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

Hugo Fort, Diego P. Vázquez1,3* and Boon Leong Lan4

Abstract
A frequent observation in plant–animal mutualistic networks is that abundant species tend to be more generalised, interacting with a broader range of interaction partners than rare species. Uncovering the causal relationship between abundance and generalisation has been hindered by a chicken-and-egg dilemma: is generalisation a by-product of being abundant, or does high abundance result from generalisation? Here, we analyse a database of plant–pollinator and plant–seed disperser networks, and provide strong evidence that the causal link between abundance and generalisation is unidirectional. 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.

Keywords
causality, generalisation, mutualistic networks, plant–animal interactions, pollination, seed dispersal, specialisation.

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 generalised, 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 generalisation lead to high abundance, or does high abundance lead to high generalisation (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 one hand, it is possible to argue that high generalisation 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 specialised on few plant species. However, the likely trade-offs between generalisation 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 generalisation and abundance. On the other hand, it is also possible to argue that high abundance should lead to high generalisation 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 generalisation if generalised species are a collection of individuals specialised 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 generalisation if pollinators forage optimally, because at high pollinator densities resources should become scarcer, pushing pollinators towards greater generalisation (Fontaine et al. 2008).

Here, we offer a solution to the abundance–generalisation 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 generalisation categories (specialist, S, or generalist, G). The four resulting classes can be represented by a $2 \times 2$ abundance–generalisation matrix:

$$
\begin{bmatrix}
F_{RS} & F_{RG} \\
F_{AS} & F_{AG}
\end{bmatrix}
$$

(1)

Each entry of the above matrix represents the fraction of (animal or plant) species in the corresponding class; thus,

---

1Physics Department, Faculty of Sciences, Universidad de la República, Ñuñoa 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, MSS502 UMA Mendoza, Argentina
4Electrical and Computer Systems Engineering, School of Engineering, Monash University, 47500 Bandar Sunway, Malaysia
*Correspondence: E-mail: dvazquez@mendoza-conicet.gob.ar

© 2015 John Wiley & Sons Ltd/CNRS
The abundance–generalisation correlation implies that the diagonal entries $F_{R,S}$ and $F_{A,G}$ should be large, whereas the non-diagonal entries should be small. In other words, the abundance–generalisation correlation implies that low abundance (or rarity, $R$) and low generalisation (or specialisation, $S$) come together as well as high abundance ($A$) and high generalisation ($G$). 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. $F_{A,S} = 0$.

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

Relationship (i) is equivalent to stating that $A$ is a sufficient condition for $G$ (and $G$ is a necessary condition for $A$), whereas 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 $F_{A,S}$ and $F_{R,G}$ 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 generalisation**

**Abundance estimates**

To compute the fractions of the $2 \times 2$ abundance–generalisation matrix (eqn 1) for each network we need appropriate measures of abundance and generalisation. 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. For animals, no equivalent estimates of abundance were available for any of the data sets; 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) were similar to those for animals. In addition, we have used two different measures of generalisation, degree and $g$ (see below), 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.

**Generalisation estimates**

For generalisation, the simplest measure is species degree (the number of species with which a given species interacts) (Vázquez & Aizen 2006). Since measuring generalisation 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 generalisation, 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 standardised measure of generalisation, 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 generalisation 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 the case for many of the datasets considered in this study, in which 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) $> 1$ 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 $104$ (Eusphealerus sp.). It makes little sense to consider a pollinator species $A$ if it has abundance $= 2$, in the same class as 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 generalisation, 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.

© 2015 John Wiley & Sons Ltd/CNRS
A conventional set or class 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 generalisations 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 \( P^C(x) = 1 \), whereas 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).

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 are degrees of vagueness or context dependence that cannot be properly expressed with clear-cut classes is the fuzzy logic classification of abundance and generalisation.
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. $x_{\min} = 1$ for the degree and the abundance and $x_{\min} = 0$ for the index $g$). In such cases we took for the threshold delimiting the class $L$ the maximum between $\mu_x - \sigma_x$ and $x_{\min}$.)

To obtain the corresponding four fuzzy logic fractions of the $2 \times 2$ abundance–generalisation 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, $p^{F^c}(x) = 1 - p^F(x)$. 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 $F'$, $HF'$, the product, which is one of the most widely used functions (Zimmermann 2010): $p^{HF'}(x) = p^F(x) \times p^{F'}(x)$. Therefore, the four fractions of the $2 \times 2$ abundance–generalisation matrix become:

$$F_{RS} \equiv p^{R \cap S}(x) = p^R(x) \times p^S(x),$$
$$F_{RG} \equiv p^{R \cap G}(x) = p^R(x) \times p^G(x),$$
$$F_{AS} \equiv p^{A \cap S}(x) = p^A(x) \times p^S(x),$$
$$F_{AG} \equiv p^{A \cap G}(x) = p^A(x) \times p^G(x).$$

Null model analysis

We compared the frequency of occurrence of species in the $2 \times 2$ abundance–generalisation matrix (eqn 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 randomised plant–animal interaction matrices for each data set 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–generalisation correlation for pollinators, seed dispersers and plants in our database. Figure 1 shows a highly positive correlation for most datasets.

We then evaluated the frequency of occurrence of species in the $2 \times 2$ abundance–generalisation matrix (eqn 1) to evaluate the predictions of logic relationships (i) and (ii) (see Introduction). Virtually no species were both abundant and specialised (i.e. $F_{AS}$ close to zero), when using both degree (Fig. 2, left column) and, especially, $g$ (Fig. 2, right column) as measures of generalisation, matching the expectation of logic relationship (i) ($A$ implies $G$). Conversely, the frequency of rare and generalised species was high, substantially higher than zero (i.e. $F_{RG}$ or $G \gg 0$) 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 generalisation categories using the mean of these variables as the threshold for classification. Using the abundance and generalisation classification based on fuzzy logic (see Methods: Calculation of abundance and generalisation), results were qualitatively similar to those obtained using the mean as threshold, especially for $g$ as our measure of generalisation, which, as we argued above (see Methods: Calculation of abundance and generalisation), is a better measure of generalisation than the degree (Fig. S1).

Given the above results, the simplest interpretation of the pervasive abundance–generalisation correlation in mutualistic networks...
networks is that abundant species are engaged in generalised 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 $2 \times 2$ abundance–generalisation matrix (eqn 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 $2 \times 2$ matrix for a majority of data sets (Fig. 3).

**DISCUSSION**

Our analysis provides strong support for the hypothesis that abundance implies generalisation, while generalisation does not appear to imply high abundance. Thus, high abundance is a sufficient (but not a necessary) condition for generalisation, while generalisation 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 generalisation in mutualistic networks is that abundant species are engaged in generalised 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 generalisation? 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 generalisation causes abundance, suggesting then that abundance causes generalisation. In other words, if there is a causal relationship between abundance and generalisation, abundance is what causes generalisation, 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 artefacts (Vázquez et al. 2009a). However, this is an unlikely explanation of our finding of $F_{A,S} \approx 0$, as such artefacts come from the lack of information for links involving rare species (Blüthgen 2010) rather than abundant species. Missing interaction links for rare spe-
Figure 3 Results of null model analyses. Each panel shows the results of the null model analysis of the fraction of species in each category in the $2 \times 2$ abundance–generalisation matrix (eqn 1) for each network in our dataset (indicated by dataset codes in the abscissa of each panel), each of the two generalisation 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 $2 \times 2$ matrix the ordinates are scaled between 0 and 1.
cies would not affect $F_{A,S}$; it would, instead, lead to an under-
estimation of $F_{R,G}$. Therefore, increased sampling effort should
lead to a larger $F_{R,G}$, thus reinforcing our conclusion. Furthermore,
as we mentioned above (see Methods, Calculation of abundance
and generalisation), using the sum of interaction weights as a proxy of
the abundances of pollinators and seed dispersers is arguably limited. However, since for the data sets
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 classifica-
tion scheme of species into abundance and generalisation cate-
gories (the binary and fuzzy logic classifications). Thus, it
seems unlikely that our findings are just an artefact of the
limitations of the abundance data for animals.

Our study sheds light on a long-standing causality dilemma
between abundance and generalisation in plant–animal mutu-
alistic networks, with important implications for the ecological
and evolutionary dynamics of these ecological systems. Fur-
thermore, the reasoning used here, which is based on first prin-
ciples of logical inference, could be applied to address similar
causality problems in ecology. For example for many ecologi-
cal 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.

ACKNOWLEDGEMENTS

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., Universi-
dad de la República. We thank Nicolás Pérez for discussions
on fuzzy logic algorithms, and Ignasi Bartomeus, Nico Blüth-
gen, James H. Brown, Jimena Dorado, Colin Fontaine,
Rodrigo Ramos-Jiliberto, Rachael Winfree and two anony-

B.L.L. contributed to the analyses and to the final version of

REFERENCES

Agrawal, A.A., Ackerly, D.D., Adler, F., Arnold, A.E., Cáceres, C.,
Doak, D.F. et al. (2007). Filling key gaps in population and
*Baird, J.W. (1980). The selection and use of fruit by birds in an eastern
boreal forest herbs. I. Breeding systems and pollination. Can. J. Bot.,
65, 2036–2046.
*Bechler, B. (1983). Frugivory and polygamy in birds of paradise. Auk,
100, 1–12.
networks of oil-flowers: a tiny world within the smallest of all worlds.
Blüthgen, N. (2010). Why network analysis is often disconnected from
community ecology: a critique and an ecologist’s guide. Basic Appl.
Ecol., 11, 185–195.
Bolnick, D.I., Amarasekare, P., Araújo, M.S., Bürger, R., Levine, J.M.,
Novak, M. et al. (2011). Why intraspecific trait variation matters in
preferences across a Puerto Rican forested landscape: pattern
plant–flower-visitor network in the high-altitude sub-alpine desert of
coastal dune forest. In: Acta XVII Congressus Internationals
Ornithologici (ed. Nohring, R.). Presented at the Deutches
Hughes, A. (2010). Disturbance and diversity: an ecological chicken and
Mountains of Australia: comparisons with montane Colorado, USA.
matorral mediterráneo del sur de España: importancia de su invernada
relationship in the primary beech forest of Ashu, Kyoto: an overview of
the flowering phenology and seasonal pattern of insect visits. Contr
Biol Lab Kyoto Univ., 27, 309–375.
*Krasnov, B.R., Poulin, R., Shenbrot, G.I., Moullidiot, D. & Khokhlova,
abundance and host specificity in fleas (Siphonaptera) parasitic on
Press, Princeton.
Lett., 2, 276–280.
81, 201–205.
Community in the Deciduous Forests of Piedmont North Carolina.
Duke University, Durham, USA.
community of a temperate deciduous forest. Ecol. Monogr., 56, 21–42.
*Noma, N. (1997). Annual fluctuations of sap fruits production and
synchronization within and inter species in a warm temperate forest on
Yakushima Island, Japan. Tropics, 6, 441–449.
*Olesen, J.M., Eskildsen, L.I. & Venkatasamy, S. (2002). Invasion of
pollination networks on oceanic islands: importance of invader


**SUPPORTING INFORMATION**

Additional Supporting Information may be downloaded via the online version of this article at Wiley Online Library (www.ecologyletters.com).

Editor, Rebecca Irwin

Manuscript received 20 July 2015
First decision made 20 August 2015
Manuscript accepted 17 September 2015