



# Nestedness and null models in ecology

Lara Kaiser

Bachelor thesis (Student ID 3302629)  
submitted to  
the Faculty of Environment & Natural Resources  
at the Albert-Ludwigs-University Freiburg



Supervisor: Prof. Dr. Carsten Dormann, Department of Biometry and Environmental System Analysis  
Co-supervisor: Dr. Michael Staab, Department of Nature

Freiburg, Jan 2015

# Contents

<b>1. Introduction</b>	<b>3</b>
1.1. Motivation . . . . .	3
1.2. Definition and history . . . . .	4
1.3. Cause of nestedness and nestedness as cause . . . . .	6
1.4. Short analysis guide . . . . .	11
1.5. Matrices . . . . .	11
1.6. Metrics . . . . .	12
1.7. Nullmodels . . . . .	16
1.8. Z-score . . . . .	22
1.9. Software . . . . .	22
1.10. Moduled compartments in networks . . . . .	24
1.11. Possible conclusions . . . . .	24
<b>2. Methods</b>	<b>27</b>
2.1. Measures of nestedness and types of null models . . . . .	29
2.2. Statistical Analysis . . . . .	29
<b>3. Results</b>	<b>31</b>
3.1. Statistical Analysis . . . . .	31
3.1.1. Metric correlation . . . . .	31
3.1.2. Null models . . . . .	31
3.1.3. Observed matrices . . . . .	32
3.1.4. Moduled matrices . . . . .	33
3.1.5. ANOVA . . . . .	33
<b>4. Discussion</b>	<b>37</b>
4.1. Literature . . . . .	37
4.1.1. Metric correlation . . . . .	37
4.1.2. Null models . . . . .	37
4.2. Statistical Analysis . . . . .	37
4.2.1. Metric correlation . . . . .	37
4.2.2. Null models . . . . .	38
4.2.3. Observed matrices . . . . .	39
4.2.4. Moduled matrices . . . . .	40
4.2.5. ANOVA . . . . .	41
4.3. Conclusions . . . . .	42
<b>Acknowledgements</b>	<b>45</b>
<b>A. Altered bipartite code</b>	<b>51</b>



## List of Figures

1.1. Example of a perfectly nested matrix . . . . .	6
1.2. Example of a moderately nested matrix . . . . .	6
1.3. Example of a not nested matrix . . . . .	6
2.1. Small matrix . . . . .	28
2.2. Medium matrix . . . . .	28
2.3. Big matrix . . . . .	28
2.4. Moduled matrix . . . . .	28
2.5. Observed matrix . . . . .	28
3.1. Scatter plot of metric correlation . . . . .	34
3.2. Scatter plot of metric correlation after null model application . . . . .	34
3.3. Comparison of small matrices and r2d generated matrices . . . . .	35
3.4. Metric-null model combination of the moduled matrices . . . . .	36
3.5. Metric-null model combination of the observed matrices . . . . .	36



## List of Tables

1.1. Causes of nestedness by site properties . . . . .	8
1.2. Causes of nestedness by species properties . . . . .	9
1.3. Causes of nestedness by neither site nor species properties . . . . .	10
1.4. Common metrics and their scaling of nestedness. . . . .	16
1.5. Common null models for presence-absence matrices and abundance matrices . . . . .	21
1.6. Software for nestedness analysis . . . . .	23
2.1. Metrics used and their equivalent in literature . . . . .	29
2.2. Null models used in my analysis . . . . .	29

# Abstract

Detecting patterns of interaction in networks is a primary goal of ecological research and a nestedness pattern has been detected frequently over the last years. Nestedness is a interaction pattern of a network that shows high occupancy of links in the upper left triangle of the ordered interaction matrix. As this shape may be the result of environmental gradients, it can be used to detect which gradients are those shaping the community and is therefore a very interesting ecological pattern. While there are many ways of interpreting a nested pattern, there are just as many ways of calculating it. More and more null models and metrics are developed, while their performance was not sufficiently tested for. While some metrics as well as some null models have been compared, there are still many unsatisfying evaluated possible combinations. It was, for example, not tested sufficiently for the influence of different matrix types. The possibility of method combinations leads to a problem in comparing the “nestedness” that was conducted by these different methods. The question arises whether different calculated nestedness even describes the same pattern. Differing in matrix ordering process, choice of metric and null model, ecologists have many options when it comes to evaluating data on whether or not it is nested. Therefore it is difficult to compare and infer from the result of nestedness analysis, if the processes in which this “nestedness” was created vary in the underlying methods. Existing literature on this topic was summarised to give an introduction on common metrics and null models and the results of previous analysis on their performance. Seven null models and ten metrics were chosen to test how they perform together. Intrinsic metric scaling was compared before applying the null models and was found to be inconsistent in their return. The same was done after applying the null models, which led to even less consistency. Metric and null model performance on modulated matrices was analysed and were all found to be unsuited for such an analysis. One very constraint model was analysed on how different metrics influence it. I found that the metrics are not consistent in their definition of a nested pattern. Especially the  $T$  metric returns results very different from all other metrics.



# 1. Introduction

## 1.1. Motivation

A major aspect in ecology is detecting patterns in networks and understanding their origin. One such pattern, that has gained more and more attention over the years, is nestedness. While there are different definitions of nestedness used, they all agree that a nested pattern consists of two properties: firstly a core of generalists species, where generalists interact with other generalist and secondly an asymmetry of interaction strength, where specialist species only interact with generalists but not other specialists (Krishna et al., 2008). If translated into a matrix representing the network, such a pattern would show a high occupancy of links in the upper left triangle after ordering rows and columns for example by marginal totals. A link is an observed interaction between two species or a sighting of a species at an area.

To quantify the degree of nestedness in such a matrix different ways of calculation have been proposed. There are multiple variations of metrics currently used. They have been developed over the last decades and authors proposed ever new different ways of calculating the nestedness of a matrix (Patterson and Atmar, 1986; Atmar and Patterson, 1993; Brualdi and Sanderson, 1999; Almeida-Neto et al., 2008; Corso et al., 2008; Staniczenko et al., 2013; Wright and Reeves, 1992; Wright et al., 1997; Cam et al., 2000; Lomolino, 1996; Cutler, 1991, 1994; Hausdorf and Hennig, 2003a). Another impotent part of nestedness analysis is the use of null models. They are a benchmarking test to indicate the significance of the degree of nestedness measured by the metric. Quantifying the degree of nestedness of a matrix and the test with a null model on significance can give insight into how nestedness arose in the given network. However, null models have always been a highly controversial topic and this is not different with regards to nestedness analysis (Gotelli and Ulrich, 2012). Clearly representing only one ecological driver is a very challenging task and therefore it is difficult to infer or conclude something from an analysis accomplished with an unclear null model.

Metrics and null models were used to calculate nestedness in different networks and nestedness was detected frequently. There is no agreed on metric-null model-combination. With networks that differ not only in size or fill but might even be nested only in part, it is highly unlikely that a overall applicable method exists at all. However, it is possible to find combinations that perform better than others (Ulrich et al., 2009). The current diversity of applied metrics and null models seems to cause more confusion than clarity and it is highly questionable whether studies that used different methods are comparable. Furthermore, networks may have been declared nested with inappropriate methods and might not be nested after all. There are combinations that detect nestedness much more frequently than it really exists and therefore significant nestedness would be detected were in fact there was none (Almeida-Neto et al., 2008). Drawing conclusions or even infer from questionable results is to be avoided in any case and therefore a reanalysis of previous studies is necessary if the methods used were inappropriate.

Which methods are inappropriate has to be analysed first and while many analyses regarding the combination of null models and metrics have been made, there are still many combinations that are not evaluated. Ulrich et al. (2009) analysed metric-null model performance but only for binary metrics. Recently, new metrics have been developed, that allow the analysis with abundance data (Staniczenko et al., 2013; Galeano et al., 2009; Almeida-Neto and Ulrich, 2011). Ulrich

and Gotelli (2010) conducted an analysis similar to Ulrich et al. (2009) for abundance data, but did not include several of the new metrics. The performance of abundance metrics has to be evaluated further with a wider range of null models. While metrics are used whenever nestedness is analysed, there is no evidence that they are consistent in their definition of nestedness. There was no study so far, that compared the intrinsic consistency of different metrics.

Comparing the various analyses is not possible if they make use of different metrics which are inconsistent in their intrinsic scaling of degree of nestedness. Therefore, commonly used metrics were chosen in the present work and their correlation in identifying degrees of nestedness was compared.

The different types of constraints in null models influence the results to a high degree (Hausdorf and Hennig, 2007). There are still gaps in knowledge as to which null model contributes how severely to the outcome. Differently constrained null models were chosen and combined with a selection of metrics to address this issue. Special interest was taken in the performance of metrics and null models on moduled matrices as there are no studies that treat this topic so far.

## 1.2. Definition and history

The concept of nestedness was first described in biogeography by Eric Hultén in 1937 and Philip J. Darlington in 1957, who observed different species sets on islands, that differed in size and shape. With nestedness, they described the pattern by which these species sets occurred together on defined spacial areas. A metacommunity was classified “nested”, if the species set found in a species poorer area, was a subset of the species set found in a richer unit (Ulrich et al., 2009). This was nicely put by Patterson and Atmar (1986) when they said that

“the species comprising a depauperate fauna should constitute a proper subset of those in richer faunas” (Patterson and Atmar, 1986, page 68).

Generally speaking, nestedness is defined by the following properties (Krishna et al., 2008):

- an asymmetry of interaction strength (specialists only interacting with generalists)
- a “core of generalists” (generalists interacting with generalists)

The sorted interaction matrix fulfilling these properties would exhibit a high occupancy of the upper left triangle. This, however, is a very vague definition as there is no clear consensus of nestedness. Figures 1.1, 1.2 and 1.3 show three matrices with a different degrees of a nested pattern.

The idea of a nested pattern spread from biogeography to other fields of application in ecology, with a slightly changed definition. New interpretations of a nested pattern were discovered and with these, nestedness gained in popularity. The first of which was introduced by Atmar and Patterson (1993); Patterson and Atmar (1986), who declared that a nested pattern can also be used to predict the order of extinction in a network. They introduced the so called matrix temperature metric  $T$ , the first metric, to calculate the degree of nestedness of a given matrix.

Since then, three major metrics types to calculate nestedness have asserted themselves in ecology. Atmar and Patterson (1995) developed the nestedness temperature calculator (NTC), the first software for nestedness calculation, which uses the matrix temperature  $T$  metric. It is still one of the most frequently used metrics to calculate nestedness. Brualdi and Sanderson (1999) introduced a “gap metric” (BR), which calculates nestedness according to the sum of necessary pushes of 1s and 0s in each row or column in order to obtain a perfect nested matrix. The third

major way to calculate nestedness is the “overlap metric”, which counts the number of species that form a perfect subset of the larger scaled set of species in a given row or column pair. The first of this type of metric was introduced by Hausdorf and Hennig (2003a) and is called HH, after the developers initials. “Matrix temperature”  $T$ , “gap metric” and “overlap metric” are the three common types of metrics that have been applied and varied by many authors. The recently introduced metric by Almeida-Neto et al. (2008) NODF (NO for paired nestedness overlap, DF for decreasing fill) is a variation of the HH metric and is the now the most commonly used “overlap metric”. Other metrics have been introduced, like  $\eta$  by Corso et al. (2008) or “nestedness” by Bastolla et al. (2009). The Weighted-Interaction Nestedness Estimator (WINE) is software using the WIN metric, which takes abundance into account. It was introduced by Galeano et al. (2009) and is the first metric using abundance data. This was an important step to future analysis as many studies showed that abundance can play a major role in networks and highly influences whether or not a network is nested (Staniczenko et al., 2013).

When Bastolla et al. (2009) claimed that nestedness is the cause for greater biodiversity in mutualistic networks by applying nestedness on plant-pollinator networks, it caused another wave of great interest in its field. Many others found these results to be contradictory to their own analysis (Kondoh et al., 2010). They also found nestedness in food webs, showing that not only mutualistic networks, but also competitive ones can show a nested structure. The occurrence of highly nested sub webs in antagonistic networks, which result in a moduled matrix was another important finding of Kondoh et al. (2010).

Another important and highly controversial part of nestedness analysis have always been the benchmarking test for significance of a nested pattern. As the number of metrics increased, new benchmarking procedures were developed. The Monte Carlo method, which uses the average of a large number of random simulations, asserted itself as the benchmarking procedure for nestedness analysis. In nestedness analysis, these simulations are created by the null models which represent different ecological gradients. The Monte Carlo method was among others used by Patterson and Atmar (1986); Lomolino (1996); Hausdorf and Hennig (2003b).

Nestedness is an important pattern in ecology, but its analysis is complex and difficult to undertake for an outsider. Therefore many software packages have been developed to make it easier to choose between the different options to evaluate nestedness. Recently, new software includes the possibility to evaluate binary and abundance matrices as well (Beckett et al., 2014; Ulrich and Gotelli, 2010; Ulrich et al., 2009; Hausdorf and Hennig, 2003b).

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	0
3	1	1	1	1	1	1	1	0	0	0
4	1	1	1	1	1	1	0	0	0	0
5	1	1	1	1	1	0	0	0	0	0
6	1	1	1	1	0	0	0	0	0	0
7	1	1	1	0	0	0	0	0	0	0
8	1	1	1	0	0	0	0	0	0	0
9	1	1	0	0	0	0	0	0	0	0
10	1	0	0	0	0	0	0	0	0	0

Figure 1.1.: Perfectly nestedness

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1	1	1	1	0	1	0	1	0	1	1
2	1	1	0	1	0	1	0	0	1	0
3	1	0	0	1	1	0	1	1	1	0
4	1	1	0	0	1	0	1	0	0	1
5	1	0	1	0	1	0	0	1	1	0
6	1	1	0	1	1	0	1	0	1	0
7	1	1	0	1	1	0	1	0	0	0
8	1	1	0	1	0	1	0	0	0	0
9	1	0	1	0	1	0	0	0	0	0
10	1	0	1	1	0	0	0	0	0	0

Figure 1.2.: Decreasing nestedness

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1	0	1	0	1	0	1	0	1	0	1
2	1	0	1	0	1	0	1	0	1	0
3	0	1	0	1	0	1	0	1	0	1
4	1	0	1	0	1	0	1	0	1	0
5	0	1	0	1	0	1	0	1	0	1
6	1	0	1	0	1	0	1	0	1	0
7	0	1	0	1	0	1	0	1	0	1
8	1	0	1	0	1	0	1	0	1	0
9	0	1	0	1	0	1	0	1	0	1
10	1	0	1	0	1	0	1	0	1	0

Figure 1.3.: Not nested

### 1.3. Causes of nestedness and nestedness as cause

There are many theories on how nestedness arises. They can be divided into three main groups:

- 1) caused by site properties, shown in table 1.1
- 2) caused by species properties, shown in table 1.2
- 3) caused by other factors, shown in table 1.3

The site properties are only applicable to metacommunities, or networks in which one species acts as a site (host-parasite). These causes often can not be seen as the singular cause of nestedness because they influence each other. Real networks are highly complex systems, each having it's own specific processes shaping the community. This makes it difficult to find a null model that unambiguously describes one of these processes. Species-species networks are also ruled by different underlying processes since there is a difference between mutualistic and antagonistic networks (Corso et al., 2008). The difficulty of finding the cause of nestedness in a network was well phrased by Patterson and Atmar (2000, page 10) :

“Yet most ecological systems are nested, and the factors influencing this ecological structure are as diverse as nature itself.”

Krishna et al. (2008) found that 60% - 70% of nestedness was explained by relative species abundance. In a purely neutral mutualistic network, generalisation leads to higher abundance

and higher abundance leads to more interaction partners. Therefore nestedness would arise solely from abundance. Johnson et al. (2013) found that degree heterogeneity and the finite size of a real network are sufficient to create the level of nestedness observed in empirical networks. They concluded that networks are nested by chance as real networks were found to be naturally disassortative and nestedness was found correlated with disassortativity. Bascompte et al. (2003) suggested that more interactions would lead to a higher degree of nestedness. They claimed that compartments in food webs would not arise from tight specialisation. Ulrich et al. (2009) summarised all the causes by proposing that they are all variations of ordered extinction or colonisation, and other causes are merely the different gradients that lead to extinction or colonisation.

**Nestedness as a cause** Bastolla et al. (2009) gained much attention and fed the interest in nestedness when he claimed that nestedness would increase the biodiversity of a network. It is stated that the maximum number of species was reached through a nested pattern. This theory has been proven incorrect by James et al. (2012) who found that the total number of links divided by the network magnitude (hence connectance), not the degree of nestedness influences persistence and the survival of a species and therefore the biodiversity. It was even found that randomization, which decreased nestedness, simultaneously increases persistence.

Table 1.1.: Causes of nestedness by site properties

Cause	Description	Source
Nested habitats	Caused by the heterogeneity of the habitats and the specialisation of species. More generalist species can occur in different habitat types, while specialists only occur in specific habitats	Wright and Reeves (1992)
Habitat quality	Habitats of higher quality are preferably colonized (for migratory birds)	Simberloff and Martin (1991) Ulrich (2006b)
Disturbance regimes	Frequency of disturbance and species resilience. The higher the disturbance frequency the fewer disturbance delicate species	Ulrich (2006b)
Hierarchical niche overlap, Fragmentation	Man made isolation of once connected areas. Similar to natural isolation but with a very steep extinction slope.	Patterson and Atmar (2000)
historical refuges	Adaptation to different climate condition and retreat to refuges in the past	Daubennire (1975)

Table 1.2.: Causes of nestedness by species properties

Cause	Description	Source
Selective colonization	Colonization as a combination of isolation of site and dispersal ability of species. Higher dispersal rate leads to colonisation of more isolated sites	Lomolino (1996) Ulrich et al. (2009)
Selective extinction	Extinction due to area size and species delicateness to extinction. Inverse order of selective colonization. With decreasing area species that are more specialised go extinct first.	Lomolino (1996) Ulrich et al. (2009)
Forbidden links	Interactions that are not existing (f.e due to diurnal and nocturnal behaviour or phenotype incompatibility)	Krishna et al. (2008)
Selective environmental tolerances along stress gradients	Combination of environmental harshness and species specific resistance. The higher the generalisation of a species the higher the resistance to environmental difficulties	Blake (1991)
Native, non-native species assemblage	Immigration in new habitats more likely by generalist species (no suited habitats for specialists)	Novak et al. (2011)
Community entering	Interaction of new species mainly with generalist residential species	Bastolla et al. (2009)
Dispersal ability	High dispersal rate allows for easier colonisation	Lewinsohn et al. (2006)
Alignment of phenotype traits	“Forced links” interaction only possible with specific species due to phenotype traits	Stang et al. (2006)
Evolutionary history	Evolutionary adoption to specific circumstances that only allow for interactions with evolutionary partner	Krishna et al. (2008)
Assemblage convergence	Coevolutionary convergence allowing other species to be connected to the network	Thompson (2005)
Optimal diet choice	Decreasing food preferences due to nutrient contents, inter specific difference in food quality intake	MacArthur and Pianka (1966)
Interspecific hierarchy	Differences of species lead to distinct prey-predator roles	Kondoh et al. (2010)
Niche differentiation in behaviour	Different lifestyle leads to forbidden interaction (migratory species, hibernation)	Kondoh et al. (2010)
Timing of resource use	Nocturnal -diurnal behaviour	Kondoh et al. (2010)
Species dependency	Relative importance of specific interaction for one partner	Krishna et al. (2008)
Phylogenetic relationships	Extinction processes set of ordered extinction of related species	Rezende et al. (2007)

Table 1.3.: Causes of nestedness by neither site nor species properties

Cause	Description	Source
Passive sampling	Species ubiquity	Bastolla et al. (2009)Ulrich et al. (2009)
Weighted abundance	Interaction strength concludes from abundance data	Krishna et al. (2008)
Too low sampling intensity	Low sampling intensity is typical for real networks and is sufficient to mimic nestedness	Bluthgen et al. (2008)
mass effect	The more individuals exist, the more interactions are observable	Ulrich et al. (2009)

## 1.4. Short analysis guide

Seven steps have to be taken in a nestedness analysis:

1. Acquire data on the network.
2. Create a matrix from the data, which will hereafter be called the “original matrix”.
3. Choose an appropriate metric.
4. Apply the metric on the original matrix and obtain the original metric value.
5. Choose an appropriate null model that represents the ecological gradient to analyse the network for.
6. Apply the null model on the original matrix and generate sufficient new matrices, which will hereafter be called “null matrices”.
7. Apply the chosen metric on each null matrix and obtain a metric value distribution.

If the original metric value is in the significant interval of the distribution, the network is significantly nested and the gradient excluded by the null model contributes to the nestedness in the network.

## 1.5. Matrices

To analyse a network, it must be translated into a matrix. Two matrix types are used for this purpose. The first is the commonly used presence-absence matrix (McCoy and Heck Jr, 1987). It can be subdivided into two matrices: a matrix whose rows and columns represent species and sites respectively, and an interaction matrix in which rows and columns both describe species. The former describes a metacommunity where a non-zero entry indicates the presence of a species at a site, whereas the latter describes an interaction matrix and a 1 indicates an interaction between the two species. If parasitic-host networks are analysed, a species can also be seen as a site (Ulrich et al., 2009; Almeida-Neto et al., 2007). Therefore interactions of any kind between two parties (species-location or species-species) are denoted with 1 and conversely, 0 denotes that no interaction has been observed (Ulrich et al., 2009). The number of non-zero entries is called the frequency of the matrix. The connectance of the matrix is the number of non-zero entries divided by the number of matrix cells. Presence/absence matrices do not make it clear whether there has been only one or multiple interactions. The abundance frequency is not evaluated and information is lost. The matrix that includes such data is an abundance matrix. The principle is the same as with the presence-absence matrix, only that instead of a binary value (0 or 1) the actual observed interaction values are noted.

A third matrix type to be mentioned here is a hybrid between the classic presence-absence matrix and the abundance matrix. This matrix is filled with 1s and 0s but based on the probability of a cell fill determined from a abundance matrix. Therefore abundance values are incorporated indirectly. It can be evaluated with classic presence-absence metrics (Ulrich et al., 2009).

A maximal nested matrix has a 50% fill and nodes are perfectly aggregated in the upper left triangle of the matrix. A perfectly nested matrix however is independent of fill as long as the interactions match the criteria of perfect nestedness of the specific metric used (Almeida-Neto et al., 2007). Depending on whether an empirical matrix is tested for nestedness or a metric quality is tested, these original matrices can either be produced from observed data, or drawn from a appropriate distribution that approximates the distribution of the network of interest.

Almeida-Neto et al. (2007) warned to use the term anti-nestedness as it is not well defined. It is used to describe a perfectly checkerboarded pattern as well as a completely random pattern.

## 1.6. Metrics

A metric is a way to calculate the degree of nestedness of the original matrix according to a defined method (Ulrich et al., 2009). According to Ulrich (2009) there are 4 properties that a good metric must provide:

1. it has to be independent of matrix fill;
2. it has to be independent to size and shape of the matrix;
3. it has to be invariant to the transposition of the matrix;
4. it has to be invariant to the inversion of presence and absence.

An other definition of a good metric was made by Gotelli and Ulrich (2012, page 4) :

“A proper nestedness metric should measure the degree to which species poor sites appear to be random subsamples of species rich sites.”

Many metrics have been proposed over the years to calculate nestedness of a non-weighted network. Generally spoken, a classic binary matrix is analysed based on the number of non-zero cells and their position in the matrix. Three main metric types have asserted themselves for the calculation of nestedness: the temperature metric ( $T$ ), gap metrics and overlap metrics. There are metrics that use different approaches, which could prove useful for future analysis. For each of these types, many variations have been proposed. However, only the most commonly used will be reviewed here.

**Temperature metric  $T$**   $T$  is the metric developed by Patterson and Atmar (1986) and is used in their NTC (Atmar and Patterson, 1995). For the calculation of  $T$ , rows and columns are both sorted first by marginal totals and then to obtain the most nested matrix according to its calculation. It is calculated using the squared distance of unexpected absence above the isocline of perfect nestedness and of unexpected presence below the isocline. It sums the squared deviations of these “surprises”, divides the sum by the maximum value possible for this matrix and then multiplies it by 100 to obtain the temperature scale from 0 to 100 (Atmar and Patterson, 1993; Almeida-Neto et al., 2007). It was believed that unexpected absence and presence are more meaningful than expected absence and presence, because they are less likely to occur and therefore should be assigned more weight. The isocline of perfect nestedness is drawn from the lower left corner to the upper right corner of the matrix, thereby dividing it into two triangles with equal size. It crosses the line (and dividing the space above and below them equally) drawn from the lower left edge to a point XY and from the point XY to the upper right edge. The point XY is chosen according to size and fill of the matrix (Ulrich, 2006b; Beckett et al., 2014). It ranks from 0°-100° where 0° represents perfect nestedness and 100° perfect anti-nestedness in the matrix (Johnson et al., 2013). Anti-nestedness meaning a perfectly random pattern in the NTC. Bascompte et al. (2003) used a normalised version to obtain values from 0 to 1 for better comparison.  $T$  is still among the most used measures although it has been proven to have many flaws. Ulrich et al. (2009) found it to be correlated to the matrix size, James et al. (2012) found it to be at least more prone to type 1 error than the BR metric and Ulrich and J Gotelli (2007) found the type 1 error rate too high as well. Corso et al. (2008) found several problems and found  $T$  to be “not well defined”. Almeida-Neto et al. (2008) even suggested that all previous studies using  $T$  should be re-evaluated because of its many flaws. Matrix dimension seems to have a bigger influence than matrix fill on this metric (Wright et al., 1997; Greve and L Chown, 2006).

**BR** The discrepancy metric (BR) introduced by Brualdi and Sanderson (1999) is used on matrices ordered by marginal totals. The various approaches of ordering the matrices are discussed in paragraph sorting. It measures the necessary corrections of presence or absence that must be made to obtain a perfectly nested matrix. It counts the ‘wrong’ 0s and 1s that have to be moved to obtain the perfectly nested matrix. This means that it counts all 0s that follow a 1 and counts how many “pushes” it takes until perfect order is acquired. It is calculated for both rows and columns and the version that requires least pushes will be used. Using it only on rows or columns would resolve in a metric that is not invariant to matrix transpose. Therefore a separate calculation of this metric for rows and columns and a later combination of both was used by Beckett et al. (2014) for the BR metric. Smaller BR scores indicate higher nestedness. The normalisation of this metric is given by dividing its value by the matrix fill (the total number of occurrences) (Ulrich, 2008). It is sensitive to ordering of rows and columns, reaching its maximum when ordered by row and column total (Ulrich and J Gotelli, 2007; James et al., 2012; Beckett et al., 2014). Ulrich and J Gotelli (2007) also found it to be less prone to type 1 error than  $T$ . Type 1 error is the wrongful detection of an effect that is not there. Type 2 error is the missing detection of an effect that is present. Further details are given in section 1.7.

Ulrich et al. (2009) analysed eight gap counting metrics and found the NC (Wright and Reeves, 1992) and BR metrics (Brualdi and Sanderson, 1999) to be most independent of matrix shape, fill and size and the UT (Cutler, 1991) and again the BR metric to be most invariant to occurrence inversion and transposition. Six other gap metrics called N0, N1, Ua, Ut, Up, D were analysed by Ulrich et al. (2009), but found to be less well operating and will not further be discussed.

**HH** The overlap metric, first introduced by Hausdorf and Hennig (2003b) which was later referred to as HH, measures nestedness regarding decreasing fill and paired overlap. This means that it counts the number of cases in which one set is a proper subset of a set of a larger scale. Smaller values indicate a nested pattern. Two identical species are not enough to count as a subset. This metric allows for the test of several nested subsets, as it was done by Hausdorf and Hennig (2007).

**NODF** The HH overlap metric is the basis for the popular NODF metric by Almeida-Neto et al. (2008) also based on decreasing fill and paired overlap. Decreasing fill is measured on a scale from 0 to 100 with 100 (perfect nestedness) when the marginal totals of a row (column) further away from most filled row (column) is lower than the one before. Otherwise it will be 0 (no nestedness). This also means that row pairs with identical marginal total (tied ranks) will be counted as 0. Therefore NODF counts such tied ranks as negative contributions to overall nestedness (Podani and Schmera, 2012). This also leads to a problem in distinguishing between a perfect checkerboard pattern and perfect compartmented pattern, as both would return 0 nestedness. Paired overlap is the count of all row (column) pairs in which a connection appears in the row (column) further away from most filled row (column) where a connection is already in the same place of the previous row (column). If this is the case for all fills in the row pair, it is measured as 100 (nested). If a fill appears in the row (column) further away from most filled row (column) it is measured as 0 (no nestedness). Then both counts for each pair are summed and averaged to obtain a nestedness measure ranging from 0 (no nestedness) to 100 (perfect nestedness). As nestedness is calculated separately for rows and columns and then combined later, it could be calculated for only one of these two. Therefore it could be used as a measure of the overlap of species from poor to rich sites as well as a measure for distribution richness (Beckett et al.,

2014; Strona et al., 2014; Ulrich and Almeida-Neto, 2012). Almeida-Neto et al. (2008) claim that it reaches maximum nestedness at 50% fill. They compared it to other commonly used metrics ( $T$ , BR, NC) and found it less sensitive to matrix size and shape and less prone to type 1 error. Beckett et al. (2014) found it to be sensitive to ordering reaching its maximum for ordering according to marginal totals. Podani and Schmera (2012) modified NODF and called the new metric NODFmax. They claimed this measure to be independent of matrix ordering. Yet in their opinion, both metrics still gave too much weight to the overlap component and too little to differences in richness.

**PRN and PRSN** Podani and Schmera (2012) introduced variations of the overlap metric. The percentage relativized nestedness (PRN) and the percentage relativized strict nestedness (PRSN). They also included the mean Simpson similarity into their analysis. According to them NODF is a variation of the mean Simpson similarity of site pairs. This is for matrices in which column margins are strictly monotonically decreasing from left to right. The PRN and PRSN metrics uses paired overlap and richness difference and are insensitive of column ordering, but count tied ranks differently. PRN counts them positively, while PRSN counts them negatively. Mean Simpson and NODFmax both don't handle richness differences specially. Mean Simpson handles tied ranks positively, NODFmax negatively. The main difference of those metrics to the NODF is that they are order-invariant. All metrics that count ties negatively have the same problem as NODF in distinguishing between completely segregated pattern (checkerboard) and pattern with different but internal completely aggregated compartments. It could also happen that nested pattern measured less nested than random pattern. They also found that their metrics performed very well with unconstrained null models and claimed that they were resistant to type1 and type 2 errors. They found that PRSN and NODFmax performed better than PRN and mean Simpson.

$\eta$  Another metric has been introduced by Corso et al. (2008) called  $\eta$ , based on the Manhattan distance MD, which is the distance of a point to another in a matrix using only vertical and horizontal lines. It sums the row and column index of each connection and assigns this sum as the weight of the connection (Beckett et al., 2014). It uses a reference point (upper left corner cell) as the "epicenter" of nestedness and sums the distance of all filled cells to this point. It is projected into a Cartesian coordination system and rescaled to the unit square to obtain a symmetric matrix which takes values from 0 (nested) to 1 (random). It then undergoes a benchmarking procedure that takes into account the maximum possible distance (random) and the minimal distance (nested). Beckett et al. (2014) found it to be sensitive to ordering and maximal for ordering according to row and column totals.

**"nestedness"** Bascompte et al. (2003) suggested a metric which they called "nestedness" ranging from 0 to 1. It averages the number of interactions that are shared by two species relative to their number of overall interactions in mutualistic networks. Johnson et al. (2013) introduced a variation of the "nestedness" metric, which will be referred to as JDM as suggested by Beckett et al. (2014). They found the "nestedness" metric to depend on the heterogeneity of the degree distribution in some cases and criticised the normalisation factor used. The new metric uses the adjacency matrix and measures the disassortivity between the nodes of the network. It normalises the overlap (calculated using the adjacency matrix) using the expected nestedness of a random graph that has the same degree distribution as the original matrix. This way it accounts for degree heterogeneity effects. It can easily be applied to bipartite networks and show the nest-

edness value of a single connection as well as the overall nestedness. It also suggests a better way of normalisation and was found to be invariant to row and column ordering (Beckett et al., 2014).

**checkerboard** Another way to analyse the matrix is for its checkerboard character is the so called C-score suggested by Stone and Roberts (1990). It is the count of all 2x2 submatrix of the original matrix, after ordering according to marginal totals, in which two species never appear together, averaged over the community. It returns values from 0 (nested) to 1 (perfect checkerboard). The structure of such a matrix would be  $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  or  $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ . The higher the C-score, the higher the species segregation. Therefore it can be used to analyse nestedness. The checkerboard score CH (Ulrich and J Gotelli, 2007) is the non-normalised version of this score. Checkerboard is the definition for anti-nestedness used by Jonsson (2001). This metric has found overall acceptance but must be used with care as it cannot distinguish between segregation and aggregation under certain null models (fixed-fixed): since it is the average of all pairwise values of different species it is not clear whether the high checkerboard character of the matrix arose from a equal distribution of segregation or from segregation and aggregation. Positively as well as negatively associated species pairs are detected by this metric.

Recently a number of new metrics were proposed by different authors, which can be used for the analysis of weighted networks. Ulrich and Gotelli (2010) tested six of these new metrics. Although they see high potential in further abundance analysis they found nearly all of these metrics in combination with various null models to have too high type 1 error rates when tested with random matrices. The Mantel test, that compares two matrices to detect non-random correlation, is mentioned as the only acceptable metric, but criticised for its poor power when testing for significant aggregation or segregation.

**WIN** Manhattan distance is the basis of the weighted metric that WIN uses in the WINE software. With this metric, ordering of rows and columns according to marginal totals puts highest totals in the lower right corner, but keeps the reference point in the upper left corner. Nestedness is highest if the distance is greatest. Values range from 0 to 1, where 0 indicates a random matrix and 1 a nested one. To account for the weight of an interaction, two adjacency matrices are calculated to show the dependency of a row element on a column element and vice versa. Just as with its binary brother *eta*, the matrix is rescaled to a unit square in a Cartesian space. This way a weighted interaction distance is calculated for all non-zero elements. The mean of this results is the nestedness metric WIN. This metric was found independent from matrix fill and size and invariant to transposition. But, like all analysis, it is sensitive to the number of sample. (Galeano et al., 2009).

**WNODF** The weighted NODF called WNODF is calculated very similar to the above mentioned NODF. It is also based on decreasing fill, meaning that if the marginal total of the upper row of a pair of rows is smaller than the lower one, the paired nestedness is zero. If not, the paired nestedness is the percentage of cells that have a lower value than the according cell in the row above. Therefore it is more strict than NODF, not only requiring an paired overlap but also greater values in this paired overlap (Beckett et al., 2014). A mean nestedness value can then be calculated from all row paired nestedness. The mean nestedness of all column paired

nestedness' is calculated likewise. WDNODF then uses both, the row paired and column paired values to calculate the overall nestedness. Just as NODF the values of WNODF rank between 0 (no nestedness) and 100 (perfect nestedness) (Almeida-Neto and Ulrich, 2011). Staniczenko et al. (2013) criticised that WNODF does not take all quantitative information available into account.

**SR** A very interesting and new approach called SR (for spectral radius) was proposed by Staniczenko et al. (2013). This metric can calculate nestedness for binary as well as for abundance matrices. They used the spectral radius of the largest eigenvalue of the adjacency matrix to calculate the nestedness of a matrix. The matrix is ordered according to the entries of the dominant eigenvalue. It is controlled for the influence of mass effect by scaling preferences relative to mass action in the binary matrix. They found that a nested structure would have the largest spectral radius for binary as well as quantitative matrices. This method was proven to be very fast and invariant to permutation of the matrix. If the largest spectral radius of the empirical matrix is greater than the lowest of a perfectly nested matrix of same type then it is highly likely for it to be nested. (Staniczenko et al., 2013).

One needs to keep in mind that nestedness is scored differently among the different metrics. The degree of nestedness increases with increasing values for NODF, SR, JDM, WIN, and WNODF but with decreasing values for  $T$ , BR, *eta*, "nestedness", and C-score. C-score, WIN, *eta*, "nestedness" have a scale from 0 to 1,  $T$ , NODF and weighted NODF from 0 to 100 (Beckett et al., 2014). This is shown in table 1.4. The diversity of scoring systems is another example for the non-uniform character of the field and the multitude of different approaches.

Table 1.4.: Common metrics and their scaling of nestedness.

Metric	Nested	Random	Example
$T$	0	100	Atmar and Patterson (1993)
nestedness	0	100	Bascompte et al. (2003)
C.score	0	1	Stone and Roberts (1990)
<i>eta</i>	0	1	Corso et al. (2008)
BR	0	>0	Brualdi and Sanderson (1999)
checker	0	>0	Dormann (2008)
NODF	100	0	Almeida-Neto et al. (2008)
wNODF	100	0	Almeida-Neto and Ulrich (2011)
PRN	100	0	Podani and Schmera (2011)
WIN	1	0	Galeano et al. (2009)
JDM	>1	1	Johnson et al. (2013)
HH	low	high	Hausdorf and Hennig (2003a)
SR	high	low	Staniczenko et al. (2013)

## 1.7. Null models

"A null model is a pattern-generating model that is based on randomization of ecological data or random sampling from a known or imagined distribution. The null model is designed with respect to some ecological or evolutionary process of interest. Certain elements of the data are held constant, and others are allowed to vary stochastically to create new assemblage patterns. The randomization is designed to

produce a pattern that would be expected in the absence of a particular ecological mechanism” (Gotelli and Graves, 1996, pages 3-4)

To analyse nestedness, null models are needed to evaluate the significance of the metric value of the original matrix. Which null model is the right choice, and whether null models are appropriate in general, has been a highly controversial topic. As a result, there are many null models that are more or less useful and more or less tested for their performance (Gotelli and Ulrich, 2012). A null model is a method to test the significance of a pattern calculated with a specific metric. The null model will generate a certain number of matrices under chosen constraints from which a metric value distribution is drawn. Excluding one specific factor will give conclusion on how this factor contributes to nestedness. The excluded factor of the null model should solely represent the one specific gradient to test for. If the original metric value is in the significant part of this distribution, the original matrix is indeed significantly nested, and the excluded factor is the reason for the nestedness of the network. To be able to draw a conclusion, this distribution has to be drawn from a sufficient number of null matrices. Galeano et al. (2009); Ulrich (2006b) found that 100 matrices should be enough to avoid undersampling, but still have a short enough calculation process. However, null models are highly debated regarding their quality of performance and also need to be analysed for their proneness to type 1 or type 2 errors. Before testing with a null model (to detect the influence of a factor) it is therefore necessary to test the null model itself and its vulnerability to be prone to type 1 or type 2 error.

Null model is prone to type 2 errors if it does not detect nestedness when there is nestedness.

Null model is prone to type I errors if nestedness is detected when a matrix is random.

(Gotelli and Graves, 1996)

**Test of the null model** (Gotelli and Ulrich, 2012) :

1. Decide on a null model to test for a particular pattern.
2. Decide on a metric.
3. Create a set of random R matrices and of structured S matrices.
4. Apply null model and metric on S matrices.
  - a) Evaluate type 2 error rate for null model and metric for structured matrices.
5. Apply null model and metric on R matrices.
  - a) Evaluate type 1 error rate for null model and metric for random matrices.
6. Decide on the quality of null model and metric.
7. If good : test on empirical data.
8. If bad: choose new null model and metric. Repeat process.

After the performance of the null model is tested, a null model can be chosen to do the actual analysis.

**Test with the null model** (Gotelli and Ulrich, 2012) :

1. Obtain empirical matrix  $M_E$ .
2. Apply chosen ordering process on  $M_E$ .

3. Apply chosen metric on  $M_E$  -> obtain metric value for  $M_E$ .
4. Apply chosen null model on  $M_E$ -> obtain new set of matrices  $M_N$ .
5. Apply chosen metric on  $M_N$  -> obtain metric values for  $M_N$ .
6. Draw significance interval (5%) from metric values of  $M_N$  distribution.
7. Compare metric value of  $M_E$  with the value distribution of  $M_N$ .
8. Draw ecological inference.

The right choice of null model is essential for the analysis. Finding the right null model is a very difficult task and depends not only on the matrix and metric used but mostly on the ecological gradient one wants to test for. There are many null models to choose from, each with its own perks and flaws and it is up to the ecologist to find the appropriate null model for his analysis. Table 1.5 lists the most commonly used null models<sup>1</sup>.

There are two things to combine, firstly the type of null model and secondly the constraints on the null model. There are three different types of null models: those using only binary data, those using abundance data to calculate the probabilities for a fill in the binary matrix, and those using the abundance data in the returning null matrices. All null models differ additionally in the way they are constrained. Constraints can for example be applied to fill, row sums, column sums, connectance, frequency, dimension, row/column connectance and probability. For abundance matrices it can additionally be an individual (fill) constraint or a population (value) constraint. (Ulrich and Gotelli, 2010).

The number of generated null models to be used on each matrix is a trade-off between accuracy and time. More null models will generate a better result, but it requires much more calculation effort. The aim is to avoid undersampling by using too few null models but to generate enough for a sound result (Beckett et al., 2014). According to Galeano et al. (2009); Ulrich (2006b), 100 matrices yield an accurate result while keeping calculation time reasonably short.

Since presence-absence matrix is the predominantly used matrix type, the binary based null models that go along with it are more common than the others. The least constrained binary null model is the equiprobable-equiprobable (EE) model. It does not constrain the marginal totals and lets individuals float within the matrix, but keeps the occurrence constrained (Ulrich, 2006b). The EE null model can be used if the matrix is independent of variations of richness or incidence and therefore independent on the marginal distribution (Moore and Swihart, 2007).

The most constrained binary null model is the fixed-fixed model, that constrains matrix size, fill, marginal totals and frequency (Strona et al., 2014; Ulrich et al., 2009).

There are null models that either constrain row sums but not column sums (EF) or vice versa (FE) (Ulrich et al., 2009).

The FF null model should be used if the factors that contributed to the marginal distribution are gradients of interest for the analysis (Gotelli et al., 2010; Moore and Swihart, 2007). Joppa et al. (2010) found that a null model with unconstrained column totals generates distributions that do not resemble nature. Ulrich et al. (2009) prefers the FF model as it is more conservative and more of the original elements are retained. He further found the EE to have a problem with its dependency of variance for different matrix sizes. Ulrich (2009) found that null models that are too liberal would give similar distributions with increasing matrix size and thereby lower the variance produced. They also found a positive influence of restrictive models on passive sampling effects. Ulrich and J Gotelli (2007) found that the FF model is better suited for nestedness analysis than the EE model. The FF model decreases the occurrence of type 1 error but it therefore

---

<sup>1</sup>A full characterisation of all null models is near impossible due to the lack of information given by the authors.

increases the occurrence of type 2 errors (Ulrich et al., 2009; Ulrich and Gotelli, 2007). It can not be applied to extremely filled or empty matrices as there are not enough possible rearrangements (Ulrich and J Gotelli, 2007). Hausdorf and Hennig (2007) showed that the constraints have a major impact on the result as they showed that 14 out of 18 matrices were not nested under constrained models but showed high nestedness when compared to unconstrained models. There is a tendency not to use the EE model as it was found to identify nestedness too frequently (Ulrich et al., 2009).

The null models that use abundance to calculate the resulting binary matrices are similar to the classic binary matrices. They use the abundance data to calculate the probability of a fill for either rows and columns or for each cell in the matrix and assign 1 and 0 to each cell according to these probabilities. Moore and Swihart (2007) found that the null model choice again was a very important influence on the results. They used the Recol model to analyse seven datasets that were already analysed with the Randnest model and identified only one as nested, although Randnest originally found all seven to be nested. Nielsen and Bascompte (2007) found that the CE model performed well and had a low rate of type 1 errors. These models are described in table 1.5.

Until recently, the commonly used presence-absence matrix only required a null model that uses binary data. The use of abundance data requires null models that can use abundance data. According to Ulrich and Gotelli (2010), such null models could be more powerful than the sole analysis of presence-absence matrices. These null models use the actual values that were observed, and so contain more of the original information. Staniczenko et al. (2013) reanalysed binary matrices that showed significant nestedness using abundance data. They found that the nestedness was no longer observable after the application of the abundance data on those matrices. Ulrich and Gotelli (2010) analysed 14 different abundance null models and found those four null models to be overall best performing that kept fill constant, but allowed individuals to vary and did not retain the original zero cells.

A main reason to use a null model is to identify the ecological pattern that structured the network. Therefore a good null model should distinctly represent one factor (ecological gradient, forbidden links, etc).

If the original matrix was identified as highly nested by the metric and the application of the null model on this matrix indicates significance of the degree of nestedness, then the factor left out by the null model played an important role in shaping the network. If it is identified as not significantly nested, then the factor has no influence on the nestedness of this network. Consider two null models that both constrain marginal totals with only one of them additionally keeping 0 fixed. This additional constraint represents “forbidden links”, that is, interactions that can never occur. These two null models are applied to the original matrix. If the first does not identify significance but the second does, one can conclude that the “forbidden links” are the reason for nestedness in the network. However, finding the correct null model is a challenging task because it is very difficult to represent only one clearly defined gradient. The essence of this problem was captured by Gotelli and Ulrich (2012) :

“Relating complex patterns in presence–absence matrices to explicit ecological mechanisms remains a difficult challenge.” (Gotelli and Ulrich, 2012, page 8)

**Sorting** Before a null model is applied, it is common to order the matrix. There are different methods to order a matrix. The most intuitive way is to order it by row and column marginal totals. Hereby rows and columns are swapped until as many interactions as possible are in the desired (normally upper left) corner (nested pattern). The rows and columns are sorted by marginal sum in descending order from top to bottom and left to right respectively. Depending on whether

one starts the ordering process with rows or columns gives a different result. Sorting rows first gives the columns more importance and vice versa. It is also possible to sort only by row (column) sums to test for species incidence (composition) (Ulrich et al., 2009). Wright and Reeves (1992) were the first to propose order free analysis, but this method has been criticized by Ulrich and Almeida-Neto (2012) as nestedness is the result of different ecological gradients of species and sites and ordering by incidence ignores this fact. Sorting with respect to a special gradient might lead to a conclusion as to what caused the nested pattern in the given network. Ecological gradients might influence each other and therefore the degree of nestedness. The sorting of rows and columns allows to test for whether or not the according gradient contributes to any nestedness detected. This was suggested by Lomolino (1996). Sorting according to area might give conclusion on the ordered extinction and ordering by isolation on the influence of colonisation (Lomolino, 1996). Therefore it is essential to pay close attention to the sorting of rows and columns as they might determine the degree of nestedness profoundly. As matrix order is an important part of the detection of the cause of nestedness it only seems logical that the metric used for the analysis must be sensitive to matrix order. Some software has a last ordering process, which depends on the metric used to quantify the nestedness (Ulrich, 2006b). Depending on the desired results, one can order the matrix along different gradients such as environmental influence or taxon specificities. Ulrich et al. (2009) suggested that ordering along gradients should be further explored.

Table 1.5.: Different null models and their constraints. “bin” stands for binary, “abu” for abundance matrices. Missing descriptions are due to lack of information in literature.

Name of matrix	Row constraint	Columns constraint	Fixed 0	Size	Fill	Type	Reference
EE, P, ER, SIM1, R00, CRO	no	no	no	no	no	bin	Ulrich et al. (2009); Almeida-Neto et al. (2007); Guimarães (2006)
FF, SIM9	yes	yes	NA	yes	yes	bin	Strona et al. (2014); Ulrich et al. (2009)
EF, LI, FC, SIM3, FCT	no	yes	NA	NA	NA	bin	Beckett et al. (2014); Ulrich et al. (2009); Guimarães (2006)
FE, COL, FR, SIM2, R0, Random0, FRT	yes	no	NA	NA	NA	bin	Ulrich et al. (2009); Guimarães (2006)
SS	no	no	no	fixed	fixes	bin	Staniczenko et al. (2013); Beckett et al. (2014)
CC	no	no	no	fixed	fixed	bin	Beckett et al. (2014)
Fixedincidence proportional	fixed	proportional to species incidences	NA	NA	NA	prob	Ulrich et al. (2009)
Randnest Abundance-proportional	none	proportional to species relative abundances	NA	NA	NA	prob	Ulrich et al. (2009)
Incidence-proportional, SIM8, Model2	proportional to species incidence	proportional to species incidence	NA	NA	NA	prob	Ulrich et al. (2009)
E-Proportional	proportional to species incidences	equiprobable	NA	NA	NA	prob	Ulrich et al. (2009); Almeida-Neto et al. (2007)
Recol, Proportional	proportional to species relative abundances	proportional to carrying capacities	NA	NA	NA	prob	Ulrich et al. (2009)
CE, DD, PRC, APO	probabilistically by node degree	probabilistically by node degree	no	fixed	fixed	bin	Bascompte et al. (2003); Joppa et al. (2010); Almeida-Neto et al. (2007); Strona et al. (2014); Guimarães (2006)
PP	proportional	proportional	NA	NA	NA	NA	Ulrich and Gotelli (2012)
Binary shuffel	no	no	yes	yes	yes	abu-fixed	Beckett et al. (2014); Staniczenko et al. (2013)
CRT	yes	no	NA	yes	yes	abu-free	Beckett et al. (2014)
CCT	no	yes	NA	yes	yes	abu-free	Beckett et al. (2014)
RCTA	yes	yes	yes	yes	yes	abu-free	Beckett et al. (2014)
rc	yes	yes	NA	yes	yes	abu-free	Ulrich (2010)
aa	NA	NA	no	no	no	abu-fixed	Ulrich (2010)
is	NA	NA	NA	NA	NA	NA	Ulrich (2010)
ss	NA	NA	NA	NA	NA	NA	Ulrich (2010)

## 1.8. Z-score

To see whether the measured degree of nestedness in the original matrix is in fact significant, the metric distribution of the null matrices is calculated. The distance of the original value from the mean of the distribution provides information on whether or not the degree of nestedness in the original matrix is significant. This distance is called z-score, which is measured in standard deviations. A z-score greater than two indicates significant nestedness if the metric scores a high degree of nestedness with a high value. Conversely, if the metric uses low values for a high degree of nestedness, then a z-score lower than two indicates significant nestedness. If the z-score is equal to zero, the original metric is the mean of the distribution. Inappropriate null models will generate distributions that always detect nestedness where there is none (type 1 error), or never detect it if there is (type 2 error). Thus, detected significance can either be correctly caused by significant nestedness in the original matrix or falsely because the null model is not constrained clearly enough.

The z-score is calculated as follows:

$$Z = \frac{E_{obs} - E_{exp}}{\sigma_{exp}} \quad (1.1)$$

where  $E_{obs}$  is the original metric value,  $E_{exp}$  is the mean of the sums of the null model metric values and  $\sigma_{exp}$  is the standard deviation of the null model distribution. (Ulrich et al., 2009).

## 1.9. Software

“A nestedness beginner would probably have a hard time in discriminating among the 78 potential different combinations of settings (and the respective Z values) obtainable with nestedness.” (Strona et al., 2014, page 6)

Nestedness is a commonly investigated pattern. However, acquiring a deep understanding of all metric-null model combinations is timeconsuming if not impossible. These are the reasons that a variety of software exists, that calculates nestedness. The software packages are limited in metric and null model options and only offer the most commonly used types. This limits the possibility of choosing a metric-null model combination more suited for the specific needs, but it offers a comfortable way of conducting a fast analysis of the data. Table 1.6 summarises the most commonly used software and the metric and null model options they provide.

The program “NODF” additionally offers four sorting options (Almeida-Neto and Ulrich, 2011). The software “Pairs” additional includes metrics for other network properties (Ulrich, 2008). There is software that solely calculates nestedness (SDRSimplex) (Podani, 2001) and needs to be combined with other software meant for the generation of null models (ecosim, matrix) (Ulrich, 2006a), NestRand (Podani and Schmera, 2012), NestTest (Podani and Schmera, 2012)). The SDR-Simplex also includes a version for abundance data (?). WINE positions the matrix value between the perfectly random and the maximally nested matrix, both generated from the original matrix. 0 indicating randomness, 1 perfect nestedness (Almeida-Neto and Ulrich, 2011).

Table 1.6.: Software and metrics/null models. This table makes no claim to completeness.

Software	Metrics	Null model	Weighted	Year	Empty rows/- columns	Program	Examples
NTC	$T$	Random00, EE	no	1993	no	NA	Atmar and Patterson (1995)
Nestedness	$N_0, N_1, N_c, UA, UP, UT, NODF, BR, MT$	FF <sub>sa</sub> , EE, FE, EF, Fixedincidence proportional, sampling model, 7th model	NA	2006	yes	FORTRAN	Ulrich (2006b)
BINMATNEST	$T$	FF, EE, EF	no	2006	no	NA	Rodriguez-Gironés (2006)
ANINHADO	NODF, T	ER, CE, COL, LI	no	2006	NA	C++	Guimarães (2006)
NeD	NODF, $T, BR$	FF, EE, FE, EF	no	2014	both	Python, Django	Strona et al. (2014)
FALCON	NODF, MD, NTC, JDM, BR, WNODEF, SR	SS, FF, CC, DD, EE, Binary shuffle, CRT, CCT, RCTA	both	2014	NA	Matlab, Octave, R	Beckett et al. (2014)
WINE	WIN	Binary shuffle	yes	2009	NA	Matlab	Galeano et al. (2009)
NODF	WIN, NODF, WNODEF, BR, modified BR, modified $T$ ,	FF, EF, FE, EE, Incidence-proportional, rc, aa, ss	both	2012	yes	FORTRAN	Ulrich (2010)
Pairs	BR, C-score, checkerboard score, others	FF <sub>sa</sub> , EE, FE, EF, Fixed-incidence proportional, sampling model, 7th model	both	2013	NA	FORTRAN	Ulrich (2008)

## 1.10. Moduled compartments in networks

Lewinsohn et al. (2006) suggested that networks should not be tested for only one nested pattern. They proposed patterns that occur along a gradient or compartmented pattern. The latter is represented by subsets that have a high interaction density within themselves but rare or no interactions between each other. In an interaction matrix these compartments are easily visible as clusters of filled cells that are clearly distinct from other such clusters. These clusters appear along the main diagonal of the matrix.

Compartmentalisation was thought of to be very unlikely to occur in food webs. This was recently falsified after reanalysis and by applying a new detection method (Kondoh et al., 2010). Generally spoken, one would expect a higher chance of detecting a compartmentalised pattern in networks with low diversity but high specificity interactions. This might be the case in co-evolved mutualistic networks, where systems highly depend on one another but not so much on others. These compartments can evolve into their own evolutionary cell in which further co-evolution or even further compartmentalisation can occur. This co-evolution might be the reason why one would observe more turnover in a nested pattern than expected between two different compartments (Lewinsohn et al., 2006). Conversely, we expect one nested compartment for high diversity but low specificity in a mutualistic network (Bascompte et al., 2003).

Given a high degree of nestedness, compartmentalisation only occurs in matrices with low connectance. With increasing connectance and simultaneous high degree of nestedness, compartmentalisation decreased (Fortuna et al., 2010). There is evidence that antagonistic systems show compartments although they also show high connectance. Such a pattern was found in antagonistic networks by Kondoh et al. (2010) in multi tropic networks. These clusters do not forbid nestedness as each cluster can show its own nested structure within the cluster. Therefore, a further test for nestedness within each compartment should be done. It was suggested that a network should first be analysed for which pattern occurs and not solely whether a network shows one distinct pattern. Fortuna et al. (2010) also found it to be important to test for different patterns to better understand the complexity underlying the architecture of networks. This implies that only communities with low connectances are likely to simultaneously present nested and modular patterns. They also proposed that a clustered structure increases the persistence of a network, as they retain perturbation within themselves and prevent influences on other compartments. This would suggest that one could analyse each cluster as an isolated network. However, this is not possible because rare interactions between clusters are often even very important for the entire network (Lewinsohn et al., 2006).

## 1.11. Possible conclusions

Inferences can be drawn from nestedness in a network. It allows for example to predict the order of extinction or colonisation. Moreover, even the influence of the extinction of one species on the whole system. Patterson and Atmar (2000) see missing species at the presence site as ideal candidates of reintroduction, whereas presences on the absence site struggle to keep their existence. They also conclude that a population lifetime is predictable by their position in the matrix. In metacommunities species are increasingly more likely to survive if they are closer to the most hospitable site, meaning that species that lie on the boundary line between absence site and presence site are endangered. Specialist species living on hospitable sites and generalist species living in hostile sites are most prone to extinction. Nestedness can also be used to predict occurrence of species. That is, if most rare species D is present at a site, it is likely that the more common species A,B and C are also present (Cutler, 1994). It is no longer only used as a way to describe colonization or extrication order but might provide insight into the evolutionary

process that led to today's assemblage. It can also give information about species distribution or serve as a tool to infer species preservation. (Patterson and Atmar, 2000).

There is an ongoing debate called SLOSS (single-large-or-several-small) about conservation of biodiversity: is it better to protect one large island, that should include all species, or several small islands that together form the equivalent species assemblage to the large island? If one assumes that networks are nested and the species richest island contains all species of any subset of a smaller island, it would be logical that the conservation of this single large island would be sufficient to maintain the biodiversity. However, perfect nestedness is very rare and the only case in which the protection of the large island is truly useful. In any case that deviates from perfect nestedness, as most natural networks do, several smaller islands contain an overall larger number of different species. Therefore it would be desirable to protect areas according to their specific circumstances and not as predicted by an algorithm (Ulrich et al., 2009; Patterson and Atmar, 2000).



## 2. Methods

As a first step, the considerable variety of approaches to the topic with different null models and metric was reviewed to gain an overview of the task at hand. The metrics for further analysis were selected on the basis of their performance and frequency of usage according to the literature.

For the analysis, two R packages were used bipartite version 2.05 (Dormann, 2008) and vegan version 2.2-0 (Oksanen et al., 2014). The bipartite package was slightly altered before the analysis and changed to include two null models from the vegan package. The altered code can be found in Appendix A. Other programming done in R is attached to this study as a CD. The analysis included ten different metrics and seven different null models (see section 2.1).

Different kinds of original matrices for the analysis were produced. The 23 empirical matrices of the bipartite package as well as 150 matrices with three different sizes (small, medium, large) and 50 moduled matrices (density 2, log normal) were chosen for the analysis. Each moduled matrix showed three different compartments of which each was generated nested. The genweb function of the bipartite package, which uses a log normal distribution, was used to generate marginal totals. 50 small matrices were produced with the r2d null model of the bipartite package. All matrices used abundance data. All matrices except the 23 for the empirical networks were generated nested. An example of a small a medium and a big matrix are shown in figure 2.1, 2.2, 2.3. An example of a moduled and an observed matrix are shown in Figure 2.4 and 2.5.

As a first step of the analysis, all seven metrics for each of the 273 original matrices were calculated. Then, all ten null models on these original matrices to generate 25 null matrices for each original matrix. Again all seven metrics were calculated for each new null matrix. From each null model-metric combination a z-score was obtained using the metric value of the original matrix and those of the null matrices calculated with the same metric. All 273 z-scores of the same metric-null model combination were used to obtain a z-score distribution for this particular combination. This was done for all 70 possible null model-metric combinations. It was accounted for the different scaling of the metrics. The seven metrics applied on the 273 matrices returned 1911 original metric values (2.1). The overall number of metric values for all matrices-metric-null model combination was 477 750 (2.2). 19110 z-scores were calculated using the z-score produced by each metric null model combination for each matrix(2.3).

$$273 * 7 = 1911 \quad (2.1)$$

273 original matrices. Seven metrics.

$$273 * 10 * 25 * 7 = 477750 \quad (2.2)$$

273 original matrices. Ten null models. 25 null matrices. Seven metrics.

$$273 * 10 * 7 = 19110 \quad (2.3)$$

273 z-scores of the metric null model combinations. Ten null models. Seven metrics.

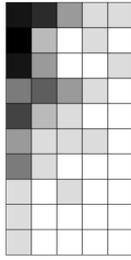


Figure 2.1.: Examlle of a small matrix used in this analysis

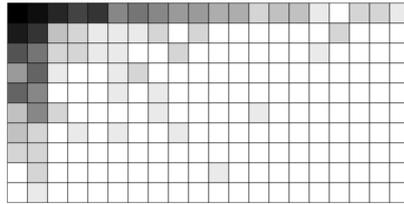


Figure 2.2.: Examlle of a medium matrix used in this analysis

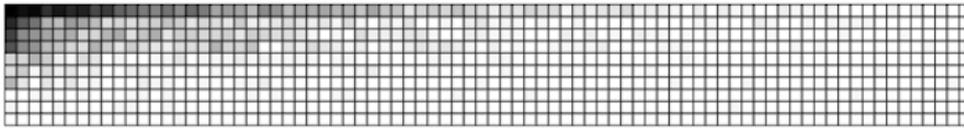


Figure 2.3.: Examlle of a big matrix used in this analysis

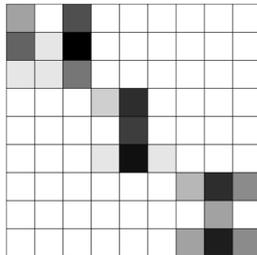


Figure 2.4.: Examlle of a moduled matrix used in this analysis

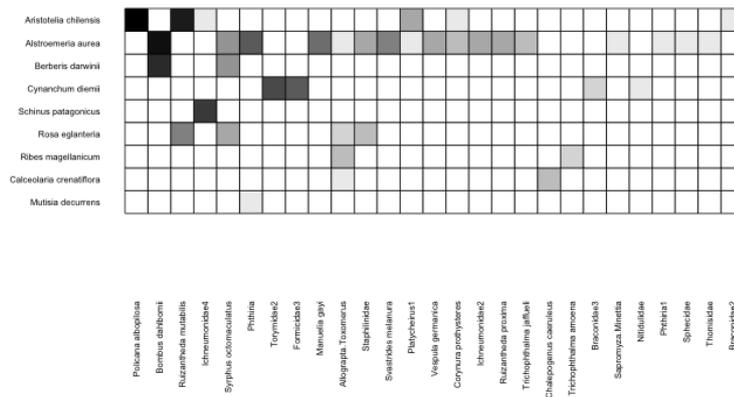


Figure 2.5.: Examlle of an observed matrix used in this analysis

Table 2.1.: Metrics used and their equivalent in literature

R name	name
binmatnest	$T$
binmatnest2	$T$
discrepancy	BR
discrepancy2	BR
C.score	C.score
checker	C.score
NODF	NODF
NODF2	NODF
weighted NODF	weighted NODF
wine	WIN

Table 2.2.: Null models used in this analysis. Showing constraints and freedom

Name	Constraint	Free	type
swap.web	connectance,marginal sums,marginal frequency	NA	abu
r2d	marginal sums	connectance,marginal frequency	abu
vaznull	connectance,marginal-probability	NA	abu
quasiswap	fill,marginal sums	marginal frequency	abu
mgen	different probabilities	connectance,marginal sums	abu
swsh	fill,marginal-frequency,row sums	NA	abu
shuffle.web	dimension,connectance,fill	marginal sums	bin/abu

## 2.1. Measures of nestedness and types of null models

To compare the intrinsic scaling and consistence in the definition of nestedness by the metrics, a scatter plot and the Pearson correlation were used. The same comparison was made both for the original matrices and after the application of null models. Table 2.1 shows the metrics and table 2.2 the null models that were used. Figures 3.1 and 3.2 show the results of this analysis.

## 2.2. Statistical Analysis

The z-scores of all ten null model types for all seven metrics for the 273 original matrices was calculated. It was accounted for the different scaling of the metrics by calculating the z-score distribution of the three NODF metrics like

$$Z = \frac{[100 - E_{obs}] - [100 - E_{exp}]}{\sigma_{exp}} \quad (2.4)$$

A similar correction was made for the wine metric.

$$Z = \frac{[1 - E_{obs}] - [1 - E_{exp}]}{\sigma_{exp}} \quad (2.5)$$

This is necessary to compare the outcome. Mean values below -2 indicate significant nestedness for all metrics after correction, mean values above 2 indicate significant non-nestedness

for all metrics. An ANOVA analysis was conducted for all matrix types to find the main reason of variance. The different null model-metric combinations were compared to find differences in their performance. Boxplots were used to visualise the results. A scatter plot was used to investigate correlation between different metrics before and after applying a null model.

## 3. Results

### 3.1. Statistical Analysis

It was found that some null models' influence was extremely severe regardless of the metric, while others' performance was influenced by the size of the matrices. The analysis showed that for moduled matrices, no satisfying null model-metric combination exists. The correlation of the different metrics showed only very little intrinsic consistency of the metrics and was even less after applying the null models. The ANOVA identified null models or their combination with a metric to be the main reason of the explained variance.

#### 3.1.1. Metric correlation

The correlation of the different metrics showed that they are very inconsistent with their intrinsic definition of nestedness. Metrics that use the same calculation method were highly correlated. These are the two binmatnest metrics, the three NODF metrics and the two discrepancy metrics. The NODF and the weighted NODF showed a higher positive correlation than NODF and NODF2 did. The binmatnest metrics showed least positive correlation with the other metrics. This is not surprising as other studies already showed that this metric depends on the matrix size, fill and dimension (Ulrich et al., 2009). There was no or even negative correlation for the binmatnest2 metric with all the other metrics except with wine. It was highly negative correlated with both discrepancy metrics. Binmatnest showed very little positive correlation with NODF and weighted NODF but 47% correlation with checker. It shows, however, a fairly high positive correlation with wine. Both discrepancy metrics showed negative correlation with wine, they were correlated positive with all the others, but showed only little correlation with checker and medium correlation with all three NODF metrics. They were somewhat correlated with the C.score. C.score and checker were only moderately correlated. While the C.score showed high correlation with all three NODF, checker showed no or moderate positive correlation with them. Cscore showed moderate positive correlation with wine, checker only little. All three NODF metrics were somewhat positively correlated with wine.

The result of the same analysis after null model application showed even less correlation. Only C.score and binmatnest were correlated, as well as checker and binmatnest2. Both discrepancy metrics still showed the highest correlation. There were only 10 negatively correlated metrics before null model application, but 21 after. NODF and weighted NODF, which had 0.99 correlation before, now only showed a weak correlation of 0.39. The high correlation between C.score and the NODF metrics almost inverted itself and now was a negative correlation. Figures 3.1 and 3.2 show the detailed results.

#### 3.1.2. Null models

It was assumed that the margin constraining r2d null model would return a set of null models that were very similar to the original matrix. This should especially be the case for matrices that were constructed using marginal totals as a way to generate nestedness. This was true for all the artificially generated matrices as a log-normal marginal distribution was used to create the fill probability for each row and column. First, 50 matrices were generated using this method,

and the r2d model was applied on those 50 matrices, generating one null model for each of the 50 matrices. These 50 new null model generated matrices were used as 50 new original matrices. Therefore, applying a row and column sum constraining null model (r2d) on row and column sum constraint generated original matrices should result in a mean z-score distribution of 0 and a standard deviation of 1 as marginal totals are supposed to have a high influence on all metrics. It was assumed that the metric values returned by the r2d null models would be normally distributed. The analysis of the 50 original matrices that were generated using the r2d null model showed the expected distribution when applying the r2d null model. The mean values for all distributions were almost 0, while 65% of each distribution was between 2 and -2. The small webs, which were used as the original matrices to produce the r2d generated matrices showed very similar results and therefore further use of the r2d generated matrices was discarded. This result shows that the r2d model produces very little distortion and the metric distribution deviates only little from the original metric. A comparison of the small and the r2d generated distributions is given in figure 3.3.

The analysis of the differently sized matrices showed that null models and metrics differed very little for different sizes. Only the swsh null model indicated significant nestedness. All other null models yielded means close to 0. The r2d model showed a slightly wider z-score distribution for decreasing size, but had always a mean very close to 0 for all metrics. Means were, however, slightly lower with increasing size. Swap.web gave a somewhat lower mean with increasing size, yet still did not show high significance with all metrics. Vaznull showed a greater difference between the metrics with increasing size, the weighted NODF reacted strongest and indicated higher segregation for bigger networks, as did wine. C.score and checker differed by almost half a standard deviation and increased their difference with increasing size. The other metrics did not react strongly. Mgen had very similar returns for the small web, but lost this similarity with increasing size. C.score and checker differed strongest for big webs. Except for C.score all metrics had a lower z-score distribution with increasing size. Wine had its lowest mean for medium sized webs. All metrics reacted evenly with quasiswap and had lower means with increasing size. Of all null models, swsh showed the highest variation over the metrics. The discrepancy metrics had highest means for small webs, this changed with increasing size and both binmatnest and C.score, as well as the two unweighted NODF had higher values with the big networks. This happened with decreasing discrepancy values and increasing values for the other metrics mentioned. Checker behaved similar to the discrepancy metrics. Weighted NODF and wine reacted very severely to the different sizes. For small webs, both had the lowest means observed, wine being the overall lowest. For the weighted NODF, this trend went on with increasing size and showed significant nestedness for medium and big webs. Wine reacted contrarily. It had the smallest mean for small webs, but steeply increased with increasing web size, showing significant segregation for the big webs. I excluded the shuffle.web null model from my findings as it indicated extremely high nestedness independent of web or metric used.

### 3.1.3. Observed matrices

The observed matrices showed a higher dependency on the null models than the different sized matrices did. The r2d model reacted very similar to the mgen model. Both returned significant segregation for all NODF metrics, the wine metric and the C.score, while all other metrics identified no significance. The swap.web model returned similar results to the vaznull model. Both identified no significance for all metrics except for the combination of the weighted NODF and wine with vaznull. This combinations identified significant segregation. The quasiswap model identified significant nestedness with all metrics. The swsh null model had means very close to zero, identifying significant nestedness only for the weighted NODF. The results are shown in figure 3.5.

### 3.1.4. Moduled matrices

The z-score results for the moduled matrices showed a high tendency for all null model-metric combinations to detect significant segregation. Only the combination of the swsh null model with the weighted NODF identified a significant nested pattern. All other models either showed no significance or highly significant segregated patterns. The r2d model as well as the mgen exclusively returned high segregation. The swsh model and the quasiswap returned results closer to zero, but still did not indicate significant nestedness except for the above mentioned case. Within each null model, there was a high variation between the different metrics. All metrics appeared to depend strongly on the null model, but the two weighted metrics, weighted NODF and wine, showed a tendency to identify less segregation than the other metrics, independent of the null model. The swap.web null model and the quasiswap returned very similar results, swap.web indicating a little more segregation for each metric. The results are shown in figure 3.4.

### 3.1.5. ANOVA

The results of the ANOVA for each matrix type identified the null models as the main reason of variance for all matrix types except for the empirical, where 66% of variance was still unexplained. The null models explained 50% for the moduled and small matrices, and 76% for medium sizes but only 70% for big matrices. No trend of higher variance by the null model for bigger matrices was observed. Medium and big matrices differed mainly in explained variance by null model-metric combination, which was only 7% for medium but 17.3% for big webs. All results changed drastically, when the shuffle.web null model was removed from the analysis. The shuffle.web null model is suspected to create such a high variance between the null models because it was already known to cause extreme z-score distributions in other analyses. After removing shuffle.web, all except the moduled matrices had a very high unexplained variance. A trend was visible for the different sized matrices. Increasing size led to increasing explained variance by the null model-metric combination, which reached 34.47% for the big matrices. For the moduled matrices, the null model still explained most variance with 49.7%. The only matrix type that reacted differently to the exclusion of shuffle.web was the observed ones. Explained variance increased to almost 50%, when it had only been one third with shuffle.web included.

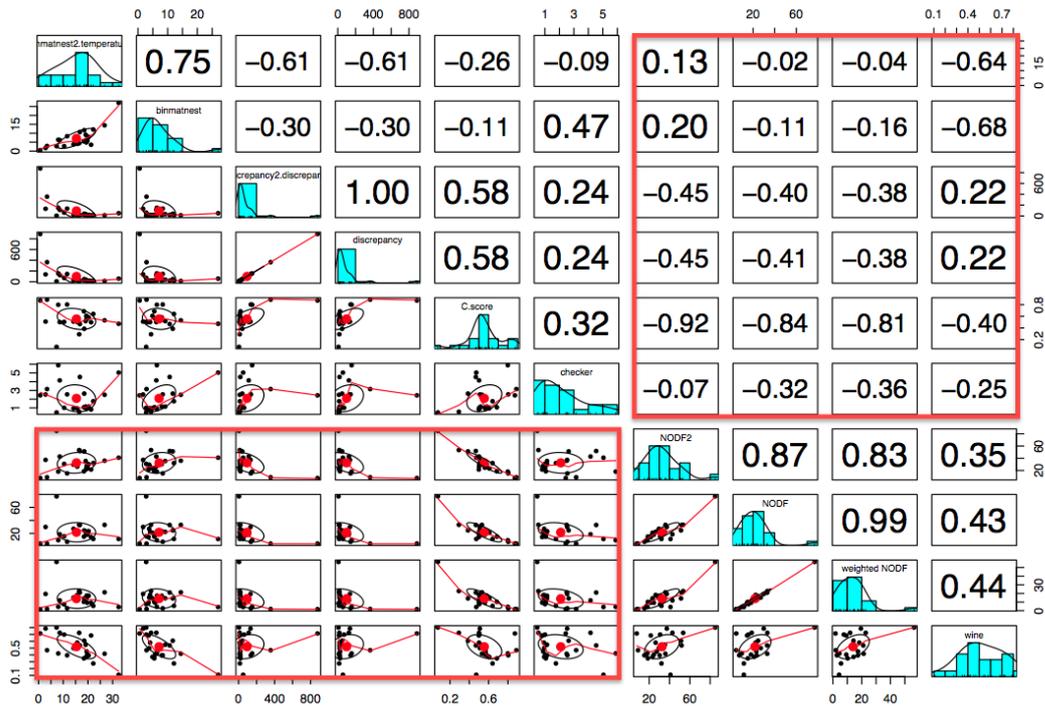


Figure 3.1.: Scatter plot and Pearson correlation for the different metrics. Different scaling is not accounted for, values within the orange box thus have to be multiplied by  $-1$  to get the standardised values. The table is “symmetric” around the diagonal. Since the values above the diagonal are calculated from the corresponding scatter plots, the plots in the marked lower left part have to be understood “in reverse”.

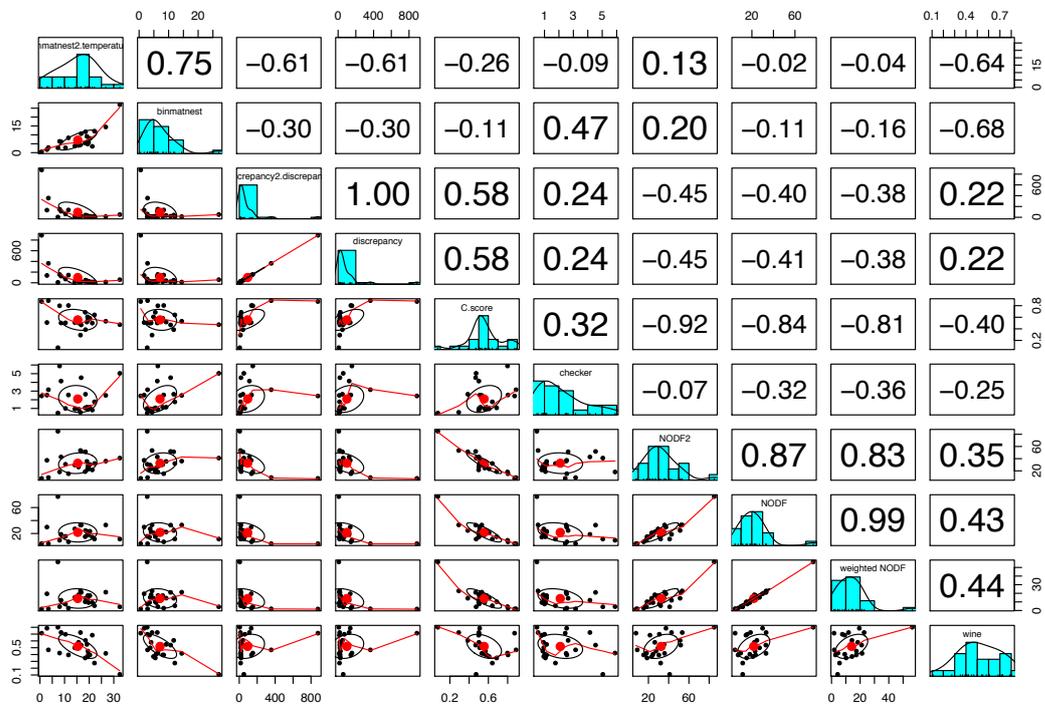


Figure 3.2.: Scatter plot and Pearson correlation for the different metrics after null models were applied. Scaling is accounted for, the values are standardised.

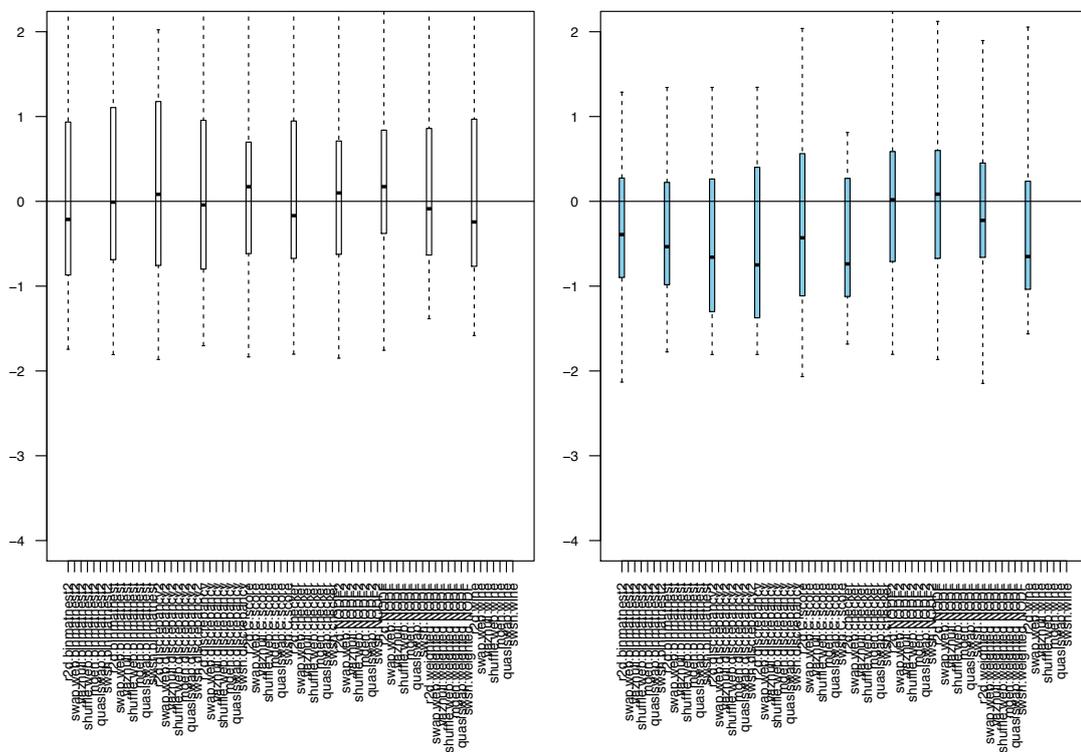


Figure 3.3.: Results for the small matrices (on the left) and the r2d generated matrices (on the right) for all metrics with the r2d model.

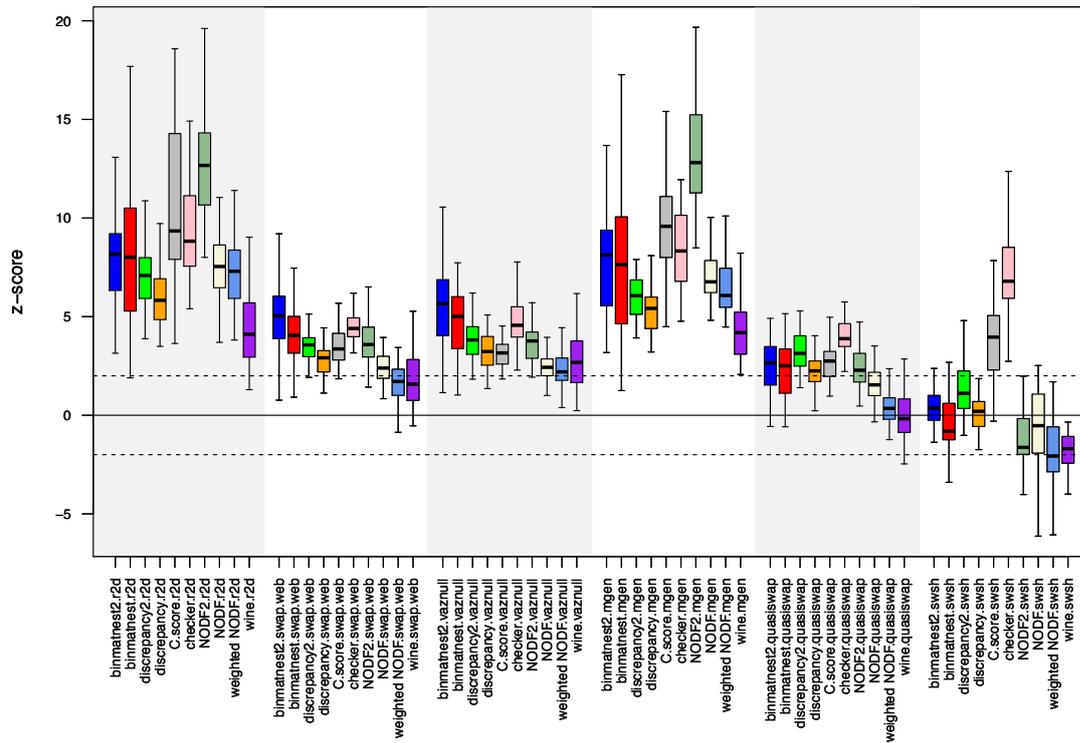


Figure 3.4.: Metric-null model combination of the modulated matrices

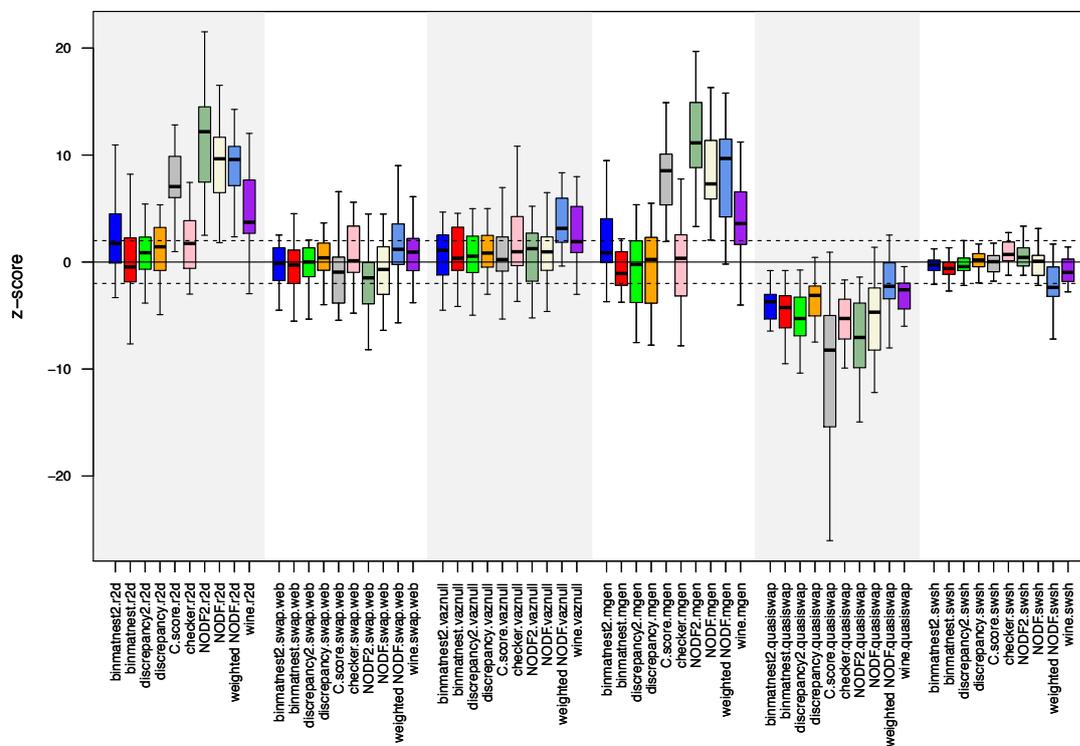


Figure 3.5.: Metric-null model combination of the observed matrices

## 4. Discussion

### 4.1. Literature

#### 4.1.1. Metric correlation

There are many metrics available, which combined with a null model leave the ecologist with a wide amount of possible combinations. Choosing the right metric in combination with the right null model is as crucial as it is difficult. Many different metrics have been analysed with many different null models, yet no conclusion has been reached to what is the best combination. The ecologist has to decide this for the specific network he wants to analyse and depending on the question that drove him to the analysis in the first place. However, there are metrics that have proven more suitable than others. BR, NODF and checkerboard seem to be the overall best performing metrics for binary networks. Despite its popularity, the temperature metric  $T$  does not perform well and should not be used in further analysis. Its only reasonable application seems to be, when comparing results with existing studies that use  $T$ . As suggested by Almeida-Neto et al. (2008), a reevaluation of the conclusions drawn of analysis basing on the  $T$  metric seems appropriate. However, Gotelli and Ulrich (2012) found that the same matrix, analysed with the same null model returned very different results for BR (nested), NODF (anti-nested), and  $T$  (random). This shows how essential further analysis and careful choice of metric are. Among the metrics used for analysis of abundance matrices, WNODF and WIN seem to be performing well, but further analysis of their performance is necessary. SR appear to be a very promising new metric, but has been analysed to little to draw final conclusions.

#### 4.1.2. Null models

More analysis of the influence of abundance data on nestedness is to be made, as it appears to be a very interesting new approach that deserved further attention. The different null models differ highly in their performance and some do not seem to be justified (Joppa et al., 2010). Null models should represent only one clearly defined ecological factor, so that the results can be interpreted equally clearly. This is not the case for very liberal null models and it is fruitless to use a null model, that always returns high significance for any network (Ulrich et al., 2009). Detecting highly significant nestedness in a network is easy using such a model but the cause of nestedness remains obscure. Such null models should therefore not be used and instead, closer attention should be paid to those that allow for a clear statement on processes that cause nestedness (Ulrich et al., 2009; Ulrich and J Gotelli, 2007).

### 4.2. Statistical Analysis

#### 4.2.1. Metric correlation

A comparison of different nestedness metrics requires the metrics to return the same relative degree of nestedness for the same matrix. That is, all metrics should score a more nested matrix higher than a less nested one, regardless of the metrics' calculation processes. In other words, all metrics should be highly correlated in the resulting degree of nestedness if they are applied

on the same matrices. If a metric is not correlated with the others, it should not be used for comparison. The question is: are all the metrics consistent in intrinsic scaling of nestedness? And if so, why does one need the wide variety of different metrics instead of a single one? To answer these questions, a scatter plot was produced and the Pearson correlation was calculated for each metric combination for the 23 empirical matrices. The same was done after applying the different null models.

Suppose that a Pearson correlation of at least 0.75 indicates sufficient correlation<sup>1</sup> and excluding all metric correlation of metrics using the same calculation process, the only sufficient correlation was found for C.score with NODF and weighted NODF. The unexpected missing correlation between C.score and checker suggests that there is an error in one of their calculation processes in the bipartite package since checker is merely the non-normalised version of C.score<sup>2</sup>. Given the discrepancy between C.score and checker, the correlation between C.score and NODF and weighted NODF has to be ignored. Thus, no two metrics show sufficient intrinsic consistency. There is no sufficiently strong correlation between intrinsic scaling to claim that these metrics measure the same pattern. Clearly, they do not agree on what nestedness is and therefore a comparison of their results seems impossible. This becomes even more prominent after application of the null models. Before applying the null models, ten combinations of metrics showed a negative correlation. After applying the null models, there were 21 negatively correlated combinations.

The fact that correlation was even less after applying null models shows that null models distort webs in such a way that it becomes even more difficult to produce unambiguous results with the metrics. Each of the null models restrains different properties of the original matrix. This distorts the matrix to a different extent. The metrics, in turn, weigh the properties of a matrix differently and react more or less sensitive to the distorted properties. One changed property can have a strong impact on one metric, leading to a greater distinction in the level of significance for this metric. Another metric, on the other hand, might be much less sensitive to this property returning less significance. Whether or not sensitivity to a specific property is desirable depends on the definition of nestedness and on to what extent this property contributes to the definition of nestedness.

It is therefore suggested that the definition of nestedness and its calculation be clearly defined. Subsequently, all metrics should be revalued. Furthermore it is proposed that all studies using different metrics to compare results be reanalysed using the same well defined metric.

#### 4.2.2. Null models

There are many different metrics to calculate nestedness and many different null models to test the metrics' significance. Ten metrics and seven null models were applied on 273 matrices that differed in size, degree of nestedness, and compartmentalisation. The mean of the z-score distribution for the 50 matrices that were generated with the r2d null model, was expected to be close to zero for all margin constraining null models. This was the case and all original matrices were identified as not significant. The small webs, that were the basis for the r2d generated matrices showed almost the same distributions. The only null model that does not constrain margins is shuffle.web. This model identified all matrices with all metrics as highly significantly nested. This shows that marginal totals play a major role in the identification of nestedness and that failing to constrain such an important contributor will always lead to identifying high nestedness, independent of metric. Using such a null model is not useful and there is no network imaginable

<sup>1</sup>The author claims that an even higher value is necessary to signify a correlation.

<sup>2</sup>This calculation error has been corrected during this analysis. The corrected version of the bipartite package is available on github.

where all parties interact randomly with each other. This null model was therefore ignored in further analysis.

Of the three different matrix sizes, that were all designed nested, significant nestedness was only identified by the combination of swsh and the weighted NODF for medium sized and big webs. All other null model-metric combinations returned mean values and 65% of the distribution within the range of 2 standard deviations from zero. This leads to the conclusion that the factor that generated the nested pattern in the original matrix was mainly conserved by all null models.

As the nestedness arose from a log-normal marginal distribution, it was expected that null models constraining marginal totals would not identify significant nestedness. This explains the distributions returned for the marginal constraint null models r2d, quasiswap, swap.web and partly for the row constraining model swsh. The null models mgen and vaznull showed more sensitivity to the null models when web size increased, but were still within the non significant interval. This can be interpreted as the result of them constraining marginal probabilities: they allow for a little more alteration of the original matrix, while keeping the probabilities of the original matrix.

#### 4.2.3. Observed matrices

The distributions of the observed networks were very inconsistent for the different null models. While r2d and mgen tended to identify them as segregated, quasiswap identified significant nestedness for all of them. The null models swsh, vaznull and swap.web tended to identify no significance. The null models swap.web and quasiswap were expected to return similar distributions, because they use very similar ways of calculation. This was not the case. They differed strongly in their returns. This leads to the conclusion that the structure of the observed matrices is highly sensitive even to small differences in the generation of the null matrices. The unexpected and severe difference between these similar null models should be further investigated.

All NODF metrics, the wine metric and the C.score identified significant segregation in combination with the r2d and the mgen null model. They showed no distinct difference to the other metrics in combination with the remaining null models. This suggests that these metrics react more sensitive to the alterations by these null models than the other metrics do. It also shows that the alterations by the remaining null models do not influence these metrics much different than the other metrics. The r2d model fully constrains margins, while the mgen only keeps probabilities constant. The similar results of these two models seem to suggest that constraining marginal sums is not as important a factor as previously thought. This idea is supported by the similarity between the margin constraining null model swap.web and the probability constraining null model vaznull. This might lead to the conclusion that constraining marginal probabilities is sufficient and that there is another property, related to marginal probability, that is more important than constraining marginal sums. This should be further analysed.

As both r2d and swap.web constrain margins and therefore decreasing fill, it must be the difference in paired overlap that causes the different reaction of the unweighted NODF metrics on these two null models. The identification of significant segregation of NODF with the r2d model but not with the swap.web suggests, that r2d creates null matrices with a much higher number of paired overlap than swap.web does. This can not be translated to the probability constraining null models vaznull and mgen, as they can also have a chance of decreasing fill.

The return of the row constraining null model swsh is different than expected. This model created the biggest differences between the different metrics for the other types of matrices in this analysis. Contrary to these findings it returned the most moderate results of all null models for the observed matrices. Means were very close to zero for all metrics but the weighted NODF and the wine. This might be a reason to over think this model, as it seems to have reasonable constraints for observed networks, although this was not seen when applying it to artificially

produced matrices.

#### 4.2.4. Moduled matrices

None of the used null model-metric combination identified the moduled matrices as nested. There was a tendency of the weighted metrics to identify more nestedness than the others. The row constraining null model *swsh* surprisingly returned distributions closest to a mean of zero.

The NODF metrics cannot detect any nestedness within the original matrix, as there is almost no chance of decreasing fill. In case there is decreasing fill (row pair between different compartments), there is no chance of paired overlap. This will always result in no nestedness detected even though it exists in the matrix. By applying a null model, even those that use constrained margins, paired overlap can occur in those rows that showed decreasing fill in the original matrix and NODF will almost always detect higher nestedness than in the original matrix. Therefore using NODF on moduled matrices should always result in detecting a highly segregated pattern in the original matrix. This is not the case for the weighted NODF because difference in abundance allows decreasing fill and therefore also counts paired overlap within compartments. The less constrained the null model, the more likely that nestedness can be identified. This explains the high segregation indicated for NODF by *mgen*, that does not constrain marginal sums. The high segregation of the margin constraining null model *r2d* can only be explained by NODF being highly sensitive to change in paired overlap. This is supported by a comparison with *swap.web*. This null model constrains the positions of zeros additionally and therefore does not allow for change in paired overlap. *Swap.web* indicates much less segregation than *r2d* and therefore shows the influence of paired overlap on NODF. The reason for *vaznull* identifying only a moderate amount of nestedness is unclear because this null model is less constraining than *r2d*. Since *swap.web* and *quasiswap* are very similar, it was assumed that they return very similar results. This was confirmed by the analysis.

A similar conclusion can be drawn about the checkerboard scores. In a compartmentalised matrix, a checkerboard pattern only occurs  $n - 1$  times, where  $n$  is the number of compartments ( $q - 1$  times when all compartments consist of only one species and  $q$  is the smallest of either row or column number). This will result in a high degree of nestedness. This is because the relation of the total number of cells to the number of rows or columns will be very high in most cases. Applying a null model will increase the probability of checkerboard pairs and therefore a lower degree of nestedness will be detected. A z-score distribution indicating significant nestedness should be returned. This did not agree with the findings as both checkerboard scores indicated high segregation for each null model.

It is impossible to infer the other metrics' returns, as they are highly dependent on the placement of the compartments within the matrix. Including abundance data clearly has an influence on the NODF metric. Including abundance data allows the detection of nestedness in the original matrix and therefore moderates the effect of the null models. This is supported by the findings that weighted NODF always indicated less segregation than NODF and NODF2. Since a binary metric using the Manhattan distance for calculation was not included, it is not possible to unambiguously trace back the moderating performance of wine to either using abundance data or the Manhattan distance.

Deviations from the expectations might be explained by the low density of the moduled matrices, allowing for zero entries in the compartments. A reanalysis using completely filled compartments might give deeper insights. It seems that there is no good way to constrain the compartmentalisation of these matrices because even the marginal constraint models have no power in maintaining this structure. A null model that constrains the placement of zeros would keep the compartmentalisation but result in an unchanged matrix for binary networks and highly similar matrices for abundance networks. The row constraining null model *swsh* returned unsignifi-

cant distributions. In general, a less constraining model tends to cause detectable nestedness. The behaviour of swsh is therefore counterintuitive because this null model is less constraining than others, yet does not cause much detectable nestedness. Even with constrained rows, the entries are allowed to change within the row. This is sufficient to create paired overlap, which in turn generates detectable nestedness for NODF. One would therefore expect NODF to indicate more nestedness than in the original matrix. However, this is not the case. Thus, the swsh model prevents the creation of detectable nestedness in combination with most metrics. C.score and checker are the only metrics that find segregation in combination with swsh and thereby also contradict the logical argument about the checkerboard metrics above. To summarise, swsh behaves contradictory to the logical arguments constructed for NODF and the checkerboard metrics.

There is a high variance in the results returned by different null model-metric combinations. All null models tear up the compartmentalisation. This allows for the detection of much higher nestedness in the null models than in the original matrix, especially for the NODF metrics. This leads to detection of significant segregation in any case for the NODF, and in many cases for the other metrics. Therefore, the exclusion of NODF for detecting nestedness in moduled matrices is suggested. As the other metrics suffer from similar problems, further analysis is required. The contrary should be the case for the checkerboard scores resulting in the detection of significant nestedness in most cases. As this theory is not supported by the findings, further investigation on the performance of the checkerboard scores is necessary.

A new metric should be introduced that first searches for compartments within the matrix and then analyses the nestedness of each compartment. Subsequently, each compartment's nestedness should be used to calculate the nestedness of the entire matrix.

#### 4.2.5. ANOVA

The results of the ANOVAs, that were done for the different matrix types on the variance created by the null model, metric or the combination of both, show a strong influence of the shuffle.web null model. Including shuffle.web, all matrix types had their main reason for variance explained by the null models. This changed drastically after shuffle.web was removed. This, once more, shows that the shuffle.web null model behaves extremely different than the other null models. After shuffle.web was removed, it was interesting that the null model explained only very little of the variance. Only for the moduled matrices most variance was still explained by the null models. This seems to be due to the different degree of conservation of the moduled structure. All other matrices had their main reason for variance unexplained. For all matrices the metric explained very little of the variance for all types of matrices. Only the moduled matrices had more than 10% of variance explained by metrics.

Some of the metrics were found to be highly unsuited for the analysis of moduled matrices, it was therefore expected that the metrics explain a much higher percentage of the variance, which was not the case. This might be due to the high variance of the null model and of the combination of null model and metric, since neither null models nor metrics seem to be suited for such analysis. Considering that very little consistency was found between the original metrics, it is strange that almost no variance was explained by the metrics.

Increasing the size of the web led to less unexplained variance, from which is concluded that null models and metrics influence increases with increasing size of the web. The combination of null model and metric showed the steepest augmentation of explained variance with increasing size. In combination with the low explained variance by the metrics this seems to be caused by one group of null models that react stronger to increasing size than the other. The cause of this is suspected to be the difference between null models that only partly constrain margins and null

models that clearly constrain marginal totals.

Although unexplained variance was still highest for the observed networks, excluding shuffle.web increased the explained variance. Variance explained by null models and the combination of null models with metrics were higher than with shuffle.web. It seems that observed webs naturally create a very high variance between all types of null models, even those that returned similar results for the artificial matrices. This might be because real webs are much less nested than those that were artificially generated, therefore marginal totals are more evenly distributed in observed webs. This allows for more possibilities when creating a null matrix resulting in higher variance between the different null models. For bigger webs, there seems to be a stronger influence of the null models on the metrics, which results in the variance explained by the null model-metric combination.

### 4.3. Conclusions

The analysis of the different metrics has shown, that they do not have a consistent intrinsic definition of nestedness. This result was even more evident after applying the null models. The  $T$  metric was found to be least correlated with all other metrics. A reanalysis of all existing metrics is therefore suggested. They should be tested for their exact definition of nestedness and only metrics with high correlated intrinsic scoring of nestedness should be used. However, high correlation defies the purpose of different metrics because all metrics in essence give the same result.

The null models vary profoundly in their results. Marginal constraint null models produced much more sensitive results and therefore allow for ecological interpretation. Null models that do not constrain margins always indicated extremely significant nestedness in the original matrix and thus neither use nor justification is seen in them. A network in which all parties interact with the same probability seems very unlikely in reality. The main purpose of a null model is to identify the factor that created the nested structure of the original matrix. A null model that solely constrains species richness is likely not to include that factor. This would only be the case if species richness was the cause of nestedness, leading to the conclusion “The more species, the higher the nestedness”. There is no reason to assume that this is the case for real networks, because they are highly complex. It is therefore suggested that these models no longer be used.

The results for the moduled matrices showed that today’s metrics and null models are not suited for such an analysis. It is therefore proposed that a special position in nestedness analysis is allocated to these matrix types. Metrics that are unable to detect any nestedness in a moduled matrix, such as the NODF metrics, should never be used for the analysis of moduled matrices. The checkerboard metrics instead would appear to be better suited, as they should detect high nestedness in the original matrix and decreasing nestedness in the null models. Yet this was not found to be the case and further investigation is recommended before their use. All null models failed to preserve the moduled structure of the matrices. Therefore they are not usable, as the moduled structure is the object of the analysis. A new metric that first detects compartments and then analyses the nestedness of each compartment to calculate overall nestedness of the network is needed. Finding an appropriate null model is challenging as it must constrain the moduled structure but at the same time allow for enough variation.

Real webs react very different to the null models than artificially nested created networks do. They seem to be structured in a way to which the null models react very different. There are null models that produce null matrices to which some metrics react very sensitive, while others react moderate to the same null matrices. The swsh null model seems to create null matrices that are very similar to the original webs, although this was contrary for all artificially generated matrix types. All metric-null model combinations return very similar results. The same 23 ma-

trices were identified as significantly nested, insignificant or significantly segregated, depending on the metric-null model combination to analyse them. This again shows how severe the choice of metric and null model can influence the analysis and how crucial it is to pay close attention when selecting them.

Metrics, null models and networks are all highly different within themselves. Finding a general method to analyse all types of matrices is impossible. A variety of null models is necessary to represent different ecological factors. Special cases of nestedness, such as moduled matrices, call for different metrics to evaluate them. However, different metrics for the same matrix type seem to have no advantages but only promote confusion and it is suggested to agree on only one unified metric, that clearly defines nestedness and calculates it strictly according to that definition. Although there are methods that seem overall better suited for nestedness analysis, it remains the challenge of the ecologist to find the best way for his specific case.



## Acknowledgements

First I want to thank my supervisor Carsten Dormann, who guided me through this thesis and was supporting me regarding the statistical as well as the literature part of my work. I want to thank Michael Staab for offering help and advice. Further more I want to thank my colleges Paul Bauche, who was a great support and shared his experience of the R program with me as well as Antje Leßmann and Severin Hauenstein for their support. Many thanks to the patient readers of my thesis, Martin Legeland , Michael and Florentine Kaiser, who provided me with constructive criticism.



# Bibliography

- Almeida-Neto, M., P. Guimaraes, P. R. Guimaraes, R. D. Loyola, and W. Ulrich (2008). A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos* 117(8), 1227–1239.
- Almeida-Neto, M., P. R. Guimarães Jr, and T. M. Lewinsohn (2007). On nestedness analyses: rethinking matrix temperature and anti-nestedness. *Oikos* 116(4), 716–722.
- Almeida-Neto, M. and W. Ulrich (2011). A straightforward computational approach for measuring nestedness using quantitative matrices. *Environmental Modelling & Software* 26(2), 173–178.
- Atmar, W. and B. Patterson (1995). The nestedness temperature calculator. *A visual basic program, including 294*.
- Atmar, W. and B. D. Patterson (1993). The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia* 96(3), 373–382.
- Bascompte, J., P. Jordano, C. J. Melián, and J. M. Olesen (2003). The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences* 100(16), 9383–9387.
- Bastolla, U., M. A. Fortuna, A. Pascual-Garcia, A. Ferrera, B. Luque, and J. Bascompte (2009). The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature* 458(7241), 1018–1020.
- Beckett, S. J., C. A. Boulton, and H. T. Williams (2014). Falcon: a software package for analysis of nestedness in bipartite networks. *F1000Research* 3.
- Blake, J. G. (1991). Nested subsets and the distribution of birds on isolated woodlots. *Conservation Biology* 5(1), 58–66.
- Bluthgen, N., J. Frund, D. P. Valdovinos, F. S. Siquez, and F. Menzel (2008). What do interaction network metrics tell us about specialization and biological traits. *Ecology* 89(12), 3387–3399.
- Brualdi, R. A. and J. G. Sanderson (1999). Nested species subsets, gaps, and discrepancy. *Oecologia* 119(2), 256–264.
- Cam, E., J. D. Nichols, J. E. Hines, and J. R. Sauer (2000). Inferences about nested subsets structure when not all species are detected. *Oikos* 91(3), 428–434.
- Corso, G., A. I. Araujo, and A. M. Almeida (2008). A new nestedness estimator in community networks. *arXiv preprint arXiv:0803.0007*.
- Cutler, A. (1991). Nested faunas and extinction in fragmented habitats. *Conservation Biology* 5(4), 496–504.
- Cutler, A. H. (1994). Nested biotas and biological conservation: metrics, mechanisms, and meaning of nestedness. *Landscape and Urban Planning* 28(1), 73–82.

- Daubenmire, R. (1975). Floristic plant geography of eastern washington and northern idaho. *Journal of Biogeography*, 1–18.
- Dormann, C.F., G. B. . F.-J. (2008). Introducing the bipartite package: Analysing ecological networks. *R news* Vol 8/2, 8 - 11.
- Fortuna, M. A., D. B. Stouffer, J. M. Olesen, P. Jordano, D. Mouillot, B. R. Krasnov, R. Poulin, and J. Bascompte (2010). Nestedness versus modularity in ecological networks: two sides of the same coin? *Journal of Animal Ecology* 79(4), 811–817.
- Galeano, J., J. M. Pastor, and J. M. Iriondo (2009). Weighted-interaction nestedness estimator (wine): A new estimator to calculate over frequency matrices. *Environmental Modelling & Software* 24(11), 1342–1346.
- Gotelli, N. J. and G. R. Graves (1996). *Null Models in Ecology*. Smithsonian Institution Press, Washington, D.C.
- Gotelli, N. J., G. R. Graves, and C. Rahbek (2010). Macroecological signals of species interactions in the danish avifauna. *Proceedings of the National Academy of Sciences* 107(11), 5030–5035.
- Gotelli, N. J. and W. Ulrich (2012). Statistical challenges in null model analysis. *Oikos* 121(2), 171–180.
- Greve, M. and S. L. Chown (2006). Endemicity biases nestedness metrics: a demonstration, explanation and solution. *Ecography* 29(3), 347–356.
- Guimarães, P. R., P. G. (2006). Improving the analyses of nestedness for large sets of matrices. *environmental modelling and software*. 21: 1512-1513.
- Hausdorf, B. and C. Hennig (2003a). Biotic element analysis in biogeography. *Systematic Biology* 52(5), 717–723.
- Hausdorf, B. and C. Hennig (2003b). Nestedness of north-west european land snail ranges as a consequence of differential immigration from pleistocene glacial refuges. *Oecologia* 135(1), 102–109.
- Hausdorf, B. and C. Hennig (2007). Null model tests of clustering of species, negative co-occurrence patterns and nestedness in meta-communities. *Oikos* 116(5), 818–828.
- James, A., J. W. Pitchford, and M. J. Plank (2012). Disentangling nestedness from models of ecological complexity. *Nature* 487(7406), 227–230.
- Johnson, S., A. V. Domínguez-García, and M. A. Munoz (2013). Factors determining nestedness in complex networks. *PLoS ONE* 8(9): e74025. doi:10.1371/journal.pone.0074025.
- Jonsson, B. G. (2001). A null model for randomization tests of nestedness in species assemblages. *Oecologia* 127(3), 309–313.
- Joppa, L. N., J. M. Montoya, R. Solé, J. Sanderson, S. L. Pimm, et al. (2010). On nestedness in ecological networks. *Evolutionary Ecology Research* 12(1), 35.
- Kondoh, M., S. Kato, and Y. Sakato (2010). Food webs are built up with nested subwebs. *Ecology* 91(11), 3123–3130.
- Krishna, A., P. R. Guimaraes Jr, P. Jordano, and J. Bascompte (2008). A neutral-niche theory of nestedness in mutualistic networks. *Oikos* 117(11), 1609–1618.

- Lewinsohn, T. M., P. Inacio Prado, P. Jordano, J. Bascompte, and J. M. Olesen (2006). Structure in plant–animal interaction assemblages. *Oikos* 113(1), 174–184.
- Lomolino, M. V. (1996). Investigating causality of nestedness of insular communities: selective immigrations or extinctions? *Journal of biogeography* 23(5), 699–703.
- MacArthur, R. H. and E. R. Pianka (1966). On optimal use of a patchy environment. *American Naturalist*, 603–609.
- McCoy, E. D. and K. L. Heck Jr (1987). Some observations on the use of taxonomic similarity in large-scale biogeography. *Journal of Biogeography*, 79–87.
- Moore, J. E. and R. K. Swihart (2007). Toward ecologically explicit null models of nestedness. *Oecologia* 152(4), 763–777.
- Nielsen, A. and J. Bascompte (2007). Ecological networks, nestedness and sampling effort. *Journal of Ecology* 95(5), 1134–1141.
- Novak, M., J. W. Moore, and R. A. Leidy (2011). Nestedness patterns and the dual nature of community reassembly in california streams: a multivariate permutation-based approach. *Global change biology* 17(12), 3714–3723.
- Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O’Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner (2014). *vegan: Community Ecology Package*. R package version 2.2-0.
- Patterson, B. and W. Atmar (2000). Analyzing species composition in fragments. *Isolated vertebrate communities in the tropics* 46, 9–24.
- Patterson, B. D. and W. Atmar (1986). Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological Journal of the Linnean Society* 28(1-2), 65–82.
- Podani, J. (2001, 3). Multilevel dendrogram comparison (dencompar) for windows-the simplex method (sdrsimplex). <http://ramet.elte.hu/podani/subindex.html>.
- Podani, J. and D. Schmera (2011). A new conceptual and methodological framework for exploring and explaining pattern in presence - absence data. *Oikos* 120(11), 1625 – 1638.
- Podani, J. and D. Schmera (2012). A comparative evaluation of pairwise nestedness measures. *Ecography* 35(10), 889–900.
- Rezende, E. L., J. E. Lavabre, P. R. Guimarães, P. Jordano, and J. Bascompte (2007). Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature* 448(7156), 925–928.
- Rodríguez-Gironés, M. A. & Santamaría, L. (2006). A new algorithm to calculate the nestedness temperature of presence-absence matrices. *Journal of Biogeography*, 33: 924-935.
- Simberloff, D. and J.-L. Martin (1991). Nestedness of insular avifaunas: simple summary statistics masking complex species patterns. *Ornis Fennica* 68(4), 178–192.
- Stang, M., P. G. Klinkhamer, and E. Van Der Meijden (2006). Size constraints and flower abundance determine the number of interactions in a plant–flower visitor web. *Oikos* 112(1), 111–121.
- Staniczenko, P. P., J. C. Kopp, and S. Allesina (2013). The ghost of nestedness in ecological networks. *Nature communications* 4, 1391.

- Stone, L. and A. Roberts (1990). The checkerboard score and species distributions. *Oecologia* 85(1), 74–79.
- Strona, G., P. Galli, D. Seveso, S. Montano, and S. Fattorini (2014). Nestedness for dummies (ned): a user friendly web interface for exploratory nestedness analysis. *Journal of Statistical Software*. Software available at: <http://ecosoft.alwaysdata.net>.
- Strona, G., D. Nappo, F. Boccacci, S. Fattorini, and J. San-Miguel-Ayanz (2014). A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. *Nature communications* 5.
- Thompson, J. N. (2005). *The geographic mosaic of coevolution*. University of Chicago Press.
- Ulrich, W. (2006a). Matrix—a fortran program for generating randomized ecological matrices. [www.keib.umk.pl/wp-content/uploads/2013/04/MatrixManual.pdf](http://www.keib.umk.pl/wp-content/uploads/2013/04/MatrixManual.pdf).
- Ulrich, W. (2006b). Nestedness—a fortran program for measuring order and disorder in ecological communities. [www.keib.umk.pl/wp-content/uploads/2013/04/NestednessManual.pdf](http://www.keib.umk.pl/wp-content/uploads/2013/04/NestednessManual.pdf).
- Ulrich, W. (2008). Pairs – a fortran program for studying pair wise species associations in ecological matrices. version 1.0. <http://www.keib.umk.pl/pairs/?lang=en>.
- Ulrich, W. (2009). Nestedness analysis as a tool to identify ecological gradients. *Ecological Questions* 11, 27–34.
- Ulrich, W. (2010). Nodf—a fortran program for nestedness analysis. [www.ulrichnet.com/nodf](http://www.ulrichnet.com/nodf).
- Ulrich, W. and M. Almeida-Neto (2012). On the meanings of nestedness: back to the basics. *Ecography* 35(10), 865–871.
- Ulrich, W., M. A.-N. Almeida-Neto, and N. J. Gotelli (2009). A consumer’s guide to nestedness analysis. *Oikos* 118, 3–17.
- Ulrich, W. and N. J. Gotelli (2007). Null model analysis of species nestedness patterns. *Ecology* 88(7), 1824–1831.
- Ulrich, W. and N. J. Gotelli (2010). Null model analysis of species associations using abundance data. *Ecology* 91(11), 3384–3397.
- Ulrich, W. and N. J. Gotelli (2012). A null model algorithm for presence–absence matrices based on proportional resampling. *Ecological Modelling* 244, 20–27.
- Ulrich, W. and N. J. Gotelli (2007). Disentangling community patterns of nestedness and species co-occurrence. *Oikos* 116(12), 2053–2061.
- Wright, D. H., B. D. Patterson, G. M. Mikkelsen, A. Cutler, and W. Atmar (1997). A comparative analysis of nested subset patterns of species composition. *Oecologia* 113(1), 1–20.
- Wright, D. H. and J. H. Reeves (1992). On the meaning and measurement of nestedness of species assemblages. *Oecologia* 92(3), 416–428.

## **A. Altered bipartite code**

```

nested <- function (web, method = "binmatnest2", rescale = FALSE, normalised = FALSE)
{
  if (!any(method %in% c("binmatnest", "discrepancy", "binmatnest2",
    "discrepancy2", "NODF", "NODF2", "weighted NODF", "wine",
    "C.score", "checker", "ALL")))
    stop("Typo? Unknown method!")
  if ("ALL" %in% method)
    index <- c("binmatnest", "discrepancy", "binmatnest2",
    "discrepancy2", "NODF", "NODF2", "weighted NODF",
    "wine", "C.score", "checker")
  else index <- method
  out <- NULL
  if ("binmatnest2" %in% index){
    nesy <- try(nestedtemp(web)$statistic, silent=TRUE)
    nesy.value <- if (inherits(nesy, "try-error")) NA else nesy
    out <- c(out, binmatnest2 = nesy.value)
  }
  if ("binmatnest" %in% index){ # NA occur if web is full (i.e. no 0s)
    nesy <- try(nestedness(web, null.models = FALSE)$temperature, silent=TRUE)
    nesy.value <- if (inherits(nesy, "try-error")) NA else nesy
    out <- c(out, binmatnest = nesy.value)
  }
  if ("discrepancy2" %in% index) {
    out <- c(out, discrepancy2 = nesteddisc(web)$statistic)
  }
  if ("discrepancy" %in% index)
    out <- c(out, discrepancy = unname(discrepancy(web)))
  if ("C.score" %in% index)
    out <- c(out, C.score = C.score(web, normalised = normalised))
  if ("checker" %in% index)
    out <- c(out, checker = nestedchecker(web)$C.score)
  if ("NODF2" %in% index)
    out <- c(out, NODF2 = unname(nestednodf(web, order = TRUE)$statistic[3]))
  if ("NODF" %in% index)
    out <- c(out, NODF = unname(nestednodf(web, order = FALSE)$statistic[3]))
  if ("weighted NODF" %in% index)
    out <- c(out, `weighted NODF` = unname(nestednodf(web,
    order = FALSE, weighted = TRUE)$statistic[3]))
  if ("wine" %in% index)
    out <- c(out, wine = wine(web)$wine)
  if (rescale & !"ALL" %in% method)
    warning("You requested rescaling, but you won't get it (unless you use method='ALL')!")
  if (rescale & "ALL" %in% method)
    out <- abs(c(100, 100, 0, 0, 0, 0, 0, 0, 0, 0) - out)
  out
}

```

```

nullmodel <- function (web, N = 1000, method = "r2d", ...)
{
  methods <- c("r2dtable", "swap.web", "vaznull", "shuffle.web",
    "mgen", "quasiswap", "swsh", "abuswap")
  if (is.numeric(method)) {
    m <- method
  } else {
    m <- pmatch(method, methods)
  }
  if (is.na(m))
    stop("Abbreviated name does not uniquely identify method.")
  if (m == 1) {
    if (all(web < 2))
      warning("This seems to be a binary web. Only methods shuffle.web and mgen should be used!\n I
      proceeded nonetheless. Read the note in the help file!")
  }
}

```

```

rs <- rowSums(web)
cs <- colSums(web)
out <- r2dtable(N, r = rs, c = cs)
}
if (m == 2) {
if (all(web < 2))
warning("This seems to be a binary web. Only methods shuffle.web and mgen should be used!\n I
proceeded nonetheless. Read the note in the help file!")
out <- swap.web(N, web, ...)
}
if (m == 3) {
if (all(web < 2))
warning("This seems to be a binary web. Only methods shuffle.web and mgen should be used!\n I
proceeded nonetheless. Read the note in the help file!")
out <- vaznull(N, web)
}
if (m == 4) {
if (any(web > 1))
out <- shuffle.web(web, N, ...)
if (all(web < 2))
out <- replicate(n = N, expr = unname(commsimulator(web,
method = "quasiswap", ...)), simplify = FALSE)
}
if (m == 5) {
out <- replicate(n = N, mgen(web, autotransform="equiprobable"),
simplify = FALSE)
}

if (m == 6){
null <- vegan::nullmodel(web, method = "quasiswap_count")
res <- simulate(null, nsim=N)
out <- apply(res, 3)
}

if (m == 7){
out <- permatswap(web, method = "swsh", times=N, fixedmar="rows")[[3]]
}

if (m == 8){
out <- replicate(n = N, expr = unname(permatswap(web, method = "abuswap", times=1,
fixedmar="rows", seed=runif(1))[[3]]), simplify = FALSE)
}

if (!(m %in% 1:8)) stop("Please choose a valid method.")
return(out)
}

```

```

mgen <- function(web, n=sum(web), keep.species=TRUE, rep.cell=TRUE, autotransform="sum",
trials=100){
# function to generate a quantitative network based on a probability matrix
# by Diego Vázquez (brushed up for a more bipartite-consistent use of names by CFD)
# web a matrix with observation probabilities, emerging from some null model considerations
external to this function; if an original network is used, this will be automatically converted to a
probability matrix by dividing it by the number of interactions (CFD); ORIGINAL: a probability
matrix
# n number of interactions to allocate into the new matrix
# autotransform: determines how a non-probability web is converted into probabilities;
# option 1: "sum": simply divide each entry by the sum of interactions in the web
# option 2: "equiprobable": product of marginal probabilities (col/rowSums divided by
sum(web) (in a sense this is the basis of the r2dtable null model, just without the 'turn into integers'
bit)

```

*# keep.species: Random assignment of interactions may lead to empty columns or rows and hence reduce*

*#the dimensions of the simulated web. By default, this is prevented from happening, i.e. each row/column will receive at least one interaction. Setting keep.species to FALSE may (but need not) cause a loss of species.# trials: allocating interactions (when rep.cell=TRUE) can be difficult or even impossible. When the number of attempts to allocate them exceeds trials \* n it will be stopped and an error message will be returned. Default is 100. Setting 'trials' to a very large value may make this function run for hours and hours. Your choice!*

```
if (sum(web) != 1) { # added by CFD
  message(paste("This is not a probability matrix! I will proceed after transforming the entries
  according to option 'autotransform':", autotransform, "!"))
  if (autotransform == "sum") {
    m <- web/sum(web)
  } else {# equiprobable, or anything else
    m <- (rowSums(web)/sum(web)) %*% t(colSums(web)/sum(web))
  }
} else m <- web

if (rep.cell == FALSE & n > (nrow(m)*ncol(m))){
  message("Argument n should be smaller than the number of cells in matrix!")
}
else{
  mac <- matrix(cumsum(m),nrow(m),ncol(m)) #Cumulative probability matrix
  mint <- matrix(0,nrow(m),ncol(m)) #Interaction matrix
  if (keep.species){
    for (i in 1:nrow(m)){
      c1 <- sample(ncol(m), replace=TRUE, prob=colSums(m))
      c1 <- c1[1]
      mint[i, c1] <- 1
    }
    for (i in 1:ncol(m)){
      if(sum(mint[,i]) == 0){
        r1 <- sample(nrow(m), replace=TRUE, prob=rowSums(m))
        r1 <- r1[1]
        mint[r1, i] <- 1
      }
    }
  }
  while.counter <- 0
  while (sum(mint) < n){
    rand <- runif(1, 0, 1)
    ri <- min(which(mac >= rand))
    if (rep.cell == TRUE) mint[ri] <- mint[ri] + 1
    if (rep.cell == FALSE) mint[ri] <- 1
    while.counter <- while.counter + 1
    if (while.counter >= trials*n) stop("Cannot allocate the requested interactions in a reasonable
    amount of time! \n Either increase 'trials' or decrease 'n'.")
  }
  mint
}
}
```

# Selbstständigkeitserklärung

## Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit ohne fremde Hilfe selbstständig verfasst und nur die angegebenen Quellen und Hilfsmittel benutzt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen habe ich unter Angabe der Quellen kenntlich gemacht.

(I hereby declare that I have composed this document unassistedly and that I only used the sources and devices I declared. Passages taken verbatim or in meaning from other sources are identified as such and the sources are acknowledged and cited.)

Freiburg, Jan 2015