Null Models

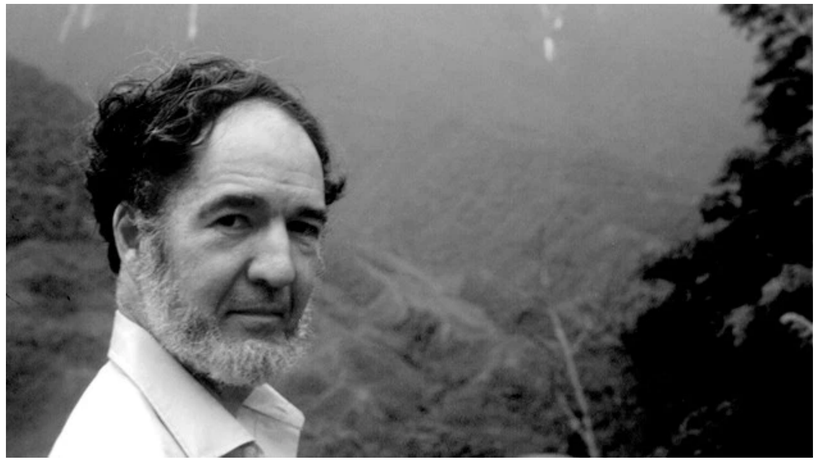
Fernando Pedraza
fernando.pedraza@ieu.uzh.ch

outline for this morning

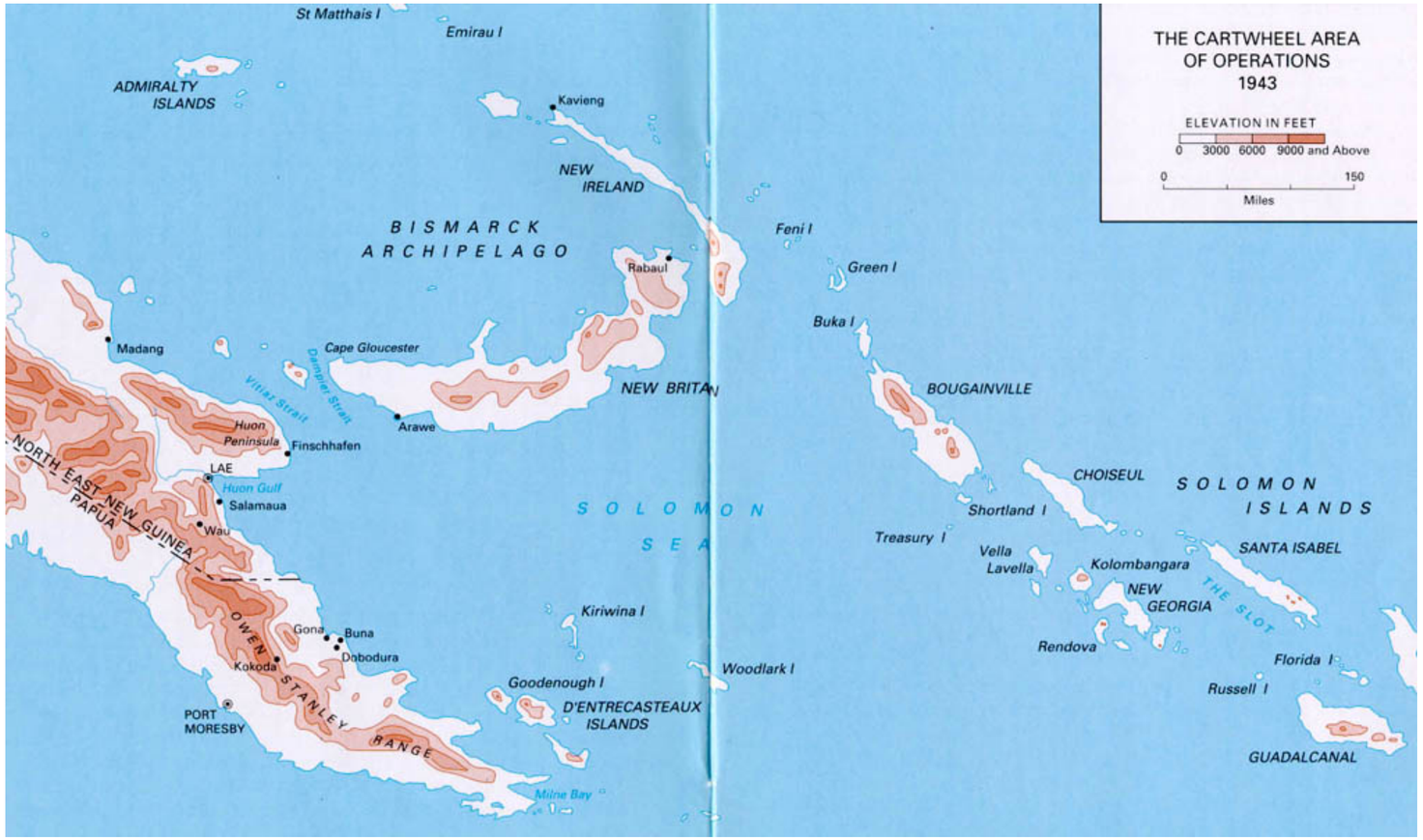
i. *Null models in ecological research*
ii. *Null models in network research*
iii. *Examples of null models*
iv. *Use cases of null models*
v. *Caveats of null models*

null models in ecological research

foreword: a tale of birds and islands



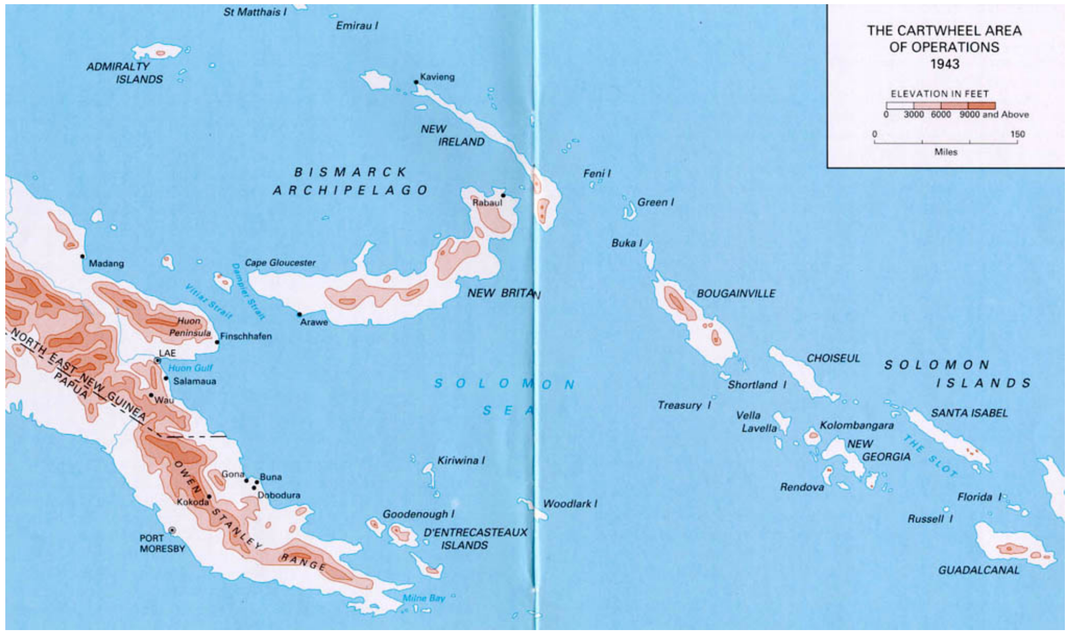
Jared Diamond



Bismark Archipelago, New Guinea

foreword: a tale of birds and islands

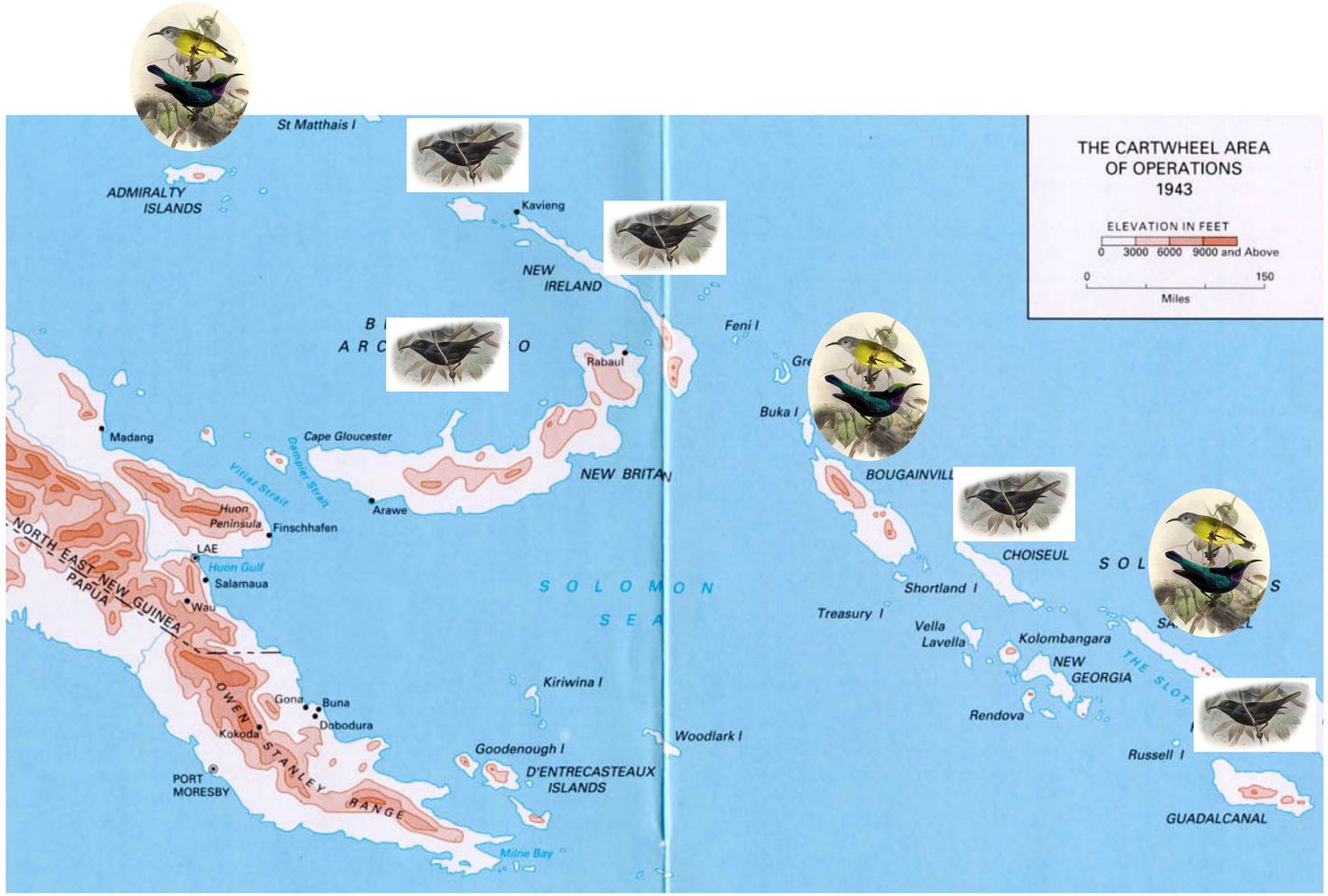
black sunbird
(found on 14 islands)



Bismark Archipelago, New Guinea

Bismarck black
myzomela
(found on 23 islands)

what explains the distribution of bird species?



Bismark Archipelago, New Guinea

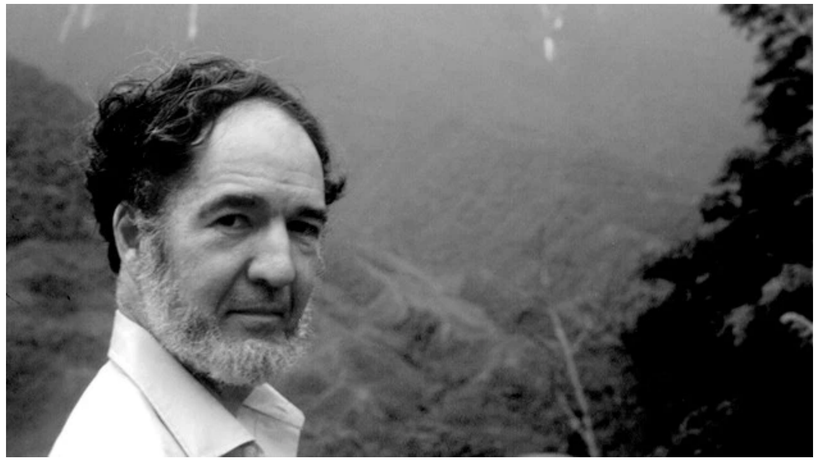
what explains the distribution of bird species?

black sunbird
(found on 14 islands)

competition leads
to exclusion

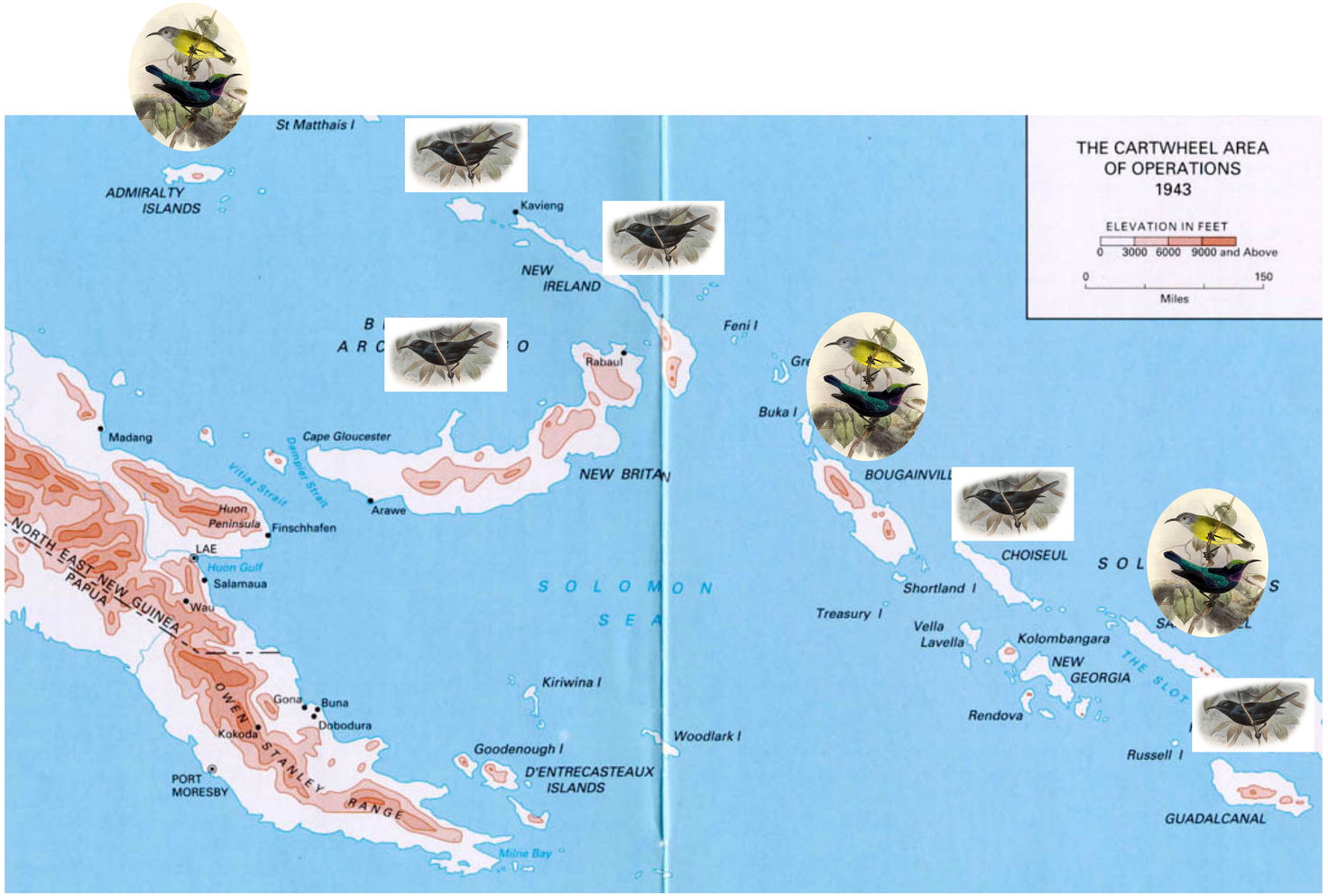
Bismarck black myzomela
(found on 23 islands)

assembly rules



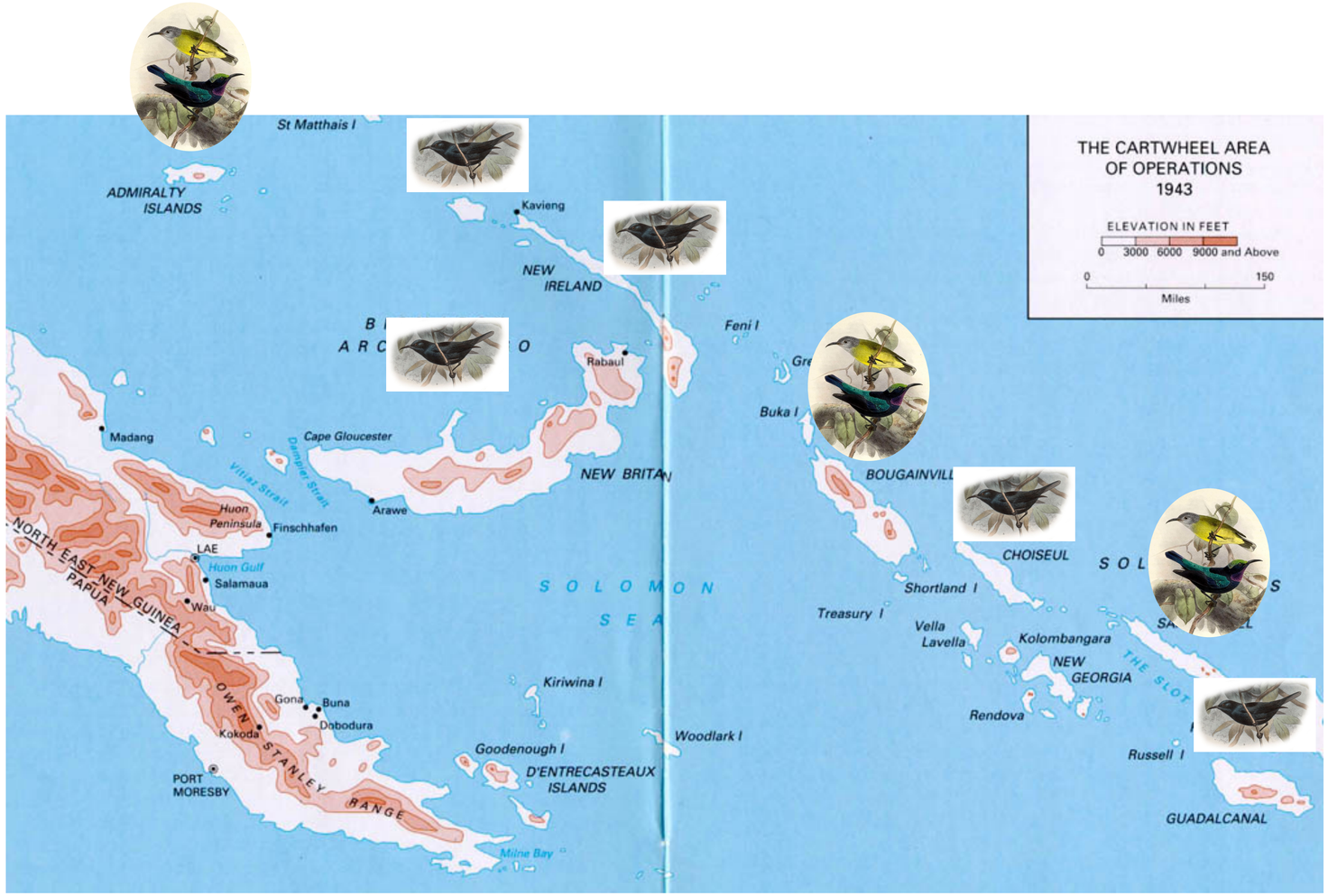
Jared Diamond

"competition is responsible for determining the patterns of assemblage composition."



Bismark Archipelago, New Guinea

what explains the distribution of bird species?

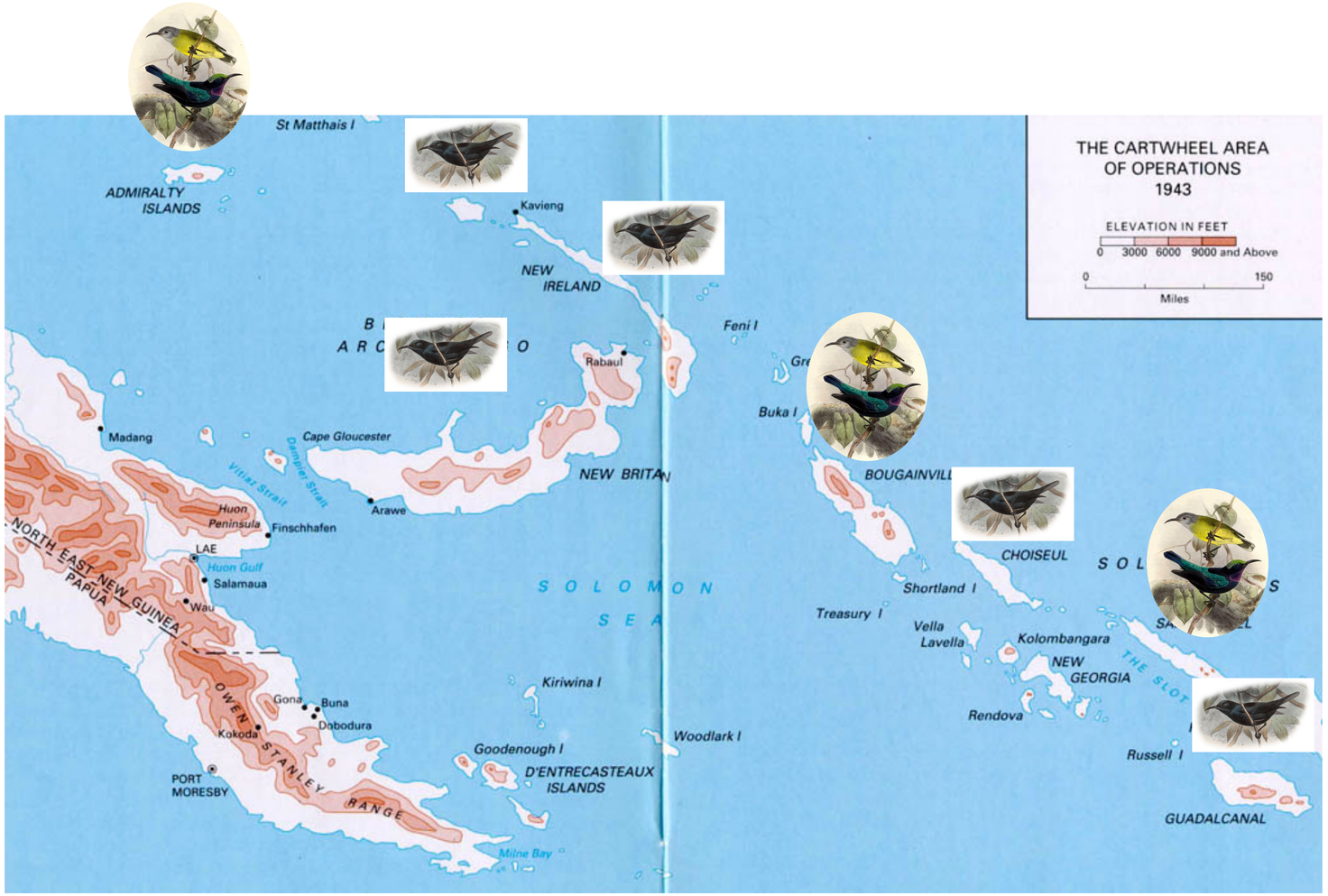


Bismark Archipelago, New Guinea

what explains the distribution of bird species?

Daniel Simberloff

“In order to to demonstrate that competition is responsible for the joint distributions of species, one would have to falsify a null hypothesis stating that the distributions are generated by the species randomly and individually colonizing an archipelago”

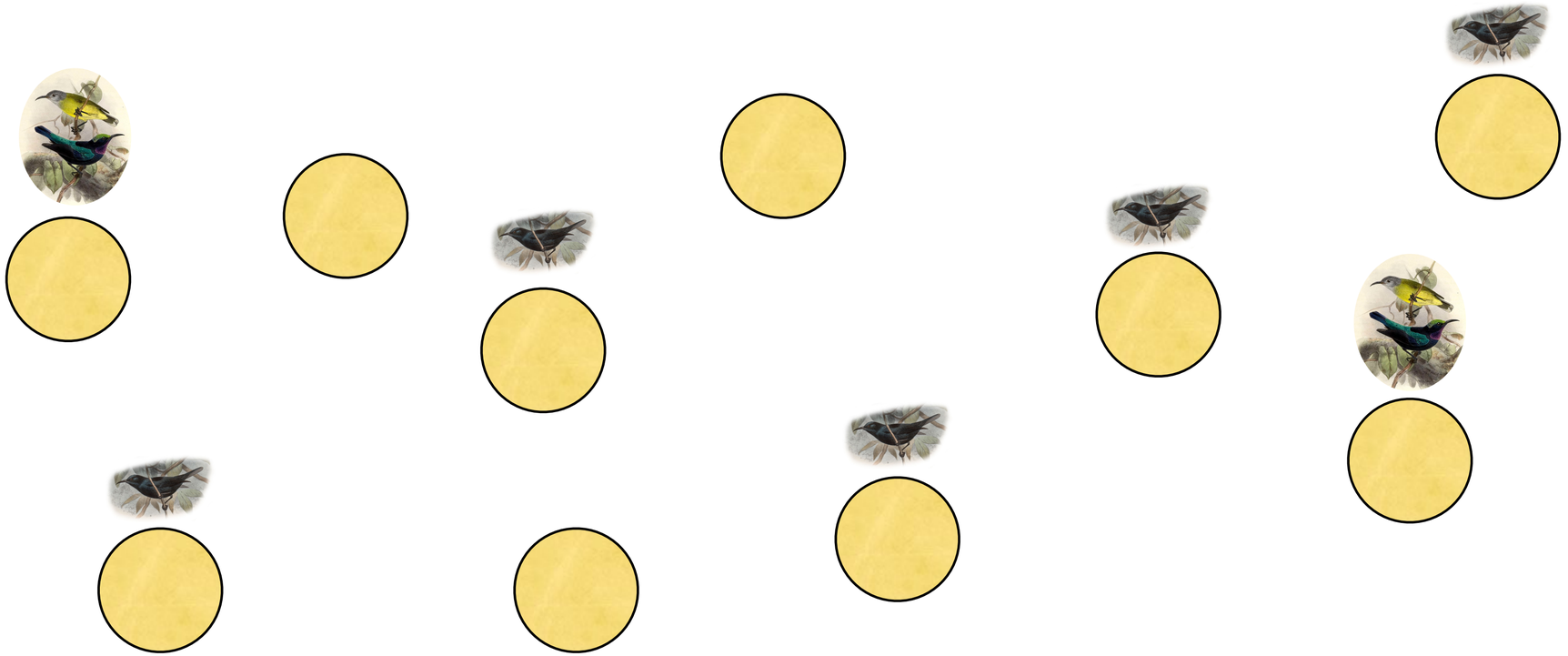


Bismark Archipelago, New Guinea

the null model

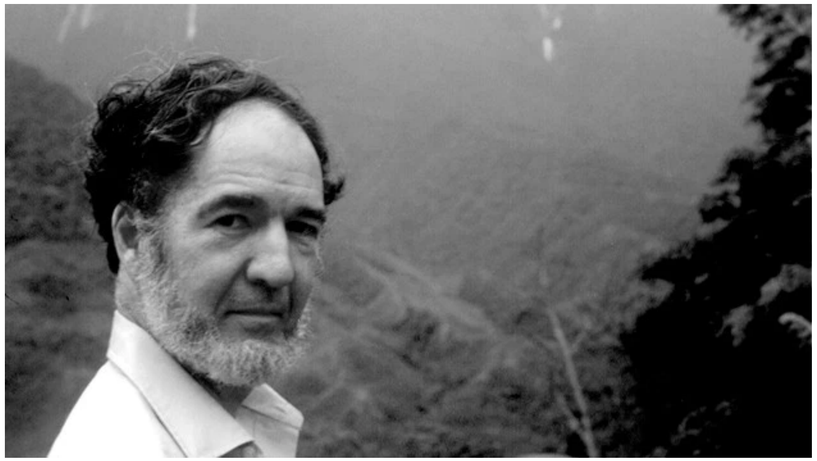
- A null model is a pattern-generating model that is based on randomisation of ecological data.
- The goal of a null model strategy is to construct a model that deliberately excludes a mechanism being tested.
- Can the patterns in the real data be reproduced in a simple model that does not incorporate biologically important mechanisms?

the null expectation



N = 10 islands, Sp1 = 5 islands, Sp2 = 2 islands
Probability of findings Sp1 = 0.5, Probability of finding Sp2 = 0.2
Probability of overlap (if independent) = 0.5 x 0.2 = 0.1

inferences from observations?



Jared Diamond

Daniel Simberloff

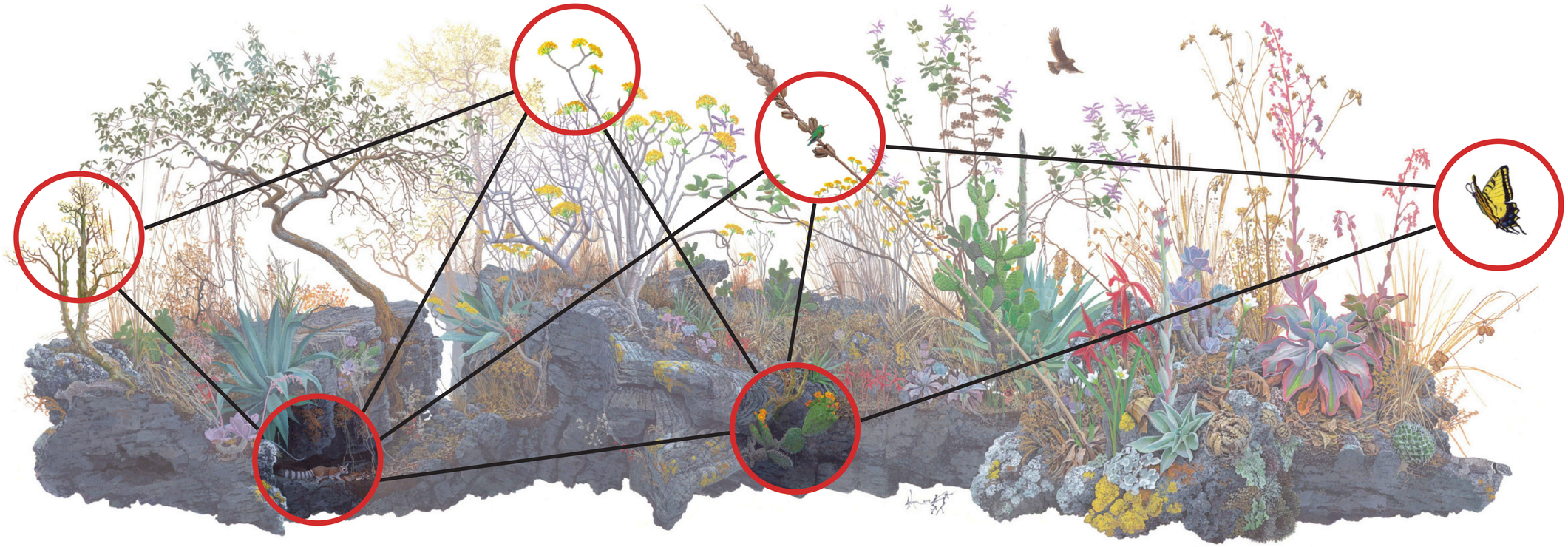
Does interspecific competition make species co-occurrences significantly non-random, and can competition be inferred from observation of spatial distribution patterns?

null models in network research

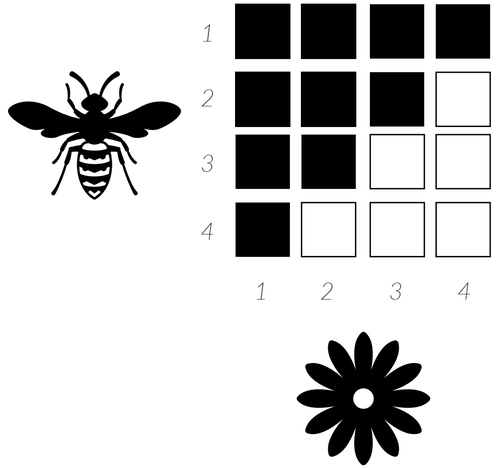
communities and the interactions between species

Artwork by Aslam Narváez Parra

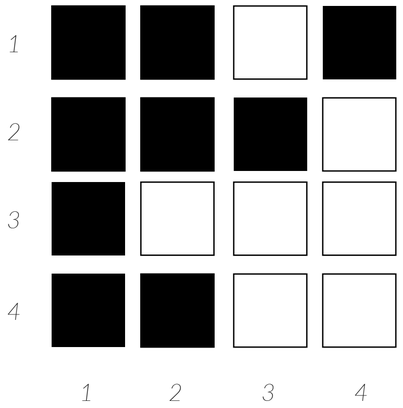
communities and the interactions between species

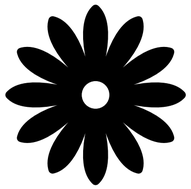


communities and the interactions between species



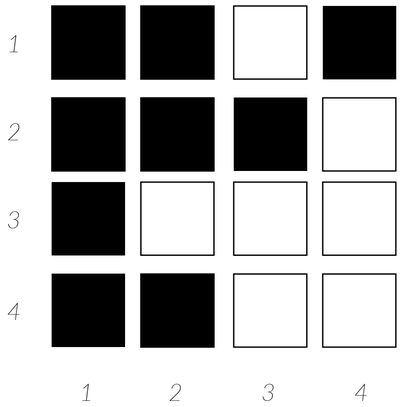
measuring the structural properties of networks

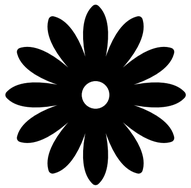




Nestedness = 0.8

assessing the significance of the structural pattern





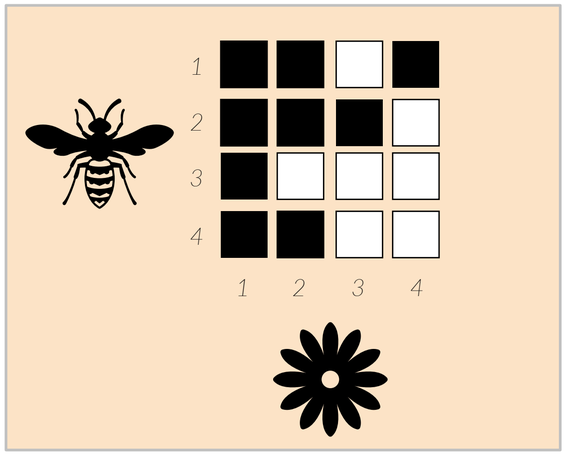
Nestedness = 0.8 ***… so what?***

null models

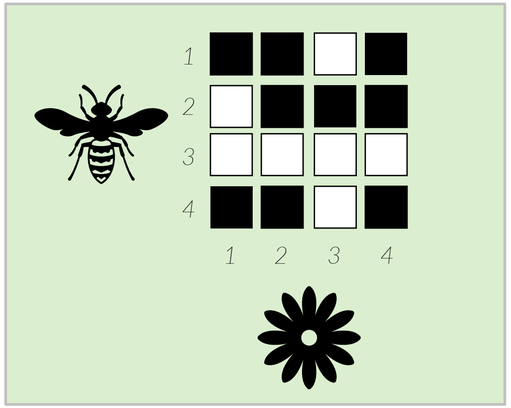
a null model is an appropriate randomisation of the observed matrix intended to serve as a benchmark to assess the significance of a given pattern. It is a pattern-generating model that deliberately excludes a mechanism of interest, and so it serves to test whether the observed level of structure can be explained out of chance.

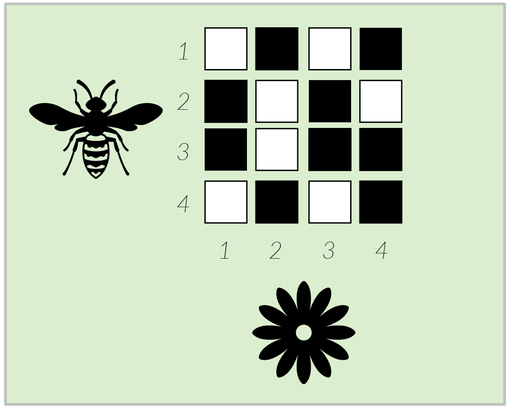
assessing the significance of the structural pattern

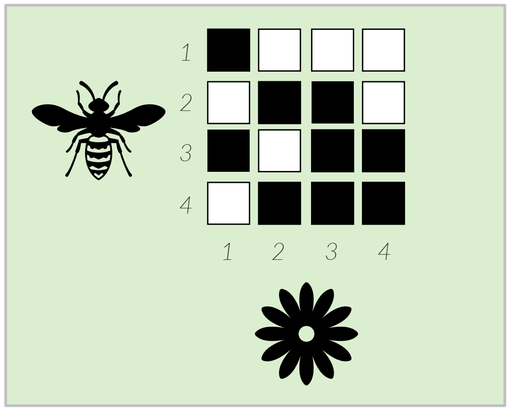
observed structure

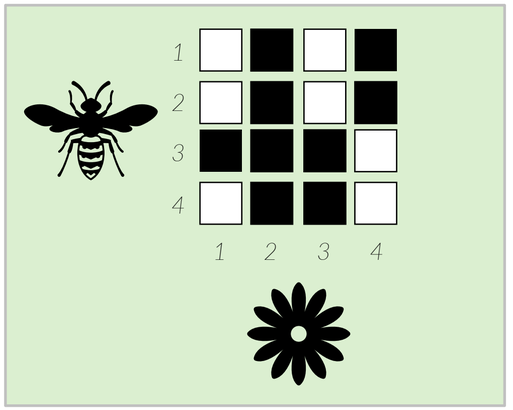


null expectations of structure









examples of null models

examples of null models

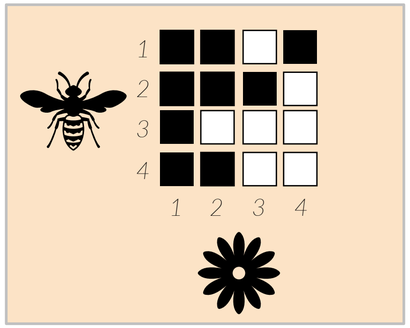
i. *equifrequent null model*

ii. *equifrequent column null model*

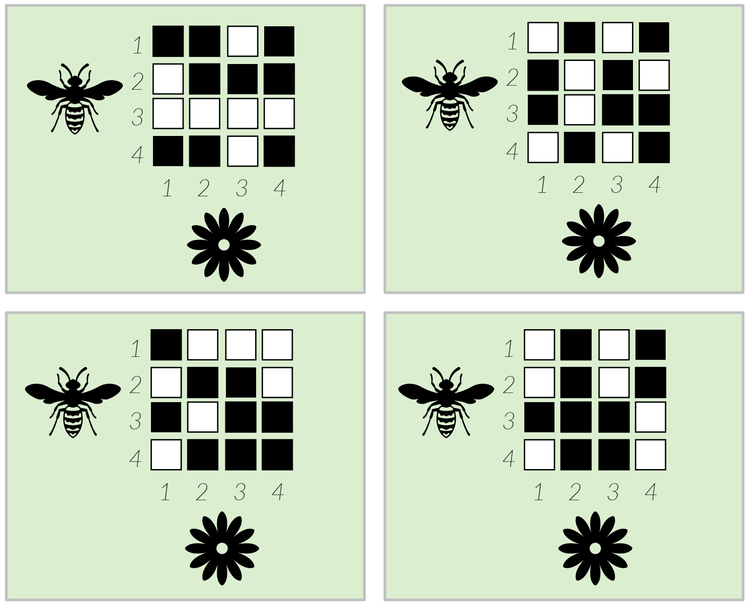
iii. *probabilistic cell null model*

iv. *swap null model*

observed structure



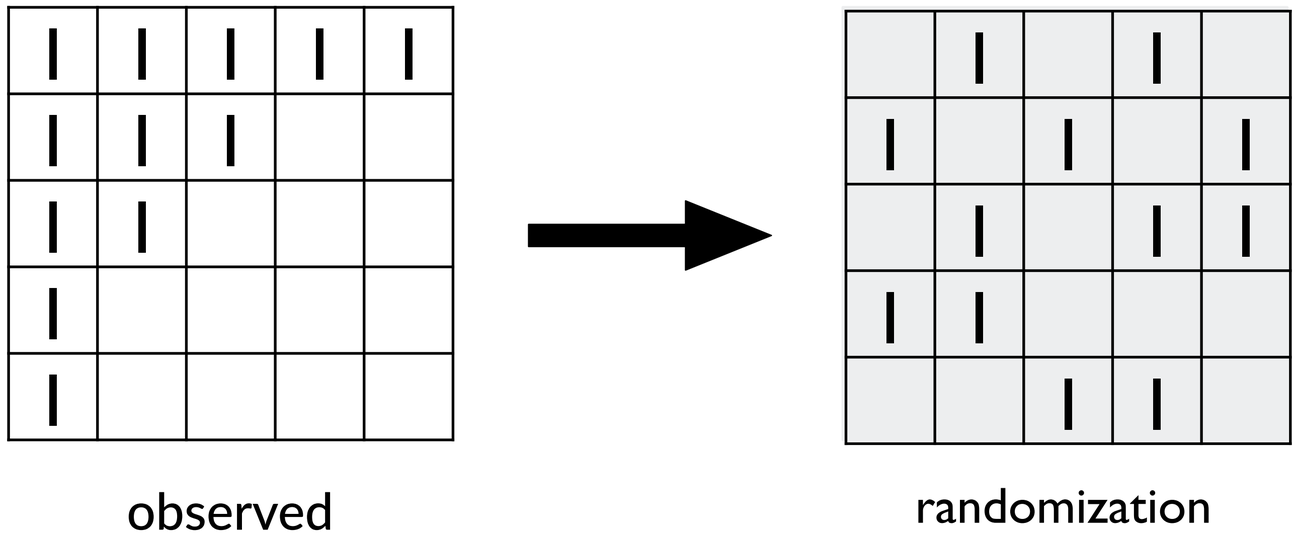
null expectations of structure



equifrequent null model

each cell ij has the same probability of having a "1" given by the total fraction of "1"s in the observed matrix

$$\rho_{ij} = \frac{\text{total number of "1"s}}{\text{potential number of "1"s}} = \frac{12}{25} = 0.48$$



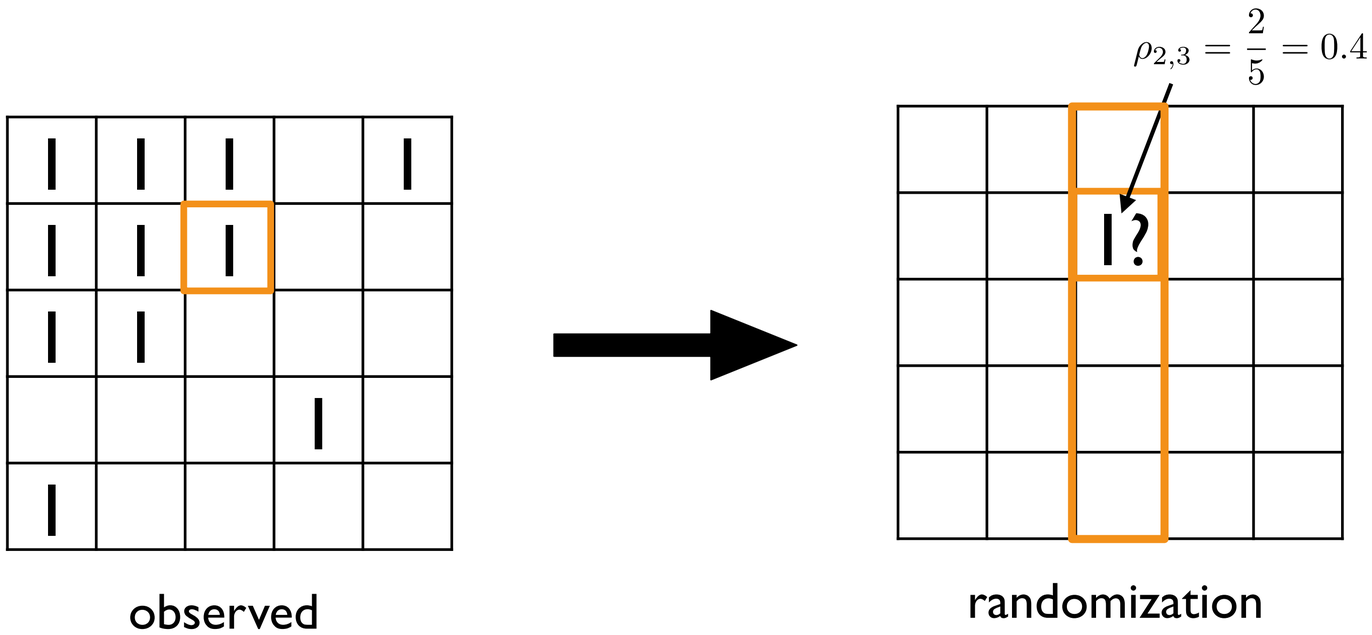
observed

randomization

equifrequent column null model

the probability $\boldsymbol{p_{ij}}$ *of drawing an interaction in cell* $\boldsymbol{ij}$ *is the fraction of "1"s in column* $\boldsymbol{j}$

$$\rho_{ij} = \text{fraction of "1"s in column j} = \frac{1}{n}\sum_{i=1}^{n} M_{ij}$$



probabilistic cell null model

the probability p_{ij} *of drawing an interaction in cell* ij *is proportional to the degree of both row* i *and column* j

$$\rho_{ij} = \frac{p_i + q_j}{2}$$

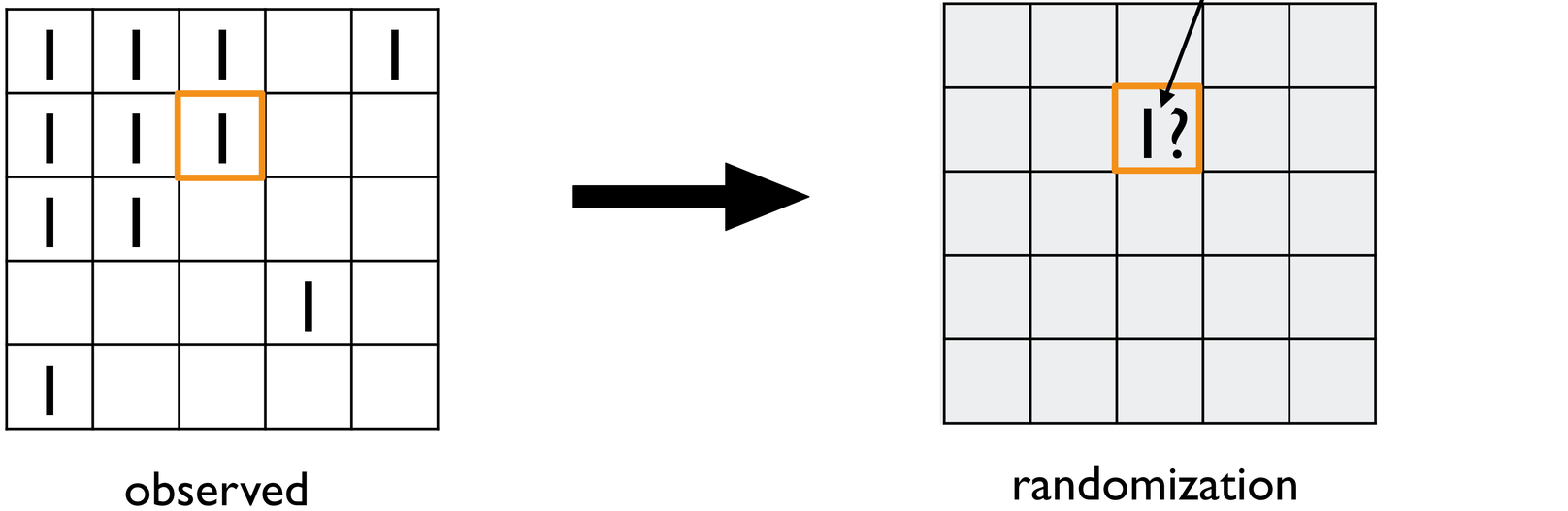
fraction of "1"s in row i

$$p_i = \frac{1}{n}\sum_{i=1}^{n} M_{ij}$$

fraction of "1"s in column

$$q_j = \frac{1}{m}\sum_{j=1}^{m} M_{ij}$$

$$\rho_{2,3} = \frac{3/5 + 2/5}{2} = 0.5$$



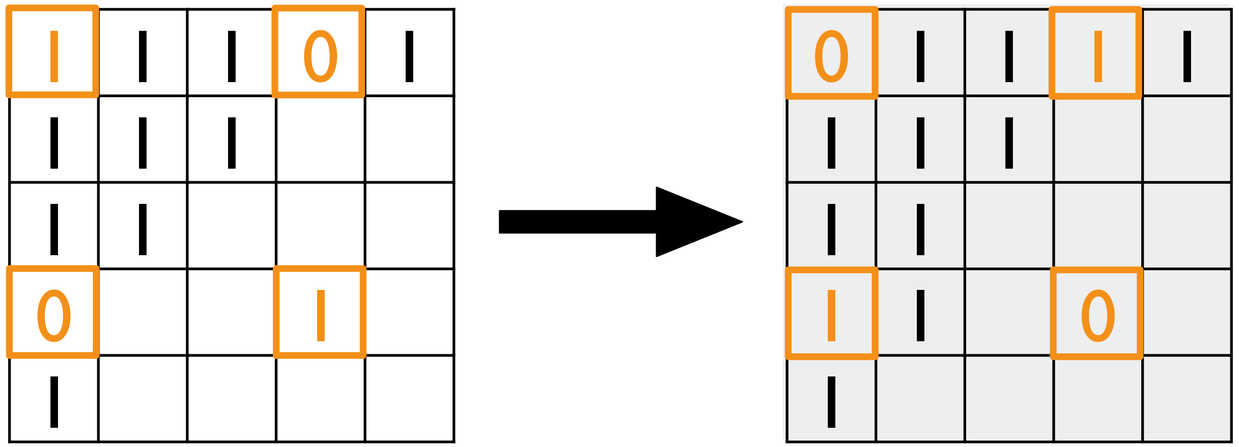
observed

randomization

swap null model

it maintains exactly the number of "1"s in both rows and columns by sequentially reshuffling 2x2 sub matrices with the same row and column total as follows:

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



examples of null models

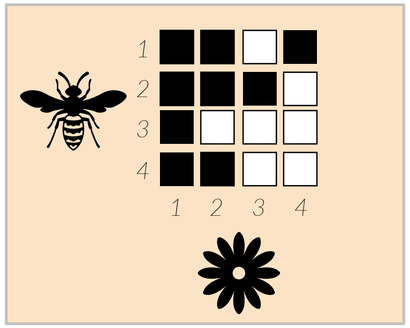
i. *equifrequent null model*

ii. *equifrequent column null model*

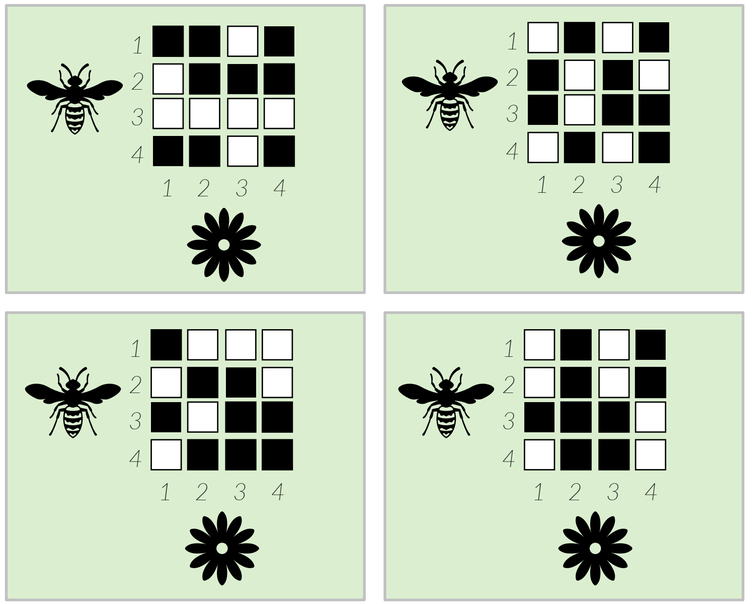
iii. *probabilistic cell null model*

iv. *swap null model*

observed structure



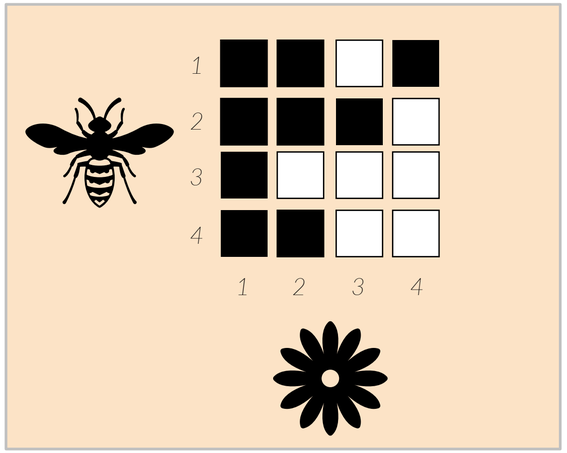
null expectations of structure



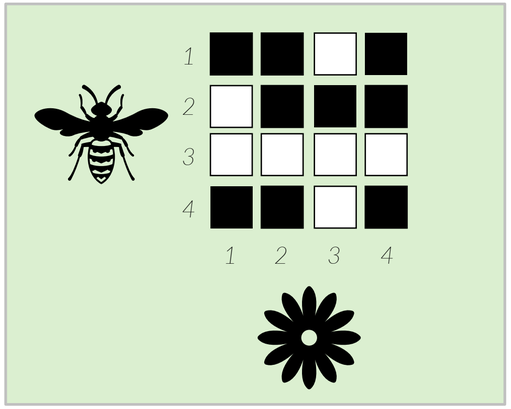
use cases of null models

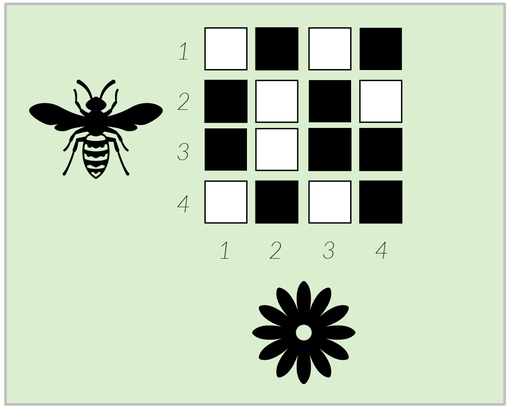
use case I: significance of a network pattern

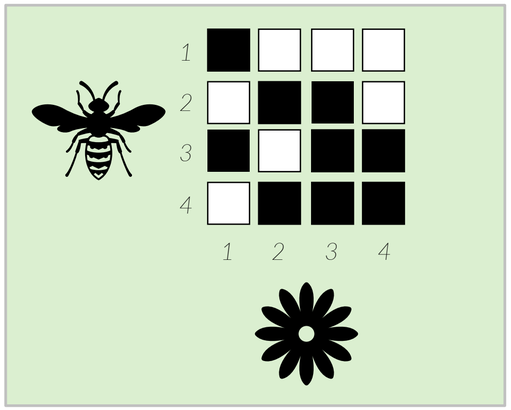
observed structure

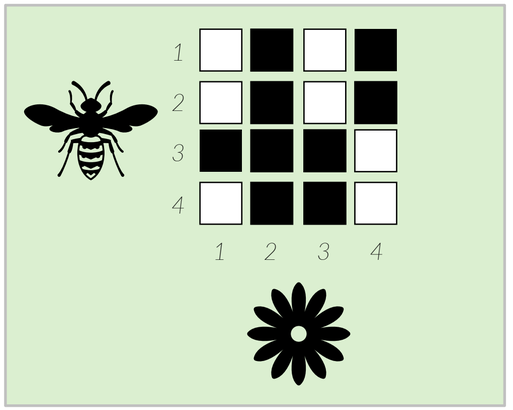


null expectations of structure



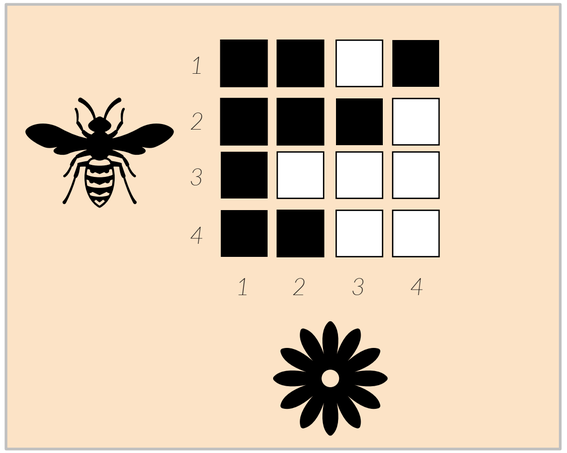




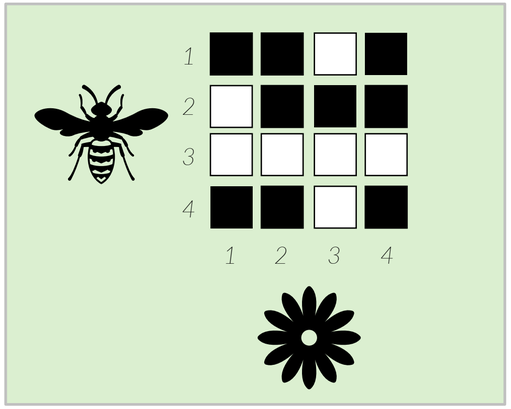


use case I: significance of a network pattern

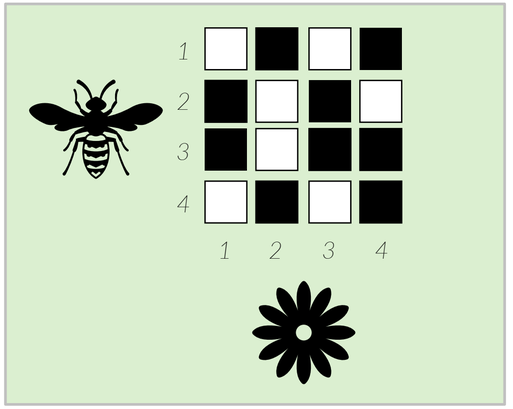
nestedness = 0.8



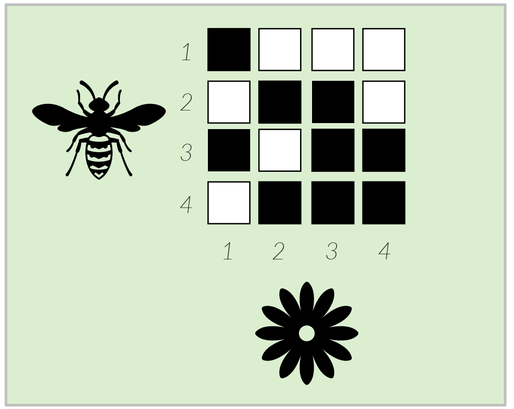
nestedness = 0.5



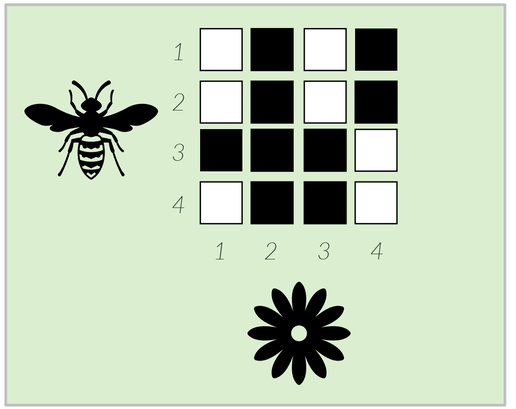
nestedness = 0.3



nestedness = 0.6



nestedness = 0.4



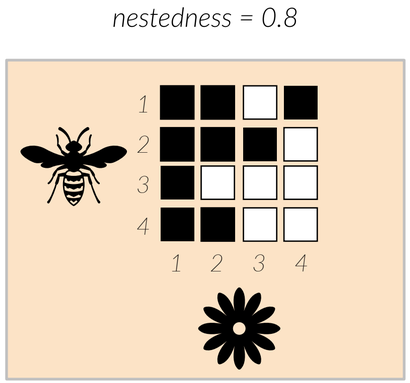
use case I: significance of a network pattern

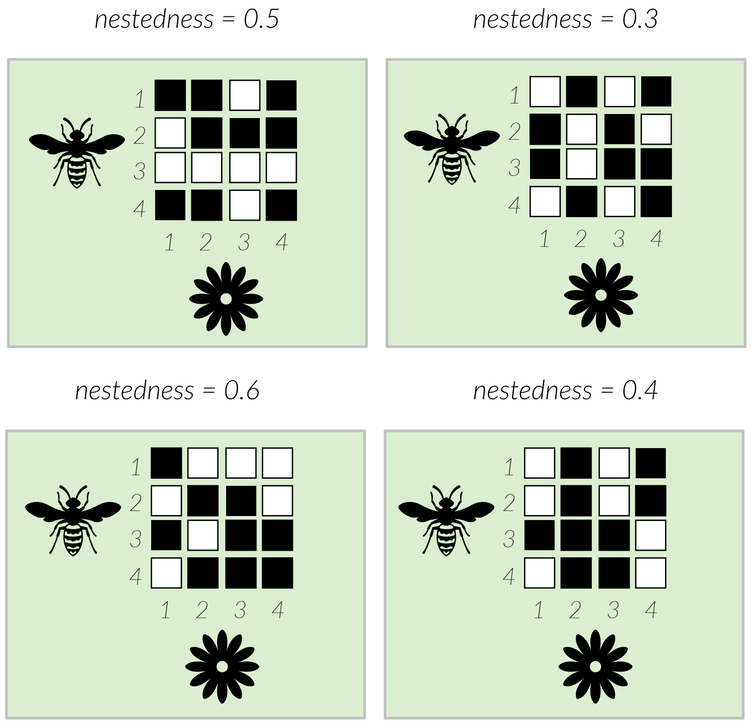
observed value

average across replicates

$$z = \frac{x_i - \overline{x}}{\sigma}$$

standard deviation across replicates





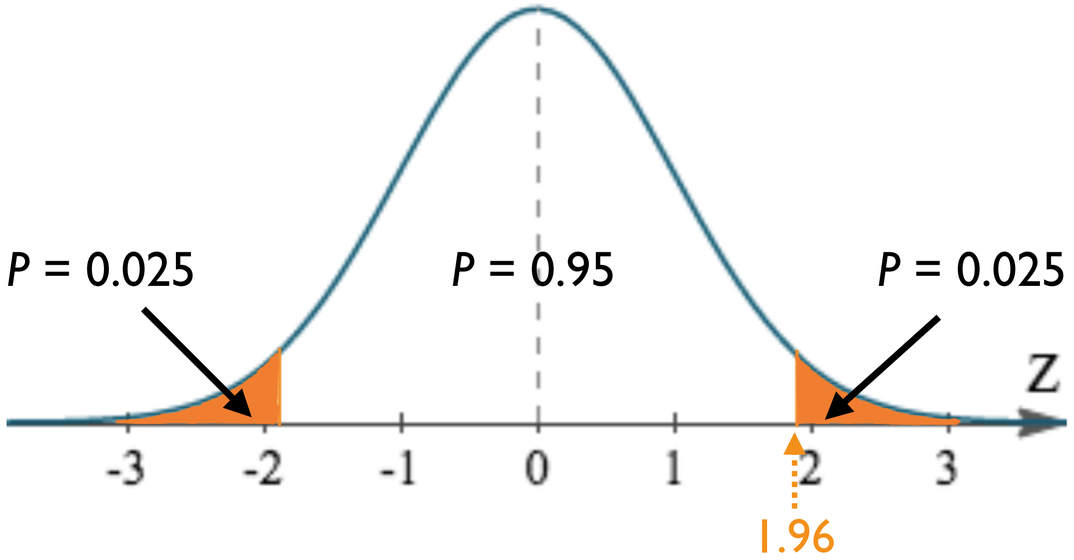
use case I: significance of a network pattern

observed value

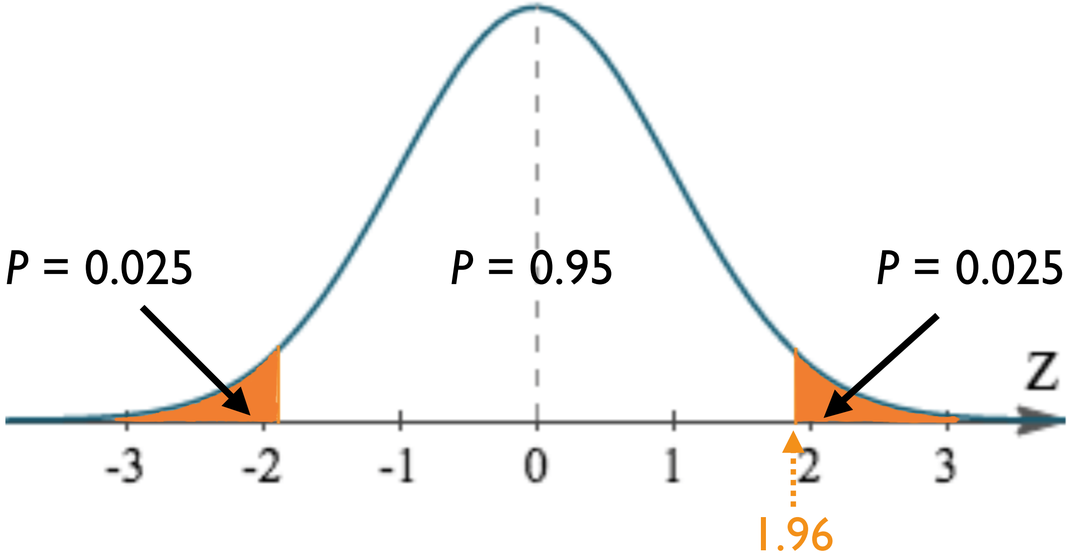
average across replicates

$$z = \frac{x_i - \overline{x}}{\sigma}$$

standard deviation across replicates



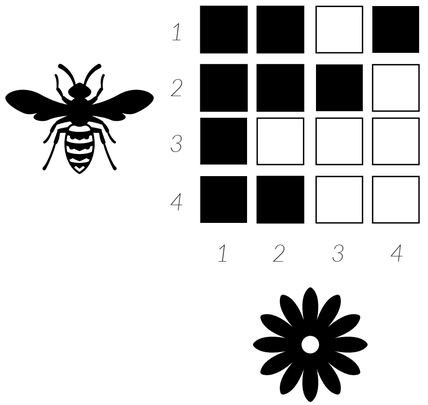
use case I: significance of a network pattern



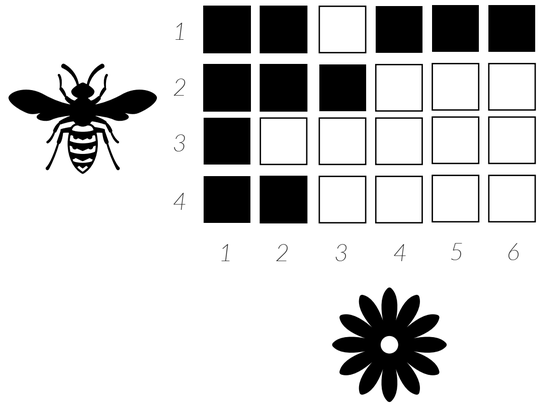
- *a z-score is a useful statistic for calculating the probability of a value occurring within a normal distribution with a mean of 0 and standard deviation of 1.*
- *a pattern is statistically significant ($p < 0.05$) if $Z > 1.96$.*
- *the p-value tells us how likely it is that the pattern we observed could have arisen by chance*

use case II: comparing across networks

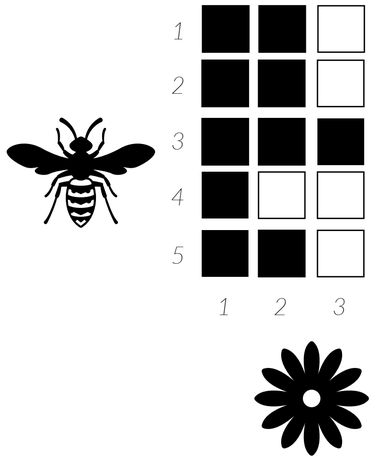
community a



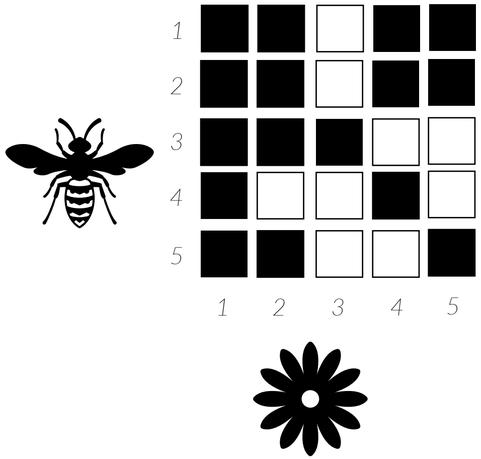
community b



community c



community d



The raw value of nestedness, for example, depends on matrix size, shape, and filling.

It cannot be compared across communities!

use case II: comparing across networks

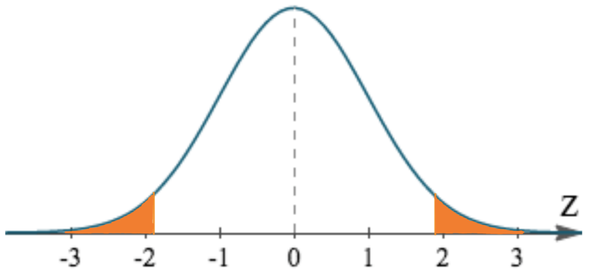
observed value

average across replicates

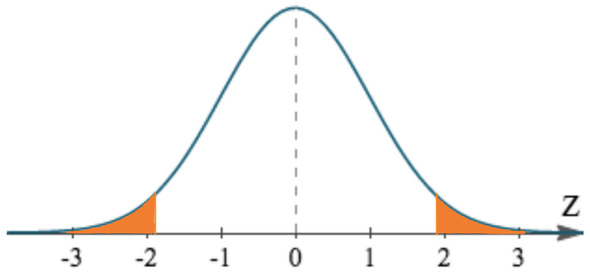
$$z = \frac{x_i - \overline{x}}{\sigma}$$

standard deviation across replicates

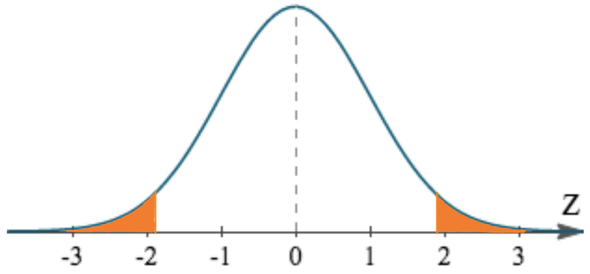
community a



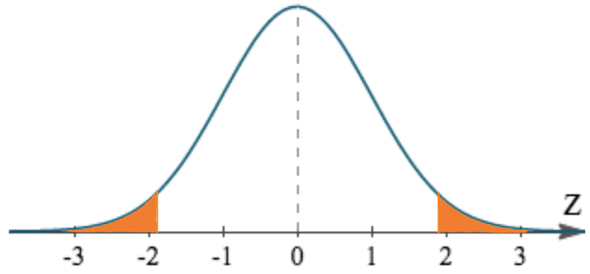
community b



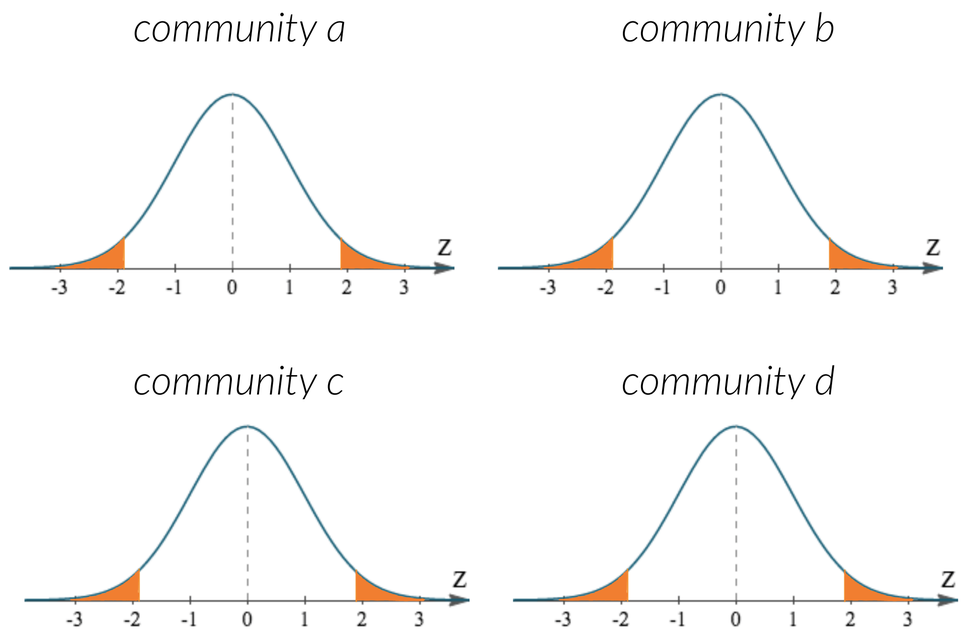
community c



community d



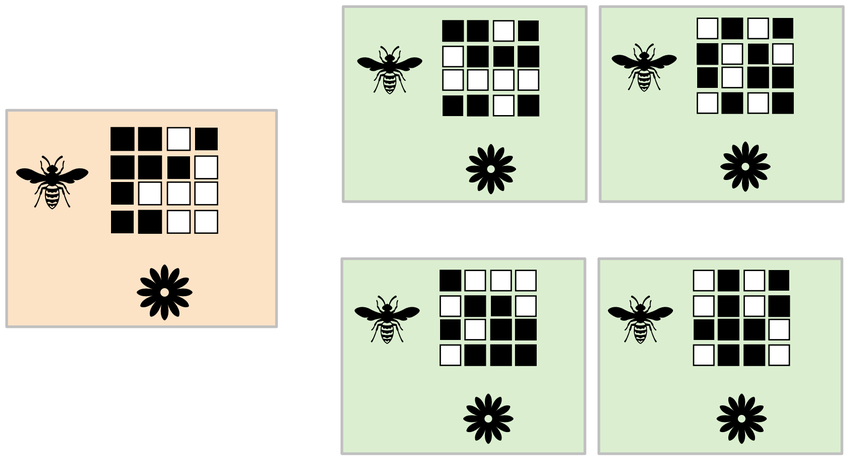
use case II: comparing across networks



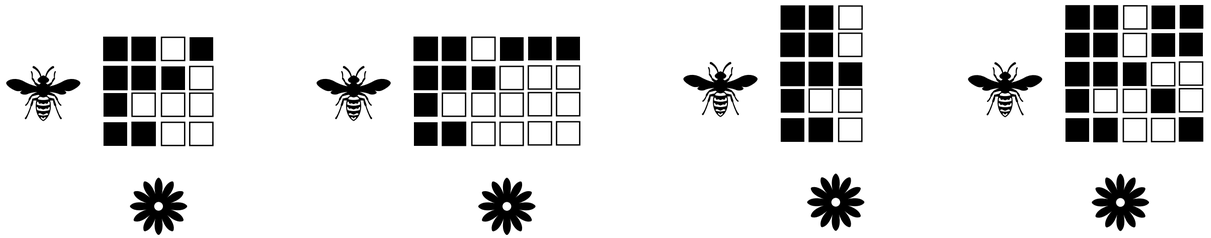
- *a z-score is a useful statistic for normalising a value and therefore making it comparable across systems.*
- *Relative nestedness (z-score) can be used to compare across communities.*

use cases of null models

significance of a network pattern



comparing across networks



caveats of null models

which null model to use?

- The goal of a null model strategy is to construct a model that deliberately excludes a mechanism being tested.

which null model to use?

- Keep in mind what we want to fix! (e.g. only the total number of “1”s, the number of “1”s per row, number of “1”s per row and column, …).

- Understand the limitations of each null model in the context of the trade-off between errors of type 1 (mistakenly rejecting a true null hypothesis) and errors of type 2 (mistakenly accepting a wrong null hypothesis).

type I and type II error

decision about null hypothesis (H_0)	null hypothesis (H_0) is true	null hypothesis (H_0) is false
accept	correct	type II error
reject	type I error	correct

The **type I error** occurs when the a true null hypothesis is rejected.

The **type II error** occurs when a false null hypothesis is erroneously failed to be rejected.

type I and type II error

decision about null hypothesis (H_0)	null hypothesis (H_0) is true	null hypothesis (H_0) is false
accept	correct	type II error
reject	type I error	correct

Example:
H0: water is contaminated
H1: water is not contaminated

The **type I error**: we erroneously conclude that water that is indeed contaminated is not contaminated (dangerous!).

The **type II error:** we erroneously conclude that water that is not contaminated is contaminated (not as bad?).

type I and type II error

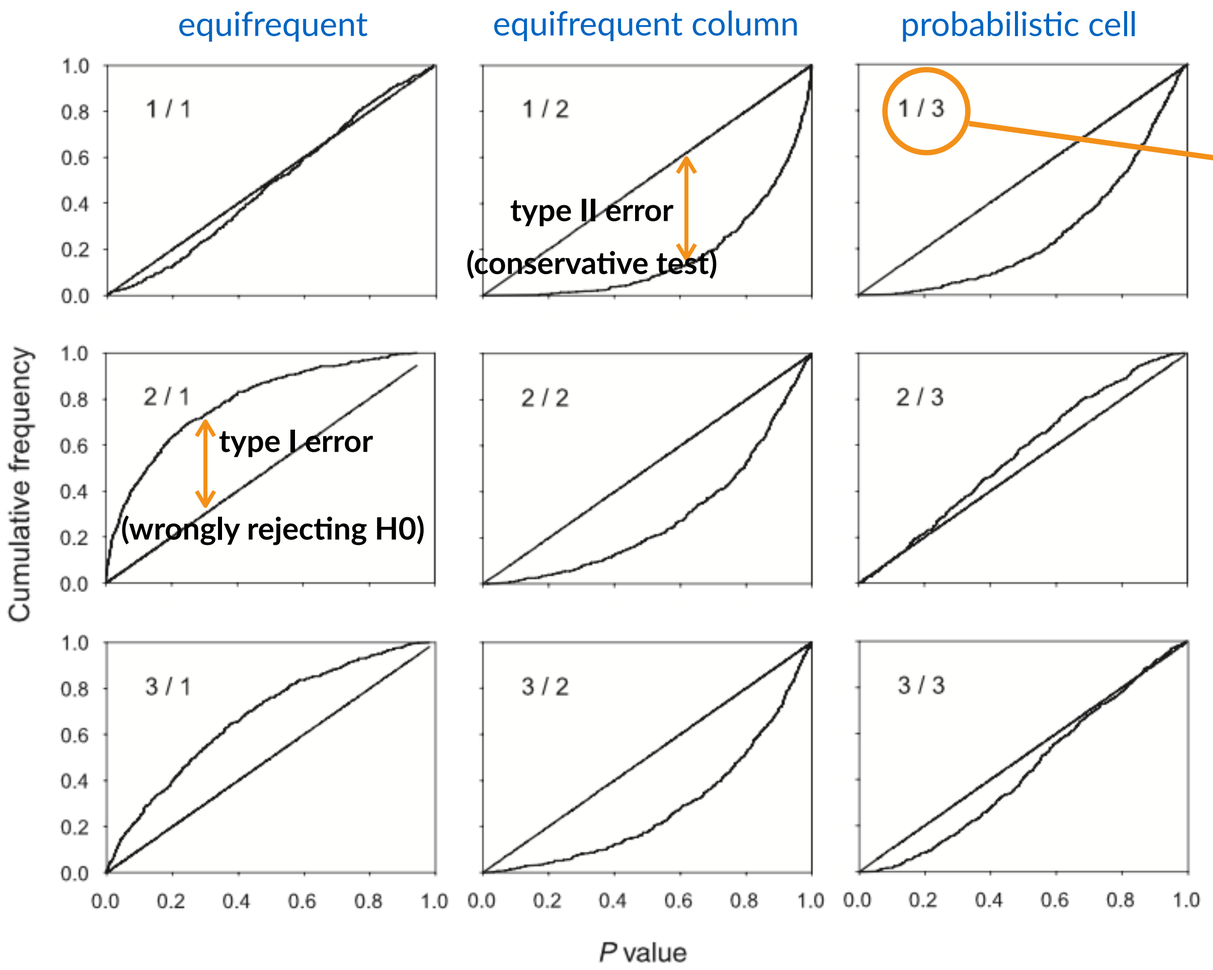
decision about null hypothesis (H_0)	null hypothesis (H_0) is true	null hypothesis (H_0) is false
accept	correct	type II error
reject	type I error	correct

Example:
H0: network is not nested
H1: network is nested

The **type I error**: we erroneously conclude that a network that is not nested is nested.

The **type II error:** we erroneously conclude that a network that is indeed nested is not nested (conservative tests).

null models and error types



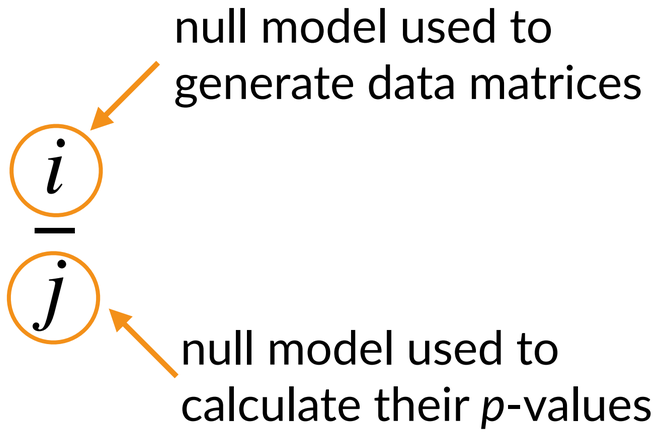


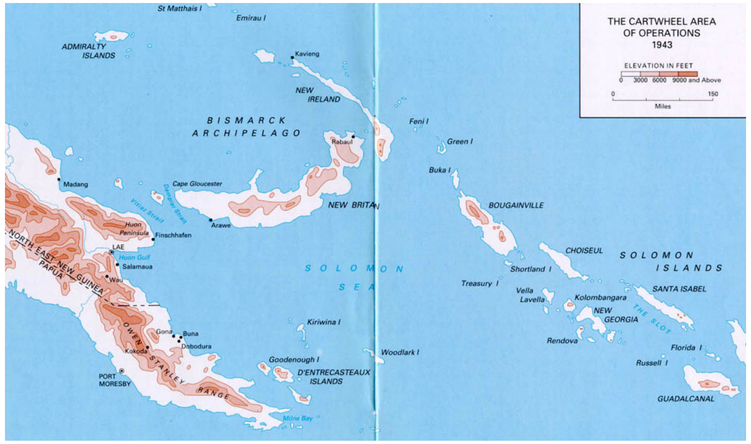
Figure 3 Cumulative frequency of the *p*-values associated with the nestedness temperature of random matrices. The label *i*/*j* in each plot indicates the null model used to generate data matrices (*i*) and to calculate their *p*-values (*j*). Straight lines have slope 1. Curves above the straight line represent a high tendency for the test to produce type I errors, while curves below the straight line represent conservative tests (high incidence of type II errors).

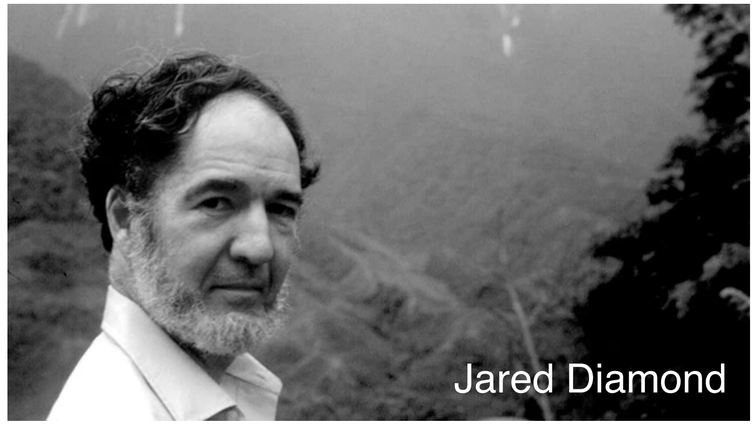
Rodríguez-Gironés & Santamaria (2006)

outline for this morning

i. *Null models in ecological research*
ii. *Null models in network research*
iii. *Examples of null models*
iv. *Use cases of null models*
v. *Caveats of null models*

back to Diamond's assembly rules





Ecology, 83(8), 2002, pp. 2091–2096

SPECIES CO-OCCURRENCE: A META-ANALYSIS OF J. M. DIAMOND'S ASSEMBLY RULES MODEL

NICHOLAS J. GOTELLI[1] AND DECLAN J. MCCABE[2]

Department of Biology, University of Vermont, Burlington, Vermont 05405 USA

Abstract. J. M. Diamond's assembly rules model predicts that competitive interactions between species lead to nonrandom co-occurrence patterns. We conducted a meta-analysis of 96 published presence–absence matrices and used a realistic "null model" to generate patterns expected in the absence of species interactions. Published matrices were highly nonrandom and matched the predictions of Diamond's model: there were fewer species combinations, more checkerboard species pairs, and less co-occurrence in real matrices than expected by chance. Moreover, nonrandom structure was greater in homeotherm vs. poikilotherm matrices. Although these analyses do not confirm the mechanisms of Diamond's controversial assembly rules model, they do establish that observed co-occurrence in most natural communities is usually less than expected by chance. These results contrast with previous analyses of species co-occurrence patterns and bridge the apparent gap between experimental and correlative studies in community ecology.

Key words: *community assembly rules; meta-analysis; null models; presence–absence matrix; species co-occurrence.*

outline for this afternoon

1. *using null models “by hand”*
2. *testing the significance of a network pattern*
3. *comparing network patterns across communities*