

EMINES — 09/06/21

# Analyse Topologique des Données

# التحليل الطوبولوجي للبيانات

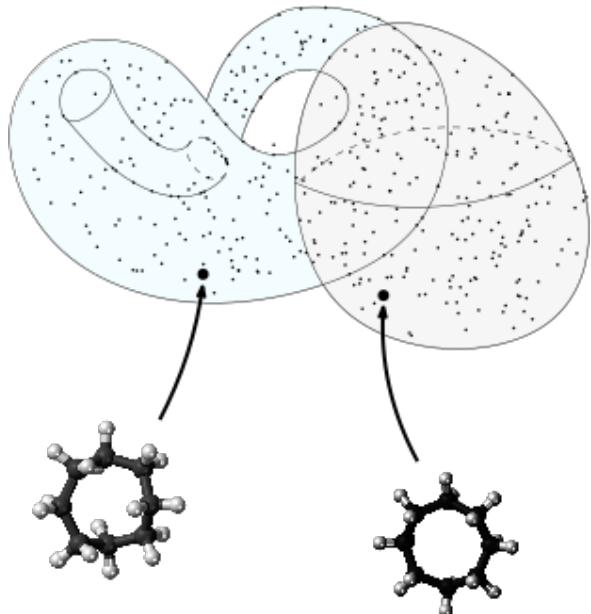
Séance II/II : Homologie persistante

<https://raphaeltinarrage.github.io>

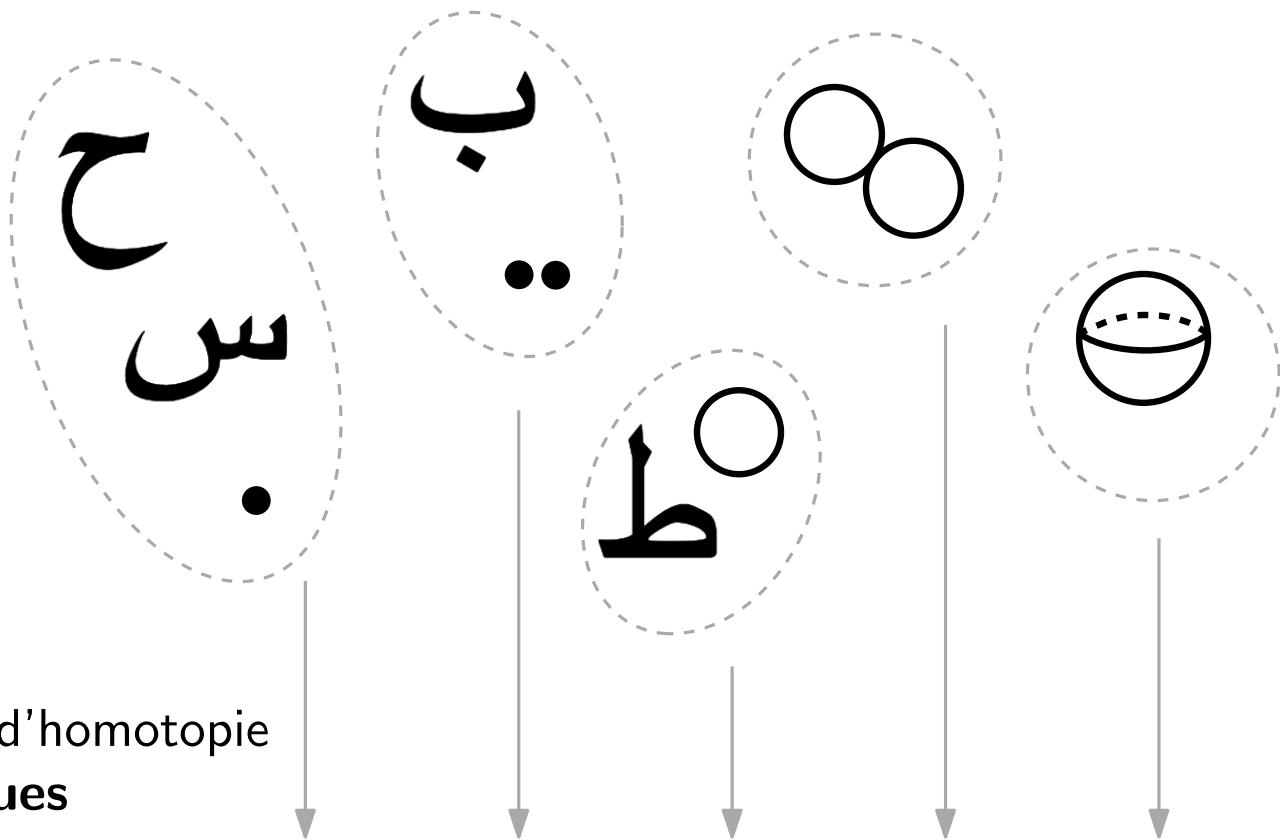
# Rappel de la première séance

2/43

- Les jeux de données peuvent avoir une topologie intéressante



- On peut classer les espaces topologiques en **classes d'homotopie**

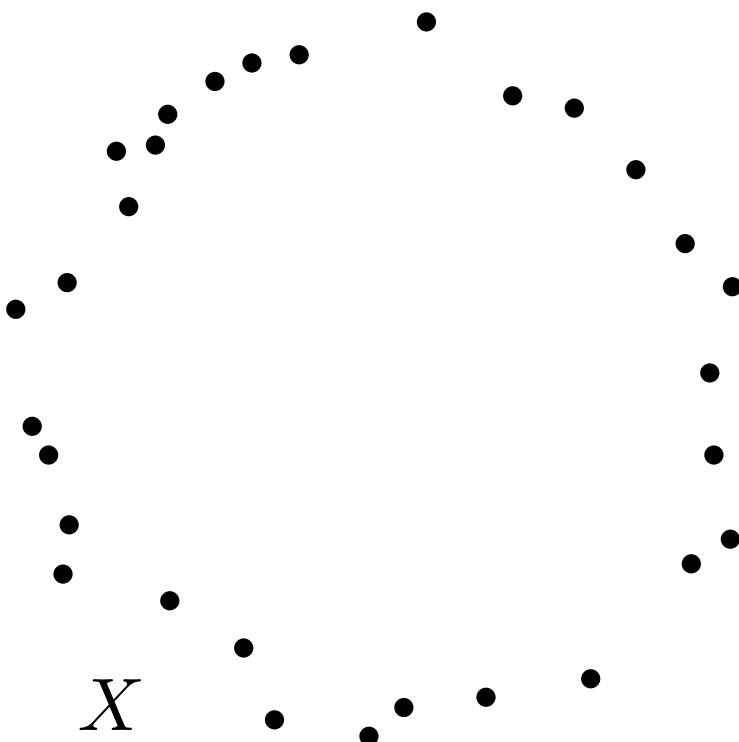


- On peut comprendre les classes d'homotopie grâce aux **invariants topologiques**

Nombres de Betti	$\beta_0(X)$	1	2	1	1	1
	$\beta_1(X)$	0	0	1	2	0
	$\beta_2(X)$	0	0	0	0	1

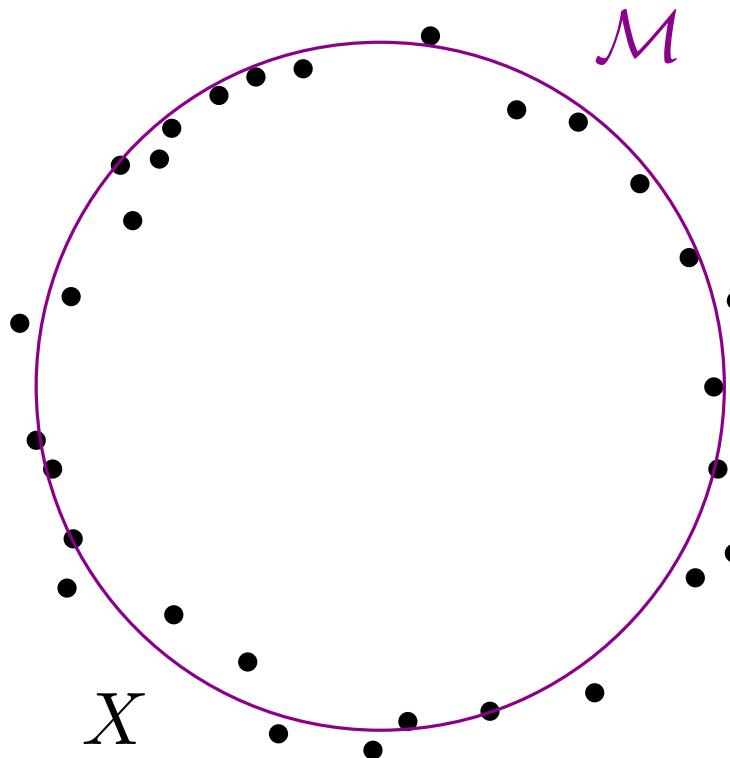
Dans pratique, on a souvent affaire à des sous-ensembles de l'espace euclidien :  $X \subset \mathbb{R}^n$ .

Bien sûr,  $X$  est fini.



Dans pratique, on a souvent affaire à des sous-ensembles de l'espace euclidien :  $X \subset \mathbb{R}^n$ .

Bien sûr,  $X$  est fini.



En Analyse Topologique des Données, on imagine que  $X$  a été échantilloné sur **un objet sous-jacent** continu,  $\mathcal{M} \subset \mathbb{R}^n$ .

Comprendre la topologie de  $\mathcal{M}$  permettrait de mieux comprendre notre jeux de données.

# I - Homologie simpliciale

≈ 15 min

1 - Groupes d'homologie

2 - Fonctorialité

# II - Inférence topologique

≈ 10 min

1 - Estimation du paramètre

2 - Nerfs

# III - Homologie persistante

≈ 20 min

1 - Modules de persistance

2 - Décomposition

3 - Stabilité

# IV - Applications

≈ 10 min

# Programme

5/43 (1/2)

Nous allons définir l'**homologie simpliciale** sur le corps  $\mathbb{Z}/2\mathbb{Z}$ .

basé sur les complexes simpliciaux

nous allons faire de l'algèbre linéaire sur  $\mathbb{Z}/2\mathbb{Z}$

Nous devons définir :

- les chaines,
- les opérateurs de bord,
- les cycles et les bords,
- les groupes d'homologie.

Le groupe  $\mathbb{Z}/2\mathbb{Z}$  peut être vu comme l'ensemble  $\{0, 1\}$  muni de l'opération

$$0 + 0 = 0$$

$$0 + 1 = 1$$

$$1 + 0 = 1$$

$$1 + 1 = 0$$

Pour tout  $n \geq 1$ , le **groupe produit**  $(\mathbb{Z}/2\mathbb{Z})^n$  est le groupe dont l'ensemble sous-jacent est

$$(\mathbb{Z}/2\mathbb{Z})^n = \{(\epsilon_1, \dots, \epsilon_n), \epsilon_1, \dots, \epsilon_n \in \mathbb{Z}/2\mathbb{Z}\}$$

et l'opération de groupe est définie par

$$(\epsilon_1, \dots, \epsilon_n) + (\epsilon'_1, \dots, \epsilon'_n) = (\epsilon_1 + \epsilon'_1, \dots, \epsilon_n + \epsilon'_n).$$

Le groupe  $\mathbb{Z}/2\mathbb{Z}$  peut être muni d'une structure de **corps**

$$0 \times 0 = 0$$

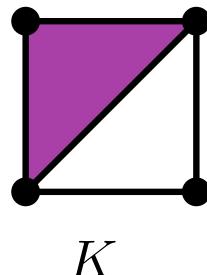
$$0 \times 1 = 0$$

$$1 \times 0 = 0$$

$$1 \times 1 = 1$$

et le produit  $(\mathbb{Z}/2\mathbb{Z})^n$  peut être vu comme un **espace vectoriel** sur le corps  $\mathbb{Z}/2\mathbb{Z}$ .

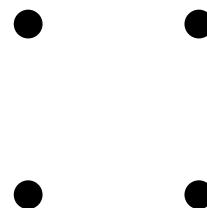
**Définition (rappel) :** Soit  $V$  un ensemble (appelé l'ensemble des sommets). Un **complexe simplicial** sur  $V$  est un ensemble  $K$  composé de sous-ensemble de  $V$  (appelés les *simplexes*) tel que, pour tout  $\sigma \in K$  et tout  $\tau \subset \sigma$  non-vide, on a  $\tau \in K$ . La dimension d'un simplexe  $\sigma \in K$  est  $\dim(\sigma) = |\sigma| - 1$ .



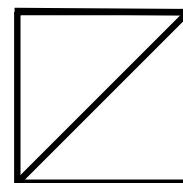
$K$

Soit  $K$  un complexe simplicial. Pour tout  $n \geq 0$ , on définit

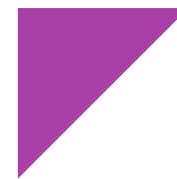
$$K_{(n)} = \{\sigma \in K, \dim(\sigma) = n\}.$$



$K_{(0)}$



$K_{(1)}$

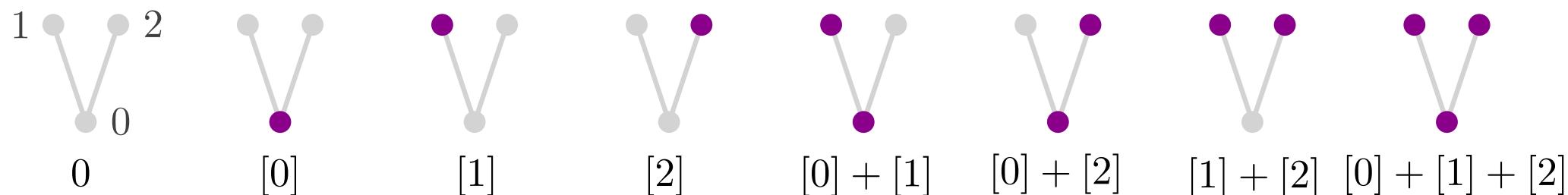


$K_{(2)}$

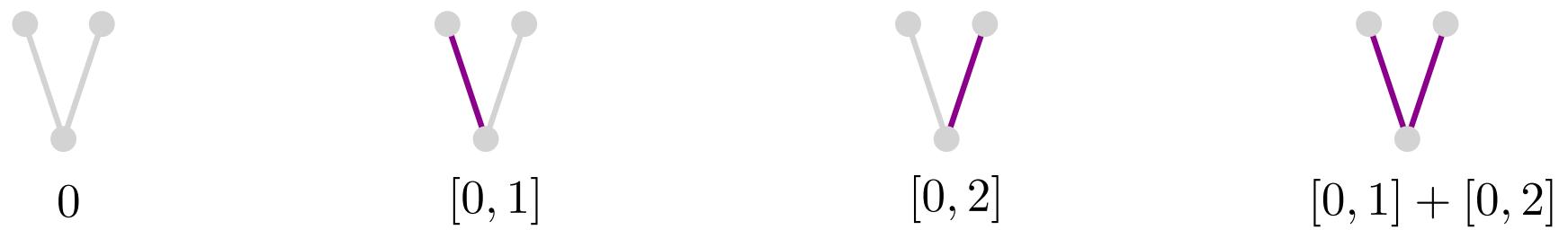
Soit  $n \geq 0$ . Les  $n$ -chaines de  $K$  est l'ensemble  $C_n(K)$  composé des sommes formelles

$$\sum_{\sigma \in K_{(n)}} \epsilon_\sigma \cdot \sigma \quad \text{où} \quad \forall \sigma \in K_{(n)}, \epsilon_\sigma \in \mathbb{Z}/2\mathbb{Z}.$$

**Exemple :** Les 0-chaines de  $K = \{[0], [1], [2], [0, 1], [0, 2]\}$  sont :



et les 1-chaines :



Soit  $n \geq 0$ . Les  $n$ -chaines de  $K$  est l'ensemble  $C_n(K)$  composé des sommes formelles

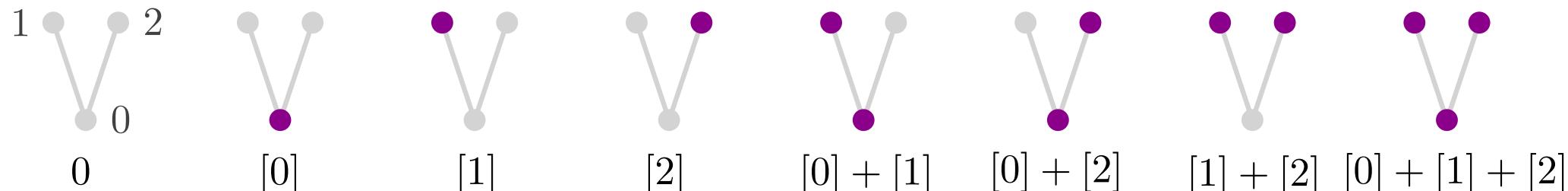
$$\sum_{\sigma \in K_{(n)}} \epsilon_\sigma \cdot \sigma \quad \text{où} \quad \forall \sigma \in K_{(n)}, \epsilon_\sigma \in \mathbb{Z}/2\mathbb{Z}.$$

On peut munir  $C_n(K)$  d'un structure de **groupe** via

$$\sum_{\sigma \in K_{(n)}} \epsilon_\sigma \cdot \sigma + \sum_{\sigma \in K_{(n)}} \eta_\sigma \cdot \sigma = \sum_{\sigma \in K_{(n)}} (\epsilon_\sigma + \eta_\sigma) \cdot \sigma.$$

De plus,  $C_n(K)$  peut être muni d'une structure de  $\mathbb{Z}/2\mathbb{Z}$ -espace vectoriel.

**Exemple :** Les 0-chaines de  $K = \{[0], [1], [2], [0, 1], [0, 2]\}$  sont :



et les 1-chaines :



Soit  $n \geq 0$ . Les  **$n$ -chaines** de  $K$  est l'ensemble  $C_n(K)$  composé des sommes formelles

$$\sum_{\sigma \in K_{(n)}} \epsilon_\sigma \cdot \sigma \quad \text{où} \quad \forall \sigma \in K_{(n)}, \epsilon_\sigma \in \mathbb{Z}/2\mathbb{Z}.$$

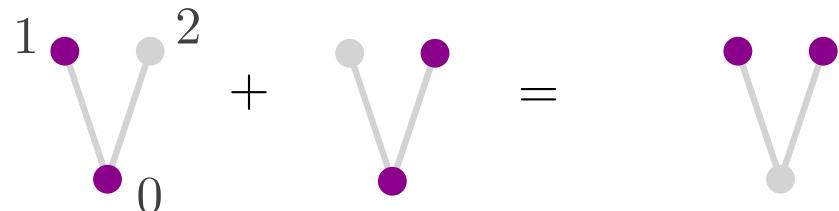
On peut munir  $C_n(K)$  d'un structure de **groupe** via

$$\sum_{\sigma \in K_{(n)}} \epsilon_\sigma \cdot \sigma + \sum_{\sigma \in K_{(n)}} \eta_\sigma \cdot \sigma = \sum_{\sigma \in K_{(n)}} (\epsilon_\sigma + \eta_\sigma) \cdot \sigma.$$

De plus,  $C_n(K)$  peut être muni d'une structure de  $\mathbb{Z}/2\mathbb{Z}$ -espace vectoriel.

**Exemple:** Dans le complexe simplicial  $K = \{[0], [1], [2], [0, 1], [0, 2]\}$ , la somme des 0-chaines  $[0] + [1]$  et  $[0] + [2]$  est  $[1] + [2]$  :

$$([0] + [1]) + ([0] + [2]) = [0] + [0] + [1] + [2] = [1] + [2].$$



# Opérateur de bord

7/43 (1/4)

Soit  $n \geq 1$ , et  $\sigma = [x_0, \dots, x_n] \in K_{(n)}$  un simplexe de dimension  $n$ . On définit son **bord** comme l'élément suivant de  $C_{n-1}(K)$ :

$$\partial_n \sigma = \sum_{\substack{\tau \subset \sigma \\ |\tau| = |\sigma| - 1}} \tau$$

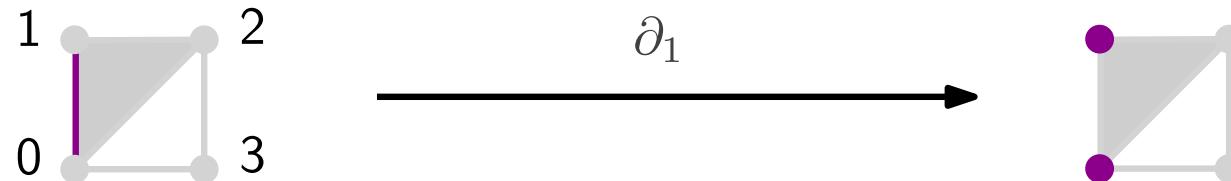
On peut étendre l'opérateur  $\partial_n$  comme une application linéaire  $\partial_n : C_n(K) \rightarrow C_{n-1}(K)$ .

**Exemple:** Soit le complexe simplicial

$$K = \{[0], [1], [2], [3], [0, 1], [0, 2], [1, 2], [1, 3], [2, 3], [0, 1, 2]\}.$$

Le simplexe  $[0, 1]$  a pour faces  $[0]$  et  $[1]$ . Donc

$$\partial_1 [0, 1] = [0] + [1].$$



# Opérateur de bord

7/43 (2/4)

Soit  $n \geq 1$ , et  $\sigma = [x_0, \dots, x_n] \in K_{(n)}$  un simplexe de dimension  $n$ . On définit son **bord** comme l'élément suivant de  $C_{n-1}(K)$ :

$$\partial_n \sigma = \sum_{\substack{\tau \subset \sigma \\ |\tau| = |\sigma| - 1}} \tau$$

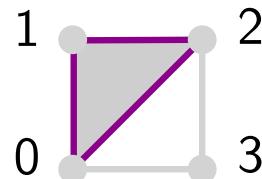
On peut étendre l'opérateur  $\partial_n$  comme une application linéaire  $\partial_n : C_n(K) \rightarrow C_{n-1}(K)$ .

**Exemple:** Soit le complexe simplicial

$$K = \{[0], [1], [2], [3], [0, 1], [0, 2], [1, 2], [1, 3], [2, 3], [0, 1, 2]\}.$$

Le bord de la 1-chaine  $[0, 1] + [1, 2] + [2, 0]$  est

$$\begin{aligned} \partial_1([0, 1] + [1, 2] + [2, 0]) &= \partial_1[0, 1] + \partial_1[1, 2] + \partial_1[2, 0] \\ &= [0] + [1] + [1] + [2] + [2] + [0] = 0 \end{aligned}$$



$\partial_1$



# Opérateur de bord

7/43 (3/4)

Soit  $n \geq 1$ , et  $\sigma = [x_0, \dots, x_n] \in K_{(n)}$  un simplexe de dimension  $n$ . On définit son **bord** comme l'élément suivant de  $C_{n-1}(K)$ :

$$\partial_n \sigma = \sum_{\substack{\tau \subset \sigma \\ |\tau| = |\sigma| - 1}} \tau$$

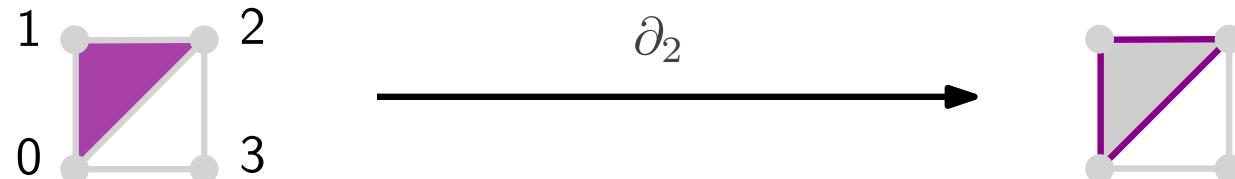
On peut étendre l'opérateur  $\partial_n$  comme une application linéaire  $\partial_n : C_n(K) \rightarrow C_{n-1}(K)$ .

**Exemple:** Soit le complexe simplicial

$$K = \{[0], [1], [2], [3], [0, 1], [0, 2], [1, 2], [1, 3], [2, 3], [0, 1, 2]\}.$$

Le simplexe  $[0, 1, 2]$  a pour faces  $[0, 1]$  et  $[1, 2]$  et  $[2, 0]$ . Donc

$$\partial_2 [0, 1, 2] = [0, 1] + [1, 2] + [2, 0].$$



# Opérateur de bord

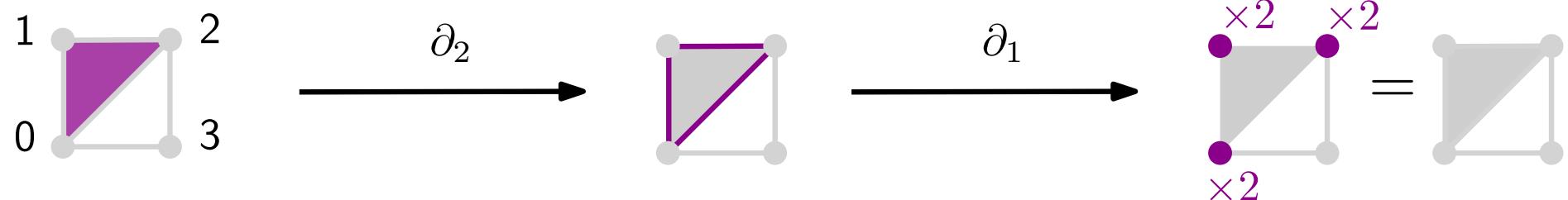
7/43 (4/4)

Soit  $n \geq 1$ , et  $\sigma = [x_0, \dots, x_n] \in K_{(n)}$  un simplexe de dimension  $n$ . On définit son **bord** comme l'élément suivant de  $C_{n-1}(K)$ :

$$\partial_n \sigma = \sum_{\substack{\tau \subset \sigma \\ |\tau| = |\sigma| - 1}} \tau$$

On peut étendre l'opérateur  $\partial_n$  comme une application linéaire  $\partial_n : C_n(K) \rightarrow C_{n-1}(K)$ .

**Proposition :** Pour tout  $n \geq 1$ , pour tout  $c \in C_n(K)$ , on a  $\partial_{n-1} \circ \partial_n(c) = 0$ .



Soit  $n \geq 0$ . On a un suite d'espaces vectoriels

$$\dots \longrightarrow C_{n+1}(K) \xrightarrow{\partial_{n+1}} C_n(K) \xrightarrow{\partial_n} C_{n-1}(K) \longrightarrow \dots$$

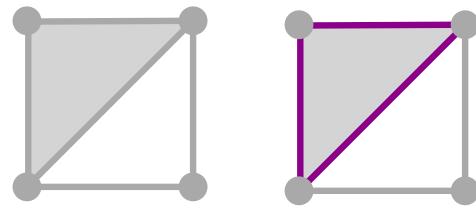
Les applications  $\partial_{n+1}$  et  $\partial_n$  sont linéaires, on peut s'intéresser à leur noyau et leur image

**Définition :** On définit:

- Les  **$n$ -cycles**:  $Z_n(K) = \text{Ker}(\partial_n)$ ,
- Les  **$n$ -bords**:  $B_n(K) = \text{Im}(\partial_{n+1})$ .

**Exemple :** Soit le complexe simplicial

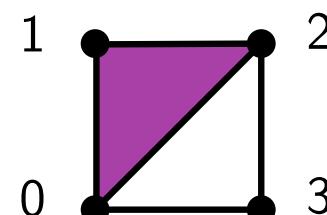
Les 1-cycles sont :



0

$$[0, 2] + [2, 3] + [0, 3]$$

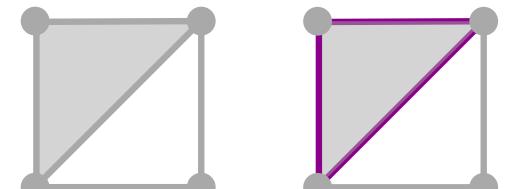
$$[0, 1] + [1, 2] + [0, 2]$$



0

$$[0, 1] + [1, 2] + [2, 3] + [0, 3].$$

Les 1-bords sont :



0

$$[0, 1] + [1, 2] + [0, 2]$$

Soit  $n \geq 0$ . On a un suite d'espaces vectoriels

$$\dots \longrightarrow C_{n+1}(K) \xrightarrow{\partial_{n+1}} C_n(K) \xrightarrow{\partial_n} C_{n-1}(K) \longrightarrow \dots$$

Les applications  $\partial_{n+1}$  et  $\partial_n$  sont linéaires, on peut s'intéresser à leur noyau et leur image

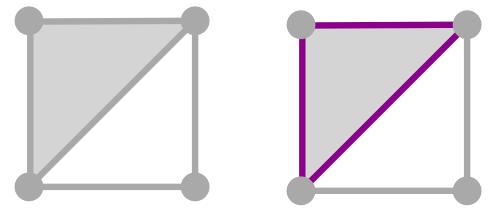
**Définition :** On définit:

- Les  **$n$ -cycles**:  $Z_n(K) = \text{Ker}(\partial_n)$ ,
- Les  **$n$ -bords**:  $B_n(K) = \text{Im}(\partial_{n+1})$ .

**Proposition :** On a  $B_n(K) \subset Z_n(K)$ .

**Exemple :** Soit le complexe simplicial

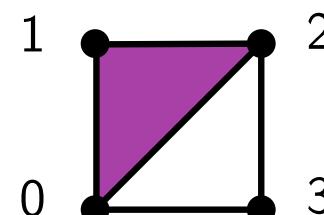
Les 1-cycles sont :



0

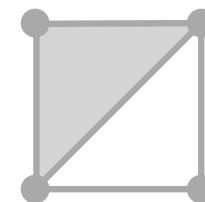
$$[0, 2] + [2, 3]$$

$$[0, 1] + [1, 2] + [0, 2]$$



0

Les 1-bords sont :



0

$$[0, 1] + [1, 2] + [0, 2]$$

$$[0, 1] + [1, 2] + [2, 3] + [0, 3].$$

Soit  $n \geq 0$ . On a un suite d'espaces vectoriels

$$\dots \longrightarrow C_{n+1}(K) \xrightarrow{\partial_{n+1}} C_n(K) \xrightarrow{\partial_n} C_{n-1}(K) \longrightarrow \dots$$

Les applications  $\partial_{n+1}$  et  $\partial_n$  sont linéaires, on peut s'intéresser à leur noyau et leur image

**Définition :** On définit:

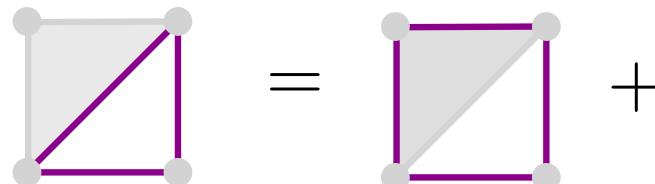
- Les  **$n$ -cycles**:  $Z_n(K) = \text{Ker}(\partial_n)$ ,
- Les  **$n$ -bords**:  $B_n(K) = \text{Im}(\partial_{n+1})$ .

**Proposition :** On a  $B_n(K) \subset Z_n(K)$ .

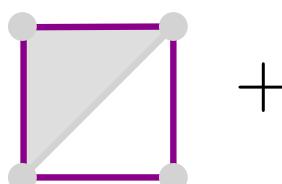
**Définition :** On dit que deux chaines  $c, c' \in C_n(K)$  sont **homologues** si il existe un bord  $b \in B_n(K)$  tel que  $c = c' + b$ .

→ deux chaines sont homologues si elles sont égales à un bord près

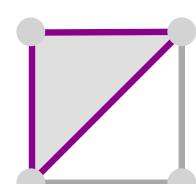
**Exemple :**



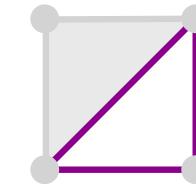
=



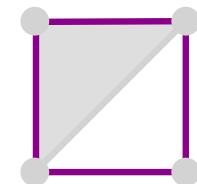
+



donc



≈



$$[0, 2] + [2, 3] + [0, 3] = [0, 1] + [1, 2] + [2, 3] + [0, 3] + [0, 1] + [0, 2] + [1, 2].$$

Soit  $n \geq 0$ . On a un suite d'espaces vectoriels

$$\dots \longrightarrow C_{n+1}(K) \xrightarrow{\partial_{n+1}} C_n(K) \xrightarrow{\partial_n} C_{n-1}(K) \longrightarrow \dots$$

Les applications  $\partial_{n+1}$  et  $\partial_n$  sont linéaires, on peut s'intéresser à leur noyau et leur image

**Définition :** On définit:

- Les  **$n$ -cycles**:  $Z_n(K) = \text{Ker}(\partial_n)$ ,
- Les  **$n$ -bords**:  $B_n(K) = \text{Im}(\partial_{n+1})$ .

**Proposition :** On a  $B_n(K) \subset Z_n(K)$ .

**Définition :** On dit que deux chaines  $c, c' \in C_n(K)$  sont **homologues** si il existe un bord  $b \in B_n(K)$  tel que  $c = c' + b$ .

—————> **Interprétation** : deux cycles sont homologues si il représentent le même trou

**Exemple :**

donc

$$[0, 2] + [2, 3] + [0, 3] = [0, 1] + [1, 2] + [2, 3] + [0, 3] + [0, 1] + [0, 2] + [1, 2].$$

On a défini une suite d'espaces vectoriel, connectés par des applications linéaires

$$\cdots \longrightarrow C_{n+1}(K) \longrightarrow C_n(K) \longrightarrow C_{n-1}(K) \longrightarrow \cdots$$

et pour tout  $n \geq 0$ , on a défini les cycles et le bords  $Z_n(K)$  et  $B_n(K)$ .

Puisque  $B_n(K) \subset Z_n(K)$ , on peut voir  $B_n(K)$  comme un sous-espace linéaire de  $Z_n(K)$ .

**Définition :** Le  $n^{\text{ème}}$  **groupe d'homologie (simpliciale}** de  $K$  est l'espace vectoriel quotient

$$H_n(K) = Z_n(K)/B_n(K).$$

On a défini une suite d'espaces vectoriel, connectés par des applications linéaires

$$\cdots \longrightarrow C_{n+1}(K) \longrightarrow C_n(K) \longrightarrow C_{n-1}(K) \longrightarrow \cdots$$

et pour tout  $n \geq 0$ , on a défini les cycles et le bords  $Z_n(K)$  et  $B_n(K)$ .

Puisque  $B_n(K) \subset Z_n(K)$ , on peut voir  $B_n(K)$  comme un sous-espace linéaire de  $Z_n(K)$ .

**Définition :** Le  $n^{\text{ème}}$  **groupe d'homologie (simpliciale}** de  $K$  est l'espace vectoriel quotient

$$H_n(K) = Z_n(K)/B_n(K).$$

**Remarque :** Un  $\mathbb{Z}/2\mathbb{Z}$ -espace vectoriel fini est isomorphique à  $(\mathbb{Z}/2\mathbb{Z})^k$  pour un certain  $k$ .

**Définition :** Soit  $K$  un complexe simplicial et  $n \geq 0$ . Son  $n^{\text{ème}}$  **nombre de Betti** est l'entier  $\beta_n(K) = \dim H_n(K)$ .

$$H_n(K) = (\mathbb{Z}/2\mathbb{Z})^k \quad \longrightarrow \quad \beta_n(K) = k$$

# Groupes d'homologie

9/43 (3/6)

On a défini une suite d'espaces vectoriel, connectés par des applications linéaires

$$\cdots \longrightarrow C_{n+1}(K) \longrightarrow C_n(K) \longrightarrow C_{n-1}(K) \longrightarrow \cdots$$

et pour tout  $n \geq 0$ , on a défini les cycles et le bords  $Z_n(K)$  et  $B_n(K)$ .

Puisque  $B_n(K) \subset Z_n(K)$ , on peut voir  $B_n(K)$  comme un sous-espace linéaire de  $Z_n(K)$ .

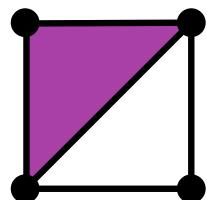
**Définition :** Le  $n^{\text{ème}}$  **groupe d'homologie (simpliciale}** de  $K$  est l'espace vectoriel quotient

$$H_n(K) = Z_n(K)/B_n(K).$$

**Remarque :** Un  $\mathbb{Z}/2\mathbb{Z}$ -espace vectoriel fini est isomorphique à  $(\mathbb{Z}/2\mathbb{Z})^k$  pour un certain  $k$ .

**Définition :** Soit  $K$  un complexe simplicial et  $n \geq 0$ . Son  $n^{\text{ème}}$  **nombre de Betti** est l'entier  $\beta_n(K) = \dim H_n(K)$ .

**Exemple :**



$$H_0(K) = \mathbb{Z}/2\mathbb{Z} \longrightarrow \beta_0(K) = 1$$

$$H_1(K) = \mathbb{Z}/2\mathbb{Z} \longrightarrow \beta_1(K) = 1$$

$$H_2(K) = 0 \longrightarrow \beta_2(K) = 0$$

On a défini une suite d'espaces vectoriel, connectés par des applications linéaires

$$\cdots \longrightarrow C_{n+1}(K) \longrightarrow C_n(K) \longrightarrow C_{n-1}(K) \longrightarrow \cdots$$

et pour tout  $n \geq 0$ , on a défini les cycles et le bords  $Z_n(K)$  et  $B_n(K)$ .

Puisque  $B_n(K) \subset Z_n(K)$ , on peut voir  $B_n(K)$  comme un sous-espace linéaire de  $Z_n(K)$ .

**Définition :** Le  $n^{\text{ème}}$  **groupe d'homologie (simpliciale}** de  $K$  est l'espace vectoriel quotient

$$H_n(K) = Z_n(K)/B_n(K).$$

**Remarque :** Un  $\mathbb{Z}/2\mathbb{Z}$ -espace vectoriel fini est isomorphique à  $(\mathbb{Z}/2\mathbb{Z})^k$  pour un certain  $k$ .

**Définition :** Soit  $K$  un complexe simplicial et  $n \geq 0$ . Son  $n^{\text{ème}}$  **nombre de Betti** est l'entier  $\beta_n(K) = \dim H_n(K)$ .

**Proposition :** Si  $X$  et  $Y$  sont deux espaces topologiques homotopiquement équivalents, alors pour tout  $n \geq 0$  ils ont les mêmes groupes d'homologie  $H_n(X) \simeq H_n(Y)$ . Par conséquent,  $\beta_n(X) = \beta_n(Y)$ .

On a défini une suite d'espaces vectoriel, connectés par des applications linéaires

$$\cdots \longrightarrow C_{n+1}(K) \longrightarrow C_n(K) \longrightarrow C_{n-1}(K) \longrightarrow \cdots$$

et pour tout  $n \geq 0$ , on a défini les cycles et le bords  $Z_n(K)$  et  $B_n(K)$ .

Puisque  $B_n(K) \subset Z_n(K)$ , on peut voir  $B_n(K)$  comme un sous-espace linéaire de  $Z_n(K)$ .

**Définition :** Le  $n^{\text{ème}}$  **groupe d'homologie (simpliciale}** de  $K$  est l'espace vectoriel quotient

$$H_n(K) = Z_n(K)/B_n(K).$$

**Remarque :** Un  $\mathbb{Z}/2\mathbb{Z}$ -espace vectoriel fini est isomorphique à  $(\mathbb{Z}/2\mathbb{Z})^k$  pour un certain  $k$ .

**Définition :** Soit  $K$  un complexe simplicial et  $n \geq 0$ . Son  $n^{\text{ème}}$  **nombre de Betti** est l'entier  $\beta_n(K) = \dim H_n(K)$ .

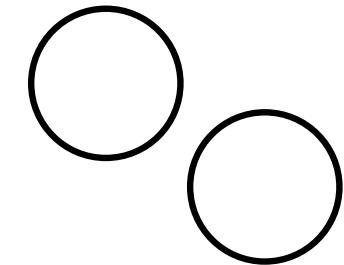
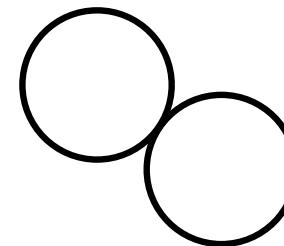
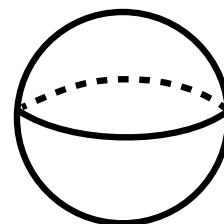
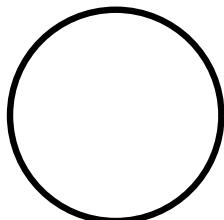
**Proposition :** Si  $X$  et  $Y$  sont deux espaces topologiques homotopiquement équivalents, alors pour tout  $n \geq 0$  ils ont les mêmes groupes d'homologie  $H_n(X) \simeq H_n(Y)$ . Par conséquent,  $\beta_n(X) = \beta_n(Y)$ .

→ (la théorie marche mieux avec l'*homologie singulière*)

# Groupes d'homologie

9/43 (6/6)

$X$



$$H_0(X) \quad \mathbb{Z}/2\mathbb{Z}$$

$$\mathbb{Z}/2\mathbb{Z}$$

$$\mathbb{Z}/2\mathbb{Z}$$

$$\mathbb{Z}/2\mathbb{Z}$$

$$(\mathbb{Z}/2\mathbb{Z})^2$$

$$\beta_0(X)$$

$$1$$

$$1$$

$$1$$

$$1$$

$$2$$

$$H_1(X)$$

$$0$$

$$\mathbb{Z}/2\mathbb{Z}$$

$$0$$

$$(\mathbb{Z}/2\mathbb{Z})^2$$

$$(\mathbb{Z}/2\mathbb{Z})^2$$

$$\beta_1(X)$$

$$0$$

$$1$$

$$0$$

$$2$$

$$2$$

$$H_2(X)$$

$$0$$

$$0$$

$$\mathbb{Z}/2\mathbb{Z}$$

$$0$$

$$0$$

$$\beta_2(X)$$

$$0$$

$$0$$

$$1$$

$$0$$

$$0$$

# I - Homologie simpliciale

1 - Groupes d'homologie

2 - Fonctorialité

# II - Inférence topologique

1 - Estimation du paramètre

2 - Nerfs

# III - Homologie persistante

1 - Modules de persistance

2 - Décomposition

3 - Stabilité

# IV - Applications

Nous avons vu que l'homologie transforme les espaces topologiques en espaces vectoriels

$$H_i : \text{Top} \longrightarrow \text{Vect}$$

$$X \longmapsto H_i(X)$$

En fait, elle transforme aussi les **applications continues** en **applications linéaires**

$$X \xrightarrow{f} Y \qquad H_n(X) \xrightarrow{H_n(f)} H_n(Y)$$

Cette opération préserve les **diagrammes commutatifs** :

$$X \xrightarrow{f} Y \xrightarrow{g} Z,$$

$$\text{with } g \circ f \text{ curved arrow from } X \text{ to } Z.$$

$$H_n(X) \xrightarrow{H_n(f)} H_n(Y) \xrightarrow{H_n(g)} H_n(Z).$$

$$\text{with } H_n(g \circ f) \text{ curved arrow from } H_n(X) \text{ to } H_n(Z).$$

$$H_n(g \circ f) = H_n(g) \circ H_n(f)$$

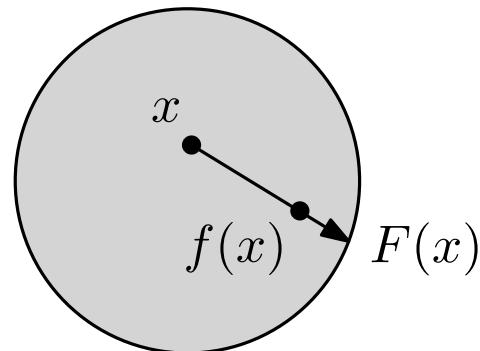
## Application : Théorème du point fixe de Brouwer

Soit  $f: \mathcal{B} \rightarrow \mathcal{B}$  une application continue, où  $\mathcal{B}$  est la boule unité fermée de  $\mathbb{R}^n$ .

Montrons que  $f$  a un point fixe ( $f(x) = x$ ).

Si ce n'était pas le cas, on pourrait définir une application  $F: \mathcal{B} \rightarrow \partial\mathcal{B}$  telle que  $F$  restreinte à  $\partial\mathcal{B}$  soit l'application identité.

En effet, on peut définir  $F(x)$  comme le premier point d'intersection entre la demi-droite  $[x, f(x))$  et  $\partial\mathcal{B}$ .



Notons l'inclusion  $i: \partial\mathcal{B} \rightarrow \mathcal{B}$ . Alors  $F \circ i: \partial\mathcal{B} \rightarrow \partial\mathcal{B}$  est l'identité.

Par fonctorialité, on a des diagrammes commutatifs

$$\begin{array}{ccccc} & & \text{id} & & \\ & \partial\mathcal{B} & \xrightarrow{i} & \mathcal{B} & \xrightarrow{F} \partial\mathcal{B}, \end{array}$$

$$\begin{array}{ccccc} & & H_i(\text{id}) & & \\ & H_i(\partial\mathcal{B}) & \xrightarrow{H_i(i)} & H_i(\mathcal{B}) & \xrightarrow{H_i(F)} H_i(\partial\mathcal{B}). \end{array}$$

Mais pour  $i = n - 1$ , on a une absurdité :

$$\begin{array}{ccccc} & & \text{id} & & \\ & \mathbb{Z}/2\mathbb{Z} & \longrightarrow & 0 & \longrightarrow \mathbb{Z}/2\mathbb{Z}. \end{array}$$

## I - Homologie simpliciale

- 1 - Groupes d'homologie
- 2 - Fonctorialité

## II - Inférence topologique

- 1 - Estimation du paramètre
- 2 - Nerfs

## III - Homologie persistante

- 1 - Modules de persistance
- 2 - Décomposition
- 3 - Stabilité

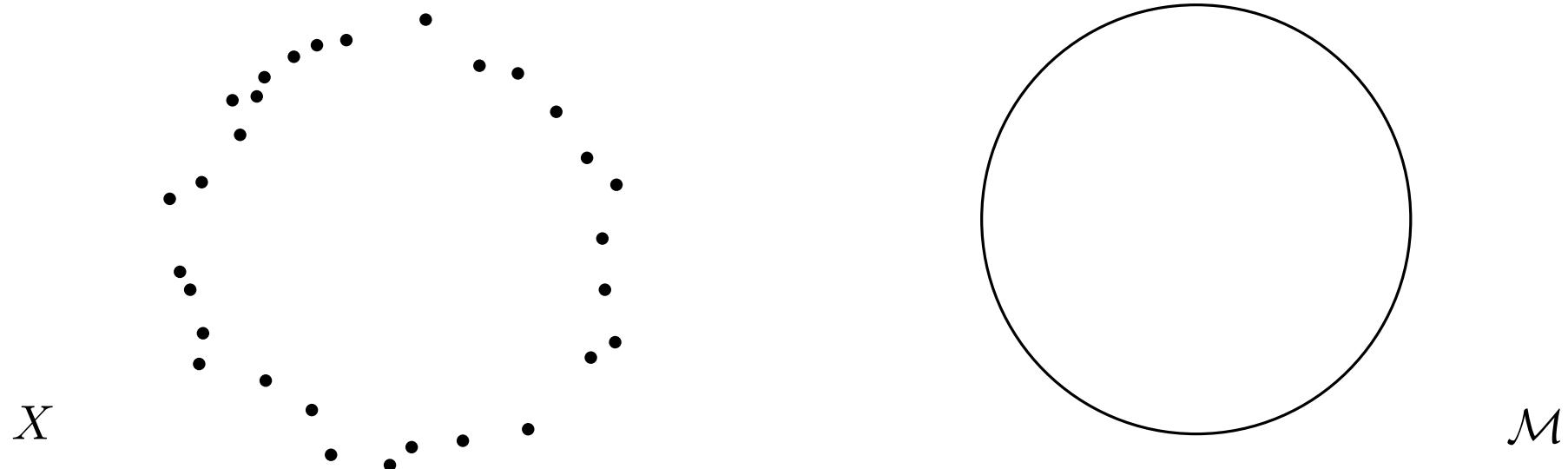
## IV - Applications

# Problème de l'inférence homologique 14/43 (1/12)

Soit  $\mathcal{M} \subset \mathbb{R}^n$  un sous-ensemble.

Supposons que l'on ait un échantillon fini  $X \subset \mathcal{M}$ .

**Problème :** Estimer les groupes d'homologie de  $\mathcal{M}$  à partir de  $X$ .

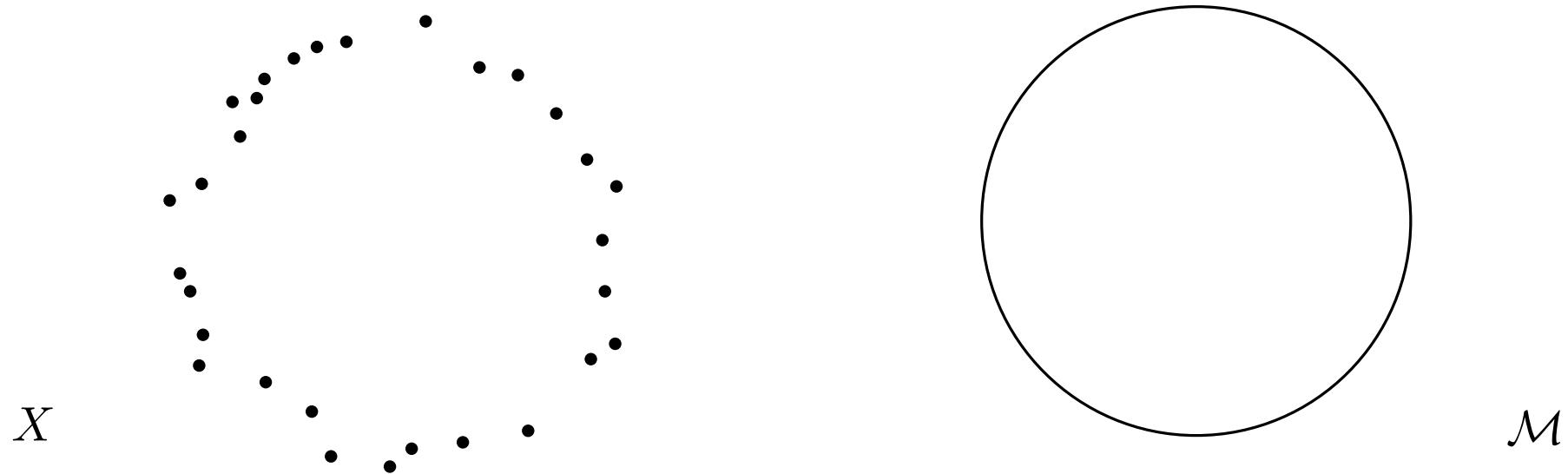


# Problème de l'inférence homologique 14/43 (2/12)

Soit  $\mathcal{M} \subset \mathbb{R}^n$  un sous-ensemble.

Supposons que l'on ait un échantillon fini  $X \subset \mathcal{M}$ .

**Problème :** Estimer les groupes d'homologie de  $\mathcal{M}$  à partir de  $X$ .



On ne peut pas utiliser  $X$  directement. Son homologie est décevante :

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

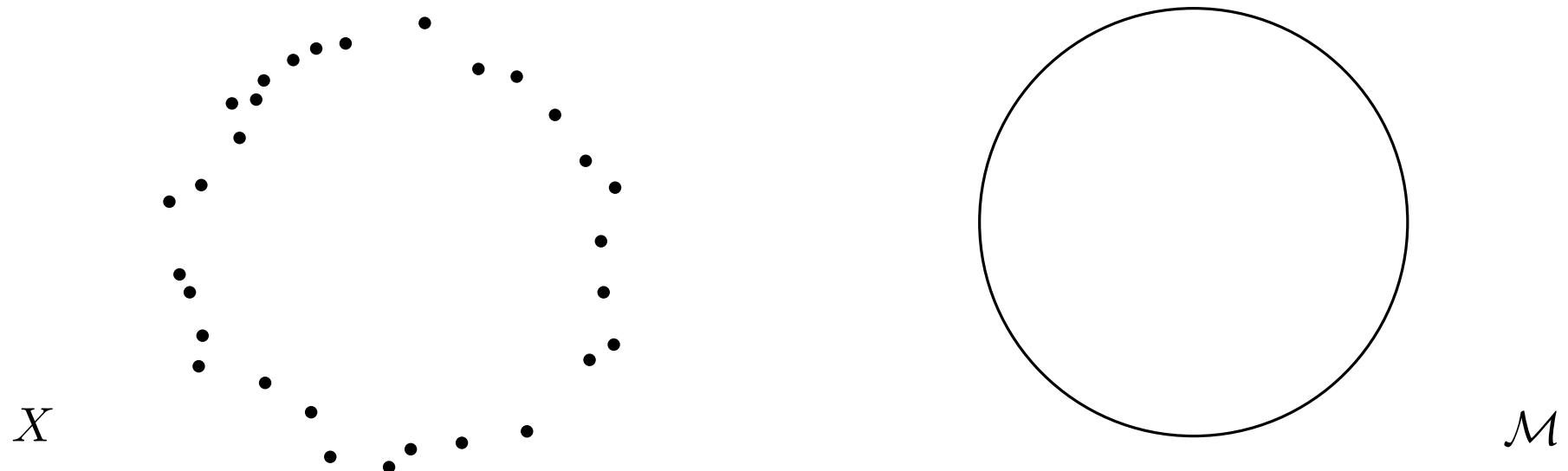
nombre de composantes connexes  
= nombre de points de  $X$

# Problème de l'inférence homologique 14/43 (3/12)

Soit  $\mathcal{M} \subset \mathbb{R}^n$  un sous-ensemble.

Supposons que l'on ait un échantillon fini  $X \subset \mathcal{M}$ .

**Problème :** Estimer les groupes d'homologie de  $\mathcal{M}$  à partir de  $X$ .



On ne peut pas utiliser  $X$  directement.

**Idée :** Épaissir  $X$ .

**Définition :** Pour chaque  $t \geq 0$ , le  $t$ -épaisissement de l'ensemble  $X$ , noté  $X^t$ , est l'ensemble des points de l'espace ambiant à distance au plus  $t$  de  $X$  :

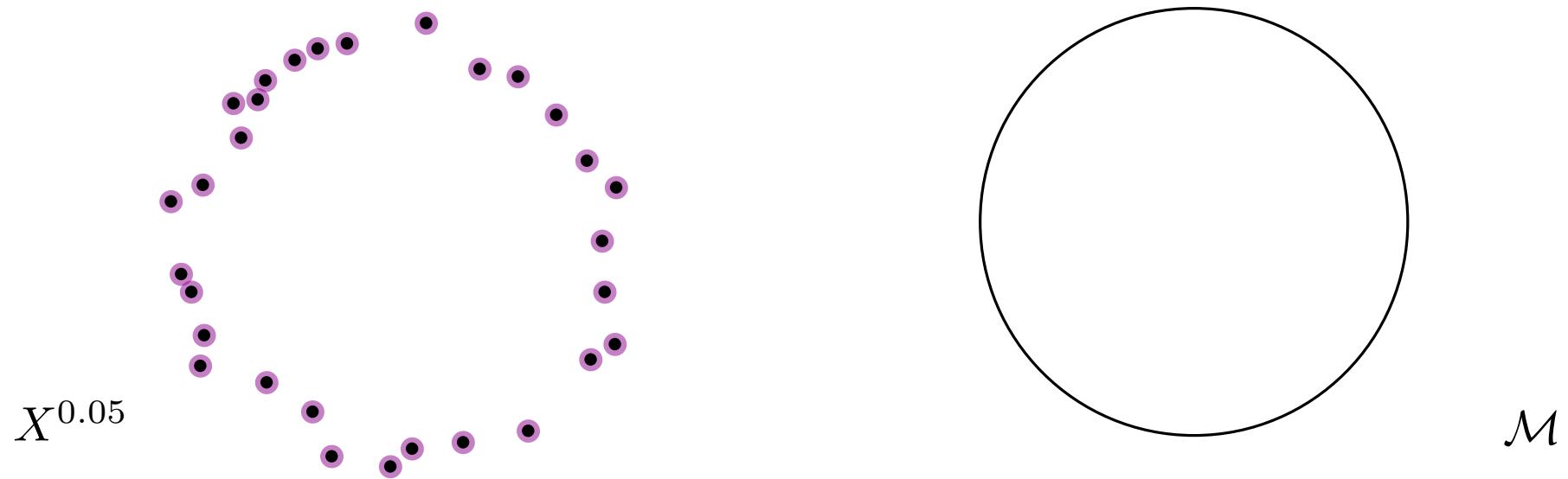
$$X^t = \{y \in \mathbb{R}^n, \exists x \in X, \|x - y\| \leq t\}.$$

# Problème de l'inférence homologique 14/43 (4/12)

Soit  $\mathcal{M} \subset \mathbb{R}^n$  un sous-ensemble.

Supposons que l'on ait un échantillon fini  $X \subset \mathcal{M}$ .

**Problème :** Estimer les groupes d'homologie de  $\mathcal{M}$  à partir de  $X$ .



On ne peut pas utiliser  $X$  directement.

**Idée :** Épaissir  $X$ .

**Définition :** Pour chaque  $t \geq 0$ , le  $t$ -épaisissement de l'ensemble  $X$ , noté  $X^t$ , est l'ensemble des points de l'espace ambiant à distance au plus  $t$  de  $X$  :

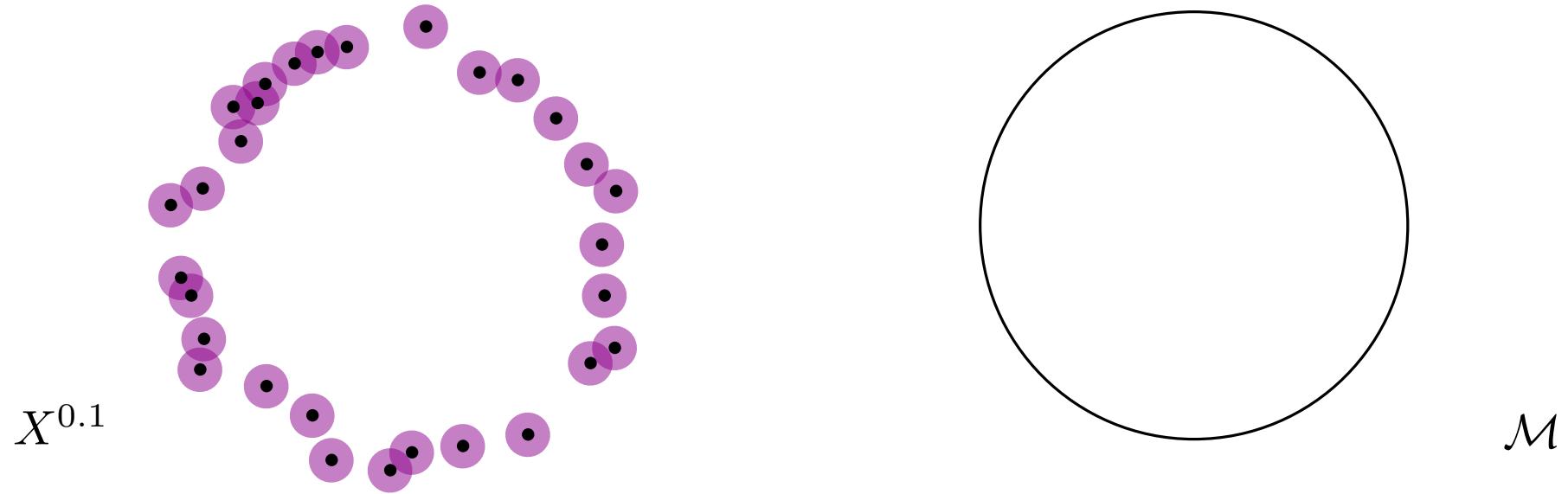
$$X^t = \{y \in \mathbb{R}^n, \exists x \in X, \|x - y\| \leq t\}.$$

# Problème de l'inférence homologique 14/43 (5/12)

Soit  $\mathcal{M} \subset \mathbb{R}^n$  un sous-ensemble.

Supposons que l'on ait un échantillon fini  $X \subset \mathcal{M}$ .

**Problème :** Estimer les groupes d'homologie de  $\mathcal{M}$  à partir de  $X$ .



On ne peut pas utiliser  $X$  directement.

**Idée :** Épaissir  $X$ .

**Définition :** Pour chaque  $t \geq 0$ , le  $t$ -épaisissement de l'ensemble  $X$ , noté  $X^t$ , est l'ensemble des points de l'espace ambiant à distance au plus  $t$  de  $X$  :

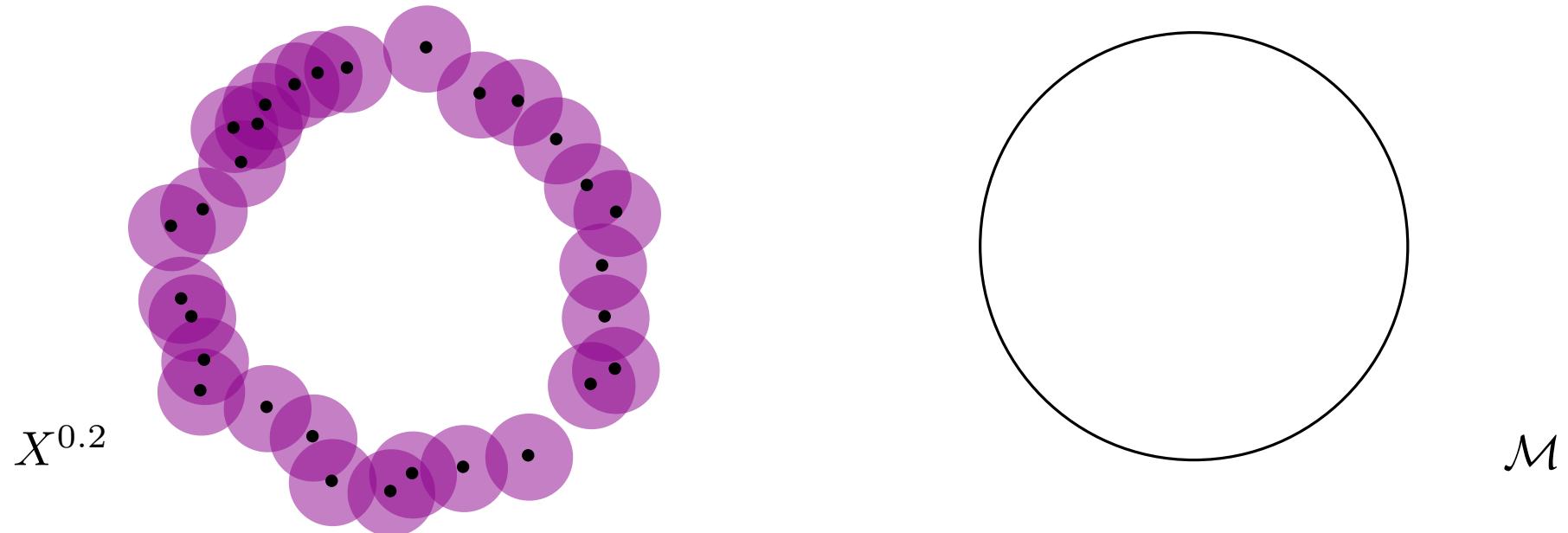
$$X^t = \{y \in \mathbb{R}^n, \exists x \in X, \|x - y\| \leq t\}.$$

# Problème de l'inférence homologique 14/43 (6/12)

Soit  $\mathcal{M} \subset \mathbb{R}^n$  un sous-ensemble.

Supposons que l'on ait un échantillon fini  $X \subset \mathcal{M}$ .

**Problème :** Estimer les groupes d'homologie de  $\mathcal{M}$  à partir de  $X$ .



On ne peut pas utiliser  $X$  directement.

**Idée :** Épaissir  $X$ .

**Définition :** Pour chaque  $t \geq 0$ , le  $t$ -épaisissement de l'ensemble  $X$ , noté  $X^t$ , est l'ensemble des points de l'espace ambiant à distance au plus  $t$  de  $X$  :

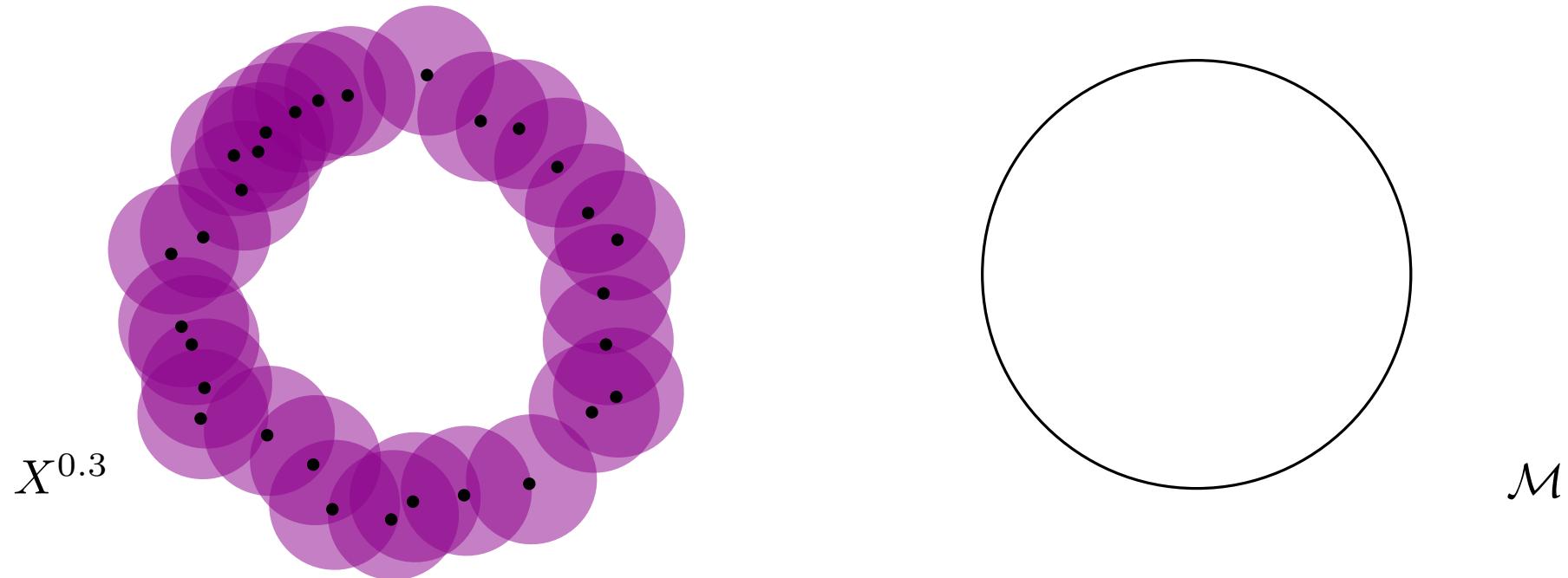
$$X^t = \{y \in \mathbb{R}^n, \exists x \in X, \|x - y\| \leq t\}.$$

# Problème de l'inférence homologique 14/43 (7/12)

Soit  $\mathcal{M} \subset \mathbb{R}^n$  un sous-ensemble.

Supposons que l'on ait un échantillon fini  $X \subset \mathcal{M}$ .

**Problème :** Estimer les groupes d'homologie de  $\mathcal{M}$  à partir de  $X$ .



On ne peut pas utiliser  $X$  directement.

**Idée :** Épaissir  $X$ .

**Définition :** Pour chaque  $t \geq 0$ , le  $t$ -épaisissement de l'ensemble  $X$ , noté  $X^t$ , est l'ensemble des points de l'espace ambiant à distance au plus  $t$  de  $X$  :

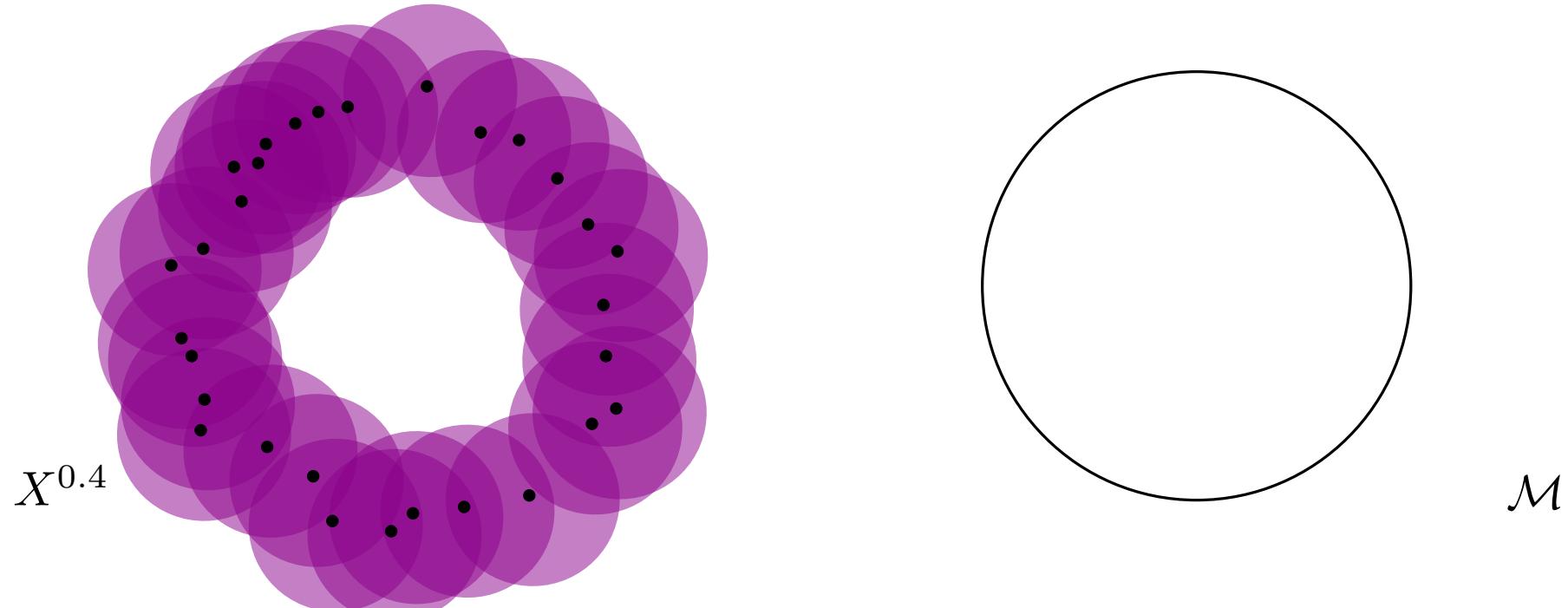
$$X^t = \{y \in \mathbb{R}^n, \exists x \in X, \|x - y\| \leq t\}.$$

# Problème de l'inférence homologique 14/43 (8/12)

Soit  $\mathcal{M} \subset \mathbb{R}^n$  un sous-ensemble.

Supposons que l'on ait un échantillon fini  $X \subset \mathcal{M}$ .

**Problème :** Estimer les groupes d'homologie de  $\mathcal{M}$  à partir de  $X$ .



On ne peut pas utiliser  $X$  directement.

**Idée :** Épaissir  $X$ .

**Définition :** Pour chaque  $t \geq 0$ , le  $t$ -épaisissement de l'ensemble  $X$ , noté  $X^t$ , est l'ensemble des points de l'espace ambiant à distance au plus  $t$  de  $X$  :

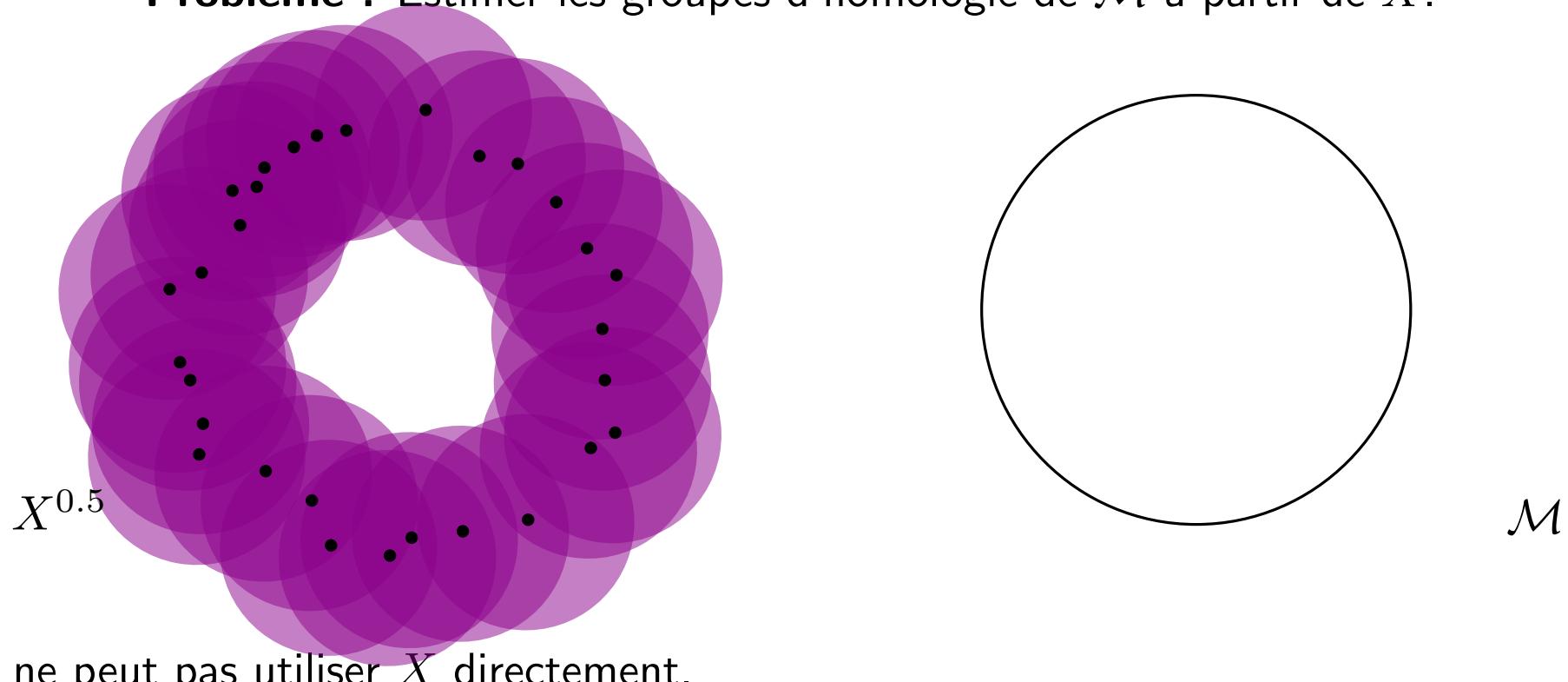
$$X^t = \{y \in \mathbb{R}^n, \exists x \in X, \|x - y\| \leq t\}.$$

# Problème de l'inférence homologique 14/43 (9/12)

Soit  $\mathcal{M} \subset \mathbb{R}^n$  un sous-ensemble.

Supposons que l'on ait un échantillon fini  $X \subset \mathcal{M}$ .

**Problème :** Estimer les groupes d'homologie de  $\mathcal{M}$  à partir de  $X$ .



On ne peut pas utiliser  $X$  directement.

**Idée :** Épaissir  $X$ .

**Définition :** Pour chaque  $t \geq 0$ , le  $t$ -épaisissement de l'ensemble  $X$ , noté  $X^t$ , est l'ensemble des points de l'espace ambiant à distance au plus  $t$  de  $X$  :

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

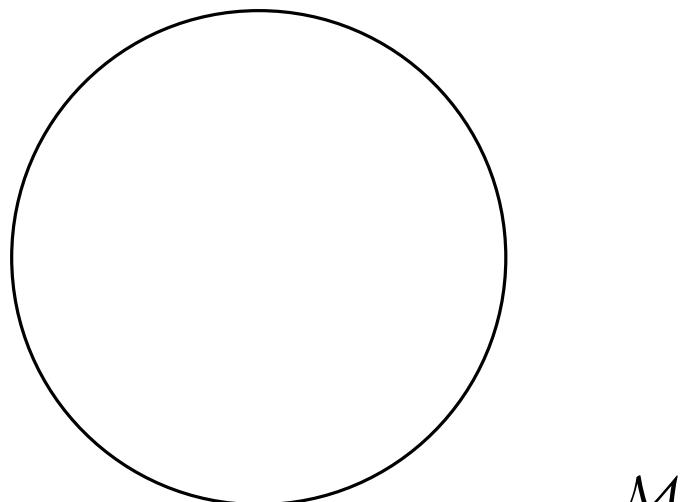
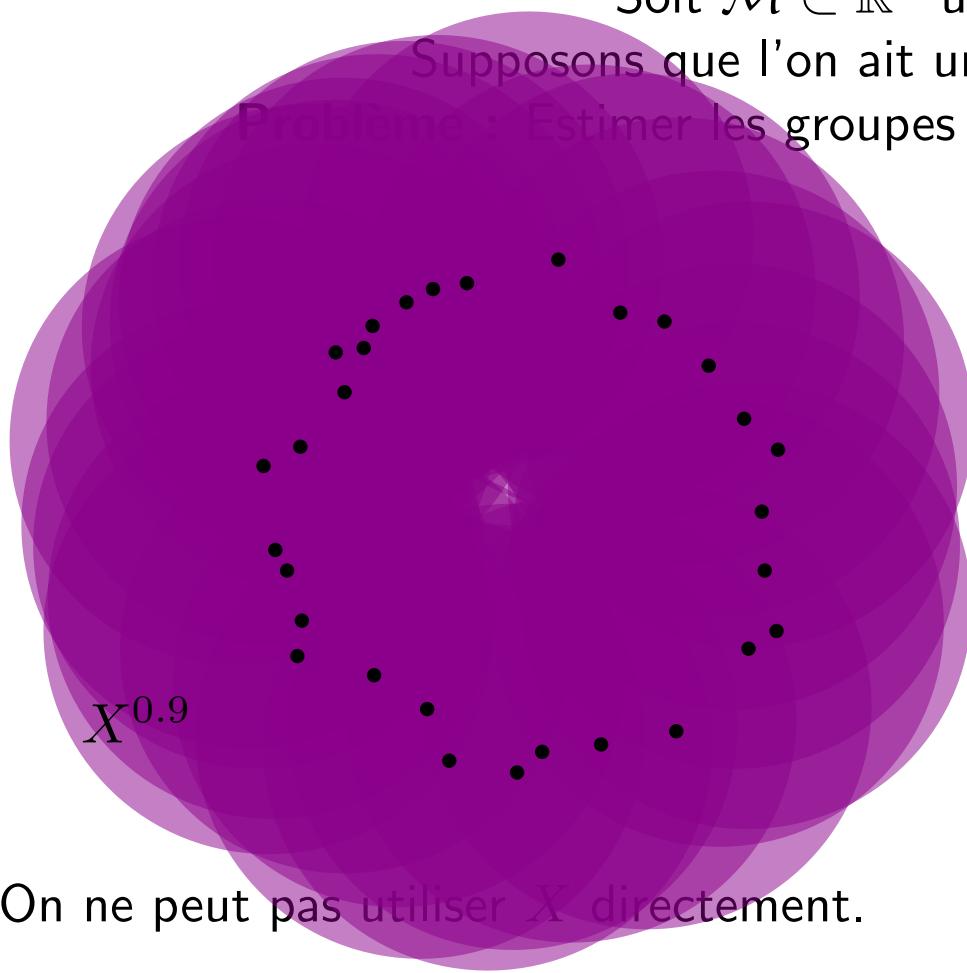
# Problème de l'inférence homologique

14/43 (10/12)

Soit  $\mathcal{M} \subset \mathbb{R}^n$  un sous-ensemble.

Supposons que l'on ait un échantillon fini  $X \subset \mathcal{M}$ .

Problème : Estimer les groupes d'homologie de  $\mathcal{M}$  à partir de  $X$ .



On ne peut pas utiliser  $X$  directement.

**Idée :** Épaissir  $X$ .

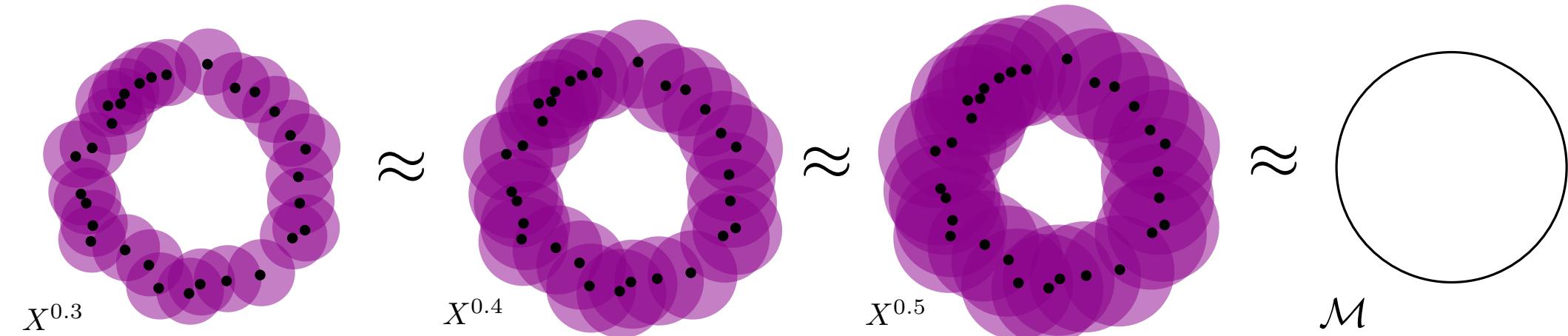
**Définition :** Pour chaque  $t \geq 0$ , le  $t$ -épaisissement de l'ensemble  $X$ , noté  $X^t$ , est l'ensemble des points de l'espace ambiant à distance au plus  $t$  de  $X$  :

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

# Problème de l'inférence homologique

14/43 (11/12)

Certains épaissements sont homotopiquement équivalents à  $\mathcal{M}$ .



On peut donc retrouver l'homologie de  $\mathcal{M}$  :

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

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

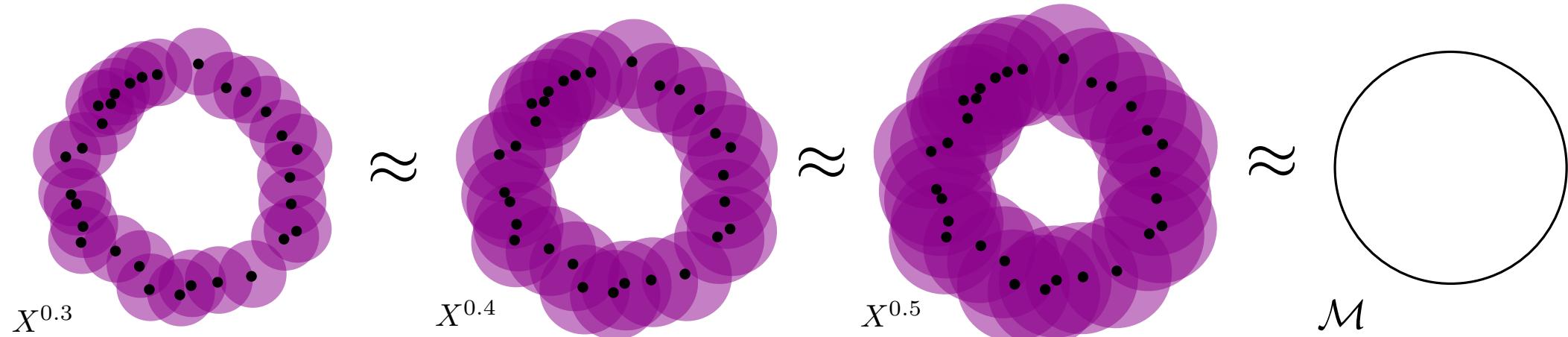
$$\beta_2(\mathcal{M}) = \beta_2(X^{0.3})$$

...

# Problème de l'inférence homologique

14/43 (12/12)

Certains épaissements sont homotopiquement équivalents à  $\mathcal{M}$ .



On peut donc retrouver l'homologie de  $\mathcal{M}$  :

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

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

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

...

Question 1 : Comment sélectionner un  $t$  tel que  $X^t \approx \mathcal{M}$  ?

Question 2 : Comment calculer les groupes d'homologie  $X^t$  ?

# I - Homologie simpliciale

- 1 - Groupes d'homologie
- 2 - Fonctorialité

# II - Inférence topologique

- 1 - Estimation du paramètre
- 2 - Nerfs

# III - Homologie persistante

- 1 - Modules de persistance
- 2 - Décomposition
- 3 - Stabilité

# IV - Applications

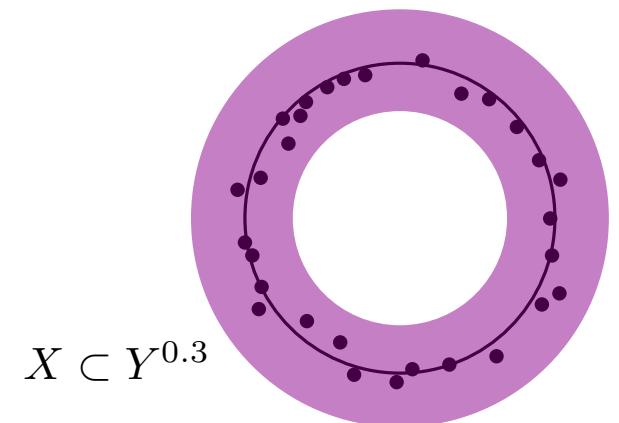
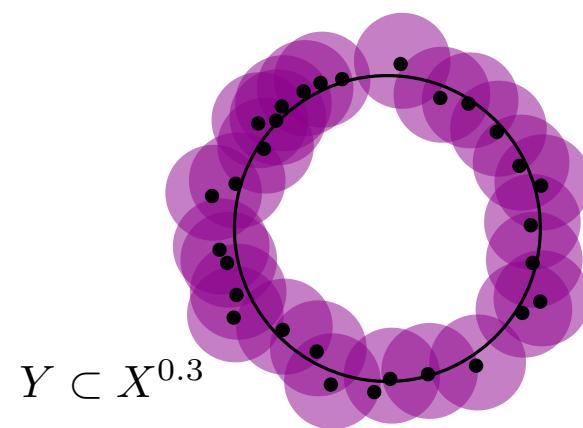
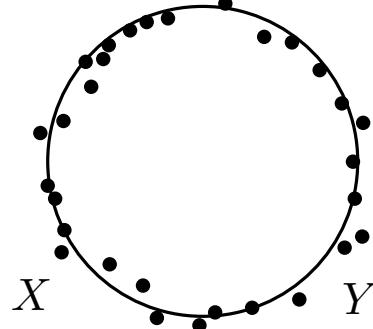
# Distance de Hausdorff

16/43

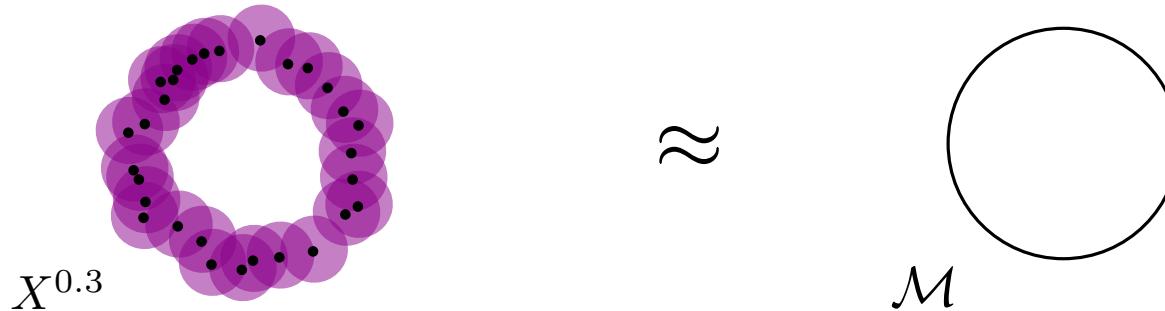
Définition : Soient  $X, Y \subset \mathbb{R}^n$  deux sous-ensembles (compacts). La **distance de Hausdorff** entre  $X$  et  $Y$  est

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

Propriété : La distance de Hausdorff est égale à  $\inf \{t \geq 0, X \subset Y^t \text{ et } Y \subset X^t\}$ .



Question 1 : Comment sélectionner un  $t$  tel que  $X^t \approx \mathcal{M}$ ?



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

Soit  $X$  et  $\mathcal{M}$  deux sous-ensembles de  $\mathbb{R}^n$ . Supposons que  $\mathcal{M}$  ait une portée strictement positive, et que  $d_H(X, \mathcal{M}) \leq \frac{1}{17} \text{portée}(\mathcal{M})$ .

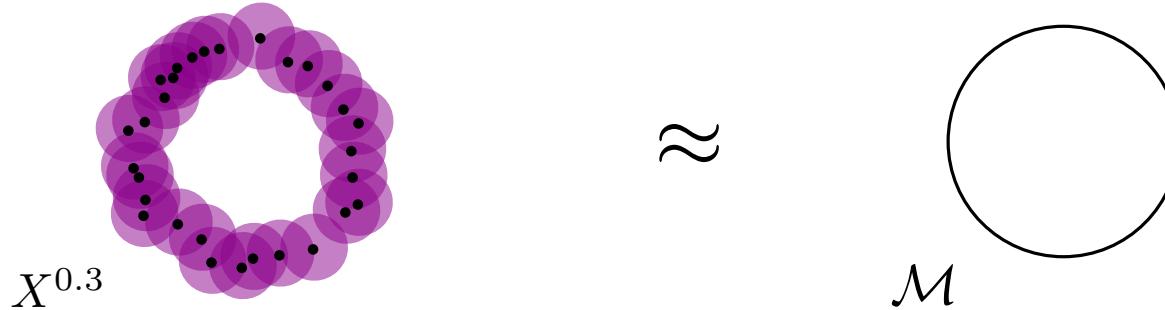
Alors  $X^t$  et  $\mathcal{M}$  sont homotopiquement équivalents, pourvu que

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

# Sélection du paramètre $t$

17/43 (2/2)

Question 1 : Comment sélectionner un  $t$  tel que  $X^t \approx \mathcal{M}$ ?



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

Soit  $X$  et  $\mathcal{M}$  deux sous-ensembles de  $\mathbb{R}^n$ . Supposons que  $\mathcal{M}$  ait une portée strictement positive, et que  $d_H(X, \mathcal{M}) \leq \frac{1}{17} \text{portée}(\mathcal{M})$ .

Alors  $X^t$  et  $\mathcal{M}$  sont homotopiquement équivalents, pourvu que

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

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

Soit  $X$  et  $\mathcal{M}$  deux sous-ensembles de  $\mathbb{R}^n$ , avec  $\mathcal{M}$  une sous-variété, et  $X$  un sous-ensemble fini de  $\mathcal{M}$ . Supposons que  $\mathcal{M}$  ait une portée strictement positive.

Alors  $X^t$  et  $\mathcal{M}$  sont homotopiquement équivalents, pourvu que

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

# I - Homologie simpliciale

- 1 - Groupes d'homologie
- 2 - Fonctorialité

# II - Inférence topologique

- 1 - Estimation du paramètre
- 2 - Nerfs

# III - Homologie persistante

- 1 - Modules de persistance
- 2 - Décomposition
- 3 - Stabilité

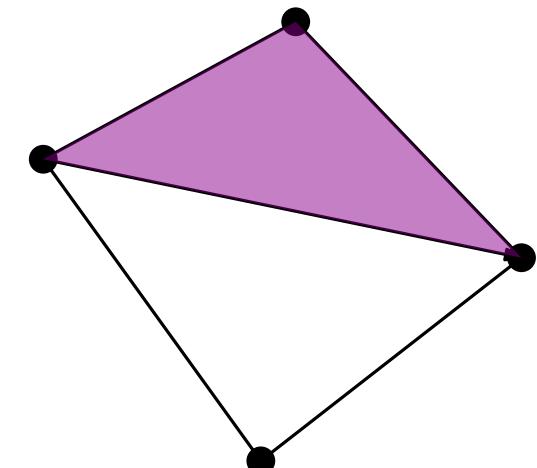
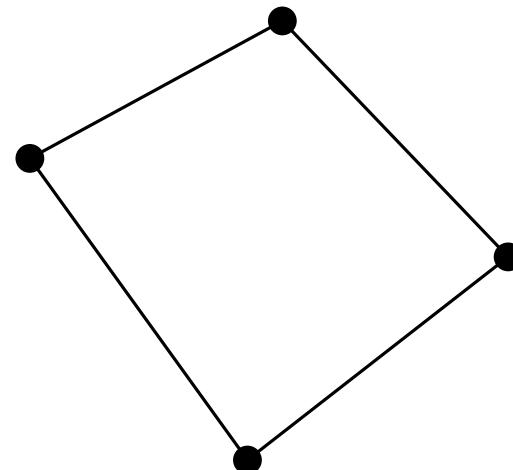
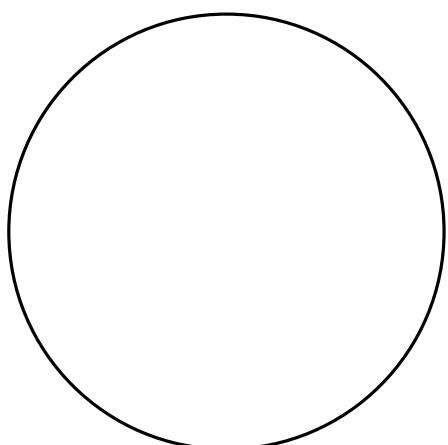
# IV - Applications

Question 2 : Comment calculer les groupes d'homologie de  $X^t$ ?

Nous avons besoin d'une **triangulation de  $X^t$** , c'est-à-dire : un complexe simplicial  $K$  homéomorphe à  $X^t$ .

En fait, on aura besoin de quelque chose de plus faible : complexe simplicial  $K$  homotopiquement équivalent à  $X^t$ .

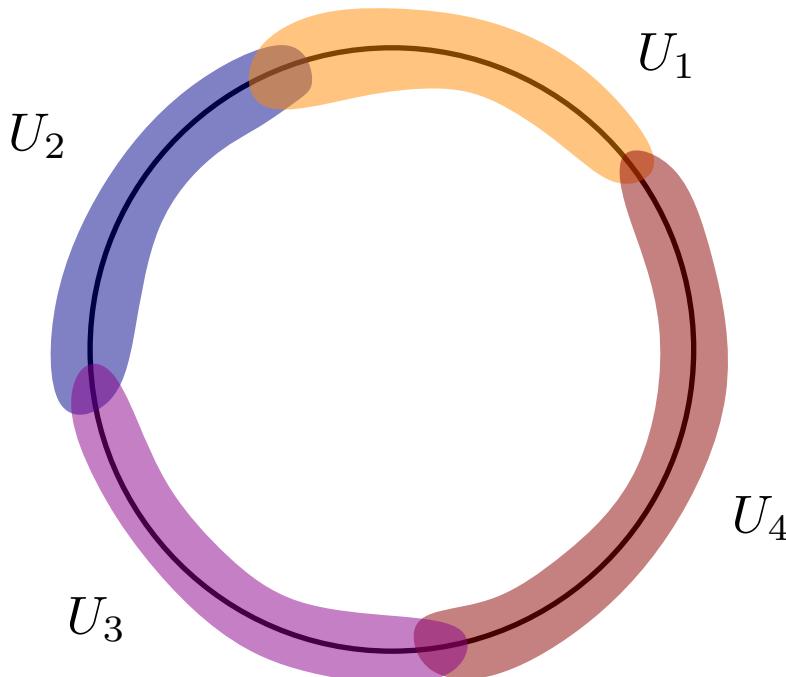
Dans les deux cas, on aura  $\beta_i(X^t) = \beta_i(K)$  pour tout  $i \geq 0$ .



Définition : Soit  $X$  un espace topologique, et  $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$  un *recouvrement* de  $X$ , c'est-à-dire une collection de sous-ensembles  $U_i \subset X$  telle que

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

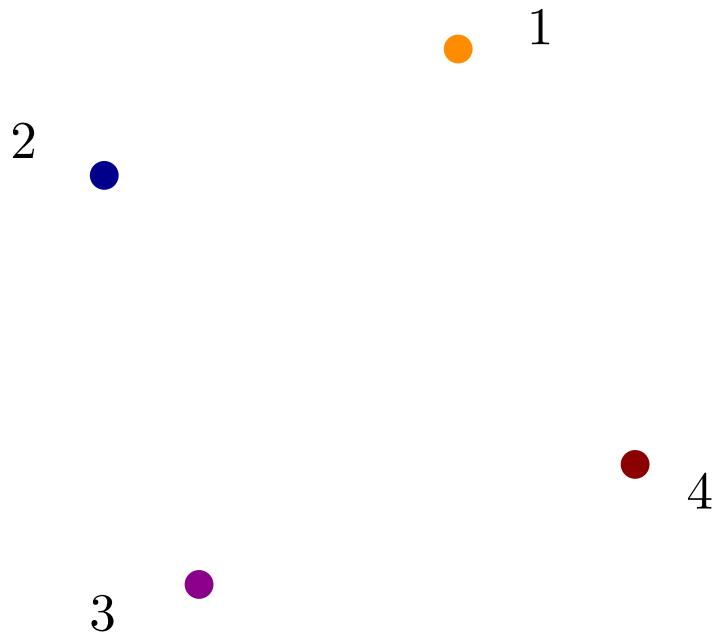
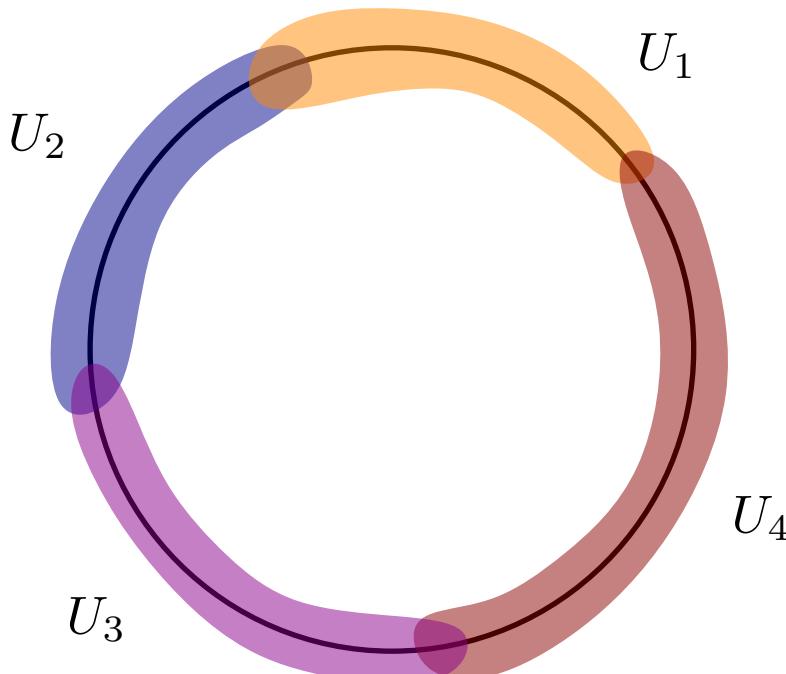
Le **nerf** de  $\mathcal{U}$  est le complexe simplicial avec ensemble de sommets  $\{1, \dots, N\}$  et dont les  $m$ -simplexes sont les sous-ensembles  $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$  tels que  $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$ . On le note  $\mathcal{N}(\mathcal{U})$ .



Définition : Soit  $X$  un espace topologique, et  $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$  un *recouvrement* de  $X$ , c'est-à-dire une collection de sous-ensembles  $U_i \subset X$  telle que

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

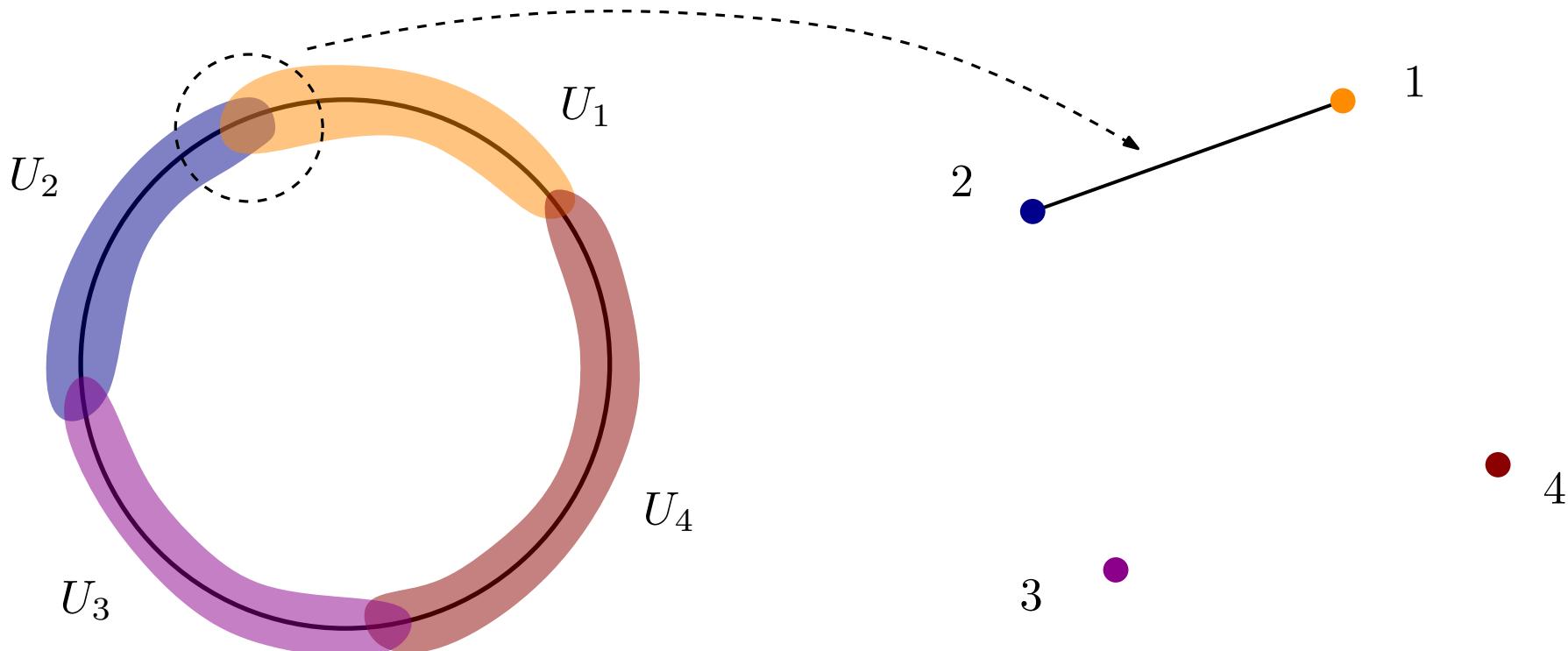
Le **nerf** de  $\mathcal{U}$  est le complexe simplicial avec ensemble de sommets  $\{1, \dots, N\}$  et dont les  $m$ -simplexes sont les sous-ensembles  $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$  tels que  $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$ . On le note  $\mathcal{N}(\mathcal{U})$ .



**Définition :** Soit  $X$  un espace topologique, et  $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$  un *recouvrement* de  $X$ , c'est-à-dire une collection de sous-ensembles  $U_i \subset X$  telle que

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

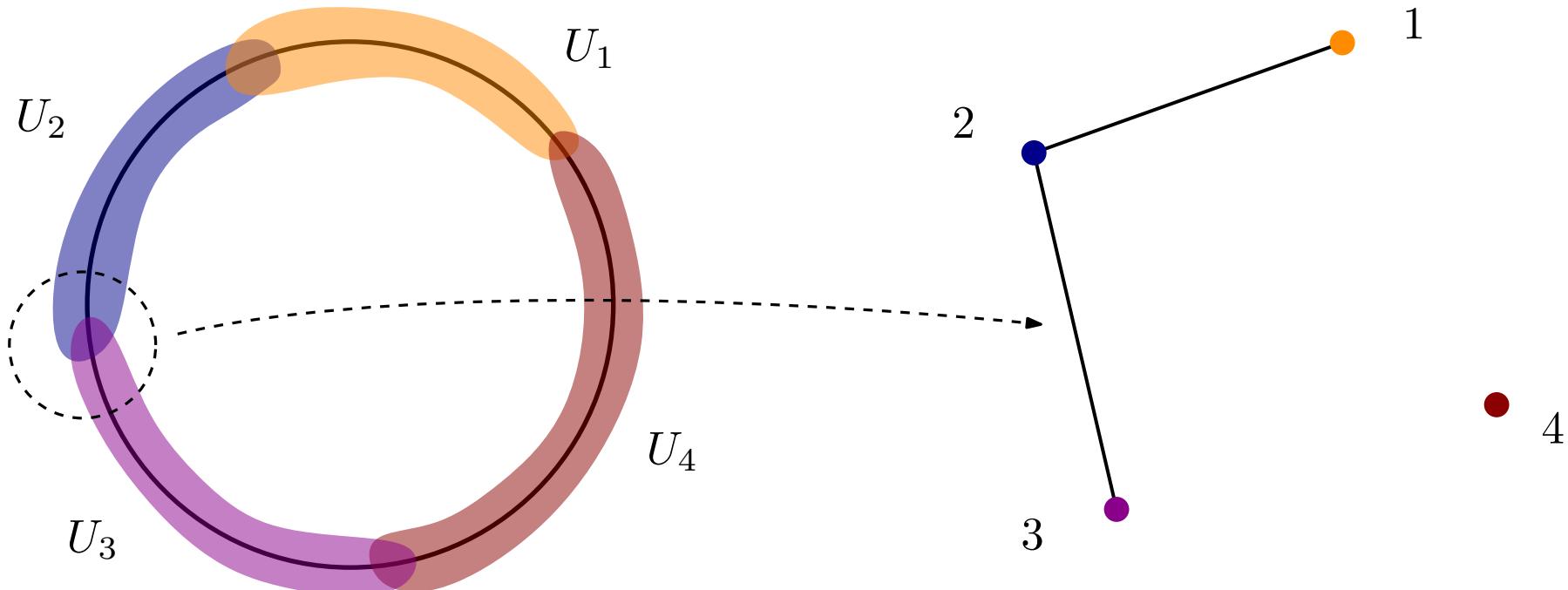
Le **nerf** de  $\mathcal{U}$  est le complexe simplicial avec ensemble de sommets  $\{1, \dots, N\}$  et dont les  $m$ -simplexes sont les sous-ensembles  $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$  tels que  $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$ . On le note  $\mathcal{N}(\mathcal{U})$ .



**Définition :** Soit  $X$  un espace topologique, et  $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$  un *recouvrement* de  $X$ , c'est-à-dire une collection de sous-ensembles  $U_i \subset X$  telle que

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

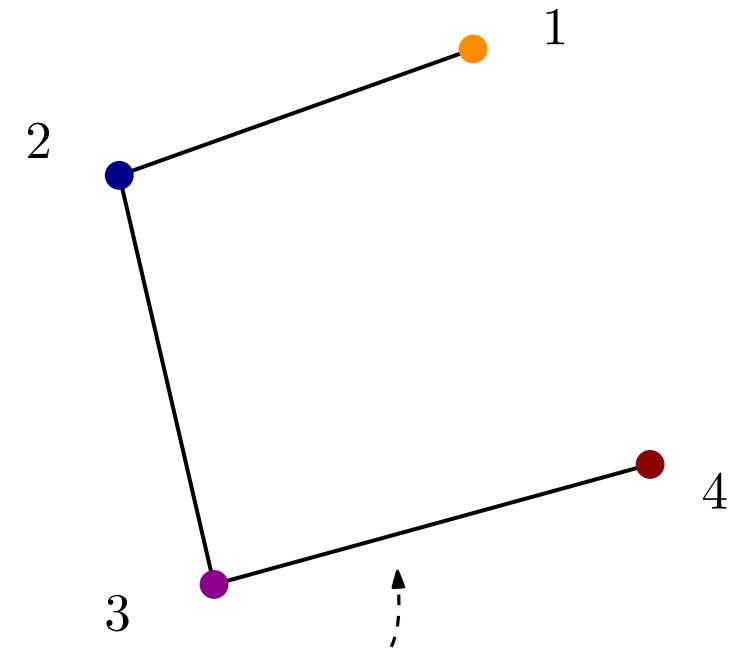
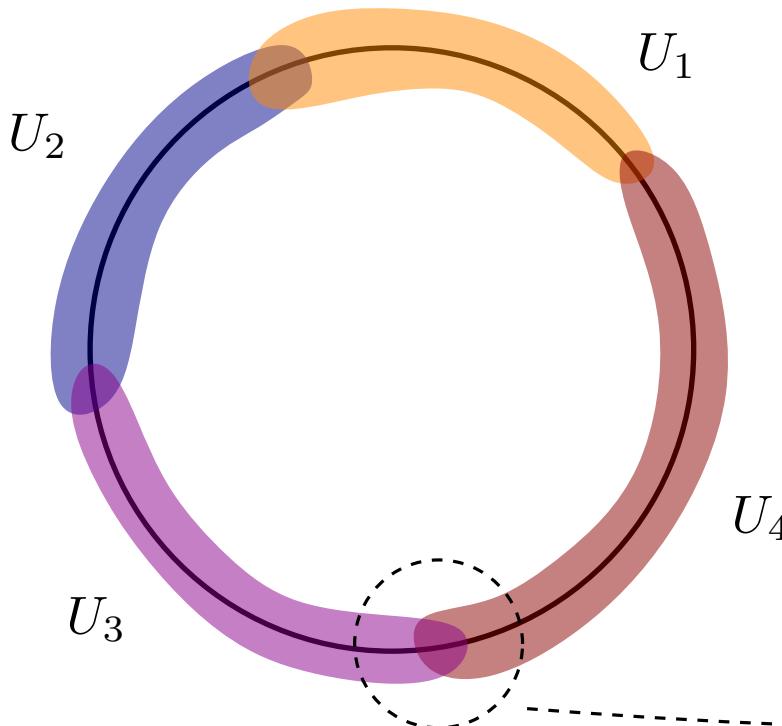
Le **nerf** de  $\mathcal{U}$  est le complexe simplicial avec ensemble de sommets  $\{1, \dots, N\}$  et dont les  $m$ -simplexes sont les sous-ensembles  $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$  tels que  $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$ . On le note  $\mathcal{N}(\mathcal{U})$ .



**Définition :** Soit  $X$  un espace topologique, et  $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$  un *recouvrement* de  $X$ , c'est-à-dire une collection de sous-ensembles  $U_i \subset X$  telle que

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

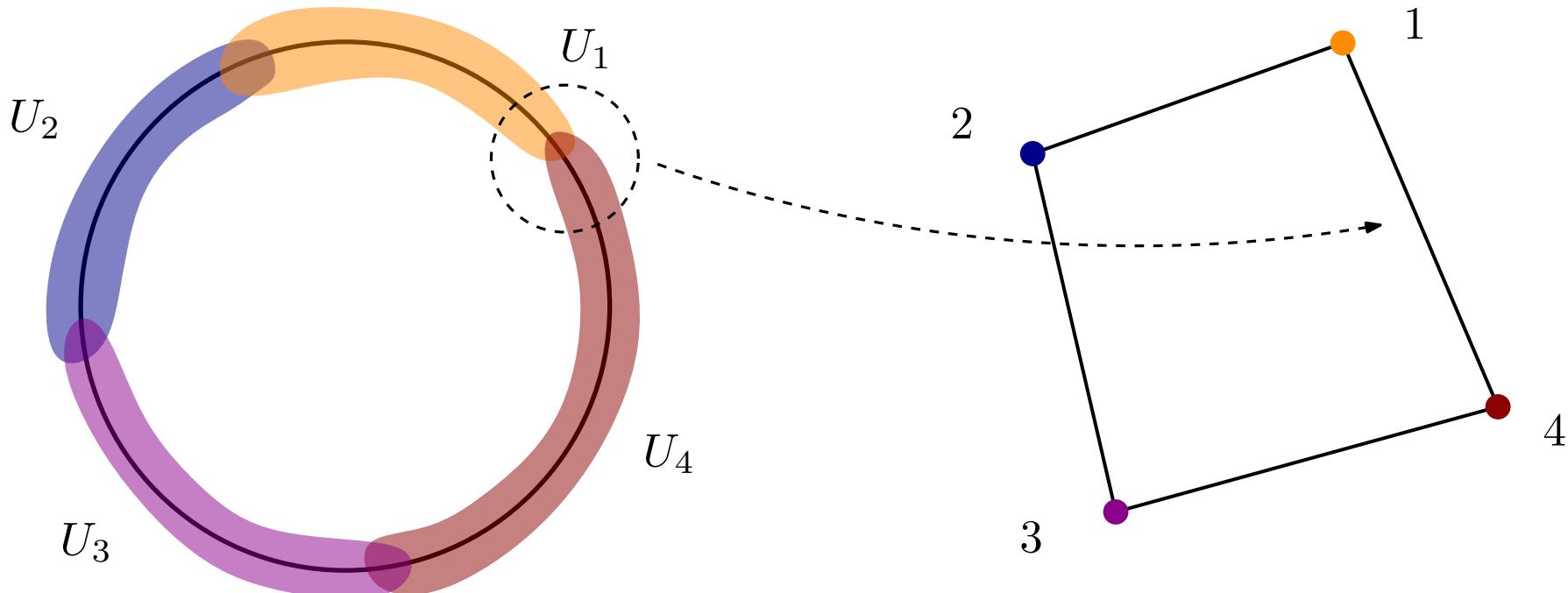
Le **nerf** de  $\mathcal{U}$  est le complexe simplicial avec ensemble de sommets  $\{1, \dots, N\}$  et dont les  $m$ -simplexes sont les sous-ensembles  $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$  tels que  $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$ . On le note  $\mathcal{N}(\mathcal{U})$ .



**Définition :** Soit  $X$  un espace topologique, et  $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$  un *recouvrement* de  $X$ , c'est-à-dire une collection de sous-ensembles  $U_i \subset X$  telle que

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

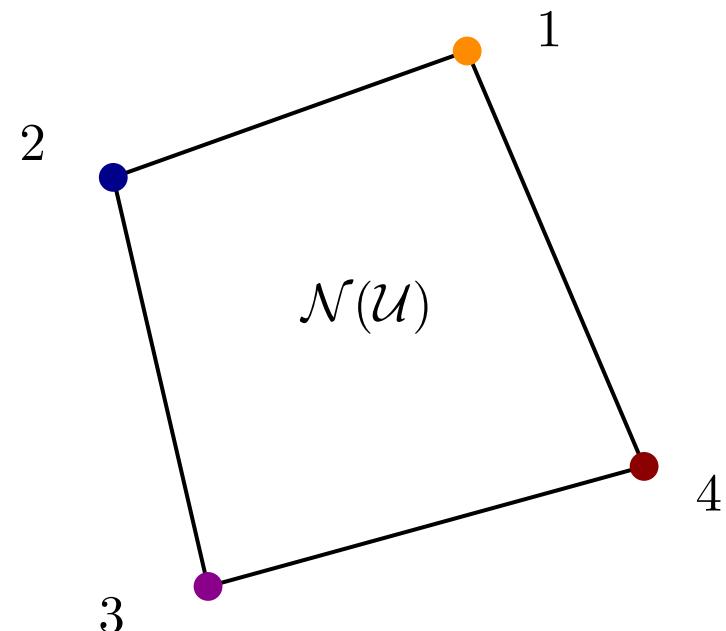
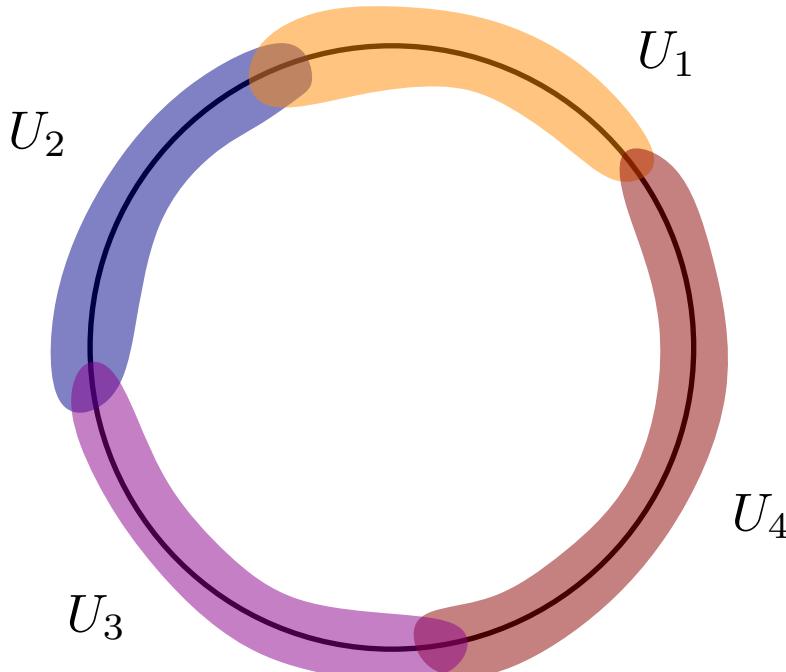
Le **nerf** de  $\mathcal{U}$  est le complexe simplicial avec ensemble de sommets  $\{1, \dots, N\}$  et dont les  $m$ -simplexes sont les sous-ensembles  $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$  tels que  $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$ . On le note  $\mathcal{N}(\mathcal{U})$ .



**Définition :** Soit  $X$  un espace topologique, et  $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$  un *recouvrement* de  $X$ , c'est-à-dire une collection de sous-ensembles  $U_i \subset X$  telle que

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

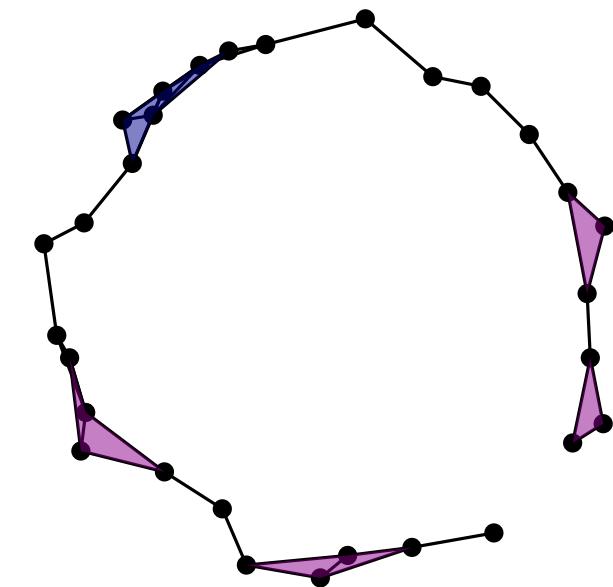
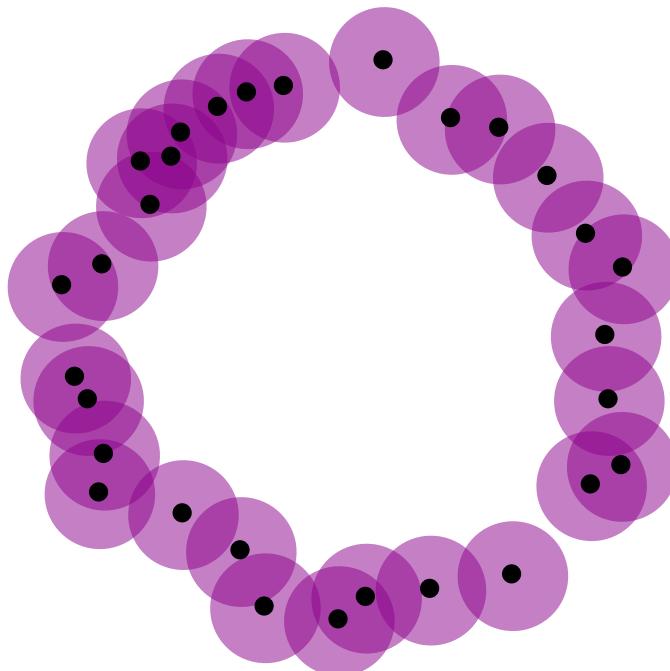
Le **nerf** de  $\mathcal{U}$  est le complexe simplicial avec ensemble de sommets  $\{1, \dots, N\}$  et dont les  $m$ -simplexes sont les sous-ensembles  $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$  tels que  $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$ . On le note  $\mathcal{N}(\mathcal{U})$ .



**Définition :** Soit  $X$  un espace topologique, et  $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$  un *recouvrement* de  $X$ , c'est-à-dire une collection de sous-ensembles  $U_i \subset X$  telle que

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

Le **nerf** de  $\mathcal{U}$  est le complexe simplicial avec ensemble de sommets  $\{1, \dots, N\}$  et dont les  $m$ -simplexes sont les sous-ensembles  $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$  tels que  $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$ . On le note  $\mathcal{N}(\mathcal{U})$ .

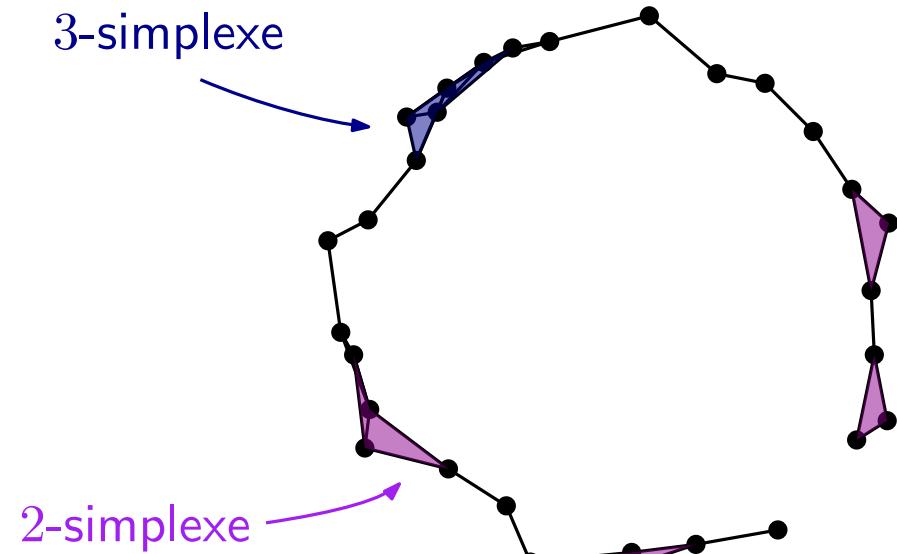
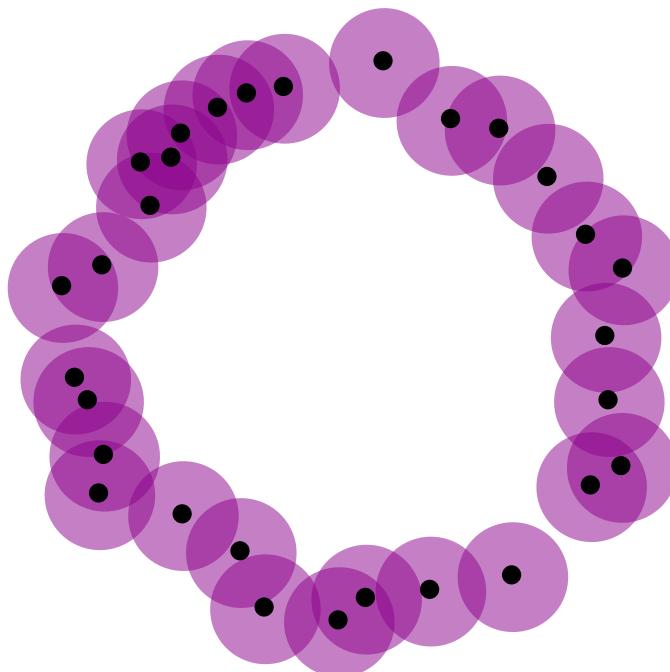


$X^{0.2} = \bigcup_{x \in X} \overline{\mathcal{B}}(x, 0.2)$  est recouvert par  $\mathcal{U} = \{\overline{\mathcal{B}}(x, 0.2), x \in X\}$

**Définition :** Soit  $X$  un espace topologique, et  $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$  un *recouvrement* de  $X$ , c'est-à-dire une collection de sous-ensembles  $U_i \subset X$  telle que

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

Le **nerf** de  $\mathcal{U}$  est le complexe simplicial avec ensemble de sommets  $\{1, \dots, N\}$  et dont les  $m$ -simplexes sont les sous-ensembles  $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$  tels que  $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$ . On le note  $\mathcal{N}(\mathcal{U})$ .

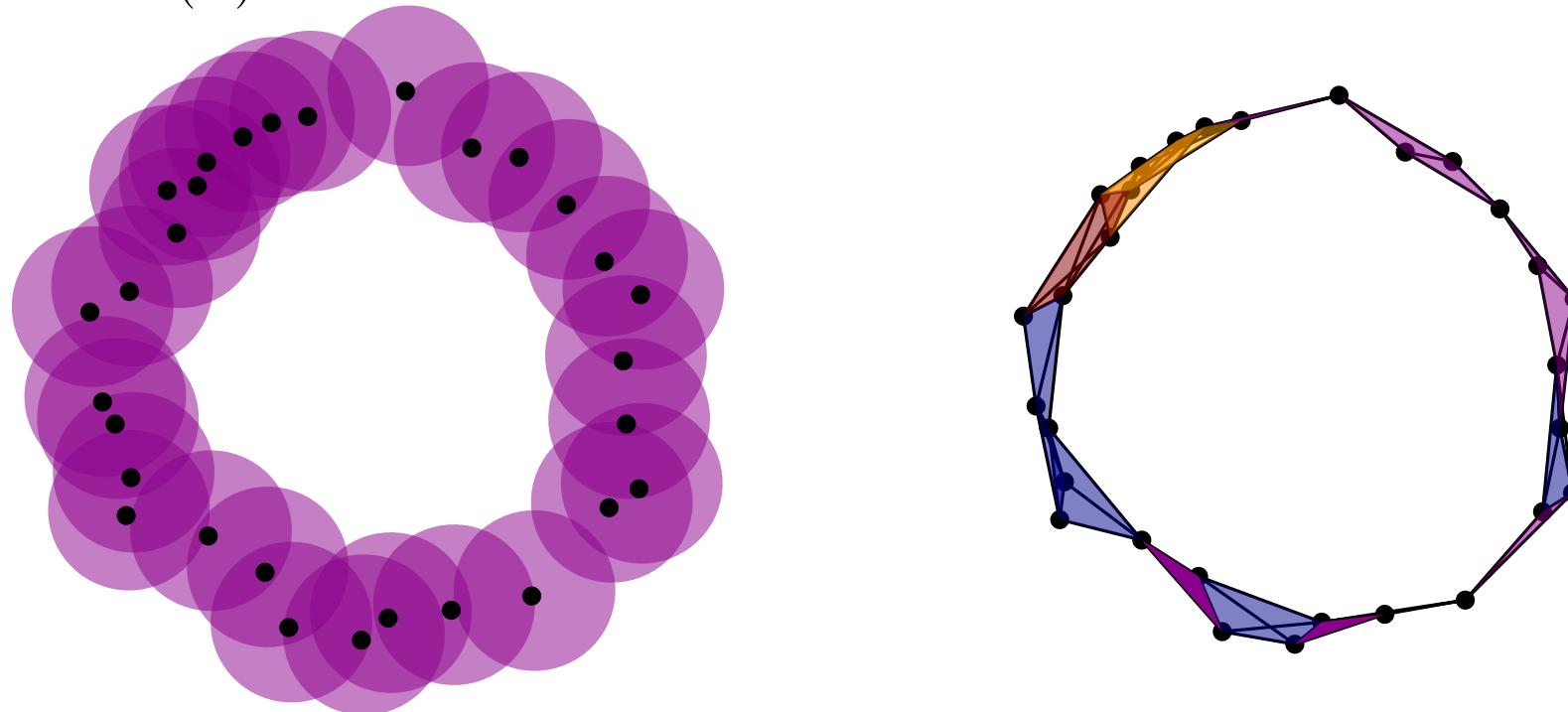


$X^{0.2} = \bigcup_{x \in X} \overline{\mathcal{B}}(x, 0.2)$  est recouvert par  $\mathcal{U} = \{\overline{\mathcal{B}}(x, 0.2), x \in X\}$

**Définition :** Soit  $X$  un espace topologique, et  $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$  un *recouvrement* de  $X$ , c'est-à-dire une collection de sous-ensembles  $U_i \subset X$  telle que

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

Le **nerf** de  $\mathcal{U}$  est le complexe simplicial avec ensemble de sommets  $\{1, \dots, N\}$  et dont les  $m$ -simplexes sont les sous-ensembles  $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$  tels que  $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$ . On le note  $\mathcal{N}(\mathcal{U})$ .

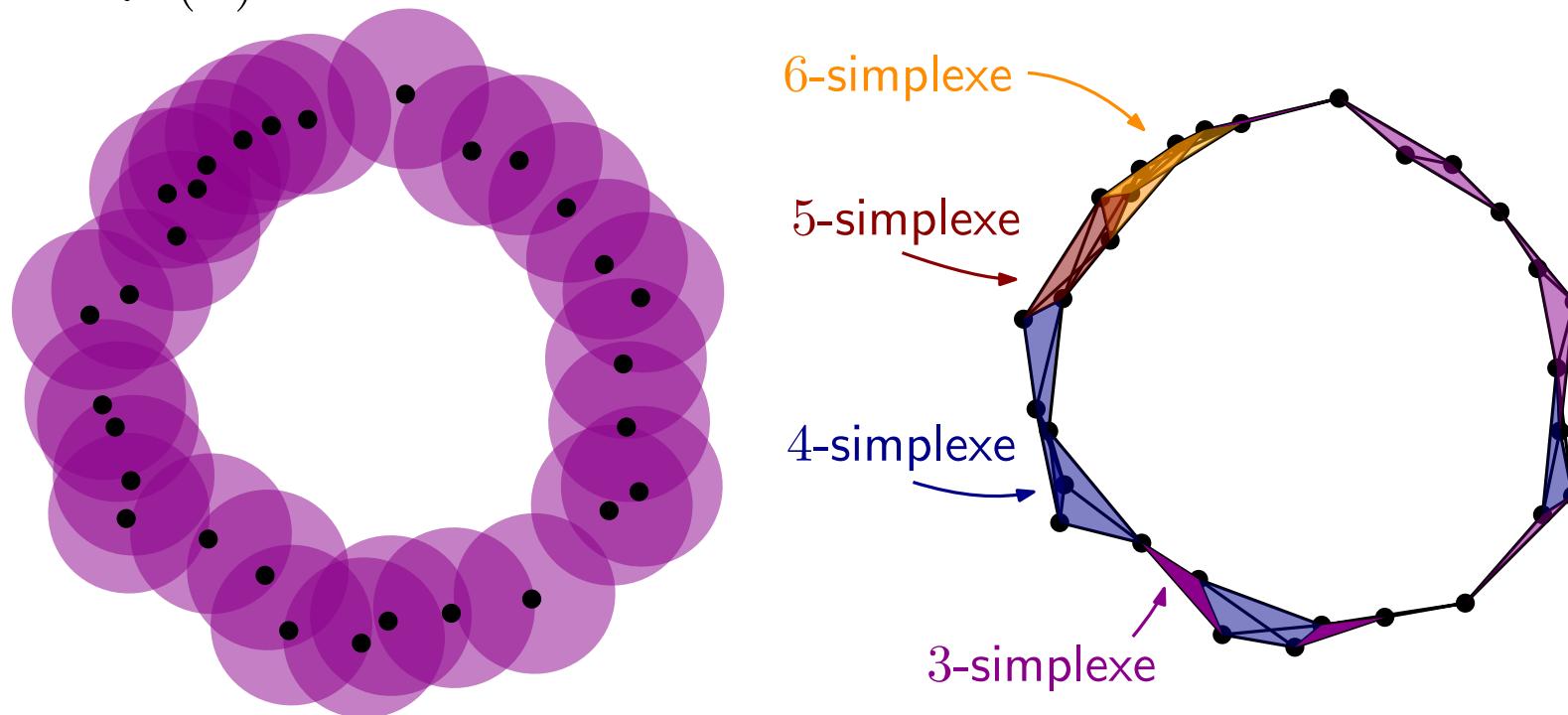


$X^{0.3} = \bigcup_{x \in X} \overline{\mathcal{B}}(x, 0.3)$  est recouvert par  $\mathcal{U} = \{\overline{\mathcal{B}}(x, 0.3), x \in X\}$

**Définition :** Soit  $X$  un espace topologique, et  $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$  un *recouvrement* de  $X$ , c'est-à-dire une collection de sous-ensembles  $U_i \subset X$  telle que

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

Le **nerf** de  $\mathcal{U}$  est le complexe simplicial avec ensemble de sommets  $\{1, \dots, N\}$  et dont les  $m$ -simplexes sont les sous-ensembles  $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$  tels que  $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$ . On le note  $\mathcal{N}(\mathcal{U})$ .



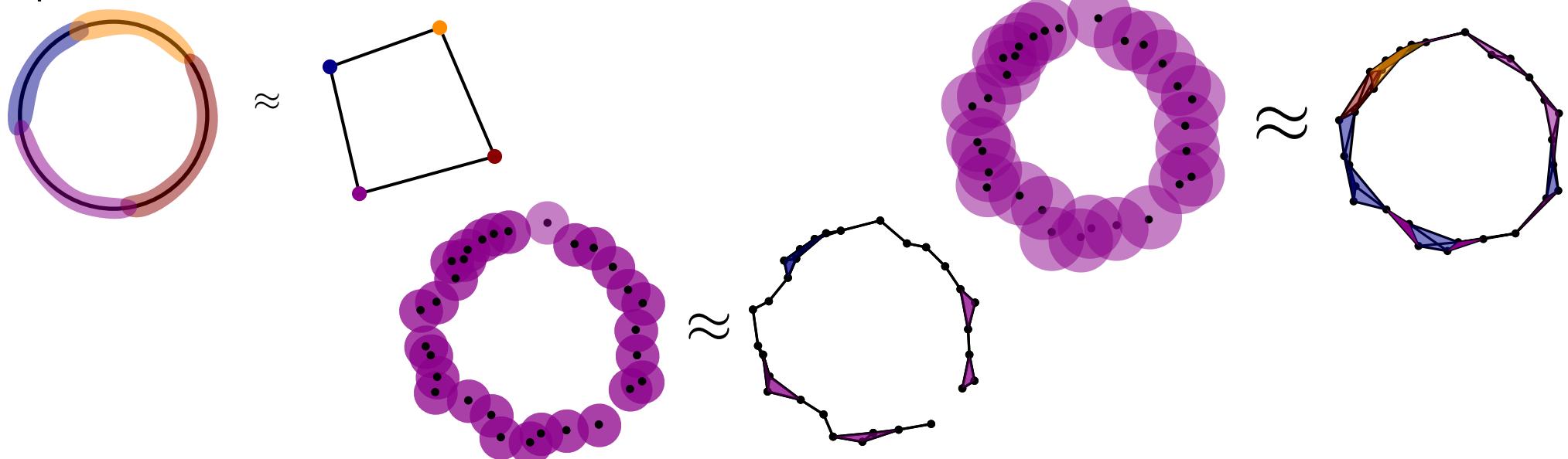
$X^{0.3} = \bigcup_{x \in X} \overline{\mathcal{B}}(x, 0.3)$  est recouvert par  $\mathcal{U} = \{\overline{\mathcal{B}}(x, 0.3), x \in X\}$

**Définition :** Soit  $X$  un espace topologique, et  $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$  un *recouvrement* de  $X$ , c'est-à-dire une collection de sous-ensembles  $U_i \subset X$  telle que

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

Le **nerf** de  $\mathcal{U}$  est le complexe simplicial avec ensemble de sommets  $\{1, \dots, N\}$  et dont les  $m$ -simplexes sont les sous-ensembles  $\{i_1, \dots, i_m\} \subset \{1, \dots, N\}$  tels que  $\bigcap_{k=0}^m U_{i_k} \neq \emptyset$ . On le note  $\mathcal{N}(\mathcal{U})$ .

**Théorème du nerf :** Soit  $X \subset \mathbb{R}^n$ . Supposons que chaque  $U_i$  soit une boule (ou plus généralement, ou sous-ensemble fermé et convexe). Alors  $\mathcal{N}(\mathcal{U})$  est homotopiquement équivalent à  $X$ .



# Complexe de Čech

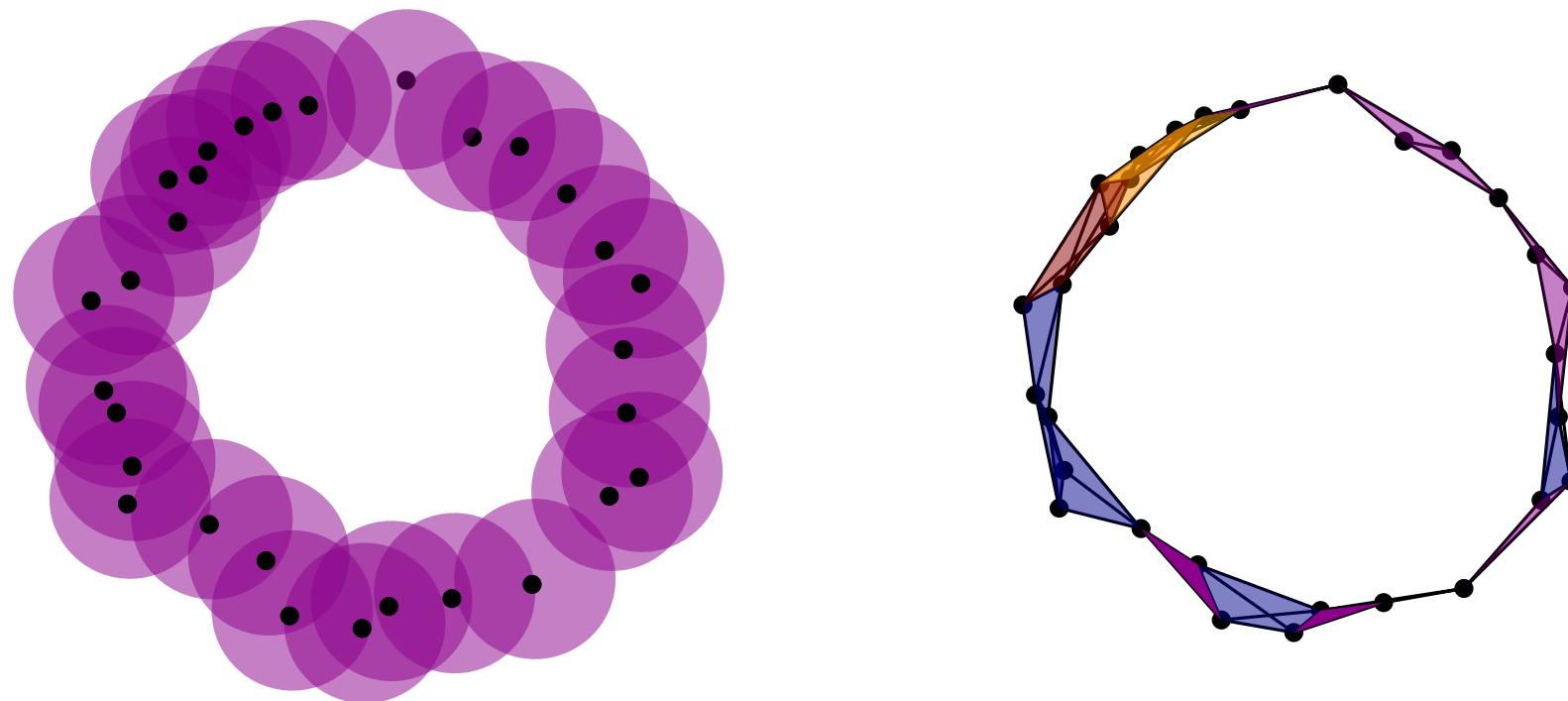
21/43 (1/2)

Soit  $X$  un sous-ensemble fini de  $\mathbb{R}^n$ , et  $t \geq 0$ . Soit la collection

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

C'est un recouvrement de l'épaisseur  $X^t$ , dont chaque élément est une boule fermée. Par le théorème du nerf, son nerf  $\mathcal{N}(\mathcal{V}^t)$  est homotopiquement équivalent à  $X^t$ .

**Définition :** Ce nerf est noté  $\check{\text{C}}\text{ech}^t(X)$  et est appelé le **complexe de Čech au temps  $t$** .

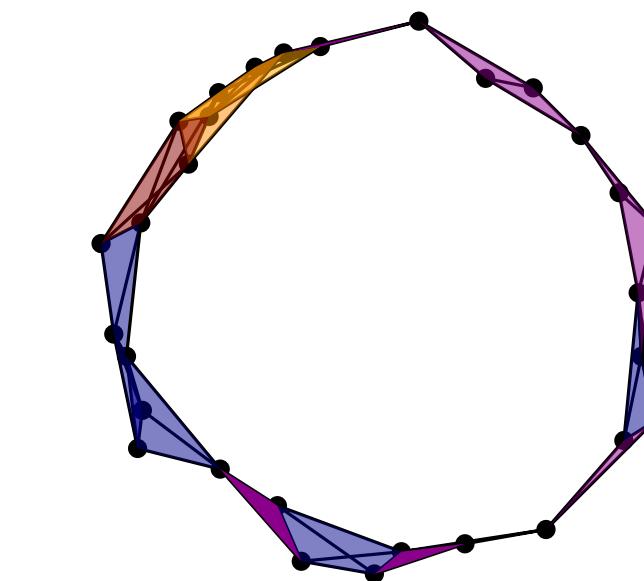
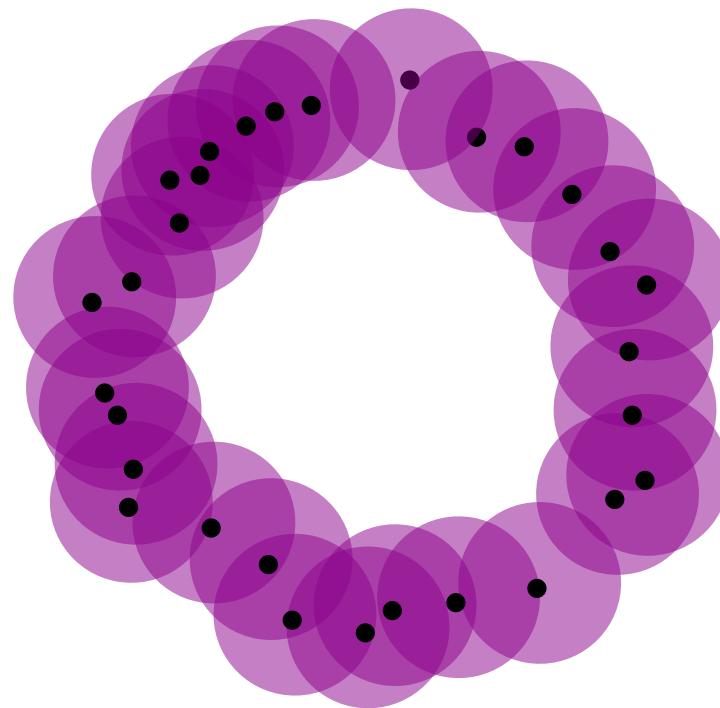


Soit  $X$  un sous-ensemble fini de  $\mathbb{R}^n$ , et  $t \geq 0$ . Soit la collection

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

C'est un recouvrement de l'épaisseur  $X^t$ , dont chaque élément est une boule fermée. Par le théorème du nerf, son nerf  $\mathcal{N}(\mathcal{V}^t)$  est homotopiquement équivalent à  $X^t$ .

**Définition :** Ce nerf est noté  $\check{\text{Cech}}^t(X)$  et est appelé le **complexe de Čech au temps  $t$** .



→ La Question 2 (Comment calculer les groupes d'homologie de  $X^t$  ?) est résolue.

# I - Homologie simpliciale

- 1 - Groupes d'homologie
- 2 - Fonctorialité

# II - Inférence topologique

- 1 - Estimation du paramètre
- 2 - Nerfs

# III - Homologie persistante

- 1 - Modules de persistance
- 2 - Décomposition
- 3 - Stabilité

# IV - Applications

# Le problème de l'échelle

23/43 (1/4)

Question 1 : Comment sélectionner un  $t$  tel que  $X^t \approx \mathcal{M}$ ?

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

Soit  $X$  et  $\mathcal{M}$  deux sous-ensembles de  $\mathbb{R}^n$ . Supposons que  $\mathcal{M}$  ait un portée positive, et que  $d_H(X, \mathcal{M}) \leq \frac{1}{17}\text{portée}(\mathcal{M})$ .

Alors  $X^t$  et  $\mathcal{M}$  sont homotopiquement équivalents, pourvu que

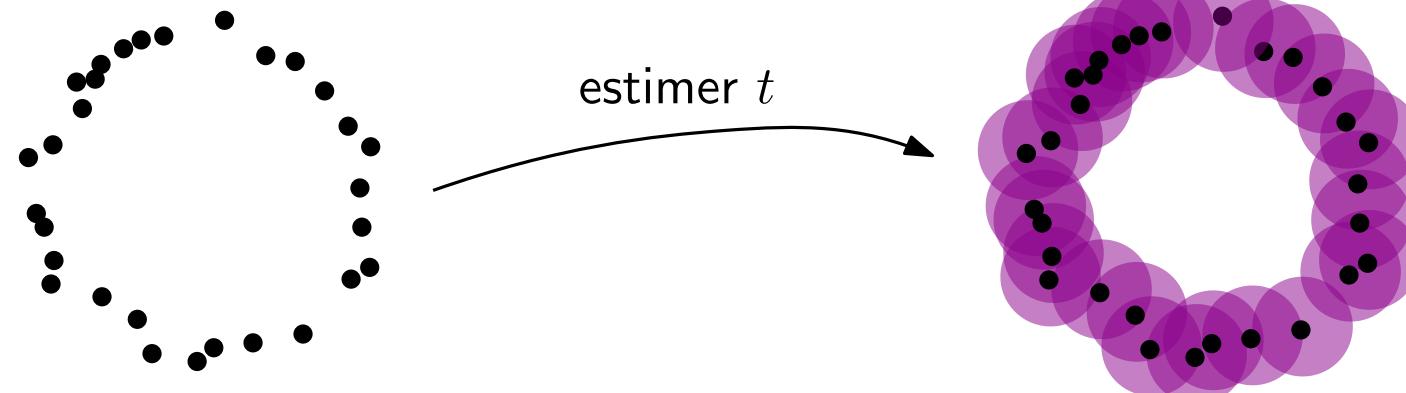
$$t \in [4d_H(X, \mathcal{M}), \text{portée}(\mathcal{M}) - 3d_H(X, \mathcal{M})].$$

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

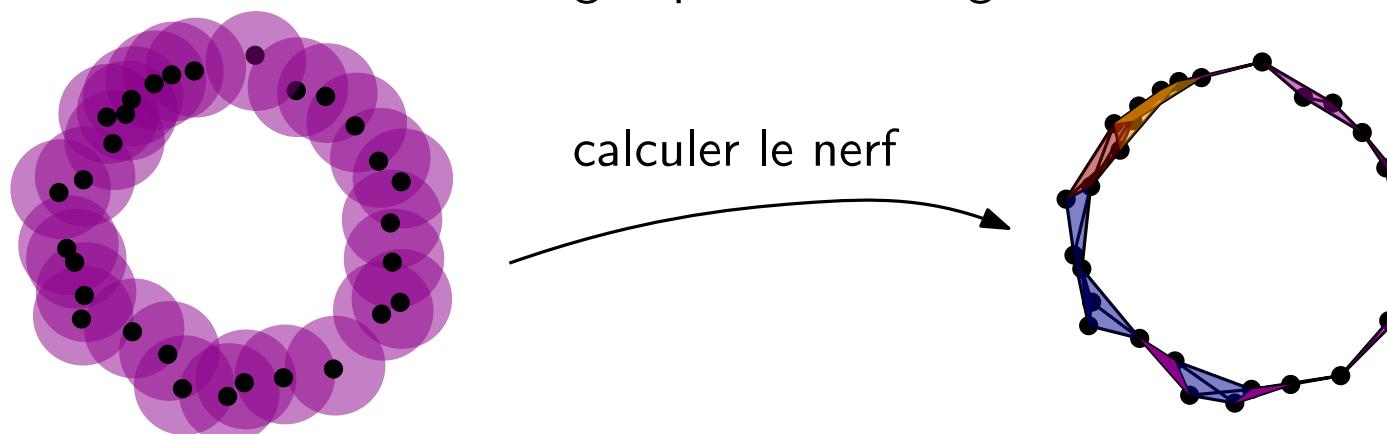
Soit  $X$  et  $\mathcal{M}$  deux sous-ensembles de  $\mathbb{R}^n$ , avec  $\mathcal{M}$  une sous-variété, et  $X$  un sous-ensemble fini de  $\mathcal{M}$ . Supposons que  $\mathcal{M}$  ait une portée positive.

Alors  $X^t$  et  $\mathcal{M}$  sont homotopiquement équivalents, pourvu que

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



Question 2 : Comment calculer les groupes d'homologie de  $X^t$ ?



# Le problème de l'échelle

23/43 (2/4)

Question 1 : Comment sélectionner un  $t$  tel que  $X^t \approx \mathcal{M}$ ?

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

Soit  $X$  et  $\mathcal{M}$  deux sous-ensembles de  $\mathbb{R}^n$ . Supposons que  $\mathcal{M}$  ait un portée positive, et que  $d_H(X, \mathcal{M}) \leq \frac{1}{17}\text{portée}(\mathcal{M})$ .

Alors  $X^t$  et  $\mathcal{M}$  sont homotopiquement équivalents, pourvu que

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

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

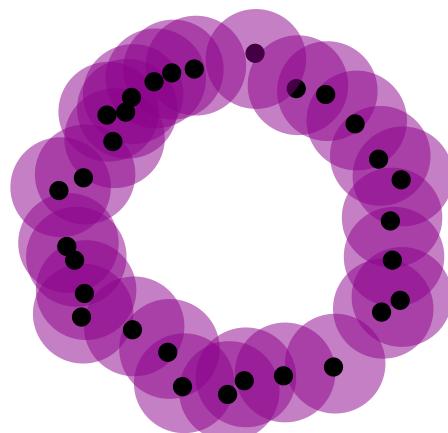
Soit  $X$  et  $\mathcal{M}$  deux sous-ensembles de  $\mathbb{R}^n$ , avec  $\mathcal{M}$  une sous-variété, et  $X$  un sous-ensemble fini de  $\mathcal{M}$ . Supposons que  $\mathcal{M}$  ait une portée positive.

Alors  $X^t$  et  $\mathcal{M}$  sont homotopiquement équivalents, pourvu que

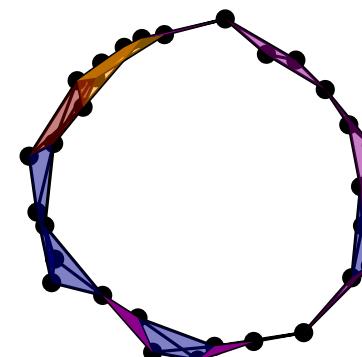
$$t \in \left[ 2d_H(X, \mathcal{M}), \sqrt{\frac{3}{5}}\text{portée}(\mathcal{M}) \right].$$

on ne connaît pas ces quantités !

Question 2 : Comment calculer les groupes d'homologie de  $X^t$ ?



calculer le nerf



# Le problème de l'échelle

23/43 (3/4)

Question 1 : Comment sélectionner un  $t$  tel que  $X^t \approx \mathcal{M}$ ?

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

Soit  $X$  et  $\mathcal{M}$  deux sous-ensembles de  $\mathbb{R}^n$ . Supposons que  $\mathcal{M}$  ait un portée positive, et que  $d_H(X, \mathcal{M}) \leq \frac{1}{17}\text{portée}(\mathcal{M})$ .

Alors  $X^t$  et  $\mathcal{M}$  sont homotopiquement équivalents, pourvu que

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



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

Soit  $X$  et  $\mathcal{M}$  deux sous-ensembles de  $\mathbb{R}^n$ , avec  $\mathcal{M}$  une sous-variété, et  $X$  un sous-ensemble fini de  $\mathcal{M}$ . Supposons que  $\mathcal{M}$  ait une portée positive.

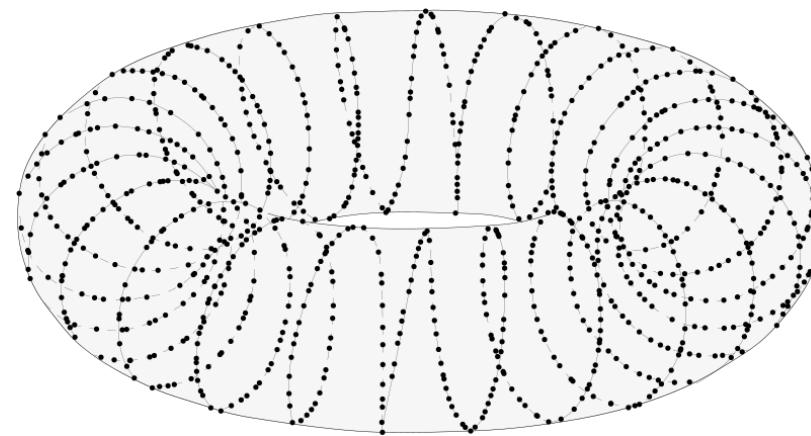
Alors  $X^t$  et  $\mathcal{M}$  sont homotopiquement équivalents, pourvu que

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



on ne connaît pas ces quantités !

Cet objet est-il de dimension 1 ou 2 ?



# Le problème de l'échelle

23/43 (4/4)

Question 1 : Comment sélectionner un  $t$  tel que  $X^t \approx \mathcal{M}$ ?

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

Soit  $X$  et  $\mathcal{M}$  deux sous-ensembles de  $\mathbb{R}^n$ . Supposons que  $\mathcal{M}$  ait un portée positive, et que  $d_H(X, \mathcal{M}) \leq \frac{1}{17}\text{portée}(\mathcal{M})$ .

Alors  $X^t$  et  $\mathcal{M}$  sont homotopiquement équivalents, pourvu que

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



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

Soit  $X$  et  $\mathcal{M}$  deux sous-ensembles de  $\mathbb{R}^n$ , avec  $\mathcal{M}$  une sous-variété, et  $X$  un sous-ensemble fini de  $\mathcal{M}$ . Supposons que  $\mathcal{M}$  ait une portée positive.

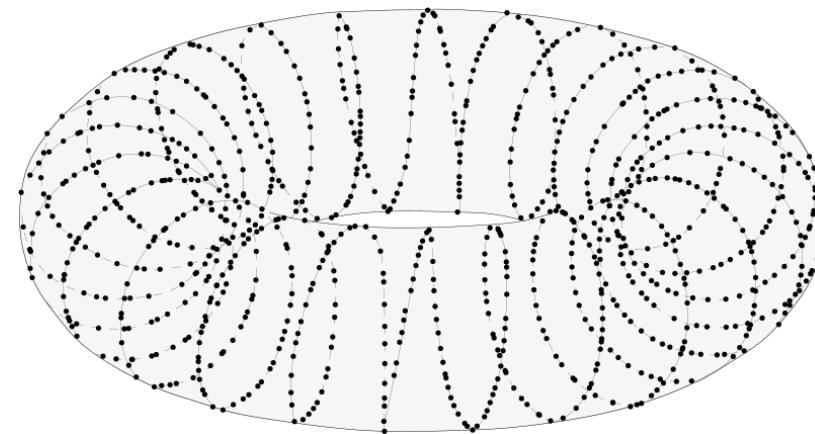
Alors  $X^t$  et  $\mathcal{M}$  sont homotopiquement équivalents, pourvu que

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



on ne connaît pas ces quantités !

Cet objet est-il de dimension 1 ou 2 ?



Idée (analyse multi-échelle) : À la place de choisir une valeur pour  $t$ , on va toutes les choisir.

Définition : La **filtration de Čech** de  $X$  est la collection  $V[X] = (X^t)_{t \geq 0}$ .

# I - Homologie simpliciale

- 1 - Groupes d'homologie
- 2 - Fonctorialité

# II - Inférence topologique

- 1 - Estimation du paramètre
- 2 - Nerfs

# III - Homologie persistante

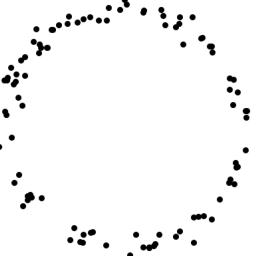
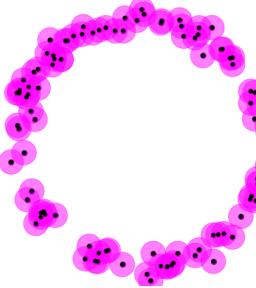
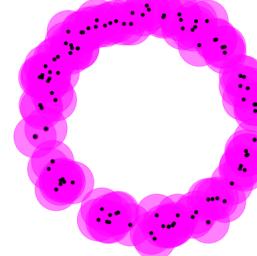
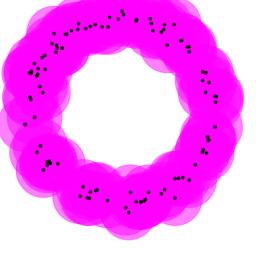
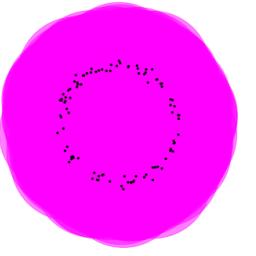
- 1 - Modules de persistance
- 2 - Décomposition
- 3 - Stabilité

# IV - Applications

# Homologie de la filtration de Čech

25/43 (1/2)

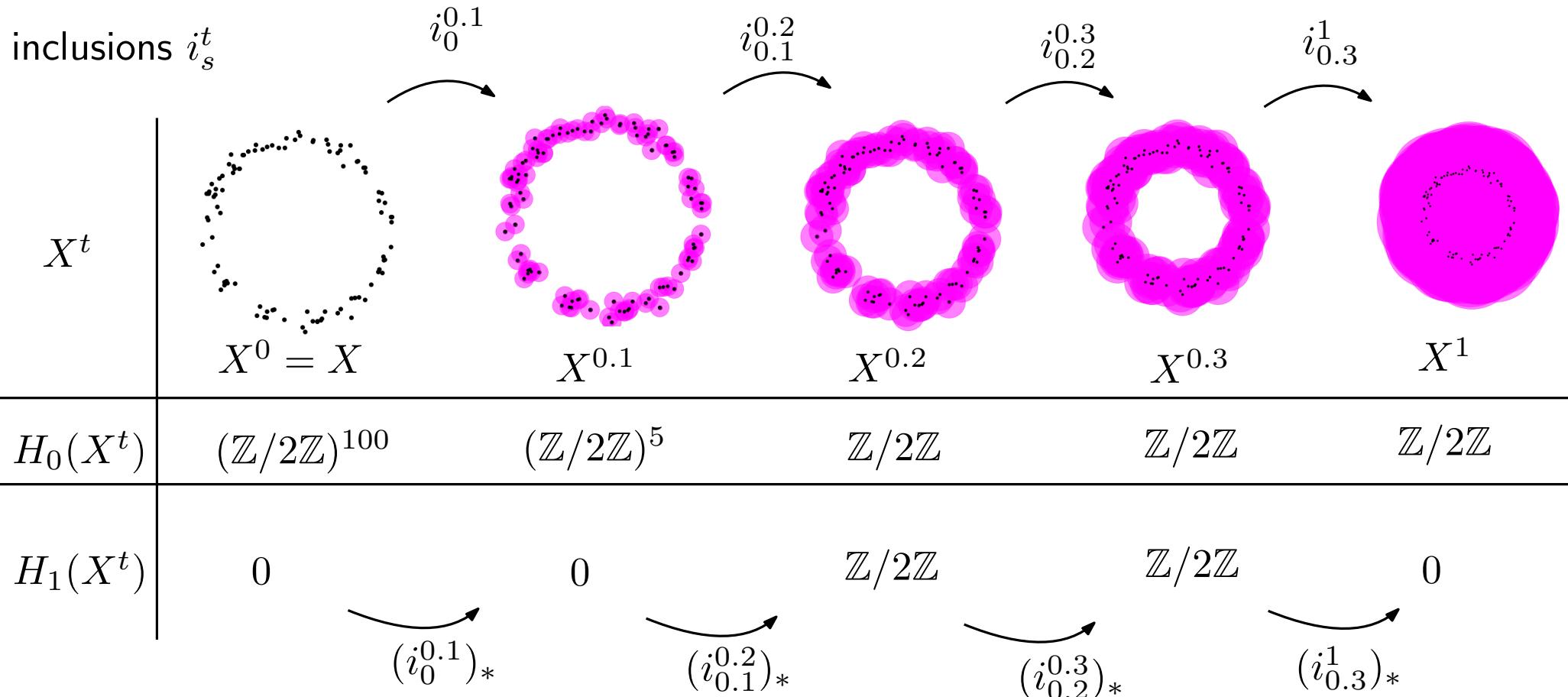
Calculons l'homologie de chaque épaissement :

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

# Homologie de la filtration de Čech

25/43 (2/2)

Calculons l'homologie de chaque épaissement :



La donnée de  $(H_i(X^t))_{t \geq 0}$  et  $((i_s^t)_*)_{s \leq t}$  est appelée un **module de persistance**.

# Modules de persistance

26/43

**Définition :** Un **module de persistance**  $\mathbb{V}$  est une paire  $(\mathbb{V}, v)$  où  $\mathbb{V} = (V^t)_{t \in \mathbb{R}^+}$  est une famille de  $\mathbb{Z}/2\mathbb{Z}$ -espaces vectoriels, et  $v = (v_s^t : V^s \rightarrow V^t)_{s \leq t \in \mathbb{R}^+}$  une famille d'applications linéaires telles que

- pour tout  $t \in \mathbb{R}^+$ ,  $v_t^t : V^t \rightarrow V^t$  est l'application identité,
- pour tout  $r, s, t \in \mathbb{R}^+$  such that  $r \leq s \leq t$ , on a  $v_s^t \circ v_r^s = v_r^t$ .

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

Module de persistance associé à la filtration de Čech :

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

# Suivre l'évolution des trous

27/43 (1/3)

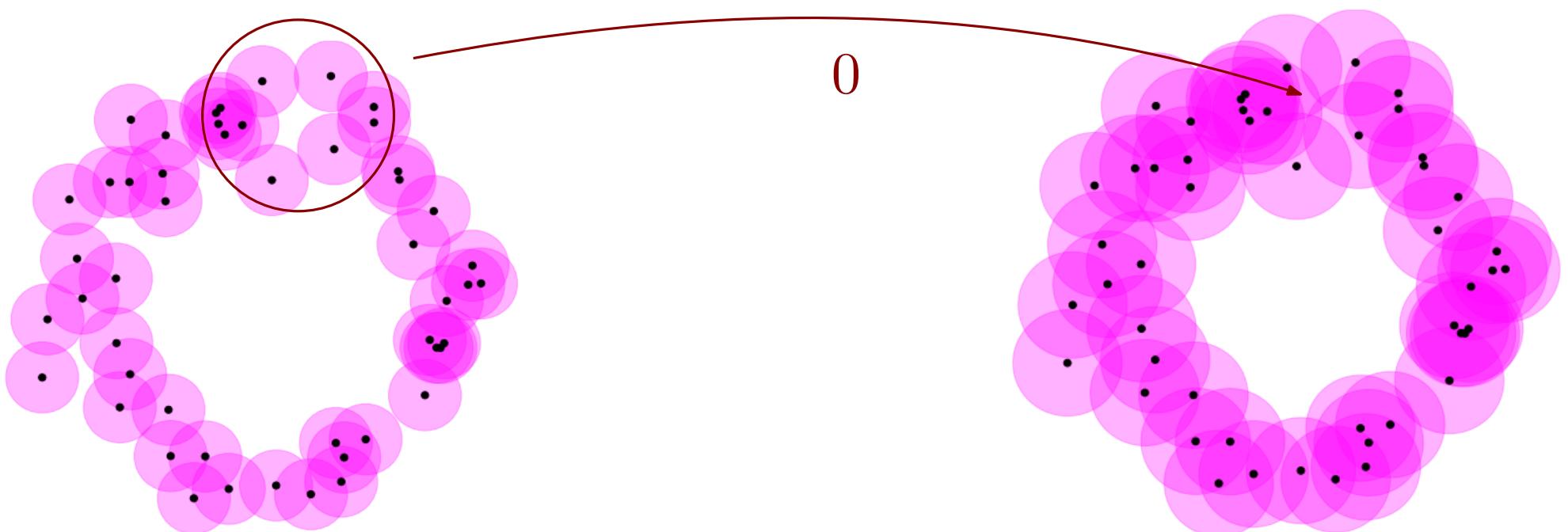
$$\dashrightarrow H_i(X^{t_1}) \xrightarrow{(i_{t_1}^{t_2})_*} H_i(X^{t_2}) \xrightarrow{(i_{t_2}^{t_3})_*} H_i(X^{t_3}) \xrightarrow{(i_{t_3}^{t_4})_*} H_i(X^{t_4}) \dashleftarrow$$

Soit  $i \geq 0$ ,  $t_0 \geq 0$  et considérons un cycle  $c \in H_i(X^{t_0})$ .

Sont **moment de mort** est :  $\sup \{t \geq t_0, (i_{t_0}^t)_*(c) \neq 0\}$ ,

Son **moment de naissance** est :  $\inf \{t \geq t_0, (i_t^{t_0})^{-1}(\{c\})_* \neq \emptyset\}$ ,

Sa **persistance** est la différence.



# Suivre l'évolution des trous

27/43 (2/3)

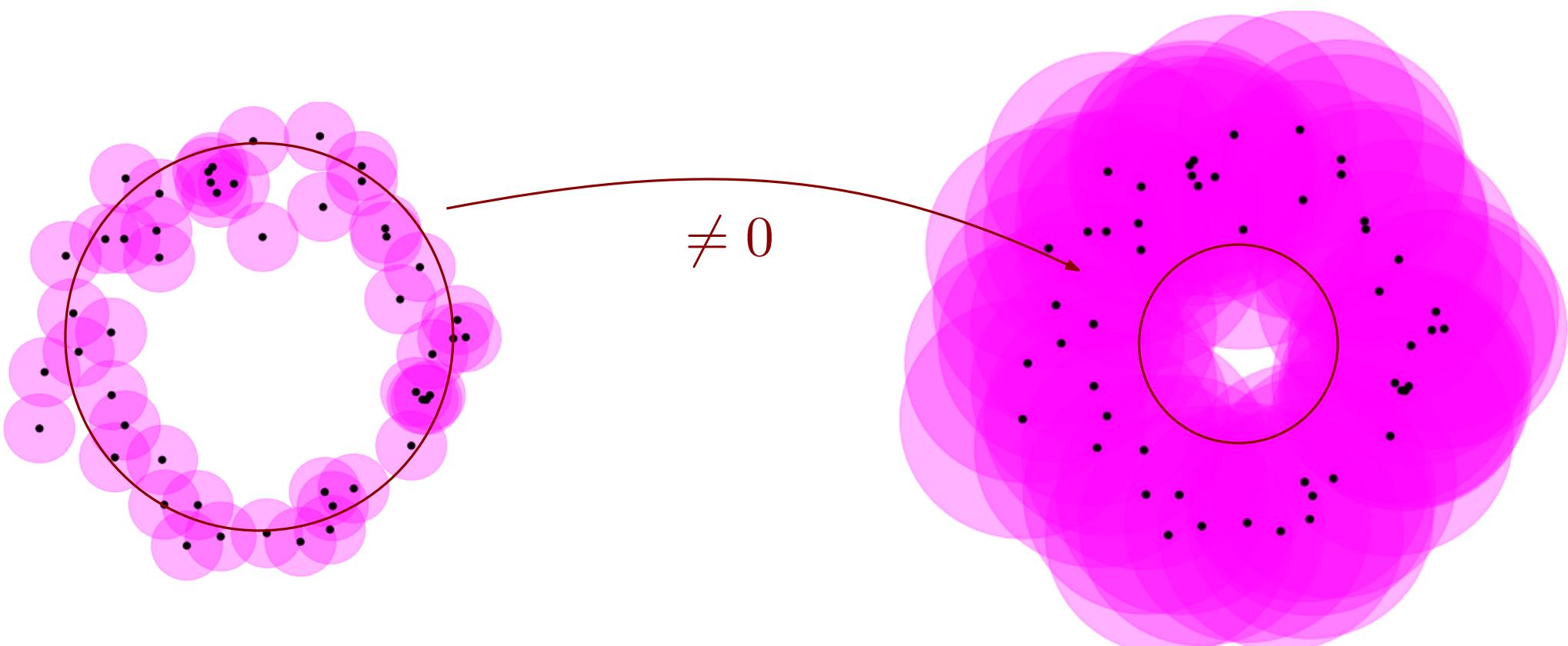
$$\dashrightarrow H_i(X^{t_1}) \xrightarrow{(i_{t_1}^{t_2})_*} H_i(X^{t_2}) \xrightarrow{(i_{t_2}^{t_3})_*} H_i(X^{t_3}) \xrightarrow{(i_{t_3}^{t_4})_*} H_i(X^{t_4}) \dashleftarrow$$

Soit  $i \geq 0$ ,  $t_0 \geq 0$  et considérons un cycle  $c \in H_i(X^{t_0})$ .

Sont **moment de mort** est :  $\sup \{t \geq t_0, (i_{t_0}^t)_*(c) \neq 0\}$ ,

Son **moment de naissance** est :  $\inf \{t \geq t_0, (i_t^{t_0})^{-1}(\{c\})_* \neq \emptyset\}$ ,

Sa **persistance** est la différence.



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

Soit  $i \geq 0$ ,  $t_0 \geq 0$  et considérons un cycle  $c \in H_i(X^{t_0})$ .

Sont **moment de mort** est :  $\sup \{t \geq t_0, (i_{t_0}^t)_*(c) \neq 0\}$ ,

Son **moment de naissance** est :  $\inf \{t \geq t_0, (i_t^{t_0})^{-1}(\{c\})_* \neq \emptyset\}$ ,

Sa **persistance** est la différence.

## Interprétation :

- les cycles avec une grande persistance correspondent à des propriétés topologiques importantes du jeu de données,
- les cycles avec une courte persistance correspondent à du bruit topologique.

# I - Homologie simpliciale

- 1 - Groupes d'homologie
- 2 - Fonctorialité

# II - Inférence topologique

- 1 - Estimation du paramètre
- 2 - Nerfs

# III - Homologie persistante

- 1 - Modules de persistance
- 2 - Décomposition
- 3 - Stabilité

# IV - Applications

Théorème (Crawley-Boevey, 2015) :

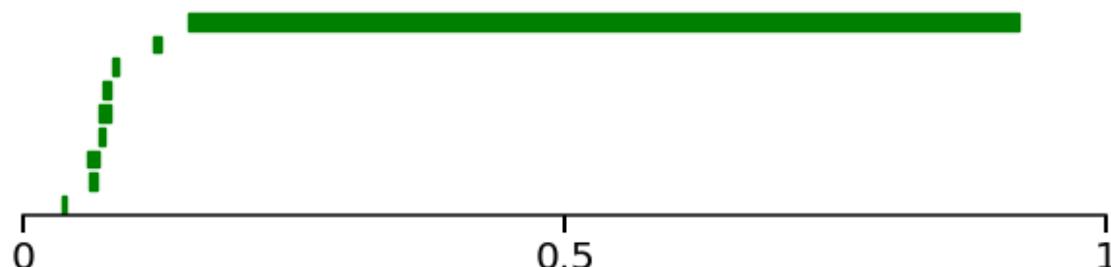
Un module de persistance (régulier) peut être décomposé en module-intervalles.

Ce multi-ensemble d'intervalles s'appelle le **code-barres**. Il décrit complètement le module de persistance.

Module de persistance :  $\mathbb{V}$

Code-barre :  $\{ [0.171, 0.897), [0.035, 0.049), [0.037, 0.046), [0.072, 0.078), [0.077, 0.083), [0.046, 0.050), [0.050, 0.054), [0.036, 0.040), [0.089, 0.092) \}$

Représentation graphique :



# Codes-barres

29/43 (2/3)

Code-barres du module de persistance associé à la filtration de Čech :  $H_0$  en rouge et  $H_1$  en vert.

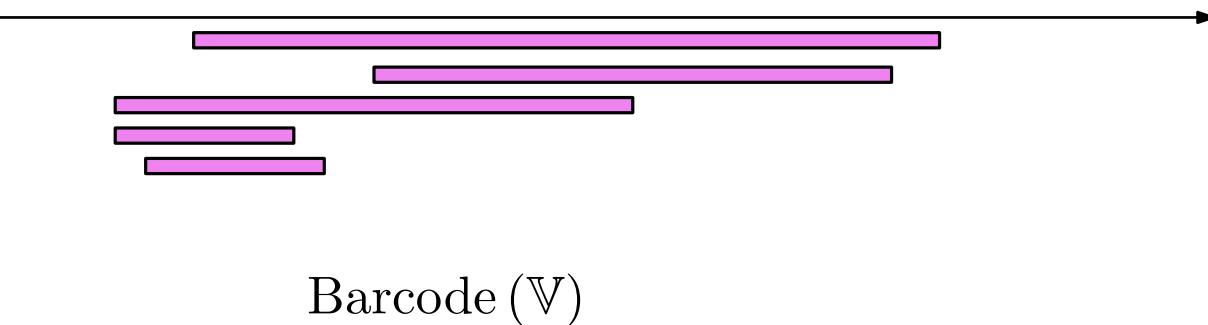
Code-barres du module de persistance associé à la filtration de Čech :  $H_0$  en rouge et  $H_1$  en vert.

Sur un code-barres, on peut lire l'homologie **à chaque moment**, et comment elle **évolue**.

# Diagrammes de persistance

30/43 (1/3)

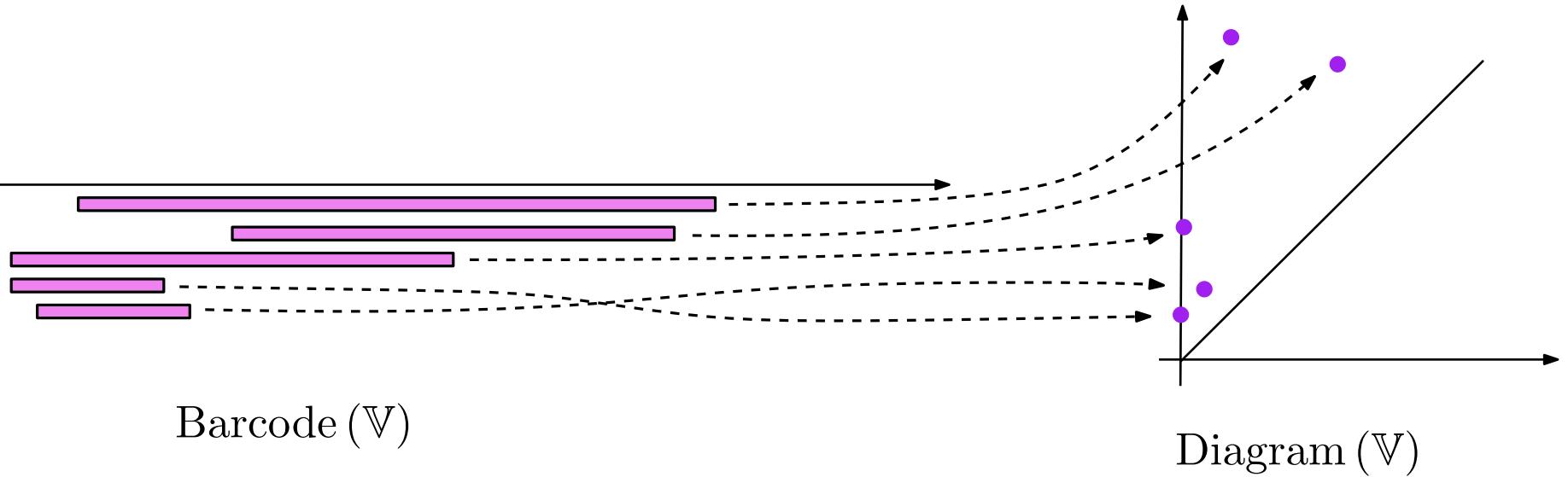
On a associé à tout module de persistance  $\mathbb{V}$  son code-barres



# Diagrammes de persistance

30/43 (2/3)

On a associé à tout module de persistance  $\mathbb{V}$  son code-barres



Pour chaque  $[a, b]$ ,  $(a, b]$ ,  $[a, b)$  ou  $(a, b)$  dans  $\text{Barcode}(\mathbb{V})$ , soit le point  $(a, b)$  de  $\mathbb{R}^2$ . La collection de tout ces points est appelée le **diagramme de persistance** de  $\mathbb{V}$ .

# Diagrammes de persistance

30/43 (3/3)

On a associé à tout module de persistance  $\mathbb{V}$  son code-barres



Barcode ( $\mathbb{V}$ )

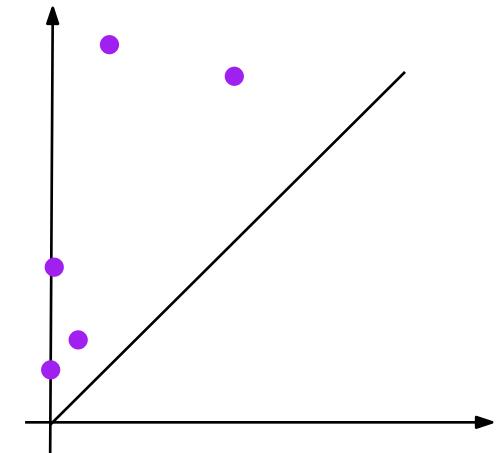


Diagram ( $\mathbb{V}$ )

Pour chaque  $[a, b]$ ,  $(a, b]$ ,  $[a, b)$  ou  $(a, b)$  dans Barcode ( $\mathbb{V}$ ), soit le point  $(a, b)$  de  $\mathbb{R}^2$ . La collection de tout ces points est appelée le **diagramme de persistance** de  $\mathbb{V}$ .

# I - Homologie simpliciale

- 1 - Groupes d'homologie
- 2 - Fonctorialité

# II - Inférence topologique

- 1 - Estimation du paramètre
- 2 - Nerfs

# III - Homologie persistante

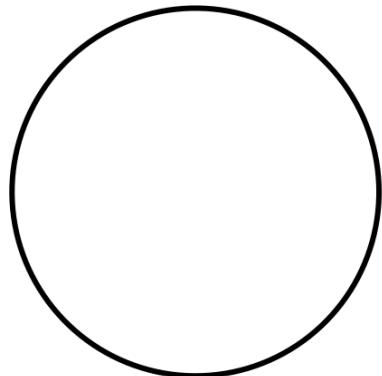
- 1 - Modules de persistance
- 2 - Décomposition
- 3 - Stabilité

# IV - Applications

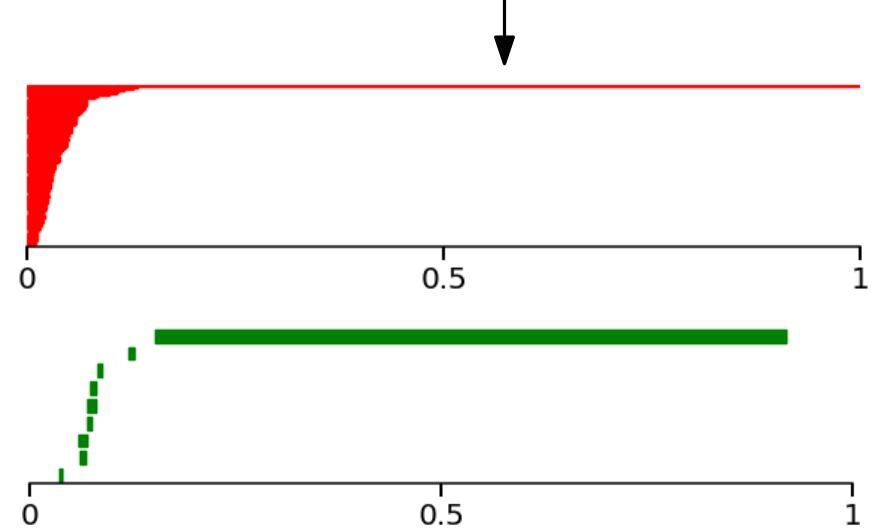
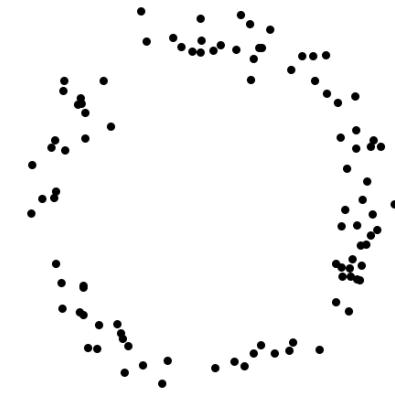
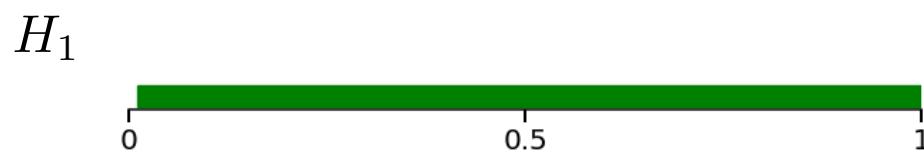
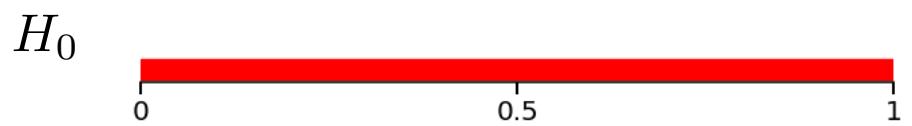
# Stabilité des modules de persistance

32/43

Soit  $X \subset \mathbb{R}^n$  un sous-ensemble fini, vu comme un échantillon de  $\mathcal{M}$ .



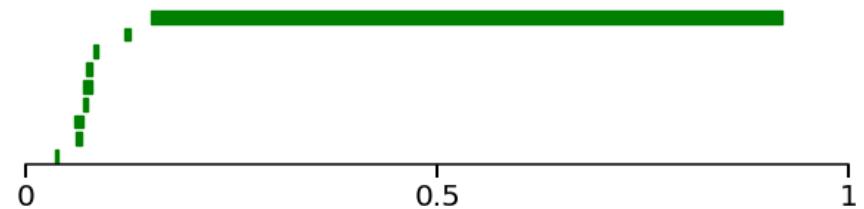
Code-barres de la  
filtration de Čech



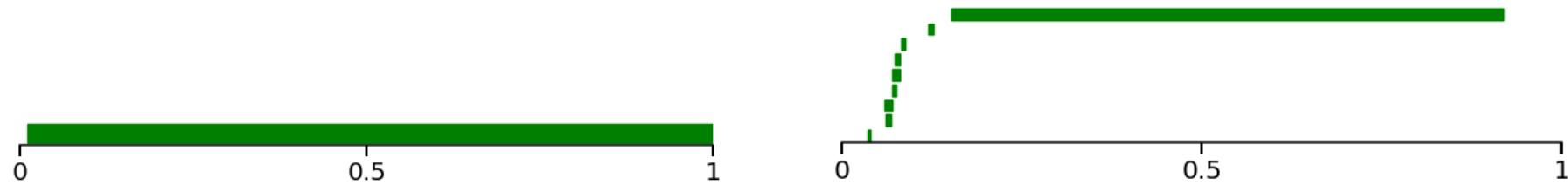
# Distance bottleneck

33/43 (1/10)

Soient deux codes-barres  $P$  et  $Q$ , c'est-à-dire des multi-ensembles d'intervalles  $\{(a_i, b_i), i \in \mathcal{I}\}$  de  $(\overline{\mathbb{R}^+})^2$  tels que  $a_i \leq b_i$  pour tout  $i \in \mathcal{I}$ .



Soient deux codes-barres  $P$  et  $Q$ , c'est-à-dire des multi-ensembles d'intervalles  $\{(a_i, b_i), i \in \mathcal{I}\}$  de  $(\overline{\mathbb{R}^+})^2$  tels que  $a_i \leq b_i$  pour tout  $i \in \mathcal{I}$ .

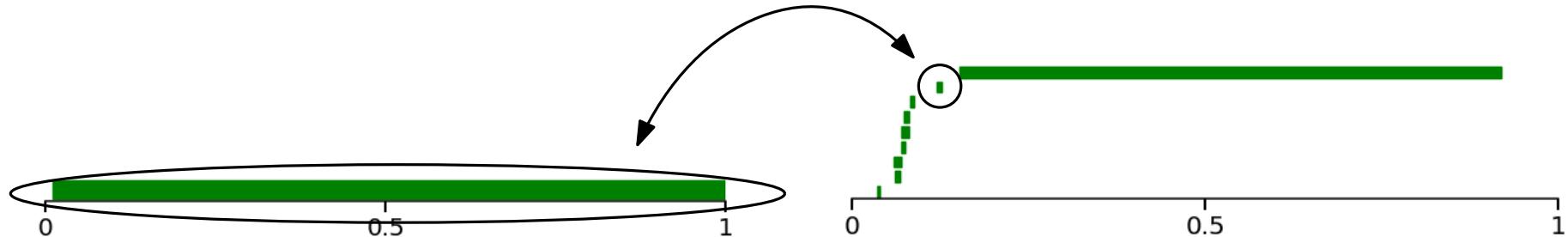


Un **appariement partiel** entre les codes-barres est un sous-ensemble  $M \subset P \times Q$  tel que

- pour tout  $p \in P$ , il existe au plus un  $q \in Q$  tel que  $(p, q) \in M$ ,
- pour tout  $q \in Q$ , il existe au plus un  $p \in P$  tel que  $(p, q) \in M$ .

Les points  $p \in P$  (resp.  $q \in Q$ ) tels qu'il existe  $q \in Q$  (resp.  $p \in P$ ) avec  $(p, q) \in M$  sont dit **appareillés** par  $M$ .

Soient deux codes-barres  $P$  et  $Q$ , c'est-à-dire des multi-ensembles d'intervalles  $\{(a_i, b_i), i \in \mathcal{I}\}$  de  $(\overline{\mathbb{R}^+})^2$  tels que  $a_i \leq b_i$  pour tout  $i \in \mathcal{I}$ .

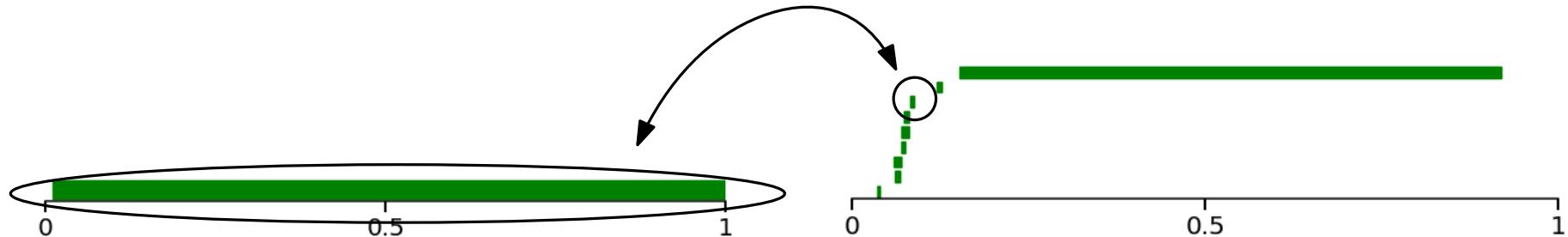


Un **appariement partiel** entre les codes-barres est un sous-ensemble  $M \subset P \times Q$  tel que

- pour tout  $p \in P$ , il existe au plus un  $q \in Q$  tel que  $(p, q) \in M$ ,
- pour tout  $q \in Q$ , il existe au plus un  $p \in P$  tel que  $(p, q) \in M$ .

Les points  $p \in P$  (resp.  $q \in Q$ ) tels qu'il existe  $q \in Q$  (resp.  $p \in P$ ) avec  $(p, q) \in M$  sont dit **appareillés** par  $M$ .

Soient deux codes-barres  $P$  et  $Q$ , c'est-à-dire des multi-ensembles d'intervalles  $\{(a_i, b_i), i \in \mathcal{I}\}$  de  $(\overline{\mathbb{R}^+})^2$  tels que  $a_i \leq b_i$  pour tout  $i \in \mathcal{I}$ .

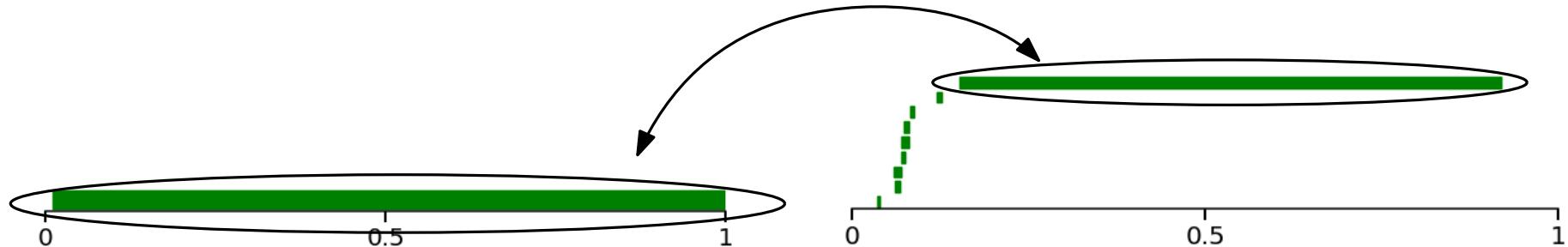


Un **appariement partiel** entre les codes-barres est un sous-ensemble  $M \subset P \times Q$  tel que

- pour tout  $p \in P$ , il existe au plus un  $q \in Q$  tel que  $(p, q) \in M$ ,
- pour tout  $q \in Q$ , il existe au plus un  $p \in P$  tel que  $(p, q) \in M$ .

Les points  $p \in P$  (resp.  $q \in Q$ ) tels qu'il existe  $q \in Q$  (resp.  $p \in P$ ) avec  $(p, q) \in M$  sont dit **appareillés** par  $M$ .

Soient deux codes-barres  $P$  et  $Q$ , c'est-à-dire des multi-ensembles d'intervalles  $\{(a_i, b_i), i \in \mathcal{I}\}$  de  $(\overline{\mathbb{R}^+})^2$  tels que  $a_i \leq b_i$  pour tout  $i \in \mathcal{I}$ .

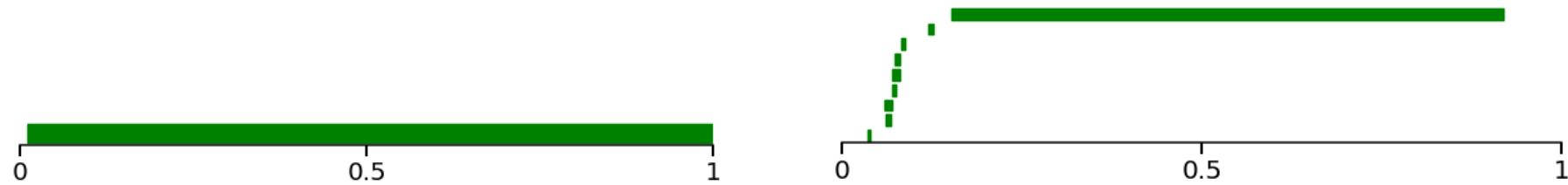


Un **appariement partiel** entre les codes-barres est un sous-ensemble  $M \subset P \times Q$  tel que

- pour tout  $p \in P$ , il existe au plus un  $q \in Q$  tel que  $(p, q) \in M$ ,
- pour tout  $q \in Q$ , il existe au plus un  $p \in P$  tel que  $(p, q) \in M$ .

Les points  $p \in P$  (resp.  $q \in Q$ ) tels qu'il existe  $q \in Q$  (resp.  $p \in P$ ) avec  $(p, q) \in M$  sont dit **appareillés** par  $M$ .

Soient deux codes-barres  $P$  et  $Q$ , c'est-à-dire des multi-ensembles d'intervalles  $\{(a_i, b_i), i \in \mathcal{I}\}$  de  $(\overline{\mathbb{R}^+})^2$  tels que  $a_i \leq b_i$  pour tout  $i \in \mathcal{I}$ .



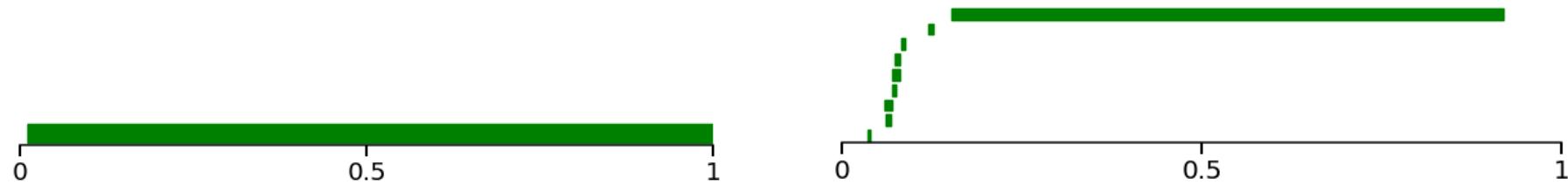
Un **appariement partiel** entre les codes-barres est un sous-ensemble  $M \subset P \times Q$  tel que

- pour tout  $p \in P$ , il existe au plus un  $q \in Q$  tel que  $(p, q) \in M$ ,
- pour tout  $q \in Q$ , il existe au plus un  $p \in P$  tel que  $(p, q) \in M$ .

Les points  $p \in P$  (resp.  $q \in Q$ ) tels qu'il existe  $q \in Q$  (resp.  $p \in P$ ) avec  $(p, q) \in M$  sont dit **appareillés** par  $M$ .

Si un point  $p \in P$  (resp.  $q \in Q$ ) n'est pas appareillé par  $M$ , on considère qu'il est appareillé avec le singleton  $\bar{p} = [\frac{p_1+p_2}{2}, \frac{p_1+p_2}{2}]$  (resp.  $\bar{q} = [\frac{q_1+q_2}{2}, \frac{q_1+q_2}{2}]$ ).

Soient deux codes-barres  $P$  et  $Q$ , c'est-à-dire des multi-ensembles d'intervalles  $\{(a_i, b_i), i \in \mathcal{I}\}$  de  $(\overline{\mathbb{R}^+})^2$  tels que  $a_i \leq b_i$  pour tout  $i \in \mathcal{I}$ .



Un **appariement partiel** entre les codes-barres est un sous-ensemble  $M \subset P \times Q$  tel que

- pour tout  $p \in P$ , il existe au plus un  $q \in Q$  tel que  $(p, q) \in M$ ,
- pour tout  $q \in Q$ , il existe au plus un  $p \in P$  tel que  $(p, q) \in M$ .

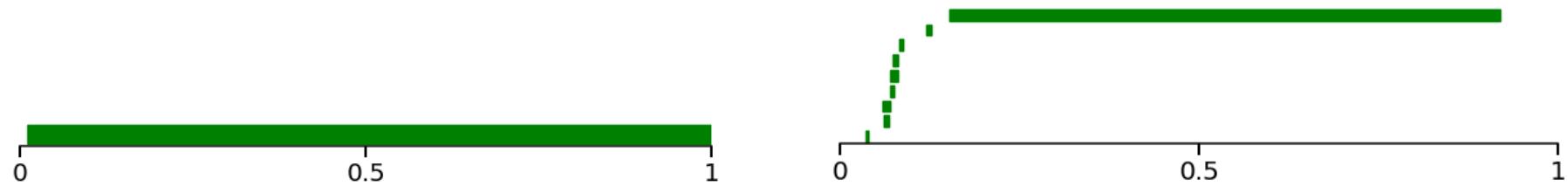
Les points  $p \in P$  (resp.  $q \in Q$ ) tels qu'il existe  $q \in Q$  (resp.  $p \in P$ ) avec  $(p, q) \in M$  sont dit **appareillés** par  $M$ .

Si un point  $p \in P$  (resp.  $q \in Q$ ) n'est pas appareillé par  $M$ , on considère qu'il est appareillé avec le singleton  $\bar{p} = [\frac{p_1+p_2}{2}, \frac{p_1+p_2}{2}]$  (resp.  $\bar{q} = [\frac{q_1+q_2}{2}, \frac{q_1+q_2}{2}]$ ).

Le **coût** d'un paire appareillée  $(p, q)$  (resp.  $(p, \bar{p})$ , resp.  $(\bar{q}, q)$ ) est la norme sup  $\|p - q\|_\infty = \sup\{|p_1 - q_1|, |p_2 - q_2|\}$  (resp.  $\|p - \bar{p}\|_\infty$ , resp.  $\|\bar{q} - q\|_\infty$ ).

Le **coût** de l'appariement partiel  $M$ , noté  $\text{cost}(M)$ , est le supremum de ses coûts.

Soient deux codes-barres  $P$  et  $Q$ , c'est-à-dire des multi-ensembles d'intervalles  $\{(a_i, b_i), i \in \mathcal{I}\}$  de  $(\overline{\mathbb{R}^+})^2$  tels que  $a_i \leq b_i$  pour tout  $i \in \mathcal{I}$ .



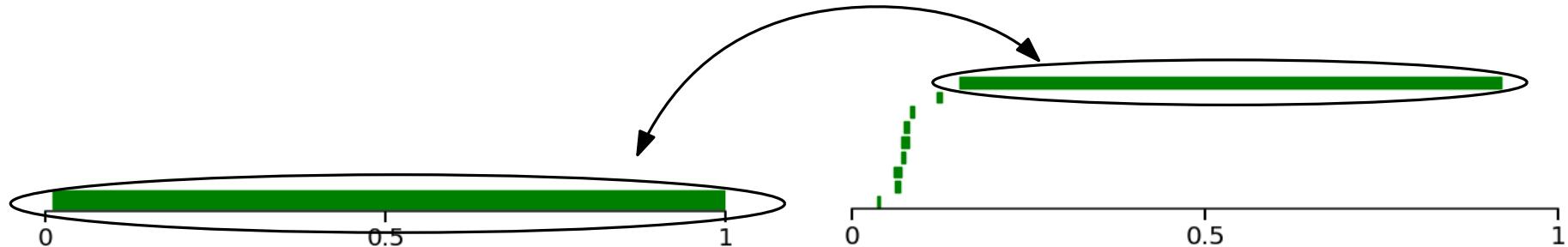
Définition : La **distance bottleneck** entre  $P$  et  $Q$  est définie comme le plus petit coût des appariements partiels :

$$d_b(P, Q) = \inf\{\text{cost}(M), M \text{ est un appariement partiel entre } P \text{ et } Q\}.$$

# Distance bottleneck

33/43 (9/10)

Soient deux codes-barres  $P$  et  $Q$ , c'est-à-dire des multi-ensembles d'intervalles  $\{(a_i, b_i), i \in \mathcal{I}\}$  de  $(\overline{\mathbb{R}^+})^2$  tels que  $a_i \leq b_i$  pour tout  $i \in \mathcal{I}$ .



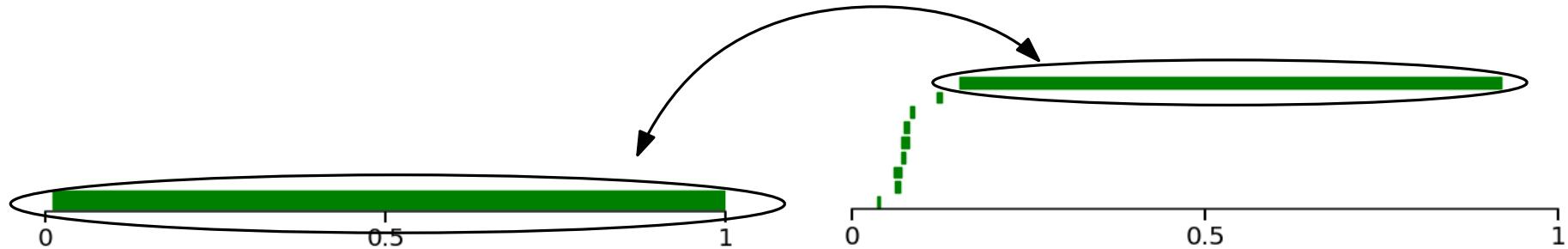
Définition : La **distance bottleneck** entre  $P$  et  $Q$  est définie comme le plus petit coût des appariements partiels :

$$d_b(P, Q) = \inf\{\text{cost}(M), M \text{ est un appariement partiel entre } P \text{ et } Q\}.$$

# Distance bottleneck

33/43 (10/10)

Soient deux codes-barres  $P$  et  $Q$ , c'est-à-dire des multi-ensembles d'intervalles  $\{(a_i, b_i), i \in \mathcal{I}\}$  de  $(\overline{\mathbb{R}^+})^2$  tels que  $a_i \leq b_i$  pour tout  $i \in \mathcal{I}$ .



Définition : La **distance bottleneck** entre  $P$  et  $Q$  est définie comme le plus petit coût des appariements partiels :

$$d_b(P, Q) = \inf\{\text{cost}(M), M \text{ est un appariement partiel entre } P \text{ et } Q\}.$$

Si  $\mathbb{U}$  and  $\mathbb{V}$  sont deux modules de persistance, on définit leur **distance bottleneck** comme

$$d_b(\mathbb{U}, \mathbb{V}) = d_b(\text{Barcode}(\mathbb{U}), \text{Barcode}(\mathbb{V})).$$

# Théorème de stabilité

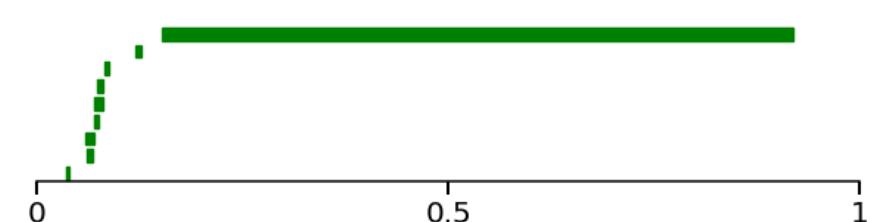
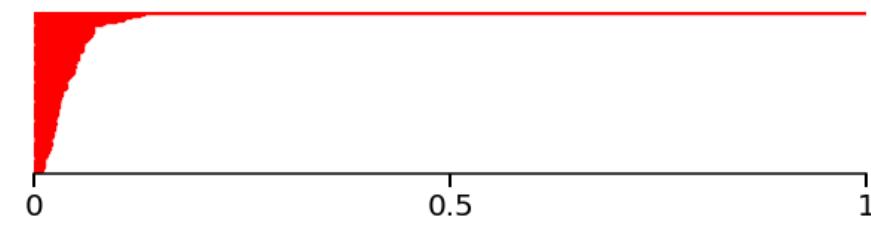
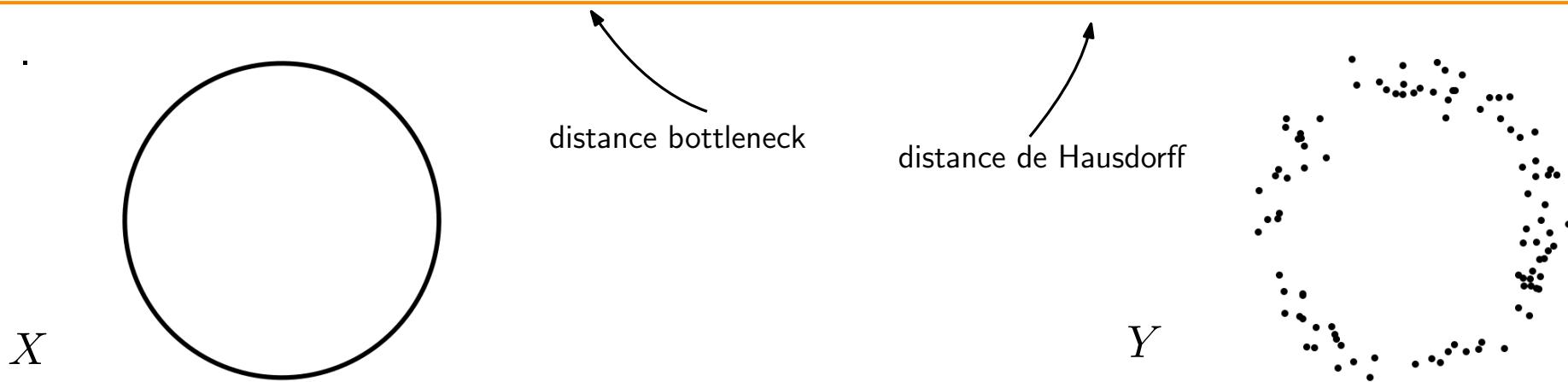
34/43 (1/2)

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

Soit  $X$  et  $Y$  deux sous-ensembles de  $\mathbb{R}^n$ .

Considérons leurs filtrations de Čech, et leurs modules de persistance de  $i^{\text{ème}}$  homologie,  $\mathbb{U}$  et  $\mathbb{V}$ . Alors

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



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

Soit  $X$  et  $Y$  deux sous-ensembles de  $\mathbb{R}^n$ .

Considérons leurs filtrations de Čech, et leurs modules de persistance de  $i^{\text{ème}}$  homologie,  $\mathbb{U}$  et  $\mathbb{V}$ . Alors

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



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

On a

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

où  $d_i$  est la **distance d'entrelacement**.

# I - Homologie simpliciale

- 1 - Groupes d'homologie
- 2 - Fonctorialité

# II - Inférence topologique

- 1 - Estimation du paramètre
- 2 - Nerfs

# III - Homologie persistante

- 1 - Modules de persistance
- 2 - Décomposition
- 3 - Stabilité

# IV - Applications

# Inférence topologique I

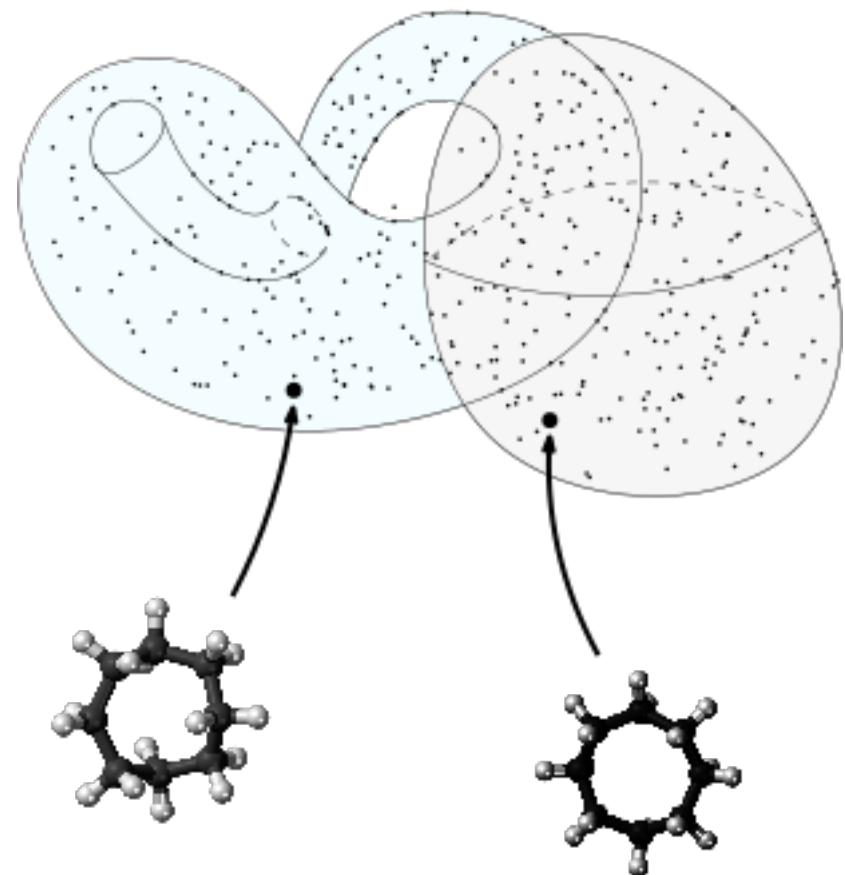
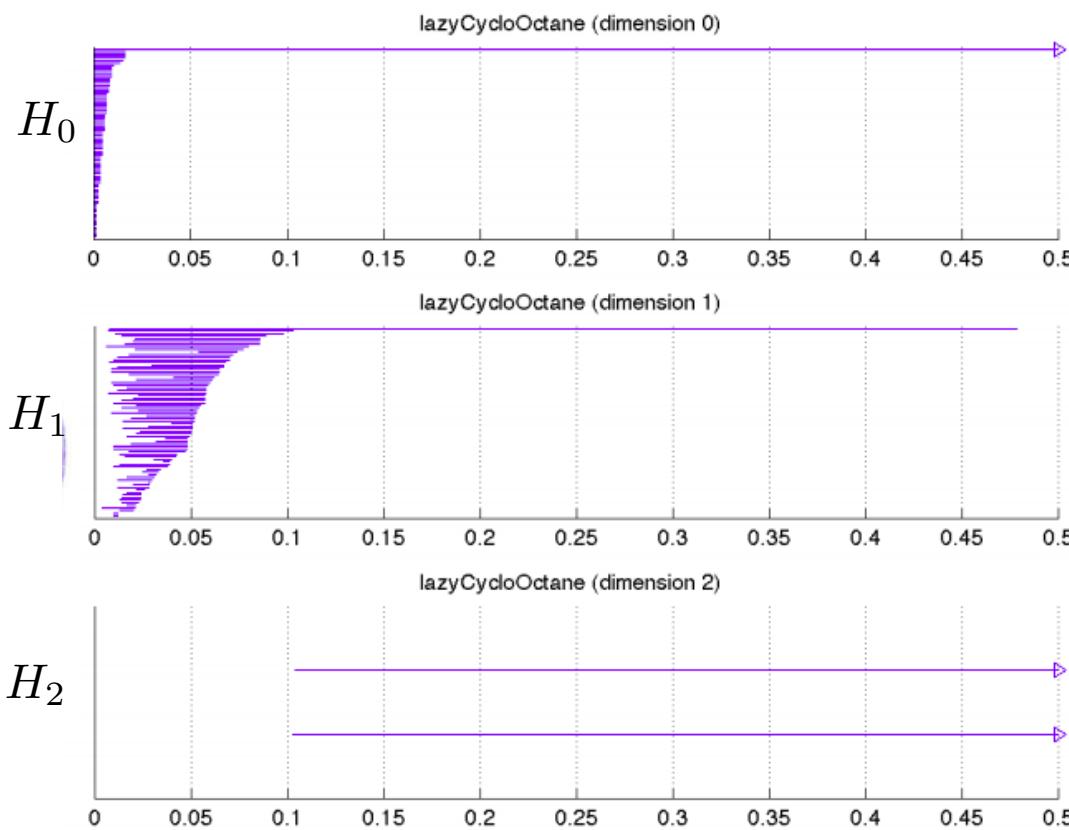
36/43 (1/2)

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](https://www.researchgate.net/publication/44697030_Topology_of_Cyclooctane_Energy_Landscape)

La molécule de cyclo-octane  $C_8H_{16}$  contient 24 atomes. En générant beaucoup de molécules, on obtient un nuage de points dans  $\mathbb{R}^{72}$  ( $3 \times 24 = 72$ ).

On calcule les codes-barres :



# Inférence topologique I

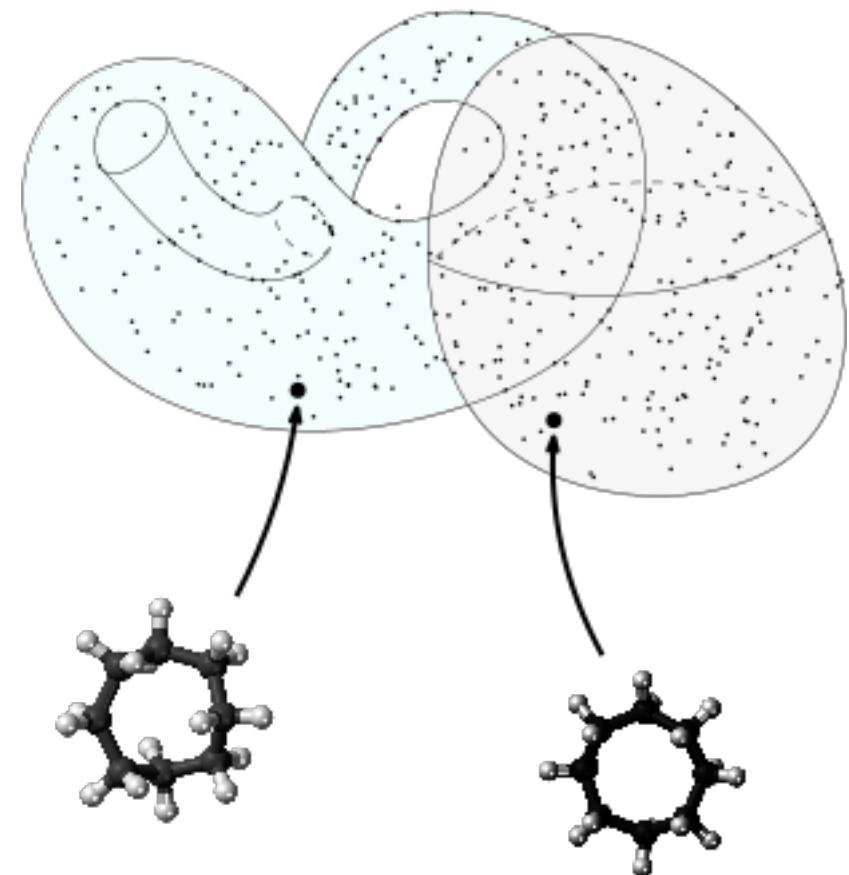
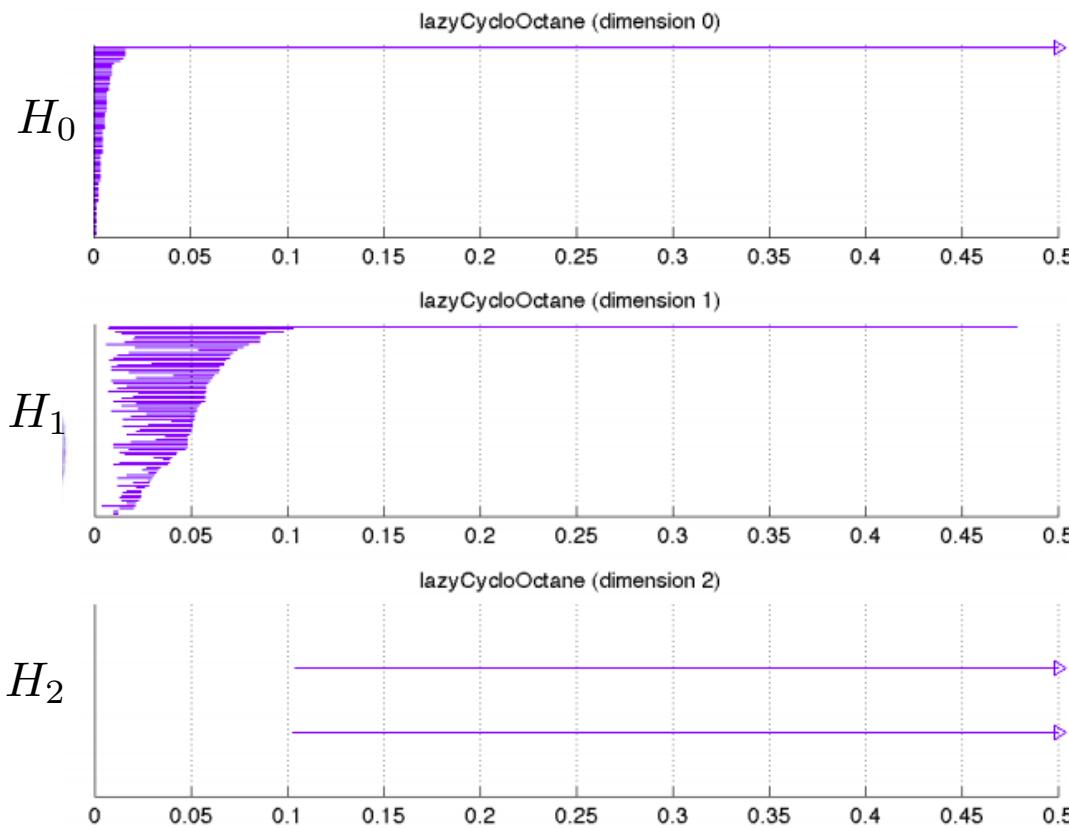
36/43 (2/2)

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](https://www.researchgate.net/publication/44697030_Topology_of_Cyclooctane_Energy_Landscape)

La molécule de cyclo-octane  $C_8H_{16}$  contient 24 atomes. En générant beaucoup de molécules, on obtient un nuage de points dans  $\mathbb{R}^{72}$  ( $3 \times 24 = 72$ ).

On calcule les codes-barres :

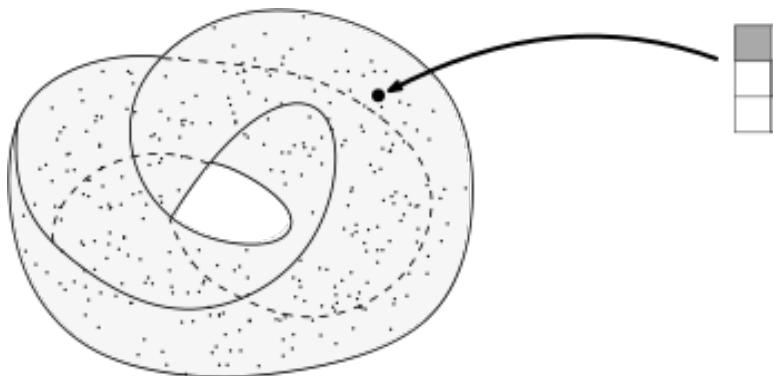


On déduit :  $H_0 = \mathbb{Z}/2\mathbb{Z}$ ,  $H_1 = \mathbb{Z}/2\mathbb{Z}$ ,  $H_2 = (\mathbb{Z}/2\mathbb{Z})^2$

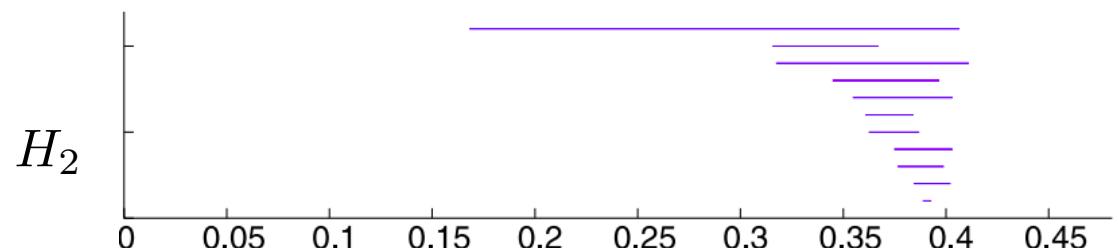
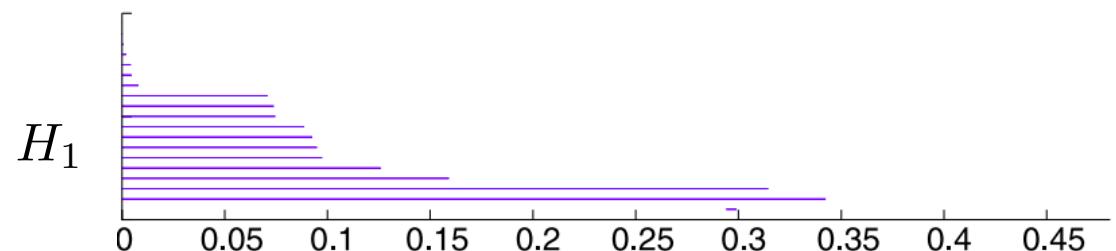
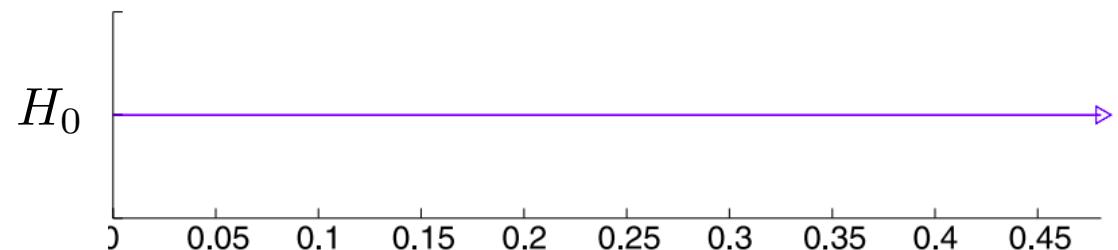
G. Carlsson, T. Ishkhanov, V. de Silva, and A. Zomorodian, [On the Local Behavior of Spaces of Natural Images](#), 2008

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

À partir d'un grande collection d'images, les auteurs extraient des imagettes de taille  $3 \times 3$ . Puisqu'elles consistent en 9 pixels, chacune de ces imagettes peut être vue comme un vecteur en dimension 9, et l'ensemble comme un nuage de points dans  $\mathbb{R}^9$ .



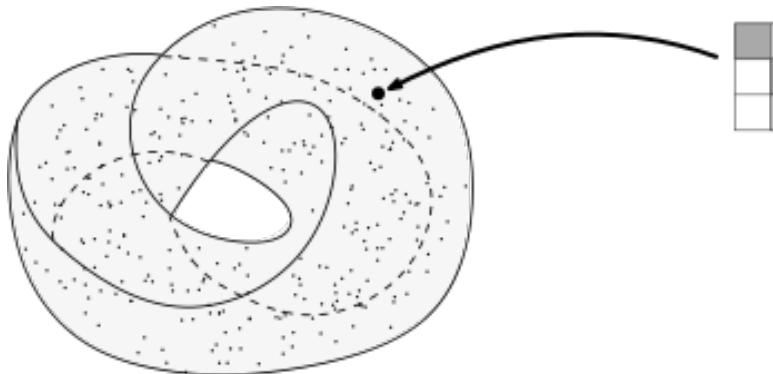
On obtient les codes-barres :



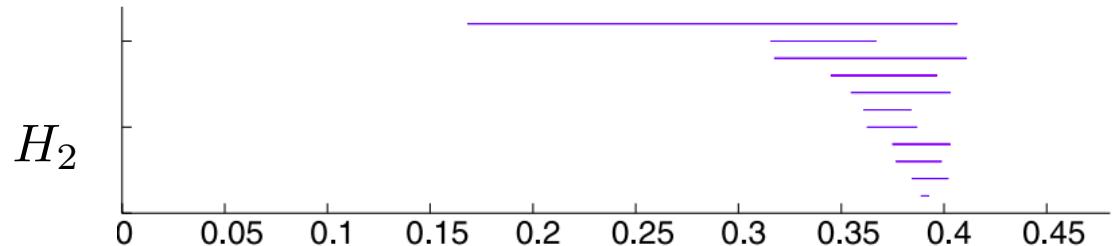
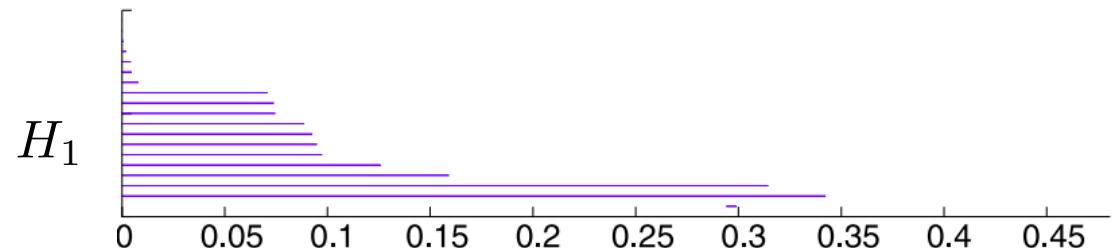
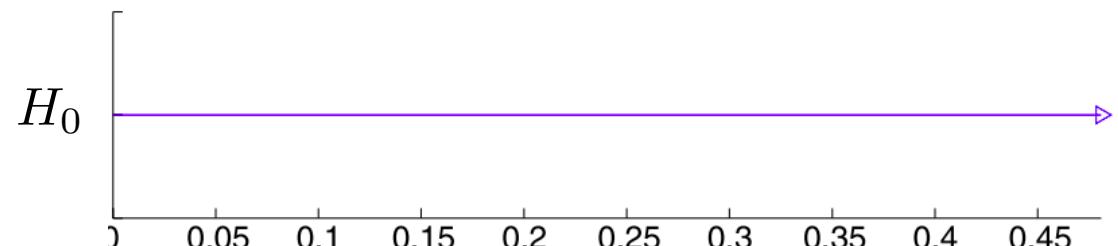
G. Carlsson, T. Ishkhanov, V. de Silva, and A. Zomorodian, [On the Local Behavior of Spaces of Natural Images](#), 2008

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

À partir d'un grande collection d'images, les auteurs extraient des imagettes de taille  $3 \times 3$ . Puisqu'elles consistent en 9 pixels, chacune de ces imagettes peut être vue comme un vecteur en dimension 9, et l'ensemble comme un nuage de points dans  $\mathbb{R}^9$ .



On obtient les codes-barres :



On déduit :

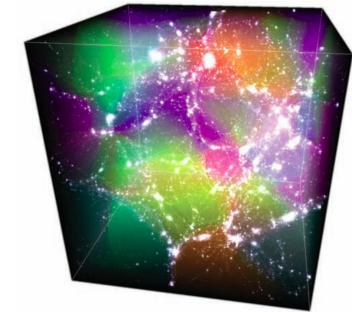
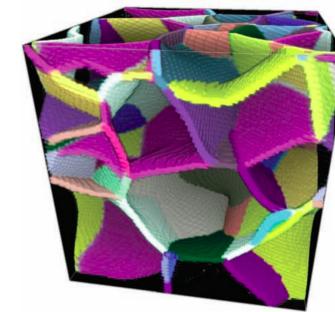
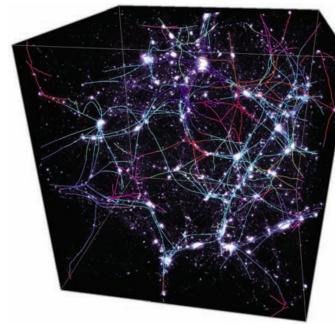
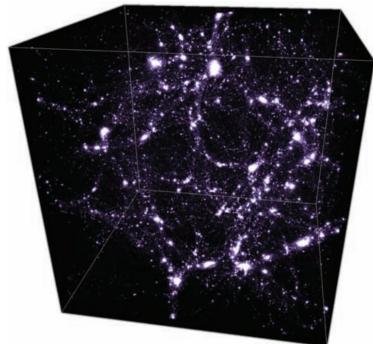
$$\begin{aligned} H_0 &= \mathbb{Z}/2\mathbb{Z}, \\ H_1 &= (\mathbb{Z}/2\mathbb{Z})^2, \\ H_2 &= \mathbb{Z}/2\mathbb{Z}. \end{aligned}$$

# Analyse multi-échelle I

38/43

T. Sousbie, The persistent cosmic web and its filamentary structure, 2011

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



vu comme un objet  
de dimension 1

de dimension 2

de dimension 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>

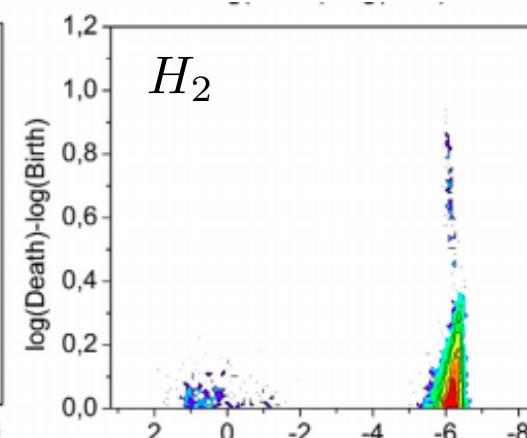
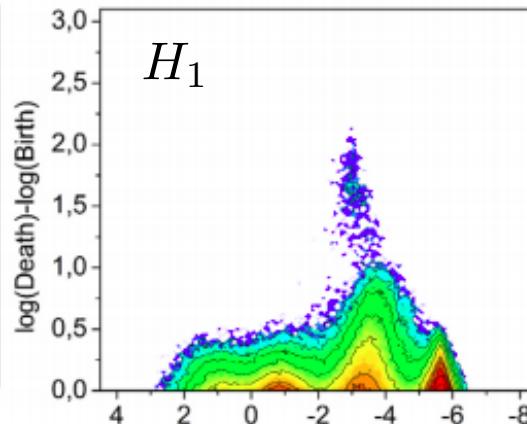
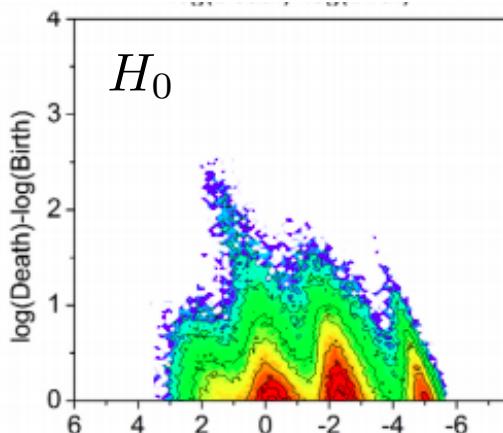


Diagramme de  
persistance moyen  
pour un modèle de  
Voronoi  
(échelle log)

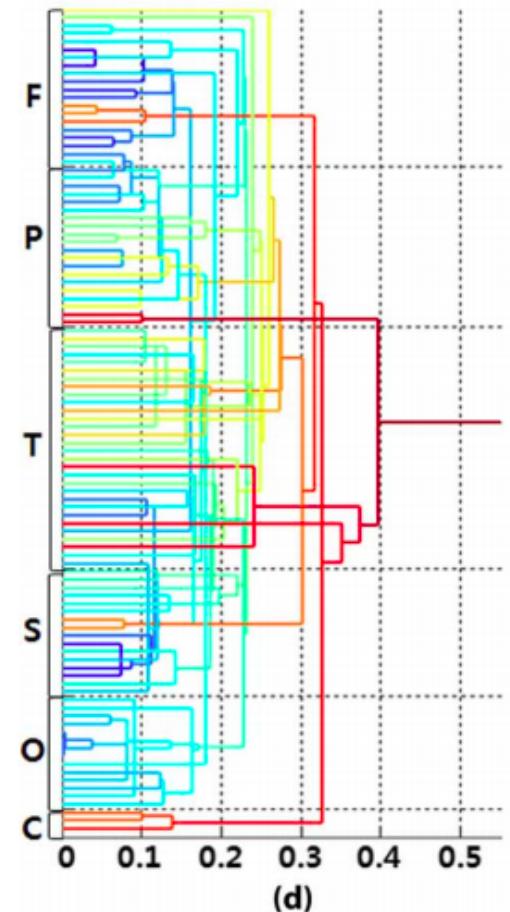
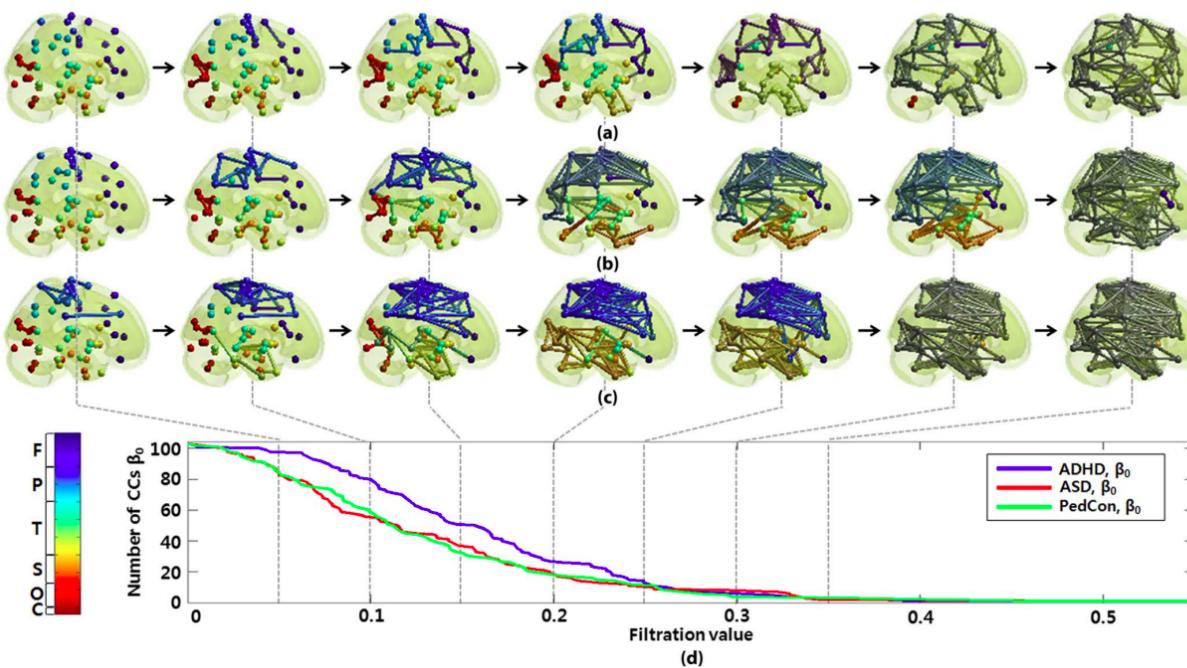
# Analyse multi-échelle II

39/43

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>

→ La  $H_0$ -homologie persistante induit une classification hiérarchique



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

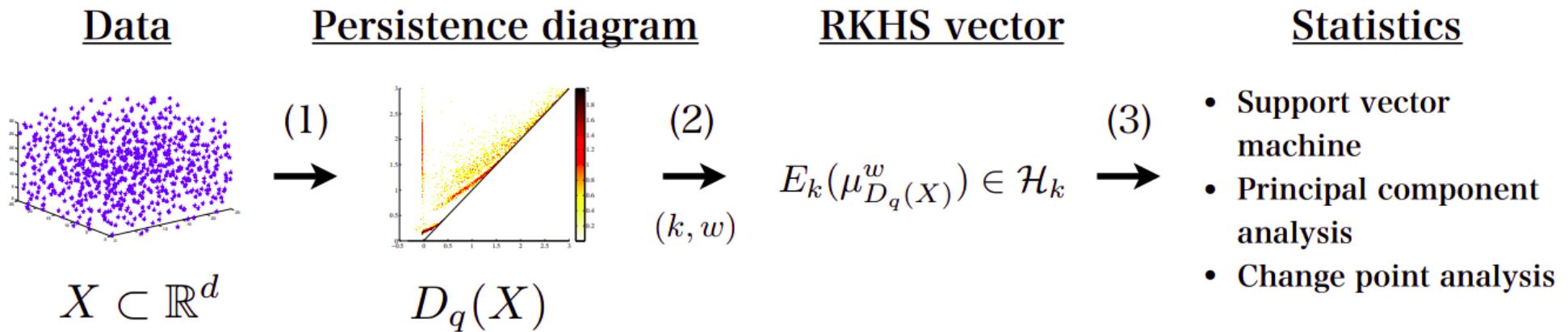
<https://arxiv.org/abs/1706.03358>

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>

Les codes-barres ne sont pas des éléments d'un espace euclidien, et ne peuvent donc pas être utilisés directement

→ on peut utiliser le **kernel trick**



# En machine learning II

41/43

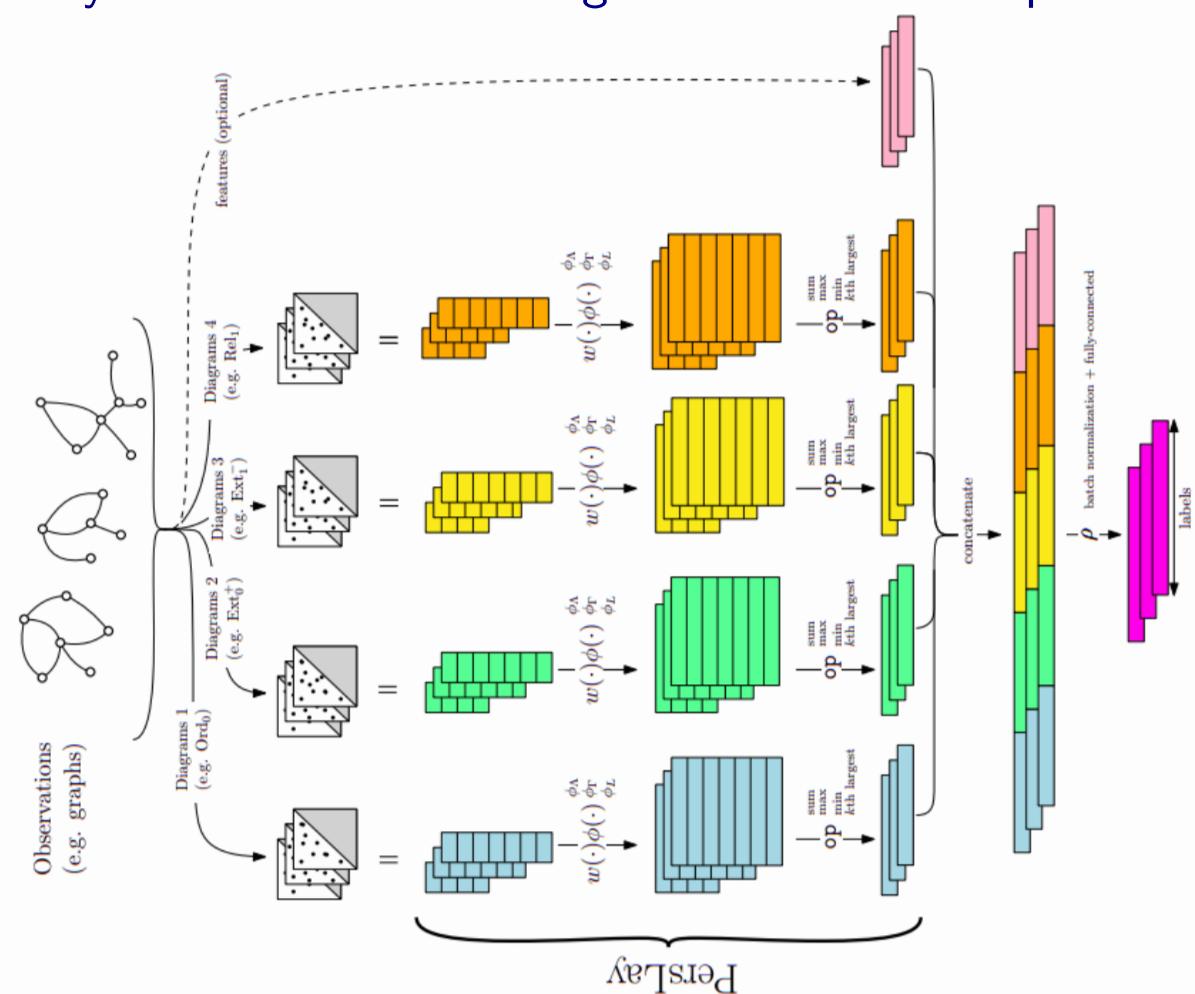
Rickard Brüel-Gabrielsson, Bradley J. Nelson, Anjan Dwaraknath, Primoz Skraba, Leonidas J. Guibas, Gunnar Carlsson, [A Topology Layer for Machine Learning](#), 2019

<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>

On peut intégrer les diagrammes de persistance à un réseau de neurones



# Classification

42/43

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>

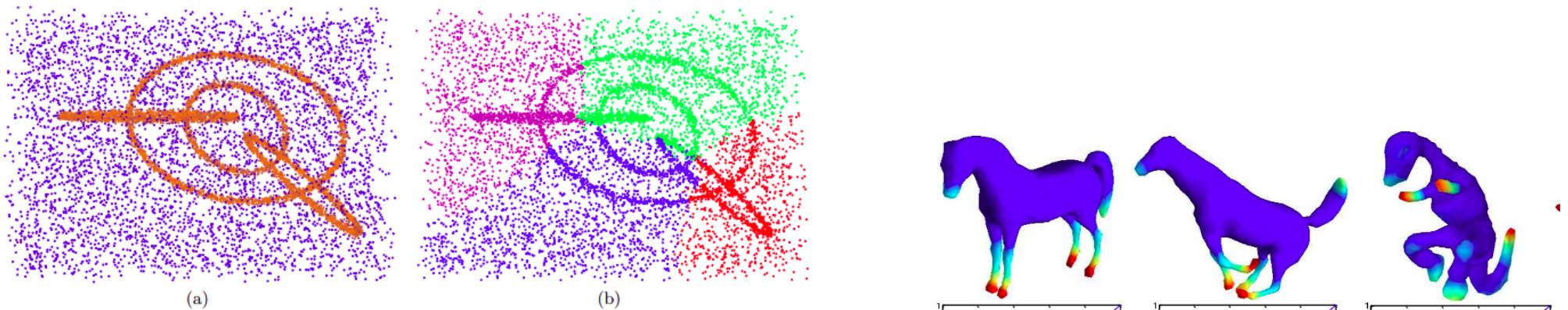
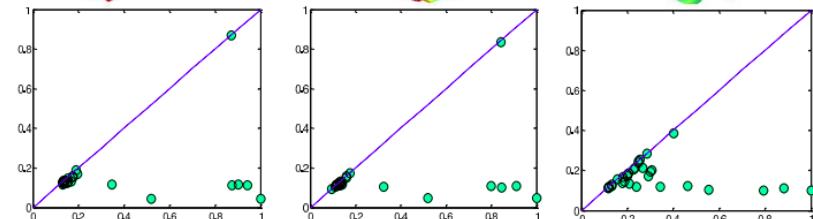
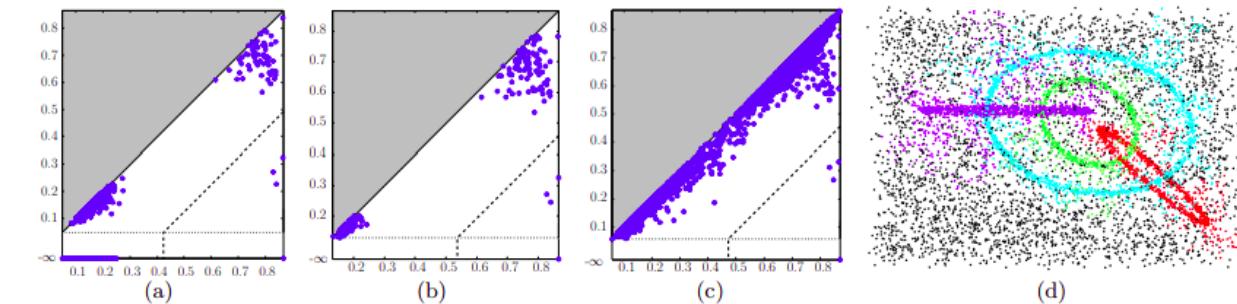
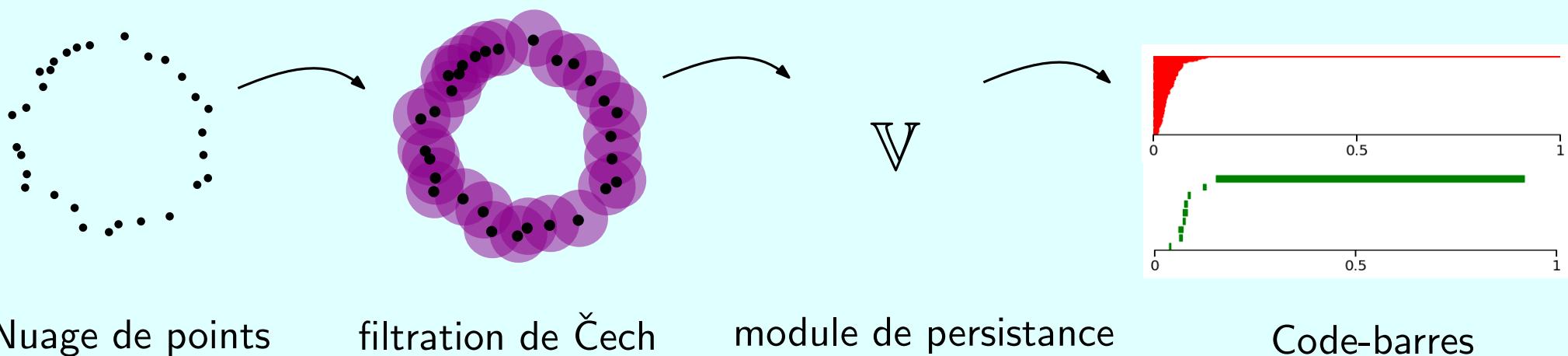


Figure 7: (a) The rings data set with the estimated density function. (b) The result obtained using spectral clustering.



# Conclusion

L'homologie persistante permet une estimation **multi-échelle** et **stable** de l'homologie des jeux de données.

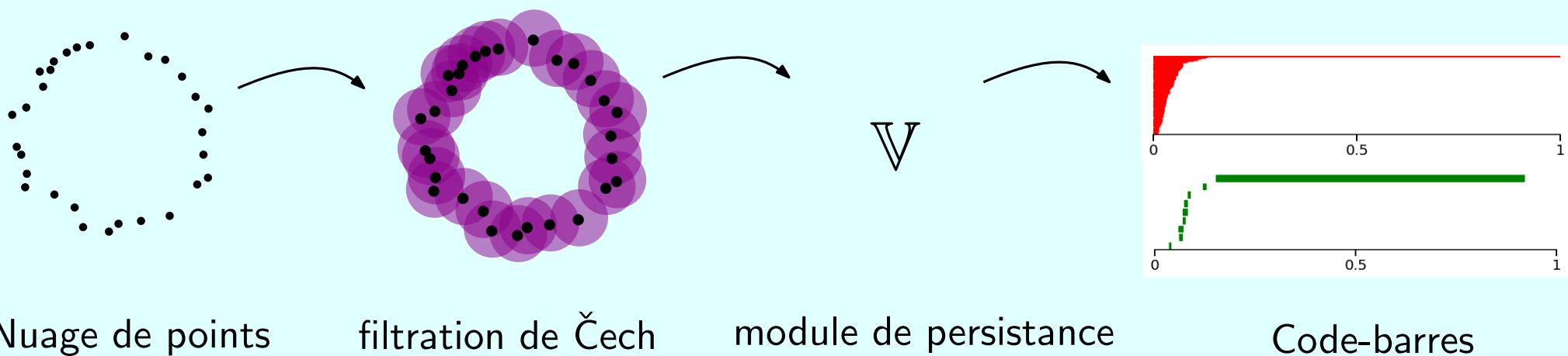


Elle permet d'analyser les données depuis un nouveau point de vue.

Un cours à propos de la TDA : <https://raphaeltinarrage.github.io/EMAp.html>

# Conclusion

L'homologie persistante permet une estimation **multi-échelle** et **stable** de l'homologie des jeux de données.



Elle permet d'analyser les données depuis un nouveau point de vue.

Un cours à propos de la TDA : <https://raphaeltinarrage.github.io/EMAp.html>

Merci !