Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 195 (2021) 376–384

Procedia Computer Science

www.elsevier.com/locate/procedia

XI Latin and American Algorithms, Graphs and Optimization Symposium

Hitting times for arc-disjoint arborescences in random digraph processes☆

Maurício Collares[a,1], Yoshiharu Kohayakawa[b,2], Taísa Martins[c,3], Roberto Parente[d,4], Victor Souza[e,5]

[a]Departamento de Matemática, Instituto de Ciências Exatas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[b]Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil
[c]Instituto de Matemática e Estatística, Universidade Federal Fluminense, Niterói, Brazil
[d]Instituto de Matemática e Estatística, Universidade Federal da Bahia, Salvador, Brazil
[e]Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge, United Kingdom

Abstract

In this work, we study hitting times for the appearance of a spanning structure in the Erdős-Rényi random directed graph processes. Namely, we are concerned with the appearance of an arborescence, a spanning digraph in which, for a vertex u called the *root* and any other vertex v, there is exactly one directed path from u to v. Let $\mathcal{D}(n,0), \mathcal{D}(n,1), \dots, \mathcal{D}(n,n(n-1))$ be the random digraph process where for every $m \in \{0,\dots,n(n-1)\}$, $\mathcal{D}(n,m)$ is a digraph with vertex set $\{1,\dots,n\}$; $\mathcal{D}(n,0)$ has no arcs and, for $1 \leq m \leq n(n-1)$, the digraph $\mathcal{D}(n,m)$ is obtained by adding an arc to $\mathcal{D}(n,m-1)$, chosen uniformly at random among the not present arcs. In this paper we determine the hitting time for the existence of k arc-disjoint arborescences when $k = k(n) \ll \sqrt{\log n}$.

Peer-review under responsibility of the scientific committee of the XI Latin and American Algorithms, Graphs and Optimization Symposium

Keywords: Hitting time; Random graph processes; Arborescences; Packing.

☆ Y. Kohayakawa was partially supported by CNPq (311412/2018-1, 423833/2018-9) and FAPESP (2018/04876-1). This study was financed in part by CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil, Finance Code 001. FAPESP is the São Paulo Research Foundation. CNPq is the National Council for Scientific and Technological Development of Brazil.
[1] Email: mauricio@collares.org
[2] Email: yoshi@ime.usp.br
[3] Email: tlmartins@id.uff.br
[4] Email: roberto.parente@ufba.br
[5] Email: vss28@cam.ac.uk

1877-0509
Peer-review under responsibility of the scientific committee of the XI Latin and American Algorithms, Graphs and Optimization Symposium
10.1016/j.procs.2021.11.046

1. Introduction

A *spanning tree* in a graph is a subgraph that includes all vertices of the graph and in which any two vertices are connected by exactly one path. Generalising this notion to the digraph setting, we obtain the notion of *arborescence*. An arborescence is a spanning digraph in which, for a vertex u called the *root* and any other vertex v, there is exactly one directed path from u to v.

Note that in an undirected graph, the existence of a spanning tree is equivalent to the graph being connected. The study of spanning trees in random graphs goes back about 80 years, to the seminal paper of Erdős and Rényi on random graphs [4]. In the paper, they started the study of the phase transition in random graphs. In particular, they determined the "threshold" value of $m = n \log n/2$ above which a random m-edge undirected graph is typically connected.

Given this threshold, a subsequent natural question is to estimate the number of spanning trees in a random graph at or above this threshold. Of particular interest is the number of edge-disjoint spanning trees. The *spanning tree packing number* or *STP number* of a graph G is the maximum number of edge-disjoint spanning trees contained in G. One of the earliest results on the STP number is a min-max relation shown by Tutte [11] and Nash-Williams [9].

Erdős and Rényi's work [4] implies that for $p = (\log n - \omega(1))/n$, the random graph $\mathcal{G}(n, p)$ is disconnected with high probability, and hence the STP number is zero. Palmer and Spencer [10] showed that with high probability the STP number of $\mathcal{G}(n, p)$ equals the minimum degree whenever such value is a constant, which typically happens when p is around $(\log n + O(\log\log n))/n$. In fact, they proved a stronger hitting time result and showed that, with high probability, the precise time when the minimum degree first becomes k (for constant k) coincides with the time when k edge-disjoint spanning trees first appear. Other ranges of p were also considered by Catlin, Chen and Palmer [2] and Chen, Li and Lian [3].

Characterising the entire range of p, Gao, Pérez-Giménez and Sato [6] have shown that the STP number is, with high probability, the minimum between δ and $\lfloor m/(n-1)\rfloor$ during the whole random graph process. They also determine the asymptotic value of p at which the STP number changes from δ to $\lfloor m/(n-1)\rfloor$.

In the directed case, much less is known. Bal, Bennett, Cooper, Frieze and Prałat [1] considered the random digraph process $\mathcal{D}(n, 0), \mathcal{D}(n, 1), \dots, \mathcal{D}(n, n(n-1))$ on common vertex set $[n] = \{1, \dots, n\}$, the stochastic process in which we start with n vertices and no arcs, and at each step, we add one new arc chosen uniformly at random from the set of missing arc. They have shown that the events $\mathcal{D}(n,m)$ *has an arborescence* and *at most one vertex of* $\mathcal{D}(n,m)$ *has in-degree zero* have the same hitting time with high probability. It is worth mentioning that they are actually concerned with the appearance of a rainbow arborescence in a randomly coloured random digraph and obtain the arborescence result as a corollary.

Concerning packing results for a digraph G, let $\lambda(G)$ denote the largest integer $k \geq 0$ such that, for all $0 \leq \ell \leq k$, we have $\sum_{i=0}^{\ell-1}(\ell - i)|\{v : d_G^{in}(v) = i\}| \leq \ell$. One can observe that $\lambda(G)$ is an upper bound on the number of arc-disjoint arborescences by noticing that in order to pack ℓ arborescences, every vertex of G whose in-degree is $\ell - i$ must be the root of at least i arborescences, since its in-degree would be exhausted. Letting $\mathcal{D}(n, p)$ denote the random digraph (defined by including each of the $n(n-1)$ arcs independently with probability p), Hoppen, Parente and Sato [7] have shown that the maximum number of arc disjoint arborescences in $\mathcal{D}(n, p)$ is $\lambda(\mathcal{D}(n, p))$ with high probability for every $0 \leq p \leq 1$. Moreover, they determined $\lambda(\mathcal{D}(n, p))$ asymptotically for values of p such that the minimum in-degree of $\mathcal{D}(n, p)$ is concentrated.

In this paper we determine the hitting time on random digraph processes of k arc-disjoint arborescences when $k = k(n) \ll \sqrt{\log n}$ (that is, if $k/\sqrt{\log n}$ tends to 0 as n tends to infinity). For this, we are concerned with the following two events.

$$\mathcal{A}_{k,m} = \{\mathcal{D}(n,m) \text{ has } k \text{ arc-disjoint arborescences}\}$$
$$\mathcal{Z}_{k,m} = \{\text{at most } k \text{ vertices of } \mathcal{D}(n,m) \text{ have in-degree less than } k\}.$$

Let $\mathcal{E}_{k,m}$ stand for one of the above events. We define a random variable, the *hitting time* of $\mathcal{E}_k$, by

$$m(\mathcal{E}_k) = \min\{m \in \mathbb{N} : \mathcal{E}_{k,m} \text{ occurs}\}.$$

Establishing an analogue of Palmer and Spencer's result in the directed setting and a strengthening of Hoppen, Parente and Sato's result for $k \ll \sqrt{\log n}$, our main result is the following.

Theorem 1.1. *For* $k = k(n) \ll \sqrt{\log n}$, *we have* $m(\mathcal{A}_k) = m(\mathcal{Z}_k)$ *with high probability.*

Structure of the paper. In Section 2 we present some basic tools about random models and a min-max result concerning the existence of k arc-disjoint arborescences due to Frank [5]. In Section 3 we present results about in-degree distributions of sets and vertices. In Section 4 we study the neighbourhood of vertices with in/out-degree close to minimum in/out-degree. In Section 5 we study the in-degree of sets in the subgraph formed by the vertices with degree at least k. Finally, in Section 6 we prove our main theorem.

2. Tools

For a digraph $D = (V, A)$ and $S, T \subset V$, let $A(S, T)$ denote the set of arcs $uv \in A$ with $u \in S$ and $v \in T$. We also write $A(T)$ to denote $A(T, T)$. Our main deterministic tool provides a necessary and sufficient condition for the existence of k disjoint arborescences in a digraph. It is a directed analogue of the classic results of Tutte and Nash–Williams, and one of the main tools in our paper. A *subpartition* of a set V is a family of pairwise disjoint subsets of V.

Theorem 2.1 (Frank [5]). *Let $D = (V, A)$ be a digraph and let $k \geq 0$ be an integer. Then D contains k arc-disjoint arborescences if, and only if, for every subpartition $\mathcal{P}$ of V, we have $\sum_{U\in\mathcal{P}} d^{in}(U) \geq k(|\mathcal{P}| - 1)$, where $d^{in}(U) = |A(V \setminus U, U)|$.*

We now state some tools from the theory of random graphs. Recall that the hypergeometric distribution $\mathrm{Hyp}(a, b, m)$ is the distribution of a random variable $|A_m \cap B|$, where A and B are fixed sets such that $|A| = a$, $B \subset A$ has size $|B| = b$, and A_m denotes a m-random subset of A. If $X \sim \mathrm{Hyp}(a, b, m)$, then $\mathbb{E}[X] = mb/a$.

Theorem 2.2 (Chernoff bounds [8]). *Let X be a binomial or hypergeometric random variable. If $\mu = \mathbb{E}[X]$,*

$$\mathbb{P}(X \geq \mu + t) \leq \exp\big(- \mu\varphi(t/\mu)\big) \leq \exp\left(-\frac{t^2}{2(\mu + t/3)}\right),$$
$$\mathbb{P}(X \leq \mu - t) \leq \exp\big(- \mu\varphi(-t/\mu)\big) \leq \exp(-t^2/2\mu),$$

where $\varphi(x) = (1 + x)\log(1 + x) - x$ for $x \geq -1$, $\varphi(-1) = 1$ and $\varphi(x) = \infty$ for $x < -1$.

Although our main result concerns the random digraph process, it will be convenient to prove some results in $\mathcal{D}(n, p)$. Let $N = n(n - 1)$. The following results allow us to transfer such results to $\mathcal{D}(n, m)$ with $m = pN$.

Definition 2.3. *A property Q of digraphs is (m, r)-*increasing *if*

$$\mathbb{P}(\mathcal{D}(n, m') \text{ has } Q) \leq \mathbb{P}(\mathcal{D}(n, m) \text{ has } Q) + o(1) \qquad \text{if } m - r < m' \leq m,$$
$$\mathbb{P}(\mathcal{D}(n, m') \text{ has } Q) \geq \mathbb{P}(\mathcal{D}(n, m) \text{ has } Q) + o(1) \qquad \text{if } m \leq m' < m + r.$$

We define (m, r)-*decreasing* properties analogously. A property is (m, r)-*monotone* if it is either (m, r)-decreasing or (m, r)-increasing. Observe that a monotone property Q is (m, r)-monotone for any $0 \leq m \leq N$ and any $r > 0$.

Proposition 2.4 (Proposition 1.13 and Remark 1.14 [8]). *Let Q be a property for digraphs and $0 \leq m \leq N$. If Q is $(m, O(\sqrt{m(N - m)/N}))$-monotone and holds with high probability in $\mathcal{D}(n, p)$ for every $p = m/N + O(\sqrt{m(N - m)/N^3})$, then Q holds for $\mathcal{D}(n, m)$ with high probability.*

We will also use some expansion properties of random graphs.

Lemma 2.5 (Lemma 3.17 [7]). *Let ζ be a positive constant. For $p \gg 1/n$, the following holds with high probability. For disjoint sets $S, T \subset [n]$, each of size at least ζn, we have $|A(S, T)| \geq \zeta^2 n^2 p/2$.*

Lemma 2.6 (Lemma 3.18 [7]). *Let ε be a positive constant. For $p \gg 1/n$, the following holds with high probability. For every $S \subset [n]$ of size $|S| \leq \varepsilon n$ we have $|A(S)| \leq 3\varepsilon pn|S|$.*

3. In-degree distribution

In this section, we define probabilities p_k and show that, although $\mathcal{D}(n, p_k)$ does not contain k arc-disjoint arborescences with high probability, the number of vertices with in-degree $k - 1$ is reasonably low. Let $X_k = |\{v\colon d^{in}(v) = k\}|$ and $X_{<k} = |\{v\colon d^{in}(v) < k\}|$.

Lemma 3.1. *Let $p \geq 0.9 \log n/(n-1)$. With high probability in $\mathcal{D}(n,p)$ and in $\mathcal{D}(n,m)$ with $m = pn(n-1)$, every vertex $v \in V$ satisfies $d^{in}(v) + d^{out}(v) > p(n-1)/5$.*

Proof. For every $0 < \alpha < 1$, using Theorem 2.2 with $t = \alpha p(n-1)$, we have

$$\mathbb{P}(\operatorname{Bin}(2(n-1), p) \leq 2\alpha p(n-1)) \leq \exp\left(-2p(n-1)(1-\alpha+\alpha\log\alpha)\right).$$

A simple calculation shows that, for $\alpha = 1/10$, $2p(n-1)\cdot(1-\alpha+\alpha\log\alpha) \geq 1.204\log n$. Therefore, the result follows by applying the union bound over all vertices. The same result for $\mathcal{D}(n,m)$ follows by applying the hypergeometric variant of Theorem 2.2. □

Definition 3.2. *Let $k = k(n) \ll \sqrt{\log n}$. Define*

$$p_k = \frac{\log n + (1 + (k-1)/\log n)((k-1)\log\log n - \log(k-1)!) - (\log\log n)/2}{n-1}.$$

Constant terms in the numerator of the above definition may be omitted if desired, which simplifies the formula slightly when $k \ll \sqrt{\log n/\log\log n}$ and considerably when k is bounded.

Lemma 3.3. *Let $k = k(n) \ll \sqrt{\log n}$. For any p of the form $p = p_k + o(1/n)$, it holds that $X_{k-1} > k$, $X_{k-1} \ll \log n$ and $\delta^{in}(\mathcal{D}(n,p)) = k-1$ with high probability.*

Proof. Observe that the events $\{\{d^{in}(v) = k-1\} : v \in V(\mathcal{D}(n,p))\}$ are independent and identically distributed. Therefore, X_{k-1} is binomially distributed and we can write

$$\mathbb{E}[X_{k-1}] = n\binom{n-1}{k-1}p^{k-1}(1-p)^{n-k}.$$

To bound the above, observe that since $k^2 \ll n$, it holds that $\binom{n-1}{k-1} \sim (n-1)^{k-1}/(k-1)!$. Moreover, since $p^2 n \ll 1$, $pk \ll 1$ and $1 - x = e^{-x+O(x^2)}$ as $x \to 0$, we have $(1-p)^{n-k} \sim e^{-p(n-1)}$. Substituting, we obtain

$$\mathbb{E}[X_{k-1}] \sim \frac{n((n-1)p)^{k-1}e^{-p(n-1)}}{(k-1)!}.$$

We now use the value of p to conclude that

$$\mathbb{E}[X_{k-1}] \sim \sqrt{\log n}\left(\frac{(n-1)p}{\log n}\right)^{k-1} \exp\left(-\frac{(k-1)((k-1)\log\log n - \log(k-1)!)}{\log n}\right). \tag{1}$$

Using that $k^2 \ll \log n$, we can bound

$$\frac{(n-1)p}{\log n} = 1 + \frac{(k-1)\log\log n - \log(k-1)!}{\log n} + O\left(\frac{\log\log n}{\log n}\right).$$

We now use $1 + x = \exp(x + O(x^2))$ (as x tends to 0) and once again $k^2 \ll \log n$ to compute

$$\left(\frac{(n-1)p}{\log n}\right)^{k-1} = \exp\left(\frac{(k-1)((k-1)\log\log n - \log(k-1)!)}{\log n} + o\left(\frac{(k-1)(\log\log n)^2}{\log n}\right)\right). \tag{2}$$

The above error term satisfies $\exp\left(o\left(\frac{(k-1)(\log\log n)^2}{\log n}\right)\right) = 1 + o(1)$, so combining (2) with (1) leads to $\mathbb{E}[X_{k-1}] \sim \sqrt{\log n}$. Therefore, we have $X_{k-1} \ll \log n$ with high probability by Markov's inequality. Moreover, since X_{k-1} is binomially distributed, it holds with high probability (by Chernoff's inequality) that $X_{k-1} \geq \sqrt{\log n}/2 > k$ if n is large. It only remains to show that $\delta^{in}(\mathcal{D}(n,p)) = k-1$ with high probability, which is equivalent to showing that $X_{<k-1} = 0$ with high probability. To do so, observe that for every $i \leq k$,

$$\mathbb{E}[X_{i-1}] = \left(\frac{i}{n-i+1}\right)\left(\frac{1-p}{p}\right)\mathbb{E}[X_i] \leq \left(\frac{2i}{np}\right)\mathbb{E}[X_i] \ll \frac{\mathbb{E}[X_i]}{\sqrt{\log n}},$$

since $np \sim \log n$ and $k \ll \sqrt{\log n}$. Summing over $i < k$, we obtain that $\mathbb{E}[X_{<k-1}] \leq \mathbb{E}[X_{k-1}]/\sqrt{\log n} \ll 1$. Therefore, $X_{<k-1} = 0$ with high probability by Markov's inequality, finishing the proof. □

4. Light vertices

The concept of light vertices will simplify the analysis of subpartitions in the next section.

Definition 4.1. *A vertex $v \in [n]$ is said to be ε-*in-light *if $d^{in}(v) \le \varepsilon \log n$ and, ε-*out-light *if $d^{out}(v) \le \varepsilon \log n$.*

Below are some properties of light vertices. Let $\Gamma^{in}(v) = \{u \in V : uv \in A\}$ and $\Gamma^{out}(v) = \{u \in V : vu \in A\}$.

Lemma 4.2. *Let $\varepsilon \le 0.09$ and $0.9 \log n/(n-1) \le p \le (1+o(1)) \log n/(n-1)$. The following holds with high probability for $\mathcal{D}(n,p)$ and for $\mathcal{D}(n,m)$ with $m = pn(n-1)$:*

(a) *there is no pair (u,v) of ε-in-light vertices such that $uv \in A$ or $\Gamma^{in}(v) \cap \Gamma^{in}(u) \ne \emptyset$;*

(b) *there is no pair (u,v) of ε-out-light vertices such that $uv \in A$ or $\Gamma^{out}(v) \cap \Gamma^{out}(u) \ne \emptyset$;*

(c) *there is no $uv \in A$ such that u is ε-out-light and v is ε-in-light.*

Proof. We will prove the result only for $\mathcal{D}(n,m)$, since the proof for $\mathcal{D}(n,p)$ can be obtained merely by replacing the hypergeometric distribution by the binomial distribution in appropriate places.

We start by proving item (a). Let $S \subseteq V$ denote the set of ε-in-light vertices. For distinct vertices $u,v \in V$, we estimate the probability that they are ε-in-light and that $uv \in A$. There are $2n-3$ arcs (other than uv) that contribute to $d^{in}(u) + d^{in}(v)$, and if the vertices are ε-in-light this sum is at most $2\varepsilon \log n$. Let $N = n(n-1)$ and $\mu = (2n-3)(m-1)/(N-1) \sim 2m/n = 2p(n-1)$. Since $2\varepsilon \log n \le p(n-1)/5$, we have

$$\mathbb{P}(\{u,v\} \subset S, uv \in A) \le p \cdot \mathbb{P}(\operatorname{Hyp}(N-1, 2n-3, m-1) \le 2\varepsilon \log n) \le pn^{-6/5},$$

where the last inequality follows from Chernoff bounds (c.f. the proof of Lemma 3.1). Therefore, the probability that there exists u,v with the above property is at most $n^2 \cdot pn^{-6/5} = o(1)$, proving the first part of (a); an adaptation of this argument proves part (c). To prove the second part of item (a), observe that for distinct vertices $u,v,z \in V$, we have by an analogous hypergeometric estimate that

$$\mathbb{P}(zu, zv \in A, u \in S \text{ and } v \in S) \le p^2 \cdot \mathbb{P}(\operatorname{Hyp}(N-2, 2n-4, m-2) \le 2\varepsilon \log n) \le p^2 n^{-6/5},$$

and the probability that there exists a triple (u,v,z) as above is at most $n^3 \cdot p^2 n^{-6/5} = o(1)$. Hence, with high probability none of the events occur, i.e. there is no pair (u,v) of ε-in-light vertices such that $uv \in A$ or $\Gamma^{in}(v) \cap \Gamma^{in}(u) \ne \emptyset$. By symmetry, the same argument proves case (b). □

The next result follows directly from Lemma 4.2.

Corollary 4.3. *Let $\varepsilon \le 0.09$ and $0.9 \log n/(n-1) \le p \le (1+o(1)) \log n/(n-1)$. The following holds with high probability in $\mathcal{D}(n,p)$:*

(a) *For every set S and every set T of ε-in-light vertices disjoint from S, $|A(S,T)| \le |S|$.*

(b) *For every set S of ε-out-light vertices and every set T of vertices disjoint from S, $|A(S,T)| \le |T|$.*

5. Frank's condition

In this section, we verify the conditions of Theorem 2.1 for the subgraph of $\mathcal{D}(n,p)$ induced by the set $V_k = \{v \in [n] : d^{in}(v) \ge k\}$.

Lemma 5.1. *Let $1 \le k = o(\log n)$ and $V_k = \{v \in [n] : d^{in}(v) \ge k\}$. If $|V_k^c| = o(\log n)$, there exists a constant $\tau > 0$ such that, for all $0.9 \log n/(n-1) \le p \le (1+o(1)) \log n/(n-1)$ the following holds with high probability: for every $S \subset V_k$ of size $2 \le |S| \le \tau n$, we have $d^{in}_{V_k}(S) \ge k$.*

Proof. Let $\varepsilon = 0.09$ and take $\tau = \varepsilon/48$. Let $S_h \subset S$ be the set of vertices of S that are not ε-in-light. By the hypothesis on k, all vertices outside of V_k are ε-in-light, and therefore there are no arcs between V_k^c and $S \setminus S_h$ by Lemma 4.2(c). If $|S_h| \geq |S|/8$, we have

$$|A(V_k \setminus S, S)| \geq |A(V_k \setminus S, S_h)| \geq |S_h|(\varepsilon \log n - |V_k^c|) - |A(S)| \geq \frac{\varepsilon}{3}|S_h| \log n,$$

where in the last inequality we used that $|A(S)| \leq 3\tau|S|\log n$ (by Lemma 2.6), together with the hypothesis on $|V_k^c|$ and the value of τ. Finally, when $|S_h| < |S|/8$, we have

$$|A(V_k \setminus S, S)| \geq |A(V_k \setminus S, S \setminus S_h)| \geq k|S \setminus S_h| - |S_h|,$$

where the last inequality follows from $|A(S, S \setminus S_h)| \leq S_h$ (by Corollary 4.3(a)) and the fact that there are no arcs between V_k^c and $S \setminus S_h$. We conclude that $|A(V_k \setminus S, S)| \geq (k+1)|S \setminus S_h| - |S_h| \geq (|S|/8)(7k-1)$, finishing the proof since $|S| \geq 2$. □

Lemma 5.2. *Let $1 \leq k = o(\log n)$ and $V_k = \{v \in [n] : d^{in}(v) \geq k\}$. If $|V_k^c| = o(\log n)$, there exists $\tau > 0$ such that for all $p \geq 0.9 \log n/(n-1)$ the following holds with high probability: for every partition (S, T, U) of V_k such that $|S| \geq (1-\tau)n$, $|T| \neq 0$ and either $|U| \leq |T|$ or $|U| \leq 6\tau p(n-1)$, we have $d^{in}_{V_k}(S) + d^{in}_{V_k}(T) \geq (|T|+1)k$.*

Proof. Take $\tau = 1/180$. By Lemma 3.1 we have for every $v \in [n]$, $d^{in}(v) + d^{out}(v) \geq p(n-1)/5$ with high probability. Note that $|T| \leq \tau n$, so if $|U| \leq |T|$ then we have $|A(T \cup U)| \leq 6\tau p(n-1)|T \cup U|$ by Lemma 2.6. Otherwise, $|U| \leq 6\tau p(n-1)$, and so $|A(T, U)| \leq \min\{|T| \cdot |U|, |A(T \cup U)|\} \leq 6\tau p(n-1)|T|$.

To bound $d^{in}_{V_k}(S) + d^{in}_{V_k}(T)$, we will ignore arcs from U to S. The remaining relevant arcs all intersect T, and therefore we may write

$$\begin{aligned} d^{in}_{V_k}(S) + d^{in}_{V_k}(T) &\geq \sum_{v \in T} \left(d^{in}(v) + d^{out}(v) - 2\left|V_k^c\right|\right) - 2|A(T)| - |A(T, U)| \\ &\geq |T|p(n-1)/5 - 2|T||V_k^c| - 18\tau|T|(n-1)p = |T|p(n-1)/10 + o(|T|\log n). \end{aligned}$$

Since $k = o(\log n)$ and $p(n-1) = \Omega(\log n)$, the result follows. □

Lemma 5.3. *There exists $\tau > 0$ such that for all $(\log n - \log(2^{-8}\log n))/(n-1) \leq p \leq (1+o(1))\log n/(n-1)$, the following holds with high probability. Let $1 \leq k = o(\log n)$ and $V_k = \{v \in [n] : d^{in}(v) \geq k\}$. For all $S \subset V_k$ such that $p(n-1)/40 \leq |S| \leq \tau n$, we have that $d^{out}_{V_k}(S) \geq k$.*

Proof. Let $\varepsilon = 0.09$ and take $\tau = \varepsilon/48$. Let S_ℓ be the set of ε-out-light vertices in S and define $S_h = S \setminus S_\ell$. We first treat the case $|S_h| \geq |S|/8$. For that, notice that $|A(S_h, V_k^c)| \leq |S_h|$ by Corollary 4.3(a). Therefore

$$d^{out}_{V_k}(S) \geq \sum_{v \in S_h} d^{out}(v) - |S_h| - |A(S)| \geq (\varepsilon \log n - 1)|S_h| - 3\tau n|S|p = \Omega(|S|\log n).$$

For the case when $|S_\ell| \geq 7|S|/8$, let us define X_0 to be the cardinality of the set of vertices of out-degree zero, i.e., $X_0 = \left|\{v \in V(\mathcal{D}(n,p)) : d^{out}(v) = 0\}\right|$. Using this, we can crudely estimate that $|A(S_\ell, [n] \setminus S_\ell)| \geq |S_\ell| - X_0$. Since there is no arc from an ε-out-light vertex to an ε-in-light vertex, $|A(S_\ell, V_k^c)| = 0$ and, by Corollary 4.3(b), $|A(S_\ell, S_h)| \leq |S_h|$. Hence, $d^{out}_{V_k}(S) \geq |A(S_\ell, V_k \setminus S)| \geq |S_\ell| - X_0 - |S_h| \geq 6|S|/8 - X_0$. We claim that $X_0 \leq |S|/4$ with high probability, which implies $d^{out}_{V_k}(S) \geq |S|/2 \geq k$, as desired. To show our claim, let $t = 2^{-8}\log n$. Using the hypothesis on p, we can compute

$$\mathbb{E}[X_0] \leq n(1-p)^{n-1} \leq 2^{-8}\log n = t. \tag{3}$$

By Theorem 2.2, we have $\mathbb{P}(X_0 \geq \mathbb{E}[X_0] + t) \leq \exp\left(-t^2/(2(\mathbb{E}[X_0] + t/3)\right) \leq \exp(-3t/8) = o(1)$. Therefore, using (3), we obtain that $\mathbb{P}(X_0 \geq 2^{-7}\log n) = o(1)$, as claimed. □

Lemma 5.4. *Let $1 \leq k = o(\log n)$ and $V_k = \{v \in [n] : d^{in}(v) \geq k\}$. For all p with $(\log n - \log(2^{-8}\log n))/(n-1) \leq p \leq (1+o(1)\log n/n$ and $|V_k^c| = o(\log n)$, the following holds with high probability: for every subpartition $\mathcal{P}$ of V_k with $|\mathcal{P}| \geq 2$ or of the form $\mathcal{P} = \{X\}$ with $|X| \leq n - (\log n)/20$, it holds that*

$$\sum_{X \in \mathcal{P}} d^{in}_{V_k}(X) \geq k|\mathcal{P}|. \tag{4}$$

Proof. Let $p \le (1 + o(1)) \log n/(n-1)$, and $\varepsilon = \min(\tau_{5.1}, \tau_{5.2}, \tau_{5.3})$. We start by noticing the following.

Claim 5.5. *If $S \in \mathcal{P}$ and $|S| < |V_k| - p(n-1)/40$, then $d^{in}_{V_k}(S) \ge k$.*

Proof. We split the analysis into cases according to the size of S. If $|S| = 1$, let v be the sole element of S. If v is ε-in-light, then by Lemma 4.2 there are no incoming arcs from V_k^c, and therefore $d^{in}_{V_k}(v) = d^{in}(v) \ge k$, since $S \subset V_k$. If v is not ε-in-light, then $d^{in}_{V_k}(v) \ge d^{in}(v) - |V_k^c| = \Omega(\log n)$.

The other cases follow directly from previous lemmas. To spell out the details, note that if $2 \le |S| \le \varepsilon n$, we can apply Lemma 5.1. Finally, if $\varepsilon n \le |S| \le (1-\varepsilon)n$, let $T = V_k \setminus S$, and observe that $|S| \ge \varepsilon n - |V_k^c| \ge \varepsilon n/2$. By Lemma 2.5, there are at least $\varepsilon^2 n^2/8$ arcs from T to S. □

In order to show (4), note that if all sets $S \in \mathcal{P}$ satisfy $|S| < (1-\varepsilon)n$, we conclude, through the claim, that each $S \in \mathcal{P}$ satisfies $d^{in}_{V_k}(S) \ge k$. If, however, there exists $S \in \mathcal{P}$ with $|S| \ge (1-\varepsilon)n$, define $T = \bigcup_{X \in \mathcal{P}} X \setminus S$ and $U = V_k \setminus (S \cup T)$. If $|T \cup U| \ge p(n-1)/40$, then Lemma 5.3 implies that $d^{in}_{V_k}(S) = d^{out}_{V_k}(T \cup U) \ge k$. Combining it with the claim applied to the sets in $\mathcal{P} \setminus \{S\}$, we obtain (4). Otherwise, note that $|T \cup U| < p(n-1)/40 < (\log n)/20$, and therefore $|\mathcal{P}| \ne 1$ by hypothesis. Therefore, since $|\mathcal{P}| \ge 2$, T is nonempty and we can therefore apply Lemma 5.2. □

6. Proof of Theorem 1.1

Proof. Since $m(\mathcal{A}_k) \ge m(\mathcal{Z}_k)$ deterministically, it is sufficient to show that $m(\mathcal{A}_k) \le m(\mathcal{Z}_k)$ with high probability. Let $m_k = \lfloor n(n-1)p_k \rfloor$. By Lemmas 3.3 and 5.4, we have that, for $p = p_k + o(1/n)$, the following four properties hold with high probability:

(*i*) $X_{<k} > k$;

(*ii*) $X_{<k} \ll \log n$;

(*iii*) $\delta^{in} \ge k - 1$;

(*iv*) for every subpartition $\mathcal{P}$ of $V_k = \{v \in V : d^{in}(v) \ge k\}$ with $|\mathcal{P}| \ge 2$ or of the form $\mathcal{P} = \{X\}$ with $|X| \le n - (\log n)/20$, it holds that $\sum_{X \in \mathcal{P}} d^{in}_{V_k}(X) \ge k|\mathcal{P}|$.

Properties (*i*), (*ii*) and (*iii*) are monotone, and therefore also hold with high probability in $\mathcal{D}(n, m_k)$ by Lemma 2.4 (observing that $\sqrt{m_k(N - m_k)/N^3} = o(1/n)$). Property (*iv*) is not monotone, since V_k can change as new arcs appear in the process, increasing the number of subpartitions that need to be checked. We will show, however, that property (*iv*) is $(m_k, o(n))$-increasing using the following claim. Let $B_k(m) = \{v \in V : d_{\mathcal{D}(n,m)}(v) < k\}$.

Claim 6.1. *Let $m' = m_k - o(n)$. The event $\mathcal{E} = \{B_k(m')$ contains no arcs at time $m_{k+1}\}$ occurs with high probability.*

Proof. Every vertex of $B = B_k(m')$ is $(\varepsilon/2)$-in-light at time m', since $k \ll \sqrt{\log n}$. We now show that, with high probability, every vertex in B receives fewer than $(\varepsilon/2)\log n$ incoming arcs between time m' and m_{k+1}. Indeed, the number of incoming arcs a vertex $v \in B$ receives is stochastically dominated by a random variable $Y \sim \mathrm{Hyp}(N - m', n-1, m_{k+1} - m')$. Since $m_{k+1} - m' = \Theta(n \log\log n)$, it holds that $\mathbb{E}[Y] = O(\log\log n)$. Therefore, by Chernoff's inequality,

$$\mathbb{P}(Y \ge (\varepsilon/2)\log n) \le n^{-\varepsilon/2}.$$

Since by $|B| \ll \log n$ by property (*ii*), no vertex of B receives more than $(\varepsilon/2)\log n$ arcs with high probability by the union bound. Therefore, all vertices of B are ε-light at time m_{k+1}. This implies the desired result, since by Lemma 4.2 no two ε-light vertices are connected at time m_{k+1} with high probability. □

We now prove that property (*iv*) is $(m_k, o(n))$-increasing. Suppose an arc uv with $v \in V_k^c$ appears, increasing $d^{in}(v)$ to k. The "new" subpartitions of $V_k \cup \{v\}$ are obtained by taking a subpartition $\mathcal{P}$ of V_k and either adding v to an existing set $U \in \mathcal{P}$ (in which case (*iv*) trivially holds) or adding $\{v\}$ to $\mathcal{P}$. In this last case, by Claim 6.1 we have $d^{in}_{V_k}(v) = d^{in}(v) = k$ if $\mathcal{E}$ holds, and therefore the addition of $\{v\}$ to $\mathcal{P}$ increases $\sum_{X \in \mathcal{P}} d^{in}_{V_k}(X)$ by k and $|\mathcal{P}|$ by 1, which does not change the validity of property (*iv*). Therefore, under event $\mathcal{E}$, (*iv*) is an increasing property, which implies that (unconditionally) the property (*iv*) is $(m_k, o(n))$-increasing as desired.

Let $m^\star$ denote the time in the process such that $|B_k(m^\star)| = k$, i.e., $m^\star = m(\mathcal{Z}_k)$. By applying Lemma 3.3 to p_{k+1} and by monotonicity we have $\delta^{in}(\mathcal{D}(n, m_{k+1})) = k$, and therefore $m_k \le m^\star < m_{k+1}$. Therefore, by Claim 6.1, no two vertices of $B_k(m^\star) \subset B_k(m')$ are connected at time $m^\star$.

We are now ready to show that Frank's condition is satisfied on $\mathcal{D}(n, m^\star)$. Let $\mathcal{P}$ be a subpartition of $V(\mathcal{D}(n, m^\star))$ and let $\mathcal{P}'$ be the subpartition of $V_k = V_k(m^\star)$ obtained by restricting each set of $\mathcal{P}$ to V_k (and deleting empty sets). We first assume $\mathcal{P}'$ satisfies the hypothesis of property (*iv*) and let ℓ be the number of sets in $\mathcal{P}$ which lie entirely in $B_k(m^\star)$. Each one of these sets has in-degree at least $k-1$ by Claim 6.1. Using that $|\mathcal{P}| = |\mathcal{P}'| + \ell$ and $k(\ell - 1) \le k\ell$ (since $\ell \le k$), we have

$$\sum_{U \in \mathcal{P}} d^{in}(U) \ge \sum_{U \in \mathcal{P}'} d^{in}(U) + k\ell \ge k|\mathcal{P}'| + k\ell \ge k(|\mathcal{P}| - 1).$$

We now deal with the case $\mathcal{P}' = \{X\}$ with $|X| > n - (\log n)/20$. In this case, $|\mathcal{P}| \le 1 + B_k(m^\star) = k + 1$. We will show that $d^{in}(X) \ge k^2$ with high probability, which will imply that Frank's condition is satisfied on $\mathcal{D}(n, m^\star)$, since $k^2 \ge (|\mathcal{P}| - 1)k$. Lemma 3.1, applied to $\mathcal{D}(n, m_k)$, implies that for every $v \in B_k(m^\star)$, $d^{out}(v) = d^{out}(v) + d^{in}(v) - (k-1) \ge p_k(n-1)/5 - (k-1) = (1/5 + o(1)) \log n$. Therefore, by the condition on $|X|$, we have $|V \setminus X| \le \log n/20$ and $d^{in}(X) \ge |A(B_k(m^\star), X)| \ge k \cdot (1/5 - 1/20 - o(1)) \log n \ge k^3$, as claimed.

Since Frank's condition is satisfied, $\mathcal{D}(n, m^\star)$ has k arborescences and we have $m(\mathcal{A}_k) \le m(\mathcal{Z}_k)$, as desired. □

The above proof of Theorem 1.1, together with a refined proof of Lemma 3.3 (and the Central Limit Theorem) gives the following corollary. Due to space limitations, we omit its proof.

Corollary 6.2. *Let* $1 \ll k(n) \ll \sqrt{\log n}$ *and*

$$m = n\Big(\log n + (1 + (k-1)/\log n)((k-1)\log\log n - \log(k-1)!) - \log(k-1) + (k-1)^{-1/2}\sigma\Big)$$

for some $\sigma = \sigma(n) \in \mathbb{R}$. *Then* $\mathbb{P}(\mathcal{D}(n,m)$ *has* k *arc-disjoint arborescences*$) \sim \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\sigma} e^{-t^2/2}\, dt$.

When k is constant, a modification of the proof (involving proving convergence of the number of vertices of in-degree $k-1$ to a Poisson distribution) gives the following.

Corollary 6.3. *Let* $k = O(1)$ *and* $m = n(\log n + (k-1)\log\log n - \log((k-1)!) + c)$. *Then it holds that* $\mathbb{P}(\mathcal{D}(n,m)$ *has* k *arc-disjoint arborescences*$) \sim \exp(-e^{-c}) \sum_{\ell=0}^{k} e^{-c\ell}/\ell!$.

Acknowledgements

This work began while the first four authors were attending the 3º Workshop Paulista em Otimização, Combinatória e Algoritmos (WoPOCA). We would like to thank the organisers of this workshop for the invitation and for providing a productive research environment.

References

[1] Bal, D., Bennett, P., Cooper, C., Frieze, A., Prałat, P., 2016. Rainbow arborescence in random digraphs. Journal of Graph Theory 83, 251–265. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/jgt.21995, doi:10.1002/jgt.21995, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/jgt.21995.
[2] Catlin, P.A., Chen, Z.H., Palmer, E.M., 1993. On the edge arboricity of a random graph. Ars Combinatoria 35, 129–134.
[3] Chen, X., Li, X., Lian, H., 2013. Note on packing of edge-disjoint spanning trees in sparse random graphs. available as arXiv:1301.1097 .
[4] Erdős, P., Rényi, A., 1960. On the evolution of random graphs. Publications of the Mathematical Institute of the Hungarian Academy of Sciences 5, 17–60.
[5] Frank, A., 1979. Covering branchings. Acta Sci. Math. (Szeged) 41, 77–81.
[6] Gao, P., Pérez-Giménez, X., Sato, C.M., 2014. Arboricity and spanning-tree packing in random graphs with an application to load balancing, in: Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics, USA. p. 317–326.

[7] Hoppen, C., Parente, R., Sato, C., 2019. Packing arborescences in random digraphs. SIAM Journal on Discrete Mathematics 33, 438–453. URL: https://doi.org/10.1137/17M1151511, doi:10.1137/17M1151511, arXiv:https://doi.org/10.1137/17M1151511.
[8] Janson, S., Łuczak, T., Ruciński, A., 2000. Random graphs. Wiley-Interscience, New York.
[9] Nash-Williams, C.S.J.A., 1961. Edge-disjoint spanning trees of finite graphs. Journal of the London Mathematical Society s1-36, 445–450. doi:10.1112/jlms/s1-36.1.445.
[10] Palmer, E.M., Spencer, J.J., 1995. Hitting time for k edge-disjoint spanning trees in a random graph. Periodica Mathematica Hungarica 31, 235–240.
[11] Tutte, W.T., 1961. On the problem of decomposing a graph into n connected factors. Journal of the London Mathematical Society s1-36, 221–230. doi:10.1112/jlms/s1-36.1.221.