaments. It is known that every digraph D contains an acyclic digraph H with $\chi(H) \geq \sqrt{\chi(D)}$ [1] (see Lemma 5.2).

Theorem 1.20. For all $\varepsilon > 0$ there exists a tournament D such that if H is an acyclic subgraph of D then

$$\chi(H) \leq \left(\frac{1}{2} + \varepsilon\right) \frac{\chi(D)}{\log \chi(D)}.$$

We also consider which trees are contained in a uniformly random orientation of a graph G of given minimum degree. Since every graph G contains a (critical) subgraph G_{crit} with minimum degree $\chi(G) - 1$, the next theorem implies that for every positive ε and graph G with chromatic number k, if k is sufficiently large, then almost surely a random orientation of G contains every oriented tree of order $(1 - \varepsilon)k/\log k$.

Theorem 1.21. For all positive ε and sufficiently large $k = k(\varepsilon)$, the following holds for every graph G be with $\delta(G) \geq k - 1$. If D is an orientation of G formed by orienting each $e \in E(G)$ uniformly at random, independently for each edge, then D contains every oriented tree of order $(1 - \varepsilon)k/\log k$ almost surely as $k \to \infty$.

Note that every graph G contains a critical subgraph G' (i.e., a subgraph G' such that $\chi(G') = \chi(G)$ but every proper subgraph of G' has chromatic number strictly less than $\chi(G)$); moreover, we have $\delta(G') \geq \chi(G') - 1 = \chi(G) - 1$ (if this was not the case, then it would be possible to delete a vertex v of degree at most $\chi(G') - 2$ from G' and then properly colour G' with $\chi(G') - 1$ colours: first colour $G' - v$ using $\chi(G') - 2$ colours, and then extend this to a proper colouring of G' using a new colour for v). In other words, if G is a graph with high chromatic number, then almost every orientation of G contains every oriented tree T whose order is close to $\chi(G)/\log\chi(G)$—which is much larger than the bound in Theorem 1.19.

We conclude the Chapter 5 with two non-probabilistic partial results towards $q^{-1}(G)$ Burr's conjecture. For any graph G, let $q^{-1}(G)$ be the largest integer t such that every orientation of G contains every oriented tree of order t. The best known general bounds imply that $q^{-1}(G)$ Our main theorem in the chapter is an approximate result towards Burr's conjecture, in the following sense: we obtain a lower bound on $q^{-1}(G)$ (depending on $\chi(G)$ and $|G|$) which for most graphs is significantly stronger than the currently known bounds.

Theorem 1.22. If D is a digraph of order n, where $n \geq 1$, then D contains every oriented tree of order $\chi(D)/\log_2(2n)$, that is,

$$q^{-1}(D) \geq \frac{\chi(D)}{\log_2(2n)}.$$

The theorem above implies that, for all $C > 0$ and sufficiently large k, if G is a graph with chromatic number k and order n, where $n \leq \mathrm{e}^{(\log k)^C}/2$, then every orientation of G contains every oriented tree of order $k/(\log k)^C$. We remark that this is typically the case. A celebrated result of Bollobás on the expected chromatic number of the binomial random graph (Theorem 5.3, page 126) implies that, for any fixed C with $0 < C < 1$, we have $|G| \leq \mathrm{e}^{(\log k)^C}/2$ for almost every graph G.

We also extend a partial result in [1], proving that Burr's conjecture holds for every orientation of every star (Theorem 5.10). This result, together with the aforementioned results for paths, suggests that there is a strong relationship between $t(T)$ and $q(T)$. I conjecture the following.

Conjecture 1.23 (Transference conjecture). If T is an oriented tree, then

$$t(T) = q(T).$$

This conjecture holds for directed paths by Theorem 1.18 and was confirmed for paths with at most two blocks (i.e., for paths with at most one non-leaf vertex which has either no inneighbours or no outneighbours) by Addario-Berry, Havet and Thomassé [2]; moreover, extending results of Addario-Berry, Havet, Sales, Reed and Thomassé [1] (Lemma 5.8) we show that this conjecture holds for stars as well (Corollary 5.11). Conjecture 1.23 is further discussed in Chapter 6.

As long as the roots are not severed, all is well. And all will be well in the garden.

Chance the Gardener

The secret of getting ahead is getting started.

Mark Twain

2 Preliminary concepts and results

We follow standard graph-theoretical notation (see, e.g., [27]). For clarity, we define some of our notation (mostly related to directed graphs) below. More specific terms are defined in later sections (all definitions are indexed on page 140).

A *directed graph* G, or *digraph* for short, is a pair (V, E) of sets: a vertex set V and an edge set E, where each edge $e \in E$ is an ordered pair of distinct vertices (more precisely, E is a set of ordered pairs $(u, v) \in V \times V$ of distinct elements of V); the *order* of G is $|V|$. We think of the edge (u, v) as being directed from u to v, and write $x{\to}y$ or $y{\leftarrow}x$ to denote the edge (x, y); if the orientation of the edge does not matter, we write $\{u, v\}$ (or $\{v, u\}$) instead. In either case, u and v are said to be the *endvertices* of $\{u, v\}$, and we also call u (respectively v) a *neighbour* of v (respectively u).

directed graph, digraph

order

endvertices

neighbour

In a digraph G, the *outneighbourhood* $N^+_G(x)$ of a vertex x is the set $\{\, y : x{\to}y \in E(G)\,\}$; the *inneighbourhood* $N^-_G(x)$ of x is $\{\, y : x{\leftarrow}y \in E(G)\,\}$. The *outdegree* and *indegree* of x in G are respectively $\deg^+_G(x) := \left|N^+_G(x)\right|$ and $\deg^-_G(x) := \left|N^-_G(x)\right|$, and the *semidegree* $\deg^0_G(x)$ of x is the minimum of the outdegree and indegree of x. We say that G is r*-regular* if for all $x \in D$ we have $\deg^-(x) = \deg^+(x) = r$. The *minimum semidegree* $\delta^0(G)$ of G is the minimum of $\deg^0_G(x)$ over all $x \in V(G)$. For any subset $Y \subseteq V(G)$, we write $\deg^-_G(x, Y)$ for $|N^-_G(x) \cap Y|$, the *indegree of x in Y*; the *outdegree of x in Y*, denoted by $\deg^+_G(x, Y)$, is defined similarly. The *semidegree of x in Y*, denoted by $\deg^0_G(x, Y)$, is the minimum of those two values. We drop the subscript when there is no danger of confusion, writing $N^-(x)$, $\deg^0(x)$, and so forth. We write $|G|$ and $e(G)$ for the number of vertices and edges of G respectively. For digraphs G and H, we call H a *subgraph* of G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$; H is said to be *spanning* if $V(H) = V(G)$. For any set $X \subseteq V(G)$, we write $G[X]$ for the subgraph of G *induced* by X, which has vertex set X and whose edges are all edges of G with both endvertices in X. If H is a subgraph of G then we write $G - H$ for $G\big[V(G) \setminus V(H)\big]$. Likewise, for a vertex v or set of vertices S, we write $G - v$ or $G - S$ for $G\big[V(G) \setminus \{v\}\big]$ or $G\big[V(G) \setminus S\big]$ respectively. Incidentally, we treat graphs as sets, writing $x \in G$ to indicate that x is a vertex of G. For disjoint subsets $X, Y \subseteq V(G)$, where G is a digraph, we denote by $G[X \to Y]$, or equivalently by $G[Y \leftarrow X]$, the subdigraph

outneighbourhood

inneighbourhood, outdegree indegree

semidegree

r-regular

minimum semidegree

degree in set

semidegree in set

subgraph

spanning

induced

of G with vertex set $X \cup Y$ and edge set

$$E\big(G[X \to Y]\big) \coloneqq \{\, x \to y \in E(G) : x \in X, y \in Y \,\}.$$

oriented

Some graphs considered in this thesis are *oriented*, meaning that there is at most one edge between each pair of vertices and there are no loops (i.e. edges with a single endvertex). Equivalently, an oriented graph G can be formed by assigning an orientation to each edge $\{u, v\}$ of some (undirected) graph H, i.e.: by replacing each $\{u, v\} \in E(H)$ by one of the possible ordered pairs (u, v) or (v, u); in this case we refer to H as the *underlying graph* of G, and say that G is an *orientation* of H. We refer to the *maximum degree* of an oriented graph G, denoted $\Delta(G)$, to mean the maximum degree of the underlying graph H.

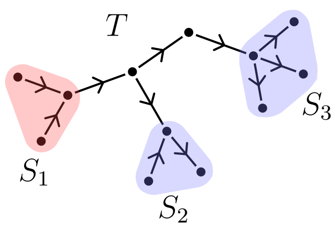
underlying graph, orientation maximum degree

directed path antidirected path strongly connected directed cycle

A *directed path* of length k is an oriented graph with vertex set $v_0, \dots, v_k$ and edges $v_{i-1} \to v_i$ for each $1 \leq i \leq k$, and an *antidirected path* of length k is an oriented graph with vertex set $v_0, \dots, v_k$ and edges $v_{i-1} \to v_i$ for odd $i \leq k$ and $v_{i-1} \leftarrow v_i$ for even $i \leq k$ (or *vice versa*). A digraph is *strongly connected* if for any ordered pair (x, y) of its vertices there exists a directed path P from x to y (i.e., all edges of P are oriented towards y). A *directed cycle* of length k is an oriented graph with vertex set $v_1, \dots, v_k$ and edges $v_i \to v_{i+1}$ for each $1 \leq i \leq k$ with addition taken modulo k.

tree, oriented tree leaf, star centre, subtree out-subtree in-subtree in-leaf, out-leaf

A *tree* is an acyclic connected graph, and an *oriented tree* is an orientation of a tree. A *leaf* in a tree or oriented tree is a vertex incident to a single edge. A *star* is a tree in which at most one vertex (the *centre*) is not a leaf. A *subtree* T' of a tree T is a subgraph of T which is also a tree, and we define subtrees of oriented trees similarly. For oriented trees T and T' we say that T' is an *out-subtree* (respectively an *in-subtree*) of T if both T' and $T - T'$ are subtrees of T, and the unique edge of T between T' and $T - T'$ is directed towards T' (respectively away from T'). In a similar way we say that a vertex is an *in-leaf* or *out-leaf* of T; in



in-star out-stars

Figure 2.1: A tree T and some of its subtrees (each formed by the vertices in shaded areas). The tree S_1 is an *in-star* (in-subtree) of T, while both S_2 and S_3 are *out-stars* (out-subtrees). Moreover, S_1 and S_3 are anti-directed stars.

particular, an *anti-directed star* is an orientation of a star S in which every vertex has either no inneighbours or no outneighbours. Now let T be a tree or oriented tree (see Figure 2.1). anti-directed star

It is often helpful to nominate a vertex r of T as the *root* of T; to emphasise this fact we sometimes refer to T as a *rooted tree*. If so, then every vertex x other than r has a unique *parent*; this is defined to be the (sole) neighbour p of x in the unique path in T from x to r, and x is said to be a *child* of p. An *ancestral order* of the vertices of a rooted tree T is an order of $V(T)$ in which the root vertex appears first and every non-root vertex appears later than its parent. Where it is clear from the context that a tree is oriented, we may refer to it simply as a tree. root rooted tree parent child, ancestral order

Let $A_1, A_2, \dots$ be a sequence of events. We say that A_n holds *asymptotically almost surely* if $\mathbb{P}(A_n) \to 1$ as $n \to \infty$. Likewise, in this thesis all occurrences of the standard asymptotic notation $\mathrm{o}(f)$ refer to sequences $f(n)$ with parameter n as $n \to \infty$. We will often have sets indexed by $\{1, 2, \dots, k\}$, such as $V_1, \dots, V_k$, and addition of indices will always be performed modulo k. Also, if $\varphi\colon A \to B$ is a function from A to B and $A' \subseteq A$, then we write $\varphi(A')$ for the image of A' under φ. We omit floors and ceilings whenever they do not affect the argument, write $a = b \pm c$ to indicate that $b - c \leq a \leq b + c$, and write abc/def for the fraction $(abc)/(def)$. For all $k \in \mathbb{N}$, (where $\mathbb{N} := \{1, 2, \dots\}$ is the set of natural numbers), we denote by $[k]$ the set $\{1, 2, \dots, k\}$, and write $\binom{S}{k}$ to denote the set of all k-element subsets of a set S. For any two disjoint sets A and B, we write $A \mathbin{\dot\cup} B$ for their union. We use the notation $x \ll y$ to indicate that for every positive y there exists a positive number x_0 such that for every $0 < x < x_0$ the subsequent statements hold. Such statements with more variables are defined similarly. We always write $\log x$ to mean the natural logarithm of x. asymptotically almost surely

2.1 Trees

In many stages of our proofs we will be required to partition a tree into subtrees, sometimes so that each piece contains a linear fraction of its edges, and sometimes so that each piece contains a linear fraction of some given subset of vertices.

Definition 2.1. Let T be a tree or oriented tree. A *tree-partition* of T is a collection $\{T_1, \dots, T_s\}$ of edge-disjoint subtrees of T such that tree-partition

$$\bigcup_{i\in[s]} V(T_i) = V(T) \qquad \text{and} \qquad \bigcup_{i\in[s]} E(T_i) = E(T).$$

Note that any two distinct trees in a tree-partition $\mathcal{P}$ have at most one vertex in common, and that if $\mathcal{P}$ contains at least 2 trees then each tree has at least one vertex in common with some other tree in $\mathcal{P}$. To obtain a partition of a tree we sometimes use a result of Kühn, Mycroft and Osthus.

Lemma 2.2. [51, Lemma 2.9] Let T be a tree on $n \geq 3$ vertices. Then there exists a tree-partition $\{T_1, T_2\}$ of T such that $e(T_1), e(T_2) \geq e(T)/3$.

We omit the proof of Lemma 2.2 because it is very similar to the proof of the next lemma—a tree-partition lemma for splitting a 'target' set of vertices (Lemma 2.3)—which we also use in our proofs.

Lemma 2.3. If T is a tree and $L \subseteq V(T)$, then T admits a tree-partition $\{T_1, T_2\}$ of T such that T_1 and T_2 each contain at least $|L|/3$ vertices of L.

Note that if T has at least one vertex, then some vertex lies in $V(T_1) \cap V(T_2)$. In particular, if $|L|$ is either 1 or 2, then it is always possible to form a tree-partition where (say) T_1 is a single vertex of L and $T_2 = T$.

For a tree T (possibly oriented), an edge $e \in E(T)$ and a vertex $v \in e$, we write $T - e$ for the forest we obtain by deleting e from T, and write C_v^e for the vertex set of the component of $T - e$ which contains v. We prove Lemma 2.3 using the following simple fact.

Fact 2.4. For all reals $c, x_1, \ldots, x_s$, if $\sum_{i=1}^{s} x_i \geq 3c$ and $0 \leq x_i \leq c$ for all $i \in [s]$, then there exists $i \in [s]$ such that $c \leq x_1 + \cdots + x_i \leq 2c$.

Proof of Lemma 2.3. Note that we may ignore the orientation of edges and assume that T is an undirected tree. Define $\ell := |L|$. For each edge $e = \{u, v\} \in E(T)$ we

heavy neighbour

say that v is a *heavy neighbour* of u if $|C_v^e \cap L| \geq \ell/3$. Observe that if u and v are both heavy neighbours of each other, then $\left\{T\left[C_u^e\right], T\left[C_v^e \cup \{u\}\right]\right\}$ is the desired tree-partition. We may therefore assume that for each edge $e = \{u, v\} \in E(T)$ either u is a heavy neighbour of v or v is a heavy neighbour of u, but not both. It follows that some vertex v has no heavy neighbours (to see this, form an auxiliary orientation of $E(T)$ with each edge directed $u \to v$ where v is a heavy neighbour of u, and choose v to be a vertex with no outneighbours). Let $C_1, \ldots, C_s$ be the vertex sets of the components of $T - v$. For each $i \in [s]$ let $\ell_i := |C_i \cap L|$, and observe that since v has no heavy neighbours we have $\ell_i < \ell/3$; since ℓ_i is an integer, we then have $\ell_i \leq (\ell - 1)/3$.

If $v \in L$, then $\ell_1 + \cdots + \ell_s = \ell - 1$ and by Fact 2.4 there exists $j \in [s]$

such that $(\ell-1)/3 \leq \ell_1+\cdots+\ell_j \leq 2(\ell-1)/3$. In this case the desired tree-partition is $\left\{T\left[\{v\}\cup\bigcup_{1\leq i\leq j}C_i\right], T\left[\{v\}\cup\bigcup_{j<i\leq s}C_i\right]\right\}$, because each of these subtrees contains v and hence contains at least $(\ell-1)/3+1>\ell/3$ vertices of L. On the other hand, if $v\notin L$, then $\ell_1+\cdots+\ell_s=\ell$, so by Fact 2.4 above there exists $j\in[s]$ with $\ell/3\leq \ell_1+\cdots+\ell_j\leq 2\ell/3$, and as before $\left\{T\left[\{v\}\cup\bigcup_{1\leq i\leq j}C_i\right], T\left[\{v\}\cup\bigcup_{j<i\leq s}C_i\right]\right\}$ is the desired tree-partition. □

Recall that an ancestral order of the vertices of a rooted tree T is an order of $V(T)$ in which the root vertex appears first and every non-root vertex appears later than its parent. This ancestral order is *tidy* if for any initial segment $\mathcal{I}$ of the order, at most $\log_2 n$ vertices in $\mathcal{I}$ have a child not in $\mathcal{I}$. (These orders were considered in [51].) tidy ancestral order

Tidy ancestral orders have applications for tree-traversal algorithms. Indeed, many such procedures operate on trees, one vertex at a time, so that (at any step) the visited vertices form a connected subtree; moreover, these procedures often keep a stack with vertices whose neighbours are yet to be visited; if we process vertices of a tree in a tidy ancestral order, keeping a stack of the vertices which have been processed but have at least one unvisited child, then the size of the stack is guaranteed to grow somewhat slowly. The following lemma states that every tree admits a tidy ancestral order.

Lemma 2.5. [51, Lemma 2.11] Every rooted tree T admits a tidy ancestral order. Moreover, if $\{T_1,T_2\}$ is a tree-partition of T such that $|T_1|\leq|T_2|$ and the root of T is the single vertex in $T_1\cap T_2$, then T admits a tidy ancestral order such that every vertex of T_1 precedes every vertex of T_2 in this order.

Proof. Let T be tree with root r and order n, and let $\{T_1,T_2\}$ be a tree-partition of T such that r is the (sole) vertex in $T_1\cap T_2$ (this tree-partition is arbitrary: take for example the single-vertex r and T as parts). We use induction in the order of T. The lemma is easy to check if T has at most 3 vertices. Suppose then that $n\geq 4$ and let $\{T_1,T_2\}$ be a tree-partition of T such that $|T_1|\leq|T_2|$ Let $H_1,\ldots,H_s$ be the components of $T-r$, labelled so that $H_1,H_2,\ldots,H_j$ lie in T_1 and $H_{j+1},\ldots,H_s$ lie in T_2, and moreover so that $|H_s|\geq|H_i|$ for all $i\in\{j+1,\ldots,s\}$. Finally, for each $i\in[s]$ let r_i be the vertex in H_i which is closest to r.

Note that for all $i\in[s-1]$ we have $|H_i|<(n-1)/2$, since $|T_1|\leq n/2$ and $|H_i|\leq|H_s|\leq n-1$; indeed, if $i\in[j]$ then $|H_i|+1\leq|T_1|\leq n/2$ and if $i\in[s-1]\setminus[j]$ then $2|H_i|\leq|H_i|+|H_s|\leq|T\setminus\{r\}|=n-1)$.

Also, by our choice of roots, we can combine ancestral orders $H_1,\ldots,H_s$: given any collection $\{\prec_i\}_{i\in[s]}$ such that $\prec_i$ is an ancestral order of H_i we can form an ancestral order $\prec$ of T such that the restriction of $\prec$ to H_i is $\prec_i$, as follows. First comes r; then come all vertices of H_1, ordered according to $\prec_1$, followed by all vertices of H_2 ordered according to $\prec_2$ and so on. Moreover, such combination preserves 'tidiness'. Indeed, suppose that $\prec_i$ is a tidy ancestral order of H_i and that $\prec$ is defined as described above. Let $\mathcal{I}$ be an initial segment $x = x_1, \ldots, x_t$ of $\prec$. If $x_t \in H_{t'}$ where $t' < s$, then the only vertices of $\mathcal{I}$ which have neighbours outside of $\mathcal{I}$ are r and the vertices in some initial segment $\mathcal{I}'$ of $H_{t'}$. Hence, by induction, at most $1 + \log_2 |H_{t'}| \leq \log_2 n$ vertices in $\mathcal{I}$ have neighbours outside $\mathcal{I}$. On the other hand, if $t' = s$, then the only vertices of $\mathcal{I}$ which have neighbours outside of $\mathcal{I}$ form an initial segment $\mathcal{I}'$ of H_s (note that the at all neighbours of r lie in $\mathcal{I}$), and so at most $\log_2 |H_s| < \log_2 n$ vertices in $\mathcal{I}$ have neighbours outside $\mathcal{I}$. □

Tree-partitions play an important role in the analysis of the allocation algorithm (Section 4.3), often through the following lemma. An important feature of the algorithm is that distant vertices are distributed almost independently of one another. We use Lemma 2.6 below to argue that any sufficiently large set of vertices will be well distributed by the algorithm: roughly speaking, the lemma states that in every tree with somewhat limited maximum degree most vertices are 'far apart'. We remark that this lemma is a strengthened version of [51, Lemma 2.10].

Lemma 2.6. For all $C, K > 0$ there exists n_0 such that for every rooted tree T of order $n \geq n_0$ with root r and $\Delta(T) \leq (\log n)^C$, there exist $s \in \mathbb{N}$, pairwise-disjoint subsets $F_1 \ldots, F_s \subseteq V(T)$, and not-necessarily-distinct vertices $v_1, \ldots, v_s$ of T with the following properties.

(i) $\left|\bigcup_{i\in[s]} F_i\right| \geq n - n^{5/12}$.

(ii) $|F_i| \leq n^{2/3}$ for each $i \in [s]$.

(iii) For each $i \in [s]$, each $x \in \{r\} \cup \bigcup_{j<i} F_j$, and each $y \in F_i$, the path from x to y in T includes v_i.

(iv) For any $i \in [s]$ and $y \in F_i$ we have $\mathrm{dist}_T(v_i, y) \geq (K \log\log n)^3$.

The original version of this lemma [51, Lemma 2.10] had constants $\Delta, \varepsilon, k > 0$ rather than $C, K > 0$, assumed additionally that $\Delta(T) \leq \Delta$, had $n - \varepsilon n$ in place of $n - n^{5/12}$ in (i) and had k in place of $K \log\log n$ in (iv). However, the form of the lemma given above can be established by an essentially identical proof, replacing each instance of k by $K \log\log n$ and each instance of Δ by $(\log n)^C$.

The crucial point is that we then replace the bound $3n^{1/3}\Delta^{k^3} \leq \varepsilon n$ by the bound $3n^{1/3}(\log n)^{C(K\log\log n)^3} \leq n^{5/12}$. These changes yield (i) and (iv) above, whilst (ii) and (iii) are unchanged.

Proof. We first partition T into a family $\mathcal{F}$ of subtrees, applying Lemma 2.2 repeatedly. More precisely, we start out with $\mathcal{F} = \{T\}$, and iterate the following step. Let T_{big} be the largest tree in $\mathcal{F}$; we stop if $|T_{\text{big}}| \leq n^{2/3}$; otherwise, let $\{T_1, T_2\}$ be a tree-partition of T_{big} such that $e(T_1), e(T_2) \geq e(T_{\text{big}})/3$ and replace T_{big} in $\mathcal{F}$ by T_1 and T_2. Note that this process stops after at most $3n^{1/3}$ steps (indeed, when we stop each tree $T' \in \mathcal{F}$ will have been obtained by partitioning a tree with at least $n^{2/3}$ vertices, so $|T'| \geq \frac{1}{3}n^{2/3}$). Moreover, this is a tree-partition, so any two trees in $\mathcal{F}$ share at most a vertex of T. Let $\prec$ be an ancestral order of T, and let $f(F)$ be the smallest vertex of F in this order, for each $F \in \mathcal{F}$. This induces an order $T_1, \ldots, T_s$ of $\mathcal{F}$ such that $f(T_i) \prec f(T_j)$ if and only if $i \leq j$.

For each $j \in [s] \setminus \{1\}$ let $T_{<j} := T\Big[\bigcup_{i<j} V(T_i)\Big]$. We first note that $T_{<j}$ is a connected tree, as this property is preserved by any tree-split we performed in $\mathcal{F}$. Secondly, note that $V(T_j) \cap V(T_{<j}) = v_j$ for each $j \in [s]$. Indeed there exists at least one vertex in this intersection, since each edge of T is contained in precisely one member of $\mathcal{F}$ and since $T_{<j+1}$ is connected. Furthermore, if x and y are distinct vertices $V(T_j) \cap V(T_{<j})$, then there exists a path P from x to y in $T_{<j}$ (since $T_{<j}$ is connected) and also a path Q from x to y in T_j (since T_j is connected). These paths are distinct, because $E(T_j) \cap E(T_{<j}) = \varnothing$; it follows that $P \cup Q$ contains a cycle, contradicting the fact that T is a tree. These two observations imply that for each $j \in [s]$, any path from $x \in V(T_{<j})$ to $y \in V(T_j)$ contains v_j.

For each $2 \leq j \leq s$, let $F_j := \Big\{ x \in V(T_j) : \text{dist}(v_j, x) \geq (K\log\log n)^3 \Big\}$. We now prove that the required properties hold. Items (ii) and (iv) hold by construction. Also note that $|V(T_j) \setminus F_j| \leq (\log n)^{CK(\log\log n)^3}$ and that $\Big|\bigcup_{j\in[s]} V(T_j)\Big| = n$. Then $|V(T) \setminus (F_1 \cup \cdots \cup F_s)| \leq 3n^{1/3}(\log n)^{CK(\log\log n)^3} < n^{5/12}$, so item (i) holds as well. Finally, for item (iii), let $j \in [s]$, $x \in \{x\} \cup \bigcup_{i<j} F_i$ and let $y \in F_j$. Then since $x \in V(T_{<j})$ and $y \in V(T_j)$ the path from x to y contains v_j as desired. $\square$

2.2 Regularity

Let G be a bipartite graph with vertex classes A and B. Loosely speaking, G is 'regular' if the edges of G are 'random-like' in the sense that they are distributed roughly uniformly. More formally, for any sets $X \subseteq A$ and $Y \subseteq B$, we write $G[X,Y]$ for the bipartite subgraph of G with vertex classes X and Y and whose edges are the edges of G with one endvertex in each of the sets X and Y, and define the *density* $d_G(X,Y)$ of edges between X and Y to be

density

$$d_G(X,Y) := \frac{e\Big(G[X,Y]\Big)}{|X||Y|}.$$

(d,ε)-regular

Then, for any $d, \varepsilon > 0$, we say that G is (d,ε)*-regular* if for every $X \subseteq A$ and every $Y \subseteq B$ such that $|X| \geq \varepsilon|A|$ and $|Y| \geq \varepsilon|B|$ we have $d_G(X,Y) = d \pm \varepsilon$. The following well-known proposition is immediate from this definition.

Lemma2.7 (Slicing lemma). Fix $\alpha, \varepsilon, d > 0$ and let G be a (d,ε)-regular bipartite graph with vertex classes A and B. If $A' \subseteq A$ and $B' \subseteq B$ have sizes $|A'| \geq \alpha|A|$ and $|B'| \geq \alpha|B|$, then $G[A',B']$ is $(d,\varepsilon/\alpha)$-regular.

$(d_{\geq},\varepsilon)$-regular

We say that G is $(d_{\geq},\varepsilon)$*-regular* if G is (d',ε)-regular for some $d' \geq d$. Another immediate consequence of the definition of regularity is that, for small ε, if G is (d,ε)-regular then almost all vertices of A have degree close to $d|B|$ in B and almost all vertices of B have degree close to $d|A|$ in A. We say that G is 'super-regular' if no vertex has degree much lower than this. More precisely, G is (d,ε)*-super-regular* if (A,B) is $(d_{\geq},\varepsilon)$-regular and also for every $a \in A$ and $b \in B$ we have $\deg(a,B) \geq (d-\varepsilon)|B|$ and $\deg(b,A) \geq (d-\varepsilon)|A|$.

(d,ε)-super-regular

To complete the embedding of a spanning oriented tree in a tournament, we will make use of the following well-known lemma, which states that every balanced super-regular bipartite graph contains a perfect matching (a bipartite graph is *balanced* if its vertex classes have equal size).

balanced

Lemma 2.8. If $d \geq 2\varepsilon > 0$ and G is a (d,ε)-super-regular balanced bipartite graph, then G contains a perfect matching.

Proof. Let A and B be the vertex classes of G; let $m := |A| = |B|$; let $S \subseteq A$ and write $N(S) \subseteq B$ for the set of vertices of G with a neighbour in S. If $|S| < \varepsilon m$, then for each $a \in S$ we have $\deg(a,B) \geq (d-\varepsilon)m \geq \varepsilon m > |S|$, so certainly $\Big|N(S)\Big| \geq |S|$. Otherwise, if $\varepsilon m \leq |S| \leq (1-\varepsilon)m$, then, since G is $(d_{\geq},\varepsilon)$-regular, at most εm vertices of B have no neighbours in S, so $\Big|N(S)\Big| \geq (1-\varepsilon)m \geq |S|$. Finally, if $|S| > (1-\varepsilon)m$, then each $b \in B$ has a neighbour in S, since $\deg(b,A) \geq (d-\varepsilon)m \geq \varepsilon m > |A \setminus S|$, so $\Big|N(S)\Big| = m \geq |S|$. In each case Hall's criterion holds (i.e., $\Big|N(S)\Big| \geq |S|$ for each subset $S \subseteq A$), so G contains a perfect matching. $\square$

The celebrated Regularity Lemma of Szemerédi [77, 78] states that every sufficiently large graph admits a partition such that almost all pairs of parts are regular in the sense we discuss here. Alon and Shapira [4] obtained a digraph analogue of their result.

Let G be a tournament, and let $X, Y \in V(G)$. We call the ordered pair (X, Y) a *directed pair* in G if there are no edges in $G[X \leftarrow Y]$, that is, if every edge between X and Y is directed towards Y. Similarly, for any $\mu \geq 0$ we call (X, Y) a *μ-almost-directed pair* if $e\Big(G[X{\leftarrow}Y]\Big) \leq \mu|X||Y|$, so any directed pair is a 0-almost-directed pair.

directed pair

μ-almost-directed pair

Observe that the underlying graph of $G[X \to Y]$ is a bipartite graph with vertex classes X and Y. We say that $G[X \to Y]$ is *(d, ε)-regular* (respectively *$(d_{\geq}, \varepsilon)$-regular* or *(d, ε)-super-regular*) to mean that this underlying graph is (d, ε)-regular (respectively $(d_{\geq}, \varepsilon)$-regular or (d, ε)-super-regular). In this way we may apply the previous results of this subsection to directed graphs.

(d, ε)-regular (digraph) $(d_{\geq}, \varepsilon)$-regular (digraph), (d, ε)-super-regular (digraph)

Lemma 2.9 (Regularity Lemma for digraphs). [4] For all positive ε, M' there exist M, n_0 such that if G is a digraph of order $n \geq n_0$ and $d \in [0, 1]$, then there exist a partition $V_0, \ldots, V_k$ of $V(G)$ and a spanning subgraph G' of G such that

(i) $M' \leq k \leq M$;

(ii) $|V_1| = \cdots = |V_k|$ and $|V_0| < \varepsilon n$;

(iii) $\deg^+_{G'}(x) \geq \deg^+_G(x) - (d + \varepsilon)n$ for all $x \in G$;

(iv) $\deg^-_{G'}(x) \geq \deg^-_G(x) - (d + \varepsilon)n$ for all $x \in G$;

(v) for all $i \in [k]$ the digraph $G'[V_i]$ has no edges;

(vi) for all distinct i, j with $1 \leq i, j \leq k$, the pair (V_i, V_j) is ε-regular and has density 0 or density at least d in $G'[V_i{\to}V_j]$.

We refer to the sets $V_1, \ldots, V_k$ as the *clusters* of G. For $d \in [0, 1]$, the *reduced graph R* with parameters ε, d and M' of G is a digraph we obtain by applying Lemma 2.9 to G with parameters ε, d and M'; the digraph R has vertex set $[k]$ and edges $i{\to}j$ precisely when $G'[V_i{\to}V_j]$ has density at least d.

clusters, reduced graph

2.3 Tournaments

A *tournament* G is an orientation of a complete graph, and a *subtournament* of G is a sub(di)graph of G which is a tournament. A *regular tournament* is a tournament in which every vertex has equal indegree and outdegree; it is easily checked that regular tournaments of order n exist for every odd $n \in \mathbb{N}$. A *transitive tournament* is a tournament whose vertices can be ordered $v_1, \ldots, v_n$ such that $v_i \to v_j$ is an

tournament, subtournament regular tournament

transitive tournament

edge for each $i < j$.

The following straightforward lemma shows that a tournament can only have a few vertices of small in- or outdegree.

Lemma 2.10. For each $d \in \mathbb{N}$, every tournament contains at most $4d - 2$ vertices with semidegree less than d.

Proof. Let G be a tournament, and let X be the set of vertices $x \in V(G)$ with $\deg^+(x) \leq d - 1$. Then

$$\binom{|X|}{2} = e\Big(G[X]\Big) \leq \sum_{x \in X} \deg^+(x) \leq (d-1)|X|,$$

where the central inequality holds because every edge of $G[X]$ contributes one to the sum. It follows that $|X| \leq 2d - 1$, that is, there are at most $2d - 1$ vertices with outdegree less than d. Essentially the same argument shows that there are at most $2d - 1$ vertices with indegree less than d, so in total at most $4d - 2$ vertices have semidegree less than d. □

2.4 Useful estimates and bounds

In this section we present various useful estimates, which we use many times throughout this thesis to analyse randomised procedures. We write $\mathcal{B}(n, p)$ to denote the binomial distribution (the result of n independent Bernoulli experiments, each with success probability p). We write $\mathbb{P}(A)$ for the probability of the event A and $\mathbb{E}X$ (respectively $\operatorname{Var} X$) for the expectation (respectively variance) of the random variable X.

Lemma 2.11. Suppose that $1/n \ll 1/k$. If $X := \mathcal{B}\Big(n, \frac{1}{2}\Big)$, then for every $r \in [k]$ we have

$$\mathbb{P}(\, X \equiv r \mod k\,) = \frac{1}{k} \pm \frac{4}{\sqrt{n}}.$$

Proof. Define $p_\mu := \max_{x \in \{0,\ldots,n\}} \mathbb{P}(X = x)$. Kühn, Mycroft and Osthus [51, proof of Lemma 2.1] gave a straightforward proof that $\mathbb{P}(\, X \equiv r \mod k\,) = \frac{1}{k} \pm 2p_\mu$, and the result then follows from a standard estimate on the binomial distribution (see, for example, [11, Section 1.2]) which states that $p_\mu \sim 1/\sqrt{\pi n/2}$. □

Another useful tool in probabilistic combinatorics (used extensively in Chapters 3 and 4) is Chebyshev's inequality (see, e.g., [5, Section 4.1]).

Theorem 2.12 (Chebyshev's inequality). For each positive λ, if X is a real-valued random variable then

$$\mathbb{P}\Big(|X - \mathbb{E}X| \geq \lambda \operatorname{Var} X\Big) \leq \frac{1}{\lambda^2}.$$

We also use the Chernoff bounds below (see, e.g., [45, Theorem 2.1]).

Theorem 2.13. For all $t \geq 0$ and all p such that $0 < p < 1$, if $X \coloneqq \mathcal{B}(n, p)$, then

$$\mathbb{P}(X \geq np + t) \leq \exp\left(-\frac{t^2}{2(np + t/3)}\right); \quad \text{and} \tag{2.1}$$

$$\mathbb{P}(X \leq np - t) \leq \exp\left(-\frac{t^2}{2np}\right) \tag{2.2}$$

Let N be an n-element set, and let $M \in \binom{N}{m}$ be a subset of N with m elements. If we choose $S \in \binom{N}{k}$ uniformly at random, then the random variable $X = |S \cap M|$ is said to have *hypergeometric distribution* with parameters n, m and k. Note that the expectation of X is then $\mathbb{E}X = km/n$.

hypergeometric distribution

Theorem 2.14. [45, Corollary 2.3 and Theorem 2.10] For every $0 < a < 3/2$, if X has binomial or hypergeometric distribution, then

$$\mathbb{P}\Big(\,|X - \mathbb{E}X| \geq a\mathbb{E}X\,\Big) \leq 2\exp(-a^2\mathbb{E}X/3).$$

We use the following Azuma-type concentration result for martingales, due to Mc Diarmid [63] (and in the form stated by Sudakov and Vondrák [74]), to analyse some randomised algorithms (Lemmas 3.3 and 4.15).

Lemma 2.15. [63, 74] Fix $n \in \mathbb{N}$ and let $X_1, \ldots, X_n$ be random variables taking values in $[0, 1]$ such that for each $i \in [n]$ we have $\mathbb{E}(\,X_i \mid X_1, \ldots, X_{i-1}\,) \leq a_i$. If $\mu \geq \sum_{i=1}^{n} a_i$, then for every δ with $0 < \delta < 1$ we have

$$\mathbb{P}\Big(\sum_{i=1}^{n} X_i > (1+\delta)\mu\Big) \leq \mathrm{e}^{-\delta^2\mu/3}.$$

We also use the well-known Cauchy–Bunyakovsky–Schwarz inequality.

Theorem 2.16. All real $u_1, v_1, \ldots, u_n, v_n$ satisfy

$$\left(\sum_{i=1}^{n} u_i v_i\right)^2 \leq \left(\sum_{i=1}^{n} u_i^2\right)\left(\sum_{i=1}^{n} v_i^2\right),$$

with equality if and only if (a) $u_1 = \cdots = u_n = 0$ or (b) $v_1 = \cdots = v_n = 0$ or (c) there exists $\alpha \neq 0$ such that $u_i = \alpha v_i$ for all $i \in [n]$.

When this is a strict inequality, we will be interested in the difference

$$\Delta_{\mathrm{err}} \coloneqq \left(\sum_{i=1}^{n} u_i^2\right)\left(\sum_{i=1}^{n} v_i^2\right) - \left(\sum_{i=1}^{n} u_i v_i\right)^2 = \sum_{1 \leq i,j \leq n} (u_i^2 v_j^2 - u_i v_i u_j v_j).$$

Note that

$$\Delta_{\mathrm{err}} = \frac{1}{2} \sum_{1 \leq i,j \leq n} (u_i^2 v_j^2 + u_j^2 v_i^2 - 2 u_i v_i u_j v_j) = \frac{1}{2} \sum_{1 \leq i,j \leq n} (u_i v_j - u_j v_i)^2. \tag{2.3}$$

2.5 Homomorphisms, allocation, embedding

Our strategy for finding a spanning tree T in a digraph G consists grosso modo of two phases: allocation (assigning pieces of T to pieces of G) and embedding (refining the assignment into an embedding). Correspondingly, at the core of the proofs of the next chapters lie two algorithms, which we introduce here.

Let H and G be digraphs. A homomorphism $\varphi : H \to G$ is an edge-preserving map from $V(H)$ to $V(G)$, so that every edge $u{\to}v \in E(H)$ is mapped to an edge $\varphi(u){\to}\varphi(v) \in E(G)$. The *$\varphi$-indegree* $\deg^-_\varphi(v)$ of $v \in H$ in G is $\left|\varphi\left(N^-_H(v)\right)\right|$; the *$\varphi$-outdegree* $\deg^+_\varphi(v)$ of v is defined similarly. The *maximum degree* of φ is $\Delta(\varphi) \coloneqq \max_{v \in H} \deg^-_\varphi(v) + \deg^+_\varphi(v)$. Note that vertices in $\varphi\left(N^-_H(v)\right) \cap \left(\varphi(N^+_H(v)\right)$ are counted twice.

φ-indegree

φ-outdegree, maximum degree

Let R be a digraph of order k which is a 'reduced graph' of G, where $1/n \ll 1/k$, so each vertex of R corresponds to a set of approximately n/k vertices of G and the edges of R correspond to regular pairs of clusters of G. If T is a tree whose order is close to n, then it is natural to look for copies of T in G by mapping many edges of T 'along' edges of R; in other words, to seek a homomorphism $\varphi : T \to R$. This is precisely the role of the allocation algorithm. If this can be done, then we embed T to G using a (deterministic) embedding algorithm which 'follows' the homomorphism embedding in turn each vertex $x \in T$ to a vertex in the cluster $\varphi(x)$, relying on the fact edges of R correspond to regular pairs.

Recall that in our applications G is only (if at all) slightly larger than T. Hence, for the above strategy to succeed we need to guarantee that vertices of T

are well distributed among clusters of G, and that the embedding avoids occupying too many neighbours of a vertex at any step.

2.5.1 Allocation

We first consider the problem of defining a homomorphism φ from a large oriented tree T to a small digraph R. If T contains a large directed path P, say, with $|P| \gg |R|$, then we must require that R contains a strongly connected component (or, equivalently, a component with semidegree at least 1), since otherwise the homomorphism image of P must be an isomorphic to P (which is not possible). However, if we insist that φ maps roughly equally many vertices to each vertex $x \in R$ the problem becomes more subtle. In this case R itself needs to be strongly connected—or we might get 'trapped' in some strongly connected component of R as in the example above—but this alone is not sufficient, as can be seen when T is an antidirected path and R is a long directed cycle.

Indeed, the main challenge in allocation is ensuring that the homomorphism we produce maps the correct number of vertices of T to each cluster. We define the homomorphism using variants of a simple randomised algorithm (described below), which yields the desired allocation when R satisfies some expansion or connectivity properties.

The two specific variants we use are Algorithms 3.2 and 4.14. Significantly, the reduced digraphs R we have in Chapters 3 and 4 are quite different, and for this reason we postpone the precise description of the allocation algorithms and analysis to the corresponding chapters.

Randomised allocation algorithm. Roughly speaking, we process the vertices of a rooted tree T in an ancestral order. When processing any vertex x (other than the root), the parent p of x in T is the only neighbour of x for which φ has been defined. Let C_p be the children of p. We define φ for all vertices in C_p at once, as follows: first choose an inneighbour v^- and an outneighbour v^+ of $\varphi(p)$ in R uniformly at random, with choices made independently of all other choices in the algorithm; then map each child inneighbour of p to v^- and each child outneighbour of p to v^+.

Note that this algorithm still fails if T is a star, in which case the centre of T is the only vertex mapped to some cluster. However, as the maximum degree of T is not too high we are able to avoid this problem. Indeed, in Lemma 2.6 we show that if $\Delta(T) \leq (\log |T|)^C$, where $1/n \ll 1/C$, then most vertices of T lie far away from one another. This lemma plays a key role in the proof of Lemmas 3.3 and 4.15.

2.5.2 Embedding

We now suppose that we are given a rooted tree T and a slightly larger tournament G, as well as a regular partition of G and a corresponding reduced graph R. Moreover, suppose that there exists an 'allocation' $\varphi : V(T) \to V(R)$ of each vertex $x \in T$ to a cluster of G. We shall embed the vertices of T one at a time, in an ancestral order, greedily embedding each vertex $x \in T$ to some vertex $\phi(x)$ in $\varphi(x)$.

The main difficulty in doing so is to avoid 'stepping on our own toes', i.e., fully occupying the neighbourhood of $\phi(x)$ before all neighbours of x are embedded—this is somewhat straightforward while a constant fraction of $V(G)$ remains unoccupied, but becomes increasingly difficult once the number of embedded vertices passes this threshold. Suppose first that we wish to embed T to G, where T is almost-spanning (i.e., $|T| = (1-\varepsilon)|G|$ where $1/n \ll \varepsilon$).

Embedding algorithm. We choose an ancestral order of T and then embed each $x \in T$ greedily to the cluster $\varphi(x)$; at each step we reserve a set of vertices for the children of x. (An important issue here is that not too many vertices can lie in reserved sets: this is handled by requiring that vertices be processed in a *tidy* ancestral order, see 2.1.)

If on the other hand $|T| = |G|$, then embedding algorithm alone cannot completely embed T. We follow a standard approach to handle this difficulty. Firstly, we reserve a small set $X \subseteq V(G)$ of vertices to be used at the end; secondly, we delete a set $L \subseteq V(T)$ from T, so that $|L| - |X|$ is sufficiently large and so that the edges incident to vertices in L have some nice properties; next we apply the embedding algorithm to $G - X$ and $T - L$ (this succeeds since there is 'room to spare'); we complete the embedding by reintegrating X using a matching-type result (such as Lemma 2.8).

We describe the embedding algorithm in Section 4.3.4 (page 94). A very similar variant is used, in a simpler setting, in Section 3.3.2 (page 37).

3 Spanning Trees of Tournaments

This chapter is organised as follows. In Section 3.1 we outline the proof of our main result of this chapter, namely that almost every tree is unavoidable. Next, in Section 3.2 we give definitions and preliminary results which we will use later on. These include structural results for tournaments. In Section 3.3 we prove Theorem 1.9 (‘trees with polylogarithmic maximum degree are almost unavoidable’) by considering the ‘random embedding algorithm’ of Kühn, Mycroft and Osthus [51] and explaining how to modify the proofs of the associated results to obtain slightly sharper bounds; we also use these sharper bounds when considering tournaments with a specific structure (cycles of cluster tournaments, see Section 3.1). In Section 3.4 we consider tournaments G whose vertex set can be partitioned into two large sets which form an almost-directed pair in G, proceeding as outlined in the proof sketch below to show that every such tournament contains a copy of every nice oriented tree of polylogarithmic maximum degree (this is Lemma 3.18, which can be interpreted as proving Theorem 1.4—a sufficient condition for unavoidability—for such tournaments). Then, in Section 3.5 we do the same for tournaments G which contain an almost-spanning cycle of cluster tournaments (Lemma 3.24), making use of the sharper estimates established in Section 3.3. In Section 3.6 we prove Theorem 1.4 by combining the results of Sections 3.4 and 3.5. We also give the proof of Theorem 1.6 (almost every tree is nice).

3.1 Proof outline for Theorem 1.4

Roughly speaking, there are two main differences between our results and those in [51]. Firstly, we require several (minor) modifications in their results to deal with trees with larger maximum degree (these are mostly consequences of changes in Lemmas 2.6 and 3.3), culminating with the proof of Theorem 1.9. Secondly, because we are concerned with spanning trees of every tournament (rather than almost-spanning trees), we need to proceed more carefully in order to guarantee that the later stages of the embedding will succeed—we develop two distinct strategies to address this problem, according to the structure of the tournament looks like.

Indeed, our proof of Theorem 1.4 uses a structural characterisation of large tournaments (Lemma 3.1) from [51]. Loosely speaking, this shows that every large tournament G has one of the following two possible structures. The first possibility is that $V(G)$ can be partitioned into two sets U and W such that almost all edges of G between U and W are directed from U to W. We refer to such a structure as an 'almost-directed pair'. The second possibility is that $V(G)$ contains disjoint subsets $V_1, \dots, V_k$ of equal size called 'clusters' whose union includes almost all vertices of G and such that the edges of G directed from V_i to V_{i+1} (with addition taken modulo k) are 'random-like'. We refer to this structure as a 'cycle of cluster tournaments'. Given a tournament G on n vertices and a nice oriented tree T on n vertices with polylogarithmic maximum degree we consider separately these two cases for the structure of G.

Almost-directed pairs. Suppose that G admits an almost-directed pair (U, W). In this case we begin by identifying the set Z of 'atypical' vertices of G, namely those which lie in too many edges directed 'the wrong way' (i.e., from W to U). Since (U, W) is an almost-directed pair Z must be small. We then choose a set S of $|Z|$ distinct vertices of T, each of which lies in an out-star of T and is adjacent to both an in-leaf and an out-leaf of T (see Figure 1.4 on page 4). We also choose a small set S^- of vertices of T, each of which lies in an in-star of T and is adjacent to an out-leaf of T, and a small set S^+ of vertices of T, each of which lies in an out-star of T and is adjacent to an in-leaf of T. The fact that T is nice ensures that we can choose such sets. Having done so, we form a subtree T' of T by removing one out-leaf adjacent to each vertex in S^-, one in-leaf adjacent to each vertex in S^+, and one in-leaf and one out-leaf adjacent to each vertex in S. We then embed T' in G; this can be achieved by *ad hoc* methods (Lemma 3.17) using the fact that G has slightly more vertices than T' to give us a little 'room to spare'. Moreover, we can insist that the image P^- of S^- under this embedding has $P^- \subseteq U$, and likewise that the image P^+ of S^+ has $P^+ \subseteq W$.

It then suffices to embed the removed leaves into the set $Q \subseteq V(G)$ of vertices of G not covered by the embedding of T'. To do this, we first embed the removed leaves adjacent to vertices of S so as to cover the set Z of atypical vertices of G. This is achieved as follows. Let b be an atypical vertex of G, choose a vertex $s \in S$, and let s^+ and s^- be the removed out-leaf and in-leaf (respectively) adjacent to s. Since s is a vertex of T', s has already been embedded in G, say to a vertex x. Let x^+ be an outneighbour of x in Q, and let x^- be an inneighbour of x in Q (our embedding of T' in G will ensure that such vertices exist). Since G is a tournament, we must have either an edge $b \to x$ or $x \to b$ in G. In the former case we embed s^- to b and s^+ to x^+, and in the latter case we embed s^+ to b and s^- to x^-; either way we have extended our embedding to cover the atypical vertex b.

Once we have taken care of all atypical vertices in this manner, let $Q^- \subseteq U$ and $Q^+ \subseteq W$ be the sets of uncovered vertices in U and W, respectively. The only vertices of T not yet embedded are the removed neighbours of vertices in $S^- \cup S^+$. We now use the fact that all vertices of Q^- and Q^+ are typical to find perfect matchings in the graphs $G[P^- \to Q^+]$ and $G[Q^- \to P^+]$ (our embedding of T' in G will ensure for this that we have $|P^-| = |Q^+| = |P^+| = |Q^-|$). Recall that each $s \in S^-$ was embedded to some vertex $p \in P^-$, which is matched to some $q \in Q^+$; we embed the removed outneighbour of s to q. Likewise, each $s \in S^+$ was embedded to some vertex $p \in P^+$, which is matched to some $q \in Q^-$; we embed the removed inneighbour of s to q. This completes the embedding of T in G.

Cycles of cluster tournaments. Suppose that G contains an almost spanning cycle of cluster tournaments with clusters $V_1, \dots, V_k$ of equal size. Again we begin by identifying the small set B of atypical vertices, which in this case are those vertices in some cluster V_i which have atypically small inneighbourhood in V_{i-1} or atypically small outneighbourhood in V_{i+1}, as well as those vertices not contained in any cluster V_i. We also choose a small set L of vertices of T each of which is adjacent to at least one in-leaf and at least one out-leaf of T (this is possible since T is nice). Following this we split T into subtrees T_1 and T_2 which partition the edge-set of T and have precisely one vertex in common, so that T_1 and T_2 each contain many vertices of L. Next we form subtrees T_1' and T_2' of T_1 and T_2 respectively by removing one in-leaf and one out-leaf adjacent to each vertex of L. Finally, we embed T into G by the following two steps.

First, we embed T_1 in G so that all atypical vertices are covered and also so that the number of vertices of T_1 embedded in each cluster V_i is approximately equal (more specifically, with an additive error on the order of $\frac{n}{\log\log n}$). To do this, we apply a 'random embedding algorithm' of Kühn, Mycroft and Osthus [51] to embed T_1' into G so that approximately the same number of vertices of each cluster are covered and also so that roughly the same number of vertices of L are embedded to each cluster. (In fact, at this point we use slightly sharper estimates on the numbers of vertices embedded in each cluster than those given in [51]; these arise from the same proofs). Then, by a similar argument to that used for covering atypical vertices in the previous case, for each $i \in [k]$ and each vertex $x \in L$ which was embedded in the cluster V_i we may use the fact that $G[V_i]$ is a tournament to choose an atypical vertex b and an uncovered vertex $y \in V_i$ so that the removed inneighbour and outneighbour of x can be embedded to b and y. This gives the desired embedding of T_1 in G.

Secondly, to complete the embedding of T in G we embed T_2 into the uncovered vertices of G (except for the single common vertex of T_1 and T_2 which is already embedded). For this we again apply the random embedding algorithm to embed T_2' in G with approximately the same number of vertices embedded within each cluster. We then carefully embed the removed inneighbours and outneighbours of a small number of vertices of L to achieve the following property. Let $U_i \subseteq V_i$ be the set of vertices of V_i which remain uncovered, and let $P_i \subseteq V_i$ be the image of vertices of L embedded to V_i whose removed inneighbour and outneighbour have not yet been embedded. We ensure that $2|P_1| = \cdots = 2|P_k| = |U_1| = \cdots = |U_k|$. Having done so, we partition each set U_i into two equal-size parts U_i^- and U_i^+, and use the fact that all vertices which remain uncovered are typical to find perfect matchings in $G[U_{i-1}^- \to P_i]$ and $G[P_i \to U_{i+1}^+]$ for each $i \in [k]$. Then, for each vertex x in L whose removed inneighbour and outneighbour have not yet been embedded, let $p \in P_i$ be the vertex to which x was embedded, and let q^- and q^+ be the vertices to which p is matched in U_{i-1} and U_{i+1} respectively. We may then embed the removed inneighbour and outneighbour of x to q^- and q^+ respectively; doing so for every $x \in L$ completes the embedding of T in G.

3.2 Preliminaries

The following structure plays a key role in our proof. Let d and ε be positive real numbers, and let G be a digraph whose vertex set is the disjoint union of sets $V_1, \dots, V_k$. We say that G is a (d,ε)*-regular cycle of cluster tournaments* if for each $i \in [k]$ the induced subgraph $G[V_i]$ is a tournament and the digraph $G[V_i \to V_{i+1}]$ is $(d_{\geq}, \varepsilon)$-regular (where addition on the subscript is taken modulo k). Likewise, we say that G is a (d,ε)*-super-regular cycle of cluster tournaments* if for each $i \in [k]$ the induced subgraph $G[V_i]$ is a tournament and the digraph $G[V_i \to V_{i+1}]$ is (d,ε)-super-regular. In either case we refer to the sets $V_1, \dots, V_k$ as the *clusters* of G.

(d,ε)-regular cycle of cluster tournaments

(d,ε)-super-regular cycle of cluster tournaments

clusters

The following lemma, a combination of two lemmas from [51] about so-called 'robust outexpanders', states that every tournament with large minimum semidegree either admits a partition $\{S, S'\}$ where S and S' are not too small and (S, S') is an almost-directed pair, or contains an almost-spanning cycle of cluster tournaments.

Lemma 3.1. [51, Lemmas 2.7 and 2.8] Suppose that $1/n \ll 1/k_1 \ll 1/k_0 \ll \varepsilon \ll d \ll \mu \ll \nu \ll \eta$, and let G be a tournament on n vertices. Then either

(a) $\delta^0(G) < \eta n$;

(b) G contains a spanning μ-almost-directed pair (S, S') with $|S|, |S'| > \nu n$; or

(c) G contains a (d, ε)-regular cycle of cluster tournaments G' with k clusters of equal size, such that $|G'| \geq (1 - \varepsilon)n$ and $k_0 \leq k \leq k_1$.

3.3 An approximate result (Theorem 1.9)

In this section we show how to obtain somewhat sharper estimates for the random allocation and embedding algorithms in [51] to embed oriented trees in slightly larger tournaments.

3.3.1 Allocation around a cycle of cluster tournaments

We begin with the random allocation algorithm of Kühn, Mycroft and Osthus [51], which is presented below as Algorithm 3.2 (page 36). Given a rooted oriented tree T and a cycle of cluster tournaments G with clusters $V_1, V_2, \ldots, V_k$, this assigns each vertex of T to a cluster of G. We allocate vertices of T one at a time in an ancestral order. This ensures that whenever we allocate a vertex x other than the root, the parent p of x has previously been allocated to some cluster V_i. We then say that x is allocated *canonically* if either $p \to x \in E(T)$ and x is allocated to the cluster V_{i+1}, or $p \leftarrow x \in E(T)$ and x is allocated to the cluster V_{i-1}. Moreover, we say that an allocation of the vertices of T to the clusters of G is *semi-canonical* if every vertex of T is either allocated canonically or allocated to the same cluster as its parent, every vertex adjacent to the root of T is allocated canonically, and for each $i \in [k]$ the set U_i of vertices allocated to V_i induces a forest $F = T[U_i]$ in which no connected component has more than $\Delta(T)$ vertices.

canonically

semi-canonical

The following lemma, a slightly modified version of [51, Lemma 3.3], states that Algorithm 3.2 will always return a semi-canonical allocation, and moreover that if T is sufficiently large then the allocation of vertices to clusters will be approximately uniform.

Algorithm 3.2: The Vertex Allocation Algorithm [51]

Input : an oriented tree T with root r and clusters $V_1, \ldots, V_k$.

Choose an ancestral order $t_1, \ldots, t_n$ of $V(T)$ (so in particular $t_1 = r$).
for $\tau = 1$ **to** n **do**
  if $\tau = 1$ **then** allocate r to V_1.
  else
    Let t_σ be the parent of t_τ.
    if $\mathrm{dist}_T(t_\tau, r)$ *is odd* **then** allocate t_τ canonically.
    else Allocate t_τ to the same cluster as t_σ with probability $1/2$ and allocate t_τ canonically with probability $1/2$, independently of all previous choices.

Lemma 3.3. Let T be an oriented tree on n vertices rooted at r. If we 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 will be semi-canonical.

(b) Let u and v be distinct vertices of T such that u lies on the path from r to v, let P be the path between u and v, and let $E \subseteq V(T)$ consist of all vertices $x \in V(P) \setminus \{u\}$ for which $\mathrm{dist}(r, x)$ is even. If we condition on the event that u is allocated to some cluster V_j, then v is allocated to cluster V_{j+R+F} (taking addition in the subscript modulo k) where $R \coloneqq \mathcal{B}\Big(|E|, \frac{1}{2}\Big)$ and F is a deterministic variable depending only on $\mathrm{dist}(r, u)$ and the orientations of edges of P (that is, F is unaffected by the random choices made by the Vertex Allocation Algorithm).

(c) Suppose that $1/n \ll 1/k$. Let u and v be vertices of T such that u lies on the path from r to v, and $\mathrm{dist}_T(u, v) \geq (\log\log n)^3$. Then for any $i, j \in [k]$,

$$\mathbb{P}(\, v \text{ is allocated to } V_i \mid u \text{ is allocated to } V_j \,) = \frac{1}{k}\left(1 \pm \frac{1}{4\log\log n}\right).$$

(d) Suppose that $1/n \ll 1/k, \alpha, 1/C$ and that $\Delta(T) \leq (\log n)^C$. Let S be a subset of $V(T)$ with at least αn vertices. Then with probability $1 - o(1)$ each of the k clusters V_i has $|S|\Big(\frac{1}{k} \pm \frac{1}{\log\log n}\Big)$ vertices of S allocated to it.

The statement above differs from the original version of the lemma in the following ways. Firstly, (b) was not stated explicitly, but was established in the original proof. Secondly, the original version of (c) instead had constants $1/k \ll \delta$, assumed that $\mathrm{dist}_T(u, v) \geq k^3$ instead of $\mathrm{dist}_T(u, v) \geq (\log\log n)^3$, and had δ in place of $\frac{1}{\log\log n}$ in the displayed equation. Finally, the original version of (d) had constants $1/n \ll 1/\Delta, 1/k \ll \delta$, assumed instead that $\Delta(T) \leq \Delta$, had δ in place of $\frac{1}{\log\log n}$, was only stated for the special case $S = V(T)$, and only provided an upper bound on the number of vertices allocated to each cluster. So our version of the lemma allows the bounds in (c) and (d) to decrease with n, and $\Delta(T)$ to grow with n, rather than being fixed constants. We now show how the original proof can be modified to establish our altered versions of (c) and (d). We include the proof for completeness.

Proof. We first observe that the algorithm allocates each edge either canonically or within a cluster. Moreover, every edge incident to r must be allocated canonically since the neighbours of r lie at odd distance from it; finally, any component allocated within a cluster is a star of order at most $\Delta(T)$ composed of a (non-root) vertex and some of its neighbours, so (a) holds.

Call an edge of P *odd* if its endpoint which lies farthest from r is at odd distance from r; we call the edge *even* otherwise. To prove (b), let $F := f_{\text{odd}} - b_{\text{odd}} - b_{\text{even}}$, where f_{odd} is the number of odd edges of P directed towards v and b_{odd} is the number of odd edges of P directed towards u and b_{even} is the number of even edges of P directed towards u. Let R be the sum of the number of even edges of P which are allocated canonically and the number of even edges of P which are *not* allocated canonically. Then R is a random variable with binomial distribution $\mathcal{B}\Big(|E|, 1/2\Big)$, where E is the set of even edges of P. Since the algorithm allocates vertices in an ancestral order, u is the first vertex of P to be allocated, and hence (conditioned to u being allocated to V_j) the cluster to which v is allocated is V_{j+F+R} as desired. odd even

To prove (c), let $\ell := \text{dist}_T(u, v)$, and define E as in (b), so $|E| = \lfloor \frac{\ell}{2} \rfloor$ or $|E| = \lceil \frac{\ell}{2} \rceil$. By (b) it suffices to show that $\mathbb{P}\Big(\mathcal{B}\Big(|E|, \frac{1}{2}\Big) = r \mod k\Big) = \frac{1}{k} \pm \frac{1}{4k\log\log n}$ for each $r \in [k]$, and since $|E| \geq \frac{1}{3}(\log\log n)^3$ this holds by Lemma 2.11.

The proof of (d) is identical to the proof of Lemma 4.15 (d) (taking $\xi = 1/2$ and using the estimates above), and so is omitted. □

3.3.2 Embedding around a cycle of cluster tournaments

Having applied the random allocation algorithm to allocate the vertices of T (an oriented tree) to G (a cycle of cluster tournaments which is slightly larger than T), Kühn, Mycroft and Osthus proceeded to embed T in G using a vertex embedding algorithm which successively embedded vertices of T in G following an ancestral order of the vertices of T, with each vertex being embedded in the cluster to which is was allocated. Studying this algorithm yields the following lemma, which is a modified form of [51, Lemma 3.4].

Lemma 3.4. Suppose that $1/n \ll 1/C$ and that $1/n \ll 1/k \ll \varepsilon \ll \gamma \ll d \ll \alpha$, and let $m := n/k$.

(1) Let T be an oriented tree on at most n vertices with root r and $\Delta(T) \leq (\log n)^C$.

(2) Let G be a (d, ε)-regular cycle of cluster tournaments on clusters $V_1, \ldots, V_k$, each of size at least $(1+\alpha)m$ and at most $3m$, and let v be a vertex of V_1 with 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 $V_1, \ldots, V_k$ so that at most $(1+\alpha/2)m$ vertices are allocated to each cluster V_i, and so that the allocation is semi-canonical.

Then G contains a copy of T in which r is embedded to v, and such that each vertex is embedded in the cluster to which it was allocated.

The differences between Lemma 3.4 as stated above and the original version in [51] are twofold. Firstly, the original assumption that $\Delta(T) \leq \Delta$ for some (fixed) Δ with $1/n \ll 1/\Delta \ll \varepsilon$ has been replaced by our assumption that $\Delta(T) \leq (\log n)^C$. Secondly, we allow the cluster sizes to vary between the bounds in (2), whereas the original form insisted that all clusters have size exactly $(1+\alpha)n$. Neither of these changes materially affects the original proof given in [51]. We include a sketch of the proof of this result, after introducing some auxiliary results and definitions.

We note that the upper bound $3m$ on the size of clusters in Lemma 3.4 (2) is not strictly needed, but allows for a simpler proof.

We require an upper bound on the size of a tournament containing an arbitrary tree of order n. Below we state the best known bound which holds for all n is due to El Sahili (but any linear bound would suffice).

Theorem 3.5. [29, Corollary 2] Every oriented tree on n vertices is contained in every tournament on $3n-3$ vertices.

Let G be a digraph such that $V(G)$ is the disjoint union of sets $V_1, \ldots, V_k$, each of size m. We say that $S \subseteq V_i$ is (c,γ)-*good* if for all $V'_{i-1} \subseteq V_{i-1}$ and all $V'_{i+1} \subseteq V_{i+1}$ with $|V'_{i-1}| \geq cm$ and $|V'_{i+1}| \geq cm$ there exist $\gamma\sqrt{m}$ vertices in S which each have at least γm inneighbours in V'_{i-1} and at least γm outneighbours in V'_{i+1}. The next lemma guarantees the presence of good sets in any small subset of a cluster.

(c,γ)-good

Lemma 3.6. [51, Lemma 2.5] Suppose that $1/m \ll \varepsilon \ll \gamma \ll c, d$. Let G be an (d, ε)-regular cycle of cluster tournaments on clusters $V_1, \dots, V_k$, each of size m. Then for any $i \in [k]$ 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 $|S| \leq \sqrt{m}$.

With these tools in place, we may describe how to embed T. Suppose $C > 0$ and let T be an oriented tree on n vertices with root r, such that $\Delta(T) \leq (\log n)^C$. Let G be a regular cycle of cluster tournaments whose clusters $V_1, \dots, V_k$ contain each slightly more than $m := n/k$ vertices. Finally, suppose we are given a semi-canonical allocation of the vertices of $V(T)$ to G. Our goal is to embed T to G according to this allocation.

As outlined in Section 3.1, we shall use a greedy algorithm (described below), which proceeds by embedding at each step a component of T formed by edges allocated within some cluster. More precisely, we form the *canonical tree* T_{canon} by contracting the edges of T which were allocated within a cluster. Each component contains at most $\Delta(T) \leq (\log n)^C$ vertices—we say that these vertices *correspond* to the vertex resulting from these contractions. (Note that no edge incident to t_1 is contracted.) We process the vertices of T_{canon} in a tidy ancestral order, embedding at each time τ all vertices corresponding to the current vertex.

canonical tree

correspond

For $v \in T$, we write $C^-(v)$ for the children of v in $N_T^-(v)$, $C^+(v)$ for the children of v in $N_T^+(v)$ and $C(v)$ for $C^-(v) \cup C^+(v)$. We write S_x for the star $T\big[\{x\} \cup C(x)\big]$ induced by x and its children.

$C^-(v)$, $C^+(v)$

$C(v)$

▸ **Embedding algorithm.** If at any point in the description below there is more than one possible choice available, we take the lexicographically first of these, so that for each input the output will be uniquely defined — thus making the algorithm deterministic. Furthermore, if at any point some required choice cannot be made, terminate with failure.

At each time τ, with $1 \leq \tau \leq n$, we shall embed a vertex t_τ to a vertex $v_\tau \in V_{\varphi(t_\tau)}$; we will also reserve sets A_τ^-, A_τ^+ for the children of t_τ. We say that a vertex t_s of T is *open* at time τ if t_s has been embedded but some child of t_s has not yet been embedded.

open

▸ **Input.** An oriented tree T with ancestral order $t_1 \prec \ldots \prec t_n$ of T, a homomorphism $\varphi : T \to R$, where R is a directed Hamilton cycle. Also, a digraph G, a partition $\mathcal{V} := \{\, V_i : i \in R \,\}$ of $V(G)$, a vertex $v_1 \in V_{\varphi(t_1)}$ and constants c and γ.

▸ **Procedure.** At each time τ, with $1 \leq \tau \leq n$, we take the following steps.

▸ **Step 1.** 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 children of open vertices, so

$$B^\tau := \{v_1, \ldots, v_{\tau-1}\} \cup \bigcup_{t_s \,:\, t_s \text{ is open}} \left(A_s^- \cup A_s^+\right)$$

For each $V_i \in \mathcal{V}$, let $V_i^\tau := V_i \setminus B^\tau$, so V_i^τ is the set of available vertices of V_i.

- **Step 2.** If $\tau = 1$ embed t_1 to v_1. Alternatively, if $\tau > 1$:

  (2.1) Let t_σ be the parent of t_τ (so A_σ^-, A_σ^+ were reserved for the children of t_σ).

  (2.2) If $t_\sigma \to t_\tau$, let $W := A_\sigma^+ \cap V_{\varphi(t_\tau)}$; otherwise let $W := A_\sigma^- \cap V_{\varphi(t_\tau)}$.

  (2.3) Let $i = \varphi(t_\sigma)$. Choose a set $W_\tau \in W$ such that $|W_\tau| \geq 3(\log n)^{2C}$ and such that for all $v \in W_\tau$

  $$\deg_G^-(v, V_{i-1}^\tau) \geq \gamma m \qquad \text{and} \qquad \deg_G^+(v, V_{i+1}^\tau) \geq \gamma m. \tag{3.1}$$

  (2.4) Embed the star of T corresponding to t_τ into v_τ (using Theorem 3.5).

- **Step 3.** In Step 2 we embedded t_τ to a set W_τ, where $W_\tau \in V_{\varphi(t_\tau)}$. For each $x \in C^-(t_\tau)$, choose a set $A_x^- \subseteq N_G^-(v_\tau) \cap V_{\varphi(x)}^\tau$ containing at most $2m^{1/2}$ vertices and which is (c, γ)-good for S_x; let A_τ^- be the union of these sets. Similarly, for each $y \in C^+(\tau)$, choose a set $A_y^+ \subseteq N_G^+(v_\tau) \cap V_{\varphi(y)}^\tau$ containing at most $2m^{1/2}$ vertices and which is (c, γ)-good for S_y; choose these sets so that they are pairwise disjoint and let A_τ^+ be their union.

- **Termination.** Terminate after every vertex of T 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.

Remarks. Briefly, this algorithm embeds the vertices of a tree in a tidy ancestral order. At any step a vertex is embedded and sets of vertices which are 'good' are reserved for the children of this vertex. To deal with vertices embedded within a cluster we use the fact that the host graph is a tournament and that the sets of reserved vertices are sufficiently large to embed any induced substar of the tree (using the linear bound of Theorem 3.5). We describe a very similar version of this algorithm in Section 4.3.4 (page 94), where a complete proof of an embedding lemma (Lemma 4.16) is given, together with a proof sketch of Lemma 3.4. The main differences there are that R might not be a cycle, and that the host graph no longer is a tournament, so we never embed subtrees within clusters; in particular, the notion of (c, γ)-good sets is replaced by a more general property of being $(\beta, \gamma, \varphi, m)$-good (see Definition 4.3 on page 79). Since the proof of Lemma 3.4 (stating the correctness of the above algorithm) is very similar to the proof of Lemma 4.16, we defer the former until Section 4.3.4 (where we prove Lemma 4.16).

Combining Lemma 3.3 and Lemma 3.4 immediately yields the following corollary, a modified version of [51, Lemma 3.2], in which the original constant bound on $\Delta(T)$ has been replaced by a polylogarithmic bound.

Corollary 3.7. Suppose that $1/n \ll 1/C$ and that $1/n \ll 1/k \ll \varepsilon \ll d \ll \alpha \leq 2$, and let $m := n/k$. Let T be an oriented tree on at most n vertices with $\Delta(T) \leq (\log n)^C$ and with root r. Also let G be a (d, ε)-regular cycle of cluster tournaments on clusters $V_1, \dots, V_k$, each of size $(1+\alpha)m$, and 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. Then G contains a copy of T in which r is embedded to v.

Proof. Apply the Vertex Allocation Algorithm (Algorithm 3.2) to allocate the vertices of T to the clusters $V_1, \dots, V_k$. Then by Lemma 3.3(a) this allocation is semi-canonical, and by Lemma 3.3(d) 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 3.4, this will successfully embed T in G with r embedded to v. □

3.3.3 Proof of Theorem 1.9

In this section we give the complete proof of Theorem 1.9. This theorem states that for all positive α, C, if T is a (sufficiently large) oriented tree of order n such that $\Delta(T) \leq (\log n)^C$ and G is a tournament of order $(1+\alpha)n$ then G contains T.

We said that Theorem 1.9 is a sharpened version of [51, Theorem 1.4(2)]. Indeed, we shall essentially follow the proof of [51, Theorem 1.4(2)], using Corollary 3.7 above in place of [51, Lemma 3.2]. More precisely, we first derive Lemma 3.12— an analogous statement to [51, Lemma 3.1] in which the bound $\Delta(T) \leq \Delta$ for constant Δ is replaced by $\Delta(T) \leq (\log n)^C$. The (short) derivation of this statement is identical to that given in [51], except that Corollary 3.7 is used in place of [51, Lemma 3.2]. We then follow the proof of [51, Theorem 1.4(2)] in Section 6 of [51], with the only changes being that we now use this modified version of [51, Lemma 3.1]. Other results we use (from Sections 2 and 5 of [51]) are applied exactly as they are. (We emphasise that the proofs of these results are quite very similar the ones in [51] to the original forms, but are included here for convenience.)

The notion of 'almost transitive' tournaments plays a key role in the strategy for proving Theorem 1.9. A tournament G on n vertices is *ε-almost-transitive* if the vertices of G can be given an order $v_1, \dots, v_n$ so that at most εn^2 edges are directed against the order of the vertices, that is, they are directed from v_j to v_i ε-almost-transitive where $i < j$. The following lemma states that any oriented tree is contained in any slightly larger almost-transitive tournament.

Lemma 3.8. [51, Lemma 5.1] For all $\alpha > 0$ there exist $\varepsilon_0 > 0$ and $n_0 \in \mathbb{N}$ such that for any $\varepsilon \leq \varepsilon_0$ and any $n \geq n_0$, any ε-almost-transitive tournament G on at least $(1+\alpha)n$ vertices contains any oriented tree T on n vertices.

We now state various other definitions and results which we use in the proof of Theorem 1.9. We make use of the following observation from [51], that if G is a regular cycle of cluster tournaments on clusters $V_1, \dots, V_k$, and we select a subset $U_j \subseteq V_j$ uniformly at random for each $j \in [k]$, then with high probability the restriction of G to these subsets is also regular. This follows from a lemma of Alon et al. [4] 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.

Lemma 3.9. [51, Lemma 2.6] Suppose that $1/m \ll 1/k \ll \varepsilon \ll \varepsilon' \ll d$ and that $m^{1/3} \leq m' \leq m$. Let G be a (d, ε)-regular cycle of cluster tournaments on clusters $V_1, \dots, V_k$, 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 $1 - o(1)$, $G[U_1 \cup \dots \cup U_k]$ is a $(d/2, \varepsilon')$-regular cycle of cluster tournaments.

μ-robust outneighbourhood

(μ,ν)-robust outexpander

Let G be a digraph on n vertices, and let μ be a positive constant. For any subset S of $V(G)$, the *μ-robust outneighbourhood* $\mathrm{RN}^+_\mu(S)$ is the set of vertices of G with at least μn inneighbours in S. For $0 < \mu \leq \nu \leq 1/2$ we say that G is a *(μ, ν)-robust outexpander* when $|\mathrm{RN}^+_\mu(S)| \geq |S| + \mu n$ for each $S \subseteq V(G)$ with $\nu n < |S| < (1-\nu)n$. The following lemma states that any tournament which is not a robust outexpander admits a vertex partition forming an almost-directed pair.

Lemma 3.10. [51, Lemma 2.8] Suppose that $1/n \ll \mu \ll \nu$. Let G be a tournament on n vertices which 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 $e(G[S \leftarrow S']) \leq 4\mu n^2$.

Proof sketch. If G is not a robust outexpander, then there exists a subset $S \subseteq V(G)$ with $\nu n < |S| < (1-\nu)n$ which 'witnesses' this fact, and we can argue that $\{S, V(G) \setminus S\}$ is the desired partition. $\square$

Kühn, Mycroft and Osthus also gave a structural property of tournaments which are robust expanders with high minimum semidegree, namely that such tournaments must contain an almost-spanning cycle of cluster tournaments (essentially, the proof proceeds by applying a version of Szemerédi's regularity lemma for digraphs, then applying a result of Kühn, Osthus and Treglown [57] to find a directed Hamilton cycle in the reduced digraph).

Lemma 3.11. [51, Lemma 2.7] 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 $\delta^0(G) \geq \eta n$. Then G contains a (d,ε)-regular cycle of cluster tournaments on clusters $V_1,\dots,V_k$ of the same size where $|\bigcup_{i=1}^k V_i| > (1-\varepsilon)n$ and $M' \leq k \leq M$.

From Lemma 3.11 and Corollary 3.7 we immediately obtain the following lemma on embedding trees of logarithmic maximum degree in robust outexpander tournaments. This is analogous to [51, Lemma 3.1], but now the maximum degree bound is polylogarithmic rather than constant.

Lemma 3.12. Suppose that $1/n \ll 1/C$ and that $1/n \ll \mu \ll \nu \ll \eta \ll \alpha$. Let G be a tournament on $(1+\alpha)n$ vertices which is a robust (μ,ν)-outexpander with $\delta^0(G) \geq \eta n$ and let T be an oriented tree on n vertices with $\Delta(T) \leq (\log n)^C$. Then G contains a copy of T.

Proof. If $\alpha > 2$ then G contains a copy of T by Theorem 3.5. So we may assume that $\alpha \leq 2$. 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 3.11, G contains a (d,ε)-regular cycle of cluster tournaments G' on clusters $V_1,\dots,V_k$, where $M' \leq k \leq M$, and $|V_1| = \cdots = |V_k| \geq (1-\varepsilon)(1+\alpha)n/k \geq (1+\alpha/2)n/k$. For each i choose $V_i' \subseteq V_i$ of size $|V_i'| = (1+\alpha/2)n/k$ uniformly at random. By Lemma 3.9 we may fix an outcome of these choices so that $G'' = G'[V_1' \cup \cdots \cup V_k']$ is a $(d/2,\varepsilon')$-regular cycle of cluster tournaments. So by Corollary 3.7 G'' contains a copy of T, so G contains T also. □

We can now give the proof of Theorem 1.9, which we first restate. The exposition here follows essentially the one of [51, Theorem 1.4]. The main difference is that we to use Lemmas 3.3 and 3.12 (rather then the corresponding versions in [51]) in order to achieve a polylogarithmic bound on $\Delta(T)$.

Theorem 1.9. For every $\alpha, C > 0$ there exists n_0 such that if T is an oriented tree on $n \geq n_0$ vertices with $\Delta(T) \leq (\log n)^C$ and G is a tournament on at least $(1+\alpha)n$ vertices, then G contains a copy of T.

Proof outline. An important step in the proof is arguing that every tournament G is either almost-transitive or is almost spanned by large vertex-disjoint tournaments $G[S^-], G[S]$ and $G[S^+]$ such that $G[S]$ is a cycle of cluster tournaments and almost all edges (of G) between these subtournaments are directed from S^- to $S \dot\cup S^+$ and from $S^- \dot\cup S$ to S^+. (This is shown using Algorithm 3.13.) We then proceed as follows. If G is almost-transitive, then G contains every almost-spanning subtree T by Lemma 3.8. Otherwise, we partition $V(T)$ into three sets V^-, V, V^+ of $V(T)$ with sizes roughly corresponding to S^-, S, S^+, so that every edge of T between these sets is directed either from V^- to $V \cup V^+$ or from $V^- \cup V$ to V^+. We then embed T to G (with the forests $T[V^-], T[V], T[V^+]$ embedded to $G[S^-], G[S], G[S^+]$, respectively). Each component of these forests is embedded greedily to its assigned partition. Importantly, slightly more vertices of T are placed in V. This means that we have enough 'room to spare' to embed $T[V^-]$ to $G[V^-]$ and $T[V^+]$ to $G[V^+]$; on the other hand, we use Corollary 3.7 to embed components of T to $G[S]$.

Proof. The proof proceeds through three steps: first we partition the vertices of G, then we partition the vertices of T to match the partition of $V(G)$, and finally we embed each part of T in the corresponding part of G.

► **Step 1.** Let A be the set of all positive α' for which the statement of theorem holds. In other words, $\alpha' \in A$ if and only if for all $C > 0$ there exists n_0 such that for any $n \geq n_0$, any tournament on at least $(1 + \alpha')n$ vertices contains any tree T on n vertices with $\Delta(T) \leq (\log n)^C$. Note that, if $\alpha' \in A$ and $\alpha'' > \alpha'$ then $\alpha'' \in A$, and that $2 \in A$ by Theorem 3.5. Let $a_{\mathrm{inf}} \coloneqq \inf A$, so Theorem 1.9 if and only if $a_{\mathrm{inf}} = 0$. We therefore suppose, looking for a contradiction, that $a_{\mathrm{inf}} > 0$. Let $C > 0$ and choose constants

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

If we set $\alpha \coloneqq a_{\mathrm{inf}} - \mu$, then $\alpha < 2$ and we may assume that $\gamma \ll \alpha$. Also $\alpha + 2\mu \in A$, so for all $n' \geq n_0'$, every tournament whose order is at least $(1 + \alpha + 2\mu)n'$ contains every tree T of order n' such that $\Delta(T) \leq (\log n')^{2C}$. Our goal is to prove that, if $n \geq n_0$, then every tournament G whose order is at least $(1 + \alpha)n$ contains every tree T of order n such that $\Delta(T) \leq (\log n)^C$. Since the choice of C is arbitrary, this implies that $\alpha \in A$, which in turn contradicts the assumption that $a_{\mathrm{inf}} > 0$.

Let G be a tournament of order at least $(1 + \alpha)n$. If $|G| \geq 3n$ then G contains every oriented tree T of order n by Theorem 3.5, so we assume that $|G| < 3n$. Our next step is to find a large subtournament $G' \subseteq G$ and a partition of $V(G')$ with some useful structural properties. More precisely, we first obtain an ordered family of subsets of $V(G)$, using Algorithm 3.13 (page 45). At any iteration, this algorithm maintains an ordered family $\mathcal{S}^\tau := (S_1^\tau, \ldots, S_\tau^\tau)$ of pairwise vertex-disjoint subsets of $V(G)$ as well as a set B^τ of 'bad' edges; if the largest element of $\mathcal{S}^\tau$ is S_q^τ, then one of the following two things happens. The algorithm stops if $|S_q^\tau| < \gamma n$ or if $G[S_q^\tau]$ is a robust outexpander with linear minimum semidegree; otherwise, S_q^τ is split in two sets S', S'' such that almost all edges of $G[S_q^\tau]$ are directed from S' to S'', we mark a few more edges as bad and delete vertices incident to too many bad edges. The key observations are contained in the next claim.

Algorithm 3.13: Vertex Partition Algorithm

Input : a tournament G on n vertices; constants μ, ν, η and γ, with $\frac{1}{n} \ll \mu \ll \nu \ll \eta \ll \gamma \ll 1$.

Let $\mathcal{S}^1 = (\,V(G)\,)$, and $B^1 = \varnothing$.
for $\tau = 1, 2, \ldots$ **do**
    Let S_q^τ be a largest member of $\mathcal{S}^\tau := (S_1^\tau, \ldots, S_\tau^\tau)$.
    if $|S_q^\tau| < \gamma n$ **then** terminate.
    if *$G[S_q^\tau]$ is a robust (μ,ν)-outexpander and $\delta^0(G[S_q^\tau]) > \eta n$* **then** terminate.
    if *there exists $v \in S_q^\tau$ with* $\deg^-_{S_q^\tau}(v) < \eta n$ **then**
        Let $S' = \{v\}$ and $S'' = S_q^\tau \setminus \{v\}$.
    else if *there exists $v \in S_q^\tau$ with* $\deg^+_{S_q^\tau}(v) < \eta n$ **then**
        Let $S' = S_q^\tau \setminus \{v\}$ and $S'' = \{v\}$.
    else if *$G[S_q^\tau]$ is not a robust (μ,ν)-outexpander* **then**
        Apply Lemma 3.10 to partition S_q^τ into sets S' and S'' such that $|S'|, |S''| > \nu |S_q^\tau|$ and $\left|E(S'' \leftarrow S')\right| \le 4\mu |S_q^\tau|^2$.
    Let $\mathcal{S}^{\tau+1} = \left(S_1^\tau, \ldots, S_{q-1}^\tau, S', S'', S_{q+1}^\tau, \ldots, S_\tau^\tau\right)$ and $B^{\tau+1} = B^\tau \cup E\left(G[S' \leftarrow S'']\right)$.
    for $i \in [\tau+1]$ **do**
        Delete all vertices in $S_i^{\tau+1}$ incident to more than $\sqrt{\eta} n$ edges in $B^{\tau+1}$.

Claim 3.14. Algorithm 3.13 terminates at a time τ_{end} with $\tau_{\text{end}} \leq |G| \leq 3n$. Moreover, if $G' := G[S_1^{\tau_{\text{end}}} \dot\cup \cdots \dot\cup S_{\tau_{\text{end}}}^{\tau_{\text{end}}}]$, then $|G'| \geq |G| - 8\sqrt{\eta}n$ and either

(i) G' is 2γ-almost transitive, or

(ii) G' admits a partition $\{S^-, S, S^+\}$ such that

  (a) S is a robust (μ, ν)-outexpander, $|S| \geq \gamma n$ and $\delta^0\Big(G[S]\Big) > \eta n$;

  (b) $deg_G^-(x, S \dot\cup S^+) \leq \sqrt{\eta}n$ for all $x \in S^-$; and

  (c) $deg_G^+(z, S^- \dot\cup S) \leq \sqrt{\eta}n$ for all $z \in S^+$.

Proof. At any step τ the algorithm either terminates at steps 4 or 5 or the condition of one of steps 6, 8 or 10 must hold. As a consequence, if the algorithm does not terminate at time τ, then $|\mathcal{S}^{\tau+1}| = |\mathcal{S}^\tau| + 1$. Since each member of $\mathcal{S}^{\tau+1}$ contains at least one vertex of G, the algorithm terminates at some time τ_{end}, with $\tau_{\text{end}} \leq |G| \leq 3n$.

If at time τ the algorithm does not terminate, then it splits S_ℓ^τ in either step 7, 9 or 11. We first argue that step 11 occurs at most $3/\gamma\nu$ times. Indeed, any set obtained by such split has size at least $\gamma\nu n$ (since $|S_\ell^\tau| \geq \gamma n$ and $|S'|, |S''| \geq \nu|S_\ell^\tau|$), so we can have at most $|G|/\gamma\nu n \leq 3/\gamma\nu$ such sets.

We next show that G' contains almost every vertex of G (i.e., not many vertices have been deleted). Note that any vertex which is deleted must lie in $S' \dot\cup S''$. For each $\tau < \tau_{\text{end}}$, if we split S_q^τ in either step 7 or 9 then $e(S' \leftarrow S'') \leq \eta n$, and if we split S_q^τ in step 11 then $e(S' \leftarrow S'') \leq 4\mu|G|^2 \leq 36\mu n^2$ by Lemma 3.10. Since $\tau_{\text{end}} \leq 3n$ and step 11 happens at most $3/\gamma\nu$ times, we have $|B^{\tau_{\text{end}}}| \leq 3\eta n^2 + 108\mu n^2/\nu\gamma \leq 4\eta n^2$. Since $B^1 \subseteq \cdots \subseteq B^{\tau_{\text{end}}}$ and every deleted vertex lies in at least $\sqrt{\eta}n$ edges of $B^{\tau_{\text{end}}}$, we have that

$$|G'| \geq |G| - 8\sqrt{\eta}n. \tag{3.2}$$

We now prove (i) and (ii). Suppose first that the algorithm terminated in step 4. Then $|S_i^{\tau_{\text{end}}}| < \gamma n$ for all $i \in [\tau_{\text{end}}]$ and we will show that G' is 2γ-almost-transitive. Indeed let $v_1, v_2, \ldots, v_{|G'|}$ be the vertices of G', ordered so that all vertices in $S_1^{\tau_{\text{end}}}$ appear first, followed by all vertices in $S_2^{\tau_{\text{end}}}$ and so on. Then any edge $v_i \leftarrow v_j$ where $i < j$ either lies in $B^{\tau_{\text{end}}}$ or has both endvertices in the same set of $\mathcal{S}^{\tau_{\text{end}}}$. In other words the number of 'backward' edges in this order is at most

$$4\eta n^2 + \sum_{S \in \mathcal{S}^{\tau_{\text{end}}}} \binom{|S|}{2} \leq 4\eta n^2 + \sum_{S \in \mathcal{S}^{\tau_{\text{end}}}} \frac{\gamma n|S|}{2} \leq 4\eta n^2 + \frac{3\gamma n^2}{2} \leq 2\gamma n^2.$$

Recall that $|G'| \geq (1 + \alpha/2)n$ by (3.2), so G' is 2γ-almost-transitive and (i) holds.

If the algorithm terminates in step 5, then for some $i \in [\tau_{\text{end}}]$ we have that $G[S_i^{\tau_{\text{end}}}]$ is a (μ, ν)-robust outexpander such that $|S_i^{\tau_{\text{end}}}| \geq \gamma n$ and $\delta^0(G[S_i^{\tau_{\text{end}}}]) \geq \eta n$. Let $S^- := \bigcup_{1 \leq j < i} S_j^{\tau_{\text{end}}}$, let $S := S_i^{\tau_{\text{end}}}$ and let $S^+ := \bigcup_{i < j \leq \tau_{\text{end}}} S_j^{\tau_{\text{end}}}$, so these sets are disjoint and $G' = G[S^- \dot\cup S \dot\cup S^+]$. Moreover, all edges from $S \dot\cup S^+$ to S^- lie in $B^{\tau_{\text{end}}}$, and so do all edges from S^+ to $S^- \dot\cup S$. To complete the proof, consider the ordered family (S^-, S, S^+). Note that by definition any vertex $v \in G'$ is incident to at most $\sqrt{\eta} n$ edges directed 'against' this order (i.e., directed from S^+ to $S^- \dot\cup S$ or from $S \dot\cup S^+$ to S^-), since otherwise the algorithm would have removed v at some time $\tau < \tau_{\text{end}}$. Thus $\{S^-, S, S^+\}$ satisfies (ii). $\diamond$

Returning to the proof of Theorem 1.9, our approach now depends on whether Claim 3.14 (i) or (ii) holds. In the first case, G' (and hence G) contains a copy of T by Lemma 3.8, so there is nothing else to do. Otherwise, we let S^-, S, S^+ and G' be as in (ii) and proceed to Step 2.

Step 2. In this step we will build a partition $\{V^-, V, V^+\}$ of $V(T)$ with part sizes corresponding approximately to the sizes in the partition $\{S^-, S, S^+\}$ of G', except that V will be slightly larger. Let β, β^+ and β^- be such that $|S| = \beta|G'|$, $|S^+| = \beta^+|G'|$ and $|S^-| = \beta^-|G'|$, so $\beta + \beta^+ + \beta^- = 1$ and $\beta \geq \gamma n/|G'| \geq \gamma/3$.

Suppose first that β^+ and β^- are both small. More precisely, $\beta^+, \beta^- \leq \alpha\beta^2/20$, so $\beta \geq 1 - \alpha/10$. In this case we embed the whole of T to $G[S]$. Recall that T has order n and that $G[S]$ is a (μ, ν)-robust outexpander with $\delta^0(G[S]) \geq \eta n$. Also,

$$|S| = \beta|G'| \overset{(3.2)}{\geq} (1 + \alpha - 8\sqrt{\eta})(1 - \alpha/10)n \geq (1 + \alpha/4)n,$$

so $G[S]$ contains T by Lemma 3.12.

We finally analyse what happens if either β^+ or β^- is greater than $\alpha\beta^2/20$. In this case $\beta \leq 1 - \alpha\beta^2/20$, and we partition $V(T)$ according to the values of β^+ and β^-.

β^- is large and β^+ is small. More precisely, $\beta^- > \alpha\beta^2/20$ and $\beta^+ \leq \alpha\beta^2/20$. Fix a partition $\{V^-, V, V^+\}$ of $V(T)$ such that V^+ is empty, $|V^-| = \beta^-(1 - \alpha\beta)n$ and all edges of T with one endpoint in V^- and another in V are directed from V^- to V. (We can form $T[V]$ greedily by successively removing a vertex of out-degree 0 from T, adding it to $T[V]$.) Since $\beta^+ + \beta + \beta^- = 1$,

$$|T[V]| = n - |T[V^-]| = \beta n(1 + \alpha - \alpha\beta) + (1 - \alpha\beta)\beta^+ n \leq \beta n(1 + \alpha - \alpha\beta) + \alpha\beta^2 n/20.$$

β^- is small and β^+ is large. More precisely, $\beta^- \leq \alpha\beta^2/20$ and $\beta^+ > \alpha\beta^2/20$. Fix a partition $\{V^-, V, V^+\}$ of $V(T)$ such that V^- is empty, $|V^+| = \beta^+(1 - \alpha\beta)n$ and all edges of T with one endpoint in V and another in V^+ are directed from V to V^+. Again $|T[V]| = n - |T[V^+]| \leq \beta n(1 + \alpha - \alpha\beta) + \alpha\beta^2 n/20$.

β^+ **and** β^- **are both large.** More precisely, $\beta^+, \beta^- > \alpha\beta^2/20$. Fix a partition $\{V^-, V, V^+\}$ of $V(T)$ such that $|V^-| = \beta^-(1 - \alpha\beta)n$, $|V^+| = \beta^+(1 - \alpha\beta)n$ and that all edges of T with endpoints in distinct sets are directed as follows: edges between V^- and any other set are directed away from V^-, and edges between any set and V^+ are directed towards V^+. Note that $|T[V]| = \beta(1 + \alpha - \alpha\beta)n$.

In all three cases above we have

$$\beta(1+\alpha-\alpha\beta)n \leq |V| \leq \beta(1+\alpha-\alpha\beta)n + \alpha\beta^2 n/20 \leq \beta(1+\alpha)n - \frac{\alpha\beta^2 n}{2}. \quad (3.3)$$

► **Step 3.** In the final step we embed the forests $T[V^-], T[V], T[V^+]$ to the tournaments $G[S^-], G[S], G[S^+]$, respectively, so that they form a copy of T in G. This will complete the proof, since a copy of T in G' is also a copy of T in G. Recall that in the previous step we chose V to be slightly larger than a proportional partition of T would yield. While this gives us enough room to spare (by our choice of α) when embedding $T[V^-]$ to $G[S^-]$ and $T[V^+]$ to $G[S^+]$, this is not the case for $T[V]$ and $G[S]$. However, crucially, $G[S]$ is a (somewhat large) robust (μ, ν)-outexpander of linear semidegree.

Let $T_1^-, \ldots, T_x^-$ be the component subtrees of $T[V^-]$, let $T_1^+, \ldots, T_y^+$ be the component subtrees of $T[V^+]$, and let $T_1, \ldots, T_z$ be the component subtrees of $T[V]$. We contract each of these component subtrees to a single vertex, and call the resulting tree the *contracted tree* T_{con} of T. Let W be the set vertices of T_{con} which correspond to components T_i with order at least n/Δ (so $|W| \leq \Delta$) and fix an order of the vertices of T_{con} which begins with the vertices in W and such that each vertex in this order has at most Δ neighbours preceding it. (To do this, fix an ancestral order of T_{con} starting at a vertex in W, then move all vertices of W to the beginning of the order.) We shall proceed greedily, as follows: we first embed the forest corresponding to the vertices in W, and then go on to embed each component corresponding to the remaining vertices of T_{con} in the order described above.

contracted tree

Note that the vertices of W correspond to a subforest F of $T[V]$ with order at most $|V|$ and such that $\Delta(F) \leq (\log n)^C \leq \left(\log|V|\right)^{2C}$. Since $G[S]$ is a robust (μ, ν)-outexpander of order

$$\beta|G'| \overset{(3.2)}{\geq} \beta(1 + \alpha - 8\sqrt{\eta})n \overset{(3.3)}{\geq} \left(1 + \frac{\alpha\beta}{10}\right)|V| \geq (1 + \gamma^2)|V|$$

and such that $\delta^0\Big(G[S]\Big) \geq \eta n$, we have that $G[S]$ contains a copy of this forest by Lemma 3.12 (with $2C$ here playing there the role of C).

Let $t^\star$ be the vertex of T_{con} corresponding to the next component we will embed. If $t^\star \in V$, then $t^\star$ corresponds to some T_i, and since T_i has not been embedded we have $|T_i| \leq n/\Delta$. Moreover, $t^\star$ has at most Δ neighbours preceding it in T_{con}. Let $t_1^-, \ldots, t_p^-$ be the inneighbours of $t^\star$ which have already been embedded, and let $t_1^+, \ldots, t_q^+$ be the outneighbours of $t^\star$ which have already been embedded. Then each t_i^- is a vertex of V^- and hence has been embedded to a vertex $v_i^- \in S^-$, and each t_j^+ is a vertex of V^+ which has been embedded to a vertex $v_j^+ \in S^+$. Let $S^\star$ be the set of unoccupied vertices in $S \cap \bigcup_{i\in[p]} N^+(v_i^-) \cup \bigcup_{j\in[q]} N^-(v_j^+)$. We want to embed T_i to $S^\star$. Note that

$$\begin{aligned}
|S^\star| &\geq |S| - (p+q)\sqrt{\eta}n - |V| \overset{(3.3)}{\geq} \beta|G'| - \Delta\sqrt{\eta}n - (\beta(1+\alpha)n - \alpha\beta^2 n/2) \\
&\overset{(3.2)}{\geq} \beta n(1+\alpha) - (8+\Delta)\sqrt{\eta}n - \beta(1+\alpha)n + \alpha\beta^2 n/2 \\
&\geq \alpha\beta^2 n/3 \geq 3n/\Delta \geq 3|T_i|,
\end{aligned}$$

and we can embed T_i to $G[S^\star]$ by Theorem 3.5.

Otherwise, if $t^\star$ corresponds to some T_i^-, then as above the vertices of T_i^- have at most Δ already-embedded neighbours, and all of these are outneighbours of vertices of T_i which have been embedded to $S \,\dot\cup\, S^+$. Let $v_1, \ldots, v_r$ be the vertices of $S \,\dot\cup\, S^+$ to which these vertices were embedded, and let $S^\star$ be the set of unoccupied vertices of $S^- \cap \bigcup_{i\in[r]} N^-(v_i)$. Since at most $|V^-| - |T_i^-|$ vertices of V^- have been embedded to S^-, we have

$$\begin{aligned}
|S^*| &\geq |S^-| - r\sqrt{\eta}n - (|T[V^-]| - |T_i^-|) \\
&\overset{(3.2)}{\geq} \beta^-(1+\alpha)n - (8+\Delta)\sqrt{\eta}n - \beta^-(1-\alpha\beta)n + |T_i^-| \\
&\geq \beta^-(\alpha + \alpha\beta/2)n + |T_i^-|. \qquad (3.4)
\end{aligned}$$

Where we that $\beta^- \geq \alpha\beta^2/20$ and $\beta \geq \gamma/3$, and hence we have $\eta, 1/\Delta \ll \gamma, \beta, \beta^-$. We conclude that $|S^*| \geq (1+\alpha+2\mu)|T_i^-|$ and therefore, if $|T_i^-| \geq \beta^-\alpha n/2$, then $|T_i^-| \geq n_0'$, and we can embed T_i^- to $G[S^*]$ by our choice of n_0' (since $\alpha + 2\mu \in A$ and $\Delta(T_i^-) \leq (\log n)^C \leq (\log|T_i^-|)^{2C}$). On the other hand, if $|T_i^-| < \beta^-\alpha n/2$ then $|S^*| \geq 3|T_i^-|$ by (3.4), and so we can embed T_i^- to $G[S^*]$ by Theorem 3.5.

The procedure in the case where $t^\star$ corresponds to some T_i^+ is symmetric and we omit the calculation. We continue in this way, embedding all components of $T[V^-], T[V]$ and $T[V^+]$. This completes an embedding of T to G (and the proof). $\square$

In the proof of Theorem 1.4 (on Section 3.6) we use the following corollary. This is a consequence of Theorem 1.9 and El Sahili's Theorem 3.5. Indeed, this corollary is simple to apply since it holds for both small and large trees.

Corollary 3.15. Suppose that $1/n \ll \alpha, 1/C$. Let T be an oriented tree on $n' \leq n$ vertices with $\Delta(T) \leq (\log n)^C$, and let G be a tournament on at least $n' + \alpha n$ vertices. Then G contains a copy of T.

Proof. Fix $\alpha, C > 0$ and choose n_0 sufficiently large to apply Theorem 1.9 with $2C$ in place of C, and also so that $\log n_0 \geq \Big(1 + \log(2/\alpha)\Big)^2$. Then we may assume that $n \geq 2n_0/\alpha$. If $n' > \alpha n/2$, then $n' > n_0$, so G contains a copy of T by Theorem 1.9, since G has at least $n' + \alpha n \geq (1+\alpha)n'$ vertices and $\Delta(T) \leq (\log n)^C \leq (\log n')^{2C}$. On the other hand, if $n' \leq \alpha n/2$, then $|G| \geq n' + \alpha n \geq 3n'$, and thus G contains a copy of T by the aforementioned theorem of El Sahili. □

3.4 Almost-directed pairs

Our aim in this section is to prove Lemma 3.18 (needed for the proof of Theorem 1.4), which states that every nice oriented tree T of polylogarithmic maximum degree is contained in every tournament whose vertex set admits a partition $\{U, W\}$ into not-too-small sets U and W such that the pair (U, W) is almost-directed.

reverse edge

We begin with a definition and two lemmas. If (X, Y) is a μ-almost-directed pair in a digraph G, we say that an edge $e \in E(G)$ is a *reverse edge* if $e \in E\Big(G[X \leftarrow Y]\Big)$ (so, by definition, an almost-directed pair has at most $\mu|X||Y|$ reverse edges). Our first lemma guarantees that we may partition the vertex set of an oriented tree T into sets A and B so that (A, B) is a directed pair in T and so that specific in-subtrees of T have all their vertices in A and specific out-subtrees of T have all their vertices in B. Moreover, we may specify the sizes of A and B (subject to the trivial necessary conditions).

Lemma 3.16. Let T be an oriented tree on n vertices. Let $\mathcal{T}^-$ be a collection of in-subtrees of T, and let $\mathcal{T}^+$ be a collection of out-subtrees of T, such that the trees in $\mathcal{T}^- \cup \mathcal{T}^+$ are pairwise vertex-disjoint. If a and b are integers with

$$a \geq \Big|\bigcup_{S\in\mathcal{T}^-} V(S)\Big|, \quad b \geq \Big|\bigcup_{S\in\mathcal{T}^+} V(S)\Big| \quad \text{and} \quad a+b=n,$$

then there exists a partition $\{A, B\}$ of $V(T)$ with $|A| = a$ and $|B| = b$ such that (A, B) is a directed pair in T and

$$\bigcup_{S\in\mathcal{T}^-} V(S) \subseteq A \quad \text{and} \quad \bigcup_{S\in\mathcal{T}^+} V(S) \subseteq B.$$

Proof. The key observation is that in every oriented forest there is a vertex with no inneighbours (since a forest has more vertices than edges). Define $V^- := \bigcup_{S\in\mathcal{T}^-} V(S)$ and $V^+ := \bigcup_{S\in\mathcal{T}^+} V(S)$, and let $k := a - |V^-|$, so $0 \leq k \leq n - |V^-| - |V^+|$. Greedily choose distinct vertices $v_1, v_2, \ldots, v_k$ of $V(T) \setminus (V^- \cup V^+)$ such that v_i has no inneighbours in $T - \big(V^- \cup V^+ \cup \{v_1, \ldots, v_{i-1}\}\big)$ for each $i \in [k]$. The desired partition is then $A := V^- \cup \{v_1, \ldots, v_k\}$ and $B := V(T) \setminus A$. Indeed, we have $V^- \subseteq A$, $V^+ \subseteq B$, $|A| = |V^-| + k = a$ and $|B| = n - |A| = b$. It remains to show that (A, B) is a directed pair in T. So suppose that $u \to v$ is an edge of T and $v \in A$. It then suffices to show that we must have $u \in A$ as well. For this, observe that since $V^+ \subseteq B$ consists of outstars of T, and $v \in A$, we cannot have $u \in V^+$. So if $v \notin V^-$, then $v = v_i$ for some $i \in [k]$, and by choice of v_i we then have $u \in V^- \cup \{v_1, \ldots, v_{i-1}\} \subseteq A$. On the other hand, if $v \in V^-$ then v is a vertex of some in-subtree of T, so u must be a vertex of the same in-subtree; it follows that $u \in A$. □

Suppose now that T is an oriented tree of polylogarithmic maximum degree whose vertex set is partitioned into sets A and B which form a directed pair (A, B) in T, and also that G is a tournament whose vertex set admits a partition into sets U and W such that (U, W) is an almost-directed pair in G. The next lemma shows that if U and W are slightly larger than A and B respectively, then under the additional assumption that every vertex of G lies in few reverse edges, we may embed T in G so that vertices of A are embedded in U and vertices of B are embedded in W. (Recalling the proof outline of Theorem 1.4, we will use this lemma to embed the subtree T' in G.)

Lemma 3.17. Suppose that $1/n \ll 1/C$ and that $1/n \ll \mu \ll \alpha$. Let T be an oriented tree with $\Delta(T) \leq (\log n)^C$ and let $\{A, B\}$ be a partition of $V(T)$ such that (A, B) is a directed pair in T. Also let G be a tournament on n vertices. If $V(G)$ admits a partition $\{U, W\}$ such that

(i) $|U| \geq |A| + \alpha n$,

(ii) $|W| \geq |B| + \alpha n$,

(iii) for each $u \in U$ we have $\deg^-(u, W) \leq \mu n$, and

(iv) for each $w \in W$ we have $\deg^+(w, U) \leq \mu n$,

then there exists a copy of T in G such that every vertex in A is embedded in U and every vertex in B is embedded in W.

Proof. Consider the oriented forest F with $F \coloneqq T[A] \cup T[B]$ (so $V(F) = V(T)$ and the edges of F are the edges of T with both endvertices in A or both endvertices in B). Let $C_1, \dots, C_s$ be the components of F, and let T' be the tree we obtain by contracting $V(C_j)$ to a single vertex v_j, for each $j \in [s]$. We may assume the components are labelled so that $v_1, \dots, v_s$ is an ancestral order of $V(T')$. We will greedily embed $C_1, \dots, C_s$ in G in that order, defining a mapping $\varphi \colon V(T) \to U \cup W$. For each $j \in [s]$, let U_j (respectively W_j) be the set of vertices of U (respectively W) which have not been covered by the embedding of $C_1, \dots, C_{j-1}$.

If $V(C_1) \subseteq A$, then by (i) we have $|U_1| = |U| \geq |A| + \alpha n \geq |C_1| + \alpha n$, so there exists a copy of C_1 in $G[U_1]$ by Corollary 3.15. By a similar argument using (ii) we may embed C_1 in $G[W_1]$ if $V(C_1) \subseteq B$. Now suppose that we have already embedded components $C_1, \dots, C_{j-1}$ for some $1 < j \leq n$, so $\varphi(v)$ is defined for every $v \in \bigcup_{i=1}^{j-1} V(C_i)$. Since we assumed that $v_1, \dots, v_s$ was an ancestral order of $V(T')$, there exists a unique integer $i \in [j-1]$ for which some vertex $u \in V(C_i)$ is adjacent to some vertex $v \in C_j$. Suppose first that $u \to v \in E(T)$. Then C_i has been embedded in U and C_j is a component of $T[B]$, and we want to embed C_j in $W_j \cap N^+\big(\varphi(u)\big)$. Note that $\varphi(u)$ has at most μn inneighbours in W by (iii), so by (ii) the number of outneighbours of $\varphi(u)$ in W which are not in the image of φ (that is, which are not covered by the embedding so far) is at least $|W_j| - \mu n \geq \big(|W| - |B| + |C_j|\big) - \mu n \geq |C_j| + \alpha n - \mu n \geq |C_j| + \alpha n/2$. We may therefore embed C_j in $G[W_j]$ by Corollary 3.15. If instead $u \leftarrow v \in E(T)$ then C_j is a component of $T[A]$ and we may embed C_j in $G[U_j]$ by a similar argument using (i) and (iv). In either case we have extended φ as desired, and so proceeding in this manner gives a copy of T in G. $\square$

We are now ready to state and prove Lemma 3.18, the main result of this section, following the approach sketched in the proof outline of Theorem 1.4 (Section 3.1).

Lemma 3.18. Suppose that $1/n \ll 1/C$ and that $1/n \ll \mu \ll \alpha, \nu$. Let G be a tournament on n vertices, and let T be an α-nice oriented tree on n vertices with $\Delta(T) \leq (\log n)^C$. If there is a partition $\{U, W\}$ of $V(G)$ with $|U|, |W| \geq \nu n$ such that (U, W) is a μ-almost-directed pair in G, then G contains a (spanning) copy of T.

Proof. Introduce new constants ψ and β so that $1/n \ll \mu \ll \psi \ll \beta \ll \alpha, \nu$. Since (U, W) is a μ-almost directed pair in G, there are at most $\mu|U||V|$ reverse edges, so at most $\sqrt{\mu}|U|$ vertices of U are incident to at least $\sqrt{\mu}|W|$ reverse edges, and at most $\sqrt{\mu}|W|$ vertices of W are incident to at least $\sqrt{\mu}|U|$ reverse edges. Let Z be the set of all such vertices, so $z := |Z| \leq \sqrt{\mu}\big(|U| + |W|\big) = \sqrt{\mu}n$. Now let $W_0 := W \setminus Z$, and let X be the set of all vertices $w \in W_0$ with $\deg^0(w, W_0) < \psi n$. Then by Lemma 2.10 we have $|X| < 4\psi n$. Choose a subset $Y \subseteq W_0$ of size ψn uniformly at random. Note that for each $w \in W_0 \setminus X$ the values of $\deg^-(w, Y)$ and of $\deg^+(w, Y)$ then have a hypergeometric distribution with expectation at least $\psi n|Y|/|W_0| \geq \psi^2 n$, so $\mathbb{P}(\deg^0(w, Y) < \psi^2 n/2)$ decreases exponentially with n by Theorem 2.14. Taking a union bound over the at most n vertices $w \in W_0 \setminus X$ we find that with positive probability every $w \in W_0 \setminus X$ has $\deg^0(w, Y) \geq \psi^2 n/2 \geq 2z$. Fix a choice of Y for which this event occurs and define $U' := U \setminus Z$ and $W' := W_0 \setminus (Y \cup X)$. Also let $n' := |U' \cup W'|$, so $n' \geq n - |X| - |Y| - |Z| \geq (1 - 6\psi)n$. Observe that we then have the following properties.

(a) Every vertex $u \in U \setminus Z$ has $\deg^-(u, W') \leq \sqrt{\mu}|W| \leq \psi n'$.

(b) Every vertex $w \in W \setminus Z$ has $\deg^+(w, U') \leq \sqrt{\mu}|U| \leq \psi n'$.

(c) Every vertex $w \in W'$ has $\deg^0(w, Y) \geq 2z$.

(d) $|U'| \geq |U| - |Z| \geq |U| - \sqrt{\mu}n$ and $|W'| \geq |W| - |X| - |Y| - |Z| \geq |W| - 6\psi n$.

(e) $\Delta(T) \leq (\log n)^C \leq (\log n')^{2C}$.

Define $t := \lceil \beta n \rceil$. Let $\mathcal{S}^-$ be the set of pendant instars of T which contain an out-leaf of T, and let $\mathcal{S}^+$ be the set of pendant outstars of T which contain both an in-leaf of T and an out-leaf of T. Observe that $\mathcal{S}^- \cup \mathcal{S}^+$ is then a set of vertex-disjoint subtrees of T. Moreover, since T is α-nice, we have $|\mathcal{S}^-|, |\mathcal{S}^+| \geq \alpha n$. We define $S_1^-, \dots, S_t^-$ to be the smallest t members of $\mathcal{S}^-$ and $S_1, \dots, S_{t+z}^+$

to be the smallest $t+z$ members of $\mathcal{S}^+$. Since $t+z \le 2\beta n$ we must then have $\left|\bigcup_{i\in[t]} V(S_i^-)\right|, \left|\bigcup_{i\in[t+z]} V(S_i^+)\right| \le 2\beta n/\alpha$. For each $i \in [t]$ let ℓ_i^+ be an out-leaf of T in S_i^- and let c_i^- be the centre of the star S_i^-, and for each $i \in [z]$ let ℓ_{t+i}^+ be an out-leaf of T in S_{t+i}^+. Similarly, for each $i \in [t+z]$ let ℓ_i^- be an in-leaf of T in S_i^+ and let c_i^+ be the centre of the star S_i^+. We can be sure that these leaves exist by definition of $\mathcal{S}^+$ and $\mathcal{S}^-$.

We now define T' to be the subtree of T obtained by deleting the leaves ℓ_i^+ and ℓ_i^- from T for each $i \in [t+z]$. So $L_i^- := S_i^- - \ell_i^+$ (respectively $L_i^+ := S_i^+ - \ell_i^-$) is an in-subtree (respectively out-subtree) of T' for each $i \in [t]$, and $L_{t+j}^+ := S_{t+j}^+ - \{\ell_{t+j}^-, \ell_{t+j}^+\}$ is an out-subtree of T' for each $j \in [z]$. Also define $a := |U| - t$ and $b := |W| - t - 2z$. Then we have $a \ge \nu n - t \ge 2\beta n/\alpha \ge \left|\bigcup_{i\in[t]} V(L_i^-)\right|$ and $b \ge \nu n - t - 2z \ge 2\beta n/\alpha \ge \left|\bigcup_{i\in[t+z]} V(L_i^+)\right|$, and also $a+b = |U|+|W|-2t-2z = |T'|$, so we may apply Lemma 3.16 to obtain a partition $\{A, B\}$ of $V(T')$ with $|A| = a$ and $|B| = b$ such that (A, B) is a directed pair in T' and so that $V(L_i^-) \subseteq A$ for each $i \in [t]$ and $V(L_i^+) \subseteq B$ for each $i \in [t+z]$. Next, since by (d) we have $|U'| \ge a + \beta n'/2$ and $|W'| \ge b + \beta n'/2$, by (a), (b) and (e) we may apply Lemma 3.17 (with $n', \psi, 2C$ and $\beta/2$ in place of n, μ, C and α respectively) to obtain an embedding φ of T' in G so that $\varphi(A) \subseteq U'$ and $\varphi(B) \subseteq W'$.

We next embed the vertices ℓ_{t+j}^+ and ℓ_{t+j}^- for $j \in [z]$ so that all vertices of Z are covered. Note that our embedding of T' in G ensured that for each $j \in [z]$ the centre c_{t+j}^+ of S_{t+j}^+ was embedded to a vertex $w_{t+j} := \varphi(c_{t+j}^+)$ in W', so in particular we have $\deg^0(w_{t+j}, Y) \ge 2z$ by (c). This means that we can greedily choose distinct vertices $y_1^-, y_1^+, \dots, y_z^-, y_z^+ \in Y$ so that for each $j \in [z]$ the vertex y_j^- is an inneighbour of w_{t+j} and y_j^+ is an outneighbour of w_{t+j}. Write $Z := \{q_1, \dots, q_z\}$, and for each $j \in [z]$ consider the orientation of the edge of G between q_j and w_{t+j}. If $q_j \to w_{t+j} \in E(G)$, then we set $\varphi(\ell_{t+j}^-) := q_j$ and $\varphi(\ell_{t+j}^+) := y_j^+$. Similarly, if $q_j \leftarrow w_{t+j} \in E(G)$, then we set $\varphi(\ell_{t+j}^-) := y_j^-$ and $\varphi(\ell_{t+j}^+) := q_j$.

Observe that we have now embedded all of the vertices of T except for the leaves $\ell_1^+, \dots, \ell_t^+$ and $\ell_1^-, \dots, \ell_t^-$. Let $P^- := \{\, \varphi(c_i^-) : i \in [t] \,\}$ and $P^+ := \{\, \varphi(c_i^+) : i \in [t] \,\}$, so $P^- \subseteq U'$ and $P^+ \subseteq W'$. Also, let Q^- be the set of uncovered vertices of U and let Q^+ be the set of uncovered vertices of W. Then $|Q^-| = |U| - a = t$, and $|Q^+| = |W| - b - 2z = t$, so we have $|P^-| = |P^+| = |Q^-| = |Q^+| = t$. Observe that since we already covered all vertices of Z, we also have $Q^- \subseteq U \setminus Z$ and $Q^+ \subseteq W \setminus Z$. Together with the fact that $t = \lceil \beta n \rceil$, by (a) and (b) it follows that $G[P^- \to Q^+]$ and $G[Q^- \to P^+]$ are both $(1, \frac{1}{2})$-super-regular, so the balanced bipartite underlying graph of each contains a perfect matching by Lemma 2.8. For each $j \in [t]$ let $\varphi(\ell_j^+) \in Q^+$ (respectively $\varphi(\ell_j^-) \in Q^-$) be the vertex matched to $\varphi(c_j^-) \in P^-$ (respectively $\varphi(c_j^+) \in P^+$); this completes the embedding φ of T

in G. $\square$

3.5 Cycles of cluster tournaments

Our goal in this section is to prove Lemma 3.24, which states that every sufficiently large tournament containing an almost-spanning regular cycle of cluster tournaments contains a spanning copy of every nice oriented tree T with polylogarithmic maximum degree. Recall from the proof sketch of Theorem 1.4 that for this we split T into two subtrees T_1 and T_2. We then embed T_1 so that all 'atypical' vertices are covered and so that roughly the same number of vertices from each cluster are covered. Since T_1 covered all atypical vertices, the vertices which remain uncovered then form a super-regular cycle of cluster tournaments, and we use this fact to embed T_2 to cover all vertices which remain uncovered and so complete the embedding of T in G. In Section 3.5.1 we focus on the embedding of T_1, showing that we can find an embedding with the desired properties (Lemma 3.19). Likewise, in Section 3.5.2 we consider the embedding of T_2, and prove that we can indeed embed T_2 so as to cover all remaining vertices, as desired (Lemma 3.20). Finally, in Section 3.5.3 we combine these results to prove Lemma 3.24 by first splitting T into subtrees T_1 and T_2 and then successively embedding these subtrees using Lemmas 3.19 and 3.20.

3.5.1 Embedding the first subtree

The subtree T_1 will have polylogarithmic maximum degree and will contain many vertices which are adjacent to at least one in-leaf and at least one out-leaf of T, and we wish to embed T_1 into a tournament G which contains an almost-spanning cycle of cluster tournaments so that approximately the same number of vertices of T_1 are embedded in each cluster. The following lemma states that we can indeed do this.

Lemma 3.19. Suppose that $1/n \ll 1/k \ll \varepsilon \ll d \ll \psi \ll \beta \ll \alpha$ and also that $1/n \ll 1/C$. Let T be an oriented tree of order n, with root r and maximum degree $\Delta(T) \leq (\log n)^C$, which contains at least βn distinct vertices that are each adjacent to at least one in-leaf and one out-leaf of T. Let G be a tournament which contains a (d, ε)-regular cycle of cluster tournaments whose clusters $V_1, \dots, V_k$ have size $(1+\alpha)\frac{n}{k} \leq |V_i| \leq \frac{3n}{k}$ for each $i \in [k]$, and assume additionally that $B \coloneqq V(G) \setminus \bigcup_{i\in[k]} V_i$ has size $|B| \leq \psi n$. Then there exists an embedding φ of T in G covering B, such that r is embedded in V_1 and such that for each $i \in [k]$ we have

$$\left|\varphi\big(V(T)\big) \cap V_i\right| = \big(n - |B|\big)\left(\frac{1}{k} \pm \frac{2}{\log\log n}\right).$$

Loosely speaking the proof proceeds as follows. We begin by selecting from each cluster V_i a large subset V_i' of vertices which each have large semidegree in $V_i \setminus V_i'$. Then $V_1', \dots, V_k'$ are the clusters of a regular cycle of cluster tournaments in $G' \coloneqq G\big[\bigcup_{i\in[k]} V_i'\big]$. We remove a small number of leaves from T to obtain a subtree T', and embed T' in G' by using the Vertex Allocation Algorithm (Algorithm 3.2) and Lemma 3.4. Lemma 3.3 then ensures that approximately the same number of vertices are embedded in each cluster. Finally, we extend the embedding of T' in G to an embedding of T in G by embedding the removed leaves so as to cover all vertices of B.

Proof. Define $m \coloneqq \frac{n}{k}$, so $(1+\alpha)m \leq |V_i| \leq 3m$ for each $i \in [k]$, and let $\delta \coloneqq \frac{1}{\log\log n}$. Let B_i be the set of all vertices $x \in V_i$ such that $\deg^0(x, V_i) < \alpha m/20$. By Lemma 2.10 we have $|B_i| < \alpha m/4$. For each $i \in [k]$, pick a subset $Y_i \subseteq V_i$ of size $|Y_i| = \alpha m/4$ uniformly at random with choices made independently for each i. Note that for each $i \in [k]$ and each $x \in V_i \setminus B_i$, the random variables $\deg^-(x, Y_i)$ and $\deg^+(x, Y_i)$ then have hypergeometric distributions with expected value at least $(\alpha m/20)|Y_i|/|V_i| > 5\beta m$, and thus $\mathbb{P}(\deg^0(x, Y_i) < 4\beta m)$ decreases exponentially with n by Theorem 2.14. Taking a union bound, we find that there is a positive probability that for every $i \in [k]$ and every $x \in V_i \setminus B_i$ we have $\deg^0(x, Y_i) \geq 4\beta m$. Fix a choice of sets $Y_1, \dots, Y_k$ such that this event occurs, and for each $i \in [k]$ let $V_i' \coloneqq V_i \setminus (Y_i \cup B_i)$, so

$$3m \geq |V_i| \geq |V_i'| \geq |V_i| - |B_i| - |Y_i| > (1+\alpha)m - \frac{\alpha m}{4} - \frac{\alpha m}{4} = \left(1 + \frac{\alpha}{2}\right)m$$

and so $\deg^0(x, Y_i) \geq 4\beta m$ for each $x \in V_i'$. Let $G' \coloneqq G[V_1' \cup \dots \cup V_k']$, and observe that since $V_1, \dots, V_k$ were the clusters of a (d, ε)-regular cycle of cluster tournaments in G, by Lemma 2.7 the sets $V_1', \dots, V_k'$ are the clusters of a spanning $(d, 3\varepsilon)$-regular cycle of cluster tournaments in G'. In particular we may choose a vertex $v \in V_1'$ with at least $(d-3\varepsilon)|V_k'|$ inneighbours in V_k' and at least $(d-3\varepsilon)|V_2'|$ outneighbours in V_2'. The tournament G', the clusters $V_1', \dots, V_k'$ and the vertex v then meet the conditions of Lemma 3.4 with $\alpha/2$ and 3ε in place of α and ε respectively (and with n playing the same role there as here).

Let $t := \lceil \beta n \rceil - 1$, and choose a set $W := \{w_1, \dots, w_t\}$ of t distinct vertices in T so that each w_i is adjacent to at least one in-leaf and at least one out-leaf of T and so that r is not a leaf of T which is adjacent to a vertex of W (such a set exists by the assumptions of the lemma). For each $j \in [t]$, let w_j^- and w_j^+ be respectively an in-leaf and an out-leaf adjacent to w_j. Let T' be the oriented tree we obtain by deleting from T the vertices w_j^- and w_j^+ for each $j \in [t]$, so $|T'| = n - 2t$ and $\Delta(T') \leq \Delta(T) \leq (\log n)^C \leq (\log(n-2t))^{2C}$. Also take r to be the root of T', and apply the Vertex Allocation Algorithm (Algorithm 3.2) to allocate the vertices of T' to the clusters $V_1', \dots, V_k'$. By Lemma 3.3(a) the obtained allocation will be semi-canonical. Moreover, by two applications of Lemma 3.3(d) (with $\beta/2$ and $2C$ in place of α and C respectively) we have with probability $1 - o(1)$ that for each $i \in [k]$ the number of vertices of T' allocated to the cluster V_i' is

$$(n-2t)\left(\frac{1}{k} \pm \frac{1}{\log\log(n-2t)}\right) = \frac{n-2t}{k} \pm \frac{3\delta n}{2}, \tag{3.5}$$

and the number of vertices of W allocated to the cluster V_i' is

$$t\left(\frac{1}{k} \pm \frac{1}{\log\log(n-2t)}\right) = \frac{t}{k} \pm \frac{3\delta t}{2}. \tag{3.6}$$

Fix an outcome of the Vertex Allocation Algorithm for which each of these events occurs, and apply Lemma 3.4 to obtain an embedding φ of T' in G' so that r is embedded to v and each vertex of T' is embedded in the cluster V_i' to which it is allocated. In particular r is embedded in V_1, as required.

We now extend φ to an embedding of T in G which covers B. Let $b := |B| \leq \psi n$, and let $q_1, \dots, q_b$ be the vertices of B. Also let $p \in [k]$ be such that $b \equiv p \bmod k$, and for each $i \in [k]$ choose $W_i \subseteq W$ such that $\varphi(W_i) \subseteq \varphi(W) \cap V_i'$ and so that $|W_i| = \lceil b/k \rceil$ if $i \in [p]$ and $|W_i| = \lfloor b/k \rfloor$ if $i \in [k] \setminus [p]$. (Since $b/k \leq \psi n/k$ and $\psi \ll \beta$, we have that (3.6) ensures that we can indeed choose such sets.) The sets $W_1, \dots, W_k$ are then vertex-disjoint and $|\bigcup_{i \in [k]} W_i| = b$, so by relabelling if necessary we may assume that $\bigcup_{i \in [k]} W_i = \{w_1, \dots, w_b\}$. For each $j \in [t]$ set $p_j := \varphi(w_j)$ and write i_j to denote the index such that $p_j \in V_{i_j}$. Greedily choose $2t$ distinct vertices $c_1^-, c_1^+, \dots, c_t^-, c_t^+$ so that for each $j \in [t]$ we have that $c_j^-, c_j^+ \in Y_{i_j}$, that c_j^- is an inneighbour of p_j and that c_j^+ is an outneighbour of p_j. It is possible to make such choices since for each $i \in [k]$ there are at most $2t/k$ vertices w_j with $i_j = i$ by (3.6), and because for each $j \in [t]$ we have $p_j \in V_{i_j}'$

(since w_j is a vertex of T'), so the semidegree of p_j in Y_{i_j} is at least $4\beta m \geq 2 \cdot (2t/k)$ by our choice of the sets Y_i.

Recall that each vertex in W is adjacent to precisely one removed in-leaf w_j^- of T and one removed out-leaf w_j^+ of T, and that these leaves have not yet been embedded. For each $s \in [b]$ we embed one of these leaves to the vertex q_s and the other to either c_s^- or c_s^+ according to the direction of the edge between q_s and p_s. For each $b+1 \leq s \leq t$ we then embed the in-leaf of w_s to c_s^- and the out-leaf of w_s to c_s^+. More precisely, for all integers s with $1 \leq s \leq b$ we set $\varphi(w_s^-) \coloneqq q_s$ and $\varphi(w_s^+) \coloneqq c_s^+$ if $q_s \to p_s \in E(G)$, and set $\varphi(w_s^+) \coloneqq q_s$ and $\varphi(w_s^-) \coloneqq c_s^-$ if $q_s \leftarrow p_s \in E(G)$. Then, for all integers s with $b < s \leq t$ we set $\varphi(w_s^-) \coloneqq c_s^-$ and $\varphi(w_s^+) \coloneqq c_s^+$. Following this extension φ is an embedding of T in G which covers every vertex in B. Moreover, for each $i \in [k]$ the number of vertices embedded in the cluster V_i is

$$\begin{aligned}\left|\varphi\Big(V(T)\Big) \cap V_i\right| &= \left(\frac{n-2t}{k} \pm \frac{3\delta n}{2}\right) + 2\left(\frac{t}{k} \pm \frac{3\delta t}{2}\right) - \left(\frac{b}{k} \pm 1\right) \\ &= \Big(n - |B|\Big)\left(\frac{1}{k} \pm 2\delta\right)\end{aligned}$$

where the first term counts the number of vertices of T' embedded in V_i (see (3.5)), and the second and third terms count the number of removed leaves embedded in V_i. Indeed, by (3.6) there are $t/k \pm 3\delta t/2$ vertices of W embedded in V_i, each of which is adjacent to two removed leaves, and these removed leaves are each embedded in V_i except for the $\lfloor b/k \rfloor$ or $\lceil b/k \rceil$ leaves embedded in B. □

3.5.2 Embedding the second subtree

Recall from the outline at the beginning of this section that, following the embedding of the first subtree T_1, the vertices which remain uncovered form a super-regular cycle of cluster tournaments. We wish to embed the second subtree T_2 so that all of these vertices are covered. The following lemma demonstrates that this is possible.

Lemma 3.20. Suppose that $1/n \ll 1/C$ and that $1/n \ll 1/k \ll \varepsilon \ll d \ll \beta$. Let T be an oriented tree on n vertices with root r, with maximum degree $\Delta(T) \leq (\log n)^C$, and which contains at least βn distinct vertices that are each adjacent to at least one in-leaf and at least one out-leaf of T. Let G be a (d, ε)-super-regular cycle of cluster tournaments on n vertices whose clusters $V_1, \dots, V_k$ each have size $\frac{n}{k} \pm \frac{2n}{\log\log n}$, and let v be a vertex of V_1. Then G contains a (spanning) copy of T in which r is embedded to v.

Loosely speaking, the proof of Lemma 3.20 begins by removing a small number of in-leaves and out-leaves of T to obtain a subtree T'. We then select small disjoint subsets X_i and Y_i of V_i for each $i \in [k]$ with the property that each vertex in V_i has many inneighbours in each of X_{i-1} and Y_{i-1} and many outneighbours in each of X_{i+1} and Y_{i+1}, and so that most vertices in V_i have large semidegree in X_i. Removing these sets from G yields a subgraph G' of G which is a regular cycle of cluster tournaments, and we embed T' in G' using Lemmas 3.3 and 3.4. It remains to embed the removed leaves of T so as to cover all vertices of G which remain uncovered. We first use the fact that the image of each vertex of T' embedded in V_i has large semidegree in X_i to embed a small number of removed leaves to equalise the numbers of uncovered vertices in each cluster and the numbers of removed leaves needing to be embedded in that cluster, before completing the embedding by using the super-regularity of G to find perfect matchings in appropriate auxiliary bipartite graphs.

Proof. Introduce new constants η and γ such that $\varepsilon \ll \eta \ll \gamma \ll d$. Also define $\delta := \frac{2}{\log\log n}$ and $m := \frac{n}{k}$, so each cluster has size $m \pm \delta n$, assume without loss of generality that $\beta \leq \frac{1}{4}$, and let $t := \lceil \beta n \rceil - 1$. Choose a set W of t distinct vertices of T so that each $w \in W$ is adjacent to at least one in-leaf of T and at least one out-leaf of T and so that r is neither in W nor a leaf of T which is adjacent to a vertex of W (our assumption on T ensures that we can choose such a set W). Let T' be the oriented tree formed by deleting from T precisely one in-leaf and one out-leaf adjacent to each vertex of W, and take r to be the root of T'. Observe that T' then has precisely $n - 2t$ vertices and maximum degree $\Delta(T') \leq \Delta(T) \leq (\log n)^C \leq \big(\log(n-2t)\big)^{2C}$; in other words, T' meets the conditions of Lemma 3.4 with $n - 2t$ and $2C$ in place of n and C respectively. We will embed T' in an appropriate subgraph of G, which we find by using the following claim.

Claim 3.21. For each $i \in [k]$ there exist sets $F_i, X_i, Y_i \subseteq V_i$ with $X_i, Y_i \subseteq F_i$ such that, writing $V_i' := V_i \setminus F_i$, we have

(i) $|F_i| \leq 3\gamma m$,

(ii) X_i and Y_i are disjoint, and $v \in V_1'$,

(iii) for each $x \in V_i' \setminus \{v\}$ we have $\deg^0(x, X_i) \geq \eta m$, and

(iv) for each $x \in V_i$ we have

$$\deg^-(x, X_{i-1}), \deg^-(x, Y_{i-1}), \deg^+(x, X_{i+1}), \deg^+(x, Y_{i+1}) \geq \eta m.$$

Proof. For each $i \in [k]$ let $D_i \subseteq V_i$ be the set of all vertices $x \in V_i$ such that $\deg^0(x, V_i) < \gamma m/5$, so $|D_i| \leq \gamma m$ by Lemma 2.10. Then $\deg^0(x, V_i) \geq \gamma m/5$ for all $x \in V_i \setminus D_i$. Also, since $V_1, \dots, V_k$ are the clusters of a (d, ε)-super-regular cycle of cluster tournaments, every vertex $x \in V_i$ has at least $(d-\varepsilon)|V_{i-1}| \geq dm/2$ inneighbours in V_{i-1} and at least $(d-\varepsilon)|V_{i+1}| \geq dm/2$ outneighbours in V_{i+1}. For each $i \in [k]$ choose disjoint subsets $X_i, Y_i \subseteq V_i$ with $|X_i| = |Y_i| = \lfloor \gamma m \rfloor$ uniformly at random and independently of all other choices. Then for each $i \in [k]$ and each $x \in V_i$ the random variables $\deg^-(x, X_{i-1}), \deg^-(x, Y_{i-1}), \deg^+(x, X_{i+1})$ and $\deg^+(x, Y_{i+1})$ each have hypergeometric distribution with expectation at least $(dm/2)\lfloor \gamma m \rfloor/(m + \delta n) \geq d\gamma m/3 \geq 2\eta m$; if additionally $x \in V_i \setminus D_i$, then the random variables $\deg^+(x, X_i)$ and $\deg^-(x, X_i)$ each have hypergeometric distribution with expectation at least $(\gamma m/5)\lfloor \gamma m \rfloor/(m + \delta n) \geq \gamma^2 m/6 \geq 2\eta m$. The probability that any given one of these random variables is less than $\eta m + 1$ therefore decreases exponentially with n by Theorem 2.14, so by taking a union bound over all of these at most $6n$ events we find that with positive probability none of these random variables is less than $\eta m + 1$. Fix a choice of the sets X_i and Y_i with this property, then removing v from the sets X_1 and Y_1 if necessary and taking $F_1 := (X_1 \cup Y_1 \cup D_1) \setminus \{v\}$ and $F_i := X_i \cup Y_i \cup D_i$ for each $2 \leq i \leq k$ gives the desired sets. ◇

Fix sets F_i, X_i, Y_i and V_i' as in Claim 3.21, and observe that

$$|V_i'| = |V_i| - |F_i| \geq m - \delta n - 3\gamma m \geq \left(1 - \frac{\beta}{4}\right) m \geq (1+\beta)\,\frac{n-2t}{k}.$$

Let $G' := G[V_1' \cup \dots \cup V_k']$, and note that by Lemma 2.7 G' is then a $(d, 2\varepsilon)$-regular cycle of cluster tournaments with clusters $V_1', \dots, V_k'$. Observe also that v has at least $(d-\varepsilon)|V_2| - |F_2| \geq \gamma m$ outneighbours in V_2' and at least $(d-\varepsilon)|V_k| - |F_k| \geq \gamma m$ inneighbours in V_k'. In other words, G' meets the conditions of Lemma 3.4 with $n - 2t$, β and 2ε in place of n, α and ε respectively (so, in particular, m there corresponds to $m - 2t/k$ here).

Apply the Vertex Allocation Algorithm (Algorithm 3.2) to allocate the vertices of T' to the clusters $V_1', \dots, V_k'$ of G'. For each $i \in [k]$ let T_i' consist of all vertices of T' allocated to the cluster V_i', and likewise let $W_i \subseteq W$ consist of all vertices of W allocated to the cluster V_i'. By Lemma 3.3(a) the allocation we obtain from the Vertex Allocation Algorithm will be semi-canonical. Furthermore, by two applications of Lemma 3.3(d) (with $n - 2t$, $2C$ and $\beta/2$ in place of n, C and α respectively) we find with probability $1 - o(1)$ that for every $i \in [k]$ we have

$$|W_i| := |W| \left(\frac{1}{k} \pm \frac{1}{\log\log(n-2t)}\right) = \frac{t}{k} \pm \delta n \qquad (3.7)$$

and

$$|T'_i| := (n-2t)\left(\frac{1}{k} \pm \frac{1}{\log\log(n-2t)}\right) = m - \frac{2t}{k} \pm \delta n. \tag{3.8}$$

Fix an allocation with these properties, and observe that this allocation then meets the conditions of Lemma 3.4 with $n-2t$ and β in place of n and α respectively (since $\delta \ll \beta$). So we may apply Lemma 3.4 to obtain an embedding φ of T' in G' such that each vertex of T' is embedded in the cluster to which it was allocated and so that $\varphi(r) = v$.

For each $i \in [k]$, let $U_i \subseteq V_i$ be the set of vertices of V_i not covered by φ and let $U := \bigcup_{i\in[k]} U_i$. Then, since every vertex was embedded in the cluster to which it was allocated, by (3.8) we have for each $i \in [k]$ that

$$|U_i| = |V_i| - |T'_i| = (m \pm \delta n) - \left(m - \frac{2t}{k} \pm \delta n\right) = \frac{2t}{k} \pm 2\delta n, \tag{3.9}$$

and since $|G| = n$ and $|T'| = n - 2t$ we have $|U| = 2t$. Also, for each $i \in [k]$, let $P_i := \varphi(W_i)$ and write $P := \bigcup_{i\in[k]} P_i$. In other words, P_i (respectively P) is the set of vertices of G to which vertices of W_i (respectively W) were embedded. So $P_i \subseteq V'_i$ and $|P_i| = |W_i|$, and similarly $|P| = |W| = t$.

Our next goal is to choose, for each $x \in P$, an inneighbour x^- of x in U and an outneighbour x^+ of x in U such that the chosen inneighbours and outneighbours are all distinct, and with this we will be able to complete the embedding. Indeed, for each vertex $w \in W$ there is a unique vertex $x \in P$ with $\varphi(w) = x$. Let w^+ and w^- denote the out-leaf and in-leaf adjacent to w which we removed when forming T'; we could then embed w^+ to x^+ and w^- to x^-, and doing so for each $w \in W$ would extend φ to an embedding of T in G, completing the proof. If for every $i \in [k]$ both $G[U_{i-1} \to P_i]$ and $G[P_i \to U_{i+1}]$ are super-regular and $|U_i| = |P_{i-1}| + |P_{i+1}|$, then (after appropriately partitioning the sets U_i) we could apply Lemma 2.8 to find, for each $i \in [k]$ and each $x \in P_i$, vertices $x^- \in U_{i-1}$ and $x^+ \in U_{i+1}$ satisfying the above properties. However, neither of these assumptions is necessarily valid. Over the following steps of the proof we embed the removed leaves adjacent to a small number of vertices of W so that these assumptions do indeed hold for the remaining vertices; we then complete the embedding of T in G in the manner described above.

Step 1: Balancing the sets W_i and ensuring super-regularity. The first step of this process is to embed the removed leaves adjacent to a small number of vertices of W so that equally many vertices in each set W_i have not had their adjacent removed leaves embedded. We also cover all vertices in each set U_i which have too few inneighbours in P_{i-1} or too few outneighbours in P_{i+1}; this will ensure that the auxiliary bipartite graphs which we consider at the end of the proof are super-regular.

For each $i \in [k]$ define $s_i := \lfloor 4\varepsilon m \rfloor + |W_i| - \min_{i\in[k]} |W_i|$, so by (3.7) we have $\lfloor 4\varepsilon m \rfloor \leq s_i \leq 4\varepsilon m + 2\delta n$. For each $i \in [k]$, let B_i^- be the set of vertices in U_i with fewer than ηm inneighbours in P_{i-1}, and let B_i^+ be the set of vertices in U_i with fewer than ηm outneighbours in P_{i+1}. Since $G[V_{i-1} \to V_i]$ is $(d_{\geq}, \varepsilon)$-regular, and $|P_{i-1}| = |W_{i-1}| \geq t/k - \delta n > \varepsilon |V_{i-1}|$ by (3.7), we must have $|B_i^-| \leq \varepsilon |V_i| \leq 2\varepsilon m$; likewise, since $G[V_i \to V_{i+1}]$ is $(d_{\geq}, \varepsilon)$-regular and $|P_{i+1}| > \varepsilon |V_{i+1}|$, we must have $|B_i^+| \leq \varepsilon |V_i| \leq 2\varepsilon m$. So we may choose for each $i \in [k]$ a subset $B_i \subseteq U_i$ of size $|B_i| = s_i$ with $B_i^- \cup B_i^+ \subseteq B_i$.

Next, for each $i \in [k]$ we proceed as follows. Let $\{b_1, \dots, b_{s_i}\}$ be the vertices in B_i, arbitrarily choose distinct vertices $w_1, \dots, w_{s_i} \in W_i$, and for each $j \in [s_i]$ let $p_j := \varphi(w_j)$, so $p_j \in P_i$. Since $W \subseteq V(T') \setminus \{r\}$, for each $j \in [s_i]$ the vertex p_j was embedded in $V_i' \setminus \{v\}$, so by Claim 3.21(iii) we have $\deg^0(p_j, X_i \setminus B_i) \geq \deg^0(p_j, X_i) - |B_i| = \eta m - s_i \geq s_i$. We may therefore choose distinct vertices $x_1, \dots, x_{s_i}$ in $X_i \setminus B_i$ such that for each $j \in [s_i]$, the vertex x_j is an inneighbour of p_j if $b_j \in N^+(p_j)$, whilst x_j is an outneighbour of p_j if $b_j \in N^-(p_j)$. For each $j \in [s_i]$ let w_j^+ be the removed out-leaf of T adjacent to w_j and let w_j^- be the removed in-leaf of T adjacent to w_j. If $b_j \in N^+(p_j)$ then we set $\varphi(w_j^+) = b_j$ and $\varphi(w_j^-) = x_j$, whilst if $b_j \in N^-(p_j)$ then we set $\varphi(w_j^-) = b_j$ and $\varphi(w_j^+) = x_j$. Observe that our choice of vertices $x_1, \dots, x_{s_i}$ ensures that these embeddings are consistent with the directions of the edges $w_j^- \to w_j$ and $w_j \to w_j^+$.

Having carried out these steps for each $i \in [k]$ we have extended the embedding φ to cover all vertices in $B_1 \cup \dots \cup B_k$. For each $i \in [k]$ we now define $W_i^0 := W_i \setminus \{w_1, \dots w_{s_i}\}$ and $P_i^0 := P_i \setminus \{p_1, \dots, p_{s_i}\}$. In other words, W_i^0 is the set of vertices of W which were embedded in V_i' and whose adjacent removed leaves have not yet been embedded, and P_i^0 is the set of vertices of G to which vertices of W_i^0 have been embedded. By (3.7) we then have

$$|P_i^0| = |W_i^0| = |W_i| - s_i = \min_{i\in[k]} |W_i| - \lfloor 4\varepsilon m \rfloor = \frac{t}{k} - 4\varepsilon m \pm \delta n, \tag{3.10}$$

so in particular we have $|W_1^0| = \dots = |W_k^0| = |P_1^0| = \dots = |P_k^0|$. Similarly, for each $i \in [k]$ we define $U_i^0 := U_i \setminus \{b_1, x_1, \dots, b_{s_i}, x_{s_i}\}$. In other words, U_i^0 is the set of vertices of V_i which have not yet been covered by φ. By (3.9) we then have

$$|U_i^0| = |U_i| - 2s_i = \frac{2t}{k} - 8\varepsilon m \pm 6\delta n. \tag{3.11}$$

Write $W^0 := \bigcup_{i\in[k]} W_i^0$, $P^0 := \bigcup_{i\in[k]} P_i^0$, and $U^0 := \bigcup_{i\in[k]} U_i^0$. So in particular U^0 is the set of vertices of G which remain uncovered. Since there are two such vertices

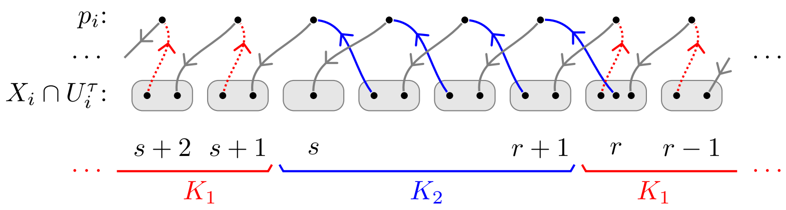


Figure 3.1: This diagram illustrates how we extend the embedding φ at each step of the balancing algorithm. The vertices $p_1, \dots, p_k$ at the top lie in the sets $P_1^\tau, \dots, P_k^\tau$ respectively, and the shaded areas represent the sets $X_i \cap U_i^\tau$ for $i \in [k]$ (that is, the vertices of X_i not yet covered by φ). The extension of φ at step τ then covers the vertices drawn in the shaded areas, so three extra vertices are covered from V_r, one from V_s, and two from each other cluster.

for each vertex of W^0, and $|W_1^0| = |W_2^0| = \cdots = |W_k^0|$, it follows that $|U^0|$ is divisible by $2k$.

Step 2: Balancing the numbers of uncovered vertices. Our next step is to embed the removed leaves adjacent to a small number of vertices of W so that, following these embeddings, there are equally many uncovered vertices within each cluster (while preserving super-regularity as well as the property $|W_1| = \cdots = |W_k|$ from Step 1).

We achieve this by applying the following 'balancing algorithm'. Each iteration of this algorithm will extend φ by embedding, for each $i \in [k]$, the removed in-leaf and out-leaf adjacent to some vertex in W_i.

The balancing algorithm proceeds as follows. For each time $\tau \geq 0$ and for each $i \in [k]$, we let $W_i^\tau \subseteq W_i$ be the set of vertices of T whose adjacent removed leaves have not yet been embedded, we let $P_i^\tau \subseteq P_i$ be the set of vertices of G to which vertices of W_i^τ have been embedded, and we let $U_i^\tau \subseteq V_i$ be the set of uncovered vertices in V_i at time τ. Observe that these definitions of W_i^0, P_i^0 and U_i^0 coincide with those given above. We also define the quantity $M^\tau := \frac{1}{k}\sum_{i\in[k]} |U_i^\tau|$, so M^τ is the average number of uncovered vertices per cluster at time τ.

Our observation above that $|U^0|$ is divisible by $2k$ ensures that M^0 is an even integer, and in fact the algorithm will ensure that M^τ is an even integer at each time $\tau \geq 0$. At time step τ, if $|\,U_i^\tau\,| = M^\tau$ for all $i \in [k]$, then we stop with success. Otherwise, since M^τ is an integer, we may choose $r, s \in [k]$ with $|U_r^\tau| \geq M^\tau + 1$ and $|U_s^\tau| \leq M^\tau - 1$. Define $K_1 := \{s+1, s+2, \dots, r-1, r\}$ and $K_2 := \{r+1, \dots, s\} = [k] \setminus K_1$, with addition taken modulo k. For each $i \in [k]$, we choose a vertex $w_i \in W_i^\tau$, and let $p_i \in P_i^\tau$ be the vertex to which w_i was embedded. We also choose a vertex $x_i^+ \in N^+(p_i) \cap X_{i+1} \cap U_{i+1}^\tau$ and, if $i \in K_1$ then we choose a vertex $x_i^- \in N^-(p_i) \cap X_i \cap U_i^\tau$, whilst if $i \in K_2$ then we choose a vertex $x_i^- \in N^-(p_i) \cap X_{i-1} \cap U_{i-1}^\tau$. We make these choices so that the $2k$ vertices $\{x_1^-, x_1^+, \dots, x_k^-, x_k^+\}$ are all distinct (if it is not possible to make such choices then we terminate with failure, but we shall see shortly that this will not happen).

For each $i \in [k]$ let w_i^+ be the removed out-leaf of T adjacent to w_i and let w_i^- be the removed in-leaf of T adjacent to w_i; we then set $\varphi(w_i^-) := x_i^-$ and $\varphi(w_i^+) := x_i^+$ (see Figure 3.1 for an illustration of this embedding). To conclude this iteration of the algorithm, for each $i \in [k]$ we update the sets W_i^τ, P_i^τ and U_i^τ by setting $W_i^{\tau+1} := W_i^\tau \setminus \bigcup_{i\in[k]}\{w_i\}$, $P_i^{\tau+1} := P_i^\tau \setminus \bigcup_{i\in[k]}\{p_i\}$, and $U_i^{\tau+1} := U_i^\tau \setminus \bigcup_{i\in[k]}\{x_i^+, x_i^-\}$. Observe that we then have

$$|U_i^{\tau+1}| = \begin{cases} |U_i^\tau| - 3 & \text{if } i = r, \\ |U_i^\tau| - 1 & \text{if } i = s, \text{ and} \\ |U_i^\tau| - 2 & \text{otherwise.} \end{cases} \tag{3.12}$$

In particular it follows that $M^{\tau+1} = M^\tau - 2$; since M^τ was an even integer it follows that $M^{\tau+1}$ is an even integer, as required.

Claim 3.22. The balancing algorithm described above stops with success after at most $3k\delta n$ iterations.

Proof. We first check that we can choose vertices w_i, p_i, x_i^- and x_i^+ as described whenever $\tau \leq 3k\delta n$. First observe that, for each $i \in [k]$, since $|W_i^0| \geq t/2k > 3k\delta n$ by (3.10), and at most one vertex is removed from W_i^τ at each step τ of the balancing algorithm (and its image is removed from P_i^τ), there are at least $|W_i^0| - \tau \geq 1$ possible choices for w_i at step $\tau \leq 3k\delta n$ of the balancing algorithm. So we may choose the vertices w_i and p_i for $i \in [k]$ as claimed. Next observe that for each $i \in [k]$ at most $2s_i \leq 8\varepsilon m + 4\delta n$ vertices were embedded in X_i in Step 1. Also, each iteration of the balancing algorithm embeds at most three vertices in X_i, so at time $\tau \leq 3k\delta n$ the total number of vertices which have so far been embedded in X_i is at most $3\tau + 8\varepsilon m + 4\delta n \leq 9k\delta n + 9\varepsilon m \leq \eta m/2$. Since $p_i \in V_i'$, it follows by Claim 3.21(iii) and (iv) that $\deg^-(p_i, X_{i-1} \cap U_{i-1}^\tau) \geq \eta m/2$, that $\deg^+(p_i, X_{i+1} \cap U_{i+1}^\tau) \geq \eta m/2$ and that $\deg^-(p_i, X_i \cap U_i^\tau) \geq \eta m/2$. So we may greedily choose the vertices x_i^- and x_i^+ for each $i \in [k]$ as desired.

It therefore suffices to prove that the algorithm stops after at most $3k\delta n$ iterations and thus, because it cannot fail in these early steps, it always stops successfully. For each $\tau \geq 0$ let $\Upsilon^\tau := \sum_{i\in[k]} \Big| |U_i^\tau| - M^\tau \Big|$, so Υ^τ is a non-negative integer. In particular by (3.11) we have

$$\Upsilon^0 = \sum_{i\in[k]} \Big| |U_i^0| - M^0 \Big| \leq 6k\delta n, \tag{3.13}$$

Also, by (3.12) we have $|U_r^{\tau+1}| = |U_r^\tau| - 3$ and $M^{\tau+1} = M^\tau - 2$; by our choice of r it follows that $\left| |U_r^{\tau+1}| - M^{\tau+1} \right| = \left| |U_r^\tau| - M^\tau \right| - 1$. Similarly we find that $\left| |U_s^{\tau+1}| - M^{\tau+1} \right| = \left| |U_s^\tau| - M^\tau \right| - 1$ and that $\left| |U_j^{\tau+1}| - M^{\tau+1} \right| = \left| |U_j^\tau| - M^\tau \right|$ for each $j \in [k] \setminus \{r, s\}$. Together these equalities imply that $\Upsilon^{\tau+1} = \Upsilon^\tau - 2$. Since Υ^τ is always non-negative, we conclude that for some $\tau \leq 3k\delta n$ we must have $\Upsilon^\tau = 0$. It follows that $|U_j^\tau| = M^\tau$ for all $j \in [k]$, and so the algorithm will stop at step τ. $\diamond$

Returning to the proof of Lemma 3.20, we conclude that the balancing algorithm will stop with success at some time τ_{end} with $\tau_{\text{end}} \leq 3k\delta n$. For each $i \in [k]$, let $W_i^* := W_i^{\tau_{\text{end}}}$, $P_i^* := P_i^{\tau_{\text{end}}}$, and $U_i^* := U_i^{\tau_{\text{end}}}$, and write $W^* := \bigcup_{i\in[k]} W_i^*$, $P^* := \bigcup_{i\in[k]} P_i^*$, and $U^* := \bigcup_{i\in[k]} U_i^*$. So the embedding φ now covers all vertices of $V(G)$ except for those in U^*, and the only vertices of T which remain to be embedded are one in-leaf and one out-leaf of each vertex of W^*. In particular we have $|U^*| = 2|P^*| = 2|W^*|$. Observe that in the execution of the balancing algorithm, at each time τ and for each $i \in [k]$ precisely one vertex was removed from W_i^τ. Therefore, since we initially had $|W_1^0| = \cdots = |W_k^0|$ by (3.10), we now have $|W_1^*| = \cdots = |W_k^*|$. We denote this common size by L, and note that by (3.10) we then have $L \geq t/k - 4\varepsilon m - \delta n - \tau_{\text{end}} \geq 2t/3k$. Also, since $\Upsilon^{\tau_{\text{end}}} = 0$, we must have $|U_1^*| = \cdots = |U_k^*| = M^{\tau_{\text{end}}}$, so

$$L = |W_1^*| = \cdots = |W_k^*| = |P_1^*| = \cdots = |P_k^*| = \frac{1}{2}|U_1^*| = \cdots = \frac{1}{2}|U_k^*| \geq \frac{2t}{3k} \geq \frac{\beta m}{2}. \tag{3.14}$$

► **Step 3: Completing the embedding.** We are now ready to complete the embedding of T in G as described previously, beginning with the following claim.

Claim 3.23. For each $i \in [k]$ each vertex in U_i^* has at least $\eta m/2$ inneighbours in P_{i-1}^* and at least $\eta m/2$ outneighbours in P_{i+1}^*, and each vertex in P_i^* has at least ηm inneighbours in U_{i-1}^* and at least ηm outneighbours in U_{i+1}^*.

Proof. Recall that the set B_i chosen in Step 1 contained all vertices of U_i with fewer than ηm inneighbours in P_{i-1} or fewer than ηm outneighbours in P_{i+1}. All vertices of B_i were covered in Step 1, so no vertex of B_i is contained in U_i^*. The first statement then follows from the fact that for each $j \in [k]$ we have

$$|P_j \setminus P_j^*| = |P_j \setminus P_j^0| + |P_j^0 \setminus P_j^*| \leq s_j + \tau_{\text{end}} \leq 4\varepsilon m + 2\delta n + 3k\delta n \leq \frac{\eta m}{2}.$$

For the second statement observe that no vertices have yet been embedded in any set Y_j, so $Y_{i-1} \subseteq U_{i-1}^*$ and $Y_{i+1} \subseteq U_{i+1}^*$. Moreover, since $P_i^* \subseteq P_i \subseteq V_i$, by

Claim 3.21(iv) every vertex of P_i^* has at least ηm inneighbours in Y_{i-1} and at least ηm outneighbours in Y_{i+1}. $\diamond$

For each $i \in [k]$ we now partition U_i^* into disjoint sets U_i^- and U_i^+ each of size L uniformly at random and independently of all other choices. Since $G[V_i \to V_{i+1}]$ is $(d_{\geq}, \varepsilon)$-regular for each $i \in [k]$, by (3.14) and Lemma 2.7 both $G[U_{i-1}^- \to P_i^*]$ and $G[P_i^* \to U_{i+1}^+]$ are then $(d_{\geq}, \varepsilon')$-regular, where $\varepsilon' := 3\varepsilon/\beta$. Also, by Claim 3.23, each $u \in U_{i-1}^-$ has $\deg^+(u, P_i^*) \geq \eta m/2 \geq \eta L/2$ and each $u \in U_{i+1}^+$ has $\deg^-(u, P_i^*) \geq \eta m/2 \geq \eta L/2$. Furthermore, for each $p \in P_i^*$ the random variables $\deg^-(p, U_{i-1}^-)$ and $\deg^+(p, U_{i+1}^+)$ each have hypergeometric distribution with expectation at least $\eta m L/2L \geq \eta L/2$. Applying Theorem 2.14 and taking a union bound we find that with positive probability we have for every $i \in [k]$ and every $p \in P_i^*$ that $\deg^-(p, U_{i-1}^-) \geq \eta L/4$ and $\deg^+(p, U_{i+1}^+) \geq \eta L/4$. Fix such an outcome of our random selection; then for each $i \in [k]$ the underlying graphs of both $G[U_{i-1}^- \to P_i^*]$ and $G[P_i^* \to U_{i+1}^+]$ are $(\eta/4, \varepsilon')$-super-regular balanced bipartite graphs with vertex classes of size L.

We may therefore apply Lemma 2.8 to obtain, for each $i \in [k]$, a perfect matching M_i^- in $G[U_{i-1}^- \to P_i^*]$ and a perfect matching M_i^+ in $G[P_i^* \to U_{i+1}^+]$. For each $i \in [k]$ and each $w \in W_i^*$ let w^- be the removed in-leaf of T adjacent to w and let w^+ be the removed out-leaf of T adjacent to w. Also let $p = \varphi(w)$ and let $q^- \in U_{i-1}^-$ and $q^+ \in U_{i+1}^+$ be the vertices matched to p in M_i^- and M_i^+ respectively, and set $\varphi(w^-) := q^-$ and $\varphi(w^+) := q^+$. Since each $p \in P_i^*$ is matched to precisely one inneighbour in U_{i-1}^- and precisely one outneighbour in U_{i+1}^+, this extends φ to an embedding of T in G. $\square$

3.5.3 Joining the pieces

As outlined at the start of this section, we will 'split' our tree T into two subtrees T_1 and T_2, which we embed successively in G using Lemmas 3.19 and 3.20. Recall from Definition 2.1 that a tree-partition splits a tree T in 2 edge-disjoint trees which together contain all vertices and edges of the original tree, and that given any set $L \subseteq V(T)$ of vertices, there is a tree-split where each of the pieces contains a third of the vertices of L (Lemma 2.3). We now state and prove the main result of this section.

Lemma 3.24. Suppose that $1/n \ll 1/C$ and that $1/n \ll 1/k \ll \varepsilon \ll d \ll \psi \ll \alpha$. Let T be an α-nice oriented tree on n vertices with maximum degree $\Delta(T) \leq (\log n)^C$. Also let G be a tournament on n vertices which contains a (d, ε)-regular cycle of cluster tournaments whose clusters $V_1, \dots, V_k$ have equal size such that $B := V(G) \setminus \bigcup_{i\in[k]} V_i$ has size $|B| \leq \psi n$. Then G contains a (spanning) copy of T.

Proof of Lemma 3.24. Introduce a new constant β with $\psi \ll \beta \ll \alpha$, and define $m := |V_1| = \cdots = |V_k| = \big(n - |B|\big)/k$ and $s := \lceil \alpha n \rceil$. Since T is α-nice we may choose a set L of s distinct vertices of T such that each vertex in L is adjacent to at least one in-leaf and at least one out-leaf of T. Apply Lemma 2.3 to obtain a tree-partition $\{T_1, T_2\}$ of T such that the subtrees T_1 and T_2 each contain at least $s/3$ vertices of L. Let r be the unique common vertex of T_1 and T_2, which we take as the root of each subtree, and observe that for each vertex $x \neq r$ every neighbour of x is contained in the same subtree as x. So in particular T_1 contains at least $s/3 - 1 \geq \alpha n/4 \geq \beta n$ vertices each adjacent to at least one in-leaf and at least one out-leaf of T_1, and likewise T_2 contains at least $\alpha n/4 \geq \beta n$ vertices each adjacent to at least one in-leaf and at least one out-leaf of T_2. Write $n_1 := |T_1|$ and $n_2 := |T_2|$, so $3\alpha n/4 \leq n_1, n_2$ and $n_1 + n_2 = n + 1$. By relabelling if necessary we may assume that $n_1 \leq n_2$. Observe also that $\Delta(T_1) \leq \Delta(T) \leq (\log n)^C \leq (\log n_1)^{2C}$ and likewise that $\Delta(T_2) \leq (\log n_2)^{2C}$. So T_1 meets the conditions of Lemma 3.19 with $2C$ and n_1 in place of C and n respectively, and likewise T_2 meets the conditions of Lemma 3.20 with $2C$ and n_2 in place of C and n respectively.

Next, proceed as follows for each $i \in [k]$. Define

$$B_i^+ := \Big\{ v \in V_i : \deg^+(v, V_{i+1}) < (d - \varepsilon)m \Big\}, \quad \text{and}$$
$$B_i^- := \Big\{ v \in V_i : \deg^-(v, V_{i-1}) < (d - \varepsilon)m \Big\}.$$

Since $G[V_{i-1} \to V_i]$ and $G[V_i \to V_{i+1}]$ are each $(d_\geq, \varepsilon)$-regular, we must then have $|B_i^-|, |B_i^+| < \varepsilon m$. Let B_i be a set of $2\varepsilon m$ vertices such that $B_i^- \cup B_i^+ \subseteq B_i \subseteq V_i$ and define $V_i' := V_i \setminus B_i$. It follows that for every vertex $x \in V_i'$ we have $\deg^-(v, V_{i-1}'), \deg^+(v, V_{i+1}') \geq (d-\varepsilon)m - 2\varepsilon m = (d - 3\varepsilon)m$. Choose a subset $Z_i \subseteq V_i'$ of size $|Z_i| = \alpha m/5$ uniformly at random and independently of all other choices. So for each $x \in V_i'$ the random variables $\deg^-(x, Z_{i-1})$ and $\deg^+(x, Z_{i+1})$ have hypergeometric distributions with expectation at least $(d-3\varepsilon)\alpha m/5$. Applying Theorem 2.14 and taking a union bound we find that with positive probability we have for every $x \in V_i'$ that $\deg^-(x, Z_{i-1}), \deg^+(x, Z_{i+1}) \geq d\alpha m/10$. Fix an outcome of the random selections for which this event occurs.

Define $B' := B \cup \bigcup_{i\in[k]} B_i$, so $|B'| = |B| + 2k\varepsilon m \leq 2\psi n$. Next choose arbitrarily a set $X_i \subseteq V_i' \setminus Z_i$ of size $(1+\alpha/4)n_1/k$ for each $i \in [k]$; this is possible since for each $i \in [k]$ we have

$$\begin{aligned}|V_i' \setminus Z_i| &= (1-2\varepsilon)m - \frac{\alpha m}{5} \\ &\geq \left(1 - \frac{\alpha}{4}\right) m \geq \left(1-\frac{\alpha}{4}\right)(1-\psi)\frac{n}{k} \geq \left(1-\frac{\alpha}{3}\right)\frac{n}{k} \geq \left(1+\frac{\alpha}{3}\right)\frac{n_1}{k},\end{aligned}$$

where the final inequality uses the fact that $n_1 \leq n+1-n_2 \leq n+1-3\alpha n/4 \leq (1-2\alpha/3)n$. Define $G_1 := G\Big[B' \cup \bigcup_{i\in[k]} X_i\Big]$. Since $G[V_i \to V_{i+1}]$ is $(d_{\geq}, \varepsilon)$-regular for each $i \in [k]$, and $n_1 \geq 3\alpha n/4$, it follows by Lemma 2.7 that the sets $X_1, \ldots, X_k$ are the clusters of a (d, ε')-regular cycle of cluster tournaments in G_1, where $\varepsilon' := 4\varepsilon/3\alpha$. The tournament G_1, the clusters X_i and the set B' therefore meet the conditions of Lemma 3.19 with $n_1, \alpha/3, \varepsilon'$ and 2ψ in place of n, α, ε and ψ respectively. So we may apply Lemma 3.19 to obtain an embedding φ of T_1 in G_1 so that r is embedded in X_1, so that every vertex of B' is covered, and so that for each $i \in [k]$ we have

$$\begin{aligned}\Big|\varphi\Big(V(T)\Big) \cap X_i\Big| &= \Big(n_1 - |B'|\Big)\left(\frac{1}{k} \pm \frac{2}{\log\log n_1}\right) \\ &= \frac{n_1 - |B'|}{k} \pm \left(\frac{2n_1}{\log\log n_1} - 2\right), \qquad (3.15)\end{aligned}$$

where the last equality holds with a lot of room to spare (since $|B'|/\log\log n_1 \gg 2$).

For convenience of notation write $E := \frac{2n_2}{\log\log n_2} \geq \frac{2n_1}{\log\log n_1}$. For each $i \in [k]$ define $U_i := V_i \setminus \varphi\Big(V(T)\Big)$, so U_i contains all vertices of V_i not covered by our embedding of T_1. Then by (3.15) we have for each $i \in [k]$ that

$$\begin{aligned}|U_i| &= \Big|V_i \setminus B_i\Big| - \frac{n_1 - |B'|}{k} \pm (E-2) = m - 2\varepsilon m - \frac{n_1}{k} + \frac{|B| + 2k\varepsilon m}{k} \pm (E-2) \\ &= \frac{n - |B|}{k} - \frac{n_1}{k} + \frac{|B|}{k} \pm (E-2) = \frac{n_2}{k} \pm (E-1),\end{aligned}$$

where the second equality uses the fact that $|B'| = |B| + 2k\varepsilon m$, and the final equality uses the fact that $n_2 = n+1-n_1$. Let $v = \varphi(r)$, so $v \in X_1$, and set $U_1^* := U_1 \cup \{v\}$ and $U_i^* := U_i$ for $2 \leq i \leq k$, so $|U_i^*| = \frac{n_2}{k} \pm E$ for each $i \in [k]$. In particular, we have $|U_i^*| \geq \alpha n/2k \geq \alpha|V_i|/2$ for each $i \in [k]$, so by Lemma 2.7 the sets $U_1^*, \ldots, U_k^*$ are the clusters of a spanning $(d, 2\varepsilon/\alpha)$-regular cycle of cluster tournaments in the tournament $G_2 := G[U_1^* \cup \cdots \cup U_k^*]$. Furthermore, for each $i \in [k]$ we have $Z_i \subseteq U_i^* \subseteq V_i'$ (since we chose X_i to be disjoint from Z_i, and every vertex of B was covered by the embedding of T_1), so every vertex $u \in U_i^*$ has $\deg^-(x, U_{i-1}^*), \deg^+(x, U_{i+1}^*) \geq d\alpha m/10$. So in fact the clusters $U_1^*, \ldots, U_k^*$ form a spanning $(d\alpha/10, 2\varepsilon/\alpha)$-super-regular cycle of cluster tournaments in G_2.

In other words, the tournament G_2, the clusters $U_1^*, \dots, U_k^*$ and the vertex v meet the conditions of Lemma 3.20 with $d\alpha/10, 2\varepsilon/\alpha$ and n_2 in place of d, ε and n respectively. Since $|G_2| = |G| - |T_1| + 1 = n - n_1 + 1 = n_2 = |T_2|$ we may therefore apply Lemma 3.20 to find a spanning copy of T_2 in G_2 in which r is embedded to v, and then the embeddings of T_1 and T_2 together form a spanning copy of T in G. $\square$

3.6 Unavoidable trees (Theorems 1.4 and 1.6)

In this section we give the proofs of Theorem 1.4 (that every large nice oriented tree of polylogarithmic maximum degree is unavoidable) and Theorem 1.6 (that a random labelled oriented tree is nice asymptotically almost surely).

3.6.1 A class of unavoidable trees (Theorem 1.4)

We begin by combining the results of the previous two sections to prove Theorem 1.4. The main task is to use Lemma 3.1 to show that we can find either an almost-directed pair in G which partitions $V(G)$ or an almost-spanning cycle of cluster tournaments in G. In the former case we then embed T in G using Lemma 3.18, whilst in the latter case we embed T in G using Lemma 3.24.

Proof of Theorem 1.4. Introduce new constants $k_0, k_1, \varepsilon, d, \mu, \eta, \omega$ and γ such that

$$\frac{1}{n} \ll \frac{1}{k_1} \ll \frac{1}{k_0} \ll \varepsilon \ll d \ll \mu \ll \eta \ll \omega \ll \gamma \ll \alpha.$$

We may also assume that $1/n \ll 1/C$. Let G be a tournament on n vertices, and let T be an α-nice tree on n vertices such that $\Delta(T) \leq (\log n)^C$. We begin by finding an almost-spanning subgraph of G induced by three vertex-disjoint subsets X, Y, Z of $V(G)$. We will argue that either (i) $G[Y]$ is a large (d, ε)-regular cycle of cluster tournaments (and hence $T \subseteq G$ by Lemma 3.24) or (ii) one of the pairs $(X, Y \cup Z), (X \cup Y, Z)$ is large and almost-directed (and hence $T \subseteq G$ by Lemma 3.18).

We choose vertex-disjoint subsets $X, Y, Z \subseteq V(G)$ such that

(a) $X \cup Y \cup Z = V(G)$,

(b) $|Y| \geq n/3$, and

(c) $e\Big(G[Y \to X]\Big) + e\Big(G[Z \to X]\Big) + e\Big(G[Z \to Y]\Big) \leq \min\Big(\eta\Big(|X| + |Z|\Big)n, 3\gamma\eta n^2\Big)$.

Moreover, we make this choice so that $|Y|$ is minimal among all choices of X, Y and Z which satisfy (a)–(c) above (taking $Y = V(G)$ and $X = Z = \varnothing$ shows that such subsets do exist).

Suppose first that $|Y| \leq (1-2\gamma)n$. Then we have either $|X| \geq \gamma n$ or $|Z| \geq \gamma n$. If $|X| \geq \gamma n$ then, taking $A \coloneqq X$ and $B \coloneqq Y \cup Z$, we have a partition $\{A, B\}$ of $V(G)$ into sets $|A|, |B| \geq \gamma n$ such that the number of edges directed from B to A is $e\Big(G[Y \to X]\Big) + e\Big(G[Z \to X]\Big) \leq 3\gamma\eta n^2 \leq \omega|A||B|$ by (c), so (A, B) is an ω-almost-directed pair in G. If instead $|Z| \geq \gamma n$ then a similar argument shows that taking $A \coloneqq X \cup Y$ and $B = Z$ gives a partition $\{A, B\}$ of $V(G)$ into sets $|A|, |B| \geq \gamma n$ such that (A, B) is an ω-almost-directed pair in G. Either way, we may then apply Lemma 3.18 (with ω and γ in place of μ and ν respectively) to find a copy of T in G.

Now suppose instead that $|Y| > (1-2\gamma)n$, and write $G' \coloneqq G[Y]$. Observe in particular that we then have $|X| + |Z| = n - |Y| < 2\gamma n$, so (c) states that $e\Big(G[Y \to X]\Big) + e\Big(G[Z \to X]\Big) + e\Big(G[Z \to Y]\Big) \leq \eta\Big(|X| + |Z|\Big)n$. If there exists a vertex $y \in Y$ with $\deg^-_{G'}(y) < \eta n$, then moving y from Y to X would increase $e\Big(G[Y \to X]\Big)$ by less than ηn whilst increasing $|X|$ by one and leaving $e\Big(G[Z \to X]\Big) + e\Big(G[Z \to Y]\Big)$ and $|Z|$ unchanged. The resulting sets would then satisfy (a), (b) and (c) with a smaller value of $|Y|$, contradicting the minimality of $|Y|$ in our choice of X, Y and Z. So every vertex $y \in Y$ must have $\deg^-_{G'}(y) \geq \eta n$. Likewise, if there exists a vertex $y \in Y$ with $\deg^+_{G'}(y) < \eta n$, then we obtain a similar contradiction by moving y from Y to Z. We conclude that every vertex $y \in Y$ must have $\deg^+_{G'}(y) \geq \eta n$, so $\delta^0(G') \geq \eta n \geq \eta|Y|$. Now suppose that there exists a partition $\{S, S'\}$ of Y such that (S, S') is a μ-almost-directed pair in G'. Observe that moving all vertices of S from Y to X would increase $e\Big(G[Y \to X]\Big)$ by at most $e(S' \to S) \leq \mu|S||S'| \leq \gamma\eta|S|n$ whilst increasing $|X|$ by $|S|$ and leaving $e\Big(G[Z \to X]\Big) + e\Big(G[Z \to Y]\Big)$ and $|Z|$ unchanged. So if $|S| \leq n/2$, then at least $|Y| - n/2 \geq n/3$ vertices would remain in Y, and so the resulting sets would satisfy (a), (b) and (c) with a smaller value of $|Y|$, again contradicting the minimality of $|Y|$. On the other hand, if $|S| > n/2$ then $|S'| \leq n/2$, and we obtain a similar contradiction by moving all vertices of S' from Y to Z. We conclude that no such partition $\{S, S'\}$ of Y exists. Therefore by Lemma 3.1 there is an integer k with $k_0 \leq k \leq k_1$ such that G' contains a (d, ε)-regular cycle of cluster tournaments with clusters $V_1, \dots, V_k$ of equal size such that $|\bigcup_{i \in [k]} V_i| > (1-\varepsilon)|Y| \geq (1-3\gamma)n$. We may therefore apply Lemma 3.24 (with 3γ in place of ψ) to obtain a copy of T in G. $\square$

3.6.2 Most oriented trees are nice (Theorem 1.6)

We now turn to the proof of Theorem 1.6, for which we use the following classical result, known as Cayley's theorem.

Theorem 3.25 (Borchardt 1860; Cayley 1889). There exist precisely n^{n-2} distinct labelled oriented trees or order n.

A *cherry* is a path of length two, and its *centre* is the vertex of degree two. In an oriented tree T we refer to an in-subtree (respectively out-subtree) which is an (oriented) cherry as an *in-cherry* (respectively *out-cherry*). Our next lemma states that most labelled undirected trees have many pendant cherries. This is a special case of a much more general result for simply generated trees due to Janson [44]. For completeness, we include a proof of the particular statement that suffices for our purposes.

cherry, centre

in-cherry, out-cherry

Lemma 3.26. Fix $\varepsilon > 0$, and let T be a tree chosen uniformly at random from the set of all labelled undirected trees with vertex set $[n]$. Then asymptotically almost surely T contains $(1 \pm \varepsilon)\frac{e^{-3}}{2}n$ pendant cherries.

Proof. For each set $S \in \binom{[n]}{3}$, let $\widehat{S}$ be the indicator random variable which has value 1 if S spans a pendant cherry in T and 0 otherwise. We first note that

$$\mathbb{P}(\,\widehat{S} = 1\,) = \frac{3(n-3)(n-3)^{n-5}}{n^{n-2}} = \frac{3}{n^2}\left(1 - \frac{3}{n}\right)^{n-4}.$$

Indeed, there are three possible choices for the centre of the cherry, this centre is adjacent to one of the $n-3$ vertices in $[n] \setminus S$, and by Theorem 3.25 there are $(n-3)^{n-5}$ distinct possibilities for the undirected labelled tree spanned by $[n] \setminus S$, giving the numerator, whilst the denominator is simply the total number of labelled undirected trees on n vertices (again by Theorem 3.25). The number of pendant cherries in T is $X := \sum_{S\in\binom{[n]}{3}} \widehat{S}$, so by linearity of expectation it follows that

$$\mathbb{E}(X) = \sum_{S\in\binom{[n]}{3}} \mathbb{P}(\widehat{S} = 1) = \binom{n}{3}\frac{3}{n^2}\left(1 - \frac{3}{n}\right)^{n-4} = \big(1 + o(1)\big)\frac{e^{-3}}{2}n. \tag{3.16}$$

It therefore suffices to show that X is concentrated around $\mathbb{E}(X)$. Consider any distinct $S, S' \in \binom{[n]}{3}$, and note that if S intersects S' then we must have $\widehat{S} \cdot \widehat{S'} = 0$. On the other hand, if S and S' are disjoint then by a similar argument as above we have

$$\mathbb{E}\big(\widehat{S} \cdot \widehat{S'}\big) = \mathbb{P}(\,\widehat{S} = \widehat{S'} = 1\,) = \frac{\big[3(n-6)\big]^2 (n-6)^{n-8}}{n^{n-2}} = \frac{9}{n^4}\left(1 - \frac{6}{n}\right)^{n-6},$$

so

$$\begin{aligned}\mathbb{E}(X^2) &= \mathbb{E}\Big(\sum_{S\in\binom{[n]}{3}} \widehat{S}^2 + \sum_{\substack{S,S'\in\binom{[n]}{3}\\ S\neq S'}} \widehat{S}\cdot\widehat{S'}\Big) = \sum_{S\in\binom{[n]}{3}} \mathbb{E}(\widehat{S}) + \sum_{\substack{S,S'\in\binom{[n]}{3}\\ S\cap S'=\varnothing}} \mathbb{E}\Big(\widehat{S}\cdot\widehat{S'}\Big)\\ &= \binom{n}{3}\frac{3}{n^2}\left(1-\frac{3}{n}\right)^{n-4} + \binom{n}{3}\binom{n-3}{3}\frac{9}{n^4}\left(1-\frac{6}{n}\right)^{n-6}. \end{aligned} \tag{3.17}$$

Combining (3.16) and (3.17) we find that

$$\begin{aligned}\mathrm{Var}(X) &= \binom{n}{3}\frac{3}{n^2}\left(1-\frac{3}{n}\right)^{n-4} + \binom{n}{3}\binom{n-3}{3}\frac{9}{n^4}\left(1-\frac{6}{n}\right)^{n-6}\\ &\quad - \left[\binom{n}{3}\frac{3}{n^2}\left(1-\frac{3}{n}\right)^{n-4}\right]^2\\ &= \Big(1+\mathrm{o}(1)\Big)\frac{\mathrm{e}^{-3}}{2}n + \Big(1+\mathrm{o}(1)\Big)\frac{\mathrm{e}^{-6}}{4}n^2 - \left(\Big(1+\mathrm{o}(1)\Big)\frac{\mathrm{e}^{-3}}{2}n\right)^2 = \mathrm{o}(n^2). \end{aligned} \tag{3.18}$$

By Theorem 2.12 (Chebyshev's inequality), (3.16) and (3.18) it follows that

$$\mathbb{P}\left(\Big|X-\mathbb{E}(X)\Big| > \frac{\varepsilon}{2}\cdot\mathbb{E}(X)\right) \leq \frac{\mathrm{Var}(X)}{\Big(\varepsilon\mathbb{E}(X)/2\Big)^2} = \mathrm{o}(1),$$

which together with (3.16) proves the lemma. □

We can now prove Theorem 1.6 (almost all labelled oriented trees are $\frac{1}{250}$-nice).

Proof of Theorem 1.6. Let $\mathcal{T}_n$ be the set of all labelled oriented trees with vertex set $[n]$. Note that we can select an oriented tree T uniformly at random from $\mathcal{T}_n$ using the following two-step random procedure: first select a tree T_0 uniformly at random from the set of all labelled undirected trees with vertex set $[n]$, then form a labelled oriented tree T by orienting each edge e of T_0 uniformly at random and independently of all other choices. Indeed, since there are n^{n-2} possibilities for T_0 by Theorem 3.25, and every tree in $\mathcal{T}_n$ has $n-1$ edges, the probability that any given labelled oriented tree T is selected by this two-step procedure is $n^{2-n}2^{1-n}$; in other words, this is a uniformly-random selection of $T \in \mathcal{T}_n$.

Let C be the number of pendant cherries of T_0, let X be the number of pendant in-cherries of T which contain an out-leaf of T, and let Y be the number of pendant out-cherries of T which contain both an in-leaf and out-leaf of T. Observe that the probability that a fixed pendant cherry of T_0 contributes to X is $3/8$, and likewise the probability that a fixed pendant cherry of T_0 contributes to Y is $1/4$. So $X \sim \mathcal{B}(C, 3/8)$ and $Y \sim \mathcal{B}(C, 1/4)$. Since by Lemma 3.26 we have $C \geq n/50$ asymptotically almost surely (where we use the fact that $\mathrm{e}^{-3}/2 > 1/50$), it follows by Theorem 2.14 that we also have $|X|, |Y| \geq C/5 \geq n/250$ asymptotically almost surely. Since no pendant cherry of T can be counted by both X and Y, it follows that T is $\frac{1}{250}$-nice. $\square$

That government is not a necessary good but an unavoidable evil

Lyn Nofziger

If luck weren't involved, I'd win every tournament!

Phil Hellmuth

There's always going to be comparisons, and that's unavoidable.

Raymond E. Feist

Mistakes in themselves are unavoidable.

Ernest Mandel

4 Spanning Structures via Semidegree

This chapter is organised as follows. We begin in Section 4.1 by outlining the proof of our main result (Theorem 1.13, a sufficient condition for a tree to be contained in every digraph of high minimum semidegree). Section 4.2 introduces notation and auxiliary results. Section 4.3 is devoted to analysing the randomised allocation algorithm, and Section 4.3.4 deals with the embedding algorithm and its analysis. Next, Section 4.4 contains the proof of Theorem 1.13 for trees with many bare paths, while Section 4.5 contains the proof for trees with many leaves. We combine these results in Section 4.6, proving Theorems 1.12 (that every large digraph with high minimum semidegree contains every spanning tree with bounded maximum degree), 1.13 and 1.15 (an extension of Theorem 1.12 for tree-like spanning digraphs).

4.1 Proof outline for Theorem 1.13

Our proof of Theorem 1.13 builds on the ideas of the previous chapter. The fact that G has large semidegree will provide us with a useful structure in the reduced graph (a regular expander), which will be crucial to achieving a good distribution of vertices among clusters in the allocation phase. Our goal is to allocate vertices of T evenly to clusters in that structure, following which we embed T using a greedy algorithm.

More precisely, the main difference in allocation is due to the fact that we can no longer embed edges of T 'within' clusters. A major consequence of this is that the allocation cannot simply proceed along a cycle in the reduced graph R, because doing so could no longer distribute vertices evenly. The high semidegree of G, however, guarantees that the reduced graph has good expansion properties, and thus the allocation proceeds mapping edges of T to edges of an expander subdigraph J of R.

Loosely speaking, if G is a large directed graph of high semidegree, then we may partition $V(G)$ into clusters $V_1, \dots, V_k$ of the same size plus a small set $V^\star$ of atypical vertices such that many pairs of clusters form super-regular pairs. Moreover, we may insist that the pairs along a cycle of clusters $V_1, V_2, \dots, V_k, V_1$ are super-regular (in the direction $1 \to \cdots \to k \to 1$). Also, every large tree T either contains many leaves or contains a collection of many 'bare paths' of bounded length. For every large digraph of order n and semidegree at least $(1/2+\alpha)n$, and for every oriented tree of order n with maximum degree $(\log n)^C$ we consider separately the two cases for the structure of T.

Suppose first that T contains many such bare paths. To define a homomorphism (i.e., an allocation) from T to the reduced graph R of G, we select three disjoint collections $\mathcal{P}_{\mathrm{A}}, \mathcal{P}_{\mathrm{B}}, \mathcal{P}_{\mathrm{C}}$ of vertex-disjoint bare paths of order 7 in T such that all paths in $\mathcal{P}_{\mathrm{A}} \dot\cup \mathcal{P}_{\mathrm{B}} \dot\cup \mathcal{P}_{\mathrm{C}}$ lie in a small subtree T' of T. We then contract all edges in these paths and apply a randomised algorithm to define a homomorphism of the contracted T' to the reduced graph of G. Next, we extend this mapping to all contracted edges so that the mapping of these paths satisfies useful properties (this completes a homomorphism of T'). In particular, we are careful when mapping the contracted paths, and use the fact that the reduced graph has very large semidegree to (A) force the paths in $\mathcal{P}_{\mathrm{A}}$ to go through all atypical vertices of G; (B) ensure that many edges of bare paths in $\mathcal{P}_{\mathrm{B}}$ are allocated along the cycle $V_1 \to \cdots \to V_k \to V_1$; and (C) ensure that bare paths in $\mathcal{P}_{\mathrm{C}}$ are mapped with freedom to be 'shifted around' while preserving the homomorphism. Now, having 'coerced' the homomorphism in this manner may have produced a somewhat uneven map of T' over the reduced graph. These 'imbalances' are too big to be fixed with (C), so instead we apply a weighed version of the allocation algorithm to the remaining vertices of T which (combined with the homomorphism of T') yields an almost even map of T over the reduced graph, with much smaller imbalances. We conclude the allocation by modifying the mapping of the vertices in $\mathcal{P}_{\mathrm{C}}$ so make the map even.

We embed most vertices of T using a greedy algorithm, and complete this with perfect matchings. This algorithm is guaranteed to work if G is slightly larger than the tree we are embedding, so a preliminary step is to delete a few vertices from T, obtaining a tree T'' which is slightly smaller than G. Roughly speaking, the embedding algorithm processes each vertex of T'' in a tidy ancestral order, embedding at each step the current vertex t and all of its siblings according to their allocation. When this is done, appropriate sets are reserved for the children of the vertices just embedded. This algorithm works so long as there is room to spare. To conclude we reintegrate the removed vertices, arguing as in Chapter 3 to show that the required matchings exist. (The allocation of bare paths in $\mathcal{P}_{\mathrm{A}}$ and $\mathcal{P}_{\mathrm{B}}$ plays a crucial role here: the embedding is done somewhat differently as we approach vertices mapped to $V^\star$, and the edges allocated along the super-regular cycle are used for the final matchings.)

If T has many leaves—more precisely, if T has many vertex-disjoint edges incident to leaves—the proof proceeds very similarly, with leaf-edges playing the role of bare paths. In particular, the 'shifting' which we required in order to balance the allocation follows the same ideas of the 'cluster balancing algorithm' of Section 3.5.2 (see Figure 3.1 on page 63), with the difference that this shifting is not done along the cycle, but using a structure defined in Section 4.2.4.

4.2 Preliminaries

The following concepts and results play an important role in the proofs in this chapter.

4.2.1 Bare paths

Let T be a tree. A *path decomposition* $\mathcal{P}$ of a tree T is a collection of edge-disjoint subpaths of T such that each edge of T is contained in precisely one path of $\mathcal{P}$; we say that $\mathcal{P}$ is *bare* if each path of $\mathcal{P}$ is bare. In other words, two paths in a bare path decomposition are only allowed to intersect at their endvertices. We write $p(T)$ for the smallest size of a bare path-decomposition of T, and $\ell(T)$ for the number of leaves of T. In particular, if T is a path then $p(T) = 1$, and if T is a star then $p(T) = \ell(T)$.

path decomposition

bare path decomposition

Lemma 4.1. If T is a tree which is not a path, then $\ell(T) \leq p(T) \leq 2\ell(T) - 3$.

Proof. Let T be a tree which is not a path. Then each bare path of T contains at most 1 leaf of T (since bare paths intersect only at endvertices), and thus $p(T) \geq \ell(T)$. By replacing bare paths by edges, we obtain a tree T' with no vertex of degree 2. Note that each bare path of T' is a single edge, so $p(T) = e(T')$ and that $\ell(T) = \ell(T')$. So it suffices to show that $e(T')$, the number of edges of T', is at most $2\ell(T) - 3$. Consider the sum of degrees of the vertices in T', writing $e \coloneqq e(T'), \ell \coloneqq \ell(T')$ and $p \coloneqq p(T')$. Then $|T'| = e + 1$ and

$$\ell + 3\Big(p + 1 - \ell\Big) = \ell + 3\Big(e + 1 - \ell\Big) \leq \sum_{v \in T'} \deg_{T'}(v) = 2e = 2p,$$

so $p(T) = p(T') \leq 2\ell(T') - 3 = 2\ell(T) - 3$ as required. □

The bounds in Lemma 4.1 are sharp in the sense that there exist trees attaining each of the bounds. Stars match the lower bound; for the upper bound, suppose T' is a tree on $e + 1$ vertices and ℓ leaves in which every vertex has degree either 1 or 3; it follows that the displayed equation in the proof above holds with equality.

4.2.2 Regularity

The next lemma will be used in the proofs of Lemmas 4.20 and 4.23 to obtain the reduced graph required by the allocation algorithm (described in Section 4.3).

Lemma 4.2. Suppose that $1/n \ll 1/k \ll \varepsilon \ll d \ll \eta \ll \alpha$. If G is a digraph of order n with $\delta^0(G) \geq \left(\frac{1}{2}+\alpha\right)n$, then there exists a partition $V_0 \dot\cup V_1 \dot\cup \cdots \dot\cup V_k$ of $V(G)$ and a digraph $R^\star$ with $V(R^\star) = V_0 \dot\cup [k]$ such that

(a) $|V_0| < \varepsilon n$ and $m := |V_1| = \cdots = |V_k|$;

(b) The pairs (V_i, V_{i+1}) are (d,ε)-super-regular for each $i \in [k]$ (where $k+1 = 1$);

(c) For each $i \in [k]$ we have (V_{i-1}, V_i) and (V_i, V_{i+1}) are (d,ε)-super-regular;

(d) For all $i, j \in [k]$ we have $i \rightarrow j \in E(R^\star)$ precisely when (V_i, V_j) is (d,ε)-regular;

(e) For all $v \in V_0$ and all $i \in [k]$ we have $v \leftarrow i \in E(R^\star)$ precisely when $\deg^-(v, V_i) \geq (1/2+\eta)m$, and $v \rightarrow i \in E(R^\star)$ precisely when $\deg^+(v, V_i) \geq (1/2+\eta)m$;

(f) For all $i \in [k]$ we have $\deg^0_{R^\star}(i, [k]) \geq (1/2+\eta)k$; and

(g) For all $v \in V_0$ we have $\deg^0_{R^\star}(v, [k]) > \alpha k$.

Proof. We use a standard argument using regularity to establish (a) and (b) as well as

(i) For each $i \in [k]$ there exist $N^-, N^+ \subseteq [k]$ such that $|N^-|, |N^+| \geq \left(\frac{1}{2}+\eta\right)k$ and for all $j^- \in N^-, j^+ \in N^+$ we have that (V_i, V_{j^+}) and (V_{j^-}, V_i) are (d,ε)-regular;

(ii) For each $x \in V_0$ there exist $N^-, N^+ \subseteq [k]$ such that $|N^-|, |N^+| > \alpha k$ and for all $j^- \in N^-, j^+ \in N^+$ we have that $\left|N^-_G(x) \cap V_{j^-}\right| > |V_{j^-}|/2$ and $\left|N^+_G(x) \cap V_{j^+}\right| > |V_{j^+}|/2$;

Indeed, we first introduce constants ε', d' with $\varepsilon' \ll \varepsilon \ll d \ll d' \ll \alpha$, apply the digraph version of the Regularity Lemma (Lemma 2.9) to G and obtain a partition $V(G) = V'_0 \dot\cup V'_1 \dot\cup \cdots \dot\cup V'_k$ satisfying (a) with ε' in place of ε and also satisfying (b) with (ε', d') in place of (d,ε). For each $i \in [k]$ the set V_i contains at most $\varepsilon' n/k$ vertices which have less than $(d-\varepsilon')|V'_{i+1}|$ outneighbours in V'_{i+1} and at most $\varepsilon' n/k$ vertices which have less than $(d-\varepsilon')|V'_{i-1}|$ inneighbours in V'_{i-1}. By moving $\varepsilon' n/k$ vertices (including all vertices with atypical degrees) from each V'_i

to V_0' we can ensure (a) and (b) hold (as stated in the lemma), and that (i) holds as in the claim statement. Finally, (ii) follows by the rather large semidegree of G.

To conclude, let $R^\star$ be a digraph with vertex set $V_0 \dot\cup [k]$ and edges as follows. For all distinct $i, j \in [k]$ we have $i \rightarrow j \in E(R^\star)$ if (V_i, V_j) is (d, ε)-regular; for all $v \in V_0$ and all $i \in [k]$, we have $v \rightarrow i \in E(R^\star)$ if $\deg^+_G(v, V_i) \geq (1/2 + \eta)m$ and $v \leftarrow i \in E(R^\star)$ if $\deg^-_G(v, V_i) \geq (1/2 + \eta)m$. Then (d) and (e) hold. It is immediate from (i) and (ii) that for all $i \in [k]$ we have $\deg^0_{R^\star}(i, [k]) \geq (1/2 + \eta)k$ and for all $v \in V_0$ we have $\deg^0_{R^\star}(v, [k]) > \alpha k$, so (f) and (g) hold as well. □

Our embedding algorithm described in Section 4.3.4 relies on a key property of large dense digraphs, stated in the next lemma. The form in which it is stated here is a generalisation of [51, Lemma 2.5]. We begin with a definition.

Definition 4.3. Let $\beta, \gamma, m > 0$, let G and R be digraphs and let S be an oriented star with centre c. Also, Let φ be a homomorphism from S to R and let $\{ V_i : i \in R \}$ be a partition of $V(G)$. Finally, let $J^- := \varphi\big(N^-_S(c)\big)$ and let $J^+ := \varphi\big(N^+_S(c)\big)$, so $|J^-| + |J^+| = \Delta(\varphi)$. A subset $V \subseteq V_{\varphi(c)}$ is $(\beta, \gamma, \varphi, m)$-*good* for S if for every collection $(\beta, \gamma, \varphi, m)$-good

$$\mathcal{V} = \{ V_j^- \subseteq V_j : j \in J^- \} \cup \{ V_j^+ \subseteq V_j : j \in J^+ \}$$

of sets of size at least βm there exists $V' \subseteq V$ such that $|V'| \geq \gamma m^{1/\Delta(\varphi)}$ and such that every vertex $v \in V'$ satisfies the following

- for all $j \in J^-$ we have $\deg^-(v, V_j^-) \geq \gamma m$, and
- for all $j \in J^+$ we have $\deg^+(v, V_j^+) \geq \gamma m$.

Here is some motivation for Definition 4.3. We will embed the vertices of the tree one by one, and, after embedding a vertex $x \in T$, we reserve sets of vertices for the children of x. If the reserved sets are always good, then this greedy embedding strategy will succeed. The next lemma states one sufficient condition (on $\beta, \gamma, G, R, m, S$ and φ) for this to occur.

Lemma 4.4. Suppose $\frac{1}{m} \ll \varepsilon \ll \gamma \ll 1/q, \beta, d$, let G and R be digraphs and let S be an oriented star with centre c. Let φ be a homomorphism from S to R and let $\{ V_j : j \in R \}$ be a partition of $V(G)$ into sets of size m. If $\Delta(\varphi) = q$ and for each edge $i \rightarrow j \in E(S)$ the pair $(V_{\varphi(i)}, V_{\varphi(j)})$ is (d, ε)-regular, then every subset $V^c \subseteq V_{\varphi(c)}$ such that $|V^c| = \gamma m/2$ contains a subset V of order at most $m^{1-1/\Delta(\varphi)}$ which is $(\beta, \gamma, \varphi, m)$-good for S.

Proof. Note that the definition of $(\beta, \gamma, \varphi, m)$-good is not affected whether the number of in-leaves (respectively out-leaves) of S to a fixed $x \in R$ is precisely 1 or another positive integer. Hence, by removing leaves if necessary, we may assume that S has no two distinct in-leaves x, y with $\varphi(x) = \varphi(y)$, as this cannot weaken the statement we are trying to prove; moreover, this implies that S has precisely q leaves.

Let $x_1, \ldots, x_q$ be an enumeration of the leaves of S, starting with the in-leaves. For each $i \in [q]$, let $S_i := S\Big[\{c, x_1, \ldots, x_i\}\Big]$, so $S_q = S$. Fix $V^c \subseteq V_{\varphi(c)}$. If $t \in [q]$ and $\bar{t} = (v_1, \ldots, v_t)$ is a t-tuple of distinct vertices $v_j \in V_j \subseteq G$ for each $j \in [t]$, and if $V \subseteq V^c$, we write $N^{S_t}(\bar{t}, V)$ for the set of $v \in V$ such that for each $j \in [t]$ the map $(x_j, c) \mapsto (v_j, v)$ is a homomorphism from the edge between x_j and c, which lies in S, to an edge between v_j and v, which lies in G; we call $N^{S_t}(\bar{t}, X)$ the S_t-*neighbourhood* of $\bar{t}$ in V. Finally, let T be the set of tuples $(v_1, \ldots, v_q)$ with $v_i \in V_i$ for each $i \in [q]$; call $\bar{t} \in T$ *bad* if $\Big|N^S(\bar{t}, V^c)\Big| < |V^c|(d/3)^q$ and *good* otherwise.

S_t-neighbourhood

bad, good

Let $V^0 := V^c$. For each $i \in [q]$, in ascending order, we proceed as follows. Suppose that $x_i \in N^-_S(c)$. Note that $|V^{i-1}| \geq |V^0|(3/d)^{i-1}$, and therefore $(V_{\varphi(x_i)}, V^{i-1})$ is $(\varepsilon'(3/d)^{i-1}, d/2)$-regular. So at most $\varepsilon' m(3/d)^{i-1}$ vertices $v_i \in V_{\varphi(x_i)}$ have less than $|V^{i-1}|(d/2 - \varepsilon'(3/d)^{i-1}) \geq |V^{i-1}|d/3$ outneighbours in V^{i-1}. Let

$$B_i := \Big\{\, v_i \in V_{\varphi(x_i)} : \deg^+(v_i, V^{i-1}) < |V^{i-1}|d/3 \,\Big\},$$

fix $v_i \in V_{\varphi(x_i)} \setminus B_i$ and let

$$V^i := N^{S_i}\Big((v_1, \ldots, v_i), V^{i-1}\Big),$$

so $|V^i| \geq |V^{i-1}|d/3 \geq |V'|(d/3)^i$. We proceed similarly if $x_i \in N^+_S(c)$.

Let $B := \{\, t \in T : \bar{t} \text{ bad} \,\}$. By the construction above, each bad t contains at least one vertex from $B_1 \cup \cdots \cup B_q$ and thus

$$|B| \leq m^{q-1} \sum |B_i| \leq 2\varepsilon m^q (3/d)^{q+1}/\gamma. \tag{4.1}$$

Claim 4.5. There exists $V \subseteq V^c$ such that $|V| \leq m^{1-1/q}$ and such that for at most $|B|$ tuples $\bar{t} \in T$ we have $\Big|N^{S_q}(\bar{t}, V)\Big| < m^{1-1/q}(d/3)^q/4$.

Proof. Let $\varepsilon' := 2\varepsilon/\gamma$. Choose $V \in V^c$ at random by including each $v \in V^c$ with probability $p = 1/\gamma m^{1/q}$ independently of all other vertices. By (2.2), with probability $1 - o(1)$ we have that $|V| < 2p|V^c| \leq m^{1-1/q}$. Let us call this event E_1.

If $\bar{t} \in T$ is good, then the probability that

$$\Big|N^S(\bar{t}, V)\Big| < p|V^c|(d/3)^q/2 = m^{1-1/q}(d/3)^q/4$$

decreases exponentially with m by (2.1), whereas $|T \setminus B| = \mathrm{O}(m^q)$. Thus, by a union bound, with probability $1 - \mathrm{o}(1)$ the randomly selected set V has the property that at most $|B|$ tuples $\bar{t} \in T$ are such that $\left|N^S(\bar{t}, V^c)\right| < m^{1-1/q}(d/3)^q/4$. Call this event E_2. We may therefore fix a choice of $V \subseteq V^c$ such that both E_1 and E_2 hold. $\diamond$

Fix V as in the claim above. It remains to show that V is $(\beta, \gamma, \varphi, m)$-good for S. Indeed, let $\mathcal{V}' = \{\, V_i' \subseteq V_{\varphi(i)} : i \in [q] \,\}$ be a collection of sets such that $|V_i'| = \beta m$ for each $i \in [q]$, and let T' be the set of tuples $(v_1, \ldots, v_q)$ with $v_i \in V_i'$ for each $i \in [q]$. Then by (4.1) there exist at least

$$\beta^q m^q - 2\varepsilon m^q (3/d)^{q+1}/\gamma \geq (\beta m)^q/2$$

tuples of vertices $\bar{t} = (v_1, \ldots, v_q) \in T'$ such that $\left|N^{S_q}(\bar{t}, V)\right| \geq m^{1-1/q}(d/3)^q/4$. So there exist at least

$$\Big((\beta m)^q/2\Big)\Big(m^{1-1/q}(d/3)^q/4\Big) = (\beta d/3)^q m^{q+1-1/q}/8$$

pairs $(v, \bar{t})$ where $\bar{t} \in T'$, $v \in V$ and $v \in N^S(\bar{t}, V)$. In particular, at least $\Big((\beta d/3)^q m^{q+1-1/q}/8\Big)/2\left|T'\right| \geq \gamma m^{1/q}$ vertices $v^\star \in W$ must lie in the neighbourhood of at least $(\beta d/3)^q m^q/16$ tuples $\bar{t} \in T'$ — otherwise there would be fewer than

$$|T'| \cdot \Big((\beta d/3)^q m^{q+1-1/q}/8\Big)/2\left|T'\right| + |V|(\beta d/3)^q m^q/16 \leq (\beta d/3)^q m^{q+1-1/q}/8$$

such pairs $(v, \bar{t})$. So each of these vertices $v^\star$ has at least $(\beta d/3)^q m^q/16 m^{q-1} \geq \gamma m$ neighbours in each $V_i' \in \mathcal{V}'$, as required. $\square$

4.2.3 Matchings

The next simple lemma is used in the proofs of Lemmas 4.8, 4.18 and 4.22.

Lemma 4.6. Let G be a bipartite graph with vertex classes V and W, and suppose every vertex in V has degree at least $\varepsilon|W|$. Then there exists a spanning subgraph $H \subseteq G$ such that

(i) $\deg_H(v) = 1$ for each $v \in V$,

(ii) $\deg_H(w) \leq 1 + \frac{|V|}{\varepsilon|W|}$ for each $w \in W$.

Proof. Let $n := |V|$. We build H iteratively, starting with an empty graph with vertex set $V \cup W$, and proceeding greedily through the vertices $v_1, \ldots, v_n$ of V, as follows. For each $i \in [n]$, let $w_i \in W$ be a neighbour of v_i in G with minimum degree in the graph $H_i := (V \cup W, \bigcup_{j\in[i-1]}\{v_i w_i\})$, and let $H := H_n$. Clearly every $v \in V$ has degree 1 in H, by construction. Moreover, if $w \in W$ is connected to a vertex in V and j is the largest index such that $v_j w \in E(H)$, then, by construction, every vertex in $W_j := N_G(v_j)$ has degree at least $\deg_{H_j}(w) - 1 = deg_H(w) - 1$ in H_j; since $|W_j| \geq \varepsilon|W|$ we have that

$$\varepsilon|W|(\deg_H(w) - 1) \leq \sum_{w'\in W_j} \deg_{H_{j-1}}(w') \leq \sum_{w'\in W_j} \deg_H(w') \leq |V|,$$

and it follows that $\deg_H(w) \leq 1 + |V|/\varepsilon|W|$; this completes the proof since the choice of w is arbitrary. □

Fact 4.7. [12, Exercise 16.1.6] Let M and N be edge-disjoint matchings of a graph G. If $|M| > |N|$, then there exist disjoint matchings M' and N' of G such that $|M'| = |M| - 1$, $|N'| = |N| + 1$ and $M' \cup N' = M \cup N$.

Proof. Note that the components $C_1, \dots, C_s$ of the subgraph of G formed by the edges in $M \cup N$ are either single edges or cycles; moreover, since $|M| > |N|$, there must be some component C_i such that

$$\Big|E(C_i) \cap M\Big| = \Big|E(C_i) \cap N\Big| + 1.$$

To obtain the desired matchings, it suffices to 'swap' the edges in C_i, i.e.: define:

$$M' := M \cup (N \cap C_i) \setminus (M \cap C_i) \quad \text{and} \quad N' := N \cup (M \cap C_i) \setminus (N \cap C_i). \ \square$$

4.2.4 Diamond-paths

The definitions of this Section play an important role in the proof of Lemma 4.18; they will be used to ensure that our vertex allocation covers exceptional vertices and also that many identically oriented bare paths are allocated along a Hamilton cycle in the reduced graph.

Let T be an oriented tree and $\prec$ be an ancestral order of T. Note that $\prec$ induces a unique ancestral order on each (oriented) subtree T' of T, namely, the restriction of $\prec$ to the vertices of T'. We shall also write $\prec$ to refer to the orders induced by $\prec$ on each of these subtrees. We say that the oriented trees T_1, T_2

$\prec$-isomorphic with ancestral orders $\prec_1$ and $\prec_2$ respectively are *$\prec$-isomorphic* if there exists an isomorphism $\rho : V(T_1) \to V(T_2)$ which preserves order.

Let $\prec$ be an ancestral order of (an oriented tree) T and suppose that P and Q are paths of T of order 7, each rooted at a leaf, labelled so that $V(P) = \{\, p_i : i \in [7]\,\}$, $V(Q) = \{\, q_i : i \in [7]\,\}$ and that $p_1 \prec p_2 \prec \ldots \prec p_7$ and $q_1 \prec q_2 \prec \ldots \prec q_7$

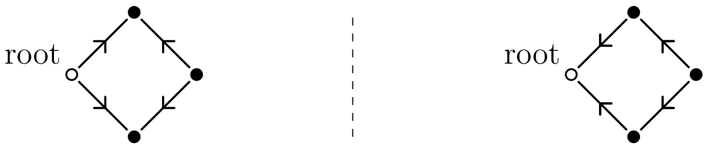


Figure 4.1: Left: a ($\circ\!\to\!\bullet\!\leftarrow\!\bullet$)-diamond. Right: a ($\circ\!\leftarrow\!\bullet\!\leftarrow\!\bullet$)-diamond ($\circ$ denotes the root of the path).

(so p_1 is the root of P, q_1 is the root of Q, and the edges of these paths connect vertices with indexes differing by 1). The *prefix section* prefix(P) of P is the (rooted) path induced by its 3 first vertices p_1, p_2, p_3; the *middle section* middle(P) of P is the path induced by its 'middle' vertices p_3, p_4, p_5; and the *suffix section* suffix(P) of P is the path induced by p_5, p_6, p_7—so, the prefix, middle and suffix sections of P are edge-disjoint rooted subpaths of P of order 3 each, whose union is P. We say that P and Q have the *same middle section* if their middle sections are $\prec$-isomorphic. In other words, P and Q have the same middle section if mapping $p_3 \mapsto q_3, p_4 \mapsto q_4$ and $p_5 \mapsto q_5$ is an isomorphism—note that this definition would make little (if any) sense if P and Q were not oriented paths.

prefix section
middle section
suffix section
same middle section

Note that these definitions depend on which ancestral order is considered, but this order will always be clear from context. The *centre* $c(P)$ of a path P of odd order is the vertex $v \in P$ which is equidistant to the two leaves of P.

centre

Let P be an oriented path of order 3. A *P-diamond* is the digraph formed from P by blowing up its middle vertex into 2 vertices. So if P is $a{\to}b{\to}c$ then the digraph H with $V(H) = \{u, v, v', w\}$ and $E(H) = \{u{\to}v, u{\to}v', v{\to}w, v'{\to}w\}$ is a P-diamond (see Figure 4.1). The paths uvw and $uv'w$ are the *branches* of the diamond. If $\prec$ is an ancestral order of P, say with $a \prec b \prec c$, then we say that the P-diamond H has *prefix* u, *middle* $\{v, v'\}$ and *suffix* w, and we shall denote the (rooted) P-diamond by $u{\preceq}^{v}_{v'}{\succeq}w$. If P and $\prec$ are clear from context, we write *diamond* instead of P-diamond. A *P-diamond path* in a digraph D is a sequence of P-diamonds $\left(u_i{\preceq}^{v_i}_{v'_i}{\succeq}w_i\right)_{i=0}^{t}$ such that $v_i = v'_{i-1}$ for each $i \in [t]$; we say that this path *connects* v_0 and v'_t. Finally, a graph G is *P-connected* if there exists a P-diamond path connecting each pair $u, v \in G$.

P-diamond
branches
prefix, middle, suffix
diamond, P-diamond path
connected (by diamond path), P-connected

Lemma 4.8 (P-connected subgraphs). Suppose $\frac{1}{n} \ll \alpha$. Let D be a digraph with of order n such that $\delta^0(D) \geq \left(\frac{1}{2} + \alpha\right)n$. If P is a rooted oriented path of order 3, then D contains a spanning P-connected subdigraph H which is the union of $n-1$ diamonds and such that $\Delta^0(H) \leq 4/\alpha$.

Proof. We may assume without loss of generality that $V(D) = [n]$. Let $i \in [n-1]$; we write $\Diamond_i$ for the set of all P-diamonds with middle $\{i, i+1\}$. Let B_{pref} be a bipartite graph with vertex classes $\Diamond := \{\Diamond_1, \dots \Diamond_{n-1}\}$ and $[n]$, with an edge between $\Diamond_i$ and $x \in [n]$ if x is a prefix of a P-diamond in $\Diamond_i$. Note that $\deg\big(\Diamond_i, [n]\big) \geq \alpha n$. Therefore, by Lemma 4.6, there exists $H_{\text{pref}} \subseteq B_{\text{pref}}$ such that each vertex of $\Diamond$ is covered by precisely one edge of H_{pref} and each vertex of $[n]$ is covered by at most $1/\alpha$ edges of H_{pref}. We define B_{suff} similarly for suffixes and obtain the corresponding graph H_{suff}. We define the spanning subgraph $H \subseteq D$ as follows. For each $i \in [n-1]$, let p_i be the neighbour of $\Diamond_i$ in H_{pref} and let s_i be the neighbour of $\Diamond_i$ in H_{suff}. Then $p_i{\preceq}{}^{i}_{i+1}{\succeq}s_i$ is a P-diamond. Let $E(H)$ be the union of the edges in the gadgets $p_i{\preceq}{}^{i}_{i+1}{\succeq}s_i$. Note that H is P-connected. Moreover, the maximum underlying degree of $x \in H$ is $4/\alpha$ (because each edge of H_{pref} and H_{suff} corresponds to 2 edges of D). $\square$

The next lemma explains the importance of P-connectedness (see Figure 4.2). Let T be a rooted oriented tree with many induced paths $\prec$-isomorphic to P, and let D be a digraph containing a spanning P-connected subgraph which is the union of P-diamonds $\Big(x_i{\preceq}{}^{y_i}_{w_i}{\succeq}z_i\Big)_{i=1}^{t}$. If a homomorphism $\varphi : T \to D$ maps a linear number of copies of P to each of the paths $x_i y_i z_i$ and $x_i w_i z_i$ rooted at x_i, then we can transform φ into a homomorphism ϱ where the number of vertices mapped to each vertex of D is slightly different. This will be an important tool when adjusting the allocation.

Lemma 4.9. Suppose that $\frac{1}{n} \ll \frac{1}{k} \ll \eta \ll \lambda < 1$. Let P be a rooted path of order 3; let T be a rooted oriented tree of order n which contains a collection $\mathcal{P}$ of λn induced subgraphs isomorphic to P; let R be a digraph of order k, and let H be a P-connected spanning subgraph of R. Finally, for each $v \in R$, let δ_v be an integer, with $|\delta_v| < \frac{n}{\log n}$ and such that $\sum_{v \in R} \delta_v = 0$. If there exists a homomorphism $\varphi : T \to R$ such that for each diamond $x_i{\preceq}{}^{y_i}_{w_i}{\succeq}z_i$ in H there are at least $\eta n/k^3$ paths in $\mathcal{P}$ which are mapped to $x_i y_i z_i$ and at least $\eta n/k^3$ paths which are mapped to $x_i w_i z_i$, then there exists a homomorphism $\varrho : T \to R$ such that $|\varrho^{-1}(v)| = |\varphi^{-1}(v)| + \delta_v$ for all $v \in R$.

Proof. We proceed greedily, as follows. Let $u, v \in R$ be such that $\delta_v < 0 < \delta_u$, and consider the P-diamond path from u to v. Let $\Big(x_i{\preceq}{}^{y_i}_{w_i}{\succeq}z_i\Big)_{i=1}^{t}$ be the sequence of diamonds in this path, so $u = y_1$ and $v = w_t$. For each $i \in [t]$, select a path in T which is mapped to $x_i y_i w_i$, and modify the mapping of this path so that it is now mapped to $x_i w_i z_i$ (see Figure 4.2). The resulting mapping ϱ' is such that $|\varrho^{-1}(u)| = |\varphi^{-1}(u)| - 1$ and $|\varrho^{-1}(v)| = |\varphi^{-1}(v)| + 1$, whereas $|\varrho^{-1}(x)| = |\varphi^{-1}(x)|$

for all $x \in R \setminus \{u, v\}$. Note that this procedure reduces by at most 1 the number of paths mapped to each diamond branch. Hence, by iterating this procedure at most $\sum_v |\delta_v| \leq kn/\log n$ times, we can 'shift weights' as needed to obtain the desired mapping ϱ. Note that it is indeed possible to carry out these steps, because each diamond has at least $\eta n/k^3$ paths allocated to each of its branches (and each step changes the size of preimage of any $x \in R$ by at most 1). $\square$

We remark that the quantity $\eta n/k^3$ in the previous lemma is much larger than necessary—the proof works with no modifications if this is replaced by $kn/\log n$—but this version will suffice for our purposes.

4.3 An approximate result

In this section we discuss the allocation and embedding algorithms which lie at the core of our proofs. The main results of this section are Lemmas 4.15, 4.16 and Theorem 4.17 (see below).

Lemma 4.15 states that the randomised allocation algorithm (Algorithm 4.14) allocates roughly the same number of vertices to each cluster in the reduced graph. This algorithm will be applied to regular graphs with good expansion properties. To prove that the algorithm works, we begin by introducing regular expanding digraphs and proving that every (sufficiently large) digraph of high semidegree contains one such subgraph. We then show that a particular kind of random walk in these digraphs (according to a non-homogeneous Markov chain) converges quickly to a uniform stationary distribution. Finally, we deduce some properties of the randomised algorithm.

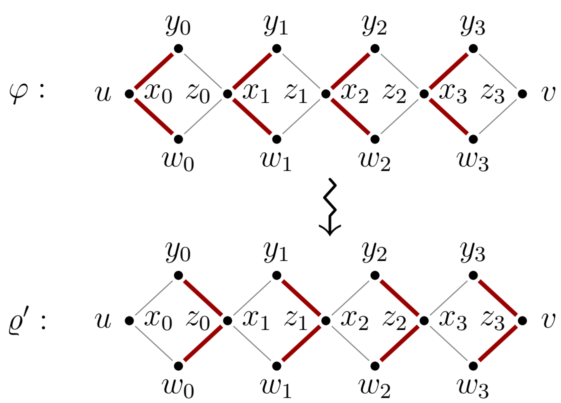


Figure 4.2: Weight-shifting from vertex u to vertex v using P-diamond.

Lemma 4.16 then states that if G is a digraph with an appropriate reduced graph R, and if we are given an appropriate allocation (such as the one guaranteed by Lemma 4.15) of a tree T to R, then the embedding algorithm successfully finds a copy of T in G.

Finally, Theorem 4.17 combines these lemmas, showing that if T is a tree with maximum degree bounded by a polylogarithmic function of $|T|$, and if G is a digraph with high semidegree and order slightly grater than $|T|$, then G contains a copy of T.

4.3.1 Existence of regular expander subdigraph

expander We begin with a definition. A digraph D is an *expander* if

$$\left|N^-(S)\right|, \left|N^+(S)\right| > |S| \qquad \text{for all nonempty proper } S \subseteq V(D).$$

Lemma 4.10. Suppose that $1/n \ll 1/f, \alpha$. Let G be a digraph of order n with $\delta^0(G) \geq (\frac{1}{2}+\alpha)n$. If $F \subseteq G$ is a subgraph of G with $\Delta^0(F) \leq f$, then G contains a spanning d-regular subdigraph H such that

(i) H contains F,

(ii) $d \leq 25n^{2/3}/\alpha$, and

(iii) H is an expander.

Proof. We first form a spanning subgraph $H_p \subseteq G$ by keeping each edge of G with probability $p := n^{-1/3}$ independently of all other edges.

Claim 4.11. With probability $1 - o(1)$

(a) every vertex of H_p has in- and outdegree at most $4n^{2/3}$, and

(b) if S is a nonempty proper subset of $V(G)$, then $\left|N^-_{H_p}(S)\right|, \left|N^+_{H_p}(S)\right| > |S|$.

Proof of claim. Let $x \in H_p$. Note that $\deg^-_{H_p}(x)$ and $\deg^+_{H_p}(x)$ are binomial random variables with expectation between $(\frac{1}{2}+\alpha)n^{2/3}$ and $n^{2/3}$. By Chernoff (2.1) (applied with $t = n^{2/3}/2$) we have

$$\begin{aligned} &\mathbb{P}\Big(\deg^-_{H_p}(x) > 4n^{2/3}\Big) \leq \exp\Big(-n^{2/3}\Big) \quad \text{and} \\ &\mathbb{P}\Big(\deg^+_{H_p}(x) > 4n^{2/3}\Big) \leq \exp\Big(-n^{2/3}\Big). \end{aligned} \tag{4.2}$$

By a union bound over all n vertices, we have that (a) holds with probability $1 - o(1)$. To prove (b), let $S \subseteq H_p$ be proper and nonempty. We will show that $|N^+_{H_p}(S)| > |S|$; the argument showing that $|N^-_{H_p}(S)| > |S|$ is symmetric. We consider four cases. If $|S| < n^{1/2}$, then for each vertex $x \in S$ the expected outdegree of x is a binomial random variable with expectation at least $(\frac{1}{2}+\alpha)n^{2/3}$, and by Chernoff (2.2) (applied with $t = np/2$) we have, for any $y \in S$, that

$$\mathbb{P}\Big(|N^+_{H_p}(S)| \le |S|\Big) \le \mathbb{P}\left(\deg^+_{H_p}(y) < \frac{np}{2}\right) \le \exp\Big(-\frac{n^{2/3}}{8}\Big), \tag{4.3}$$

If $n^{1/2} \le |S| < n/2$, then $\Big|N^+_{H_p}(S)\Big| \le |S|$ if and only if there exists $T \subseteq V(H_p)$ such that $|T| \ge n - |S|$ and $N^+_{H_p}(S) \cap T = \varnothing$. Since $|S| < n/2$, we have $|T| \ge n/2$ and thus each vertex $x \in S$ has at least αn outneighbours in T. In particular, the number $e_{H_p}(S,T)$ of edges from S to T in H_p is a binomial random variable with expectation at least $|S|\alpha n^{2/3} \ge \alpha n^{7/6}$. By Chernoff (2.2), applied with $t = \alpha n^{7/6}/2$, we have $\mathbb{P}\Big(e_{H_p}(S,T) = 0\Big) \le \exp\Big(-\alpha n^{7/6}/8\Big)$. So if B_T is the event '$e_{H_p}(S,T) = 0$' then

$$\begin{aligned}\mathbb{P}\Big(\Big|N^+_{H_p}(S)\Big| \le |S|\Big) &= \mathbb{P}\Big(\bigcup_{|T|\ge n-|S|} B_T\Big)\\ &\le 2^n \exp\Big(-\frac{\alpha n^{7/6}}{8}\Big) \le \exp\Big(-\frac{\alpha n^{7/6}}{10}\Big).\end{aligned} \tag{4.4}$$

If $n/2 \le |S| < n - n^{1/2}$, then, as before, $\Big|N^+_{H_p}(S)\Big| \le |S|$ if and only if there exists $T \subseteq V(H_p)$ such that $|T| \ge n - |S| \ge n^{1/2}$ and $N^+_{H_p}(S) \cap T = \varnothing$. Since $|S| \ge n/2$ for all $x \in G$ we have $\deg^-_G(x,S) \ge \alpha n$. In particular, the number $e_{H_p}(S,T)$ of edges from S to T is a binomial random variable with expectation at least $|T|\alpha n^{2/3} \ge \alpha n^{7/6}$, and thus (4.4) holds. Finally, if $|S| \ge n - n^{1/2}$, then for all $x \in H_p$ we have $\deg^-_G(x,S) \ge n/2$, and thus $\deg^-_{H_p}(x,S)$ is a binomial random variable with expectation at least $n^{2/3}/2$. By Chernoff (2.2)

$$\mathbb{P}\Big(N^+_{H_p}(S) \ne V(H_p)\Big) \le \sum_{x\in H_p} \mathbb{P}\Big(\deg^-_{H_p}(x) \le n^{1/2}\Big) \le \exp\Big(-\frac{n^{2/3}}{40}\Big). \tag{4.5}$$

For all proper and nonempty $S \in V(H_p)$, let $P_S := \mathbb{P}\Big(|N^+_{H_p}(S)| \le |S|\Big)$. By a union bound, we obtain

$$\begin{aligned}\mathbb{P}\Big(\text{(b) does not hold}\Big) &= \sum_{0<|S|<\sqrt{n}} P_S + \sum_{\sqrt{n}\le|S|<n-\sqrt{n}} P_S + \sum_{n-\sqrt{n}\le|S|<n} P_S\\ &\le 3\binom{n}{n^{1/2}} \underbrace{\exp\left(-\frac{n^{2/3}}{40}\right)}_{(4.3)\text{ and }(4.5)} + \underbrace{2^n \exp\left(-\frac{\alpha n^{7/6}}{10}\right)}_{(4.4)} = o(1),\end{aligned}$$

where the sums are over proper and nonempty subsets $S \subseteq H_p$ and we use the bound $\binom{n}{\sqrt{n}} \le (\sqrt{n}\, e)^{\sqrt{n}} \le \exp\Big(\sqrt{n}\ln n\Big)$. Therefore (b) holds with probability $1 - o(1)$, as desired. ◇

We now come back to the proof of the lemma. By the claim above, we can fix an outcome H_p of the random choices such that (a) and (b) both hold. Let $H' \coloneqq H_p \cup F$. We will build a regular digraph H which contains H' and satisfies (ii). Note that H will satisfy (iii) as well, by (b). Since $\Delta^0(H') \le 4n^{2/3} + f \le 5n^{2/3}$, there exists a proper equitable edge colouring of H' using at most $5n^{2/3}+1$ colours (by Lemma 4.7). In other words, there exists a partition $\mathcal{M}$ of $E(H')$ into at most $5n^{2/3}+1$ matchings with sizes as equal as possible; in particular, since $\left|E(H')\right| \le 5n^{5/3}$, each matching in this partition contains at most $5n^{5/3} \le n$ edges. We choose an equitable partition $M_1, \dots, M_d$ of the edges of H' (refining $\mathcal{M}$) so that for each $i \in [d]$ the matching M_i has at most $\alpha n/6$ edges and $d \le 32n^{2/3}/\alpha$.

Let $G_0 \coloneqq G - E(H')$. For each $i \in [d]$, we proceed as follows. Greedily choose a cycle C_i in $G_{i-1} \cup M_i$, such that C_i has length $3|M_i|$ and covers all edges in M_i; let C'_i be a Hamilton cycle in $G_{i-1} \setminus V(C_i)$ and let $G_i \coloneqq G_{i-1} \setminus \big(E(C_i) \cup E(C'_i)\big)$. We set $H \coloneqq \bigcup_{i\in[d]} (C_i \cup C'_i)$. Note that $H' \subseteq H$ and that H is the union of spanning regular subdigraphs $C_i \cup C'_i$ of G, so H is a spanning regular subgraph of G. Furthermore, for each vertex $x \in H$ we have $\deg^-_H(x) = \deg^+_H(x) = d \le 32n^{2/3}/\alpha$.

We need to argue that it is indeed possible to carry out the steps above. It suffices to show that C_i and C'_i exist. Let $i \in [d]$, let $r \coloneqq |M_i| \le \alpha n/6$ and let $u_j{\to}v_j$ be the edges in M_i for each $j \in [r]$. Note that

$$\delta^0(G_i) \ge \delta^0(G_0) - d \ge (1/2 + 3\alpha/4)n,$$

so, for each pair of vertices $x, y \in G_i$ there exists at least $3\alpha n/4$ vertices z such that $z \in N^+_{G_i}(x) \cap N^-_{G_i}(y)$. We can therefore choose distinct vertices $z_j \in N^+_{G_i}(v_j) \cap N^-_{G_i}(u_{j+1})$ (addition modulo r), which form the cycle C_i: $u_1v_1z_1u_2, \dots u_rv_rz_ru_1$. Since $|C_i| < 3|M_i| \le \alpha n/2$, it follows that $\delta^0\big(G_i \setminus V(C_i)\big) \ge (1/2 + \alpha/4)n$ and thus $G_i \setminus V(C_i)$ contains a directed Hamilton cycle C'_i. □

4.3.2 Random walks

Let D be a digraph, let P be an oriented path rooted in one of its leaves, and let $v_0, v_1, \dots, v_r$ be an ancestral order of P. For each $v \in D$, a *random P-walk* $\mathcal{W}_{P,D}(v)$ on D, starting at v, is a random walk $X_0, X_1, \dots, X_r$ starting at $X_0 = v$ and such that for each $i \in [r]$ we choose X_{i+1} as follows: if v_{i+1} is an outneighbour of v_i in P, then choose X_{i+1} uniformly at random from $N^+_D(X_i)$; otherwise choose X_{i+1} uniformly at random from $N^-_D(X_i)$, with all choices made independently for all i. (This random walk is a non-homogeneous Markov chain.) The following lemma will be used to establish a crucial property of random P-walks.

random P-walk

Lemma 4.12. Let D be an expander digraph of order $n \geq 3$, let $f : V(D) \to [0,1]$ and let $M := \max_{x,y \in D} f(x) - f(y)$. If $M > 0$, then there exists $u, x, y \in D$ such that $x, y \in N^-(u)$ and $f(y) - f(x) \geq M/(n-1)$ and, similarly there exists $v, w, z \in D$ such that $w, z \in N^+(v)$ and $f(w) - f(z) \geq M/(n-1)$.

Proof. We prove only the existence result for u, x, y, the statement for v, w, z follows by symmetry. Let $S_1, \ldots, S_r$ be a partition of $V(D)$ such that for all $x, y \in D$ we have $f(x) = f(y)$ if and only if $x, y \in S_i$ for some $i \in [r]$. Clearly, $1 < r \leq n$. Since f is a constant in each set of this partition, we write $f(i)$ for the common value of f over all $x \in S_i$. We can assume that the sets are labelled so that $f(i) < f(j)$ whenever $i < j$. Note that $M = f(r) - f(1)$, and therefore $f(j+1) - f(j) \geq M/(r-1) \geq M/(n-1)$ for some $j \in [r-1]$. Let $X := S_1 \cup \cdots \cup S_j$ and let $Y := S_{j+1} \cup \cdots \cup S_n$. Since D is an expander, we have that $\left|N^+(X)\right| > |X|$ and $\left|N^+(Y)\right| > |Y|$. Because $|X| + |Y| = n$ there must be a vertex $u \in N^+(X) \cap N^+(Y)$. Let $x \in X$ and $y \in Y$ be inneighbours of u. Then $f(y) - f(x) \geq f(j+1) - f(j) \geq M/(n-1)$ as desired. $\square$

Lemma 4.13. Let D be an d-regular expander digraph of order k. Also, let P be an oriented path of order n, rooted in one of its leaves. Then for all $v \in V(D)$, if $X_0, \ldots, X_n$ is a random P-walk $\mathcal{W}_{P,D}(v)$, then

$$\max_{x \in D} \left(\mathbb{P}(X_n = x) - \frac{1}{k}\right)^2 \leq \left(1 - \frac{1}{2k^3}\right)^n. \tag{4.6}$$

In particular, for all $x \in V(D)$ we have that $\mathbb{P}(X_n = x) \to 1/k$ as $n \to \infty$.

Proof. For a random variable X with range D, define

$$m(X) := \sum_{x \in D} \left(\mathbb{P}(X = x) - \frac{1}{k}\right)^2,$$

so $m(X) = 0$ if and only if $\mathbb{P}(X = x) = 1/k$ for all $x \in D$. We first prove that

$$m(X) \leq (1 - 1/k)^2 + (k-1)/k^2 = 1 - k^{-1} < 1. \tag{4.7}$$

Proof of (4.7). Note that $1 - 1/k$ is the value $m(\cdot)$ attains at a random variable which concentrates all 'weight' in a single vertex of D. Let X be a random variable which maximises $m(X)$ over all random variables with range D and suppose, looking for a contradiction, that there exist distinct u and v in D with 'positive weight', i.e., such that $\mathbb{P}(X = u), \mathbb{P}(X = v) > 0$; we may assume, by averaging, that $0 < \mathbb{P}(X = u) \leq 1/k$ and $1/k \leq \mathbb{P}(X = v) < 1$. Now consider the random variable Y defined as

$$Y = \begin{cases} X & \text{if } X \neq u \\ v & \text{otherwise} \end{cases}$$

then $m(Y) > m(X) + 2\Big(\mathbb{P}(X = u) - 1/k\Big)^2 > m(X)$, contradicting the choice of X. $\diamond$

We now return to the proof of the lemma. Let μ be a probability distribution over D, let $v \in D$ be selected according to μ, and let $X_0, \ldots, X_n$ be a random P-walk $\mathcal{W}_{P,D}(v)$. We will show that for all $i \in [n]$

$$m(X_{i-1}) \geq m(X_i),$$

with strict inequality if $m(X_{i-1}) > 0$. Let $i \in [n]$, suppose that v_{i-1} is an inneighbour of v_i in P and define $f(y) \coloneqq \mathbb{P}(X_{i-1} = y) - 1/k$ for all $y \in D$. Then

$$\begin{aligned} m(X_i) &= \sum_{x \in D} \left(\mathbb{P}(X_i = x) - \frac{1}{k}\right)^2 = \sum_{x \in D} \left(\left(\sum_{y \in N_D^-(x)} \frac{\mathbb{P}(X_{i-1} = y)}{d} \right) - \frac{1}{k} \right)^2 \\ &= \sum_{x \in D} \left(\sum_{y \in N_D^-(x)} \frac{f(y)}{d} \right)^2 \overset{(\natural)}{\leq} \sum_{\substack{x \in D \\ y \in N_D^-(x)}} d \left(\frac{f(y)}{d} \right)^2 \overset{(\sharp)}{=} \sum_{x \in D} f(x)^2 = m(X_{i-1}) \end{aligned}$$

Where ($\natural$) follows by Theorem 2.16 (with $u_i = 1/d$ and $v_i = f(y)$) because $|N_D^-(x)| = d$ for all $x \in D$ and ($\sharp$) because D is d-regular (so the term $f(y)$ appears precisely d times for each $y \in D$). In particular, it also follows by Theorem 2.16 that equality holds in ($\natural$) if and only if there exists $\alpha \neq 0$ such that $f(y) = \alpha/d$ for all i—that is, if and only if X_{i-1} is uniformly distributed over D. Moreover, let $M_i \coloneqq \max_{x \in D} \Big| f(y) \Big|$. We have that

$$\begin{aligned} m(X_i) &= \sum_{x \in D} \left(\sum_{y \in N_D^-(x)} \frac{f(y)}{d} \right)^2 \\ &\overset{(2.3)}{=} \sum_{x \in D} \left(\sum_{y \in N_D^-(x)} d \left(\frac{f(y)}{d} \right)^2 - \sum_{z,w \in N_D^-(x)} \frac{\Big(f(z) - f(w)\Big)^2}{2d^2} \right) \\ &= m(X_{i-1}) - \sum_{\substack{x \in D \\ z,w \in N_D^-(x)}} \frac{\Big(f(z) - f(w)\Big)^2}{2d^2} \leq m(X_{i-1}) - \frac{M_i^2}{2d^2} \end{aligned}$$

where the inequality follows from Lemma 4.12. Note that if $m(X_{i-1}) = \varepsilon$ then $|f(x)|^2 \geq \varepsilon/k$ for some $x \in D$, and thus $M_i^2 \geq \varepsilon/k$, so

$$m(X_i) \leq m(X_{i-1}) - \varepsilon/2kd^2 \leq m(X_{i-1})(1 - 1/2k^3).$$

Therefore,

$$\max_{x\in D}\left(\mathbb{P}(X_n = x) - \frac{1}{k}\right)^2 \le m(X_n) \le m(X_0)\left(1 - \frac{1}{2l^3}\right)^n$$
$$\overset{(4.7)}{\le} \left(1 - \frac{1}{2k^3}\right)^n . \qquad \square$$

4.3.3 Allocation algorithm

Let F, R be digraphs. An *allocation* of F to R is a homomorphism from F to R, i.e., a map $\varphi : V(F) \to V(R)$ such that every edge $u{\to}v \in E(F)$ is mapped to an edge $\varphi(u){\to}\varphi(v) \in E(R)$. In our applications, R will usually be a suitable reduced digraph of the host graph G. allocation

Algorithm 4.14 below is a randomised procedure which defines an allocation of a rooted tree T to a digraph D, and is inspired by the Vertex Allocation Algorithm [51, Section 3.2]. The algorithm in [51], however, is only applied when D is a directed cycle with loops in every vertex (i.e., $v{\to}v \in E(D)$ for all $v \in D$), whereas here there are no loops but we require that $\delta^0(D) \ge 1$. Essentially, Algorithm 4.14 steps through the vertices of T in an ancestral order, and defines the homomorphism $\varphi : T \to D$ uniformly at random at each step, with the restriction that siblings (i.e., vertices with the same parent) are mapped to the same vertex if the edge between them and the parent has the same orientation.

Algorithm 4.14: The Vertex Allocation Algorithm

Input : an oriented tree T of order n with root r, an ancestral order $r = t_1, \dots, t_n$ of T, a digraph D with $\delta^0(D) \ge 1$ and $x_1 \in V(D)$.

```
for τ = 1 to n do
    if τ = 1 then define φ(r) := x_1.
    else if φ(t_τ) is undefined then
        Let t_σ be the parent of t_τ and let x_σ = φ(t_σ).
        Choose x_τ^+ ∈ N_D^+(x_σ) and x_τ^- ∈ N_D^-(x_σ) uniformly at random
          independently of all other choices.
        Let t_τ^1, ..., t_τ^s be the children of t_σ.
        for i = 1 to s do
            if t_τ^i ∈ N_T^+(t_σ) then define φ(t_τ^i) := x_τ^+.
            else define φ(t_τ^i) := x_τ^-.
```

The next lemma states that Algorithm 4.14 will always build a homomorphism and, moreover, that if T is sufficiently large and D is a regular expander, then roughly the same number of vertices of T is mapped to each vertex of D.

Lemma 4.15. Let T be an oriented tree of order n rooted at r, let D be a regular expander digraph of order k, and let $x \in D$. If φ is the allocation we obtain by applying the Algorithm 4.14 to T and D, then the following properties hold.

(a) φ is a homomorphism $\varphi : T \to D$ and $\Delta(\varphi) \leq 3$.

(b) Let $u, v \in V(T)$, where u lies on the path from r to v, let P be the path between u and v, and let $W \coloneqq V(P) \setminus \{u\}$. For all $j \in D$, the allocation of W, conditioned on the event '$\varphi(u) = j$' is a random P-walk on D.

(c) Suppose that $1/n \ll 1/k$. Let $u, v \in V(T)$ be such that u lies on the path from r to v, and $\mathrm{dist}_T(u, v) \geq 5k^3 \log\log\log n$. Then for all $i, j \in D$,

$$\mathbb{P}(\, v \text{ is allocated to } i \mid u \text{ is allocated to } j\,) = \frac{1}{k}\left(1 \pm \frac{1}{4\log\log n}\right).$$

(d) Suppose that $1/n \ll 1/k, \zeta, 1/C$ and that $\Delta(T) \leq (\log n)^C$. Let S be a subset of $V(T)$ with at least $n^{2/3+\zeta}$ vertices. Then with probability $1-\mathrm{o}(1)$ each of the vertices of D has $|S|\left(\frac{1}{k} \pm \frac{1}{\log\log n}\right)$ vertices of S allocated to it.

(e) Suppose that $1/n \ll 1/k, \beta, 1/C$, that D is d-regular, and let $Q \subseteq E(D)$. If $S \subseteq E(T)$ contains at least βn vertex-disjoint edges, then φ allocates at least $|S|/4kd$ edges of S to each edge of Q with probability $1 - \mathrm{o}(1)$.

Proof. Note that every edge of T is mapped to an edge of D; moreover, for all $v \in T$, the neighbours of v fall in 3 categories: parent, in- or outchild of v, and all vertices in each of these categories are allocated to the same cluster, so (a) holds. In particular, (b) also holds, since the allocation of vertices along any path match the choices that would be made in a random P-walk. From this point onward, let us assume $V(D) = [k]$. We now prove item (c). Suppose $1/n \ll 1/k$. Let $P(u, v)$ be the path from u to v in T, and let ℓ be the length of $P(u, v)$, so $\ell \geq 5k^3 \log\log\log n$. Let $u = v_0, v_1, \ldots, v_\ell = v$ be the vertices of $P(u, v)$. Suppose that u has been allocated to $x_0 \in D$, and for all $i \in \{0, 1, \ldots, \ell\}$ let X_i be the vertex to which v_i is allocated (so $X_0 = x_0$). These variables form a random P-walk (by the previous item), and therefore by Lemma 4.13 for all $x \in D$ we have that

$$\mathbb{P}(X_n = x) = \frac{1}{k} \pm \left(1 - \frac{1}{2k^3}\right)^{\ell/2} = \frac{1}{k} \pm \mathrm{e}^{-\ell/4k^3} = \frac{1}{k}\left(1 \pm \frac{1}{4\log\log n}\right),$$

which proves (c).

We now prove (d). By Lemma 2.6, there exists an integer $s \leq 3n^{1/3}$, vertices $v_1, \ldots, v_s \in V(T)$ and pairwise-disjoint subsets $F_1, \ldots, F_s$ of $V(T)$ such that $|\bigcup_{i=1}^{s} F_i| \geq n - n^{5/12}$ and $|F_i| \leq n^{2/3}$ for each $i \in [k]$, such that if $j < i$, then any path from r or any vertex of F_j to any vertex of F_i passes through the vertex v_i, and also such that $\mathrm{dist}(v_i, F_i) \geq 5k^3 \log\log\log n$. Write $\delta_n := \frac{1}{\log\log n}$; we shall prove that

(†) with probability $1 - o(1)$, for all $j \in [k]$ at most $|S|(\frac{1}{k} + \frac{\delta_n}{2k})$ vertices from $\bigcup_{i\in[s]} F_i \cap S$ are allocated to cluster V_j.

Note that (†) implies (d). Indeed, since the number of vertices of T not contained in any of the sets F_i is at most $n^{5/12} \leq \delta_n|S|/2k$, if (†) holds then for any $j \in [k]$ in total at most $|S|(1+\delta_n)/k$ vertices of S are allocated to V_j. It follows that at least $|S| - (k-1)|S|(1+\delta_n)/k \geq |S|(1/k - \delta_n)$ vertices of S are allocated to V_j, so (d) holds.

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

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

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

$$\begin{aligned}\mathbb{E}(\, X_i^j \mid X_{i-1}^j, \dots, X_1^j\,) &\leq \max_{q\in[k]} \mathbb{E}(\, X_i^j \mid X_{i-1}^j, \dots, X_1^j, v_i \in V_q\,)\\ &= \max_{q\in[k]} \mathbb{E}(\, X_i^j \mid v_i \in V_q\,) = \max_{q\in[k]} \frac{\sum_{x\in F_i\cap S} \mathbb{P}(\, x \in V_j \mid v_i \in V_q\,)}{n^{2/3}}\\ &\overset{(c)}{\leq} \left(\frac{1}{k} + \frac{\delta_n}{4k}\right)\frac{|F_i \cap S|}{n^{2/3}}.\end{aligned}$$

We apply Lemma 2.15 with

$$\mu := \left(\frac{1}{k} + \frac{\delta_n}{4k}\right)\frac{|S|}{n^{2/3}} \geq \left(\frac{1}{k} + \frac{\delta_n}{4k}\right)\sum_{i\in[s]}\frac{|F_i \cap S|}{n^{2/3}},$$

to obtain

$$\begin{aligned}\mathbb{P}\Big(\sum_{i\in[s]} X_i^j > (1+\delta_n/8)\mu\Big) &\leq \exp\left(\frac{-(\delta_n/8)^2\mu}{3}\right)\\ &= \exp\left(-\frac{\delta_n^2(1+\delta_n/4)|S|}{192kn^{2/3}}\right) \leq \exp\big(-n^{\zeta/2}\big)\end{aligned}$$

where the second inequality holds since we assumed that $1/n \ll 1/k, \zeta$. Taking a union bound, we find that with probability $1 - o(1)$ we have for each $j \in [k]$ that

$$n^{2/3} \sum_{i \in [s]} X_i^j \leq n^{2/3}(1 + \delta_n/8)\mu \leq |S| \left(\frac{1}{k} + \frac{\delta_n}{2k} \right).$$

In other words, for each $j \in [k]$ there are at most $|S|\left(\frac{1}{k} + \frac{\delta_n}{2k}\right)$ vertices of $\bigcup_{i=1}^r F_i \cap S$ allocated to V_j, so (†) holds.

To conclude, let $\prec$ be an ancestral order of T and, for each edge $e \in S$, let x_e and y_e be the endvertices of e, labelled so that $x_e \prec y_e$. At least half of the edges in S have the same orientation with respect to $\prec$, so let S' be a subset of S with at least $|S|/2$ edges which are all oriented, say, from x_e to y_e. Let $u_1 \to v_1, \ldots, u_q \to v_q$ be the edges in Q. By item (d), with probability $1 - o(1)$ there are at least $|S'|\left(\frac{1}{k} \pm \frac{1}{\log \log n}\right)$ vertices $x_e \in e \in S'$ allocated to each u_j for all $j \in [q]$. Let us call this event E_1. Conditioned on the occurrence of E_1, and for each $j \in [q]$, let Z_j be the number of edges of S' which are allocated to $u_j \to v_j$; then, for any fixed $j \in [q]$, since D is d-regular (and d is bounded), it follows that Z_j is a binomial random variable with expectation at least $|S'|/d$, so the probability that $Z_j < |S'|/2d$ decreases exponentially with n. By a union bound (over these q events), it follows that with probability $1 - o(1)$ we have that $Z_j \geq |S'|/2d \geq |S|/4kd$ for all $j \in [q]$; we call this event E_2. Since both E_1 and $E_2|E_1$ (i.e., E_2 conditioned on the occurrence of E_1) happen with probability $1 - o(1)$, we conclude that item (e) holds as required. □

4.3.4 Embedding

In this section we describe an algorithm for embedding trees in dense digraphs. This algorithm will be used (with a few modifications) in the proof of many of our results (such as Lemmas 4.23 and 4.20 and Theorem 1.15). It receives as input a tree T, digraphs G (where T is to be embedded) and R (a reduced graph of G), a partition $\mathcal{V} \coloneqq \{ V_i : i \in R \}$ of $V(G)$, and a homomorphism $\varphi : T \to R$. It embeds vertices greedily, one at a time, so that each vertex $x \in T$ is embedded to the set $V_{\varphi(x)}$. The main result of this section is Lemma 4.16, which (roughly speaking) states that if $\Delta(\varphi)$ is bounded, the number of vertices φ maps to any V_i is always somewhat smaller than $|V_i|$, and if the edges of R correspond to regular pairs in G, then the algorithm successfully embeds T to G.

This algorithm has grown out of an embedding algorithm used by Kühn, Mycroft and Osthus [51] in their solution to Conjecture 1.7, to embed trees in tournaments. In their application, however, R was always a cycle (so $\Delta(\varphi) \leq 2$) and some edges were embedded within the V_i.

For $v \in T$, we write $C^-(v)$ for the children of v in $N^-_T(v)$, $C^+(v)$ for the children of v in $N^+_T(v)$ and $C(v)$ for $C^-(v) \cup C^+(v)$. We write S_x for the star $T\Big[\{x\} \cup C(x)\Big]$ induced by x and its children. $C^-(v)$, $C^+(v)$ $C(v)$

Embedding algorithm. If at any point in the description below there is more than one possible choice available, we take the lexicographically first of these, so that for each input the output will be uniquely defined — thus making the algorithm deterministic. Furthermore, if at any point some required choice cannot be made, terminate with failure.

At each time τ, with $1 \leq \tau \leq n$, we shall embed a vertex t_τ to a vertex $v_\tau \in V_{\varphi(t_\tau)}$; we will also reserve sets A^-_τ, A^+_τ for the children of t_τ. We say that a vertex t_s of T is *open* at time τ if t_s has been embedded but some child of t_s has not yet been embedded. open

Input. An oriented tree T with ancestral order $t_1 \prec \ldots \prec t_n$ of T, a digraph R, a homomorphism $\varphi : T \to R$. Also, a digraph G, a partition $\mathcal{V} := \{\, V_i : i \in R \,\}$ of $V(G)$, a vertex $v_1 \in V_{\varphi(t_1)}$ and constants β and γ.

Procedure. At each time τ, with $1 \leq \tau \leq n$, we take the following steps.

Step 1. 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 children of open vertices, so

$$B^\tau := \{v_1, \ldots, v_{\tau-1}\} \cup \bigcup_{t_s \,:\, t_s \text{ is open}} \Big(A^-_s \cup A^+_s\Big)$$

For each $V_i \in \mathcal{V}$, let $V^\tau_i := V_i \setminus B^\tau$, so V^τ_i is the set of available vertices of V_i.

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

(2.1) Let t_σ be the parent of t_τ (so A^-_σ, A^+_σ were reserved for the children of t_σ).

(2.2) If $t_\sigma \to t_\tau$, let $W := A^+_\sigma \cap V_{\varphi(t_\tau)}$; otherwise let $W := A^-_\sigma \cap V_{\varphi(t_\tau)}$.

(2.3) Choose $v_\tau \in W$ such that

$$\begin{aligned} \deg^-_G(v_\tau, V^\tau_i) &\geq \gamma m \qquad \text{for all } i \in \varphi\Big(C^-(t_\tau)\Big), \\ \deg^+_G(v_\tau, V^\tau_j) &\geq \gamma m \qquad \text{for all } j \in \varphi\Big(C^+(t_\tau)\Big). \end{aligned} \tag{4.8}$$

(2.4) Embed t_τ to v_τ.

Step 3. In Step 2 we embedded t_τ to a vertex $v_\tau \in W$. For each $x \in C^-(t_\tau)$, choose a set $A^-_x \subseteq N^-_G(v_\tau) \cap V^\tau_{\varphi(x)}$ containing at most $2m^{1-1/\Delta(\varphi)}$ vertices and which is $(\beta, \gamma, \varphi, m)$-good for S_x; let A^-_τ be the union of these sets. Similarly, for each $y \in C^+(\tau)$, choose a set $A^+_y \subseteq N^+_G(v_\tau) \cap V^\tau_{\varphi(y)}$ containing at most $2m^{1-1/\Delta(\varphi)}$

vertices and which is $(\beta, \gamma, \varphi, m)$-good for S_y; choose these sets so that they are pairwise disjoint and let A^+_τ be their union.

- **Termination.** Terminate after every vertex of T 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 4.16. Suppose that $1/n \ll 1/C$, that $1/n \ll 1/k \ll \varepsilon \ll \gamma \ll \beta \ll d \ll \alpha \leq 2$ and let $m := n/k$.

(i) Let T be an oriented tree of order at most n with root t_1 and $\Delta(T) \leq (\log n)^C$, and let $\prec$ be a tidy ancestral order of T.

(ii) Let R and G be digraphs, with $|R| = k$, and suppose there exists a partition $\{ V_i : i \in R \}$ of $V(G)$ such that $(1+\alpha)m \leq |V_i| \leq 3m$ for each $i \in R$.

(iii) Suppose that φ is a homomorphism from T to R such that each edge $x{\to}y \in E(T)$ is mapped to an (d, ε)-regular pair $G[V_{\varphi(x)}{\to}V_{\varphi(y)}]$, such that $\Delta(\varphi) \leq 4$ and which maps at most $(1 + \alpha/2)m$ vertices to each V_i.

(iv) Let $v_1 \in V_{\varphi(t_1)}$ be a vertex such that $\deg^-_G(v_1, V_{\varphi(x)}), \deg^+_G(v_1, V_{\varphi(y)}) \geq \gamma m$ for all $x \in C^-(t_1)$ and all $y \in C^+(t_1)$.

Under these assumptions, the embedding algorithm (with parameters β, γ) successfully embeds T to G.

We remark that any fixed constant q with $\alpha \ll 1/q$ could replace 4 in the bound $\Delta(\varphi) \leq 4$ in (iii) above.

Proof. The Vertex Embedding Algorithm will only fail if at some point it is not possible to make a required choice. We will show this is never the case, hence the algorithm succeeds.

Consider the set of unavailable vertices B^τ at some time τ. Since the algorithm embeds each vertex $x \in T$ to $V_{\varphi(x)}$, we know that at most $(1 + \alpha/2)m$ vertices of each V_j are already occupied. Furthermore, suppose that a vertex $t_\sigma \in T$ is open at time τ. Then $\sigma < \tau$ and t_σ has a child t_ρ with $\tau \leq \rho$. Since we are processing the vertices of T in a tidy order, there can be at most $\Delta(T) \log_2 n$ of these children vertices $t_\rho \in T$.

Recall that each reserved set has size $2m^{1-1/\Delta(\varphi)}$, and thus, at any time τ, the total number of vertices in reserved sets is at most $(2m^{1-1/\Delta_\varphi})\Delta(T)(\log_2 n) \leq \alpha m/4$. 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 $|V^\tau_j| \geq \alpha m/4$.

We now argue that all required choices can be made. Indeed, in Step (2.3) we choose $v_\tau \in W$ satisfying (4.8). But W is $(\beta, \gamma, \varphi, m)$-good for S_{t_τ} — it has been reserved at Step 3 when processing vertex t_σ. In particular, W contains at least $\gamma m^{1/\Delta(\varphi)}$ vertices z such that $\deg^-_G(z, V^\tau_{\varphi(x)}) \geq \gamma m$ and $\deg^+_G(z, V^\tau_{\varphi(y)}) \geq \gamma m$ for all $x \in C^-(t_\tau)$ and all $y \in C^+(t_\tau)$. Moreover, t_σ has been open since time $\sigma < \tau$ and hence the only vertices which have been embedded to W are children of t_σ (of which there are at most $\Delta(T) < \gamma m^{1/\Delta(\varphi)}/2$), so we can choose v_τ as required.

In Step 3 we wish to choose $A^-_x \subseteq N^-_G(v_\tau) \cap V^\tau_{\varphi(x)}$ and $A^+_y \subseteq N^+_G(v_\tau) \cap V^\tau_{\varphi(y)}$ for each $x \in C^-(t_\tau)$ and each $y \in C^+(t_\tau)$, such that A^-_x is $(\beta, \gamma, \varphi, m)$-good for S_x and A^+_y is $(\beta, \gamma, \varphi, m)$-good for S_y. By (4.8), v_τ has at least γm inneighbours in $V^\tau_{\varphi(x)}$ for each $x \in C^-(t_\tau)$ and at least γm outneighbours in $V^\tau_{\varphi(y)}$ for each $y \in C^+(t_\tau)$; for $\tau = 1$ this holds by our hypothesis (iv) instead.

Since $\left|C^-(t_\tau)\right| + \left|C^+(t_\tau)\right| \leq \Delta(T)$, the total number of vertices which are reserved in Step 3 (at any given time τ) is at most $\Delta(T)2m^{1-1/\Delta(\varphi)} \leq \gamma m/3$. Therefore, for each $x \in C^-(t_\tau)$ we have that at any point during Step 3, $V^\tau_{\varphi(x)} \cap N^-_G(v_\tau)$ contains at least $\gamma m - \gamma m/3 > \gamma m/2$ unreserved vertices. Similarly, for each $y \in C^+(t_\tau)$ we have that at any point during Step 3, $V^\tau_{\varphi(y)} \cap N^-_G(v_\tau)$ contains at least $\gamma m - \gamma m/3 > \gamma m/2$ unreserved vertices. Hence, by Lemma 4.4, it follows that for all $x \in C^-(t_\tau)$ and all $y \in C^+(t_\tau)$ there exist $(\beta, \gamma, \varphi, m)$-good sets for S_x and S_y, each of size at most $2m^{1/\Delta(\varphi)}$ and containing only unreserved vertices of $V^\tau_{\varphi(x)}$ and $V^\tau_{\varphi(y)}$, respectively. Hence all choices in Step 3 can be made. □

As promised (in Section 3.3.2), we now sketch of the proof of Lemma 3.4, since this is based on the argument above. We recall that Lemma 3.4 states that given a semi-canonical allocation of a tree T (whose maximum degree is at most polylogarithmic on $|T|$) to a cycle of cluster tournaments, there exists an embedding of T which respects this allocation.

A key difference to the proof above (related to the fact that in Lemma 3.4 is that we sometimes embed edges within clusters) is that $\Delta(T_{\text{canon}}) \leq (\log n)^{2C}$, so the components of T corresponding to vertices of T_{canon} are small; this ensures the crucial property that the number of vertices in reserved sets remains is sublinear at any step (see embedding algorithm and discussion on page 39).

Proof sketch (Lemma 3.4). As mentioned above this proof proceeds by arguing (as in Lemma 4.16) that the choices required by the algorithm can always be made. Informally, the argument goes as follows: at any step of the algorithm, there are many unused vertices; moreover, the sets reserved for the children of any vertex are large enough that we can always find (c, γ)-good sets in them; moreover, these good sets are large enough that we can embed any star of T corresponding to the currently processed vertex of T_{canon}.

We build T_{canon} as described above, let R be the directed cycle $1\to 2\to\cdots\to k\to 1$, let $\beta = c$ and $m = n/k$, and note that with these definitions a (β,γ,φ,m)-good (Definition 4.3 on page 79) set is precisely a (c,γ)-good set (since R is a directed cycle). To deal with components of T which are embedded within a cluster, we apply Theorem 3.5. More precisely, our modified version of the embedding algorithm only fails if we cannot find a copy of S in the unused vertices reserved for v. But this star contains at most $\Delta(T) \leq (\log n)^C$ vertices, and as argued in the proof of Lemma 4.16 there are at least $\gamma m^{1/4}/2 \geq 3|S|$ unused vertices in the reserved set; this means that we can find the desired copy of S in the unused vertices reserved for v, by Theorem 3.5. □

4.3.5 An approximate result (Theorem 4.17)

As a corollary of Lemmas 4.15 and 4.16 we obtain an approximate result for embedding almost-spanning trees in digraphs of high semidegree.

Theorem 4.17. For all positive α, ε, C with $\varepsilon \ll \alpha$ there exists n_0 such that the following holds for all $n \geq n_0$. If G is a digraph of order $(1+\varepsilon)n$ and minimum semidegree at least $(1/2+\alpha)n$, then G contains every oriented tree of order n and maximum degree at most $(\log n)^C$.

Proof. Let G be a digraph of order $(1+\varepsilon)n$ such that $\delta^0(G) \geq (1/2+\alpha)n$. We introduce constants ε', d, k, η such that $\frac{1}{n} \ll \frac{1}{k} \ll \varepsilon' \ll d \ll \eta \ll \alpha, \varepsilon, C$, and apply Lemma 4.2 to obtain a partition $V_0 \,\dot\cup\, V_1 \,\dot\cup \cdots \dot\cup\, V_k$ and a digraph $R^\star$ with $V(R^\star) = [k]$ such that

(a) $|V_0| < \varepsilon' n$ and $m := |V_1| = \cdots = |V_k|$;

(b) For each $i \in [k]$ we have $G[V_{i-1}\to V_i]$ and $G[V_i \to V_{i+1}]$ are (d,ε')-super-regular;

(c) For all $i, j \in [k]$ we have $i\to j \in E(R^\star)$ precisely when $G[V_i \to V_j]$ is (d,ε')-regular.

(d) For all $i \in [k]$ we have $\deg^0_{R^\star}(i,[k]) \geq (1/2+\eta)k$.

We also introduce constants β, γ such that $1/n \ll 1/k \ll \varepsilon' \ll \gamma \ll \beta, d \ll \eta$. Note that $R^\star$ contains a regular expander D by Lemma 4.10, and also that $m = (1+\eta)n/k$. Fix a vertex $r \in T$ and a vertex $x \in D$. Then T, r, D and x satisfy the hypothesis of Lemma 4.15, so if we apply the Algorithm 4.14 we obtain a homomorphism $\varphi : T \to D$ such that $\Delta(\varphi) \leq 3$ and such that for all $i \in [k]$ we have $\left|\varphi^{-1}(V_i)\right| \leq n/k \pm n/\log\log n \leq (1+\eta/2)n/k$. Moreover, by Lemma 4.4 there exists a vertex $v_1 \in V_{\varphi(r)}$ such that for each $x \in C^-(r)$ we have $\deg^-_G(x, V_{\varphi(x)}) \geq \gamma m$ and for each $y \in C^+(r)$ we have $\deg^+_G(y, V_{\varphi(x)}) \geq \gamma m$. Hence, T, x, G, v_1, D, φ and the constants above (with η here as α) satisfy the hypothesis of Lemma 4.16, so if we apply the embedding algorithm with these parameters it successfully finds a copy of T in G. □

4.4 Trees with many bare paths

Our goal is to prove Lemma 4.20, which is a version of Theorem 1.13 for trees with 'many bare paths'. Our main tools are the Allocation Algorithm (Algorithm 4.14) and the Embedding Algorithm (described in Section 4.3.4). However, since the trees here will be spanning trees of the host graph, we need to adapt both procedures. Lemma 4.18 below guarantees the existence of a special allocation of vertices; the proof of Lemma 4.20 describes how to modify the embedding algorithm.

4.4.1 Allocation

Roughly speaking, Lemma 4.18 states that given a tree T (containing many vertex-disjoint bare paths), a graph $R^\star$ on $V_0 \,\dot\cup\, [k]$ (with linear minimum semidegree in $[k]$), and a Hamilton cycle $H \subseteq R^\star\big[[k]\big]$, there exists a homomorphism φ from T to $R^\star$ which satisfies many properties: $\Delta(\varphi)$ is bounded and maps a large collection of $\prec$-isomorphic bare paths of T evenly along H; φ covers V_0 bijectively, mapping to V_0 centres of bare paths and, moreover, φ distributes the neighbours of the vertices mapped to V_0 somewhat evenly over $[k]$; and $\left|\varphi^{-1}(i)\right| = \left|\varphi^{-1}(j)\right|$ for all $i, j \in [k]$.

Lemma 4.18. Suppose $\frac{1}{n} \ll \frac{1}{k} \ll \varepsilon \ll \lambda \ll \eta \ll \alpha$ and that $\frac{1}{n} \ll \frac{1}{C}$. Let T be an oriented tree of order n such that $\Delta(T) \leq (\log n)^C$. Suppose that there exists a collection $\mathcal{P}$ of λn vertex-disjoint bare paths of T with order 7. Let $R^\star$ be a graph with vertex set $V_0 \dot\cup [k]$, such that $|V_0| < \varepsilon n$, where $n - |V_0| \equiv 0 \bmod k$, and such that for each $i \in [k]$ we have $\deg^0_{R^\star}(i, [k]) \geq (1/2 + \eta)k$ and for each $v \in V_0$ we have $\deg^0_{R^\star}(v, [k]) > \alpha k$. Suppose also that H is a directed Hamilton cycle in $R^\star\big[[k]\big]$. Then there exists a tidy ancestral order $\prec$ of T, disjoint subsets $\mathcal{P}^0, \mathcal{P}^H \subseteq \mathcal{P}$ and a homomorphism $\varphi : T \to R^\star$ such that

(i) $\Delta(\varphi) \leq 4$;

(ii) $|\mathcal{P}^0| = |V_0|$ and for each $P \in \mathcal{P}^0$ the centre of P is mapped to V_0;

(iii) φ maps precisely one vertex of T to each $v \in V_0$;

(iv) φ maps at least $\lambda n/24k$ centres of paths in $\mathcal{P}^H$ to each $i \in [k]$;

(v) If $N := \{\, N^-_T(x) \cup N^+_T(x) : \varphi(x) \in V_0 \,\}$ are the neighbours of vertices φ maps to V_0, then φ maps at most $2\varepsilon n/\alpha k$ vertices of N to each $i \in [k]$;

(vi) For each $P \in \mathcal{P}^H$, the restriction of φ to middle(P) is a homomorphism from middle(P) to H;

(vii) $\left|\varphi^{-1}(1)\right| = \left|\varphi^{-1}(2)\right| = \cdots = \left|\varphi^{-1}(k)\right|$.

The proof is divided in four stages, which we now outline. In the **setup** we fix a tree partition $\{T_1, T_2\}$ of T such that T_1 contains a collection $\mathcal{P}'$ of many bare paths whose middles are $\prec$-isomorphic to some rooted path P_{ref}. We also fix a partition $\mathcal{P}' = \mathcal{P}^0 \dot\cup \mathcal{P}^H \dot\cup \mathcal{P}^\diamond$ such that $|\mathcal{P}^0| = |V_0|$ and $|\mathcal{P}^H|, |\mathcal{P}^\diamond| \geq |\mathcal{P}'|/4$, and map middle$(P)$ for each path in $\mathcal{P}'$ in the following manner. For each $P \in \mathcal{P}^0$, we map middle(P) so that $c(P)$ is mapped (bijectively) to a vertex in V_0 and the neighbours of such vertices are mapped to vertices in $[k]$ somewhat evenly. For each $P \in \mathcal{P}^H$, we map middle(P) along H so that the same number of centres are mapped to each $i \in [k]$. Next, we find a spanning subgraph $J^\diamond$ of $R^\star\big[[k]\big]$ which is P_{ref}-connected, and map middle(P) for all $P \in \mathcal{P}^\diamond$ so that for each diamond $x {\preceq^{y}_{w}\succeq} z \in J^\diamond$ a linear number of (middles of) paths is mapped to xyz and a linear number of (middles of) paths is mapped to xwz—this will be useful in the last stage of the proof. Finally, we let T_1' be the tree we obtain by contracting all edges in each path $P \in \mathcal{P}'$.

In **phase 1** we apply Lemma 4.10 to find a regular expander $J \subseteq R^\star$ and then allocate T_1' to J using Algorithm 4.14. We concatenate this allocation to the maps defined in the setup and complete (greedily) the allocation of the contracted paths, obtaining a homomorphism of T_1 into $R^\star$. By Lemmas 4.6 and 4.15, this allocation of T_1 is almost uniform over $[k]$, with error $\mathrm{O}(\lambda n/\eta k)$.

In **phase 2** we apply Algorithm 4.14 again, this time allocating T_2 to a regular expander J^{blow} which is a subgraph of a weighted blow-up of $R^\star\big[[k]\big]$. The allocation we obtain is again almost uniform, but biased so as to correct the linear errors introduced in the setup stage when embedding paths in $\mathcal{P}'$. Altogether, the maps we defined form an allocation of T to $R^\star$. We argue that the resulting allocation of T satisfies all of the properties stated in the lemma, except perhaps (vii). However, by Lemma 4.15, we have at most sublinear errors in the order of the preimages of each $i \in [k]$.

To conclude the proof, in **phase 3** we modify the mapping of the centres of paths in $\mathcal{P}^\diamond$ along $J^\diamond$, so as to ensure (vii). This requires only a sublinear number of changes, and thus is possible by Lemma 4.9.

Proof. As outlined above, this proof is divided in four parts.

► **Setup.** By Lemma 2.3, there exists a tree-partition $\{T_1, T_2\}$ of T such that $|T_1| \le 2n/3$ and we may assume, without loss of generality, that T_1 contains $\lambda n/2$ paths of $\mathcal{P}$. Let r be the common vertex of T_1 and T_2; by Lemma 2.5 we can fix a tidy ancestral order $\prec$ of T starting with r and such that each vertex of T_1 precedes all vertices in $V(T_2) \setminus \{r\}$ in this order.

Let $\mathcal{P}'$ be a collection of at least $\lambda n/16$ and at most $\lambda n/8$ paths of $\mathcal{P}$ in T_1 whose middle sections are $\prec$-isomorphic to some rooted path P_{ref} of order 3. For each $P \in \mathcal{P}'$ we write $v_1^P, v_2^P, \dots, v_7^P$ to denote the vertices of P, labelled so that $v_1^P \prec v_2^P \prec \cdots \prec v_7^P$.

Fix (arbitrarily) a partition $\mathcal{P}' = \mathcal{P}^0 \mathbin{\dot\cup} \mathcal{P}^H \mathbin{\dot\cup} \mathcal{P}^\diamond$ such that $|\mathcal{P}^0| = |V_0| < \varepsilon n \le |\mathcal{P}'|/3$, such that $|\mathcal{P}^H|, |\mathcal{P}^\diamond| \ge |\mathcal{P}'|/3$, such that $|\mathcal{P}^H| \equiv 0 \bmod k$ and such that $|\mathcal{P}^\diamond| \equiv 0 \bmod 2(k-1)$. We shall define a homomorphism φ mapping the middle of each path $P \in \mathcal{P}'$ to $R^\star$. We do this separately for $\mathcal{P}^0, \mathcal{P}^H$ and $\mathcal{P}^\diamond$, as follows.

Let $\varphi_0 : \{\, c(P) : P \in \mathcal{P}^0 \,\} \to V_0$ be an arbitrary bijection. For each $P \in \mathcal{P}^0$, we proceed as follows. Write $c_P \coloneqq \varphi_0(v_4^P)$, so $c_P \in V_0$. We shall map each $\{v_3^P, v_5^P\}$ to $[k]$, extending φ_0 so that the preimage of each $i \in [k]$ has size at most

$$\left|(\varphi_0)^{-1}(i)\right| \le \frac{2|V_0|}{\alpha k}. \tag{4.9}$$

To do so, let B_3 be a bipartite graph with vertex classes V_0 and $[k]$, with an edge connecting $i \in [k]$ to $c_P \in V_0$ if mapping $v_3^P \mapsto i$ would create a homomorphism from $P\big[\{v_3^P, v_4^P\}\big]$ to $R^\star$. Note that each $x \in V_0$ has degree at least αk, so, by Lemma 4.6, there exists a subgraph B_3' of B_3 containing V_0, such that each $x \in V_0$ has degree 1 and each vertex in $[k]$ has degree at most $|V_0|/\alpha k$ in B_3'. For each edge $ic_P \in E(B'_3)$, where $i \in [k]$ and $c_P \in V_0$, we set $\varphi_0(v_3^P) = i$. We proceed similarly to define $\varphi_0(v_5^P)$ for each $P \in \mathcal{P}^0$. Note that φ_0 satisfies (iii) and (v).

Fix a partition $\mathcal{P}^H = \mathcal{P}_1^H \dot\cup \cdots \dot\cup \mathcal{P}_k^H$ with parts of equal size. Note that for each $P \in \mathcal{P}^H$ and for each $i \in [k]$ there is a unique homomorphism $\varphi_{H,P}$ from middle(P) to H such that $\varphi_{H,P}\big(c(P)\big) = i$. We set φ_H to be the union of all homomorphisms $\varphi_{H,P}$. Note that φ_H satisfies (iv) and (vi).

By Lemma 4.8, $R^\star$ contains a P_{ref}-connected subgraph $J^\diamond$ with $\Delta(J^\diamond) \le 4/\eta$ which is the union of P_{ref}-diamonds $\left(x_i \substack{\prec y_i \succ \\ \prec w_i \succ} z_i\right)_{i=1}^{k-1}$. We fix a partition $\mathcal{P}^\diamond = \mathcal{P}_1^y \dot\cup \mathcal{P}_1^w \dot\cup \cdots \dot\cup \mathcal{P}_{k-1}^y \dot\cup \mathcal{P}_{k-1}^w$ with parts of equal sizes, that is of size $|\mathcal{P}^\diamond|/2(k-1) \ge |\mathcal{P}^\diamond|/3k$ each. For each $i \in [k-1]$, each $P^y \in \mathcal{P}_i^y$ and each $P^w \in \mathcal{P}_i^w$, we define $\varphi_\diamond$ as the unique $\prec$-isomorphisms from middle(P^y) to $x_iy_iz_i$ and from middle(P^w) to $x_iw_iz_i$. Note that for each $i \in [k]$ we have that the size of the pre-image $(\varphi_\diamond)^{-1}(i)$ is

$$\left|(\varphi_\diamond)^{-1}(i)\right| \le \Delta(J^\diamond)\frac{|\mathcal{P}^\diamond|}{2(k-1)} \le \frac{8}{\eta}\left(\frac{2}{3}|\mathcal{P}'|\right) \Big/ 2(k-1) \le \frac{\lambda n}{2\eta k} \tag{4.10}$$

To conclude the setup, we define φ as the disjoint union of the maps φ_0, φ_H and $\varphi_\diamond$ (this definition will be extended later), and form a tree T'_1 by contracting the edges of each path in $\mathcal{P}'$ (so that each path becomes a single vertex). If the endvertices of a path P so contracted were v_1^P and v_7^P, we write $v_{1,7}^P$ for the vertex resulting from the contraction.

- **Phase 1**. By Lemma 4.10, $R^\star$ contains a $25k^{2/3}/\eta$-regular expander J. We apply Algorithm 4.14 to T'_1 and J. This yields a homomorphism ϱ_1 from the former to the latter. For each $P \in \mathcal{P}'$, we proceed as follows. Recall that $\varphi(v_3^P), \varphi(v_4^P)$ and $\varphi(v_5^P)$ have already been defined in the setup stage, i.e., that middle(P) has already been mapped. We define $\varphi(v_1^P) = \varphi(v_7^P) = \varrho_1(x)$, where $c_P = v_{1,7}^P$ is the vertex resulting from the contraction of P.

We complete the mapping of T_1 by defining $\varphi(v_2^P)$ and $\varphi(v_6^P)$ greedily, using Lemma 4.6, as follows. We form an auxiliary bipartite graph B_2 with classes $\mathcal{P}'$ and $[k]$, with edges joining $j \in [k]$ to $P \in \mathcal{P}'$ whenever setting $\varphi(v_2^P) = j$ completes a homomorphism from $P\Big[\{v_1^P, v_2^P, v_3^P\}\Big]$ to $R^k := R^\star\Big[[k]\Big]$. Note that each $P \in \mathcal{P}'$ has at least ηk neighbours in $[k]$, so by Lemma 4.6 there exists $B'_2 \subseteq B_2$ which contains $\mathcal{P}'$, such that each $P \in \mathcal{P}'$ has degree 1 and each $j \in [k]$ has degree at most $|\mathcal{P}'|/\eta k \le \lambda n/8\eta k$. We can thus extend φ by setting $\varphi(v_2^P) = j$ for each edge Pj in B'_2. We proceed similarly to define $\varphi(v_6^P)$ for all $P \in \mathcal{P}'$ (i.e., define an auxiliary bipartite graph B_6, apply Lemma 4.6 and then set $\varphi(v_6^P)$ according to the subgraph B'_6 thus obtained). Let $P_{2,6} := \bigcup_{P \in \mathcal{P}'}\{v_2^P, v_6^P\}$, and note that for each $i \in [k]$ we have

$$\left|\varphi^{-1}(i) \cap P_{2,6}\right| \leq \frac{\lambda n}{4\eta k}. \tag{4.11}$$

Claim 4.19. For each $i \in [k]$ we have $\left|\varphi^{-1}(i)\right| = |T_1| \pm 6\lambda|T_2|/\eta k$.

Proof. Let $P_X := \{\, v_x^P \in P : x \in X \text{ and } P \in \mathcal{P}' \,\}$; we shall abuse notation, writing P_i for $P_{\{i\}}$ and $P_{i,j}$ for $P_{\{i,j\}}$. We partition the vertices of T_1 into five sets: $\mathcal{A} := \{A, P_7, P_{2,6}, P_{3,5}, P_4\}$, where $A := P_1 \,\dot\cup\, (T_1' \cap T)$ (so, in other words, A consists of all vertices of T_1 except for v_i^P such that $i \in \{2, 3, \ldots, 7\}$ and $P \in \mathcal{P}'$). For each $i \in [k]$, and each $S \in \mathcal{A}$, let $m(S) := \min_{j \in [k]}\left|\varphi^{-1}(j) \cap S\right|$, let $\delta_i(S) := |\varphi^{-1}(i) \cap S| - m(S)$ and $\delta_i := \sum_{S \in \mathcal{A}} \delta_i(S)$; so. Since $|S| \geq \lambda n/8$ for all $S \in \mathcal{A}$, we have

$$\begin{aligned} 0 \leq \delta_i &\leq \underbrace{\frac{2|T_1'|}{k \log\log|T_1'|} + \frac{2\lambda n}{k \log\log(\lambda n)}}_{A,\, P_7\text{: Lemma 4.15}} + \overbrace{\frac{\lambda n}{4\eta k} + \frac{2\varepsilon n}{\alpha k} + \frac{\lambda n}{2\eta k}}^{P_{2,6},\, P_{3,5}\text{: (4.11) + (4.9) + (4.10)}} + \underbrace{\frac{\lambda n}{2\eta k}}_{P_4\text{: (4.10)}} \\ &\leq \frac{2\lambda n}{\eta k} \leq \frac{6\lambda|T_2|}{\eta k}. \end{aligned} \tag{4.12}$$

Let us go through the calculations above. We distribute vertices of A roughly uniformly by Lemma 4.15 (d), since φ is an allocation of $T[A]$ obtained by the randomised allocation procedure; moreover, also by Lemma 4.15 (d), the vertices in P_1 are allocated quite evenly (recall that $P_1 \subseteq A$), and so are the vertices in P_7 (since their allocation is the same as the vertices in P_1). We have already calculated the 'unevenness' in the allocation of $P_{2,6}$ in (4.11). Things are a little more complicated for the allocation of the vertices in $P_{3,5}$: their allocation may have larger deviations from the uniform distribution and moreover these deviations depend on whether the vertices lie in a path in $\mathcal{P}^0$ (calculated in (4.9)) or $\mathcal{P}^\diamond$ (calculated in (4.10)) or $\mathcal{P}^H$ (no error because these are allocated symmetrically along H). Finally, vertices in P_4 are distributed evenly if they come from paths in $\mathcal{P}^0 \cup \mathcal{P}^H$; otherwise the error is given by (4.10) as well. This completes the proof of the claim since $\left|\varphi^{-1}(i)\right| = |T_1| \pm \delta_i$. ◇

▸ **Phase 2**. We now define φ for the remaining vertices of T. We first find a homomorphism ϱ_2 from T_2 to a weighted blow-up of $R^\star\big[[k]\big]$. For each $i \in [k]$, define

$$\delta_i := \left|\varphi^{-1}(i) \cap V(T_1)\right| - \min_{j \in [k]} \left|\varphi^{-1}(j) \cap V(T_1)\right|$$

(as in the claim above) and

$$f_i \coloneqq \begin{cases} |T_2|/k - \delta_i & \text{if } i \neq 1, \\ |T_2|/k - \delta_i + 1 & \text{otherwise,} \end{cases}$$

so $f_i = |T_2|(1 \pm 6\lambda/\eta)/k$ by (4.12). Note that f_i is precisely the number of vertices of T_2 that φ should map to cluster i if we are to satisfy (vii). We use an auxiliary graph B which is a blow-up of $R^\star\Big[\,[k]\,\Big]$, where each $i \in [k]$ is replaced by a set B_i, and $x \to y \in E(B)$ for all $x \in B_i$ and all $y \in B_j$ such that $i \to j \in E(R^\star)$. In fact, we will restrict our attention to a regular expander subgraph J^{blow} of B.

Choose numbers $b_1, \ldots, b_k$ as follows. Let $n' \coloneqq |T_2|$ and $\ell\!\ell\!\ell(n') \coloneqq \log\log\log n'$. For each $i \in [k]$, set $b_i \coloneqq f_i \ell\!\ell\!\ell(n')/n'$. Consider the digraph B which is a blow-up of $R^\star\Big[\,[k]\,\Big]$, as described above, where vertex i is replaced by a set B_i of b_i vertices. Note that $|B| = \ell\!\ell\!\ell(n')$ and that $f_i \geq (1 - 6\lambda/\eta)n'/k$, so

$$\delta^0(B) \geq \delta^0\Big(R^\star\Big[\,[k]\,\Big]\Big) k \frac{f_i \ell\!\ell\!\ell(n')}{n'} \geq \left(\frac{1}{2} + \eta\right) k \frac{(1 - 6\lambda/\eta)\ell\!\ell\!\ell(n')}{k} \geq \frac{|B|}{2}\left(1 + \frac{\eta}{2}\right),$$

and therefore B contains an expanding regular subdigraph J^{blow} by Lemma 4.10.

We allocate vertices of T_2 to vertices of J^{blow} as follows. Choose a vertex x_r in $B_{\varphi(r)}$ (where r is the root of T_2, the unique vertex in $T_1 \cap T_2$, and also the unique vertex of T_2 for which φ has been defined). Recall that we have an ancestral order $\prec$ of T_2; we apply the Allocation Algorithm 4.14, which produces a homomorphism $\varrho_2 : T_2 \to J^{\text{blow}}$ mapping approximately the same number of vertices to each $x \in J^{\text{blow}}$ and such that $\varrho_2(r) = x_r$. Indeed, by Lemma 4.15, it follows that the number of vertices allocated to each vertex of J^{blow} is $n'\big(\frac{1}{|J^{\text{blow}}|} \pm \frac{1}{\log\log n'}\big)$. Since $|J^{\text{blow}}| = |B| = \ell\!\ell\!\ell(n')$, the number of vertices allocated to B_i is

$$b_i \cdot \frac{n'}{\ell\!\ell\!\ell(n')}\left(1 \pm \frac{\ell\!\ell\!\ell(n')}{\log\log n'}\right) = f_i\left(1 \pm \frac{\ell\!\ell\!\ell(n')}{\log\log n'}\right).$$

Using ϱ_2 we define a homomorphism $\varphi' : T_2 \to R^\star\Big[\,[k]\,\Big]$ as follows: for each $x \in T_2$, set $\varphi'(x) = i$ if $\varrho_2(x) \in B_i$. Notice that $\varphi(r) = \varphi'(r)$, and thus, setting $\varphi(x) = \varphi'(x)$ for all $x \in T_2$ we obtain the desired extension of φ.

▸ **Phase 3**. To conclude the proof we slightly modify φ so as to satisfy (vii), by changing $\varphi(v_4^P)$ for $P \in \mathcal{P}^\diamond$ while preserving the other properties. Let $\Delta_{\min} \coloneqq \min_{i\in[k]}\left|\varphi^{-1}(i)\right|$ and for each $i \in [k]$ let $\Delta_i \coloneqq \left|\varphi^{-1}(i)\right| - \Delta_{\min}$. We proceed greedily, as follows. Suppose that $i, j \in [k]$ maximise $\left|\varphi^{-1}(j)\right| - \left|\varphi^{-1}(i)\right| > 0$. Choose a P_{ref}-diamond path $\left(x_s \prec^{y_s}_{w_s} \succ z_s\right)_{s=1}^{s^\star}$ in $J^\diamond$ connecting i and j, and for each $s \in \{1, i+1, \ldots, s^\star\}$ select a path $P \in \mathcal{P}^\diamond$ whose middle $v_s^x v_s^w v_s^z$ is $\prec$-isomorphic to P_{ref} and such that $\varphi(v_s^x) = x_s$, $\varphi(v_s^w) = w_s$ and $\varphi(v_s^z) = z_s$. We change φ

so that $\varphi(v_s^w) = y_s$ along this diamond path. This decreases $\sum_i \Delta_i$ by 2 and changes the allocation of at most k paths in $\mathcal{P}^\diamond$. Furthermore, since at the start $\Delta_i \le n\ell\!\ell(n')/k \log\log n'$ for all $i \in [k]$, we can reduce all Δ_i to 0 in at most $\sum_i k\Delta_i/2 \le k^2 n\ell\!\ell(n')/\log\log n' \le \lambda n/25k$ iterations. Since $|\mathcal{P}_i^y|, |\mathcal{P}_i^w| \ge \lambda n/24k$ for all $i \in [k]$, these changes can be done, and the modified φ satisfies (vii).

It remains to show that $\Delta(\varphi) \le 4$. We first note we that T_1' and T_2 are allocated according to the allocation algorithm, so, considering only the restriction of φ to $T\Big[V(T_1') \cup V(T_2)\Big]$ we have that $\Delta(\varphi)$ is at most 3. This accounts for all edges of T except those of paths in $\mathcal{P}'$, so, to conclude, we consider the mapping of these bare paths. Let $P \in \mathcal{P}'$ and let $v_1 \prec \cdots \prec v_7$ be the vertices of P. Note that the φ-degree of any vertex is bounded above by their total degree, hence the interior vertices $v_2, \ldots, v_6$ have φ-degree at most 2 (because P is a bare path). As for v_1 and v_7, their φ-degree is at most 4, since their φ-degree in $T\Big[V(T_1') \cup V(T_2)\Big]$ is at most 3 but their neighbour in P may have been allocated to a different vertex of $R^\star$ in Phase 1; therefore (i) holds as well. $\square$

4.4.2 Proof of Lemma 4.20

We now describe how to modify the embedding algorithm of Section 4.3.4 and prove Lemma 4.20, which states that we can always embed T (a tree with polylogarithmic maximum degree and many bare paths of order 7) to G (a digraph with high semidegree). Roughly speaking, all sufficiently large digraphs with large semidegree admit a regular partition with a 'suitable reduced graph' $R^\star$ satisfying the hypothesis of Lemma 4.18. We apply Lemma 4.18 to obtain a special homomorphism $\varphi : T \to R^\star$, and then apply the embedding algorithm described in Section 4.3.4 (with some modifications), to obtain an embedding of almost all of T in G. Finally, we complete the embedding using perfect matchings.

Lemma 4.20. Suppose that $1/n \ll 1/C$ and that $1/n \ll \lambda \ll \alpha$. Let T be an oriented tree of order n and maximum degree $\Delta(T) \le (\log n)^C$, and let G be a digraph of order n with minimum semidegree $\delta^0(G) \ge \left(\frac{1}{2} + \alpha\right)n$. If T contains a collection $\mathcal{P}$ of λn vertex-disjoint bare paths of order 7, then G contains a (spanning) copy of T.

Here is a brief outline of the proof. We first use Lemma 4.2 to define an auxiliary graph $R^\star$ which satisfies all properties required by the allocation lemma. Next, we allocate vertices of T (Lemma 4.18). Before embedding the tree, we reserve small subsets in each cluster for dealing with exceptional vertices and for the final matching, and choose $v_1 \in G$ where the embedding will begin. We then apply a slightly modified version of the Embedding Algorithm (we skip centres of some paths and embed paths covering exceptional vertices in a different manner); this successfully embeds almost all of T following the chosen allocation (this is similar to the proof of Lemma 4.16). We complete the embedding with perfect matchings (Lemma 2.8).

Proof of Lemma 4.20. Let G be a digraph of order n such that $\delta^0(G) \geq (1/2+\alpha)n$. We introduce constants ε, d, k, η such that $\frac{1}{n} \ll \frac{1}{k} \ll \varepsilon \ll d \ll \lambda \ll \eta \ll \alpha$. We apply Lemma 4.2 to obtain a partition $V_0 \dot\cup V_1 \dot\cup \cdots \dot\cup V_k$ and a digraph $R^\star$ with $V(R^\star) = V_0 \dot\cup [k]$ such that

(a) $|V_0| < \varepsilon n$ and $m := |V_1| = \cdots = |V_k|$;

(b) For each $i \in [k]$ we have $G[V_{i-1} \to V_i]$ and $G[V_i \to V_{i+1}]$ are (d, ε)-super-regular;

(c) For all $i, j \in [k]$ we have $i \to j \in E(R^\star)$ precisely when $G[V_i \to V_j]$ is (d, ε)-regular.

(d) have $v \leftarrow i \in E(R^\star)$ precisely when $\deg^-(v, V_i) \geq (1/2 + \eta)m$, and $v \to i \in E(R^\star)$ precisely when $\deg^+(v, V_i) \geq (1/2 + \eta)m$

(e) For all $i \in [k]$ we have $\deg^0_{R^\star}(i, [k]) \geq (1/2 + \eta)k$; and

(f) For all $v \in V_0$ we have $\deg^0_{R^\star}(v, [k]) > \alpha k$.

Let $\mathcal{P}$ be a collection of λn vertex-disjoint bare paths of T with order 7 and let H be the Hamilton cycle $1 \to 2 \to \ldots \to k \to 1$ in $R^\star$. Note that $H, \mathcal{P}, R^\star$ and T satisfy the conditions of Lemma 4.18, and hence we may fix a tidy ancestral order $\prec$ of T, disjoint sets $\mathcal{P}^0, \mathcal{P}^H \subseteq \mathcal{P}$ and a homomorphism $\varphi : T \to R^\star$ with the following properties

(i) $\Delta(\varphi) \leq 4$;

(ii) $|\mathcal{P}^0| = |V_0|$ and for each $P \in \mathcal{P}^0$ the centre of P is mapped to V_0;

(iii) φ maps precisely one vertex of T to each $v \in V_0$;

(iv) φ maps at least $\lambda n/24k$ centres of paths in $\mathcal{P}^H$ to each $i \in [k]$;

(v) If $N := \{ N^-_T(x) \cup N^+_T(x) : \varphi(x) \in V_0 \}$ are the neighbours of vertices φ maps to V_0, then φ maps at most $2\varepsilon n/\alpha k$ vertices of N to each $i \in [k]$;

(vi) For each $P \in \mathcal{P}^H$, the restriction of φ to middle(P) is a homomorphism from middle(P) to H;

(vii) $\left|\varphi^{-1}(1)\right| = \left|\varphi^{-1}(2)\right| = \cdots = \left|\varphi^{-1}(k)\right|$.

Finally, before we embed T, we reserve some sets of vertices of G with good properties. We introduce a new constant γ, with $1/n \ll 1/k \ll \varepsilon \ll \gamma \ll d$.

Claim 4.21. There exist $v_1 \in V_1$ and, for each $i \in [k]$, disjoint sets $X_i, Y_i \subseteq V_i$ with $|X_i| = |Y_i| = \lambda m/50$ such that

(i) If $U \subseteq V_i$ with $|U| \geq \lambda m/24$, then $G[X_{i-1} \to U]$ and $G[U \to X_{i+1}]$ are both $\left(\frac{50\varepsilon}{\lambda}, \frac{d}{32}\right)$-super-regular;

(ii) For all $x \in \varphi^{-1}(V_0)$, if φ maps an inneighbour of x to V_j, then

$$\deg^-\Big(\varphi(x), Y_j\Big) \geq \lambda m/200\,;$$

(iii) For all $y \in \varphi^{-1}(V_0)$, if φ maps an outneighbour of y to V_j, then

$$\deg^+\Big(\varphi(y), Y_j\Big) \geq \lambda m/200\,;$$

(iv) For each $y \in C^-(t_1)$ and each $z \in C^+(t_1)$ we have

$$\deg^-(v_1, V_{\varphi(y)}) \geq \gamma m \quad \text{ and } \quad \deg^+(v_1, V_{\varphi(z)}) \geq \gamma m.$$

Proof. Choose $Z_i \subseteq V_i$ with $|Z_i| = \lambda m/25$ uniformly at random, independently for each $i \in [k]$; choose $X_i \subseteq Z_i$ with $|X_i| = \lambda m/50$ uniformly at random and independently of all other choices and let $Y_i := Z_i \setminus X_i$. We will show that with high probability these sets satisfy all required properties.

Let $i \in [k]$. Recall that $G[V_{i-1} \to V_i]$ and $G[V_i \to V_{i+1}]$ are both (d, ε)-super-regular, so for each $x \in V_{i-1}$ and each $y \in V_{i+1}$ we have that $\deg^+(x, X_i)$ and $\deg^-(y, X_i)$ are random variables with hypergeometric distribution and expectation at least $|X_i|(d-\varepsilon)/2 \geq \lambda dm/100$. By Lemma 2.14, the probability that any one of these random variables has value strictly less than $\lambda dm/200$ decreases exponentially with n. By a union bound (over $2n$ events) it follows that with probability $1 - \mathrm{o}(1)$ all these random variables have value at least $\lambda dm/200$, and thus (i) holds with probability $1 - \mathrm{o}(1)$.

Let $v \in V_0$, let $t := \varphi^{-1}(v)$ and let x be a neighbour of t in T. If $x \in N_T^-(t)$, then $\deg^-(v, V_{\varphi(x)}) \geq m/2$, so $\deg^-(v, Y_{\varphi(x)})$ is a random variable with hypergeometric distribution and expectation at least $|X_i|/2 = \lambda m/100$ and by Lemma 2.14 the probability that $\deg^-(v, Y_{\varphi(x)})$ is less than $\lambda m/200$ decreases exponentially with n. Similarly, if $x \in N_T^+(t)$, then the probability that $\deg^+(v, Y_{\varphi(x)})$ is less than $\lambda m/200$ decreases exponentially with n. Again by a union bound (ii) and (iii) both hold with probability $1 - \mathrm{o}(1)$.

Let $I \subseteq [k]$ be such that all children of t_1 are mapped to V_i with $i \in I$. Let S be the star consisting of t_1 and its children. Since $\Delta(\varphi) \leq 4$, we have that $|I| \leq 4$, and so by Lemma 4.4 (applied with S and new a constant β such that

$1/n \ll 1/k \ll \varepsilon \ll \gamma \ll \beta, d)$ there exists $v_1 \in V_1$ which satisfies (iv). $\diamond$

Returning to the proof of the lemma, we introduce a constant β (as above) with $1/n \ll 1/k \ll \varepsilon \ll \gamma \ll \beta, d \ll \lambda$. We next greedily embed almost all of T to G. Recall that the embedding algorithm processes the vertices of T in a tidy ancestral order, reserving good sets (as in Definition 4.3) for the children of each vertex. We would like to apply the embedding algorithm of Section 4.3.4, but Lemma 4.16 (which guarantees the success of the algorithm) cannot be immediately used, for two reasons: firstly, $|T| = |G|$, so there is no room to spare; secondly, the edges of $R^\star$ between $[k]$ and V_0 do not correspond to regular pairs. To take care of these issues, we slightly modify the algorithm so it successfully embeds all vertices of T (except centres of some bare paths) according to φ. The changes only affect how the algorithm processes a small set of bare paths of T: we do not embed the centres $c(P)$ of paths in $\mathcal{P}^H$ and also use a different procedure to embed the middle sections of paths $P \in \mathcal{P}^0$. More precisely, we apply the embedding algorithm with input $T, \prec, R^\star, \varphi, G \setminus \bigcup_{i\in[k]} X_i, \mathcal{V} \coloneqq \{\, V_i \setminus X_i : i \in R^\star \,\}, v_1, \gamma$ and β with the changes detailed below. For each $x \in \varphi^{-1}(V_0)$ we write $P^x \in \mathcal{P}^0$ for the bare path with centre x, and write $p_1^x \prec \cdots \prec p_7^x$ for the vertices of P^x, so $x = p_4^x$.

Step 1. For each $i \in [k]$ write Y_i^τ for the available vertices of Y_i, write V_i^τ for the available vertices in $V_i \setminus Y_i$, and change the definition of B^τ so that it now includes $Y_1^\tau \cup \cdots \cup Y_k^\tau$, i.e. let

$$B^\tau \coloneqq \{v_1, \ldots, v_{\tau-1}\} \cup Y_1^\tau \cup \cdots \cup Y_k^\tau \cup \bigcup_{t_s\colon t_s \text{ is open}} \Big(A_s^- \cup A_s^+\Big),$$

so for all $\tau \geq 1$ and all $i \in [k]$ we have $V_i^\tau \cap Y_i = \varnothing$.

Step 2. If t_τ is a centre of some $P \in \mathcal{P}^H$, then we skip it (rather than embedding it) and proceed to Step 3.

If t_τ is the 'second vertex' of a path $P^x \in \mathcal{P}^0$, that is, if $t_\tau = v_2^x$ for some $x \in \varphi^{-1}(V_0)$, then instead of steps (2.3) and (2.4) we embed all vertices of P^x at once, as follows. For simplicity, we suppose that P^x is a path with vertices $t_\tau = p_1^x, p_2^x, p_3^x, x = p_4^x, p_5^x, p_6^x, p_7^x$ with $p_1^x \prec \cdots \prec p_7^x$ and whose edges are directed from p_s^x to p_{s+1}^x for each $s \in [6]$; the argument proceeds similarly otherwise. When p_1^x was embedded—say to a vertex u_1—we reserved a good set $A^+_{p_2^x} \subseteq V_{\varphi(p_2^x)} \cap N^+(p_1^x)$ for p_2^x. Let $U_3 \coloneqq N^-\Big(\varphi(x)\Big) \cap Y^\tau_{\varphi(p_3^x)}$. Note that the only vertices embedded to $Y_{\varphi(p_3^x)}$ are p_3^y, p_5^y for some $P^y \in \mathcal{P}^0$, and there are at most $2\varepsilon m/\alpha \leq \lambda m/400$ of these by item (v) of Lemma 4.18. Hence, by (ii) and (iii), it follows that $|U_3| \geq \lambda m/400$. Since $A^+_{p_2^x}$ is $(\beta, \gamma, \varphi, m)$-good for $S_{v_2^x}$, we conclude that $A^+_{p_2^x}$ contains a vertex u_2 with high outdegree in U_3. Let u_3 be a vertex in $U_3 \cap N^+_G(u_2)$. Note that $u_1 \to u_2 \to u_3 \to u_4$ is an embedding of half of P^x. Let $U_5 := N^+\big(\varphi(x)\big) \cap Y^\tau_{\varphi(p_5^x)}$. Note that the only vertices embedded to $Y_{\varphi(p_5^x)}$ are p_3^y, p_4^y for some $P^y \in \mathcal{P}^0$, and there are at most $2\varepsilon m/\alpha \leq \lambda m/400$ of these by item (v) of Lemma 4.18. Hence, by (ii) and (iii), it follows that $|U_5| \geq \lambda m/400$.

We can embed the second half of P^x using the original embedding algorithm, iterating steps 2 and 3 for the remaining vertices of P^x. Briefly, we do the following. We reserve a set $A^+_{p_5^x} \subseteq U_5$ containing at most $2m^{3/4}$ vertices and which is $(\beta, \gamma, \varphi, m)$-good for $S_{p_5^x}$, and choose a vertex $u_5 \in A^+_{p_5^x}$ with at least γm outneighbours in $V^\tau_{\varphi(p_6^x)}$. Then, we reserve a set $A^+_{p_6^x} \subseteq V^\tau_{\varphi(p_6^x)}$ containing at most $2m^{3/4}$ vertices and which is $(\beta, \gamma, \varphi, m)$-good for $S_{p_6^x}$, and choose a vertex $u_6 \in A^+_{p_6^x}$ with at least γm outneighbours in $V^\tau_{\varphi(p_7^x)}$. Finally, we reserve a set $A^+_{p_7^x} \subseteq V^\tau_{\varphi(p_7^x)}$ containing at most $2m^{3/4}$ vertices and which is $(\beta, \gamma, \varphi, m)$-good for $S_{p_7^x}$, and choose a vertex $u_7 \in A^+_{p_7^x}$ such that for each $z \in C^-(p_7^x)$ we have that $\varphi(p_7^x)$ has at least γm inneighbours in $V_{\varphi(z)}$ and for each $w \in C^+(p_7^x)$ we have that $\varphi(p_7^x)$ has at least γm outneighbours in $V_{\varphi(w)}$. (We do not reserve any sets for the children of this vertex, as this will be done in Step 3.) We set $\varphi(p_i^x) := u_i$ for all $i \in [k]$ and note that this extends the embedding of T while embedding P^x.

Step 3. If in Step 2 we embedded a parent t_τ of a vertex $v \in T$ which is the centre of some $P \in \mathcal{P}^H$, then we reserve a set for the only *child* y of v (rather than the child of t_τ) as follows: if $y \in C^-(v)$ we choose a set $A^-_y \subseteq N^-_G(v) \cap V^\tau_{\varphi(x)}$ containing at most $2m^{3/4}$ vertices and which is $(\beta, \gamma, \varphi, m)$-good for S_y, and let A^-_τ be the union of these sets; if $y \in C^+(\tau)$ we choose a set $A^+_y \subseteq N^+_G(v) \cap V^\tau_{\varphi(y)}$ containing at most $2m^{3/4}$ vertices and which is $(\beta, \gamma, \varphi, m)$-good for S_y, and let A^+_τ be the union of these sets.

If in Step 2 we embedded a path $P^x \in \mathcal{P}^0$ then $t_\tau = v_2^x$ and for each $z \in C^-(p_7^x)$ we know that $\varphi(p_7^x)$ has at least γm inneighbours in $V_{\varphi(z)}$, and for each $w \in C^+(p_7^x)$ we know that $\varphi(p_7^x)$ has at least γm outneighbours in $V_{\varphi(w)}$. We reserve good sets for the children of v_7^x as in the original algorithm (i.e., proceed as the original algorithm would if t_τ was v_7^x).

To prove that this procedure works, let $F := \{\, c(P) : P \in \mathcal{P}^H \,\}$ be the centres of paths in $\mathcal{P}^H$ and recall that $|\varphi(F) \cap V_i| = |\mathcal{P}^H|/k$ for all $i \in [k]$. These are the sole vertices which are not embedded by the modified algorithm. Let $T^\star$ be the tree we obtain from T by contracting the 3 vertices in the middle section of each path in $\mathcal{P}^H$ and the vertices $p_2^x, \dots, p_7^x$ of each path $P^x \in \mathcal{P}^0$ so that each bare path induced by those vertices becomes a single vertex. We now argue that all other vertices of T are successfully embedded by this modified version of the embedding algorithm. Since the paths $P^x \in \mathcal{P}^0$ are bare, the number of open vertices at any step is not greater than the number of open vertices we would have by applying the original algorithm to $T^\star$, so (as in the proof of Lemma 4.16) the number of vertices in reserved sets at each time τ is at most $(2m^{3/4})(\log_2 n)\Delta(T) \leq \varepsilon m$. Note also that since we never embed a vertex of F it follows that for all $\tau \geq 1$ and all $i \in [k]$ we have $|V_i^\tau| \geq |\mathcal{P}^H|/k - |X_i| - |Y_i| \geq \lambda m/50$ vertices in V_i which are available for the embedding (recall that V_i^τ is the set of vertices which were neither used nor reserved for children of open vertices). Finally, because each edge $i \to j$ of $R^\star\big[[k]\big]$ corresponds to an (ε', d')-regular pair (V_i, V_j) in G, we can indeed reserve sets in Step 3 as required (this follows from essentially the same argument we used in the proof of Lemma 4.16).

At this point, every vertex has been embedded according to φ, and the only vertices yet to be embedded are the centres of paths in $\mathcal{P}^H$, i.e., the vertices in F. For each $i \in [k]$, let M_i be the unused vertices in V_i (so M_i contains X_i); let U_i be the set of vertices to which we embedded the parents of vertices in F, and let W_i be the set of vertices to which we embedded the children of all vertices in F. By Claim 4.21 (i) we have that (M_{i-1}, U_i), (M_{i+1}, W_i), (U_i, M_{i+1}) and (W_i, M_{i+1}) are super-regular, and by items (vi) and (iv) of Lemma 4.18 we have that $|U_{i-1}| = |M_i| = |W_{i+1}|$. By Lemma 2.8, there exists a perfect matching of edges oriented from U_{i-1} to M_i and another perfect matching of edges oriented from M_i to W_{i+1}. These matchings complete the embedding of T to G. □

4.5 Trees with many leaves

This section has the same structure as the previous one. The main difference here is that leaf-edges will take the place of middle-sections of bare-paths: we use them to ensure the embedding covers exceptional vertices, as well as to set the matchings at the end of the embedding.

Let T be an oriented tree with polylogarithmic maximum degree and many vertex disjoint leaf-edges and let G be a digraph with $\delta^0(G) \geq (1/2 + \alpha)n$, where $|T| = |G| = n$ and $1/n \ll \alpha$. The main result of this section is Lemma 4.23, which states that Theorem 1.13 holds for such T. To prove this we also prove an auxiliary lemma, Lemma 4.22, which states that there exists a 'suitable' allocation of T to an extended reduced graph of G.

4.5.1 Allocation

A *leaf-edge* of an oriented tree T is an edge containing a leaf vertex. leaf-edge

Lemma 4.22. Suppose that $1/n \ll 1/k \ll \varepsilon \ll \lambda \ll \alpha$ and that $1/n \ll C$. Let $R^\star$ be a digraph with vertex set $V_0 \dot\cup [k]$, where $|V_0| < \varepsilon n$ and $n - |V_0| \equiv 0 \mod k$, and such that for all $i \in [k]$ and all $v \in V_0$ we have $\deg^0(i,[k]) \geq (1/2+\alpha)k$ and $\deg^0(v,[k]) \geq \alpha k$. Also, suppose that H is a cycle $1 \to 2 \to \cdots \to k \to 1$ in $R^\star\big[[k]\big]$. If T is an oriented tree of order n with $\Delta(T) \leq (\log n)^C$ and at least λn vertex-disjoint leaf-edges, then there exists a homomorphism $\varphi : T \to R^\star$ and a collection $\mathcal{E}$ of vertex-disjoint leaf-edges of T such that the following hold.

(i) Either all edges in $\mathcal{E}$ are oriented towards the leaf vertex, or all edges in $\mathcal{E}$ are oriented away from the leaf vertex.

(ii) φ maps precisely one leaf of T to each $v \in V_0$;

(iii) φ maps at least $\lambda n/32k$ leaf edges in $\mathcal{E}$ to each edge of H; and

(iv) $\left|\varphi^{-1}(1)\right| = \left|\varphi^{-1}(2)\right| = \cdots = \left|\varphi^{-1}(k)\right|$.

Proof. We assume that T contains at least $\lambda n/2$ vertex-disjoint leaf-edges which are oriented towards their leaf-vertex—we call these *out-leaf-edges*—the proof is symmetric otherwise. out-leaf-edges

Apply Lemma 2.3 to obtain a partition $\{T_1, T_2\}$ of T such that $|T_1|, |T_2| \geq \lambda n/6$ and such that $V(T_1)$ contains a set $E \subseteq E(T_1)$ of at least $\lambda n/7$ vertex-disjoint out-leaf-edges of T. (To do so, consider the vertex-disjoint leaf-edges of T and let P be the collection of non-leaves of T in those edges; apply Lemma 2.3 to T and P.) Let r be the intersection of T_1 and T_2. We can assume that $|T_1| \leq |T_2|$. Let E^1 and E^2 be disjoint subsets of E with $|E^1| = |E^2| = \lambda n/15$; for each $j \in \{1,2\}$, let P^j be the set of parents of leaves of T in E^j, and let L^j be the set of leaves of T in E^j. Let $T' = T_1 \setminus L^1$.

By Lemma 4.8, $R^\star\big[[k]\big]$ contains a subgraph J^s such that $\Delta(J^s) \leq 8/\alpha$ and such that for all $i \in [k-1]$ there exists $v_{ij} \in [k]$ with $i, i+1 \in N^+_{R^\star}(v_{ij})$. By Lemma 4.10, $R^\star\big[[k]\big]$ contains a spanning regular expander subgraph J which contains J^s as a subgraph and such that $\Delta(J) \leq 25n^{2/3}/\alpha$.

Apply the allocation algorithm to T', J and $x_1 = 1$; and let $\varphi_1 : T' \to J$ be the allocation it generates. By Lemma 4.15 (e) (applied with $2C$ in place of C and $|L_2|/n$ as β), at least $|L_2|\alpha/100k^{5/3} \geq n/k^2$ leaf-edges are mapped to each edge of J^s; moreover, by Lemma 4.15 (d) (applied with $2C$ in place of C and $\zeta = 1/3$), for each $i \in [k]$ at least $|L_1|/2k$ parents of leaves in L_1 are mapped to i. Let $P_i := \varphi_1^{-1}(V_i) \cap P^1$ be the set of parents of leaves in L_1 which are mapped to i. By Lemma 4.15, for all $i \in [k]$ we have that

$$\left|\varphi_1^{-1}(i)\right| = |T'| \left(\frac{1}{k} + \frac{1}{\log\log|T'|}\right). \tag{4.13}$$

We next extend φ_1, defining the allocation of some leaves in L^1 so that they are mapped to V_0 bijectively. First, note that by Lemma 4.6 there exists a bipartite subgraph B_0 of $R^\star$ with vertex classes $[k]$ and V_0 such that for all $i \in [k]$ and all $v \in V_0$ we have $\deg^-_{B_0}(v, [i]) = 1$ and $\deg^+_{B_0}(i, V_0) \leq \varepsilon m/2\alpha$. For each edge vj of B_0, where $v \in V_0$ and $j \in [k]$, we proceed as follows: let $p \in P_j$ be a parent of a leaf-edge $e \in E^1$ such that the leaf x of T in e has yet to be allocated; we set $\varphi_1(x) = v$.

At this point, all vertices of T_1 except for some leaves in L^1 have been allocated. We shall allocate the remaining out-leaf-edges of T_1 along H. Note that, for each $i \in [k]$, there are at least $|L_1|/2k - \varepsilon m/2\alpha \geq \lambda m/32$ vertices $p \in P_i$ in edges $p{\rightarrow}y \in L^1$ such that $\varphi_1(y)$ has yet to be defined; we set $\varphi_1(y) = i + 1$ for all such y (so $p{\rightarrow}y$ is allocated along an edge of H). Note that $\varphi_1 : T_1 \to R^\star$ is a homomorphism and that (i)–(iii) hold for φ_1.

We now define an allocation $\varphi_2 : T_2 \to R^\star$; our goal will be to combine φ_1 and φ_2 to form the desired allocation φ. We first obtain an allocation ϱ_2 from T_2 to a weighted blow-up of $R^\star\big[[k]\big]$ (as in the proof of Lemma 4.18). Let $\mu := \min_{i\in[k]}\left|\varphi_1^{-1}(i)\right|$ and, for each $i \in [k]$, let $\delta_i := \left|\varphi_1^{-1}(i)\right| - \mu$, so

$$0 \leq \delta_i \leq \frac{2|T'|}{k\log\log|T'|} + \frac{\varepsilon m}{2\alpha} \leq \frac{\lambda n}{2k} \leq \frac{\lambda|T_2|}{2k}. \tag{4.14}$$

For each $i \in [k]$, define

$$f_i := \begin{cases} |T_2|/k - \delta_i & \text{if } i \neq 1, \\ |T_2|/k - \delta_i + 1 & \text{otherwise,} \end{cases}$$

so $f_i = |T_2|(1 \pm \lambda/2)/k$ by (4.14). Note that f_i is precisely the number of vertices of T_2 that we should map to cluster i if φ_1 together with φ_2 are to satisfy (iv). We use an auxiliary graph B which is a blow-up of $R^\star\big[\,[k]\,\big]$, where each $i \in [k]$ is replaced by a set B_i, and $x{\rightarrow}y \in E(B)$ for all $x \in B_i$ and all $y \in B_j$ such that $i{\rightarrow}j \in E(R^\star)$. In fact, we will restrict our attention to a regular expander subgraph J^{blow} of B.

Choose numbers $b_1, \dots b_k$ as follows. Let $n' := |T_2|$ and $\ell\!\ell\!\ell(n') := \log\log\log n'$. For each $i \in [k]$, set $b_i := f_i \ell\!\ell\!\ell(n')/n'$. Consider the digraph B which is a blow-up of $R^\star\big[\,[k]\,\big]$, as described above, where vertex i is replaced by a set B_i of b_i vertices. Note that $|B| = \ell\!\ell\!\ell(n')$ and that $f_i \geq (1-\lambda/2)n'/k$, so

$$\delta^0(B) \geq \delta^0\Big(R^\star\big[\,[k]\,\big]\Big) k \frac{f_i \ell\!\ell\!\ell(n')}{n'} \geq \left(\frac{1}{2}+\eta\right) k \frac{(1-6\lambda/\eta)\ell\!\ell\!\ell(n')}{k} \geq \frac{|B|}{2}\left(1+\frac{\eta}{2}\right),$$

and therefore B contains an expanding regular subdigraph J^{blow} by Lemma 4.10.

We allocate vertices of T_2 to vertices of J^{blow} as follows. Choose a vertex x_r in $B_{\varphi(r)}$ (where r is the root of T_2, the unique vertex in $T_1 \cap T_2$, and also the unique vertex of T_2 for which φ has been defined). Recall that we have an ancestral order $\prec$ of T_2; we apply the Allocation Algorithm 4.14, which produces a homomorphism $\varrho_2 : T_2 \to J^{\text{blow}}$ mapping approximately the same number of vertices to each $x \in J^{\text{blow}}$ and such that $\varrho_2(r) = x_r$. Indeed, by Lemma 4.15 (applied with $2C$ in place of C), it follows that the number of vertices allocated to each vertex of J^{blow} is $n'\big(\frac{1}{|J^{\text{blow}}|} \pm \frac{1}{\log\log n}\big)$. Since $|J^{\text{blow}}| = |B| = \ell\!\ell\!\ell(n')$, the number of vertices allocated to B_i is

$$b_i \cdot \frac{n'}{\ell\!\ell\!\ell(n')}\left(1 \pm \frac{\ell\!\ell\!\ell(n')}{\log\log n}\right) = f_i\left(1 \pm \frac{\ell\!\ell\!\ell(n')}{\log\log n}\right).$$

Using ϱ_2 we define a homomorphism $\varphi_2 : T_2 \to R^\star\big[\,[k]\,\big]$ as follows: for each $x \in T_2$, set $\varphi_2(x) = i$ if $\varrho_2(x) \in B_i$. Note that $\varphi_1(r) = \varphi_2(r)$, and thus, setting $\varphi(x) = \varphi_1(x)$ for all $x \in T_1$ and $\varphi(y) = \varphi_2(y)$ for all $y \in T_2$ we obtain a homomorphism $\varphi : T \to R^\star$ which satisfies (i)–(iii).

The only property φ still might lack is (iv). Our final step is to adjust φ, changing the allocation of some leaf-edges mapped to edges of J^s, proceeding as we do in the proof of Lemma 4.9. For each $i \in [k]$, let $\gamma_i := \big|\varphi^{-1}(i)\big| - m$, so γ_i is positive if φ allocates too many vertices to i and negative if φ allocates too few vertices to i; in particular, all γ_i are zero if and only if the allocation satisfies (iv). Note that $\gamma_i \leq n/\log n$ and $\sum_{i\in[k]} \gamma_i = 0$.

We proceed greedily, as follows. Let $u, v \in [k]$ be such that $\gamma_v < 0 < \gamma_u$; let $(u_i, w_i, v_i)_{i=0}^t$ be a sequence of vertices in $[k]$, where $t < k$, such for all $i \in \{0, 1, \dots, t\}$ we have that $w_i \to u_i, w_i \to v_i \in E(J^s)$, and $u = u_0, v = v_t$ and also that $v_{i-1} = u_i$ for all $j \in [t]$ For each $j \in [t]$, select an out-leaf-edge in T which is mapped to $w_i \to u_i$, and modify the mapping of this path so that it is now mapped to $w_i \to v_i$. The modified map $\hat\varphi$ we obtain is such that $|\hat\varphi^{-1}(u)| = |\varphi^{-1}(u)| - 1$ and $|\hat\varphi^{-1}(v)| = |\varphi^{-1}(v)| + 1$, whereas $|\hat\varphi^{-1}(x)| = |\varphi^{-1}(x)|$ for all $x \in R^\star \setminus \{u, v\}$. Note that this procedure reduces by at most 1 the number of out-leaf-edges mapped to each edge of J^s. Hence, by iterating this procedure at most $\sum_v |\delta_v| \leq kn/\log n$

times, we can 'shift weights' as needed to obtain the desired mapping φ. (Note that it is indeed possible to carry out these steps, because each edge of J^s has at least $\eta n/k^3$ out-leaf-edges allocated to it.) After these changes, φ satisfies (iv) and still satisfies (i)–(iii). $\square$

4.5.2 Proof of Lemma 4.23

In this section we describe how to modify the embedding algorithm of Section 4.3.4 so that it successfully embeds a spanning tree T with many vertex-disjoint leaf-edges to a digraph G of high semidegree (thus proving Lemma 4.23).

Lemma 4.23. Suppose that $\frac{1}{n} \ll \lambda \ll \alpha$ and that $\frac{1}{n} \ll \frac{1}{C}$. If T is an oriented tree of order n with $\Delta(T) \leq (\log n)^C$ and at least λn vertex-disjoint leaf-edges, and if G is a digraph of order n with minimum semidegree $(\frac{1}{2} + \alpha)n$ then G contains a (spanning) copy of T.

Proof. As in the proof of Lemma 4.20, we begin by constructing a suitable regular partition of G. We introduce constants ε, d, η with $1/n \ll 1/k \ll \varepsilon \ll d \ll \lambda \ll \eta \ll \alpha$ and apply Lemma 4.2 to obtain a partition $V_0 \mathbin{\dot\cup} V_1 \mathbin{\dot\cup} \cdots \mathbin{\dot\cup} V_k$ and a digraph $R^\star$ with $V(R^\star) = V_0 \mathbin{\dot\cup} [k]$ such that

(a) $|V_0| < \varepsilon n$ and $m := |V_1| = \cdots = |V_k|$;

(b) For each $i \in [k]$ we have $G[V_{i-1} \to V_i]$ and $G[V_i \to V_{i+1}]$ are (d, ε)-super-regular;

(c) For all $i, j \in [k]$ we have $i \to j \in E(R^\star)$ precisely when $G[V_i \to V_j]$ is (d, ε)-regular.

(d) For all $v \in V_0$ and all $i \in [k]$ we have $v \leftarrow i \in E(R^\star)$ precisely when $\deg^-(v, V_i) \geq (1/2 + \eta)m$, and $v \to i \in E(R^\star)$ precisely when $\deg^+(v, V_i) \geq (1/2 + \eta)m$;

(e) For all $i \in [k]$ we have $\deg^0_{R^\star}(i, [k]) \geq (1/2 + \eta)k$; and

(f) For all $v \in V_0$ we have $\deg^0_{R^\star}(v, [k]) > \alpha k$.

Let $H \subseteq R^\star$ be the directed cycle $1 \to 2 \to \cdots \to k \to 1$. Note that H, T and $R^\star$ satisfy the conditions of Lemma 4.22 (applied taking the value of η for α there, with remaining constants as here), so there exists an allocation φ of the vertices of T such that

(i) Either all edges in $\mathcal{E}$ are oriented towards the leaf vertex, or all edges in $\mathcal{E}$ are oriented away from the leaf vertex.

(ii) φ maps precisely one leaf of T to each to $v \in V_0$;

(iii) φ maps at least $\lambda n/32k$ leaf edges in $\mathcal{E}$ to each edge of H; and

(iv) $\left|\varphi^{-1}(1)\right| = \left|\varphi^{-1}(2)\right| = \cdots = \left|\varphi^{-1}(k)\right|$.

We assume that all edges in $\mathcal{E}$ are oriented towards the leaf vertex; the proof is similar otherwise. Let $L^0 \coloneqq \varphi^{-1}(V_0)$ be the set of leaves which are mapped to V_0, and let P^0 be the set of parents of those leaves, so $|V_0| = |L^0|$. For each $i \in [k]$, let L_i be a set containing precisely $\lambda m/32$ leaves mapped to V_i whose parents have been mapped to V_{i-1}, let P_i be the set of parents of L_i, and let $L^H = \bigcup_{i\in[k]} L_i$.

We embed $T' \coloneqq T \setminus L^H$ to G by applying a (slightly modified) version of the vertex embedding algorithm. Before doing so, we reserve some sets of vertices of G which have good properties and which we will use to complete the embedding later on. We introduce a new constant γ with $1/n \ll 1/k \ll \varepsilon \ll \gamma \ll d$. For each $i \in [k]$ reserve sets X_i (for the final matching) and Y_i for parents of whose leaves will be embedded to V_0.

Claim 4.24. There exist $v_1 \in V_1$ and, for each $i \in [k]$, disjoint sets $X_i, Y_i \subseteq V_i$ with $|X_i| = |Y_i| = \lambda m/100$ such that

(i) If $U \subseteq V_i$ with $|U| \geq \lambda m/24$, then $G[X_{i-1} \to U]$ and $G[U \to X_{i+1}]$ are both $\left(\frac{100\varepsilon}{\lambda}, \frac{d}{32}\right)$-super-regular;

(ii) For all $x \in \varphi^{-1}(V_0)$, if φ maps an inneighbour of x to V_j, then
$$\deg^-\left(\varphi(x), Y_j\right) \geq \lambda m/400\,;$$

(iii) For all $y \in \varphi^{-1}(V_0)$, if φ maps an outneighbour of x to V_j, then
$$\deg^+\left(\varphi(y), Y_j\right) \geq \lambda m/400\,;$$

(iv) For each $y \in C^-(t_1)$ and each $z \in C^+(t_1)$ we have
$$\deg^-(v_1, V_{\varphi(y)}) \geq \gamma m \quad \text{and} \quad \deg^+(v_1, V_{\varphi(z)}) \geq \gamma m.$$

Proof. This Claim (and its proof) are almost identical to Claim 4.21, so the proof is omitted. ◇

Returning to the proof of the lemma, (as in the proof of Lemma 4.20) we introduce a constant β (as above) with $1/n \ll 1/k \ll \varepsilon \ll \gamma \ll \beta, d$. We next apply the embedding algorithm to $T' \coloneqq T \setminus L^H$ to allocate T' to $G \setminus \bigcup_{i\in[k]} X_i$ (being careful when close to V_0 to embed in the sets Y_i all parents of leaves mapped to V_0). Roughly speaking, as in Lemma 4.20, we embed each vertex of

T' to $G' \coloneqq G \setminus \bigcup_{i\in[k]} X_i$ according to the allocation φ as dictated by the vertex embedding algorithm, except for the leaves in L^0 and their parents in P^0, which we embed to V_0 and $\bigcup_{i\in[k]} Y_i$, respectively. More precisely, we apply the following changes to the embedding algorithm.

Step 1. For each $i \in [k]$ write Y_i^τ for the available vertices of Y_i, write V_i^τ for the available vertices in $V_i \setminus (X_i \mathbin{\dot\cup} Y_i)$ and change the definition of B^τ so that it now includes $Y_1^\tau \cup \cdots \cup Y_k^\tau$, i.e. let

$$B^\tau \coloneqq \{v_1, \ldots, v_{\tau-1}\} \cup Y_1^\tau \cup \cdots \cup Y_k^\tau \cup \bigcup_{t_s\colon t_s \text{ is open}} \left(A_s^- \cup A_s^+\right),$$

so for all $\tau \geq 1$ and all $i \in [k]$ we have $V_i^\tau \cap Y_i = \varnothing$.

Step 2. Nothing changes in this step.

Step 3. We only modify this step if t_τ is either a vertex in P^0 or a parent of such vertex, otherwise we proceed as in the original algorithm.

If in Step 2 we embedded $t_\tau \in P^0$ to a vertex v_τ, then t_τ is adjacent to a leaf $\ell \in L^0$ with $w^\ell \coloneqq \varphi(\ell) \in V_0$; moreover, t_τ was embedded to an inneighbour of w^ℓ in $Y^\tau_{\varphi(t_\tau)}$. We reserve a set $A_\ell^+ \coloneqq \{w^\ell\}$ for the child of t_τ, and let A_τ^- and A_τ^+ be the union of the sets reserved for the other children of t_τ, which we select as in Step 3 of the original algorithm.

If in Step 2 we embedded a vertex t_τ which is a parent of a vertex $p \in P^0$, we reserve sets for the other children of t_τ, as in the original algorithm, but reserve the set A_p^+ (or A_p^-) for p in a different manner, so that it is guaranteed to lie in $Y^\tau_{\varphi(p)} \cap N^-\big(\varphi(\ell)\big)$, where ℓ is the only leaf in L^0 connected to p in T: if $p \in C^-(t_\tau)$, choose a set $A^-_{t_\tau} \subseteq N^-_G(v_\tau) \cap Y^\tau_{\varphi(p)} \cap N^-\big(\varphi(\ell)\big)$ containing at most $2m^{3/4}$ vertices and which is $(\beta, \gamma, \varphi, m)$-good for S_p; and if $p \in C^+(t_\tau)$, choose a set $A^+_{t_\tau} \subseteq N^+_G(v_\tau) \cap Y^\tau_{\varphi(p)} \cap N^-\big(\varphi(\ell)\big)$ containing at most $2m^{3/4}$ vertices and which is $(\beta, \gamma, \varphi, m)$-good for S_p. (In either case we let A_τ^- and A_τ^+ be the union of the sets reserved for the children of t_τ, as in the original algorithm.)

If neither of these conditions apply, we follow Step 3 of the original algorithm.

We now argue that the modified embedding algorithm successfully embeds T' to $G \setminus (X_1 \cup \cdots \cup X_k)$ and that every vertex of P^0 is embedded to a vertex in $Y_1 \mathbin{\dot\cup} \cdots \mathbin{\dot\cup} Y_k$. Note first that every vertex is embedded according to φ (i.e., for all $x \in T$ we embed x to $\varphi(x)$ if $\varphi(x) \in V_0$, and embed x to a vertex in $V_{\varphi(x)}$ otherwise). Following the proof of Lemma 4.16 all that we need to prove is that the choices of the embedding algorithm can be carried out as required. Note that if the choices in Step 3 can be done, then all choices in Step 2 (all of which are made according to the original algorithm) can be made; as a consequence, we only need to consider what happens in Step 3. Recall that T' has none of the leaves in L^H and that for each $i \in [k]$ we have $|\varphi(L^H) \cap V_i| = |L^H|/k$; therefore, for all $\tau \geq 1$ and all $i \in [k]$ we have $|V_i^\tau| \geq |L_i| - |X_i| - |Y_i| \geq \lambda m/50$; moreover, for each $p \in P^0$ connected to a leaf $\ell \in L^0$, the number of vertices in reserved sets at any time τ is at most $2m^{3/4}(\log n)\Delta(T) \leq \varepsilon m \leq |Y^\tau_{\varphi(p)} \cap N^-\big(\varphi(\ell)\big)|/2$. When the algorithm reaches Step 3, we have just embedded a vertex $t_\tau \in T'$ to a vertex $v_\tau \in G \setminus \bigcup_{i\in[k]} X_i$. We consider the following 3 cases.

If $t_\tau \in P^0$, then t_τ is adjacent to a single leaf $\ell \in L^0$, and this leaf is mapped to a vertex $w^\ell := \varphi(\ell) \in V_0$. Since $\varphi^{-1}(w^\ell) = \{\ell\}$ and $v_\tau \to w^\ell \in E(G)$ we can reserve the desired set A_ℓ^+. Moreover, since $|V_i^\tau| \geq |L_i| - |X_i| - |Y_i| \geq \lambda m/50$ we can apply Lemma 4.4 to find and reserve good sets for each of the remaining children of t_τ.

Now suppose t_τ is a parent of a vertex $p \in P^0$. As before, p is adjacent to a single leaf $\ell \in L^0$, and this leaf is mapped to a vertex $w^\ell := \varphi(\ell) \in V_0$. We wish to reserve sets for the children of t_τ with the restriction that the set reserved for p should lie in $Y_i^\tau \cap N_G^-(w^\ell)$. Recall that the only vertices we ever embed to Y_i lie in P^0, so the number of vertices of Y_i unavailable for embedding is at most $|P^0| + 2m^{3/4}(\log_2 n)\Delta(T) \leq 2\varepsilon m$. By (ii), it follows that $|Y_i^\tau \cap N_G^-(w^\ell)| \geq \lambda m/400 - 2\varepsilon m \geq \gamma m$, so we can reserve a set for p which is good for S_p as required as well.

Lastly, if neither of the previous conditions holds, then we reserve sets for the children of t_τ as in the original algorithm; this can be done since

$$|V_i^\tau| \geq |L_i| - |X_i| - |Y_i| \geq \lambda m/50.$$

Since it is always possible to reserve the desired sets, we conclude that the modified embedding algorithms successfully embeds T' to $G \setminus \bigcup_{i\in[k]} X_i$.

Let P_{i-1} be the parents of L_i, so $|P_i| = |L_i|$ and, every vertex in P_{i-1} has been embedded to $i-1$. For each $i \in [k]$, let W_i be the set of vertices of $V_i \subseteq G$ to which no vertex has been embedded yet, and let U_i be the set of vertices to which the vertices in P_i have been embedded. Since $|W_i| = |L_i| = |U_{i-1}|$ and $X_i \subseteq L_i$ we have that there exists a perfect matching of edges directed from U_{i-1} to W_i by Claim cl:reserve-again (i) and Lemma 2.8. This completes the embedding of T to G. $\square$

4.6 Proof of Theorems 1.12, 1.13 and 1.15

Suppose that $1/n \ll \lambda \ll \alpha$ and let G be a digraph of order n with semidegree at least $(1/2+\alpha)n$. We conclude this chapter with the proofs of Theorems 1.13, 1.12 and 1.15 (see below). The two first are simple consequences of Lemmas 4.20 and 4.23.

We begin with Theorem 1.13, which states that G contains every spanning oriented tree T of order n with $\Delta(T) \leq (\log n)^C$ if T contains either λn vertex-disjoint bare paths of order 7 or λn vertex-disjoint edges incident to leaves.

Proof of Theorem 1.13. Suppose that $1/n \ll 1/C$ and that $1/n \ll \lambda \ll \alpha$. Let G be a digraph of order n with $\delta^0(G) \geq (1/2+\alpha)n$. Let T be an oriented tree of order n with $\Delta(T) \leq (\log n)^C$. If T contains λn vertex-disjoint bare paths of order 7, then G contains T by Lemma 4.20; otherwise, if T contains λn vertex-disjoint edges incident to leaves, then G contains T by Lemma 4.23. □

Theorem 1.12 asserts that if $1/n \ll \Delta$, then G contains every oriented tree T of order n with maximum degree at most Δ.

Proof of Theorem 1.12. Let T be an oriented tree of order n with $\Delta(T) \leq \Delta$ and let G be a digraph with $\delta^0(G) \geq (1/2+\alpha)n$. We introduce new constants λ and λ' with $1/n \ll \lambda \ll \lambda' \ll \alpha, \Delta$. If T contains at least $\lambda' n$ leaves, then T contains at least $\lambda' n/\Delta > \lambda n$ edge-disjoint leaf-edges, so $T \subseteq G$ by Lemma 4.23. Otherwise, by Lemma 4.1, T contains a bare path decomposition into at most $2\lambda' n$ paths. Let $x_1, \ldots, x_s$ be the lengths of these paths (i.e., the number of edges in each of them), so $x_1 + \cdots + x_s = n-1$. Then, for all $t > 0$, we have

$$\sum_{i=1}^{s} \left\lfloor \frac{x_i}{t} \right\rfloor \geq \sum_{i=1}^{s} \left(\frac{x_i}{t} - 1 \right) = \frac{n-1}{t} - s.$$

Choosing $t = 8$, it follows that T contains are at least $(n-1)/8 - 2\lambda' n \geq n/10$ bare paths of length 8 (and order 9). Therefore, T contains at least $n/10$ vertex-disjoint bare paths of order 7, so $T \subseteq G$ by Lemma 4.20. □

Theorem 1.15 states a sufficient condition which ensures that a tree-like digraph H is a spanning subgraph of every digraph with high semidegree. Its proof combines ideas from all previous sections.

4.6.1 Tree-like spanning subdigraphs

We conclude this chapter by discussing how the method we developed can be used to embed tree-like spanning digraphs. We note that the next theorem is stated so as to demonstrate how the ideas can be applied, and thus the bounds we obtain are not best possible, and the class of digraphs we embed is not the most general possible either. We discuss how the statement can be strengthened after sketching the proof.

Theorem 1.15. Suppose that $1/n \ll \alpha \ll C$. Let G be a digraph of order n with $\delta^0(G) \geq (1/2 + \alpha)n$. If Q is a subdivision of a graph Q_{under} such that

(i) $|Q_{\text{under}}| \leq (\log n)^C$;

(ii) each edge of Q_{under} has been subdivided at least $\log n$ times; and

(iii) $|Q| = n$;

then G contains every orientation of Q.

We include below only a proof sketch of Theorem 1.15 because it is quite similar to the proofs in the previous sections.

Proof sketch. We fix an orientation of Q and, abusing notation, write Q to denote this oriented subdivision.

As in the proofs of Lemmas 4.20 and 4.23, we begin by constructing a suitable regular partition of G. We introduce constants ε, d, η with $1/n \ll 1/k \ll \varepsilon \ll d \ll \lambda \ll \eta \ll \alpha$ and apply Lemma 4.2 to obtain a partition $V_0 \,\dot\cup\, V_1 \,\dot\cup \cdots \dot\cup\, V_k$ and a digraph $R^\star$ with $V(R^\star) = V_0 \,\dot\cup\, [k]$ such that

(a) $|V_0| < \varepsilon n$ and $m := |V_1| = \cdots = |V_k|$;

(b) For each $i \in [k]$ we have $G[V_{i-1} \rightarrow V_i]$ and $G[V_i \rightarrow V_{i+1}]$ are (d, ε)-super-regular;

(c) For all $i, j \in [k]$ we have $i \rightarrow j \in E(R^\star)$ precisely when $G[V_i \rightarrow V_j]$ is (d, ε)-regular.

(d) For all $v \in V_0$ and all $i \in [k]$ we have $v \leftarrow i \in E(R^\star)$ precisely when $\deg^-(v, V_i) \geq (1/2 + \eta)m$, and $v \rightarrow i \in E(R^\star)$ precisely when $\deg^+(v, V_i) \geq (1/2 + \eta)m$;

(e) For all $i \in [k]$ we have $\deg^0_{R^\star}(i, [k]) \geq (1/2 + \eta)k$; and

(f) For all $v \in V_0$ we have $\deg^0_{R^\star}(v, [k]) > \alpha k$.

Let $H \subseteq R^\star$ be the directed cycle $1 \rightarrow 2 \rightarrow \cdots \rightarrow k \rightarrow 1$.

Let $x_1, \ldots, x_q$ be the vertices of Q which 'correspond' to vertices in Q_{under}: that is, these are vertices which do not arise from the subdivision of the edges of Q_{under}. By relabelling if necessary we may assume without loss of generality that $V(Q_{\text{under}}) = \{x_1, \ldots, x_q\}$. For each edge $x_i x_j \in E(Q_{\text{under}})$ where $i < j$, write P_{ij} for the path from x_i to x_j in Q. (Note that this is an oriented path.) Let $\{P_{ij}^1, P_{ij}^2\}$ be a tree-partition of P_{ij} such that P_{ij}^1 contains x_i and P_{ij}^2 contains x_j, and moreover choose this tree-partition so that each path has at least $|P_{ij}|/2$ vertices and so that $|P_{ij}^1| \le |P_{ij}^2|$. Also, let c_{ij} be the sole vertex in $P_{ij}^1 \cap P_{ij}^2$. We let x_1 be the root of P_{ij} and P_{ij}^1, let c_{ij} be the root of P_{ij}^2, and let $\prec$ be the (unique) ancestral order of P_{ij}. Note that P_{ij}^1 contains a collection $\mathcal{P}_{ij}$ of at least $\lambda|P_{ij}|$ vertex-disjoint bare paths of order 7 whose middles are $\prec$-isomorphic, and moreover we may assume that none of these paths contains x_i. Note that for all distinct $e, e' \in E(Q_{\text{under}})$, for all $P \in \mathcal{P}_e$ and all $P' \in \mathcal{P}_{e'}$ the middle sections of P and P' are vertex-disjoint, and so we shall (again) abuse notation and say that these middle sections are $\prec$-isomorphic whenever they are isomorphic according to the respective ancestral orders $\prec_e$ and $\prec_{e'}$ of P_e and $P_{e'}$ respectively.

Claim 4.25. There exists a subset E_{good} of $E(Q_{\text{under}})$ such that $\mathcal{P}_{\text{good}}$, defined as $\mathcal{P}_{\text{good}} \coloneqq \bigcup_{e\in E_{\text{good}}} \mathcal{P}_e$ is a collection of at least $\lambda n/4$ vertex-disjoint bare paths of Q whose middles are pairwise $\prec$-isomorphic.

Proof. Let $\mathcal{P}' \coloneqq \bigcup_{e\in E(Q_{\text{under}})} \mathcal{P}_e$. Since this is a collection of vertex-disjoint bare paths of Q, we have $|\mathcal{P}'| \ge \sum_{e\in E(Q_{\text{under}})} \lambda|P_e| \ge \lambda n$. Note that 'being $\prec$-isomorphic' is an equivalence relation and that there are only four possible distinct orientations for the middle of any path in $\mathcal{P}'$; it follows (by averaging) that at least $\lambda n/4$ of these paths are pairwise $\prec$-isomorphic to some rooted path P_{ref} of order 3. We therefore define E_{good} to be the set of edges $e \in E(Q)$ such that $\mathcal{P}_e$ contains paths $\prec$-isomorphic to P_{ref}, completing the proof. $\diamond$

Returning to the proof of the lemma, let E_{good} and $\mathcal{P}_{\text{good}}$ be as in the claim, and fix a partition $\{V_0^e\}_{e\in E_{\text{good}}}$ of V_0 such that $|V_0^e| = |\mathcal{P}_e||V_0|/|\mathcal{P}_{\text{good}}|$, i.e., such that each part contains a number of vertices proportional to the number of paths $\mathcal{P}_e$ contributes to $\mathcal{P}_{\text{good}}$. Since $|\mathcal{P}_{\text{good}}| \ge \lambda n$ and $|V_0| < \varepsilon n$, we have that $|V_0^e| \le |\mathcal{P}_e|\varepsilon n/\lambda n \le \lambda|\mathcal{P}_e|/5$.

With this partition in place, our strategy is as follows. Firstly, allocate all vertices $x_1, \dots, x_q$ to V_1 and fix another cluster V_r such that $G[V_1{\to}V_r]$ and $G[V_1{\leftarrow}V_r]$ are both (ε, d)-regular. Secondly, allocate all paths of Q which correspond to 'bad' edges of Q_{under}, i.e., paths P_e for all $e \in E(Q_{\text{under}}) \setminus E_{\text{good}}$. These paths are allocated to $R^\star\big[[k]\big]$ using Algorithm 4.14, with the only restriction that $x_1, \dots, x_q$ are allocated to V_1 and any neighbour of a vertex in $\{x_1, \dots, x_q\}$ is allocated to V_r. Note that these paths are long (compared to k), and hence they are allocated quite evenly: more precisely, the total error in allocating all bad paths is at most $4n/\log\log n$ (by Lemma 4.15).

Then each good path P_e (where $e \in E_{\text{good}}$) is allocated following closely the four stages in the proof of Lemma 4.18. Begin by defining a (possibly uneven) allocation of P_e^1, as follows. We partition the $\prec$-isomorphic bare paths in $\mathcal{P}_e$ in three groups $\mathcal{P}_e^0, \mathcal{P}_e^H$ and $\mathcal{P}_e^\diamond$, each containing a fraction of the bare paths in $\mathcal{P}_e$. (Recall that all those paths lie in P_e^1.) Again as in Lemma 4.18, the centres of paths in $\mathcal{P}_e^0$ are used to cover exceptional vertices (more precisely, the exceptional vertices in V_0^e) so that the neighbours of these exceptional vertices are mapped to adequate vertices in $[k]$; moreover middles of paths in $\mathcal{P}_e^H$ are allocated along the cycle $1 \to \cdots \to k \to 1$ and the middles of paths in $\mathcal{P}_e^\diamond$ are allocated to diamonds in a diamond-connected subgraph $J^\diamond$ of $R^\star$. We complete the allocation of P_e^1 by first contracting the bare paths in $\mathcal{P}_e$ and then applying the allocation algorithm to the path arising from this contraction—this produces an almost even allocation, which we extend using the prefix and suffix of each path in $\mathcal{P}_e$ to obtain an allocation of P_e^1. This allocation of P_e^1 may be somewhat uneven (due to the 'jumping around' to reach middles of paths in $\mathcal{P}_e$ and to connect the path to V_r at the end). Crucially, we fix this imbalance by allocating P_e^2 to a weighted blow-up of $R^\star\big[[k]\big]$, which compensates for the uneven usage of clusters. By doing so we complete an allocation of the whole path P_e which is almost even, with error at most $4|P_e|/\log\log|P_e|$ (by Lemma 4.15 and an argument similar to the proofs of Lemmas 4.18).

After all paths of Q (bad and good) have been allocated, we have the following situation: $\varphi : Q \to R^\star$ is a homomorphism which maps precisely one vertex to each exceptional vertex, a linear number of $\prec$-isomorphic middles of paths is allocated along H and a linear umber of $\prec$-isomorphic middles of paths is allocated to the 'switching diamonds'. However, the allocation is not perfect, and some clusters may have too few or too many vertices allocated to them. Still, φ is 'close enough' to being perfect, as the total error in the allocation of the whole of Q is at most $8n/\log\log n$. We can therefore use the diamond-switching strategy we applied in the Lemma 4.18, modifying φ to obtain a 'perfect' allocation of Q to $R^\star$.

Again mimicking the approach in Section 4.4, we proceed to embed these paths, now following closely the proof of Lemma 4.20. The key difference is that rather than having a single root vertex for the embedding we now have the whole set $\{x_1, \ldots, x_q\}$. We choose vertices $v_1, \ldots, v_q$ in V_1 such that each one of these has very large semidegree in V_r, and reserve good sets of vertices in V_r for the neighbours of each x_i. We also reserve sets of vertices to help embed neighbours of vertices allocated to V_0 and some sets of vertices to be used in the final matching. Recall that many centres of $\prec$-isomorphic paths were embedded along H: let W_H

be the set of these centres. We apply the vertex embedding algorithm, skipping vertices in W_H and being careful to embed neighbours of vertices mapped to V_0 to the correct reserved sets; we are also careful when embedding neighbours of $x_1, \dots, x_q$: these should be embedded to reserved sets in V_r. Since we never embed vertices in W_H, the algorithm proceeds with enough 'room to spare' and hence the embedding succeeds: we embed each vertex according to the allocation, while avoiding sets reserved for the final matching. Our final step is to embed the vertices in W_H so that the required perfect matchings exist, completing the embedding of Q to G. $\triangle$

We note that a much larger class of spanning digraphs can be handled with a method similar to the one discussed above. For example, suppose that $1/n \ll \lambda \ll \alpha, C$, let G be a digraph of order n with $\delta^0(G) \geq (1/2+\alpha)n$ and let Q be a digraph of order n with $\Delta(Q) \leq (\log n)^C$. Suppose that there exists a small set S of $V(Q)$, say with $|S| \leq n^{1/3}$ and a component Q' of $Q - S$ such that

- $Q[S]$ has no edges,
- each component of $Q - S$ is a tree,
- for all $x, y \in S$ we have $\mathrm{dist}(x, y) \geq 7$,
- some component Q' of $Q - S$ contains at least $n^{1/2}$ vertices, and
- Q' contains either λn vertex-disjoint leaf edges or λn vertex-disjoint bare paths of order 7.

Remark. Note that the hypothesis of Theorem 1.15 imply all of the above. On the other hand, there are many ways in which this generalises Theorem 1.15: for instance, we may form Q (as in Theorem 1.15) from a graph Q_{under} with order at most $(\log n)^C$ and subdivide every edge at least 7 times (rather than $\log n$ times); another example is form Q from a forest F of trees with maximum degree at most $(\log n)^C$ which contains at least one tree satisfying the hypothesis of Theorem 1.13 and join these trees by at most $n^{1/3}$ vertex-disjoint paths of order at least 7 each. In either case, every orientation of Q can be shown to be contained in a digraph G of high semidegree by the argument we outline below.

Under these assumptions, we can follow a strategy very similar to the proof of Theorem 1.15 and conclude that Q is a spanning subgraph of G. Let $Q_1, \dots, Q_p$ be the components of $Q - S$, labelled so that $|Q_1| \leq \dots \leq |Q_p|$ and let s be such that $Q_s = Q'$. We introduce ε, d with $1/n \ll 1/k \ll \varepsilon \ll d \ll \lambda$ and apply the Regularity lemma (Lemma 4.2) to the host graph G to obtain a partition $V_0 \dot\cup V_1 \dot\cup \cdots \dot\cup V_k$ of $V(G)$, a reduced graph $R^\star$ with $V(R^\star) = V_0 \dot\cup [k]$, and a super-regular cycle $H \in R^\star\big[[k]\big]$ of clusters (so $G[V_i{\to}V_{i+1}]$ is (ε, d)-super-regular for all $i \in [k]$ with addition modulo k). Next, we define an allocation of $V(Q)$: the vertices in S are allocated to V_1 and some cluster V_r is chosen for the neighbours of vertices in S (this is possible since no short cycle of Q involves a vertex in S and also because the high semidegree of the reduced graph ensures that each vertex lies in a 2-cycle). Note that p (the number of components of $Q - S$) satisfies $p \leq \Delta(Q)|S|$, and hence at most $2\Delta(Q)|S|(\log n) \leq 2(\log n)^{C+1}n^{1/3}$ vertices of Q may lie in components Q_i of order less than $\log n$. Let j be the largest index such that $|Q_i| < \log n$ for all $i \leq j$. For each $i \leq j$ we allocate the vertices of Q_i greedily with the sole restriction that vertices in S are allocated to V_1 and their neighbours to V_r. The allocation of these components is likely to be uneven, but the total imbalance they create is sublinear. We embed each larger component (except for Q_s) using the random allocation algorithm, while respecting the allocation of S and their neighbours. To conclude the allocation, we follow the proof of either Lemma 4.18 or 4.22: we first split the tree, then use either bare paths or leaves to (a) cover exceptional vertices, (b) prepare switching devices (diamonds or leaves) and (c) set up edges allocated along H; we allocate the second half of the tree so as to decrease imbalances, making them sublinear (we can consider imbalances of the whole allocation when setting the weights for the blow-up of $R^\star\big[[k]\big]$) and finally use the switching devices to make this allocation perfect.

The embedding algorithm also proceeds as outlined in the proof sketch above: in a preliminary step we reserve vertices for the final matching, for embedding each vertex of S, and for the neighbours of the vertices in S. The embedding then follows the greedy vertex embedding algorithm, except that we skip vertices for the final matching and that we embed neighbours of vertices of S to the appropriate reserved sets. Since we embed each vertex according to its allocation, we may complete the embedding using a perfect matching.

> Mad, adj. Affected with a high degree of intellectual independence.
>
> ---
>
> Ambrose Bierce

> A tree that is unbending, is easily broken.
>
> ---
>
> Lao Tzu

5 Subdigraphs via chromatic number

For every graph G, let $\mathcal{D}(G)$ denote the family of all orientations of G. In this chapter we investigate a few questions related to Burr's conjecture (Conjecture 1.17), which broadly fall under the following theme. $\mathcal{D}(G)$

Question 5.1. Which digraphs H must be contained in every $D \in \mathcal{D}(G)$?

Since all graphs admit an acyclic orientation, it follows that any such H must be acyclic. We are mostly interested in the case when H is a tree, but we first consider the general setting.

In Section 5.1 we gather results related (either by statement or proof) to random orientations of graphs. We give an upper bound on the largest integer k such that every tournament of order n contains a k-chromatic acyclic subgraph, and obtain a probabilistic result about the order of trees which must be contained almost surely in an orientation of a graph with given chromatic number.

In Section 5.2 we prove two non-probabilistic results, the first is an approximate result towards Burr's conjecture (Theorem 1.22) and the second is a proof of Burr's conjecture for stars (Theorem 5.10).

5.1 Typical behaviour

We begin with an observation from [1], whose proof we include because it illustrates a technique which we use in other results of this section.

Lemma 5.2. [1] Each digraph D contains an acyclic subgraph H with

$$\chi(H) \geq \sqrt{\chi(D)}.$$

Proof. Fix an order $v_1, \ldots, v_n$ of the vertices of D, and paint each edge $v_i \to v_j$ red if $i < j$ and blue otherwise; let R be the digraph formed by the red edges and let B be the one formed by the blue edges, and note that both R and B are acyclic. Note that $\chi(R) \cdot \chi(B) \geq \chi(D)$: indeed, if c_R is a proper colouring of R with $\chi(R)$ colours and c_B is a proper colouring of B with $\chi(B)$ colours, then the colouring $c : x \mapsto \big(c_R(x), c_B(x)\big)$ for all $x \in D$ is a proper colouring of D with at most $\chi(R)\cdot\chi(B)$ colours. It follows that either R or B must have chromatic number at least $\sqrt{\chi(D)}$. □

If H is contained in every $D \in \mathcal{D}(G)$, how high can $\chi(H)$ be, as a function of $\chi(G)$? A natural approach to this question is to use the probabilistic method and consider a 'typical' orientation: does a random orientation of G contain an acyclic digraph with 'high' chromatic number? We obtain a simple sublinear bound for the highest integer $a(k)$ such that every orientation of a k-chromatic graph contains an acyclic subgraph with chromatic number $a(k)$, by considering random tournaments. More precisely, the next lemma implies that for every $\varepsilon > 0$ and sufficiently large k

$a(k)$

$$\inf_G \max\Big\{\, q : \text{each } D \in \mathcal{D}(G) \text{ contains some } H_D \text{ with } \chi(H_D) \geq q \,\Big\} \leq (1+\varepsilon)\frac{k}{2\log k}.$$

where the infimum is taken over all graphs G with $\chi(G) = k$.

Theorem 1.20. For all $\varepsilon > 0$ there exists a tournament D such that if H is an acyclic subgraph of D then

$$\chi(H) \leq \left(\frac{1}{2}+\varepsilon\right)\frac{\chi(D)}{\log\chi(D)}.$$

To prove this theorem we use a celebrated result due to Bollobás [10] about the chromatic number of the binomial random graph $G(n,p)$. The *binomial random graph* $G(n,p)$ is the random labelled graph of order n which is constructed by including each possible edge with probability p, with choices made independently for each edge.

binomial random graph

Theorem 5.3. [10] For all p, with $0 < p < 1$, we have that

$$\chi\Big(G(n,p)\Big) = \left(\frac{1}{2}+\mathrm{o}(1)\right)\log\left(\frac{1}{1-p}\right)\frac{n}{\log n}$$

with probability at least $1 - 2^n\exp\Big(-\Theta(n^2)\Big)$.

Proof of Theorem 1.20. Let G be the complete graph K_k of order k and fix a total order $v_1 \prec \cdots \prec v_k$ of $V(G)$. For any orientation D' of G, we paint each edge $v_i \to v_j$ of D' red if $i < j$ and blue otherwise. Let R be the underlying (undirected) graph formed by the red edges and B be the underlying graph formed by the blue edges. (Note that these graphs depend on the choice of $\prec$ and D').

If we choose D' uniformly at random, by setting the orientation of each edge independently of all other edges, and if we define R and B as above, then

R and B are identically (but not independently) distributed as the binomial random graph $G(k, 1/2)$. For any fixed order $\prec$ and $\varepsilon > 0$, let $E_\varepsilon^\prec$ be the event that '$\max\{\chi(R), \chi(B)\} > (1/2+\varepsilon)k/\log k$'. Then, for all $\varepsilon > 0$ and all $\prec$ we have

$$\mathbb{P}(E_\varepsilon^\prec) \le 2^{k+1}\exp\Bigl(-\Theta(k^2)\Bigr) \quad \text{as } k \to \infty$$

by Theorem 5.3. Note that there are $k! \le k^k = \exp(k\log k)$ possible total orders of $V(G)$ and thus, using a union bound over every order of $V(G)$ we have that

$$\mathbb{P}\left(\bigcup_{\prec} E_\varepsilon^\prec\right) \le \sum_{\prec}(2k)^k\exp\Bigl(-\Theta(k^2)\Bigr) \le \exp\Bigl(k\ln(2k^2) - \Theta(k^2)\Bigr) = \mathrm{o}(1),$$

and so with positive probability we have $\max_\prec\{\chi(R), \chi(B)\} \le (1/2+\varepsilon)k/\log k$ where $\prec$ ranges over all total orders of $V(G)$. To conclude, note that if A is an acyclic subdigraph of D', then for some order $\prec$ we have that the underlying graph of A is a subgraph of R. This means that for some orientation D of the edges of K_k the maximum chromatic number of an acyclic subgraph of D is at most $(\frac{1}{2}+\varepsilon)k/\log k$. □

Next, we narrow our focus to the case when H is a tree. More precisely, let G be a graph: what is the largest integer t such that each $D \in \mathcal{D}(G)$ contains every oriented tree of order t? Recall that $q(T)$ is the smallest integer such that every orientation of a $q(T)$-chromatic graph contains a copy of T. Addario-Berry, Havet, Sales, Reed, Thomassé [1] have obtained the following useful lemma.

Lemma 5.4. [1] Every acyclic digraph G contains all oriented trees of order $\chi(G)$.

It follows immediately from Lemmas 5.2 and 5.4 that every orientation of a graph G contains every oriented tree of order $\sqrt{\chi(G)}$, but there is still quite a big gap between this value and the value $\chi(G)/2+1$ in Burr's conjecture. In an effort to advance towards the conjecture, we again investigate the typical behaviour using the probabilistic method: what trees do we find (almost surely) in a uniformly-random orientation of a graph G?

Theorem 1.21. For all positive ε and sufficiently large k, the following holds for every graph G with $\delta(G) \ge k-1$. If D is an orientation of G formed by orienting each $e \in E(G)$ uniformly at random, independently for each edge, then D contains every oriented tree of order $(1-\varepsilon)k/\log k$ almost surely as $k \to \infty$.

Remark. The theorem above is related to chromatic number through the following fact: every graph G contains a subgraph G' with $\delta(G') \ge \chi(G) - 1$.

Proof. Fix ε with $0 < \varepsilon < 1$ (we may assume the upper bound as otherwise the lemma is trivial). For each $v \in V(G)$, let $W_v \subseteq N(v)$ be a set of $k-1$ neighbours of v, chosen arbitrarily. Let $x \in V(G)$, and let t be a positive integer to be chosen later. Let $V_0 := \{x\}$, and for each $i \in [t]$ let $V_i := \bigcup_{y \in V_{i-1}} W_y$. Let D be an orientation of G formed by assigning an orientation to each of its edges uniformly at random and independently from all other edges. We will show that almost surely D contains a subgraph D' which contains every oriented tree of order $(1-\varepsilon)k/\log k$. Let $i \in \{0, \ldots, t-1\}$ and $y \in V_i$, and define $X_y^+ := |N_D^+(y) \cap W_y|$ and $X_y^- := |N_D^-(y) \cap W_y|$. Therefore the semidegree of y in V_{i+1} is at least $\min\{X_y^+, X_y^-\}$, and $X_y^+, X_y^- \sim \mathrm{Bin}\big(k-1, \frac{1}{2}\big)$, so $\mathbb{E}X_y^+ = \mathbb{E}X_y^- = \frac{k-1}{2}$. Note that for each $i \in \{0, \ldots, t-1\}$ and each $y \in V_i$ we have

$$\begin{aligned}\mathbb{P}[X_y^+ < t] &\leq \mathbb{P}\Big[\,\mathbb{E}X_y^+ - X_y^+ > \mathbb{E}X_y^+ - t\,\Big]\\ &\leq \mathbb{P}\Big[\,|\mathbb{E}X_y^+ - X_y^+| > \mathbb{E}X_y^+ - t\,\Big]\\ &\leq 2\exp\left(-2\frac{(\mathbb{E}X_y^+ - t)^2}{k-1}\right)\end{aligned}$$

where the last inequality follows by the Chernoff bound (Theorem 2.13). Now, let $V := V_0 \cup \cdots \cup V_{t-1}$. Then $|V| \leq (k-1)^t = \exp\big(t\ln(k-1)\big)$. By the union bound,

$$\begin{aligned}\mathbb{P}\left[\bigcap_{y\in V}\{X_y^+ \geq t\}\right] &\leq 1 - \sum_{y\in V} 2\exp\left(-2\frac{(\mathbb{E}X_y^+ - t)^2}{k-1}\right)\\ &\leq 1 - 2\exp\left(t\ln(k-1) - 2\frac{(\frac{k-1}{2} - t)^2}{k-1}\right)\\ &\leq 1 - 2\exp\left(t\ln(k-1) - \frac{k-1}{2}\cdot\exp\left(\frac{4t}{k-1}\right)\right).\end{aligned}$$

We now choose t so that this probability tends to 1. It suffices to show that the exponent in the last expression above tends to $-\infty$. If $t = \frac{k-1}{2\omega(k)}$ where we think of ω as a function which tends slowly to infinity, we have

$$t\ln(k-1) - \frac{k-1}{2}\cdot\exp\left(\frac{4t}{k-1}\right) = \frac{k-1}{2}\left(\frac{\ln(k-1)}{\omega(k)} - \exp\left(\frac{2}{\omega(k)}\right)\right).$$

Note that setting $\omega(k) = \ln k/(1-\varepsilon')$ for any constant ε' with $0 < \varepsilon' < 1$ gives us the desired limit, and note that we cannot take ω to be sublogarithmic. Moreover a similar calculation (with $\omega(k) = \ln k/(1-\varepsilon')$) yields

$$\mathbb{P}\left[\bigcap_{y\in V}\{X_y^- \geq t\}\right] \to 1 \quad \text{as} \quad k \to \infty.$$

So if $t = (1-\varepsilon)\frac{k}{\ln k}$, then for each $i \in \{0, \ldots t-1\}$ and each $v \in V_i$ the semidegree of v in V_{i+1} is at least t asymptotically almost surely. Let $D' := D[V_0 \cup \cdots \cup V_t]$.

Claim 5.5. If T is an oriented tree on t vertices, then D' contains T almost surely.

Proof. Suppose we fix an outcome of the random orientation such that for each $i \in \{0, \dots t-1\}$ and each $v \in V_i$ the semidegree of v in V_{i+1} is at least t. Let $x_0, x_1, \dots, x_{t-1}$ be an ancestral order of $V(T)$. We shall embed T to D' greedily.

We first embed x_0 to the single vertex $v_0 \in V_0$. Now suppose that we have embedded all vertices up to x_{i-1}, and wish to embed x_i. Then x_i has a single neighbour x which has already been embedded, say to some vertex $v \in V_j$. The semidegree of v in V_{j+1} is at least t, and at most $i-1 < t$ neighbours of v have been used by vertices already embedded, so v has at least one inneighbour and one outneighbour in V_j which have not been used. This means we can embed v_i to some unused vertex in V_j, extending the embedding. Proceeding in this manner we obtain a copy of T in D'. Since the random orientation almost surely has the required properties, the claim follows. $\diamond$

The claim above implies that if $t = (1-\varepsilon)k/\ln k$ then a random orientation of G contains a copy of every oriented tree on t vertices almost surely. $\square$

5.2 Non-probabilistic results

To conclude this chapter, we prove two results for specific classes of graphs and trees. In Theorem 1.22 we consider graphs G such that $\chi(G)$ is at least a polylogarithmic function of $|G|$, and in Theorem 5.10 we verify Burr's conjecture all but two orientations of every star, which—together with a Lemma in [1]—implies that Burr's conjecture holds for every orientation of a star.

To prove Theorem 1.22 we require a few definitions and two auxiliary lemmas. Let D be a digraph. If $S \subseteq V(D)$, then we write $N^+(S)$ for the set of vertices in $D - S$ which have an inneighbour in S, and $N^-(S)$ for the set of vertices in $D - S$ which have an outneighbour in S. If $N^+(S) = D - S$ then we say that S is a *dominating set* of D; similarly, if $N^-(S) = D - S$ then S is an *anti-dominating set* of D. Our first lemma shows that we can always find an independent set with 'large' outneighbourhood.

$N^+(S)$

dominating set, anti-dominating set

Lemma 5.6. Every digraph D on n vertices contains a maximal independent set S such that $|N^+(S)| \geq |N^-(S)|$.

Proof. Fix an enumeration $v_1, \dots, v_n$ of $V_0 := V(D)$ which maximises the number of forward arcs, i.e., of arcs $v_i \to v_j$ where $i < j$. For each $i \geq 1$, if $V_{i-1} \neq \varnothing$, proceed as follows. Let s_i be the vertex v_j of minimum index j in V_{i-1} and let $V_i = V_{i-1} \setminus \Big(\{s_i\} \cup N^+(s_i) \cup N^-(s_i)\Big)$. Finally, let S be the set of all s_i. By construction $D[S]$ is a maximal independent set of D. Furthermore, by the labelling of the vertices, we have that $|N^+(S)| \geq |N^-(S)|$, since the outneighbourhood of s_i in V_{i-1} is at least as large as its inneighbourhood in V_{i-1}, otherwise s_i could be moved to the end of the order, yielding an order with more forward edges, a contradiction. □

By applying this lemma repeatedly, we obtain the following.

Lemma 5.7. Let D be a digraph on n vertices. Then there exists a dominating set S of D such that $\chi\Big(D[S]\Big) \leq \log_2 n + 1$. Furthermore, $D[S]$ is acyclic.

Proof. This follows by induction on $n := |V(D)|$. If $n = 1$ the lemma holds trivially, so suppose $n > 1$. By Lemma 5.6, D contains an maximal independent set I such that $|N^+(I)| \geq |N^-(I)|$, that is $|N^-(I)| \leq (n - |I|)/2$. Let

$$D' := D - \Big(I \cup N^+(I)\Big) = D\Big[N^-(I) \setminus N^+(I)\Big],$$

where the second inequality follows by the maximality of I. By induction there exists $S' \subseteq V(D')$ such that $D'[S']$ is acyclic and $\chi(D'[S']) \leq \log_2(|V(D')|) + 1 \leq \log_2(n/2) + 1$ and such that S' is a dominating set of D'. Let $S := D[I \cup S']$. Clearly S is a dominating set of D and $\chi(D[S]) \leq 1 + \chi(D[S']) \leq \log_2 n + 1$. Moreover, since $V(D') = N^-(I) \setminus N^+(I)$, every edge between S' and I is directed towards the latter; so any directed cycle in $D[S]$ must lie entirely in $D[S']$. Since $D[S']$ is acyclic, we conclude that $D[S]$ is acyclic as well. □

Theorem 1.22. If D is a digraph of order n, where $n \geq 1$, then D contains every oriented tree of order $\chi(D)/\log_2(2n)$, that is,

$$q^{-1}(D) \geq \frac{\chi(D)}{\log_2(2n)}.$$

Proof. Let D be an oriented graph on $n \geq 1$ vertices and let T be an oriented tree on t vertices, where $t \leq \chi(D)/\log_2(2n)$. We will show that D contains a copy of T. The proof is by induction on n. The theorem is trivial if $t = 1$ (since every digraph with one vertex contains a copy of a tree with at most 1 vertex), and so we suppose that $t > 1$. Suppose further that T contains an out-leaf (respectively, in-leaf) v. By Lemma 5.7, there exists an anti-dominating (respectively, dominating) set S of D such that $\chi(D[S]) \leq \log_2(2n)$. Let $D' = D - S$, and $n' := |V(D')| = n - |S|$. Note that since $t > 1$ we have $n' \geq \chi(D) - \log_2(2n) > 0$ and $\chi(D') \geq \chi(D) - \log_2(2n)$.

By induction, D' contains a copy of every oriented tree on $\chi(D')/\log_2(2n')$ vertices. Note that

$$\frac{\chi(D')}{\log_2(2n')} \geq \frac{\chi(D)-\log_2(2n)}{\log_2\big(2(n-|S|)\big)} \geq \frac{\chi(D)-\log_2(2n)}{\log_2(2n)} \geq \frac{\chi(D)}{\log_2(2n)} - 1 \geq t-1,$$

and thus D' contains a copy of $T-v$. Since every vertex of D' has an outneighbour (respectively, inneighbour) in S, it follows that D contains a copy of T. □

5.2.1 Burr's conjecture holds for stars

As noted in [1], every graph G with $\chi(G)=2t-2$ contains a subgraph G' with minimum degree $2t-3$. So for every orientation D of G' we have that the average out-degree of D is

$$\sum_{v\in V(D)} \frac{\deg^+_D(v)}{|V(D)|} \geq \frac{|E(D)|}{|V(D)|} \geq t-1-\frac{1}{2},$$

and thus D contains a vertex with outdegree at least $t-1$, and (by symmetry) also a vertex of in-degree at least $t-1$. Therefore, every orientation of G contains the two anti-directed stars of order t (i.e., a star in whose centre has either outdegree $t-1$ or indegree $t-1$). To complement this result, we introduce some notation. Let $S(a,b)$ be an oriented star on $a+b+1$ vertices whose centre has in-degree a and out-degree b. The observation above can be stated as the following lemma—note that since $|S(0,r)|=|S(r,0)|=r+1$, this establishes Burr's conjecture for these two types of stars.

Lemma 5.8. [1] Let D be a digraph such that $\chi(D)\geq 2r$. Then D contains a copy of $S(r,0)$ and a copy of $S(0,r)$.

We also need the following fact.

Fact 5.9. If G is a graph such that $V(G)=A_1\cup A_2\cup\cdots\cup A_k$, then

$$\chi(G)\leq\sum_{i=1}^{k}\chi\big(G[A_i]\big).$$

Using Lemma 5.8 and Fact 5.9 we can give a simple proof of Burr's conjecture for all remaining orientations of stars.

Theorem 5.10. Let a and b be positive integers. If D is a digraph such that $\chi(D)\geq 2(a+b+1)-3=2(a+b)-1$, then D contains a copy of $S(a,b)$.

Proof. For all $S \subseteq V(D)$, we write $\chi(S)$ for $\chi(D[S])$. Let A be the set of vertices of D which have in-degree less than a, let B be the set of vertices of D which have out-degree less than b, and let $C = V(D) \setminus (A \cup B)$. We will show that $C \neq \varnothing$, and thus D contains a copy of $S(a, b)$. By Lemma 5.8 we have that $\chi(A) \leq 2a - 1$ and $\chi(B) \leq 2b - 1$, so by Fact 5.9

$$\chi(D) \leq \chi(A) + \chi(B) + \chi(C) \leq 2a - 1 + 2b - 1 + \chi(C) = 2(a + b) - 2 + \chi(C).$$

Therefore $\chi(C) > 0$, and so $C \neq \varnothing$, completing the proof. □

Corollary 5.11. Burr's conjecture holds for every orientation of a star.

You are so much more than your orientation, you know it and I know it.

Adam Lambert

Certainly, as a guitarist, I was aware of descending chromatic lines and arpeggios long before 1968.

Jimmy Page

6 Further directions

Spanning trees are a central topic in graph theory. The area is full of open problems, some of which seem to require a substantially different approach when transferred to the digraph setting. We discuss a few directions for further research.

6.1 Trees in tournaments

Recall that Theorem 1.4 states that all large nice oriented trees of polylogarithmic maximum degree are unavoidable. Together with Moon's theorem on the maximum degree of a random labelled tree (Theorem 1.5) and our proof that almost all labelled oriented trees are nice (Theorem 1.6) this established Theorem 1.2, that almost all labelled oriented trees are unavoidable.

The same method can be used to show that other classes of random oriented trees are asymptotically almost surely unavoidable. More precisely, let $\mathcal{T}$ be a class of undirected trees, let $\mathcal{T}_n$ consist of all members of $\mathcal{T}$ with n vertices, and let T be a tree selected uniformly at random from $\mathcal{T}_n$. If we can show, for some constants C and ξ, that

(a) $\Delta(T) \leq (\log n)^C$ asymptotically almost surely, and

(b) T has at least ξn pendant stars asymptotically almost surely,

then by a similar argument to the proof of Theorem 1.6 it follows that a uniformly-random orientation T^* of T is asymptotically almost surely α-nice (where $\alpha \ll \xi$), and therefore by Theorem 1.4 that T^* is asymptotically almost surely unavoidable. Following the methods of Janson [44] it is not hard to show that (a) and (b) hold for many classes $\mathcal{T}$ of simply-generated random trees, such as uniformly-random ordered trees (see [44, Example 10.1]), binary trees (see [44, Example 10.3]) and d-ary trees for a fixed integer $d \geq 3$ (see [44, Example 10.6]) In the same way Theorem 1.4 directly shows that for many fixed trees T, such as not-too-unbalanced d-ary trees for a fixed integer $d \geq 3$, a random orientation of T is unavoidable asymptotically almost surely. Finally we note that for many oriented trees it is straightforward to check directly that the conditions of Theorem 1.4 are satisfied, for instance in the case of balanced antidirected binary trees, in which every non-leaf vertex has one child as an inneighbour and one child as an outneighbour.

However, there do exist oriented trees which are not nice but which are unavoidable, such as the paths and claws discussed in Chapter 1. In this context it is natural to ask whether the property of being unavoidable can be succinctly characterised or easily tested.

Question 6.1.

(i) Is there a concise characterisation of unavoidable oriented trees?

(ii) Can we determine in polynomial time if an oriented tree is unavoidable?

We suspect that it would be very difficult to establish such a characterisation. As a more attainable goal, it would be interesting to establish further classes of unavoidable oriented trees. For example, say that an oriented tree T with root r is *outbranching* if for every vertex $v \in V(T)$ the path in T from r to v is directed from r to v. In particular, if the root of T is not a leaf then T then has no in-leaves at all, so T is not α-nice for any $\alpha > 0$.

outbranching

Problem 6.2. What conditions are sufficient to ensure that an outbranching oriented tree T is unavoidable?

To shed some light on this problem it may help to consider the outbranching balanced binary trees B_d on $2^{d+1} - 1$ vertices, in which every non-leaf vertex has two children as outneighbours and every leaf is at distance precisely d from the root.

Conjecture 6.3. B_d is unavoidable for d sufficiently large (maybe $d > 1$ suffices).

It seems that further new ideas and techniques would be necessary to prove Conjecture 6.3, since the existence of both many in-leaves and many out-leaves of T is crucial to our approach. In a similar vein, Lu, Chang, Wang, Lin and Wong [61] have shown that for every integer k every sufficiently large tournament contains at least one spanning k-ary tree (a *k-ary tree* is an outbranching where every non-leaf vertex has out-degree precisely k except perhaps one, which is allowed to have smaller out-degree).

k-ary tree

Finally, recall that in Chapter 3 we defined $g(T)$ for an oriented tree T to be the smallest integer such that every tournament on $g(T)$ vertices contains a copy of T. So T is unavoidable if and only if $g(T) = |T|$. As noted earlier, if T is an anti-directed star on n vertices then $g(T) \geq 2n - 2$, and Kühn, Mycroft and

Osthus's proof of Conjecture 1.7 for large trees shows that this is the maximum possible value of $g(T)$ for large n. That is, every oriented tree T on n vertices, where n is large, has $g(T) \leq 2n-2$. The following 'double-star' construction due to Allen and Cooley (see [51]) also yields an oriented tree T for which $g(T)$ is significantly larger than $|T|$. Fix $a, b, c \in \mathbb{N}$ with $a+b+c=n$, and let T be the oriented tree on n vertices formed from a directed path P on b vertices by adding a new vertices as inneighbours of the initial vertex of P and adding c new vertices as outneighbours of the terminal vertex of P. Now take disjoint sets of vertices A, B and C of sizes $2a-1, b-1$ and $2c-1$ respectively, and let G be the tournament in which $G[A]$ and $G[C]$ are regular tournaments, $G[B]$ is an arbitrary tournament, and all remaining edges of G are directed from A to B, from B to C or from A to C. So G has $2a+b+2c-3 = 2n-b-3$ vertices, but G does not contain a copy of T, since then (as $|B| < b$) either the initial vertex of P would be in A, which cannot occur since each vertex of A has only $a-1$ inneighbours, or the terminal vertex of P would be in C, which cannot occur since each vertex of C has only $c-1$ outneighbours. So $g(T) \geq 2n-b-2$ (and it is not too hard to check that in fact $g(T) = 2n-b-2$).

For any $\Delta, n \in \mathbb{N}$, taking $a = c = \Delta - 1$ and $b = n - 2\Delta + 2$ in the above construction yields an oriented tree T on n vertices with $\Delta(T) = \Delta$ and $g(T) = n + 2\Delta - 4$. In other words, for any $n \in \mathbb{N}$ and any $\Delta \geq 3$ there exist oriented trees on n vertices with maximum degree at most Δ which are not contained in some tournament on $n + 2\Delta - 5$ vertices. On the other hand, Theorem 1.9 shows that every oriented tree whose maximum degree is at most polylogarithmic in n is contained in every tournament on $n + o(n)$ vertices. Kühn, Mycroft and Osthus [51] asked whether this $o(n)$ term can be replaced by a constant for oriented trees whose maximum degree is at most a constant Δ, and the previous construction shows that a constant of $2\Delta - 4$ would be best possible. More generally it would be interesting to know whether the previous construction is extremal for any bound on $\Delta(T)$ (as a function of n), with the exception of the antidirected paths $\check{P}_3, \check{P}_5$ and $\check{P}_7$ on 3, 5 and 7 vertices respectively — as described in the introduction, these three paths are avoidable and so are not contained in any tournament on the same number of vertices.

Question 6.4. With the exception of $\check{P}_3, \check{P}_5$ and $\check{P}_7$, is every oriented tree T on n vertices contained in every tournament on $n + 2\Delta(T) - 4$ vertices?

6.2 Trees in digraphs via semidegree

The main theorems in Chapter 4 describe large families of spanning trees which must be contained in every digraph G whose semidegree is slightly above the minimum threshold for connectivity. We prove, among others, Theorem 1.12—a digraph analogue of Komlós, Sárközy and Szemerédi's classical theorem (Theorem 1.11). We also prove an approximate result for almost spanning trees T (of every such G), where $\Delta(T)$ is allowed to be a polylogarithmic function of T. Finally, we demonstrate how our techniques can be used to embed spanning tree-like digraphs with polylogarithmic maximum degree.

It would also be interesting to clarify whether the polylogarithmic bound we require for the degree of the trees in many of our results could be improved without a substantially different approach. I believe that the bottleneck for such improvement is related to the limits of Lemma 2.6: this is a tree-partition lemma which lies at the core of our analysis of the allocation algorithm (Algorithm 4.14). While this bound on the degree is sufficient high to encompass almost every tree, the polylogaritmic bound on the degree is still far from the known bounds for graphs (see below).

Another natural question is whether similar results hold when the minimum semidegree of the host graph G is closer to $n/2$. Csaba, Levitt, Nagy-György and Szemerédi [21] have shown that this is possible for bounded-degree (undirected) trees.

Theorem 6.5. [21] For all Δ there exists n_0 and c_Δ (which both depend only on Δ) such that the following holds for all $n \geq n_0$. If T is a tree of order n and maximum degree Δ and G is a graph of order n and minimum degree $\delta(G) \geq n/2 + c_\Delta \log n$, then $T \subseteq G$. Furthermore, the bound on $\delta(G)$ is tight: for sufficiently large n, there exists a graph G with $\delta(G) \geq n/2 + \log n/17$ such that the complete ternary tree of order n is not a subgraph of G.

Many variations and extensions of Theorem 1.11 have been developed for graphs. These include, among others, allowing $\Delta(T)$ to be $\mathrm{O}(n/\log n)$ [50] and generalising the class of spanning subgraphs to graphs with bounded bandwidth [14] and arrangeability [15]. The presence of spanning trees has also been studied in resilience [6], random perturbation [13, 64] and maker-breaker [20] scenarios. It is very natural to ask which of these proofs can be adapted to the realm of digraphs (and which cannot), and, conversely, whether techniques such as those described in this thesis could help describe further classes of spanning (undirected) graphs.

6.3 Trees via chromatic number

Let T be an oriented tree. Recall that $t(T)$ is the smallest integer such that every tournament on $t(T)$ vertices contains a copy of T, and that $q(T)$ is the smallest integer such that every orientation of a $q(T)$-chromatic graph contains a copy of T. Then $t(T) \leq q(T)$. I conjecture the following.

Conjecture 1.23 (Transference conjecture). If T is an oriented tree, then

$$t(T) = q(T).$$

This is known to hold for any path P with at most 2 blocks [2, 32, 38, 69, 28, 82] (where $t(P) = q(P) = |V(P)|$ by Lemma 5.8) and for anti-directed stars [1] (that is, for stars S such that each vertex has either no inneighbours or no outneighbours, in which case we have $t(S) = q(S) = 2|S| - 2$). In Chapter 5 we completed the proof for stars with Theorem 5.10, showing that Conjecture 1.23 holds for every orientation of a star (where if S is a star which is not antidirected then $t(S) = q(S) = 2|S| - 3$).

To see why these results settle cases of Conjecture 1.23, it suffices to argue that Lemma 5.8 and Theorem 5.10 are best possible. Indeed, since a regular tournament on $2r - 3$ vertices does not contain a copy of the antidirected stars $S(0, r), S(r, 0)$ of order $r + 1$, it follows that Lemma 5.8 is best possible. As for Theorem 5.10, consider the following construction. Let A and B be regular tournaments on $2a - 1$ and $2b - 1$ vertices respectively and construct a tournament T on $2(a + b + 1) - 4$ vertices by taking vertex-disjoint copies of A and B, and adding all edges from A to B; then every vertex of A has in-degree less than a and every vertex in b has out-degree less than b, so T contains no copy of $S(a, b)$.

Theorem 6.6. The transference conjecture holds for every oriented star.

Conjecture 1.23, if confirmed, transfers many of the results discussed in this thesis (such as Theorems 1.2 and 1.9) from the realm of tournaments into results towards Burr's conjecture. In particular, it would imply that Burr's conjecture holds for sufficiently large values of n.

We note that a similar question has been considered by Bialostocki and Gyárfás recently [9]. Following their terminology, we define 'Ramsey niceness' as follows. If $\mathcal{F}$ is a family of graphs, we write $R_k(\mathcal{F})$ to denote the smallest integer n such that every k-colouring of K_n contains a monochromatic copy of some $F \in \mathcal{F}$. The family is *k-nice* if for every graph G with $\chi(G) = R_k(\mathcal{F})$ and every k-colouring k-nice of $E(G)$ there exists a monochromatic copy of some $F \in \mathcal{F}$. Aharoni et. al. [3] have since provided some support for the following conjecture: for every $\mathcal{F}$, there exists an infinite set of values of k (perhaps even all $k \geq k_0(\mathcal{F})$) for which $\mathcal{F}$ is k-nice. This is very similar in spirit to the numbers captured by $t(T)$ and $q(T)$, with colourings in place of orientations and by considering families of graphs rather than trees.

The greater our knowledge increases
the more our ignorance unfolds.

John F. Kennedy

All you need in this life is ignorance
and confidence, and then success is
sure.

Mark Twain

If ignorance is bliss, there should be
more happy people.

Victor Cousin

Index of Definitions

'When I use a word,' Humpty Dumpty said in rather a scornful tone, 'it means just what I choose it to mean—neither more nor less.'

Lewis Carroll

Bibliography

[1] L. Addario-Berry, F. Havet, C. L. Sales, B. Reed, and S. Thomassé. *Oriented trees in digraphs*. Discrete Mathematics, **313**, 2013, 967–974 (see iii, v, 13, 14, 125, 127, 129, 131, 137).

[2] L. Addario-Berry, F. Havet, and S. Thomassé. *Paths with two blocks in n-chromatic digraphs*. Journal of Combinatorial Theory, Series B, **4**, 2007, 620–626 (see 12, 14, 137).

[3] R. Aharoni, N. Alon, M. Amir, P. Haxell, D. Hefetz, Z. Jiang, G. Kronenberg, and A. Naor. *Ramsey-nice families of graphs*. European Journal of Combinatorics, **72**, 2018, 29–44 (see 138).

[4] 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 (see iv, 25, 42).

[5] N. Alon and J. H. Spencer. *The probabilistic method*, With an appendix by Paul Erdős, A Wiley-Interscience Publication. New York: John Wiley & Sons Inc., 1992, xvi+254 (see iii, iv, 11, 26).

[6] J. Balogh, B. Csaba, and W. Samotij. *Local resilience of almost spanning trees in random graphs*. Random Structures & Algorithms, **38**, 2011, 121–139 (see 136).

[7] J. Bang-Jensen and G. Gutin. *Digraphs: Theory, Algorithms and Applications*, London: Springer-Verlag, 2001 (see 7).

[8] E. A. Bender and N. C. Wormald. *Random trees in random graphs*. Proceedings of the American Mathematical Society, **103**, 1988, 314–320 (see iii, 3).

[9] A. Bialostocki and A. Gyárfás. *Replacing the host K_n by n-chromatic graphs in Ramsey-type results*. Open Access Library, 2015. arXiv: `1506.04495` (see 137).

[10] B. Bollobás. *The chromatic number of random graphs*. Combinatorica, **8**, 1988, 49–55 (see v, 126).

[11] B. Bollobás. *Random graphs*, vol. 73. Cambridge Studies in Advanced Mathematics. Cambridge: Cambridge University Press, 2001, xviii+498 (see 26).

[12] J. A. Bondy and U. S. R. Murty. *Graph theory*, vol. 244. Graduate Texts in Mathematics. New York: Springer-Verlag, 2008, xii+651 (see v, 82).

[13] J. Böttcher, J. Hàn, Y. Kohayakawa, R. Montgomery, O. Parczyk, and Y. Person. *Universality for bounded degree spanning trees in randomly perturbed graphs*, Feb. 2018. arXiv: 1802.04707 (see 136).

[14] J. Böttcher, M. Schacht, and A. Taraz. *Proof of the bandwidth conjecture of Bollobás and Komlós*. Mathematische Annalen, **343**, 2009, 175–205 (see 136).

[15] J. Böttcher, A. Taraz, and A. Würfl. *Spanning embeddings of arrangeable graphs with sublinear bandwidth*. Random Structures & Algorithms, **48**, 2015, 270–289 (see 136).

[16] S. A. Burr. *Subtrees of directed graphs and hypergraphs*, Proceedings of the eleventh Southeastern Conference on Combinatorics, graph theory and Computing (Boca Raton), 1980, 227–239 (see 11, 12).

[17] L. Caccetta and R. Häggkvist. *On minimal digraphs with given girth*, Proceedings of the Ninth Southeastern Conference on Combinatorics, Graph Theory and Computing (Florida Atlantic University, Boca Raton, Florida), vol. 21. 1978 (see iii, 7).

[18] M. C. Cai. *A counterexample to a conjecture of Grant*. Discrete Mathematics, **44**, 1983, 111 (see 9).

[19] F. R. K. Chung. *A note on subtrees in tournaments*. Bell Laboratories Internal Memorandum. 1982. http://www.math.ucsd.edu/~fan/mypaps/fanpap/59anosit.PDF (see 5).

[20] D. Clemens, A. Ferber, R. Glebov, D. Hefetz, and A. Liebenau. *Building spanning trees quickly in Maker-Breaker games*. SIAM Journal on Discrete Mathematics, **29**, 2015, 1683–1705 (see 136).

[21] B. Csaba, J. Nagy-György, I. Levitt, and E. Szemerédi. *Tight Bounds for Embedding Bounded Degree Trees*, Fete of Combinatorics and Computer Science, Berlin (Heidelberg): Springer, 2010, 95–137. https://doi.org/10.1007%2F978-3-642-13580-4_5 (see v, 136).

[22] B. Cuckler. *On the number of short cycles in regular tournaments*. unpublished manuscript (2008). (see 7).

[23] A. Czygrinow, L. DeBiasio, T. Molla, and A. Treglown. *Tiling directed graphs with tournaments*. Forum of Mathematics, Sigma, **6**, 2018, 53 pages (see 7).

[24] A. Czygrinow, H. A. Kierstead, and T. Molla. *On directed versions of the Corrádi–Hajnal corollary*. European Journal of Combinatorics, **42**, 2014, 1–14 (see 7).

[25] L. DeBiasio, D. Kühn, T. Molla, D. Osthus, and A. Taylor. *Arbitrary orientations of Hamilton cycles in digraphs*. SIAM Journal on Discrete Mathematics, **29**, 2015, 1553–1584 (see 8).

[26] L. DeBiasio and T. Molla. *Semi-degree threshold for anti-directed Hamiltonian cycles*. Electronic Journal of Combinatorics, **22**, 2015 (see 8, 9).

[27] R. Diestel. *Graph theory*, vol. 173. Graduate Texts in Mathematics. Translated from the 1996 German original. New York: Springer-Verlag, 1997, xiv+289 (see 17).

[28] A. El Sahili and M. Kouider. *About paths with two blocks*. Journal of Graph Theory, **3**, 2007, 221–226 (see 12, 137).

[29] A. El Sahili. *Trees in tournaments*. Journal of Combinatorial Theory, Series B, **92**, 2004, 183–187 (see iv, 5, 38).

[30] Erdős. *Graph Theory and Probability*. Canadian Journal of Mathematics, **11**, 1959, 34–38 (see 11).

[31] P. Erdős. *Some problems in graph theory*, Theory of Graphs and its Applications, ed. by M. Fielder. New York: Academic Press, 1965, 29–36 (see 11).

[32] T. Gallai. *On directed paths and circuits*, Theory of graphs (Proceedings of Colloquium held at Tihany, 1966), New York: Academic press, 1968, 15–118 (see iii, 12, 137).

[33] B. Gendron, A. Hertz, and P. St-Louis. *On a generalization of the Gallai–Roy–Vitaver theorem to the bandwidth coloring problem*. Operations Research Letters, **36**, 2008, 345–350 (see 12).

[34] A. Gyárfás. *On Ramsey covering numbers*, Infinite and Finite Sets, New York: North-Holland/American Elsevier, 1973, 801–816 (see 11).

[35] R. Häggkvist and A. Thomason. *Trees in tournaments*. Combinatorica, **11**, 1991, 123–130 (see 5).

[36] A. Hajnal and E. Szemerédi. *Proof of a conjecture of P. Erdős*, Combinatorial theory and its applications, II (Proc. Colloq., Balatonfüred, 1969), Amsterdam: North-Holland, 1970, 601–623 (see 7).

[37] Y. O. Hamidoune. *A note on minimal directed graphs with given girth*. Journal of Combinatorial Theory, Series B, **43**, 1987, 343–348 (see 7).

[38] M. Hasse. *Zur algebraischen Begründung der Graphentheorie I*. Mathematische Nachrichten, **28**, 1964, 275–290 (see iii, 12, 137).

[39] F. Havet. *Trees in tournaments*. Discrete mathematics, **243**, 2002, 121–134 (see iii, 5).

[40] F. Havet and S. Thomassé. *Median orders of tournaments: A tool for the second neighborhood problem and Sumner's conjecture*. Journal of Graph Theory, **35**, 2000, 244–256 (see 5).

[41] 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 (see 2).

[42] C. T. Hoáng and B. Reed. *A note on short cycles in digraphs*. Discrete Mathematics, **66**, 1987, 103–107 (see 7).

[43] G. Houri. *Une condition suffisante d'existence d'un circuit Hamiltonien*. Comptes Rendus de l'Académie des Sciences, **25**, 1960, 495–497 (see 8).

[44] S. Janson. *Simply generated trees, conditioned Galton-Watson trees, random allocations and condensation*. Probability Surveys, **9**, 2012, 103–252 (see iv, 71, 133).

[45] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs*, Wiley-Interscience, New York, 2000, xii+333 (see iv, 27).

[46] M. Kabiya and R. Yuster. *Packing Transitive Triples in a Tournament*. Annals of Combinatorics, **12**, 2009, 291–306 (see 7).

[47] P. Keevash and B. Sudakov. *Triangle packings and 1-factors in oriented graphs*. Journal of Combinatorial Theory, Series B, **99**, 2009, 709–727 (see 7).

[48] 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 (see 7, 8).

[49] J. Komlós, G. N. Sárközy, and E. Szemerédi. *Proof of a packing conjecture of Bollobás*. Combinatorics, Probability, and Computing, **4**, 1995, 241–255 (see iii, 8).

[50] J. Komlós, G. N. Sárközy, and E. Szemerédi. *Spanning trees in dense graphs*. Combinatorics, Probability, and Computing, **10**, 2001, 397–416 (see 9, 136).

[51] D. Kühn, R. Mycroft, and D. Osthus. *An approximate version of Sumner's universal tournament conjecture*. Journal of Combinatorial Theory, Series B, **101**, 2011, 415–447 (see iv, 5, 6, 20–22, 26, 31–39, 41–43, 79, 91, 94, 135).

[52] D. Kühn, R. Mycroft, and D. Osthus. *A proof of Sumner's universal tournament conjecture for large tournaments*. Proceedings of the London Mathematical Society, **102**, 2010, 731–766 (see 5).

[53] D. Kühn and D. Osthus. *Embedding large subgraphs into dense graphs*, ed. by S. Huczynska, J. D. Mitchell, and C. M. Roney-Dougal. Vol. 365. London Mathematical Society Lecture Notes. Cambridge University Press, 2009, 137–167 (see 8).

[54] D. Kühn and D. Osthus. *A survey on Hamilton cycles in directed graphs*. European Journal of Combinatorics, **33**, 2012, 750–766 (see 7).

[55] D. Kühn and D. Osthus. *Hamilton cycles in graphs and hypergraphs: an extremal perspective*, Proceedings of the International Congress of Mathematicians (Seoul, Korea), vol. 4. 2014, 381–406 (see 8).

[56] D. Kühn, D. Osthus, and T. Townsend. *Proof of a tournament partition conjecture and an application to 1-factors with prescribed cycle lengths*. Combinatorica, **36**, 4 2016, 451–169 (see 8).

[57] D. Kühn, D. Osthus, and A. Treglown. *Hamiltonian degree sequences in digraphs*. Journal of Combinatorial Theory, Series B, **100**, 2010, 367–380 (see 43).

[58] L. Li and T. Molla. *Cyclic triangle factors in regular tournaments*. ArXiv e-prints, 2018. arXiv: `1806.06903 [math.CO]` (see 7).

[59] X. Lu. *On claws belonging to every tournament*. Combinatorica, **11**, 1991, 173–179 (see 2).

[60] X. Lu. *Claws contained in all n-tournaments*. Discrete Mathematics, **119**, 1993, 107–111 (see 2).

[61] X. Lu, D.-W. Wang, G. J. Chang, I.-J. Lin, and C.-K. Wong. *On k-ary spanning trees of tournaments*. Journal of Graph Theory, **30**, 1999, 167–176 (see 134).

[62] X. Lu, D.-W. Wang, and C.-K. Wong. *On avoidable and unavoidable claws*. Discrete Mathematics, **184**, 1998, 259–265 (see 2).

[63] C. McDiarmid. *Concentration*, Probabilistic methods for algorithmic discrete mathematics, vol. 16. Algorithms Combin. Berlin: Springer, 1998, 195–248 (see iv, 27).

[64] B. S. Michael Krivelevich Matthew Kwan. *Bounded-Degree Spanning Trees in Randomly Perturbed Graphs*. SIAM Journal on Discrete Mathematics, **31**, 2017, 155–171 (see 136).

[65] J. W. Moon. *Counting labelled trees*, Canadian Mathematical Monographs. Canadian Mathematical Congress, 1970 (see iii, 5).

[66] R. Mycroft and T. Naia. *Spanning trees of dense directed graphs*. In preparation (see 1).

[67] R. Mycroft and T. Naia. *Unavoidable trees in tournaments*. Random Structures & Algorithms, **53**, 2018, 352–385 (see 1).

[68] L. Rédei. *Ein kombinatorischer Satz*. Acta Litteraria Szeged, **7**, 1934, 39–43 (see 2).

[69] B. Roy. *Nombre chromatique et plus longs chemins d'un graphe*. Revue Française d'Informatique et de Recherche Opérationelle, **1**, 1967, 29–132 (see iii, 12, 137).

[70] M. E. Saks and V. T. Sós. *On unavoidable subgraphs of tournaments*, Infinite and finite sets, ed. by A. Hajnal, L. Lovász, and V. T. Sós. Vol. 2. Colloquia Mathematica Societatis János Bolyai. North-Holland, 1981, 663–674 (see 2).

[71] A. D. Scott. *Induced trees in graphs of large chromatic number*. J. Graph Theory, **24**, 1997, 297–311 (see 11).

[72] A. D. Scott and P. Seymour. *Induced subgraphs of graphs with large chromatic number. XIII. New brooms*. ArXiv e-prints, 2018. Submitted. arXiv: `1807.03768` (see 11).

[73] J. Shen. *On the girth of digraphs*. Discrete Mathematics, **211**, 2000, 167–181 (see 7).

[74] B. Sudakov and J. Vondrák. *A randomized embedding algorithm for trees*. Combinatorica, **30**, 2010, 445–470 (see iv, 27).

[75] B. D. Sullivan. *A Summary of Problems and Results related to the Caccetta-Haggkvist Conjecture*. ArXiv preprint, May 2006. arXiv: `math/0605646` (see 7).

[76] D. P. Sumner. *Subtrees of a graph and chromatic number*. The theory and applications of graphs, 1981, 557–576 (see 11).

[77] E. Szemerédi. *On sets of integers containing no k elements in arithmetic progression*. Acta Arithmetica, **27**, 1975. Collection of articles in memory of Juriĭ Vladimirovič Linnik, 199–245 (see 24).

[78] E. Szemerédi. *Regular partitions of graphs*, Problèmes combinatoires et théorie des graphes (Colloques Internationaux CNRS, Université Orsay, Orsay, 1976), Paris: CNRS, 1978, 399–401 (see 24).

[79] A. Thomason. *Paths and cycles in tournaments*. Transactions of the American Mathematical Society, **296**, 1986, 167–180 (see 2).

[80] A. Treglown. *A note on some embedding problems for oriented graphs*. Journal of Graph Theory, **69**, 2012, 330–336 (see 7).

[81] A. Treglown. *On directed versions of the Hajnal-Szemerédi theorem*. Combinatorics, Probability, and Computing, **24**, 2015, 873–928 (see 7).

[82] L. M. Vitaver. *Determination of minimal coloring of vertices of a graph by means of boolean powers of the incidence matrix*. Proceedings of the USSR Academy of Sciences, **147**, 1962 (see iii, 12, 137).

[83] N. C. Wormald. *Subtrees of large tournaments*, Combinatorial Mathematics X, ed. by L. R. A. Casse. Vol. 1036. Lecture Notes in Mathematics. Berlin (Heidelberg): Springer, 1983, 417–419 (see iii, 5, 11).

[84] R. Yuster. *Combinatorial and computational aspects of graph packing and graph decomposition*. Comput. Sci. Rev., **1**, 2007 (see 7).

This document was typeset with LaTeX, using the memoir document class, the Latin Modern font and the microtype package. BibLaTeX was used for the bibliography, Tikz for the diagrams, and ... well, the list would never end. My heartfelt thanks to the LaTeX community, for so much more than TeXnicalities! ❤

TNL