Abstract

Given an instance of the preferential attachment graph $G_n = ([n], E_n)$, we would like to find vertex 1, using only ‘local’ information about the graph. Borgs et al gave an $O(\log^4 n)$ time algorithm, which is local in the sense that it grows a connected set of vertices which will contain vertex 1 while still of size $O(\log^4 n)$. We give an $O(\omega \log^{7/2} n)$ algorithm to find vertex 1 (where $\omega \to \infty$ is arbitrary), which is local in the strong sense that it traverses the graph vertex-by-vertex, and operates only on the neighborhood of its current vertex.

1 Introduction

The Preferential Attachment Graph $G_n$ was first discussed by Barabási and Albert [2] and then rigorously analysed by Bollobás, Riordan, Spencer and Tusnády [3]. It is perhaps the simplest model of a natural process that produces a graph with a power law degree sequence.

The Preferential Attachment Graph can be viewed as a sequence of random graphs $G_1, G_2, \ldots, G_n$ where $G_{t+1}$ is obtained from $G_t$ as follows: Given $G_t$, we add vertex $t + 1$ and $m$ random edges $\{e_i = (t + 1, u_i) : 1 \leq i \leq m\}$ incident with vertex $t + 1$. Here the constant $m$ is a parameter of the model. The vertices $u_i$ are not chosen uniformly from $V_t$, instead they are chosen proportional to their degrees. This tends to generate some very high degree vertices, compared with what one would expect in Erdős-Rényi models with the same edge-density. We refer to $u_1, u_2, \ldots, u_m$ as the left choices of vertex $t + 1$. We also say that $t + 1$ is a right neighbor of $u_i$ for $i = 1, 2, \ldots, m$. 

\*Research supported in part by NSF grant CCF1013110
\†Research supported in part by NSF grant DMS
We consider the problem of searching through the preferential attachment graph looking for vertex number 1, using only local information. This was considered by Borgs, Brautbar, Chayes, Khanna and Lucire [5] in the context of the Preferential Attachment Graph $G_n = (V_n, E_n)$, defined below. Here $V_n = [n] = \{1, 2, \ldots, n\}$. They present the following local algorithm that searches for vertex 1, in a graph which may be too large to hold in memory in its entirety. (It is assumed that once the search reaches vertex 1, it will recognise the fact.)

1: Initialize a list $L$ to contain an arbitrary node $u$ in the graph.

2: While $L$ does not contain node 1 do
3: Add a node of maximum degree in $N(L)$ to $L$;
4: return $L$

Here for vertex set $L$, we let $N(L) = \{w \notin L : \exists v \in L \text{ s.t. } \{v, w\} \in E_n\}$.

They show that w.h.p. the algorithm succeeds in reaching vertex 1 in $O(\log^4 n)$ steps. In [5], they also show how a local algorithm to find vertex 1 can be used to give local algorithms for some other problems.

We should note that, as the maximum degree in $G_n$ is $n^{1/2-o(1)}$ w.h.p., one cannot hope to have a polylog($n$) time algorithm if we have to check the degrees of the neighbors as we progress. Thus the algorithm above operates on the assumption that we can find the highest-degree neighbor of a vertex in $O(1)$ time. This would be the case, for example, if the neighborhood of a vertex is stored as a linked-list which is sorted by degrees. In the same situation, we would also be able to find, say, the 10 highest-degree neighbors of a vertex in constant time. In this setting, each of steps 2-7 of the following Degree Climbing Algorithm takes constant time.

**Algorithm DCA:**

The algorithm generates a sequence of vertices $v_1, v_2, \ldots$, until vertex 1 is reached. We let $\omega \rightarrow \infty$ arbitrarily slowly.

1: Let $v_1$ be a vertex of degree at least $\omega \log^{5/2} n$ chosen from (close to) the distribution $\pi$ where $\pi_v = \frac{d_n(v)}{2mn}$. (See Remark 1.3 to see how to implement this step).
2: $i \leftarrow 1$.
3: repeat
4: Let $W = \{w_1, w_2, \ldots, w_m/2\}$ be the $m/2$ neighbors of $v_i$ of largest degree, excluding $v_{i-1}$.
5: Choose $v_{i+1}$ randomly from $W$.
6: $i \leftarrow i + 1$.
7: until $d(v_i) \geq \frac{n^{1/2}}{\log^{1/100} n}$.
Starting from $v_T$, where $T$ is the value of $t$ at this point, do a random walk on the vertices of degree at least $\frac{n^{1/2}}{\log^{1/20} n}$ until vertex 1 is reached.

Note that our DCA algorithm is a local algorithm in a strong sense: the algorithm only requires access to the current vertex and its neighborhood. (Unlike the algorithm from [5], it does not need access to the entire set $v_1, \ldots v_t$ of vertices visited so far.) Our main result is the following:

**Theorem 1.1.** Assuming that $m$ is sufficiently large, algorithm DCA finds vertex 1 in $G_n$ in $O(\omega \log^{7/2} n)$ time, w.h.p.

The bulk of our proof consists of showing that the execution of Steps 2–7 requires only time $O(\log n)$ w.h.p. This analysis requires a careful accounting of conditional probabilities. This is facilitated by the conditional model of the preferential attachment graph due to Bollobás and Riordan [4], which we recast in terms of sums of independent copies of the mean one exponential random variable.

We begin with Step 1 because our analysis relies on the algorithm having a good start, i.e. finding a vertex $v_1 \leq \frac{n}{\omega \log^{5/2} n}$. Though this step is easily implemented as in Remark 1.3, it is in this step that we pay the biggest price in running time. It seems likely that our proof can be improved to make Step 1 unnecessary:

**Conjecture 1.2.** Steps 2-7 of Algorithm DCA find vertex 1 in $G_n$ in $O(\log n)$ time, w.h.p.

**Remark 1.3.** Step 1 can be implemented, starting at any vertex, via a random walk. It is known that w.h.p. the mixing time of such a walk is $O(\log n)$, see Mihail, Papadimitriou and Saberi [11]. Also, the stationary measure of vertices of degree at least $\omega \log^{5/2} n$ is $\Omega \left( \frac{1}{\omega \log^{7/2} n} \right)$, as can be seen from Lemma 3.4 below. So a random walk will w.h.p. find $v_1$ in $O(\omega \log^{7/2} n)$ time.

Furthermore, if the degree of $v_1$ is at least $\omega \log^{7/2} n$ then the same lemma shows that w.h.p.

$$v_1 \leq \frac{n}{\omega \log^{5/2} n}.$$  

We will assume this bound on $v_1$ throughout.

## 2 Outline of paper

In Section 3 we reformulate the construction of Bollobás and Riordan [4] in terms of sums of independent copies of the exponential random variable of mean one. This enables us to show that for $i \leq n_0$ (see (5)), the degree $d_n(i)$ can w.h.p. be expressed asymptotically as $\eta_i \left( \frac{n}{i} \right)^{1/2}$. Here $\eta_i$ is the sum of $m$ independent exponential random variables, subject to some mild conditioning.

Section 5 is the heart of the paper. The aim is to show that if $v_{t-1}$ is not too small, then the ratio $v_t/v_{t-1}$ is bounded above by $9/10$ in conditional expectation. We deduce from this that
w.h.p. the main loop Steps 3–6 only takes $O(\log n)$ rounds. The idea is to determine a degree $\Delta$ such that most of $v_{t-1}$’s left neighbors have degree at least $\Delta$, while only few of $v_{t-1}$’s right neighbors have degree at least $\Delta$. In this way, $v_t$ is likely to be significantly smaller than $v_{t-1}$.

There are many technicalities involved in overcoming the conditioning incurred by running DCA and this explains the length of the proof.

Once we find a vertex $v_T$ of high enough degree, then we know that w.h.p. $v_T$ is not very large and lies in a small connected subgraph of vertices of high degree that contains vertex one. Then a simple argument based on the worst-case covertime of a graph suffices to show that only $o(\log n)$ more steps are required.

### 3 Conditional Model

Bollobás and Riordan [4] gave an ingenious construction equivalent to the preferential attachment graph model. We choose $x_1, x_2, \ldots, x_{2mn}$ independently and uniformly from $[0,1]$. We then let $\{\ell_i, r_i\} = \{x_{2i-1}, x_{2i}\}$ where $\ell_i < r_i$ for $i = 1, 2, \ldots, mn$. We then sort the $r_i$ in increasing order $R_1 < R_2 < \cdots < R_{mn}$ and let $R_0 = 0$. We then let

$$W_j = R_{mj} \quad \text{and} \quad w_j = W_j - W_{j-1} \quad \text{and} \quad I_j = (W_{j-1}, W_j)$$

for $j = 1, 2, \ldots, n$. Given this we can define $G_n$ as follows: It has vertex set $V_n = [n]$ and an edge $\{x, y\}, x \leq y$ for each pair $\ell_i, r_i$, where $\ell_i \in I_x$ and $r_i \in I_y$.

We can generate the sequence $R_1, R_2, \ldots, R_{mn}$ by letting

$$R_i = \left( \frac{\Upsilon_i}{\Upsilon_{mn+1}} \right)^{1/2}. \quad (1)$$

where

$$\Upsilon_N = \xi_1 + \xi_2 + \cdots + \xi_N \quad \text{for} \quad N \geq 1$$

and $\xi_1, \xi_2, \ldots, \xi_{mn+1}$ are independent exponential mean one random variables i.e. $\Pr(\xi_i \geq x) = e^{-x}$ for all $i$.

**Explanation:** The order statistics of $N$ independent uniform $[0,1]$ random variables can be expressed as the ratios $\Upsilon_i/\Upsilon_{N+1}$ for $1 \leq i \leq N$. Then we note that $r_1^2, r_2^2, \ldots, r_{mn}^2$ are independent uniform $[0,1]$ random variables.

We note that the probability density $f(x)$ of the sum $\eta$ of $m$ independent exponential mean one random variables is given by $f(x) = \frac{x^{m-1}e^{-x}}{(m-1)!}$ i.e.

$$\Pr(a \leq \eta \leq b) = \int_a^b \frac{x^{m-1}e^{-x}}{(m-1)!} \, dx. \quad (2)$$

This is a standard result. It can be verified by induction on $m$. See for example Exercise 4.1.4.10 of Grimmett and Stirzaker [8].

We will also use the following estimates that can easily be verified from (2):
Lemma 3.1.

(a) If $\eta = (1 - \alpha)m$ for $0 \leq \alpha \leq 1$ then
\[
\frac{\eta^{m-1}e^{-\eta}}{(m-1)!} \leq e^{-\alpha^2 m/2}.
\] (3)

(b) If $\eta = (1 + \alpha)m$ for $0 < \alpha$ then
\[
\frac{\eta^{m-1}e^{-\eta}}{(m-1)!} \leq \begin{cases} 
  e^{-(m-1)\alpha^2/3} & 0 \leq \alpha \leq 1 \\
  e^{-(m-1)/3} & 1 \leq \alpha \leq 3 \\
  e^{m(1-\alpha/2)} & 3 \leq \alpha
\end{cases}.
\] (4)

(c) It follows from (b) that
\[
\int_{\eta=(1+\alpha)m}^{2m} \frac{\eta^{m-1}e^{-\eta}}{(m-1)!} d\eta \leq me^{-(m-1)\alpha^2/3} 0 \leq \alpha \leq 1.
\]
\[
\int_{\eta=2m}^{3m} \frac{\eta^{m-1}e^{-\eta}}{(m-1)!} d\eta \leq me^{-(m-1)/3}.
\]
\[
\int_{\eta=(1+\alpha)m}^{\infty} \frac{\eta^{m-1}e^{-\eta}}{(m-1)!} d\eta \leq 2e^{-(\alpha-1)m/2} 3 \leq \alpha.
\]

(d) If $x \leq 1$ then
\[
\Pr(\eta \leq mx) \leq (xe^{1-x})^m.
\]
\[
\square
\]

(e) If $Z$ is the sum of $N$ independent exponential mean one random variables then
\[
\Pr(|Z - N| \geq \alpha N) \leq 2e^{-\alpha^2 N/3} for 0 < \alpha < 1.
\]

(f)
\[
\Pr(Z \geq \beta N) \leq \left( \frac{1}{\beta e^{\beta-1}} \right)^N for \beta \geq 2.
\]
\[
\square
\]

We now prove high probability bounds on various parameters of $G_n$. Let
\[
n_0 = \frac{n}{\omega \log^{2+4/m} n} and n_1 = \frac{n}{\log^5 n}.
\] (5)
Lemma 3.2. Let $L$ be a large constant and let $\omega$ be a slowly growing function of $n$ and let $\mathcal{E}_1$ be the event that

$$
\Upsilon_k \in k \left[ 1 \pm \frac{L\theta_k^{1/2}}{3k^{1/2}} \right] \quad \text{for} \quad k = \frac{m + 1}{m} \in \{\omega, \omega + 1, \ldots, n\} \quad \text{or} \quad k = mn + 1
$$

where $\Upsilon_k$ is as in (1) and

$$
\theta_k = \begin{cases} 
\frac{k^{1/2}}{2^{7/2}}, & k \leq n^{2/5} \\
\frac{k^{3/2} \log n}{n^{1/2}}, & n^{2/5} < k \leq n_0 \\
\frac{n}{\omega^{3/2} \log^2 n}, & n_0 < k.
\end{cases}
$$

Then we have:

(a) $\Pr(\neg \mathcal{E}_1) = o(1)$.

(b) Let $\eta_i = \xi_{(i-1)m+1} + \xi_{(i-1)m+2} + \cdots + \xi_{im}$. If $\mathcal{E}_1$ occurs then

\begin{enumerate}
\item $W_i \sim \left( \frac{i}{n} \right)^{1/2}$ for $\omega \leq i \leq n$, and
\item $w_i \sim \frac{\eta_i}{2m(in)^{1/2}}$ for $\omega \leq i \leq n$.
\end{enumerate}

(c) $\eta_i \leq \log n$ for $i \in [n]$ w.h.p.

(d) $\eta_i \leq \log \log n$ for $i \in [\log^{10} n]$ w.h.p.

(e) If $\omega \leq i < j \leq n$ then $\Pr(\text{edge } ij \text{ exists}) \sim \frac{\eta_i}{2(ij)^{1/2}}$.

Proof. (a) Applying Lemma 3.1(e) to (1) for $i \geq 1$ we see that

$$
\Pr(\neg \mathcal{E}_1) \\
\leq 2 \sum_{i=\omega}^{n} \exp \left\{ -\frac{L^2 \theta_i}{27} \right\} + 2 \exp \left\{ -\frac{L^2 \theta_{mn+1}}{27} \right\} \\
= 2 \sum_{i=\omega}^{n^{2/5}} \exp \left\{ -\frac{L^2 i^{1/2}}{27} \right\} + 2 \sum_{i=n^{2/5}+1}^{n/\omega \log^2 n} \exp \left\{ -\frac{L^2 \gamma^{3/2} \log^2 n}{27 n^{1/2}} \right\} + 2 \sum_{i=n/\omega \log^2 n+1}^{n+1} \exp \left\{ -\frac{L^2 n}{27 \omega^{3/2} \log^2 n} \right\} \\
= o(1).
$$

(b) For this we use

$$W_i = \left( \frac{\Upsilon_{mi}}{\Upsilon_{mn+1}} \right)^{1/2}. $$

Then,

$$W_i \notin \left( \frac{i}{n} \right)^{1/2} \left[ 1 \pm \frac{L\theta_i^{1/2}}{i^{1/2}} \right].$$
implies that either 

\[ \Upsilon_{mn+1} \notin (mn + 1) \left[ 1 \pm \frac{L\theta_i^{1/2}}{3(mn + 1)^{1/2}} \right] \] or \[ \Upsilon_{mi} \notin mi \left[ 1 \pm \frac{L\theta_i^{1/2}}{3i^{1/2}} \right]. \]

These events are ruled out by the occurrence of \( \mathcal{E}_1 \).

We now estimate the \( w_i \)'s. We use \( (1 + x)^{1/2} \leq 1 + \frac{x}{2} \) for \( 0 \leq |x| \leq 1 \). Then,

\[
w_i = \left( \frac{\Upsilon_{mi}}{\Upsilon_{mn+1}} \right)^{1/2} - \left( \frac{\Upsilon_{m(i-1)}}{\Upsilon_{mn+1}} \right)^{1/2} \\
= \left( \frac{\Upsilon_{m(i-1)}}{\Upsilon_{mn+1}} \right)^{1/2} \left( \left( 1 + \frac{\eta}{\Upsilon_{m(i-1)}} \right)^{1/2} - 1 \right) \\
\leq \left( \frac{m(i-1)}{(mn+1)} \left( 1 + \frac{L\theta_i^{1/2}}{3m^{1/2}(i-1)^{1/2}} \right)^{1/2} \right)^{1/2} \frac{\eta_i}{2m(i-1)} \left( 1 - \frac{L\theta_i^{1/2}}{3m^{1/2}(i-1)^{1/2}} \right) \\
\leq \frac{\eta_i}{2m(i-1)^{1/2}} \left( 1 + \frac{2L\theta_i^{1/2}}{m^{1/2}i^{1/2}} \right).
\]

A similar calculation gives

\[
w_i \geq \frac{\eta_i}{2m(i-1)^{1/2}} \left( 1 - \frac{2L\theta_i^{1/2}}{m^{1/2}i^{1/2}} \right).
\]

(c),(d) These follow from Lemma 3.1(f).

(e) This follows from (b) and the fact that the edge exists with probability asymptotic to 
\( 1 - \left( 1 - \frac{w_i}{W_i} \right)^n \).

At this point we let \( \mathcal{E}_2 = \mathcal{E}_1 \) intersected with the high probability events described in Lemma 3.2(c),(d).

Let

\[
\lambda_0 = \frac{1}{\log^{4/m} n}.
\]

We need to control the proliferation of vertices \( v \) with \( \eta_v \leq \lambda_0 \). For \( i \in [n] \) we let

\[
B_i = \sum_{v=\omega}^{i} X_v \text{ where } X_v = \frac{\lambda_0 \times 1_{\eta_v \leq \lambda_0}}{2m(vn)^{1/2}}.
\]

Note that given \( \mathcal{E}_2 \), \( B_i \) is asymptotically equal to the total length \( \sum_{v=\omega}^{i} w_v \times 1_{\eta_v \leq \lambda_0} \) of the intervals \( I_v, v \leq i \) that have \( \eta_v \leq \lambda_0 \).
Lemma 3.3.

(a) W.h.p., \( B_i = o \left( \frac{W_i}{\log n} \right) \) for all \( i \geq \log^3 n \).

(b) \( \eta_v > \lambda_0 \) for \( v \leq \log^3 n \).

Proof. (a) It follows from Lemma 3.1(d) and Lemma 3.2(b) that

\[
\mathbb{E}(B_i \mid \mathcal{E}_2) \leq (1 + o(1)) \frac{\lambda_0^m}{2m(Vn)^{1/2}} \leq \left( \frac{\lambda_0}{m} \right)^{1/2} \frac{i}{n}.
\]

Now \( B_i \) is the sum of independent bounded random variables \( X_v \). Applying Hoeffding’s theorem [9] we see that

\[
\Pr \left( B_i \geq 2 \frac{\lambda_0^m}{m} \left( \frac{i}{n} \right)^{1/2} \right) \leq \exp \left\{ -\frac{\lambda_0^{2m}i}{m^2n} \left( \sum_{v=\omega}^{i} \frac{\lambda_0}{4n^2} \right)^{-1} \right\} \leq \exp \left\{ -\frac{\lambda_0^{2m}i}{\lambda_0^2 \log n} \right\} = o(n^{-1}).
\]

(b) The expected number of \( v \leq \log^3 n \) with \( \eta_v \leq \lambda_0 \) is at most \( \log^3 n \times \lambda_0^m = o(1) \) and the result follows. \( \square \)

At this point we let \( \mathcal{E}_3 \) be the intersection of \( \mathcal{E}_2 \) with the high probability events described in Lemma 3.3.

Suppose that we fix the values for \( W_1, W_2, \ldots, W_n \). Then the degree \( d_n(i) \) of vertex \( i \) can be expressed

\[
d_n(i) = m + \sum_{j=1}^{n} \sum_{k=1}^{m} \zeta_{j,k}
\]

where the \( \zeta_{j,k} \) are independent Bernouilli random variables such that

\[
\Pr(\zeta_{j,k} = 1) \in \left[ \frac{w_i}{W_j}, \frac{w_i}{W_{j-1}} \right].
\]

So, putting

\[
\bar{d}_n(i) = \mathbb{E}(d_n(i) \mid \mathcal{E}_3)
\]

we have

\[
mw_i \left( 1 + \sum_{j=i}^{n} \frac{1}{W_j} \right) \leq \bar{d}_n(i) - m \leq mw_i \left( 1 + \sum_{j=i-1}^{n} \frac{1}{W_j} \right).
\]

Now assuming \( \mathcal{E}_3 \), we have for \( i \geq \omega \),

\[
\sum_{j=i}^{n} \frac{1}{W_j} \geq \sum_{j=i}^{n} \left( \frac{n}{j} \right)^{1/2} \left( 1 - \frac{2L\theta_0^{1/2}}{j^{1/2}} \right).
\] (7)
But
\[
\sum_{j=\omega}^{n} \theta_j^{1/2} \leq \sum_{j=\omega}^{n^{2/5}} \frac{1}{j^{3/4}} + \sum_{j=n^{2/5}+1}^{n/\omega \log^2 n} \frac{\log^{1/2} n}{n^{1/4} j^{1/4}} + \sum_{j=n/\omega \log^2 n}^{n} \frac{n^{1/2}}{j \omega^{3/2} \log^2 n} \\
\leq 4n^{1/10} + \frac{4n^{1/2}}{3\omega^{3/4} \log n} + \frac{3n^{1/2} \log \log n}{\omega^{3/2} \log^2 n} \tag{8}
\]

It follows that
\[
\bar{d}_n(i) \geq m + m\omega_i n^{1/2} \left(1 + 2(n^{1/2} - (i + 1)^{1/2}) - \frac{2Ln^{1/2}}{\omega^{3/2} \log n}\right) \\
\geq m + \eta_i \left(\frac{n}{\bar{i}}\right)^{1/2} \left(1 - \left(\frac{i}{n}\right)^{1/2} - \frac{2L}{\omega^{3/2} \log n} - \frac{2\theta_i^{1/2}}{m^{1/2} \bar{i}^{1/2}}\right). \tag{9}
\]

Here we use a lower bound of $1/k^{1/2}$ for $\theta_k/k$ to bound the error term coming from Lemma 3.2(b2).

A similar calculation gives a similar upper bound for $\bar{d}_n(i)$ and this proves that
\[
\bar{d}_n(i) \in m + \eta_i \left(\frac{n}{\bar{i}}\right)^{1/2} \left[1 - \left(\frac{i}{n}\right)^{1/2} \pm \left(\frac{3L}{m\omega^{3/4}} + \frac{3L}{\omega^{3/2} \log n}\right)\right]. \tag{10}
\]

Now $d_n(i) - m$ is the sum of independent 0, 1 random variables. So, from Hoeffding [9],
\[
\Pr(|d_n(i) - \bar{d}_n(i)| \geq \delta(\bar{d}_1(i) - m) | E_3) \leq 2\Pr(E_3)^{-1} e^{-\delta^2(d_n(i) - m)/3} \leq (2 + o(1)) e^{-\delta^2(d_n(i) - m)/3}. \tag{11}
\]

The next lemma summarises what we need to know about $d_n(i)$.

**Lemma 3.4.** With high probability:

(a) $\eta_i \geq \lambda_0$ and $i \leq n_0 \implies d_n(i) \sim \eta_i \left(\frac{n}{\bar{i}}\right)^{1/2}$.

(b) $\eta_i \geq \lambda_1$ and $i \leq n^{1/2} \implies d_n(i) \sim \eta_i \left(\frac{n}{\bar{i}}\right)^{1/2}$.

(c) $i \leq n_0 \implies d_n(i) \leq (1 + o(1)) \max\{1, \eta_i\} \left(\frac{n}{\bar{i}}\right)^{1/2}$.

(d) $d_n(i) \geq \frac{n^{1/2}}{\log^{1/20} n}$ implies that $i \leq \frac{n}{\log^{1/39} n}$.

**Proof.** Putting $\delta = \frac{L}{\omega^{3/4}}$ into (10) we get that
\[
\Pr \left( d_n(i) \notin \eta_i \left(\frac{n}{\bar{i}}\right)^{1/4} \left[1 \pm \frac{4L}{\omega^{3/4}}\right] | E_3, \eta_i \right) \leq \frac{2}{\Pr(E_3)} \exp \left\{ -\frac{L^2 \eta_i n^{1/2}}{i^{1/2} \omega^{1/2}} \right\}. \tag{11}
\]
Here we have
\[
\exp\left\{ -\frac{L^2 \eta_n}{i^{1/2} \omega^{1/2}} \right\} \leq \exp\left\{ -\frac{L^2 \lambda_0 \omega^{1/2} \log^{1+2/m} n}{\omega^{1/2}} \right\} = o(n^{-1}).
\]

(b) Here we have
\[
\exp\left\{ -\frac{L^2 \eta_n}{4i^{1/2} \omega} \right\} \leq \exp\left\{ -\frac{L^2 n^{1/4}}{\omega \log n} \right\} = o(n^{-1}).
\]

(c) We substitute \( \max\{1, \eta_i\} \) for \( \eta_i \) into the probability bound for the upper limit on \( d_n(i) \) in (11).

(d) We can use (c) and Lemma 3.2(d).

Remark 3.5. Lemma 3.4 implies that w.h.p. \( v_1 \leq \frac{n}{\omega \log^+ n} \).

4 A structural lemma

We need to deal with small cycles.

Lemma 4.1. Let a cycle \( C \) be \( \rho_N \)-small if \( |C| \leq \rho_N = 1000 \log_{10/9} \log N \).

(a) Let \( M = e^{20000(\log_{10/9} \log N)^2} \). Then w.h.p., no two \( \rho_N \)-small cycles \( C_1, C_2 \) where one of them \( C_1 \subseteq [M, N] \) in \( G_N \) are within distance \( \rho_N \) of each other.

(b) W.h.p. the number of vertices within distance \( \rho_N \) of a \( \rho_N \)-small cycle is \( N^{o(1)} \).

Proof. (a) We observe that the existence of two small cycles close together implies the existence of a path \( P \) of length at most \( 3\rho_N \) such that the two endpoints both have neighbors in \( P \), other than their \( P \)-neighbors. The expected number of such paths can be bounded by

\[
\sum_{t=4}^{3\rho_N} \sum_{i_1, i_2, \ldots, i_t} \prod_{r=1}^{t} \frac{\log N}{(i_{r-1} i_r)^{1/2}} \left( \sum_{1 \leq r, s \leq t} \frac{\log N}{i_r^{1/2}} \times \frac{\log N}{i_s^{1/2}} \right) \leq \frac{\log^2 N}{\max \{i_r, i_s\}^{1/2}} \sum_{t=4}^{3\rho_N} t^2 \left( \log N \sum_{i=1}^{N} \frac{1}{i} \right)^t \leq \frac{10\rho_N^2}{M^{1/2}} (\log N)^{6\rho_N + 2} = o(1).
\]

Here \( \frac{\log N}{(ij)^{1/2}} \) bounds the probability that the edge \( ij \) exists, regardless of other edges. It follows from Lemma 3.2(c) and (e).

(b) We observe that the number of vertices \( x \geq M \) within \( \rho_N \) of a small cycle is bounded by the number of paths \( P \) with endpoint \( x \) of length at most \( 2\rho_N \) such that the other endpoint \( y \) has a
neighbor in \( P \), other than its \( P \)-neighbors. The expected number of such paths can be bounded by

\[
\sum_{t=4}^{2\rho N} \sum_{i_1, i_2, \ldots, i_t} \prod_{i=1}^{t} \frac{\log N}{(i_{r-1} i_r)^{1/2}} \sum_{1 \leq s \leq t} \frac{\log N}{i_s^{1/2}} \leq \log N \sum_{t=4}^{2\rho N} t \left( \log N \sum_{i=1}^{N} \frac{1}{i} \right)^t \leq 5\rho_N^2 (\log N)^{4\rho_N + 1} \leq N^{o(1)}.
\]

Now use the Markov inequality. \( \square \)

5 Analysis of the main loop

The main loop consists of Steps 3–6. Let \( P_s = (v_1, v_2, \ldots, v_s) \) for \( s \geq 1 \) and let \( P_T \) be the sequence of vertices followed by the algorithm. Let \( Z_i = v_i/v_{i-1} \). We will argue (roughly) that if \( v_{i-1} \notin \{v_1, v_2, \ldots, v_{i-2}\} \) then

\[
E(Z_i \mid P_{i-1}, E) \leq \frac{9}{10}.
\]

So if \( r = 2 \log_{10/9} n \) and \( \mu \) is the number of steps in the main loop, and there are no repetitions of vertices, then

\[
\Pr(\mu \geq r) \leq \Pr \left( Z_1 Z_2 \cdots Z_r \geq \frac{1}{n} \right) \leq n E(Z_1 Z_2 \cdots Z_r) \leq \frac{1}{n}
\]

and so w.h.p. the algorithm will exit the main loop within \( 2 \log_{10/9} n \) steps. Having proved this, we will use Lemma 4.1 to take care of cycles and prove a bound of \( O(\log n) \) on the execution time of the main loop.

We will use a method of deferred decisions, exposing various parameters of \( G_n \) as we proceed.

We do this by considering two separate entities which are exploring the graph \( G_n \). Apart from \textit{Algorithm} which is attempting to find vertex 1, we imagine there is an \textit{Analyzer}, who exposes bits of information.

As Algorithm progresses, we allow the Analyzer access to more and more information. In particular, until Step 7 has been reached, whenever Algorithm chooses vertex \( v_t \):

\textbf{E1} The Analyzer learns the index \( v_t \) and the interval \( I_{v_t} \).

\textbf{E2} The Analyzer learns the left-choices \( \lambda(v_t, 1), \lambda(v_t, 2), \ldots, \lambda(v_t, m) \) and the corresponding left neighbors \( u_1, u_2, \ldots, u_m \). Recall that, a priori, the \( \lambda(v_t, i) \) are random points chosen uniformly from the interval \([0, \rho(v_t, k)]\) for some \( \rho(v_t, k) \in I_{v_t} \) for \( k = 1, 2, \ldots, m \) (see Section 3). The analyzer also learns the intervals \( I_{u_i} \) and the values \( \eta_{u_i} \) for \( i = 1, 2, \ldots, m \).
The Analyzer learns the list $u'_1, u'_2, \ldots, u'_s$ of all vertices $u'_i$ which have the properties that:

(a) $v_t < u'_i \leq \frac{n}{\log^2 n}$.
(b) $u'_i \leq v_t \log^3 n$ (see (22)).
(c) $\lambda(u'_i, k) \in I_{v_t}$ for some $1 \leq k \leq m$.

These are the right-neighbors of $v_t$ out to index $\frac{n}{\log n}$. The analyzer also learns $I_{u'_k}$ and $\eta_{u'_k}$ for $k = 1, 2, \ldots, s$.

Note that at a typical point in the running of the algorithm, Algorithm and the Analyzer have access to incomparable sets of information. For example, the Analyzer knows the indices of vertices examined by Algorithm (knowledge which would make Algorithm's task trivial). On the other hand, Algorithm knows the precise degrees of neighbors of its current vertex (or at least, the precise list of the $m/2$ neighboring vertices of maximum degree).

Our proof will analyze the progress of the Algorithm from Analyzer’s perspective. In particular, we will use the random variables exposed in E1-E3 to prove that Algorithm reaches Step 7 w.h.p. Limiting the information available to the Analyzer is useful to reduce the necessary conditioning. (I.e., we must now prove that Algorithm succeeds using only E1-E3, but also now need only condition on E1-E3.)

Note that our proofs below will sometimes include bounds on probabilities of events involving random variables not exposed in E1-E3 (for example, events concerning the degree of a vertex). In all cases, however, the probabilities of these events are bounded by the probabilities of events involving only the variables from E1-E3 using inequalities (such as those in Lemma 3.4) which hold simultaneously for all vertices in the range of interest in the underlying graph, given $\mathcal{E}$. In particular, though Analyzer has not exposed random variables such as the degrees of vertices:

1. At each step of the analysis, he can compute bounds on some events involving random variables exposed in E1-E3 (as these are exposed only when they first appear in calculations, there is control over the effect of conditioning);
2. He knows that given $\mathcal{E}$, all of these events (simultaneously) give bounds on corresponding events he is interested in (perhaps involving other random variables, such as the degree, which we have not exposed). See (18) and (19) for an example of this.

Remark 5.1. We emphasize that at any point in the running of DCA, it is not the case that the distribution of random variables which Analyzer has not yet learned have identical distributions throughout the process. For example, having learned the left neighbors of all visited vertices so far conditions down slightly the lengths of intervals $I_v$ he has not yet seen yet, for vertices $v$ which are not a known left-neighbor of any vertex. Our proof will have to handle this kind of extra conditioning.
We let $N_L(i)$ denote the left neighbors of $i$ i.e. those $j < i$ that are neighbors of $i$. The distribution of all parameters not exposed so far is random, subject to conditioning on what we know (Remark 5.1) and the continued occurrence of $\mathcal{E}$. In particular, we have

**Observation 5.2.** It follows from Lemma 3.3 that w.h.p. for $t = O(\log n)$, all of the left neighbors $w$ of $v_t \geq \log^3 n$ will satisfy $\eta_w \geq \lambda_0$. (The probability that there exists $t \leq T$ such that $\eta_w < \lambda_0$ is bounded by $O\left(\frac{tB_{\text{ev}}}{W_{\text{ev}}}\right) = o(1)$).

Let

$$M_1 = e^{20000(\log_{10}9 \log n)^2}$$

and

$$\rho_t := \frac{v_t}{v_{t-1}}.$$

**Lemma 5.3.** Suppose that $M_1 \leq v_t \leq n_1$. Then,

$$E(\rho_t | P_{t-1}, \mathcal{E}, \mathcal{L}) \leq \frac{9}{10},$$

where $\mathcal{L} = \{v_k \neq v_\ell : 1 \leq k < \ell \leq T\}$.

**Proof.** To simplify the analysis, we will sometimes compute contributions to $E(\rho_t)$ under the assumption that there is no conditioning other than $\mathcal{E}$. When these calculations are completed, we will show how to correct them for the extra conditioning alluded to in Remark 5.1. For the rest of the proof of this lemma, we let $i = v_{t-1}$ and denote by $j$ a candidate for $v_t$.

We will begin by proving that

$$E(\rho_t | P_{t-1}, \mathcal{E}, \mathcal{D}_t) \leq \frac{9}{10},$$

where $\mathcal{D}_t$ is the event that

$$v_{t-1} \notin \{v_1, v_2, \ldots, v_{t-2}\} \cup N(v_{t-3}) \cup N_2(\{v_1, v_2, \ldots, v_{t-4}\})$$

Here $N(S)$ is the set of neighbors and $N_2(S)$ is the set of vertices within distance two of $S \subseteq [n]$.

The calculation of the ratio $\rho_t$ takes contributions from two cases: that where $v_t$ is a left-neighbor of $v_{t-1}$, and that where $v_t$ is a right-neighbor of $v_{t-1}$. Note that in the former case, since $G_i$ has total degree $2m_i$ and in addition $G_{i/2}$ has total degree $im$, each random choice $j$ by vertex $i$ has probability at least $1/2$ of satisfying $j \leq i/2$. This implies that

$$E(\rho_t | v_t < v_{t-1}) \leq \frac{3}{4}.$$

Next we wish to check that $v_t < v_{t-1}$ is reasonably likely. (Later, we will bound $E(\rho_t | v_t > v_{t-1})$.)
Let \( \Lambda \) denote the degree of a candidate \( j \in N_L(i) \); we wish \( \Lambda \) to be large to bound \( \Pr(v_t = j) \) from below. For \( \gamma > \lambda_0 \) we define

\[
\Delta_i(\gamma) := m + \gamma m \left( \frac{n}{i} \right)^{1/2}.
\]

This is a degree threshold. For a suitable parameter \( \gamma \), we wish it to be known to Analyzer that there should be many left-neighbors but few right-neighbors which have degree > \( \Delta_i(\gamma) \). Given \( \mathcal{E}_1 \), we have

\[
W_i \sim \left( \frac{i}{n} \right)^{1/2}
\]

and for \( j < i \) we have

\[
w_j \sim \frac{\eta_j}{2m(jn)^{1/2}}
\]

and so

\[
\frac{w_j}{W_i} \sim \frac{\eta_i}{2m(ij)^{1/2}}.
\]

Under the assumption that \( \mathcal{D}_t \) holds, the left choices \( \lambda(i, k), k = 1, 2, \ldots, m \) will be uniformly chosen from \([0, W_i]\), except for a few intervals forbidden by \( \mathcal{D}_t \). Lemma 5.7 shows that the amount forbidden is \( o(W_i) \) w.h.p. Then, using Observation 5.2 and Lemma 3.4 we see that

\[
\Pr(\Lambda \leq \Delta_i(\gamma) \mid \mathcal{E}) \leq \frac{1 + o(1)}{W_i} \sum_{j=1}^{i} w_j \Pr(d_n(j) \leq \Delta_i(\gamma) \mid \mathcal{E}, \mathcal{L})
\]

\[
\leq \frac{1 + o(1)}{2mi^{3/2}} \sum_{j=1}^{i} \frac{1}{j^{1/2}} \Pr \left( \eta_j \leq (1 + o(1))m^2 \left( \frac{j}{i} \right)^{1/2} \mid \mathcal{E} \right)
\]

\[
\leq \frac{1 + o(1)}{2mi^{3/2}} \sum_{j=1}^{i} \frac{1}{j^{1/2}} m (\gamma e^{1-\gamma})^m
\]

\[
\leq (1 + o(1))(\gamma e^{1-\gamma})^m.
\]

Let us now consider the degrees of the neighbors \( j \) of \( i \) for which \( j > i \). Let \( R_K(\gamma), K = 0, 1, \ldots, \) denote the number of \( j \) in the interval

\[
J_K(\gamma) = \begin{cases} 
[i, \gamma^{-2}i] & K = 0 \\
[K^2 \gamma^{-2}i, (K + 1)^2 \gamma^{-2}i] & K \geq 1
\end{cases}
\]

with degree at least \( \Delta_i(\gamma) \).

**Remark 5.4.** We are concerned only with \( \gamma > \lambda_0 \) and \( i \leq n_1 \). In which case we can assume that \( R_K = 0 \) for \( K > \frac{\log n}{\lambda_0} \). Indeed, if \( j \in J_K, K > \frac{\log n}{\lambda_0} \) then, from Lemma 3.2(d),

\[
d_n(j) \leq \left( \frac{n}{j} \right)^{1/2} \log n \leq \frac{\gamma}{K} \left( \frac{n}{i} \right)^{1/2} \log n < \lambda_0 \left( \frac{n}{i} \right)^{1/2} \leq \Delta_i(\gamma).
\]
We can therefore assume that $J_K \subseteq [n_0]$. This is because we have

$$i \frac{\log^2 n}{\lambda_0^3} \ll n_0.$$ 

We can therefore assume that

$$v_{t-1} \geq \log^3 n \text{ implies that } v_t \leq v_{t-1} \log^3 n.$$ \hspace{1cm} (22)

Here we can write

$$E(R_K(\gamma) \mid \mathcal{E}, \mathcal{L}) \leq (1 + o(1)) \sum_{j \in J_K(\gamma)} \frac{m}{2(ij)^{1/2}} \int_{\eta_j = \gamma \cdot m(j/i)^{1/2}}^{\infty} \frac{\eta_j^{m-1} e^{-\eta_j}}{(m-1)!} d\eta_j$$

$$= (1 + o(1)) \sum_{j \in J_K(\gamma)} \frac{m}{2(ij)^{1/2}} \int_{\eta_j = \gamma \cdot m(j/i)^{1/2}}^{\infty} \frac{\eta_j^{m-1} e^{-\eta_j}}{(m-1)!} d\eta_j$$ \hspace{1cm} (23)

**Explanation:** The factor $(1 + o(1))$ accounts for the conditioning on $\mathcal{E}_1$, through $\Pr(A \mid B) \leq \Pr(A)/\Pr(B)$ for events $A, B$. We “fix” $\eta_i$ and sum over relevant $j$ and “fix” $\eta_j$. We multiply by the densities of $\eta_i, \eta_j$ and integrate. The factor $\frac{\eta_j}{2(ij)^{1/2}} \approx \frac{m w_i}{W_j}$ is asymptotically equal to the expected number of times $j$ chooses $i$ as a neighbor. We integrate over $\eta_j \geq \gamma m(j/i)^{1/2}$ to get $d_n(j) \geq \Delta_i(\gamma)$, given $\mathcal{E}_2$.

Note that some conditioning has been omitted in the computation of $E(\eta_i)$. This will be accounted for later.

Thus

$$E(R_K(\gamma) \mid \mathcal{E}) \leq (1 + o(1)) \sum_{j \in J_K(\gamma)} \frac{m}{2(ij)^{1/2}} I_j$$

where

$$I_j = \int_{\eta_j = \gamma \cdot m(j/i)^{1/2}}^{\infty} \frac{\eta_j^{m-1} e^{-\eta_j}}{(m-1)!} d\eta_j \leq 1.$$ 

Thus, if $m$ is large then

$$E(R_0(\gamma)) \leq (1 + o(1)) \sum_{j \in J_0(\gamma)} \frac{m}{2(ij)^{1/2}} \leq (1 + o(1)) m \frac{1 - \gamma}{\gamma}.$$ \hspace{1cm} (25)

Continuing,

$$E(R_1(\gamma) \mid \mathcal{E}) \leq \sum_{j \in J_K(\gamma)} \frac{m}{2(ij)^{1/2}} \int_{\eta_j = \gamma \cdot m(j/i)^{1/2}}^{\infty} \frac{\eta_j^{m-1} e^{-\eta_j}}{(m-1)!} d\eta_j$$

$$\leq \frac{m}{2i^{1/2}} \int_{x = \gamma^{-2} i}^{2 \gamma^{-2} i} \frac{1}{x^{1/2}} \int_{\eta_j = \gamma \cdot m(x/i)^{1/2}}^{\infty} \frac{\eta_j^{m-1} e^{-\eta_j}}{(m-1)!} d\eta_j dx$$
\[ E(R_K(\gamma) | \mathcal{E}) \leq \sum_{j \in J_K(\gamma)} \frac{m}{2(i_j)^{1/2}} \int_{\eta_j = K_m}^{\infty} \frac{\eta_j^{m-1}e^{-\eta_j}}{(m-1)!} d\eta_j \leq \sum_{j \in J_K} \frac{m e^{-(m-1)/3}}{2(i_j)^{1/2}} \leq \frac{4m e^{-(m-1)/3}}{\gamma} \] (27)

for \( K = 2, 3 \)

\[ E(R_K(\gamma) | \mathcal{E}) \leq \sum_{j \in J_K(\gamma)} \frac{m}{2(i_j)^{1/2}} \int_{\eta_j = K_m}^{\infty} \frac{\eta_j^{m-1}e^{-\eta_j}}{(m-1)!} d\eta_j \leq \sum_{j \in J_K} \frac{m e^{(3-K)/2}}{2(i_j)^{1/2}} \leq \frac{m(K + 1)^2e^{(3-K)/2}}{K\gamma}, \] (28)

for \( K \geq 4 \).

Next let

\[ S_i(\gamma) = \frac{2}{m} \sum_{K=0}^{\infty} \frac{R_K(\gamma)(K + 1)^2}{\gamma^2}. \]

This is an upper bound on the expected ratio \( \rho_t \) if the neighbor \( j > i \) and \( \gamma^* \leq \gamma \).

It follows from (25) – (28) that if \( \varepsilon \) is a small positive constant and if \( m \gg 1/\varepsilon \) is sufficiently large then

\[ E(R_{K \geq 0}(1 - \varepsilon)) \leq 2\varepsilon m \text{ and } E(S_i(1 - \varepsilon)) \leq 2 + 6\varepsilon. \] (29)

In general, if \( \gamma < 1 \) then

\[ E(R_{K \geq 0}(\gamma)) \leq \frac{2m}{\gamma} \text{ and } E(S_i(\gamma)) \leq \frac{3}{\gamma^2}. \] (30)

Now let

\[ \gamma^*_i = \max \{ \gamma : | \{j \in N_L(i) : d_n(i) < \Delta_i(\gamma) \} | \leq m/2 \}. \]

It follows from (21) that

\[ \text{Pr}(\gamma^*_i \leq \theta | \mathcal{E}) \leq (1 + o(1))(m/2)(\theta e^{1-\theta}m^{2/2}) \leq (2(\theta e^{1-\theta})m/2)^m. \] (31)

Let \( f(\theta) = dP(\gamma^* \leq \theta) \). Then, \[ E(\rho_t) = \int_{\theta=0}^{1} E(\rho_t | \gamma^*_i = \theta) f(\theta) d\theta \]
\[
\begin{align*}
&\leq \int_{\theta=0}^{1-\varepsilon} \mathbb{E}(\rho_t \mid \gamma_i^* = \theta) f(\theta) d\theta + \int_{\theta=1-\varepsilon}^{1} \mathbb{E}(\rho_t \mid \gamma_i^* = \theta) f(\theta) d\theta \\
&\leq \int_{\theta=0}^{1-\varepsilon} \frac{3}{\theta^2} f(\theta) d\theta + \int_{\theta=1-\varepsilon}^{1} \frac{2 \varepsilon m}{m/2} \times (2 + 6 \varepsilon) + \frac{m/2 - 2 \varepsilon m}{m/2} \times \frac{3}{4} f(\theta) d\theta \\
&\leq \int_{\theta=0}^{1-\varepsilon} \frac{3}{\theta^2} f(\theta) d\theta + \frac{7}{8} \\
&\leq \left[ \frac{3}{\theta^2} \Pr(\gamma_i^* \leq \theta) \right]_{0}^{1-\varepsilon} + \int_{\theta=0}^{1-\varepsilon} \frac{6}{\theta^3} \Pr(\gamma_i^* \leq \theta) d\theta + \frac{7}{8} \\
&\leq \frac{3}{(1 - \varepsilon)^2} \left( 2((1 - \varepsilon)e^{\varepsilon})^{m/2} \right)^m + \int_{\theta=0}^{1-\varepsilon} \frac{6}{\theta^3} (2(\theta e^{1-\theta})^{m/2})^m d\theta + \frac{7}{8} \\
&\leq \frac{8}{9}. 
\end{align*}
\]

**Remark 5.5.** Note that in the above calculations we use

\[
\Pr(v_t > v_{t-1} \mid \mathcal{E}, D_t, P_{t-1}) \leq 4\varepsilon.
\]

Let us now account for the extra conditioning. There are three sources to deal with:

**Remark 5.6.** 

(i) We learn that some vertices are not included in the left choices of other vertices. In particular, when we construct the $R_K(\gamma)$ we learn that $v_i$ is not a left neighbor of certain vertices. This affects (19) and (21). We have already indicated that this is dealt with in Lemma 5.7. Also, vertex $v$ not being chosen as the left neighbor of a $v_i$ conditions the value $\eta_v$ downwards. This affects equation (20). Lemma 5.7 also shows that for any fixed $v$, $\Pr(v \text{ not chosen}) = 1 - o(1)$ and so the net effect on (20) is another $1 + o(1)$ factor.

(ii) If $i = v_{t-1} \in N_L(v_{t-2})$ then $\eta_i$ will be conditioned upwards. This affects (23) and (24). In particular the value of $\mathbb{E}(\eta_i)$ needs to be increased. Lemma 5.8 shows that it is only necessary to inflate this by at most $3m$. The effect of this is in the estimates for the $R_K(\gamma)$. Because these estimates drop exponentially fast with $m$, multiplying them by $3m$ has a negligible effect. It will perhaps increase the needed size of $m$.

(iii) If $v_{t-2} \in N_L(i = v_{t-1})$ then then $\eta_i$ will be conditioned upwards. Lemma 5.8 shows that it is only necessary to inflate the estimated value of $\mathbb{E}(\eta_i)$ by at most $3m$ too.

This completes the proof of (13). We now deal with the possibility that DCA visits a vertex $v_\ell$ within distance two of a previously visited vertex $v_k$. Let us call this a *clash* between $k$ and $\ell$. Of course $v_\ell$ is within distance two of $v_{\ell-2}$, but this is excluded from the definition.

Let a vertex $v$ be *unusual* if it has a left neighbor $w$ where $w \leq \frac{v}{\log_{10} n}$. We show in Lemma 5.9 that w.h.p. none of $v_1, v_2, \ldots, v_T$ are unusual.
Suppose first that there is a clash between $k$ and $\ell$ where $\ell - k > \rho_n = 1000 \log_{10/9} \log n$ and that there are no clashes between $k$ and $\ell - 1$. Then, by (13)

$$E\left(\frac{v_\ell}{v_k}\right) \leq \frac{1}{\log^{1000} n}.$$ 

So,

$$\Pr\left(\frac{v_\ell}{v_k} \geq \frac{1}{\log^{50} n}\right) \leq \frac{1}{\log^{50} n}. \tag{32}$$

We can inflate the R.H.S. of (32) by $O(\log^2 n)$ to account for the number of possible choices for $k, \ell$. We can therefore assume that $v_\ell/v_k \leq 1/\log^{50} n$.

This rules out $v_k = v_\ell$. If $v_\ell, v_k$ are neighbors then Remark 5.4 implies that $v_k$ does not affect the choice of $v_{\ell+1}$. Suppose then that $v_k, v_\ell$ are neighbors of some vertex $x$. Then either $x \leq \frac{v_k}{\log^{25} n}$ and then $v_k$ is unusual. Or, $x \geq v_\ell \log^{25} n$ and then Remark 5.4 implies that $x$ does not affect the choice of $v_{\ell+1}$. So w.h.p. either there are no clashes of this type or (13) is unaffected by the clash. We call this the long path property.

Suppose now that $\ell - k \leq \rho_n$. $v_k = v_\ell$ is ruled out by our assumption that $L$ occurs. Suppose next that $v_k$ and $v_\ell$ are neighbors. Let us now consider the effect of $v_k$ on the choice of $v_{\ell+1}$. If we compute $v_{\ell+1}$, ignoring $v_k$ then with a minor adjustment to our calculations for (13) we get that $E(v_{\ell+1}/v_\ell) \leq 0.89 + O(1/m)$. The 0.89 comes from not rounding up to 0.9 and the $O(1/m)$ accounts for excluding $v_k$, if indeed $v_k < v_\ell$. The effect of adding $v_k$ to the mix can only result in a reduction of this expectation. The same argument applies in the case where $v_k, v_\ell$ have a common neighbor $x$. We ignore $x$ and then put it back into the mix.

5.1 Finishing the proof of Theorem 1.1

Let

$$X = \{j \leq T_1 : v_j \in \{v_1, v_2, \ldots, v_{j-1}\}\} = \{j_1 < j_2 < \cdots < j_s\}.$$ 

In the following imagine first that $N = n_1$, we need to change it later to $M_1 \log^{10} n$. We claim that w.h.p.

**P1** $j \in X$ implies that $j > \rho_N$.

**P2** $j, k \in X$ implies that $|j - k| > \rho_N$.

**P3** $j \notin X$ implies that $E(v_{j+1} | P_j, \mathcal{E}) \leq (0.9)v_j$.

**P4** $v_{j+1} \leq v_j \log^3 N$ for $j \geq 0$.

Assume for the moment that $P = \{P1, P2, P3, P4\}$ are all true. We let

$$r(i) = \begin{cases} 0 & i < j_1 \\ \max\{j \in X : j \leq i\} & j \geq j_1 \end{cases} \tag{33}$$
and let $\mathcal{B}_i$ denote the event 

$$
\mathcal{B}_i = \left\{ v_j \leq \frac{(0.9)^j \log^{3r(j)+1} N}{\omega}, 1 \leq j < i \right\}.
$$

Then $\mathcal{P}$ implies that for $\ell \geq 0$ we have 

$$
E(v_i | \mathcal{B}_i) \leq \frac{(0.9)^i \log^{3r(i)+1} N}{\omega}. \tag{34}
$$

So, from the Markov inequality, we have 

$$
\Pr(\neg \mathcal{B}_{i+1} | \mathcal{B}_i) \leq \frac{(0.9)^i \log^{3r(i)+1} N}{\omega}.
$$

Now we have 

$$
\Pr(\neg \mathcal{B}_{i+1}) \leq \Pr(\neg \mathcal{B}_{i+1} | \mathcal{B}_i) + \Pr(\neg \mathcal{B}_i)
$$

and so for any fixed $\tau > 0$ we have 

$$
\Pr(\neg \mathcal{B}_\tau) \leq \sum_{i=1}^{\tau} \frac{(0.9)^i \log^{3r(i)+1} N}{\omega}.
$$

But $X_1, X_2$ imply that $r(i) \leq i/\rho N$. In which case $\log^{r(i)/i} N = \exp \left\{ \log^{\log N/\rho N} \right\} \leq 0.01$ and so 

$$
\Pr(\neg \mathcal{B}_\tau) \leq \sum_{i=1}^{\tau} \frac{(0.95)^i}{\omega} = o(1). \tag{35}
$$

It follows that after $O(\log n)$ steps DCA will w.h.p. reach a vertex $v \in [1, M_1]$. We can then repeat this argument with $N = M_1 \log^{10} n$ to argue that w.h.p. DCA will reach $v_t \leq M_2 = e^{20000(\log \log N)^2} = e^{O((\log \log \log n)^2)} \leq \log^{1/20} n$.

We now verify $\mathcal{P}$. The reader should be aware that we have to verify it for two phases: Phase 1, where $v_t \geq M_1$ and Phase 2, where $v_t \leq M_1 \log^4 n$. Let the sequence of vertices visited in Phase 2 be $w_1, w_2, \ldots$. In the second Phase $w_1$ replaces $v_*$ in the definitions.

For $P1$ in Phase 1, we observe that in $G_n$ there are w.h.p. $n^{1-o(1)}$ vertices with degree at least $\omega \log^{5/2} n$. This follows Lemma 3.4. Property P1 follows from this and Lemma 4.1(b) and the fact that we choose $v_1$ from near the steady state.

In Phase 2, we have less control over the start vertex $w_1$. We avoid this problem by scrapping the assumption of P1 for this phase. We only really needed P1 to avoid an early move to a vertex $v > n_1$. In Phase 2 we know that $w_1 \leq N/\log^{10} n$. Without P1 we replace (34) by 

$$
E(w_i | \mathcal{B}_i) \leq \frac{(0.9)^i \log^{3r(i)+1} N}{\log^{10} n}. 
$$
We now have \( r(i) \leq 1 + i/\rho_N \) (where \( r(i) \) is as defined in (33)). This inflates the numerator in the RHS of (34) by \( \log^3 N \). But this is amply compensated by increasing the denominator by \( \omega^{-1}\log^{10} n \). The proof of (35) goes through unchanged.

Property P2 follows from Lemma 4.1(a). Property P3 follows from Lemma 5.3 and Property P4 follows from (22). It is important to realise that we are replacing \( n \) by \( N \) in the use of (22) in Phase 2. Our definition of \( N \) then implies that w.r.t. \( G_N \), we start at a vertex with index less than \( N/\log^{4} N \).

Using Lemmas 3.3 and 3.4 we see that

\[
\Pr \left( \exists i \in [\log^{1/4} n] : d_n(i) \notin \left( \frac{n}{T} \right)^{1/2} \left[ \frac{1}{\log^{1/10} n}, \log \log n \right] \right) = o(1)
\]

for large \( m \).

In summary, it follows that w.h.p. DCA reaches Step 7 in \( O(\log n) \) time. Also, at this time \( v_T \leq \log^{1/49} n \). The random walk will w.h.p. take place on \([\log^{1/39} n]\). This follows from Lemma 3.4(d). Vertex 1 will be in the same component as \( v_t \) in the subgraph of \( G_n \) induced by vertices of degree at least \( \frac{n^{1/2}}{\log^{1/20} n} \). This is because there is a path from \( v_T \) to vertex 1 through vertices in \([v_T]\) only and furthermore it follows from Lemmas 3.2(c) and 3.4(a) that w.h.p. every vertex on this path has degree at least \( \frac{n^{1/2}}{\log^{1/20} n} \). The expected time to visit all vertices of a graph with \( \nu \) vertices is \( O(\nu^3) \), see for example Aleliunas, Karp, Lipton, Lovász and Rackoff [1]. Consequently, vertex 1 will be reached in a further \( o(\log n) \) steps w.h.p.

This completes the proof of Theorem 1.1.

\( \square \)

### 5.2 Auxilliary Lemmas

Let

\[
S_T = \sum_{t=1}^{T} w_{v_t}.
\]

For some vertices \( j \), we must replace \( \frac{w_j}{W_i} \) by \( \theta \in \left[ \frac{w_j}{W_i}, \frac{w_j}{W_i - S_T} \right] \) as the probability of an edge \( ij \). We now prove that \( S_T = o(W_j) \) for \( j \geq \log^{1/10} n \), showing that this does not affect (17). We also show that

**Lemma 5.7.**

(a) \( S_T = o(W_j) \) for \( j \geq \log^{1/10} n \), w.h.p.

(b) For any fixed \( v \), \( \Pr(v \text{ is not a left choice of a } v_i | \mathcal{E}) = 1 - o(1) \).
Proof. (a) We argue first that w.h.p.,
\[
\max \{ \eta_{v_1}, \eta_{v_2}, \ldots, \eta_{v_T} \} \leq 20 \log \log n. 
\]  
(36)

We begin with \(v_1\) which is produced a little differently from the rest. Let
\[
Z_1 = \sum_{i \leq n_0} d_n(i) \times 1_{\eta_i \geq (1-\epsilon)m}.
\]

It follows from Lemma 3.4(a) that w.h.p.
\[
Z_1 \geq Z_2 = \sum_{i \leq n_0} \log n \times 1_{\eta_i \geq (1-\epsilon)m}.
\]

It then follows from Lemma 3.1(d) and Chernoff bounds that w.h.p.
\[
Z_2 \geq \frac{n \log n}{2\omega \log^{2+1/m} n} \geq \frac{n}{\log^2 n}.
\]

On the other hand, let
\[
Z_3(k, \ell) = \sum_{i=k}^{\ell} d_n(i) \times 1_{\eta_i \geq 20 \log \log n}
\]
and
\[
Z_3 = Z_3(1, n_0) + Z_3(n_0 + 1, n).
\]

Here \(Z_3\) bounds the probability that \(\eta_{v_1} \geq 20 \log \log n\).

It follows from Lemma 3.4 that
\[
E(Z_3(1, n_0)) \leq (1 + o(1)) \sum_{i=1}^{n_0} \frac{n^2}{i} \left( \frac{n_i}{i} \right)^{1/2} \int_{\eta=20 \log \log n}^{\infty} \frac{\eta^m e^{-\eta}}{(m-1)!} d\eta
\]
\[
\leq \frac{n}{\log^{10} n}.
\]

Now we can use the rather weak bound from Lemma 3.4(c) to show that
\[
Z_3(n_0 + 1, n) \leq \omega \log^2 n \sum_{i=n_0 + 1}^{n} 1_{\eta_i \geq 10 \log \log n}.
\]

It follows from Lemma 3.1(f) and the Chernoff bounds that w.h.p. we have
\[
Z_3(n_0 + 1, n) \leq \omega \log^2 n \times \frac{n}{\log^g n}.
\]

The above analysis shows that w.h.p. \(Z_3 = o(Z_1)\) and so \(\eta_{v_1} \leq 20 \log \log n\) w.h.p.

For \(t \geq 2\) we can use Lemma 3.1(d) directly to argue that
\[
\Pr(\eta_{v_t} \geq 20 \log \log n) \leq \frac{1}{\log^{19} n}.
\]
This completes the proof of (36).

It follows from (36) that w.h.p.

\[ S_T \leq (20 + o(1)) \log \log n \sum_{t=1}^{T} \frac{1}{2m(v_t n)^{1/2}}. \]

Let \( \tau = \min \{ t : v_t \leq \log^{10} n \} \). We claim that w.h.p.

\[ T - \tau \leq 20 \log_2 \log n. \] (37)

Given (37) we have

\[ S_T \leq \frac{(2o + o(1))T \log \log n}{2mn^{1/2} \log^5 n} + \frac{20 \log_2 \log n}{2mn^{1/2}} \ll W_j = (1 + o(1)) \left( \frac{j}{n} \right)^{1/2} \]

for \( j \geq \log^{1/10} n \).

But (37) follows directly from the Markov inequality and the fact that \( \mathbb{E}(v_{r+k}) \leq 2^{-k} \log^{10} n \).

(b) We have that

\[ \Pr(v \text{ is not a left choice of a } v_i \mid \mathcal{E}) \leq (1 + o(1)) \sum_{i=1}^{T} \frac{w_v}{W_{v_i}} \leq (1 + o(1)) \sum_{i=1}^{T} \frac{\eta_v}{\sqrt{2m(vv_i)}}. \]

Now w.h.p. there are \( O(\log \log n) \) indices \( i \) such that \( v_i \leq \log^{10} n \). This follows from Lemma 5.3.

We will of course have to use this inductively, in order to avoid a circular argument. In which case we have, from Lemma 3.2(c),(d) that

\[ \sum_{i=1}^{T} \frac{\eta_v}{\sqrt{2m(vv_i)}} \leq O(\log n) \times \frac{\log n}{2m \log^5 n} + O(\log \log n) \times \frac{\log \log n}{2m \log^{1/40} n} = o(1). \]

\[ \square \]

**Lemma 5.8.** Let \( i = v_{t-1} \) and \( j = v_{t-2} \). Then

\[ \bar{E} = \mathbb{E}(\eta_i \mid P_{t-1}, \mathcal{D}_{t-1}, \mathcal{E}) \leq 3m^2. \]

**Proof.** We first assume that \( i < j \). In this case we have

\[ \bar{E} \leq \mathbb{E} \left( \sum_{v \in N_L(j)} \eta_v \mid \mathcal{E} \right) \leq (2 + o(1))m^2. \]
Recall that a vertex is unusual if it has a left neighbor $w$ where $w \leq \frac{v}{\log^{10} n}$.

**Lemma 5.9.** W.h.p., none of $v_1, v_2, \ldots, v_T$ are unusual.

**Proof.** First consider the possibility that $i = v_t$ has a unusual left neighbor. Let $i_0 = \frac{i}{\log^{10} n}$. Assuming $\mathcal{E}$ we find that $W_{i_0}/W_i \leq \frac{1+o(1)}{\log^{10} n}$. So the probability that $i$ chooses a left neighbor $j \leq i_0$ is $O\left(\frac{1}{\log^{10} n}\right)$. Now inflate by $O(\log n)$ to account for all possible values of $t$. 

\[ E = E(\eta_i | P_{t-1}, \mathcal{D}_{t-1}, \mathcal{E}) \leq 3m^2. \]

We write
\[
E \leq \int_{\gamma=0}^{1-\varepsilon} \Pr(\gamma_j^* = \gamma | \mathcal{E}) \sum_{v>j} \frac{E(\eta_{v-2} | \cdots)}{2m(vn)^{1/2}} \int_{\eta_v=(1-o(1))\gamma m(v/j)^{1/2}}^{\infty} \frac{\eta_v^m e^{-\eta_v}}{(m-1)!} d\eta_v. 
\]

Thus
\[
\tilde{E} = E_1 + E_2 + E_3 \leq (1 + 2m^2 e^{-m\varepsilon^2/4}) E(\eta_{v-2} | \cdots). 
\]

Applying the inequality of Remark 5.5 and induction on $t$, we see that
\[
\tilde{E} \leq (1 - 4\varepsilon) \times (2 + o(1))m^2 + 4\varepsilon \times 3(1 + 2m^2 e^{-m\varepsilon^2/4})m^2 \leq 3m^2. 
\]

Recall that a vertex $v$ is unusual if it has a left neighbor $w$ where $w \leq \frac{v}{\log^{10} n}$.
6 Concluding remarks

We have described an algorithm that finds a distinguished vertex using $O(1)$ space. It would be nice to improve the running time to $O(\log n)$. Unless we can find a vertex of degree $\omega \log^{5/2} n$ time more quickly, this will involve dealing with vertices of high index, where the degrees are not so concentrated.

It would be nice to extend the result to other more general models of web graphs e.g. Cooper and Frieze [6]. In this case, we would not be able to use the model described in Section 3.

As a final observation, the algorithm DCA could be used to find the vertex of largest degree. Leastwise, if we replace Step 8 by “Do the random walk for $\log n$ steps and output the vertex of largest degree encountered” then w.h.p. this will produce a vertex of highest degree. This is because $\log n$ will be enough time to visit all vertices $v \leq \log^{1/39} n$, wherein the maximum degree vertex lies.

References


