Visualizing Multi-dimensional Persistent Homology

Matthew L. Wright

in collaboration with Michael Lesnick

Institute for Mathematics and its Applications
University of Minnesota

What is persistent homology?

Persistent homology is an algebraic method for discerning $\boxed{\text{topological features}}$ of $\boxed{\text{data}}$.

e.g. components,
holes,
graph structure

e.g. set of discrete
points, with a metric

Persistent homology emerged in the past 20 years due to the work of:

Frosini, Ferri, et. al. (Bologna, Italy)

Robins (Boulder, Colorado, USA)

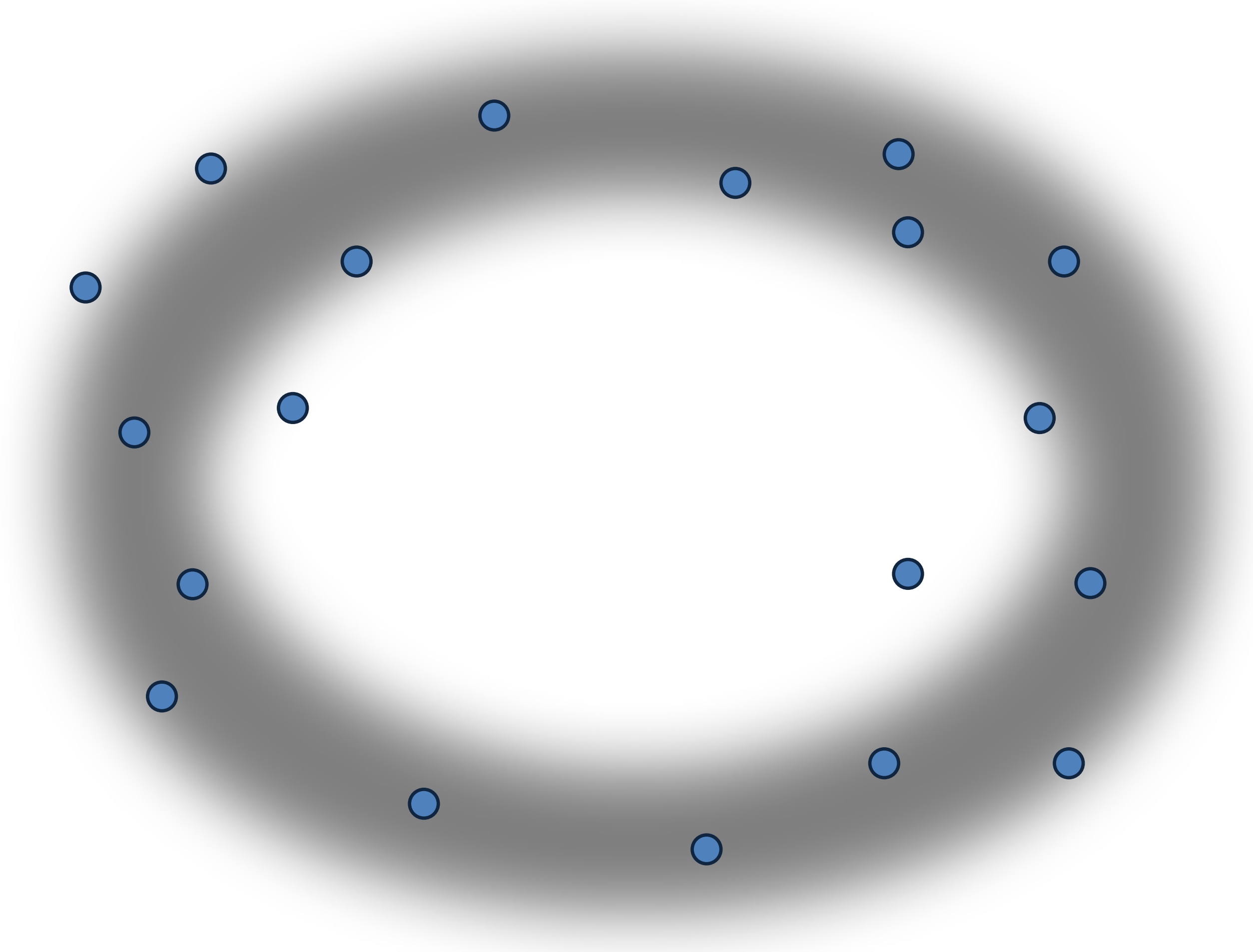
Edelsbrunner (Duke, North Carolina, USA)

Carlsson, de Silva, et. al. (Stanford, California, USA)

Zomorodian (Dartmouth, New Hampshire, USA)

and others

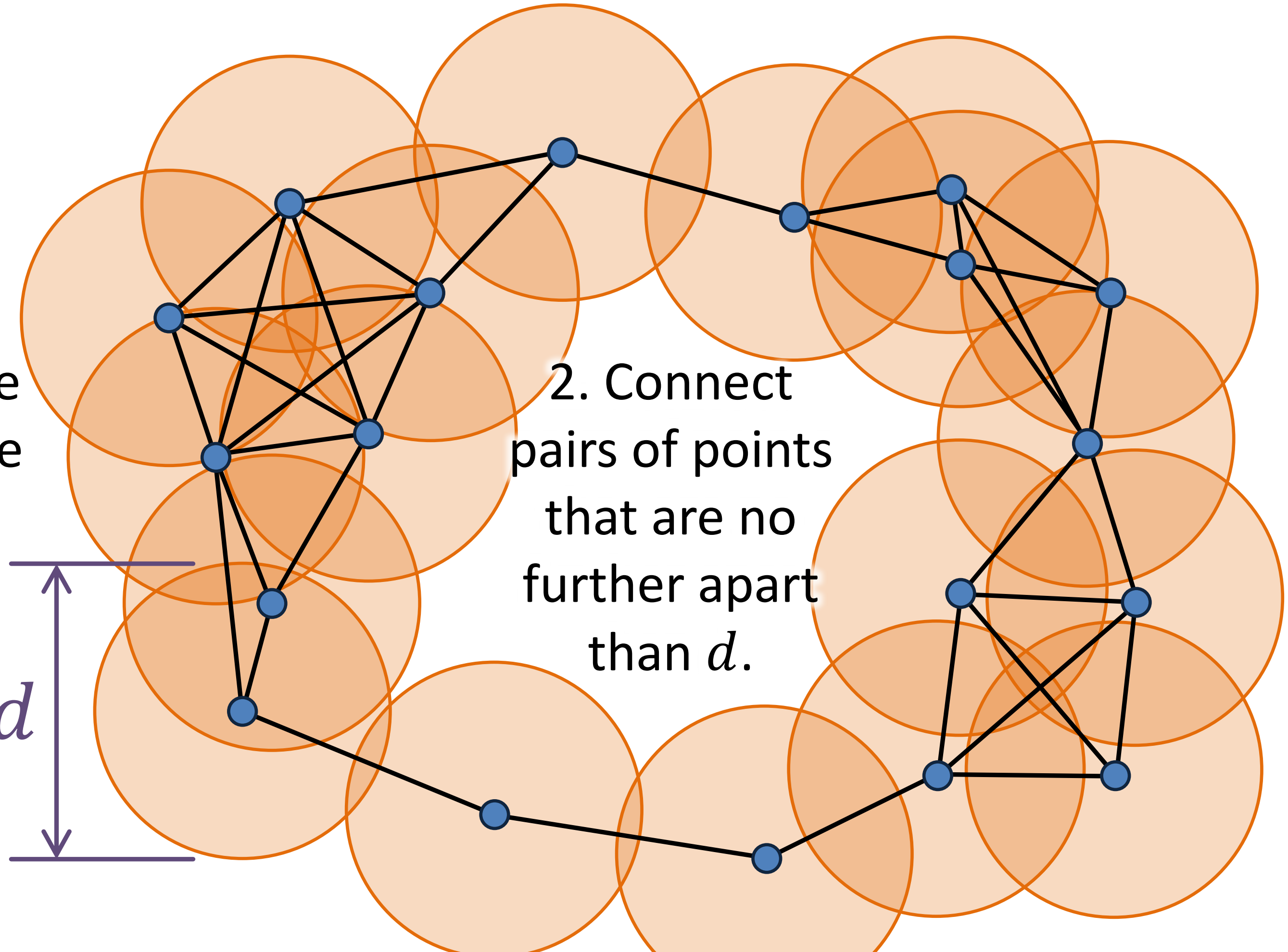
Example: What is the shape of the data?



Problem: Discrete points have trivial topology.

Idea: Connect nearby points.

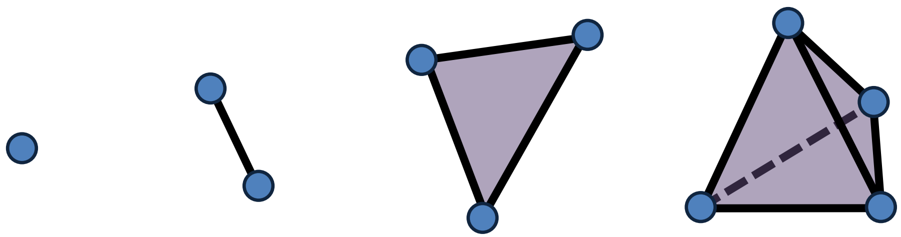
1. Choose a distance d.



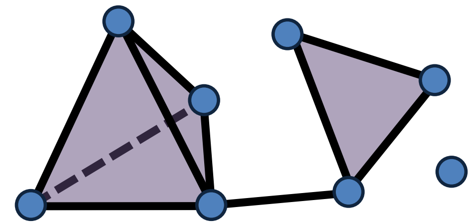
Problem: A graph captures connectivity, but ignores higher-order features, such as holes.

Background

A **simplicial complex** is built from points, edges, triangular faces, etc.

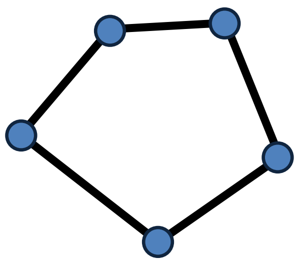


0-simplex 1-simplex 2-simplex 3-simplex (solid)

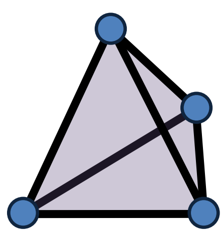


example of a simplicial complex

Homology counts components, holds, voids, etc.



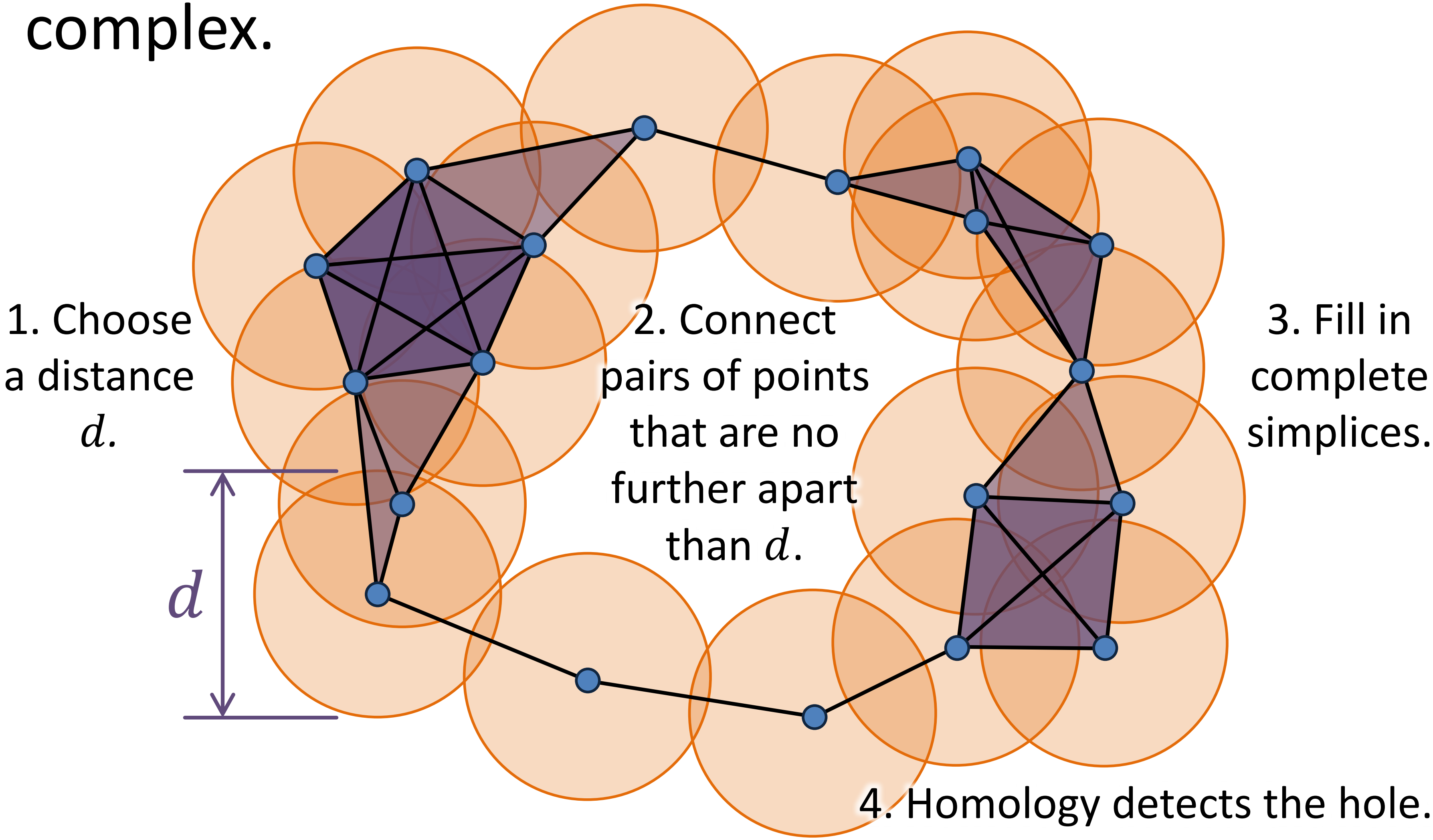
hole



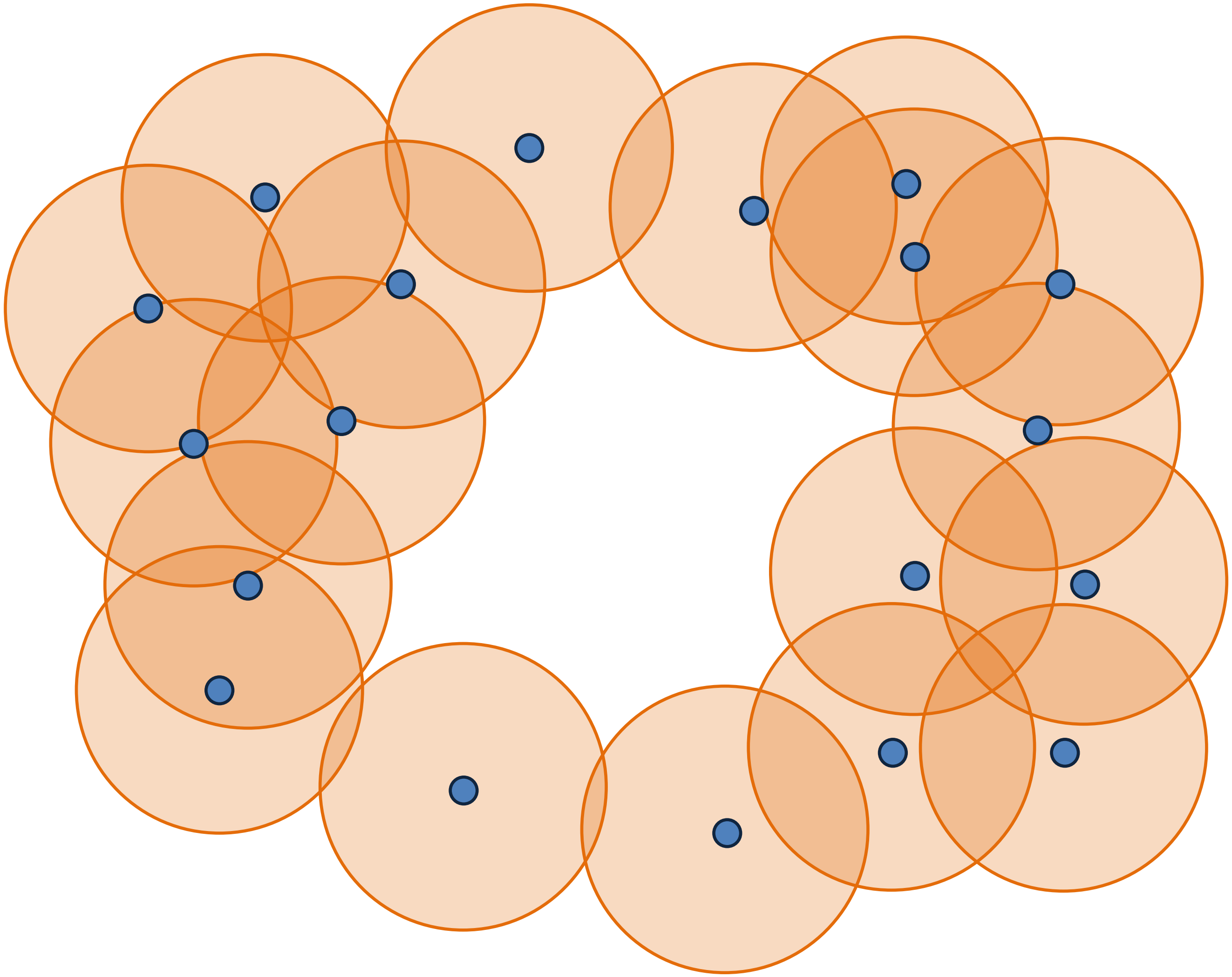
void
(contains faces but empty interior)

Homology of a simplicial complex is computable via linear algebra.

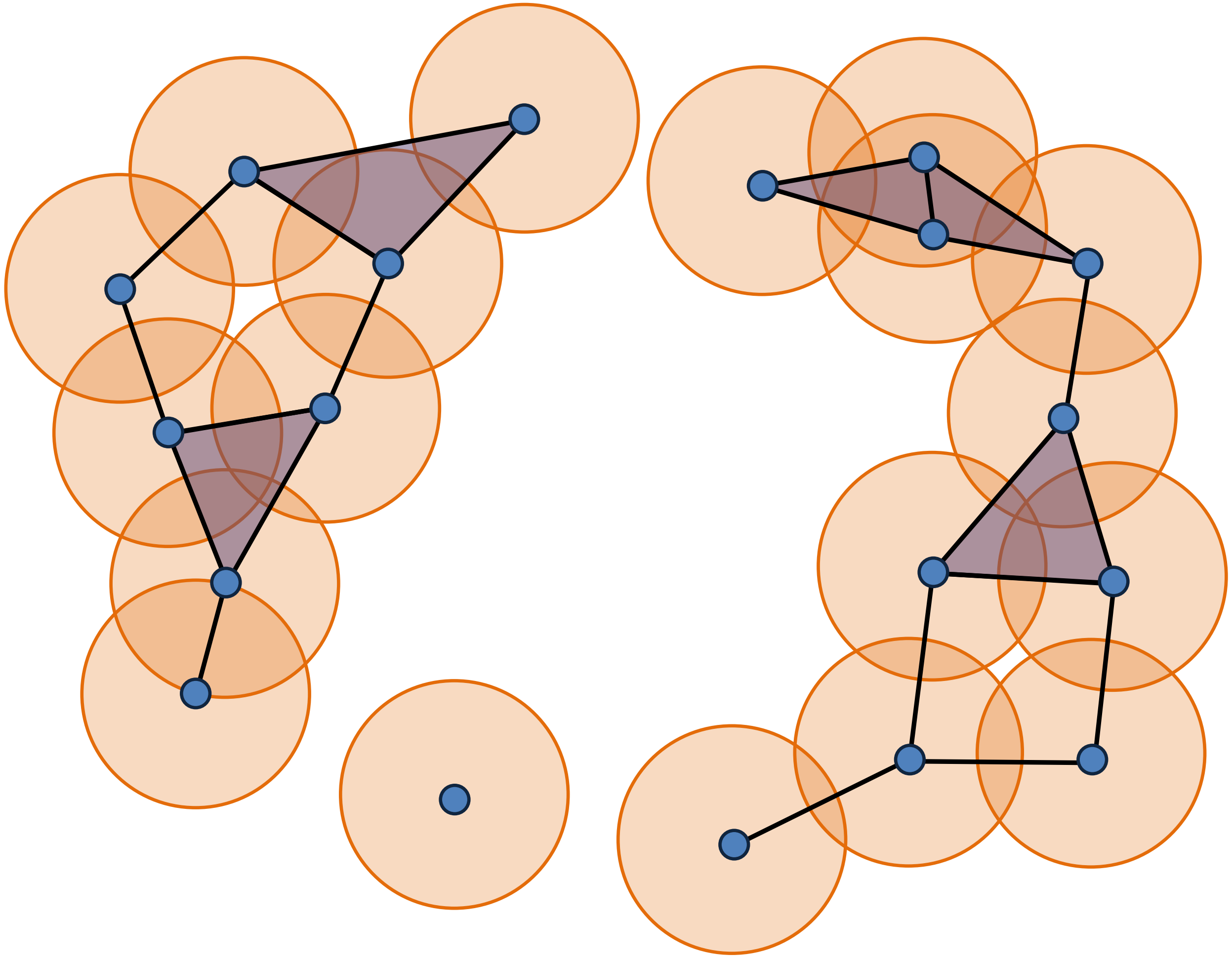
Idea: Connect nearby points, build a simplicial complex.



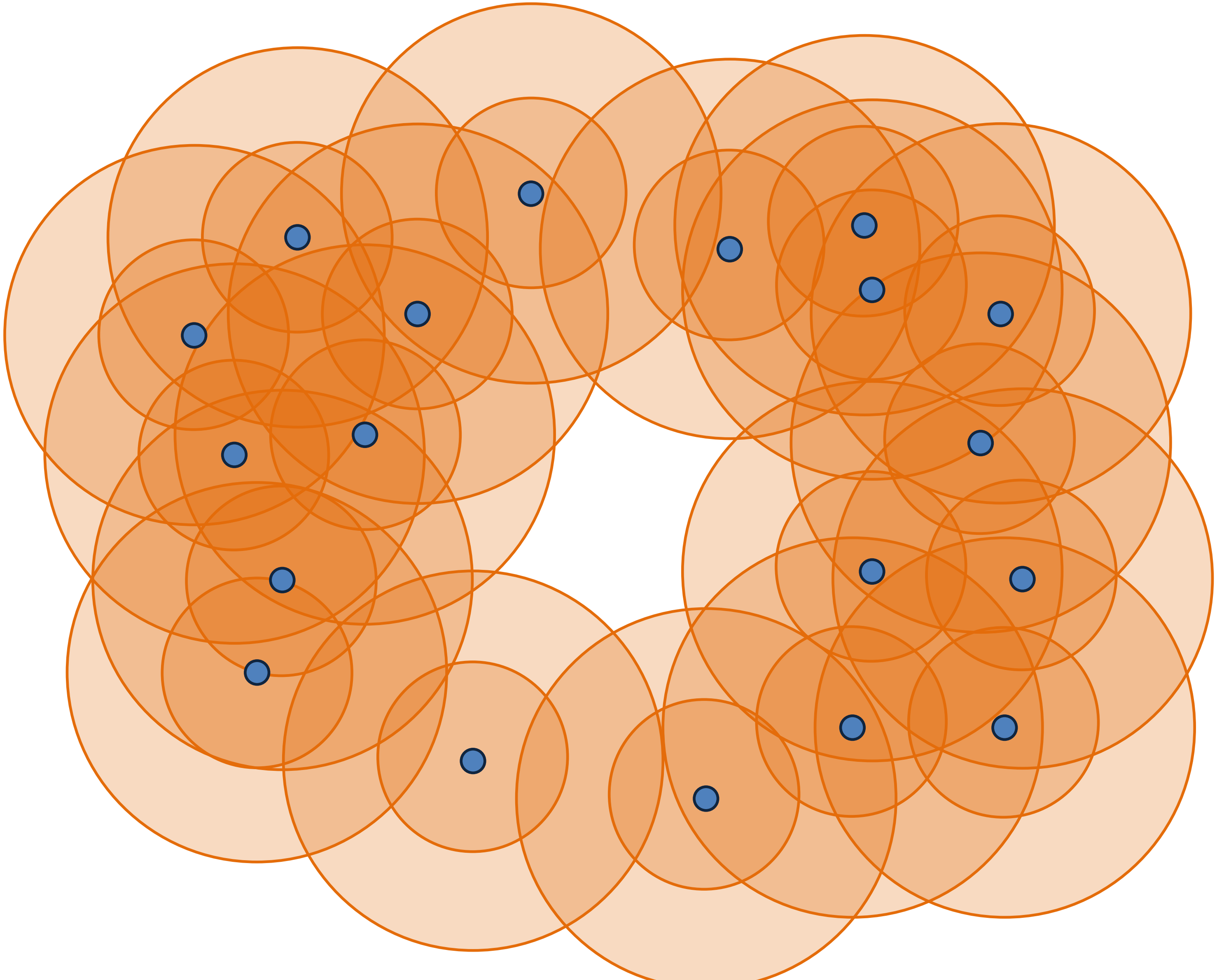
Problem: How do we choose distance d?



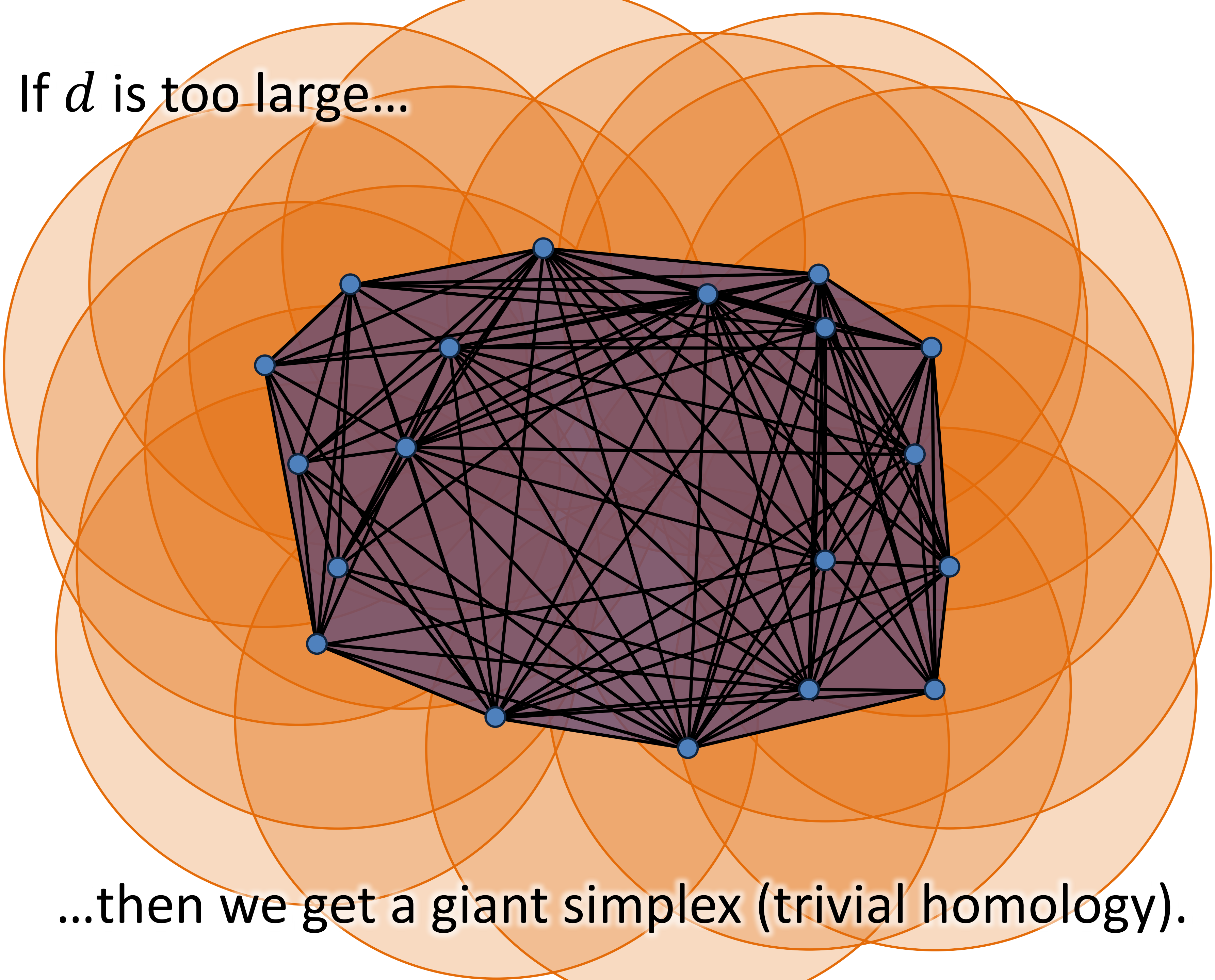
If d is too small...



...then we detect noise.

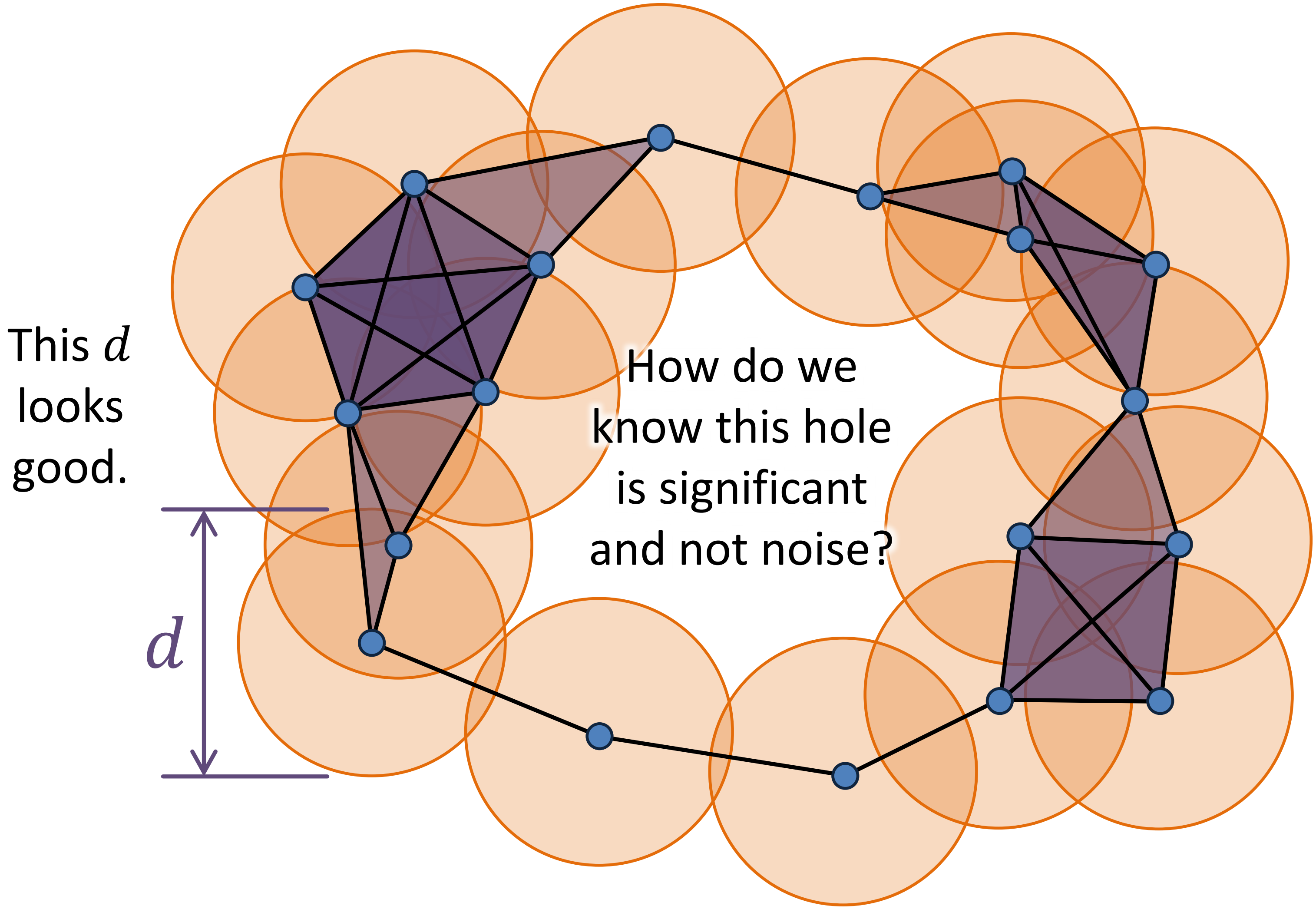


If d is too large…



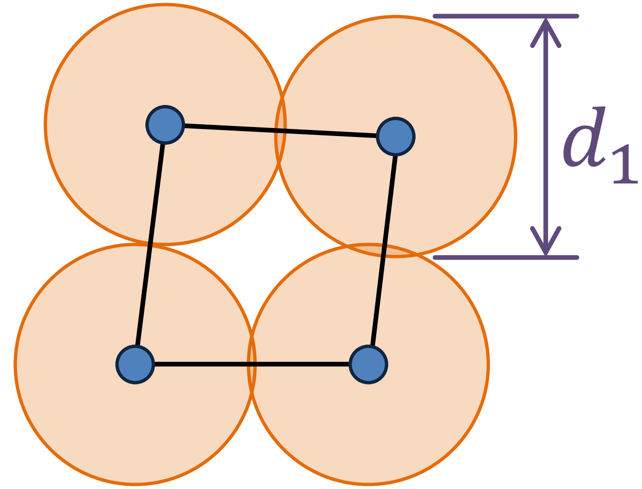
…then we get a giant simplex (trivial homology).

Problem: How do we choose distance d?



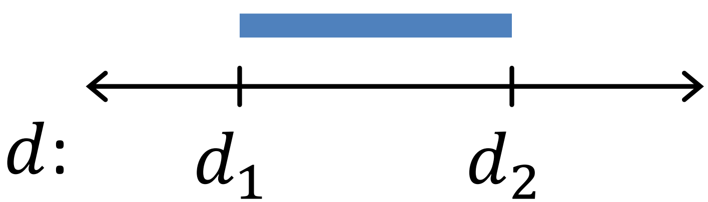
Idea: Consider *all* distances d.

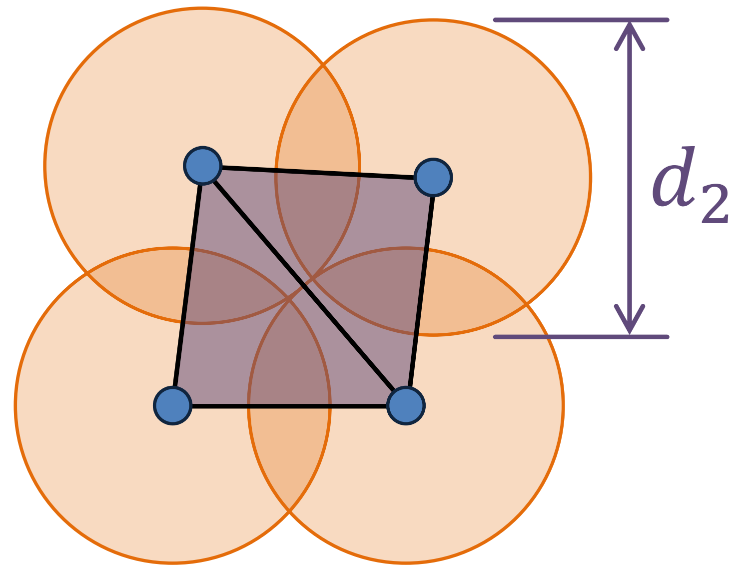
Each hole appears at a particular value of d and disappears at another value of d.



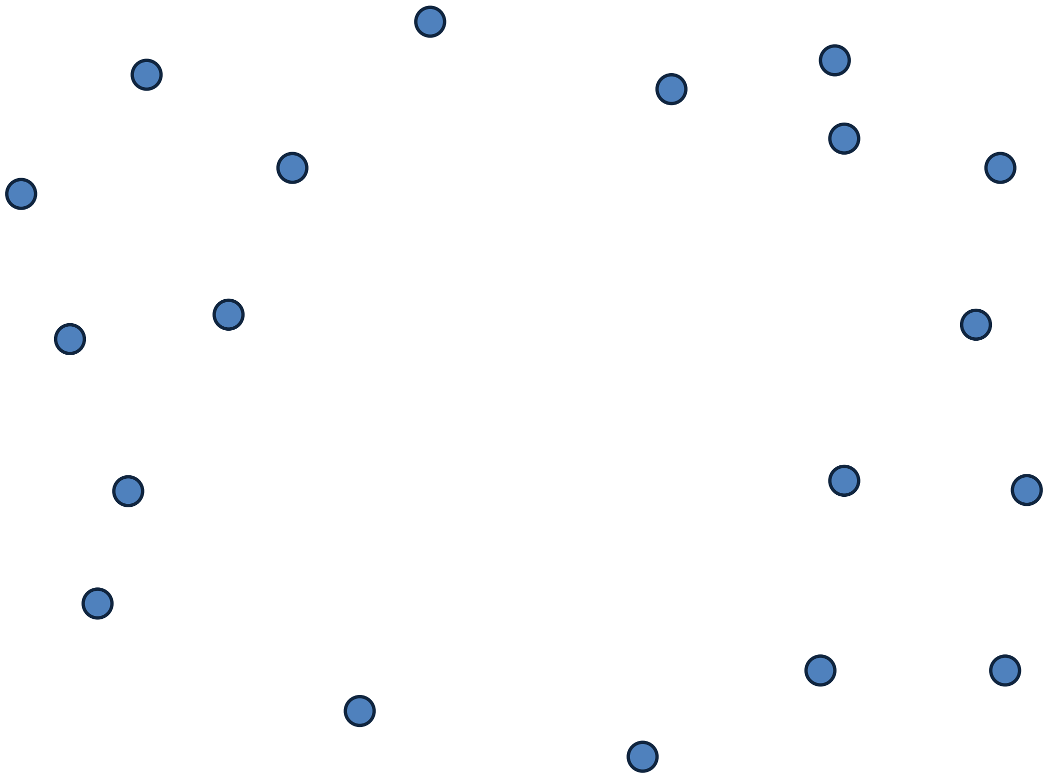
We can represent the **persistence** of this hole as a pair (d_1, d_2).

We visualize this pair as a bar from d_1 to d_2:

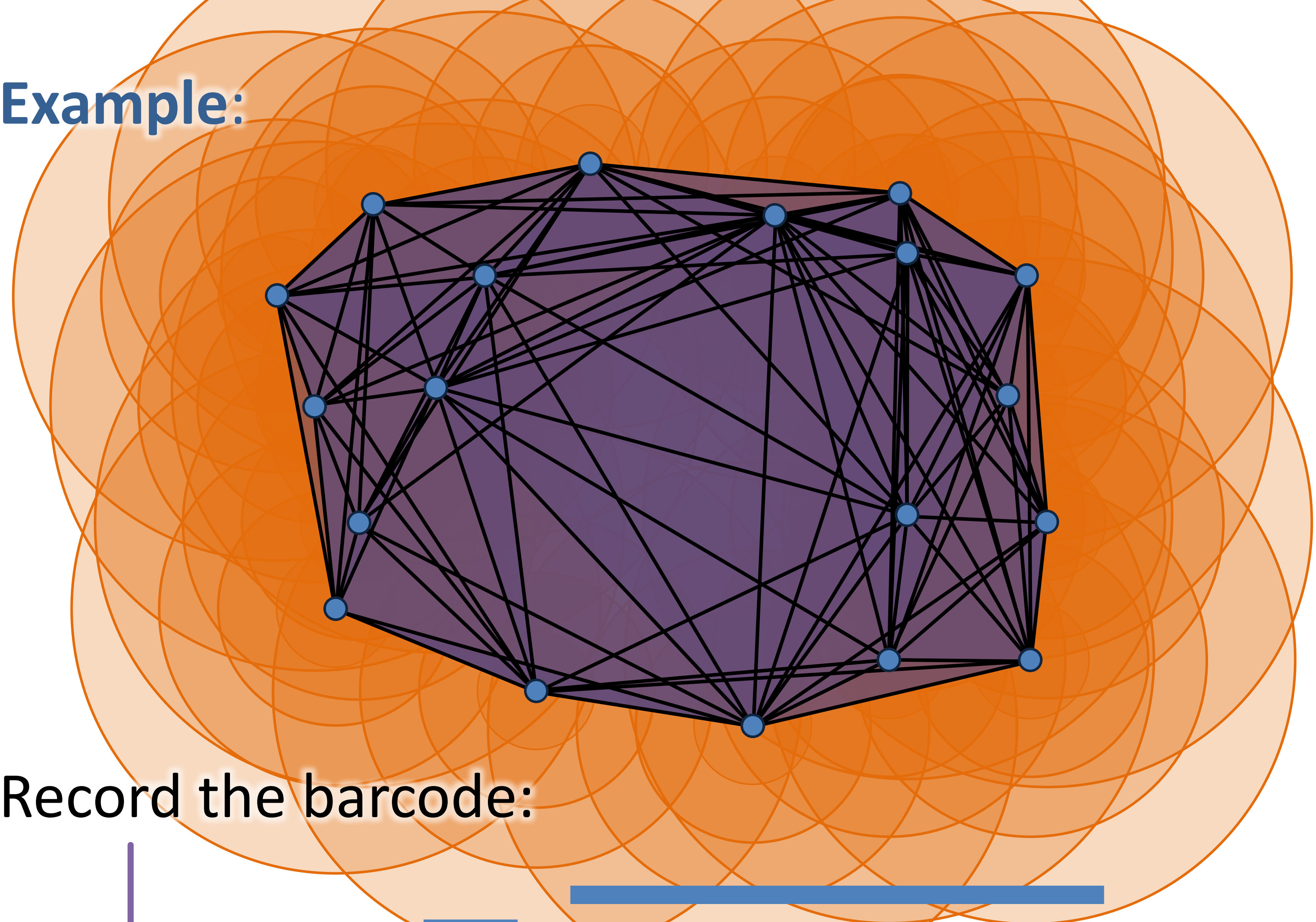




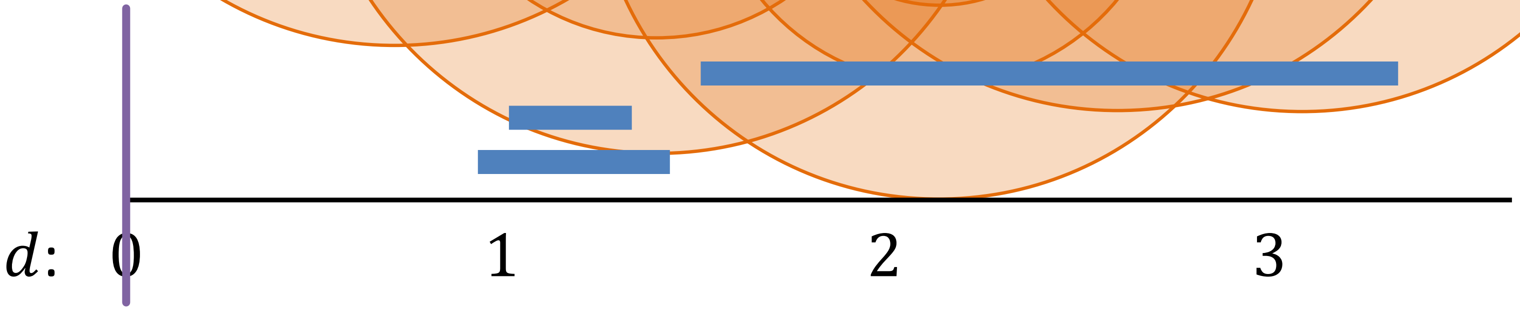
A collection of bars is a **barcode**.



Example:



Record the barcode:

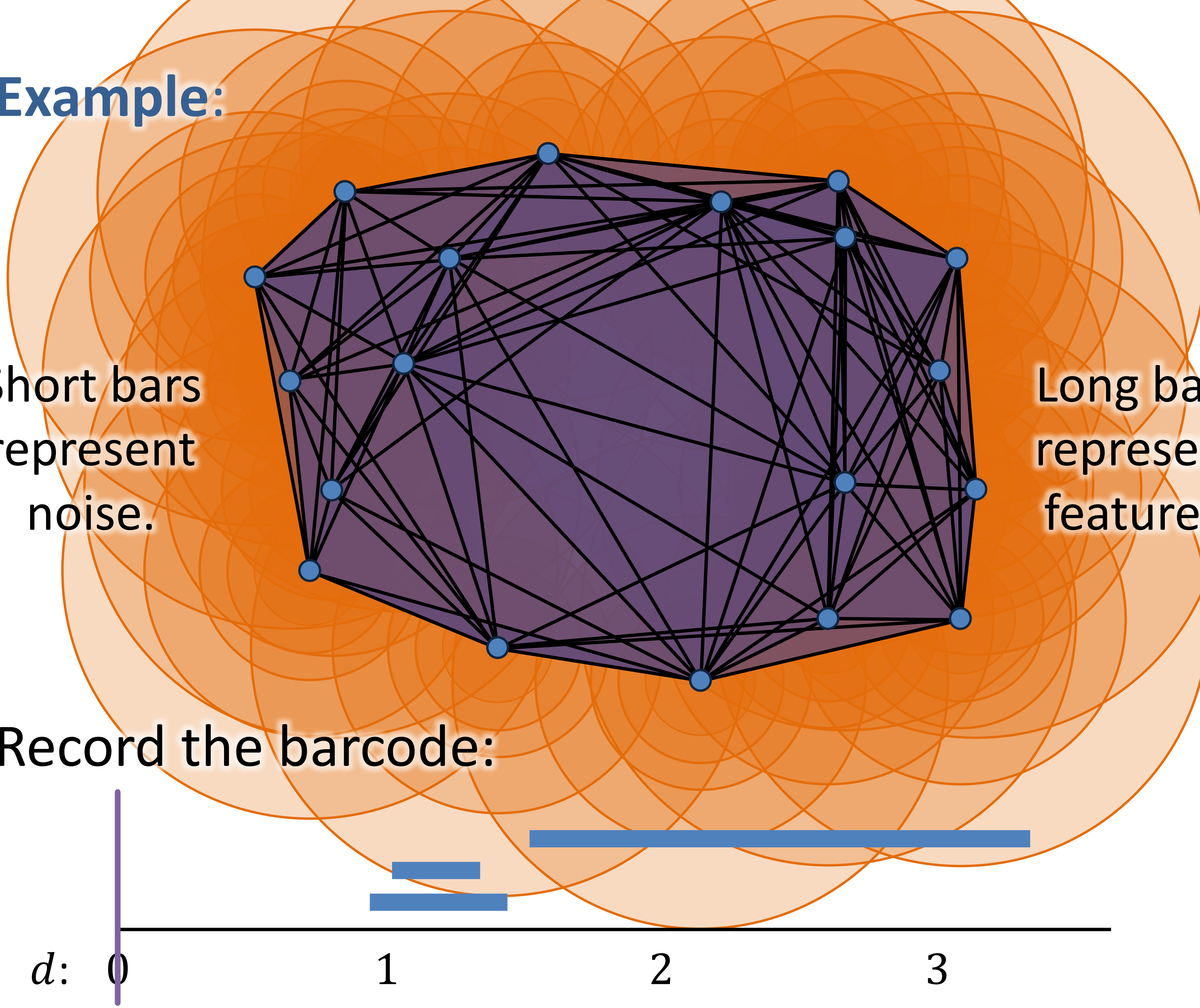


Example:

Short bars represent noise.

Long bars represent features.

Record the barcode:

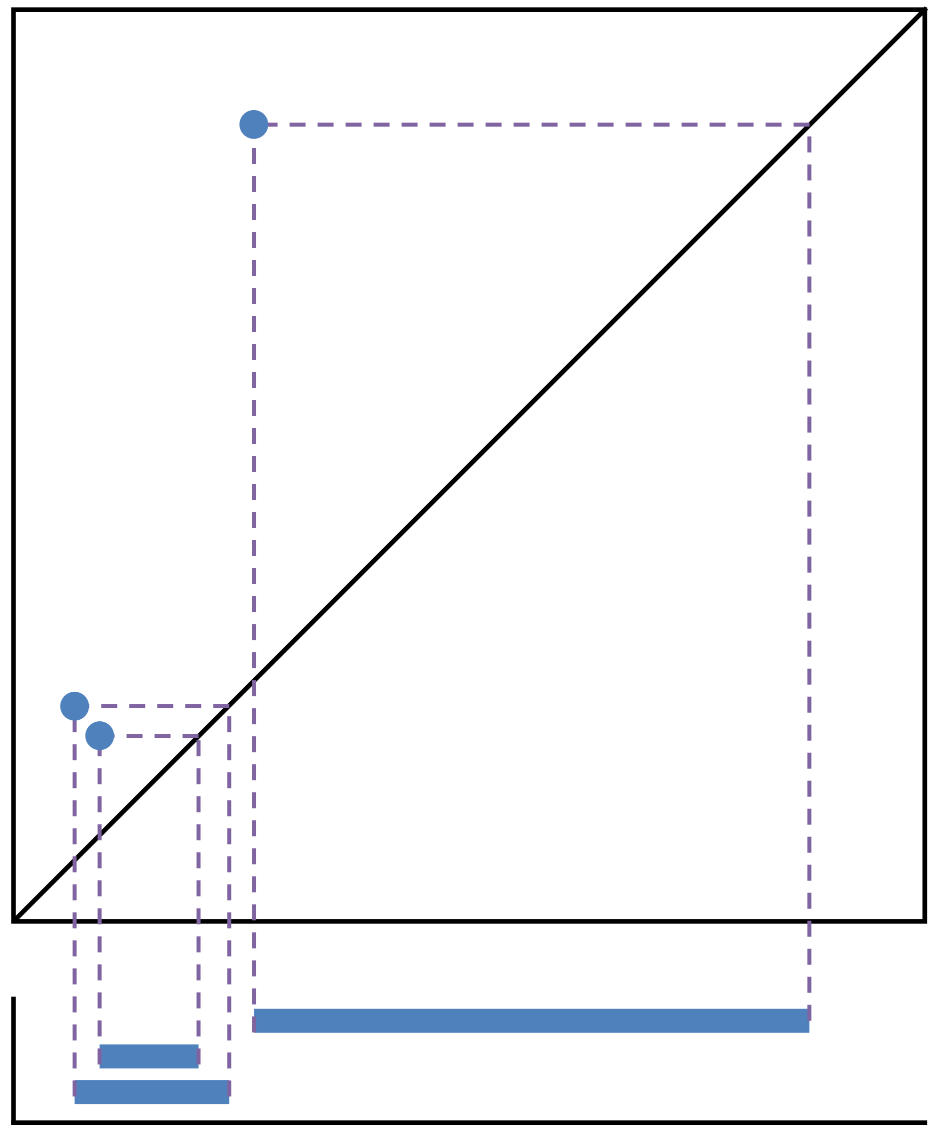


A **persistence diagram** is an alternate depiction of a barcode.

Instead of drawing (a, b) as a bar from a to b, draw a dot at coordinates (a, b).

Dots far from the diagonal represent features.

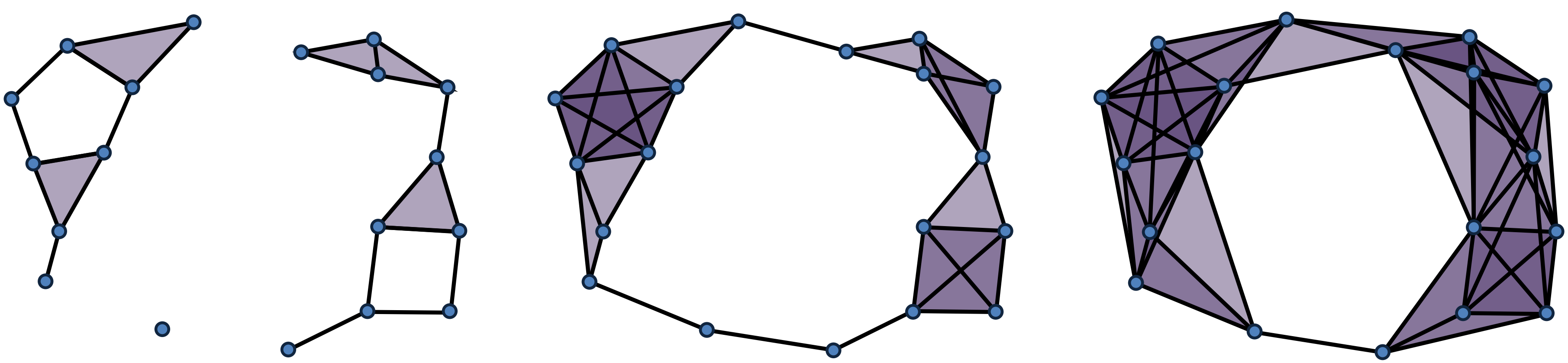
Dots near the diagonal represent noise.



A barcode is a visualization of an algebraic structure.

Consider the sequence (C_i) of complexes associated to a point cloud for an sequence of distance values:

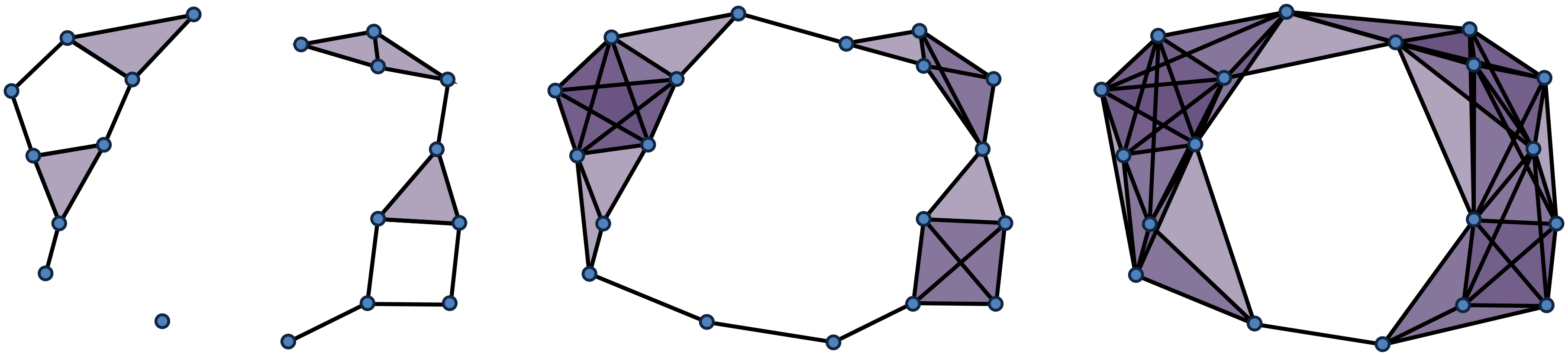
$$C_1 \xhookrightarrow{\iota} C_2 \xhookrightarrow{\iota} C_3$$



A barcode is a visualization of an algebraic structure.

Consider the sequence (C_i) of complexes associated to a point cloud for an sequence of distance values:

$$\cdots \hookrightarrow C_1 \hookrightarrow C_2 \hookrightarrow C_3 \hookrightarrow C_4 \hookrightarrow C_5 \hookrightarrow C_6 \hookrightarrow C_7 \hookrightarrow \cdots$$



This sequence of complexes, with maps, is a **filtration**.

A barcode is a visualization of an algebraic structure.

Filtration: $C_1 \hookrightarrow C_2 \hookrightarrow \cdots \hookrightarrow C_m$

Homology with coefficients from a field F:

$$H_*(C_1) \to H_*(C_2) \to \cdots \to H_*(C_m)$$

Let $M = H_*(C_1) \oplus H_*(C_2) \oplus \cdots \oplus H_*(C_m)$.

For $i \leq j$, the map $f_i^j : H_*(C_i) \to H_*(C_j)$ is induced by the inclusion $C_i \hookrightarrow C_j$.

Let $F[x]$ act on M by $x^k \alpha = f_i^{i+k}(\alpha)$ for any $\alpha \in H_*(C_i)$.

i.e. x acts as a shift map $x : H_*(C_i) \to H_*(C_{i+1})$

Then M is a graded $F[x]$-module, called a **persistence module**.

A barcode is a visualization of an algebraic structure.

Let $M = H_*(C_1) \oplus H_*(C_2) \oplus \cdots \oplus H_*(C_m)$.

Then M is a graded $F[x]$-module.

The structure theorem for finitely generated modules over PIDs implies:

$$M \cong \bigoplus_i \underbrace{x^{t_i} \cdot F[x]}_{\text{homology generators that appear at } t_i \text{ and persist forever after}} \oplus \left(\bigoplus_j \underbrace{x^{r_j} \cdot \left({F[x]} \big/ {x^{s_j} \cdot F[x]} \right)}_{\text{homology generators that appear at } r_j \text{ and persist until } r_j + s_j} \right)$$

i.e. bars of the form (t_j, ∞)

i.e. bars of the form (r_j, s_j)

Thus, the barcode is a complete discrete invariant.

Stability:

Persistence barcodes are stable with respect to pertubations of the data.

Cohen-Steiner, Edelsbrunner, Harer (2007)

Computation:

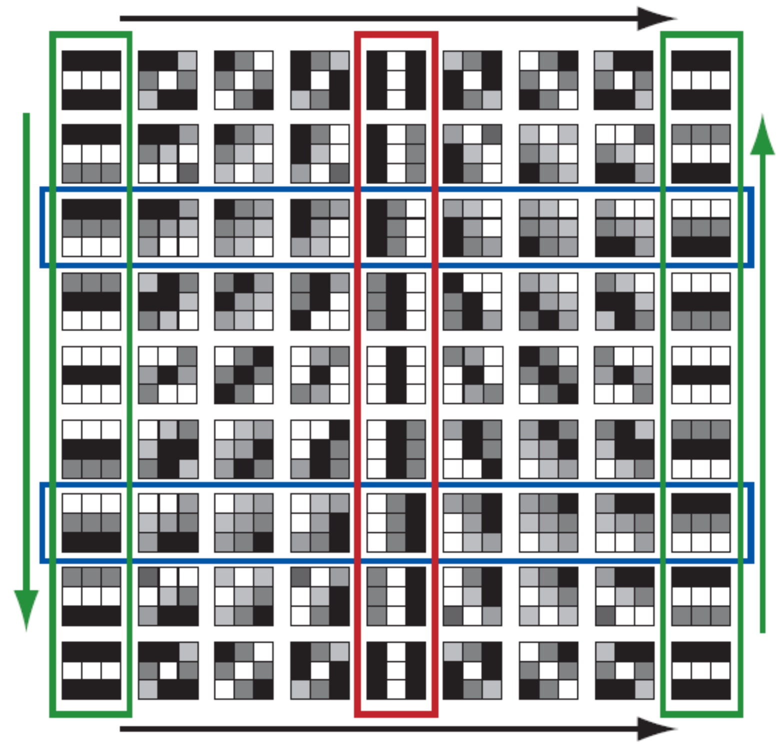
The barcode is computable via linear algebra on the boundary matrix. Runtime is $O(n^3)$, where n is the number of simplices.

Zomorodian and Carlsson (2005)

Where has persistent homology been used?

Image Processing

The space of 3x3
high-contrast patches
from digital images
has the topology of a
Klein bottle.

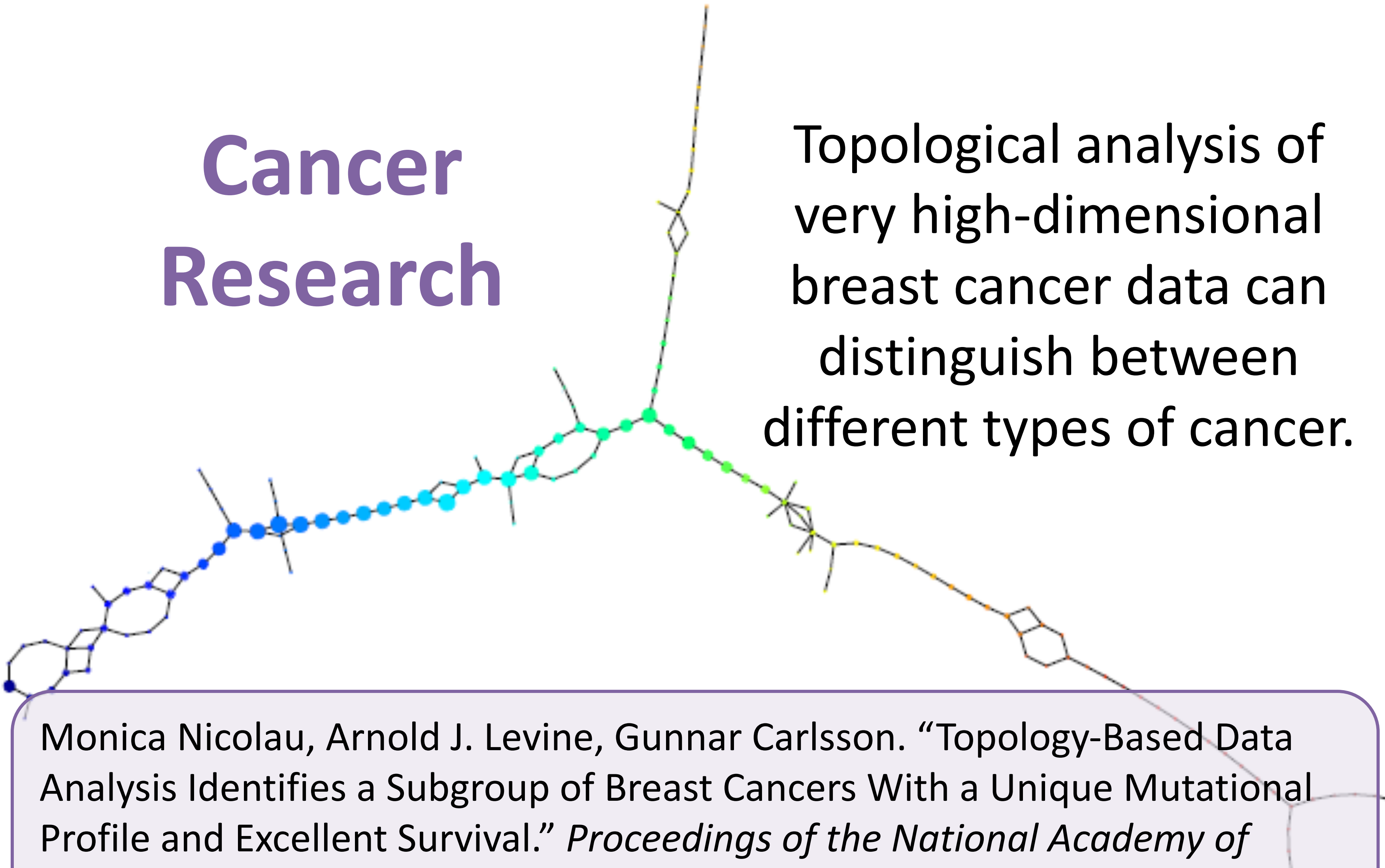


Gunnar Carlsson, Tigran Ishkhanov, Vin de Silva, Afra Zomorodian. “On the Local Behavior of Spaces of Natural Images.” *Journal of Computer Vision*. Vol. 76, No. 1, 2008, p. 1 – 12.

Image credit: Robert Ghrist. “Barcodes: The Persistent Topology of Data.” *Bulletin of the American Mathematical Society*. Vol. 45, no. 1, 2008, p. 61-75.

Where has persistent homology been used?

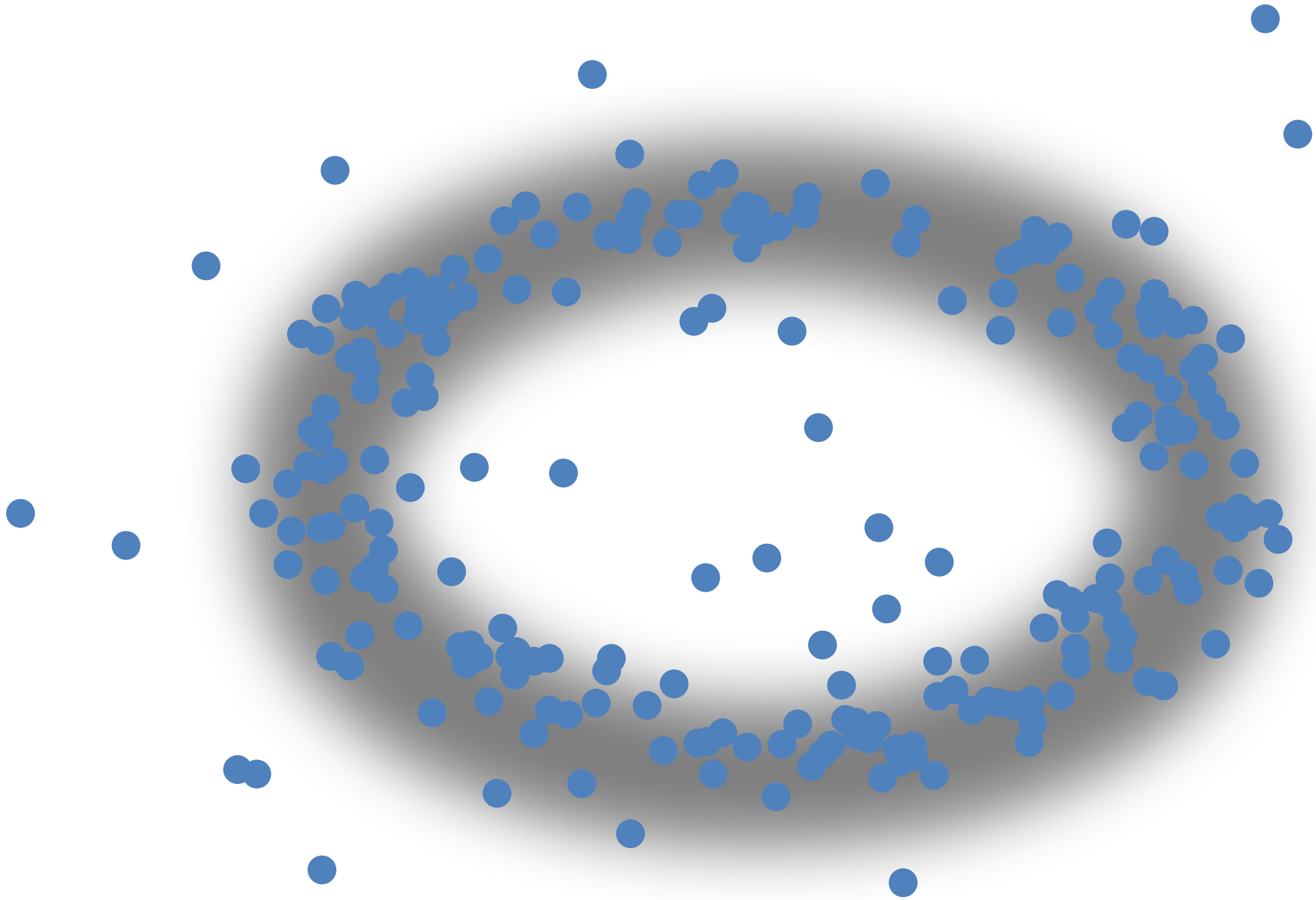
Cancer Research



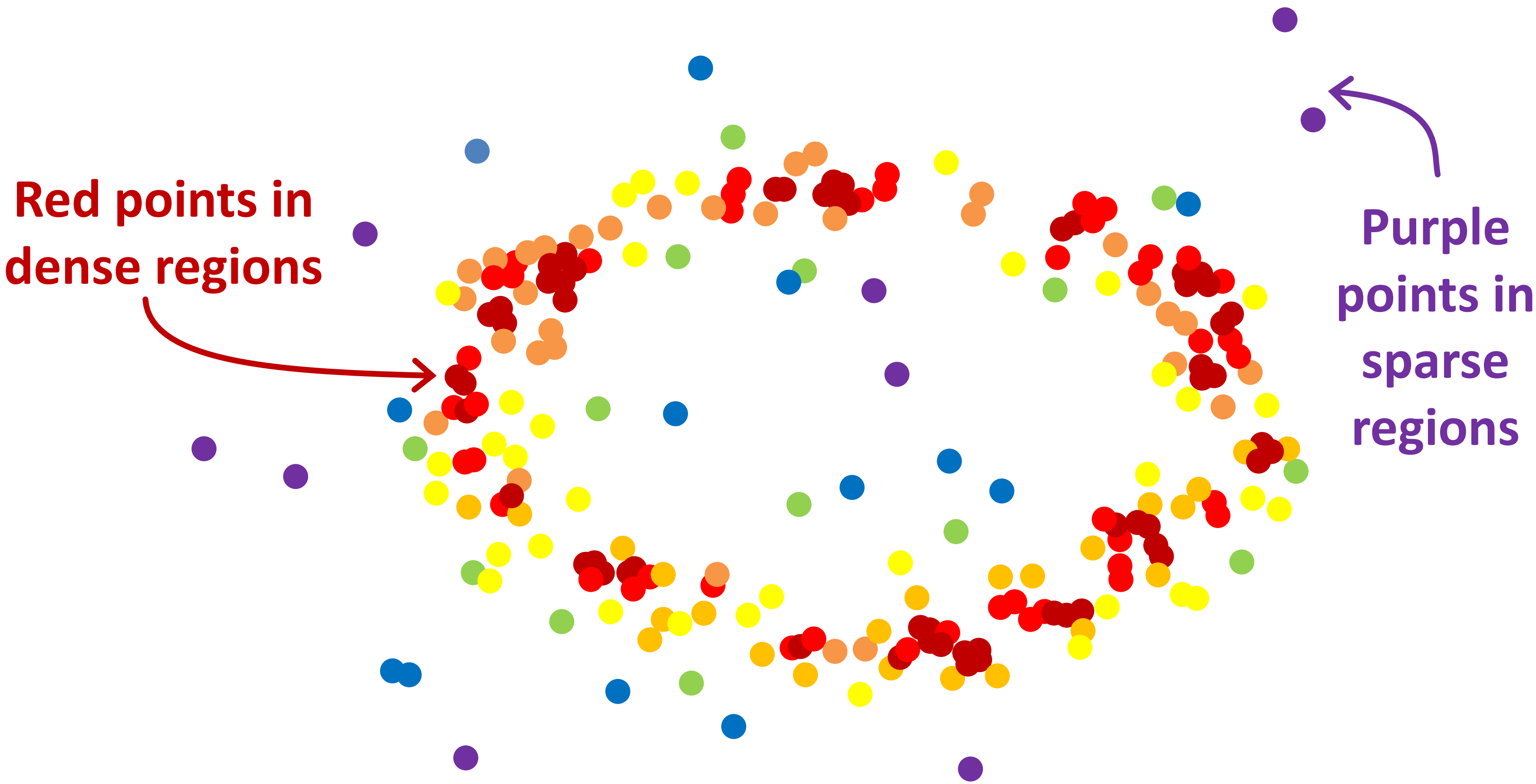
Topological analysis of very high-dimensional breast cancer data can distinguish between different types of cancer.

Monica Nicolau, Arnold J. Levine, Gunnar Carlsson. “Topology-Based Data Analysis Identifies a Subgroup of Breast Cancers With a Unique Mutational Profile and Excellent Survival.” *Proceedings of the National Academy of Sciences*. Vol. 108, No. 17, 2011, p. 7265 – 7270.

Problem: Persistent homology is sensitive to outliers.



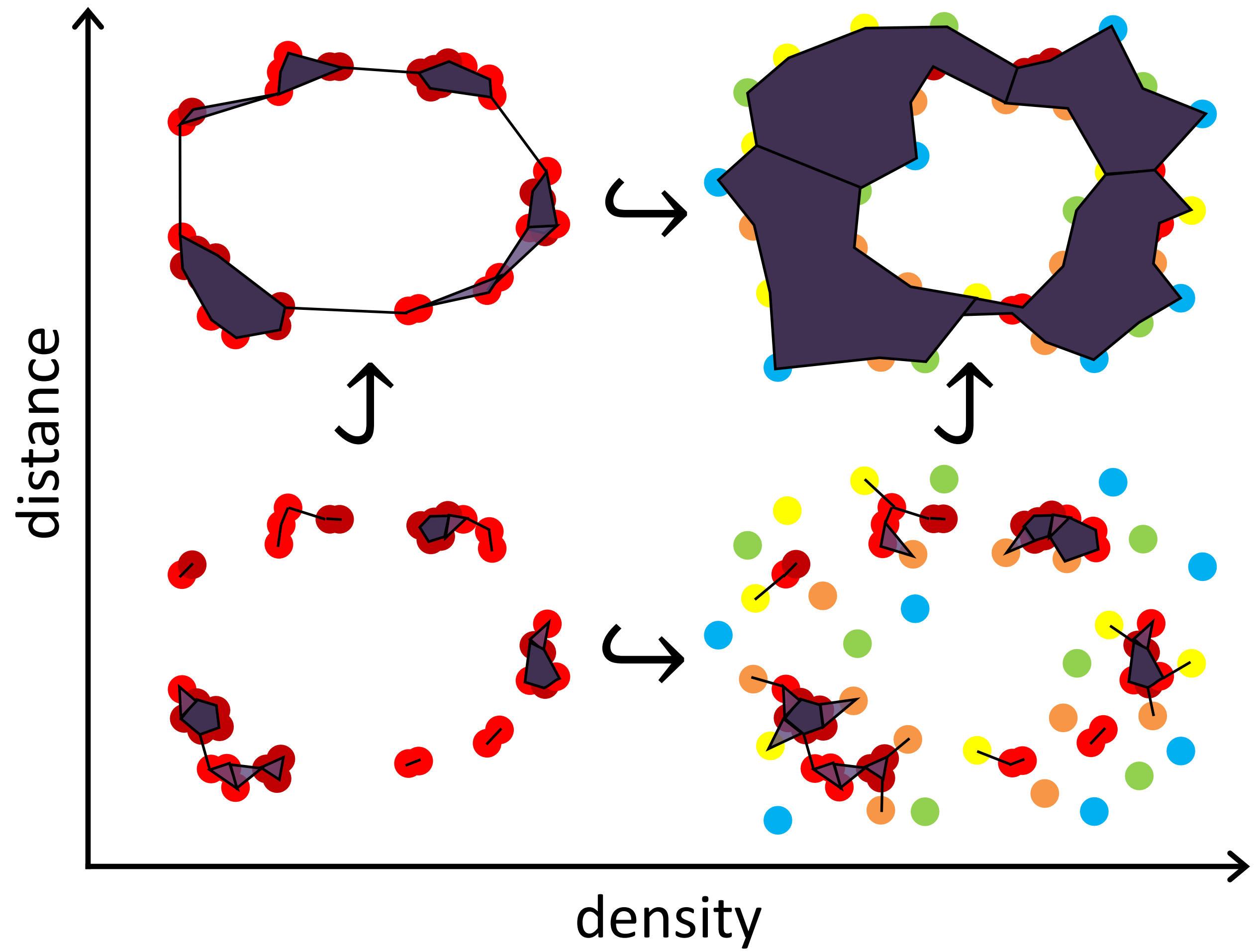
Problem: Persistent homology is sensitive to outliers.



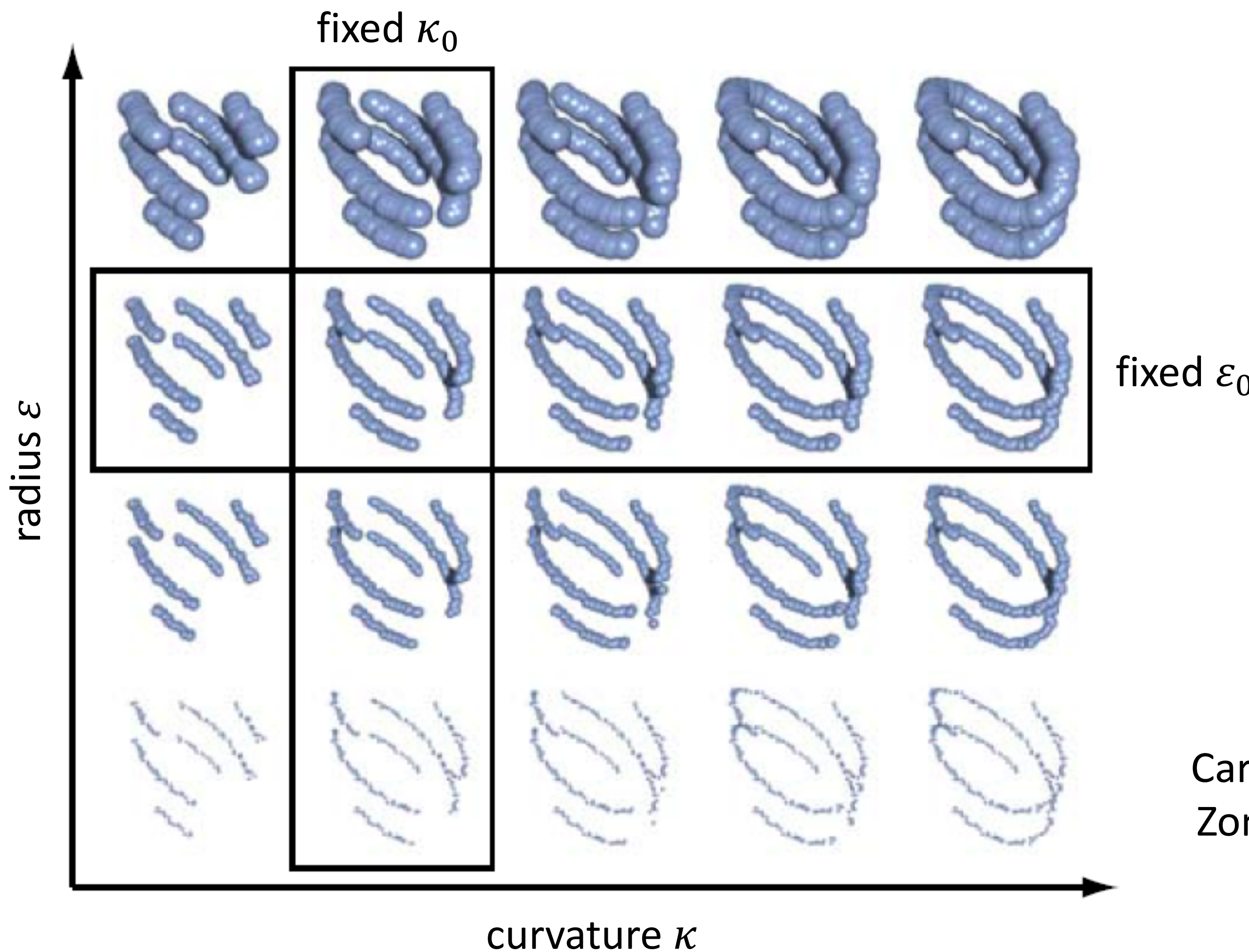
Do we have to threshold by density?

Multi-dimensional persistence: Allows us to work with data indexed by two parameters, such as distance and density.

We obtain a **bifiltration**: a set of simplicial complexes indexed by *two* parameters.



Example: A bifiltration indexed by curvature κ and radius ε.



Carlsson and Zomorodian (2009)

Ordinary persistence requires fixing either κ or ε.

Algebraic Structure of Multi-dimensional Persistence

The homology of a bifiltered simplicial complex is a finitely-generated bigraded module: i.e. a 2-graded module over $F[x, y]$ for a field F.

We call this a **2-dimensional persistence module**.

Problem: The structure of multi-graded modules is much more complicated than that of graded modules.

There is no complete, discrete invariant for multi-dimensional persistence modules (Carlsson and Zomorodian, 2007).

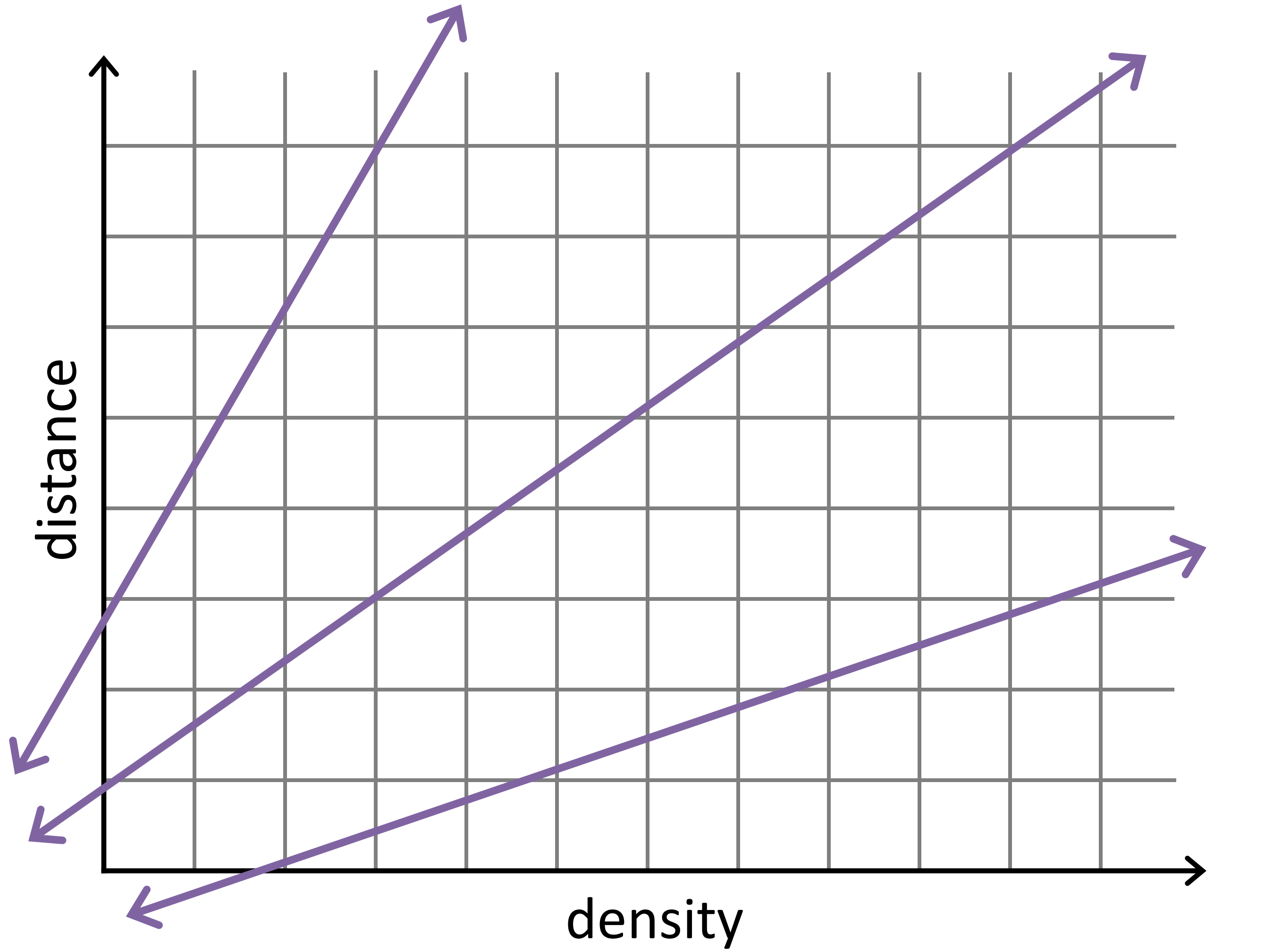
Thus, *there is no multi-dimensional barcode*.

Question: How can we visualize multi-dimensional persistence?

Concept: Visualize a barcode along any one-dimensional slice of a multi-dimensional parameter space.

Example:

Along any one-dimensional slice, a barcode exists.

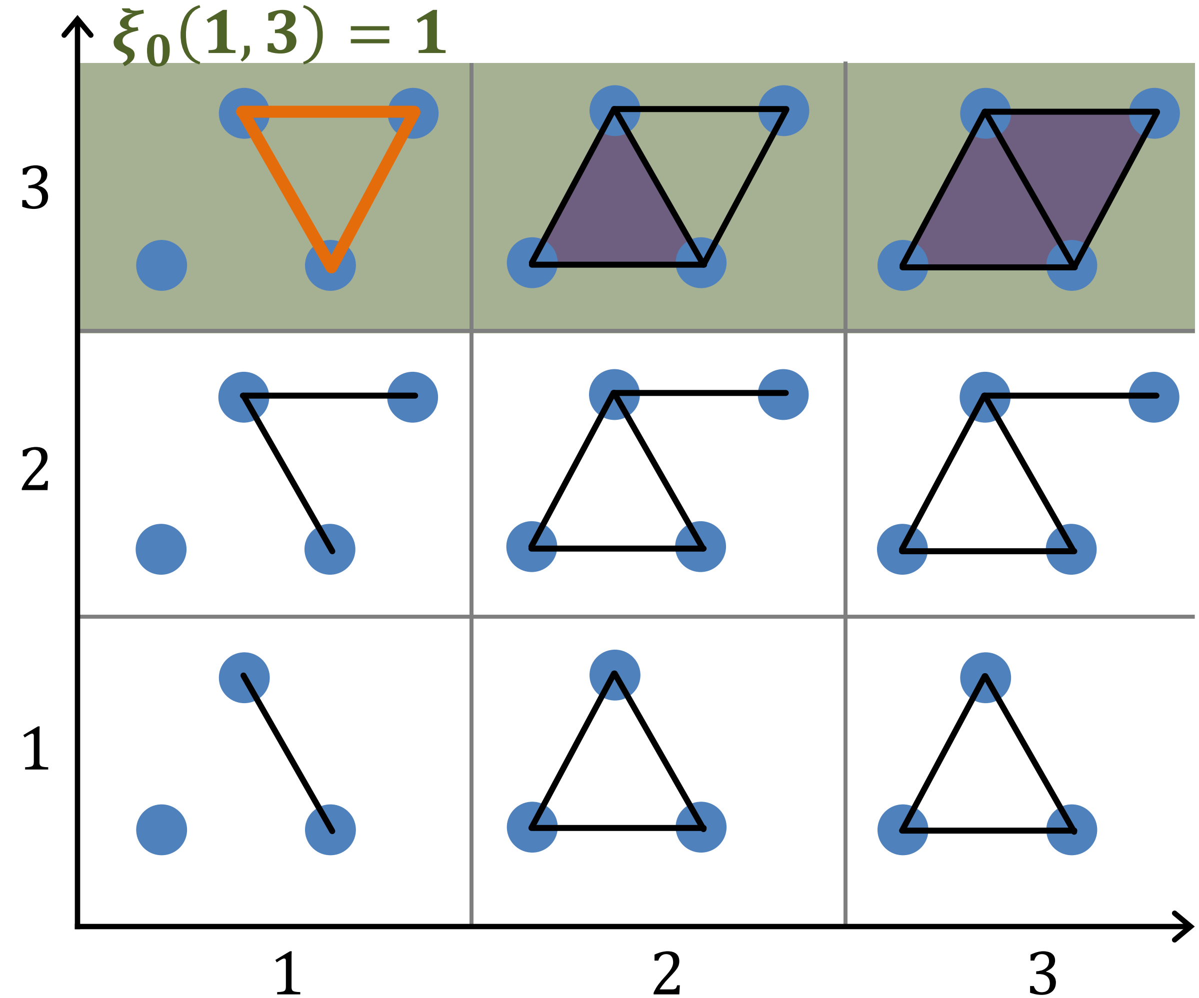


Bi-graded Betti numbers ξ_0 and ξ_1

These are functions,
$\xi_0, \xi_1 : \mathbb{N}^2 \to \mathbb{N}$

ξ_0 indicates coordinates at which homology appears

Example: 1st homology (holes)

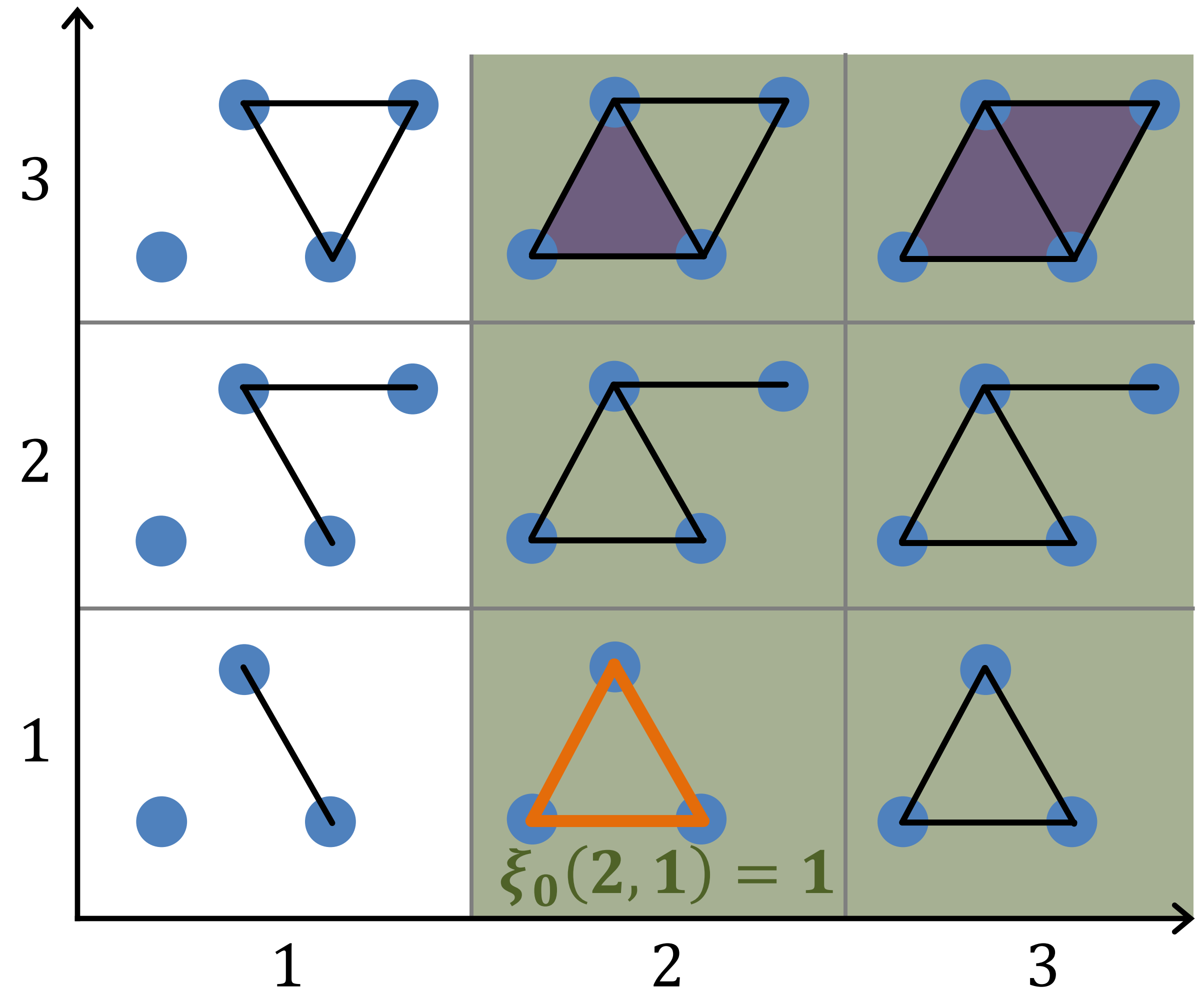


Bi-graded Betti numbers ξ_0 and ξ_1

These are functions,
$\xi_0, \xi_1 : \mathbb{N}^2 \to \mathbb{N}$

ξ_0 indicates coordinates at which homology appears

Example: 1st homology (holes)



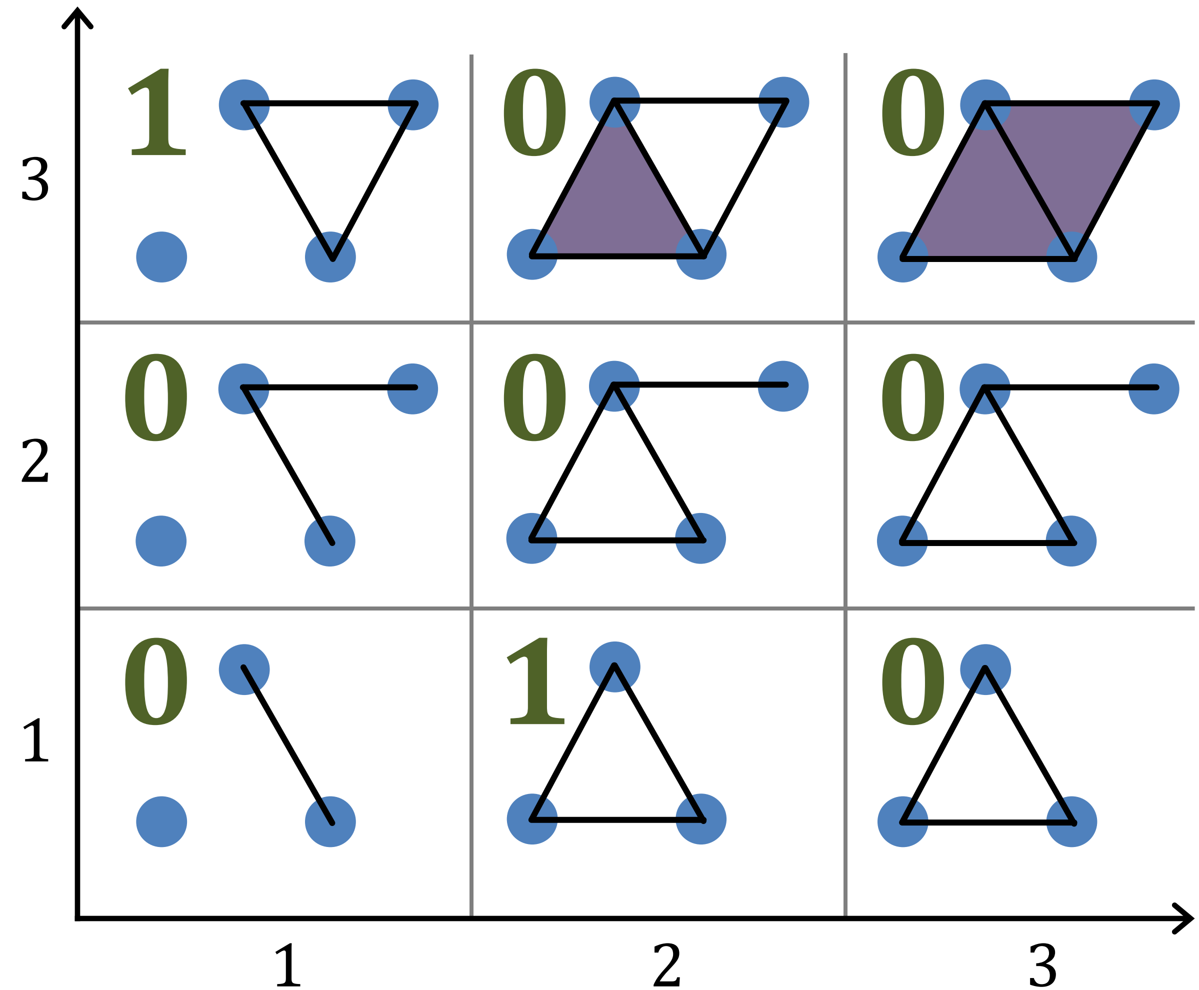
Bi-graded Betti numbers ξ_0 and ξ_1

These are functions,
$\xi_0, \xi_1 : \mathbb{N}^2 \to \mathbb{N}$

ξ_0 indicates coordinates at which homology appears

values of ξ_0 in green

Example: 1st homology (holes)



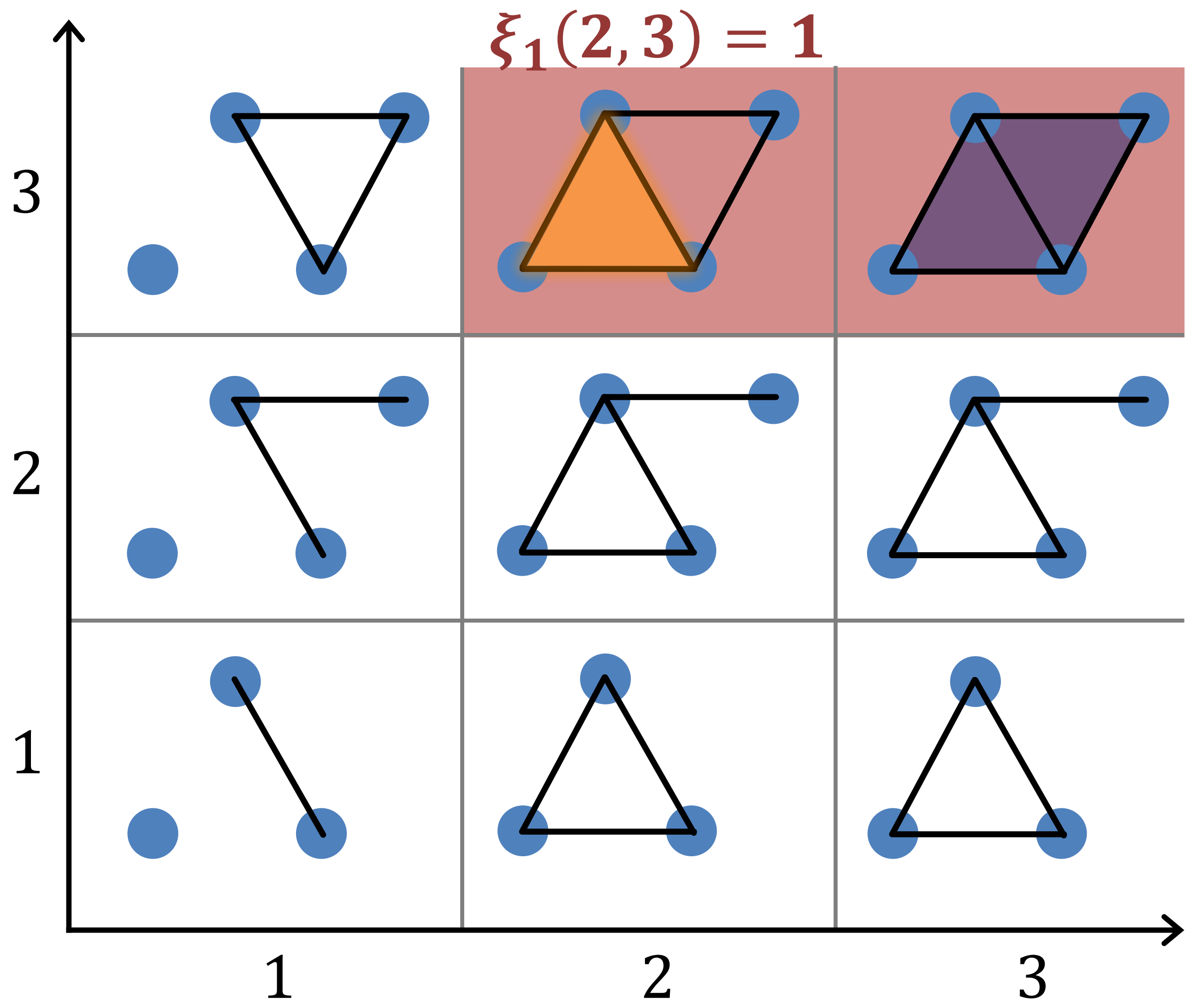
Bi-graded Betti numbers ξ_0 and ξ_1

These are functions,
$\xi_0, \xi_1 : \mathbb{N}^2 \to \mathbb{N}$

ξ_0 indicates coordinates at which homology appears

ξ_1 indicates coordinates at which homology disappears

Example: 1st homology (holes)



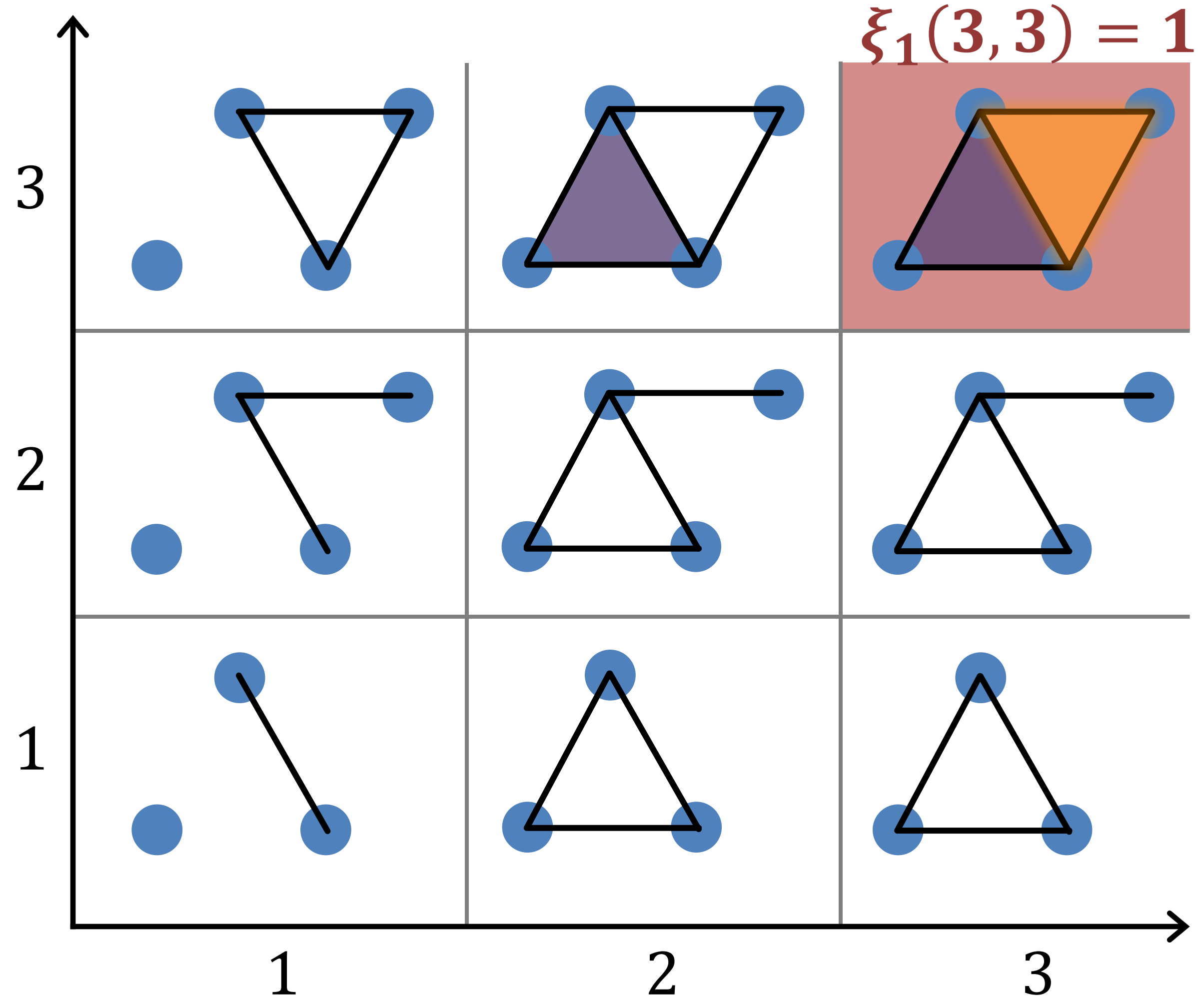
Bi-graded Betti numbers ξ_0 and ξ_1

These are functions,
$\xi_0, \xi_1 : \mathbb{N}^2 \to \mathbb{N}$

ξ_0 indicates coordinates at which homology appears

ξ_1 indicates coordinates at which homology disappears

Example: 1st homology (holes)



Bi-graded Betti numbers ξ_0 and ξ_1

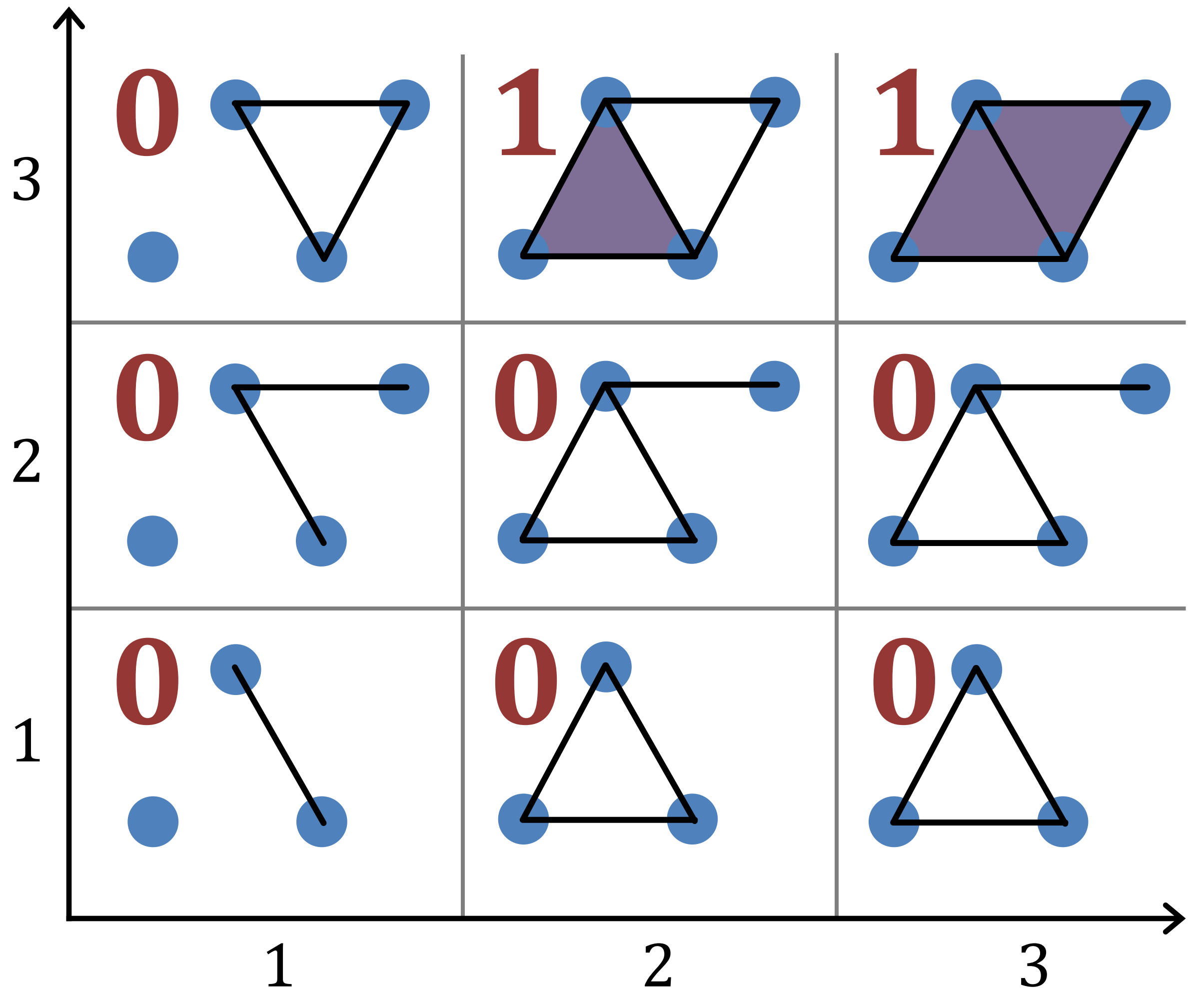
These are functions,
$\xi_0, \xi_1 : \mathbb{N}^2 \to \mathbb{N}$

ξ_0 indicates coordinates at which homology appears

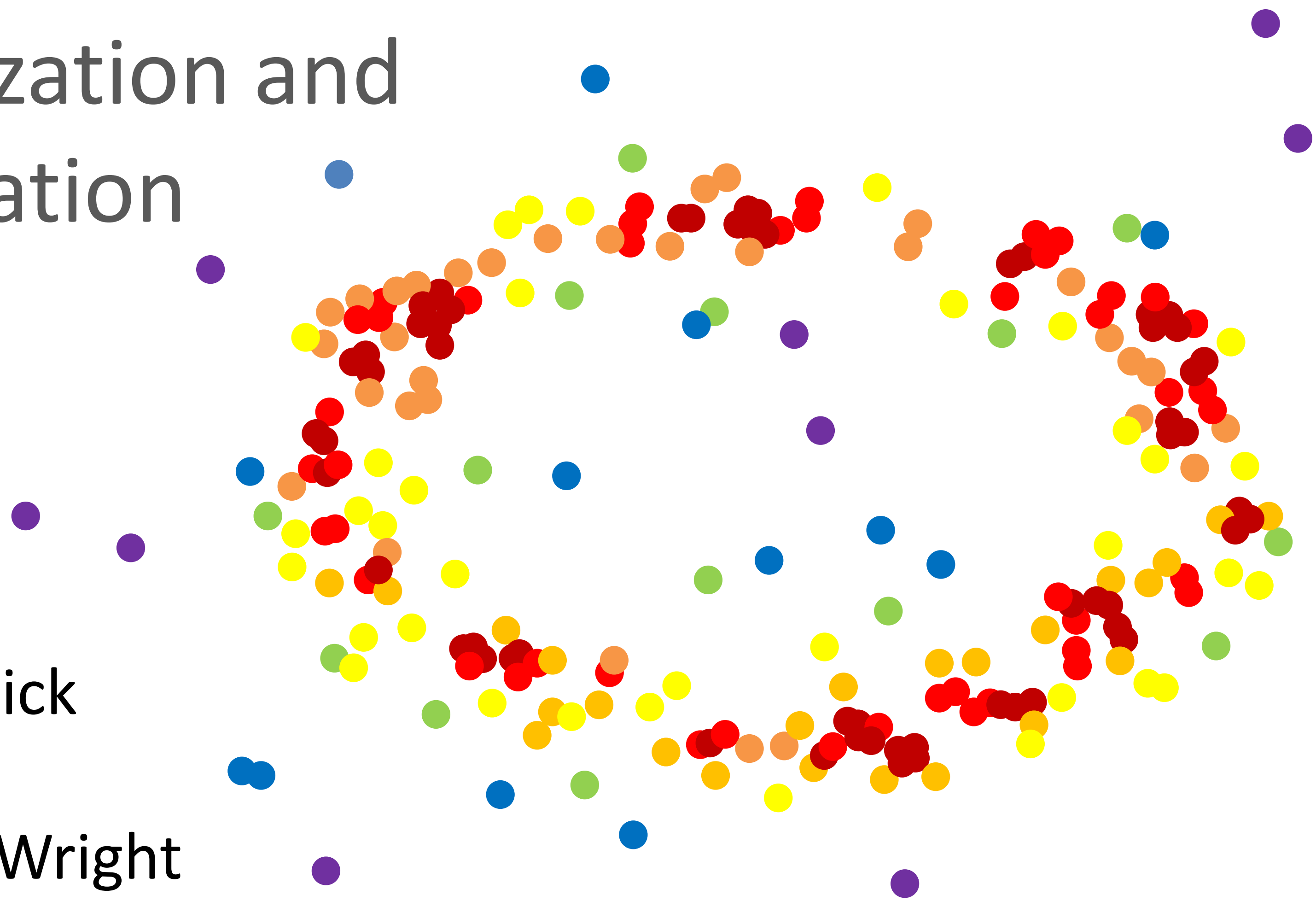
ξ_1 indicates coordinates at which homology disappears

values of ξ_1 in red

Example: 1st homology (holes)



Rank
Invariant
Visualization and
Exploration
Tool

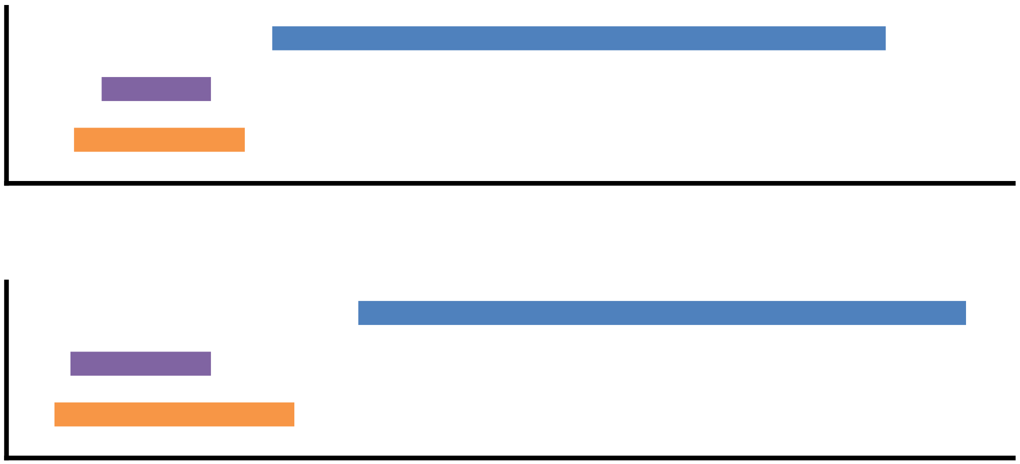


Mike Lesnick
and
Matthew Wright

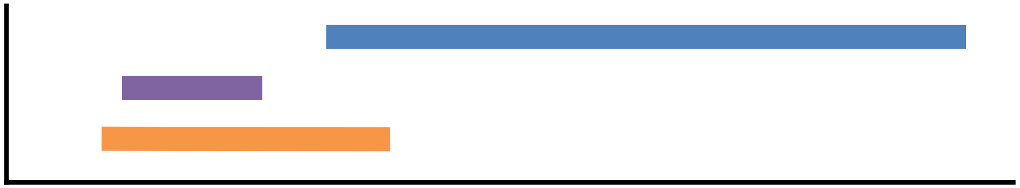
How RIVET Works

RIVET pre-computes a relatively small number of discrete barcodes, from which it draws barcodes in real-time.

Endpoints of bars appear in the same order in each of these two barcodes.

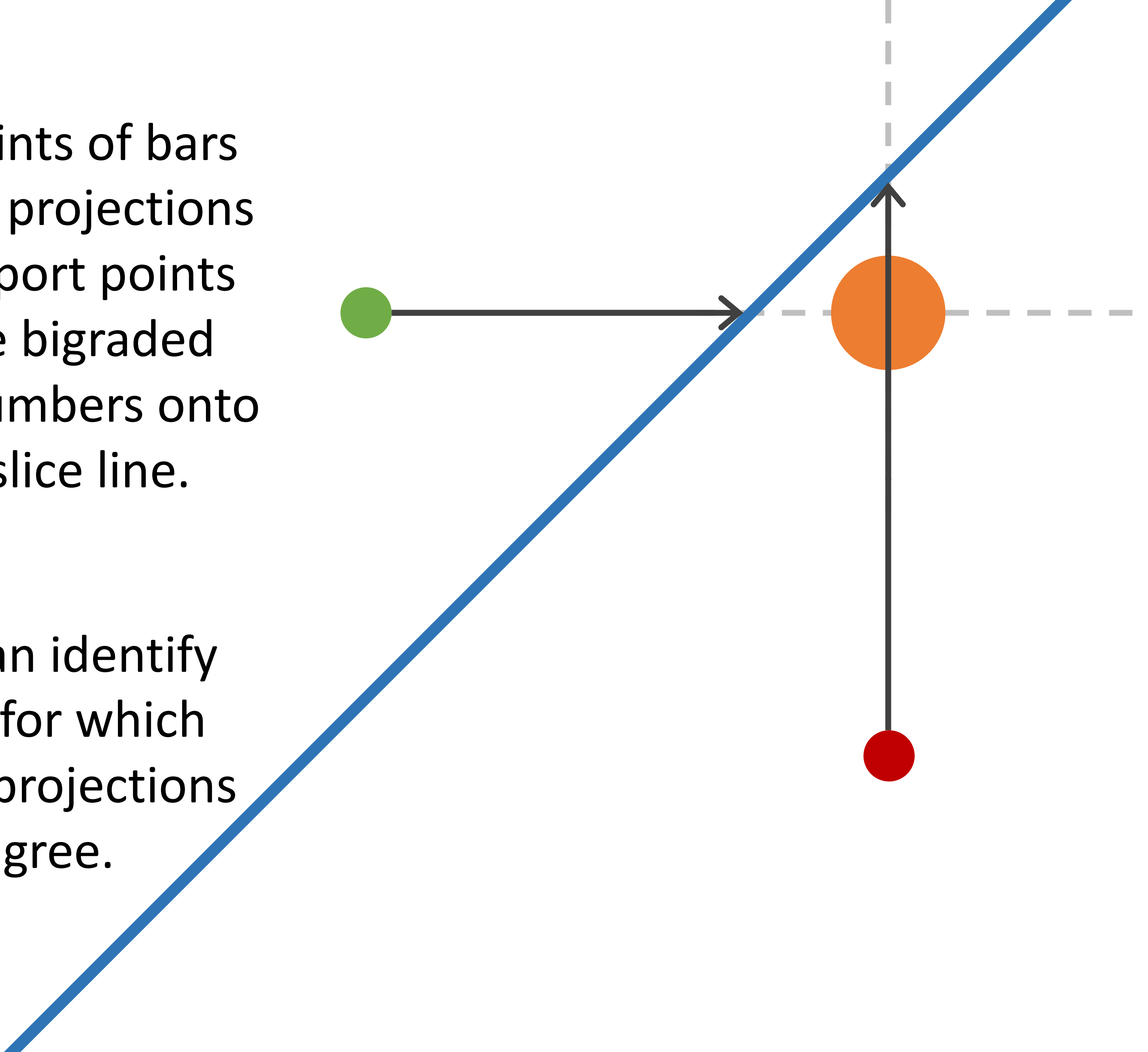


Endpoints of bars in this barcode have a different order.



Endpoints of bars are the projections of support points of the bigraded Betti numbers onto the slice line.

We can identify lines for which these projections agree.



Data Structure

At the core of RIVET is a line arrangement.

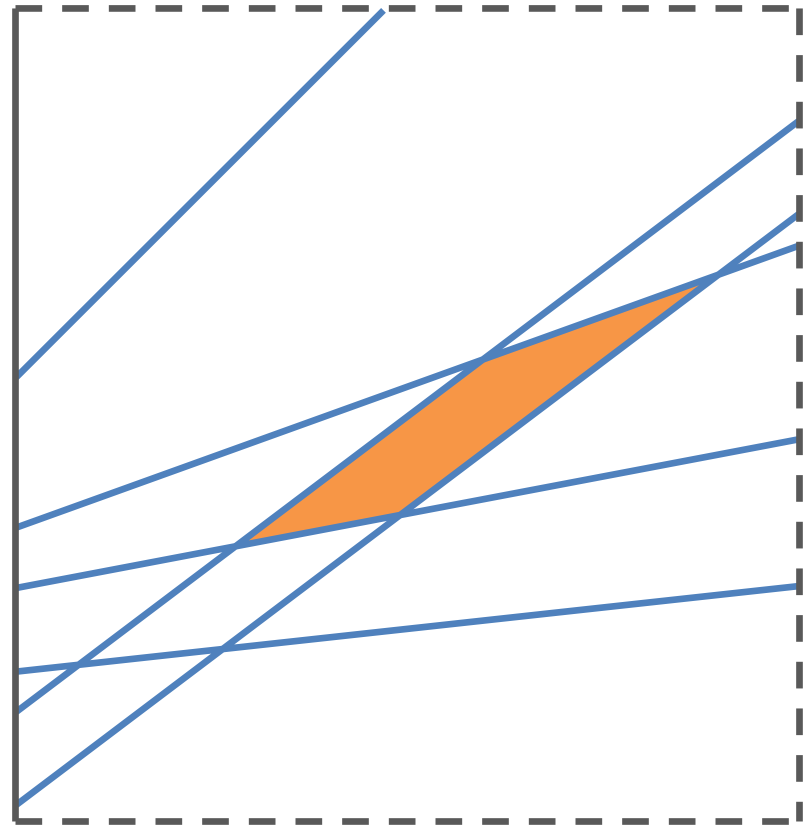
Each line corresponds to a point where projections of two support points agree.

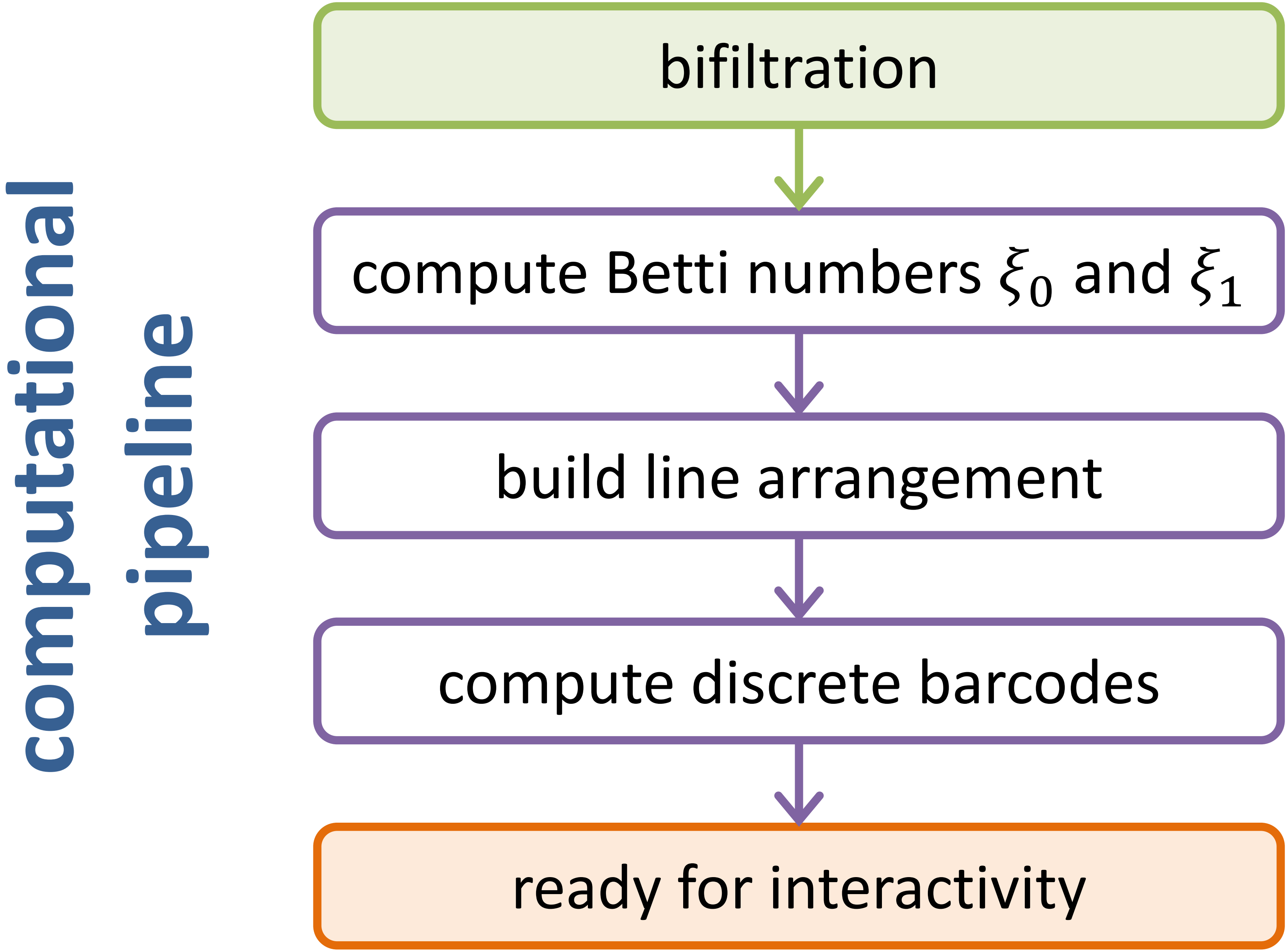
Cells correspond to families of lines with the same discrete barcode.

When the user selects a slice line, the appropriate cell is found, and its discrete barcode is re-scaled and displayed.

point-line duality:

$$(a, b) \ \leftrightarrow y = ax - b$$




computational pipeline
bifiltration
compute Betti numbers ξ_0 and ξ_1
build line arrangement
compute discrete barcodes
ready for interactivity

Performance

Suppose we are interested in i^{th} homology.

Let n be the total number of simplices of dimensions $i-1$, i, and $i+1$ in the bifiltration.

Let k be the number of multigrades.

Then the time required to compute the line arrangement and all discrete barcodes is

$$O(k^2 \log k + nk^2 + n^3).$$

Then the time required to find a cell is $O(\log k)$.

For more information:

Robert Ghrist. “Barcodes: The Persistent Topology of Data.” *Bulletin of the American Mathematical Society*. Vol. 45, no. 1, 2008, p. 61-75.

Gunnar Carlsson and Afra Zomorodian. “The Theory of Multidimensional Persistence.” *Discrete and Computational Geometry*. Vol. 42, 2009, p. 71-93.

Michael Lesnick and Matthew Wright. “Efficient Representation and Visualization of 2-D Persistent Homology.” *in preparation*.