

CCMN — 18/06/2024

Análise Topológica de Dados e suas aplicações, lição II/II

<https://raphaeltinarrage.github.io>

I - Comparar espaços topológicos

- 1 - Equivalência de homeomorfia
- 2 - Equivalência de homotopia

II - Invariante topológico

- 1 - Número de componentes conexas
- 2 - Característica de Euler
- 3 - Números de Betti
- 4 - Homologia (simplicial sobre $\mathbb{Z}/2\mathbb{Z}$)

III - Inferência topológica

- 1 - Estimação do parâmetro t
- 2 - Nervos

IV - Homologia persistente

- 1 - Módulos de persistência
- 2 - Decomposição
- 3 - Estabilidade

V - Aplicações

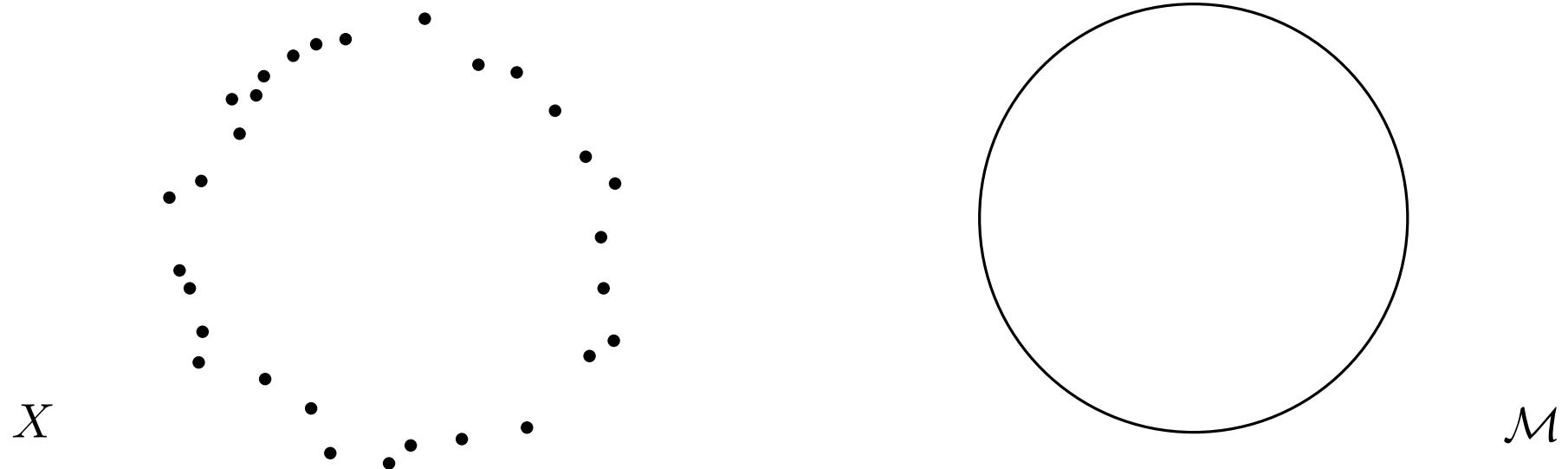
O problema da inferência homológica

3/38 (1/12)

Seja $\mathcal{M} \subset \mathbb{R}^n$ um subconjunto.

Suponha que tenhamos uma amostra finita $X \subset \mathcal{M}$.

Problema: Estimar os grupos de homologia de \mathcal{M} a partir de X .



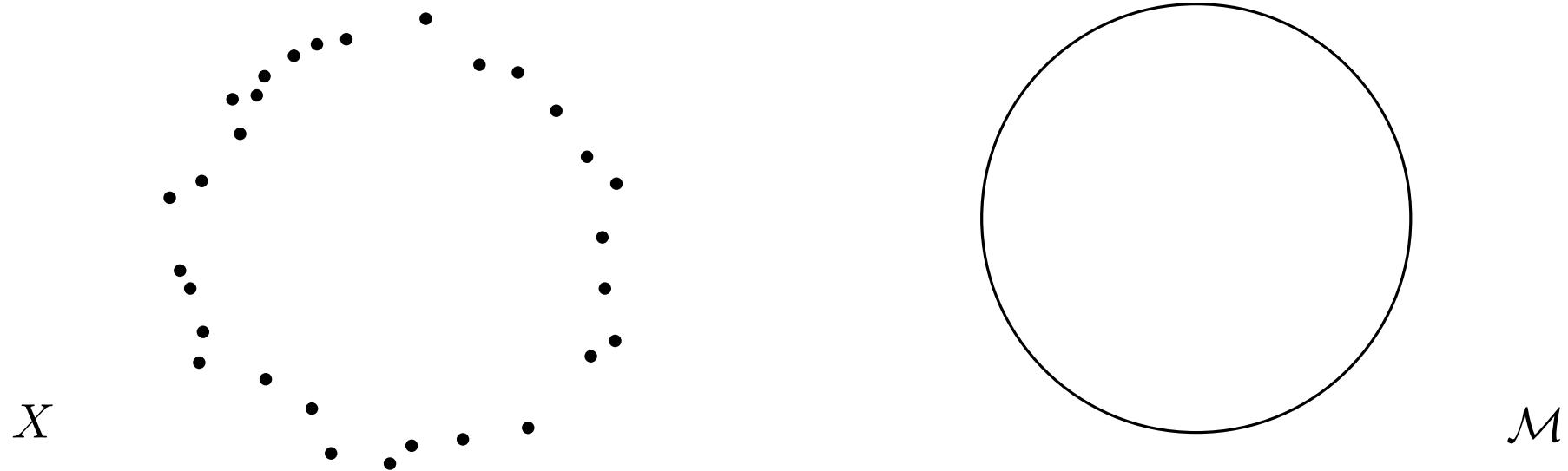
O problema da inferência homológica

3/38 (2/12)

Seja $\mathcal{M} \subset \mathbb{R}^n$ um subconjunto.

Suponha que tenhamos uma amostra finita $X \subset \mathcal{M}$.

Problema: Estimar os grupos de homologia de \mathcal{M} a partir de X .



Não podemos usar X diretamente.

Sua homologia é decepcionante:

$$\beta_0(X) = 30 \quad \text{e} \quad \beta_i(X) = 0 \quad \text{para } i \geq 1$$

número de componentes conexas
= número de pontos de X

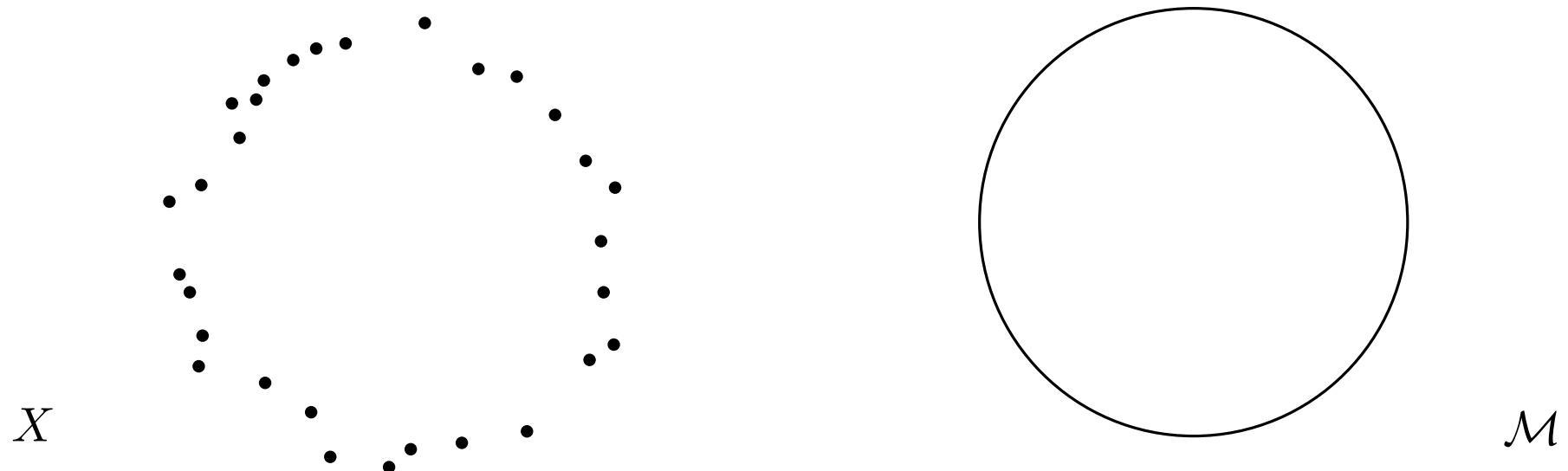
O problema da inferência homológica

3/38 (3/12)

Seja $\mathcal{M} \subset \mathbb{R}^n$ um subconjunto.

Suponha que tenhamos uma amostra finita $X \subset \mathcal{M}$.

Problema: Estimar os grupos de homologia de \mathcal{M} a partir de X .



Não podemos usar X diretamente.

Idéia: Espessar X .

Definição: Por todo $t \geq 0$, o **t -espessamento** do conjunto X , denotado X^t , é o conjunto de pontos do espaço ambiente a uma distância máxima de t de X :

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

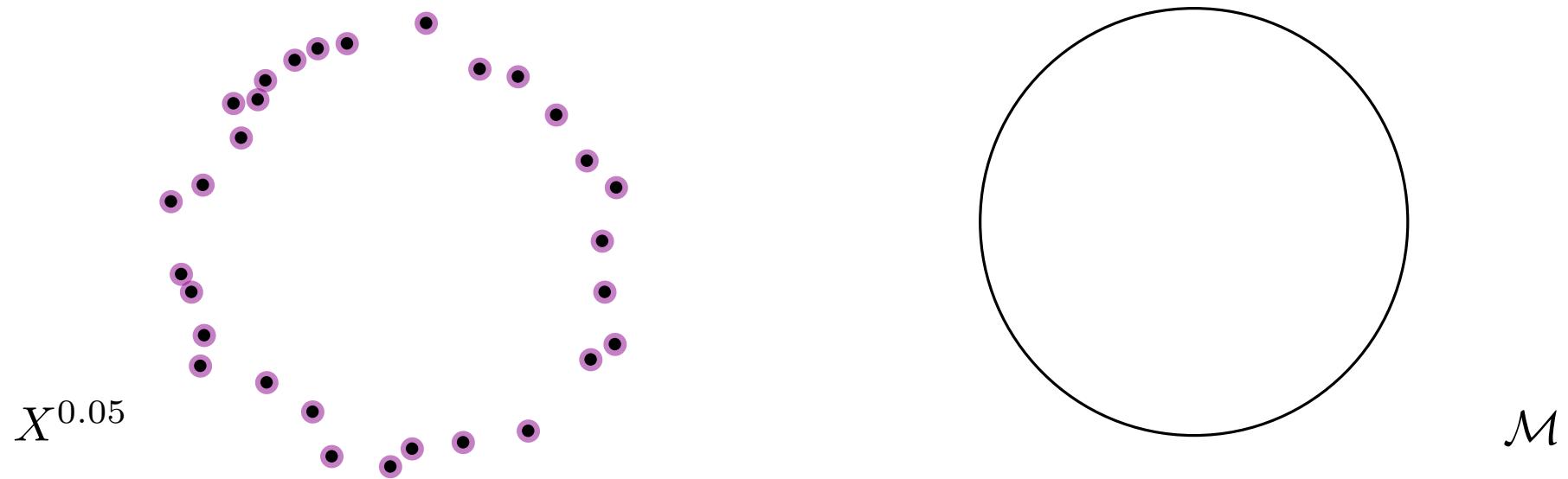
O problema da inferência homológica

3/38 (4/12)

Seja $\mathcal{M} \subset \mathbb{R}^n$ um subconjunto.

Suponha que tenhamos uma amostra finita $X \subset \mathcal{M}$.

Problema: Estimar os grupos de homologia de \mathcal{M} a partir de X .



Não podemos usar X diretamente.

Idéia: Espessar X .

Definição: Por todo $t \geq 0$, o **t -espessamento** do conjunto X , denotado X^t , é o conjunto de pontos do espaço ambiente a uma distância máxima de t de X :

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

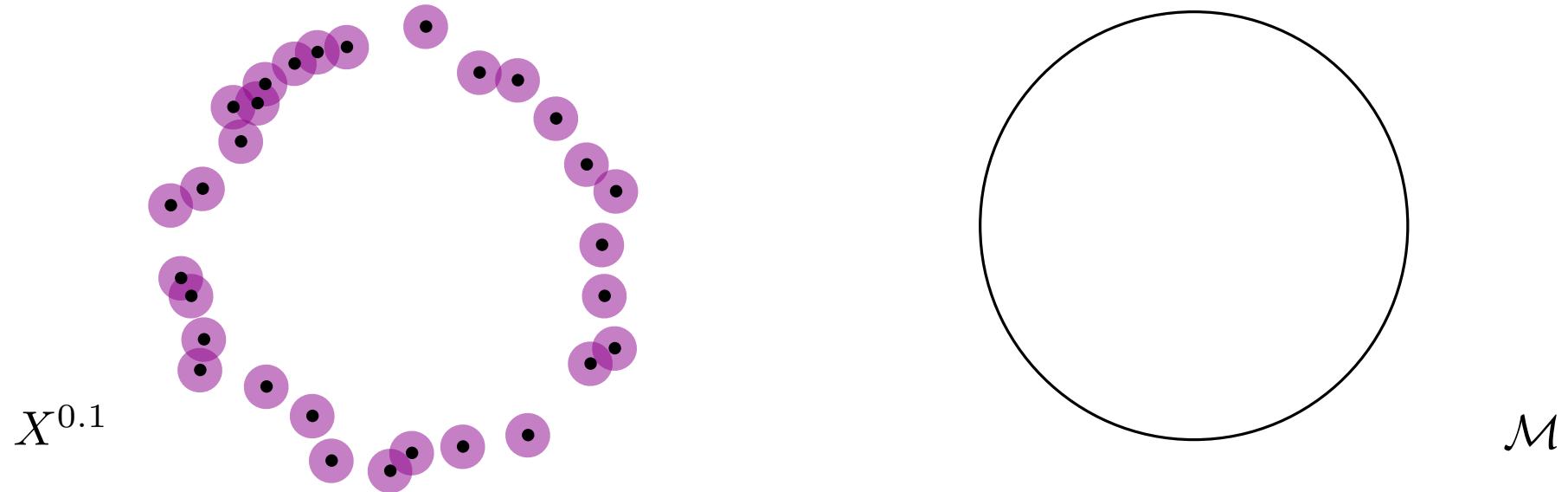
O problema da inferência homológica

3/38 (5/12)

Seja $\mathcal{M} \subset \mathbb{R}^n$ um subconjunto.

Suponha que tenhamos uma amostra finita $X \subset \mathcal{M}$.

Problema: Estimar os grupos de homologia de \mathcal{M} a partir de X .



Não podemos usar X diretamente.

Idéia: Espessar X .

Definição: Por todo $t \geq 0$, o **t -espessamento** do conjunto X , denotado X^t , é o conjunto de pontos do espaço ambiente a uma distância máxima de t de X :

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

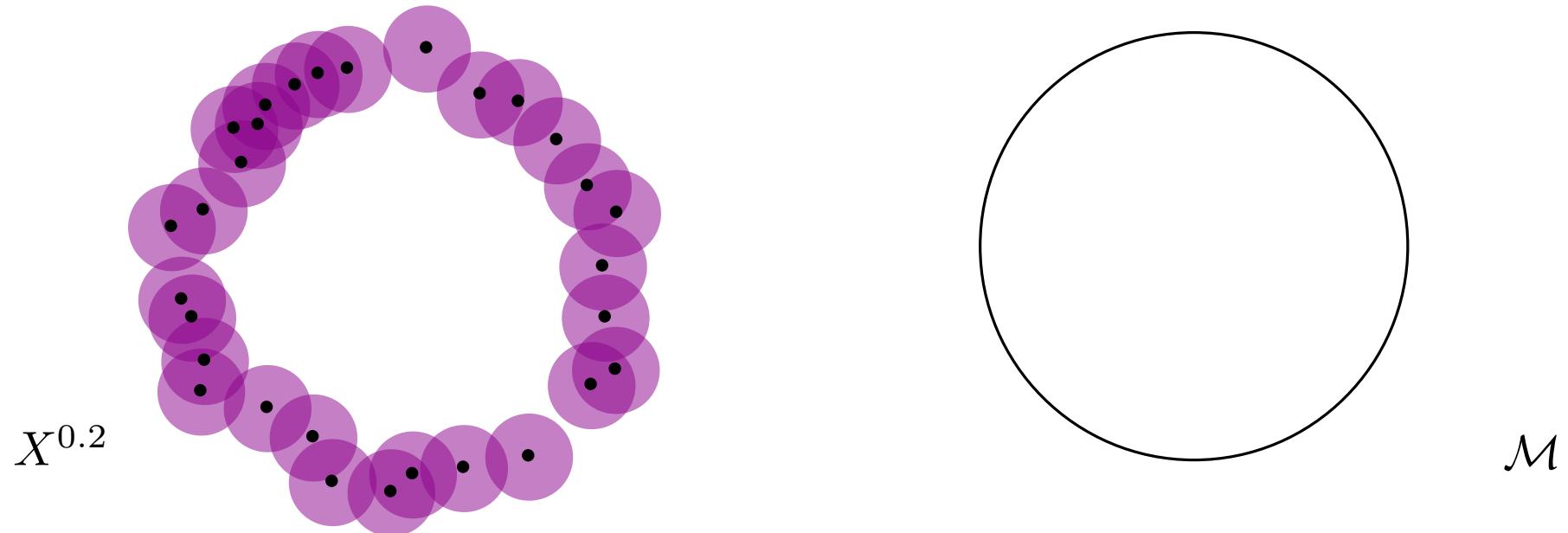
O problema da inferência homológica

3/38 (6/12)

Seja $\mathcal{M} \subset \mathbb{R}^n$ um subconjunto.

Suponha que tenhamos uma amostra finita $X \subset \mathcal{M}$.

Problema: Estimar os grupos de homologia de \mathcal{M} a partir de X .



Não podemos usar X diretamente.

Idéia: Espessar X .

Definição: Por todo $t \geq 0$, o **t -espessamento** do conjunto X , denotado X^t , é o conjunto de pontos do espaço ambiente a uma distância máxima de t de X :

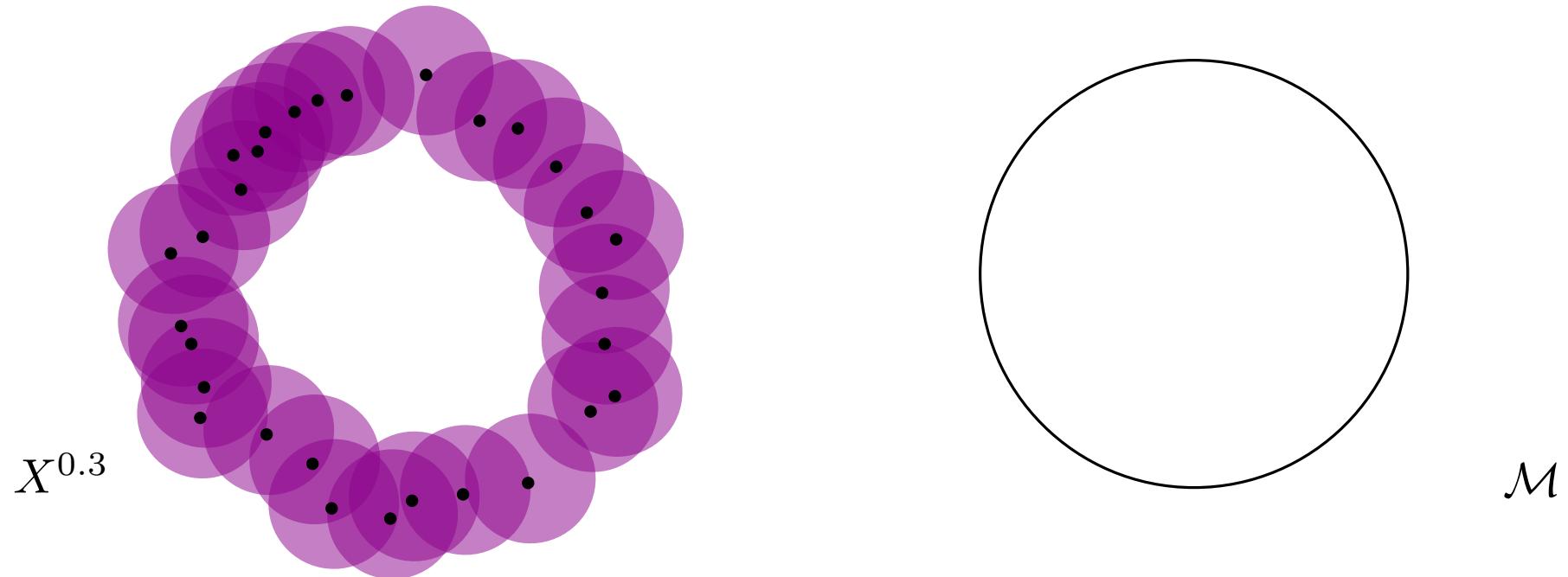
$$X^t = \{y \in \mathbb{R}^n \mid \exists x \in X, \|x - y\| \leq t\}.$$

O problema da inferência homológica 3/38 (7/12)

Seja $\mathcal{M} \subset \mathbb{R}^n$ um subconjunto.

Suponha que tenhamos uma amostra finita $X \subset \mathcal{M}$.

Problema: Estimar os grupos de homologia de \mathcal{M} a partir de X .



Não podemos usar X diretamente.

Idéia: Espessar X .

Definição: Por todo $t \geq 0$, o **t -espessamento** do conjunto X , denotado X^t , é o conjunto de pontos do espaço ambiente a uma distância máxima de t de X :

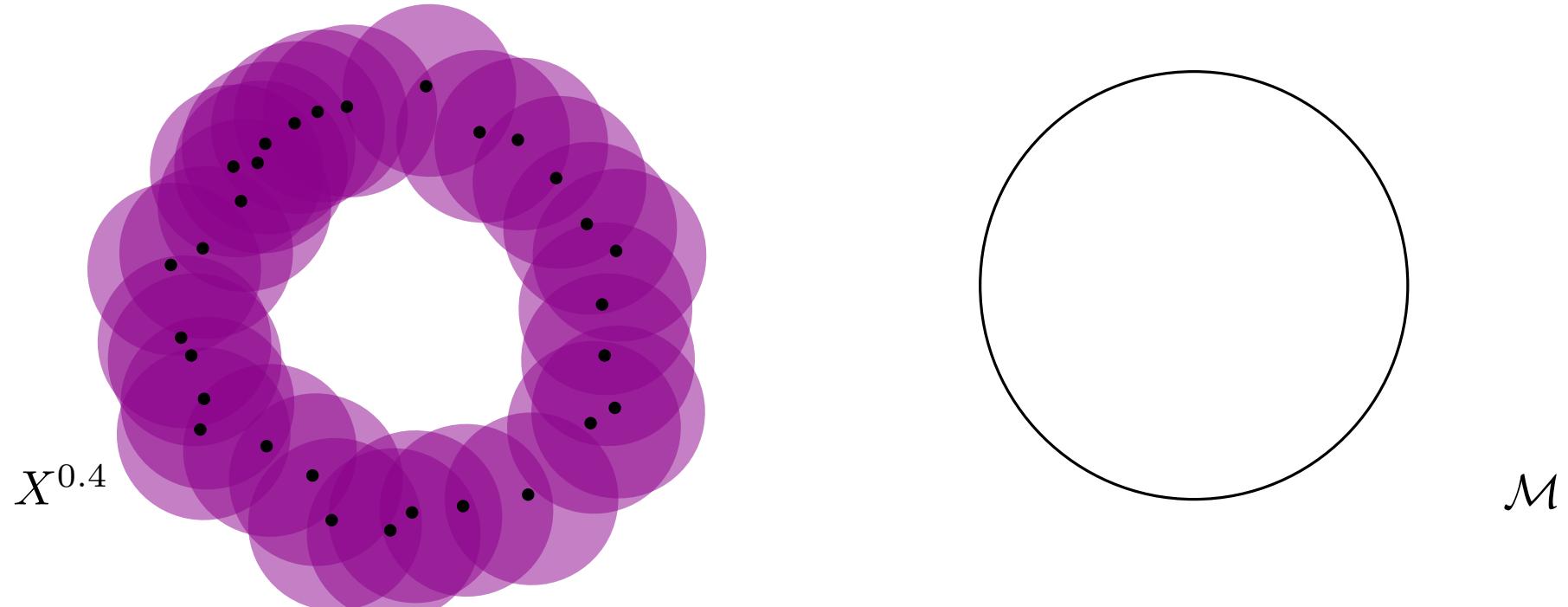
$$X^t = \{y \in \mathbb{R}^n \mid \exists x \in X, \|x - y\| \leq t\}.$$

O problema da inferência homológica 3/38 (8/12)

Seja $\mathcal{M} \subset \mathbb{R}^n$ um subconjunto.

Suponha que tenhamos uma amostra finita $X \subset \mathcal{M}$.

Problema: Estimar os grupos de homologia de \mathcal{M} a partir de X .



Não podemos usar X diretamente.

Idéia: Espessar X .

Definição: Por todo $t \geq 0$, o **t -espessamento** do conjunto X , denotado X^t , é o conjunto de pontos do espaço ambiente a uma distância máxima de t de X :

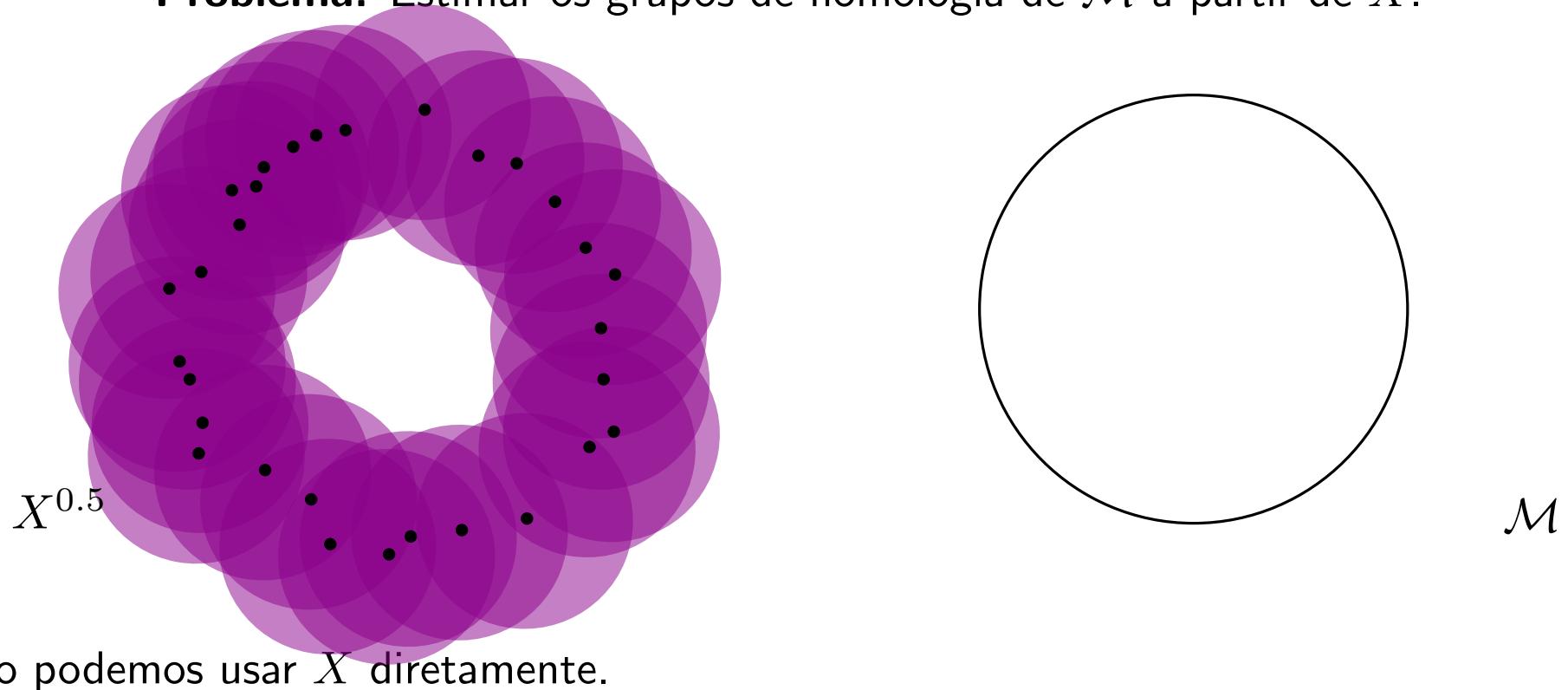
$$X^t = \{y \in \mathbb{R}^n \mid \exists x \in X, \|x - y\| \leq t\}.$$

O problema da inferência homológica 3/38 (9/12)

Seja $\mathcal{M} \subset \mathbb{R}^n$ um subconjunto.

Suponha que tenhamos uma amostra finita $X \subset \mathcal{M}$.

Problema: Estimar os grupos de homologia de \mathcal{M} a partir de X .



Idéia: Espessar X .

Definição: Por todo $t \geq 0$, o **t -espessamento** do conjunto X , denotado X^t , é o conjunto de pontos do espaço ambiente a uma distância máxima de t de X :

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

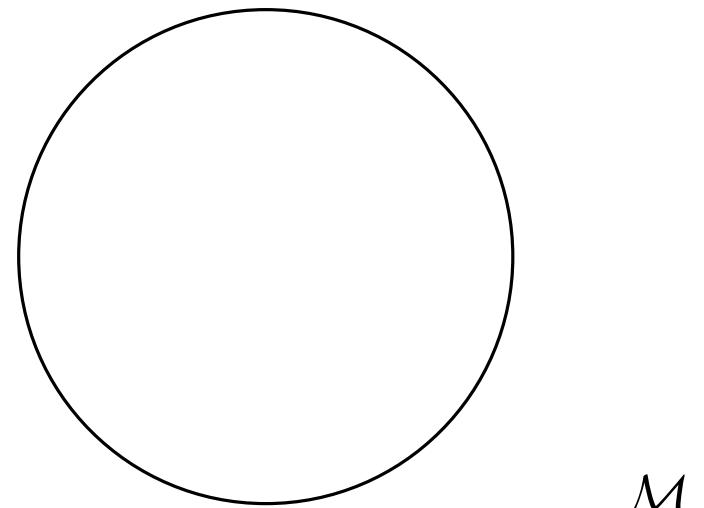
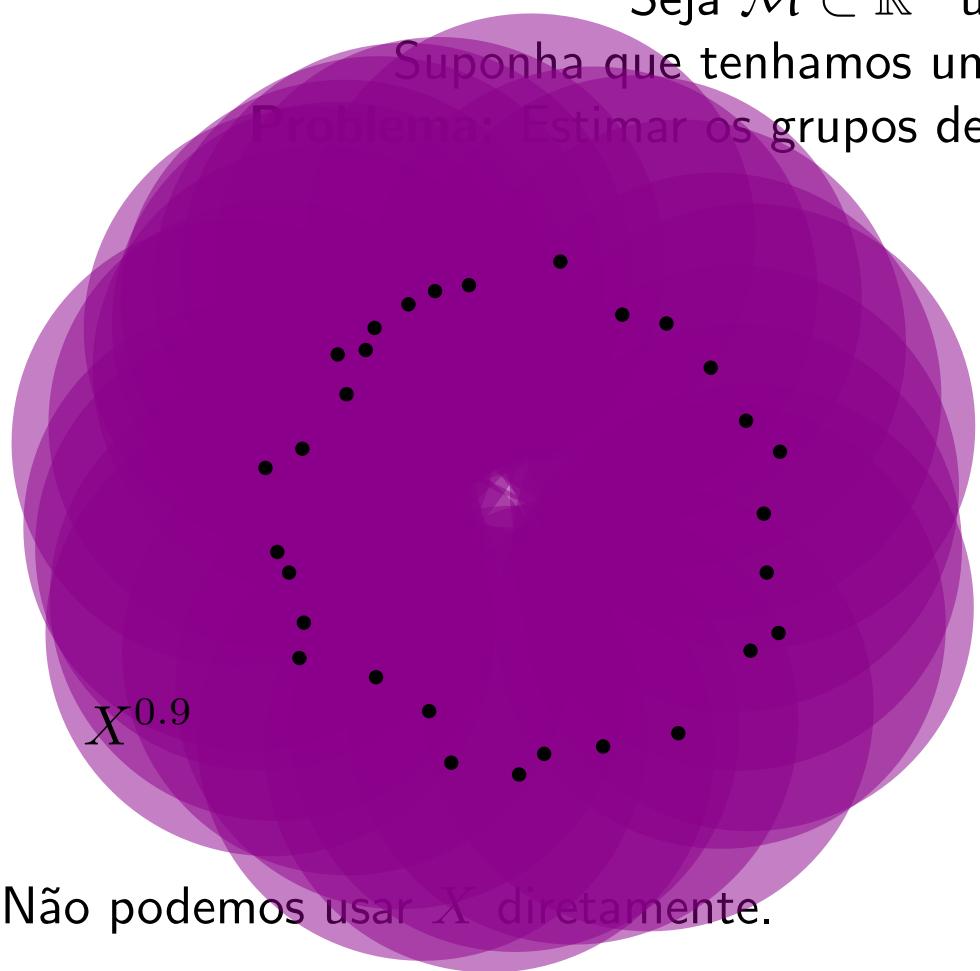
O problema da inferência homológica

3/38 (10/12)

Seja $\mathcal{M} \subset \mathbb{R}^n$ um subconjunto.

Suponha que tenhamos uma amostra finita $X \subset \mathcal{M}$.

Problema: Estimar os grupos de homologia de \mathcal{M} a partir de X .



Não podemos usar X diretamente.

Idéia: Espessar X .

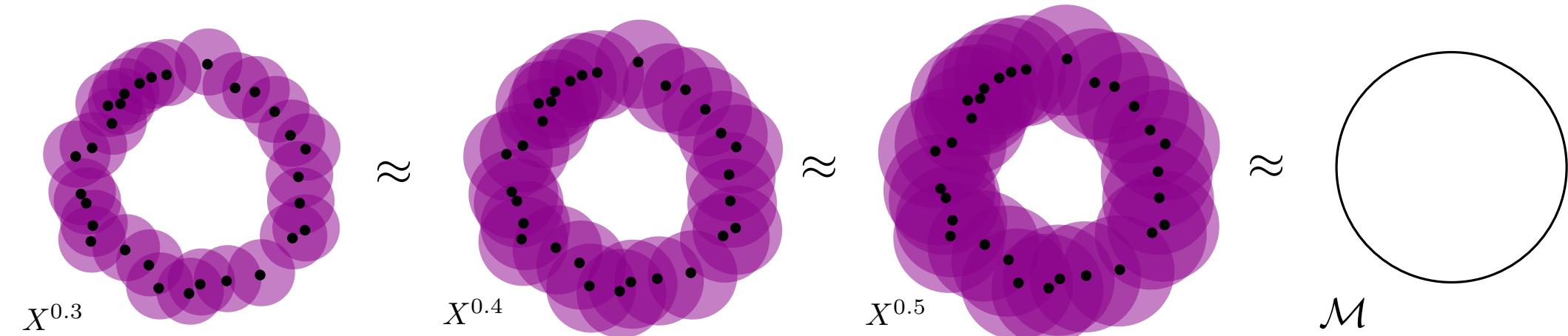
Definição: Por todo $t \geq 0$, o **t -espessamento** do conjunto X , denotado X^t , é o conjunto de pontos do espaço ambiente a uma distância máxima de t de X :

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

O problema da inferência homológica

3/38 (11/12)

Alguns espessamentos são homotopicamente equivalentes a \mathcal{M} .



Portanto, podemos encontrar a homologia de \mathcal{M} :

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

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

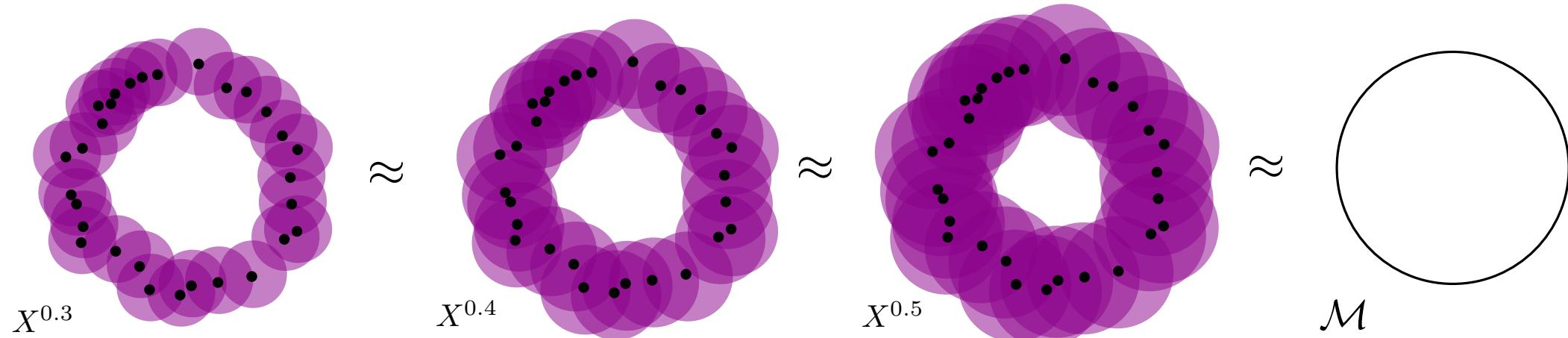
$$\beta_2(\mathcal{M}) = \beta_2(X^{0.3})$$

...

O problema da inferência homológica

3/38 (12/12)

Alguns espessamentos são homotopicamente equivalentes a \mathcal{M} .



Portanto, podemos encontrar a homologia de \mathcal{M} :

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

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

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

...

Questão 1: Como selecionar um t tal que $X^t \approx \mathcal{M}$?

Questão 2: Como calcular os grupos de homologia de X^t ?

I - Comparar espaços topológicos

- 1 - Equivalência de homeomorfia
- 2 - Equivalência de homotopia

II - Invariante topológico

- 1 - Número de componentes conexas
- 2 - Característica de Euler
- 3 - Números de Betti
- 4 - Homologia (simplicial sobre $\mathbb{Z}/2\mathbb{Z}$)

III - Inferência topológica

- 1 - Estimação do parâmetro t
- 2 - Nervos

IV - Homologia persistente

- 1 - Módulos de persistência
- 2 - Decomposição
- 3 - Estabilidade

V - Aplicações

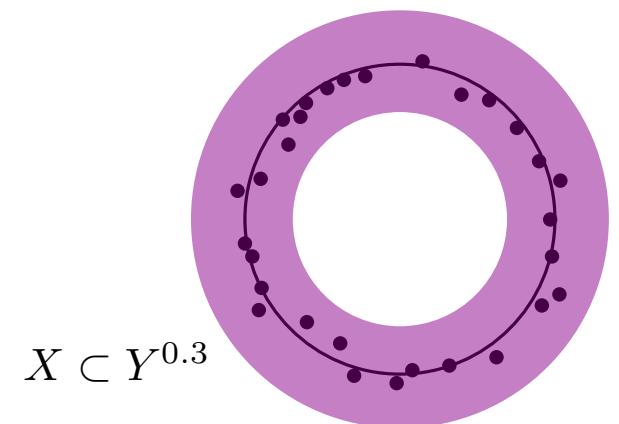
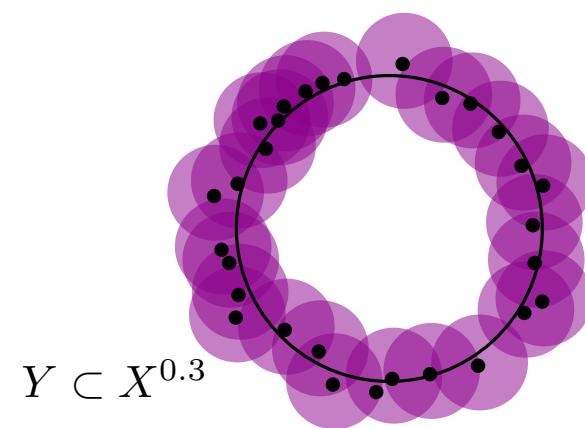
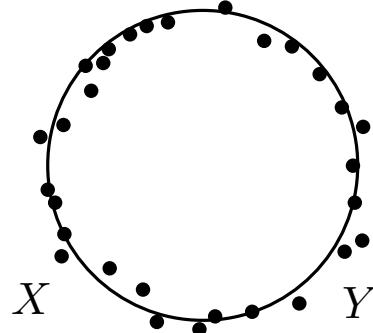
Distância de Hausdorff

5/38

Definição: Sejam $X, Y \subset \mathbb{R}^n$ dois subconjuntos compactos. A **Distância de Hausdorff** entre X e Y é

$$d_H(X, Y) = \max \left\{ \sup_{y \in Y} \inf_{x \in X} \|x - y\|, \sup_{x \in X} \inf_{y \in Y} \|x - y\| \right\}.$$

Propriedade: A distância de Hausdorff é igual a $\inf \{t \geq 0 \mid X \subset Y^t \text{ e } Y \subset X^t\}$.

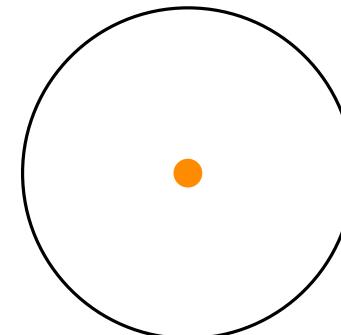


O **eixo mediano** (medial axis) de X é o subconjunto $\text{med}(X) \subset \mathbb{R}^n$ que consiste em pontos $y \in \mathbb{R}^n$ que possuem pelo menos duas projeções em X :

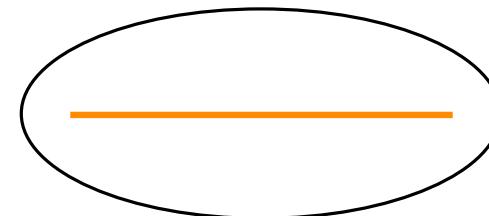
$$\text{med}(X) = \{y \in \mathbb{R}^n \mid \exists x, x' \in X, x \neq x', \|y - x\| = \|y - x'\| = \text{dist}(y, X)\}.$$

Exemplos:

O eixo mediano do círculo é a origem.



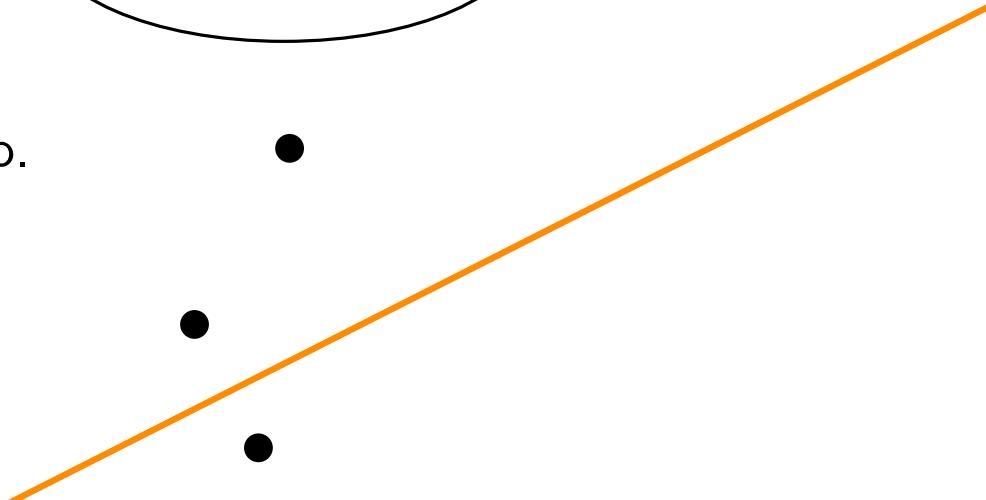
O eixo mediano de uma elipse é um segmento.



O eixo mediano de um ponto é o conjunto vazio.



O eixo mediano de dois pontos é a bissetriz.

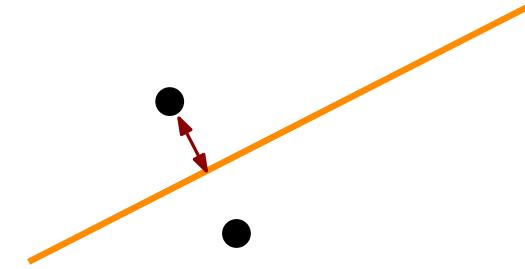
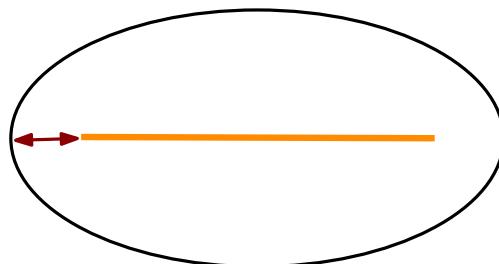
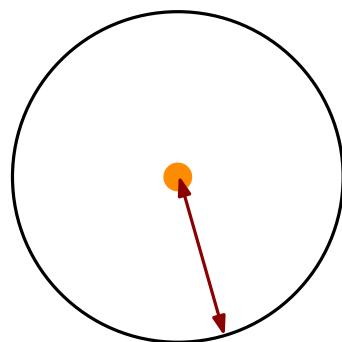


O **eixo mediano** (medial axis) de X é o subconjunto $\text{med}(X) \subset \mathbb{R}^n$ que consiste em pontos $y \in \mathbb{R}^n$ que possuem pelo menos duas projeções em X :

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

O **reach** (alcance) de X é

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



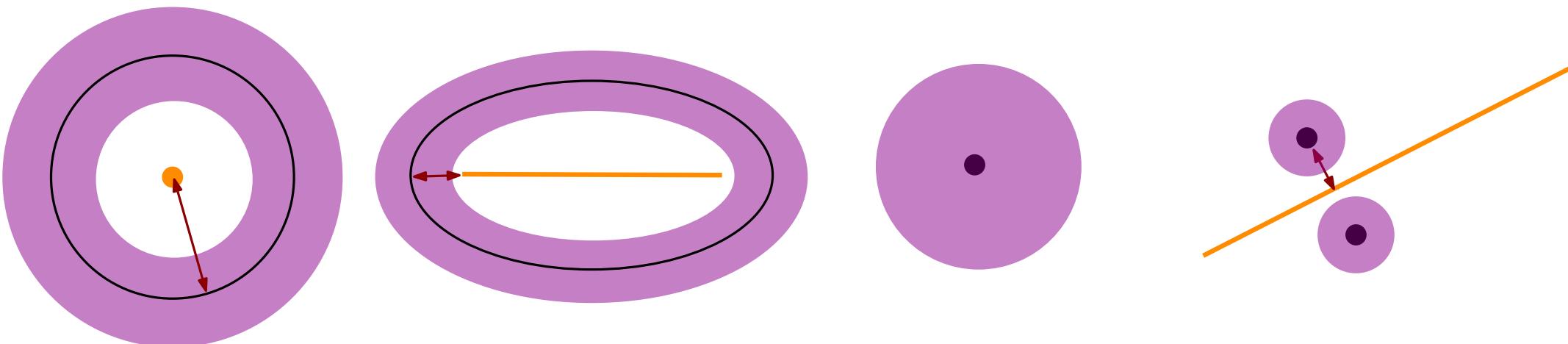
Proposição: Para todo $t \in [0, \text{reach}(X))$, os espaços X e X^t são homotopicamente equivalentes.

O **eixo mediano** (medial axis) de X é o subconjunto $\text{med}(X) \subset \mathbb{R}^n$ que consiste em pontos $y \in \mathbb{R}^n$ que possuem pelo menos duas projeções em X :

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

O **reach** (alcance) de X é

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



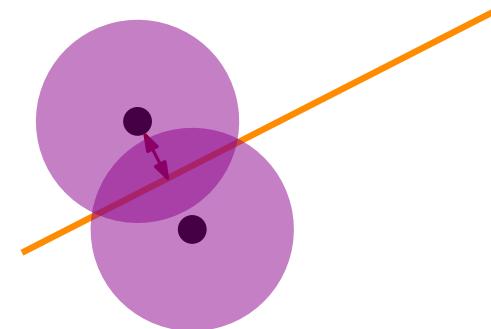
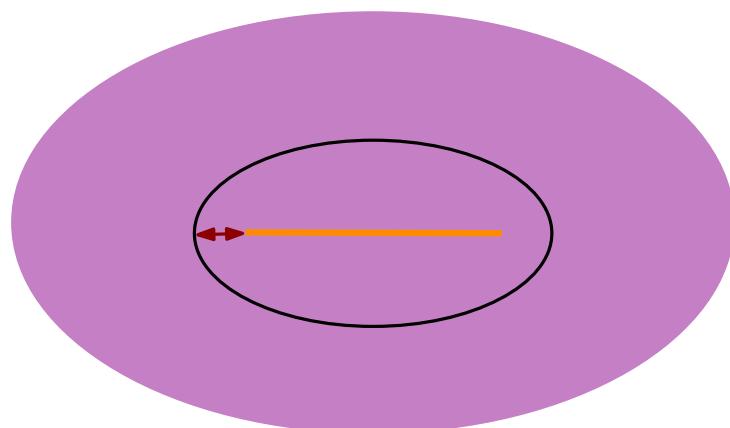
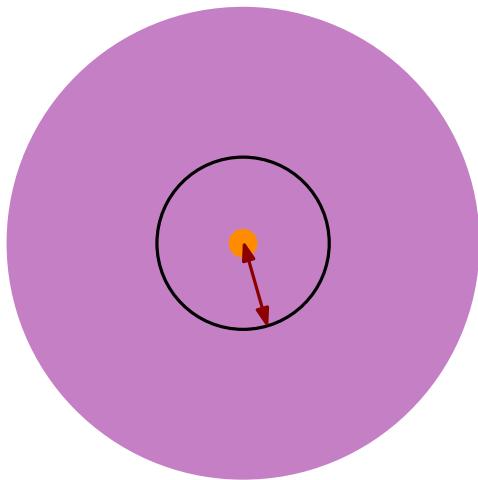
Proposição: Para todo $t \in [0, \text{reach}(X))$, os espaços X e X^t são homotopicamente equivalentes.

O **eixo mediano** (medial axis) de X é o subconjunto $\text{med}(X) \subset \mathbb{R}^n$ que consiste em pontos $y \in \mathbb{R}^n$ que possuem pelo menos duas projeções em X :

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

O **reach** (alcance) de X é

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



Proposição: Para todo $t \in [0, \text{reach}(X))$, os espaços X e X^t são homotopicamente equivalentes.

Se $t \geq \text{reach}(X)$, os conjuntos X e X^t podem não ser homotopicamente equivalentes.

Proposição: Para todo $t \in [0, \text{reach}(X))$, os espaços X e X^t são homotopicamente equivalentes.

Prova: Para todo $t \in [0, \text{reach}(X))$, o espessamento X^t retrai-se em X . Uma homotopia é dada pela seguinte função:

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

De fato, a projeção $\text{proj}(x, X)$ está bem definida (é única).

Questão 1: Como selecionar um t tal que $X^t \approx \mathcal{M}$?



Teorema (Frédéric Chazal, David Cohen-Steiner et André Lieutier, 2009):

Sejam X e \mathcal{M} dois subconjuntos de \mathbb{R}^n . Suponha que \mathcal{M} tenha um *reach* estritamente positivo, e que $d_H(X, \mathcal{M}) \leq \frac{1}{17} \text{reach}(\mathcal{M})$.

Então X^t e \mathcal{M} são homotopicamente equivalentes, desde que

$$t \in [4d_H(X, \mathcal{M}), \text{reach}(\mathcal{M}) - 3d_H(X, \mathcal{M})].$$

Questão 1: Como selecionar um t tal que $X^t \approx \mathcal{M}$?



Teorema (Frédéric Chazal, David Cohen-Steiner et André Lieutier, 2009):

Sejam X e \mathcal{M} dois subconjuntos de \mathbb{R}^n . Suponha que \mathcal{M} tenha um *reach* estritamente positivo, e que $d_H(X, \mathcal{M}) \leq \frac{1}{17} \text{reach}(\mathcal{M})$.

Então X^t e \mathcal{M} são homotopicamente equivalentes, desde que

$$t \in [4d_H(X, \mathcal{M}), \text{reach}(\mathcal{M}) - 3d_H(X, \mathcal{M})].$$

Teorema (Partha Niyogi, Stephen Smale et Shmuel Weinberger, 2008):

Sejam X e \mathcal{M} dois subconjuntos de \mathbb{R}^n , com \mathcal{M} uma subvariedade, e X um subconjunto finito de \mathcal{M} . Suponha que \mathcal{M} tenha um *reach* estritamente positivo. Então X^t e \mathcal{M} são homotopicamente equivalentes, desde que

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

I - Comparar espaços topológicos

- 1 - Equivalência de homeomorfia
- 2 - Equivalência de homotopia

II - Invariante topológico

- 1 - Número de componentes conexas
- 2 - Característica de Euler
- 3 - Números de Betti
- 4 - Homologia (simplicial sobre $\mathbb{Z}/2\mathbb{Z}$)

III - Inferência topológica

- 1 - Estimação do parâmetro t
- 2 - Nervos

IV - Homologia persistente

- 1 - Módulos de persistência
- 2 - Decomposição
- 3 - Estabilidade

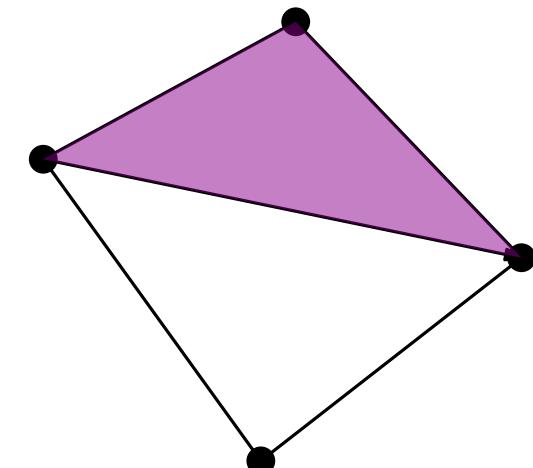
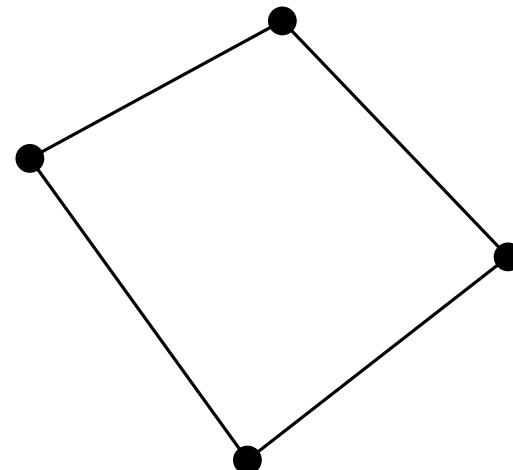
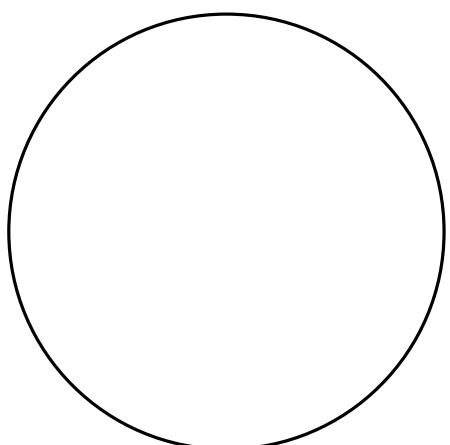
V - Aplicações

Questão 2: Como calcular os grupos de homologia de X^t ?

Precisamos de uma triangulação de X^t , isto é, um complexo simplicial de K homeomórfico a X^t .

Na verdade, precisaremos de algo mais fraco: um complexo simplicial de K homotopicamente equivalente a X^t .

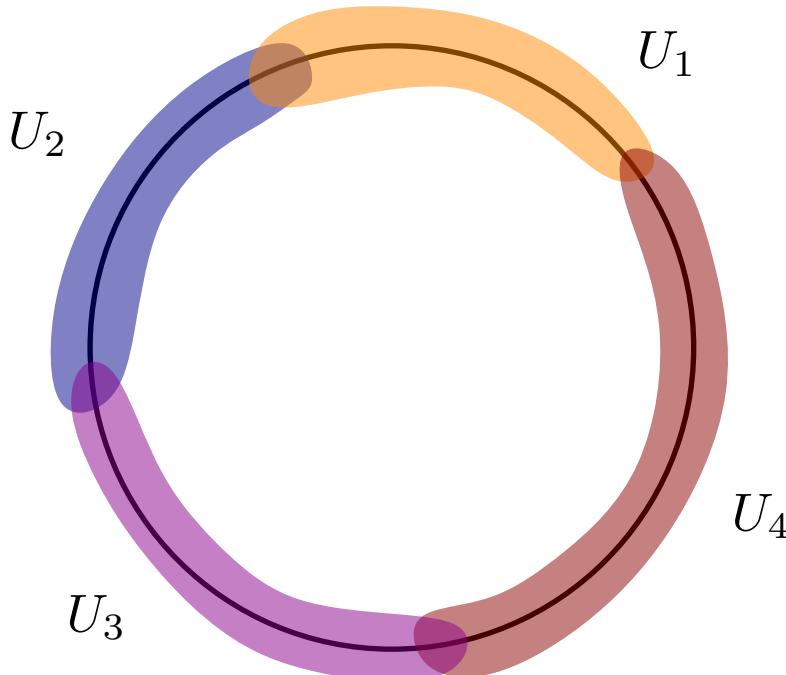
Em ambos casos, tem $\beta_i(X^t) = \beta_i(K)$ por todo $i \geq 0$.



Definição: Seja X um espaço topológico, e $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$ uma *cobertura* de X , isto é, uma coleção de subconjuntos $U_i \subset X$ tal que

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

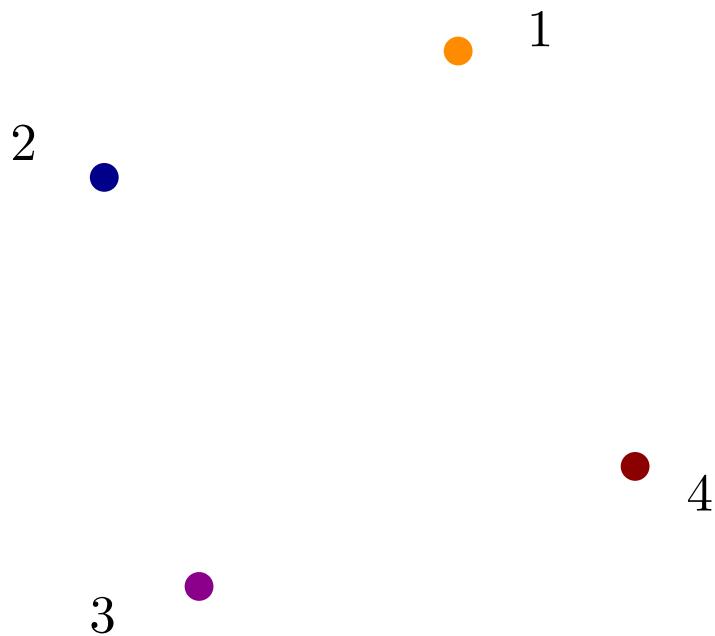
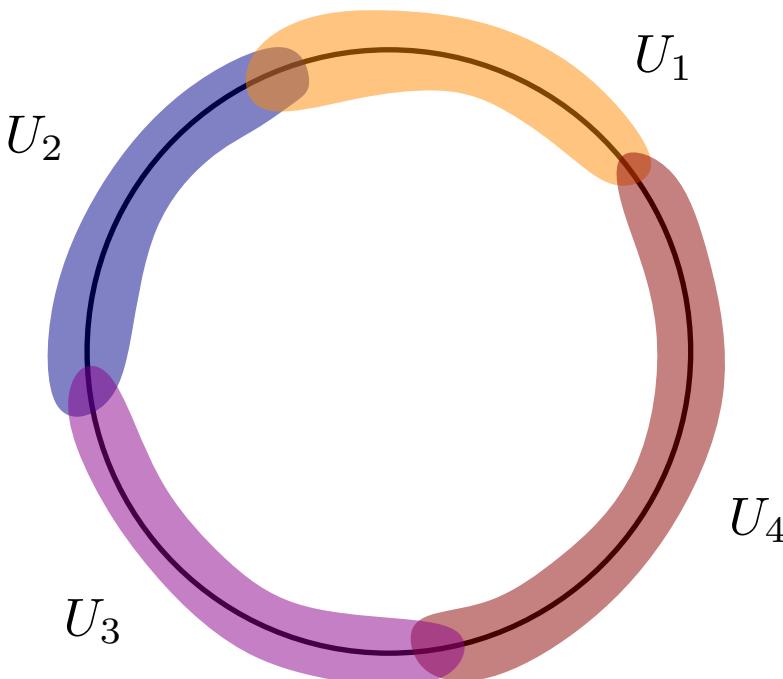
O **nervo** de \mathcal{U} é o complexo simplicial com conjunto de vértices $\{1, \dots, N\}$ e cujos m -simplexos são os subconjuntos $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$ tais que $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$. Ele é denotado $\mathcal{N}(\mathcal{U})$.



Definição: Seja X um espaço topológico, e $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$ uma *cobertura* de X , isto é, uma coleção de subconjuntos $U_i \subset X$ tal que

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

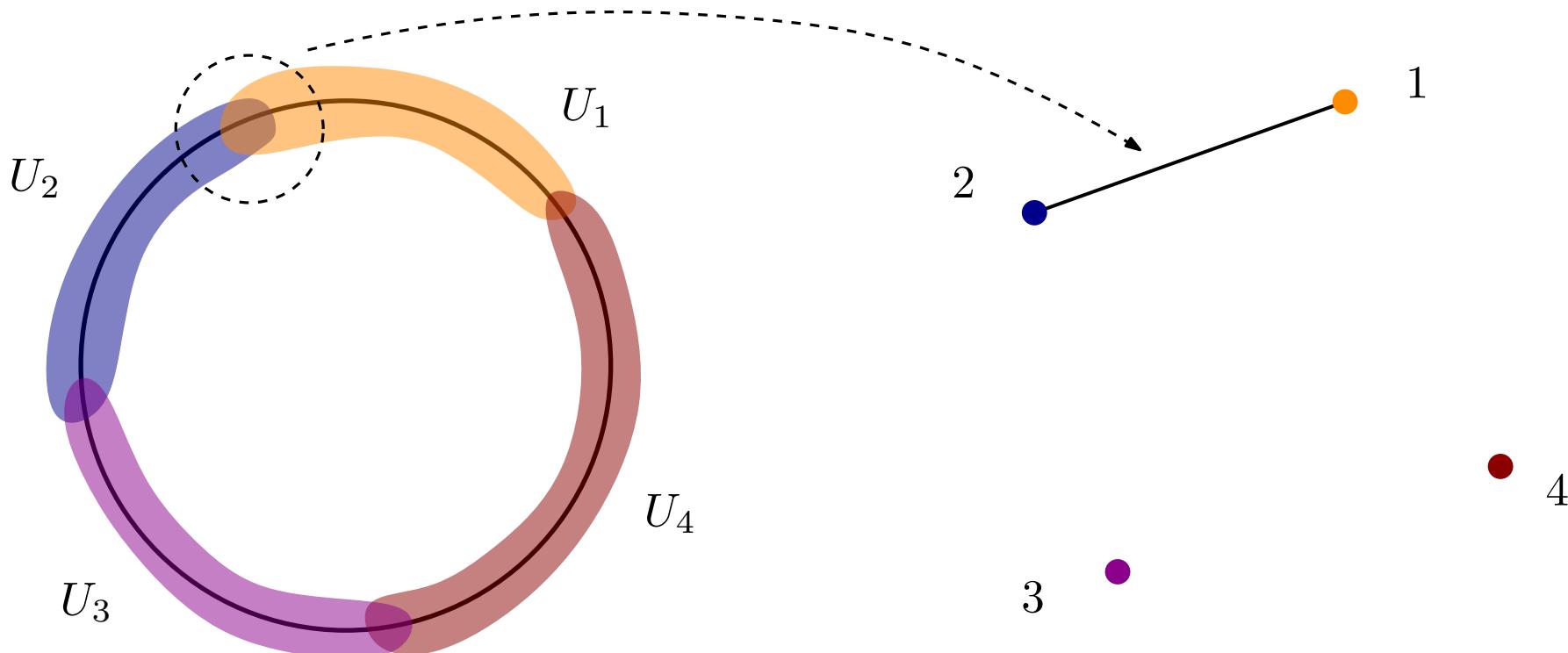
O **nervo** de \mathcal{U} é o complexo simplicial com conjunto de vértices $\{1, \dots, N\}$ e cujos m -simplexos são os subconjuntos $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$ tais que $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$. Ele é denotado $\mathcal{N}(\mathcal{U})$.



Definição: Seja X um espaço topológico, e $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$ uma *cobertura* de X , isto é, uma coleção de subconjuntos $U_i \subset X$ tal que

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

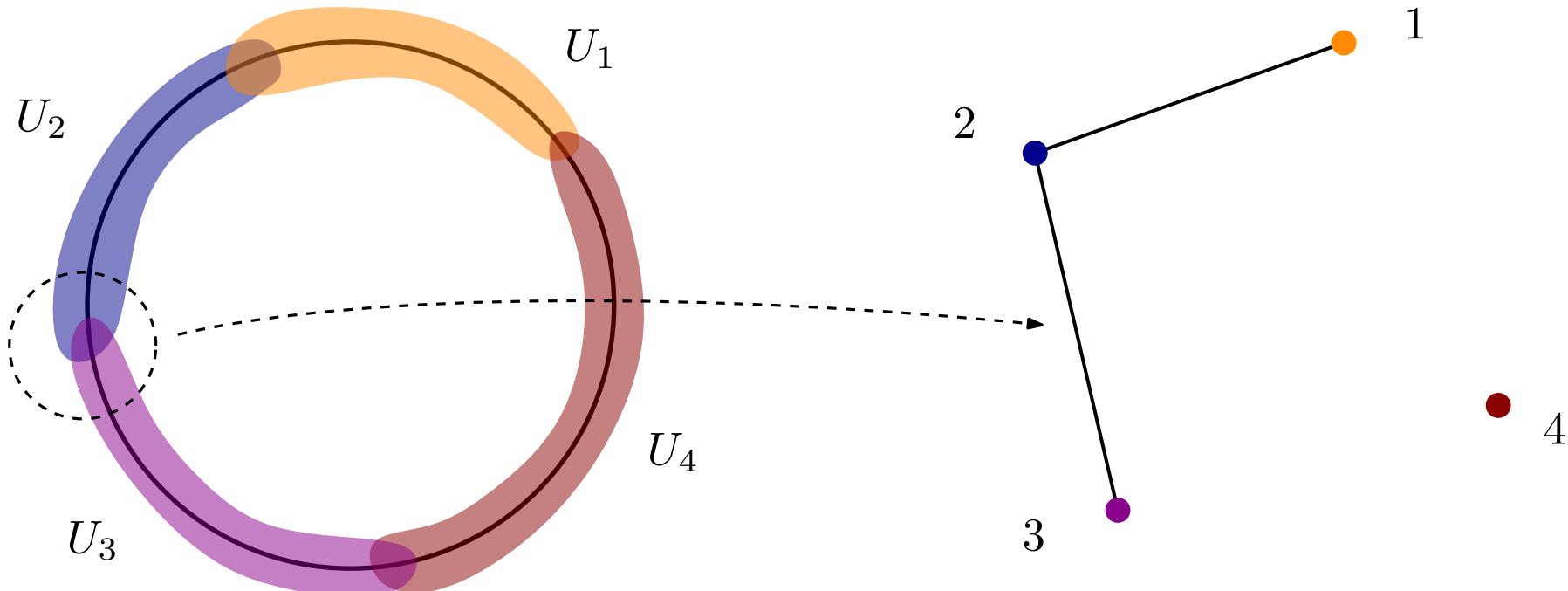
O **nervo** de \mathcal{U} é o complexo simplicial com conjunto de vértices $\{1, \dots, N\}$ e cujos m -simplexos são os subconjuntos $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$ tais que $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$. Ele é denotado $\mathcal{N}(\mathcal{U})$.



Definição: Seja X um espaço topológico, e $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$ uma *cobertura* de X , isto é, uma coleção de subconjuntos $U_i \subset X$ tal que

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

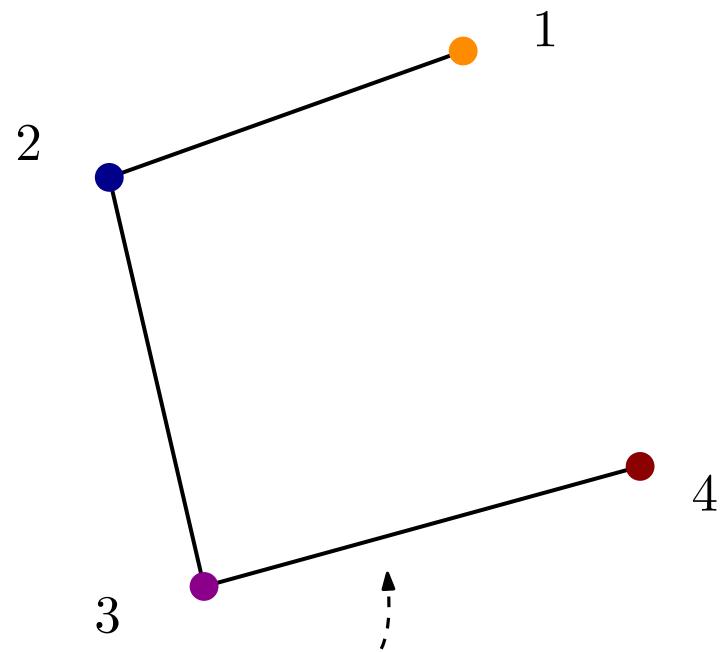
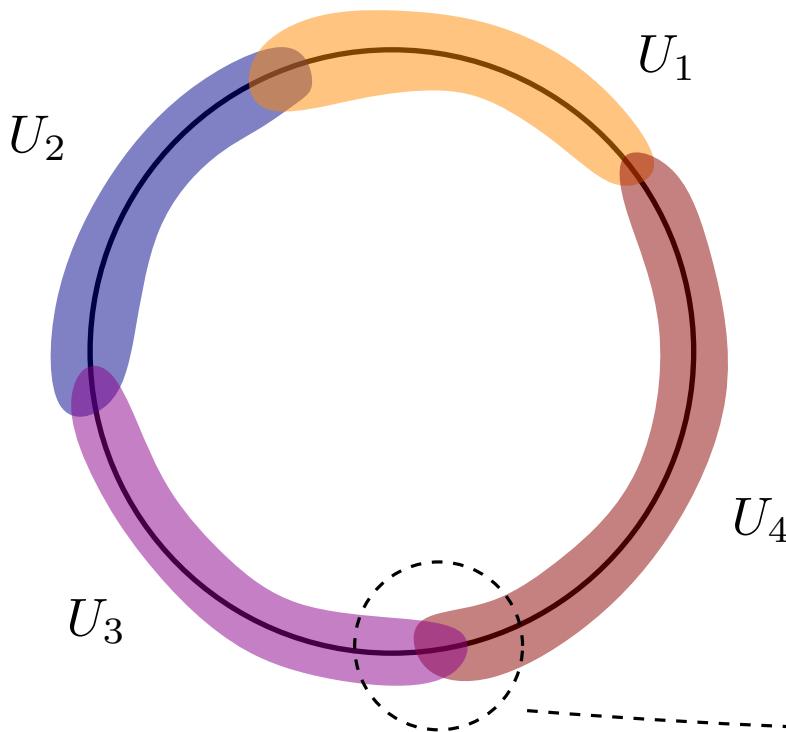
O **nervo** de \mathcal{U} é o complexo simplicial com conjunto de vértices $\{1, \dots, N\}$ e cujos m -simplexos são os subconjuntos $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$ tais que $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$. Ele é denotado $\mathcal{N}(\mathcal{U})$.



Definição: Seja X um espaço topológico, e $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$ uma *cobertura* de X , isto é, uma coleção de subconjuntos $U_i \subset X$ tal que

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

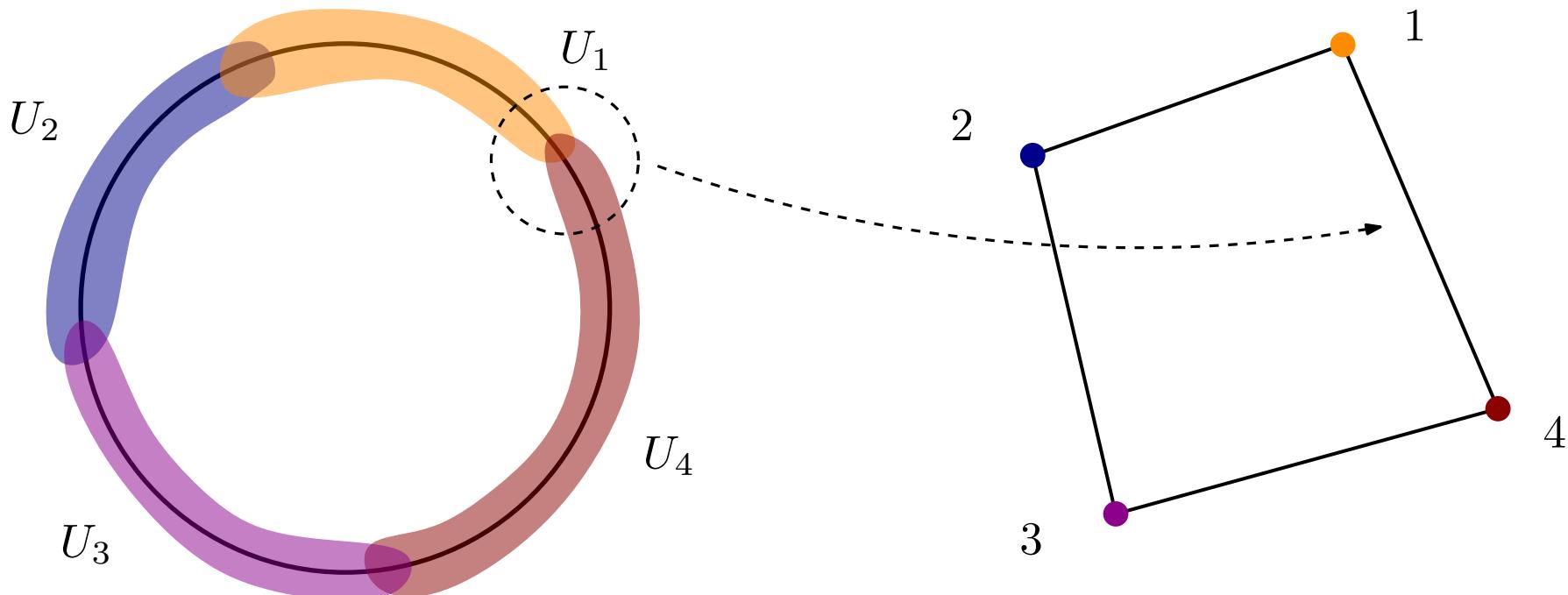
O **nervo** de \mathcal{U} é o complexo simplicial com conjunto de vértices $\{1, \dots, N\}$ e cujos m -simplexos são os subconjuntos $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$ tais que $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$. Ele é denotado $\mathcal{N}(\mathcal{U})$.



Definição: Seja X um espaço topológico, e $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$ uma *cobertura* de X , isto é, uma coleção de subconjuntos $U_i \subset X$ tal que

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

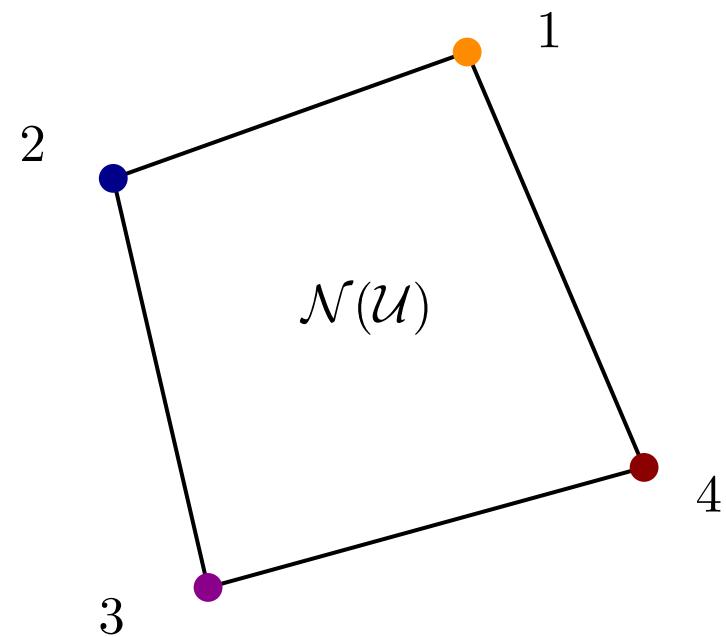
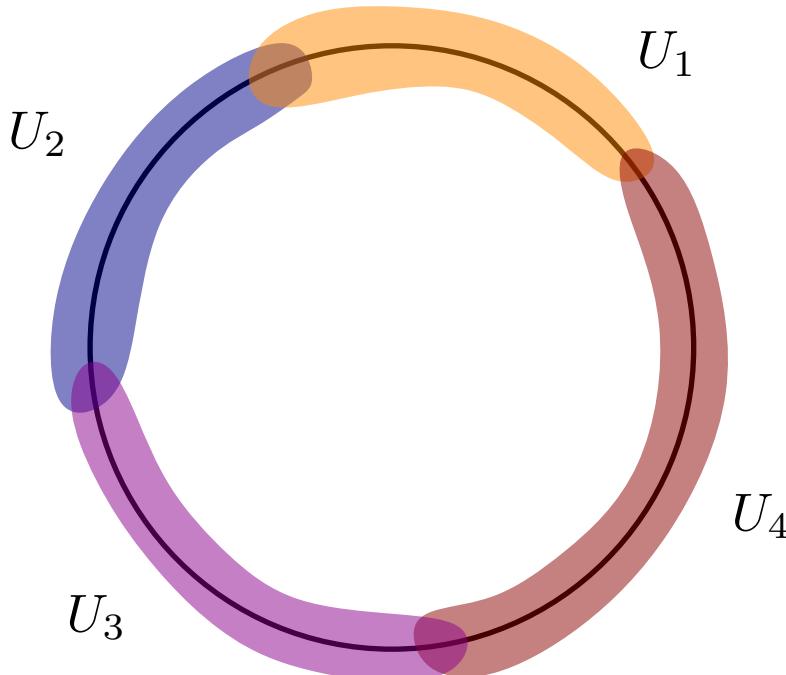
O **nervo** de \mathcal{U} é o complexo simplicial com conjunto de vértices $\{1, \dots, N\}$ e cujos m -simplexos são os subconjuntos $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$ tais que $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$. Ele é denotado $\mathcal{N}(\mathcal{U})$.



Definição: Seja X um espaço topológico, e $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$ uma *cobertura* de X , isto é, uma coleção de subconjuntos $U_i \subset X$ tal que

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

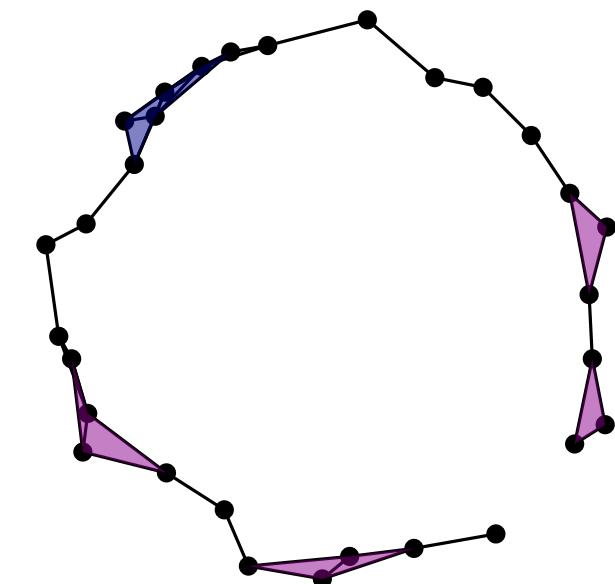
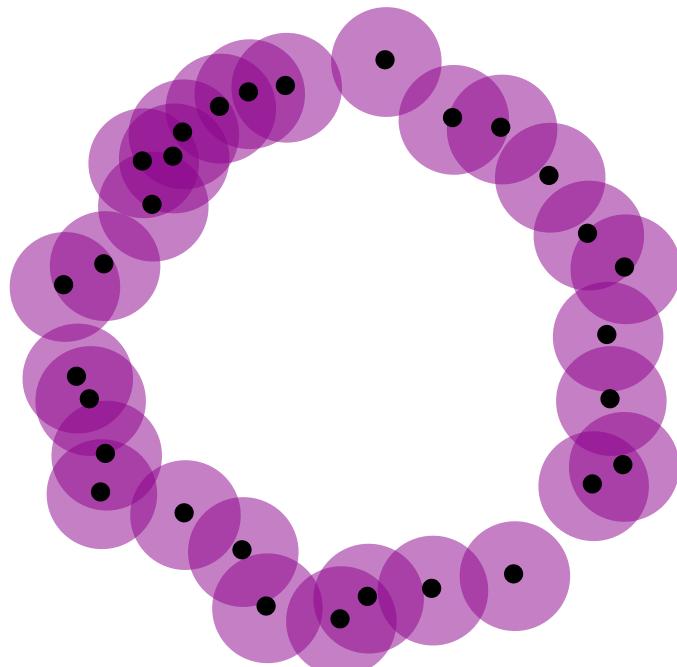
O **nervo** de \mathcal{U} é o complexo simplicial com conjunto de vértices $\{1, \dots, N\}$ e cujos m -simplexos são os subconjuntos $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$ tais que $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$. Ele é denotado $\mathcal{N}(\mathcal{U})$.



Definição: Seja X um espaço topológico, e $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$ uma *cobertura* de X , isto é, uma coleção de subconjuntos $U_i \subset X$ tal que

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

O **nervo** de \mathcal{U} é o complexo simplicial com conjunto de vértices $\{1, \dots, N\}$ e cujos m -simplexos são os subconjuntos $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$ tais que $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$. Ele é denotado $\mathcal{N}(\mathcal{U})$.

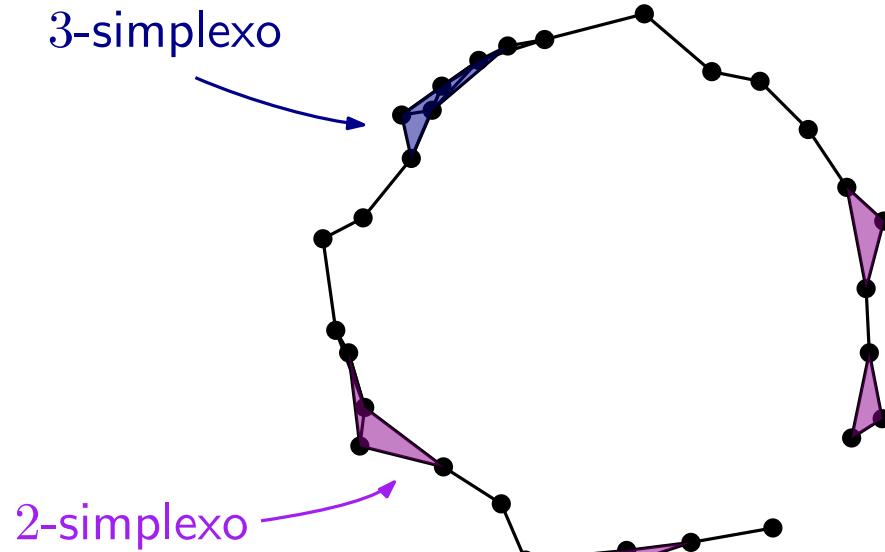
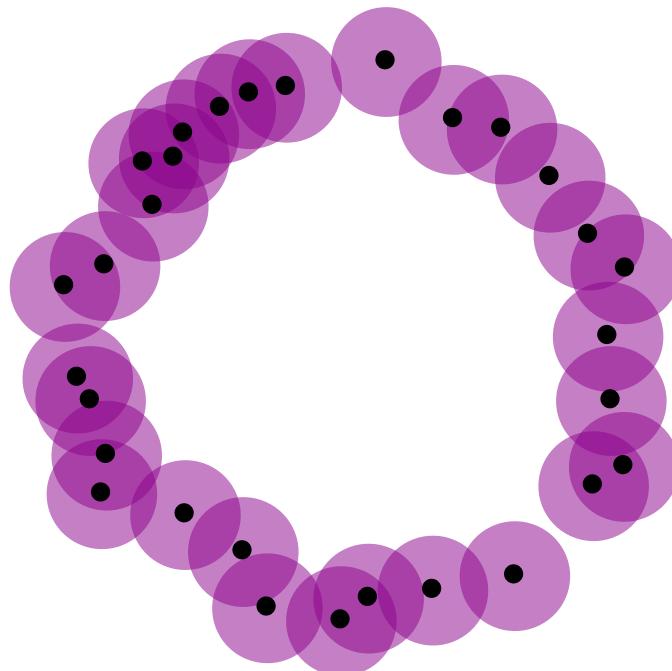


$X^{0.2} = \bigcup_{x \in X} \overline{\mathcal{B}}(x, 0.2)$ é coberto por $\mathcal{U} = \{\overline{\mathcal{B}}(x, 0.2), x \in X\}$

Definição: Seja X um espaço topológico, e $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$ uma *cobertura* de X , isto é, uma coleção de subconjuntos $U_i \subset X$ tal que

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

O **nervo** de \mathcal{U} é o complexo simplicial com conjunto de vértices $\{1, \dots, N\}$ e cujos m -simplexos são os subconjuntos $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$ tais que $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$. Ele é denotado $\mathcal{N}(\mathcal{U})$.

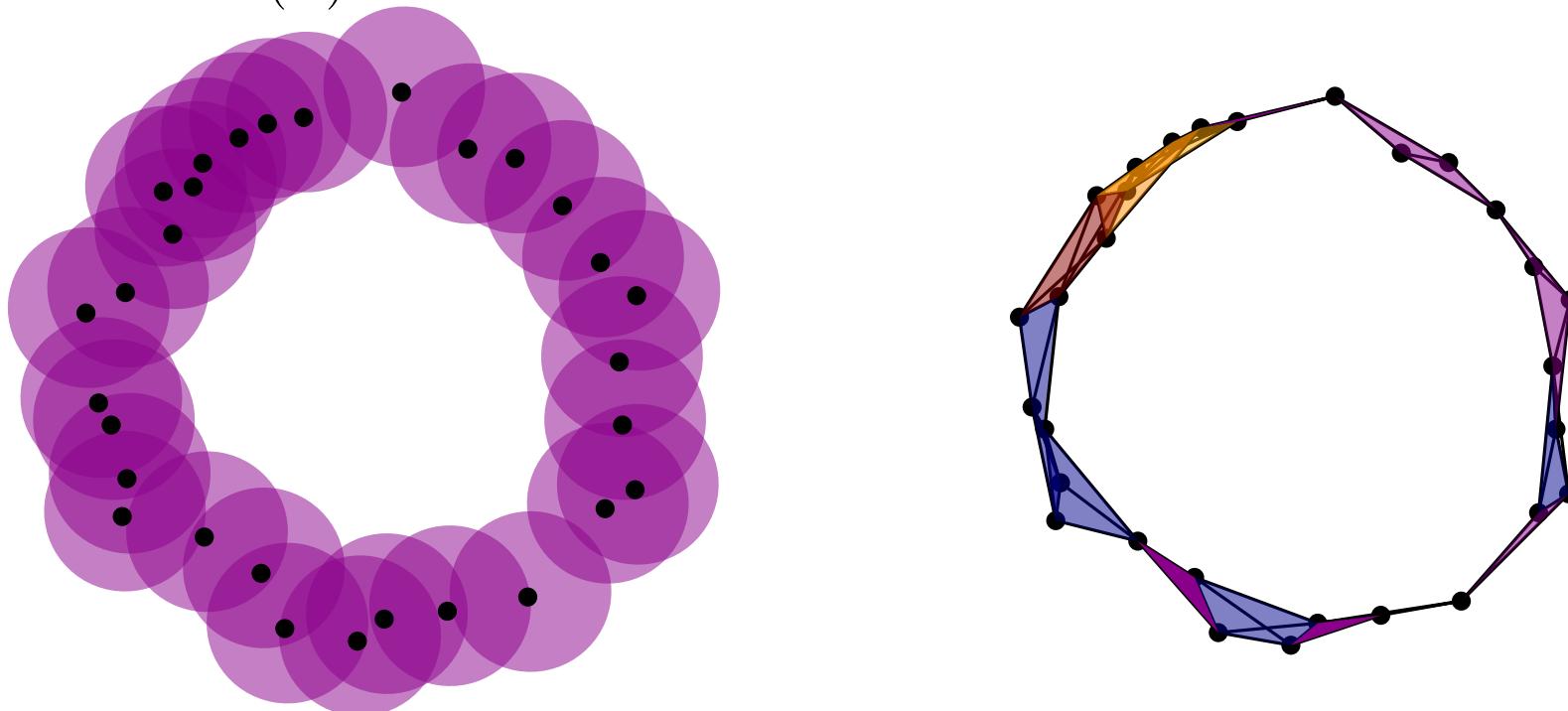


$X^{0.2} = \bigcup_{x \in X} \overline{\mathcal{B}}(x, 0.2)$ é coberto por $\mathcal{U} = \{\overline{\mathcal{B}}(x, 0.2), x \in X\}$

Definição: Seja X um espaço topológico, e $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$ uma *cobertura* de X , isto é, uma coleção de subconjuntos $U_i \subset X$ tal que

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

O **nervo** de \mathcal{U} é o complexo simplicial com conjunto de vértices $\{1, \dots, N\}$ e cujos m -simplexos são os subconjuntos $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$ tais que $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$. Ele é denotado $\mathcal{N}(\mathcal{U})$.

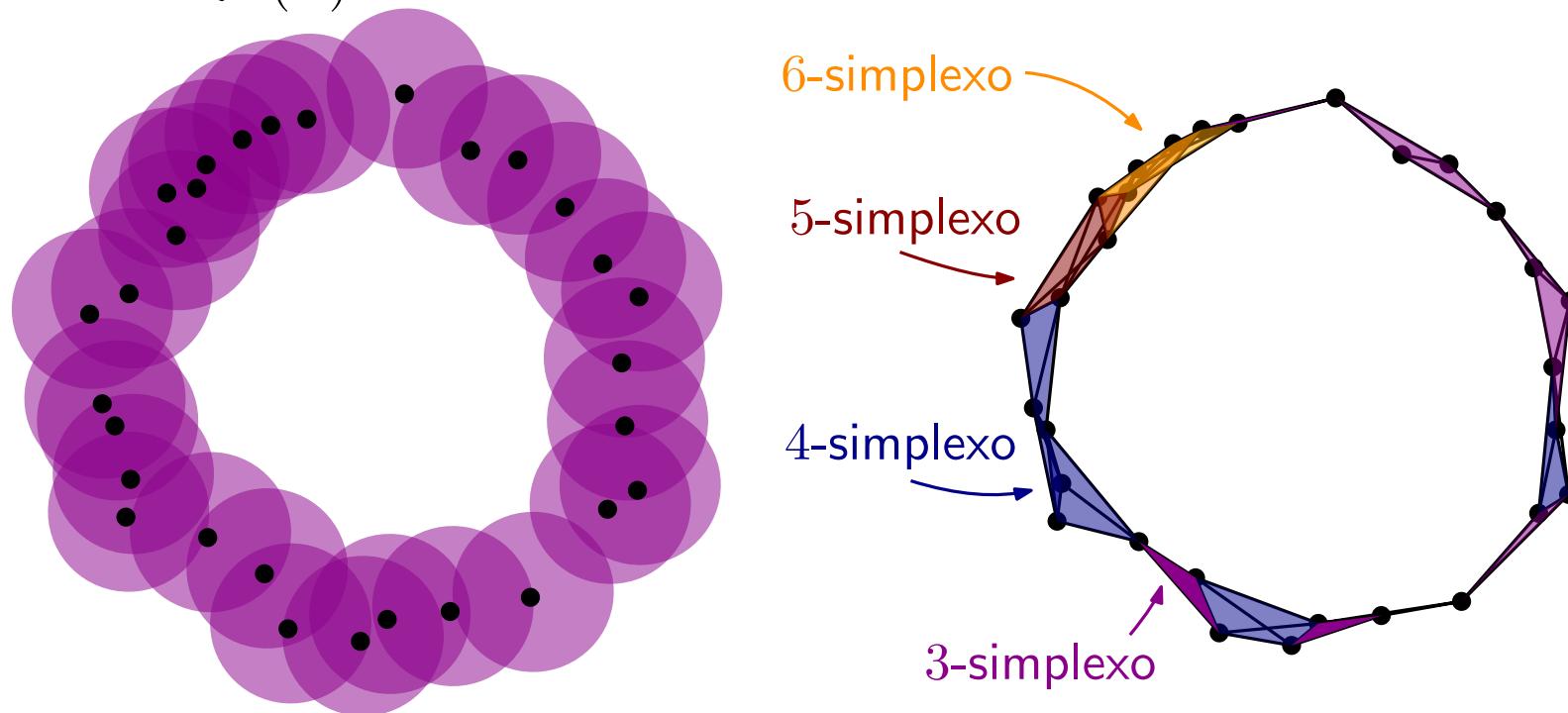


$X^{0.3} = \bigcup_{x \in X} \overline{\mathcal{B}}(x, 0.3)$ é coberto por $\mathcal{U} = \{\overline{\mathcal{B}}(x, 0.3), x \in X\}$

Definição: Seja X um espaço topológico, e $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$ uma *cobertura* de X , isto é, uma coleção de subconjuntos $U_i \subset X$ tal que

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

O **nervo** de \mathcal{U} é o complexo simplicial com conjunto de vértices $\{1, \dots, N\}$ e cujos m -simplexos são os subconjuntos $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$ tais que $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$. Ele é denotado $\mathcal{N}(\mathcal{U})$.



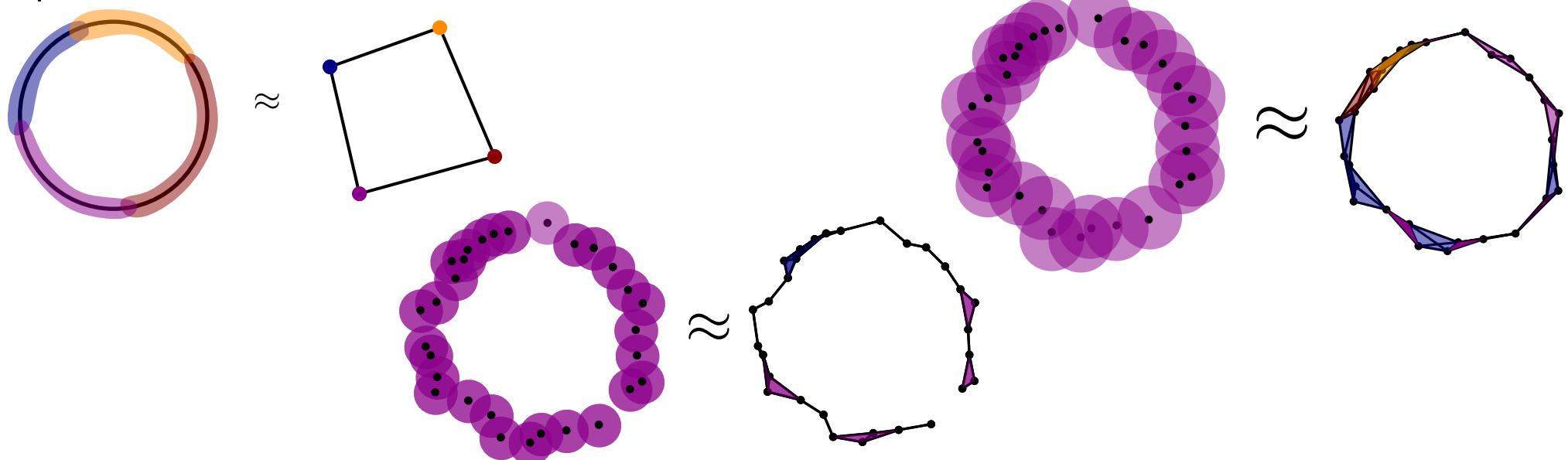
$X^{0.3} = \bigcup_{x \in X} \overline{\mathcal{B}}(x, 0.3)$ é coberto por $\mathcal{U} = \{\overline{\mathcal{B}}(x, 0.3), x \in X\}$

Definição: Seja X um espaço topológico, e $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$ uma *cobertura* de X , isto é, uma coleção de subconjuntos $U_i \subset X$ tal que

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

O **nervo** de \mathcal{U} é o complexo simplicial com conjunto de vértices $\{1, \dots, N\}$ e cujos m -simplexos são os subconjuntos $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$ tais que $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$. Ele é denotado $\mathcal{N}(\mathcal{U})$.

Teorema do nervo: Seja $X \subset \mathbb{R}^n$. Suponha que cada U_i seja uma bola (ou mais geralmente, um subconjunto fechado e convexo). Então $\mathcal{N}(\mathcal{U})$ é homotopicamente equivalente a X .

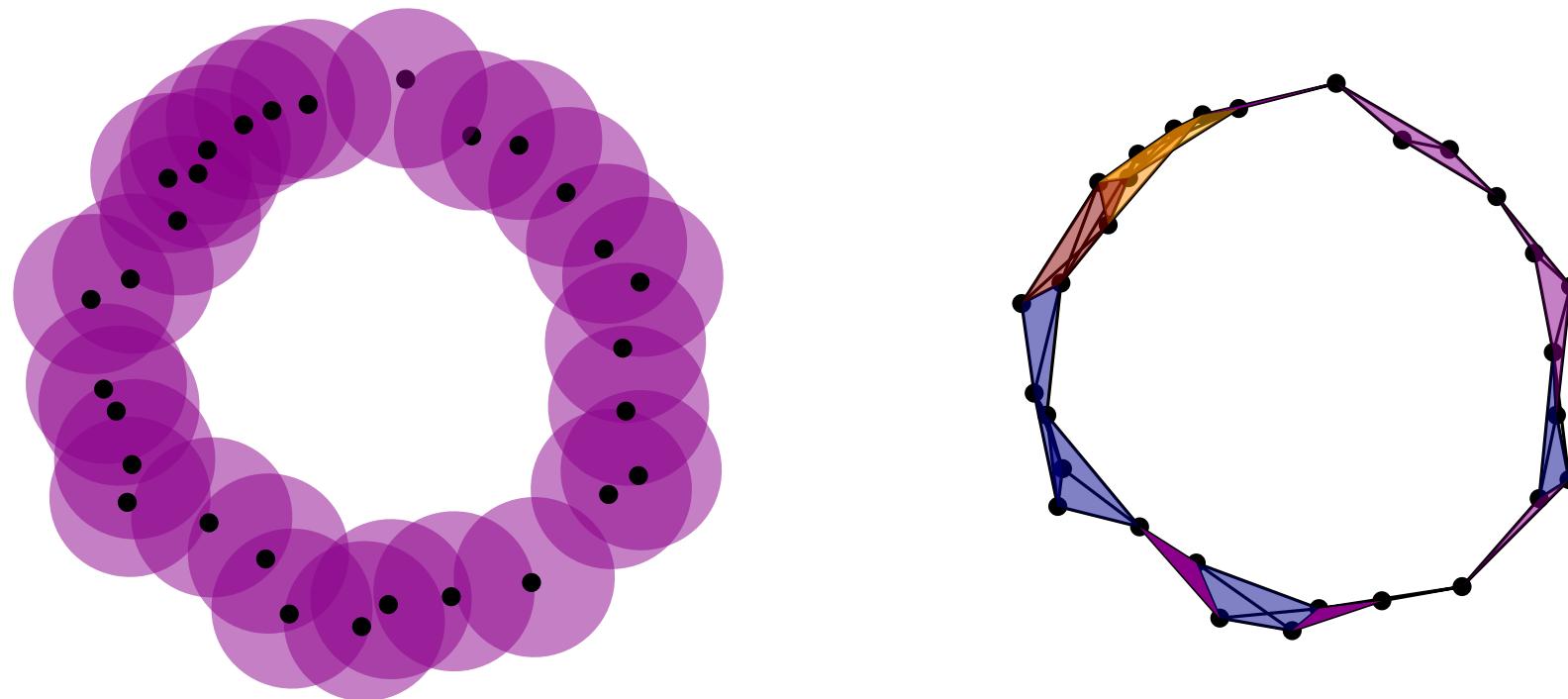


Seja X um subconjunto finito de \mathbb{R}^n , et $t \geq 0$. Seja a coleção

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

É uma cobertura do espessamento X^t , onde cada elemento é uma bola fechada. Pelo teorema do nervo, seu nervo $\mathcal{N}(\mathcal{V}^t)$ é homotopicamente equivalente a X^t .

Definição: Este nervo é notado $\check{\text{Cech}}^t(X)$ e é chamado de **complexo de Čech no tempo t** .

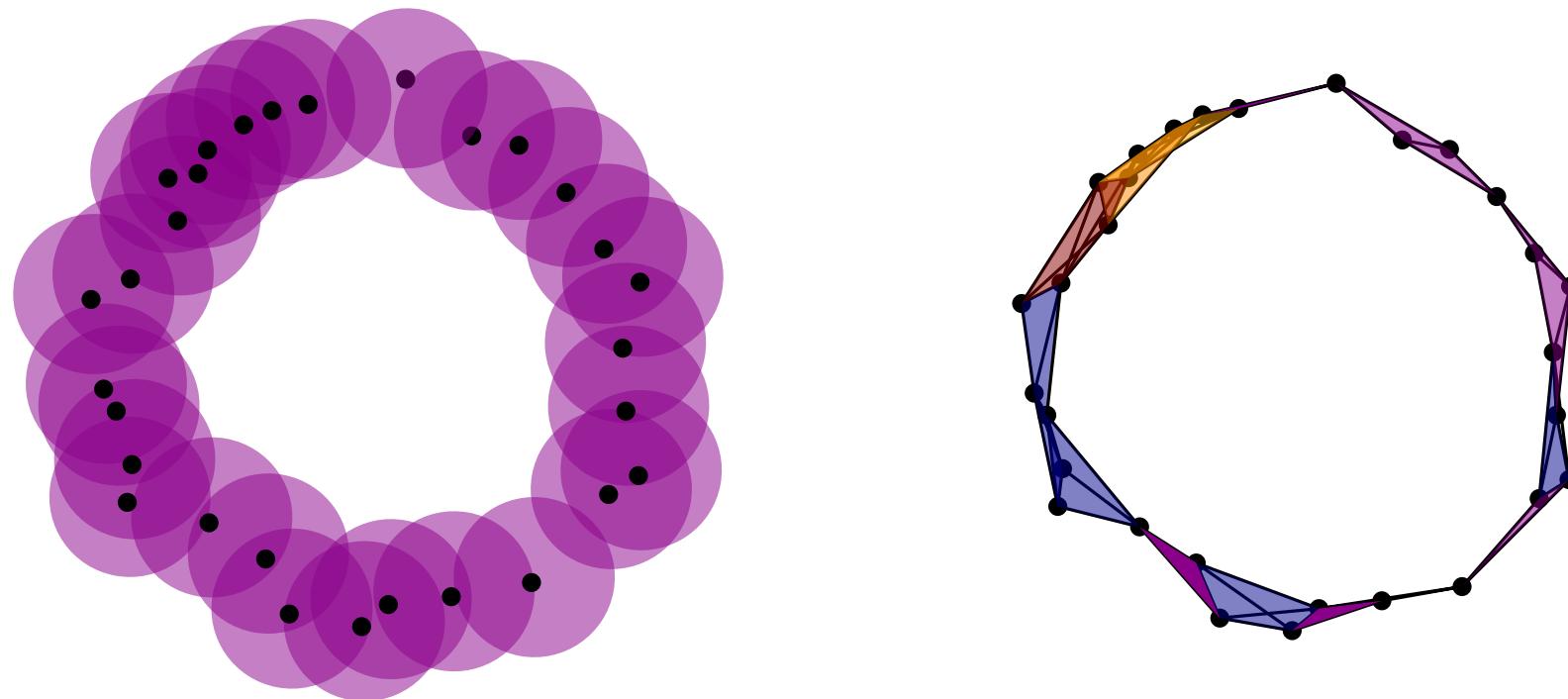


Seja X um subconjunto finito de \mathbb{R}^n , et $t \geq 0$. Seja a coleção

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

É uma cobertura do espessamento X^t , onde cada elemento é uma bola fechada. Pelo teorema do nervo, seu nervo $\mathcal{N}(\mathcal{V}^t)$ é homotopicamente equivalente a X^t .

Definição: Este nervo é notado $\check{\text{Cech}}^t(X)$ e é chamado de **complexo de Čech no tempo t** .



→ A Questão 2 (Como calcular os grupos de homologia de X^t ?) é resolvida.

Seja $X = \{x_1, \dots, x_N\} \subset \mathbb{R}^n$ be finito, $t \geq 0$ e considere o t -espessamento

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

Por definição, seu nervo, $\check{\text{Cech}}^t(X)$, o complexo de Čech no tempo t , é um complexo simplicial nos vértices $\{1, \dots, N\}$ cujos simplices são os subconjuntos $\{i_1, \dots, i_m\}$ tal que

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

Seja $X = \{x_1, \dots, x_N\} \subset \mathbb{R}^n$ be finito, $t \geq 0$ e considere o t -espessamento

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

Por definição, seu nervo, $\check{\text{Cech}}^t(X)$, o complexo de Čech no tempo t , é um complexo simplicial nos vértices $\{1, \dots, N\}$ cujos simplices são os subconjuntos $\{i_1, \dots, i_m\}$ tal que

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

Portanto, o cálculo do complexo de Čech se baseia no seguinte predicado geométrico:

Dadas m bolas fechadas de \mathbb{R}^n , elas se intersectam?

Esse problema é conhecido como o *smallest circle problem*.

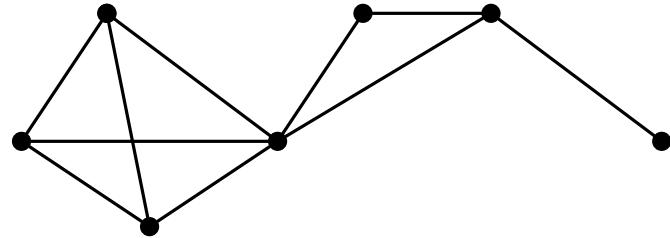
Ele pode ser resolvido em um tempo de $O(m)$, onde m é o número de pontos.

Complexo de clique

13/38 (1/6)

Seja G um grafo.

Chamamos de **clique** de G um conjunto de vértices v_1, \dots, v_m tal que, para cada $i, j \in \llbracket 1, m \rrbracket$ com $i \neq j$, a aresta $[v_i, v_j]$ pertence a G .

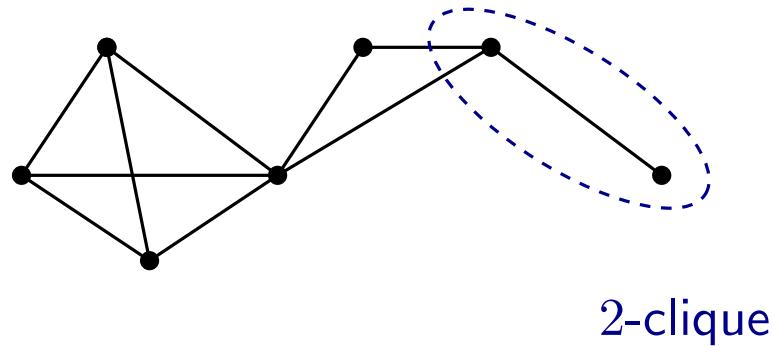


Complexo de clique

13/38 (2/6)

Seja G um grafo.

Chamamos de **clique** de G um conjunto de vértices v_1, \dots, v_m tal que, para cada $i, j \in \llbracket 1, m \rrbracket$ com $i \neq j$, a aresta $[v_i, v_j]$ pertence a G .

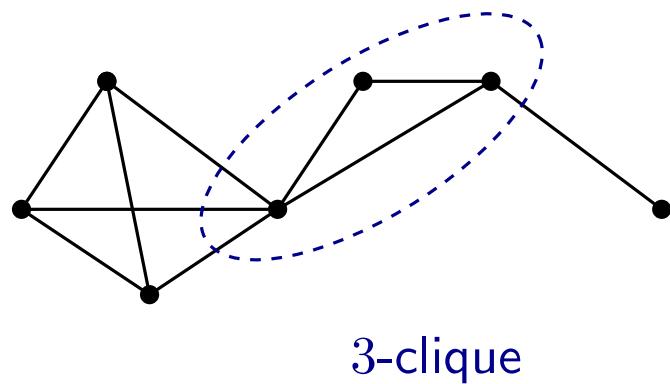


Complexo de clique

13/38 (3/6)

Seja G um grafo.

Chamamos de **clique** de G um conjunto de vértices v_1, \dots, v_m tal que, para cada $i, j \in \llbracket 1, m \rrbracket$ com $i \neq j$, a aresta $[v_i, v_j]$ pertence a G .

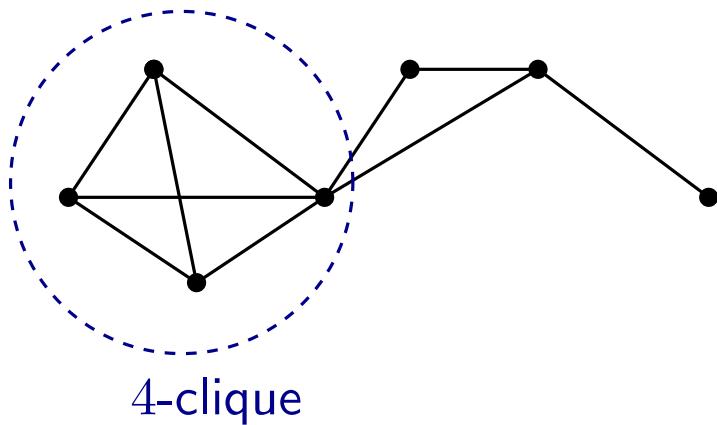


Complexo de clique

13/38 (4/6)

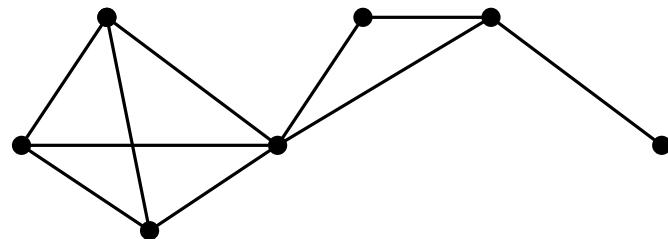
Seja G um grafo.

Chamamos de **clique** de G um conjunto de vértices v_1, \dots, v_m tal que, para cada $i, j \in \llbracket 1, m \rrbracket$ com $i \neq j$, a aresta $[v_i, v_j]$ pertence a G .



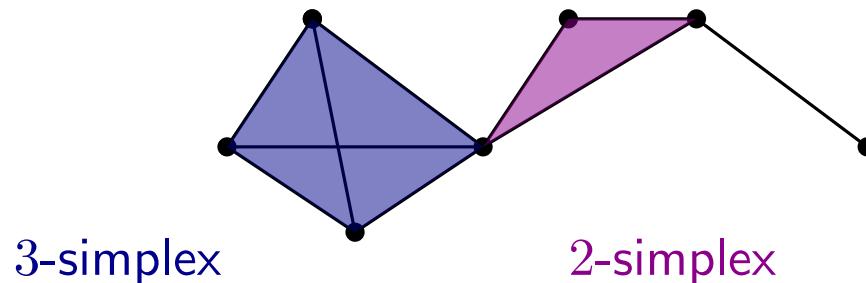
Seja G um grafo.

Chamamos de **clique** de G um conjunto de vértices v_1, \dots, v_m tal que, para cada $i, j \in \llbracket 1, m \rrbracket$ com $i \neq j$, a aresta $[v_i, v_j]$ pertence a G .



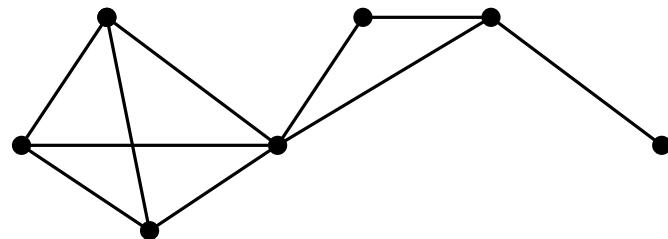
Definição: Dado um grafo G , o **complexo de clique** associado é o complexo simplicial

- cujos vértices são os vértices de G ,
- cujos simplices são os conjuntos de vértices dos cliques de G .



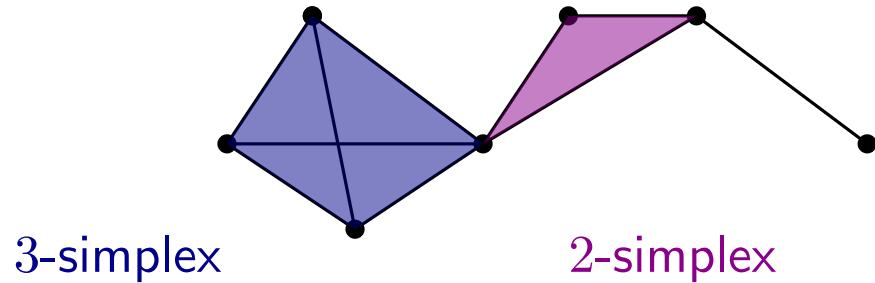
Seja G um grafo.

Chamamos de **clique** de G um conjunto de vértices v_1, \dots, v_m tal que, para cada $i, j \in \llbracket 1, m \rrbracket$ com $i \neq j$, a aresta $[v_i, v_j]$ pertence a G .



Definição: Dado um grafo G , o **complexo de clique** associado é o complexo simplicial

- cujos vértices são os vértices de G ,
- cujos simplices são os conjuntos de vértices dos cliques de G .

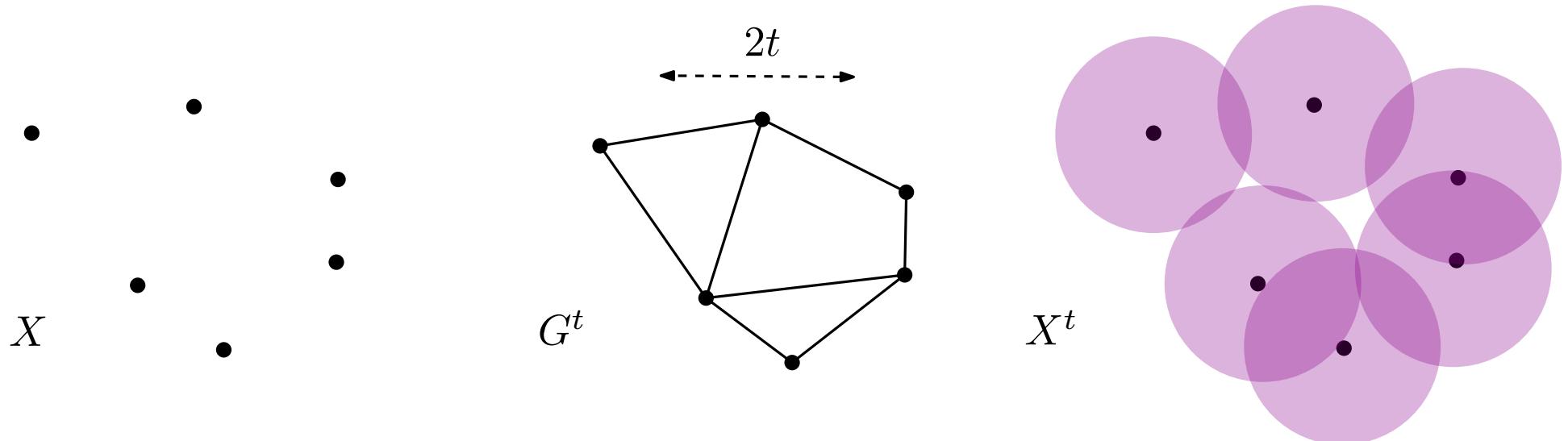


Observação: O complexo de clique de um grafo é um complexo simplicial.

Seja $X = \{x_1, \dots, x_N\} \subset \mathbb{R}^n$ e $t \geq 0$.

Considere o grafo G^t cujo conjunto de vértices é $\{1, \dots, N\}$ e cujas arestas são os pares (i, j) tais que $\|x_i - x_j\| \leq 2t$.

De modo equivalente, G^t pode ser visto como o 1-esqueleto do complexo $\check{\text{C}}\text{ech}^t(X)$.



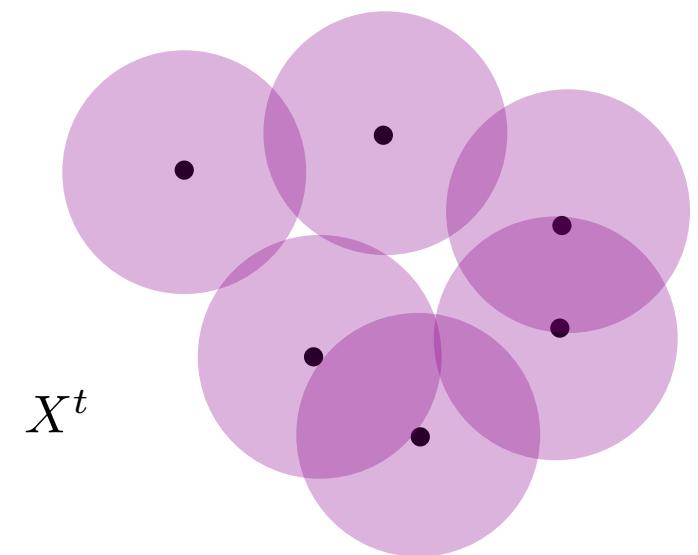
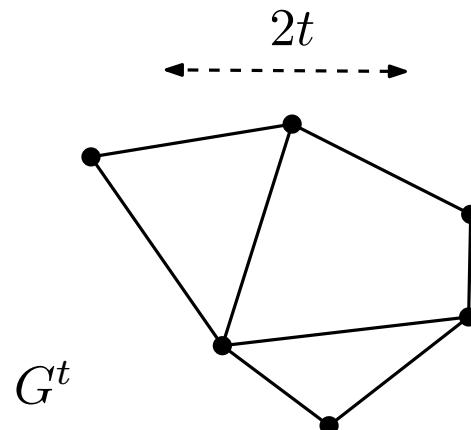
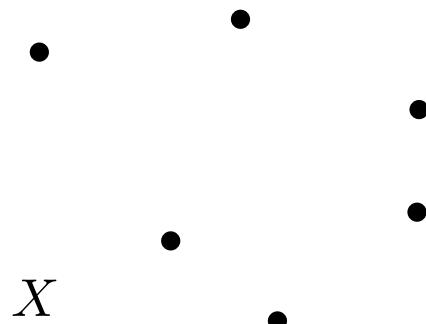
Complexo de Rips

14/38 (2/6)

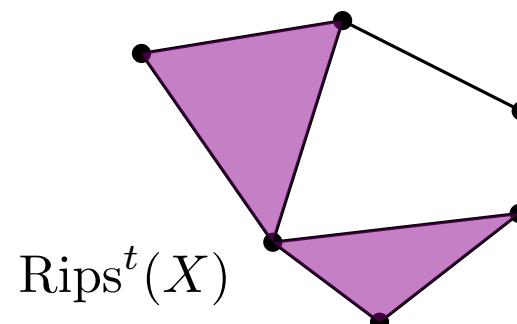
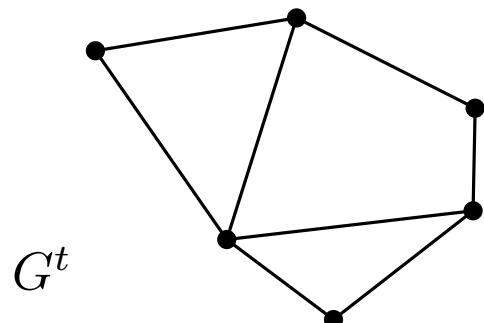
Seja $X = \{x_1, \dots, x_N\} \subset \mathbb{R}^n$ e $t \geq 0$.

Considere o grafo G^t cujo conjunto de vértices é $\{1, \dots, N\}$ e cujas arestas são os pares (i, j) tais que $\|x_i - x_j\| \leq 2t$.

De modo equivalente, G^t pode ser visto como o 1-esqueleto do complexo $\check{\text{C}}\text{ech}^t(X)$.



Definição: O **complexo de Rips** de X no tempo t é o complexo de clique do grafo G^t . Denotamos esse complexo por $\text{Rips}^t(X)$.



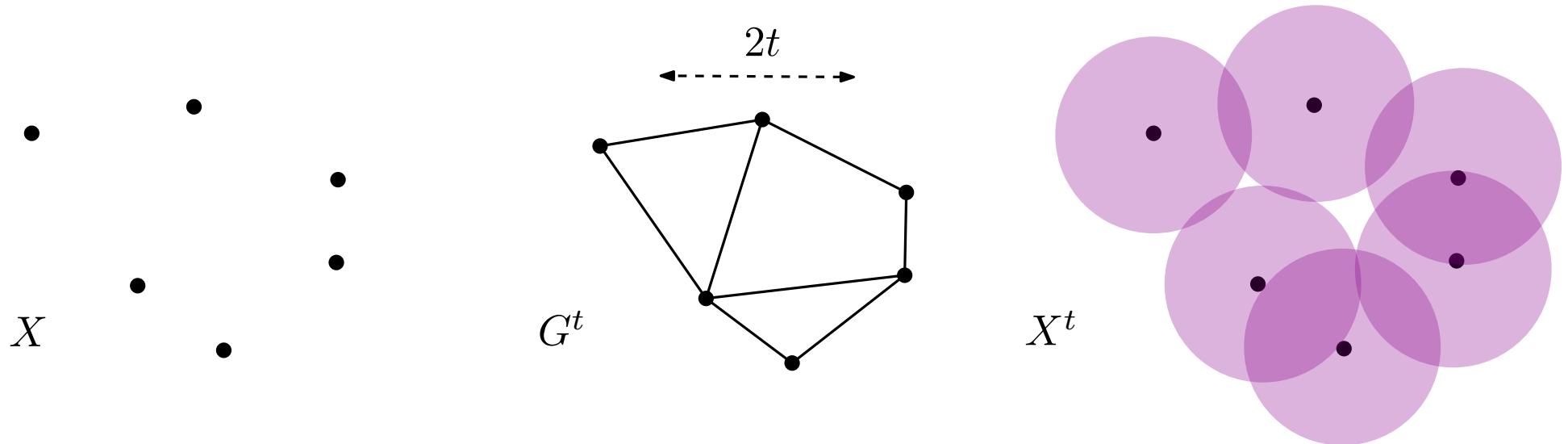
Complexo de Rips

14/38 (3/6)

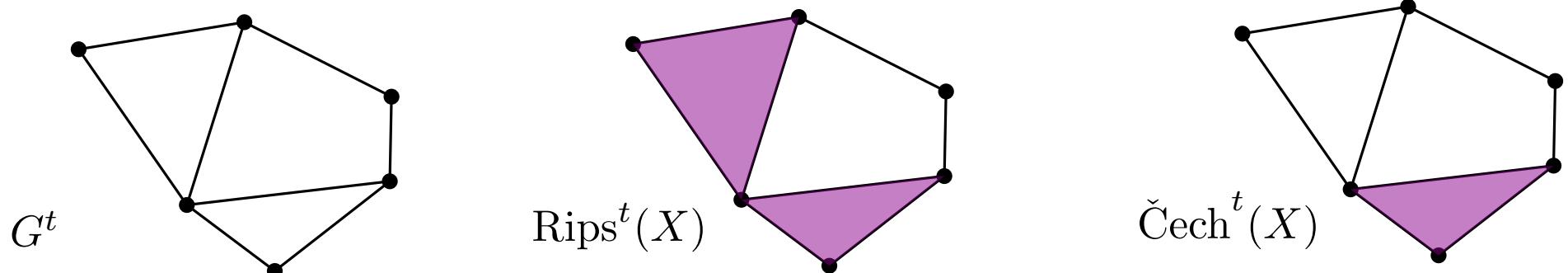
Seja $X = \{x_1, \dots, x_N\} \subset \mathbb{R}^n$ e $t \geq 0$.

Considere o grafo G^t cujo conjunto de vértices é $\{1, \dots, N\}$ e cujas arestas são os pares (i, j) tais que $\|x_i - x_j\| \leq 2t$.

De modo equivalente, G^t pode ser visto como o 1-esqueleto do complexo $\check{\text{C}}\text{ech}^t(X)$.



Definição: O **complexo de Rips** de X no tempo t é o complexo de clique do grafo G^t . Denotamos esse complexo por $\text{Rips}^t(X)$.



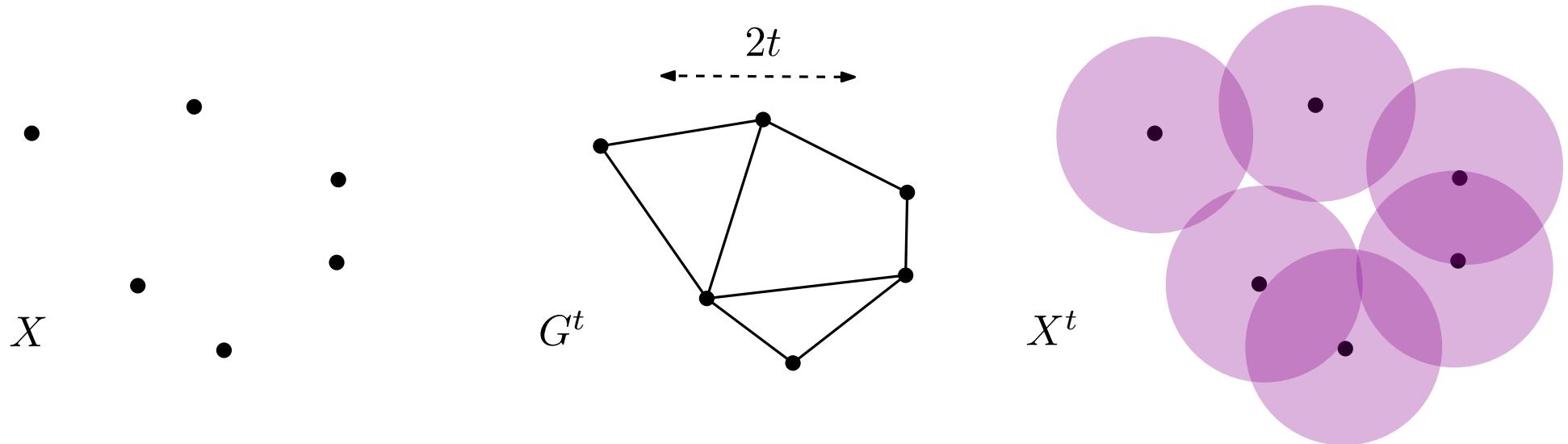
Complexo de Rips

14/38 (4/6)

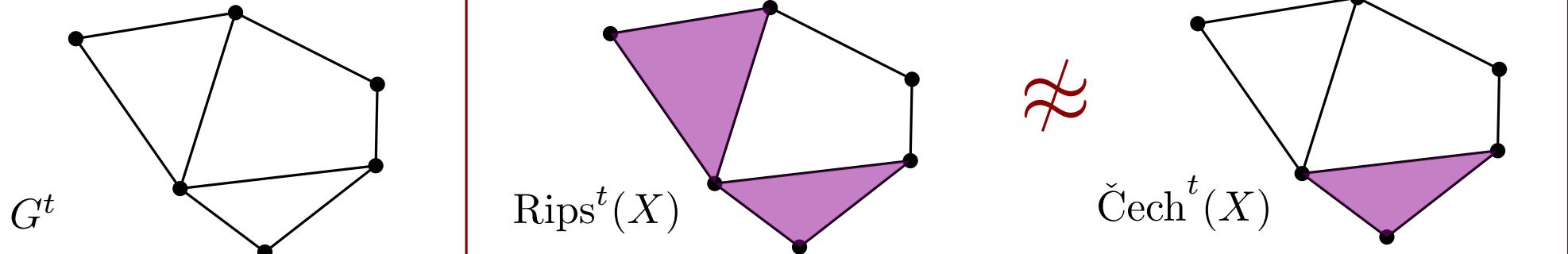
Seja $X = \{x_1, \dots, x_N\} \subset \mathbb{R}^n$ e $t \geq 0$.

Considere o grafo G^t cujo conjunto de vértices é $\{1, \dots, N\}$ e cujas arestas são os pares (i, j) tais que $\|x_i - x_j\| \leq 2t$.

De modo equivalente, G^t pode ser visto como o 1-esqueleto do complexo $\check{\text{C}}\text{ech}^t(X)$.



Definição: O **complexo de Rips** de X no tempo t é o complexo de clique do grafo G^t . Denotamos esse complexo por $\text{Rips}^t(X)$.

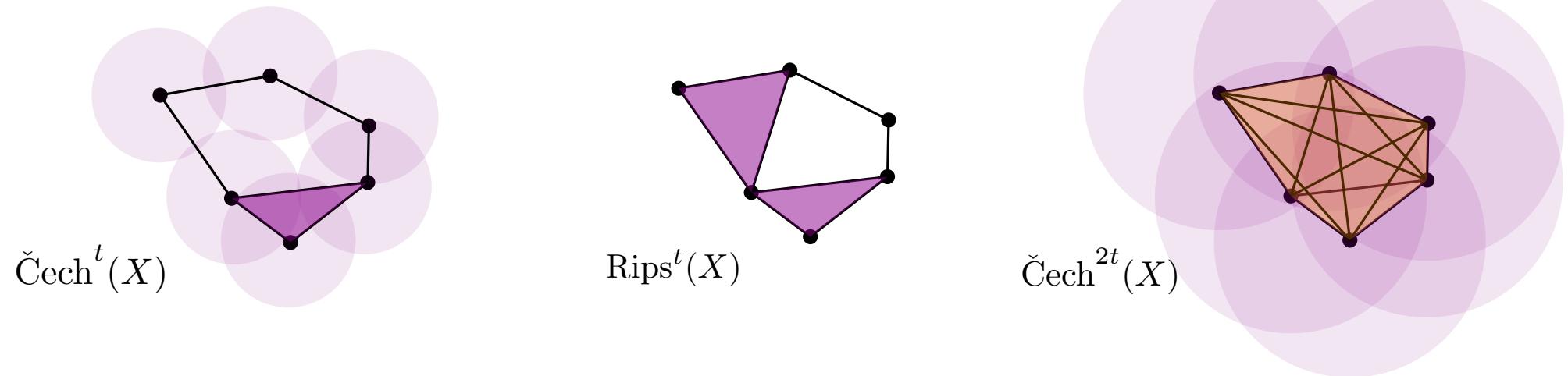


Complexo de Rips

14/38 (5/6)

Proposição: Para todo $t \geq 0$, temos

$$\check{\text{Cech}}^t(X) \subset \text{Rips}^t(X) \subset \check{\text{Cech}}^{2t}(X).$$

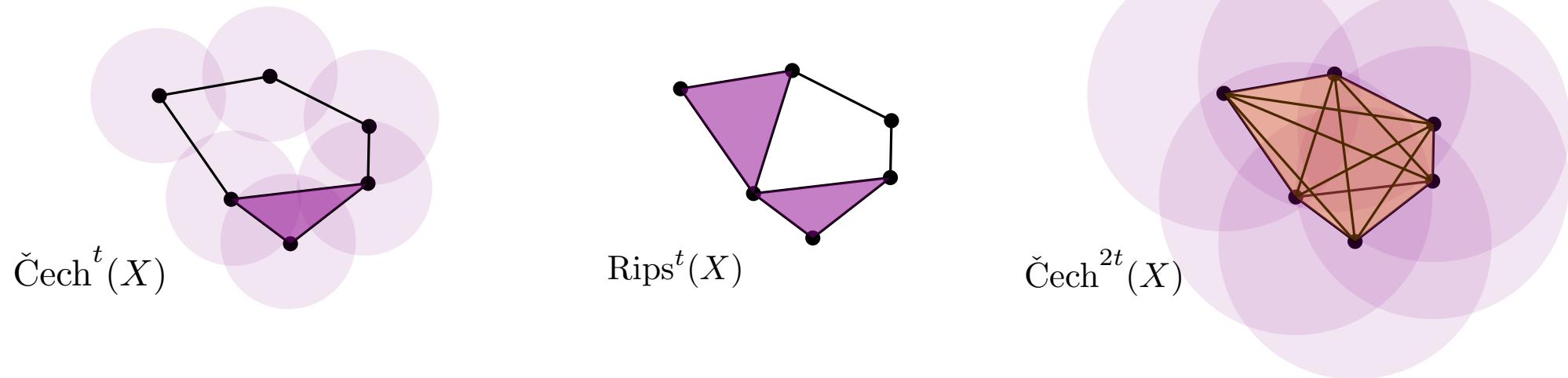


Complexo de Rips

14/38 (6/6)

Proposição: Para todo $t \geq 0$, temos

$$\check{\text{Cech}}^t(X) \subset \text{Rips}^t(X) \subset \check{\text{Cech}}^{2t}(X).$$



Proof: Seja $t \geq 0$. A primeira inclusão decorre do fato de que $\text{Rips}^t(X)$ é o complexo de clique de $\check{\text{Cech}}^t(X)$.

Para provar o segundo, escolha um simplex $\sigma \in \text{Rips}^t(X)$. Vamos provar que $\omega \in \check{\text{Cech}}^{2t}(X)$.

Seja $x \in \sigma$ um vértice qualquer. Observe que $\forall y \in \sigma$, temos $\|x - y\| \leq 2t$ pela definição do complexo de Rips. Assim, temos

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

Como a interseção não é vazia, deduzimos $\sigma \in \check{\text{Cech}}^{2t}(X)$.

I - Comparar espaços topológicos

- 1 - Equivalência de homeomorfia
- 2 - Equivalência de homotopia

II - Invariante topológico

- 1 - Número de componentes conexas
- 2 - Característica de Euler
- 3 - Números de Betti
- 4 - Homologia (simplicial sobre $\mathbb{Z}/2\mathbb{Z}$)

III - Inferência topológica

- 1 - Estimação do parâmetro t
- 2 - Nervos

IV - Homologia persistente

- 1 - Módulos de persistência
- 2 - Decomposição
- 3 - Estabilidade

V - Aplicações

O problema da escala

16/38 (1/3)

Questão 1 : Como selecionar um t tal que $X^t \approx \mathcal{M}$?

Teorema (Frédéric Chazal, David Cohen-Steiner et André Lieutier, 2009):

Sejam X e \mathcal{M} dois subconjuntos de \mathbb{R}^n . Suponha que \mathcal{M} tenha um *reach* estritamente positivo, e que $d_H(X, \mathcal{M}) \leq \frac{1}{17} \text{reach}(\mathcal{M})$.

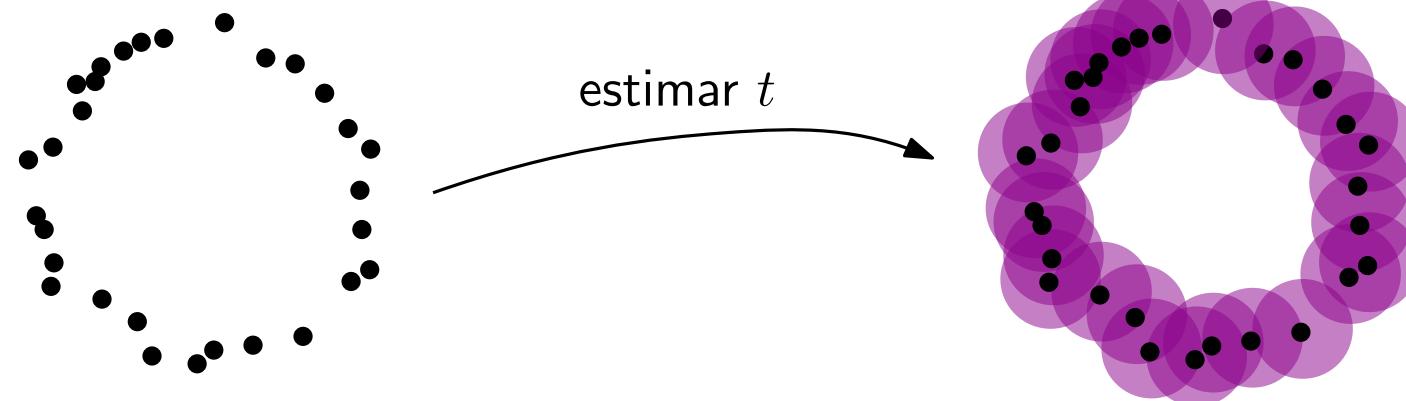
Então X^t e \mathcal{M} são homotopicamente equivalentes, desde que

$$t \in [4d_H(X, \mathcal{M}), \text{reach}(\mathcal{M}) - 3d_H(X, \mathcal{M})].$$

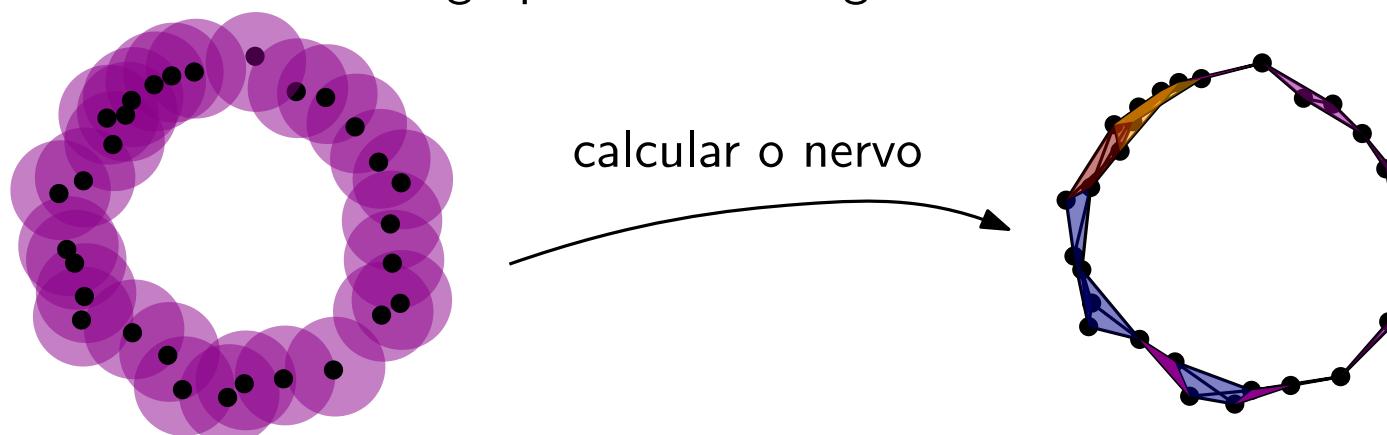
Teorema (Partha Niyogi, Stephen Smale et Shmuel Weinberger, 2008):

Sejam X e \mathcal{M} dois subconjuntos de \mathbb{R}^n , com \mathcal{M} uma subvariedade, e X um subconjunto finito de \mathcal{M} . Suponha que \mathcal{M} tenha um *reach* estritamente positivo. Então X^t e \mathcal{M} são homotopicamente equivalentes, desde que

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



Questão 2: Como calcular os grupos de homologia de X^t ?



O problema da escala

16/38 (2/3)

Questão 1 : Como selecionar um t tal que $X^t \approx \mathcal{M}$?

Teorema (Frédéric Chazal, David Cohen-Steiner et André Lieutier, 2009):

Sejam X e \mathcal{M} dois subconjuntos de \mathbb{R}^n . Suponha que \mathcal{M} tenha um *reach* estritamente positivo, e que $d_H(X, \mathcal{M}) \leq \frac{1}{17} \text{reach}(\mathcal{M})$.

Então X^t e \mathcal{M} são homotopicamente equivalentes, desde que

$$t \in [4d_H(X, \mathcal{M}), \text{reach}(\mathcal{M}) - 3d_H(X, \mathcal{M})].$$

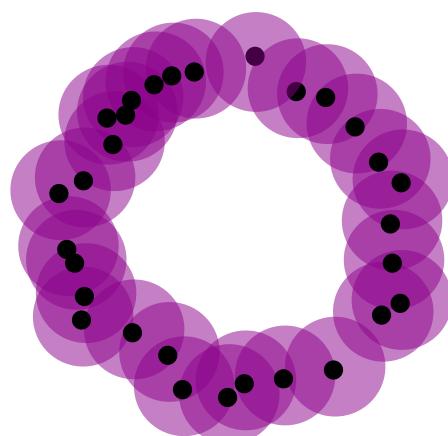
Teorema (Partha Niyogi, Stephen Smale et Shmuel Weinberger, 2008):

Sejam X e \mathcal{M} dois subconjuntos de \mathbb{R}^n , com \mathcal{M} uma subvariedade, e X um subconjunto finito de \mathcal{M} . Suponha que \mathcal{M} tenha um *reach* estritamente positivo. Então X^t e \mathcal{M} são homotopicamente equivalentes, desde que

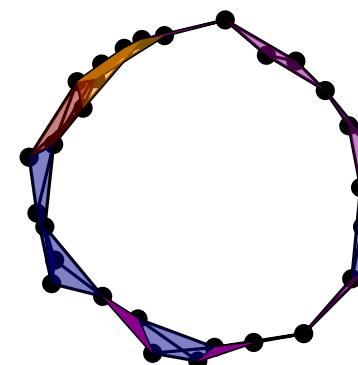
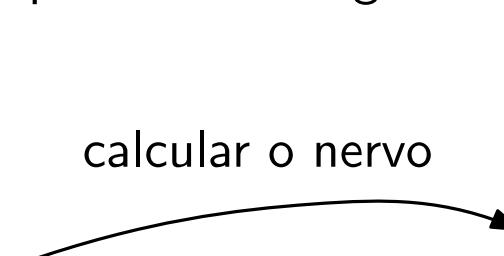
$$t \in \left[2d_H(X, \mathcal{M}), \sqrt{\frac{3}{5}} \text{reach}(\mathcal{M}) \right].$$

↑
não conhecemos estas quantidades!

Questão 2: Como calcular os grupos de homologia de X^t ?



calcular o nervo



O problema da escala

16/38 (3/3)

Questão 1 : Como selecionar um t tal que $X^t \approx \mathcal{M}$?

Teorema (Frédéric Chazal, David Cohen-Steiner et André Lieutier, 2009):

Sejam X e \mathcal{M} dois subconjuntos de \mathbb{R}^n . Suponha que \mathcal{M} tenha um *reach* estritamente positivo, e que $d_H(X, \mathcal{M}) \leq \frac{1}{17} \text{reach}(\mathcal{M})$.

Então X^t e \mathcal{M} são homotopicamente equivalentes, desde que

$$t \in [4d_H(X, \mathcal{M}), \text{reach}(\mathcal{M}) - 3d_H(X, \mathcal{M})].$$

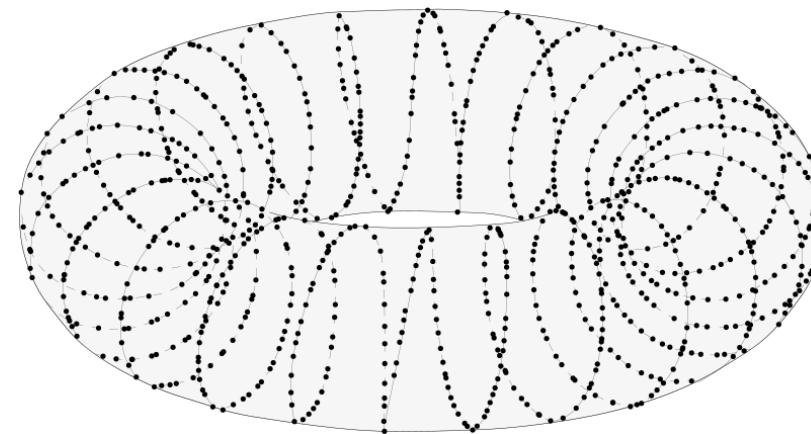
Teorema (Partha Niyogi, Stephen Smale et Shmuel Weinberger, 2008):

Sejam X e \mathcal{M} dois subconjuntos de \mathbb{R}^n , com \mathcal{M} uma subvariedade, e X um subconjunto finito de \mathcal{M} . Suponha que \mathcal{M} tenha um *reach* estritamente positivo. Então X^t e \mathcal{M} são homotopicamente equivalentes, desde que

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

↑
não conhecemos estas quantidades!

Este objeto é de dimensão 1 ou 2?



Idéia (análise multiescala): Em vez de escolher um valor por t , escolheremos todos eles.

Definição: A **filtração de Čech** de X é a coleção $V[X] = (X^t)_{t \geq 0}$.

I - Comparar espaços topológicos

- 1 - Equivalência de homeomorfia
- 2 - Equivalência de homotopia

II - Invariante topológico

- 1 - Número de componentes conexas
- 2 - Característica de Euler
- 3 - Números de Betti
- 4 - Homologia (simplicial sobre $\mathbb{Z}/2\mathbb{Z}$)

III - Inferência topológica

- 1 - Estimação do parâmetro t
- 2 - Nervos

IV - Homologia persistente

- 1 - Módulos de persistência
- 2 - Decomposição
- 3 - Estabilidade

V - Aplicações

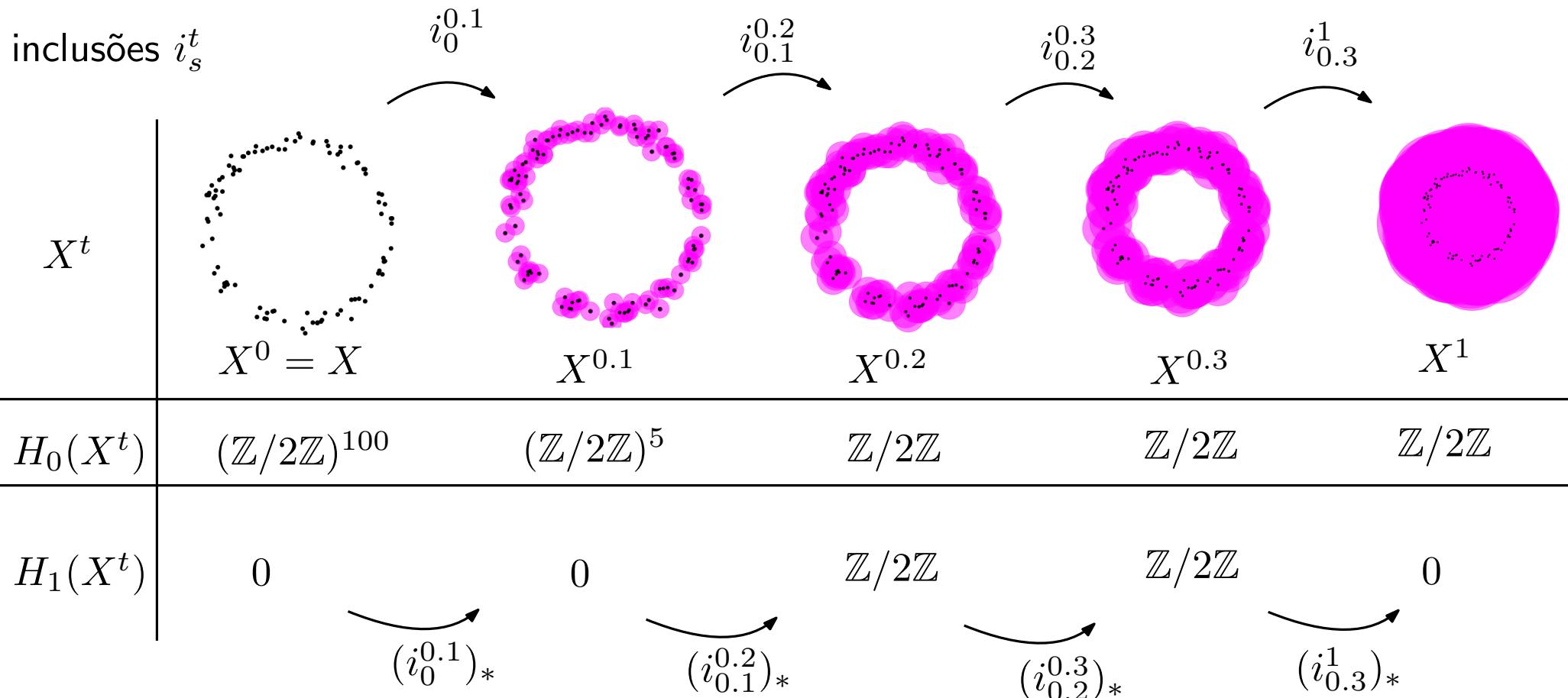
Calculemos a homologia de cada espessamento:

X^t	$X^0 = X$	$X^{0.1}$	$X^{0.2}$	$X^{0.3}$	X^1
$H_0(X^t)$	$(\mathbb{Z}/2\mathbb{Z})^{100}$	$(\mathbb{Z}/2\mathbb{Z})^5$	$\mathbb{Z}/2\mathbb{Z}$	$\mathbb{Z}/2\mathbb{Z}$	$\mathbb{Z}/2\mathbb{Z}$
$H_1(X^t)$	0	0	$\mathbb{Z}/2\mathbb{Z}$	$\mathbb{Z}/2\mathbb{Z}$	0

Homologia da filtração de Cech

18/38 (2/2)

Calculemos a homologia de cada espessamento:



Os dados de $(H_i(X^t))_{t \geq 0}$ e $((i_s^t)_*)_{s \leq t}$ é chamado de **módulo de persistência**.

Definição: Um **módulo de persistência** \mathbb{V} é um par (\mathbb{V}, v) onde $\mathbb{V} = (V^t)_{t \in \mathbb{R}^+}$ é uma coleção de $\mathbb{Z}/2\mathbb{Z}$ -espaços vetoriais, e $v = (v_s^t : V^s \rightarrow V^t)_{s \leq t \in \mathbb{R}^+}$ é uma coleção de transformações lineares tais que

- por todo $t \in \mathbb{R}^+$, $v_t^t : V^t \rightarrow V^t$ é a aplicação identidade,
- por todo $r, s, t \in \mathbb{R}^+$ tais como $r \leq s \leq t$, temos $v_s^t \circ v_r^s = v_r^t$.

$$\begin{array}{ccccc}
 & & v_r^s & & \\
 & V^r & \xrightarrow{\hspace{2cm}} & V^s & \xrightarrow{\hspace{2cm}} V^t \\
 & & \searrow & & \\
 & & v_r^t & &
 \end{array}$$

Módulo de persistência associado à filtração de Cech :

$$\begin{array}{ccccccc}
 \dashrightarrow & X^{t_1} & \xrightarrow{i_{t_1}^{t_2}} & X^{t_2} & \xrightarrow{i_{t_2}^{t_3}} & X^{t_3} & \xrightarrow{i_{t_3}^{t_4}} X^{t_4} & \dashleftarrow \\
 \dashrightarrow & H_i(X^{t_1}) & \xrightarrow{(i_{t_1}^{t_2})_*} & H_i(X^{t_2}) & \xrightarrow{(i_{t_2}^{t_3})_*} & H_i(X^{t_3}) & \xrightarrow{(i_{t_3}^{t_4})_*} H_i(X^{t_4}) & \dashleftarrow
 \end{array}$$

Monitorando a evolução dos buracos

20/38 (1/3)

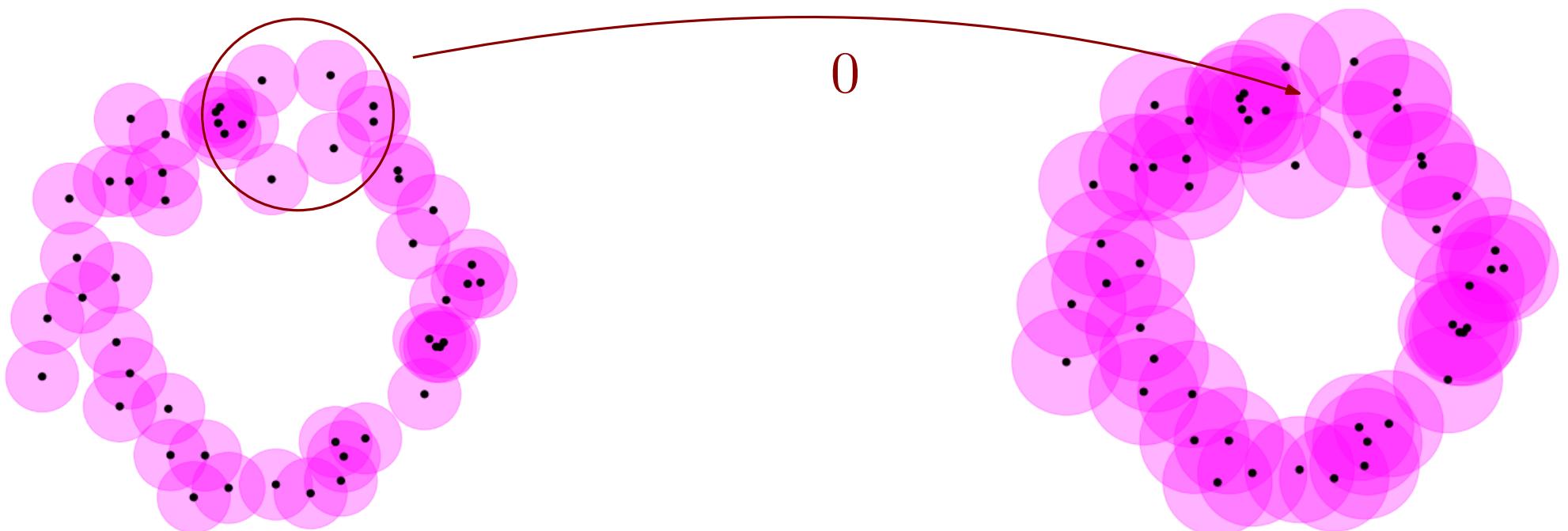
$$\dashrightarrow H_i(X^{t_1}) \xrightarrow{(i_{t_1}^{t_2})_*} H_i(X^{t_2}) \xrightarrow{(i_{t_2}^{t_3})_*} H_i(X^{t_3}) \xrightarrow{(i_{t_3}^{t_4})_*} H_i(X^{t_4}) \dashleftarrow$$

Seja $i \geq 0$, $t_0 \geq 0$ e considere um ciclo $c \in H_i(X^{t_0})$.

Seu **momento de morte** é: $\sup \{t \geq t_0 \mid (i_{t_0}^t)_*(c) \neq 0\}$,

Seu **momento de nascimento** é: $\inf \{t \geq t_0 \mid (i_t^{t_0})^{-1}(\{c\})_* \neq \emptyset\}$,

Sua **persistência** é a diferença.



Monitorando a evolução dos buracos

20/38 (2/3)

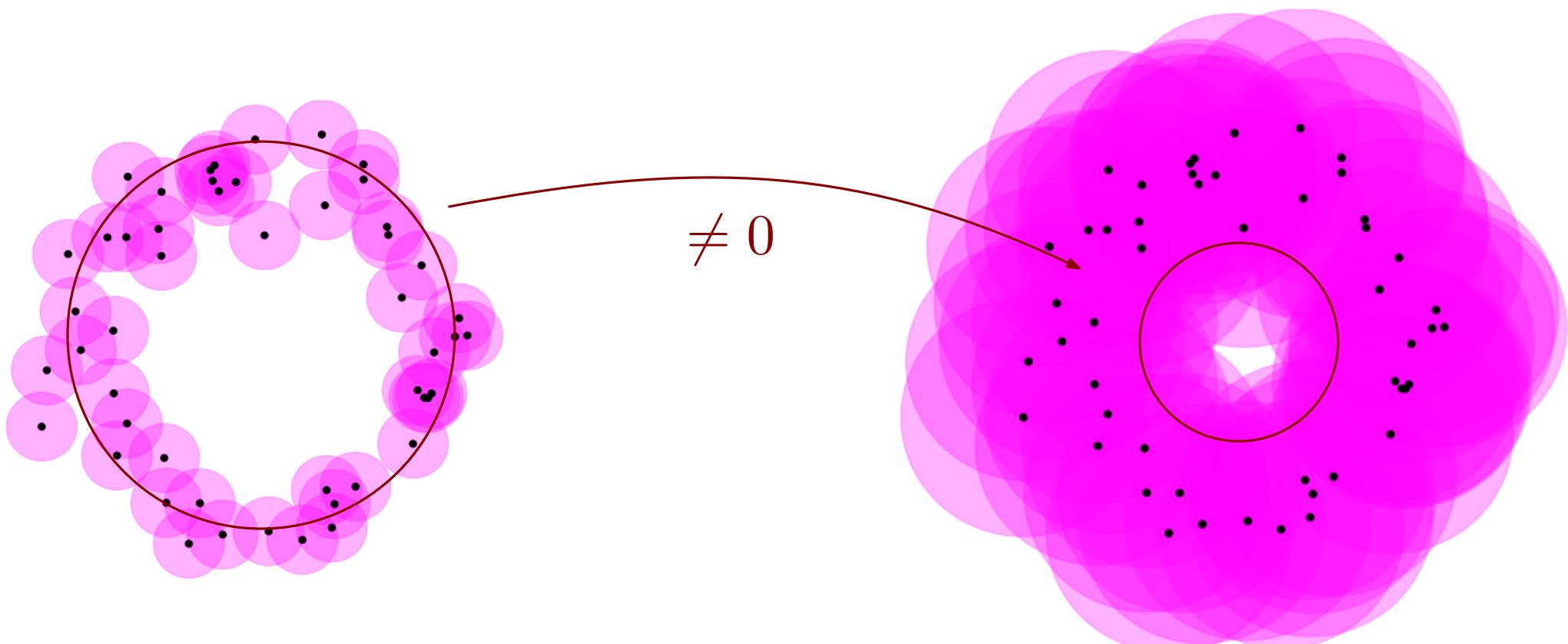
$$\dashrightarrow H_i(X^{t_1}) \xrightarrow{(i_{t_1}^{t_2})_*} H_i(X^{t_2}) \xrightarrow{(i_{t_2}^{t_3})_*} H_i(X^{t_3}) \xrightarrow{(i_{t_3}^{t_4})_*} H_i(X^{t_4}) \dashleftarrow$$

Seja $i \geq 0$, $t_0 \geq 0$ e considere um ciclo $c \in H_i(X^{t_0})$.

Seu **momento de morte** é: $\sup \{t \geq t_0 \mid (i_{t_0}^t)_*(c) \neq 0\}$,

Seu **momento de nascimento** é: $\inf \{t \geq t_0 \mid (i_t^{t_0})^{-1}(\{c\})_* \neq \emptyset\}$,

Sua **persistência** é a diferença.



$$\dashrightarrow H_i(X^{t_1}) \xrightarrow{(i_{t_1}^{t_2})_*} H_i(X^{t_2}) \xrightarrow{(i_{t_2}^{t_3})_*} H_i(X^{t_3}) \xrightarrow{(i_{t_3}^{t_4})_*} H_i(X^{t_4}) \dashleftarrow$$

Seja $i \geq 0$, $t_0 \geq 0$ e considere um ciclo $c \in H_i(X^{t_0})$.

Seu **momento de morte** é: $\sup \{t \geq t_0 \mid (i_{t_0}^t)_*(c) \neq 0\}$,

Seu **momento de nascimento** é: $\inf \{t \geq t_0 \mid (i_t^{t_0})^{-1}(\{c\})_* \neq \emptyset\}$,

Sua **persistência** é a diferença.

Interpretação:

- ciclos com alta persistência correspondem a importantes propriedades topológicas do conjunto de dados,
- ciclos com curta persistência correspondem a ruído topológico.

I - Comparar espaços topológicos

- 1 - Equivalência de homeomorfia
- 2 - Equivalência de homotopia

II - Invariante topológico

- 1 - Número de componentes conexas
- 2 - Característica de Euler
- 3 - Números de Betti
- 4 - Homologia (simplicial sobre $\mathbb{Z}/2\mathbb{Z}$)

III - Inferência topológica

- 1 - Estimação do parâmetro t
- 2 - Nervos

IV - Homologia persistente

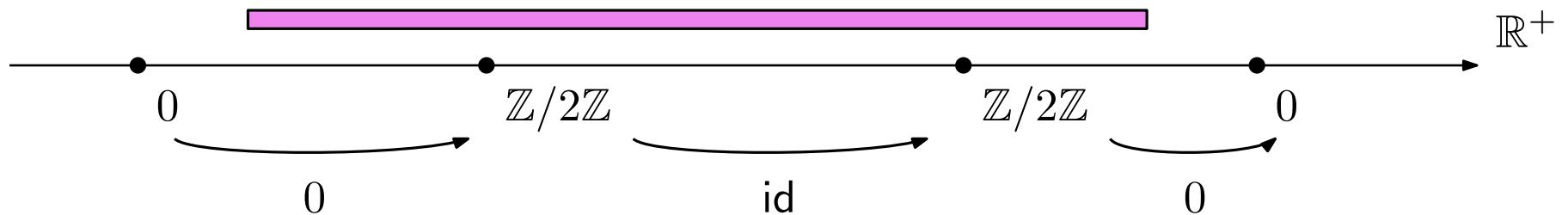
- 1 - Módulos de persistência
- 2 - Decomposição
- 3 - Estabilidade

V - Aplicações

Definição: Seja $I \subset \mathbb{R}^+$ um intervalo: $[a, b]$, $(a, b]$, $[a, b)$ ou (a, b) , com $a, b \in \mathbb{R}^+$ de modo que $a \leq b$, e com potencialmente $a = -\infty$ ou $b = +\infty$.

O **módulo-intervalo** associado a I é o módulo de persistência $\mathbb{B}[I]$ com espaços vetoriais $\mathbb{B}^t[I]$ e mapas lineares $v_s^t: \mathbb{B}^s[I] \rightarrow \mathbb{B}^t[I]$ definidos como

$$\mathbb{B}^t[I] = \begin{cases} \mathbb{Z}/2\mathbb{Z} & \text{se } t \in I, \\ 0 & \text{se não,} \end{cases} \quad \text{e} \quad v_s^t = \begin{cases} \text{id} & \text{se } s, t \in I, \\ 0 & \text{se não.} \end{cases}$$

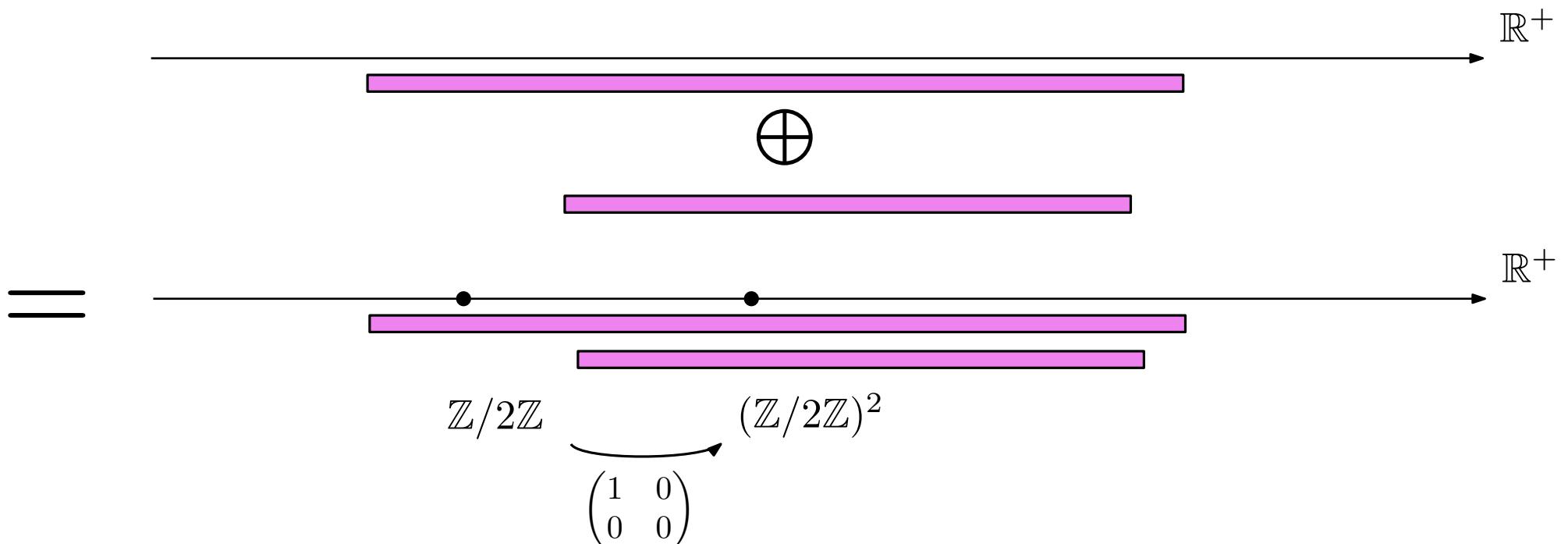


Definição: Seja $I \subset \mathbb{R}^+$ um intervalo: $[a, b]$, $(a, b]$, $[a, b)$ ou (a, b) , com $a, b \in \mathbb{R}^+$ de modo que $a \leq b$, e com potencialmente $a = -\infty$ ou $b = +\infty$.

O **módulo-intervalo** associado a I é o módulo de persistência $\mathbb{B}[I]$ com espaços vetoriais $\mathbb{B}^t[I]$ e mapas lineares $v_s^t: \mathbb{B}^s[I] \rightarrow \mathbb{B}^t[I]$ definidos como

$$\mathbb{B}^t[I] = \begin{cases} \mathbb{Z}/2\mathbb{Z} & \text{se } t \in I, \\ 0 & \text{se não,} \end{cases} \quad \text{e} \quad v_s^t = \begin{cases} \text{id} & \text{se } s, t \in I, \\ 0 & \text{se não.} \end{cases}$$

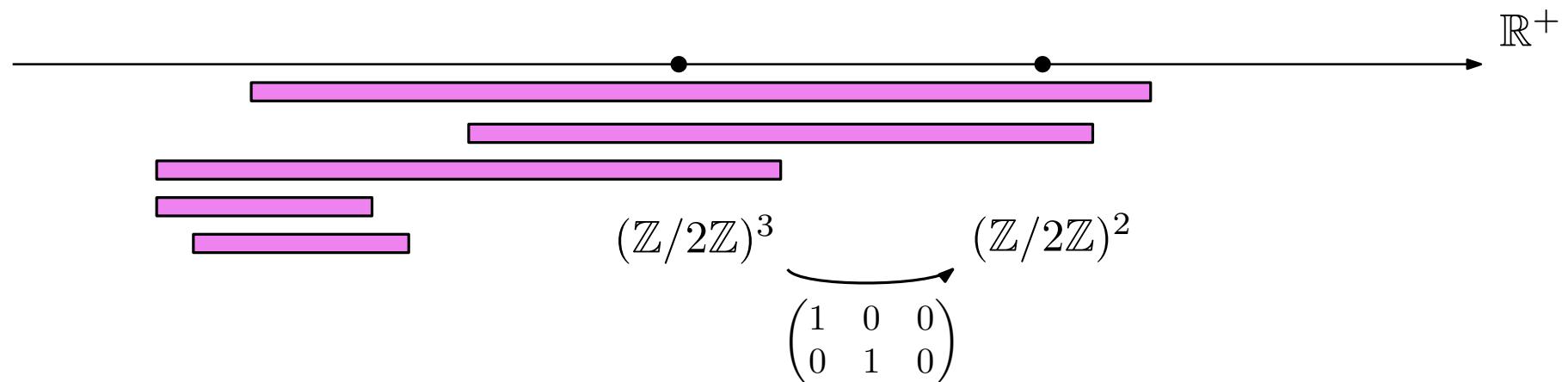
Podemos somar módulos-intervalo:



Definição: Seja $I \subset \mathbb{R}^+$ um intervalo: $[a, b]$, $(a, b]$, $[a, b)$ ou (a, b) , com $a, b \in \mathbb{R}^+$ de modo que $a \leq b$, e com potencialmente $a = -\infty$ ou $b = +\infty$.

O **módulo-intervalo** associado a I é o módulo de persistência $\mathbb{B}[I]$ com espaços vetoriais $\mathbb{B}^t[I]$ e mapas lineares $v_s^t: \mathbb{B}^s[I] \rightarrow \mathbb{B}^t[I]$ definidos como

$$\mathbb{B}^t[I] = \begin{cases} \mathbb{Z}/2\mathbb{Z} & \text{se } t \in I, \\ 0 & \text{se não,} \end{cases} \quad \text{e} \quad v_s^t = \begin{cases} \text{id} & \text{se } s, t \in I, \\ 0 & \text{se não.} \end{cases}$$



Um módulo de persistência \mathbb{V} **decompõe-se em módulos-intervalo** se existir um multiconjunto \mathcal{I} de intervalos de T tal que

$$\mathbb{V} \simeq \bigoplus_{I \in \mathcal{I}} \mathbb{B}[I].$$

Multiconjunto significa que \mathcal{I} pode conter várias cópias do mesmo intervalo I .

Teorema (consequência de Krull-Remak-Schmidt-Azumaya): Se um módulo de persistência se decompõe em módulos-intervalo, então o multiconjunto \mathcal{I} é único.

Nesse caso, \mathcal{I} é chamado de **barcode de persistência** de \mathbb{V} , e é denotado $\text{Barcode}(\mathbb{V})$.



Um módulo de persistência \mathbb{V} **decompõe-se em módulos-intervalo** se existir um multiconjunto \mathcal{I} de intervalos de T tal que

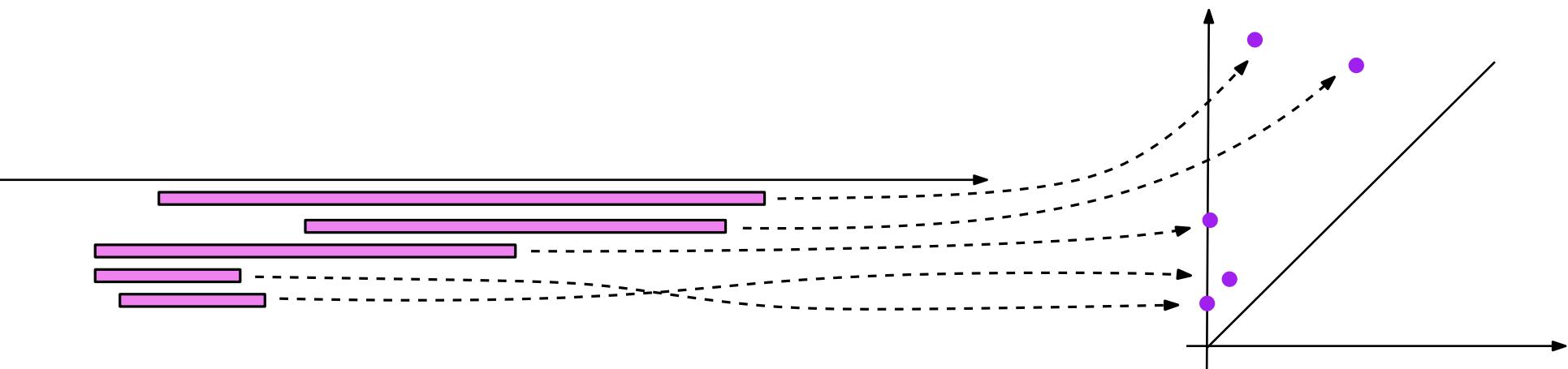
$$\mathbb{V} \simeq \bigoplus_{I \in \mathcal{I}} \mathbb{B}[I].$$

Multiconjunto significa que \mathcal{I} pode conter várias cópias do mesmo intervalo I .

Teorema (consequência de Krull-Remak-Schmidt-Azumaya): Se um módulo de persistência se decompõe em módulos-intervalo, então o multiconjunto \mathcal{I} é único.

Nesse caso, \mathcal{I} é chamado de **barcode de persistência** de \mathbb{V} , e é denotado $\text{Barcode}(\mathbb{V})$.

Para cada $[a, b]$, $(a, b]$, $[a, b)$ ou (a, b) em $\text{Barcode}(\mathbb{V})$, considere o ponto (a, b) de \mathbb{R}^2 . A coleção de todos esses pontos é o **diagrama de persistência** de \mathbb{V} .



Um módulo de persistência \mathbb{V} **decompõe-se em módulos-intervalo** se existir um multiconjunto \mathcal{I} de intervalos de T tal que

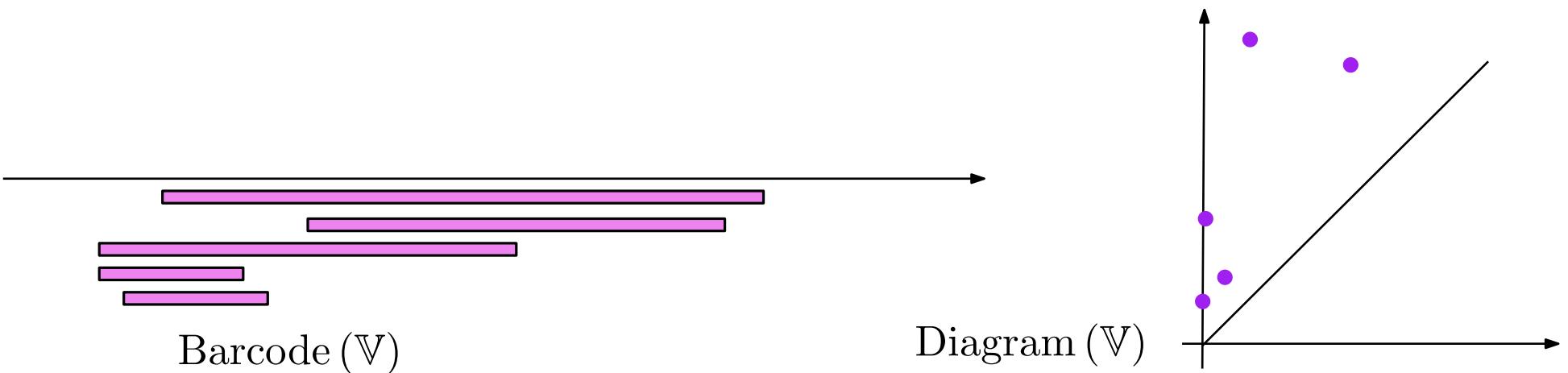
$$\mathbb{V} \simeq \bigoplus_{I \in \mathcal{I}} \mathbb{B}[I].$$

Multiconjunto significa que \mathcal{I} pode conter várias cópias do mesmo intervalo I .

Teorema (consequência de Krull-Remak-Schmidt-Azumaya): Se um módulo de persistência se decompõe em módulos-intervalo, então o multiconjunto \mathcal{I} é único.

Nesse caso, \mathcal{I} é chamado de **barcode de persistência** de \mathbb{V} , e é denotado $\text{Barcode}(\mathbb{V})$.

Para cada $[a, b]$, $(a, b]$, $[a, b)$ ou (a, b) em $\text{Barcode}(\mathbb{V})$, considere o ponto (a, b) de \mathbb{R}^2 . A coleção de todos esses pontos é o **diagrama de persistência** de \mathbb{V} .



Um módulo de persistência \mathbb{V} é **de dimensão pontualmente finita** se $\dim V^t < +\infty$ para todo t .

Teorema (Crawley-Boevey, 2015): Todo módulo de persistência de dimensão pontualmente finita se decompõe em módulos-intervalo.

Um módulo de persistência \mathbb{V} é **de dimensão pontualmente finita** se $\dim V^t < +\infty$ para todo t .

Teorema (Crawley-Boevey, 2015): Todo módulo de persistência de dimensão pontualmente finita se decompõe em módulos-intervalo.

Prova (Zomorodian, Carlsson, 2005): Caso mais simples: o módulo de persistência é de dimensão puntualmente finita e tem um número finito de termos.

Podemos escrever nosso módulo de persistência como

$$V^1 \xrightarrow{v_1^2} V^2 \xrightarrow{v_2^3} V^3 \xrightarrow{v_3^4} V^4 \longrightarrow \dots \longrightarrow V^n$$

Considere o espaço vetorial $\mathcal{V} = \bigotimes_{1 \leq i \leq n} V^i = V^1 \times \dots \times V^n$.

Seja $\mathbb{Z}/2\mathbb{Z}[x]$ o espaço de polinômios com coeficientes em $\mathbb{Z}/2\mathbb{Z}$. Dotamos \mathcal{V} de uma ação de $\mathbb{Z}/2\mathbb{Z}[x]$ por meio de

$$x \cdot (a^1, a^2, \dots, a^n) = (0, v_1^2(a^1), v_2^3(a^2), \dots, v_{n-1}^n(a^{n-1})).$$

Portanto, \mathcal{V} pode ser visto como um módulo gerado finitamente sobre o domínio principal $\mathbb{Z}/2\mathbb{Z}[x]$. Pela classificação, \mathcal{V} é isomorfo a uma soma

$$\mathcal{V} \simeq \bigoplus_{i \in I} \mathbb{Z}/2\mathbb{Z}[x]/(x^i \cdot \mathbb{Z}/2\mathbb{Z}[x]).$$

Identificamos os componentes $\mathbb{Z}/2\mathbb{Z}[x]/(x^i \cdot \mathbb{Z}/2\mathbb{Z}[x])$ com barras de comprimento i .

Em um código de barras, podemos ler a homologia **em cada etapa** e ver como ela **evolui**.

I - Comparar espaços topológicos

- 1 - Equivalência de homeomorfia
- 2 - Equivalência de homotopia

II - Invariante topológico

- 1 - Número de componentes conexas
- 2 - Característica de Euler
- 3 - Números de Betti
- 4 - Homologia (simplicial sobre $\mathbb{Z}/2\mathbb{Z}$)

III - Inferência topológica

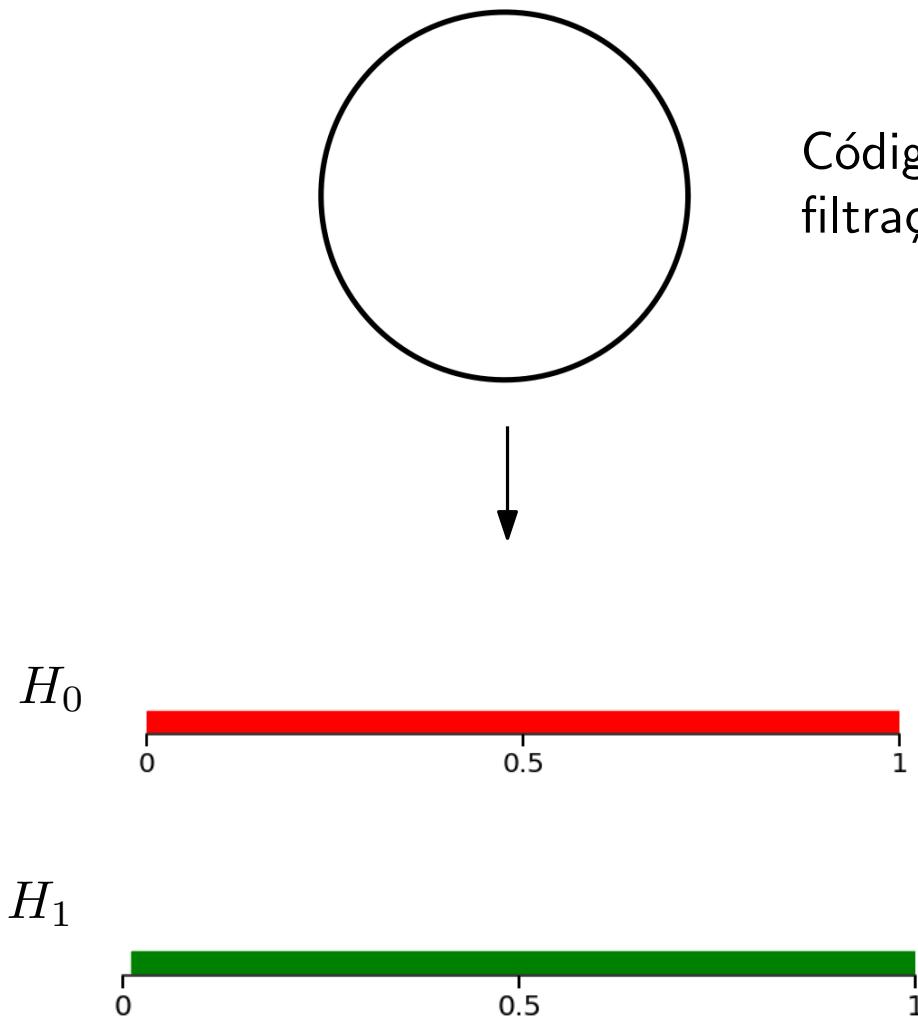
- 1 - Estimação do parâmetro t
- 2 - Nervos

IV - Homologia persistente

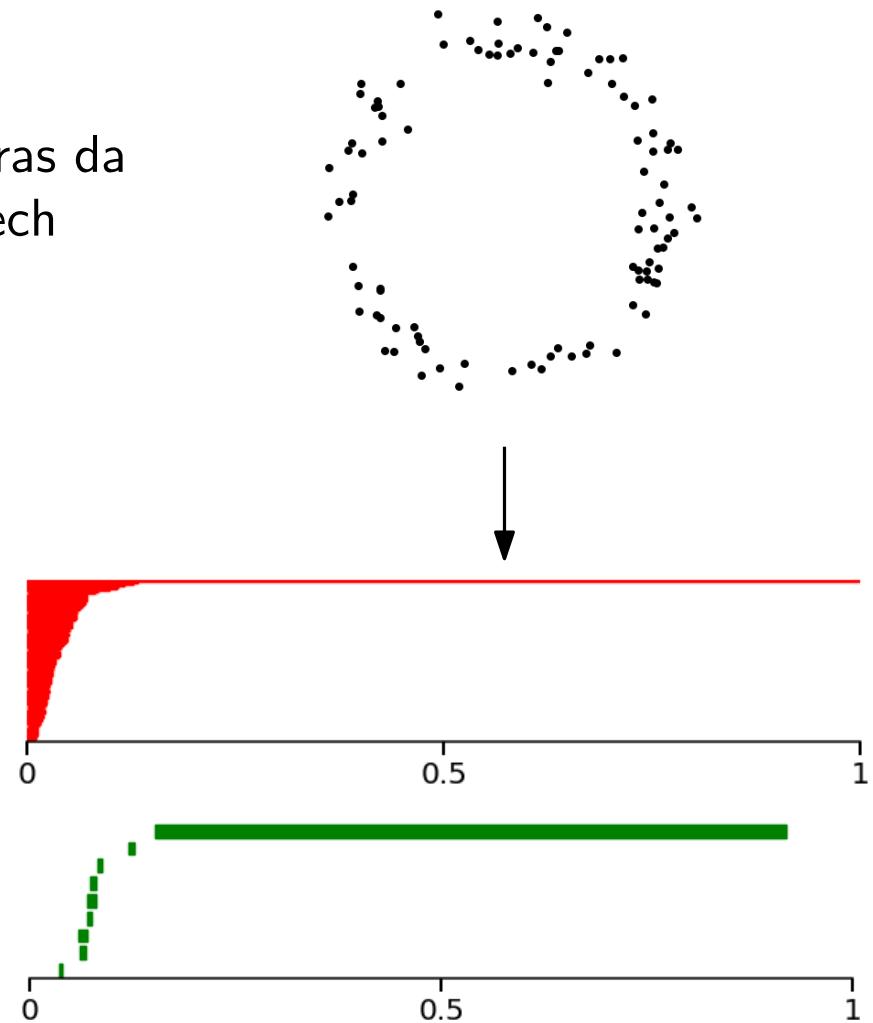
- 1 - Módulos de persistência
- 2 - Decomposição
- 3 - Estabilidade

V - Aplicações

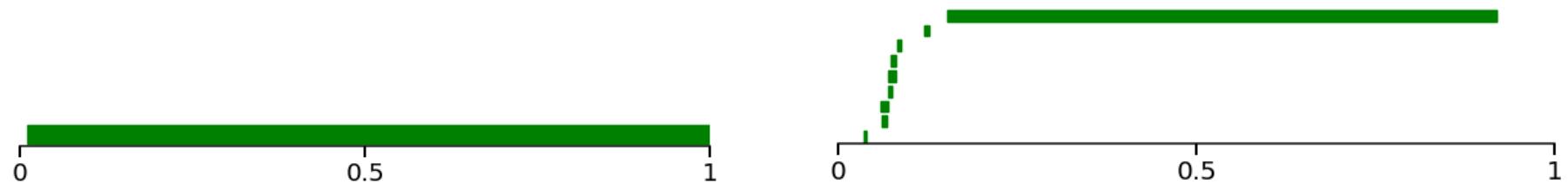
Seja $X \subset \mathbb{R}^n$ um subconjunto finito, visto como uma amostra de \mathcal{M} .



Código de barras da
filtração de Čech



Sejam dois códigos de barras P e Q .



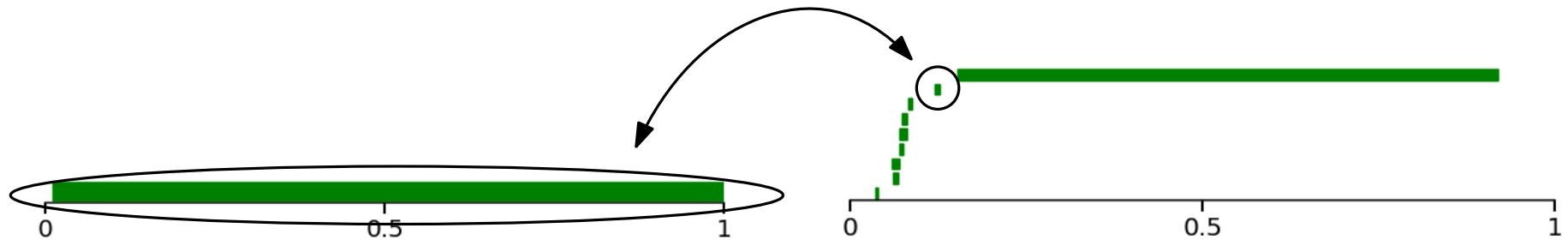
Uma **correspondência parcial** entre P e Q é um subconjunto $M \subset P \times Q$ tal como

- por todo $p \in P$, existe no máximo um $q \in Q$ tal como $(p, q) \in M$,
- por todo $q \in Q$, existe no máximo um $p \in P$ tal como $(p, q) \in M$.

As barras $p \in P$ (resp. $q \in Q$) tal como existe $q \in Q$ (resp. $p \in P$) com $(p, q) \in M$ são referidas como **correspondentes** por M .

Se uma barra $p \in P$ (resp. $q \in Q$) não for correspondida por M , digamos que é correspondida com o conjunto unitário $\bar{p} = [\frac{p_1+p_2}{2}, \frac{p_1+p_2}{2}]$ (resp. $\bar{q} = [\frac{q_1+q_2}{2}, \frac{q_1+q_2}{2}]$).

Sejam dois códigos de barras P e Q .



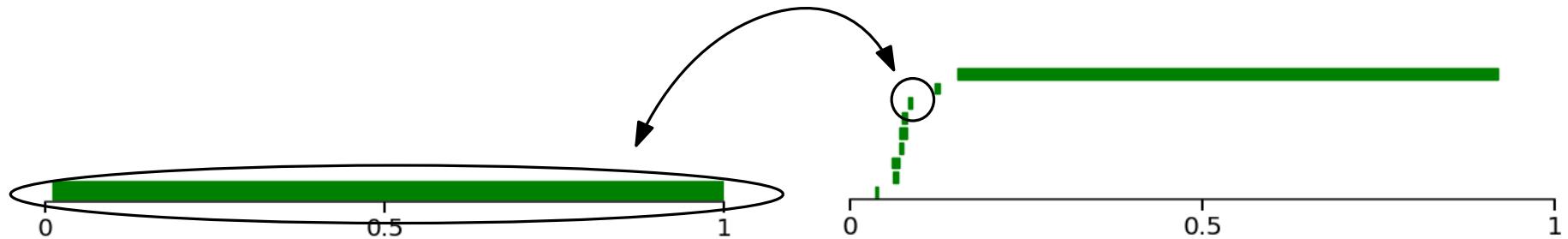
Uma **correspondência parcial** entre P e Q é um subconjunto $M \subset P \times Q$ tal como

- por todo $p \in P$, existe no máximo um $q \in Q$ tal como $(p, q) \in M$,
- por todo $q \in Q$, existe no máximo um $p \in P$ tal como $(p, q) \in M$.

As barras $p \in P$ (resp. $q \in Q$) tal como existe $q \in Q$ (resp. $p \in P$) com $(p, q) \in M$ são referidas como **correspondentes** por M .

Se uma barra $p \in P$ (resp. $q \in Q$) não for correspondida por M , digamos que é correspondida com o conjunto unitário $\bar{p} = [\frac{p_1+p_2}{2}, \frac{p_1+p_2}{2}]$ (resp. $\bar{q} = [\frac{q_1+q_2}{2}, \frac{q_1+q_2}{2}]$).

Sejam dois códigos de barras P e Q .



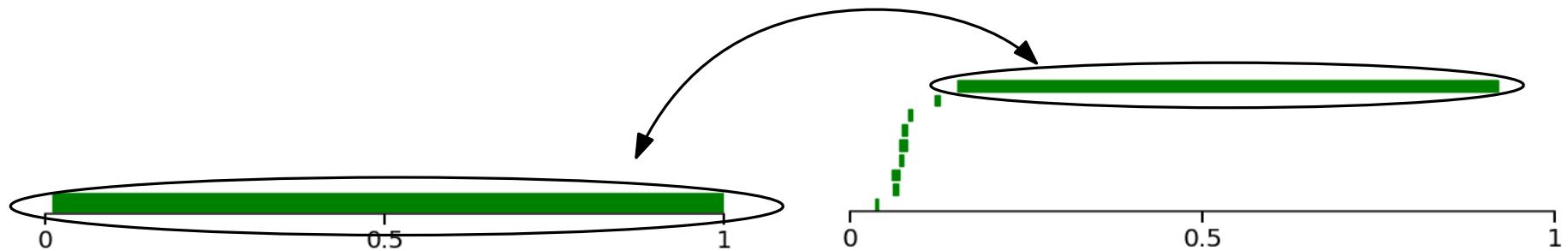
Uma **correspondência parcial** entre P e Q é um subconjunto $M \subset P \times Q$ tal como

- por todo $p \in P$, existe no máximo um $q \in Q$ tal como $(p, q) \in M$,
- por todo $q \in Q$, existe no máximo um $p \in P$ tal como $(p, q) \in M$.

As barras $p \in P$ (resp. $q \in Q$) tal como existe $q \in Q$ (resp. $p \in P$) com $(p, q) \in M$ são referidas como **correspondentes** por M .

Se uma barra $p \in P$ (resp. $q \in Q$) não for correspondida por M , digamos que é correspondida com o conjunto unitário $\bar{p} = \left[\frac{p_1+p_2}{2}, \frac{p_1+p_2}{2} \right]$ (resp. $\bar{q} = \left[\frac{q_1+q_2}{2}, \frac{q_1+q_2}{2} \right]$).

Sejam dois códigos de barras P e Q .



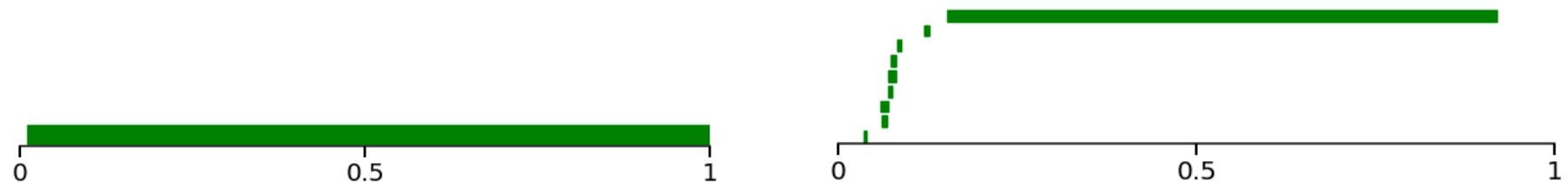
Uma **correspondência parcial** entre P e Q é um subconjunto $M \subset P \times Q$ tal como

- por todo $p \in P$, existe no máximo um $q \in Q$ tal como $(p, q) \in M$,
- por todo $q \in Q$, existe no máximo um $p \in P$ tal como $(p, q) \in M$.

As barras $p \in P$ (resp. $q \in Q$) tal como existe $q \in Q$ (resp. $p \in P$) com $(p, q) \in M$ são referidas como **correspondentes** por M .

Se uma barra $p \in P$ (resp. $q \in Q$) não for correspondida por M , digamos que é correspondida com o conjunto unitário $\bar{p} = \left[\frac{p_1+p_2}{2}, \frac{p_1+p_2}{2} \right]$ (resp. $\bar{q} = \left[\frac{q_1+q_2}{2}, \frac{q_1+q_2}{2} \right]$).

Sejam dois códigos de barras P e Q .



Uma **correspondência parcial** entre P e Q é um subconjunto $M \subset P \times Q$ tal como

- por todo $p \in P$, existe no máximo um $q \in Q$ tal como $(p, q) \in M$,
- por todo $q \in Q$, existe no máximo um $p \in P$ tal como $(p, q) \in M$.

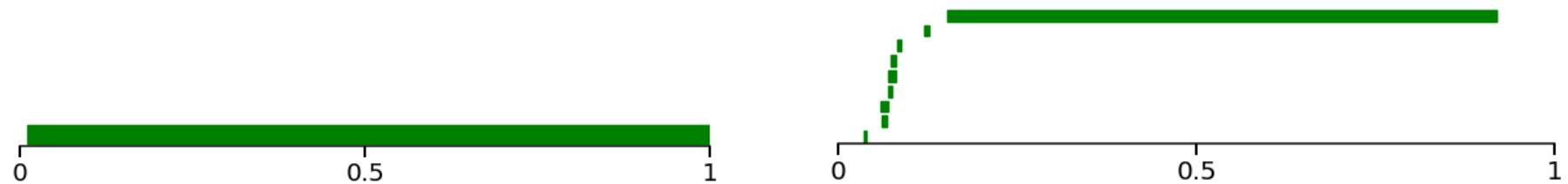
As barras $p \in P$ (resp. $q \in Q$) tal como existe $q \in Q$ (resp. $p \in P$) com $(p, q) \in M$ são referidas como **correspondentes** por M .

Se uma barra $p \in P$ (resp. $q \in Q$) não for correspondida por M , digamos que é correspondida com o conjunto unitário $\bar{p} = [\frac{p_1+p_2}{2}, \frac{p_1+p_2}{2}]$ (resp. $\bar{q} = [\frac{q_1+q_2}{2}, \frac{q_1+q_2}{2}]$).

O **custo** de um par (p, q) (resp. (p, \bar{p}) , resp. (\bar{q}, q)) é a norma sup $\|p - q\|_\infty = \sup\{|p_1 - q_1|, |p_2 - q_2|\}$ (resp. $\|p - \bar{p}\|_\infty$, resp. $\|\bar{q} - q\|_\infty$).

O **custo** da correspondência M , referido como $\text{cost}(M)$, é o supremo dos custos das pares.

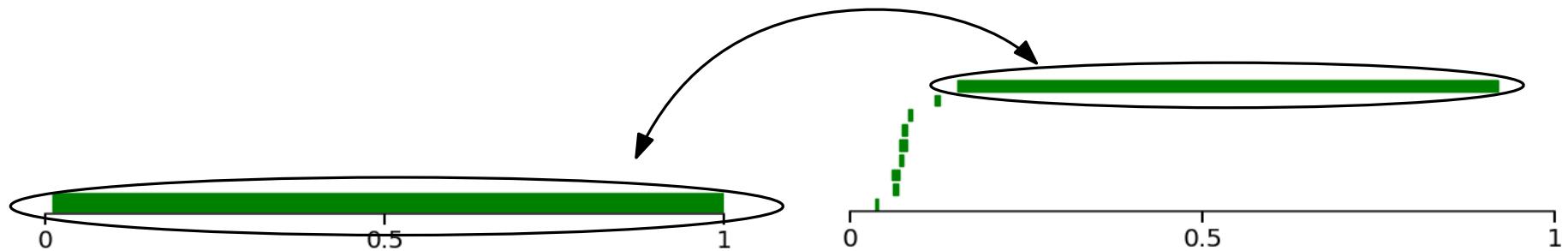
Sejam dois códigos de barras P e Q .



Definição : A **distância bottleneck** entre P e Q é definida como o menor custo de qualquer correspondência:

$$d_b(P, Q) = \inf\{\text{cost}(M), M \text{ é uma correspondência entre } P \text{ e } Q\}.$$

Sejam dois códigos de barras P e Q .



Definição : A **distância bottleneck** entre P e Q é definida como o menor custo de qualquer correspondência:

$$d_b(P, Q) = \inf\{\text{cost}(M), M \text{ é uma correspondência entre } P \text{ e } Q\}.$$

Teorema da estabilidade

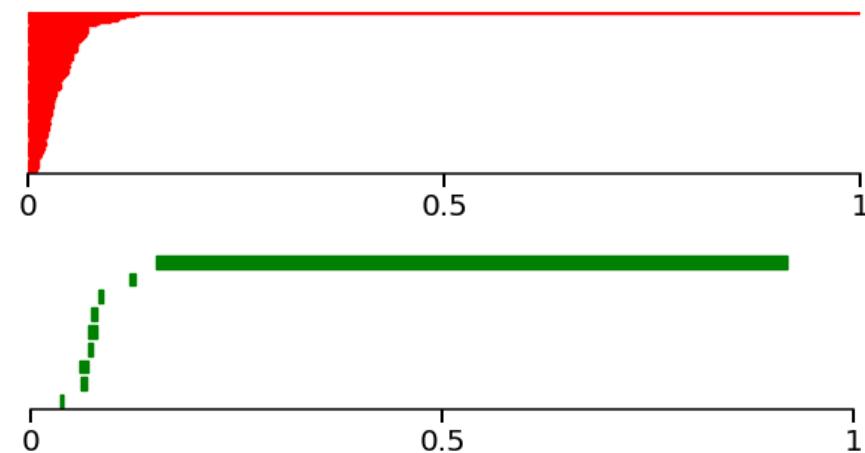
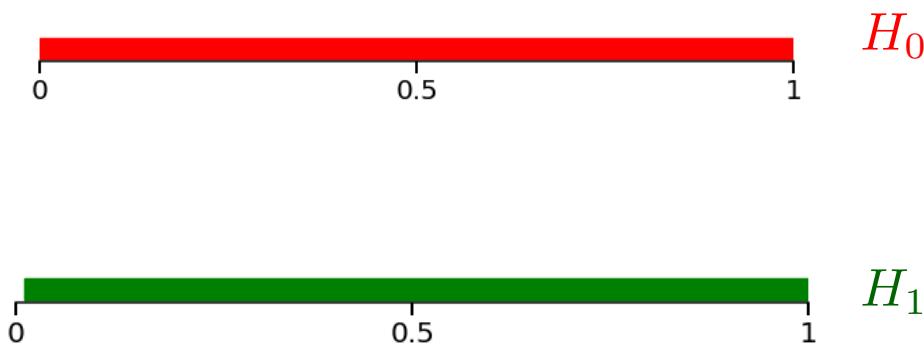
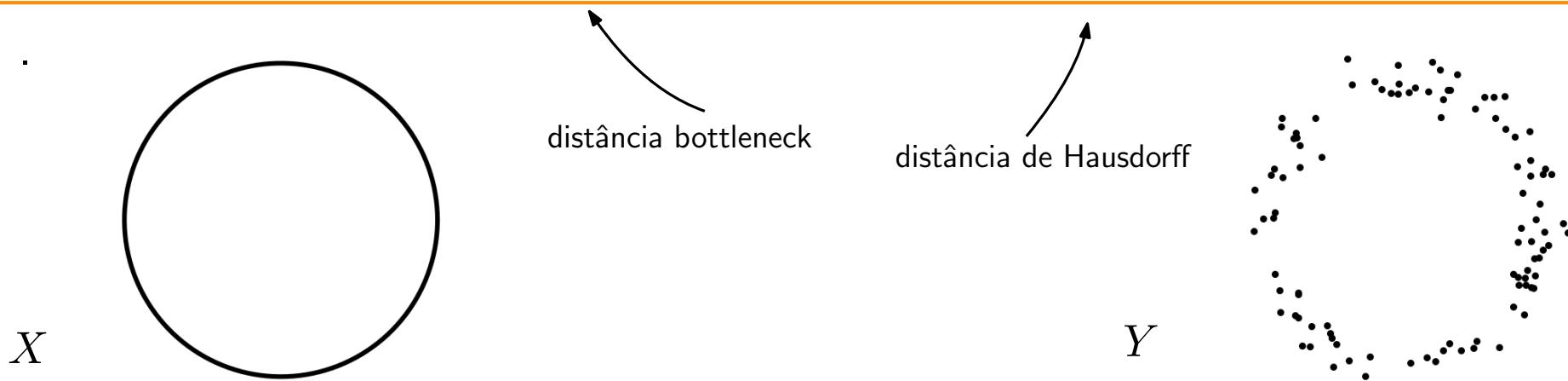
27/38 (1/2)

Teorema (Cohen-Steiner, Edelsbrunner et Harer, 2005):

Sejam X e Y dois subconjuntos do \mathbb{R}^n .

Considerem suas filtrações de Čech, e seus módulos de persistência de i -ésima homologia, notados \mathbb{U} e \mathbb{V} . Então

$$d_b(\text{Barcode}(\mathbb{U}), \text{Barcode}(\mathbb{V})) \leq d_H(X, Y)$$



Teorema da estabilidade

27/38 (2/2)

Teorema (Cohen-Steiner, Edelsbrunner et Harer, 2005):

Sejam X e Y dois subconjuntos do \mathbb{R}^n .

Considerem suas filtrações de Čech, e seus módulos de persistência de i -ésima homologia, notados \mathbb{U} e \mathbb{V} . Então

$$d_b(\text{Barcode}(\mathbb{U}), \text{Barcode}(\mathbb{V})) \leq d_H(X, Y)$$



Teorema (Chazal, de Silva, Glisse et Oudot, 2009) :

Vale

$$d_i(\mathbb{U}, \mathbb{V}) = d_b(\mathbb{U}, \mathbb{V})$$

onde d_i é a **distância de entrelaçamento**.

I - Comparar espaços topológicos

- 1 - Equivalência de homeomorfia
- 2 - Equivalência de homotopia

II - Invariante topológico

- 1 - Número de componentes conexas
- 2 - Característica de Euler
- 3 - Números de Betti
- 4 - Homologia (simplicial sobre $\mathbb{Z}/2\mathbb{Z}$)

III - Inferência topológica

- 1 - Estimação do parâmetro t
- 2 - Nervos

IV - Homologia persistente

- 1 - Módulos de persistência
- 2 - Decomposição
- 3 - Estabilidade

V - Aplicações

Inferência topológica I

29/38 (1/2)

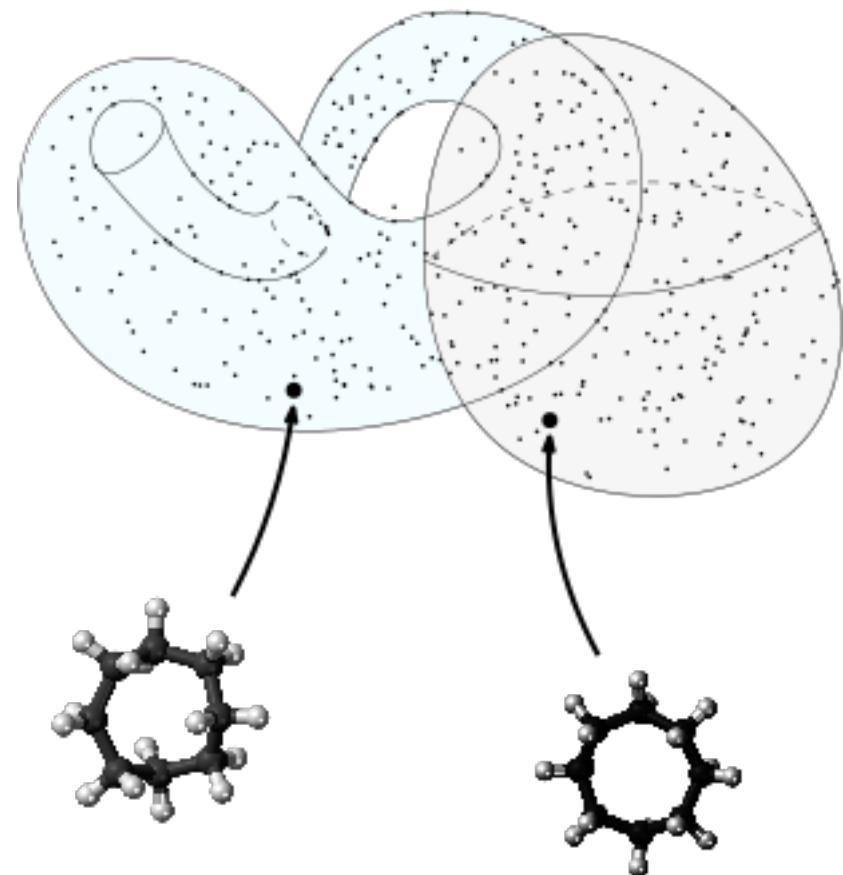
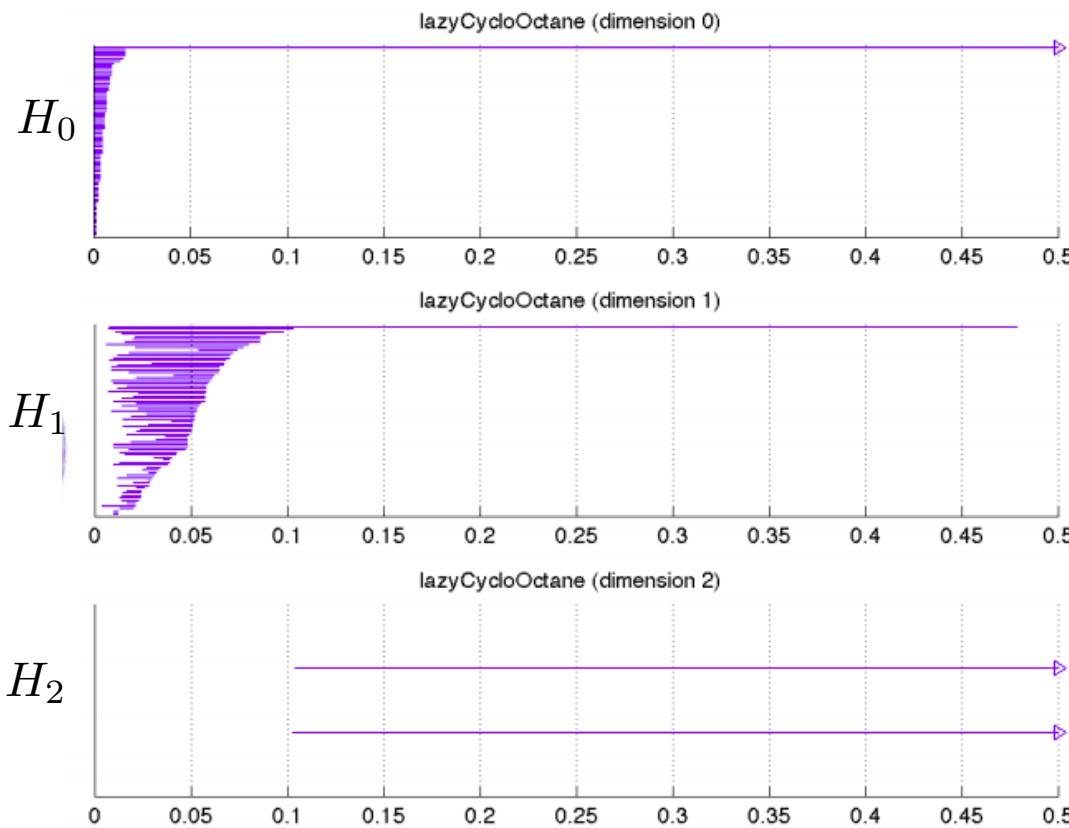
[S. Martin, A. Thompson, E. A. Coutsias, and J-P. Watson, [Topology of cyclo-octane energy landscape](#), 2010]

https://www.researchgate.net/publication/44697030_Topology_of_Cyclooctane_Energy_Landscape

A molécula de ciclo-octano C_8H_{16} contém 24 átomos.

Ao gerar muitas moléculas, obtemos uma nuvem de pontos em \mathbb{R}^{72} ($3 \times 24 = 72$).

Os códigos de barras são calculados:



Inferência topológica I

29/38 (2/2)

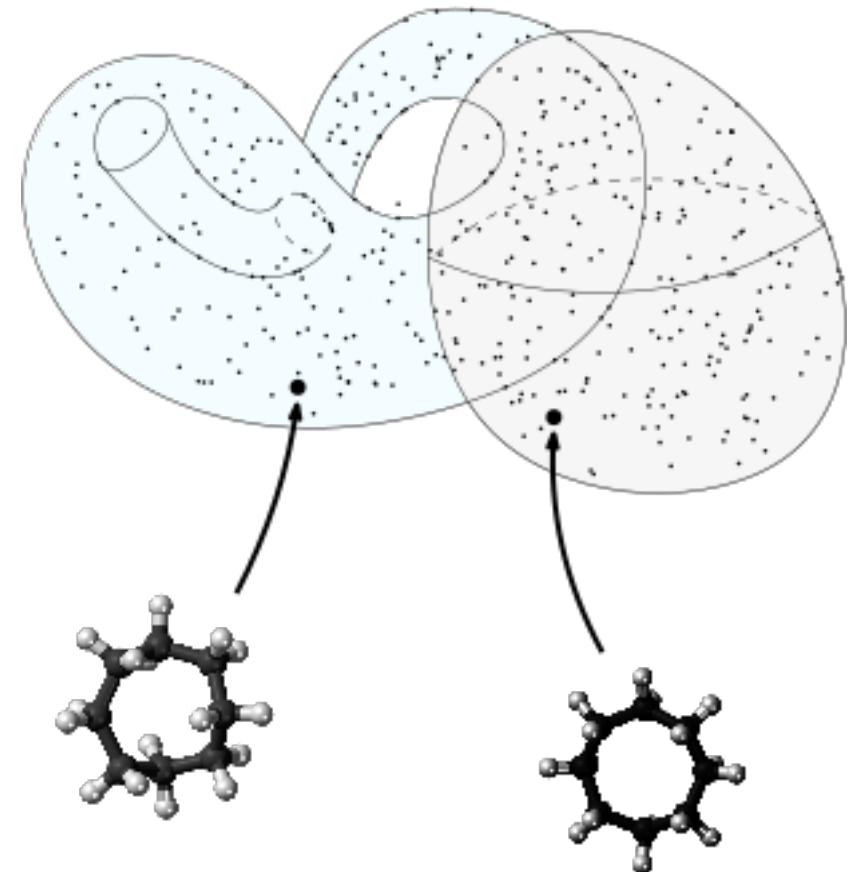
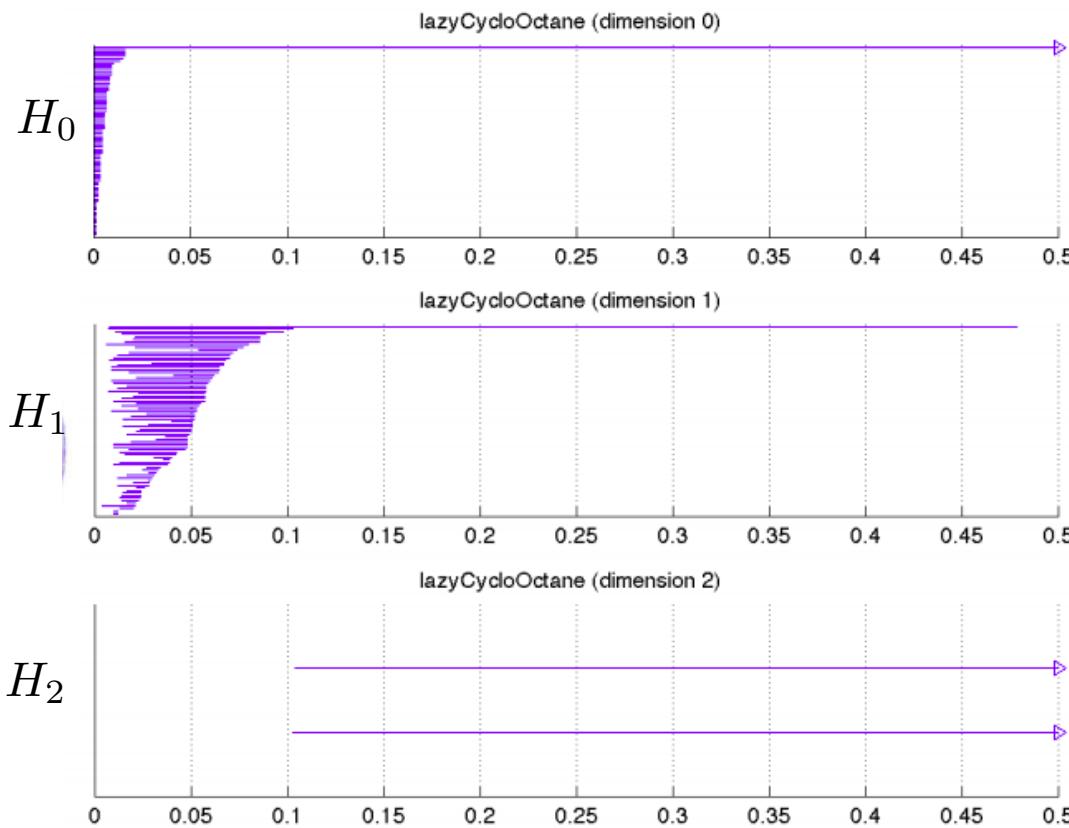
[S. Martin, A. Thompson, E. A. Coutsias, and J-P. Watson, [Topology of cyclo-octane energy landscape](#), 2010]

https://www.researchgate.net/publication/44697030_Topology_of_Cyclooctane_Energy_Landscape

A molécula de ciclo-octano C_8H_{16} contém 24 átomos.

Ao gerar muitas moléculas, obtemos uma nuvem de pontos em \mathbb{R}^{72} ($3 \times 24 = 72$).

Os códigos de barras são calculados:

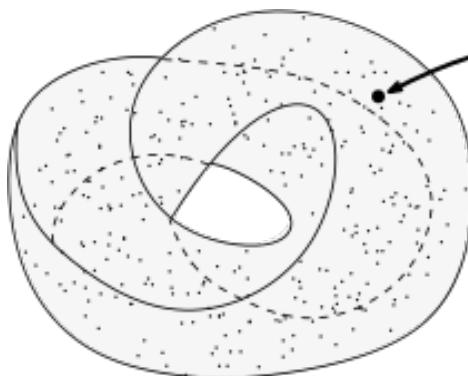


Deduzimos: $H_0 = \mathbb{Z}/2\mathbb{Z}$, $H_1 = \mathbb{Z}/2\mathbb{Z}$, $H_2 = (\mathbb{Z}/2\mathbb{Z})^2$

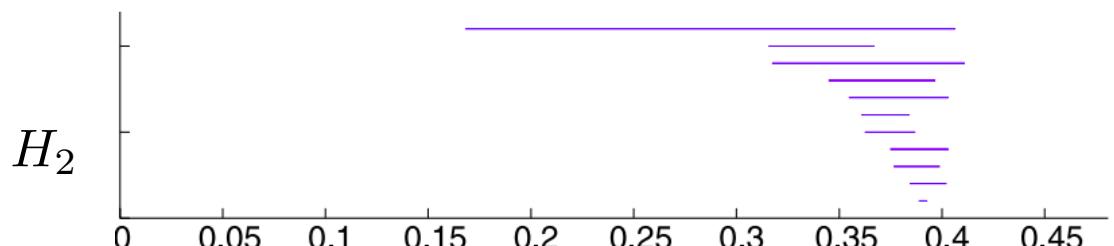
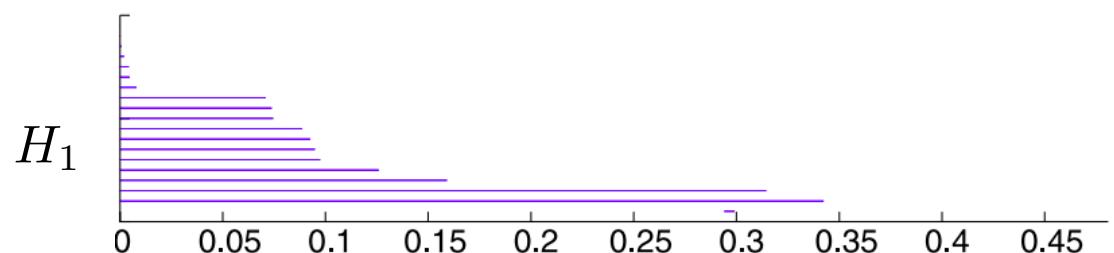
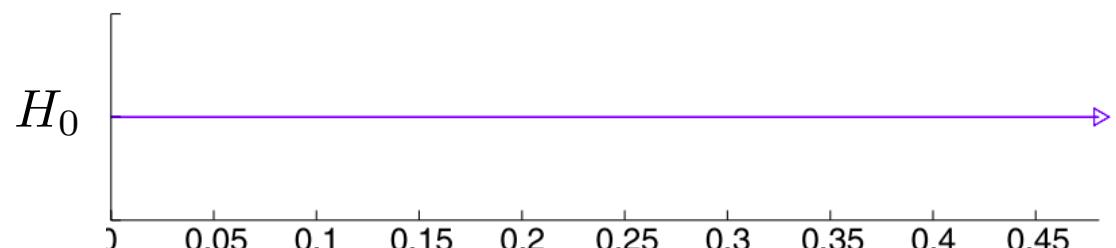
[G. Carlsson, T. Ishkhanov, V. de Silva, and A. Zomorodian, *On the Local Behavior of Spaces of Natural Images*, 2008]

<https://link.springer.com/article/10.1007/s11263-007-0056-x>

A partir de uma grande coleção de imagens, os autores extraem patches de tamanho 3×3 . Como consistem em 9 pixels, cada uma destes patches pode ser vista como um vetor em dimensão 9, e o todo como uma nuvem de pontos em \mathbb{R}^9 .



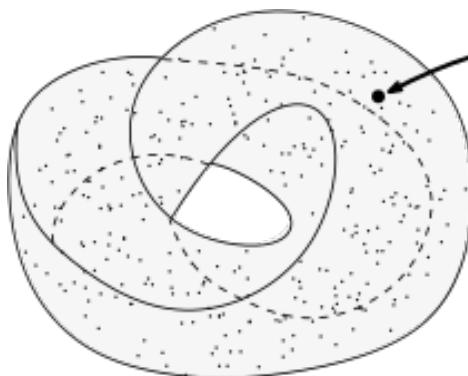
Os códigos de barras são obtidos:



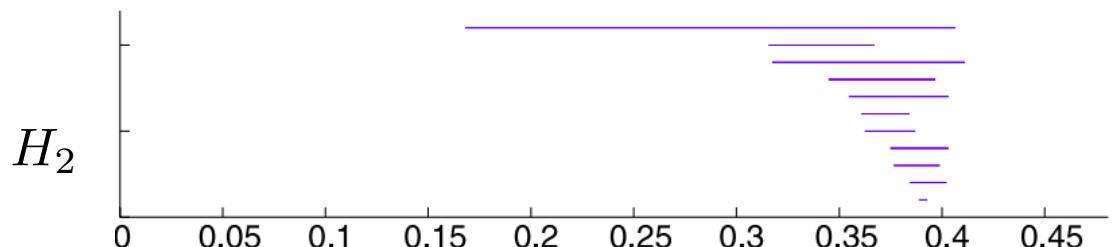
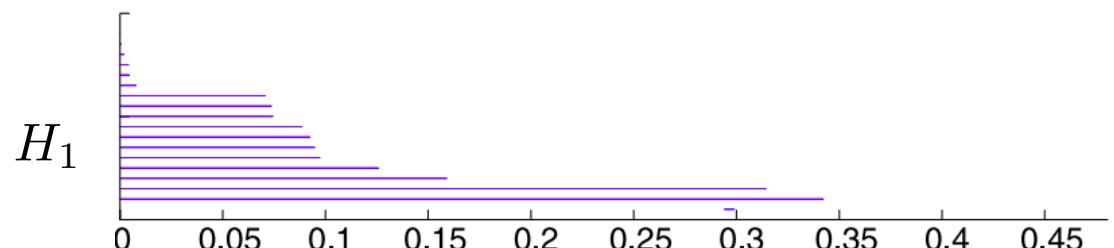
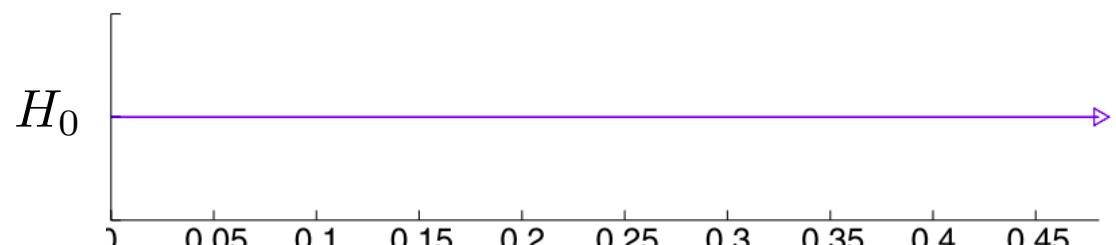
[G. Carlsson, T. Ishkhanov, V. de Silva, and A. Zomorodian, *On the Local Behavior of Spaces of Natural Images*, 2008]

<https://link.springer.com/article/10.1007/s11263-007-0056-x>

A partir de uma grande coleção de imagens, os autores extraem patches de tamanho 3×3 . Como consistem em 9 pixels, cada uma destes patches pode ser vista como um vetor em dimensão 9, e o todo como uma nuvem de pontos em \mathbb{R}^9 .



Os códigos de barras são obtidos:



Deduzimos:

$$H_0 = \mathbb{Z}/2\mathbb{Z},$$

$$H_1 = (\mathbb{Z}/2\mathbb{Z})^2,$$

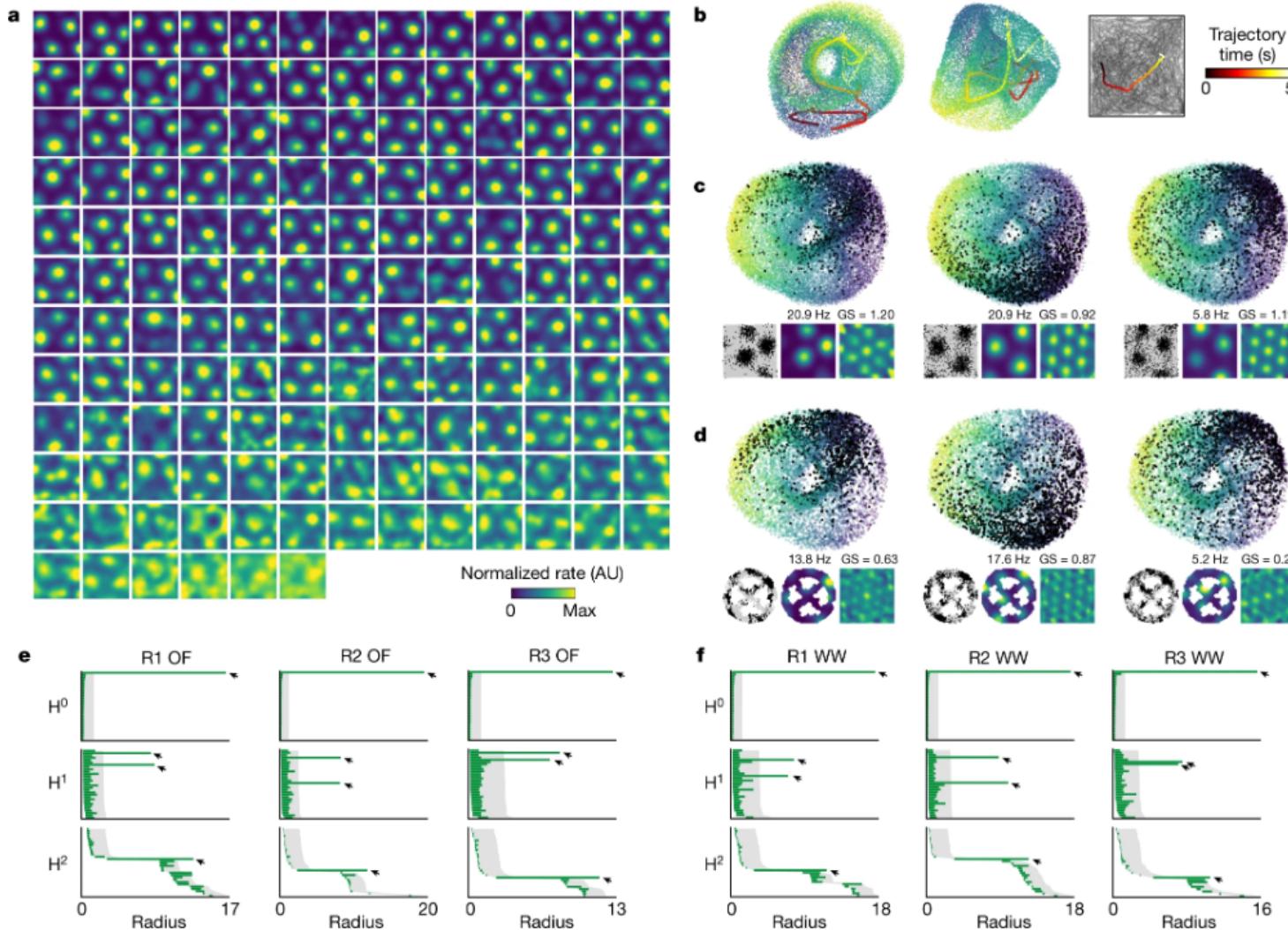
$$H_2 = \mathbb{Z}/2\mathbb{Z}.$$

Inferência topológica III

31/38

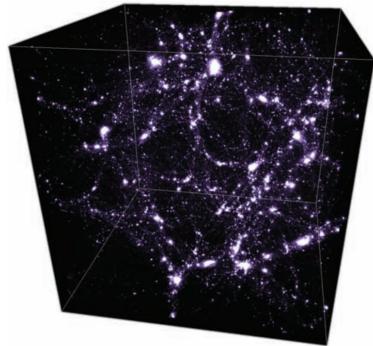
[Richard J. Gardner et al, Toroidal topology of population activity in grid cells , 2022]

Os autores registraram spikes de grid-cells de ratos, e aplicaram redução de dimensionalidade na matriz de firing. Ao aplicar a homologia persistente a esta nuvem de pontos, observamos a homologia de um toro.

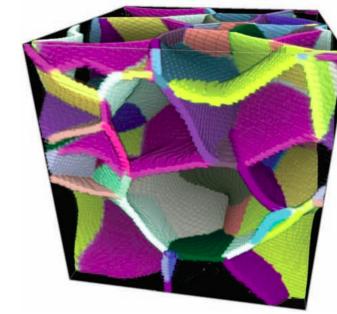


T. Sousbie, [The persistent cosmic web and its filamentary structure](#), 2011

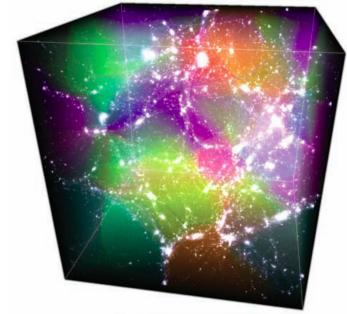
<https://www.giss.nasa.gov/staff/mway/cluster/sousbie2011mnras.pdf>



visto como um
objeto de dimensão 1



de dimensão 2



de dimensão 3

P. Pranav, H. Edelsbrunner, R. de Weygaert, G. Vegter, M. Kerber, B. Jones and M. Wintraecken, [The topology of the cosmic web in terms of persistent Betti numbers](#), 2016

<https://arxiv.org/pdf/1608.04519.pdf>

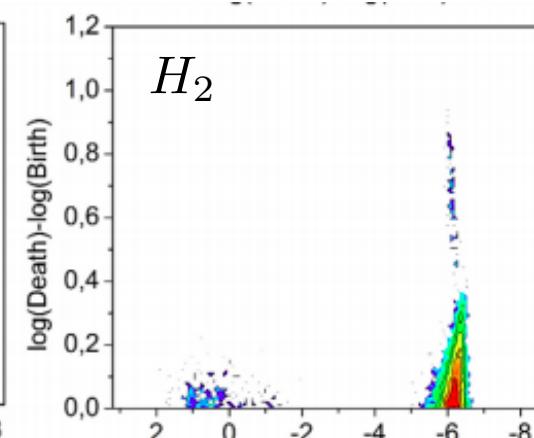
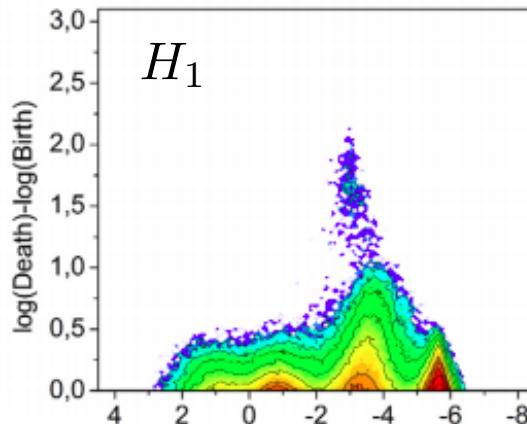
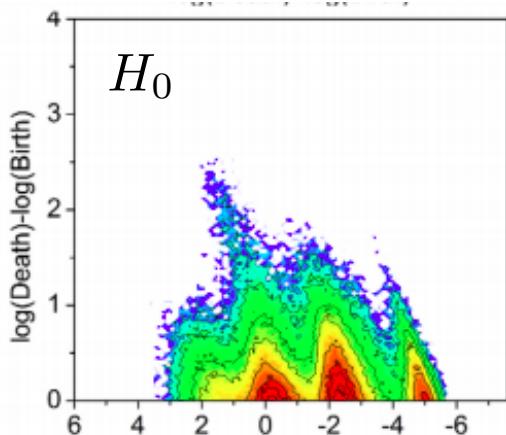
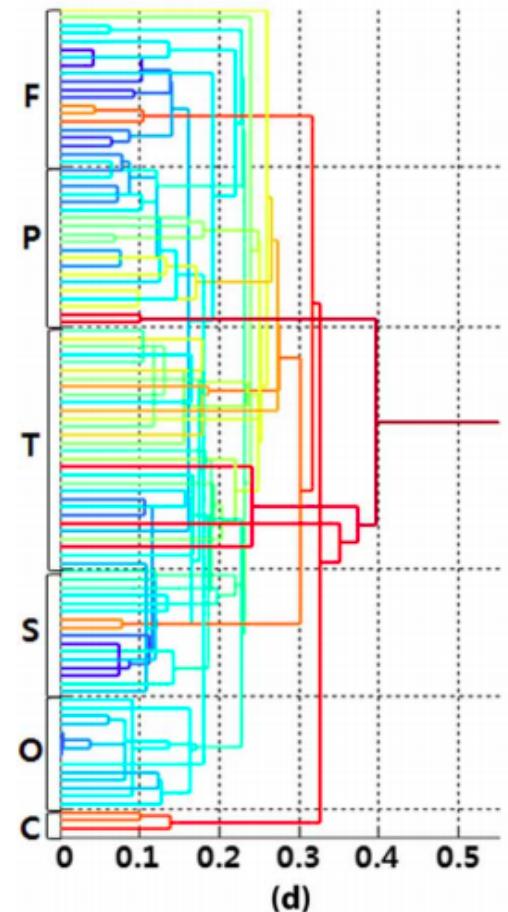
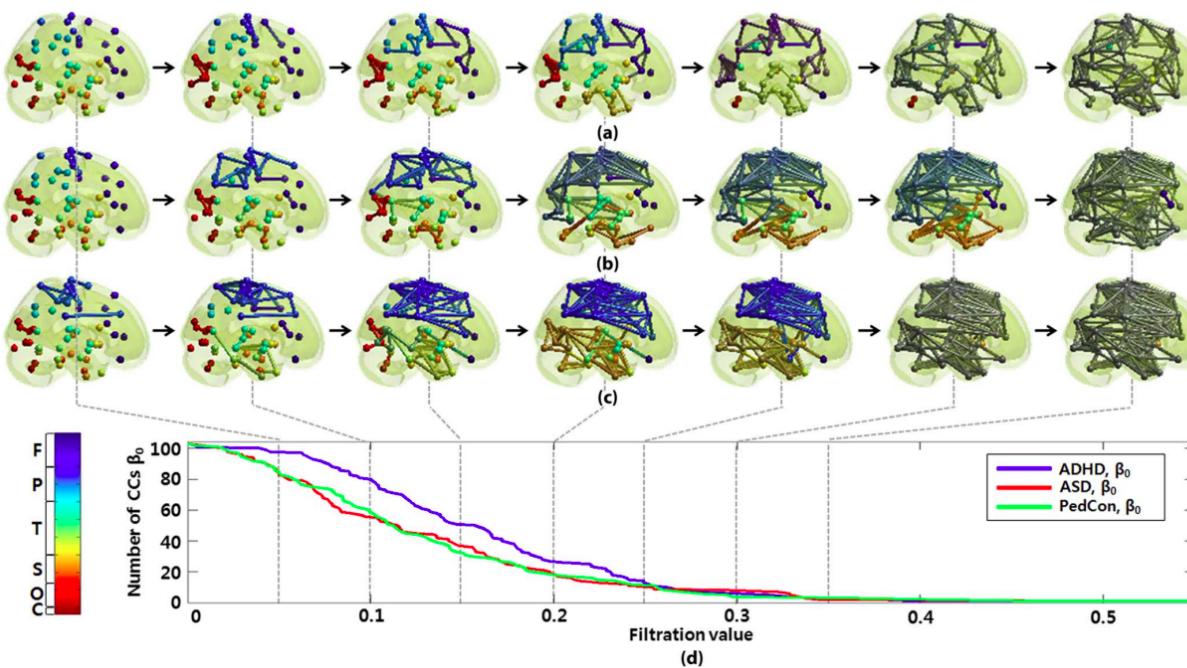


Diagrama de
persistência média
para um modelo
Voronoi
(escala log)

Hyekyoung Lee, Hyejin Kang, Moo K Chung, Bung-Nyun Kim, Dong Soo Lee,
 Persistent brain network homology from the perspective of dendrogram, 2012

<http://pages.stat.wisc.edu/~mchung/papers/lee.2012.TMI.pdf>

→ A H_0 -homologia persistente induz uma classificação hierárquica



Mathieu Carrière, Marco Cuturi, Steve Oudot, Sliced Wasserstein Kernel for Persistence Diagrams, 2017

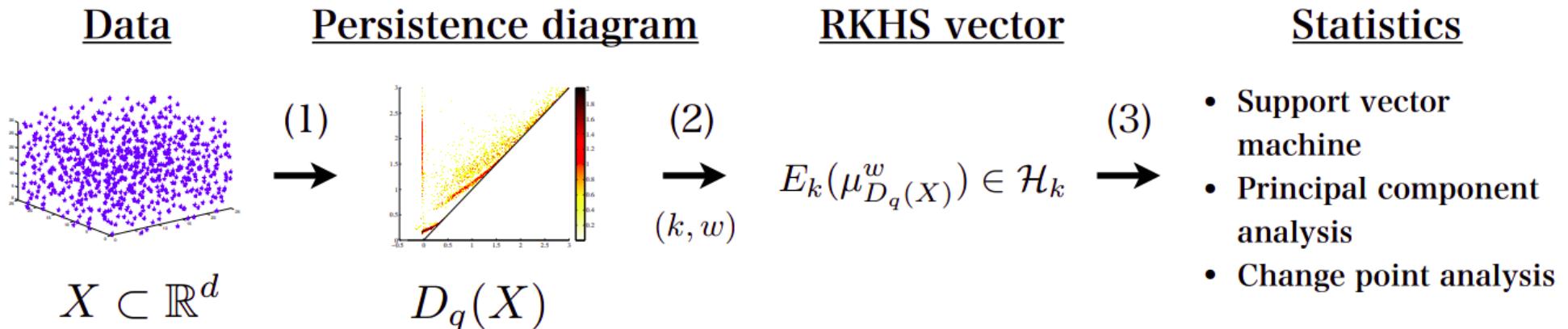
<https://arxiv.org/abs/1706.03358>

Genki Kusano, Kenji Fukumizu, Yasuaki Hiraoka, Kernel Method for Persistence Diagrams via Kernel Embedding and Weight Factor, 2018

<https://www.jmlr.org/papers/volume18/17-317/17-317.pdf>

Os códigos de barras não são elementos de um espaço euclidiano e, portanto, não podem ser usados diretamente

→ podemos usar o **kernel trick**



No machine learning II

35/38

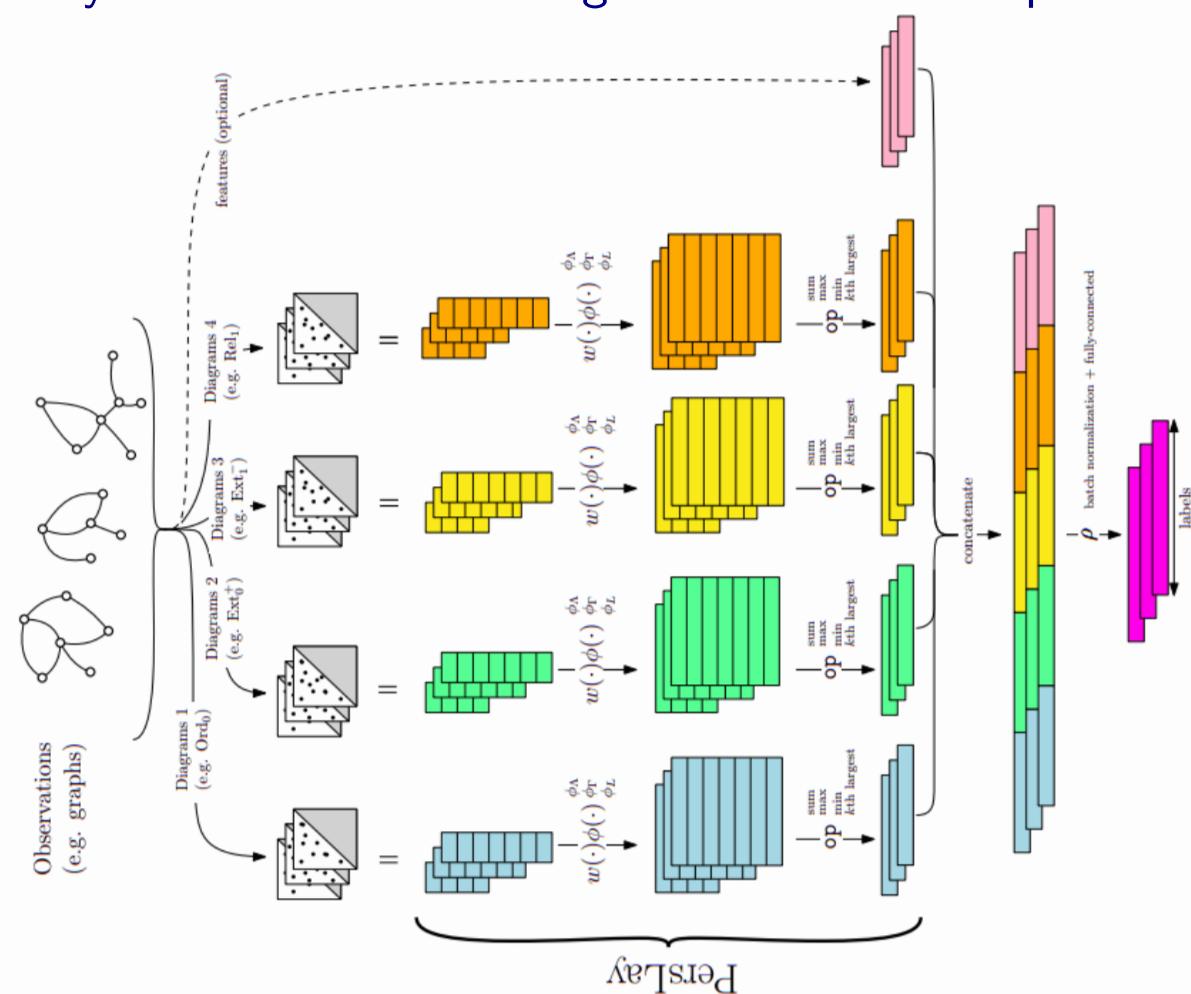
Rickard Brüel-Gabrielsson, Bradley J. Nelson, Anjan Dwaraknath, Primoz Skraba, Leonidas J. Guibas, Gunnar Carlsson, [A Topology Layer for Machine Learning](#), 2019

<https://arxiv.org/abs/1905.12200>

Mathieu Carrière, Frédéric Chazal, Yuichi Ike, Théo Lacombe, Martin Royer, Yuhei Umeda, [PersLay: A Neural Network Layer for Persistence Diagrams and New Graph Topological Signatures](#), 2019

<https://arxiv.org/abs/1904.09378>

Os diagramas de persistência podem ser integrados com uma rede neural

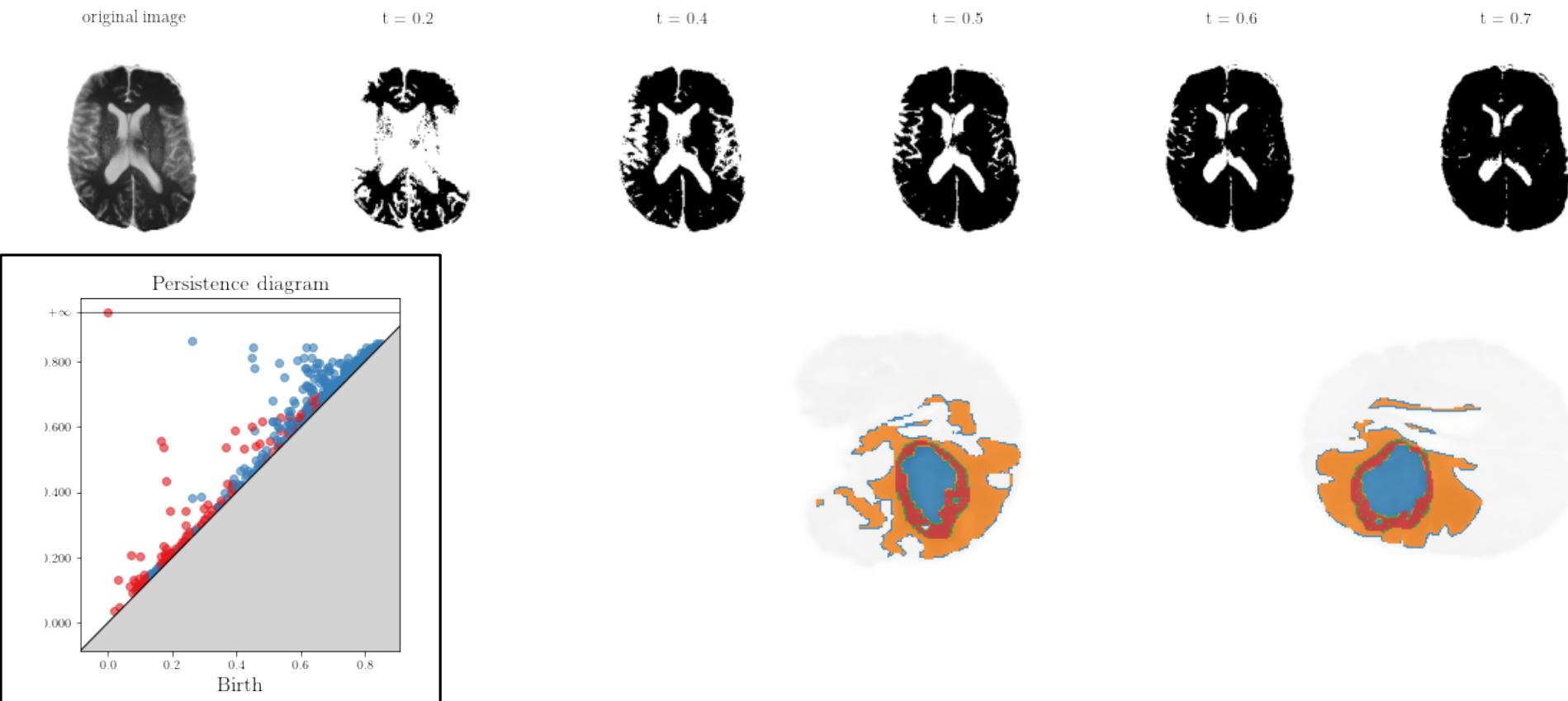


Em colaboração com Anton François.

O **glioblastoma** é o tumor cerebral mais comum, difuso, de grau variável de agressividade, e cujo prognóstico médico é difícil de estabelecer.

Neste contexto, o problema da **segmentação** consiste na demarcação automática das três regiões que formam o tumor (edema, núcleo necrótico e enhancing tumor).

Podemos usar a **homologia persistente cúbica**, especialmente definida para imagens.



Frédéric Chazal, Steve Oudot, Primoz Skraba, Leonidas J. Guibas, Persistence-Based Clustering in Riemannian Manifolds, 2011

<https://geometrica.saclay.inria.fr/team/Fred.Chazal/papers/cgos-pbc-09/cgos-pbcrm-11.pdf>

Chunyuan Li, Maks Ovsjanikov, Frederic Chazal, Persistence-based Structural Recognition, 2014

<https://geometrica.saclay.inria.fr/team/Fred.Chazal/papers/loc-pbsr-14/CVPR2014.pdf>

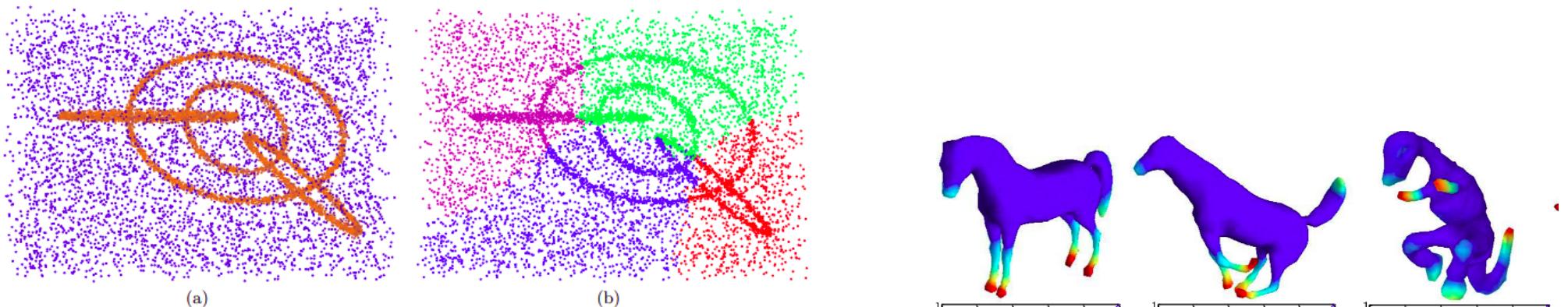
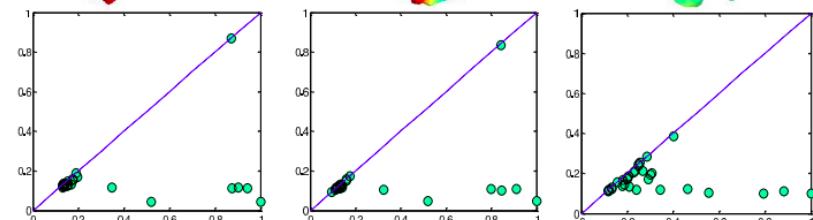
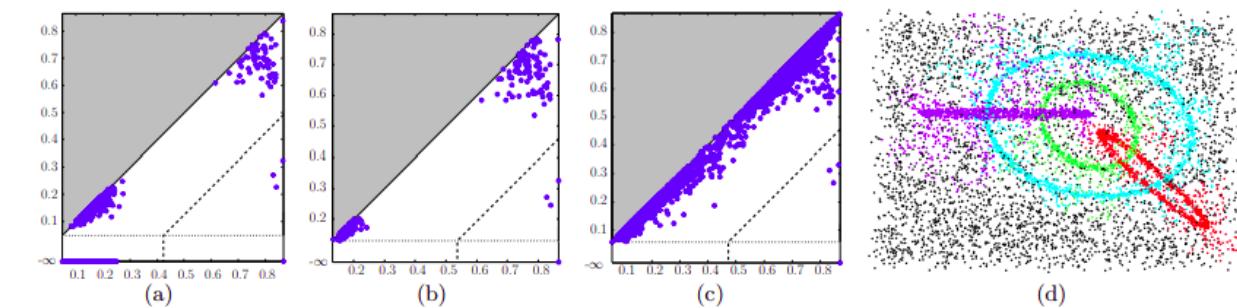
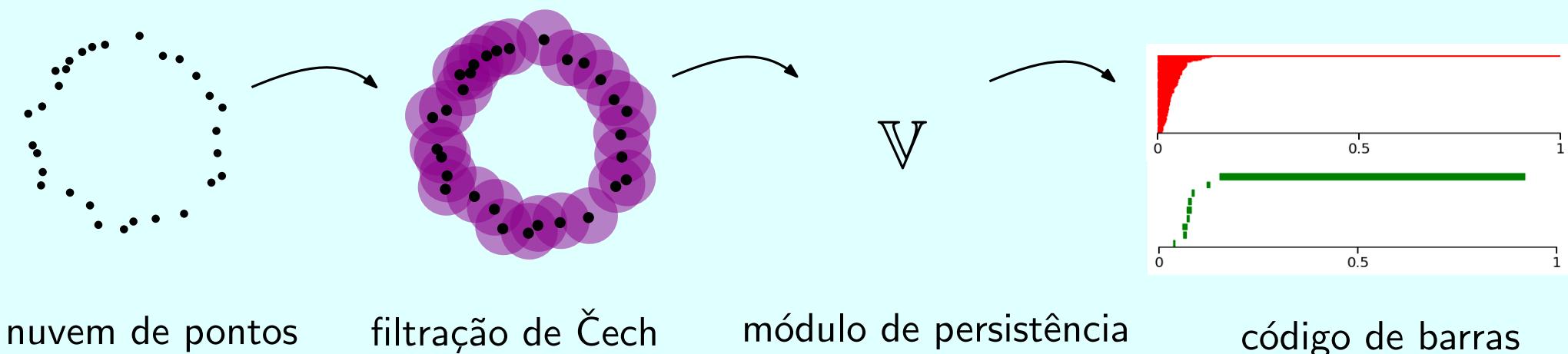


Figure 7: (a) The rings data set with the estimated density function. (b) The result obtained using spectral clustering.



Conclusão

A homologia persistente permite uma estimativa **multi-escala** e **estável** da homologia dos conjuntos de dados.



Permite analisar os dados a partir de uma nova perspectiva.

Um curso de TDA: <https://raphaeltinarrage.github.io>