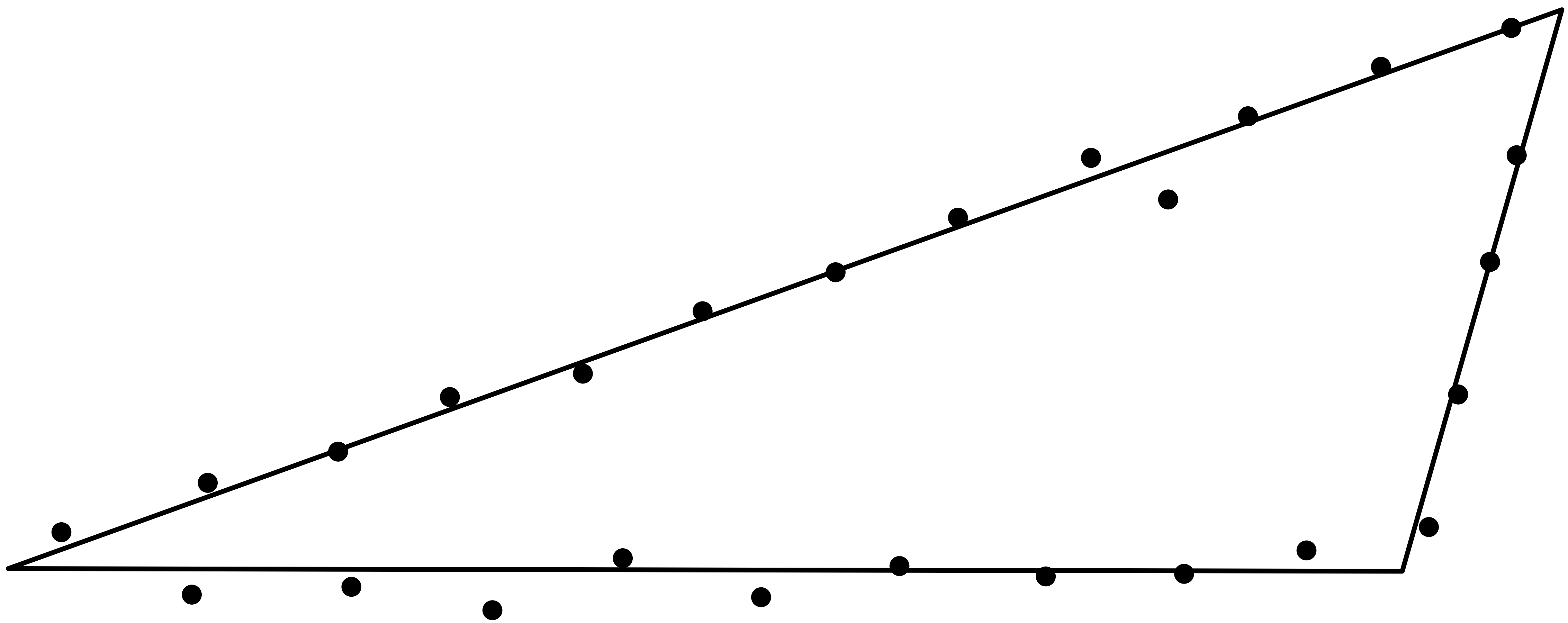
Sophia-Antipolis - January 2017

Homology and homology inference

F. Chazal
DataShape Group
INRIA Saclay

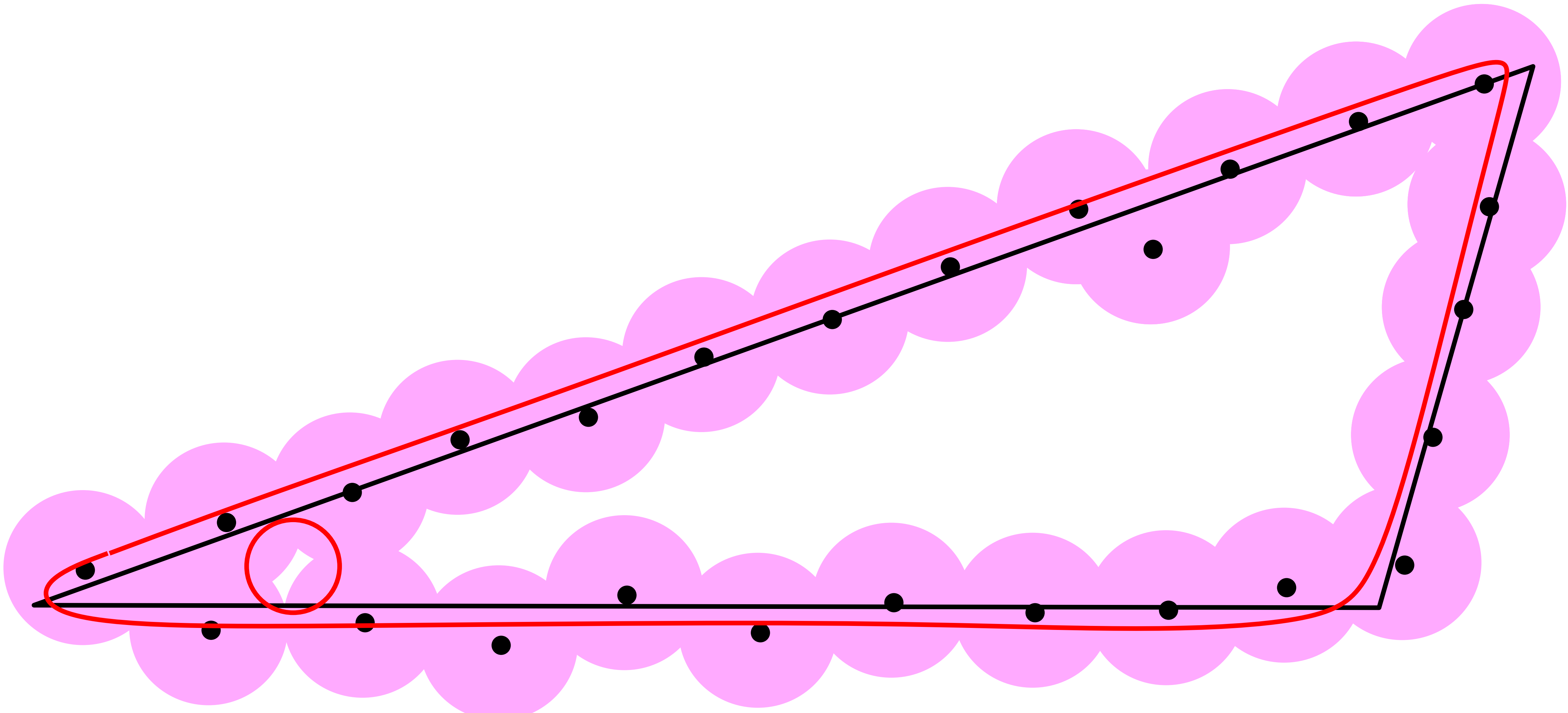
informatics mathematics

Motivation: getting topological information without reconstructing



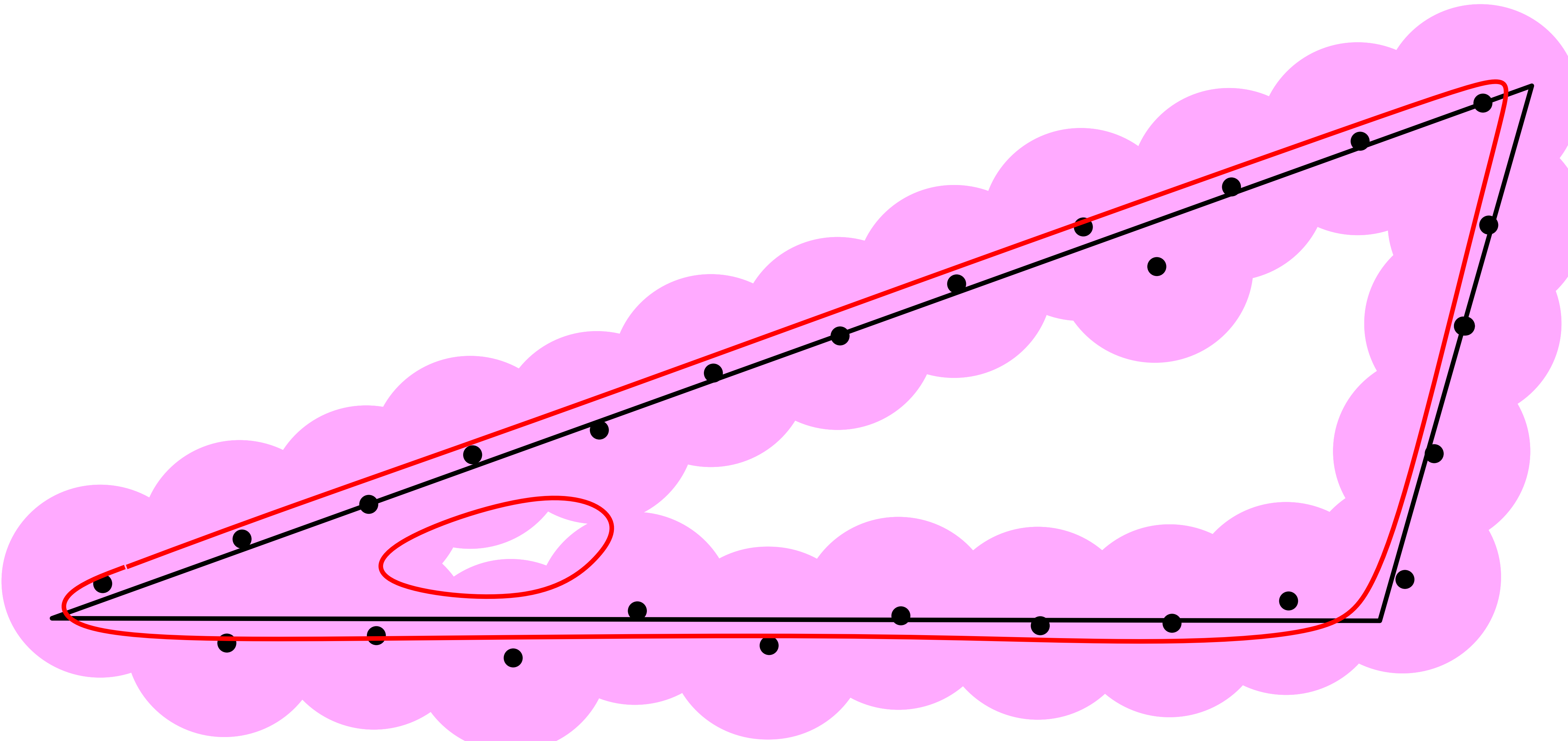
How to determine the number of “cycles” of the underlying shape from the point cloud approximation?

Motivation: getting topological information without reconstructing



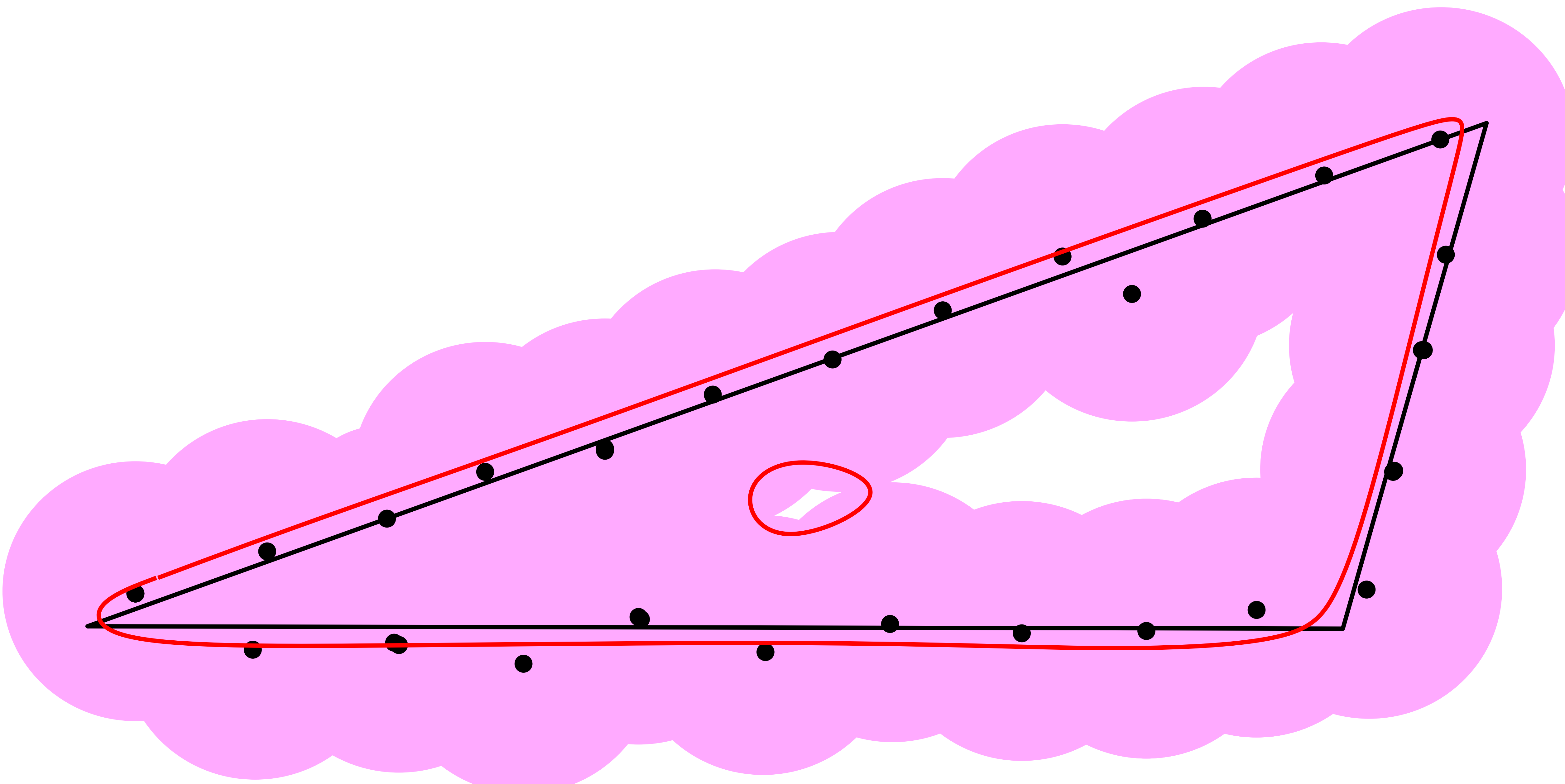
How to determine the number of "cycles" of the underlying shape from the point cloud approximation?

Motivation: getting topological information without reconstructing



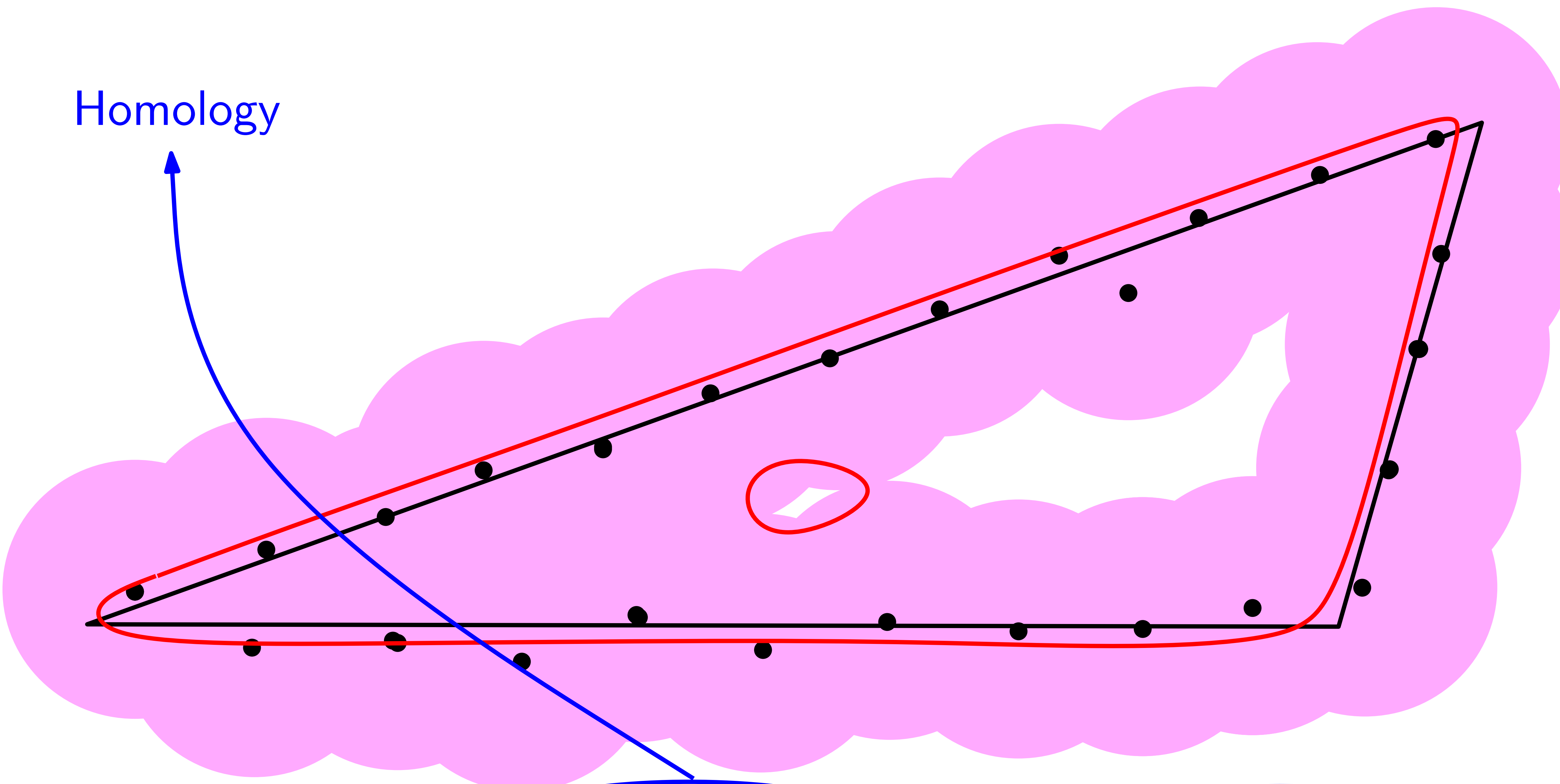
How to determine the number of “cycles” of the underlying shape from the point cloud approximation?

Motivation: getting topological information without reconstructing



How to determine the number of “cycles” of the underlying shape from the point cloud approximation?

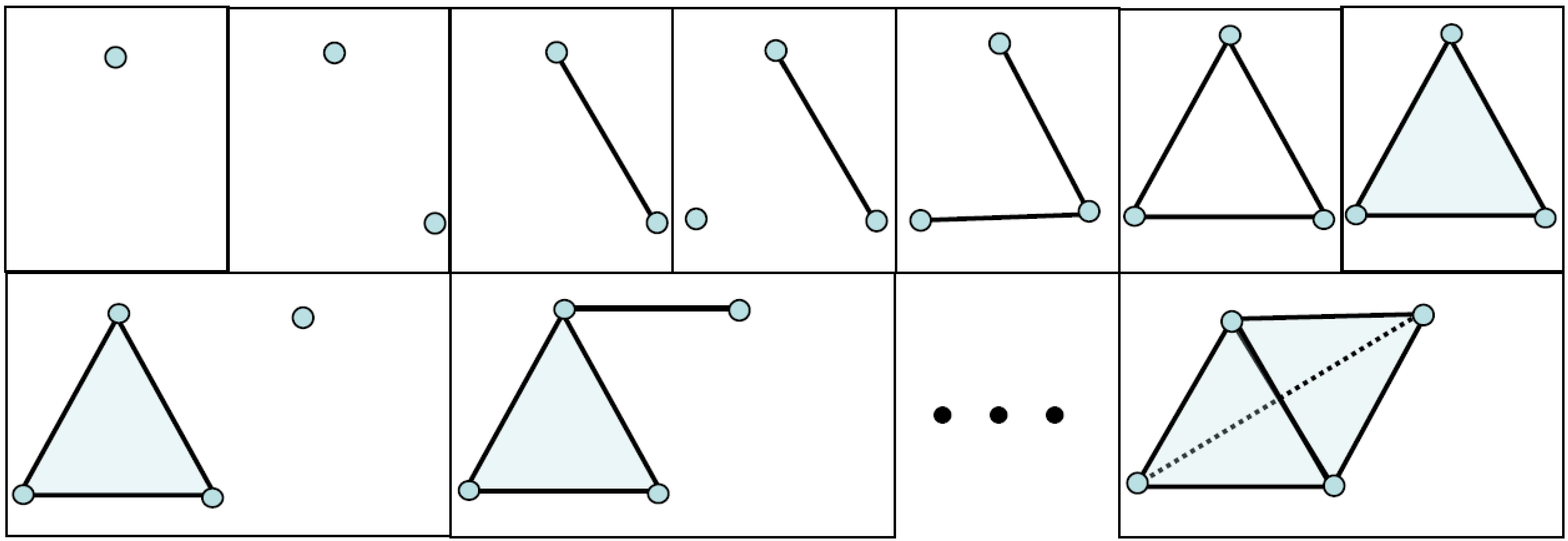
Motivation: getting topological information without reconstructing



How to determine the number of "cycles" of the underlying shape from the point cloud approximation?

Persistent homology

Filtrations of simplicial complexes



A filtration of a (finite) simplicial complex K is a sequence of subcomplexes such that

i) $\emptyset = K^0 \subset K^1 \subset \cdots \subset K^m = K$,

ii) $K^{i+1} = K^i \cup \sigma^{i+1}$ where σ^{i+1} is a simplex of K.

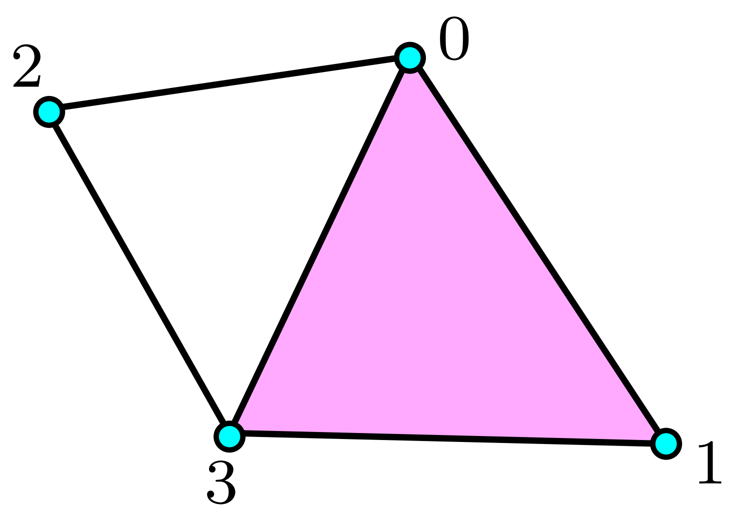
Example: filtration associated to a function

- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: filtration associated to a function

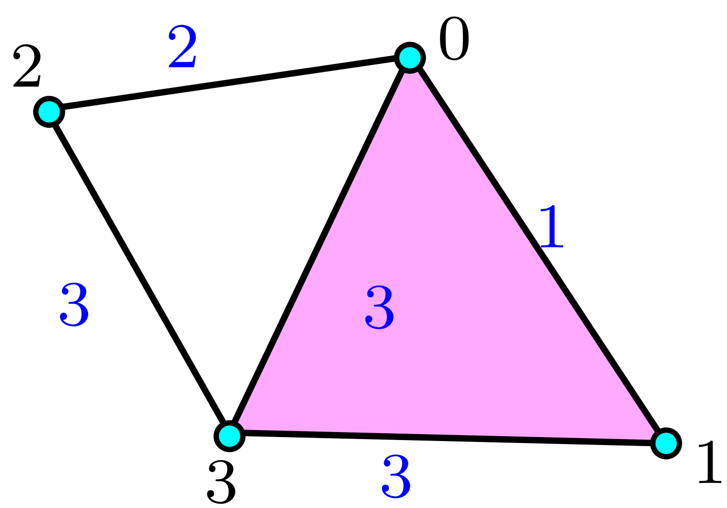


- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: filtration associated to a function

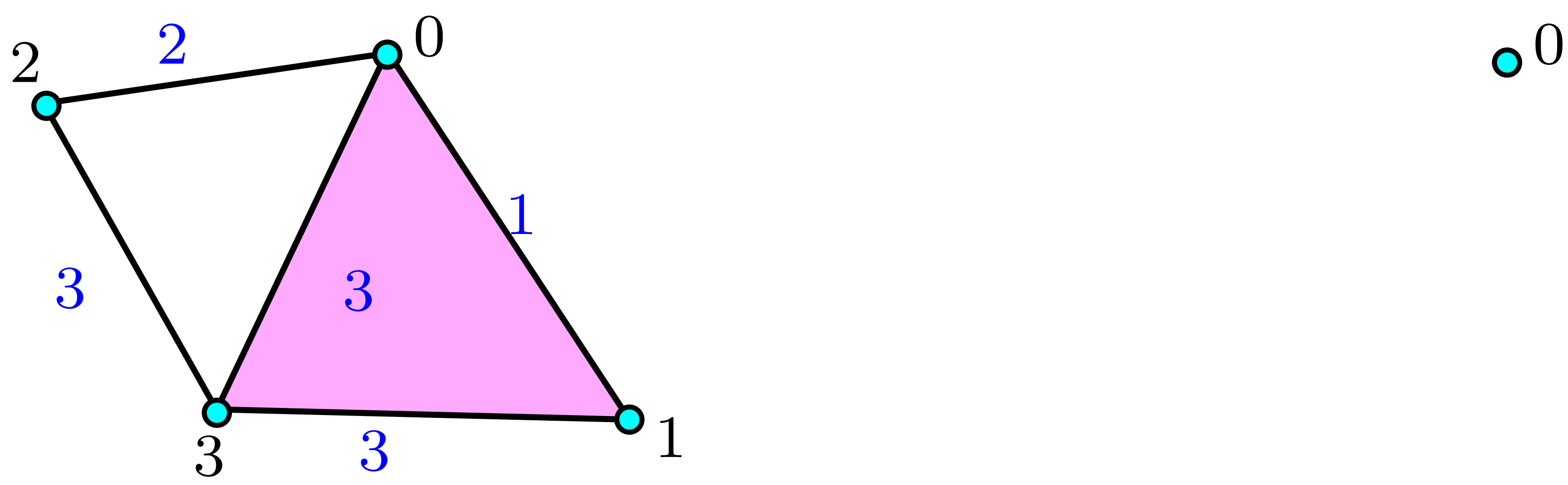


- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: filtration associated to a function

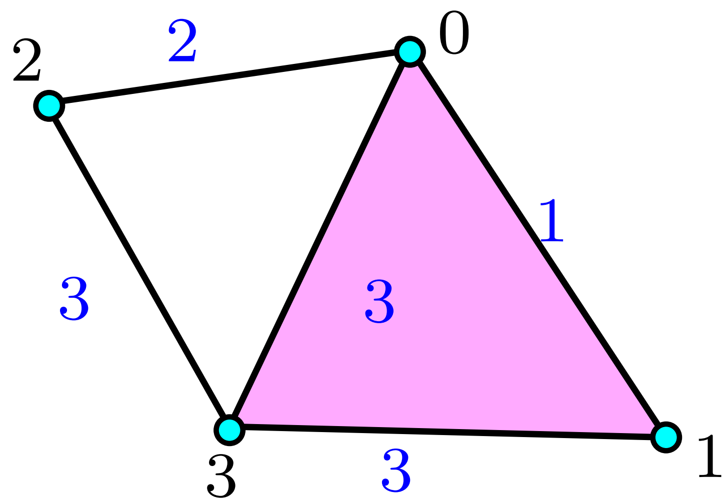


- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: filtration associated to a function

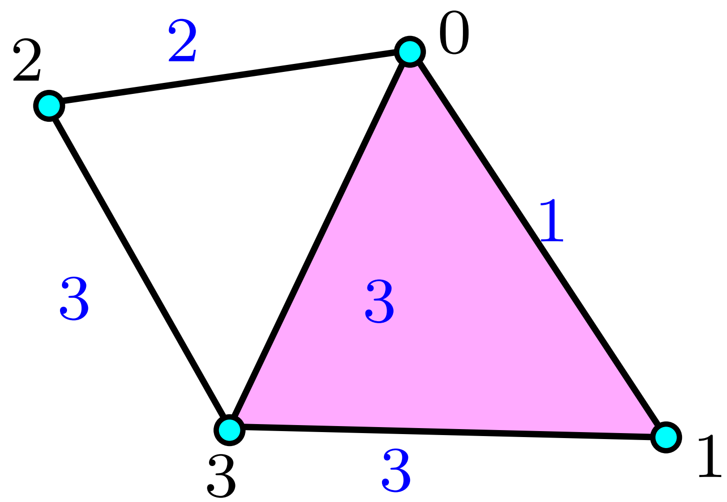


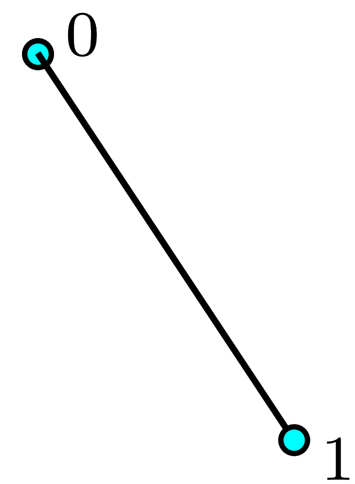
- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: filtration associated to a function



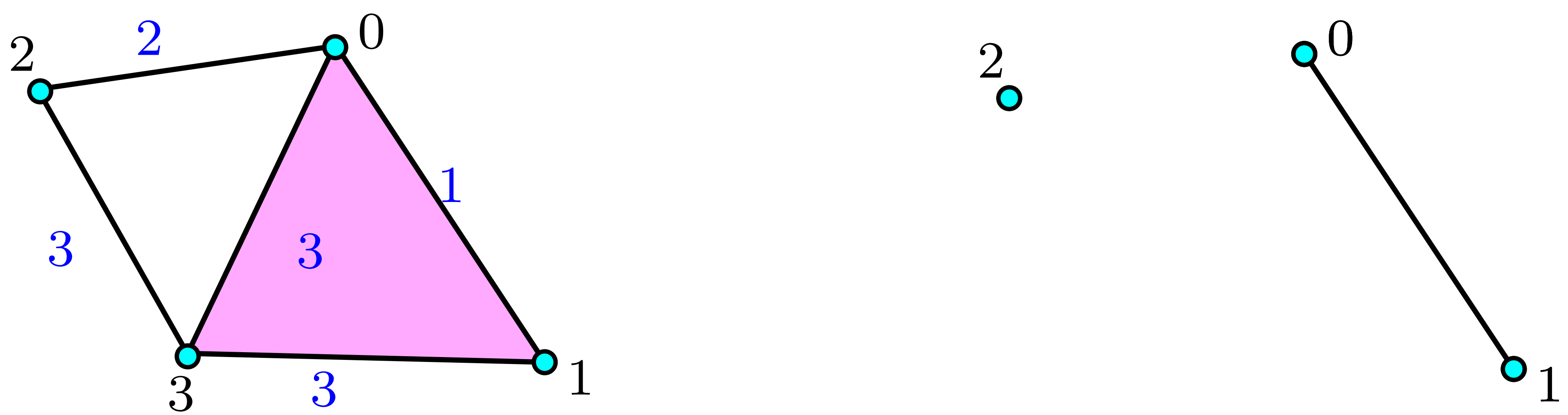


- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: filtration associated to a function

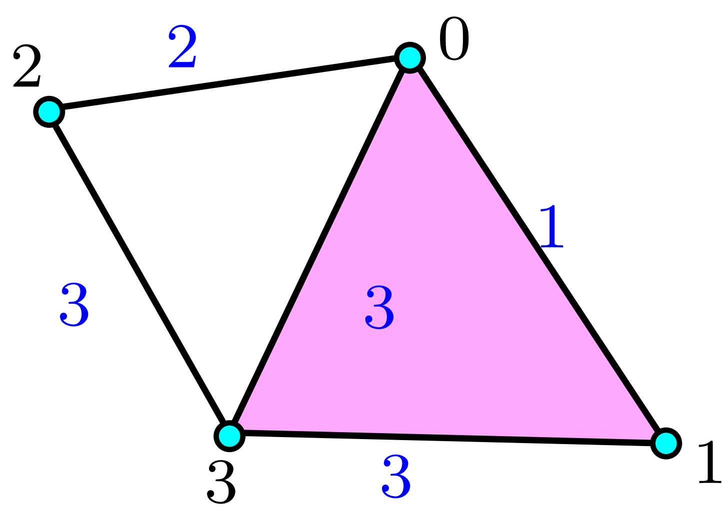


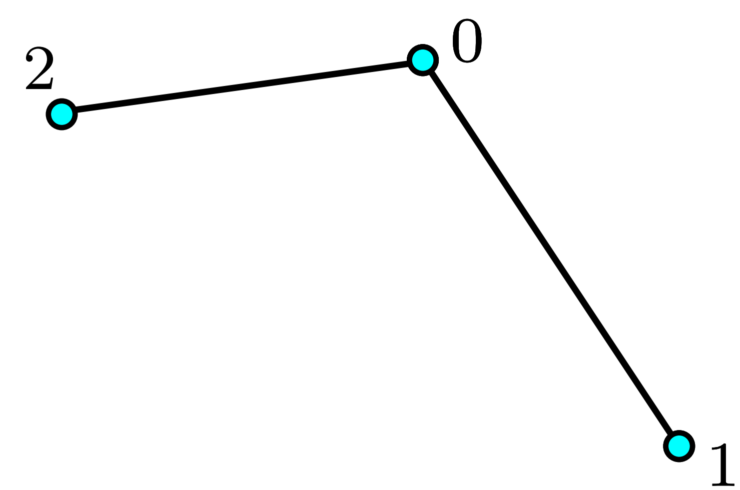
- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: filtration associated to a function



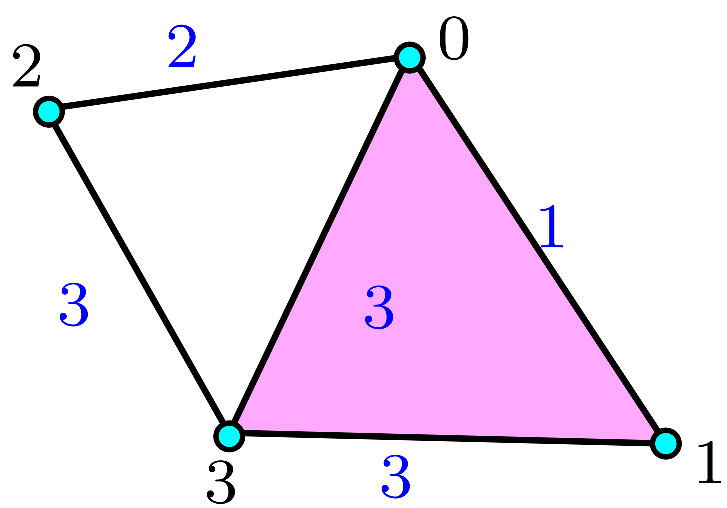


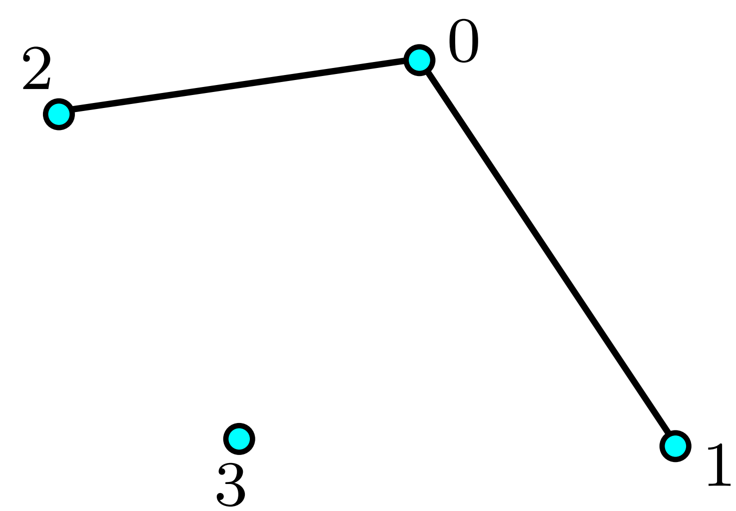
- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: filtration associated to a function



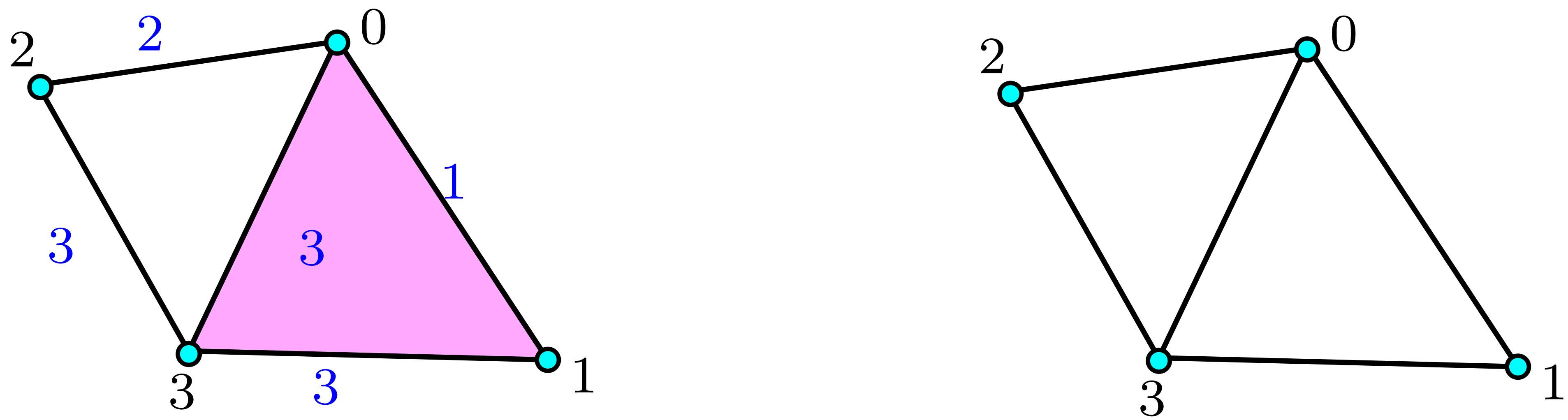


- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: filtration associated to a function

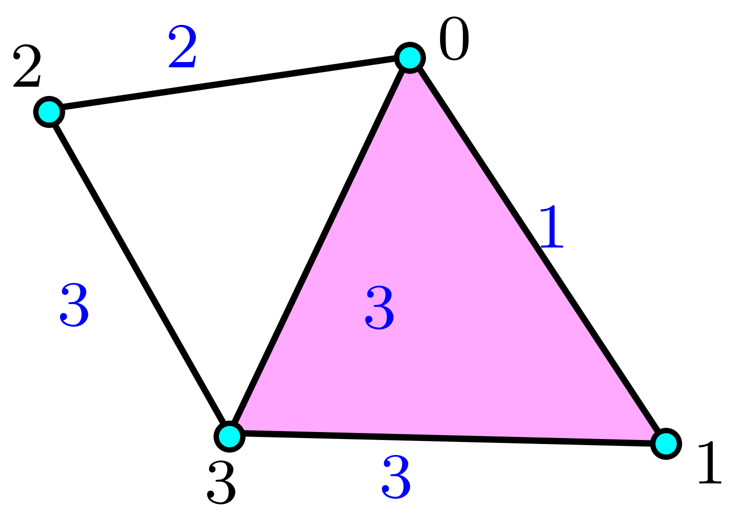


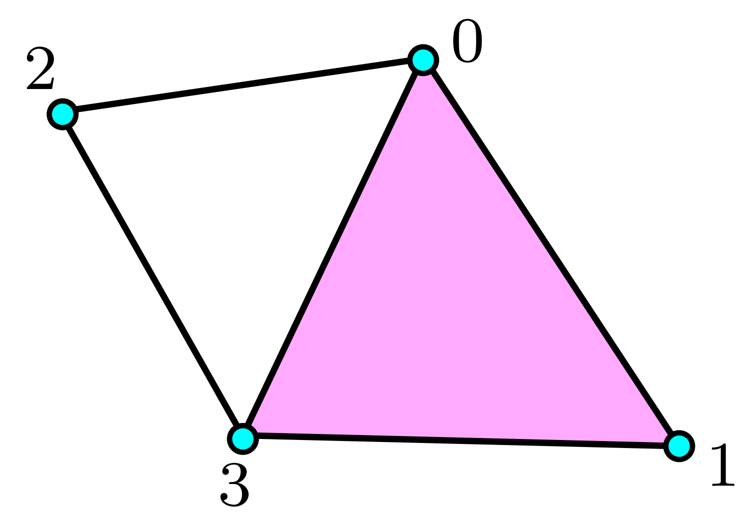
- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: filtration associated to a function



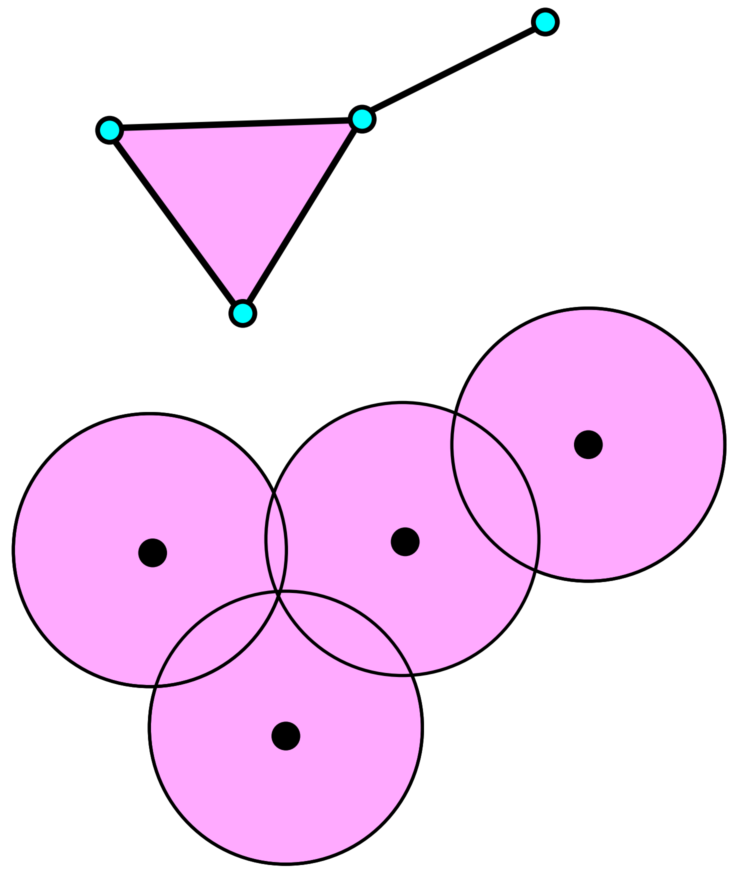


- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: The Cěch filtration

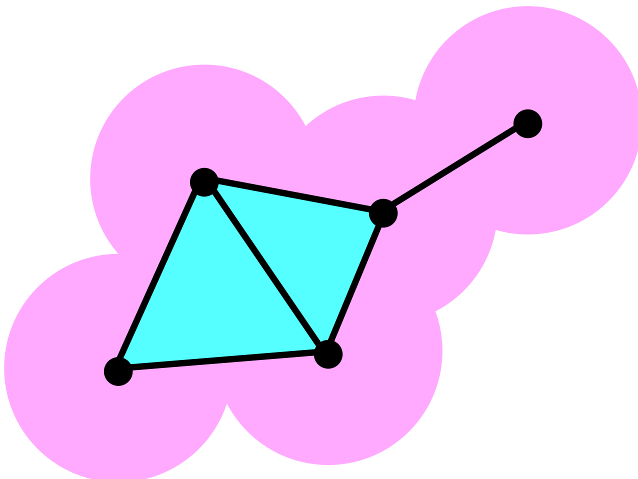


Let $P = \{p_0, \cdots p_n\}$ be a (finite) point cloud (in a metric space).

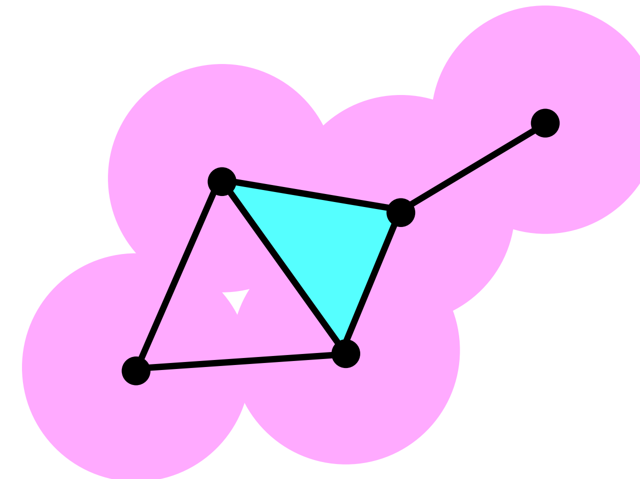
The Cěch complex $\mathcal{C}^\alpha(P)$: for $p_0, \cdots p_k \in P$,

$$\sigma = [p_0 p_1 \cdots p_k] \in \mathcal{C}^\alpha(P) \quad \text{iff} \quad \cap_{i=0}^{k} B(p_i, \alpha) \neq \emptyset$$

Example: the Rips complex



Rips vs Čech



Let $P = \{p_0, \cdots p_n\}$ be a (finite) point cloud (in a metric space).

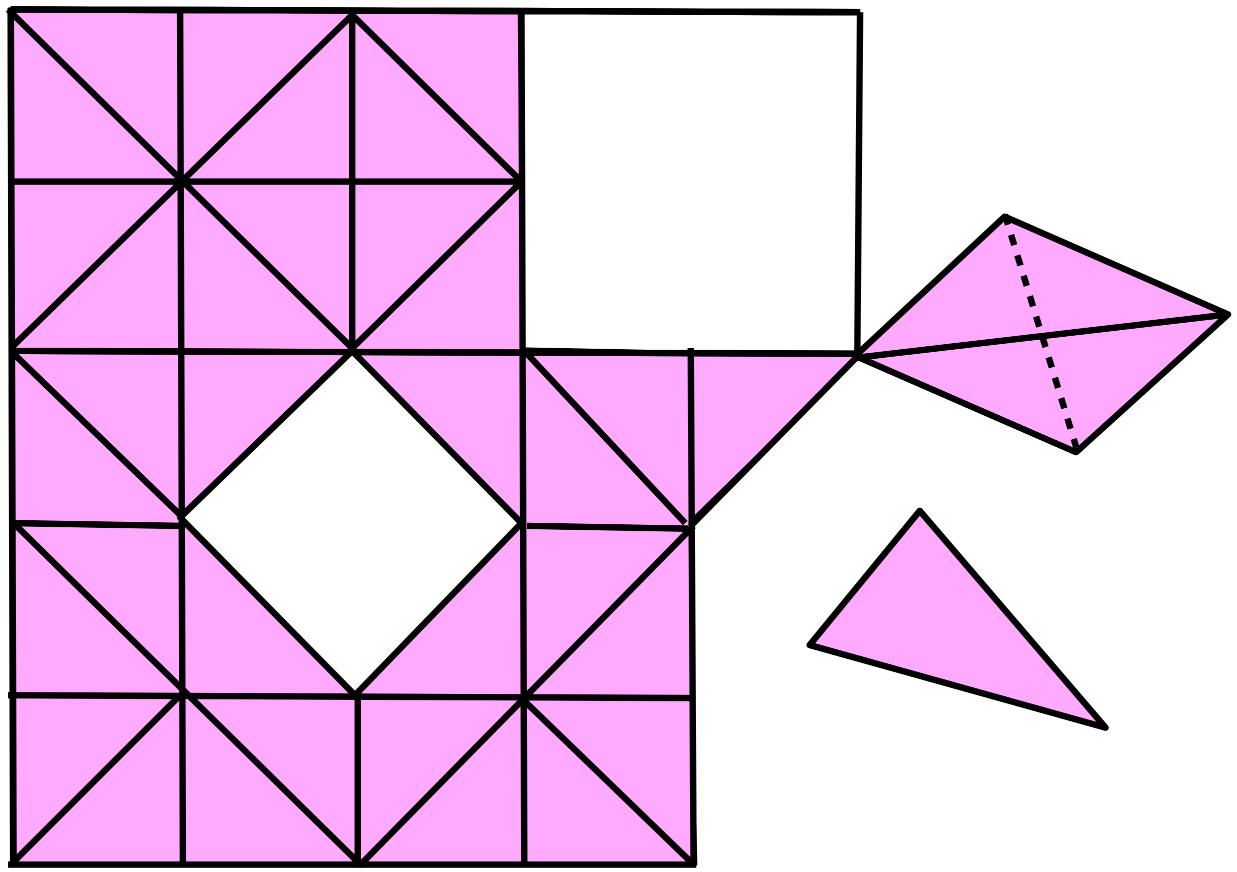
The Rips complex $\mathcal{R}^\alpha(P)$: for $p_0, \cdots p_k \in P$,

$$\sigma = [p_0 p_1 \cdots p_k] \in \mathcal{R}^\alpha(P) \;\text{ iff }\; \forall i, j \in \{0, \cdots k\},\; d(p_i, p_j) \leq \alpha$$

- Easy to compute and fully determined by its 1-skeleton
- Rips-Čech interleaving: for any $\alpha > 0$,

$$\mathcal{C}^{\frac{\alpha}{2}}(P) \subseteq \mathcal{R}^\alpha(P) \subseteq \mathcal{C}^\alpha(P) \subseteq \mathcal{R}^{2\alpha}(P) \subseteq \cdots$$

Homology of simplicial complexes



- 2 connected components
- Intuitively: 2 cycles

Topological invariants:
- Number of connected components
- Number of cycles: how to define a cycle?
- Number of voids: how to define a void?
- ...

(Simplicial) homology and Betti numbers

In the following: homology with coefficient in $\mathbb{Z}/2$

Refs: J.R. Munkres, *Elements of Algebraic Topology*, Addison-Wesley, 1984.
A. Hatcher, *Algebraic Topology*, Cambridge University Press 2002.

The space of k-chains

Let K be a d-dimensional simplicial complex. Let $k \in \{0, 1, \cdots, d\}$ and $\{\sigma_1, \cdots, \sigma_p\}$ be the set of k-simplices of K.

k-chain:

$$c = \sum_{i=1}^{p} \varepsilon_i \sigma_i \quad \text{with} \quad \varepsilon_i \in \mathbb{Z}/2\mathbb{Z} = \{0, 1\}$$

Sum of k-chains:

$$c + c' = \sum_{i=1}^{p} (\varepsilon_i + \varepsilon_i') \sigma_i \quad \text{and} \quad \lambda.c = \sum_{i=1}^{p} (\lambda \varepsilon_i') \sigma_i$$

where the sums $\varepsilon_i + \varepsilon_i'$ and the products $\lambda \varepsilon_i$ are modulo 2.

The space of k-chains

Let K be a d-dimensional simplicial complex. Let $k \in \{0, 1, \cdots, d\}$ and $\{\sigma_1, \cdots, \sigma_p\}$ be the set of k-simplices of K.

k-chain:

$$c = \sum_{i=1}^{p} \varepsilon_i \sigma_i \quad \text{with} \quad \varepsilon_i \in \mathbb{Z}/2\mathbb{Z} = \{0, 1\}$$

Sum of k-chains:

$$c + c' = \sum_{i=1}^{p} (\varepsilon_i + \varepsilon'_i)\sigma_i \quad \text{and} \quad \lambda.c = \sum_{i=1}^{p} (\lambda \varepsilon'_i)\sigma_i$$

The space $\mathcal{C}_k(K)$ of k-chains is a $\mathbb{Z}/2$-vector space

where the sums $\varepsilon_i + \varepsilon'_i$ and the products $\lambda\varepsilon_i$ are modulo 2.

Geometric interpretation:
k-chain $=$ union of k-simplices
sum $c + c'$ $=$ symmetric difference

The space of k-chains

Let K be a d-dimensional simplicial complex. Let $k \in \{0, 1, \cdots, d\}$ and $\{\sigma_1, \cdots, \sigma_p\}$ be the set of k-simplices of K.

k-chain:

$$c = \sum_{i=1}^{p} \varepsilon_i \sigma_i \quad \text{with} \quad \varepsilon_i \in \mathbb{Z}/2\mathbb{Z} = \{0, 1\}$$

Sum of k-chains:

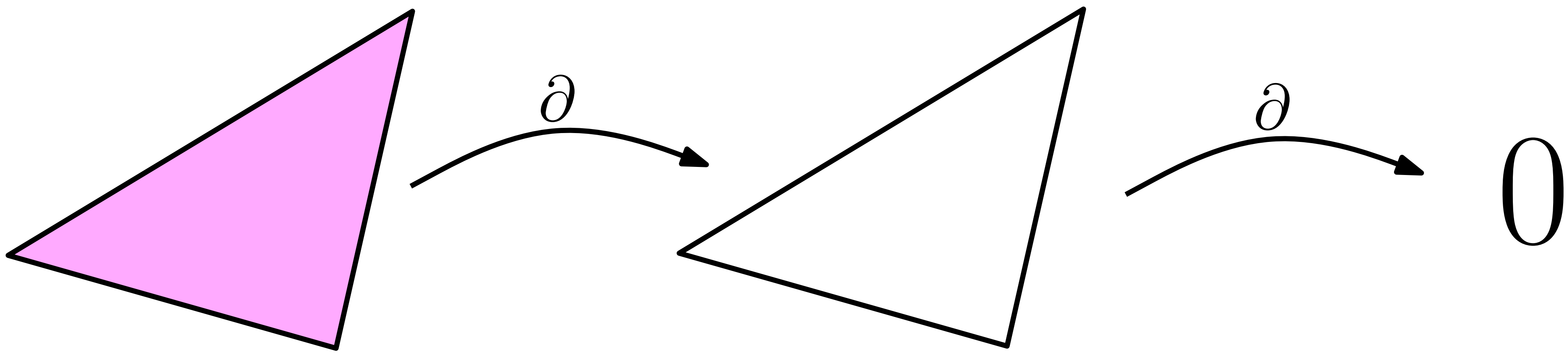
$$c + c' = \sum_{i=1}^{p} (\varepsilon_i + \varepsilon'_i)\sigma_i \quad \text{and} \quad \lambda.c = \sum_{i=1}^{p} (\lambda \varepsilon'_i)\sigma_i$$

The space $\mathcal{C}_k(K)$ of k-chains is a $\mathbb{Z}/2$-vector space

where the sums $\varepsilon_i + \varepsilon'_i$ and the products $\lambda\varepsilon_i$ are modulo 2.

Geometric interpretation:
k-chain $=$ union of k-simplices
sum $c + c'$ $=$ symmetric difference

The boundary operator



The boundary $\partial\sigma$ of a k-simplex σ is the sum of its $(k-1)$-faces. This is a $(k-1)$-chain.

$$If \sigma = [v_0, \cdots, v_k] \;\; \text{then} \;\; \partial\sigma = \sum_{i=0}^{k} [v_0 \cdots \hat{v}_i \cdots v_k]$$

$(-1)^i$ if not with coeff in $\mathbb{Z}/2$!

The boundary operator is the linear map defined by

$$\begin{array}{rccl} \partial: & \mathcal{C}_k(K) & \to & \mathcal{C}_{k-1}(K) \\ & c & \to & \partial c = \sum_{\sigma \in c} \partial\sigma \end{array}$$

Fundamental property of the boundary operator

$$\partial\partial := \partial \circ \partial = 0$$

Proof: by linearity it is just necessary to prove it for a simplex.

$$\begin{aligned}
\partial\partial\sigma &= \partial\left(\sum_{i=0}^{k}[v_0 \cdots \hat{v}_i \cdots v_k]\right) \\
&= \sum_{i=0}^{k}\partial[v_0 \cdots \hat{v}_i \cdots v_k] \\
&= \sum_{j<i}[v_0 \cdots \hat{v}_j \cdots \hat{v}_i \cdots v_k] + \sum_{j>i}[v_0 \cdots \hat{v}_i \cdots \hat{v}_j \cdots v_k] \\
&= 0
\end{aligned}$$

Cycles and boundaries

The chain complex associated to a complex K of dimension d

$$\emptyset \to \mathcal{C}_d(K) \xrightarrow{\partial} \mathcal{C}_{d-1}(K) \xrightarrow{\partial} \cdots \mathcal{C}_{k+1}(K) \xrightarrow{\partial} \mathcal{C}_k(K) \xrightarrow{\partial} \cdots \mathcal{C}_1(K) \xrightarrow{\partial} \mathcal{C}_0(K) \xrightarrow{\partial} \emptyset$$

k-cycles:

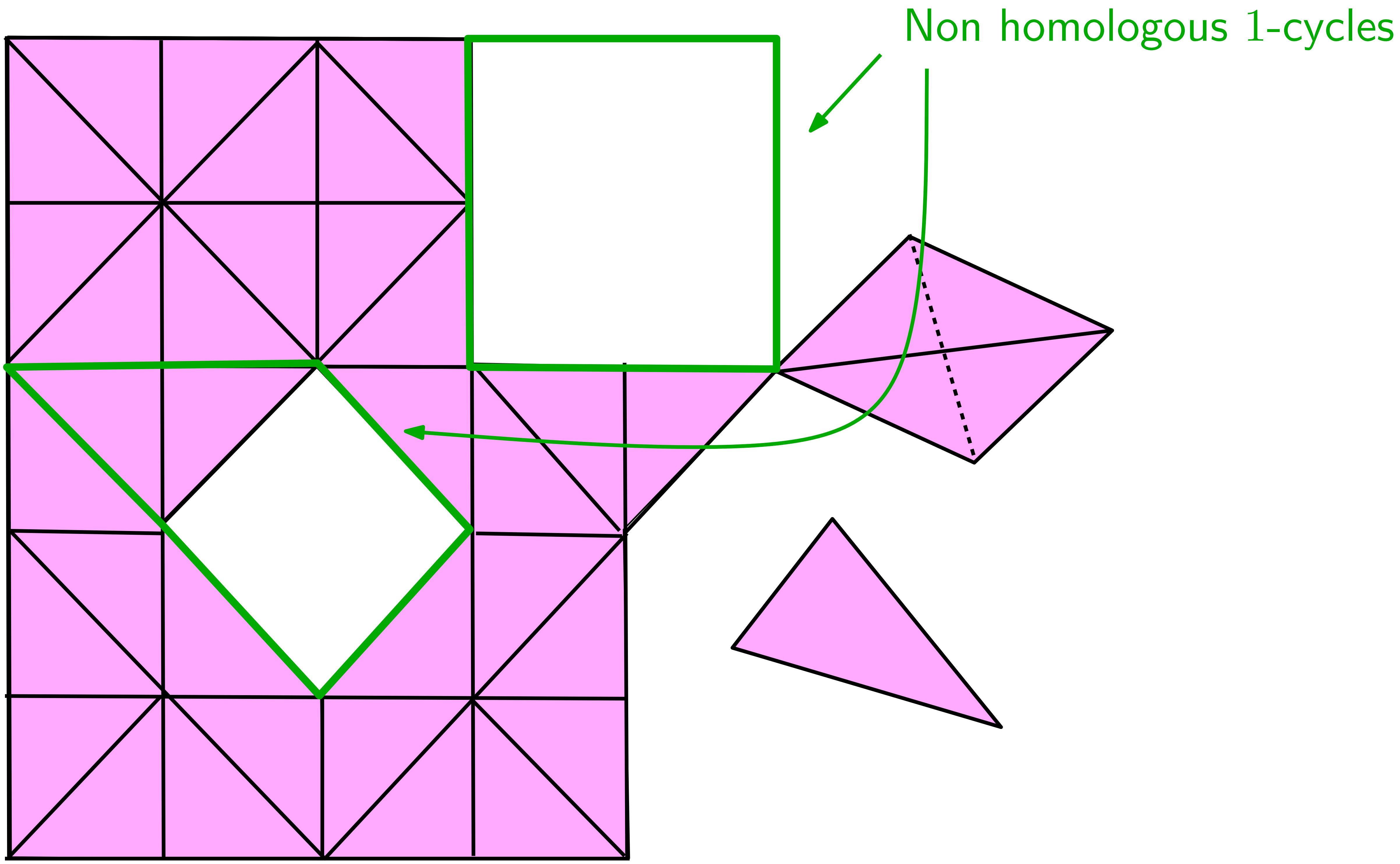
$$Z_k(K) := \ker(\partial : \mathcal{C}_k \to \mathcal{C}_{k-1}) = \{c \in \mathcal{C}_k : \partial c = \emptyset\}$$

k-boundaries:

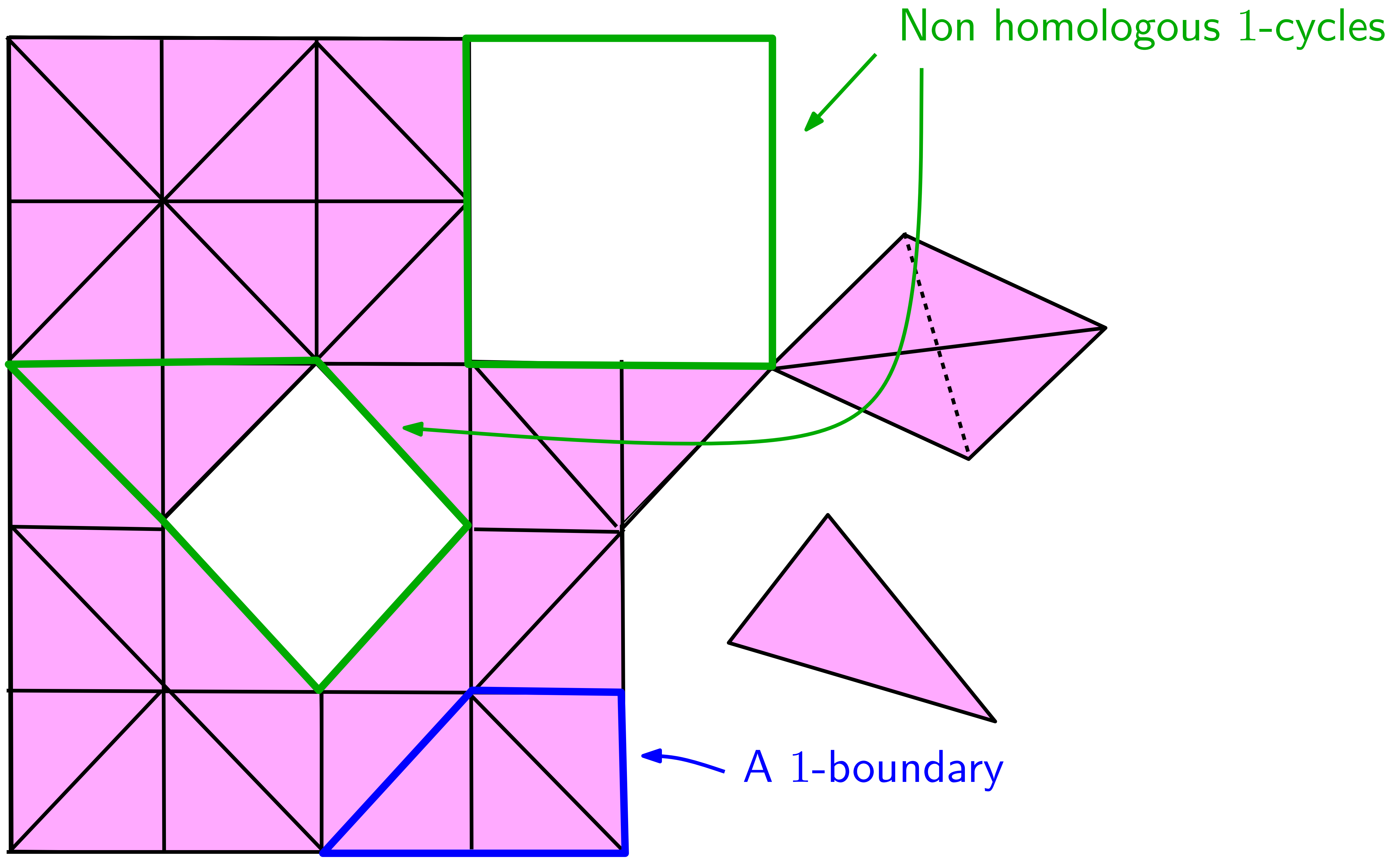
$$B_k(K) := im(\partial : \mathcal{C}_{k+1} \to \mathcal{C}_k) = \{c \in \mathcal{C}_k : \exists c' \in \mathcal{C}_{k+1}, c = \partial c'\}$$

$$B_k(K) \subset Z_k(K) \subset \mathcal{C}_k(K)$$

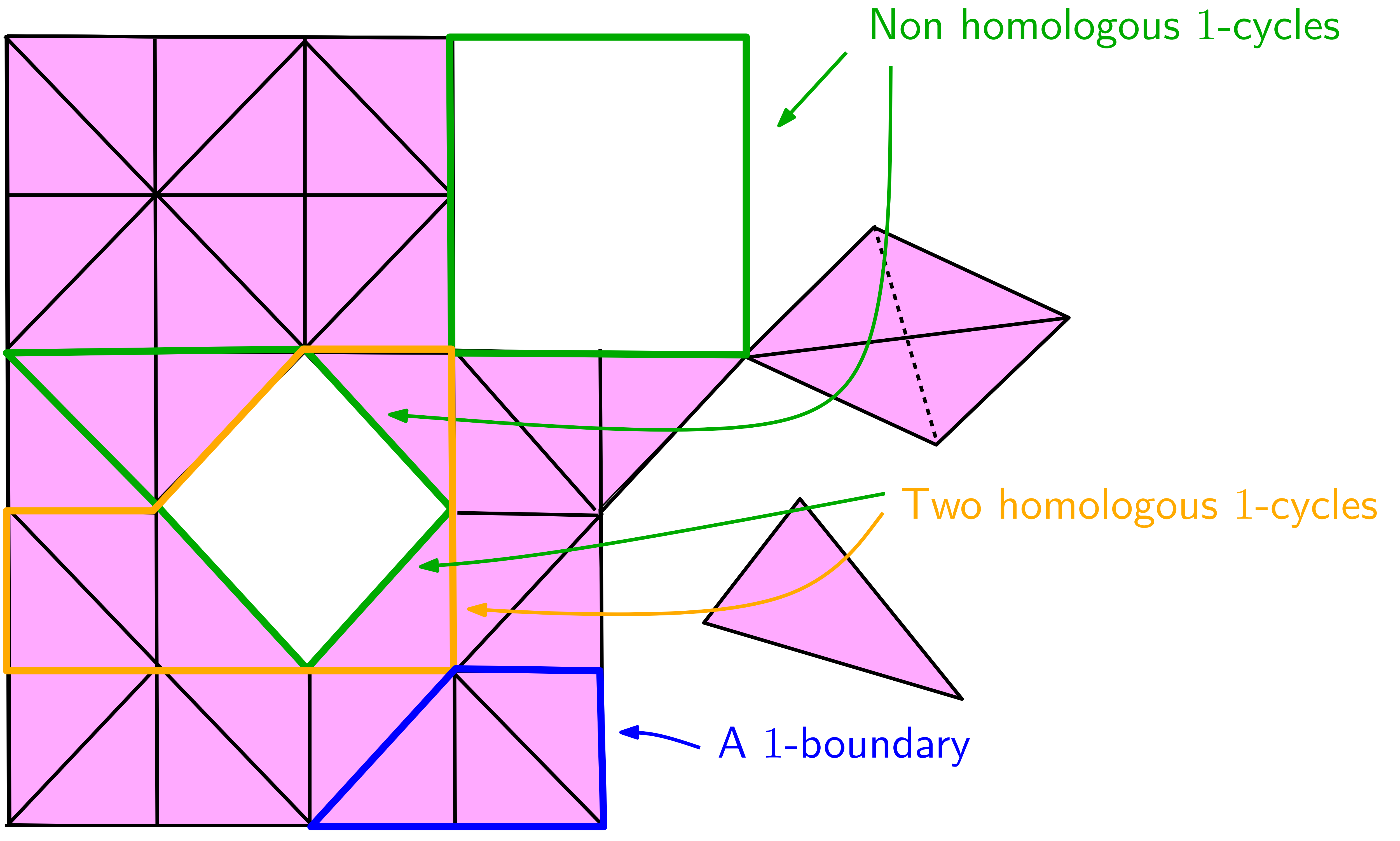
Cycles and boundaries



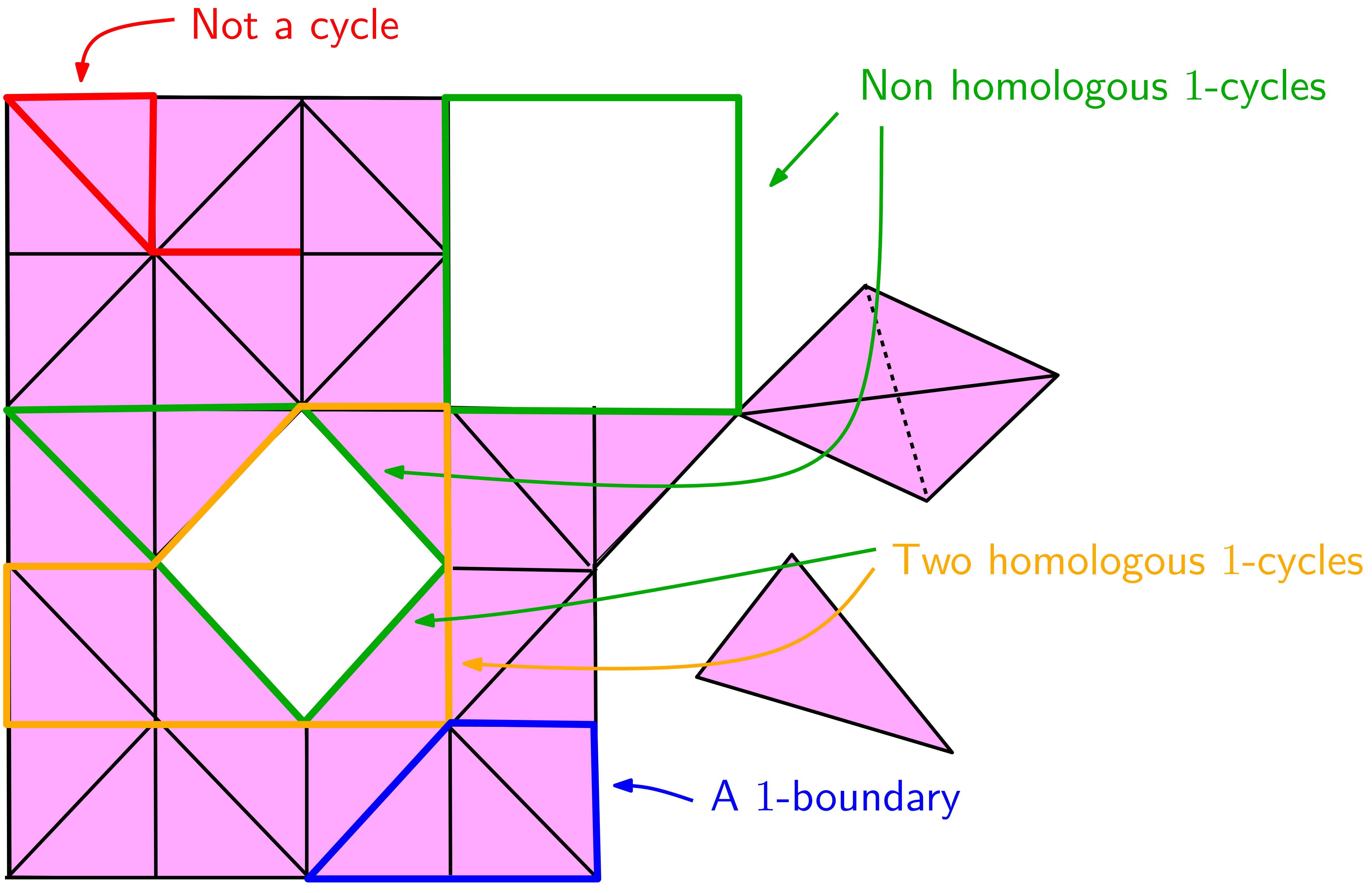
Cycles and boundaries



Cycles and boundaries



Cycles and boundaries



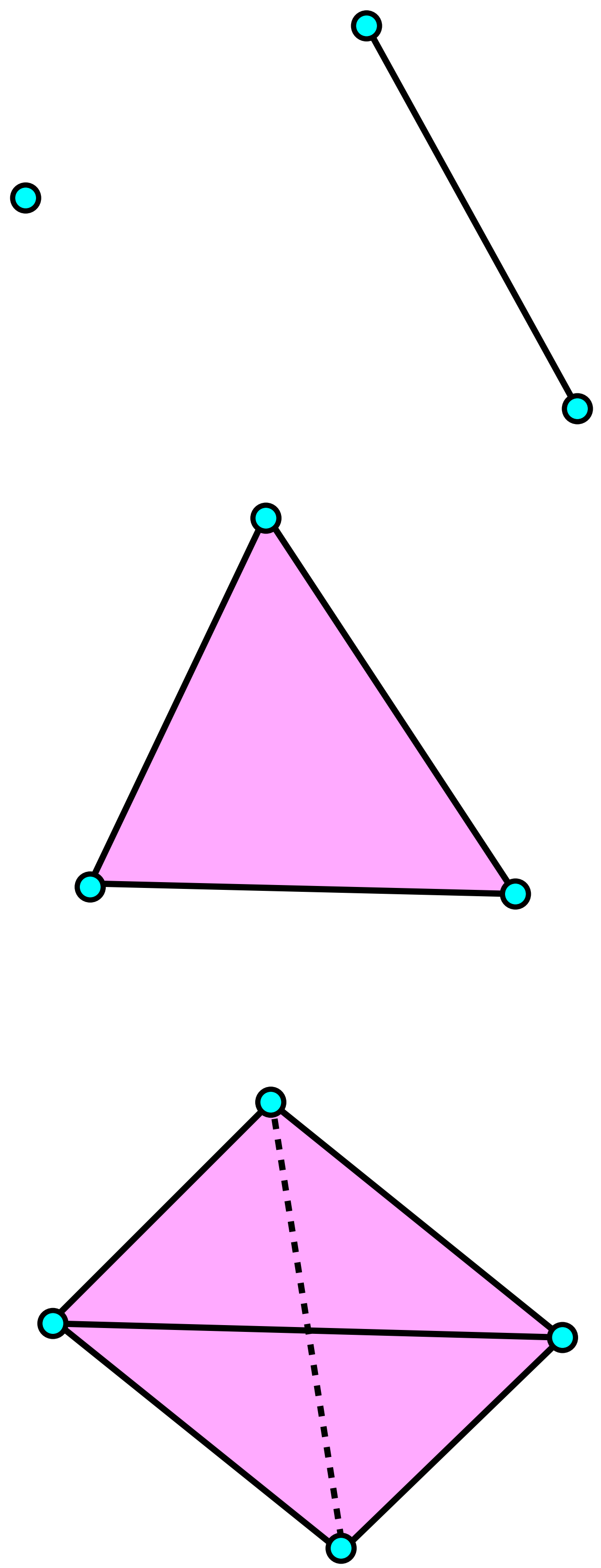
Homology groups and Betti numbers

$$B_k(K) \subset Z_k(K) \subset \mathcal{C}_k(K)$$

- The k^{th} homology group of K: $H_k(K) = Z_k / B_k$
- Tout each cycle $c \in Z_k(K)$ corresponds its homology class $c + B_k(K) = \{c + b : b \in B_k(K)\}$.
- Two cycles c, c' are homologous if they are in the same homology class: $\exists b \in B_k(K)$ s. t. $b = c' - c (= c' + c)$.
- The k^{th} Betti number of K: $\beta_k(K) = \dim(H_k(K))$.

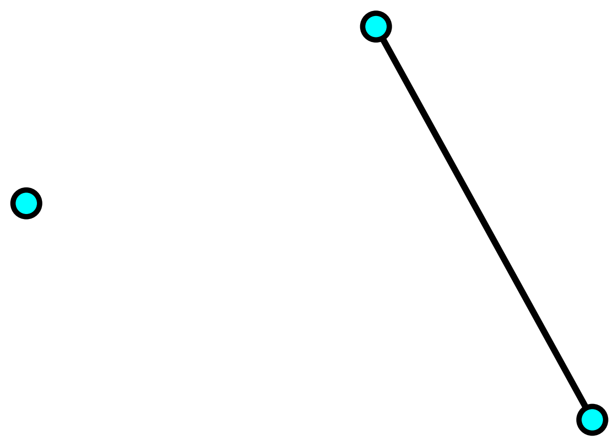
Elementary examples

Remark: $\beta_0 =$ number of connected components of K

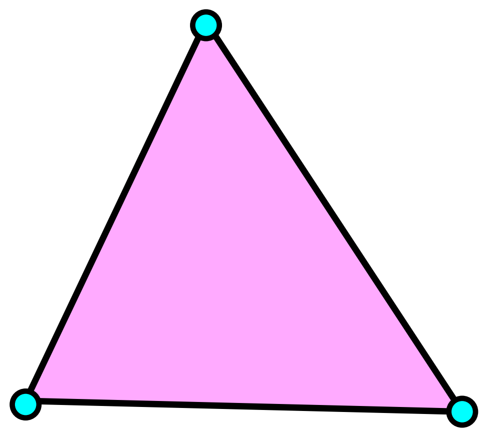


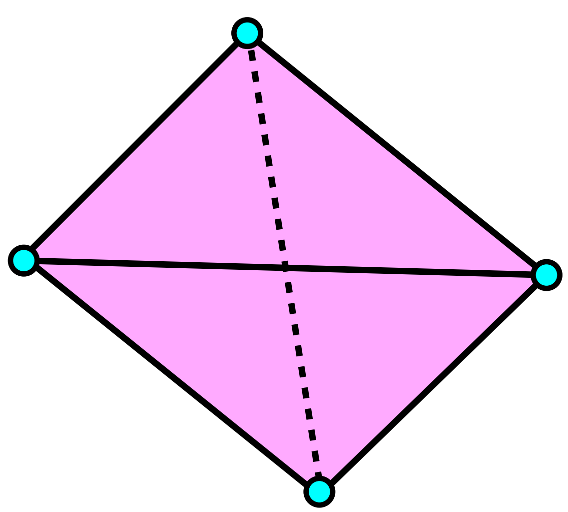
Elementary examples

Remark: $\beta_0 =$ number of connected components of K



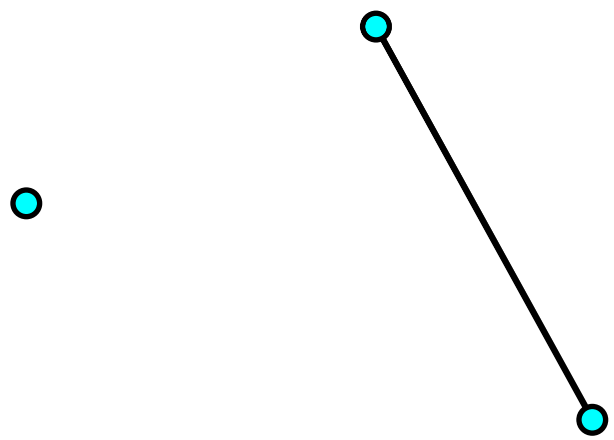
$$\beta_0 = 2$$
$$\beta_1 = 0$$
$$\beta_2 = 0$$



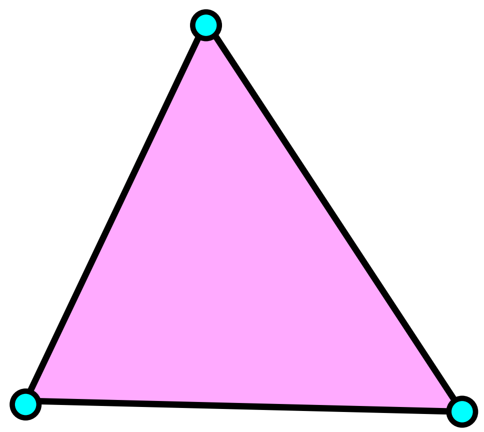


Elementary examples

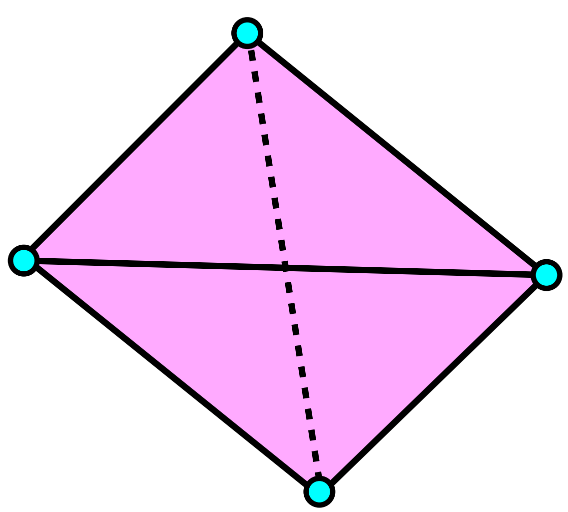
Remark: $\beta_0 =$ number of connected components of K



$\beta_0 = 2$
$\beta_1 = 0$
$\beta_2 = 0$

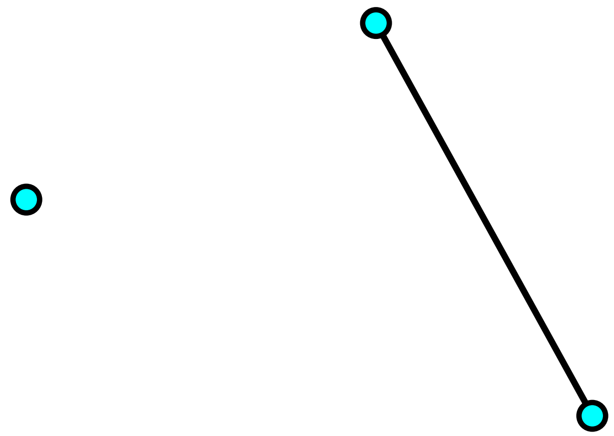


$\beta_0 = 1$
$\beta_1 = 0$
$\beta_2 = 0$

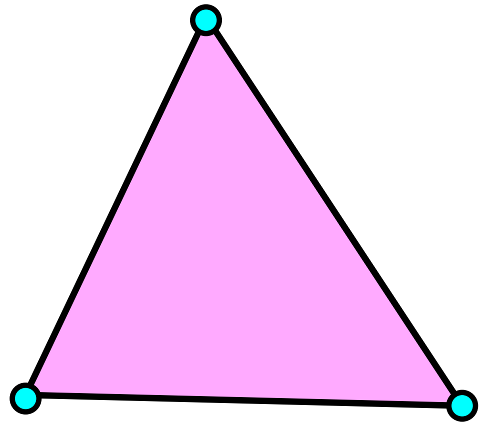


Elementary examples

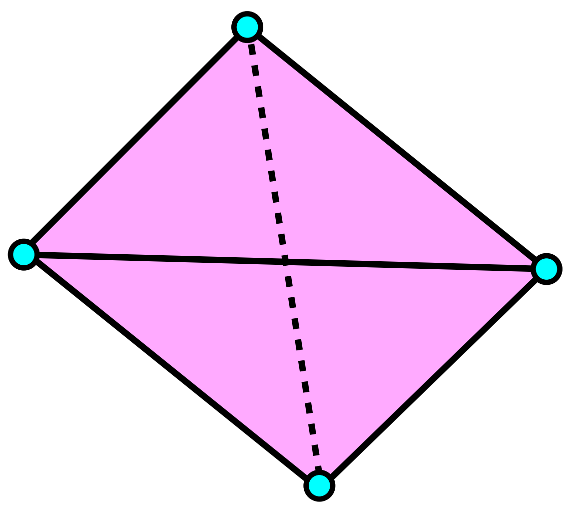
Remark: $\beta_0 =$ number of connected components of K



$\beta_0 = 2$
$\beta_1 = 0$
$\beta_2 = 0$

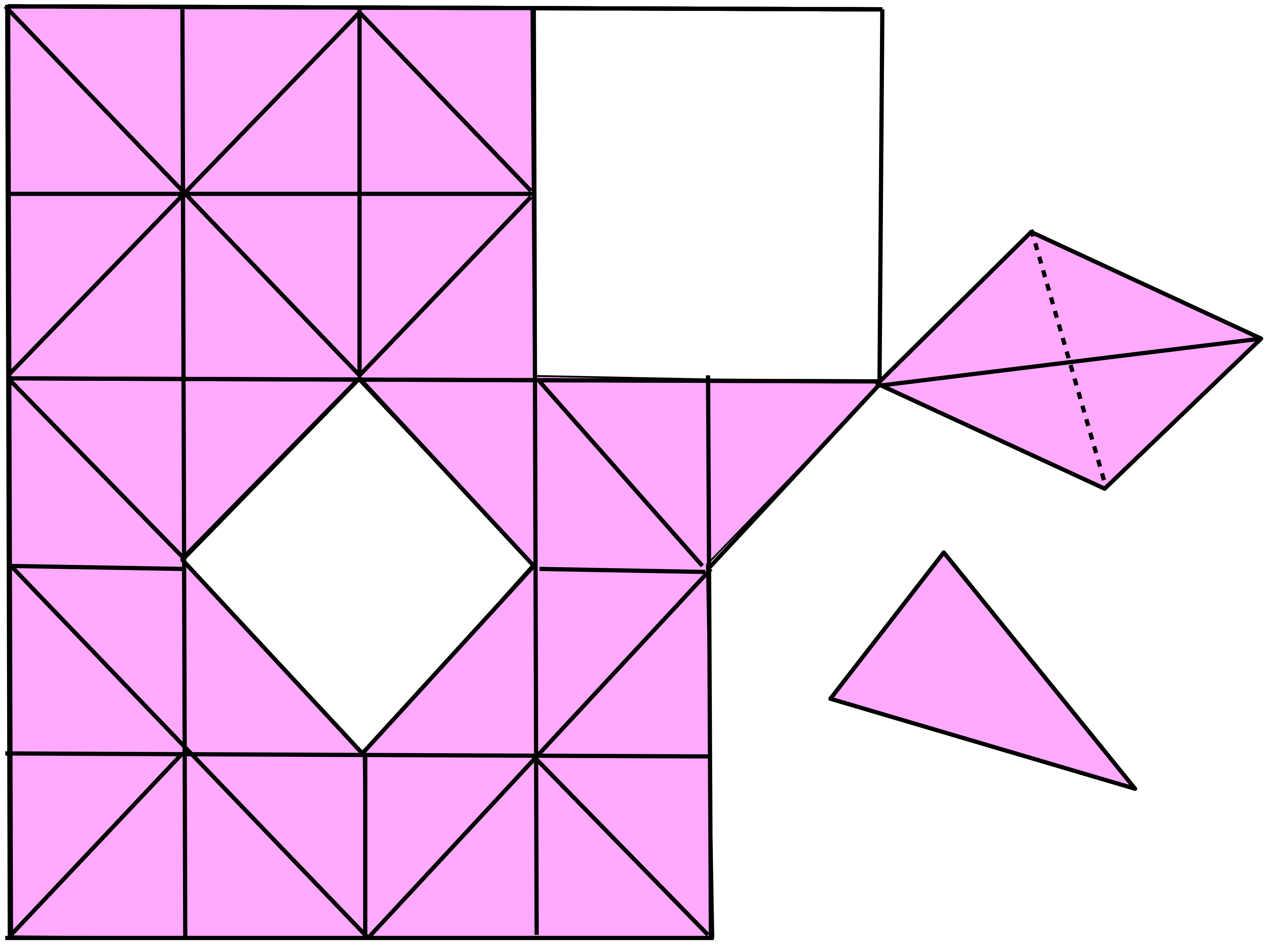


$\beta_0 = 1$
$\beta_1 = 0$
$\beta_2 = 0$

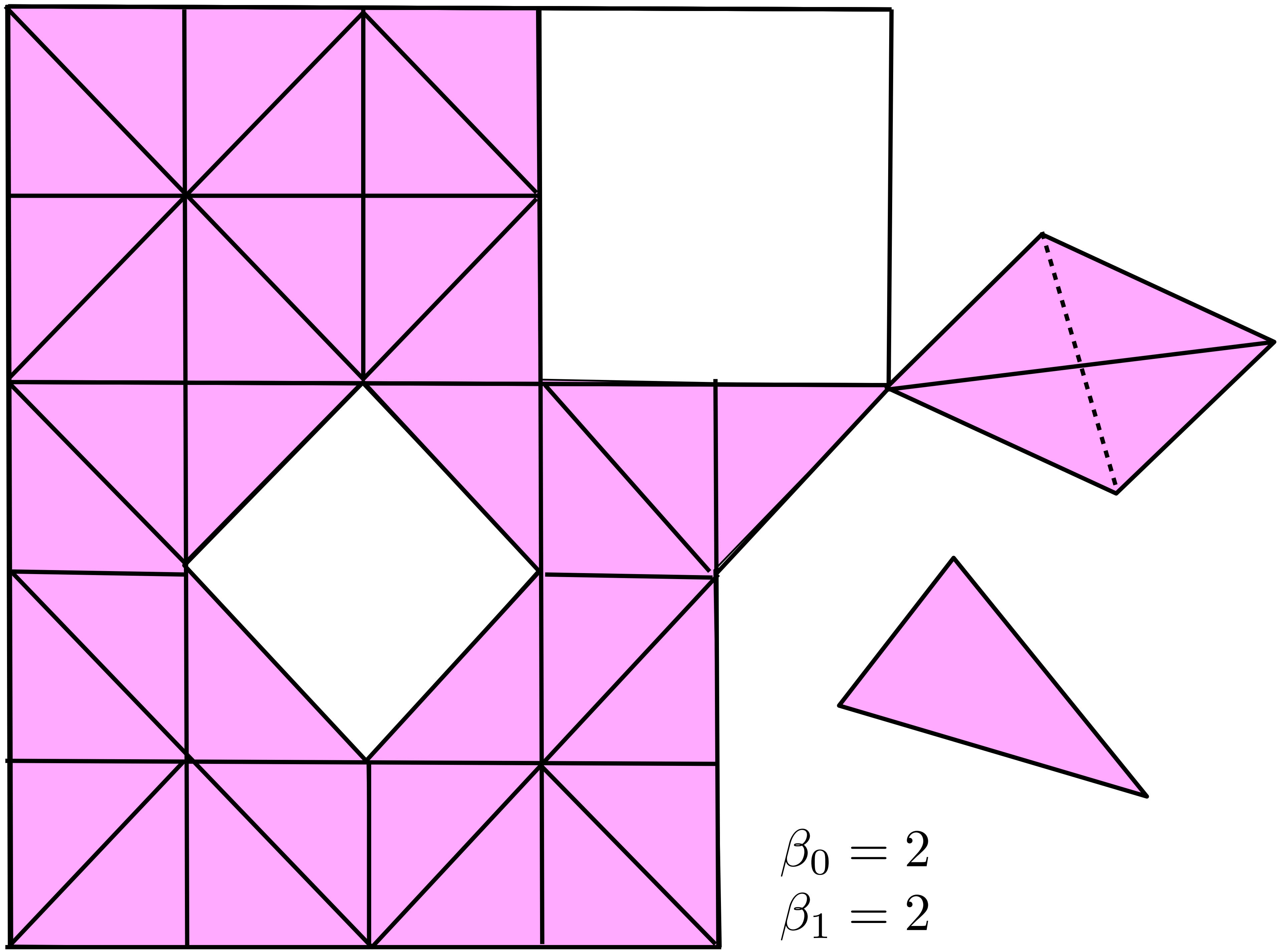


$\beta_0 = 1$
$\beta_1 = 0$
$\beta_2 = 1$ if empty and $\beta_2 = 0$ if filled
$\beta_3 = 0$

Elementary examples

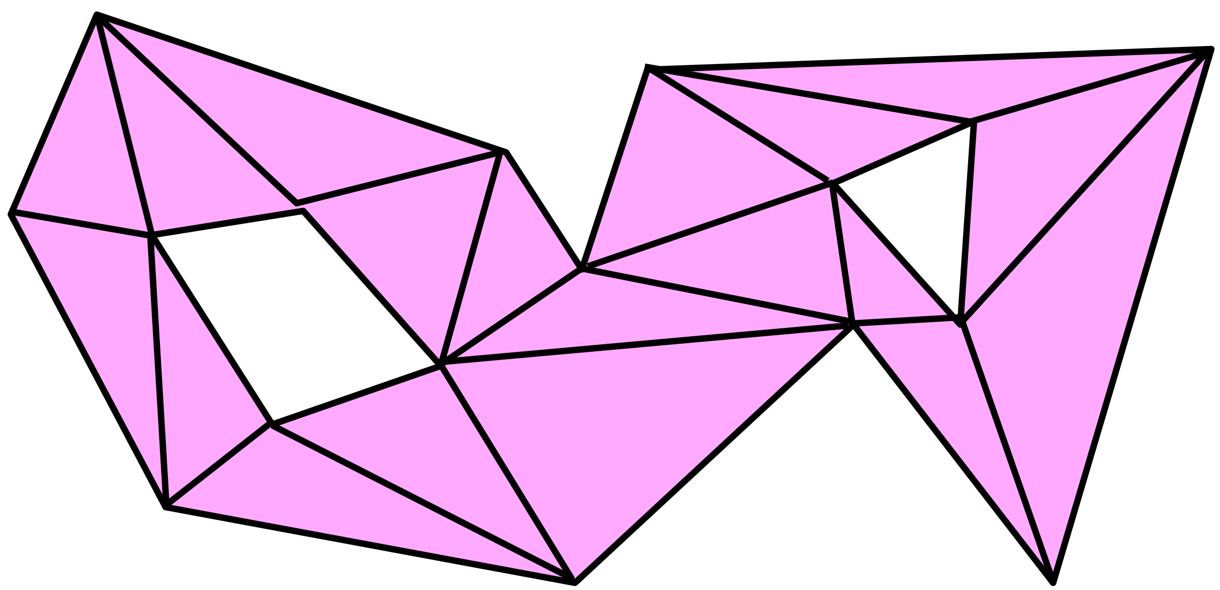


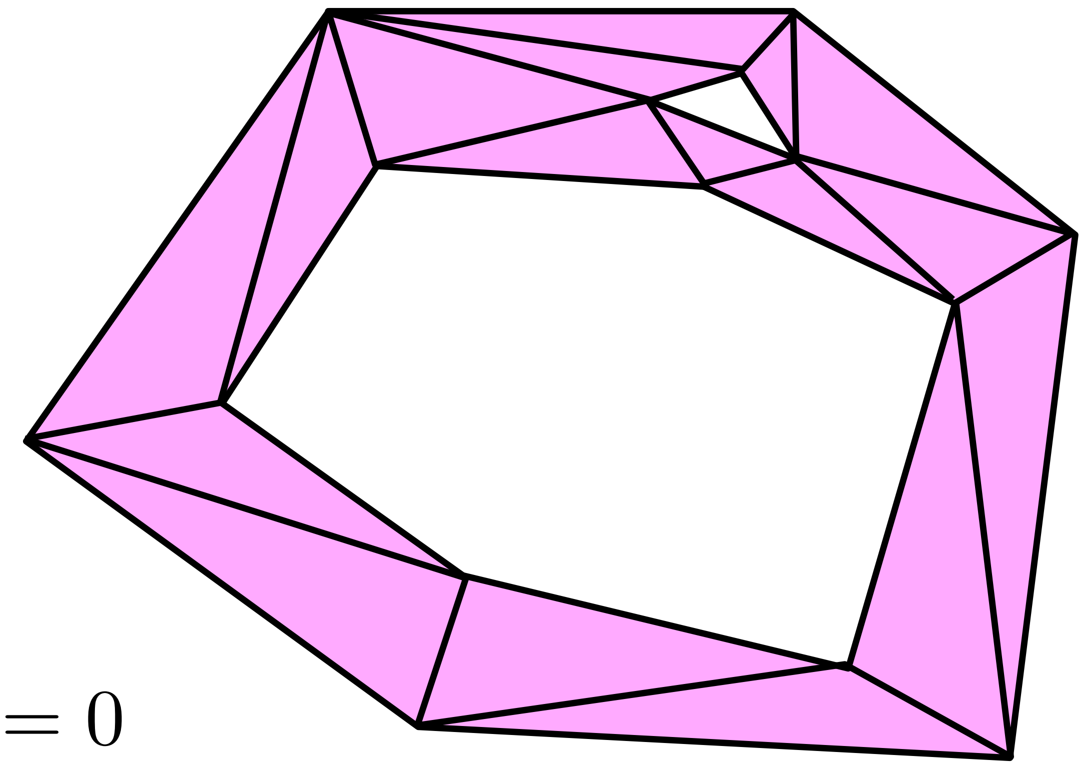
Elementary examples



$\beta_0 = 2$
$\beta_1 = 2$
$\beta_2 = 1$ if empty and $\beta_2 = 0$ if filled
$\beta_3 = 0$

Topological invariance and singular homology



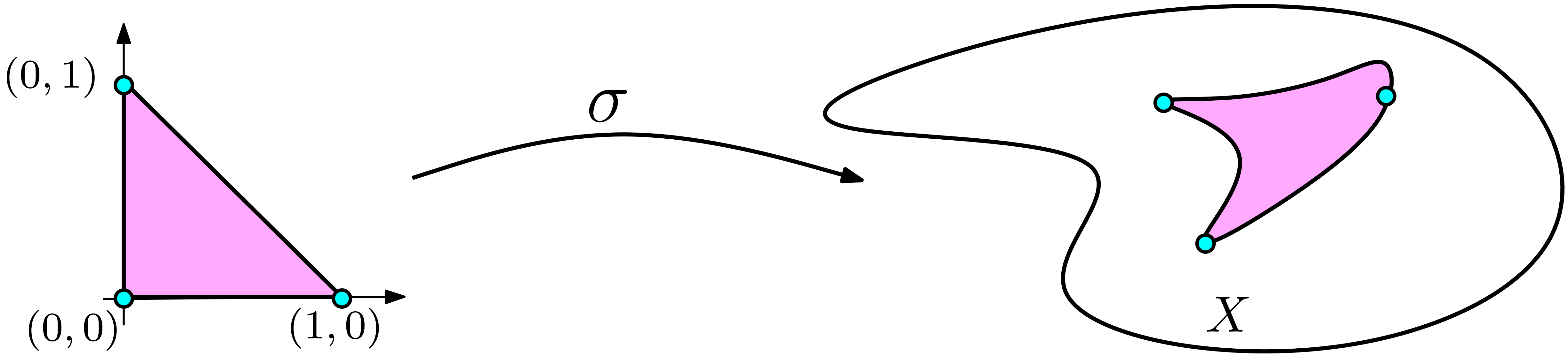


$\beta_0 = 1,\ \beta_1 = 2, \beta_2 = 0$

Theorem: If K and K' are two simplicial complexes such that $|K|$ and $|K'|$ are homeomorphic, then their homology groups are isomorphic and their Betti numbers are equal.

- This is a classical result in algebraic topology but the proof is not obvious.
- Rely on the notion of singular homology $\rightarrow$ defined for any topological space.

Topological invariance and singular homology



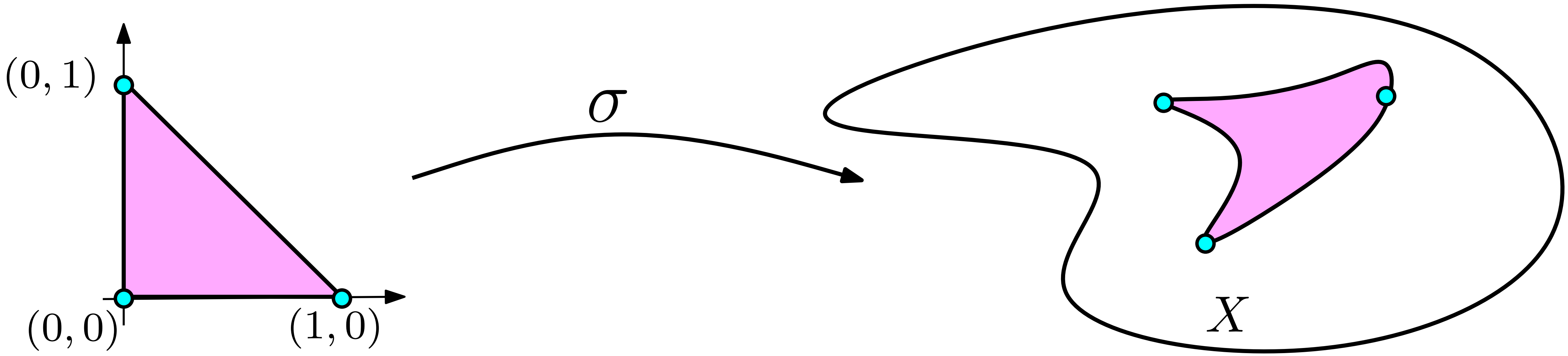
Let Δ_k be the standard simplex in $\mathbb{R}^k$. A singular k-simplex in a topological space X is a continuous map $\sigma : \Delta_k \to X$.

The same construction as for simplicial homology can be done with singular complexes $\to$ Singular homology

Important properties:

- Singular homology is defined for any topological space X.
- If X is homotopy equivalent to the support of a simplicial complex, then the singular and simplicial homology coincide!

Topological invariance and singular homology



Let Δ_k be the standard simplex in $\mathbb{R}^k$. A singular k-simplex in a topological space X is a continuous map $\sigma : \Delta_k \to X$.

Homology and continuous maps:

- if $f : X \to Y$ is a continuous map and $\sigma : \Delta_k \to X$ a simplex in X, then $f \circ \sigma : \Delta_k \to Y$ is a simplex in $Y \Rightarrow f$ induces a linear maps between homology groups:

$$f_\sharp : H_k(X) \to H_k(Y)$$

- if $f : X \to Y$ is an homeomorphism or an homotopy equivalence then $f_\sharp$ is an isomorphism.

An algorithm for geometric inference

- $X \subset \mathbb{R}^d$ be a compact set such that $\mathrm{wfs}(X) > 0$.
- $L \subset \mathbb{R}^d$ be a finite set such that $d_H(X, L) < \varepsilon$ for some $\varepsilon > 0$.

An algorithm for geometric inference

- $X \subset \mathbb{R}^d$ be a compact set such that wfs$(X) > 0$.
- $L \subset \mathbb{R}^d$ be a finite set such that $d_H(X, L) < \varepsilon$ for some $\varepsilon > 0$.

Goal: Compute the Betti numbers of X^r for $0 < r < \mathrm{wfs}(X)$ from L.

An algorithm for geometric inference

- $X \subset \mathbb{R}^d$ be a compact set such that wfs$(X) > 0$.
- $L \subset \mathbb{R}^d$ be a finite set such that $d_H(X, L) < \varepsilon$ for some $\varepsilon > 0$.

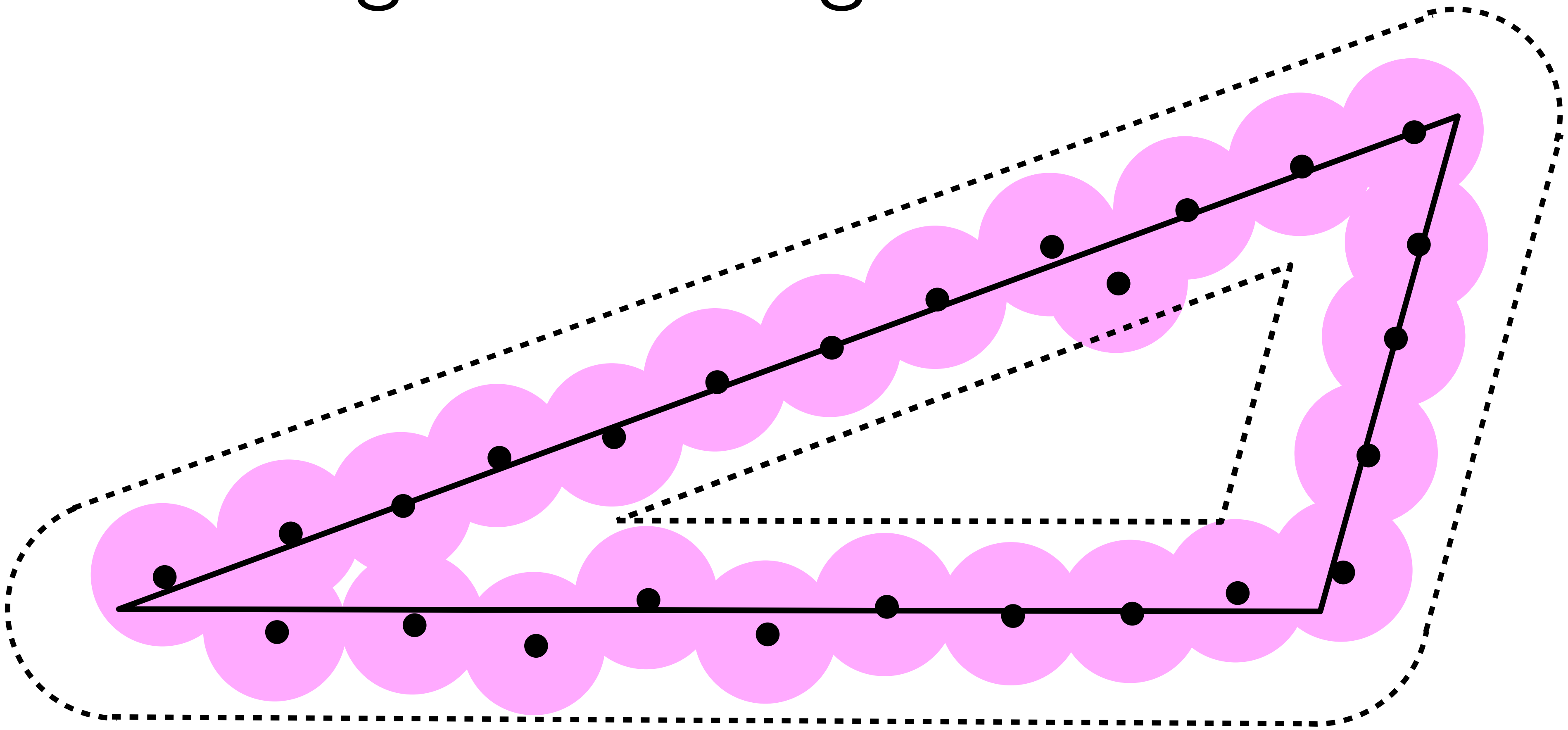
Goal: Compute the Betti numbers of X^r for $0 < r < \mathrm{wfs}(X)$ from L.

Theorem: [CL'05 - CSEH'05]
Assume that wfs$(X) > 4\varepsilon$. For $\alpha > 0$ s.t. $\alpha + 4\varepsilon <$ wfs(X), let $i : L^{\alpha+\varepsilon} \hookrightarrow L^{\alpha+3\varepsilon}$ be the canonical inclusion.For any $0 < r < \mathrm{wfs}(X)$,

$$H_k(X^r) \cong im\left(i_* : H_k(L^{\alpha+\varepsilon}) \to H_k(L^{\alpha+3\varepsilon})\right)$$

An algorithm for geometric inference

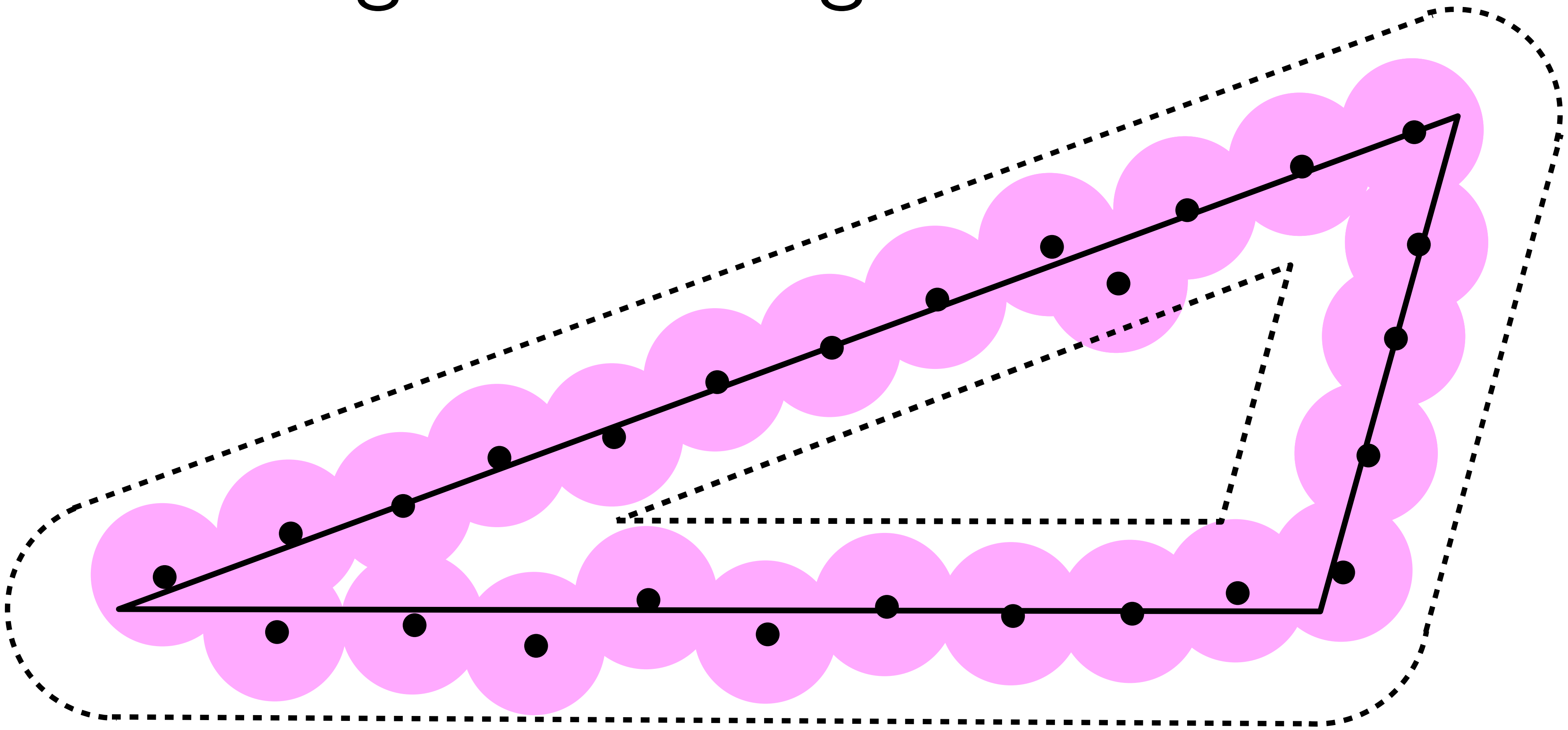
Proof:



For any $\alpha > 0$, $\quad X^{\alpha} \subseteq L^{\alpha+\varepsilon} \subseteq X^{\alpha+2\varepsilon} \subseteq L^{\alpha+3\varepsilon} \subseteq X^{\alpha+4\varepsilon} \subseteq \cdots$

An algorithm for geometric inference

Proof:



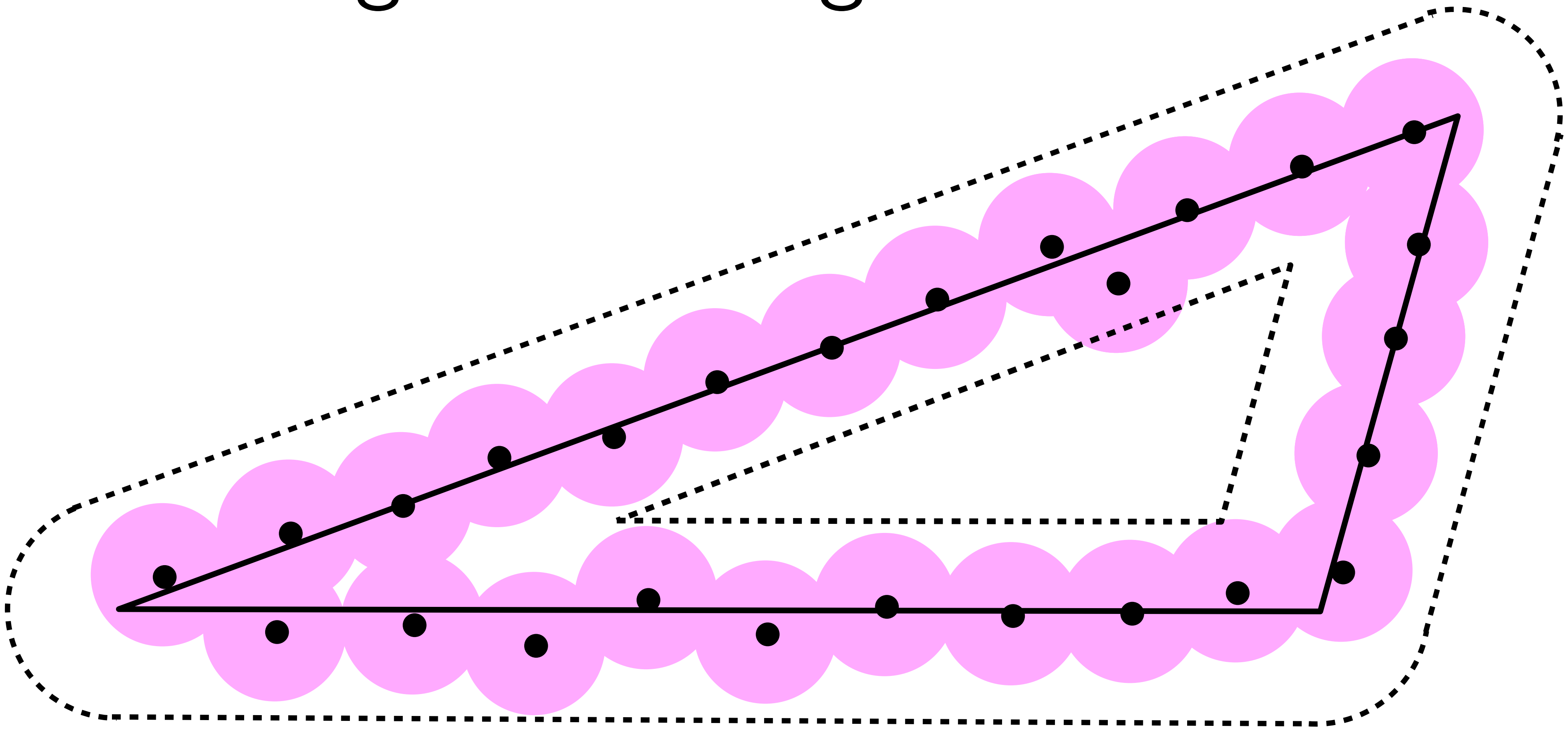
For any $\alpha > 0$, $\quad X^{\alpha} \subseteq L^{\alpha+\varepsilon} \subseteq X^{\alpha+2\varepsilon} \subseteq L^{\alpha+3\varepsilon} \subseteq X^{\alpha+4\varepsilon} \subseteq \cdots$

At homology level:

$$H_k(X^{\alpha}) \to H_k(L^{\alpha+\varepsilon}) \to H_k(X^{\alpha+2\varepsilon}) \to H_k(L^{\alpha+3\varepsilon}) \to H_k(X^{\alpha+4\varepsilon}) \to \cdots$$

An algorithm for geometric inference

Proof:



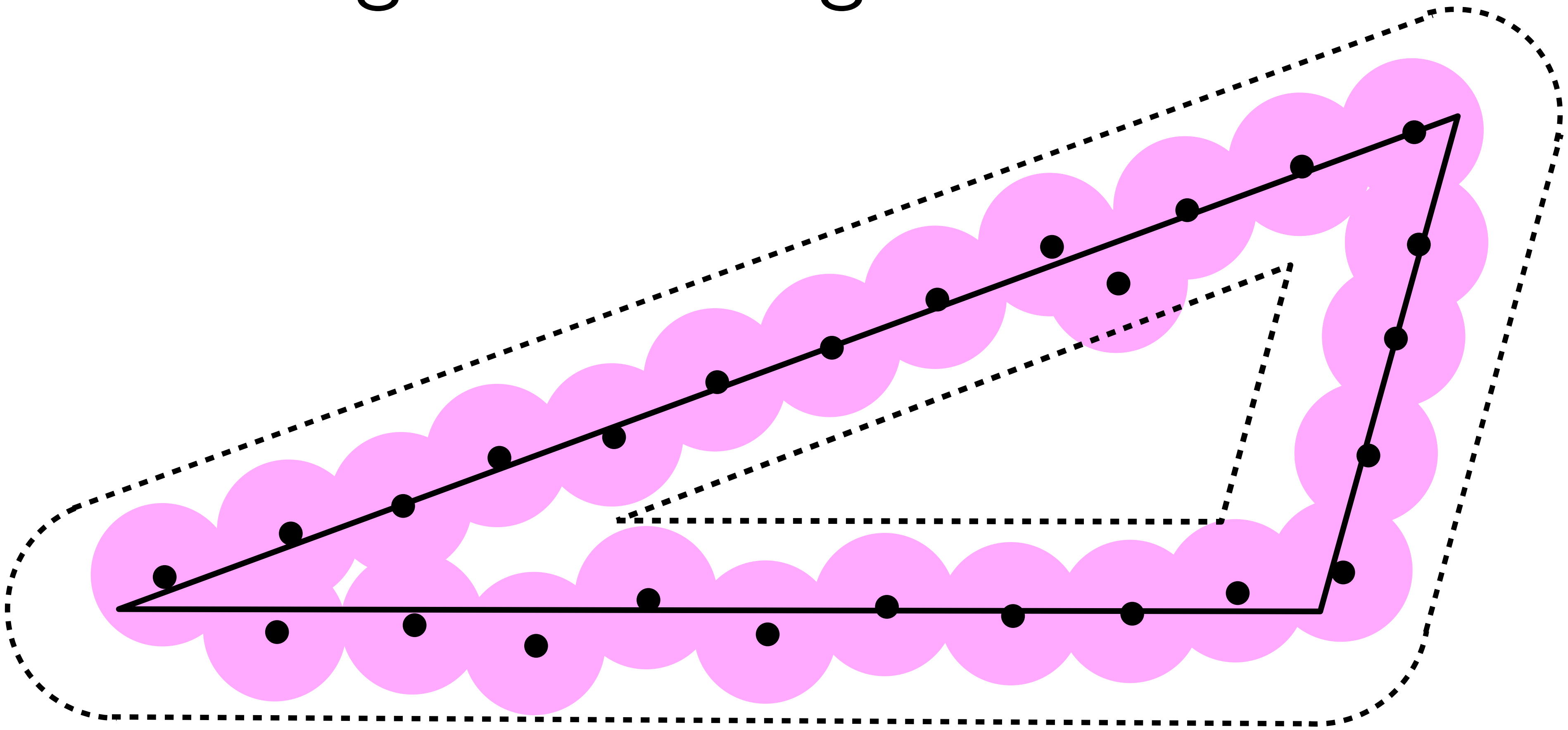
For any $\alpha > 0$, $\quad X^\alpha \subseteq L^{\alpha+\varepsilon} \subseteq X^{\alpha+2\varepsilon} \subseteq L^{\alpha+3\varepsilon} \subseteq X^{\alpha+4\varepsilon} \subseteq \cdots$

At homology level:

$$\begin{tikzcd}
H_k(X^\alpha) \arrow[r] \arrow[rr, bend right, "\text{isomorphism}"'] & H_k(L^{\alpha+\varepsilon}) \arrow[r] \arrow[rr, bend left, "\text{rank} = \dim H_k(X^\alpha)"] & H_k(X^{\alpha+2\varepsilon}) \arrow[r] \arrow[rr, bend right, "\text{isomorphism}"'] & H_k(L^{\alpha+3\varepsilon}) \arrow[r] & H_k(X^{\alpha+4\varepsilon}) \arrow[r] & \cdots
\end{tikzcd}$$

An algorithm for geometric inference

Proof:



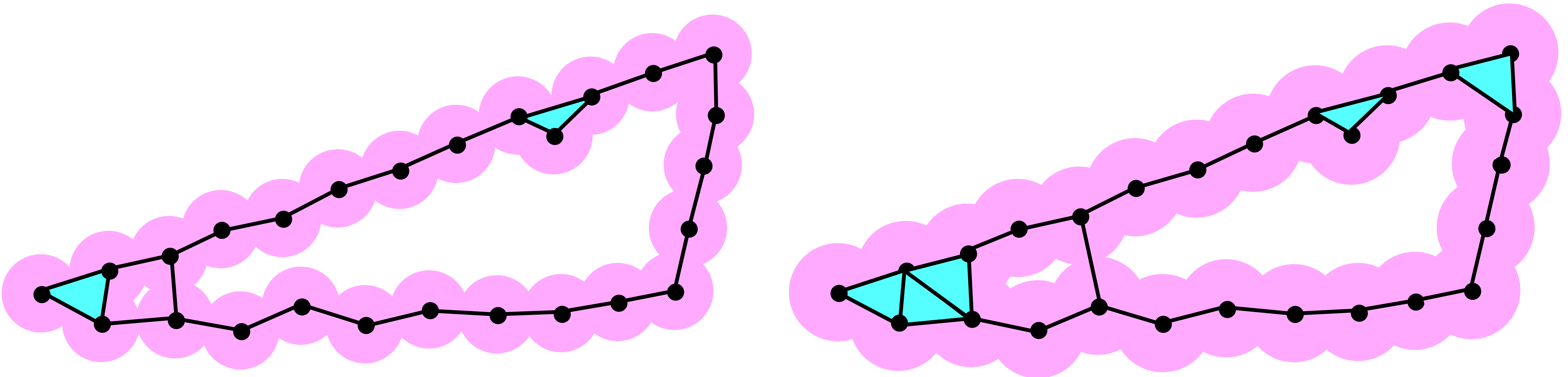
For any $\alpha > 0$, $\quad X^{\alpha} \subseteq L^{\alpha+\varepsilon} \subseteq X^{\alpha+2\varepsilon} \subseteq L^{\alpha+3\varepsilon} \subseteq X^{\alpha+4\varepsilon} \subseteq \cdots$

At homology level:

Cannot be directly computed !

$$\begin{tikzcd}
H_k(X^{\alpha}) \arrow[r] \arrow[rr, bend right, "\text{isomorphism}"'] & H_k(L^{\alpha+\varepsilon}) \arrow[r] \arrow[rr, bend left, "\text{rank} = \dim H_k(X^{\alpha})"] & H_k(X^{\alpha+2\varepsilon}) \arrow[r] \arrow[rr, bend right, "\text{isomorphism}"'] & H_k(L^{\alpha+3\varepsilon}) \arrow[r] & H_k(X^{\alpha+4\varepsilon}) \arrow[r] & \cdots
\end{tikzcd}$$

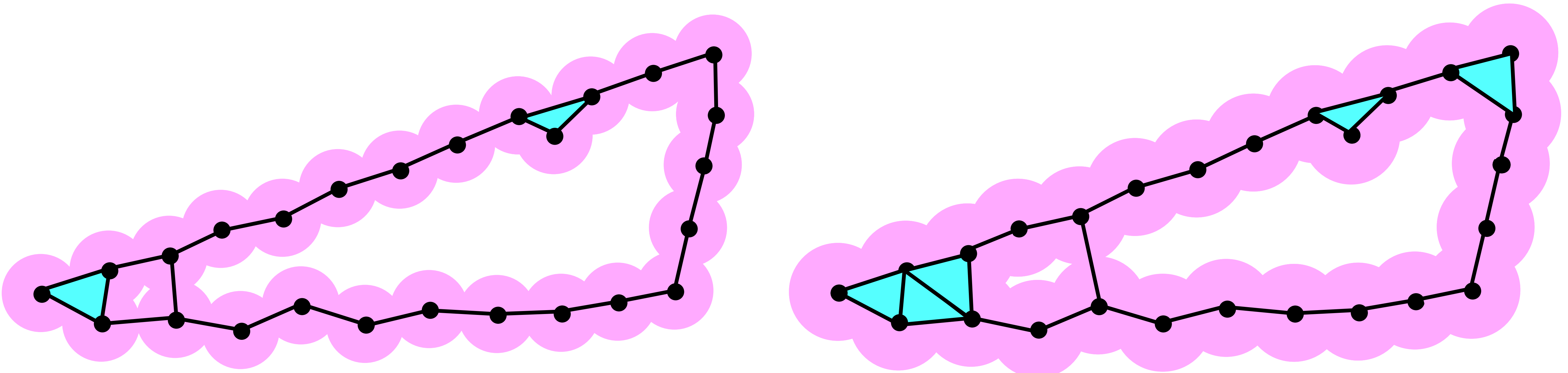
Using the Čech complex



The Čech complex $\mathcal{C}^\alpha(L)$:

$$\text{for } p_0, \cdots p_k \in L, \quad \sigma = [p_0 p_1 \cdots p_k] \in \mathcal{C}^\alpha(L) \;\text{ iff }\; \bigcap_{i=0}^{k} B(p_i, \alpha) \neq \emptyset$$

Using the Čech complex

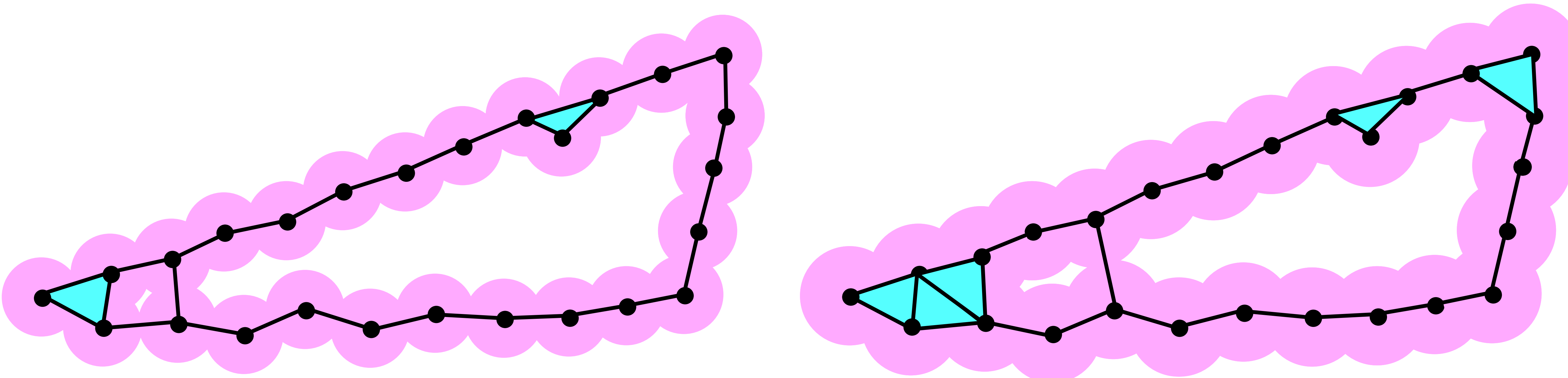


The Čech complex $\mathcal{C}^\alpha(L)$:

$$\text{for } p_0, \cdots p_k \in L, \quad \sigma = [p_0 p_1 \cdots p_k] \in \mathcal{C}^\alpha(L) \quad \text{iff} \quad \bigcap_{i=0}^{k} B(p_i, \alpha) \neq \emptyset$$

Nerve theorem: For any $\alpha > 0$, L^α and $\mathcal{C}^\alpha(L)$ are homotopy equivalent and the homotopy equivalences can be chosen to commute with inclusions.

Using the Čech complex



The Čech complex $\mathcal{C}^\alpha(L)$:

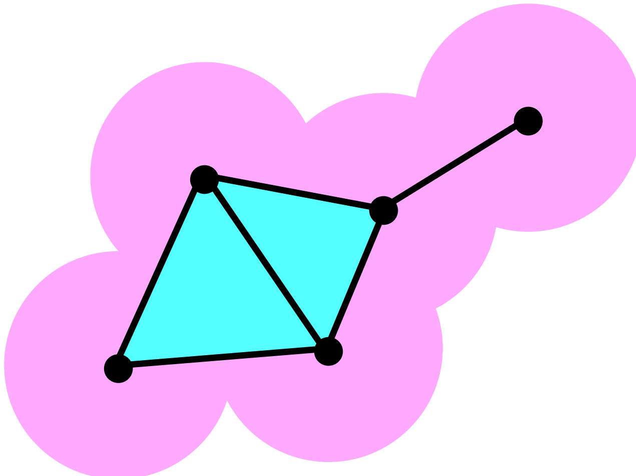
$$\text{for } p_0, \cdots p_k \in L, \quad \sigma = [p_0 p_1 \cdots p_k] \in \mathcal{C}^\alpha(L) \ \text{ iff } \ \bigcap_{i=0}^{k} B(p_i, \alpha) \neq \emptyset$$

Nerve theorem: For any $\alpha > 0$, L^α and $\mathcal{C}^\alpha(L)$ are homotopy equivalent and the homotopy equivalences can be chosen to commute with inclusions.

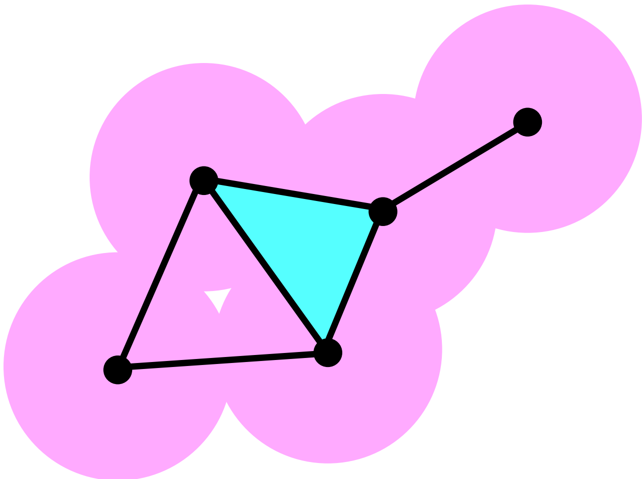
$$\begin{tikzcd}
\cdots \arrow[r] & H_k(L^{\alpha+\varepsilon}) \arrow[r] \arrow[d] & H_k(L^{\alpha+3\varepsilon}) \arrow[r] \arrow[d] & \cdots \\
\cdots \arrow[r] & H_k(\mathcal{C}^{\alpha+\varepsilon}(L)) \arrow[r] & H_k(\mathcal{C}^{\alpha+3\varepsilon}(L)) \arrow[r] & \cdots
\end{tikzcd}$$

Allow to work with simplicial complexes but... still too difficult to compute

Using the Rips complex



Rips vs Čech



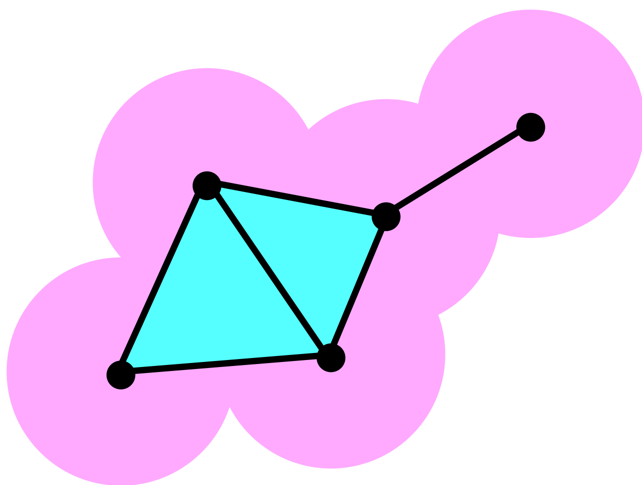
The Rips complex $\mathcal{R}^\alpha(L)$: for $p_0, \cdots p_k \in L$,

$$\sigma = [p_0 p_1 \cdots p_k] \in \mathcal{R}^\alpha(L) \text{ iff } \forall i, j \in \{0, \cdots k\},\ d(p_i, p_j) \leq \alpha$$

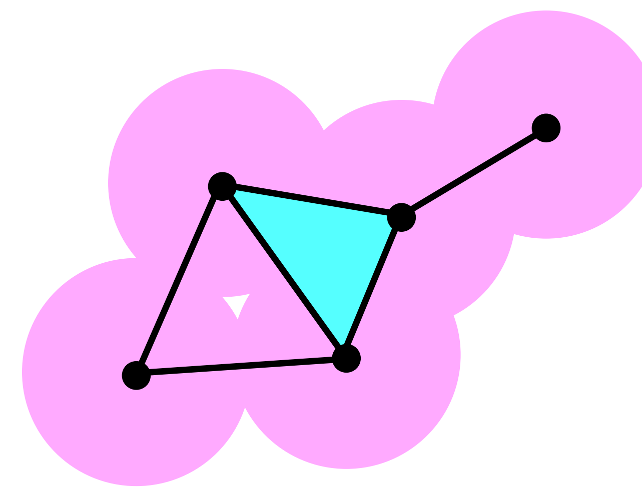
- Easy to compute and fully determined by its 1-skeleton
- Rips-Čech interleaving: for any $\alpha > 0$,

$$\mathcal{C}^{\frac{\alpha}{2}}(L) \subseteq \mathcal{R}^\alpha(L) \subseteq \mathcal{C}^\alpha(L) \subseteq \mathcal{R}^{2\alpha}(L) \subseteq \cdots$$

Using the Rips complex



Rips vs Čech



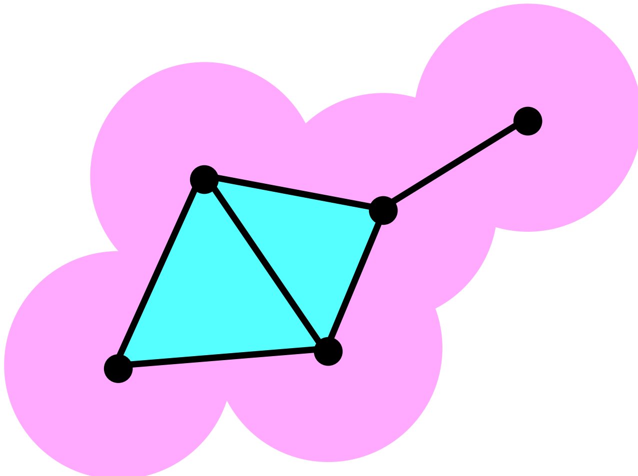
The Rips complex $\mathcal{R}^\alpha(L)$: for $p_0, \cdots p_k \in L$,

$$\sigma = [p_0 p_1 \cdots p_k] \in \mathcal{R}^\alpha(L) \text{ iff } \forall i,j \in \{0, \cdots k\},\ d(p_i, p_j) \leq \alpha$$

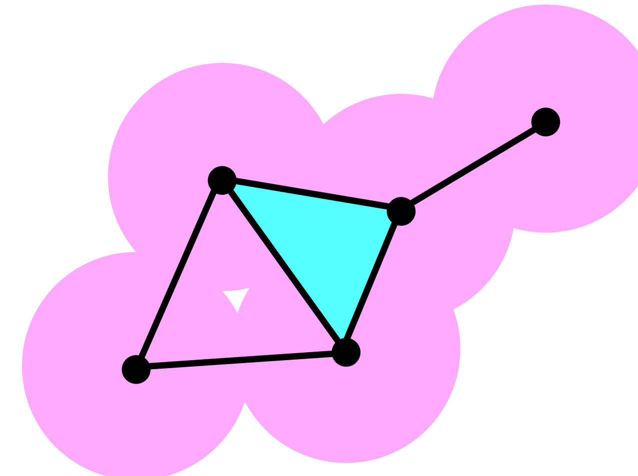
Theorem: [C-Oudot'08]
Let $X \subset \mathbb{R}^d$ be a compact set and $L \subset \mathbb{R}^d$ a finite set such that $d_H(X, L) < \varepsilon$ for some $\varepsilon < \frac{1}{9}$ wfs(X). Then for all $\alpha \in [2\varepsilon, \frac{1}{4}(\mathrm{wfs}(X) - \varepsilon)]$ and all $\lambda \in (0, \mathrm{wfs}(X))$, one has: $\forall k \in \mathbb{N}$

$$\beta_k(X^\lambda) = \dim(H_k(X^\lambda)) = \mathrm{rk}(\mathcal{R}^\alpha(L) \to \mathcal{R}^{4\alpha}(L))$$

Using the Rips complex



Rips vs Čech



The Rips complex $\mathcal{R}^\alpha(L)$: for $p_0, \cdots p_k \in L$,

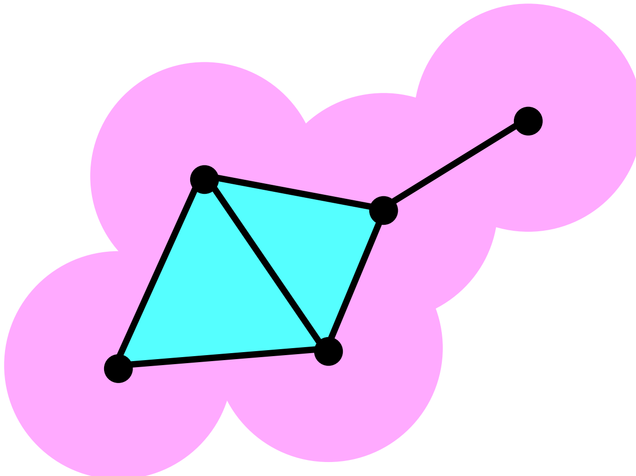
$$\sigma = [p_0 p_1 \cdots p_k] \in \mathcal{R}^\alpha(L) \ \text{ iff } \ \forall i,j \in \{0, \cdots k\},\ d(p_i, p_j) \leq \alpha$$

Theorem: [C-Oudot'08]
Let $X \subset \mathbb{R}^d$ be a compact set and $L \subset \mathbb{R}^d$ a finite set such that $d_H(X, L) < \varepsilon$ for some $\varepsilon < \frac{1}{9}$ wfs(X). Then for all $\alpha \in [2\varepsilon, \frac{1}{4}(\mathrm{wfs}(X) - \varepsilon)]$ and all $\lambda \in (0, \mathrm{wfs}(X))$, one has: $\forall k \in \mathbb{N}$

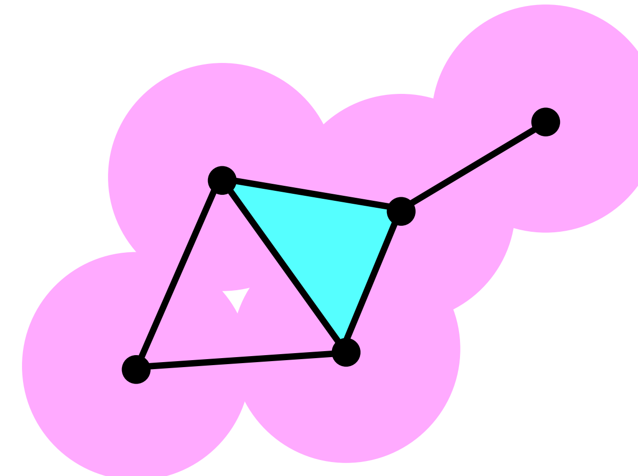
$$\beta_k(X^\lambda) = \dim(H_k(X^\lambda)) = \boxed{\mathrm{rk}(\mathcal{R}^\alpha(L) \to \mathcal{R}^{4\alpha}(L))}$$

Easy to compute using persistence algo.

Using the Rips complex



Rips vs Čech



The Rips complex $\mathcal{R}^\alpha(L)$: for $p_0, \cdots p_k \in L$,

$$\sigma = [p_0 p_1 \cdots p_k] \in \mathcal{R}^\alpha(L) \text{ iff } \forall i, j \in \{0, \cdots k\},\ d(p_i, p_j) \leq \alpha$$

Theorem: [C-Oudot'08]
Let $X \subset \mathbb{R}^d$ be a compact set and $L \subset \mathbb{R}^d$ a finite set such that $d_H(X, L) < \varepsilon$ for some $\varepsilon < \frac{1}{9}$ wfs(X). Then for all $\alpha \in [2\varepsilon, \frac{1}{4}(\mathrm{wfs}(X) - \varepsilon)]$ and all $\lambda \in (0, \mathrm{wfs}(X))$, one has: $\forall k \in \mathbb{N}$

$$\beta_k(X^\lambda) = \dim(H_k(X^\lambda)) = \mathrm{rk}(\mathcal{R}^\alpha(L) \rightarrow \mathcal{R}^{4\alpha}(L))$$

Choice of α when wfs(X) is unknown: see [C-Oudot 2008]

An algorithm to compute Betti numbers

Input: A filtration of a simplicial complex $\emptyset = K^0 \subset K^1 \subset \cdots \subset K^m = K$, s. t. $K^{i+1} = K^i \cup \sigma^{i+1}$ where σ^{i+1} is a simplex of K.

Output: The Betti numbers $\beta_0, \beta_1, \cdots, \beta_d$ of K.

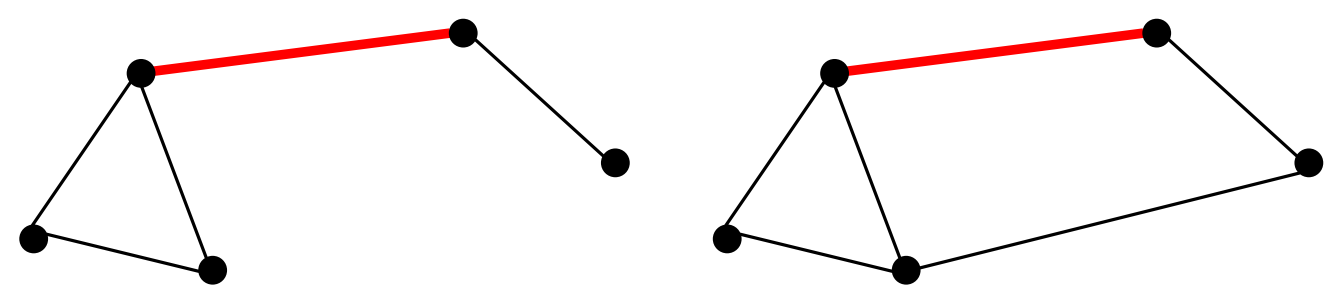
> $\beta_0 = \beta_1 = \cdots = \beta_d = 0$;
> for $i = 1$ to m
> $k = \dim \sigma^i - 1$;
>   if σ^i is contained in a $(k+1)$-cycle in K^i
>   then $\beta_{k+1} = \beta_{k+1} + 1$;
>   else $\beta_k = \beta_k - 1$;
>   end if;
> end for;
> output $(\beta_0, \beta_1, \cdots, \beta_d)$;

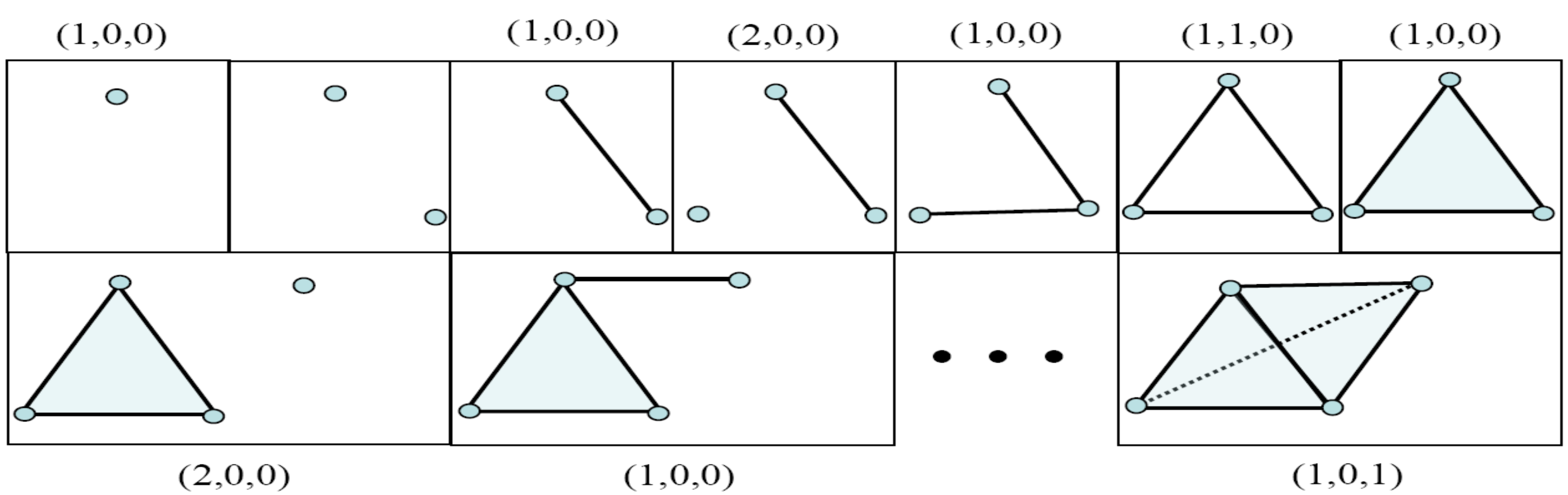
An algorithm to compute Betti numbers

Input: A filtration of a simplicial complex $\emptyset = K^0 \subset K^1 \subset \cdots \subset K^m = K$, s. t. $K^{i+1} = K^i \cup \sigma^{i+1}$ where σ^{i+1} is a simplex of K.

Output: The Betti numbers $\beta_0, \beta_1, \cdots, \beta_d$ of K.

$\beta_0 = \beta_1 = \cdots = \beta_d = 0$;
for $i = 1$ to m
    $k = \dim \sigma^i - 1$;
    if σ^i is contained in a $(k+1)$-cycle in K^i
        then $\beta_{k+1} = \beta_{k+1} + 1$;
        else $\beta_k = \beta_k - 1$;
    end if;
end for;
output $(\beta_0, \beta_1, \cdots, \beta_d)$;



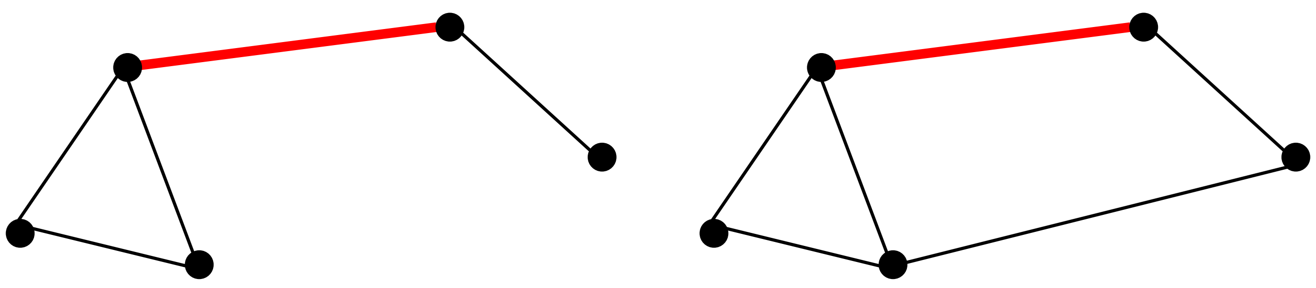


An algorithm to compute Betti numbers

Input: A filtration of a simplicial complex $\emptyset = K^0 \subset K^1 \subset \cdots \subset K^m = K$, s. t. $K^{i+1} = K^i \cup \sigma^{i+1}$ where σ^{i+1} is a simplex of K.

Output: The Betti numbers $\beta_0, \beta_1, \cdots, \beta_d$ of K.

```
β_0 = β_1 = ··· = β_d = 0;
for i = 1 to m
   k = dim σ^i − 1;
   if σ^i is contained in a (k + 1)-cycle in K^i
      then β_{k+1} = β_{k+1} + 1;
      else β_k = β_k − 1;
   end if;
end for;
output (β_0, β_1, ··· , β_d);
```



Remark: At the i^{th} step of the algorithm, the vector $(\beta_0, \cdots, \beta_d)$ stores the Betti numbers of K^i.

Proof

- If σ^i is contained in a $(k+1)$-cycle in K^i, this cycle is not a boundary in K^i.

- If σ^i is contained in a $(k+1)$-cycle c in K^i, then c cannot be homologous to a cycle in K^{i-1}

$$\Rightarrow \beta_{k+1}(K^i) \geq \beta_{k+1}(K^{i-1}) + 1$$

- If σ^i is not contained in a $(k+1)$-cycle c in K^i, then $\partial\sigma^i$ is not a boundary in K^{i-1}

$$\Rightarrow \beta_k(K^i) \leq \beta_k(K^{i-1}) - 1$$

- the previous inequalities are equalities.

Getting more information

Definition: A $(k+1)$-simplex σ^i is positive if it is contained in a $(k+1)$-cycle in K^i. It is negative otherwise.

Create a new $(k+1)$-cycle in K^i

Destroy a k-cycle in K^i

$$\beta_k(K) = \sharp(\text{positive simplices}) \; - \; \sharp(\text{negative simplices})$$

- How to keep track of the evolution of the homology all along the filtration?
- What are the created/destroyed cycles?
- What is the lifetime of a cycle?
- How to compute rank$(H_k(K^i) \rightarrow H_k(K^j))$?

Getting more information

Definition: A $(k+1)$-simplex σ^i is positive if it is contained in a $(k+1)$-cycle in K^i. It is negative otherwise.

Create a new $(k+1)$-cycle in K^i

Destroy a k-cycle in K^i

$$\beta_k(K) = \sharp(\text{positive simplices}) \; - \; \sharp(\text{negative simplices})$$

- How to keep track of the evolution of the homology all along the filtration?
- What are the created/destroyed cycles?
- What is the lifetime of a cycle?
- How to compute rank$(H_k(K^i) \to H_k(K^j))$?

This is where topological persistence comes into play!

Multiscale inference

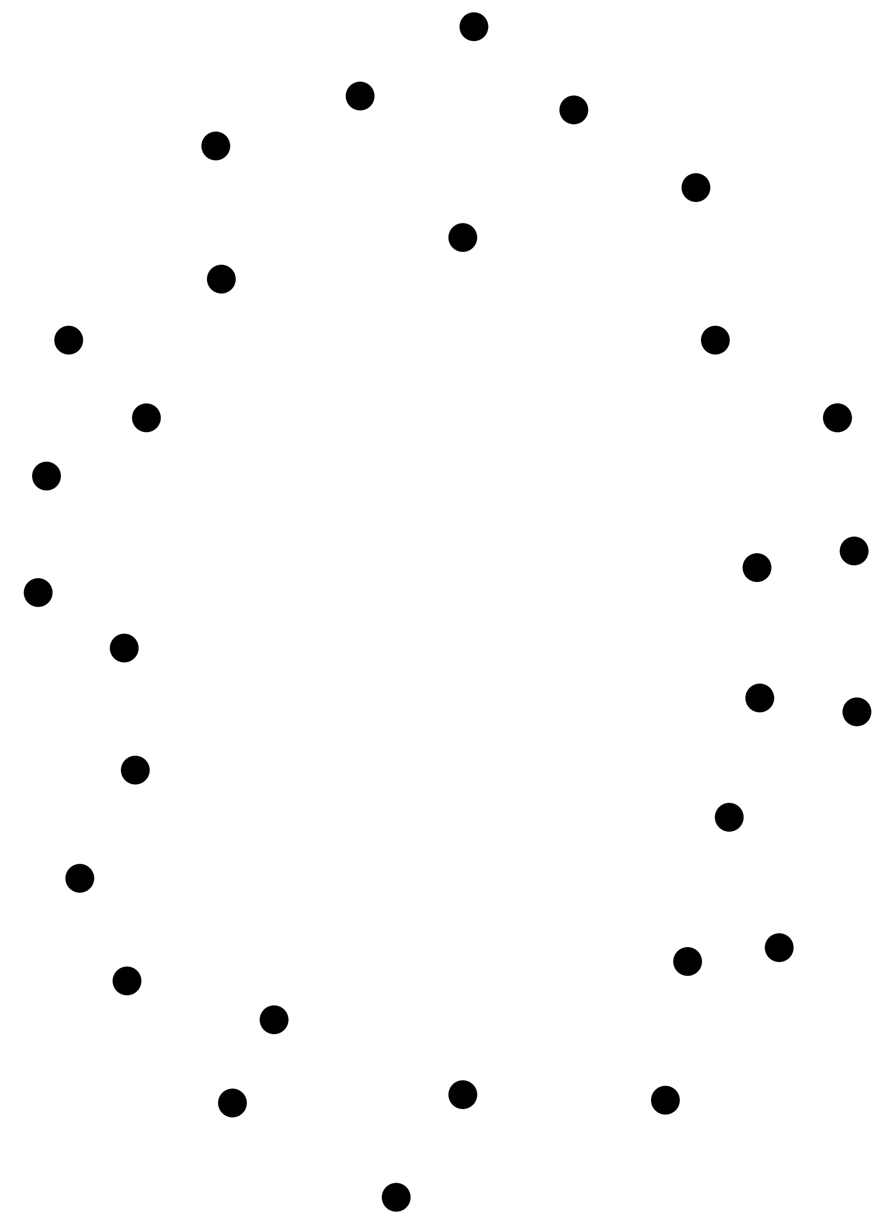
Input: A point cloud W and its pairewise distances $\{d(w,w')\}_{w,w'\in W}$.
$\rightarrow$ Maintain a nested pair $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$ where $L = L(\varepsilon)$.

Init.: $L = \emptyset$; $\varepsilon = +\infty$

WHILE $L \subset W$
insert $p = argmax_{w\in W} d(w, L)$ in L
update $\varepsilon = \max_{w\in W} d(w, L)$
update $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$
Persistence($\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$)
END_WHILE

Rank of the map induced at homology level

Output: Sequence of persistent Betti numbers of $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$



Multiscale inference

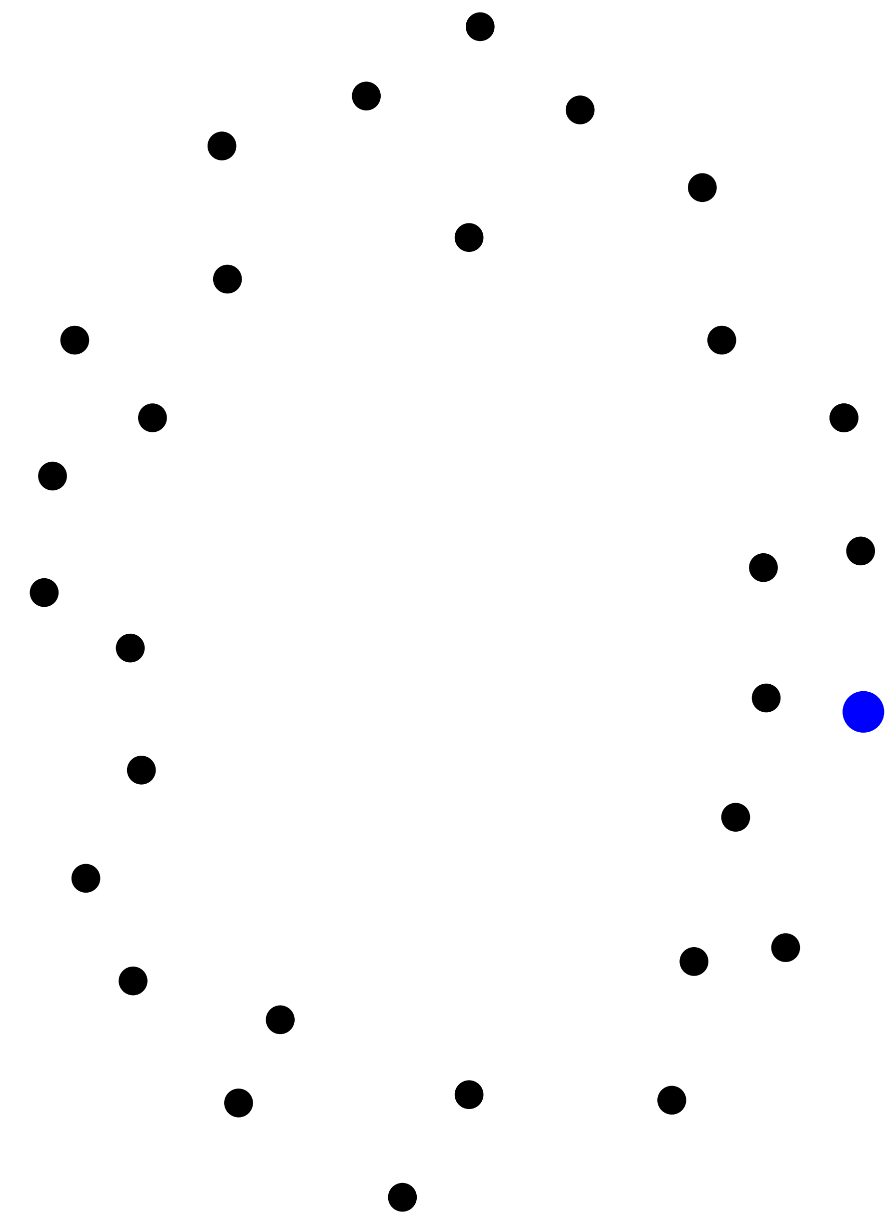
Input: A point cloud W and its pairewise distances $\{d(w,w')\}_{w,w'\in W}$.
$\to$ Maintain a nested pair $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$ where $L = L(\varepsilon)$.

Init.: $L = \emptyset$; $\varepsilon = +\infty$

WHILE $L \subset W$
insert $p = argmax_{w\in W} d(w,L)$ in L
update $\varepsilon = \max_{w\in W} d(w,L)$
update $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$
Persistence($\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$)
END_WHILE

Rank of the map induced at homology level

Output: Sequence of persistent Betti numbers of $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$



Multiscale inference

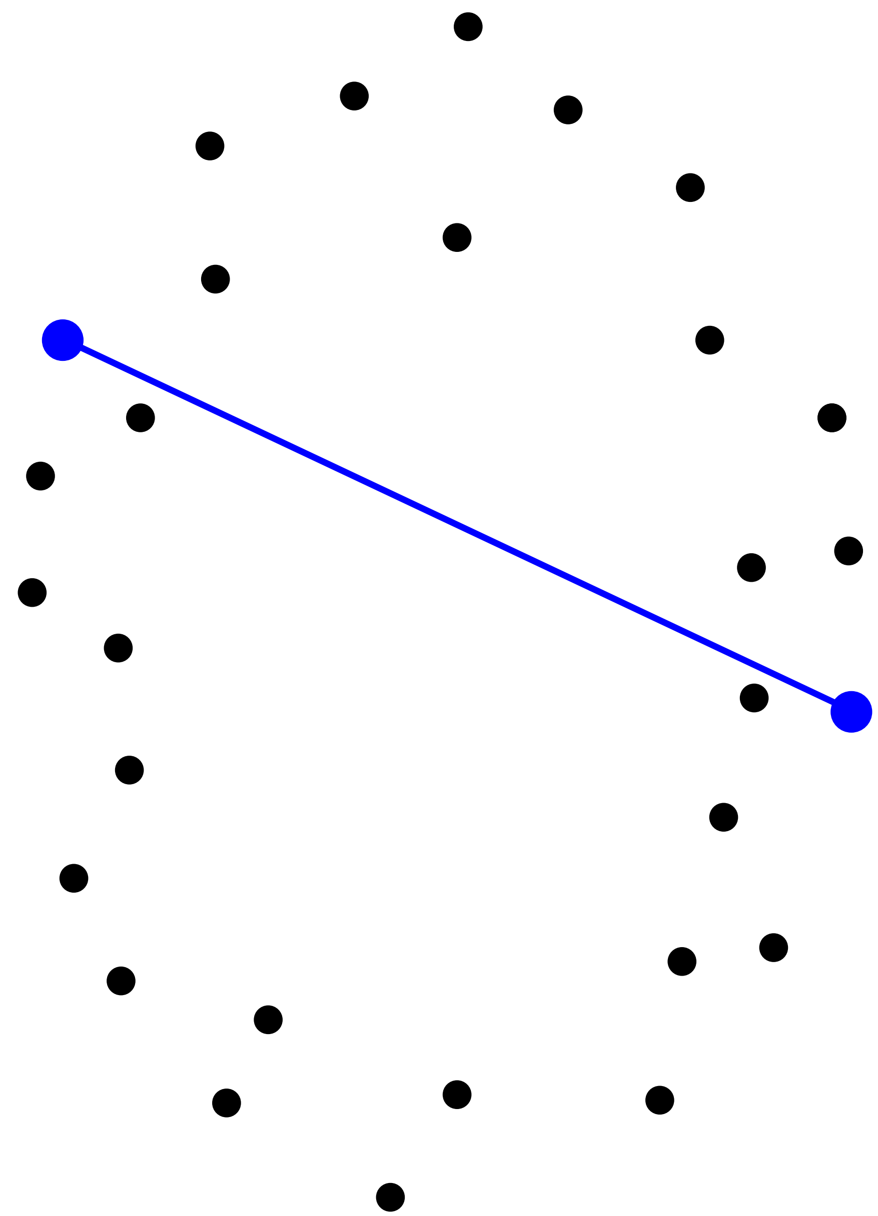
Input: A point cloud W and its pairewise distances $\{d(w,w')\}_{w,w' \in W}$.
$\rightarrow$ Maintain a nested pair $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$ where $L = L(\varepsilon)$.

Init.: $L = \emptyset$; $\varepsilon = +\infty$

WHILE $L \subset W$
insert $p = argmax_{w \in W} d(w, L)$ in L
update $\varepsilon = \max_{w \in W} d(w, L)$
update $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$
Persistence($\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$)
END_WHILE

Rank of the map induced at homology level

Output: Sequence of persistent Betti numbers of $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$



Multiscale inference

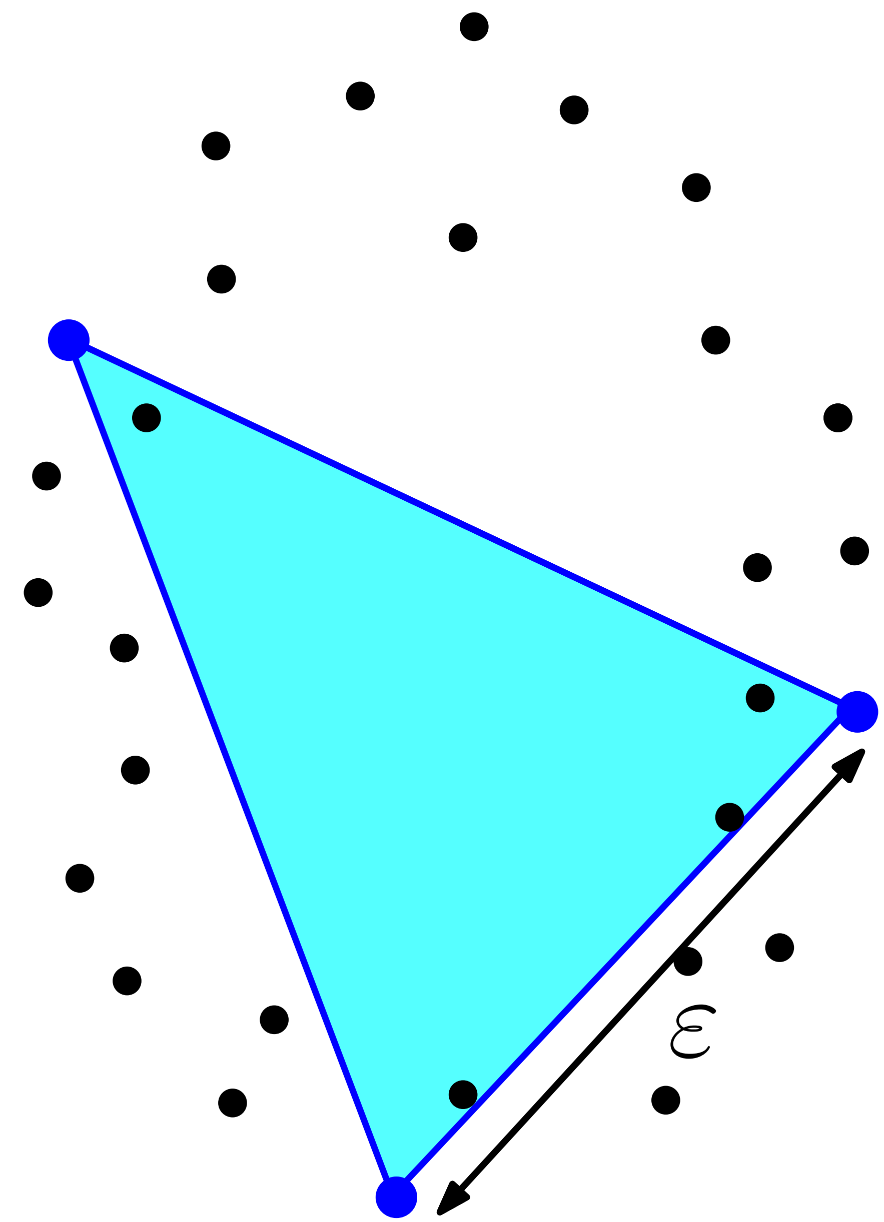
Input: A point cloud W and its pairewise distances $\{d(w,w')\}_{w,w'\in W}$.
$\rightarrow$ Maintain a nested pair $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$ where $L = L(\varepsilon)$.

Init.: $L = \emptyset$; $\varepsilon = +\infty$

WHILE $L \subset W$
insert $p = argmax_{w\in W} d(w,L)$ in L
update $\varepsilon = \max_{w\in W} d(w,L)$
update $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$
Persistence($\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$)
END_WHILE

Rank of the map induced at homology level

Output: Sequence of persistent Betti numbers
of $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$



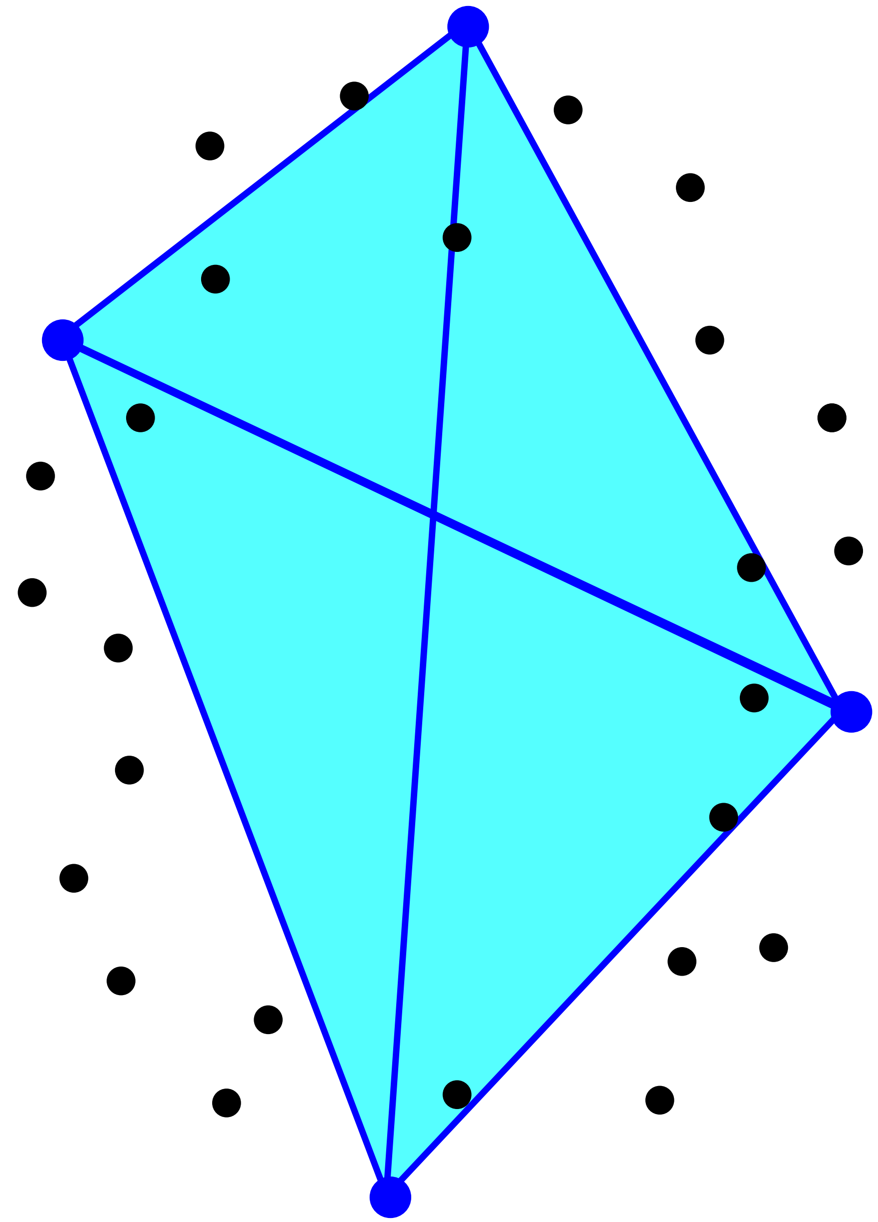
Multiscale inference

Input: A point cloud W and its pairewise distances $\{d(w,w')\}_{w,w' \in W}$.
$\rightarrow$ Maintain a nested pair $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$ where $L = L(\varepsilon)$.

Init.: $L = \emptyset$; $\varepsilon = +\infty$

WHILE $L \subset W$
insert $p = argmax_{w \in W} d(w, L)$ in L
update $\varepsilon = \max_{w \in W} d(w, L)$
update $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$
Persistence($\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$)
Rank of the map induced at homology level
END_WHILE

Output: Sequence of persistent Betti numbers of $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$



Multiscale inference

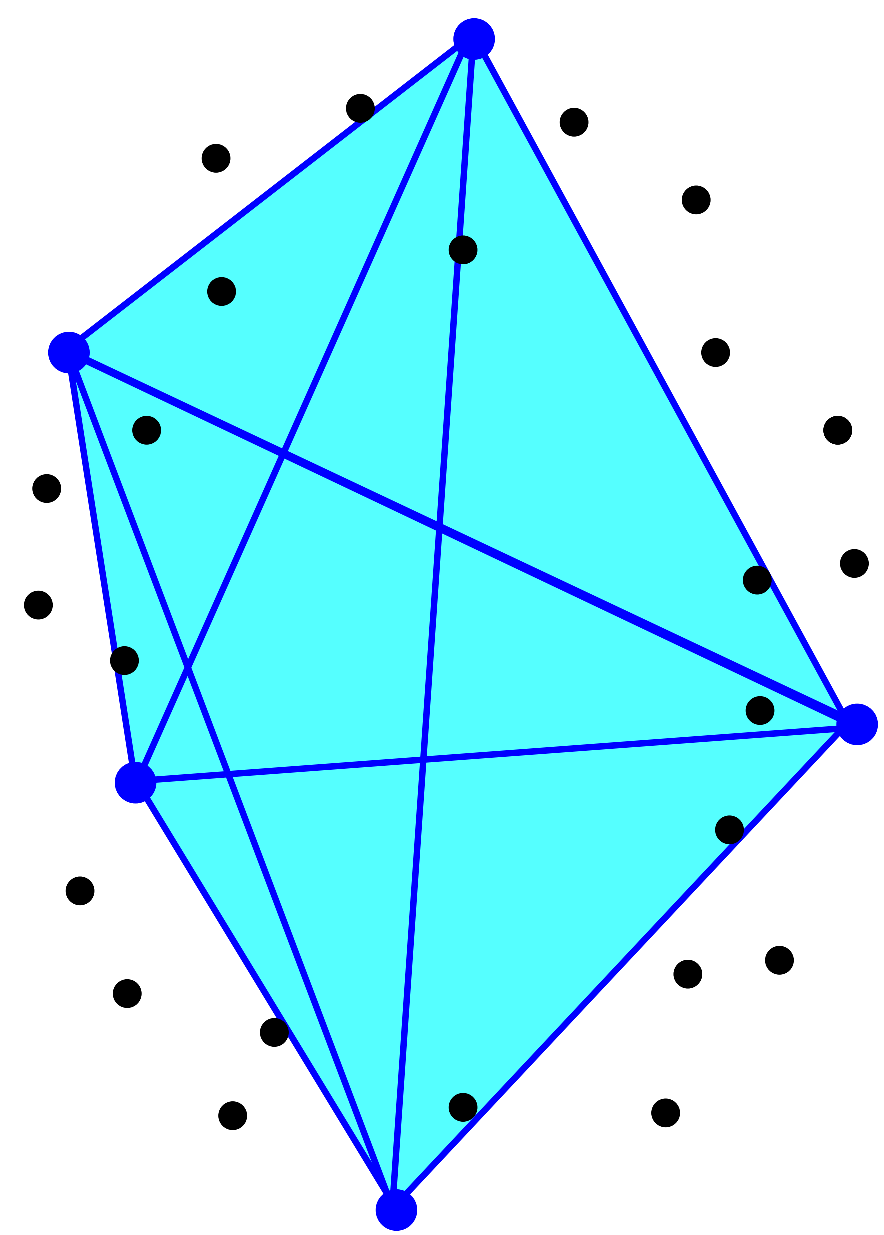
Input: A point cloud W and its pairewise distances $\{d(w,w')\}_{w,w'\in W}$.
$\rightarrow$ Maintain a nested pair $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$ where $L = L(\varepsilon)$.

Init.: $L = \emptyset$; $\varepsilon = +\infty$

WHILE $L \subset W$
insert $p = argmax_{w\in W} d(w, L)$ in L
update $\varepsilon = \max_{w\in W} d(w, L)$
update $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$
Persistence($\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$)
END_WHILE

Rank of the map induced at homology level

Output: Sequence of persistent Betti numbers of $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$



Multiscale inference

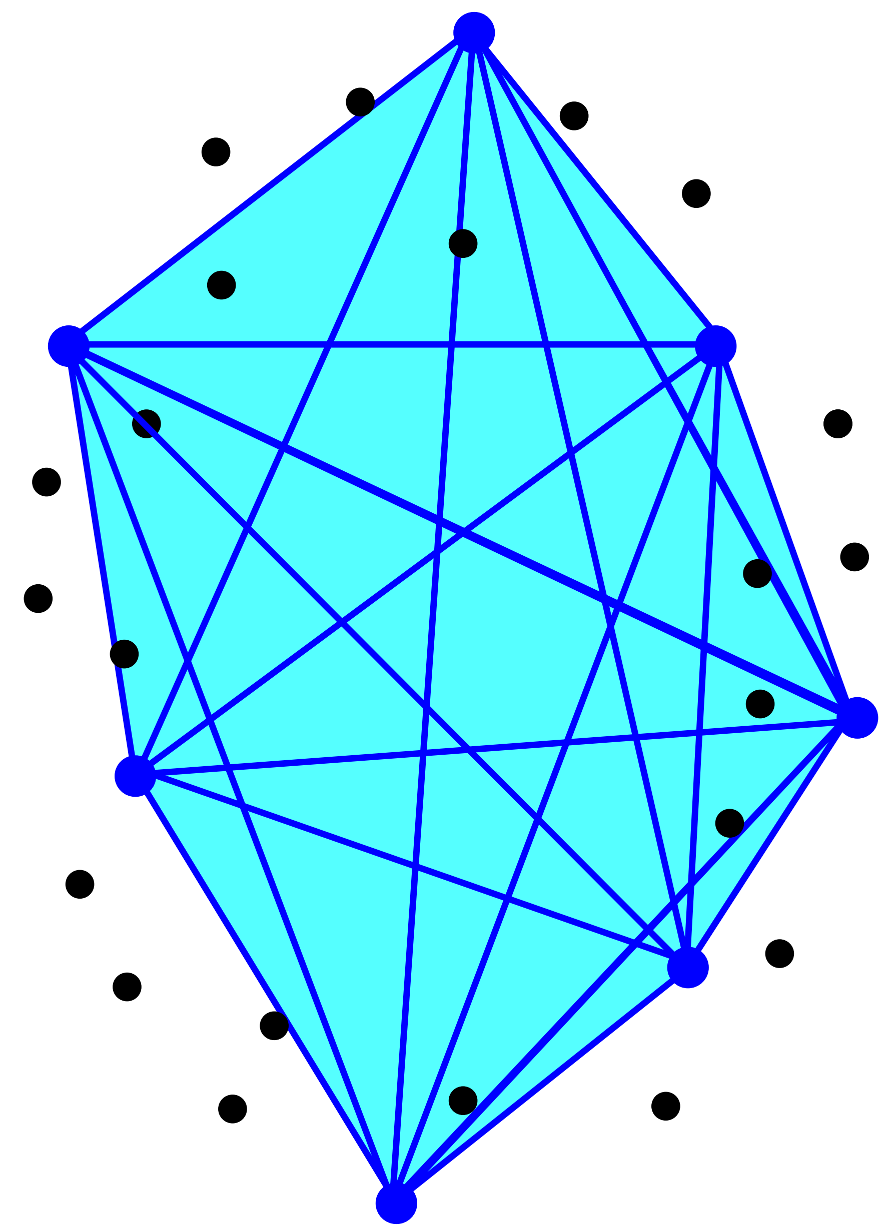
Input: A point cloud W and its pairewise distances $\{d(w,w')\}_{w,w'\in W}$.
$\rightarrow$ Maintain a nested pair $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$ where $L = L(\varepsilon)$.

Init.: $L = \emptyset$; $\varepsilon = +\infty$

WHILE $L \subset W$
insert $p = argmax_{w\in W} d(w,L)$ in L
update $\varepsilon = \max_{w\in W} d(w,L)$
update $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$
Persistence($\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$)
END_WHILE

Rank of the map induced at homology level

Output: Sequence of persistent Betti numbers of $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$



Multiscale inference

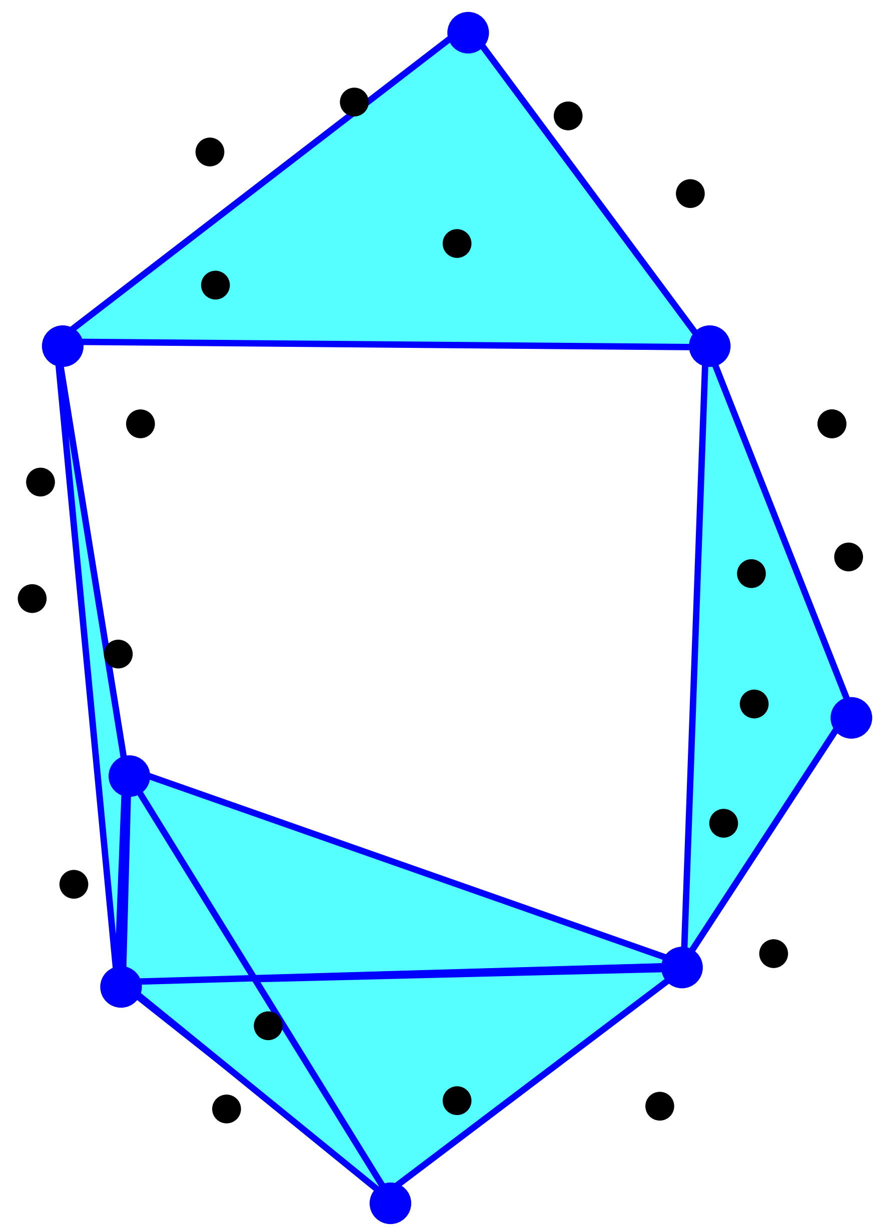
Input: A point cloud W and its pairewise distances $\{d(w,w')\}_{w,w'\in W}$.
$\rightarrow$ Maintain a nested pair $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$ where $L = L(\varepsilon)$.

Init.: $L = \emptyset$; $\varepsilon = +\infty$

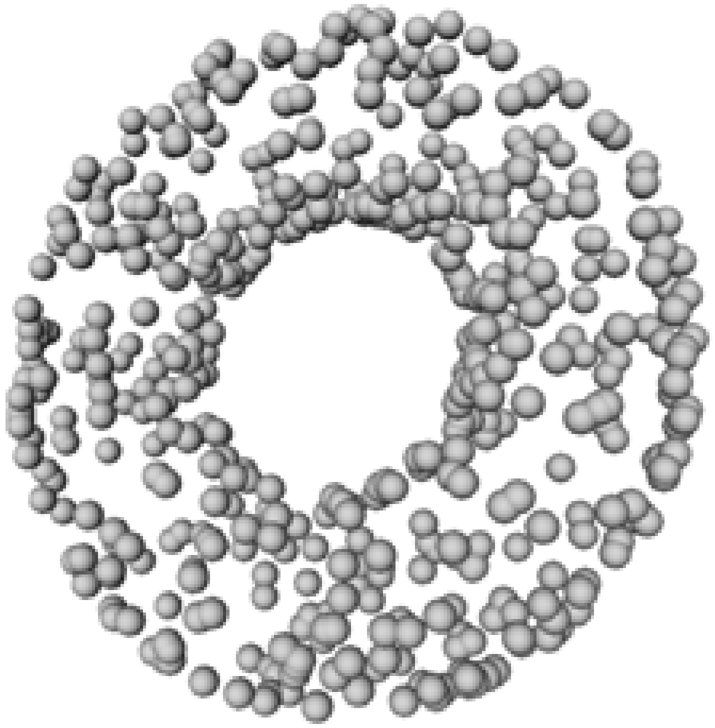
WHILE $L \subset W$
insert $p = argmax_{w\in W} d(w,L)$ in L
update $\varepsilon = \max_{w\in W} d(w,L)$
update $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$
Persistence($\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$)
END_WHILE

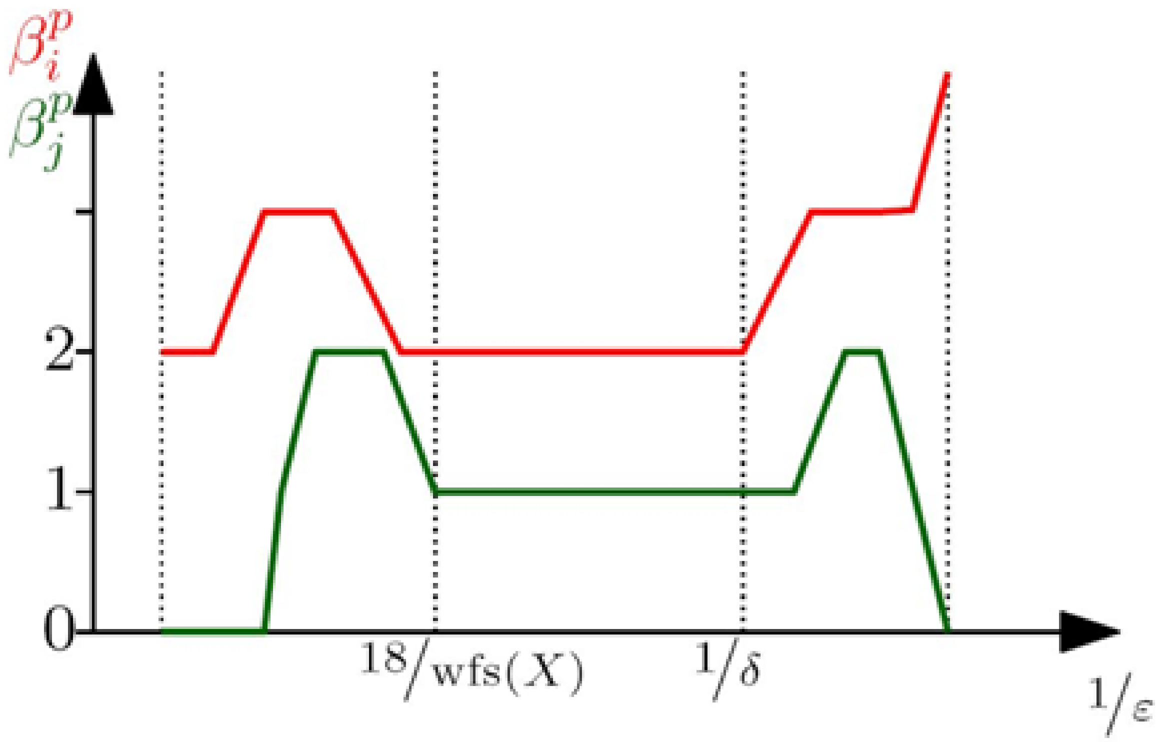
Rank of the map induced at homology level

Output: Sequence of persistent Betti numbers of $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$



Multiscale inference



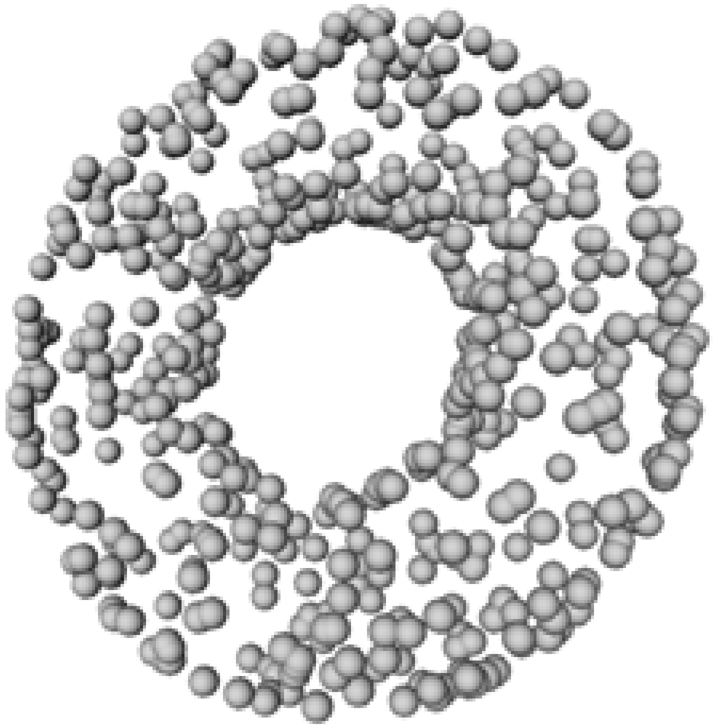


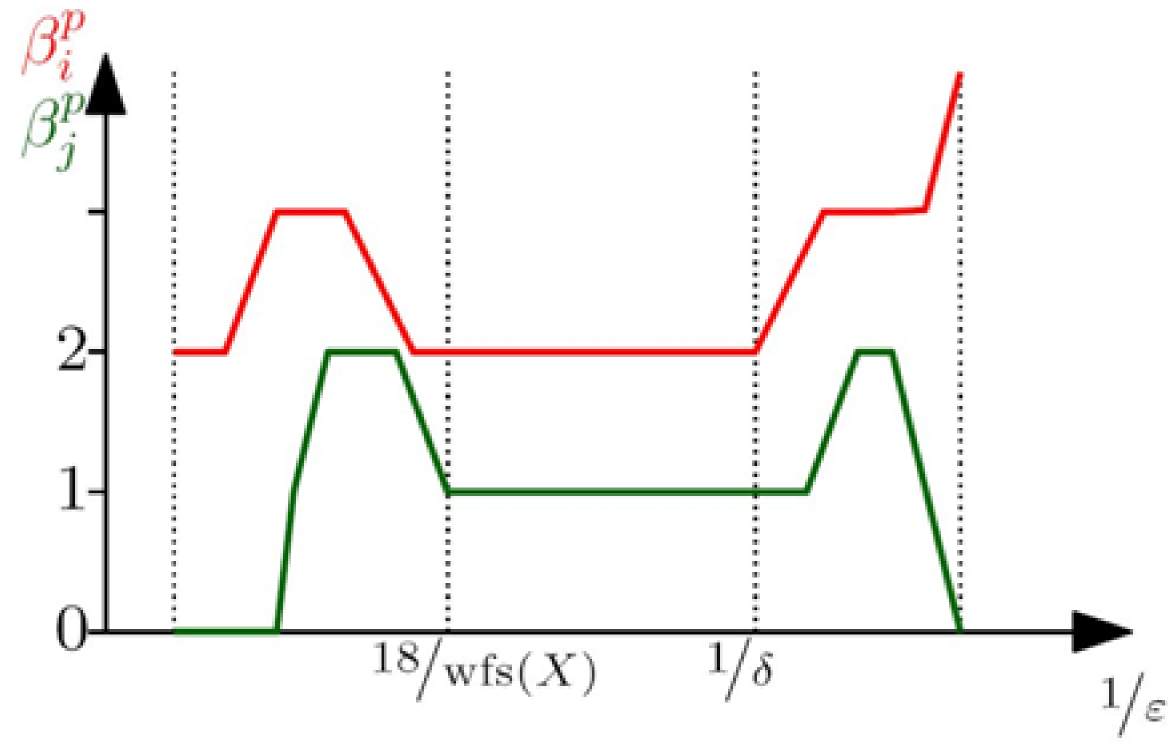
Theorem: [C-Oudot'08]
If $d_H(W, X) < \delta$ for $\delta < \frac{1}{18}\mathrm{wfs}(X)$, then at every iteration of the algorithm such that $\delta < \varepsilon < \frac{1}{18}\mathrm{wfs}(X)$,

$$\beta_k(X^\lambda) = \dim H_k(X^\lambda) = rk(H_k(\mathcal{R}^{4\varepsilon}(L)) \to H_k(\mathcal{R}^{4\varepsilon}(L)))$$

for any $\lambda \in (0, \mathrm{wfs}(X))$ and any $k \in \mathbb{N}$.

Multiscale inference





Complexity of the algorithm:

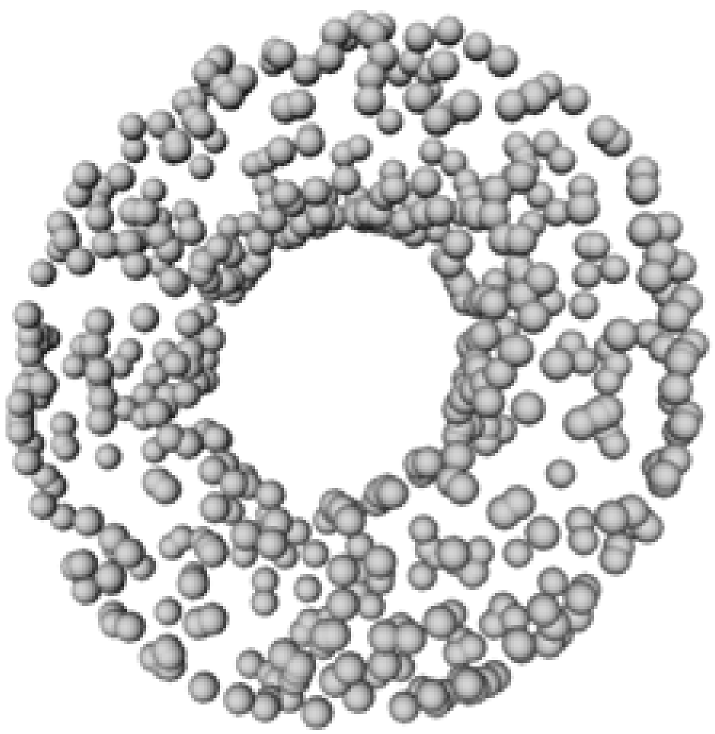
- If $X \subset \mathbb{R}^d$ is non smooth the running time of the algorithm is

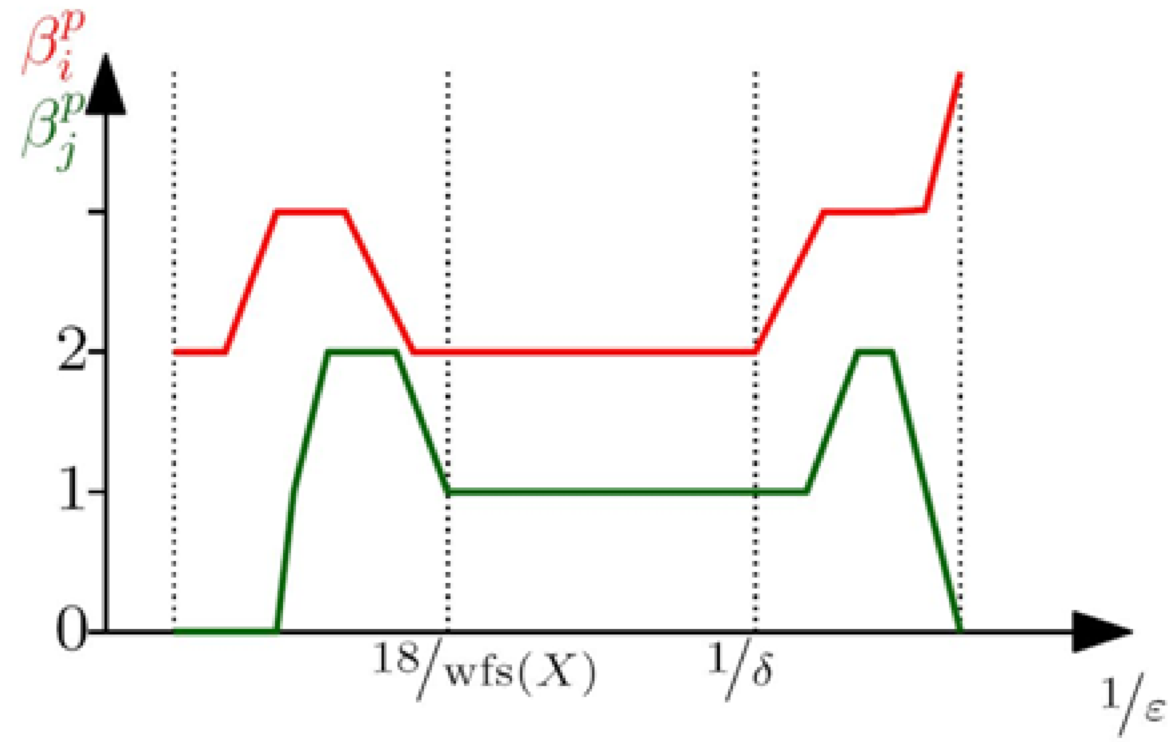
$$O(8^{33^d}|W|^5)$$

- If X is a smooth submanifold of $\mathbb{R}^d$ dimension m the running time is

$$O(8^{35^m}|W|)$$

Multiscale inference





Complexity of the algorithm:

- If $X \subset \mathbb{R}^d$ is non smooth the running time of the algorithm is

$$O(8^{33^d}|W|^5)$$

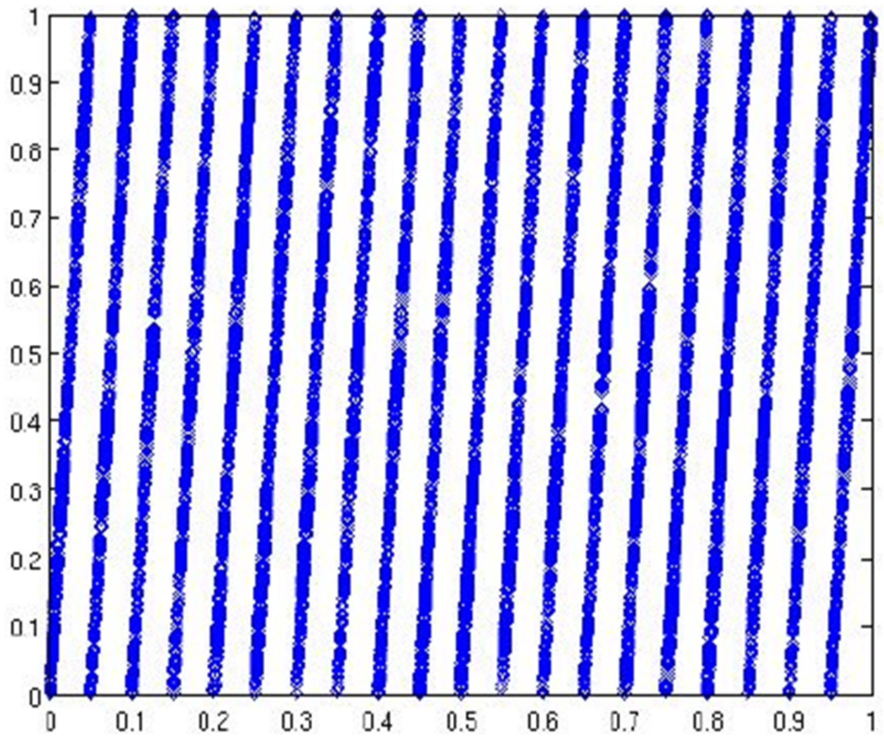
- If X is a smooth submanifold of $\mathbb{R}^d$ dimension m the running time is

$$O(8^{35^m}|W|)$$

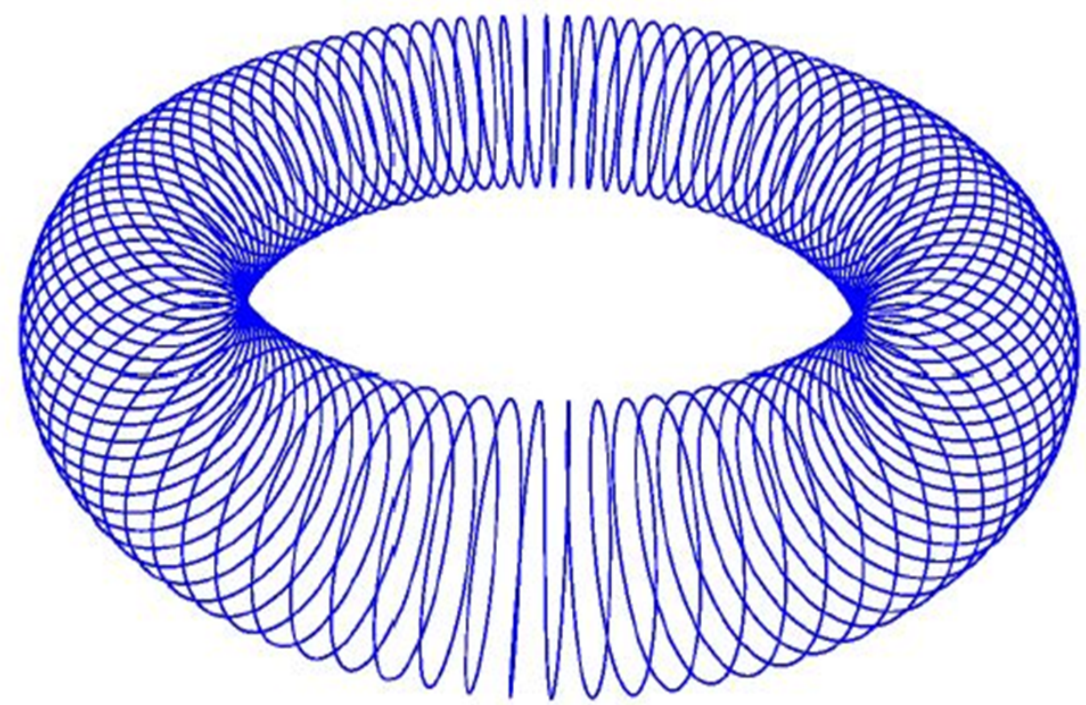
Depend on the intrinsic dimension of X

A synthetic example

$[0,1] \times [0,1]$



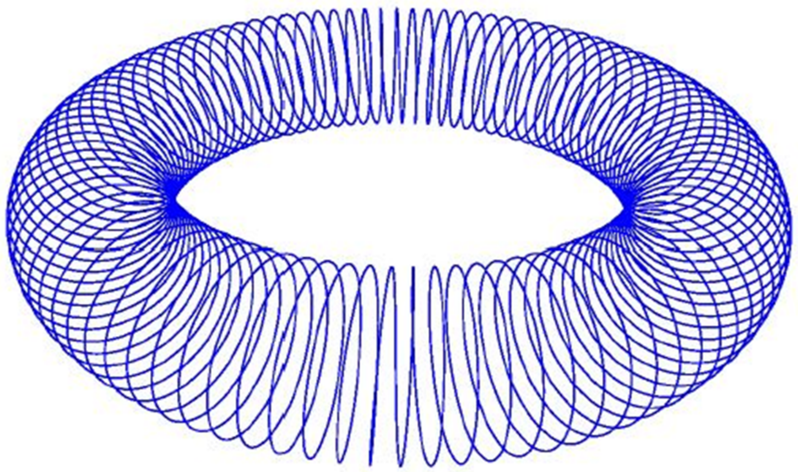
$\mathbb{R}^{1000}$

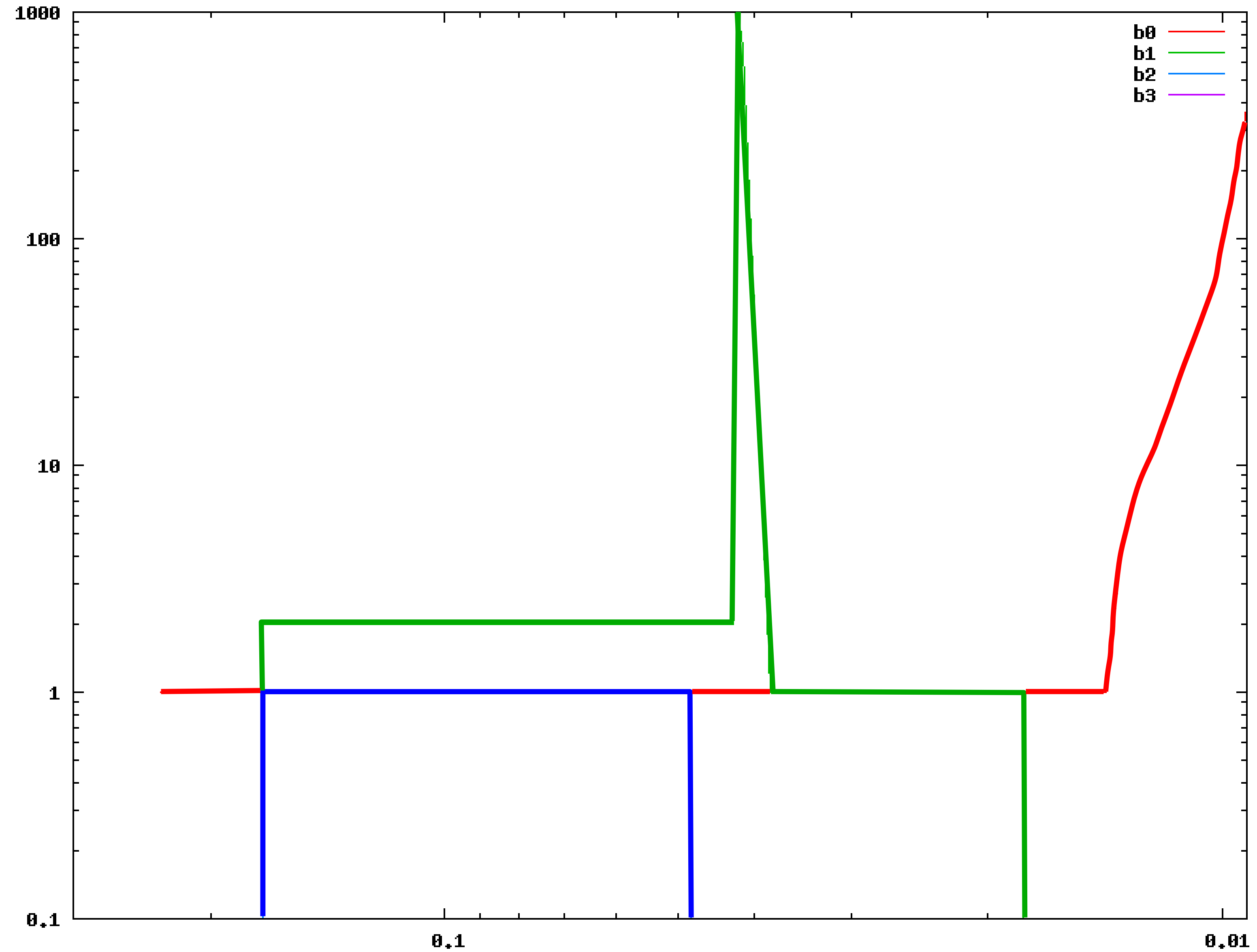


Non-linear embedding of $\mathbb{S}^1 \times \mathbb{S}^1$ in $\mathbb{R}^{1000}$

$50,000$ points sampled uniformly at random from a curve drawn on the 2-torus $\mathbb{S}^1 \times \mathbb{S}^1$.

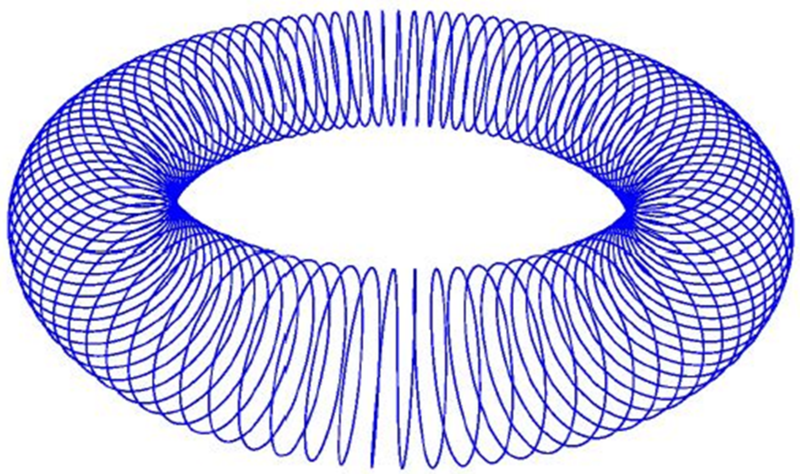
A synthetic example

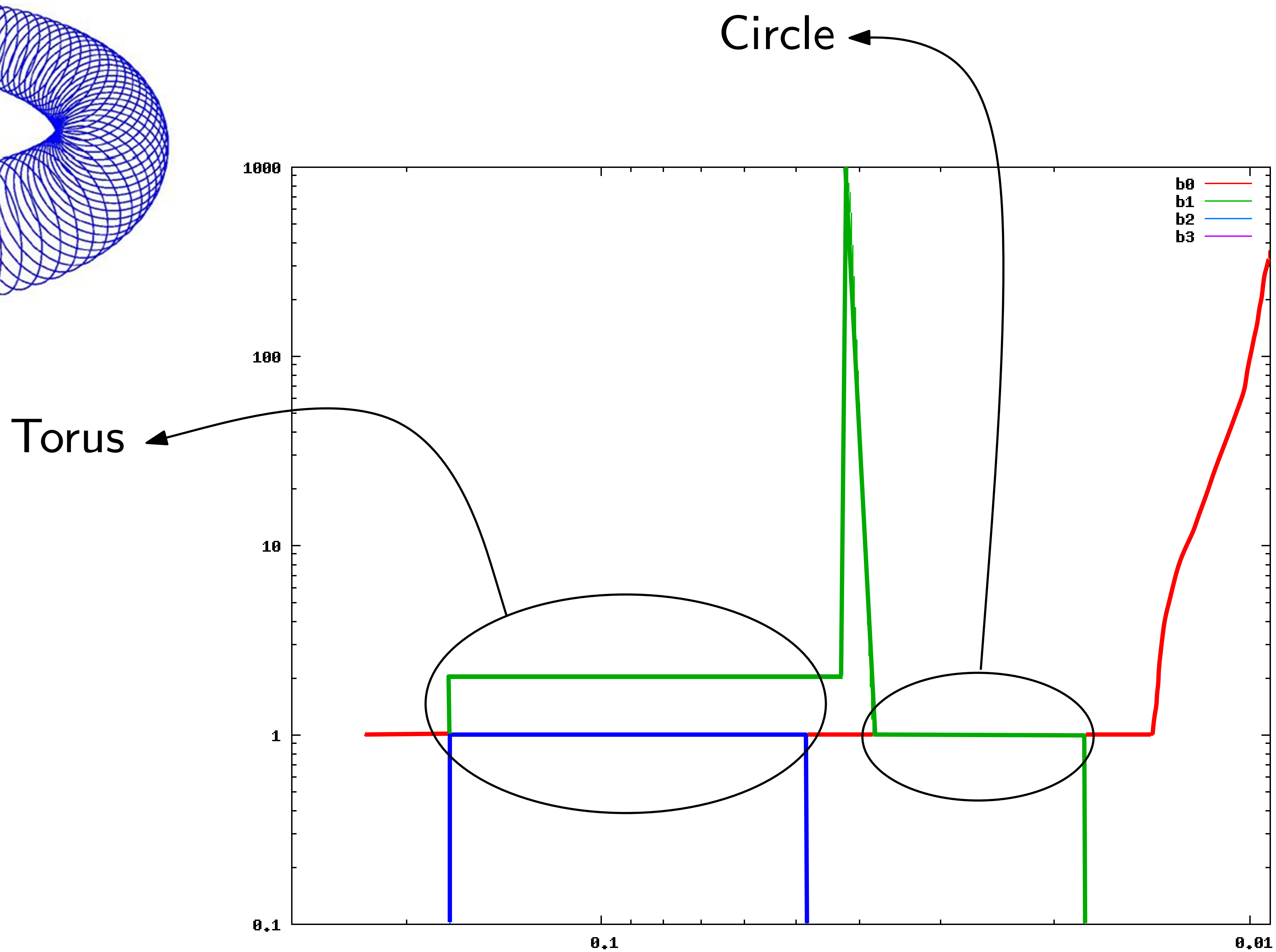




Output: sequence of Betti numbers on a log-log scale

A synthetic example





Output: sequence of Betti numbers on a log-log scale