Persistent Homology of Plant-Pollinator Interactions: The Importance of Flowering Plant Surveys

Melinda Kleczynski [1], Chad Giusti [2], Anthony Kearsley [1]

[1] National Institute of Standards and Technology
[2] Oregon State University

Applied Algebraic Topology Research Network (AATRN) Online Seminar
August 21, 2024

NIST NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

Acknowledgements

- This research was performed while M. Kleczynski held a National Institute of Standards and Technology (NIST) National Research Council (NRC) Research Postdoctoral Associateship Award at the NIST Information Technology Laboratory.
- Data were provided by the H.J. Andrews Experimental Forest and Long Term Ecological Research (LTER) program, administered cooperatively by Oregon State University, the USDA Forest Service Pacific Northwest Research Station, and the Willamette National Forest. This material is based upon work supported by the National Science Foundation under the grant LTER8 DEB-2025755.

E. Seo, J. A. Jones, R. A. Hutchinson, and V. W. Pfeiffer, *Plant pollinator data at HJ Andrews Experimental Forest, 2011 to 2021*. https://andlter.forestry.oregonstate.edu/data/abstract.aspx?dbcode=SA026, Aug. 2022.

Overview

- There is increasing consumer interest in plant species which are beneficial to pollinators.

Overview

- There is increasing consumer interest in plant species which are beneficial to pollinators.
- Preference metrics provide a way to quantify how much a type of plant is favored by a type of pollinator.

Overview

- There is increasing consumer interest in plant species which are beneficial to pollinators.
- Preference metrics provide a way to quantify how much a type of plant is favored by a type of pollinator.
- Computing preferences requires quantitative interaction data and information about floral abundance.

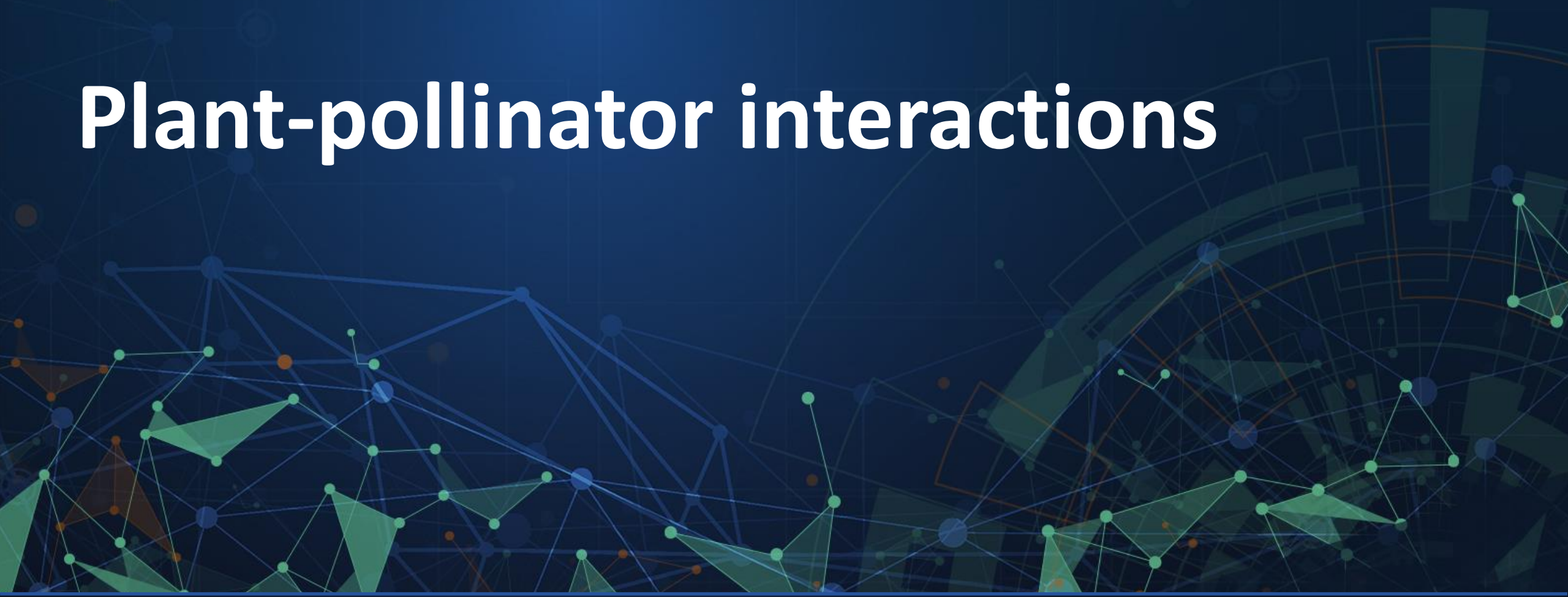
Overview

- There is increasing consumer interest in plant species which are beneficial to pollinators.
- Preference metrics provide a way to quantify how much a type of plant is favored by a type of pollinator.
- Computing preferences requires quantitative interaction data and information about floral abundance.
- Floral abundance can be estimated using interaction data or measured by performing a flowering plant survey.

Overview

- There is increasing consumer interest in plant species which are beneficial to pollinators.
- Preference metrics provide a way to quantify how much a type of plant is favored by a type of pollinator.
- Computing preferences requires quantitative interaction data and information about floral abundance.
- Floral abundance can be estimated using interaction data or measured by performing a flowering plant survey.
- We use topological data analysis to investigate whether the use of floral survey data affects the mathematical structure of observed plant-pollinator interactions.

Plant-pollinator interactions



NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

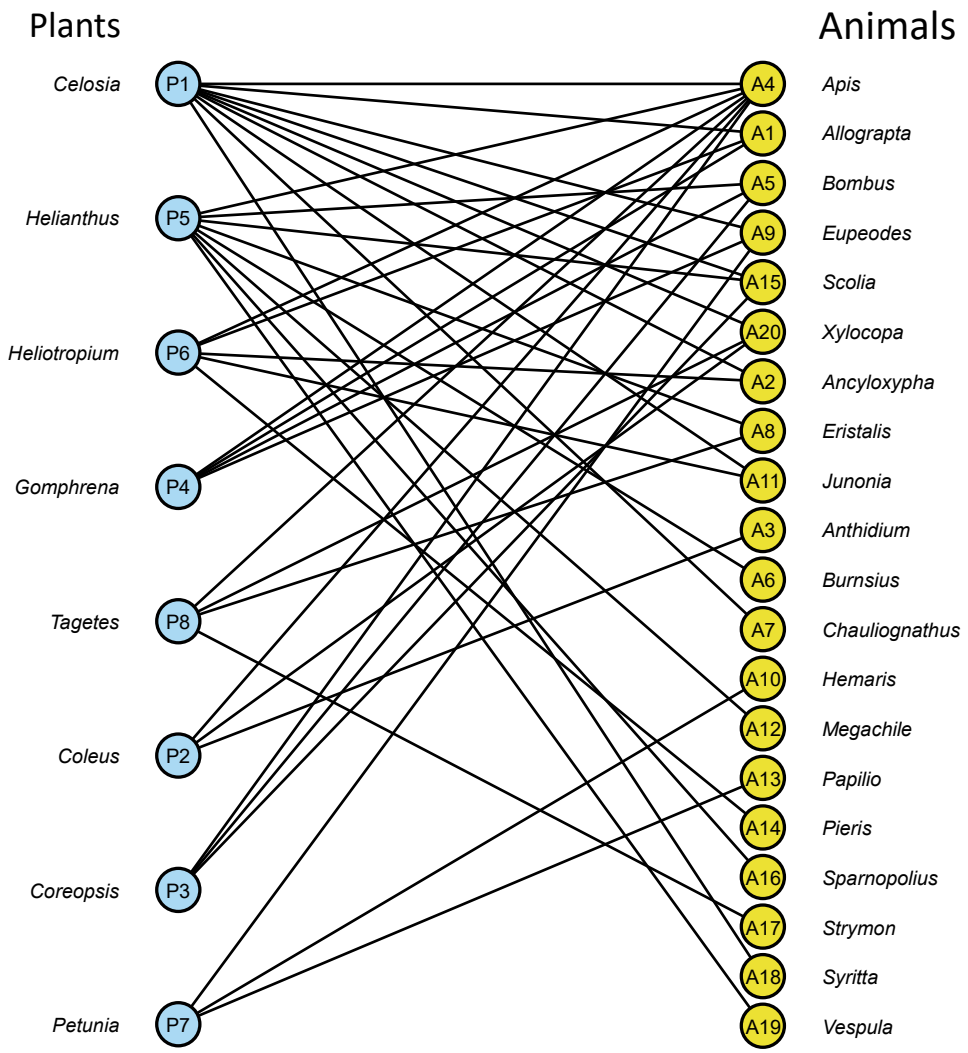
Terminology

- We use flower visitation data to study plant-pollinator interactions, but not all flower visits result in pollination

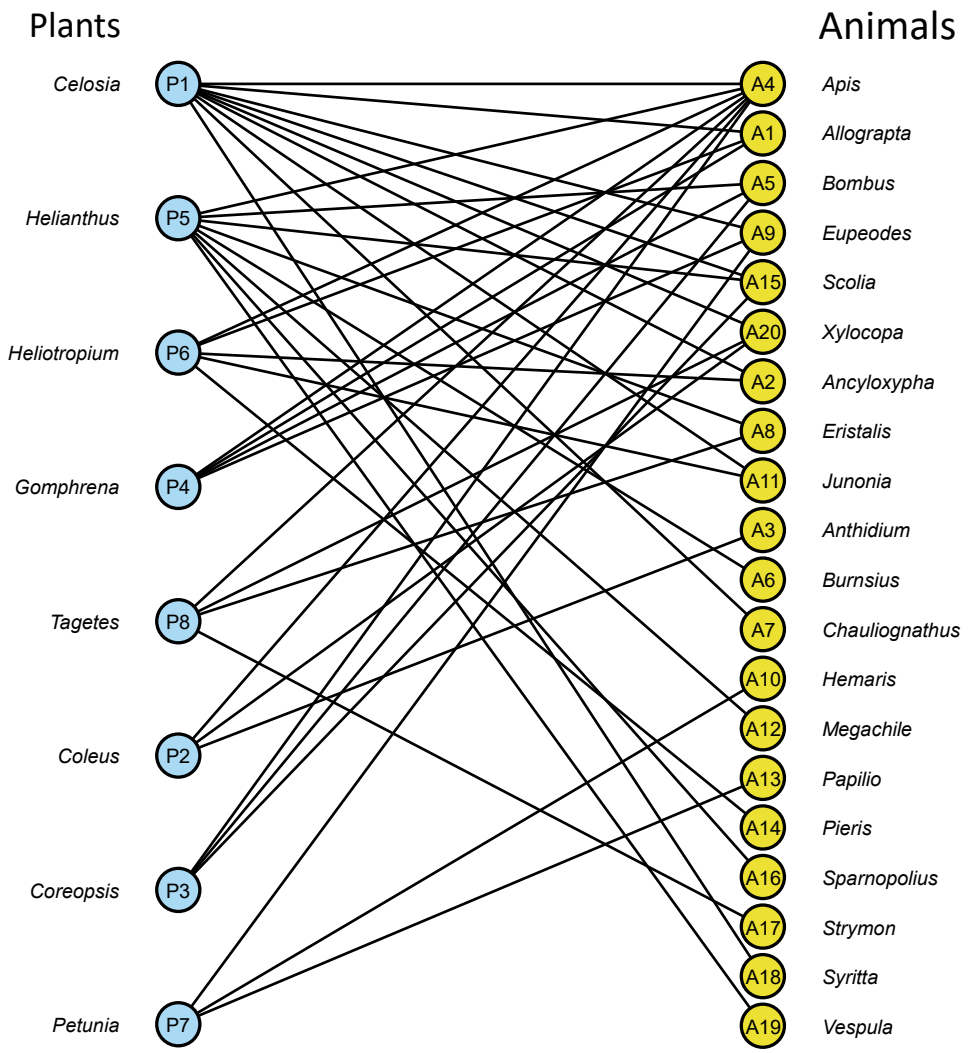
Terminology

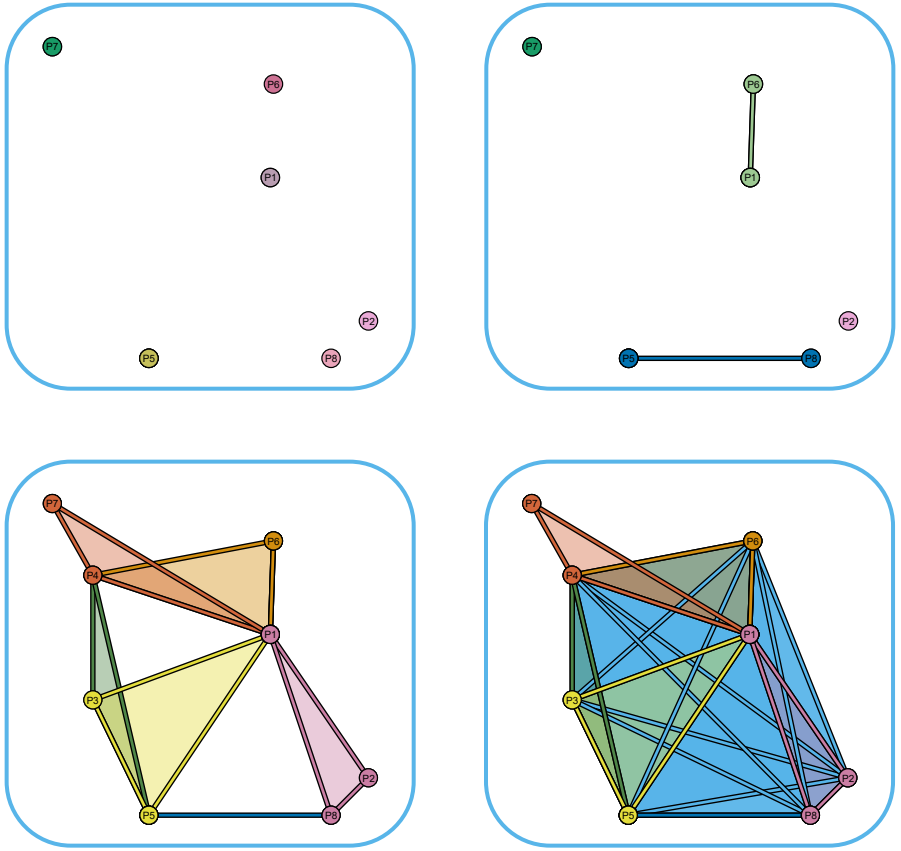
- We use flower visitation data to study plant-pollinator interactions, but not all flower visits result in pollination
- In this talk, we use the following terms interchangeably:
  - Plant / Flowering plant
  - Pollinator / Visitor / Insect / Animal
  - Interaction / Flower visitation

Binary interactions



Binary interactions





M. Kleczynski, D. Guillot, and C. Giusti, *The shape of ecological communities: From pollinators to purple martins*, 2023.

Quantitative flower visitation data

Flower visitation

Plant	Visitor	Number of interactions
Eriophyllum lanatum (common woolly sunflower)	*Apis mellifera* (honey bee)	8
Orthocarpus imbricatus (mountain owl's clover)	*Apis mellifera* (honey bee)	1
Orthocarpus imbricatus (mountain owl's clover)	*Bombus bifarius* (black-notched bumble bee)	2

Quantitative flower visitation data

Flower visitation

Plant	Visitor	Number of interactions
Eriophyllum lanatum (common woolly sunflower)	*Apis mellifera* (honey bee)	8
Orthocarpus imbricatus (mountain owl's clover)	*Apis mellifera* (honey bee)	1
Orthocarpus imbricatus (mountain owl's clover)	*Bombus bifarius* (black-notched bumble bee)	2

Flower survey

Plant	Number of flowers
Eriophyllum lanatum (common woolly sunflower)	30
Orthocarpus imbricatus (mountain owl's clover)	74

Preference index

$$PI_{ij}^{\text{survey}} = \frac{A_{ij}/A_{*j}}{F_i/F} \qquad PI_{ij}^{\text{no survey}} = \frac{A_{ij}/A_{*j}}{A_{i*}/A}$$

A_{ij}	total number of interactions between flower type i and visitor type j
A_{*j}	total number of interactions involving visitor type j
F_i	total number of flowers of type i
A_{i*}	total number of interactions involving flower type i
F	total number of flowers
A	total number of interactions

A. R. Kells, J. M. Holland, and D. Goulson, *The value of uncropped field margins for foraging bumblebees*, Journal of Insect Conservation, 5 (2001), pp. 283–291.

J. Fründ, K. E. Linsenmair, and N. Blüthgen, *Pollinator diversity and specialization in relation to flower diversity*, Oikos, 119 (2010), pp. 1581–1590.

Apis mellifera - *Eriophyllum lanatum* preference

Flower visitation

Plant	Visitor	Number of interactions
Eriophyllum lanatum (common woolly sunflower)	*Apis mellifera* (honey bee)	8
Orthocarpus imbricatus (mountain owl's clover)	*Apis mellifera* (honey bee)	1
Orthocarpus imbricatus (mountain owl's clover)	*Bombus bifarius* (black-notched bumble bee)	2

Flower survey

Plant	Number of flowers
Eriophyllum lanatum (common woolly sunflower)	30
Orthocarpus imbricatus (mountain owl's clover)	74

$$PI^{\text{survey}} = \frac{8/9}{30/104} = 3.1$$

$$PI^{\text{no survey}} = \frac{8/9}{8/11} = 1.2$$

Simplicial complexes

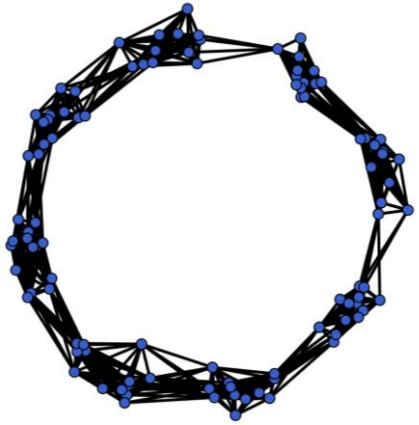
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

Graphs

Graphs

A graph is an ordered pair (V, E) such that V and E are sets and each element of E is a subset of V with two elements.

We call the elements of V vertices and the elements of E edges.

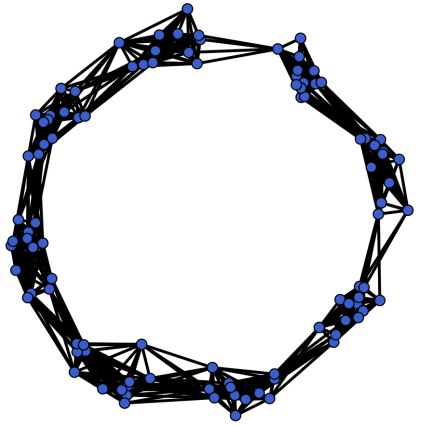


Graphs and simplicial complexes

Graphs

A graph is an ordered pair (V, E) such that V and E are sets and each element of E is a subset of V with two elements.

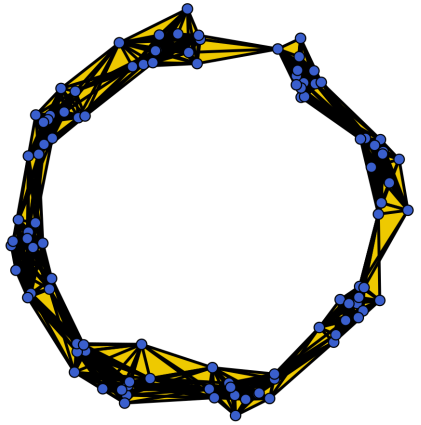
We call the elements of V vertices and the elements of E edges.



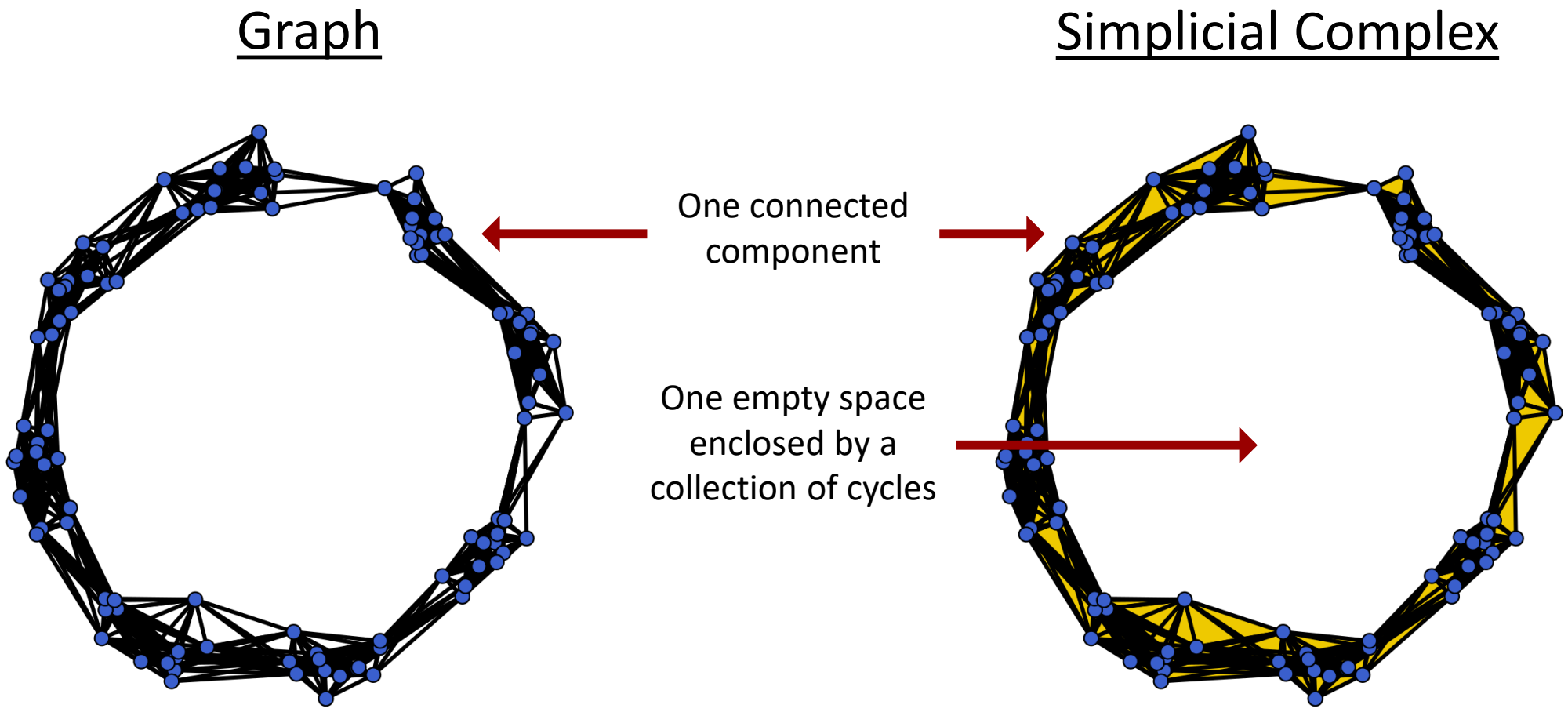
Simplicial complexes

Given a discrete set X, a simplicial complex K_X is a collection of finite subsets of X such that if $\tau \in K_X$ and $\sigma \subset \tau$ then $\sigma \in K_X$.

We call the elements of X vertices and the elements of K_X simplices. A simplex σ with $|\sigma| = k + 1$ is a k-simplex.



Structural features



Characterizing simplicial complexes

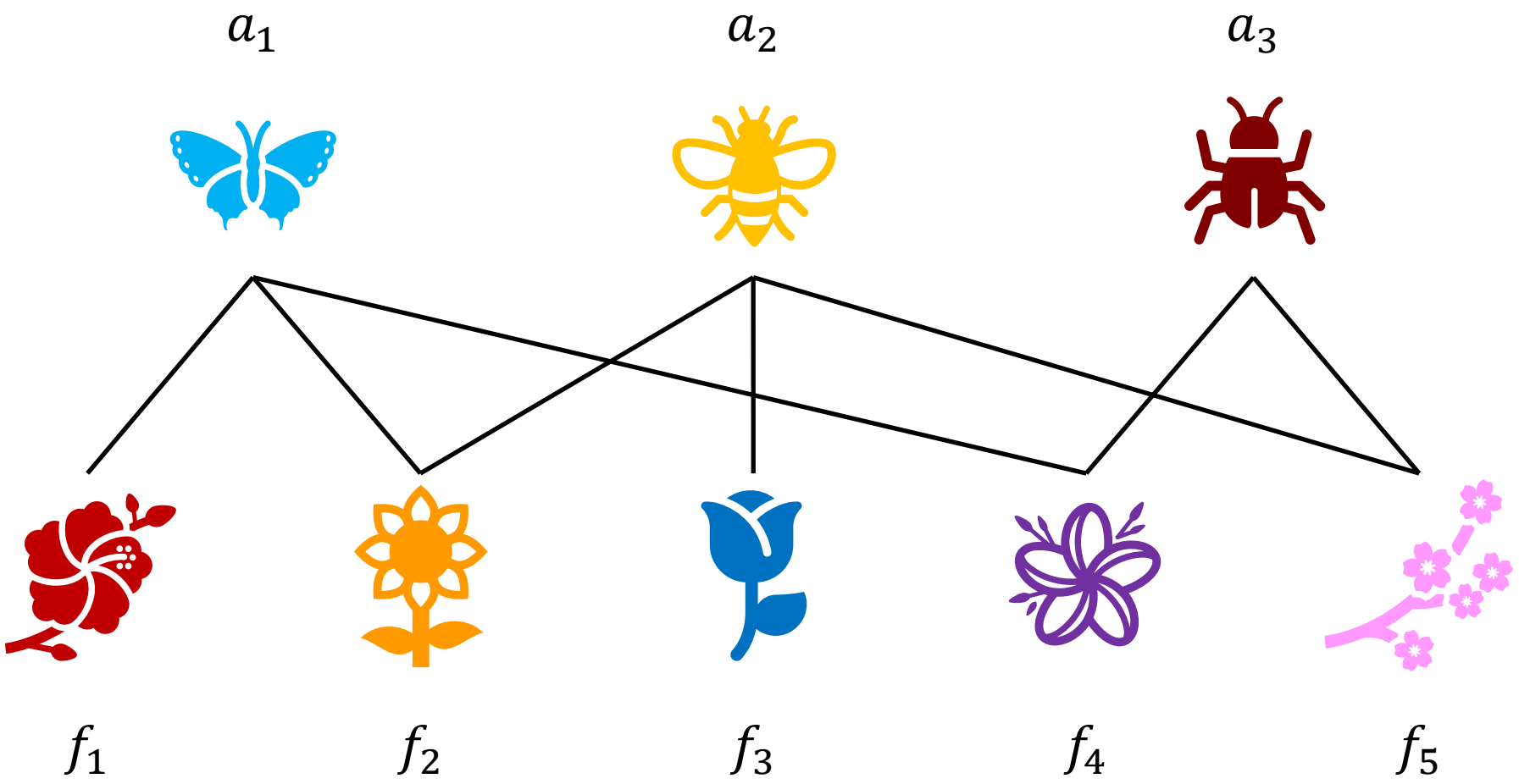
Let n be a nonnegative integer. A simplicial complex generates a homology group H_n. For coefficients over a field, H_n is a quotient vector space. Its elements are equivalence classes of cycles.

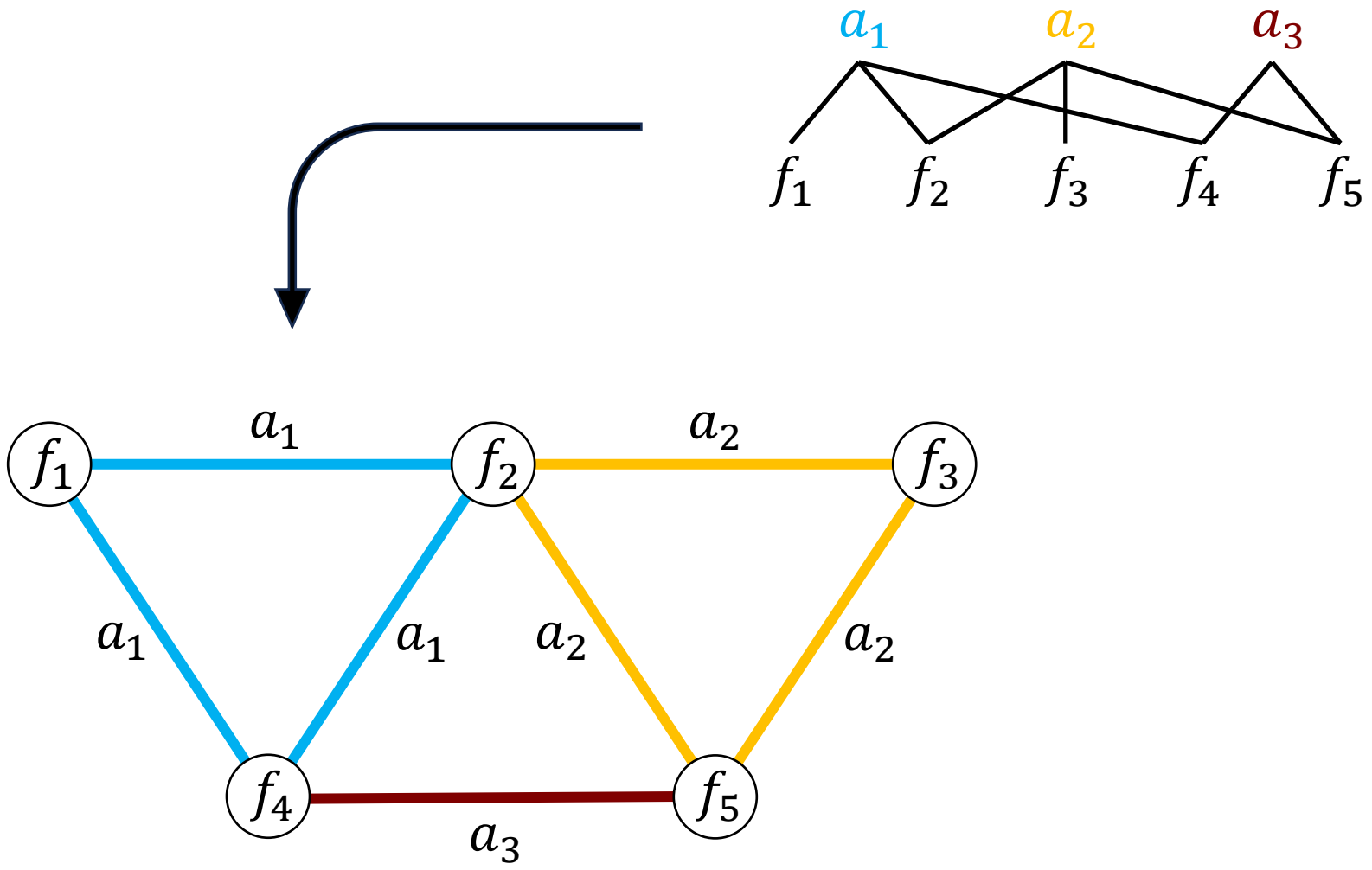
- $\dim(H_0)$ is the number of connected components.
- $\dim(H_1)$ is the number of empty spaces which are bounded by 1-simplices (like windows).
- $\dim(H_2)$ is the number of empty spaces which are bounded by 2-simplices (like bubbles).

Simplicial complexes from plant-pollinator data

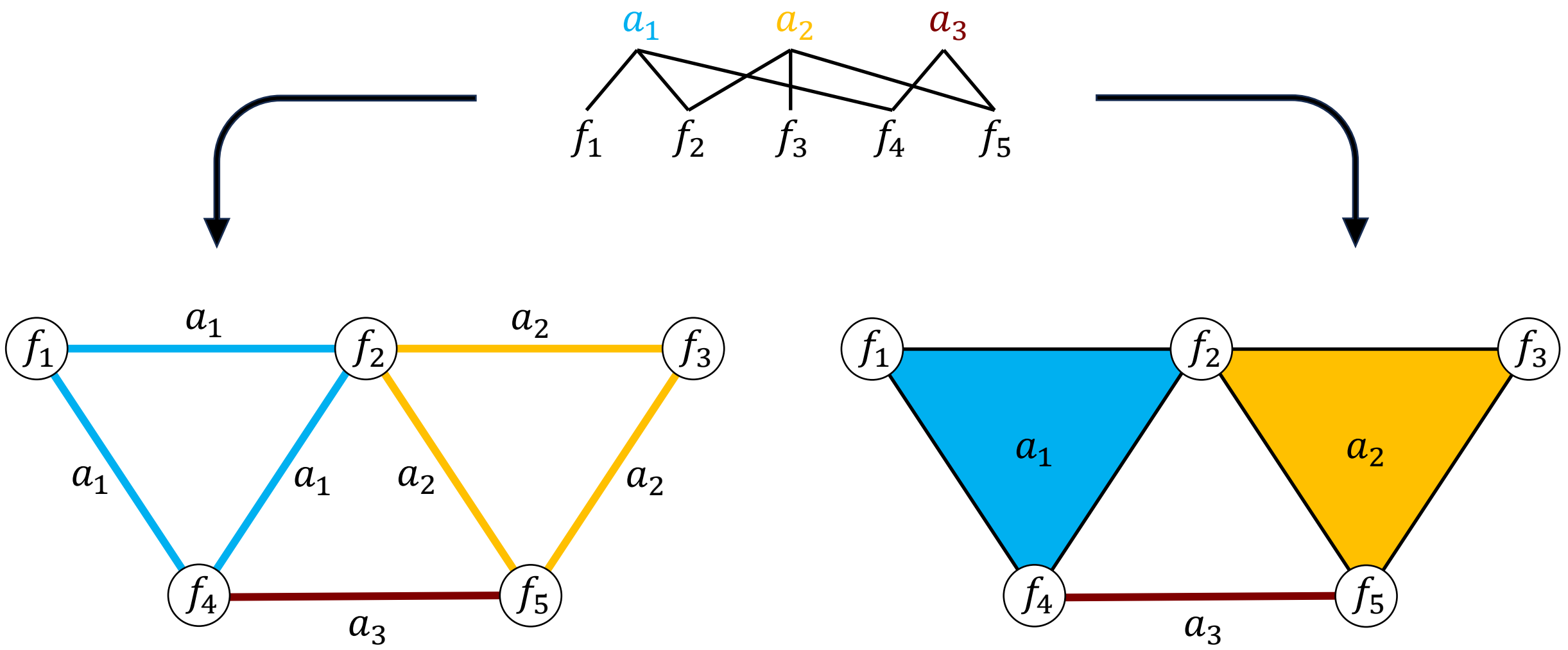
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

Simple flower visitation community





B. Peleg, A. Shmida, and S. Ellner, *Foraging graphs: Constraint rules on matching between bees and flowers in a two-sided pollination market*, Journal of Theoretical Biology, 157(1992), pp. 191–201.



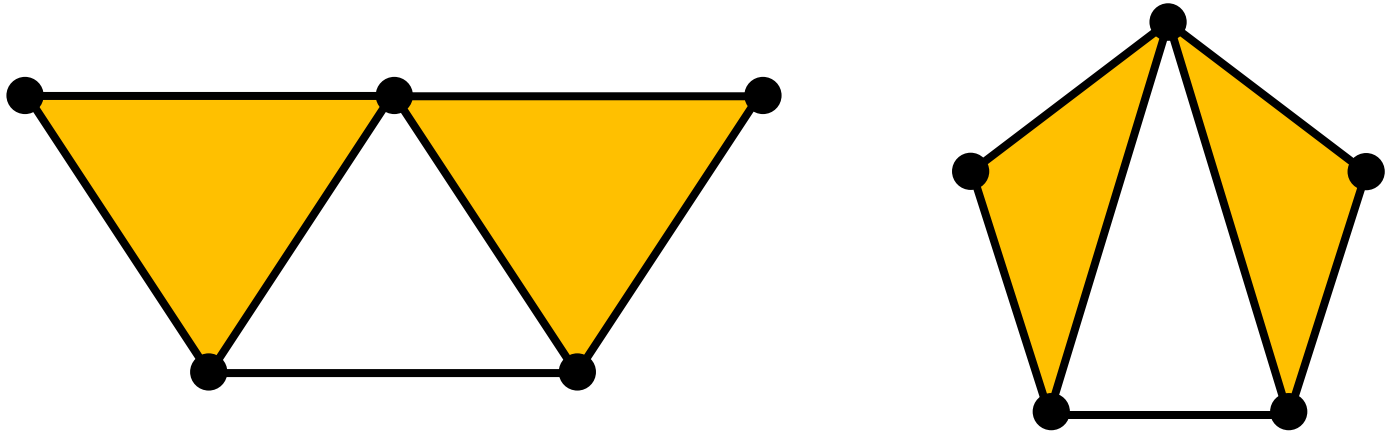
B. Peleg, A. Shmida, and S. Ellner, *Foraging graphs: Constraint rules on matching between bees and flowers in a two-sided pollination market*, Journal of Theoretical Biology, 157(1992), pp. 191–201.

C. H. Dowker, *Homology groups of relations*, Annals of Mathematics, 56 (1952), pp. 84–95.

G. Sugihara, *Niche Hierarchy: Structure, Organization, and Assembly in Natural Systems*, J. Ross Publishing, 2017.

Vertex coordinates

- The vertices of these simplicial complexes do not have defined coordinates. For example, these are the same simplicial complex.



- We may choose vertex coordinates for visualization, but we do not use them for computations.

Edge weights

Scaled preference

$$\widetilde{PI}_{ij} = \frac{PI_{ij}}{\max_{m,n} PI_{mn}}$$

	a_1	a_2	a_3
f_1	0.7	0.0	0.0
f_2	0.2	0.8	0.0
f_3	0.0	0.3	0.0
f_4	0.3	0.0	0.1
f_5	0.0	0.2	1.0

Cross-dissimilarity =
1 – Scaled preference

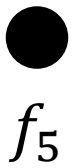
	a_1	a_2	a_3
f_1	0.3	1.0	1.0
f_2	0.8	0.2	1.0
f_3	1.0	0.7	1.0
f_4	0.7	1.0	0.9
f_5	1.0	0.8	0.0

Sequence of simplicial complexes

$$\varepsilon = 0.0$$

Cross-dissimilarity

	a_1	a_2	a_3
f_1	0.3	1.0	1.0
f_2	0.8	0.2	1.0
f_3	1.0	0.7	1.0
f_4	0.7	1.0	0.9
f_5	1.0	0.8	0.0



Sequence of simplicial complexes

$\varepsilon = 0.2$

Cross-dissimilarity

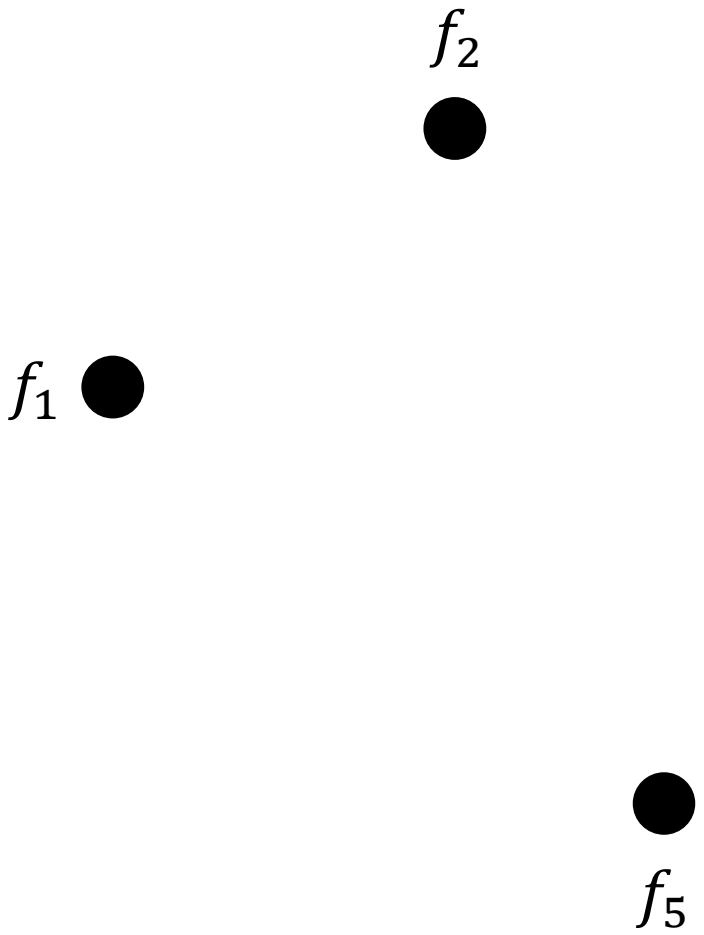
	a_1	a_2	a_3
f_1	0.3	1.0	1.0
f_2	0.8	0.2	1.0
f_3	1.0	0.7	1.0
f_4	0.7	1.0	0.9
f_5	1.0	0.8	0.0

Sequence of simplicial complexes

$\varepsilon = 0.3$

Cross-dissimilarity

	a_1	a_2	a_3
f_1	0.3	1.0	1.0
f_2	0.8	0.2	1.0
f_3	1.0	0.7	1.0
f_4	0.7	1.0	0.9
f_5	1.0	0.8	0.0

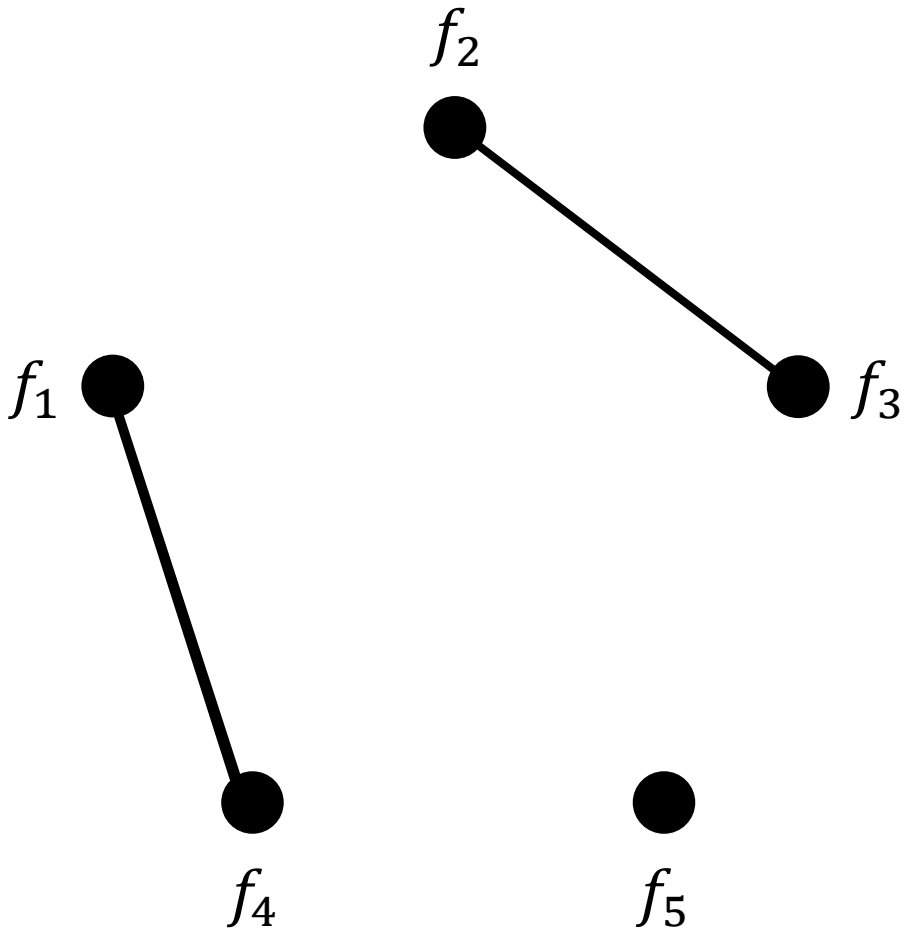


Sequence of simplicial complexes

$\varepsilon = 0.7$

Cross-dissimilarity

	a_1	a_2	a_3
f_1	0.3	1.0	1.0
f_2	0.8	0.2	1.0
f_3	1.0	0.7	1.0
f_4	0.7	1.0	0.9
f_5	1.0	0.8	0.0

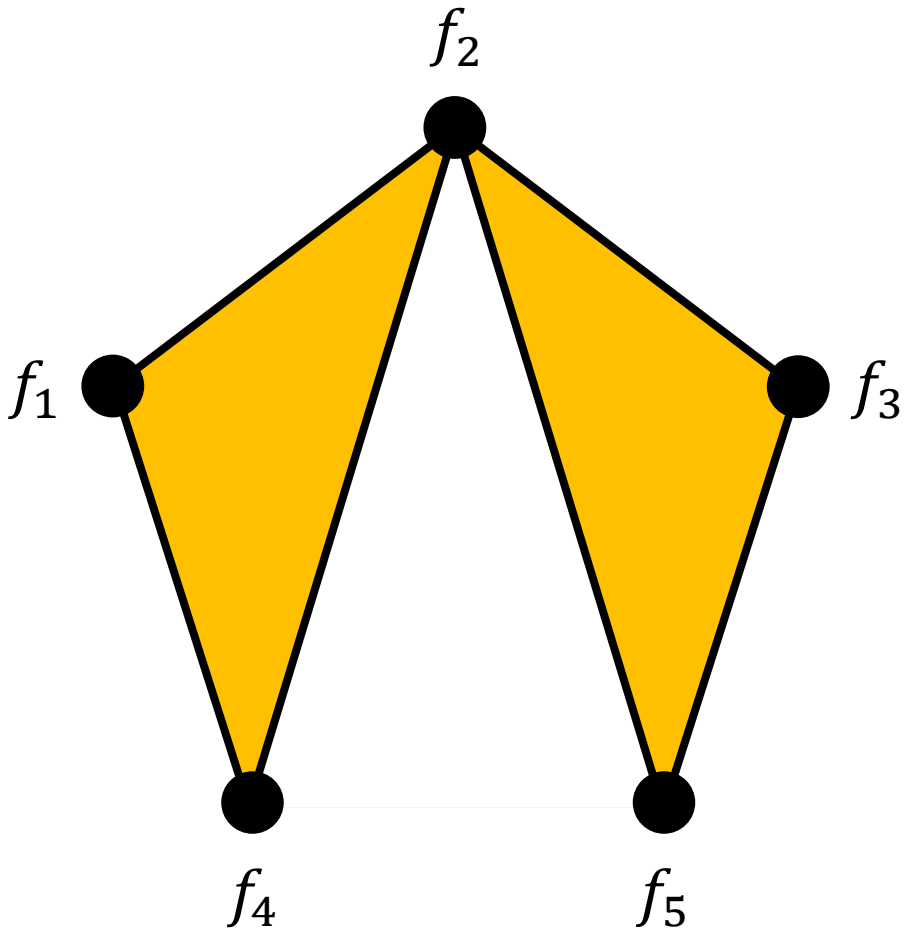


Sequence of simplicial complexes

$\varepsilon = 0.8$

Cross-dissimilarity

	a_1	a_2	a_3
f_1	0.3	1.0	1.0
f_2	0.8	0.2	1.0
f_3	1.0	0.7	1.0
f_4	0.7	1.0	0.9
f_5	1.0	0.8	0.0

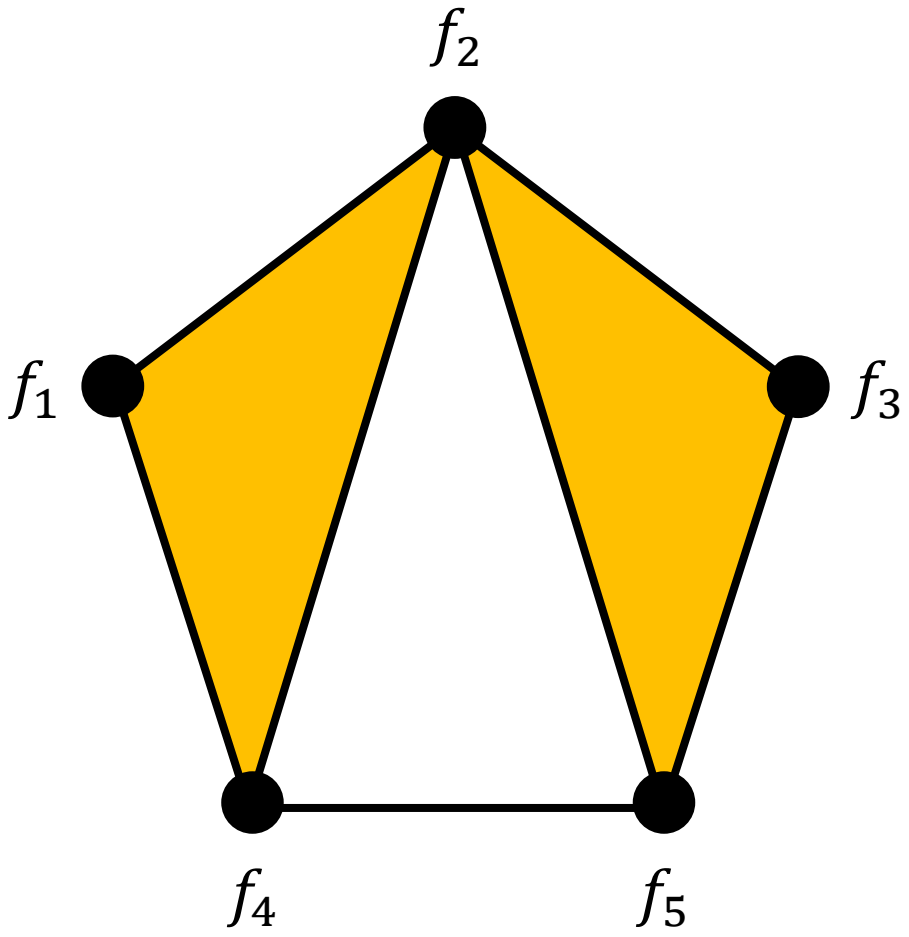


Sequence of simplicial complexes

$\varepsilon = 0.9$

Cross-dissimilarity

	a_1	a_2	a_3
f_1	0.3	1.0	1.0
f_2	0.8	0.2	1.0
f_3	1.0	0.7	1.0
f_4	0.7	1.0	0.9
f_5	1.0	0.8	0.0

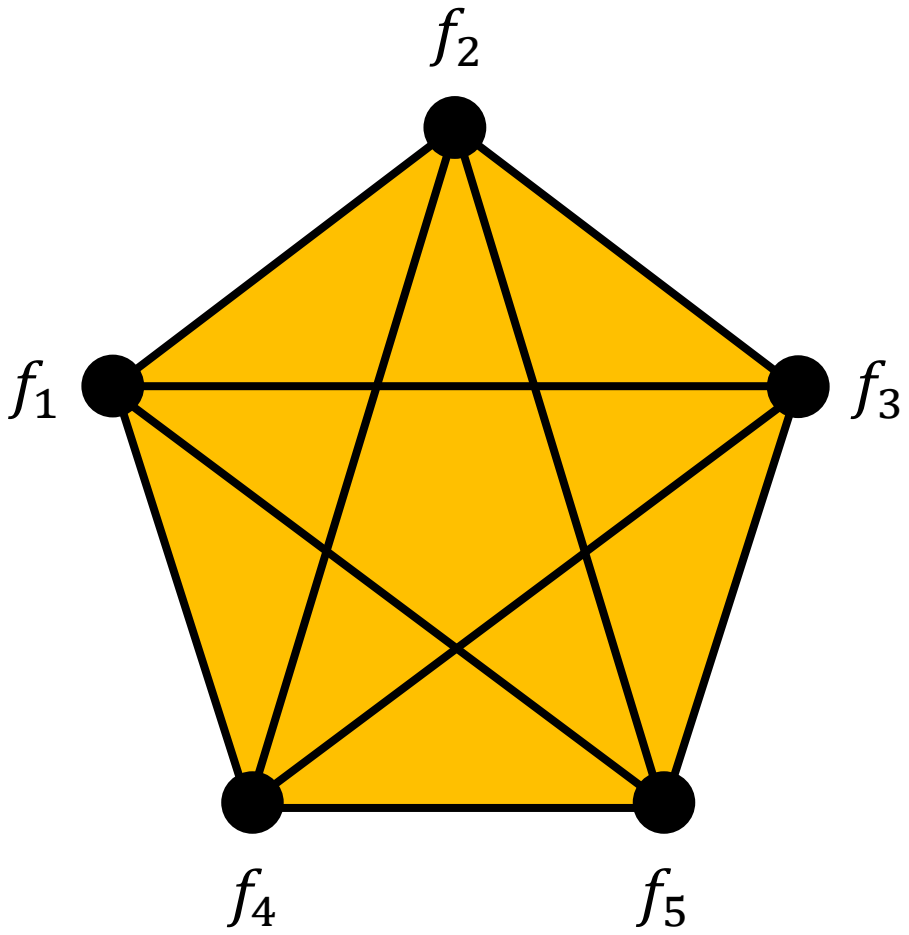


Sequence of simplicial complexes

$\varepsilon = 1.0$

Cross-dissimilarity

	a_1	a_2	a_3
f_1	0.3	1.0	1.0
f_2	0.8	0.2	1.0
f_3	1.0	0.7	1.0
f_4	0.7	1.0	0.9
f_5	1.0	0.8	0.0

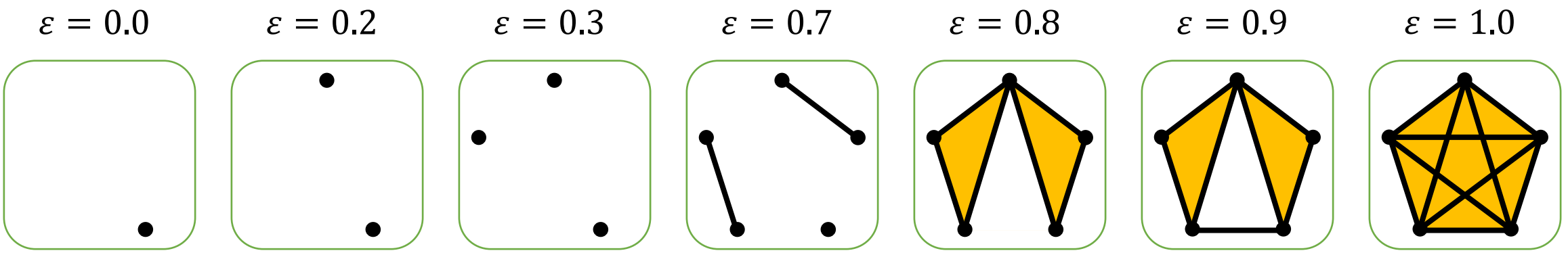


Topological summaries and distances

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

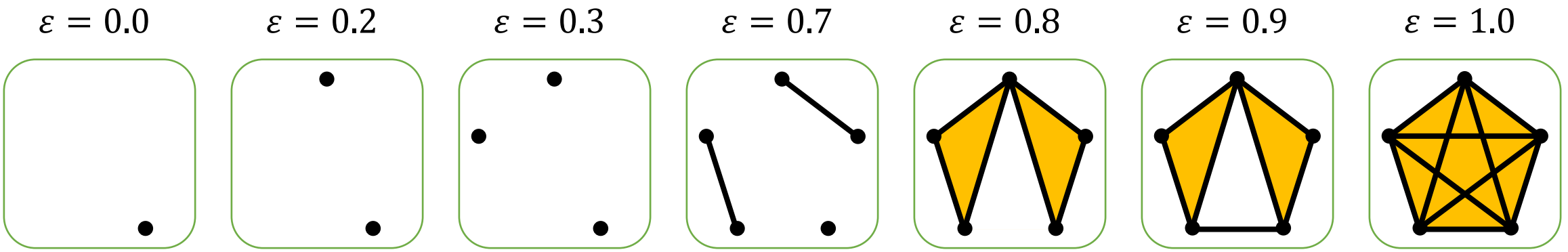
Dimension 1 barcode

Sequence of Simplicial Complexes



Dimension 1 barcode

Sequence of Simplicial Complexes

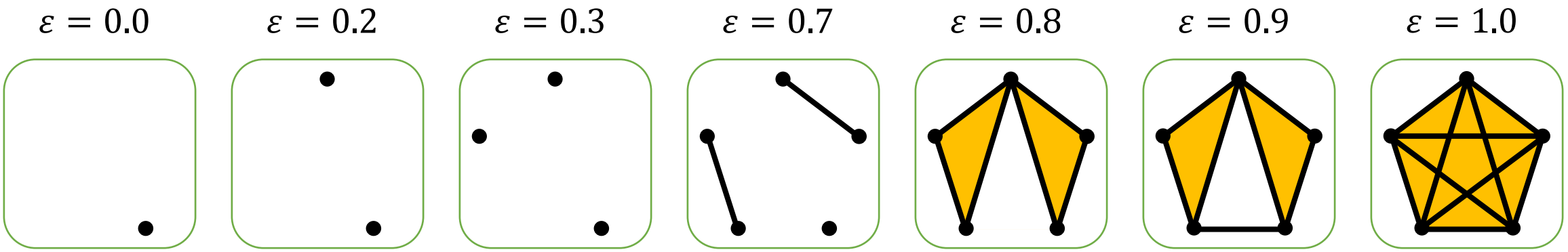


Feature
Birth
$b = 0.9$

Feature
Death
$d = 1.0$

Dimension 1 barcode

Sequence of Simplicial Complexes



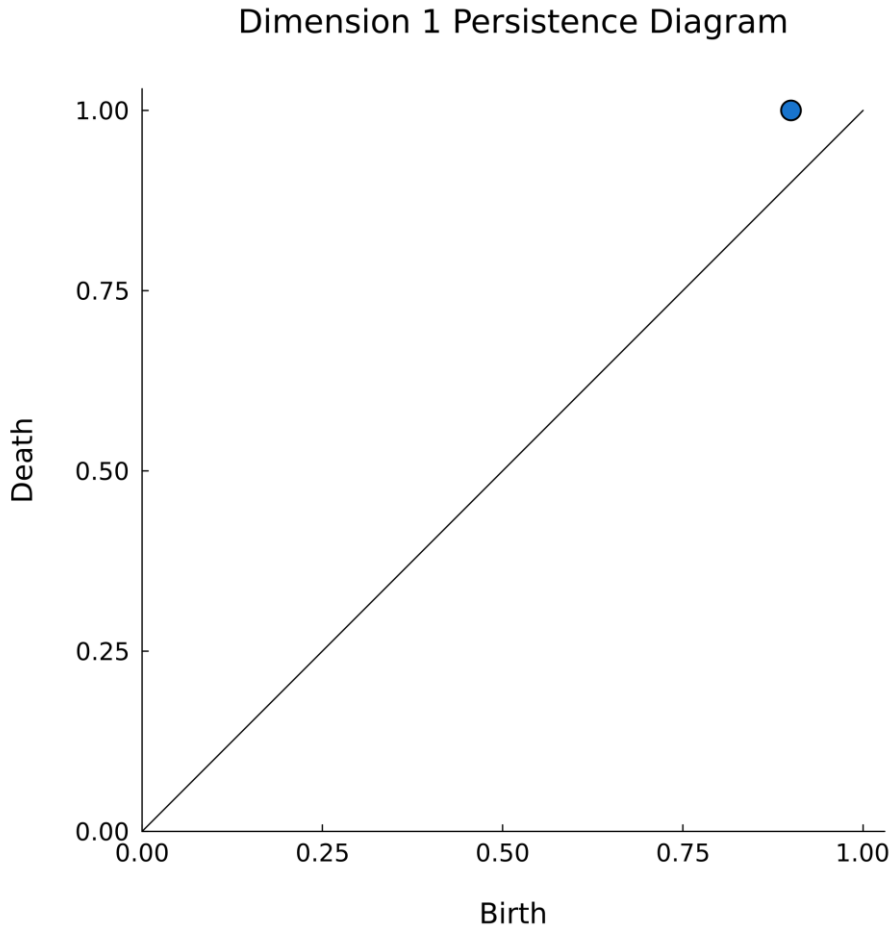
- Summarize each topological feature as an interval $[b, d)$, commonly called a bar.
- Repeat this process for each topological feature in a given dimension to obtain a multiset of intervals called a barcode.
- The dimension 1 barcode for this example is $\{[0.9, 1.0)\}$.

Persistence diagrams

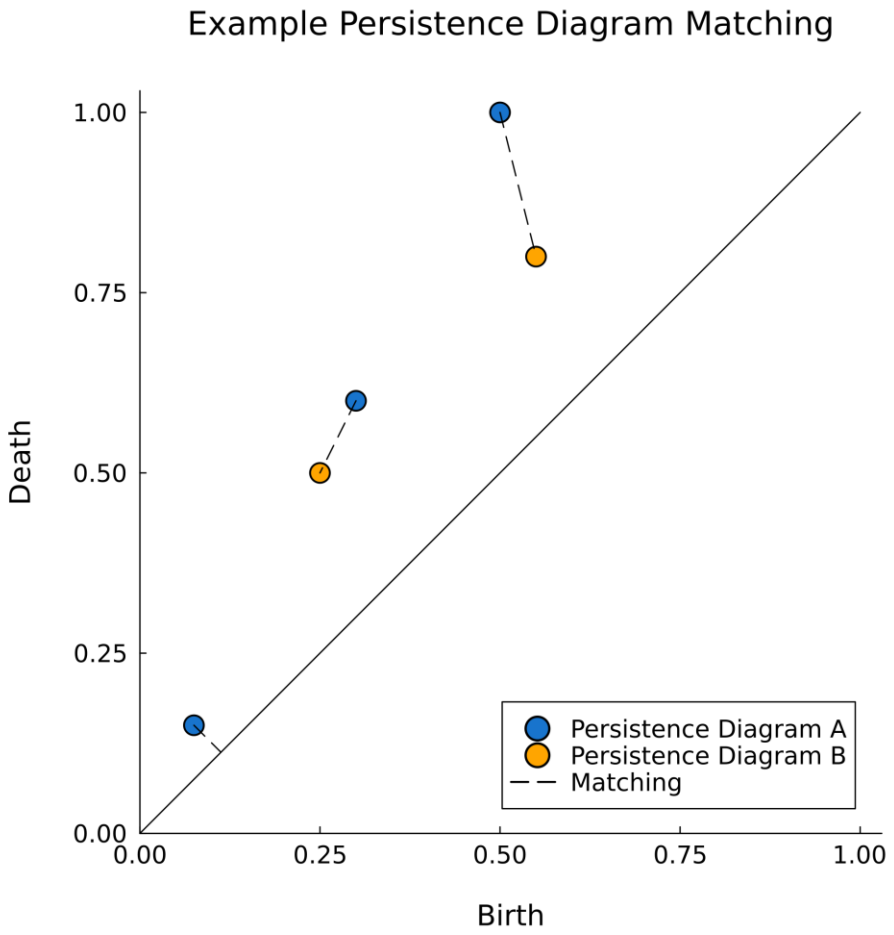
A persistence diagram is a multiset with elements

- (b, d) for each interval $[b, d)$ in the barcode (with the same multiplicity as in the barcode)
- (x, x) for all $x \in \mathbb{R}$ (with countably infinite multiplicity)

* For the current analysis, barcodes only contain finite intervals. Our definition of the persistence diagram reflects this.



(p, q)-Wasserstein distance



Let S and T be multisets of points in $\mathbb{R}^2$. Let $\text{bij}(S, T)$ denote the set of all bijections from S to T.

The (p, q)-Wasserstein distance between S and T is

$$\inf_{f \in \text{bij}(S,T)} \left(\sum_{s \in S} \|s - f(s)\|_q^p \right)^{1/p}$$

Methods

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

Data processing

- We only consider insect visitors and data from the years 2016, 2017, and 2018.

Data processing

- We only consider insect visitors and data from the years 2016, 2017, and 2018.
- During these years, data collection occurs in three meadow complexes.

Data processing

- We only consider insect visitors and data from the years 2016, 2017, and 2018.
- During these years, data collection occurs in three meadow complexes.
- Each meadow complex contains four meadows, and there are five observation periods each year (about once a week).

Data processing

- We only consider insect visitors and data from the years 2016, 2017, and 2018.
- During these years, data collection occurs in three meadow complexes.
- Each meadow complex contains four meadows, and there are five observation periods each year (about once a week).
- We compute preferences for each meadow-year-watch.

Data processing

- We only consider insect visitors and data from the years 2016, 2017, and 2018.
- During these years, data collection occurs in three meadow complexes.
- Each meadow complex contains four meadows, and there are five observation periods each year (about once a week).
- We compute preferences for each meadow-year-watch.
- We compute persistent homology for each complex-year. We use median preferences (out of all meadow-year-watches) for the main analysis and randomly chosen preferences (out of all meadow-year-watches) for bootstrap samples.

Bootstrap samples

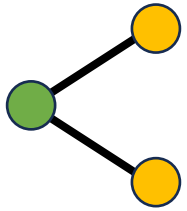
Simplified example data

Plant	Visitor	Total number of interactions	Weight
Common woolly sunflower (CWS)	Honey bee	4	0.8
Common woolly sunflower (CWS)	California bumble bee	1	0.2

Weighted subsample with replacement

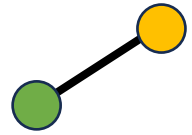
<u>Sample 1:</u>
CWS - Honey bee
CWS - Honey bee
CWS - Honey bee
CWS - California bumble bee
CWS - Honey bee

Plants Visitors



<u>Sample 2:</u>
CWS - Honey bee
CWS - Honey bee
CWS - Honey bee
CWS - Honey bee
CWS - Honey bee

Plants Visitors



Topological rankings of plant types

- Compute the cross-dissimilarities of all plant-pollinator pairs

Topological rankings of plant types

- Compute the cross-dissimilarities of all plant-pollinator pairs
- Compute the dimension 1 persistence diagram

Topological rankings of plant types

- Compute the cross-dissimilarities of all plant-pollinator pairs
- Compute the dimension 1 persistence diagram
- For each plant type
  - Remove the plant from the cross-dissimilarity matrix (do not recompute the remaining cross-dissimilarities)
  - Compute the dimension 1 persistence diagram
  - Take the Wasserstein distance between the original persistence diagram and the persistence diagram obtained after removing the plant

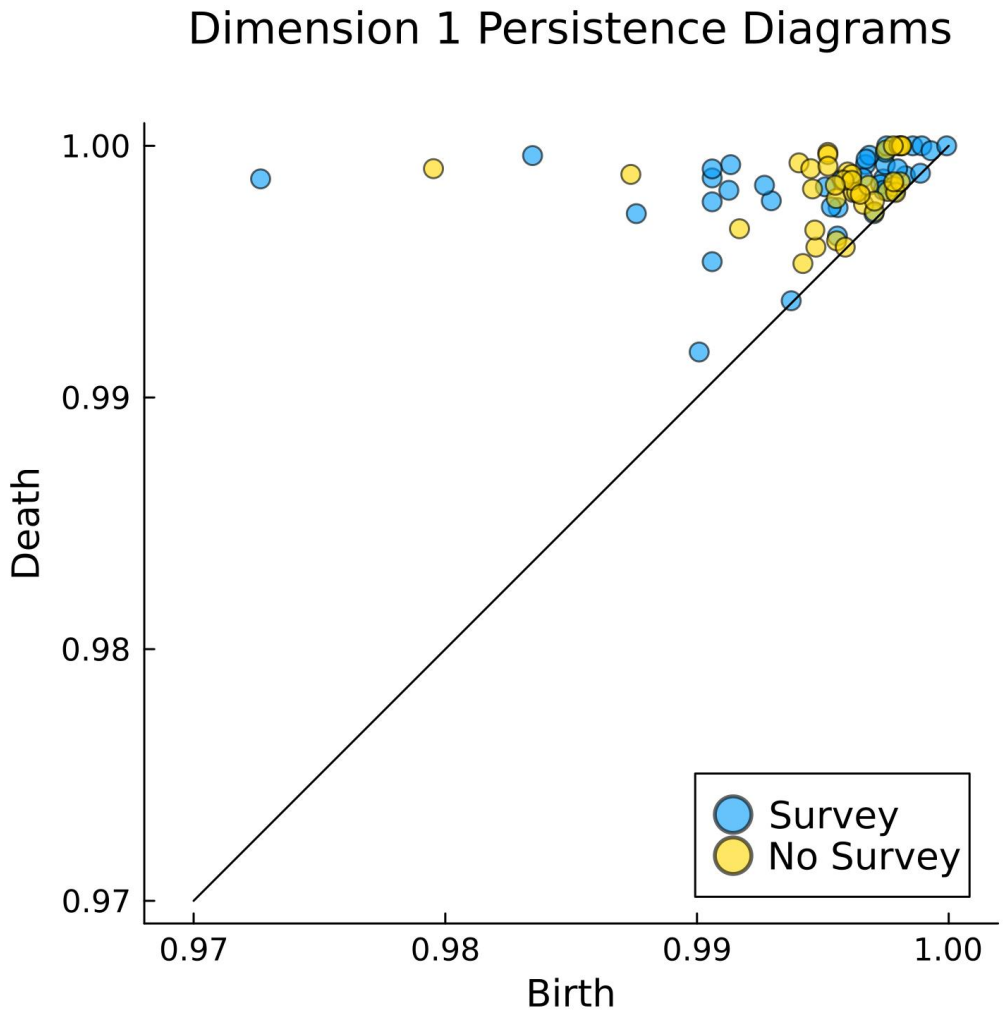
Topological rankings of plant types

- Compute the cross-dissimilarities of all plant-pollinator pairs
- Compute the dimension 1 persistence diagram
- For each plant type
  - Remove the plant from the cross-dissimilarity matrix (do not recompute the remaining cross-dissimilarities)
  - Compute the dimension 1 persistence diagram
  - Take the Wasserstein distance between the original persistence diagram and the persistence diagram obtained after removing the plant
- Plants with higher Wasserstein distances receive higher rankings

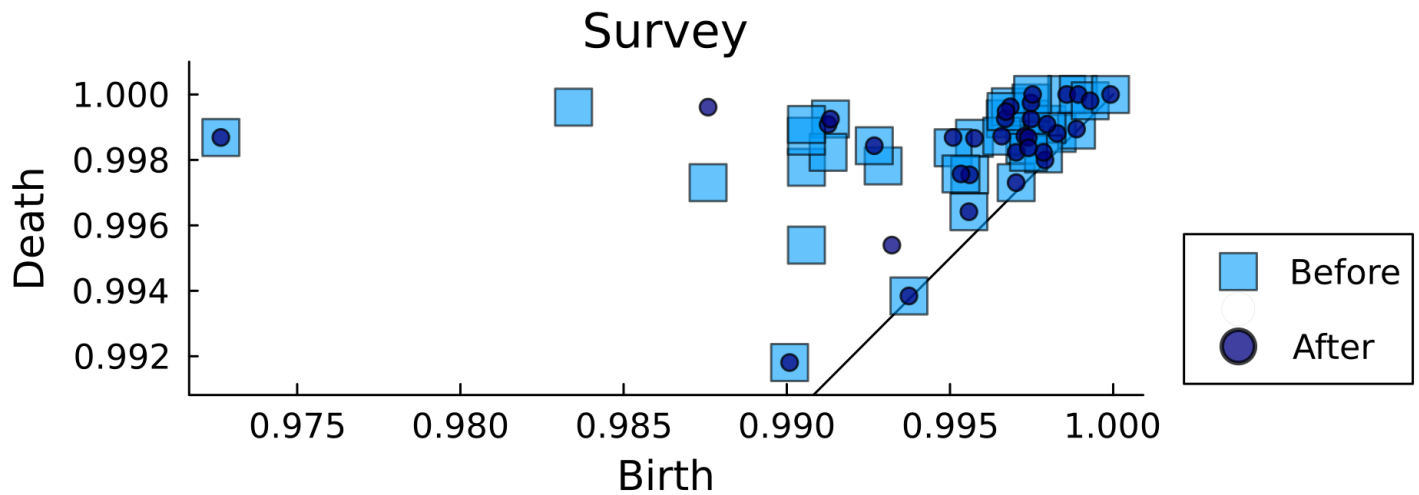
Preliminary results

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE

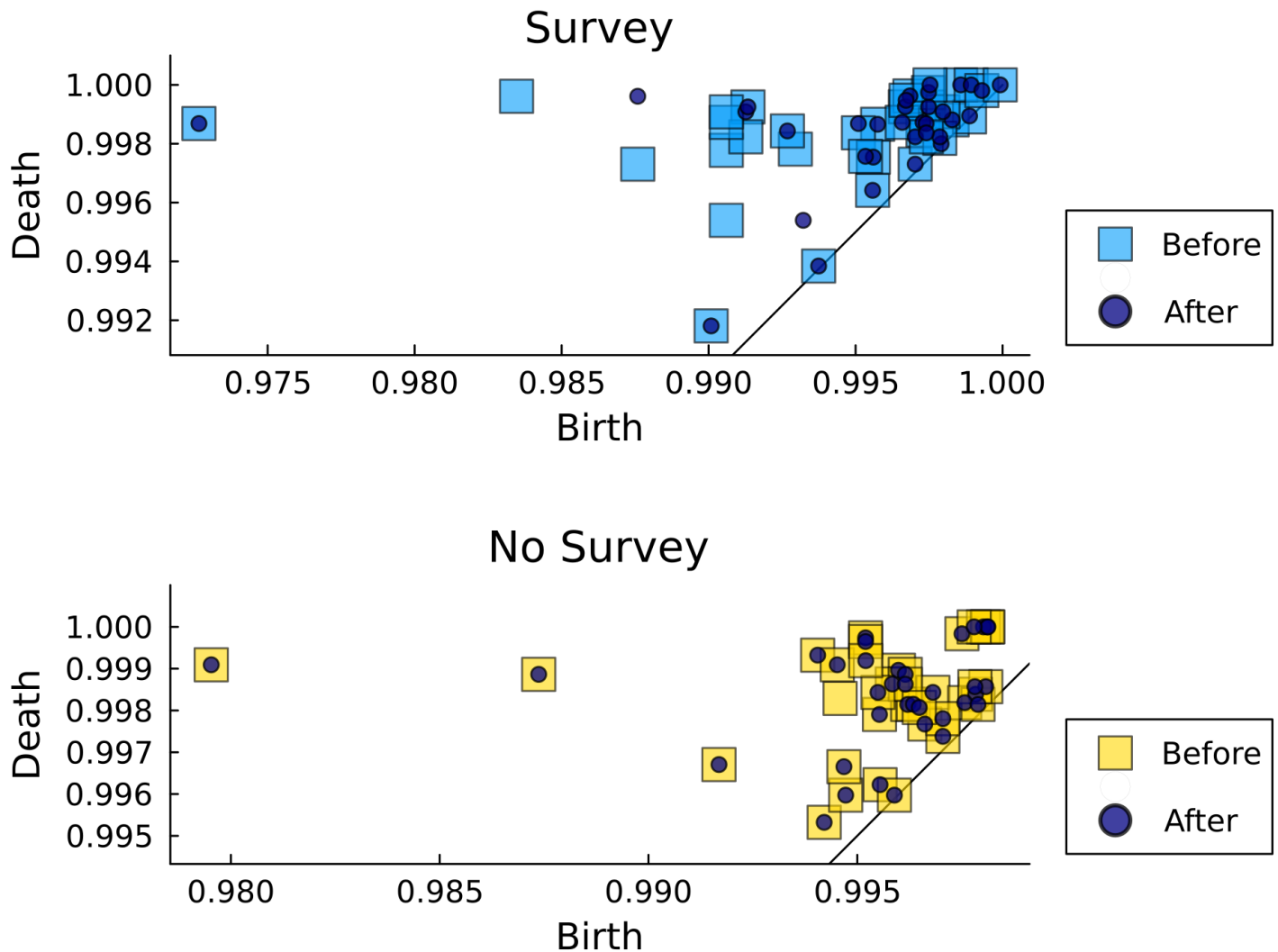
Carpenter meadow complex, 2017



Dimension 1 persistence diagrams before and after removing *Angelica arguta*



Dimension 1 persistence diagrams before and after removing *Angelica arguta*



Topological rankings

Survey

Plant name	Wasserstein distance (p = q = 1)	Topological ranking
Angelica arguta	0.0427	1
Fragaria virginiana	0.0376	2
Erigeron aliceae	0.0338	3
Erysimum capitatum	0.0227	4
Ligusticum grayi	0.0178	5

Topological rankings

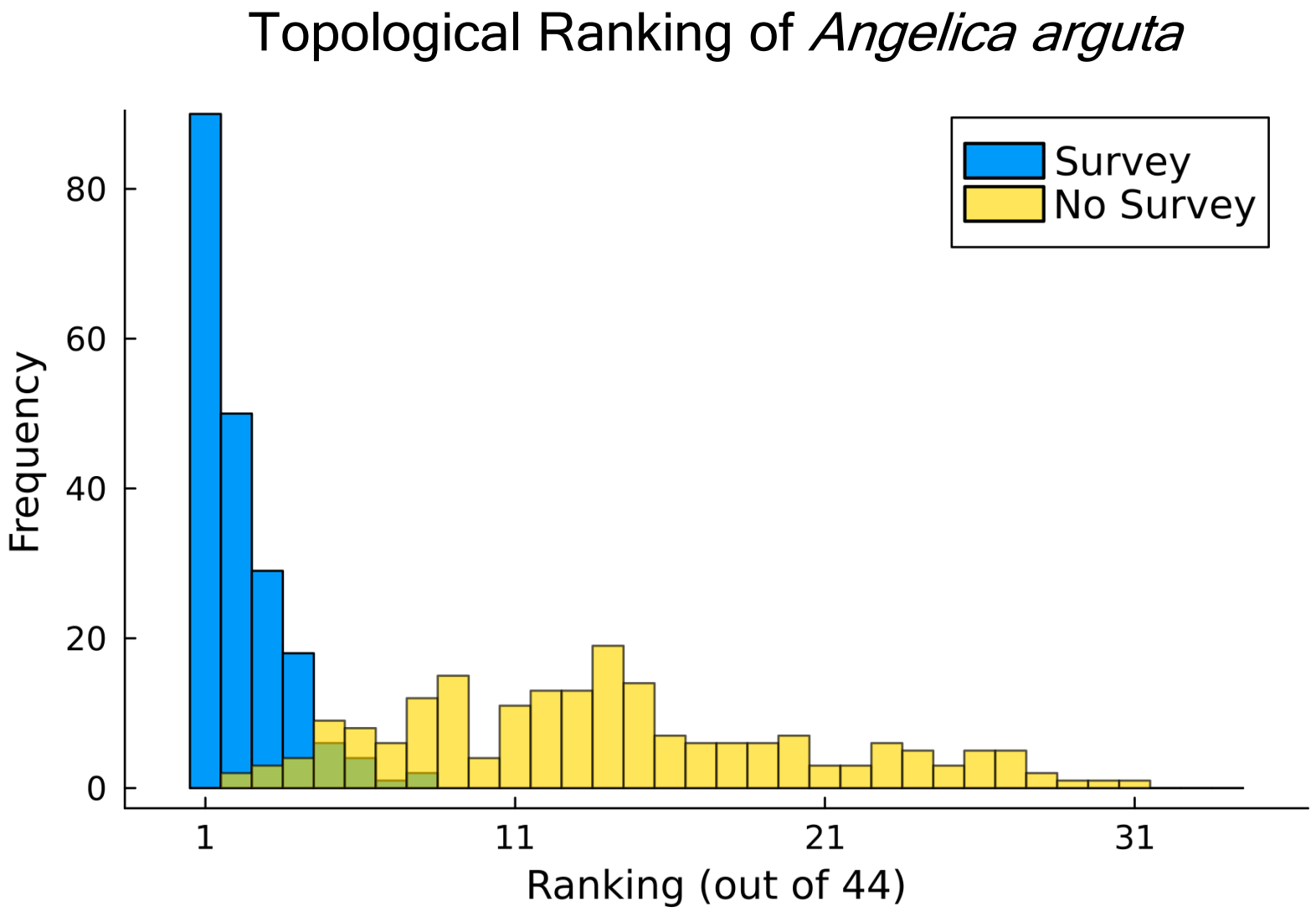
Survey

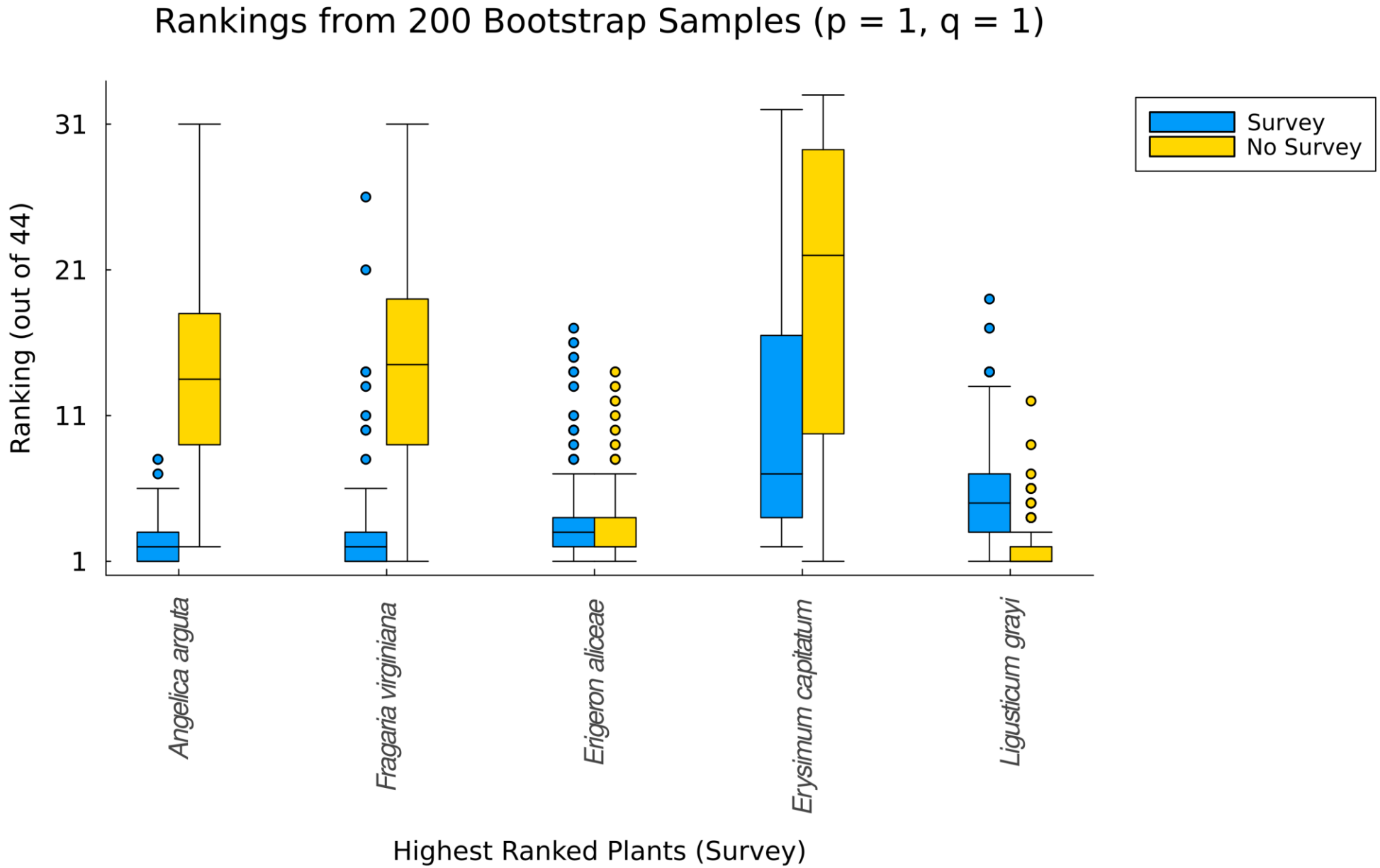
Plant name	Wasserstein distance (p = q = 1)	Topological ranking
Angelica arguta	0.0427	1
Fragaria virginiana	0.0376	2
Erigeron aliceae	0.0338	3
Erysimum capitatum	0.0227	4
Ligusticum grayi	0.0178	5

No survey

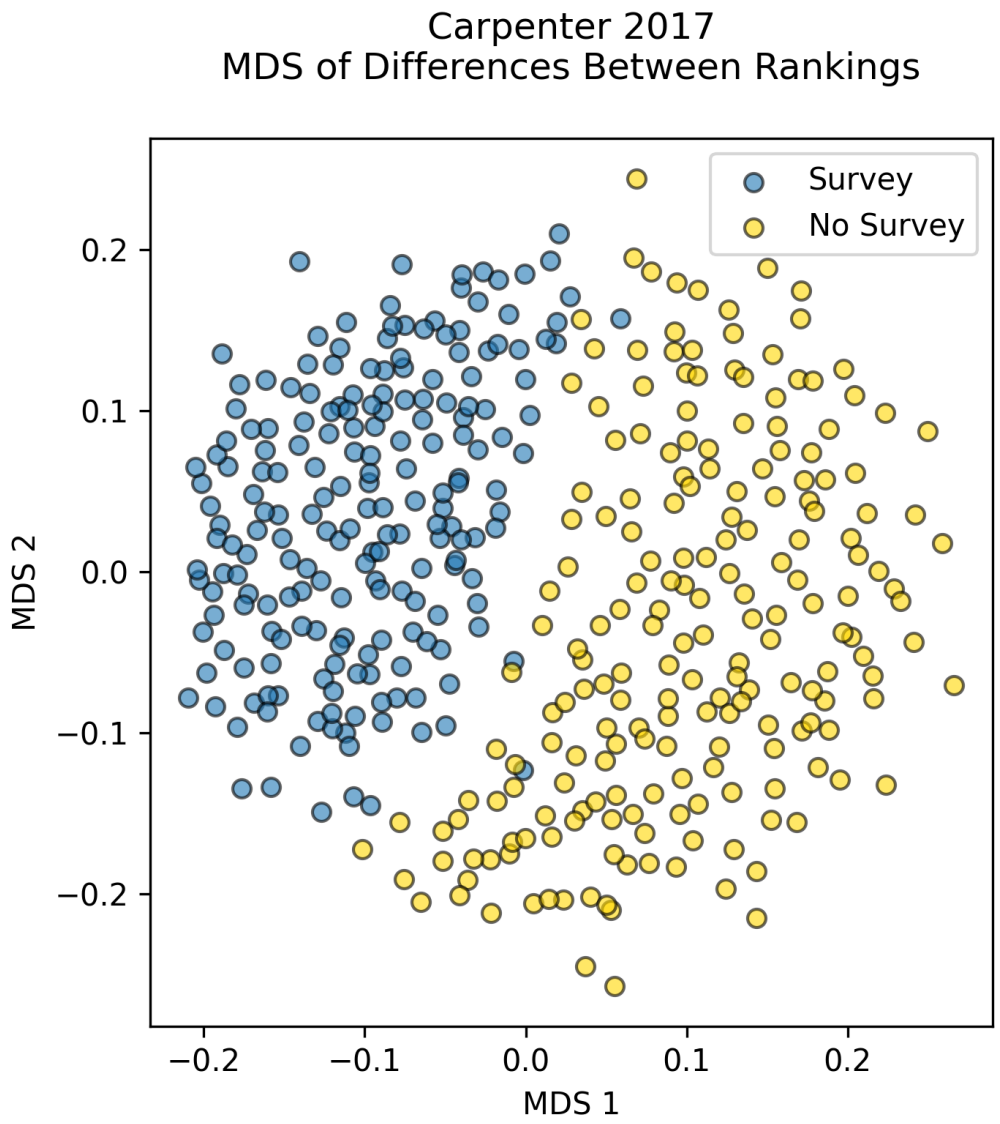
Plant name	Wasserstein distance (p = q = 1)	Topological ranking
Erigeron aliceae	0.0268	1
Erigeron foliosus	0.0214	2
Ligusticum grayi	0.0178	3
Claytonia lanceolata	0.0167	4
Erysimum capitatum	0.0164	5

Rankings from 200 bootstrap samples (p = q = 1)

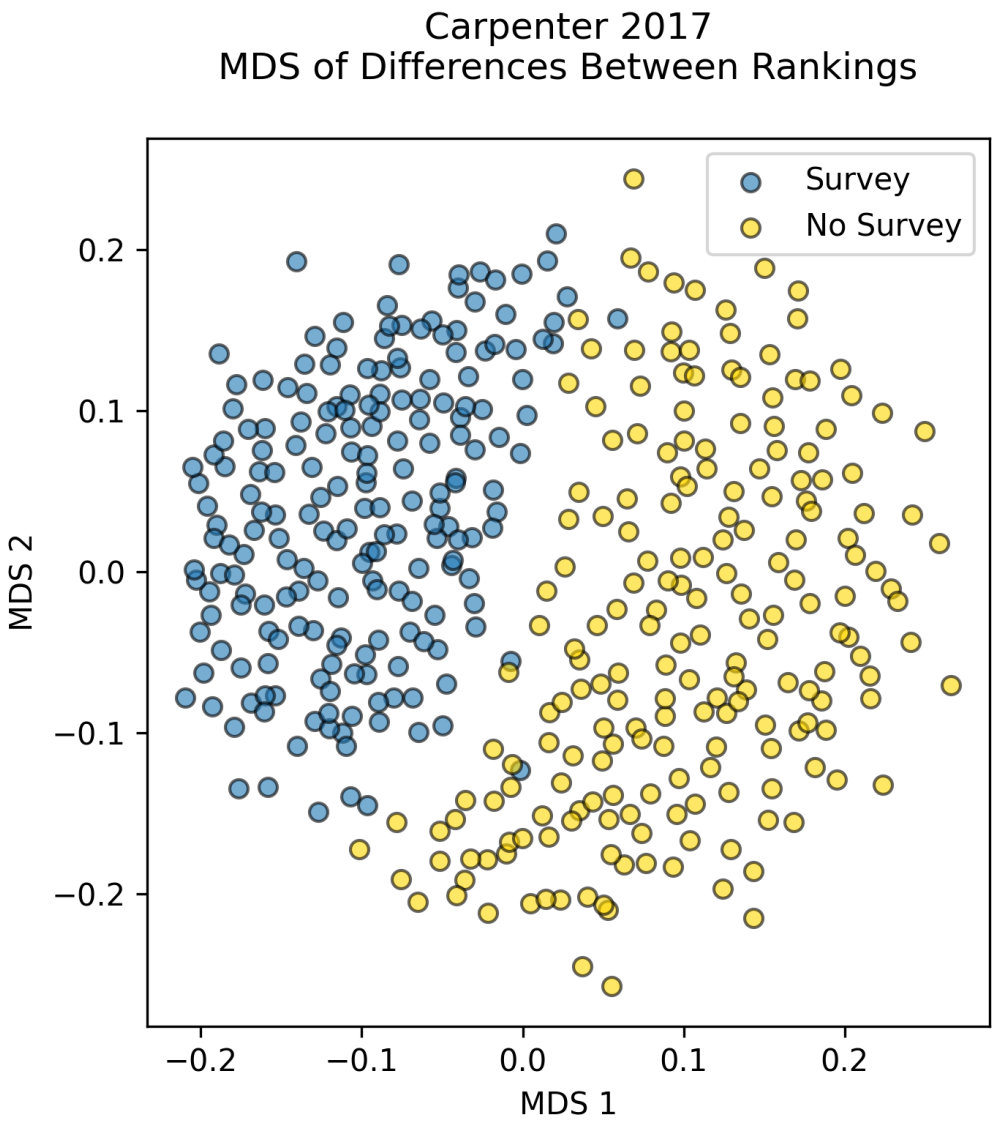



Rankings from 200 Bootstrap Samples (p = 1, q = 1)
Ranking (out of 44)
31
21
11
1
Survey
No Survey
Angelica arguta
Fragaria virginiana
Erigeron aliceae
Erysimum capitatum
Ligusticum grayi
Highest Ranked Plants (Survey)

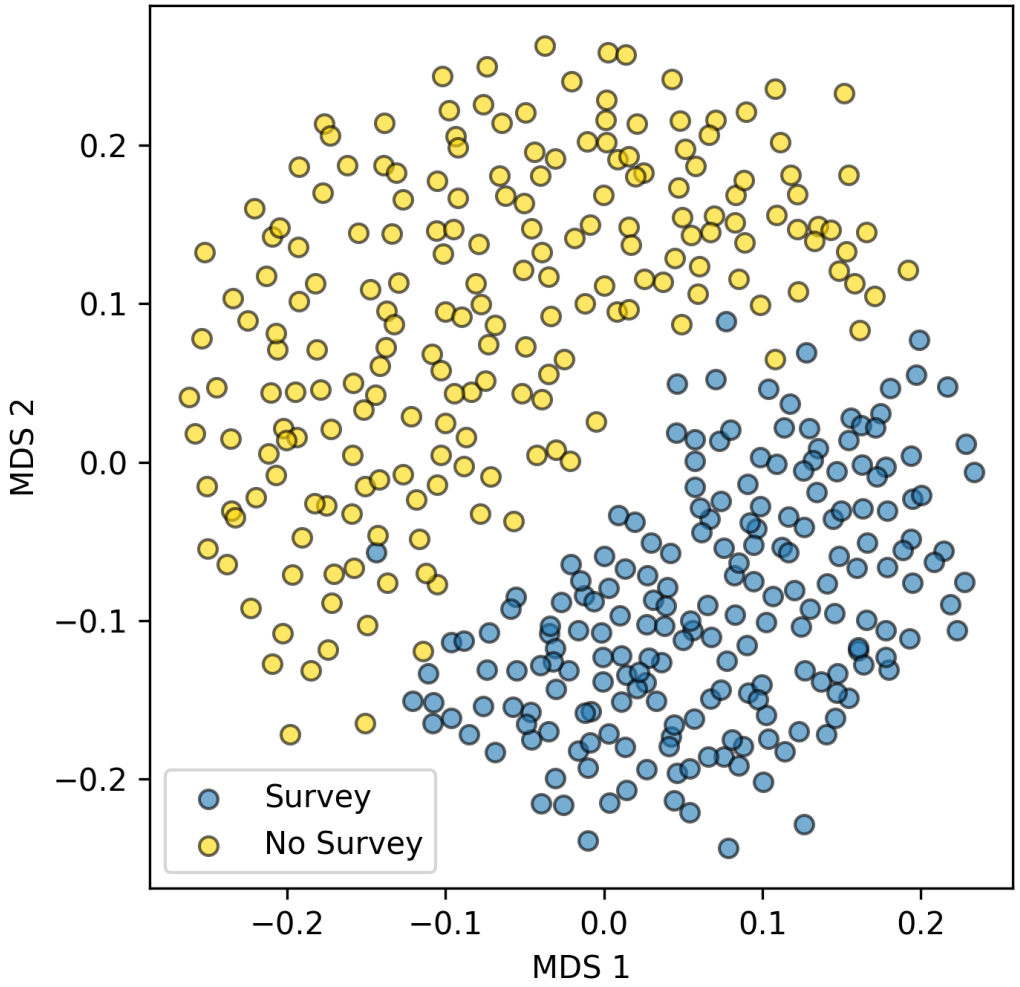
Rankings from (1, 1)-Wasserstein Distances



Rankings from (1, 1)-Wasserstein Distances



Rankings from (2, 2)-Wasserstein Distances



Conclusions

- Data-driven techniques are of interest for real-world interventions such as developing seed mixes which are beneficial to pollinators.

W. Glenny, J. Runyon, and L. Burkle, *Assessing pollinator friendliness of plants and designing mixes to restore habitat for bees*, Jan. 2022.

Conclusions

- Data-driven techniques are of interest for real-world interventions such as developing seed mixes which are beneficial to pollinators.

  W. Glenny, J. Runyon, and L. Burkle, *Assessing pollinator friendliness of plants and designing mixes to restore habitat for bees*, Jan. 2022.

- Understanding the importance of different data types helps guide optimal strategies for data collection.

Conclusions

- Data-driven techniques are of interest for real-world interventions such as developing seed mixes which are beneficial to pollinators.

  W. Glenny, J. Runyon, and L. Burkle, *Assessing pollinator friendliness of plants and designing mixes to restore habitat for bees*, Jan. 2022.

- Understanding the importance of different data types helps guide optimal strategies for data collection.
- We find that flower survey data can significantly alter the mathematical structure of plant-pollinator communities, demonstrating the importance of flowering plant surveys.

Thank you!

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
U.S. DEPARTMENT OF COMMERCE