Supplementary Information

A simple model of bipartite cooperation for ecological and organizational networks

Serguei Saavedra[1], Felix Reed-Tsochas[2], and Brian Uzzi[3,4]

[1]Department of Engineering Science and CABDyN Complexity Centre, Oxford University, UK [2]James Martin Institute and CABDyN Complexity Centre, Saïd Business School, Oxford University, UK [3]Kellogg School of Management and Northwestern Institute on Complex Systems, Northwestern University, USA [4]Haas School of Business, University of California at Berkeley, USA

Supplementary Text 1: Comparison of mutualistic models

The mixed model (Santamaría and Rodríguez-Gironés)

Santamaría & Rodríguez-Gironés[S1] have proposed a family of mutualistic models to test the importance of trait complementarities and exploitation barriers in generating the structure of real-world mutualistic networks. As they show, their best model (the mixed model) is able to reproduce the nested configuration of a significant number of mutualistic networks. This model forms links between a given number of plants P and animals A (where P and A are the input parameters) if an established condition is met. This condition corresponds to the match between trait variables uniformly distributed in the interval [0,1] (two for each plant and two for each animal) with some relaxation given by two additional variables uniformly distributed in the interval [0,0.25] (two for each plant and two for each animal), and finally a difference between two barrier variables uniformly distributed in the interval [0,1] (two for each plant and two for each animal). Thus, a total of 6 input variables are required for each plant and for each animal.

Table S1 shows that although the model is able to replicate the nestedness of some of the observed mutualistic networks to a good approximation, it does not successfully produce the degree distribution and modularity of such networks. We have run the mixed model to observe its overall goodness-of-fit for nestedness, degree distribution and modularity, and find that it is able to generate only 14% (4 out of 35) of the total number of observed metrics (see Table S1 and S3).

The differential limiting size model (Guimarães et al.)

Guimarães et al.[S2] propose a family of mutualistic models built on preferential attachment to show that broad-scale degree distributions similar to those observed in real mutualistic networks are generated when two conditions are met. These conditions are (1) a similar or different growing rate and (2) differences in set size (number of plant species/number of animal species). Following the proposed model that is best suited to reproducing the features of real networks (the differential limiting size model) we require the following input parameters: (i) The size of population A and population P. (ii) The initial number (*No*) of nodes in population A and population P to start the simulation with. They show that this number does not affect the results and we used *No*=10 as suggested. Just in the case of the Deciduous forest network, we used No=5 due to the small size of the network. (iii) The number of links (m) attached to newcomer nodes. We use m=Round(L/(P+A-2No)), where L is the total number of links in the network in order to match the connectance of the observed networks. (iv) The growing ratio (p=0.5), which determines the assigned population (A or P) of each newcomer node. (v) Each newcomer i attaches to an incumbent j from the opposite population (A or P) according to a probability $P(j)$ following preferential attachment.

Table S2 shows that although the model is in fact able to give good approximations to the degree distribution (specially for plants) for most of the observed mutualistic networks, it performs poorly when trying to reproduce the nestedness and modularity of such networks. We have run the differential limiting size model to observe its overall goodness-of-fit for nestedness, degree distribution and modularity, and find that it is able to generate only 25% (9 out of 35) of the total number of observed metrics (see Table S2 and S3).

Global model comparisons (overall goodness-of-fit)

We record the percentage from the total number of calculated metrics (degree distribution, nestedness and modularity) that are reproduced to a good or excellent approximation (Kolmogorov-Smirnov probability>0.05 or normalized errors<2 model s.d.) by the mutualistic models. For ecological networks, Table S3 shows that the mixed model has an overall goodness-of-fit of 14% (4 out of 35 variables), the differential limiting size model has an overall goodness-of-fit of 25% (9 out of 35 variables), and the bipartite cooperation model has an overall goodness-of-fit of 77% (27 out of 35 variables). For organizational networks, Table S4 shows that the mixed model has an overall goodness-of-fit of 0% (0 out of 16 variables), the differential limiting size model has an overall goodness-of-fit of 25% (4 out of 16 variables), and the bipartite cooperation model has an overall goodness-of-fit of 69% (11 out of 16 variables). It is important to note that the bipartite cooperation model does not only provide better overall goodness-of-fit, but also outperforms the two competing models on every individual network measure in both domains.

Supplementary Table S1: The mixed model

Dataset – Environment	*L*	*P*	*A*	$(KS_P\text{--}KS_A)$	*N*	*Q*
Marsh, Japan	430	64	187	0.000* -- 0.452††	0.976**(0.953)	0.551*(0.752)
Grassland, Cass, New Zealand	374	41	139	0.000* -- 0.000*	0.957**(0.932)	0.474*(0.664)
Subalpine forest/meadow, Japan	865	90	354	0.000* -- 0.001*	0.985*(0.966)	0.545*(0.683)
Subalpine, Arthur's, New Zealand	120	18	60	0.000* -- 0.000*	0.858††(0.860)	0.553*(0.228)
Subalpine, Craigieburn, New Zealand	346	49	118	0.000* -- 0.026**	0.961†(0.933)	0.480*(0.696)
Tundra, Canada	179	29	81	0.000* -- 0.000*	0.971**(0.880)	nm
Scrub/snow gum forest, Australia	252	36	81	0.000* -- 0.000*	0.935*(0.870)	nm
Deciduous forest, USA	65	7	33	0.000* -- 0.001*	0.953*(0.836)	nm
Arctic tundra, Greenland	453	31	75	0.000* -- 0.005*	0.793**(0.881)	nm
Subarctic rock slope, Sweden	242	24	118	0.000* -- 0.022**	0.927††(0.918)	nm
NYGI 1985	7250	823	2562	0.000* -- 0.000*	0.997*(0.986)	0.598**(0.630)
NYGI 1991	3981	325	1590	0.000* -- 0.000*	0.994*(0.982)	0.601*(0.689)
NYGI 1997	1450	148	700	0.000* -- 0.001*	0.990*(0.976)	0.653**(0.706)
NYGI 2003	228	62	128	0.000* -- 0.000*	0.976**(0.948)	0.711*(0.601)

Table S1. Empirical values and statistical significance for the mixed model[S1]. For each pollination dataset and four organizational networks used in this paper, the table presents its environment/location; total number of links *L*., *P* and *A* are the number of nodes in class P and class A respectively. For the degree distributions, (KS_P-KS_A) shows the combined Kolmogorov-Smirnov (KS) probability using the two-group equivalence KS test between the empirical and model-generated distributions for class P and class A respectively. *N* and *Q* correspond to the observed nestedness and mean modularity values respectively, along with the normalized errors (z-scores) for the comparison between the empirical and model-generated values. The model-generated mean values for *N* and *Q* are shown inside the parentheses. Five of the observed pollination networks have already been found to be non-modular (nm)[S3]. All comparisons are based on 1000 model simulations. Note that the model reproduces <14% of the overall number of observed metrics with a good or excellent fit (4 out of 35 and 0 out of 16 for the ecological and organizational networks respectively).

††: KS≥0.30, normalized errors <1 model s.d. (excellent fit). †: KS<0.30, normalized errors between 1 and 2 model s.d. (good fit). **: KS<0.05, normalized errors between 2 and 3 model s.d. (poor fit). *: KS<0.01, normalized errors >3 model s.d. (bad fit).

Supplementary Table S2: The differential limiting size model

Dataset – Environment	L	P	A	$(KS_P$--$KS_A)$	N	Q
Marsh, Japan	430	64	187	0.000* -- 0.001*	0.976††(0.974)	0.551*(0.722)
Grassland, Cass, New Zealand	374	41	139	0.344†† -- 0.000*	0.957*(0.865)	0.474*(0.504)
Subalpine forest/meadow, Japan	865	90	354	0.000* -- 0.000*	0.985*(0.933)	0.545†(0.533)
Subalpine, Arthur's, New Zealand	120	18	60	0.809†† -- 0.000*	0.858*(0.775)	0.553*(0.481)
Subalpine, Craigieburn, New Zealand	346	49	118	0.854†† -- 0.002*	0.961*(0.880)	0.480*(0.518)
Tundra, Canada	179	29	81	0.53†† -- 0.005*	0.971*(0.932)	nm
Scrub/snow gum forest, Australia	252	36	81	0.121† -- 0.000*	0.935*(0.841)	nm
Deciduous forest, USA	65	7	33	0.216† -- 0.001*	0.953*(0.729)	nm
Arctic tundra, Greenland	453	31	75	0.99†† -- 0.001*	0.793*(0.639)	nm
Subarctic rock slope, Sweden	242	24	118	0.016** -- 0.000*	0.927*(0.941)	nm
NYGI 1985	7250	823	2562	0.000* -- 0.000*	0.997*(0.992)	0.598*(0.510)
NYGI 1991	3981	325	1590	0.000* -- 0.000*	0.994*(0.974)	0.601*(0.503)
NYGI 1997	1450	148	700	0.000* -- 0.000*	0.990**(0.981)	0.653*(0.834)
NYGI 2003	228	62	128	0.101† -- 0.352††	0.976†(0.971)	0.711††(0.725)

Table S2. Empirical values and statistical significance for the differential limiting size model[S2]. For each pollination dataset and four organizational networks used in this paper, the table presents its environment/location; total number of links L., P and A are the number of nodes in class P and class A respectively. For the degree distributions, (KS_P-KS_A) shows the combined Kolmogorov-Smirnov (KS) probability using the two-group equivalence KS test between the empirical and model-generated distributions for class P and class A respectively. N and Q correspond to the observed nestedness and mean modularity values respectively, along with the normalized errors (z-scores) for the comparison between the empirical and model-generated values. The model-generated mean values for N and Q are shown inside the parentheses. Five of the observed pollination networks have already been found to be non-modular (nm)[S3]. All comparisons are based on 1000 model simulations. Note that the model reproduces 25% of the overall number of observed metrics with a good or excellent fit (9 out of 35 and 4 out of 16 for the ecological and organizational networks respectively).

††: $KS \geq 0.30$, normalized errors <1 model s.d. (excellent fit). †: $KS<0.30$, normalized errors between 1 and 2 model s.d. (good fit). **: $KS<0.05$, normalized errors between 2 and 3 model s.d. (poor fit). *: $KS<0.01$, normalized errors >3 model s.d. (bad fit).

Supplementary Table S3: Global model comparisons for ecological networks

Network Properties	#Calculated variables	#Reproduced variables (BCM)	#Reproduced variables (MM)	#Reproduced variables (DLSM)
Degree distribution	20	16	1	7
Nestedness	10	7	3	1
Modularity	5	4	0	1
Total	35	27	4	9
Overall goodness-of-fit		**77%**	**14%**	**25%**

Table S3. Global model comparisons for ecological networks. The table shows for each of the observed network properties (degree distribution, nestedness and modularity) the number of calculated variables, and the number of those variables reproduced to a good or excellent approximation (Kolmogorov-Smirnov probability>0.05 or normalized errors<2 model s.d., see Table 1,S1-S2) following the bipartite cooperation model (BCM), the mixed model[S1] (MM) and the differential limiting size model[S2]. Overall goodness-of-fit corresponds to the ratio between the total number of reproduced metrics and the total number of calculated metrics.

Supplementary Table S4: Global model comparisons for organizational networks

Network Properties	#Calculated variables	#Reproduced variables (BCM)	#Reproduced variables (MM)	#Reproduced variables (DLSM)
Degree distribution	8	6	0	2
Nestedness	4	4	0	1
Modularity	4	1	0	1
Total	16	11	0	4
Overall goodness-of-fit		**69%**	**0%**	**25%**

Table S4. Global model comparisons for organizational networks. The table shows for each of the observed network properties (degree distribution, nestedness and modularity) the number of calculated variables, and the number of those variables reproduced to a good or excellent approximation (Kolmogorov-Smirnov probability>0.05 or normalized errors<2 model s.d., see Table 1,S1-S2) following the bipartite cooperation model (BCM), the mixed model[S1] (MM) and the differential limiting size model[S2]. Overall goodness-of-fit corresponds to the ratio between the total number of reproduced metrics and the total number of calculated metrics.

Supplementary Table S5: Empirical values and statistical significance for the bipartite cooperation model using a beta distribution

Dataset – Environment	*L*	*P*	*A*	$(KS_P\text{--}KS_A)$	*N*	*Q*
Marsh, Japan	430	64	187	0.331†† -- 0.305††	0.976†† (0.970)	0.551†† (0.558)
Grassland, Cass, New Zealand	374	41	139	0.850†† -- 0.324††	0.957†† (0.955)	0.474† (0.456)
Subalpine forest/meadow, Japan	865	90	354	0.682†† -- 0.001*	0.985** (0.977)	0.545** (0.525)
Subalpine, Arthur's, New Zealand	120	18	60	0.200† -- 0.998††	0.858* (0.943)	0.553** (0.525)
Subalpine, Craigieburn, New Zealand	346	49	118	0.043** -- 0.000*	0.961†† (0.951)	0.480† (0.471)
Tundra, Canada	179	29	81	0.066† -- 0.992††	0.971† (0.948)	nm
Scrub/snow gum forest, Australia	252	36	81	0.793†† -- 0.056†	0.935†† (0.944)	nm
Deciduous forest, USA	65	7	33	0.901†† -- 0.622††	0.953† (0.933)	nm
Arctic tundra, Greenland	453	31	75	0.033** -- 0.106†	0.793** (0.894)	nm
Subarctic rock slope, Sweden	242	24	118	0.338† -- 0.004**	0.927† (0.949)	nm
NYGI 1985	7250	823	2562	0.053† -- 0.099†	0.997† (0.996)	0.598* (0.491)
NYGI 1991	3981	325	1590	0.101† -- 0.430††	0.994†† (0.993)	0.601* (0.515)
NYGI 1997	1450	148	700	0.009** -- 0.213†	0.990† (0.987)	0.653** (0.602)
NYGI 2003	228	62	128	0.311† -- 0.002**	0.976** (0.963)	0.711† (0.698)

Table S5. Empirical values and statistical significance for the bipartite cooperation model using a beta distribution. For each pollination dataset and four organizational networks used in this paper, the table presents its environment/location; total number of links *L*., *P* and *A* are the number of nodes in class P and class A respectively. For the degree distributions, (KS_P-KS_A) shows the combined Kolmogorov-Smirnov (KS) probability using the two-group equivalence KS test between the empirical and model-generated distributions for class P and class A respectively. *N* and *Q* correspond to the observed nestedness and mean modularity values respectively, along with the normalized errors (z-scores) for the comparison between the empirical and model-generated values. The model-generated mean values for *N* and *Q* are shown inside the parentheses. Five of the observed pollination networks have already been found to be non-modular (nm)[S3]. All comparisons are based on 1000 model simulations using the beta distribution proposed by Williams & Martinez[S4], defined by $p(x) = \beta \cdot (1-x)^{(\beta-1)}$ with $\beta = \frac{P \cdot (A-1)}{2 \cdot (L-P)} - 1$. Note that the model reproduces more than 70% of the overall number of observed metrics with a good or excellent fit (25 out of 35 and 10 out of 16 for the ecological and organizational networks respectively).

††: KS≥0.30, normalized errors <1 model s.d. (excellent fit). †: KS<0.30, normalized errors between 1 and 2 model s.d. (good fit). **: KS<0.05, normalized errors between 2 and 3 model s.d. (poor fit). *: KS<0.01, normalized errors >3 model s.d. (bad fit).

Supplementary Text 2: Pollination-network datasets

We studied ten of the largest pollination-network datasets compiled in the literature. For further information, please refer to Bascompte et al.[S5]:

1. Arctic tundra, Greenland: Elberling H. & Olesen J.M., unpublished data.

2. Subarctic rock slope, Sweden: Elberling H. & Olesen J.M. (1999). The structure of a high latitude plantflower visitor system: the dominance of flies. *Ecography* 22, 314-323.

3. Tundra, Canada: Hocking B. (1968). Insect flower associations in the high Arctic with special reference to nectar. *Oikos* **19**, 359-387

4. Scrub/snow gum forest, Australia: Inouye, D. W. & Pyke, G. H. (1988). Pollination biology in the Snowy Mountains of Australia: comparisons with montane Colorado. *Aust. J. Ecol.* **13**,191-210.

5. Marsh, Japan: Kato M. & Miura R. (1996). Flowering phenology and anthophilous insect community at a threatened natural lowland marsh at Nakaikemi in Tsuruga, Japan. *Contrib. Biol. Lab., Kyoto Univ.* **29**, 1-48.

6. Subalpine forest/meadow, Japan: Kato M., Matsumoto M. & Kato T. (1993). Flowering phenology and anthophilous insect community in the cool-temperate subalpine forests and meadows at Mt. Kushigata in the Central part of Japan. *Contrib. Biol. Lab., Kyoto Univ.* **28**, 119-172.

7. Subalpine, Arthur's Pass, New Zealand: Primack R. B. (1983). Insect pollination in the New Zealand mountain flora. *New Zealand J. Bot.* **21**, 317-333.

8. Grassland, Cass, New Zealand: Primack R.B. (1983). Insect pollination in the New Zealand mountain flora. *New Zealand J. Bot.* **21**, 317-333.

9. Subalpine, Craigieburn, New Zealand: Primack R.B. (1983). Insect pollination in the New Zealand mountain flora. *New Zealand J. Bot.* **21**, 317-333.

10. Deciduous forest, USA: Schemske, D., Willson, M.F., Melampy, M., Miller, L., Verner, L., Schemske, K. & Best, L. (1978). Flowering ecology of some spring woodland herbs. *Ecology* **59**, 351-366.

Supplementary Text 3: The New York garment industry network

This is extensive empirical dataset on the manufacturer-contractor network of the New York City garment industry from 1985-2003. Our data include approximately 700,000 transactions from January 1985 to December 2003. Here, a link connecting two firms is formed if there is a trading transaction between these two for a particular year. This generates yearly snapshots of a bipartite network with firms having one or multiple manufacturers or contractors. In this paper, we focus on the properties of the largest connected component which contains ~95% of the population. From the total number of firms in the NYGI, approximately 30% are manufacturers, 68% are contractors and only 2% are hybrid firms (both activities). Hence, due to the low number of hybrids, we only take into account manufacturers and contractors to build up a bipartite network for each of the 19 years. This unique dataset has been collected and made available by UNITE (the Union of Needle Trades and Industrial and Textile Employees), which has organized ~90% of the firms in the NYGI and has developed a highly reliable record system[S6].

From 1985 to 2003 the NYGI network experienced a population decrease of an order of magnitude. However, each type of firm followed a different decline. Figure S1 shows the population contraction normalized for each type of firm. This shows that contractors may follow a linear contraction whereas manufacturers and hybrid firms exhibit a more rapid decline. Nevertheless, we find that although the number of firms in the network varies from year to year, the relation between the total number of firms and links follows a constant relation. The inset in Figure S1 shows that this relation is defined by a power-law with γ=1.22±0.01 (R^2=0.98). Note that Bascompte et al.[S5] have found a very similar relation for mutualistic networks between the total number of links and species defined by a power law with exponent γ=1.13. This suggests that in common with species in ecological systems, firms in the NYGI exhibit a surprising stable relation over the years between the total number of firms and links in the network. Although there is considerable debate about the universal properties of food webs, it is accepted that the number of species and links are two parameters that play a key role in determining the topological properties of ecological networks[S7].

Supplementary Figure S1: Nonlinear decline

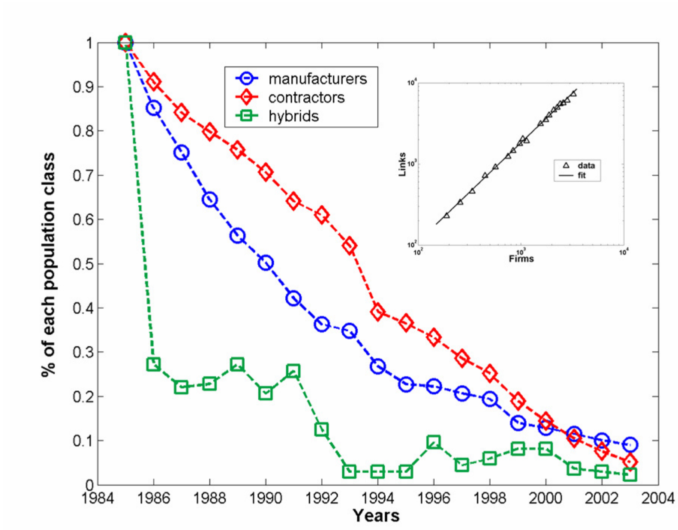

Figure S1. Nonlinear decline. This shows the different declining trajectories over the years for manufacturers (blue circles), contractors (red diamonds), and hybrids (green squares). The number of firms is normalized to its corresponding value in 1985. The inset shows the relationship in a log-log scale between the total number of firms and links in the network. The solid line is the fit to the data defined by $\gamma=1.22\pm0.01$ ($R^2=0.98$).

Supplementary Figure S2: Degree distribution

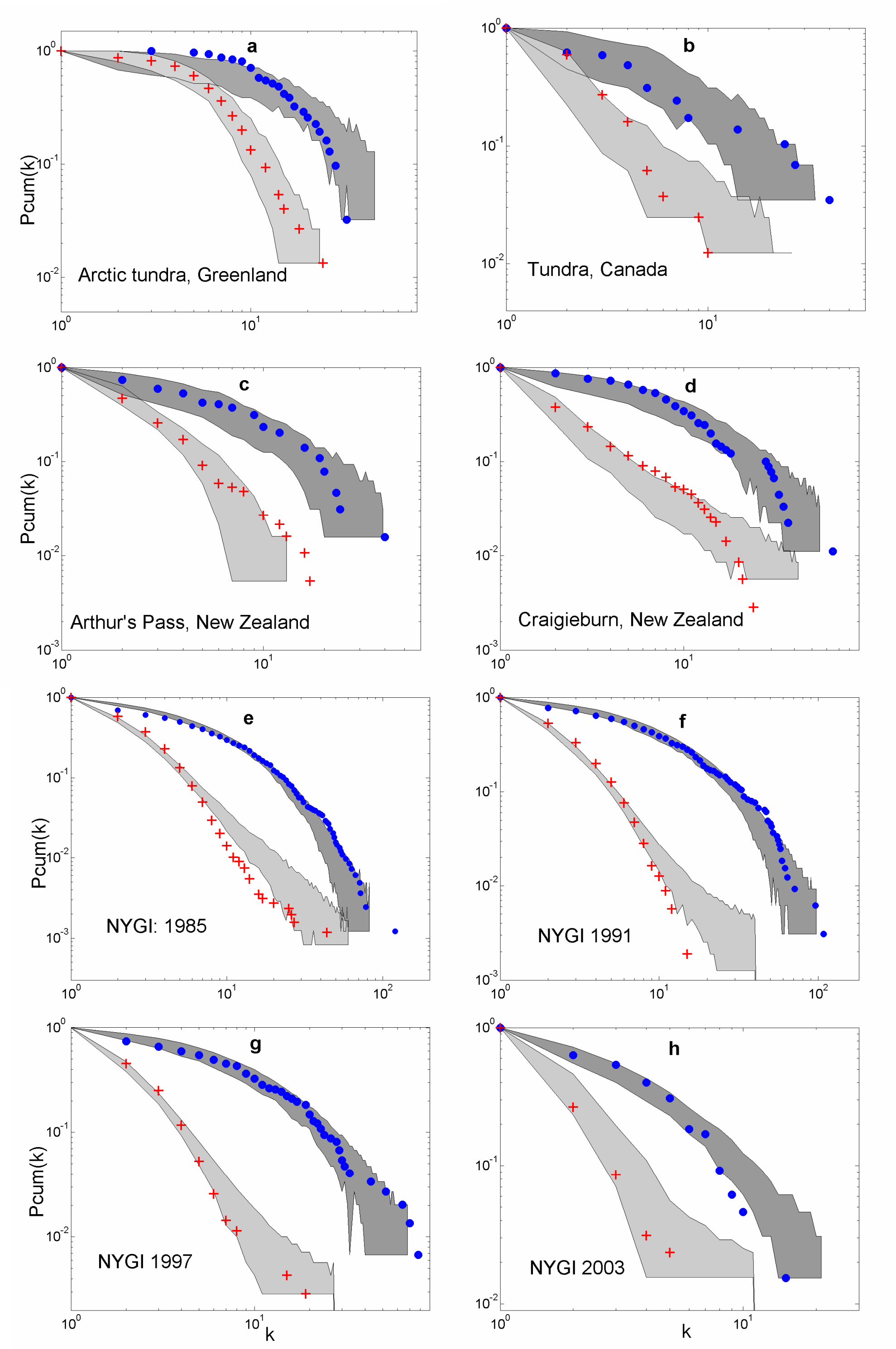

Figure S2. Degree distribution. Figures **a-d** and **e-h** shows the cumulative degree distribution *Pcum*(*k*), a widely used statistical metric that measures the probability that a node has up to *k* network connections, for four pollination networks and four organizational networks respectively. Blue dots correspond to the degree distribution for plants/manufacturers and red crosses to the degree distribution for animals/contractors. The distributions are plotted on a log-log scale. The gray region corresponds to the 95% confidence over 1000 simulations using the model-generated degree distributions.

Supplementary Figure S3: Model-generated matrices

a

Animals

Plants

100000000000000000000000000000000000000000000000000000000000
100000000000000000000000000000000000000000000000000000000000
111000000000000000000000000000000000000000000000000000000000
100100000000000000000000000000000000000000000000000000000000
111110000000000000000000000000000000000000000000000000000000
001000000000000000000000000000000000000000000000000000000000
111111000000000000000000000000000000000000000000000000000000
101000000000000000000000000000000000000000000000000000000000
111000100000000000000000000000000000000000000000000000000000
010010011110000000000000000000000000000000000000000000000000
111000100001111111100000000000000000000000000000000000000000
111000000100000100011111100000000000000000000000000000000000
000000000000000000000000011000000000000000000000000000000000
100100000000000000000000000111111000000000000000000000000000
000010000000000000000000000000000110000000000000000000000000
010010000000000000000000000000000001111111111110000000000000
000010000100000000000000000000000110000000000001111111100000
111110000010000101100000001000000100000000010000010000111111

b

Animals

Plants

100000000000000000000000000000000
111110000000000000000000000000000
111111000000000000000000000000000
100110111000000000000000000000000
100000001100000000000000000000000
100010000001111111111111100000000
111110011110101010110010011111111

Figure S3. Model-generated matrices. Figures **a** and **b** show the adjacency bipartite matrices generated following the bipartite cooperation model for the pollination networks Subalpine Arthur's, New Zealand and Deciduous forest, USA respectively. Rows correspond to plants and columns to animals. Plants are sorted in ascending order according to their reward trait t_{Rp}, and animals are sorted in descending order according to their foraging trait t_{Fa}. A value of 1 corresponds to an interaction between a plant and an animal, otherwise there is no interaction. Note that the model's algorithm produces highly nested configurations, which account for a great part of the goodness of fit to the data.

Supplementary References

S1. Santamaría, L. & Rodríguez-Gironés, A. Linkage rules for plant-pollinator networks: Trait complementarity or exploitation barriers? *PLoS Biology* **5**, 354-362 (2007).

S2. Guimarães, P. R. et al. Building-up mechanisms determining the topology of mutualistic networks. *J. Theor. Biol.* **249**, 181-189 (2007).

S3. Olesen, J. M., Bascompte J., Dupont, Y. L. & Jordano, P. The modularity of pollination networks. *Proc. Natl. Acad. Sci. USA* **104**, 19891-19896 (2007).

S4. Williams, R. J. & Martinez, N. Simple rules yield complex food webs. *Nature* **404**, 180-183 (2000).

S5. Bascompte J., Jordano, P., Melián C. J. & Olesen, J. M. The nested assembly of plant-animal mutualistic networks. *Proc. Natl. Acad. Sci. USA* **100**, 16 (2003).

S6. Uzzi, B. The sources and consequences of embeddedness for the economic performance of organizations: the network effect. *Amer. Sociological Rev.* **61**, 674-698 (1996).

S7. Dunne, J. A., Williams, R. J. & Martinez, N. D. Food-web structure and network theory: The role of connectance and size. *Proc. Natl. Acad. Sci.* **99**, 12917-12922 (2002).