Ecological Network Structure: Data, Models, and Inference

Jennifer A. Dunne
Santa Fe Institute
Pacific Ecoinformatics & Computational Ecology Lab

PEaCE Lab: www.foodwebs.org
Why is network anatomy so important to characterize?  
Because structure always affects function.

Technological Networks  
- Road Maps
- Internet Connectivity
- Circuit Boards

Social Networks  
- The Kevin Bacon Game
- Support Network for a Homeless Woman

Biological Networks  
- Proteins
- Neurons
- Ecosystems

Nodes = Species/Taxa
Edges = Directed trophic links

- predation
- herbivory
- detritivory
- parasitism
- cannibalism
1950's Paradigm:
Complex communities MORE stable than simple communities

1970's Challenge:
Complex communities LESS stable than simple communities

Current & Future Research:
“Devious strategies” that promote stability and species coexistence
1. Food-Web Data
2. General Patterns Across Webs
3. Network Structure Models
4. Confronting Models with Data
5. Case Study: Ancient Food Webs
1. Food-Web Data
Earliest known graph of feeding relations

Lorenzo Camerano, 1880

Network of 15 taxa:
- Amphibians
- Reptiles
- Fish
- Birds
- Mammals
- Worms
- Crustaceans
- Spiders
- Various insects
- Plants
- Parasitic plants
1st generation data: 1920s-1980s:

Summerhayes & Elton 1923: Food web of Bear Island
Bear Island

1 bacteria, 4 autotrophs, 13 invertebrates, 6 birds, 4 mammals

Directed Connectance ($C$): Proportion of possible links ($S^2$) that are realized ($L$)

- $S$ (# taxa) = 28
- $L$ (# links) = 59
- $L/S$ (links/species) = 2.1
- $C$ (connectance; $L/S^2$) = 0.075
- TL (mean trophic level) = 2.07
2nd generation data: 1990s-present

Food Web of Little Rock Lake, Wisconsin

Martinez 1991
Examples of currently used datasets

$S \sim 25$ to 180, $C \sim 0.03$ to 0.3

Lake & Pond Webs

Bridge Brook Lake, Little Rock Lake, Skipwith Pond

Stream Webs

Canton Creek, Stony Stream

Estuary Webs

St. Marks Seagrass, Chesapeake Bay, Ythan Estuary

Terrestrial Webs

El Verde Rainforest, St. Martin Island, Coachella Valley

Marine Webs

Benguela, Caribbean Reef, NE US Shelf
3rd generation” data: 2009 and beyond

Antarctic Weddell Sea Food Web

Highly & Evenly Resolved

Original species = 492
62 autotrophs
4 mixotrophs
345 invertebrates
48 ectotherm vertebrates
29 endotherm vertebrates
3 detritus
1 bacteria

S = 290
L = 7200
L/S = 24.8
C = 0.086
Mean TL = 3.79

Data compiled by Ute Jacob
2. General Patterns Across Webs
Apparent Complexity
Beyond S and C: Link distributions

Little Rock Lake

Exponential, not Power Law
Apparent complexity

Raw data for 16 webs

# of trophic links

Cumulative distribution

- a. Grassland
- b. Scotch Broom
- c. Salton Sea
- d. Yehan Estuary 1
- e. Yehan Estuary 2
- f. El Verde Rain forest
- g. Stony Stream
- h. Chesapeake Bay
- i. St. Marks Seagrass
- j. St. Marys Island
- k. Little Rock Lake
- l. Lake Taboo
- m. Mirror Lake
- n. Bridge Brook Lake
- o. Coachella Valley
- p. Skippers Pond

$r^2$ values:

- a. 0.95
- b. 0.97
- c. 0.95
- d. 0.98
- e. 0.99
- f. 0.99
- g. 0.96
- h. 0.99
- i. 0.98
- j. 0.95
- k. 0.99
- l. 0.96
- m. 0.96
- n. 0.99
- o. 0.95
- p. 0.97
Apparent complexity → Underlying simplicity

Raw data for 16 webs

Normalized data for 16 webs

Camacho et al. 2002, Dunne et al. 2002
Whence the exponential?

**Principle of Maximum Entropy (MaxEnt, Jaynes 1957)**

The probability distribution with the maximum information entropy is the least biased probability distribution which satisfies a set of information-containing constraints.

69%

Percent of 51 food webs with degree distributions not significantly different from MaxEnt distributions constrained by # species, links, and top or basal spp.

Why should observed distributions tend toward MaxEnt distributions?

Observed large-scale patterns typically arise from aggregation of many small-scale processes, whose fluctuations tend to cancel each other out. A few constraints set the pattern in the aggregate, everything else tends to the greatest randomness (Frank, in press).

Williams 2009
Beyond degree distribution

Types of Organisms:

- % Top spp. = 1.1
- % Intermediate spp. = 85.9
- % Basal spp. = 13.0
- % Cannibal spp. = 14.1
- % Herbivore spp. = 37.0
- % Omnivore sp. = 39.1
- % Species in loops = 26.1

Linkage Metrics:

- Mean food chain length = 7.28
- SD food chain length = 1.31
- Log number of chains = 5.75
- Mean trophic level = 2.40
- Mean max. trophic simil. = 0.74
- SD vulnerability (#pred.) = 0.60
- SD generality (#prey) = 1.42
- SD links (#total links) = 0.71
- Mean shortest path = 1.91
- Clustering coefficient = 0.18
Scale dependence with $S$ & $C$

Data from 7 Food Webs

3. Network Structure Models
Empirical regularities $\rightarrow$ modeling opportunities

Simple, stochastic, single-dimensional models of food-web structure

Explain "the phenomenology of observed food web structure, using a minimum of hypotheses"

1) Two Parameters: $S$ (species richness) and $C$ (connectance)
2) Assign each species $i$ a uniform random "niche value" $n_i$ of 0 to 1
3) Simple rules distribute links from consumers to resources

Cascade model

Link distribution rules:
→ Each species $i$ has probability $P = \frac{2CS}{(S-1)}$ of consuming resource species $j$ with lower niche values ($n_j < n_i$)

Effect of rules:
→ Creates strict hierarchy of feeding (cannibalism & longer cycles prohibited)

Cohen & Newman 1985
Niche model

Link distribution rules:

→ Species i is assigned a feeding range $r_i$ (drawn from beta distribution)
→ The center $c_i$ of $r_i$ is a uniform random number between $r_i/2$ and min $(n_i, 1-r_i/2)$
→ Species i feeds on all species that fall within $r_i$

Effect of rules:

→ Beta distribution generates exponential-type degree distributions
→ The feeding hierarchy is slightly relaxed (cycles can occur)
→ Webs are "interval" (species feed on contiguous sets of species)

Williams & Martinez 2000
Nested hierarchy model

Link distribution rules:

→ Each consumer i’s number of resource species j assigned using beta distribution

→ Resources j chosen randomly from species with \( n_j < n_i \) until all links are assigned or a j is obtained which already has at least one consumer

→ Species i links to j and joins j’s “consumer group”

→ Subsequent j chosen randomly from the set of j of this group until all of i’s links are assigned or all j of the consumer group have been chosen

→ Subsequent j chosen from remaining species with no consumers and \( n_j < n_i \)

→ Subsequent j chosen randomly from species with \( n_j \geq n_i \)

Effect of rules:

→ Rules meant to mimic phylogenetic effects (guilds of related consumers)

→ Food webs are not “interval”

→ Hierarchy relaxed in principle, in practice rarely violated

Cattin et al. 2004
Generalized cascade model

Link distribution rules:

→ Species $i$ consumes resources species $j$ with $n_j \leq n_i$ with a probability equal to a random number with mean $2C$ drawn from a beta distribution

Effect of rules:

→ Create a simple, non-interval, beta-distributed hierarchical model that allows cannibalism

Stouffer et al. 2005
'Relaxed' niche models

**Link distribution rules:**

→ Same as niche model, but allow for gaps in a slightly expanded feeding range or for links external to feeding range

1. Generalized niche model (Stouffer et al. 2005)
2. Relaxed niche model (Williams & Martinez 2008)
3. Minimum potential niche model (Allesina et al. 2008)

**Effect of rules:**

→ Relax the intervalsity constraint of the niche model

*Stouffer et al. 2005, Williams & Martinez 2008, Allesina et al. 2008*
Random models

**Link distribution rules:**

→ Distribute links randomly
   1. Random model (Williams & Martinez 2000): \( P = C \)
   2. Random beta model (Dunne *et al.* 2008): beta distribution

**Effect of rules:**

→ Minimal constraints
   1. Random: no hierarchy, no intervality, no beta distribution
   2. Random beta: no hierarchy, no intervality

*Williams & Martinez 2000, Dunne *et al.* 2008*
### Summary of model constraints

<table>
<thead>
<tr>
<th>Model</th>
<th>beta distribution</th>
<th>intervality</th>
<th>hierarchical feeding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>hierarchy</td>
</tr>
<tr>
<td>Random</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Random beta</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Cascade</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Generalized cascade</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Niche</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Relaxed niche</td>
<td>yes</td>
<td>no*</td>
<td>yes</td>
</tr>
<tr>
<td>Nested hierarchy</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>
4. Confronting Models with Data
Inference methods

1. Degree distribution
2. Suite of properties
3. Likelihood
1) Degree distribution

Data Normalization & Analytical Analyses

<table>
<thead>
<tr>
<th>Property</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distribution of number of prey</td>
<td>$p_{\text{prey}}(k) = (1/2z)E_1(k/2z)$</td>
</tr>
<tr>
<td>Distribution of number of predators</td>
<td>$p_{\text{pred}}(m) = (1/2z)\gamma(m + 1, 2z)$</td>
</tr>
</tbody>
</table>
| Distribution of number of links | $p_{\text{links}}(r) = \int_0^r p_{\text{prey}}(t)p_{\text{pred}}(r-t)dt$  
                              | $\quad = 1/(2z)^2\Gamma_1(r/2z)\gamma(r-t+1, 2z)dt$          |
| Fraction of top species        | $T = (1 - e^{-z^2})/2z$                        |
| Fraction of basal species      | $B = \ln(1 + 2z)/2z$                           |
| Standard deviation of the vulnerability | $\sigma_V = \sqrt{1/3 + 1/z}$                |
| Standard deviation of the generality | $\sigma_G = \sqrt{8/(3 + 6C) - 1}$            |

Camacho et al. 2002, Stouffer et al. 2005
<table>
<thead>
<tr>
<th>Model</th>
<th>beta distribution</th>
<th>intervality</th>
<th>hierarchical feeding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>hierarchy exceptions</td>
</tr>
<tr>
<td>Random</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Cascade</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Generalized cascade</td>
<td>yes</td>
<td>no</td>
<td>yes ( n_j = n_i )</td>
</tr>
<tr>
<td>Niche</td>
<td>yes</td>
<td>yes</td>
<td>yes ( n_j \geq n_i )</td>
</tr>
<tr>
<td>Nested hierarchy</td>
<td>yes</td>
<td>no</td>
<td>yes ( n_j \geq n_i^* )</td>
</tr>
</tbody>
</table>
Beyond degree distribution...

- **Test**: against the structure of empirical food webs
- **Assess**: a suite of structural properties
- **Generate**: a set of 1000 model webs for each empirical web
- **Evaluate**: how well does the model perform?

  → Normalized Model Error = (empirical value - model mean) / (model median value - value at upper or lower 95% boundary of model distribution)
  
  → MEs ≤ |1| show 'good' fit of model mean to empirical value

Williams & Martinez 2000
Recent test (10 webs, 5 models)

Path Length:
- The average of the shortest chain of links between each pair of species.
- Most models significantly underestimate path length.

Williams & Martinez 2008
1) Mean $\text{ME} \leq |1|$ for all models: effect of hierarchy + beta distribution constraints

2) Niche: lowest ME mean & SD, most properties closest to 0, fewest properties $>|1|$

3) All models drastically underestimate herbivory/detritivory

Summary Stats

| ME | ME | % ME > |1| |
|----|----|--------|
| Gen Cas | -0.57 | 2.37 | 46% |
| Gen Nic | -0.50 | 1.40 | 39% |
| Nes Hier | -0.53 | 1.45 | 26% |
| Niche | -0.10 | 1.32 | 25% |
| Rel Nic | -0.40 | 1.58 | 33% |

Williams & Martinez 2008
### 'Suite of properties' summary

<table>
<thead>
<tr>
<th>Model</th>
<th>beta distribution</th>
<th>intervality</th>
<th>hierarchical feeding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>hierarchy</td>
</tr>
<tr>
<td>Random</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Random beta</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Cascade</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Generalized cascade</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Niche</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Relaxed niche</td>
<td>yes</td>
<td>no*</td>
<td>yes</td>
</tr>
<tr>
<td>Nested hierarchy</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>
3) Likelihood: topology as a whole

1) 3 models (Cascade, Niche, Nested hierarchy) and 10 datasets analyzed. All webs have links that violate assumptions of each model.

2) Use a genetic algorithm to order species to minimize violating links (Matrix $A \rightarrow A^*$).

3) Split datasets into compatible links (Matrix $N$) and incompatible links (Matrix $K$).

4) Calculate probability of obtaining Matrix $N$ with the model and Matrix $K$ with a random graph.

5) Product of those P’s gives a “total likelihood” (Tot $L$) of that model for that dataset.

Allesina et al. 2008
Alternate model: The minimum potential niche model

**Link distribution rules:**
- Same as niche model, but define a feeding range where the consumer has a probability of <1 of feeding on species in that range.

**Effect of link distribution rules:**
- Relax the intervality constraint of the niche model
- No empirical links are incompatible with the models
- Introduces an extra model parameter
Minimum potential (relaxed) niche model performs best:

→ no irreproducible links (Niche model has most)
→ marginally better Tot $\mathcal{L}$ than the Niche model on every dataset
→ much better Tot $\mathcal{L}$ than Nested hierarchy or Cascade models
## 'Likelihood' summary

<table>
<thead>
<tr>
<th>Model</th>
<th>beta distribution</th>
<th>intervality</th>
<th>hierarchical feeding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>hierarchy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>exceptions</td>
</tr>
<tr>
<td>Cascade</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>no</td>
</tr>
<tr>
<td>Niche</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$n_j \geq n_i$</td>
</tr>
<tr>
<td>Relaxed niche</td>
<td>yes</td>
<td>no*</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$n_j \geq n_i$</td>
</tr>
<tr>
<td>Nested hierarchy</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$n_j \geq n_i^*$</td>
</tr>
</tbody>
</table>
Pros and cons of inference approaches

1) **Degree Distributions**
   - **Pros:** Characterizes a central tendency of structure
   - **Cons:** Very minimal notion of “structure”

2) **Suite of Properties**
   - **Pros:** Allows assessment of details of how/why structure differs
   - **Cons:** Properties are not independent, making overall evaluation problematic

3) **Likelihood**
   - **Pros:** Based on full structure of network
   - **Cons:** How to understand details of how/why structure differs?
     How to interpret magnitude of differences in Tot $L$?
Apparent Complexity

- Marine
- Estuary
- Lake
- Rainforest
- Desert
Underlying Simplicity

Two Parameters \((C,S)\)

Simple Link Distribution Rules

Successful Prediction of Network Structure
'Complex' food webs aren't intractably complex: underlying shared scale-dependent structure.

MaxEnt provides a simple null model for a central feature of data and models, the degree distribution.

The Niche model and its variants, but not the Cascade model, do a good job of predicting many aspects of empirical web structure.

- Hierarchical Feeding + Beta/Exponential Distribution

The Niche and Relaxed Niche models fit data better than other variants.

- Intervality + Cycles

Evaluation of models largely indifferent to inference method.
5. Case Study: Ancient Food Webs
Are species interactions structured differently in ancient versus modern ecosystems? (can we even put convincing data together?)

- Have food webs become more complex since the beginning of the Phanerozoic?
- What do differences/similarities in ecological network structure suggest about fundamental constraints on species interactions?
- Does food-web complexity or structure change across extinction boundaries?
- Do major evolutionary innovations ramify throughout food webs?
- How does community structure respond to major environmental perturbations?
**Geologic Time Scale**

<table>
<thead>
<tr>
<th>PHANEROZOIC</th>
<th>MESOZOIC</th>
<th>PALEOZOIC</th>
<th>PRECAMBRIAN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TERTIARY</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PALEOGENE</td>
<td>23.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NEOGENE</td>
<td>1.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QUATERNARY</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HOLOCENE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PLEISTOCENE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PLIOCENE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MIocene</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>OLIGOCENE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>EOCENE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PALEOGENE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PALEOZOIC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEVONIAN</td>
<td>355</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SILURIAN</td>
<td>418</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORDOVICIAN</td>
<td>441</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAMBRIAN</td>
<td>490</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MESOZOIC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRIASSIC</td>
<td>200</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JURASSIC</td>
<td>144.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRETACEOUS</td>
<td>65</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burgess Shale (505 Ma)</td>
<td>Chengjiang Shale (520 Ma)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Lagerstätten: Fossil assemblages with exceptional soft-tissue preservation
Burgess Shale Biota

Wiwaxia

Waptia

Marella

Hallucigenia

Opabinia

Anomalocaris

Ollenoides

Pikaia

Ottoia
Lines of evidence for feeding interactions

Every link is a hypothesis based on inferences

- Gut contents
- Body size
- By analogy with associated taxa
- Damage patterns
- Environmental deposition
- Functional morphology
- Stable isotopes
- Trace fossils
- Coprolites
- The occasional smoking gun...

**Certainty:**
1 = possible
2 = probable
3 = certain
Functional morphology + damage patterns

Image by Ken Dowd (via NSF): *Anomalocaris canadensis* hunting trilobites
Burgess Shale Food Web

\[ S = 85, \ L = 559, \ L/S = 6.6, \ C = 0.08, \ TL = 2.99 \]
Normalized link distributions

- Chengjiang
- Burgess
- Bridge Brook
- Skipwith
- Benguela
- Coachella
- Chesapeake
- St. Martin
- St. Mark's
- Reef

cumulative distribution

# trophic links/2(L/S)
Niche model errors

Model Errors for 17 properties for 10 food webs

Results generally robust to removal of uncertain links
A few potentially meaningful differences?

- **Higher LinkSD in both Cambrian webs**
  - Reflects higher vulnerability to predation

- **Longer Path lengths in Chengjiang web**
  - Reflects lower integration among taxa

- **More taxa in Loops in Chengjiang web**
  - Reflects less hierarchical trophic organization

*C = Chengjiang shale*
*B = Burgess shale*
*M = 8 modern Webs*
Detailed species interaction data compiled for ancient ecosystems from the early Phanerozoic (> 500 MA).

The structure of Cambrian & modern webs is very similar and well-predicted by the niche model.

Results are robust to removal of uncertain or random links.

Differences in Cambrian structure may reflect a rapid transition to more stable, constrained, hierarchical, integrated, trophic organization following the Cambrian “explosion” of diversity, body plans, and trophic roles.

Shared architecture across habitats and deep time is suggestive of constraints on trophic organization.

- null: MaxEnt
- thermodynamic constraints
- dynamical stability of complex systems
- evolutionary processes

Dunne et al. 2008


Reviews:
