Correction

ECOLOGY, STATISTICS

Correction for "A unifying approach for food webs, phylogeny, social networks, and statistics," by Grace S. Chiu and Anton H. Westveld, which appeared in issue 38, September 20, 2011, of *Proc Natl Acad Sci USA* (108:15881–15886; first published September 6, 2011; 10.1073/pnas.1015359108).

The authors note that the data for three of the eight datasets analyzed in the article (Skipwith, St. Martin, and Ythan) were incorrectly processed. As a result, Fig. 2, Table 1, and Table 2 appeared incorrectly. The corrected figure, its corrected corresponding legend, and the corrected tables appear below. These errors do not affect the conclusions of the article.

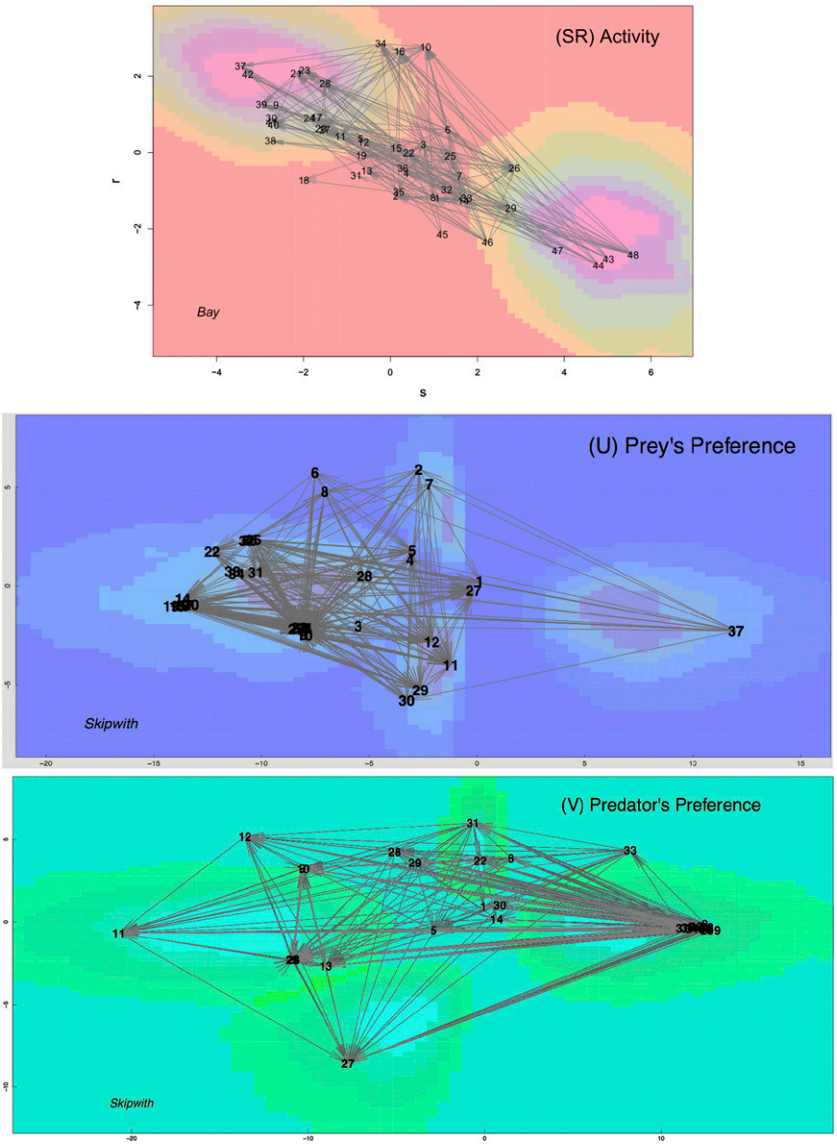

Fig. 2. Food web graphs displaying trophic structure from fitting model **1** with phylogeny measure x_{ij} as the predictor. (SR) Feeding activity in *Bay*. The *s* axis refers to activity as prey, and the *r* axis, as predator. Label of node *i* is located at the mean of the bivariate posterior distribution (an "estimate") of $[s_i, r_i]$; this distribution describes the probability of the position of $[s_i, r_i]$ on the *sr* plane, given the knowledge of the observed food web. Distribution density appears as heat map for benthos-eating birds ("37") and detritus ("48"). Legend for node labels appears in *SI Text*. (U, V) Preference of being consumed/consuming in *Skipwith*. They are similarly interpreted as (SR), but referring to u_i vectors for small oligochaetes ("2"), *C. praeusta* ("14"), *L. marmoratus* ("30"), and detritus ("37"), and v_i vectors for *A. juncea* ("11"), *A. germari* ("19"), great diving beetle ("27"), and *P. sagittalis* ("31"). Nodes far apart in the latent *u* or *v* space differ substantially with respect to feeding preference.

Table 1. Bayesian inference numerical summaries for selected food webs from fitting statistical social network model 1, with phylogenetic similarity x_{ij} as the predictor

			Credible interval		
Food web*	Parameter	Posterior median†	Lower limit	Upper limit	Credibility‡
Bay	β_1	5.10	1.20	10.46	0.95
	ρ_{sr}	−0.46	−0.75	−0.07	0.95
	ρ	−0.82	−0.95	−0.11	0.75
Reef	β_1	2.16	0.06	4.21	0.90
	ρ_{sr}	−0.08	(interval includes 0)		0.50
	ρ	−0.29	−0.52	−0.02	0.60
Skipwith	β_1	13.08	(interval includes 0)		0.50
	ρ_{sr}	−0.94	−0.99	−0.11	0.85
	ρ	−0.97	−0.99	−0.67	0.99
St. Martin	β_1	−2.05	−4.23	−0.10	0.60
	ρ_{sr}	−0.34	−0.61	−0.02	0.85
	ρ	−0.30	(interval includes 0)		0.50

*Other food webs and z_{ij} appear in Table S2.
†The posterior median can be considered a parameter "estimate."
‡Credible intervals presented have approximately the highest credibility without including 0. High credibility for an interval excluding 0 indicates statistical importance of the corresponding parameter to feeding potential (p_{ij}).

Table 2. Goodness-of-fit summaries for selected food webs (others in Table S2)

Model	Predictor	GoF*	Model	Predictor	GoF
	Bay			*Reef*	
1	x_{ij}	367	**1**	x_{ij}	524
1	z_{ij}	386	**1**	z_{ij}	526
1	—	394	**1**	—	533
naive†	x_{ij}	719	naive	x_{ij}	1273
naive	z_{ij}	719	naive	z_{ij}	1287
	Skipwith			*St. Martin*	
1	x_{ij}	42	**1**	x_{ij}	286
1	z_{ij}	36	**1**	z_{ij}	275
1	—	33	**1**	—	286
naive	x_{ij}	734	naive	x_{ij}	679
naive	z_{ij}	737	naive	z_{ij}	678

*Derived from the Bayes factor on the model's ability to predict the act of feeding (y_{ij}). When comparing between models, noticeably smaller GoF values suggest better fit.
†Simple logistic regression ignoring network dependence—i.e., naively setting $s_i + r_j + u_i' v_j + \varepsilon_{ij} = 0$ for all i,j in model **1**.

www.pnas.org/cgi/doi/10.1073/pnas.1212275109

CORRECTION

A unifying approach for food webs, phylogeny, social networks, and statistics

Grace S. Chiu[a,1] and Anton H. Westveld[b]

[a]Commonwealth Scientific and Industrial Research Organisation (CSIRO) Mathematics, Informatics and Statistics, GPO Box 664, Canberra, ACT 2601, Australia; and [b]Department of Mathematical Sciences, University of Nevada, Las Vegas, NV 89154

Edited* by Adrian Raftery, University of Washington, Seattle, WA, and approved July 7, 2011 (received for review October 13, 2010)

A food web consists of nodes, each consisting of one or more species. The role of each node as predator or prey determines the trophic relations that weave the web. Much effort in trophic food web research is given to understand the connectivity structure, or the nature and degree of dependence among nodes. Social network analysis (SNA) techniques—quantitative methods commonly used in the social sciences to understand network relational structure—have been used for this purpose, although postanalysis effort or biological theory is still required to determine what natural factors contribute to the feeding behavior. Thus, a conventional SNA alone provides limited insight into trophic structure. Here we show that by using novel statistical modeling methodologies to express network links as the random response of within- and internode characteristics (predictors), we gain a much deeper understanding of food web structure and its contributing factors through a unified statistical SNA. We do so for eight empirical food webs: Phylogeny is shown to have nontrivial influence on trophic relations in many webs, and for each web trophic clustering based on feeding activity and on feeding preference can differ substantially. These and other conclusions about network features are purely empirical, based entirely on observed network attributes while accounting for biological information built directly into the model. Thus, statistical SNA techniques, through statistical inference for feeding activity and preference, provide an alternative perspective of trophic clustering to yield comprehensive insight into food web structure.

Bayesian hierarchical modeling | food web connectance | latent space models | network data | Procrustes problem

Fig. 1 depicts a simple food web. Food webs are network structures consisting of nodes, each containing one (e.g., Human) or various species (e.g., Ticks). For a trophic food web, nodes are interwoven by directed links that conventionally point from prey to predator (1). Certain patterns among trophic relations suggest the clustering of nodes; the ability to unveil these patterns can facilitate other aspects of food web research, such as the identification of functional groups, trophic levels, and keystone species. This in turn can provide information about the stability of the web under perturbations (e.g., species extinction). Conventional social network analysis (SNA) techniques have been applied to trophic research for this purpose (e.g., refs. 2 and 3). These methods typically seek optimal partitioning of the network into compartments of nodes subject to prespecified mathematical criteria (4, 5). After food web features have been identified, a natural question follows: What factors contributed to the feeding behavior among nodes that gave rise to those features? Conventional SNA frameworks provide no direct means to address this question.

Recent advancement in statistical regression methodologies that model complex network structures (6–9) has allowed researchers, mostly from the social sciences, to unravel valuable information entwined in the relational links observed empirically among nodes. Standard regression regarding nodes as independent has been used to connect various within-node characteristics in a food web (1): Because predator-prey links were not part of the regression, they were used post hoc to qualitatively explain the relationship among nodal characteristics. Instead, a key feature of the more novel statistical SNA methods is one's ability to utilize network dependency and explicitly express predator-prey links as a regression function of both within- and internode characteristics—e.g., biomass of the node, the role of the node as predator or prey, and phylogenetic similarity between nodes. This allows direct inference for what makes a given node a predator or prey and for dependence features including the tendency for predator-prey role reversal between a given pair of nodes. A fundamental principle of regression modeling is the appropriate use of available predictor variables (here, nodal characteristics) and the dependency among data to improve the accuracy and precision of inference drawn for the behavior of the response variable (here, the feeding links, which exhibit complex dependencies). This type of statistical modeling of trophic (feeding) relations is not to be confused with that in ref. 10, which uses compartment membership to explain between-node similarity, nor with that in ref. 11, which employs Bayesian melding (12) to model intercompartmental energy-matter flows subject to mass balance.

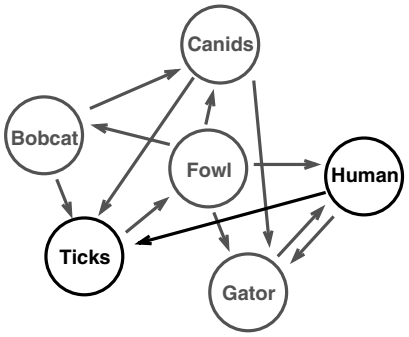

Fig. 1. Pictorial representation of feeding links pointing from prey to predator. Nodes (hypothetical) are shown in an arbitrary arrangement.

We apply an existing statistical SNA framework (13) known as latent space modeling (7, 8) to analyze eight empirical trophic food webs that have been previously studied (14). These webs were observed from Goose Creek Bay in the St. Marks National Wildlife Refuge in the southeastern United States (15), the Benguela marine ecosystem off the South African coast (16), the grasslands in England and Wales (17), the Caribbean coral reefs with a reduced set of nodes (18), the northeast US continental shelf (19), a pond on Skipwith Common in England (20), the

Author contributions: G.S.C. and A.H.W. designed research; G.S.C. and A.H.W. performed research; G.S.C. and A.H.W. contributed new analytic tools; G.S.C. and A.H.W. analyzed data; G.S.C. conceived methodological application; A.H.W. provided editorial input to drafts written by G.S.C.; and G.S.C. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]To whom correspondence should be addressed. E-mail: grace.chiu@csiro.au.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1015359108/-/DCSupplemental.

Caribbean island of St. Martin (21), and Ythan Estuary in Scotland (22). Henceforth, we refer to them respectively as *Bay*, *Benguela*, *Grass*, *Reef*, *Shelf*, *Skipwith*, *St. Martin*, and *Ythan*. These webs correspond to various types of aquatic and terrestrial communities, through which we demonstrate the practicality of statistical SNA in a rather general context of food web ecology. Our analyses illustrate the new perspectives of trophic patterns, presented through three configurations of the food web graph, according to the statistical inference for node activity and feeding preference. We also demonstrate that the statistical framework can provide rigorous quantitative evidence for the contribution of phylogenetic information to predator-prey relations. We discuss how our findings reflect the recent debate over the mean trophic level (MTL) in ref. 23.

Modeling Framework

For each food web, associated with any pair of nodes is a numerical value y_{ij} representing trophic linkage in the form of sending activity, or the activity of being consumed, directed from the ith node to the jth. The dataset takes on one of two forms. The first is presence-absence data, where $y_{ij} = 1$ if the link $i \to j$ is present, and $y_{ij} = 0$ otherwise. For example, in Fig. 1, Ticks are observed to predate on Human, but the converse is not true, and thus $y_{\mathrm{HT}} = 1$ and $y_{\mathrm{TH}} = 0$. The other form is weighted data, where y_{ij} is the magnitude or weight of some consumption measure (e.g., volume) for the predation of i by j (24). The observed y_{ij}s can be displayed in a square matrix called the diet matrix.

Not all eight food webs being analyzed are associated with weights. Thus, we only consider the presence and absence of predation between pairs, using logistic regression. That is, the odds $p_{ij}/(1-p_{ij})$ represent the underlying predation behavior of j on i, where p_{ij} is the probability of the event $\{y_{ij} = 1\}$. For predicting predation, taxonomy is the sole nonfeeding information that is readily available for all eight of our food webs. We define two measures of phylogenetic similarity, x_{ij} and z_{ij}, to quantify the taxonomic likeness between i and j (*Materials and Methods*). Then, to describe internodal relations, we consider the mixed-effects model

$$\log\frac{p_{ij}}{1-p_{ij}} = \mu_{ij} + s_i + r_j + u_i'v_j + \varepsilon_{ij}, \qquad i \neq j \qquad [1]$$

where the log-odds is expressed as a fixed mean μ_{ij} ($=\beta_0 + \beta_1 x_{ij}$, $\beta_0 + \beta_1 z_{ij}$, or simply β_0), plus random deviation from the mean. The total random deviation is decomposed into four mean-zero components: s_i due to i in the role of the sender, r_j due to j in the role of the receiver, inner product $u_i'v_j$ due to the interaction between i and j, and ε_{ij}, which is the remainder not attributable to the former three components. For example, Human's activity level as prey and as predator is represented by s_{H} and r_{H}, respectively. For Human and Ticks, u_{H} (v_{H}) being close to u_{T} (v_{T}) in the latent two-dimensional u space (v space) would indicate that Human and Ticks are similarly preferred as prey by other nodes (have similar preference for prey nodes). Network dependence not addressed by parameters in Eq. **1** is modeled through $\rho_{sr} = \mathrm{Correlation}(s_i, r_i)$ for all i, and through $\rho = \mathrm{Correlation}(\varepsilon_{ij}, \varepsilon_{ji})$ for all $i \neq j$. Phylogenetic similarity is relevant to feeding when the regression coefficient $\beta_1 \neq 0$. Another way to view the influence of this predictor is the points of reference it provides when interpreting model parameters. For example, without predictors in Eq. **1**, having $s_i > s_j$ is equivalent to i being consumed by more nodes in the food web than j. With x_{ij} or z_{ij}, the interpretation changes: i is more actively consumed than j when (i,j) is compared against those pairs of nodes sharing the same phylogenetic similarity. See *Materials and Methods* for detailed interpretation of all model parameters.

"The largest gains in estimating regression coefficients often come from specifying structure in the model" (25). To demonstrate the informational gain in the food web inference from specifying network structure, we compare model goodness-of-fit between model **1** and simple logistic regression, which naively ignores all dependence among y_{ij}s. Model **1** and its complex correlation structure can be expressed in a Bayesian hierarchical framework (13), which was implemented as "gbme.asym.r" (http://www.stat.washington.edu/hoff/Code/GBME/) to extend earlier models in ref. 7. We employed this software to perform Bayesian statistical inference (25) for all model parameters. Invariance of $u_i'v_j$ under orthogonal transformation of u_i or v_j prompted us to work out a suitable Procrustes transformation of u_i and v_j so that their estimates produced by "gbme.asym.r" were interpretable (7, 8). See *Materials and Methods* for details.

Results

Visual Representation of Trophic Features. We first summarize the statistical inference by three graphs in Fig. 2. Unlike that of Fig. 1, the arrangement of nodes in the graphs labeled SR (s_i vs. r_i), U (u_i vectors), and V (v_i vectors) is due to the fitting of model **1**. Respectively, the graphs reflect connectivity structure from the perspectives of sender-receiver activity, sending preference, and receiving preference. The graphs allow immediate visual assessment of predator-prey connectedness in the web, and the extent of trophic clustering (tight or loose clusters, and how many). For *Bay*, SR shows a fair number of feeding links, although not particularly dense. It also suggests roughly four clusters, comprising nodes that are (*i*) most actively consumed but least active as consumers (*Halodule wrightii*: "43," micro epiphytes: "44," phytoplankton: "47," detritus: "48"), (*ii*) the most active consumers but average on the scale of being consumed (ominivorous crabs: "10," predatory shrimps: "16," predatory worms: "34"), (*iii*) least actively consumed but are very active consumers (benthos-eating birds: "37," herbivorous ducks: "42"), and (*iv*) the rest of the web, possibly divided further depending on the clustering resolution. Clusters *i* to *iii* appear tighter than *iv*, but the majority of nodes are evenly scattered. Thus, from the perspective of feeding activity, some trophic levels (TLs) are not clearly distinguishable from each other. For *Skipwith*, U and V each shows a very tight cluster made up of numerous nodes, with the remaining nodes forming isolated small clusters. Thus, overall, some TLs from the perspective of feeding preference as prey or as predator can be clearly distinguished. For node- or cluster-specific insight, we see that the large clusters in U and V are made up of different sets of nodes; thus, trophic clusters depend on the perspective of feeding behavior from which clustering is viewed. In U, detritus ("37") is greatly isolated from the rest of the web—i.e., consumers of detritus differ substantially from those of any other node in the web. Particularly, consumers of *Notonecta glauca* ("13"), *Hydroporus erythrocephalus* ("22"), and *Sialis lutaria* ("28") differ the most. Indeed, network links or arrows that originate from "37" land in very different parts of the web compared to arrows originating from "13," "22," or "28." Similarly, the consumers of *Chydorus latus* ("7"), *Corynoneura scutellata* ("32"), and *Tanytarsus bruchonius* ("35") differ the most from those of *Lumbriculus variegatus* ("3"). Graph V can be similarly interpreted but with respect to prey items. For example, the clusters and arrows indicate that *Acanthocyclops vernalis* ("8") consumes very different items than do the diving bell spider ("5") and great diving beetle ("27"). Descriptions of SR, U, and V for all eight webs appear in Table S1. As can be seen in at least one of the three graphs, the notion of "trophic levels" is consistently ambiguous for all eight webs. This agrees with the findings in ref. 23 that TL estimates have substantial uncertainty and that catch MTL is a poor biodiversity indicator for marine ecosystems. Our results extend the argument beyond marine environments and are conveniently available from a single empirical analysis applied to each of a small number of webs.

Insight into other aspects of network dependency is also available from the statistical inference. For *Bay*, the negative trend

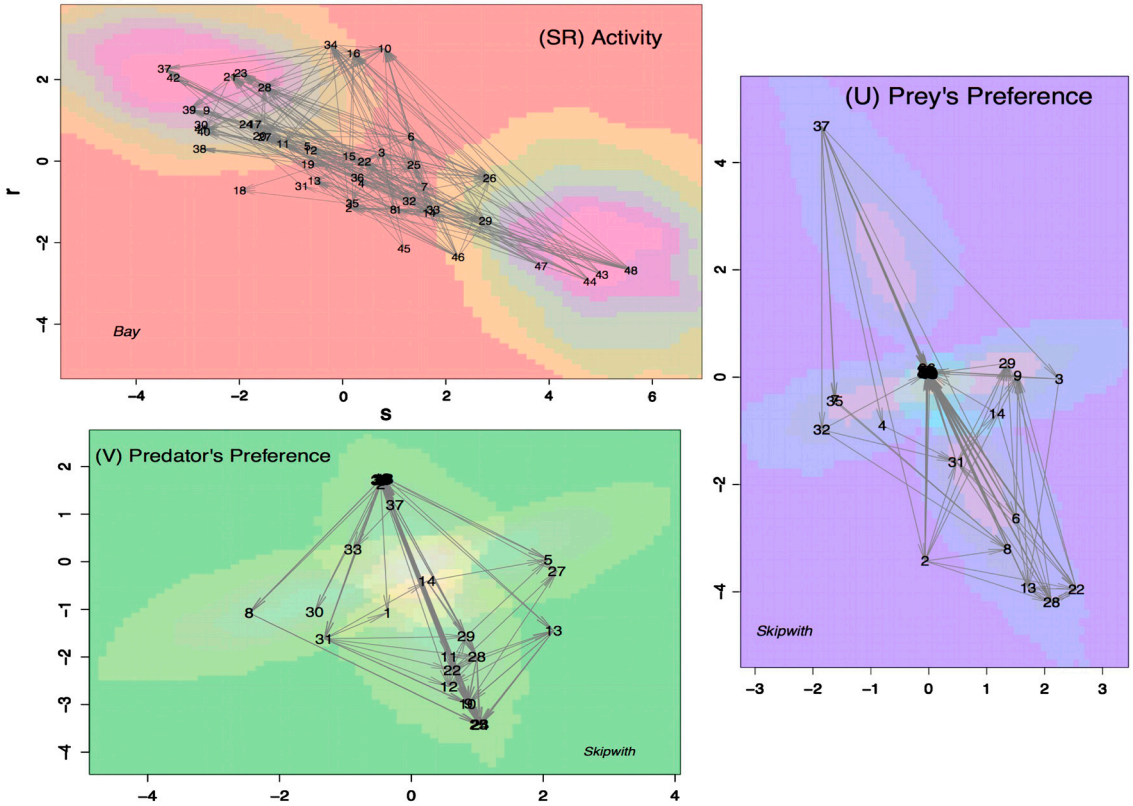

Fig. 2. Food web graphs displaying trophic structure from fitting model **1** with phylogeny measure x_{ij} as the predictor. (SR) Feeding activity in *Bay*. The *s* axis refers to activity as prey, and the *r* axis, as predator. Label of node *i* is located at the mean of the bivariate posterior distribution (an "estimate") of $[s_i, r_i]$; this distribution describes the probability of the position of $[s_i, r_i]$ on the *sr* plane, given the knowledge of the observed food web. Distribution density appears as heat map for benthos-eating birds ("37") and detritus ("48"). Legend for node labels appears in *SI Text*. (U, V) Preference of being consumed/consuming in *Skipwith*. They are similarly interpreted as (SR), but referring to u_i vectors for *L. variegatus* ("3"), *S. lutaria* ("28"), *C. scutellata* ("32"), and detritus ("37"), and v_i vectors for diving bell spider ("5"), *A. vernalis* ("8"), *S. scoticum* ("12"), and detritus. Nodes far apart in the latent *u* or *v* space differ substantially with respect to feeding preference.

in SR provides evidence for $\rho_{sr} < 0$—i.e., active predators are unlikely to be active prey, and vice versa. Inference summary statistics for various model parameters provide complementary information to the graphs.

Numerical Summaries of Inference and Goodness-of-Fit. Tables 1 and 2 feature *Bay*, *Reef*, *Skipwith*, and *St. Martin*. They represent a range of scenarios on (*a*) the importance of phylogenetic information, (*b*) the extent of predator-prey reciprocity, and (*c*) the

Table 1. Bayesian inference numerical summaries for selected food webs from fitting statistical social network model 1, with phylogenetic similarity x_{ij} as the predictor

			Credible interval		
Food web*	Parameter	Posterior median†	Lower limit	Upper limit	Credibility‡
Bay	β_1	5.10	1.20	10.46	0.95
	ρ_{sr}	−0.46	−0.75	−0.07	0.95
	ρ	−0.82	−0.95	−0.11	0.75
Reef	β_1	2.16	0.06	4.21	0.90
	ρ_{sr}	−0.08	(interval includes 0)		0.50
	ρ	−0.29	−0.52	−0.02	0.60
Skipwith	β_1	−1.07	(interval includes 0)		0.50
	ρ_{sr}	−0.04	(interval includes 0)		0.50
	ρ	−0.39	−0.67	−0.04	0.50
St. Martin	β_1	0.39	(interval includes 0)		0.50
	ρ_{sr}	−0.52	−0.80	−0.07	0.95
	ρ	−0.98	−1.00	−0.24	0.99

*Other food webs and z_{ij} appear in Table S2.
†The posterior median can be considered a parameter "estimate."
‡Credible intervals presented have approximately the highest credibility without including 0. High credibility for an interval excluding 0 indicates statistical importance of the corresponding parameter to feeding potential (p_{ij}).

Table 2. Goodness-of-fit summaries for selected food webs (others in Table S2)

Model	Predictor	GoF*	Model	Predictor	GoF
	Bay			*Reef*	
1	x_{ij}	367	**1**	x_{ij}	524
1	z_{ij}	386	**1**	z_{ij}	526
1	—	394	**1**	—	533
naive†	x_{ij}	719	naive	x_{ij}	1273
naive	z_{ij}	719	naive	z_{ij}	1287
	Skipwith			*St. Martin*	
1	x_{ij}	138	**1**	x_{ij}	252
1	z_{ij}	138	**1**	z_{ij}	204
1	—	142	**1**	—	287
naive	x_{ij}	319	naive	x_{ij}	482
naive	z_{ij}	321	naive	z_{ij}	481

*Derived from the Bayes factor on the model's ability to predict the act of feeding (y_{ij}). When comparing between models, noticeably smaller GoF values suggest better fit.

†Simple logistic regression ignoring network dependence—i.e., naively setting $s_i + r_j + u_i'v_j + \varepsilon_{ij} = 0$ for all i,j in model **1**.

association between an individual's sending and receiving activity. Based on Table 1, *Bay*/*Skipwith* is high/low for all *a*–*c*, and *Reef*/*St. Martin* is high/low for *a* but low/high for *b*–*c*. Through Table 2, these webs showcase the superior goodness-of-fit of model **1** relative to the naive model. Inference for model parameters appears in Tables 1 and 2 and Table S2: All eight webs show evidence that (*i*) phylogeny is noticeably relevant to feeding [moderate to strong evidence that $\beta_1 \neq 0$ from at least one perspective of feeding (see *Materials and Methods*)], and (*ii*) $\rho_{sr} < 0$, although weakly for *Grass* (x_{ij}), *Reef*, or *Skipwith*. *Bay, Benguela, Reef*, and *Shelf* show evidence of varying strength that phylogenetically similar nodes are more likely to yield feeding interaction ($\beta_1 > 0$). *Grass* and *Ythan* show the opposite tendency with reasonable evidence; both webs exhibit minimal connectance ($C = \sum_i \sum_j (y_{ij} > 0)/[n(n-1)]$), thus the prevalence of $y_{ij} = 0$ may explain the evidence for $\beta_1 < 0$. Moderate to strong evidence that predation is unlikely to reciprocate between nodes ($\rho < 0$) is seen in all webs except *Grass* and *Skipwith*, whose weak ρ and ρ_{sr} may reflect their heavy dominance by insects with similar taxonomy. A weak ρ_{sr} also among *Reef* nodes can be due to their unusually coarse taxonomic classification.

When analyzing a social network with n nodes, it is common to provide descriptive indices of network features (e.g., refs. 26 and 27) such as the connectance index, C, for trophic food webs. Popular indices for our eight food webs appear in ref. 14. Our statistical SNA can provide probabilistic interpretations associated with these indices by accounting for the natural variability inherent in feeding behavior, via the posterior predictive distribution (25) of the index (Fig. 3 and Table S3). This distribution provides information about the uncertainty in the index under plausible scenarios that may alter the links in the food web. It can further be used for model validation (*Materials and Methods*). In our case, validation of model **1** suggested that the model is indeed consistent with the observed food webs.

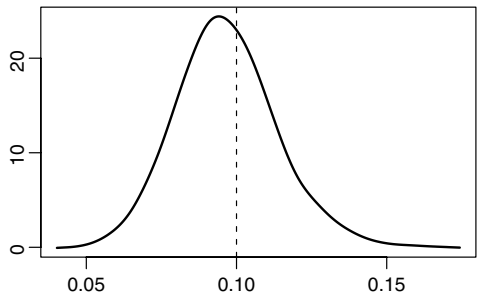

Fig. 3. Posterior predictive distribution of the connectance index C, from fitting model **1** for the *Bay* food web. For the presently observed food web, $C = 0.10$ (dashed line).

Discussion

Using statistical SNA methods to study trophic relations, we have demonstrated a new direction for quantitative analysis of food webs. The statistical inference framework provides a common thread to the understanding of feeding patterns and broad-sense connectivity, the relevance of nodal attributes (e.g., phylogenetic similarity) to feeding behavior, trophic clustering from various perspectives of feeding behavior, and the uncertainty of food web descriptive measures (e.g., MTL, C). In general, food web features are subject to natural variability (uncertainty due to natural events in which feeding behavior may vary) and to observation error such as in the identification of predator's stomach content. Under this framework, we can conveniently visualize numerous food web features through a small set of graphical displays and use probabilistic statements about quantitative parameters in the regression model to summarize these features and assess the extent of the influence of nodal attributes on the trophic structure. It facilitates a comprehensive understanding of trophic structure that is readily achievable with a single, unified quantitative food web analysis.

Materials and Methods

Data. The eight food webs that we analyzed represent a variety of aquatic and terrestrial commmunities (14) (*SI Text*). Each of the corresponding eight source articles (15–22) contains information from which the diet matrix for the presence (1) and absence (0) of feeding linkage between nodes can be deduced. Cannibalism data (y_{ii}) are not modeled by Eq. **1**; in this context, of interest is the structure of codependence among the different nodes rather than a node's self influence. Taxonomic information of nodes accompanies each article, although their formats differ. For some webs, this information contains common names only, but for others it contains Latin names at various levels of the phylogenetic tree. For convenience, we converted all eight sets of taxonomic information into Linnaean trees of a common format, with the ranks of (*i*) domain, (*ii*) kingdom, (*iii*) phylum, (*iv*) class, (*v*) order, (*vi*) family, (*vii*) genus, and (*viii*) species. Thus, information about subclass, suborder, etc., was only used to determine missing information about superceding ranks. Six of the eight webs comprise nodes (e.g., detritus) that entirely consist of organic but inanimate matter. For these six, we appended the Linnaean tree to an artificial superceding rank of "animacy," which took on one of two values, *animate* or *inanimate*. We consulted the online resource Integrated Taxonomic Information System (ITIS, http://www.itis.gov) to make the conversions; various other online sources (via Internet search engines) were used only when the original taxonomic information contained Latin names that were not found within the ITIS. Based on the resulting phylogenetic trees, we computed two different measures of phylogenetic similarity for each food web: similarity x that addresses missing taxonomic topology, and a more conservative version z that regards missing topology as implying different phylogeny.

Constructing Phylogenetic Similarity Measures. The taxonomic classification (e.g., Linnaean) of an organism is a set of polychotomous qualitative characters. In the absence of a universal measure of phylogenetic similarity between two organisms according to their taxonomic classification, Gower's general coefficient of similarity and its variants (28) may be used to quantify the comparison. In the case of complete topological information on the Linnaean tree, the path length between two species (29) is a special case of Gower's measure. However, the nodes in each of our eight food webs result from aggregation of species at uneven taxonomic resolutions, so that full topology is unavailable/inapplicable.

Uneven aggregation is common in food web studies (15, 22, 30). For example, consider the following four nodes in the *Benguela* food web: Node "3" is identified as "bacteria"; "4," as "benthic carnivores"; "23," as "kob"; and "26," as "whales and dolphins." If mapped to the eight-character Linnaean classification of ranks *i*–*viii* from above, then these nodes are identified at four different resolutions. Specifically, let "NA" denote "not applicable" or missing. Then, listed in the order of *i*–*viii*, "3" is *Bacteria*-NA-NA-NA-NA-NA-NA-NA; "4" is *Eukaryota-Animalia*-NA-NA-NA-NA-NA-NA; "23" is *Eukaryota-Animalia-Chordata-Actinopterygii-Perciformes-Sciaenidae-Argyrosomus-A. hololepidotus*; and "26" is *Eukaryota-Animalia-Chordata-*

Mammalia-Cetacean-NA-NA-NA. To quantify the phylogenetic similarity between any two nodes, let $a_{ijk} = 1$ if nodes i and j match on character k, and $a_{ijk} = 0$ otherwise. Also let $w_{ijk} = 1$ if information exists for the kth character for both i and j, and $w_{ijk} = 0$ otherwise. Typical use of Gower's measure removes from consideration any character that involves one or more NAs. Thus, a common variant of Gower's similarity coefficient between i and j is

$$A_{ij} = \frac{a_{ij1}w_{ij1} + \ldots + a_{ijK}w_{ijK}}{w_{ij1} + \ldots + w_{ijK}}$$

where $K = 8$ for this example. The measure ranges between 0 and 1. This definition gives $A_{3,4} = A_{3,23} = A_{3,26} = 0$, $A_{4,23} = A_{4,26} = 1$ (maximum possible value), and $A_{23,26} = 3/5$. However, $A_{23,26} < A_{4,23} = A_{4,26}$ appears counterintuitive and remains so as long as $w_{ijk} = 0$ for any k involving NAs. As an alternative similarity measure, we propose taking

$$w_{ijk} = \begin{cases} 1 & \text{for } k = 1,\ldots,m \\ 0 & \text{for } k = m+1,\ldots,K \end{cases}$$

where $m = \max_k \{i\text{'s character } k \text{ is not NA}, j\text{'s character } k \text{ is not NA}\}$. For example, $m = 8$ for computing $A_{4,23}$, but $m = 5$ for computing $A_{4,26}$. Then, $A_{3,4} = A_{3,23} = A_{3,26} = 0$, $A_{4,23} = 2/8 = 0.25$, $A_{4,26} = 2/5 = 0.4$, and $A_{23,26} = 3/8 = 0.375$.

Our measure attempts to address missing phylogenetic topology and is intended to yield less lopsided values than those based on the conventional practice of discrediting NAs. However, the occasional pair with an equal number of NAs may still be problematic. For example, with the above eight-character classification, the nodes "birds" and "sharks" are only identified to the class level (*Aves* and *Chondrichthyes*, respectively). Here, their similarity coefficient is $3/4 = 0.75 = 2 \times A_{23,26}$, which may be unrealistic. A more conservative alternative is to regard all characters involving NAs to be different between nodes. Thus, we would replace the above $A_{4,23}$ with 0.25, and the coefficient between birds and sharks would be $3/8 = 0.375$. The flip side of this conservative alternative is an unrealistically small value for comparing, say, microzooplankton and mesozooplankton, both of which are only identified to the kingdom level (*Animalia*), so that their similarity coefficient is 0.25 (conservative) as opposed to 1, the latter of which is obtained using either our less conservative method or the conventional approach of discarding NAs altogether. While it may be of general interest, a more sophisticated similarity measure in the presence of "jagged" phylogenetic topology is not the focus of our work. Indeed, comparisons similar to that of birds and sharks, or of micro- and macrozooplankton, are in the minority for the food webs analyzed here.

For statistical SNA using model **1**, we considered (*a*) the similarity coefficient as proposed above to address missing topology, and (*b*) the more conservative measure that regards NAs as implying difference. We took the logarithm of each measure (1 was added to all A_{ij} values prior to transformation to avoid taking the log of 0) and denoted them by x_{ij} and z_{ij}, respectively. Such transformation reduced skewness of the semiqualitative covariate, thus increasing its ability to distinguish among cases (pairs of nodes here) to help predict the response. Note that $x_{ij} = x_{ji}$ and $z_{ij} = z_{ji}$.

Statistical Analyses and Model Validation. Given either x_{ij} or z_{ij} in Eq. **1** for a food web with n nodes, we performed Bayesian estimation of the parameters β_0, β_1, ρ_{sr}, and ρ, as well as $[s_i,r_i]$, u_i, and v_i for all $i = 1,\ldots,n$. A food web with n nodes consists of $n(n-1)$ pairwise directed links y_{ij} for $i \neq j$. In Eq. **1**, the sender and receiver effects, s_i and r_j, can be interpreted as sending and receiving activity. Thus, $s_i > s_j$ implies that node i is more active as prey than node j. Similarly, $r_i > r_j$ implies that i is more active as predator than j. Given node i, its activity level in the food web as either prey or predator is conveniently described by the vector $[s_i,r_i]$. The interaction term $u_i'v_j$ in Eq. **1** is the inner product of k-dimensional vectors $u_i = [u_{i1},\ldots,u_{ik}]$ and $v_j = [v_{j1},\ldots,v_{jk}]$. In our analyses, we took $k = 2$ for easy visualization of these latent spaces, although one could define criteria for selecting an optimal k (7, 13). In essence, if u_a and u_b are neighbors in the u space, then the sending behavior (other than sending activity) of a to c is similar to that of b to c, for all nodes c. The same interpretation applies to the receiving behavior (other than receiving activity) of neighbors v_a and v_b in the v space. In the food web context, the u space then refers to preference of being consumed, and v space, to preference of consuming. Eq. **1** alone constitutes a two-way analysis-of-variance model with random row, column, and row-column interaction effects (31). Dependence features within the network are represented by additionally specifying that the vectors $[s_i,r_i]$ and $[\varepsilon_{ij},\varepsilon_{ji}]$ each has some nonzero covariance. For this, we assume $[s_1,r_1],\ldots,[s_n,r_n]$ are independent and identically distributed (iid) as bivariate mean-zero Gaussian vectors, with variance-covariance matrix Σ, and $[\varepsilon_{ij},\varepsilon_{ji}]$ for all $i \neq j$ are iid bivariate mean-zero Gaussian vectors, with variance-covariance matrix Ω, where

$$\Sigma = \begin{bmatrix} \sigma_s^2 & \rho_{sr}\sigma_s\sigma_r \\ \rho_{sr}\sigma_s\sigma_r & \sigma_r^2 \end{bmatrix}, \qquad \Omega = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

For a food web, ρ_{sr} describes the tendency for any given node to be active as both prey and predator—e.g., $\rho_{sr} < 0$ implies that if a node is active as prey, then it is unlikely to be active as predator, and vice versa. The parameter ρ is the correlation between ε_{ij} and ε_{ji}. It describes the tendency of predator-prey role reversal within any given pair of nodes—e.g., $\rho < 0$ implies that predation is unlikely to be reciprocated within any given pair. Finally, writing the 2D preference vectors as $u_i = [u_{i1},u_{i2}]$ and $v_j = [v_{j1},v_{j2}]$, we assume that $u_{1q},\ldots,u_{nq}$ are iid Gaussian with mean 0 and variance σ_{uq}^2 for $q = 1,2$, and $v_{1q},\ldots,v_{nq}$ are similarly distributed but with variance σ_{vq}^2. This is the default assumption that is implemented in "gbme.asym.r" for performing statistical SNA.

Bayesian estimation of the set of parameters $\{[\beta_0,\beta_1,\sigma_s^2,\sigma_r^2,\rho_{sr},\sigma^2,\rho,\sigma_{u1}^2,\sigma_{u2}^2,\sigma_{v1}^2,\sigma_{v2}^2],[s_i,r_i,u_{i1},u_{i2},v_{i1},v_{i2}]_{i=1,\ldots,n}\}$ then proceeded for each food web dataset by applying the default prior distributions built into the "gbme.asym.r" software to produce Markov chain Monte Carlo (MCMC) draws (25) from the joint posterior distribution (jpd) of these parameters. The underlying jpd is the basis of Bayesian inference; we approximated it empirically by the distribution of the MCMC draws. For location parameters $[s_i,r_i]$, u_i, and v_i, the mean of the MCMC draws (after Procrustes transformation, if applicable—see next section) was taken as the parameter estimate and used to produce network graphs in Fig. 2. Nodes far apart in SR/U/V differ substantially with respect to feeding activity/being preferred as prey by others/preference of consuming others. Each heat map in Fig. 2 is the density of the approximate jpd marginalized over all parameters except those that correspond to the displayed bivariate plane for the particular node of interest. For example, the upper-left part of the SR heat map is the density of the jpd marginalized over all parameters except $[s_{37},r_{37}]$; it describes the probability of the location of $[s_{37},r_{37}]$ on the sr plane, given the knowledge of the observed food web. Alternatively, quantiles of the jpd can be used as numerical summaries of the density and to assess uncertainty.

For nonlocation parameters, the median and upper and lower αth quantiles of the MCMC draws were used to summarize the Bayesian inference in the respective form of a parameter estimate and a 100(1–2α)% credible interval (Bayesian analog of the confidence interval). Credibility or credible level 1–2α is the probability, based on the observed food web, that the parameter lies inside the associated credible interval. The conventional practice is to present all intervals at an arbitrary but high credible level—e.g., 95%. Then, often a 95% credible interval for, say, β_1 may include 0, although a 90% credible interval may not. Because 90% is also high, one ought not to regard phylogenetic similarity as a statistically unimportant predictor for feeding behavior simply because the 95% credible interval for β_1 includes 0. The same argument applies to the credible intervals for ρ and ρ_{sr}, which, when excluding 0, imply the statistical importance, respectively, of the tendency for predator-prey role reversal and of the association between activity as prey and as predator. To avoid being misled by the arbitrary cutoff for credible levels, we computed credible intervals for each model parameter in Table 1 and Table S2 at credible levels of 50% and above, in 5% increments up to 95%, then finally at 99%. Among these intervals, the one that excluded 0 with the highest credibility is reported. If its associated credible level is above 50%, then there is some statistical evidence that the parameter is important; the higher the credible level, the stronger the evidence. We do not report the actual interval if it included 0 with 50% credibility: In this case, there is a high posterior probability that the parameter falls in a range that includes 0, indicating little statistical importance of the corresponding web feature with respect to p_{ij}. This, however, does not preclude a web feature's importance with respect to y_{ij}. In *SI Text*, we explain the two perspectives of statistical evidence for $\beta_1 \neq 0$. The former corresponds to the direct relationship, as given by Eq. **1**, between phylogenetic information and feeding potential (p_{ij}). The latter, though, corresponds to the amount of statistical gain in describing the act of feeding (y_{ij}) by including the predictor. A gain is evidenced by a noticeable reduction in GoF in Table 2 and Table S2 between models. The tables demonstrate the relevance of x and/or z to the act of feeding (except for *Ythan*), even when the relevance to feeding potential is inconsequential (*Reef*: z; *Skipwith, St. Martin*: x,z). GoF and credible intervals for β_1 together provide evidence for all eight webs that phylogenetic similarity as the predictor in Eq. **1** yields statistical gain to the understanding of one or both forms of feeding behavior.

Because our statistical SNA provides probabilistic interpretations of trophic structure, such interpretations can be applied to any measure that

is computed from the diet matrix to describe trophic relational patterns, which are subject to natural variability. Here, we focus on the connectance index, C; the same principles are applicable to other descriptive measures. The index C is the fraction of observed pairwise links out of all possible pairs in the food web. Although C alone is not designed to describe broad-sense connectivity, it formalizes the notion of network link density that can be visualized in Fig. 2. The posterior predictive distribution of C (Fig. 3) describes the likelihood for the values of C, given the same set of n nodes, under plausible scenarios (due to natural variability) in addition to those that gave rise to the presently observed food web. For example, while $C = 0.10$ for the current *Bay* web, the probability is 0.95 for C to fall in the credible interval between 0.07 and 0.13. This inference for C also provides an assessment of uncertainty: A high probability for a short credible interval indicates little uncertainty in C. Similar probability statements for the other seven food webs appear in Table S3. Fig. 3 can also be viewed as a tool for model validation: Model **1** predicts the *Bay* food web to yield a C that would most likely lie in the range surrounding the mode of this distribution; because the presently observed C indeed lies in this range, our model is consistent with the observed. This consistency was seen in all eight food webs.

Procrustes Transformation. Due to the invariance of $u_i'v_j$ under certain reflection/rotation of the u space and v space, we worked out the mathematics for the special handling of u_i and v_j so that the inference was readily interpretable.

Bayesian inference for model **1** via MCMC involves the simulation of T draws from the jpd of $\{[\beta_0,\beta_1,\sigma_s^2,\sigma_r^2,\rho_{sr},\sigma^2,\rho,\sigma_{u1}^2,\sigma_{u2}^2,\sigma_{v1}^2,\sigma_{v2}^2], [s_i,r_i,u_{i1},u_{i2},v_{i1},v_{i2}]_{i=1,\ldots,n}\}$. In what follows, we let the superscript "(t)" denote the tth MCMC sample, for $t = 1,2,\ldots,T$. The inner product $\langle u_i,v_j\rangle = u_i'v_j$ is a parameter in Eq. **1** and is associated with a specific orientation of the u space and v space. The set of MCMC inner products consisting of $\langle u_i^{(t)},v_j^{(t)}\rangle$ for $t = 1,2,\ldots,T$ forms the inference for $\langle u_i,v_j\rangle$, though our interest lies in the inference for u_i and v_j instead of their product. In this case, basing the inference for u_i directly on $\{u_i^{(1)},\ldots,u_i^{(T)}\}$ is problematic, and similarly for v_j. The reason is as follows. Each $u_i^{(t)}$ and $v_j^{(t)}$ arises from the u space and v space under the tth orientation. The tth orientation may drastically differ from the $(t+1)$st, yet $\langle u_i^{(t)},v_j^{(t)}\rangle$ may be identical to $\langle u_i^{(t+1)},v_j^{(t+1)}\rangle$. Thus, directly pooling the $u_i^{(t)}$s from all T different orientations leads to uninterpretable inference for u_i, which resides in the u space under a specific orientation. The same argument applies to v_j. Sensible inference for u_i and v_j may be obtained by forcing each tth pair of $u_i^{(t)}$ and $v_j^{(t)}$ to take on a common orientation via a two-sets orthogonal Procrustes transformation (32). This generalizes the Procrustes transformation of refs. 7 and 8 used for the symmetric case where $u_i = v_i$ for all nodes i. For our model under asymmetry, we take $u_{i0} = \sum_{t=1}^{T} u_i^{(t)}/T$ and $v_{i0} = \sum_{t=1}^{T} v_i^{(t)}/T$ as the arbitrary "default" orientations of the u space and v space, respectively, for each ith node. For each t, the objective is to find a 2×2 orthogonal transformation matrix, $Q^{(t)}$, such that $Q^{(t)}u_i^{(t)}$ aligns with u_{i0} as closely as possible and $Q^{(t)}v_i^{(t)}$ aligns with v_{i0} as closely as possible, for all i. Temporarily drop the superscript "(t)" to reduce clutter, and define U as the $2\times n$ matrix whose ith column is u_i; define U_0, V, and V_0 similarly. Therefore, for each MCMC sample,

$$\begin{aligned} Q &= \arg\min_A\{\|AU - U_0\| + \|AV - V_0\|\} \\ &= \arg\max_A \operatorname{trace}\{(LDM')A\} \\ &= \arg\max_A \operatorname{trace}\{DM'AL\} = ML' \end{aligned}$$

where LDM' is the singular value decomposition of $C = UU_0' + VV_0'$, i.e., D is the diagonal matrix whose entries are the singular values of C, L is the orthogonal matrix whose columns are the eigenvectors of CC', and M is the orthogonal matrix whose columns are the eigenvectors of $C'C$. Each tth MCMC sample has a unique $U^{(t)}$ and $V^{(t)}$, and hence, $Q^{(t)}$ derived as above.

Due to the orthogonality of $Q^{(t)}$, we have $\langle Q^{(t)}u_i^{(t)},Q^{(t)}v_j^{(t)}\rangle = \langle u_i^{(t)},v_j^{(t)}\rangle$; thus, the posterior distribution of (hence, inference for) $u_i'v_j$ is invariant under the transformation $Q^{(t)}$. The sets $\{Q^{(1)}u_i^{(1)},\ldots,Q^{(T)}u_i^{(T)}\}$ and $\{Q^{(1)}v_j^{(1)},\ldots,Q^{(T)}v_j^{(T)}\}$ form the Bayesian statistical inference for u_i and v_j, respectively, and are the basis of (U) and (V) in Fig. 2.

ACKNOWLEDGMENTS. We thank J. A. Dunne for data files; C. Bondavalli, J. A. Dunne, and R. E. Ulanowicz for reference material; G. R. Hosack and A. B. Zwart for suggestions; and S. Barry, J. M. Dambacher, K. R. Hayes, and G. R. Hosack for their moral support. A.H.W. thanks Commonwealth Scientific and Industrial Research Organisation (CSIRO) Mathematics, Informatics and Statistics for its financial support toward his Visiting Researcher position during 2010 when this work was primarily conducted.

1. Cohen JE, Jonsson T, Carpenter SR (2003) Ecological community description using the food web, species abundance, and body size. *Proc Natl Acad Sci USA* 100:1781–1786.
2. Dambacher JM, et al. (2010) Analyzing pelagic food webs leading to top predators in the Pacific Ocean: A graph-theoretic approach. *Prog Oceanogr* 86:152–165.
3. Krause AE, Frank KA, Mason DM, Ulanowic RE, Taylor WW (2003) Compartments revealed in food-web structure. *Nature* 426:282–285.
4. Mucha PJ, Richardson T, Macon K, Porter MA, Onnela J-P (2010) Community structure in time-dependent, multiscale, and multiplex networks. *Science* 328:876–878.
5. Wasserman S, Faust K (1994) *Social Network Analysis: Methods and Applications* (Cambridge Univ Press, Cambridge).
6. Westveld AH, Hoff PD (2011) A mixed effects model for longitudinal relational and network data, with applications to international trade and conflict. *Ann Appl Stat* 5:843–872.
7. Hoff PD (2005) Bilinear mixed-effects models for dyadic data. *J Am Stat Assoc* 100:286–295.
8. Hoff PD, Raftery AE, Handcock MS (2002) Latent space approaches to social network analysis. *J Am Stat Assoc* 97:1090–1098.
9. Gill PS, Swartz TB (2001) Statistical analyses for round robin interaction data. *Can J Stat* 29:321–331.
10. Mariadassou M, Robin S, Vacher C (2010) Uncovering latent structure in valued graphs: A variational approach. *Ann Appl Stat* 4:715–742.
11. Chiu GS, Gould JM (2010) Statistical inference for food webs with emphasis on ecological networks via Bayesian melding. *Environmetrics* 21:728–740.
12. Poole D, Raftery AE (2000) Inference for deterministic simulation models: The Bayesian melding approach. *J Am Stat Assoc* 95:1244–1255.
13. Ward MD, Siverson RM, Cao X (2007) Disputes, democracies, and dependencies: A reexamination of the Kantian Peace. *Am J Pol Sci* 51:583–601.
14. Dunne JA, Williams RJ, Martinez ND (2004) Network structure and robustness of marine food webs. *Mar Ecol Prog Ser* 273:291–302.
15. Christian RR, Luczkovich JJ (1999) Organizing and understanding a winter's seagrass foodweb network through effective trophic levels. *Ecol Modell* 117:99–124.
16. Yodzis P (1998) Local trophodynamics and the interaction of marine mammals and fisheries in the Benguela ecosystem. *J Anim Ecol* 67:635–658.
17. Martinez ND, Hawkins BA, Dawah HA, Feifarek BP (1999) Effects of sampling effort on characterization of food-web structure. *Ecology* 80:1044–1055.
18. Opitz S (1996) *Trophic Relations in Caribbean Coral Reefs* (ICLARM, Manila).
19. Link J (2002) Does food web theory work for marine ecosystems? *Mar Ecol Prog Ser* 230:1–9.
20. Warren PH (1989) Spatial and temporal variation in the structure of a freshwater food web. *Oikos* 55:299–311.
21. Goldwasser L, Roughgarden J (1993) Construction and analysis of large Caribbean food web. *Ecology* 74:1216–1233.
22. Hall SJ, Raffaelli D (1991) Food-web patterns: Lessons from a species-rich web. *J Anim Ecol* 60:823–842.
23. Branch TA, et al. (2010) The trophic fingerprint of marine fisheries. *Nature* 468:431–435.
24. Chiu GS, Westveld AH (2010) A statistical social network model for consumption data in food webs. arXiv:1006.4432v2.
25. Gelman A, Carlin JB, Stern HS, Rubin DB (2004) *Bayesian Data Analysis* (Chapman and Hall/CRC, Boca Raton, FL), 2nd ed.
26. Beyer K, Gozlan RE, Copp GH (2010) Social network properties within a fish assemblage invaded by non-native sunbleak Leucaspius delineatus. *Ecol Modell* 221:2118–2122.
27. Kones JK, Soetaert K, van Oevelen D, Owino JO (2009) Are network indices robust indicators of food web functioning? A Monte Carlo approach. *Ecol Modell* 220:370–382.
28. Gower JC (1971) A general coefficient of similarity and some of its properties. *Biometrics* 27:857–874.
29. Clarke KR, Warwick RM (1998) A taxonomic distinctness index and its statistical properties. *J Appl Ecol* 35:523–531.
30. Dunne JA, Williams RJ, Martinez ND (2002) Network structure and biodiversity loss in food webs: Robustness increases with connectance. *Ecol Lett* 5:558–567.
31. Lomax RG (2001) *Statistical Concepts: A Second Course for Education and the Behavioral Sciences* (Lawrence Erlbaum Associates, Mahwah, NJ), 2nd ed.
32. Gower JC, Dijksterhuis GB (2004) *Procrustes Problems* (Oxford Univ Press, Oxford).