Recent Progress and Problems in Food Web Theory

Joel E. Cohen

Introduction

In 1968 I proposed a way to set lower bounds on the dimensionality of trophic niche space by using data on food webs. In 1978 I showed that, in community food webs in a single habitat, the data on trophic niche overlap were consistent with a one-dimensional trophic niche space more often than might be expected from some simplified random models of food webs.

This note has three parts. First, I will comment on some recent explanations and criticisms of my empirical finding of a high frequency of interval food webs (Sect. 2). Second, I will review some recent mathematical progress in the graph theory of food webs (Sect. 3). Third, I will state an outstanding theoretical question concerning the combinatorial structure of food webs (Sect. 4): to explain the observed distributions of the length of maximal food chains. I conclude with a list of open problems.

Intervality of food webs: fact or artifact

Let me review the general structure of the argument I used in 1978. I collected observed food webs. I tested each observed food web to see if it had or did not have a certain property, which I call property P. (The property of interest was intervality, but that is irrelevant for the moment.) I also constructed a model of food webs that described successfully properties $P_1, \ldots, P_d$ other than property P of food webs. [For example, in model 5 of Cohen (1978, p. 60), the number of kinds of predators was taken as given (P_1), the number of kinds of prey was taken as given (P_2), the number of predator-prey interactions, or directed arcs in the food web directed graph, was taken as given (P_3), and the number of overlaps among the diets of predators, or number of edges in the niche overlap graph, was shown (Cohen, 1978, p. 69) to be very well described by the model (P_4).] Relative to this model, the observed frequency of property P (intervality) substantially exceeded the expected frequency. I inferred that P remained in need of explanation. ["It is concluded that the high observed frequency of arrangements of niche overlap that can be represented in a one-dimensional niche space [property P] does not result from the operation, within the framework of several plausible models [in particular, of model 5 that describes $P_1, \ldots, P_4$], of chance alone" (Cohen, 1978, p. 92).]

More than one commentary has gone like this. "Ah yes, but if you adjoin to your model additional properties $P_{d+1}, P_{d+2}, \ldots, P_{d+D}$ then the model with properties $P_1, \ldots, P_{d+D}$ fully explains the observed frequency of property P. Therefore, property P is an artifact of the omission of $P_{d+1}, \ldots, P_{d+D}$ from your model."

For example, Yodzis (1981a) found that a model, which he interpreted in terms of energetic constraints on the assembly of communities, could not explain the observed frequency of intervality. When Yodzis (1982) constrained the artificial food webs generated from the model to have a number of dominant cliques equal to the number observed, the frequency of intervality could be explained. [When Yodzis (1981c) chose different parameter values for his original model, he was also able to account for the observed frequency of intervality, but Lawton and Pimm (1982) pointed out that these new parameter values appear to be biologically senseless.]

Similarly, Critchlow and Stearns (1982, p. 495) observed that the number of block submatrices in observed food web matrices is generally higher, and the number of distinct diet

overlaps in observed food web matrices is generally lower, than predicted by model 5. They asserted (p. 496), without any calculation to justify their claim: "Failure to account for block submatrices and distinct diet overlaps introduced a bias into the analysis that led directly to the conclusion that interval food webs are improbably frequent. No deeper explanation need be sought." To justify this claim, they would have to show that a model that correctly described the number of block submatrices and the number of distinct diet overlaps would also correctly predict the frequency of intervality. They did not show this.

In yet another approach, Sugihara (1982) showed that the frequency of interval food webs could be accounted for by requiring that the niche overlap graph be a rigid circuit graph, or by requiring that a certain simplicial complex, constructed to reflect the sharing of food resources by predators, have no holes.

He incidentally showed that the existence of block submatrices may not suffice to explain the excess frequency of interval food webs. The "food webs" he used are in fact connected components from real community food webs. When he randomly assembled model food webs to allow small holes in the simplicial complex that describes resource use, he found substantially fewer interval food webs in the artificial sample than in the real food webs. This finding appears to contradict at least that part of the unsupported claim of Critchlow and Stearns pertaining to block submatrices. However, the random model of Sugihara differs from my model 5, on which Critchlow and Stearns based their other calculations, in requiring that each compartment ("food web" in his language) be connected. Whether the frequency of intervality would be correctly predicted by my model 5 if a block submatrix structure were imposed on it remains open, but doubtful, in the light of Sugihara's results.

Let me pretend that the logical error in the argument of Critchlow and Stearns did not exist, and respond as if all the above workers had proved their points.

(a) Such explanations are exactly what I want; congratulations! Indeed, one purpose of modeling is to explain everything we think we know about food webs. It is a happy event when the task of explaining one empirical discovery drives the incorporation of further empirical discoveries into a model.

Now, however, that we have five explanations (two offered by Yodzis, two by Sugihara, and one, asserted only, by Critchlow and Stearns), plus qualitative ones offered by others, which, if any, is correct? Can they all be correct? If so, how are they to be reconciled? A large number of proofs of the existence of God raises doubts about the infallibility of any one of them, though not of course about the existence of God itself; the feeling is similar here.

(b) Some of the explanations offered appear to push the problem of theoretical explanation one step back. Suppose, as Yodzis (1982) claimed, the frequency of intervality is explained by constraining the number of dominant cliques, or, as Critchlow and Stearns (1982) claimed, by constraining the number of block submatrices and distinct diet overlaps. Then what explains the low number of dominant cliques and distinct diet overlaps and the high number of block submatrices in real food webs? Suppose you were trying to explain the variation in men's heights, and you observed that almost all the variation could be explained if you knew the heights of their ears. This represents some progress, but it would be nice to be able to explain the heights of their ears. In this regard the proposals of Sugihara (1982) are of special interest, because he has tried to derive imputed community properties from an underlying theory of population dynamics.

(c) The word "artifact" is loaded. The fraction of food webs that are interval is determined by the food web data. If failure to resolve different kinds of organisms biases food web data in the direction of intervality, the solution is to refine the observations. If the fraction of food webs that are interval can be explained by a tested model that also accounts for other features of food webs, it means that our understanding has been enlarged, not that intervality is an artifact.

In the second half of this section, I would like to point out two further misunderstandings in the article of Critchlow and Stearns (1982).

They asserted (p. 479) that "Cohen's techniques have no necessary connection to niches." If two different kinds of predators take a single kind of prey organism under similar circumstances at one stage of the prey's life cycle, I cannot see how any useful definition of niche would fail to imply that the niches of the two predators must overlap. It seems to me that information about dietary overlap *does* provide some information about niches, even if this information is not necessary and sufficient to describe all aspects of niche overlap.

They also asserted (p. 479) that "the interval graph algorithm selected by Cohen is the wrong technique." They amplified this assertion later (pp. 493-494): "The sets which overlap in a food web are the diets of the predators. These sets are directly accessible and should be tested for interval representation with the interval diets criterion. The interval diet overlaps criterion, i.e., Cohen's interval food web criterion, is an inherently more complicated and less precise method of testing a food web for interval diets . . ."

The interval graph algorithm I used is indeed the wrong technique for answering the question Critchlow and Stearns asked whether diets themselves have an interval representation. But that is a different question from the slightly more abstract one I was asking. To pretend, as Critchlow and Stearns have done, that they asked the same question I asked, only better, is to generate confusion rather than progress.

In brief, Critchlow and Stearns applied the consecutive one's tests to the food web matrix directly. I applied the consecutive one's test to the dominant clique matrix derived from the food web matrix. Critchlow and Stearns have repeated some well known examples in which the two approaches give different answers. Unfortunately these examples are too simple to be of help in deciding which approach is more useful. So consider the following example.

Suppose a food web C contains 7 kinds of predators $p_1, p_2, \ldots, p_7$. (The argument would hold for any number of predators.) Suppose C contains $\binom{7}{2} = 21$ kinds of prey. For each pair p_i, p_j of distinct predators, suppose there is a prey species that constitutes the shared diet of p_i and of p_j, and that the prey species is different for each distinct pair of predators. (This assumption is much stronger and less realistic than what is required to make the following point, but is a simple way of describing a food web matrix without the consecutive one's property that has an overlap graph that is a complete graph on 7 points.) Then the diet of every predator in C overlaps with the diet of every other predator in C. The consecutive diets test of Critchlow and Stearns would infer that the trophic niche space in this case could not be one-dimensional because the food web matrix fails the consecutive diets test. I believe it is more plausible to infer from the exhaustive pairwise overlapping of diets that there is some common region in trophic niche space that all 7 predators share, and therefore that the dimensionality of the trophic niche space need not exceed 1.

I hope this example makes it unmistakably clear that the two techniques really are different.

Graph theory of food webs

In this section I will describe some mathematical facts developed in the last decade concerning food webs, interval graphs and their generalizations. Many of these results are due to Fred S. Roberts and his students. The best single review known to me is by his doctoral student Cozzens, from whose thesis (1981) I learned of many of the results stated here.

A digraph (= directed graph) is a finite set of vertices (corresponding here to kinds of organisms) and arcs or directed edges. If (x, y) is an edge, then organism x preys on organism y. Digraphs are of ecological interest as an approximation to the combinatorial structure of food webs. It will be assumed here that all digraphs are acyclic and loopless. The niche overlap graph G of a digraph D is a graph with the same vertices as D and an undirected edge between vertices x and y if and only if, in D, there is a vertex z such that

(x, z) and (y, z) are arcs of D. In other words, the trophic niches of x and y overlap if and only if they take some prey in common.

The results I shall describe fall into two classes. The first class deals with the question: When is a given graph the niche overlap graph of some digraph? That is, which niche overlap graphs could arise from some food web? The second class deals with the question: Given a niche overlap graph, when is it the intersection graph of some given family of sets? E.g., when is it an interval graph?

Roberts (1978) showed that the niche overlap graph of an (acyclic loopless) digraph always has at least one isolated vertex, which is a vertex with no neighbors. Moreover, if G is any graph, there is a finite number k such that the union of G with k isolated vertices is the niche overlap graph of a digraph. The competition number k(G) of a graph G is the smallest integer k that works, i.e., the least number of isolated vertices that must be adjoined to G to make the union into the niche overlap graph of some (acyclic) digraph. Recently Opsut (1982) showed that the computation of a graph's competition number, and hence the characterization of niche overlap graphs of (acyclic loopless) digraphs, is an NP-complete problem. This means that the problem is hard.

Lundgren and Maybee (1982) have characterized niche overlap graphs. They define a collection S of sets of vertices of G to be a line cover of G if each set in S is either a clique in G or the empty set, and every vertex of G is in at least one of the cliques in S. Then they prove: a graph G with n vertices is a niche overlap graph if and only if the vertices of G can be labeled $v_1, v_2, \ldots, v_n$ so that G has a line cover $S = \{C_1, \ldots, C_n\}$ such that if the vertex v_i is in the set C_j then $i > j$. It follows that if G is a rigid circuit graph with an isolated vertex, then G is a niche overlap graph of some (acyclic loopless) digraph. In particular, every interval graph with an isolated vertex is a niche overlap graph.

Dutton and Brigham (1982) characterized the niche overlap graphs of digraphs that may have both cycles and loops. Roberts and Steif (1982) characterized the niche overlap graphs of digraphs that are loopless but may have cycles.

Roberts (1982) reviewed a variety of other applications of the mathematical ideas behind niche overlap graphs. It is illuminating to see the ecological applications in this broader context.

I now turn to the characterization of an undirected graph G (in the present context, G is a niche overlap graph) as the intersection graph of some family of sets. Recall that G is an intersection graph of a family of sets S if there exists a function f that assigns to each vertex v of G a set f(v) in S such that for all vertices x and y of G, (x, y) is an undirected edge of G if and only if f(x) and f(y) intersect, i.e., have a nonempty intersection.

An intersection graph of the family of intervals of the real line is an interval graph. The boxicity of a graph G, denoted b(G), is the smallest integer k such that G is an intersection graph of boxes (rectangular regions, open or closed) in the space of k-dimensional real vectors (Roberts, 1969). Obviously if G is an interval graph, b(G) = 1. Roberts (1978) showed that the boxicity of a niche overlap graph can be arbitrarily high. So an observation that many real niche overlap graphs have boxicity equal to 1 is not an artifact caused by considering niche overlap graphs. Although a linear-time algorithm is known for deciding whether a graph is an interval graph (Booth and Lueker, 1976), Cozzens (1981) proved that the computation of the boxicity of a graph is an NP-complete problem.

Given a graph G with n vertices, its adjacency matrix is an $n \times n$ matrix A with 1 everywhere on the main diagonal and, for $i \neq j$, with $a_{ij} = 1$ when (v_i, v_j) is an edge of G, $a_{ij} = 0$ otherwise. Mirkin (1972) showed that G is an interval graph if and only if the rows of its A can be permuted so that the consecutive one's property holds for the entries within a column at and below the main diagonal.

Another well-known characterization is that G is an interval graph if and only if none of its induced subgraphs is in a specified list of forbidden induced subgraphs. Steif (1982) showed that if D is the acyclic digraph of a niche overlap graph G that is interval, there

can be no forbidden induced subdigraph list that characterizes D. However, there does exist a list of forbidden *sink* induced subdigraphs for digraphs with interval niche overlap graphs. Here a sink induced subdigraph is an induced subdigraph such that for every vertex x in the subdigraph, if (x, y) is an arc in the main digraph, then y is also a vertex in the subdigraph. [This corresponds to my (1978) notion of a sink subweb.] Thus it is possible to determine whether a food web is interval by checking all its sink induced subdigraphs (sink subwebs) against a finite list.

Lundgren and Maybee (1982) suggested the nice idea of studying a food web's common enemy graph, which is the obvious dual to its niche overlap graph: if D is the original food web digraph and D' is the graph obtained by reversing the direction of all the arcs of D, then the common enemy graph of D is the niche overlap graph of D', i.e., two kinds of organisms have a common enemy if and only if there is a predator that preys on both of them.

I showed in 1978 that a community food web has a niche overlap graph that is interval if and only if every sink food web in it has a niche overlap graph that is interval. Dually, Lundgren and Maybee (1982) showed that a community food web has a common enemy graph that is interval if and only if every source food web contained in it does.

Another idea of potential ecological relevance is that of an overlap graph. A graph G is an overlap graph if there is a function from the vertices of G to intervals on the real line such that two vertices are adjacent in G if and only if the corresponding intervals intersect *and* neither interval properly contains the other. Obviously every overlap graph is an interval graph but not conversely. If one thinks of the intervals corresponding to vertices as one-dimensional regions in the trophic niche space, overlap graphs have the nice property of not allowing the trophic niche of one kind of organism to be contained in the trophic niche of another. Gavril (1973) proved that every overlap graph is a circle graph and conversely. A circle graph is defined as the intersection graph of a finite collection of chords on a circle.

Food chain lengths

The problem of explaining or predicting the length of food chains has long intrigued ecologists. The recent collection of substantial numbers of food webs in machine-readable form presents a new opportunity and challenge for quantitative tests of food chain models. So far as I know, not a single model has been published that predicts quantitatively the frequency distribution of lengths of maximal food chains in a single food web or collection of food webs.

Before anyone rushes out to do battle, let me sound a note of caution regarding the data. I published the frequency distribution of lengths of maximal food chains in four food webs in 1978 (p. 59). When I learned from a manuscript of Peter Yodzis that he had computed these frequency distributions for all 40 food webs in the collection of F. Briand, I asked for and he kindly sent all 40 frequency distributions. Three of my four analyzed food webs are among those in Briand's collection, so I compared the frequency distributions computed by Yodzis and me for those three. The statistics for my case 28.12 (Briand's 38), "Rocky shore, Lake Nyasa," were identical. The statistics for the other two cases, shown in Table 1, were not.

The differences are probably due to Briand's correction of the data I used. In the food web matrix of the Aspen forest, Canada, Briand (1983, Appendix) reported modifying entries that would explain why Yodzis showed two fewer prey than I did. This change affects substantially the frequency of maximal food chains of length two. Similarly, in the food web matrix of the sandy shore, Lake Nyasa, Briand reported deleting two columns and one row (these were duplications I did not detect). This change has little effect on the frequency distribution of chain lengths.

Table 1. Comparison of frequency distributions of lengths of maximal food chains in two food webs, as computed by Cohen (1978) and Yodzis

Food web	Aspen forest, Canada		Sandy shore, Lake Nyasa	
Author	Cohen	Yodzis	Cohen	Yodzis
Identification number	1.3	25	28.2	39
Predators	22	22	34	32
Prey	19	17	18	17
Interval	Yes	?	No	?
Length of chain	Frequency			
1	2	3	21	20
2	9	17	37	36
3	6	5	11	11
4	7	1	0	0
>4	0	0	0	0

In light of the possible sensitivity of the frequency distribution of maximal food chain length to the details of the food web data, it is important to make those data as reliable as possible before too much effort is invested in explaining the quantitative details of these frequency distributions.

In spite of this caveat, I believe that quantitative tests such as I propose are essential for sharpening our models and for getting rid of the host of speculations that are biologically plausible, qualitatively correct and actually wrong.

Conclusion: open problems

Aside from general exhortations, the preceding remarks identify some specific open problems.

1. Suppose random food webs are generated by my model 5, and all of them in which the number of block submatrices differs from the observed number are discarded. Among those remaining, will the frequency of intervality approximate the observed frequency of intervality? In other words, will conditioning on the observed number of block submatrices really explain the frequency of intervality, as Critchlow and Stearns (1982) claimed?

2. Similarly, will conditioning on the observed number of distinct diets explain the observed frequency of intervality?

3. Will conditioning on the number of block submatrices and the number of distinct diets explain intervality?

4. Are the common enemy graphs of real food webs interval about as frequently as would be expected from random models?

5. Are the niche overlap graphs of real food webs overlap graphs (i.e., representable without proper inclusion of intervals) about as frequently as would be expected from random models?

6. Can dynamic models, of the kind recommended by Lawton and Pimm, quantitatively explain the frequency distribution of maximal food chain lengths? Can they quantitatively account for the frequency of intervality?

Acknowledgments

REFERENCES

Booth, K., and G. Lueker. 1976. Testing for the consecutive ones property, interval graphs, and graph planarity using PQ-tree algorithms. *Journal of Computer and Systems Science* 13, 335-379.

Briand, F. 1983. Environmental control of food web structure. *Ecology* 64, 253-263.

Cohen, J. E. 1968. Interval graphs and food webs: a finding and a problem. RAND Corporation Document D-17696-PR, 30 August 1968.

Cohen, J. E. 1978. Food Webs and Niche Space. Princeton: Princeton University Press.

Cozzens, M. B. 1981. Higher and Multi-Dimensional Analogues of Interval Graphs. Ph.D. thesis, Rutgers University, New Brunswick, N.J. March 1981.

Critchlow, R. E., and S. C. Stearns. 1982. The structure of food webs. *American Naturalist* 120, 478-499.

Dutton, R. D., and R. C. Brigham. 1982. A characterization of competition graphs. In press, *Discrete Applied Mathematics*.

Gavril, F. 1973. Algorithms for a maximum clique and a maximum independent set of a circle graph. *Networks* 4, 261-273.

Lawton, J. H., and S. L. Pimm. 1982. Pattern and process in food webs. Manuscript, May 1982.

Lundgren, J. R., and J. S. Maybee. 1982. Food webs, competition graphs and associated matrices. In press, *Discrete Applied Mathematics*.

Mirkin, B. G. 1972. Description of some relations on the set of real-line intervals. *Journal of Mathematical Psychology* 9, 243-252.

Opsut, R. J. 1982. On the computation of the competition number of a graph. In press, *SIAM Journal on Algebraic and Discrete Methods*.

Roberts, F. S. 1969. On the boxicity and cubicity of a graph. In: Recent Progress in Combinatorics (W. T. Tutte, ed.), New York, Academic Press, 301-310.

Roberts, F. S. 1978. Food webs, competition graphs, and the boxicity of ecological phase space. In: Theory and Applications of Graphs (Y. Alavi and D. Lick, eds.), New York, Springer-Verlag, 477-490.

Roberts, F. S. 1982. Application of edge coverings by cliques. Manuscript, June 1982.

Roberts, F. S., and J. E. Steif. 1982. A characterization of competition graphs of arbitrary digraphs. In press, *Discrete Applied Mathematics*.

Steif, J. E. 1982. Frame Dimension, Generalized Competition Graphs, and Forbidden Sublist Characterizations. B.A. thesis, Rutgers College, New Brunswick, N.J. May 1982.

Sugihara, G. 1982. Niche Hierarchy: Structure, Organization and an Assembly in Natural Communities. Ph.D. thesis, Princeton University, Princeton, N.J. May 1982.

Yodzis, P. 1981a. The structure of assembled communities. *Journal of Theoretical Biology* 92, 103-117.

Yodzis, P. 1982. The compartmentalization of real and assembled ecosystems. *American Naturalist* 120, 551-570.

Yodzis, P. 1981c. Why are so many niche spaces one-dimensional? Manuscript, August 1981.