Topological Data Analysis for Machine Learning
Lecture 1: Algebraic Topology

Bastian Rieck

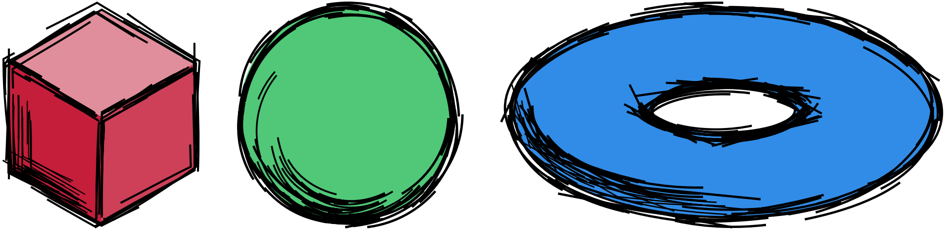
Pseudomanifold

Preliminaries

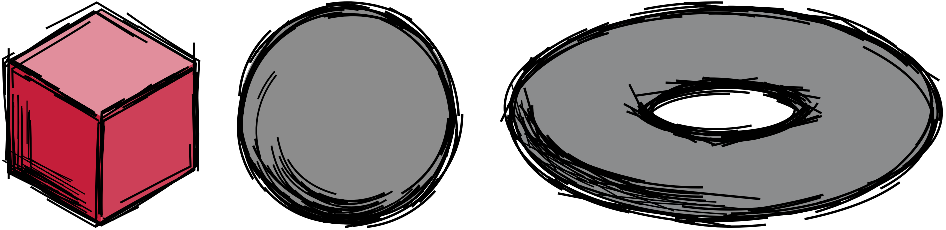
Do you have feedback or any questions? Write to bastian.rieck@bsse.ethz.ch or reach out to @Pseudomanifold on Twitter. You can find the slides and additional information with links to more literature here:

https://topology.rocks/ecml_pkdd_2020

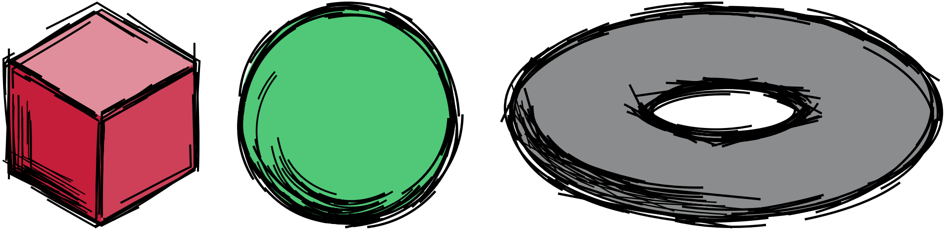
What is computational topology?



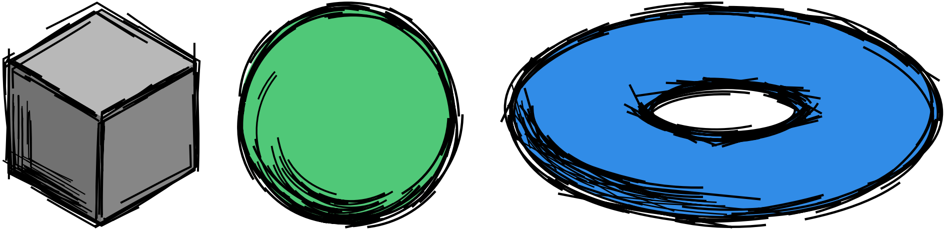
What is computational topology?



What is computational topology?



What is computational topology?



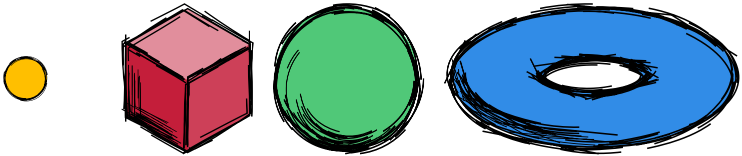
Which qualities of the sphere make it *different* from the torus?

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:

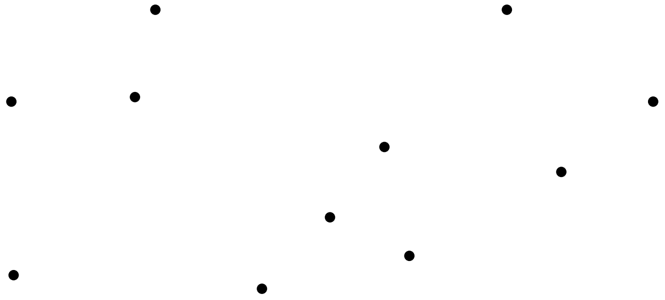
1. $\{v\} \in S$ for all $v \in \mathrm{K}$.
2. If $\sigma \in S$ and $\tau \subseteq \sigma$, then $\tau \in \mathrm{K}$.

Terminology

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

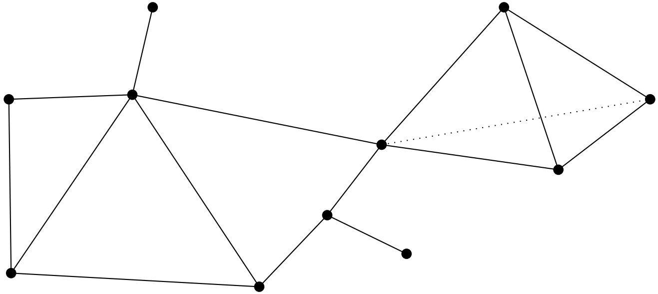
Simplicial complexes
Example

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



Simplicial complexes
Example

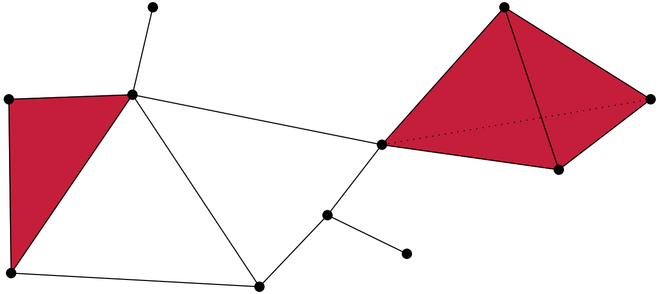
Simplicial complexes can be decomposed into their skeletons, which only contain simplices of a certain dimension.



Simplicial complexes

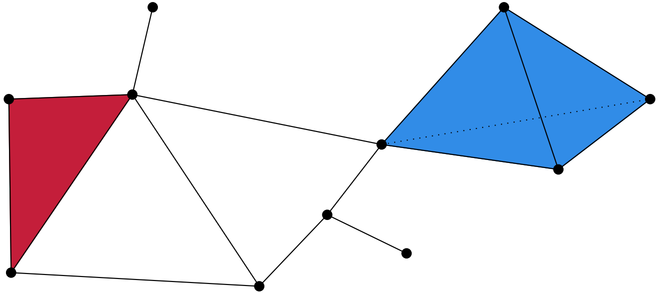
Example

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



Simplicial complexes
Example

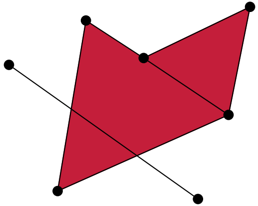
Simplicial complexes can be decomposed into their skeletons, which only contain simplices of a certain dimension.



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.

Digression

Groups

Definition

A *group* is a set G with a binary operation $\cdot$ that combines two elements to yield another one, such that $(G, \cdot)$ has the following properties:

1. The operation is *closed*, i.e. $a \cdot b \in G$ for $a, b \in G$.
2. The operation is *associative*, i.e. $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for $a, b, c \in G$.
3. There is an *identity element* $e \in G$ such that $e \cdot a = a \cdot e = a$ for $a \in G$.
4. Each $a \in G$ has an *inverse element* $a^{-1} \in G$ such that $a \cdot a^{-1} = e = a^{-1} \cdot a$.

The operation $\cdot$ is not required to be commutative. In general, $a \cdot b = b \cdot a$ is *not* required to hold. However, the groups that we will encounter are commutative!

Groups

Examples and non-examples

- The set with only two elements and addition modulo 2 is group, called $\mathbb{Z}_2$.[1]
- The set of integers $\mathbb{Z}$ with the usual addition is a group.
- The set of $\mathbb{R}$-valued quadratic matrices with elementwise addition is a group.
- The set of $\mathbb{R}$-valued quadratic matrices with non-zero determinant together with matrix multiplication is a group.
- The natural numbers $\mathbb{N}$ with addition do *not* form a group (why?).

[1] It is actually also a *field*, the smallest non-trivial field.

Back to simplicial complexes

Chain groups

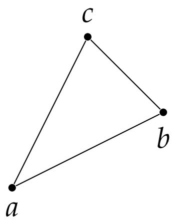
Definition

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 $\sigma = \{v_0, \ldots, v_p\} \in \mathrm{K}$ to its *boundary*:

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

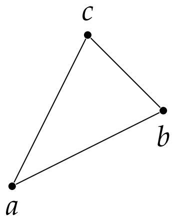
In the equation above, $\widehat{v}_i$ indicates that the set does *not* contain the i^{th} vertex. The function $\partial_p \colon C_p \to C_{p-1}$ is thus a homomorphism between the chain groups.

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*:

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

Digression
Kernel and image

Definition

The *kernel* of a homomorphism $f\colon A \to B$ is the set of all elements that are mapped to the zero element, i.e. $\ker f := \{a \in A \mid f(a) = 0\} \subseteq A$. The *image* of f is the set of all its outputs, i.e. $\operatorname{im} f := \{f(a) \mid a \in A\} \subseteq B$.

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.

(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…)

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.

Quotient groups

Example

$2\mathbb{Z} \subseteq \mathbb{Z}$ is the subgroup of $\mathbb{Z}$ defined by being a multiple of 2. Hence, $\mathbb{Z}_2 := \mathbb{Z}/2\mathbb{Z}$ consists of only 0 and 1.

Why quotient groups?

Quotient groups 'reduce' a group by partitioning it into equivalence classes that are defined by another subgroup.

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:

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

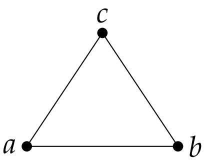
The rank is a generating set of the smallest cardinality. We will see how to calculate this easily!

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

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

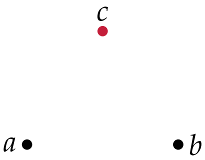
Example

Boundary matrix calculation

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

Example

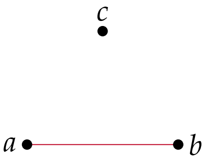
Boundary matrix calculation



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

Example

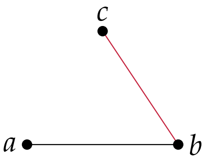
Boundary matrix calculation



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

Example

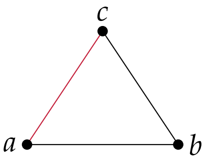
Boundary matrix calculation



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

Example

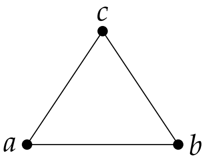
Boundary matrix calculation



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

Example

Boundary matrix calculation



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

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

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}.

Take-away messages

- Homology groups characterise topological objects.
- They can be easily expressed as linear operators.
- The calculation of homology groups boils down to linear algebra.

https://topology.rocks/ecml_pkdd_2020