The non-random assembly of indirect interactions in plant-pollinator networks

Jose B. Lanuza[1,2] | **Alfonso Allen-Perkins[2,3]** | **Ignasi Bartomeus[2]**

1 School of Environmental and Rural Science, University of New England, Armidale, New South Wales 2350, Australia.

2 Estación Biológica de Doñana (EBD-CSIC), E-41092 Seville, Spain.

3 Departamento de Ingeniería Eléctrica, Electrónica, Automática y Física Aplicada, ETSIDI, Universidad Politécnica de Madrid, 28040 Madrid, Spain.

Abstract

Ecological processes leave distinct structural imprints on the species interactions that shape the topology of animal-plant mutualistic networks. Detecting how direct and indirect interactions between animals and plants are organized is not trivial since they go beyond pairwise interactions, but may get blurred when considering full network descriptors. Recent work has shown that the meso-scale, the intermediate level of network complexity between the species and the full network, can capture this important information. The meso-scale describes network subgraphs representing patterns of direct and indirect interactions between a small number of species and when these network subgraphs differ statistically from a benchmark, they are often referred to as "network motifs." Despite motifs can capture relevant ecological information of species interactions, they remain overlooked in natural plant-pollinator networks. By exploring 60 empirical plant-pollinator networks from 18 different studies with wide geographical coverage we show that some network subgraphs are consistently under- or over-represented, suggesting the presence of worldwide network motifs in plant-pollinator networks. In addition, we found a higher proportion of densely connected network subgraphs, which could reflect that neutral effects are the predominant ecological process governing the structure of the meso-scale. We found that distinct subgraph positions describing species ecological roles (e.g., generalisation and number of indirect interactions) are occupied by different floral visitor and plant groups on both trophic levels. For example, bees tend to appear in subgraph positions with a lower number of indirect interactions than the rest of floral visitor groups. Interestingly, the realized combinations of different species groups within a network subgraph can not be retrieved from their joint probability distributions, indicating that plant and floral visitor group combinations on network subgraphs are not random. Our results highlight the non-random structure of the key building blocks shaping plant-pollinators networks and the association of certain network subgraphs and their positions with specific groups of floral visitors and plants.

Keywords: pollinators, plant-pollinator networks, plant-pollinator interactions, network motifs, indirect interactions, floral visitors

Introduction

Ecological communities are formed by a plethora of interacting species that form networks of interactions. Because of the tremendous complexity of these ecological networks, species interactions are generally condensed in metrics that summarize this information (Guimarães, 2020). Plant-pollinator interactions are no exception, and they are often studied with indices that aggregate the information at the network (macro-scale) and species level (micro-scale). For example, plant-pollinator network level approaches have identified common invariant structural properties across mutualistic networks, including a degree distribution that decays as a power law (Jordano, 1987), nestedness (Bascompte et al., 2003), or modularity (Olesen et al., 2007). In addition, the local species position within the network can define its degree of specialisation (Blüthgen et al., 2006) or its role in connecting the rest of the community (Olesen et al., 2007). Despite the unquestionable progress with the use of these metrics at both network and species level, condensing complex information into a single metric implies the loss of relevant ecological information that obscures the understanding of species interactions (Cirtwill et al., 2018; Simmons, Cirtwill, et al., 2019).

Traditionally, plant-pollinator research has focused on direct interactions but overlooked indirect interactions (i.e., the mediated effect between two species by a third species; Holt & Lawton, 1994), such as facilitative or competitive interactions between plants for pollinators (Carvalheiro et al., 2014; Moeller, 2004; Sargent & Ackerly, 2008). Despite the widespread nature of indirect interactions in ecological communities (Strauss, 1991), plant-pollinator research often fails to finely capture those indirect interactions with the conventional analytical tools that condense the information either by species (e.g., degree) or in global topological indices (e.g., nestedness). Nonetheless, the analysis of network subgraphs at the meso-scale, the intermediate level of complexity between the full network and the species level, allows to consider both direct and indirect interactions (Simmons, Cirtwill, et al., 2019). The network subgraphs that significantly depart from an expected benchmark are often defined as 'network motifs' (Milo et al., 2002). The analysis of network motifs in plant-pollinator networks have revealed that the different drivers that govern species interactions (e.g., species abundances versus trait-matching) can lead to different patterns of indirect interactions (Simmons et al., 2020). Yet, the global patterns of indirect interactions in real plant-pollinator networks are still unknown (e.g., over- and under- represented network subgraphs).

By linking the structural properties of the meso-scale with the species' ecology we can advance in our understanding on the determinants of species interactions. For instance, different motifs can have different ecological meanings (Simmons, Cirtwill, et al., 2019) and the position within a motif can determine the species' functional role (Baker et al., 2015; Stouffer et al., 2012). Yet, our understanding of the ecological

2

implications of network motifs in real plant-pollinator networks remains in its infancy. Moreover, it is unclear how the species ecology and life-history traits determine its functional role within the network of interactions (Coux et al., 2016). For example, large pollinators can forage larger distances (Greenleaf et al., 2007), deposit greater pollen quantities (Földesi et al., 2021) and handle complex zygomorphic flowers in comparison with small pollinators that are restricted to lower floral complexity (Gong & Huang, 2009). Despite these obvious differences in pollinator behaviour, how contrasting species' ecology translates into different interaction patterns is still unknown. Similarly, recent empirical findings indicate that the meso-scale is the best descriptor of plant reproductive success (Allen-Perkins et al., 2021), but little is known about how plants reproductive strategies shape their position within the network of interactions. Although some studies have evaluated plant reproductive strategies in plant-pollinator networks (Lázaro et al., 2020; Tur et al., 2013), this is an often overlooked aspect in a community context (Devaux et al., 2014) and rarely incorporated into plant-pollinator network studies. Hence, exploring how the main plant and floral visitor life-history strategies integrate with the emergent framework of network motifs could help to progress knowledge on the ecological processes that shape plant-pollinator interactions.

Here, we used 60 empirical networks from 18 different studies and 14 countries to evaluate the structure of the meso-scale in plant-pollinator networks. From these networks we extracted plant and floral visitors species and determined the principal groups that define their main life-history strategies. To obtain plant groups, we used a comprehensive dataset that included floral, reproductive and vegetative traits compiled in a larger set of plant species. Floral visitor groups were divided by the main taxonomic groups that differed in life form and behaviour. Once we split the different plant-pollinator networks into their subgraphs elements, we explored: (i) if there is a common invariant structural property in the overall subgraphs from these networks (i.e., over- and under-represented subgraphs); (ii) which plant and floral visitor groups are over- and under-represented in different subgraphs positions associated with specific ecological roles; and, (iii) if there are over- and under-represented plant and floral visitor group combinations within a subgraph.

Methods

Overview

To investigate the structure of the meso-scale in plant-pollinator networks, we gathered published plant-pollinator networks with worldwide distribution. Then, we classified the different plant and floral visitor species from these networks into groups that summarise their main life-history strategies. After, we calculated the normalised subgraph frequencies and investigated if these frequencies are similar or not to the ones

expected by chance with the help of null models. Then, for each group of plants and floral visitors we investigated the ecological implications of their overall representation on the different subgraph positions, by comparing, on the one hand, the observed (normalised) frequencies of each group in the different subgraph positions with the frequencies of their simulated counterparts, and, on the other hand, the differences among the patterns of connections of subgraph positions. Finally, we compared if there are differences in the observed frequencies of subgraph combinations from empirical networks and the expected frequencies calculated as the product of the probabilities of the different plant and floral visitor groups within the network subgraph.

Plant-pollinator studies

We have compiled 60 plant-pollinator networks from 18 different studies (Table S1). All studies sampled plant-pollinator interactions in natural systems and were selected based on wide geographical coverage (Fig. S1) and presence of interaction frequency as a measure of interaction strength. In total, there were 503 plant species, 1,111 floral visitors species and 6,248 pairwise interactions registered. For ease of data manipulation, plant and floral visitor species names were standardize with the help of the R package *taxize* version *0.9.99* (Chamberlain et al., 2020). All analyses and data manipulation were conducted in R *version 4.0.5* (R Core Team, 2021).

Plant and floral visitor groups

First, plant species were grouped through hierarchical clustering into the optimal number of functional groups that summarized the main plant reproductive strategies. For this, we used the trait dataset collated in Lanuza et al. (2021) that comprised a total of 1506 species including the 503 species considered in this study (Table S2). This dataset consisted of 8 floral, 4 reproductive biology and 3 plant morphology traits excluding traits with high percentage of missing values (over 30%; Table S3). We opted to calculate the plant functional groups on this larger set of species because of the higher accuracy when delimiting functional groups with that many variables and species (Dolnicar et al., 2014). To feed the clustering analysis, we calculated the distance between the different qualitative and scaled quantitative variables with Gower distance (Gower, 1971). For this, we used the function *gowdis* with method *ward.D2* from the *FD* package version *1.0-12* (Laliberté et al., 2014). Finally, we conducted hierarchical clustering with the function *hclust* from the *R* stats package version *4.0.5* and calculated the optimal number of clusters with the function *kgs* from the *maptree* package version *1.4-7* (White & Gramacy, 2009).

Second, floral visitors were aggregated into taxonomic groups based on taxonomic rank as done similarly in other plant-pollinator studies (Fenster et al., 2004; Ollerton et al., 2009). We opted to divide floral visitors on the taxonomic rank level and not with functional traits because the main orders of floral visitors differed in

form and behaviour and had lower superior taxonomic complexity (i.e., floral visitors had 6 orders versus plants that had 38). Thus, this allowed us to group floral visitors into groups that represented approximately the main life strategies of the possible pollinators: (i) bees (Hymenoptera-Anthophila), (ii) non-bee Hymenoptera (Hymenoptera-non-Anthophila), (iii) syrphids (Syrphidae-Diptera), (iv) non-syrphids-Diptera, (v) Lepidoptera and (vi) Coleoptera. However, a minor set of species belonged to other taxonomic groups that were considered in analyses but not discussed further because of their low representation in the full set of networks (3.55% of the total interactions recorded). These taxonomic groups were 'lizards,' 'birds' and 'other insects.' This last group was formed by a mix of uncommon insect taxa on the full set of networks.

Overall network subgraphs patterns

Following previous work (Simmons, Cirtwill, et al., 2019; Simmons et al., 2020), we broke down the plant-pollinator networks into their constituent network subgraphs that can capture both direct and indirect interactions. Prior to analyses, we turned the quantitative networks into qualitative (or binary) ones, where interactions are present or absent. This approach was preferred since quantitative subgraph analysis have been developed just recently and more work is needed to fully understand and interpret the information that they can convey (Simmons et al., 2020). All analyses were run considering singletons (species with only one interaction detected) but we conducted a second exploration without singletons (64.98% of interactions) to evaluate the effect of rare species on the observed patterns. Results without singletons are qualitatively consistent and are not further discussed in the main text (Figs S2 and S3).

We calculated the frequency of all subgraphs up to five nodes (17 different subgraphs in total; see Fig. 1) for each empirical network, by using the *bmotif* package version *2.0.2* (Simmons, Sweering, et al., 2019). We focused exclusively on subgraphs up to five nodes in our analyses for computational reasons and for the ease of visualization and interpretation of a reduced set of less complex network structures. Throughout the entire manuscript, floral visitors occupy the top nodes of the different network subgraphs and plants the bottom ones. To control for variation in network size, subgraph frequencies were normalised as a proportion of the total number of subgraphs within each subgraph class (i.e., 2-species, 3-species, 4-species and 5-species network subgraphs). In addition, we excluded 2-species (or nodes) subgraphs from our analyses because its normalised frequencies would always equal one. Note that the aim is not comparing different networks, which may differ in sampling effort, but correctly describing the structure of each individual network. Like Simmons et al. (2020), we classified the different subgraphs by their average path length (mean number of links between all pairs of nodes) into four different groups: (i) complete or strong ($\overline{x} = 1.38$), (ii) fan or medium-strong ($\overline{x} = 1.48$), (iii) assymmetric complete or medium-weak ($\overline{x} = 1.60$) and (iv) core-peripheral or weak ($\overline{x} = 1.85$), where $\overline{x}$ represents the mean of the average path length across subgraphs in the same
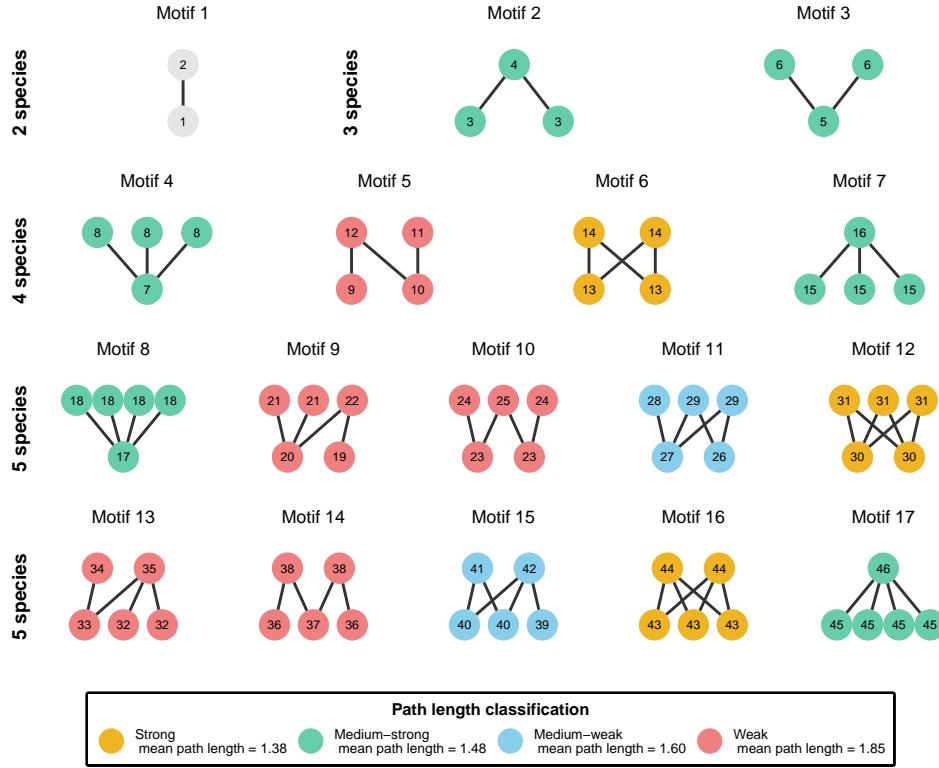
5

Figure 1. Adapted figure of Simmons et al., (2019). All possible network subgraphs from two to five species in bipartite networks grouped by the different subgraph classes (2-species, 3-species, 4-species and 5-species). There is a total of 17 possible subgraphs with 46 different positions denoted within each node. The different subgraphs are aggregated by the broad indirect interaction structures they represent on four different groups (strong, medium-strong, medium-weak and weak). For each group, we show in the legend the mean value of the average path length of all subgraphs within it. The different path length groups range from densely connected network structures (strong subgraphs where all possible interactions are realised and the mean path length is minimum) to weakly connected ones (weak subgraphs where the mean path length is maximum).

group. In general terms, plant and floral visitor groups on more densely connected network subgraphs (e.g., complete or fan) will experience lower number of indirect interactions in contrast to groups on less densely connected structures (e.g., medium-weak or weak).

To assess the significance of the observed frequencies, we created 1,000 simulated networks for each binary network using the *nullmodel* function and the *vaznull* model in the *bipartite* package version *2.14* (Dormann et al., 2009). Generated networks had the same number of plants and floral visitors, as well as the same connectance of their corresponding empirical networks. We discarded null models with more constraints (e.g., constraining for species degree) because the permutations did not generate sufficient variability and provided highly similar or identical networks to the original ones. After extracting the subgraph frequencies from the simulated networks, for each subgraph type and empirical network, we calculated the percentage of simulated networks whose frequencies were smaller than the ones observed, that is, we estimated the percentile of the observed subgraph frequencies. Subgraphs whose percentile is close to 0 or 100 are under- or

over-represented in empirical networks, respectively, and indicate the presence of network motifs as they differ statistically from the network structures generated with simulated networks. To summarize general patterns across networks, we used an intercept-only linear mixed model (LMM) per subgraph with the help of the *lme4* package version *1.1-21* (Bates et al., 2015), where the response variable was the observed subgraph percentile per network. Note that as we are only interested in modelling the intercept value, no predictors were specified. In these models, we used the study identifiers in Table S1 as a random intercept. By doing so, we obtained estimates of the average subgraph frequency, but controlling for the variation at the study level.

Over- and under-represented groups on subgraph positions

We calculated which plant and floral visitor groups were over- or under-represented in the different subgraph positions by comparing position frequencies of empirical networks with those of their corresponding simulated counterparts. To estimate the position frequencies of each plant and floral visitor group in a given network, we added the frequencies of those species that belong to the same group, and then, we normalised the resulting frequencies by dividing them by the total number of times that a corresponding group appears in any position within the same subgraph size class (i.e., 2, 3, 4 or 5 species motifs). Then, we calculated the percentile of the observed position frequencies for each group and network, just like we did with subgraph frequencies. To outline the general patterns of position frequencies across networks and groups, we fit a LMM per subgraph position, where the response variable was the observed position percentile per network. We used the group identifier as an explanatory variable and the study identifiers as a random intercept. Hence, we assessed the average subgraph frequency per species group, after controlling for the variation at the study level. Finally, we visualized with the help of the *ComplexHeatmap* package version *2.6.2* (Gu et al., 2016) over- and under-representation of plant and floral visitor groups on the different subgraph positions.

To understand if the different plant and floral visitor groups are associated with subgraph positions that represent different network roles, we characterized the roles of each subgraph position for plant and floral visitor groups in terms of their specialisation and number of indirect interactions. To that end, we calculated the specificity of each subgraph position following Poisot et al. (2015) as $s_i = (L - l_i)/(L - 1)$, where L is the total number of groups and l_i the number of interaction partners of the group i in a given subgraph. Values of the vector *si*, range between 1, complete specialisation (1 partner within the subgraph) and 0 complete generalism (all possible partners within the subgraph). Then, because all nodes within a motif are connected, we calculated the number of indirect interactions (n_i) for each motif position as $n_i = z_i - k_i - 1$, where z_i is the total number of nodes in a given motif, and k_i represents the degree of the corresponding motif position, that is, the total number of links connected to such position. Thus, we considered all possible indirect interactions for each position. Finally, we analysed in a single model how specialization and the

7

₂₀₂ number of indirect interactions were associated with the observed percentiles for each trophic guild and

₂₀₃ subgraph position (see paragraph above). For this, we fitted a LMM model with the help again of the *lme4*

₂₀₄ package where we considered specialization and number of indirect interactions as response variables with the

₂₀₅ interaction by plant and floral visitor group for both response variables. Further, we considered as random

₂₀₆ factor subgraph position nested within study system. All model diagnostics were conducted with the help of

₂₀₇ the *Dharma* package (Hartig & Hartig, 2017) version *0.3.3.0* and model outputs were visualized with the

₂₀₈ *visreg* package (Breheny et al., 2020) version 2.7.0.

Over- and under-represented group combinations of subgraphs

₂₁₀ Finally, we studied which combinations of plant and floral visitor groups tend to appear together within

₂₁₁ the same subgraphs (up to five nodes). That is, which combinations are over- or under-represented. This

₂₁₂ analysis used 57 out of the 60 networks available due to computational limitations to identify all the nodes

₂₁₃ in the subgraphs of the three networks with the highest number of links. To do so, for each of the 53,250

₂₁₄ possible subgraph combinations, we estimated the observed and the expected probability of finding those

₂₁₅ combinations in empirical networks, respectively. Then, we determined whether the observed probabilities

₂₁₆ are likely to come from the expected probabilities or not. To calculate the observed probability (p_i^O) of the

₂₁₇ different groups within a subgraph i (e.g., subgraph 3, 'bee' + 'bee' + 'selfing herbs'), we divided the number

₂₁₈ of times that i appears in our set of empirical networks (n_i^O) by the sum of the number of times that each

₂₁₉ possible combination appeared: $p_i^O = n_i^O / \sum_{k=1}^{53,250} n_k^O$. To estimate the expected probability (p_i^E) of a given

₂₂₀ subgraph combination i, we assumed that the probability of finding a given group x in the position α is

₂₂₁ independent of the probabilities associated with other positions and groups in that subgraph. Therefore,

₂₂₂ the expected probability (p_i^E) of the subgraph combination i is given by the product of the probabilities

₂₂₃ of its constituents, i.e., $p_i^E = \prod_{(x,\alpha) \in i} p_i(x, \alpha)$, where $p_i(x, \alpha)$ is the probability of finding a given group

₂₂₄ x in the position α of i. Then, by controlling for the variation at the network level with an LMM (see

₂₂₅ below), we calculated the average number of times that the group x appears in the position α (denoted

₂₂₆ as $n_i(x, \alpha)$). To obtain the expected probability $p_i(x, \alpha)$ we divided the number of times that the group x

₂₂₇ appears in the position α by the the total number of times that any group was observed in that position (that

₂₂₈ is, $p_i(x, \alpha) = n_i(x, \alpha) / \sum_k n_i(k, \alpha)$). The average number of times that the group x appears in the position

₂₂₉ α $(n_i(x, \alpha))$ was calculated with a LMM for each position, where the response variable was the number of

₂₃₀ times that a group was observed at a given position in each real-world network, the explanatory variable

₂₃₁ was the group identifier, and the random intercept was given by network identifiers nested within the study

₂₃₂ identifiers.

₂₃₃ Once we obtained p_i^O and p_i^E, we used a simulation approach to determine whether the former is likely

234 to come from the latter or not. This approach was preferred since the large number of possible subgraph
235 combinations and the small probabilities for some of them advise against using an exact test of goodness-of-fit
236 or a Chi-square one. Specifically, we created 1,000 random samples with repetition of possible subgraph
237 combinations, where each sample contained 10 million elements and, for each combination, the probability of
238 being selected was equal to its expected probability. From those random samples, we extracted the mean
239 and the standard deviation of the expected probability of i, denoted as $E\left[p_i^E\right]$ and $\sigma\left[p_i^E\right]$, respectively,
240 and calculated the z-scores of p_i^O as $z_i^O = \left(p_i^O - E\left[p_i^E\right]\right)/\sigma\left[p_i^E\right]$, for those motif combinations with $p_i^O > 0$.
241 According to the usual interpretation of z-scores, combinations with $z_i^O > 1.96$ are over-represented, whereas
242 those with $z_i^O < -1.96$ are under-represented, at the 95% confidence level. Notice that we focused on
243 combinations with $p_i^O > 10^{-7}$ because we do not have enough numerical resolution to accurately detect
244 whether combinations that took place only once are under-represented (due to the limited size of our random
245 samples).

RESULTS

Plant and floral visitor groups

248 The hierarchical cluster analysis divided the dataset into five different groups with different and overlapping
249 characteristics (Figs S4 and S5). The first group referred to as 'selfing herbs' consisted of herbs with
250 hermaphroditic flowers with high levels of autonomous selfing. The second group named 'small outcrossing
251 perennials' had small perennial species with a mix of life forms (i.e., trees shrubs and herbs) with outcrossing
252 hermaphroditic flowers. The third group referred to as 'self-incompatible perennials with large flowers'
253 comprised perennial species with a mixed of life forms and large self-incompatible hermaphroditic flowers
254 with high number of ovules. The fourth group named 'tall plants with small unisexual flowers' had the
255 tallest species, highest proportion of shrub and tree life forms, dioecious and monoecious breeding systems,
256 small flowers and the highest numbers of flowers and inflorescences per plant. Finally, the fifth group named
257 'short-lived outcrossers with long zygomorphic flowers' consisted of small perennial and short-lived herbs with
258 long self-compatible zygomorphic flowers that were unable to self-pollinate.

259 In total, there were 1,126 species of floral visitors with 6,325 interactions recorded with plants. Most
260 plants interacted with bees (2,256 interactions) and non-syrphid Diptera (1,768) followed by syrphids (845),
261 Lepidoptera (437), Coleoptera (432) and non-bee Hymenoptera (362).

Overall network subgraphs patterns

263 Most network subgraphs were under- or over-represented (close to the 1st and 99th percentile, respectively)

9

in the comparison between empirical and simulated networks (Fig. 2). Network subgraphs 3, 5, 9, 10 and 14 were under-represented in empirical networks, that is, all were close to the 1st percentile and under the 25th percentile. Interestingly, four out of five of these motifs belonged to the largest path length classification (i.e., core-peripheral). In addition, network subgraphs 2, 6, 7, 16 and 17 were over-represented, all over the 75th percentile. In contrast to the under-represented motifs, over-represented motifs belonged to the two shortest path length groups (i.e., complete and fan). The remaining network subgraphs (i.e., 4, 8, 11, 12, 13, and 15) were between the 25th and 75th percentile.
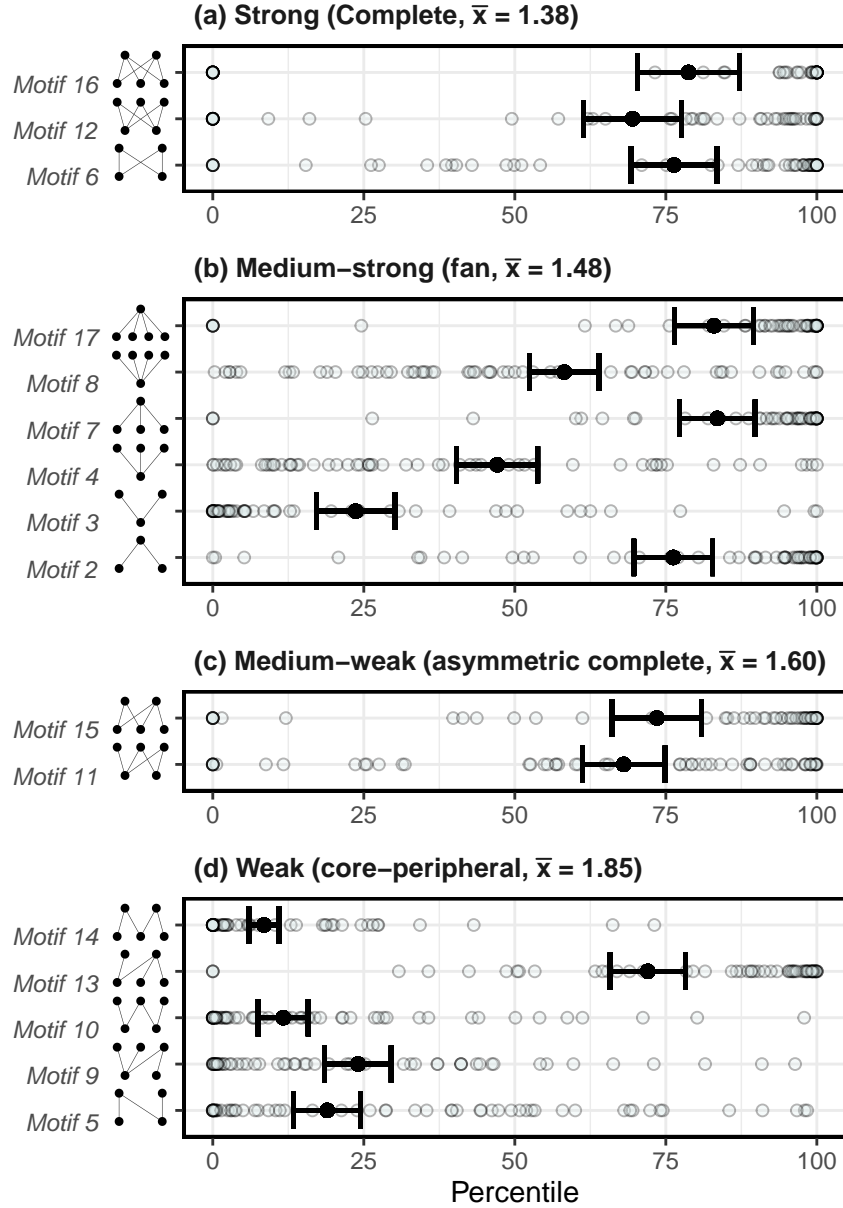
Figure 2. Comparison of network subgraph frequencies between empirical and simulated networks grouped by average path length (plots a, b, c and d) where $\overline{x}$ indicates the mean of the average path lengths of all subgraphs within each group (strong, medium-strong, medium-weak and weak). This comparison is shown with the mean percentage of subgraph frequencies in empirical networks that were over the subgraph frequencies of the simulated ones (percentiles). This was done by network (light blue transparent dots) and then averaged for all networks (black dots with error bars that correspond to the standard deviation).

Over- and under-represented groups on subgraph positions

The comparison of the plant and floral visitor group frequencies per subgraph position between empirical and simulated networks showed a 6.36% and 25.38% of over- ($>75^{th}$) and under-represented ($<$25th) groups in the different positions, respectively (Fig. 3). Floral visitors showed a total of 4.55% and 32.58% of over- and under-represented groups in the different positions and plants 8.18% and 18.18% of over- and

11

under-represented groups, respectively. Notably, the differences across groups were more marked for floral visitors than for plant groups (the differences between min and max percentiles per position were generally two-three times larger for floral visitors). From most over- to under-represented floral visitors groups on the different subgraph positions (indicated with the dendrogram order in Fig. 3), we found: bees, non-syrphids Diptera, syrphids, Coleoptera, non-bee Hymenoptera and Lepidoptera. Although plant functional groups showed less differences between them, there were also more represented groups than others, thus from most over- to under-represented groups on the different subgraph positions we found: self incompatible perennials with large flowers, small outcrossing perennials, tall plants with unisexual flowers, selfing herbs and short lived outcrossers with long zygomorphic flowers.

For both floral visitor and plant groups we found over- and under-representation on different subgraph positions that implied different ecological roles. First, for floral visitors, we found that bees were less frequent on subgraph positions that involved specialization (i.e., lower specificity values) while the rest of taxonomic groups (non-syrphid Diptera, syrphids, Lepidoptera, Coleoptera and non-bee Hymenoptera) were infrequent in generalised subgraph positions with low number of indirect interactions (Fig. S6 A). Although plants did not show differences with null models on their frequencies, all plant but tall plants with small unisexual flowers showed a slight tendency to be over-represented on specialised subgraph positions (Fig. S6 B). Second, regarding the number of indirect interactions, bees were over-represented in subgraph positions with low number of indirect interactions (Fig. S7 A) and the rest of floral visitor groups showed no differences across positions in comparison with simulated networks. For plants, all groups showed a slight trend to be over-represented on subgraph positions with higher number of indirect interactions (Fig. S7 B). When we further explore the observed frequencies by number of indirect interactions for floral visitor and plant groups, we found that both bees and tall plants with small unisexual flowers were the only groups with high frequencies on subgraph positions associated with low number interactions in comparison with the other floral visitor and plant groups (Figs S8 and S9).
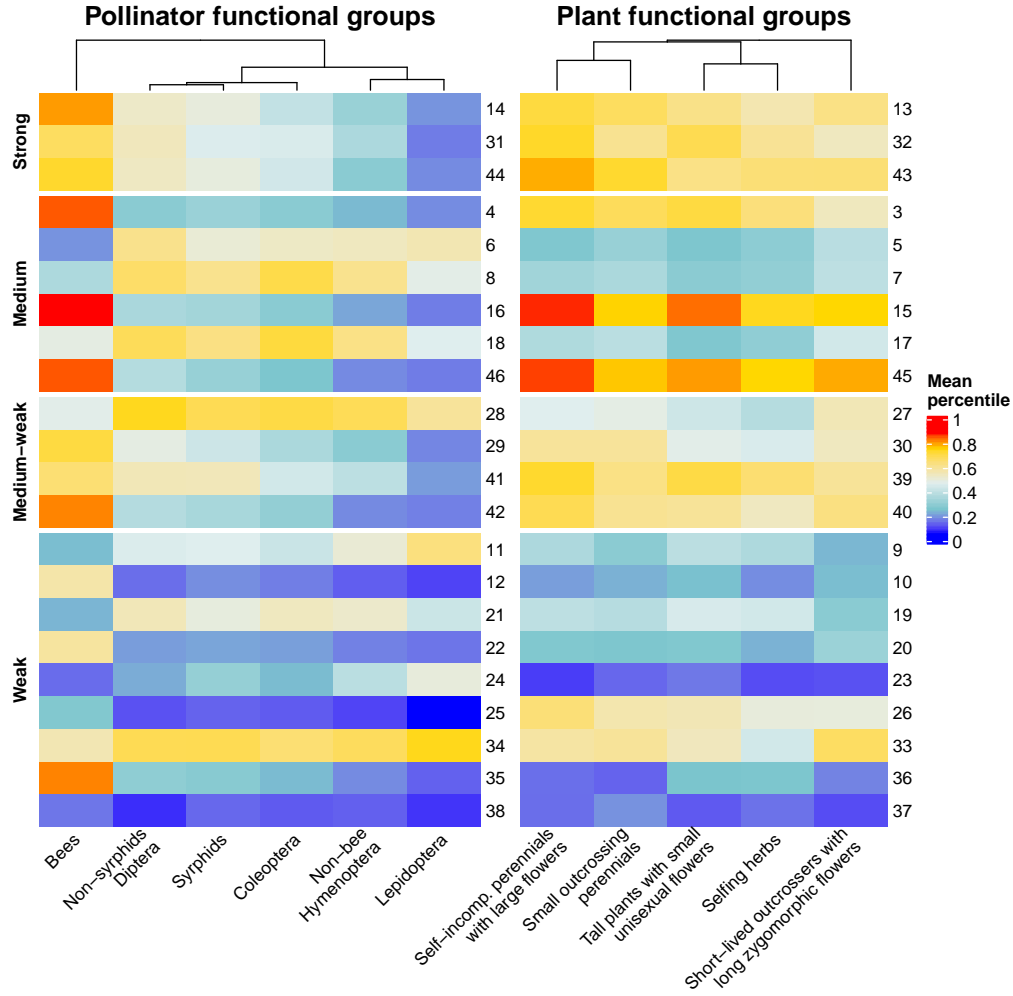
Figure 3. Heatmap indicating under- and over-representation of floral visitor and plant groups in the different subgraph positions (numbers on the right of each heatmap). The different subgraph positions are dividied by the average path length clasification determined by Simmons et al. (2020) and they are ordered from the most densely connected network structures (strong or complete) to the less connected ones (weak or core-peripheral). The superior dendrogram indicates the differences across groups with the more separated groups showing larger differences.

Over- and under-represented group combinations of subgraphs

The different observed frequencies on the different subgraph positions showed a hierarchical order of probabilities for floral visitors (Fig. 4). That is, the taxonomic group of 'bees' was the most abundant one on all subgraph positions followed by non-syrphid Diptera which was after bees, the second most frequent group on all positions. Then, the group of syrphids was the most abundant in all positions but two (positions 28 and 31 from motifs 11 and 12, respectively). After syrphids, Coleoptera was the most frequent group on the different positions followed by non-bee Hymenoptera and finally, Lepidoptera which was the less frequent group on all positions but two (positions 12 and 22 from motifs 5 and 9, respectively). Remarkably, plant functional groups were more variable on the different subgraph positions. However, there were also

13

predominant functional groups found across the different subgraph positions. The three most frequent groups were 'tall plants with small unisexual flowers,' 'self-incompatible perennials with large flowers' and 'small outcrossing perennials.' On the contrary, the groups of 'selfing herbs' and 'short lived outcrossers with long zygomorphic flowers' had the lowest probability to be present on the different subgraph plant positions.
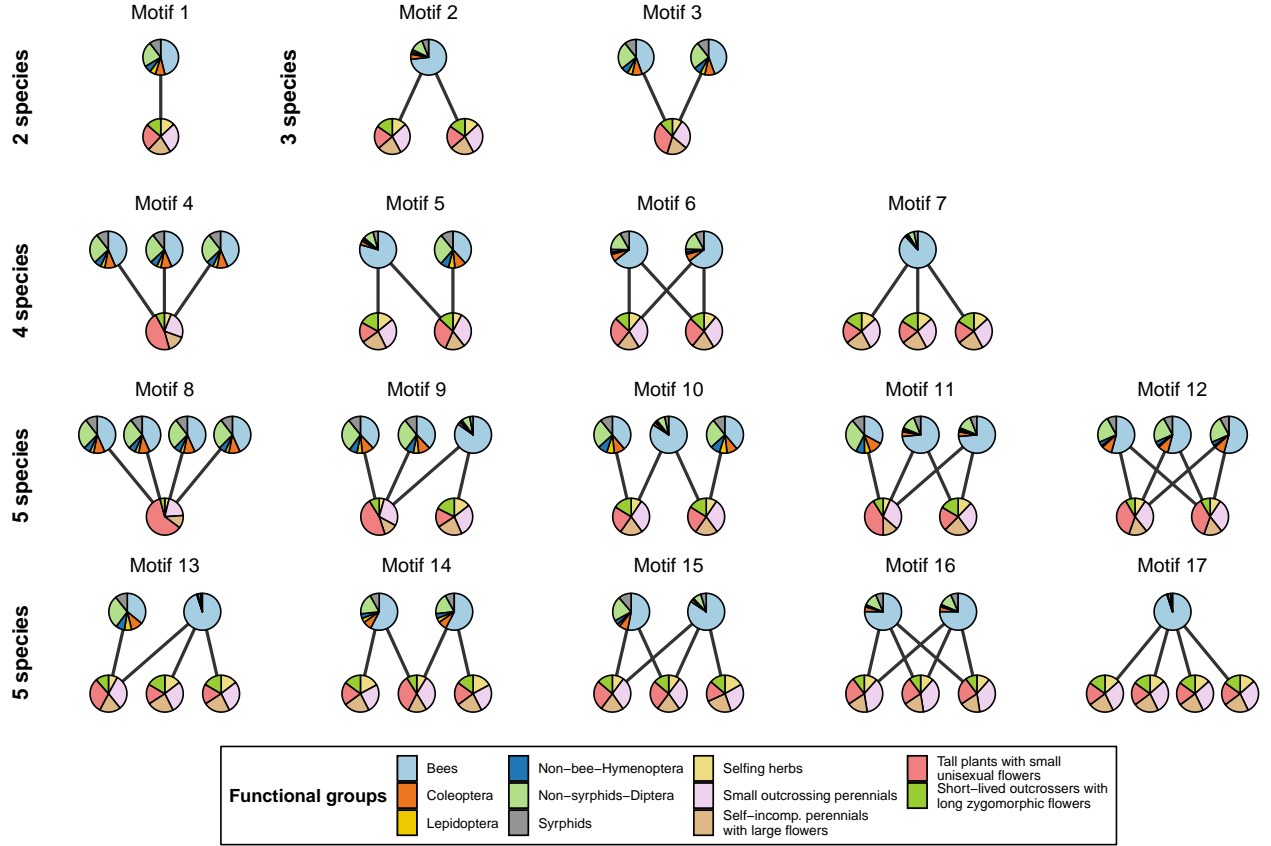

Figure 4. Graphical representation of the probability of finding a given plant and floral visitor group x in the position α of motif i, $p_i(x, \alpha)$, for all the possible network subgraphs from two to five species in bipartite networks. The slices in the nodes for a given functional group x are proportional to the corresponding value of $p_i(x, \alpha)$.

The statistical comparison with z-scores between the observed and expected plant and floral visitor group combinations showed that we can not recover the observed combinations from its probability of occurrence. In fact, most combinations are under- or over- represented and follow a modified Gaussian distribution (Johnson's S_U) with 65% of under-represented group combinations, 20% of no statistical difference and 15% of under-represented ones (Fig. 5). In addition, subgraphs with small node combinations (i.e., 2 and 3 nodes) only appeared as under-represented. Network subgraphs with 4 and 5 node combinations appeared in the three statistical categories but 4 node subgraphs had a higher proportion in the under-represented category (10% in comparison with the 1% and the 3% of the no statistical difference and over-represented categories, respectively). Finally, 5 node subgraphs were in similar proportions in the no statistical difference

³²² and over-represented categories (99% and 97%, respectively) but they appeared in lower proportion in the

³²³ under-represented category (89%). When we explored the identity on the different positions of the most

³²⁴ probable subgraph combinations we found that the most common plant and floral visitor groups were also

³²⁵ the most probable on these subgraph combinations (lower panel Fig. 5).
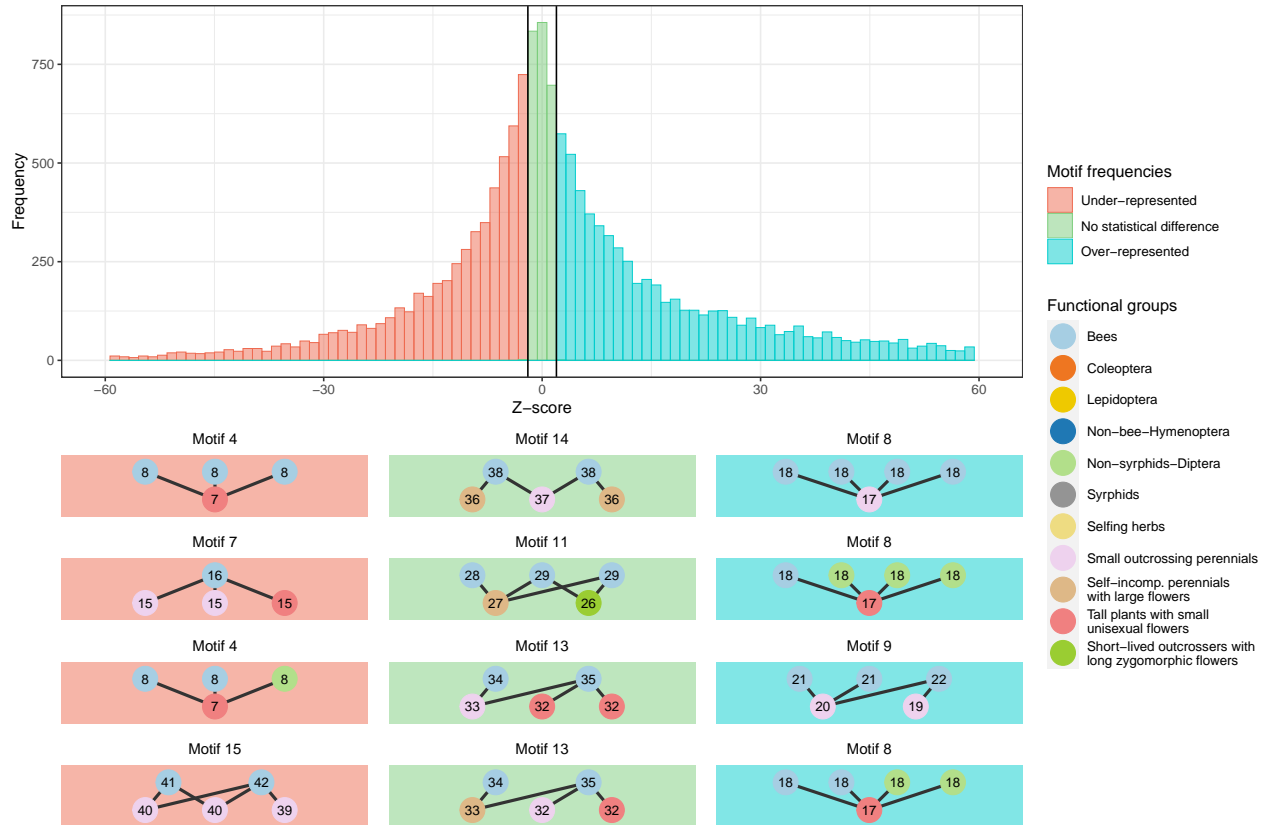

Figure 5. Distribution of z-scores values after the statistical comparison of the observed probabilities of finding a subgraph combination in real networks and the expected probabilities (obtained by computer simulations). The resulting z-scores frequencies of this statiscal comparison are coloured based on a 95% confidence level: under-represented ($z_i^O < -1.96$, red), no statistical difference ($-1.96 \geq z_i^O \leq 1.96$, green) and over-represented ($z_i^O > 1.96$, blue). For instance, for a z-score value of 0 there were approximately 800 subgraph combinations that showed no statistical difference in the comparison between real and simulated networks. In addition, four specific graphical examples of the most probable subgraphs being observed on each category are also provided and the different plant and floral visitor groups are represented with distinct colour codes as indicated in the legend.

Discussion

³²⁷ Understanding the global patterns of network structure is key to unravel the processes that govern community

³²⁸ dynamics (Bascompte & Jordano, 2007; Bastolla et al., 2009; Guimarães, 2020). However, common analytical

³²⁹ approaches at the node or at the full network level are unable to capture the full complexity of ecological

³³⁰ systems (Bergamo et al., 2021; Simmons, Cirtwill, et al., 2019). Because network subgraphs can capture

both direct and indirect interactions, by analysing their over- and under-representation we were able to explore in great detail the structure of interactions between animal pollinators and the plants they visit. We have explored for the first time the overall structure and presence of network motifs in real plant-pollinator networks and shown that the different subgraphs up to five nodes are consistently over- or under-represented in comparison with simulated networks. Despite the presence of multiple ecological processes (e.g., neutral effects, morphological matching or phenological overlap) shaping the meso-scale, we found a tendency for over-represented motifs that were more densely connected (i.e., complete, fan and asymmetric complete). As suggested in Simmons et al. (2020), a higher proportion of these densely connected motifs could reflect that neutral effects are the predominant ecological process governing the structure of the meso-scale. Thus, as indicated for other network properties like nestedness at the macro-scale level (Krishna et al., 2008; Suweis et al., 2013), or interaction probability at the species level (Bartomeus et al., 2016), species relative abundances (i.e., neutral effects) may be also the main driver governing this intermediate level of complexity on plant-pollinator networks.

Overall, generalist network motifs for floral visitors were consistently over-represented (e.g., fan-type subgraphs; Fig. 2). For instance, when we compare network subgraphs with equal structure for both trophic levels, we found that floral visitors had always over-representation in the positions that implied the highest number of direct links (e.g., positions 16 and 46 from network subgraphs 7 and 17, respectively). Despite this being in good agreement with the current view of the generalist nature of pollinators (Olesen & Jordano, 2002; Ollerton, 2017; Waser et al., 1996), we found that this over-representation of generalist network motifs was driven by the taxonomic group of bees while the rest of the taxonomic groups tended to be under-represented (Fig. S6 A). In contrast, non-bee groups were over-represented in subgraph positions associated with specialised roles. This large effect of bees on the overall patterns is no surprise, as bees were the group with the largest number of interaction records in the whole set of networks as also shown for other broad scale plant-pollinator network studies (Carvalheiro et al., 2014; Ollerton, 2017). Interestingly, bees also were the group with lower number of indirect interactions while the rest of floral visitor groups showed clearly higher number of indirect interactions indicating that these non-bee floral visitor groups may experience higher competitive interactions for floral resources (Thomson & Page, 2020). Despite plant functional groups showed little differences between them on their subgraph frequencies, all groups showed a tendency to be over-represented on subgraph positions that were associated with higher levels of specialisation. This tendency towards specialist roles may be the result of co-evolutionary processes (e.g., trait-matching) between plants and pollinators to optimise pollen transfer (Moreira-Hernández & Muchhala, 2019). Remarkably, only tall plants with small unisexual flowers showed higher frequencies on subgraphs that were associated with low number of indirect interactions with

likely direct implications on pollen transfer for this group that needs to guarantee cross-pollination.

The probabilities of appearance of each functional group on the different subgraph positions were insufficient to predict most of the observed network position combinations (Fig. 5). This implies that not just the representation of network subgraphs types but also the identity of the elements of their structure is non-random. Notably, we found that just 30% of the possible 53,250 subgraph combinations are realized. Thus, similarly to the presence of forbidden links at the network level (Jordano, 2016; Olesen et al., 2011), we also found the existence of forbidden subgraph positions and those represented almost two thirds of the total possible subgraph combinations. This concurs well with the general idea that plant-pollinator networks have a high proportion of non-realized direct interactions (Bascompte & Jordano, 2013; Jordano et al., 2003), and as a result, also a high proportion of non-realized combinations of indirect interactions. Some of this non-observed direct interactions are probably due to under-sampling (Dorado et al., 2011; Jordano, 2016). However, a large fraction of the possible interactions do not take place because of life-history constraints such as species co-occurrence, trait matching processes and phenological uncoupling (Bartomeus et al., 2016; Jordano, 2016; Olesen et al., 2011). Despite the advances in the understanding on the ecological processes that influence species pairwise interactions (Bartomeus et al., 2016; Peralta et al., 2020), we still know little about how these processes determine competitive and facilitative interactions in mutualistic networks. Thus, as suggested by Simmons et al. (2020), exploring the proportion of these realized network subgraphs that are competitive or facilitative is a necessary next step to improve our understanding of the architecture of biodiversity (Bascompte & Jordano, 2007).

Despite the obvious usefulness of functional groups in summarizing species ecological roles (e.g., similar reproductive biology for plants and behaviour for pollinators), delimiting species into functional groups is often not trivial. For instance, taxonomic diversity and trait availability limit our ability to properly represent the wide range of reproductive strategies (Dellinger, 2020) and different clustering analysis can lead to distinct outcomes (e.g., k-means and hierarchical clustering). Further, here we divided floral visitors on the taxonomic rank level because it represented adequately pollinator ecological roles without the tremendous effort of compiling trait data for over 1000 species of floral visitors. However, further groupings with traits could help revealing greater relevance of trait-matching processes on indirect interactions. We believe that this could be investigated by exploring which proportion of over- and under-represented motifs present high and low trait-matching, respectively. Moreover, we analysed network subgraphs focusing exclusively on presence-absence interactions which assume that all interactions are equally important as suggested by Simmons et al. (2020), further work should investigate how to account for the quantitative measurements of the strength of plant-pollinator interactions. Finally, despite the wide geographical coverage of our plant-pollinator dataset,

17

there are major gaps for some regions of the world with a clear bias towards Europe, America and islands which have long interested ecologist (MacArthur & Wilson, 1963; Warren et al., 2015). These biases are also found in global datasets of networks where the data is highly fragmented across space (Poisot et al., 2021) and more efforts are needed to capture global patterns of plant-pollinator interactions.

In conclusion, by following the framework of Simmons et al. (2020) we have been able to explore competitive and facilitative interactions in plant-pollinator networks worldwide and found the presence of worldwide motifs that highlight the non-random structure of indirect interactions. Further, we found that plant and floral visitor groups are associated with different subgraph positions that involve distinct ecological roles (e.g., generalisation versus specialisation) and that the different subgraph combinations are strongly constrained. Although we are still far to predict how species interact, our work is a first step to elucidate beyond pairwise interactions.

Acknowledgements

Author contributions

IB, AA-P and JBL designed the idea. JBL collected the data. AA-P lead the analysis with contributions of JBL and guidance of IB. JBL wrote the manuscript with contributions of IB and AA-P.

Data Availability Statement

All data and code are publicly available at Github (https://github.com/JoseBSL/FunctionalMotifs) and Zenodo (https://doi.org/10.5281/zenodo.6489213).

References

Allen-Perkins, A., Hurtado, M., Garcia-Callejas, D., Godoy, O., & Bartomeus, I. (2021). Community level individual-based plant-pollinator networks unveil pollen flow dynamics and plant reproductive success. *BioRxiv*.

Baker, N. J., Kaartinen, R., Roslin, T., & Stouffer, D. B. (2015). Species' roles in food webs show fidelity across a highly variable oak forest. *Ecography*, *38*(2), 130–139.

Bartomeus, I., Gravel, D., Tylianakis, J. M., Aizen, M. A., Dickie, I. A., & Bernard-Verdier, M. (2016). A common framework for identifying linkage rules across different types of interactions. *Functional Ecology*, *30*(12), 1894–1903.

Bascompte, J., & Jordano, P. (2007). Plant-animal mutualistic networks: The architecture of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *38*, 567–593.

Bascompte, J., & Jordano, P. (2013). *Mutualistic networks*. Princeton University Press.

Bascompte, J., Jordano, P., Melián, C. J., & Olesen, J. M. (2003). The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences*, *100*(16), 9383–9387.

Bastolla, U., Fortuna, M. A., Pascual-García, A., Ferrera, A., Luque, B., & Bascompte, J. (2009). The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature*, *458*(7241), 1018–1020.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bergamo, P. J., Traveset, A., & Lázaro, A. (2021). Pollinator-mediated indirect effects on plant fecundity revealed by network indices. *The American Naturalist*, *198*(6), 734–749.

Blüthgen, N., Menzel, F., & Blüthgen, N. (2006). Measuring specialization in species interaction networks. *BMC Ecology*, *6*(1), 9.

Breheny, P., Burchett, W., & Breheny, M. P. (2020). Package 'visreg.' *R Package Version 2.7.0*. https://CRAN.R-project.org/package=visreg

Carvalheiro, L. G., Biesmeijer, J. C., Benadi, G., Fründ, J., Stang, M., Bartomeus, I., Kaiser-Bunbury, C. N., Baude, M., Gomes, S. I. F., Merckx, V., Baldock, K. C. R., Bennett, A. T. D., Boada, R., Bommarco, R., Cartar, R., Chacoff, N., Dänhardt, J., Dicks, L. V., Dormann, C. F., . . . Kunin, W. E. (2014). The potential for indirect effects between co-flowering plants via shared pollinators depends on resource abundance, accessibility and relatedness. *Ecology Letters*, *17*(11), 1389–1399.

Chamberlain, S., Szoecs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., Bartomeus, I., Baumgartner, J., O'Donnell, J., Oksanen, J., Tzovaras, B. G., Marchand, P., Tran, V., Salmon, M., Li, G., & Grenié,

M. (2020). Taxize: Taxonomic information from around the web. *R Package Version 0.9.99.* https://CRAN.R-project.org/package=taxize

Cirtwill, A. R., Roslin, T., Rasmussen, C., Olesen, J. M., & Stouffer, D. B. (2018). Between-year changes in community composition shape species' roles in an arctic plant–pollinator network. *Oikos, 127*(8), 1163–1176.

Coux, C., Rader, R., Bartomeus, I., & Tylianakis, J. M. (2016). Linking species functional roles to their network roles. *Ecology Letters, 19*(7), 762–770.

Dellinger, A. S. (2020). Pollination syndromes in the 21st century: Where do we stand and where may we go? *New Phytologist, 228*(4), 1193–1213.

Devaux, C., Lepers, C., & Porcher, E. (2014). Constraints imposed by pollinator behaviour on the ecology and evolution of plant mating systems. *Journal of Evolutionary Biology, 27*(7), 1413–1430.

Dolnicar, S., Grün, B., Leisch, F., & Schmidt, K. (2014). Required sample sizes for data-driven market segmentation analyses in tourism. *Journal of Travel Research, 53*(3), 296–306.

Dorado, J., Vázquez, D. P., Stevani, E. L., & Chacoff, N. P. (2011). Rareness and specialization in plant–pollinator networks. *Ecology, 92*(1), 19–25.

Dormann, C. F., Frueund, J., Bluethgen, N., & Gruber, B. (2009). Indices, graphs and null models: Analyzing bipartite ecological networks. *The Open Ecology Journal, 2*, 7–24.

Fenster, C. B., Armbruster, W. S., Wilson, P., Dudash, M. R., & Thomson, J. D. (2004). Pollination Syndromes and Floral Specialization. *Annual Review of Ecology, Evolution, and Systematics, 35*(1), 375–403.

Földesi, R., Howlett, B. G., Grass, I., & Batáry, P. (2021). Larger pollinators deposit more pollen on stigmas across multiple plant species—a meta-analysis. *Journal of Applied Ecology, 58*(4), 699–707.

Gong, Y.-B., & Huang, S.-Q. (2009). Floral symmetry: Pollinator-mediated stabilizing selection on flower size in bilateral species. *Proceedings of the Royal Society B: Biological Sciences, 276*(1675), 4013–4020.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 857–871.

Greenleaf, S. S., Williams, N. M., Winfree, R., & Kremen, C. (2007). Bee foraging ranges and their relationship to body size. *Oecologia, 153*(3), 589–596.

Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics, 32*(18), 2847–2849.

Guimarães, P. R. J. (2020). The structure of ecological networks across levels of organization. *Annual Review of Ecology, Evolution, and Systematics, 51*, 433–460.

Hartig, F., & Hartig, M. F. (2017). Package 'DHARMa.' *R Package Version 0.3.3.0.* https://CRAN.R-project.org/package=DHARMa

Holt, R. a., & Lawton, J. (1994). The ecological consequences of shared natural enemies. *Annual Review of Ecology and Systematics*, *25*(1), 495–520.

Jordano, P. (1987). Patterns of mutualistic interactions in pollination and seed dispersal: Connectance, dependence asymmetries, and coevolution. *The American Naturalist*, *129*(5), 657–677.

Jordano, P. (2016). Sampling networks of ecological interactions. *Functional Ecology*, *30*(12), 1883–1893.

Jordano, P., Bascompte, J., & Olesen, J. M. (2003). Invariant properties in coevolutionary networks of plant–animal interactions. *Ecology Letters*, *6*(1), 69–81.

Krishna, A., Guimaraes Jr, P. R., Jordano, P., & Bascompte, J. (2008). A neutral-niche theory of nestedness in mutualistic networks. *Oikos*, *117*(11), 1609–1618.

Laliberté, E., Legendre, P., Shipley, B., & Laliberté, M. E. (2014). Package 'FD.' Measuring functional diversity from multiple traits, and other tools for functional ecology. *R Package Version 1.0-12.* https://CRAN.R-project.org/package=FD

Lanuza, J. B., Rader, R., Stavert, J., Kendall, L. K., Saunders, M. E., & Bartomeus, I. (2021). Trade-offs among plant reproductive traits determine interactions with floral visitors. *BioRxiv*.

Lázaro, A., Gómez-Martínez, C., Alomar, D., González-Estévez, M. A., & Traveset, A. (2020). Linking species-level network metrics to flower traits and plant fitness. *Journal of Ecology*, *108*(4), 1287–1298.

MacArthur, R. H., & Wilson, E. O. (1963). An equilibrium theory of insular zoogeography. *Evolution*, *17*(4), 373–387.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science*, *298*(5594), 824–827.

Moeller, D. A. (2004). Facilitative interactions among plants via shared pollinators. *Ecology*, *85*(12), 3289–3301.

Moreira-Hernández, J. I., & Muchhala, N. (2019). Importance of pollinator-mediated interspecific pollen transfer for angiosperm evolution. *Annual Review of Ecology, Evolution, and Systematics*, *50*, 191–217.

Olesen, J. M., Bascompte, J., Dupont, Y. L., Elberling, H., Rasmussen, C., & Jordano, P. (2011). Missing and forbidden links in mutualistic networks. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1706), 725–732.

Olesen, J. M., Bascompte, J., Dupont, Y. L., & Jordano, P. (2007). The modularity of pollination networks. *Proceedings of the National Academy of Sciences*, *104*(50), 19891–19896.

Olesen, J. M., & Jordano, P. (2002). Geographic Patterns in Plant–Pollinator Mutualistic Networks. *Ecology*, *83*(9), 2416–2424.

Ollerton, J. (2017). Pollinator diversity: Distribution, ecological function, and conservation. *Annual Review of Ecology, Evolution, and Systematics*, *48*, 353–376.

514 Ollerton, J., Alarcón, R., Waser, N. M., Price, M. V., Watts, S., Cranmer, L., Hingston, A., Peter, C. I., &

515 Rotenberry, J. (2009). A global test of the pollination syndrome hypothesis. *Annals of Botany*, *103*(9),

516 1471–1480.

517 Peralta, G., Vázquez, D. P., Chacoff, N. P., Lomáscolo, S. B., Perry, G. L., & Tylianakis, J. M. (2020).

518 Trait matching and phenological overlap increase the spatio-temporal stability and functionality of

519 plant–pollinator interactions. *Ecology Letters*, *23*(7), 1107–1116.

520 Poisot, T., Bergeron, G., Cazelles, K., Dallas, T., Gravel, D., MacDonald, A., Mercier, B., Violet, C., &

521 Vissault, S. (2021). Global knowledge gaps in species interaction networks data. *Journal of Biogeography.*

522 Poisot, T., Kéfi, S., Morand, S., Stanko, M., Marquet, P. A., & Hochberg, M. E. (2015). A continuum of

523 specialists and generalists in empirical communities. *PloS One*, *10*(5), e0114674.

524 R Core Team. (2021). R: A language and environment for statistical computing. In *R Foundation for*

525 *Statistical Computing.* https://www.R-project.org/

526 Sargent, R. D., & Ackerly, D. D. (2008). Plant–pollinator interactions and the assembly of plant communities.

527 *Trends in Ecology & Evolution*, *23*(3), 123–130.

528 Simmons, B. I., Beckerman, A. P., Hansen, K., Maruyama, P. K., Televantos, C., Vizentin-Bugoni, J., &

529 Dalsgaard, B. (2020). Niche and neutral processes leave distinct structural imprints on indirect interactions

530 in mutualistic networks. *Functional Ecology.*

531 Simmons, B. I., Cirtwill, A. R., Baker, N. J., Wauchope, H. S., Dicks, L. V., Stouffer, D. B., & Sutherland,

532 W. J. (2019). Motifs in bipartite ecological networks: Uncovering indirect interactions. *Oikos*, *128*(2),

533 154–170.

534 Simmons, B. I., Sweering, M. J., Schillinger, M., Dicks, L. V., Sutherland, W. J., & Di Clemente, R. (2019).

535 Bmotif: A package for motif analyses of bipartite networks. *Methods in Ecology Evolution*, *10*(5), 695–701.

536 Stouffer, D. B., Sales-Pardo, M., Sirer, M. I., & Bascompte, J. (2012). Evolutionary conservation of species'

537 roles in food webs. *Science*, *335*(6075), 1489–1492.

538 Strauss, S. Y. (1991). Indirect effects in community ecology: Their definition, study and importance. *Trends*

539 *in Ecology & Evolution*, *6*(7), 206–210.

540 Suweis, S., Simini, F., Banavar, J. R., & Maritan, A. (2013). Emergence of structural and dynamical

541 properties of ecological mutualistic networks. *Nature*, *500*(7463), 449–452.

542 Thomson, D. M., & Page, M. L. (2020). The importance of competition between insect pollinators in the

543 anthropocene. *Current Opininion in Insect Science*, *38*, 55–62.

544 Tur, C., Castro-Urgal, R., & Traveset, A. (2013). Linking plant specialization to dependence in interactions

545 for seed set in pollination networks. *PloS One*, *8*(10), e78294.

546 Warren, B. H., Simberloff, D., Ricklefs, R. E., Aguilée, R., Condamine, F. L., Gravel, D., Morlon, H., Mouquet,

547   N., Rosindell, J., Casquet, J.others. (2015). Islands as model systems in ecology and evolution: Prospects

548   fifty years after MacArthur-wilson. *Ecology Letters*, *18*(2), 200–217.

549 Waser, N. M., Chittka, L., Price, M. V., Williams, N. M., & Ollerton, J. (1996). Generalization in pollination

550   systems, and why it matters. *Ecology*, *77*(4), 1043–1060.

551 White, D., & Gramacy, R. (2009). Maptree: Mapping, pruning, and graphing tree models. *R Package Version*

552   *1.4-5.* https://CRAN.R-project.org/package=maptree