Covariate-Informed Latent Interaction Models: Addressing Geographic & Taxonomic Bias in Predicting Bird–Plant Interactions

Georgia Papadogeorgou\textsuperscript{a}, Carolina Bello\textsuperscript{b}, Otso Ovaskainen\textsuperscript{c,de}, and David B. Dunson\textsuperscript{f}

\textsuperscript{a}Department of Statistics, University of Florida, Gainesville, FL; \textsuperscript{b}Department of Environmental Systems Science, ETH Zürich, Zurich, Switzerland; \textsuperscript{c}Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland; \textsuperscript{d}Organismal and Evolutionary Biology Research Programme, University of Helsinki, Helsinki, Finland; \textsuperscript{e}Centre for Biodiversity Dynamics, Department of Biology, Norwegian University of Science and Technology, Trondheim, Norway; \textsuperscript{f}Department of Statistical Science, Duke University, Durham, NC

ABSTRACT

Reductions in natural habitats urge that we better understand species’ interconnection and how biological communities respond to environmental changes. However, ecological studies of species’ interactions are limited by their geographic and taxonomic focus which can distort our understanding of interaction dynamics. We focus on bird–plant interactions that refer to situations of potential fruit consumption and seed dispersal. We develop an approach for predicting species’ interactions that accounts for errors in the recorded interaction networks, addresses the geographic and taxonomic biases of existing studies, is based on latent factors to increase flexibility and borrow information across species, incorporates covariates in a flexible manner to inform the latent factors, and uses a meta-analysis dataset from 85 individual studies. We focus on interactions among 232 birds and 511 plants in the Atlantic Forest, and identify 5% of pairs of species with an unrecorded interaction, but posterior probability that the interaction is possible over 80%. Finally, we develop a permutation-based variable importance procedure for latent factor network models and identify that a bird’s body mass and a plant’s fruit diameter are important in driving the presence of species interactions, with a multiplicative relationship that exhibits both a thresholding and a matching characteristics lead to over-representation of a subset of species and under-representation of others implying that the resulting measured networks are taxonomically and geographically biased. Even if measured networks from individual studies are compiled into one overarching network including all recorded interactions, these biases will propagate since cryptic species that occur in uncharted regions, or are not the explicit focus of the individual studies, will remain under-represented. Even though these biases and their implications are well-recognized (Báldi and Montoya 2001), their robustness to species extinctions (Dunne, Williams, and Martinez 2002; Aizen, Sabatino, and Tylianakis 2012), and their resilience in the face of environmental change (Tylianakis et al. 2008). Climate change and the reduction in species’ natural habitats necessitate that we urgently understand species’ interdependence in order to better predict how environmental changes will affect species’ equilibrium and co-existence.

Predicting and understanding species interactions is a long-standing question in ecology. However, accessing all the possible interactions in a mutualistic network is a huge task that requires significant experimental effort (Jordano 2016). Individual studies might focus on recording the interactions of only a given set of species. Even for the species under study, most measured networks are recorded in a specific geographical area where only a subset of species occurs. As a result, recorded networks are substantially incomplete and not comprehensively representative of species interactions, irrespective of the researchers’ observational effort. These individual study

1. Introduction

Animal–plant interactions have played an important role in the generation of Earth’s biodiversity (Ehrlich and Raven 1964). Hundreds of species form complex networks of interdependences whose structure has important implications for the stability of ecosystems (Solé and Montoya 2001), their robustness to species extinctions (Dunne, Williams, and Martinez 2002; Aizen, Sabatino, and Tylianakis 2012), and their resilience in the face of environmental change (Tylianakis et al. 2008). Climate change and the reduction in species’ natural habitats necessitate that we urgently understand species’ interdependence in order to better predict how environmental changes will affect species’ equilibrium and co-existence.

Predicting and understanding species interactions is a long standing question in ecology. However, accessing all the possible interactions in a mutualistic network is a huge task that requires significant experimental effort (Jordano 2016). Individual studies might focus on recording the interactions of only a given set of species. Even for the species under study, most measured networks are recorded in a specific geographical area where only a subset of species occurs. As a result, recorded networks are substantially incomplete and not comprehensively representative of species interactions, irrespective of the researchers’ observational effort. These individual study characteristics lead to over-representation of a subset of species and under-representation of others implying that the resulting measured networks are taxonomically and geographically biased. Even if measured networks from individual studies are compiled into one overarching network including all recorded interactions, these biases will propagate since cryptic species that occur in uncharted regions, or are not the explicit focus of the individual studies, will remain under-represented. Even though these biases and their implications are well-recognized (Báldi and McCollin 2003; Seddon, Soorae, and Launay 2005; Pyšek et al. 2008; Hale and Swearer 2016), most models for species interactions do not account for them (e.g., Bartomeus 2013). Some advances are emerging in the literature (Weinstein and Graham 2017; Graham and Weinstein 2018; Cirtwill et al. 2019), but the models therein do not provide a comprehensive treatment of species’ traits and phylogenetic information. Our goal is to use incomplete networks to understand whether a given bird would eat the fruit of a given plant if given the opportunity, and to learn which species traits are important in forming these interactions.

From a statistical perspective, a bird–plant interaction network can be conceptualized as a bipartite graph, where the birds and plants form separate sets of nodes, and an edge connects one node from each set. If a certain animal-plant interaction has been recorded, the corresponding edge necessarily exists.
However, absence of a recorded interaction does not mean that the interaction is not possible and the networks are measured with error. Modeling the probability of connections on a graph measured without error has received a lot of attention in the statistics literature, and examples stretch across social (Newman, Watts, and Strogatz 2002; Wu et al. 2010), biological (Chen and Yuan 2006; Bullmore and Sporns 2009), and ecological (Croft, Krause, and James 2004; Blonder and Dornhaus 2011) networks, among others. Since the literature on network modeling is vast, we focus on approaches for bipartite graphs. In an early approach, Skvoretz and Faust (1999) adapted the p* network models to the bipartite setting. Community detection in bipartite graphs (referred to as co-clustering) was first introduced in Hartigan (1972) and it has flourished in the last couple of decades (e.g., Dhillon, Mallela, and Modha 2003; Shan and Banerjee 2008; Wang et al. 2011; Razaei, Amini, and Li 2019).

Our approach is more closely related to network modeling using latent factors (Hoff, Raftery, and Handcock 2002; Handcock, Raftery, and Tantrum 2007) and its extension to multilinear relationships (Hoff 2005, 2011, 2015), where the nodes are embedded in a Euclidean space and the presence of an edge depends on the nodes’ relative distance in the latent space. Since our observed networks have missing edges, our approach also has ties to modeling noisy observed networks (Jiang, Gold, and Kolaczyk 2011; Wang et al. 2012; Priebe et al. 2015; Chang, Kolaczyk, and Yao 2022).

Our goals are to complete the bipartite graph of species interactions given the recorded, error-prone networks from individual studies, and to understand which covariates are most important for driving species interdependence. We develop a Bayesian approach to modeling the probability that a bird–plant interaction is possible based on a meta-analysis dataset from 85 studies on the Atlantic Forest. The proposed approach (a) models the probability of a link in the bipartite graph, (b) incorporates the missingness mechanism caused by the taxonomic and geographic bias of individual studies, and the possibility that an interaction was not detected, (c) uses covariate information to inform the network model and improve precision, (d) employs a latent variable approach to link the model components, (e) quantifies our uncertainty around the estimated graph, and (f) uses posterior samples in a permutation approach to acquire a variable importance metric. To our knowledge, our approach is the first to employ latent network models for noisy networks, to use covariates to inform the latent factors via separate models instead of including them in the network model directly, and to study variable importance in latent factor models.

2. A Multi-Study Dataset of Bird–Plant Interactions in the Atlantic Forest

The Atlantic Forest is threatened due to overexploitation of its natural resources, and it currently includes only 12% of its original biome (Ribeiro et al. 2009). In this biome, plants rely heavily on frugivore animals for their seed dispersal, and reductions in frugivore populations lead to disruptions in the regeneration of ecosystems. To better understand species’ interactions and how biological communities respond to environmental changes, we study bird–plant interactions in the Atlantic Forest. We use an extensive dataset which includes frugivore–plant interactions from 166 studies for five frugivore classes (Bello et al. 2017). A recorded interaction represents a setting where a frugivore was involved in the plant's seed dispersal process, in that it handled a fruit in a manner that may have ended in consumption and subsequent dispersal of the seed. Other types of fruit handling that could not have led to seed dispersal were excluded from the dataset, wherever this information was available. Since we focus on bird–plant interactions (excluding mammals or other classes), we maintain 85 studies that include at least one such interaction. These 85 studies recorded interactions for 232 birds and 511 plant species, but only 458 of the plant species were involved in an interaction with a bird (Supplement J includes the list of species). The number of unique recorded bird–plant interactions was 3804.

One of the key characteristics of our data is that unobserved interactions might be possible. For an interaction to be recorded there has to exist at least one study for which both species co-occur at the study site, they interact, and the interaction was detected and recorded. However, individual studies are often limited in terms of the species or geographical area they focus on. Species-oriented studies record only a subset of the interactions that are detected: an animal-oriented study focuses on a given animal’s diet whereas a plant-oriented study focuses on learning which animals eat the fruits of a given plant. Hence, measured networks from such studies do not represent species comprehensively and are taxonomically biased. In contrast, network studies record any interaction that is observed. However, studies of either type often focus on a small area where not all animal and plant species occur, and are hence geographically biased. As a result, a complete record of interactions is almost impossible to acquire, even for species of explicit interest.

The taxonomic and geographical biases of the individual studies propagate when compiling the recorded interactions into one combined network. Since most studies are located in the southeast Atlantic Forest (see Figure 1(a)), interactions among species that do not co-occur in this area are less likely to be detected. Out of the eight bioregions of the Atlantic Forest biome, 45% of interactions were recorded in the Serra do Mar bioregion, and there was no recorded interaction in the São Francisco bioregion. Therefore, the combined network will over-represent species that occur in the regions that are heavily studied and under-represent species that do not, implying that the combined network is itself geographically biased. Out of the 85 studies in our data, 19 were animal-oriented, 45 were plant-oriented, and 19 were network studies (the remaining two were a combination). Figure 1(b) shows the number of unique species observed in each study by study type. Animal-oriented studies have recorded interactions on a much smaller number of bird species than plant-oriented studies, and the reverse is true for plant species. These trends in over- and under-representation of certain species will persist in the combined network, as species that were the focus of species-oriented studies will be more heavily represented. In fact, our data over-represent trees and shrubs and under-represent other types of plants, whereas birds with recorded interactions correspond to only 27.1% of the birds residing in the Atlantic Forest (Bello et al. 2017). An analysis of the number of recorded interactions in Bello et al. (2017) indicates that new studies continue to discover previously
undetected interactions, implying that the recorded interactions are only a subset of those that are possible.

Our data include key bird and plant physical traits such as the diameter and color of the plant’s fruit, and the bird’s body mass and gape size, which are available with varying amounts of missingness. These covariates may influence the success of a frugivory interaction, and researchers are interested in understanding this relationship (Rossberg 2013; Fenster et al. 2015; Dehling et al. 2016; Descombes et al. 2019). In ecological studies, it is often assumed that species that are more genetically related have more similar traits and share more interactions. Phylogenetic trees have been used to represent such correlations across species (Ives and Helmus 2011), and incorporating species’ phylogenetic information can improve our understanding of species interactivity (Benadi et al. 2022). In some cases, phylogenetic information has agreed with observed correlations in species’ traits or interaction profiles (Mariadassou, Robin, and Vacher 2010), but not in others (Rezende et al. 2007). We acquire phylogenetic information for bird species from https://birdtree.org (Jetz et al. 2012) and for plant species using the V. PhyloMaker R package (Jin and Qian 2019).

3. Learning Species Interactions Addressing Geographic and Taxonomic Bias

We use $i = 1, 2, \ldots, n_B$ and $j = 1, 2, \ldots, n_P$ to represent birds and plants, respectively. For every bird $i$, $X_i = (X_{i1}, X_{i2}, \ldots, X_{ip_B})'$ represents $p_B$ measured physical traits, and similarly $W_j = (W_{j1}, W_{j2}, \ldots, W_{jp_P})'$ for plant $j$. Each species has an individual detectability score representing the probability that it would be detected to interact if the interaction occurred. We denote this as $p_i$ for birds and $q_j$ for plants. Each study $s = 1, 2, \ldots, S$ has recorded an interaction for each pair of species, or not. We compile the record of measured interactions across all studies in a three-dimensional array $A$ of dimension $n_B \times n_P \times S$ where the $ijs$ entry, $A_{ijs}$, is equal to 1 if study $s$ recorded an $(i,j)$ interaction, and equal to 0 otherwise. We are interested in inferring the $n_B \times n_P$ matrix $L$, the entries of which represent whether bird $i$ would interact with plant $j$ if given the opportunity ($L_{ij} = 1$), or not ($L_{ij} = 0$). We assume that there was no human error in recording interactions, and a recorded interaction is truly possible (hence, if $A_{ijs} = 1$ for at least one $s$, then $L_{ij} = 1$ necessarily). In contrast, pairs without any recorded interaction might still be interactive. Our goal is to infer the value of $L_{ij}$ for pairs $(i,j)$ without a recorded interaction. In our study, $n_B = 232$, $n_P = 511$, and $S = 85$. A glossary is included in Supplement A.

3.1. Study Focus and Species Co-occurrence

To elucidate a model for the probability that two species are interactive, we first investigate the conditions under which a specific pair would be recorded to interact in a given study. In a measured network, an interaction would be recorded if all of the following held: (a) the species interact if given the opportunity, (b) the species are of interest in the particular study, (c) the species co-occur in the study area, and (d) the researchers detected the two species interacting. If any of the above does not happen, the given study necessarily would not record the specific interaction. Violations of (b) and (c) capture the potential of taxonomic and geographical bias of the given study. To address these biases, one should take the focus and species occurrence for each study into consideration. We let $F$ denote a three-dimensional binary array of dimension $n_B \times n_P \times S$ representing the focus of each study. In general, $F_{ijs} = 1$ implies that if an interaction between $i$ and $j$ was detected, it would have been recorded, and $F_{ijs} = 0$ only occurs when study $s$ is animal- or plant-oriented, and the species $i$ or $j$ are not species of interest. For species occurrence, we let $O$ be a similarly defined array where $O_{ijs} = 1$ indicates that species $i$ and $j$ both occur in the geographic area of study $s$, and $O_{ijs} = 0$, otherwise. The focus of each study is known. In contrast, even if the individual study area is well-defined, lack of perfect knowledge of which species exist in that area keeps us from knowing which interactions are even possible to be observed (Poisot, Stouffer, and Gravel 2015).

In our study, we consider $O$ fixed and known. We discuss its choice in Section 6 and extensions that assume $O$ unknown where applicable.

3.2. The Covariate-Informed Latent Interaction Model

Our approach is based on linking measured networks, species true interactivity, detectability, and trait information using latent factors $U_i = (U_{i1}, U_{i2}, \ldots, U_{iH})'$ for bird $i$ and $V_j = (V_{j1}, V_{j2}, \ldots, V_{jH})'$ for plant $j$. The number of latent factors $H$ can be conceptualized as very large to include all important species’ traits, measured or not. To elucidate a likelihood for
the measured networks and trait information, \( P(A, \{X\}, \{W\} \mid F, O) \), we make assumptions that are summarized below and discussed in detail in Supplement B.

### 3.2.1. The Likelihood for the Measured Networks

We allow for species’ (measured or latent) covariates to drive detectability and the interactions that they are able to form. For example, a bird’s size might be informative of both as larger birds are more visible and their larger beaks allow them to consume fruits of all sizes. Conditional on the species’ detectability scores and their true interaction profiles, we assume that species traits do not inform which interactions are recorded in any other way.

Measured networks might exhibit dependence across species or study sites in the following ways. Studies that are geographically close or focus on similar species will have similar patterns of recorded interactions. An impossible interaction will be unrecorded across all studies. Species that are hard to detect will have a low number of recorded interactions across all measured networks. We assume that these are all types of dependencies that can manifest across the measured networks, and measured networks are independent across species and studies once we condition on study focus, species occurrence and detectability, and the true underlying interaction matrix. The assumption that observing a possible interaction is conditionally independent across studies has been previously employed within a related context (Weinstein and Graham 2017). We assume that the record of each interaction depends only on the individual species and study characteristics. These assumptions allow us to write the likelihood of the measured networks conditional on (measured and latent) covariates, the true interaction matrix, study focus, species occurrence and networks conditional on (measured and latent) covariates, the assumptions allow us to write the likelihood of the measured

\[
\prod_{i,j} P(A_{ij} = a_{ij} \mid L_{ij}, F_{ij}, O_{ij}, p_i, q_j)
\]

(see Supplement B.1). We specify

\[
P(A_{ij} = 1 \mid L_{ij} = 1, F_{ij} = f, O_{ij} = o, p_i, q_j) = \begin{cases} 0, & \text{if } f \in o, \\
p_i q_j, & \text{if } f \not\in o, \end{cases}
\]

(1)

which implies that the conditional likelihood for the measured networks simplifies to

\[
\prod_{i,j} (p_i q_j)^{a_{ij}} (1 - p_i q_j)^{1 - a_{ij}} \prod_{i,j} I(a_{ij} = 0).
\]

(2)

Equation (1) acts as a specification for the “missingness mechanism” for the unrecorded interactions. It expresses that an impossible interaction will never be recorded. It also specifies that a study is informative of whether an interaction is possible only if the species co-occur in the area and they are part of the study focus, as a way to account for geographic and taxonomic biases. Even if all of these hold, an interaction can still be unrecorded in the study with probability 1 − \( p_i q_j \). From (2) we see that the measured data are informative about species detectability through how often a given species is recorded to interact versus not among all studies and species of the other type for which such interaction is possible to be observed. Therefore, detectability scores are not informed by measured networks from studies for which the species is not of focus or does not interact with the focal species.

### 3.2.2. The Latent Factors for Model Specification

The likelihood discussed above cannot be used directly since it conditions on unmeasured quantities (the true interaction matrix, the unmeasured detectability scores and the latent covariates) along with the measured ones. Building toward an observed data likelihood, we specify a joint distribution over the unmeasured variables conditional on the measured ones (see Supplements B.2 and B.3 for the mathematical details). We assume that the indicators of species’ true interactions are independent conditional on species’ characteristics. Therefore, we ignore the possibility that species co-occurrence and competition might imply that an interaction occurs in one location might not occur in another. For the species’ detectability, we assume that it is independent across species, and only depends on individual characteristics and not those of other species. To specify the distribution of the latent features conditional on the measured covariates, we combine it with the likelihood of the measured traits, and specify instead the likelihood of the measured traits given the latent factors, and the marginal distribution of the latent features. Under these assumptions, we can write the distribution of the latent parameters times the likelihood of measured covariates,

\[
\prod_{i,j} p(q_j \mid V_j, W_j) p(W_j \mid V_j) \prod_{i} p(p_i \mid U_i, X_i) p(X_i \mid U_i)
\]

\[
\times \prod_{i,j} p(L_{ij} \mid U_i, V_j, X_i, W_j) p((U))p((V)),
\]

(3)

which is combined with (2) for the full distribution over our measured and latent variables. In Supplement B.3 we discuss how one could simultaneously model \( O \), and incorporate geographical covariates as predictors for species co-occurrence in different locations.

For appropriately chosen link functions \( f_m \) and \( g_l \), we assume the trait submodel:

\[
f_m^{-1}(E(X_{im} \mid U_i)) = \beta_{m0} + \sum_{m} U_i^{m} \beta_m,
\]

for \( \beta_{m0} \in \mathbb{R}, \beta_m \in \mathbb{R}^H, m = 1, 2, \ldots, p_B \), and

\[
g_l^{-1}(E(W_{jl} \mid V_j)) = \gamma_{l0} + \sum_{l} V_j^{l} \gamma_l,
\]

for \( \gamma_{l0} \in \mathbb{R}, \gamma_l \in \mathbb{R}^H, l = 1, 2, \ldots, p_P \).

(4)

We adopt logistic link functions for binary traits. For continuous traits, we use the identity link function, and we incorporate a parameter for the residual variance. Therefore, the latent factors are specified to be the driving force of birds’ and plants’ physical traits, and they can be conceived as low-dimensional summaries of the species’ traits. As long as the important information in the measured traits for detectability and species interactions is captured by the lower-dimensional latent factors, then the conditional distributions for \( p_i, q_j \), and \( L_{ij} \) in (3) can be specified to depend only on the latent factors. We do so below.

We specify the interaction submodel as

\[
\text{logit}P(L_{ij} = 1 \mid X_i, U_i, W_j, V_j) = \lambda_0 + \sum_{h=1}^{H} \lambda_h U_{ih} V_{jh}, \quad \text{for } \lambda_h \in \mathbb{R}, h = 0, 1, \ldots, H.
\]

(5)
In (5), the latent factors are used as in classic bipartite network models (e.g., Hoff 2011). Alternatively, one could allow for a different number of latent factors for each set of species and include them linearly in the interaction submodel. However, using the same number of factors $H$ allows us to conceive the interaction submodel (5) as a flexible representation of species’ interactions driven by interactions among the species’ “effective” traits. Since the role of traits in an ecological network is believed to be interactive (Fenster et al. 2015), we prefer this over the alternative.

The detection of species is believed to depend on species traits such as size and behavior (Garrard et al. 2013; Troscianko, Skelhorn, and Stevens 2017). A bird’s body mass, whether they are solitary or gregarious, and a plant’s height might affect whether their interactions are easily detected or not. For that reason, we specify the detection submodel to depend on the species’ latent factors (which act as a low-dimensional summary of the covariates) as

$$E[\text{logit}(p_h) \mid U_h, X_h] = \delta_0 + U_h^T \delta, \quad \text{and} \quad E[\text{logit}(q_h) \mid V_h, W_h] = \xi_0 + V_h^T \xi,$$

(6)

for $\delta_0, \xi_0 \in \mathbb{R}$, and $\delta, \xi \in \mathbb{R}^H$. We assume that $\text{logit}(p_h)$ and $\text{logit}(q_h)$ have conditional normal distributions with mean as in (6) and residual variance $\sigma_{\beta,h}^2$ and $\sigma_{\gamma,h}^2$, respectively.

Even though all latent factors are allowed to be drivers of traits in (4), true interactions in (5), and detectability in (6), different factors can be more or less important in each model component, and they might effectively contribute to only a subset of them if their corresponding coefficient is small (Supplement B.4). We discuss this further in Section 3.3.

### 3.3. Bayesian Inference

Our approach is placed within the Bayesian paradigm which allows for uncertainty quantification on the probability of truly possible interactions. The prior on the latent factors specifies that (a) the marginal variance of the latent factors is equal to 1, (b) a given species’ $H$ latent factors are independent, and (c) the latent factors across species are dependent with correlation that depends on their phylogeny. Parts (a) and (b) are common in latent factor models: since the latent factors are not identifiable parameters, restricting their scale does not affect model fit. Assuming that latent factors are a priori independent across $h$ allows them to capture different aspects of the species’ latent features, though it does not restrict them to being independent a posteriori. Latent factors are instead specified to be dependent across species: If $U_h = (U_{1h}, \ldots, U_{nh})^T$ and $V_h = (V_{1h}, \ldots, V_{nh})^T$ represent the collection of the $h$th coefficient across species, we specify $U_h \sim N(0, \Sigma_U)$ and $V_h \sim N(0, \Sigma_V)$, independently across $h$, but for $\Sigma_U = \rho_U \Sigma_U + (1 - \rho_U)I$, and similarly for $\Sigma_V$ where $I$ is diagonal and $\Sigma_U, \Sigma_V$ are the phylogenetic correlation matrices discussed in Section 2. We specify $\rho_U, \rho_V \sim \text{Beta}(a_\rho, b_\rho)$ with values near 0 or 1 representing close-to-independence and almost perfect phylogenetic dependence of the species’ latent factors, respectively.

Prior distributions need to be adopted for the remaining parameters which include the intercept, variance terms, and the coefficients of the latent factors in the models (4), (5), and (6). Due to the complete model’s high dimensionality for a moderate value of $H$, we adopt a prior distribution on model parameters which assigns increasing weight to values close to zero as the index $h$ increases. Specifically, we specify

$$\begin{align*}
\beta_{mh} | \tau_{mh}^{\beta}, \theta_h &\sim N(0, \tau_{mh}^{\beta}), \\
\lambda_{h} | \tau_{h}^{\lambda}, \theta_h &\sim N(0, \tau_{h}^{\lambda}), \\
\gamma_{h} | \tau_{h}^{\gamma}, \theta_h &\sim N(0, \tau_{h}^{\gamma}), \\
\delta_{h} | \tau_{h}^{\delta}, \theta_h &\sim N(0, \tau_{h}^{\delta}),
\end{align*}$$

(7)

where $\tau_{mh}^{\beta}, \tau_{h}^{\lambda}, \tau_{h}^{\gamma}, \tau_{h}^{\delta} \sim IG(v/2, v/2)$, and $\theta_h | \pi_h \sim (1 - \pi_h) P_0 + \pi_h \delta_{h, \infty}$. $\pi_h = \sum_{l=1}^{H} \omega_l$, $\omega_l = v_l \prod_{l=1}^{H}(1 - v_l)$, $v_l \sim \text{Beta}(1, \alpha)$, $t < H$ and $v_H = 1$.

In (7), the prior variance of model coefficients is specified using parameter-specific variance terms $\tau$ and overall variance terms $\theta$. Equation (8) specifies the truncated increasing shrinkage prior of Legramanti, Durante, and Dunson (2020), which uses a stick-breaking specification to define the mixing probabilities of a spike-and-slab prior distribution on $\theta_h$, where $P_0$ is a slab distribution, and $\delta_{h, \infty}$ represents a point-mass at $\theta_\infty$. We set $P_0$ to be an inverse gamma distribution, and set $\delta_{h, \infty}$ close to zero. This specification results in prior distributions for $\theta_h$ which assign larger weight to the point-mass rather than the slab distribution and are therefore concentrated closer to zero for larger values of $h$. The parameter-specific variance terms $\tau$ are centered at 1 and provide flexibility to the $l$th coefficient from each model to deviate from a $N(0, \theta_h)$ prior if this prior would lead to over-shrinkage of the corresponding coefficient. Therefore, the prior on $\theta_h$ is used to penalize more heavily the contribution of latent factors corresponding to a higher index $h$, essentially implying that not all the species’ information represented in the species’ latent factors will be important for detectability and species’ interactions, while the parameters $\tau$ adjust the prior variance to allow for additional flexibility in the coefficient of the latent factors across submodels. Inverse-gamma prior distributions are also assumed for the remaining residual variance parameters. Hyperparameter values are reported in Table S.2.

### 3.4. Posterior Computation

We sample from the posterior distribution of model parameters using Markov chain Monte Carlo (MCMC). Here we describe the algorithm at a high-level, but all details are included in Supplement C. At each MCMC step, the entries of the true interaction matrix for pairs with a recorded interaction are set to 1. The remaining entries are set to 1 or 0 with weights resembling the current values of (5) while reflecting that an unrecorded interaction among species that co-existed in multiple studies is more likely to be impossible. The parameters of the interaction model in (5) are updated using the Pólya-Gamma data augmentation scheme under which Pólya-Gamma random variables are drawn for all $n_g \times n_p$ pairs, conditional on which the posterior distributions of model parameters are normally distributed (Polson, Scott, and Windle 2013). Parameters of the models for binary traits are updated similarly. Despite the involvement of the latent factors in all submodels, the latent factors have normal posterior
distributions conditional on all other quantities. Updates for the parameters in the increasing shrinkage prior are adapted to our setting from Legramanti, Durante, and Dunson (2020). Species’ detectability scores are updated employing Metropolis–Hastings steps with a Beta proposal distribution centered at the current value. Despite the large number $(n_B + n_P)$ of parameters updated this way, these updates required minimal tuning. The parameters $\rho_U, \rho_V$ are updated similarly. Imputation of missing covariate values is based on (4).

We investigated the impact of out-of-sample species, and developed an algorithm which combines samples from the posterior distribution using the original data and an importance sampling step to predict interactions for these species. To avoid distraction from our main focus, we refer interested readers to supplement G.

4. Variable Importance in Latent Interaction Models

We propose a permutation-based approach to measure a covariate’s importance in latent factor network models. In our study, this procedure will inform us of the relative importance of species traits for forming interactions. We briefly discuss the approach here, though further details are included in Supplement D. We are interested in studying the importance of the $k$th bird trait. We use $X_k$ to denote the vector of the $k$th covariate across all bird species. For each $(i,j)$ pair of species, let $\mathbf{I}^{(r)}_{ij}$ be the logit of the $r$th posterior sample for the probability of interaction in (5), and $\mathbf{I}^{(r)}_{j}$ be the vector of these probabilities across $i$, $(\mathbf{I}^{(r)}_{ij1}, \mathbf{I}^{(r)}_{ij2}, \ldots, \mathbf{I}^{(r)}_{ijn_B})^T$. For each posterior sample $r$ and plant species $j$, we calculate the squared correlation between the predicted interaction probabilities $\mathbf{I}^{(r)}_{j}$ and the covariate $X_k$. We average these values over all plant species and posterior samples. For a large number of permutations $B$, we reorder the entries in $X_k$ and repeat this process. We use the number of standard deviations away from the mean of the permuted test statistics that the observed test statistic falls as a measure of variable importance. A similar approach is followed for the plant species $W$.

In our latent factor model, the latent factors and their coefficients are not identifiable parameters, and as a result we cannot interpret the magnitude of these coefficients as a variable importance metric. Even though the interaction probabilities are conditionally defined, and resampling methods belong generally outside the Bayesian paradigm, we find this resampling procedure to perform well in practice.

5. Simulations

5.1. The Setup: Data Generative Mechanisms Imitating the Observed Data

We perform simulations to study the impact of ignoring the taxonomic and geographical biases, to evaluate our approach under a variety of data generative mechanisms (DGMs), and compare its performance to that of alternative approaches. We consider 24 scenarios that are combinations of the following: (a) the same or different covariates drive interactions and detectability, (b) the important covariates are observed, some are observed and some unobserved, or all are unobserved, (c) the correlation among covariates is 0 or 0.3, and (d) there is low or high information, corresponding to species co-occurrence and recorded interactions that are more or less sparse. Choices of (a) allow us to evaluate whether the performance of our model is hindered by the fact that it uses the same latent factors in all submodels. Detailed information on the DGMs and additional simulation results including simulations on the variable importance metric are included in Supplement F, and are summarized below.

Our simulations are based on our data on recorded bird–plant interactions in terms of the observed number of species and studies, and the structure of measured covariates. We generate covariates $\tilde{X}, \tilde{W}$ from a matrix-normal distribution, with correlation across covariates equal to 0 or 0.3, and correlation across species resembling the species’ phylogenetic correlation matrices. Some of the covariates were then transformed to binary variables using their initial values as linear predictors in a Bernoulli distribution with a logistic link function. Only a subset of the generated covariates are available in the simulated data, and the rest are considered unmeasured. For the measured covariates, we maintain the same structure and proportion of missingness as in the observed data: 2 continuous and 3 binary covariates with proportion of missing values varying from 0%–32% for bird species, and 4 continuous and 8 binary covariates with proportion of missing values varying from 0% to 80% for plant species. The interaction submodel and the detectability submodels are specified as multiplicative and linear in $\tilde{X}, \tilde{W}$, respectively. The important covariates in the models can be the same or different, measured or unmeasured, and measured covariates might be interacting with unmeasured covariates. For example, in DGM2 the measured $\tilde{X}_{ij}$ interacts with the unmeasured $\tilde{W}_{j13}$. The set of unmeasured covariates includes the same number of binary and continuous covariates as the set of measured ones. Across all scenarios, the true interaction model achieves AUROC equal to 0.78. The six combinations of (a)–(b) correspond to DGM1–6 shown in Table 1, where we also show which covariates are included in the interaction and detectability submodels.

5.2. The Setup: Alternative Approaches

We focus on comparing the proposed approach to an alternative approach which uses covariates directly. We also considered alterations of our model where versions of it (a) fix the number of latent factors $H$, (b) exclude the parameters $\tau$, and (c) allow the covariates to inform the latent factors only through (4), cutting the feedback from the interaction and detectability submodels (Jacob et al. 2017). We also considered an approach that uses both covariates and latent factors in the interaction submodel, and our approach and the covariates approach while assuming that the observed interaction network is not measured with error and bias correction is not performed. We present all models in Supplement E. Due to space constraints, we include the results from these approaches in detail in the supplement, and we summarize them below. Note that the competing method that uses covariates directly is our own construction and it does not exist in the literature, and that the models that are based directly on the covariates do not incorporate phylogenetic information.
Table 1. Simulations setup. Variables included in the interaction and detectability submodels. ✓ indicates that the covariate was used in the interaction model. Shaded cells indicate that the covariate was used in the model for species detectability.

| Description | Bird covariates | Plant covariates |
|-------------|-----------------|------------------|
|             | Cont. | Binary | Cont. | Binary |
| DGM1        | same & measured | measured | unmeasured | ✓ | ✓ | ✓ | ✓ | ✓ |
| DGM2        | same & measured | measured | unmeasured | ✓ | ✓ | ✓ | ✓ | ✓ |
| DGM3        | same & measured | measured | unmeasured | ✓ | ✓ | ✓ | ✓ | ✓ |
| DGM4        | different & measured | measured | unmeasured | ✓ | ✓ | ✓ | ✓ | ✓ |
| DGM5        | different & measured | measured | unmeasured | ✓ | ✓ | ✓ | ✓ | ✓ |
| DGM6        | different & measured | measured | unmeasured | ✓ | ✓ | ✓ | ✓ | ✓ |

Figure 2. Predictive performance in simulations. The methods considered use latent factors or observed covariates (horizontal axis). The method using latent factors is the proposed approach. The columns represent the 6 DGMs in Table 1. The rows correspond to combinations of the high and low signal scenarios and the two correlation values. Results are shown by observational effort for pairs of species by color.

5.3. Simulation Results

Methods were evaluated in terms of their predictive power in identifying true interactions. Figure 2 shows the simulation results in terms of the AUROC (area under the receiver operating characteristics curve) when predicting the values of $L$ among pairs with unrecorded interactions, separately by DGM, and amount of observational effort defined as the number of studies that could have recorded the interaction if it was observed, $\sum_i F_{ij} O^{ij}_i O^{ij}_j$. Even though the AUROC is not a Bayesian criterion, it has been used before in a related setting (Sosa and Betancourt 2022), and it is not clear how one could use alternatives like the WAIC (Watanabe 2010) since its computation for network data is complicated (Gelman, Hwang, and Vehtari 2014) and our network of interest (the matrix $L$) is latent.

First, we notice that the performance of both methods improves with higher observational effort, implying that both models accommodate that an unrecorded interaction that was possible to be recorded across many studies is most likely not possible. Therefore, focusing on pairs of species on the lower end of observational effort compares the model structure more directly. Across all 24 scenarios considered and across the spectrum of observational effort, our approach that uses latent
factors performs better than or comparably to the method that uses covariates directly, and all alternatives considered in the supplement. The improvement by using the latent factors is most visible for the low observational effort pairs. The performance of our approach is essentially unaltered by whether the same or different covariates drive the different submodels, irrespective of whether these covariates are measured or correlated (DGM 1 vs. 4, 2 vs. 5, and 3 vs. 6). Also, its performance is very similar when the covariates are correlated or not (comparing rows 1–2, and 3–4). The only exception is in the low information setting when all important covariates are unmeasured (rows 3 and 4, DGM 3 and 6), in which case correlation among covariates might improve the performance of the latent factor model. When the important covariates are measured, the performance of our model improves (comparing DGMs 1 through 3, and 4 through 6), though in the high information setting the model can learn interaction profiles even when the important covariates are all unmeasured and uncorrelated with the measured ones (DGM6, High, Uncorrelated). These results inform us that the latent factor model performs better than using covariates directly across a variety of scenarios, and agree with prior work discussing in Section 4 accurately identifies the covariates that are important for forming interactions, without specifying the functional form in which covariates drive interactivity. We find that variable importance should be interpreted separately for continuous and binary covariates. Our approach to variable importance is based on resampling techniques, and as a result it is arguably not-fully Bayesian. As an alternative we investigated variable importance for the model that includes covariates and latent factors. Apart from requiring a parametric specification of how covariates are included in the model, we find that using the coefficients of the covariates from this model for variable importance is flawed. This issue is related to spatial confounding in the spatial literature and arises due to collinearity of the phylogenetically-correlated covariates and latent factors (see Van Ee, Ivan, and Hooten 2022, for a discussion on spatial confounding in ecology, and references therein).

Across our simulations, we found that 1000 MCMC iterations took on average 89 min. In the supplement, we also investigate the computational time of the proposed approach when varying the number of species and number of individual studies.

### 6. Bird–Plant Interactions in the Atlantic Forest

We considered the two approaches that correct for taxonomic and geographic bias discussed in Section 5, and we use the increasing shrinkage prior on the latent factor coefficients. We specify \( O_{ij} = O^B_{ij} O^P_{ij} \), where \( O^B \) is an \( n_S \times n_S \) binary occurrence matrix for birds, and similarly for \( O^P \). These are assumed known with entries equal to 1 if the species has a recorded interaction in the study and 0 otherwise. Due to a large number of recorded interactions with missing coordinate information, we are unable to include environmental or geographical covariates, though we discuss extensions in that direction in supplement B.2. We ran four chains of 80,000 iterations each, with a 40,000 burn in, and kept every 40th iteration. For our approach, 1000 iterations took on average 86 min. MCMC convergence was investigated by studying traceplots and running means for identifiable parameters. Convergence diagnostics are shown in Supplement I. Based on similar diagnostics, we found that the MCMC of the alternative approach failed to converge based on the same number of iterations. For that reason, we excluded from this analysis the two traits of the plant species with the largest amounts of missingness (seed length and whether the species is threatened for extinction) which led to no detectable lack of convergence.

In Figure 3, we show estimates for the probability of interaction for species in the largest taxonomic families. According to our model (Figure 3(a)), species in the same family form similar interactions as evidenced by the taxonomically-structured posterior interaction probabilities where the blue lines separate them in clusters with similar values. In contrast, results from the alternative approach that employs covariates directly (Figure 3(b)) indicate that some species interact with most other species and some species with none, as evidenced by rows and columns that are mostly close to one or zero. Since we do not expect this “all or none” structure in species interactions, results from the covariate approach seem untrustworthy, and indicate that it might rely on covariates too heavily. Species within the same family can belong to different genera, though genera are not shown in the figure to ease visualization. However, we observed that clusters of posterior interaction probabilities from the latent factor model within the depicted taxonomic families generally correspond to species organization by genera, supporting that interactions are taxonomically structured. The taxonomic structure is further supported by posterior means (95% credible intervals) for \( \rho_U \) and \( \rho_V \), which were 0.97 (0.947, 0.988) and 0.95 (0.935, 0.97), respectively. In Supplement H we show that the results remain unchanged when using an alternative specification of the intra-species correlation matrix based on species’ taxonomic relationships.

### 5.4. Results from Additional Simulation Studies

Here, we summarize some additional simulation results, all of which are shown in the supplement. We have found that our approach performs better (or equally as well) compared to all other approaches considered. In all scenarios, approaches that ignore the taxonomic and geographical biases lead to very poor performance for predicting missing interactions which deteriorates for species with a higher observational effort, and a smaller number of predicted possible interactions compared to their counterparts with bias correction (also noted by Weinstein and Graham 2017; Graham and Weinstein 2018). Using a higher number of latent factors \( H \) with sufficient shrinkage, and incorporating the variance parameters \( \tau \) improve the performance of the proposed approach. Cutting the feedback among the submodels for informing the latent factors performed better only in sparse settings and for pairs of species with high observational effort, indicating that the measured interactions can be helpful in informing the latent factors for predicting missing interactions.

We also found that the variable importance metric introduced in Section 4 accurately identifies the covariates that are important for forming interactions, without specifying the functional form in which covariates drive interactivity. We find that variable importance should be interpreted separately for continuous and binary covariates. Our approach to variable importance is based on resampling techniques, and as a result it is arguably not-fully Bayesian. As an alternative we investigated variable importance for the model that includes covariates and latent factors. Apart from requiring a parametric specification of how covariates are included in the model, we find that using the coefficients of the covariates from this model for variable importance is flawed. This issue is related to spatial confounding in the spatial literature and arises due to collinearity of the phylogenetically-correlated covariates and latent factors (see Van Ee, Ivan, and Hooten 2022, for a discussion on spatial confounding in ecology, and references therein).
The two approaches often return opposite conclusions about species’ interactions. The latent factor approach almost always returns probabilities of interaction that are lower than those from the covariate approach. The covariate approach predicts that 18% of pairs interact (posterior probability above 80%), and only 9% of pairs do not (posterior probability below 10%), both unrealistic. In contrast, the latent factor model predicts that 5% of pairs interact, and 41% do not. The vast majority of pairs that are predicted to not interact under the covariate model are also predicted to not interact based on the latent factor model, though the latent factor approach has substantially lower posterior standard deviation in these predictions. A more in-depth comparison is given in Supplement H.

To compare model performance directly, we applied a variant of cross-validation. We randomly choose 100 recorded interactions, we set their corresponding values in the observed interaction matrix equal to 0, and we predict their probability of interaction. We repeat this procedure 30 times, each time holding out a different subset of recorded interactions. Our setting forbids us from comparing model performance based on unrecorded interactions, since those interactions are not certainly impossible. Our comparison is based on how well each approach can differentiate the held-out pairs of species that truly interact from the group of all pairs, which necessarily includes pairs that do not interact. Since the two approaches return drastically different prevalence of interactions, we evaluate the relative magnitude of posterior interaction probabilities in the held-out and in the overall data. For the covariate approach, the mean and median posterior probability of interaction for the held out pairs was on average 1.21 and 1.36 times higher than the corresponding value across all pairs of species. In contrast, those numbers where substantially higher and equal to 1.85 and 3.19 for the mean and median, respectively, for our approach. Therefore, our approach is much more effective in differentiating the pairs that are truly interactive from the set of all pairs compared to the approach that uses covariates directly.

### 6.2. The Importance of Traits and Phylogeny for Species Interactivity

Apart from understanding which pairs of species are interactive, ecologists are also interested in understanding the traits which make species interactions possible (Garrard et al. 2013; Bastazini et al. 2017; Troscianko, Skelhorn, and Stevens 2017). Figure 4(a)–(b) show the variable importance metric described in Section 4 for bird and plant traits. We identify a bird’s body mass and a plant’s fruit diameter as the most important continuous traits in forming interactions. In Figure 4(c) we plot the posterior
probabilities of interaction reordering the species in increasing values of the two covariates. High posterior probabilities are concentrated on the upper left triangle, indicating that a given bird would interact with most plant species that are smaller than some threshold size, in line with the current ecological literature (Fenster et al. 2015). At the same time, there seems to be some preference for larger birds to not consume fruits that are too small, indicating a matching-size type of behavior for forming interactions. These results illustrate that our approach can identify complicated interactive relationships without having to specify these trends parameterically.

We studied the overall importance of these traits and phylogenetic information using the cross-validation technique discussed in Section 6.1. For trait importance, we excluded each of the traits from the available information, separately, and for the phylogenetic information we set $\rho_U, \rho_V$ to 0 which forces the latent factors to not be phylogenetically structured a priori. Excluding bird mass or fruit diameter returned on average a median posterior probability of interaction for the held out pairs 3.17 and 3.06 times higher, respectively, than the corresponding value across all pairs, compared to 3.19 when all traits are included. Therefore, fruit diameter can be an important covariate to measure for predicting species interactions. When ignoring phylogenetic information the corresponding value was 1.66 illustrating that phylogenetic information is crucial for predicting missing interactions.

7. Discussion

We introduced an approach based on latent factors that uses species traits and recorded interactions to complete the bipartite graph of species interdependence accounting for the taxonomic and geographic biases of individual studies, and we proposed an approach to study variable importance in latent network models. We found that using covariates to inform the latent factors performs better in predicting pairs of species that do not interact and separating those that interact from the rest, compared to using the covariates directly. Even though using the covariates in the proposed manner complicates the investigation of variable importance, we proposed a variable importance metric which performed well in simulations and identified important physical traits for species interdependence that are in line with ecological knowledge.

A possible extension to our model could accommodate simultaneous modeling of species co-occurrence, that would allow us to incorporate geographic information and other environmental variables that define the environmental niche of the species such as temperature, precipitation, and evapotranspiration (Gravel et al. 2019). Even though we provide an overview of such an approach in the supplement, studying the co-existence of species across space is a hard problem in itself and it is the topic of joint species distribution modeling in ecology (Ovaskainen and Abrego 2020). Importantly, modeling species co-occurrence and interactivity simultaneously and allowing for different interaction profiles based on environmental and geographical covariates would open the road to investigating the importance of species abundance, co-occurrence and competition in forming interactions. We find this to be an exciting line of future work.

Supplementary Materials

Code and data for replicating the study results are available at https://github.com/gpapadog/Bird_Plant_Interactions. An R package that implements the proposed method is available at https://github.com/gpapadog/BiasedNetwork. The supplementary materials include supporting information, mathematical derivations, and additional simulation and study results. Supplement A includes a glossary. Supplement B includes mathematical derivations of the observed data likelihood and prior distributions. In Supplement C we provide the details of our MCMC procedure. In Supplement D we describe the variable importance metric and corresponding computations in more detail. In Supplement E we describe all the alternative models that we consider in our simulations, and Supplement F includes additional simulation results. Supplement G includes a discussion about the impact of out-of-sample species. In Supplement H, we include additional results from our study on bird-plant interactions, MCMC convergence diagnostics are shown in Supplement I, and a list of all bird and plant species in our study is given in Supplement J.

Disclosure Statement

The authors report that there are no competing interests to declare.

Funding

This project has received funding from the European Research Council (ERC) under the European Union Horizon 2020 research and innovation programme (grant agreement No 856506; ERC-synergy project LIFEPLAN). Otso Ovaskainen was funded by Academy of Finland (grant no. 309581), Jane and Aatos Erkko Foundation, Research Council of Norway through its Centres of Excellence Funding Scheme (223257). Carolina Bello acknowledges funding support from the European Research Council (ERC) under the European union’s Horizon 2020 research and innovation programme (grant agreement No 787638) and the Swiss National Science Foundation (grant No. 173342), both granted to Catherine Graham.

ORCID

Georgia Papadog Georgiou http://orcid.org/0000-0002-1982-2245

References

Aizen, M. A., Sabatino, M., and Tylianakis, J. M. (2012), "Specialization and Rarity Predict Nonrandom Loss of Interactions from Mutualist Networks," Science, 335, 1486–1489. [2250]

Báldi, A., and McCollin, D. (2003), "Island Ecology and Contingent Theory: The Role of Spatial Scale and Taxonomic Bias," Global Ecology and Biogeography, 12, 1–3. [2250]

Bartomeus, I. (2013), “Understanding Linkage Rules in Plant-Pollinator Networks by Using Hierarchical Models that Incorporate Pollinator Detectability and Plant Taits,” PloS One, 8, e69200. [2250]

Bastazini, V. A., Ferreira, P. M., Azambuja, B. O., Casas, G., Debastiani, V. J., Guimarães, P. R., and Pillar, V. D. (2017), "Untangling the Tangled Bank: A Novel Method for Partitioning the Effects of Phylogenies and Traits on Ecological Networks," Evolutionary Biology, 44, 312–324. [2258]

Bello, C., Galetti, M., Pizo, M. A., Marigliana, T. C., Culot, L., Bufalo, F., Labeca, F., Pedroso, F., Constantini, R., Emer, C., Silva, W. R., Silva, F. R., Ovaskainen, O., and Jordano, P. (2017), "ATLANTIC-FRUGIVORY: A Plant-Interaction Dataset for the Atlantic Forest," Ecology, 98, 1729. [2251]

Benadi, G., Dormann, C. F., Fründ, J., Stephan, R., and Vázquez, D. P. (2022), “Quantitative Prediction of Interactions in Bipartite Networks based on Traits, Abundances, and Phylogeny,” The American Naturalist, 199, 841–854. [2252]

Blonder, B., and Dornhaus, A. (2011), “Time-Ordered Networks Reveal Limitations to Information Flow in Ant Colonies," PloS One, 6, e101031. [2251]

Bullmore, E., and Sporns, O. (2009), “Complex Brain Networks: Graph Theoretical Analysis of Structural and Functional Systems," Nature Reviews Neuroscience, 10, 186–198. [2251]
Van Ee, J. J., Ivan, J. S., and Hooten, M. B. (2022), “Community Confounding in Joint Species Distribution Models,” Scientific Reports, 12, 1–14. [2257]

Wang, D. J., Shi, X., McFarland, D. A., and Leskovec, J. (2012), “Measurement Error in Network Data: A Re-Classification,” Social Networks, 34, 396–409. [2251]

Wang, P., Laskey, K. B., Domeniconi, C., and Jordan, M. I. (2011), “Nonparametric Bayesian Co-clustering Ensembles,” in Proceedings of the 2011 SIAM International Conference on Data Mining, Society for Industrial and Applied Mathematics, pp. 331–342. [2251]

Watanabe, S. (2010), “Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory,” Journal of Machine Learning Research, 11, 3571–3594. [2256]

Weinstein, B. G., and Graham, C. H. (2017), “On Comparing Taits and Abundance for Predicting Species Interactions with Imperfect Detection,” Food Webs, 11, 17–25. [2250,2253,2257]

Wu, Y., Zhou, C., Xiao, J., Kurths, J., and Schellnhuber, H. J. (2010), “Evidence for a Bimodal Distribution in Human Communication,” Proceedings of the National Academy of Sciences, 107, 18803–18808. [2251]