Adv. Appl. Prob. **24**, 845–857 (1992)
Printed in N. Ireland

GIANT COMPONENTS IN THREE-PARAMETER RANDOM DIRECTED GRAPHS

TOMASZ ŁUCZAK* AND
JOEL E. COHEN** *Rockefeller University*

Abstract

A three-parameter model of a random directed graph (digraph) is specified by the probability of 'up arrows' from vertex i to vertex j where $i<j$, the probability of 'down arrows' from i to j where $i>j$, and the probability of bidirectional arrows between i and j. In this model, a phase transition—the abrupt appearance of a giant strongly connected component—takes place as the parameters cross a critical surface. The critical surface is determined explicitly. Before the giant component appears, almost surely all non-trivial components are small cycles. The asymptotic probability that the digraph contains no cycles of length 3 or more is computed explicitly. This model and its analysis are motivated by the theory of food webs in ecology.

PHASE TRANSITION; ACYCLIC RANDOM DIGRAPHS; RANDOM GRAPHS; FOOD WEB; ECOLOGY

AMS 1991 SUBJECT CLASSIFICATION: PRIMARY 05C80
SECONDARY 92D40

1. Introduction

Erdős and Rényi (1960) observed that the structure of a random graph changes discontinuously as a function of c, the average degree of a vertex. When $c<1$, a typical random graph consists of a large number of small components. When $c>1$, the probability that a graph has a giant component that contains a positive fraction of all the vertices tends to 1 as $n\to\infty$. Abrupt transitions in structure have since been discovered in other randomized models of discrete objects like random cubes (Ajtai et al. (1982)), random hypergraphs (Schmidt-Pruzan and Shamir (1985)) and random digraphs (Karp (1990), Łuczak (1990)). In percolation theory (Grimmett (1989), Kesten (1982)), this kind of behavior is called a phase transition, because of connections with physics.

In this paper, we find the critical surface on which the phase transition takes place in a three-parameter model of a random digraph. Few multiparameter models have been considered previously. Our motivation for defining and analyzing the three-parameter model of a directed graph which follows is the theory of food webs in

Received 24 July 1989; revision received 12 September 1991.
* Permanent postal address: Institute of Mathematics, Adam Mickiewicz University, Matejki 48/49, 60-769 Poznan, Poland.
** Postal address: The Rockefeller University, 1230 York Avenue, Box 20, New York, NY 10021, USA.

ecology (Cohen et al. (1990)). In defining the critical surface, a crucial role is played by a function $x:[1,\infty)\to(0,1]$ defined as the smallest root of the equation

$$x(c)\exp(-x(c))=c\exp(-c). \tag{1}$$

This function unexpectedly links our result with the first paper on phase transitions in graph theory (Erdös and Rényi (1960)) where the same function is important.

2. Main result

$D_n(r,s,t)$ denotes a digraph on the set of vertices $[n]=\{1,2,\cdots,n\}$ in which independently for each pair of vertices i, j with $i>j$, the arrow (i,j) from i to j occurs (and the reverse arrow does not occur) with probability r/n; the arrow (j,i) from j to i occurs (and the reverse arrow does not occur) with probability s/n; both arrows, (i,j) and (j,i) occur simultaneously with probability t/n; and finally neither (i,j) nor (j,i) occurs with probability $1-(r+s+t)/n$. Throughout we assume that r, s, t are non-negative real numbers which do *not* depend on n. We say that $D_n(r,s,t)$ has a property almost surely if the probability that $D_n(r,s,t)$ has this property tends to 1 as $n\to\infty$.

Our goal is to find those triples (r,s,t) for which almost surely all components of $D_n(r,s,t)$ are either trivial (i.e. consist of one vertex) or very small cycles, as well as the triples (r,s,t) such that $D_n(r,s,t)$ almost surely contains a large component of order n. Here and below 'component' means a strongly connected component and a 'non-trivial' component is a component with at least two vertices.

Our main result is the following.

Theorem. Let $\omega(n)\to\infty$ and $x:[1,\infty)\to(0,1]$ be the function defined by (1).

(i) If

$$r+t<1 \quad\text{and}\quad s+t<1 \tag{2}$$

or

$$r+t\geqq 1 \quad\text{but}\quad s+t<x(r+t) \tag{3}$$

or

$$s+t\geqq 1 \quad\text{but}\quad r+t<x(s+t), \tag{4}$$

then almost surely all components of $D_n(r,s,t)$ are of size less than $\omega(n)\log n$,

$$P_0\equiv\lim_{n\to\infty}\Pr\{D_n(r,s,t)\text{ contains no cycles of length at least }3\}>0$$

and when $(r+t)(s+t)>0$ then also $P_0<1$. Moreover almost surely all cycles of length at least 3 are shorter than $\omega(n)$ and no two such cycles share a vertex. In particular, if $t=0$, all non-trivial components are cycles of length at most $\omega(n)$.

(ii) If

$$r+t\geqq 1 \quad\text{and}\quad s+t>x(r+t) \tag{5}$$

or

$$(6) \qquad s+t \geqq 1 \quad \text{and} \quad r+t > x(s+t)$$

then there exists a positive constant $\alpha = \alpha(r, s, t)$, which does not depend on n, such that almost surely $D_n(r, s, t)$ contains a large component on at least αn vertices. Moreover, almost surely this component contains a cycle of length at least 3, i.e. $P_0 = 0$.

The results of Karp (1990) and Łuczak (1990) show that, in a one-parameter model of a random digraph, almost surely after the phase transition, the large strongly connected component is unique and every other non-trivial component is a cycle of length less than $\omega(n)$. One may adapt arguments from Łuczak (1990) to show that these conclusions remain true also in our multiparameter model. However, the proof is long and the result is not necessary for our ecological application. So we omit these results for brevity.

The results of our theorem have been extended in two respects (Cohen et al. (1990)). First, under the hypotheses of part (i) of the theorem,

$$P_0 = \begin{cases} \exp\left((r+t)(s+t)/2\right) \dfrac{(r+t)\exp(s+t) - (s+t)\exp(r+t)}{r-s}, & r \neq s. \\ \exp\left((r+t)^2/2\right) \exp(r+t)(1-r-t), & r = s. \end{cases}$$

Second, if the parameters fall on the critical surface $r+t \geqq 1$ and $s+t = x(r+t)$, or $s+t \geqq 1$ and $r+t = x(s+t)$, then $P_0 = 0$.

3. Proof of main result

Because Theorem 1 is symmetric with respect to r and s, we shall assume from here on that $r \geqq s$.

Consider first the following procedure. Choose vertex i, such that $i/n = z$, and join i to vertices j smaller than i independently with probability $(r+t)/n$ (i.e. add arrows (i, j) where $j < i$) and with probability $(s+t)/n$ put in arrows (i, j) from i to each larger vertex $j > i$. Denote the expected number of outneighbours of i that result from this procedure by $f_1^n(z)$. Now join each outneighbour i_1 of i with a vertex $j \neq i_1$ independently with probability $(r+t)/n$ if $j < i_1$ and with probability $(s+t)/n$ if $j > i_1$. The expected number of outneighbours of all outneighbours of i (i.e. the second-generation descendants of i) is denoted by $f_2^n(z)$. Similarly, define $f_k^n(z)$ to be the number of vertices in the kth generation of outneighbours of i for $k = 3, 4, \cdots$. Each descendant of i chooses its children independently of the previous stages of the process, i.e. a vertex may appear in more than one generation of descendants. Thus $f_k^n(z)$ is not equal to the expected number of vertices of $D_n(r, s, t)$ which are reachable from i in exactly k steps but clearly $f_k^n(z)$ bounds this value from above.

Unfortunately, as a function with domain $\{i/n : 1 \leqq i \leqq n\}$, f_1^n is rather inconvenient to study. Thus we shall approximate f_k^n by some functions f_k defined on the whole interval $[0, 1]$.

For $z \in [0, 1]$ set $f_1(z) = z(r+t) + (1-z)(s+t)$ and, when $k \geqq 1$,

$$f_{k+1}(z) = (r+t)\int_0^z f_k(x)\,dx + (s+t)\int_z^1 f_k(x)\,dx$$
$$= (r-s)\int_0^z f_k(x)\,dx + (s+t)\int_0^1 f_k(x)\,dx.$$

We shall show that for some constant C_k and every natural number n

$$|f_k^n(z) - f_k(z)| \leqq C_k n^{-1}, \tag{7}$$

provided z belongs to the domain of f_k^n.

Indeed, if $k = 1$ then

$$f_1^n(z) = z(r+t) + (1-z)(s+t) - (r+t)/n$$

so (7) holds. Now suppose that (7) is valid for some $k \geqq 1$. Then

$$|f_{k+1}^n(z) - f_{k+1}(z)| = \left| \sum_{i=1}^{zn-1} \frac{r+t}{n} f_k^n(i/n) + \sum_{zn+1}^{n} \frac{s+t}{n} f_k^n(i/n) - f_{k+1}(z) \right|$$
$$\leqq \left| \sum_{i=1}^{zn-1} \frac{r+t}{n} f_k(i/n) + \sum_{zn+1}^{n} \frac{s+t}{n} f_k(i/n) - f_{k+1}(z) \right| + C_k(r+t)/n.$$

One can easily check that all functions f_k are positive, increasing and bounded from above in the interval $[0, 1]$. So, for suitable constants $C' = C'(k)$ and $C'' = C''(k)$, we have

$$\left| \sum_{i=1}^{zn-1} f_k(i/n)/n - \int_0^z f_k(x)\,dx \right| < C'/n$$

and

$$\left| \sum_{zn+1}^{n} f_k(i/n)/n - \int_z^1 f_k(x)\,dx \right| < C''/n.$$

Hence

$$|f_{k+1}^n(z) - f_{k+1}(z)| \leqq (C_k + C' + C'')(r+t)/n,$$

and, by induction, (7) follows.

The following result, characterizing the behavior of f_k for large k, is proved in the Appendix.

Lemma 1. Let a sequence of functions $\{f_i\}_{i=0}^{\infty}$, $f_i : [0, 1] \to \mathbb{R}$ be defined by $f_0 \equiv 1$ and

$$f_k(x) = (a-b)\int_0^x f_{k-1}(t)\,dt + b\int_0^1 f_{k-1}(t)\,dt, \tag{8}$$

where a and b denote some positive real parameters. Let $x = x(c)$ be the smallest root of (1).

(i) If

$$a < 1 \quad \text{and} \quad b < 1 \tag{2'}$$

or

$$a \geqq 1 \quad \text{but} \quad b < x(a) \tag{3'}$$

or

$$b \geqq 1 \quad \text{but} \quad a < x(b) \tag{4'}$$

then

$$\exists A = A(a, b) < 1, \qquad \exists N = N(a, b), \qquad \forall k \geqq N, \qquad \forall x \in [0, 1]: f_k(x) \leqq A^k.$$

(ii) If

$$a \geqq 1 \quad \text{and} \quad b > x(a) \tag{5'}$$

or

$$b \geqq 1 \quad \text{and} \quad a > x(b) \tag{6'}$$

then

$$\forall C > 0, \qquad \exists N = N(a, b, C), \qquad \forall k \geqq N, \qquad \forall x \in [0, 1]: f_k(x) \geqq f_{k-1}(x) > C.$$

Now we are ready to prove the first part of the theorem.

Proof of theorem, Part (i). Let us look first at cycles of length at least 3 contained in $D_n(r, s, t)$. Denote by $X_k = X_k(n, r, s, t)$ the number of cycles of length k and let $Y_k^v = Y_k^v(n, r, s, t)$ be the number of vertices reachable from v in exactly k steps. Then $\boldsymbol{E}X_k$ is less than the expected number of vertices contained in cycles of length k, which in turn is bounded from above by the number of arrows (i, j), where i belongs to the $(k-1)$th generation of descendants of j. The probability that such an arrow exists is at most $\max\{(r+t)/n, (s+t)/n\} \leqq (r+t)/n$ since $r \geqq s$. Thus

$$\begin{aligned} \boldsymbol{E}X_k &\leqq \sum_{v \in [n]} \sum_{l=1}^{n-1} \Pr\{Y_{k-1}^v = l\} l \frac{r+t}{n} \\ &\leqq (r+t) \max\{\boldsymbol{E}Y_{k-1}^v : v \in [n]\} \\ &\leqq (r+t) \max\{f_{k-1}^n(x) : x \in [0, 1]\}. \end{aligned}$$

But when one of the conditions (2), (3) or (4) is fulfilled, then, due to Lemma 1(i) with $a = r+t$, $b = s+t$, the last maximum decreases exponentially, i.e. for some $A < 1$, $\boldsymbol{E}X_k \leqq (r+t)A^{k-1}$.

Thus, the probability that there exists a cycle of length larger than $\omega(n)$, when $\omega(n) \to \infty$, is bounded above by

$$\sum_{k \geqq \omega(n)}^{n} \boldsymbol{E}X_k \leqq (r+t) \sum_{k=\omega(n)}^{\infty} A^{k-1} = \frac{r+t}{A(1-A)} A^{\omega(n)} \to 0.$$

Moreover, for the expectation of the number X of all cycles which are less than, say, $\log\log n$ in length, we have

$$EX = \sum_{k=3}^{\log\log n} EX_k \leqq (r+t)\sum_{k=3}^{\infty} A^{k-1} \leqq \frac{A^2(r+t)}{1-A}. \tag{9}$$

On the other hand,

$$EX_k = \binom{n}{k}\sum_{i=1}^{k-1}\sigma_i^k\left(\frac{r+t}{n}\right)^i\left(\frac{s+t}{n}\right)^{k-i} = \left(1+O\left(\frac{1}{n}\right)\right)\frac{1}{k!}\sum_{i=1}^{k-1}\sigma_i^k(r+t)^i(s+t)^{k-i}$$

where σ_i^k denotes the number of ways one can build a directed cycle on the set $[k]=\{1, 2, \cdots, k\}$ with exactly i arrows going from a larger vertex to a smaller one. Hence

$$EX = \sum_{k=3}^{\log\log n} EX_k = (1+o(1))\sum_{k=3}^{\infty}\frac{1}{k!}\sum_{i=1}^{k-1}\sigma_i^k(r+t)^i(s+t)^{k-i} = (1+o(1))\lambda.$$

From (9) we know that λ exists, i.e. that the above series converges. Clearly when $(r+t)(s+t)>0$ then also $\lambda>0$. Furthermore, one can easily check that for each natural number l

$$E[X(X-1)\cdots(X-l+1)] = (1+o(1))(EX)^l = (1+o(1))\lambda^l$$

so the distribution of X tends to the Poisson distribution with mean λ, i.e., for each natural m

$$\lim_{n\to\infty} \Pr\{X=m\} = \frac{\lambda^m}{m!}\exp(-\lambda).$$

Thus $P_0 = \lim_{n\to\infty}\Pr\{X=0\} = \exp(-\lambda)>0$ and if $(r+t)(s+t)>0$ then $P_0<1$.

Now, to show that all cycles of length between 3 and $\omega(n)$ of $D_n(r, s, t)$ are almost surely vertex disjoint, note that if two such cycles of a directed graph share a vertex the graph must contain a subdigraph with less than $2\omega(n)-1$ vertices and more arrows than vertices. The smallest possible such subdigraph which contains two cycles has $k=4$ vertices. When k denotes the number of vertices and l the number of arrows in the subdigraph, there are $\binom{n}{k}$ ways of choosing the vertices, $\binom{k(k-1)}{l}$ ways of choosing the arrows, and the probability that two cycles of length less than $\log\log n$ share a vertex is less than the expected number of subdigraphs with more arrows than vertices, which is bounded above by

$$\sum_{k=4}^{2\log\log n}\binom{n}{k}\sum_{l=k+1}^{k(k-1)}\binom{k(k-1)}{l}\left[\max\left\{\frac{r+t}{n}, \frac{s+t}{n}\right\}\right]^l \leqq \sum_{k=4}^{2\log\log n} n^k 2^{k(k-1)}k^2\left(\frac{r+t}{n}\right)^{k+1}$$
$$\leqq n^{-1}2^{(2\log\log n)^2}(2\log\log n)^3\max\{1, (r+t)^{2\log\log n+1}\} = o(1),$$

for each triple (r, s, t).

Finally, to complete the proof of (i), note that $t<1$, so almost surely each component of the graph induced by all 'double' arrows of $D_n(r, s, t)$ contains fewer than $\log n/(t-1-\log t)$ vertices (see Erdös and Rényi (1960)). Moreover, we have just proved that $D_n(r, s, t)$ almost surely contains less than $\omega(n)$ cycles longer than 2 so the largest component of $D_n(r, s, t)$ is almost surely smaller than $\omega(n) \log n/(t-1-\log t)$.

Remark. In fact, for $t>0$ the structure of $D_n(r, s, t)$ could be characterized more precisely. The argument from our proof, applied a bit more carefully, shows that almost surely all cycles of length at least 3 are contained in components of size less than $\omega(n)$, and, consequently, all other components are 'trees' of size less than $\log n/(t-1-\log t)$ in which each pair of vertices i, j is either non-adjacent or connected by a pair of arrows (i, j) and (j, i).

To show the second part of the theorem we shall find a sequence of functions $\bar{f}_k(z)$ which satisfy recursive relation (8) and which bound from below the expected number of descendants in the kth generation of a vertex $i = zn$ of $D_n(r, s, t)$; then we use Lemma 1(ii). To do so, we shall simply omit vertices that have been chosen in any earlier generation when picking the vertices which are to be the children or outneighbours of each new generation. However, carrying out this idea requires some technical arguments. We first state a further consequence of Lemma 1(ii) and show how it implies the second part of the theorem.

Lemma 2. Let r, s, t be such that one of conditions (5) and (6) holds. Then there exists a positive constant $\alpha' = \alpha'(r, s, t)$, which does not depend on n, such that almost surely $D_n(r, s, t)$ contains a vertex v and sets S^-, S^+ such that

(i) $S^- \cap S^+ = \varnothing$, and

(ii) $|S^-| = |S^+| = \lfloor \alpha' n \rfloor$, and

(iii) for every $w^- \in S^-[w^+ \in S^+] D_n(r, s, t)$ contains a directed path from w^- to v [from v to w^+].

A proof of Lemma 2 comes later.

Proof of theorem, Part (ii). Let r, s, t be such that one of conditions (5) and (6) holds. In fact, since we assume that $r \geqq s$, we consider only (5). Moreover, $x(1)=1$, so $r+t>1$. Choose r', s', t' in such a way that $r' \leqq r$, $s' \leqq s$, $t' \leqq t$, $r'+t'<r+t$, $s'+t'<s+t$, but the assumption (5) is fulfilled also for r', s', t'. Then one can construct $D_n(r, s, t)$ to be a supergraph of $D_n(r', s', t')$ by examining all pairs of vertices i, j, $i>j$, for which neither arrow (i, j) nor (j, i) exists in $D_n(r', s', t')$ and adding arrows (i, j) or (j, i) or both of them or neither of them with probabilities $(r-r')/n$, $(s-s')/n$, $(t-t')/n$, and $1-(r+s+t-r'-s'-t')/n$, respectively.

Lemma 2 implies the existence of $\alpha' = \alpha'(r', s', t')>0$ such that almost surely $D_n(r', s', t')$ contains two disjoint subsets S^-, S^+, of size $\lfloor \alpha' n \rfloor$ and from some vertex v there exist paths to all vertices of S^+ and v is reachable from each vertex of S^-. Moreover, almost surely the sum of the indegree and the outdegree of each vertex of $D_n(r', s', t')$ is less than $\log n$, since

$$n \sum_{i=\log n}^{n-1} \binom{n}{i} \left(\frac{r+s+t}{n}\right)^i \leqq n \sum_{i=\log n}^{n-1} \left(\frac{e(r+s+t)}{i}\right)^i \to 0.$$

Now construct $D_n(r, s, t)$ from $D_n(r', s', t')$. Then for each vertex $w^+ \in S^+$, the probability that the arrow (w^+, w^-) will be contained in $D_n(r, s, t)$ for some $w^- \in S^-$ is at least

$$1 - \left(1 - \min\left\{\frac{r+t}{n}, \frac{s+t}{n}\right\}\right)^{|S^+|-\log n} > 1 - \exp(-\alpha'(s+t)) + o(1)$$

$$> 0{\cdot}9 - 0{\cdot}9 \exp(-\alpha'(x+t)).$$

Since these events are independent for different vertices in S^+, the number of vertices from S^+ with outneighbours in S^- is bounded from below by a binomially distributed random variable with parameters $(\lfloor \alpha' n \rfloor,\ 0{\cdot}9 - 0{\cdot}9 \exp(-\alpha'(s+t)))$, which with probability tending to 1 as $n \to \infty$ is larger than αn where $\alpha = 0{\cdot}8\alpha'(1 - \exp(-\alpha'(s+t)))$. Hence almost surely at least $\lfloor \alpha n \rfloor$ vertices of S^+ belong to the same component of $D_n(r, s, t)$ as the vertex v and lie on cycles of length at least 3.

Proof of Lemma 2. Let us start with some notation. If $i \in [n]$ and $M \subseteq [n]$ then by the height of i in M we shall mean the number of vertices of M which are smaller than i. The height of i in M is denoted by $h(i, M)$. For a quadruple (i, j, m, M) where $i, j, m \in [n]$, $M \subseteq [n]$, $i \notin M$ and $j \leqq h(i, M) \leqq j + m \leqq |M|$, $F(i, j, m, M)$ is a set obtained from M by deleting its $h(i, M) - j$ smallest elements and $m + j - h(i, M)$ largest ones. Thus $F(i, j, k, M)$ has exactly $|M| - m$ elements and $h(i, F(i, j, m, M)) = j$.

Let i be a vertex of $D_n(r, s, t)$, $j, m \in [n]$, and $S \subseteq [n]$ be a set of 'spoiled' vertices. Then an outneighbour [inneighbour] v of i is defined to be (j, m, S)-proper if it lies in $F(i, j, m, [n] \backslash S)$.

Note that $F(i, j, m, [n] \backslash S)$ always exists whenever

$$|S| \leqq 0{\cdot}3m, \tag{10}$$

$$|j - (h(i, [n] \backslash S) - 0{\cdot}5m)| \leqq |S| + 0{\cdot}1m \tag{11}$$

and

$$j \leqq n - m - |S|, \tag{12}$$

since from (10) and (11) it follows that

$$j \leqq j + 0{\cdot}4m - |S| \leqq h(i, [n] \backslash S) \leqq j + 0{\cdot}6m + |S| \leqq j + m.$$

In a process, which will be crucial for our considerations, at each step in discrete time, we shall choose only neighbours that are proper in the sense above, adding to the set of spoiled vertices all descendants picked up previously. To be precise, let $\varepsilon > 0$ be a small positive constant such that the condition (5) holds also for some $r' \leqq r$, $s' \leqq s$, $t' \leqq t$ where $r' + t' = (1-\varepsilon)(r+t)$ and $s' + t' = (1-\varepsilon)(s+t)$. Once and for all, set $m = \lfloor \varepsilon n \rfloor$. Now we shall define an (i, j, m, S)-process for each $|S| \leqq 0{\cdot}2m$, $i \in S$, $0{\cdot}3m \leqq i \leqq n - 0{\cdot}3m$ and $|j - (i - 0{\cdot}5m)| \leqq |S|$. In the first step of an (i, j, m, S)-process, find all (j, m, S)-proper outneighbours of i. (We assume i is connected with all smaller vertices with probability $(r+t)/n$ and with all larger vertices with probability $(s+t)/n$.) Denote the set of all (j, m, S)-proper outneighbours by N_1 and for each $i' \in N_1$ set $h(i') = h(i', F(i, j, m, [n] \backslash S))$. Set $S = S \cup N_1$. If $|S| \geqq 0{\cdot}3m$, then stop. If not, pick any $i_1 \in N_1$ and find the set N_2 of all $(h(i_1), m, S)$-proper outneighbours of i_1, putting, for each $i'' \in N_2$, $h(i'') = h(i'', F(i_1, h(i_1), m, [n] \backslash S))$. Set $S = S \cup N_2$. Then when $|S| \leqq 0{\cdot}3m$, pick another vertex from the outneighbours of i, say i_2, find the set N_3 of all $(h(i''), m, S)$-proper outneighbours of i_2 and set $h(i''') = h(i''', F(i_2, h(i_2), m, [n] \backslash S))$ for all $i''' \in N_3$, and set $S = S \cup N_3$. Repeat this procedure until either $|S| > 0{\cdot}3m$ or no unspoiled descendants of i remain. A reverse (i, j, m, S)-process is defined in the same way, but instead of outneighbours we look for proper inneighbours at every step.

Let $i = zn$ and let $\bar{f}_k^n(z)$ denote the expected number of all descendants of i found in the kth generation (not in the kth step!) in the (i, j, m, S)-process. Clearly, $\bar{f}_k^n(z)$ is smaller than the expected number of vertices reachable from i in k steps in $D_n(r, s, t)$. (By contrast with the process in the proof of the first part of the theorem, k can never be arbitrarily large.) Moreover, our process is constructed in such a way that each descendant $\bar{i}$ of i chooses its outneighbours from a set that has the same number of vertices (namely, $n - m - 1$), and in this set $\bar{i}$ has the same height as $\bar{i}$ had in the set $\bar{i}$ was chosen from by i. Thus, arguing as in the proof of the first part of the theorem, one arrives at a recursive formula for $\bar{f}_k(z) = \lim_{z_n \to z} \lim_{n \to \infty} \bar{f}_k^n(z_n)$:

$$\begin{aligned}\bar{f}_k(z) &= \frac{r+t}{n}(n-m)\int_0^z \bar{f}_{k-1}(x)\,dx + \frac{s+t}{n}(n-m)\int_z^1 \bar{f}_{k-1}(x)\,dx \\ &\geqq (r'+t')\int_0^z \bar{f}_{k-1}(x)\,dx + (s'+t')\int_z^1 \bar{f}_{k-1}(x)\,dx \\ &= (r'-s')\int_0^z \bar{f}_{k-1}(x)\,dx + (s'+t')\int_0^1 \bar{f}_{k-1}(x)\,dx.\end{aligned}$$

The same inequality with r' and s' interchanged is valid for $\hat{f}_k(z)$, the asymptotic expected number of kth-generation ancestors of i found in a reverse (i, j, m, S)-process.

Now let $T^+(i)[T^-(i)]$ be the number of all descendants [ancestors] of i found in a [reverse] (i, j, m, S)-process.

Fact. Let $S \subseteq [n]$, $|S| \leqq 0{\cdot}2m$, $i \notin S$, $|j - (h(i, [n] \backslash S) - 0{\cdot}5m)| \leqq |S| + 0{\cdot}1m$ and $j \leqq n - m - |S|$. Then there exist positive constants β and β', which do not depend on n, such that the probabilities that $T^+(i) \geqq 0{\cdot}1m$, and $T^-(i) \geqq 0{\cdot}1m$ are larger than β and β', respectively. Moreover, with positive probability $\gamma > 0[\gamma' > 0]$, the number of descendants [ancestors] of i in the $(\log\log n)^2$th generation is at least $\log^2 n$ while the total number of all descendants [ancestors] up to the $(\log\log n)^2$th generation is less than $n^{1/2}$.

Proof of fact. Due to Lemma 1(ii) we may choose k_0 large enough to have $\bar{f}_{k_0}(z) > 2$ for all $z \in [0, 1]$. Let $\bar{G}(i)$ be the (i, j, m, S)-process described in the proof of Lemma 2. Let $\bar{G}_{k_0}(i)$ denote the process starting from i in which the lth generation of descendants of i in $\bar{G}_{k_0}(i)$ equals the lk_0th generation of descendants in $\bar{G}(i)$ for $l = 1, 2, \cdots$. Since the expected number of offspring in $\bar{G}_{k_0}(i)$ is larger than two for every i, a well-known theorem of branching processes (see, for example, Harris (1963), Theorem I.6.1) guarantees that the probability of extinction of the line of descent in $\bar{G}_{k_0}(i)$ is strictly smaller than 1, i.e. with positive probability the number of descendants in $\bar{G}(i)$ will grow to exceed $0{\cdot}1m$. Moreover, the expectation of the square of the number of offspring in $\bar{G}_{k_0}(i)$ is bounded above by the expectation of the square of the number of offspring for the analogous process defined for the digraph $D_n(0, 0, r+s+t)$, which can easily be shown to be finite. Thus, results from the theory of branching processes (Harris (1963), Theorem I.8.1) assure us that with probability $1 - o(1)$ the number of offspring in the $\lfloor(\log\log n)^2/k_0\rfloor$ generation of $\bar{G}_{k_0}(i)$ is at least

$$\bar{f}_{k_0}(z)^{(\log\log n)^2/\log\log\log n}/\log\log\log n > \log^2 n,$$

where the inequality holds for large enough n. (In this formula, '$\log\log\log n$' actually stands for anything larger than a constant.) Finally, the expectation of the total number of offspring (outneighbours) of i up to the $(\log\log n)^2$ generation in $\bar{G}(i)$ equals

$$\sum_{i=1}^{(\log\log n)^2} f_i(z) \leqq (\log\log n)^2 f_{k_0}(z)^{(\log\log n)^2} < n^{0\cdot 4},$$

where the inequality on the right holds for large enough n. So, from Markov's inequality, the probability that the total number of offspring [outneighbours] of i up to $(\log\log n)^2$ generations is less than $n^{1/2}$ is at least $1 - O(n^{-0\cdot 1})$. Analogous arguments apply in the reverse case.

The strength of the above result is that it allows us to start with any set of spoiled vertices, provided it is not too large.

Now we prove Lemma 2. Set $i_1 = \lfloor 0{\cdot}5n \rfloor$, $j_1 = i - \lfloor 0{\cdot}5m \rfloor$, $S = \{i_1\}$ and perform an (i_1, j_1, m, S)-process until the total number of descendants of i_1 is larger than $n^{1/2}$ or you find all descendants in the $(\log\log n)^2$th generation. If you succeed, i.e. the number of descendants in the $(\log\log n)^2$th generation is larger than $\log^2 n$ and the total number of descendants is smaller than $n^{1/2}$, then perform the reverse (i_1, j_1, m, S)-process but start from the set S which contains all the vertices spoiled in the previous step. If you fail in either the forward process or the reverse process, set $i_2 = \min\{i \in [n] \backslash S : i > 0{\cdot}5m\}$, $j_2 = i_2 - \lfloor 0{\cdot}5m \rfloor$, add to the set S of already spoiled vertices the vertex i_2 and perform the (i_2, j_2, m, S)-process until the $(\log\log n)^2$th generation or the number of descendants is too large. Continue this procedure until you find a 'good' vertex, i.e. a vertex v such that the numbers of both descendants and ancestors in the $(\log\log n)^2$th generation are at least $\log^2 n$ and with less than $2n^{1/2}$ relatives found so far.

With probability $1 - o(1)$ one finds a good vertex before examining $\log n$ vertices. Indeed, from the Fact, the probability that a vertex is not good is less than $1 - \gamma\gamma'$, so the probability that we fail $\log n$ times is less than $(1 - \gamma\gamma')^{\log n} \to 0$.

Thus, we have found a good vertex v spoiling less than $2n^{1/2} \log n$ vertices. We show that $T^-(v) \geqq 0{\cdot}05m$ and $T^+(v) \geqq 0{\cdot}05m$. Indeed, examine first each of $\log^2 n$ descendants of v not examined yet and stop when the total number of descendants of v reaches $0{\cdot}05m$. The probability that this does not happen is, due to the Fact, less than $(1 - \beta)^{\log^2 n} \to 0$. Thus, with probability $1 - o(1)$ we have checked that $T^+(v) > 0{\cdot}05m$, having spoiled so far less than $0{\cdot}05m + 2n^{1/2} \log n < 0{\cdot}1m$ vertices. Now examine each of the ancestors of v in the $(\log\log n)^2$th generation of inneighbours of v, until the total number of all ancestors of v is larger than $0{\cdot}05m$. The probability that this does not happen, due to the Fact, is less than $(1 - \beta')^{\log^2 n} \to 0$, since each time we look for the parents of a new ancestor of v the total number of spoiled vertices is less than $0{\cdot}1m + 0{\cdot}05m < 0{\cdot}2m$. Hence with probability tending to 1 as $n \to \infty$ we have found a vertex v in $D_n(r, s, t)$ with many ancestors and descendants, and the assertion of Lemma 2 holds.

Appendix

Proof of Lemma 1. When $a = b$ then $f_k(x) \equiv a^k$ and the assertion holds. Moreover, since for a function $g_k(x)$ defined by $g_k(x) = f_k(1 - x)$ we have

$$g_k(x) = (b - a)\int_0^x g_{k-1}(t)\, dt + a\int_0^1 g_{k-1}(t)\, dt,$$

we may assume that $a > b$.

It is not hard to see that $f_k(x)$ is a polynomial of degree k. Let

$$f_k(x) = \sum_{i=0}^{k} c(k; i)x^i.$$

Then $c(0; 0) = 1$, $c(0; i) = 0$ for $i \geqq 1$ and when $k \geqq 1$

$$c(k; 0) = b \sum_{i=0}^{k-1} \frac{c(k-1, i)}{i+1},$$

and when $1 \leqq i \leqq n$

$$c(k; i) = \frac{a - b}{i} c(k - 1; i - 1).$$

Hence

$$c(k;i)=\frac{(a-b)^i}{i!}c(k-i;0)$$

for $k \geqq 1$, $i \geqq 1$ and

$$c(k;0)=b\sum_{i=0}^{k-1}\frac{(a-b)^i}{(i+1)!}c(k-i-1,0)=b\sum_{i=1}^{k}\frac{(a-b)^{i-1}}{i!}c(k-i;0).$$

Thus assume that $b<x(a)$, i.e. that

$$b\sum_{i=1}^{\infty}\frac{(a-b)^{i-1}}{i!}=b\frac{e^{a-b}-1}{a-b}<1.$$

Then, for some $\varepsilon>0$,

$$b\frac{\exp\left(\dfrac{a-b}{1-\varepsilon}\right)-1}{a-b}=1. \tag{13}$$

We shall show that $c(k;0)\leqq(1-\varepsilon)^k$. It is true for $c(0;0)$. For $c(k;0)$ we have

$$c(k;0)\leqq(1-\varepsilon)^k b\sum_{i=1}^{\infty}\frac{(a-b)^{i-1}}{i!}\frac{1}{(1-\varepsilon)^i}=(1-\varepsilon)^k b\frac{\exp\left(\dfrac{a-b}{1-\varepsilon}\right)-1}{a-b}=(1-\varepsilon)^k.$$

Thus

$$\begin{aligned}\max\{f_x(x):x\in[0,1]\}&=\sum_{i=0}^{k}c(k;i)=\sum_{i=0}^{k}\frac{(a-b)}{i!}c(k-i;0)\\&\leqq\sum_{i=0}^{k}\frac{(a-b)^i}{i!}(1-\varepsilon)^{k-i}\leqq(1-\varepsilon)^k\sum_{i=0}^{\infty}\frac{1}{i!}\left(\frac{a-b}{1-\varepsilon}\right)^i\\&=(1-\varepsilon)^k\exp\left(\frac{a-b}{1-\varepsilon}\right)=\frac{a}{b}(1-\varepsilon)^k,\end{aligned}$$

where the last equality follows from (13). Thus the assertion of Lemma 1(i) holds for every A such that $1-\varepsilon<A<1$.

Now let $a>1$ and $b>x(a)$ so that

$$b\frac{e^{a-b}-1}{a-b}=1+2\varepsilon>0$$

for some $\varepsilon>0$ and choose M large enough that

$$b\sum_{i=1}^{M}\frac{(a-b)^{i-1}}{i!}\geqq 1+\varepsilon.$$

Then, for $k \geqq M$,

$$c(k;0) \geqq b \sum_{i=1}^{M} \frac{(a-b)^{i-1}}{i!} c(k-i;0) \geqq (1+\varepsilon) \min\{c(k-i;0) : 1 \leqq i \leqq M\}.$$

Thus the sequence $c(k;0)$ (and therefore $f_k(x)$) increases for $k \geqq M$. For each $k \geqq 2M$,

$$c(k;0) \geqq (1+\varepsilon)c(k-M;0) \geqq (1+\varepsilon)^{\lfloor k/M \rfloor -1} c(k;0).$$

Hence, for $k \geqq 2M$,

$$\min\{f_k(x) : x \in [0,1]\} = c(k;0) \geqq (1+\varepsilon)^{\lfloor k/M \rfloor -1} c(k;0).$$

This proves Lemma 1(ii).

Acknowledgements

We thank C. M. Newman and the referee for reviewing the paper and making helpful suggestions. This work was supported in part by U.S. National Science Foundation grant BSR 87-05047 and the hospitality of Mr and Mrs William T. Golden.

References

Ajtai, M., Komlós, J. and Szemerédi, E. (1982) Largest random component of a k-cube. *Combinatorica* **2**, 1–7.

Cohen, J. E., Luczak, T., Newman, C. M. and Zhou, Z. (1990) Stochastic structure and nonlinear dynamics of food webs: qualitative stability in a Lotka–Volterra cascade model. *Proc. R. Soc. London* B **240**, 607–627.

Erdös, P. and Rényi, A. (1960) On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5**, 17–61.

Grimmett, G. R. (1989) *Percolation.* Springer-Verlag, Berlin.

Harris, T. E. (1963) *The Theory of Branching Processes*: Springer-Verlag, Berlin.

Karp, R. M. (1990) The transitive closure of a random digraph. *Random Structures and Algorithms* **1**, 73–94.

Kesten, H. (1982) *Percolation Theory for Mathematicians.* Birkhaüser, Basle.

Łuczak, T. (1990) The phase transition in the evolution of random digraphs. *J. Graph Theory* **14**, 217–223.

Schmidt-Pruzan, J. and Shamir, E. (1985) Components structure in the evolution of random hypergraphs. *Combinatorica* **5**, 81–94.