Effects of sampling completeness on the structure of plant–pollinator networks

A. Rivera-Hutinel,[1] R. O. Bustamante,[2,3] V. H. Marín,[2] and R. Medel[2,4]

[1]*Instituto de Entomología, Facultad de Ciencias Básicas, UMCE, José Pedro Alessandri 774, Santiago, Chile*
[2]*Departamento de Ciencias Ecológicas, Facultad de Ciencias, Universidad de Chile, Las Palmeras 3425, Casilla 653, Santiago, Chile*
[3]*Institute of Ecology and Biodiversity, Facultad de Ciencias, Universidad de Chile, Las Palmeras 3425, Casilla 653, Santiago, Chile*

Abstract. Plant–animal interaction networks provide important information on community organization. One of the most critical assumptions of network analysis is that the observed interaction patterns constitute an adequate sample of the set of interactions present in plant–animal communities. In spite of its importance, few studies have evaluated this assumption, and in consequence, there is no consensus on the sensitivity of network metrics to sampling methodological shortcomings. In this study we examined how variation in sampling completeness influences the estimation of six network metrics frequently used in the literature (connectance, nestedness, modularity, robustness to species loss, path length, and centralization). We analyzed data of 186 flowering plants and 336 pollinator species in 10 networks from a forest-fragmented system in central Chile. Using species-based accumulation curves, we estimated the deviation of network metrics in undersampled communities with respect to exhaustively sampled communities and the effect of network size and sampling evenness on network metrics. Our results indicate that: (1) most metrics were affected by sampling completeness but differed in their sensitivity to sampling effort; (2) nestedness, modularity, and robustness to species loss were less influenced by insufficient sampling than connectance, path length, and centralization; (3) robustness was mildly influenced by sampling evenness. These results caution studies that summarize information from databases with high, or unknown, heterogeneity in sampling effort per species and should stimulate researchers to report sampling intensity to standardize its effects in the search for broad patterns in plant–pollinator networks.

Key words: accumulation curves; Clench model; ecological networks; Los Ruiles National Reserve, Chile; network size; plant–pollinator network metrics; sampling completeness; sampling effort; sampling evenness.

INTRODUCTION

The analysis of plant–animal interaction networks has received considerable attention in recent years, in part because it represents a tractable way to inquire into the complex nature of community organization. By covering such diverse topics as the relationship between biodiversity and stability (e.g., Memmott et al. 2004, Bascompte et al. 2006), ecological generalization (e.g., Bascompte et al. 2003, Jordano et al. 2003), interaction syndromes (e.g., Olesen et al. 2007), assembly rules (e.g., Guimarães et al. 2007, Santamaria and Rodríguez-Gironés 2007), conservation (e.g., Fortuna and Bascompte 2006, Tylianakis et al. 2010, Gonzalez et al. 2011), and phylogenetic structure (e.g., Rezende et al. 2007), among others, the analysis of plant–animal interaction networks has provided new insights in ecology, evolution, and conservation of communities (see review in Bascompte and Jordano 2007). Networks

are often depicted in a matrix where columns and rows represent animal and plant species, respectively, and the elements are the set of the realized interactions between the two groups. Among the most commonly used metrics to describe the structure of plant–animal networks are connectance, nestedness, modularity, and robustness (reviewed in Dormann et al. 2009). Examination of metrics, and therefore of network structure, in different ecological systems and under different ecological scenarios has produced a significant advance in the understanding of general properties of plant–pollinator networks. Nevertheless, an important shortcoming of these metrics is their sensitivity to network size, that is, to the total number of interacting species (animals plus plants). This scale sensitivity implies that some network properties do not necessarily remain constant with changes in the number of species belonging to a network (Bersier et al. 1999, Banasek-Richter et al. 2004). For example, it has been described that, although modularity and robustness are largely insensitive to network size (e.g., Dunne et al. 2002, Olesen et al. 2007), an increase in species richness tends to decrease connectance and to increase nestedness (e.g., Jordano 1987, Bascompte et al.

2003). In consequence, direct comparisons among communities need to be interpreted with caution, especially when network metrics are presented in unstandardized form and network size effects are not properly controlled (Jordano 1987, Bascompte et al. 2003, Olesen et al. 2006, 2007, Almeida-Neto et al. 2007). A second methodological shortcoming relates to the sensitivity of network metrics to sampling intensity (Blüthgen et al. 2006, Vázquez and Aizen 2006, Nielsen and Bascompte 2007). Even though sampling effects are not only circumscribed to network analyses, but also pervade most empirical studies in ecology, conclusions of the few studies that have addressed this issue suggest that sampling intensity does not affect all network metrics in the same way. For example, Nielsen and Bascompte (2007) analyzed the sensitivity of connectance and nestedness metrics to variation in sampling effort, concluding that nestedness was less prone to sampling bias compared to the number of species and links in the network. Similarly, Petanidou et al. (2008) showed that estimates of the degree of specialization of animal and plant species depend strongly on sampling effort, and tend to be inflated in studies with limited sampling.

There are two broad approaches to examine the effect of sampling bias on network structure. First, some studies have focused on the analysis of fieldwork sampling techniques and the extent to which conclusions are influenced by the way samples are collected. For example, Gibson et al. (2010) analyzed the potential bias in network metric estimates when using time-based observations or transects. Although connectance and nestedness did not differ between the two sampling procedures, time-based observations tend to record a higher number of unique interactions and lower web asymmetries than transect samplings. In the same vein, Bosch et al. (2009) estimated network metrics from both a phyto- and a zoocentric perspective, concluding that analyses that rely on pollination visitation patterns to record interactions often tend to underestimate connectance and nestedness, compared to analyses that identify interactions according to the pollen load transported by pollinators. Second, data manipulation after sampling permits one to evaluate whether network statistics are susceptible to bias because of limited sampling or confounding variables. In this regard, two categories of analysis can be identified in the literature: rarefaction analysis and accumulation curve analysis.

Rarefaction analysis consists of the random removal of an increasing number of interactions in the network, simulating a reduction in the number of interactions recorded (e.g., Jordano 1987, Banasek-Richter et al. 2004, Blüthgen et al. 2006, Vázquez and Aizen 2006). This is the only available method when researchers have no direct access to the original data set. The procedure has the disadvantage that it does not provide information on observations that do not record any visit to a plant, which we call failed sampling units (FSU,

hereafter). Because by definition FSU are not incorporated into ecological network analysis (because only interactions are recorded), the network contains necessarily a fraction of the information collected. Because FSU values provide valuable information on the time spent by researchers to record interactions (Fig. 1), it is necessary to include such events in estimates of potential bias attributable to limited sampling effort.

Accumulation curve analysis consists of the recording of new species (or interactions) for a community as a function of increased sampling effort (e.g., Nielsen and Bascompte 2007, Jordano et al. 2009, Chacoff et al. 2011). Unlike rarefaction analysis, this approach requires direct access to the original data set. In the context of network analysis, species accumulation curves (or sample-based rarefaction curves; sensu Colwell et al. 2004) permit estimation of the expected number of species or interactions as a function of the number of observations (see Fig. 1c). When the accumulation curve reaches an asymptote, it means that increasing sampling effort will not increase the number of species or interactions recorded (Colwell et al. 2004). To evaluate sampling completeness, it is possible to calculate how far the observed curve is from the expected asymptotic value. Asymptote estimation may be performed by fitting the observed accumulation curve to an asymptotic parametric model (Soberón and Llorente 1993), or through a nonparametric estimator (e.g., Chao 2; see Chacoff et al. 2011).

Accumulation curves may be useful descriptors of sampling completeness for both species and interaction levels. One shortcoming of accumulation curves based on interactions, however, is that the number of interactions can reach an asymptote even though some plant species have been undersampled for pollinator species. This phenomenon may occur when the sampling effort is highly heterogeneous (i.e., low sampling evenness). Poorly sampled plant species would thus be associated with one or a few pollinator species, overestimating the presence of specialists, inflating some network metrics such as nestedness, and deflating others such as connectance (Vázquez and Aizen 2006, Blüthgen et al. 2008). This inconvenience may be overcome by using accumulation curves on a per plant species basis to evaluate the sampling completeness of the pollinator assemblage visiting each plant species (e.g., Chacoff et al. 2011). This analysis has some advantages: (1) it is commensurate to the scale where observations are performed (i.e., the plant species); (2) it permits quantification of the heterogeneity in sampling effort among plant species; and (3) it permits assessment of levels of generalization–specialization among plant species in relation to interspecific differences in visitation rates (e.g., Herrera 2005). In this study we use species-based accumulation curves to examine the sensitivity of six widely used network metrics to variation in sampling completeness per plant species in 10 plant–pollinator networks. More specifically, we attempt to answer the
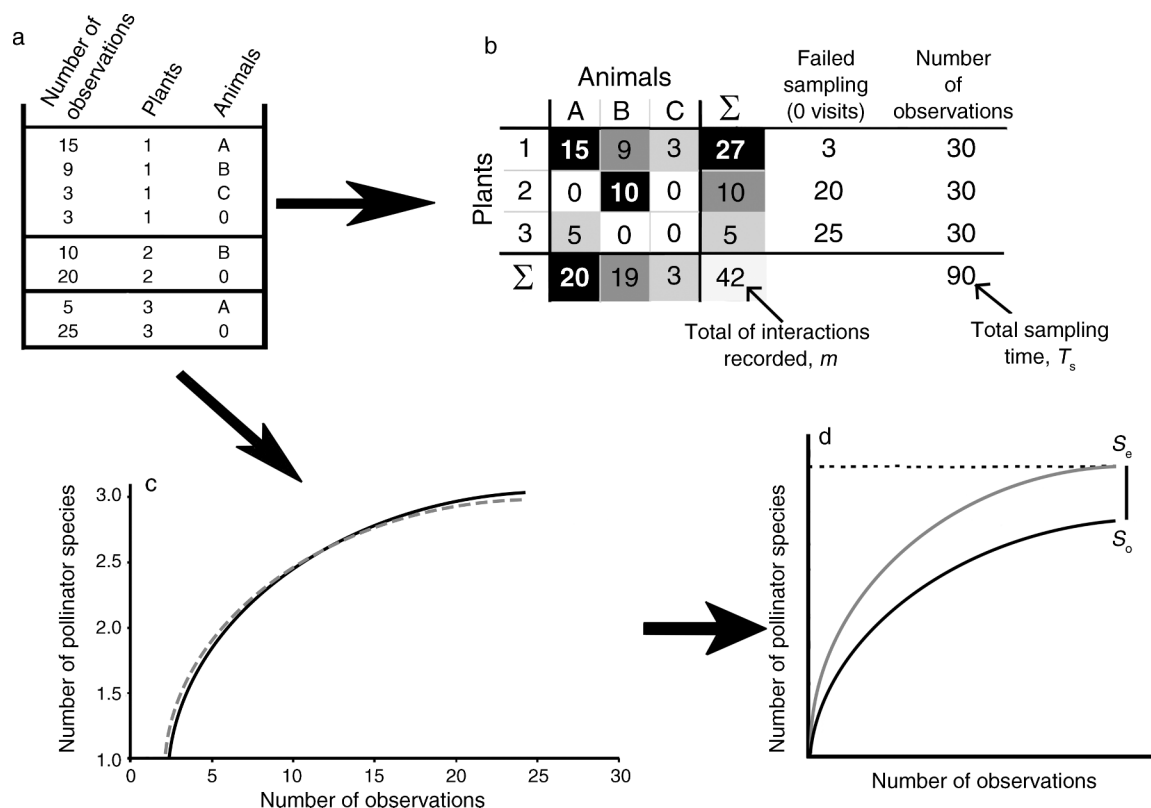
Fig. 1. Hypothetical example representing the overall procedure used in the study of plant–animal networks, and the methodology used in this study. (a) Observed interactions between plant species (1, 2, 3) and animal species (A, B, C). The number of time-based recordings for each interaction, including the null interactions (i.e., observations with no pollinator detected, 0's), is depicted in the first column. (b) Interaction matrix using the recordings in panel (a): the total numbers (Σ) of interactions for every plant and animal species, and the total number of interactions recorded (m). Because null interactions (failed sampling; i.e., no visits observed) are not included in the interaction matrix, the total sampling time (T_s) is not congruent with the number of interactions recorded (m). (c) Species accumulation curve for plant species 1. The curve shows the number of pollinator species observed (S_o) at increasing sampling effort. If the curve reaches an asymptote, it implies that the sampling effort is adequate to record the whole pollinator assemblage. The Clench asymptotic model is shown by the dashed line (see *Analyses of sampling effort* for details). (d) By adjusting the accumulation curve (S_o, lower curve) to the asymptotic models, it is possible to estimate the maximum number of pollinators expected (S_e, upper curve) and sampling completeness, that is, the fraction of the expected pollinator assemblage that is recorded (S_o/S_e).

following questions. (1) Does sampling completeness influence network metric estimates? (2) Which metrics are more robust to variation in sampling completeness? (3) To what extent is the effect of sampling completeness on network metrics influenced by the number of species in the network and by sampling evenness?

MATERIALS AND METHODS

Fieldwork and sampling procedure

This study was conducted in an area near Los Ruiles National Reserve (35°50′ S, 72°30′ W), a reserve aimed to protect a rare and endemic deciduous forest in the coastal range of central Chile. The Ruil forest is a highly fragmented ecosystem distributed along 100 km of the coastal range in central Chile (35–36° S latitude) (Bustamante and Castor 1998). Currently, the landscape is a mosaic in which the original forest has been almost completely substituted by plantations of *Pinus radiata*

(Monterrey pine), an exotic tree originally from California. The dominant species are the caducifolious broad-leaved trees *Nothofagus alessandrii* and *N. glauca* living with other subdominant evergreen trees such as *Cryptocarya alba*, *Aetoxicum punctatum*, and *Gevuina avellana*, among others (San Martin et al. 1984). Our study focused on 10 fragments of native forest ranging from 5.6 to 49.3 ha. Sampling of species and interactions was performed during the austral spring and summer seasons, when most flowering occurs in the study system. Fieldwork was carried out during September 2007 through February 2008 and September 2008 through February 2009. On average, there were 85.8 ± 11.8 plant species per fragment (mean ± SD; range 67–103) and 146.4 ± 21.0 pollinator species per fragment (mean ± SD; range 113–189). Observations were performed by 3.8 persons/day, on average, and always on sunny days. The sampling procedure consisted of recording the flower visitors to plants during a 10-min

observation period (see Plate 1). Observations on plant species were performed at random according to their abundance in the study site. Overall, there were 1390 ± 232 time-based observations per fragment (mean ± SD; range 1215–2004; $n = 10$ observation periods), equivalent to 231 hours of observation per fragment. Only insect species that contacted the anthers or stigma of flowers or entered the flower tube were considered as legitimate visitors. We collected insects in the first year of the study for subsequent taxonomic identification in the laboratory. In total, we identified 186 flowering plant species and 336 pollinator species in the overall system. Analyses of accumulation curves and subsequent fitting to the Clench asymptotic model revealed that our sampling effort captured 96.9% of the expected plant species richness (observed number of species, $S_o = 186$; expected number of species, $S_e = 192$; fit to Clench's model: $r^2 = 0.99$) and 96.3% of the expected pollinator species richness in the global system ($S_o = 336$, $S_e = 349$; $r^2 = 0.99$).

Analyses of sampling effort

To investigate how sampling effort influences network metrics, we used species-based accumulation curves for the number of pollinator species visiting each plant species. We calculated accumulation curves for every plant species, including replicated analyses for the same plant species in different sites. In total, we analyzed 687 species-based curves. For each accumulation curve we estimated the mean and standard deviation from 100 random permutations of the data (Gotelli and Colwell 2001). The maximum number of pollinators at the asymptote of each accumulation curve was estimated through two different asymptotic models. (1) The Clench model, a modification of the Michaelis-Mentel equation, is defined as $S(t) = at/(1 + bt)$; it is appropriate when the probability of adding new species decreases with the number of time-based observations (Soberón and Llorente 1993). (2) The negative exponential model is defined as $S(t) = (a/b)[1 - e^{-bt}]$. In the two models, $S(t)$ represents the predicted number of species at sampling effort t (Fig. 1c), a is the rate of increase at the beginning of the sampling, and b is a fitted constant that controls the shape of the curve. In both cases, when $t \rightarrow \infty$, $S(t) \rightarrow a/b$. In consequence, the expected number of pollinator species in the asymptote (S_e) was calculated as $S_e = a/b$ (Soberón and Llorente 1993). To estimate the parameters a and b, the functions must be adjusted to the observed accumulation curves through a nonlinear estimation procedure. We used the Gauss-Newton algorithm for parameter estimations (Fletcher 2000). The model that showed the best fit to the observed data, as evaluated by Akaike's information criterion, was selected for subsequent analyses (Fox 2002).

Once we estimated the expected number of pollinators per plant species at the asymptote, the sampling completeness (SC) for each plant species was calculated as SC = S_o/S_e, where S_o is the observed number of

species and S_e is the maximum number of species expected according to the model (Fig. 1d). Based on their sampling completeness, the plant species from each site were assigned to one of five SC categories: 0–0.30, 0.31–0.50, 0.51–0.70, 0.71–0.90, and 0.91–1.00. For example, subnetworks in the category 0.31–0.50 of completeness include those plant species for which 31–50% of the pollinator species expected by the asymptotic model were recorded. The width of categories was determined by finding the combination of ranges that provided the lowest variation in the number of plant species across categories. Finally, we calculated network statistics for the set of species belonging to the same SC category, totaling five subnetworks per fragment. Because the same procedure was repeated for the 10 fragments here analyzed, our database consisted of 50 subnetworks, 10 per sampling completeness category. A summary of network characteristics per sampling category is shown in Table 1.

Colwell and Coddington (1994) reviewed diverse estimators of species richness in samples, concluding that parametric estimators are not necessarily independent of the number of samples, especially in small sample sizes. The authors found that Chao 2 and Jackknife estimators provide the least biased estimators for a small number of samples. To ensure that the parametric procedure used in this study does not underestimate species richness at small sample effort categories, we examined the covariation of the expected maximum species richness under Clench's model and Chao 2 procedure with ANCOVA, using sampling completeness categories as treatment effect. All variables were log-transformed. Results revealed an overall association between estimates ($F_{1, 384} = 211.08$, $P < 0.001$, $\beta = 1.099 \pm 0.075$, mean ± SE) and lack of sampling completeness effect ($F_{4, 384} = 1.558$, $P = 0.185$). The slopes for the different sampling categories did not differ ($F_{4, 383} = 0.927$, $P = 0.448$). Overall, these results indicate that both procedures provide about the same richness estimates, thus validating use of the parametric procedure in this study.

Network metrics estimation and statistical analyses

We calculated the effect of sampling completeness on six frequently used network metrics:

1) Connectance (C), the proportion of possible links actually observed in a web, with values from 0 to 1. In bipartite networks it is calculated as $C = I/(P \times A)$, where I is the number of interactions, P is the number of plant species, and A is the number of animal species in the network. Connectance is often considered a measure of network complexity and redundancy (Jordano 1987, Bascompte and Jordano 2006).

2) Nestedness (N), representing one kind of asymmetry of interactions, characterized by a core of highly connected species (generalists) that interact mainly with each other, and a group of specialist species that

TABLE 1.   Network summary statistics (mean $\pm$ 2 SE) for the 50 subnetworks at Reserva Nacional Los Ruiles, Chile, 2007–2008.

Network characteristics	Sampling completeness category				
	0–0.30	0.31–0.50	0.51–0.70	0.71–0.90	0.91–1.00
Sampling completeness per species, SC	0.15 $\pm$ 0.08	0.41 $\pm$ 0.06	0.60 $\pm$ 0.04	0.80 $\pm$ 0.06	0.97 $\pm$ 0.02
Sampling evenness, J'_{obs}	0.88 $\pm$ 0.03	0.95 $\pm$ 0.04	0.93 $\pm$ 0.03	0.92 $\pm$ 0.03	0.88 $\pm$ 0.03
Network size, S	78.0 $\pm$ 13.3	36.9 $\pm$ 6.3	49.6 $\pm$ 8.5	81.0 $\pm$ 13.8	83.8 $\pm$ 14.3
Network asymmetry, NA	1.42 $\pm$ 1.20	5.24 $\pm$ 1.20	5.29 $\pm$ 1.20	6.33 $\pm$ 1.19	8.49 $\pm$ 1.19
Number of plant species, P	32.4 $\pm$ 1.2	6.9 $\pm$ 1.5	8.0 $\pm$ 1.2	10.8 $\pm$ 1.3	10.0 $\pm$ 1.9
Number of animal species, A	45.1 $\pm$ 1.2	29.9 $\pm$ 1.5	41.3 $\pm$ 1.4	70.1 $\pm$ 1.3	71.2 $\pm$ 1.2
Number of interactions, I	63.2 $\pm$ 4.5	92.2 $\pm$ 8.3	93.7 $\pm$ 6.6	105.7 $\pm$ 8.1	93.7 $\pm$ 7.2

Notes: $N = 10$ subnetworks per sampling completeness category. The sampling completeness per species (SC), the fraction of the expected pollinator assemblage that is recorded, was calculated from accumulation curves as $SC = S_o/S_e$, where S_o is the observed number of species and S_e is the expected number of species in the asymptote of Clench or exponential models. For example, subnetworks in the category SC = 0.31–0.50 include those plant species for which 31–50% of the pollinator species expected by the asymptotic model were recorded. Sampling evenness (J'_{obs}) quantifies the heterogeneity of sampling effort among plant species. When J'_{obs} approaches 0, the effort distribution is highly heterogeneous among species; when J'_{obs} approaches 1, the sampling effort is homogeneously distributed among species. The network size (S) within sampling completeness categories was calculated as the sum of the plant (P) and animal (A) species belonging to the interval. Network asymmetry (NA) was calculated from the ratio (A/P).

interact mainly with the generalist species (Bascompte et al. 2003). We used the NODF (nestedness based on overlap and decreasing fills) algorithm proposed by Almeida-Neto et al. (2008), which, unlike previous metrics, corrects for matrix dimensionality (Ulrich et al. 2009). NODF values of 0 and 100 indicate lack of nestedness and perfect nesting, respectively.

3) Modularity (M), providing information on the existence of groups composed of strongly interactive species that interact weakly with species belonging to other modules (Guimerà and Amaral 2005). Diverse algorithms have been suggested to estimate modularity (see Olesen et al. 2007). Here, we used the spinglass algorithm (Reichardt and Bornholdt 2006). M can take values between 0 and ~1, and reaches its maximum when all network interactions are within modules and no link exists among modules.

4) Average path length (L), defined as the average number of steps along the shortest paths for all possible pairs of species in the network, or the fewest number of steps required to connect species i with species j, averaged for all pairs of species (Albert and Barabasi 2002, Olesen et al. 2006). Although not frequently used in studies of mutualistic networks (but see Lundgren and Olesen 2005, Olesen et al. 2006, Bezerra et al. 2009), this metric provides important information on network cohesiveness. Low L values are associated with high levels of species cohesion, a property characteristic of "small-world" systems (Albert and Barabasi 2002).

5) Centralization (E), which permits identification of central species around which peripheral species are assembled. We calculated E as eigenvalue centrality (Bonacich 1972), defined as the principal eigenvalue of the adjacency matrix of a network, with values between 0 and ∞. Eigenvector centrality assigns a value to each species in the network. A species receives a high E value if it is strongly correlated with

many other species that are themselves central within the network. The presence of central nodes confers high network cohesion or information flow (Borgatti and Everett 1997).

6) Robustness (R), providing information on the resilience of networks to losing species, based on the concept of the attack tolerance curve (Albert and Barabasi 2002). Curves are constructed by sequentially removing plant or animal species without replacement. After each species removal (e.g., pollinators), the number of species in the other group (e.g., plants) that remain in the network is calculated. The curve indicates the fraction of species remaining according to the fraction of removed species in the alternative group. The area under this curve (values between 0 and 1) corresponds to the resilience level of the overall system (Burgos et al. 2007). If the decay curve is concave ($R < 0.5$), removing a small fraction of species produces a high number of secondary extinctions. If the curve is convex ($R > 0.5$), the network is highly robust. In this study, we used a random species removal criterion, and estimated independently the robustness after plant removal (R_p) and animal removal (R_a).

Because the sensitivity of metrics to sampling completeness may vary depending on the number of species in the network (network size) and sampling evenness (the homogeneity of sampling effort among plant species), we examined the pattern of covariation of these variables with network metrics. The network size within sampling completeness categories was calculated as the sum of the plant and animal species belonging to the interval. To quantify the heterogeneity of sampling effort among plant species, we used an evenness index (J'_{obs}) based on Shannon diversity (Blüthgen et al. 2008). J'_{obs} approaches 0 when sampling effort distribution is highly heterogeneous, and approaches 1 when sampling effort is homogeneously distributed among species. We

TABLE 2. Summary of ANCOVA results for sampling completeness on network metrics.

Network metrics	Sampling completeness $F_{4,43}$	Network size		Sampling evenness		R^2
		$F_{1,43}$	Trend	$F_{1,43}$	Trend	
Connectance, C	**34.77*****	**13.57*****	−	0.09	0	0.79
Nestedness, NODF	**26.67*****	**16.92*****	+	1.87	0	0.70
Modularity, M	**30.87*****	0.44	0	0.05	0	0.72
Robustness to plant loss, R_p	1.76	**21.58*****	+	4.92*	+	0.63
Robustness to animal loss, R_a	**26.62*****	0.86	0	5.24*	+	0.73
Path length, L	**26.62*****	0.52	0	0.65	0	0.72
Centralization, E	**43.53*****	**43.53*****	+	1.26	0	0.61

Notes: Network size and sampling evenness were used as covariates in every analysis. Symbols + and − indicate positive and negative patterns of covariation with sampling completeness, respectively. Symbol 0 indicates absence of significant covariation. Network size, connectance, mean path length, and centralization values were log-transformed before analysis. Boldface indicates values that retain significance after Bonferroni adjustment.

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

analyzed the effect of sampling completeness on the metric of interest with a one-way ANCOVA, using the completeness category of 0.91–1.00 as reference for comparisons and network size and sampling evenness as covariates. A posteriori contrasts were performed using Tukey HSD. All analyses were performed in R version 2.10 (R Development Core Team 2009) and associated modules: *car* version 1.2–16 (Fox 2002), *vegan* version 1.15–4 (Oksanen et al. 2009), *igraph* version 0.5.3 (Csardi and Nepusz 2006) and *bipartite* version 1.06 (Dormann et al. 2008), all *available online*.[5]

RESULTS

The exponential asymptotic model provided the best fit in 495 accumulation curves (72% of cases). The number of observations per species was variable in the overall system (mean: 26 time-based observations per species, range: 1–283). In spite of this, when data were analyzed at the subnetwork scale, the mean sampling evenness, 0.91 ± 0.05 (mean ± SD), was greater than those reported for other pollination networks (0.67 ± 0.14, $n = 21$ networks; see Blüthgen et al. 2008), indicating that the species belonging to our sampling completeness categories were more homogeneously sampled than commonly reported. The mean completeness of subnetworks did not relate to sampling evenness ($F_{1,48} = 0.34$, $R^2 = 0.007$, $P = 0.560$), suggesting that the heterogeneity in the number of time-based observations among species was associated with the distribution of abundance rather than with variation in sampling effort (Vázquez and Aizen 2006). Network size in the 50 subnetworks was 65.2 ± 24.3 species (mean ± SD), and the animal/plant ratio (network asymmetry) was 5.35 ± 2.92 (mean ± SD). The A/P ratio (network asymmetry) ranged from 1.4 in the category with the lowest sampling completeness (0–0.30) to 8.5 in subnetworks containing the best-sampled plant species (0.91–1.00), indicating that well-characterized networks tend to present a higher asymmetry in the number of interactive species than low-sampled networks (see Table 1).

Connectance values differed among completeness categories, showing an overall increase toward higher values of sampling completeness (Table 2, Fig. 2a). Subnetworks with species belonging to the three categories with lowest completeness (ranging from 0 to 0.70) showed low connectance values and differed statistically from the best-sampled subnetworks (Fig. 2a). Because by definition connectance values result from the number of observed interactions relative to the number of potential interactions ($C = I/P \times A$), it follows that an increase in connectance will result from an increased number of recorded interactions with respect to the total number of species recorded. In fact, the number of interactions (I) increased with sampling effort as expected ($F_{1,48} = 14.8$, $R^2 = 0.22$, $P < 0.001$). Network size, i.e., the total number animal and plant species within a network, was a significant covariate that related inversely to network connectance (Table 2), as shown in several analyses of plant–pollinator networks (Bascompte and Jordano 2007). Sampling evenness, by contrast, showed no relationship with connectance (Table 2; $P = 0.766$), indicating that this metric, one of the most important descriptors of network structure, was largely insensitive to the heterogeneity of sampling effort among plant species.

Nestedness values differed among sampling categories and tended to increase with sampling completeness (Fig. 2b). The only nestedness value differing from the reference was observed in the SC category 0–0.3, where nestedness was underestimated by 78% with respect to the best-sampled category. Our findings are consistent with previous conclusions that nestedness tends to stabilize quickly with increasing sampling effort (Nielsen and Bascompte 2007). In this study, nestedness values increased with network size, as revealed by the significant covariate effect (Table 2; $P < 0.001$). Regarding sampling evenness, our results did not show influence of heterogeneity of sampling effort on nestedness (Table 2). It has been suggested that incomplete and heterogeneous samplings overestimate specialist interactions and underestimate generalist interactions (Blüthgen et al. 2008, Dorado et al. 2011). However, our
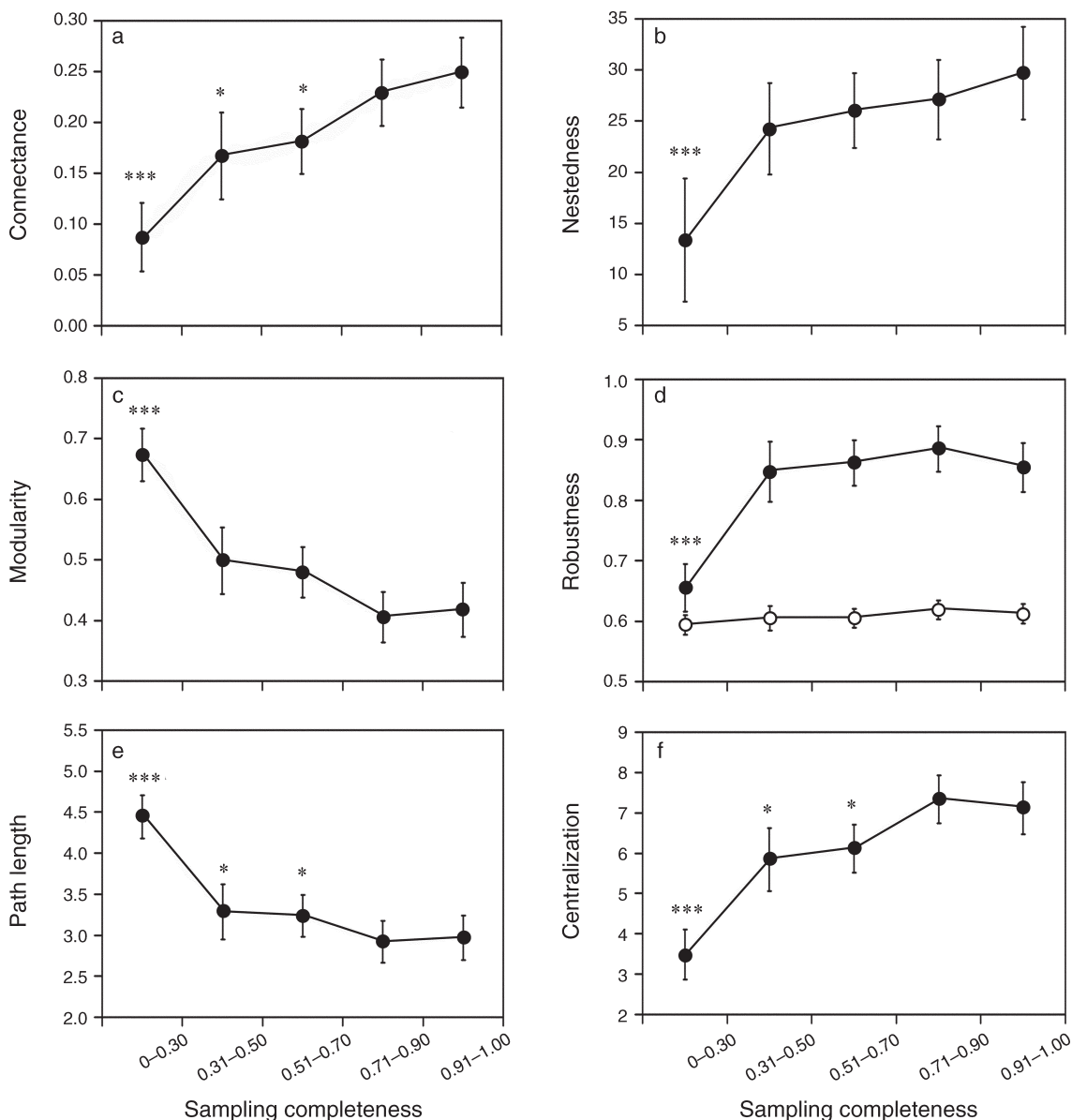
FIG. 2.   Network metric profiles across sampling completeness categories (the fraction of the expected pollinator assemblage that is recorded, an indicator of sampling effort). All metric values are corrected by network size and sampling evenness. Bars around means are 95% confidence intervals from 10 subnetworks contained in completeness intervals. Network metrics include (a) connectance, (b) nestedness, (c) modularity, (d) network robustness to the loss of animal (solid circles) and plant (open circles) species, (e) mean path length, and (f) centralization. The values provided by the most intensively sampled category (0.91–1.00 completeness as revealed by species-based accumulation curves) were used as reference for statistical comparisons in one-way ANCOVA.

* $P < 0.05$; *** $P < 0.001$.

results suggest that the NODF nestedness metric is robust to heterogeneous sampling and shows deviations only at very low sampling effort. In consequence, the NODF metric seems to be more appropriate than temperature-based nestedness, as suggested by Almeida-Neto et al. (2008) and Ulrich et al. (2009).

Mean modularity differed among categories and showed a tendency to decrease with sampling completeness (Table 2), which suggests that detection of

modularity in undersampled networks may represent an artifact associated with insufficient sampling effort. The only category showing a modularity level different from the sampling reference was that in the lowest sampling completeness (Fig. 2c). In principle, low-sampled communities will record a limited number of links, which may result in a pattern of low-connected modules that inflates network modularity in comparison to better-sampled communities. Neither network size

PLATE 1. Plant–pollinator interaction between the solitary bee *Manuelia postica* (Apidae, Xilocopinae) and *Olsynium junceum* (Iridaceae) at Los Ruiles Natural Reserve, Chile. Photo credit: R. Medel.

nor sampling evenness influenced modularity, as revealed by the nonsignificant covariation of variables in ANCOVA (Table 2; $P = 0.511$ and $P = 0.824$, respectively). In general, modularity showed a robust pattern to variation in sampling effort.

The robustness to animal loss ($R_a = 0.82 \pm 0.10$, mean $\pm$ SD) was greater than the robustness to plant loss ($R_p = 0.61 \pm 0.04$) ($t_{98} = 13.69$; $P < 0.001$; Fig. 2d), indicating that animal and plant species removal had different effects on network resilience. More specifically, networks were more resilient when animals were removed, in part because the number of animal species exceeded the number of plant species in the overall system (336 vs. 186 species, respectively), and the effect of species loss may be more easily compensated in the animal rather than plant species assemblage. Regarding the effect of variation in sampling completeness, our results indicate that, except for the low robustness to animal loss detected at very low sampling effort (below 0.30 completeness), where networks were 23% less resilient than the reference value (Fig. 2d), sampling effort had a small impact on network resilience. Network size influenced only the robustness to plant loss (Table 2). The increased resilience of plant species to

increasing network size probably results from the numerical asymmetry between animal and plant species. If robustness estimates are influenced by the total number of species used in analyses, networks composed of a low number of plant species will show a low robustness because only a few plant deletions are required to remove many pollinator species, leading to an overall system destabilization. In consequence, any increase in the number of plant species will have a disproportional effect on plant redundancy, turning the network more robust to plant species loss. On the other hand, if the high number of animal species implies a high redundancy in the plant species with which they interact, any increase in the number of animal species will have a negligible impact on the already high network resilience. Regarding sampling heterogeneity, our results did not show a clear effect on network robustness. Although both R_p and R_a tended to increase with sampling evenness, the statistical significance disappeared after Bonferroni adjustment (Table 2).

Mean path length (L) and centralization (E) metrics were highly sensitive to sampling effort. Values of L and E calculated from subnetworks containing species with completeness lower than SC = 0.70 differed significantly

from the best-sampled subnetworks (Fig. 2e and f, respectively). In general, increased sampling completeness conferred high network cohesion, as revealed by the decreasing mean path length and increasing centralization values toward higher categories of sampling completeness. Because more interactions are recorded in well-sampled communities, species will be more connected, on average, which implies that a lower distance is necessary to connect pairs of interacting species. Similarly, the greater number of interactions recorded in well-sampled communities will increase the importance of generalist species in determining network cohesiveness. ANCOVA results indicate that centralization, but not average path length, was influenced by network size (Table 2), suggesting that the importance of generalist species for network cohesion tends to increase in large-sized communities. Neither path length nor centralization was affected by sampling evenness.

DISCUSSION

Except for the robustness to plant loss, all of the network metrics examined here were sensitive to some extent to sampling effort. The most sensitive metrics were connectance, path length, and centralization, which gave underestimated values below sampling completeness of 0.7 (Fig. 2a, e, f). Nestedness, modularity, and robustness to animal loss were the most robust metrics, differing from well-sampled communities only in those subnetworks with plant species showing <30% of the expected pollinator assemblage (Fig. 2b–d). Although nestedness and robustness to animal loss showed underestimated values in the lowest sampling interval, modularity was overestimated by almost 40% (Fig. 2c). Regarding network size effects, our results varied depending on the network metric under assessment. For example, connectance, robustness to animal loss, and centralization showed high sensitivity to variation in network size, but nestedness, modularity, robustness to plant species loss, and path lengths were not affected. Effects of sampling evenness were less prevalent in our analysis. The only metrics influenced by the heterogeneity of sampling effort among plant species were those related to robustness (Table 2). Specifically, our results indicate that communities where sampling effort is more homogeneously distributed among plant species will show a higher robustness to species loss.

Our results assume that the most important cause responsible for metric behavior across sampling effort categories is sampling completeness per plant species. However, subnetworks belonging to the same sampling category may include species that share biological properties such as phenology and floral characters, among other variables, which might account, in part, for the overall pattern. We examined this possibility by comparing the similarity in species composition, as estimated by Jaccard's index, among and within sampling categories. Subnetworks within sampling categories were more similar in plant species composi-

tion than those belonging to different sampling categories (ANOSIM test with 1000 permutations, $F_{4,45} = 2.24$, $P < 0.001$). To gain insight into the factors influencing similarity in species composition, we used partial Mantel tests to examine whether the correlation in the sampling completeness distance was accounted for by sampling effort (number of time-based observations per plant species), taxonomic distance (phenotypic overlap), or phenology distance (temporal overlap). The statistical significance of such correlations was assessed by 1000 permutations. Our results indicate that variation in sampling completeness was accounted for by sampling effort ($r_m = 0.327$, $P = 0.001$) but not taxonomy ($r_m = -0.012$, $P = 0.759$) and phenology ($r_m = -0.001$, $P = 0.509$). These results suggest that similarity in species composition within categories was accounted for by the relative abundance of species in the overall system, which in turn influenced the sampling effort received by each plant species, rather than by taxonomic and phenological factors.

The aim of this study was to examine the extent to which a number of frequently used network metrics are influenced by sampling intensity, network size, and sampling evenness. Although other studies have previously evaluated the effect of network size on the structure of plant–animal relationships (e.g., Jordano 1987, Ollerton and Cranmer 2002, Blüthgen et al. 2006, Vázquez and Aizen 2006, Almeida-Neto et al. 2007, Nielsen and Bascompte 2007), studies examining the contribution of different sources of bias to network metric estimates have been notoriously absent in the plant–pollinator literature. Even though species-based accumulation curves have been used to examine the impact of sampling effort on single network metrics (e.g., nestedness as in Nielsen and Bascompte 2007) and the level of generalization of individual plants in populations of single species (e.g., Herrera 2005), this is the first study using species-based accumulation curves to evaluate the behavior of network descriptors to variation in sampling effort in plant–pollinator networks.

Regarding the effect of network size on network metrics, there is an important conceptual distinction between our descriptive approach and scale–variance analysis. It is one thing to state that a given metric is statistically sensitive to variations in network size (or any other aspect of network structure), which is often examined using random matrices and investigating how the values recorded for a given metric vary with network size (e.g., Bersier et al. 1999, Vázquez and Aizen 2006, Ulrich et al. 2009). It is another thing to state that a given pattern has scale variance, which is often analyzed by examining the statistical relationship of a metric with network size at different ranges of network size, often implying an underlying biological process (e.g., Martinez 1992). In this work, we have described the statistical sensitivity of network metrics to sampling effort, keeping constant network size and sampling evenness.

Because we do not know the statistical behavior of most network metrics in response to network size, we cannot ascribe causality to the network size dependency shown by the metrics here analyzed (Table 2).

Results from this study have important consequences for the way in which field studies on plant pollinator should be designed. Because network metrics have different sensitivities to sampling effort and network size, studies having low sampling effort need to be interpreted with caution. For instance, although nestedness, modularity, and robustness to species loss seem to be robust estimators that show bias only at very low sampling effort, conclusions on connectance, average path length, and centralization may be more sensitive to low sampling because they rely directly on the number of interactions recorded. The lack of effect of sampling evenness on most metrics suggests that heterogeneous sampling effort across plant species is likely to have a small influence on network metric biases. In consequence, the reliability of network metrics may be independent of whether sampling is homogeneous (i.e., a fixed an equal number of time-based observations per plant; Lundgren and Olesen 2005) or random (i.e., recordings proportional to the relative abundance of plant species, such as in this study) across species.

By identifying the network metrics most sensitive to sampling effort, our results provide a baseline from which to overcome the inherent sampling limitations of most empirical studies. For instance, consideration of additional variables such as the phenology of species (e.g., Olesen et al. 2008, Petanidou et al. 2008), pollen load (e.g., Bosch et al. 2009), pollen limitation (e.g., Power and Stout 2011), species abundance (e.g., Vázquez and Aizen 2006), invasion status of interacting species (e.g., Aizen et al. 2008, Vilà et al. 2009), and other factors may help to provide more complete descriptions of plant–pollinator networks. We strongly encourage researchers to document sampling effort using accumulation curves at either network or species levels. This information will permit standardization of network metrics to variation in sampling procedures and control of such effects in the search for broad patterns in plant–pollinator networks.

Literature Cited

Aizen, M. A., C. L. Morales, and J. M. Morales. 2008. Invasive mutualists erode native pollination webs. PLoS Biology 6:396–403.

Albert, R., and A. L. Barabasi. 2002. Statistical mechanics of complex networks. Reviews of Modern Physics 74:47–97.

Almeida-Neto, M., P. Guimarães, P. Guimarães, Jr., R. Loyola, and W. Ulrich. 2008. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. Oikos 117:1227–1239.

Almeida-Neto, M., P. R. Guimarães, and T. M. Lewinsohn. 2007. On nestedness analyses: rethinking matrix temperature and anti-nestedness. Oikos 116:716–722.

Banasek-Richter, C., M. F. Cattin, and L. F. Bersier. 2004. Sampling effects and the robustness of quantitative and qualitative food-web descriptors. Journal of Theoretical Biology 226:23–32.

Bascompte, J., and P. Jordano. 2006. The structure of plant–animal mutualistic networks. Pages 143–159 in M. Pascual and J. Dunne, editors. Ecological networks: linking structure to dynamics in food webs. Oxford University Press, Oxford, UK.

Bascompte, J., and P. Jordano. 2007. Plant–animal mutualistic networks: the architecture of biodiversity. Annual Review of Ecology, Evolution, and Systematics 38:567–593.

Bascompte, J., P. Jordano, C. J. Melian, and J. M. Olesen. 2003. The nested assembly of plant–animal mutualistic networks. Proceedings of the National Academy of Sciences USA 100:9383–9387.

Bascompte, J., P. Jordano, and J. M. Olesen. 2006. Asymmetric coevolutionary networks facilitate biodiversity maintenance. Science 312:431–433.

Bersier, L., P. Dixon, and G. Sugihara. 1999. Scale-invariant or scale-dependent behavior of the link density property in food webs: a matter of sampling effort? American Naturalist 153:676–682.

Bezerra, E. L., I. C. Machado, and M. A. R. Mello. 2009. Pollination networks of oil flowers: a tiny world within the smallest of all worlds. Journal of Animal Ecology 78:1096–1101.

Blüthgen, N., J. Fründ, D. P. Vázquez, and F. Menzel. 2008. What do interaction network metrics tell us about specialization and biological traits? Ecology 89:3387–3399.

Blüthgen, N., F. Menzel, and N. Blüthgen. 2006. Measuring specialization in species interaction networks. BMC [Bio Med Central] Ecology 6:9.

Bonacich, P. 1972. Factoring and weighting approaches to status scores and clique identification. Journal of Mathematical Sociology 2:113–120.

Borgatti, S. P., and M. G. Everett. 1997. Network analysis of 2-mode data. Social Networks 19:243–270.

Bosch, J., A. M. Martín González, A. Rodrigo, and D. Navarro. 2009. Plant–pollinator networks: adding the pollinator's perspective. Ecology Letters 12:409–419.

Burgos, E., H. Ceva, R. P. J. Perazzo, M. Devoto, D. Medan, M. Zimmermann, and A. M. Delbue. 2007. Why nestedness in mutualistic networks? Journal of Theoretical Biology 249:307–313.

Bustamante, R. O., and C. Castor. 1998. The decline of an endangered temperate ecosystem: the ruil (Nothofagus alessandrii) forest in central Chile. Biodiversity and Conservation 7:1607–1626.

Chacoff, N. P., D. P. Vázquez, S. B. Lomáscolo, E. L. Stevani, J. Dorado, and B. Padrón. 2011. Evaluating sampling completeness in a desert plant–pollinator network. Journal of Animal Ecology 81:190–200.

Colwell, R., and J. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society of London B 345:101–118.

Colwell, R. K., C. X. Mao, and J. Chang. 2004. Interpolating, extrapolating and comparing incidence-based species accumulation curves. Ecology 85:2717–2727.

Csardi, G., and T. Nepusz. 2006. The igraph software package for complex network research. InterJournal, Complex Systems 1695. http://igraph.sf.net

Dorado, J., D. P. Vázquez, E. Stevani, and N. P. Chacoff. 2011. Rareness and specialization in plant–pollinator networks. Ecology 92:19–25.

Dormann, C. F., B. Gruber, N. Blüthgen, and J. Fründ. 2009. Indices, graphs and null models: analyzing bipartite ecological networks. The Open Ecology Journal 2:7–24.

Dormann, C. F., B. Gruber, and J. Fründ. 2008. Introducing the bipartite package: analysing ecological networks. R News 8:8–11.

Dunne, J. A., R. J. Williams, and N. D. Martinez. 2002. Network structure and biodiversity loss in food webs: robustness increases with connectance. Ecology Letters 5:558–567.

Fletcher, R. 2000. Practical methods of optimization. Second edition. John Wiley, Padstow, UK.

Fortuna, M., and J. J. Bascompte. 2006. Habitat loss and the structure of plant–animal mutualistic networks. Ecology Letters 9:278–283.

Fox, J. 2002. An R and S-PLUS companion to applied regression. Sage Publications, Thousand Oaks, California, USA.

Gibson, R. H., B. Knott, T. Eberlein, and J. Memmott. 2010. Sampling method influences the structure of plant–pollinator networks. Oikos 120:822–831.

Gonzalez, A., B. Rayfield, and Z. Lindo. 2011. The disentangled bank: how loss of habitat fragments and disassembles ecological networks. American Journal of Botany 98:503–516.

Gotelli, N. J., and R. K. Colwell. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecology Letters 4:379–391.

Guimarães, P. R., G. Machado, M. A. M. de Aguiar, P. Jordano, J. Bascompte, A. Pinheiro, and S. F. dos Reis. 2007. Build-up mechanisms determining the topology of mutualistic networks. Journal of Theoretical Biology 249:181–189.

Guimerà, R., and L. A. N. Amaral. 2005. Cartography of complex networks: modules and universal roles. Journal of Statistical Mechanics. Theory and Experiment, article P02001.

Herrera, C. M. 2005. Plant generalization on pollinators: species property or local phenomenon? American Journal of Botany 92:13–20.

Jordano, P. 1987. Patterns of mutualistic interactions in pollination and seed dispersal: connectance, dependence asymmetries, and coevolution. American Naturalist 129:657–677.

Jordano, P., J. Bascompte, and J. M. Olesen. 2003. Invariant properties in coevolutionary networks of plant–animal interactions. Ecology Letters 6:69–81.

Jordano, P., D. Vázquez, and J. Bascompte. 2009. Redes complejas de interacciones mutualistas planta–animal. Pages 17–41 in R. Medel, M. A. Aizen, and R. Zamora, editors. Ecología y evolución de interacciones planta–animal. Editorial Universitaria, Santiago, Chile.

Lundgren, R., and J. M. Olesen. 2005. The dense and highly connected world of Greenland's plants and their pollinators. Arctic, Antarctic, and Alpine Research 37:514–520.

Martinez, N. D. 1992. Constant connectance in community food webs. American Naturalist 139:1208–1218.

Memmott, J., N. M. Waser, and M. V. Price. 2004. Tolerance of pollination networks to species extinctions. Proceedings of the Royal Society of London B 271:2605–2611.

Nielsen, A., and J. Bascompte. 2007. Ecological networks, nestedness and sampling effort. Journal of Ecology 95:1134–1141.

Oksanen, J., R. Kindt, P. Legendre, B. O'Hara, G. L. Simpson, P. Solymos, M. Henry, H. Stevens, and H. Wagner. 2009. Vegan: community ecology package. R package version 1.15-4. http://www.cran.r-project.org/package=vegan

Olesen, J. M., J. Bascompte, Y. L. Dupont, and P. Jordano. 2006. The smallest of all worlds: pollination networks. Journal of Theoretical Biology 240:270–276.

Olesen, J. M., J. Bascompte, Y. L. Dupont, and P. Jordano. 2007. The modularity of pollination networks. Proceedings of the National Academy of Sciences USA 104:19891–19896.

Olesen, J. M., J. Bascompte, H. Elberling, and P. Jordano. 2008. Temporal dynamics in a pollination network. Ecology 89:1573–1582.

Ollerton, J., and L. Cranmer. 2002. Latitudinal trends in plant–pollinator interactions: are tropical plants more specialised? Oikos 98:340–350.

Petanidou, T., A. S. Kallimanis, J. Tzanopoulos, S. P. Sgardelis, and J. D. Pantis. 2008. Long-term observation of a pollination network: fluctuation in species and interactions, relative invariance of network structure and implications for estimates of specialization. Ecology Letters 11:564–575.

Power, E. F., and J. C. Stout. 2011. Organic dairy farming: impacts on insect–flower interaction networks and pollination. Journal of Applied Ecology 48:561–569.

R Development Core Team. 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reichardt, J., and S. Bornholdt. 2006. Statistical mechanics of community detection. Physical Review E 74:016110.

Rezende, E. L., J. E. Lavabre, P. R. Guimarães, Jr., P. Jordano, and J. Bascompte. 2007. Nonrandom coextinctions in phylogenetically structured mutualistic networks. Nature 448:925–929.

San Martin, J., H. Figueroa, and C. Ramirez. 1984. Fitosociología de los bosques de ruil (Nothofagus alessandrii Espinosa) en Chile. Revista Chilena de Historia Natural 57:171–200.

Santamaria, L., and M. A. Rodríguez-Gironés. 2007. Linkage rules for plant–pollinator networks: trait complementarity or exploitation barriers? PLoS Biology 5:354–362.

Soberón, J., and J. Llorente. 1993. The use of species accumulation functions for the prediction of species richness. Conservation Biology 7:480–488.

Tylianakis, J. M., E. Laliberte, A. Nielsen, and J. Bascompte. 2010. Conservation of species interaction networks. Biological Conservation 143:2270–2279.

Ulrich, W., M. Almeida-Neto, and N. J. Gotelli. 2009. A consumer's guide to nestedness analysis. Oikos 118:3–17.

Vázquez, D. P., and M. A. Aizen. 2006. Community-wide patterns of specialization in plant–pollinator interactions revealed by null models. Pages 201–219 in N. M. Waser and J. Ollerton, editors. Plant–pollinator interactions: from specialization to generalization. University of Chicago Press, Chicago, Illinois, USA.

Vilà, M., I. Bartomeus, A. C. Dietzsch, T. Petanidou, I. Steffan-Dewenter, J. C. Stout, and T. Tscheulin. 2009. Invasive plant integration into native plant–pollinator networks across Europe. Proceedings of the Royal Society of London B 276:3887–3893.