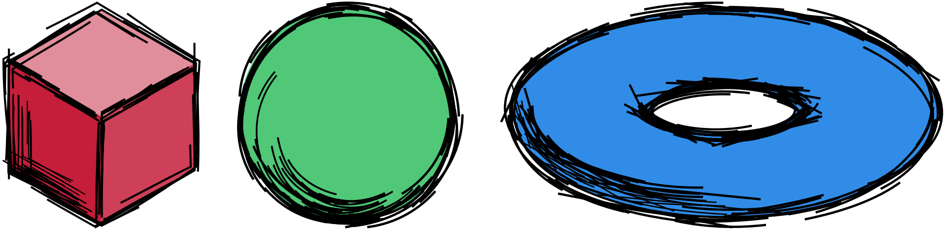
A Primer in Topological Data Analysis
Lecture 1: Computational Topology & Persistent Homology

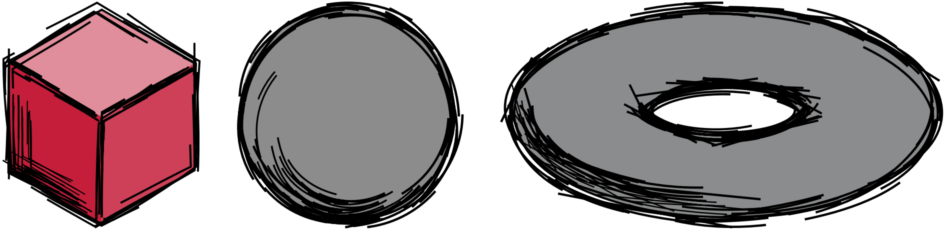
Bastian Rieck

Pseudomanifold

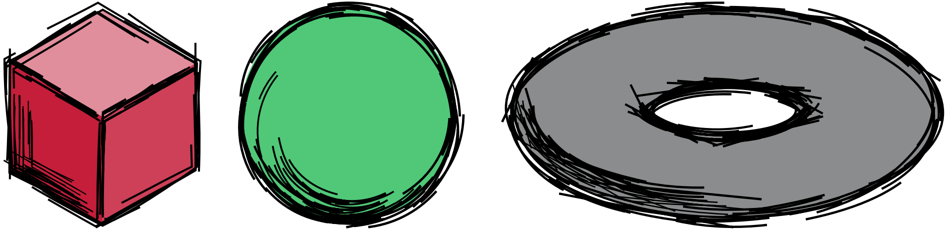
What is computational topology?



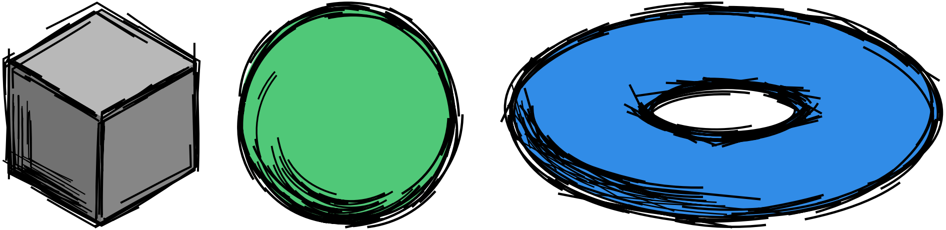
What is computational topology?



What is computational topology?



What is computational topology?



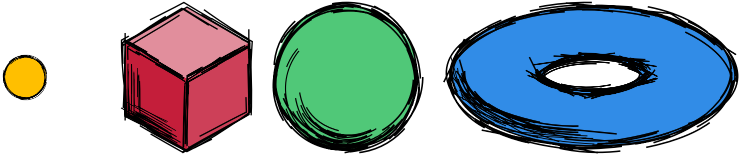
Which qualities of the sphere make it *different* from the torus?

Betti numbers

The d^{th} Betti number counts the number of d-dimensional holes. It can be used to distinguish between spaces.

β_0	Connected components
β_1	Tunnels
β_2	Voids

Space	β_0	β_1	β_2
Point	1	0	0
Cube	1	0	1
Sphere	1	0	1
Torus	1	2	1



Agenda

1. Use *simplicial complex* to model a space.
2. Define boundary operators and maps.
3. Calculate Betti numbers using matrix reduction.

Simplicial complexes

Definition

We call a non-empty family of sets K with a collection of non-empty subsets S an *abstract simplicial complex* if:

1. $\{v\} \in S$ for all $v \in \mathrm{K}$.
2. If $\sigma \in S$ and $\tau \subseteq \sigma$, then $\tau \in \mathrm{K}$.

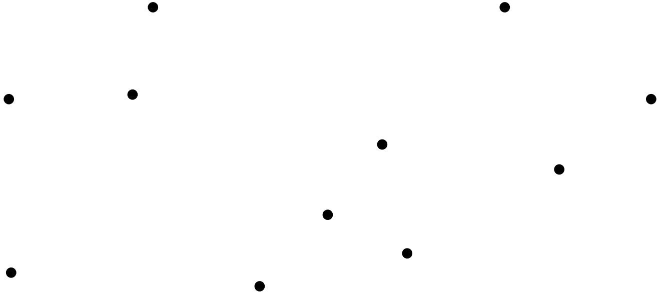
Terminology

The elements of a simplicial complex K are called *simplices*. A k-simplex consists of $k+1$ vertices.

Simplicial complexes

Example

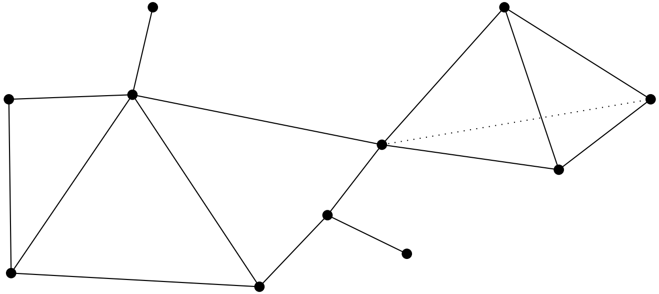
Simplicial complexes can be decomposed into their skeletons, which only contain simplices of a certain dimension.



Simplicial complexes

Example

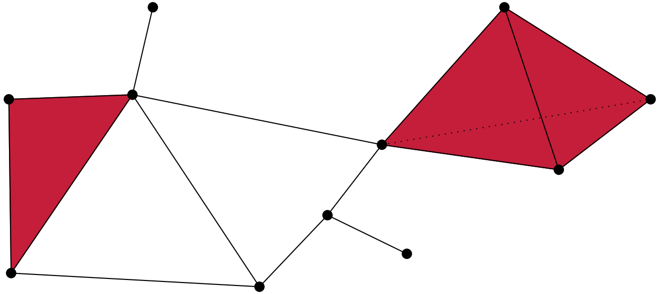
Simplicial complexes can be decomposed into their skeletons, which only contain simplices of a certain dimension.



Simplicial complexes

Example

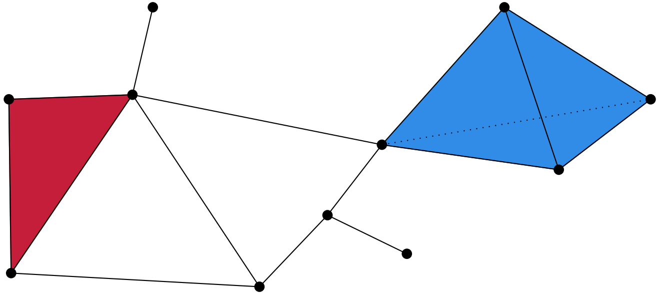
Simplicial complexes can be decomposed into their skeletons, which only contain simplices of a certain dimension.



Simplicial complexes

Example

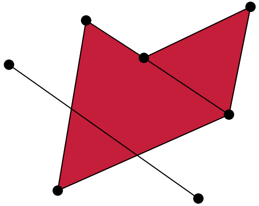
Simplicial complexes can be decomposed into their skeletons, which only contain simplices of a certain dimension.



Simplicial complexes

Non-example

This is *not* a simplicial complex because some higher-dimensional simplices do not intersect in a lower-dimensional one!



Simplicial complexes

More examples

- Graphs can be considered (low-dimensional) simplicial complexes.
- Simplicial complexes can be obtained from point clouds (more about this later).
- *Hypergraphs* can be converted to simplicial complexes.

Back to simplicial complexes

Chain groups

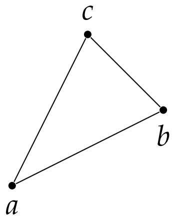
Definition

Given a simplicial complex K, the p^{th} chain group C_p of K consists of all combinations of p-simplices in the complex. Coefficients are in $\mathbb{Z}_2$, hence all elements of C_p are of the form $\sum_j \sigma_j$, for $\sigma_j \in \mathrm{K}$. The group operation is addition with $\mathbb{Z}_2$ coefficients.

$\mathbb{Z}_2$ is convenient for implementation reasons because *addition* can be implemented as *symmetric difference*. Other choices are possible!

We need chain groups to algebraically express the concept of a *boundary*.

Simplicial chains



Let $\mathrm{K} = \{\{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}, \{a, c\}, \{a, b, c\}\}$. Some valid simplicial 1-chains of K are:

- $\{a, b\}$
- $\{a, c\}$
- $\{b, c\}$
- $\{a, b\} + \{a, c\}$
- $\{a, b\} + \{b, c\}$
- $\{a, c\} + \{b, c\}$
- $\{b, c\} + \{a, c\} + \{a, b\}$

Boundary homomorphism

Given a simplicial complex K, the p^{th} boundary homomorphism is a function that assigns each simplex $\sigma = \{v_0, \dots, v_p\} \in \mathrm{K}$ to its *boundary*:

$$\partial_p \sigma = \sum_i \{v_0, \dots, \widehat{v}_i, \dots, v_p\}$$

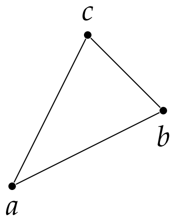
In the equation above, $\widehat{v}_i$ indicates that the set does *not* contain the i^{th} vertex. The function $\partial_p \colon C_p \to C_{p-1}$ is thus a homomorphism between the chain groups.

Caveat

With other coefficients, the boundary homomorphism is slightly more complex, involving alternating signs for the different terms. Over $\mathbb{Z}_2$, signs can be ignored.

Boundary homomorphism

Example



Let $\mathrm{K} = \{\{a\}, \{b\}, \{c\}, \{a,b\}, \{b,c\}, \{a,c\}, \{a,b,c\}\}$. The boundary of the triangle is non-trivial:

$$\partial_2\{a,b,c\} = \{b,c\} + \{a,c\} + \{a,b\}$$

The boundary of its edges is trivial, though, because duplicate simplices cancel each other out:

$$\begin{aligned}\partial_1\left(\{b,c\} + \{a,c\} + \{a,b\}\right) &= \{c\} + \{b\} + \{c\} + \{a\} + \{b\} + \{a\} \\ &= 0\end{aligned}$$

Chain complex

For all p, we have $\partial_{p-1} \circ \partial_p = 0$: *Boundaries do not have a boundary themselves*. This leads to the *chain complex*:

$$0 \xrightarrow{\partial_{n+1}} C_n \xrightarrow{\partial_n} C_{n-1} \xrightarrow{\partial_{n-1}} \dots \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0$$

Cycle and boundary groups

$$\text{Cycle group } Z_p = \ker \partial_p$$
$$\text{Boundary group } B_p = \operatorname{im} \partial_{p+1}$$

We have $B_p \subseteq Z_p$ in the group-theoretical sense. In other words, every boundary is also a cycle. These groups are abelian groups.

(The fact that these sets are groups is a consequence of some deep theorems in group theory! Unfortunately, we cannot cover all of these things here...)

Digression

Normal subgroup and quotient group

Normal subgroup

Let G be a group and N be a subgroup. N is a *normal subgroup* if $gng^{-1} \in N$ for all $g \in G$ and $n \in N$.

For an abelian group, every subgroup is normal!

Definition

Let G be a group and N be a normal subgroup of G. Then the *quotient group* is defined as $G/N := \{gN \mid g \in G\}$, partitioning G into equivalence classes.

Intuitively, G/N consists of all elements in G that are *not* in N.

Homology groups & Betti numbers

The p^{th} homology group H_p is a quotient group, defined by 'removing' cycles that are boundaries from a higher dimension:

$$H_p = Z_p / B_p = \ker \partial_p / \operatorname{im} \partial_{p+1},$$

With this definition, we may finally calculate the p^{th} Betti number:

$$\beta_p = \operatorname{rank} H_p$$

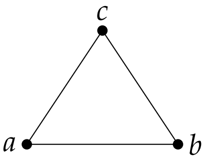
The rank is a generating set of the smallest cardinality. We will see how to calculate this easily!

Intuition

Calculate all boundaries, remove the boundaries that come from higher-dimensional objects, and count what is left.

Example

Simplicial complex



$$\mathrm{K} = \{\{a\},\{b\},\{c\},\{a,b\},\{b,c\},\{a,c\}\}$$

Notice that K does not contain the 2-simplex $\{a,b,c\}$. Next, we will see how to calculate the boundary matrix of K and its homology groups!

Example

Boundary matrix calculation

$$M = \begin{array}{cccccccc} & a & b & c & ab & bc & ac & \\ \Bigg(& 0 & 0 & 0 & 0 & 0 & 0 & \Bigg) \; a \\ & 0 & 0 & 0 & 0 & 0 & 0 & \; b \\ & 0 & 0 & 0 & 0 & 0 & 0 & \; c \\ & 0 & 0 & 0 & 0 & 0 & 0 & \; ab \\ & 0 & 0 & 0 & 0 & 0 & 0 & \; bc \\ & 0 & 0 & 0 & 0 & 0 & 0 & \; ac \end{array}$$

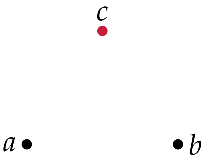
Example

Boundary matrix calculation

$$M = \begin{array}{cccccccc} & a & b & c & ab & bc & ac & \\ \left(\right. & 0 & 0 & 0 & 0 & 0 & 0 & \left.\right) a \\ & 0 & 0 & 0 & 0 & 0 & 0 & b \\ & 0 & 0 & 0 & 0 & 0 & 0 & c \\ & 0 & 0 & 0 & 0 & 0 & 0 & ab \\ & 0 & 0 & 0 & 0 & 0 & 0 & bc \\ & 0 & 0 & 0 & 0 & 0 & 0 & ac \end{array}$$

Example

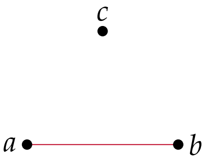
Boundary matrix calculation



$$M = \begin{array}{c@{}c}
\begin{array}{cccccc} a & b & c & ab & bc & ac \end{array} & \\
\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} & \begin{array}{c} a \\ b \\ c \\ ab \\ bc \\ ac \end{array}
\end{array}$$

Example

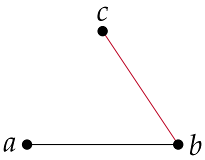
Boundary matrix calculation



$$M = \begin{array}{c c}
\begin{pmatrix}
\begin{array}{cccccc}
a & b & c & ab & bc & ac \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{array}
\end{pmatrix}
\end{array}$$

Example

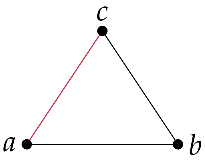
Boundary matrix calculation



$$M = \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \end{array}\begin{array}{cccccc|c} a & b & c & ab & bc & ac & \\ 0 & 0 & 0 & 1 & 0 & 0 & a \\ 0 & 0 & 0 & 1 & 1 & 0 & b \\ 0 & 0 & 0 & 0 & 1 & 0 & c \\ 0 & 0 & 0 & 0 & 0 & 0 & ab \\ 0 & 0 & 0 & 0 & 0 & 0 & bc \\ 0 & 0 & 0 & 0 & 0 & 0 & ac \end{array}$$

Example

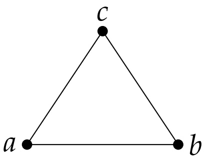
Boundary matrix calculation



$$M = \begin{array}{c|cccccc|c} & a & b & c & ab & bc & ac & \\ \hline & 0 & 0 & 0 & 1 & 0 & 1 & a \\ & 0 & 0 & 0 & 1 & 1 & 0 & b \\ & 0 & 0 & 0 & 0 & 1 & 1 & c \\ & 0 & 0 & 0 & 0 & 0 & 0 & ab \\ & 0 & 0 & 0 & 0 & 0 & 0 & bc \\ & 0 & 0 & 0 & 0 & 0 & 0 & ac \end{array}$$

Example

Boundary matrix calculation



$$M = \begin{array}{c} \\ \\ \end{array}\begin{array}{cccccccc} & a & b & c & ab & bc & ac & \\ \left(\right. & 0 & 0 & 0 & 1 & 0 & 1 & \left.\right) a \\ & 0 & 0 & 0 & 1 & 1 & 0 & b \\ & 0 & 0 & 0 & 0 & 1 & 1 & c \\ & 0 & 0 & 0 & 0 & 0 & 0 & ab \\ & 0 & 0 & 0 & 0 & 0 & 0 & bc \\ & 0 & 0 & 0 & 0 & 0 & 0 & ac \end{array}$$

Example

Dimension 0

To compute H_0, we need to calculate $Z_0 = \ker \partial_0$ and $B_0 = \operatorname{im} \partial_1$.

Calculating Z_0

We have $Z_0 = \ker \partial_0 = \operatorname{span}(\{a\}, \{b\}, \{c\})$, because each one of these simplices is mapped to zero. Since we cannot express any one of these simplices as a linear combination of the others, we have $Z_0 = (\mathbb{Z}/2\mathbb{Z})^3$,

Calculating B_0

We have $B_0 = \operatorname{im} \partial_1 = \operatorname{span}(\{a\} + \{b\}, \{b\} + \{c\}, \{a\} + \{c\})$. However, since $\{a\} + \{b\} + \{b\} + \{c\} = \{a\} + \{c\}$, there are only *two* independent elements, i.e. $\operatorname{im} \partial_1 = \operatorname{span}(\{a\} + \{b\}, \{b\} + \{c\})$. Hence, $B_0 = (\mathbb{Z}/2\mathbb{Z})^2$.

Example

Dimension 0, continued

- By definition, $H_0 = Z_0 / B_0 = (\mathbb{Z}/2\mathbb{Z})^3 / (\mathbb{Z}/2\mathbb{Z})^2 = \mathbb{Z}/2\mathbb{Z}$.
- Hence, $\beta_0 = \operatorname{rank} H_0 = 1$.

Intuition

Our calculation tells us that the simplicial complex has a *single* connected component!

Example

Dimension 1

To compute H_1, we need to calculate $Z_1 = \ker \partial_1$ and $B_1 = \operatorname{im} \partial_2$.

Calculating Z_1

We have $Z_1 = \ker \partial_1 = \operatorname{span}(\{a, b\} + \{b, c\} + \{a, c\})$. This is the *only* cycle in K; we can verify this by inspection or pure combinatorics. Hence, $Z_1 = \mathbb{Z}/2\mathbb{Z}$.

Calculating B_1

There are *no* 2-simplices in K, so $B_1 = \operatorname{im} \partial_2 = \{0\}$.

Example

Dimension 1, continued

- By definition, $H_1 = Z_1 / B_1 = (\mathbb{Z}/2\mathbb{Z}) / \{0\} = \mathbb{Z}/2\mathbb{Z}$.
- Hence, $\beta_1 = \operatorname{rank} H_1 = 1$.

Intuition

Our calculation tells us that the simplicial complex has a *single* cycle!

This is one of the few situations in which a ‘division by zero’ is well-defined! By the definition of the quotient group, this means we are *not* removing any elements from the group.

Homology calculations in practice

Smith normal form

Let M be an $n \times m$ matrix with at least one non-zero entry over some field $\mathbb{F}$. There are invertible matrices S and T such that the matrix product SMT has the form

$$SMT = \begin{pmatrix} b_0 & 0 & 0 & & \cdots & & 0 \\ 0 & b_1 & 0 & & \cdots & & 0 \\ 0 & 0 & \ddots & & & & 0 \\ \vdots & & & b_k & & & \vdots \\ & & & & 0 & & \\ & & & & & \ddots & \\ 0 & & & \cdots & & & 0 \end{pmatrix},$$

where all the entries b_i satisfy $b_i \geq 1$ and divide each other, i.e. $b_i \mid b_{i+1}$. All b_i are unique up to multiplication by a unit.

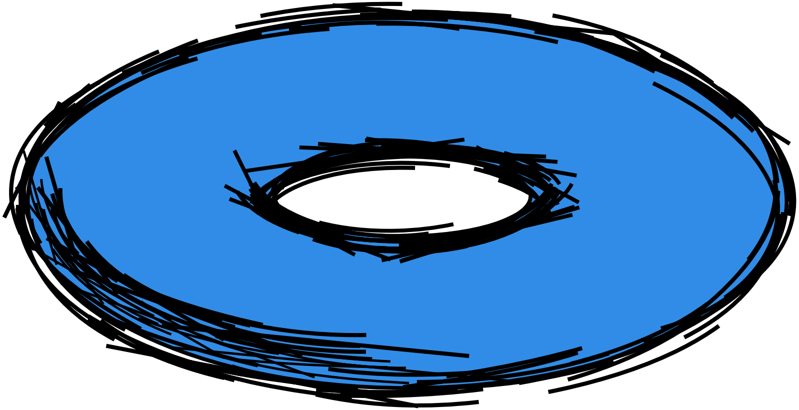
Homology calculations in practice

1. Calculate boundary operator matrices.
2. Bring each matrix into Smith normal form (similar to Gaussian elimination).
3. Read off description of p^{th} homology group.

We have:

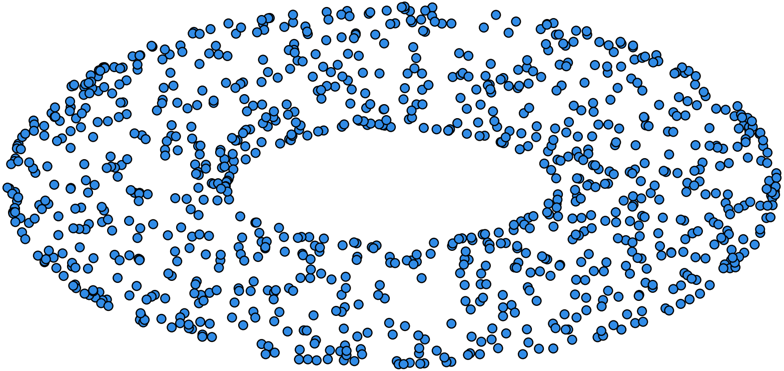
- $\operatorname{rank} Z_p$ is the number of zero columns of the boundary matrix of ∂_p.
- $\operatorname{rank} B_p$ is the number of non-zero rows of the boundary matrix of ∂_{p+1}.

Going from theory to practice



What we see

Going from theory to practice



What we get

From point clouds to simplicial complexes

Vietoris–Rips complex

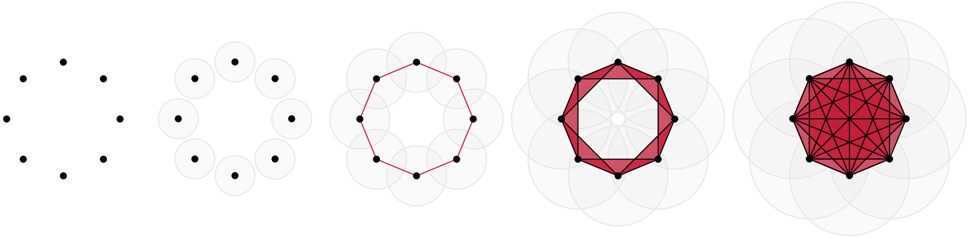
Given a set of points $\mathcal{X} = \{x_1, \dots, x_n\}$ and a metric dist such as the Euclidean distance, pick a threshold ϵ and build the Vietoris–Rips complex $\mathcal{V}_\epsilon$ defined as:

$$\mathcal{V}_\epsilon(\mathcal{X}) := \{\sigma \subseteq \mathcal{X} \mid \forall u, v \in \sigma : \mathrm{dist}(u, v) \leq \epsilon\}$$

Equivalently, $\mathcal{V}_\epsilon$ contains all simplices whose *diameter* is less than or equal to ϵ.

Example

Vietoris–Rips construction

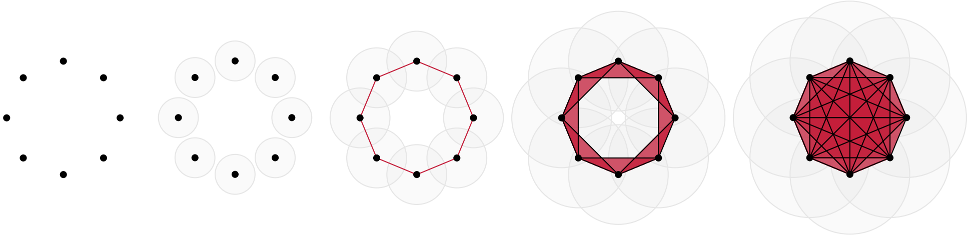


$\epsilon = 0.0$ $\epsilon = 0.1$ $\epsilon = 0.2$ $\epsilon = 0.5$ $\epsilon = 1.0$

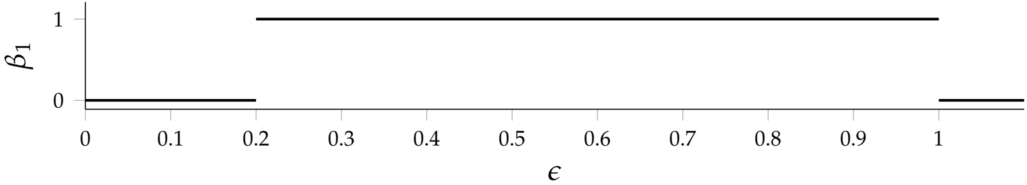
Draw Euclidean balls (circles) of diameter ϵ and create a k-simplex σ for each subset of $k+1$ points that intersect pairwise.

Example

The Betti numbers of a Vietoris–Rips complex



$\epsilon = 0.0$ $\epsilon = 0.1$ $\epsilon = 0.2$ $\epsilon = 0.5$ $\epsilon = 1.0$

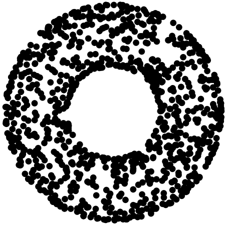


Issues with this approach

- How to pick ϵ?
- There might not be one ‘correct’ value for ϵ.
- Computationally inefficient; matrix reduction has to be performed for *every* simplicial complex.

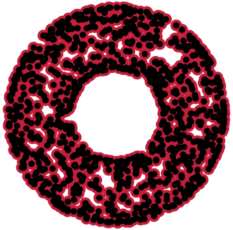
Intuition

Go through all scales and *track* topological features



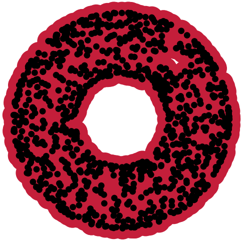
Intuition

Go through all scales and *track* topological features



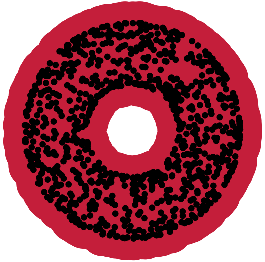
Intuition

Go through all scales and *track* topological features



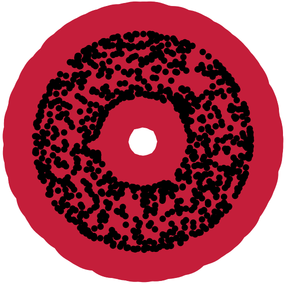
Intuition

Go through all scales and *track* topological features



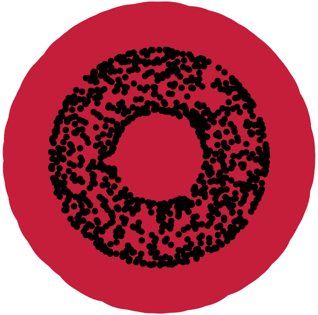
Intuition

Go through all scales and *track* topological features



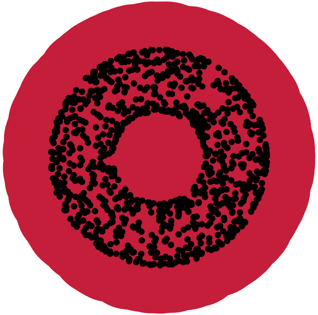
Intuition

Go through all scales and *track* topological features

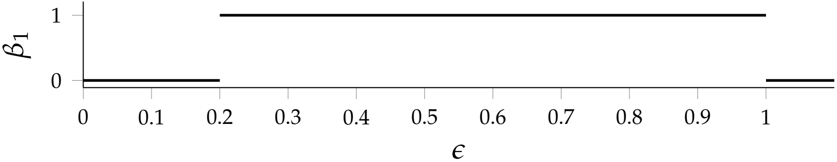


Intuition

Go through all scales and *track* topological features



Filtrations



The Betti number of the data *persists* over a range of the threshold parameter ϵ. To formalise this, assume that simplices in the Vietoris–Rips complex are added one after the other. This gives rise to a *filtration* , i.e.

$$\emptyset = \mathrm{K}_0 \subseteq \mathrm{K}_1 \subseteq \cdots \subseteq \mathrm{K}_{n-1} \subseteq \mathrm{K}_n = \mathcal{V}_\epsilon,$$

where each K_i is a valid simplicial subcomplex of $\mathcal{V}_\epsilon$.

Chain complexes and filtrations

Since $\mathrm{K}_i \subseteq \mathrm{K}_j$ for $i \leq j$, we obtain a sequence of homomorphisms connecting the homology groups of each simplicial complex, i.e.

$$f_p^{i,j} \colon H_p(\mathrm{K}_i) \to H_p(\mathrm{K}_j),$$

which in turn gives rise to a sequence of homology groups, i.e.

$$0 = H_p(\mathrm{K}_0) \xrightarrow{f_p^{0,1}} H_p(\mathrm{K}_1) \xrightarrow{f_p^{1,2}} \dots \xrightarrow{f_p^{n-2,n-1}} H_p(\mathrm{K}_{n-1}) \xrightarrow{f_p^{n-1,n}} H_p(\mathrm{K}_n) = H_p(\mathcal{V}_\epsilon),$$

with p denoting the dimension of the corresponding homology group.

Persistent homology group

Given two indices $i \leq j$, the p^{th} persistent homology group $H_p^{i,j}$ is defined as

$$H_p^{i,j} := Z_p\left(\mathrm{K}_i\right) / \left(B_p\left(\mathrm{K}_j\right) \cap Z_p\left(\mathrm{K}_i\right)\right),$$

which contains all the homology classes of K_i that are still present in K_j.

Implication

We can calculate a new set of homology groups alongside the filtration and assign a ‘duration’ to each topological feature.

Persistent homology

Tracking of topological features

- *Creation* in K_i: $c \in H_p\left(\mathrm{K}_i\right)$, but $c \notin H_p^{i-1,i}$
- *Destruction* in K_j: c is created in K_i, with $f_p^{i,j-1}(c) \notin H_p^{i-1,j-1}$ and $f_p^{i,j}(c) \in H_p^{i-1,j}$

The *persistence* of a class c that is created in K_i and destroyed in K_j is defined as

$$\operatorname{pers}(c) := |\,\mathrm{w}(j) - \mathrm{w}(i)|,$$

where $\mathrm{w}\colon \mathbb{Z} \to \mathbb{R}$ assigns each simplicial complex of the filtration a weight, such as an associated distance, or an index. Persistence thus measures the ‘scale’ at which a certain topological feature occurs.

Standard filtrations

The distance filtration

Given a distance metric dist, such as the Euclidean metric, the *distance filtration* assigns weights based on pairwise distances between points:

$$\mathrm{w}(\sigma) := \begin{cases} 0 & \text{if } \sigma \text{ is a vertex} \\ \mathrm{dist}(u, v) & \text{if } \sigma = \{u, v\} \\ \max_{\tau \subseteq \sigma} \mathrm{w}(\tau) & \text{else} \end{cases}$$

Simplices need to be sorted in *ascending* order of their weights; in case of a tie, faces precede co-faces.

Persistent homology is capable of *preserving* distances under random projections[1].

[1] D. R. Sheehy, 'The Persistent Homology of Distance Functions under Random Projection', *Proceedings of the 30th Annual Symposium on Computational Geometry*, 2014, pp. 328–334

Example

Boundary matrix calculation alongside a filtration

$$M = \begin{array}{c} \begin{array}{ccccccc} a & b & c & ab & bc & ac & abc \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \begin{array}{l} \\ a \\ b \\ c \\ ab \\ bc \\ ac \\ abc \end{array}$$

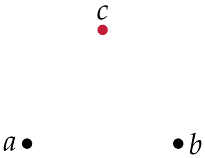
Example

Boundary matrix calculation alongside a filtration

$$\begin{array}{rl} & \begin{array}{ccccccc} a & b & c & ab & bc & ac & abc \end{array} \\ M = & \left(\begin{array}{ccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \begin{array}{l} a \\ b \\ c \\ ab \\ bc \\ ac \\ abc \end{array} \end{array}$$

Example

Boundary matrix calculation alongside a filtration

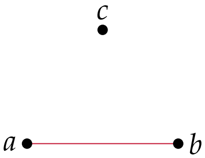


$$M = \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \end{array}\begin{array}{cccccccc} a & b & c & ab & bc & ac & abc & \\ \end{array}$$

$$M = \begin{array}{c|ccccccc} & a & b & c & ab & bc & ac & abc \\ \hline a & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ b & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ c & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ ab & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ bc & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ ac & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ abc & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}$$

Example

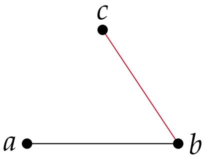
Boundary matrix calculation alongside a filtration



$$M = \begin{array}{c} \\ \\ \end{array}\begin{array}{cccccccc} a & b & c & ab & bc & ac & abc & \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & a \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & b \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & c \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & ab \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & bc \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & ac \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & abc \end{array}$$

Example

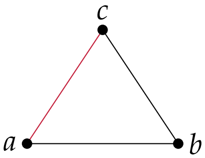
Boundary matrix calculation alongside a filtration



$$M = \begin{array}{c} \\ \\ \end{array}\begin{array}{cccccccc} a & b & c & ab & bc & ac & abc & \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & a \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & b \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & c \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & ab \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & bc \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & ac \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & abc \end{array}$$

Example

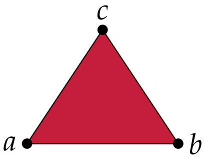
Boundary matrix calculation alongside a filtration



$$M = \begin{array}{c c c c c c c c}
a & b & c & ab & bc & ac & abc & \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & a \\
0 & 0 & 0 & 1 & 1 & 0 & 0 & b \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & c \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & ab \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & bc \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & ac \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & abc
\end{array}$$

Example

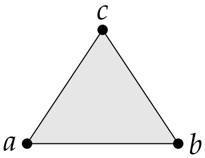
Boundary matrix calculation alongside a filtration



$$M = \begin{array}{c} \begin{array}{ccccccc} a & b & c & ab & bc & ac & abc \end{array} \\ \left(\begin{array}{ccccccc} 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \end{array} \begin{array}{l} \\ a \\ b \\ c \\ ab \\ bc \\ ac \\ abc \end{array}$$

Example

Boundary matrix calculation alongside a filtration



$$M = \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \end{array}\begin{array}{ccccccc|c} a & b & c & ab & bc & ac & abc & \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & a \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & b \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & c \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & ab \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & bc \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & ac \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & abc \end{array}$$

Boundary matrix reduction by column operations

```
Let $M$ be a *boundary matrix*
for $i = 1$ do
    while $\exists i' < i : \mathrm{low}(i') = \mathrm{low}(i) \neq 0$
    do
        $M(i) = M(i) \oplus M(i')$
    end while
end for
```

$$M = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Boundary matrix reduction by column operations

```
Let $M$ be a *boundary matrix*
for $i = 1$ do
    while $\exists i' < i : \mathrm{low}(i') = \mathrm{low}(i) \neq 0$
    do
        $M(i) = M(i) \oplus M(i')$
    end while
end for
```

$$M = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Boundary matrix reduction by column operations

```
Let $M$ be a *boundary matrix*
for $i = 1$ do
    while $\exists i' < i : \mathrm{low}(i') = \mathrm{low}(i) \neq 0$
    do
        $M(i) = M(i) \oplus M(i')$
    end while
end for
```

$$M = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Boundary matrix reduction by column operations

```
Let $M$ be a *boundary matrix*
for $i = 1$ do
    while $\exists i' < i : \operatorname{low}(i') = \operatorname{low}(i) \neq 0$
    do
        $M(i) = M(i) \oplus M(i')$
    end while
end for
```

$$M = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Boundary matrix reduction by column operations

```
Let $M$ be a *boundary matrix*
for $i = 1$ do
    while $\exists i' < i : \mathrm{low}(i') = \mathrm{low}(i) \neq 0$
    do
        $M(i) = M(i) \oplus M(i')$
    end while
end for
```

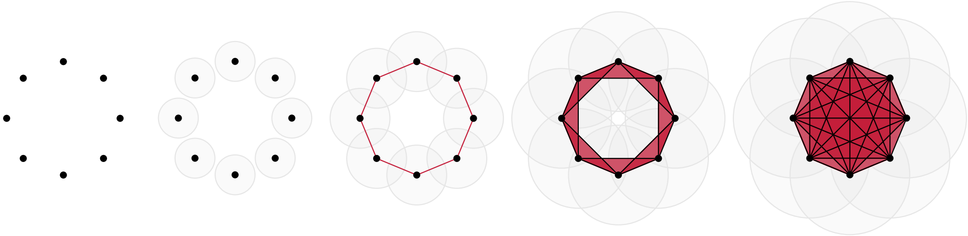
$$M = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Using the reduced boundary matrix

$$\begin{array}{cccccccc}
a & b & c & ab & bc & ac & abc & \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & a \\
0 & 0 & 0 & 1 & 1 & 0 & 0 & b \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & c \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & ab \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & bc \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & ac \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & abc
\end{array}$$

- If column i is empty, then σ_i is a *positive* simplex that *creates* a topological feature.
- If column j is non-empty with $\mathrm{low}(j) = k$, then σ_j is a *negative* simplex that *destroys* the topological feature created by σ_k.
- For example, simplex abc destroys the cycle created by ac.

Illustrative example



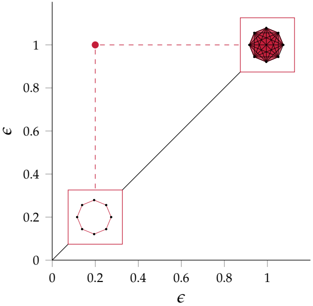
$\epsilon = 0.0$ $\epsilon = 0.1$ $\epsilon = 0.2$ $\epsilon = 0.5$ $\epsilon = 1.0$

Here, the topological feature is the circle that underlies that data. Since it persists from $\epsilon = 0.20$ to $\epsilon = 1.0$, its persistence is $\mathrm{pers} = 1.0 - 0.20 = 0.80$.

Topological features and how to track them

Types of topological features

- Dimension 0: *connected components*
- Dimension 1: *cycles*
- Dimension 2: *voids*



Given a topological feature with associated simplicial complexes K_i and K_j, store the point $(\mathrm{w}(i), \mathrm{w}(j))$ in a *persistence diagram*.

If a feature is *never* destroyed, we assign it a weight of ∞.