EMAp Summer Course

Topological Data Analysis with Persistent Homology

https://raphaeltinarrage.github.io/EMAp.html

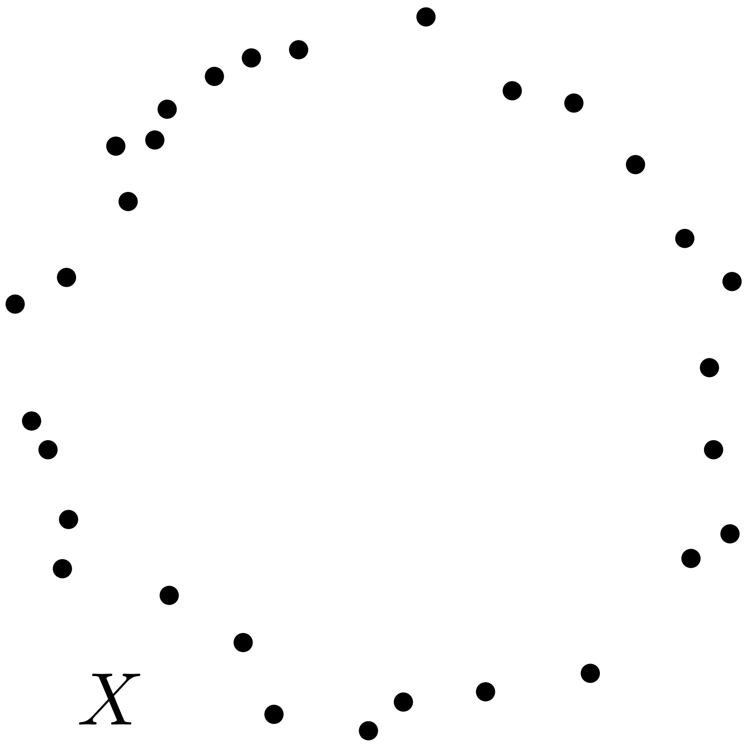
Lesson 7: Topological inference

Last update: February 3, 2021

Introduction

In real life, we are often given datasets that are subsets of the Euclidean space: $X \subset \mathbb{R}^n$.

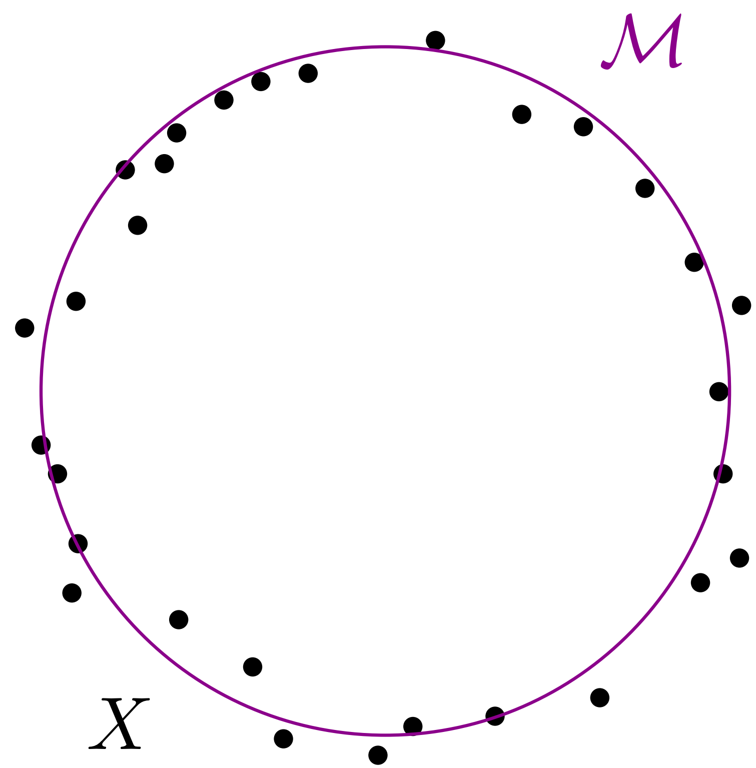
Of course, X is finite.



Introduction

In real life, we are often given datasets that are subsets of the Euclidean space: $X \subset \mathbb{R}^n$.

Of course, X is finite.



In Topological Data Analysis, we think of X as being a sample of an underlying continuous object, $\mathcal{M} \subset \mathbb{R}^n$.

Understanding the topology of $\mathcal{M}$ would give us interesting insights about our dataset.

I - Thickenings

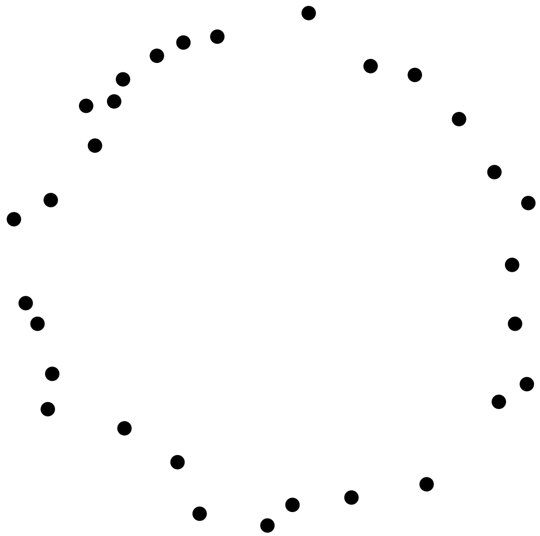
II - Čech complex

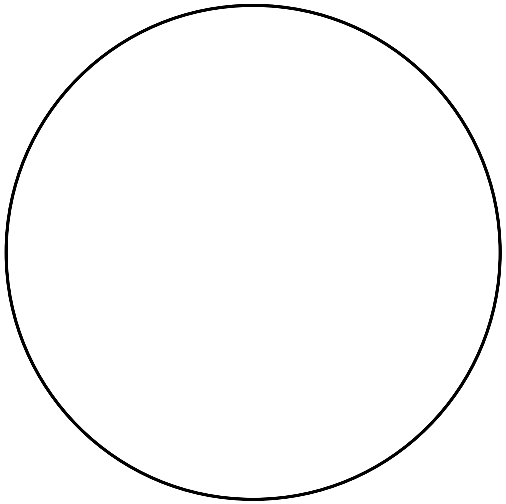
III - Rips complex

The Topological Inference problem

Let $\mathcal{M} \subset \mathbb{R}^n$ be a bounded subset.
Suppose that we are given a finite sample $X \subset \mathcal{M}$.
Estimate the homology groups of $\mathcal{M}$ from X.

X

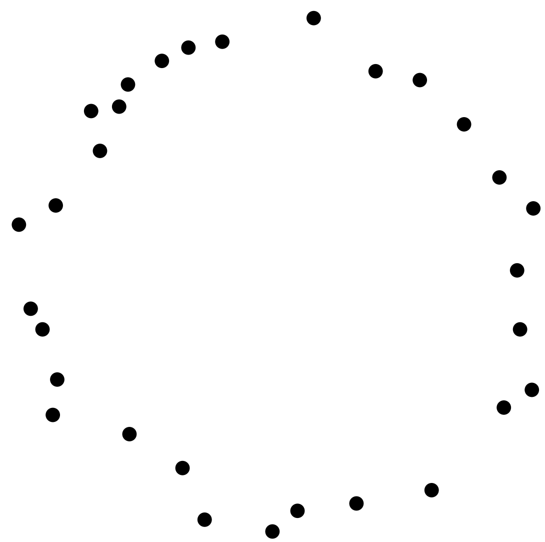




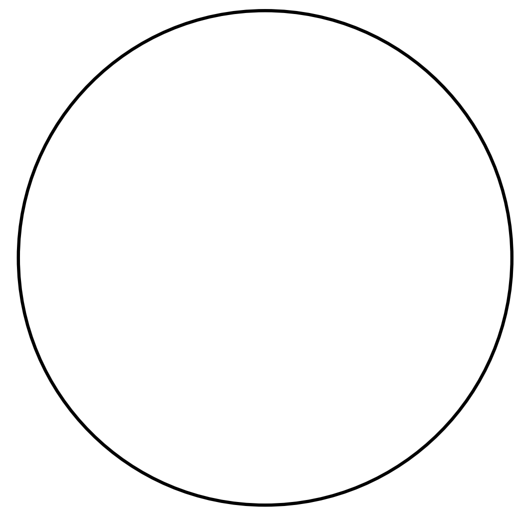
$\mathcal{M}$

The Topological Inference problem

Let $\mathcal{M} \subset \mathbb{R}^n$ be a bounded subset.
Suppose that we are given a finite sample $X \subset \mathcal{M}$.
Estimate the homology groups of $\mathcal{M}$ from X.



X



$\mathcal{M}$

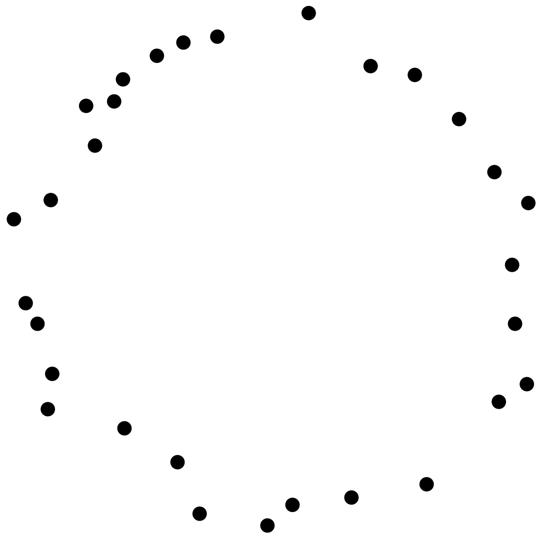
We cannot use X directly. Its homology is disapointing:

$$\beta_0(X) = 30 \quad \text{and} \quad \beta_i(X) \text{ for } i \geq 1$$

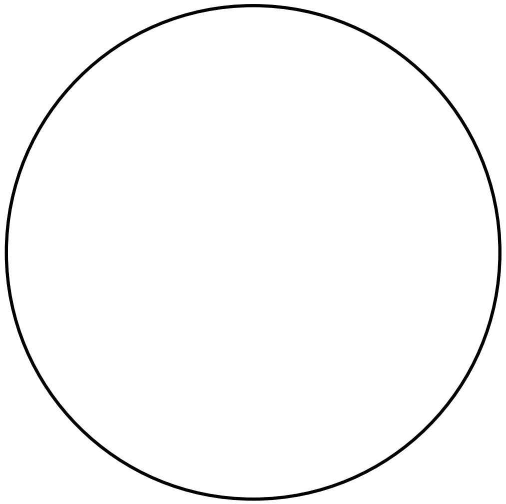
number of connected components
$=$ number of points of X

The Topological Inference problem

Let $\mathcal{M} \subset \mathbb{R}^n$ be a bounded subset.
Suppose that we are given a finite sample $X \subset \mathcal{M}$.
Estimate the homology groups of $\mathcal{M}$ from X.



X



$\mathcal{M}$

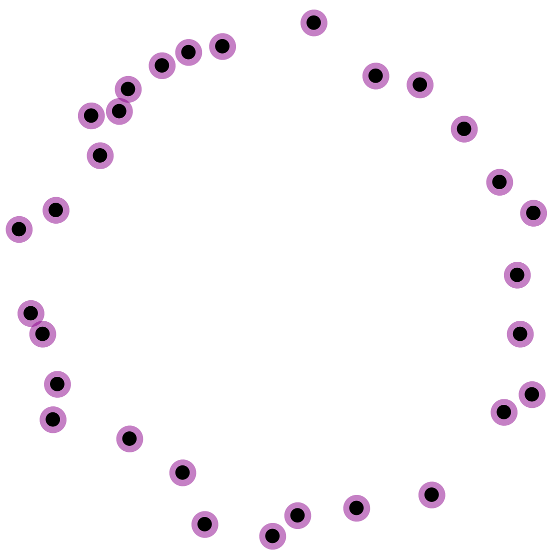
We cannot use X directly.

Idea: Thicken X.

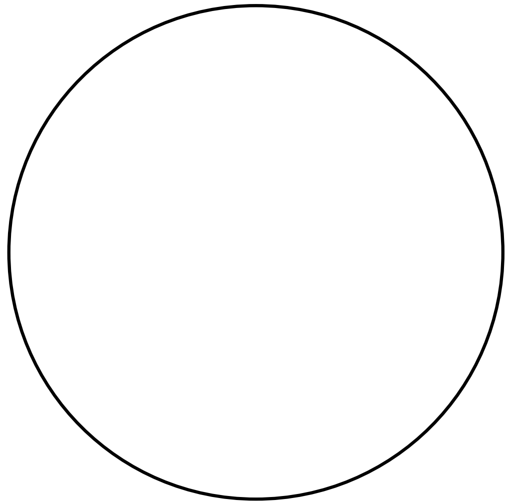
Definition: For every $t \geq 0$, the *t-thickening* of the set X, denoted X^t, is the set of points of the ambient space with distance at most t from X:

$$X^t = \{y \in \mathbb{R}^n, \exists x \in X, \|x - y\| \leq t\}.$$

Let $\mathcal{M} \subset \mathbb{R}^n$ be a bounded subset.
Suppose that we are given a finite sample $X \subset \mathcal{M}$.
Estimate the homology groups of $\mathcal{M}$ from X.



$X^{0.05}$



$\mathcal{M}$

We cannot use X directly.

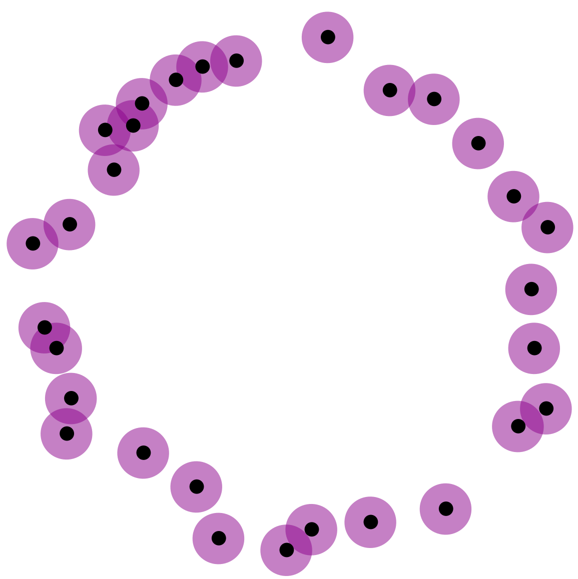
Idea: Thicken X.

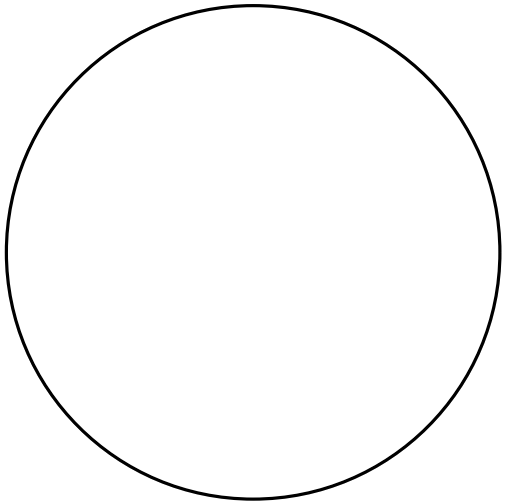
Definition: For every $t \geq 0$, the *t-thickening* of the set X, denoted X^t, is the set of points of the ambient space with distance at most t from X:

$$X^t = \{y \in \mathbb{R}^n, \exists x \in X, \|x - y\| \leq t\}.$$

The Topological Inference problem

Let $\mathcal{M} \subset \mathbb{R}^n$ be a bounded subset.
Suppose that we are given a finite sample $X \subset \mathcal{M}$.
Estimate the homology groups of $\mathcal{M}$ from X.





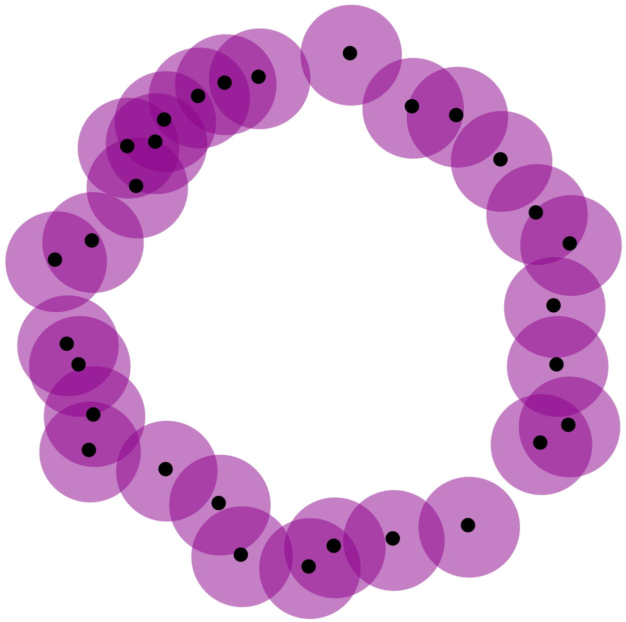
We cannot use X directly.

Idea: Thicken X.

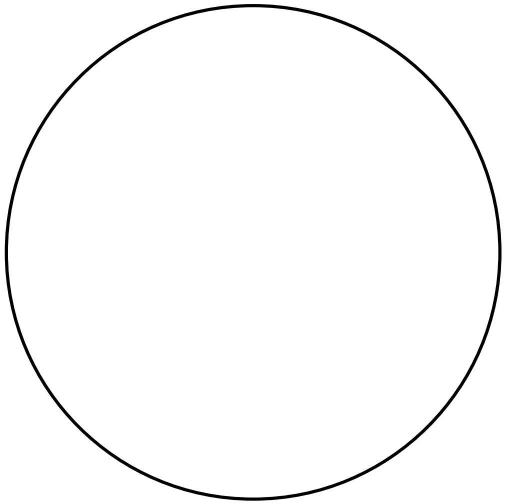
Definition: For every $t \geq 0$, the *t-thickening* of the set X, denoted X^t, is the set of points of the ambient space with distance at most t from X:

$$X^t = \{y \in \mathbb{R}^n, \exists x \in X, \|x - y\| \leq t\}.$$

Let $\mathcal{M} \subset \mathbb{R}^n$ be a bounded subset.
Suppose that we are given a finite sample $X \subset \mathcal{M}$.
Estimate the homology groups of $\mathcal{M}$ from X.



$X^{0.2}$



$\mathcal{M}$

We cannot use X directly.

Idea: Thicken X.

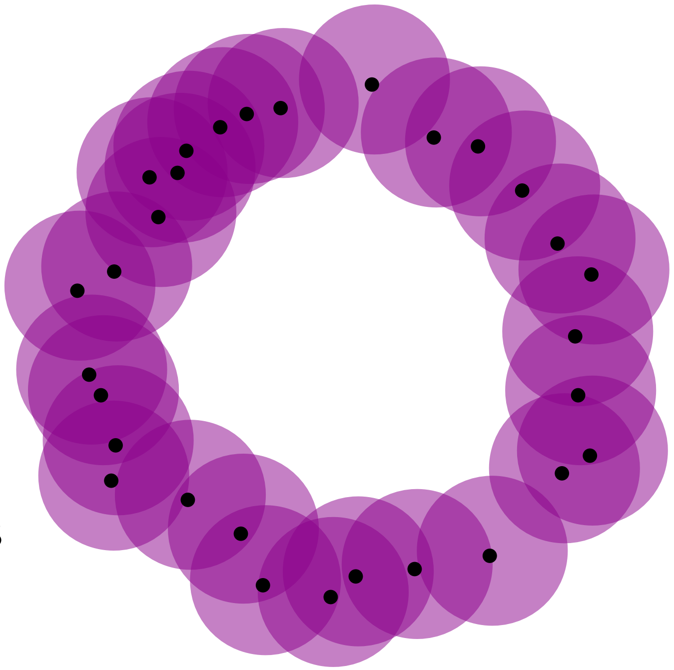
Definition: For every $t \geq 0$, the *t-thickening* of the set X, denoted X^t, is the set of points of the ambient space with distance at most t from X:

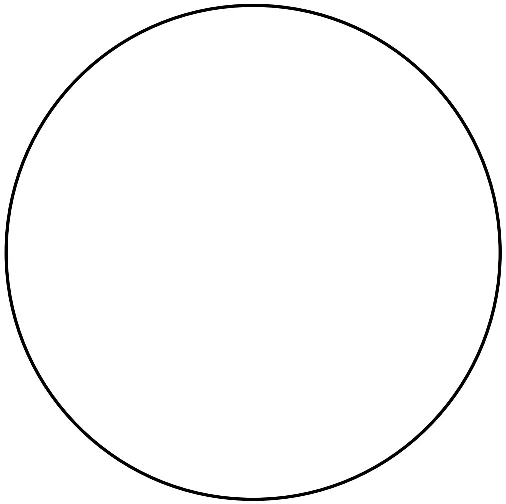
$$X^t = \{y \in \mathbb{R}^n, \exists x \in X, \|x - y\| \leq t\}.$$

The Topological Inference problem

Let $\mathcal{M} \subset \mathbb{R}^n$ be a bounded subset.
Suppose that we are given a finite sample $X \subset \mathcal{M}$.
Estimate the homology groups of $\mathcal{M}$ from X.

$X^{0.3}$





$\mathcal{M}$

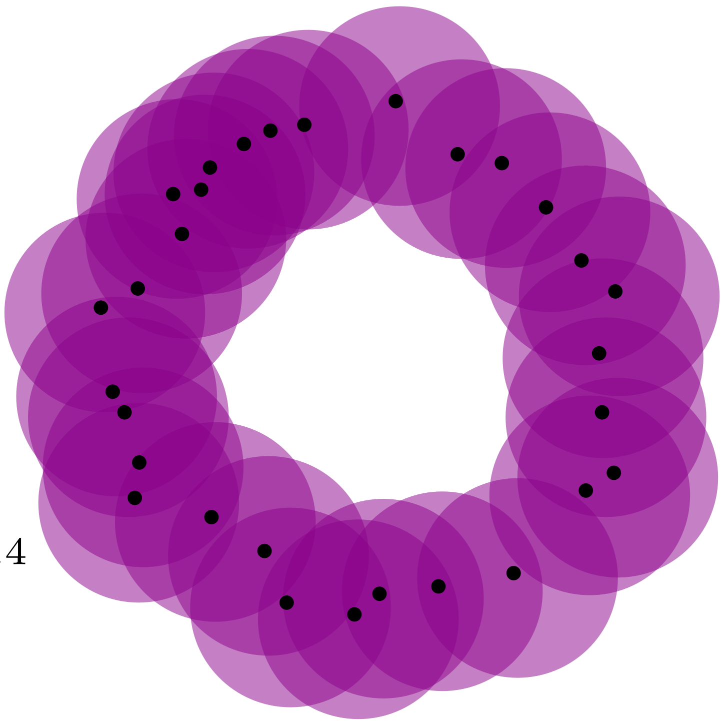
We cannot use X directly.

Idea: Thicken X.

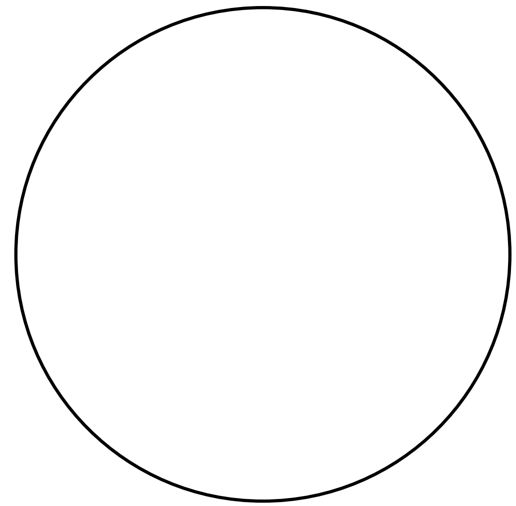
Definition: For every $t \geq 0$, the *t-thickening* of the set X, denoted X^t, is the set of points of the ambient space with distance at most t from X:

$$X^t = \{y \in \mathbb{R}^n, \exists x \in X, \|x - y\| \leq t\}.$$

Let $\mathcal{M} \subset \mathbb{R}^n$ be a bounded subset.
Suppose that we are given a finite sample $X \subset \mathcal{M}$.
Estimate the homology groups of $\mathcal{M}$ from X.



$X^{0.4}$



$\mathcal{M}$

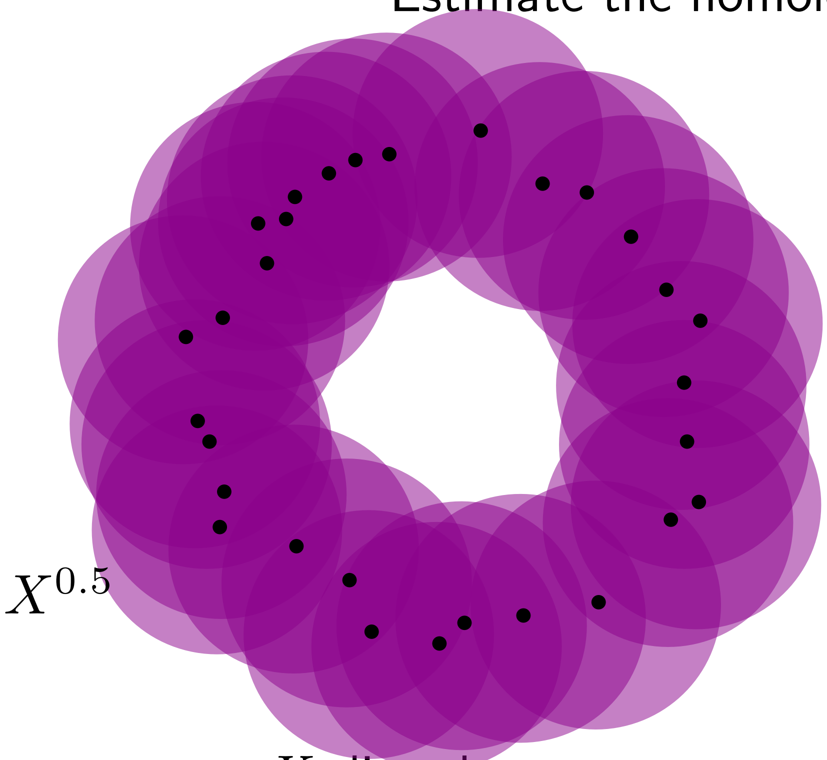
We cannot use X directly.

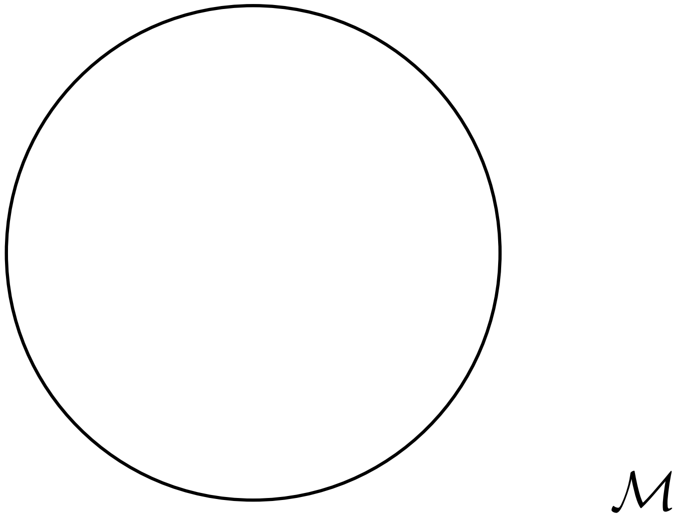
Idea: Thicken X.

Definition: For every $t \geq 0$, the *t-thickening* of the set X, denoted X^t, is the set of points of the ambient space with distance at most t from X:

$$X^t = \{y \in \mathbb{R}^n, \exists x \in X, \|x - y\| \leq t\}.$$

Let $\mathcal{M} \subset \mathbb{R}^n$ be a bounded subset.
Suppose that we are given a finite sample $X \subset \mathcal{M}$.
Estimate the homology groups of $\mathcal{M}$ from X.





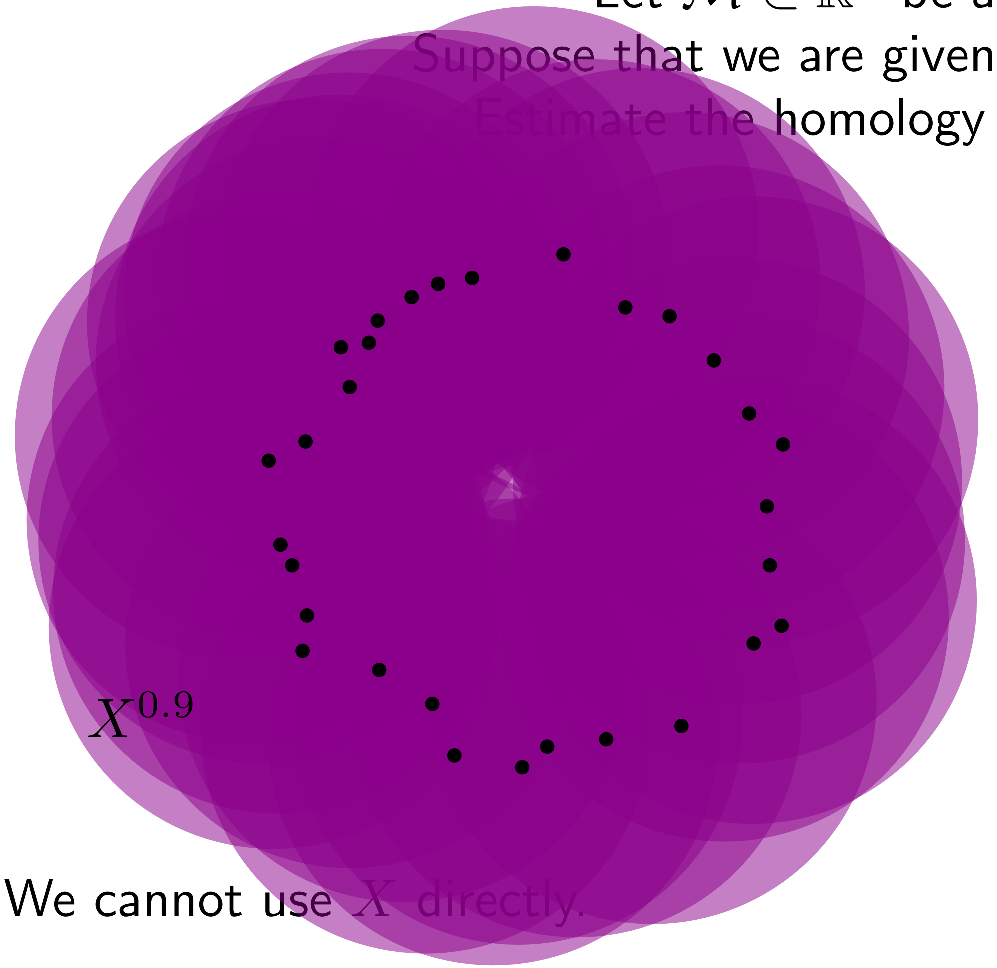
We cannot use X directly.

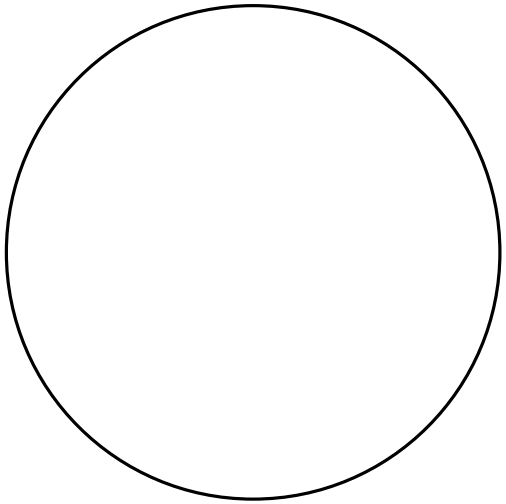
Idea: Thicken X.

Definition: For every $t \geq 0$, the *t-thickening* of the set X, denoted X^t, is the set of points of the ambient space with distance at most t from X:

$$X^t = \{y \in \mathbb{R}^n, \exists x \in X, \|x - y\| \leq t\}.$$

Let $\mathcal{M} \subset \mathbb{R}^n$ be a bounded subset.
Suppose that we are given a finite sample $X \subset \mathcal{M}$.
Estimate the homology groups of $\mathcal{M}$ from X.





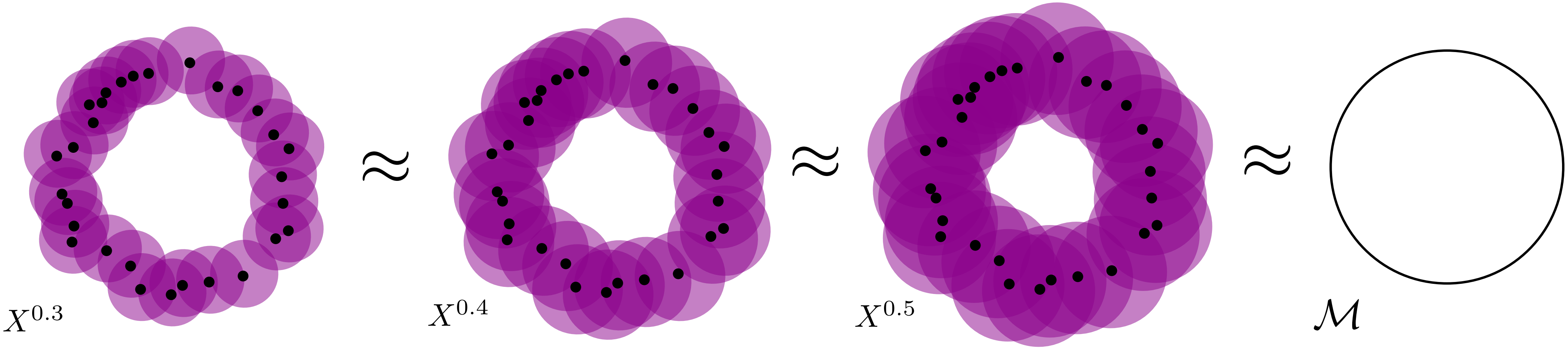
We cannot use X directly.

Idea: Thicken X.

Definition: For every $t \geq 0$, the *t-thickening* of the set X, denoted X^t, is the set of points of the ambient space with distance at most t from X:

$$X^t = \{y \in \mathbb{R}^n, \exists x \in X, \|x - y\| \leq t\}.$$

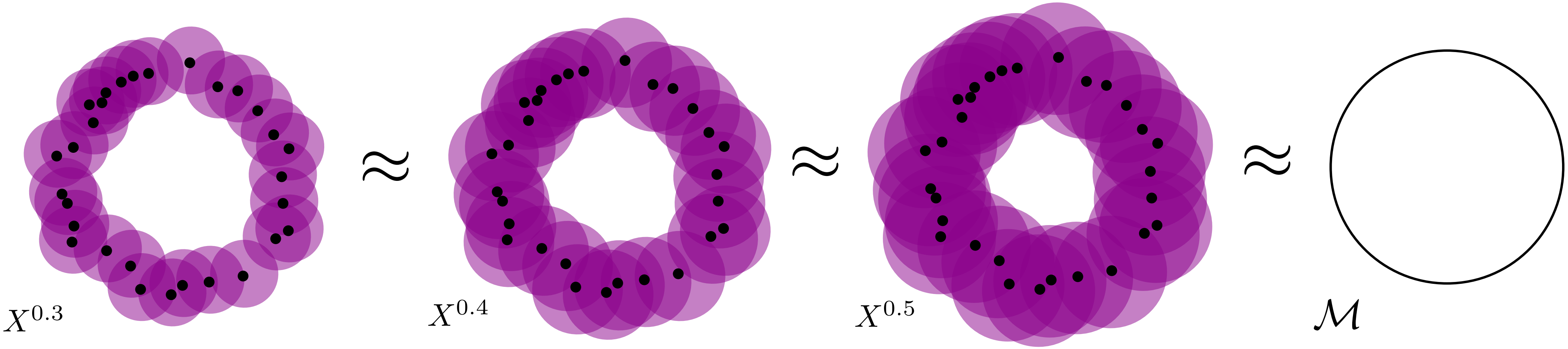
Some thickenings are homotopy equivalent to $\mathcal{M}$.



Hence we can recover the homology of $\mathcal{M}$:

$$\beta_0(\mathcal{M}) = \beta_0(X^{0.3})$$
$$\beta_1(\mathcal{M}) = \beta_1(X^{0.3})$$
$$\beta_2(\mathcal{M}) = \beta_2(X^{0.3})$$
$$\ldots$$

Some thickenings are homotopy equivalent to $\mathcal{M}$.



Hence we can recover the homology of $\mathcal{M}$:

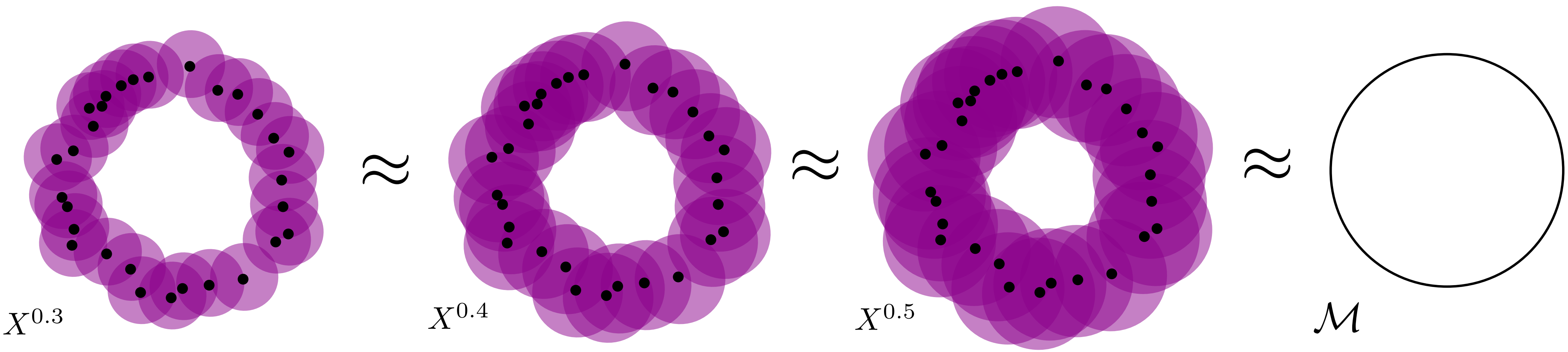
$$\beta_0(\mathcal{M}) = \beta_0(X^{0.3})$$
$$\beta_1(\mathcal{M}) = \beta_1(X^{0.3})$$
$$\beta_2(\mathcal{M}) = \beta_2(X^{0.3})$$
$$\dots$$

Question 1: How to select a t such that $X^t \approx \mathcal{M}$?

Question 2: How to compute the homology groups of X^t?

The Topological Inference problem

Some thickenings are homotopy equivalent to $\mathcal{M}$.



Hence we can recover the homology of $\mathcal{M}$:

$$\beta_0(\mathcal{M}) = \beta_0(X^{0.3})$$
$$\beta_1(\mathcal{M}) = \beta_1(X^{0.3})$$
$$\beta_2(\mathcal{M}) = \beta_2(X^{0.3})$$
$$\dots$$

Question 1: How to select a t such that $X^t \approx \mathcal{M}$?

Hausdorff distance

Reach

Question 2: How to compute the homology groups of X^t?

Hausdorff distance

Let X be any subset of $\mathbb{R}^n$. The function *distance to* X is the map

$$\begin{aligned} \operatorname{dist}(\cdot, X) : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ x &\longmapsto \operatorname{dist}(x, X) = \inf\{\|y - x\|, x \in X\} \end{aligned}$$

A *projection* of $y \in \mathbb{R}^n$ on X is a point $x \in X$ which attains this infimum.

Hausdorff distance

Let X be any subset of $\mathbb{R}^n$. The function *distance to* X is the map

$$\begin{aligned} \operatorname{dist}(\cdot, X) : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ x &\longmapsto \operatorname{dist}(x, X) = \inf\{\|y - x\|, x \in X\} \end{aligned}$$

A *projection* of $y \in \mathbb{R}^n$ on X is a point $x \in X$ which attains this infimum.

Definition: Let $Y \subset \mathbb{R}^n$ be another subset. The *Hausdorff distance* between X and Y is

$$\begin{aligned} \mathrm{d_H}(X, Y) &= \max\left\{\sup_{y \in Y} \operatorname{dist}(y, X),\ \sup_{x \in X} \operatorname{dist}(x, Y)\right\} \\ &= \max\left\{\sup_{y \in Y} \inf_{x \in X} \|x - y\|,\ \sup_{x \in X} \inf_{y \in Y} \|x - y\|\right\}. \end{aligned}$$

Hausdorff distance

Let X be any subset of $\mathbb{R}^n$. The function *distance to* X is the map

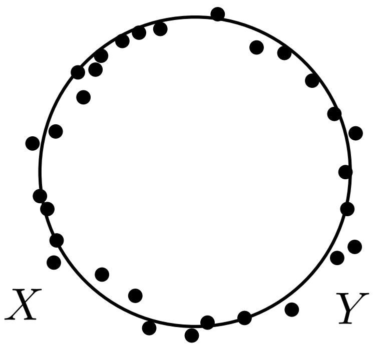
$$\begin{aligned} \operatorname{dist}(\cdot, X) : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ x &\longmapsto \operatorname{dist}(x, X) = \inf\{\|y - x\|, x \in X\} \end{aligned}$$

A *projection* of $y \in \mathbb{R}^n$ on X is a point $x \in X$ which attains this infimum.

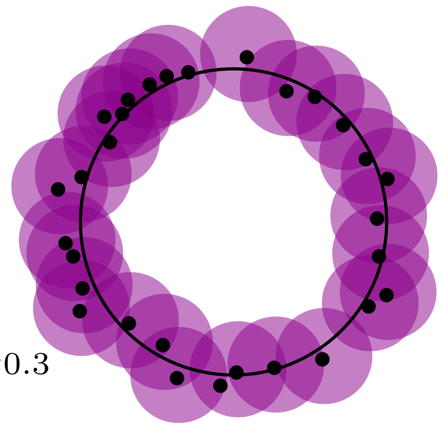
Definition: Let $Y \subset \mathbb{R}^n$ be another subset. The *Hausdorff distance* between X and Y is

$$\begin{aligned} d_H(X, Y) &= \max\left\{\sup_{y \in Y} \operatorname{dist}(y, X), \ \sup_{x \in X} \operatorname{dist}(x, Y)\right\} \\ &= \max\left\{\sup_{y \in Y} \inf_{x \in X} \|x - y\|, \ \sup_{x \in X} \inf_{y \in Y} \|x - y\|\right\}. \end{aligned}$$

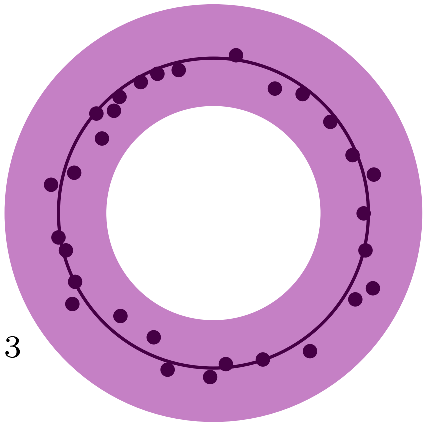
Exercise: Show that the Hausdorff distance is equal to $\inf\{t \geq 0, X \subset Y^t \text{ and } Y \subset X^t\}$.



X Y



$Y \subset X^{0.3}$



$X \subset Y^{0.3}$

Medial axis and reach

The *medial axis* of X is the subset $\operatorname{med}(X) \subset \mathbb{R}^n$ which consists of points $y \in \mathbb{R}^n$ that admit at least two projections on X:

$$\operatorname{med}(X) = \{y \in \mathbb{R}^n, \exists x, x' \in X, x \neq x', \|y - x\| = \|y - x'\| = \operatorname{dist}(y, X)\}.$$

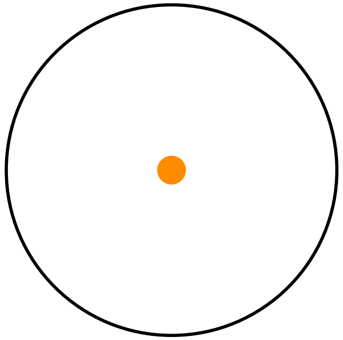
Medial axis and reach

The *medial axis* of X is the subset $\mathrm{med}\,(X) \subset \mathbb{R}^n$ which consists of points $y \in \mathbb{R}^n$ that admit at least two projections on X:

$$\mathrm{med}\,(X) = \{y \in \mathbb{R}^n, \exists x, x' \in X, x \neq x', \|y - x\| = \|y - x'\| = \mathrm{dist}\,(y, X)\}\,.$$

Examples:

The medial axis of the unit circle is the origin



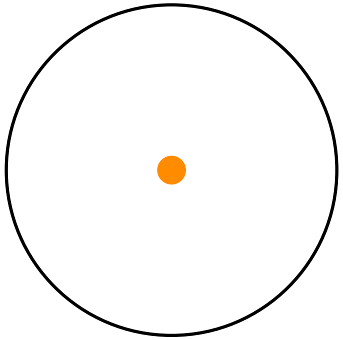
Medial axis and reach

The *medial axis* of X is the subset $\mathrm{med}\,(X) \subset \mathbb{R}^n$ which consists of points $y \in \mathbb{R}^n$ that admit at least two projections on X:

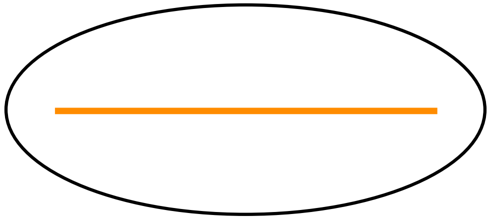
$$\mathrm{med}\,(X) = \{y \in \mathbb{R}^n, \exists x, x' \in X, x \neq x', \|y - x\| = \|y - x'\| = \mathrm{dist}\,(y, X)\}\,.$$

Examples:

The medial axis of the unit circle is the origin



The medial axis of an ellipse is a segment



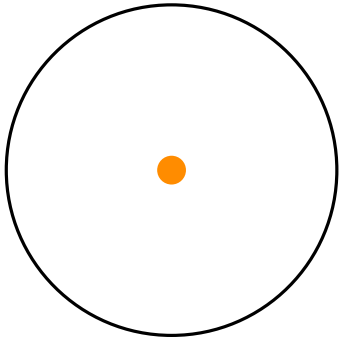
Medial axis and reach

The *medial axis* of X is the subset $\operatorname{med}(X) \subset \mathbb{R}^n$ which consists of points $y \in \mathbb{R}^n$ that admit at least two projections on X:

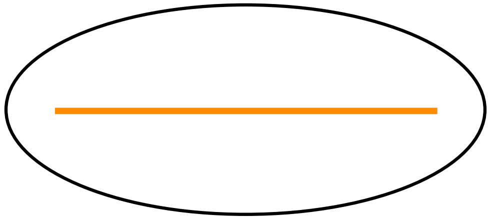
$$\operatorname{med}(X) = \{y \in \mathbb{R}^n, \exists x, x' \in X, x \neq x', \|y - x\| = \|y - x'\| = \operatorname{dist}(y, X)\}.$$

Examples:

The medial axis of the unit circle is the origin



The medial axis of an ellipse is a segment



The medial axis of a point is the empty set

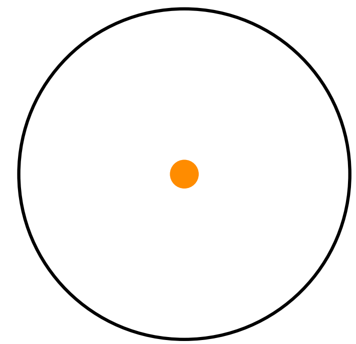
Medial axis and reach

The *medial axis* of X is the subset $\operatorname{med}(X) \subset \mathbb{R}^n$ which consists of points $y \in \mathbb{R}^n$ that admit at least two projections on X:

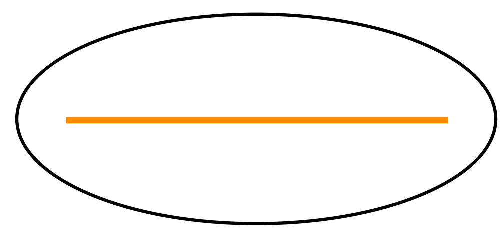
$$\operatorname{med}(X) = \{y \in \mathbb{R}^n, \exists x, x' \in X, x \neq x', \|y - x\| = \|y - x'\| = \operatorname{dist}(y, X)\}.$$

Examples:

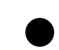
The medial axis of the unit circle is the origin



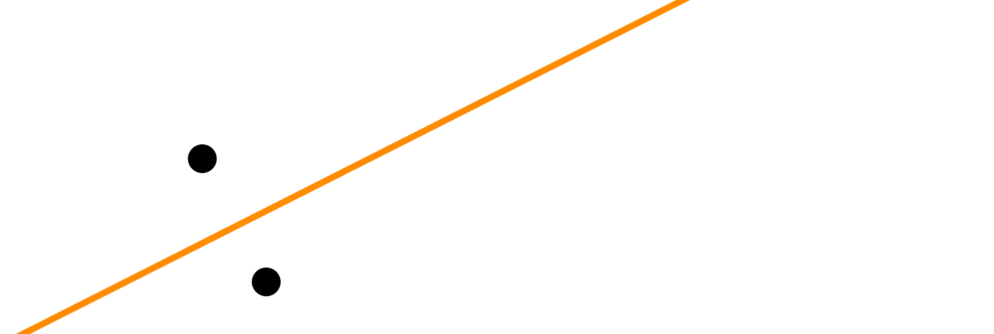
The medial axis of an ellipse is a segment



The medial axis of a point is the empty set



The medial axis of two points is their bisector



Medial axis and reach

The *medial axis* of X is the subset $\operatorname{med}(X) \subset \mathbb{R}^n$ which consists of points $y \in \mathbb{R}^n$ that admit at least two projections on X:

$$\operatorname{med}(X) = \{y \in \mathbb{R}^n, \exists x, x' \in X, x \neq x', \|y - x\| = \|y - x'\| = \operatorname{dist}(y, X)\}.$$

The *reach* of X is

$$\begin{aligned}
\operatorname{reach}(X) &= \inf\{\operatorname{dist}(y, X), y \in \operatorname{med}(X)\} \\
&= \inf\{\|x - y\|, x \in X, y \in \operatorname{med}(X)\}.
\end{aligned}$$

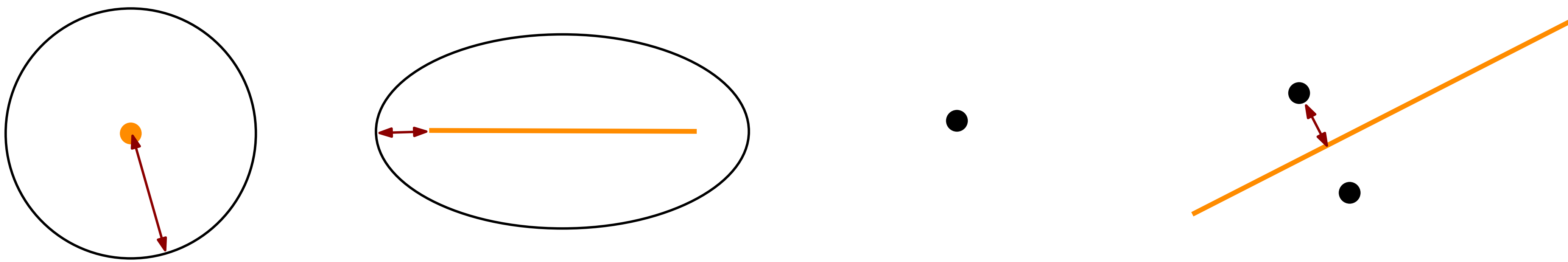
Medial axis and reach

The *medial axis* of X is the subset $\mathrm{med}\,(X) \subset \mathbb{R}^n$ which consists of points $y \in \mathbb{R}^n$ that admit at least two projections on X:

$$\mathrm{med}\,(X) = \{y \in \mathbb{R}^n, \exists x, x' \in X, x \neq x', \|y - x\| = \|y - x'\| = \mathrm{dist}\,(y, X)\}\,.$$

The *reach* of X is

$$\begin{aligned}\mathrm{reach}\,(X) &= \inf\{\mathrm{dist}\,(y, X)\,, y \in \mathrm{med}\,(X)\}\\ &= \inf\{\|x - y\|\,, x \in X, y \in \mathrm{med}\,(X)\}\,.\end{aligned}$$



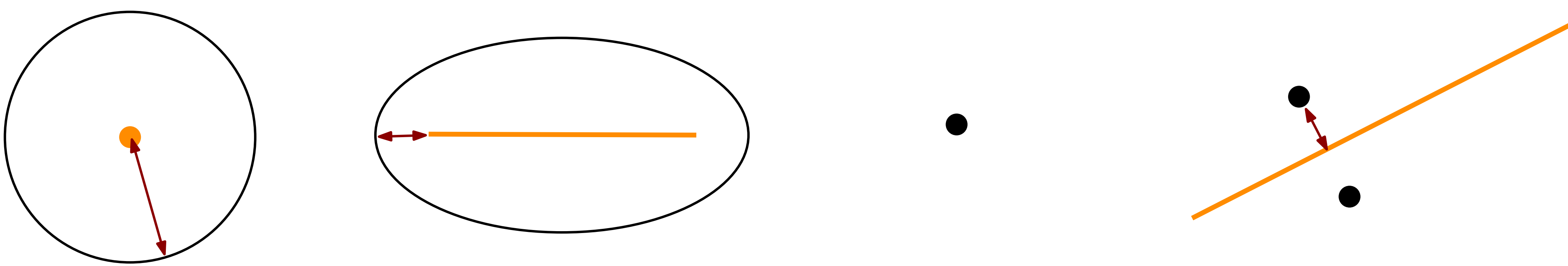
Medial axis and reach

The *medial axis* of X is the subset $\mathrm{med}\,(X) \subset \mathbb{R}^n$ which consists of points $y \in \mathbb{R}^n$ that admit at least two projections on X:

$$\mathrm{med}\,(X) = \{y \in \mathbb{R}^n, \exists x, x' \in X, x \neq x', \|y - x\| = \|y - x'\| = \mathrm{dist}\,(y, X)\}\,.$$

The *reach* of X is

$$\begin{aligned} \mathrm{reach}\,(X) &= \inf\,\{\mathrm{dist}\,(y, X)\,, y \in \mathrm{med}\,(X)\} \\ &= \inf\,\{\|x - y\|\,, x \in X, y \in \mathrm{med}\,(X)\}\,. \end{aligned}$$



Proposition: For every $t \in [0, \mathrm{reach}\,(X))$, the spaces X and X^t are homotopy equivalent.

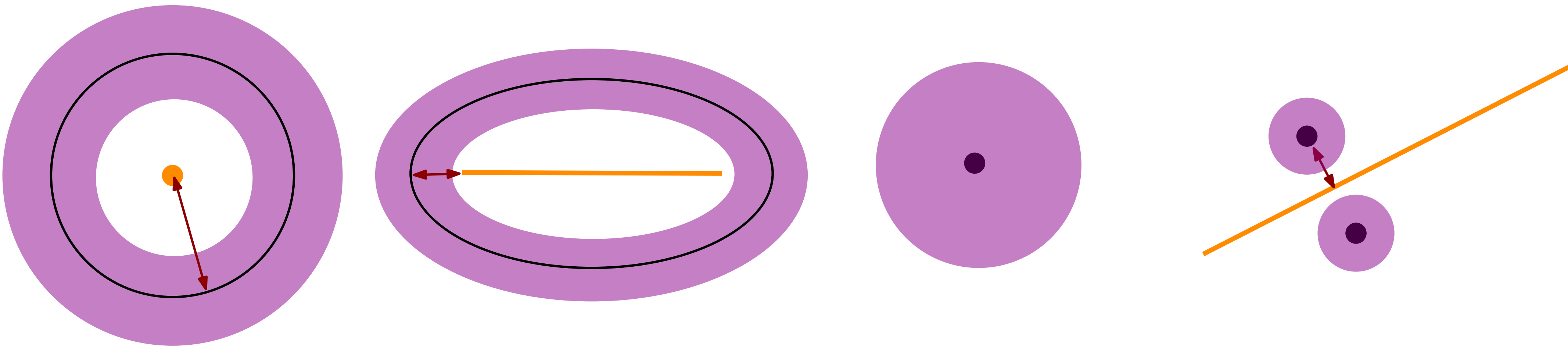
Medial axis and reach

The *medial axis* of X is the subset $\mathrm{med}(X) \subset \mathbb{R}^n$ which consists of points $y \in \mathbb{R}^n$ that admit at least two projections on X:

$$\mathrm{med}(X) = \{y \in \mathbb{R}^n, \exists x, x' \in X, x \neq x', \|y - x\| = \|y - x'\| = \mathrm{dist}(y, X)\}.$$

The *reach* of X is

$$\begin{aligned}\mathrm{reach}(X) &= \inf\{\mathrm{dist}(y, X), y \in \mathrm{med}(X)\} \\ &= \inf\{\|x - y\|, x \in X, y \in \mathrm{med}(X)\}.\end{aligned}$$



Proposition: For every $t \in [0, \mathrm{reach}(X))$, the spaces X and X^t are homotopy equivalent.

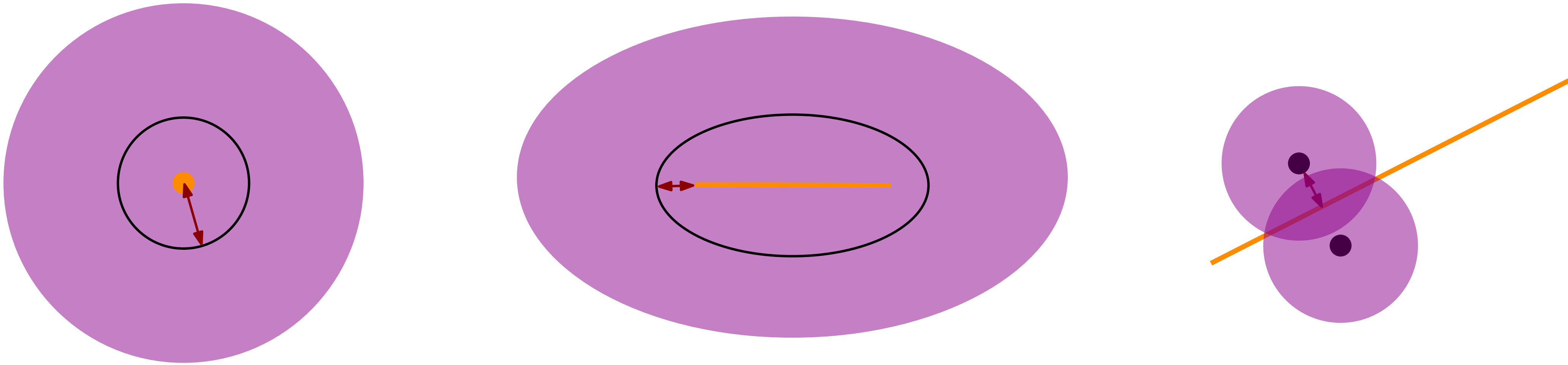
Medial axis and reach

The *medial axis* of X is the subset $\mathrm{med}\,(X) \subset \mathbb{R}^n$ which consists of points $y \in \mathbb{R}^n$ that admit at least two projections on X:

$$\mathrm{med}\,(X) = \{y \in \mathbb{R}^n, \exists x, x' \in X, x \neq x', \|y - x\| = \|y - x'\| = \mathrm{dist}\,(y, X)\}\,.$$

The *reach* of X is

$$\begin{aligned}\mathrm{reach}\,(X) &= \inf\{\mathrm{dist}\,(y, X)\,, y \in \mathrm{med}\,(X)\}\\ &= \inf\{\|x - y\|\,, x \in X, y \in \mathrm{med}\,(X)\}\,.\end{aligned}$$



Proposition: For every $t \in [0, \mathrm{reach}\,(X))$, the spaces X and X^t are homotopy equivalent.

If $t \geq \mathrm{reach}\,(X)$, the sets X and X^t may not be homotopy equivalent.

Proposition: For every $t \in [0, \mathrm{reach}\,(X))$, the spaces X and X^t are homotopy equivalent.

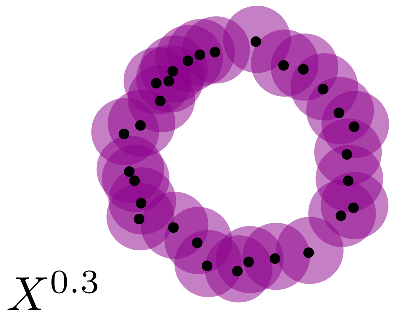
Proof: For every $t \in [0, \mathrm{reach}\,(X))$, the thickening X^t deform retracts onto X. A homotopy is given by the following map:

$$\begin{aligned} X^t \times [0,1] &\longrightarrow X^t \\ (x,t) &\longmapsto (1-t)x + t \cdot \mathrm{proj}\,(x, X)\,. \end{aligned}$$

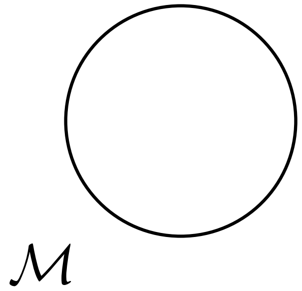
Indeed, the projection $\mathrm{proj}\,(x, X)$ is well defined (it is unique).

Selection of the parameter t

Remember Question 1: How to select a t such that $X^t \approx \mathcal{M}$?



$\approx$

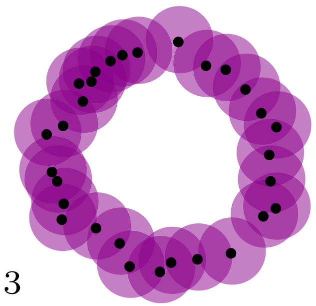


Theorem (Frédéric Chazal, David Cohen-Steiner, and André Lieutier, 2009):
Let X and $\mathcal{M}$ be subsets of $\mathbb{R}^n$. Suppose that $\mathcal{M}$ has positive reach, and that $\mathrm{d_H}\left(X, \mathcal{M}\right) \leq \frac{1}{17}\mathrm{reach}\left(\mathcal{M}\right)$.
Then X^t and $\mathcal{M}$ are homotopic equivalent, provided that

$$t \in \left[4\mathrm{d_H}\left(X, \mathcal{M}\right), \mathrm{reach}\left(\mathcal{M}\right) - 3\mathrm{d_H}\left(X, \mathcal{M}\right)\right).$$

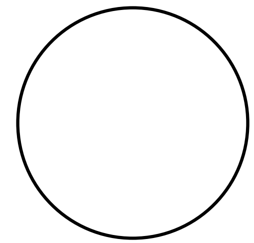
Selection of the parameter t

Remember Question 1: How to select a t such that $X^t \approx \mathcal{M}$?



$X^{0.3}$

$\approx$



$\mathcal{M}$

Theorem (Frédéric Chazal, David Cohen-Steiner, and André Lieutier, 2009):
Let X and $\mathcal{M}$ be subsets of $\mathbb{R}^n$. Suppose that $\mathcal{M}$ has positive reach, and that $\mathrm{d_H}(X, \mathcal{M}) \leq \frac{1}{17}\mathrm{reach}(\mathcal{M})$.
Then X^t and $\mathcal{M}$ are homotopic equivalent, provided that

$$t \in \left[4\mathrm{d_H}(X, \mathcal{M}), \mathrm{reach}(\mathcal{M}) - 3\mathrm{d_H}(X, \mathcal{M})\right).$$

Theorem (Partha Niyogi, Stephen Smale, and Shmuel Weinberger, 2008):
Let X and $\mathcal{M}$ be subsets of $\mathbb{R}^n$, with $\mathcal{M}$ a submanifold, and X a finite subset of $\mathcal{M}$.
Suppose that $\mathcal{M}$ has positive reach.
Then X^t and $\mathcal{M}$ are homotopic equivalent, provided that

$$t \in \left[2\mathrm{d_H}(X, \mathcal{M}), \sqrt{\frac{3}{5}}\mathrm{reach}(\mathcal{M})\right).$$

I - Thickenings

II - Čech complex

III - Rips complex

(Weak) triangulations

Let us consider Question 2: How to compute the homology groups of X^t?

We must a triangulation of X^t, that is: a simplicial complex K homeomorphic to X.

Actually, we will define something weaker: a simplicial complex K that is homotopy equivalent to X.

(Weak) triangulations

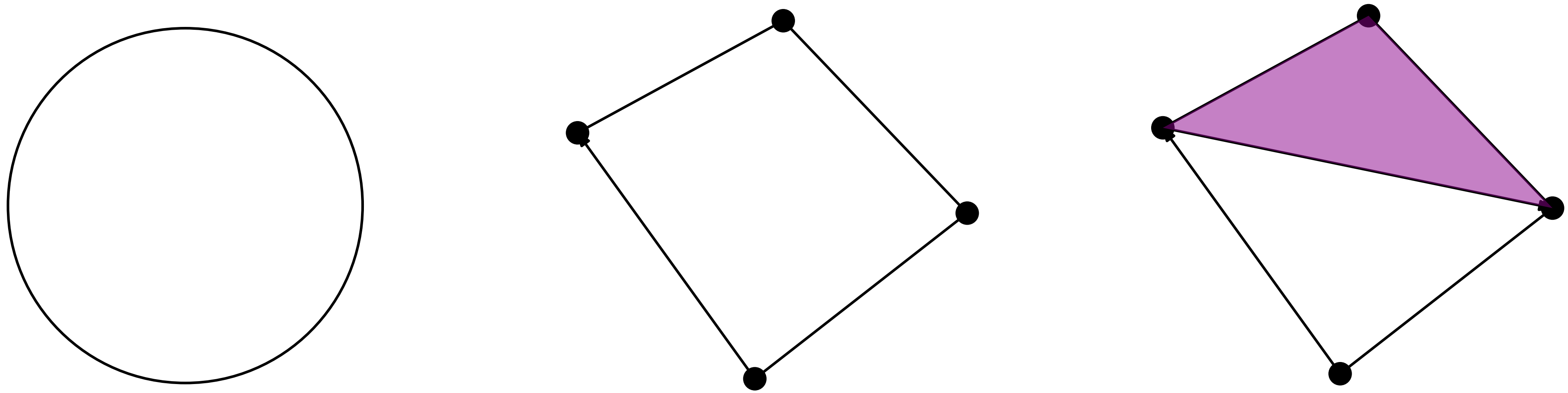
Let us consider Question 2: How to compute the homology groups of X^t?

We must a triangulation of X^t, that is: a simplicial complex K homeomorphic to X.

Actually, we will define something weaker: a simplicial complex K that is homotopy equivalent to X.

weak triangulation

Either case, we will have $\beta_i(X) = \beta_i(K)$ for all $i \geq 0$.

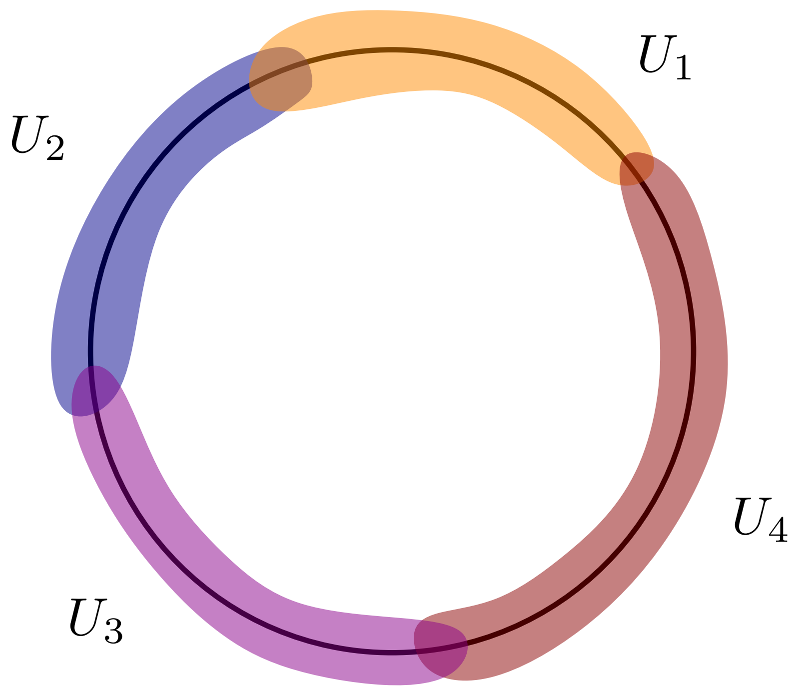


Nerves

Definition: Let X be a topological space, and $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$ a cover of X, that is, a collection of subsets $U_i \subset X$ such that

$$\bigcup_{1 \leq i \leq N} U_i = X.$$

The *nerve* of $\mathcal{U}$ is the simplicial complex with vertex set $\{1, ..., N\}$ and whose m-simplices are the subsets $\{i_1, ..., i_m\} \subset \{1, ..., N\}$ such that $\bigcap_{k=0}^{m} U_{i_k} \neq \emptyset$. It is denoted $\mathcal{N}(\mathcal{U})$.

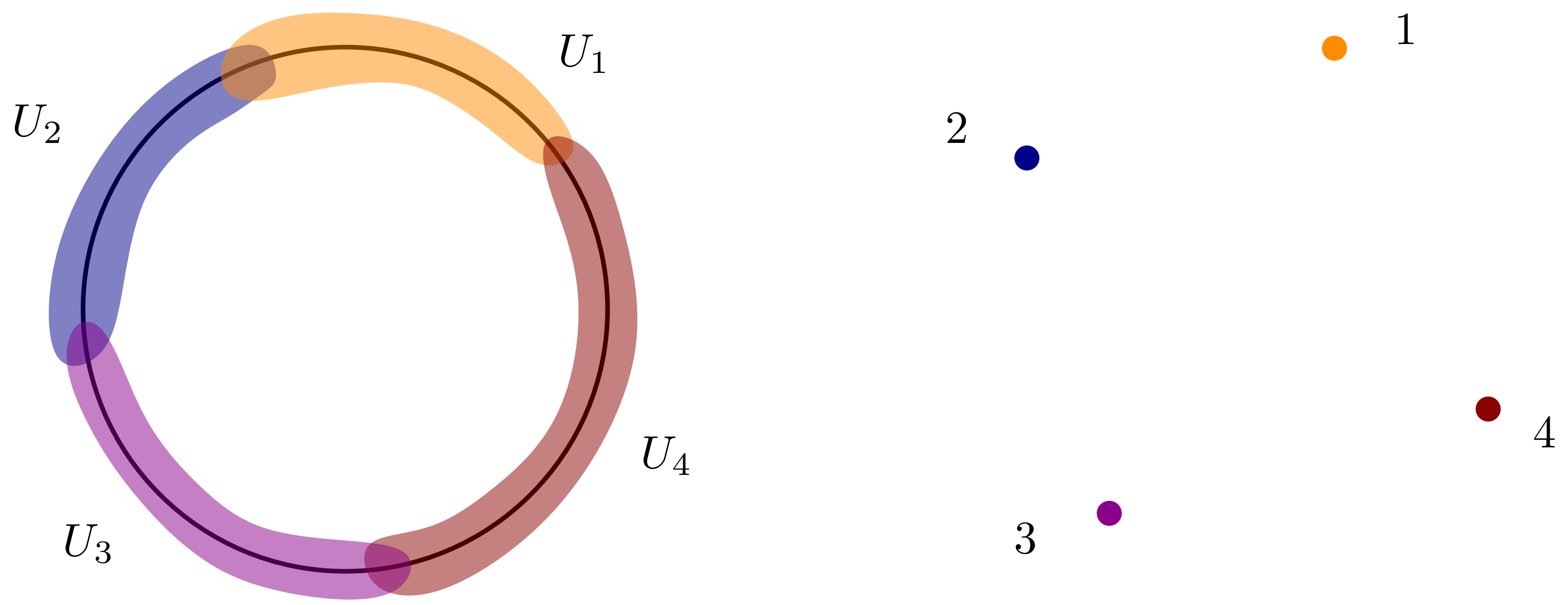


Nerves

Definition: Let X be a topological space, and $\mathcal{U} = \{U_i\}_{1 \le i \le N}$ a cover of X, that is, a collection of subsets $U_i \subset X$ such that

$$\bigcup_{1 \le i \le N} U_i = X.$$

The *nerve* of $\mathcal{U}$ is the simplicial complex with vertex set $\{1, ..., N\}$ and whose m-simplices are the subsets $\{i_1, ..., i_m\} \subset \{1, ..., N\}$ such that $\bigcap_{k=0}^{m} U_{i_k} \neq \emptyset$. It is denoted $\mathcal{N}(\mathcal{U})$.

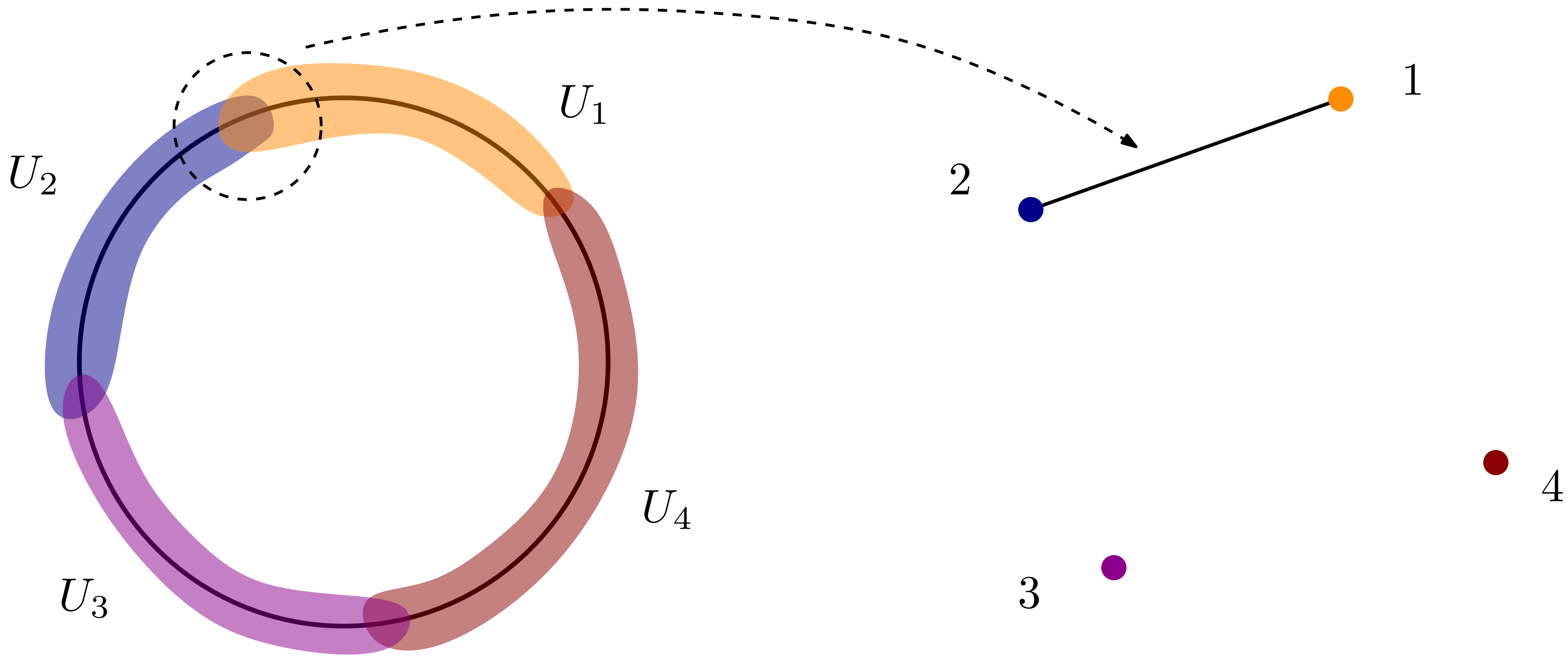


Nerves

Definition: Let X be a topological space, and $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$ a cover of X, that is, a collection of subsets $U_i \subset X$ such that

$$\bigcup_{1 \leq i \leq N} U_i = X.$$

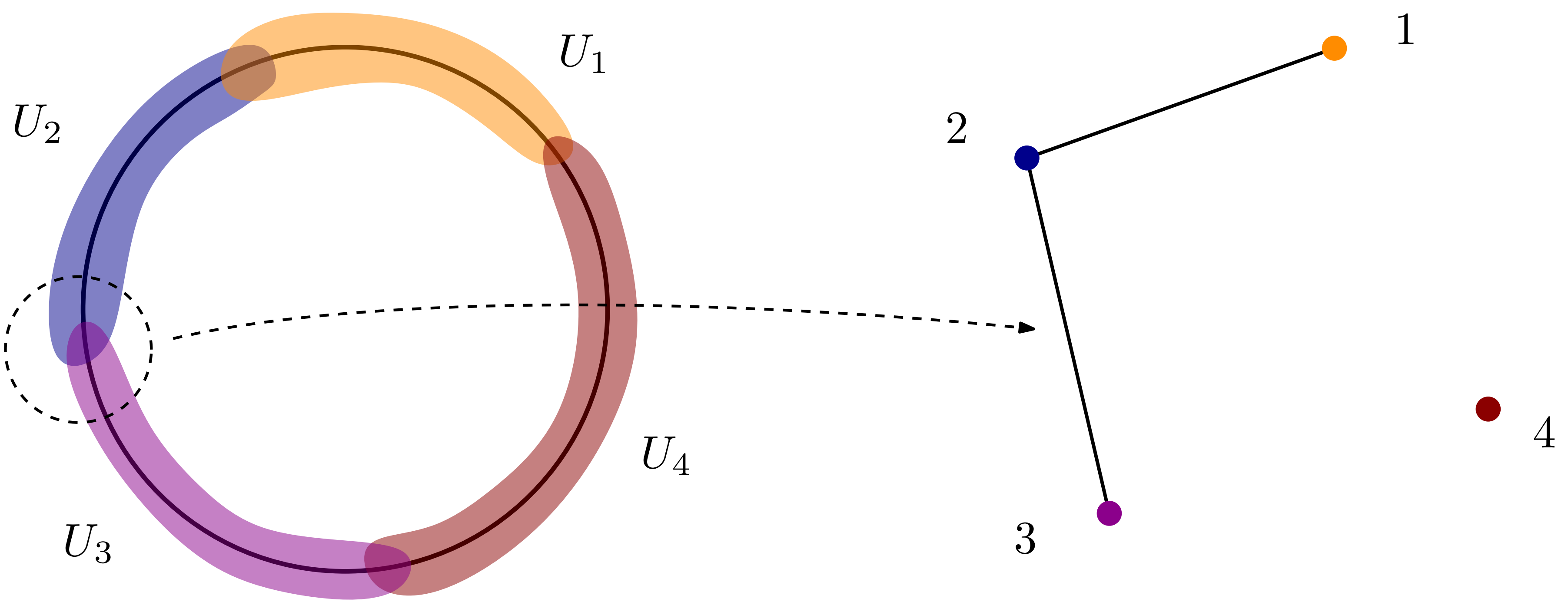
The *nerve* of $\mathcal{U}$ is the simplicial complex with vertex set $\{1, \ldots, N\}$ and whose m-simplices are the subsets $\{i_1, \ldots, i_m\} \subset \{1, \ldots, N\}$ such that $\bigcap_{k=0}^{m} U_{i_k} \neq \emptyset$. It is denoted $\mathcal{N}(\mathcal{U})$.



Definition: Let X be a topological space, and $\mathcal{U} = \{U_i\}_{1 \le i \le N}$ a cover of X, that is, a collection of subsets $U_i \subset X$ such that

$$\bigcup_{1 \le i \le N} U_i = X.$$

The *nerve* of $\mathcal{U}$ is the simplicial complex with vertex set $\{1, ..., N\}$ and whose m-simplices are the subsets $\{i_1, ..., i_m\} \subset \{1, ..., N\}$ such that $\bigcap_{k=0}^{m} U_{i_k} \neq \emptyset$. It is denoted $\mathcal{N}(\mathcal{U})$.

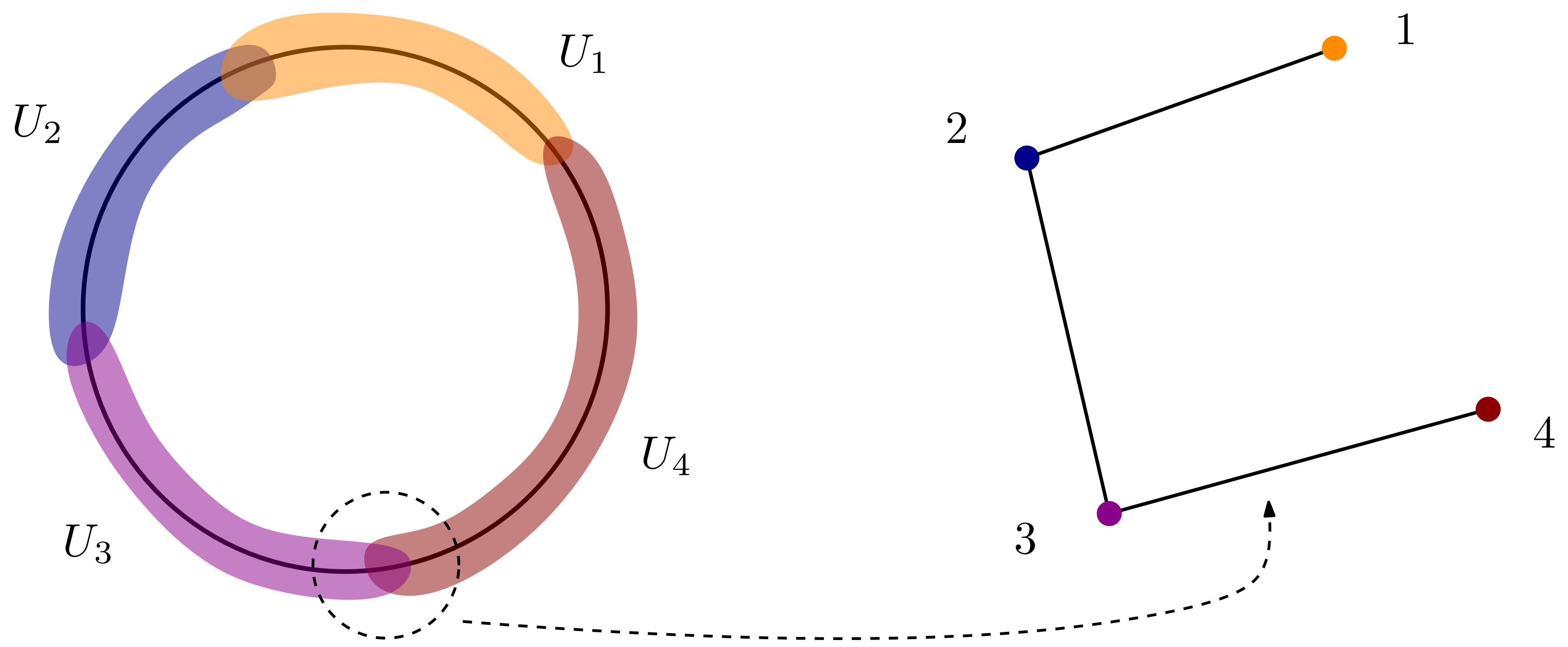


Nerves

Definition: Let X be a topological space, and $\mathcal{U} = \{U_i\}_{1 \le i \le N}$ a cover of X, that is, a collection of subsets $U_i \subset X$ such that

$$\bigcup_{1 \le i \le N} U_i = X.$$

The *nerve* of $\mathcal{U}$ is the simplicial complex with vertex set $\{1, ..., N\}$ and whose m-simplices are the subsets $\{i_1, ..., i_m\} \subset \{1, ..., N\}$ such that $\bigcap_{k=0}^{m} U_{i_k} \neq \emptyset$. It is denoted $\mathcal{N}(\mathcal{U})$.

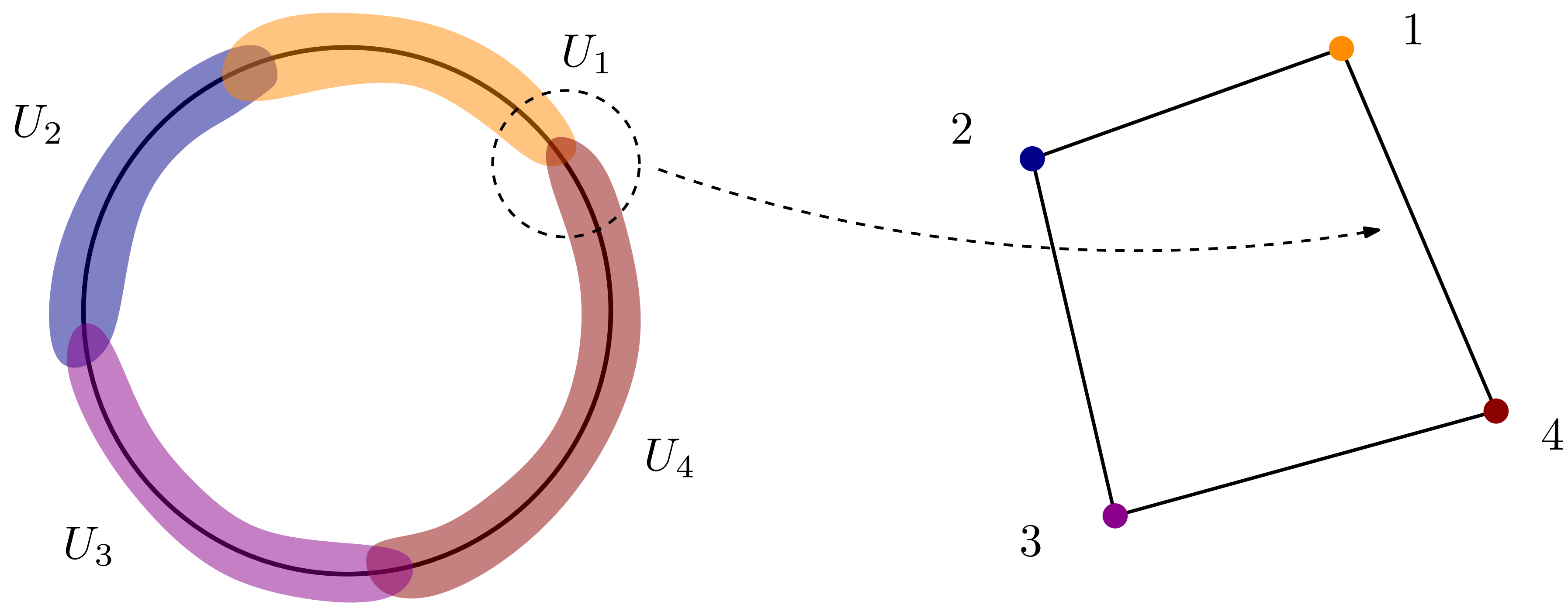


Nerves

Definition: Let X be a topological space, and $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$ a cover of X, that is, a collection of subsets $U_i \subset X$ such that

$$\bigcup_{1 \leq i \leq N} U_i = X.$$

The *nerve* of $\mathcal{U}$ is the simplicial complex with vertex set $\{1, ..., N\}$ and whose m-simplices are the subsets $\{i_1, ..., i_m\} \subset \{1, ..., N\}$ such that $\bigcap_{k=0}^{m} U_{i_k} \neq \emptyset$. It is denoted $\mathcal{N}(\mathcal{U})$.

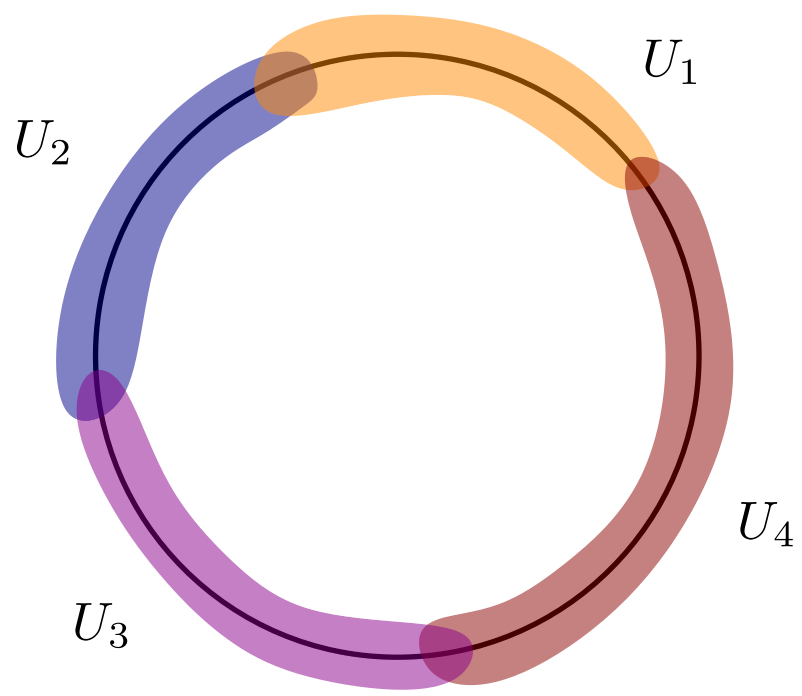


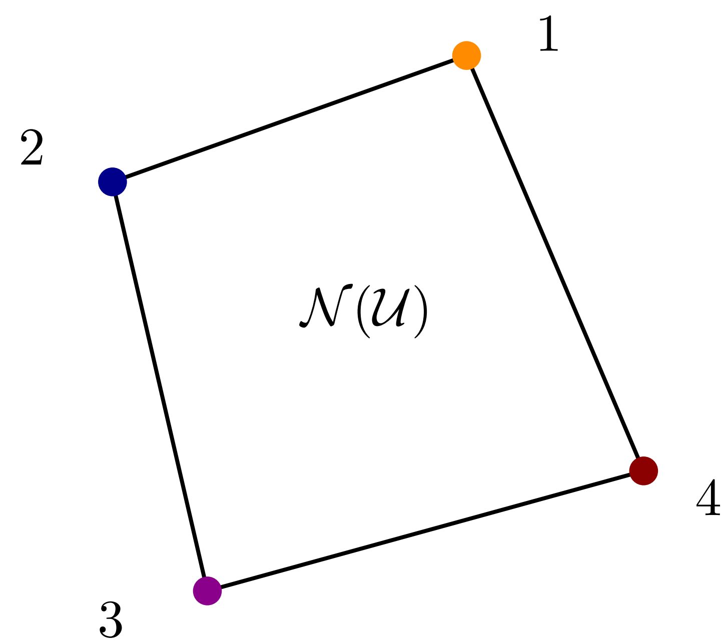
Nerves

Definition: Let X be a topological space, and $\mathcal{U} = \{U_i\}_{1 \le i \le N}$ a cover of X, that is, a collection of subsets $U_i \subset X$ such that

$$\bigcup_{1 \le i \le N} U_i = X.$$

The *nerve* of $\mathcal{U}$ is the simplicial complex with vertex set $\{1, ..., N\}$ and whose m-simplices are the subsets $\{i_1, ..., i_m\} \subset \{1, ..., N\}$ such that $\bigcap_{k=0}^{m} U_{i_k} \neq \emptyset$. It is denoted $\mathcal{N}(\mathcal{U})$.



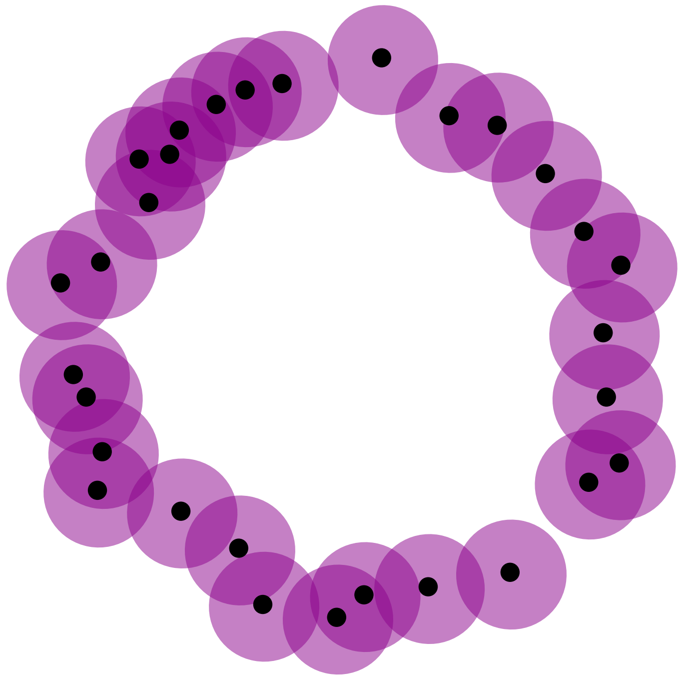


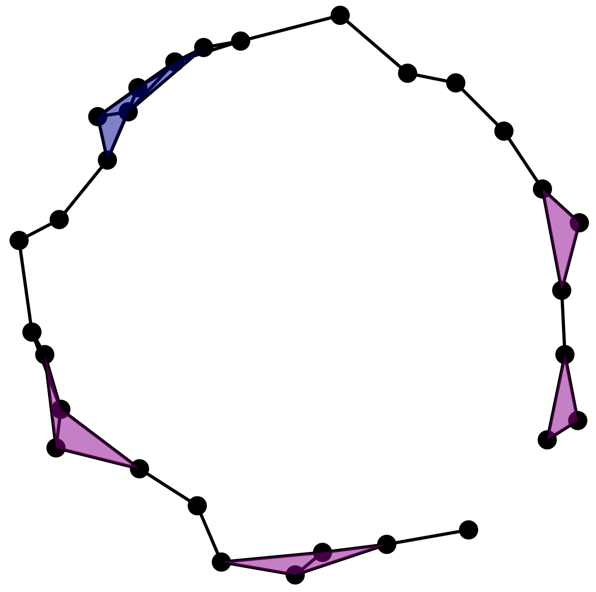
Nerves

Definition: Let X be a topological space, and $\mathcal{U} = \{U_i\}_{1\leq i\leq N}$ a cover of X, that is, a collection of subsets $U_i \subset X$ such that

$$\bigcup_{1\leq i\leq N} U_i = X.$$

The *nerve* of $\mathcal{U}$ is the simplicial complex with vertex set $\{1, ..., N\}$ and whose m-simplices are the subsets $\{i_1, ..., i_m\} \subset \{1, ..., N\}$ such that $\bigcap_{k=0}^{m} U_{i_k} \neq \emptyset$. It is denoted $\mathcal{N}(\mathcal{U})$.





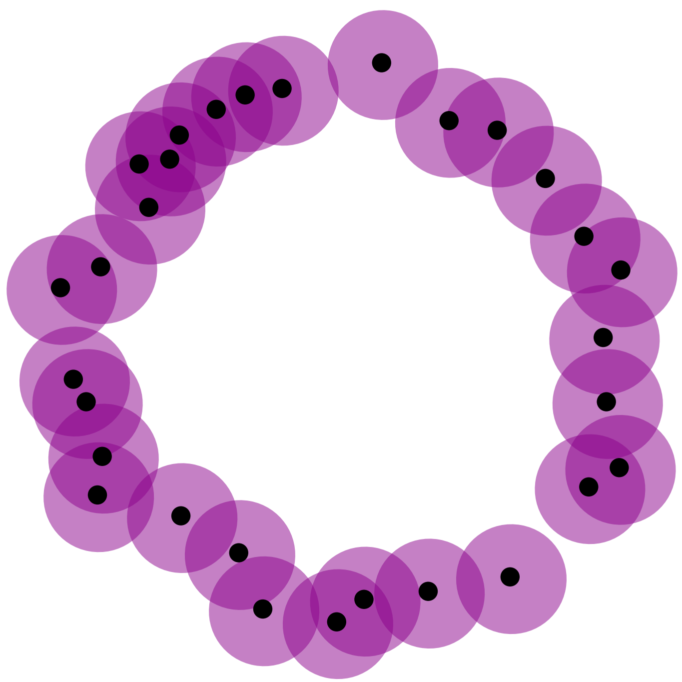
$X^{0.2} = \bigcup_{x\in X} \overline{\mathcal{B}}\left(x, 0.2\right)$ is covered by $\mathcal{U} = \left\{\overline{\mathcal{B}}\left(x, 0.2\right), x \in X\right\}$

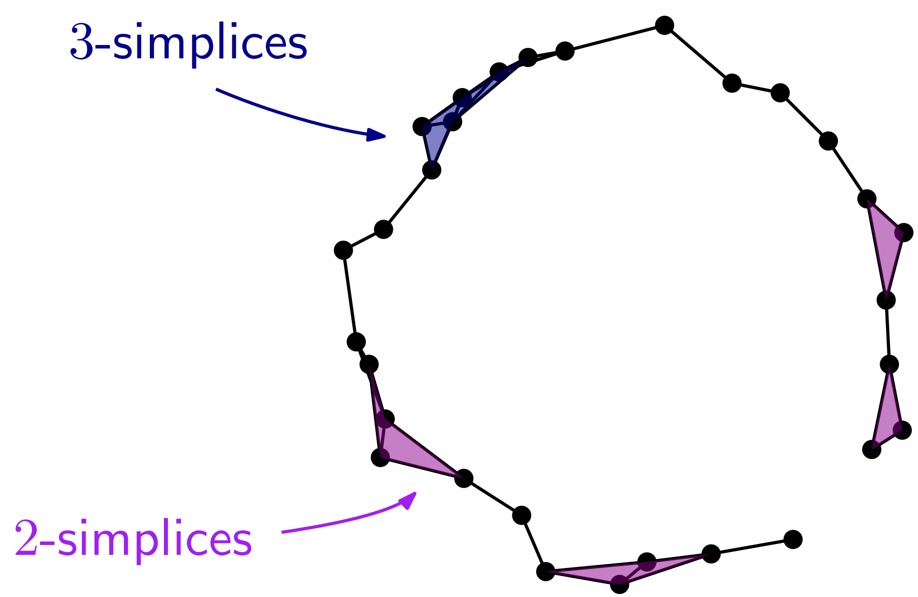
Nerves

Definition: Let X be a topological space, and $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$ a cover of X, that is, a collection of subsets $U_i \subset X$ such that

$$\bigcup_{1 \leq i \leq N} U_i = X.$$

The *nerve* of $\mathcal{U}$ is the simplicial complex with vertex set $\{1, ..., N\}$ and whose m-simplices are the subsets $\{i_1, ..., i_m\} \subset \{1, ..., N\}$ such that $\bigcap_{k=0}^{m} U_{i_k} \neq \emptyset$. It is denoted $\mathcal{N}(\mathcal{U})$.





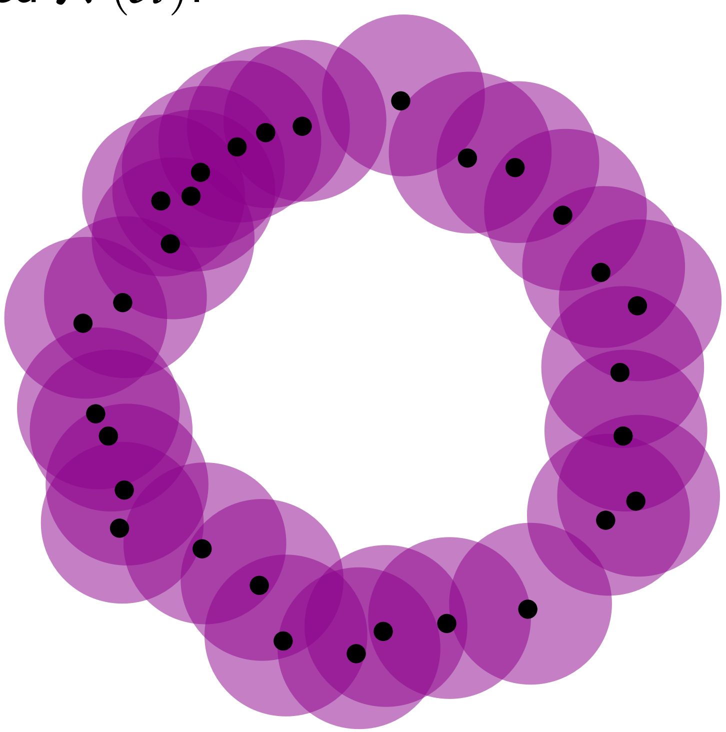
$X^{0.2} = \bigcup_{x \in X} \overline{\mathcal{B}}(x, 0.2)$ is covered by $\mathcal{U} = \{\overline{\mathcal{B}}(x, 0.2), x \in X\}$

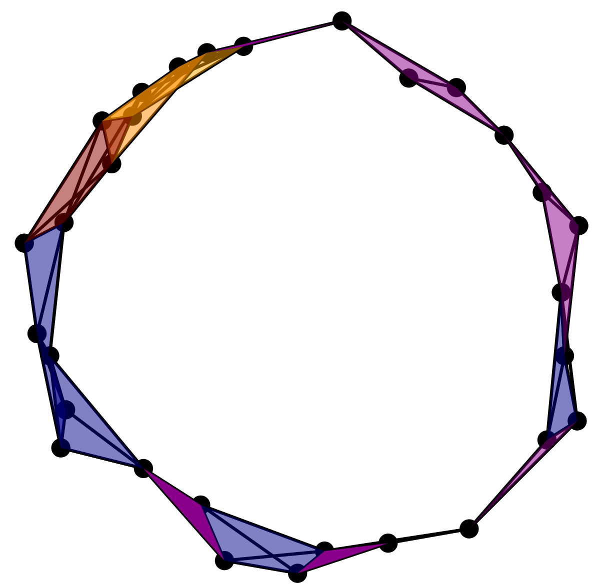
Nerves

Definition: Let X be a topological space, and $\mathcal{U} = \{U_i\}_{1 \le i \le N}$ a cover of X, that is, a collection of subsets $U_i \subset X$ such that

$$\bigcup_{1 \le i \le N} U_i = X.$$

The *nerve* of $\mathcal{U}$ is the simplicial complex with vertex set $\{1, ..., N\}$ and whose m-simplices are the subsets $\{i_1, ..., i_m\} \subset \{1, ..., N\}$ such that $\bigcap_{k=0}^{m} U_{i_k} \neq \emptyset$. It is denoted $\mathcal{N}(\mathcal{U})$.





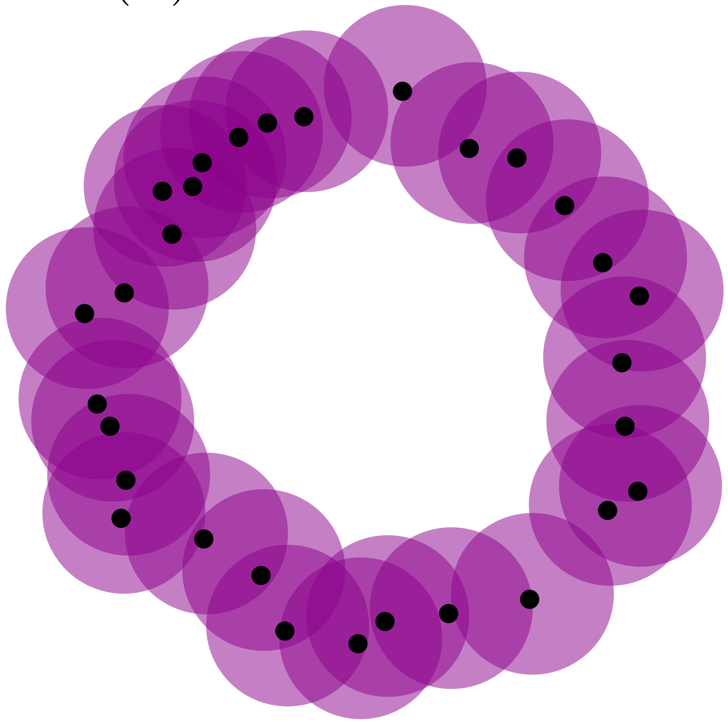
$X^{0.3} = \bigcup_{x \in X} \overline{\mathcal{B}}\,(x, 0.3)$ is covered by $\mathcal{U} = \left\{\overline{\mathcal{B}}\,(x, 0.3)\,, x \in X\right\}$

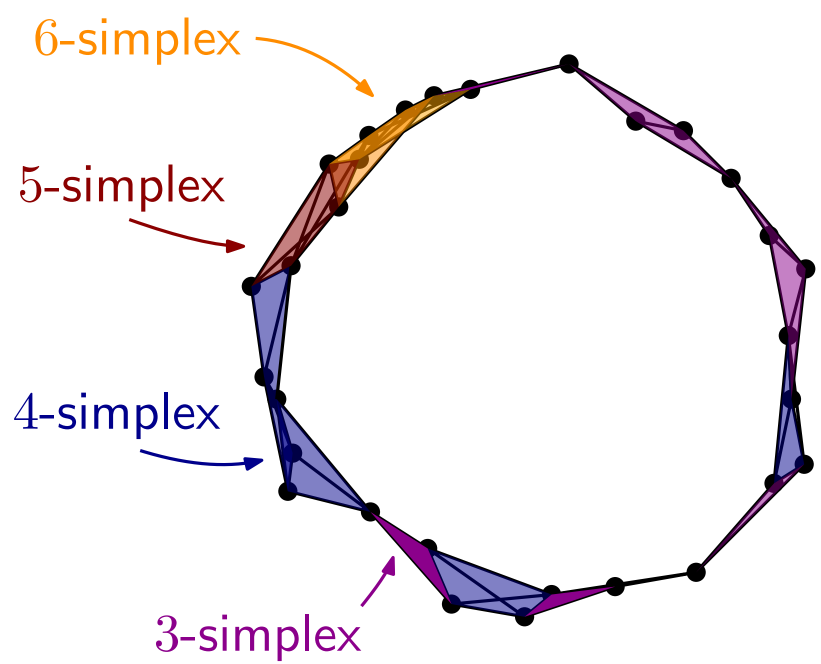
Nerves

Definition: Let X be a topological space, and $\mathcal{U} = \{U_i\}_{1 \le i \le N}$ a cover of X, that is, a collection of subsets $U_i \subset X$ such that

$$\bigcup_{1 \le i \le N} U_i = X.$$

The *nerve* of $\mathcal{U}$ is the simplicial complex with vertex set $\{1, ..., N\}$ and whose m-simplices are the subsets $\{i_1, ..., i_m\} \subset \{1, ..., N\}$ such that $\bigcap_{k=0}^{m} U_{i_k} \neq \emptyset$. It is denoted $\mathcal{N}(\mathcal{U})$.





$X^{0.3} = \bigcup_{x \in X} \overline{\mathcal{B}}(x, 0.3)$ is covered by $\mathcal{U} = \left\{\overline{\mathcal{B}}(x, 0.3), x \in X\right\}$

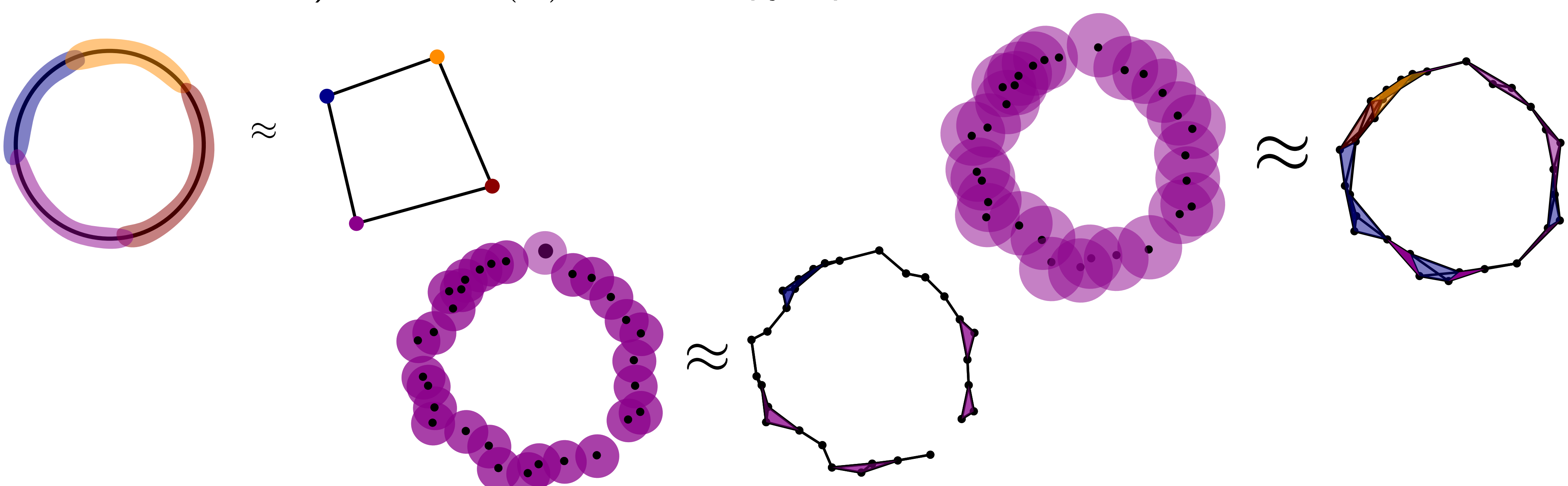
Nerves

Definition: Let X be a topological space, and $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$ a cover of X, that is, a collection of subsets $U_i \subset X$ such that

$$\bigcup_{1 \leq i \leq N} U_i = X.$$

The *nerve* of $\mathcal{U}$ is the simplicial complex with vertex set $\{1, ..., N\}$ and whose m-simplices are the subsets $\{i_1, ..., i_m\} \subset \{1, ..., N\}$ such that $\bigcap_{k=0}^{m} U_{i_k} \neq \emptyset$. It is denoted $\mathcal{N}(\mathcal{U})$.

Nerve theorem: Consider $X \subset \mathbb{R}^n$. Suppose that each U_i are balls (or more generally, closed and convex). Then $\mathcal{N}(\mathcal{U})$ is homotopy equivalent to X.



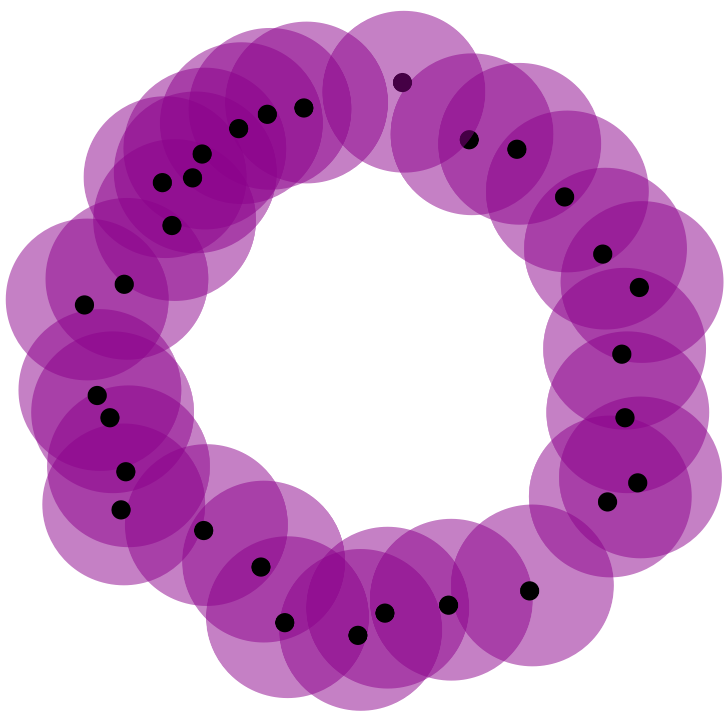
Čech complex

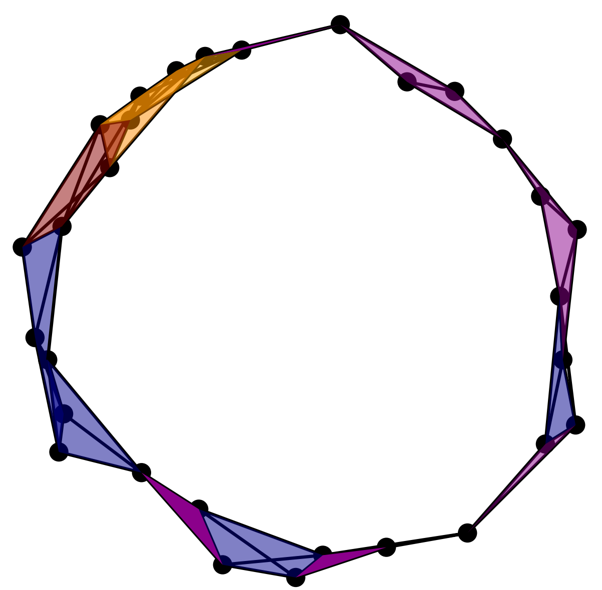
Let X be a finite subset of $\mathbb{R}^n$, and $t \geq 0$. Consider the collection

$$\mathcal{V}^t = \left\{ \overline{\mathcal{B}}\left(x, t\right), x \in X \right\}.$$

This is a cover of the thickening X^t, and each components are closed balls.
By Nerve Theorem, its nerve $\mathcal{N}(\mathcal{V}^t)$ has the homotopy type of X^t.

Definition: This nerve is denoted $\check{\mathrm{C}}\mathrm{ech}^t(X)$ and is called the *Čech complex of X at time t*.





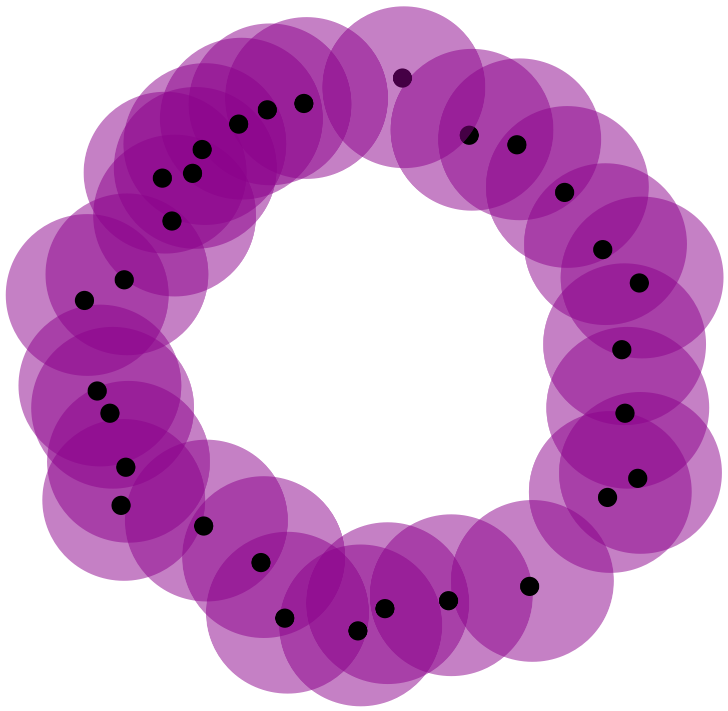
Čech complex

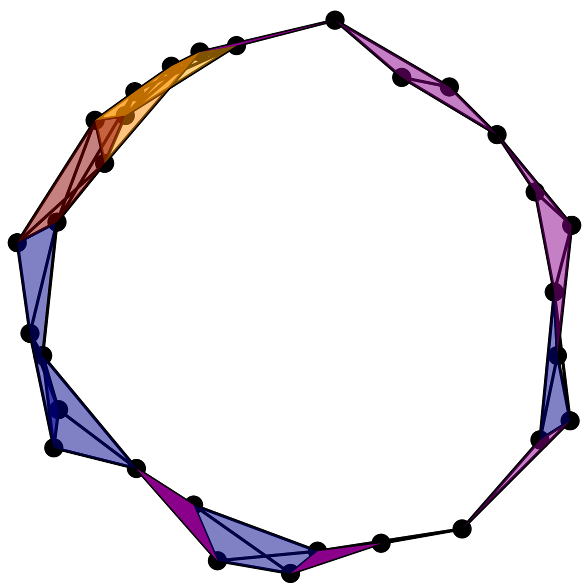
Let X be a finite subset of $\mathbb{R}^n$, and $t \geq 0$. Consider the collection

$$\mathcal{V}^t = \left\{ \overline{\mathcal{B}}\left(x,t\right), x \in X \right\}.$$

This is a cover of the thickening X^t, and each components are closed balls.
By Nerve Theorem, its nerve $\mathcal{N}(\mathcal{V}^t)$ has the homotopy type of X^t.

Definition: This nerve is denoted $\check{\mathrm{C}}\mathrm{ech}^t(X)$ and is called the *Čech complex of X at time t*.





$\longrightarrow$ The Question 2 (How to compute the homology groups of X^t?) is solved.

I - Thickenings

II - Čech complex

III - Rips complex

Computation of the Čech complex

Let $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^n$ be finite, let $t \geq 0$ and consider the t-thickening

$$X^t = \bigcup_{x \in X} \overline{\mathcal{B}}\,(x, t)\,.$$

By definition, its nerve, $\check{\mathrm{C}}\mathrm{ech}^t(X)$, the Čech complex at time t, is a simplicial complex on the vertices $\{1, \ldots, N\}$ whose simplices are the subsets $\{i_1, ..., i_m\}$ such that

$$\bigcap_{1 \leq k \leq m} \overline{\mathcal{B}}\,(x_{i_k}, t) \neq \emptyset.$$

Computation of the Čech complex

Let $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^n$ be finite, let $t \geq 0$ and consider the t-thickening

$$X^t = \bigcup_{x \in X} \overline{\mathcal{B}}\,(x, t)\,.$$

By definition, its nerve, $\check{\mathrm{C}}\mathrm{ech}^t(X)$, the Čech complex at time t, is a simplicial complex on the vertices $\{1, \ldots, N\}$ whose simplices are the subsets $\{i_1, ..., i_m\}$ such that

$$\bigcap_{1 \leq k \leq m} \overline{\mathcal{B}}\,(x_{i_k}, t) \neq \emptyset.$$

Therefore, computing the Čech complex relies on the following geometric predicate:

Given m closed balls of $\mathbb{R}^n$, do they intersect?

This problem is known as the *smallest circle problem*.
It can can be solved in $O(m)$ time, where m is the number of points.

Computation of the Čech complex

Let $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^n$ be finite, let $t \geq 0$ and consider the t-thickening

$$X^t = \bigcup_{x \in X} \overline{\mathcal{B}}\,(x, t)\,.$$

By definition, its nerve, $\check{\mathrm{C}}\mathrm{ech}^t(X)$, the Čech complex at time t, is a simplicial complex on the vertices $\{1, \ldots, N\}$ whose simplices are the subsets $\{i_1, ..., i_m\}$ such that

$$\bigcap_{1 \leq k \leq m} \overline{\mathcal{B}}\,(x_{i_k}, t) \neq \emptyset.$$

Therefore, computing the Čech complex relies on the following geometric predicate:

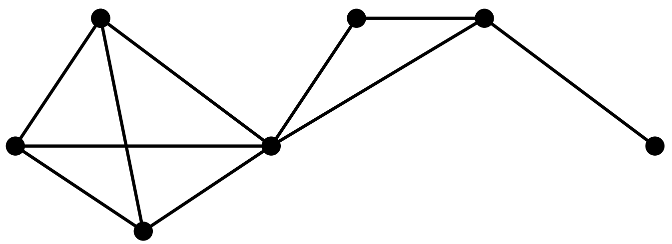
Given m closed balls of $\mathbb{R}^n$, do they intersect?

This problem is known as the *smallest circle problem*.
It can can be solved in $O(m)$ time, where m is the number of points.

$\longrightarrow$ in practice, we prefer a more simple version

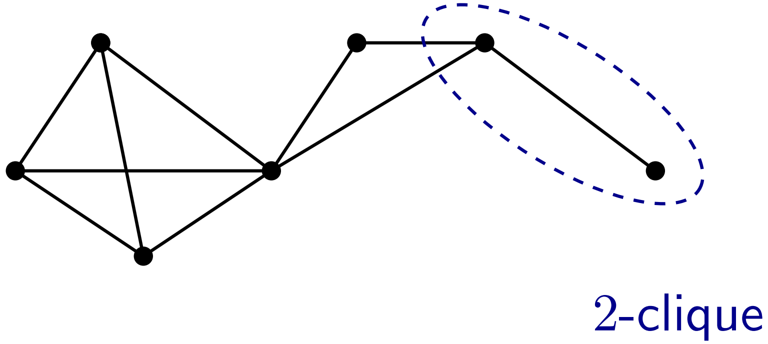
Clique complex

Let G be a graph.
We call a *clique* of G a set of vertices $v_1, ..., v_m$ such that for every $i, j \in [\![1, m]\!]$ with $i \neq j$, the edge $[v_1, v_j]$ belongs to G.



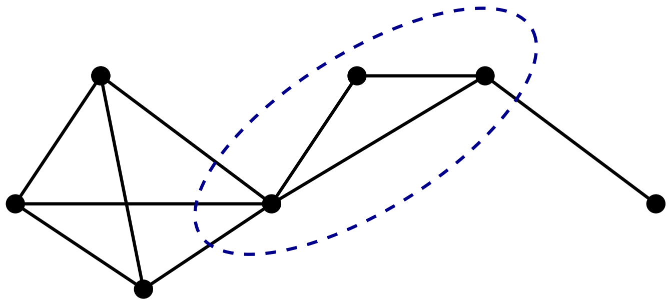
Clique complex

Let G be a graph.
We call a *clique* of G a set of vertices $v_1, ..., v_m$ such that for every $i, j \in [\![1, m]\!]$ with $i \neq j$, the edge $[v_1, v_j]$ belongs to G.



Clique complex

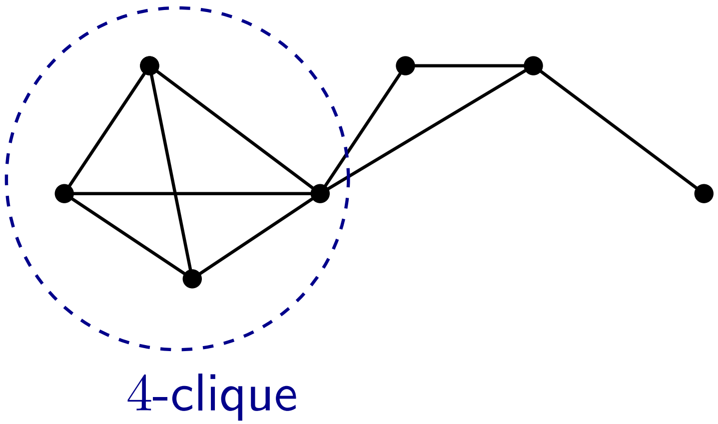
Let G be a graph.
We call a *clique* of G a set of vertices $v_1, ..., v_m$ such that for every $i, j \in [\![1, m]\!]$ with $i \neq j$, the edge $[v_1, v_j]$ belongs to G.



3-clique

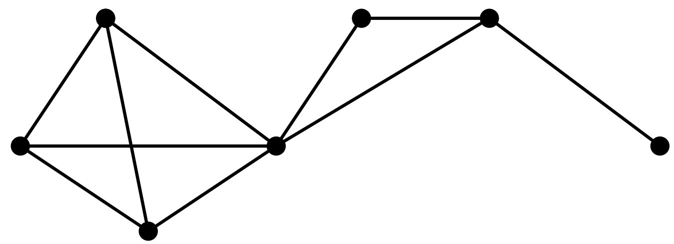
Clique complex

Let G be a graph.
We call a *clique* of G a set of vertices $v_1, ..., v_m$ such that for every $i, j \in [\![1, m]\!]$ with $i \neq j$, the edge $[v_1, v_j]$ belongs to G.



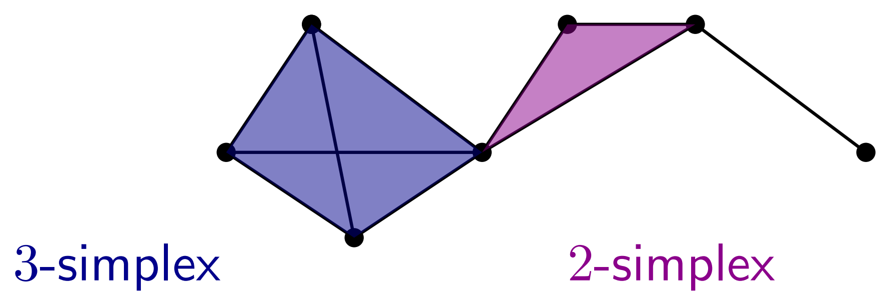
Clique complex

Let G be a graph.
We call a *clique* of G a set of vertices $v_1, ..., v_m$ such that for every $i, j \in [\![1, m]\!]$ with $i \neq j$, the edge $[v_1, v_j]$ belongs to G.



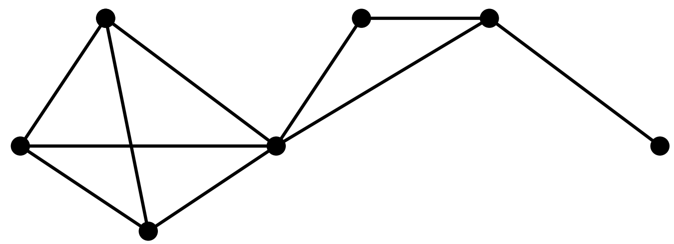
Definition: Given a graph G, the corresponding *clique complex* is the simplicial complex whose

- vertices are the vertices of G,
- simplices are the sets of vertices of the cliques of G.



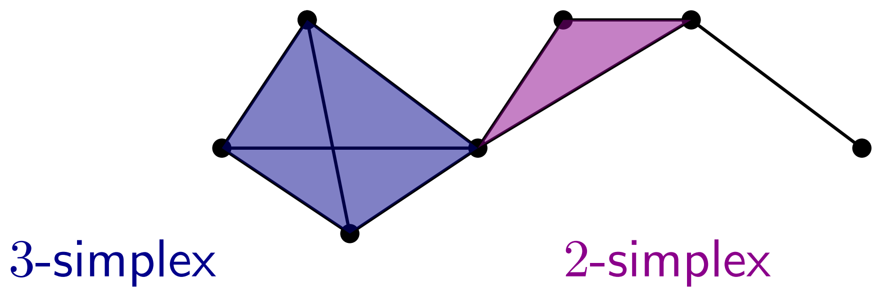
Clique complex

Let G be a graph.
We call a *clique* of G a set of vertices $v_1, ..., v_m$ such that for every $i, j \in [\![1, m]\!]$ with $i \neq j$, the edge $[v_1, v_j]$ belongs to G.



Definition: Given a graph G, the corresponding *clique complex* is the simplicial complex whose

- vertices are the vertices of G,
- simplices are the sets of vertices of the cliques of G.



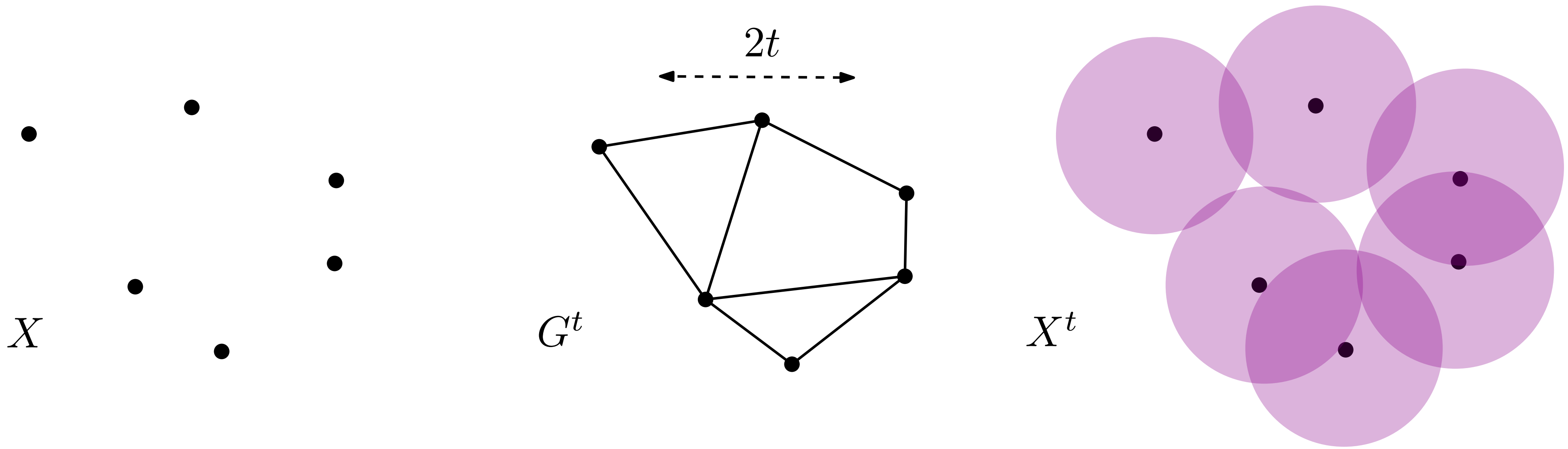
Exercise: Prove that the clique complex of a graph is a simplicial complex.

Rips complex

Let $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^n$ and $t \geq 0$.

Consider the graph G^t whose vertex set is $\{1, \ldots, N\}$, and whose edges are the pairs (i, j) such that $\|x_i - x_j\| \leq 2t$.

Alternatively, G^t can be seen as the 1-skeleton of the Čech complex $\check{\mathrm{C}}\mathrm{ech}^t(X)$.

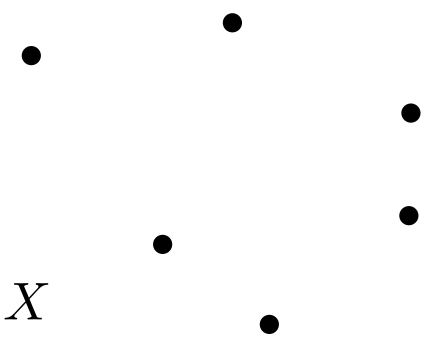


Rips complex

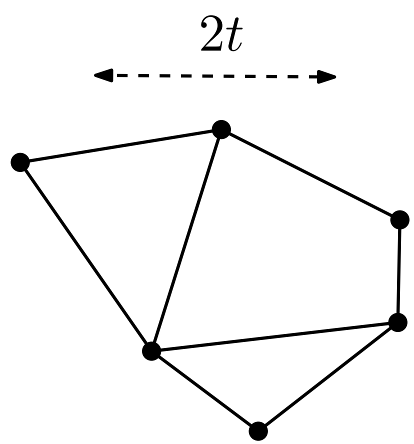
Let $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^n$ and $t \geq 0$.

Consider the graph G^t whose vertex set is $\{1, \ldots, N\}$, and whose edges are the pairs (i, j) such that $\|x_i - x_j\| \leq 2t$.

Alternatively, G^t can be seen as the 1-skeleton of the Čech complex $\check{\mathrm{C}}\mathrm{ech}^t(X)$.

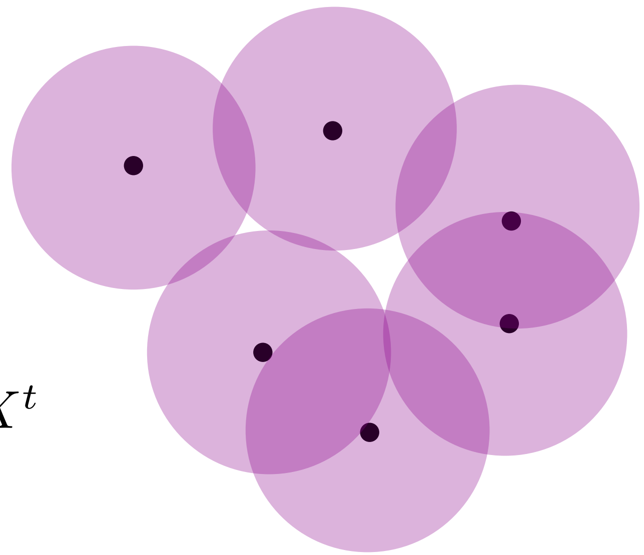


X



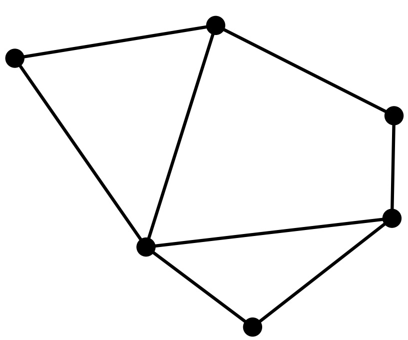
G^t

X^t

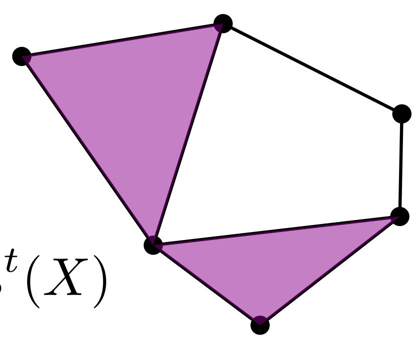


Definition: The *Rips complex of X at time t* is the clique complex of the graph G_t. We denote it $\mathrm{Rips}^t(X)$.

G^t



$\mathrm{Rips}^t(X)$

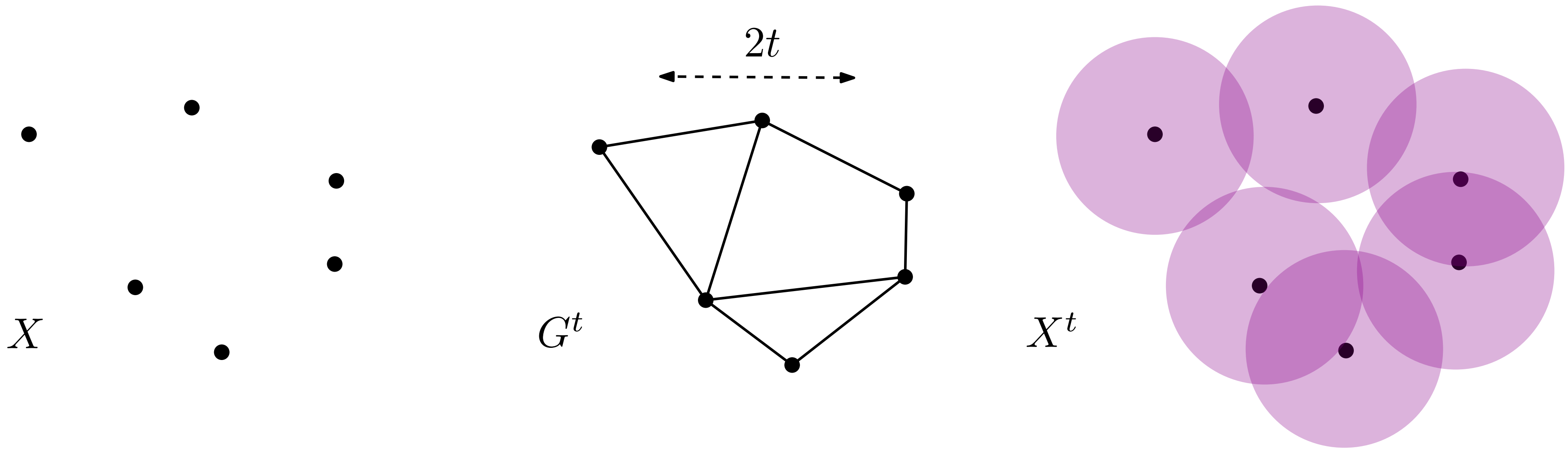


Rips complex

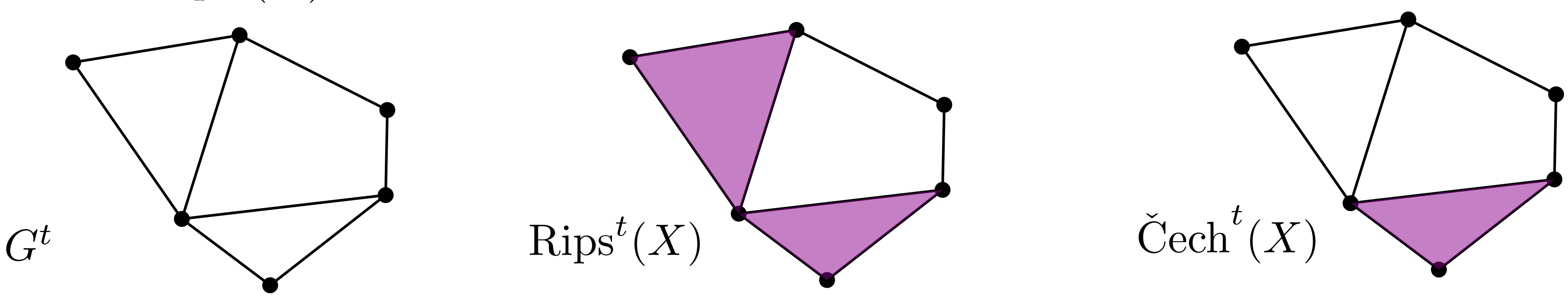
Let $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^n$ and $t \geq 0$.

Consider the graph G^t whose vertex set is $\{1, \ldots, N\}$, and whose edges are the pairs (i, j) such that $\|x_i - x_j\| \leq 2t$.

Alternatively, G^t can be seen as the 1-skeleton of the Čech complex $\check{\mathrm{C}}\mathrm{ech}^t(X)$.



Definition: The *Rips complex of X at time t* is the clique complex of the graph G_t. We denote it $\mathrm{Rips}^t(X)$.

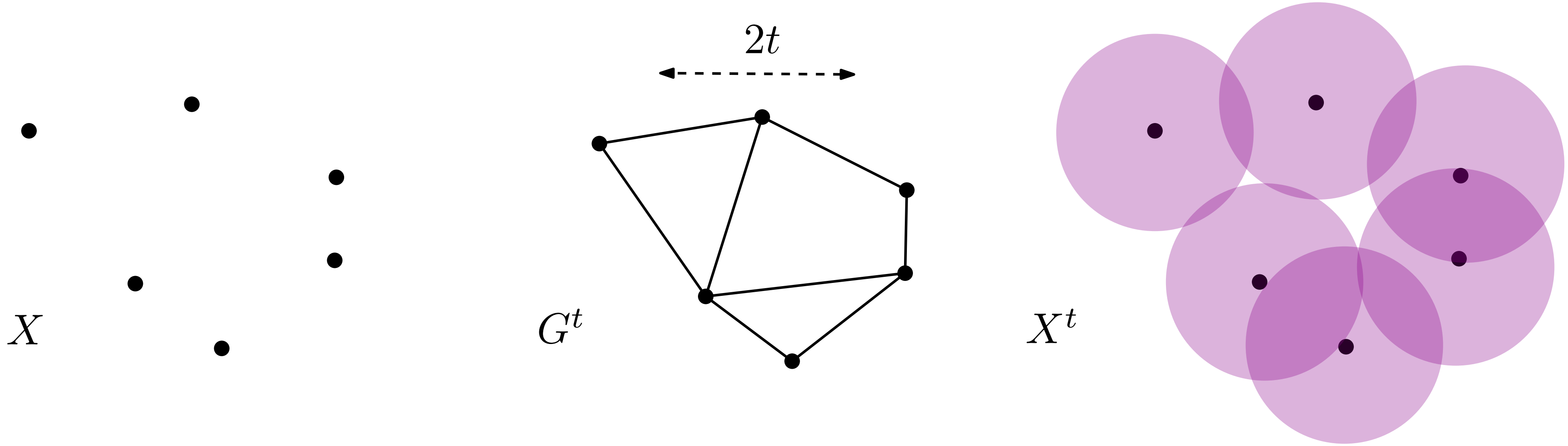


Rips complex

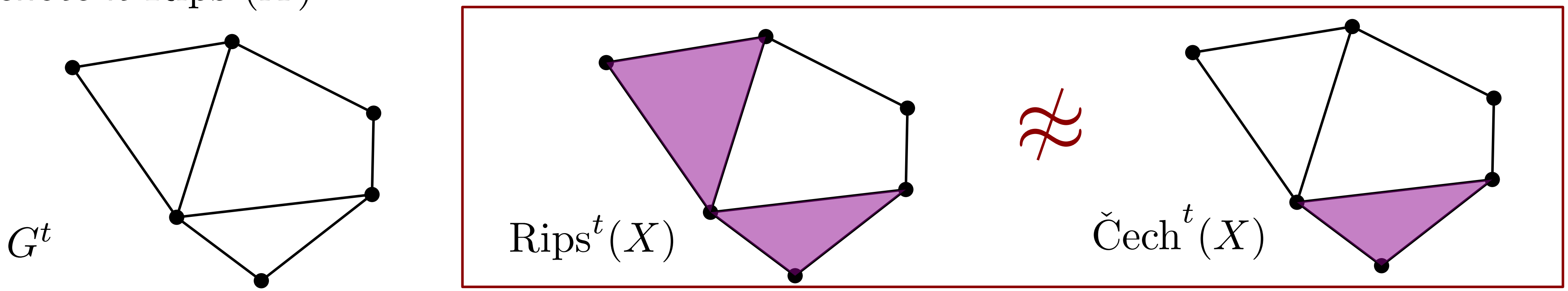
Let $X = \{x_1, \ldots, x_N\} \subset \mathbb{R}^n$ and $t \geq 0$.

Consider the graph G^t whose vertex set is $\{1, \ldots, N\}$, and whose edges are the pairs (i,j) such that $\|x_i - x_j\| \leq 2t$.

Alternatively, G^t can be seen as the 1-skeleton of the Čech complex $\check{\mathrm{C}}\mathrm{ech}^t(X)$.



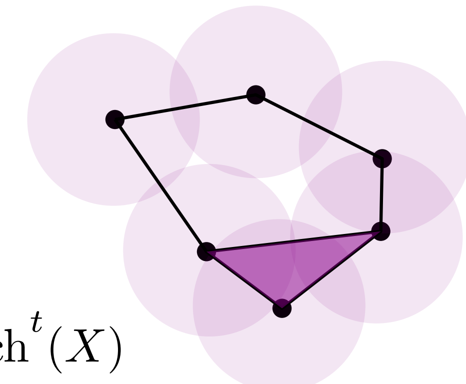
Definition: The *Rips complex of X at time t* is the clique complex of the graph G_t. We denote it $\mathrm{Rips}^t(X)$.



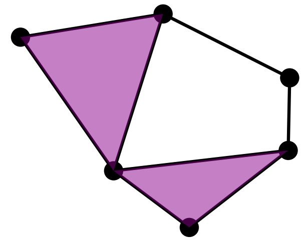
Rips complex

Proposition: For every $t \geq 0$, we have

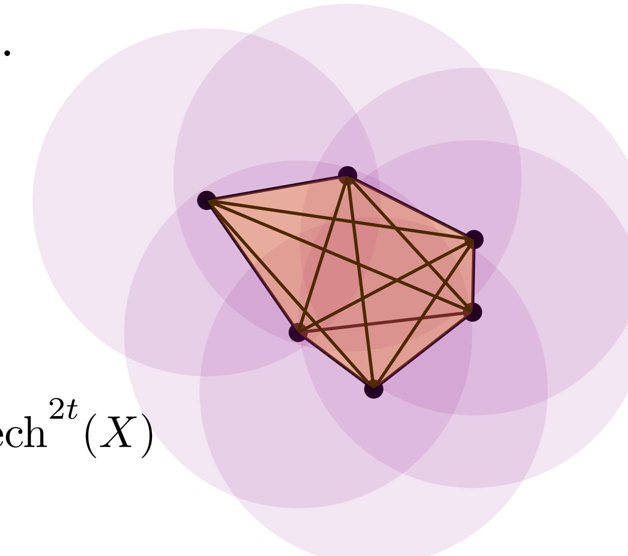
$$\check{\mathrm{C}}\mathrm{ech}^t(X) \subset \mathrm{Rips}^t(X) \subset \check{\mathrm{C}}\mathrm{ech}^{2t}(X).$$



$\check{\mathrm{C}}\mathrm{ech}^t(X)$



$\mathrm{Rips}^t(X)$

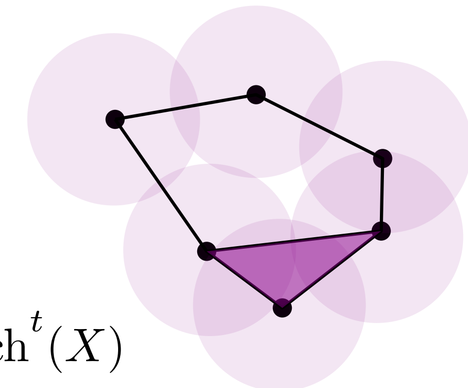


$\check{\mathrm{C}}\mathrm{ech}^{2t}(X)$

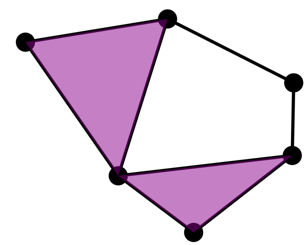
Rips complex

Proposition: For every $t \geq 0$, we have

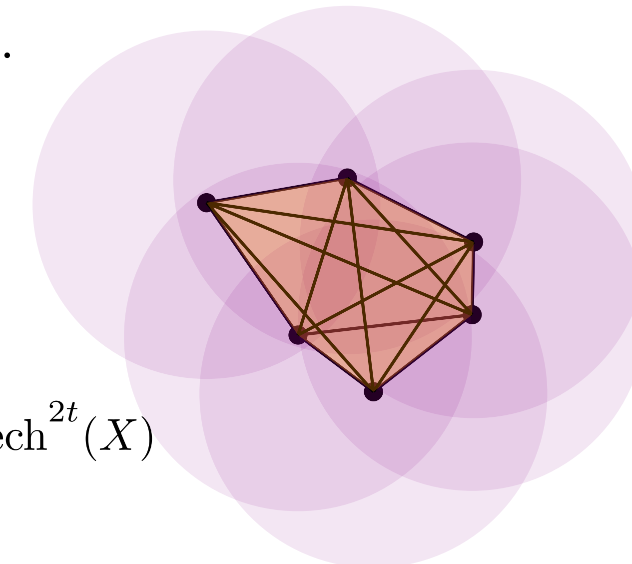
$$\check{\mathrm{C}}\mathrm{ech}^t(X) \subset \mathrm{Rips}^t(X) \subset \check{\mathrm{C}}\mathrm{ech}^{2t}(X).$$



$\check{\mathrm{C}}\mathrm{ech}^t(X)$



$\mathrm{Rips}^t(X)$



$\check{\mathrm{C}}\mathrm{ech}^{2t}(X)$

Proof: Let $t \geq 0$. The first inclusion follows from the fact that $\mathrm{Rips}^t(X)$ is the clique complex of $\check{\mathrm{C}}\mathrm{ech}^t(X)$.

To prove the second one, choose a simplex $\sigma \in \mathrm{Rips}^t(X)$. Let us prove that $\omega \in \check{\mathrm{C}}\mathrm{ech}^{2t}(X)$.

Let $x \in \sigma$ be any vertex. Note that $\forall y \in \sigma$, we have $\|x - y\| \leq 2t$ by definition of the Rips complex. Hence

$$x \in \bigcap_{y \in \sigma} \overline{\mathcal{B}}\left(y, 2t\right).$$

The intersection being non-empty, we deduce $\sigma \in \check{\mathrm{C}}\mathrm{ech}^{2t}(X)$.

Conclusion

We considered the problem of topological inference, and studied the solution by thickenings.

We've seen that a nice thickening exists, and that its homology can be computed via the Čech complex.

For computational reasons, we introduced the Rips complex.

Homework: Exercise 37

Facultative: Exercises 39, 40, 41

Conclusion

We considered the problem of topological inference, and studied the solution by thickenings.

We've seen that a nice thickening exists, and that its homology can be computed via the Čech complex.

For computational reasons, we introduced the Rips complex.

Theorem (Frédéric Chazal, David Cohen-Steiner, and André Lieutier, 2009):
Let X and $\mathcal{M}$ be subsets of $\mathbb{R}^n$. Suppose that $\mathcal{M}$ has positive reach, and that $\mathrm{d_H}(X, \mathcal{M}) \leq \frac{1}{17}\mathrm{reach}(\mathcal{M})$.
Then X^t and $\mathcal{M}$ are homotopic equivalent, provided that

$$t \in \left[4\mathrm{d_H}(X, \mathcal{M}), \mathrm{reach}(\mathcal{M}) - 3\mathrm{d_H}(X, \mathcal{M})\right).$$

Theorem (Partha Niyogi, Stephen Smale, and Shmuel Weinberger, 2008):
Let X and $\mathcal{M}$ be subsets of $\mathbb{R}^n$, with $\mathcal{M}$ a submanifold, and X a finite subset of $\mathcal{M}$.
Suppose that $\mathcal{M}$ has positive reach.
Then X^t and $\mathcal{M}$ are homotopic equivalent, provided that

$$t \in \left[2\mathrm{d_H}(X, \mathcal{M}), \sqrt{\frac{3}{5}}\mathrm{reach}(\mathcal{M})\right).$$

Conclusion

We considered the problem of topological inference, and studied the solution by thickenings.

We've seen that a nice thickening exists, and that its homology can be computed via the Čech complex.

For computational reasons, we introduced the Rips complex.

Theorem (Frédéric Chazal, David Cohen-Steiner, and André Lieutier, 2009):
Let X and $\mathcal{M}$ be subsets of $\mathbb{R}^n$. Suppose that $\mathcal{M}$ has positive reach, and that $\mathrm{d_H}(X, \mathcal{M}) \leq \frac{1}{17}\mathrm{reach}(\mathcal{M})$.
Then X^t and $\mathcal{M}$ are homotopic equivalent, provided that

unknown in practice

$$t \in \left[4\mathrm{d_H}(X, \mathcal{M}), \mathrm{reach}(\mathcal{M}) - 3\mathrm{d_H}(X, \mathcal{M})\right).$$

Theorem (Partha Niyogi, Stephen Smale, and Shmuel Weinberger, 2008):
Let X and $\mathcal{M}$ be subsets of $\mathbb{R}^n$, with $\mathcal{M}$ a submanifold, and X a finite subset of $\mathcal{M}$.
Suppose that $\mathcal{M}$ has positive reach.
Then X^t and $\mathcal{M}$ are homotopic equivalent, provided that

$$t \in \left[2\mathrm{d_H}(X, \mathcal{M}), \sqrt{\frac{3}{5}}\mathrm{reach}(\mathcal{M})\right).$$

Conclusion

We considered the problem of topological inference, and studied the solution by thickenings.

We've seen that a nice thickening exists, and that its homology can be computed via the Čech complex.

For computational reasons, we introduced the Rips complex.

Homework: Exercise 37

Facultative: Exercises 39, 40, 41

Merci !