Reprinted from JOURNAL OF COMBINATORIAL THEORY, Series B
All Rights Reserved by Academic Press, New York and London

Vol. 48, No. 1, February 1990
Printed in Belgium

The Limiting Probability That α-In, β-Out Is Strongly Connected

C. COOPER

Department of Computing, Mathematics and Statistics,
Polytechnic of North London,
Holloway Road, London N7 8DB, England

AND

A. M. FRIEZE

Department of Mathematics, Carnegie Mellon University,
Pittsburgh, Pennsylvania 15213-3890

Communicated by the Managing Editors

Received June 3, 1987; revised May 11, 1988

We consider a random digraph $D_{\alpha,\beta}(n)$ with vertex set $\{1, 2, ..., n\}$ in which each vertex v independently chooses α random arcs entering v and β random arcs leaving v. We compute the limiting probability that $D_{\alpha,\beta}(n)$ is strongly connected as n tends to infinity. This solves an open problem from [2].

1. INTRODUCTION

This paper is concerned with the connectivity of a particular model of a random digraph. Let α, β be positive real constants. The random digraph $D_{\alpha,\beta}(n)$ is constructed as follows: it has vertex set $V_n = \{1, 2, ..., n\}$ and each $v \in V_n$ randomly chooses 2 sets of vertices IN(v) and OUT(v) as follows. For IN(v), choose $x_1, x_2, ..., x_{\lfloor \alpha \rfloor} \in V_n$ independently of each other and then with probability $\alpha - \lfloor \alpha \rfloor$ choose a further random vertex in V_n. OUT(v) is independently chosen in the same manner.

$$\text{The arcs of } D_{\alpha,\beta}(n) \text{ are } \{(v, w) : w \in \text{OUT}(v), v \in V_n\} \cup \{(w, v) : w \in \text{IN}(v), v \in V_n\}.$$

Fenner and Frieze [2] considered a class of digraphs $D_k(n)$ for fixed

0095-8956/90 $3.00

positive integer k. $D_k(n)$ is closely related to $D_{k,k}(n)$. One of the results of that paper is the following:

$$\lim_{n\to\infty} \Pr(D_k(n) \text{ is } k\text{-strongly connected}) = 1 \text{ for } k \geq 2.$$

(k-strongly connected means that one must remove at least k vertices before the remaining digraph is no longer strongly connected. A digraph is strongly connected if there is a directed path joining each pair of vertices.)

The case $k=1$ was left unresolved by that paper. The following theorem includes this case.

Thus far we have described α, β as constants. In our theorem we will also deal with the case where α say, is a function of n that tends to 1 as n tends to infinity.

THEOREM 1. *Let*

(a) $\alpha = 1-\omega/n$ *where* $\omega=\omega(n)\to\infty$ *then*

$$\lim_{n\to\infty} \Pr(D_{\alpha,\beta}(n) \textit{ is strongly connected}) = 0.$$

(b) $\alpha = 1-a/n$, $\beta = 1-b/n$ *for* $a, b \geq 0$ *then*

$$\lim_{n\to\infty} \Pr(D_{\alpha,\beta}(n) \textit{ is strongly connected}) = (1-e^{-1})^2\, e^{-(a+b)e^{-1}}.$$

(c) $\alpha = 1-a/n$ *for* $a \geq 0$, $1 \leq \beta \leq 2$ *then*

$$\lim_{n\to\infty} \Pr(D_{\alpha,\beta}(n) \textit{ is strongly connected}) = e^{-ae^{-1}}(1-e^{-\beta})(1-(2-\beta)\,e^{-1}).$$

(d) $\alpha = 1-a/n$ *for* $a \geq 0$, $\beta \geq 2$ *then*

$$\lim_{n\to\infty} \Pr(D_{\alpha,b}(n) \textit{ is strongly connected}) = e^{-ae^{-1}}(1-e^{-\beta}).$$

(e) $1 \leq \alpha \leq 2$, $\beta \geq 2$ *then*

$$\lim_{n\to\infty} \Pr(D_{\alpha,\beta}(n) \textit{ is strongly connected}) = 1-(2-\alpha)\,e^{-\beta}.$$

(f) $1 \leq \alpha$, $\beta \leq 2$, *then*

$$\lim_{n\to\infty} \Pr(D_{\alpha,\beta}(n) \textit{ is strongly connected}) = (1-(2-\alpha)\,e^{-\beta})(1-(2-\beta)\,e^{-\alpha}).$$

We first prove Theorem 1 for the special case where $\alpha=\beta=1$, having done this it will be easy to show how to obtain the complete result.

The next section introduces connectivity blocking substructures and computes the probability that one exists. Section 3 deals with the main problem in showing that nothing else is likely to prevent strong connectivity.

2. Small Strong Components

From now on let D denote the random digraph $D_{1,1}(n)$ and A denote its arc-set. Let now $\mathrm{OUT}(v)=\{\mathrm{out}(v)\}$ and $\mathrm{IN}(v)=\{\mathrm{in}(v)\}$. Let $A_{\mathrm{out}}=\{(v, \mathrm{out}(v)): v\in V_n\}$ and let A_{in} be defined similarly. Let D_{out} be the sub-digraph (V_n, A_{out}) and let D_{in} be defined similarly. For $S\subseteq V_n$ define

$$\mathrm{IN}(S)=\{\mathrm{in}(v): v\in S\} \text{ and } \mathrm{OUT}(S)=\{\mathrm{out}(v): v\in S\}$$

$$N^+(S)=\{w\notin S: \exists v\in S \text{ s.t. } (v, w)\in A\} \text{ and}$$

$$N^-(S)=\{w\notin S: \exists v\in S \text{ s.t. } (w, v)\in A\}.$$

The subgraph D_{out} constitutes a random functional digraph. The properties of such digraphs are well established (see, e.g., Bollobas [1]). The components are unicyclic and the directed edges of any component are "directed towards" the unique cycle. Similar results obtain for D_{in}.

If D is not strongly connected then there exists $S\subseteq V_n$, $1\leqslant |S|\leqslant n-1$ such that $N^+(S)=N^-(V_n-S)=\varnothing$.

Let $C_{\mathrm{out}}=\{C\subseteq V_n: \{(v, \mathrm{out}(v)): v\in C\}$ is a cycle of $D\}$ and let C_{in} be defined similarly.

Lemma 2.1. *Suppose* $S\subseteq V_n$.

If $N^+(S)=\varnothing$ *then* $\exists C\in C_{\mathrm{out}}$ *such that* $C\subseteq S$.

If $N^-(S)=\varnothing$ *then* $\exists C\in C_{\mathrm{in}}$ *such that* $C\subseteq S$.

Proof. We need only consider the case $N^+(S)=\varnothing$. If $v\in S$, the unicyclic component of D_{out} which contains v contains a unique cycle. As $N^+(S)=\varnothing$ this cycle must be contained in S. ∎

We now define cycle sets $\hat{C}_{\mathrm{out}}$, $\hat{C}_{\mathrm{in}}$ as follows:

$$\hat{C}_{\mathrm{out}}=\{C\in C_{\mathrm{out}}: C\cap \mathrm{in}(V_n)=\varnothing, |C|\leqslant \sqrt{n}/\log n\},$$

$$\hat{C}_{\mathrm{in}}=\{C\in C_{\mathrm{in}}: C\cap \mathrm{out}(V_n)=\varnothing, |C|\leqslant \sqrt{n}/\log n\},$$

Let $\hat{C}=\hat{C}_{\mathrm{in}}\cup\hat{C}_{\mathrm{out}}$. We note that if $\hat{C}\neq\varnothing$ then the digraph is not strongly connected. For example if $C\in\hat{C}_{\mathrm{out}}$ then $N^+(C)=\varnothing$ as C is a cycle in D_{out} and $C\cap \mathrm{in}(V_n)=\varnothing$.

We further define sets

$$C'_{\text{out}} = \{C \in C_{\text{out}} : \exists S \text{ s.t. } C \subseteq S,\ N^+(S) = \varnothing \text{ and } |S| \leqslant \sqrt{n}/\log n\},$$
$$C'_{\text{in}} = \{C \in C_{\text{in}} : \exists S \text{ s.t. } C \subseteq S,\ N^-(S) = \varnothing \text{ and } |S| \leqslant \sqrt{n}/\log n\}.$$

Let $C' = C'_{\text{in}} \cup C'_{\text{out}}$. We go on to show that almost every (a.e.) D has a large strongly connected component and that in a.e. D, $C' = \hat{C}$. Note that $\hat{C} \subseteq C'$ always.

LEMMA 2.2. *If D is not strongly connected then at least one of the following events occur*:

(a) $E_1 = \{\exists S: \sqrt{n}/\log n < |S| \leqslant n/2$ *and either* $N^+(S) = \varnothing$ *or* $N^-(S) = \varnothing\}$

(b) $E_2 = \{C' \neq \hat{C}\}$

(c) $E_3 = \{\hat{C} \neq \varnothing\}$.

Proof. If D is not strongly connected there is a se set S, $|S| \leqslant n/2$ such that $N^+(S) = \varnothing$ or $N^-(S) = \varnothing$. If E_1 does not occur then assume w.l.o.g. that $N^+(S) = \varnothing$ and $|S| \leqslant \sqrt{n}/\log n$. By Lemma 2.1 S contains a cycle $C \in C'_{\text{out}}$. Either $C \in \hat{C}$ and E_3 occurs or $C \in C' - \hat{C}$ and E_2 occurs. ∎

LEMMA 2.3. *The limiting probabilities for the events in Lemma 2.2 are*:

(a) $\lim_{n \to \infty} \Pr(E_1) = 0$,

(b) $\lim_{n \to \infty} \Pr(E_2) = 0$,

(c) $\lim_{n \to \infty} \Pr(E_3) = 1 - (1 - e^{-1})^2$.

Proof. (a) We prove in the next section that for a.e. D, vertex 1 is connected by directed paths to and from all but at most $4(\log n)^3$ vertices.

(b) To show that $\Pr(E_2) \to 0$ we count the expected number N of cycles in C'-$\hat{C}$. This satisfies

$$N \leqslant 2 \sum_{s=1}^{\lfloor \sqrt{n}/\log n \rfloor} \sum_{k=1}^{s} \binom{n}{s}\binom{s}{k} \frac{(k-1)!}{n^k} \left(\frac{s}{n}\right)^{s-k} \left(\frac{n-s}{n}\right)^{n-s} \left(1 - \left(\frac{n-k}{n}\right)^{s}\right). \tag{2.1}$$

To see this, suppose S satisfies $N^+(S) = \varnothing$ and $|S| = s$. From S choose a cycle C of size k contained in A_{out}. The probability of such a cycle is $(k-1)!/n^k$. The remaining $s-k$ vertices make their out choices in S and no vertex outside S makes its in choice in S. As $C \in C' - \hat{C}$ at least one vertex

$v \in S$ chooses in(v) from C. The case $N^-(S)=\varnothing$ is covered by a similar argument. Thus for large n

$$N \leqslant 2 \sum_{s=1}^{\lfloor \sqrt{n}/\log n \rfloor} \binom{n}{s} s! \frac{(n-s)^{n-s}}{n^n} \frac{s}{n} \sum_{s-k=0}^{s-1} \frac{s^{s-k}}{(s-k)!}$$
$$\leqslant 3 \sum_{s=1}^{\lfloor \sqrt{n}/\log n \rfloor} \frac{s}{n} = O((\log n)^{-2}).$$

Thus $\Pr(E_2) \to 0$.

(c) Let $X = |\hat{C}_{\text{in}}| + |\hat{C}_{\text{out}}|$. We compute $\Pr(X=0)$. If $C_1 \in \hat{C}_{\text{in}}$, $C_2 \in \hat{C}_{\text{out}}$ then $C_1 \cap C_2 = \varnothing$ for if not $\exists c \in C_2$ such that $c = \text{inv}(v)$ for some $v \in V_1$ which contradicts $C_2 \cap \text{in}(V_n) = \varnothing$. Let $E_t(X)$ be the tth factorial moment of X. We show that

$$\lim_{n \to \infty} E_t(X) = \left(2 \log\left(\frac{e}{e-1}\right)\right)^t$$

and so (see, e.g., Bollobas [1a]) X is asymptotically Poisson with mean $2\log(e/(e-1))$ and so $\Pr(X=0) \to (1-e^{-1})^2$.

$$E_t(X) = \sum_{i=0}^{t} \binom{t}{i} \sum_{k=i}^{n} \sum_{s=t-i}^{n-k} \sum_{\substack{k_1+\cdots+k_i=k \\ s_{i+1}+\cdots+s_t=s}} \binom{n}{k_1, ..., k_i} \binom{n-k}{s_{i+1}, ..., s_t}$$
$$\times \left(\frac{n-k}{n}\right)^{n-s} \left(\frac{n-s}{n}\right)^{n-k} \prod_{j=1}^{i} \frac{(k_j-1)!}{n^{k_j}} \prod_{j=i+1}^{t} \frac{(s_j-1)!}{n^{s_j}},$$

where $k_1, ..., k_i$ are the sizes of i cycles from $\hat{C}_{\text{in}}$ an $s_{i+1}, ..., s_t$ are the sizes of $t-i$ cycles from $\hat{C}_{\text{out}}$. As k_j, s_j are at most $\sqrt{n}/\log n$ and t is fixed, we have for example

$$\frac{n^k}{k_1! \cdots k_i!}\left(1 - \frac{t^2}{(\log n)^2}\right) \leqslant \binom{n}{k_1 \cdots k_i} \leqslant \frac{n^k}{k_1! \cdots k_i!}.$$

We thus replace the multinomial coefficients by $(n^{s+k}/(k_1! \cdots s_t!))$ $(1+o(1))$ and then

$$E_t(X) = (1+o(1)) \sum_{i=0}^{t} \binom{t}{i} \sum_{k=i}^{\infty} \sum_{s=t-i}^{\infty} \sum_{k_1, ..., s_t} \frac{e^{-k_1}}{k_1} \cdots \frac{e^{-s_t}}{s_t}$$
$$= (1+o(1)) \sum_{i=0}^{t} \binom{t}{i} \left(\log\left(\frac{e}{e-1}\right)\right)^t$$
$$= (1+o(1)) \left(2 \log\left(\frac{e}{e-1}\right)\right)^t. \quad \blacksquare$$

Theorem 1 (for $\alpha = \beta = 1$) follows immediately from Lemmas 2.2 and 2.3 since

$$\Pr(E_3) \leqslant \Pr(D(n) \text{ is not strongly connected}) \leqslant \Pr(E_1 \cup E_2 \cup E_3). \quad (2.2)$$

If we suppress all loops in our digraph so that in(v) and out(v) are chosen randomly from $V_n - \{v\}$ for each v we obtain a random digraph D'.

COROLLARY 1. $\lim_{n\to\infty} \Pr(D' \textit{ is strongly connected}) = ((1-e^{-1})e^{e^{-1}})e^{e^{-1}})^2$.

Proof. If we suppress loops then the summation over the cyle sizes in $E_t(X)$ is from $2 \leqslant k_j, s_j \leqslant \sqrt{n}/\log n$. ∎

3. AN ALGORITHM

We must show that $\Pr(E_1) \to 0$ in Lemma 2.2a. In order to do this we consider the following sets:

$$X_+ = \{j \in V_n : D \text{ contains a directed path from 1 to } j\}$$
$$X_- = \{j \in V_n : D \text{ contains a directed path from } j \text{ to } 1\}.$$

The main result of this section is

THEOREM 3.1. $\lim_{n\to\infty} \Pr(\min\{|X_+|, |X_-|\} < n - 2(\log n)^3) = 0.$

The proof of this theorem is based on the analysis of Algorithm CONNECT, which attempts to construct a subset of X_+ of size at least $n - 2(\log n)^3 n$ in the following manner. During Pass j of the algorithm there are two Steps, 1 and 2. Step 1 is an iterative process whereby at iteration i the set $V(i-1, j)$ of vertices not so far identified as belonging to X_+ is examined to determine those $v \in V(i-1, j)$ with in(v) $\notin V(i-1, j)$. These can be added to X_+. However, to simplify the proof, only a subset $S(i, j)$ of a specific size is chosen. This process is continued until the total number of vertices acquired during the current stage is sufficiently large. This occurs at Step i_j^*. The algorithm then goes on to Step 2 where it takes $T(j)$ the set of vertices acquired at Step 1 on this Pass j, and finds the vertices $v \in V(i_j^*, j) \cap \text{out}(T(j))$. A fixed size subset $S(0, j+1)$ of these vertices is then used to start Step 1 at the next Pass($j+1$) of the algorithm.

We now define the constants required by Algorithm CONNECT and in its subsequent analysis.

3.2. *Notation*

The algorithm has three Passes, $j = 1, 2, 3$. We let $s(i, j) = |S(i, j)|$ and $v(i, j) = |V(i, j)|$ for $j = 1, 2, 3$ and $i \geqslant 0$.

In general $v(i-1, j) = v(i, j) + s(i, j)$. When $i = 0$ this has no meaning unless we define $v(-1, j)$ to be $v(0, j) + s(0, j)$, which we do.

At each Pass the size of the set of vertices on which it starts Step 1 will be required to be

$$s(0, 1) = \lceil \sqrt{n}/\log n \rceil, \qquad s(0, 2) = \lceil n^{.72} \rceil, \qquad s(0, 3) = \lceil n/10 \rceil. \tag{3.1a}$$

$V(0, j)$ is the unacquired vertex set at the start of Step 1, Pass j. Thus

$$\begin{aligned} v(0, 1) &= n - s(0, 1) \\ v(0, 2) &= n - (s(0, 1) + l_1 + s(0, 2)) \\ v(0, 3) &= n - (s(0, 1) + l_1 + s(0, 2) + l_2 + s(0, 3)), \end{aligned} \tag{3.1b}$$

where the l_j are the stopping sizes for vertex acquisition at Step 1 of Pass j and are given by

$$l_1 = \lceil n^{.73} \rceil, \qquad l_2 = \lceil .4n \rceil, \qquad l_3 = v(0, 3) - \lceil 2 \log^3 n \rceil.$$

ε_j and t_j are (respectively) constants relating to probability inequalities and bounds on the number of iterations at Pass j and are given by

$$\begin{aligned} \varepsilon_1 &= \log n/n^{.24}, & t_1 &= \lceil 1/\varepsilon_1 \rceil \\ \varepsilon_2 &= s(0, 2)/n, & t_2 &= \lceil 1/\varepsilon_2 \rceil \\ \varepsilon_3 t_3 &= \frac{1}{2} \log\left(\frac{v(-1, 3)}{v(0, 3)}\right), & t_3 &= \left\lceil \frac{4(\log n - 3 \log\log n)}{3 \log(v(-1, 3))/v(0, 3)} \right\rceil \end{aligned}$$

and hence $\varepsilon_3 \geqslant 1/100 \log n$.

3.3. *Algorithm CONNECT*

begin
Initialise:
  $M := \{1, \text{out}(1), \text{out}^2(1), \ldots, \text{out}^{\lceil \sqrt{n}/\log n \rceil}(1)\}$
  if $|M| < \lceil \sqrt{n}/\log n \rceil$ **then** *fail*
  else $S(0, 1) := M$; $V(0, 1) := V_n - S(0, 1)$; $T(1) := \varnothing$
Passes $j = 1, 2, 3$:
  for $j := 1, 2, 3$ **do**
  begin
    Step 1:
      $i := 0$
      while $|T(j)| < l_j$ **do**
      begin
        $i := i + 1$
        $P(i, j) := \{v \in V(i-1, j) : \text{in}(v) \notin V(i-1, j)\}$
        Choose $S(i, j) \subseteq P(i, j)$ where

for $j=1,2$

$$|S(i,j)| = \left\lceil (1-\varepsilon_j)\frac{s(i-1,j)\,v(i-1,j)}{v(i-2,j)} \right\rceil$$

for $j=3$

$$|S(i,3)| = v(i-1,3) - \left\lfloor (1+\varepsilon_3)\frac{v(i-1,3)^2}{v(i-2,3)} \right\rfloor$$

else *fail* {i.e. if $|P(i,j)|$ is too small}

if $|S(i,j)| \leqslant l_j - |T(j)|$ **then** $S := S(i,j)$ **else** S is a random $(l_j - |T(j)|)$-subset of $S(i,j)$

$T(j) := T(j) \cup S$

$V(i,j) := V(i-1,j) - S$

end

if $j=3$ **then stop else**

begin

Step 2:

$i_j^* := i$

$S' := \{v \in V(i_j^*, j) : \exists x \in T(j) \text{ such that } v = \text{out}(x)\}$

if $|S'| < s(0, j+1)$ **then** *fail* **else**

$T(j+1) := \varnothing$

$V(0, j+1) := V(i_j^*, j) - S(0, j+1)$ where $S(0, j+1)$ is a randomly chosen $s(0,j)$-subset of S'

end

end

end

It should be observed that if the algorithms does not fail then the set sizes $s(i,j)$ and $v(i,j)$ are not in fact random. They are the values generated by the following recurrence relations:
for $j=1,2,3$ $s(0,j)$, $v(0,j)$ and $v(-1,j) = s(0,j) + v(0,j)$ are defined by (3.1) and then

$$s(i,j) = \left\lceil (1-\varepsilon_j)\frac{s(i-1,j)\,v(i-1,j)}{v(i-2,j)} \right\rceil \qquad \text{for} \quad j=1,2 \text{ and } i \geqslant 1, \tag{3.2a}$$

$$s(i,3) = v(i,3) - \left\lfloor (1+\varepsilon_3)\frac{v(i-1,3)^2}{v(i-2,3)} \right\rfloor \qquad \text{for} \quad i \geqslant 1, \tag{3.2b}$$

$$v(i,j) = v(i-1,j) - s(i,j) \qquad \text{for} \quad i \geqslant 1. \tag{3.2c}$$

From now on s and v refer explicitly to the values generated by these recurrences. Thus for i sufficiently large, CONNECT will not produce a set $S(i,j)$ of size $s(i,j)$. The next lemma gives the salient properties of these sequences.

LEMMA 3.4. (a) *For* $0 \leqslant i \leqslant t_j$ *and* $j=1,2$,

$$s(i,j) \geqslant s(0,j)\left((1-\varepsilon_j)\frac{v(0,j)}{v(-1,j)}\right)^i \tag{3.3i, j}$$

(b) *For* $0 \leqslant i \leqslant t_j$ *and* $j=1, 2,$

$$\varepsilon_j^2 \frac{s(i,j)\,v(i,j)}{v(i-1,j)} \geqslant n^{\theta_j}, \tag{3.4i, j}$$

where $\theta_1 = .02$ *and* $\theta_2 = .15$.

(c) $v(i,3) \geqslant 2(\log n)^3$ *implies* $\varepsilon_3^2(v(i,3)^2/v(i-1,3)) \geqslant \log n/6250$.

(d) $v(i,3) \leqslant v(0,3)(1+\varepsilon_3)^{i(i+1)/2}(v(0,3)/v(-1,3))^i$ *for* $i \geqslant 0$.

(e) $\sum_{i=1}^{t_j} s(i,j) \geqslant l_j$ *for* $j=1, 2,$ *and* $v(t_3, 3) \leqslant 2(\log n)^3$.

Proof. (a) and (b). Fix $j=1$ or 2. We show by induction on i that for $i \geqslant 0$ we have

$$(3.3i, j), (3.4i, j), \quad \text{and} \quad \frac{s(i,j)}{v(i-1,j)} \leqslant \frac{s(0,j)}{v(-1,j)}.$$

This is easy to check for $i=0$, and inductively assume it is true for some $i \geqslant 0$. Then

$$s(i+1,j) = \left\lceil (1-\varepsilon_j)\frac{s(i,j)\,v(i,j)}{v(i-1,j)} \right\rceil \leqslant \frac{s(i,j)\,v(i,j)}{v(i-1,j)} \quad \text{as} \quad \varepsilon_j \frac{s(i,j)\,v(i,j)}{v(i-1,j)} \to \infty.$$

And so

$$\frac{s(i+1,j)}{v(i,j)} \leqslant \frac{s(i,j)}{v(i-1,j)} \leqslant \frac{s(0,j)}{v(-1,j)}$$

and

$$\begin{aligned} s(i+1,j) &\geqslant (1-\varepsilon_j)\frac{s(i,j)\,v(i,j)}{v(i-1,j)} = (1-\varepsilon_j)\,s(i,j)\frac{v(i-1,j)-s(i,j)}{v(i-1,j)} \\ &\geqslant s(0,j)(1-\varepsilon_j)^{i+1}\left(\frac{v(0,j)}{v(-1,j)}\right)^{i+1}. \end{aligned}$$

Furthermore,

$$\begin{aligned} \frac{s(i,1)\,v(i,1)}{v(i-1,1)} \geqslant s(i+1,1) &\geqslant \frac{\sqrt{n}}{\log n}\left(1-\frac{2\log n}{n^{.24}}\right)^{\lceil n^{.24}/\log n \rceil + 1} \\ &\geqslant \frac{\sqrt{n}}{2e^2 \log n} \end{aligned}$$

and then

$$\varepsilon_1^2 s(i+1,1) \geqslant n^{.02} \log n/(2e^2).$$

Similarly

$$s(i+1,2) \leqslant (1-\varepsilon_2)^{\lceil 1/\varepsilon_2 \rceil + 1} \left(1 - \frac{s(0,2)}{n-o(n)}\right)^{\lceil 1/\varepsilon_2 \rceil + 1} n^{.72}.$$

$\varepsilon_2 = s(0, 2)/n$ and so

$$s(i+1,2) \geqslant n^{.72} e^{-2}/2$$

giving

$$\varepsilon_2^2 s(i+1,2) \geqslant n^{.16} e^{-2}/2,$$

which completes the induction.

(c) We first observe that $v(i+1,3) = \lfloor (1+\varepsilon_3)(v(i,3)^2/v(i-1,3)) \rfloor$ for $i \geqslant 0$ and hence

$$\frac{v(i+1,3)}{v(i,3)} \geqslant \frac{v(i,3)}{v(i-1,3)} + \left(\frac{\varepsilon_3 v(i,3)^2}{v(i-1,3)} - 1\right) v(i,3)^{-1}. \tag{3.5}$$

We show now by induction that $v(i,3) > 2(\log n)^3$ implies

$$\frac{v(i,3)}{v(i-1,3)} \geqslant \frac{v(0,3)}{v(-1,3)} \geqslant \frac{4}{5}. \tag{3.6}$$

This is true when $i = 0$ and assume it true for some $i \geqslant 0$. From (3.5) we have

$$\frac{v(i+1,3)}{v(i,3)} \geqslant \frac{v(0,3)}{v(-1,3)} + \left(\frac{\varepsilon_3 v(i,3)\, v(0,3)}{v(-1,3)} - 1\right) v(i,3)^{-1}. \tag{3.7}$$

Now if $v(i,3) \geqslant v(i+1,3) > 2(\log n)^3$ then $\varepsilon_3 v(i,3) \geqslant (\log n)^2/50$. Hence (3.7) implies (3.6) for $i+1$ completing the induction.

So (3.6) implies

$$\varepsilon_3^2 \frac{v(i,3)^2}{v(i-1,3)} \geqslant \frac{4}{5} \varepsilon_3^2 v(i,3) \geqslant \frac{\log n}{6250}.$$

(d) We first show by induction on i that for $i \geqslant 0$

$$v(i,3) \leqslant (1+\varepsilon_3)^i \frac{v(0,3)}{v(-1,3)} v(i-1,3). \tag{3.8}$$

This is clearly true for $i=0$ and inductively assume it is true for some $i \geqslant 0$. Then

$$v(i+1,3) \leqslant (1+\varepsilon_3)\frac{v(i,3)^2}{v(i-1,3)}$$

implying

$$v(i+1,3) \leqslant (1+\varepsilon_3)^{i+1}\frac{v(0,3)}{v(-1,3)}\, v(i,3)$$

as required. We obtain (d) by iterating (3.8).

(e) For $j=1,2$ and $t \geqslant 1$ we use (a) to show

$$\sum_{i=1}^{t} s(i,j) \geqslant a(t,j) = \sum_{i=1}^{t} (1-\varepsilon_j)^i \left(\frac{v(0,j)}{v(-1,j)}\right)^i s(0,j)$$

$$= (1-\varepsilon_j)\frac{s(0,j)\,v(0,j)}{s(0,j)+\varepsilon_j v(0,j)}\left(1-(1-\varepsilon_j)^t\left(\frac{v(0,j)}{v(-1,j)}\right)^t\right).$$

Then $a(t_1,1) \geqslant n^{.74}/2\log n)^2$ and $a(t_2,2) \geqslant (1-o(1))((1-e^{-2})/2)n \geqslant .4n$. For $j=3$ we use (d) to show

$$v(t_3,3) \leqslant v(0,3)\, e^{\varepsilon_3 t_3/2}(e^{\varepsilon_3 t_3})^{t_3/2} e^{t_3 \log(v(0,3)/v(-1,3))}$$

$$= v(0,3)\, e^{-\log n + 3\log\log n + \log(v(-1,3)/v(0,3))/4} \leqslant 2(\log n)^3. \quad \blacksquare$$

The following lemma is easily derivable from Theorem 1 of Hoeffding [3] and is used to bound probabilities:

LEMMA 3.5. *Let X_i, $i=1, \ldots, m$ be independent random variables taking values in* [0, 1] *and let $E(\sum_{i=1}^{m} X_i) = m\mu$. Then for $\varepsilon \in (0,1)$*

$$\Pr\left(\sum_{i=1}^{m} X_i \leqslant (1-\varepsilon)\, m\mu\right) \leqslant e^{-(\varepsilon^2/2)m\mu}$$

$$\Pr\left(\sum_{i=1}^{m} X_i \geqslant (1+\varepsilon)\, m\mu\right) \leqslant e^{-(\varepsilon^2/3)m\mu}.$$

We can prove

LEMMA 3.6. *Given i successful iterations of Step* 1 *of Algorithm CONNECT during Pass j*

(a) *for* $j=1,2$

$$\Pr\left(|S(i+1,j)|\leqslant\left\lceil(1-\varepsilon_j)\frac{v(i,j)\,s(i,j)}{v(i-1,j)}\right\rceil\right)\leqslant e^{-(\varepsilon_j^2/2)(v(i,j)s(i,j)/v(i-1,j))}$$

(b) *for* $j=3$

$$\Pr\left(|V(i+1,3)|\geqslant\left\lfloor(1+\varepsilon_3)\frac{v(i,3)^2}{v(i-1,3)}\right\rfloor\right)\leqslant e^{-(\varepsilon_3^2/3)(v(i,3)^2/v(i-1,3))}.$$

Proof. (a) Suppose we have just successfully completed the ith iteration of Step 1 at Pass j. Let $R(i,j)=P(i,j)-S(i,j)$ where $P(i,j)=\{v\in V(i-1,j):\text{inv}(v)\notin V(i-1,j)\}$ as defined in CONNECT and thus

$$R(i,j)=\{v\in V(i-1,j):\text{inv}(v)\notin V(i-1,j),\, v\notin S(i,j)\}$$

is the set of vertices found to be attached but not selected at the ith iteration. For $v\in V(i,j)$ define a random variable Z_v as follows

$$Z_v=1 \qquad \text{if in}(v)\notin V(i,j)$$
$$Z_v=0 \qquad \text{otherwise.}$$

If $v\in R(i,j)$ then $\Pr(Z_v=1)=1$. If $v\in V(i,j)-(i,j)$ then $\Pr(Z_v=1)=s(i,j)/v(i-1,j)$ as $\text{in}(v)$ can only be chosen from $V(i-1,j)=V(i,j)\cup S(i,j)$ in this case. Now consider a random variable X_v such that

$$X_v=Z_v, \qquad v\in V(i,j)-R(i,j)$$
$$X_v=1 \text{ with probability } \frac{s(i,j)}{v(i-1,j)} \qquad \text{for} \quad v\in R(i,j).$$

Certainly $\sum_{v\in V(i,j)} Z_v\geqslant Z_{v\in V(i,j)} X_v$ but by Lemma 3.5 we have

$$\Pr\left(\sum_{v\in V(i,j)} X_v\leqslant(1-\varepsilon_j)\frac{v(i,j)\,s(i,j)}{v(i-1,j)}\right)\leqslant e^{-(\varepsilon_j^2/2)(v(i,j)s(i,j)/v(i-1,j))} \tag{3.9}$$

and part (a) follows.

(b) Let $U_v=1-Z_v$, and let $Y_v=1-X_v$ then $\sum_{v\in V(i,3)} Y_v\geqslant\sum_{v\in V(i,3)} U_v$ and the Y_v satisfy the conditions of Lemma 3.5 and so

$$\Pr\left(\sum_{v\in V(i,3)} Y_v\geqslant(1+\varepsilon_3)\frac{v(i,3)^2}{v(i-1,3)}\right)\leqslant e^{-(\varepsilon_3^2/3)(v(i,3)^2/v(i-1,3))} \tag{3.10}$$

and the result follows. ∎

LEMMA 3.7. *Algorithm CONNECT terminates successfully on a.e. D.*

Proof. Initialize. We form the sequence of out vertices starting from vertex 1, given by $(1, \text{out}(1), \text{out}(\text{out}(1)), ..., \text{out}^k(1))$.

The probability that the vertices are distinct up to and including the kth iteration is at least $1-k^2/n$. Thus if $k=\lceil \sqrt{n}/\log n \rceil$ we will complete Initialize successfully with probability at least $1-2/\log n$.

Pass 1. *Step* 1. By Lemma 3.4(b) and Lemma 2.6(a)

$$\Pr(\text{Step 1 terminates unsuccessfully at some iteration } t \leqslant t_1) \leqslant t_1 e^{-n^{.02}/2} = o(1).$$

By Lemma 3.4(e) the stopping size of $T(1)$ in Step 1 has been achieved by iteration t_1.

Pass 1. *Step* 2. The expected number of vertices $x \in T(1)$ which choose $\text{out}(x)$ in $S(0,1) \cup T(1)$ is

$$\frac{|T(1)|(|S(0,1)| + |T(1)|}{n} \leqslant 2n^{.46}. \tag{3.11}$$

Let $M = \{x \in T(1) : \text{out}(x) \in V(i_1^*, 1)\}$ and $M_1 = \{x \in M : \exists y \in M - \{x\} \text{ s.t. } \text{out}(y) = \text{out}(x)\}$. The Markov inequality applied to the expectation in (3.11) implies $\Pr(|M| \geqslant n^{.73}/2) = 1 - o(1)$. But for $x \in T(1)$ $\Pr(x \in M_1) \leqslant |T(1)|/n$ and so $E(|M_1|) \leqslant |T(1)|^2/n \leqslant 2n^{.46}$. Thus $\Pr(|M_1| \geqslant \sqrt{n}) = o(1)$ and $|\text{out}(M)| \geqslant |M - M_1| \geqslant n^{.73}/3$ with probability $1-o(1)$.

Pass 2. *Step* 1. As before,

$$\Pr(\text{Step 1 terminates unsuccessfully at iteration } t \leqslant t_2) \leqslant t_2 e^{-n^{.15}/2}$$

which tends to zero as required and thus we achieve the stopping size for $T(2)$.

Pass 2. *Step* 2. Let us condition on the set $T(2)$ and let

$$Z = |\{v \in V(i_2^*, 2) : \text{there does not exist } x \in T(2) \text{ s.t. } \text{out}(x) = v\}|$$

then Z is a random variable counting the number of vertices $v \in V(i_2^*, 2)$ which are not attached to an out edge of $T(2)$. Let $\phi(Z)$ be the number of k-subsets of such vertices we can form, then $\phi(m) = \max\{\binom{m}{k}, 0\}$ and thus, as $\phi(Z)$ is nonnegative and monotone nondecreasing we can apply the generalized Markov inequality

$$\phi(m) \Pr(Z \geqslant m \mid T(2)) \leqslant E(\phi(Z) \mid T(2))$$

giving

$$\Pr(Z \geqslant m \mid T(2)) < \frac{\binom{|V(i_2^*, 2)|}{k}\left(\frac{n-k}{n}\right)^{|T(2)|}}{\binom{m}{k}} \quad \text{for} \quad m \geqslant k.$$

Putting $|T(2)| = \lceil .4n \rceil$, $|V(i_2^*, 2)| = .6n - o(n)$, $m = .45n$ and $k = \log n$ gives $\Pr(Z \geqslant m) = o(1)$, and thus with probability $1 - o(1)$ we can choose $\lceil n/10 \rceil$ vertices as required.

Pass 3. *Step* 1. We see from Lemma 3.4(c), (d), and (e) and Lemma 3.6(b) that

$$\Pr(\text{Pass 3 Step 1 terminates unsuccessfully}) \leqslant t_3 e^{-\log n/18750} = o(1).$$

and thus we conclude that with probability $1 - o(1)$ the algorithm can be successfully completed. ∎

Proof of Theorem 3.1. We have thus shown that X_+ is almost always sufficiently large. That X_- is almost always sufficiently large follows by symmetry. ∎

4. Proof of Theorem 1

We observe that (2.2) continues to hold in its original form except in the case where α or β is less than 1. In this case further complications arise as Lemma 2.1 no longer holds and we can not rely on the components of the in and out subdigraphs being unicyclic, as some may now lack cycles altogether. To prove the theorem in the general case we consider the events in Lemma 2.2 but now define

$$E_3 = E_{31} \cup E_{32},$$

where

$$\begin{aligned} E_{31} &= \{\hat{C} \neq \varnothing\} \\ E_{32} &= \{\exists v : \text{IN}(v) = \varnothing \text{ and } v \notin \text{OUT}(V_n)\} \\ &\qquad \cup \{\exists v : \text{OUT}(v) = \varnothing \text{ and } v \notin \text{IN}(V_n)\} \end{aligned}$$

and define a further event $E_4 = E_{41} \cup E_{42}$ by

$$E_{41} = \Big\{ \exists v\colon \mathrm{IN}(v) = \varnothing \text{ but } v \in \mathrm{OUT}(V_n) \text{ and } v \in S \text{ s.t. } N^-(S) = \varnothing,\ |S| \leqslant \frac{\sqrt{n}}{\log n} \Big\}$$

$$E_{42} = \Big\{ \exists v\colon \mathrm{OUT}(v) = \varnothing \text{ but } v \in \mathrm{IN}(V_n) \text{ and } v \in S \text{ s.t. } N^+(S) = \varnothing,\ |S| \leqslant \frac{\sqrt{n}}{\log n} \Big\}$$

and generalize (2.2) to

$$\begin{aligned}\Pr(E_3) &\leqslant \Pr(D_{\alpha,\beta}(n) \text{ is not strongly connected}) \\ &\leqslant \Pr(E_1 \cup E_2 \cup E_3 \cup E_4). \qquad (4.1)\end{aligned}$$

Proof of (a). Assume $\alpha = 1 - \omega/n$ and let $X = \{v \in V_n\colon \mathrm{IN}(v) = \varnothing$ and $v \notin \mathrm{OUT}(V_n)\}$. Now

$$E(|X|) = n\frac{\omega}{n}\left(1 - \frac{1}{n}\right)^{\lfloor \beta \rfloor n}\left(1 - \frac{\beta - \lfloor \beta \rfloor}{n}\right)^n \approx \omega e^{-\beta}$$

and

$$\begin{aligned}E(|X|(|X| - 1)) &= n(n-1)\left(\frac{\omega}{n}\right)^2\left(1 - \frac{2}{n}\right)^{\lfloor \beta \rfloor n}\left(1 - \frac{2(\beta - \lfloor \beta \rfloor)}{n}\right)^n \\ &\approx \omega^2 e^{-2\beta}.\end{aligned}$$

Thus the Chebycheff inequality can be used to show $\Pr(X \neq \varnothing) \to 1$ and (a) follows. For the remainder of the proof we first observe that (lengthy) calculations similar to those of Lemma 2.3(c) show that $\Pr(E_3)$ tends to the claimed limits. Thus to prove the theorem it suffices to prove that $\Pr(E_1 \cup E_2 \cup E_4) \to 0$. We deal with these in reverse order.

Proof that $\Pr(E_4) \to 0$. We show that $\Pr(E_4) \to 0$ in the case where $\alpha, \beta < 1$, and note that the proof follows a fortiori for the other cases. Consider E_{41}

$$\begin{aligned}\Pr(E_{41}) \leqslant \sum_{k=2}^{\sqrt{n}/\log n} &\binom{n}{k}\left(\frac{b}{n} + \left(1 - \frac{b}{n}\right)\left(1 - \frac{k}{n}\right)\right)^{n-k} \\ &\times k(k-1)\frac{a}{n^2}\left(\frac{a}{n} + \left(1 + \frac{a}{n}\right)\frac{k}{n}\right)^{k-1}.\end{aligned}$$

To see this let $|S|=k$ in the definition of E_{41}, then $(b/n+(1-b/n)(1-k/n))^{n-k}$ accounts for $S\cap \mathrm{OUT}(V_n-S)=\varnothing$, $k(k-1)(a/n^2)$ accounts for a pair $v, w\in S$ with $\mathrm{IN}(v)=\varnothing$ and $\mathrm{out}(w)=v$ and finally $(a/n+(1-a/n)(k/n))^{k-1}$ accounts for $\mathrm{IN}(s-\{v\})\subseteq S$. Thus

$$\begin{aligned}\Pr(E_{41}) &\leqslant \sum_{k=2}^{\sqrt{n}/\log n}\left(\frac{ne}{k}\right)^k e^{-k}\frac{k^2a}{n^2}\left(\frac{k}{n}\right)^{k-1} e^{a+o(1)}\\ &= ae^{a+o(1)}\sum_{k=2}^{\sqrt{n}/\log n}\frac{k}{n}\\ &= o(1)\end{aligned}$$

A similar result holds for E_{42} and we conclude that $\Pr(E_4)\to 0$ as required.

Proof that $\Pr(E_2)\to 0$. To show $\Pr(E_2)\to 0$ we count the expected number N_{in} of cycles in $C'_{\mathrm{in}}-\hat{C}_{\mathrm{in}}$. This satisfies

$$\begin{aligned}N_{\mathrm{in}} \leqslant & \sum_{s=1}^{\lfloor\sqrt{n}/\log n\rfloor}\sum_{k=1}^{s}\binom{n}{s}\binom{s}{k}(k-1)!\left(\frac{\alpha}{n}\right)^k\left(\frac{s}{n}\right)^{\lfloor\alpha\rfloor(s-k)}\\ &\times\left((1-(\alpha-\lfloor\alpha\rfloor))\left(1-\frac{s}{n}\right)\right)^{s-k}\left(1-\frac{s}{n}\right)^{\lfloor\beta\rfloor(n-s)}\\ &\times\left((1-(\beta-\lfloor\beta\rfloor))\frac{s}{n}\right)^{n-s}\frac{bsk}{n}.\end{aligned}$$

This looks more complex than the expression in (2.1) but if we consider the cases $\alpha=1-a/n$, $\alpha\geqslant 1$, $\beta=1-b/n$, $\beta\geqslant 1$ separately and ignore $(1-(\alpha-\lfloor\alpha\rfloor)(1-s/n))^{s-k}$ when $\alpha\geqslant 1$ and $(1-(\beta-\lfloor\beta\rfloor)(s/n))^{n-s}$ when $\beta\geqslant 1$ then the calculations are much as before and we fond that $N_{\mathrm{in}}\to 0$. A similar result holds for $C'_{\mathrm{out}}-\hat{C}_{\mathrm{out}}$ and $\Pr(E_2)\to 0$ as required.

Proof that $\Pr(E_1)\to 0$. We now have to indicate why $\Pr(E_1)\to 0$. We need only consider case (b) for the more edges we have, the more likely CONNECT is to succeed.

Consider Lemma 3.6(a) and the definition of Z_v. If the "otherwise" case includes $\mathrm{IN}(v)=\varnothing$ then

$$\Pr(Z_v=1)=\frac{s(i,j)}{v(i-1,j)}\left(1-\frac{a}{n}\right).$$

The effect of this in (3.9) is to replace the ε_j^2 term in the RHS by $(\varepsilon_j-a/n)^2/(1-a/n)^2$ and is negligible.

In Lemma 3.6(b) we now have

$$\Pr(Y_v = 1) = \frac{v(i,3) + \frac{a}{n} s(i,3)}{v(i-1,3)}.$$

The effect of this on (3.10) is to replace the ε_3^2 terms in the RHS by

$$\left(\frac{\varepsilon_3 v(i,3) - \frac{a}{n} s(i,3)}{v(i,3) + \frac{a}{n} s(i,3)} \right)^2 \geqslant \varepsilon_3^2 \left(1 - \frac{a}{25(\log n)^2} \right)^2$$

which is again negligible.

Consider now Lemma 3.7 and in particular the effect of the changes on the various paragraphs of Algorithm CONNECT.

Initialise. The probability that we hit a vertex v with $\mathrm{OUT}(v) = \varnothing$ is at most $ka/n \to 0$.

Pass 1. *Step* 1. We have shown that (3.9) is only weakened in an insignificant manner.

Pass 1. *Step* 2. (3.11) holds a fortiori, and also $\Pr(|M_1| \geqslant \sqrt{n}) = o(1)$ a fortiori. Since $E(|\{v: \mathrm{OUT}(v) = \varnothing\}|) = a$ we also have $\Pr(|M| \geqslant n^{.73}/2) = 1 - o(1)$ and the proof goes through.

Pass 2. *Step* 1. As for Pass 1. Step 1.

Pass 2. *Step* 2. Put $m = .45n - \log n$ to handle the vertices with $\mathrm{OUT}(v) = \varnothing$.

Pass 3. *Step* 1. We have shown that (3.10) is only weakened in an insignificant manner.

Thus we have shown that $\Pr(E_1 \cup E_2 \cup E_4) \to 0$ as required in the general case and applying this to (4.1) completes the proof of Theorem 1. ∎

ACKNOWLEDGMENTS

When we first considered $D_{1,1}(n)$ it was not clear if the limit in the above theorem should be 0 or 1 or something in between, as it has turned out. The fact that the limit could not be 1 was pointed out to us at the Conference on Random Graphs in Poznan (1985) by W. Fernandez de la Vega. In fact he pointed out sub-structures that will occur with positive probability and prevent strong connectivity. We show here that in the case of $D_{1,1}(n)$ there is nothing else likely to prevent strong connectivity and we are pleased to acknowledge his insight.

REFERENCES

1. B. BOLLOBAS, "Random Graphs," pp. 364–373, Academic Press, San Diego, 1985.
1a. B. BOLLOBAS, "Random Graphs," pp. 16–17, Academic Press, San Diego, 1985.
2. T. I. FENNER AND A. M. FRIEZE, On the connectivity of random m-orientable graphs and digraphs, *Combinatorica* **2** (1982), 347–359.
3. W. HOEFFDING, Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.* **58** (1963), 13–30.

Printed by Catherine Press, Ltd., Tempelhof 41, B-8000 Brugge, Belgium