Submitted to the Annals of Applied Probability

DIFFUSION LIMITED AGGREGATION ON THE BOOLEAN LATTICE

By Alan Frieze* Wesley Pegden†

Carnegie Mellon University

In the Diffusion Limited Aggregation (DLA) process on $\mathbb{Z}^2$, or more generally $\mathbb{Z}^d$, particles aggregate to an initially occupied origin by arrivals on a random walk. The scaling limit of the result, empirically, is a fractal with dimension strictly less than d. Very little has been shown rigorously about the process, however.

We study an analogous process on the Boolean lattice $\{0,1\}^n$, in which particles take random decreasing walks from $(1,\dots,1)$, and stick at the last vertex before they encounter an occupied site for the first time; the vertex $(0,\dots,0)$ is initially occupied. In this model, we can rigorously prove that lower levels of the lattice become full, and that the process ends by producing an isolated path of unbounded length reaching $(1,\dots,1)$.

1. Introduction. In the classical model of Diffusion Limited Aggregation (DLA), we begin with a single particle cluster placed at the origin of our space, and then, one-at-a-time, let particles take random walks "from infinity" until they collide with, and then stick to, the existing cluster; when the space is not recurrent, some care is required to make this precise.

Introduced by Witten and Sander in 1981 [10], the process is particularly natural in Euclidean space (with particles taking Brownian motions) or on d-dimensional lattices; in these cases, the process is empirically observed to produce structures with fractal dimensions strictly less than the dimension of the space (e.g., roughly 1.7 for $d = 2$, with slight but seemingly nonnegligible dependence on details such as the choice of underlying lattice or the precise "sticking" condition).

Strikingly little has been proved rigorously about the model, however. Kesten [8] proved an a.s. asymptotic upper bound of $Cn^{2/3}$ on the radius of the n-particle cluster for the lattice $\mathbb{Z}^2$, for example, but no nontrivial lower bounds are known. In particular, it is not even known rigorously that the process does not have a scaling limit with positive density. (Eldan showed that an analogous process in the hyperbolic plane *does* aggregate to positive density [6]). Eberz-Wagner showed at least that the process leaves infinitely many holes [5]. For some more recent results, see Benjamini and Yadin [1].

In this paper, we study an analogous aggregation process on the Boolean lattice $\mathcal{B} = \{0,1\}^n$, which evolves at discrete times $t = 0, 1, \dots$, each of which has an associated cluster C_t. C_0 consists of just the vertex $\mathbf{0} = (0,\dots,0) \in \mathcal{B}$. Then, for $t > 0$, C_t is produced from C_{t-1} by choosing a random decreasing walk ρ_t from $\mathbf{1} = (1,\dots,1)$, letting v be the last vertex of the longest initial segment

*Research supported in part by NSF grant CCF1013110. **Email:** alan@random.math.cmu.edu.
†Research supported in part by NSF grant DMS1363136. **Email:** wes@math.cmu.edu.
MSC 2010 subject classifications: Primary 60C05; Secondary 82C24

of ρ_t which is disjoint from C_{t-1}, and setting $C_t = C_{t-1} \cup \{v\}$. The process terminates at the first time t_{end} when $C_{t_{\text{end}}} \ni \mathbf{1}$.

In particular, the clusters C_t grow from $\mathbf{0}$ by aggregation of decreasing random walks from $\mathbf{1}$. Our initial motivation for considering this model was to evaluate the impact of very large dimensionality on a DLA-like process. (An analogous motivation underlies work on percolation in the Boolean lattice; see for example [3, 7, 9].) We will see, however, that the Boolean lattice also allows strong rigorous (and perhaps, surprising) statements to be made about the structure of the aggregate. In particular, let $\mathcal{L}_k = \{x \in \mathcal{B} \mid |x| = k\}$ denote the kth level of $\mathcal{B}$, so that $|\mathcal{L}_k| = \binom{n}{k}$. We will prove the following.

THEOREM 1.1. *There exists $c_0 > 0$ such that w.h.p.*[1]*, for all*

$$k \le k_0 := \sqrt{\tfrac{c_0 n}{\log n}},$$

we have

$$\mathcal{L}_k \subset C_{t_{\text{end}}}.$$

THEOREM 1.2. *For any $K > 0$, we have w.h.p. that for all $k \le \frac{n}{150K \log n}$, we have*

$$|\mathcal{L}_k \cap C_{t_{\text{end}}}| \ge \left(1 - \tfrac{1}{n^K}\right) \cdot |\mathcal{L}_k|.$$

THEOREM 1.3. *For all $\varepsilon \le \frac{1}{100}$, we have w.h.p. that for all $k < \varepsilon^3 n$,*

$$|\mathcal{L}_k \cap C_{t_{\text{end}}}| \ge (1 - 2\varepsilon) \cdot |\mathcal{L}_k|.$$

THEOREM 1.4. *There is a constant c_1 such that for all $\varepsilon < \frac{1}{100}$, we have w.h.p. that for all*

$$k_1 = \sqrt{\tfrac{c_1 n}{\log n}} \le k < \varepsilon^3 n,$$

we have

$$|\mathcal{L}_k \setminus C_{t_{\text{end}}}| \ge \left(\frac{(1-\rho)e^{-\frac{10n}{9k}}}{2}\right) \cdot |\mathcal{L}_k|,$$

where

$$\rho = \max\left(1 - \left(\frac{k+1}{en}\right)^{100}, 1 - \left(\frac{1}{10e}\right)^{10}\right).$$

Thus Theorems 1.1, 1.2, and 1.3 provide progressively weaker statements as k increases about the fullness of the level $\mathcal{L}_k$ at the end of the process; Theorem 1.4 shows that Theorems 1.1, 1.2 and 1.3 are qualitatively best-possible. A key contrast from classical DLA is that the process does "fill" parts of the cube, and moreover, that this can be proved. Note also that the boundary between full and not full levels occurs w.h.p. at around $k = \sqrt{\frac{n}{\log n}}$.

A striking (unproved) feature of the classical DLA processes is a rich-get-richer phenomenon, where long arms of the process seem to grow at a rate significantly faster than $t^{1/d}$. In the Boolean lattice, we observe an extreme version of this kind of runaway growth:

[1] A sequence of events $\mathcal{E}_n, n \ge 0$ occurs with high probability (w.h.p.) if $\lim_{n\to\infty} \mathbf{Pr}(\mathcal{E}_n) = 1$.

THEOREM 1.5. *If $a < \frac{1}{2}$ then w.h.p. for all $k \geq n - n^a$ we have that*

$$|\mathcal{L}_k \cap C_{t_{\text{end}}}| = 1.$$

Recall that our DLA process on $\mathcal{B}$ ends once $\mathbf{1}$ is occupied; Theorem 1.5 implies that $\mathbf{1}$ becomes occupied as the terminal vertex on an isolated path of occupied vertices whose length is at least n^a.

Notation. In what follows we use the notation $A_n \approx B_n$ to mean that $A_n = (1 + o(1))B_n$ as $n \to \infty$ and $A_n \lesssim B_n$ to mean that $A_n \leq (1+o(1))B_n$ as $n \to \infty$; we write $A_n \approx_b B_n$ to mean that A_n/B_n is bounded above and below by positive absolute constants as $n \to \infty$. In some places we give expressions for integer quantities that may not be integer; in cases where we do this, it does not matter whether we round up or down.

2. Lower levels. In this section, we prove Theorems 1.1, 1.2, 1.3 and 1.4. We define

$$\tau_{k,\varepsilon} = \frac{\varepsilon}{4}\binom{n}{k+1} = \omega_{k,\varepsilon}\binom{n}{k} \quad \text{for} \quad \omega_{k,\varepsilon} = \frac{\varepsilon}{4}\frac{n-k}{k+1}. \tag{1}$$

Roughly speaking, we expect that at time $\tau_{k,\varepsilon}$, the level $\mathcal{L}_k$ is mostly full, while higher levels are empty enough to have little effect on the process at this time. We will prove a sequence of lemmas confirming this general picture. First, we establish an upper bound on the height of the cluster at a time $\tau_{k,\varepsilon}$:

LEMMA 2.1. *Let $\phi = 1 + \sqrt{3}$ and $0 < \varepsilon < 1$. If $k < \frac{n}{(1+\phi)e^{\phi}}$, then, for all $\delta > 0$, we have with probability $1 - o(n^{-1})$ that*

$$\mathcal{L}_j \cap C_{\tau_{k,\varepsilon}} = \varnothing \text{ for all } j \geq (1 + \phi + \delta)k.$$

PROOF. Consider a fixed vertex v in $\mathcal{L}_{k+s}$. If it becomes occupied by time $\tau_{k,\varepsilon}$, then there is a sequence of times $t_{k+1} < t_{k+2} < \cdots < t_{k+s} \leq \tau_{k,\varepsilon}$ such that

$$\rho_{t_s} \cap \mathcal{L}_{k+s} = \{v\} \text{ and } \rho_{t_i} \cap \mathcal{L}_{k+i-1} = \rho_{t_{i-1}} \cap \mathcal{L}_{k+i-1} \tag{2}$$

for $i = 1, \ldots, s$.

By considering the $\binom{\tau_{k,\varepsilon}}{s}$ possible choices of the times $t_{k+1} < \cdots < t_{k+s} \leq \tau_{k,\varepsilon}$, the probability that each ρ_{t_i} satisfies the intersection conditions (2) for $i = s, s-1, \ldots,$ we have that

$$\begin{aligned}\mathbf{Pr}(v \text{ is occupied at time } \tau_{k,\varepsilon}) &\leq \binom{\omega_{k,\varepsilon}\binom{n}{k}}{s}\prod_{i=1}^{s}\frac{1}{\binom{n}{k+i}} \\ &\leq \left(\frac{\omega_{k,\varepsilon}\left(\frac{ne}{k}\right)^k e}{s}\right)^s \prod_{i=1}^{s}\frac{(k+i)^{k+i}}{n^{k+i}}\end{aligned}$$

$$
\begin{aligned}
&= \omega_{k,\varepsilon}^s n^{sk-sk-s(s+1)/2} \left(\frac{e^{k+1}}{k^k s}\right)^s \prod_{i=1}^{s} (k+i)^{k+i} \\
&\le \omega_{k,\varepsilon}^s n^{-s(s+1)/2} \left(\frac{e^{k+1}}{k^k s}\right)^s (k+s)^{s(k+(s+1)/2)} \\
&= \left(\frac{\omega_{k,\varepsilon}}{s}\right)^s n^{-s(s+1)/2} e^{(k+1)s} \left(1+\frac{s}{k}\right)^{ks} (k+s)^{s(s+1)/2} \\
&\le \left(\frac{\omega_{k,\varepsilon}}{s}\right)^s \left(\frac{(k+s)^{1/2} \cdot e^{1+k/(s+1)}}{n^{1/2}}\right)^{s(s+1)}.
\end{aligned}
$$

So, multiplying by $\binom{n}{k+s}$ we see that

$$
\begin{aligned}
&\mathbf{Pr}(\exists v \in \mathcal{L}_{k+s} : v \text{ is occupied at time } \tau_{k,\varepsilon}) \\
&\le \binom{n}{k+s} \left(\frac{\omega_{k,\varepsilon}}{s}\right)^s \left(\frac{(k+s)^{1/2} \cdot e^{k/(s+1)+1}}{n^{1/2}}\right)^{s(s+1)} \qquad (3) \\
&\le \left(\frac{\omega_{k,\varepsilon}}{s} \left(\frac{k+s}{n}\right)^{(s+1)/2-1-k/s} e^{2+k/s+k+s}\right)^s.
\end{aligned}
$$

Suppose now that $k = \alpha n, s = \beta n \ge 1$. Then the above expression becomes

$$
\left((\alpha+\beta)^{\beta/2} e^{\alpha+\beta} \left(\omega_{k,\varepsilon} s^{-1} n^{1/n} (\alpha+\beta)^{-1/2-\alpha/\beta} e^{2+\alpha/\beta}\right)^{1/n}\right)^{\beta n^2}.
$$

We insist that $\beta \ge \alpha$, in which case

$$
\left(\omega_{k,\varepsilon} s^{-1} n^{1/n} (\alpha+\beta)^{-1/2-\alpha/\beta} e^{2+\alpha/\beta+\alpha+\beta}\right)^{1/n} \le 1 + o(1),
$$

which implies that the expression in (3) is $o(n^{-1})$ so long as $(\alpha+\beta)^{\beta/2} e^{\alpha+\beta} < 1$.

Let $\beta = \gamma\alpha$ ($\gamma \ge 1$). Then our requirement is that $(\alpha(1+\gamma))^{\gamma/2} e^{1+\gamma} < 1$ or $\alpha < \frac{1}{e^{2+2/\gamma}(1+\gamma)}$. Now $e^{2+2/\gamma}(1+\gamma)$ is minimized at the solution to $\gamma^2 = 2(\gamma+1)$, which is $\phi = 1 + 3^{1/2}$. In summary, if $\alpha < \frac{1}{(1+\phi)e^{\phi}}$ then with probability $1 - o(n^{-1})$ all levels above $\alpha(1+\phi+o(1))n$ are empty at time $\omega_{k,\varepsilon}\binom{n}{\alpha n}$, which gives the Lemma. □

Now we define $\Phi_{v,t}$ to be the fraction of (monotone) paths between $\mathbf{1}$ and v which have at least one occupied vertex other than v at time t. The following Lemma implies that levels above $\mathcal{L}_k$ play a small role when analyzing level $\mathcal{L}_k$ at time $\tau_{k,\varepsilon}$.

LEMMA 2.2. *For all fixed $\varepsilon > 0$ and all $\alpha = \frac{k}{n} < \min(\frac{\varepsilon^2}{4}, \frac{1}{100})$, we have*

$$
\mathbf{Pr}\left(\exists v \in \mathcal{L}_k \text{ s.t. } \Phi_{v,\tau_{k,\varepsilon}} \ge \varepsilon\right) = o\left(\tfrac{1}{n}\right). \qquad (4)
$$

PROOF. Recall that the particle in $C_t \setminus C_{t-1}$ is deposited by the decreasing walk ρ_t. We fix a vertex $v \in \mathcal{L}_k$, choose some λ such that $k + \lambda \le \frac{n}{2}$, and define, for each $t = 1, \ldots, \tau_{k,\varepsilon}$, a random variable

$\xi_{v,t} \in [0,1]$ equal to the fraction of paths between v and $\mathcal{L}_{k+\lambda+1}$ whose interiors intersect the path ρ_t. Let $\mathcal{U}_v$ denote the set of all 2^{n-k} ancestors $x > v$ of v. Note that $\xi_{v,t}$ is determined by the minimum $\zeta \geq 1$ such that ρ_t visits $\mathcal{L}_{k+\zeta} \cap \mathcal{U}_v$, and, with respect to this random variable ζ, can be bounded by

$$\xi_{v,t} \leq \xi(\zeta) := \begin{cases} \displaystyle\sum_{i=\zeta}^{\lambda} \frac{1}{\binom{n-k}{i}} \leq \frac{3}{2\binom{n-k}{\zeta}} & \zeta \leq \lambda \\ 0 & \zeta > \lambda. \end{cases} \tag{5}$$

Moreover, we have for $s \leq k+\lambda$ that

$$\mathbf{Pr}(\zeta = s) = \frac{\binom{n-k}{s}}{\binom{n}{k+s}} \cdot \frac{k}{k+s}, \tag{6}$$

since this is the probability that ρ_t visits $\mathcal{L}_{k+s} \cap \mathcal{U}_v$, and then on the next step, moves outside of $\mathcal{U}_v$. In particular, we have that

$$\begin{aligned} \mathbf{E}(\xi(\zeta)) \leq \frac{3}{2} \sum_{s=1}^{\lambda} \frac{\binom{n-k}{s}}{\binom{n}{k+s}} \frac{k}{k+s} \frac{1}{\binom{n-k}{s}} \leq \frac{3}{2} \sum_{s=1}^{\lambda} \frac{1}{\binom{n}{k+s}} \\ = \frac{3}{2} \left(\frac{1 + \sum_{s=2}^{\lambda} \frac{(k+2)\cdots(k+s)}{(n-k-1)\cdots(n-k-s+1)}}{\binom{n}{k+1}} \right) \leq \frac{2}{\binom{n}{k+1}}, \end{aligned} \tag{7}$$

for $k + \lambda < \frac{1}{2}(n-k-\lambda+1)$, or $k+\lambda < \frac{n+1}{3}$.

We will use the following concentration inequality for nonnegative and bounded independent random variables; we show in Appendix A that this is an easy consequence of Bernstein's inequality.

LEMMA 2.3. *Let $X_1, \ldots, X_N$ be independent random variables such that, for all i, $\mathbf{E}(X_i) \leq E$ and $X_i \in [0, C]$ almost surely. Then for $S_N = \sum_{i=1}^{N} X_i$, $E_N = \mathbf{E}(S_N)$, and for all $t \leq NE$, we have that*

$$\mathbf{Pr}\left(|S_n - E_n| > t\right) < 2\exp\left(-\frac{t^2/4}{NEC}\right).$$

Note that in the same situation, Hoeffding's inequality gives $2e^{-2t^2/NC^2}$, which, ignoring constant factors in the exponent, is always worse; the point is that we are interested in the case where $E \ll C$. And though we have stated the lemma here with the condition $t \leq NE$, one could drop it and still obtain the bound $2\exp\left(-\frac{t^2/2}{NEC+Ct}\right)$, an analogous improvement over Hoeffding anytime $t \ll NC$.

To apply Lemma 2.3, notice that from (5) that $\xi(s) \leq \frac{3}{2(n-k)}$ always. If $\Xi_{v,T} = \sum_{t=1}^{T} \xi_{v,t}$ then $\Xi_{v,T}$ is stochastically dominated by a sum $Z_T = \xi(\zeta_1) + \xi(\zeta_2) + \cdots + \xi(\zeta_T)$ where each ζ_j is an independent copy of a random variable ζ satisfying (6). Now (7) implies that $T\mathbf{E}(\xi(\zeta_i)) = \mathbf{E}(Z_T) \leq \frac{2T}{\binom{n}{k+1}}$, and thus Lemma 2.3 with $t = \theta, N = \tau_{k,\varepsilon}, E \leq \frac{2}{\binom{n}{k+1}}, C = \frac{3}{2(n-k)}$ gives that

$$\mathbf{Pr}\left(\Xi_{v,\tau_{k,\varepsilon}} \geq \frac{2\tau_{k,\varepsilon}}{\binom{n}{k+1}} + \theta\right) \leq \mathbf{Pr}\left(Z_{\tau_{k,\varepsilon}} \geq \frac{2\tau_{k,\varepsilon}}{\binom{n}{k+1}} + \theta\right) \leq \exp\left(-\frac{\theta^2/4}{\frac{2\tau_{k,\varepsilon}}{\binom{n}{k+1}} \cdot \frac{3}{2(n-k)}}\right) = e^{-\theta^2(n-k)/(3\varepsilon)} = e^{-\theta(n-k)/6},$$

for

$$\theta = \varepsilon - \frac{2\tau_{k,\varepsilon}}{\binom{n}{k+1}} = \frac{\varepsilon}{2} \qquad (\text{ from (1)}).$$

Now we have that

$$\mathbf{Pr}\left(\exists v \in \mathcal{L}_k,\, \Xi_{v,\tau_{k,\varepsilon}} \geq \frac{2\tau_{k,\varepsilon}}{\binom{n}{k+1}} + \theta\right) \leq \left(\frac{en}{k}\right)^k e^{-\theta(n-k)/6}.$$

Writing $k = \alpha n$, we have that

$$\left(\frac{en}{k}\right)^k e^{-\theta(n-k)/6} = e^{\ln(e/\alpha)\alpha n - \theta n(1-\alpha)/6} = o\left(\tfrac{1}{n}\right),$$

for any $\alpha < \min(\theta^2, \frac{1}{100})$, say. Thus for $\alpha < \min(\frac{\varepsilon^2}{4}, \frac{1}{100})$, we have that

$$\mathbf{Pr}\left(\exists v \in \mathcal{L}_k,\, \Xi_{v,\tau_{k,\varepsilon}} \geq \varepsilon\right) = o\left(\tfrac{1}{n}\right).$$

Now, by taking $\lambda > 2\phi k$, we may assume that the levels above level λ are still empty at time $\tau_{k,\varepsilon}$, so that $\Phi_{v,\tau_{k,\varepsilon}} \leq \Xi_{v,\tau_{k,\varepsilon}}$, completing the proof of the Lemma. □

Now we define $\Upsilon_{v,t}$ for $v \in \mathcal{L}_k$ to be the fraction of down-neighbors of v which are unoccupied at time t. By controlling $\Upsilon_{v,t}$ and $\Xi_{v,t}$ simultaneously, we can make the behavior of the cluster with respect to v sufficiently predictable.

LEMMA 2.4. *Suppose that $0 < \varepsilon \leq \frac{1}{100}$ and $k \leq \varepsilon^3 n$ is fixed. Then,*

$$\mathbf{Pr}\left(\exists t \in [\tau_{k-1,\varepsilon}, \tau_{k,\varepsilon}],\, v \in \mathcal{L}_k,\, \Phi_{v,t} + \Upsilon_{v,t} \geq 2\varepsilon\right) = o\left(\tfrac{1}{n}\right). \tag{8}$$

Lemma 2.4 will be proved by induction on k. Before giving the proof, we use it to prove Theorems 1.1, 1.2, and 1.3.

PROOF OF THEOREM 1.1. Let Λ_k denote the set of vertices in $\mathcal{L}_k$ which are still unoccupied by particles at time $\tau_{k,\varepsilon}$. We fix $\varepsilon = \frac{1}{100}$ and apply Lemma 2.4. Since $\tau_{k,\varepsilon} - \tau_{k-1,\varepsilon} > \frac{1}{2}\tau_{k,\varepsilon}$, we have for any vertex $v \in \mathcal{L}_k$ that if $\Psi = \max\{\Phi_{v,t} + \Upsilon_{v,t} : t \in [\tau_{k-1,\varepsilon}, \tau_{k,\varepsilon}]\}$,

$$\mathbf{Pr}\left(v \text{ not occupied by } \tau_{k,\varepsilon} \mid \Psi \leq 2\varepsilon\right) \lesssim \left(1 - \frac{1-2\varepsilon}{\binom{n}{k}}\right)^{\frac{1}{2}\omega_{k,\varepsilon}\binom{n}{k}} \leq e^{-\frac{1}{3}\omega_{k,\varepsilon}}.$$

Explanation: For a fixed time t for which $\Phi_{v,t} + \Upsilon_{v,t} \leq 2\varepsilon$, the term $\frac{1}{\binom{n}{k}} \cdot (1 - 2\varepsilon)$ is a lower bound on the probability that ρ_t chooses to go through v on level k, avoids occupied vertices on

the way to v and then chooses an occupied vertex in level $k-1$. Conditioning on $\Psi \le 2\varepsilon$ (i.e., on the 2ε condition for all t's simultaneously) inflates these probabilities by at most a factor of $\mathbf{Pr}(\Phi \le 2\varepsilon)^{-1} = 1 + o(1)$.

Thus

$$\mathbf{E}\left(|\Lambda_k| \mid \Psi \le 2\varepsilon\right) \lesssim \binom{n}{k} e^{-\omega_{k,\varepsilon}/3}$$

and so by the Markov inequality,

$$\mathbf{Pr}\left(\Lambda_k \ne \varnothing \mid \Psi \le 2\varepsilon\right) \le \binom{n}{k} e^{-\omega_{k,\varepsilon}/4} = o(n^{-1})$$

as long as

$$\omega_{k,\varepsilon} = \frac{\varepsilon}{4}\frac{n-k}{k+1} \ge 5k\log(ne/k).$$

In particular, this holds for

$$k \le \sqrt{\frac{\varepsilon n}{10\log n}}$$

and gives the desired statement (recalling that $\varepsilon = 1/100$). □

Proof of Theorem 1.2. Again by the Markov inequality applied to $|\Lambda_k|$, we have

$$\mathbf{Pr}\left(|\Lambda_k| \ge \beta\binom{n}{k} \;\middle|\; \Psi \le 2\varepsilon\right) \lesssim \frac{e^{-\omega_{k,\varepsilon}/3}}{\beta}.$$

This is $o(n^{-1})$, assuming that $\varepsilon = \frac{1}{100}$ and

$$k \le \frac{n}{150K\log n}, \quad \beta = \frac{1}{n^K},$$

for any constant $K > 0$, giving the theorem. □

Theorem 1.3 is a consequence of the following slightly stronger statement:

Lemma 2.5. *For all $\varepsilon > 0$, we have w.h.p. that for all $k \le \varepsilon^3 n$, $\mathcal{L}_k$ is at least $(1-2\varepsilon)$ occupied at time $\tau_{k,\varepsilon}$.*

Proof. This follows directly from Lemma 2.4. Indeed, if there are ℓ occupied sites in $\mathcal{L}_k$ at time $\tau_{k,\varepsilon}$, and m edges between $\mathcal{L}_{k+1}$ and occupied sites in $\mathcal{L}_k$, then assuming that $\Phi_{v,t} + \Upsilon_{v,t} \le 2\varepsilon$ for $v \in \mathcal{L}_{k+1}$, (from (8)), the degrees of vertices in $\mathcal{L}_k, \mathcal{L}_{k+1}$ gives that with probability $1 - o(n^{-1})$,

$$\binom{n}{k+1} \cdot (1-2\varepsilon)(k+1) \le m \le \ell(n-k),$$

so that $\ell \ge (1-2\varepsilon)\binom{n}{k}$. □

We now prove Lemma 2.4, by induction on k.

PROOF OF LEMMA 2.4. In particular, assuming that (8) holds for some k, we aim to prove that if $0 < \varepsilon \le \frac{1}{100}$ and $k \le \varepsilon^3 n$ then

$$\mathbf{Pr}\left(\exists t \in [\tau_{k-1,\varepsilon}, \tau_{k,\varepsilon}],\, w \in \mathcal{L}_k,\, \Phi_{v,t} + \Upsilon_{v,t} \ge 2\varepsilon\right) \le \frac{2k}{n^3}. \tag{9}$$

Observe that since $\Phi_{v,t}$ is increasing in t and $\Upsilon_{v,t}$ is decreasing in t, (9) can be proved by showing

$$0 < \varepsilon \le \frac{1}{100}, k \le \varepsilon^3 n \text{ implies that } \mathbf{Pr}\left(\exists w \in \mathcal{L}_k,\, \Upsilon_{w,\tau_{k,\varepsilon}} \ge \varepsilon\right) \le \frac{k}{n^3}, \tag{10}$$

and

$$0 < \varepsilon \le \frac{1}{100}, k \le \varepsilon^3 n \text{ implies that } \mathbf{Pr}\left(\exists w \in \mathcal{L}_k,\, \Phi_{w,\tau_{k,\varepsilon}} \ge \varepsilon\right) \le \frac{k}{n^3}. \tag{11}$$

Of course, (11) follows from (4), so we just need to show (10). For the sake of conditioning in the induction, define the event

$$\mathcal{A}_{k,\varepsilon,n} = \left(\forall t \in [\tau_{k-1,\varepsilon}, \tau_{k,\varepsilon}],\, w \in \mathcal{L}_k,\, \Phi_{v,t} + \Upsilon_{v,t} < 2\varepsilon\right),$$

so that we are aiming to prove inductively that

$$\mathbf{Pr}(\mathcal{A}_{k,\varepsilon,n}) \ge 1 - \frac{k}{n^3}.$$

As a base case we take $k = 1$ which trivially satisfies (9). Assume $k \ge 1$ and fix some vertex $w \in \mathcal{L}_k$, and let $N_w^- \subset \mathcal{L}_{k-1}$ be the down-neighborhood of w. If we fix a set $D \subset N_w^-$ of size $|D| = \Delta$, then we have, since $k \le \varepsilon^3 n$, that

$$\mathbf{Pr}\left(D \cap C_{\tau_{k,\varepsilon}} = \varnothing \mid \mathcal{A}_{k-1,\varepsilon,n}\right) \le \left(1 - \frac{4\Delta}{5\binom{n}{k-1}}\right)^{\tau_{k,\varepsilon} - \tau_{k-1,\varepsilon}} \le e^{-\frac{\Delta\omega_{k,\varepsilon}}{2}}. \tag{12}$$

Explanation: The first inequality arises because each path ρ_t for $t \in [\tau_{k-1,\frac{1}{10}} + 1, \tau_{k,\varepsilon}]$ has probability $\frac{\Delta}{\binom{n}{k-1}}$ of intersecting D, and conditioned on that event, applying (8) inductively with $\varepsilon = \frac{1}{100}$ ensures that with probability at least $\frac{4}{5}$, a particle will occupy at least one site v of D after step t for $\tau_{k-1,\varepsilon} < t \le \tau_{k,\varepsilon}$ (either because v was already occupied before step t, or because ρ_t deposits a particle at v.) The second inequality arises because $k \le \varepsilon^3 n$ implies that $\tau_{k,\varepsilon} - \tau_{k-1,\varepsilon} \ge \frac{\omega_{k,\varepsilon}}{2}\binom{n}{k}$.

Thus we have that

$$\begin{aligned}
&\mathbf{Pr}\left(\exists w \in \mathcal{L}_k, D \subset N_w^-, \text{ s.t. } |D| = \Delta, D \cap C_{\tau_{k,\varepsilon}} = \varnothing\right) \\
&\le \binom{n}{k}\binom{k}{\Delta} e^{-\Delta\omega_{k,\varepsilon}/2} \le \left(\frac{ne}{k}\right)^k \left(\frac{ke}{\Delta}\right)^\Delta e^{-\Delta\omega_{k,\varepsilon}/2} \\
&= e^{k\log(ne/k) + \Delta\log(ke/\Delta) - \Delta\omega_{k,\varepsilon}/2} \le \frac{1}{n^3},
\end{aligned}$$

if either $k \le n^{1/2}$ or (ii) $k > n^{1/2}$ and $\Delta \le k$ and $\Delta\omega_{k,\varepsilon} > 4\max(k\log(ne/k), \Delta\log(ke/\Delta))$.

For (ii), recalling that $\omega_{k,\varepsilon} = \frac{\varepsilon}{4} \cdot \frac{n-k}{k+1}$, we can take

$$\Delta(n-k) > \tfrac{16}{\varepsilon} \cdot k(k+1)\log(ne/k),$$

and

$$n-k > \tfrac{4}{\varepsilon} \cdot k\log(ke/\Delta),$$

which, for $\frac{\Delta}{k} \geq \varepsilon$, would follow from

$$\frac{20 \cdot k\log(n/k)}{n-k} < \varepsilon^2, \tag{13}$$

and

$$e^{1-\varepsilon\frac{n-k}{4k}} < \varepsilon, \tag{14}$$

respectively. Both (13) and (14) are satisfied when $k < \varepsilon^3 n$ and n is large. □

Lemma 2.4 is not quite strong enough to prove Theorem 1.4. For that purpose, we prove the following Lemma, which allows stronger statements when k is linear in n:

LEMMA 2.6. *Suppose that $k \geq k_0$. Let t_k^ρ be the first time when a ρ fraction of the vertices in $\mathcal{L}_k$ are occupied. We have that*

$$\mathbf{Pr}\left(\exists w \in \mathcal{L}_{k+1},\, \Upsilon_{w,t_k^\rho} \geq \tfrac{1}{10}\right) = o\left(\tfrac{1}{n}\right),$$

provided that $\varepsilon = \frac{1}{100}, \alpha = \frac{k}{n} \leq \varepsilon^3$ and

$$\rho = 1 - \left(\frac{k+1}{en}\right)^{100}. \tag{15}$$

PROOF. For any constant K and sufficiently large n, we have that

$$\mathbf{Pr}\left(t_k^\rho > \frac{1}{2}\binom{n}{k}\log\left(\frac{1}{1-\rho}\right)\right) > 1 - \frac{1}{n^K}.$$

This is because the number of vertices in $\mathcal{L}_k$ that are occupied at time t is dominated by the number of occupied bins when t balls are placed randomly into $\binom{n}{k}$ bins. Note that the expected number of occupied bins in the latter experiment is

$$\binom{n}{k}\left(1 - \left(1 - \frac{1}{\binom{n}{k}}\right)^t\right) \approx (1-(1-\rho)^{1/2})\binom{n}{k} \leq \frac{\rho}{2}\binom{n}{k} \text{ when } t = \frac{1}{2}\binom{n}{k}\log\left(\frac{1}{1-\rho}\right).$$

Note also that the number of occupied boxes is highly concentrated. This can be verified through a simple application of McDiarmid's inequality, see [2].

In particular,

$$t_k^\rho - \tau_{k-1,\varepsilon} \geq H_{\rho,k} := \frac{\log(1/(1-\rho))}{2}\binom{n}{k}$$

with probability at least $\left(1-\frac{1}{n^K}\right)$. This follows from the fact that $\rho \geq 1-e^{-\varepsilon/4}$.

From (15) we also have that

$$H_{\rho,k} = 50\log\left(\frac{en}{k+1}\right)\cdot\binom{n}{k},$$

and so we have $\tau_{k-1,\varepsilon} + H_{\rho,k} \leq \tau_{k,\varepsilon}$ for sufficiently large n; see (1). In particular, we can apply Lemma 2.4 in the entire range $[\tau_{k-1,\varepsilon}, \tau_{k-1,\varepsilon} + H_{\rho,k}]$.

To do this, we fix some vertex $w \in \mathcal{L}_{k+1}$, and let $N_w^- \subset \mathcal{L}_k$ be the down-neighborhood of w. If we fix a set $D \subset N_w^-$ of size $|D| = \Delta$, then we have for $k \leq \varepsilon^3 n$ that

$$\begin{aligned}\mathbf{Pr}\left(D \cap C_{t_k^\rho} = \varnothing \mid t_k^\rho \geq \tau_{k-1,\varepsilon} + H_{\rho,k}\right) &\leq \left(1-\frac{1}{n^K}\right)^{-1}\left(1-\frac{4\Delta}{5\binom{n}{k}}\right)^{H_{\rho,k}} \\ &\leq 2e^{-\frac{\Delta\cdot\log(1/(1-\rho))}{5}} = 2(1-\rho)^{\Delta/5}.\end{aligned}$$

Explanation: We repeat the argument for (12) and multiply by $(1-n^{-K})^{-1}$ to account for conditioning on $t_k^\rho \geq \tau_{k-1,\varepsilon} + H_{\rho,k}$.

Thus we have that

$$\begin{aligned}&\mathbf{Pr}\left(\exists w \in \mathcal{L}_{k+1}, D \subset N_w^-, \text{ s.t. } |D| = \Delta = (k+1)/10, D \cap C_{\tau_{k,\varepsilon}} = \varnothing \mid t_k^\rho \geq \tau_{k-1,\varepsilon} + H_{\rho,k}\right) \\ &\leq \binom{n}{k+1}\binom{k+1}{\Delta} \times 2(1-\rho)^{\frac{\Delta}{5}} \leq 2\left(\frac{ne}{k+1}\right)^{k+1}\left(\frac{(k+1)e}{\Delta}\right)^{\Delta}(1-\rho)^{\Delta/5}. \\ &= 2e^{(k+1)\log(ne/k+1)+\Delta\log((k+1)e/\Delta)+(\Delta/5)\log(1-\rho)} = o\left(\frac{1}{n}\right),\end{aligned}$$

if $\Delta = (k+1)/10$ and $\Delta\log(\frac{1}{1-\rho}) \geq 10\max\{(k+1)\log(ne/(k+1)), \Delta\log((k+1)e/\Delta)\}$.

For these cases, we require that

$$\rho \geq 1 - \left(\frac{k+1}{en}\right)^{100}$$

and

$$\rho \geq 1 - \left(\frac{\Delta}{e(k+1)}\right)^{10} = 1 - \left(\frac{1}{10e}\right)^{10},$$

respectively, both of which follow from our choice of ρ. □

We are now ready to prove Theorem 1.4.

Proof of Theorem 1.4. We apply Lemma 2.6 with $\varepsilon = \frac{1}{100}$ and ρ satisfying (15). Condition on the event $\mathcal{F} = \{\forall v \in \mathcal{L}_k, \Upsilon_{v,t_k^\rho} < \frac{1}{10}\}$. Let $O_k = \mathcal{L}_k \cap C_{t_k^\rho}$ be the set of occupied vertices in $\mathcal{L}_k$, so that $|O_k| = \left\lceil \rho\binom{n}{k}\right\rceil$. Fix a vertex $v \in U_k = \mathcal{L}_k \setminus O_k$ and let N_v^+ be its neighborhood in $\mathcal{L}_{k+1}$. For each $w \in N_v^+$, we define events $\mathcal{E}_w^1$ and $\mathcal{E}_w^2$, respectively by

1. Every path ρ_t, $t > t_k^\rho$ which contains w avoids both O_k and v, and
2. The first path ρ_t, $t > t_k^\rho$ which intersects w and then hits $O_k \cup \{v\}$ hits O_k and not v,

and set $\mathcal{E}_w = \mathcal{E}_w^1 \cup \mathcal{E}_w^2$. Let $\mathcal{A}_v$ be the event that $\mathcal{E}_w$ occurs for each $w \in N_v^+$. Observe that if $\mathcal{A}_v$ occurs then vertex v remains unoccupied on termination. Moreover, if we fix some set $W_0 \subset N_v^+$, then the events $\mathcal{E}_w$ for $w \in W_0$ are conditionally independent, given the event that $W_0 = \mathcal{W} := \{w \in N_v^+ \text{ such that } \neg\mathcal{E}_w^1 \text{ occurs}\}$. (Indeed, given that $\neg\mathcal{E}_w^1$ occurs, we know that at least one path goes through w. Moreover, the event $\mathcal{E}_w^2$ depends on just the first path with this property, and the choice this path makes below w is independent of choices made by paths not going through w.) Now, for any choice of W_0 and any $w \in W_0$, we have

$$\mathbf{Pr}(\mathcal{E}_w \mid \mathcal{F}, \mathcal{W} = W_0) \geq 1 - \frac{1}{\frac{9}{10}k},$$

since $|(O_k \cap N_w^-) \cup \{v\}| \geq \frac{9}{10}k$, and $\mathcal{E}_w^2$ implies that the first path through w choosing among $(O_k \cap N_w^-) \cup \{v\}$ chooses v. Now using the conditional independence of the $\mathcal{E}_w$'s given W_0, we have that

$$\mathbf{Pr}(\mathcal{A}_v \mid \mathcal{F}, \mathcal{W} = W_0) \geq \left(1 - \frac{10}{9k}\right)^{n-k} \geq p := e^{-1/\alpha}.$$

Finally, since this is true for any fixed W_0, we have that

$$\mathbf{Pr}(\mathcal{A}_v \mid \mathcal{F}) \geq \left(1 - \frac{10}{9k}\right)^{n-k} \geq p. \tag{16}$$

It follows from (16) that on termination, conditioning on $\mathcal{F}$, there are in expectation at least $(1-\rho)p\binom{n}{k}$ vertices of U_k that remain unoccupied at the end of the process.

Let now Z_k denote the number of v such that $\mathcal{A}_v$ occurs. Now Z_k is determined by at most $(k+1)\binom{n}{k+1}$ random choices viz. the paths from $\mathbf{1}$ to $\mathcal{L}_{k+1}$ that give rise to a first visit to a vertex of $\mathcal{L}_{k+1}$ that continues on to O_k. More precisely, we partition the paths from $\mathbf{1}$ to $\mathbf{0}$ according to which member of $\mathcal{L}_{k+1}$ they visit. Z_k is determined by an independent choice of a path from each part of the partition followed by a choice of vertex in $\mathcal{L}_k$. Changing one of these choices, changes Z_k by at most one and so applying McDiarmid's inequality we get

$$\mathbf{Pr}\left(Z_k \leq \frac{1-\rho}{2}p\binom{n}{k}\right) \leq \exp\left\{-\frac{(1-\rho)^2p^2\binom{n}{k}^2}{2(k+1)\binom{n}{k+1}}\right\} \leq \exp\left\{-\frac{\left(\frac{k}{en}\right)^{100}e^{-2n/k}\left(\frac{n}{k}\right)^k}{n}\right\}.$$

Now

$$\log\left(\frac{\left(\frac{k}{en}\right)^{100}e^{-2n/k}\left(\frac{n}{k}\right)^k}{n}\right) = (k-100)\log(n/k) - 100 - \log n - \frac{2n}{k} \geq 2\log n$$

if $k_1 \leq k \leq \varepsilon^3 n$. This proves the Theorem. □

3. Long path. In this section we prove Theorem 1.5

3.1. *Setup.* We begin our proof by fixing certain parameters a, b, c. Recall that n^a is the length of the path that we prove exists. n^b will be a bound on the expected value of a level at a certain time, and the exponent c will occur in error bounds in our concentration analysis.

Our proof will require that these parameters satisfy the following constraints:

1. $a < 1 - 2c$. This is needed to ensure that the probability in (44) is $o(n^{-1})$ as claimed.
2. $2c < 1 - a$. This is needed to ensure that the RHS of (38) is $o(\mu_0)$.
3. $a < 2c$. This is needed to ensure that δ_j in (39) to be $o(1)$.
4. $a + b < 1$. This is needed to ensure that the LHS of (48) is o(1).
5. $a < \frac{1}{2}$. This is also needed to ensure that the LHS of (48) is o(1).
6. $a > b$. This is needed in (22).

We choose a as large as possible here. So we take

$$a = \frac{1}{2} - \varepsilon,\ b = \frac{1}{2} - 2\varepsilon,\ c = \frac{1}{4} - \frac{\varepsilon}{3}$$

for some arbitrarily small $\varepsilon > 0$.

We then let

$$\ell = n^a \text{ and } k = n - \ell \tag{17}$$

and assume that ℓ is an integer. We let $\mathcal{O}_{j,t} = C_t \cap \mathcal{L}_j$, the set of occupied vertices on level j at time t.

A considerable difficulty facing our proof of Theorem 1.5 is that we do not understand the "intermediate" behavior of the cube; that is, our Theorems 1.1, 1.2, 1.3, 1.4 lose their bite well below level $\frac{n}{2}$, say. Thus the proof must be agnostic to the behavoir of the process in the middle layers of the cube. One natural idea to handle this would be to to assume a "worst-case" behavior for the intermediate levels of the cube; say, that level k becomes full while levels $k + j$ $(j \geq 1)$ are still empty, and show that even in this scenario, a path of length nearly $n - k$ will still grow, for sufficiently large k. However, the DLA process is not monotone in a clean way, preventing us from arguing directly that having level k full while higher levels are empty is truly a worst-case scenario from the standpoint of the probability that a long isolated path reaches $\mathbf{1}$.

Instead, we proceed by defining a stopping time. We run the DLA process on the empty cube, until time τ_0 when there first exists $0 \leq j^* < j_0$ such that

$$|\mathcal{O}_{k+j^*,\tau_0}| \geq \zeta(j^*, \mu_0).$$

Here μ_0 and j_0 are parameters which will chosen later, and we define

$$\zeta(j, \mu_0) = \frac{\binom{\mu_0}{j_0+1}}{j_0 \binom{\mu_0}{j_0-j} \eta(j)},$$

and

$$\eta(j) = \prod_{s=0}^{j-1} \binom{n}{\ell - s}.$$

μ_0 will be an upper bound estimate for the time after τ_0 when we can expect the process to end, and $j_0 \approx \sqrt{2\ell}$ will be the level from which we show the isolated path with grow. We will see in (22) that $\zeta(j_0 - 1, \mu_0) < 1$, so that the stopping time τ_0 always occurs.

Roughly speaking, by beginning our analysis from this stopping time, we begin from a situation where we have some (carefully chosen) useful bounds on the sizes of levels, which makes an analysis of the remainder of the process possible.

For the purpose of analyzing the growth of the DLA process in expectation, it is useful to allow the process to continue past the point when vertex **1** becomes occupied. To do this, we extend the DLA process past time t_{end} by letting Θ_t be the number of particles stuck "above" **1**. In particular, $\Theta_t = \max\{0, t - t_{\text{end}}\}$, and occupancies of vertices $v \in \mathcal{B}$ at times $t > t_{\text{end}}$ are the same as at time t_{end}.

Now we let $X_{j,t} = |\mathcal{O}_{k+j,\tau_0+t}|$ for $0 \le j, t$ and let $Y_{j,t} = \Theta_t^2 + \sum_{r \ge j^*+j} X_{r,t}$. (It would be natural to replace Θ_t^2 with Θ_t here, but using Θ_t^2—or any fast-enough growing function of Θ_t—ensures that the following recurrence for $Y_{j,t}$ will not be broken by the cases where $t \gg t_{\text{end}}$.) Then we have that for $j \ge 1$,

$$\mathbf{E}(X_{j,t} - X_{j,t-1} \mid X_{j-1,t-1}) \le \frac{X_{j-1,t-1}}{\binom{n}{\ell-j+1}}, \quad t \ge 1. \tag{18}$$

$$\mathbf{E}(Y_{j,t} - Y_{j,t-1} \mid Y_{j-1,t-1}) \ge \frac{Y_{j-1,t-1}}{\sum_{r \ge j^*+j} \binom{n}{\ell-r+1}} \approx \frac{Y_{j-1,t-1}}{\binom{n}{\ell-j^*-j+1}}, \quad t \ge 1. \tag{19}$$

Explanation
The RHS of (18) is the probability that a particle chooses an occupied position on level $k + j$. It is an upper bound for the increase because it does not account for the particle being blocked higher up in the cube.

For the middle term in (19), observe that there are $Y_{j-1,t-1}$ occupied vertices among the $\sum_{r \ge j^*+j} \binom{n}{\ell-r+1}$ vertices at or above level $j - 1$; thus the middle term gives the probability that a randomly chosen vertex from $\rho_t \cap \bigcup_{r \ge j^*+j} \mathcal{L}_{\ell-r+1}$ is occupied, and the occurrence of this event implies that $Y_{j,t}$ increases by one. This explains the first inequality.

Removing the conditioning in (18), (19) we obtain for $j \ge 1$,

$$\mathbf{E}(X_{j,t} - X_{j,t-1}) \le \frac{\mathbf{E}(X_{j-1,t-1})}{\binom{n}{\ell-j+1}}, \quad t \ge 1. \tag{20}$$

$$\mathbf{E}(Y_{j,t} - Y_{j,t-1}) \gtrsim \frac{\mathbf{E}(Y_{j-1,t-1})}{\binom{n}{\ell-j^*-j+1}}, \quad t \ge 1. \tag{21}$$

The recurrences (20), (21) yield upper and lower bounds as on the expectations of $X_{j,t}, Y_{j,t}$, which will be analyzed in Section 3.3.

To prove that a path grows from j_0, we will first show that after

$$\mu_1 := \mu_0 - \omega_1 \binom{n}{\ell - j_0}$$

steps for

$$\omega_1 = \log^2 n,$$

we will have that w.h.p. $Y_{j_0,\tau_0+\mu_1} \gtrsim n^b$ and $X_{j_0-1,\tau_0+\mu_1} \lesssim \omega_1 n^b$. Observe that this implies that for the minimum $t_{\rm fin}$ for which $Y_{j_0,t_{\rm fin}} = 1$, we have $X_{j_0-1,t_{\rm fin}} \lesssim \omega_1 n^b$, and that we have that $|\mathcal{O}_{k+j_0,t_{\rm fin}}| = 1$. In particular, we will prove that the DLA process can quickly produce a path from to $\mathbf{1}$ after $t_{\rm fin}$; that $X_{j_0-1,t}$ does not increase quickly after $t_{\rm fin}$, and that the small value of $X_{j_0-1,t}$ for t near $t_{\rm fin}$ implies that no particles stick at j_0 while the path to $\mathbf{1}$ is being created.

3.2. *Choice of* μ_0, μ_1, j_0. In this section we define μ_0, μ_1, j_0 and compute various quantities associated with them for later use. In particular, we let

$$\xi(j,t) = \frac{\binom{\mu_0}{j_0+1}}{\eta(j)} \cdot \frac{\binom{t}{j}}{\binom{\mu_0}{j_0}} = \zeta(j,\mu_0) \cdot \frac{j_0\binom{\mu_0}{j_0-j}\binom{t}{j}}{\binom{\mu_0}{j_0}}.$$

Note from (17) that

$$\zeta(j_0-1,\mu_0) = \xi(j_0-1,\mu_0)\frac{\mu_0 - j_0 + 1}{j_0^2\mu_0}.$$

This, together with lines (23), (27) and (30), below, will imply then that

$$\zeta(j_0-1,\mu_0) = O\left(\frac{n^b}{n^a}\right) = o(1). \tag{22}$$

Roughly speaking, $\xi(j,t)$ is an approximate target for comparison with $|\mathcal{O}_{k+j,\tau_0+t}|$. In particular, we will choose μ_0, j_0 and prove that

P1: $\mathbf{E}(X_{j_0,\mu_1}) \lesssim \xi(j_0,\mu_0) \approx n^b$ – see (27) and (36).
P2: $\mathbf{E}(X_{j_0-1,\mu_1}) \lesssim \xi(j_0-1,\mu_0) \approx 2e^{\frac{4}{3}} n^b$ – see (37).
P3: $\mathbf{E}(Y_{j_0,\mu_1}) \gtrsim \frac{\xi(j_0,\mu_0)}{j_0}$ – see (41).

We choose

$$\omega = (1-a)\log n$$

and then j_0 by

$$j_0 = \min\left\{j : j(j+3) \geq 2\ell + \frac{4\ell}{\omega - 1}\right\}.$$

Now $j(j+3) - (j-1)(j+2) = 2j+2$ and so we have that

$$2\ell + \frac{4\ell}{\omega-1} \leq j_0(j_0+3) \leq 2\ell + \frac{4\ell}{\omega-1} + 2j_0 + 2.$$

Thus

$$j_0^2 = 2\ell + \frac{4\ell}{\omega-1} + \theta_0, \tag{23}$$

where $|\theta_0| \leq 3j_0$.

Next we prove an asymptotic estimate for $\eta(j)$.

LEMMA 3.1. *If $\ell \gg j$, $n-\ell+j \gg 1$, and $\ell = o(n^{1/2})$, then*

$$\eta(j) \approx \binom{n}{\ell}^{j(\ell-(j-1)/2)} (2\pi\ell)^{-\frac{1}{2}j} e^{\ell j + 3\ell^2 j/2n + \varepsilon_a j - \varepsilon_b j^2},$$

where

$$\varepsilon_a = \frac{j}{2\ell} + O\left(\left(\frac{j}{\ell}\right)^2\right), \quad \varepsilon_b = \frac{j}{6\ell} + \frac{j^2}{24\ell^2} + O\left(\left(\frac{j}{\ell}\right)^3\right).$$

PROOF. We let

$$\phi(\ell) = \prod_{r=1}^{\ell} r!$$

be the superfactorial function. It is known that

$$\phi(\ell) \approx C_1 \ell^{\frac{1}{2}\ell^2+\ell+\frac{5}{12}} e^{-\frac{3}{4}\ell^2-\ell} (2\pi)^{\frac{1}{2}\ell}$$

for some absolute constant $C_1 > 0$. (See, for example, Adamchik [4]. We use the asymptotic expression for the Barnes function $G(z)$ on page 2. Note also that $\phi(\ell) = G(\ell+2)$.)

We need to estimate $\frac{\phi(\ell-x)}{\phi(\ell)}$ where $x = O(\ell^{1/2})$. In preparation we observe that if $x = O(\ell^{1/2})$ then

$$\left(1-\frac{x}{\ell}\right)^{\ell-x} = \exp\left\{-x + \sum_{i=2}^{\infty} \frac{x^i}{i(i-1)\ell^{i-1}}\right\} = e^{-x+\varepsilon_a x},$$

where $\varepsilon_a = \sum_{i=2}^{\infty} \frac{x^{i-1}}{i(i-1)\ell^{i-1}} = \frac{x}{2\ell} + O\left(\left(\frac{x}{\ell}\right)^2\right)$, and

$$\left(1-\frac{x}{\ell}\right)^{\frac{1}{2}(\ell-x)^2} = \exp\left\{-\frac{1}{2}x\ell + \frac{3}{4}x^2 - \sum_{i=3}^{\infty} \frac{x^i}{i(i-1(i-2))\ell^{i-2}}\right\} = e^{-\frac{1}{2}x\ell+\frac{3}{4}x^2-\varepsilon_b x^2},$$

where $\varepsilon_b = \sum_{i=3}^{\infty} \frac{x^{i-2}}{i(i-1(i-2))\ell^{i-2}} = \frac{x}{6\ell} + \frac{x^2}{24\ell^2} + O\left(\left(\frac{x}{\ell}\right)^3\right)$.

Thus, if $x = O(\ell^{1/2})$ then

$$\begin{aligned}
\frac{\phi(\ell-x)}{\phi(\ell)} &\approx \frac{\left(1-\frac{x}{\ell}\right)^{\frac{1}{2}(\ell-x)^2}\left(1-\frac{x}{\ell}\right)^{\ell-x}\left(1-\frac{x}{\ell}\right)^{\frac{5}{12}} \ell^{\frac{1}{2}(\ell-x)^2+\ell-x+\frac{5}{12}} e^{-\frac{3}{4}(\ell-x)^2-(\ell-x)} (2\pi)^{\frac{1}{2}(\ell-x)}}{\ell^{\frac{1}{2}\ell^2+\ell+\frac{5}{12}} e^{-\frac{3}{4}\ell^2-\ell} (2\pi)^{\frac{1}{2}\ell}} \\
&= \left(1-\frac{x}{\ell}\right)^{\frac{1}{2}(\ell-x)^2}\left(1-\frac{x}{\ell}\right)^{\ell-x}\left(1-\frac{x}{\ell}\right)^{\frac{5}{12}} \ell^{-\ell x+\frac{1}{2}x^2-x} e^{\frac{3}{2}\ell x-\frac{3}{4}x^2+x} (2\pi)^{-\frac{1}{2}x} \\
&\approx e^{-x-\frac{1}{2}x\ell+\frac{3}{4}x^2+\varepsilon_a x-\varepsilon_b x^2} \times \ell^{-\ell x+\frac{1}{2}x^2-x} e^{\frac{3}{2}\ell x-\frac{3}{4}x^2+x} (2\pi)^{-\frac{1}{2}x} \\
&= \ell^{-\ell x+\frac{1}{2}x^2-x} e^{\ell x+\varepsilon_a x-\varepsilon_b x^2} (2\pi)^{-\frac{1}{2}x}.
\end{aligned}$$

Observe that if $s^3 = O(m^2)$ for some $m \to \infty$ with n then

$$\binom{m}{s} = \frac{m^s}{s!} \exp\left\{-\frac{s^2}{2m} + O\left(\frac{s^3}{m^2}\right)\right\}. \tag{24}$$

So, if $\ell \gg j, n-\ell+j \gg 1$ and $\ell = o(n^{1/2})$ then

$$\eta(j) \approx \frac{n^{j(\ell-(j-1)/2)}e^{-3\ell^2 j/2n}\phi(\ell-j)}{\phi(\ell)} \approx \left(\frac{n}{\ell}\right)^{j(\ell-(j-1)/2)}(2\pi\ell)^{-\frac{1}{2}j}e^{\ell j+3\ell^2 j/2n+\varepsilon_a j-\varepsilon_b j^2}, \tag{25}$$

as desired. □

As a consequence, we have:

LEMMA 3.2. *If $\mu_0 \geq \mu_1 \gg \ell^2$ then*

$$\xi(j_0,\mu_0) \approx \frac{\mu_0^{j_0+1}}{(j_0+1)!\eta(j_0)} \approx \frac{(\mu_0 e)^{j_0+1}(2\pi)^{\frac{1}{2}j_0}\ell^{j_0(\ell-\frac{1}{2}j_0+1)}}{\sqrt{2\pi j_0}(j_0+1)^{j_0+1}n^{j_0(\ell-\frac{1}{2}(j_0-1))}e^{(\ell-\frac{1}{3}+\varepsilon_0)j_0}},$$

where $\varepsilon_0 = o(1)$.

PROOF. Using (23), we compute

$$\begin{aligned}\varepsilon_a j_0 - \varepsilon_b^2 j_0^2 &= \frac{j_0^2}{2\ell} - \frac{j_0^3}{6\ell} - \frac{j_0^4}{24\ell^2} + O(j_0^{-1})\\ &= \left(1+O\left(\frac{1}{\omega}\right)\right) - \left(\frac{j_0}{3}+O\left(\frac{j_0}{\omega}\right)\right) + o(1)\\ &= -\left(\frac{1}{3}-\varepsilon_0\right)j_0 \text{ where } \varepsilon_0 = o(1).\end{aligned}$$

The lemma now follows by using (24) to deal with $\binom{\mu_0}{j_0+1}$ and Stirling's approximation and (25) to deal with $\eta(j_0)$. The factor $e^{3\ell^2 j_0/2n}$ can be absorbed into the $\varepsilon_0 j_0$ term. □

Now choose μ_0 as

$$\begin{aligned}\mu_0 &:= \frac{1}{e}\left(\frac{(j_0+1)^{1+\frac{1}{j_0}}n^{\ell-\frac{1}{2}(j_0-1)}e^{\ell-\frac{1}{3}+\frac{4}{3j_0}}j_0^{\frac{1}{2j_0}}n^{\frac{b}{j_0}}}{(2\pi)^{\frac{j_0-1}{2j_0}}\ell^{\ell-\frac{1}{2}j_0+1}}\right)^{\frac{j_0}{j_0+1}}\\ &\approx \frac{j_0}{e^{\frac{4}{3}}\sqrt{2\pi\ell}}\left(\frac{n^{\ell-\frac{1}{2}(j_0-1)}e^{\ell}}{\ell^{\ell-\frac{1}{2}(j_0-1)}}\right)^{\frac{j_0}{j_0+1}} \qquad (26)\\ &= \left(\frac{n}{\ell}\right)^{\ell+o(\ell)}.\end{aligned}$$

Observe that with this choice, we have from Lemma 3.2,

$$\xi(j_0,\mu_0) \approx n^b. \tag{27}$$

We now compare μ_0 and μ_1.

LEMMA 3.3.

$$\mu_1 \geq \mu_0(1-e^{-j_0}). \tag{28}$$

PROOF. Evaluating the exponents in (26), we see from (23) that

$$\frac{(\ell-\frac{1}{2}(j_0-1))j_0}{j_0+1}-(\ell-j_0)=\frac{\frac{1}{2}(j_0^2+j_0)-\ell+j_0}{j_0+1}=\frac{2\ell}{(\omega-1)(j_0+1)}-O(1).$$

In particular this says that

$$\mu_0\gtrsim\frac{j_0}{e^{\frac{4}{3}}\sqrt{2\pi\ell}}\left(\frac{n}{\ell}\right)^{\ell-j_0+\frac{2\ell}{(\omega-1)(j_0+1)}-O(1)}e^{\frac{\ell j_0}{j_0+1}}.$$

It follows from this that

$$\frac{\mu_0}{\binom{n}{\ell-j_0}}\geq\frac{\mu_0\ell^{\ell-j_0}e^{j_0-\ell}}{n^{\ell-j_0}}\gtrsim\frac{j_0}{e^{\frac{4}{3}}\sqrt{2\pi\ell}}\left(\frac{n}{\ell}\right)^{\frac{2\ell}{(\omega-1)(j_0+1)}-O(1)}e^{j_0-\frac{\ell}{j_0+1}}\geq \tag{29}$$

$$\frac{j_0n^{-O(1)}}{e^{\frac{4}{3}}\sqrt{2\pi\ell}}\left(\left(\frac{n}{\ell}\right)^{\frac{2}{\omega}}e^{-1}\right)^{\frac{\ell}{j_0+1}}e^{j_0}=\frac{j_0n^{-O(1)}e^{\frac{\ell}{j_0+1}+j_0}}{e^{\frac{4}{3}}\sqrt{2\pi\ell}}\geq e^{j_0},$$

and the lemma follows. □

We now compare $\xi(j_0,\mu_0)$ and $\xi(j_0-1,\mu_0)$.

LEMMA 3.4.

$$\frac{\xi(j_0-1,\mu_0)}{\xi(j_0,\mu_0)}=\Theta(1). \tag{30}$$

PROOF. We have that

$$\frac{\xi(j_0-1,\mu_0)}{\xi(j_0,\mu_0)}=\frac{\binom{\mu_0}{j_0-1}\binom{n}{\ell-j_0+1}}{\binom{\mu_0}{j_0}}.$$

Using (24) and applying Stirling's formula to $(\ell-j_0+1)!$, we get that

$$\frac{\xi(j_0-1,\mu_0)}{\xi(j_0,\mu_0)}\approx\frac{j_0(ne)^{\ell-j_0+1}}{\sqrt{2\pi\ell}\mu_0(\ell-j_0+1)^{\ell-j_0+1}}$$

$$\approx\frac{j_0(ne)^{\ell-j_0+1}}{\sqrt{2\pi\ell}(\ell-j_0+1)^{\ell-j_0+1}}\cdot\frac{e^{\frac{4}{3}}\sqrt{2\pi\ell}}{j_0}\cdot\left(\frac{\ell^{\ell-\frac{1}{2}(j_0-1)}}{n^{\ell-\frac{1}{2}(j_0-1)}e^{\ell}}\right)^{\frac{j_0}{j_0+1}},$$

where at the end we have used (26). Now

$$\begin{aligned}(\ell-j_0+1)^{\ell-j_0+1}&=\ell^{\ell-j_0+1}\left(1-\frac{j_0-1}{\ell}\right)^{\ell-j_0+1}\\&=\ell^{\ell-j_0+1}\exp\left\{-(\ell-j_0+1)\left(\frac{j_0-1}{\ell}+\frac{(j_0-1)^2}{2\ell^2}+O\left(\frac{j^3}{\ell^3}\right)\right)\right\}\\&\approx\ell^{\ell-j_0+1}e^{-j_0}.\end{aligned}$$

So, we can write

$$\frac{\xi(j_0-1,\mu_0)}{\xi(j_0,\mu_0)}\approx e^{\frac{4}{3}}\left(\frac{\ell}{ne}\right)^{(\ell-\frac{1}{2}(j_0-1))\frac{j_0}{j_0+1}-(\ell-j_0+1)}e^{j_0-\frac{1}{2}(j_0-1)\frac{j_0}{j_0+1}}$$

$$= e^{\frac{4}{3}} e^{\frac{j_0(j_0+3)}{2(j_0+1)}} \left(\frac{\ell}{ne}\right)^{\frac{2\ell}{(\omega-1)(j_0+1)}+O(1)}.$$

Our choice of ω implies that

$$\frac{2\ell}{\omega-1}\log\Big(\frac{ne}{\ell}\Big) = \frac{2\ell(\omega+1)}{\omega-1}.$$

And then we see that

$$\begin{aligned}
&\frac{\xi(j_0-1,\mu_0)}{\xi(j_0,\mu_0)}\\
&\approx e^{\frac{4}{3}}\cdot\exp\left\{\frac{1}{2(j_0+1)}\left(j_0(j_0+3)-\frac{2\ell}{\omega-1}\log\Big(\frac{ne}{\ell}\Big)+O(\log n)\right)\right\}\\
&\approx e^{\frac{4}{3}}\cdot\exp\left\{\frac{1}{2(j_0+1)}\left(j_0(j_0+3)-\frac{2\ell(\omega+1)}{\omega-1}\right)\right\}\\
&=\Theta(1).
\end{aligned}$$

Here we have used (23). □

3.3. *Expected occupancies.* Now we analyze the random variables $X_{j,t}$ and $Y_{j,t}$ in expectation. The next lemma helps us deal with the recurrences (20), (21).

LEMMA 3.5. *Let $\alpha_{-1}=\beta_{-1}=1$.*

(a) *Let $x_{j,t}$ satisfy (i) $x_{0,t}\le\alpha_0+t$, (ii) $x_{j,0}\le\alpha_j$, (iii) $x_{j,t}-x_{j,t-1}\le\beta_{j-1}x_{j-1,t-1}$ for $j\ge1$, where $\beta_j\ge0$ for $j\ge0$. Then when $j\ge1$ we have*

$$x_{j,t}\le\sum_{i=0}^{j+1}\alpha_{j-i}\binom{t}{i}\prod_{s=j-i}^{j-1}\beta_s. \tag{31}$$

(b) *Let $y_{j,t}$ satisfy (i) $y_{j^*,t}\ge\alpha_{j^*}$, (ii) $y_{j,0}\ge0$ for $j>j^*$, and (iii) $y_{j,t}-y_{j,t-1}\ge\beta_{j-1}y_{j-1,t-1}$ for $j>j^*, t\ge1$. Then for $j\ge j^*$ we have*

$$y_{j,t}\ge\alpha_{j^*}\binom{t}{j-j^*}\prod_{s=j^*}^{j-1}\beta_s. \tag{32}$$

PROOF. (a) Now we have

$$x_{1,t}\le\alpha_1+\beta_0\sum_{\tau=1}^{t}(\alpha_0+\tau-1)=\alpha_1+\alpha_0\beta_0 t+\beta_0\binom{t}{2}.$$

So equation (31) is true for $j=1$. Assume inductively that it is true for $j-1$ where $j\ge2$. Then

$$\begin{aligned}
x_{j,t}&\le\alpha_j+\beta_{j-1}\sum_{\tau=1}^{t}x_{j-1.\tau-1}\\
&\le\alpha_j+\beta_{j-1}\sum_{\tau=1}^{t}\sum_{i=0}^{j}\alpha_{j-1-i}\binom{\tau-1}{i}\prod_{s=j-i-1}^{j-2}\beta_s
\end{aligned}$$

$$
\begin{aligned}
&= \alpha_j + \beta_{j-1} \sum_{i=0}^{j} \alpha_{j-1-i} \binom{t}{i+1} \prod_{s=j-i-1}^{j-2} \beta_s \\
&= \alpha_j + \sum_{i=0}^{j} \alpha_{j-1-i} \binom{t}{i+1} \prod_{s=j-i-1}^{j-1} \beta_s \\
&= \alpha_j + \sum_{i=1}^{j+1} \alpha_{j-i} \binom{t}{i} \prod_{s=j-i-1}^{j-1} \beta_s.
\end{aligned}
$$

(b) We have $y_{j^*,t} \geq \alpha_{j^*}$ and so equation (32) is true for $j = j^*$. Assume inductively that it is true for $j-1$ where $j > j^*$. Then,

$$
\begin{aligned}
y_{j,t} &\geq \beta_{j-1} \sum_{\tau=1}^{t} x_{j-1.\tau-1} \\
&\geq \beta_{j-1} \sum_{\tau=1}^{t} \alpha_{j^*} \binom{\tau-1}{j-1-j^*} \prod_{s=j^*}^{j-2} \beta_s \\
&= \alpha_{j^*} \binom{t}{j-j^*} \prod_{s=j^*}^{j-1} \beta_s.
\end{aligned}
$$

□

3.3.1. *Upper Bound.* To use Lemma 3.5 for an upper bound on $\mathbf{E}(X_{j,\tau_0+t}), j \geq 0$ we use the definition of the stopping time τ_0 to define

$$
\alpha_j = |\mathcal{O}_{k+j,\tau_0}| \leq \zeta(j,\mu_0) = \frac{\binom{\mu_0}{j_0+1}}{j_0 \binom{\mu_0}{j_0-j} \eta(j)} \text{ and } \beta_j = \frac{1}{\binom{n}{\ell-j}} \text{ for } j \geq 0.
$$

Thus, for $0 \leq j \leq j_0$ and $\mu_1 \leq t \leq \mu_0$,

$$
\begin{aligned}
\mathbf{E}(X_{j,\tau_0+t}) &\leq \sum_{i=0}^{j+1} |\mathcal{O}_{k+j-i,\tau_0}| \binom{t}{i} \prod_{s=j-i}^{j-1} \frac{1}{\binom{n}{\ell-s}} \\
&\leq \sum_{i=0}^{j+1} \zeta(j-i,\mu_0) \binom{t}{i} \prod_{s=j-i}^{j-1} \frac{1}{\binom{n}{\ell-s}} \\
&= \sum_{i=0}^{j+1} \frac{\binom{\mu_0}{j_0+1}}{j_0 \binom{\mu_0}{j_0-j+i} \eta(j-i)} \binom{t}{i} \prod_{s=j-i}^{j-1} \frac{1}{\binom{n}{\ell-s}} \qquad (33) \\
&\lesssim \frac{\binom{\mu_0}{j_0+1}}{\binom{\mu_0}{j_0}} \binom{t}{j} \prod_{s=0}^{j-1} \frac{1}{\binom{n}{\ell-s}} \qquad (34) \\
&= \xi(j,t).
\end{aligned}
$$

To go from (33) to (34) we let u_i denote the summand in (33) and observe that

$$
\frac{u_{i+1}}{u_i} = \frac{t-i}{i+1} \cdot \frac{j_0 - j + i + 1}{\mu_0 - j_0 + j - i}. \qquad (35)
$$

This implies that

$$\frac{u_j}{u_i} \geq \left(1 - \frac{\omega_1 \binom{n}{\ell - j_0}}{\mu_1}\right)^{j_0} = 1 - o(1),$$

whenever $i < j$. Here we have used (28). Furthermore, (35) implies that $u_{j+1} \lesssim u_j$ for $j \leq j_0$. This verifies (34).

So:

LEMMA 3.6. *We have*

$$\mathbf{E}(X_{j_0,\tau_0+\mu_1}) \lesssim \xi(j_0,\mu_1) \approx \xi(j_0,\mu_0) \approx n^b \tag{36}$$

and

$$\mathbf{E}(X_{j_0-1,\tau_0+\mu_1}) \lesssim \xi(j_0-1,\mu_0) \approx \xi(j_0,\mu_0) \times \Theta(1) = \Theta(n^b). \tag{37}$$

PROOF. These come from (27), (28), (30) and (34). □

3.3.2. *Lower Bound.* To use Lemma 3.5 for a lower bound on $\mathbf{E}(Y_{j,t}), j \geq 0$ we use $y_{j,t} = Y_{j,t}$ for $j \geq j^*$, and take

$$\alpha_{j^*} = |\mathcal{O}_{k+j^*,\tau_0}| \geq \zeta(j^*,\mu_0) \text{ and } \beta_j = \frac{1}{\binom{n}{\ell-j}},$$

to get:

LEMMA 3.7. *For $j \geq 0$,*

$$\mathbf{E}(Y_{j,t}) \geq \zeta(j^*,\mu_0)\binom{t}{j-j^*}\prod_{s=0}^{j-j^*-1}\frac{1}{\binom{n}{\ell-j^*-s}} = \frac{\binom{\mu_0}{j_0+1}}{j_0\eta(j)}\frac{\binom{t}{j-j^*}}{\binom{\mu_0}{j_0-j^*}}.$$

3.4. *Concentration.* We can obtain w.h.p. upper bounds on the sizes of sets $\mathcal{O}_{j_0+j,t}$ by applying Markov's inequality to the random variables $X_{j,t}$. In this section, we obtain suitable w.h.p. lower bounds on the random variables $Y_{j,t}$. Let

$$N_j = \binom{n}{\ell - j} \text{ and } L_j = \frac{\binom{\mu_0}{j_0+1}}{j_0\eta(j)\binom{\mu_0}{j_0-j^*}} \text{ for } j \geq 0.$$

Observe from Lemma 3.7 that we have

$$\mathbf{E}(Y_{j+1,t}) \geq \frac{\binom{t}{j+1-j^*}L_j}{N_j}.$$

We will establish lower concentration of the level sizes inductively, starting from level j^*+1. For each level $j > j^*$, there will be a time t_j past which we have a good w.h.p. lower bound on the level size, which can then be used inductively for the next level.

We define $\widehat{t}_j, j^*+1 \le j \le j_0-1$ by

$$\widehat{t}_j = \min\left\{ t : \frac{L_j\binom{t}{j-j^*}}{N_j} \ge n \right\}.$$

And then define $t_j, j^*+1 \le j \le j_0-1$ by

$$t_j = \max\left\{\widehat{t}_j, \omega_2 t_{j-1}\right\}, \quad \omega_2 = n^c.$$

We also let $t_{j^*} = \tau_0$.

Our definition ensures that $t_j \ll t_{j+1}$ for all $j^* < j < j_0 - 1$. The following Lemma shows that the t_j's don't grow to large.

LEMMA 3.8. *For $j^*+1 \le j \le j_0-1$,*

$$\left(\frac{n}{\ell}\right)^{\ell - O(j_0)} \le t_j \le \omega_2^{j-j^*}\left(\frac{\ell}{n}\right)^{j_0/3}\mu_0 \ll \mu_0. \tag{38}$$

PROOF. We will first need to bound $\widehat{t}_j$ from above and below. Thus we estimate

$$\begin{aligned}
\frac{L_j}{N_j} &= \frac{\binom{\mu_0}{j_0+1}}{j_0\eta(j+1)\binom{\mu_0}{j_0-j^*}} \\
&\approx \frac{\mu_0^{j^*+1}(j_0-j^*)!}{j_0(j_0+1)!\eta(j+1)} \\
&\approx \frac{(j_0-j^*)!}{j_0(j_0+1)!}\left(\frac{j_0}{e^{\frac{4}{3}+o(1)}\sqrt{2\pi\ell}}\left(\frac{n^{\ell-\frac{1}{2}(j_0-1)}e^{\ell}}{\ell^{\ell-\frac{1}{2}(j_0-1)}}\right)^{\frac{j_0}{j_0+1}}\right)^{j^*+1} \\
&\qquad \times \left(\frac{\ell}{n}\right)^{(j+1)(\ell-j/2)}(2\pi\ell)^{\frac{1}{2}(j+1)}e^{-\ell(j+1)-3\ell^2(j+1)/2n-\varepsilon_a(j+1)+\varepsilon_b(j+1)^2} \\
&\le \left(\frac{\ell}{n}\right)^{\ell(j-j^*)+\frac{1}{2}(j_0j^*-j^2)-O(j_0)}.
\end{aligned}$$

Thus,

$$t_j \ge \widehat{t}_j \ge \left(\frac{n}{\ell}\right)^{\ell - \frac{j^2-j_0j^*}{2(j-j^*)}-O(j_0)} \ge \left(\frac{n}{\ell}\right)^{\ell-O(j_0)},$$

since clearly $\frac{j^2-j_0j^*}{j-j^*} \le j_0$.

On the other hand, because $\widehat{t}_j$ is large we can write

$$n \approx \frac{L_j\binom{\widehat{t}_j}{j-j^*}}{N_j} \approx \left(\frac{\widehat{t}_j}{\mu_0}\right)^{j-j^*}\frac{L_j\binom{\mu_0}{j-j^*}}{N_j} = \left(\frac{\widehat{t}_j}{\mu_0}\right)^{j-j^*}\frac{\xi(j_0,\mu_0)\eta(j_0)}{j_0\eta(j+1)} \approx$$

$$\left(\frac{\widehat{t}_j}{\mu_0}\right)^{j-j^*} \frac{n^b}{j_0} \prod_{i=j+1}^{j_0-1} \binom{n}{\ell-i}.$$

We see from this that

$$\frac{\mu_0}{\widehat{t}_j} \gtrsim \left(\frac{n^{b-1}}{j_0} \prod_{i=j+1}^{j_0-1} \left(\frac{n}{\ell}\right)^{\ell-i}\right)^{1/(j-j^*)} =$$

$$\left(\frac{n^{b-1}}{j_0}\right)^{1/(j-j^*)} \left(\frac{n}{\ell}\right)^{(j_0-j-1)(\ell-\frac{1}{2}(j_0+j))/(j-j^*)} \geq \left(\frac{n}{\ell}\right)^{j_0/3}.$$

Consequently,

$$t_j \leq \omega_2^{j-j^*} \widehat{t}_j \leq \left(\frac{\ell}{n}\right)^{j_0/3} \mu_0,$$

which completes the proof of the lemma. □

Our next task is to obtain a high probability lower bound on the random variables $Y_{j^*+1,t}$. Define, for $j^* \leq j \leq j_0 - 1$,

$$\delta_j = \frac{j-j^*}{\omega_2} \tag{39}$$

We define $\mathcal{E}_j$ to be the event that there is a $\tau \in [t_j, \mu_0]$ such that $Y_{j,\tau} < (1-\delta_j) L_j\binom{\tau}{j-j^*}$.

LEMMA 3.9. *For all $j^* \leq j \leq j_0 - 1$, we have*

$$\mathbf{Pr}\left(\mathcal{E}_j\right) \leq \frac{j-j^*}{n^2}. \tag{40}$$

Moreover, we have that w.h.p.

$$Y_{j_0,\mu_1} \gtrsim \frac{\xi(j_0,\mu_0)}{j_0}. \tag{41}$$

PROOF. We prove (40) by induction. The base case $j = j^*$ is trivial because all we assume is that $Y_{j^*,t} \geq L_{j^*} = \zeta(j^*, \mu_0)$ for $t \geq \tau_0$.

Assume now that $j^* + 1 \leq j + 1 \leq j_0 - 1$. We write

$$\tilde{\mathcal{L}}_j = \bigcup_{j' \geq j} \mathcal{L}_{j'}$$

and

$$\tilde{N}_j = |\tilde{\mathcal{L}}_j| = \sum_{j' \geq j} \binom{n}{\ell - j'}.$$

We define a new random variable $Z_{j+1,t} = \theta_{j+1,t_j} + \theta_{j+1,t_j+1} + \cdots + \theta_{j+1,t}$, where the $\theta_{j+1,\tau}$'s are independent $\{0,1\}$ random variables where

$$\mathbf{E}(\theta_{j+1,\tau}) = \frac{\left\lceil (1-\delta_j) L_j\binom{\tau}{j-j^*}\right\rceil}{\tilde{N}_j} \gtrsim \frac{(1-\delta_j) L_j\binom{\tau}{j-j^*}}{N_j} \tag{42}$$

We will define these variables so that

$$\neg\mathcal{E}_j \quad \text{implies} \quad Y_{j+1,t} \geq Z_{j+1,t} \text{ for } t \in [t_j, \mu_1]. \tag{43}$$

For each $\tau \geq t_j$, we define $S_{j,\tau}$ to be the lexicographically first subset of $\bar{\mathcal{L}}_j$ among subsets of size $\left\lceil (1-\delta_j)L_j\binom{\tau}{j-j^*}\right\rceil$ for which a maximum possible number of vertices at level j are occupied. (In particular, $\neg\mathcal{E}_j$ implies that $S_{j,\tau}$ is full.) We let $\theta_{j+1,\tau}$ be the indicator random variable for the event that the path ρ_τ used at step τ intersects $S_{j,\tau}$. Observe that the $\theta_{j+1,\tau}$'s are independent for $\tau \geq t_j$ and also that (42) and (43) hold.

Now for $t \geq t_{j+1}$ we have

$$\begin{aligned}
\mathbf{E}(Z_{j+1,t}) &\gtrsim \frac{(1-\delta_j)L_j}{N_j}\sum_{\tau=t_j}^{t-1}\binom{\tau}{j-j^*} \\
&= \frac{(1-\delta_j)L_j}{N_j}\left(\binom{t}{j-j^*+1} - \binom{t_j+1}{j-j^*+1}\right) \\
&\geq \frac{(1-\delta_j)L_j\binom{t}{j-j^*+1}}{N_j}\left(1-\left(\frac{t_j+1}{t}\right)^{j-j^*+1}\right) \\
&\geq \frac{(1-\delta_j)L_j\binom{t}{j-j^*+1}}{N_j}\left(1-\left(\frac{1}{\omega_2}\right)^{j-j^*+1}\right) \\
&\geq \frac{\left(1-\delta_j-\frac{1}{2\omega_2}\right)L_j\binom{t}{j-j^*+1}}{N_j}
\end{aligned}$$

And applying Hoeffding's theorem, we see that

$$\begin{aligned}
\mathbf{Pr}\left(\exists t \in [t_{j+1},\mu_1] : Z_{j+1,t} \leq \frac{(1-\delta_{j+1})L_j\binom{t}{j+1-j^*}}{N_j}\right) &\leq \sum_{t=t_{j+1}}^{\mu_1}\exp\left\{-\frac{L_j\binom{t}{j+1-j^*}}{10\omega_2^2 N_j}\right\} \\
&\leq \mu_1\exp\left\{-\frac{n}{10\omega_2^2}\right\} = \left(\frac{n}{\ell}\right)^{\ell+o(\ell)}\exp\left\{-\frac{n^{1-2c}}{10}\right\}.
\end{aligned} \tag{44}$$

Thus

$$\mathbf{Pr}(\mathcal{E}_{j+1}) \leq \mathbf{Pr}(\mathcal{E}_j) + \left(\frac{n}{\ell}\right)^{\ell+o(\ell)}\exp\left\{-\frac{n^{1-2c}}{10}\right\} \leq \frac{j-j^*+1}{n^2},$$

completing the inductive proof of (40).

Evaluating just the $t=\mu_1$ term from (44) with $j = j_0-1$, $\delta_{j_0} = n^{-c/4}$ gives

$$\begin{aligned}
\mathbf{Pr}\left(Y_{j_0,\mu_1} \leq \frac{(1-\delta_{j_0})L_{j_0-1}\binom{\mu_0}{j_0-j^*}}{N_{j_0-1}}\right) &\leq \mathbf{Pr}\left(Z_{j_0,\mu_1} \leq \frac{(1-\delta_{j_0})L_{j_0-1}\binom{\mu_0}{j_0-j^*}}{N_{j_0-1}}\right) + \mathbf{Pr}(\mathcal{E}_{j_0-1}) \\
&\leq \exp\left\{-\frac{L_{j_0-1}\binom{\mu_0}{j_0-j^*}}{10n^{c/2}N_{j_0-1}}\right\} + o(n^{-1}) = \exp\left\{-\frac{\xi(j_0,\mu_0)}{10j_0n^{c/2}}\right\} + o(n^{-1}) \\
&\leq \exp\left\{-\frac{n^b}{20n^{(c+a)/2}}\right\} + o(n^{-1}) = o(n^{-1}).
\end{aligned}$$

(Notice we only proved concentration for one value of t in the j_0 case, as opposed to an interval as in (44).) So we have that w.h.p.

$$Y_{j_0,\mu_1} \gtrsim \frac{L_{j_0-1}\mu_0^{j_0-j^*}}{N_{j_0-1}(j_0-j^*)!} \approx \frac{\binom{\mu_0}{j_0+1}\mu_0^{j_0-j^*}}{j_0\binom{\mu_0}{j_0-j^*}\eta(j_0)(j_0-j^*)!} \approx \frac{\binom{\mu_0}{j_0+1}}{j_0\eta(j_0)} = \frac{\xi(j_0,\mu_0)}{j_0},$$

giving (41). □

So now let $\tau_0^* \le \mu_1$ be the first time that level $k+j_0$ is non-empty. Equation (41) shows that τ_0^* exists w.h.p. Now consider the next $\omega_1\binom{n}{\ell-j_0}$ particles. We argue that w.h.p. these particles create an isolated path from level $k+j_0$ to the top.

Observe that we have w.h.p. that

$$|\mathcal{O}_{k+j_0,\tau_0^*}| = 1 \text{ and } |\mathcal{O}_{k+j_0-1,\tau_0^*}| \le n^b \log n.$$

For the second bound we have used the Markov inequality and (37) from Lemma 3.6. Note that we have $\mathcal{O}_{j,\tau_0^*} = \varnothing$ for $j > k+j_0$. Next let $\tau_i^*, i \ge 1$ be the time when the first particle occupies $\mathcal{L}_{k+j_0+i}$. We observe that

$$\mathbf{Pr}\left(\tau_i^* - \tau_{i-1}^* \ge \omega_1\binom{n}{k+j_0+i-1}\right) \le \left(1 - \frac{1}{\binom{n}{k+j_0+i-1}}\right)^{\omega_1\binom{n}{k+j_0+i-1}} \le e^{-\omega_1}. \tag{45}$$

We observe next that for $i \ge 0$ we have

$$\omega_1 \sum_{r=i}^{n-k-j_0} \binom{n}{k+j_0+r} \le 2\omega_1\binom{n}{k+j_0+i} \tag{46}$$

This implies that for $i \le n-k-j_0$ we that w.h.p.,

$$\tau_i^* \le \tau_0^* + 2\omega_1\binom{n}{k+j_0} \ll \mu_0. \tag{47}$$

For the final inequality we used (29).

In particular, $t_{\text{end}} - \tau_0^* = \tau_{n-k-j_0}^* - \tau_0^* \ll \mu_0$. Finally, let us consider the probability that $|\mathcal{O}_{j,t_{\text{end}}}| \ge 2$ for some $i \ge 0, j \ge k+j_0$. At a fixed time $t \in [\tau_0^*, \mu_0]$, the probability that a particle lands on level $k+j_0$ is at most $\frac{n^b \log n}{\binom{n}{k+j_0-1}}$ and the probability that a particle at time $t \in (\tau_i^*, t_{\text{end}}]$ lands at level $k+j_0+i$ ($i>0$) by colliding with the *first* particle which landed at level $k+j_0+i-1$ is at most $\frac{1}{\binom{n}{k+j_0+i-1}}$. Thus, using (45), (46) and (47), the probability the particle at time $t \in [\tau_0^*, t_{\text{end}}]$ becomes the second particle to occupy a level $j \in [k+j_0, n]$ is at most

$$2\omega_1\binom{n}{k+j_0}\frac{n^b\log n}{\binom{n}{k+j_0-1}} + \sum_{i=1}^{n-k-j_0-1} 2\omega_1\binom{n}{k+j_0+i}\frac{1}{\binom{n}{k+j_0+i-1}} + o(1) = o(1). \tag{48}$$

□

4. Further Questions. In some sense, our theorems characterize the beginning and end of the process under consideration. Understanding the behavior of the process in the middle of the cube seems like a major challenge. On the other hand, it is likely to be a prerequisite for an understanding of some basic parameters of the model. For example, from empirical evidence, the following seems likely:

CONJECTURE 4.1. $t_{\text{end}} = o(2^n)$.

Of course an extremely natural target is the following:

QUESTION 4.2. *How large can the parameter a be in Theorem 1.5?*

There are also some interesting modifications of the model to consider. For example, what happens if the random walks are not monotone? In the Boolean lattice this may seem a bit unnatural. It may be interesting to consider hypercubes $[m]^n$, in which case the behavior of the process relative to the relationship between m and n can be explored.

REFERENCES

[1] I. Benjamini and A. Yadin, Upper bounds on the growth rate of Diffusion Limited Aggregation, Arxiv:1705.06095v1, 2017.

[2] S. Boucheron, G. Lugosi and P. Massart, Concentration Inequalities: A Nonasymptotic Theory of Independence, Oxford University Press, Oxford, UK, 2013.

[3] B. Bollobás and Y. Kohayakawa, On Richardsons model on the hypercube, *Combinatorics, geometry and probability* (1993), Cambridge Univ. Press, Cambridge, 1997, pp. 129137.

[4] V. Adamchik, Symbolic and Numeric Computations of the Barnes Function, *Computer Physics Communications*, 157 (2004) 181-190.

[5] D.M. Eberz-Wagner, Discrete growth models, Ph.D Thesis (1999).

[6] R. Eldan, Diffusion-limited aggregation on the hyperbolic plane, *The Annals of Probability* 43 (2015) 2084-2118.

[7] J.A. Fill, and R. Pemantle, Percolation, first-passage percolation and covering times for Richardson's model on the n-cube, *The Annals of Applied Probability* (1993) 593-629.

[8] H. Kesten, How long are the arms in DLA?, *Journal of Physics A* 20 (1987) 29-33.

[9] A. Martinsson, Unoriented first-passage percolation on the n-cube, *The Annals of Applied Probability* 26 (2016) 2597-2625.

[10] T. Witten and L. Sander, Diffusion-limited aggregation, *Physical review B* 27 (1983) 5686-5697.

APPENDIX A: CONCENTRATION INEQUALITY

In this section we derive Lemma 2.3. Recall the statement:

LEMMA A.1 (Lemma 2.3). *Let $X_1, \ldots, X_N$ be independent random variables such that, for all i, $\mathbf{E}(X_i) \leq E$ and $X_i \in [0, C]$ almost surely. Then for $S_N = \sum_{i=1}^N X_i$, $E_N = \mathbf{E}(S_N)$, and for all $t \leq NE$, we have that*

$$\mathbf{Pr}\left(|S_N - E_N| > t\right) < 2\exp\left(-\frac{t^2}{4NEC}\right).$$

This is an immediate consequence of Bernstein's inequality (see, e.g., [2]):

LEMMA A.2 (Bernstein). *Let $X_1, \ldots, X_N$ be independent random variables and V, C be constants such that*

$$\sum_{i=1}^{N} \mathbf{E}(X_i^2) \leq V$$

and for all $q \geq 3$,

$$\sum_{i=1}^{N} \mathbf{E}(\max(X_i^q, 0)) \leq \frac{q!}{2} V C^{q-2}.$$

Then for $S_N = \sum_{i=1}^{N} X_i$, $E_N = \mathbf{E}(S_N)$, we have that

$$\mathbf{Pr}\left(S_N - E_N > t\right) < \exp\left(\frac{-t^2}{2(V + Ct)}\right). \qquad \square$$

PROOF OF LEMMA 2.3. In the setting of Lemma 2.3, the conditions of Lemma A.2 hold for the random variables X_i (as well as for the random variables $-X_i$) by taking C as given, and taking $V = NCE$, since $0 \leq X_i \leq C$. So we have that

$$\mathbf{Pr}\left(S_N - E_N > t\right) < \exp\left(\frac{-t^2/2}{NCE + Ct}\right) \leq \exp\left(\frac{-t^2/2}{2NCE}\right),$$

assuming $t \leq EN$. The analogous statement holds for $-S_N$ also, giving the Lemma. $\square$

DEPARTMENT OF MATHEMATICAL SCIENCES
CARNEGIE MELLON UNIVERSITY
PITTSBURGH PA15213
U.S.A.