Abundance and generalization in mutualistic networks: solving the chicken-and-egg dilemma

Hugo Fort¹, Diego P. Vázquez²,³*, Boon Leong Lan⁴

¹Physics Department, Faculty of Sciences, Universidad de la República, Igua 4225, Montevideo 11400, Uruguay.

²Argentine Institute for Dryland Research, CONICET, CC 507, 5500 Mendoza, Argentina.

³Faculty of Exact and Natural Sciences, National University of Cuyo, M5502JMA Mendoza, Argentina.

⁴School of Engineering, Monash University, 46150 Bandar Sunway, Selangor, Malaysia.

*Corresponding author: Phone: +54-261-524-4121. E-mail: dvazquez@mendoza-conicet.gob.ar.

Statement of authorship: H.F. conceived the study. H.F. and D.V. jointly compiled the data, performed the analyses and wrote the paper. B.L.L. contributed to the analyses and to the final version of the manuscript.

Running title: Abundance and generalization in mutualistic networks

Manuscript type: Letter

20  **Key words:** causality, generalization, mutualistic networks, plant-animal interactions, pollination, seed dispersal, specialization
Abstract: A frequent observation in plant-animal mutualistic networks is that abundant species tend to be more generalized, interacting with a broader range of interaction partners than rare species. Uncovering the causal relationship between abundance and generalization has been hindered by a chicken-and-egg dilemma: is generalization a by-product of being abundant, or does high abundance result from generalization? Here we analyze a database of plant-pollinator and plant-seed disperser networks, and provide strong evidence that the causal link between the abundance and generalization is uni-directional. Specifically, species appear to be generalists because they are more abundant, but the converse, that is, that species become more abundant because they are generalists, is not supported by our analysis. Furthermore, null model analyses suggest that abundant species interact with many other species simply because they are more likely to encounter potential interaction partners.
Introduction

Understanding the causal relationships among entities in natural systems is one of the major philosophical and scientific challenges of all times (Pearl 2000; Shipley 2000). The challenge is particularly great in ecology, given that ecological systems are usually driven by multiple causality, and experimentation is not always possible (Pickett et al. 1994; Shipley 2000). The frequent observation in plant-animal mutualistic networks that abundant species tend to be more generalized, interacting with a broader range of interaction partners than rare species (Dupont et al. 2003; Vázquez & Aizen 2003), is a case in point: does high generalization lead to high abundance, or does high abundance lead to high generalization (Santamaría & Rodríguez-Gironés 2007; Fontaine 2013)?

Solving the above chicken-and-egg dilemma has been difficult, as there are good reasons to argue both ways. On the one hand, it is possible to argue that high generalization should lead to high local abundance (as well as broad geographic distributions) because generalists are able to exploit a broad range of resources, thus giving them an advantage over specialists (Brown 1984). For instance, an animal that can exploit flowers or fruits of many plant species should attain a higher local abundance than an animal specialized on few plant species. However, the likely trade-offs between generalization and the ability to exploit successfully any given resource---the jack-of-all-trades is a master of none (MacArthur 1972; Krasnov et al. 2004)---would blur the positive correlation between generalization and abundance. On the other hand, it is also possible to argue that high abundance should lead to high generalization since abundant species would interact with many other species simply because they are more likely to encounter potential interaction partners (Vázquez et al. 2007). In addition, abundance may lead to
generalization if generalized species are a collection of individuals specialized on distinct sets of resources, so that a greater number of individuals results in a larger set of resources exploited by the population (Araújo et al. 2011; Bolnick et al. 2011). We should also expect abundance to determine generalization if pollinators forage optimally, because at high pollinator densities resources should become scarcer, pushing pollinators towards greater generalization (Fontaine et al. 2008).

Here we offer a solution to the abundance-generalization causality dilemma in plant-animal mutualistic networks by evaluating the logical consequences of the above alternative hypotheses. We start by classifying plant and animal species in a network into two abundance categories (rare, \( R \), or abundant, \( A \)) and two generalization categories (specialist, \( S \), or generalist, \( G \)). The four resulting classes can be represented by a 2×2 abundance-generalization matrix:

\[
\begin{bmatrix}
F_{R,S} & F_{R,G} \\
F_{A,S} & F_{A,G}
\end{bmatrix}
\] (1)

Each entry of the above matrix represents the fraction of (animal or plant) species in the corresponding class; thus, \( F_{R,S} + F_{R,G} + F_{A,S} + F_{A,G} = 1 \). The abundance-generalization correlation implies that the diagonal entries \( F_{R,S} \) and \( F_{A,G} \) should be large, while the non diagonal entries should be small. In other words, the abundance-generalization correlation implies that low abundance (or rarity, \( R \)) and low generalization (or specialization, \( S \)) come together as well as high abundance (\( A \)) and high generalization (\( G \)). Therefore the diagonal matrix entries, \( F_{R,S} \) and \( F_{A,G} \) must be larger than the non-diagonal ones, \( F_{R,G} \) and \( F_{A,S} \). The question we want to answer can be formulated in terms of the following two alternative logic relationships between \( A \) and \( G \):
(i) If $A$ implies $G$, then no abundant species can be a specialist, i.e., $F_{AS} = 0$.

(ii) If $G$ implies $A$, then no generalist species can be rare, i.e., $F_{RG} = 0$.

Relationship (i) is equivalent to stating that $A$ is a sufficient condition for $G$ (and $G$ is a necessary condition for $A$), while relationship (ii) is equivalent to stating that $G$ is a sufficient condition for $A$ (and $A$ is a necessary condition for $G$).

We compiled a database on plant-pollinator and plant-seed disperser networks from local communities around the world, and then classified species in each network as either abundant or rare and either specialist or generalist to evaluate whether the frequencies $F_{AS}$ and $F_{RG}$ observed in plant-animal mutualistic networks matched the above predictions. We also conducted a null model analysis to assess whether the observed network patterns can be reproduced by assuming that abundant species are more likely to encounter potential interaction partners than rare species.

**Methods**

**Data**

The data consisted in 35 quantitative bipartite mutualistic networks (22 for plant-pollinator interactions and 13 for plant-seed disperser interactions), with broad geographic and taxonomic spans, with link weights represented as animal visitation frequency to plants, and the abundance of each animal or plant species in the network (Table 1).

**Calculation of abundance and generalization**

*Abundance estimates*
To compute the fractions of the $2 \times 2$ abundance-generalization matrix (eq. 1) for each network we need appropriate measures of abundance and generalization. Regarding abundance, for plants, ten of the plant-pollinator networks included estimates of plant abundance independent from the interaction observations, from transect and quadrat sampling (see Supplementary Information). For animals, no equivalent estimates of abundance were available for any of the datasets; we thus estimated the abundance of each animal species from the quantitative interaction networks by summing across the link weights (representing animal visitation frequency to plants). These animal abundance data are arguably limited, as they are not independent from the interactions; but these are the best data available to evaluate our question. Furthermore, the plant data do include independent estimates of abundance, and results for those datasets (see below) we similar to those for animals. In addition, we have used two different measures of generalization, degree and $g$, for both of which we got similar results. All of this suggests that our results are robust to the methodological limitations of the animal abundance data.

**Generalization estimates**

For generalization, the simplest measure is species degree (the number of species with which a given species interacts) (Vázquez & Aizen 2006). Since measuring generalization in this way ignores important information about interaction frequency and availability of interaction partners, and depends strongly on network size, we also used an alternative measure of generalization, based on the Kullback-Leibler distance $d$, which overcomes some of these limitations (Blüthgen et al. 2006). The Kullback–Leibler (K–L) distance or relative entropy is a
non-symmetric measure of the difference between two probability distributions $P$ and $Q$ (the K–L divergence from $P$ to $Q$ is generally not the same as that from $Q$ to $P$). For a given animal or plant, $P$ corresponds to the distribution of the interactions with each partner (respectively, plants or animals) and $Q$ corresponds to the overall partner availability (Blüthgen et al. 2006). We defined $g = 1 - d/d_{\text{max}}$ as a standardized measure of generalization, where $d_{\text{max}}$ is the natural logarithm of the sum of all link weights in the network (i.e., the grand sum of the bipartite interaction matrix) and is the maximum theoretically possible value for $d$ (corresponding to the case when two species interact exclusively with each other).

**Binary classification of abundance and generalization using the mean as threshold**

For any distribution there are two standard reference points that in principle could be used to provide binary classifications of variables: the median and the mean. The median, by its very definition, fails in the case of variables that follow very skewed distributions, such as abundance or degree. That is, for many of the datasets considered in this study more than half of the species have abundance and/or degree equal to its minimum possible value, i.e. 1. Therefore the corresponding median is also 1 and any species with abundance (degree) greater than 1 would be classified as $A (G)$. Such a classification is highly unsatisfactory. Take for instance the dataset number 1, code name 'bah' (Barrett & Helenurm 1987) in Table 1. The median abundance for pollinator species is 1 and the maximum abundance is 104 ($Eusphalerum$ sp.). It makes little sense to consider a pollinator species $A$ if it has abundance = 2, in the same class that a species with abundance = 104. Indeed a species with abundance = 2 seems to be closer to a species with abundance = 1 than to another species with abundance = 104. On the other hand, the mean does a
better job than the median (in the same example the mean abundance is equal to 5.4 and then A species are those with an abundance of at least 6). Hence, we classified species in each network as A if their abundance was equal or greater than mean abundance in the network, or R if their abundance was lower than the mean. Similarly, for generalization, we classified species as G if their degree or g was at or above the mean, and S if their degree or g was below the mean.

Fuzzy logic classification of abundance and generalization

Relying on a sharp threshold for defining categories may be problematic, as most so-called opposites—tall or short, hot or cold, etc.—are not separated by a sharp line; instead, they are ends of a continuum that involves subtle shadings. A useful alternative for classifying things or properties for which there is a degree of vagueness or context dependence that cannot be properly expressed with clear-cut classes is the fuzzy logic formalism (Zadeh 1965). We have conducted a fuzzy logic analysis, which leads to the same qualitative conclusions we reached using the simpler and more drastic classification with the mean as threshold. Fuzzy logic uses fuzzy sets or classes, which are generalizations of the conventional sets or 'crisp' sets. A conventional set or class \( C \) can be defined by a membership function \( P_C(x) \) that specifies whether an element \( x \) belongs to \( C \) or not. If an element \( x \) belongs to \( C \) then \( P_C(x) = 1 \) while if it does not belong to \( C \) then \( P_C(x) = 0 \). In contrast, a fuzzy set or class \( F \) is defined through a membership function \( P_F(x) \) that is not binary; rather it varies from 0 to 1. For example, a fuzzy set, representing a fuzzy concept such as 'tall', can be defined by assigning to each possible element within a certain domain (e.g., all the people in a country) a membership grade between 0 and 1.
that denotes the extent to which that element belongs to the fuzzy set (i.e., the extent to which that particular individual is tall).

Thus, we considered the following less drastic fuzzy classification: a species is in the lower class \( L \) (\( S \) or \( R \)), with membership grade \( = 1 \), if its property \( x \) (generalization or abundance, respectively) is below or equal to the mean minus one standard deviation of this quantity, \( \mu_x - \sigma_x \). In a completely equivalent way, a species is in the upper class \( U \) (\( G \) or \( A \)), with membership grade \( = 1 \), if its corresponding property is above the mean plus one standard deviation, \( \mu_x + \sigma_x \). Those species in-between \( \mu_x - \sigma_x \) and \( \mu_x + \sigma_x \) have a membership grade to class \( L \) (\( S \) or \( R \)) that is given by a linear membership function \( P_L(x) \) interpolating between 0, for \( x = \mu_x + \sigma_x \), and 1, for \( x = \mu_x - \sigma_x \). [In some of the plant-animal networks considered in our study \( \mu_x - \sigma_x \) was below the minimum possible value for the variable \( x \) (i.e., \( x_{\text{min}} = 1 \) for the degree and the abundance and \( x_{\text{min}} = 0 \) for the index \( g \)). In such cases we took for the threshold delimiting the class \( L \) the maximum between \( \mu_x - \sigma_x \) and \( x_{\text{min}} \).

To obtain the corresponding four fuzzy logic fractions of the 2\( \times \)2 abundance-generalization matrix it remains to specify how to compute the membership function of the complements of classes \( R \) and \( S \), i.e., \( A \) and \( G \) respectively, as well as the membership function for the intersections \( R \cap S \), \( R \cap G \), \( A \cap S \), and \( A \cap G \). As membership function of the complement \( F^C \) of a fuzzy set \( F \) (e.g., \( G \) in the case of \( S \), \( A \) in the case of \( R \)) the natural and most widely used function is the additive complement, \( P^F(x) = 1 - P^F(x) \). There are many functions that can be used to compute the intersection of fuzzy sets (Zimmermann 2010). We used as membership function for the intersection of two fuzzy sets \( F \) and \( F' \), \( F \cap F' \), the product, which is one of the
most widely used functions (Zimmermann 2010): \( P^{F\cap F'}(x) = P^F(x) \times P^{F'}(x) \). Therefore, the four fractions of the 2×2 abundance-generalization matrix become:

\[
F_{r,s} \equiv P^{R\cap S}(x) = P^R(x) \times P^S(x), \quad F_{r,g} \equiv P^{R\cap G}(x) = P^R(x) \times P^G(x),
\]

\[
F_{a,s} \equiv P^{A\cap S}(x) = P^A(x) \times P^S(x), \quad F_{a,g} \equiv P^{A\cap G}(x) = P^A(x) \times P^G(x).
\]

**Null model analysis**

We compared the frequency of occurrence of species in the 2×2 abundance-generalization matrix (eq. 1) with those predicted by a null model that assumes neutrality of interactions (Vázquez et al. 2009b), so that individuals interact randomly, regardless of their taxonomic identity. The null model generated 1000 randomized plant-animal interaction matrices for each dataset by assigning interactions according to an interaction probability matrix \( N \) constructed by multiplying the relative abundances of each pair of plant and animal species in the network, with the only constraint that each species had at least one interaction (see Vázquez et al. 2009b for details).

**Results**

We started by confirming the abundance-generalization correlation for pollinators, seed dispersers and plants in our database. Fig. 1 shows a highly positive correlation for most datasets.

We then evaluated the frequency of occurrence of species in the 2×2 abundance-generalization matrix (eq. 1) to evaluate the predictions of logic relationships (i) and (ii) (see *Introduction*). Virtually no species were both abundant and specialized (i.e., \( F_{a,s} \) close to zero),
when using both degree (Fig. 2, left column) and, especially, \( g \) (Fig. 2, right column) as measures of generalization, matching the expectation of logic relationship (i) (\( A \) implies \( G \)). Conversely, the frequency of rare and generalized species was high, substantially higher than zero (i.e., \( F_{R,G} \gg 0 \)) for both degree and \( g \) (Fig. 2), which does not match the expectation of logic relationship (ii) (\( G \) implies \( A \)).

The above results were based on a classification of species into abundance and generalization categories using the mean of these variables as the threshold for classification. Using the abundance and generalization classification based on fuzzy logic (see Methods: Calculation of abundance and generalization), results were qualitatively similar to those obtained using the mean as threshold, especially for \( g \) as our measure of generalization, which, as we argued above (see Methods: Calculation of abundance and generalization), is a better measure of generalization than the degree (Fig. S1).

Given the above results, the simplest interpretation of the pervasive abundance-generalization correlation in mutualistic networks is that abundant species are engaged in generalized interactions simply because they are more likely to encounter potential interaction partners (Vázquez et al. 2007). To evaluate this conjecture we compared the observed frequencies in the 2×2 abundance-generalization matrix (eq. 1) with the frequencies predicted by the null model that assumes random interactions among individuals. The predictions of the null model match closely the observed frequencies of occurrence in the 2×2 matrix for a majority of datasets (Fig. 3).

**Discussion**
Our analysis provides strong support for the hypothesis that abundance implies generalization, while generalization does not appear to imply high abundance. Thus, high abundance is a sufficient (but not a necessary) condition for generalization, while generalization is a necessary (but not sufficient) condition for a species to be abundant. Furthermore, our null model analysis indicates that the simplest interpretation of the pervasive correlation between abundance and generalization in mutualistic networks is that abundant species are engaged in generalized interactions simply because they are more likely to encounter potential interaction partners.

Based on these results, can we make a statement about the causal relationship between abundance and generalization? It is well known that, given two propositions \( p \) and \( q \), logical implications of the kind “\( p \) implies \( q \)” do not imply cause and effect; in other words, we can infer that “\( A \) implies \( G \),” but not that “\( A \) causes \( G \)”. Cause-and-effect assertions are predictive hypotheses that cannot be proved by statistical analysis, only disproved (Panik 2012). In that sense, our findings provide evidence against the proposition that generalization causes abundance, suggesting then that abundance causes generalization. In other words, if there is a causal relationship between abundance and generalization, abundance is what causes generalization, not the other way around.

When studying the structure of ecological interaction networks, it is important to bear in mind that the observed structure may partly result from sampling artifacts (Vázquez et al. 2009a). However, this is an unlikely explanation of our finding of \( F_{AS} \approx 0 \), as such artifacts come from the lack of information for links involving rare species (Blüthgen 2010) rather than abundant species. Missing interaction links for rare species would not affect \( F_{AS} \); it would, instead, lead to an underestimation of \( F_{RG} \). Therefore, increased sampling effort should lead to a
larger $F_{RG}$, thus reinforcing our conclusion. Furthermore, as we mentioned above (see Methods, Calculation of abundance and generalization), using the sum of interaction weights as a proxy of the abundances of pollinators and seed dispersers is arguably limited. However, since for the datasets available for our study there were no independent measurements of abundances for animals, these are the best estimates one can obtain. In any event, for plants, for which we did have estimates of abundances independent from visits, the results are similar to those for animals, confirming the general trend we found. Our findings also seem independent of the classification scheme of species into abundance and generalization categories (the binary and fuzzy logic classifications). Thus, it seems unlikely that our findings are just an artifact of the limitations of the abundance data for animals.

Our study sheds light on a long-standing causality dilemma between abundance and generalization in plant-animal mutualistic networks, with important implications for the ecological and evolutionary dynamics of these ecological systems. Furthermore, the reasoning used here, which is based on first principles of logical inference, could be applied to address similar causality problems in ecology. For example, for many ecological relationships, such as the relationship between species diversity and disturbance (Hughes 2010), it is unclear whether effects are uni or bi-directional, and to what extent feedbacks influence dynamics (Agrawal et al. 2007). Our approach could be used to offer solutions to such causality dilemmas.

Acknowledgments
HF was supported in part by PEDECIBA-Uruguay and ANII (SNI), DVP by FONCYT (PICT-2010-2779), and BLL by TWAS-Unesco as an Associate in the Physics Dept., Universidad de la República. We thank Nicolás Pérez for discussions on fuzzy logic algorithms, and Ignasi Bartomeus, Nico Blüthgen, James H. Brown, Jimena Dorado, Colin Fontaine, Rodrigo Ramos-Jiliberto, Rachael Winfree and two anonymous reviewers for helpful comments on the manuscript.

References


13.

14.

15.

16.

17.

18.

19.
20.  

21.  

22.  

23.  

24.  

25.  

26.  
27. 

28. 

29. 

30. 

31. 

32. 

33. 


Table 1. Datasets used in the study.

<table>
<thead>
<tr>
<th>Dataset no.</th>
<th>Dataset code</th>
<th>Interaction.type</th>
<th>Location</th>
<th>No. animal species</th>
<th>No. plant species</th>
<th>Ref.*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>bah</td>
<td>Plant-pollinator</td>
<td>Central New Brunswick, Canada</td>
<td>102</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>bez</td>
<td>Plant-pollinator</td>
<td>Pernambuco State, Brazil</td>
<td>13</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>dih</td>
<td>Plant-pollinator</td>
<td>Hickling, Norfolk, UK</td>
<td>61</td>
<td>17</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>dis</td>
<td>Plant-pollinator</td>
<td>Shelfanger, Norfolk, UK</td>
<td>36</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>ino</td>
<td>Plant-pollinator</td>
<td>Snowy Mountains, Australia</td>
<td>85</td>
<td>42</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>kat</td>
<td>Plant-pollinator</td>
<td>Ashu, Kyoto, Japan</td>
<td>679</td>
<td>91</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>mem</td>
<td>Plant-pollinator</td>
<td>Bristol, England</td>
<td>79</td>
<td>25</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>mos</td>
<td>Plant-pollinator</td>
<td>Melville Island, Canada</td>
<td>18</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>mot</td>
<td>Plant-pollinator</td>
<td>North Carolina, USA</td>
<td>44</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>10</td>
<td>ole</td>
<td>Plant-pollinator</td>
<td>Mauritian Ile aux Aigrettes</td>
<td>13</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>oll</td>
<td>Plant-pollinator</td>
<td>KwaZulu-Natal region, South Africa</td>
<td>56</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>12</td>
<td>sch</td>
<td>Plant-pollinator</td>
<td>Brownfield, Illinois, USA</td>
<td>32</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>13</td>
<td>sma</td>
<td>Plant-pollinator</td>
<td>Ottawa, Canada</td>
<td>34</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>14</td>
<td>vag</td>
<td>Plant-pollinator</td>
<td>Arroyo Goye, neighborhood of Nahuel Huapi National Park, Argentina</td>
<td>29</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>15</td>
<td>vcl</td>
<td>Plant-pollinator</td>
<td>Cerro Lópe, neighborhood of Nahuel Huapi National Park, Argentina</td>
<td>33</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>16</td>
<td>vll</td>
<td>Plant-pollinator</td>
<td>Llao-Llao Municipal Reserve, Bariloche, Argentina</td>
<td>29</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>17</td>
<td>vmn</td>
<td>Plant-pollinator</td>
<td>Lago Mascardi, Nahuel Huapi National Park, Argentina</td>
<td>26</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>18</td>
<td>vmh</td>
<td>Plant-pollinator</td>
<td>Lago Mascardi, Nahuel Huapi National Park, Argentina</td>
<td>35</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>19</td>
<td>vqh</td>
<td>Plant-pollinator</td>
<td>Península Quetrihué, Nahuel Huapi National Park, Argentina</td>
<td>27</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>20</td>
<td>vqn</td>
<td>Plant-pollinator</td>
<td>Península Quetrihué, Nahuel Huapi National Park, Argentina</td>
<td>24</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>21</td>
<td>vsa</td>
<td>Plant-pollinator</td>
<td>Península Quetrihué, Nahuel Huapi National Park, Argentina</td>
<td>27</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>22</td>
<td>vvi</td>
<td>Plant-pollinator</td>
<td>Villavicencio Nature Reserve, Mendoza,</td>
<td>97</td>
<td>41</td>
<td>14</td>
</tr>
<tr>
<td>23</td>
<td>bai</td>
<td>Plant-seed disperser</td>
<td>Princeton, Mercer, New Jersey, USA</td>
<td>21</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>24</td>
<td>bee</td>
<td>Plant-seed disperser</td>
<td>Mount Missim, Morobe Prov, New Guinea</td>
<td>9</td>
<td>31</td>
<td>16</td>
</tr>
<tr>
<td>25</td>
<td>ccg</td>
<td>Plant-seed disperser</td>
<td>Caguana, Puerto Rico</td>
<td>16</td>
<td>25</td>
<td>17</td>
</tr>
<tr>
<td>26</td>
<td>cci</td>
<td>Plant-seed disperser</td>
<td>Cialitos, Puerto Rico</td>
<td>20</td>
<td>34</td>
<td>17</td>
</tr>
<tr>
<td>27</td>
<td>cco</td>
<td>Plant-seed disperser</td>
<td>Cordillera, Puerto Rico</td>
<td>13</td>
<td>25</td>
<td>17</td>
</tr>
<tr>
<td>28</td>
<td>cfr</td>
<td>Plant-seed disperser</td>
<td>Fronton, Puerto Rico</td>
<td>15</td>
<td>21</td>
<td>17</td>
</tr>
<tr>
<td>29</td>
<td>fro</td>
<td>Plant-seed disperser</td>
<td>Mtunzini, South Africa</td>
<td>10</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td>30</td>
<td>ge1</td>
<td>Plant-seed disperser</td>
<td>Santa Genebra Reserve T1 SE, Brazil</td>
<td>18</td>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td>31</td>
<td>ge2</td>
<td>Plant-seed disperser</td>
<td>Santa Genebra Reserve T2 SE, Brazil</td>
<td>29</td>
<td>35</td>
<td>19</td>
</tr>
<tr>
<td>32</td>
<td>hra</td>
<td>Plant-seed disperser</td>
<td>Hato Ratón, Sevilla Spain</td>
<td>17</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>33</td>
<td>nco</td>
<td>Plant-seed disperser</td>
<td>Nava Correhuelas S Cazorla, SE Spain</td>
<td>33</td>
<td>25</td>
<td>21</td>
</tr>
<tr>
<td>34</td>
<td>sap</td>
<td>Plant-seed disperser</td>
<td>Yakushima Island, Japan</td>
<td>8</td>
<td>15</td>
<td>22</td>
</tr>
<tr>
<td>35</td>
<td>sno</td>
<td>Plant-seed disperser</td>
<td>Tropical rainforest, Trinidad</td>
<td>14</td>
<td>50</td>
<td>23</td>
</tr>
</tbody>
</table>

* Dataset references: 1, Barrett & Helenurm (1987); 2, Bezerra et al. (2009); 3, Dicks et al. (2002); 4, Inouye & Pyke (1988); 5, Kato et al. (1990); 6, Memmott (1999); 7, Mosquin & Martin (1967); 8, Motten (1982, 1986); 9, Olesen et al. (2002); 10, Ollerton et al. (2003); 11, Schemske et al. (1978); 12, Small (1976); 13, Vázquez & Simberloff (2003); 14, Vázquez et al. (2009b); 15, Baird (1980); 16, Beehler (1983); 17, Carlo et al. (2003); 18, Frost et al. (1980); 19, Galetti & Pizo (1996); 20, Jordano (1985); 21, Olesen et al. (2010); 22, Noma (1997); 23, Snow & Snow (1971).
Figure legends

**Figure 1.** Correlation between abundance and generalization. Plots show the distribution of Spearman rank correlation coefficients between abundance and one of two measures of generalization, degree or $g$ (where $g = 1 - \frac{d}{d_{\text{max}}}$ and $d$ is Kullback-Leibler distance (Blüthgen et al. 2006; see text) for pollinators (left), seed dispersers (center) and plants (right) in mutualistic networks. In all panels, each box-and-whisker plot, the horizontal line parting each box indicates the median, box limits are first and third distributional quartiles, whiskers extend to most extreme data point within 1.5 times the interquartile range, and circles indicate outlying data points.

**Figure 2.** The distribution of species in abundance-generalization classes. Plots show the fraction of species in each of four abundance-generalization classes for animal species in plant-seed disperser networks (top), and for animal (middle) and plant (bottom) species in plant-pollinator networks, using degree (left) or $g$ (right) as measures of generalization, using mean abundance and generalization as thresholds for defining categories (see Fig. S1 for an alternative using fuzzy logic). In all panels, $R$ are rare species, $A$ abundant species, $S$ specialized species, and $G$ generalized species. In each box-and-whisker plot, the thick horizontal line parting each box indicates the median, notches above and below indicate confidence limits of the median (precisely $1.58 \frac{\text{IQR}}{\sqrt{n}}$, where IQR is the interquartile range and $n$ is the sample size), whiskers extend to most extreme data point within 1.5 times the interquartile range, and circles indicate outlying data points.
indicate outlying data points. Medians of boxes in which notches do not overlap are considered to be significantly different (Crawley 2007).

**Figure 3.** Results of null model analyses. Each panel shows the results of the null model model analysis of the fraction of species in each category in the 2×2 abundance-generalization matrix (eq. 1) for each network in our dataset (indicated by dataset codes in the abcissa of each panel), each of the two generalization measures (degree, top three panels, and $g$, bottom three panels), and each group of studied species (plants and animals in plant-pollinator networks and animals in plant-seed disperser networks). For each network, observed fractions are represented by empty circles, and 95% confidence intervals of null model fractions are represented by error bars. Thus, an overlap between a circle and an error bar means no significant differences between observed and predicted fractions. For each category in the 2×2 matrix the ordinates are scaled between 0 and 1.
[Figure 2]
[Figure 3, continued]

**Plant–pollinator: Plants, g**

<table>
<thead>
<tr>
<th>Fraction in category</th>
<th>AG</th>
<th>AS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>vag</th>
<th>vcl</th>
<th>vll</th>
<th>vmh</th>
<th>vmn</th>
<th>vqh</th>
<th>vqn</th>
<th>vsa</th>
<th>vvi</th>
</tr>
</thead>
</table>

**Plant–seed disperser: Animals, g**

<table>
<thead>
<tr>
<th>Fraction in category</th>
<th>AG</th>
<th>AS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Dataset | bai | bee | ccf | cci | cco | cfr | fro | ge1 | ge2 | hra | nco | sap | sno |

**Plant–pollinator: Animals, g**

<table>
<thead>
<tr>
<th>Fraction in category</th>
<th>AG</th>
<th>AS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Dataset | bar | bez | dih | dis | ino | kat | mem | mos | mot | ole | oll | sch | sma | vag | vcl | vll | vmh | vmn | vqh | vqn | vsa | vvi |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
Abundance and generalization in mutualistic networks:
solving the chicken-and-egg dilemma

Hugo Fort¹, Diego P. Vázquez²,³*, Boon Leong Lan⁴

Supplementary figures: Figure legends

Figure S1. The distribution of species in abundance-generalization classes, using fuzzy logic to
define criteria for defining abundance and generalization categories (see Methods: Calculation of
abundance and generalization). Other conventions as in Fig. 2 (main text).