COMBINATORICA
Akadémiai Kiadó – Springer-Verlag

COMBINATORICA 13 (2) (1993) 209–229

ON DEPTH FIRST SEARCH TREES IN m-OUT DIGRAPHS

W. C. STEPHEN SUEN

Received March 10, 1989
Revised September 27, 1992

We consider depth first search (DFS for short) trees in a class of random digraphs: a m-out model. Let π_i be the i^{th} vertex encountered by DFS and $L(i,m,n)$ be the height of π_i in the corresponding DFS tree. We show that if $i/n \to \alpha$ as $n \to \infty$, then there exists a constant $a(\alpha,m)$, to be defined later, such that $L(i,m,n)/n$ converges in probability to $a(\alpha,m)$ as $n \to \infty$. We also obtain results concerning the number of vertices and the number of leaves in a DFS tree.

1. Introduction and Main Results

DFS algorithms are very useful in graph theory because they are efficient [10]. This paper is inspired by the papers [1,4] where algorithms and methods similar to depth first search were used to find long paths in random graphs. Here, we consider DFS trees in the context of a random m-out digraph model which is defined as follows. We say that a digraph is m-out if every vertex in the digraph has exactly m out-going edges (including loops and multiple edges if any) which are numbered from 1 to m. The random digraph $D(m,n)$ is obtained by randomly picking a digraph, with equal probability, from the set of all m-out digraphs with n vertices $v_1, v_2, \dots, v_n$. In fact, it is easily shown that $D(m,n)$ can be obtained from the following constructive method. We assume that each vertex v in $D(m,n)$ has m different coupons. Each coupon entitles v to pick a vertex from $\{v_1, v_2, \dots, v_n\}$ randomly and independently of others with equal probability $1/n$. If w is the vertex picked using the i^{th} coupon, then we say that $D(m,n)$ contains an edge directed from vertex v to w with label i.

We apply the following DFS algorithm to $D(m,n)$. Note that unlike usual notation for directed edges, we use (i,v) to denote the edge with label i directed from vertex v. We also use $\varphi(i,v)$ to denote the vertex that edge (i,v) is incident to.

AMS subject classification code (1991): 05 C 20, 68 R 10

```
algorithm DFS  (see [5])
Procedure DF (v)
Begin
        DFI (v) := i
        i := i+1
        for j := 1 to m do
            if DFI (φ(j,v)) = 0 then
            begin
                F := F ∪ {edge (j,v)}
                DF (φ(j,v))
            end
end {DF}
Begin {MainProgram}
        DFI (v) := 0 for all vertices v in the graph
        F := empty set ∅
        i := 1
        for k := 1 to n do
            if DFI (v_k) = 0 then DF (v_k)
        output F which contains the edges of the forest of DFS trees
end {MainProgram}
```

The DFS algorithm uses a depth first index DFI(v), where DFI$(v) = k$ iff v is the k^{th} vertex picked by the DFS algorithm. We use π_k to denote the k^{th} vertex picked (ie. DFI$(\pi_k) = k$) and so DFS in fact gives us a sequence $\Pi = \{\pi_1, \pi_2, \ldots, \pi_n\}$, where $\pi_1 = v_1$. Let $S_k = V_n \setminus \{\pi_1, \pi_2, \ldots, \pi_k\}$. Note that DFS picks π_{k+1} from S_k by examining edges in $\{(i, \pi_j) : i = 1, \ldots, m,\ j = 1, \ldots, k\}$ that have not been examined before. Thus each edge inspection is independent of previous edge inspections. At each edge inspection, the edge (i, π_j) is inspected if j is the maximum and i is the minimum among all other edges available for inspection. In examining (i, π_j), if $\varphi(i, \pi_j) \in S_k$ (in which case the examination is said to be successful) then $\pi_{k+1} = \varphi(i, \pi_j)$, otherwise DFS examines another edge. If at any stage when the edges from a vertex π_j $(j \leq k)$ have all been examined and yet π_{k+1} has not been picked, then DFS backtracks to the parent of π_j in the DFS tree. In the case where all edges emanating from $\pi_1, \ldots, \pi_k$ have all been examined without having picked π_{k+1}, then the algorithm "restarts" by simply picking a new vertex v_l where l is the minimum so that v_l has not been picked before. The spanning subgraph of $D(m,n)$ containing only those edges in F obtained using algorithm DFS is therefore a forest.

When $m = 1$, we have a random mapping and the probability of having a DFS tree with $k+1$ vertices (and with a given root) is bounded above by

$$(1 - 1/n)(1 - 2/n) \ldots (1 - k/n) \leq e^{-k^2/2n},$$

which goes to 0 as $n \to \infty$ if $k \geq n^{1/2+\varepsilon}$, $\varepsilon > 0$. Since we are interested in order n properties, we shall focus our attention on $m \geq 2$. The following results will be shown in subsequent sections.

Theorem 1. *Let $N(m,n)$ be the number of vertices in the* DFS *tree with root $v_1 = \pi_1$. Then for any $\varepsilon > 0$,*

$$P(|n^{-1}N(m,n) - 1 + y(m)| \geq \varepsilon) = O(n^{-m}) \quad \text{as} \quad n \to \infty,$$

where $y(m)$ is the smallest root of the equation: $y = \exp\{m(y-1)\}$.

Theorem 2. *Let $N_L(m,n)$ be the number of leaves in the* DFS *tree with root π_1. Then for any $\varepsilon > 0$,*

$$P(|n^{-1}N_L(m,n) - (1-y)^{m+1}/(m+1)| \geq \varepsilon = O(n^{-1}) \quad \text{as} \quad n \to \infty,$$

where $y = y(m)$ is the smallest root of $y = e^{m(y-1)}$.

Standard analysis (e. g. Lagrange expansion formula in [6]) shows that

$$y(m) = \begin{cases} 1 & \text{if } 0 \leq m \leq 1 \\ m^{-1} \sum\limits_{k=1}^{\infty} k^{k-1}(me^{-m})^k/k! < 1 & \text{if } m > 1. \end{cases}$$

For any given $i = 1, 2, \ldots, n$, let $\Pi = \Pi_i = \{\pi_1, \pi_2, \ldots, \pi_i\}$ be the sequence of vertices encountered by DFS until vertex π_i is picked. In Π, there are vertices past which DFS has backtracked. We use black to colour those vertices. Let $L'(i,m,n)$ be the number of non-black vertices in Π. Then we have the following two theorems on $L'(i,m,n)$.

Theorem 3. *Suppose that $i/n \to \lambda$ as $n \to \infty$. If $\lambda \in (0, 1-1/m)$, then for any $\varepsilon > 0$,*

$$P(|n^{-1}L'(i,m,n) - b(\lambda,m)| \geq \varepsilon) = O(n^{-m}) \quad \text{as} \quad n \to \infty,$$

where

$$b(\zeta,m) = \int_0^{\zeta} (1 - x(\beta,m))d\beta,$$

and $x(\beta,m)$ is the smallest positive root of $x = (\beta + x - x\beta)^m$. Note that for β in $(0, 1-1/m)$,

$$x(\beta,m) = \beta^m \sum_{k=0}^{\infty} \frac{1}{k+1} \binom{mk+m}{k} (\beta^{m-1}(1-\beta))^k.$$

Theorem 4. *Suppose that $i/n \to \lambda$ as $n \to \infty$. If $\lambda \in (1-1/m, 1-y(m))$, then for any $\varepsilon > 0$,*

$$P(|n^{-1}L'(i,m,n) - a(\lambda,m)| \geq \varepsilon) = O(n^{-m}) \quad \text{as} \quad n \to \infty,$$

where $a(\lambda,m) = b(1-(1-\lambda)\gamma, m)$ and γ is the largest root of the equation $\gamma = e^{m(1-\lambda)(\gamma-1)}$.

Define $L(i,m,n)$ as the height of π_i in the DFS tree with root π_1. If π_i is in the DFS tree with root π_1, then $L(i,m,n) = L'(i,m,n) - 1$. Using this observation and the facts that $b(\alpha,m)$ is continuous in α and $b(0,m) = 0$, the following result follows immediately from Theorems 1, 3 and 4.

Theorem 5. *Let $L(i,m,n)$ be the height of π_i in the* DFS *tree with root π_1. Suppose that $i/n \to \lambda$ as $n \to \infty$. If $\lambda \in [0, 1-y(m))$, then*

$$P(|n^{-1}L(i,m,n) - a(\lambda,m)| \geq \varepsilon) = O(n^{-m}) \quad \text{as} \quad n \to \infty,$$

where

$$a(\lambda,m) = \begin{cases} b(\lambda,m) & \text{if } \lambda \in [0, 1-1/m] \\ b(1-(1-\lambda)\gamma, m) & \text{if } \lambda \in (1-1/m, 1-y(m)) \end{cases}$$

and γ is the largest root of $\gamma = e^{m(1-\lambda)(\gamma-1)}$.

Also, $(1-\lambda)\gamma$ increases as λ increases from $1-1/m$ (see Lemma 7 in Section 5). Therefore $a(\lambda,m)$ is at maximum when $\lambda = 1-1/m$. We can draw the following conclusion from Theorems 1, 3 and 4.

Theorem 6. *Let $H(m,n)$ be the height of the* DFS *tree with root π_1, then for any $\varepsilon > 0$,*

$$P(|n^{-1}H(m,n) - b(1-1/m, m)| \geq \varepsilon) = O(n^{-m}) \quad \text{as} \quad n \to \infty.$$

Also, if π_τ is the vertex at which the DFS *tree with root π_1 attains its maximum height, then for any $\varepsilon > 0$,*

$$P(|\tau - n(1-1/m)| \geq \varepsilon n) = O(n^{-m}) \quad \text{as} \quad n \to \infty.$$

The random digraph $D(m,n)$ studied in this paper is different from, but very similar to, the usual m-out digraphs appeared, for example, in [3]. The reason for choosing $D(m,n)$ here is to minimize unnecessary notation. Note that Theorems 5 and 6 imply that if DFS is used to find a directed path in $D(m,n)$, then the length of the longest path found is about $nb(1-1/m, m)$. The methods in this paper can be applied to investigate the DFS trees in other random graph models. In particular in [7], DFS is applied to the well known $G_{n,p(n)}$ model (see for example Bollobás [2] and a result on long paths similar to those in [1,4] is obtained. For related results and references on long paths and long cycles in sparse random graphs, please refer to [2]. DFS can also be used to study the largest strongly connected subgraph in $D(m,n)$ (see [8]) as well as long induced paths and large induced trees in $G_{n,p(n)}$ (see[9]). The reader is referred to [5] for references and results related to DFS on graphs.

Theorems 1, 2, 3, and 4 are shown respectively in Sections 2, 3, 4 and 5. Theorems 5 and 6 follow immediately from Theorems 1, 3 and 4.

2. Proof of Theorem 1

Note that after the first DFS tree is found, DFS "restarts" by simply picking a new vertex. This creates some complication because DFS normally picks a new vertex by examining edges. In order to avoid this complications, we add out-going edges (j, v_1), $j \geq m+1$ to the vertex $\pi_1 = v_1$ where each $\varphi(j, v_1)$ is independently chosen from $\{v_1,\dots,v_n\}$ with equal probability. We call this new random graph model $D'(m,n)$. After π_i is obtained, let $Z_{i,n}$ be the number of additional edge inspections used before π_{i+1} is picked in $D'(m,n)$. Then as each edge inspection is independent of previous edge inspections, the quantity $Z_{i,n}$ is geometrically distributed with distribution given by

$$P(Z_{i,n}=j)=(1-i/n)(i/n)^{j-1}, \quad j=1,2,\dots.$$

and $\{Z_{i,n}: i=1,2,\dots,n-1\}$ is a set of mutually independent random variables. Let $U_n(k)$ be the number of edges inspected when vertex π_{k+1} is picked. Then

$$U_n(k)=\sum_{i=1}^{k} Z_{i,n}.$$

Note that π_{k+1} is in the DFS tree with root π_1 in $D(m,n)$ if and only if none of the edges in $\{(i,\pi_1): i\geq m+1\}$ have been inspected while picking $\pi_2,\pi_3,\dots,\pi_{k+1}$. That is, $N(m,n)>k$ if and only if $U_n(1)\leq m,\ U_n(2)\leq 2m,\dots,U_n(k)\leq km$. Hence for $k\geq 1$,

$$P(N(m,n)>k)=P(U_n(1)\leq m,\ U_n(2)\leq 2m,\ \dots,\ U_n(k)\leq km). \tag{1}$$

We shall use equation (1) and the following estimates of $U_n(k)$ to show Theorem 1.

Note that throughout the rest of this paper, we shall use ε to denote a generic positive constant, ϱ to denote a number in $(0,1)$ and η to denote a positive number. The numbers ϱ and η may depend on some other constants but not on n.

Lemma 1. *Let $\varepsilon>0$ and $\delta\in(0,1)$ be positive constants. Suppose that $k=k(n)$ satisfies $k/n\to\alpha\in(0,\delta)$. Then there exists $\varrho=\varrho(\varepsilon,\delta)$ in $(0,1)$ such that for all sufficiently large n, we have*

$$P(|U_n(k)+n\log(1-k/n)|\geq\varepsilon n)\leq\varrho^n.$$

Proof. For $t\leq-\log(k/n)=-\log\alpha+o(1)$,

$$\begin{aligned}&E[\exp(tU_n(k)+nt\log(1-k/n))]\\&\quad=\prod_{l=1}^{k}\{e^t(1-l/n)(1-e^t l/n)^{-1}\}e^{nt\log(1-k/n)}=\{\exp(f_n(t))\}^n,\end{aligned}$$

where $f_n(t)=\frac{kt}{n}+\frac{1}{n}\sum_{l=1}^{k}\log(1-l/n)-\frac{1}{n}\sum_{l=1}^{k}\log(1-e^t l/n)+t\log(1-k/n)$.

Since $\frac{1}{n}\sum_{l=1}^{k}\log(1-l/n)$ is bounded below and above by

$$-1-\left(1-\frac{k+1}{n}\right)\log\left(1-\frac{k+1}{n}\right)\quad\text{and}\quad-1-\left(1-\frac{k}{n}\right)\log\left(1-\frac{k}{n}\right),$$

and $\frac{1}{n}\sum_{l=1}^{k}\log(1-e^t l/n)$ is bounded below and above by

$$-1-e^{-t}\left(1-e^t\frac{k+1}{n}\right)\log\left(1-e^t\frac{k+1}{n}\right) \quad \text{and}$$
$$-1-e^{-t}\left(1-e^t\frac{k}{n}\right)\log\left(1-e^t\frac{k}{n}\right),$$

we have that

$$\begin{aligned} f(\alpha,t) &= \lim_{n\to\infty} f_n(t) \\ &= \alpha t-(1-\alpha)\log(1-\alpha)+e^{-t}(1-e^t\alpha)\log(1-e^t\alpha)+t\log(1-\alpha). \end{aligned}$$

Note that since $\frac{\partial f(\alpha,t)}{\partial\alpha}>0$ for all $t\neq 0$, we have that

$$f(\alpha,t)\le \delta t-(1-\delta)\log(1-\delta)+e^{-t}(1-e^{-t}\delta)\log(1-e^t\delta)+t\log(1-\delta)=f(\delta,t).$$

Now by Markov's inequality, we have for any $t>0$,

$$\begin{aligned} P(U_n(k)+n\log(1-k/n)\ge\varepsilon n) &= P(\exp(tU_n(k)+nt\log(1-k/n))\ge e^{\varepsilon nt}) \\ &\le e^{-\varepsilon nt}E[\exp(tU_n(k)+nt\log(1-k/n))] \\ &\le \exp(-\varepsilon nt+nf_n(t)) \\ &\le (\exp(-\varepsilon t+f(\delta,t)+o(1)))^n, \end{aligned}$$

and similarly

$$\begin{aligned} P(U_n(k)+n\log(1-k/n)\le-\varepsilon n) &= P(-U_n(k)-n\log(1-k/n)\ge\varepsilon n) \\ &\le \exp(-\varepsilon nt+nf_n(-t)) \\ &= (\exp(-\varepsilon t+f(\delta,-t)+o(1)))^n. \end{aligned}$$

Since $f(\delta,t)=O(t^2)$ as $t\to 0$, the quantity $f(\delta,t)-\varepsilon|t|$ is strictly negative for all non-zero t sufficiently close to 0. Choosing a suitable t gives that there exists $\varrho=\varrho(\varepsilon,\delta)$ in $(0,1)$ such that for all large n,

$$P(|U_n(k)+n\log(1-k/n)|\ge\varepsilon n)\le\varrho^n. \qquad \blacksquare$$

Before we show Theorem 1, we observe that by plotting $f(y)=y$ and $f(y)=e^{m(y-1)}$, if y is the smallest root of $y=e^{m(y-1)}$, then for small $\varepsilon>0$,

(2) $\quad y-\varepsilon<e^{m(y-\varepsilon-1)}$ or $m(1-y+\varepsilon)+\log(y-\varepsilon)=\eta_1(\varepsilon)<0$, and

(3) $\quad y+\varepsilon>e^{m(y+\varepsilon-1)}$ or $m(1-y-\varepsilon)+\log(y+\varepsilon)=\eta_2(\varepsilon)>0$.

We shall now show Theorem 1. If $k=[n-ny(m)+\varepsilon n]-1$, then

$$P(N(m,n)\ge n-ny(m)+\varepsilon n)=P(N(m,n)>k)$$

which, from (1), is bounded above by

$$P(U_n(k) \leq km) = P(U_n(k) + n\log(1-k/n) \leq km + n\log(1-k/n))$$
$$\leq P(U_n(k) + n\log(1-k/n) \leq mn(1-y+\varepsilon) + n\log(y-\varepsilon) + O(1)).$$

Hence from (2) and Lemma 1, for large n,

$$(4) \qquad P(N(m,n) \geq n - ny(m) + \varepsilon n)$$
$$\leq P(U_n(k) + n\log(1-k/n) \leq -n\eta(\varepsilon) + O(1)) \leq \varrho^n,$$

We shall next show that for any $\varepsilon > 0$,

$$(5) \qquad P(N(m,n) \leq n - ny(m) - \varepsilon n) = O(n^{-m}) \quad \text{as} \quad n \to \infty.$$

Let $l = [n/40]$. Note that $1/40 < 1 - y(m)$ for all $m \geq 2$. Then

$P(N(m,n) \leq l) \leq P($there is a set W of vertices not including v_1 such that $0 \leq |W| < l$ and every edge directed from $\{v_1\} \cup W$ leads to a vertex in $\{v_1\} \cup W)$

$$\leq \sum_{j=0}^{l-1} \binom{n-1}{j} ((j+1)/n)^{(j+1)m} = \sum_{j=0}^{l-1} a_j, \quad \text{say.}$$

It is easy to check that for $j = 1$ to $l-2$, $a_{j+1}/a_j \leq (1/40)^{m-1} e^{m+1}$, which is less than 1 because $m \geq 2$. Hence as $n \to \infty$,

$$(6) \qquad P(N(m,n) \leq l) \leq a_0 + na_1 = O(n^{-m}) + O(n^{2-2m}) = O(n^{-m}).$$

Let $k = \lfloor n - ny(m) - \varepsilon n \rfloor$. If $k \leq l$, then (5) follows immediately from (6). Therefore assume $k > l$. Then using (1), as $n \to \infty$,

$$P(N(m,n) \leq n - ny(m) - \varepsilon n)$$
$$= P(N(m,n) \leq l) + \sum_{j=l+1}^{k} P(N(m,n) = j)$$
$$\leq O(n^{-m}) + \sum_{j=l+1}^{k} P(U_n(j) > jm)$$
$$= O(n^{-m}) + \sum_{j=l+1}^{k} P(U_n(j) + n\log(1-j/n) > jm + n\log(1-j/n)).$$

Let $g(t) = mt + \log(1-t)$. Since $g(t)$ is a concave function and since j satisfies $1/40 \leq j/n \leq 1 - y - \varepsilon$, we have from (3) that

$$g(j/n) \geq \eta(\varepsilon) = \min(g(1/40), g(1-y-\varepsilon)) > 0.$$

Hence by Lemma 1, for large n,

$$P(U_n(j) + n\log(1-j/n) > jm + n\log(1-j/n))$$
$$\leq P(U_n(j) + n\log(1-j/n) > n\eta(\varepsilon)) \leq \varrho^n,$$

It therefore follows that

$$P(N(m,n) \geq n - y(m)n - \varepsilon n) \leq O(n^{-m}) + n\varrho^n = O(n^{-m}) \quad \text{as} \quad n \to \infty.$$

Our proof of Theorem 1 is therefore complete. ∎

3. Proof of Theorem 2

With digraph $D'(m,n)$ and variable $Z_{j,n}$ defined as in Section 2, let

$$W_j = \begin{cases} 1 & \text{if } Z_{j,n} > m \\ 0 & \text{otherwise.} \end{cases}$$

Now π_j is a leaf in DFS tree with root π_1 in $D(m,n)$ if and only if $W_j = 1$ and π_j is in the DFS tree. Thus

$$N_L(m,n) = \sum_{j=1}^{N(m,n)} W_j. \tag{7}$$

We require the following lemma.

Lemma 2. *Suppose that k is such that $k/n \to \alpha \in (0,1)$ as $n \to \infty$. Then for any $\varepsilon >$ 0, there is ϱ in $(0,1)$ such that for all large n,*

$$P\left(\left|\sum_{j=1}^{k} W_j - n(k/n)^{m+1}/(m+1)\right| \geq \varepsilon n\right) \leq \varrho^n.$$

Proof. Our proof here is similar to that of Lemma 1. Note that for sufficiently small $t > 0$,

$$\begin{aligned} E[\exp(tW_j)] &= e^t(j/n)^m + 1 - (j/n)^m \leq 1 + (t+t^2)(j/n)^m \\ &\leq \exp((t+t^2)(j/n)^m), \end{aligned}$$

and

$$\begin{aligned} E[\exp(-tW_j)] &= e^{-t}(j/n)^m + 1 - (j/n)^m \leq 1 - (t-t^2)(j/n)^m \\ &\leq \exp(-(t-t^2)(j/n)^m). \end{aligned}$$

Thus for sufficiently small $t > 0$,

$$\begin{aligned} E\left[\exp\left(t\sum_{j=1}^{k} W_j\right)\right] &\leq \exp\left((t+t^2)\sum_{j=1}^{k}(j/n)^m\right) \\ &= \exp((t+t^2)n(k/n)^{m+1}/(m+1) + o(n)) \\ &= \{\exp((t+t^2)(k/n)^{m+1}/(m+1) + o(1))\}^n, \end{aligned}$$

and similarly for $t > 0$,

$$\begin{aligned} E\left[\exp\left(-t\sum_{j=1}^{k} W_j\right)\right] &\leq \exp\left(-(t-t^2)\sum_{j=1}^{k}(j/n)^m\right) \\ &= \{\exp(-(t-t^2)(k/n)^{m+1}/(m+1) + o(1))\}^n. \end{aligned}$$

Therefore, for suitable $t>0$ and for large enough n,

$$P\left(\sum_{j=1}^{k} W_j - n(k/n)^{m+1}/(m+1) \geq \varepsilon n\right)$$
$$\leq E\left[\exp\left(t\sum_{j=1}^{k} W_j\right)\right] \exp(-tn(k/n)^{m+1}/(m+1))\exp(-\varepsilon nt)$$
$$= \{\exp(-\varepsilon t + t^2(k/n)^{m+1}/(m+1) + o(1))\}^n \leq \varrho^n,$$

and

$$P\left(\sum_{j=1}^{k} W_j - n(k/n)^{m+1}/(m+1) \leq -\varepsilon n\right)$$
$$\leq E\left[\exp\left(-t\sum_{j=1}^{k} W_j\right)\right] \exp(tn(k/n)^{m+1}/(m+1))\exp(-\varepsilon nt)$$
$$= \{\exp(-\varepsilon t + t^2(k/n)^{m+1}/(m+1) + o(1))\}^n \leq \varrho^n.$$

Lemma 2 now follows. ∎

Since, from Theorem 1, for any $\varepsilon>0$,

$$P(|N(m,n) - (1-y)n| \geq \varepsilon n) = O(n^{-m}) \quad \text{as} \quad n \to \infty,$$

Theorem 2 now follows very easily from Lemma 2 and (7).

4. Proof of Theorem 3

Assume $i/n \to \lambda$ in $(0, 1-1/m)$ as given in Theorem 3. Consider the sequence $\Pi = \{\pi_1, \pi_2, \ldots, \pi_i\}$ of vertices picked by DFS . Recall from Section 1 that we colour the vertices in Π so that π_j is black if and only if π_j is a vertex past which DFS has backtracked. The sequence Π is now separated into runs of black vertices. If $\pi_j, \pi_{j+1}, \ldots, \pi_k$ is a run of black vertices (with π_{j-1} and π_{k+1} non-black), we say that π_j (resp. π_k) is the left end (resp. right end) of the run of black vertices. Note that $L'(i,m,n) = i-$number of black vertices. For $s<t$, let $\Pi(s,t)$ denote the subsequence $\{\pi_s, \ldots, \pi_t\}$ of Π and let $B(s,t)$ be the number of black vertices in $\Pi(s,t)$. We shall obtain the following estimates of $B(s,t)$, proof of which will be given later, and use these estimates to show Theorem 3.

Lemma 3. *Suppose that $\alpha n \leq s < t \leq \beta n$ and $(t-s)/n \to c$ as $n \to \infty$. Then for any $\varepsilon_1 > 0$, there is $\varrho_1 \in (0,1)$ such that for all large n,*

$$P(cnx(\alpha,m) - \varepsilon_1 n \leq B(s,t) \leq cnx(\beta,m) + \varepsilon_1 n) > 1 - \varrho_1^n.$$

Proof of Theorem 3. Note that Lemma 3 suggests that $B(\lfloor \beta n\rfloor, \lfloor(\beta+\Delta\beta)n\rfloor)/\Delta\beta \approx x(\beta,m)n$ when $\Delta\beta$ is small. Thus, intuitively, $L'(i,m,n)$ is about $n\int_0^\lambda (1-x(\beta,m))d\beta$. We shall make this idea rigorous as follows. Let ξ be a large positive integer to be defined later. For $l=1,2,\ldots,\xi$, let

$$\begin{aligned} \alpha_l &= \lambda l/\xi \qquad \text{with} \quad \alpha_0 = 0, \\ t_l &= \lfloor \alpha_l n \rfloor \qquad \text{with} \quad t_0 = 0, \\ \Pi_l &= \{t_{l-1}+1,\ldots,t_l\}, \\ B_l &= \text{number of black vertices in} \quad \Pi_l. \end{aligned}$$

Note that if B is the number of black vertices in Π, then

$$B_2 + \ldots + B_\xi \le B \le B_1 + B_2 + \ldots + B_\xi + o(n).$$

From Lemma 3, we have that for any $\varepsilon_1 > 0$, there is a $\varrho(l)$ in $(0,1)$ so that for large n,

$$P\left(\frac{\lambda}{\xi} nx(\alpha_{l-1},m) - \varepsilon_1 n \le B_l \le \frac{\lambda}{\xi} nx(\alpha_l,m) + \varepsilon_1, n\right) > 1 - \varrho(l)^n.$$

Hence for any $\varepsilon_1 > 0$, we may choose ϱ in $(0,1)$ and ϱ depends on ξ such that for large n,

$$P\left(\text{for} \quad l=1,\ldots,\xi,\ \frac{\lambda}{\xi} nx(\alpha_{l-1},m) - \varepsilon_1 n \le B_l \le \frac{\lambda}{\xi} nx(\alpha_l,m) + \varepsilon_1 n\right) > 1 - \varrho^n,$$

which implies that

$$P\left(\sum_{l=1}^{\xi-1} \frac{\lambda}{\xi} nx(\alpha_l,m) - \varepsilon_1 n\xi \le \sum_{l=1}^{\xi} B_l \le \sum_{l=1}^{\xi} \frac{\lambda}{\xi} nx(\alpha_l,m) + \varepsilon_1 n\xi\right) > 1 - \varrho^n.$$

For any ε_2 and ε_3, since $x(\alpha,m)$ is increasing with α (for $\alpha < 1-1/m$), we can choose ξ large enough so that

$$\sum_{l=1}^{\xi-1} \frac{\lambda}{\xi} x(\alpha_l,m) + \varepsilon_2 \ge \int_0^\lambda x(\beta,m)d\beta \ge \sum_{l=1}^{\xi} \frac{\lambda}{\xi} x(\alpha_l,m) - \varepsilon_3.$$

Hence we have for any ε_1 and for any $\varepsilon_2, \varepsilon_3 > 0$, we can choose ξ large enough so that there is ϱ in $(0,1)$,

$$P\left(n\int_0^\lambda x(\beta,m)d\beta - (\varepsilon_1\xi + \varepsilon_2)n \le B \le n\int_0^\lambda x(\beta,m)d\beta + (\varepsilon_1\xi + \varepsilon_3)n\right) > 1 - \varrho^n$$

for all large n. Since $i = \lambda n + o(n)$, Theorem 3 now follows from the above (since ε_1, ε_2, ε_3 are all arbitrary). ∎

It therefore remains to show Lemma 3. Let us first make the following observation. For edge e in $\{(j,v):j=1,2,\ldots,m, v=\pi_1,\pi_2,\ldots,\pi_i\}$, let

$$X_e = \begin{cases} 1 & \text{if } e \text{ has been succesfully inspected} \\ 0 & \text{otherwise.} \end{cases}$$

Note that conditional on the event that edge e is examined when picking vertex π_j, we have $P(X_e=0)=(j-1)/n$. Let L_j be the set $\{(l,\pi_j):l=1,2,\ldots,m\}$. Define

$$Y_j = \sum_{e\in L_j} X_e.$$

Suppose that π_{k+1} is uncoloured. Then π_k is coloured if and only if $Y_k = 0$. Also, if π_k is the right end of a run of black vertices in Π (that is, $Y_k = 0$ and π_{k+1} is uncoloured), then the event that the vertices $\pi_{j+1},\pi_{j+2},\ldots,\pi_{k-1}$ are black corresponds to the event E that

(8) $$Y_{k-1} \le 1, Y_{k-1}+Y_{k-2} \le 2, \ldots, Y_{k-1}+\ldots,+Y_{j+1} \le k-j-1.$$

We have (8) because if π_k is the right end of a run of black vertices, then the vertices $\pi_{j+1},\pi_{j+2},\ldots,\pi_{k-1}$ are black if and only if for each l from $j+1$ to $k-1$, every edge directed from $\pi_l,\pi_{l+1},\ldots,\pi_{k-1}$ has been examined and at most $(k-1-l)+1=k-l$ of these edge inspections are successful. Suppose now that $k\le\beta n$ and $j\ge\alpha n$. Then since $\alpha\le P(X_e=0)\le\beta$, for all edge e directed from $\pi_{j+1},\ldots,\pi_{k-1}$, and since each edge in $D'(m,n)$ is directed independently of others, we have for $j=1,2,\ldots,m$, and for $j\le l\le k$ that $P_\alpha(Y_l\le j)\le P(Y_l\le j)\le P_\beta(Y_l\le j)$. Hence

(9) $$P_\alpha(E) \le P(E) \le P_\beta(E),$$

where P_ζ is the probability law corresponding to the probability space in which $P_\zeta(X_e=0)=\zeta$ and the random variables X_e's are independent.

To find an upper bound for $B(s,t)$ in Lemma 3, we perform the following experiment. Suppose that $l\le t$ is the largest integer such that π_l is non-black. If there is no such l, let $l=s-1$. That is, $\pi_{l+1},\pi_{l+2},\ldots,\pi_t$ are black while π_l is not. We use green to colour the vertices in $\Pi(s,t)$ according to the following rules:

(i) if $l<t$, colour the vertices $\pi_{l+1},\ldots,\pi_t$ and do not colour π_l,

(ii) start colouring from π_{l-1} to π_s as follows:

  (a) if π_{k+1} is uncoloured, colour π_k if $Y_k=0$, otherwise do not colour π_k, where $Y_k=\sum_{e\in L_k} X_e$ and X_e's are sampled with the probability law P_β,

  (b) if π_k is coloured but π_{k+1} is not, then we colour the vertices $\pi_{j+1},\pi_{j+2},\ldots,\pi_{k-1}$ but do not colour π_j if

(10) $$Y_{k-1}\le 1,\quad Y_{k-1}+Y_{k-2}\le 2,\ldots,Y_{k-1}+\ldots+Y_{j+1}\le k-j-1$$
$$\text{and}\quad Y_{k-1}+\ldots+Y_j \ge k-j+1,$$

(iii) the colouring process is stopped once π_s is reached.

Let $G(s,t)$ be the number of green vertices in $\Pi(s,t)$. Since by (9), the length of a run of green vertices in $\Pi(s,t)$ is greater in distribution than that of black vertices, $G(s,t)$ is greater than $B(s,t)$ in distribution, that is, for any $l\ge 0$,

(11) $$P_\beta(G(s,t)\ge l) \le P(B(s,t)\ge l).$$

Similarly for a lower bound of $B(s,t)$, we can repeat the experiment using the red colour and probability law P_α instead of P_β. Let $R(s,t)$ be the number of red vertices in $\Pi(s,t)$. Then we have for any $l \geq 0$,

$$P_\alpha(R(s,t) \leq l) \leq P(B(s,t) \leq l). \tag{12}$$

We require the following result which together with (11) and (12), gives Lemma 3 immediately.

Lemma 4. *Suppose that $\alpha n \leq s < t \leq \beta n$ and $(t-s)/n \to c$ as $n \to \infty$. Then for any $\varepsilon_1 > 0$, there is $\varrho \in (0,1)$ such that for all large n,*

$$P_\beta(G(s,t) \geq cnx(\beta,m) + \varepsilon_1 n) \leq \varrho^n, \tag{13}$$

$$P_\alpha(R(s,t) \leq cnx(\alpha,m) - \varepsilon_1 n) \leq \varrho^n. \tag{14}$$

In order to show Lemma 4 and (13) in particular, we consider the following. Let $Z_{1,g}$ denote the length of the run of uncoloured vertices with π_k as its right end. Then for $j = 1,2,\ldots$

$$\begin{aligned} P_\beta(Z_{1,g} = j) &= P_\beta(\pi_{k-j} \quad \text{is green and} \quad \pi_{k-j+1},\ldots,\pi_{k-1} \quad \text{are uncoloured}) \\ &= P_\beta(Y_{k-j} = 0, Y_{k-j+1} \geq 1, \ldots, Y_{k-1} \geq 1) \\ &= (1-\beta^m)^{j-1}\beta^m. \end{aligned}$$

Suppose that π_{k-j+1} is uncoloured but π_{k-j} is green, that is, π_{k-j} is the right end of a run of green vertices. Let $Z_{2,g}$ denote the length of this run of green vertices. From (10) we have that for $l \geq 1$, $P_\beta(Z_{2,g} = l)$ is equal to the probability that

$$Y_1' \leq 1, \quad Y_1' + Y_2' \leq 2, \ldots, \quad Y_1' + \ldots + Y_{l-1}' \leq l-1 \quad \text{and} \quad Y_1' + \ldots + Y_l' \geq l+1,$$

where $Y_1', Y_2', \ldots, Y_l'$ are independent and each is the sum of m independent Bernoulli variables X_e and $P_\beta(X_e = 0) = \beta$. Define $Z_{3,g} = Z_{1,g} + Z_{2,g}$. Then given that π_k is uncoloured, $\{Z_{3,g} = \nu\}$ corresponds to the event that the subsequence $\pi_{k-\nu+1}, \ldots, \pi_{k-1}$ of $\Pi(s,t)$ contains exactly one run of green vertices with $\pi_{k-\nu+1}$ being the left end of the run of green vertices. (Note that $\pi_{k-\nu}$ is uncoloured.) Similarly, we have $Z_{1,r}, Z_{2,r}$ and $Z_{3,r}$ for red vertices. For $k = 1,2,3$, define $\mu_{k,r} = E_\alpha[Z_{k,r}]$ and $\mu_{k,g} = E_\beta[Z_{k,g}]$, where E_ζ is the expectation operator in the probability space with law P_ζ. We shall use the following result, to be proved later, to obtain estimates for $R(s,t)$ and $G(s,t)$.

Lemma 5. *There exist positive constants σ_1 and σ_2 such that*

$$E_\beta[\exp(\theta Z_{1,g})] = e^\theta \beta^m / [1 - e^\theta(1-\beta^m)] \qquad \textit{for} \quad \theta < \sigma_1$$

$$E_\beta[\exp(\theta Z_{2,g})] = 1 + \beta^{-m} \frac{x_1(e^\theta)(e^\theta - 1)}{e^\theta(1 - x_1(e^\theta))} \qquad \textit{for} \quad \theta < \sigma_2,$$

where $x_1(y)$ is the smallest positive root of $(x+\beta-x\beta)^m=x/y$. Also,

$$\mu_{1,g}=\beta^{-m}, \qquad \mu_{2,g}=\beta^{-m}x(\beta,m)/(1-x(\beta,m)),$$
$$\mu_{3,g}=\beta^{-m}/(1-x(\beta,m)).$$

If we replace β with α in the above, we obtain corresponding estimates for red vertices.

Corollary 6. *For any $\varepsilon>0$, there is $\varrho\in(0,1)$ such that for all large n,*

$$P(\text{there is a run of black vertices in } \Pi \text{ with length at least } \varepsilon n)\le \varrho^n.$$

Proof of Corollary 6. Let β be such that $\lambda<\beta<1-1/m$. Then by choosing suitable $\theta>0$, there is ϱ_1 in $(0,1)$ so that

$$P_\beta(Z_{2,g}\ge \varepsilon n)\le E_\beta[\exp(\theta Z_{2,g})]\exp(-\varepsilon n\theta)\le \varrho_1^n \tag{15}$$

for all large n. Since $\beta>i/n$ for large n, it follows from (9) that,

$$\begin{aligned}
&P(\text{there is a run of black vertices in } \Pi \text{ with length at least } \varepsilon n)\\
&\quad\le P_\beta(\text{there is a run of green vertices in } \Pi \text{ with length at least} \varepsilon n)\\
&\quad= \sum_{i\ge k\ge \varepsilon n} P_\beta(\text{there is a run of green vertices in } \Pi \text{ with length at least } \varepsilon n\\
&\qquad\qquad\qquad\qquad\qquad\qquad\text{and with } \pi_k \text{ as its right end})\\
&\quad\le \sum_{i\ge k\ge \varepsilon n} P_\beta(Z_{2,g}\ge \varepsilon n)
\end{aligned}$$

which, from (15), is bounded above by $n\varrho_1^n\le\varrho^n$. ∎

Proof of Lemma 4. We shall show (13). Let τ be the number of runs of green vertices in $\Pi(s,t)$ excluding the runs (if any) containing π_s or π_t. Then

$$P_\beta(\tau\ge\nu)\le P_\beta\left(\sum_{k=1}^{\nu} Z_{3,g}^{(k)}\le t-s\right)$$

where the variables $Z_{3,g}^{(k)}$ are independent with the same distribution as $Z_{3,g}$ defined before Lemma 5. For $\varepsilon>0$, let $\nu=\lfloor \varepsilon n+cn/\mu_{3,g}\rfloor$ so that as $n\to\infty, \nu\mu_{3,g}\ge t-s+\eta(\varepsilon)\nu+o(n)$ where $\eta(\varepsilon)>0$. Hence

$$P_\beta(\tau\ge\nu)\le P_\beta\left(\sum_{k=1}^{\nu} Z_{3,g}^{(k)}\le \nu\mu_{3,g}-\eta(\varepsilon)\nu+o(n)\right). \tag{16}$$

Since from Lemma 5, $E_\beta[\exp(\theta Z_{3,g})]$ is defined for θ in an open interval containing the origin, we have as $\theta\to 0$,

$$E_\beta[\exp(-\theta Z_{3,g})]=1-\theta\mu_{3,g}+O(\theta^2)\le\exp(-\theta\mu_{3,g}+O(\theta^2))$$

and so from (16) for positive θ sufficiently close to 0,

$$(17)\qquad \begin{aligned} P_\beta(\tau \geq \nu) &\leq \{E_\beta[\exp(-\theta Z_{3,g})]\exp(\theta\mu_{3,g} - \theta\eta(\varepsilon) + o(1))\}^\nu \\ &\leq \{\exp(-\eta(\varepsilon)\theta + O(\theta^2) + o(1))\}^\nu \leq \varrho_1^\nu \leq \varrho_2^n \end{aligned}$$

where ϱ_2 is in $(0,1)$ and n is large. Suppose that $0 < 2\varepsilon_2 < \varepsilon_1$ where ε_1 is given in Lemma 4. Now from Corollary 6, there is $\varrho_3 \in (0,1)$ so that

$$G(s,t) \leq \sum_{k=1}^{\tau} Z_{2,g} + 2\varepsilon_2 n$$

with probability at least $1 - \varrho_3^n$ for large n. If $\nu = \lfloor \varepsilon_3 n + cn/\mu_{3,g} \rfloor$ where ε_3 is such that $\eta = \varepsilon_1 - 2\varepsilon_2 - \varepsilon_3\mu_{2,g} > 0$, then since $\mu_{2,g}/\mu_{3,g} = x(\beta,m)$,

$$\nu\mu_{2,g} = \varepsilon_3\mu_{2,g}n + cnx(\beta,m) + O(1).$$

Using (17) in which we put $\varepsilon = \varepsilon_3$, we have for all large n,

$$\begin{aligned} &P_\beta(G(s,t) \geq cnx(\beta,m) + \varepsilon_1 n) \\ &\quad\leq P_\beta\left(\sum_{k=1}^{\tau} Z_{2,g} \geq cnx(\beta,m) + \varepsilon_1 n - 2\varepsilon_2 n\right) + \varrho_3^n \\ &\quad\leq P_\beta\left(\sum_{k=1}^{\nu} Z_{2,g} \geq cnx(\beta,m) + \varepsilon_1 n - 2\varepsilon_2 n\right) + \varrho_2^n + \varrho_3^n \\ &\quad\leq P_\beta\left(\sum_{k=1}^{\nu} Z_{2,g} \geq \nu\mu_{2,g} + \eta n + O(1)\right) + \varrho_2^n + \varrho_3^n. \end{aligned}$$

Using methods similar to those used in showing (17) from (16), we have

$$P_\beta\left(\sum_{k=1}^{\nu} Z_{2,g} \geq \nu\mu_{2,g} + \eta n + O(1)\right) \leq \varrho_4^n, \quad \text{where} \quad \varrho_4 \in (0,1).$$

We now have (13) because

$$P_\beta(G(s,t) \geq cnx(\beta,m) + \varepsilon_1 n) \leq \varrho_2^n + \varrho_3^n + \varrho_4^n \leq \varrho_5^n, \quad \text{for large} \quad n.$$

Our proof of inequality (14) is omitted as it is very similar to that of (13). ∎

Proof of Lemma 5. Note that for $\theta < -\log(1-\beta^m) = \sigma_1 > 0$,

$$E_\beta[\exp(\theta Z_{1,g})] = \prod_{j=1}^{\infty}(1-\beta^m)^{j-1}\beta^m e^{j\theta} = e^\theta\beta^m/[1 - e^\theta(1-\beta^m)].$$

Finding $E_\beta[\exp(\theta Z_{2,g})]$ is not as easy. For $l = 1,2,\ldots$ and $k = 0,1,2,\ldots$, let $P_{l,k}$ stand for the probability that

$$Y_1' \leq k, \quad Y_1' + Y_2' \leq k+1, \ldots, \quad Y_1' + \ldots + Y_{l-1}' \leq k+l-2 \quad \text{and} \quad Y_1' + \ldots + Y_l' \geq k+l.$$

We shall use generating functions to find

$$G(e^{\theta}) = E_{\beta}[\exp(\theta Z_{2,g})] = \sum_{l=1}^{\infty} e^{\theta l} P_{l,1}.$$

Note that we shall adopt the convention that $\binom{l}{j}=0$ for $l<j$. Now

$$P_{1,k} = P_{\beta}(Y_1 \geq k+1) = \sum_{j=k+1}^{m} \binom{m}{j}(1-\beta)^j \beta^{m-j},$$

$$P_{l,k} = \sum_{j=0}^{k} P_{\beta}(Y_1' = j) P_{l-1,k+1-j} = \sum_{j=0}^{k} \binom{m}{j}(1-\beta)^j \beta^{m-j} P_{l-1,k+1-j}, \quad \text{for } l \geq 2.$$

Hence

$$\begin{aligned}
\sum_{l=2}^{\infty}\sum_{k=0}^{\infty} P_{l,k} y^l x^k &= \sum_{l=2}^{\infty}\sum_{k=0}^{\infty}\sum_{j=0}^{k} \binom{m}{j}(1-\beta)^j \beta^{m-j} P_{l-1,k+1-j} y^l x^k \\
&= \sum_{l=2}^{\infty}\sum_{j=0}^{\infty}\sum_{k=j}^{\infty} \binom{m}{j}(1-\beta)^j \beta^{m-j} P_{l-1,k+1-j} y^l x^k \\
&= \sum_{l=2}^{\infty}\sum_{j=0}^{m}\sum_{k=0}^{\infty} \binom{m}{j}(1-\beta)^j \beta^{m-j} P_{l-1,k+1} y^l x^{k+j} \\
&= \frac{y}{x}(\beta + x - x\beta)^m \sum_{l=1}^{\infty}\sum_{k=1}^{\infty} P_{l,k} y^l x^k.
\end{aligned}$$

Let $F(x,y)=\sum_{l=1}^{\infty}\sum_{k=0}^{\infty} P_{l,k} y^l x^k$, and $H(x)=\sum_{k=0}^{\infty} P_{1,k} x^k$. Then

$$F(x,y) - yH(x) = \frac{y}{x}(\beta + x - x\beta)^m \left(F(x,y) - \sum_{l=1}^{\infty} P_{l,0} y^l \right), \quad \text{and} \tag{18}$$

$$H(x) = [1 - (\beta + x - x\beta)^m]/(1-x). \tag{19}$$

Since $P_{l,k} \in [0,1]$, $F(x,y)$ converges for $|x|$ and $|y|$ in $(0,1)$. Now if $x=x(y)$ is a root of the equation

$$(\beta + x - x\beta)^m = x/y, \tag{20}$$

then from (18) and (19),

$$\sum_{l=1}^{\infty} P_{l,0} y^l = yH(x) = (y - x(y))/(1 - x(y)).$$

Let $x_1(y)$ be the smallest positive root of equation (20). Note that

$$x_1(y) = \sum_{k=1}^{\infty} \frac{1}{k} \binom{mk}{k-1} y^k (1-\beta)^{k-1} \beta^{mk-k+1}.$$

and that x_1 is the only root of (20) such that $x_1(y) \to 0$ as $y \to 0$, giving

$$\lim_{y\to 0} \sum_{l=1}^{\infty} P_{l,0} y^l = \lim_{y\to 0} (y - x_1(y))/(1 - x_1(y)) = 0.$$

Therefore, we have

$$\sum_{l=1}^{\infty} P_{l,0} y^l = (y - x_1(y))/(1 - x_1(y)).$$

Since $P_{l,0} = \beta^m P_{l-1,1}$ for $l \geq 2$ and since $P_{1,0} = P_\beta(Y_1 \geq 1) = 1 - \beta^m$.

$$G(y) = \sum_{l=1}^{\infty} P_{l,1} y^l = \frac{\beta^{-m}}{y} \sum_{l=1}^{\infty} P_{l,0} y^l - \beta^{-m} P_{1,0} = 1 + \beta^{-m} \frac{x_1(y)(y-1)}{y(1 - x_1(y))}.$$

Since $x_1(1) \neq 1$, we have $\lim_{y\to 1} x_1(y)(y-1)/[y(1-x_1(y))] = 0$, and so $G(1) = 1$, implying that the variable $Z_{2,g}$ is non-defective. Also, $x_1(y)$ is defined for $y < \zeta_0$ where ζ_0 is the unique number such that the line $z = x/\zeta_0$ is tangent to the curve $z = (x + \beta - x\beta)^m$. We find that

$$\zeta_0 = \frac{[(m-1)/(m\beta)]^{m-1}}{(1-\beta)m} > 1.$$

Hence $E_\beta[\exp(\theta Z_{2,g})] = G(e^\theta)$ is defined for $\theta < \sigma_2$ where $\sigma_2 = \log \zeta_0 > 0$. Note that $\mu_{1,g} = \beta^{-m}$ is obvious and $\mu_{2,g} = G'(1)$. Now

$$G(y) = 1 + \beta^{-m} \left(\frac{1-y}{y} - (1-y)/(y - yx_1(y)) \right)$$

and so $\frac{d}{dy} G(y) = \beta^{-m}(-y^{-2} + 1/(y - yx_1(y)) - (1-y)\frac{d}{dy}(y - yx_1(y))^{-1}$, giving that $G'(1) = \beta^{-m} + \beta^{-m}/(1 - x_1(1))$. Since $x_1(1) = x(\beta, m)$, we have

$$\mu_{2,g} = \beta^{-m} x(\beta,m)/(1 - x(\beta,m)), \quad \mu_{3,g} = \mu_{1,g} + \mu_{2,g} = \beta^{-m}/(1 - x(\beta,m)).$$

The other half of the lemma is proved similarly. ∎

5. Proof of Theorem 4

We shall use the model $d'(m,n)$ defined in Section 2 to show Theorem 4. Note that for vertex π_i in the DFS tree with root π_1 in $D(m,n)$, $L'(i,m,n)$ in $D(m,n)$ is equal to $L'(i,m,n)$ in $D'(m,n)$. Consider the situation after π_i has just been picked. Let j'' be the largest integer less than $(1-1/m)n$ so that $\pi_{j''}$ is non-black. Let i'' be the smallest integer not less than $(1-1/m)n$ so that $\pi_{i''}$ is non-black. Then since the vertices $\pi_{j''+1},\dots,\pi_{i''-1}$ are all black, we have that

$$L'(i,m,n) \le L'(j'',m,n) + (i - i'' + 1) \tag{21}$$

$$L'(i,m,n) \ge L'(j'',m,n). \tag{22}$$

Let $\gamma = \gamma(\zeta)$ be the l;argest root of $z = \exp(\zeta(z-1))$. Let i be as given in the hypothesis of the theorem, that is, $i/n \to \lambda$ in $(1-1/m, 1-y(m))$. Let j' be the unigue integer such that

$$1 - (j'+1)/n < (1-i/n)\gamma(m(1-i/n)) \le 1 - j'/n.$$

The following lemma contains some observations on j' which will be useful later.

Lemma 7. *Suppose that ε_1, ε_2 and ε_3 are positive constants.*

(i) *If $l = \lceil j' - \varepsilon n \rceil$, then for all large n,*

$$m(j'-l)/n + \log((1-j'/n)/(1-l/n)) \ge \eta = \eta(\varepsilon) > 0. \tag{23}$$

(ii) *If j satisfies $j' + \varepsilon_1 n \le j \le i - \varepsilon_2 n$, then for all large n,*

$$m(i-j)/n + \log((1-i/n)/(1-j/n)) \le -\zeta = -\zeta(\varepsilon_1, \varepsilon_2) < 0. \tag{24}$$

(iii) *if k satisfies $j' + 1 \le 1$, then*

$$m(k-j')/n + \log((1-k/n)/(1-j'/n)) \ge 0. \tag{25}$$

Proof of Lemma 7. Suppose that γ is the largest root of $z = e^{\xi(z-1)}$. Then by plotting the functions $f(z) = z$ and $f(z) = e^{\xi(z-1)}$, we see that for $\zeta < 1$,the gradient of $f(z) = e^{\zeta(z-1)}$ is bigger than 1 when $z = \gamma$, and so

$$z > e^{\xi(z-1)} \qquad \text{for} \quad z \in (1,\gamma), \tag{26}$$

$$z < e^{\xi(z-1)} \qquad \text{for} \quad z \in (\gamma,\infty), \tag{27}$$

$$\gamma\xi = \frac{d}{dz} e^{\xi(z-1)} \Big|_{z=\gamma} > 1. \tag{28}$$

Also, if $\psi = \zeta\gamma$, then $\log\psi - \log\xi = \psi - \xi$, giving that

$$\frac{d\psi}{d\xi} = \frac{\psi(\xi - 1)}{\xi(\psi - 1)} < 0 \tag{29}$$

and that ψ is a decreasing function in ξ. To show (23), note that

$$\begin{aligned} m(j' - l)/n + \log((1 - j'/n)/(1 - l/n)) \\ &= m(j'/n - l/n) - \log(1 + (j'/n - l/n)/(1 - j'/n)) \\ &\geq m(j'/n - l/n) - (j'/n - l/n)/(1 - j'/n) \\ &= (j'/n - l/n)[m - 1/(1 - j'/n)]. \end{aligned}$$

Since $m(1-i/n) = m(1-\lambda) + o(1)$ and $m(1-\lambda) < 1$,

$$1 - j'/n \geq (1 - i/n)\gamma(i/n) + o(1) = (1 - \lambda)\gamma(m(1 - \lambda)) + o(1).$$

Taking $\xi = m(1-\lambda)$, it follows from (28) that $1 - j'/n > 1/m$, and (23) now follows. To show (24), note that $(1-j/n)/(1-i/n) \in (1+c_2, \gamma(m(1-\lambda)) - c_1)$ where c_1 and c_2 are positive and c_1 depends on ε_1 and c_2 depends on ε_2. Hence from (26), taking $\xi = m(1-\lambda)$, there is $c_3 > 1$ so that

$$\begin{aligned} (1 - j/n)/(1 - i/n) &> c_3 e^{m(1-\lambda)[1-j/n)/(1-i/n)-1]} \\ &= c_3 e^{m(i/n - j/n)} + o(1). \end{aligned}$$

Inequality (24) now follows. To show (25), let $\xi = m(1-i/n)$ and $\xi_2 = (1-k/n)$ and γ_1 and γ_2 be the largest roots of $\gamma = e^{\xi(\gamma-1)}$ where $\xi = \xi_1$ and ξ_2 respectively. As $\xi_1 < \xi_2$, (29) now gives

$$(1 - j'/n)/(1 - k/n) \geq \gamma_1\varepsilon_1/\varepsilon_2 > \gamma_2.$$

By virtue of (27), taking $\xi = \xi_2$,

$$(1 - j'/n)/(1 - k/n) \leq \exp\{\xi_2[(1 - j'/n)/(1 - k/n)01]\} = e^{m(k/n - j'/n)}.$$

Inequality (25) now follows. ∎

The following lemma contains some useful estimates of j'' and i''. We shall use these estimates and (21), (22) to prove Theorem 4 first; proof of the lemma will be given later.

Lemma 8. *For any $\varepsilon > 0$, there is ϱ in $(0,1)$ such that for all large n,*

$$P(|j' - j''| \leq \varepsilon n) \geq 1 - \varrho^n \tag{30}$$

$$P(|i - i''| \leq \varepsilon n) \geq 1 - \varrho^n. \tag{31}$$

Proof of Theorem 4. From (21), we have for any $\varepsilon > 0$,

$$\begin{aligned}
&P(L'(i,m,n) \geq b(j'/n,m)n + \varepsilon n) \\
&\leq P(L'(j'',m,n) + (i - i'' + 1) \geq b(j'/n,m)n + \varepsilon n)) \\
&\leq P(L'(j'',m,n) \geq b(j''/n,m)n + (b(j'/n,m) - b(j''/n,m)n + \varepsilon n - (i - i'' + 1)) \\
&\leq P(L'(j'',m,n) \geq b(j''/n,m)n + \varepsilon_1 n) \\
&+ P(b(j'/n,m) - b(j''/n,m) \leq -\varepsilon_2) + P(i - i'' + 1 \geq \varepsilon_3 n),
\end{aligned}$$

where ε_1, ε_2, ε_3 are positive and satisfy $\varepsilon_1 = \varepsilon_2 - \varepsilon_3$. Now since from Lemma 8, j''/n converges in probability to the constant $\lim_{n\to\infty} j'/n$ which is less than $1 - 1/m$, we have from Theorem 3 that for large n,

$$P(L'(j'',m,n) \geq b(j''/n,m)n + \varepsilon_1 n) \leq \varrho_1^n.$$

Since for any $\varepsilon_2 > 0$, $|b(j''/n,m) - b(j'/n,m)| \leq \varepsilon_2$ implies that there is $\varepsilon_4 > 0$ such that $|j' - j'' \leq \varepsilon_4 n$. Hence from (30),

$$P(b(j'/n,m) - b(j''/n,m) \leq -\varepsilon_2) \leq \varrho_2^n.$$

Thus, together with (31), there is ϱ in $(0,1)$ such that

$$P(L'(i,m,n) \geq b(j'/n,m)n + \varepsilon n) \leq \varrho^n. \tag{32}$$

From (22), for any $\varepsilon > 0$,

$$\begin{aligned}
&P(L'(i,m,n) \leq b(j'/n,m)n - \varepsilon n) \\
&\leq P(L'(j'',m,n) \leq b(j'/n,m)n - \varepsilon n) \\
&= P(L'(j'',m,n) \leq b(j''/n,m)n + (b(j'/n,m) - b(j''/n,m))n - \varepsilon n) \\
&\leq P(L'(j'',m,n) \leq b(j''/n,m)n - \varepsilon_1 n) + P((b(j'/n,m) - b(j''/n,m))n - \geq \varepsilon_2 n)
\end{aligned}$$

where ε_1 and ε_2 are positive and $\varepsilon_1 = \varepsilon - \varepsilon_2$. But as in showing (32), each term above is less than ϱ_3^n for large n. Hence, there is ϱ in $(0,1)$ so that

$$P(L'(i,m,n) \leq b(j'.m)n - \varepsilon n) \leq \varrho^n. \tag{33}$$

Theorem 4 now follows from (32) and (33) because $j'/n \to 1(1-\lambda)\gamma$ where $\gamma = \gamma(m(1-\lambda))$ is the largest root of $\gamma = \exp((1-\lambda)(\gamma - 1))$ and $b(j',m)$ tends to $b(1-(1-\lambda)\gamma, m)$ as $n \to \infty$. ∎

It therefore remains to show Lemma 8. For $k > j$, let $U_{n,j}(k)$ be the number of edge inspections required by DFS to get to π_k from π_j. That is, $U_{n,j}(k) = U_n(k) - U_n(j)$ where $U_n(l)$ is defined in Section 2. Our proof of Lemma 8 requires the next lemma which follows immediately from Lemma 1.

Lemma 9. *Suppose that $k > j$ and that k/n and j/n converge to constants which are less than δ and $\delta \in (0,1)$. Then there is ϱ in $(0,1)$ such that for all large n,*

$$P(|U_{n,j}(k) + n\log(1 - k/n) - n\log(1 - j/n)| \geq \varepsilon n) \leq \varrho^n.$$ ∎

Proof of Lemma 8. Let $l = [j' - \varepsilon n]$. Consider picking vertices after $pi_{j'}$ has been picked. We know that there are $m(j'-l) - U_{n,l}(j')$ (if it si not negative) unused edges directed from $\pi_l, \ldots, \pi_{j'}$. Thus $j'' \geq l$ if and only if not all of these edges have been inspected while picking $\pi_{j'+1}, \ldots, \pi_i$. Hence

$$\begin{aligned} P(j'' > j' - \varepsilon n) = P(m(j'-l) - U_{n,l}(j') \geq 0 \quad &\text{and for} \quad k = j'+1 \quad \text{to} \quad i, \\ U_{n,j'}(k) \leq m(k-j') + m(j'-l) - U_{n,l}(j')) \\ \geq 1 - \sum_{k=j'+1}^{i} P(U_{n,j'}(k) > m(k-j') + m(j'-l) - U_{n,l}(j')) \\ - P(m(j'-l) - U_{n,l}(j') < 0). \end{aligned}$$

From (23) there is η such that for large n,

$$m(j'-l)/n + \log((1-j'/n)/(1-l/n)) \geq \eta(\varepsilon) > 0.$$

Also for any ε_1 in $(0, \eta(\varepsilon))$, we have from Lemma 9 that for large n,

$$P(U_{n,l}(j') + n\log((1-j'/n)/(1-l/n)) \geq \varepsilon_1 n) \leq \varrho_1^n.$$

Hence, with (25), Lemma 9 and the above,

$$\begin{aligned} (34) \quad P(j'' > j' - \varepsilon n) &\geq 1 - \varrho_1^n - \sum_{k=j'+1}^{i} P(U_{n,j'}(k) > m(k-j') + n(\eta(\varepsilon) - \varepsilon_1)) \\ &\geq 1 - \varrho_1^n - \sum_{k=j'+1}^{i} P(U_{n,j'}(k) + n\log((1-k/n)/(1-j'/n)) > n(\eta(\varepsilon) - \varepsilon_1)) \\ &\geq 1 - \varrho_1^n - n\varrho_2^n \geq 1 - \varrho^n, \end{aligned}$$

for large n. Also

$$\begin{aligned} P(j'' \geq j' + \varepsilon n \quad &\text{or} \quad i'' \leq i - \varepsilon n) \\ &\leq P(\text{there is } j \text{ so that } j' + \varepsilon n \leq j < i - \varepsilon n \text{ and } U_{n,j}(i) \leq m(i-j)) \\ &\leq \sum_{j=l}^{h} P(U_{n,j}(i) \leq m(i-j)) \end{aligned}$$

where $h = \lceil 1 - \varepsilon n \rceil$ and $l = \lfloor j' + \varepsilon n \rfloor$. From Lemma 9 and (24), there is $\eta > 0$ so that for $l \leq j \leq h$ and for large n,

$$P(U_{n,j}(i) \leq m(i-j)) = P(U_{n,j}(i) + n\log((1-i/n)/(1-j/n) \leq -\eta n) \leq \varrho_3^n,$$

which implies that

$$P(j'' \geq j' + \varepsilon n \quad \text{or} \quad i'' \leq i - \varepsilon n) \leq n\varrho_3^n \leq \varrho^n.$$

Inequalities (30) and (31) follow from (34) and the above. ∎

Acknowledgement. I would like to thank the referees for their helpful comments and suggestions. I would like to thank them for pointing out the various typographical errors in the original manuscript.

References

[1] AJTAI, M., KOMLÓS, J., and SZEMERÉDI, E.: The longest path in a random graph, *Combinatorica* **1** (1981), 1–12.

[2] BOLLOBÁS, B. *Random Graphs*, Academic Press, (1985).

[3] FENNER, T. I., and FRIEZE, A. M. On large matchings and cycles in sparse random graphs, *Discrete Mathematics* **59** (1986), 243–256.

[4] FERNANDEZ DE LA VEGA, W. Long paths in random graphs, *Studia Sci. Math. Hung.* **14** (1979), 335–340.

[5] GIBBONS, A. *Algorithmic Graph Theory*, Cambridge University Press, (1985).

[6] GOURSAT, E. *A Course in Mathematical Analysis*, Vol. 1 Dover Publ., New York, (1959).

[7] SUEN, S. PhD dissertation, University of Bristol, UK, (1985).

[8] SUEN, S. On the largest strong components in m-out digraphs, *Discrete Math* **94** (1991), 45–52.

[9] SUEN, S. On large induced tree and long induced cycles in sparse random graphs, *J. Combinat. Th. Ser. B.* **56** (1992), 250–262.

[10] TARJAN, R. Depth first and linear graph algorithms, *SIAM. J. Comput* **1** (1972), 146–160.

W. C. Stephen Suen

Department of Mathematics
Carnegie Mellon University
Pittsburg, PA 15213 U.S.A.
ss8b@andrew.cmu.edu