

Probability and  
Stochastics Series

---

Topics in  
Contemporary  
Probability and  
Its Applications

---

Edited by  
J. Laurie Snell



CRC Press

Boca Raton New York London Tokyo

**Library of Congress Cataloging-in-Publication Data**

Topics in contemporary probability and its application / [compiled  
by] J. Laurie Snell

p. cm. -- (Probability and stochastics series)

Includes bibliographical references and index.

ISBN 0-8493-8073-1 (alk. paper).

1. Stochastic processes. 2. Probabilities. I. Snell, J. Laurie  
(James Laurie), 1925— . II. Series.

QA274.T65 1994

519.2--dc20

94-23472

CIP

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

Neither this book nor any part may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, microfilming, and recording, or by any information storage or retrieval system, without prior permission in writing from the publisher.

All rights reserved. Authorization to photocopy items for internal or personal use, or the personal or internal use of specific clients, may be granted by CRC Press, Inc., provided that \$.50 per page photocopied is paid directly to Copyright Clearance Center, 27 Congress Street, Salem, MA 01970 USA. The fee code for users of the Transactional Reporting Service is ISBN 0-8493-8073-1/95/\$0.00+\$.50. The fee is subject to change without notice. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

CRC Press, Inc.'s consent does not extend to copying for general distribution, for promotion, for creating new works, or for resale. Specific permission must be obtained in writing from CRC Press for such copying.

Direct all inquiries to CRC Press, Inc., 2000 Corporate Blvd., N.W., Boca Raton, Florida 33431.

© 1995 by CRC Press, Inc.

No claim to original U.S. Government works

International Standard Book Number 0-8493-8073-1

Library of Congress Card Number 94-23472

1 2 3 4 5 6 7 8 9 0

Printed on acid-free paper

# RANDOM GRAPHS IN ECOLOGY

Joel E. Cohen  
Rockefeller University

### ABSTRACT

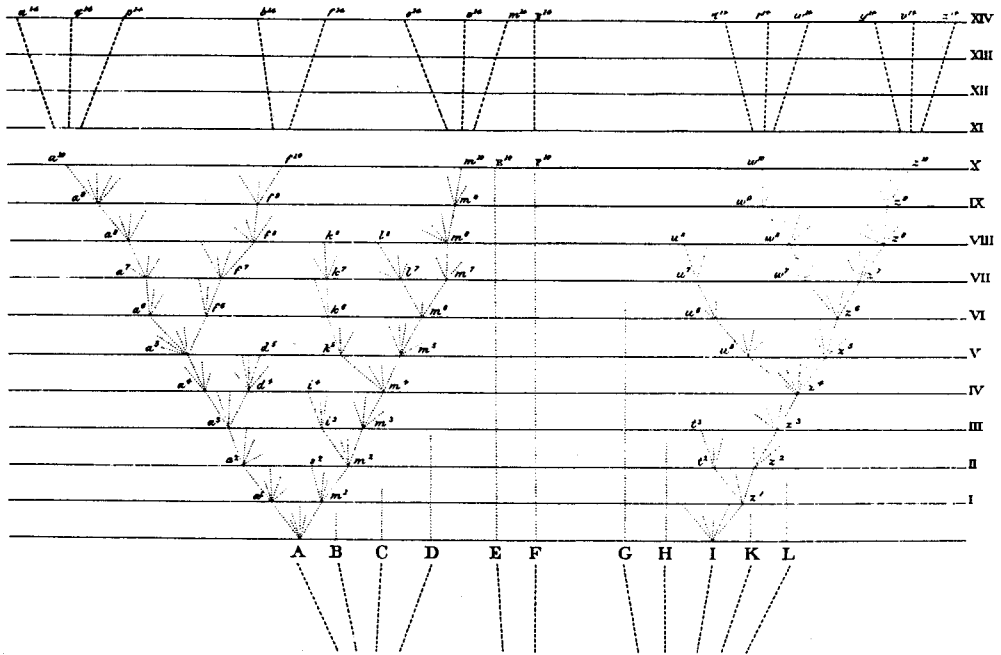
Random graphs have described relations among biological species at least since Charles Darwin's *Origin of Species* (1859). In ecology, random directed graphs called food webs describe which species eat which species. When large numbers of real food webs are viewed as an ensemble, empirical patterns appear. Stochastic models of food webs can explain and unify some of these empirical patterns. This chapter gives an elementary account of how stochastic models of food webs are constructed, analyzed and interpreted. Connections are described with models of parallel computing and the qualitative of large systems of nonlinear ordinary differential equations.

### 1. HISTORICAL BACKGROUND

The first edition of Charles Darwin's book *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, contains only one picture. The picture is a fold-out inserted between pages 116 and 117. Because the first edition has been reissued in a facsimile edition (Darwin 1859 [1964], [14]), it is not necessary to collect rare books to enjoy the picture (Figure 1).

The picture is a diagrammatic sketch of a small portion of what Darwin later (p. 130) calls "the great Tree of Life," showing how (p. 123) "two or more genera are produced by descent, with modification, from two or more species of the same genus." One node of the tree is labelled (A). Darwin explains: "Let (A) be a common, widely-diffused, and varying species, belonging to a genus large in its own country. The little fan of diverging dotted lines of unequal lengths proceeding from (A), may represent its varying offspring. The variations are supposed to be extremely slight, but of the most diversified nature; they are not supposed all to appear simultaneously, but often after long intervals of time; nor are they all supposed to endure for equal periods. Only those variations which are in some way profitable will be preserved or naturally selected." Some of the dotted lines terminate in extinction; others

give rise to new varieties or species. Lines emanating from these new varieties or species give rise to other new varieties or species, which in turn eventually give rise to new genera.



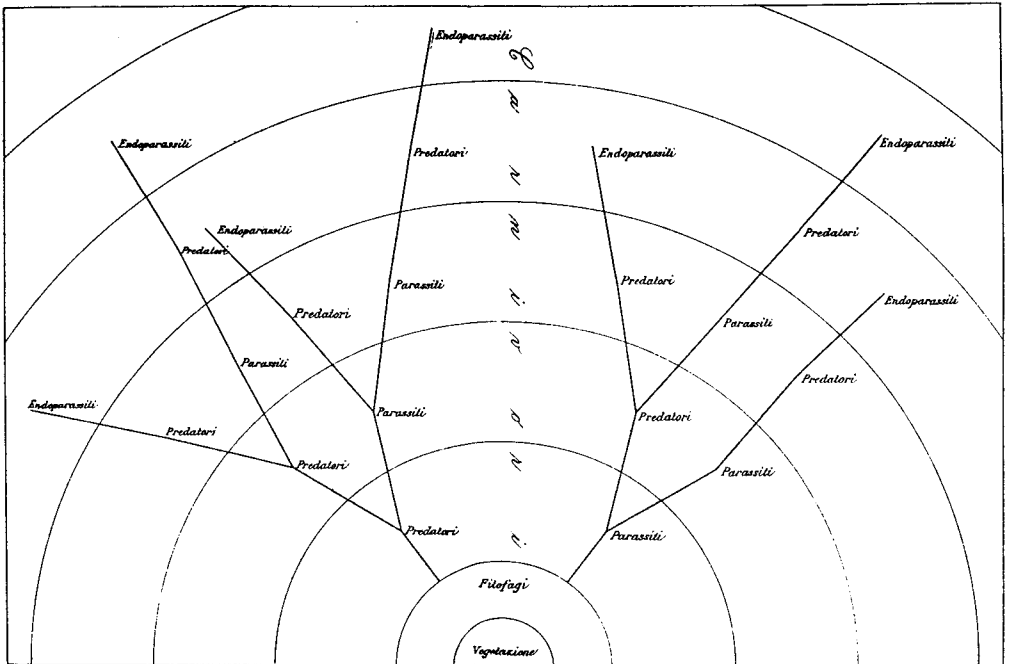
**FIGURE 1.** Evolution of new genera from old, according to the sole illustration of Charles Darwin's *Origin of Species*. Source: Darwin (1859 [1964], between p. 116 and 117).

The picture appeared in 1859, when the theory of probability was relatively young (Pierre Simon Laplace had first published his *Théorie des Probabilités* in 1812, but Isaac Todhunter was not to publish his *History of the Mathematical Theory of Probability* until 1865), the theory of graphs (apart from Leonhard Euler's great initiatives) was little developed, and the theory of random graphs lay a century in the future.

Yet to modern eyes, the picture is clearly a realization of a random directed graph. It is in fact a random directed tree. Charles Darwin made a simplified graph-theoretic model of nature, though I do not know whether he was the first to do so. Darwin symbolized all the complexity of a species or variety by a node of a graph, and all the complexity of a relation between two species or varieties (in Darwin's case, the relation of descent) by a directed edge, or arc, of the graph.

Darwin's picture has been as influential scientifically as the book it illustrated. His picture provides the prototype of today's phylogenetic trees, which are drawn based on results of molecular biology and numerical taxonomy that Darwin could not even imagine. Beyond summarizing evolutionary relationships, the picture's way of representing relationships among varieties or species influenced ecologists, though the words "ecology" and "ecologist" had not been invented when *Origin* appeared. We shall pursue the ecological thread of the story.

In the 1870s, in America, according to Stephen Alfred Forbes, "under the influence of Darwin and Agassiz and Huxley, a transforming wave of progress was sweeping through college and school, a wave whose strong upward swing was a joy to those fortunate enough to ride on its crest, but which smothered miserably many an unfortunate whose feet were mired in marsh mud" (Forbes 1977, [16], p. 10), in a retrospective speech he gave in 1907). Forbes, then at the Illinois State Normal University, published detailed empirical studies of the food of birds (1877) and fishes (1878) of Illinois and the Great Lakes. Forbes's major papers from 1878 onward have recently been reissued in facsimile (Forbes 1977, [16]).



**FIGURE 2.** Camerano's schema of feeding relations. *Vegetazione* = vegetation, *Fitofagi* = phytophagous animals, *Predatori* = predators, *Parassiti* = parasites, *Carnivori* = carnivores, *Endoparasiti* = endoparasites. Source: Camerano (1880, [3]), plate IX).

In 1880, Forbes summarized long lists of the food of Illinois fishes in tabular form. He listed categories of food eaten by fishes as the headings of columns, and the names of particular species of fishes as the labels of rows. As the entries in each row, he reported how many of each kind of fish he examined and how frequently he found each kind of fish eating each kind of food. In later papers, he exchanged the roles of rows and columns, listing predators as the column headings, prey as the row labels, and the frequencies of predators consuming prey items as entries of the tables. This is still the format of modern so-called "predation matrices".

Also in 1880, a 24-year-old Italian entomologist named Lorenzo Camerano, working at the Museum of Zoology of the University of Turin, used a relabeled form of Darwin's random directed tree to describe the feeding relations in a community of plants and animals (Figure 2).

Camerano did not refer explicitly to Darwin, so I do not know whether he borrowed Darwin's graphical idea or came upon it independently. The nodes of Camerano's tree represent groups of species, and an arc goes from one group to another if members of the second consume members of the first. For example, "predators" eat "phytophages", "parasites" eat "predators", and so on. To my knowledge, this is the first picture of a food web as a directed graph.

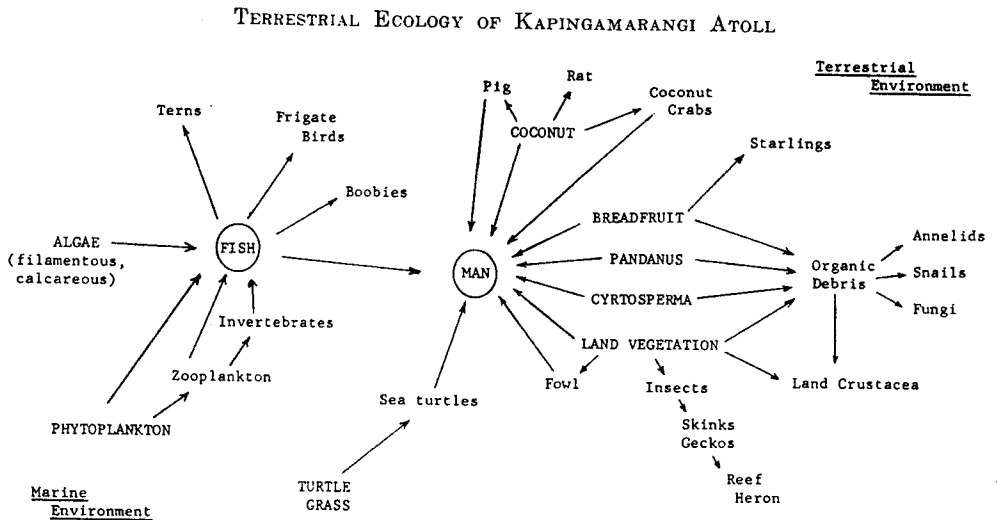


FIGURE 3. A contemporary food web: Niering's web of the Kapingamarangi Atoll. Source: Niering (1963, [29],p. 157), reprinted with permission.

These brief historical comments are not intended as exhaustive scholarship, and certainly not as establishing priorities, but rather as evidence that the mathematical models of food webs below are rooted in biological experience and thought.

In the century and more since Forbes described food webs in empirical detail and Camerano described generalized, abstract food webs, other ecologists have reported many hundreds, perhaps thousands, of food webs. A minority of reported webs is quantitative, like Forbes's tables. Most show only the feeding links among species (e.g., Figure 3). Henceforth, except in section 7, we shall be concerned exclusively with qualitative food webs that show who eats whom, but not how much or how often.

If a collection of species living in one place is like a city, a web is like a street map of the city; it shows where road traffic can and does go. A street map usually omits many important details, e.g., the flow of pedestrian and bicycle traffic, how much traffic flows along the available streets, what kind of vehicular traffic it is, the reasons for the traffic, the laws governing traffic flow, rush hours, and the origin of the vehicles. By analogy, a web often omits small flows of food or predation on minor species, the quantities of food or energy consumed, the chemical composition of food flows, the behavioral and physical constraints on predation, variations over time, whether periodic or random, in eating, and the population dynamics of the species involved. Thus a web gives at best very sketchy information about the functioning of a community. But just as a map provides a helpful framework for organizing more detailed information, a web helps biologists picture how a community works.

Until the second half of this century, each web was treated much as a species of butterfly was treated before Darwin — as unique, complex, and beautiful, perhaps sharing some similarities with other webs from similar habitats, but essentially individual and lawless. In recent decades, in addition to relishing the features that set each individual web apart, ecologists have given more attention to the patterns and order to be found in collections of a large number of food webs. Some recent accounts of the results obtained from this relatively new perspective are by Pimm (1982, [31]), Yodzis (1989, Chap. 8, [37]), Lawton (1989, [24]), Schoener (1989, [36]), Cohen, Briand, and Newman (1990, [9]), and Pimm, Lawton, and Cohen (1991, [33]).

This paper will review some theoretical ideas about random directed graphs that have been used to interpret data on ensembles of food webs. Little attention will be given to comparing the results of theoretical derivations with data, though such comparisons are the vital scientific rationale for theoretical work.

The mathematical custom, which will be followed here, of showing only theoretical derivations without the detailed empirical research that leads a scientist to require them is a traditional but unfortunate fraud. If data analysis is thought of as one level of research, and mathematical analysis is thought of as another level of research (let us take the level of mathematical analysis as

higher, to flatter mathematicians), then the trajectory of my own research in this area is something like the letter M, starting from data and moving up to theory, then back down to data, then again up to theory, and so on. During a long-term effort to collect, analyze, and summarize real food webs (starting in 1968, and still continuing), I noticed along the way some remarkably simple patterns; for example, the ratio of the number of predators to the number of prey, and the ratio of the total number of links to the total number of species, are roughly independent of the total number of species in different webs. These and several other data-analytic patterns required theoretical explanation. I constructed and, with mathematical colleagues, analyzed dozens of models to see which ones could explain the patterns quantitatively and qualitatively. The few surviving models made novel predictions not tested previously, so I returned to data analysis. When the data suggested modifications in the models, more mathematics was required. Most of my part of the story so far appears in gruesome detail in two books (Cohen 1978, [4], Cohen, Briand, Newman 1990, [9]), but the story is far from complete.

Before turning to some mathematical ideas, I would like to urge budding mathematicians and mathematical scientists to study at least one area of empirical science. I do not mean other people's mathematical theories about empirical science. Rather I urge contact with some domain of nature, contact that is as direct as possible. If that contact turns out to be fun, then a strenuous effort to formulate, analyze and interpret it by using mathematical language is almost guaranteed to lead to novel mathematical questions, and may well lead to new scientific insights.

## 2. CHARACTERISTICS OF DIRECTED GRAPHS USEFUL FOR MEASUREMENT AND THEORY

Some characteristics of directed graphs will now be defined. The purpose of these definitions is to specify properties of food webs, which will be treated as directed graphs, that can be measured in the real world and compared with predictions calculated from probabilistic food web models. This section will give definitions; the next, some models; and, the following sections, some calculated properties of the models leading to testable predictions.

A directed graph (henceforth abbreviated to "digraph") consists of a finite set of vertices together with a set of arcs, or directed edges (Robinson and Foulds 1980, [35]). Each vertex corresponds to a species, or sometimes to a stage in the life-cycle of a species (for example, the larvae of a certain insect or the eggs of a certain fish), or sometimes to a group of species whose diet and predators are not distinguished (for example, a size-class of phytoplankton or zooplankton species). For simplicity we shall just say that each vertex represents a species. Let the number of vertices (species) be  $S > 1$ . We label the vertices (and species)  $1, 2, \dots, S$  and call the set of vertices  $[S] = \{1, 2, \dots, S\}$ .



If  $i \in [S]$ ,  $j \in [S]$ , we interpret the arc  $(i, j)$  to mean that species  $i$  is eaten by species  $j$ . Graphically, as in Figure 3, we draw an arrow from  $i$  to  $j$ , showing the direction that the food flows. An arc of a food web is called a trophic link, or a link, in ecology. Denote by  $A$  the set of arcs or links or directed edges,  $A \subseteq [S] \times [S]$ . Then a digraph  $D_S$  on  $S$  vertices is the ordered pair  $([S], A)$ .

The adjacency matrix  $W = W(D_S)$  of  $D_S$  is an  $S \times S$  matrix such that  $w_{ij} = 1$  if and only if  $(i, j) \in A$  and  $w_{ij} = 0$  otherwise (Robinson and Foulds 1980, [35] p. 169). Ecologists call this matrix the predation matrix. Its column  $j$  has an element equal to 1 in each row that represents a species in the diet of species  $j$ , and 0 elsewhere. Its row  $i$  has an element equal to 1 in each column that represents a predator or consumer of species  $i$ , and 0 elsewhere. Sometimes we shall call  $W$  the web. The digraph and the adjacency matrix contain equivalent information. Sometimes it is more convenient to think in terms of a graphical representation of a web, other times in terms of a matrix representation.

A basal species is a species that eats no other species. A basal species is called a source in digraph theory, and is identified by a column of  $W$  in which all elements are equal to zero, or by a vertex of the digraph that has no incoming arcs. A top species is a species that is eaten by no other species. A top species is called a sink in digraph theory, and is identified by a row of  $W$  in which all elements are equal to zero, or by a vertex of the digraph that has no outgoing arcs. A species that is simultaneously basal and top is said to be isolated. A basal species that is not isolated is said to be a proper basal species. A proper top species is a top species that is not isolated. An intermediate species is one that eats and is eaten by other species.

A walk in a digraph is a sequence of alternating vertices and edges, starting and ending with vertices. For example, if  $(i, j) \in A$ ,  $(j, k) \in A$ , then  $i, (i, j), j, (j, k), k$  is a walk from vertex  $i$  to vertex  $k$ . Cannibalism is represented by a walk of the form  $i, (i, i), i$ , and is present if  $W$  has any nonzero diagonal element. The length of a walk is the number of edges in it. An  $n$ -walk is a walk of length  $n$ . A digraph is acyclic if and only if no vertex (or species) appears as a vertex more than once in any walk in the digraph. Thus an acyclic web has no cannibalism, no 2-loops such as  $i, (i, j), j, (j, i), i$ , where  $i \neq j$ , nor any longer loops.

A chain is a walk from a basal species to a top species. A chain in this sense is identical to a "maximal food chain" as defined by Cohen (1978, [4] p. 56). An  $n$ -chain is a chain of length  $n$ , i.e., a chain with  $n$  links. The height of a web is the length of the longest chain.

A real web can be described by its number of species, links, top species, intermediate species, basal species, number of chains of each length  $n = 1, 2, \dots, S - 1$ , and other characteristics. These observations can then be compared with predictions of stochastic models of webs, to which we now turn.

### 3. MODELS

In modeling an ensemble of food webs, the elements  $w_{ij}$  of  $W$ , which represent the presence or absence of feeding on species  $i$  by species  $j$ , are treated as random variables. This step in modeling food webs is very troubling for some ecologists. After all, much of ecology is devoted to documenting the exquisite adaptation of organisms to their diet and to their predators. The teeth of the tiger adapt it to carnivory. The teeth of the antelope adapt it to herbivory. Some plants provide refuge and food to certain species of ants that protect the plants against more deleterious consumers. What then is the sense of modeling the presence or absence of feeding between two species as a random variable? While the interactions between species may be very complex in many feeding relations, it may be possible to summarize the pattern resulting from these interactions in simple ways. By analogy, each driver on the congested highways of New York has his or her own special personal, cultural and ethnic history, his or her own reasons for urgency or patience, but the traffic jams of morning and evening rush hour are substantially predictable. Traffic engineers have long found simplified models of driver behavior useful approximations for understanding repeatable phenomena of traffic flow.

The probability that  $w_{ij} = 1$ , i.e., that species  $j$  eats species  $i$ , is written  $P\{w_{ij} = 1\}$ . The heterogeneous cascade model (Cohen 1990, [7]) assumes that  $P\{w_{ij} = 1\} = p_{ij}$ ,  $0 \leq p_{ij} \leq 1$ , where

$$p_{ij} = 0 \quad \text{if } i \geq j, \quad (3.1)$$

$$p_{ij} > 0 \quad \text{if } i < j, \quad \text{and} \quad (3.2)$$

the events  $\{w_{ij} = 1\}$  are mutually independent for all  $1 \leq i < j \leq S$ . (3.3)

The matrix  $P$  with elements  $p_{ij}$  is called the predation probability matrix. Assumptions (3.1) and (3.2) guarantee that a heterogeneous cascade digraph is acyclic with probability one, because food is permitted to flow only from a species with a lower label  $i$  to a species with a higher label  $j$ . There is thus a cascade, or hierarchy, of feeding relations. The cascade is heterogeneous because it is not yet assumed that all nonzero elements  $p_{ij}$  are equal to one another.

More general cascade models could weaken the assumption (3.3) of independence among trophic links or relax (3.2) to  $p_{ij} \geq 0$  for  $i < j$ . Such extensions will not be considered here.

It seems difficult to estimate separately all the elements of the predation probability matrix  $P$ . Indeed, even if one had a large number of webs with the same  $S$ , it might not be obvious how to label species, because, if links were relatively sparse, there could be multiple orderings of the species compatible

with the assumption of a cascade. Therefore, it is convenient to assume that certain elements of  $P$  are equal in order to estimate their common value.

Many special cases of the heterogeneous cascade model are of biological interest (Cohen 1990, [7]). For example, suppose that each consumer is equally likely to consume any of the prey species available to it, but, because of different behavioral capacities and morphology, different predators have different probabilities of preying on the species available to them. Then the model is said to be predator-dominant, and  $p_{ij} = b_j > 0$ , for  $j = 2, \dots, S$  and  $i < j$ . For a second example, suppose that the chances of predation are determined by the relative abundances or defensive abilities of the available prey species, so that different prey species have different probabilities of being preyed on but, as a first approximation, each prey is equally like to be preyed on by any of its possible consumers. Then the model is said to be prey-dominant, and  $p_{ij} = a_i > 0$ , for  $j = 2, \dots, S$  and  $i < j$ .

The simplest model assumes  $p_{ij} = p > 0$ , for  $1 \leq i < j \leq S$ . This homogeneous cascade model  $C(S, p)$  on  $S$  species (vertices) with link (arc) probability  $p$  is the only case that will be considered further here.

Because real webs have widely varying numbers of species, it has been of interest to describe empirically and explain theoretically how the properties of webs vary as the number  $S$  of species varies. It is found empirically that the average number of links (averaging over different webs with roughly the same number of species) increases with the number of species. The increase in the expected number of links is at least proportional to the number of species. Clearly, the expected number of links cannot increase faster than  $S^2$  asymptotically for large  $S$ , because  $W$  has only  $S^2$  elements. There is still some uncertainty among ecologists whether the expected number of links increases linearly with  $S$  or superlinearly, that is, as some power, greater than one, of  $S$ . The homogeneous cascade model can encompass both of these possibilities, as follows.

Fix the positive real number  $c$  and the nonpositive real number  $d$ ,  $0 \geq d \geq -1$ . For  $1 \leq i < j \leq S$ , let the probability of any arc be  $p = cS^d$ . Then the expected number of arcs is

$$E(L) = \binom{S}{2} p \sim \frac{c}{2} S^{2+d} \tag{3.4}$$

asymptotically. When  $d = -1$ ,  $E(L)$  is linear in  $S$ , hence this special case is called the linear cascade model; when  $d > -1$ ,  $E(L)$  rises faster than linearly in  $S$ , hence this special case is called the superlinear cascade model (Cohen 1990, [7]).

#### 4. CONNECTIONS

The homogeneous cascade model  $C(S, p)$  on  $S$  species (vertices) with link (arc) probability  $p_{ij} = p > 0$ , for  $1 \leq i < j \leq S$ , is identical to a model that was

proposed independently for parallel computation. The homogeneous cascade model is also closely connected with a basic model of random (undirected) graphs.

The homogeneous cascade model may be interpreted as a model of parallel computation by supposing that each vertex represents a task that must be processed, and that an arc goes from task  $i$  to task  $j$  if task  $i$  must be completed before task  $j$ ; thus the arcs represent precedence (Gelenbe, Nelson, Philips, Tantawi, 1986, [18]). If each task requires unit time, then the duration of the entire computation is just one plus the length (number of arcs) in the longest chain, or one plus the height of the digraph. If each task requires an amount of time given by some random variable, then the duration of the entire computation is the maximum of the sum of the random task durations, where the sum is taken along each chain and the maximum is taken over all chains.

A basic model  $G(S, p)$  in the theory of random (undirected) graphs (Erdős and Rényi 1960, [15]; see also e.g., Bollobás 1985, [1], p. 32; Palmer 1985, [30], p. 6) constructs a random graph on the vertex set  $[S]$  by placing an undirected edge  $\{i, j\}$  between vertices  $i$  and  $j$  with probability  $p$  and no edge between vertices  $i$  and  $j$  with probability  $1 - p$ , independently for all  $i$  and  $j$  with  $1 \leq i < j \leq S$ . To go from the random digraphs specified by  $C(S, p)$  to the random graphs specified by  $G(S, p)$ , simply drop the orientation of arcs, i.e., replace the arc  $(i, j)$  from  $i$  to  $j$  by the edge  $\{i, j\}$  between  $i$  and  $j$ . To go from  $G(S, p)$  to  $C(S, p)$ , do the reverse, i.e., orient each edge from the vertex with the lower number to the vertex with the higher number.

Suppose the edge probability  $p = p(S)$  in  $G(S, p)$  depends on  $S$  as the number  $S$  of vertices increases without bound. The marvellous accomplishment of Erdős and Rényi (1960) was to discover that the structure of  $G(S, p)$  changes quite suddenly as  $p$  changes smoothly from 0 to 1, and to invent methods for calculating these changes in structure (see Bollobás [1985], [1]). Without going into the methods, it is worthwhile to give a simple example of the powerful results. The complete graph  $K_n$  is defined as the graph on  $n$  vertices in which all possible  $n(n-1)/2$  edges are present. Erdős and Rényi (1960) proved that, in the limit as  $S \rightarrow \infty$ , if  $pS^{2/(n-1)} \rightarrow 0$ , then the probability that the random graph  $G(S, p)$  contains  $K_n$  approaches zero, while, if  $pS^{2/(n-1)} \rightarrow \infty$ , then the probability that the random graph  $G(S, p)$  contains  $K_n$  approaches one. Thus, loosely speaking, complete graphs on  $n = 4$  vertices are virtually absent from large random graphs if  $p(S) < S^{-2/3}$  for large  $S$ , but suddenly appear as soon as  $p(S) > S^{-2/3}$ .

Define the competition graph  $G(D) = ([S], E)$  of a digraph  $D = ([S], A)$  as an undirected simple graph on the vertex set  $[S]$  with edge set  $E$ , where  $\{i, j\} \in E$  if and only if there exists a vertex  $k$  in  $[S]$  such that both  $(k, i) \in A$ ,  $(k, j) \in A$ . Thus consumers  $i$  and  $j$  are joined by an edge in the competition graph if and only if there is at least one prey species that both consumers eat. [Ecologists also call  $G(D)$  the trophic niche overlap graph of  $D$ .] When the random digraph is specified by the homogeneous cascade model  $C(S, p)$ ,

the competition graph  $G(C(S, p))$  is a random graph model that differs from the classical model  $G(S, p)$ . Like  $G(S, p)$ ,  $G(C(S, p))$  displays abrupt changes in structure, but the changes occur for different values of  $p(S)$  (Cohen and Palka 1990, [13]). For example, in the limit as  $S \rightarrow \infty$ , if  $pS^{1+1/n} \rightarrow 0$ , then the probability that the random competition graph  $G(C(S, p))$  contains  $K_n$  approaches zero, while if  $pS^{1+1/n} \rightarrow \infty$ , then the probability that the random competition graph  $G(C(S, p))$  contains  $K_n$  approaches one. Thus, loosely speaking, complete graphs on  $n = 4$  vertices are virtually absent from large random competition graphs derived from the homogeneous cascade model if  $p(S) < S^{-5/4}$  for large  $S$ , but suddenly appear as soon as  $p(S) > S^{-5/4}$ . Some structural properties of the competition graph of the homogeneous cascade model explain facts about observed competition graphs (those derived from real food webs) that have previously lacked a quantitative explanation (Cohen and Palka 1990, [13]).

### 5. BASIC PROPERTIES OF THE HETEROGENEOUS CASCADE MODEL FOR FINITES

Define

$$q_{ij} = 1 - p_{ij} = P\{w_{ij} = 0\}, \quad \text{for } i, j = 1, \dots, S, \quad (5.1)$$

$$r_S = 1, \quad r_i = \prod_{j=i+1}^S q_{ij}, \quad \text{for } i = 1, \dots, S - 1, \quad (5.2)$$

$$c_1 = 1, \quad c_j = \prod_{i=1}^{j-1} q_{ij}, \quad \text{for } j = 2, \dots, S. \quad (5.3)$$

The probability that row  $i$  of  $W$  is entirely 0 is given by  $r_i$  and the probability that column  $j$  of  $W$  is entirely 0 is given by  $c_j$ . For  $i = 1, \dots, S$ , it follows that

- $P\{\text{species } i \text{ is not isolated}\} = 1 - r_i c_i,$
- $P\{\text{species } i \text{ is top}\} = r_i,$
- $P\{\text{species } i \text{ is basal}\} = c_i,$
- $P\{\text{species } i \text{ is intermediate}\} = (1 - r_i)(1 - c_i) = 1 - r_i - c_i + r_i c_i,$
- $P\{\text{species } i \text{ is proper top}\} = r_i(1 - c_i),$
- $P\{\text{species } i \text{ is proper basal}\} = c_i(1 - r_i).$

Consequently, if  $N, T, B, I, T_P,$  and  $B_P$  respectively denote the numbers of not isolated, top, basal, intermediate, proper top, and proper basal species (these numbers are random variables), then

$$E(N) = S - \sum_{i=1}^S r_i c_i, \quad (5.4)$$

$$E(T) = \sum_{i=1}^S r_i, \quad \text{var}(T) = \sum_{i=1}^S r_i(1 - r_i), \quad (5.5)$$

$$E(B) = \sum_{i=1}^S c_i, \quad \text{var}(B) = \sum_{i=1}^S c_i(1 - c_i), \quad (5.6)$$

$$E(I) = 2S - E(T) - E(B) - E(N), \quad (5.7)$$

$$E(T_P) = E(T) + E(N) - S, \quad (5.8)$$

$$E(B_P) = E(B) + E(N) - S. \quad (5.9)$$

A weak component is a maximal set of species linked to each other, directly or indirectly, by arcs regardless of their orientation. A web has more than one weak component with very low probability in the models to be developed and in the data, so this possibility receives no special treatment here.

Let the number of links, i.e., the number of elements of  $W$  equal to 1, be

$$L = |A| = \sum_{1 \leq i < j \leq S} w_{ij}. \quad (5.10)$$

$L_{BT}$  the number of basal-top links (links from some basal species to some top species),  $L_{BI}$  the number of basal-intermediate links,  $L_{IT}$  the number of intermediate-top links, and  $L_{II}$  the number of intermediate-intermediate links. For  $i < j$ , there is a basal-top link from  $i$  to  $j$  if and only if there is a link from  $i$  to  $j$  (with probability  $p_{ij}$ ) and  $j$  is top (with probability  $r_j$ ) and  $i$  is basal (with probability  $c_i$ ). By such arguments, it is obvious that

$$E(L) = \sum_{j=2}^S \sum_{i=1}^{j-1} p_{ij}, \quad (5.11)$$

$$\text{var}(L) = \sum_{j=2}^S \sum_{i=1}^{j-1} p_{ij}q_{ij}, \quad (5.12)$$

$$E(L_{BT}) = \sum_{j=2}^S \sum_{i=1}^{j-1} p_{ij}c_i r_j, \quad (5.13)$$

$$E(L_{BI}) = \sum_{j=2}^S \sum_{i=1}^{j-1} p_{ij}c_i(1 - r_j), \quad (5.14)$$

$$E(L_{IT}) = \sum_{j=2}^S \sum_{i=1}^{j-1} p_{ij}(1 - c_i)r_j, \quad (5.15)$$

$$E(L_{II}) = \sum_{j=2}^S \sum_{i=1}^{j-1} p_{ij}(1 - c_i)(1 - r_j). \quad (5.16)$$

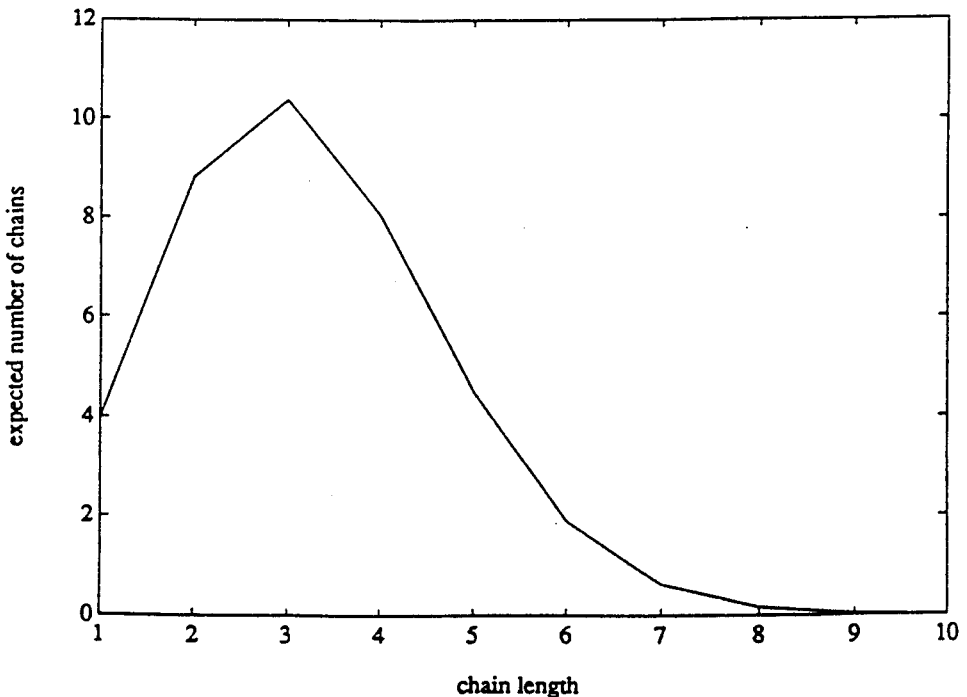
There is a chain of length  $n$ , where  $n$  counts the number of links, that involves the  $n + 1$  species  $i_0, i_1, \dots, i_n$  if and only if the following conditions all hold:  $1 \leq i_0 < i_1 < \dots < i_n \leq S$ ; and there is a link from  $i_h$  to  $i_{h+1}$  for all  $h = 0, \dots, n - 1$ ; and  $i_0$  is basal; and  $i_n$  is top. Let  $C_n$  denote the number of chains of length  $n$ . Then

$$E(C_n) = \sum_{i_0=1}^{S-n} \sum_{i_1=i_0+1}^{S-n+1} \dots \sum_{i_h=i_{h-1}+1}^{S-n+h} \dots \sum_{i_n=i_{n-1}+1}^S \prod_{h=0}^{n-1} p_{i_h i_{h+1}} c_{i_0} r_{i_n} \quad (5.17)$$

or

$$E(C_n) = \sum_{i_0=1}^{S-n} c_{i_0} \sum_{i_1=i_0+1}^{S-n+1} p_{i_0 i_1} \dots \sum_{i_h=i_{h-1}+1}^{S-n+h} p_{i_{h-1} i_h} \dots \sum_{i_n=i_{n-1}+1}^S p_{i_{n-1} i_n} r_{i_n} \quad (5.18)$$

For different values of  $n$ , different numbers of summations are required to evaluate  $E(C_n)$ . For numerical and symbolic computation, a recursive function is convenient.

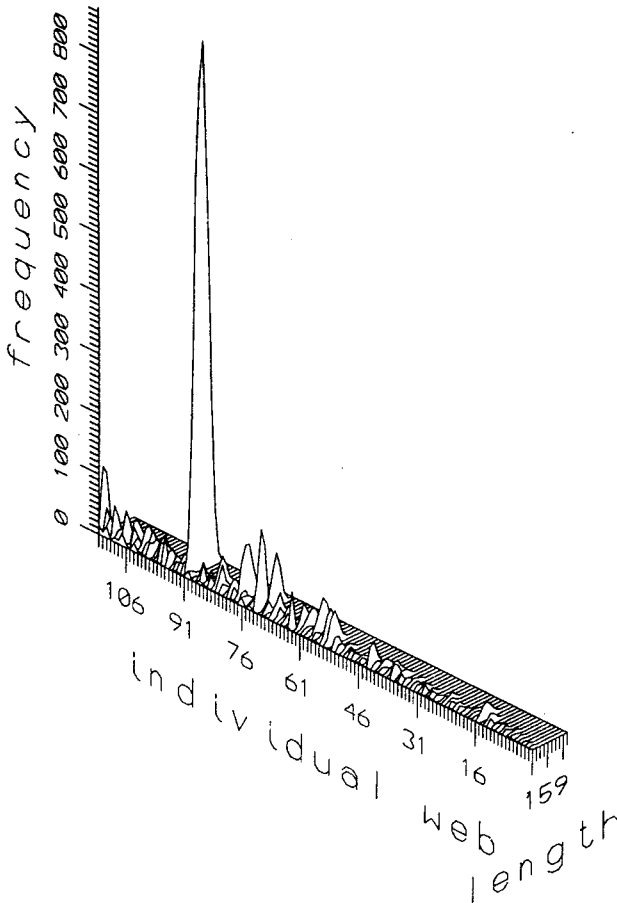


**FIGURE 4.** Expected number of chains of each length according to the linear cascade model of a web with  $S = 17$  species and link probability  $4/17$ , based on  $c = 4$ . The expected total number of chains is 38.458. The expected numbers of chains of length 11 through 16 are not shown because their sum is less than 0.001.

**Exercise.** Simplify the above formulas (5.1) to (5.18) when  $p_{ij} = p > 0$ , for  $1 \leq i < j \leq S$ . In particular, prove (Cohen, Briand and Newman 1990, [9], p. 119) that (5.18) can be transformed to

$$E(C_n) = p^n q^{S-1} \sum_{k=n}^{S-1} (S-k) \binom{k-1}{n-1} q^{-k}, \quad n = 1, 2, \dots, S-1. \quad (5.19)$$

Figure 4 plots (5.19), the expected number of chains of each length according to the linear cascade model, for a web with  $S = 17$  species and link probability  $4/17$ , based on  $c = 4.0$ . The choice of this value of  $c$  is explained at (6.8).



**FIGURE 5.** Observed numbers of chains with one to nine links in 113 community food webs. No chains longer than nine links were observed. Each slice parallel to the axis labeled “length” is the observed frequency distribution for one web. The height of a slice at each value of length is the observed number of chains of that length. The webs are ranked from the smallest (web 1) to the largest (web 113) number of species. The mean number of species was close to 17. Source: Cohen (1990, [7], p. 59), reprinted with permission.



For qualitative comparison, Figure 5 shows the observed number of chains of each length in 113 real webs. The raw data are given in Cohen, Briand, Newman (1990, Chap. 4, [9]). A detailed quantitative comparison between the observations and the predictions of the linear cascade model (Cohen, Briand, Newman 1990, [9], Chap. 3.4) shows that the predictions are not bad for all but 16 or 17 of the webs.

## 6. LIMIT THEORY OF THE LINEAR CASCADE MODEL FOR LARGE $S$

One reason for developing the linear cascade model was to explain a simple empirical pattern — as the number of species in a web increases, the fraction of all species that are top species seems neither to increase nor to decrease systematically, but fluctuates around a fixed proportion. (Caveat! This finding was based on webs with at most 48 species. The finding is under challenge by the very few webs reported so far with more than 48 species. In a science as young as this part of ecology, models and theories often aim at a shifting target.) How can the linear cascade model account for this (tentative) finding?

Giving away part of the answer to the previous exercise, the expected number of top species and the expected number of basal species in the linear cascade model are

$$E(T) = E(B) = [1 - q^S]/p, \tag{6.1}$$

where  $q = 1 - p$ . Hence, since  $p = c/S$ , asymptotically

$$\lim_{S \rightarrow \infty} E(T)/S = \lim_{S \rightarrow \infty} E(B)/S = (1/c)(1 - e^{-c}), \tag{6.2}$$

**Exercise.** Prove that

$$\text{var}(T) = \text{var}(B) = (1 - q^S)/p - (1 - q^{2S})/(1 - q^2) \tag{6.3}$$

and hence

$$\lim_{S \rightarrow \infty} \text{var}(T/S) = \lim_{S \rightarrow \infty} \text{var}(B/S) = 0. \tag{6.4}$$

The right side of (6.2) is a constant, independent of  $S$ . The model predicts that the proportion of species that are top species will asymptotically be independent of  $S$  and will [according to (6.4)] have vanishing variability, relative to  $S$ .

For quantitative comparison with actual data, it is necessary to examine the number of proper top species as a fraction of nonisolated species, because isolated species have been removed from the data. The right quantity to calculate would be  $E(T_P/N)$ , but that appears to be too difficult!

**Exercise.** Show that

$$E(T_P) = E(B_P) = S[(1 - q^S)/c - q^{S-1}], \quad E(N) = S(1 - q^{S-1}), \tag{6.5}$$

and asymptotically

$$\lim_{S \rightarrow \infty} \frac{E(T_P)}{E(N)} = \lim_{S \rightarrow \infty} \frac{E(B_P)}{E(N)} = \frac{(1 - e^{-c})/c - e^{-c}}{1 - e^{-c}}. \quad (6.6)$$

Table 1 compares the linear cascade model's predicted asymptotic (as  $S \rightarrow \infty$ ) fractions of nonisolated species that are proper basal, proper top, and intermediate with the observed fractions. It also compares the predicted asymptotic fractions of links of each possible kind with the observed fractions. To calculate these numerical predictions requires not only the formulas above but also a specific numerical value for the parameter  $c$ . The asymptotic formula (3.4) with  $d = -1$  gives a rough-and-ready estimate

$$\hat{c} = 2E(L)/S. \quad (6.7)$$

The predictions in Table 1 take the ratio of links to species as exactly 2, though the actual ratio is 1.99, hence  $\hat{c} = 4.0$ .

113 Community food webs	Observed number	Observed fraction	Linear cascade model predicted fraction
proper basal species	353	0.186	0.231
intermediate species	1038	0.546	0.537
proper top species	511	0.269	0.231
all species	1902	1.000	1.000
basal-intermediate links	1029	0.272	0.264
basal-top links	230	0.061	0.114
intermediate-intermediate links	1194	0.316	0.359
intermediate-top links	1327	0.351	0.264
all links	3780	1.000	1.000

**TABLE 1.** Observed fractions of nonisolated species that are proper top, intermediate and proper basal, and observed fractions of links of each kind, in 113 community food webs. Source: Cohen (1990, [7], p. 56).

The fact that the right side of (6.7) is independent of  $S$  illustrates a property of the linear cascade model, and possibly a property of the real data, called scale-invariance — the form of the web (as measured by the proportion of its nonisolated species that are proper top species) is asymptotically independent of its scale, or size, or number of species. In the superlinear cascade model, by contrast, the proportion of nonisolated species that are proper top species slowly but markedly declines as the number of species increases (Cohen 1990, [7], p. 70). Whether this is a virtue or a defect of the superlinear cascade model will depend on large food webs now being collected.

Chain lengths may also be analyzed in detail, but the analysis is substantially more difficult (Newman and Cohen 1986, [28]; reprinted in Cohen, Briand, Newman 1990, [9], pp. 149-173). Let

$$C = \sum_{n=1}^{S-1} C_n \tag{6.8}$$

be the total number of chains of all possible lengths. It may be shown that

$$\lim_{S \rightarrow \infty} E(C_n/C) = \lim_{S \rightarrow \infty} E(C_n)/E(C), \quad n = 1, 2, \dots, S - 1, \tag{6.9}$$

and both limits may be calculated from the coefficients of a power-series expansion of an explicitly stated generating function. Because, in the linear cascade model,

$$\lim_{S \rightarrow \infty} \text{var}(C_n/S) = 0, \tag{6.10}$$

the random variable  $C_n/C$  converges in probability as  $S \rightarrow \infty$  to the limit in (6.9). For large  $c$ , say  $c = 10$ , the asymptotic relative frequency of chains of each length is normally distributed with mean equal to  $c$  and variance equal to  $c - 1/2$  (Cohen, Briand and Newman 1990, [9], p. 155). For realistic values of  $c$  around 4.0, the asymptotic relative frequency of chains of each length is skewed to the right, with modal value close to  $c$ . Apparently  $S = 17$  is close to infinity, because the asymptotic (large  $S$ ) distribution of chain lengths closely resembles that shown for  $S = 17$  in Figure 4.

The linear cascade model explains qualitatively an observation made independently by many biologists, that the longest chain length is short relative to the number of species in a web, even for webs with very large numbers of species. Let  $M$  be the maximum chain length, or height, in a realization of the linear cascade model;  $M$  is a random variable. Then for extremely large  $S$ ,  $M$  grows like  $\ln(S)/\ln(\ln(S))$ , even more slowly than  $\ln(S)$ . The exact results (Newman and Cohen 1986, [28]) are a bit more surprising than this.

Define  $m^*$  to be the smallest positive integer  $m$  such that  $c^{m+1}S/(m+2)! \leq (m+2)^{-1/2}$ . Then, for large enough  $S$ ,  $m^*$  is a nondecreasing sequence such that

$$\lim_{S \rightarrow \infty} \frac{m^*}{\ln S / \ln \ln S} = 1, \tag{6.11}$$

and

$$\lim_{S \rightarrow \infty} P[M = m^* \text{ or } M = m^* - 1] = 1. \tag{6.12}$$

That is, the distribution of the height is concentrated on just two numbers (which depend on  $S$ ),  $m^*$  and  $m^* - 1$ . For very, very large numbers  $S$  of species,  $m^*$  grows at a rate that is essentially independent of  $c$  or  $p = c/S$  (provided  $c > 0$ ) and depends only on  $S$ . For extremely large  $S$ ,  $m^*$  is approximately  $\ln S / \ln \ln S$  in the sense that their ratio approaches 1. This does not imply

that the difference between  $M$  and  $\ln(S)/\ln(\ln(S))$  is arbitrarily small with high probability, because  $M$  and  $\ln(S)/\ln(\ln(S))$  could differ by an amount that goes to infinity more slowly than  $\ln(S)/\ln(\ln(S))$ . Higher order correction terms to the asymptotic expansion (6.11) are given by Newman and Cohen (1986, [28]).

The estimated rate of convergence of  $P[M = m^* \text{ or } M = m^* - 1]$  to 1 is very slow, namely,

$$1 - P[M = m^* \text{ or } M = m^* - 1] = O(m^{*-1/2}). \quad (6.13)$$

A qualitatively similar phenomenon has been observed elsewhere in the theory of random graphs. Bollobás and Erdős (1976, [2]) proved that the size of the maximal complete subgraph (clique) in a random graph takes one of at most two values (that depend on the size of the random graph) with a probability that approaches 1 as the random graph gets large, when the edge probability is held fixed, independent of the number of vertices. However, the asymptotic behavior of the at most two possible values for the size of the largest clique does depend on the fixed probability that there is an edge between any two given vertices, according to Bollobás and Erdős (1976, [2]).

According to the superlinear cascade model (Newman 1991, [27]; Cohen and Newman 1991, [12], when  $S$  becomes very large, the ratio between  $M$  and  $epS$  is arbitrarily close to one, where  $p = cS^d$ , with a probability that approaches one. So  $M$  grows like  $ecS^{1+d}$ , much faster than in the linear cascade model.

The average chain length  $\mu(S)$  in the homogeneous cascade model web with  $S$  species and probability  $p$  of a link is the same for both the linear and the superlinear cascade models (Newman and Cohen 1986, [28], p. 358):

$$\mu(S) = Sp \frac{(1+p)^{S-1} - (1-p)^{S-1}}{(1+p)^S - (1-p)^S - 2Sp(1-p)^{S-1}}. \quad (6.14)$$

In the linear cascade model, the average chain length approaches a finite limit for large  $S$ :

$$\lim_{S \rightarrow \infty} \mu(S) = \frac{c(1 - e^{-2c})}{1 - e^{-2c}(1 + 2c)}, \quad (6.15)$$

which is approximately 4.01 if  $c = 4.0$ . This limit is approached fairly rapidly. For example, the average chain length is predicted to exceed 3.5 as soon as  $S \geq 30$ ; if  $S = 100$ ,  $M(100) = 3.86$ .

Comparing (6.11) and (6.15) shows that the ratio of the height to the average chain length in the linear cascade model increases without limit as the number of species gets large. Here is an intuitive explanation for this phenomenon. First, as mentioned above, there is a fixed limiting distribution of chain lengths as  $S$  gets large. The limiting average chain length given by (6.15) is the mean of this fixed distribution. Second, as  $S$  gets large, more and more chains are sampled from this fixed distribution of chain lengths, so that

very long chains, with a fixed low probability, are more likely to be observed and the height increases without limit. The particular form of the increase given by (6.11) is expected with a sample size proportional to  $S$  taken from a fixed distribution whose tail probabilities behave approximately like those of a Poisson distribution. Such a tail behavior can be proved by using the exact formulas for the generating function of the distribution. The expected number of chains of all lengths, which represents the sample size, is proportional to  $S$  (Newman and Cohen 1986, [28]).

In the superlinear cascade model, as  $S$  increases without limit, the ratio of the average chain length to  $Sp$  approaches 1 (Cohen and Newman 1991, [12]). The ratio of the height to the average chain length in the superlinear cascade model approaches the limit  $e$  as the number of species gets large. Thus there is a qualitative difference in the ratio of the height to the average chain length between the cascade and the superlinear homogeneous models.

## 7. DYNAMICS OF FOOD WEBS

When food webs were still young, few people had hopes of comprehending their structural and dynamic complexity in simple models. Instead, early explorers focused on a single feeding link, and on the dynamics of the two populations interacting through that link. Kingsland (1985, [22]) tells the history in an interesting way. Alfred J. Lotka, trained as a physical chemist, borrowed the style of the kinetic equations of chemistry to model a single predatory population eating a single prey population (his great 1924 book [25] summarizes his work). If  $N(t)$  is the number or biomass (quantity of living material) of the resource population and  $P(t)$  is the number or biomass of the predatory population, the Lotka-Volterra equations for predator-prey interactions are (May 1981, Chap. 5, [26]):

$$\frac{dN}{dt} = aN(t) - \alpha N(t)P(t), \quad (7.1)$$

$$\frac{dP}{dt} = -bP(t) + \beta N(t)P(t), \quad (7.2)$$

where the coefficients  $a, b, \alpha$  and  $\beta$  are all assumed positive. Among other simplifications, these equations ignore genetic variations within the predator and prey populations, differences in ages within a population with respect to probabilities or rates of predation, reproduction and mortality, the spatial distribution of predators and prey, the possible influences of environmental changes (weather, diurnal cycles, volcanoes, tides) on both species and their interactions, all other species that might interact with either or both species, and any physical and chemical constraints (such as limited nutrients or limited space) that affect either or both species.

Dynamic equations of the same embarrassing simplicity as (7.1) and (7.2) are often presented as more realistic than the structural models described in

the previous sections, and are often accepted as such by both ecologists and mathematicians. The reason for this acceptance is simply familiarity, not evidence. In general, the empirical support for dynamic equations such as the Lotka-Volterra predator-prey equations, though non-zero (Gause 1935, [17]; see Krebs 1972, [23], pp. 250-272 for a review), is slimmer than that for the cascade models.

Realistic or not, the Lotka-Volterra predator-prey equations (hereafter abbreviated to LV's) can obviously be generalized to  $S$  interacting species. If  $u_i(t)$  is the abundance or biomass of the  $i$ th species at time  $t$ , then it is assumed that there exist a real  $S \times S$  matrix  $B = (b_{ij})$  and a real  $S \times 1$  vector  $e = (e_i)$ , with both  $B$  and  $e$  independent of time  $t$ , such that, for all  $t \geq 0$ ,

$$\frac{du_i}{dt} = u_i \left( e_i + \sum_{j=1}^S b_{ij} u_j \right), \quad u_i(0) > 0, \quad i = 1, \dots, S. \quad (7.3)$$

The coefficient  $b_{ij}$  measures the effect of species  $j$  on the growth rate of species  $i$ . It is assumed that (7.3) has a stationary solution in the first orthant, i.e., that there exists a constant  $S \times 1$  vector  $q = (q_i)$  such that

$$0 = e_i + \sum_{j=1}^S b_{ij} q_j, \quad q_i > 0, \quad i = 1, \dots, S. \quad (7.4)$$

LVs make several further assumptions that could be shortcomings empirically. The LVs assume that there are no mutualistic interactions between species, that all interactions among species are strictly pairwise, and that the pairwise interactions follow a simple mass-action law specified by the product of abundances or biomasses.

Nevertheless, LVs have been extensively used for food web theory (Pimm 1982, [31]; Cohen, Luczak, Newman, Zhou 1990, [10]; Pimm, Lawton, Cohen 1991, [33]). A major problem is that the LVs have  $S^2 + S$  parameters (the  $b_{ij}$  and the  $e_i$ ), not counting the initial conditions. Yet, in the real world, there are only roughly twice as many links as species. It seems unlikely that all the possible interactions really matter dynamically.

Since the values of the interaction coefficients in the matrix  $B$  in (7.3) can never be known exactly, but the signs of the coefficients can be estimated more reliably, it is natural to consider the behavior of (7.3) when the value of any interaction coefficient  $b_{ij}$  is changed to some other number with the same sign, while an interaction coefficient that is zero is left at zero. Under certain conditions, some of which will be described in a moment, the LVs bequeath their stability to the whole family of equations obtained by replacing the interaction coefficients by others with the same signs. Under such conditions, (7.3) is said to be qualitatively stable. Qualitative stability is a natural concept for linking dynamic models with structural models based on random graphs, because structural models deal with the presence or absence of certain interactions,

and not with the magnitudes of those interactions. What is surprising is that it is possible to say things about qualitative stability that connect usefully with structural food web models.

A recently proposed hybrid of the LVs and the linear cascade model, called the Lotka-Volterra cascade model (LVCM), assumes the population dynamics of the LVs while letting a refinement of the cascade model determine the interactions between species (Cohen, Luczak, Newman, and Zhou 1990, [10]). Mathematical analysis of the LVCM combines the theory of large random digraphs with the qualitative theory of nonlinear differential equations to characterize the global asymptotic stability of ecological communities in the double limit of large time and large numbers of species.

The LVCM links the cascade model to the Lotka-Volterra model by discriminating among the possible population dynamical effects caused by each feeding link. In principle, when species  $j$  eats species  $i$ , there could be a positive, a negative, or no effect on the population growth rate of species  $j$ , and a positive, a negative, or no effect on the population growth rate of species  $i$ , for a total of nine possible pairs of effects. The LVCM ignores the five possible pairs of effects where  $j$  eating  $i$  hurts the population growth rate of species  $j$  or helps the population growth rate of species  $i$ . Thus if species  $j$  eats species  $i$ , the LVCM supposes that one of four biological effects occurs:

- (i) the feeding has no effect on the growth of species  $j$  but hurts the growth of species  $i$ ; or
- (ii) the feeding helps the growth of species  $j$  but has no effect on the growth of species  $i$ ; or
- (iii) the feeding helps species  $j$  and hurts species  $i$ ; or
- (iv) the feeding has no effect on the growth of either  $j$  or  $i$ .

Corresponding to each biological effect, assume that:

- (i)  $b_{ji} = 0$  and  $b_{ij} < 0$ ; or
- (ii)  $b_{ji} > 0$  and  $b_{ij} = 0$ ; or
- (iii)  $b_{ji} > 0$  and  $b_{ij} < 0$ ; or
- (iv)  $b_{ji} = b_{ij} = 0$ .

Because of (iv), the event:

(iv') there is no dynamic interaction of any kind between species  $i$  and  $j$  represents two biologically distinct situations: predation without dynamic effects [described by (iv)] (e.g., the old lady who accidentally swallowed a fly), and the absence of predation [i.e., the absence of an edge  $(i, j)$  in  $W$ ].

The LVCM assumes that events (i), (ii), (iii), and (iv') occur independently for each pair  $i, j = 1, \dots, S$  such that  $i < j$ , with probabilities, respectively,  $r/S, s/S, t/S$  and  $1 - (r + s + t)/S$ , where  $r, s, t$  are nonnegative constants that do not depend on  $S$ . [Predation without dynamic effects (iv) occurs with probability  $c/S - (r + s + t)/S$ .]

More formally, for  $S = 1, 2, \dots$ , let  $N_S$  be the system (7.3) with randomly chosen coefficients where, with probability 1,  $b_{ii} < 0$  for  $i = 1, \dots, S$  and the

pairs  $\{b_{ji}, b_{ij}\}$  for each  $i, j = 1, \dots, S$  with  $i < j$  are chosen independently with probabilities

$$P \{b_{ji} = 0 \text{ and } b_{ij} < 0\} = r/S, \tag{i}$$

$$P \{b_{ji} > 0 \text{ and } b_{ij} = 0\} = s/S, \tag{ii}$$

$$P \{b_{ji} > 0 \text{ and } b_{ij} < 0\} = t/S, \tag{iii}$$

$$P \{b_{ji} = 0 \text{ and } b_{ij} = 0\} = 1 - (r + s + t)/S. \tag{iv'}$$

and the vector  $e_s = (e_i)_{i=1}^s$  is chosen (depending on  $b_s = (b_{ij})_{i,j=1}^s$ ) so that for some vector  $q_s > 0$  (also depending on  $b_s$ ),  $0 = e_s + b_s q_s$ . The sequence of systems  $\{N_s\}_{s=1}^\infty$  defines the LVCM.

To state the interesting facts about the LVCM requires some further definitions. For any real finite scalar  $s$ , define  $\text{sign}(s) = +1$  if  $s > 0$ ,  $\text{sign}(s) = -1$  if  $s < 0$  and  $\text{sign}(s) = 0$  if  $s = 0$ . Define  $\bar{B} \sim B$  (read:  $\bar{B}$  is sign-equivalent to  $B$ ) if and only if, for all  $i, j$ ,  $\text{sign}(b_{ij}) = \text{sign}(\bar{b}_{ij})$ . Let  $\bar{N}$  refer to the family of equations (7.3) when (i)  $B$  is replaced by any  $S \times S$  matrix  $\bar{B} \sim B$ , and (ii)  $e$  is replaced by any  $S \times 1$  vector  $\bar{e}$  such that  $0 = \bar{e} + \bar{B}\bar{q}$  has a positive solution  $\bar{q} > 0$ . Positive initial conditions  $u(0) > 0$  are assumed throughout. A result is considered "qualitative" if it refers to all of  $\bar{N}$ .

Define LVs (7.3) to be qualitatively globally asymptotically stable (q.g.a.s.) if and only if every solution of every system in the family  $\bar{N}$  is bounded and has a limit as  $t \rightarrow \infty$  and that limit is independent of the initial conditions. This property may formalize what some ecologists mean by ecological stability (Pimm 1984, [32]) for some ecological communities, but no claim is made that it formalizes ecologists' concept of ecological stability for all communities.

The LVCM has what physicists call a phase transition. As the parameters of the LVCM cross a certain critical surface, the probability of being qualitatively globally asymptotically stable changes from positive to zero. Cohen, Luczak, Newman, and Zhou (1990, [10]) show where the critical surface is and how the LVCM behaves on either side of the surface as well as on it.

Let  $x: [1, \infty) \rightarrow (0, 1]$  be the smallest root of  $x(z)e^{-x(z)} = ze^{-z}$ .

(i) If

$$r + t < 1 \text{ and } s + t < 1, \text{ or} \tag{7.5}$$

$$r + t \geq 1 \text{ but } s + t < x(r + t), \text{ or} \tag{7.6}$$

$$s + t \geq 1 \text{ but } r + t < x(s + t) \tag{7.7}$$

then

$$\lim_{S \rightarrow \infty} P \{N_S \text{ is q.g.a.s.}\} = \rho > 0, \text{ where} \tag{7.8}$$

$$\rho = \begin{cases} e^{(r+t)(s+t)/2} \frac{(r+t)e^{s+t} - (s+t)e^{r+t}}{r-s}, & r \neq s, \\ e^{(r+t)^2/2} e^{r+t} (1-r-t), & r = s. \end{cases} \tag{7.9}$$



When  $(r + t)(s + t) > 0$ , then  $\rho < 1$ .

(ii) If

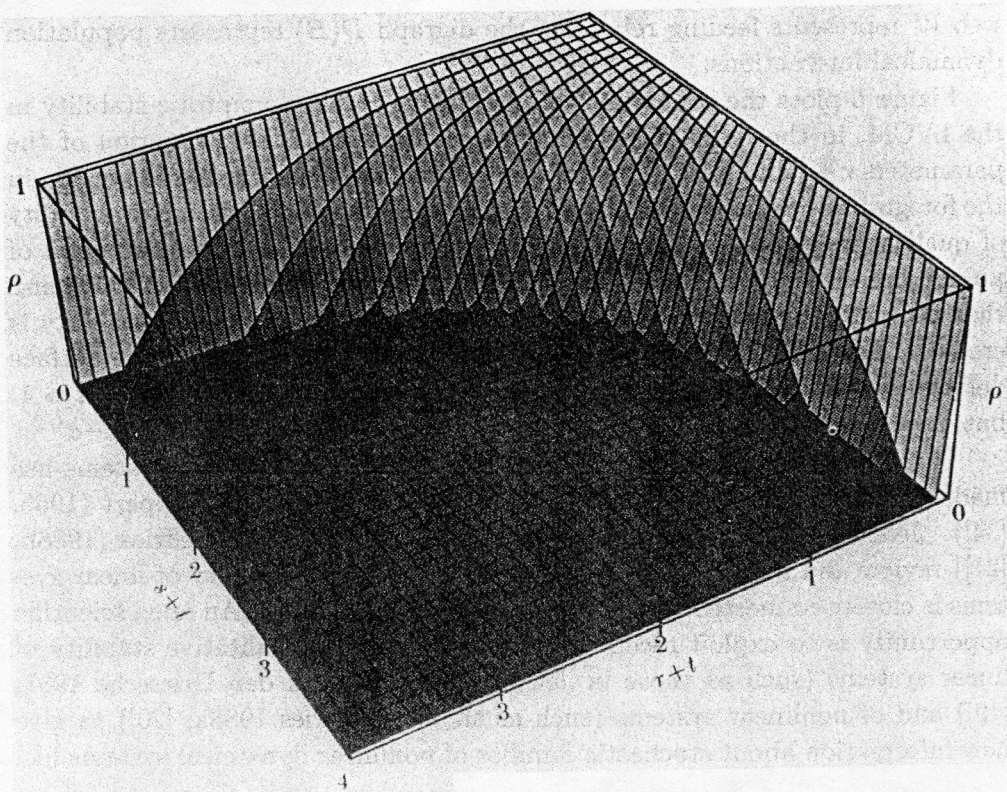
$$r + t \geq 1 \text{ and } s + t \geq x(r + t), \text{ or} \tag{7.10}$$

$$s + t \geq 1 \text{ and } r + t \geq x(s + t) \tag{7.11}$$

then

$$\lim_{S \rightarrow \infty} P \{N_S \text{ is q.g.a.s.}\} = 0. \tag{7.12}$$

The proof of this theorem is difficult and will not be attempted here. The proof depends crucially on facts, first developed by Quirk and Ruppert (1965, [34]), about the qualitative stability of linear systems.



**FIGURE 6.** Perspective view of the probability of qualitative global asymptotic stability in the Lotka-Volterra cascade model (LVCM), in the limit as the number of species  $S$  approaches infinity, as a function of the parameters  $r + t$  and  $s + t$ , which are defined in the text. The limiting probability is 0 in the flat region in the foreground. Small values of  $r + t$  and  $s + t$  assure a high limiting probability of qualitative global asymptotic stability. Source: Cohen, Luczak, Newman, Zhou (1990, [10], p. 616), reprinted with permission.

Informally speaking, a critical surface divides the three-dimensional parameter space  $\{(r, s, t) | r \geq 0, s \geq 0, t \geq 0\}$  of the LVCM into two regions. In region (i), where (7.5) or (7.6) or (7.7) holds, as the number of species becomes large, the probability that the LVCM is qualitatively globally asymptotically

stable (q.g.a.s.) approaches a positive limit. This limiting probability is given explicitly by (7.9). In region (ii), where (7.10) or (7.11) holds, as the number of species becomes large, the probability that the LVCM is q.g.a.s. approaches zero. The critical surface for the phase transition in the stability of the LVCM is exactly the same as the critical surface for a phase transition when a giant strongly connected component suddenly appears in a certain digraph  $D(B)$ , called the interaction digraph, associated with the matrix  $B$  of interaction coefficients.  $D(B)$  has  $S$  vertices, an arrow  $(j, i)$  from vertex  $j$  to vertex  $i$  if  $b_{ij} \neq 0$  and  $b_{ji} = 0$ , and a bidirectional arrow  $\{i, j\}$  between  $i$  and  $j$  (equivalent to a pair of unidirectional arrows  $(j, i)$  and  $(i, j)$ ) if  $b_{ij}b_{ji} < 0$ . (The case  $b_{ij}b_{ji} > 0$  with  $i \neq j$  was excluded in constructing the LVCM.) Whereas the web  $W$  represents feeding relations, the digraph  $D(B)$  represents population dynamical interactions.

Figure 6 plots the probability of qualitative global asymptotic stability in the LVCM, in the limit of large numbers of species  $S$ , as a function of the parameters  $r+t$  and  $s+t$ . The limiting probability is zero in the flat region in the foreground. Small values of  $r+t$  and  $s+t$  assure a high limiting probability of qualitative global asymptotic stability. The transition from the region of positive probability to the region of zero probability is abrupt in the sense that the derivative of  $\rho$  changes discontinuously as the frontier of stability is crossed. For example, along the diagonal cross-section through the surface defined by  $r+t = s+t$ , the derivative of  $\rho$  is evidently zero when  $r+t > 1$ , but, as  $r+t$  approaches 1 from below, the derivative of  $\rho$  approaches  $-e^{3/2}$ .

The theory of the qualitative stability of linear and nonlinear systems has made great progress since the pioneering paper of Quirk and Ruppert (1965, [34]). Jeffries, Klee, and van den Driessche (1987, [19]) and Jeffries (1988b, [21]) review and extend this progress. The qualitative stability of linear systems is closely connected to the qualitative stability of LVs. An open scientific opportunity is to exploit recent discoveries about the qualitative stability of linear systems (such as those in Jeffries, Klee and van den Driessche 1987, [19]) and of nonlinear systems (such as those in Jeffries 1988a, [20]) to give new information about stochastic families of nonlinear dynamical systems like the LVCM.

Information about qualitative global asymptotic stability could assist the design of managed ecological systems such as closed ecological life support systems for space travel, nature reserves, and complex chemostats, microcosms, and mesocosms. Qualitatively globally asymptotically stable systems may be desirable for practical and aesthetic reasons, because perturbations that do not change the signs of the interactions between species will not alter the existence of a long-run globally stable equilibrium. If future empirical studies confirm its usefulness, the LVCM would suggest designs that maximize (subject to some constraints) the probability of being q.g.a.s., asymptotically for large numbers of species. They would be those designs that satisfy the hypotheses (7.5) to (7.7) with large values of  $\rho$ . The LVCM would suggest avoiding ecological

designs that satisfy (7.10) or (7.11), which have little chance of being q.g.a.s., asymptotically for large numbers of species.

Apart from its potential uses in ecological design, the LVCM warns of the possibility that gradual, smooth changes in the probabilities  $r$ ,  $s$ , and  $t$  of various kinds of dynamic interactions related to feeding can have abrupt effects on the long-run probability of qualitative global stability or instability of ecological communities. Such phase transitions in stability are driven by phase transitions in the structure of underlying random graphs or digraphs, and are found in many large, random structures (see Cohen [1988], [5] for an expository account).

### ACKNOWLEDGMENTS

J.E.C. acknowledges helpful comments from Charles M. Newman, the support of U.S. National Science Foundation grant BSR87-05047, and the hospitality of Mr. and Mrs. William T. Golden.

### References

- [1] Bollobás, B., *Random Graphs*. Academic Press, London, 1985.
- [2] Bollobás B. and Erdős, P., Cliques in random graphs, *Math. Proc. Camb. Phil. Soc.*, **80**, 419–427, 1976.
- [3] Camerano, L., Dell'equilibrio dei viventi mercè la reciproca distruzione, *Atti della Reale Accademia delle Scienze di Torino*, **15**, 393–414, 1880.
- [4] Cohen, J. E., *Food Webs and Niche Space*. Princeton University Press, Princeton, NJ. Monographs in Population Biology, Vol. 11, 1978, 190.
- [5] Cohen, J. E., Threshold phenomena in random structures, *Discrete Applied Mathematics*, 113–128, 1988. Reprinted in *Applications of Graphs in Chemistry and Physics*, Ed. J. W. Kennedy and L. V. Quintas, Amsterdam, North-Holland, 113–128, 1988.
- [6] Cohen, J. E., Food webs and community structure. in *Perspectives on Ecological Theory*, Eds. J. Roughgarden, R. M. May, S. Levin, Princeton University Press, Princeton, 1989, 181–202.
- [7] Cohen, J. E., A stochastic theory of community food webs. VI. Heterogeneous alternatives to the cascade model, *Theoretical Population Biology*, **37**, 55–90, 1990.
- [8] Cohen, J. E., Briand, F. and Newman, C. M., A stochastic theory of community food webs. III. Predicted and observed lengths of food chains, *Proc. Roy. Soc.*, (London), **B. 228**, 317–353, 1986.

- [9] Cohen, J. E., Briand, F. and Newman, C. M., *Community Food Webs: Data and Theory*, Biomathematics, **20**, Springer-Verlag, Heidelberg, Berlin, New York, 1990.
- [10] Cohen, J. E., Luczak, T., Newman, C. M. and Zhou, Z. M., Stochastic structure and nonlinear dynamics of food webs: qualitative stability in a Lotka-Volterra cascade model, *Proc. Roy. Soc.*, (London), series B. **240**, 607–627, 1990.
- [11] Cohen, J. E., and Newman, C. M., A stochastic theory of community food webs. I. Models and aggregated data, *Proc. Roy. Soc.*, (London), **B. 224**, 421–448, 1985.
- [12] Cohen, J. E. and Newman, C. M., 1991 Community area and food chain length: theoretical predictions, *American Naturalist*, **138**, 1542–1554, 1991.
- [13] Cohen, J. E. and Palka, Z. J., A stochastic theory of community food webs: V. Intervality and triangulation in the trophic niche overlap graph, *American Naturalist*, **135(3)**, 435–463, 1990.
- [14] Darwin, Charles, 1859 *On the Origin of Species*. Facsimile of 1st ed. with Intro. by Ernst Mayr. Harvard University Press, Cambridge, MA, 1964.
- [15] Erdős, P. and Rényi, A., On the evolution of random graphs, *Publ. Math. Inst. Hung. Acad. Sci.*, **5**, 17–61, 1960.
- [16] Forbes, S. A., *Ecological Investigations of Stephen Alfred Forbes*, [Reprints.] Arno Press, New York, 1977.
- [17] Gause, G. F., 1934 *The Struggle for Existence*. Reprinted 1964. Hafner, New York.
- [18] Gelenbe, E., Nelson, R., Philips, T. and Tantawi, A., An approximation of the processing time for a random graph model of parallel computation, in *Proceedings of the Fall Joint Computer Conference*, H. S. Stone, Ed., IEEE Computer Society Press, Washington, DC, 691–697, 1986.
- [19] Jeffries, C., Klee, V. and van den Driessche, P., Qualitative stability of linear systems, *Linear Algebra and Its Applications*, **87**, 1–48, 1987.
- [20] Jeffries, C., Ecosystem modelling: qualitative stability. in M. G. Singh, *Systems and Control Encyclopedia; Theory, Technology, Applications*, Pergamon, Oxford, 1988, 1348–1351.

- [21] Jeffries, C., Eigenvalues, stability, and color tests, *Linear Algebra and Its Applications*, **107**, 1988, 65–76.
- [22] Kingsland, S. E., *Modeling Nature: Episodes in the History of Population Ecology*. University of Chicago Press, Chicago, 1985.
- [23] Krebs, Charles J., *Ecology: The Experimental Analysis of Distribution and Abundance*. Harper and Row, New York, 1972.
- [24] Lawton, J. H., Food webs. In J. M. Cherrett, Ed., *Ecological Concepts: The Contribution of Ecology to an Understanding of the Natural World*, Blackwell Scientific, Oxford, 1989, 43–78.
- [25] Lotka, A. J., 1924 *Elements of Physical Biology*. Williams and Wilkins, Baltimore. Rep. 1956: *Elements of Mathematical Biology*. Dover, New York.
- [26] May, R. M., ed. *Theoretical Ecology: Principles and Applications*. 2d ed. Blackwell Scientific, Oxford, 1981.
- [27] Newman, C. M., Chain lengths in certain random directed graphs, *Random Structures and Algorithms*, 1991. In Press.
- [28] Newman, C. M. and Cohen, J. E., A stochastic theory of community food webs. IV. Theory of food chain lengths in large webs, *Proc. Roy. Soc.*, (London), **B. 228**, 355–377, 1986.
- [29] Niering, W. A., Terrestrial ecology of Kapingamarangi Atoll, Caroline Islands, *Ecological Monographs*, **33**, 131–160, 1963.
- [30] Palmer, E. M., *Graphical Evolution: An Introduction to the Theory of Random Graphs*. John Wiley & Sons, New York, 1985.
- [31] Pimm, S. L., *Food Webs*. Chapman and Hall, London, 1982.
- [32] Pimm, S. L., The complexity and stability of ecosystems, *Nature*, **307**, 321–326, 1984.
- [33] Pimm, S., Lawton, J., Cohen, J. E., Food web patterns and their consequences, *Nature*, **350**, 669–674, 1991.
- [34] Quirk, J. and Ruppert, R., Qualitative economics and the stability of equilibrium, *Review of Economic Studies*, **32**, 311–326, 1965.
- [35] Robinson, D. F. and Foulds, L. R., *Digraphs: Theory and Techniques*. Gordon and Breach, New York, 1980.
- [36] Schoener, T. W., Food webs from the small to the large, *Ecology*, **70**, 1559–1589, 1989.

- [37] Yodzis, P., *Introduction to Theoretical Ecology*. Harper and Row, New York, 1989.