

# Modeling the Building Blocks of Biodiversity

Lucas N. Joppa<sup>1\*</sup>, Rich Williams<sup>2</sup>

**1** Computational Ecology and Environmental Sciences, Microsoft Research, Cambridge, United Kingdom, **2** Quid.com, San Francisco, California, United States of America

## Abstract

**Background:** Networks of single interaction types, such as plant-pollinator mutualisms, are biodiversity's "building blocks". Yet, the structure of mutualistic and antagonistic networks differs, leaving no unified modeling framework across biodiversity's component pieces.

**Methods/Principal Findings:** We use a one-dimensional "niche model" to predict antagonistic and mutualistic species interactions, finding that accuracy decreases with the size of the network. We show that properties of the modeled network structure closely approximate empirical properties even where individual interactions are poorly predicted. Further, some aspects of the structure of the niche space were consistently different between network classes.

**Conclusions/Significance:** These novel results reveal fundamental differences between the ability to predict ecologically important features of the overall structure of a network and the ability to predict pair-wise species interactions.

**Citation:** Joppa LN, Williams R (2013) Modeling the Building Blocks of Biodiversity. PLoS ONE 8(2): e56277. doi:10.1371/journal.pone.0056277

**Editor:** Jerome Chave, Centre National de la Recherche Scientifique, France

**Received:** March 23, 2012; **Accepted:** January 12, 2013; **Published:** February 27, 2013

**Copyright:** © 2013 Joppa, Williams. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The authors have no support or funding to report.

**Competing Interests:** All authors were employed by Microsoft Research when this study was done. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

\* E-mail: [ljoppa@microsoft.com](mailto:ljoppa@microsoft.com)

## Introduction

Ecological networks describe who interacts with whom in ecological communities [1]. Subsets of these networks where a single type of interaction is considered, such as mutualistic interactions between plants and their pollinators [2], or antagonistic interactions between parasites and their hosts [3], are considered to be the true 'building blocks' of biodiversity [2–5]. These 'building blocks' provide vital ecosystem services [6], and understanding the way they are structured is important for gaining insight into everything from ecosystem function [7–10], to the response of ecosystems to species extinctions [11,12], invasions [13,14], and the influence of keystone species [15].

Theory suggests that the architecture of mutualistic and antagonistic species interaction networks should differ [2,3]. Empirical evidence supports this, and highlights structural differences between these two types of networks in ways important to their stability and fragility [2,3]. Moreover, there is a practice of using different structural models for mutualistic [16–19] and multi-trophic networks (i.e., 'food webs') [20–22], even though doing so is a purely methodological convenience unrelated to the true nature of species interactions [23]. A simple structural model capable of replicating these different types of networks would provide a single conceptual framework for linking these 'building blocks' in studies of population dynamics [24], ecosystem regime shifts [25], and other basic and applied ecological questions.

Here we fit a simple, probabilistic niche-structured model [26,27] to 151 empirical networks to replicate the link structure of three different classes of bipartite ecological networks (mutualistic:  $n = 67$ , parasitic – antagonistic:  $n = 40$ , herbivory – antagonistic:  $n = 44$ ). Further details on the datasets analyzed have been published previously [27–29]. While the structure of these classes

of networks differs [2,30], the rationale for a general model is simple: since a simple one-dimensional niche axis can consistently explain the structure of multi-trophic level food webs [20,31], one might also expect it to explain the structure of antagonistic plant-animal networks derived from these networks. Moreover, just as body size is a common explanatory trait for niche position in "food web" antagonistic networks [32], interactions in networks of mutualistic relationships like plants and their pollinators are often described as being controlled by one or a few traits in both the plant and the pollinator (e.g. length of flower corolla pollinator proboscis [33]). These examples suggest that a low-dimensional niche-structured model might explain much of the structure of all three classes of networks. The objective of this paper is to address whether or not such a low-dimensional model sufficiently explains the structure of ecological networks.

## Model

### Bipartite Probabilistic Niche Model

We developed a simple extension of the probabilistic niche model [26,27] (PNM) for bipartite networks. Our model is complementary to one proposed earlier [34] by making use of a niche axis on which each species is positioned and where each species interacts preferentially with species nearby on this axis. That ours is a probabilistic version brings it in line with another recent model [35], although our model differs in two major ways as our unit of interaction is the species while [35] considers individuals, and our interaction function is much more general.

By freeing up the constraints on niche parameters imposed by the original niche model [20], the PNM provides significantly better fit to empirical food web data than previous models, and allows for analysis of the resulting niche-related model parameters.

In our bipartite probabilistic niche model (BPNM; Figure 1) each resource is placed on a niche axis and the probability that a consumer interacts with a resource is a function of the relative distance between the resource's position on the axis and the position of the consumer's niche on that axis. We consider bipartite networks with  $L$  links between  $S_C$  consumer species and  $S_R$  resource species. The bipartite probabilistic niche model defines the probability of interaction between resource species  $i$  and consumer species  $j$ . The resource species are placed on a one-dimensional niche axis; the niche position of resource species  $j$  is defined as  $n_j$ . Each consumer species  $i$  has two traits, its niche center  $c_i$  and its niche range  $r_i$ . Figure 1 shows a one-dimensional version of this model. The probability of consumer  $i$  interacting with resource  $j$  is a function of the relative distance between the resource's niche position and the consumer's niche center:

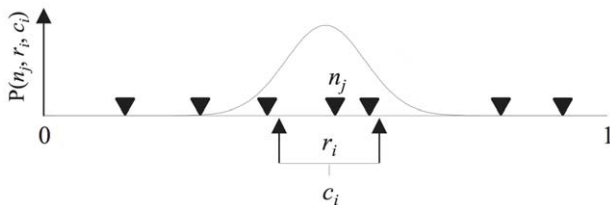
$$P(i,j|\theta) = \alpha \exp\left(-\left|\frac{n_j - c_i}{r_i/2}\right|^e\right) \quad (1)$$

where  $P(i,j|\theta)$  is the probability that consumer species  $i$  interacts with resource species  $j$  given a particular parameter set  $\theta$  where  $\theta = \{n_1 \dots n_{S_R}, c_1 \dots c_{S_C}, r_1 \dots r_{S_C}, e\}$ ; the parameter  $n_j$  is the niche position of resource species  $j$ ; the parameter  $c_i$  is the center of the niche of consumer species  $i$ ; the parameter  $r_i$  is the width of the niche of consumer species  $i$ ; the parameter  $e$  varies the cutoff rate of the niche probability function (for larger values, the niche probability function is flatter in the center and cuts off more quickly at the edges of the niche); and the parameter  $\alpha$  is the probability that  $i$  eats  $j$  when  $j$  is exactly on  $i$ 's niche optimum (i.e. when  $n_j = c_i$ ).

**Statistical methods.** The set of model parameter values for a network with  $S_C$  consumer species and  $S_R$  resource species is given by  $\theta$ , while  $X$  is the data, i.e.,  $X$  is an  $S_C \times S_R$  connection matrix containing an observation  $X_{ij}$  for each link  $i, j$  ( $X_{ij} = 1$  means  $i$  interacts with  $j$ ;  $X_{ij} = 0$  means  $i$  does not interact with  $j$ ). The log-likelihood is defined as:

$$\ell(X|\theta) = \sum_i \sum_j \ln \left\{ \begin{array}{l} P(i,j|\theta) \text{ if } X_{ij} = 1 \\ 1 - P(i,j|\theta) \text{ if } X_{ij} = 0 \end{array} \right\} \quad (2)$$

In each of the models we use, every possible interaction between species pairs is assigned a probability of occurring. This allows model performance to be calculated in a straightforward likelihood framework, and for all models, we find the maximum likelihood parameter values for each empirical network. Using likelihoods to



**Figure 1. Diagram of the probabilistic niche model for bipartite networks.** Each resource species has a single parameter  $n_j$ ; each consumer species has two parameters,  $c_i$  giving the location on the niche axis where it has the highest probability of interacting with a resource species and a distribution width  $r_i$ . The probability that consumer species  $i$  interacts with resource species  $j$  is defined by the probability  $P(n_j, r_i, c_i)$ , here a Gaussian. doi:10.1371/journal.pone.0056277.g001

evaluate model performance effectively separates the model's performance evaluation from any summary metrics of network structure, a distinction that we will show to be extremely important.

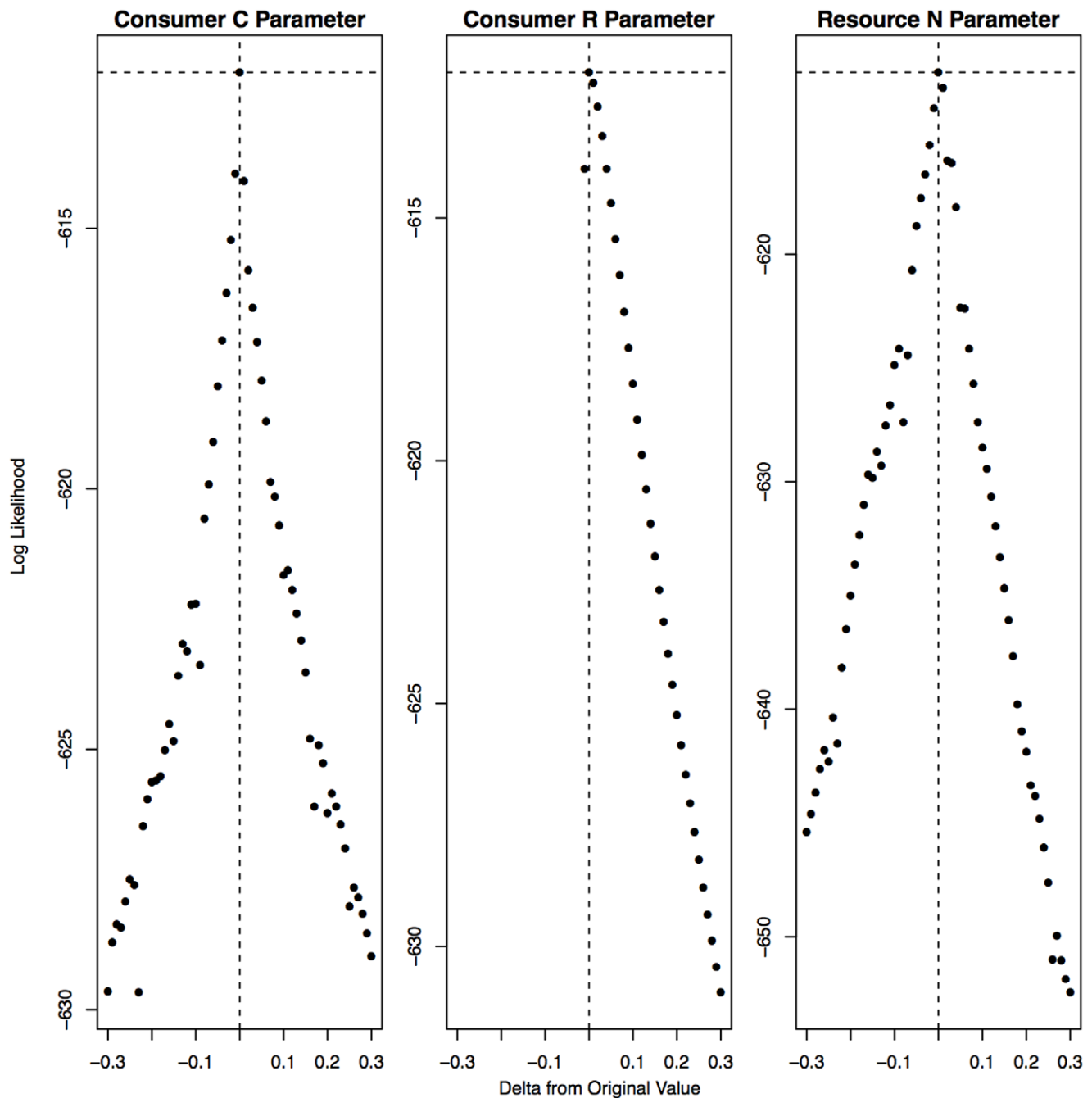
### Potential Overparameterization

For all models we use simulated annealing [36] to find the maximum likelihood parameter set given the observed feeding relationships. The number of parameters in the BPNM scales as  $(2 * S_C) + S_R + 2$ , and the number of binary observations scales as  $S_C * S_R$ . This means that there are relatively few observations per parameter for some of the less species-rich datasets we analyze, which leads to a potentially overparameterized model. Overparameterization might cause the BPNM to not be the best minimal model for explaining all aspects of ecological bipartite networks, and some parameters might be poorly estimated – necessitating caution when interpreting parameter values. The BPNM also has, to varying degrees, different numbers of parameters than the two other models we compare it to. We deal with these issues in several ways.

First, since the different models in this analysis have different numbers of parameters, their relative performance on a single network is compared using AICc rather than by direct comparison of likelihoods. AICc allows the comparison of models with different numbers of parameters and includes a correction for small sample size. Second, we visually inspected the sensitivity of parameters by choosing several networks at random and, for each consumer and each resource, changing a particular parameter ( $r, c$ , or  $n$ ) from  $-0.3$  to  $+0.3$  away from the maximum likelihood value, while holding all other parameters constant at their maximum likelihood values. One can assess sensitivity by analyzing the change in the likelihood surface moving towards and away from the maximum likelihood parameter. In Figure 2 we show examples of this for the  $r$  and  $c$  parameters of a randomly chosen consumer and the  $n$  parameter of a randomly chosen resource species. In these three cases one sees the characteristic peak expected at the maximum likelihood value. Not all parameters are as sensitive as these, but even so, the problem of overparameterization is only likely to make our central conclusion below (that the BPNM outperforms other simpler models) overly conservative. Finally, previous work [26] explored an extension of the PNM, reducing parameters by making certain parameters functions of others, and when comparing models using AIC found that the PNM with its full parameter set consistently outperformed other, more minimal, simplifications of the same model.

### Other Models

We compared the performance of the BPNM against two simpler models. The simplest is a random model, in which every link occurs with constant probability  $P = C_B = \frac{L}{S_R * S_C}$ . A recent study suggested that a bipartite version of the cascade model [21] often provides a good model of mutualistic networks, and so we developed a probabilistic bipartite cascade model. Each consumer species is assigned a parameter  $C_i$  and each resource species is assigned a parameter  $R_j$ . An interaction between consumer and resource is only possible if  $R_j < C_i$  and the number of possible interactions  $N$  is the number of plant-animal pairs in which  $R_j < C_i$ . To force the expected number of links to be equal to the observed number of links we define  $P(i,j|\theta) = p_c$  if  $R_j < C_i$  and  $P(i,j|\theta)$  otherwise. The fixed link probability  $P_C = \frac{L}{N}$  if  $N > L$  and  $P_C = 1$  otherwise. As in the BPNM, we find the maximum likelihood set of model parameters  $\theta = \{R_1 \dots R_{S_R}, C_1 \dots C_{S_C}\}$ .

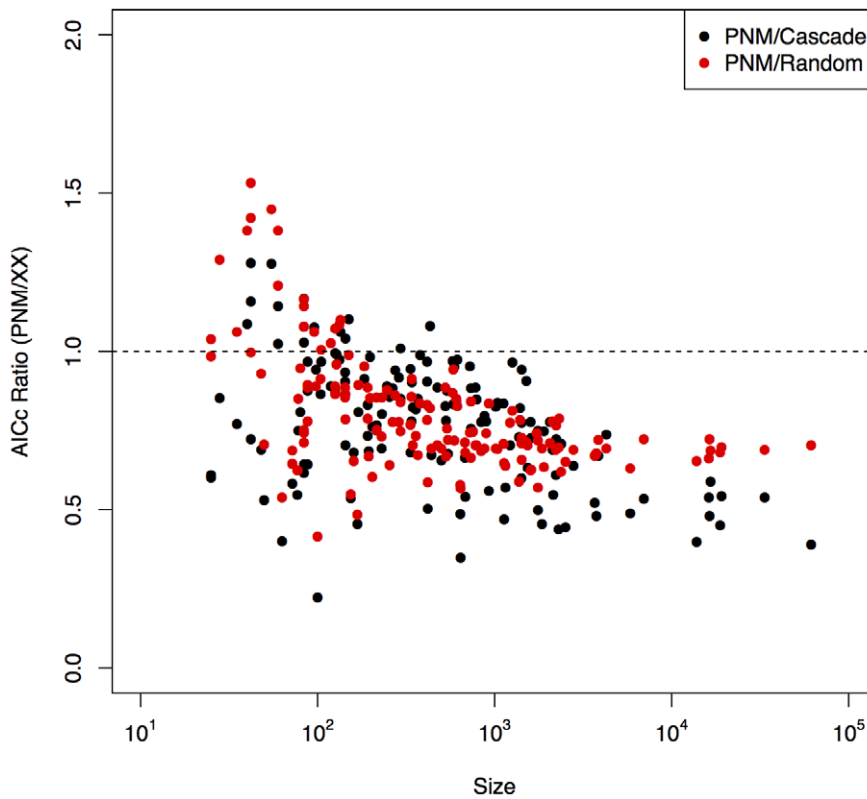


**Figure 2. Parameter sensitivity for one randomly chosen consumer, and one randomly chosen resource species.** The maximum likelihood value of the consumer parameters ( $c$  – left,  $r$  – middle) and resource parameter ( $n$  – right) is indicated by the vertical dashed line. Each parameter was perturbed across the continuum (x-axis) of  $-0.3$  to  $0.3$  away from the maximum likelihood value (while holding all other parameters in the model constant). The corresponding likelihoods of each assessed parameter value are on the y-axis. Some parameters were already close to the value 0 and could not be changed to a smaller value, explaining why the distributions are not symmetrical in the centre plot. doi:10.1371/journal.pone.0056277.g002

### Niche Overlap Overview

The degree of dietary overlap between all consumers is a commonly measured structural property of ecological networks. In particular, ‘nestedness’ is a measure of the structure of a binary matrix, where, in this case, the rows represent resource species ( $S_R$ ), the columns represent consumer species ( $S_C$ ), and a “1” in the matrix records the interaction between consumer  $i$  and resource  $j$ . There are various metrics to calculate nestedness, all of

which to some degree attempt to measure the niche overlap between consumers, or in other words, the extent to which the resource species of specialist consumers are proper subsets of the resource species of more generalist consumers (and vice-versa). To determine the BPNM’s predictive skill for the network aggregate metric of niche overlap, we first calculated the niche overlap of consumers in the empirical network using a standard nestedness



**Figure 3. Performance of the random (red) and cascade (black) models as measured by AICc score.** Results are presented relative to the PNM, where the x-axis shows the PNM AICc divided by the alternative model’s AICc. The random model outperforms the PNM only once, while the cascade model does so three times, all of which occur on the lower 50% of the size distribution of networks we analyze. doi:10.1371/journal.pone.0056277.g003

metric, and compared that to two different approaches to calculate the niche overlap of the model-derived network.

**Empirical Niche Overlap**

For our analysis we utilize the metric “NODF”, due to the transparency and statistical properties of the metric [30,37–40] to measure the niche overlap of empirical networks. The original NODF algorithm calculates nestedness for the entire matrix, and independently for only rows (resources) or columns (consumers). The BPNM only calculates the niche range for consumers, and so we record only the consumer NODF value for each matrix (NODF<sub>C</sub>). Each bipartite network was sorted in descending order by both row and column marginal totals. The NODF value of a community of consumer species is a function of the paired overlap between all combinations of the  $S_C$  columns. The number of possible combinations is:

$$P_C = S_C \left( \frac{S_C - 1}{2} \right)$$

The paired overlap ( $PO_{Cij}$ ) between any two ( $i, j$ ) consumers is the fraction of resources taken by consumer  $j$  that are also taken by consumer  $i$ . Note that the sorting described above means that it is always the case that the number of resources taken by consumer  $i$  is always greater than or equal to the number of resources taken by consumer  $j$ . Defining  $N_R$  as the number of resource species taken

either by consumer  $i$  or  $j$ , NODF then considers the paired overlap in consumer pairs, defining

$NP_{Cij} = PO_{Cij}$  if  $N_{Rj} > N_{Ri}$ , where  $NP_{Cij}$  is not recorded (“NA”) in the event of  $N_{Ri} = N_{Rj}$ .

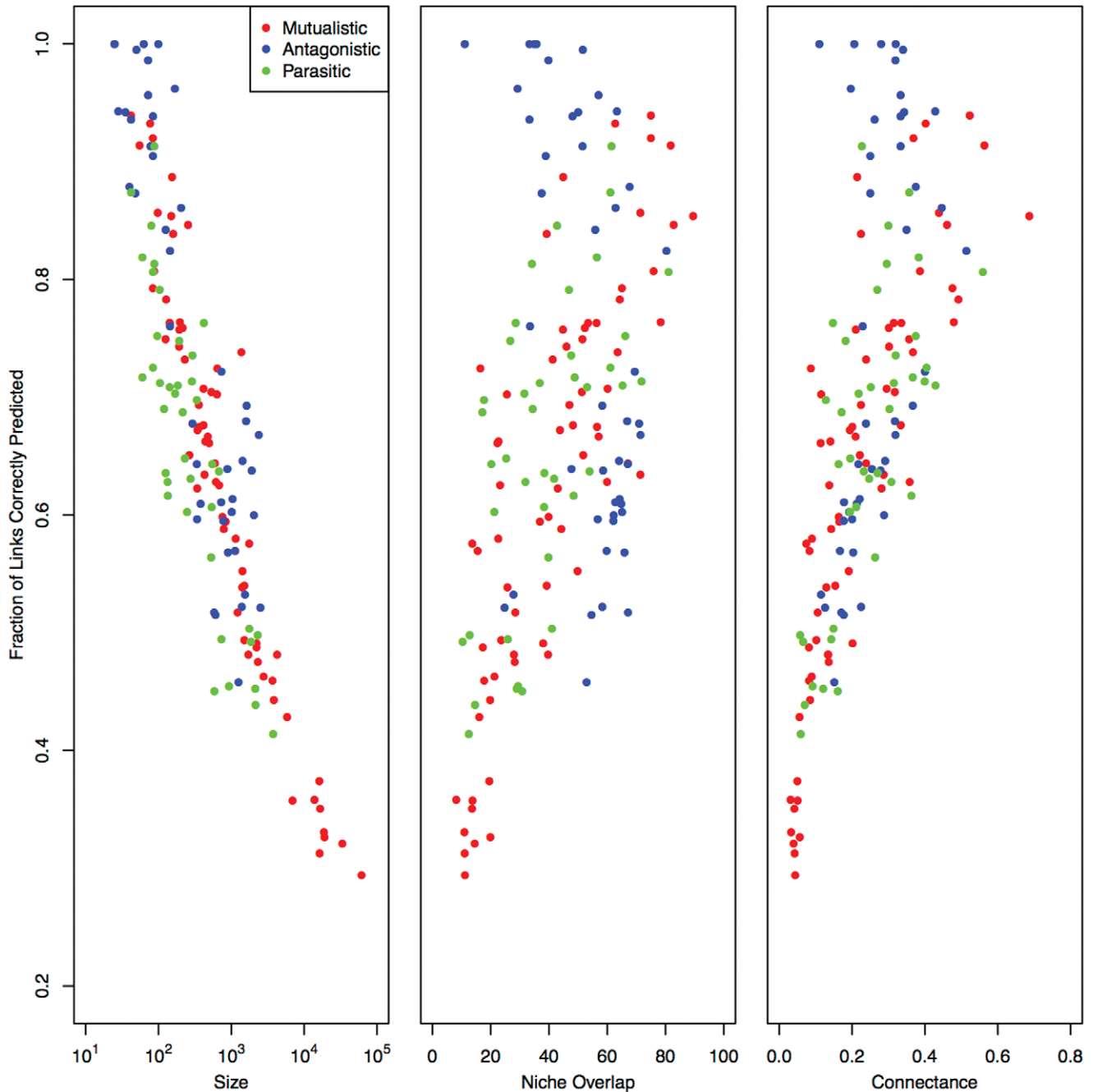
We note that the original NODF algorithm records  $NP_{Cij} = 0$  when two consumer’s feeding ranges perfectly overlap, yet it is debatable whether or not perfectly overlapping feeding ranges should result in a maximum deduction from nestedness, and thus for the purposes of this analysis we ignore those instances both in our estimating of empirical nestedness and calculation of overlap in model derived networks (although qualitative comparisons did not indicate a large discrepancy between the two).

NODF for consumers is then defined as the average of the  $NP_{Cij}$  values:  $NODF_C = 100 * \left( \frac{\sum NP_{Cij}}{P_C} \right)$ ; where 100 indicates perfect nestedness.

**Model Derived Niche Overlap**

Because the BPNM provides a probability of every link, it is not possible to directly calculate standard nestedness scores for model-derived networks. To address this, we employed two different approaches. In the first, we simulated 100 random networks, where for any network the realization of any interaction between a consumer and resource was the result of a random binomial draw against the BPNM derived probability. We then calculated the niche overlap for each simulated network using the NODF algorithm as above, and compared the empirical niche overlap with the mean niche overlap across all 100 simulated networks.

We also note that like most structural properties NODF can be significantly sensitive to changes in only a few interactions across



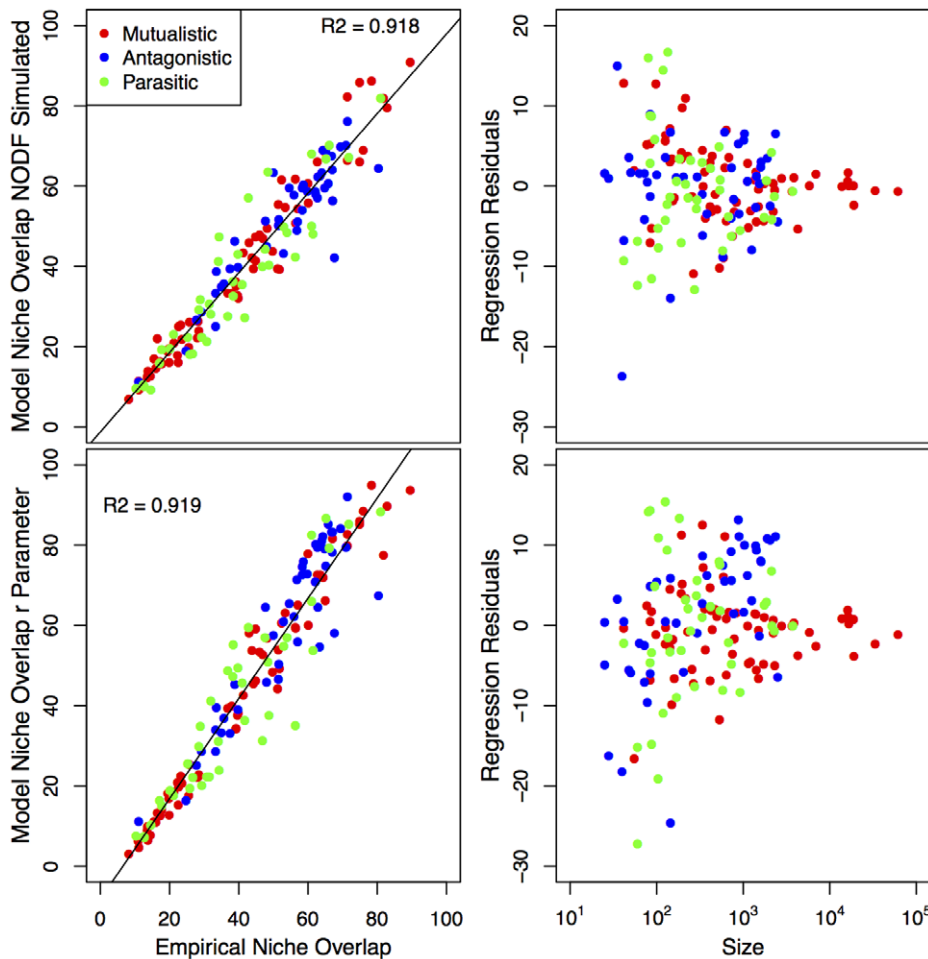
**Figure 4. Left) Fraction of links predicted versus network size (#Resource Species \* #Consumer Species); Centre ) Fraction of links ( $f_l$ ) correctly predicted (x-axis) versus empirical niche overlap (y-axis); Right) Fraction of links ( $f_l$ ) correctly predicted (x-axis) versus empirical connectance [#Links/(#Consumers \* #Resources)] (y-axis).** See Methods Summary for details on calculations of niche overlap. doi:10.1371/journal.pone.0056277.g004

an entire network [37]. To avoid the problems with this due to simulating networks as above, we calculated a metric on the BPNM derived networks complementary to NODF. In the BPNM, the one-dimensional niche employed represents the resource dimension available to consumers. For this reason the NODF method of niche overlap measurement is complementary to measuring the overlap between the  $r$  parameter of each consumer in a community. Thus, the consumer's dietary range ("r" parameter) can be used to make a comparison between the niche overlap of the consumer community in the empirical

network and the best-fit model derived community. To accomplish this we sort all  $S_C$  consumers by decreasing width of dietary range and for each consumer calculate the position of the  $r$  parameter cutoffs along the one dimensional niche axis:

$$r_{imin} = c_i - \left(\frac{r_i}{2}\right) \text{ and } r_{imax} = c_i + \left(\frac{r_i}{2}\right).$$

For any two consumers ( $i, j$ ), we calculate the percent of consumer  $i$ 's dietary range that overlaps with consumer  $j$ 's dietary



**Figure 5. Top Left) Empirical niche overlap (as measured by NODF – see Methods) versus model derived niche overlap (as measured by the average NODF score of 100 simulated networks – see Methods); Top Right) Residuals about the opposing linear regression versus network size; Bottom Left) Empirical niche overlap (as measured by NODF – see Methods) versus model derived niche overlap (as measured by the overlap in consumer  $r$  parameters – see Methods); Bottom Right) Residuals about the opposing linear regression versus network size.**

doi:10.1371/journal.pone.0056277.g005

range ( $PO_{Cij}$ , the continuous equivalent of the original NODF algorithm [40]), where  $NP_{Cij} = PO_{Cij}$  if  $r_j > r_i$ , and where  $NP_{Cij}$  is not recorded (“NA”) in the event of  $r_j = r_i$ .

The number of possible pairwise comparisons ( $P_C$ ) is the same as calculated above, and the overall niche overlap for consumers is then defined as the average of the  $NP_{Cij}$  values:

$Overlap_C = 100 * (\frac{\sum NP_{Cij}}{P_C})$ ; where 100 indicates perfect overlap. Thus,  $Overlap_C$  is a BPNM-specific method of measuring niche overlap on the same scale as  $NODF_C$ .

### Connectance

We measured the size of the networks as  $S_R * S_C$ , and the connectance of the empirical networks as  $\frac{\sum X}{S_R * S_C}$ . We measured the connectance of the model derived networks in two different ways. For the approach using 100 network simulations, we calculated model derived connectance as the average connectance of all 100 simulated networks. Avoiding simulating the networks resulted in the connectance metric  $\frac{\sum P_{ij}}{S_R * S_C}$ .

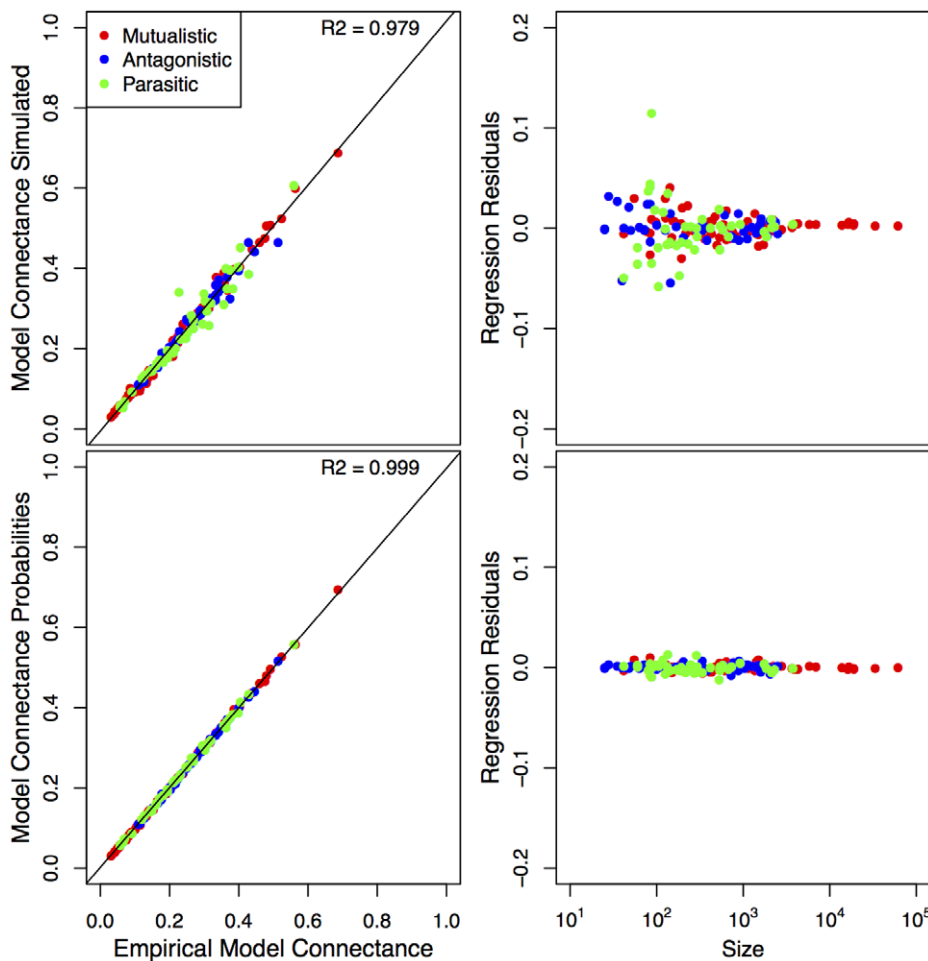
## Results

### Replicating Consumer-resource Interactions

In Figure 3 we show that the probabilistic niche model consistently outperforms a random (red) model in all but 18 of our 151 networks. The BPNM outperforms a bipartite cascade model in all but 15 of the 151 networks (black points in Figure 3). The few cases where the BPNM model does not outperform the cascade model are those networks in the lower end of the range of sizes we analyze. Given the uniformly superior performance of the BPNM, the rest of the paper analyzes only the performance of that model in detail.

In all networks, the expected number of links produced by the model, which is equal to the sum of the link probabilities, is very close to the number of interactions in the empirical dataset. As described and implemented previously by others [26,31], we can thus use the fraction of links ( $f_L$ ) correctly predicted ( $\frac{\sum P_{ij}}{X}$ ) as a simple and easily understood measure of model performance comparable to previous work. We note, however, that if the total model predicted links is significantly different from the total number of links in the network then this measure of performance is





**Figure 6. Top Left) Empirical connectance versus model derived connectance (as measured by the average connectance of 100 simulated networks – see Methods); Top Right) Residuals about the opposing linear regression versus network size; Bottom Left) Empirical connectance versus model derived connectance (sum of the model probabilities/network size – see Methods); Bottom Right) Residuals about the opposing linear regression versus network size.**

doi:10.1371/journal.pone.0056277.g006

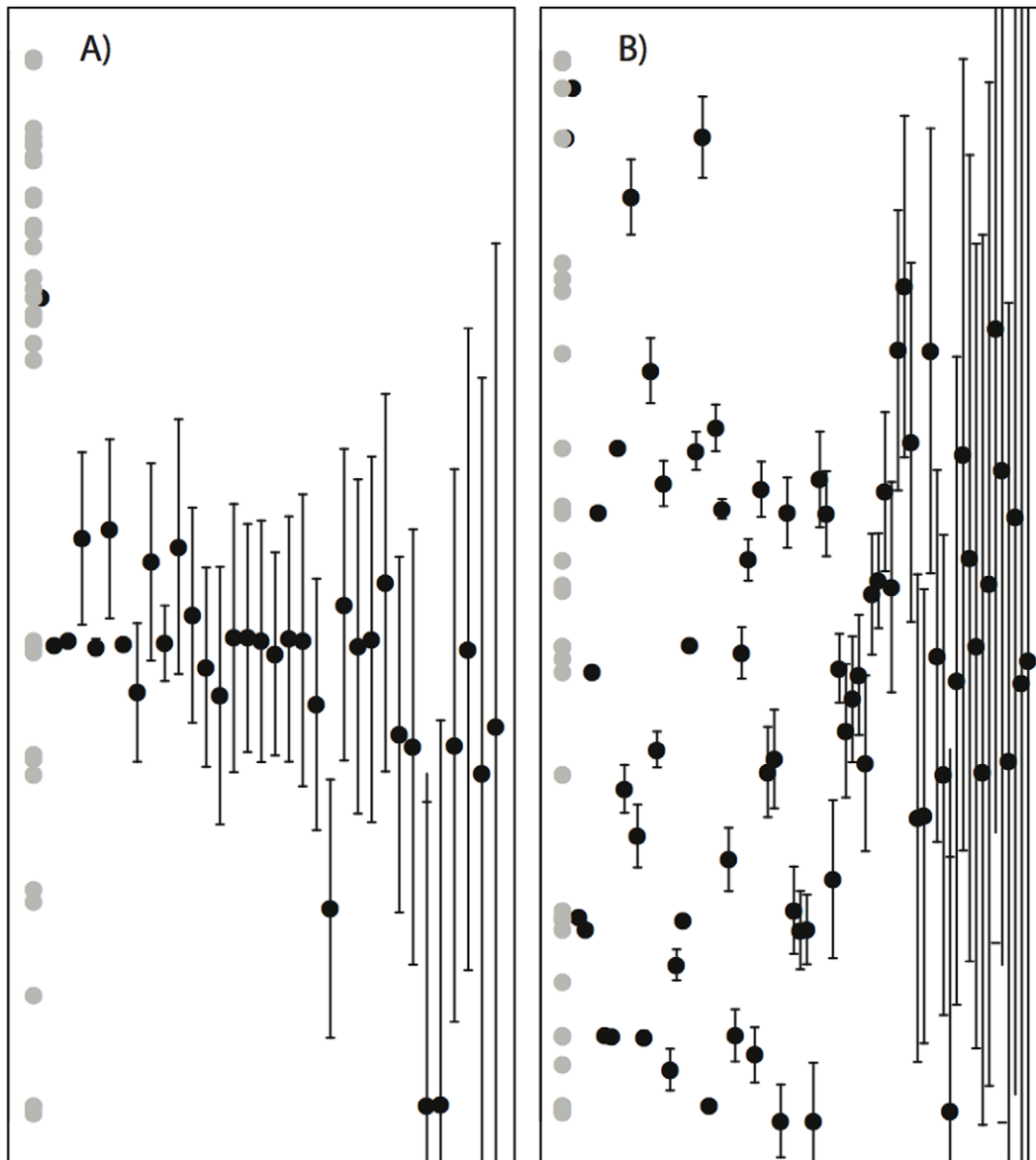
insufficient, as a model predicting every link as realized would by necessity produce a perfect fit – while also incorrectly predicting every missing link. Figure 4 (left) shows how our measure of model performance ( $f_L$ ) scales with network size. The decrease in  $f_L$  with size shows that a single niche dimension is sufficient for explaining individual interactions within small networks, but becomes insufficient for very large ones. This scale dependence in model fit exists for all three categories of interaction networks. Given the single dimension employed by the BPNM, the scale dependence in model fit is perhaps not surprising, as by random chance one expects larger networks to include greater heterogeneity in species trait distributions, and thus more dimensions for a niche space to be defined within.

### Replicating Consumer-resource Network Structure

A model that replicates the interactions between consumers and resources by necessity also replicates overall network structure, such as connectance, degree distribution, specialist-to-generalist consumer ordering, and niche overlap. One might assume that a model failing to replicate the interactions might also fail to replicate metrics of overall network structure, although analysis of empirical data has shown that temporal variability in individual interactions did not affect aggregated properties of network

architecture such as nestedness [41]. We show that the BPNM is capable of reproducing three ecologically important aspects of network structure (connectance, niche overlap, and specialist-to-generalist consumer ordering) for all categories of networks regardless of how well the model does at predicting individual interactions within the network.

Niche overlap is a structural quality of a bipartite network and as a general ecological concept simply addresses the question: to what extent are the resources taken by specialist consumers proper subsets of the resources taken by more generalist consumers [37–40]? We focus on niche overlap (as measured by the nestedness metric NODF; see *Methods* for implementation details) for two reasons. First, it is claimed that consumers in mutualistic networks are more highly overlapping than antagonistic consumers due to differing coevolutionary pressures [3] and thus we might expect our model results to differ between these two interaction types. More importantly, we investigate niche overlap because it is closely related to the niche concept underlying the BPNM. The usual picture of a nested network is highly niche structured in a single niche dimension. In nested networks, a narrower niche (e.g. specialized consumer feeding range) is nested within broader ones (e.g. generalist consumer). In this view of nested networks, narrower niches have a high level of overlap with broader ones.



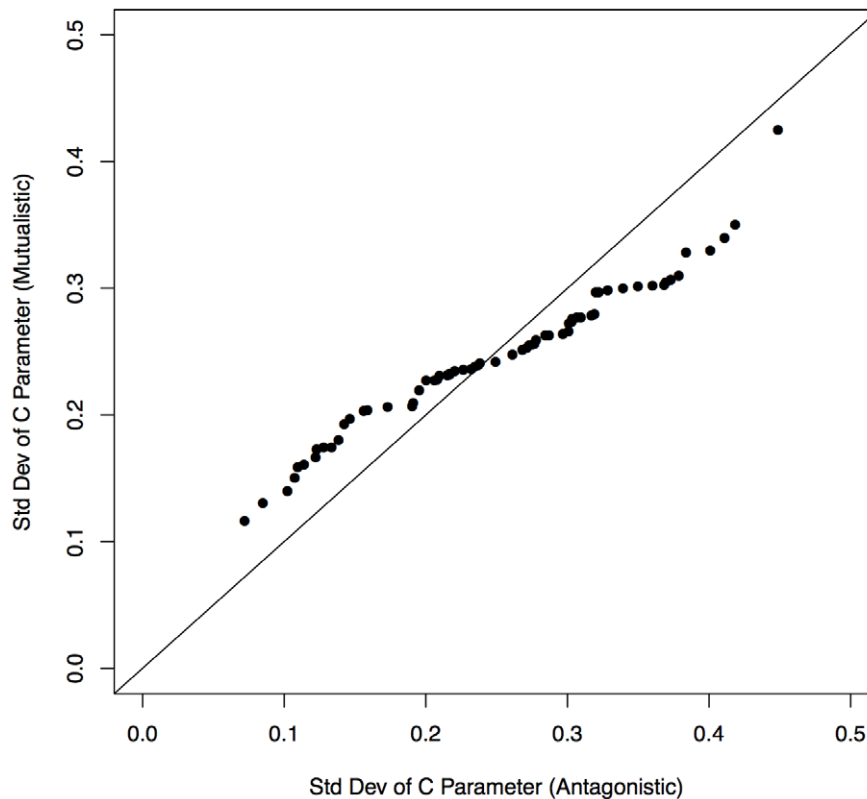
**Figure 7. Plots of best-fit model parameters for two empirical networks.** The y-axis represents the single dimensional axis assumed in the BPNM, scaling from 0 to 1. Light grey circles represent the niche positions ( $n$  parameters) of resource species along the single niche dimension. Black circles represent consumer niche center positions ( $c$  parameters), and the black bars represent the 97.5% range of consumer's dietary widths ( $r$  parameters). Consumers have been ordered equidistant from one another along the x-axis, in order (left to right) from most specialized (consume the fewest resource species) to most generalist (consume the greatest number of resource species) according their rank within the empirical interaction network. Left: An antagonistic network, where the best-fit model achieves 57% fraction of links ( $f_L$ ) correct. Right: A mutualistic network, where the best-fit model achieves 49% fraction of links ( $f_L$ ) correct.  
doi:10.1371/journal.pone.0056277.g007

In keeping with the theoretical niche-structured assumptions of the BPNM, in Figure 4 (middle and right) we show that the BPNM performs best on the most highly nested and connected (across all networks Spearman's correlation between connectance and nestedness = 0.80, p-value < 0.001) empirical networks. This shows that connectance, the niche overlap, and the niche concept as implemented by the niche model are related, such that all highly nested (or connected) ecological networks are well-explained by a one dimensional niche structured model.

We further explore the relationship between niche structure and niche overlap in Figure 5 (top and bottom left) by plotting the niche overlap of the empirical network against the niche overlap of

the model-derived network (see Methods Summary for our two alternative niche overlap calculations). Figure 5 shows the BPNM is capable of replicating the niche overlap structure even in networks that have a relatively low level of niche overlap and that, at the level of individual interactions, are poorly explained by the BPNM (have a low value of  $f_L$ ). This means that regardless of how well the model performs at predicting specific links; it consistently replicates the niche overlap structure of the empirical networks. The linear regressions in Figure 5, and the relationship between the residuals about those regressions and the size of the network shows that model niche overlap most departs from empirical overlap in small networks with high degrees of overlap (Both





**Figure 8. Q-Q plot of the standard deviation of consumer  $c$  parameter values in antagonistic (including parasitic) and mutualistic networks.**

doi:10.1371/journal.pone.0056277.g008

regressions:  $R^2 = \sim 0.92$ ,  $p < 2 \times 10^{-16}$ ), and the range of the residuals decreases with network size.

Figure 6 replicates Figure 5, but shows the relationship between empirical connectance and two measures of model derived connectance (see Methods for details). In Figure 6 we show the same results to be true for connectance as for niche overlap, although even stronger. This replicates the strong correlation between those two measures in the empirical dataset. It can be argued that connectance is a more parsimonious descriptor of network structure than calculating more complicated niche overlap metrics such as nestedness. Yet, connectance and niche overlap need not be strongly related (except when approaching the limit of all interactions realized). We do not attempt to discern whether either is driving the results in Figures 4 and 5, as it is sufficiently interesting to note that the BPNM replicates both measures.

Further, we show in Figure 7 that the BPNM, regardless of its ability to predict individual links, is capable of replicating another fundamental structural characteristic of ecological networks, which is the generalist to specialist consumer ordering. In Figure 7, the model-estimated niche axis position of the resources is plotted on the y-axis; while on the x-axis we have shown the model-estimated niche centroid ( $c_i$ ) and niche width ( $r_i$ ) of consumers in a mutualistic (Figure 7 right) and antagonistic (Figure 7 left) network. In both networks, however, we have ordered the consumers, from most specialized to most generalized, along the x-axis according to their empirical rank. It is immediately obvious that such an ordering closely matches the model ordering (measured as the consumer's  $r$  value). Indeed, the average spearman's rank-order correlation

(0.90, 92% of p-values  $< 0.05$ ) across all empirical and model-derived networks shows this to be true.

### Differences between Mutualistic and Antagonistic Interaction Networks

In Figure 7 (left) it is clear that the center of consumer's niches in the antagonistic network is much more strongly clumped together than in the mutualistic network (Figure 7 right), although the distribution of the resources across the single axis remains fairly uniform. We examined whether this pattern recurred across all the networks by examining the distributions of niche positions (" $c$ ") in all the networks and comparing the distributions between mutualistic and antagonistic (both predator-prey and parasite) networks. Distribution width was measured using the standard deviation of the  $c$  values, and the distribution of the standard deviation of  $c$  in mutualistic and antagonistic networks is significantly different (KS test,  $p = 0.029$ ). Q-Q plots show the relationship between two probability distributions, where a one-to-one relationship on the x and y axes indicates the two distributions are the same. Thus, a Q-Q plot of the standard deviation of  $c$  for mutualistic and antagonistic networks (Figure 8) shows that  $c$  standard deviation of mutualistic networks is more likely to be intermediate-valued, corresponding to a near-uniform distribution. In contrast, the  $c$  standard deviation of an antagonistic network is more likely to be relatively small or large, corresponding to a more highly peaked distribution (as in Figure 7 left) or a distribution with multiple spread peaks respectively.

## Discussion

A recent study [38] suggested that a bipartite cascade model consistently gave as good or better results than a niche-structured model. Our results do not support this, instead showing that an optimally parameterized niche model almost always outperforms an optimally parameterized cascade model. This difference in results probably occurs because the rules used to assign niche positions and interaction niche parameters (center and width) previously used [38] are highly non-optimal. Importantly, when maximum likelihood parameter choices are made, the BPNM performs consistently better than the bipartite cascade model.

Santamaría and Rodríguez-Gironés [34] showed previously a combination of complementarity and cascade rules to be necessary to reproduce observed network properties. We show here that a unidimensional complementarity model is sufficient. One possible explanation for this discrepancy might be that the BPNM is probabilistic, while that of Santamaría and Rodríguez-Gironés is not. However, the discrepancy is more likely due to the fact that Santamaría and Rodríguez-Gironés used fixed uniform distributions for their model, while our results show that the niche value distributions within empirical networks are highly non-uniform.

We show that a simple one-dimensional probabilistic niche model is capable of closely replicating the structure of three broad categories of ecological bipartite networks, regardless of fundamentally different structures across these networks, and that the niche-structured model is a consistently better model than models without constrained niches. For the first time we show that two commonly identified themes within the ecological network

literature, nestedness and niche, are closely related, and that using the niche concept is a simple and successful approach to modeling the structure of multiple types of bipartite ecological interaction networks.

Yet, we also find that there are clear differences in the niche structure of mutualistic and antagonistic networks. In particular, the distributions of niche centers are significantly different between mutualistic and antagonistic networks. It is likely that the consistent differences we show in the niche structure of mutualistic and antagonistic networks describe how mutualistic and antagonistic processes produce different constraints on species behaviors and interactions. This, in turn, directly influences our ability to predict and preserve earth's natural systems. Interactions between species, and the structure of ecological networks those interactions create, both have important consequences for the functioning and robustness of ecosystems [2,3]. The ability of a simple one-dimensional niche model to predict the structure of a wide range of ecological interaction networks demonstrates that common principles structure these systems. Discovering the order and consistent variability underlying what sometimes appear to be overwhelmingly complex systems is a vital step towards developing our ability to predict and preserve species, ecosystems, and their services.

## Author Contributions

Conceived and designed the experiments: LNJ RW. Performed the experiments: LNJ RW. Analyzed the data: LNJ RW. Contributed reagents/materials/analysis tools: LNJ RW. Wrote the paper: LNJ RW.

## References

- Pimm SL (1982) Food webs. Chapman and Hall. 240 p.
- Bascompte J, Jordano P (2007) Plant-animal mutualistic networks: the architecture of biodiversity. *Annu Rev Ecol Evol Syst* 38: 567–593.
- Thébault E, Fontaine C (2010) Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science* 329: 853–856.
- Lafferty KD, Allesina S, Arim M, Briggs CJ, Leo De L, et al. (2008) Parasites in food webs: the ultimate missing links. *Ecol Lett* 11: 533–546.
- Montoya JM, Pimm SL, Solé RV (2006) Ecological networks and their fragility. *Nature* 442: 260–264.
- Kearns CK, Inouye DW, Waser NM (1998) Endangered mutualisms: the conservation biology of plant-pollinator interactions. *Annu Rev Ecol Evol Syst* 29: 83–112.
- Belgrano A, Scharler U, Dunne JA, Ulanowicz RE (2005) Aquatic Food Webs: An Ecosystem Approach. London: Oxford University Press. 272 p.
- deRuiter PC, Walters V, Moore JC (2006) Dynamic food webs: multispecies assemblages, ecosystem development and environmental change. *Theoretical Ecology Series*. San Diego: Academic Press. 608 p.
- Pascual M, Dunne JA (2006) Ecological Networks: Linking Structure to Dynamics in Food Webs. New York: Oxford University Press. 404 p.
- May RM (2006) Network structure and the biology of populations. *Trends Ecol Evol* 21: 394–399.
- Dunne JA, Williams RJ, Martinez MD (2002) Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecol Lett* 5: 558–567.
- Srivasan UT, Dunne JA, Harte J, Martinez ND (2007) Response of complex food webs to realistic extinction sequences. *Ecology* 88: 671–682.
- Mitchell CE, Power AG (2003) Release of invasive plants from fungal and viral pathogens. *Nature* 421: 625–627.
- Torchin ME, Lafferty KD, Dobson AP, McKenzie VJ, Kuris AM (2003) Introduced species and their missing parasites. *Nature* 421: 628–630.
- Paine RT (1980) Food webs – linkage, interaction strength and community infrastructure. The 3rd Tansley lecture. *J Anim Ecol* 49: 667–685.
- Guimarães PR Jr, Ric-Gray V, Oliveira PS, Izzo TJ, dos Reis SF, et al. (2007) Interaction intimacy affects structure and coevolutionary dynamics in mutualistic networks. *Curr Biol* 17: 1–7.
- Santamaría L, Rodríguez-Gironés MA (2007) Linkage rules for plant-pollinator networks: trait complementarity or exploitation barriers. *PLoS Biol* 5: 354–362.
- Saavedra S, Reed-Tsochas F, Uzzi B (2009) A simple model of bipartite cooperation for ecological and organizational networks. *Nature* 457: 463–466.
- Campbell C, Yang S, Albert R, Shea K (2010) A network model for plant-pollinator community assembly. *Proc Nat Acad Sci USA* 108: 197–202.
- Williams RJ, Martinez N (2000) Simple rules yield complex food webs. *Nature* 404 6774: 180–183.
- Cohen JE, Briand F, Newman CM (1990) Community Food Webs: Data and Theory. Berlin: Springer. 308 p.
- Allesina S, Alonso D, Pascual M (2008) A general model for food web structure. *Science* 320: 658–661.
- Fontaine C, Guimarães PR Jr, Kéfi S, Loculle N, Memmott J, et al. (2011) The ecological and evolutionary implications of merging different types of network. *Ecol Lett* 14: 1170–1181.
- Martinez ND, Williams RJ, Dunne JA (2006) Diversity, complexity, and persistence in large model ecosystems. *Ecological Networks: Linking Structure to Dynamics in Food Webs* (eds M. Pascual & J.A. Dunne). New York: Oxford University Press. pp 163–182.
- Brose U, Berlow EL, Martinez ND (2005) Scaling up keystone effects from simple to complex ecological networks. *Ecol Lett* 8: 1317–1325.
- Williams RJ, Anandanadesan A, Purves DW (2010) The Probabilistic Niche Model Reveals the Niche Structure and Role of Body Size in a Complex Food Web. *PLoS One*: 5(8): e12092. doi:10.1371/journal.pone.0012092.
- Williams RJ, Purves DW (2011) The probabilistic niche model reveals substantial variation in the niche structure of empirical food webs. *Ecology* 92: 1849–1857.
- Joppa LN, Bascompte J, Montoya JM, Sole RV, Sanderson J, et al. (2009) Reciprocal specialization in ecological networks. *Ecol Lett* 12(9): 961–969.
- Williams RJ (2010) Simple MaxEnt models explain food web degree distributions. *Theor Ecol* 3: 45–52.
- Joppa LN, Montoya J, Sole R, Sanderson J, Pimm SL, (2010) On nestedness in ecological networks. *Evol Ecol Res* 12: 35–46.
- Williams RJ, Martinez N (2008) Success and its limits among structural models of complex food webs. *J Anim Ecol* 77: 512–519.
- Petchey OL, Beckerman AP, Riede JO, Warren PH (2008) Size, foraging, and food web structure. *Proc Nat Acad Sci USA* 105: 4191–4196.
- Grant V, Grant KL (1965) Flower Pollination in the Phlox Family. New York: Columbia University Press.
- Santamaría L, Rodríguez-Gironés MA (2007) Linkage rules for plant-pollinator networks: Trait complementarity or exploitation barriers? *PLoS Biol* 5(2): e31. doi:10.1371/journal.pbio.0050031.
- Jabot F, Bascompte J (2012) Bitrophic interactions shape biodiversity in space. *Proc Nat Acad Sci USA* 109: 4521–4526.
- Kirkpatrick S, Gelatt CD, Vecchi MP (1983) Optimization by simulated annealing. *Science* 220: 671–680.
- Joppa LN, Williams RJ (2011) The influence of single elements on nested community structure. *Methods Ecol Evol* 2: 541–549.
- Pires M, Prado P, Guimaraes P (2011) Do food web models reproduce the structure of mutualistic networks? *PLoS ONE* 6(11): e27280.

39. Ulrich W, Almeida-Neto M, Gotelli N (2009) A consumer's guide to nestedness analysis. *Oikos* 118: 3–17.
40. Almeida-Neto M, Guimaraes P, Guimaraes P Jr, Loyola R, Ulrich W (2008) A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* 117: 1227–1239.
41. Petanidou T, Kallimanis AS, Tzanopoulos J, Sgardelis SP, Pantis JD, et al. (2008) Long-term observation of a pollination network: fluctuation of species and interactions, relative invariance of network structure and implications for estimates of specialization. *Ecol Lett* 11: 564–575.