Graph Theory, Homology and Food Webs

George Sugihara

1. INTRODUCTION

Ecological systems consist of large networks of interrelated parts. The living components of these networks are linked largely by trophic interactions involving predator-prey encounters and competition.

Our aim here will be to show how the formal tools of graph theory and algebraic topology may be brought to bear to uncover patterns in the structure of real food webs. We will begin by examining the various ways that food webs can be represented, and will then apply these representations to search for regularities in the structure of natural ecosystems.

The pattern of interactions in food webs is of scientific interest because it determines the flow of energy and materials (e.g., nutrients and pollutants) through ecological systems as well as the dynamics and stability of species populations. Knowledge that real ecosystems are contained within a specific subset of possible topologies has fundamental interest as it helps to illuminate those narrow regions in a wide landscape of mathematical possibilities that warrant further study.

2. CHARACTERIZING FOOD WEB STRUCTURE

The basic information that is used to construct a food web is contained in the so-called food web matrix $A$ (Fig. 1). This is a binary matrix whose columns $C$ correspond to the set of consumer species and whose rows $R$ correspond to the set of resources or prey in the system. The 1's in the matrix signify

---

1Research supported by an Ogden Porter Jacobus Prize from Princeton University and a Eugene P. Wigner Fellowship from Oak Ridge National Laboratory (operated by Union Carbide Corporation under contract W-7405-eng-26 with the U.S. Department of Energy), Oak Ridge, Tennessee. Publication No. 2229, Environmental Sciences Division, ORNL.
that a given consumer uses a given prey item which, in turn, generates the ordered pairs \((C_i,R_j)\). Notice that a given column \(C_i\) in \(A\) identifies the subset of resources used by that consumer and each row \(R_j\) identifies the subset of consumers that share a given resource.

Ordered pairs in the food web matrix can be put into graphical form by constructing the digraph of trophic flows (Fig. 2). The digraph \(D(V,E^*)\) consists of a finite nonempty set \(V\) of vertices formed from the union of consumers \(C\) and resources \(R\), along with an arc set \(E^*\) containing the ordered pairs \((C_i,R_j)\) from \(V\). The vertices \(V\) in \(D(V,E^*)\), therefore, consist of the set of all species in the food web (both predators and prey) and the directed edges \(E^*\) indicate who eats whom. The digraph is the classical standard for representing food webs (e.g., see Shelford 1913).

Two newer representations involving undirected graphs (no ordered pairs) may be constructed by choosing either consumers or resources, separately, as

<table>
<thead>
<tr>
<th>Community Food Web Matrix (Knysna Estuary)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14</td>
</tr>
</tbody>
</table>

**LEGEND**
2. attached plants 10. Arenicola
5. Hypergamphus 11. Hymenosoma
7. Upogebia 13. Lithognathus
8. Lamna 14. Rhabdosargus
9. Solen 15. Hypacanthus

**FIGURE 1.** Data from Day (1967) (Sugiha 1982).
object sets. These undirected graphs are examples of intersection graphs defined as follows.

Let $S = (S_1, \ldots, S_p)$ be an object set consisting of a nonempty family of distinct subsets. The intersection graph $G(S)$ is formed by identifying $S$ with the vertices of $G$ and making $S_i$ and $S_j$ adjacent whenever $i \neq j$ and $S_i \cap S_j \neq \emptyset$. Hence an edge is drawn between two vertices in $G(S)$ whenever the subsets that they represent overlap. Mardzewski (1945) has shown that every graph is an intersection graph.

The intersection graph that is formed by choosing $C = (C_1, \ldots, C_p)$ as the object set is Cohen's (1979) consumer overlap graph $G(C)$ (Fig. 3). Here, whenever two consumers (subsets of resources) $C_i$ and $C_j$ overlap with respect to at least one resource an edge is drawn between them. The vertices in $G(C)$ represent consumer species and the edges signify overlap in consumer resource use. Such overlap may be determined mechanically by checking the columns of $A$ for nonzero dot products.

The consumer overlap graph provides a portrait of the potential competitive structure of a food web. It is a simplified picture, however, in that a single edge joining two consumer species may represent overlap with respect to multiple resources.

A fuller characterization of consumer interrelationships may be obtained by taking the formal conjugate of $G(C)$ (Sugihara 1982). Herein let $R$ be the object set and form the intersection graph $G(R)$ over $R$ by joining those
resources that belong to a single consumer species. This involves checking for nonzero dot products between the rows of $A$.

The resulting resource graph $G(R)$ (Fig. 4) gives a more complete picture than $G(C)$ for two reasons. First, whereas a consumer species in $G(C)$ is represented as a single vertex, in $G(R)$ it is represented by the constellation of resources used. That is to say, the $n$-dimensional trophic niche of a species in $G(C)$ corresponds to a clique (maximally connected subgraph) over its $n$ resource vertices. A species that uses 5 resources is represented in $G(R)$ by its 5-dimensional niche, that is, as a clique over its 5 labeled resource vertices (see shaded area in Fig. 4). Secondly, unlike $G(C)$, $G(R)$ contains information as to the number of resources involved in the niche overlaps. As can be seen in Figure 4 the intersection of species niches may involve multiple resources.

For pedagogical purposes it may be useful to construct the multigraph (a graph having multiple edges) version of $G(R)$ out of tinker toys. Balls of modeling clay may be used to represent resource vertices, and an individual species niche may be represented by struts of a given color maximally connecting the resource set. Such a model is always possible due to the fact that any graph can be imbedded (with no edges crossing) in 3-space using only straight lines (e.g., Boesch 1981).

The information in the resource graph $G(R)$ may be given a more robust geometry if the notions of clique and graph are replaced with the terms simplex and complex as defined below.
Let $v_1, \ldots, v_n$ be vectors in $\mathbb{R}^n$. A vector $V$ is said to be affinely dependent on them if there exist real numbers $\lambda_0, \ldots, \lambda_n$ called barycentric coordinates such that

$$\sum_{i=0}^{n} \lambda_i = 1 \quad \text{and} \quad V = \lambda_0 v_0 + \ldots + \lambda_n v_n.$$  

Suppose further that $v_0, \ldots, v_n$ are affinely independent (none affinely dependent on the rest). Then, a simplex (closed) having vertices $v_0, \ldots, v_n$ is defined to be the set of points affinely dependent on $v_0, \ldots, v_n$ such that every barycentric coordinate is $\geq 0$. The boundary of a simplex consists of those points which have at least one barycentric coordinate equal to 0, and its dimension $n$ is simply the number of vertices minus 1.

Notice how the notion of simplex corresponds with the graphical term "clique". Both are similar in that vertices and edges in the simplex identify with vertices and edges in the clique, however in dimensions above 1 the simplex characteristics of interior, exterior and boundary, as defined by the barycentric coordinates, differ from those in the graph. Harary (1972) defines a graph as a complex of simplexes of dimension 0 or 1. A simplex, therefore, can be viewed
as a clique which has been inflated in a higher dimensional space to have an interior that is filled, and a clique conversely can be viewed as the vertices and edges of a simplex, or the $l$-skeleton of a simplex. Notice that because a clique is convex, so is the simplex. That is, if $u$ and $w$ are two vertices of the simplex then every straight line segment joining $u$ and $w$ belongs to the simplex. Indeed the simplex is the smallest convex set containing its vertices which is expressed by saying it is the convex hull of its vertices.
Paired with our graphical representation of the species niche in $G(R)$ as an $n$-pointed clique; it is also possible to represent the $n$-dimensional niche of a species as a simplex over $n$ resource vertices. In geometrical terms this gives rise to a solid tinker toy model where a species niche may be thought of as a convex polyhedron whose $n$ vertices correspond to the $n$ resources used by that species. This generates a powerful architectural description of communities where the polyhedral species niches are joined through shared resource vertices to form a simplicial complex $K(R)$ (Fig. 6).

According to this model a community hypervolume may be broadly visualized as a multidimensional crystal having clusters of crystal faces which correspond to individual species niches. The simplicial complex model allows one to visualize clearly the geometry by which species niches are fitted together in communities.

3. TOPOLOGY OF REAL FOOD WEBS

We shall now apply these characterizations to a collection of 40 community food web matrices recently extracted from the literature (Sugihara 1982, after Briand 1983). In all, this collection contains 73 nontrivial communities, defined here as connected components (all vertices linked through some sequence of edges) in $G(C)$ containing 2 or more consumer species.

3.1 HOLES: THE GEOMETRY OF SPECIES PACKING

A question of biological interest concerns how closely species niches are packed together in $K(R)$. That is, are species niches packed tightly so that each community in $K(R)$ is a simple solid (Fig. 6B), or are they fitted together more loosely so that $K(R)$ resembles a multidimensional swiss cheese (Fig. 6A)? In biological terms, we are asking whether there are any minimal constraints acting on the geometry of species packing.

3.1.1 TECHNICAL PRELIMINARIES

The mathematics which we will use to address this question is homology theory or the study of multidimensional holes. The full technical details of these ideas are rather baroque, and the interested reader is referred to the excellent expositions given by Atkin (1974) and Giblin (1977). It will suffice to say, however, that basic information as of the holiness of a complex is contained in the $p^{th}$ order betti number, $\beta_p(K)$. The $p^{th}$ order betti number represents the number of $p$-dimensional holes in a complex and is calculated as the rank of the $p^{th}$ homology group

$$\beta_p(K) = \text{rank } H_p(K).$$
This is formed as the quotient group of the family of all $p$-dimensional cycles, $Z_p(K)$ and the family of bounding cycles $B_p(K)$ in $K$,

$$H_p(K) = Z_p(K)/B_p(K).$$

$H_p(K)$, therefore, consists of non-bounding $p$-cycles and the $p^{th}$ order betti number is the number of non-bounding $p$-dimensional cycles in $K$. These $p$-cycles are the kernel of the standard boundary homomorphism.

Some intuition may be gained by demonstrating how to compute the first order betti number for complexes having only 1-dimensional holes (Fig. 7). Put briefly, this involves finding the number of independent 1-cycles $[\text{rank } Z_1(G)]$ in $K(R)$ and subtracting out the number of those that are in the boundary of a species simplex $[\text{rank } B_1(K)]$,

$$\beta_0(K) = \text{Rank } Z_1(G) - \text{Rank } B_1(K)$$

The computation of the first order betti number, therefore, involves finding the number of non-bounding 1-cycles in the 1-skeleton of $K(R)$. Because $G(R)$ is the 1-skeleton of $K(R)$, the number of independent 1-cycles of $K(R)$ is the same as the rank of the cycle basis of $G(R)$.

$$\text{Rank } Z_1(G) = \text{Rank } \{\text{cycle basis } G(R)\}$$
The cardinality of the cycle basis for $G(R)$ is calculated from the elegant Euler-Poincare equation

$$\text{Rank } Z_1(G) = \text{Edges} - \text{Vertices} + 1 .$$

In Figure 8A below, for example, \(\text{rank } Z_1(G) = 9 - 6 + 1 = 4\), hence there are 4 independent 1-cycles in this complex. For 8B and C there are 6 and 8 independent 1-cycles respectively.

The number of independent 1-cycles in the boundaries of all species simplexes \(\text{rank } B_1(K)\) is calculated from the union of independent 1-cycles in the species cliques in $G(R)$ corresponding to maximal simplexes in $K(R)$ (contained no larger simplex). For example, in Figure 8B the maximal simplexes are the 2 shaded triangles and the solid tetrahedron. Applying equation 2 to the cliques corresponding to each maximal simplex (1-skeleton), the number of independent bounding 1-cycles is computed as

$$\text{Rank } B_1(K) = (3 - 3 + 1) + (3 - 3 + 1) + (6 - 4 + 1) = 5 .$$

As calculated earlier, the total number of independent 1-cycles (both bounding and non-bounding), \(\text{rank } Z_1(G)\), is 6. Hence applying equation 1 we find

$$\beta_1(K) = 6 - 5 = 1 .$$
This indicates the presence of a single 1-dimensional hole in $K(\mathcal{R})$. For simplices that are joined at 3 or more points and whose intersection thereby contains cycles, rank $B_1(K)$ is computed as the union of bounding cycles for each simplex, calculated by working the inclusion-exclusion principle in the usual way to avoid counting cycles more than once. For the complex in Figure 8C, which consists of 2 tetrahedra and 2 triangles, the bounding cycles may be calculated as

$$\text{Rank } B_1(K) = (3)_a + (3 - 1)_b + (1)_c + (1)_d = 7$$

where 1 is subtracted from $b$ to account for the shared cycle in the intersection of simplices $a$ and $b$. In this example, $\beta_1(G) = 8$ so that $\beta_1(K) = 1$, again demonstrating the existence of a single 1-dimensional hole in the complex.

3.1.2 RESULTS

A representative simplicial complex illustrating Bird's (1930) Aspen Parkland community is pictured below (Fig. 9). This community complex is typical in that although it possesses a rich mosaic pattern, from a homological standpoint it is rather simple and contains no holes of dimension 1 or greater. In all,
out of the the 73 community complexes tested, none had $p$-dimensional holes with $p \geq 2$ bounded by spheres or hyperspheres. Of the 60 communities that could possibly have 1-dimensional holes only two in fact did; although these, it has been argued, should be held with some reserve (Sugihara 1982). The data suggest, therefore, that holes are extremely rare in real niche spaces.

Aside from its interest as a fundamental topological property of communities, the deeper biological significance of this result is the implication that
resources in the environment are ordered or correlated with each other by various means (e.g., spatially, taxonomically, by size), and that this ordering is perceived similarly by all of the species involved. For example, consider an insectivorous lizard community where the prey sizes may be grouped into the categories large, middle, and small. As seen in Figure 10, the appearance of a lizard on the scene that eats only large and small insects, while avoiding the middle-sized ones, will create a hole. On the other hand, if this species obeyed the size ordering and also took in middle-sized prey items, the hole would be filled. One should not be misled by this grossly simple example because some of the orderings are observed to be quite subtle (Sugihara 1982).

![Diagram](image)

**Figure 10.** $K(R)$ for a community where each species uses 2 resource sizes. Here, the species which uses large and small insects while avoiding the middle-sized ones creates a hole (Sugihara 1982).

### 3.2 Assembly Rules for $G(C)$

Turning now to the condensed picture of niche overlaps given by $G(C)$, we will attempt to obtain insight into the process of community assembly from the observation that real consumer overlap graphs possess the rigid circuit property (Sugihara 1982). A connected graph $G$ is said to be *rigid circuit* or *chordal* if every circuitous path $P_n \in G$ of length $n \geq 4$ is shortened by a chord (Dirac...
Such graphs have also been called triangulated (Rose et al. 1967) since all generated subgraphs contain no more than triangular circuits.

The following results for rigid circuit graphs lead to biologically reasonable assembly rules for $G(C)$.

---

**Figure 11.** An illustration of the rigid circuit property. Notice in graph A that b and d are extreme vertices (neighbors form a clique) whereas a and c are not (Sugihara 1982).
LEMMA 1. (Dirac 1961). Every rigid circuit graph possesses at least one extreme vertex. That is, a vertex whose neighbors form a complete graph or clique.

LEMMA 2. If $G$ is a rigid circuit graph, then $G - V$ is also rigid.

PROOF. This follows trivially from the definition of triangulated graphs since every generated subgraph of $G$ must be rigid.

A perfect elimination ordering (p.e.o.) (Fig. 12) is an ordering on the vertex set $V$

$$\alpha: V \rightarrow \{1, 2, \ldots, n\}$$

such that for every pair of vertices $(v_i, v_j)$ not joined by an edge there exists no minimal path from $v_i$ to $v_j$ containing only $v_i$ and $v_j$ and vertices numbered less than

![Diagram](image)

Figure 12. A perfect elimination ordering (Sugihara 1982).
min[\alpha(v_i), \alpha(v_j)]$. Therefore collapsing a graph by following a p.e.o. involves eliminating extreme vertices to generate the sequence of graphs

$$G_n, G_{n-1}, ..., G_1$$

having $n, ..., 1$ vertices.

**Theorem 1.** $G$ is a rigid circuit graph if and only if it possesses a perfect elimination ordering.

**Proof.** Lemmas 1 and 2 lead directly to the result that rigid circuit graphs have a p.e.o. Conversely, because a p.e.o. involves eliminating extreme vertices there can be no generated subgraph containing $P_k$, $n \geq 4$ (see also Rose et al. 1976).

The reverse of a perfect elimination ordering is a perfect addition ordering (p.a.o.) (Fig. 13),

$$\gamma(V_i) = n - \alpha(V_i) + 1$$

This generates the sequence of graphs

$$G_1, G_2, ..., G_n$$

![Diagram](image)

**Figure 13.** An assembly rule deduced from the rigid circuit property (Sugihara 1982).
by adding rather than deleting extreme vertices. Clearly theorem 1 holds for
perfect addition as well as perfect elimination.

The fact that real niche overlap graphs $G(C)$ are rigid circuit guarantees the
existence of a p.a.o. However, to say that such an ordering exists is not to say
that real niche overlap graphs were constructed in this way. The following bio-
logical arguments may lead to such an ordering.

ARGUMENT 1. It is conventional wisdom that species tend to enter communi-
ties in order of increasing (non-decreasing) specialization. That is, generalist
species using a wide variety of resources enter first with specialists having nar-
rower requirements coming in later on. Adding sequential specialists to $G(C)$
may generate a p.a.o. by requiring incoming species to attach as extreme ver-
tices within single cliques. A clique in $G(C)$ corresponds with the biological term
guild, or group of functionally related species. If an incoming species bridges
two different functional groups it will likely be more generalized than any one
of the species it overlaps. Therefore, if an incoming species is a specialist it will
tend to attach within a single guild, as an extreme vertex. This, in turn gen-
erates a p.a.o.

A more rigorous and general argument follows.

ARGUMENT 2. Suppose specialization at the $i^{th}$ stage is associated with fewer
niche overlaps or competitors in $G(C)$. A specialist, therefore, would have lower
point degree (fewer radiating edges) than a generalist. Suppose further that
each species subsequent to the $i^{th}$ one is a greater (or equal) specialist with
respect to the species already present at the $i^{th}$ stage $(v_1, \ldots, v_{i-1})$. That is, no
subsequent species $v_j, j > i$ will overlap more of the vertices $(v_1, \ldots, v_{i-1})$ than
$v_i$. This produces the following rule for ordering vertices in $G$.

\hspace{1cm} (P1). As the next vertex $v_i$ to number choose the one adjacent to the most
labeled vertices $(v_1, \ldots, v_{i-1})$.

To see that this condition generates a p.a.o. we will use the following property
of orderings which is strong enough to imply p.e.o. on triangulated graphs (Tar-
jan 1978).

\hspace{1cm} (P2). If $\alpha(v_1) < \alpha(v_2), (v_1, v_2) \in E$ and $(v_1, v_2) \notin E$, then there exists a
vertex $v^*$ such that $\alpha(v_1) < \alpha(v^*), (v_1, v^*) \in E$ and $(v_1, v^*) \notin E$.

LEMMA 3. (Tarjan 1978). If $G$ is triangulated, any ordering that satisfies P2 is
a p.e.o.

THEOREM 2. If $G$ is triangulated, any ordering generated by P1 is a p.a.o.
GRAPH, THEORY, HOMOLOGY AND FOOD WEBS

PROOF. Let \( \alpha \) be a reverse ordering generated by \( P1 \) where vertices are numbered from \( n \) to 1. Due to lemma 3, it will suffice to show that \( \alpha \) satisfies \( P2 \). Suppose \( \alpha(v_3) < \alpha(v_1) < \alpha(v_2), (v_3,v_2) \in E \) but \( (v_1,v_2) \notin E \). According to \( P1 \) when \( v_1 \) is labeled it must be adjacent to at least as many vertices numbered before \( v_1 \) as \( v_3 \). Since \( (v_3,v_2) \in E \) but \( (v_1,v_2) \notin E \), there must be some vertex \( v^* \) numbered before \( v_1 \) to which \( v_1 \) is adjacent but not \( v_3 \). Thus \( \alpha \) satisfies \( P2 \), and \( P1 \) generates a p.a.o. (see also Tarjan 1978).

Therefore whether one argues loosely as in argument 1 or more precisely as in argument 2, increasing specialization combined with the rigid circuit property leads to the rule that species enter communities by attaching within individual guilds or cliques (perfect addition) rather than across multiple guilds. Given the conditions set forth in argument 2, this rule is both necessary and sufficient.

3.3 ASSEMBLY AND HOLES

The following consequence of the above assembly rule can explain the lack of holes in \( K(R) \) discussed in Sect. 3.1.

LEMMA 4. Suppose that a guild or a clique in \( G(C) \) is defined by a common resource \( v_e \) in \( K(R) \). (True in most observed cases.) Then, perfect addition in \( G(C) \) pre-empts holes in \( K(R) \).

PROOF. It is easy to see that requiring all species; say the \( i^{\text{th}} \), to attach to single guilds (p.a.o.) implies that the distance (number of edges) in \( K(R) \) between \( v_e \) and some other resource \( v_j \) in the attachment set must be

\[
d(v_e,v_j) < 2.
\]

This restriction pre-empts the possibility of holes in a community complex.

Therefore rigidity in niche overlaps \( G(C) \) and the absence of holes may both be explained by our derived assembly rule.

3.4 ASSEMBLY AND INTERVALITY

A graph \( G \) is said to be interval (Fig. 14) if it is isomorphic to some graph \( \Omega(F) \) where \( F \) is a family of intervals. That is,

\[
v_i \text{ adj } v_j \text{ in } G \text{ iff } f_i \cap f_j \neq \emptyset \text{ in } \Omega
\]

Intervality is a special case of boxicity (Roberts 1969) which is the minimum number of dimensions required of boxes to represent adjacency in \( G \). Interval graphs, therefore, have boxicity 1.
The importance of intervality for the study of food webs derives from Cohen's (1978) provocative finding that nature contains an excess of interval niche overlap graphs $G(C)$. To find natural food webs contained in this narrow subset of possible topologies implies a peculiar nonrandomness in the construction of ecological systems. Despite numerous attempts to find an explanation, no biologically plausible mechanism has been discovered for this curious phenomenon (Cohen 1978, Yodzis 1982). One can gain some insight into the problem, however, from the following necessary and sufficient conditions for interval graphs.

**Theorem 3.** (Lekkerkerker and Boland 1962). A graph $G$ is interval iff

1. $G$ is a rigid circuit graph.
2. $G$ is non-asteroidal (i.e., $G$ does not contain 3 distinct points $v_o, v_1, v_2$ and three paths connecting them $W_o, W_1, W_2$ such that each point $v_i$ of this triple of points is not a neighbor of any path $W_i$ connecting the other two).

Because real niche overlap graphs are rigid circuit it seems worthwhile to test whether conditioning on rigidity alone may account for the high frequency of interval niche overlap graphs. That is, could the excess of interval niche overlap graphs in nature be a simple consequence of the fact that they are triangulated? In particular could it be a consequence of the assembly rule for rigid circuit graphs deduced earlier?
GRAPH, THEORY, HOMOLOGY AND FOOD WEBS

Out of our collection of 73 nontrivial communities \( G(C) \) (connected components with 2 or more species), 63 were observed to be interval. When these communities are numerically assembled by a random perfect addition ordering, 62.36 are found to be interval, and the difference is not significant (\( z = 0.33 \)). These random rigid circuit graphs have the same number of vertices and edges as the observed communities however they are assembled by a perfect addition ordering which uses uniform distributions over appropriate limits to determine the sizes and identities of the attachment sets at each step. Therefore the assembly rule deduced in Sect. 3.2 appears to be sufficient to account for the high frequency of interval niche overlap graphs observed in nature.

4. REFERENCES


ENVIRONMENTAL SCIENCES DIVISION
OAK RIDGE NATIONAL LABORATORY
OAK RIDGE, TENNESSEE 37830