ECOLOGY LETTERS

Ecology Letters, (2013) doi: 10.1111/ele.12081

LETTER

The dimensionality of ecological networks

Anna Eklöf,[1†*] Ute Jacob,[2] Jason Kopp,[1] Jordi Bosch,[3] Rocío Castro-Urgal,[4] Natacha P. Chacoff,[5] Bo Dalsgaard,[6] Claudio de Sassi,[7] Mauro Galetti,[8] Paulo R. Guimarães,[9] Silvia Beatriz Lomáscolo,[5,10] Ana M. Martín González,[6,11] Marco Aurelio Pizo,[12] Romina Rader,[13] Anselm Rodrigo,[3] Jason M. Tylianakis,[7] Diego P. Vázquez,[5,10] and Stefano Allesina,[1,14]

Abstract

How many dimensions (trait-axes) are required to predict whether two species interact? This unanswered question originated with the idea of ecological niches, and yet bears relevance today for understanding what determines network structure. Here, we analyse a set of 200 ecological networks, including food webs, antagonistic and mutualistic networks, and find that the number of dimensions needed to completely explain all interactions is small (< 10), with model selection favouring less than five. Using 18 high-quality webs including several species traits, we identify which traits contribute the most to explaining network structure. We show that accounting for a few traits dramatically improves our understanding of the structure of ecological networks. Matching traits for resources and consumers, for example, fruit size and bill gape, are the most successful combinations. These results link ecologically important species attributes to large-scale community structure.

Keywords

Ecological networks, food web structure, intervality, niche space, species traits, scaling.

Ecology Letters (2013)

INTRODUCTION

Will two individuals of different species interact if given the opportunity? If the two species have matching traits, then an interaction is possible, for example, a moth's proboscis is long enough for a flower corolla, a predatory fish's jaw can accommodate its prey or a bird's beak can crack the seed. Otherwise, the interaction is averted – a so called 'forbidden link' (Jordano *et al.* 2003). Evolution continuously alters species traits to favour or prevent interactions (Thompson 2005). Rewards and deception tend to favour interactions, while chemical defences and shells tend to prevent them (Gilman *et al.* 2012). Clearly, several traits might need to match for an interaction to occur. Thus, we can imagine species as embedded in a multi-dimensional space, where each axis represents a given trait (e.g. size, colour, mobility, phenology) and each species interacts with all others falling within a certain volume of this space. This is the familiar idea of multidimensional ecological niche pioneered by Grinell, Elton and Hutchinson (Chase & Leibold 2003). Consequently, if we were to know all the relevant traits for a set of species, connecting each species with those embedded in its corresponding volume would result in an ecological interaction network. We refer to the minimum number of trait-axes – dimensions – needed to fully reproduce such a network as its 'dimensionality', $\mathcal{D}$.

Knowing the maximum number of dimensions needed to fully describe complex ecological networks is important, as current ecological theory implicitly relies on the assumption that few dimensions are needed: species compete for few limiting factors in models explaining coexistence at a single trophic level (Silvertown 2004); macro-ecological approaches explain several ecological patterns using species' body mass and metabolic rate (Brown *et al.* 2004); and phenology drives recent models for the effect of climate change (Forrest *et al.* 2010; Diamond *et al.* 2011). Knowing the dimensionality of ecological networks would greatly contribute to our fundamental understanding of what determines species interactions and thereby the structure of ecological networks. Following this, we may also be able to predict ecological networks based on knowledge of the number of dimensions and identity of the most important traits determining structure. Moreover, knowing the dimensionality of ecological networks is vital for constructing realistic and reliable models of ecological systems.

The search for the dimensionality of ecological networks has earlier taken one of two directions: investigation of the number of niche-axes (hypothetical traits) needed to explain a particular network (Cohen 1968; Roberts 1978; Stouffer *et al.* 2006; Allesina *et al.* 2008), and the identification of the species traits that best explain observed interactions (Stouffer *et al.* 2011; Zook *et al.* 2011). We refer to 'dimensions' in the former case and reserve the word

[1]*Department of Ecology & Evolution, University of Chicago, Chicago, IL,USA*
[2]*Institute for Hydrobiology and Fisheries Science, Hamburg, Germany*
[3]*CREAF – Ecology Unit, Universitat Autónoma de Barcelona, Barcelona, Spain*
[4]*Institut Mediterrani d'Estudis Avanc¸ats (CSIC-UIB), Mallorca, Balearic Islands, Spain*
[5]*Instituto Argentino de Investigaciones de las Zonas Áridas, CONICET, Mendoza, Argentina*
[6]*Center for Macroecology, Evolution and Climate, Department of Biology, University of Copenhagen, Copenhagen, Denmark*
[7]*School of Biological Sciences, University of Canterbury, Canterbury, New Zealand*
[8]*Departamento de Ecologia, Universidade Estadual Paulista, Rio Claro, Brazil*
[9]*Departamento de Ecologia, I.B, Universidade de São Paulo, Sao Paulo, Brazil*
[10]*Instituto de Ciencias Básicas, Universidad Nacional de Cuyo, Mendoza, Argentina*
[11]*Pacific Ecoinformatics and Computational Ecology Lab, Berkeley, CA, USA*
[12]*Departamento de Zoologia, Universidade Estadual Paulista, São Paulo, Brazil*
[13]*Department of Physical Geography and Quaternary Geology, Stockholm University, Stockholm, Sweden*
[14]*Computation Institute, University of Chicago, Chicago, USA*
†*Present address: Department of Physics, Chemistry and Biology, Linköping University, SE-581 83 Linköping, Sweden.*
**Correspondence: E-mail: anna.eklof@liu.se*

'trait' for the latter. In fact, each idealised dimension could represent a combination of several empirical traits – for example, taking into account correlations between traits to create orthogonal dimensions using principal component analysis or ordination methods.

In graph theory, networks that can be completely described in one dimension are known as 'interval' (Fig. 1a). The first set of published ecological networks were small food webs that could be suitably described by one dimension (Cohen 1968) – the species can be ordered such that each predator consumes a consecutive range of prey. To date, most models of ecological network structure are based on a single dimension (Williams & Martinez 2000; Cattin *et al.* 2004; Stouffer *et al.* 2006; Allesina *et al.* 2008), even though it has long been proved that most ecological networks are only close to interval (Cattin *et al.* 2004; Stouffer *et al.* 2006; Mouillot *et al.* 2008). Interestingly, it has been shown that networks constructed from many dimensions can also appear close to interval (Brännström *et al.* 2011). Thus, quasi-intervality does not guarantee that ecological networks can be described using few dimensions, and there is as yet no rigorous estimate of the dimensionality of ecological networks.

Here, we provide, for the first time, an upper bound to the dimensionality of ecological networks, and identify which traits contribute the most to explaining interactions. To answer these fundamental questions we perform two analyses. First, we measure how many idealised dimensions are needed to completely describe the structure of a large set of empirical networks. Second, for a subset of the networks that include traits for all species, we estimate the fraction of connections explained by each trait or a combination of traits. We can thereby give a sound answer to the actual number of dimensions needed to completely describe ecological networks and quantify the actual importance of specified species traits.

We find that less than 10 dimensions are sufficient to completely describe ecological networks. Moreover, performing model selection we show that all networks are best explained by less than five dimensions. The analysis of traits shows that empirically measurable characteristics largely account for ecological interactions.

METHODS

Estimating an upper bound of dimensionality

Figure 1a shows a food web in which species i consumes all species whose body size is larger than b_i but smaller than B_i – the diet of species i is described by a range of body sizes. If this holds for all species, then the food web is described by a single dimension (body size). Fig. 1b gives an example for two dimensions: predator i consumes all species with body size within the range $[b_i, B_i]$ and are found in the depth range $[d_i, D_i]$ in the water column. Species falling within the body size range, but living at a depth outside the predator's range are not consumed. Such a food web requires two dimensions, and is therefore not interval (Cohen 1968; Stouffer *et al.* 2006; Allesina *et al.* 2008; Williams & Purves 2011). If two traits are sufficient to completely describe the diet of each predator (all of the predator's prey fall within the rectangle, see Fig. 1b) the network dimensionality is two. If three traits are sufficient, prey are embedded in a cube, and so forth. Therefore, the total number of parameters required for the model is three (minimum value, maximum value and actual trait value) for each species and dimension.

To identify the minimum number of dimensions needed to completely describe all species interactions in an ecological network, we need to consider all possible arrangements of the species in several

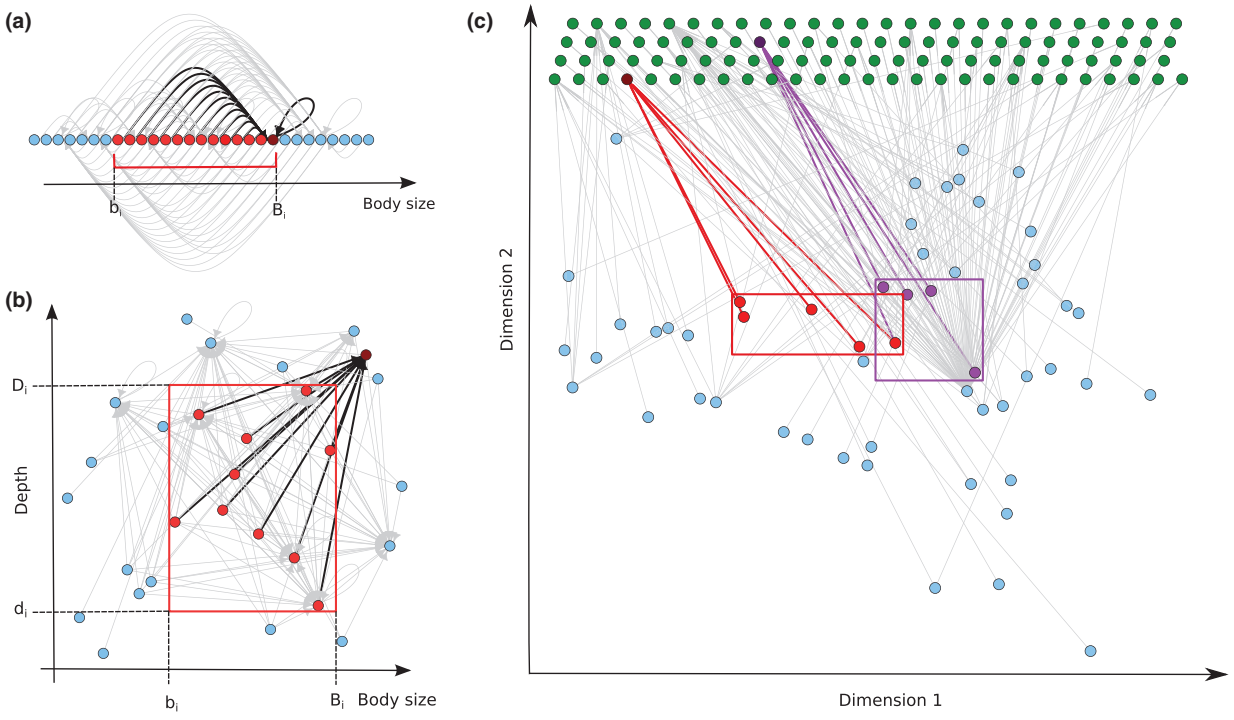

Figure 1. Schematic description of network arrangement in one (a) and two dimensions (b). In (a) the focal species (red) interacts with all species in the body size range $[b_iB_i]$. In (b) the focal species interacts with the species in the body size range $[b_iB_i]$ but only if they are also present at depth $[d_iD_i]$. Arrangement of the species in two dimensions for an empirical networks: the pollination network of the Galapagos Island (McCullen 1993) (c). The network has $\mathcal{D} = 2$; only the pollinators (blue) visiting each flower (green) are contained in each box (two highlighted).

dimensions. For S species, there are $S!$ possible arrangements in one dimension, $(S!)^2$ arrangements in two dimensions and $(S!)^X$ in X dimensions. This makes the task of determining the exact dimension of large networks computationally unfeasible. However, we can estimate the upper limit of dimensionality using heuristics (e.g. simulated annealing or genetic algorithms, see Supporting Information).

Suppose that we want to arrange species in two dimensions ($X = 2$, Fig. 1b-c). For each predator i, we define the 'prey box' as the smallest possible rectangle among all the arrangements in the plane such that all prey of i are contained in the box. If for each predator the corresponding box contains only its prey, the network can be completely explained by two dimensions. If, on the other hand, non-prey species are present in one or more of the boxes, we count them as errors, E_2. We define $\mathcal{D}_{\text{Pred}}$ as the minimum number of dimensions X such that $E_X = 0$ for the complete network. Similarly, we define the 'predator box' of species i as the smallest rectangle containing all the predators of i, and define $\mathcal{D}_{\text{Prey}}$ as the minimum number of dimensions in this direction. A 'predator-box' or 'prey-box' can contain predators/prey that are common for several species. Fig. 1c shows an example of a plant–pollinator network (Ueckert & Hansen 1971) with $\mathcal{D}_{\text{Pred}} = 2$. For plant–animal interaction networks, we consider the 'Pred' direction to be that of the animals choosing plants, and for parasite–host networks that of parasites choosing hosts.

A third way of arranging species, in addition to the predator box and prey box arrangements described above, would be for the predator box to contain only the predators and, at the same time, for the prey box to contain only the prey of a given species. Such an arrangement would lead to $\mathcal{D}_{\text{Pred and Prey}}$. However, necessarily $\mathcal{D}_{\text{Pred and Prey}} \geq \max(\mathcal{D}_{\text{Pred}}, \mathcal{D}_{\text{Prey}})$. Given that we are interested in the minimum number of dimensions needed to explain an empirical web, we consider only the first two cases, and define $\mathcal{D} = \min(\mathcal{D}_{\text{Pred}}, \mathcal{D}_{\text{Prey}})$. In graph theory, this quantity is called boxicity (Roberts 1969).

We analysed 200 ecological networks, including food webs, mutualistic and antagonistic networks spanning over a wide geographical range, and estimated $\mathcal{D}$ using several global optimisation algorithms (SI), as no polynomial-time algorithm can estimate boxicity (Kratochvíl 1994). The strategy, similar to that employed in studies of intervality (Stouffer *et al.* 2006; Stouffer *et al.* 2011; Zook *et al.* 2011), is to use search algorithms that attempt to minimise the number of erroneously predicted interactions E_X for the network by repeatedly swapping the position of any two species in a given dimension. Since it is unfeasible to try all possible arrangements, this process necessarily overestimates E_X and therefore the minimum number of dimensions $\mathcal{D}$.

Model selection

As stated above, $\mathcal{D}$ is an overestimate of the true number of dimensions. Actually, a single error ($E_X = 1$) is sufficient to reject $\mathcal{D} = X$. Because of this sensitivity, it is important to embed some level of uncertainty in the measure of $\mathcal{D}$. We use a probabilistic model in which a species included in the X-dimensional prey box of a predator is consumed with probability p_X; otherwise, it is not consumed. This probabilistic model, which can be seen as a simple multidimensional niche model (Allesina *et al.* 2008; Williams & Purves 2011), allows us to choose the most appropriate number of dimensions for modelling the network, $\mathcal{D}^* \leq \mathcal{D}$. If L is the number of interactions in the empirical data and E is the number of erroneous interactions that our model predicts, we can set the probability to its maximum likelihood estimate $\hat{p}_X = L/(L + E_X)$ and use the AIC (Akaike information criterion) (Burnham & Anderson 2002) to determine $\mathcal{D}^*$. The probabilistic model is more robust than $\mathcal{D}$ to sampling problems in the data or any inaccuracy due to the optimisation routine: E_X must decrease by a large amount to reduce $\mathcal{D}^*$ from $X + 1$ to X.

Given that we are analysing different types of networks, it is also of interest to investigate whether different network types have different scaling properties. To this end, we regressed the logarithm of the number of connections ($\log(L)$) against $\mathcal{D}$, $\mathcal{D}^*$ and $\log(\mathcal{AIC})$ (Fig. 2 main text, Fig. S1). For each regression [linear for $\log(\mathcal{AIC})$, Poisson in the other two cases], we contrasted two models: one in which all webs were grouped together, the other in which the networks were divided into three coarse grained groups: food webs, bipartite mutualistic networks (e.g. plant–pollinator) and bipartite antagonistic networks (e.g. host–parasite). We then use model selection (AIC) to investigate whether the network types are best modeled together or separately.

Predicting structure using traits

The dimensions analysed above can not directly be related to empirical traits, since each dimension could potentially represent a combination of several traits. Therefore, we additionally analyse how well empirically measurable and ecologically important traits can predict the structure of ecological networks. In other words, we want to predict species interactions based exclusively on species traits. We focus on traits that can be measured using individual specimens, and not factors pertaining to the population as a whole (e.g. abundances). Ideally, trait values should be simple to collect for published data as well as measurable in the field or laboratory as new data are collected (for detailed information on the traits used, see Supporting Information). Ideally, measuring relevant species' traits on a few sampled individuals should predict the position of the species in an ecological network.

We compiled a database of traits for 18 highly resolved networks spanning different interaction types and a wide geographic area (see Supporting Information). All traits are either continuous (e.g. body size, corolla depth), categorical (e.g. metabolic category, flower colour), or spatial and temporal match (presence–absence in a certain habitat or month) (see SI). Ordinal traits are treated as categorical. In bipartite mutualistic networks (pollination networks and frugivore networks), the traits are described either for the plants (resources) or the animals (consumers), for example, flower colour and proboscis length respectively. In bipartite antagonistic networks and food webs, the same traits are measured for both resources and consumers, for example, body mass and habitat.

For continuous traits, a species i interacts with a species j if the trait value η_j is included in the interval bounded by the minimum (m_i) and maximum (M_i) trait value for species i. For categorical variables, if i interacts with species that belong to one of the categories $\{a,b,c\}$, then i will interact with all species whose category falls in this set. Finally, for spatial and temporal match traits, suppose that species i is observed in {Jan, Feb, Mar}: species i will then interact with all species present in at least one of these months. In bipartite networks, the nodes are – by definition – partitioned in two groups and interactions occur exclusively between groups. Therefore, plants can only interact with animals, even if a trait, such as preferred habitat, is common to both groups.

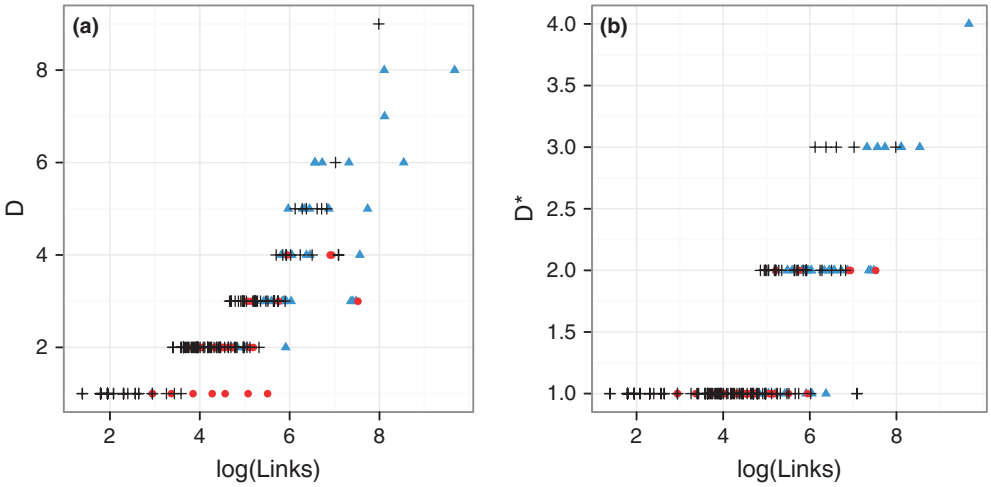

Figure 2. $\mathcal{D}$ (the number of dimensions, left) and $\mathcal{D}^*$ (the number of dimensions chosen as 'best' by model selection, right) as a function of the number of links in the network. Blue triangles represent food webs, black crosses represent bipartite mutualistic networks and red circles represent bipartite antagonistic networks. The minimum dimension $\mathcal{D}$ spans 1–9, with mean 2.665 and variance 2.1. The by model selection chosen as the 'best' dimension, $\mathcal{D}^*$ has a tighter distribution, with mean 1.395 and variance 0.38.

For all networks and each combination of traits, we measured the proportion of correctly predicted connections. For example, suppose we are using three traits to describe a food web: body size, metabolic category and mobility. In the empirical data, represented by the adjacency matrix A, all the prey of predator i have size in the interval [2,4]cm, are *invertebrates* or *ectotherm vertebrates* and have *low* or *medium* mobility. We can then build a new network (matrix) A' in which i preys upon all the species satisfying all of these three conditions. Clearly, A' contains all the connections in A, but potentially contains additional erroneous connections (species that satisfy the requirements on size, metabolism and mobility, but are not actually consumed by the predator). If E is the number of erroneous connections and L the number of empirically observed connections, we can estimate the performance of this combination of traits by computing $p = L/(L + E)$, the proportion of correctly predicted links, also known as the overlap (Petchey *et al.* 2008; Allesina 2011). For each network, we tested all possible combinations of up to seven traits plus all traits combined.

RESULTS

Dimensionality

In 35 cases, all of which are webs with less than 250 connections, we found the networks to be exactly interval, and thus $\mathcal{D} = 1$. In the other 165 networks, two or more dimensions were needed (Fig. 2). The maximum number of dimensions needed to describe all interactions correctly was always smaller than 10 (Fig. 2). In general, $\mathcal{D}$ scales almost linearly with the logarithm of the number of connections in the web, $\log(L)$ (Fig. 2). The largest number of dimensions needed is for the Phrygana pollination network (Petanidou 1991) with $\mathcal{D} = 9$. All but four networks require $\mathcal{D} \leq 6$.

As stated above, $\mathcal{D}$ is an overestimate of the true number of dimensions and even a single error ($E_X = 1$) is sufficient to reject $\mathcal{D} = X$. For example, the Weddell Sea food web (Jacob 2005) (488 species, 15 880 connections) has $\mathcal{D} = 8$ and $E_7 = 1$ (Table S4 and S5, SI). Thus, adding a single link to the 15 880 already present would reduce $\mathcal{D}$ to 7. Using the probabilistic approach as model selection, we find that all networks are best modeled using four or fewer dimensions (Fig. 2b).

We additionally tested whether network type influences dimensionality. Although the variation between groups in Fig. 2 might seem negligible, generalised linear models show that food webs, bipartite mutualistic and bipartite antagonistic networks yield different relations between $\mathcal{D}$ and $\log(L)$. We find that model selection (AIC) consistently favours keeping the groups separated, supporting type-specific scaling (Table S1).

Predicting structure using traits

When analysing how well network structure can be explained by empirical traits, we find that as few as three empirical traits can explain about one third of the interactions in the worst case (the food webs Caribbean reef and Loughhyne, Table 1). A single trait can predict a relatively large proportion of network structure (between 11 and 100%, Table 1), and always performs much better than a random graph (which would yield an expected overlap equal to the network connectance). Combining several traits necessarily increases the overlap, often considerably. However, since we observe diminishing marginal returns as traits are added, the results for three traits are often close to those obtained when all traits are considered together (Table 1). The results we show are consistent with the theoretical bounds proved above, demonstrating that network structure can be revealed using few traits.

The traits with the largest explanatory power tend to be nested: the best single trait is also present in the best combination of two traits, which are contained in the best combination of three traits, and so forth. This holds for all the networks except two (the food webs St. Marks and Kongsfjorden). The variation in the number of links that the traits can correctly predict can be ascribed to the the different types and sizes of networks and also to differences in the type of traits recorded (Table 1, SI).

Table 1. Overlap for the single trait giving the highest overlap (best1), best combination of two (best2), three (best3) and all traits (bestAll, in parenthesis the number of traits combined). Letter combinations in parenthesis identify the traits. The lower-case letter indicates whether the trait is coupled to the resources (r) or consumers (c). For a detailed description of the traits, see SI. bim – bipartite mutualistic network, bia – bipartite antagonistic network and fw – food web. S – number of species, in bipartite networks the number belonging to each group is stated. C – network connectance

Network	type	S	C	best1	best2	best3	bestAll
Puerto Rico, highland†(Dalsgaard *et al.* 2009)	bim	11 + 2	0.59	1 (BILc/BMc)	1 (BILc/BMc + any)	1 (BILc/BMc + any + any)	1 (9)
NZ landuse*	bim	15 + 16	0.17	0.44 (BWc)	0.79 (BWc + NDr)	0.91 (BWc + NDr + STlr)	0.98 (19)
Santa Genebra (Galetti & Pizo, 1996)	bim	29 + 33	0.14	0.40 (BIGc)	0.58 (BIGc + FSr)	0.65 (BiGc + FSr + BMc)	0.89 (11)
Villavicencio (Chacoff *et al.* 2012)	bim	41 + 80	0.18	0.35 (ORr)	0.46 (ORr + SLr)	0.56 (ORr + SLr + PWc)	0.90 (21)
Garraf (Bosch *et al.* 2009)	bim	19 + 165	0.26	0.62 (FBLr)	0.79 (FBLr + BLDr)	0.87 (FBLr + BLDr + POLr)	0.95 (12)
Ecuador LU-gradient (Tylianakis *et al.* 2007)	bia	29 + 9	0.18	0.58 (BLc)	0.75 (BLc + BLr)	0.75 (BLc + BLr + DPc)	0.75 (8)
NZ alpine grassland*	bia	38 + 31	0.085	0.34 (BMc)	0.74 (BMc + BMr)	0.75 (BMc + BMr + TMPc)	0.75 (6)
Ythan (Cohen *et al.* 2009)	fw	92	0.049	0.32 (BMc)	0.55 (BMc + BMr)	0.61 (BMc + BMr + HBc)	0.74 (12)
StMarks (Christian & Luczkovich, 1999)	fw	143	0.086	0.25 (BMc)	0.45 (BMc + MBr)	0.55 (BMr + MBr + HBc)	0.82 (12)
Caribbean reef (Optiz 1996)	fw	249	0.053	0.17 (BMr)	0.26 (BMr + BMc)	0.33 (BMr + BMc + MBr)	0.42 (12)
Kongsfjorden (Jacob *et al.* 2011)	fw	270	0.023	0.11 (MCr)	0.25 (HBr + BMc)	0.39 (MCr + BMc + MBr)	0.69 (12)
Loughhyne (Riede *et al.* 2010)	fw	349	0.042	0.15 (BMr)	0.24 (BMr + BMc)	0.33 (BMr + BMc + MBr)	0.47 (12)
Weddell (Jacob, 2005)	fw	488	0.067	0.20 (MBr)	0.30 (BMr + MBr)	0.40 (BMr + BMc + MBr)	0.61 (12)

*the data are available in Supporting Information.
†see Supporting Information for additional networks of the same type.
Trait identifiers: BIL, bill length; BM, body mass; BW, body width; ND, amount nectar; STl, flower type; BIG, bill gape; FS, fruit size; OR, orientation; SL, stamen length; PW, proboscis width; FBL, first bloom; BLD, bloom duration; POL, pollen volume per flower; BL, body length; DP, dates present; TMP, temperature envelope; HB, habitat; MB, mobility and MC, metabolic category.

DISCUSSION

Our results show that ecological networks are structured by few dimensions, with model selection suggesting that four or fewer dimensions can largely account for network structure. This supports the traditional idea of species interactions being determined by low-dimensional 'niche-spaces'.

Models for food web structure have shown that information on a single species trait, such as body size or trophic level, can describe the structure of empirical networks fairly accurately (Williams & Martinez 2000; Stouffer *et al.* 2006; Petchey *et al.* 2008; Zook *et al.* 2011). However, a complete description of the network requires multiple traits (Cattin *et al.* 2004; Allesina *et al.* 2008; Allesina 2011). Rossberg *et al.* (2010) showed that the degree of intervality tends to increase with the number of underlying dimensions. Although several studies set a lower bound for the dimensionality of ecological networks (Cattin *et al.* 2004; Stouffer *et al.* 2006), we were here able to determine an upper bound, and the most likely number of dimensions needed. Our analysis allows us to give an estimate of how complex different ecological networks are.

Our results can be compared to the minimum number of dimensions needed to explain the structure of a network in graph theory – the graph's boxicity (Roberts 1969). The boxicity of an undirected graph composed of S nodes is bounded from above by $\min\left(\frac{S}{2}, (\Delta + 2)\log(S)\right)$, where Δ is the number of connections of the most connected species (Sunil & Ashik 2009). For example, the Greek phrygana pollination network (Petanidou 1991) contains 797 species and 2933 interactions, $\Delta = 124$, and thus the upper bound for the dimension is $\mathcal{D} \leq \min(399, 842) = 399$. We find 9. For an undirected random graph with S nodes and probability of connection C, the boxicity is almost surely $SC(1-C)$ (Adiga *et al.* 2008). Because the boxicity of a directed graph should be lower than that of its undirected version, this should overestimate the true dimension D. Take Weddell Sea, where we have 488 species and a connectance $C = 0.067$. When we make the graph undirected, we basically double the connectance $C' = 0.134$. Hence, we should expect D to be less than 57. We find 8. For a balanced undirected bipartite graph (i.e. where the set of plants of pollinators are of the same size), with $2S$ nodes and L links (such that $L < S^2/3$) the boxicity is almost surely approximately L/S (Adiga *et al.* 2008). In the Chilean pollination network studied (Arroyo *et al.* 1982), this would translate a dimension close to 4. We find that three dimensions are sufficient.

When analysing empirical traits, we showed that by using three traits we can predict the position of possible ecological interactions, and that more than a third of the predicted connections are in fact realised. This fraction might seem low, but one has to recognise that using simple information we are ruling out most of the interactions. Take for example, the Weddell Sea food web composed of 488 species. In this network, there could be up to 238 144 connections. Using three traits, we can focus on 39 700 connections, of which 15 580 are realised. Thus, with the use of three traits (body mass of the consumer, body mass of the resource and mobility of the resource), we ruled out more than 198 000 coefficients as 'forbidden links'.

The proportion of interactions that can be explained using empirical traits varies considerably between networks. This can be ascribed mainly to two factors. First, networks belong to different types, describing plant–pollinator, host–parasitoid, plant–herbivore

and predator–prey interactions. Second, the traits collected vary between networks both in number and identity (Table S2). Although one might think that analysing the same traits for all networks would be ideal, very few traits are shared by all species, and those that are might not be informative in all contexts. For example, flower colour or corolla size could be important for pollination, but not for herbivory. In our analysis, we focused on traits that were collected exactly because they are thought to explain the interactions occurring in the network.

Despite the variation in traits, we find some consistency in the results for the most important traits. For example, for food webs – which include a wide variety of organisms – body size is always among the selected traits. However, for two of the six food webs, 'simpler' traits (i.e. categorical traits such as mobility and metabolic category) actually yield a higher overlap between the predicted and empirical interactions. Traits related to body size are also selected in the two parasitoid networks. For one of the pollination networks (NZ landuse) body width of the pollinator is the single trait giving the largest overlap, and in the plant–hummingbird networks hummingbird bill length and hummingbird body mass both describe all the interactions correctly. In the other pollination networks, different types of traits play the most important role, with attributes of the plants being the strongest predictors.

Several studies have shown that species abundance is a good predictor of aggregate properties of ecological networks (Vázquez *et al.* 2009; Cagnolo *et al.* 2011), yet we have not considered abundance here. The reason is that abundance is not a trait itself, but rather an emergent population level property determined by individual traits constraining population size, growth and success. Thus, traits could influence interactions either directly (e.g. through the trait matching rules that shape interactions) or indirectly (e.g. through the traits that determine species abundance).

Models aiming at predicting the structure of ecological networks usually focus on the consumer's perspective (Cattin *et al.* 2004; Petchey *et al.* 2008; Jacob *et al.* 2011): the traits of the consumer define whether an interaction will be observed. Our results question this approach in two ways. First, the best combination of two traits typically involves the matching of consumer and resource traits, for example, the body mass of consumer combined with the body mass for the resource (in five food webs), the bill gape of the consumers combined with the fruit size (in the frugivore network). Second, for six of the networks the single trait predicting most of the structure is based on resource characteristics. In addition, in three of the networks the best combination of two traits involves only resource traits, and in none of the networks does the best combination of both traits pertain to the consumer. This highlights that focusing exclusively on consumer traits could limit progress in ecological networks.

The importance of matching traits is perhaps not surprising, but nevertheless not self-evident. We can hypothesise that other, non-matching traits could be important for one of the interactors. For example, in a pollination network, we would *a priori* expect that the corolla depth of the flower and the proboscis length of the pollinator need to match for an interaction to be possible. However, some pollinators bite holes at the base of deep-corolla flowers to reach the nectaries in spite of their short proboscis (e.g. Inouye 1980). Therefore, the most important trait determining visit may instead be, for example, flower colour.

In summary, the analysis of the empirical traits reveals that (1) body size is important, but the proportion of the interactions that can be explained by this trait increases drastically when used in combination with additional traits, (2) combining the consumer and resource perspectives, that is, matching traits, can dramatically improve the accuracy of the predictions and (3) in different types of networks different traits are selected, highlighting that different types of networks require distinct modelling approaches.

Species phylogenies are strongly coupled to species traits and have been shown to provide important information on the structure of food webs (Bersier & Kehrli 2008; Rossberg *et al.* 2010; Eklöf *et al.* 2012). In fact, taxonomic and phylogenetic information can be seen as a summary of several traits shared by closely related species. However, whether an interaction between two organisms might occur or not depends exclusively on phenotypic traits. Thus, even though phylogeny can be used as a surrogate for species traits, gaining an understanding of which phenotypic traits drive interactions in ecological networks is important.

We have found that dimensionality scales with network complexity – completely explaining the structure of larger networks requires more dimensions compared to smaller networks. The main reason is that larger networks contain a larger variety of species and interactions. For example, the Caribbean hummingbird networks [Puerto Rico, highland in Table 1, and additional networks in Table S3, Dalsgaard *et al.* (2009)] describe pollination between a family of pollinators and a few plant species, while the Weddell Sea food web includes hundreds of species, spanning different Kingdoms and interaction modes (herbivory, carnivory, omnivory, detritivory) (Jacob 2005). Naturally, large and highly resolved networks, integrating interactions over larger areas and time spans, will include a more diverse set of interactions, possibly driven by different traits, leading to higher dimensionality. However, we found that dimensionality increases slowly with network complexity. This means that for even very large systems the number of dimensions will be fairly low.

ACKNOWLEDGEMENTS

We thank P. Staniczenko, M. Smith and S. Tang for comments. We also thank Dominique Gravel and one anonymous reviewer for comments and suggestions. AE and JCK supported by NSF EF # 0827493. AE also by SOEB. SA supported by DEB #1148867. BD supported by Carlsberg foundation and the Danish Council for Independent research. AMMG supported by Spanish Ministry of Education, BVA-2010-0845. PRG supported by FAPESP. MAP supported by CNPq. JMT supported by a Rutherford Discovery Fellowship, RSNZ. RCU, NPC, SBL and DPV supported by CONICET (PIP 2781 and 6564), FONCyT (PICT 1471 and 20805) and BBVA Foundation (BIOCON03-162). JB and AR supported by Spanish MICNN (CICYT CGL2005-00491) and CONSOLIDER (csd2008-0040).

AUTHORSHIP

AE and SA: Originated the project, selected authors, wrote the code, ran the high-performance computations, wrote the manuscript, wrote the Supplement. JCK: Organised the data, ran the code, edited the manuscript. Other Authors: contributed original data, commented on the manuscript, wrote part of the Supplement.

REFERENCES

Adiga, A., Chandran, L.S. & Sivadasan, N. (2008). Lower bounds for boxicity. arXiv preprint arXiv:0806.3175. Only in the arXiv.

Allesina, S. (2011). Predicting trophic relations in ecological networks: a test of the allometric diet breadth model. *J. Theor. Biol.*, 279, 161–168.

Allesina, S., Alonso, D. & Pascual, M. (2008). A general model for food web structure. *Science*, 320, 658–661.

Arroyo, M.T.K., Primack, R. & Armesto, J. (1982). Community studies in pollination ecology in the high temperate andes of central chile. i. pollination mechanisms and altitudinal variation. *Am. J. Bot.*, 69, 82–97.

Bersier, L.F. & Kehrli, P. (2008). The signature of phylogenetic constraints on food-web structure. *Ecol. Compl.*, 5, 132–139.

Bosch, J., Martín González, A.M., Rodrigo, A. & Navarro, D. (2009). Plant–pollinator networks: adding the pollinators perspective. *Ecol. Lett.*, 12, 409–419.

Brännström, Å., Carlsson, L. & Rossberg, A.G. (2011). Rigorous conditions for food-web intervality in high-dimensional trophic niche spaces. *J. Math. Biol.*, 63, 575–592.

Brown, J.H., Gillooly, J.F., Allen, A.P., Savage, V.M. & West, G.B. (2004). Toward a metabolic theory of ecology. *Ecology*, 85, 1771–1789.

Burnham, K.P. & Anderson, D.R. (2002). *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*. Springer Verlag, New York.

Cagnolo, L., Salvo, A. & Valladares, G. (2011). Network topology: patterns and mechanisms in plant-herbivore and host-parasitoid food webs. *J. Anim. Ecol.*, 80, 342–351.

Cattin, M.F., Bersier, L.F., Banas˘ek-Richter, C., Baltensperger, R. & Gabriel, J.P. (2004). Phylogenetic constraints and adaptation explain food-web structure. *Nature*, 427, 835–839.

Chacoff, N.P., Vázquez, D.P., Lomáscolo, S.B., Stevani, E.L., Dorado, J. & Padrón, B. (2012). Evaluating sampling completeness in a desert plant–pollinator network. *J. Anim. Ecol.*, 81, 190–200.

Chase, J.M. & Leibold, M.A. (2003). *Ecological Niches: Linking Classical and Contemporary Approaches*. University of Chicago Press, Chicago, IL.

Christian, R.R. & Luczkovich, J.J. (1999). Organizing and understanding a winter's seagrass foodweb network through effective trophic levels. *Ecol. Model.*, 117, 99–124.

Cohen, J.E. (1968). Interval graphs and food webs: a finding and a problem. RAND Corporation Document 17696-PR, Santa Monica, CA.

Cohen, J.E., Schittler, D.N., Raffaelli, D.G. & Reuman, D.C. (2009). Food webs are more than the sum of their tritrophic parts. *P. Nat. Acad. Sci. U.S.A.*, 106, 22335–22340.

Dalsgaard, B., Martín González, A.M., Olesen, J.M., Ollerton, J., Timmermann, A., Andersen, L.H. & Tossas, A.G. (2009). Plant–hummingbird interactions in the west indies: floral specialisation gradients associated with environment and hummingbird size. *Oecologia*, 159, 757–766.

Diamond, S.E., Frame, A.M., Martin, R.A. & Buckley, L.B. (2011). Species' traits predict phenological responses to climate change in butterflies. *Ecology*, 92, 1005–1012.

Eklöf, A., Helmus, M.R., Moore, M. & Allesina, S. (2012). Relevance of evolutionary history for food web structure. *P. Roy. Soc. Lond. B: Bio.*, 279, 1588–1596.

Forrest, J., Miller-Rushing, A.J., Forrest, J. & Miller-Rushing, A.J. (2010). Toward a synthetic understanding of the role of phenology in ecology and evolution. *Philos. T. Roy. Soc. B*, 365, 3101–3112.

Galetti, M. & Pizo, M.A. (1996). Fruit eating by birds in a forest fragment in southeastern brazil. *Ararajuba*, 4, 71–79.

Gilman, R.T., Nuismer, S.L. & Jhwueng, D.C. (2012). Coevolution in multidimensional trait space favours escape from parasites and pathogens. *Nature*, 483, 328–330.

Inouye, D.W. (1980). The terminology of floral larceny. *Ecology*, 61, 1251–1253.

Jacob, U. (2005). *Trophic Dynamics of Antarctic Shelf Ecosystems – Food Webs and Energy Flow Budgets, Thesis*. University of Bremen, Germany.

Jacob, U., Thierry, A., Brose, U., Arntz, W.E., Berg, S., Brey, T., Fetzer, I., Jonsson, T., Mintenbeck, K., Mollmann, C. *et al.* (2011). The role of body size in complex food webs: a cold case. *Adv. Ecol. Res*, 45, 181–223.

Jordano, P., Bascompte, J. & Olesen, J.M. (2003). Invariant properties in coevolutionary networks of plant–animal interactions. *Ecol. Lett.*, 6, 69–81.

Kratochvíl, J. (1994). A special planar satisfiability problem and a consequence of its np-completeness. *Discrete Appl. Math.*, 52, 233–252.

McCullen, C.K. (1993). Flower-visiting insects of the galapagos islands. *Pan-Pac. Entomol.*, 69, 95–106.

Mouillot, D., Krasnov, B.R. & Poulin, R. (2008). High intervality explained by phylogenetic constraints in host-parasite webs. *Ecology*, 89, 2043–2051.

Optiz, S. (1996). *Trophic Interactions in Caribbean Coral Reefs*. Tech. Rep. 43, ICLARM, Manila.

Petanidou, T. (1991). *Pollination Ecology in a Phryganic Ecosystem, Thesis*. Aristotelian University, Greece.

Petchey, O.L., Beckerman, A.P., Riede, J.O. & Warren, P.H. (2008). Size, foraging, and food web structure. *P. Nat. Acad. Sci. U.S.A.*, 105, 4191–4196.

Riede, J.O., Rall, B.C., Banasek-Richter, C., Navarrete, S.A., Wieters, E.A., Emmerson, M.C., Jacob, U. & Brose, U. (2010). Scaling of food-web properties with diversity and complexity across ecosystems. *Adv. Ecol. Res.*, 42, 139–170.

Roberts, F.S. (1969). *On the Boxicity and Cubicity of a Graph*. Academic Press, New York.

Roberts, F. (1978). Food webs, competition graphs, and the boxicity of ecological phase space. Y. Alavi, D. Lick (Eds.), Theory and Applications of Graphs, Springer, New York, pp. 477–490.

Rossberg, A.G., Brännström, Å. & Dieckmann, U. (2010). Food-web structure in low-and high-dimensional trophic niche spaces. *J. Roy. Soc. Int.*, 7, 1735–1743.

Silvertown, J. (2004). Plant coexistence and the niche. *Trends Ecol. Evol.*, 19, 605–611.

Stouffer, D.B., Camacho, J. & Amaral, L.A.N. (2006). A robust measure of food web intervality. *P. Nat. Acad. Sci. U.S.A.*, 103, 19015–19020.

Stouffer, D.B., Rezende, E.L. & Amaral, L.A.N. (2011). The role of body mass in diet contiguity and food-web structure. *J. Anim. Ecol.*, 80, 632–639.

Sunil, C.L. & Ashik, M.K. (2009). An upper bound for cubicity in terms of boxicity. *Discrete Math.*, 309, 2571–2574.

Thompson, J.N. (2005). *The Geographic Mosaic of Coevolution*. University of Chicago Press, Chicago, IL.

Tylianakis, J.M., Tscharntke, T. & Lewis, O.T. (2007). Habitat modification alters the structure of tropical host–parasitoid food webs. *Nature*, 445, 202–205.

Ueckert, D.N. & Hansen, R.M. (1971). Dietary overlap of grasshoppers on sandhill rangeland in northeastern colorado. *Oecologia*, 8, 276–295.

Vázquez, D.P., Blüthgen, N., Cagnolo, L. & Chacoff, N.P. (2009). Uniting pattern and process in plant–animal mutualistic networks: a review. *Ann. Bot. London*, 103, 1445–1457.

Williams, R.J. & Martinez, N.D. (2000). Simple rules yield complex food webs. *Nature*, 404, 180–183.

Williams, R.J. & Purves, D.W. (2011). The probabilistic niche model reveals substantial variation in the niche structure of empirical food webs. *Ecology*, 92, 1849–1857.

Zook, A.E., Eklöf, A., Jacob, U. & Allesina, S. (2011). Food webs: ordering species according to body size yields high degree of intervality. *J. Theor. Biol.*, 271, 106–113.

SUPPORTING INFORMATION

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

Editor, Jennifer Dunne
Manuscript received 29 October 2012
First decision made 6 December 2012
Manuscript accepted 4 January 2013