# VRDI TDA Breakout Session: Introduction to Persistent Homology

Moon Duchin, Tom Needham, Thomas Weighill

Voting Rights Data Institute
June 25, 2019

# Concepts from Topology

Topology is a field of math which studies geometrical objects up to loose notions of "equivalence".

Each such object is called a (topological) space, denoted $X$.

Roughly, spaces are equivalent if one can be deformed into the other via stretching and bending, without creating or closing holes.

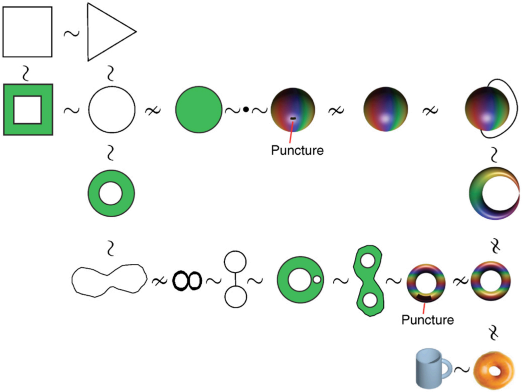



Figure: Homotopy equivalence, from Singh et. al. 2008.

# Concepts from Topology

Algebraic topology is a subfield of topology where one computes invariants of a space which distinguish it from other spaces.

To each space $X$, we can associate a vector space $H_k(X)$ called the $k$th homology vector space of $X$.

Its dimension $\beta_k(X)$ is called the $k$th Betti number of $X$.

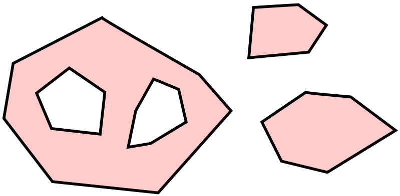
**Punchline**

If $X$ and $Y$ have $\beta_k(X) \neq \beta_k(Y)$ for some $k$, then $X$ and $Y$ are not equivalent!

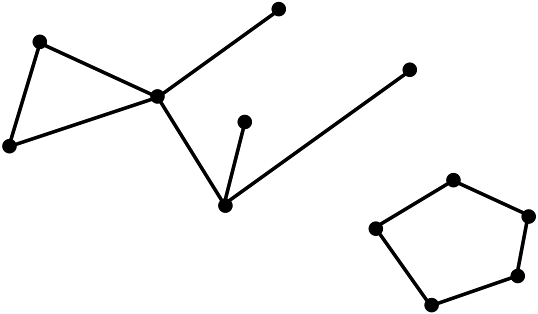
# Concepts from Topology

The Betti number $\beta_k(X)$ counts "$k$-dimensional holes" in $X$:

- $\beta_0(X)$ — # of connected pieces
- $\beta_1(X)$ — # of unfilled loops
- $\beta_2(X)$ — # of unfilled "voids" (interior of a basketball)
- $\beta_k(X)$ — well-defined concept we can't visualize for $k > 2$



# Examples of Betti Numbers



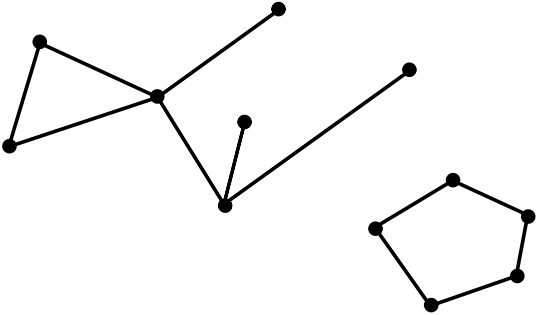
$\beta_0 =$

$\beta_1 =$

$\beta_2 =$

Figure: Disconnected graph.

# Examples of Betti Numbers



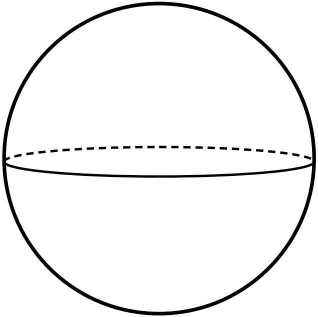
$$\beta_0 = 2$$

$$\beta_1 = 2$$

$$\beta_2 = 0$$

Figure: Disconnected graph.

# Examples of Betti Numbers



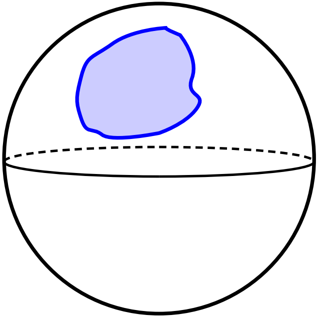
$\beta_0 =$

$\beta_1 =$

$\beta_2 =$

Figure: Surface of a sphere.

# Examples of Betti Numbers



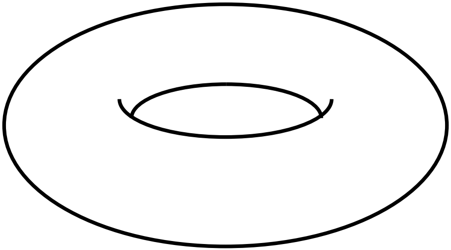
$\beta_0 = 1$

$\beta_1 = 0$

$\beta_2 = 1$

Figure: Any loop on the sphere can be filled in with a disk.

# Examples of Betti Numbers



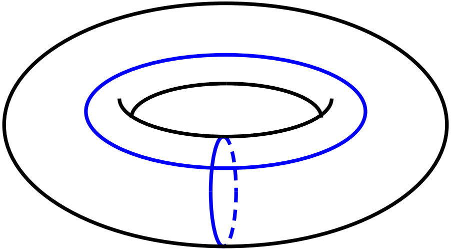
$\beta_0 =$

$\beta_1 =$

$\beta_2 =$

Figure: Torus (surface of a donut).

# Examples of Betti Numbers



$\beta_0 = 1$

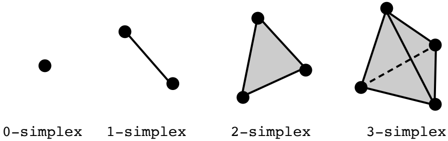
$\beta_1 = 2$

$\beta_2 = 1$

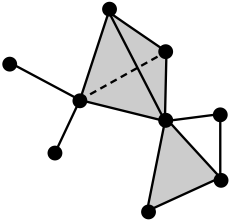
Figure: Blue loops can't be filled by disks that stay in the surface.

# Simplicial Homology

A *k*-simplex is a *k*-dimensional generalization of a triangle.




A simplicial complex is a space obtained by gluing together simplices along lower-dimensional faces.

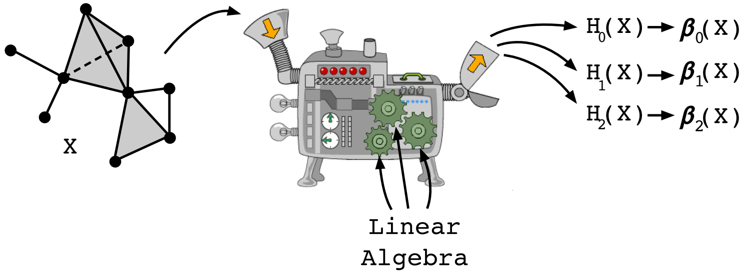


# Simplicial Homology

Computing homology/Betti numbers of simplicial complexes is easy!

Boils down to linear algebra:

- Gluing process is described by linear maps.
- Homology is computed from kernels and images of these maps.



# How Does This Apply to Data?

The most common type of data is a point cloud — a set of vectors $X = \{\vec{x}_1, \ldots, \vec{x}_N\}$, each $\vec{x}_j \in \mathbb{R}^d$.

This is a simplicial complex with only 0-dimensional simplices and no interesting topology; i.e.,

$$\beta_0 = N, \quad \beta_1, \beta_2, \ldots = 0.$$

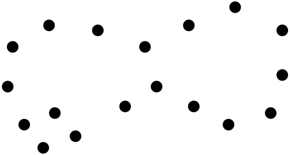
### Idea

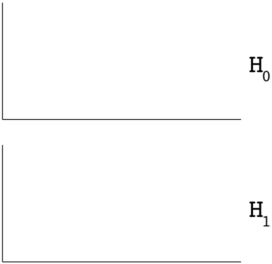
Construct a family of simplicial complexes.

Keep track of “births" and “deaths" of topological features (holes of various dimensions) along the family.

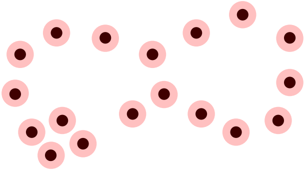
This leads to the main tool in TDA: persistent homology.

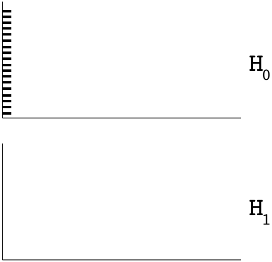
# Persistent Homology - An Example



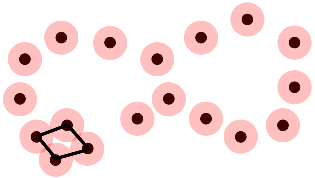


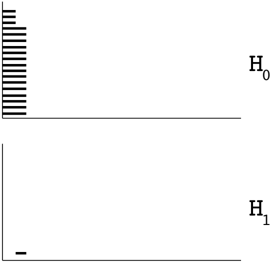
# Persistent Homology - An Example



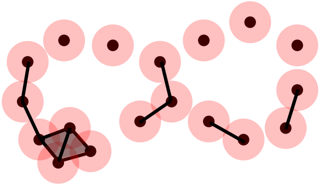


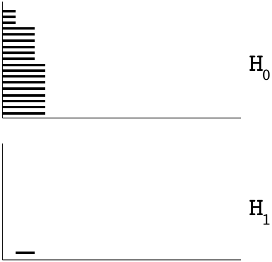
# Persistent Homology - An Example



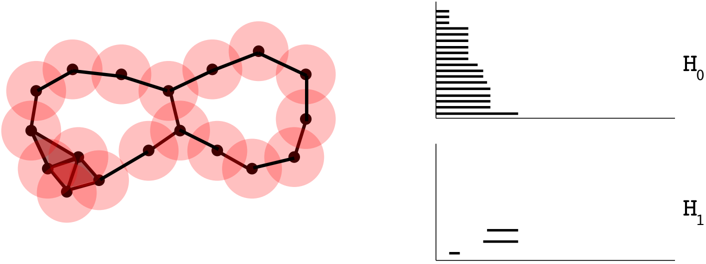


# Persistent Homology - An Example

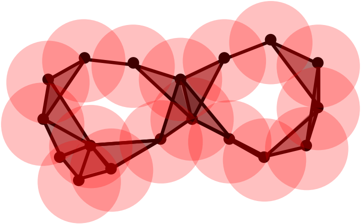


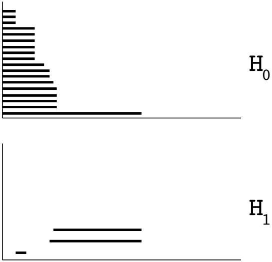


# Persistent Homology - An Example

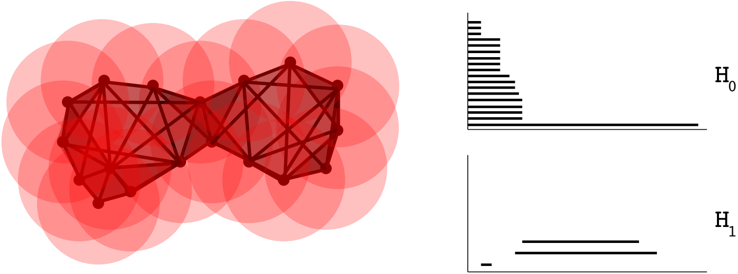


# Persistent Homology - An Example



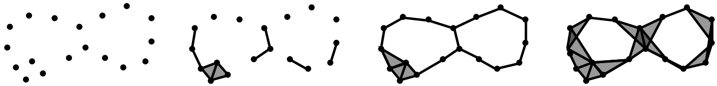


# Persistent Homology - An Example



# Terminology

Such a family is called a filtered simplicial complex.



There are many techniques for creating them.

The previous example is called a Čech complex.

In computational examples, we'll use a related construction called a Vietoris-Rips complex.

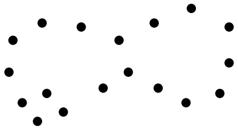
# Vietoris-Rips Complex

Let $X = \{\vec{x}_1, \ldots, \vec{x}_N\} \subset \mathbb{R}^d$. The Vietoris-Rips complex of $X$ is the filtered simplicial complex given at time $r$ by:

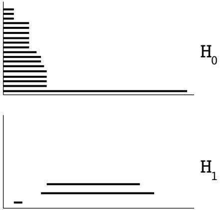
- Vertex set at time $r$ is $X$
- A subset $S = \{\vec{x}_{i_0}, \vec{x}_{i_2}, \ldots, \vec{x}_{i_k}\}$ forms a $k$-simplex at time $r$ if and only if
$$\|\vec{x} - \vec{y}\| \leq r \;\; \forall \;\; \vec{x}, \vec{y} \in S.$$

# Terminology

The topological signatures we get from persistent homology are called barcodes.



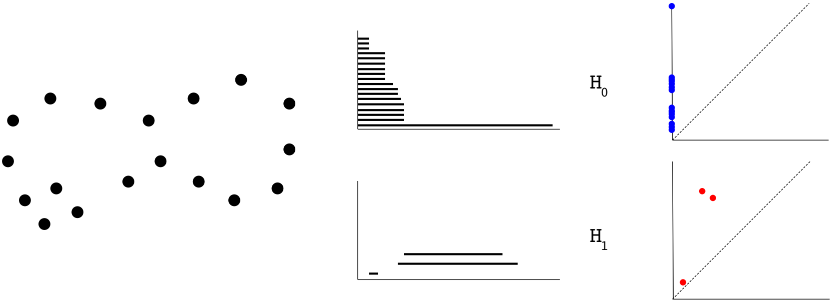
Dataset X




Barcodes for X

# Terminology

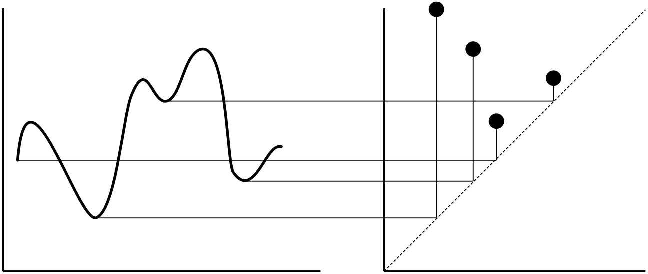
We can record the "birth time" and "death time" of each topological feature to get a persistence diagram.




Dataset X    Barcodes for X    Persistence Diagrams for X

# Level Set Filtrations

Another common filtration is by sublevel sets of the graph of a function.



# TDA Workflow