Persistent homology & algebraic foundations

Mikael Vejdemo-Johansson
CUNY College of Staten Island

Outline

- Topological Data Analysis and Persistent Homology
- The two persistences
- Algebraic foundations: the more we know…
- Sheaves

Outline

- Topological Data Analysis and Persistent Homology
- The two persistences
- Algebraic foundations: the more we know…
- Sheaves

Topological Data Analysis

Fundamental idea: use topology to understand data.

Data has shape. Shape carries meaning.

Key idea: how to create meaningful topology from discrete observations.

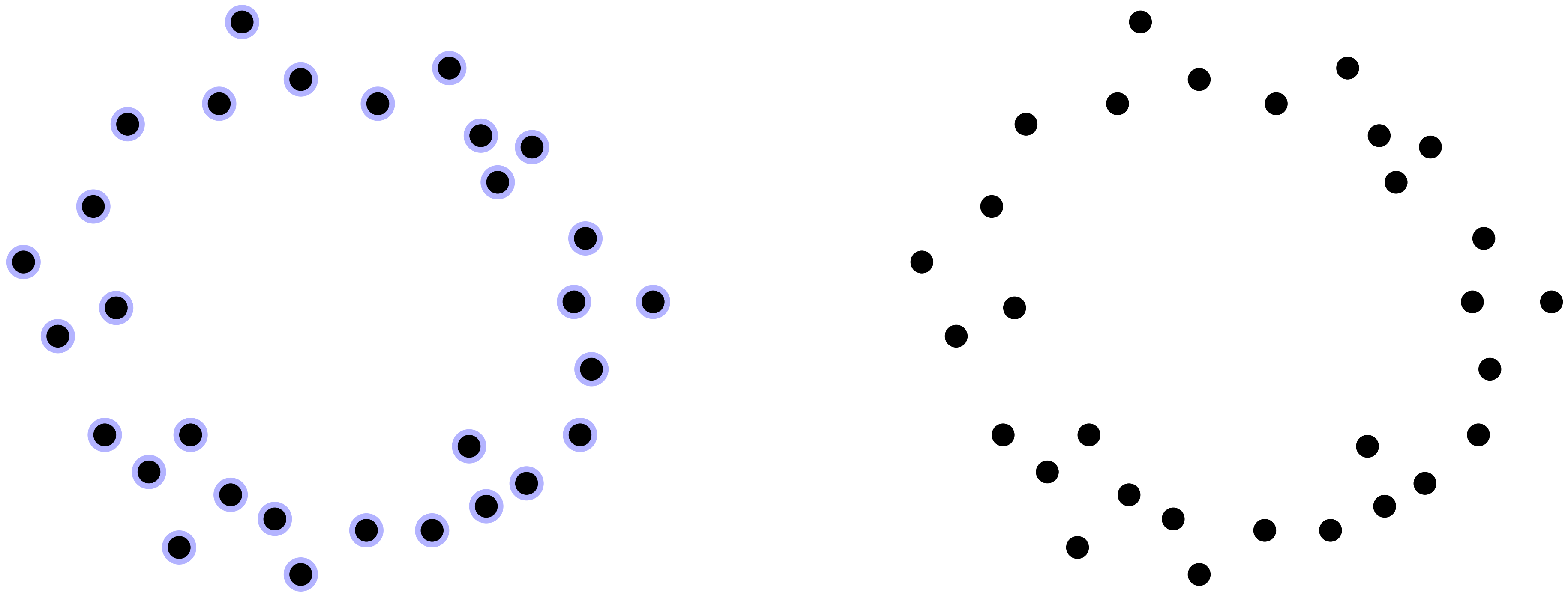
Several approaches

- Mapper / Reeb graphs / Reeb spaces
  Create topological models of data by breaking up fibres of a map

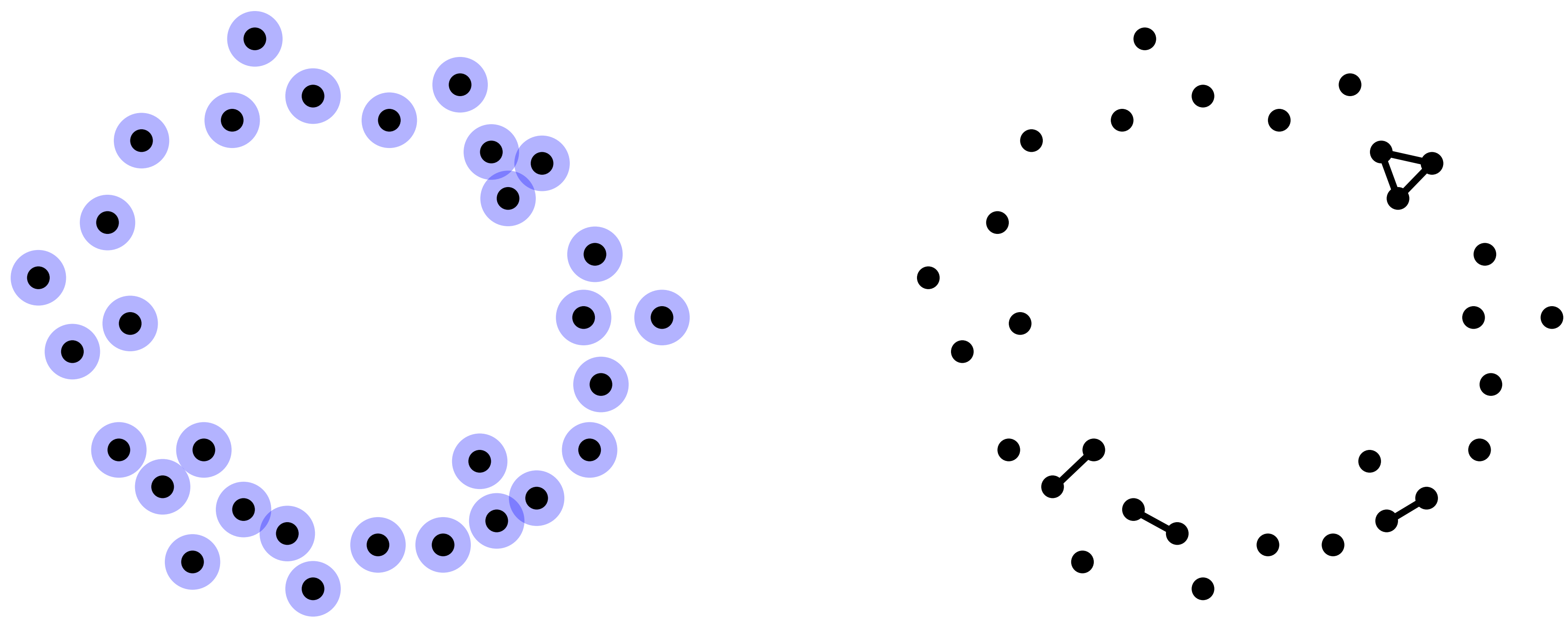
- Persistent (co)homology
  Sweep across parametrized family of topological spaces, summarize changes in their (co)homology

  Works through creating simplicial complexes that carry shape
  Vietoris-Rips; Čech; α-shapes; witness complexes…

Čech complex

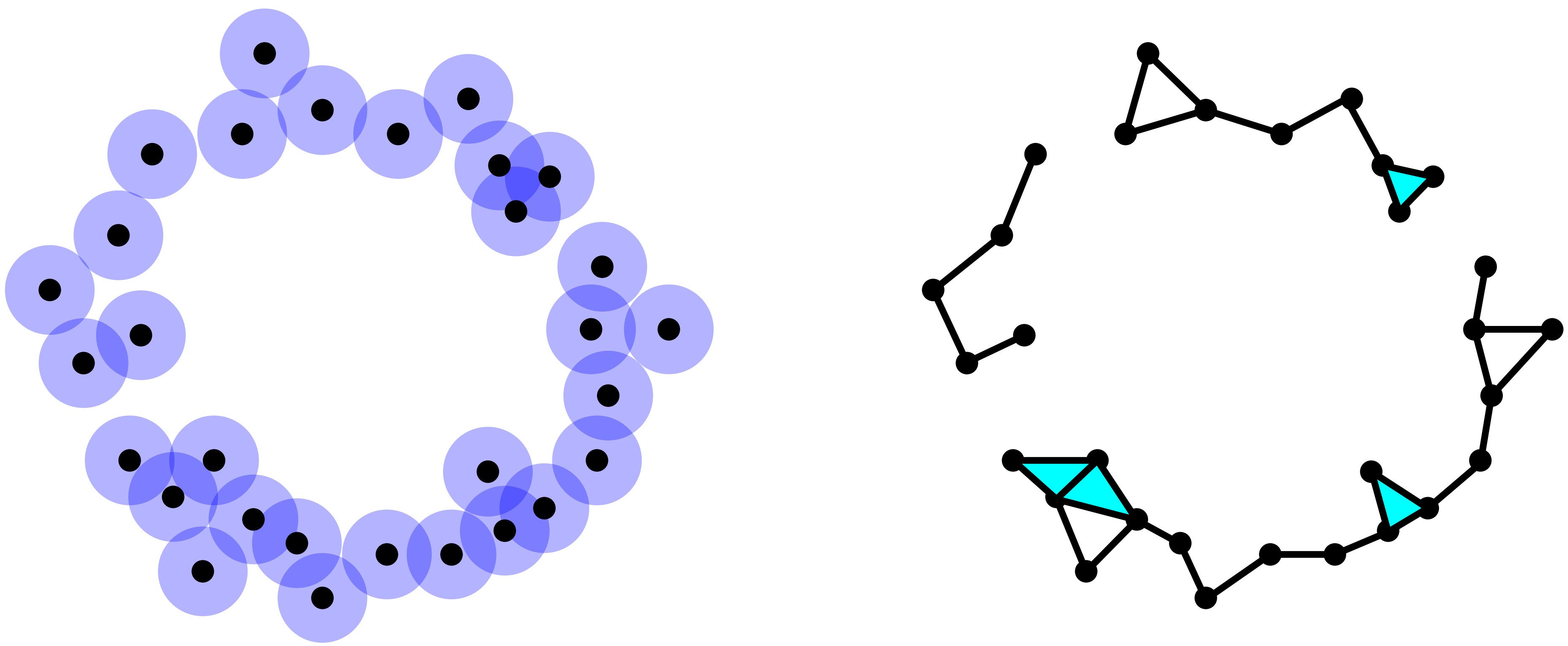


Čech complex



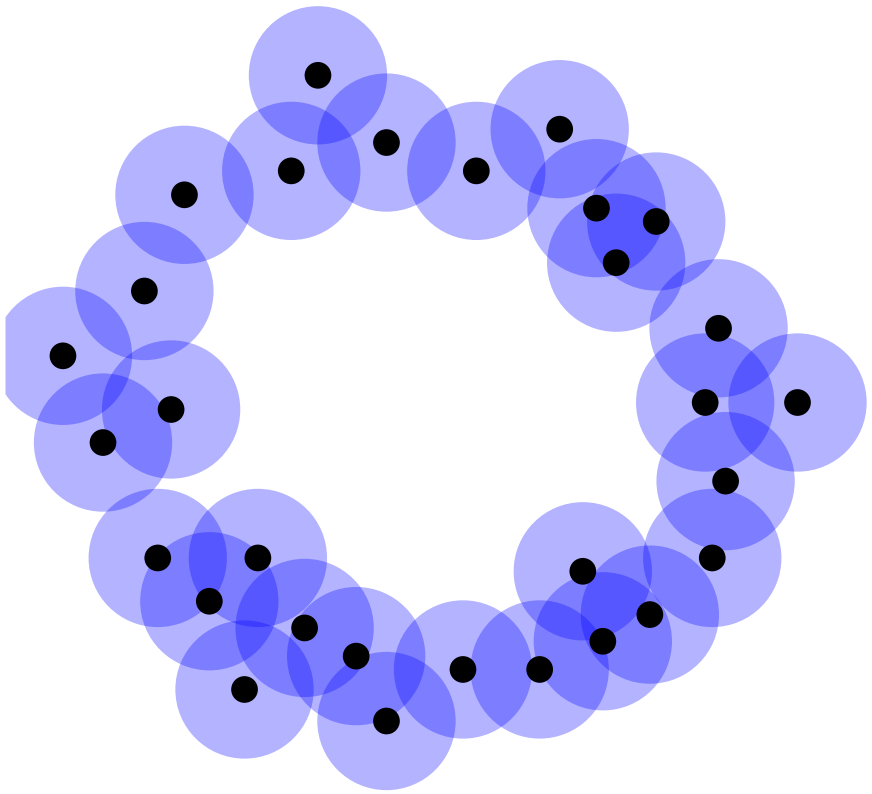
1 loop

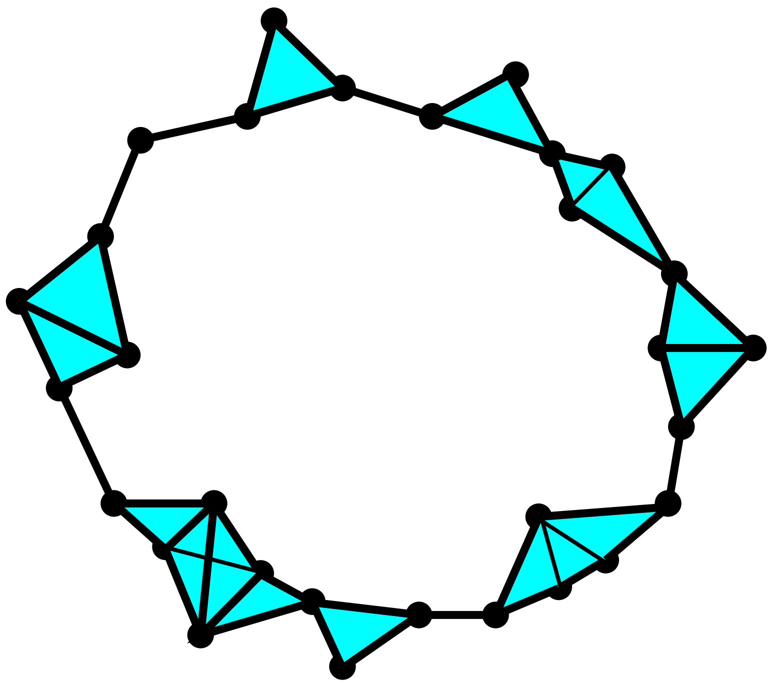
Čech complex



3 loops

Čech complex





1 loop

Outline

- Topological Data Analysis and Persistent Homology
- The two persistences
- Algebraic foundations: the more we know…
- Sheaves

Two approaches to persistent homology

Two approaches to persistent homology

Persistent homology is fundamentally about…

Two approaches to persistent homology

Persistent homology is fundamentally about…

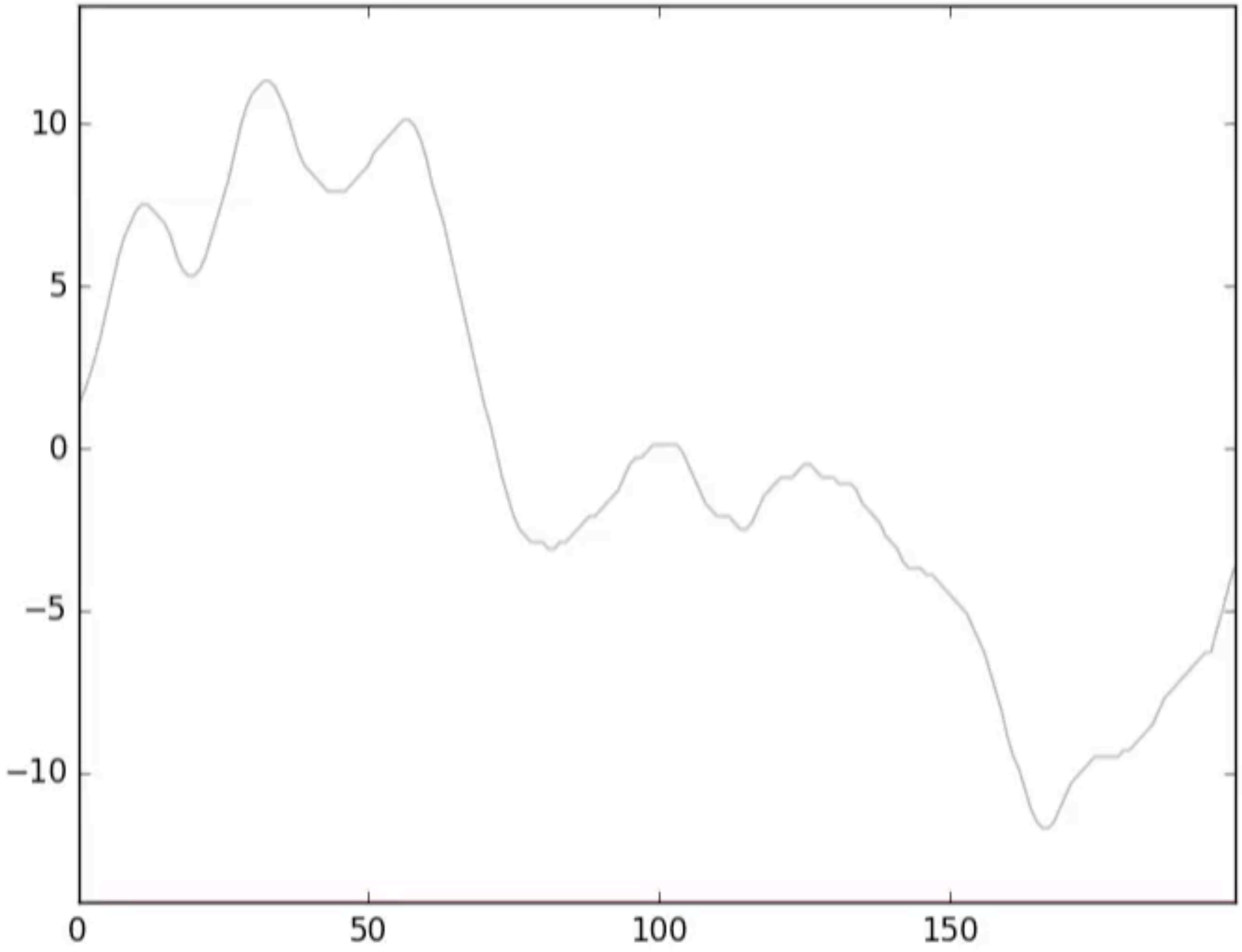
- Sublevel sets of some function on some manifold

Two approaches to persistent homology

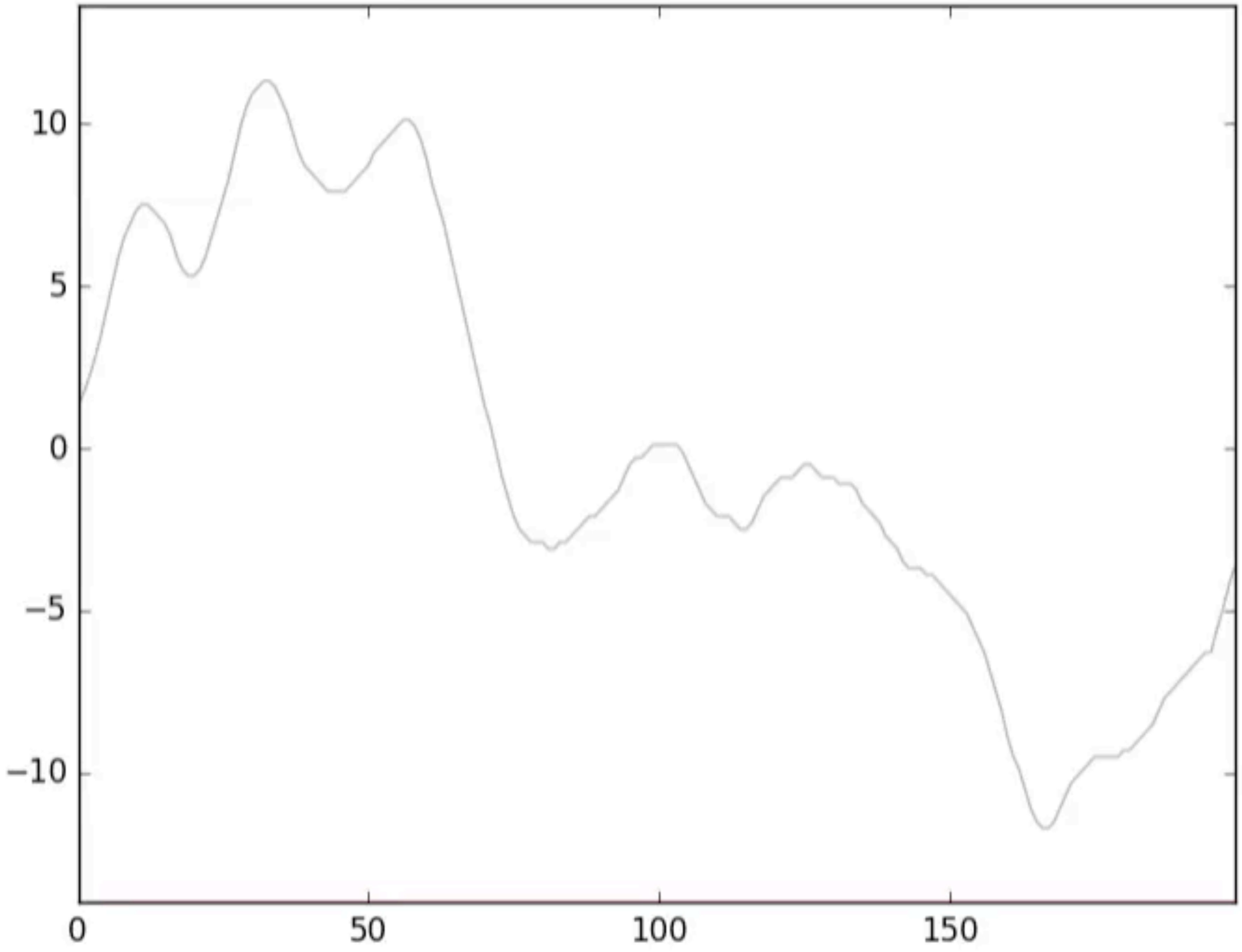
Persistent homology is fundamentally about…

- Sublevel sets of some function on some manifold
- Filtered (parametrized) families of topological spaces

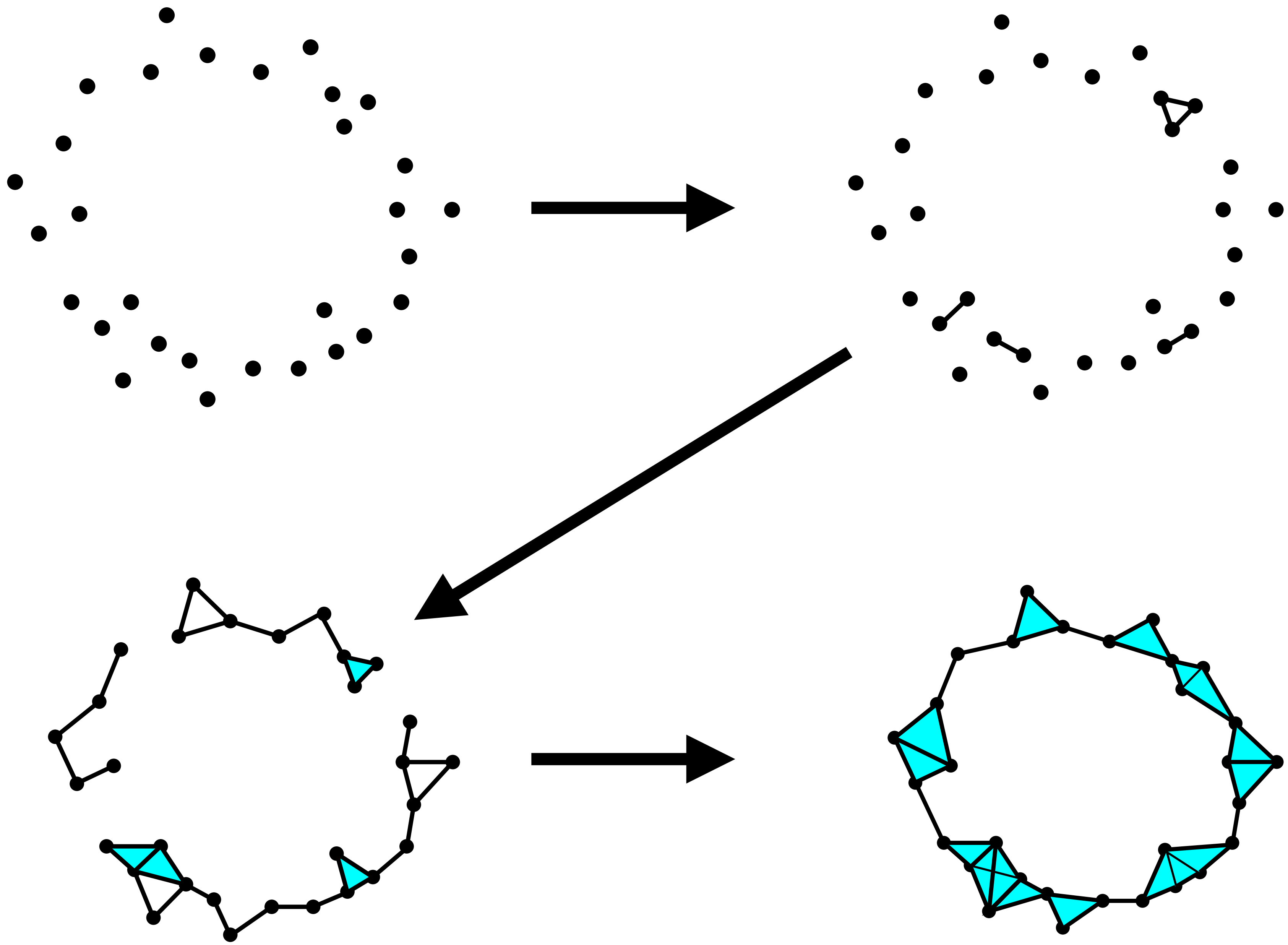
Sublevel sets



Sublevel sets



Filtered spaces



Perspectives are (kinda) equivalent

Vietoris-Rips and Čech capture sublevel sets of the **distance-to-data** function

Manifold can be discretized, produces filtered topological spaces.

Outline

- Topological Data Analysis and Persistent Homology
- The two persistences
- Algebraic foundations: the more we know…
- Sheaves

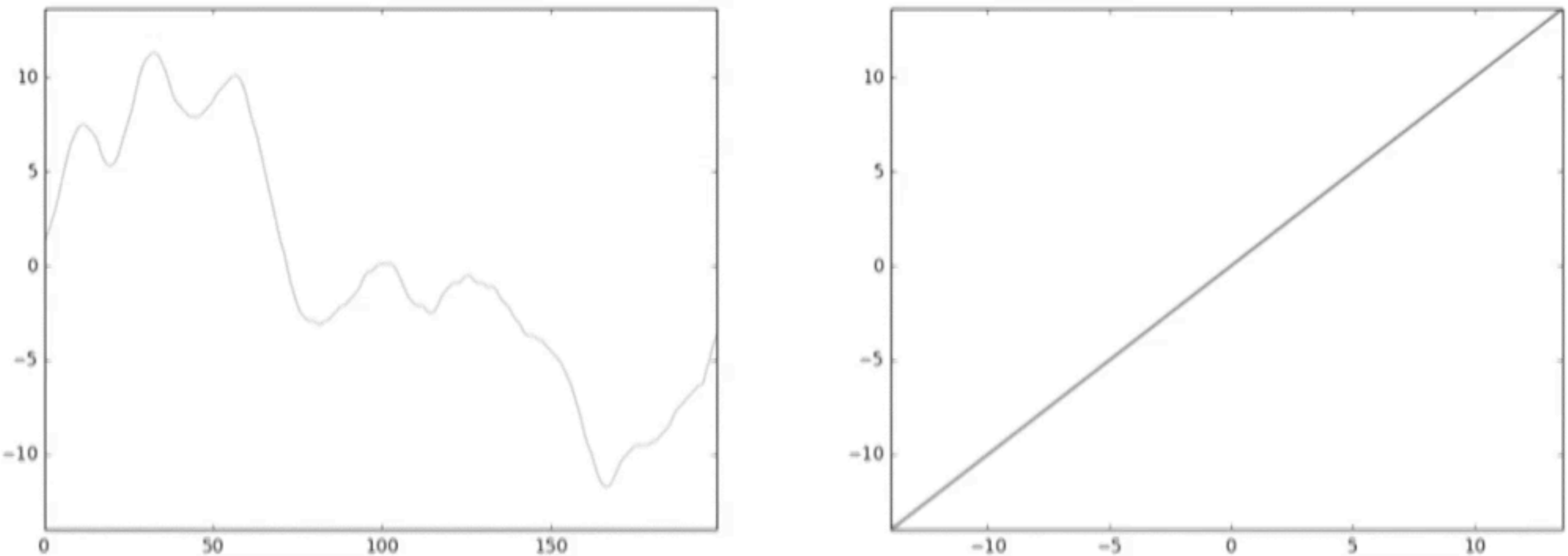
Edelsbrunner, Letscher and Zomorodian (2000; 2002)

Topological persistence and simplification

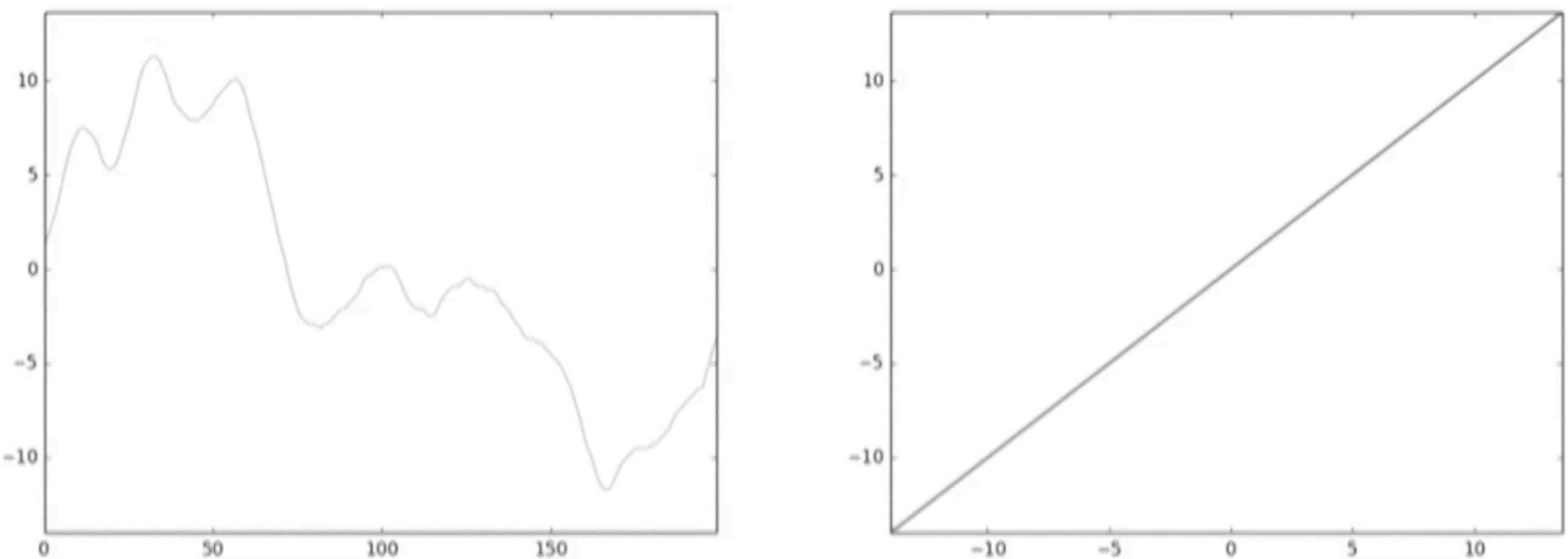
Defines **persistent homology** and provides an algorithm for computing with coefficients in $\mathbb{Z}_2$

With each span of values for a function on a manifold is associated the group of homological features that exist in all sublevel sets

We know what to capture



We know what to capture



Carlsson and Zomorodian (2005)

Computing Persistent Homology

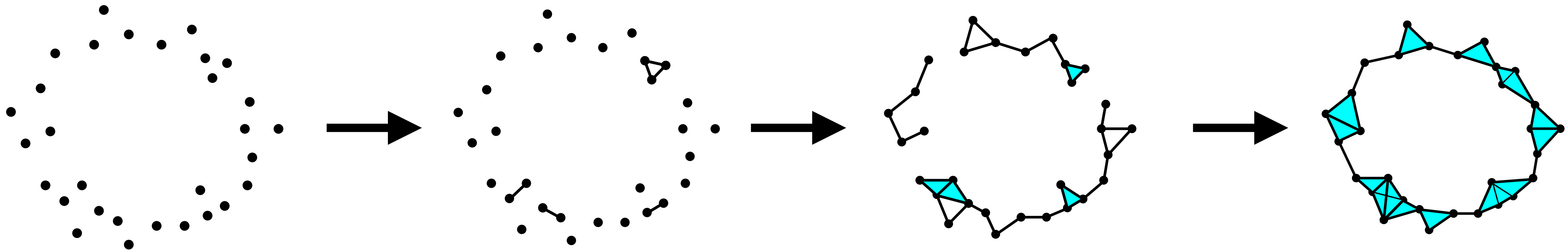
Observes there is a functor in play in the constructions in **Edelsbrunner, Letscher and Zomorodian (2000)**.

Suggests that the structure to be captured can be described by a module over $\mathbb{Z}_2[t]$.

Hence, arbitrary field coefficients $\Bbbk$ work using $\Bbbk[t]$.

Functoriality

$\mathbb{X}$



$H_1\mathbb{X}$ $\quad 0 \longrightarrow \mathbb{k} \longrightarrow \mathbb{k}^3 \longrightarrow \mathbb{k}$

$\mathbb{k}[t]$ module

$H_1\mathbb{X}$ $0 \longrightarrow \mathbb{k} \longrightarrow \mathbb{k}^3 \longrightarrow \mathbb{k}$

$\mathbb{k}[t]$ module

$H_1\mathbb{X}$ $\quad 0 \longrightarrow \mathbb{k} \longrightarrow \mathbb{k}^3 \longrightarrow \mathbb{k}$

Take direct sum $\quad H_1\mathbb{X} = \mathbb{k} \oplus \mathbb{k}^3 \oplus \mathbb{k}$

$\mathbb{k}[t]$ module

$H_1\mathbb{X}$ $0 \longrightarrow \mathbb{k} \longrightarrow \mathbb{k}^3 \longrightarrow \mathbb{k}$

Take direct sum $H_1\mathbb{X} = \mathbb{k} \oplus \mathbb{k}^3 \oplus \mathbb{k}$

Put each summand in module degree corresponding to position in sequence

$\mathbb{k}[t]$ module

$$H_1\mathbb{X} \quad 0 \xrightarrow{\cdot t} \mathbb{k} \xrightarrow{\cdot t} \mathbb{k}^3 \xrightarrow{\cdot t} \mathbb{k}$$

Take direct sum $H_1\mathbb{X} = \mathbb{k} \oplus \mathbb{k}^3 \oplus \mathbb{k}$

Put each summand in module degree corresponding to position in sequence

Module structure by defining $\cdot t$ to be the induced map from the functor action

Carlsson and Zomorodian (2009)

The Theory of Multidimensional Persistence

Also: **Carlsson, Singh and Zomorodian (2009)** and **Biasotti, Cerri, Frosini, Giorgi and Landi (2008)**

Suggests that tracking multiple parameters corresponds to working over $\mathbb{k}[t_1, t_2, \ldots, t_d]$

We are still today searching for a good invariant to describe the resulting modules.

Carlsson and de Silva (2008)

Zigzag Persistence

Notices most use cases do not use the structure from a $\mathbb{k}[t]$-module: finite diagram most common.

Recalls **Gabriel (1975)**, classifying tame representations of quivers

Recognizes persistence as a representation of the quiver A_n

Suggests strict filtration not necessary: arrows can go both ways

Burghelea and Dey (2011)

Persistence for circle valued maps

Parameters on S_1 instead of in $\mathbb{R}$.
Connect level-sets.

Discretizes to working with cyclic quiver representations: no longer tame, requires Jordan blocks for an invariant.

Chambers (2014)

Persistent homology over a DAG

Uses a potentially branching directed acyclic graph as underlying parameter space for persistent homology.

Outline

- Topological Data Analysis and Persistent Homology
- The two persistences
- Algebraic foundations: the more we know…
- Sheaves

Curry (2013)

Sheaves, cosheaves and applications

Observes that persistent (co)homology has a natural structure as cosheaves over a discretization of $\mathbb{R}$.

Why sheaves?

- **Edelsbrunner, Letscher and Zomorodian**
  Restricted to $\mathbb{Z}_2$ coefficients.

- $\mathbb{k}$[t]-modules
  Requires discretization

- Quiver representations
  Requires finite discretization

- Multidimensional, circles and DAGs
  Indicate more shapes are interesting

Why sheaves?

Sheaves allow for very high generality in the **shape** of persistence:
picking a base space fixes the shape

The original definition, producing a vector space for any specified interval [b,d], echoes the definition of a (pre)sheaf.

How sheaves?

- **Curry** makes a case for cosheaves of vector spaces
- **Vejdemo-Johansson, Škraba and Pita-Costa** (in preparation) make the case for topoi:
  by introducing the parameter dependency at the set theory level persistence emerges as semi-simplicial homology over this new set theory

What sheaves?

- Patel argues for using sheaves over the classical (Borel) topology on $\mathbb{R}$ for persistence
  This invites wild representation theories: removing the hope for discrete and field-agnostic invariants
- VJ-Š-PC argue for building a new “topology” on $\mathbb{R}$ using only connected intervals

The Persistence Topos

Heyting Algebras are partial orders similar enough to topologies for the sheaf axioms

For persistence, we wish to design a Heyting algebra with:

- Elements: intervals [b,d] in $\mathbb{R}$
- Meet (intersection): usual intersection
- Join (union): covering interval

The Persistence Topos

Naïve approach fails: the structure is not distributive
Key issue: empty intersections

Solution: orient all intervals, introduce «negative» intervals for empty intersections.

$[a,b] \cup [c,d] = [\min(a,c), \max(b,d)]$
$[a,b] \cap [c,d] = [\max(a,c), \min(b,d)]$

The Persistence Topos

These oriented intervals occur in
Cohen-Steiner, Edelsbrunner and Harer (2009)
Extending Persistence using Poincaré and Lefschetz Duality

Negative intervals can usefully represent

- Relative homology: $H_k(\mathbb{X}, \mathbb{X} \setminus \mathbb{X}_n)$
- Existence of features:
  the set of all homology classes with any presence

Other shapes

- We (VJ-Š-PC) also have Heyting algebras developed for zigzag, 1-critical multidimensional, 1-critical multidimensional zigzag, convex footprint multidimensional, circular persistent homologies.

Thank you!