

# A simple model of bipartite cooperation for ecological and organizational networks

Serguei Saavedra<sup>1,2,3</sup>, Felix Reed-Tsochas<sup>2,4</sup> & Brian Uzzi<sup>5,6</sup>

**In theoretical ecology, simple stochastic models that satisfy two basic conditions about the distribution of niche values and feeding ranges have proved successful in reproducing the overall structural properties of real food webs, using species richness and connectance as the only input parameters<sup>1–4</sup>. Recently, more detailed models have incorporated higher levels of constraint in order to reproduce the actual links observed in real food webs<sup>5,6</sup>. Here, building on previous stochastic models of consumer–resource interactions between species<sup>1–3</sup>, we propose a highly parsimonious model that can reproduce the overall bipartite structure of cooperative partner–partner interactions, as exemplified by plant–animal mutualistic networks<sup>7</sup>. Our stochastic model of bipartite cooperation uses simple specialization and interaction rules, and only requires three empirical input parameters. We test the bipartite cooperation model on ten large pollination data sets that have been compiled in the literature, and find that it successfully replicates the degree distribution, nestedness and modularity of the empirical networks. These properties are regarded as key to understanding cooperation in mutualistic networks<sup>8–10</sup>. We also apply our model to an extensive data set of two classes of company engaged in joint production in the garment industry. Using the same metrics, we find that the network of manufacturer–contractor interactions exhibits similar structural patterns to plant–animal pollination networks. This surprising correspondence between ecological and organizational networks suggests that the simple rules of cooperation that generate bipartite networks may be generic, and could prove relevant in many different domains, ranging from biological systems to human society<sup>11–14</sup>.**

In ecology, the collection and analysis of empirical data on mutualistic networks<sup>7</sup> can help identify the most significant features that have evolved in bipartite cooperative networks. In these networks, species in one class (class *A*, for animals) cooperate with species in a second class (class *P*, for plants) to mutual advantage. Species in class *P* offer rewards with certain characteristics determined by reward traits, which may also have evolved to reduce exploitation and favour mutualism<sup>15</sup>. Species in class *A* foraging for resources can benefit from the rewards offered by a given species in class *P* if the respective foraging traits (for example efficiency, morphology and behaviour) and reward traits (for example quantity, quality and availability) are complementary<sup>8,16</sup>. External factors such as the environmental context (for example population density and geographic and temporal variation) attenuate or amplify the value of reward and foraging traits, and have an impact on the number of potential partners with which a given species cooperates<sup>17–19</sup>. Furthermore, mutualistic networks exhibit constraints introduced by phylogenetic relationships between species in the same class, which impact mutualistic interaction patterns by favouring ecological similarity<sup>20</sup>.

Recently, mutualistic models<sup>21,22</sup> have been proposed which incorporate the effects of complementarity and exploitation or growth barriers on interactions between species in classes *A* and *P* (Supplementary

Information). Although a mixed model<sup>21</sup> and a differential limiting size model<sup>22</sup> have proved successful at generating structural patterns similar to those observed in real mutualistic networks, we find that they are able to reproduce less than 25% of the total number of observed metrics in different ecological networks (Supplementary Tables 1–4).

In organizational networks, cooperation between two different classes of company (manufacturers and contractors) is subject to equivalent structural constraints, which depend on the traits of the companies and the complementarity between traits of potential partners, as well as hierarchical relationships between companies in the same class<sup>23</sup>. Companies are characterized by a set of organizational traits<sup>24</sup> (for example company size, competitive niche space and brand positioning), and face interaction barriers generated by differences in status<sup>25</sup>, which limit the number and range of potential partners. Again, the values associated with trait complementarity and interaction barriers are not absolute, but are modulated by the specific market context in which the companies operate<sup>24,25</sup>.

Inspired by previous food-web models<sup>1–3</sup>, we develop a new model of bipartite cooperation that can reproduce more than 70% of the total number of observed metrics in both ecological and organizational networks (Table 1, Supplementary Tables 3, 4). In the corresponding ecological (species-based) and organizational (company-based) contexts, plants and manufacturers are treated as members of class *P*, and animals and contractors as members of class *A*. The three inputs for the model are the size of class *A*, the size of class *P* and the total number of links, *L*, all of which are given directly by the empirical data (Table 1). The model consists of two mechanisms (see Methods for a description of the model).

The first mechanism is specialization. The specialization rule determines how many partners,  $l_p$ , each member  $p \in P$  will cooperate with. This number is determined by the reward value associated with  $p \in P$ , which is given by the reward trait,  $t_{Rp}$ , attenuated or amplified by an external factor  $\lambda_p$  that accounts for effects such as geographic variation and population diversity. Higher reward values increase the number of potential interactions. Reward traits  $t_{Rp}$  are the result of a hierarchical process, which in our model corresponds to the generation of an ordered sequence in trait space, so  $t_{Rp}$  has a role equivalent to that of the niche value in the niche model<sup>1</sup> or nested-hierarchy model<sup>2</sup>. Note that the specialization rule is only associated with class *P*. This is consistent with previous findings which show that external factors affect the level of specialization more strongly in plants than in animals<sup>18</sup>.

The second mechanism is interaction. The interaction rule determines which members  $a \in A$  cooperate with each member  $p \in P$ . Here interactions are limited by the complementarity between the reward traits,  $t_{Rp}$ , for  $p \in P$  and foraging traits (organizational traits),  $t_{Fa}$ , for  $a \in A$ . The foraging trait  $t_{Fa}$  limits the range of possible partners for each member of class *A*, but again external factors  $\lambda_p$  such as temporal variation and population density can modify these

<sup>1</sup>Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK. <sup>2</sup>CABDyN Complexity Centre, <sup>3</sup>Corporate Reputation Centre, <sup>4</sup>James Martin Institute, Saïd Business School, University of Oxford, Oxford OX1 1HP, UK. <sup>5</sup>Kellogg School of Management and Northwestern Institute on Complex Systems, Northwestern University, Evanston, Illinois 60208, USA. <sup>6</sup>Haas School of Business, University of California, Berkeley, California 94720, USA.

**Table 1 | Empirical values and model statistical significance**

Data set/environment	$L$	$ P $	$ A $	$KS_P-KS_A$	$N$	$Q$
Marsh (Japan)	430	64	187	0.326*-0.438*	0.976* (0.969)	0.551* (0.553)
Grassland (Cass, New Zealand)	374	41	139	0.633*-0.385*	0.957* (0.960)	0.474† (0.465)
Subalpine forest/meadow (Japan)	865	90	354	0.552*-0.001§	0.985† (0.976)	0.545† (0.532)
Subalpine (Arthur's Pass, New Zealand)	120	18	60	0.108†-0.999*	0.858§ (0.936)	0.553‡ (0.527)
Subalpine (Craigieburn, New Zealand)	346	49	118	0.002§-0.001§	0.961* (0.955)	0.480† (0.468)
Tundra (Canada)	179	29	81	0.097†-0.989*	0.971† (0.950)	NM
Scrub/snow gum forest (Australia)	252	36	81	0.608*-0.076†	0.935* (0.949)	NM
Deciduous forest (USA)	65	7	33	0.911*-0.642*	0.953† (0.930)	NM
Arctic tundra (Greenland)	453	31	75	0.038‡-0.118†	0.793§ (0.914)	NM
Subarctic rock slope (Sweden)	242	24	118	0.223†-0.005‡	0.927† (0.952)	NM
New York garment industry, 1985	7,250	823	2,562	0.061†-0.115†	0.997† (0.996)	0.598§ (0.502)
New York garment industry, 1991	3,981	325	1,590	0.101†-0.531‡	0.994* (0.993)	0.601§ (0.529)
New York garment industry, 1997	1,450	148	700	0.003‡-0.264†	0.990† (0.988)	0.653‡ (0.625)
New York garment industry, 2003	228	62	128	0.370†-0.002‡	0.976† (0.969)	0.711† (0.700)

For each pollination data set and the four organizational networks used in this paper, the table presents its environment/location; the total number of links,  $L$ ; the respective numbers of nodes,  $|P|$  and  $|A|$ , in classes  $P$  and  $A$ ; the combined Kolmogorov–Smirnov (KS) probabilities,  $KS_P-KS_A$ , calculated for the degree distributions using the two-group equivalence KS test between the empirical and model-generated distributions for classes  $P$  and  $A$ , respectively; the observed nestedness,  $N$ ; and the observed mean modularity value,  $Q$ . Note that all networks have a ratio  $|P|/|A| < 0.5$ , which has been found to be an important factor limiting the appearance of scale-free distributions<sup>22</sup>. The model-generated mean values for  $N$  and  $Q$  are shown inside parentheses. Five of the observed pollination networks have already been found to be non-modular (NM)<sup>10</sup>. All comparisons are based on 1,000 model simulations. Note that the model reproduces more than 70% of the overall number of observed metrics with a good or excellent fit (27 out of 35 and 11 out of 16 for the ecological and organizational networks, respectively).

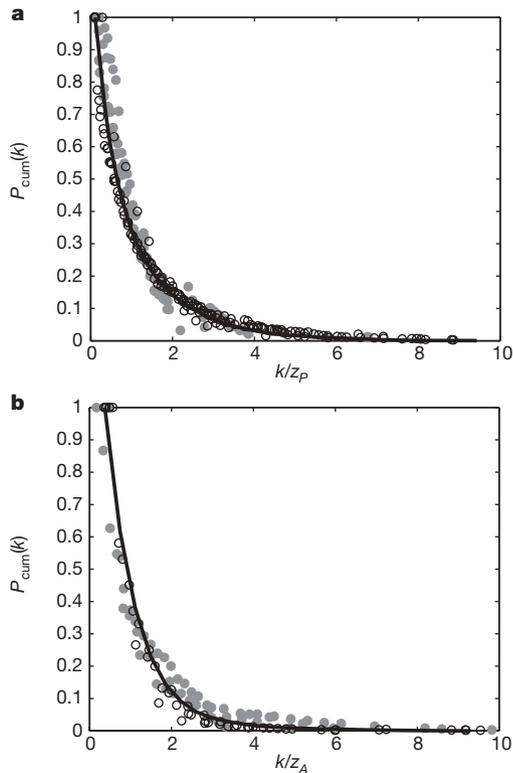
\*  $KS \geq 0.30$ , normalized errors less than one model s.d. (excellent fit). †  $KS < 0.30$ , normalized errors between one and two model s.d. (good fit). ‡  $KS < 0.05$ , normalized errors between two and three model s.d. (poor fit). §  $KS < 0.01$ , normalized errors greater than three model s.d. (bad fit).

interaction barriers. In both the specialization and interaction rules, we assume that traits are normally distributed and that the effect of external factors on members is inhomogeneous and a function of the diversity of members and the density of interactions in the network.

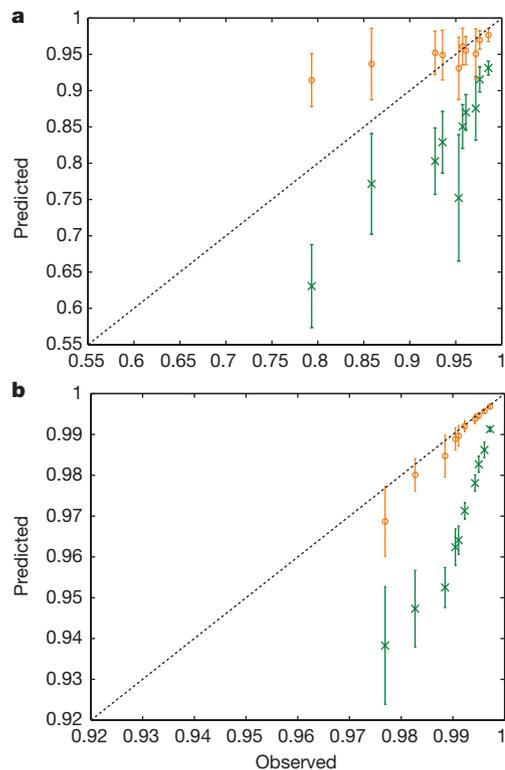
We test our model using data from real-world ecological and organizational networks. For the ecological networks, we use a diverse set of ten extensive plant–animal pollination networks compiled in the literature (Supplementary Information), which can clearly be distinguished from random assemblages<sup>9</sup>. For organizational networks, we

use a data set of approximately 700,000 yearly bilateral production transactions between more than 10,000 manufacturers and contractors in the New York garment industry (NYGI) from January 1985 to December 2003 (Supplementary Information)<sup>23</sup>. The NYGI exhibits a significant turnover of companies each year with different declining trajectories for each class (Supplementary Fig. 1), so the network structure at one time does not trivially map into the structure at other times. For each network, we investigate three key features of bipartite cooperation<sup>8–10</sup>: (1) degree distribution<sup>26</sup>, (2) nestedness<sup>27,28</sup> and (3) modularity<sup>29</sup>.

First we consider the degree distribution. In Fig. 1a and b we respectively show the cumulative distributions for members of class  $P$



**Figure 1 | Scaled degree distribution.** **a**, The cumulative degree distribution,  $P_{cum}(k)$ , for members of class  $P$  (plants, manufacturers); **b**,  $P_{cum}(k)$  for members of class  $A$  (animals, contractors). The number of partners,  $k$ , is scaled by a multiplicative factor of  $1/z_P$  for members of  $P$  and  $1/z_A$  for members of  $A$ , where  $z_P = L/P$  and  $z_A = L/A$ . Filled symbols correspond to pollination networks and open symbols to organizational networks. Note that all distributions collapse into a single curve. The solid line corresponds to the model-generated distributions averaged over 1,000 simulations (Supplementary Fig. 2).



**Figure 2 | Nestedness.** **a**, **b**, Comparisons of the randomly generated (green crosses) and model-generated (orange circles) nestedness values (mean  $\pm$  2 s.d.) with the observed nestedness values for all the pollination (**a**) and the NYGI (**b**) networks. The dashed lines correspond to there being perfect agreement between predicted and observed values. Note that a matrix with a nestedness value of one is perfectly nested.

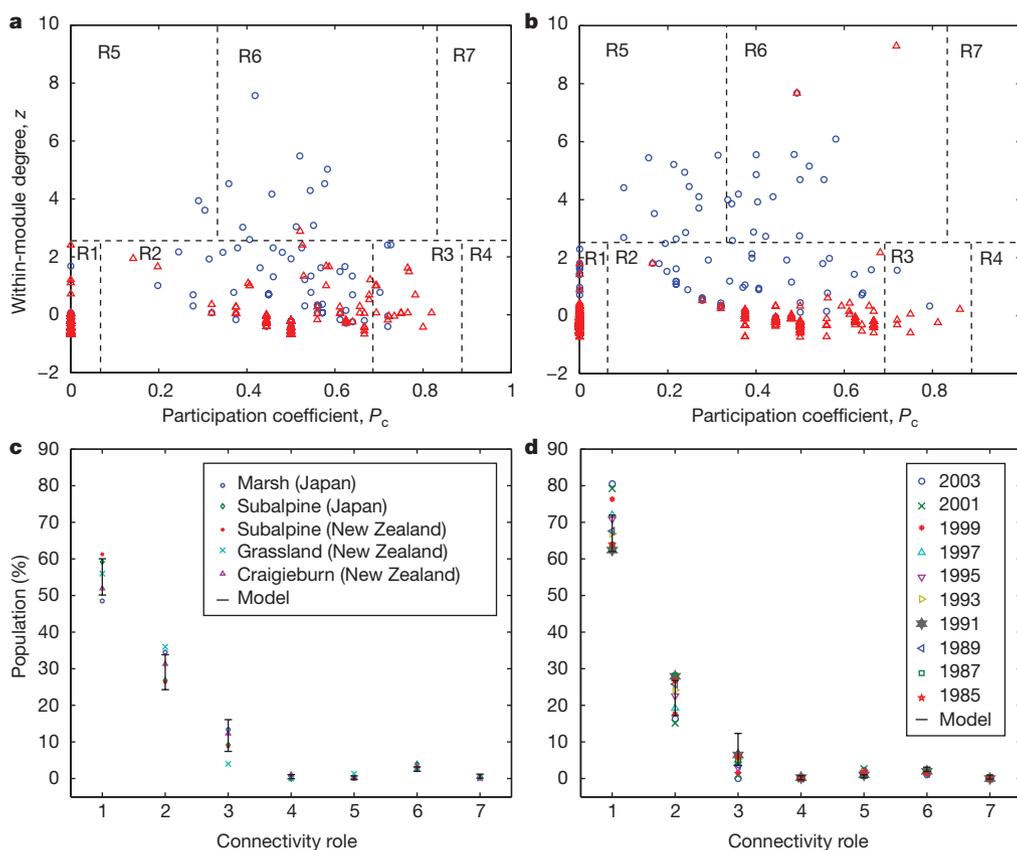
and members of class *A*. Note that pollination networks (filled symbols) and organizational networks (open symbols) exhibit the same patterns in both degree distributions. The solid line corresponds to the model-generated degree distributions (see also Supplementary Fig. 2). Table 1 provides statistical evidence for the correspondence between the empirical and model-generated distributions for both ecological and organizational networks.

To calculate the nestedness level,  $N$ , of the empirical and model-generated networks, we use the BINMATNEST program<sup>28</sup>. Here nestedness is defined in the interval  $[0, 1]$ , where 1 corresponds to a perfectly nested network<sup>9</sup>. Figure 2 shows the nestedness values for empirical networks (dashed line), random assemblages generated following the null model II proposed in ref. 9 (green crosses) and model-generated networks (orange circles). Note that the model always significantly outperforms random assemblages. Table 1 presents tests of statistical significance and shows that there is a high level of correspondence between the model and data for most of the networks (Supplementary Fig. 3).

Modularity values,  $Q$ , for the networks are calculated using the one-mode optimization algorithm proposed in ref. 30, because we want to extract only cooperative units from the network. This algorithm has shown that only five of the observed ecological networks are truly modular<sup>10</sup> (Table 1). For these five networks, we find a strong correspondence between the empirical and model-generated modularity

values (Table 1). For each organizational network, we find a modularity value that is higher than that generated from the corresponding random assemblages<sup>30</sup> ( $P < 10^{-6}$ ), and this empirical behaviour is replicated by the model (Table 1). Nodes have been shown to have different connectivity roles (for example global or local hubs) depending on how they are embedded within their module and participate in other modules<sup>30</sup>. Here we also test the ability of our model to accurately reproduce the empirically observed number of nodes within each connectivity role<sup>30</sup> (Fig. 3). To do so, we measure the Pearson correlation coefficient,  $r$ , and the ratio of the connectivity role norms,  $d$ , for the observed and model-generated networks (Methods). We consistently find values aligned with the empirical measurements for the ecological and organizational networks ( $r = 0.98$ ,  $0.9 < d < 1.1$ ).

Our study identifies striking similarities in the general structural characteristics of networks that are formed as a result of cooperative mechanisms operating in radically different contexts, linking partners in ecological and socio-economic systems, respectively. This empirical finding motivates the proposed simple model for bipartite cooperation, which captures the most important generic features of mutualistic interaction patterns starting from a minimal set of input parameters. At the level of partner–partner interactions, equivalent behaviour in different systems appears to be driven by similar types of interaction constraints. These correspond to complementarity in traits or characteristics, a hierarchical organization limiting the range



**Figure 3 | Connectivity roles.** **a, b**, Division of connectivity roles (R1–R7) for members of the subalpine (Japan) network (**a**) and the 1997 NYGI network (**b**). Plants and manufacturers are indicated with blue circles and animals and contractors with red triangles. Nodes with a normalized within-module degree ( $z$  score)  $z \geq 2.5$  have been heuristically defined as module hubs, and nodes with  $z < 2.5$  as non-hub nodes<sup>30</sup>. In addition, hubs and non-hub nodes are classified according to a participation coefficient,  $P_c$ , which gives the level of interaction of a node with the rest of the modules. The complete classification is as follows. For non-hub nodes, R1 comprises nodes with all their links connected within their own module ( $P_c \leq 0.05$ ), R2 comprises nodes with most of their links connected within their own module ( $0.05 < P_c \leq 0.62$ ), R3 comprises nodes with many links connected to other

modules ( $0.62 < P_c \leq 0.80$ ) and R4 comprises nodes with their links homogeneously connected to all other modules ( $P_c > 0.80$ ). For hub-nodes, R5 comprises nodes with most of their links connected to their own module ( $P_c \leq 0.30$ ), R6 comprises nodes with most of their links connected to other modules ( $0.30 < P_c \leq 0.75$ ) and R7 comprises nodes with their links homogeneously connected to all other modules ( $P_c > 0.75$ ). **c, d**, Proportions of nodes from the population within each connectivity role for the modular pollination networks (**c**) and the NYGI networks (**d**). Points (various symbols) correspond to the empirical values and bars (mean  $\pm$  s.d.) correspond to the model-generated values. Note that the model accurately replicates the proportion of nodes within each connectivity role for both types of network.

of potential partners, and the environmental context. The success of this simple stochastic model in generating the overall structural characteristics of mutualistic networks makes it a suitable starting point for the development of more elaborate ecological models, with the aim of addressing the important question of reproducing the actual links observed in real mutualistic webs. Such an approach would require more comprehensive statistical comparisons across different network metrics using maximum-likelihood methods<sup>6</sup>. Beyond this ecological context, and the specific organizational network that has been considered, the generic nature of the bipartite cooperation model and its underlying assembly rules suggests that it should be relevant to many other cooperative networks in domains beyond those considered here.

## METHODS SUMMARY

**Model.** For the specialization rule, the number of links,  $l_p$ , for each node  $p \in P$  is defined by  $l_p = 1 + \text{Round}((L - |P|)t_{Rp}\lambda_p / \sum_j t_{Rj}\lambda_j)$ , where the reward trait  $t_{Rp}$  is uniformly drawn from  $[0, 1]$ , the external factor  $\lambda_p$  is randomly drawn from an exponential distribution,  $|\cdot|$  denotes set cardinality and  $\text{Round}(\cdot)$  is the nearest-integer function.

For the interaction rule, nodes from class  $P$  are sorted according to their reward trait  $t_{Rp}$  in ascending order. Nodes from class  $A$  are sorted in descending order according to their foraging traits  $t_{Fa}$ , which are uniformly drawn from  $[0, 1]$ . Starting from the first node,  $p_p$ , and continuing sequentially subject to  $t_{Rp_i} > \lambda_{l_{p_i}}$ , each link  $l_{p_i}$  is connected to the first node  $a' \in A'$ , where  $A'$  is the subset of nodes in  $A$  that have not already been linked to by another node  $p \neq p_i$ . Here  $\lambda_{l_{p_i}}$  is an external factor drawn randomly from the same exponential distribution as  $\lambda_{p_i}$ . If  $t_{Rp_i} \leq \lambda_{l_{p_i}}$  then the link is randomly connected to another node  $a'' \in A''$ , where  $A''$  is the subset of nodes in  $A$  that have been linked in a previous time step. If the supply of nodes in either  $A'$  or  $A''$  is exhausted before all  $l_{p_i}$  links have been allocated, then nodes in the other subset are linked to instead. The model is initialized by connecting the first node,  $p_p$ , to  $l_{p_i}$  nodes in  $A'$ .

The exponential distribution used for  $\lambda_p$  and  $\lambda_{l_p}$  is of the form  $p(x) = \beta \exp(-\beta x)$ , with  $\beta = |P|(|A| - 1)/(2(L - |P|)) - 1$ . If we generalize our formalism and use a beta distribution<sup>13</sup>, we do not find significantly different results (Supplementary Table 5).

**Ratio of connectivity role norms.** The ratio of the norms,  $d$ , is defined by  $d = |x|/|y|$ , where  $|x| = (\sum_{i=1}^m x_i^2)^{1/2}$ ,  $|y| = (\sum_{i=1}^m y_i^2)^{1/2}$ ,  $m$  is the number of connectivity roles and  $x_i$  and  $y_i$  are the numbers of nodes within each role  $i$  for the empirical and, respectively, model-generated networks. Ratios within (0.9, 1.1) are fractions with a norm comparable to the empirical data.

Received 5 June; accepted 10 October 2008.

Published online 3 December 2008.

- Williams, R. J. & Martinez, N. Simple rules yield complex food webs. *Nature* **404**, 180–183 (2000).
- Cattin, M. F., Bersier, L. F., Banasek-Richter, C., Baltensperger, R. & Gabriel, J. P. Phylogenetic constraints and adaptation explain food-web structure. *Nature* **427**, 835–839 (2004).
- Stouffer, D. B., Camacho, J., Guimerà, R., Ng, C. A. & Nunes Amaral, L. A. Quantitative patterns in the structure of model and empirical food webs. *Ecology* **86**, 1301–1311 (2005).
- Dunne, J. A., Williams, R. J., Martinez, N. D., Wood, R. A. & Erwin, D. H. Compilation and network analyses of Cambrian food webs. *PLoS Biol.* **6**, 693–708 (2008).
- Petchey, O. L., Beckerman, A. P., Riede, J. O. & Warren, P. H. Size, foraging, and food web structure. *Proc. Natl Acad. Sci. USA* **105**, 4191–4196 (2008).
- Allesina, S., Alonso, D. & Pascual, M. A general model for food web structure. *Science* **320**, 658–661 (2008).
- Bascompte, J. & Jordano, P. in *Ecological Networks: Linking Structure to Dynamics in Food Webs* (eds Pascual, M. & Dunne, J. A.) 143–159 (Oxford Univ. Press, 2006).
- Jordano, P., Bascompte, J. & Olesen, J. M. Invariant properties in coevolutionary networks of plant-animal interactions. *Ecol. Lett.* **6**, 69–81 (2003).

- Bascompte, J., Jordano, P., Melián, C. J. & Olesen, J. M. The nested assembly of plant-animal mutualistic networks. *Proc. Natl Acad. Sci. USA* **100**, 9383–9387 (2003).
- Olesen, J. M., Bascompte, J., Dupont, Y. L. & Jordano, P. The modularity of pollination networks. *Proc. Natl Acad. Sci. USA* **104**, 19891–19896 (2007).
- Sampson, R. J., Raudenbush, S. W. & Earls, F. Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science* **277**, 918–924 (1997).
- Ostrom, E., Burger, J., Field, C. B., Norgaard, R. B. & Policansky, D. Revisiting the commons: Local lessons, global challenges. *Science* **284**, 278–282 (1999).
- Hammerstein, P. (ed.) *Genetic and Cultural Evolution of Cooperation* (MIT Press, 2003).
- Ohtsuki, H., Hauert, C., Lieberman, E. & Nowak, M. A. A simple rule for the evolution of cooperation on graphs and social networks. *Nature* **441**, 502–505 (2006).
- Bronstein, J. L. The exploitation of mutualisms. *Ecol. Lett.* **4**, 277–287 (2001).
- Waser, N. M., Chittka, L., Price, M. V., Williams, N. M. & Ollerton, J. Generalization in pollination systems, and why it matters. *Ecology* **77**, 1043–1060 (1996).
- Noë, R. & Hammerstein, P. Biological markets: supply and demand determine the effect of partner choice in cooperation, mutualism and mating. *Behav. Ecol. Sociobiol.* **35**, 1–11 (1994).
- Olesen, J. M. & Jordano, P. Geographic patterns in plant-pollinator mutualistic networks. *Ecology* **83**, 2416–2424 (2002).
- Guimarães, P. R., Rico-Gray, V., Furtado dos Reis, S. & Thompson, J. N. Asymmetries in specialization in ant-plant mutualistic networks. *Proc. R. Soc. Lond. B* **273**, 2041–2047 (2006).
- Rezende, E. L., Lavabre, J. E., Guimarães, P. R., Jordano, P. & Bascompte, J. Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature* **448**, 925–928 (2007).
- Santamaría, L. & Rodríguez-Gironés, A. Linkage rules for plant-pollinator networks: Trait complementarity or exploitation barriers? *PLoS Biol.* **5**, 354–362 (2007).
- Guimarães, P. R. Jr et al. Building-up mechanisms determining the topology of mutualistic networks. *J. Theor. Biol.* **249**, 181–189 (2007).
- Uzzi, B. The sources and consequences of embeddedness for the economic performance of organizations: the network effect. *Am. Sociol. Rev.* **61**, 674–698 (1996).
- Carroll, G. R. & Hannan, M. T. *The Demography of Corporations and Industries* (Princeton Univ. Press, 2004).
- Podolny, J. M. *Status Signals: A Sociological Study of Market Competition* (Princeton Univ. Press, 2005).
- Amaral, L. A. N., Scala, A., Barthélémy, M. & Stanley, H. E. Classes of small-world networks. *Proc. Natl Acad. Sci. USA* **97**, 11149–11152 (2000).
- Moody, J. & White, D. R. Structural cohesion and embeddedness: A hierarchical concept of social groups. *Am. Sociol. Rev.* **68**, 103–127 (2003).
- Rodríguez-Gironés, M. A. & Santamaría, L. A new algorithm to calculate the nestedness temperature of presence-absence matrices. *J. Biogeography* **33**, 924–935 (2006).
- Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proc. Natl Acad. Sci. USA* **99**, 7821–7826 (2002).
- Guimerà, R. & Nunes Amaral, L. A. Functional cartography of complex metabolic networks. *Nature* **433**, 895–900 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Dunne, R. Guimerà, J. Kertész, M. Sales-Pardo, D. Stouffer and R. Williams for comments and suggestions. F.R.-T. acknowledges funding from the European Commission under the FP6 NEST Pathfinder Initiative 'Tackling Complexity in Science' (MMCOMNET project, contract no. 012999). S.S. held a Doctoral Research Studentship funded by MMCOMNET and CONACYT, and currently is supported by a Postdoctoral Fellowship at the Oxford University Corporate Reputation Centre in conjunction with the CABDyN Complexity Centre.

**Author Contributions** B.U. provided the NYGI data; F.R.-T. designed the research; S.S., F.R.-T. and B.U. analysed the data; S.S. ran the simulations; S.S. and F.R.-T. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to F.R.-T. ([felix.reed-tsochas@sbs.ox.ac.uk](mailto:felix.reed-tsochas@sbs.ox.ac.uk)).