DIAMETER AND STATIONARY DISTRIBUTION OF RANDOM r-OUT DIGRAPHS

LOUIGI ADDARIO-BERRY*, BORJA BALLE†, AND GUILLEM PERARNAU*

ABSTRACT. Let $D(n,r)$ be a random r-out regular directed multigraph on the set of vertices $\{1,\ldots,n\}$. In this work, we establish that for every $r \geq 2$, there exists $\eta_r > 0$ such that $\mathrm{diam}(D(n,r)) = (1 + \eta_r + o(1)) \log_r n$. The constant η_r is related to branching processes and also appears in other models of random undirected graphs. Our techniques also allow us to bound some extremal quantities related to the stationary distribution of a simple random walk on $D(n,r)$. In particular, we determine the asymptotic behaviour of $\pi_{\max}$ and $\pi_{\min}$, the maximum and the minimum values of the stationary distribution. We show that with high probability $\pi_{\max} = n^{-1+o(1)}$ and $\pi_{\min} = n^{-(1+\eta_r)+o(1)}$. Our proof shows that the vertices with $\pi(v)$ near to $\pi_{\min}$ lie at the top of "narrow, slippery towers"; such vertices are also responsible for increasing the diameter from $(1+o(1)) \log_r n$ to $(1 + \eta_r + o(1)) \log_r n$.

1. INTRODUCTION

Call a random directed graph D with vertices $V(D) = \{v_1, \ldots, v_n\}$ a *random r-out digraph* if each vertex in $V(D)$ has out-degree r, and the nr heads of edges in $E(D)$ are iid and uniformly distributed over $V(D)$. We allow digraphs to have multiple edges and loops. It is useful to have a canonical construction: for each pair $(i,j) \in [n] \times [r]$, let $L_{i,j}$ be a uniformly random element of $[n]$, and write $D(n,d)$ for the random r-out digraph with vertex set $[n] = \{1,\ldots,n\}$ and edge set $\{(i, L_{i,j}) : (i,j) \in [n] \times [r]\}$.

Given a digraph D, for $u, v \in V(D)$ we write $\mathrm{dist}(u,v) = \mathrm{dist}_D(u,v)$ for the number of edges in a shortest oriented path from u to v, or set $\mathrm{dist}(u,v) = \infty$ if there exists no such path. The *diameter* of D is

$$\mathrm{diam}(D) = \max\{\mathrm{dist}(u,v) : u, v \in [n], \mathrm{dist}(u,v) < \infty\}. \tag{1.1}$$

Say D is *strongly connected* if $\mathrm{dist}(u,v) < \infty$ for all $u, v \in V(D)$. An induced subgraph $D[S]$ of D is a *strongly connected component* of D if $D[S]$ is strongly connected but for all S' with $S \subset S'$, $D[S']$ is not strongly connected. In order to study cases where D is not necessarily strongly connected, we write $D_0 = D_0(n,r)$ for the strongly connected component of $D(n,r)$ with the largest number of vertices (if there is more than one such component, D_0 is the one whose smallest labelled vertex is minimal).

Let $\lambda_r = \max\{\lambda : 1 - \lambda = e^{-r\lambda}\}$, and let

$$\eta_r = \frac{1}{\log_r(1 - \lambda_r)^{-1} - 1} = \frac{\log r}{\lambda_r r - \log r} \ . \tag{1.2}$$

Observe that $\lambda_r \to 1$ and $\eta_r \to 0$ when $r \to \infty$.

The main contribution of this paper is to determine $\mathrm{diam}(D(n,r))$ in probability[1].

---

*DEPARTMENT OF MATHEMATICS AND STATISTICS, MCGILL UNIVERSITY, MONTREAL, CANADA
†WORK DONE AT SCHOOL OF COMPUTER SCIENCE, MCGILL UNIVERSITY, MONTREAL, CANADA (NOW AT DEEPMIND, LONDON, UK)
*UNIVERSITAT POLITCNICA DE CATALUNYA, BARCELONA, SPAIN
E-mail addresses: louigi.addario@mcgill.ca, borja.balle@gmail.com, guillem.perarnau@upc.edu.

[1]A sequence of random variables X_n converges to X *in probability* if for every $\epsilon > 0$, $\mathbb{P}(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$. If $X_n/Y_n \to X$ in probability then we also write $X_n = (X + o_p(1))Y_n$.

Theorem 1.1. *For every $r \geq 2$, we have $diam(D(n,r)) = (1 + \eta_r + o_p(1)) \log_r n$ and $diam(D_0(n,r)) = (1 + \eta_r + o_p(1)) \log_r n$.*

As a consequence of our analysis we can obtain estimates for the smallest and largest probability in the stationary distribution of a random walk on $D(n,r)$. Before stating our result we give a few classical definitions that can be found in [26, 30].

Given $S \subset V(D)$, say that $D[S]$ is *attractive* if for all $v \in V(D)$ there is a directed path from v to S. It is easily seen that a digraph can contain at most one attractive strongly connected component $D[S]$. Grusho [19] proved that D_0 is the unique attractive strongly connected component of $D(n,r)$, with high probability[2]. Recently, Balle [5] showed that D_0 is ergodic whp.

It is well-known that if D has an attractive and ergodic strongly connected component D_0, then a simple random walk on D has a unique stationary distribution $\pi = \pi_D$ with support in D_0. We write $\pi_{\max}(D) = \max\{\pi_D(v) : v \in V(D)\}$ and $\pi_{\min}(D) = \min\{\pi_D(v) : \pi_D(v) > 0, v \in V(D)\}$, respectively.

Unlike in the undirected case, the stationary distribution of a directed graph is not usually determined by the degree sequence and one can find pathological examples where $\pi_{\min}(D)$ is exponentially small in the number of vertices of D. Moreover, having a large $\pi_{\min}$ is crucial to ensure that the mixing time of a random walk in D is small.

In the second part of the paper we will prove the following estimates for $\pi_{\max}(D(n,r))$ and $\pi_{\min}(D(n,r))$.

Theorem 1.2. *For every $r \geq 2$, we have $\pi_{\max}(D(n,r)) = n^{-1+o_p(1)}$ and $\pi_{\min}(D(n,r)) = n^{-(1+\eta_r)+o_p(1)}$.*

One of our main motivations for the study of $D(n,r)$ comes from the analysis of random Deterministic Finite Automata (DFA). DFA can be described using r-out regular directed graphs, where each of the r out-going arcs is labeled with distinct symbols from an alphabet of size r. A possible model of random DFA with n states can be obtained by considering a random r-out directed graph $D(n,r)$ and by assigning random labels on the out-going arcs. This particular model has been used in the literature on automata theory to study properties of "generic" DFA and average behavior of DFA algorithms (see Section 2.1 for a more detailed account of related work). In particular, the first analysis of the diameter of $D(n,r)$ appeared in the classic book of Trakhtenbrot and Barzdin [37, Theorem 5.5]. The authors of [37] showed that for every $r \geq 2$ there exists a constant $C_r \geq 1$ such that with high probability $diam(D) \leq C_r \log_r n$. This upper bound has been recently used, together with a lower bound on $\pi_{\min}(D(n,r))$, to stablish the possibility of efficently learning random DFA [1]. Thus, our results also have direct consequences on the complexity of learning random DFA.

Here we present some other graph-theoretic implications of Theorems 1.1 and 1.2:

- The results obtained can be easily transferred to random *simple* r-out digraphs. Let $D_{\text{sim}}(n,r)$ be chosen uniformly at random from the set of directed simple graphs (no loops or multiple edges) with vertex set $[n]$ such that each vertex has out-degree r. The conditional distribution of $D(n,r)$, given that it is simple, is precisely that of $D_{\text{sim}}(n,r)$. Furthermore, it is not hard to show (see [27]) that

$$\mathbb{P}(D(n,r) \text{ is simple}) = e^{-\Theta(r^2)}.$$

  In particular, this probability is bounded away from zero for fixed r, so any property that holds whp for $D(n,r)$ also holds whp for $D_{\text{sim}}(n,r)$.
- It is not hard to deduce from our arguments that for all $u, v \in V(D(n,r))$, conditional on the event that $v \in D_0(n,r)$, we have $\text{dist}_{D(n,r)}(u,v) = (1 + o_p(1)) \log_r n$.

---

[2]Here and for the remainder of the paper, *with high probability*, or *whp*, means with probability tending to 1 as $n \to \infty$.

This shows that the *typical distance* in $D(n,r)$ is $(1 + o_p(1)) \log_r n$. We leave the details to the interested reader.

- The random r-out, s-in digraph $D(n,r,s)$ is defined similarly to $D(n,r)$, but each vertex chooses s in-neighbours as well as r out-neighbours, all independently and uniformly at random; see [17]. In particular, $D(n,r) \stackrel{\mathrm{d}}{=} D(n,r,0)$[3]. It may be interesting to consider the diameter and the stationary distribution for $D(n,r,s)$ when $s \neq 0$. One case follows from Theorem 1.1: since the diameter of a digraph is the same as the diameter of the digraph obtained by flipping the direction of all the edges, $\mathrm{diam}(D(n,0,r)) \stackrel{\mathrm{d}}{=} \mathrm{diam}(D(n,r,0))$. In contrast, studying the stationary distribution of $D(n,0,r)$ seems less interesting: typically there will be many vertices with no out-edges where a simple random walk will eventually become stuck.

Outline. The paper is organized as follows. We start in Section 2 by discussing our motivation for addressing these problems and by putting our results in the context of other models for random (di)graphs. In Section 3 we introduce the notation that will be used throughout the paper and state some basic concentration inequalities and facts about branching processes. In Section 4.1 we finish the proof of the upper bound on the diameter of $D(n,r)$ (Theorem 1.1) assuming some technical estimates. The breadth-first search procedure that will be used to explore the graph is described in Section 5. In Section 6 we study the behaviour of the in-neighbourhoods of $D(n,r)$ by comparing them with Poisson Galton-Watson trees, while in Section 7 we study its out-neighbourhoods. In Section 8, we prove the technical estimates, completing the proof of the upper bound given in Section 4.1. The proof of the lower bound on the diameter of $D(n,r)$ (Theorem 1.1) occupies Section 9. We conclude the paper by proving Theorem 1.2 in Section 10.

2. Related Work

In this section we survey related work, both arising from literature on random DFA and from work on other models of random graphs.

2.1. **Random Deterministic Finite Automata.** A *deterministic finite automaton* (DFA) over an alphabet $\Sigma = \{\sigma_1, \ldots, \sigma_r\}$ is given by a set $V = \{v_1, \ldots, v_n\}$ and a function $L : [n] \times [r] \to [n]$. We think of the pair (V,L) as specifying a directed multigraph D with vertices V and edges $\{(v_i, L(i,j)) : i \in [n], j \in [r]\}$; every vertex of D has out-degree r, and the r edges leaving a vertex v are labeled with distinct symbols from Σ. In addition, a DFA is equipped with a distinguished vertex s called the *initial* state, and with a binary labelling $B : V(D) \to \{0,1\}$; the vertices in $B^{-1}(\{1\})$ are the *accepting* states of the DFA. The DFA is formally given by the tuple $Q = (V, \Sigma, L, s, B)$.

Let $\Sigma^\star$ denote the set of all finite strings with symbols in Σ. Words $w = w_1 w_2 \ldots w_t \in \Sigma^\star$ correspond to walks $x_0(w), x_1(w), \ldots, x_t(w)$ on V: $x_0 = s$ and, for $1 \leq i \leq t$, x_i is reached from x_{i-1} by following the edge with label w_i. We write $Q(w) = x_t(w)$ for the final state of the walk. The DFA *accepts* the word w if $B(Q(w)) = 1$. The set $L(Q) = \{w \in \Sigma^\star : B(Q(w)) = 1\}$ is the *language* recognized by the DFA. The set of languages recognized by some DFA are precisely the *regular languages*.

To see the connection with random out-regular graphs, observe that we may build a uniformly random DFA with n labelled states and alphabet of size r as follows. Let $D(n,r)$ be as in the first paragraph of the paper, using the random variables $(L_{i,j} : (i,j) \in [n] \times [r])$. Then for $(i,j) \in [n] \times [r]$, let $L(i,j) = L_{i,j}$; equivalently, assign label σ_i to edge $(i, L_{i,j})$. Choose the starting state s uniformly at random from $[n]$, and choose B uniformly at random from the set of functions $f : [n] \to \{0,1\}$.

---

[3]For any two random variables X and Y, we use the notation $X \stackrel{\mathrm{d}}{=} Y$ to denote that the corresponding probability distributions are equal.

DFA and regular languages play a crucial role in language theory and there is a vast literature on algorithms over DFA, ranging from minimization and property testing, to synthesis, learning and composition [34]. Next we describe the problem of learning random DFA from uniformly sampled examples, whose complexity is related to the diameter and stationary distribution of $D(n, r)$ [1].

Learning regular languages from different sources of information is a prominent problem in computational learning theory [24], which is most often studied within the context of so-called grammatical inference problems [16]. An important problem in this area concerns the possibility of learning regular languages under the *probably approximately correct (PAC)* learning model introduced by Valiant [38]. Roughly speaking, this asks for an efficient algorithm such that, when supplied with a large enough sample containing iid strings drawn from some arbitrary probability distribution μ on $\Sigma^\star$ and labels indicating whether each string belongs to some hidden regular language, the algorithm outputs a representation of a regular language (e.g. a DFA) which is close to the hidden regular language in a sense that depends on the distribution which generated the sample strings. Several results from the 90's indicate that, in its full generality, PAC learning of DFA is hard due to complexity-theoretic as well as cryptographic reasons [23, 32] (see also the recent strengthened result [13]). A natural question to ask in such a scenario is whether there exists a reasonable simplification of the problem for which a positive answer is possible. This requires one to come up with scenarios that rule out the worst-case problems arising from specially crafted regular languages and distributions over examples appearing in the proofs of the aforementioned lower bounds.

One possibility is to study the average case. This approach can be formalized by considering regular languages defined by random DFA. In particular, one can ask for an algorithm that with high probability (as the number of the states in the DFA goes to infinity) can learn the regular language recognized by a random DFA. There exists evidence suggesting that such relaxation might not be enough to achieve efficient learning in general: it was recently showed by Angluin et al. that generic instances of DFA (as well as decision trees and DNF formulas) are hard to learn from statistical queries when examples can be sampled from an arbitrary distribution [3]. Nevertheless, prior to Angluin et al.'s result it was showed that generic decision trees and generic DNF formulas can be efficiently learned when samples are drawn according to the uniform distribution [20, 35].

In view of the panorama described in the previous paragraphs, a natural question to ask is whether random DFA can be efficiently learned when sample strings are drawn from the uniform distribution. More precisely, one would like to answer the following sorts of questions. Fix a uniformly random DFA Q with states $[n]$ and alphabet $[r]$. Then fix $m \in \mathbb{N}$ and let $(\mathbf{x}_i, i \geq 1)$ be iid words sampled uniformly at random from $[r]^m$.

    (1) Given the sequences $(\mathbf{x}_i, i \geq 1)$ and $(B(Q(\mathbf{x}_i)), i \geq 1)$, is it possible to construct a DFA $\hat{Q}$ that recognizes the same language as Q with high probability?

    (2) Given the sequences $(\mathbf{x}_i, i \geq 1)$, $(Q(\mathbf{x}_i)), i \geq 1)$ and $(B(Q(\mathbf{x}_i)), i \geq 1)$, is it possible to construct a DFA $\hat{Q}$ that recognizes the same language as Q with high probability?

In both cases, if the answer is yes then it is natural to ask for *efficient* algorithms (average case running time polynomial in n, m, r, and any other parameters involved). The questions can be weakened by only requiring that $\hat{Q}$ recognizes the same set of words *of length m*. A further weakening is to only require that $\mathbb{P}(\hat{Q}(\mathbf{y}) = Q(\mathbf{y})) > 1 - \epsilon$ when $\mathbf{y}$ is uniformly distributed over $[r]^m$.

The results in [1] establish that in order to answer the second question, it would be sufficient to understand several specific properties of a random walk on a randomly generated DFA. When a string is sampled from the uniform distribution over $[r]^m$ and is labeled according to the state that it reaches, the label immediately corresponds to the final state of

a simple random walk of length m over the DFA starting from the initial state. Thus, the analysis of the algorithm in [1] relies on bounds on the diameter, stationary distribution, and mixing time on random r-out regular digraphs.

Angluin and Chen base their analysis on previous results about the strongly connected component D_0 of $D(n, r)$. The first of such results is due to Grusho, who showed that the strongly connected component D_0 is attractive and its order satisfies $|V(D_0)| = (1 + o_p(1))\lambda_r \cdot n$ [19] (we remark that its order is also the asymptotic order of the giant component in the Erdős-Rényi random graph $G(n, r/n)$). The average-case analysis of algorithms for the minimization of Deterministic Finite Automata (DFA) has led to multiple rediscoveries of Grusho's result [7, 11, 12]. As mentioned above, the diameter of $D(n, r)$ was first studied by Trakhtenbrot and Barzdin in [37], who showed that for every $r \geq 2$, we have $\mathrm{diam}(D(n, r)) = O(\log_r n)$, whp.

Several other properties of random DFA have been studied, both in learning theory and in other contexts, using the $D(n, r)$ model. For example, first Korshunov's group, and later Nicaud's group, have studied the probability that random DFA exhibit particular structures, mainly motivated by the analysis of sample and reject algorithms for enumeration of subclasses of automata (see [28] and references therein). Motivated by worst-case hardness results for learning a DFA, Angluin and co-authors have used properties of random DFA to study the problem of learning a generic DFA [2, 3]. The average-case complexity of DFA minimization algorithms has also received some attention recently [6, 15]. Finally, a series of results have led to a solution of the long-standing Černý conjecture about synchronization of finite automata in the case of random DFA [8, 29, 36].

2.2. **Diameter and stationary distribution of other random graph models.** In this subsection we describe some previous results on the diameter and the stationary distribution of certain random graph models and relate them to Theorem 1.1 and to Theorem 1.2. This provides an intuition for the results we have obtained on the diameter of $D(n, r)$. We consider the following models of random (di)graphs.

- For $p \in [0, 1)$, $G(n, p)$ is the random graph with vertex set $[n]$ in which every edge is included independently with probability p.
- For $d \in \mathbb{N}$, $G(n, d)$ is the random d-regular simple graph with vertex set $[n]$ chosen uniformly at random among all such graphs.
- For $p \in [0, 1)$, $D(n, p)$ is the random digraph with vertex set $[n]$ in which every oriented edge is included independently with probability p.

For an undirected graph $G = (V, E)$ and $u, v \in V$ we write $\mathrm{dist}_G(u, v)$ for the minimum number of edges in a path from u to v, or set $\mathrm{dist}_G(u, v) = \infty$ if there exists no such path. The diameter of G is then defined just as in (1.1). Bollobás and Fernandez de la Vega [10] studied the diameter of $G(n, d)$ and showed that for every integer $r \geq 2$, we have

$$\mathrm{diam}(G(n, r + 1)) = (1 + o_p(1))\log_r n . \tag{2.1}$$

The diameter of $G(n, p)$ was recently studied by Riordan and Wormald [33], who showed that for every constant $r > 0$, we have

$$\mathrm{diam}(G(n, r/n)) = (1 + 2\eta_r + o_p(1))\log_r n . \tag{2.2}$$

In fact, they proved a stronger result, showing convergence in distribution of the diameter after appropriate recentering and rescaling. Although, in this paper we only determine the first order asymptotic behaviour of $\mathrm{diam}(D(n, r))$, it is possible that similar techniques as the ones presented in [33] could be of use to determine its second order term. The extra term $2\eta_r$ is essentially due to the existence of "remote" vertices in the giant component of $G(n, r/n)$, whose neighbourhoods are exceptionally small up to distance about $\eta_r \log_r n$.

Our result on the diameter of $D(n, r)$ from Theorem 1.1 can be related to (2.1) and (2.2) in the following way. Given $u, v \in [n]$, one way to determine $\mathrm{dist}_{D(n,r)}(u, v)$ is to perform an

outward breadth-first search (BFS) starting at u, to perform an *inward* BFS (i.e. following edges from head to tail) starting at v, and to stop at the first time the two searches uncover a common vertex. (See Section 5.1 for a careful definition of breadth-first search.) This technique was used by Bollobás and Fernandez de la Vega in [10]. Since the BFS explores vertices in order of distance, such a procedure is guaranteed to build a shortest path from u to v.

On the one hand, in the outward BFS of $D(n, r)$ starting from u, every vertex has exactly r out-edges when explored. Similarly, in a BFS exploration of $G(n, r + 1)$, when a vertex v is discovered via an edge from one of its neighbours, this leaves r edges to unveil when v is itself explored (unless v is discovered multiple times, which at least at the start of the BFS is unlikely). Thus, a BFS of $G(n, r + 1)$ looks similar to an outward BFS of $D(n, r)$.

On the other hand, in the inward BFS of $D(n, r)$ starting from v (or at least near the start of the process) the number of in-edges arriving at a vertex are roughly distributed as a Binomial random variable with n trials and success probability r/n. Thus, a BFS of $G(n, r/n)$ looks similar to an inward BFS of $D(n, r)$.

The preceding paragraphs suggest that shortest paths in $D(n, r)$ are in some sense hybrids of shortest paths in $G(n, r + 1)$ and in $G(n, r/n)$. This, together with (2.1) and (2.2), provides some intuition for the value of the diameter of $D(n, r)$ from Theorem 1.1: it is the average of the limit values in those formulae.

An undirected model which is closely related to $D(n, r)$ is the *random r-out graph*, obtained by selecting a directed graph according to $D(n, r)$ and then forgetting about the directions of the edges. This model was introduced by Fenner and Frieze [17] and its diameter is whp $(1 + o_p(1)) \log_{2r} n$ (see [18, Exercise 16.5.3]).

There is interesting related work on distances in graphs with random edge weights. We mention in particular the paper of Janson [21] on typical and extreme distances in randomly edge-weighted complete graphs, and the subsequent work by Bhamidi and van der Hofstad [9], which establishes distributional convergence for the diameter.

To conclude this section, we discuss the stationary distribution of a simple random walk in these other models. While in undirected graphs the stationary distribution (if it exists) is completely determined by the degrees of the vertices, this is not the case in directed graphs. Cooper and Frieze [14] give a very precise description of the stationary distribution of $D(n, c/n)$ when $c = c(n) > (1 + \epsilon) \log n$, for any constant $\epsilon > 0$, and use their result to compute the cover time of $D(n, c/n)$. It is worth noticing that for such values of c, both the in-degrees and out-degrees are of logarithmic order and concentrated around their expected values, which turns to be very useful for the analysis. It seems harder to find an interesting question about the stationary distribution of $D(n, c/n)$ when $c = c(n) < (1 - \epsilon) \log n$ since like in random r-in regular digraphs, typically there are vertices with no out-edges.

3. Notation and preliminaries

We write $[n] = \{1, 2, \ldots, n\}$, $\mathbb{N} = \{1, 2, \ldots\}$, and $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$. The notation $A \subset B$ allows that $A = B$; we write $A \subsetneq B$ for strict containment. Unless we explicitly indicate otherwise, asymptotic notations will always refer to the case $n \to \infty$. We omit floors and ceilings when doing so improves readability. All logarithms are natural unless a subscript specifies otherwise.

For any two random variables X and Y, we use the notation $X \stackrel{\mathrm{d}}{=} Y$ to denote that the corresponding probability distributions are equal. For random variables X, Y, we write $X \preceq Y$, and say X is stochastically dominated by Y, if $\mathbb{P}(X \leq t) \geq \mathbb{P}(Y \leq t)$ for all $t \in \mathbb{R}$. We say $X_1, \ldots, X_k$ are *independently* stochastically dominated by $Y_1, \ldots, Y_t$ if $\mathbb{P}(X_i \leq t_i, 1 \leq i \leq k) \geq \prod_{i=1}^{k} \mathbb{P}(Y_i \leq t_i)$, for all $(t_1, \ldots, t_k) \in \mathbb{R}^k$.

3.1. **Digraphs.** Let $D = (V(D), E(D))$ be a directed graph. For $S, S' \subseteq V(D)$ let

$$E(S, S') = E_D(S, S') = \big\{ (u, v) \in E \,\big|\, u \in S, v \in S' \big\} \,.$$

Given $S \subset [n]$, $D[S] = (S, E(S, S))$ is *the subgraph of D induced by S*.

Given $u \in [n]$ and an integer $k \geq 0$, write $N_k^+(D, u) = \{v \in [n] : \mathrm{dist}(u, v) = k\}$ and $N_{\leq k}^+(D, u) = \cup_{j \leq k} N_j^+(u)$. Similarly, let $N_k^-(D, u) = \{v \in [n] : \mathrm{dist}(v, u) = k\}$ and $N_{\leq k}^-(D, u) = \cup_{j \leq k} N_j^-(u)$. We write $N^+(D, u) = N_1^+(D, u)$ and $N^-(D, u) = N_1^-(D, u)$. We also let $d_k^+(D, u) = |N_k^+(D, u)|$, and define $d_{\leq k}^+(D, u)$, $d_k^-(D, u)$, and $d_{\leq k}^-(D, u)$ correspondingly. We write $N_k^+(u) = N_k^+(D, u)$, etcetera, when D is clear from context.

3.2. **Concentration Inequalities.** We write $\mathbf{Bin}(N, p)$ to denote a Binomial random variable with N trials and success probability p. We write $\mathbf{Po}(r)$ to denote a Poisson random variables with parameter r. We also write $\mathbf{Ber}(p)$ to denote a Bernoulli random variable with success probability p.

We will use the following version of Chernoff's bound for large deviations that can be found in [22].

Lemma 3.1 (Chernoff's inequality)**.** *For any $t \geq 0$ we have*

$$\mathbb{P}(\mathbf{Bin}(N, p) \geq Np + t) \leq e^{-\frac{t^2}{2(Np + t/3)}} \,, \tag{3.1}$$

and

$$\mathbb{P}(\mathbf{Bin}(N, p) \leq Np - t) \leq e^{-\frac{t^2}{2Np}} \,. \tag{3.2}$$

We will also use Chebyshev's inequality: for any random variable X and any $t \geq 0$, $\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\sigma^2}{t^2}$, where $\sigma^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.

3.3. **Trees and branching processes.** In any rooted tree, we view edges as oriented from child to parent. Fix a rooted tree T with root $v = v(T)$. Then for all $u \in V(T)$, $N^-(T, u)$ is the set of children of u. For $u \neq v$, let $p(u) = p_T(u)$ be the parent of u in T, so $N^+(T, u) = \{p(u)\}$. For $k \geq 0$, let $T_{\leq k}$ be the subtree of T induced by $N_{\leq k}^-(T, v) = \{u \in V(T) : \mathrm{dist}_T(u, v) \leq k\}$; we view $T_{\leq k}$ as rooted at v. Also, write $T_k = N_k^-(T, v) = \{u \in V(T) : \mathrm{dist}_T(u, v) = k\}$.

A *plane tree* is a rooted tree in which the children of each node have a left-to-right order. Given a plane tree T, there is a canonical labelling of $V(T)$ by distinct elements of $\{\emptyset\} \cup \bigcup_{i \geq 1} \mathbb{N}^i$, as follows. The root v has label $\emptyset$; its children are labelled from left to right as $1, \ldots, |N^-(T, v)|$. Given $u \in V(T_k)$ with label $w_1 w_2 \ldots w_k$, the children of u are labelled from left to right as $(w_1 \ldots w_k i, 1 \leq i \leq |N^-(T, u)|)$.

Conversely, given a rooted tree T with t vertices and an ordering of $V(T)$ as $v_1, \ldots, v_t$, say $w \in V(T)$ has index j if $w = v_j$, for $1 \leq j \leq t$. We view $V(T)$ as a plane tree using the convention that the children of each vertex are listed from left to right in increasing order of index. If $V(T) \subset \mathbb{N}$ then we always use the ordering inherited from $\mathbb{N}$. Thus, for a *rooted* tree T with $V(T) \subset \mathbb{N}$, and a *plane* tree T', we say T and T' are isomorphic, and write $T \cong T'$, if T and T' are identical when viewed as plane trees.

Finally, fix a non-negative, integer-valued random variable ξ. A *Galton-Watson tree* with branching mechanism ξ is the random, potentially infinite family tree $\mathcal{T}^\xi$ of a branching process started from a single individual, in which each individual reproduces independently according to ξ (i.e. the number of offspring of each individual has the distribution of ξ). The random tree $\mathcal{T}^\xi$ is naturally viewed as a plane tree; see [25] for details and a careful construction. If ξ is $\mathbf{Po}(r)$ distributed we call $\mathcal{T}^\xi$ a Poisson(r) Galton-Watson tree.

4. Proof of Theorem 1.1

In this section we describe our proof technique for Theorem 1.1, and prove it assuming some technical estimates. We prove such estimates in Sections 8 and 9.

4.1. **Upper bound.** This subsection sketches the proof of our upper bound on the diameter of $D(n,r)$. For the remainder of this subsection, let $D = D(n,r)$ and write $d_k^-(v) = d_k^-(D,v)$, $N_k^-(v) = N_k^-(D,v)$ etcetera.

In order to derive an upper bound on the diameter of D we first show in Lemma 4.1 that for any fixed vertex v, the in-neighbourhood $N_k^-(v)$ is either empty or large, for some k slightly larger than $\eta_r \log_r n$, whp. Then, in Proposition 4.2, we show that conditional on $N_k^-(v)$ being large (and a set of other technical conditions), the distance between any vertex u and $N_{\leq k}^-(v)$ is at most $\log_r n$, whp. Putting both things together, we are able to prove the upper bound for $\mathrm{diam}(D)$.

Lemma 4.1. *Fix* $v \in [n]$, *let* $k_0(v) = \min\{k : \ d_k^-(v) \notin (0, \log^4 n)\}$. *Then for every* $\epsilon \in (0, 1/10)$, *there exists* $\delta > 0$ *such that*

$$\mathbb{P}\left(k_0(v) > (\eta_r + \epsilon)\log_r n \ \text{or} \ d_{\leq k_0(v)}^-(v) \geq \log^7 n\right) = O(n^{-(1+\delta)}).$$

Next, if $v \in [n]$ has $N^+(v) = \{v\}$ (i.e., if all edges leaving v are self-loops) then call v a *loop vertex*. Let E_{SL} be the event that D contains some loop vertex. Each vertex is independently a loop vertex with probability n^{-r}, so

$$\mathbb{P}(E_{\mathrm{SL}}) = 1 - (1 - n^{-r})^n = \Theta(n^{1-r}). \tag{4.1}$$

Note that if $r \geq 3$, the probability of a given vertex being a loop vertex is $O(n^{-3})$. This bound is small enough that it would allow union bounds over pairs of vertices, which would simplify some proofs. Since we aim to prove our result also in the case $r = 2$, we need to be a bit more careful in our computations.

Proposition 4.2. *Fix* $k, n \in \mathbb{N}$ *and* $u, v \in [n]$. *Let* $E_k = \{d_k^-(v) \geq \log^4 n, d_{\leq k}^-(v) \leq \log^7 n\}$, *and fix a graph* H *with* $v \in V(H) \subset [n]$ *such that* $\mathbb{P}(D[N_{\leq k}^-(v)] = H, E_k, \overline{E_{SL}}) > 0$. *Then*

$$\mathbb{P}(\mathrm{dist}(u, N_{\leq k}^-(v)) > \log_r n - \log_r \log_r n, \overline{E_{SL}} \mid D[N_{\leq k}^-(v)] = H) = O(n^{-3}),$$

the preceding bound holding uniformly over k *and over all* H *satisfying the above conditions.*

We prove Lemma 4.1 and Proposition 4.2 in Section 8. In the remainder of the section, we finish the proof of the upper bound from Theorem 1.1, assuming Lemma 4.1 and Proposition 4.2.

Proof of the upper bound in Theorem 1.1. Fix $\epsilon > 0$. We show that $\mathbb{P}(\mathrm{diam}(D) \leq (1+\eta_r + \epsilon)\log_r n) = 1 - o(1)$. Since $\mathrm{diam}(D_0(n,r)) \leq \mathrm{diam}(D(n,r))$, the same bound immediately holds for $D_0(n,r)$.

Let $k^* = (\eta_r + \epsilon)\log_r n$ and let $\ell^* = \log_r n - \log_r \log_r n - 1$. By (4.1) for every $r \geq 2$, we have that $\mathbb{P}(E_{\mathrm{SL}}) = O(n^{-1})$. So

$$\mathbb{P}(\exists u, v \in [n], \ \mathrm{dist}(u,v) \in (k^* + \ell^*, \infty))$$
$$\leq O(n^{-1}) + \sum_{v \in [n]} \mathbb{P}(\exists u \in [n], \ \mathrm{dist}(u,v) \in (k^* + \ell^*, \infty), \overline{E_{\mathrm{SL}}}).$$

Define $k_0 = k_0(v)$ as in Lemma 4.1, and let $E = \{k_0 \leq k^*, d^-_{\leq k_0}(v) < \log^7 n\}$. Then by Lemma 4.1,

$$\mathbb{P}(\exists u \in [n], \text{dist}(u,v) \in (k^* + \ell^*, \infty), \overline{E_{\text{SL}}})$$
$$\leq \mathbb{P}(\overline{E}, \overline{E_{\text{SL}}}) + \mathbb{P}(\exists u \in [n], \ \text{dist}(u,v) \in (k^* + \ell^*, \infty), E, \overline{E_{\text{SL}}})$$
$$= O(n^{-(1+\delta)}) + \mathbb{P}(\exists u \in [n], \ \text{dist}(u,v) \in (k^* + \ell^*, \infty), E, \overline{E_{\text{SL}}}),$$

Now let $\mathcal{H}$ be the set of graphs H with $v \in V(H)$ such that $\mathbb{P}(D[N^-_{\leq k_0}(v)] = H, E, \overline{E_{\text{SL}}}) > 0$. Then

$$\mathbb{P}(\exists u \in [n], \ \text{dist}(u,v) \in (k^* + \ell^*, \infty), E, \overline{E_{\text{SL}}})$$
$$\leq \sup_{H \in \mathcal{H}} \mathbb{P}(\exists u \in [n], \ \text{dist}(u,v) \in (k^* + \ell^*, \infty), \overline{E_{\text{SL}}} \mid D[N^-_{\leq k_0}(v)] = H)$$
$$\leq \sup_{H \in \mathcal{H}} \sum_{u \in [n]} \mathbb{P}(\text{dist}(u,v) \in (k^* + \ell^*, \infty), \overline{E_{\text{SL}}} \mid D[N^-_{\leq k_0}(v)] = H).$$

For each $H \in \mathcal{H}$ there is a (non-random) constant $k = k(H)$ such that if $D[N^-_{\leq k_0}(v)] = H$ then $k_0 = k(H)$, so the events $D[N^-_{\leq k_0}(v)] = H$ and $D[N^-_{\leq k}(v)] = H$ are identical. Furthermore, given that $D[N^-_{\leq k}(v)] = H$ we either have $d^-_k(v) = 0$ or $d^-_k(v) \geq \log^4 n$. In the latter, if $D[N^-_{\leq k}(v)] = H$ then E_k occurs, so $\mathbb{P}(D[N^-_{\leq k}(v)] = H, E_k, \overline{E_{\text{SL}}}) > 0$, so we can apply Proposition 4.2 (with $k = k(H) = k_0$). In both cases we obtain

$$\mathbb{P}(\text{dist}(u,v) \in (k^* + \ell^*, \infty), \overline{E_{\text{SL}}} \mid D[N^-_{\leq k_0}(v)] = H)$$
$$= \mathbb{P}(\text{dist}(u,v) \in (k^* + \ell^*, \infty), \overline{E_{\text{SL}}} \mid D[N^-_{\leq k}(v)] = H)$$
$$= O(n^{-3}),$$

and conclude that

$$\mathbb{P}(\exists u, v \in [n], \ \text{dist}(u,v) \in (k^* + \ell^*, \infty))$$
$$= O(n^{-1}) + \sum_{v \in [n]} O(n^{-(1+\delta)}) + \sum_{u,v \in [n]} O(n^{-3}) = O(n^{-\delta}).$$

It follows that with high probability $\text{diam}(D) \leq k^* + \ell^* \leq (1 + \eta_r + \epsilon) \log_r n$. $\qquad\square$

4.2. **Lower bound.** This subsection sketches the proof of our lower bound on the diameter of $D(n,r)$. Together with the proof in Subsection 4.1, this concludes the proof of Theorem 1.1.

In order to derive a lower bound on the diameter of D we first introduce the concept of flags; that is, vertices v which have atypically small but non-empty in-neighbourhoods $N^-_k(v)$ that induce a tree in D, for values of k slightly smaller than $\eta_r \log_r n$ (see Definition 4.3). In Lemma 4.4 we show that the probability that a given vertex is a flag is relatively large, and in Corollary 4.5 we show that whp there exists many flags. Moreover, there are no flags outside of D_0 (Lemma 4.6). Using the existence of flags in D_0, we conclude the lower bound for $\text{diam}(D)$.

Let us now precisely define the notion of flag.

Definition 4.3. For $v \in [n]$, let $k_1 = k_1(v) = \min\{k : \ d^-_k(v) \geq \log^4 n\}$; this is ∞ if $d^-_k(v) < \log^4 n$ for all k. A vertex v is an ϵ-*flag* (or simply a *flag*, if ϵ is clear from context) if $k_1 \in [(\eta_r - \epsilon/2) \log_r n, \infty)$, $|N^-_{\leq k_1}(v)| \leq \log^7 n$, and $D[N^-_{\leq k_1}(v)]$ is a tree. We write $F = F(\epsilon) \subseteq [n]$ for the set of ϵ-flags.

The condition that $D[N^-_{\leq k_1}(v)]$ is a tree means that along any shortest path from $N^-_{k_1}(v)$ to v, at each node w there are $(r-1)$ possible "wrong turns" that lead to $[n] \setminus N^-_{k_1}(v)$. We will use this when bounding $\pi_{\min}$ in Section 10.

In order to find a vertex $v \in [n]$ such that there exists $u \in [n]$ with $\mathrm{dist}(u, v) \geq (1 + \eta_r - \epsilon) \log_r n$, we will look at $F(\epsilon)$. The following lemma shows that the probability a given vertex is an ϵ-flag is relatively large. Its proof uses the same ideas on Poisson branching processes also displayed in the proof of Lemma 4.1.

Lemma 4.4. *For $\epsilon > 0$ sufficiently small, there is a $\beta > 0$ such that $\mathbb{P}(v \in F(\epsilon)) \geq n^{\beta-1}$.*

To ascertain that $D(n, r)$ will contain ϵ-flags whp, we proceed as follows. In Lemma 4.4 we showed that the expected number of ϵ-flags goes to infinity as $n \to \infty$. Then, we obtain the following upper bound on the probability that two vertices are simultaneously ϵ-flags (Corollary 9.2),

$$\mathbb{P}(u, v \in F(\epsilon)) \leq (1 + o(1)) \left(\mathbb{P}(u \in F(\epsilon)) + \frac{\log^{15} n}{n} \right) \mathbb{P}(v \in F(\epsilon)) \ . \tag{4.2}$$

The previous inequality allows us to use a second moment argument to prove the following:

Corollary 4.5. *For all $\epsilon > 0$, $\mathbb{P}(F(\epsilon) = \emptyset) = o(1)$.*

Proof. By Lemma 4.4 and linearity of expectation there is $\beta > 0$ such that $\mathbb{E}(|F|) = n\mathbb{P}(1 \in F) \geq n^\beta$. Next, by (4.2), for n large we have

$$\begin{aligned}
\mathbb{E}(|F|^2) &= \sum_{u,v \in [n]} \mathbb{P}(u, v \in F) \\
&= n(n-1)\mathbb{P}(1, 2 \in F) + n\mathbb{P}(1 \in F) \\
&\leq (1 + o(1))n(n-1)\mathbb{P}(1 \in F)(\mathbb{P}(2 \in F) + \log^{15} n/n) + n\mathbb{P}(1 \in F) \\
&\leq (1 + o(1)(n\mathbb{P}(1 \in F))^2 + (n \log^{15} n)\mathbb{P}(1 \in F) \\
&= (1 + o(1))(n\mathbb{P}(1 \in F))^2 .
\end{aligned}$$

The result follows by Chebyshev's inequality. $\qquad\square$

Once we know that whp there are many ϵ-flags, we prove that they belong to the unique attractive component.

Lemma 4.6. *For every $\epsilon > 0$, $\mathbb{P}(F(\epsilon) \setminus V(D_0) \neq \emptyset) = o(1)$.*

We use the previous results to conclude the proof of the lower bound.

Proof of Theorem 1.1 (Lower Bound). Fix $\epsilon > 0$ and write $k^* = (\eta_r - \epsilon/2) \log_r n$. Suppose that D_0 is attractive, that $|D_0| \geq n/2$, and that $F(\epsilon) \subset D_0$. Suppose that for all $v \in [n]$ and $j \geq 0$, $d_{\leq j}^-(v) \leq (r + \epsilon/2)^j \log_r^2 n$. Under these assumptions, if $u \in F(\epsilon)$, then $u \in D_0$ and, as $k^* = (\eta_r - \epsilon/2) \log_r n \leq k_1$, we also have $d_{\leq k^*}^-(u) \leq \log^7 n$. So for all $j \geq 0$,

$$d_{\leq k^*+j}^-(u) \leq (r + \epsilon/2)^j \log_r^9 n.$$

Writing $j_0 = \inf\{j : V(D_0) \subset N_{\leq k^*+j}^-(u)\}$, it follows that $(r + \epsilon/2)^{j_0} \log_r^9 n \geq n/2$. Provided ϵ is chosen small enough, for n large this implies that $j_0 \geq (1 - \epsilon/2) \log_r n$, so there is some node $v \in V(D_0)$ with $\mathrm{dist}(v, u) \geq k^* + (1 - \epsilon/2) \log_r n = (\eta_r + 1 - \epsilon) \log_r n$. Altogether, this yields

$$\begin{aligned}
&\mathbb{P}(\mathrm{diam}(D_0) < (\eta_r + 1 - \epsilon) \log_r n) \\
&\leq \mathbb{P}(D_0 \text{ is not attractive}) + \mathbb{P}(|V(D_0)| < n/2) \\
&\quad + \mathbb{P}(F(\epsilon) = \emptyset) + \mathbb{P}(\exists u \in F(\epsilon) \setminus D_0) + \mathbb{P}(\exists v \in [n], j \geq 0 : d_j^-(v) > (r + \epsilon/2)^j \log_r^2 n)
\end{aligned}$$

The first two probabilities were shown to tend to 0 in [19]. The third tends to 0 by Corollary 4.5, and the fourth by Lemma 4.6. Proposition 6.1, which provides an estimate on the growth of the in-neighbourhoods in D, shows that the last tends to 0. As $\epsilon > 0$ was arbitrarily small, the lower bound on $\mathrm{diam}(D_0)$ follows; since $\mathrm{diam}(D) \geq \mathrm{diam}(D_0)$ so does the lower bound on $\mathrm{diam}(D)$. $\qquad\square$

5. Breadth-first search and conditioning

In this section we describe the breadth-first search (BFS) procedures, which are fundamental to our analysis, and use them to prove a handful of stochastic domination results for neighbourhood sizes in $D(n, r)$.

5.1. **Outward and inward breadth-first search.** Fix a digraph D together with an ordering of its vertices $V(D)$ as $(v_1, \ldots, v_n)$. The *outward breadth-first search* (oBFS) starting from node $v \in V(D)$ is a deterministic process $((R_i^+(D, v), S_i^+(D, v)), i \geq 0)$, defined as follows. At time i, $R_i^+ = R_i^+(D, v)$ is the set of explored vertices and $S_i^+ = S_i^+(D, v)$ is the *sequence* of discovered but not yet explored vertices; S is treated as a first-in first-out queue. Node $w \in V(D)$ has *index* j if $w = v_j$.

Begin with $R_0^+ = \emptyset$ and $S_0^+ = (v)$. Now fix $i \geq 0$ and suppose (R_i^+, S_i^+) are already defined. Step i of the process is defined as follows. If $S_i^+ = (s_{i,1}, \ldots, s_{i,j})$ has positive length then write $u_i^+ = u_i^+(D, v) = s_{i,1}$ and $C_i^+(D, v) = N^+(D, u_i^+) \setminus (R_i^+ \cup S_i^+)$. List the elements of $C_i^+(D, v)$ in increasing order of index as $w_{i,1}, \ldots, w_{i,k}$; it is possible that $k = 0$. Then set

$$R_{i+1}^+ = R_i^+ \cup \{s_{i,1}\}, \quad \text{and} \quad S_{i+1}^+ = (s_{i,2}, \ldots, s_{i,j}, w_{i,1}, \ldots, w_{i,k}).$$

In words, at step i, $u_i^+ = s_{i,1}$ is explored, and $w_{i,1}, \ldots, w_{i,k}$ are discovered and added to the back of the queue for later exploration. If S_i^+ has zero length (i.e., $S_i^+ = ()$), then $S_{i+1}^+ = S_i^+$ and $R_{i+1}^+ = R_i^+$.

Writing $i^+ = i^+(D, v) = \min\{i : S_{i+1}^+ = S_i^+\}$, then $R_{i^+}^+(D, v)$ is precisely the set of vertices w with $\mathrm{dist}_D(v, w) < \infty$. The *oBFS tree* $T^+(D, v)$ has root v and vertices $R_{i^+}(D, v)$; the children of $s_{i,1}$ are precisely the vertices $w_{i,1}, \ldots, w_{i,k}$ newly discovered in step i. We write $T^+(D, v, m)$ for the subtree of $T^+(D, v)$ with vertices $R_m^+(D, v) \cup S_m^+(D, v)$. Note that if $S_m^+(D, v)$ has length ℓ then its elements are precisely $(u_{m+i}^+(D, v), 0 \leq i < \ell)$, because oBFS explores these vertices before any others. We therefore have $R_m^+(D, v) \cup S_m^+(D, v) = \{u_i^+(D, v), 0 \leq i < m + \ell\}$,

In the *inward breadth-first search* (iBFS) process $((R_i^-(D, v), S_i^-(D, v)), i \geq 0)$, the sets $C_i^-(D, v)$ and the terminal time $i^- = i^-(D, v)$ are defined in just the same manner but exploring in-neighbourhoods rather than out-neighbourhoods to discover vertices; in particular $R_{i^-}^-(D, v) = \{w : \mathrm{dist}_D(w, v) < \infty\}$. We also write $T^-(D, v, m)$ for the subtree of $T^-(D, v)$ with vertices $R_m^-(D, v) \cup S_m^-(D, v)$.

Observe that using the notation from Section 3.3, we have $T_k^+(D, v) = N_k^+(D, v)$ and $T_k^-(D, v) = N_k^-(D, v)$ for all k.

5.2. **Conditioning on neighbourhoods and the BFS exploration in** $D(n, r)$**.** We next describe the effect of iBFS on the law of $D(n, r)$. For the remainder of the section we write $D = D(n, r)$ and fix $v \in [n]$. We write $N_i^- = N_i^-(D, v)$, $u_i^- = u_i^-(D, v)$, $T^-(m) = T^-(D, v, m)$, etcetera. Informally, the point of this section may be summarized as follows: if vertex u_i^- is discovered at step j then all we know about u_i^- is that it has an edge to u_j^- and has no edges to u_k for $k < j$.

We mostly focus on the iBFS process. First, in Lemma 5.1 we control the number of edges from a vertex to the set of vertices already discovered by the process. We use these results to bound the in-degree of an undiscovered vertex to an in-neighbhourhood (Corollary 5.2), to control the one-step growth of in-neighbhourhoods (Corollary 5.3), and to bound the out-degree of a discovered vertex to the undiscovered set (Corollary 5.4). In Lemma 5.5, we show that the iBFS process is unlikely to hit the vertices of an already small exposed part of the digraph. Finally we state Lemma 5.6 which is the analogue of Lemma 5.1 for the oBFS process.

We now state and prove some useful stochastic identities and inequalities which result from this.

Lemma 5.1. *Fix $m \in \mathbb{N}_0$. Conditional on $((R_i^-, S_i^-), 0 \le i \le m)$, independently for all $w \in [n]$ we have*

$$|E(w, R_m^- \cup S_m^-)| \overset{\mathrm{d}}{=} \mathbf{Bin}\left(r, \frac{|S_m^-|}{n-m}\right)$$

if $w \notin R_m^- \cup S_m^-$,

$$1 + \mathbf{Bin}\left(r-1, \frac{|S_m^-|}{n-m}\right) \preceq |E(w, R_m^- \cup S_m^-)| \preceq 1 + \mathbf{Bin}\left(r-1, \frac{m+|S_m^-|}{n}\right)$$

if $w \in R_m^- \cup S_m^- \setminus \{v\}$, and $|E(v, R_m^- \cup S_m^-)| \overset{\mathrm{d}}{=} \mathbf{Bin}\left(r, \frac{m+|S_m^-|}{n}\right)$.

Proof. Recall the canonical construction of $D = D(n,r)$ from the introduction, and for each $x, y \in [n]$ let $\rho(x,y) = \min\{q : L_{x,q} = y\}$. Then $\rho(x,y) \le r$ precisely if there is a copy of the oriented edge xy in $D(n,r)$, and otherwise $\rho(x,y) = \infty$.

Now fix $w \in [n]$. Suppose $w \notin R_m^- \cup S_m^-$; then there are no edges from w to R_m^-. In other words, for each $1 \le j \le r$, we have $L_{w,j} \notin R_m^- = \{u_i^-, 0 \le i < m\}$. It follows that the conditional law of $|E(w, R_m^- \cup S_m^-)|$ given $(R_i^-, 0 \le i \le m)$ and $(S_i^-, 0 \le i \le m)$ is $\mathbf{Bin}(r, |S_m^-|/(n - |R_m^-|))$. The result follows in this case since $|R_m^-| = m$.

Now suppose $w \in R_m^- \cup S_m^- \setminus \{v\}$; then the parent $p_{T^-(m)}(w)$ lies in R_m^- so satisfies $p_{T^-(m)}(w) = u_j^-$ for some $0 \le j < m$. We have $p_{T^-(m)}(w) = u_j^-$ precisely if w has no edges to $\{u_i^- : 0 \le i < j\}$ but has an edge to u_j^-; equivalently, $\rho(w, u_i^-) = \infty$ for each $0 \le i < j$, and $\rho(w, u_j^-) = k$ for some $1 \le k \le r$. The heads of the first $(k-1)$ out-edges from w are then uniformly distributed over $[n] \setminus \{u_i^- : 0 \le i < j\}$, and the heads of the $r - k$ last out-edges from w are uniformly distributed over $[n] \setminus \{u_i^- : 0 \le i \le j\}$.

The index j is determined by $((R_i^-, S_i^-), 0 \le i \le m)$. Given that $w \in R_m^- \cup S_m^- \setminus \{v\}$, we thus have

$$1 + \mathbf{Bin}\left(r-1, \frac{|R_m^- \cup S_m^-| - (j+1)}{n - (j+1)}\right) \preceq |E(w, R_m^- \cup S_m^-)| \preceq 1 + \mathbf{Bin}\left(r-1, \frac{|R_m^- \cup S_m^-| - j}{n - j}\right).$$

Since $0 \le j < m$, and $|R_m^-| = m$, the second claim follows. The argument when $w = v$ is similar but easier. Finally, the independence asserted by the lemma follows from the independence of the random variables $(L_{w,p} : (w,p) \in [n] \times [r])$. $\qquad \square$

For the next corollary, recall that $d_{\le j}^- = |N_{\le j}^-| = |N_{\le j}^-(D,v)|$.

Corollary 5.2. *Fix $j \in \mathbb{N}_0$. Conditional on $(N_i^-, 0 \le i \le j)$, independently for all $w \in [n]$ we have $|E(w, N_{\le j}^-)| \overset{\mathrm{d}}{=} \mathbf{Bin}(r, d_{\le j}^-/n)$ if $w = v$, $|E(w, N_{\le j}^-)| \overset{\mathrm{d}}{=} \mathbf{Bin}(r, d_j^-/(n - d_{\le j-1}^-))$ if $w \notin N_{\le j}^-$, and*

$$1 + \mathbf{Bin}\left(r-1, \frac{d_j^-}{n - d_{\le j-1}^-}\right) \preceq |E(w, N_{\le j}^-)| \preceq 1 + \mathbf{Bin}\left(r-1, \frac{d_{\le j}^-}{n}\right)$$

if $w \in N_{\le j}^- \setminus \{v\}$.

Proof. Apply Lemma 5.1 at time $m = d_{\le j-1}^-$. $\qquad \square$

Corollary 5.3. *For all $j, q, p \in \mathbb{N}_0$, given that $d_j^- = q$ and $d_{\le j}^- = p$, $d_{j+1}^- \overset{\mathrm{d}}{=} \mathbf{Bin}(n - p, 1 - (1 - q/(n-p))^r)$ and $d_{j+1}^- \preceq \mathbf{Bin}\left(r(n-p), \frac{q}{n-p+q}\right)$.*

Proof. If $w \in [n] \backslash N^-_{\leq j}$ then $E(w, N^-_{\leq j}) = E(w, N^-_j)$. By Corollary 5.2, the number of edges from w to N^-_j then has conditional law $\mathbf{Bin}(r, q/(n-p))$, so is non-zero with probability $1 - (1 - q/(n-p))^r$. The distributional identity follows since $|[n] \backslash N^-_{\leq j}| = n - d^-_{\leq j} = n - p$.

Next, note that $\mathbf{Bin}(m, 1 - (1-x)^r)$ is stochastically dominated by $\mathbf{Bin}(rm, x)$. To see this, observe that the former is the number of columns containing at least one 1 in an $r \times m$ matrix whose entries are iid $\mathbf{Ber}(x)$ random variables, while the latter is the law of the number of ones in such a matrix. The pigeonhole principle then yields the second claim of the lemma. $\qquad\square$

Recall that $C^-_i(D, v) = N^-(D, u^-_i) \backslash (R^-_i \cup S^-_i)$ is the set of vertices discovered at step i of iBFS.

Corollary 5.4. *Fix $i, q \in \{0, 1, \ldots, n\}$. Conditioned on $|S^-_i(D, v)| = q$, we have*

$$|C^-_i(D, v)| \stackrel{\mathrm{d}}{=} \mathbf{Bin}\left(n - i - q, 1 - \left(1 - \frac{1}{n-i}\right)^r\right).$$

We omit the proof since it is very similar to those given above.

The next lemma formalizes the intuitively clear picture that it is unlikely for the early stages of iBFS to encounter a fixed, small subgraph of D not containing the starting vertex.

Lemma 5.5. *Fix a digraph G with $V(G) \subset [n]$ and $v \notin V(G)$. Fix $s \in \mathbb{N}$ and let $i_0 = \inf\{i : |R^-_i \cup S^-_i| > s\}$. Then*

$$\mathbb{P}((R^-_{i_0-1} \cup S^-_{i_0-1}) \cap V(G) \neq \emptyset \mid D[V(G)] = G) \leq \frac{r(s+1)|V(G)|}{n - |V(G)| - s}.$$

Proof. Since $|R^-_i| = i$, we clearly have $i_0 \leq s + 1$. Writing $\tau = \inf\{t : (R^-_i \cup S^-_i) \cap V(G) \neq \emptyset\}$, the probability we aim to bound is thus at most

$$\mathbb{P}(\tau < i_0 \mid D[V(G)] = G) = \sum_{i=0}^{s} \mathbb{P}(\tau = i, i_0 > i \mid D[V(G)] = G)$$

$$\leq \sum_{i=0}^{s} \mathbb{P}(\tau = i \mid i_0 > i, \tau \geq i, D[V(G)] = G)$$

Given that $D[V(G)] = G$, there are $r|V(G)| - |E(G)|$ edges from $V(G)$ to $[n] \setminus V(G)$; the heads of such edges are uniformly distributed over $[n] \setminus V(G)$. For $i > 0$, given that $(R^-_{i-1} \cup S^-_{i-1}) \cap V(G) = \emptyset$, the heads of these edges are uniformly distributed over $[n] \setminus (V(G) \cup R^-_{i-1} \cup S^-_{i-1})$. Thus,

$$\mathbb{P}((R^-_i \cup S^-_i) \cap V(G) \neq \emptyset \mid R^-_{i-1}, S^-_{i-1}, D[V(G)] = G, (R^-_{i-1} \cup S^-_{i-1}) \cap V(G) = \emptyset)$$
$$\leq \frac{r|V(G)|}{n - |V(G)| - |R^-_{i-1} \cup S^-_{i-1}|},$$

since, in this situation, the only way that $(R^-_i \cup S^-_i) \cap V(G) \neq \emptyset$ is in the case that some edge with tail in $V(G)$ has as a head the vertex u^-_{i-1}.

Given that $i_0 > i$ and $\tau \geq i$ we indeed have $(R^-_{i-1} \cup S^-_{i-1}) \cap V(G) = \emptyset$, and also have $|R^-_i \cup S^-_i| \leq s$, so $\mathbb{P}(\tau = i \mid i_0 > i, \tau \geq i, D[V(G)] = G) \leq r|V(G)|/(n - |V(G)| - s)$. Using this bound in the above sum, the result follows. $\qquad\square$

Finally, we require the following, rather simple result for oBFS.

Lemma 5.6. *Fix $m \in \mathbb{N}_0$. Conditional on R^+_m and S^+_m, independently for all $w \in [n] \setminus R^+_m$ we have*

$$|E(w, R^+_m \cup S^+_m)| \preceq \mathbf{Bin}(r, (|R^+_m| + |S^+_m|)/n) \preceq \mathbf{Bin}(r, (rm+1)/n).$$

Proof. We omit the proof of the first inequality, which parallels that of Lemma 5.1. For the second, note that $|R_m^+ \cup S_m^+| \leq rm + 1$ since D is r-out regular. □

6. In-neighbourhoods: technical lemmas

In this section we gather a few basic estimates that describe the size and structure of in-neighbourhoods of vertices in $D = D(n, r)$.

Proposition 6.1 shows that in-neighbourhoods of $D(n, r)$ cannot be too large. Lemma 6.2 controls the probability that the sequence $(d_k^-(v), k \geq 1)$ exhibits a large decrease in value for relatively small values of k. In the last part of the section, we focus on the probability that an in-neighbourhood neither exponentially expands nor dies out quickly. Lemma 6.3, shows that local in-neighbourhoods of a vertex are well-approximated by Poisson Galton-Watson trees. We use this lemma with a result of Riordan and Wormald (Lemma 6.4) to control the probability that long and thin in-neighbourhoods exist in Proposition 6.5.

Proposition 6.1. *For all $\alpha > 0$,*

$$\mathbb{P}\left(\exists v \in [n], \exists j \geq 0 \text{ such that } d_j^-(v) \geq (r+\alpha)^j \log_r^2 n\right) = O(n^{-4}).$$

Proof. Fix $v \in [n]$. We prove that

$$\mathbb{P}\left(\exists j \geq 0 \text{ such that } d_j^-(v) \geq (r+\alpha)^j \log_r^2 n\right) = O(n^{-5}); \tag{6.1}$$

a union bound over $v \in [n]$ then proves the proposition.

For $j \geq 0$, let $E_j = \{d_j^-(v) < (r+\alpha)^j \log_r^2 n\}$. Then

$$\mathbb{P}\left(\exists j \geq 0 \text{ such that } d_j^-(v) \geq (r+\alpha)^j \log_r^2 n\right) = \mathbb{P}\left(\cup_{j=1}^n \overline{E_j}\right) \leq \sum_{j=1}^n \mathbb{P}(\overline{E_j} \mid \cap_{j'<j} E_{j'}). \tag{6.2}$$

By Corollary 5.3, for every $p \geq q$ and every $a \geq 0$,

$$\mathbb{P}(d_j^-(v) \geq a \mid d_{j-1}^-(v) = q, d_{\leq j-1}^-(v) = p) \leq \mathbb{P}\left(\mathbf{Bin}\left(r(n-p), \frac{q}{n-p+q}\right) \geq a\right).$$

Note that $r(n-p) \cdot \frac{q}{n-p+q} \leq rq$. Set $a = (r+\alpha)^j \log_r^2 n$, and observe that if E_{j-1} occurs then $d_{j-1}^-(v) = q \leq q_0 := (r+\alpha)^{j-1} \log_r^2 n$. Finally, for such q we have $a = (r+\alpha)q_0 \geq (r+\alpha)q$, so

$$\begin{aligned}
\mathbb{P}(\overline{E_j} \mid \cap_{j'<j} E_{j'}) &= \mathbb{P}(d_j^-(v) \geq a \mid \cap_{j'<j} E_{j'}) \\
&\leq \sup_{q \leq q_0, p} \mathbb{P}(d_j^-(v) \geq a \mid d_{j-1}^-(v) = q, d_{\leq j-1}^-(v) = p) \\
&\leq \sup_{q \leq q_0, p} \mathbb{P}(d_j^-(v) \geq rq + \alpha q \mid d_{j-1}^-(v) = q, d_{\leq j-1}^-(v) = p) \\
&\leq \sup_{q \leq q_0} e^{-\frac{\alpha^2 q^2}{2(r+\alpha/3)q}} = e^{-\Omega(\log_r^2 n)} = O(n^{-6}),
\end{aligned}$$

where we used the Chernoff bound (3.1). Using this bound in (6.2) proves (6.1). □

Lemma 6.2. *Fix $v \in [n]$. Uniformly in $k \leq 0.99 \log_r n$ and $\omega \geq \log_r^3 n$, we have*

$$\mathbb{P}(d_{\leq k}^-(v) \geq \omega^2, d_k^-(v) \leq \omega) = O(n^{-3}).$$

Proof. Fix k and ω as above, and $v \in [n]$. Let $\tau = \min\{j : \ d_j^-(v) \geq \omega^2/k\}$. If $d_{\leq k}^-(v) \geq \omega^2$ and $d_k^-(v) \leq \omega$, then $\tau < k$, so

$$
\begin{aligned}
\mathbb{P}(d_{\leq k}^-(v) \geq \omega^2, d_k^-(v) \leq \omega) &= \mathbb{P}(d_{\leq k}^-(v) \geq \omega^2, d_k^-(v) \leq \omega, \tau < k) \\
&= \sum_{j=1}^{k-1} \mathbb{P}(d_{\leq k}^-(v) \geq \omega^2, d_k^-(v) \leq \omega, \tau = j) \\
&\leq \sum_{j=1}^{k-1} \mathbb{P}(d_j^-(v) \geq \omega^2/k, d_k^-(v) \leq \omega) .
\end{aligned} \tag{6.3}
$$

Let $\sigma = \min\{i \geq \tau : \ d_{i+1}^-(v) \leq d_i^-(v)\}$. For any $j < k$, if $d_j^-(v) \geq \omega^2/k$ and $d_k^-(v) \leq \omega$, then $\sigma < k$, so

$$
\mathbb{P}(d_j^-(v) \geq \omega^2/k, d_k^-(v) \leq \omega) \leq \mathbb{P}(d_j^-(v) \geq \omega^2/k, \ \sigma < k) . \tag{6.4}
$$

Now fix $\alpha > 0$ small enough that $(r+\alpha)^k \log_r^2 n < n^{0.999}$ for n large; this is possible by our choice of k. Also, for $\ell \in \mathbb{N}$ let $E_\ell = \{\forall i \leq \ell, d_i^-(v) < (r+\alpha)^i \log_r^2 n\}$, and let $E = \bigcap_{\ell \geq 1} E_\ell$. By Proposition 6.1, we have $\mathbb{P}(\overline{E}) = O(n^{-4})$, so for all $1 \leq j \leq k - 1$,

$$
\begin{aligned}
\mathbb{P}(d_j^-(v) \geq \omega^2/k, \ \sigma < k) &\leq \mathbb{P}(\overline{E}) + \mathbb{P}(d_j^-(v) \geq \omega^2/k, \ \sigma < k, \ E) \\
&\leq O(n^{-4}) + \sum_{\ell=j}^{k-1} \mathbb{P}(d_j^-(v) \geq \omega^2/k, \ \sigma = \ell, \ E) \\
&\leq O(n^{-4}) + \sum_{\ell=j}^{k-1} \mathbb{P}(d_\ell^-(v) \geq \omega^2/k, \ d_{\ell+1}^-(v) \leq d_\ell^-(v), \ E_\ell) \\
&\leq O(n^{-4}) + \sum_{\ell=j}^{k-1} \mathbb{P}(d_{\ell+1}^-(v) \leq d_\ell^-(v) \mid d_\ell^-(v) \geq \omega^2/k, \ E_\ell) . \tag{6.5}
\end{aligned}
$$

On E_ℓ we have $d_{\leq \ell}^-(v) < (r+\alpha)^{\ell+1} \log_r^2 n \leq (r+\alpha)^k \log_r^2 n < n^{0.999}$.

Now fix $0 < q \leq p \leq n^{0.999}$ and let X be distributed as $\mathbf{Bin}\left(n - p, 1 - \left(1 - \frac{q}{n-p+q}\right)^r\right)$. Using that $(1-x)^i \leq 1 - ix + \binom{i}{2}x^2$ for $x \in (0, 1)$ and $i \in \mathbb{N}$, for we have

$$
\begin{aligned}
\mathbb{E}(X) &\geq (n - p)\left(\frac{rq}{n-p+q} - \frac{\binom{r}{2}q^2}{(n-p+q)^2}\right) = rq \cdot \frac{(n-p)(n - p - \frac{r-3}{2}q)}{(n-p+q)^2} \\
&> rq \cdot \left(1 - \frac{3rp}{n}\right) > \frac{2rq}{3},
\end{aligned}
$$

the last inequality for n large since $p \leq n^{0.999}$. Now write $q_0 = \omega^2/k$ and $p_0 = (r + \alpha)^{\ell+1} \log_r n < n^{0.999}$. By Corollary 5.3 and the Chernoff bound (3.2), we have

$$
\begin{aligned}
&\mathbb{P}(d_{\ell+1}^-(v) \leq d_\ell^-(v) \mid d_\ell^-(v) \geq \omega^2/k, \ E_\ell) \\
&\leq \sup_{q_0 \leq q \leq p \leq p_0} \mathbb{P}(d_{\ell+1}^-(v) \leq d_\ell^-(v) \mid d_\ell^-(v) = q, d_{\leq\ell}^-(v) = p) \\
&\leq \sup_{q_0 \leq q \leq p \leq p_0} \mathbb{P}(d_{\ell+1}^-(v) \leq q \mid d_\ell^-(v) = q, d_{\leq\ell}^-(v) = p) \\
&\leq \sup_{q_0 \leq q \leq p \leq p_0} \mathbb{P}(X \leq q) \\
&\leq \sup_{q_0 \leq q \leq p \leq p_0} \mathbb{P}(X \leq \mathbb{E}(X) - (2r/3 - 1)q) \\
&\leq \sup_{q_0 \leq q \leq p \leq p_0} e^{-\frac{(2r/3-1)^2 q^2}{2(rq + (2r/3-1)q/3)}} \leq e^{-\frac{q_0}{18r+2}} \,,
\end{aligned}
$$

in the last line using that $r \geq 2$. Finally, $q_0 = \omega^2/k \geq \log_r^2 n$, so $e^{-q_0/(18r+2)} = O(n^{-4})$. Combining the preceding inequality with (6.3), (6.4) and (6.5) yields

$$
\begin{aligned}
\mathbb{P}(d_{\leq k}^-(v) \geq \omega^2, d_k^-(v) \leq \omega) &\leq \sum_{j=1}^{k-1} \sum_{\ell=j}^{k-1} \mathbb{P}(d_{\ell+1}^-(v) \leq d_\ell^-(v) \mid d_\ell^-(v) \geq \omega^2/k, \ E) + O(kn^{-4}) \\
&\leq O(k^2 n^{-4}) = O(n^{-3}) \,. \qquad \square
\end{aligned}
$$

The next lemma compares the law of $T^-(v)$ to that of a Poisson(r) Galton-Watson tree. A similar result, in the setting of undirected graphs, can be found in [33, Lemma 2.2].

Lemma 6.3. *Let $r \geq 2$ and let T' be a v-rooted plane directed tree where all edges point to the root. Suppose that $|V(T')| \leq n/2$. Then, for any $k \geq 0$ we have*

$$
\mathbb{P}(T_{\leq k}^-(D, v) \cong T') = e^{O(|V(T')|^2/n)} \mathbb{P}(\mathcal{T}_{\leq k} \cong T') \,.
$$

where $\mathcal{T}$ is a Galton-Watson branching tree whose offspring is Poisson with parameter r.

Proof. Fix $k \in \mathbb{N}_0$ and a plane tree T' of height at most k, and write $t = |V(T')|$. Recall the canonical labelling of $V(T')$ with labels from $\emptyset \cup \bigcup_{i \geq 1} \mathbb{N}^i$ introduced in Section 3.3. Consider the iBFS procedure on T' started at its root v. To make sense of this, we must specify the order in which the children of a vertex u are added to the set of discovered vertices. We use the left-to-right order: so if u is explored at step i (i.e. $u_i^-(T', v) = u$) then the rightmost child of u is the last element of $S_i^-(T', v)$.

Let $a_i = |C_i^-(T', v)|$ be the number of children of $u_i^-(T', v)$, and let $s = |V(T'_{\leq k-1})|$ be the number of vertices of T' at distance at most $k - 1$ from the root. In order to check if $T_{\leq k}(D, v)$ and T' are isomorphic, it suffices to perform s steps of the iBFS exploration from v in D. We then have

$$
\begin{aligned}
\mathbb{P}(T_{\leq k}^-(D, v) \cong T') &= \mathbb{P}\left(|C_i^-(D, v)| = a_i, 0 \leq i < s\right) \\
&= \prod_{i=0}^{s-1} \mathbb{P}\left(|C_i^-(D, v)| = a_i \mid |C_j^-(D, v)| = a_j, 0 \leq j < i\right) \\
&= \prod_{i=0}^{s-1} \mathbb{P}\left(|C_i^-(D, v)| = a_i \mid |S_i^-(D, v)| = 1 + \sum_{j=0}^{i-1}(a_j - 1)\right) \,,
\end{aligned}
$$

where the last line is due to the symmetry of the model.

Writing $q_i = 1 + \sum_{j=0}^{i-1}(a_j - 1)$, by Corollary 5.4 we then have

$$\mathbb{P}(|C_i^-(D,v)| = a_i \mid |S_i^-(D,v)| = q_i)$$

$$= \binom{n-i-q_i}{a_i}\left(1 - \left(1 - \frac{1}{n-i}\right)^r\right)^{a_i}\left(1 - \frac{1}{n-i}\right)^{r(n-q_i-a_i-i)}$$

$$= e^{O\left(\frac{a_i(a_i+r)}{n-i-q_i}\right)}\frac{(n-i-q_i)^{a_i}}{a_i!}\left(\frac{r}{n-i}\right)^{a_i}\left(1 - \frac{1}{n-i}\right)^{r(n-q_i-a_i-i)}.$$

Now let $\mathcal{T}$ be a Poisson(r) Galton-Watson tree; write ρ for the root of $\mathcal{T}$. Build $\mathcal{T}$ via iBFS starting from ρ. In this manner, we may couple $\mathcal{T}$ with a sequence $(\xi_i, i \geq 0)$ of iid $\mathbf{Po}(r)$ random variables so that for $0 \leq i < |V(\mathcal{T})|$ we have $|C_i^-(\mathcal{T},\rho)| = \xi_i$. It follows that

$$\mathbb{P}(\mathcal{T}_{\leq k} \cong T') = \mathbb{P}\left(\cap_{i=0}^{s-1}\xi_i = a_i\right) = \prod_{i=0}^{s-1}\mathbb{P}(\xi_i = a_i) = \prod_{i=0}^{s-1}e^{-r}\frac{r^{a_i}}{a_i!}.$$

Using that $1 + x \leq e^x$, this gives

$$\frac{\mathbb{P}(|C_i^-(D,v)| = a_i \mid |S_i^-(D,v)| = q_i)}{\mathbb{P}(\xi_i = a_i)} = e^{O\left(\frac{a_i(a_i+r)}{n-i-q_i}\right)}\frac{(n-i-q_i)^{a_i}}{(n-i)^{a_i}} \cdot \frac{\left(1 - \frac{1}{n-i}\right)^{r(n-q_i-a_i-i)}}{e^{-r}}$$

$$= e^{O\left(\frac{q_i+a_i^2}{n-i-q_i}\right)}.$$

Since $i + q_i \leq t \leq n/2$, $\sum_{i=0}^{s-1}a_i^2 \leq t^2$, and $s \leq t$, we have

$$\sum_{i=0}^{s-1}\frac{q_i + a_i^2}{n-i-q_i} \leq \sum_{i=0}^{s-1}\frac{t + a_i^2}{n/2} = O\left(\frac{t^2}{n}\right).$$

It follows that

$$\frac{\mathbb{P}(T_{\leq k}^-(D,v) \cong T')}{\mathbb{P}(\mathcal{T}_{\leq k} \cong T')} = \prod_{i=0}^{s-1}\frac{\mathbb{P}(|C_i^-(D,v)| = a_i \mid |S_i^-(D,v)| = q_i)}{\mathbb{P}(\xi_i = a_i)} = e^{O(t^2/n)}. \qquad \square$$

Lemma 6.4 ([33], Lemma 2.1). *Let $\mathcal{T}$ be a Poisson(r) Galton-Watson tree. There exist constants $c, C > 0$ such that for every $\omega \geq 2$ and $k \geq 1$ we have*

$$c \cdot \min\{(r(1-\lambda_r))^{k-k'}, 1\} \leq \mathbb{P}(0 < |\mathcal{T}_k| < \omega) \leq C(r(1-\lambda_r))^{k-k'},$$

where $k' = \lfloor \log_r \omega \rfloor$.

Recall that the probability of survival in $\mathcal{T}$ is $\mathbb{P}\left(\sum_{k \geq 0}|\mathcal{T}_k| = \infty\right) \in (0,1)$. Essentially, the preceding lemma states that given the branching process survives for the first k generations, the probability that $|\mathcal{T}_k| < \omega$ decays exponentially in k (provided that ω is small enough with respect to k). The final and principal result of this section is to prove a corresponding bound with $d_k^-(v)$ in place of $|\mathcal{T}_k|$.

Proposition 6.5. *For every $v \in [n]$, $k \leq 0.99 \log_r n$ and $\log_r^3 n \leq \omega \leq n^{1/6}$*

$$\mathbb{P}(0 < d_k^-(v) < \omega) = (1+o(1))\mathbb{P}(0 < |\mathcal{T}_k| < \omega) + O(n^{-3}) = \Theta((r(1-\lambda_r))^{k-k'}) + O(n^{-3}).$$

where $k' = \lfloor \log_r \omega \rfloor$.

Proof. Write $Z_k = |\mathcal{T}_k|$. We first prove an upper bound on $\mathbb{P}(0 < d_k^-(v) < \omega)$. By Lemma 6.3,

$$\mathbb{P}(0 < d_k^-(v) < \omega) = \sum_{\{T':|T'_k|\in(0,\omega)\}} \mathbb{P}\left(T_{\leq k}(D,v) \cong T'\right)$$

$$\leq \sum_{\substack{\{T':|T'_k|\in(0,\omega) \\ |T'|\leq n^{1/3}\}}} e^{O\left(\frac{|V(T')|^2}{n}\right)} \mathbb{P}(\mathcal{T}_{\leq k} \cong T') + \mathbb{P}(d_{\leq k}^-(v) \geq \omega^2, d_k^-(v) \in (0,\omega))$$

$$\leq (1+o(1))\mathbb{P}(0 < Z_k < \omega) + O(n^{-3}),$$

where in the last inequality we used Lemma 6.2.

We now turn to the lower bound. A similar argument to that above gives

$$\mathbb{P}(0 < d_k^-(v) < \omega) \geq (1+o(1))\mathbb{P}(0 < Z_k < \omega) - \mathbb{P}(\sum_{j=0}^k Z_j \geq \omega^2, Z_k < \omega).$$

Bounding the second probability is straightforward. First, fix $j < k$ and $a, b \in \mathbb{N}$. Given that $Z_j = a$, by the branching property, each of the a subtrees of $\mathcal{T}$ rooted at a node in $\mathcal{T}_j$ survives independently with probability $p := \mathbb{P}(|\mathcal{T}| = \infty)$. Note that $p > 0$ is independent of n. But Z_k is at least the number of such subtrees which survive, so $\mathbb{P}(Z_k < b \mid Z_j = a) \leq \mathbb{P}(\mathbf{Bin}(a,p) < b)$.

Finally, if $\sum_{j=0}^k Z_j \geq \omega^2$ then $\max_{0\leq j\leq k} Z_j \geq \omega^2/(k+1)$. In other words, letting $j_0 = \inf\{j : Z_j \geq \omega^2/(k+1)\}$, we must have $j_0 \leq k$. It follows from the preceding paragraph (conditioning on the value of $j_0 \leq k$) that

$$\mathbb{P}(\sum_{j=0}^k Z_j \geq \omega^2, Z_k < \omega) \leq \mathbb{P}(\mathbf{Bin}(\omega^2/(k+1),p) < \omega) = O(n^{-3}),$$

the final inequality by a Chernoff bound since $\omega^2/(k+1) \geq \omega \log^2 n$. The proposition follows from Lemma 6.4. $\square$

7. OUT-NEIGHBOURHOODS: TECHNICAL LEMMAS

Recall that E_{SL} is the event that D contains no loop vertices. As in the statement of Proposition 4.2 we now fix $k \in [n]$, let $E_k = \{d_k^-(v) \geq \log^4 n, d_{\leq k}^-(v) \leq \log^7 n\}$, and fix a graph H with $v \in V(H) \subset [n]$ such that $\mathbb{P}(D[N_{\leq k}^-(v)] = H, E_k, \overline{E_{\mathrm{SL}}}) > 0$. It is useful to write $B_H = N_k^-(H,v)$; note that this is a deterministic set since H is deterministic, and on $D[N_{\leq k}^-] = H$ we have $N_k^-(v) = B_H$. Let $\hat{n} = n - |V(H)| + |B_H|$. By the assumptions on H we have $\hat{n} \geq n - \log^7 n + \log^4 n$.

For any event A, write $\mathbb{P}^H(A) = \mathbb{P}(A \mid D[N_{\leq k}^-] = H)$. The following fact describes the distribution of D under $\mathbb{P}^H$. Its proof follows from straightforward considerations and is omitted.

Fact 7.1. *Given that $D[N_{\leq k}^-] = H$, the conditional distribution of $D(n,r)$ is that of the graph $\hat{D}$ defined as follows. First, $\hat{D}[V(H)] = H$.*

Next, independently for each $w \notin V(H)$, let $\hat{L}_w = (\hat{L}_{w,1},\ldots,\hat{L}_{w,r})$ be a vector chosen uniformly at random from the $\hat{n}^r$ vectors $(s_1,\ldots,s_r) \in (([n] \setminus V(H)) \cup B_H)^r$. Then for each $i \in [r]$ add a directed edge from w to $\hat{L}_{w,i}$.

Finally, for $w \in V(H)$, let $t_w = r - |E_H(w,H)|$; this is the number of edges with tail w and head not in H. Independently for each $w \in V(H)$, let $\hat{L}_w = (\hat{L}_{w,1},\ldots,\hat{L}_{w,t_w})$ be a vector chosen uniformly at random from $([n] \setminus V(H))^{t_w}$, and for each $i \in [t_w]$ add a directed edge from w to $\hat{L}_{w,i}$.

For the remainder of the section, fix $u \in [n]$ and write $N_j^* = N_j^+(u) \setminus V(H)$, $d_j^* = |N_j^*|$. We continue with a simple lemma.

Lemma 7.2. *We have* $\mathbb{P}^H(d_5^* < 5, N_{\leq 5}^+(u) \cap V(H) = \emptyset, \overline{E_{SL}}) = O(n^{-3})$

Proof. If $N_{\leq 5}^+(u) \cap V(H) = \emptyset$ then $d_5^* = d_5^+$. In this case, since $r \geq 2$, it is a simple combinatorial exercise to check that if also $d_5^+ < 5$ then $D[N_{\leq 5}^*]$ has at least two more edges than vertices. For any fixed digraph $\widehat{D}$ with at least two more edges than vertices and with no self loops, it is easily seen that $\mathbb{P}^H(D[N_{\leq 5}^+] \cong \hat{D}, N_{\leq 5}^+(u) \cap V(H) = \emptyset) = O(n^{-3})$. (This is not true for digraphs with self-loops if $r = 2$; the probability v itself is a loop vertex is $O(n^{-2})$ and in this case $d_5^* = 0$.) The number of isomorphism classes of digraphs with diameter at most 5 and maximum out-degree r is bounded, and the result follows. $\square$

We next show that with high probability, each generation N_j^* is approximately r times larger than the last, until j is nearly $(\log^r n)/2$.

Lemma 7.3. *Let* $\sigma = \inf\{i \geq 5 : d_i^* < r^{i-5} + 5\}$. *Then* $\mathbb{P}^H(5 < \sigma \leq (\log_r n)/8) = O(n^{-3})$.

Proof. Fix $j > 5$. If $d_5^* \geq 5$ and $d_{i+1}^* \geq rd_i^* - 4$ for every $5 \leq i < j$, then by induction $d_j^* \geq r^{j-5} + 5$. We thus have

$$\mathbb{P}^H(5 < \sigma \leq j) \leq \mathbb{P}^H\left(d_5^* \geq 5, \bigcup_{i=5}^{j}\{|d_i^*| < r^{i-5} + 5\}\right) \leq \sum_{i=5}^{j-1} \mathbb{P}(d_{i+1}^* \leq rd_i^* - 4).$$

Now fix $5 \leq i < j$. Condition on $N_{\leq i}^*$, and recall that the random variables $\{L_{w,m} : w \in N_i^*, m \in [r]\}$ are the heads of edges from vertices in N_i^*. Reveal the values of these random variables one-at-a-time; say a conflict occurs if $L_{w,m} \in N_{\leq i}^* \cup V(H)$ or $L_{w,m} = L_{w',m'}$ for a previously revealed $L_{w',m'}$. If $d_{i+1}^* \leq rd_i^* - 4$ then at least 4 conflicts occur.

Under $\mathbb{P}^H$, the random variables $L_{w,m}$ are independent and uniform over $([n] \setminus V(H)) \cup B_H$. When $L_{w,m}$ is revealed there are less than $r^{i+2} + |B_H|$ locations that can cause a conflict, since $|N_{\leq i+1}^*| < r^{i+2}$, so the probability of a conflict is less than $(r^{i+2} + |B_H|)/\hat{n}$. The set $\{L_{w,m} : w \in N_i^*, m \in [r]\}$ has size at most r^{i+1}, and $|B_H| \leq \log^7 n$; it follows that

$$\mathbb{P}^H(d_{i+1}^* \leq rd_i^* - 4) \leq \mathbb{P}\left(\mathbf{Bin}(r^{i+1}, (r^{i+2} + |B_H|)/\hat{n}) \geq 4\right)$$

$$\leq \binom{r^{i+1}}{4} \cdot \left(\frac{r^{i+2} + \log^7 n}{\hat{n}}\right)^4 \leq \frac{C(r^{8i} + \log^{28} n)}{n^4},$$

where in the last inequality we used that $\hat{n} \geq n - \log^7 n$. For $j \leq (\log_r n)/8$ we thus have

$$\mathbb{P}^H\left(d_5^* \geq 5, \bigcup_{i=5}^{j}\{|d_i^*| < r^{i-5} + 5\}\right) \leq \sum_{i=5}^{j-1} \frac{C(r^{8i} + \log^{28} n)}{n^4} = O(n^{-3}). \qquad \square$$

The third lemma of the section shows that out-neighbourhoods continue to grow rapidly until they reach size close to $n/\log n$.

Lemma 7.4. *There is $C' > 0$ such that for all i with $r^i \leq n/\log_r n - 2\log^7 n$,*

$$\mathbb{P}^H\left(d_i^* \geq \log_r^3 n, d_{i+1}^* \leq rd_i^* \cdot \left(1 - \frac{2r^2}{\log_r n}\right)\right) \leq e^{-C'\log^2 n}.$$

Proof. We have

$$\mathbb{P}^H\left(d_i^* \geq \log_r^3 n, d_{i+1}^* \leq rd_i^* \cdot \left(1 - \frac{2r^2}{\log_r n}\right)\right)$$

$$\leq \sup_{a \in [\log_r^3 n, r^{i+1}]} \mathbb{P}^H\left(d_{i+1}^* \leq rd_i^* \cdot \left(1 - \frac{2r^2}{\log_r n}\right) \mid d_i^* = a\right).$$

Condition on $N^*_{\leq i}$, and reveal the random variables $\{L_{w,m} : w \in N^*_i, m \in [r]\}$ one-at-a-time as in the previous proof. When $L_{w,m}$ is revealed there are less than $r^{i+2} + |B_H|$ locations that can cause a conflict, so under $\mathbb{P}^H$ the probability of a conflict is at most $(r^{i+2} + |B_H|)/\hat{n}$. If $d^*_{i+1} \leq rd^*_i - t$ then at least t conflicts occur, so we obtain

$$\mathbb{P}^H\left(d^*_{i+1} \leq rd^*_i \cdot \left(1 - \frac{2r^2}{\log_r n}\right) \mid d^*_i = a\right) \leq \mathbb{P}\left(\mathbf{Bin}\left(ar, \frac{r^{i+2} + |B_H|}{\hat{n}}\right) \geq \frac{2r^3 a}{\log_r n}\right).$$

Using that $r^i \leq n/\log_r n - 2\log^7 n$ and that $|B_H| \leq \log^7 n$ and $\hat{n} \geq n - \log^7 n$, it is straightforward to verify that $ar(r^{i+2} + |B_H|)/\hat{n} \leq r^3 a/\log_r n$. A Chernoff bound then gives

$$\mathbb{P}^H\left(d^*_{i+1} \leq rd^*_i \cdot \left(1 - \frac{2r^2}{\log_r n}\right) \mid d^*_i = a\right) \leq e^{-3r^3 a/(8\log_r n)} \leq e^{-C'\log^2 n},$$

for some constant $C' = C'(r)$. The latter inequality follows since $a \geq \log^3_r n$. $\square$

The following is an easy consequence of the preceding lemma, and concludes the section.

Corollary 7.5. *Let* $j^* = 3\log_r \log_r n + 5$ *and let* $\ell^* = \log_r n - \log_r \log_r n - 1$. *Then there are* $c, C > 0$ *such that*

$$\mathbb{P}^H\left(d^*_{j^*} \geq \log^3_r n, d^*_{\ell^*} \leq \frac{cn}{\log_r n}\right) \leq e^{-C\log^2 n}.$$

Proof. If $d^*_i \geq \log^3_r n$ and $d^*_{i+1} \geq rd^*_i \cdot \left(1 - \frac{2r^2}{\log_r n}\right)$ then $d^*_{i+1} \geq \log^3_r n$. Since $\ell^* - j^* = \log_r n - 4\log_r \log_r n - 6$ we also have

$$\left(1 - \frac{2r^2}{\log_r n}\right)^{\ell^* - j^*} r^{\ell^* - j^*} \log^3 n \geq \left(1 - \frac{2r^2}{\log_r n}\right)^{\log_r n - 1} \frac{n}{r^6 \log^4_r n} \cdot \log^3_r n \geq \frac{e^{-2r^2}}{r^6} \cdot \frac{n}{\log_r n},$$

where we used that $(1 - b/x)^{x-1} \geq e^{-bx}$. With $c = \frac{e^{-2r^2}}{2r^6}$, it follows from the preceding inequalities that if $d^*_{j^*} \geq \log^3_r n$ but $d^*_{\ell^*} < cn/\log_r n$ then there is $i \in [j^*, \ell^* - 1]$ such that $d^*_i \geq \log^3_r n$ and $d^*_{i+1} < rd^*_i \cdot \left(1 - \frac{2r^2}{\log_r n}\right)$. By Lemma 7.4, there exists some C' such that

$$\mathbb{P}^H\left(d^*_{j^*} \geq \log^3_r n, d^*_{\ell^*} < \frac{cn}{\log_r n}\right) \leq \sum_{i=j^*}^{\ell^*-1} \mathbb{P}^H\left(d^*_i \geq \log^3_r n, d^*_{i+1} \leq rd^*_i \cdot \left(1 - \frac{2r^2}{\log_r n}\right)\right)$$

$$\leq (\ell^* - j^*) \cdot e^{-C'\log^2 n}$$

$$\leq e^{-C\log^2 n},$$

$\square$

for some constant $C < C'$.

8. Upper bound on the Diameter

In this section we prove Lemma 4.1 and Proposition 4.2 from Subsection 4.1. Throughout the section, we fix $u, v \in [n]$.

Recall that $k^* = (\eta_r + \epsilon)\log_r n$ and that $\ell^* = \log_r n - \log_r \log_r n$.

Proof of Lemma 4.1. It suffices to prove the lemma assuming $\epsilon < 1/10$. Let $k^* = (\eta_r + \epsilon)\log_r n$, and let $\delta = \frac{\epsilon}{2\eta_r}$. An easy computation shows that $\eta_r \leq 4/5$ for every $r \geq 2$, so $k^* \leq 0.9\log_r n$. Recall the definition $k_0 = k_0(v) = \min\{k : d^-_k(v) \notin (0, \log^4 n)\}$. If $k_0 > k^*$

then $0 < d_{k^*}^-(v) < \log^4 n$, so by Proposition 6.5 we have

$$
\begin{aligned}
\mathbb{P}(k_0 > k^*) &= \mathbb{P}(0 < d_{k^*}^-(v) < \log^4 n) \\
&\leq O\left((r(1-\lambda_r))^{k^*-4\log_r \log n}\right) + O(n^{-3}) \\
&\leq O\left(r^{(\log_r(r(1-\lambda_r)))(\eta_r+\epsilon+o(1))\log_r n}\right) + O(n^{-3}) \\
&= O\left(r^{-(1+2\delta+o(1))\log_r n}\right) + O(n^{-3}) \\
&= O\left(n^{-(1+\delta)}\right),
\end{aligned}
$$

where we used that $1 + \log_r(1 - \lambda_r) = -\eta_r^{-1}$. For all $i < k_0$ we have $d_i^-(v) < \log^4 n$, so if $k_0 \leq k^*$ then $d_{\leq k_0-1}^-(v) < k^*\log^4 n \leq \log^6 n$. In this case, for n large, to have $d_{\leq k_0}^-(v) \geq \log^7 n$ we must have $d_{k_0}^- \geq (\log^7 n)/2$. It follows by Corollary 5.3 and a Chernoff bound

$$
\begin{aligned}
&\mathbb{P}(k_0 \leq k^*, d_{\leq k_0}^-(v) \geq (\log^7 n)/2) \\
&\leq \sup_{k \leq k^*-1} \sup_{q < \log^4 n} \sup_{p < \log^6 n} \mathbb{P}(d_{k+1}^-(v) \geq (\log^7 n)/2 \mid d_k^-(v) = q, d_{\leq k}^-(v) = p) \\
&\leq \mathbb{P}(\mathbf{Bin}(rn/2, 2\log^4 n/n) \geq (\log^7 n)/2) \\
&\leq e^{-(\log^7 n)/8}.
\end{aligned}
$$

Combining the two preceding bounds, the lemma follows. $\qquad\square$

The proof of Proposition 4.2 occupies the remainder of the section. Let $\tau = \min\{j \geq 1 : N_j^+(u) \cap N_{\leq k}^-(v) \neq \emptyset\}$, so in particular $\text{dist}(u, v) = \tau + k$.

Lemma 8.1. *Fix $\epsilon > 0$ and let $\tau' = \min\{j \geq 1 : |N_j^+(u) \setminus V(H)| \geq \epsilon n/\log n\}$. Then for n large,*

$$
\mathbb{P}^H(\tau > \tau' + 1) = e^{-\epsilon r \log^3 n}
$$

Proof. First,

$$
\begin{aligned}
&\mathbb{P}(D[N_{\leq k}^-(v)] = H, \tau > \tau' + 1) \\
&= \sum_F \mathbb{P}(D[N_{\leq k}^-(v)] = H, D[N_{\leq \tau'}^+(u) \setminus V(H)] = F, \tau > \tau' + 1) \\
&= \sum_F \mathbb{P}(D[N_{\leq k}^-(v)] = H, D[N_{\leq \tau'}^+(u) \setminus V(H)] = F) \\
&\qquad \cdot \mathbb{P}(\tau > \tau' + 1 \mid D[N_{\leq k}^-] = H, D[N_{\leq \tau'}^+(u) \setminus V(H)] = F).
\end{aligned}
$$

where the sums are over graphs F with $V(F) \cap V(H) = \emptyset$, such that $u \in V(F)$ and such that, for some $\ell > 0$, $V(F) = \bigcup_{j=0}^\ell N_j^+(F, u)$ and $\ell = \min\{i : |N_i^+(F, u)| \geq \epsilon n/\log n\}$. We now bound the final probability. Under such conditioning, the out-edges from $N_{\tau'}^+$ are uniformly distributed over $([n] \setminus V(H)) \cup B_H$. There are more than $r(\epsilon n/\log n)$ such out-edges; to have $\tau > \tau' + 1$ the heads of such edges must all avoid B_H; so

$$
\begin{aligned}
\mathbb{P}(\tau > \tau' + 1 \mid D[N_{\leq k}^-(v)] = H, D[N_{\leq \tau'}^+(u) \setminus V(H)] = F) &\leq \left(1 - \frac{|B_H|}{\hat{n}}\right)^{r(\epsilon n/\log n)} \\
&\leq \left(1 - \frac{\log^4 n}{n}\right)^{\epsilon r(n/\log n)},
\end{aligned}
$$

the last inequality since $|B_H| \geq \log^4 n$ and $\hat{n} < n$. Using that $1 - x \leq e^{-x}$ this gives

$$\mathbb{P}(D[N^-_{\leq k}] = H, \tau > \tau' + 1) \leq e^{-\epsilon r \log^3 n} \sum_F \mathbb{P}(D[N^-_{\leq k}(v)] = H, D[N^+_{\leq \tau'}(u) \setminus V(H)] = F)$$

$$\leq e^{-\epsilon r \log^3 n} \mathbb{P}(D[N^-_{\leq k}(v)] = H).$$

The result follows. □

Proof of Proposition 4.2. Recall that we set $\ell^* = \log_r n - \log_r \log_r n - 1$, and the notation $N^*_j = N^+_j(u) \setminus V(H)$, $d^*_j = |N^*_j|$ from Section 7. Once n is large enough that $\ell^* + 1 > 5$ we have

$$\mathbb{P}^H(\text{dist}(u, N^-_{\leq k}(v)) > \ell^*+1, \overline{E_{\text{SL}}}) = \mathbb{P}^H(\text{dist}(u, N^-_{\leq k}(v)) > \ell^*+1, N^+_{\leq 5}(u) \cap V(H) = \emptyset, \overline{E_{\text{SL}}}),$$

and we focus on the latter probability. It is convenient to use the shorthand $E = \{N^+_{\leq 5}(u) \cap V(H) = \emptyset\} \cap \overline{E_{\text{SL}}}$.

Now take $\epsilon \in (0, c)$, where c is the constant from Corollary 7.5. and let τ' be as in Lemma 8.1. Then by that lemma,

$$\mathbb{P}^H(\text{dist}(u, N^-_{\leq k}(v)) > \ell^* + 1, E) \leq \mathbb{P}^H(\tau' > \ell^*, E) + e^{-\epsilon r \log^3 n}.$$

We bound the second probability by

$$\mathbb{P}^H(\tau' > \ell^*, E) \leq \mathbb{P}^H(\tau' > \ell^*, d^*_{j^*} \geq \log^3_r n) + \mathbb{P}^H(d^*_{j^*} < \log^3_r n, E).$$

The first term on the right is at most $e^{-C \log^2 n}$ by Corollary 7.5. We further divide the second as

$$\mathbb{P}^H(d^*_{j^*} < \log^3_r n, E) \leq \mathbb{P}^H(d^*_{j^*} < \log^3_r n, d^*_5 \geq 5) + \mathbb{P}^H(d^*_5 < 5, E).$$

Recall that $\sigma = \inf\{i \geq 5 : d^*_i < r^{i-5} + 5\}$. If $d^*_{j^*} < \log^3_r n$ then $d^*_{j^*} < r^{j^*-5}$; if also $d^*_5 \geq 5$ then $5 < \sigma \leq j^*$. Lemma 7.3 then implies that the first probability on the right is $O(n^{-3})$. By the definition of E and by Lemma 7.2, the second probability is also $O(n^{-3})$. Combining all these bounds we obtain $\mathbb{P}^H(\text{dist}(u, N^-_{\leq k}(v)) > \ell^* + 1, \overline{E_{\text{SL}}}) = O(n^{-3})$, as required. □

9. Lower Bound on the Diameter

In this section we prove Lemma 4.4, Equation (4.2) and Lemma 4.6 from Subsection 4.2. Recall that $k^* = k^*(n, \epsilon) = (\eta_r - \epsilon/2) \log_r n$.

Proof of Lemma 4.4. We assume n large throughout. Given a tree T, let $k_1(T) = \inf\{k : |T_k| \geq \log^4 n\}$, let $A(T)$ be the event that $|T_{k^*}| \in [\epsilon \log^3 n, \log^3 n]$, let $B(T)$ be the event that $\max_{i \leq k^*} |T_i| \leq \log^6 n$, and $C(T)$ be the event that $k_1 \leq k^* + 5 \log_r \log n$ and $|T_{k_1}| \leq \log^5 n$. (We may view a deterministic tree as a random tree in the same way as we may view a constant as a random variable, so it is reasonable to call $A(T)$, $B(T)$ and $C(T)$ events even if T is deterministic.)

We first bound the probability that A, B and C occur for a Poisson(r) Galton-Watson tree $\mathcal{T}$.

Let $\omega = \log^3 n$ and $k_0 = \log_r \omega$. By the Kesten-Stigum theorem (see e.g. [4, pp. 24–29]) $r^{-k}|T_k|$ converges almost surely to an absolutely continuous random variable on $(0, \infty)$. As it also converges in probability, for every fixed $0 < c_1 < c_2$,

$$\inf_k \mathbb{P}(c_1 r^k \leq |T_k| \leq c_2 r^k) > 0,$$

where the infimum is taken over all k such that $[c_1 r^k, c_2 r^k] \cap \mathbb{N}$ is non-empty. Taking $c_1 = \epsilon$ and $c_2 = 1$, there exists $c_0 > 0$ such that

$$\mathbb{P}(|T_{k_0}| \in [\epsilon \omega, \omega]) > c_0.$$

By Lemma 6.4 with $\omega = 2$ and $k = k^* - k_0$, there is $a > 0$ such that

$$
\begin{aligned}
\mathbb{P}(|T_{k^*-k_0}| = 1) &\geq a\left(r(1-\lambda_r)\right)^{k^*-k_0} \\
&\geq ar^{\log_r(r(1-\lambda_r))(\eta_r - \epsilon/2 - o(1))\log_r n} \\
&\geq ar^{-(1-\epsilon/2\eta_r - o(1))\log_r n} \\
&\geq ar^{-(1-\epsilon/2)\log_r n} \\
&= an^{\epsilon/2-1},
\end{aligned}
$$

where we used that $\log_r\left(r(1-\lambda_r)\right) = -\eta_r^{-1}$ and that $\eta_r < 1$. We conclude that

$$
\mathbb{P}(A(\mathcal{T})) \geq \mathbb{P}(|T_{k^*-k_0}| = 1)\mathbb{P}(|T_{k_0}| \in [\epsilon\omega, \omega]) > ac_0 n^{\epsilon/2-1}.
$$

Next, if $\overline{B(\mathcal{T})}$ occurs then let $i \leq k^*$ be minimal such that $|\mathcal{T}_i| > \log^6 n$. In order for $A(\mathcal{T})$ to additionally occur the number of descendants of $\mathcal{T}_i$ alive at time k^* must be less than $\log^3 n$. Writing p for the survival probability of a Poisson(r) branching process, it follows as in the proof of Proposition 6.5 that

$$
\mathbb{P}(A(\mathcal{T}), \overline{B(\mathcal{T})}) \leq \mathbb{P}(\mathbf{Bin}(\log^6 n, p) \leq \log^3 n) \leq n^{-3},
$$

the last inequality by a Chernoff bound.

To bound the probability of $\overline{C(\mathcal{T})}$, let $N = N(\mathcal{T})$ be the number of vertices in $\mathcal{T}_{k^*}$ with at least one descendant in $\mathcal{T}_{k^*+5\log_r \log n}$; if $\mathcal{T}_{k^*} = \emptyset$ then $N = 0$. If $C(\mathcal{T})$ does not occur then one of the following *must* occur.

(a) $N < \log^2 n$.
(b) $N \geq \log^2 n$ but $k_1 > k^* + 5\log_r \log n$.
(c) $|\mathcal{T}_{k_1}| > \log^5 n$.

If $A(\mathcal{T})$ occurs then $|\mathcal{T}_{k^*}| \geq \epsilon\log^3 n$, so by the branching property (i.e. the independence of subtrees rooted at elements of $\mathcal{T}_{k^*}$), we have

$$
\mathbb{P}(A(\mathcal{T}), N < \log^2 n) \leq \mathbb{P}(\mathbf{Bin}(\epsilon\log^3 n, p) \leq \log^2 n) < n^{-3}
$$

for large n, by a Chernoff bound. Next, to have $k_1 > k^* + 5\log_r \log n$, every vertex in $\mathcal{T}_k^*$ must have fewer than $\log^4 n$ descendants in $\mathcal{T}_{k^*+5\log_r \log n}$, so by Lemma 6.4 and the branching property we have

$$
\begin{aligned}
\mathbb{P}(N \geq \log^2 n, k_1 > k^* + 5\log_r \log n) &\leq (\mathbb{P}(|\mathcal{T}_{5\log_r \log n}| \in (0, \log^4 n)))^{\log^2 n} \\
&\leq \left(C(r(1-\lambda_r))^{5\log_r \log n - \log_r \log^4 n}\right)^{\log^2 n} \\
&\leq \left(C(r(1-\lambda_r))^{\log \log n}\right)^{\log^2 n} \\
&\leq n^{-3}
\end{aligned}
$$

for large n, the last inequality because $r(1-\lambda_r) < 1$. Finally, by the Markov property and the definition of k_1, writing $\mathbf{Po}(t)$ for a Poisson(t) random variable, we have

$$
\begin{aligned}
\mathbb{P}(|\mathcal{T}_{k_1}| > \log^5 n) &\leq \sup_{m < \log^4 n} \mathbb{P}(\mathbf{Po}(rm) > \log^5 n \mid \mathbf{Po}(rm) \geq \log^4 n) \\
&\leq \mathbb{P}(\mathbf{Po}(r\log^4 n) > \log^5 n - \log^4 n).
\end{aligned}
$$

Standard estimates for the Poisson upper tail (see, e.g., [31], Lemma 1.2) then yield $\mathbb{P}(|\mathcal{T}_{k_1}| > \log^5 n) \leq n^{-3}$. Combining these bounds, we obtain that, for n large,

$$
\mathbb{P}(A(\mathcal{T}), B(\mathcal{T}), \overline{C(\mathcal{T})}) \leq \mathbb{P}(\overline{C(\mathcal{T})} \mid A(\mathcal{T}), B(\mathcal{T})) \leq 3n^{-3}.
$$

Combining inequalities, and choosing $\beta > 0$ appropriately, yields

$$
\mathbb{P}(A(\mathcal{T}), B(\mathcal{T}), C(\mathcal{T})) \geq an^{\epsilon/2-1} - 4n^{-3} \geq 3n^{\beta-1}.
$$

Now, if $A(\mathcal{T}), B(\mathcal{T})$ and $C(\mathcal{T})$ all occur then $|\mathcal{T}_{\leq k_1}| \leq k^* \log^6 n < \log^7 n$, so we may use Lemma 6.3 to transfer our bound from the Poisson(r) Galton-Watson tree to the tree $T^-_{\leq k_1}(v)$. We obtain

$$\mathbb{P}(A(T^-_{\leq k_1}(v)), B(T^-_{\leq k_1}(v)), C(T^-_{\leq k_1}(v))) = \left(1 + O\left(\frac{\log^{14} n}{n}\right)\right) \mathbb{P}(A(\mathcal{T}), B(\mathcal{T}), C(\mathcal{T}))$$
$$\geq 2n^{\beta - 1}.$$

Given that $A(T^-_{\leq k_1}(v)), B(T^-_{\leq k_1}(v))$ and $C(T^-_{\leq k_1}(v)$ all occur, in order to have $v \in F(\epsilon)$ it is sufficient that $D[N^-_{\leq k_1}(v)]$ is a tree, i.e. that $D[N^-_{\leq k_1}(v)] = T^-_{\leq k_1}(v)$.

Finally, when $A(T^-_{\leq k_1}(v)), B(T^-_{\leq k_1}(v))$ and $C(T^-_{\leq k_1}(v))$ all occur we have $|V(T^-_{\leq k_1}(v))| < \log^7 n$. By Corollary 5.2, in this case for each element $u \in V(T^-_{\leq k_1}(v))$, the probability that there is a non-tree edge u to $V(T^-_{\leq k_1}(v))$ is at most $(r \log^7 n)/n$. It follows that

$$\mathbb{P}(A(T^-_{\leq k_1}(v)), B(T^-_{\leq k_1}(v)), C(T^-_{\leq k_1}(v)), D[N^-_{\leq k_1}(v)] \neq T^-_{\leq k_1}(v)) \leq \frac{r \log^{14} n}{n} < n^{\beta - 1}.$$

The result follows. $\qquad \square$

The next lemma is our key tool for controlling joint probabilities of in-neighbourhoods of distinct vertices. Results of this type are standard in sparse random undirected graphs.

Lemma 9.1. *Fix* $u, v \in [n]$ *and trees* T, T', *with roots* u *and* v, *respectively, and with* $V(T) \cup V(T') \subset [n]$ *and* $V(T) \cap V(T') = \emptyset$. *Then*

$$\mathbb{P}(D[N^-_{\leq h}(u)] = T, D[N^-_{\leq h'}(v)] = T') = \left(1 + O\left(\frac{|V(T)|^2}{n - |V(T')|} + \frac{|V(T')|^2}{n - |V(T)|}\right)\right)$$
$$\cdot \mathbb{P}(D[N^-_{\leq h}(u)] = T) \cdot \mathbb{P}(D[N^-_{\leq h'}(v)] = T').$$

Proof. Recall that T_i is the i-th generation of tree T. Write h and h' for the respective heights of T and T', and t and t' for their respective sizes. In order that $D[N^-_{\leq h}(u)] = T$, it is necessary and sufficient that the following events occur.

- For each $x \in V(T) \setminus \{u\}$, there is an edge from x to $p_T(x)$ in D; call this event $A_1(u, T)$.
- There are no other edges within $D[V(T)]$; call this event $A_2(u, T)$.
- There are no edges from $[n] \setminus D[V(T)]$ to $V(T) \setminus T_h$; call this event $A_3(u, T)$.

Note that A_3 is independent of A_1 and A_2, so we have

$$\mathbb{P}(D[N^-_{\leq h}(u)] = T) = \mathbb{P}(A_1(u, T), A_2(u, T)) \cdot \left(\frac{n - t + |T_h|}{n}\right)^{r(n-t)}. \tag{9.1}$$

We now consider two such events simultaneously. Observe that if T' has root v and height h', and $V(T') \cap V(T) = \emptyset$, then $A_1(u, T) \cap A_2(u, T)$ is independent of $A_1(v, T') \cap A_2(v, T')$. We thus have

$$\mathbb{P}(D[N^-_{\leq h}(u)] = T, D[N^-_{\leq h'}(v)] = T')$$
$$= \mathbb{P}(A_1(u, T), A_2(u, T)) \cdot \mathbb{P}(A_1(v, T'), A_2(v, T'))$$
$$\cdot \mathbb{P}(A_3(u, T), A_3(v, T') \mid A_1(u, T), A_2(u, T), A_1(v, T'), A_2(v, T')). \tag{9.2}$$

Given that $A_1(u, T)$ and $A_2(u, T)$ occur, there are precisely $1 + (r-1)t$ edges leaving $V(T)$, and the heads of these edges are uniformly distributed over $[n] \setminus V(T)$. The conditional probability no such edges have head in $V(T') \setminus T'_{h'}$ is

$$\left(\frac{n - t - t' + |T'_{h'}|}{n - t}\right)^{1 + (r-1)t}.$$

Similar considerations for edges leaving $V(T')$ and edges with tail in $[n] \setminus (V(T) \cup V(T'))$ yield the identity

$$\mathbb{P}(A_3(u,T), A_3(v,T') \mid A_1(u,T), A_2(u,T), A_1(v,T'), A_2(v,T'))$$
$$= \left(1 - \frac{t' - |T'_{h'}|}{n-t}\right)^{1+(r-1)t} \cdot \left(1 - \frac{t - |T_h|}{n-t'}\right)^{1+(r-1)t'} \cdot \left(1 - \frac{t + t' - |T_h| - |T'_{h'}|}{n}\right)^{r(n-t-t')}.$$

Combined with (9.1) and (9.2), straightforward arguments give

$$\mathbb{P}(D[N^-_{\leq h}(u)] = T, D[N^-_{\leq h'}(v)] = T')$$
$$= \left(1 + O\left(\frac{t^2}{n-t'} + \frac{(t')^2}{n-t}\right)\right) \mathbb{P}(D[N^-_{\leq h}(u)] = T) \cdot \mathbb{P}(D[N^-_{\leq h'}(v)] = T'). \qquad \square$$

The following corollary provides a proof for Equation (4.2).

Corollary 9.2. *For distinct* $u, v \in [n]$ *we have* $\mathbb{P}(u, v \in F) \leq (1 + o(1))(\mathbb{P}(u \in F) + \log^{15} n/n)\mathbb{P}(v \in F)$.

Proof. We first divide according to whether or not $N^-_{\leq k_1(u)}(u) \cap N^-_{\leq k_1(v)}(v)$ is empty:

$$\mathbb{P}(u, v \in F)$$
$$= \mathbb{P}(u, v \in F, N^-_{\leq k_1(u)}(u) \cap N^-_{\leq k_1(v)}(v) \neq \emptyset) + \mathbb{P}(u, v \in F, N^-_{\leq k_1(u)}(u) \cap N^-_{\leq k_1(v)}(v) = \emptyset).$$

We start with the first term on the right. If $v \in F$ then $|N^-_{\leq k_1(v)}(v))| \leq \log^7 n$, so by symmetry

$$\mathbb{P}(v \in F, u \in N^-_{\leq k_1(v)}(v)) \leq \frac{\log^7 n}{n} \cdot \mathbb{P}(v \in F).$$

Next, by conditioning on $N^-_{\leq k_1(v)}(v)$ we have

$$\mathbb{P}(u, v \in F, u \notin N^-_{\leq k_1(v)}(v), N^-_{\leq k_1(u)}(u) \cap N^-_{\leq k_1(v)}(v) \neq \emptyset)$$
$$\leq \sum_{\substack{\{T \in \mathcal{T}(v): \\ u \notin V(T)\}}} \mathbb{P}(D[N^-_{\leq k_1(v)}(v)] = T) \cdot \mathbb{P}(u \in F, N^-_{\leq k_1(u)}(u) \cap V(T) \neq \emptyset \mid D[N^-_{\leq k_1(v)}(v)] = T).$$

In order to bound the final probability, first fix T as in the supremum and suppose that $D[N^-_{\leq k_1(v)}(v)] = T$. Consider the iBFS procedure starting from u. Recall that at step i, R^-_i is the set of explored vertices and S^-_i is the set of discovered vertices. Let $i_0 = \min\{i : |R^-_i \cup S^-_i| > \log^7 n\}$. If $u \in F$ then $|N^-_{\leq k_1(u)}(u)| \leq \log^7 n$, so to have $N^-_{\leq k_1(u)}(u) \cap V(T) = \emptyset$ it suffices that $(R^-_{i_0-1} \cup S^-_{i_0-1}) \cap V(T) = \emptyset$. Since $|V(T)| \leq \log^7 n$, by Lemma 5.5 we thus have

$$\mathbb{P}(u \in F, N^-_{\leq k_1(u)}(u) \cap V(T) \neq \emptyset \mid D[N^-_{\leq k_1(v)}(v)] = T) \leq \frac{r(\log^7 n + 1)\log^7 n}{n - 2\log^7 n}.$$

Together with the two preceding displayed equations, for n large this gives

$$\mathbb{P}(u, v \in F, N^-_{\leq k_1(u)}(u) \cap N^-_{\leq k_1(v)}(v) \neq \emptyset)$$
$$\leq \frac{\log^7 n}{n} \cdot \mathbb{P}(v \in F) + \sum_{\{T \in \mathcal{T}(v): u \notin V(T)\}} \mathbb{P}(D[N^-_{\leq k_1(v)}(v)] = T) \cdot \frac{r(\log^7 n + 1)\log^7 n}{n - 2\log^7 n}$$
$$\leq \frac{\log^{15} n}{n} \cdot \mathbb{P}(v \in F), \tag{9.3}$$

the last inequality since

$$\sum_{\{T \in \mathcal{T}(v): u \notin V(T)\}} \mathbb{P}(D[N^-_{\leq k_1(v)}(v)] = T) = \mathbb{P}(v \in F, u \notin N^-_{\leq k_1(v)}(v)).$$

We now turn to the case that $N^-_{\leq k_1(u)}(u)$ and $N^-_{\leq k_1(v)}(v)$ are disjoint. We have

$$\mathbb{P}(u,v \in F, N^-_{\leq k_1(u)}(u) \cap N^-_{\leq k_1(v)}(v) = \emptyset)$$

$$= \sum_{\substack{\{(T,T')\in\mathcal{T}(u)\times\mathcal{T}(v):\\ V(T)\cap V(T')=\emptyset\}}} \mathbb{P}(D[N^-_{\leq k_1(u)}(u)] = T, D[N^-_{\leq k_1(v)}(v)] = T').$$

$$= \left(1 + O\left(\frac{\log^{14} n}{n}\right)\right) \sum_{\substack{\{(T,T')\in\mathcal{T}(u)\times\mathcal{T}(v):\\ V(T)\cap V(T')=\emptyset\}}} \mathbb{P}(D[N^-_{\leq k_1(u)}(u)] = T) \cdot \mathbb{P}(D[N^-_{\leq k_1(v)}(v)] = T'),$$

the last line by Lemma 9.1. (Although $k_1(u)$ and $k_1(v)$ are random, by the same argument as in Lemma 4.6 we may replace them by the deterministic values $h(T)$ and $h(T')$ without affecting the probability, so Lemma 9.1 indeed applies.) Summing over all pairs $(T,T') \in \mathcal{T}(u) \times \mathcal{T}(v)$ gives an upper bound, so we obtain

$$\mathbb{P}(u,v \in F, N^-_{\leq k_1(u)}(u) \cap N^-_{\leq k_1(v)}(v) = \emptyset) \leq (1+o(1))\mathbb{P}(u \in F)\mathbb{P}(v \in F).$$

Together with (9.3) this completes the proof. $\square$

We finally prove Lemma 4.6.

Proof of Lemma 4.6. Fix $\epsilon > 0$ and write $F = F(\epsilon)$. If D_0 is attractive then with high probability every vertex v with $\max_{u\in[n]} \text{dist}(u,v) < \infty$ satisfies $v \in V(D_0)$. Since D_0 is attractive whp [19], in order to show $\mathbb{P}(F \subset V(D_0)) = 1 - o(1)$, it suffices to show that whp, for all $v \in F$ and $u \in [n]$ we have $\text{dist}(u,v) < \infty$.

Let $\mathcal{T}(v)$ be the set of digraphs T with $v \in V(T)$ and $V(T) \subset [n]$ such that if $D[N^-_{\leq k_1}(v)] = T$ then v is a flag. By the definition of a flag, all the elements of $\mathcal{T}(v)$ are rooted at v. If $D[N^-_{\leq k_1}(v)] = T$ then $N^-_{\leq k_1}(v)$ contains no loop vertices. It follows that $\mathcal{T}(v)$ is precisely the set of graphs T such that $\mathbb{P}(D[N^-_{\leq k_1}(v)] = T, v \in F, \overline{E_{\text{SL}}}) > 0$.

For $T \in \mathcal{T}(v)$ we thus have

$$\mathbb{P}(D[N^-_{\leq k_1}(v)] = T, v \in F, \overline{E_{\text{SL}}}) = \mathbb{P}(D[N^-_{\leq k_1}(v)] = T, \overline{E_{\text{SL}}}).$$

It follows that

$$\mathbb{P}(\exists u, v \in [n]: v \in F, \text{dist}(u,v) = \infty)$$

$$\leq \mathbb{P}(E_{\text{SL}}) + \sum_{u,v\in[n]} \mathbb{P}(v \in F, \text{dist}(u,v) = \infty, \overline{E_{\text{SL}}})$$

$$= \mathbb{P}(E_{\text{SL}}) + \sum_{u,v\in[n]} \sum_{T\in\mathcal{T}} \mathbb{P}(\text{dist}(u,v) = \infty, \overline{E_{\text{SL}}}, D[N^-_{\leq k_1}(v)] = T)$$

We bound the inner sum by writing

$$\sum_{T\in\mathcal{T}} \mathbb{P}(\text{dist}(u,v) = \infty, \overline{E_{\text{SL}}}, D[N^-_{\leq k_1}(v)] = T)$$

$$= \sum_{T\in\mathcal{T}} \mathbb{P}(D[N^-_{\leq k_1}(v)] = T) \cdot \mathbb{P}(\text{dist}(u,v) = \infty, \overline{E_{\text{SL}}} \mid D[N^-_{\leq k_1}(v)] = T)$$

$$\leq \sup_{T\in\mathcal{T}} \mathbb{P}(\text{dist}(u,v) = \infty, \overline{E_{\text{SL}}} \mid D[N^-_{\leq k_1}(v)] = T) \cdot \sum_{T\in\mathcal{T}} \mathbb{P}(D[N^-_{\leq k_1}(v)] = T)$$

$$\leq \sup_{T\in\mathcal{T}} \mathbb{P}(\text{dist}(u,v) = \infty, \overline{E_{\text{SL}}} \mid D[N^-_{\leq k_1}(v)] = T),$$

the final bound because a sum of probabilities of disjoint events is at most one.

Now fix $T \in \mathscr{T}$ and write $h = h(T)$ for the height of T (i.e., the greatest number of edges on a path ending at the root v). Observe that if $D[N^-_{\leq k_1}(v)] = T$ then $k_1 = h(T)$. We thus have the equality of events

$$\{D[N^-_{\leq k_1}(v)] = T\} \cap \{v \in F\} = \{D[N^-_{\leq k_1}(v)] = T\} = \{D[N^-_{\leq h}(v)] = T\}.$$

If $D[N^-_{\leq h}(v)] = T$ then the event $E_h = \{d^-_h(v) \geq \log^4 n, d^-_{\leq h}(v) \leq \log^7 n\}$ from Proposition 4.2 occurs (since in this case v is a flag), so

$$\{D[N^-_{\leq h}(v)] = T\} = \{D[N^-_{\leq h}(v)] = T\} \cap E_h,$$

so $\mathbb{P}(D[N^-_{\leq h}(v)] = T}, E_h, \overline{E_{\mathrm{SL}}}) > 0$. It follows by Proposition 4.2 that $\mathbb{P}(\mathrm{dist}(u,v) = \infty, \overline{E_{\mathrm{SL}}} \mid D[N^-_{\leq k_1}(v)] = T) = O(n^{-3})$. Using this bound, the result follows from the two preceding inequalities and the fact that $\mathbb{P}(E_{\mathrm{SL}}) = \Theta(n^{1-r}) = O(n^{-1})$. $\square$

10. The Stationary Distribution

In this section we prove Theorem 1.2. Recall that $D_0 = D_0(n,r)$ is the largest strongly connected component of D and that with high probability D_0 is attractive [19] and ergodic [5]. Write $\pi_{\max} = \pi_{\max}(D)$ and $\pi_{\min} = \pi_{\min}(D)$. Also, write $\mathrm{X} = (X_k, k \geq 0)$ for simple random walk on $D = D(n,r)$.

It is important to distinguish the randomness of the graph D from that of the walk X. For $v \in V(D) = [n]$, write $\mathbf{P}_v$ for the (random) probability measure under which X has the law of simple random walk on D with $X_0 = v$, and $\mathbf{E}_v$ for the corresponding expectation operator. It is handy to have a concrete description of X under $\mathbf{P}_v$, as follows. Recall that D has edges $\{(i, L_{i,j}), (i,j) \in [n] \times [r]\}$ (this is the "canonical construction" from the introduction). Let $(U_k, k \geq 0)$ be independent and uniformly distributed over $\{1, \ldots, r\}$. Then set $X_0 = v$ and for $k \geq 0$ let $X_{k+1} = L_{X_k, U_k}$.

10.1. **Bounding $\pi_{\max}$.** Fix $k \geq 1$ and view $D[N^-_{\leq k}(v)]$ as a *maze*, which a random walk attempting to reach v must navigate. The maze entrances are the elements of $N^-_k(v)$, and the treasure lies at v. Suppose that the random walk follows an edge e from $N^-_{\leq k}(v)$ to its complement. After following the edge, the random walk's position has distance greater than k from v. Since the distance to v decreases by at most one in a single random walk step, this means that in order to reach v after leaving $N^-_{\leq k}(v)$, the random walk must pass through $N^-_k(v)$: it must restart at one of the maze entrances.

With the preceding paragraph in mind, for positive integer h we say that $D[N^-_{\leq k}(v)]$ is *h-hard* if for every directed path P from $N^-_k(v)$ to v within $D[N^-_{\leq k}(v)]$, we have

$$\#\{u \in V(P), |E(u, N^-_{\leq k}(v))| = 1\} \geq h.$$

Perhaps more picturesque: the maze is h-hard if no matter what entrance is chosen, along any potential path to the treasure there are at least h locations where only a single direction stays within the maze; the other $(r-1)$ possibilities deposit the searcher outside of the maze walls.

For $S \subset [n]$ let $\tau_S = \inf\{k \geq 0 : X_k \in S\}$ and let $\tau_S^+ = \inf\{k > 0 : X_k \in S\}$.

Lemma 10.1. *For $k \geq 1$, if $D[N^-_{\leq k}(v)]$ is h-hard then*

$$\pi(v) \leq \frac{1}{r^h \cdot \mathbf{P}_v\left(\tau_{[n] \setminus N^-_{\leq k}(v)} \leq \tau_v^+\right)}.$$

Proof. If the maze is h-hard then from any $u \in N^-_k(v)$,

$$\mathbf{P}_u\left(\tau_v < \tau_{[n] \setminus N^-_{\leq k}(v)}\right) \leq r^{-h}. \tag{10.1}$$

To see this, simply note that in order to have $\tau_v < \tau_{[n]\setminus N^-_{\leq k}(v)}$ the walk must visit at least h vertices $w \in N^-_{\leq k}(v)$ with $|E(w, N^-_{\leq k}(v))| = 1$. But for such a vertex w we have $\mathbf{P}_w\left(X_1 \in N^-_{\leq k}(v)\right) = 1/r$ and the inequality follows by the Markov property.

We now use that

$$\frac{1}{\pi(v)} = \mathbf{E}_v\left(\tau_v^+\right) \geq \mathbf{E}_v\left(\tau_v^+ \mid \tau_{[n]\setminus N^-_{\leq k}(v)} \leq \tau_v^+\right) \cdot \mathbf{P}_v\left(\tau_{[n]\setminus N^-_{\leq k}(v)} \leq \tau_v^+\right).$$

Let K be the number of visits to $N^-_k(v)$ before the walk visits v. Since the inequality (10.1) holds for all $u \in N^-_k(v)$, it follows that for all $w \in [n]\setminus N^-_{\leq k}(v)$ we have $\mathbf{E}_w\left(\tau_v\right) \geq \mathbf{E}_w\left(K\right) \geq r^h$. Therefore

$$\mathbf{E}_v\left(\tau_v^+ \mid \tau_{[n]\setminus N^-_{\leq k}(v)} \leq \tau_v^+\right) \geq \inf_{w \in [n]\setminus N^-_{\leq k}(v)} \mathbf{E}_w\left(\tau_v\right) \geq r^h,$$

and the result follows. $\qquad\square$

Lemma 10.2. *Fix $\delta > 0$ and let $\ell^* = (1-\delta)\log_r n$. Then*

$$\mathbb{P}\left(|N^-_{\leq \ell^*}(v)| \geq n^{1-\delta/2}\right) = O(n^{-4}).$$

Proof. Choose $\alpha > 0$ small, and let A be the event that for all $k \geq 0$ and all $v \in [n]$ we have $d^-_k(v) \leq (r+\alpha)^k \log^2 n$. By Proposition 6.1, we have $\mathbb{P}(\overline{A}) = O(n^{-4})$. Assuming α is small enough with respect to δ, we also have

$$|N^-_{\leq \ell^*}(v)| = d^-_{\leq \ell^*}(v) = \sum_{k=0}^{\ell^*} d^-_k(v) \leq (r+\alpha)^{\ell^*+1} \log^2 n < n^{1-\delta/2},$$

so

$$\mathbb{P}(|N^-_{\leq \ell^*}(v)| \geq n^{1-\delta/2}) \leq \mathbb{P}(|N^-_{\leq \ell^*}(v)| \geq n^{1-\delta/2} \mid A) + \mathbb{P}(\overline{A}) = O(n^{-4}). \qquad\square$$

Proposition 10.3. *Fix $\delta > 0$ and let $\ell^* = (1-\delta)\log_r n$ and $h = (1-2\delta)\log_r n$. Then*

$$\mathbb{P}\left(\bigcap_{v \in [n]} D[N^-_{\leq \ell^*}(v)] \ \ is \ h\text{-}hard\right) = 1 - O(n^{-3}).$$

Proof. Recall from the introduction that $D = D(n,r)$ has edges $\{(i, L_{i,j}) : (i,j) \in [n]\times[r]\}$. Fix $v \in [n]$. For each $k \geq 1$ let $T^-_{\leq k} = T^-_{\leq k}(D, v)$ be the iBFS tree of $D[N^-_{\leq k}(v)]$ described in Section 3. For $w \in [n]$, $w \neq v$ let $Y(w) = |E(w, N^-_{\leq \ell^*}(v))| - 1$, and let $Y(v) = |E(v, N^-_{\leq \ell^*}(v))|$. Observe that there may be multiple edges from w to a vertex $u \in N^+(w)$, so $Y(w)$ may not equal $|N^+(w) \cap N^-_{\leq \ell^*}(v)| - 1$.

For $w \in N^-_{\leq \ell^*}(v)$ with $w \neq v$, the parent of w in $T^-_{\leq \ell^*}$ lies in $N^+(w) \cap N^-_{\leq \ell^*-1}(v) \subset N^-_{\leq \ell^*}(v)$, so $Y(w) \geq 0$. The key insight of the proof is that if $D[N^-_{\leq \ell^*}(v)]$ is not h-hard then there is some simple path P in $D[N^-_{\leq \ell^*}(v)]$ from $N^-_{\ell^*}(v)$ to v along which at least $|P| - h$ vertices w have $Y(w) > 0$. To show that $\mathbb{P}(D[N^-_{\leq \ell^*}(v)]$ is not h-hard) is small it thus suffices to show that with high probability no such path exists.

The sets $\{Y(w) : w \in N^-_{\leq \ell^*}(v)\}$ are conditionally independent given $T^-_{\leq \ell^*}$. Furthermore, given $T^-_{\leq \ell^*}$, by Corollary 5.2 we also have $Y(w) \preceq \mathbf{Bin}(r-1, |N^-_{\leq \ell^*}(v)|/n)$ for all $w \in N^-_{\leq \ell^*}(v)$.

Let A be the event that $|N^-_{\leq \ell^*}(v)| \leq n^{1-\delta/2}$. For $S \subset [n]$, it follows that on A the random variable $B(S) = |\{w \in \overline{S} : Y(w) > 0\}|$ is stochastically dominated by $\mathbf{Bin}(|S|, (r-1)n^{-\delta/2})$. Furthermore, by Lemma 10.2 we have $\mathbb{P}(A) = 1 - O(n^{-4})$.

We would like to conclude as follows. Let S be any path from $T_{\ell^*}^-$ (the last generation of $T_{\leq\ell^*}^-$) to v. The arguments of the preceding paragraphs suggest the bound

$$\mathbb{P}(B(S) \geq |S| - h \mid A) \leq \mathbb{P}(\mathbf{Bin}(|S|, rn^{-\delta/2}) \geq |S| - h) \leq 2^{|S|}(rn^{-\delta/2})^{|S|-h}.$$

On A we have $|T_{\ell^*}^-| < |T_{\leq\ell^*}^-| \leq n^{1-\delta/2}$, so there are less than $m \cdot r^t$ paths of length t from $T_{\ell^*}^-$ to v. Now use the preceding inequality and a union bound over paths of length t and over $t \geq \ell^*$.

To make the preceding argument rigorous, we need to deal with the fact that the set of paths from $T_{\ell^*}^-$ to v are random (even conditional on $T_{\ell^*}^-$, as such paths may follow edges of $D[N_{\leq\ell^*}^-(v)]$ which are not edges of $T_{\ell^*}^-$). To do so, condition on $T_{\leq\ell^*}^-$, fix $w \in T_{\ell^*}^- = N_{\ell^*}^-(v)$ and a string $s = s_1 s_2 \dots s_t \in [r]^t$ of length $|s| = t$. This string uniquely specifies a path $P = P(w,s) = (p_i(w,s), 0 \leq i \leq t)$ in D: at step i follow the s_i-th edge leaving the current vertex. Formally, we let $p_0 = w$ and, for $1 \leq i \leq t$, let $p_i = L_{p_{i-1}, s_i}$.

We reveal the path P edge-by-edge, starting from w. By the independence of the sets $Y(u)$, for each $0 \leq i < t$, given that the sub-path $p_0, \dots, p_i$ is simple (in particular $p_i \notin \{p_0, \dots, p_{i-1}\}$) then $Y(p_i)$ is conditionally independent of $p_0, \dots, p_{i-1}$ and of $Y(p_0), \dots, Y(p_{i-1})$. It follows that

$$\mathbb{P}(Y(p_i) > 0 \mid T_{\leq\ell^*}, A, (p_0, \dots, p_i) \text{ simple path in } D[N_{\leq\ell^*}^-(v)], (Y(p_j), j < i)) \leq rn^{-\delta/2}.$$

By repeated conditioning, we obtain

$$\mathbb{P}(P(w,s) \text{ is a simple path in } D[N_{\leq\ell^*}^-(v)], B(P(w,s)) \geq t - h \mid T_{\leq\ell^*}, A)$$
$$\leq \mathbb{P}(\mathbf{Bin}(t, rn^{-\delta/2}) \geq t - h)$$
$$\leq 2^t(rn^{-\delta/2})^{t-h}.$$

Now let $A_w(t)$ be the event that there is a simple path P of length t starting from w and staying within $D[N_{\leq\ell^*}^-(v)]$, for which $B(P) \geq t - h$. All possible such paths are described by a string $s \in [r]^t$, so by the preceding inequality and a union bound,

$$\mathbb{P}(A_w(t) \mid T_{\leq\ell^*}, A) \leq (2r)^t(rn^{-\delta/2})^{t-h} \leq \frac{(2r^2)^t}{n^{\delta(t-h)/2}}.$$

Since $\ell^* - h \geq \delta \log n$, this yields that

$$\mathbb{P}\left(\bigcup_{w \in N_{\ell^*}^-(v)} \bigcup_{t \geq \ell^*} A_w(t) \mid T_{\leq\ell^*}, A\right) \leq n \frac{(2r^2)^{\ell^*}}{n^{\delta(\ell^*-h)/2}} \sum_{j \geq 0} \left(\frac{2r^2}{n^{\delta/2}}\right)^j = O(n^{-4}).$$

Since $\mathbb{P}(A) = 1 - O(n^{-4})$ this bound also holds unconditionally. But, as described in the first two paragraphs of the proof,

$$\{D[N_{\leq\ell^*}^-(v)] \text{ is not } h\text{-hard}\} \subset \bigcup_{w \in N_{\ell^*}^-(v)} \bigcup_{t \geq \ell^*} A_w(t),$$

so $\mathbb{P}(D[N_{\leq\ell^*}^-(v)] \text{ is not } h\text{-hard}) = O(n^{-4})$. A union bound over $v \in [n]$ completes the proof. $\qquad\square$

Before proving our bounds on $\pi_{\max}$ we require one final result, which says that with high probability there is at least one escape route along each path from $N_{\log\log n}^-(v)$ to v, for all v.

Lemma 10.4. *For $v \in [n]$ let E_v be the event that each path from $N^-_{\log \log n}(v)$ to v contains at least one vertex w with $|N^+(w) \cap N^-_{\leq \ell^*}(v)| = 1$. Then*

$$\mathbb{P}\left(\bigcap_{v \in [n]} E_v\right) = 1 - O(n^{-3}).$$

Note that the event E_v is not the event that $D[N^-_{\leq \log \log n}(v)]$ is 1-hard: in E_v we require the vertex w to send the searcher not outside of $D[N^-_{\leq \log \log n}(v)]$ but rather out of the larger maze $D[N^-_{\leq \ell^*}(v)]$. The proof of Lemma 10.4 follows the same lines as that of Proposition 10.3 but is simpler, and is omitted.

Theorem 10.5. *For every $\epsilon > 0$, with high probability we have*

$$\frac{1}{n} \leq \pi_{\max} \leq \frac{1}{n^{1-\epsilon}},$$

Proof. The lower bound holds deterministically since $\sum_{v \in [n]} \pi(v) = 1$.

To prove the upper bound, fix $v \in [n]$ and $\delta \in (0, \epsilon/2)$. Write $\ell^* = (1 - \delta) \log_r n$, and $T_{\leq \ell^*} = T_{\leq \ell^*}(D, v)$.

First, note that if $N^+(v) \setminus N^-_{\leq \log \log n}(v) \neq \emptyset$ then $\mathbf{P}_v\left(X_1 \notin N^-_{\leq \log \log n}(v)\right) \geq 1/r$. We have

$$\mathbb{P}(N^+(v) \setminus N^-_{\leq \log \log n}(v) = \emptyset)$$
$$\leq \mathbb{P}(|N^-_{\leq \log \log n}(v)| \geq n^{1/3}) + \mathbb{P}(N^+(v) \setminus N^-_{\leq \log \log n}(v) = \emptyset \mid |N^-_{\leq \log \log n}(v)| < n^{1/3})$$
$$= O(n^{-3}) + O(n^{-4/3}),$$

the first by Proposition 6.1 and the second by Corollary 5.2.

Next, if the event E_v from Lemma 10.4 occurs then for all $w \notin N^-_{\leq \log \log n}(v)$ we have $\mathbf{P}_w\left(\tau_{[n] \setminus N^-_{\leq \ell^*}(v)} \leq \tau_v\right) \geq 1/r$. By the Markov property, it follows that if $N^+(v) \setminus N^-_{\leq \log \log n}(v) \neq \emptyset$ and E_v, then

$$\mathbf{P}_v\left(\tau_{[n] \setminus N^-_{\leq \ell^*}(v)} \leq \tau_v^+\right) \geq \frac{1}{r}\mathbf{P}_v\left(X_1 \notin N^-_{\leq \log \log n}(v)\right) \geq \frac{1}{r^2}.$$

By the preceding paragraph and Lemma 10.4, we thus have $\mathbf{P}_v\left(\tau_{[n] \setminus N^-_{\leq \ell^*}} \leq \tau_v^+\right) \geq r^{-2}$ with $\mathbb{P}$-probability $1 - O(n^{-4/3})$.

Finally, if in addition $D[N^-_{\leq \ell^*}(v)]$ is h-hard then by Lemma 10.1 we obtain that $\pi(v) \leq r^{h-2} \leq (1 + o(1))n^{-(1-2\delta)}$. By Proposition 10.3 we thus have $\pi(v) \leq (1 + o(1))n^{-(1-2\delta)}$ with probability $1 - O(n^{-4/3})$. A union bound over $v \in [n]$ then completes the proof. $\square$

10.2. **Bounding $\pi_{\min}$.** We bound $\pi_{\min}$ from below using the following lemma.

Lemma 10.6. *Let D be any r-out regular digraph. If D is ergodic and has diameter $\mathrm{diam}(D) \leq d$, then*

$$\pi_{\min} \geq \frac{1}{1 + dr^d}.$$

Proof. Fix $v \in V(D)$. For any $k \in [d]$ and $u \in N^-_k(D, v)$, let $K(u, k) \geq 1$ be the number of directed paths of length k from u to v. Observe that the probability of following each such path is precisely r^{-k}, since D is r-out regular. Furthermore, since π is stationary, it satisfies

$$\pi(v) \geq \sum_{u \in N^-_k(v)} \pi(u) \cdot \frac{K(u, k)}{r^k} \geq \sum_{u \in N^-_k(v)} \frac{\pi(u)}{r^k}.$$

By averaging over $k \in [d]$ we have

$$\pi(v) \geq \frac{1}{d} \sum_{k=1}^{d} \sum_{u \in N_k^-(v)} \frac{\pi(u)}{r^k} \geq \frac{1 - \pi(v)}{dr^d} \ ,$$

the last inequality since $\operatorname{diam}(D) \leq d$ so $\bigcup_{k=1}^{d} N_k^-(v) = V(D) \setminus \{v\}$. The lemma follows. $\square$

Theorem 10.7. *For every $\epsilon > 0$ we have*

$$\frac{1}{n^{1+\eta_r+\epsilon}} \leq \pi_{\min} \leq \frac{1}{n^{1+\eta_r-\epsilon}} \ ,$$

with high probability.

Proof. Fix $\epsilon > 0$ small. It is a straightforward consequence of Theorem 1.1 and Lemma 10.6 that $\pi_{\min} \geq n^{-(1+\eta_r+\epsilon)}$ with high probability. It remains to show that $\pi_{\min}$ is small with high probability.

Let $k^* = (\eta_r - \epsilon/2) \log_r n$, and recall from Section 9 the definition of the set $F = F(\epsilon)$ of ϵ-flags. In particular, if $v \in F(\epsilon)$ then $D[N_{\leq k^*}^-(v)]$ is a tree. It follows that if $v \in F(\epsilon)$ then $D[N_{\leq k^*}^-(v)]$ is k^*-hard.

Let A be the event that $\pi_{\max} > n^{-(1-\epsilon/6)}$. By Corollary 4.5, $\mathbb{P}(F = \emptyset) = o(1)$, and by Theorem 10.5, $\mathbb{P}(\overline{A}) = o(1)$. Therefore

$$\mathbb{P}(\pi_{\min} > n^{-(1+\eta_r-\epsilon)}) = \mathbb{P}(\pi_{\min} > n^{-(1+\eta_r-\epsilon)}, F \neq \emptyset, A) + o(1) \ . \qquad (10.2)$$

Fix $v \in [n]$, and for $u \in [n]$ let $K(u)$ be the number of paths of length k^* from u to v. Using the stationarity of π we have

$$\pi(v) = \sum_{u \in [n]} \frac{K(u)}{r^{k^*}} \cdot \pi(u) \ .$$

If v is a flag then $D[N_{\leq k^*}^-(v)]$ is a tree, $K(u) = 0$ for $u \notin N_{k^*}^-(v)$ and $K(u) = 1$ for $u \in N_{k^*}^-(v)$. In this case we also have $|N_{k^*}^-(v)| \leq \log^7 n$. Finally, on A we have $\pi(u) \leq n^{-(1-\epsilon/6)}$. On the event $\{v \in F\} \cap A$, we thus obtain the bound

$$\pi(v) \leq |N_{k^*}^-(v)| \cdot \frac{1}{r^{k^*}} \cdot n^{-(1-\epsilon/6)} \leq \frac{\log^7 n}{n^{1+\eta_r-2\epsilon/3}} \leq \frac{1}{n^{1+\eta_r-\epsilon}} \ .$$

In other words, on A, every vertex $v \in F$ *deterministically* satisfies $\pi(v) \leq n^{-(1+\eta_r-\epsilon)}$, so in this case if F is non-empty then $\pi_{\min} \leq n^{-(1+\eta_r-\epsilon)}$. It follows that the probability on the right of (10.2) is zero, so $\mathbb{P}(\pi_{\min} > n^{-(1+\eta_r-\epsilon)}) = o(1)$, as required. $\square$

Proof of Theorem 1.2. The theorem is now an immediate consequence of Theorems 10.5 and 10.7. $\square$

Acknowledgements

References

[1] D. Angluin and D. Chen. Learning a random dfa from uniform strings and state information. In *International Conference on Algorithmic Learning Theory (ALT)*, 2015.

[2] D. Angluin, L. Becerra-Bonache, A. H. Dediu, and L. Reyzin. Learning finite automata using label queries. *International Conference on Algorithmic Learning Theory (ALT)*, 2009.

[3] D. Angluin, D. Eisenstat, L. A. Kontorovich, and L. Reyzin. Lower bounds on learning random structures with statistical queries. *International Conference on Algorithmic Learning Theory (ALT)*, 2010.

[4] K. B. Athreya and P. E. Ney. *Branching Processes*. Grundlehren Der Mathematischen Wissenschaften. Springer-Verlag, Berlin Heidelberg, 1972.

[5] B. Balle. Ergodicity of random walks on random DFA. *arXiv:1311.6830*, 2013.

[6] F. Bassino, J. David, and C. Nicaud. Average case analysis of Moore's state minimization algorithm. *Algorithmica*, 2012.

[7] D. Berend and A. Kontorovich. The state complexity of random dfas. *Theoretical Computer Science*, 652:102–108, 2016.

[8] M. V. Berlinkov. On the probability of being synchronizable. In *Conference on Algorithms and Discrete Applied Mathematics*, pages 73–84. Springer, 2016.

[9] S. Bhamidi and R. Van der Hofstad. Diameter of the stochastic mean-field model of distance. *Combinatorics, Probability and Computing*, 26(6):797–825, 2017.

[10] B. Bollobás and W. Fernandez de la Vega. The diameter of random regular graphs. *Combinatorica*, 2(2):125–134, 1982.

[11] X. S. Cai and L. Devroye. The graph structure of a deterministic automaton chosen at random. *Random Structures & Algorithms*, 51(3):428–458, 2017.

[12] A. Carayol and C. Nicaud. Distribution of the number of accessible states in a random deterministic automaton. *Symposium on Theoretical Aspects of Computer Science (STACS)*, 2012.

[13] P. Chalermsook, B. Laekhanukit, and D. Nanongkai. Pre-reduction graph products: Hardnesses of properly learning dfas and approximating edp on dags. In *Foundations of Computer Science (FOCS)*, 2014.

[14] C. Cooper and A. Frieze. Stationary distribution and cover time of random walks on random digraphs. *Journal of Combinatorial Theory, Series B*, 102(2):329–362, 2012.

[15] S. De Felice and C. Nicaud. Brzozowski algorithm is generically super-polynomial for deterministic automata. *International Conference on Developments in Language Theory (DLT)*, 2013.

[16] C. De la Higuera. *Grammatical inference: learning automata and grammars*. Cambridge University Press, 2010.

[17] T. I. Fenner and A. M. Frieze. On the connectivity of random m-orientable graphs and digraphs. *Combinatorica*, 2(4):347–359, 1982.

[18] A. Frieze and M. Karoński. *Introduction to random graphs*. Cambridge University Press, 2015.

[19] A. A. Grusho. Limit distributions of certain characteristics of random automaton graphs. *Mathematical Notes of the Academy of Sciences of the USSR*, 14(1):633–637, 1973.

[20] J. C. Jackson and R. A. Servedio. Learning random log-depth decision trees under uniform distribution. *SIAM Journal on Computing*, 2005.

[21] S. Janson. One, two and three times log n/n for paths in a complete graph with random weights. *Combinatorics, Probability and Computing*, 8(04):347–361, 1999.

[22] S. Janson, T. Luczak, and A. Rucinski. *Random graphs*, volume 45. John Wiley & Sons, 2011.

[23] M. J. Kearns and L. G. Valiant. Cryptographic limitations on learning boolean formulae and finite automata. *Journal of the ACM*, 1994.

[24] M. J. Kearns and U. V. Vazirani. *An introduction to computational learning theory.* MIT press, 1994.

[25] J.-F. Le Gall. Random real trees. *Ann. Fac. Sci. Toulouse Série 6*, XV:35–62, 2006.

[26] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times.* American Mathematical Soc., 2009.

[27] B. D. McKay. Asymptotics for 0-1 matrices with prescribed line sums. *Enumeration and Design,(Academic Press, 1984)*, pages 225–238, 1984.

[28] C. Nicaud. Random deterministic automata. In *Mathematical Foundations of Computer Science (MFCS)*, 2014.

[29] C. Nicaud. Fast synchronization of random automata. In *APPROX/RANDOM 2016*, volume 60, pages 43–1, 2016.

[30] J. R. Norris. *Markov chains.* Cambridge university press, 1998.

[31] M. Penrose. *Random geometric graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 2003.

[32] L. Pitt and M. K. Warmuth. The minimum consistent DFA problem cannot be approximated within any polynomial. *Journal of the ACM*, 1993.

[33] O. Riordan and N. Wormald. The diameter of sparse random graphs. *Combinatorics, Probability and Computing*, 19(5-6):835–926, 2010.

[34] J. Sakarovitch. *Elements of automata theory.* Cambridge University Press, 2009.

[35] L. Sellie. Exact learning of random dnf over the uniform distribution. *ACM Symposium on Theory of Computing (STOC)*, 2009.

[36] E. Skvortsov and Y. Zaks. Synchronizing random automata. *Discrete Mathematics & Theoretical Computer Science*, 2010.

[37] B. A. Trakhtenbrot and Y. M. Barzdin. *Finite automata : behavior and synthesis.* North-Holland, 1973.

[38] L. G. Valiant. A theory of the learnable. *Communications of the ACM*, 1984.