The Annals of Applied Probability
1993, Vol. 3, No. 2, 403–420

STRATIGRAPHY OF A RANDOM ACYCLIC DIRECTED GRAPH: THE SIZE OF TROPHIC LEVELS IN THE CASCADE MODEL

BY TOMASZ ŁUCZAK[1] AND JOEL E. COHEN

Rockefeller University

When an ecological food web is described by an acyclic directed graph, the trophic level of a species of plant or animal may be described by the length of the shortest (or the longest) food chain from the species to a green plant or to a top predator. Here we analyze the number of vertices in different levels in a stochastic model of acyclic directed graphs called the cascade model. This model describes several features of real food webs.

For an acyclic directed graph D, define the ith lower (upper) level as the set of all vertices v of D such that the length of the shortest (longest) maximal path starting from v equals i, $i = 0, 1 \dots$. In this article, we compute the sizes of the levels of a random digraph $D(n, c)$ obtained from a random graph on the set $\{1, 2, \dots, n\}$ of vertices in which each edge appears independently with probability c/n, by directing all edges from a larger vertex to a smaller one. The number of edges between any two levels of $D(n, c)$ is also found.

1. Introduction. Ecological food webs are often described by acyclic directed graphs. The trophic level of a species of plant or animal may be described by the length of the shortest (or the longest) food chain from the species to a green plant or to a top predator. Here we analyze the distribution of species (or vertices) by trophic levels (defined in various ways) in a stochastic model of acyclic directed graphs, called the cascade model. This model describes several features of real food webs [Cohen, Briand and Newman (1990)].

Let G be a graph with the set of vertices $[n] = \{1, 2, \dots, n\}$. A digraph (directed graph) $D = D(G)$ has the same set of vertices and (v, w) (or sometimes simply vw) is an arrow, arc or directed edge of D if and only if $\{v, w\}$ is an edge of G and $v > w$. [In Cohen, Briand and Newman (1990) and elsewhere, we adopt the reverse convention that (v, w) is an arc of D if and only if $\{v, w\}$ is an edge of G and $v < w$. Because of the up–down symmetry in the stochastic model we shall define, the difference is immaterial for the calculations we carry out here.]

For each vertex v of $D(G)$ let $h(v)$, $H(v)$ denote the lengths (counting arcs, not vertices) of the shortest and the longest maximal path starting from v. (A maximal path is one that is not contained properly in any other path.)

Received January 1991; revised June 1992.
[1] Supported in part by NSF Grant BSR 87-05047.
AMS 1991 *subject classifications.* 05C80, 05C20, 92D40
Key words and phrases. Digraph, acyclic, random graph, food web, trophic level

Because $D(G)$ contains no directed cycles, $h(v)$ and $H(v)$ are always well defined and $0 \le h(v) \le H(v) \le n-1$. For an integer i, $0 \le i \le n-1$, let the *ith lower* (*upper*) *level* be the set $h_i = h_i(D)$ $[H_i = H_i(D)]$ of all vertices v of D for which $h(v) = i$ $[H(v) = i]$. We shall call the sequences $\{h_i\}_{i=0}^{n-1}$ and $\{H_i\}_{i=0}^{n-1}$ the *lower* and *upper stratigraphy* of D.

In this article we study the stratigraphies of a random digraph $D(n,c) = D(G(n,c/n))$, where c is some positive constant that does not depend on n and $G(n,c/n)$ denotes the random graph with vertex set $[n]$ in which each pair of vertices appears as an edge independently with probability c/n. [Many properties of $G(n,c/n)$ have been studied, for example, Bollobás (1985), as have some properties of $D(n,c)$ other than those considered here, for example, Cohen, Briand and Newman (1990)]. In Section 2, we introduce an easy-to-analyze random process with properties very similar to those of $D(n,c)$ and find the probability that $h(v) = i$ for a given vertex v of $D(n,c)$. In Section 3, we deduce from this, and the fact that the events $h(v) = i$ and $h(w) = i$ for two given vertices v and w are nearly independent, the size of the ith level of the lower and upper stratigraphy. In Section 4, we evaluate numerically the asymptotic formulas for the expected fractions of vertices in different levels and compare the results to those from simulations. For 20 or more vertices, the asymptotic formulas give a good approximation to the sample means, in the sense that the sample means fall within a 95% confidence interval around the values obtained from the asymptotic formulas. For 80 or more vertices, the asymptotic formulas give an excellent approximation, in the sense that the asymptotic fractions and the sample mean fractions differ by less than 0.01. In Section 5, we compute the number of edges between different levels of a stratigraphy. Finally, in Section 6, we characterize very briefly the stratigraphy of $D(G(n,p))$, when $np \to \infty$ but $p \to 0$.

For food webs [Cohen and Newman (1985); Cohen, Briand and Newman (1990)], stratigraphies are possible precise meanings of "trophic levels," a term which is loosely used in ecology. Detailed ecological applications of the concepts developed here appear in Cohen and Łuczak (1992). It may be helpful to sketch the motivation here. In the ecological interpretation of an acyclic directed graph, the vertices correspond to species or other groups of organisms, and the arcs correspond to predator–prey feeding relations. When the arc (v,w) means that species v eats species w, then the vertices of level 0 eat no others and are called *basal* species by ecologists. When the arc (v,w) means that species v is eaten by species w (as is standard in the cascade model [Cohen, Briand and Newman (1990)], then the vertices of level 0 are eaten by no others and are called *top* species by ecologists. The expected proportions of basal and top vertices in the cascade model have been calculated previously [Cohen and Newman (1985)]. The sizes and related properties of the other trophic levels are calculated for the first time in this paper.

The distribution of the number of edges between (and within) levels of a stratigraphy may provide information about some of the factors that regulate the number and numerical abundance of species in a level. If the average number of species that eat species in a given level is large, it seems more likely that species on the given level could be controlled by their predators. This information may have practical use in selecting situations where biological control (the control of pests by their natural enemies) is most likely to be effective. On the other hand, if the species of a given level eat, on average, a small number of prey species, it seems more likely that there might be competition among the species on the given level. Thus stratigraphies may shed light on the regulatory roles of predation and competition, a matter of fundamental interest for current ecological theory [e.g., Schoener (1982)].

2. The level of a vertex. Let us introduce a probabilistic model of a system of communication with properties that are very similar to those of $D(n, c)$. Consider n persons labeled by numbers 1 to n according to their rank. Each of them makes a mailing list by including each person w of higher rank (i.e., labeled by a number smaller than his own) independently with probability $p = c/n$ and sends his own petition to all persons on the list at the same time. Now, every person v who receives a petition makes his own mailing list in which each person w, $w < v$, is present with probability p and forwards the petition, modified a bit, to all persons of the list. (A person makes a new mailing list each time he gets a new petition or another version of the same petition.) Such a procedure for handling petitions we shall call an *unrestricted bureaucratic process*. A *k-restricted bureaucratic process* is defined similarly but each time a person receives a petition that has been received k times he does not make a list of possible recipients but instead puts the petition in the files. We say that $\hat{h}(v) \le k$ $[\hat{H}(v) \le k]$ if some recipient of the petition of v who had received it changed less than k times made an empty mailing list (no copy of a petition of v was changed more than $k - 1$ times).

LEMMA 1. (i) *Define a sequence of functions* $g_k\colon [0,1] \to [0,1]$ *recursively, setting*

$$g_{-1} \equiv 0,$$
$$g_k = 1 + \exp(-cx) - \exp\left(-c\int_0^x g_{k-1}(t)\,dt\right) \quad \text{for } k = 0, 1 \dots . \tag{1}$$

Then for every v, $1 \le v \le n$, *and any fixed* $k = 0, 1, \dots,$

$$\operatorname{Prob}\{\hat{h}(v) \le k\} \le g_k(v/n) + O(n^{-1}).$$

(ii) *Let* $G_k\colon [0,1] \to [0,1]$ *be a sequence of functions defined by*

$$G_{-1} \equiv 0,$$
$$G_k = \exp\left(c\int_0^x G_{k-1}(t)\,dt - cx\right) \quad \text{for } k = 0, 1 \dots . \tag{2}$$

Then for every v, $1 \le v \le n$, *and any fixed* $k = 0, 1, \ldots,$

$$\mathrm{Prob}\{\hat{H}(v) \le k\} \le G_k(v/n) + O(n^{-1}).$$

PROOF. We shall prove the lemma using induction on k. Clearly,

$$\begin{aligned}\mathrm{Prob}\{\hat{h} = 0\} &= \mathrm{Prob}\{\hat{H} = 0\} = (1 - c/n)^{v-1}\\ &= \exp(-cv/n) + O(n^{-1}) = g_0(v/n) + O(n^{-1}).\end{aligned}$$

Now fix $k \ge 1$ and for $w < v$ let $A(v, w)$ be the event that w was on the list of v when v circulated his own petition. Then

$$\begin{aligned}1 - \mathrm{Prob}\{\hat{h}(v) \le k\} &= \mathrm{Prob}\left\{\bigcap_{w<v}\{A(v,w) \text{ and } \hat{h}(w) \ge k-1 \text{ or } \neg A(v,w)\}\right\}\\ &\quad - \mathrm{Prob}\{\text{list of } v \text{ is empty}\}\\ &= \prod_{w<v} \mathrm{Prob}\{A(v,w) \text{ and } \hat{h}(w) \ge k-1 \text{ or } \neg A(v,w)\}\\ &\quad - \mathrm{Prob}\{\text{list of } v \text{ is empty}\}\\ &= \prod_{w<v} (p(1 - g_{k-1}(w/n)) + 1 - p)\\ &\quad -(1-p)^{v-1} + O(n^{-1})\\ &= \exp\left(-\sum_{w<v} g_{k-1}(w/n)\right) - \exp(-cv/n) + O(n^{-1})\\ &= \exp\left(-c\int_0^{v/n} g_{k-1}(t)\,dt\right) - \exp(-cv/n) + O(n^{-1})\\ &= g_k(v/n) + O(n^{-1}).\end{aligned}$$

Similarly

$$\begin{aligned}\mathrm{Prob}\{\hat{H}(v) \le k\} &= \mathrm{Prob}\left\{\bigcap_{w<v}\{A(v,w) \text{ and } \hat{H}(w) \le k-1 \text{ or } \neg A(v,w)\}\right\}\\ &= \prod_{w<v} \mathrm{Prob}\{A(v,w) \text{ and } \hat{H}(w) \le k-1 \text{ or } \neg A(v,w)\}\\ &= \prod_{w<v} (pG_{k-1}(w/n) + 1 - p)\\ &= \exp\left((c/n)\sum_{w<v} (G_{k-1}(w/n) - 1)\right) + O(n^{-1})\\ &= \exp\left(c\int_0^{v/n} G_{k-1}(t)\,dt - cv/n\right) + O(n^{-1})\\ &= G_k(v/n) + O(n^{-1}). \qquad \square\end{aligned}$$

An analogous result holds also for a digraph $D(n, c)$.

THEOREM 1. *Let v be a vertex of $D(n,c)$, k be a nonnegative integer and $\{g_k\}_{k=-1}^{\infty}$ and $\{G_k\}_{k=-1}^{\infty}$ denote sequences of functions defined by* (1) *and* (2). *Then*

$$\operatorname{Prob}\{h(v) \le k\} \le g_k(v/n) + O(n^{-1})$$

and

$$\operatorname{Prob}\{H(v) \le k\} \le G_k(v/n) + O(n^{-1}).$$

PROOF. Fix v and $k \ge 0$ and generate recipients of the petition of v in the k-restricted process until either some person receives the petition twice (we call this event EMERGENCY STOP) or we generate all the recipients of the petition (we call this event NATURAL END). Now, for all w who made a mailing list during the foregoing process, join w by an arc to all recipients of the copy of the petition w sent, and for all other vertices w' join w' with w'' by an arc with probability $p = c/n$, independently for all $w'' < w'$. Clearly, a digraph obtained in such a way can be identified with $D(n,c)$. If the process ends naturally, then the event $h(v) \le k$ $[H(v) \le k]$ is equivalent to $\hat{h}(v) \le k$ $[\hat{H}(v) \le k]$.

Furthermore, the probability of EMERGENCY STOP is precisely the same as the probability of the event $C = C(k,v)$ that in $D(n,p)$ there exists a vertex w such that v is joined with w by two different directed paths, each of size at most k. We shall show later that $\operatorname{Prob}\{C\} = O(n^{-1})$. Thus

$$\begin{aligned}\operatorname{Prob}\{h(v) \le k\} &= \operatorname{Prob}\{h(v) \le k \text{ and } \neg C\} + \operatorname{Prob}\{h(v) \le k \text{ and } C\}\\ &= \operatorname{Prob}\{\hat{h}(v) \le k \text{ and } \neg C\} + O(n^{-1})\\ &= \operatorname{Prob}\{\hat{h}(v) \le k\} + O(n^{-1}) = g_k(v/n) + O(n^{-1})\end{aligned}$$

and

$$\begin{aligned}\operatorname{Prob}\{H(v) \le k\} &= \operatorname{Prob}\{H(v) \le k \text{ and } \neg C\} + \operatorname{Prob}\{H(v) \le k \text{ and } C\}\\ &= \operatorname{Prob}\{\hat{H}(v) \le k \text{ and } \neg C\} + O(n^{-1})\\ &= \operatorname{Prob}\{\hat{H}(v) \le k\} + O(n^{-1}) = G_k(v/n) + O(n^{-1}).\end{aligned}$$

Thus to complete the proof it is enough to check whether $\operatorname{Prob}\{C\} = O(n^{-1})$. We shall prove a slightly stronger result.

FACT. Let $G(n,c/n)$ be an underlying graph of $D(n,c)$. Then the probability that a given vertex of $G(n,c/n)$ is connected by a path of length at most k with some cycle of length at most $2k$ is bounded from above by $3k(1 + c^{3k})/n$. (This includes also the case when v belongs to a cycle; that is, the path is empty.)

PROOF OF THE FACT. Let v be a fixed vertex of $G(n,c/n)$ and let X count subgraphs that consist of a cycle of length i, $i \le 2k$, and a path of length j,

$j \leq k$, that joins v to the cycle. Then

$$\begin{aligned}\text{Prob}\{X > 0\} &\leq EX \leq \sum_{i=3}^{2k} \binom{n}{i} \frac{(i-1)!}{2} \left(\frac{c}{n}\right)^i \sum_{j=0}^{k} i \binom{n}{j-1} (j-1)! \left(\frac{c}{n}\right)^j \\ &\leq \frac{1}{n} \sum_{i=3}^{2k} \sum_{j=0}^{k} c^{i+j} \leq 3k \frac{1 + c^{3k}}{n},\end{aligned}$$

where, for convenience, we set $\binom{n}{-1} = 1$ and $(-1)! = 1$. This completes the proof of the Fact and of Theorem 1. □

REMARK. It is not hard to find explicit formulas for the first few functions $g_k(x), G_k(x)$:

$$\begin{aligned} g_0(x) &= G_0(x) = \exp(-cx), \\ g_1(x) &= 1 + \exp(-cx) - \exp(\exp(-cx) - 1), \\ g_2(x) &= 1 + \exp(-cx) - \exp(-1 - cx + e^{-cx} + e^{-1}(\mathrm{Ei}(1) - \mathrm{Ei}(e^{-cx}))) \end{aligned}$$

and

$$\begin{aligned} G_1(x) &= \exp(1 - cx - e^{-cx}), \\ G_2(x) &= \exp(\exp(1 - e^{-cx}) - cx - 1), \\ G_3(x) &= \exp(e^{-1}(\mathrm{Ei}(\exp(1 - e^{-cx})) - \mathrm{Ei}(1)) - cx), \end{aligned}$$

where Ei denotes the exponential integral, which satisfies, for every $a, b > 0$,

$$\mathrm{Ei}(b) - \mathrm{Ei}(a) = \int_a^b \frac{e^t}{t} \, dt.$$

Because $\mathrm{Ei}(x)$ is not an elementary function, neither are $g_2(x)$ and $G_3(x)$. Moreover, if $g_{k+1}(x)$ $[G_{k+1}(x)]$ is elementary, then so is $g_k(x)$ $[G_k(x)]$. Therefore, none of the functions $g_k(x)$ for $k \geq 2$ and $G_k(x)$ for $k \geq 3$ is elementary.

Nevertheless, one can easily deduce some properties of $g_k(x)$ and $G_k(x)$ from the defining recursive relations. For example, for every $k \geq 0$, $g_k(0) = G_k(0) = 1$. All $g_k(x)$ and $G_k(x)$ are strictly positive and decreasing in x for fixed $k \geq 0$. For every $x_0 \in (0, 1)$, the sequences $\{g_k(x_0)\}_{k=-1}^{\infty}$ and $\{G_k(x_0)\}_{k=-1}^{\infty}$ are increasing. To illustrate, Figure 1 plots $g_k(x)$ and Figure 2 plots $G_k(x)$, $k = 0, 1, 2, 3, 4$. Theorem 2 will show that the area between the curve for k and the curve for $k - 1$ is the asymptotic fraction (as $n \to \infty$) of vertices in level k of the lower and upper stratigraphy, respectively. [For $k = 0$, recall that $g_{-1}(x) = G_{-1}(x) = 0$.]

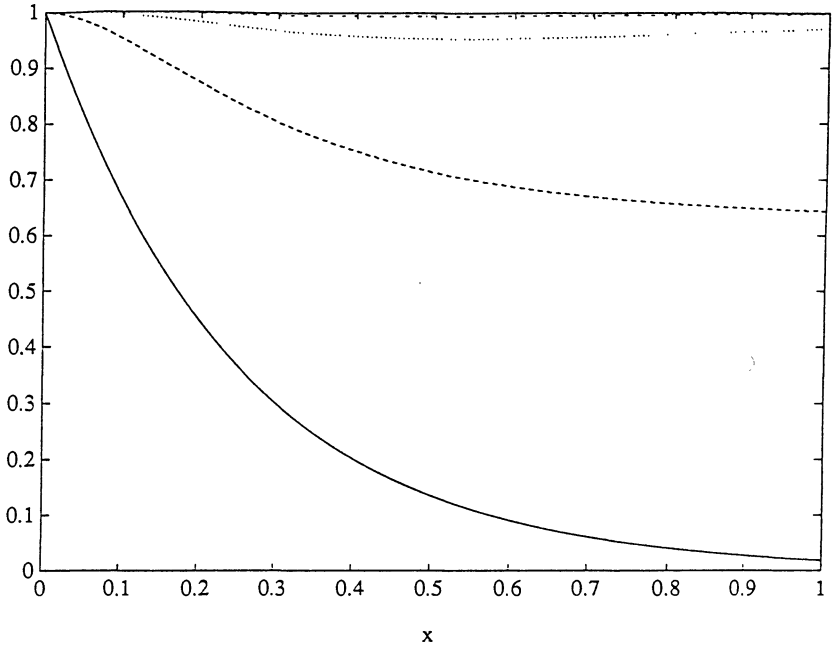

FIG. 1. *Plots of $g_0(x)$ (solid lower curve) and $g_k(x)$, $k = 1, 2, 3, 4$ (remaining curves in ascending order). The area under $g_0(x)$ is the asymptotic fraction (as $n \to \infty$) of vertices in level 0 of the lower stratigraphy, the area between $g_0(x)$ and $g_1(x)$ is the asymptotic fraction of vertices in level 1 of the lower stratigraphy, and similarly for the areas between the following pairs of curves. The curve for $g_4(x)$ is indistinguishable from the horizontal top border of the figure; that is, $g_4(x) \approx 1$.*

3. The size of a level. The sizes of the levels of the lower and upper stratigraphies are as follows:

THEOREM 2. *For any $k = 0, 1, \ldots,$ the number of vertices on the kth level of the lower stratigraphy is a.s.*

$$n\int_0^1 (g_k(x) - g_{k-1}(x))\,dx + o(n^{0.6}).$$

For an upper stratigraphy, the number of vertices of the kth level is a.s.

$$n\int_0^1 (G_k(x) - G_{k-1}(x))\,dx + o(n^{0.6}),$$

where the functions $g_k(x)$ and $G_k(x)$ are defined by (1) *and* (2) *and hereafter a.s. means "with probability tending to* 1 *as $n \to \infty$."*

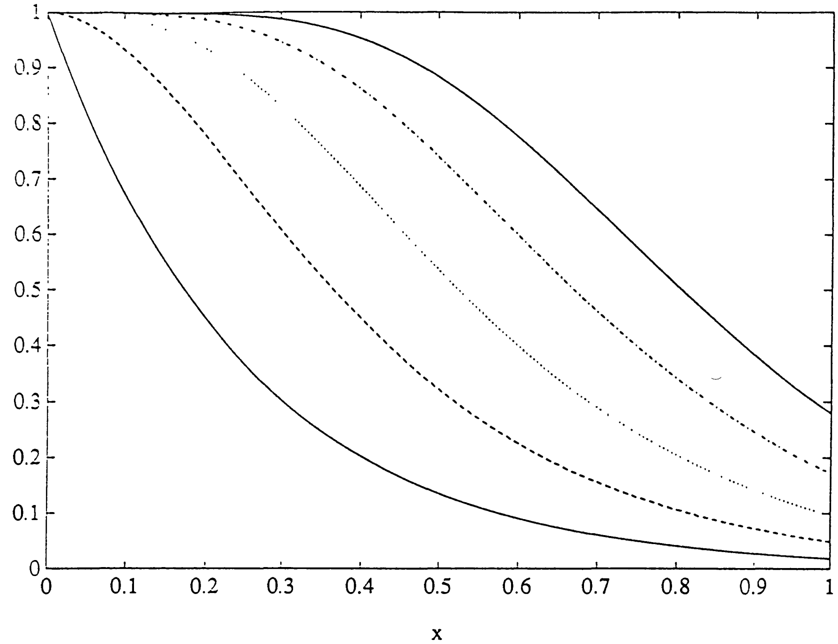

FIG. 2. *Plots of $G_0(x)$ (solid lower curve) and $G_k(x)$, $k = 1,2,3,4$ (remaining curves in ascending order). The area under $G_0(x)$ is the asymptotic fraction (as $n \to \infty$) of vertices in level* 0 *of the upper stratigraphy, the area between $G_0(x)$ and $G_1(x)$ is the asymptotic fraction of vertices in level* 1 *of the upper stratigraphy, and similarly for the areas between the following pairs of curves.*

PROOF. Let X_k be the number of vertices on the kth level of a lower stratigraphy. Then, from Theorem 1,

$$EX_k = \sum_{v \in [n]} \left(g_k\left(\frac{v}{n}\right) - g_{k-1}\left(\frac{v}{n}\right)\right) + O(1)$$

$$= n\int_0^1 (g_k(x) - g_{k-1}(x))\,dx + O(1).$$

We find next the second factorial moment of X_k:

$$E_2X_k = \sum_{\substack{(v,w)\\ v,w \in [n],\, v \neq w}} \mathrm{Prob}\{h(v) = k \text{ and } h(w) = k\}.$$

To estimate $\mathrm{Prob}\{h(v) = k$ and $h(w) = k\}$ we shall use k-restricted bureaucratic processes. As in the proof of Theorem 1, let us generate all recipients of petitions of v or w, stopping whenever one of the recipients is forced to make a second mailing list; then we generate the digraph $D(n,c)$ as described in the proof of Theorem 1. The probability of EMERGENCY STOP is the same as the probability of the event $C' = C'(n,c;v,w)$ that $D(n,c)$ contains a vertex w' and two different directed paths P_1, P_2 of length at most

k such that P_i, $i = 1, 2$, starts at v or w and ends at w'. The probability of C' can be bounded from above by $2\,\mathrm{Prob}\{C'''\} + \mathrm{Prob}\{C''\}$, where C''' is the event defined in the Fact and C'' denotes the probability that v and w are joined in $G(n, c/n)$ by a path of length at most $2k$. Then, as in the Fact,

$$\mathrm{Prob}\{C''\} \le \sum_{i=0}^{2k} \binom{n}{i} i! \left(\frac{c}{n}\right)^{i+1} \le k(1 + c^{2k+1})/n,$$

so

$$\mathrm{Prob}\{C'\} \le 10k(1 + c^{3k})/n = O(n^{-1}).$$

Thus, arguing as in the proof of Theorem 1, we get

$$\begin{aligned} &\mathrm{Prob}\{h(v) = k \text{ and } h(w) = k\} \\ &\quad = \mathrm{Prob}\{\hat{h}(v) = k \text{ and } \hat{h}(w) = k\} + O(n^{-1}) \\ &\quad = \mathrm{Prob}\{\hat{h}(v) = k\}\mathrm{Prob}\{\hat{h}(w) = k\} + O(n^{-1}) \\ &\quad = (g_k(v/n) - g_{k-1}(v/n))(g_k(w/n) - g_{k-1}(w/n)) + O(n^{-1}). \end{aligned}$$

Hence

$$\begin{aligned} E_2 X_k &= \sum_{v \in [n]} (g_k(v/n) - g_{k-1}(v/n)) \sum_{w \in [n], w \ne v} (g_k(w/n) \\ &\qquad\qquad - g_{k-1}(w/n)) + O(n) \\ &= \left(n \int_0^1 (g_k(x) - g_{k-1}(x))\, dx\right)^2 + O(n) = (EX_k)^2 + O(n). \end{aligned}$$

The variance of X_k is then

$$\mathrm{Var}\, X_k = E_2 X_k - (EX_k)^2 + EX_k + O(n).$$

From Chebyshev's inequality we get

$$\mathrm{Prob}\{|X_k - EX_k| \ge n^{0.55}\} \le \frac{\mathrm{Var}\, X_k}{n^{1.1}} = O(n^{-1}).$$

A similar argument yields the asserted formula for the size of levels of the upper stratigraphy. □

REMARK 1. It follows that the asymptotic fraction of vertices in level 0 of both the lower and the upper stratigraphy is

$$\lim_{n \to \infty} EX_0/n = \int_0^1 g_0(x)\, dx = \int_0^1 G_0(x)\, dx = \int_0^1 e^{-cx}\, dx = c^{-1}(1 - e^{-c}),$$

which confirms formula (6.2a) of Cohen and Newman [(1985), page 434]. By elementary calculations,

$$\lim_{n\to\infty} \frac{EX_1}{n} = 1 - \frac{\mathrm{Ei}(1) - \mathrm{Ei}(e^{-c})}{ec},$$

$$\lim_{n\to\infty} \frac{EX_2}{n} = \frac{\mathrm{Ei}(1) - \mathrm{Ei}(e^{-c})}{ec} - \int_0^1 \exp\{e^{-cx} - cx - 1 + e^{-1}[\mathrm{Ei}(1) - \mathrm{Ei}(e^{-cx})]\}\,dx.$$

If Y_k is the number of vertices in level k of the upper stratigraphy, then

$$\lim_{n\to\infty} EY_1/n = c^{-1}(-2 + e^{-c} + \exp[1 - e^{-c}]),$$

$$\lim_{n\to\infty} EY_2/n = c^{-1}\{1 - \exp[1 - e^{-c}] - e^{-1}[\mathrm{Ei}(1) - \mathrm{Ei}(\exp[1 - e^{-c}])]\},$$

$$\lim_{n\to\infty} EY_3/n = (ec)^{-1}[\mathrm{Ei}(1) - \mathrm{Ei}(\exp[1 - e^{-c}])] + \int_0^1 \exp\{-cx + e^{-1}[\mathrm{Ei}(\exp[1 - e^{-cx}]) - \mathrm{Ei}(1)]\}\,dx.$$

REMARK 2. Although Theorem 2 was proved only for a level number k that does not depend on n, an analogous result holds for k tending to infinity as n increases. We omitted this case here because the fraction of arcs of $D(n, c)$ with one end in vertices v for which either $h(v) \geq \omega(n)$ or $H(v) \geq \omega(n)$ tends to 0 as $\omega(n) \to \infty$.

REMARK 3. What is the height (i.e., maximum level) of a stratigraphy? For an upper stratigraphy, the height is asymptotically of order $\log n/\log\log n$ a.s. [Newman and Cohen (1986)]. For a lower stratigraphy it follows from counting isolated paths in $D(n, c)$ that a.s. for some vertices v we have $h(v) \geq (1 - o(1))\log n/\log\log n$, whereas Newman and Cohen (1986) implies that a.s., for every v,

$$h(v) \leq H(v) \leq (1 + o(1))\frac{\log n}{\log\log n}.$$

Thus, $h(v) \sim H(v) \sim \log n/\log\log n$.

4. The size of a level: Numerical analysis. The asymptotic expected fractions of vertices in each upper and lower level were evaluated numerically using the recursive relations for $g_k(x)$ and $G_k(x)$. Numerical quadrature was carried out using the trapezoidal rule and a discretization of $[0, 1]$ into 4000 equal subintervals. The parameter c was taken as 4.0, which is the best current estimate based on available ecological data [Cohen (1990)]. Tables 1 and 2 show the results for upper and lower levels, respectively, under the column headed $n \to \infty$. For example, from Table 2, on average, for

TABLE 1
Fraction of vertices in each upper level in the cascade model $D(n, c)$ with $c = 4.0$. For $n \to \infty$, fractions are computed by numerical quadrature. For each finite n, fractions are computed from 500 simulations of the cascade model

Level number	$n \to \infty$	Vertices n			
		10	20	40	80
0	0.2454	0.2496	0.2474	0.2464	0.2467
1	0.1720	0.2108	0.1961	0.1800	0.1771
2	0.1391	0.1866	0.1663	0.1514	0.1468
3	0.1172	0.1592	0.1459	0.1291	0.1261
4	0.0984	0.1110	0.1126	0.1044	0.1064
5	0.0798	0.0580	0.0717	0.0820	0.0822
6	0.0608	0.0184	0.0366	0.0542	0.0554
7	0.0424	0.0050	0.0167	0.0306	0.0331
8	0.0263	0.0014	0.0058	0.0146	0.0161
9	0.0143	0.0	0.0007	0.0051	0.0065
10	0.0068	—	0.0002	0.0015	0.0024
11	0.0028	—	0.0	0.0007	0.0008
12	0.0010	—	0.0	0.0002	0.0003
13	0.0003	—	0.0	0.0001	0.0001
14	0.0001	—	0.0	0.0	0.0001
Total	1.0068	1.0000	1.0000	1.0000	1.0000

large n, 0.5128 of vertices fall in lower level 1. Remarkably, as the number of vertices becomes arbitrarily large, 97% of vertices fall in the first three levels of the lower stratigraphy, and more than 77% percent of vertices fall in the first five levels of the upper stratigraphy.

For practical applications, it is essential to know how rapidly the limiting fractions are approached when n is finite. Tables 1 and 2 also show the average fractions of vertices in each upper and lower level, respectively, in 500 simulations each for $n = 10, 20, 40, 80$. Each simulation consisted of

TABLE 2
Fraction of vertices in each lower level in the cascade model $D(n, c)$ with $c = 4.0$. For $n \to \infty$, fractions are computed by numerical quadrature. For each finite n, fractions are computed from 500 simulations of the cascade model

Level number	$n \to \infty$	Vertices n			
		10	20	40	80
0	0.2454	0.2496	0.2474	0.2464	0.2467
1	0.5128	0.5514	0.5329	0.5221	0.5150
2	0.2085	0.1834	0.1949	0.2000	0.2060
3	0.0291	0.0142	0.0221	0.0276	0.0281
4	0.0037	0.0014	0.0021	0.0034	0.0036
5	0.0005	0.0	0.0006	0.0006	0.0006
Total	1.0001	1.0000	1.0000	1.0000	1.0000

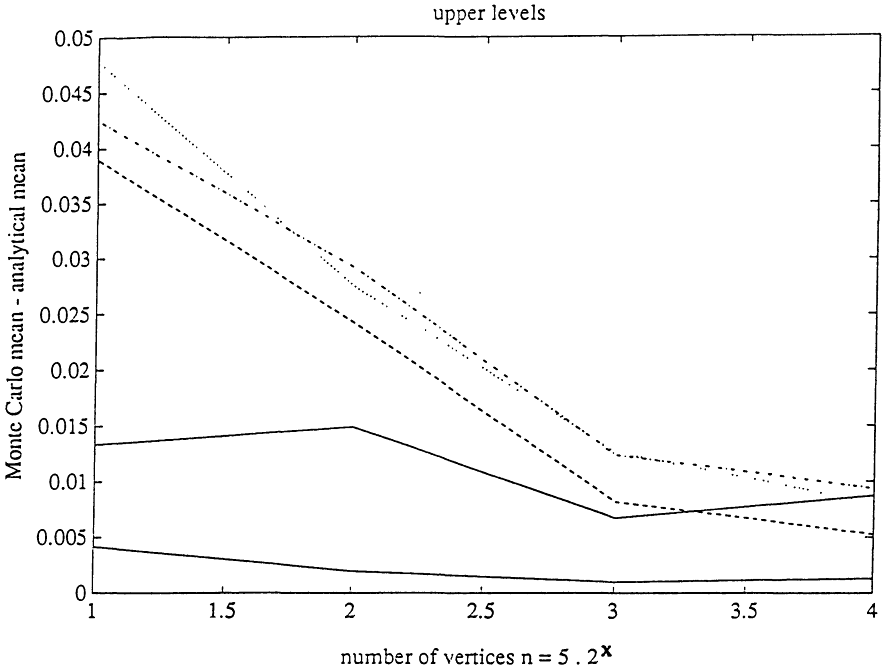

FIG. 3. *Difference between the sample mean fraction* (500 *simulations*) *and the asymptotic mean fraction of vertices in each upper level, as a function of the number n of vertices. For* $x = 1, 2, 3, 4$, $n = 5 \times 2^x = 10, 20, 40, 80$. *Reading from top to bottom at the left side of the figure, the lines correspond to upper levels* 2, 3, 1, 4 *and* 0.

generating a random acyclic directed graph according to the cascade model $D(n, 4)$, computing the number of vertices in each upper level by a minor modification of the algorithm described by F. R. K. Chung [Cohen, Briand and Newman (1990), page 148] and computing the number of vertices in each lower level by the standard algorithm for shortest paths [Robinson and Foulds (1980), page 143]. No vertices were observed to have upper level more than 14 or lower level more than 5. Figures 3 and 4 plot the difference between the simulated value for finite n and the analytical value for infinite n, as a function of the number n of vertices. The simulated values converge rapidly to the limiting values.

To provide a quantitative basis for comparing the simulation results with those from the asymptotic formulas, the 95% confidence intervals (corrected for continuity [Snedecor and Cochran (1967), page 211]) corresponding to each asymptotic value were computed. For example, in Table 2, the probability is 0.95 that, in a sample of 500 with probability 0.5128 (the expected fraction of vertices in lower level 1), the sample mean will fall between 0.4680 and 0.5576. For $n = 10$ vertices, several sample mean fractions (e.g., for upper levels 1, 2, and 3) fall outside of the corresponding 95% confidence intervals. For $n = 20$ vertices, all the sample fractions fall within (or almost

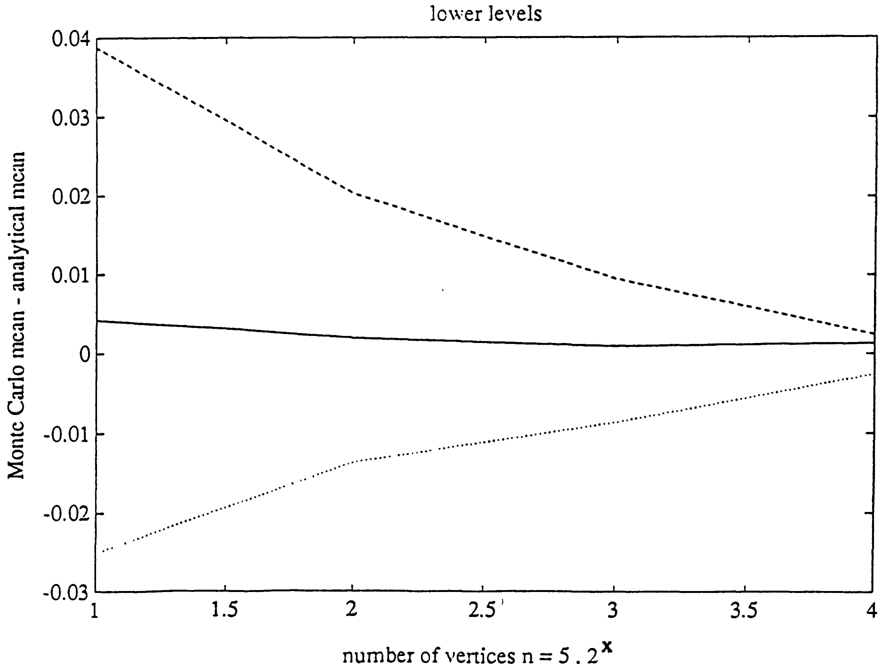

FIG. 4. *Difference between the sample mean fraction* (500 *simulations*) *and the asymptotic mean fraction of vertices in each lower level, as a function of the number n of vertices. For* $x = 1, 2, 3, 4$, $n = 5 \times 2^x = 10, 20, 40, 80$. *Reading from top to bottom at the left side of the figure, the lines correspond to lower levels* 1, 0 *and* 2.

within, in the case of upper level 3) the corresponding 95% confidence intervals. For $n = 80$ vertices, the simulated sample means and the values obtained from analytical formulas for large n agree within 0.01 (except for upper level 8, which differs a tiny bit more), providing numerical confirmation of the correctness of both the asymptotic formulas and the simulation algorithms.

All numerical calculations and simulations were performed using Matlab, version 3.5g, running on a Sperry PC/IT (clone of an IBM PC/AT) under MS-DOS 3.10 version 1.20. The elapsed time per simulation varied as a function of n very nearly as $t = 0.0527n^{1.7}$ sec.

The conclusion from this analysis is that, for 20 or more vertices, the asymptotic formulas hold to good approximation, and for 80 or more vertices, the asymptotic formulas give an excellent approximation.

5. Edges between different levels. The number of edges between two levels of a stratigraphy can be found by the same method as before. Recall that vertex w is an in-neighbour of v if (w, v) is an arc of the digraph.

THEOREM 3. *Let* i, k, l *be integers such that* $i \geq 0$ *and* $0 < l - 1 \leq k$. *Then a.s., the number of vertices* v *such that* $h(v) = k$ *and* v *has exactly* i

in-neighbours at the lth level of a lower stratigraphy is

$$n\int_0^1 (g_k(v/n) - g_{k-1}(v/n))\frac{1}{i!}[r(x,l)]^i \exp(-r(x,l))\,dx + o(n^{-0.6}),$$

where

$$(3)\qquad r(x,l) = \begin{cases} c\int_x^1 (g_l(y) - g_{l-1}(y))\,dy, & \text{for } l \le k,\\ c\int_x^1 (1 - g_k(y) + g_0(y))\,dy, & \text{for } l = k+1.\end{cases}$$

PROOF. Let $v = \lfloor xn \rfloor \in [n] = \{1, 2, \dots, n\}$ and for each $w \in [n]$, $w > v$, let us draw an arc (w, v) independently with probability $p = c/n$. For $w > v$ we denote by $B_k(w, v)$ the event that vertex $\hat{h}(w) = k$ in the bureaucratic process on the set $[n] - \{v\}$ and

$$C_{k,l}(w,v) = \begin{cases} B_k(w,v), & \text{if } l \le k,\\ B_0(w,v) \cup \bigcup_{i=k+1}^{\infty} B_i(w,v), & \text{if } l = k+1.\end{cases}$$

[The special treatment of the case $l = k + 1$ follows from the fact that if v belongs to the kth level of a lower stratigraphy and w is an in-neighbour of v, then at least one maximal path of length $k + 1$ starts at w. Thus, such a w belongs to the $(k + 1)$th level of the stratigraphy if and only if either w is joined by an arc only to v or each maximal path starting from w that does not contain v has length as least $k + 1$.] Furthermore, let $F_{k,l} = F_{k,l}(n, c; v, i)$ denote the event that v has precisely i in-neighbours $w_1, w_2, \dots, w_i$, for which $C_{k,l}(w_i, v)$ and $\hat{h}(v) = k$ in the bureaucratic process on the set $[n]$. We shall show first that the probability of the event $F_{k,l}$ is given by

$$(4)\qquad \left(g_k\left(\frac{v}{n}\right) - g_{k-1}\left(\frac{v}{n}\right)\right)\frac{1}{i!}[r(x,l)]^i \exp(-r(x,l))\,dx + O(n^{-1}).$$

Indeed, from the definition of $F_{k,l}$,

$$F_{k,l} = \left\{\hat{h}(v) = k\right\} \text{ and } \bigcup_{w_1 > w_2 > \cdots > w_i > v} \left\{ \bigcap_{j=1}^{k} \left\{A(w_j, v) \text{ and } C_{k,l}(w_j, v)\right\} \right.$$

$$\left. \text{and } \bigcap_{\substack{w > v\\ w \ne w_j\\ j = 1, 2, \dots, i}} \left\{\neg A(w,v) \text{ or } \left\{A(w,v) \text{ and } C_{k,l}(w_j,v)\right\}\right\}\right\},$$

where $A(w, v)$ denotes the event that w and v are joined by an arc. Using the independence properties of the bureaucratic process and Lemma 1, we get

$$\begin{aligned}\operatorname{Prob}\{F_{k,l}\} &= \left(g_k\left(\frac{v}{n}\right) - g_{k-1}\left(\frac{v}{n}\right)\right) \sum_{w_1 > w_2 > \cdots > w_i > v} \left(\prod_{j=1}^{k} p \operatorname{Prob}\{C_{k,l}(w_j, v)\} \right. \\ &\qquad \left. \times \prod_{\substack{w > v \\ w \neq w_j \\ j = 1, 2, \ldots, i}} \left(1 - p + p \operatorname{Prob}\{C_{k,l}(w_j, v)\}\right) \right) + O(n^{-1}) \\ &= \left(g_k\left(\frac{v}{n}\right) - g_{k-1}\left(\frac{v}{n}\right)\right) \frac{1}{i!} \left(c \int_x^1 \operatorname{Prob}\{C_{k,l}(\lfloor yn \rfloor, v)\}\, dy\right)^i \\ &\qquad \times \exp\left(-c \int_x^1 \operatorname{Prob}\{C_{k,l}(\lfloor yn \rfloor, v)\}\, dy\right) + O(n^{-1}).\end{aligned}$$

But, due to Lemma 1,

$$\operatorname{Prob}\{C_{k,l}(w,v)\} = \begin{cases} g_l(w/n) - g_{l-1}(w/n) + O(n^{-1}), & \text{if } l \leq k, \\ 1 - g_k(w/n) + g_0(w/n) + O(n^{-1}), & \text{if } l = k+1, \end{cases}$$

so (4) follows.

Now we may generate the random digraph $D(n, c)$ similarly to the way we did in the proofs of Theorems 1 and 2. We have already found all arcs (w, v), $w > v$. Generate all recipients of petitions of v in the k-restricted bureaucratic process on the set $[n]$ and all recipients of petitions issued by in-neighbours of v in the l-restricted bureaucratic process on the set $[n]$. The process stops if either some person receives a petition twice (or v or one of its in-neighbours gets a petition issued by another person) or we generate all recipients of petitions. Join all persons w to persons on the mailing list of w (if he made one) and, finally, draw arcs between all "nonexamined" pairs w', w'', independently with probability $p = c/n$. Clearly, the digraph we get could be identified with $D(n, c)$ and one may show, using the stopping time argument, that (4) gives the probability that for a vertex v we have $h(v) = k$ and v has exactly i in-neighbours among vertices on the lth upper level. Then, to get the assertion, it is enough to compute the expectation and variance of the number of such vertices and use Chebyshev's inequality. Because the method is almost identical with that employed in Theorems 1 and 2, we omit the details. □

The analogous result for an upper stratigraphy can be proved similarly.

THEOREM 4. *Let i, k, l be integers such that $i \geq 0$, $0 \leq k \leq l - 1$. Then a.s. the number of vertices on the kth level with exactly i in-neighbours on the lth* *level of an upper stratigraphy is given by*

$$n\int_0^1\left(G_k\left(\frac{v}{n}\right)-G_{k-1}\left(\frac{v}{n}\right)\right)\frac{1}{i!}[R(x,l)]^i\exp(-R(x,l))\,dx+o(n^{-0.6}),$$

where

$$R(x,l)=\begin{cases}c\int_x^1(G_l(y)-G_{l-1}(y))\,dy, & \text{when } l>k+1,\\ c\int_x^1 G_{k+1}(y)\,dy, & \text{when } l=k+1.\end{cases}$$

REMARK. The same method can be used to determine the number of vertices v in level k with exactly $i_1, i_2, \dots, i_m$ in-neighbours in each of levels $l_1, l_2, \dots, l_m$.

To get the number of arrows between two levels, sum up the appropriate values given by Theorems 3 and 4.

COROLLARY 1. *The number of arrows between levels k and $k+1$ a.s. equals*

$$cn\int_0^1(G_k(x)-G_{k-1}(x))\int_x^1 G_{k+1}(y)\,dy\,dx+o(n^{0.6})$$

for an upper statigraphy and

$$cn\int_0^1(g_k(x)-g_{k-1}(x))\int_x^1(1-g_k(y)+g_0(y))\,dy\,dx$$
$$+cn\int_0^1(g_{k+1}(x)-g_k(x))\int_x^1(g_k(y)-g_{k-1}(y))\,dy\,dx+o(n^{0.6})$$

for a lower stratigraphy, where the first term counts arrows from the $(k+1)$th level to the kth level and the second term (which one should omit when $k=0$) counts arrows that go in the reverse direction.

COROLLARY 2. *The number of arrows between level k and level l, where $k \geq l+2$, a.s. equals*

$$cn\int_0^1(g_k(x)-g_{k-1}(x))\int_x^1(g_l(y)-g_{l-1}(y))\,dy\,dx+o(n^{0.6})$$

for a lower stratigraphy. Similarly, for an upper stratigraphy, the number of arrows between level l and level k, where $l \geq k+2$, a.s. equals the same expression with g replaced by G.

REMARK. In the lower stratigraphy, all these arrows go from higher to lower levels, whereas in the upper stratigraphy, the arrows go from lower to higher levels.

COROLLARY 3. *The number of arrows between vertices within the kth level, $k \geq 1$, of a lower stratigraphy is a.s.*

$$cn\int_0^1 (g_k(x) - g_{k-1}(x)) \int_x^1 (g_k(y) - g_{k-1}(y))\, dy\, dx + o(n^{0.6}).$$

6. Stratigraphy of dense random acyclic digraphs. So far we have considered only the cascade model $D(G(n,p))$, where $p = c/n$ for some positive constant c. Biological observations [Cohen (1990)] suggest considering, in addition, the superlinear case when pn grows with n. Suppose the arc probability p depends on n according to a function $p(n)$ such that $p(n) \to 0$ but $np(n) \to \infty$ as $n \to \infty$. Then a.s. the level 0 of a digraph $D(G(n,p))$ contains $(1 + o(1))/p$ vertices. Thus the fraction of vertices in level 0 is a.s. $(1 + o(1))/(pn) \to 0$ in both the upper and the lower stratigraphies. However, the behavior of the levels higher than 0 differs in the two stratigraphies.

In the upper stratigraphy, if k is a fixed finite integer, the size of the kth level tends to c_k/p, where c_k is a positive constant for given k. However, this constant tends to 0 with increasing k, so the level number, say, $\log\log n$ has size of order $1/p$. What is more interesting is the fact that the profile of the upper stratigraphy (i.e., the ratio between the sizes of levels k and l) does not depend very much on the function $p(n)$. For example, the number of vertices in level 1 is $e - 2 + o(1)$ times the number of vertices in level 0.

In the lower stratigraphy, the situation is quite different. There are a.s. $(1 - e^{-1} - o(1))n$ vertices in the first level and $(e^{-1} + o(1))n$ vertices on the second level. Thus, the levels higher than 2 contain only $o(n)$ vertices combined. The numerical data in Table 2 show that the fraction of vertices in lower levels 3 and higher is quite small even for n as small as 10 and $np(n)$ as small as 4. These derivations from the superlinear cascade model may explain why many ecologists intuitively describe natural food webs in terms of three trophic levels. Corresponding to level 0 are the species of green plants, sometimes called primary producers. Corresponding to level 1 are the herbivores or consumers of green plants. Corresponding to level 2 are the species of carnivores that eat herbivores. Even though the superlinear cascade model does not exclude longest chains of considerable length [Newman (1992)], the model predicts that a small fraction of species should be more than 2 feeding links distant (by the shortest path) from the green plants. Further, this small fraction of species in level 3 and higher should vanish as $n \to \infty$ and $np(n) \to \infty$. A substantial fraction of species may be observed at level 0 in the available data because, in these data, $np(n)$ is only about 4, rather than extremely large.

Acknowledgments. We thank Joel Spencer and an anonymous referee who pointed out a serious problem in a previous draft. We are very grateful for the hospitality of Mr. and Mrs. William T. Golden and the friendly support of the computing staff of the Institute for Advanced Study, Princeton.

REFERENCES

BOLLOBÁS, B. (1985). *Random Graphs*. Academic, New York.

COHEN, J. E. (1990). A stochastic theory of community food webs: VI Heterogeneous alternatives to the cascade model. *Theoret. Population Biol.* **37** 55–90.

COHEN, J. E., BRIAND, F. and NEWMAN, C. M. (1990). *Community Food Webs: Data and Theory. Biomathematics* **20**. Springer, Berlin.

COHEN, J. E. and ŁUCZAK, T. (1992). Trophic levels in community food webs. *Evolutionary Ecology* **6** 73–89.

COHEN, J. E. AND NEWMAN, C. M. (1985). A stochastic theory of community food webs: I. Models and aggregated data. *Proc. Roy. Soc. London Ser. B* **224** 421–448.

NEWMAN, C. M. (1992). Chain lengths in certain random directed graphs. *Random Structures and Algorithms* **3** 243–253.

NEWMAN, C. M. and COHEN, J. E. (1986). A stochastic theory of community food webs: IV. Theory of food chain lengths in large webs. *Proc. Roy. Soc. London B* **228** 335–377.

ROBINSON, D. F. and FOULDS, L. R. (1980). *Digraphs: Theory and Techniques*. Gordon and Breach, New York.

SCHOENER, T. W. (1982). The controversy over interspecific competition. *Amer. Sci.* **70** 586–595.

SNEDECOR, G. W. and COCHRAN, W. G. (1967). *Statistical Methods*, 6th ed. Univ. Iowa Press, Ames.

DEPARTMENT OF DISCRETE MATHEMATICS
ADAM MICKIEWICZ UNIVERSITY
MATEJKI 48149
60-769 POZNAŃ
POLAND

LABORATORY OF POPULATIONS
ROCKEFELLER UNIVERSITY
1230 YORK AVENUE, BOX 20
NEW YORK, NEW YORK 10021-6399