Microsoft Research

Neutral Degree Distributions in Complex Food Webs

Richard Williams

Head, Computational Ecology and Environmental Science Group

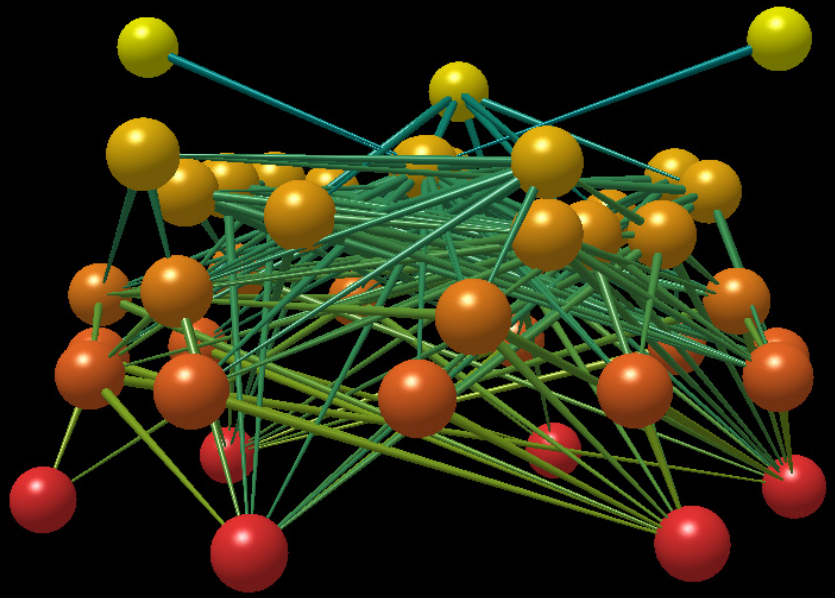
Microsoft Research, Cambridge

Degree Distributions in Complex Networks

- Degree distribution is the distribution of fraction of nodes in a network with a particular number of links.
- It is useful to have a null model driven by simple statistical considerations and to explain deviations from this null model.
- Erdős and Rényi random graphs have been widely used as a null model for complex networks.
- Random graphs have a strongly-peaked binomial degree distribution, but empirical networks are usually broadly distributed, with long right-tails that are scale-free or exponential.

Food Webs

- Directed network of predator-prey interactions
- Simplest food webs record "only" presence and absence of species and feeding interactions
- Provide a tractable overview of ecosystem structure
- Networks are quite small (O(100) nodes) but are densely connected ($C = L/S^2 = O(0.1)$)



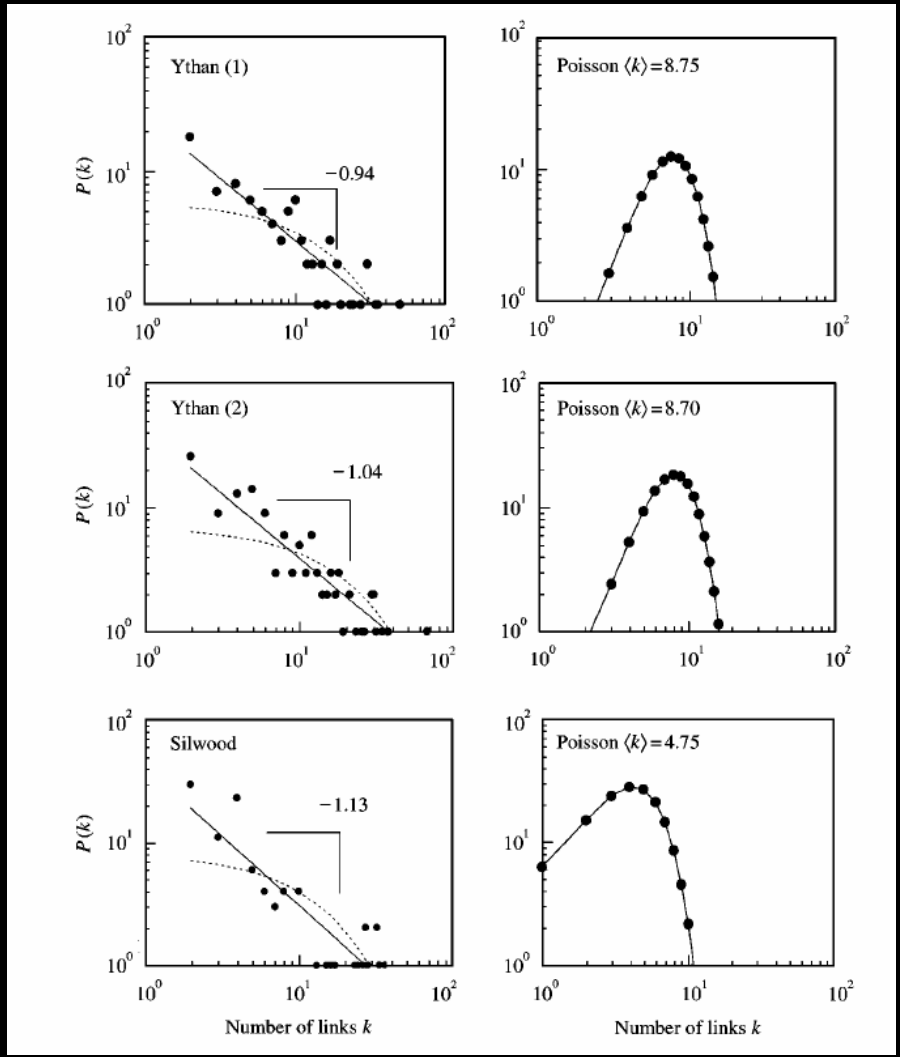
Food web from St Martin Island

Degree Distributions in Food Webs

- Montoya and Sole (2002): found long tails "indicative of power law scaling" in three food webs
- Camacho et al (2002): "there is a characteristic scale for the linkage density, i.e., food webs do *not have a* scale-free structure" in seven food webs (some of the same data sets)
- Dunne et al (2002): "while some food webs have scale-free structure, most do not if they exceed a relatively low level of connectance" in 16 food webs

Degree Distributions in Food Webs

- Montoya and Sole, JTB 2002: a “surprising feature” of the distributions is that they show strongly non-Poissonian behavior.



Reconsidering Null Models for Degree Distributions

- The equal link probability E&R model imposes unbiological and unnecessary constraints
- A Maximum Entropy Distribution (MaxEnt) gives the least biased probability distribution which satisfies a set of information containing constraints (Jaynes, 1957)

- The ecological constraints on the numbers of consumers and resources are probably different

A Null Model for Degree Distributions

- Number of basal and top species is highly variable and possibly highly methodology dependent.
- Only knowledge used is the number of nodes (S) and the number of links (L), the number of top species (T) and the number of basal species (B)

- Consumer distribution: L links distributed across (S – T) resource species with possible values of {1,…,S} consumers per species
- Resource distribution: L links distributed across (S – B) consumer species with possible values of {1,…,S} resources per species

A Null Model for Degree Distributions

- In both cases, require a discrete distribution on the set of values $\{x_1,\ldots,x_n\}$ with known mean μ
- Consumer Distr: values $\{1,\ldots S\}$, $\mu = L/(S - T)$
- Resource Distr: values $\{1,\ldots S\}$, $\mu = L/(S - B)$

- MaxEnt distribution is $P_i = P(X = x_i) = Ce^{\lambda x_i}$ for $i = 1,\ldots,n$. The constants C and λ are determined by the requirements that the probabilities sum to 1 and have mean μ:

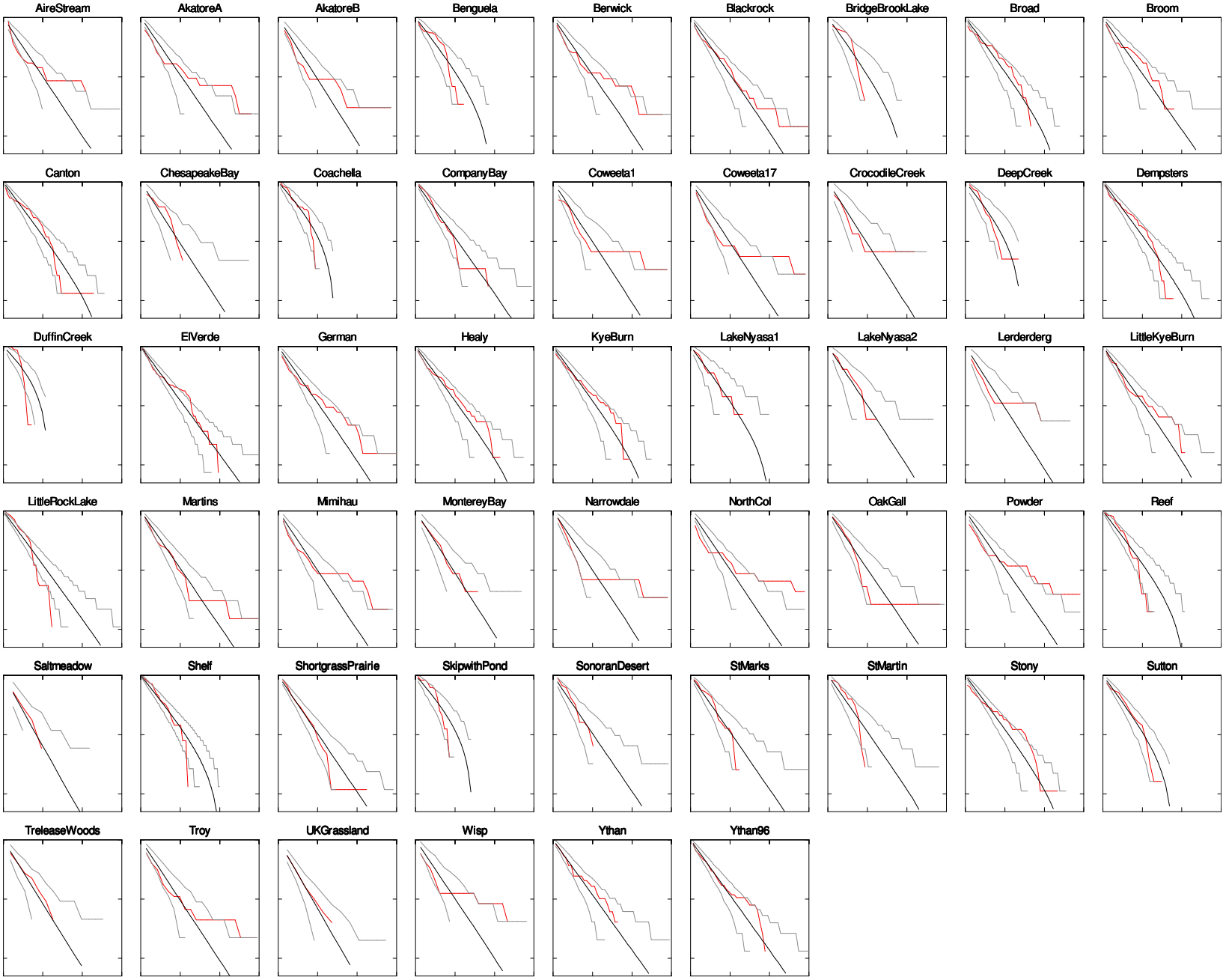
$$\sum_i P_i = 1 \text{ and } \sum_i x_i P_i = \mu$$

An Expanded Set of Food Webs

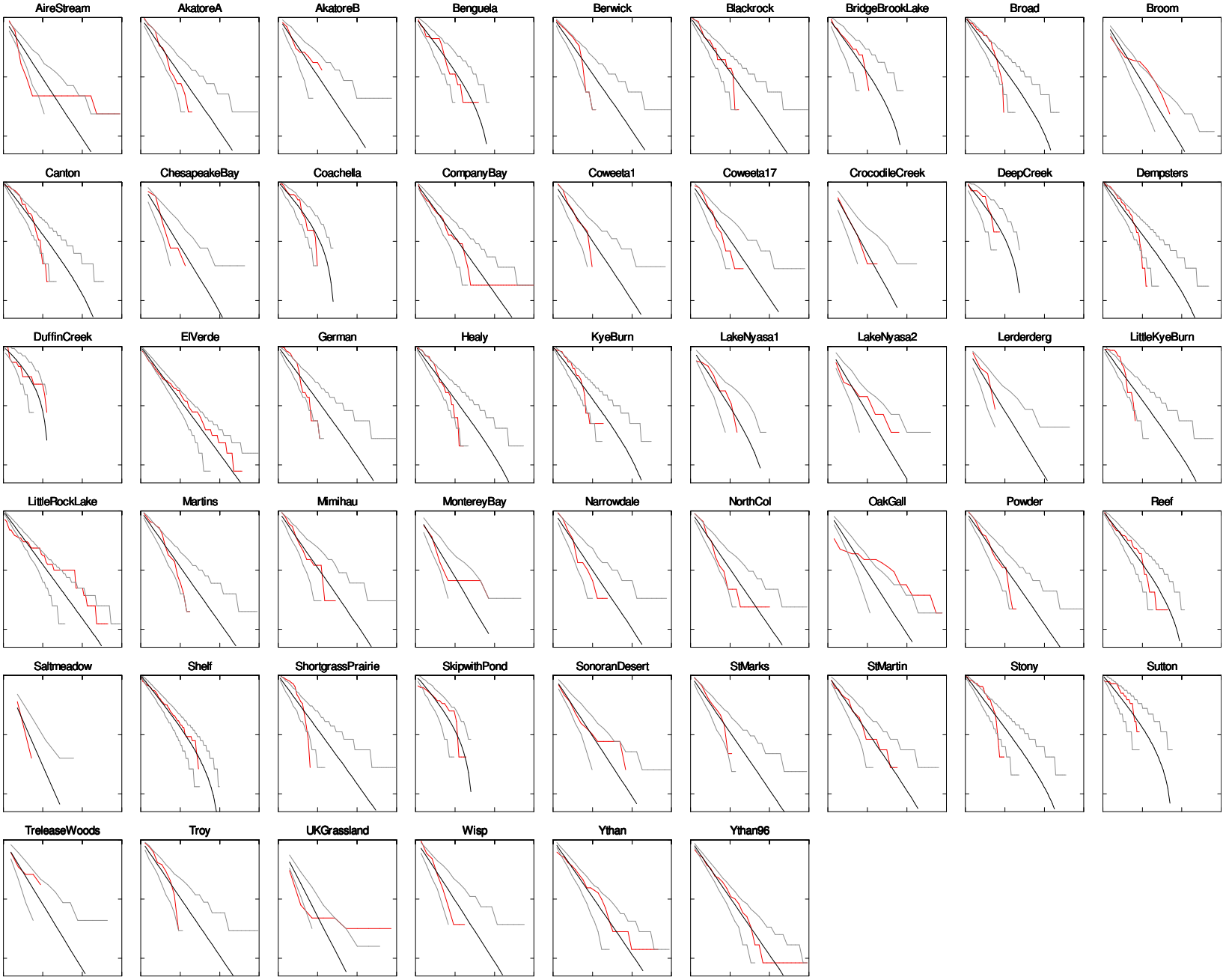
- 51 food webs with 25 or more trophic (functionally distinct) species.
- Thanks to Ross Thompson and Jen Dunne for much of the work in pulling these data together from various researchers.
- Thanks to everyone for their willingness to share these hard-won data sets.
- There are still many problems in these data – inconsistent taxonomic resolution/aggregation, inconsistent methodologies, non-independence…

- Press on despite these problems.

Cumulative Consumer Distributions



Cumulative Resource Distributions



Testing Goodness of Fit

- Previous studies of food web degree distribution have relied on visual comparison of empirical distribution and some functional form.

- Use likelihood ratio to measure difference between degree distributions: $$G = 2\sum_i O_i \ln(O_i / E_i)$$

where O_i is the observed frequency and E_i is the expected (MaxEnt) frequency.

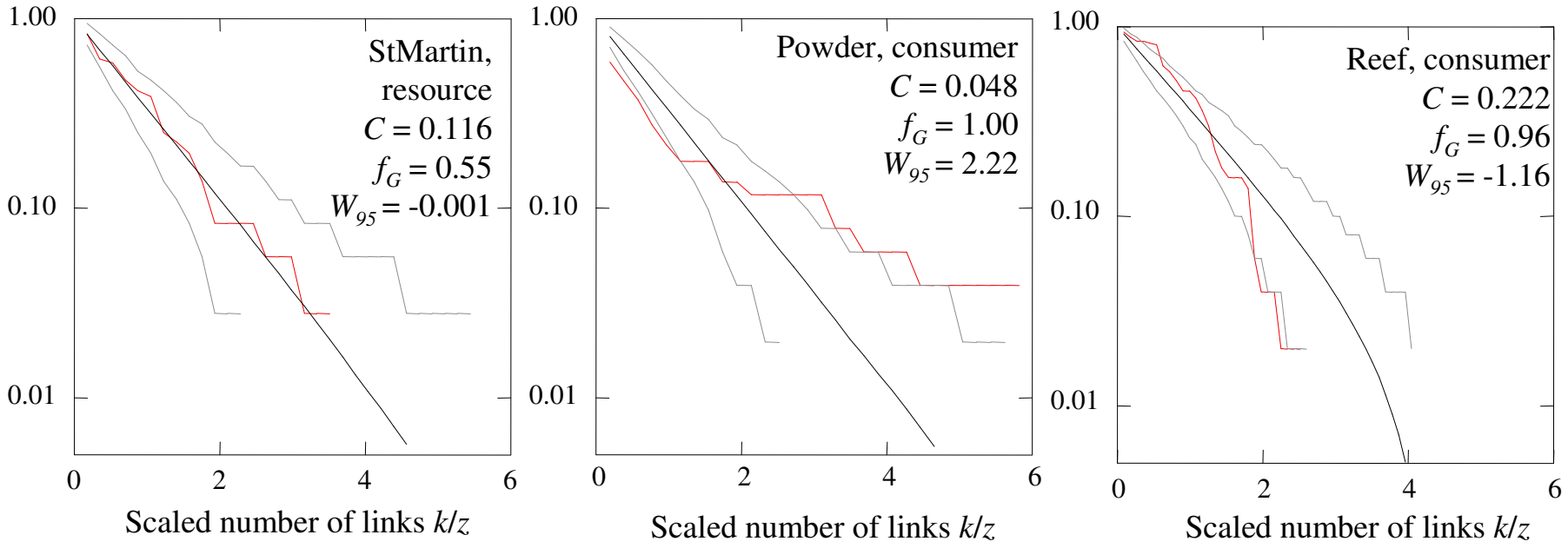
Testing Goodness of Fit

- Do randomization test due to sparseness of the distributions
  - Draw a data set from the MaxEnt distribution
  - Compare G of drawn data to G of empirical data set
  - Record the fraction of trials in which the G value of the empirical data is greater than the G value of the data drawn from the maximum entropy distribution, f_G.
  - Empirical distribution is significantly different from the maximum entropy distribution if $f_G > 0.95$

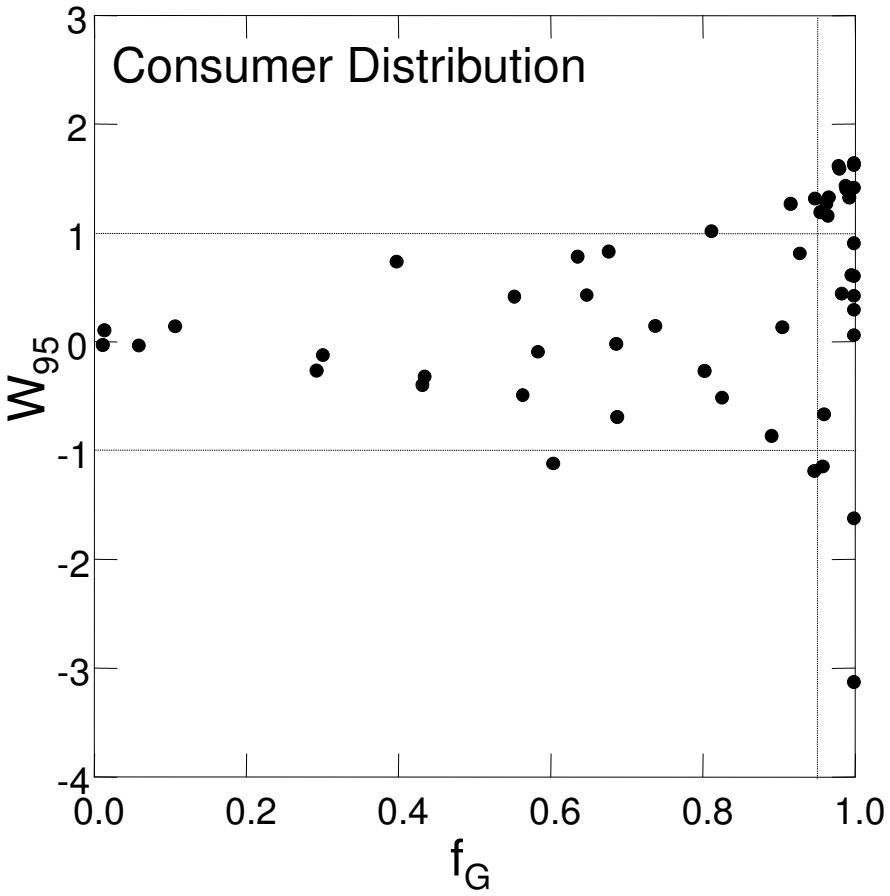
Testing Goodness of Fit

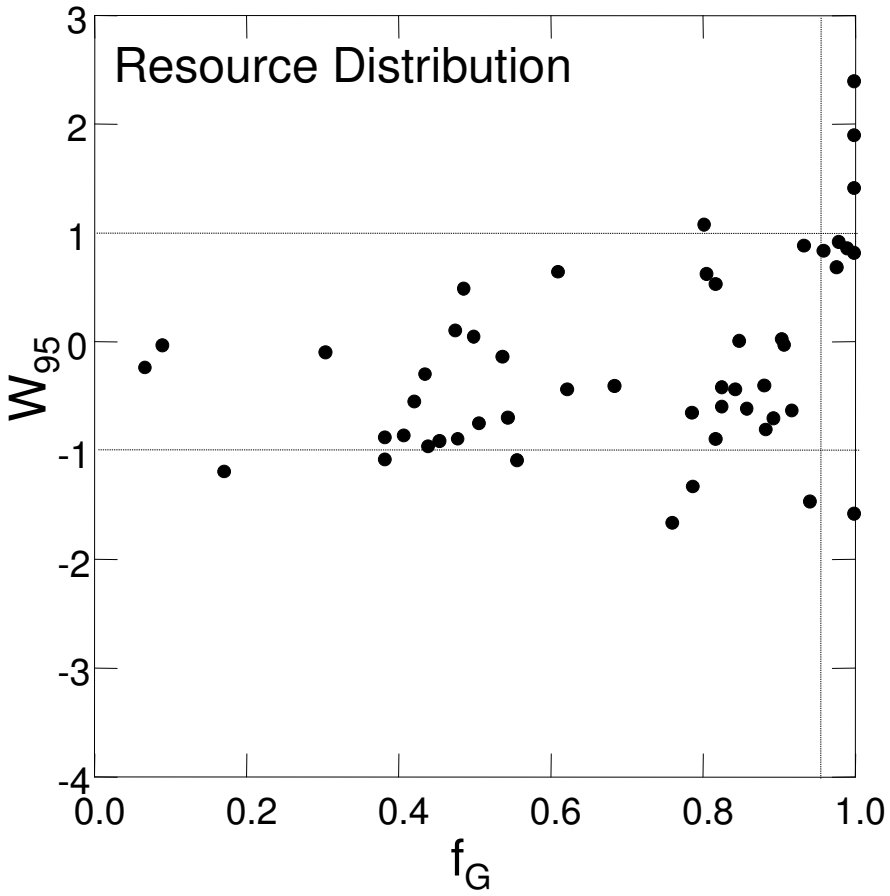
- This measure does not distinguish the relative wide of the empirical distribution, whether it is more broadly or narrowly distributed than the MaxEnt model.
- Use distribution standard deviation as a simple measure of relative width
  - Draw 10000 data sets from the MaxEnt distribution
  - Compute distribution of $\log(\sigma_{ME})$
  - Compute relative distance of $\log(\sigma_E)$ from median, normalized by width of 95% central interval
  - $W_{95} < -1$ and $W_{95} > 1$ are significantly narrower or broader than Maxent distribution respectively

Degree Distribution Examples



Empirical Web Goodness of Fit





Empirical Web Goodness of Fit

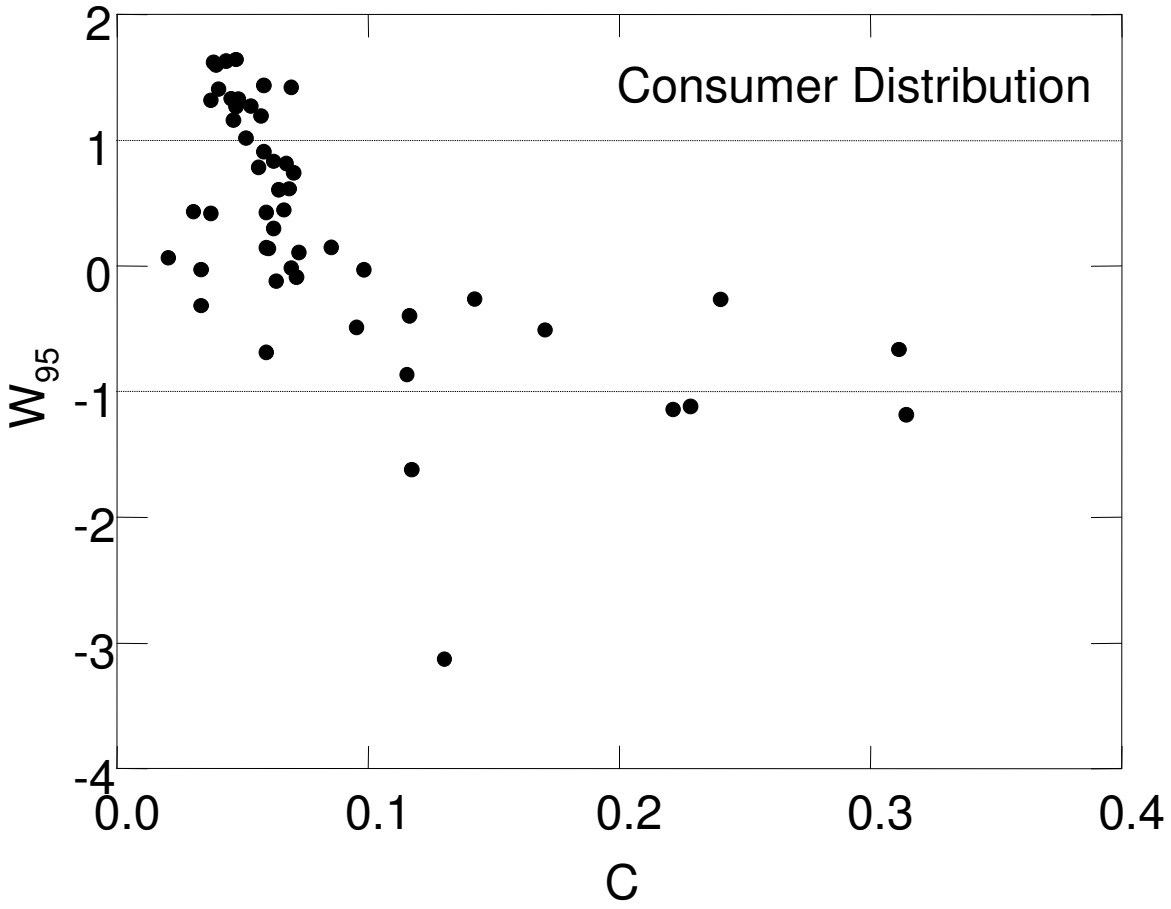
Number and (Fraction) of webs whose degree distributions are not significantly different from the model distribution

Criteria	Consumer Distr	Resource Distr
$f_G < 0.95$	28 (0.55)	42 (0.82)
$W_{95} > -1$ and $W_{95} < 1$	31 (0.61)	40 (0.78)
$W_{95} > -1$, $W_{95} < 1$ and $f_G < 0.95$	23 (0.45)	35 (0.69)
$f_G < 0.99$	39 (0.76)	47 (0.92)
Binomial $f_G < 0.99$	1 (0.02)	4 (0.08)

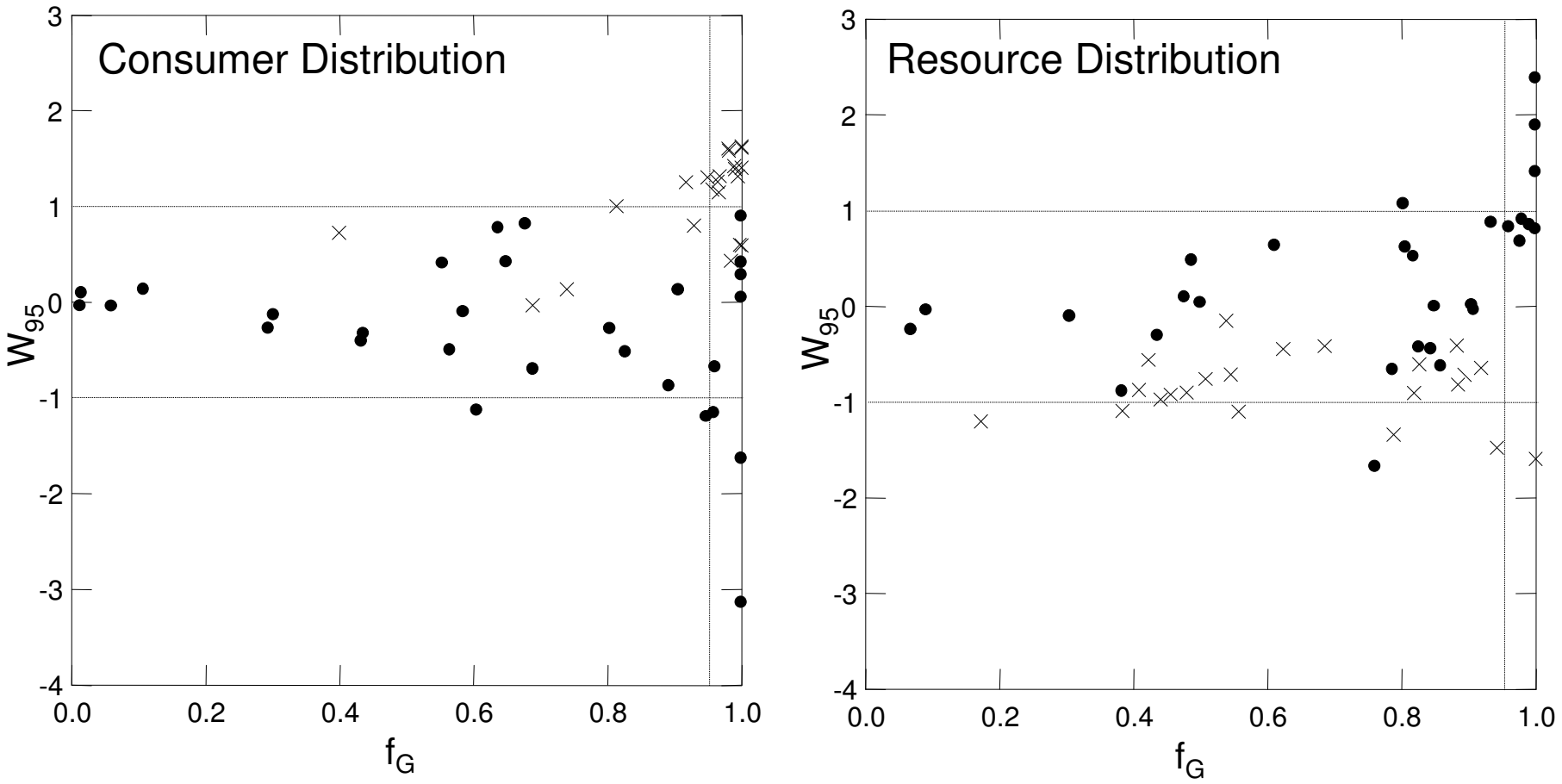
- MaxEnt model performs much better than ER random model
- While both distributions are surprisingly well-fit, there is considerable asymmetry in the goodness of fit

Scale and Complexity Dependence

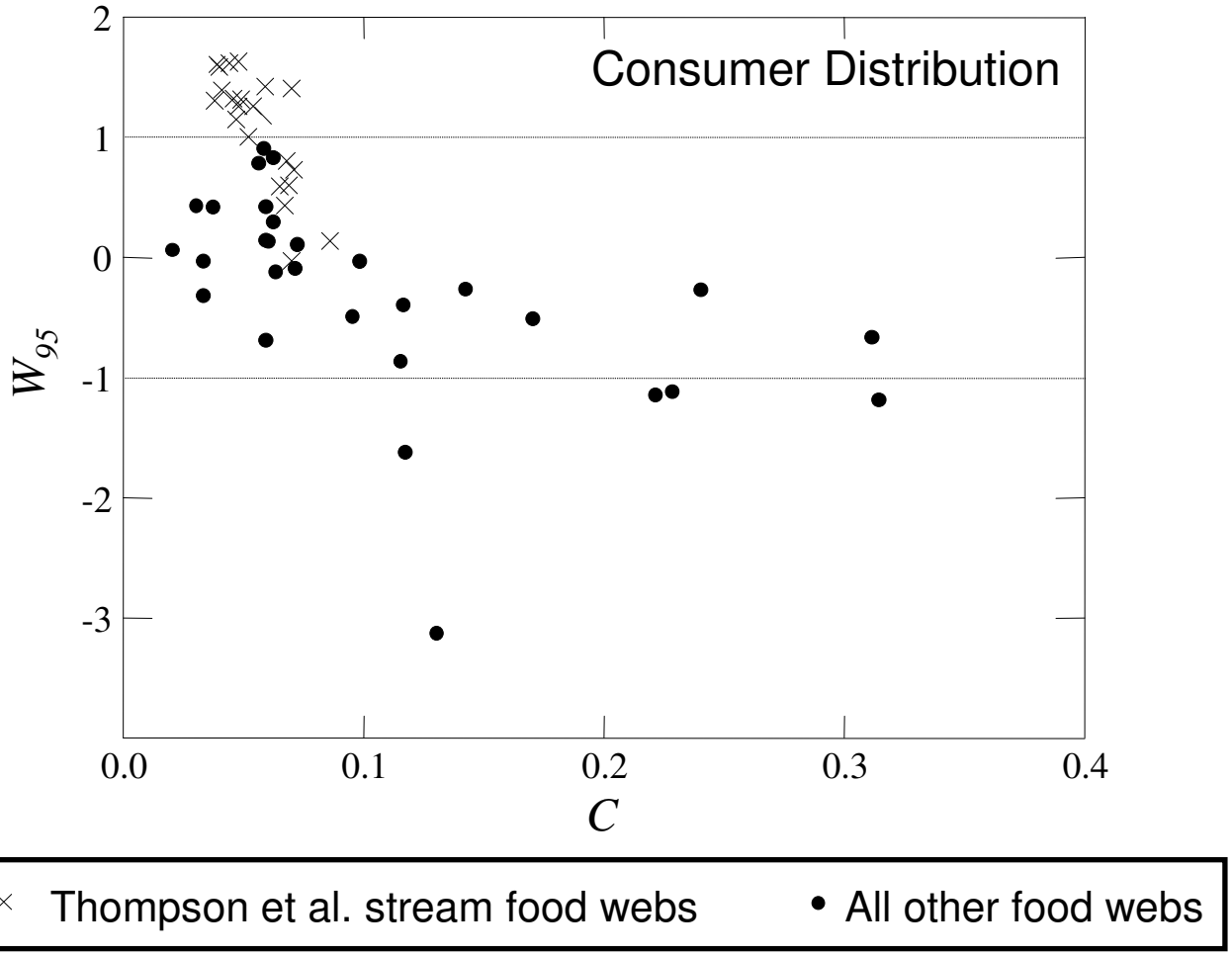
- No significant variation with S or C except between relative width W_{95} and connectivity C for the consumer distribution.



Methodological Bias?



Methodological Bias?



Key Points: Food Webs

- The relative abundance of specialists and generalists can largely be predicted by a simple statistical model rather than a mechanistic ecological model.
- The considerable asymmetry between the consumer and resource distributions requires explanation, either ecological or methodological.
- This work pushes the limits of the current data; empiricists are busy creating a new generation of more highly resolved and evenly sampled data sets.

Key Points: Network Theory

- Network theory traditionally assumes that degree distributions different from the distribution of E&R random webs require explanation.
- Given information only about the number of nodes and edges, a less constrained MaxEnt distribution is a more logical null model
- Some broad-scale empirical degree distributions are well-described by this simple null model, and deviations are typically much smaller than deviations from the E&R random model.

Thanks!