Spanning trees in dense directed graphs

Amarja Kathapurkar* Richard Montgomery†

Abstract

In 2001, Komlós, Sárközy and Szemerédi proved that, for each $\alpha > 0$, there is some $c > 0$ and n_0 such that, if $n \geq n_0$, then every n-vertex graph with minimum degree at least $(1/2 + \alpha)n$ contains a copy of every n-vertex tree with maximum degree at most $cn/\log n$. We prove the corresponding result for directed graphs. That is, for each $\alpha > 0$, there is some $c > 0$ and n_0 such that, if $n \geq n_0$, then every n-vertex directed graph with minimum semi-degree at least $(1/2 + \alpha)n$ contains a copy of every n-vertex oriented tree whose underlying maximum degree is at most $cn/\log n$.

As with Komlós, Sárközy and Szemerédi's theorem, this is tight up to the value of c. Our result improves a recent result of Mycroft and Naia, which requires the oriented trees to have underlying maximum degree at most Δ, for any constant $\Delta \in \mathbb{N}$ and sufficiently large n. In contrast to the previous work on spanning trees in dense directed or undirected graphs, our methods do not use Szemerédi's regularity lemma.

1 Introduction

Given two graphs H and G, when may we expect to find a copy of H in G? In general, this decision problem is NP-complete, and therefore we seek simple conditions on G which imply it contains a copy of H. An important early result is Dirac's theorem from 1952 that, when $n \geq 3$, any n-vertex graph with minimum degree at least $n/2$ contains a cycle through every vertex, that is, a Hamilton cycle. This is a particular instance of the following meta-question, which has seen much subsequent study. Given an n-vertex graph H, what is the lowest minimum degree condition on an n-vertex graph G which guarantees it contains a copy of H? As such a copy of H would contain every vertex in G, we say it is a *spanning* copy of H.

This question has been studied for many different graphs H, for example when H is a K-factor for some small fixed graph K [8, 14], the k-th power of a Hamilton cycle for any $k \geq 2$ [11] and when H has bounded chromatic number and maximum degree, and sublinear bandwith [4]. For more details on these results, and those for other graphs, see the survey by Kühn and Osthus [13]. Here, we will concentrate on the minimum degree required to guarantee different spanning trees.

Komlós, Sárközy and Szemerédi [10] proved in 1995 that, for each $\alpha, \Delta > 0$, there is some n_0 such that, if $n \geq n_0$, then every n-vertex graph with minimum degree at least $(1/2+\alpha)n$ contains a copy of every n-vertex tree with maximum degree at most Δ, thus confirming a conjecture of Bollobás [2]. This result is furthermore notable as one of the earliest applications of the blow-up lemma. In 2001, Komlós, Sárközy and Szemerédi [12] relaxed the maximum degree condition, showing that, for each $\alpha > 0$, there is some $c > 0$ and n_0 such that, if $n \geq n_0$, then every

*University of Birmingham, Birmingham, B15 2TT, UK. Email: AXK516@student.bham.ac.uk.

†University of Birmingham, Birmingham, B15 2TT, UK. Email: r.h.montgomery@bham.ac.uk. Supported by the European Research Council (ERC) under the European Union Horizon 2020 research and innovation programme (grant agreement No. 947978).

n-vertex graph with minimum degree at least $(1/2+\alpha)n$ contains a copy of every n-vertex tree with maximum degree at most $cn/\log n$. This is tight up to the constant c. In this paper, we will prove the corresponding version of this result for *directed graphs (digraphs)*.

The *minimum semidegree* of a digraph D, denoted by $\delta^0(D)$, is the smallest in- or out-degree over the vertices in D, that is, $\delta^0(D) = \min_{v\in V(D),\diamond\in\{+,-\}} d^\diamond(v)$. Ghouila-Houri [7] solved the minimum semidegree problem for the directed Hamilton cycle, showing that, if an n-vertex digraph D has $\delta^0(D) \geq n/2$, then it contains a directed Hamilton cycle. That is, an n-vertex cycle with the edges oriented in the same direction. DeBiasio, Kühn, Molla, Osthus and Taylor [5] showed that, when n is sufficiently large, this holds in fact for any n-vertex cycle with any orientations on its edges, except for when the edges change direction at every vertex around the cycle. This latter cycle, known as the *anti-directed Hamilton cycle*, is only guaranteed to appear if $\delta^0(D) \geq n/2+1$, as shown by DeBiasio and Molla [6].

Recently, Mycroft and Naia [16, 17] gave the first bound on the minimum semidegree required for the appearance of different spanning trees. Here, H is an oriented n-vertex tree, with some bound on the degree of its underlying (undirected) tree. Mycroft and Naia [16, 17] proved that, for each $\alpha, \Delta > 0$, there is some n_0 such that, if $n \geq n_0$, then every n-vertex digraph with minimum semidegree at least $(1/2+\alpha)n$ contains a copy of every oriented n-vertex tree T with $\Delta^\pm(T) \leq \Delta$. Moreover, their result holds for a slightly wider class of trees, allowing them to show that, for each $\alpha > 0$, almost every labelled oriented n-vertex tree appears in every n-vertex digraph with minimum semidegree at least $(1/2+\alpha)n$.

In this paper, we introduce new methods to embed oriented trees in digraphs, relaxing the maximum degree condition to give a full directed version of Komlós, Sárközy and Szemerédi's result, as follows.

Theorem 1.1. *For each $\alpha > 0$, there exists $c > 0$ and $n_0 \in \mathbb{N}$ such that the following holds for every $n \geq n_0$. Any n-vertex digraph D with $\delta^0(D) \geq (1/2+\alpha)n$ contains a copy of every oriented n-vertex tree T with $\Delta^\pm(T) \leq cn/\log n$.*

We note that the undirected version follows immediately from Theorem 1.1. Indeed, given any n-vertex tree T and an n-vertex graph G, we can apply Theorem 1.1 to a copy of T with each edge oriented arbitrarily and a digraph formed from G by replacing each edge uv with an edge from u to v and an edge from v to u. This demonstrates that, as with Komlós, Sárközy and Szemerédi's result, Theorem 1.1 is tight up to the constant c. Furthermore, through Theorem 1.1 we give a new proof of the undirected result without using Szemerédi's regularity lemma, in contrast to the work of both Komlós, Sárközy and Szemerédi [10], and Mycroft and Naia [16, 17]. Key to our result is to use a random embedding of part of the tree using 'guide sets' and embedding many leaves (and small subtrees) of the tree using 'guide graphs'. This replaces the regularity methods of [10, 16, 17], and is sketched in Section 2, where we also outline the rest of this paper.

2 Preliminaries

2.1 Notation

Let D be a digraph. We denote by $V(D)$ and $E(D)$ the vertex set and edge set of D, respectively, where every element of the edge set of D is an ordered pair of vertices. We let $|D| = |V(D)|$, which we call the *size* of D, and let $e(D) = |E(D)|$. Letting $u, v \in V(D)$, if $uv \in E(D)$, then we say that u is an *in-neighbour* of v and v is an *out-neighbour* of u. Denote by $N^-_D(v)$ and $N^+_D(v)$, respectively, the set of all in- and out-neighbours of v. We let $d^-_D(v) = \left|N^-_D(v)\right|$ and $d^+_D(v) = \left|N^+_D(v)\right|$, and we refer to these as the *in-* and *out-degree of* v, respectively. For each

$\diamond \in \{+,-\}$, we let $\delta^{\diamond}(D)$ and $\Delta^{\diamond}(D)$ be, respectively, the *minimum* and *maximum* $\diamond$*-degree* of D. For any $A, B \subseteq V(D)$, and each $\diamond \in \{+,-\}$, let $N^{\diamond}_D(A,B) = \bigcup_{a\in A}(N^{\diamond}_D(a) \cap B)$, and let $d^{\diamond}_D(A,B) = |N^{\diamond}_D(A,B)|$. We omit the subscript when the graph is clear from context. Note that, for simplicity of notation, we use '$-$' and 'in' interchangeably, and, similarly, we use '$+$' and 'out' interchangeably. We use '$\pm$' to represent that a property holds for both '$-$' and '$+$'.

Suppose that A and B are disjoint subsets of $V(D)$. We write $D[A]$ to mean D induced on the set A, that is, the graph obtained from D by deleting all vertices which are not in A. For each $\diamond \in \{+,-\}$, a $\diamond$*-matching* from A into B is a set of vertex-disjoint edges such that every edge in the set has one endpoint in A and one endpoint in B, and the endpoint in B is a $\diamond$-neighbour of the endpoint in A, that is, every edge is a $\diamond$-edge from A into B. We say this matching *covers* A if every vertex of A belongs to some edge in the matching, and we call this a *perfect* $\diamond$*-matching* if it covers both A and B. A *bare path* of length m in a tree is a path with m edges such that each of the internal vertices have degree 2 in the tree. When P is a path in D, we let $D - P$ denote the subgraph of D obtained by removing the internal vertices of P.

For any $n \in \mathbb{N}$, we let $[n] := \{1,\dots,n\}$. In order to simplify notation, we use hierarchies to state our results. That is, for $a, b \in (0,1]$, whenever we write that a statement holds for $a \ll b$ (or $b \gg a$), we mean that there exists a non-decreasing function $f\colon (0,1] \to (0,1]$ such that the statement holds whenever $a \le f(b)$. We define similar expressions with multiple variables analogously. We say a random event occurs *with high probability* if the probability of the event occurring tends to 1 as n tends to infinity. In our proofs, when we have shown that a property holds with high probability, we will implicitly assume that this property holds from that point onwards. For simplicity, we ignore floors and ceilings wherever this does not affect the argument.

2.2 Proof sketch

When $1/n \ll c \ll \alpha$, we will embed any oriented n-vertex tree T with $\Delta^{\pm}(T) \le cn/\log n$ into any n-vertex digraph D with $\delta^0(D) \ge (1/2+\alpha)n$. We embed T using the absorption method, an approach first introduced in general by Rödl, Ruciński and Szemerédi [18] which has been effective on a range of embedding problems for spanning graphs and digraphs (see, for example, the survey [3]). We first partially embed a subtree T'' of T into a set A such that, given any subset $B \subset V(D)$ with $A \subset B$ and $|B| = |T''|$, we can complete this embedding of T'' into $D[B]$ (see Theorem 2.1).

We then use an almost-spanning embedding to embed the vertices in $V(T) \setminus V(T'')$ to extend the partial embedding of T'' (see Theorem 2.2). We will have chosen T'' so that in this stage a tree, called T', is attached to an embedded vertex of T''. Using the property of the partial embedding of T'', we then complete the embedding of T'' with the unused vertices in D. The decomposition of T that we need follows from a simple proposition (Proposition 2.3).

In Section 2.2.1, we state these three results, Theorem 2.1, Theorem 2.2 and Proposition 2.3, before deducing Theorem 1.1 from them. In Section 2.2.2, we then discuss in detail the proof of Theorem 2.2, which is the major challenge overcome by this paper.

In the rest of Section 2, we restate the probabilistic tools we will use, and give a basic structural decomposition of trees and some simple results on matchings. In Section 3, we prove Theorem 2.2. In Section 4, we prove Theorem 2.1.

2.2.1 Main tools and deduction of Theorem 1.1

For Theorem 1.1, we will first find a suitable subtree $T'' \subset T$ and a set $A \subset V(D)$ with slightly fewer than $|T''|$ vertices, so that, given any set B of $|T''|$ vertices containing A, we can embed

T'' in $D[B]$. Furthermore, we will ensure that some pre-specified vertex $t \in V(T'')$ is always embedded to some fixed vertex $v \in A$, as follows.

Theorem 2.1. *Let $1/n \ll c \ll \varepsilon \ll \mu \ll \alpha$. Let D be an n-vertex digraph with minimum semidegree at least $(1/2+\alpha)n$. Let T be an oriented tree with μn vertices and $\Delta^{\pm}(T) \leq cn/\log n$, and let $t \in V(T)$.*

Then, $V(D)$ contains a vertex set A with size $(\mu - \varepsilon)n$ containing a vertex $v \in A$ such that the following holds. For any set $B \subset V(D)$ with $A \subset B$ and $|B| = \mu n$, $D[B]$ contains a copy of T in which t is copied to v.

Theorem 2.1 is proved in Section 4 by randomly embedding most of T and taking A to be the image of this embedding. We then show that the partial embedding of T can be extended using any new vertex in $y \in V(D) \setminus A$ by switching y into the partial embedding in place of some vertex in A that can instead be used to embed a new vertex of T. Repeatedly doing this will allow the embedding of T to be completed using any set of $|T| - |A|$ new vertices in $V(D) \setminus A$. This is sketched in more detail at the start of Section 4, before Theorem 2.1 is proved.

We will embed the majority of the tree for Theorem 1.1, using the following almost-spanning embedding.

Theorem 2.2. *Let $1/n \ll c \ll \varepsilon, \alpha$. Let D be an n-vertex digraph with minimum semidegree at least $(1/2 + \alpha)n$ and let $v \in V(D)$. Let T be an oriented tree with at most $(1 - \varepsilon)n$ vertices and $\Delta^{\pm}(T) \leq cn/\log n$, and let $t \in V(T)$.*

Then, D contains a copy of T in which t is copied to v.

Using in addition the following simple proposition (see, for example, [15, Proposition 3.22]), we can now deduce Theorem 1.1.

Proposition 2.3. *Let $n, m \in \mathbb{N}$ satisfy $1 \leq m \leq n/3$. Given any n-vertex tree T, there are two edge-disjoint trees $T_1, T_2 \subset T$ such that $E(T_1) \cup E(T_2) = E(T)$ and $m \leq |T_2| \leq 3m$.*

Proof of Theorem 1.1 from Theorems 2.1 and 2.2. Let $\varepsilon, \mu > 0$ be such that $c \ll \varepsilon \ll \mu \ll \alpha$. Let D be an n-vertex digraph with $\delta^0(D) \geq (1/2 + \alpha)n$. Let T be an oriented n-vertex tree with $\Delta^{\pm}(T) \leq cn/\log n$.

Using Proposition 2.3 with $m = \mu n$, find edge-disjoint trees $T', T'' \subset T$ such that $E(T') \cup E(T'') = E(T)$ and $\mu n \leq |T''| \leq 3\mu n$. Let t be the vertex which is in both T' and T''. By Theorem 2.1 applied with $\mu' = |T''|/n$, there is a set $A \subset V(D)$ such that $|A| = |T''| - \varepsilon n$, and a vertex $v \in A$ such that, for any set $B \subset V(D)$ with $A \subset B$ and $|B| = |T''|$, $D[B]$ contains a copy of T'' in which t is copied to v.

Let $D' = D - (A \setminus \{v\})$. Let $n' = |D'|$, so that $(1 - 3\mu)n \leq n' \leq n$. Let α' be such that D' has minimum semidegree $(1/2 + \alpha')n'$. Note that $(1/2 + \alpha')n \geq (1/2 + \alpha')n' \geq (1/2 + \alpha - 3\mu)n$, so that $\alpha' \geq \alpha/2$. Furthermore, $n' = n - |T''| + \varepsilon n + 1 = |T'| + \varepsilon n$, and therefore

$$\frac{|T'|}{n'} = \frac{|T'|}{|T'| + \varepsilon n} \leq \frac{|T'|}{|T'|(1 + \varepsilon)} \leq 1 - \varepsilon/2.$$

Thus, by Theorem 2.2, we can find a copy, S' say, of T' in D' in which t is copied to v. By applying the property of A from Theorem 2.1, we can then find a copy of T'' in $D - (V(S') \setminus \{v\})$ in which t is copied to v. Together, these give us a copy of T. □

2.2.2 Proof Sketch of Theorem 2.2

We will embed a $(1-\varepsilon)n$-vertex tree T for Theorem 2.2 by dividing most of T into a small core forest $T_0 \subset T$ and a collection of constant-sized subtrees, which are either attached to T_0 by a single edge or by two short paths. It is the trees attached to T_0 by a single edge that will be the most challenging to embed, and so we dedicate most of our attention in the proof sketch to this.

More precisely, we will find a tree $T' \subset T$, containing a *core forest* $T_0 \subset T'$ and vertex-disjoint trees $S_1, \dots, S_\ell \subset T' - V(T_0)$, for some $\ell \in \mathbb{N}$, such that T' is formed from T_0 by, for each $i \in [\ell]$,

(1) either adding S_i to T_0 using two bare paths with length 2,

(2) or adding S_i to T_0 with a single edge.

Furthermore, for some $\mu > 0$ and $K \in \mathbb{N}$, with $1/n \ll 1/K, \mu \ll \alpha, \varepsilon$, we will have that

- $|T_0| \leq \mu n$ (i.e., T_0 is small),
- $|T'| \geq |T| - \mu n$ (i.e., T' is most of T),
- there are at most μn trees S_i which are in Case (1), and
- each tree S_i has at most K vertices.

In Case (1), we say S_i is added to T_0 as a path, and in Case (2) we say S_i is added to T_0 as a leaf. The crux of our method is to embed T_0 along with the trees S_i in Case (2) connected to the embedding of T_0 by the appropriate edge. This is encapsulated in the following lemma, which is proved in Section 3.1.

Lemma 2.4. *Let $1/n \ll c \ll \mu \ll \alpha, \varepsilon$, let $c \ll 1/K$ and let $\ell \in \mathbb{N}$. Suppose D is an n-vertex digraph with $\delta^0(D) \geq (1/2+\alpha)n$ and $v \in V(D)$.*

Suppose that T is an oriented tree with $|T| \leq (1-\varepsilon)n$ and $\Delta^{\pm}(T) \leq cn/\log n$. Suppose that $T', S_1, \dots S_\ell \subset T$ are vertex-disjoint subtrees with $|T'| \leq \mu n$, and $|S_i| \leq K$ for each $i \in [\ell]$. Suppose that T is formed from T' by attaching each S_i, $i \in [\ell]$, to T' by an edge. Finally, let $t \in V(T')$.

Then, D contains a copy of T in which t is copied to v.

We will now briefly sketch how Theorem 2.2 can be proved from Lemma 2.4. Let m be the total number of vertices that appear in the trees S_i in Case (1) above. To embed these trees, we use the fact that two random sets in D of the same (linear) size are likely to have a perfect matching from one to the other (see Proposition 2.12). Taking $p \gg 1/n$ and $Kp \leq 1$, we can, with high probability, find pn copies of an oriented tree with K vertices in a random set of Kpn vertices in D by taking randomly K disjoint subsets within this set of size pn and finding appropriate matchings between them (see Section 2.5). Collecting isomorphic trees S_i together, and applying this to each of the constantly many (depending on K) isomorphism classes, allows us to embed the trees S_i in Case (1) with high probability in a random set with size $m + \varepsilon n/4$. Here, the extra $\varepsilon n/4$ vertices allow us to find a linear number of trees in each isomorphism class by finding some additional trees if required.

Thus, in a partition of $V(D)$ into sets $V_1 \cup V_2 \cup V_3 \cup V_4$ chosen uniformly at random so that $|V_1| = n - m - 3\varepsilon n/4$, $|V_2| = m + \varepsilon n/4$, $|V_3| = |V_4| = \varepsilon n/4$, with high probability, the following occur.

- $\delta^{\pm}(D[V_1]) \geq (1/2 + \alpha/2)|V_1|$, so that, applying Lemma 2.4, we can embed T_0 along with the trees S_i in Case (2) connected to the embedding of T_0 by the appropriate edge.

- We can embed the trees S_i in Case (1) in $D[V_2]$.
- Then, using that there are at most μn trees in Case (1), we can greedily attach them to the embedding of T_0 using two paths with length 2 whose interior vertex is an unused vertex in V_3 (see Section 3.2).
- Finally, as $|T| - |T'| \leq \mu n$, we can greedily extend the resulting embedding of T' to one of T, by adding a sequence of leaves using vertices in V_4 (see Section 3.3).

Here, the last two steps are (with high probability) possible using the semi-degree condition of D. Note that, as $\mu \ll \varepsilon$, we only embed a small proportion of vertices into V_3 and V_4.

We will now give a detailed proof sketch of Lemma 2.4.

Proof sketch of Lemma 2.4

To simplify our discussion, let us assume that each tree S_i in Lemma 2.4 consists of only a single vertex, which is an out-neighbour in the tree T of a vertex of T_0, and that every vertex in T_0 is attached to exactly one such tree. That is, T consists of T_0 with an out-matching attached. Our embedding of T_0 is randomised, which will allow the methods described to be used to find matchings attached from different subsets of the image of the embedding of T_0 to different random sets. This will allow the embedding below for T_0 to be used for the general case.

Let us detail the example situation precisely. Suppose we have a μn-vertex tree T_0 and choose two disjoint random sets $V_0, V_1 \subset V(D)$ with size $p_0 n$ and $p_1 n$ respectively, where $p_0 \gg \mu$ and $p_1 = (1 + o(1))\mu$. We will randomly embed T_0 into V_0, so that there is an out-matching from the vertex set of the embedding of T_0 into V_1. Note that there are many spare vertices in V_0, possible as in the general case we embed T_0 once. However, as we then find (potentially) many different matchings, we need to do this with few spare vertices, and therefore use most of the vertices in V_1 (as p_1 is only a little larger than μ).

We will embed T_0 vertex-by-vertex, say in order $t_1, \ldots, t_\ell$, so that each new vertex is embedded as an in- or out-leaf of the previously embedded subtree. Having chosen the random sets V_0, V_1, and before beginning the embedding, we will find *guide sets* $A_{v,\diamond} \subset N^{\diamond}_D(v, V_0)$, $v \in V_0$ and $\diamond \in \{+, -\}$, which we use to guide the random embedding. We then start the random embedding, under the rule that if, for some $v \in V_0$ and $\diamond \in \{+, -\}$, we are attaching a $\diamond$-edge as a leaf to v, then we choose this leaf uniformly at random from the unused vertices in $A_{v,\diamond}$.

The guide sets ensure that, with high probability, there will be a matching from the embedding of T_0 into V_1. These guide sets are found using Lemma 3.5, and they exist (with high probability for the choice of V_0, V_1) due to the semi-degree condition in D. Essentially, for some constants β, γ, we find, for each $v \in V(D)$ and $\diamond \in \{+, -\}$, a set $A_{v,\diamond} \subset N^{\diamond}_D(v, V_0)$ with size βn and bipartite digraphs $H^{\circ}_{v,\diamond} \subset D^{\circ}[A_{v,\diamond}, V_1]$, $\circ \in \{+, -\}$, so that in $H^{\circ}_{v,\diamond}$ each vertex in $A_{v,\diamond}$ has around $\gamma p_1 n$ $\circ$-neighbours in V_1, and each vertex in V_1 has around $\gamma \beta n$ $\circ$-edges leading into it. That is, $H^{\circ}_{v,\diamond}$ is approximately regular on each side with edge density approximately γ.

The guide graphs $H^{+}_{v,\diamond}$ can be used to find the matching from the embedding of T_0 as follows. When a vertex t_i is embedded using a guide set $A_{v_i,\diamond_i}$, to some vertex s_i say, we add the edges in $H^{+}_{v_i,\diamond_i}$ adjacent to s_i to an auxiliary graph K – note that approximately $\gamma p_1 n$ edges are added next to s_i. Note further that, as most of the vertices in $A_{v_i,\diamond_i}$ will be unused, each $w \in V_1$ will have an edge added from s_i to w with probability approximately

$$\frac{d^{-}_{H^{+}_{v_i,\diamond_i}}(w)}{|A_{v_i,\diamond_i}|} \approx \frac{\gamma \beta n}{\beta n} = \gamma. \tag{1}$$

When this is complete, K is a bipartite digraph with vertex classes $\{s_1, \dots, s_\ell\}$ and V_1. Each vertex s_i will have out-degree approximately $\gamma p_1 n$, and, due to the randomness of the embedding and (1), each vertex in V_1 will have in-degree which is approximately $\gamma\ell = \gamma|T_0| \approx \gamma p_1 n$.

Thus, K will be a bipartite graph with the in-degrees in one vertex class approximately equal to the out-degrees in the other. Via Hall's matching criterion, an out-matching will exist from $\{s_1, \dots, s_\ell\}$ to V_1 which covers most of the vertices in $\{s_1, \dots, s_\ell\}$. By ensuring that V_1 is likely to be a little larger than ℓ, we in fact will get with high probability that such an out-matching can cover $\{s_1, \dots, s_\ell\}$.

Note that, in the sketch above, we do not use the graph $H^-_{v,\diamond}$. However, in practice, we find such guide sets and guide graphs with $V_1 = V(D) \setminus V_0$ (see Lemma 3.3), before taking random subsets of V_1. We will find out-matchings into some of these random sets, and in-matchings into some others. Therefore, it is important to have both guide graphs $H^-_{v,\diamond}$ and $H^+_{v,\diamond}$, and, furthermore, that the same set $A_{v,\diamond}$ is used for both graphs.

Finally, let us note where the condition $\Delta^\pm(T) \leq cn/\log n$ is used in our proof of Lemma 2.4. In the sketch above the set V_1 will always have size which is linear in n, but we may need to attach the trees in Lemma 2.4 to few vertices in T. The maximum in- or out-degree condition on T ensures, that, if the trees S_i in Lemma 2.4 together comprise linearly (in n) many vertices in T, then they are attached to at least $C \log n$ different vertices, for some large constant C, which gives us sufficient probability concentrations when these vertices are randomly embedded for the corresponding versions of Hall's criterion to hold (see the proof of Claim 3.7).

2.3 Probabilistic tools

Let $n, m, k \in \mathbb{N}$ be such that $\max\{m, k\} \leq n$. Let A be a set of size n, and $B \subseteq A$ be such that $|B| = m$. Let A' be a uniformly random subset of A of size k. Then the random variable $X = |A' \cap B|$ is said to have hypergeometric distribution with parameters n, m and k, which we denote by $X \sim \mathrm{Hyp}(n, m, k)$. We will use the following Chernoff-type bound.

Lemma 2.5 (see, for example, [9]). *Suppose $X \sim \mathrm{Hyp}(n, m, k)$. Then for any $0 < \alpha < 3/2$, we have*

$$\mathbb{P}\left[|X - \mathbb{E}[X]| \geq \alpha\mathbb{E}[X]\right] \leq 2\exp\left(\alpha^2\mathbb{E}[X]/3\right).$$

A sequence of random variables $(X_i)_{i\geq 0}$ is a martingale if $\mathbb{E}[X_{i+1} \mid X_0, \dots, X_i] = X_i$ for each $i \geq 0$. We will use the following Azuma-type bound for martingales.

Lemma 2.6 (see, for example, [1]). *Let $(X_i)_{i\geq 0}$ be a martingale and let $c_i > 0$ for each $i \geq 1$. If $|X_i - X_{i-1}| < c_i$ for each $i \geq 1$, then, for each $n \geq 1$,*

$$\mathbb{P}[|X_n - X_0| \geq t] \leq 2\exp\left(-\frac{t^2}{\sum_{i=1}^n c_i^2}\right).$$

We will use this bound for supermartingales and submartingales. A sequence of random variables $(X_i)_{i\geq 0}$ is a supermartingale if $\mathbb{E}[X_{i+1} \mid X_0, \dots, X_i] \leq X_i$ for each $i \geq 0$, and a submartingale if $\mathbb{E}[X_{i+1} \mid X_0, \dots, X_i] \geq X_i$ for each $i \geq 0$. The bound on the upper tail in Lemma 2.6 holds for supermartingales, while the bound on the lower tail holds for submartingales.

2.4 Structural lemmas

In this section we decompose undirected trees. Note that we will later apply this to directed trees as the edge directions do not affect the decompositions. We will use the following simple but useful lemma (see [15, Lemma 4.1]) which tells us that either a tree has many leaves, or it has many bare paths.

Lemma 2.7. *Let $t, m \geq 2$, and suppose that T is a tree with at most t leaves. Then there is some s and some vertex-disjoint bare paths P_i, $i \in [s]$, in T with length m so that $|T - P_1 - \cdots - P_s| \leq 6mt + 2|T|/(m+1)$.*

We can now prove the following key lemma, in which we decompose a tree for our embedding.

Lemma 2.8. *Let $0 \ll 1/n \ll 1/K \ll 1/k \ll \eta$. Let T be a tree on n vertices with $t \in V(T)$. Then, T contains forests $T_0 \subset T_1 \subset T_2 \subset T_3 = T$, such that T_2 is a tree, and the following hold.*

- **P1** *$|T_0| \leq \eta n$ and $t \in V(T_0)$.*
- **P2** *T_1 is formed from T_0 by the vertex-disjoint addition of trees, S_v, $v \in V(T_0)$, so that, for each $v \in V(T_0)$, $S_v - v$ is a forest of trees with size at most K.*
- **P3** *T_2 is formed from T_1 by the addition of trees with size at least k and at most K attached to T_1 with exactly two bare paths of length 2.*
- **P4** *$|T_3| - |T_2| \leq \eta n$.*

Proof. Let ε satisfy $1/K \ll \varepsilon \ll 1/k$, and let $S_0 = T$. Do the following for $i = 0, 1, 2 \ldots$ as far as possible, where a set of independent leaves is a set of leaves which pairwise have no common neighbours in the tree. If S_i has at least εn independent leaves, the set L_i say, then remove $L_i \setminus \{t\}$ from S_i to get the tree S_{i+1}. Suppose this finishes with S_ℓ, which does not have at least εn independent leaves. Note that $\ell \leq 1/\varepsilon + 1$. We will show the following claim.

Claim 2.9. *To get from T to S_ℓ, for each $v \in V(S_\ell)$, there is a tree removed from v which has at most 2^ℓ vertices.*

Proof of Claim 2.9. We will show by induction on $i = 0, 1, \ldots, \ell$, that, to get from $S_{\ell-i}$ to S_ℓ, for each $v \in V(S_\ell)$, there is a tree removed from v which has at most 2^i vertices. Thus the claim follows when $i = \ell$. Note that this is trivially true for $i = 0$ and label the tree removed from $v \in V(S_\ell)$ to get from $S_{\ell-i}$ to S_ℓ as $T_{v,i}$.

Now, let $0 \leq i < \ell$, and assume that $|T_{v,i}| \leq 2^i$ for each $v \in V(S_\ell)$. As we remove a set of independent leaves from $S_{\ell-i-1}$ to get to $S_{\ell-i}$, for each $v \in V(S_\ell)$, we remove a set of independent leaves of $T_{v,i+1}$ to get $T_{v,i}$. Therefore, for each $v \in V(S_\ell)$, $|T_{v,i+1}| \leq 2|T_{v,i}| \leq 2^{i+1}$, as required. □

Let $L(S_\ell)$ be the set of leaves of S_ℓ. Remove $L(S_\ell) \setminus \{t\}$ and call the resulting tree S'. Note that, as S_ℓ does not have at least εn independent leaves, S' does not have at least εn leaves. Thus, by Lemma 2.7, for some $m \leq n/(k+1)$, S' contains vertex disjoint bare paths $P_1, \ldots, P_m$ with length k such that $t \notin V(P_i)$ for each $i \in [k]$ and

$$|S' - P_1 - \cdots - P_m| \leq 6k \cdot \varepsilon n + 2n/(k+1) + k + 1 \leq \eta n/4. \tag{2}$$

For each path P_i, $i \in [m]$, if possible, find within P_i a path P'_i with length at least $k - 2\eta^3 k$, such that, labelling its endvertices x_i and y_i the following hold.

- (i) Each of x_i and y_i had a tree with size at most $\eta k/4$ removed from them in T to reach S'.
- (ii) Letting Q_i be the component of $T - \{x_i, y_i\}$ containing $P'_i - \{x_i, y_i\}$, we have $|Q_i| \leq K$.

Say, with relabelling, these paths are $P'_1, \dots, P'_{m'}$. We will show that $m' \geq m - \eta n/2k$. Note first that the number of $i \in [m]$ with no vertices x_i and $y_i \in V(P_i)$ respectively within $\eta^3 k$ of the two endvertices of P_i, so that each of x_i and y_i had a tree with at most $\eta k/4$ vertices deleted from them, is at most $n/(\eta^3 k \cdot \eta k/4) \leq \eta n/4k$. Note further that the number of $i \in [m]$ with at least K vertices in $V(P_i)$ or in a component of $T - E(P_i)$ containing an interior vertex of P_i is at most $n/K \leq \eta n/4k$. Therefore, we can find such a path P'_i for all but at most $\eta n/2k$ values of $i \in [m]$, so that $m' \geq m - \eta n/2k$.

Letting $T_0 = S' - P'_1 - \dots - P'_{m'}$, we will now show that $|T_0| \leq \eta n$. Note that, for each $i \in [m']$, $|V(P_i) \setminus V(P'_i)| \leq 2\eta^3 k$. Therefore, as $m \leq n/k$,

$$|T_0| \leq |S' - P_1 - \dots - P_m| + k \cdot \eta n/2k + m \cdot 2\eta^3 k \overset{(2)}{\leq} \eta n.$$

Furthermore, clearly $t \in V(T_0)$, and thus **P1** holds.

Note that, by Claim 2.9, for each $i \in [m']$, $|Q_i| \leq k2^\ell \leq K$. For each $v \in V(T_0)$, let R_v be the tree containing v in $T[(V(T) \setminus V(T_0)) \cup \{v\}]$, without any of the x_i, y_i as neighbours. Now, by Claim 2.9, $R_v - v$ consists of trees with at most $2^\ell \leq K$ vertices. Let $T_1 = T_0 \cup (\bigcup_{v \in V(T_0)} R_v)$. Thus, **P2** holds.

Let $T_3 = T$ and let T_2 be $T[V(T_1) \cup (\bigcup_{i \in [m']} (\{x_i, y_i\} \cup V(Q_i)))]$. Note that **P3** holds by construction, and as $|Q_i| \leq K$ for each $i \in [m']$. Furthermore, the only missing vertices from T are those in $R_v - v$, for each $v \in \{x_i, y_i : i \in [m']\}$, and thus T_2 is a tree. For each such v, $|R_v| \leq \eta k/4$ by (i). Therefore, $|T_3| - |T_2| \leq (n/k) \cdot (2\eta k/4) \leq \eta n$, and hence **P4** holds. □

2.5 Matchings between random sets

With high probability, any random subset of vertices in the digraph in Theorem 1.1 satisfies a similar minimum semidegree condition, as follows.

Lemma 2.10. *0Let $1/n \ll c, \alpha$, and suppose D is an n-vertex digraph with $\delta^0(D) \geq (1/2 + \alpha)n$. Let $A \subseteq V(D)$ be chosen uniformly at random subject to $|A| = cn$. Then, with high probability, for every vertex $v \in V(D)$, we have $\left|N_D^{\pm}(v, A)\right| \geq (1/2 + \alpha/2)\,|A|$.*

Proof. Let v be an arbitrary vertex of D and let $A \subseteq V(D)$ be a uniformly random subset with $|A| = cn$. For $\diamond \in \{+, -\}$, we let $Z_v^\diamond$ be the random variable which measures $|N^\diamond(v) \cap A|$. Then $Z_v^\diamond$ has hypergeometric distribution with expectation

$$\mathbb{E}[Z_v^\diamond] = \frac{|N^\diamond(v)|\,|A|}{n} \geq \left(\frac{1}{2} + \alpha\right) cn.$$

Therefore, by Lemma 2.5, we have

$$\mathbb{P}\left[|Z_v^\diamond - \mathbb{E}[Z_x^\diamond]| > \frac{\alpha/2}{1/2 + \alpha}(1/2 + \alpha)cn\right] \leq 2\exp\left(-\left(\frac{\alpha/2}{1/2 + \alpha}\right)^2 \frac{(1/2 + \alpha)cn}{3}\right)$$
$$= 2\exp\left(\frac{-\alpha^2 cn}{6 + 12\alpha}\right).$$

Then, applying a union bound, with probability at least $1 - 2n\exp\left(-\alpha^2 cn/(6 + 12\alpha)\right) = 1 - o(1)$, we have that $Z_v^\diamond \geq (1/2 + \alpha/2)\,|A|$ for each $\diamond \in \{+, -\}$ and $v \in V(D)$. □

The following digraph version of Hall's matching criterion implies a matching exists, as follows directly from the same result for undirected graphs.

Lemma 2.11. *Let D be a bipartite digraph with vertex classes A and B, and let $\diamond \in \{+,-\}$. Suppose that for every $S \subset A$, $|N^\diamond_D(S,B)| \geq |S|$. Then there is a $\diamond$-matching from A into B which covers A.*

We will refer to the condition in Lemma 2.11 as *Hall's criterion*. In combination with Lemma 2.10, Lemma 2.11 shows that with high probability there is a perfect matching between a large random pair of disjoint equal-sized vertex subsets in the digraph, as follows.

Proposition 2.12. *Let $1/n \ll p, \alpha$, and suppose D is an n-vertex digraph with $\delta^0(D) \geq (1/2+\alpha)n$. Let A, B be chosen uniformly at random from all disjoint pairs of subsets of $V(D)$, each with size pn, and let $\diamond \in \{+,-\}$. Then, with high probability, there is a perfect $\diamond$-matching from A into B.*

Proof. By Lemma 2.10, with high probability we can assume the following. For all $v \in A$, we have $|N^\pm(v,B)| \geq (1/2+\alpha/2)\,|B|$, and, for all $v \in B$, we have $|N^\pm(v,A)| \geq (1/2+\alpha/2)\,|A|$. We will now show that Hall's criterion holds.

Let $S \subseteq A$, such that $S \neq \emptyset$ and $|S| \leq (1/2+\alpha/2)pn$, and let $x \in S$. Then, $|N^\diamond(S,B)| \geq |N^\diamond(x,B)| \geq (1/2+\alpha/2)pn \geq |S|$, so Hall's condition is trivially satisfied. Now take $S \subseteq A$, $|S| > (1/2+\alpha/2)pn$, and assume for a contradiction that $|N^\diamond(S,B)| < |S|$. Then in particular, $B \setminus N^\diamond(S,B) \neq \emptyset$. Take $b \in B \setminus N^\diamond(S,B)$, and let $\circ \in \{+,-\}$ be such that $\circ \neq \diamond$. We have $|N^\circ(b,A)| \geq (1/2+\alpha/2)pn$. However, since $b \notin N^\diamond(S,B)$, we have $N^\circ(b,A) \cap S = \emptyset$. So,

$$pn = |A| \geq |N^\circ(b,A)| + |S| \geq (1/2+\alpha/2)pn + (1/2+\alpha/2)pn = (1+\alpha)pn > pn,$$

giving a contradiction. Thus, Hall's criterion is satisfied for all $S \subseteq A$ and so, since $|A| = |B|$, by Lemma 2.11, there is a perfect $\diamond$-matching from A into B. □

We use Proposition 2.12 to embed many vertex disjoint small trees, via the following two lemmas. In Lemma 2.13, we embed linearly many copies of a given constant-sized tree into specified subsets of our digraph. In Lemma 2.13, we embed a forest of constant-sized trees covering almost all the vertices in our digraph.

Lemma 2.13. *Let $1/n \ll 1/K, p, \alpha$ with $pK \leq 1$. Suppose T is an oriented K-vertex tree containing $t \in V(T)$. Let D be an n-vertex digraph with $\delta^0(D) \geq (1/2+\alpha)n$. Let V_1, V_2 be vertex disjoint subsets of $V(D)$ chosen uniformly at random subject to $|V_1| = pn$ and $|V_2| = (K-1)pn$.*

Then, with high probability, $D[V_1 \cup V_2]$ contains pn vertex disjoint copies of T, in which t is copied into V_1 in each copy of T.

Proof. Let $V_1 = U_1$, and let $U_2 \cup \cdots \cup U_K$ be a partition of V_2 chosen uniformly at random so that $|U_i| = pn$ for each $i \in \{2,\dots,K\}$. Note that the distribution of any pair of sets U_i, U_j with $1 \leq i < j \leq K$ is that of two disjoint vertex sets with size pn in $V(D)$, uniformly at random drawn from all such pairs.

Label the vertices of T by $t_1,\dots,t_K$ so that $t_1 = t$ and $T[\{t_1,\dots,t_i\}]$ is a tree for each $i \in \{1,\dots,K\}$. For each $i \in \{2,\dots,K\}$, let $j_i \in \{1,\dots,i-1\}$ be such that t_{j_i} is the in- or out-neighbour in $T[\{t_1,\dots,t_{i-1}\}]$ of the vertex t_i, and let $\diamond_i \in \{+,-\}$ be such that $t_i \in N^{\diamond_i}_T(t_{j_i})$.

Now by Proposition 2.12, for each $i \in \{2,\dots,K\}$, with high probability, we can find a $\diamond_i$-matching from U_{j_i} into U_i. By applying a union bound, we see that, with high probability, for every $i \in \{2,\dots,K\}$, there is a $\diamond_i$-matching, M_i say, from U_{j_i} into U_i.

Note that the union of these matchings, $\bigcup_{2\leq i\leq K} M_i \subset D[V_1 \cup V_2]$ is the disjoint union of pn copies of T, in which, for each $i \in [K]$, the copy of t_i is in V_i. Thus, in each of these pn copies of T, $t = t_1$ is copied into $V_1 = U_1$, as required. □

Lemma 2.14. *Let $1/n \ll 1/K, \varepsilon$, and suppose F is a digraph with at most $(1-\varepsilon)n$ vertices which is the disjoint union of trees with size at most K. Let D be an n-vertex digraph with $\delta^0(D) \geq (1/2+\alpha)n$. Then, with high probability, D contains a copy of F.*

Proof. Arrange the components of F into isomorphic classes of trees $\mathcal{R}_1, \ldots, \mathcal{R}_\ell$, noting that we may take $\ell \leq (2K)^{K-1}$. For each $i \in [\ell]$, let $t_i = |\mathcal{R}_i|$ and let s_i be the size of each component in $\mathcal{R}_i$. Uniformly at random, take, in $V(D)$, disjoint subsets $V_{i,1}$ and $V_{i,2}$, $i \in [\ell]$, with $|V_{i,1}| = p_i n$ and $|V_{i,2}| = (s_i - 1)p_i n$, where $p_i = t_i/n + \varepsilon/\ell s_i$, for each $i \in [\ell]$. Note that this is possible, since

$$\sum_{i=1}^{\ell} s_i p_i n = \sum_{i=1}^{\ell} \left(s_i t_i + \frac{\varepsilon n}{\ell} \right) \leq n.$$

For each $i \in [\ell]$, we can apply Lemma 2.13 to show that, with high probability, there are $p_i n$ copies of the underlying tree of $\mathcal{R}_i$ in $D_i = D[V_{i,1} \cup V_{i,2}]$. Since $p_i n \geq t_i$, this implies that with high probability, we can find a copy of $\mathcal{R}_i$ in D_i for each $i \in [\ell]$. By applying a union bound and using that $1/n \ll 1/\ell$, we have, with high probability, that there is a copy of F in D. □

3 Almost-spanning trees

The key aim of this section is to prove Theorem 2.2, that is, to prove we can embed an almost-spanning tree T in our digraph. By Lemma 2.8, we can find $T_0 \subset T_1 \subset T_2 \subset T_3 = T$, satisfying **P1** to **P4**. In Section 3.1, we show that we can embed T_1. In Section 3.2, we show that we can embed $T_2 \setminus T_1$, and $T_3 \setminus T_2$. We conclude in Section 3.3 by combining this to obtain an embedding of T.

3.1 Embedding constant-sized trees as stars

As sketched in Section 2.2, we will embed T_0 randomly, leaf by leaf, using a guide set to embed each new vertex. Each guide set has an accompanying guide graph, which we later use to find a matching. The property of the guide graph that we use to find the matching is that it is *skew-bounded*, as follows.

Definition 3.1. *A digraph D with vertex sets $A, B \subset V(D)$ is* $(a, b, \diamond)$-skew-bounded *on (A, B) if $d^{\diamond}_D(v, B) \geq a$ for each $v \in A$ and $d^{\circ}_D(v, A) \leq b$ for each $v \in B$, where $\circ \in \{+, -\}$ and $\circ \neq \diamond$.*

This property can imply a matching exists via Hall's criterion, as follows.

Proposition 3.2. *Let $a \geq b$ and $\diamond \in \{+, -\}$. Suppose D is a digraph containing disjoint vertex sets $A, B \subset V(D)$, such that D is $(a, b, \diamond)$-skew-bounded on (A, B). Then, there is a $\diamond$-matching from A into B in D.*

Proof. Let $U \subset A$. As D is $(a, b, \diamond)$-skew-bounded on (A, B), there are at least $a\,|U|$ and at most $b\,|N^{\diamond}_D(U, B)|$ $\diamond$-edges from U to $N^{\diamond}_D(U, B)$. Thus, $|N^{\diamond}_D(U, B)| \geq a\,|U|\,/b \geq |U|$. Therefore, by Lemma 2.11, there is a $\diamond$-matching from A into B. □

In the following lemmas, we find our guide sets and guide graphs. We start by finding in D, for each $v \in V(D)$ and $\diamond \in \{+, -\}$, a guide set A and guide graphs which are skew-bounded on $(A, V(D))$.

Lemma 3.3. *Let $1/n \ll \varepsilon \ll \alpha, \eta \leq 1$ and $1/n \ll \mu \leq \alpha^2/2$. Let D be an n-vertex digraph with $\delta^0(D) \geq (1/2+\alpha)n$, let $v \in V(D)$ and let $\diamond \in \{+,-\}$.*

Then, there is a set $A \subset N^{\diamond}_D(v)$ with size μn and digraphs $H^+, H^- \subset D$ such that, for each $\circ \in \{+,-\}$, H° is $(\varepsilon n, (1+\eta)\mu\varepsilon n, \circ)$-skew-bounded on $(A, V(D))$.

Proof. We start by showing that we can label the vertices of $V(D)$ as $V(D) = \{x_1, \ldots, x_n\} = \{y_1, \ldots, y_n\}$ so that, for each $i \in [n]$,

$$|N^-_D(x_i) \cap N^{\diamond}_D(v) \cap N^+_D(y_i)| \geq \alpha^2 n. \tag{3}$$

To do this, create an auxiliary graph, as follows. For each $w \in V(D)$, create distinct new vertices w^- and w^+, and let $V^+ = \{w^+ : w \in V(D)\}$ and $V^- = \{w^- : w \in V(D)\}$. Consider the auxiliary bipartite graph H with vertex set $V^+ \cup V^-$, where for each $x, y \in V(D)$, there is an edge between x^+ and y^- if and only if $\left|N^-_D(x) \cap N^{\diamond}_D(v) \cap N^+_D(y)\right| \geq \alpha^2 n$.

Claim 3.4. *$\delta(H) \geq (1/2 + \alpha/2)n$.*

Proof of Claim 3.4. Let $x \in V(D)$. We have $|N^-_D(x) \cap N^{\diamond}_D(v)| \geq n - (n - d^-_D(x)) - (n - d^{\diamond}_D(v)) \geq 2\alpha n$. Let $B = N^-_D(x) \cap N^{\diamond}_D(v)$ and $Y = \{y \in V(D) : |N^+_D(y) \cap B| \geq \alpha^2 n\}$, and note that $d_H(x^+) = |Y|$.

For each $u \in B$, we have $|N^-_D(u)| \geq (1/2+\alpha)n$, and thus $e_D(V(D), B) \geq (1/2+\alpha)|B|n$. By the choice of Y, we have $e_D(V(D), B) \leq |Y||B| + \alpha^2 n^2$. Therefore, as, in addition, $2\alpha n \leq |B|$, we have

$$(1/2+\alpha)|B|n \leq |Y||B| + \alpha^2 n^2 \leq |Y||B| + \alpha|B|n/2.$$

Thus, $(1/2+\alpha/2)|B|n \leq |Y||B|$, so that $|Y| \geq (1/2+\alpha/2)n$. Therefore, $d_H(x^+) = |Y| \geq (1/2+\alpha/2)n$.

A similar argument, with the signs reversed, shows that $d_H(y^-) \geq (1/2+\alpha/2)n$ for each $y \in V(D)$, completing the proof of the claim. □

As in the proof of Proposition 2.12, Claim 3.4 easily implies that Hall's criterion is satisfied, so that there is a matching from V^+ to V^- in H. That is, we can label the vertices of $V(D)$ as $V(D) = \{x_1, \ldots, x_n\} = \{y_1, \ldots, y_n\}$ so that, for each $i \in [n]$, (3) holds.

We will now show by induction that, for each $0 \leq i \leq \mu n$, there is a set $A_i \subset N^{\diamond}_D(v)$ with size i and graphs $H^+_i, H^-_i \subset D$ such that, for each $\circ \in \{+,-\}$, H°_i is $(\varepsilon n, (1+\eta)\mu\varepsilon n, \circ)$-skew-bounded on $(A_i, V(D))$, $e(H^\circ_i) = i\varepsilon n$, and, for each $j \in [n]$, $d^-_{H^+_i}(x_j) = d^+_{H^-_i}(y_j)$.

Note that if $A_0 = \emptyset$ and if H^+_0, H^-_0 have no edges and vertex set $V(D)$, then the conditions hold, so assume that $0 \leq i < \mu n$ and we have $A_i \subset N^{\diamond}_D(v)$ and $H^+_i, H^-_i \subset D$ as described.

Let $J_i \subset [n]$ be the set of $j \in [n]$ for which $d^-_{H^+_i}(x_j) = d^+_{H^-_i}(y_j) \leq (1+\eta/2)\mu\varepsilon n$. Note that, as $e(H^+_i) = e(H^-_i) = i\varepsilon n \leq \mu\varepsilon n^2$, we have

$$(n - |J_i|)(1+\eta/2)\mu\varepsilon n \leq \mu\varepsilon n^2.$$

Thus, as $\eta \leq 1$, $(n - |J_i|) \leq n/(1+\eta/2) \leq n(1-\eta/4)$, so that $|J_i| \geq \eta n/4$.

For each $j \in J_i$, let $W_{i,j} = (N^-_D(x_j) \cap N^{\diamond}_D(v) \cap N^+_D(y_j)) \setminus A_i$, noting that, by (3), $|W_{i,j}| \geq \alpha^2 n - i > \alpha^2 n - \mu n \geq \alpha^2 n/2$. By averaging, choose some $w_i \in V(D)$ such that

$$|\{j \in J_i : w_i \in W_{i,j}\}| \geq \frac{\sum_{j \in J_i} |W_{i,j}|}{n} \geq \frac{\eta n/4 \cdot \alpha^2 n/2}{n} \geq \varepsilon n,$$

using that $\alpha, \eta \gg \varepsilon$. Choose a set $J'_i \subset \{j \in J_i : w_i \in W_{i,j}\}$ with size εn. Let $A_{i+1} = A_i \cup \{w_i\}$. Let H^+_{i+1} be the digraph H^+_i with edges $w_i x_j$, $j \in J'_i$, added. Note that, as $d^-_{H^+_i}(x_j) \leq (1+\eta/2)\mu\varepsilon n$

for each $j \in J_i'$, H_{i+1}^+ is $(\varepsilon n, (1+\eta)\mu\varepsilon n, +)$-skew-bounded on $(A_{i+1}, V(D))$. Furthermore, by the definition of $W_{i,j}$, the edges added to H_i^+ are in D, and therefore $H_{i+1}^+ \subset D$.

Let H_{i+1}^- be the digraph H_i^- with the edges y_jw_i, $j \in J_i'$, added. Note that, similarly, H_{i+1}^- is $(\varepsilon n, (1+\eta)\mu\varepsilon n, -)$-skew-bounded on $(A_{i+1}, V(D))$. Finally, noting that A_{i+1} has size $i+1$, that $e(H_{i+1}^+) = e(H_{i+1}^-) = (i+1)\varepsilon n$ and that, for each $j \in [n]$, $d^-_{H_{i+1}^+}(x_j) = d^+_{H_{i+1}^-}(y_j)$, completes the inductive step, and hence the proof. □

We now show that the guide sets and guide graphs found by Lemma 3.3 have a similar skew-bounded property when restricted to random vertex subsets, as follows.

Lemma 3.5. *Let $1/n \ll \varepsilon \ll \alpha, \eta \leq 1$ and $1/n \ll 1/k, p_0, p_1, \ldots, p_k \leq 1$. Let $\mu = \alpha^2 p_0/4$. Let D be an n-vertex digraph with $\delta^0(D) \geq (1/2+\alpha)n$. Let $V_0, V_1, \ldots, V_k \subset V(D)$ be disjoint random sets chosen uniformly at random subject to $|V_i| = p_i n$ for each $i \in \{0, \ldots, k\}$.*

Then, with high probability, for each $v \in V(D)$ and $\diamond \in \{+,-\}$, there is a set $A_{v,\diamond} \subset N_D^\diamond(v) \cap V_0$ with size μn and digraphs $H^\circ_{v,\diamond} \subset D$, $\circ \in \{+,-\}$, such that, for each $\circ \in \{+,-\}$ and $i \in [k]$, $H^\circ_{v,\diamond}$ is $(\varepsilon p_i n, (1+\eta)\varepsilon\mu n, \circ)$-skew-bounded on $(A_{v,\diamond}, V_i)$.

Proof. By Lemma 3.3, applied with $\varepsilon' = (1+\eta/4)\varepsilon$, $\eta' = \eta/4$ and $\mu' = (1+\eta/4)\alpha^2/4$, for each $v \in V(D)$ and $\diamond \in \{+,-\}$, there is a set $\bar{A}_{v,\diamond} \subset N_D^\diamond(v)$ with size $(1+\eta/4)\alpha^2 n/4$ and digraphs $H^+_{v,\diamond}, H^-_{v,\diamond} \subset D$ such that, for each $\circ \in \{+,-\}$, $H^\circ_{v,\diamond}$ is $((1+\eta/4)\varepsilon n, (1+\eta/4)^3\varepsilon\alpha^2 n/4, \circ)$-skew-bounded on $(\bar{A}_{v,\diamond}, V(D))$.

Select $V_0, V_1, \ldots, V_k \subset V(D)$ according to the distribution in the lemma. Using Lemma 2.5, and a union bound, we have that, with high probability, the following hold.

Q1 For each $v \in V(D)$ and $\diamond \in \{+,-\}$, $|\bar{A}_{v,\diamond} \cap V_0| \geq \alpha^2 p_0 n/4 = \mu n$.

Q2 For each $v \in V(D)$, $\diamond, \circ \in \{+,-\}$, and $w \in \bar{A}_{v,\diamond}$, $|N^\circ_{H^\circ_{v,\diamond}}(w, V_i)| \geq \varepsilon p_i n$.

Q3 For each $v \in V(D)$, $\diamond, \circ \in \{+,-\}$, and $w \in V(D)$, $|N^{\bar{\circ}}_{H^\circ_{v,\diamond}}(w, \bar{A}_{v,\diamond}) \cap V_0)| \leq (1+\eta)\varepsilon\alpha^2 p_0 n/4 = (1+\eta)\varepsilon\mu n$, where $\bar{\circ} \in \{+,-\}$ is such that $\bar{\circ} \neq \circ$.

Indeed, by Lemma 2.5, as $\varepsilon, \eta, \alpha, p_0, p_1, \ldots, p_k \gg 1/n$, for any instance of $v \in V(D)$, $\diamond, \circ \in \{+,-\}$, and $w \in V(D)$, the property **Q1** above holds with probability $1 - \exp(-\Omega(n))$, and the same is true for **Q2** and **Q3**. Therefore, by a union bound, with high probability, the properties **Q1**, **Q2** and **Q3** hold.

Now, for each $v \in V(D)$ and $\diamond \in \{+,-\}$, using **Q1**, choose $A_{v,\diamond} \subset \bar{A}_{v,\diamond} \cap V_0$ with $|A_{v,\diamond}| = \mu n$. By **Q2** and **Q3**, we have, for each $\circ \in \{+,-\}$ and $i \in [k]$, that $H^\circ_{v,\diamond}$ is $(\varepsilon p_i n, (1+\eta)\varepsilon\mu n, \circ)$-skew-bounded on $(A_{v,\diamond}, V_i)$, as required, □

We will now use the guide sets produced by Lemma 3.5 to randomly embed T_0, the small core of the original tree, and then use the guide graphs to find matchings from certain subsets of the image of the embedding to other random sets, as follows.

Lemma 3.6. *Let $1/n \ll c \ll \beta \ll \varepsilon, q, \alpha \leq 1$ and $1/n \ll c \ll p \ll 1/m$. Let D be an n-vertex digraph with $\delta^0(D) \geq (1/2+\alpha)n$.*

Let T be an oriented tree with $\Delta^\pm(T) \leq cn/\log n$ consisting of a subtree $T_0 \subset T$ with $|T_0| \leq \beta n$, such that every vertex in $V(T) \setminus V(T_0)$ is attached as a leaf to T_0. Let $t \in V(T_0)$. Let $U_0 = V(T_0)$ and let $U_1 \cup \ldots \cup U_m$ be a partition of $V(T) \setminus V(T_0)$ such that, for each $i \in [m]$, either $e_T(V(T_0), U_i) = 0$ or $e_T(U_i, V(T_0)) = 0$. Let $V_0, V_1, \ldots, V_m \subset V(D)$ be disjoint random sets chosen uniformly at random subject to $|V_0| = qn$, and, for each $i \in [m]$, $|V_i| = \lfloor (1+\varepsilon)|U_i| \rfloor + pn$.

Then, with high probability, for each $s \in V_0$, there is an embedding of T into D such that t is embedded to s, and, for each $i \in \{0, 1, \ldots, m\}$, U_i is embedded into V_i.

Proof. Choose μ such that $c, \beta \ll \mu \ll \varepsilon, q, \alpha$. For each $j \in [m]$, let $p_j = (\lfloor (1+\varepsilon)|U_i| \rfloor / n) + p$. Choose $V_0, V_1, \ldots, V_m$ according to the distribution in the lemma. By Lemma 3.5 with $p_0 = q$, with high probability, for each $v \in V(D)$ and $\diamond \in \{+,-\}$, there is

R1 a set $A_{v,\diamond} \subset N^{\diamond}_D(v) \cap V_0$ with size $q\alpha^2 n/4$, and

R2 digraphs $H^{\circ}_{v,\diamond} \subset D$, $\circ \in \{+,-\}$, such that, for each $j \in [m]$ and $\circ \in \{+,-\}$, $H^{\circ}_{v,\diamond}$ is $(\mu p_j n, (1+\varepsilon/2)\mu q \alpha^2 n/4, \circ)$-skew-bounded on $(A_{v,\diamond}, V_j)$.

We will now show that, given only **R1** and **R2**, we can embed T as required in the lemma for each $s \in V_0$. Let then $s \in V_0$. We will randomly embed T_0 into $D[V_0]$, as follows, before showing that, with positive probability, it can be extended into the required copy of T. Let $\ell = |T_0|$ and label $V(T_0) = \{t_1, \ldots, t_\ell\}$, so that $t_1 = t$ and $T_0[\{t_1, \ldots, t_i\}]$ is a tree for each $i \in [\ell]$. Let $s_1 = s$ and embed t_1 to s_1. For each $i \in \{2, \ldots, \ell\}$ in turn, let $j_i \in \{1, \ldots, i-1\}$ be such that t_{j_i} is the in- or out-neighbour of t_i in $T_0[\{t_1, \ldots, t_i\}]$ and let $\diamond_i \in \{+,-\}$ be such $t_i \in N^{\diamond_i}_{T_0}(t_{j_i})$, and, uniformly at random, embed t_i to $s_i \in A_{s_{j_i},\diamond_i} \setminus \{s_1, \ldots, s_{i-1}\}$.

Claim 3.7. *For each $j \in [m]$, with high probability, the embedding of T_0 can be extended to an embedding of $T[V(T_0) \cup U_j]$ by embedding U_j into V_j.*

As $p \gg 1/n$, and $m \leq 1/p$, we can take a union bound over all $j \in [m]$, to show that, with positive probability, for each $j \in [m]$, the embedding of T_0 can be extended to $T[V(T_0) \cup U_j]$ by embedding U_j into V_j, and hence T can be embedded as required in the lemma. Therefore, there is some choice of the embedding of T_0 for which this can be done. It is left then to prove Claim 3.7.

Proof of Claim 3.7. Let $j \in [m]$ and let $\circ_j \in \{+,-\}$ be such that all the edges from $V(T_0)$ to U_j in T are $\circ_j$-edges. For each $i \in [\ell]$, let $d_{j,i} = |N^{\circ_j}_T(t_i, U_j)|$. For each $i \in [\ell]$, take $d_{j,i}$ new vertices and call them $w_{j,i,i'}$, $i' \in [d_{j,i}]$. Let $W_j = \{w_{j,i,i'} : i \in [\ell], i' \in [d_{j,i}]\}$. Let K_j be the directed graph with vertex set $W_j \cup V_j$, containing only $\circ_j$-edges from W_j to V_j, and where, for each $i \in [\ell]$, $i' \in [d_{j,i}]$ and $v \in V_j$, there is a $\circ_j$-edge from $w_{j,i,i'}$ to v in K_j if, and only if, $s_i v \in E(H^{\circ_j}_{s_{j_i},\diamond_i})$.

We will show that, with high probability, K_j is $(\mu p_j n, \mu p_j n, \circ_j)$-skew-bounded on (W_j, V_j). This is enough to prove the claim, as, by Proposition 3.2, there is a perfect $\circ_j$-matching from W_j into V_j in K_j. Thus, we can label distinct vertices $v'_{j,i,i'}$, $i \in [\ell]$, $i' \in [d_{j,i}]$ in V_j so that $w_{j,i,i'} v'_{j,i,i'}$, $i \in [\ell]$ and $i' \in [d_{j,i}]$, is a matching in K_j. For each $i \in [\ell]$, use the vertices $v'_{j,i,i'}$, $i' \in [d_{j,i}]$, to embed $d_{j,i}$ $\circ_j$-neighbours of t_i in U_j into V_j. This is possible as, by the definition of K_j and $H^{\circ_j}_{s_{j_i},\diamond_i}$, $s_i v'_{j,i,i'}$ is a $\circ_j$-edge in D. Therefore, this extends the embedding of T_0 to an embedding of $T_0 \cup T[U_j]$ with U_j embedded into V_j, as required.

Thus, it is sufficient to prove that, with high probability, K_j is $(\mu p_j n, \mu p_j n, \circ_j)$-skew-bounded on (W_j, V_j). Now, for each $i \in [\ell]$, $s_i \in A_{j_i,\diamond_i}$, and therefore s_i has at least $\mu p_j n$ $\circ_j$-neighbours in V_j in $H^{\circ_j}_{s_{j_i},\diamond_i}$ by **R2**. Therefore, for each $i \in [\ell]$ and $i' \in [d_{j,i}]$, $w_{j,i,i'}$ has at least $\mu p_i n$ $\circ_j$-neighbours in K_j. That is, each $v \in W_j$ has at least $\mu p_j n$ $\circ_j$-neighbours in K_j. Thus, letting $\bar{\circ}_j \in \{+,-\}$ with $\bar{\circ}_j \neq \circ_j$, it is sufficient to prove that, for each $v \in V_j$, with probability $1 - o(n^{-1})$, $d^{\bar{\circ}_j}_{K_j}(v, W_j) \leq \mu p_j n$.

Let then $v \in V_j$. For each $i \in [\ell]$, let

$$X_i^{j,v} = \begin{cases} d_{j,i} & \text{if } s_i v \in E(H^{\circ_j}_{s_{j_i},\diamond_i}) \\ 0 & \text{otherwise,} \end{cases}$$

so that $d^{\bar{\circ}_j}_{K_j}(v, W_j) = \sum_{i \in [\ell]} X_i^{j,v}$. Note that, when $s_i \in A_{s_{j_i},\diamond_i} \setminus \{s_1, \ldots, s_{i-1}\}$ is chosen uniformly at random, by **R1** and **R2**, and as $\beta \ll \varepsilon, \alpha, q$ and $i \leq \ell \leq \beta n$, if $d_{i,j} > 0$, then $X_i^{j,v} = d_{j,i}$ with probability at most

$$\frac{d_{H^{\circ_j}_{s_{j_i},\diamond_i}}(v)}{|A_{s_{j_i},\diamond_i} \setminus \{s_1,\dots,s_{i-1}\}|} \le \frac{(1+\varepsilon/2)\mu q\alpha^2 n/4}{q\alpha^2 n/4-(i-1)} \le (1+\varepsilon)\mu.$$

Let $\gamma=(1+\varepsilon)\mu$. Then, for each $i\in[\ell]$, $\mathbb{E}(X_i^{j,v}|X_1^{j,v},\dots,X_{i-1}^{j,v}) \le \gamma\cdot d_{j,i}$.

Let $Y_i^{j,v}=0$ and, for each $i\in[\ell]$, let $Y_i^{j,v}=\sum_{1\le i'\le i}(X_{i'}^{j,v}-\gamma d_{j,i})$. Then, $Y_0^{j,v},Y_1^{j,v},\dots,Y_\ell^{j,v}$ is a supermartingale, since $\mathbb{E}[Y_{i+1}^{j,v}\mid Y_1,\dots,Y_i]=\mathbb{E}[X_{i+1}^{j,v}\mid Y_1,\dots,Y_i]-\gamma d_{j,i+1}+Y_i^{j,v}\le \gamma d_{j,i+1}-\gamma d_{j,i+1}+Y_i^{j,v}=Y_i^{j.v}$ for each $i\ge 0$. Note further that $|Y_{i+1}^{j,v}-Y_i^{j,v}|=|X_{i+1}^{j,v}-\gamma d_{j,i+1}|\le d_{j,i+1}$ for each $j\in[\ell-1]$. Furthermore, as $d_{j,i}\le cn/\log n$ for each $i\in[\ell]$, and $\sum_{i\in[\ell]}d_{j,i}\le |U_j|\le n$, we have $\sum_{i\in[\ell]}d_{j,i}^2\le cn^2/\log n$. Therefore, by Azuma's inequality (Lemma 2.6) with $t=pn/3$, and using that $c\ll p$,

$$\mathbb{P}(Y_\ell^{j,v}\ge pn/3)\le 2\exp(-p^2n^2\log n/9cn^2)=o(n^{-1}).$$

Thus, with probability $1-o(n^{-1})$, we have $Y_\ell^{j,v}<pn/3$, so that

$$\begin{aligned} d^{\bar{\circ}_j}_{K_j}(v,W_j) &= \sum_{i\in[\ell]}X_i^{j,v} = \gamma\cdot\left(\sum_{i\in[\ell]}d_{j,i}\right)+Y_\ell^{j,v}=\gamma|U_j|+Y_\ell^{j,v}<\gamma|U_j|+pn/3\\ &\le \gamma p_j n/(1+\varepsilon)=(1+\varepsilon)\mu p_j n/(1+\varepsilon)=\mu p_j n,\end{aligned}$$

completing the proof of the claim, and hence the lemma. □ □

Finally, by combining Lemma 3.6 and Lemma 2.13, we can prove Lemma 2.4.

Proof of Lemma 2.4. Let p satisfy $1/n\ll c\ll p\ll 1/K$. For each $j\in[\ell]$, let s_j be the vertex of S_j with an in- or out-neighbour in $V(T')$ in T. Let $\mathcal{R}$ be a maximal set of pairs (R,r) for which R is a directed tree with at most K edges and $r\in V(R)$, such that the pairs (R,r) are unique up to isomorphism. Let $m=|\mathcal{R}|$ and enumerate $\mathcal{R}$ as $(R_1,r_1),\dots,(R_m,r_m)$. Note that $p\ll 1/m$

Let $T''=T[V(T')\cup N_T^+(V(T'))\cup N_T^-(V(T'))]$. For each $i\in[m]$ and $\diamond\in\{+,-\}$, let $U_{i,\diamond}\subset V(T'')$ be the set of vertices s_j, $j\in[\ell]$, for which (S_j,s_j) is isomorphic to (R_i,r_i) and the edge from $V(T')$ to s_j in T is a $\diamond$-edge.

In $V(D)$, take disjoint random sets V_0 and $V_{i,\diamond,j}$, $i\in[m]$, $\diamond\in\{+,-\}$ and $j\in\{1,2\}$, uniformly at random subject to the following.

- $|V_0|=\varepsilon n/2$.
- For each $i\in[m]$ and $\diamond\in\{+,-\}$, we have that $|V_{i,\diamond,1}|=\lfloor(1+\varepsilon/6)|U_{i,\diamond}|\rfloor+pn$ and $|V_{i,\diamond,2}|=(\lfloor(1+\varepsilon/6)|U_{i,\diamond}|\rfloor+pn)(|R_i|-1)$.

Note that this is possible, as

$$\begin{aligned}|V_0|+\sum_{i\in[m],\diamond\in\{+,-\}}(|V_{i,\diamond,1}|+|V_{i,\diamond,2}|) &= |V_0|+\sum_{i\in[m],\diamond\in\{+,-\}}(\lfloor(1+\varepsilon/6)|U_{i,\diamond}|\rfloor+pn)|R_i|\\ &\le \varepsilon n/2+(1+\varepsilon/6)\sum_{j\in[\ell]}|S_j|+\sum_{i\in[m]}2pn\cdot|R_i|\\ &\le \varepsilon n/2+(1+\varepsilon/6)|T|+(2pn)\cdot m\cdot K\le n.\end{aligned}$$

Now, with probability $\varepsilon/2$, $v \in V_0$. By Lemma 3.6, with high probability, if $v \in V_0$, then there is an embedding of T'' into D such that t is embedded to v, $V(T') \subset V_0$, and, for each $i \in [m]$ and $\diamond \in \{+,-\}$, $U_{i,\diamond}$ is embedded into $V_{i,\diamond,1}$. By Lemma 2.13, for each $i \in [m]$ and $\diamond \in \{+,-\}$, $D[V_{i,\diamond,1} \cup V_{i,\diamond,2}]$ contains $|V_{i,\diamond,1}|$ vertex disjoint copies of R_i, in which r_i is copied into $V_{i,\diamond,1}$. For each $i \in [m]$ and $\diamond \in \{+,-\}$, add each copy of R_i containing an embedded vertex of $U_{i,\diamond}$ to the embedding of T''. Note that this results in a copy of T. □

3.2 Embedding constant-sized trees as paths

Given our decomposition $T_0 \subset T_1 \subset T_2 \subset T_3 = T$, we have now embedded T_1. We now embed the vertices from $V(T_2) \setminus V(T_1)$, recalling that we obtain T_2 from T_1 by adding constant-sized trees, where each tree is attached to T_1 by exactly two bare paths of length 2. In the following lemma, we embed $T_2 \setminus T_1$ so that the vertices in $V(T_2) \cap V(T_1)$ are embedded to preselected vertices (labelled a_i, b_i, $i \in [\ell]$). This allows us to extend our embedding of T_1 to one of T_2.

Lemma 3.8. *Let $1/n \ll 1/K \leq 1/k \ll \alpha, \varepsilon$. Suppose T is a forest formed of vertex-disjoint oriented trees T_i, $i \in [\ell]$, with at most $(1-\varepsilon)n$ vertices in total, and so that $k \leq |T_i| \leq K$, for each $i \in [\ell]$, and each tree T_i contains distinct vertices r_i and s_i which are leaves in T_i whose neighbour has total in- and out-degree 2.*

Suppose D is an n-vertex digraph with $\delta^0(D) \geq (1/2+\alpha)n$, containing the distinct vertices a_i, b_i, $i \in [\ell]$. Then, D contains a copy of T in which, for each $i \in [\ell]$, r_i is embedded to a_i and s_i is embedded to b_i.

Proof. Let β be such that $1/k \ll \beta \ll \alpha, \varepsilon$. For each $i \in [\ell]$, let r_i' and s_i' be the neighbours in T_i of r_i and s_i, respectively, and let $T_i' = T_i - \{r_i, r_i', s_i, s_i'\}$. Let T' be the forest composed of connected components T_i', $i \in [\ell]$, so that $|T_i'| \leq (1-\varepsilon)n$. Let $A = \{a_i, b_i : i \in [\ell]\}$. Then $|A| = 2\ell \leq 2n/k$. Let $B \subset V(D) \setminus A$ be a random subset of vertices with $|B| = \beta n$.

Let $D' = D - A - B$. As $1/k, \beta \ll \alpha, \varepsilon$, we have $|D'| \geq (1-\varepsilon/4)n$ and $\delta^0(D') \geq (1/2+\alpha/2)\,|D'|$. Since

$$\left|T'\right| \leq (1-\varepsilon)n \leq \frac{(1-\varepsilon)}{(1-\varepsilon/4)}|D'| \leq (1-\varepsilon/2)\left|D'\right|,$$

we have, by Lemma 2.14, with high probability we can find a copy, S' say, of T' inside D'.

Let r_i'' and s_i'' be the neighbours in T' of r_i' and s_i', respectively, for each $i \in [\ell]$, and let a_i'' and b_i'' be the copy of r_i'' and s_i'' in S', respectively.

Claim 3.9. *The following holds with high probability. For any pair of vertices $u, v \in V(D)$ and $\diamond, \circ \in \{+,-\}$, we have that $|N^\diamond(u) \cap N^\circ(v) \cap B| \geq \alpha\beta n$.*

Proof of Claim 3.9. Let $u, v \in V(D)$ and $\diamond, \circ \in \{+,-\}$. Note that, by the semi-degree condition on D, $|N^\diamond(u) \cap N^\circ(v)| \geq 2\alpha n$, and hence $|N^\diamond(u) \cap N^\circ(v) \cap B|$ has a hypergeometric distribution with $\mathbb{E}\,|N^\diamond(u) \cap N^\circ(v) \cap B| \geq 2\alpha\beta n$. By Lemma 2.5, and a union bound over all pairs $u, v \in D$ and $\diamond, \circ \in \{+,-\}$, the statement in the claim thus holds with probability $1 - o(1)$. □

Thus, with high probability, we can assume the property in the claim holds. Now, for each $i \in [\ell]$, embed a_i and b_i to r_i and s_i, respectively. Let $\diamond_i, \circ_i, \diamond_i', \circ_i' \in \{+,-\}$ be such that $r_i' \in N^{\diamond_i}(r_i) \cap N^{\circ_i}(r_i'')$, and $s_i' \in N^{\diamond_i'}(s_i) \cap N^{\circ_i'}(s_i'')$. Greedily and disjointly, for each $i \in [r]$, embed r_i' to a vertex in $N^{\diamond_i}(a_i) \cap N^{\circ_i}(a_i'') \cap B$ and embed s_i' to a vertex in $N^{\diamond_i'}(b_i) \cap N^{\circ_i'}(b_i'') \cap B$. Note that this is possible, since, from the property in the claim we have, for each $i \in [r]$

$$\left|N^{\diamond_i}(a_i) \cap N^{\circ_i}(a_i'') \cap B\right|, \left|N^{\diamond_i'}(b_i) \cap N^{\circ_i'}(b_i'') \cap B\right| \geq \alpha\beta n \geq \frac{2n}{k} \geq 2r.$$

This completes the embedding of T with the property required in the lemma. □

3.3 Proof of Theorem 2.2

We now combine Lemma 2.4 and Lemma 3.8 to find a copy of any almost-spanning tree.

Proof of Theorem 2.2. Take K, k and η so that $1/n \ll 1/K \ll 1/k \ll \eta \ll \varepsilon, \alpha$. Let D be an n-vertex graph with $\delta^0(D) \geq (1/2 + \alpha)n$. Let T be an oriented tree on at most $(1-\varepsilon)n$ vertices with $\Delta^{\pm}(T) \leq cn/\log n$. By Lemma 2.8, we can find forests $T_0 \subset T_1 \subset T_2 \subset T_3 =$ satisfying **P1** to **P4**. Randomly partition $V(D)$ into three parts, $V(D) = V_1 \cup V_2 \cup V_3$ so that $|V_1| = |T_1| + \varepsilon n/3$, $|V_2| = |T_2| - |T_1| + \varepsilon n/3$, and $|V_3| = |T| - |T_2| + \varepsilon n/3$. Note that, with probability at least $\varepsilon/3$, we have $v \in V_1$.

By applying Lemma 2.10 with $A = V_1$, with high probability we have $\delta^0(D[V_1]) \geq (1/2 + \alpha/2)\,|V_1|$. Thus, by applying Lemma 2.4 to $D = D[V_1]$ and $T = T_1$, we can find a copy of T_1 in V_1 in which t is copied to v. By **P3**, for some $\ell \in \mathbb{N}$, T_2 is formed from T_1 by the addition of trees F_i, $i \in [\ell]$, where $k \leq |F_i| \leq K$, which are each attached to T_1 by exactly two bare paths of length 2, P_i and Q_i say. For each $i \in [\ell]$, let p_i and q_i be the endpoint of P_i and Q_i, respectively, which belongs to T_1. Let a_i and b_i be the embedding in V_1 of p_i and q_i, respectively, and let $A = \{a_i, b_i : i \in [\ell]\}$.

By Lemma 2.10 again, we have, with high probability, $\delta^0(D[A \cup V_2]) \geq (1/2 + \alpha/2)\,|A \cup V_2|$. Applying Lemma 3.8 to $D[A \cup V_2]$ with $T_i = F_i \cup P_i \cup Q_i$, $r_i = p_i$, and $s_i = q_i$, for each $i \in [\ell]$, we can find a copy of T_2 in $D[V_1 \cup V_2]$. Now since T_2 is a tree, any vertex in $T_3 \setminus T_2$ can have at most one neighbour in T_2. Note that, by Lemma 2.10, we know that with high probability every vertex in D has at least $(1/2 + \alpha/2)\,|V_3| \geq \eta n$ in-neighbours in V_3 and at least $(1/2 + \alpha/2)\,|V_3| \geq \eta n$ out-neighbours in V_3. Let $j = |T_3| - |T_2| \leq \eta n$ and order the vertices of $T_3 \setminus T_2$ by $u_1, \ldots, u_j$, so that $T[V(T_2) \cup \{u_1, \ldots, u_i\}]$ is a tree for each $i \in [j]$. Embed the vertices $u_1, \ldots, u_j$ greedily into V_3, to complete the copy of T in D. Noting that this embedding was successful with probability at least $\varepsilon/3 - o(1) > 0$, there must always be such a copy of T. $\square$

4 Absorption from switching

The aim of this section is to prove Theorem 2.1. The main idea is as follows. Given a small tree T, we split it into two trees T' and T'' and randomly embed T' vertex by vertex. With positive probability, the resulting tree is such that, given the right number of other vertices in the graph, we can embed T'' to extend this into a copy of T while making some small modifications to the copy of T'. Essentially, we show that, for each vertex y, there are many vertices in the embedding of T' which we can switch with y and still get a copy of T. We then embed T'' vertex-by-vertex, at each step switching an unused vertex into the copy of T' in place of a vertex which we can instead use to extend the (partial) embedding of T''.

Proof of Theorem 2.1. Take λ such that $\varepsilon \ll \lambda \ll \mu$. Using Proposition 2.3, let $T = T' \cup T''$, where $t \in V(T')$ and $\varepsilon n < |T''| \leq 3\varepsilon n$. Let $\ell = |T'|$, and label $V(T')$ as $t_1, \ldots, t_\ell$ so that $t_1 = t$, $T'[t_1, \ldots, t_i]$ is a tree for each $i \in [\ell]$, and the leaves of T' appear last in this order (except for t) and in any bare path of length 6 the middle 3 vertices appear consecutively. For each $i \in [\ell]$, let $T_i = T'[\{t_1, \ldots, t_i\}]$.

Pick an arbitrary vertex $v \in V(D)$, and let R_1 be the graph with only the vertex v. For each $i = 2, \ldots, \ell$, do the following. Let $\diamond_i \in \{+, -\}$ be such that $N^{\diamond_i}_{T_i}(t_i)$ is non-empty (and thus contains exactly one vertex. Let $\circ_i \in \{+, -\}$ with $\circ_i \neq \diamond_i$. Take R_{i-1}, which is a copy of T_{i-1}, and let w_i be the copy of the sole vertex in $N^{\circ_i}_{T_i}(t_i)$ in R_{i-1}. Pick a vertex v_i independently at random from $N^{\circ_i}_D(w_i) \setminus V(R_{i-1})$. Embed t_i to v_i to get R_i, a copy of T_i.

Note that this process always ends with a copy of T', as $N_D^{\circ_i}(w_i) \setminus V(R_{i-1})$ always has size at least $d_D^{\circ_i}(w_i) - |T| \geq (1/2+\alpha)n - \mu n$ and $\mu \ll \alpha$. Let $R = R_\ell$, so that R is a copy of T'. We will show that, with positive probability the following property holds.

S For each distinct $x, y \in V(D)$ and $\diamond \in \{+,-\}$,

$$|\{i \in [\ell] : v_i \in N_D^{\diamond}(x) \text{ and } N_R^{\pm}(v_i) \subset N_D^{\pm}(y)\}| \geq \lambda n.$$

Noting $|R| = |T'| \leq |T| - |T''| + 1 \leq (\mu - \varepsilon)n$, let $A \subset V(D)$ contain $V(R)$ so that $|A| = (\mu - \varepsilon)n$, and let v be the copy of t. We will show in two claims that, with positive probability **S** holds, and that, if **S** holds, then A and v satisfy the property in the theorem. Thus, the theorem follows from these two claims.

Claim 4.1. *With positive probability,* ***S*** *holds.*

Proof of Claim 4.1. Fix $x, y \in V(D)$ and $\diamond \in \{+,-\}$ with $x \neq y$. We will show that **S** holds for x, y and $\diamond$ with probability at least $1 - 1/4n^2$, so that the result follows by a union bound.

For convenience, let us take two cases. Either T' has $2\mu^2 n$ leaves (Case I) or $\mu^2 n$ vertex-disjoint bare paths with length 6 (Case II). One of these cases must hold, as, suppose that Case I does not hold and thus T' has fewer than $2\mu^2 n$ leaves. Then, by Lemma 2.7, we know that there is some s and some vertex-disjoint bare paths P_i, $i \in [s]$, in T' of length 6 so that $|T' - P_1 - \cdots - P_s| \leq 72\mu^2 n + 2\ell/7$. Removing the internal vertices of each path P_i, $i \in [s]$, from T' removes 5 vertices, and $|T'| = \ell$, so that $\ell - 5s \leq 72\mu^2 n + 2\ell/7$, and therefore

$$s \geq (\ell - 2\ell/7)/5 - 72\mu^2 n/5 \geq \ell/7 - 15\mu^2 n \geq (\mu - 3\varepsilon)n/7 - 15\mu^2 n \geq \mu^2 n,$$

where the final inequality holds since $\varepsilon \ll \mu$.

Case I. Assume that at least $\mu^2 n$ leaves of T' are out-leaves, where the proof whenever T' has at least $\mu^2 n$ in-leaves follows similarly. Let ℓ' be the smallest integer such that, for each $i > \ell'$, t_i is a leaf of T'. We will analyse the embedding of T' in two stages. First, for the embedding of $t_1, \ldots, t_{\ell'}$, we show that with high probability there will be plenty of these vertices which are adjacent to out-leaves in $t_{\ell'+1}, \ldots, t_\ell$ that are embedded to in-neighbours of y. Then, we will analyse the embedding of $t_{\ell'+1}, \ldots, t_\ell$, and show that plenty of these vertices whose in-neighbour in $t_1, \ldots, t_{\ell'}$ was embedded to an in-neighbour of y are themselves embedded to a $\diamond$-neighbour of x.

For each $i \in [\ell']$, let c_i be the number of out-leaves of t_i in T'. For each $i \in [\ell']$, let X_i be the random variable which takes value c_i if $v_i \in N_D^-(y)$, and 0 otherwise. Note that, for each $i \in [\ell]$, if $c_i > 0$, then, when the process selects v_i, having chosen $v_1, \ldots, v_{i-1}$, $X_i = c_i$ with probability at least

$$\frac{|(N_D^{\circ_i}(w_i) \setminus V(R_i)) \cap N_D^-(y)|}{n} \geq \frac{|(N_D^{\circ_i}(w_i)) \cap N_D^-(y)| - |R_i|}{n} \geq \frac{2\alpha n - \mu n}{n} \geq \alpha, \tag{4}$$

as $\alpha \gg \mu$. Thus, for each $i \in [\ell]$, $\mathbb{E}[X_i \mid X_1, \ldots\; X_{i-1}] \geq \alpha c_i$.

Note that $\sum_{i \in [\ell']} c_i$ is the number of out-leaves of T', so that $\sum_{i \in [\ell']} c_i \geq \mu^2 n$ and, as $\Delta(T) \leq cn/\log n$, $\sum_{i \in [\ell']} c_i^2 \leq cn^2/\log n$. Let $Z_0 = 0$ and, for each $i \in [\ell']$, let $Z_i = \sum_{j \leq i}(X_j - \alpha c_j)$. Then, $(Z_i)_{i \geq 0}$ is a submartingale, since $\mathbb{E}[Z_{i+1} \mid Z_1, \ldots, Z_i] = Z_i + \mathbb{E}[X_{i+1} - \alpha c_{i+1} \mid X_1, \ldots\; X_i] \geq Z_i$ for each $i \in [\ell']$. Furthermore, for each $i \in [\ell']$, we have $|Z_i - Z_{i-1}| = |X_i - \alpha c_i| \leq c_i$. Therefore, by Azuma's inequality (Lemma 2.6) with $t = \alpha\mu^2 n/2$, we have

$$\mathbb{P}\left[\sum_{i \in [\ell']}(X_i - \alpha c_i) \leq -t\right] \leq 2\exp\left(\frac{-t^2}{\sum_{i \in [\ell']} c_i^2}\right) \leq 2\exp\left(\frac{-t^2 \log n}{cn^2}\right) \leq \frac{1}{8n^2}. \tag{5}$$

Here, the final inequality holds because $c \ll \mu, \alpha$. Therefore, with probability at least $1 - 1/8n^2$, we have $\sum_{i\in[\ell']} X_i \geq \sum_{i\in[\ell']} \alpha c_i - \alpha\mu^2 n/2 \geq \alpha\mu^2 n/2$.

Let $m = \sum_{i\in[\ell']} X_i \geq \alpha\mu^2 n/2$. Consider now the embedding of $t_{\ell'+1}, \dots, t_\ell$. Let $j_1, \dots, j_m \in \{\ell'+1, \dots, \ell\}$ be such that t_{j_i} is an out-leaf of T' and the image of $N^-_{T'}(t_{j_i})$ is an in-neighbour of y for each $i \in [m]$. For each $i \in [m]$, let Y_i be the random variable which takes value 1 if v_{j_i} is in $N^\diamond_D(x)$, and 0 otherwise. Note that, similarly to the calculation in (4), $\mathbb{E}[Y_i \mid Y_1, \dots Y_{i-1}] \geq \alpha$ for each $i \in [m]$. Let $Z_0 = 0$ and, for each $i \in [m]$, let $Z_i = \sum_{j\leq i}(Y_j - \alpha)$. Then, $(Z_i)_{i\geq 0}$ is a submartingale, since $\mathbb{E}[Z_{i+1} \mid Z_1, \dots, Z_i] = Z_i + \mathbb{E}[Y_i - \alpha \mid Y_1, \dots, Y_{i-1}] \geq Z_i$ for each $i \in [m]$. Furthermore, $|Z_i - Z_{i-1}| = |Y_i - \alpha| \leq 1$ for each $i \in [m]$. Therefore, by Azuma's inequality (Lemma 2.6) with $t = \alpha m/2$, we see that

$$\mathbb{P}\left[\sum_{i\in[m]} Y_i - \alpha < -t\right] \leq 2\exp\left(\frac{-t^2}{(1-\alpha)^2 m}\right) \leq \frac{1}{8n^2}, \tag{6}$$

where the final inequality holds because $1/n \ll \mu, \alpha$. Hence, with probability at least $1 - 1/8n^2$, we have $\sum_{i\in[m]} Y_i \geq \alpha m/2$. Note that $\left|\{i \in [\ell] : v_i \in N^\diamond_D(x) \text{ and } N^\pm_R(v_i) \subset N^\pm_D(y)\}\right| \geq \sum_i Y_i$.

Thus, by taking a simple union bound over the events in (5) and (6) and using $\lambda \ll \alpha, \mu$, we see that in total, with probability at least $1 - 1/4n^2$,

$$\left|\{i \in [\ell]\colon v_i \in N^\diamond_D(x) \text{ and } N^\pm_R(v_i) \subset N^\pm_D(y)\}\right| \geq \alpha m/2 \geq \lambda n.$$

Taking a union bound over all possible $x, y \in V(D)$ and $\diamond \in \{+, -\}$, we see that in this case $\mathbf{S}$ holds with probability at least $1/2$.

Case II. Let $m = \mu^2 n$. Let $P_1, \dots, P_m$ be vertex disjoint paths of length 6 in T, so that, if, for each $i \in [m]$, j_i is such that t_{j_i} is the middle vertex of P_i, then the vertices t_{j_i} appear in order in $t_1, \dots, t_\ell$.

For each $i \in [m]$, let X_i be the random variable taking value 1 if

$$v_{j_i} \in N^\diamond_D(x) \text{ and } N^\pm_R(v_{j_i}) \subset N^\pm_D(y) \tag{7}$$

and 0 otherwise. Note that, by virtue of the labelling of the $t_1, \dots, t_\ell$, the vertices that appear in $N^\pm_R(v_{j_i})$ are exactly the vertices v_{j_i-1} and v_{j_i+1}. When we choose each of $v_{j_i-1}, v_{j_i}, v_{j_i+1}$, the probability that it satisfies its condition in (7) (however the previous vertices $v_{i'}$ are chosen) is at least α, in a calculation similar to (4). Therefore, we have, for each $i \in [m]$, that $\mathbb{E}[X_i \mid X_1, \dots\ X_{i-1}] \geq \alpha^3$.

Now, let $Z_0 = 0$ and, for each $i \in [m]$, let $Z_i = \sum_{j\leq i}(X_j - \alpha^3)$. Then, $\mathbb{E}\left[Z_{i+1} \mid Z_1, \dots, Z_i\right] = Z_i + \mathbb{E}[X_{i+1} \mid X_1, \dots, X_i] - \alpha^3 \geq Z_i$ for each $i \in [m]$, and thus $(Z_i)_{i\geq 0}$ is a submartingale. Furthermore, $|Z_i - Z_{i-1}| = \left|X_i - \alpha^3\right| \leq 1$ for each $i \in [m]$. Thus, by Azuma's inequality (Lemma 2.6), letting $t = \alpha^3 m/2$, we have

$$\mathbb{P}[Z_m \leq -t] \leq 2\exp\left(\frac{-t^2}{m}\right) = 2\exp\left(\frac{-\alpha^6 m}{4}\right) \leq \frac{1}{4n^2},$$

as $1/n \ll \alpha, \mu$. Therefore, with probability at least $1 - 1/4n^2$, we have $Z_m > -t$, so that, as $\lambda \ll \mu, \alpha$,

$$\begin{aligned}\left|\{i \in [\ell]\colon v_i \in N^\diamond_D(x) \text{ and } N^\pm_R(v_i) \subset N^\pm_D(y)\}\right| &\geq \left|\{i \in [m] : v_{j_i} \in N^\diamond_D(x) \text{ and } N^\pm_R(v_{j_i}) \subset N^\pm_D(y)\}\right| \\ &= \sum_{i\in[m]} X_i = Z_m + \alpha^3 m \geq \alpha^3 m - t \geq \lambda n.\end{aligned}$$

Taking a union bound over all possible $x, y \in V(D)$ and $\diamond \in \{+, -\}$, we see that in this case **S** holds with probability at least 1/2. □

Claim 4.2. *If **S** holds then A and v satisfy the property in the theorem.*

Proof of Claim 4.2. Let $B \subset V(D)$ with $A \subset B$ and $|B| = \mu n$. Let $k = |T''| - 1 \leq 3\varepsilon n$ and label the vertices of $V(T'') \setminus V(T')$ as $s_1, \ldots, s_k$, so that, for each $i \in [k]$, $T'_i := T' \cup T''[\{s_1, \ldots, s_i\}]$ is a tree. Note that $|B \setminus V(R)| = k$ and label the vertices of $B \setminus V(R)$ as $y_1, \ldots, y_k$.

Let $S_0 = R$. Now, for each $i = 1, \ldots, k$ in turn, do the following. Let $x_i \in V(S_{i-1})$ and $\diamond_i \in \{+, -\}$ be such that we need to add a $\diamond_i$-neighbour to x_i as a leaf to get a copy of T'_i. Choose some $j'_i \in [\ell] \setminus \{1, j'_1, \ldots, j'_{i-1}\}$ such that

$$v_{j'_i} \in N^{\diamond_i}_D(x_i) \text{ and } N^{\pm}_{S_{i-1}}(v_{j'_{i-1}}) \subset N^{\pm}_D(y_i) \text{ and } d^+_{S_{i-1}}(v_{j'_i}) + d^-_{S_{i-1}}(v_{j'_i}) \leq 4/\lambda.$$

Replace $v_{j'_i}$ with y_i in S_{i-1} and add $v_{j'_i}$ as a $\diamond_i$-neighbour of x_i to get S_i, a copy of T'_i with vertex sets $V(S_{i-1}) \cup \{y_i\}$.

We need only show that there is such a vertex $v_{j'_i}$ in each case, as if this process finds S_k, then we have a copy of $T'_k = T$. Fix then $i \in [k]$. Note that there are at most $(4/\lambda) \cdot 3\varepsilon n \leq \lambda n/4$ vertices next to the vertices $v_{j'_1}, \ldots, v_{j'_i}$ in R_{i-1}, and therefore $N^+_R(v_{i'}) = N^+_{R_{i-1}}(v_{i'})$ and $N^-_R(v_{i'}) = N^-_{R_{i-1}}(v_{i'})$ for all but at most $\lambda n/4$ values of $i' \in [\ell]$. Furthermore, as $\sum_{i' \in [\ell]}(d^+_T(t_{i'}) + d^-_T(t_{i'})) \leq 2n$, at most $\lambda n/2$ values of $i \in [k]$ can have $d^+_{S_{i-1}}(v_{j'_i}) + d^-_{S_{i-1}}(v_{j'_i}) > 4/\lambda$. Thus, such an j'_i will always exist by **S**. □ □

References

[1] N. Alon and J. H. Spencer. **The Probabilistic Method**. John Wiley & Sons, 2004.

[2] B. Bollobás. **Extremal Graph Theory**. Courier Corporation, 2004.

[3] J. Böttcher. Large-scale structures in random graphs. *Surveys in Combinatorics*, 440(2017):87, 2017.

[4] J. Böttcher, M. Schacht, and A. Taraz. Proof of the bandwidth conjecture of Bollobás and Komlós. *Mathematische Annalen*, 343(1):175–205, 2009.

[5] L. DeBiasio, D. Kühn, T. Molla, D. Osthus, and A. Taylor. Arbitrary orientations of Hamilton cycles in digraphs. *SIAM Journal on Discrete Mathematics*, 29(3):1553–1584, 2015.

[6] L. DeBiasio and T. Molla. Semi-degree threshold for anti-directed Hamiltonian cycles. *The Electronic Journal of Combinatorics*, P4.34, 2015.

[7] A. Ghouila-Houri. Une condition suffisante d'existence d'un circuit Hamiltonien. *C.R. Acad. Sci. Paris*, 251(4):495–497, 1960.

[8] A. Hajnal and E. Szemerédi. Proof of a conjecture of P. Erdős. *Combinatorial theory and its applications*, 2:601–623, 1970.

[9] S. Janson, T. Łuczak, and A. Ruciński. **Random Graphs**. John Wiley & Sons, 2011.

[10] J. Komlós, G. N. Sárközy, and E. Szemerédi. Proof of a packing conjecture of Bollobás. *Combinatorics, Probability and Computing*, 4(3):241–255, 1995.

[11] J. Komlós, G. N. Sárközy, and E. Szemerédi. Proof of the Seymour conjecture for large graphs. *Annals of Combinatorics*, 2(1):43–60, 1998.

[12] J. Komlós, G. N. Sárközy, and E. Szemerédi. Spanning trees in dense graphs. *Combinatorics, Probability & Computing*, 10(5):397, 2001.

[13] D. Kühn and D. Osthus. Embedding large subgraphs into dense graphs. *Surveys in Combinatorics*, pages 137–167, 2009.

[14] D. Kühn and D. Osthus. The minimum degree threshold for perfect graph packings. *Combinatorica*, 29(1):65–107, 2009.

[15] R. Montgomery. Spanning trees in random graphs. *Advances in Mathematics*, 356:106793, 2019.

[16] R. Mycroft and T. Naia. Spanning trees of dense directed graphs. *Electronic Notes in Theoretical Computer Science*, 346:645–654, 2019.

[17] R. Mycroft and T. Naia. Trees and tree-like structures in dense digraphs, 2020. *arXiv preprint:* 2012.09201.

[18] V. Rödl, E. Szemerédi, and A. Ruciński. An approximate Dirac-type theorem for k-uniform hypergraphs. *Combinatorica*, 28(2):229–260, 2008.