APPENDIX F

Null Models for Assessing Network Structure

Looking for network patterns requires having a benchmark to asses the significance of a meaningful variable such as nestedness or modularity. The degree of absolute nestedness, for example, depends on a long list of properties, such as network size, number of interactions, connectivity distribution, and so on. In this regard, research on network structure depends very much on appropriate null models. A null model is a randomization that deliberately excludes all biological mechanisms to test whether the observed level of structure (e.g., nestedness or modularity) can be explained just out of chance using a derived probability of cell occupancy (Gotelli, 2001; Gotelli and McCabe, 2002). This has been the subject of a broad debate in island biogeography, from which research on ecological networks has benefited.

For example, Diamond (1975) published pioneering work on the patterns of distribution of birds across the Bismarck Archipelago, arguing that interspecific competition was a driving force of the observed patterns of co-occurrences between species. However, with a finite and noisy sample, a pattern may be produced by chance, and inferring competition (or any other biological mechanism) may be misleading. How do we decide? The trick is to calculate how many times we get the observed pattern by chance. Only if we never obtain this pattern by chance or it happens in less than 5% of the randomizations, we can conclude that there is some meaningful mechanism underlying our data. Needless to say, null models can never point out this mechanism. However, null models make some explicit assumptions about what they preserve and have distinct statistical behaviors; inferences on the real data may largely depend on the properties of the null model used.

Following this rationale, Diamond's finding of assembly rules was questioned by Connor and Simberloff (1979). These authors claimed that many of Diamond's previous results attributed to competition could arise by chance. This debate—kept alive for about two decades—was settled more recently by

Gotelli and McCabe, who approached the problem by using appropriate null models. Their findings showed that in the majority of communities studied, there was a significantly lower number of species co-occurrence than expected by chance, therefore supporting Diamond's assembly-rule mechanism (Gotelli and McCabe, 2002).

Similar examples of this debate about whether the significance of a result depends on the type of null model arose in the context of the measures of nestedness in island biogeography. As noted in Appendix C, the nestedness temperature calculator (NTC) is not insensitive to matrix size, shape, and filling, so we need to put it in perspective. The NTC uses a null model in which each cell of the interacting matrix has the same probability of having a 1. This probability is estimated by the connectivity of the matrix (fraction of all cells occupied). The problem is that this null model has a large type I error, that is, it mistakenly rejects a true null hypothesis. Consider that this null model tends to distribute on average the same number of interactions per species. We already know that the distribution of interactions per species in mutualistic networks is quite heterogeneous (some species interact only with one or a few other species, whereas a few species are supergeneralists). Thus, it would be easy to find statistical differences between the observed matrix and the randomizations, but it would probably be wrong to adduce nestedness for this difference. There are several potential explanations. If we want to focus on nestedness, then we have to control for these differences in the number of interactions per species. This problem has been widely discussed in the context of island biogeography, and several alternative null models have been introduced (Roberts and Stone, 1990; Sanderson, Moulton, et al., 1998; Cook and Quinn, 1998; Fischer and Lindenmayer, 2002; Rodríguez-Gironés and Santamaría, 2006; Ulrich and Gotelli, 2007; Miklós and Podani, 2004; Gotelli and Entsminger, 2001).

For example, Fischer and Lindenmayer (2002) developed a null model where cells were occupied on the basis of the actual probability of encountering a given species in the real data. A species observed in 60% of the sites has a probability of 0.6 of occupying any site in the randomization. Thus, occupancy probabilities varied across species but were kept constant across sites for a given species (Fischer and Lindenmayer). A similar strategy for maintaining the observed interspecies variability but in a deterministic way (i.e., exactly fixing the number of interactions per species rather than using this information to calculate the probability) was used by Prado and Lewinsohn (2004). These authors were testing whether interactions between plants and their insect herbivores were organized in compartments. Their null model reassigned at random the host plants of each insect species, thus preserving

the host breadth of insects but randomizing the identity of their interactors. This null model was originally proposed by Sale (1974).

However, not only does the number of islands occupied by a species change across species, the number of species per island also changes across islands. Thus, why not fix rows instead of columns? Or even better, why not fix rows and columns simultaneously? If, in the latter case, nestedness is significant, we will be sure this definitely speaks about nestedness because all species have the same number of interactions as the real matrix. This sounds great and seems like it should be the end of our quest for the ideal null model. Unfortunately, this is not so easy. There are two main difficulties. First, there is not only a single way to implement a null model fixing both rows and columns simultaneously, and the different implementations lead to considerably different results. Second, fixing both rows and columns imposes huge constraints, so that the randomization becomes very conservative. At the extreme of a perfectly nested matrix, for example, no other matrix rearrangements are allowed (Ulrich and Gotelli). In general, it is the type II error (i.e., incorrectly accepting a false null hypothesis) that becomes very high this time (Ulrich and Gotelli). Thus, the probability of detecting a pattern when it is actually present is very low. We will explain these two problems next.

Fixing exactly the number of interactions of rows and columns has been implemented through two major algorithmic families. One is called the *random-fill* procedure. This approach starts with a matrix containing only 0s and proceeds by sequentially choosing a 0 and replacing it by a 1 as long as the marginal row and column sums do not exceed the observed marginals in the original matrix (Sanderson, Moulton, et al., 1998). One difficulty associated with this method is that the process may get trapped so that no additional 1s can be replaced without failing to preserve the observed marginals. In this case, the algorithm would go backward and try to find a different path to keep filling 1s.

The second major strategy for randomizing a matrix by fixing the number of 1s in both rows and columns is the *swap algorithm*; it starts with the actual matrix and then proceeds by sequentially reshuffling 2×2 submatrices with the same row and column total, as follows:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Some discussion has arisen on whether swap or filling algorithms perform better on the basis of their statistical properties (Sanderson, Moulton, et al., 1998; Gotelli and Entsminger, 2001), with Gotelli and Entsminger concluding that swapping methods are more efficient and reliable.

The situation is more complicated because even if we chose the swap algorithm, there are several implementations. Gotelli and Entsminger (2001) considered two versions of the swap algorithm: sequential swap and independent swap. The difference is that the sequential swap uses a very large number of initial transpositions of the original matrix—let's say 30,000—to ensure that every submatrix has enough chances to be swapped. Next, each subsequent transposition is kept as a different null matrix. In the independent swap, on the other hand, each null matrix is created through a series of independent swaps of the original matrix (Gotelli and Entsminger). Again, results may vary, depending on the method used to implement the swapping procedure. At any rate, both methods have a methodological problem: they do not sample the potential presence-absence matrices uniformly (Miklós and Podani, 2004; Rodríguez-Gironés and Santamaría, 2006). As a consequence, the distribution of any statistic measured over this population of randomizations will be biased because of the dependence of these randomizations on the initial configuration. To try to remedy this problem, Miklós and Podani proposed a third swapping procedure, the *trial swap* (for details see Miklós and Podani, 2004). These authors claim that the procedure leads to a uniform distribution, but its convergence is slow for large matrices. Miklós and Podani finally, suggested a combined approach to overcome this slowness, where a fast initialization is first used, followed by the trial swap. Some concerns, however, have been raised about how efficiently this combination achieves the goal (Rodríguez-Gironés and Santamaría).

The second problem with the null models fixing both the number of 1s simultaneously in both rows and columns has to do with the type I versus type II error. Null models fixing both the observed row and column totals have very low type I errors. This means that when they reject the null hypothesis, we can be strongly convinced that there is, indeed, a significant pattern. The problem is that this low type I error comes at a price: fixed-fixed null models have a high type II error. This means that it is very difficult to detect a pattern such as nestedness even when this pattern is present (Cook and Quinn, 1998; Ulrich and Gotelli, 2007). Specifically, Ulrich and Gotelli made a compelling study of the type I versus type II statistical errors of several randomization schemes. These authors created perfectly nested matrices and then progressively increased randomness. In this scenario, a good null model should detect nestedness, but the most constrained null models only did so in less than 25% of the cases (Ulrich and Gotelli).

On top of the preceding constraints of the fixed-fixed null model, other authors have presented more conceptual concerns. These authors would argue that an observation would be just one realization of a random process if several

replicates were allowed, and, therefore, we should allow marginal totals to vary to some extent (Cook and Quinn, 1998).

The equivalent to fixing both rows and columns is to probabilistically do so. This is the null model 2 in Bascompte, Jordano, et al. (2003). In this case, the probability ρ_{ij} of drawing an interaction in cell w_{ij} is proportional to the generalization level of plant i and animal j. Specifically, this probability can be described as the arithmetic mean of the fraction of occupied cells on row (plant) i (p_i) and on column (animal) j (q_j). This can be written as

$$\rho_{ij} = \frac{p_i + q_j}{2}. \tag{F.1}$$

It can be proven that this null model gives an unbiased estimate of the overall connectivity. The following demonstration is due to Joel E. Cohen (pers. com.) during a visit to our lab.

Consider a matrix with P plant species in rows and A animal species in columns. Each realized element of this matrix will be w_{ij}=1 if plant i and animal j interact and 0 otherwise. $L = \sum_{i=1}^{P} \sum_{j=1}^{A} w_{ij}$ is the total number of interactions (1s) in the matrix. Now, let x_{ij} be the predicted element of row i and column j under null model (F.1). The interaction probabilities of plant i (p_i) and of animal j (q_j) are estimated by

$$p_i = \frac{\sum_{j=1}^{A} w_{ij}}{A}, \qquad q_j = \frac{\sum_{i=1}^{P} w_{ij}}{P}.$$

Then, by definition we have that $\sum_{i=1}^{P} p_i = L/A$ and $\sum_{j=1}^{A} q_j = L/P$. Thus, the expected number of links if $x_{ij} \sim B[(p_i + q_j)/2]$ is

$$E\left(\sum_{i,j} x_{ij}\right) = \frac{1}{2}\sum_{i=1}^{P}\sum_{j=1}^{A}(p_i + q_j) = \frac{1}{2}\left(A\frac{L}{A} + P\frac{L}{P}\right) = \frac{1}{2}2L = L. \tag{F.2}$$

Therefore, the expected number of links under null model (F.1) is the observed number of links.

In a comparative analysis of several probabilistic null models used in interaction networks, Rodríguez-Gironés and Santamaría (2006) concluded that null model (F.1) is the one that performs better in the sense of simultaneously having smaller type I and type II errors. Their clever statistical analysis was based on building a "data matrix" with each one of the three probabilistic null models and, for each, calculating the p-value with each of the three null models. This provided 3×3 comparisons of the null model used to generate

the matrix and the one used to estimate the p-value. Because the original data were already random, the p-values are expected to be uniformly distributed. Deviations from this expectation speak of the tendency of the null model to be more or less conservative, allowing the quantification of both type I and type II errors (Rodríguez-Gironés and Santamaría, 2006).

Perhaps we have to conclude that there is no magic null model, all have pros and cons, and the best strategy as advocated by Melián and Bascompte (2004) is to use a suite of null models. The most robust results are those obtained consistently across a gradient of null models. For example, the original result on the nested structure (Bascompte, Jordano, et al., 2003) and asymmetric specialization (D. P. Vázquez and Aizen, 2004) in mutualistic networks was confirmed with the most conservative null model maintaining both row and column marginals (Joppa, Bascompte, et al. [2009] and Joppa, Montoya, et al. [2010]).

So far, we have focused on the traditional debate of presence-absence null models derived from island biogeography. Within the context of mutualistic networks, other sources of null models have been generated that consider the frequency of interactions.

For example, a recent null model by Blüthgen, Fründ, et al. (2008) treats the interaction matrix as a contingency table, in an effort to generate expected frequencies of interaction. These authors use Patefield's algorithm to estimate the expected frequencies of an $R \times C$ table (treating the interaction matrix as a table of R rows and C columns). The problem is that Patefield's algorithm merely keeps the marginals of the table fixed. So, for any row species with 25 interactions recorded, we can allocate them either in a single cell (with $k = 1$ and $A_{ij} = 25$) or among five cells (with $k = 5$ and $A_{ij} = 5$ for each of the five cells). The algorithm maintains neither the degree distribution nor the overall connectivity of the table. Obviously, this results in inflated estimated interaction frequencies for a large number of cells, especially those in the lower-right corner of the interaction matrix, once the interaction matrix has been reordered by row-column sorting in decreasing order of both the row and column totals. Thus, Blüthgen, Fründ, et al. (2008) use a Kullback-Kleiber distance to obtain a *residual* interaction value that is simply a weighted frequency: the observed interaction frequency minus the expected from the Patefield's algorithm. They find that even for rare species, there are more interactions than expected by the null model and so specialization is more frequent than previously acknowledged. This is not surprising. Patefield's procedure is adequate for statistical analysis of contingency tables that do not necessarily require preserving a given frequency of filled cells (i.e., do not necessarily preserve a k value, as is badly needed when dealing with

an interaction matrix). The null matrices obtained invariably have larger connectance values than real matrices because the null model does not impose any constraint on the k-values (the frequency of nonempty cells, where it allocates a given row-column combination of interactions).

The basic rationale for generating appropriate null models is that abundance is commonly considered as a nuisance factor that needs to be controlled for. Thus, another problem with more elaborate null models attempting to control for species abundance is the circular reasoning underlying them: species abundance is derived from the interaction frequency data to estimate the expected interaction frequencies. This leads to an impossibility to reject the null hypothesis and to the unavoidable conclusion that abundance is all that matters to explain network patterns (Vázquez, 2005; Blüthgen, Fründ, et al., 2008). Species abundance has become an important property of community structure due to the influence of Steve Hubbell's neutral theory of biodiversity (Hubbell, 2001). Abundance is certainly an important component of network patterns, as stressed in early work by Jordano (1987), but we claim that it has to be considered as another ecological correlate or attribute of the species in the interaction network and taken into account simultaneously with other trait variables controlling their effects (see Chapter 4). Subtracting abundance because it is the basis of neutrality is as wrong as subtracting phylogenetic effects in comparative analysis as a way to "control" the undesirable effects of phylogeny. Blüthgen, Fründ, et al. confound the abundance effects with the bias due to sampling, which might itself be the consequence of an incorrectly designed sampling protocol that is influenced by abundance variation among species. The solution when we need to factor out abundance variation across species is to have species-abundance data derived independently of the interaction data itself. This can be done by carrying out separate censuses specifically designed for estimating abundance and separate routine focal observations at plants to record visits.

A most useful approach would be to have robust estimates of sampling bias and to incorporate them in the analysis together with independently derived abundance estimates. Vázquez, Poulin, et al. (2005) used path analysis for the Cold Lake host-parasite interaction dataset and concluded that even when the sampling effect is considered separately from host abundance, the indirect effect of host abundance on parasite richness is substantially high. Their results adequately account for sampling effects and are consistent with the hypothesis that observed patterns of interaction in host–parasite interaction networks result partly from abundance variation among species.