

# Spanning Trees of Dense Directed Graphs

Richard Mycroft<sup>1</sup> Tássio Naia<sup>2,3</sup>

*School of Mathematics  
University of Birmingham  
Birmingham, United Kingdom*

---

## Abstract

In the nineties, Komlós, Sárközy and Szemerédi confirmed a conjecture of Bollobás, showing that for every positive  $\alpha, \Delta$  and sufficiently large  $n$ , every graph with minimum degree  $(1/2 + \alpha)n$  contains every tree of order  $n$  and maximum degree at most  $\Delta$ . We obtain a directed graph analogue of their result, where the minimum degree is replaced by minimum semidegree (which is the minimum of all in- and out-degrees over all vertices) and the maximum degree is replaced by the maximum degree of the underlying graph (i.e., the maximum degree of the graph we obtain by ignoring the orientation of edges).

In fact, we prove a stronger result, which states a sufficient condition for a tree of order  $n$  to be contained in every directed graph of order  $n$  and minimum semidegree  $(1/2 + \alpha)n$ . This result implies that for all positive real  $\alpha$  and sufficiently large  $n$  every directed graph of order  $n$  with minimum semidegree  $(1/2 + \alpha)n$  contains almost every spanning oriented tree  $T$  of order  $n$ .

**Keywords:** graph theory, spanning trees, directed graphs.

---

## 1 Introduction

Let  $F$  and  $G$  be graphs of order  $n$ . How large must  $\delta(G)$  be to ensure that  $G$  contains a copy of  $F$ ? (We write  $\delta(G)$  to denote the minimum degree of  $G$ ). Perhaps one of the most well-known results in graph theory, Dirac's theorem states that if  $n \geq 3$  and  $F$  is a Hamilton cycle, then  $n/2$  is the answer. We are interested in a Dirac-type question for directed graphs (*digraphs*), replacing the notion of minimum degree by that of *minimum semidegree*  $\delta^0(G)$  of a digraph  $G$ , which is the minimum of all in- and outdegrees of vertices of  $G$ , i.e.,

$$\delta^0(G) := \min_{v \in V(G)} \min\{\deg^-(v), \deg^+(v)\},$$

---

<sup>1</sup> Email: [r.mycroft@bham.ac.uk](mailto:r.mycroft@bham.ac.uk)

<sup>2</sup> Email: [tnaia@member.fsf.org](mailto:tnaia@member.fsf.org)

<sup>3</sup> Research supported by CNPq grant 201114/2014-3.

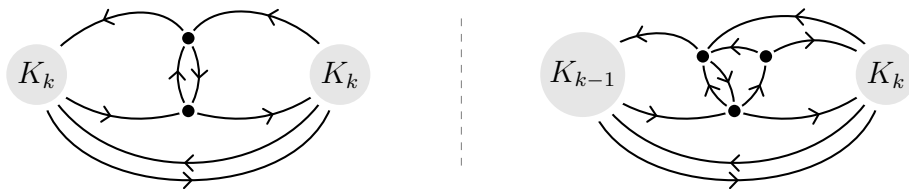


Fig. 1. Two digraphs with order  $n$  and semidegree  $n/2$  which do not contain an antidirected Hamilton cycle, due to Cai [5] (left) and to DeBiasio and Molla [8] (right). In the figure,  $K_p$  denotes a digraph of order  $p$  where each pair of vertices is connected by two edges, one in each direction, and arrowed-lines mean that all possible edges (in the corresponding direction) are present.

where  $\deg_G^-(v)$  is the indegree of  $v$  and  $\deg_G^+(v)$  is the outdegree of  $v$ . An *oriented graph* is a digraph obtained from a non-directed graph by choosing a direction for each of its edges.

**Problem 1.1** Let  $\mathcal{F}$  be a family of oriented trees of order  $n$  (for instance, the set of all oriented paths of order  $n$ ). What is the smallest 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  $\mathcal{F}$ ?

Problem 1.1 has been solved for all sufficiently large values of  $n$  when  $\mathcal{F}$  contains only orientations of paths. Ghouila-Houri [9] proved that if  $\delta^0(G) \geq n/2$ , then  $G$  contains a directed Hamilton cycle (i.e., a spanning cycle  $v_1 \rightarrow \dots \rightarrow v_n \rightarrow v_1$ ), and thus contains a spanning directed path (i.e., a spanning path  $v_1 \rightarrow \dots \rightarrow v_n$ ). This has been extended by DeBiasio, Kühn, Molla, Osthus and Taylor [7], 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 known 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 [8] (see Figure 1).

These results for paths are sharp in the sense that the value of  $\delta_{n,\mathcal{F}}$  is determined precisely. Indeed, if  $\delta^0(G) = \lceil n/2 \rceil - 1$  then  $G$  may not even be connected (see Figure 2), so  $\delta_{n,\mathcal{F}} \geq n/2$  whenever  $\mathcal{F}$  is not empty (see also Figure 1). Hence if  $G$  has semidegree at least  $n/2$  then  $G$  contains every orientation of a Hamilton path (consider a cycle of order  $n$  which is not anti-directed but contains a given orientation of an  $n$ -vertex 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  $G$  (although, for instance, more is known about the presence of Hamilton cycles in  $G$  if  $G$  is an oriented graph [10] or if  $G$  is strongly connected [17]).

The situation in the graph setting is quite different (see, e.g.: [15,16]). Of partic-



Fig. 2. 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  $n = 2k + 1$ .

ular interest to us, a classical theorem by Komlós, Sárközy and Szemerédi [11] states that if  $\delta(G) \geq (1/2 + o(1))n$  then  $G$  contains every tree with bounded degree.

**Theorem 1.2** [11] *For every positive integer  $\Delta$ , every real  $0 < \alpha < 1/2$  and sufficiently large integer  $n$ , every graph of order  $n$  and minimum degree at least  $(\frac{1}{2} + \alpha)n$  contains every spanning tree  $T$  with maximum degree at most  $\Delta$ .*

Komlós, Sárközy and Szemerédi have later strengthened Theorem 1.2, replacing the constant bound  $\Delta$  by  $cn/\log n$ , where  $c$  is some constant depending on  $\alpha$  [12]. Several other variations and extensions of Theorem 1.2 exist (see, e.g., [2,3,4,6,18]).

## 2 Our contribution

We investigate directed graph analogues of Theorem 1.2 where the host digraph has high semidegree.

**Theorem 2.1** *For all positive real  $\alpha, \Delta$  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 2.1 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  $C > 0$ , let  $T$  be an oriented tree of order  $n$ , where  $n$  is sufficiently large, and suppose that  $T$  has maximum degree  $\Delta(T) \leq (\log n)^C$  (where the maximum degree of  $T$  is the maximum degree of the tree we obtain from  $T$  by ignoring the orientation of edges). 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 2.1 follows because every (sufficiently large) oriented tree of bounded degree satisfies this condition (this is proved at the end of this section).

Before we state our main result, we introduce some definitions. A path  $P$  in a digraph  $T$  is *bare* if every edge  $e \in E(T) \setminus E(P)$  either does not intersect  $P$  or intersects  $P$  at one of its endvertices (which are the vertices in  $P$  with a single neighbour in  $P$ ). 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  $x$ , and omit floors and ceilings when this does not affect the arguments.

**Theorem 2.2** *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  $\lambda n$  vertex-disjoint bare paths of order 7; or*
- (ii) *at least  $\lambda n$  vertex-disjoint edges incident to leaves,*

*then  $G$  contains a copy of  $T$ .*

(We sketch the proof of Theorem 2.2 in the next section.)

We remark that not every tree satisfies one of the conditions above. Indeed, consider the following construction. 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 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 (i) or (ii) of Theorem 2.2.

On the other hand, almost all labelled trees have sub-logarithmic maximum degree, as proved by Moon.

**Theorem 2.3** [19, 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 2.3 remains valid if we replace ‘labelled tree’ by ‘labelled oriented tree’.

Moreover, we can prove (by a simple application of the second moment method) that almost every tree satisfies case (ii) of Theorem 2.2. This fact, combined with Theorem 2.3, immediately implies that almost every tree is contained in every digraph of high semidegree.

**Theorem 2.4** *Let  $\mathcal{T}_n$  be the set of all labelled oriented trees of order  $n$ . For all positive  $\alpha, \varepsilon$ , 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 this section describing how we can deduce Theorem 2.1 from Theorem 2.2. We first introduce a few definitions and a lemma. 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 of a tree 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  $p(T) = \ell(T)$ . It is not difficult to prove the following lemma.

**Lemma 2.5** *If  $T$  is a tree which is not a path, then  $\ell(T) \leq p(T) \leq 2\ell(T) - 3$ .*

We are now in position to derive Theorem 2.1 using Theorem 2.2 and Lemma 2.5.

**Proof.** [Proof of Theorem 2.1] 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  $\lambda$

and  $\lambda'$  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  $G$  contains a copy of  $T$  by Theorem 2.2 (i). Otherwise, by Lemma 2.5,  $T$  contains a bare path decomposition  $\mathcal{P}$  into at most  $2\lambda'n$  paths. Note that any bare path in  $T$  is a subpath of some  $P \in \mathcal{P}$ , and that if  $P$  has length  $x$  (i.e.,  $x$  edges) then  $P$  contains at least  $\lfloor x/t \rfloor$  edge-disjoint bare paths (of  $T$ ) of length  $t$  (and order  $t+1$ ). Let  $x_1, \dots, x_s$  be the lengths of the paths in  $\mathcal{P}$ , so  $x_1 + \dots + x_s = n-1$ . Then, for every positive integer  $t$ , we have

The number of edge-disjoint bare-paths of length  $t$  in

$$T \geq \sum_{i=1}^s \left\lfloor \frac{x_i}{t} \right\rfloor \geq \sum_{i=1}^s \left( \frac{x_i}{t} - 1 \right) = \frac{n-1}{t} - s.$$

Choosing  $t = 8$ , it follows that  $T$  contains at least  $(n-1)/8 - 2\lambda'n \geq n/10$  edge-disjoint bare paths of length 8 (and order 9). Therefore,  $T$  contains at least  $n/10$  vertex-disjoint bare paths of order 7, so  $G$  contains a copy of  $T$  by Theorem 2.2 (ii).  $\square$

### 3 Outline of proof of Theorem 2.2

In this section we briefly sketch the proof of Theorem 2.2. A detailed argument is contained in the second author's PhD thesis [21]. Our proof of Theorem 2.2 builds on results of Kühn, Mycroft and Osthus [13,14] as well as on a previous paper by the authors [20]. An important ingredient shared by these results is the regularity method of Szemerédi [22,23], which we briefly introduce in the next section. (A version of Szemerédi's regularity lemma for digraphs was developed by Alon and Shapira [1].)

#### 3.1 A word on regularity

Roughly speaking, the regularity lemma states that if  $G$  is a (di)graph with many vertices, then there exists an equipartition  $V_1, \dots, V_k$  of  $V(G)$  into a bounded number of sets—often called *clusters*—such that the distribution of edges between almost all pairs  $(V_i, V_j)$  is random-like. More precisely, for all  $X \subseteq V_i$  and all  $Y \subseteq V_j$ , we write  $e_G(X, Y)$  to denote the number of edges in  $G$  which are directed from  $X$  to  $Y$ ; we say that  $(V_i, V_j)$  is an  $\varepsilon$ -regular pair (informally, a random-like pair) if

$$\left| \frac{e_G(X, Y)}{|X||Y|} - \frac{e_G(V_i, V_j)}{|V_i||V_j|} \right| < \varepsilon$$

for all  $X \subseteq V_i$  and all  $Y \subseteq V_j$  with  $|X| \geq \varepsilon|V_i|$  and  $|Y| \geq \varepsilon|V_j|$ .

The 'reduced graph' is a graph whose vertices correspond to clusters of  $G$  and whose edges correspond to the pairs of clusters which are connected in this random-like fashion and which are sufficiently 'dense'. Formally, the *reduced graph*  $R = R(d)$  of  $G$  (with parameter  $d$  and with respect to a given vertex-partition  $V_1, \dots, V_k$ ) is defined as follows. Fix  $d$  with  $0 < d < 1$ , and let  $R$  be the graph with vertex set  $\{1, \dots, k\}$  such that for all  $i, j \in R$  we have  $i \rightarrow j$  if and only if  $e_G(V_i, V_j)/(|V_i||V_j|) > d$ .

### 3.2 Proof outline

Recall that the statement of Theorem 2.2 has two cases for the structure of  $T$ . The proof is different for each case, although both follow a similar outline. For the sake of simplicity, below we assume  $T$  falls into case (i) of Theorem 2.2.

Our strategy for finding a spanning tree  $T$  in a digraph  $G$  consists of three steps: ‘allocation’, ‘embedding’ and ‘extension’. As a preliminary step, we delete edges from  $T$ , obtaining a forest  $T'$ ; the deleted edges are chosen carefully, so that they are far apart and so that one component of  $T'$  is large. The goal of the two first steps (allocation and embedding) is to find a copy of  $T'$  in  $G$ . The final step (extension) adds the missing matching.

The bulk of our proof consists of describing how to embed  $T'$  to  $G$  in such a way that the extension step can be made. At the core of our proofs lie two algorithms: a randomised algorithm for allocation and a greedy algorithm for embedding  $T'$ .

Let  $R$  be the reduced graph of  $G$  (note that  $R$  is a digraph), let  $k := |V(R)|$  and  $n := |V(G)|$ . Then 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  $G$ . We may assume that  $1/n \ll 1/k \ll \alpha$ , and, using the fact that  $G$  has high semidegree, we can prove that  $R$  contains a spanning regular subdigraph  $R'$  which is an expander, and (relabelling the clusters if necessary) we may assume that  $R'$  has a directed Hamilton cycle  $1 \rightarrow 2 \rightarrow \dots \rightarrow k \rightarrow 1$ . (A digraph  $D$  is an expander if, for each nonempty proper subset  $S$  of  $V(D)$ , both the in- and outneighbourhoods of  $S$  in  $D$  are strictly larger than  $S$ .)

Informally, an *allocation* of  $T'$  to  $G$  is a homomorphism from  $T'$  to  $R'$ . This homomorphism is constructed by the allocation algorithm (in fact, we alter the allocation of a few vertices at some stage, so that the number of vertices mapped to each cluster  $V_i$  equals  $|V_i|$ ). We note that every vertex of  $G$  is allocated to some cluster in the allocation phase.

Once  $T'$  has been allocated, we embed  $T'$  to  $G$  using a (deterministic and greedy) embedding algorithm which finds a copy of  $T'$  in  $G$  by 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. We take special care, while embedding, to ensure that vertices incident to edges in the removed matching of  $T$  are mapped to ‘good’ vertices. More precisely, let  $M$  be the edges which were removed from  $T$  at the start (i.e.,  $T' := T - E(M)$ ). The embedding is such that, writing  $M_i$  for the vertices of  $M$  embedded to cluster  $V_i$ , and  $V'_i$  for the image of  $M_i$ , the following hold.

- (i)  $|M_i| = |M|/k$ ;
- (ii) either
  - (a) for each  $x \in M_i$ , the neighbour of  $x$  in  $M_i$  is embedded to  $V'_{i+1}$  for all  $i \in \{1, \dots, k\}$ , where addition is modulo  $k$ ; or
  - (b) for each  $x \in M_i$ , the neighbour of  $x$  in  $M_i$  is embedded to  $V'_{i-1}$  for all  $i \in \{1, \dots, k\}$ , where subtraction is modulo  $k$ ;
- (iii)  $(V'_i, V'_{i+1})$  and  $(V'_i, V'_{i-1})$  are regular pairs such that each vertex in  $V'_i$  has many outneighbours in  $V'_{i+1}$  and many inneighbours in  $V'_{i-1}$ .

The conditions above ensure that we may find a copy of  $M$  which *extends* the embedding of  $T'$  to an embedding of  $T$  in  $G$ .

### 3.3 Allocation

The main challenge in allocation is ensuring that the homomorphism we produce maps the correct number of vertices of  $T$  to each cluster. This is achieved by a careful allocation of bare paths of  $T$  (or vertex-disjoint edges incident to leaves of  $T$  respectively) and by using a randomised allocation algorithm which produces (with high probability) an almost uniform distribution of vertices of a large tree to a small regular expander. (The allocation algorithm is described at the end of this section.)

Here is how we proceed. We partition  $T$  into two large edge-disjoint subtrees  $T_1$  and  $T_2$ ; this can be done so that  $|V(T_1)| \leq |V(T_2)|$  and so that  $T_1$  contains many bare paths. We fix two disjoint families  $\mathcal{P}_1, \mathcal{P}_2$  of pairwise vertex-disjoint bare paths in  $T_1$ , and contract all paths in  $\mathcal{P}_1 \cup \mathcal{P}_2$ , yielding a tree  $T^*$ . Let  $T'_1$  be the tree we obtain from  $T_1$  by this contraction. We allocate  $T'_1$  using the randomised allocation algorithm, and allocate the paths in  $\mathcal{P}_1$  and  $\mathcal{P}_2$  in a different manner (described below). The randomised algorithm produces a homomorphism of a tree  $T$  to a digraph  $R$ ; we can prove that this algorithm yields, with high probability, an almost uniform distribution of vertices of  $T$  to vertices of  $R$  if  $R$  contains a spanning regular sub(di)graph  $R'$  which satisfies some expansion properties and if  $T$  has maximum degree bounded (as in the hypothesis of Theorem 2.2).

The allocation of  $T'_1$  produced by the algorithm, together with the allocation of paths in  $\mathcal{P}_1 \cup \mathcal{P}_2$  may not be very uniform (due to the allocation of the paths in  $\mathcal{P}_1$  and  $\mathcal{P}_2$ ). We recover almost-uniformity by allocating  $T_2$  using a weighted version of the allocation algorithm. The resulting allocation of  $T$  is then almost uniform.

The families of contracted paths serve distinct purposes. Paths in  $\mathcal{P}_1$  are allocated so that they have more than one valid allocation (i.e., so that some of their vertices can be allocated to either one of at least two clusters while still yielding a homomorphism). In other words, we have some flexibility when choosing the allocation of these vertices, which we use to refine the almost-uniform allocation of  $T_1 \cup T_2$  into a precise allocation which matches the sizes of each cluster. Paths in  $\mathcal{P}_2$ , on the other hand, are meant to be used in the last step of the embedding phase: extension. Roughly speaking, the paths in  $\mathcal{P}_2$  are mapped so as to ensure that many edges of certain orientation are allocated along certain edges of  $R$ ; this will be useful when finding the desired matching  $M$ .

**Allocation algorithm.** Let  $r$  be the sole vertex at the intersection of  $T_1$  and  $T_2$ , and consider  $T^*$  as rooted at  $r$ . We build a homomorphism from  $T^*$  to  $R'$  one vertex at a time, starting at  $r$ , so that no vertex is processed before its parent and so that at each step we map a vertex  $v \in V(T^*)$  to a vertex of  $R'$ . (The *parent* of a vertex  $v$  is the neighbour of  $v$  in the path from  $v$  to  $r$ ). When allocating any vertex  $x$  (other than the root), the parent  $p$  of  $x$  in  $T^*$  is the only neighbour of  $x$  for which  $\varphi$  has been defined. Let  $C_p$  be the children of  $p$ . We define  $\varphi$  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  $v^+$ .

We note that the algorithm above cannot ensure a good distribution of vertices for trees with small diameter. For instance, if  $T^*$  is a star, then the centre of  $T^*$  is the only vertex mapped to some vertex of  $R$ . However, by the hypothesis of Theorem 2.2, the maximum degree of  $T^*$  is never too high, and we avoid this problem. Indeed, an important step in the proof is to show that if  $\Delta(T^*) \leq (\log |T|)^C$ , where  $1/n \ll 1/C$ , then most vertices of  $T$  lie (sufficiently) far away from one another.

### 3.4 Embedding

Most of the embedding is determined by the algorithm below. More precisely, we apply the embedding algorithm to a subtree of  $T$ , which has been obtained by contracting some edges belonging to paths in  $\mathcal{P}_2$  (we are careful in doing this so that the restriction of the allocation to the contracted tree is still a homomorphism). Moreover, we do not apply the embedding algorithm to the whole of  $G$ , but to a subdigraph  $G' \subseteq G$  (we reserve some sets of vertices in each cluster for the matchings at the end). We prove that the embedding algorithm succeeds and that the embedding it produces can be extended to a full copy of  $T$  in  $G$  in the extension phase.

We conclude this section with an outline of the embedding algorithm. We note that the description below is very simplified, because a key property driving the algorithm, denoted  $(\beta, \gamma, \varphi, m)$ -goodness in [21], is somewhat technical and hence was omitted from this note.

**Embedding algorithm.** Similarly to the allocation algorithm, we consider the tree  $T^*$  which we wish to embed as being rooted; we embed the vertices of  $T^*$  greedily, so that siblings are embedded at the same step and every vertex is embedded after its parent. (The precise description of the algorithm is rather technical and hence we omit it.) In a nutshell, for each  $x \in T^*$ , we embed  $x$  to the cluster  $\varphi(x)$ , and when we do so we also 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 ensured by choosing carefully the order in which vertices are processed.

## 4 Conclusion

Our Theorems 2.1, 2.2 and 2.4 describe a large family of spanning trees of every digraph of high semidegree; as it turns out this is a property of almost every tree, and of every sufficiently large tree whose maximum degree is bounded by a constant. This answers many instances of Problem 1.1. Let us note two very natural ways in which to generalise this problem. One direction is to consider weaker notions of degree, such as the *total degree*  $\delta_{\text{tot}}(G)$ , defined as

$$\delta_{\text{tot}}(G) := \min_{v \in V(G)} (\deg_G^-(v) + \deg_G^+(v)).$$



**Problem 4.1** Do Theorems 2.1, 2.2 and 2.4 remain true if we replace  $\delta^0(G)$  by  $\delta_{\text{tot}}(G)/2$ ?

It is also natural to consider more general classes of spanning subdigraphs (as has been done for undirected graphs), seeking directed graph analogues of generalisations of Theorem 1.2.

**Problem 4.2** Which spanning subdigraphs must be contained in every graph of high semidegree?

It is often the case that directed versions of (undirected) graph problems require substantially distinct approaches in order to be solved. We believe that some of tools we introduce for Problem 1.1 can be adapted to obtain results towards Problems 4.1 and 4.2.

## References

- [1] Alon, N. and A. Shapira, *Testing subgraphs in directed graphs*, Journal of Computer and System Sciences **69** (2004), pp. 354–382.
- [2] Balogh, J., B. Csaba and W. Samotij, *Local resilience of almost spanning trees in random graphs*, Random Structures & Algorithms **38** (2011), pp. 121–139.
- [3] Böttcher, J., J. Hàn, Y. Kohayakawa, R. Montgomery, O. Parczyk and Y. Person, *Universality for bounded degree spanning trees in randomly perturbed graphs*, ArXiv e-prints (2018).
- [4] Böttcher, J., M. Schacht and A. Taraz, *Proof of the bandwidth conjecture of Bollobás and Komlós*, Mathematische Annalen **343** (2009), pp. 175–205.
- [5] Cai, M. C., *A counterexample to a conjecture of Grant*, Discrete Mathematics **44** (1983), p. 111.
- [6] Clemens, D., A. Ferber, R. Glebov, D. Hefetz and A. Liebenau, *Building spanning trees quickly in maker-breaker games*, SIAM Journal on Discrete Mathematics **29** (2015), pp. 1683–1705.
- [7] DeBiasio, L., D. Kühn, T. Molla, D. Osthus and A. Taylor, *Arbitrary orientations of Hamilton cycles in digraphs*, SIAM Journal on Discrete Mathematics **29** (2015), pp. 1553–1584.
- [8] DeBiasio, L. and T. Molla, *Semi-degree threshold for anti-directed Hamiltonian cycles*, Electronic Journal of Combinatorics **22** (2015).
- [9] Houri, G., *Une condition suffisante d’existence d’un circuit Hamiltonien*, Comptes Rendus de l’Académie des Sciences **25** (1960), pp. 495–497.
- [10] Keevash, P., D. Kühn and D. Osthus, *An exact minimum degree condition for Hamilton cycles in oriented graphs*, Journal of the London Mathematical Society **79** (2009), pp. 144–166.
- [11] Komlós, J., G. N. Sárközy and E. Szemerédi, *Proof of a packing conjecture of Bollobás*, Combinatorics, Probability, and Computing **4** (1995), pp. 241–255.
- [12] Komlós, J., G. N. Sárközy and E. Szemerédi, *Spanning trees in dense graphs*, Combinatorics, Probability, and Computing **10** (2001), pp. 397–416.
- [13] Kühn, D., R. Mycroft and D. Osthus, *A proof of Sumner’s universal tournament conjecture for large tournaments*, Proceedings of the London Mathematical Society **102** (2010), pp. 731–766.
- [14] Kühn, D., R. Mycroft and D. Osthus, *An approximate version of Sumner’s universal tournament conjecture*, Journal of Combinatorial Theory, Series B **101** (2011), pp. 415–447.
- [15] Kühn, D. and D. Osthus, “Embedding large subgraphs into dense graphs,” London Mathematical Society Lecture Notes **365**, Cambridge University Press, 2009 pp. 137–167.

- [16] Kühn, D. and D. Osthus, *Hamilton cycles in graphs and hypergraphs: an extremal perspective*, in: *Proceedings of the International Congress of Mathematicians (Seoul, Korea)*, ICMC 2014 **4**, 2014, pp. 381–406.
- [17] Kühn, D., D. Osthus and T. Townsend, *Proof of a tournament partition conjecture and an application to 1-factors with prescribed cycle lengths*, *Combinatorica* **36** (2016), pp. 451–169.
- [18] Michael Krivelevich, B. S., Matthew Kwan, *Bounded-degree spanning trees in randomly perturbed graphs*, *SIAM J. Discrete Math.* **31** (2017), pp. 155—171.
- [19] Moon, J. W., “Counting labelled trees,” Number 1 in *Canadian Mathematical Monographs*, Canadian Mathematical Congress, 1970.
- [20] Mycroft, R. and T. Naia, *Unavoidable trees in tournaments*, *Random Structures & Algorithms* **53** (2018), pp. 352–385.
- [21] Naia, T., “Large Structures in Dense Directed Graphs,” Ph.D. thesis, University of Birmingham (2018).
- [22] Szemerédi, E., *On sets of integers containing no  $k$  elements in arithmetic progression*, *Acta Arith.* **27** (1975), pp. 199–245, collection of articles in memory of Juriĭ Vladimirovič Linnik.
- [23] Szemerédi, E., *Regular partitions of graphs*, in: *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, CNRS, Paris, 1978 pp. 399–401.