SUPERVISED LEARNING METHODS TO PREDICT SPECIES INTERACTIONS

UNIVERSITEIT GENT

FACULTY OF BIOSCIENCE ENGINEERING

KERMIT

fwo

Michiel Stock
@michielstock

SPECIES INTERACTION NETWORKS

Networks in ecology:

food webs

parasitism

pollination

SPECIES INTERACTION NETWORKS

Networks in ecology:

food webs

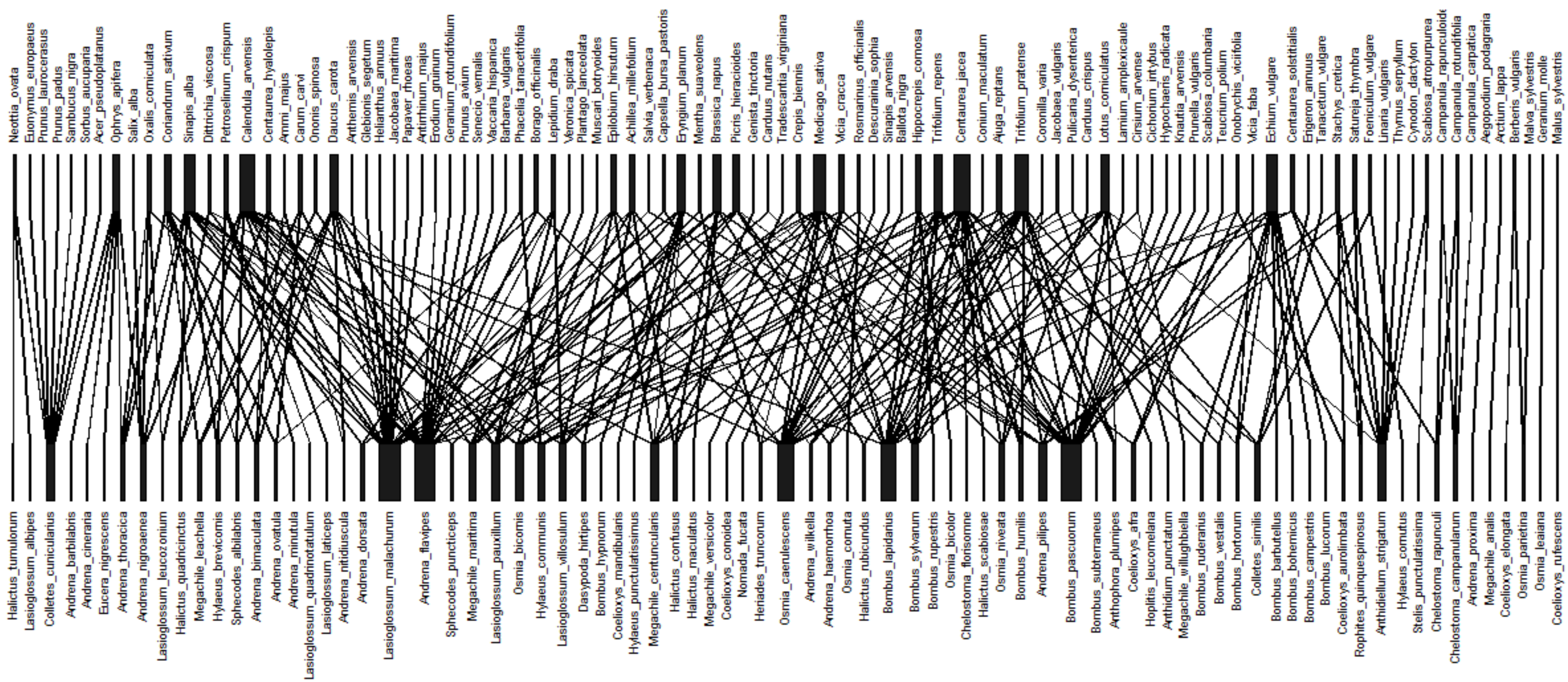
parasitism

pollination

Sampling of species interaction networks:

FLORABEILLES POLLINATION DATASET

Large pollination network containing 305 pollinator species (bees) and 452 plant species.



Density: 1.10%

POLLINATION NETWORK AS A MATRIX

1	1	0	0	0	0	1	0
1	0	0	1	0	0	1	1
1	1	1	0	0	0	0	0
0	0	0	0	0	0	0	1
1	0	1	0	1	0	0	0
1	1	1	0	0	1	1	1
0	0	0	1	1	0	0	0
0	0	0	0	0	1	0	0
1	1	1	0	0	0	0	0
1	1	1	0	1	0	0	1
1	1	1	1	0	1	0	0

POLLINATION NETWORK AS A MATRIX

Y

1	1	0	0	0	0	1	0
1	0	0	1	0	0	1	1
1	1	1	0	0	0	0	0
0	0	0	0	0	0	0	1
1	0	1	0	1	0	0	0
1	1	1	0	0	1	1	1
0	0	0	1	1	0	0	0
0	0	0	0	0	1	0	0
1	1	1	0	0	0	0	0
1	1	1	0	1	0	0	1
1	1	1	1	0	1	0	0

POLLINATION NETWORK AS A MATRIX

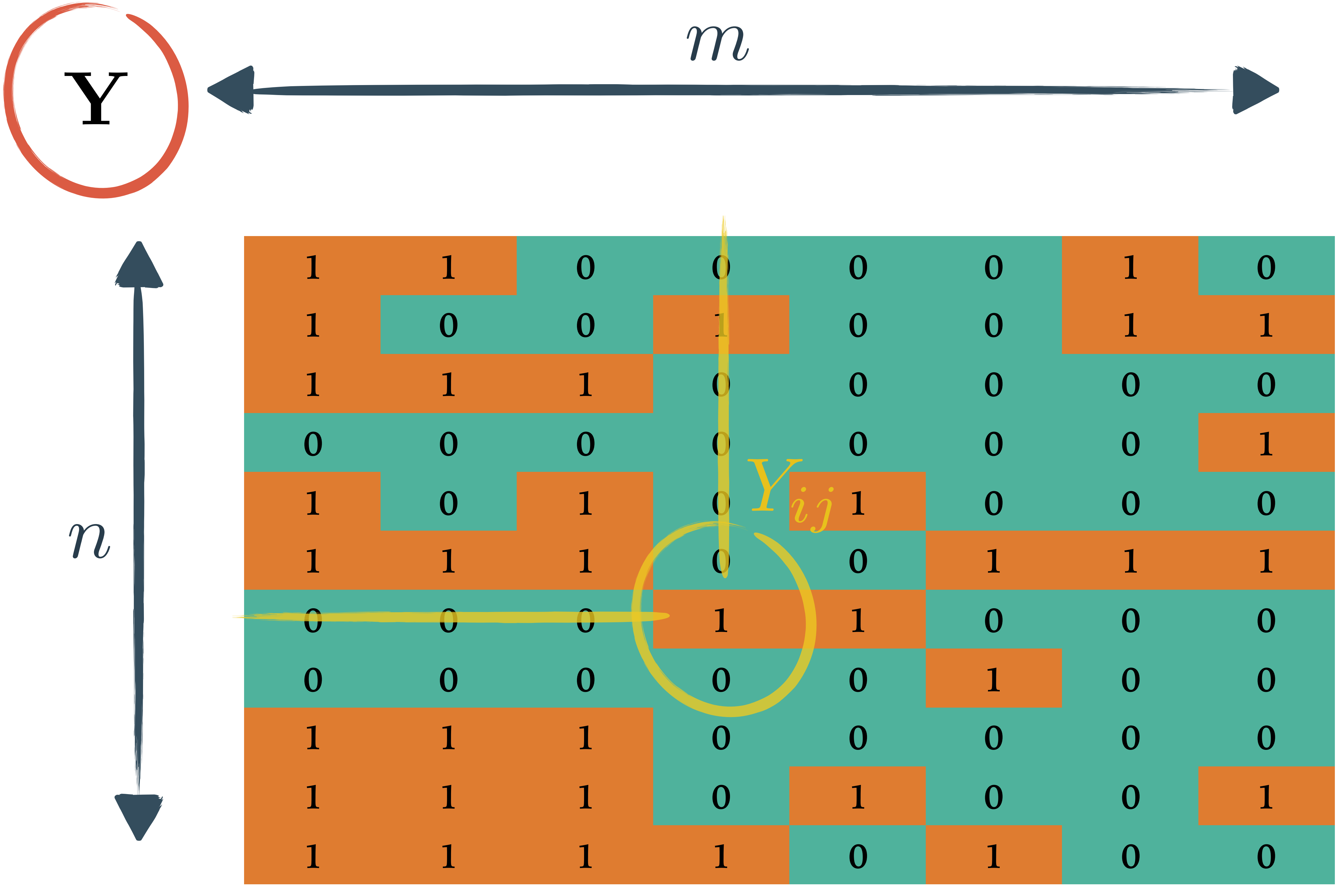
$\mathbf{Y}$

m

n

1	1	0	0	0	0	1	0
1	0	0	1	0	0	1	1
1	1	1	0	0	0	0	0
0	0	0	0	0	0	0	1
1	0	1	0	1	0	0	0
1	1	1	0	0	1	1	1
0	0	0	1	1	0	0	0
0	0	0	0	0	1	0	0
1	1	1	0	0	0	0	0
1	1	1	0	1	0	0	1
1	1	1	1	0	1	0	0

POLLINATION NETWORK AS A MATRIX



1	1	0	0	0	0	1	0
1	0	0	1	0	0	1	1
1	1	1	0	0	0	0	0
0	0	0	0	0	0	0	1
1	0	1	0	1	0	0	0
1	1	1	0	0	1	1	1
0	0	0	1	1	0	0	0
0	0	0	0	0	1	0	0
1	1	1	0	0	0	0	0
1	1	1	0	1	0	0	1
1	1	1	1	0	1	0	0

A PAIRWISE MODEL

'Learn' a pairwise function based on observed data:

$f($  $,$  $)$

such that a high score indicates two species will interact.

A PAIRWISE MODEL

‘Learn’ a pairwise function based on observed data:

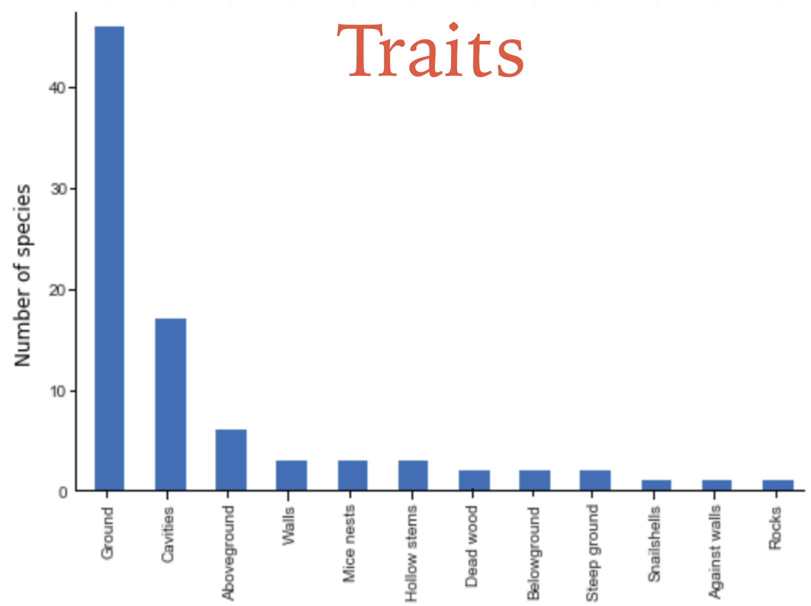
f( , )

such that a high score indicates two species will interact.

How to describe the species?

DESCRIBING THE BEES

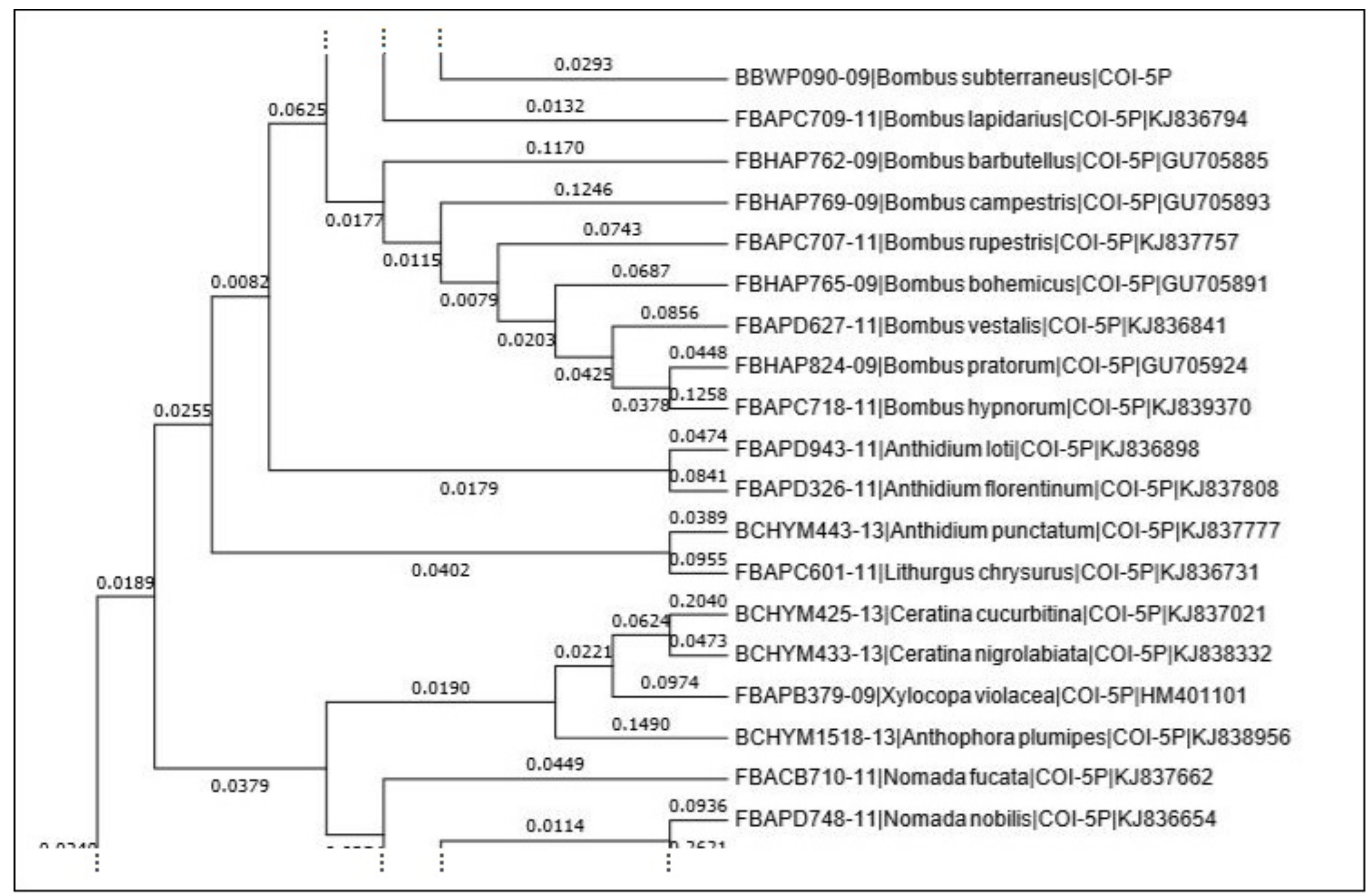
Traits



nesting times

flight times

Phylogeny



based on Cytochrome c oxidase

DESCRIBING THE PLANTS

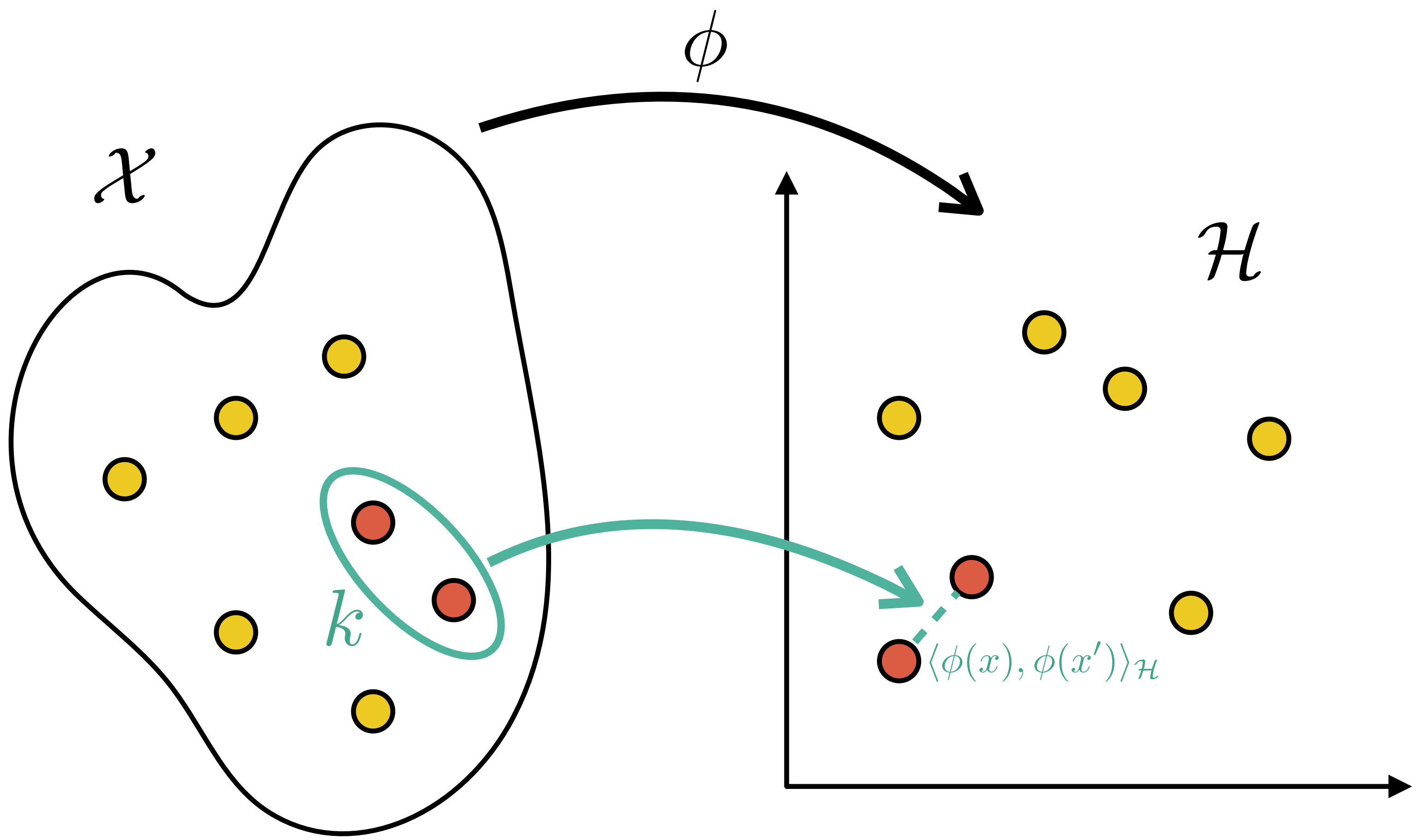
Traits

Growth habit	Categorical	Shrub
Minimum height (cm)	Numerical	50
Maximum height (cm)	Numerical	200
Mean height (cm)	Numerical	125
Blooming period	Dummy variables	[0,0,0,0,0,1,1,0,0,0,0,0]
Duration	Categorical	Perennial
Category	Categorical	Dicot
Flower colour	Dummy variables	[1,0,0,0,0,0,0,0]
Phyllotaxis	Categorical	Opposite decussated
Flower symmetry	Categorical	Versatile symmetrical
Position ovary	Categorical	Superior
Number of styles	Numerical	1
Number of stamens	Numerical	2

Phylogeny based on the *rcbL* gene.

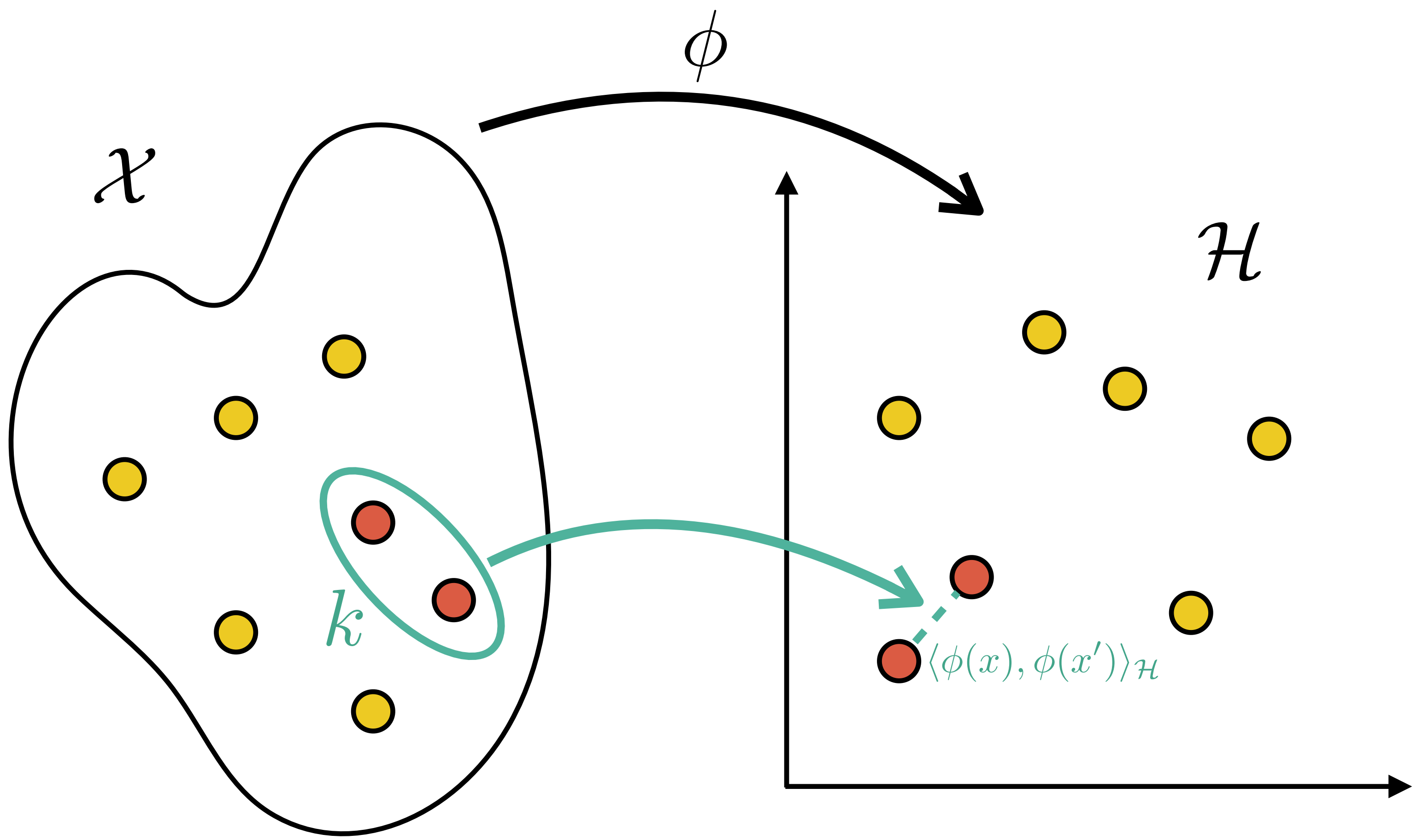
THE KERNEL TRICK

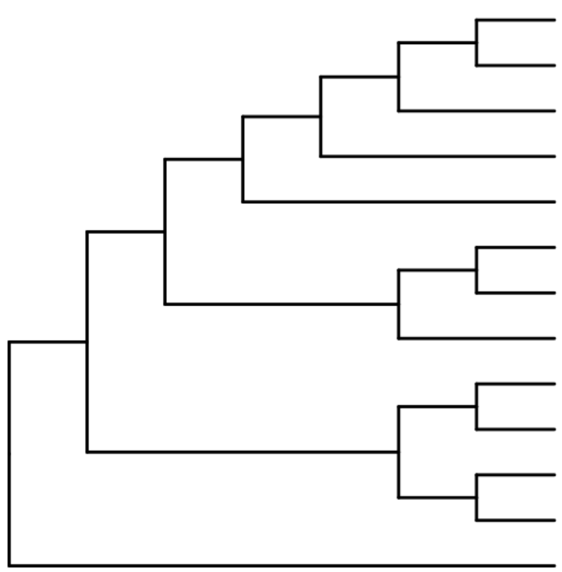
Kernels represent species by implied products in high-dimensional space.



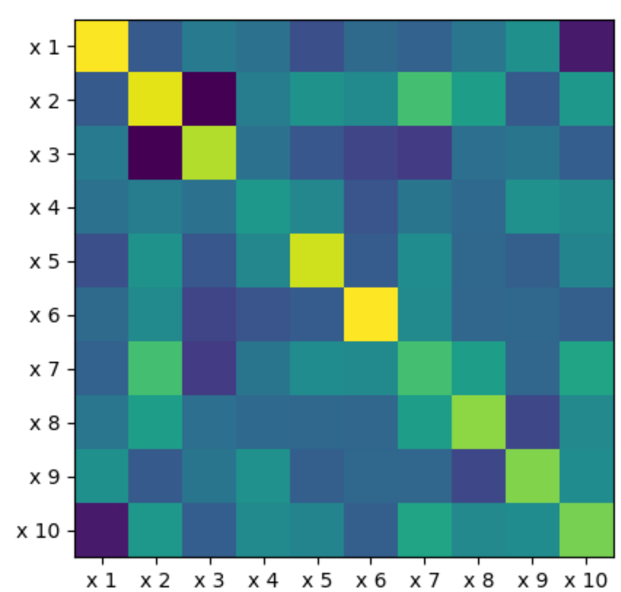
THE KERNEL TRICK

Kernels represent species by implied products in high-dimensional space.





numerical or structured description of species



Kernel matrix describes similarity between species i and j.

TWO-STEP KERNEL RIDGE REGRESSION

Let u and v denote the bees and plants, respectively.

We learn a pairwise function of the form

$$f(u, v) = \sum_{i,j} W_{ij} k(u, u_i) g(v, v_j)\,.$$

TWO-STEP KERNEL RIDGE REGRESSION

Let u and v denote the bees and plants, respectively.

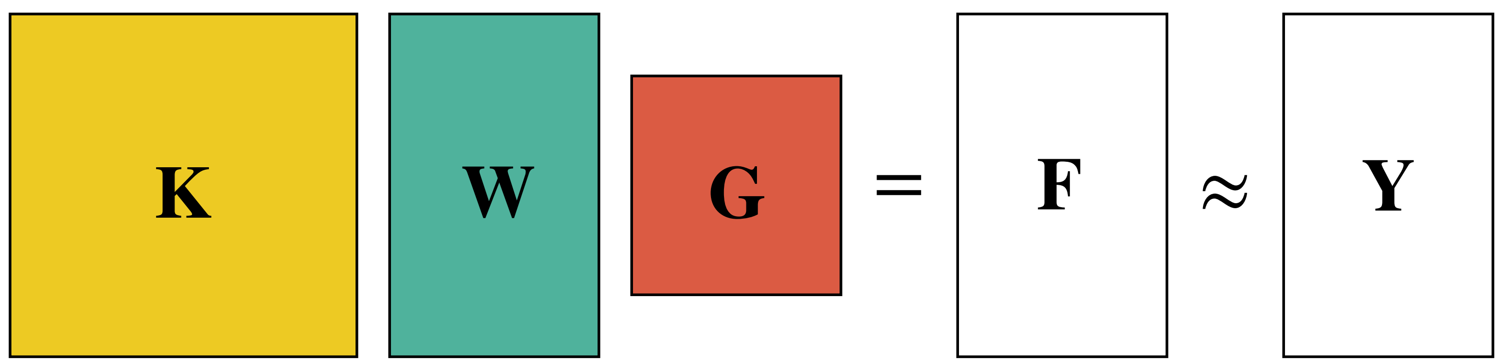
We learn a pairwise function of the form

kernel over bees

$$f(u,v) = \sum_{i,j} W_{ij} k(u,u_i) g(v,v_j)\,.$$

weights

kernel over plants



TWO-STEP KERNEL RIDGE REGRESSION

Let u and v denote the bees and plants, respectively.

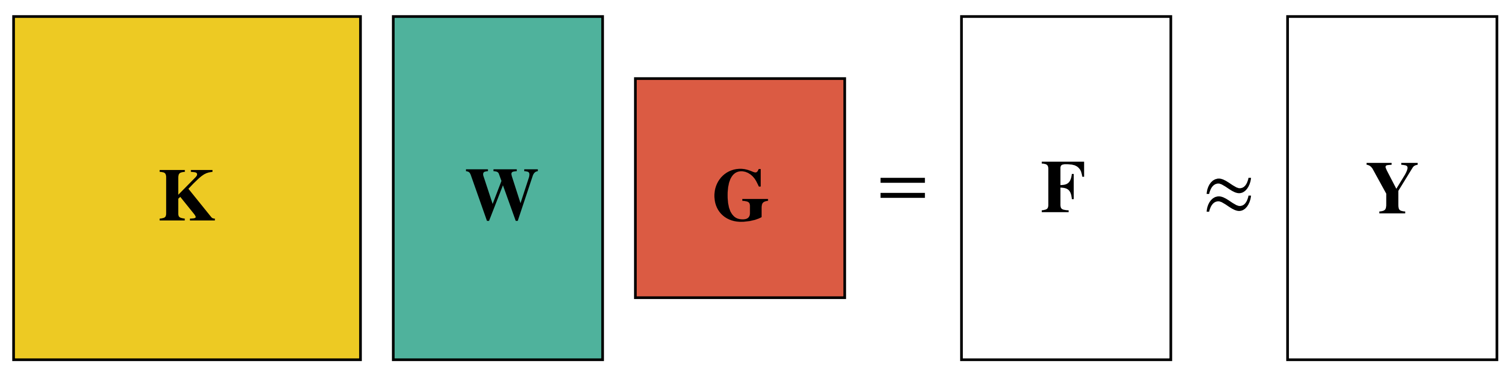
We learn a pairwise function of the form

kernel over bees

$$f(u,v) = \sum_{i,j} W_{ij} k(u, u_i) g(v, v_j)\,.$$

weights

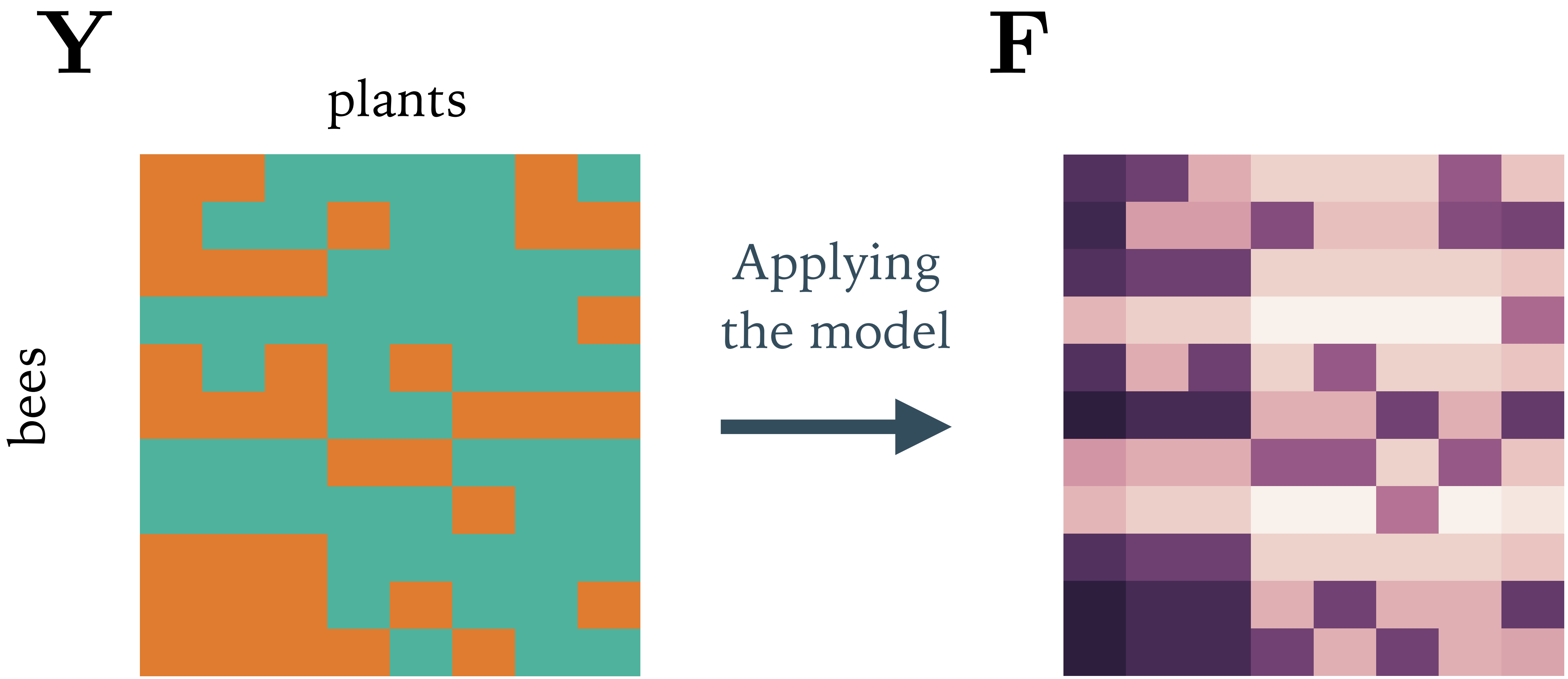
kernel over plants



The weights can be found by computing:

$$\mathbf{W} = (K + \lambda_u \mathbb{I})^{-1} \mathbf{Y} (G + \lambda_v \mathbb{I})^{-1}\,.$$

FROM OBSERVATIONS TO PREDICTIONS



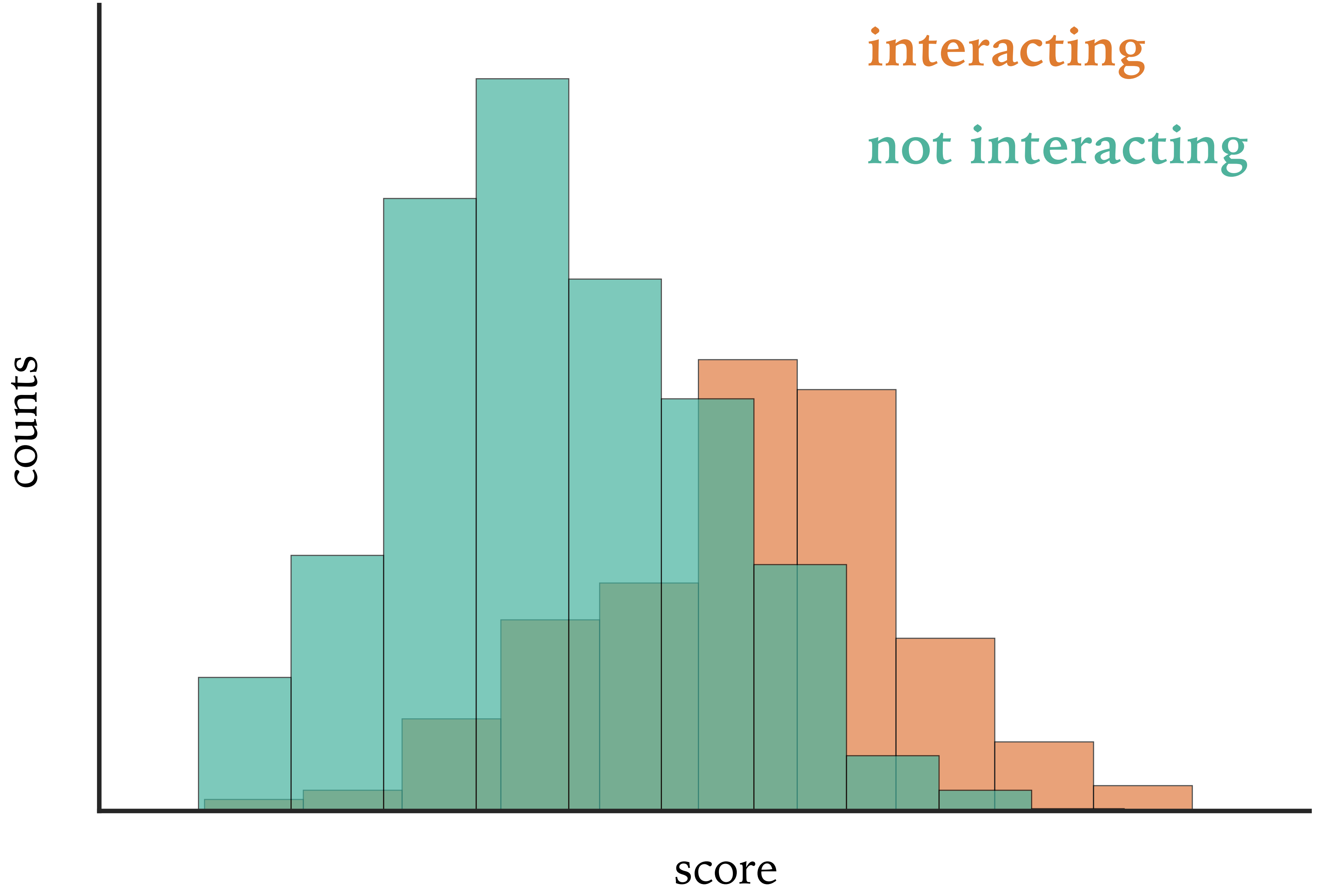
Observed interactions

Rescored interactions

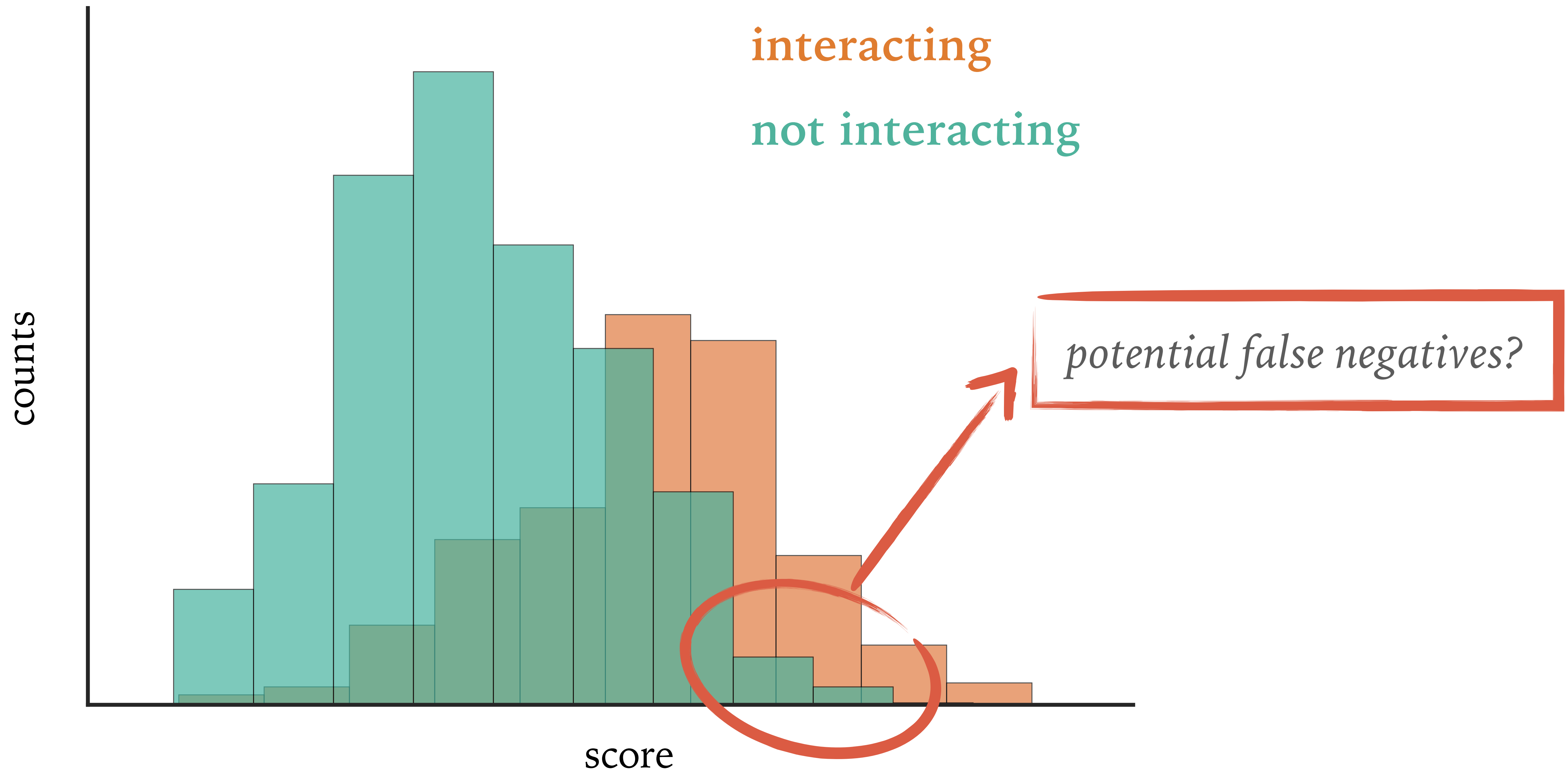
observed interaction

no interaction

USING THE SCORED INTERACTIONS



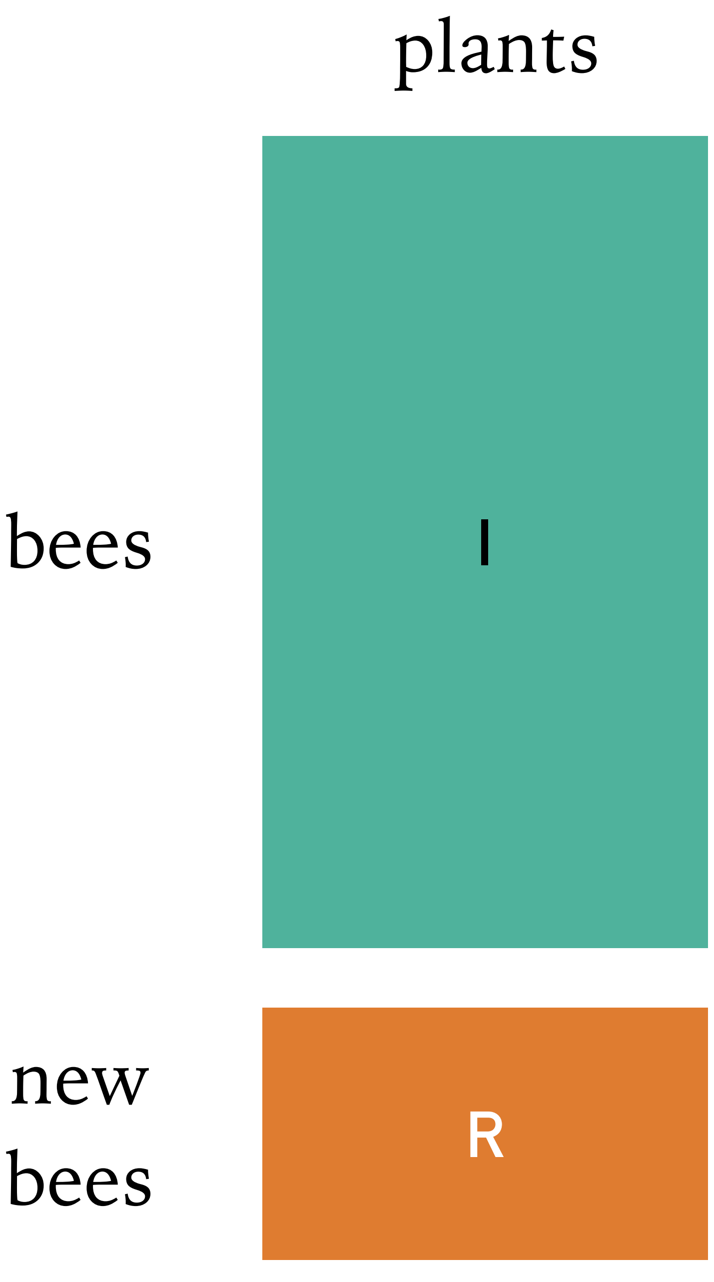
USING THE SCORED INTERACTIONS



FOUR SETTINGS FOR PAIRWISE PREDICTION

Setting I: same bees, same plants

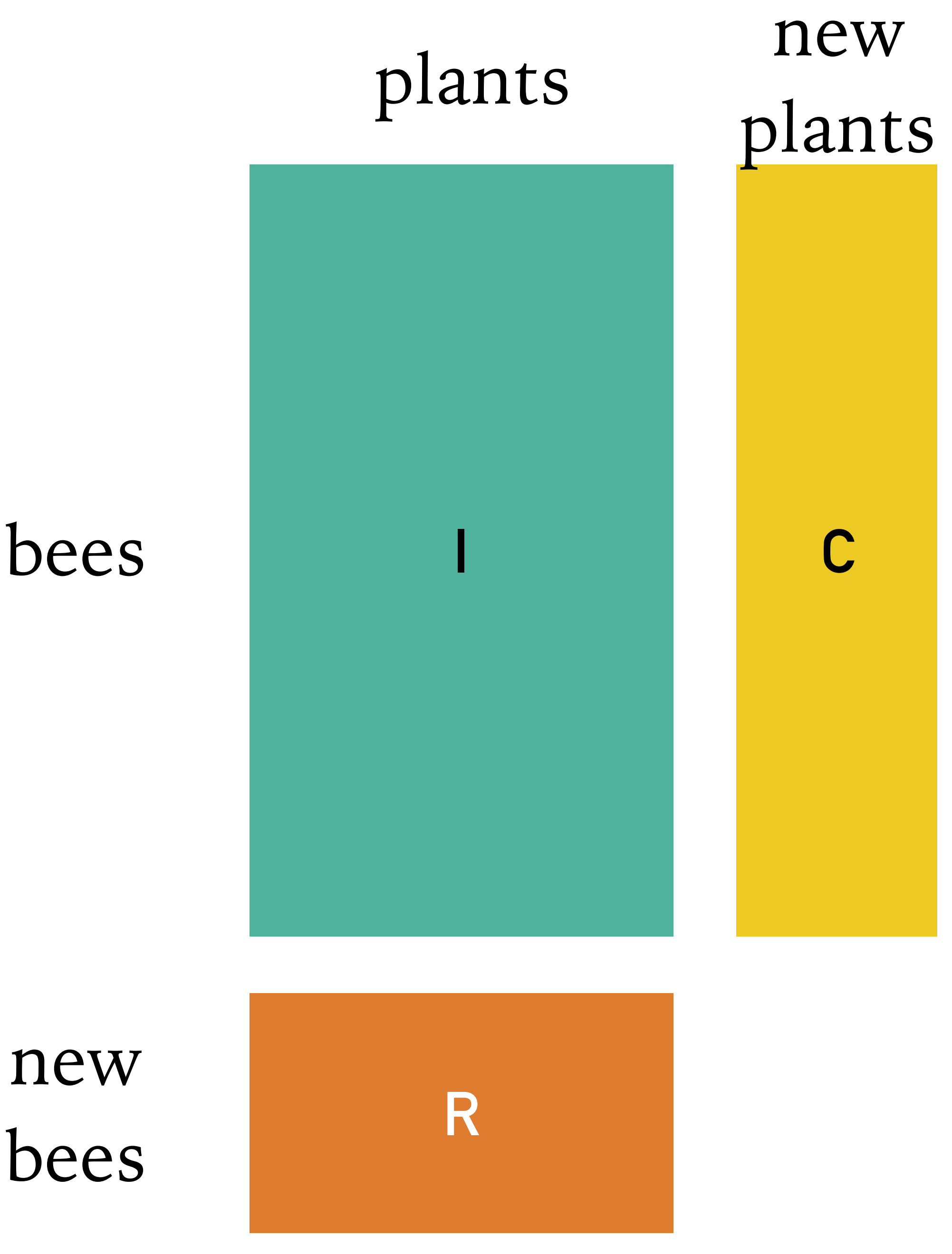
FOUR SETTINGS FOR PAIRWISE PREDICTION



Setting I: same bees, same plants

Setting R: new bees, same plants

FOUR SETTINGS FOR PAIRWISE PREDICTION

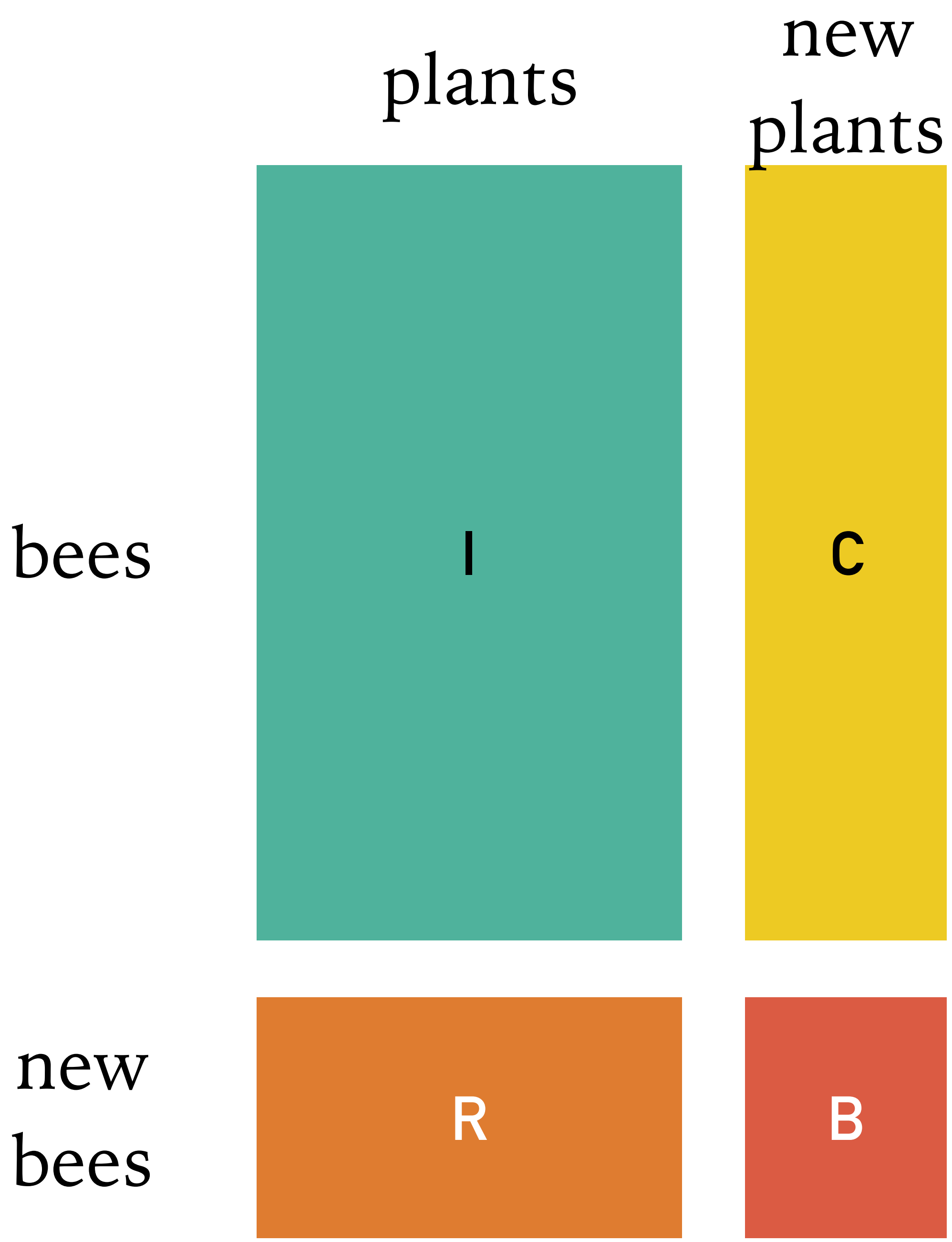


Setting I: same bees, same plants

Setting R: new bees, same plants

Setting C: same bees, new plants

FOUR SETTINGS FOR PAIRWISE PREDICTION



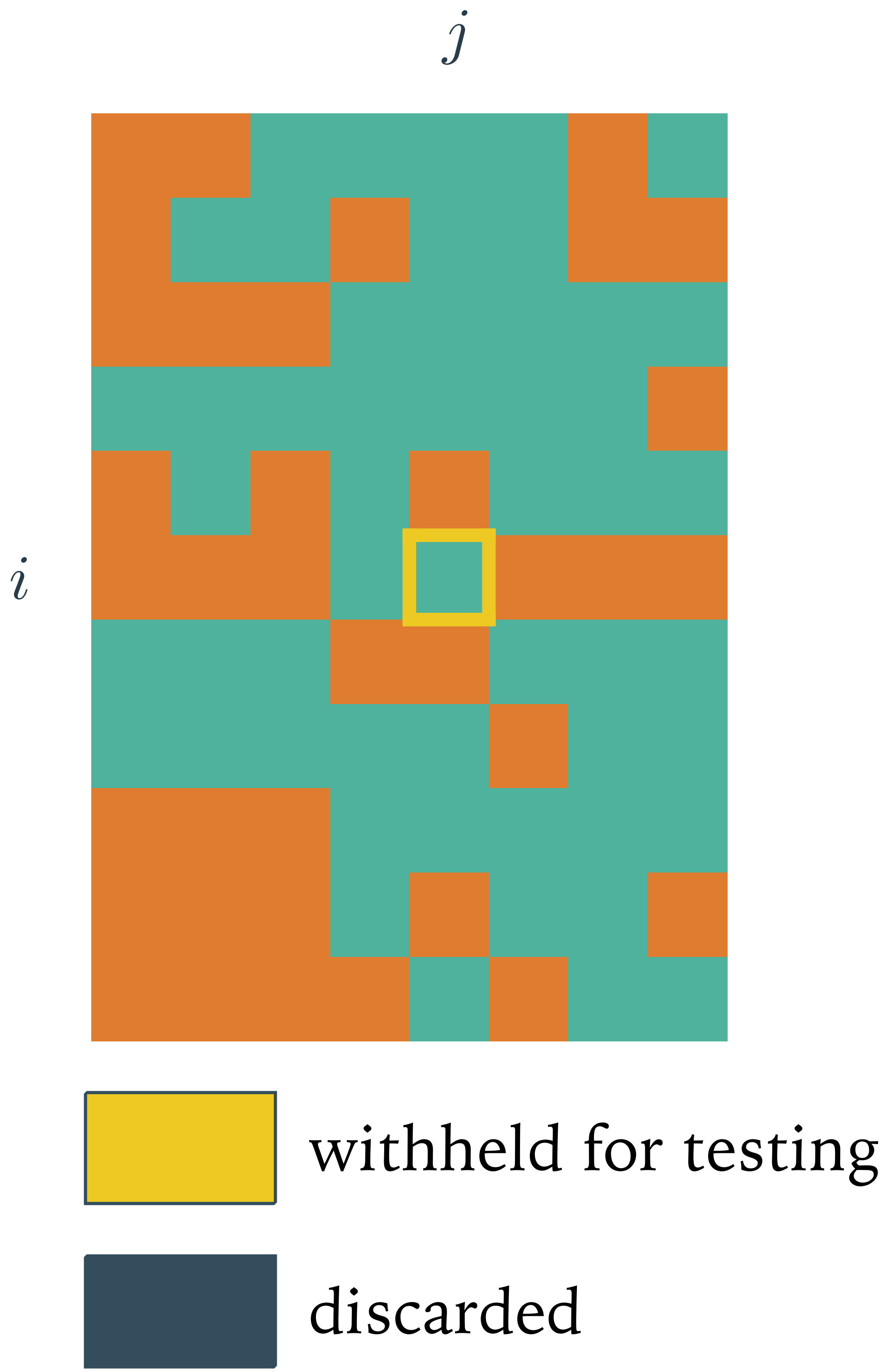
Setting I: same bees, same plants

Setting R: new bees, same plants

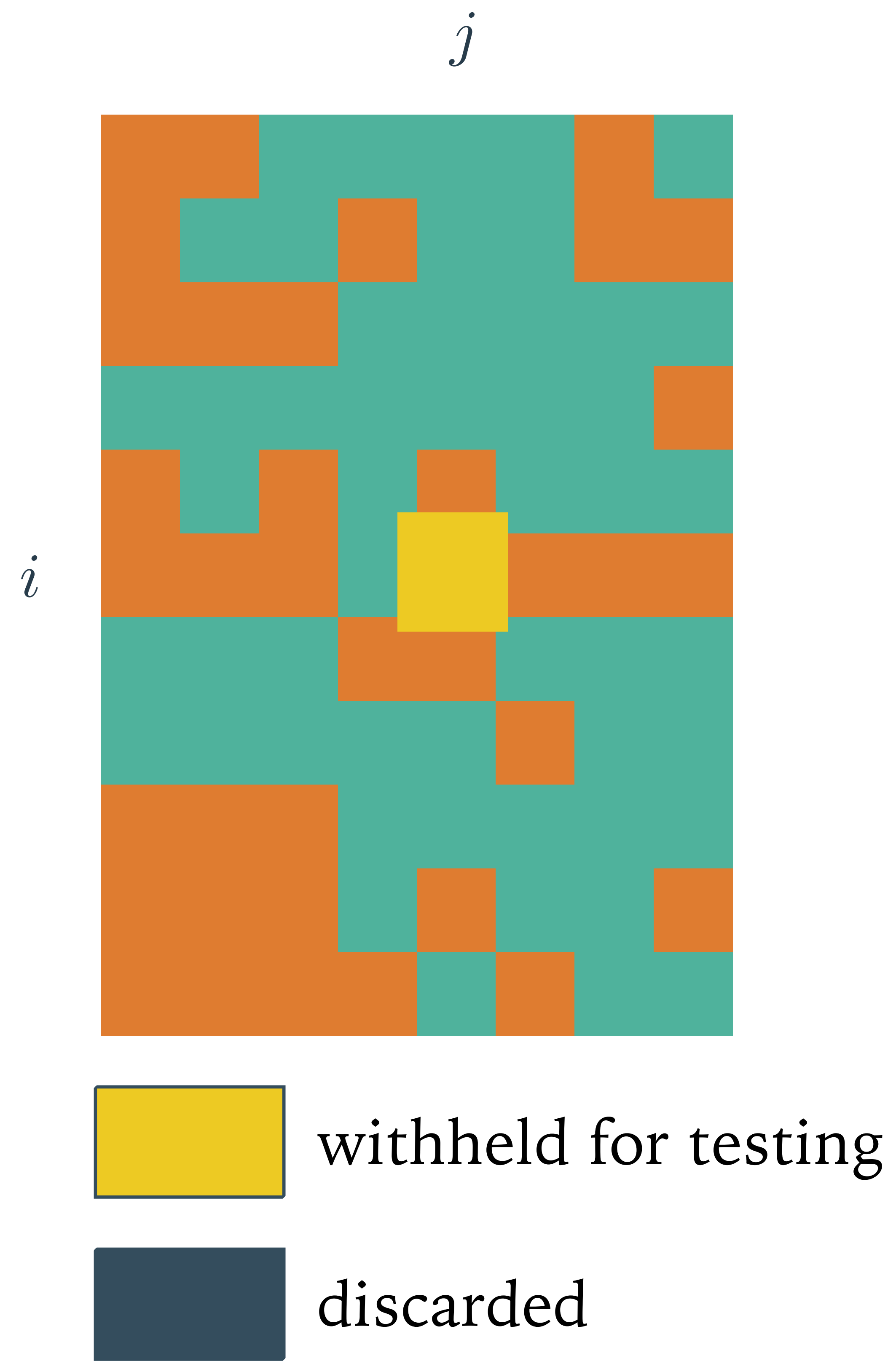
Setting C: same bees, new plants

Setting B: new bees, new plants

CROSS-VALIDATION IN THE FOUR SETTINGS

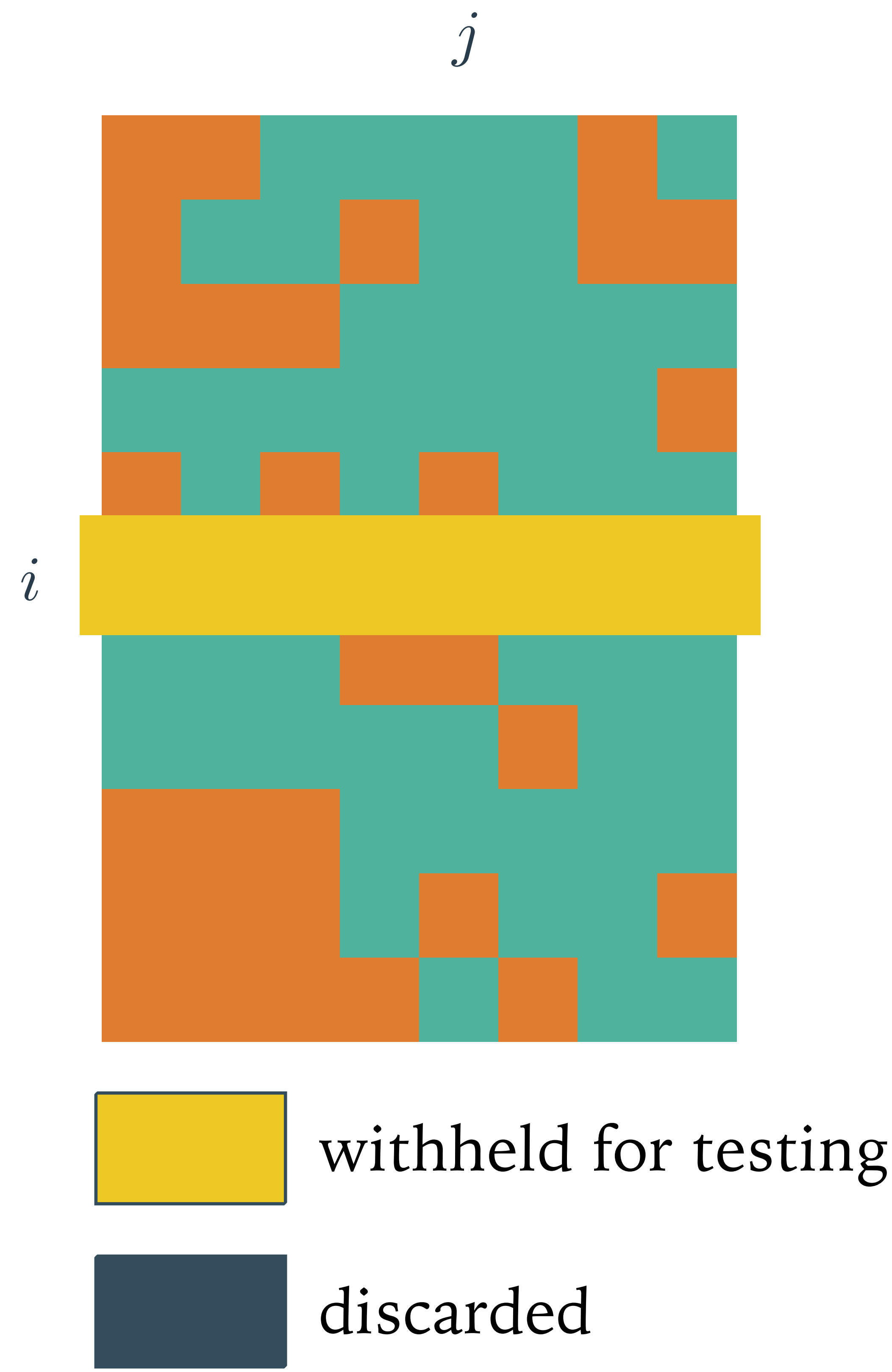


CROSS-VALIDATION IN THE FOUR SETTINGS



Setting I: leave out individual pairs

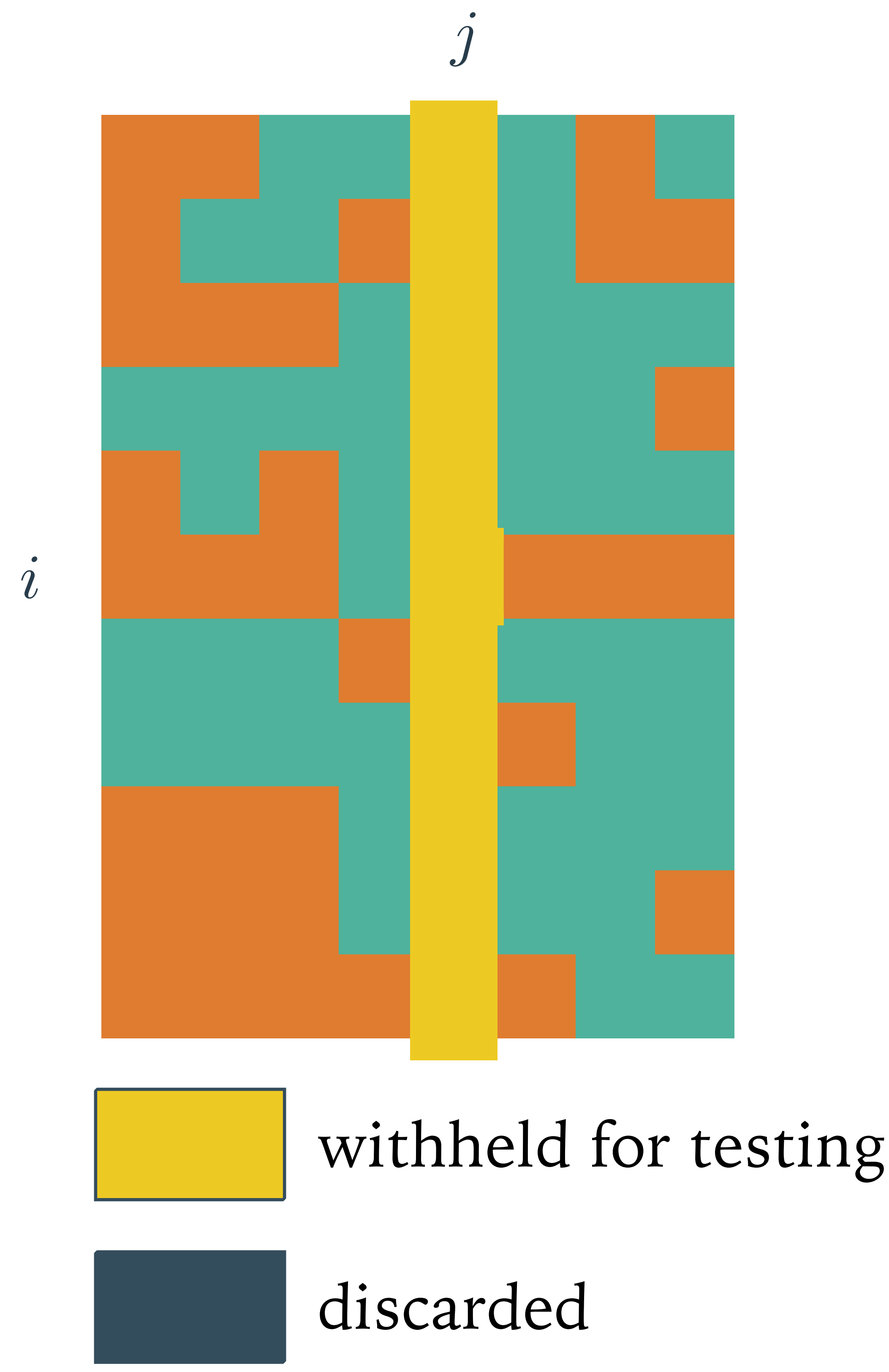
CROSS-VALIDATION IN THE FOUR SETTINGS



Setting I: leave out individual pairs

Setting R: leave out rows

CROSS-VALIDATION IN THE FOUR SETTINGS

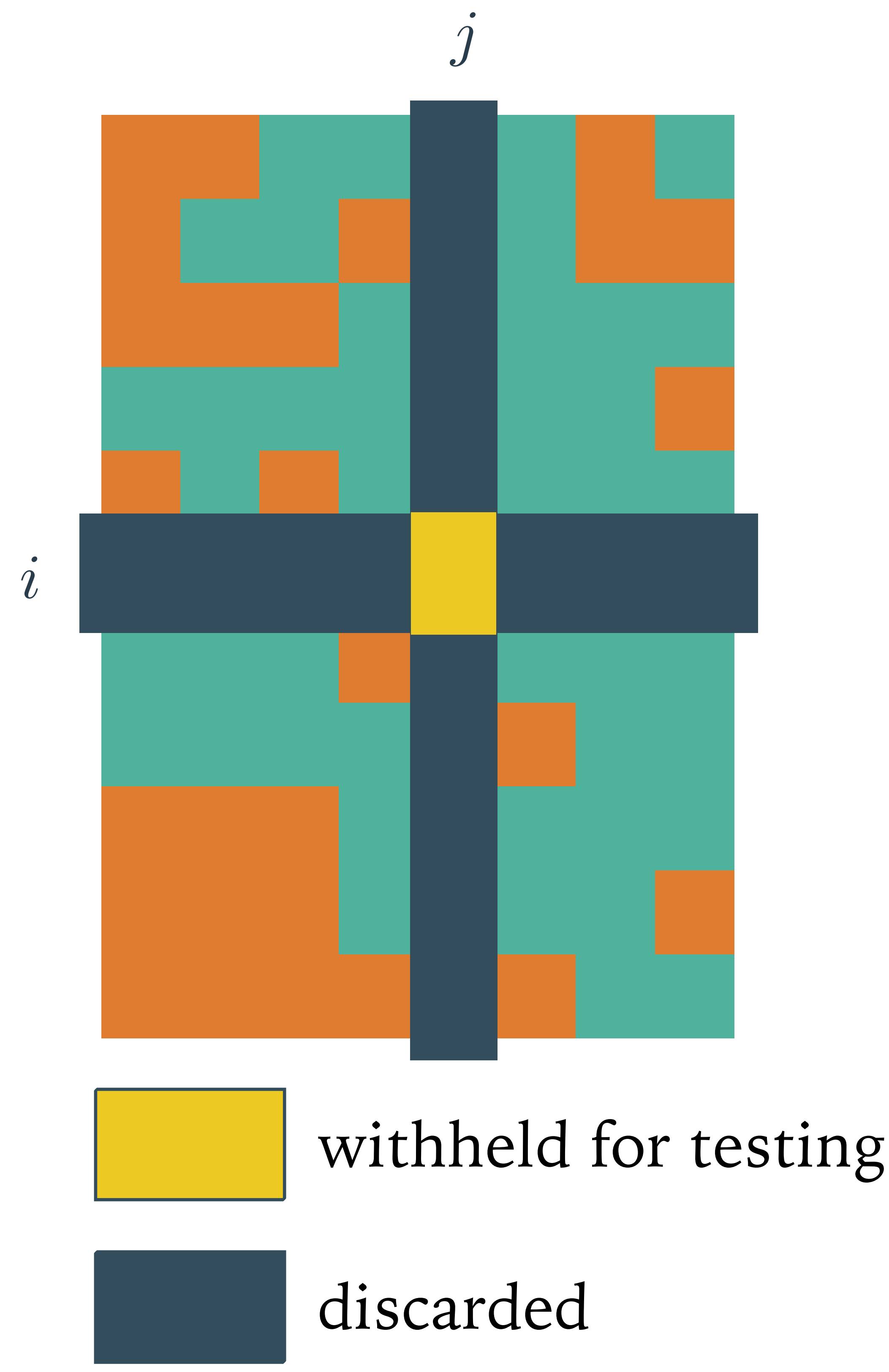


Setting I: leave out individual pairs

Setting R: leave out rows

Setting C: leave out columns

CROSS-VALIDATION IN THE FOUR SETTINGS



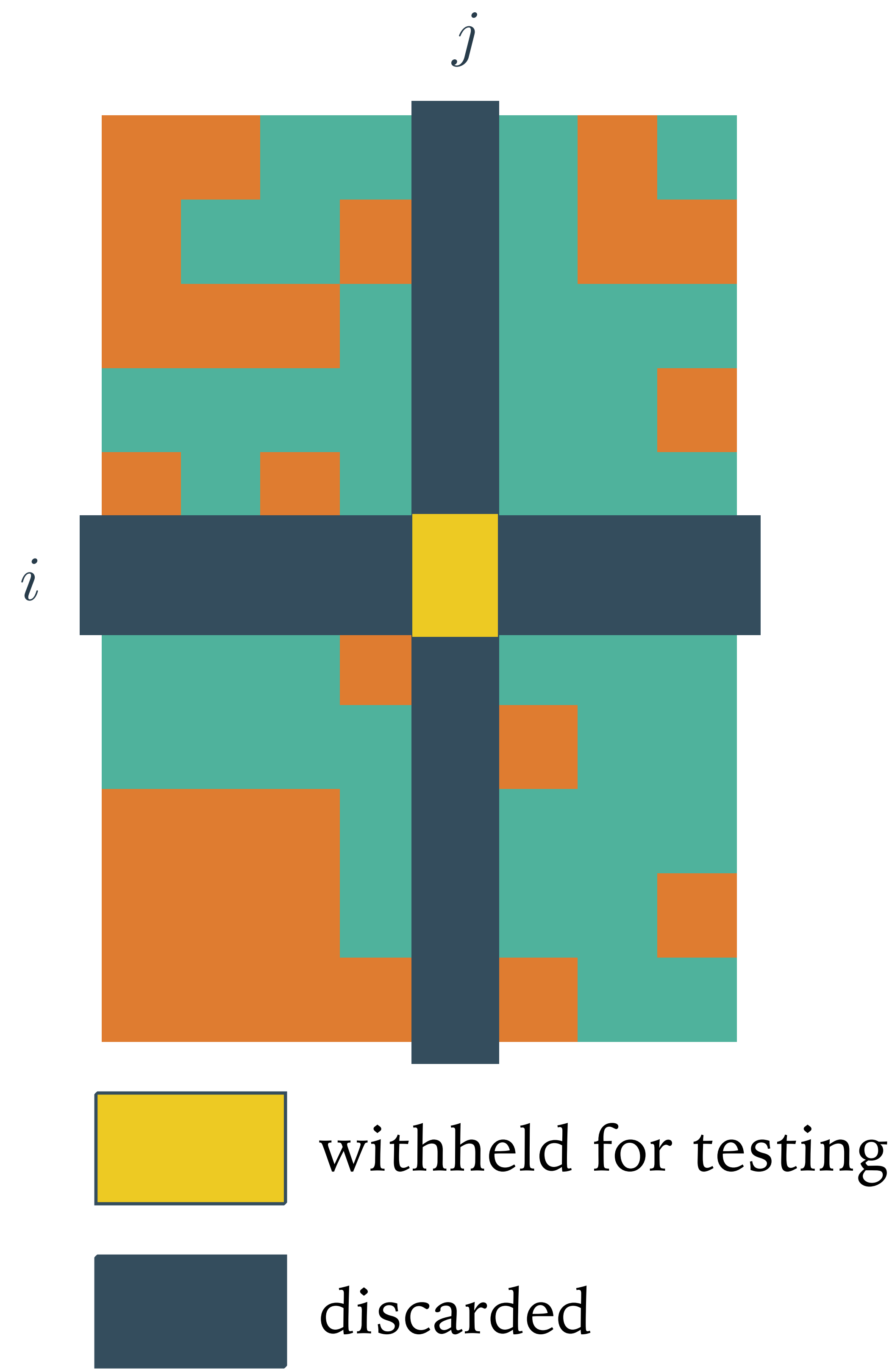
Setting I: leave out individual pairs

Setting R: leave out rows

Setting C: leave out columns

Setting B: leave out each pair, discard other pairs in row and column

CROSS-VALIDATION IN THE FOUR SETTINGS



Setting I: leave out individual pairs

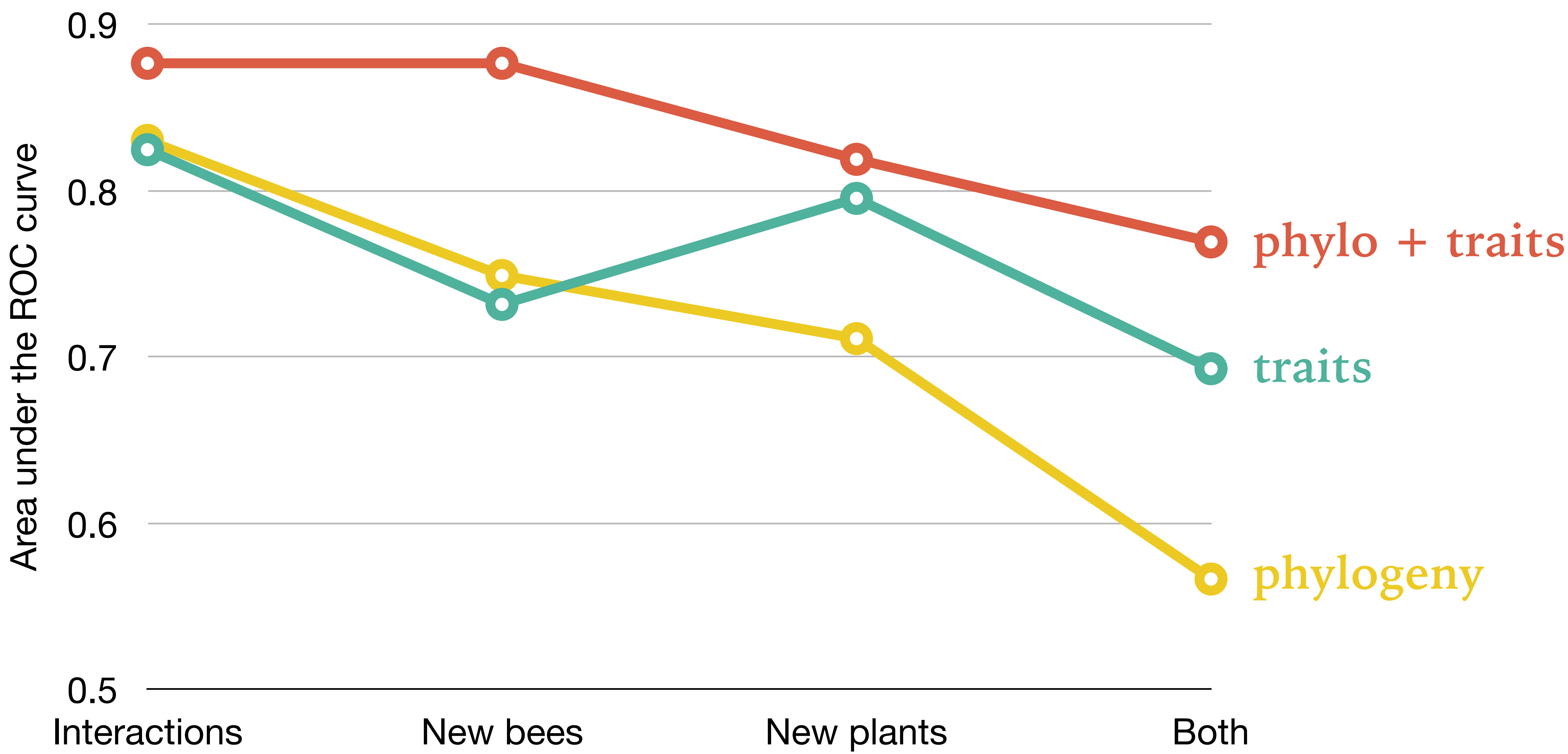
Setting R: leave out rows

Setting C: leave out columns

Setting B: leave out each pair, discard other pairs in row and column

Exact and efficient formulas for computing the leave-one-out values!

PERFORMANCE PREDICTING THE INTERACTIONS



CONCLUSIONS

1. Supervised network prediction based on kernels.
2. Two-step kernel ridge regression: a simple though powerful method.
3. Different prediction settings: use structured cross-validation methods!

CONCLUSIONS

1. Supervised network prediction based on kernels.
2. Two-step kernel ridge regression: a simple though powerful method.
3. Different prediction settings: use structured cross-validation methods!

xnet: an R-package for pairwise learning and cross-validation

ACKNOWLEDGEMENTS AND REFERENCES

pairwise learning

Sarah Vanbesien

Bernard De Baets

pollination case study

Niels Piot

Guy Smagghe

xnet

Joris Meys

ACKNOWLEDGEMENTS AND REFERENCES

pairwise learning | pollination case study | xnet





Sarah Vanbesien | **Bernard De Baets** | **Niels Piot** | **Guy Smagghe** | **Joris Meys**

- M. Stock, T. Poisot, W. Waegeman and B. De Baets, **Linear filtering reveals false negatives in species interaction data**, Scientific Reports **7** (2017), 45908.
- M. Stock, T. Pahikkala, A. Airola, B. De Baets and W. Waegeman, **A comparative study of pairwise learning methods based on Kernel Ridge Regression**, Neural Computation **30** (2018), 2245-2283.
- M. Stock, T. Pahikkala, A. Airola, W. Waegeman and B. De Baets, **Algebraic shortcuts for leave-one-out cross-validation in supervised network inference**, Briefings in Bioinformatics, accepted Sep 2018.