

EMINES — 26/05/21

# Analyse Topologique des Données

# التحليل الطوبولوجي للبيانات

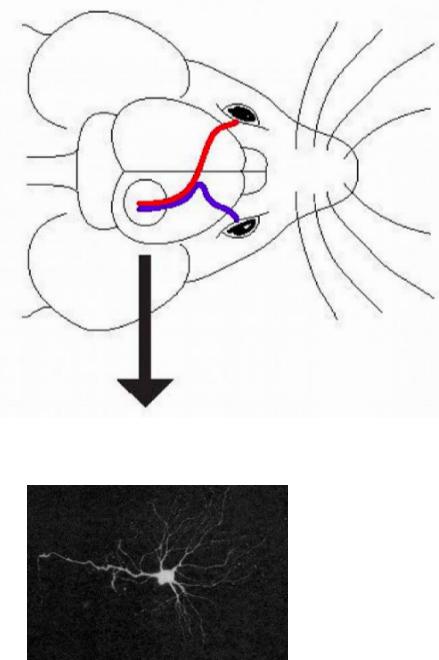
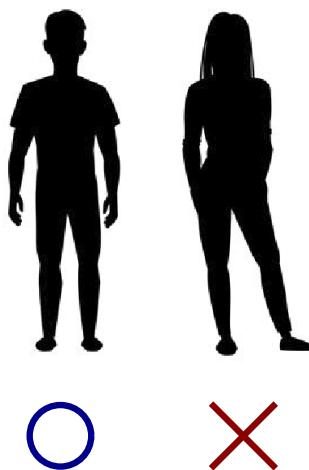
Séance I/II : Invariants topologiques

<https://raphaeltinarrage.github.io>

# Intro I - Statistiques inférentielles

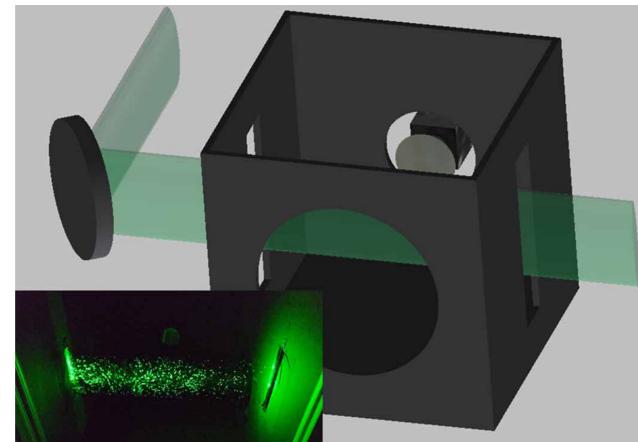
2/30 (1/10)

On va commencer par observer trois jeux de données.



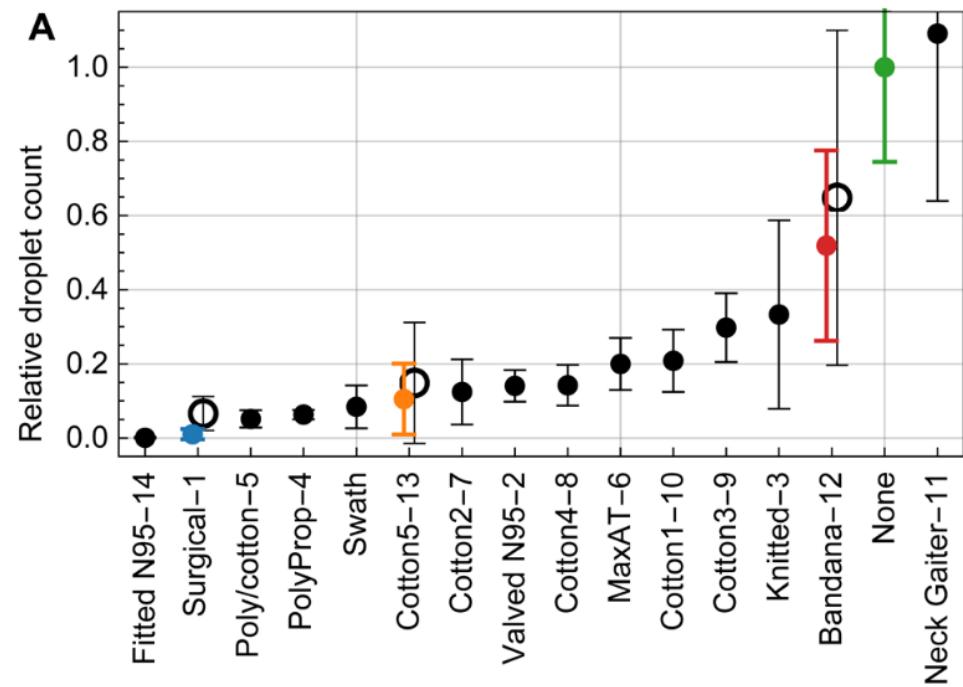
[Low-cost measurement of face mask efficacy for filtering expelled droplets during speech, Emma P. Fischer, Martin C. Fischer, David Grass, Isaac Henrion, Warren S. Warren, Eric Westman, *Science Advances*, Sep 2020]

Diverses transmission du covid : goutelettes et aérosols.



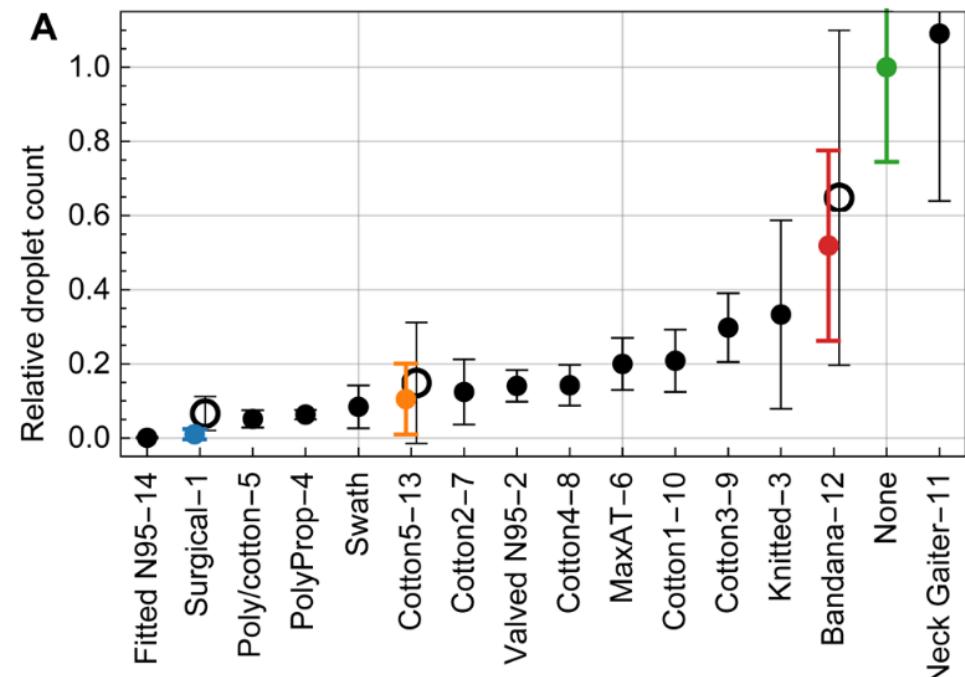
[Low-cost measurement of face mask efficacy for filtering expelled droplets during speech, Emma P. Fischer, Martin C. Fischer, David Grass, Isaac Henrion, Warren S. Warren, Eric Westman, *Science Advances*, Sep 2020]

Diverses transmission du covid : goutelettes et aérosols.



[Low-cost measurement of face mask efficacy for filtering expelled droplets during speech, Emma P. Fischer, Martin C. Fischer, David Grass, Isaac Henrion, Warren S. Warren, Eric Westman, *Science Advances*, Sep 2020]

Diverses transmission du covid : goutelettes et aérosols.

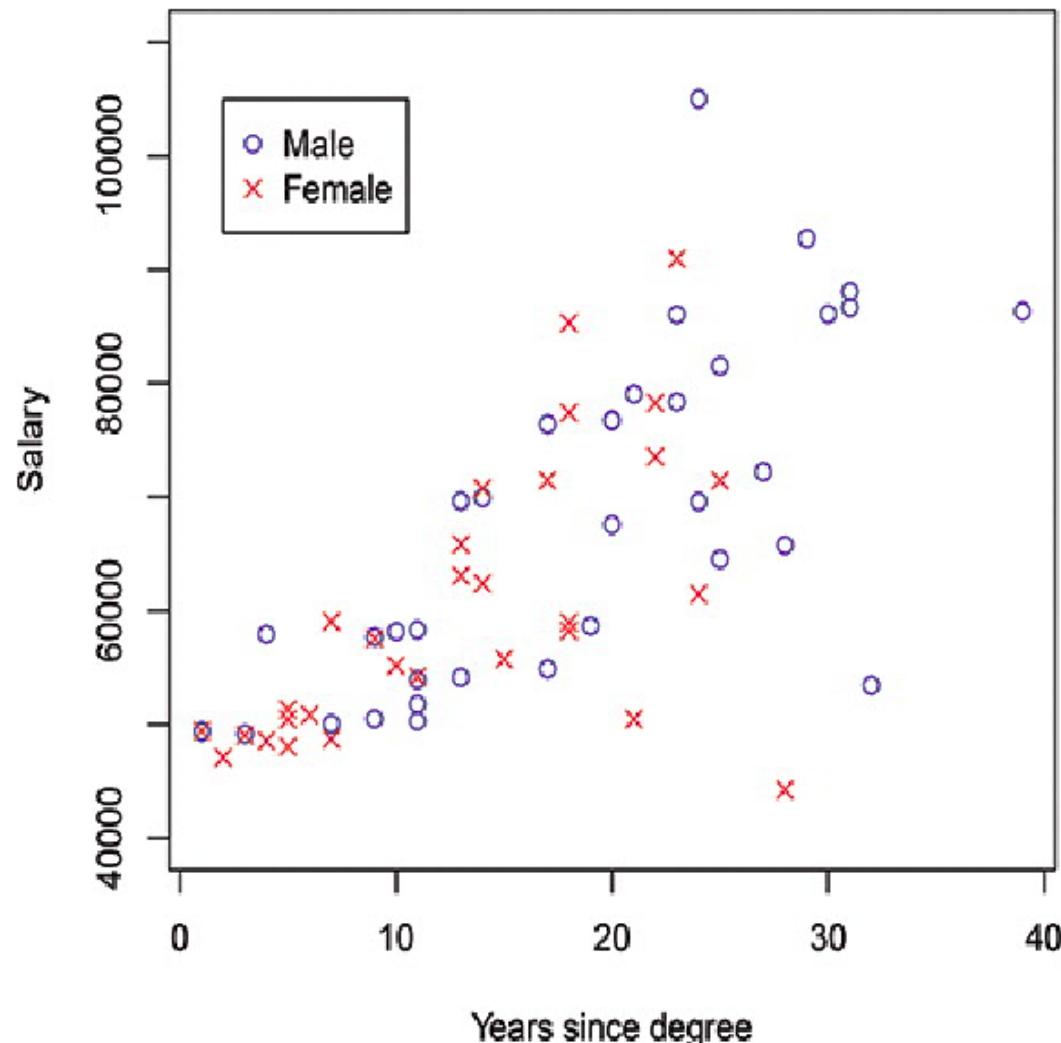
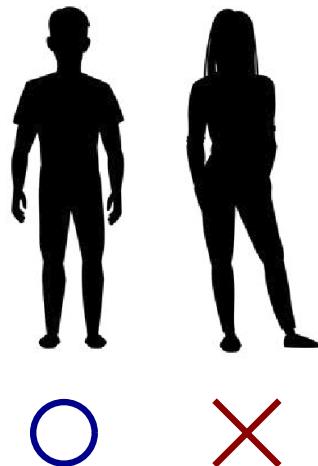


Comparaison de données de dimension 1

# Intro I - Statistiques inférentielles

2/30 (5/10)

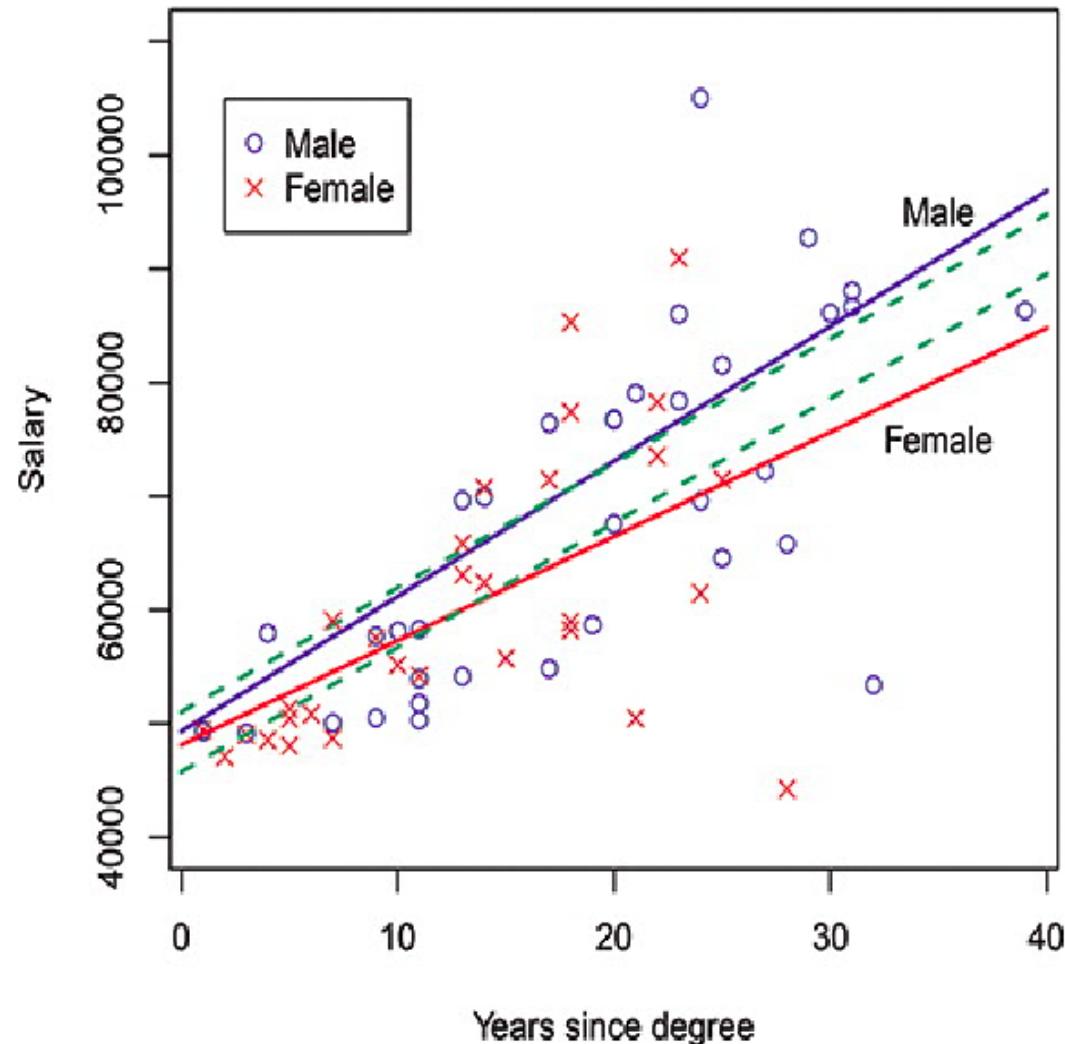
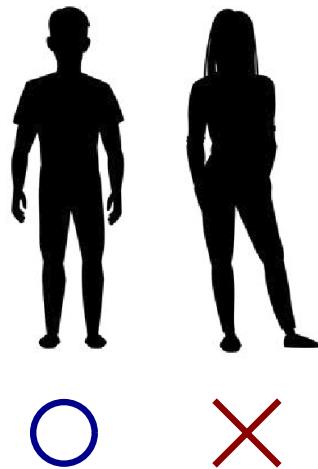
[Study of Salary Differentials by Gender and Discipline, L. Billard, *Statistics and Public Policy*, June 2017]



# Intro I - Statistiques inférentielles

2/30 (6/10)

[Study of Salary Differentials by Gender and Discipline, L. Billard, *Statistics and Public Policy*, June 2017]

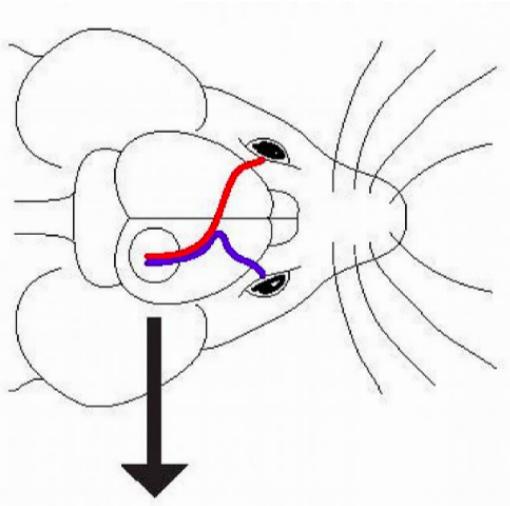


Comparaison de données de dimension 2 structurées temporellement

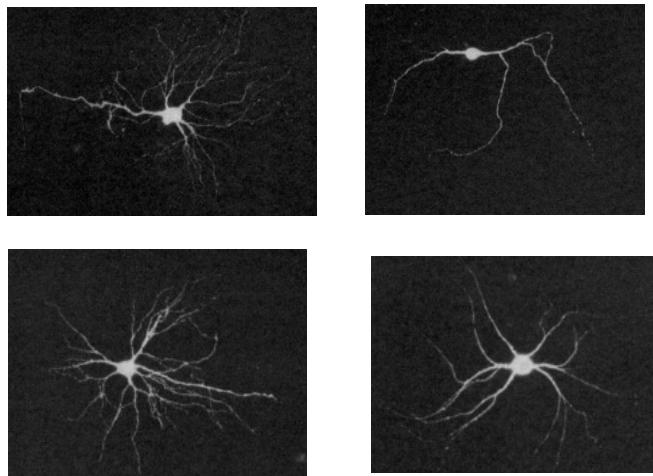
# Intro I - Statistiques inférentielles

2/30 (7/10)

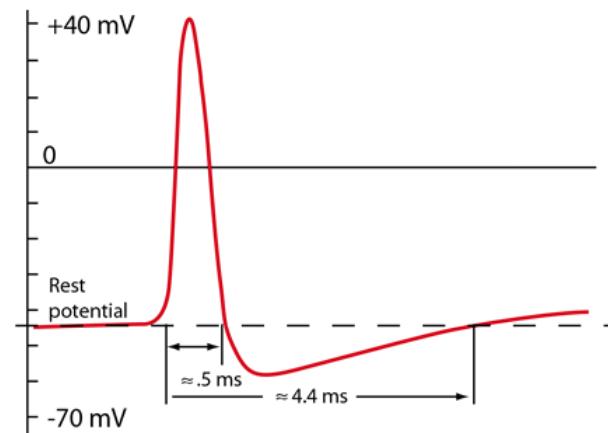
[Classification of electrophysiological and morphological neuron types in the mouse visual cortex, Nathan W. Gouwens et al, *Nature neuroscience*, Nov 2019]



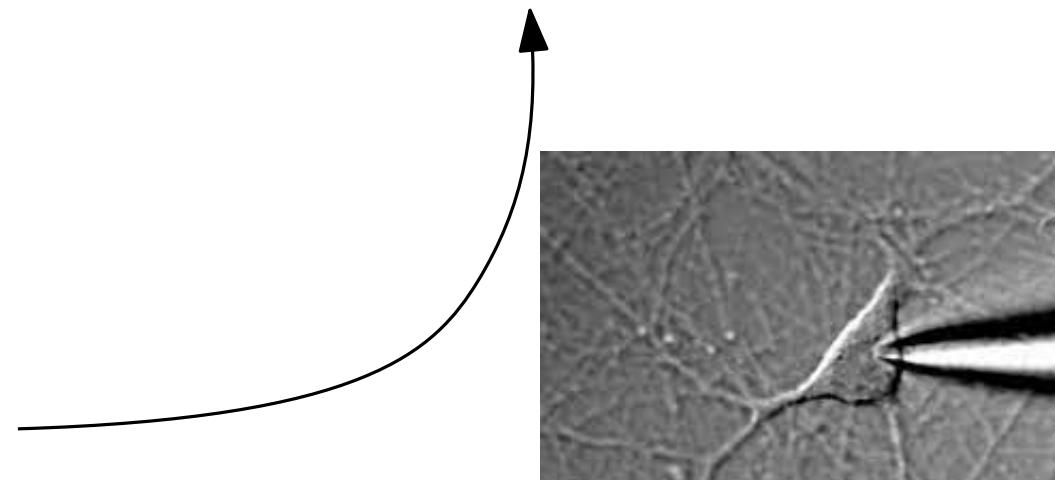
une souris



des neurones



potentiels d'action émis par les neurones

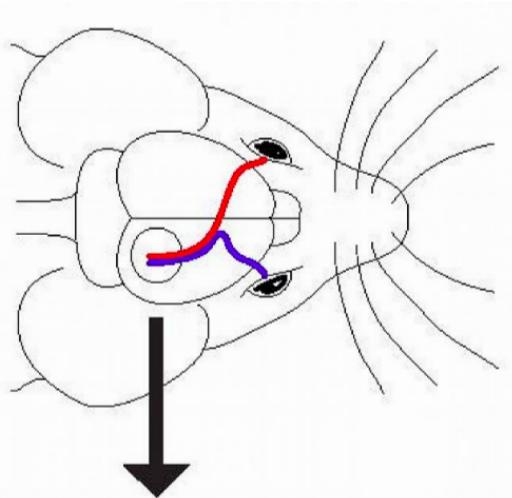


enregistrement du signal par patch-clamp

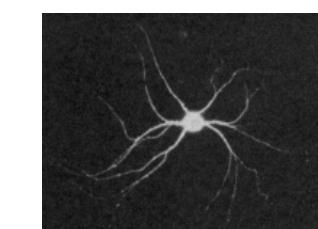
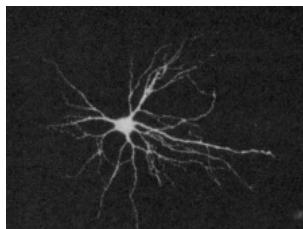
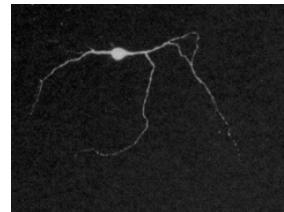
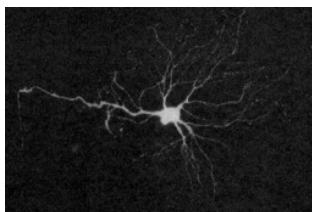
# Intro I - Statistiques inférentielles

2/30 (8/10)

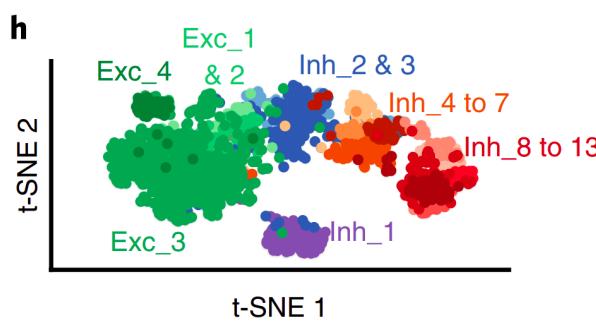
[Classification of electrophysiological and morphological neuron types in the mouse visual cortex, Nathan W. Gouwens et al, *Nature neuroscience*, Nov 2019]



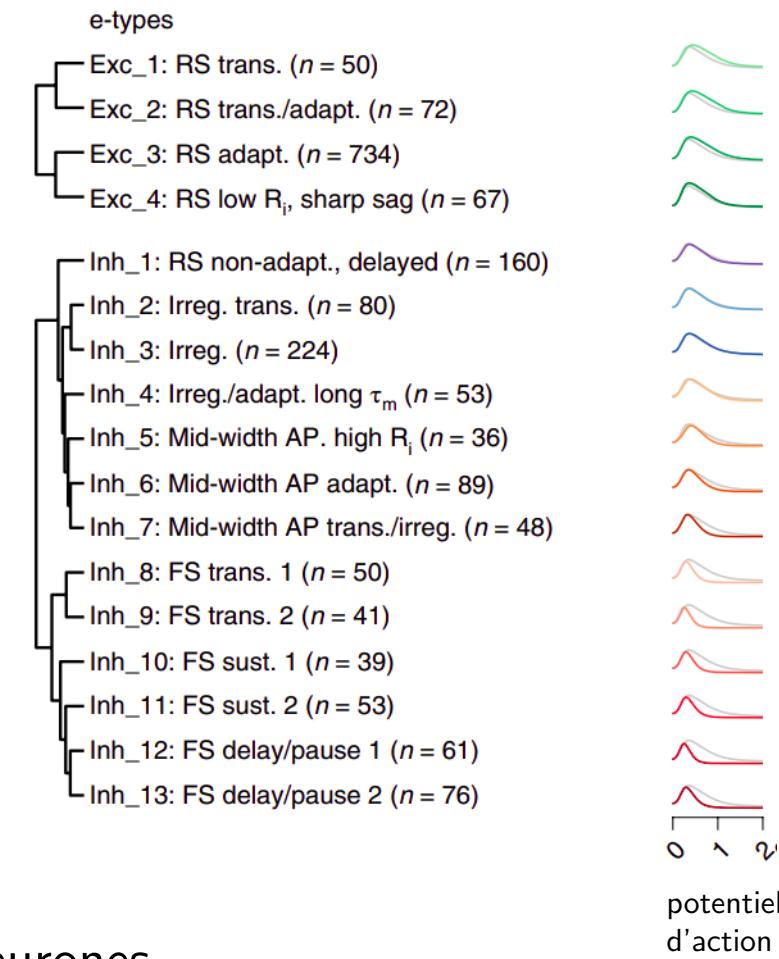
une souris



des neurones



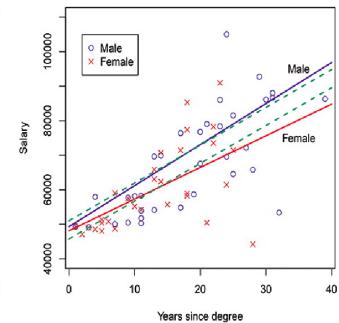
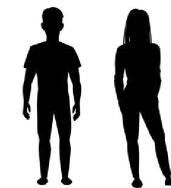
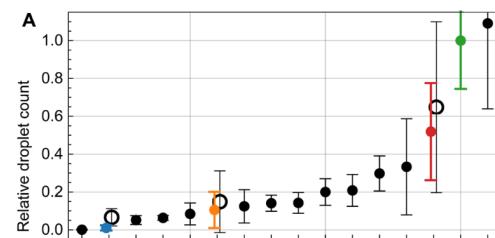
classification des neurones



Classification de données de grande dimension

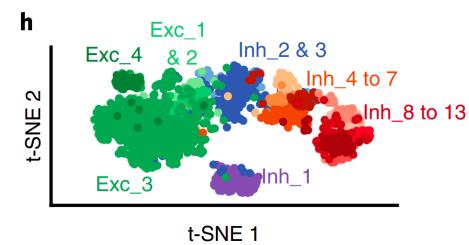
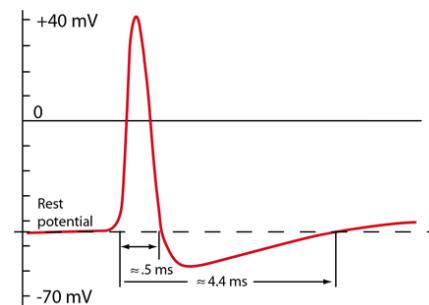
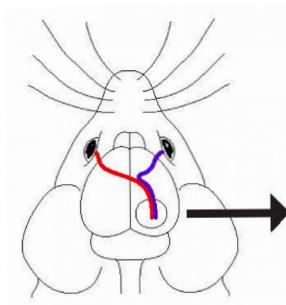
# Intro I - Statistiques inférentielles

2/30 (9/10)



**Comparaison** de données de dimension 1

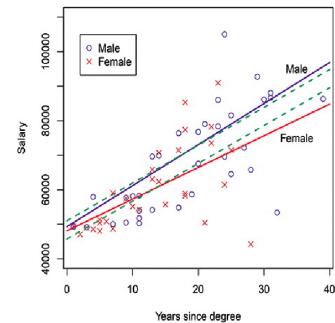
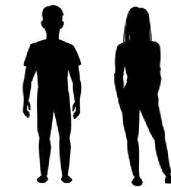
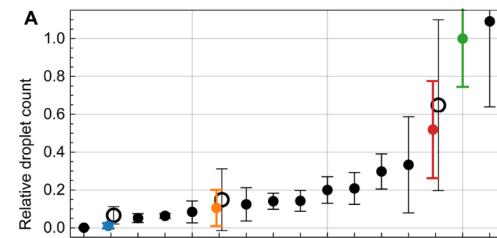
**Comparaison** de données de dimension 2



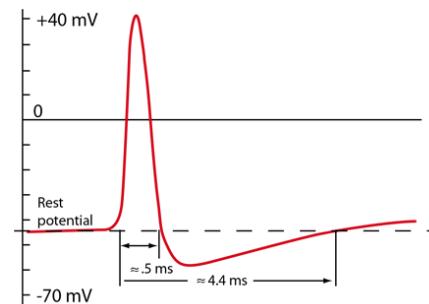
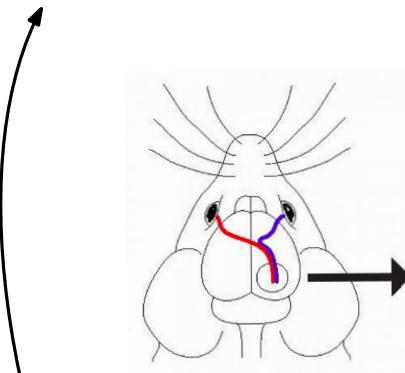
**Classification** de données de grande dimension

# Intro I - Statistiques inférentielles

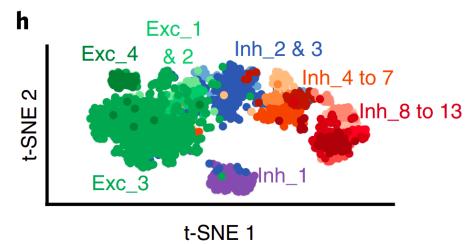
2/30 (10/10)



## Comparaison de données de dimension 1



## Comparaison de données de dimension 2



## Classification de données de grande dimension

1 - Poser une question

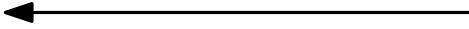
2 - Proposer un modèle statistique pour y répondre

# Intro II - Géométrie dans les données 3/30 (1/7)

[Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival, Monica Nicolau, Arnold J Levine, and Gunnar Carlsson, *Proceedings of the National Academy of Sciences*, 2011]

Les auteurs étudient les tissus de patientes infectées par le cancer du sein. Ils obtiennent 262 variables génomiques par patientes.

$$(x_1, x_2, \dots, x_{262})$$



# Intro II - Géométrie dans les données 3/30 (2/7)

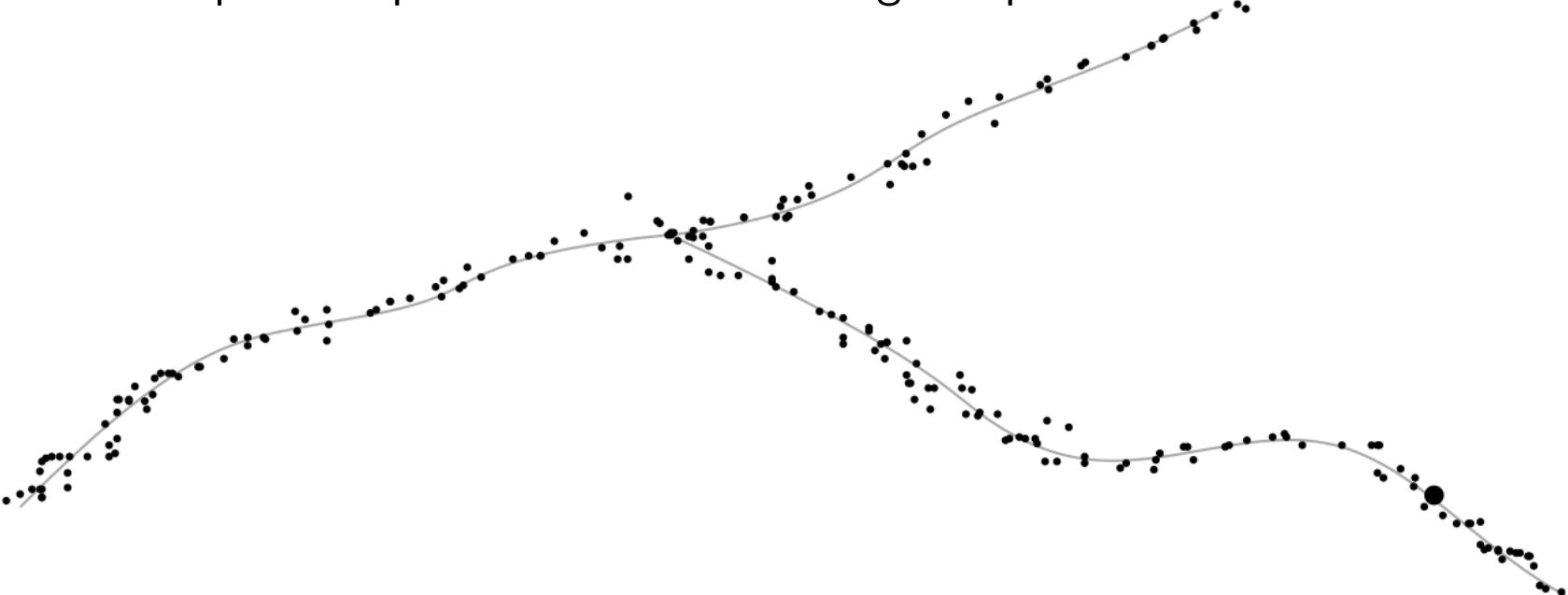
[Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival, Monica Nicolau, Arnold J Levine, and Gunnar Carlsson, *Proceedings of the National Academy of Sciences*, 2011]

Les auteurs étudient les tissus de patientes infectées par le cancer du sein. Ils obtiennent 262 variables génomiques par patientes.

$$(x_1, x_2, \dots, x_{262})$$



Rassembler plein de patientes donne un nuage de points dans  $\mathbb{R}^{262}$ .



# Intro II - Géométrie dans les données 3/30 (3/7)

[Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival, Monica Nicolau, Arnold J Levine, and Gunnar Carlsson, *Proceedings of the National Academy of Sciences*, 2011]

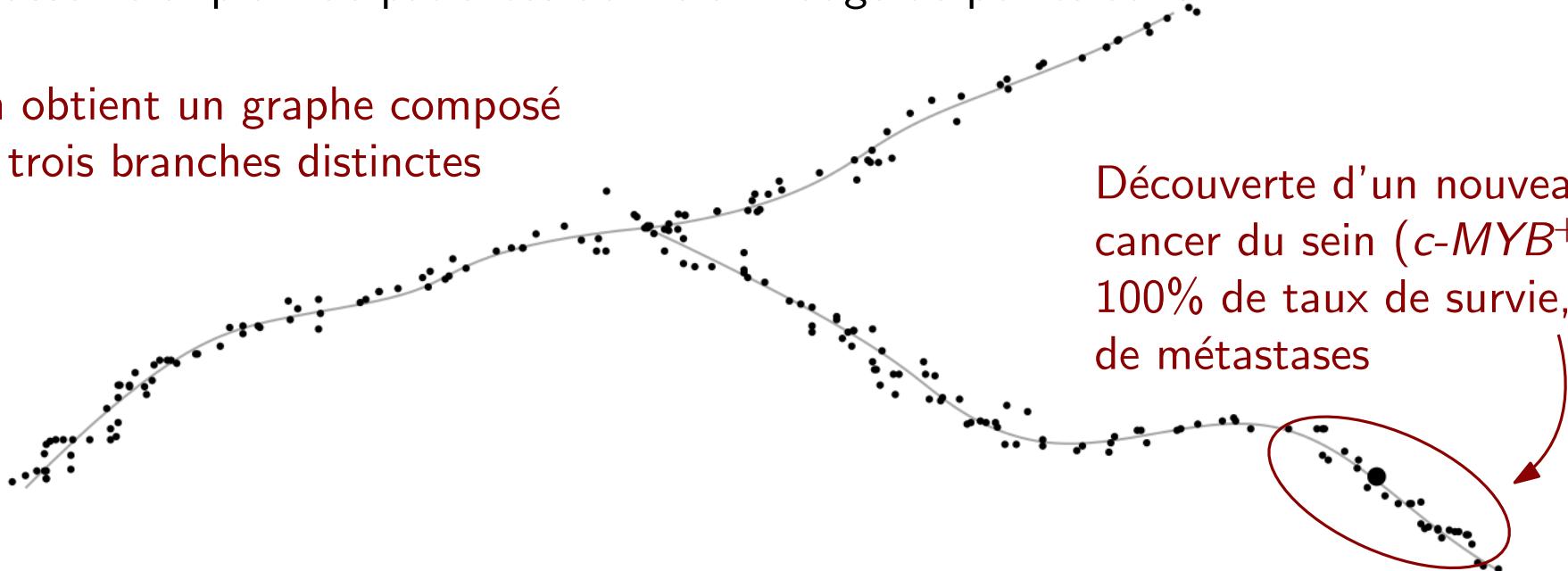
Les auteurs étudient les tissus de patientes infectées par le cancer du sein. Ils obtiennent 262 variables génomiques par patientes.

$$(x_1, x_2, \dots, x_{262})$$



Rassembler plein de patientes donne un nuage de points dans  $\mathbb{R}^{262}$ .

On obtient un graphe composé de trois branches distinctes



Découverte d'un nouveau type de cancer du sein ( $c\text{-}MYB^+$ )  
100% de taux de survie, et pas de métastases

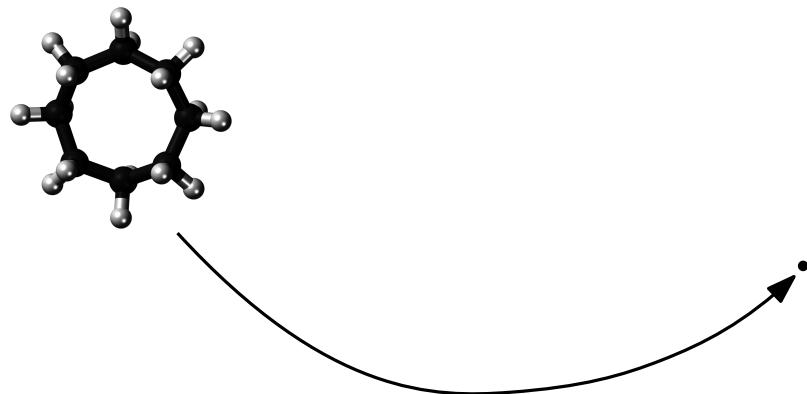
# Intro II - Géométrie dans les données 3/30 (4/7)

[Topology of cyclo-octane energy landscape, S. Martin, A. Thompson, E. A. Coutsias, J-P. Watson, *The Journal of Chemical Physics*, 2010]

La molécule de cyclo-octane  $C_8H_{16}$  contient 24 atomes.

Chaque atome a 3 coordonnées spatiales.

Ainsi, la conformation d'une molécule peut être résumée en **un point** dans  $\mathbb{R}^{72}$  ( $3 \times 24 = 72$ ).



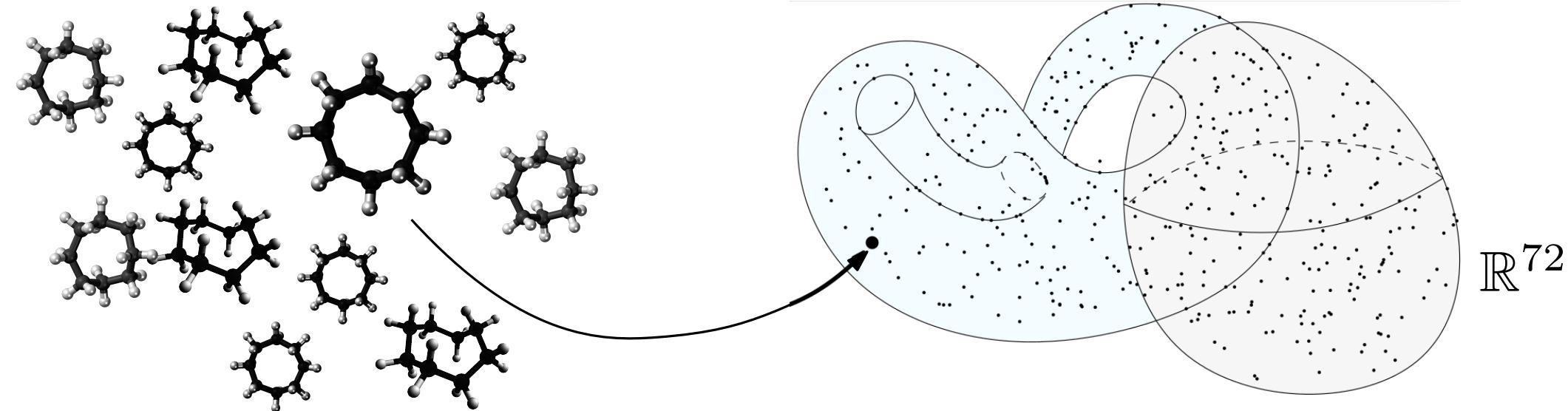
# Intro II - Géométrie dans les données 3/30 (5/7)

[Topology of cyclo-octane energy landscape, S. Martin, A. Thompson, E. A. Coutsias, J-P. Watson, *The Journal of Chemical Physics*, 2010]

La molécule de cyclo-octane  $C_8H_{16}$  contient 24 atomes.

Chaque atome a 3 coordonnées spatiales.

Ainsi, la conformation d'une molécule peut être résumée en **un point** dans  $\mathbb{R}^{72}$  ( $3 \times 24 = 72$ ).



En étudiant plein de molécules, on obtient un **nuage de points** dans  $\mathbb{R}^{72}$ .

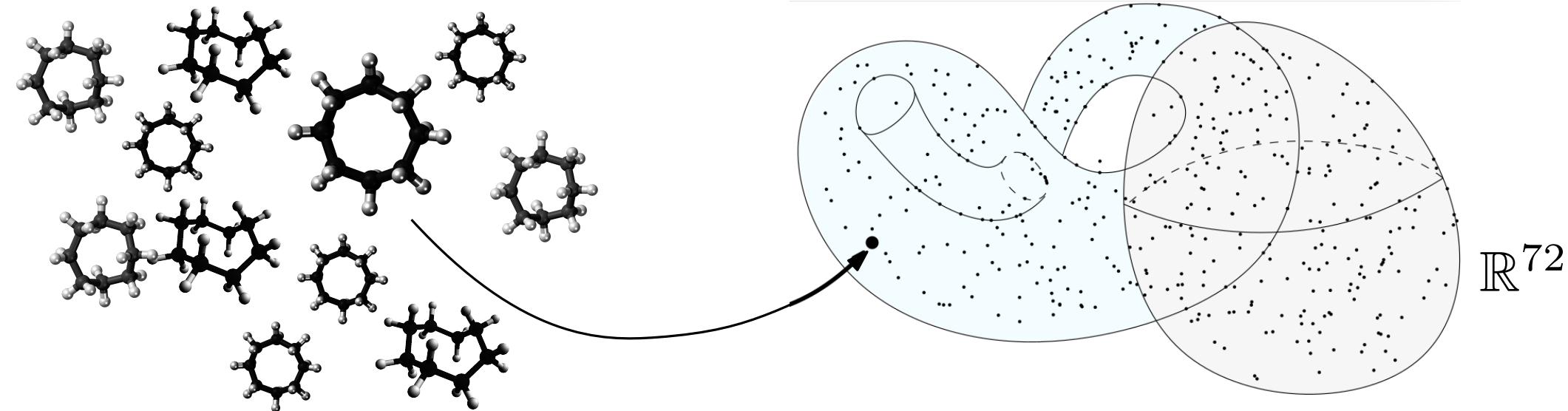
# Intro II - Géométrie dans les données 3/30 (6/7)

[Topology of cyclo-octane energy landscape, S. Martin, A. Thompson, E. A. Coutsias, J-P. Watson, *The Journal of Chemical Physics*, 2010]

La molécule de cyclo-octane  $C_8H_{16}$  contient 24 atomes.

Chaque atome a 3 coordonnées spatiales.

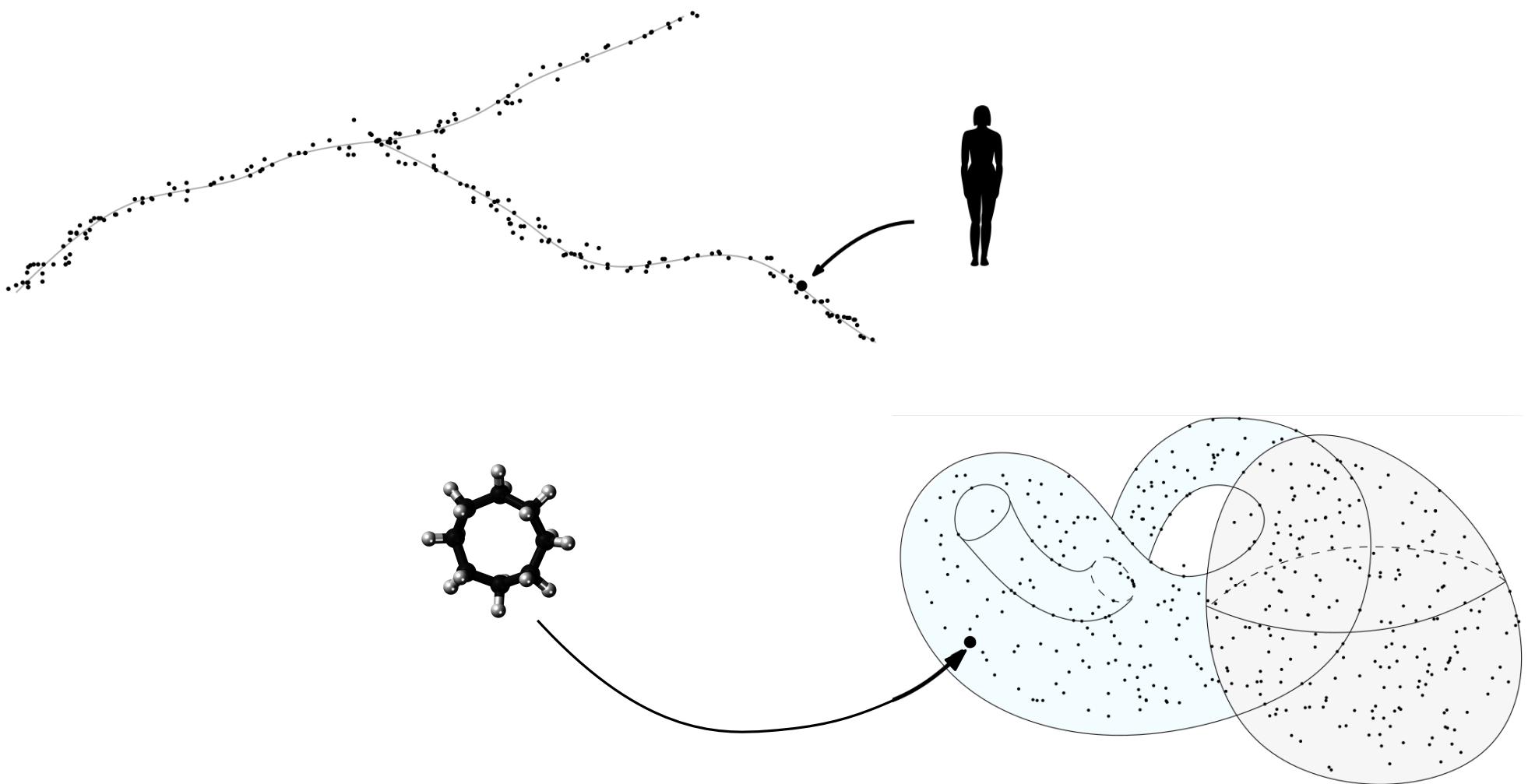
Ainsi, la conformation d'une molécule peut être résumée en **un point** dans  $\mathbb{R}^{72}$  ( $3 \times 24 = 72$ ).



En étudiant plein de molécules, on obtient un **nuage de points** dans  $\mathbb{R}^{72}$ .

Les auteurs montrent que ce nuage de points est proche d'un objet de petite dimension : **l'union d'une sphère et d'une bouteille de Klein**.

# Intro II - Géométrie dans les données 3/30 (7/7)



Question : quelle est la forme des données ?

ما هو شكل البيانات ؟

# I - Comparer les espaces topologiques

- 1 - Équivalence d'homéomorphie
- 2 - Équivalence d'homotopie

# II - Invariants topologiques

- 1 - Nombre de composantes connexes
- 2 - Caractéristique d'Euler
- 3 - Nombres de Betti

(Prochaine séance - Homologie persistante)

En topologie, nous étudions les **espaces topologiques**.

**Définition** : un espace topologique est un ensemble  $X$  muni d'une collection d'**ouverts**  $\{O_\alpha, \alpha \in A\}$ , avec  $O_\alpha \subset X$ , tels que

- $\emptyset$  et  $X$  sont des ouverts,
- une union infinie d'ouverts est un ouvert,
- une intersection finie d'ouverts est un ouvert.

**Définition** : Étant donné deux espaces topologiques  $X$  et  $Y$ , une application  $f: X \rightarrow Y$  est dite **continue** si pour tout ouvert  $O \subset Y$ , la pré-image  $f^{-1}(O)$  est un ouvert de  $X$ .

$$X \xrightarrow{\hspace{1cm}} Y$$

En topologie, nous étudions les **espaces topologiques**.

**Définition** : un espace topologique est un ensemble  $X$  muni d'une collection d'**ouverts**  $\{O_\alpha, \alpha \in A\}$ , avec  $O_\alpha \subset X$ , tels que

- $\emptyset$  et  $X$  sont des ouverts,
- une union infinie d'ouverts est un ouvert,
- une intersection finie d'ouverts est un ouvert.

**Définition** : Étant donné deux espaces topologiques  $X$  et  $Y$ , une application  $f: X \rightarrow Y$  est dite **continue** si pour tout ouvert  $O \subset Y$ , la pré-image  $f^{-1}(O)$  est un ouvert de  $X$ .

$$X \xrightarrow{\hspace{1cm}} Y$$

traduction dans le  
formalisme  $\epsilon$ - $\delta$

On peut penser par exemple aux **sous-ensembles**  $X \subset \mathbb{R}^n$  et  $Y \subset \mathbb{R}^m$ ,

et aux applications  $f: X \rightarrow Y$  **continues** au sens suivant :

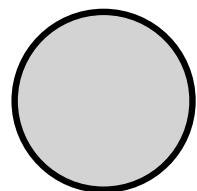
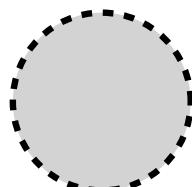
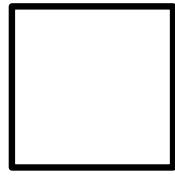
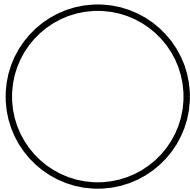
$$\forall x \in X, \forall \epsilon > 0, \exists \eta > 0, \forall y \in X, \|x - y\| < \eta \implies \|f(x) - f(y)\| < \epsilon.$$

# Exemples d'espaces topologiques

6/30 (1/3)

Dans  $\mathbb{R}^n$ , on peut définir :

- la sphère unité  $\mathbb{S}_{n-1} = \{x \in \mathbb{R}^n, \|x\| = 1\}$
- le cube unité  $\mathcal{C}_{n-1} = \{x = (x_1, \dots, x_n) \in \mathbb{R}^n, \max(|x_1|, \dots, |x_n|) = 1\}$
- les boules ouvertes  $\mathcal{B}(x, r) = \{y \in \mathbb{R}^n, \|x - y\| < r\}$
- les boules fermées  $\overline{\mathcal{B}}(x, r) = \{y \in \mathbb{R}^n, \|x - y\| \leq r\}$



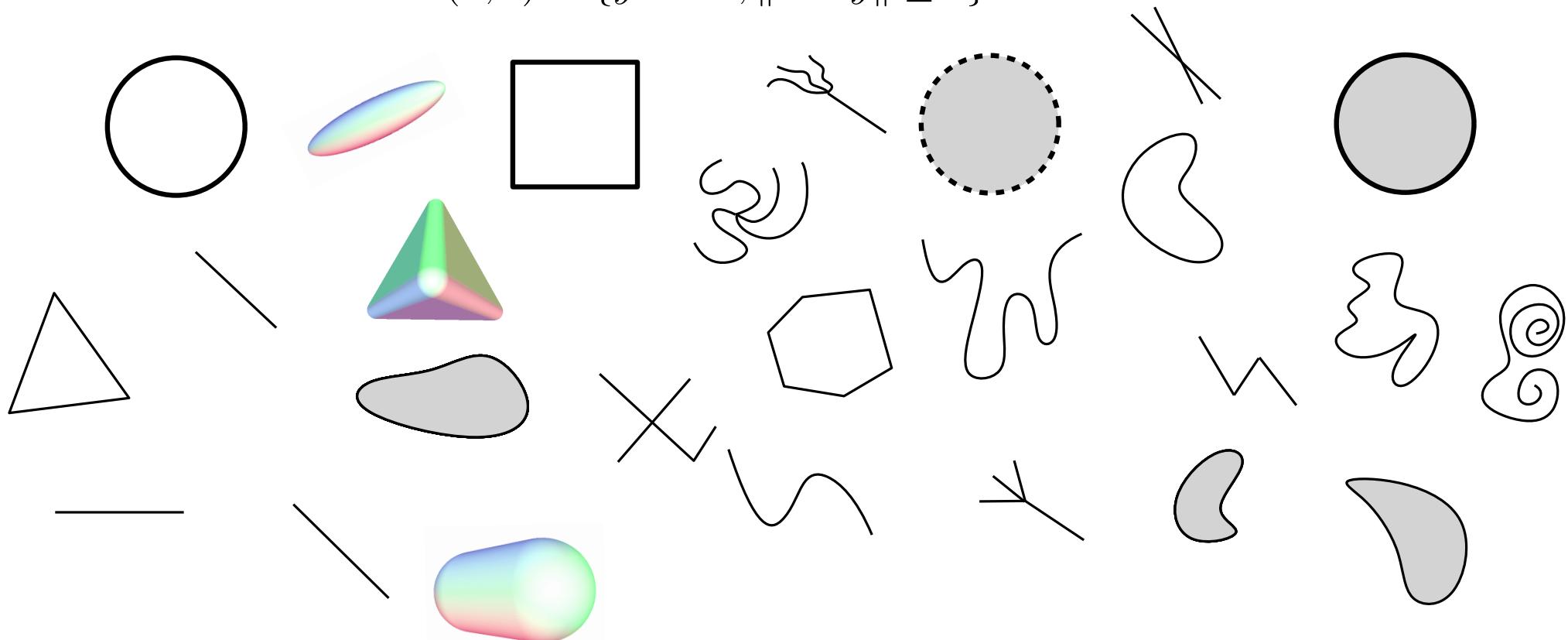
La notation  $\|x\|$  représente la norme euclidienne : si  $x = (x_1, \dots, x_n)$ , alors  $\|x\| = \sqrt{x_1^2 + \dots + x_n^2}$ .

# Exemples d'espaces topologiques

6/30 (2/3)

Dans  $\mathbb{R}^n$ , on peut définir :

- la sphère unité  $\mathbb{S}_{n-1} = \{x \in \mathbb{R}^n, \|x\| = 1\}$
- le cube unité  $\mathcal{C}_{n-1} = \{x = (x_1, \dots, x_n) \in \mathbb{R}^n, \max(|x_1|, \dots, |x_n|) = 1\}$
- les boules ouvertes  $\mathcal{B}(x, r) = \{y \in \mathbb{R}^n, \|x - y\| < r\}$
- les boules fermées  $\overline{\mathcal{B}}(x, r) = \{y \in \mathbb{R}^n, \|x - y\| \leq r\}$



En général, les espaces topologiques n'admettent pas de jolie définition algébrique...

# Exemples d'espaces topologiques

6/30 (3/3)

Dans  $\mathbb{R}^n$ , on peut définir :

- la sphère unité  $\mathbb{S}_{n-1} = \{x \in \mathbb{R}^n, \|x\| = 1\}$
- le cube unité  $\mathcal{C}_{n-1} = \{x = (x_1, \dots, x_n) \in \mathbb{R}^n, \max(|x_1|, \dots, |x_n|) = 1\}$
- les boules ouvertes  $\mathcal{B}(x, r) = \{y \in \mathbb{R}^n, \|x - y\| < r\}$
- les boules fermées  $\overline{\mathcal{B}}(x, r) = \{y \in \mathbb{R}^n, \|x - y\| \leq r\}$



En général, les espaces topologiques n'admettent pas de jolie définition algébrique...

# I - Comparer les espaces topologiques

1 - Équivalence d'homéomorphie

2 - Équivalence d'homotopie

# II - Invariants topologiques

1 - Nombre de composantes connexes

2 - Caractéristique d'Euler

3 - Nombres de Betti

(Prochaine séance - Homologie persistante)

**Définition :** Soient  $X$  et  $Y$  deux espaces topologiques, et  $f: X \rightarrow Y$  une application.

On dit que  $f$  est un **homéomorphisme** si

- $f$  est bijective,
- $f: X \rightarrow Y$  est continue,
- $f^{-1}: Y \rightarrow X$  est continue.

S'il existe un tel homéomorphisme, on dit que les deux espaces topologiques sont **homéomorphes**.

# Homéomorphismes

8/30 (2/3)

**Définition :** Soient  $X$  et  $Y$  deux espaces topologiques, et  $f: X \rightarrow Y$  une application.

On dit que  $f$  est un **homéomorphisme** si

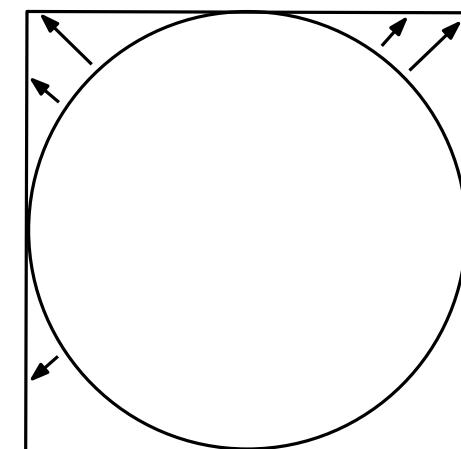
- $f$  est bijective,
- $f: X \rightarrow Y$  est continue,
- $f^{-1}: Y \rightarrow X$  est continue.

S'il existe un tel homéomorphisme, on dit que les deux espaces topologiques sont **homéomorphes**.

**Exemple :** Le cercle et le carré sont homéomorphes, et un homéomorphisme est donné par

$f: \text{Cercle} \longrightarrow \text{Carré}$

$$(x_1, x_2) \longmapsto \frac{1}{\max(|x_1|, |x_2|)}(x_1, x_2)$$



**Interprétation:** Les homéomorphismes réalisent des *déformations continues*.

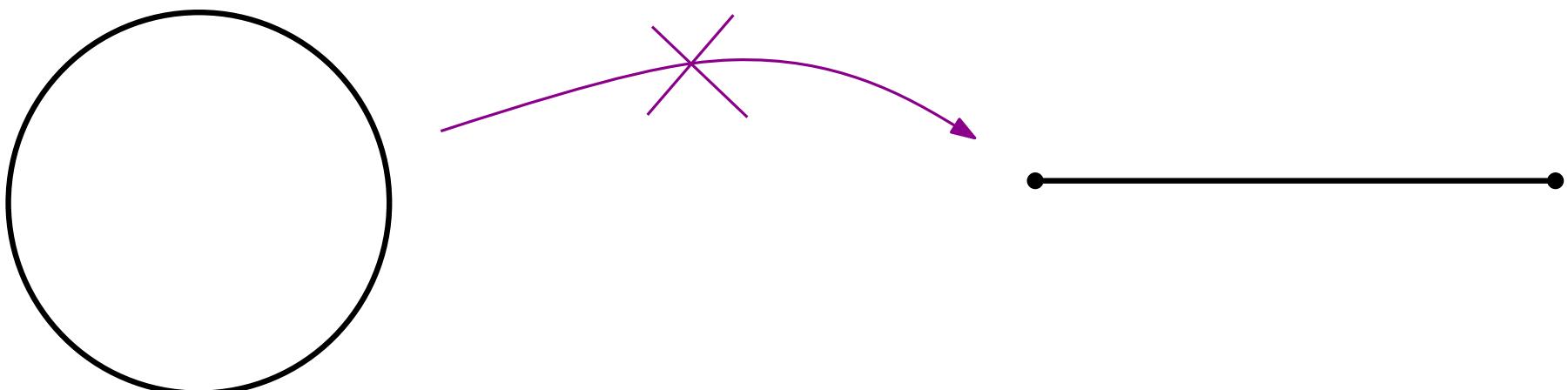
Définition : Soient  $X$  et  $Y$  deux espaces topologiques, et  $f: X \rightarrow Y$  une application.

On dit que  $f$  est un **homéomorphisme** si

- $f$  est bijective,
- $f: X \rightarrow Y$  est continue,
- $f^{-1}: Y \rightarrow X$  est continue.

S'il existe un tel homéomorphisme, on dit que les deux espaces topologiques sont **homéomorphes**.

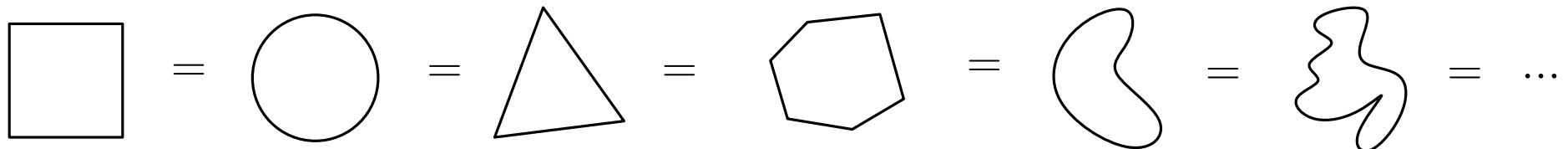
Exemple : Le cercle et l'intervalle  $[0, 1]$  ne sont pas homéomorphes.



# Classes d'homéomorphie

9/30 (1/5)

On peut rassembler les espaces topologiques qui sont homéomorphes :

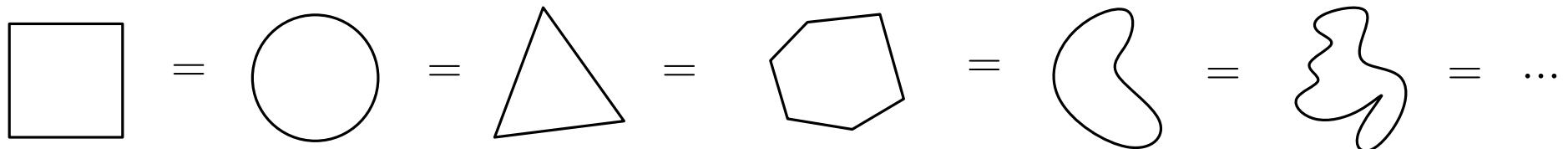


la classe des cercles

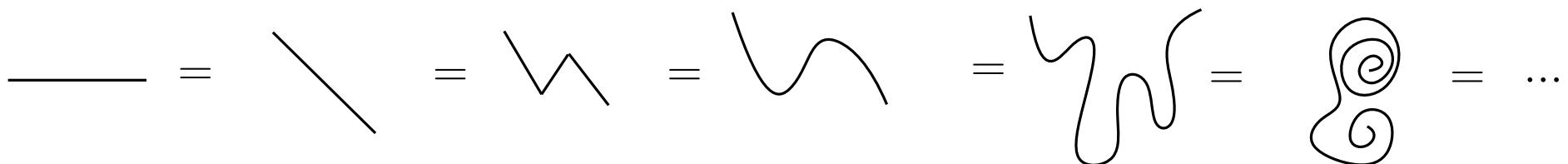
# Classes d'homéomorphie

9/30 (2/5)

On peut rassembler les espaces topologiques qui sont homéomorphes :



la classe des cercles

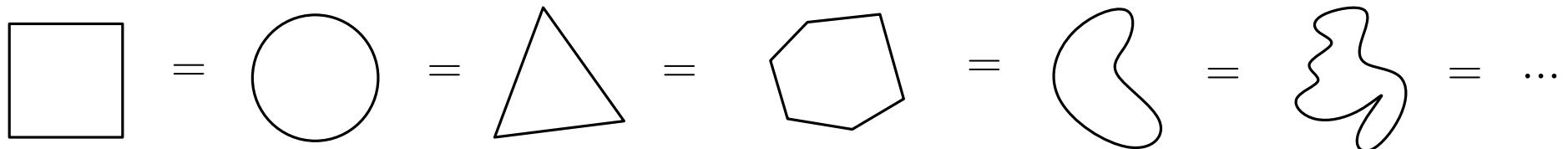


la classe des intervalles

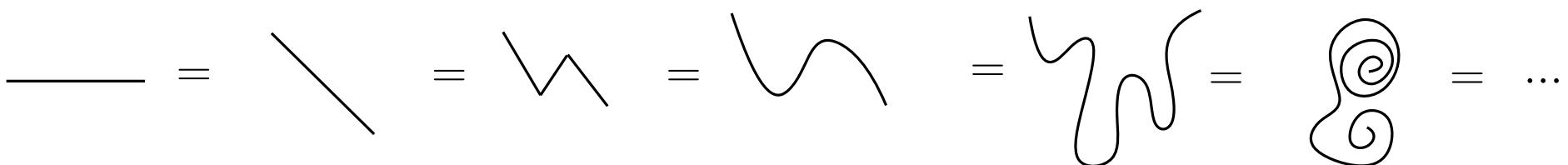
# Classes d'homéomorphie

9/30 (3/5)

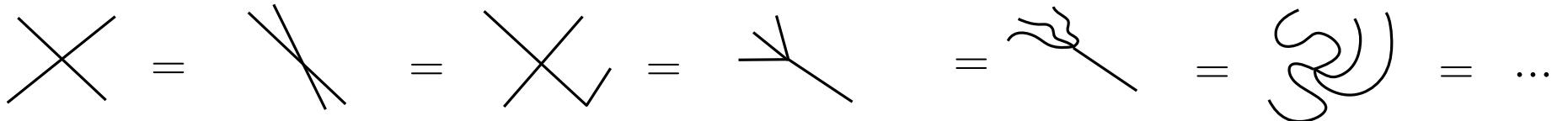
On peut rassembler les espaces topologiques qui sont homéomorphes :



la classe des cercles



la classe des intervalles

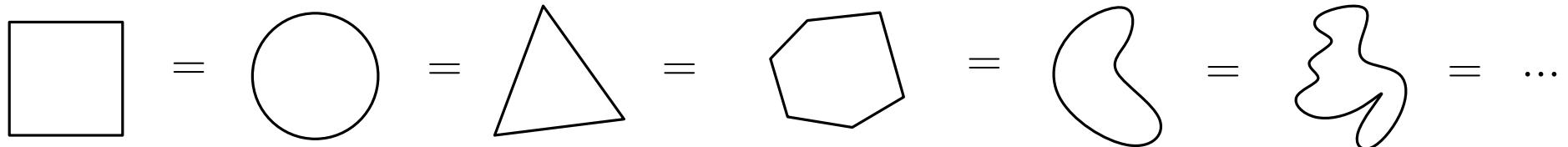


la classe des croix

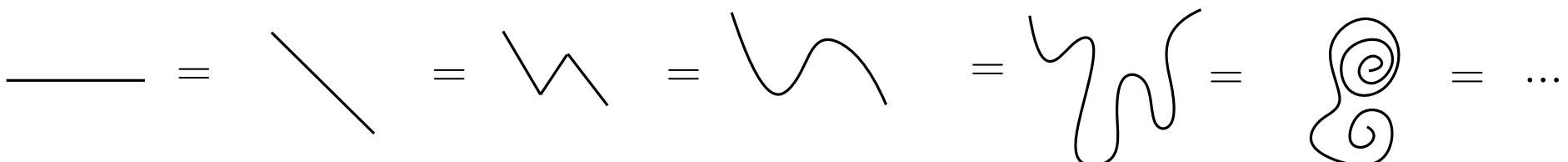
# Classes d'homéomorphie

9/30 (4/5)

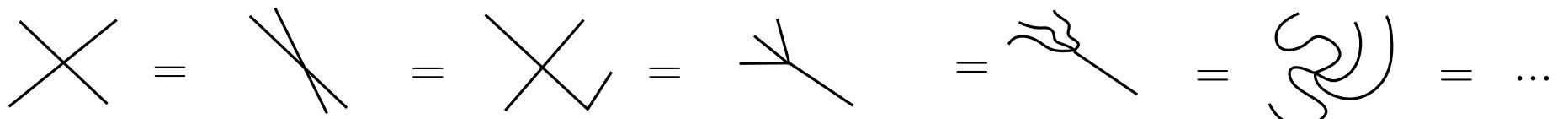
On peut rassembler les espaces topologiques qui sont homéomorphes :



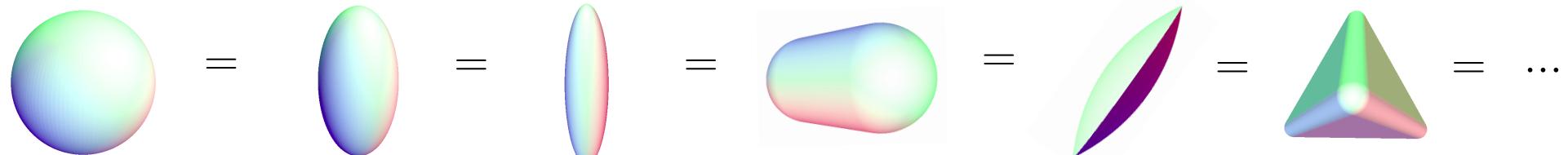
la classe des cercles



la classe des intervalles



la classe des croix

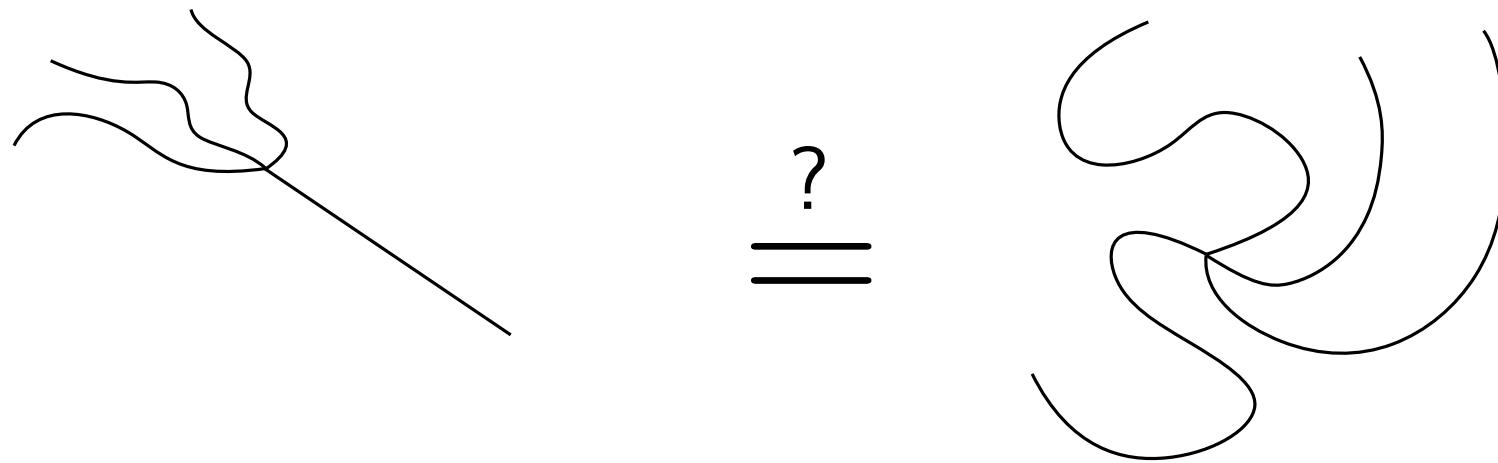


la classe des sphères

# Classes d'homéomorphie

9/30 (5/5)

En général, il est difficile de déterminer si deux espaces donnés sont homéomorphes.



Pour cela, nous allons utiliser la notion d'**invariant**.

# I - Comparer les espaces topologiques

1 - Équivalence d'homéomorphie

2 - Équivalence d'homotopie

# II - Invariants topologiques

1 - Nombre de composantes connexes

2 - Caractéristique d'Euler

3 - Nombres de Betti

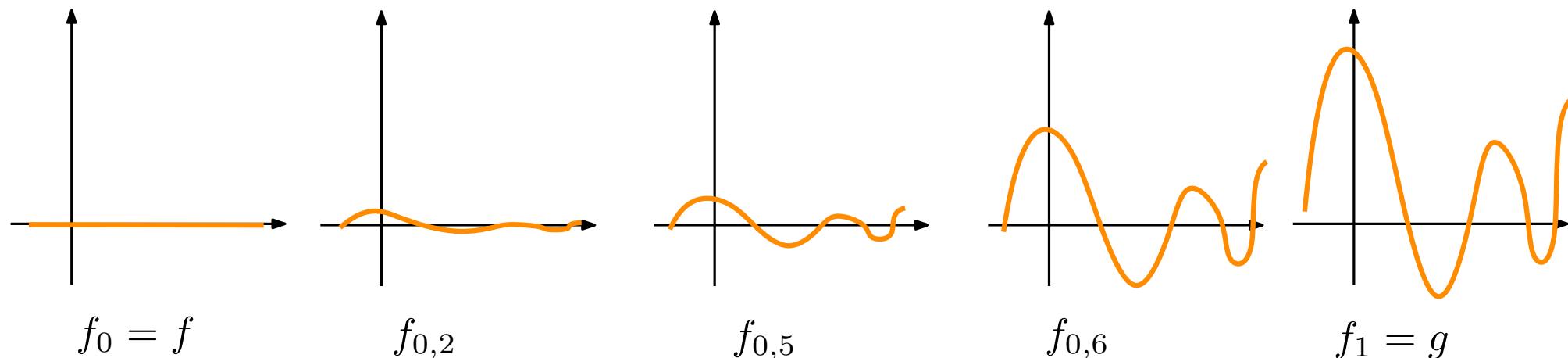
(Prochaine séance - Homologie persistante)

**Définition :** Soient  $X, Y$  deux espaces topologiques et  $f, g: X \rightarrow Y$  deux applications continues. Une **homotopie** entre  $f$  et  $g$  est une famille d'applications  $(f_t: X \rightarrow Y)_{t \in [0,1]}$  telle que :

- $f_0 = f$ ,
- $f_1 = g$ ,
- $F: X \times [0, 1] \rightarrow Y$  définie par  $F(x, t) = f_t(x)$  est continue.

Si une telle homotopie existe, on dit que les applications  $f$  et  $g$  sont **homotopes**.

**Exemple :** Une homotopie entre  $f: \mathbb{R} \rightarrow \mathbb{R}$  et  $g: \mathbb{R} \rightarrow \mathbb{R}$ .



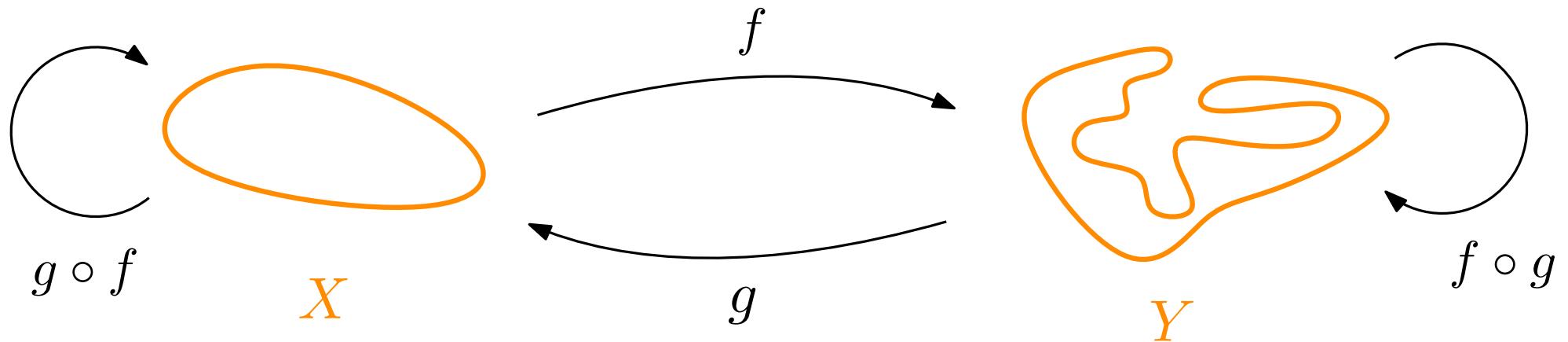
# Équivalence d'homotopie

12/30 (1/3)

**Définition :** Soient  $X$  et  $Y$  deux espaces topologiques. Une **équivalence d'homotopie** entre  $X$  et  $Y$  est une paire d'applications continues  $f: X \rightarrow Y$  et  $g: Y \rightarrow X$  telles que :

- $g \circ f: X \rightarrow X$  est homotope à l'application identité  $\text{id}: X \rightarrow X$ ,
- $f \circ g: Y \rightarrow Y$  est homotope à l'application identité  $\text{id}: Y \rightarrow Y$ .

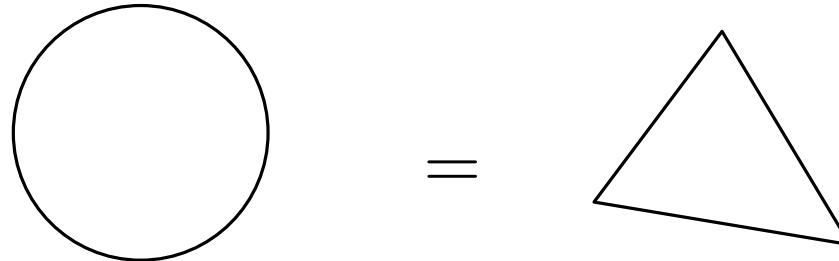
Si une telle équivalence d'homotopie existe, on dit que  $X$  et  $Y$  sont **homotopiquement équivalents**.



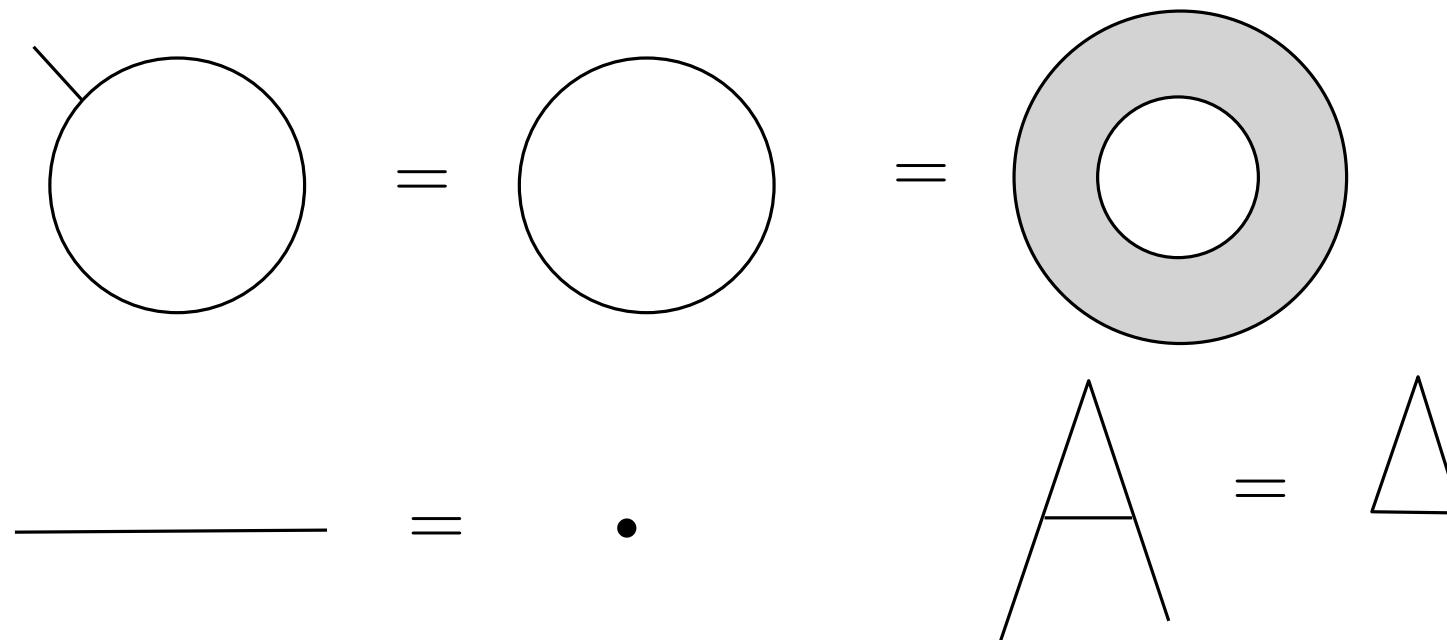
# Équivalence d'homotopie

12/30 (2/3)

L'équivalence d'homotopie permet de **déformer** continûment l'espace



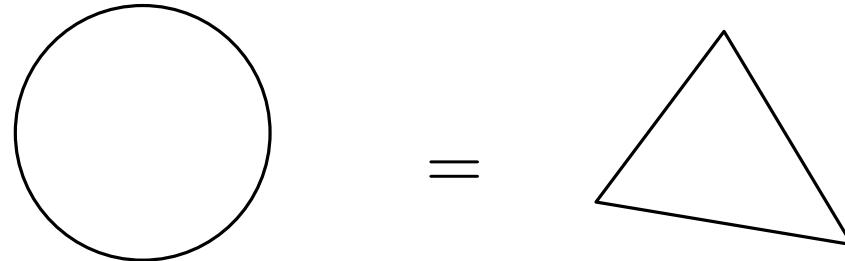
mais aussi de le **rétracter**.



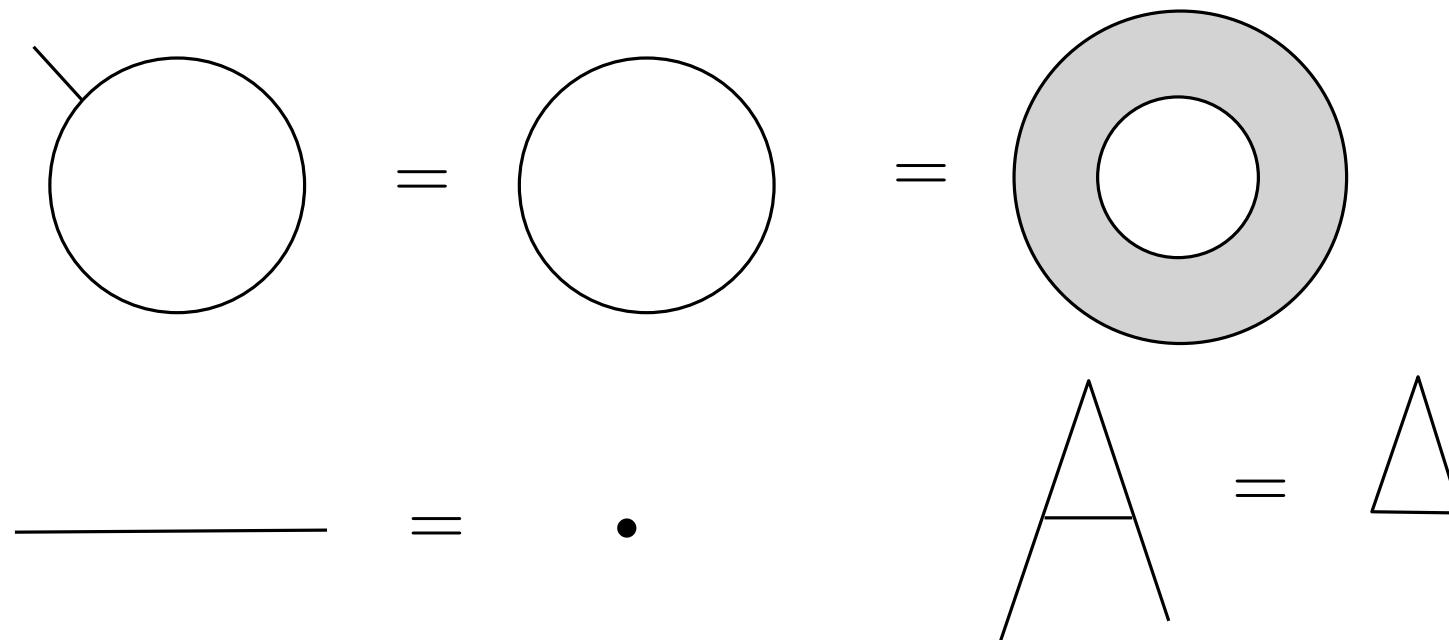
# Équivalence d'homotopie

12/30 (3/3)

L'équivalence d'homotopie permet de **déformer** continûment l'espace



mais aussi de le **rétracter**.

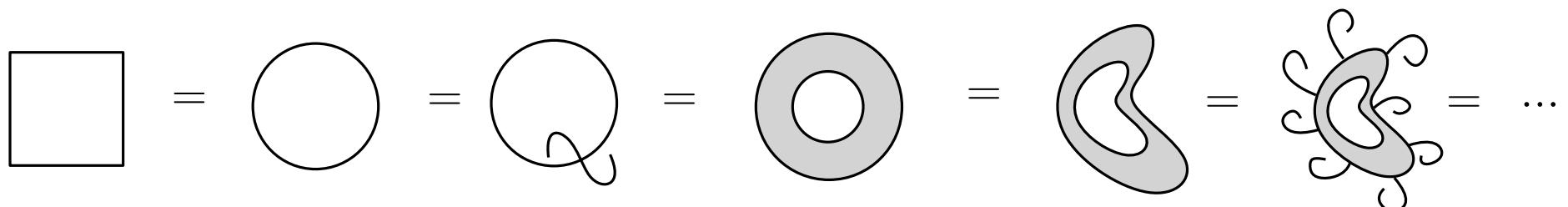


**Proposition :** Si deux espaces sont homéomorphes, alors ils sont homotopiquement équivalents.

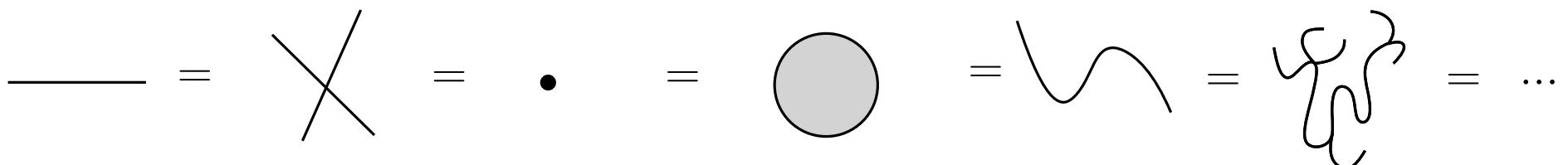
# Classes d'homotopies

13/30 (1/4)

Comme tout à l'heure, on peut rassembler les espaces homotopiquement équivalents, et obtenir des **classes d'homotopie** :



la classe des cercles



la classe des points

la classe des sphères, la classe des tores, la classe des bouteilles de Klein, ...

Exemple : Classification, pour l'équivalence d'homotopie, des lettres de l'alphabet.

A      B      C      D      E      F

G      H      I      J      K      L

M      N      O      P      Q      R

S      T      U      V      W      X

Y      Z

# Classes d'homotopies

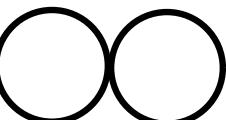
13/30 (3/4)

Exemple : Classification, pour l'équivalence d'homotopie, des lettres de l'alphabet.

A D O P Q R

$\approx$  

B

$\approx$  

C E F G H I J K L

M N S T U V W X Y Z

$\approx$  

Exercice : Classer, pour l'équivalence d'homotopie, les lettres de l'alphabet arabe.

ا ب ت ث ح ح خ

د ذ ر ز س ش ص

ض ط ظ ع غ ف ق

ك ل م ن ه و ي

## I - Comparer les espaces topologiques

- 1 - Équivalence d'homéomorphie
- 2 - Équivalence d'homotopie

## II - Invariants topologiques

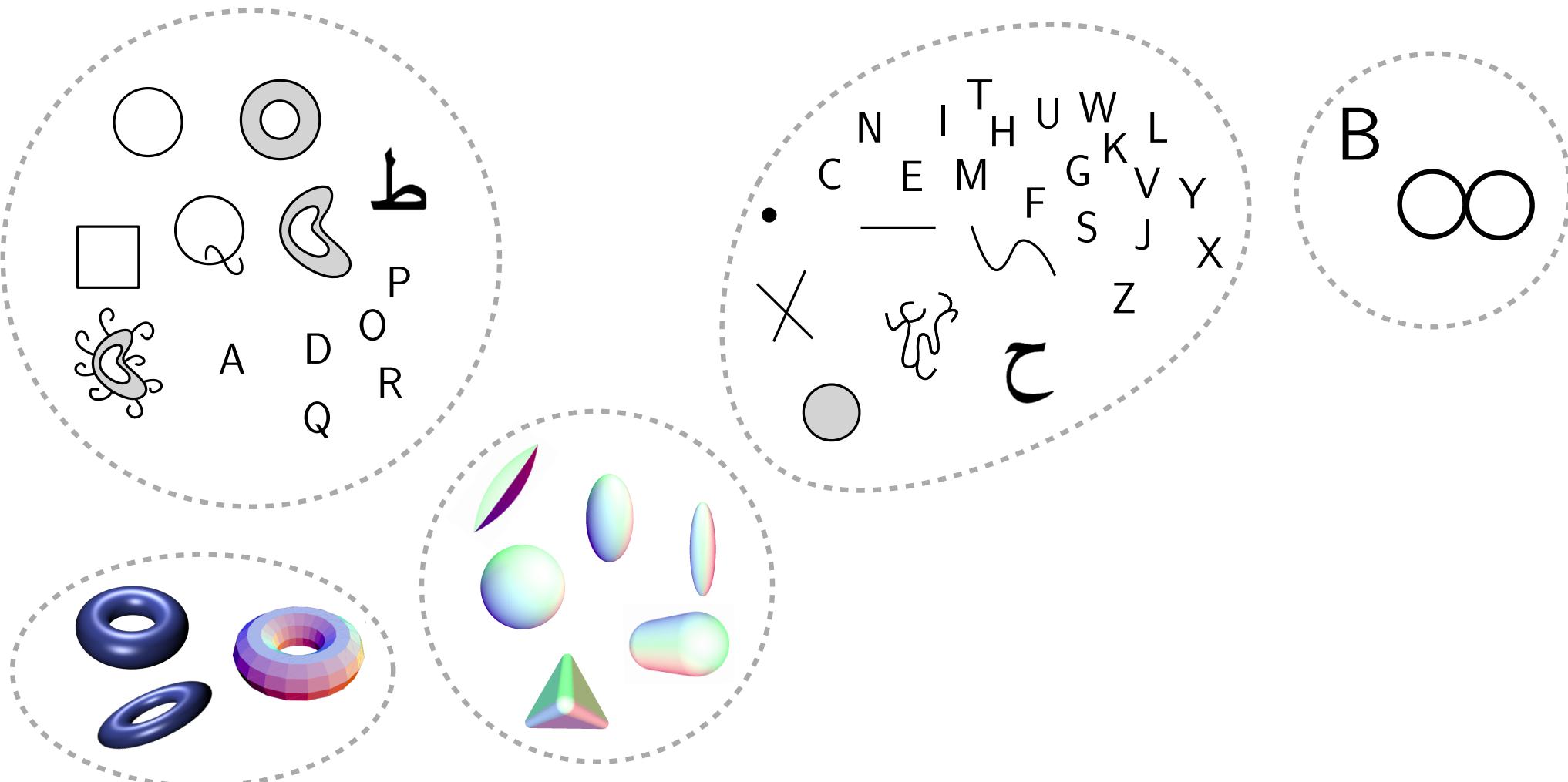
- 1 - Nombre de composantes connexes
- 2 - Caractéristique d'Euler
- 3 - Nombres de Betti

(Prochaine séance - Homologie persistante)

# Propriétés invariantes

15/30 (1/2)

Nous avons rassemblé les espaces topologiques en classes d'homotopie.

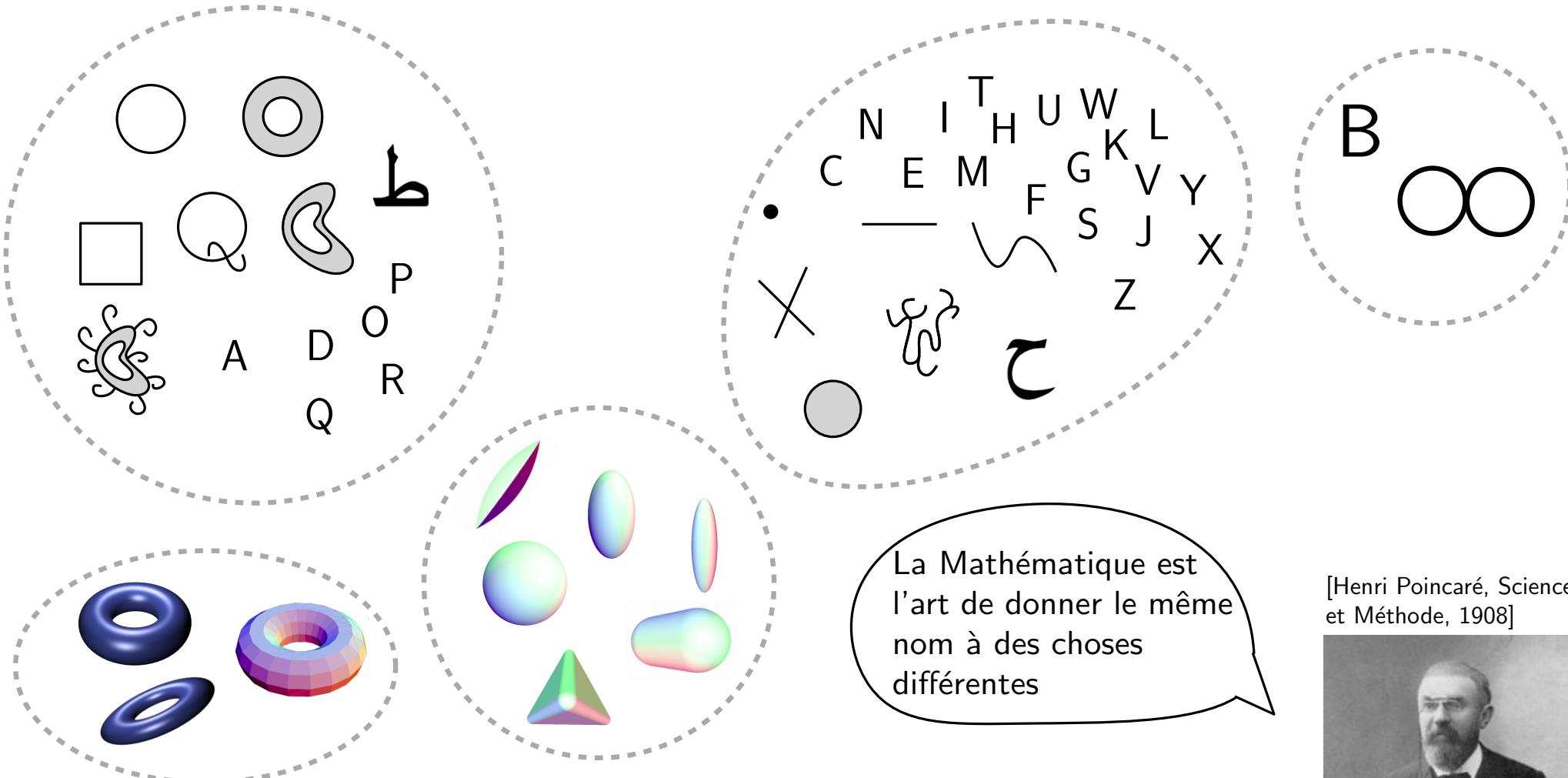


- Étant donné un espace topologique  $X$ , comment reconnaître dans quelle classe il appartient ?
- Quelles sont les **propriétés** communes des espaces d'une même classe ?

# Propriétés invariantes

15/30 (2/2)

Nous avons rassemblé les espaces topologiques en classes d'homotopie.



[Henri Poincaré, Science et Méthode, 1908]



- Étant donné un espace topologique  $X$ , comment reconnaître dans quelle classe il appartient ?
- Quelles sont les **propriétés** communes des espaces d'une même classe ?

## I - Comparer les espaces topologiques

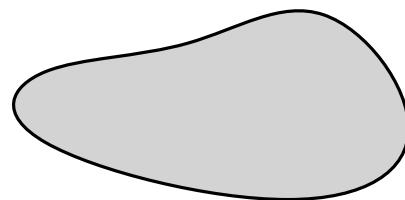
- 1 - Équivalence d'homéomorphie
- 2 - Équivalence d'homotopie

## II - Invariants topologiques

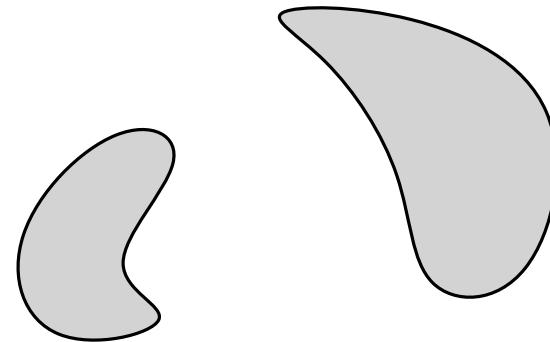
- 1 - Nombre de composantes connexes
- 2 - Caractéristique d'Euler
- 3 - Nombres de Betti

(Prochaine séance - Homologie persistante)

Définition : Un sous-ensemble  $X \subset \mathbb{R}^n$  est **connexe** (par arcs) si pour tout  $x, y \in X$ , il existe une application continue  $f: [0, 1] \rightarrow X$  telle que  $f(0) = x$  et  $f(1) = y$ .



espace connexe

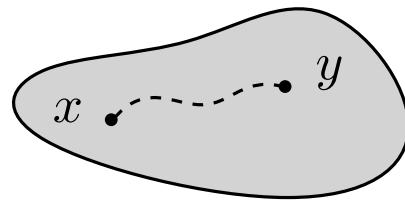


espace non-connexe

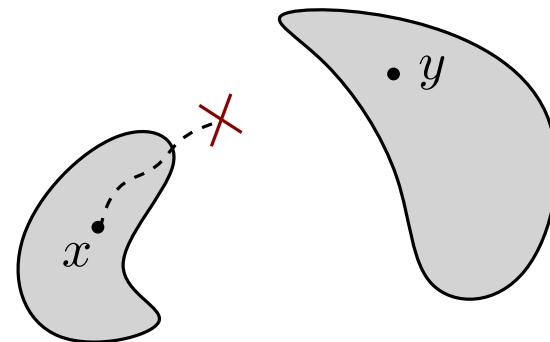
# Connexité

17/30 (2/4)

Définition : Un sous-ensemble  $X \subset \mathbb{R}^n$  est **connexe** (par arcs) si pour tout  $x, y \in X$ , il existe une application continue  $f: [0, 1] \rightarrow X$  telle que  $f(0) = x$  et  $f(1) = y$ .

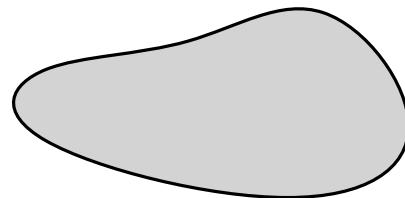


espace connexe

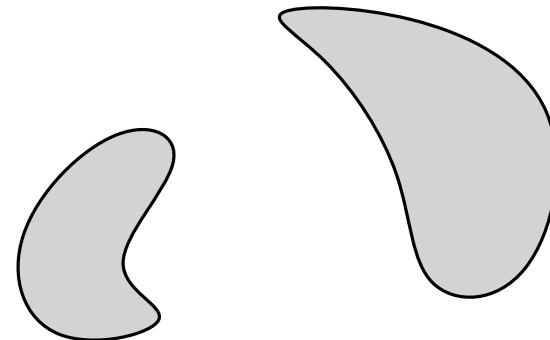


espace non-connexe

Définition : Un sous-ensemble  $X \subset \mathbb{R}^n$  est **connexe** (par arcs) si pour tout  $x, y \in X$ , il existe une application continue  $f: [0, 1] \rightarrow X$  telle que  $f(0) = x$  et  $f(1) = y$ .

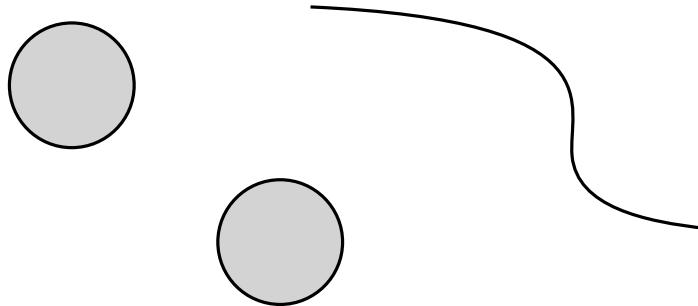
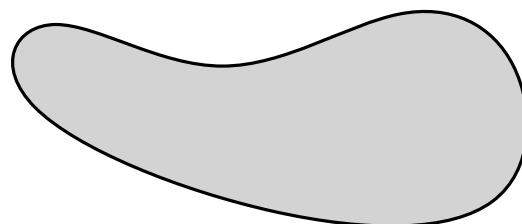


espace connexe

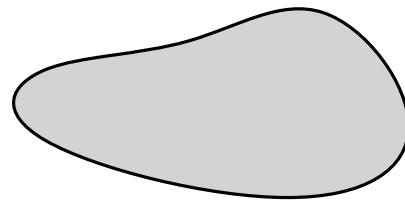


espace non-connexe

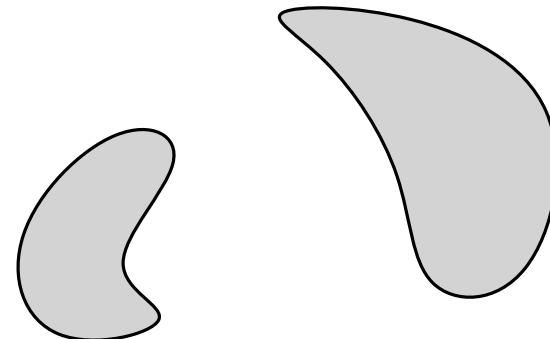
Un espace  $X$  peut être découpé en **composantes connexes** (ce sont les sous-ensembles connexes maximaux).



**Définition :** Un sous-ensemble  $X \subset \mathbb{R}^n$  est **connexe** (par arcs) si pour tout  $x, y \in X$ , il existe une application continue  $f: [0, 1] \rightarrow X$  telle que  $f(0) = x$  et  $f(1) = y$ .

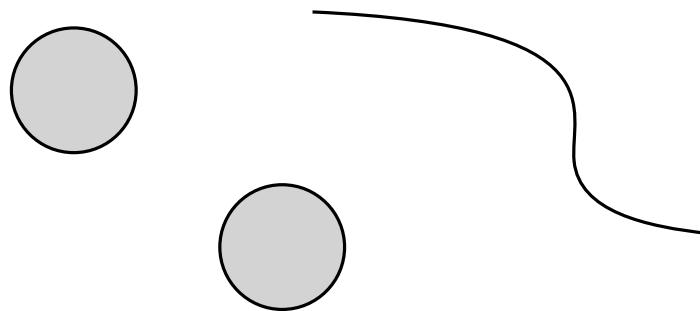
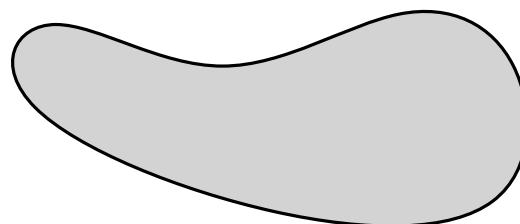


espace connexe



espace non-connexe

Un espace  $X$  peut être découpé en **composantes connexes** (ce sont les sous-ensembles connexes maximaux).



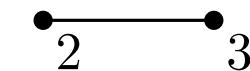
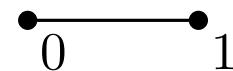
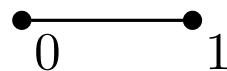
**Proposition :** Si deux espaces  $X$  et  $Y$  sont homotopiquement équivalents, alors ils ont le même nombre de composantes connexes.

**Proposition :** Si deux espaces  $X$  et  $Y$  sont homotopiquement équivalents, alors ils ont le même nombre de composantes connexes.

**Corollaire :** Si deux espaces  $X$  et  $Y$  sont homéomorphes, alors ils ont le même nombre de composantes connexes.

**Exemple :** Les sous-ensembles  $[0, 1]$  et  $[0, 1] \cup [2, 3]$  de  $\mathbb{R}$  ne sont pas homéomorphes, ni homotopiquement équivalents.

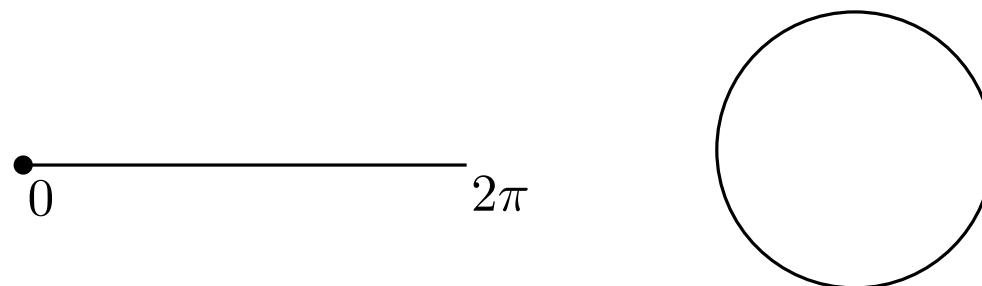
En effet, le premier n'a qu'une seule composante connexe, et le deuxième en a deux.



**Proposition :** Si deux espaces  $X$  et  $Y$  sont homotopiquement équivalents, alors ils ont le même nombre de composantes connexes.

**Exemple :** L'intervalle  $[0, 2\pi)$  et le cercle  $\mathbb{S}_1 \subset \mathbb{R}^2$  ne sont pas homéomorphes.

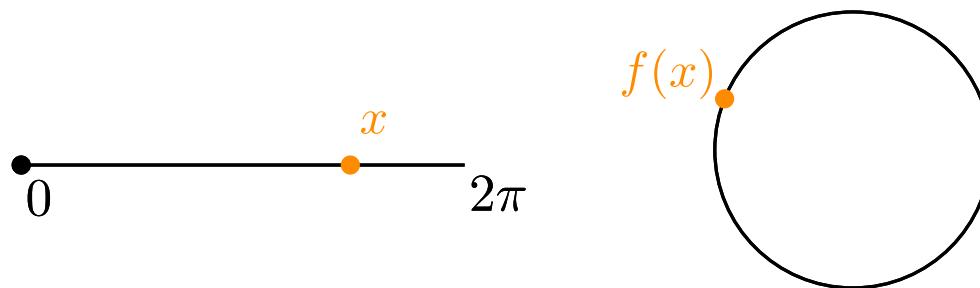
Nous allons le prouver par contradiction. Supposons qu'ils sont homéomorphes. Par définition, cela signifie qu'il existe une application  $f: [0, 2\pi) \rightarrow \mathbb{S}_1$  continue, bijective et d'inverse continue.



**Proposition :** Si deux espaces  $X$  et  $Y$  sont homotopiquement équivalents, alors ils ont le même nombre de composantes connexes.

**Exemple :** L'intervalle  $[0, 2\pi)$  et le cercle  $\mathbb{S}_1 \subset \mathbb{R}^2$  ne sont pas homéomorphes.

Nous allons le prouver par contradiction. Supposons qu'ils sont homéomorphes. Par définition, cela signifie qu'il existe une application  $f: [0, 2\pi) \rightarrow \mathbb{S}_1$  continue, bijective et d'inverse continue.



Soit  $x \in [0, 2\pi)$  tel que  $x \neq 0$ . Considérons les sous-ensembles  $[0, 2\pi) \setminus \{x\} \subset [0, 2\pi)$  et  $\mathbb{S}_1 \setminus \{f(x)\} \subset \mathbb{S}_1$ , et l'application induite

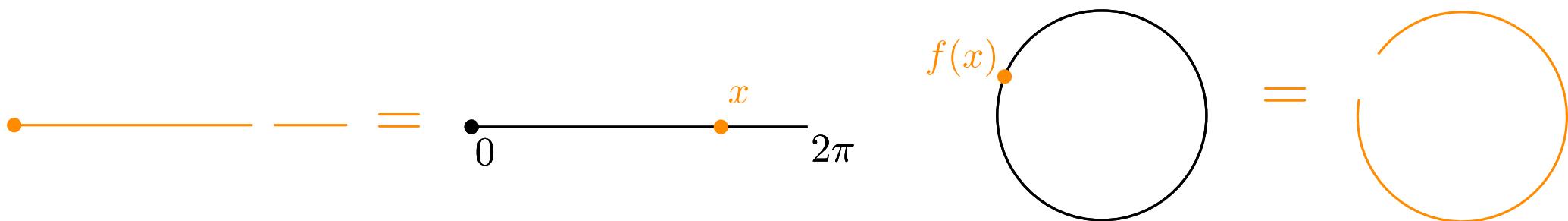
$$g: [0, 2\pi) \setminus \{x\} \rightarrow \mathbb{S}_1 \setminus \{f(x)\}.$$

L'application  $g$  est un homéomorphisme.

**Proposition :** Si deux espaces  $X$  et  $Y$  sont homotopiquement équivalents, alors ils ont le même nombre de composantes connexes.

**Exemple :** L'intervalle  $[0, 2\pi)$  et le cercle  $\mathbb{S}_1 \subset \mathbb{R}^2$  ne sont pas homéomorphes.

Nous allons le prouver par contradiction. Supposons qu'ils sont homéomorphes. Par définition, cela signifie qu'il existe une application  $f: [0, 2\pi) \rightarrow \mathbb{S}_1$  continue, bijective et d'inverse continue.



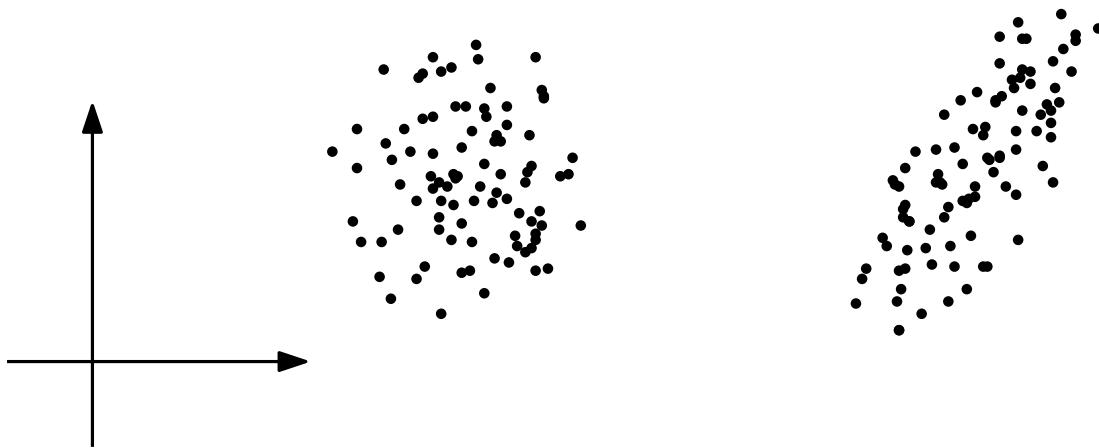
Soit  $x \in [0, 2\pi)$  tel que  $x \neq 0$ . Considérons les sous-ensembles  $[0, 2\pi) \setminus \{x\} \subset [0, 2\pi)$  et  $\mathbb{S}_1 \setminus \{f(x)\} \subset \mathbb{S}_1$ , et l'application induite

$$g: [0, 2\pi) \setminus \{x\} \rightarrow \mathbb{S}_1 \setminus \{f(x)\}.$$

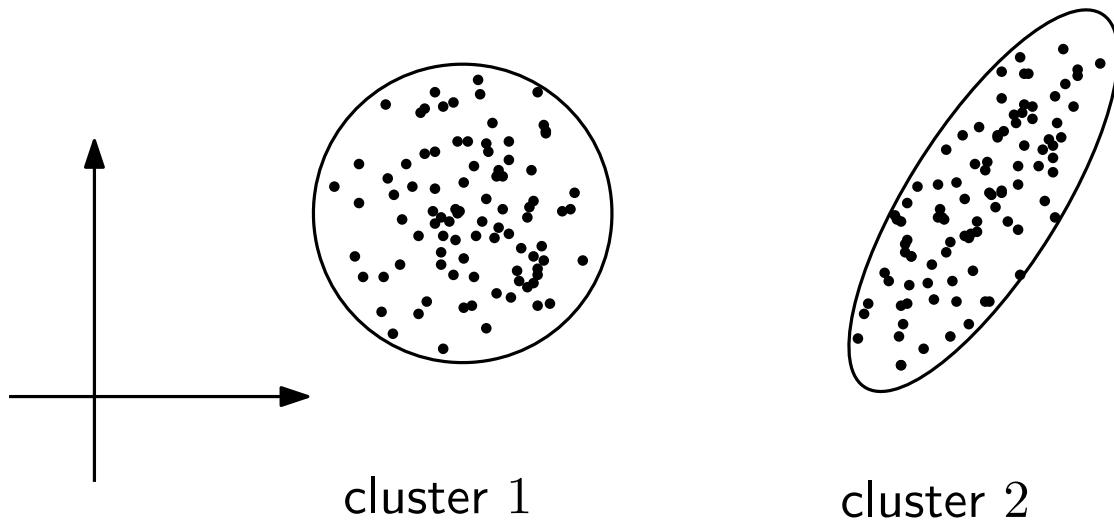
L'application  $g$  est un homéomorphisme.

De plus,  $[0, 2\pi) \setminus \{x\}$  a deux composantes connexes, mais  $\mathbb{S}_1 \setminus \{f(x)\}$  n'en a qu'une. C'est absurde.

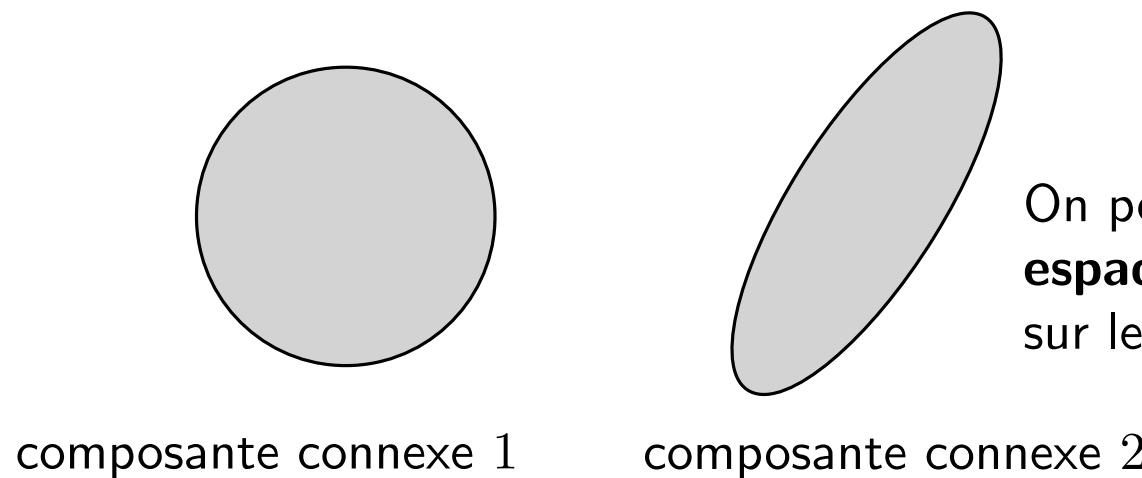
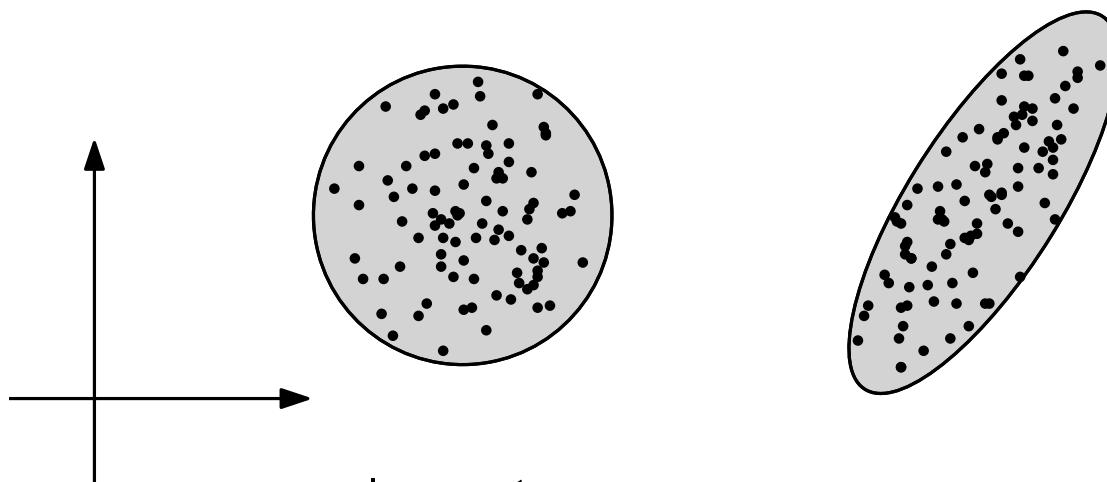
Dans les applications, trouver les composantes connexes correspond à une tâche de **classification**.



Dans les applications, trouver les composantes connexes correspond à une tâche de **classification**.



Dans les applications, trouver les composantes connexes correspond à une tâche de **classification**.



On peut voir ces ensembles comme un **espace topologique sous-jacent**, sur lequel les points ont été tirés.

## I - Comparer les espaces topologiques

- 1 - Équivalence d'homéomorphie
- 2 - Équivalence d'homotopie

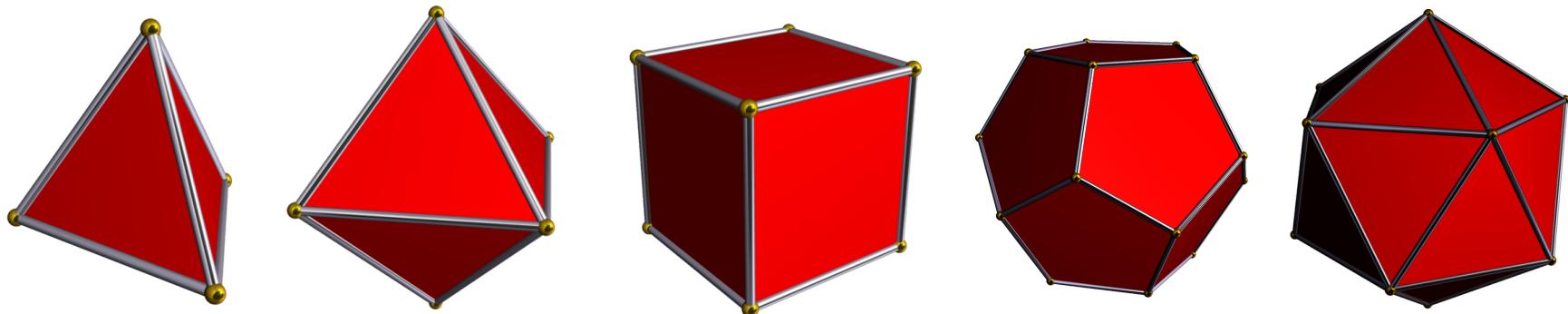
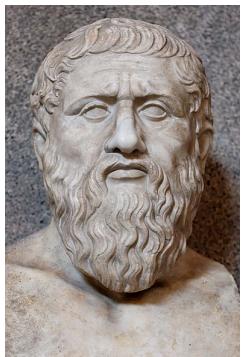
## II - Invariants topologiques

- 1 - Nombre de composantes connexes
- 2 - Caractéristique d'Euler
- 3 - Nombres de Betti

(Prochaine séance - Homologie persistante)

# Caractéristique d'Euler

21/30 (1/2)



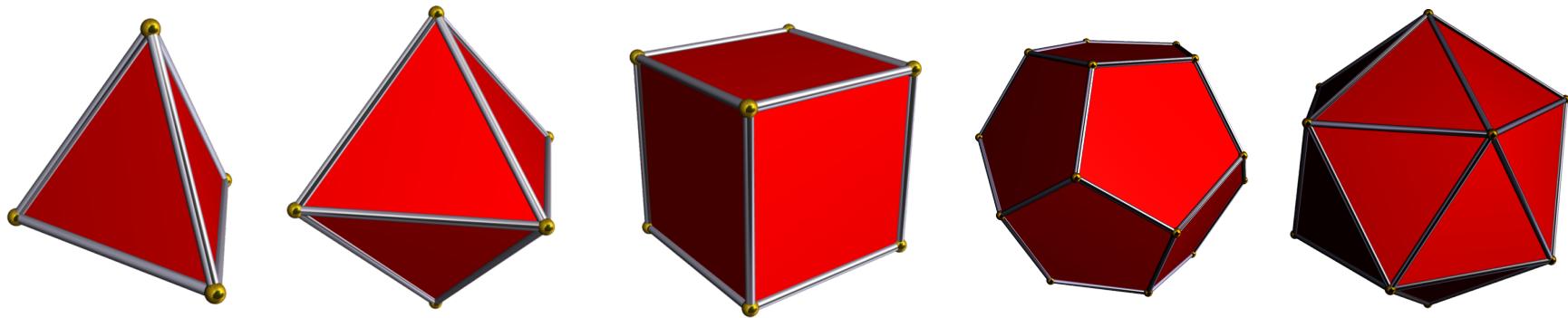
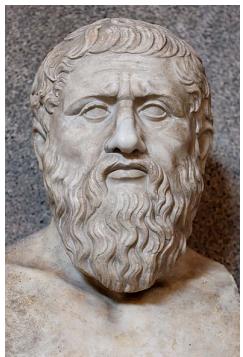
nombre de faces	4	8	6	12	20
-----------------	---	---	---	----	----

nombre d'arêtes	6	12	12	30	30
-----------------	---	----	----	----	----

nombre de sommets	4	6	8	20	12
-------------------	---	---	---	----	----

# Caractéristique d'Euler

21/30 (2/2)



nombre de faces	4	8	6	12	20
nombre d'arêtes	6	12	12	30	30
nombre de sommets	4	6	8	20	12
$\chi$	2	2	2	2	2



**Proposition [Euler, 1758]** : Dans tout polyèdre convexe, on a :  
nombre de faces – nombre d'arêtes + nombre de sommets = 2.

**Définition :** Soit  $V$  un ensemble (appelé l'ensemble des *sommets*). Un **complexe simplicial** sur  $V$  est un ensemble  $K$  de sous-ensembles de  $V$  (appelés les *simplexes*) tel que, pour tout  $\sigma \in K$  et tout  $\tau \subset \sigma$  non-vide, on ait  $\tau \in K$ .

La **dimension** d'un simplexe  $\sigma \in K$  est définie comme égale à  $|\sigma| - 1$ .

# Complexes simpliciaux

22/30 (2/6)

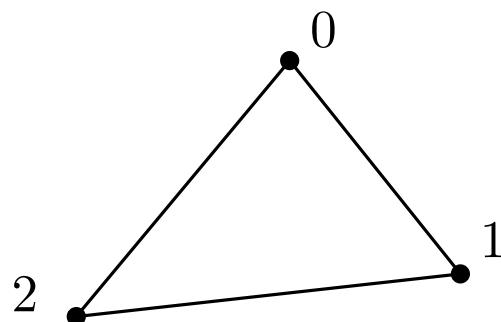
**Définition :** Soit  $V$  un ensemble (appelé l'ensemble des *sommets*). Un **complexe simplicial** sur  $V$  est un ensemble  $K$  de sous-ensembles de  $V$  (appelés les *simplexes*) tel que, pour tout  $\sigma \in K$  et tout  $\tau \subset \sigma$  non-vide, on ait  $\tau \in K$ .

La **dimension** d'un simplexe  $\sigma \in K$  est définie comme égale à  $|\sigma| - 1$ .

**Exemple :** Soit  $V = \{0, 1, 2\}$  et

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

C'est un complexe simplicial.



Il contient trois simplexes de dimension 0 ( $[0]$ ,  $[1]$  et  $[2]$ ) et trois simplexes de dimension 1 ( $[0, 1]$ ,  $[1, 2]$  et  $[0, 2]$ ).

# Complexes simpliciaux

22/30 (3/6)

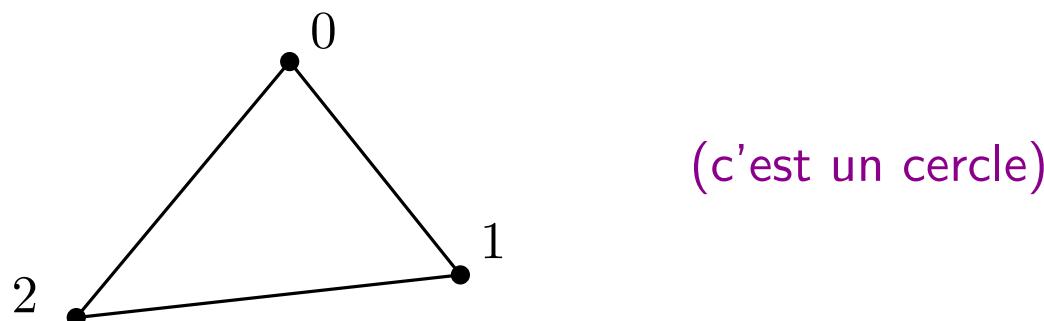
**Définition :** Soit  $V$  un ensemble (appelé l'ensemble des *sommets*). Un **complexe simplicial** sur  $V$  est un ensemble  $K$  de sous-ensembles de  $V$  (appelés les *simplexes*) tel que, pour tout  $\sigma \in K$  et tout  $\tau \subset \sigma$  non-vide, on ait  $\tau \in K$ .

La **dimension** d'un simplexe  $\sigma \in K$  est définie comme égale à  $|\sigma| - 1$ .

**Exemple :** Soit  $V = \{0, 1, 2\}$  et

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

C'est un complexe simplicial.



Il contient trois simplexes de dimension 0 ( $[0]$ ,  $[1]$  et  $[2]$ ) et trois simplexes de dimension 1 ( $[0, 1]$ ,  $[1, 2]$  et  $[0, 2]$ ).

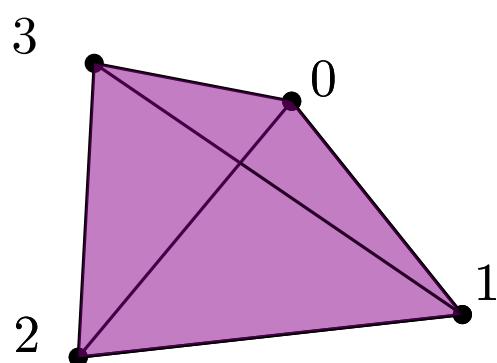
**Définition :** Soit  $V$  un ensemble (appelé l'ensemble des *sommets*). Un **complexe simplicial** sur  $V$  est un ensemble  $K$  de sous-ensembles de  $V$  (appelés les *simplexes*) tel que, pour tout  $\sigma \in K$  et tout  $\tau \subset \sigma$  non-vide, on ait  $\tau \in K$ .

La **dimension** d'un simplexe  $\sigma \in K$  est définie comme égale à  $|\sigma| - 1$ .

**Exemple :** Soit  $V = \{0, 1, 2, 3\}$  et

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

C'est un complexe simplicial.



Il contient quatre simplexes de dimension 0 ( $[0]$ ,  $[1]$ ,  $[2]$  et  $[3]$ ), six simplexes de dimension 1 ( $[0, 1]$ ,  $[1, 2]$ ,  $[2, 3]$ ,  $[3, 0]$ ,  $[0, 2]$  et  $[1, 3]$ ) et quatre simplexes de dimension 2 ( $[0, 1, 2]$ ,  $[0, 1, 3]$ ,  $[0, 2, 3]$  et  $[1, 2, 3]$ ).

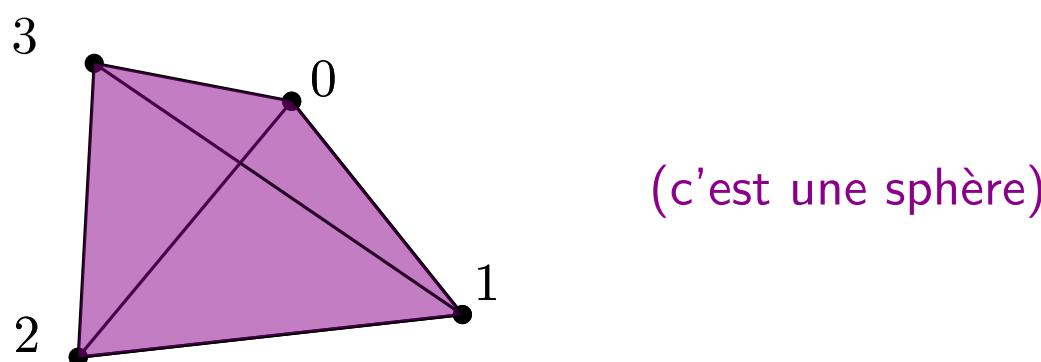
**Définition :** Soit  $V$  un ensemble (appelé l'ensemble des *sommets*). Un **complexe simplicial** sur  $V$  est un ensemble  $K$  de sous-ensembles de  $V$  (appelés les *simplexes*) tel que, pour tout  $\sigma \in K$  et tout  $\tau \subset \sigma$  non-vide, on ait  $\tau \in K$ .

La **dimension** d'un simplexe  $\sigma \in K$  est définie comme égale à  $|\sigma| - 1$ .

**Exemple :** Soit  $V = \{0, 1, 2, 3\}$  et

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

C'est un complexe simplicial.



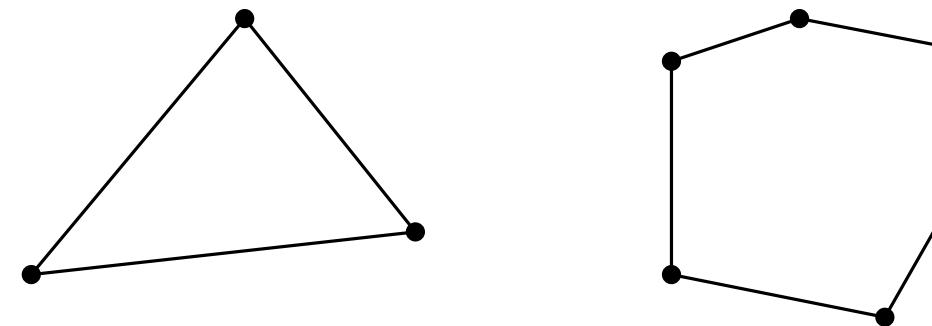
Il contient quatre simplexes de dimension 0 ( $[0]$ ,  $[1]$ ,  $[2]$  et  $[3]$ ), six simplexes de dimension 1 ( $[0, 1]$ ,  $[1, 2]$ ,  $[2, 3]$ ,  $[3, 0]$ ,  $[0, 2]$  et  $[1, 3]$ ) et quatre simplexes de dimension 2 ( $[0, 1, 2]$ ,  $[0, 1, 3]$ ,  $[0, 2, 3]$  et  $[1, 2, 3]$ ).

# Complexes simpliciaux

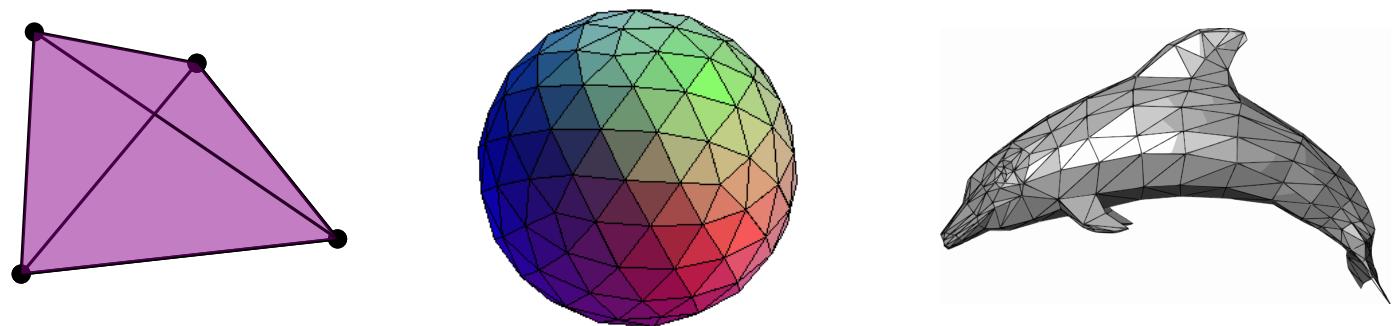
22/30 (6/6)

Les complexes simpliciaux permettent de représenter simplement des espaces topologiques.

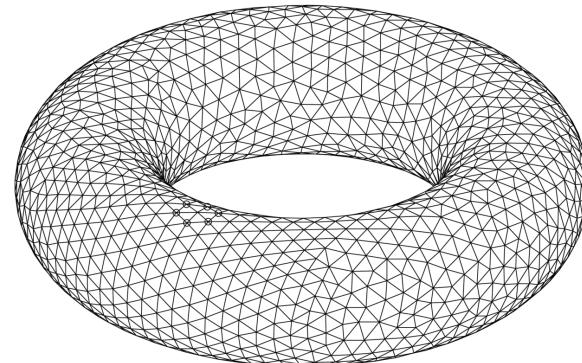
Triangulation du cercle :



Triangulation de la sphère :



Triangulation du tore :



# Caractéristique d'Euler (simpliciale)

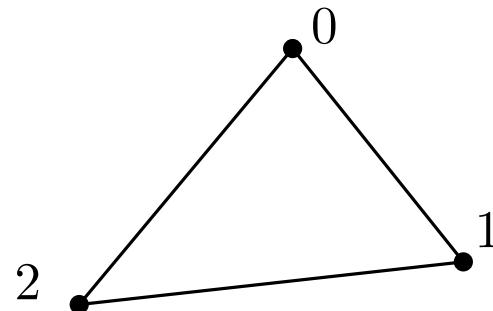
23/30 (1/2)

**Définition** : Soit  $K$  un complexe simplicial de dimension  $n$ . Sa **caractéristique d'Euler** est l'entier

$$\chi(K) = \sum_{0 \leq i \leq n} (-1)^i \cdot (\text{nombre de simplexes de dimension } i).$$

**Exemple** : Le complexe simplicial  $K = \{[0], [1], [2], [0, 1], [1, 2], [2, 0]\}$  a pour caractéristique d'Euler :

$$\chi(K) = 3 - 3 = 0$$



# Caractéristique d'Euler (simpliciale)

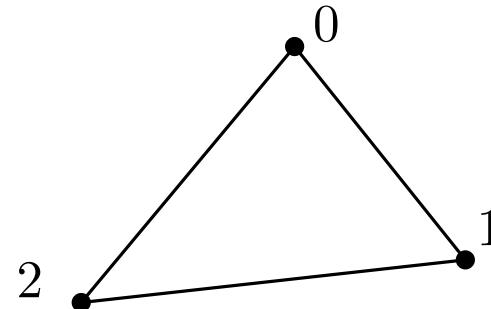
23/30 (2/2)

**Définition :** Soit  $K$  un complexe simplicial de dimension  $n$ . Sa **caractéristique d'Euler** est l'entier

$$\chi(K) = \sum_{0 \leq i \leq n} (-1)^i \cdot (\text{nombre de simplexes de dimension } i).$$

**Exemple :** Le complexe simplicial  $K = \{[0], [1], [2], [0, 1], [1, 2], [2, 0]\}$  a pour caractéristique d'Euler :

$$\chi(K) = 3 - 3 = 0$$

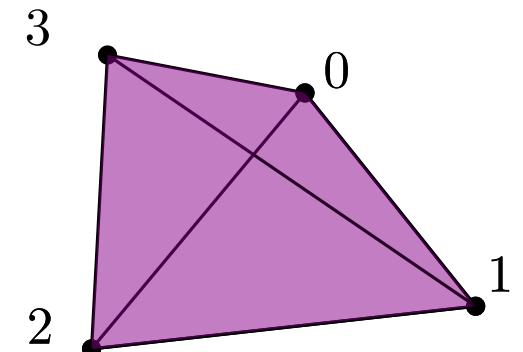


**Exemple :** Le complexe simplicial

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

a pour caractéristique d'Euler :

$$\chi(K) = 4 - 6 + 4 = 2$$



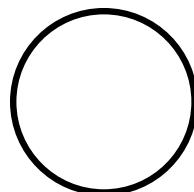
**Proposition :** Si  $X$  et  $Y$  sont des espaces topologiques homotopiquement équivalents, alors  $\chi(X) = \chi(Y)$ .

La caractéristique d'Euler est donc un **invariant** de classes d'équivalence d'homotopie.

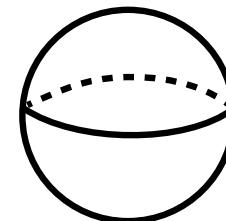
On peut utiliser cette information pour prouver que deux espaces ne sont pas homotopiquement équivalents.

**Exemple :** Le cercle a pour caractéristique d'Euler 0, et la sphère 2.

Donc il ne sont pas homotopiquement équivalents.



$$\chi(\mathbb{S}_1) = 0$$



$$\chi(\mathbb{S}_2) = 2$$

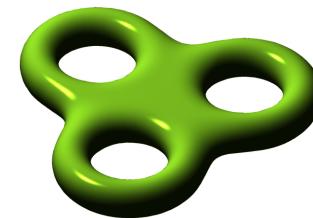
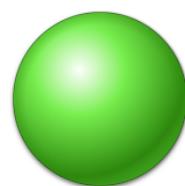
**Proposition :** Si  $X$  et  $Y$  sont des espaces topologiques homotopiquement équivalents, alors  $\chi(X) = \chi(Y)$ .

La caractéristique d'Euler est donc un **invariant** de classes d'équivalence d'homotopie.

On peut utiliser cette information pour prouver que deux espaces ne sont pas homotopiquement équivalents.

**Exemple :** Classification des surfaces.

Les classes d'homéomorphisme des *surfaces compactes et orientées* sont classifiées par leur caractéristique d'Euler.



...

$\chi$

2

0

-2

-4

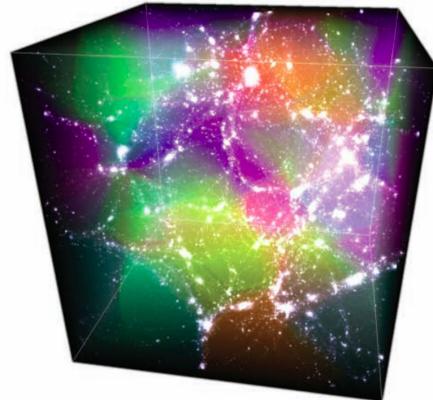
$2 - 2 \times \text{genre}$

# Invariance - dans les applications

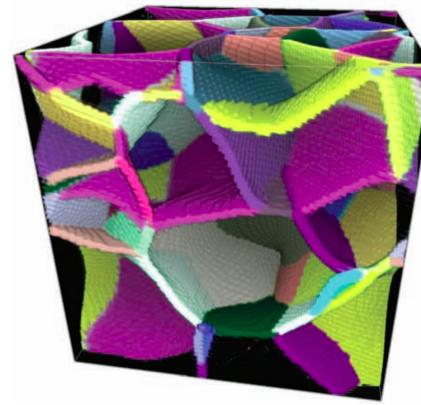
25/30 (1/2)

La caractéristique d'Euler contient des informations à propos de la classe d'homéomorphisme (et d'homotopie) de l'espace considéré.

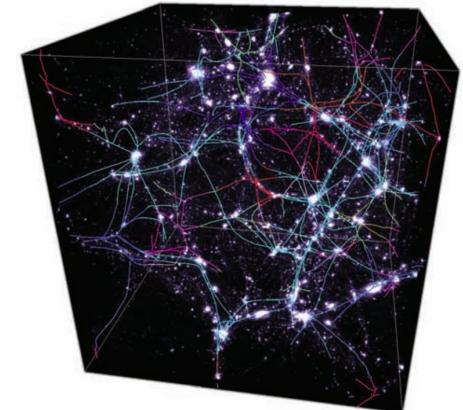
[T. Sousbie, The persistent cosmic web and its filamentary structure, 2011]



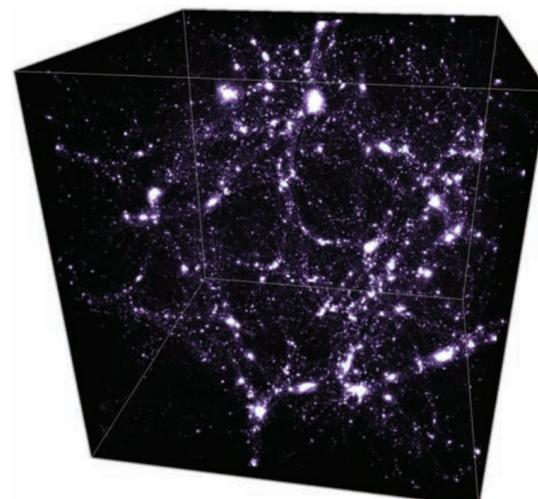
vu comme un objet de dimension 3



de dimension 2



de dimension 1

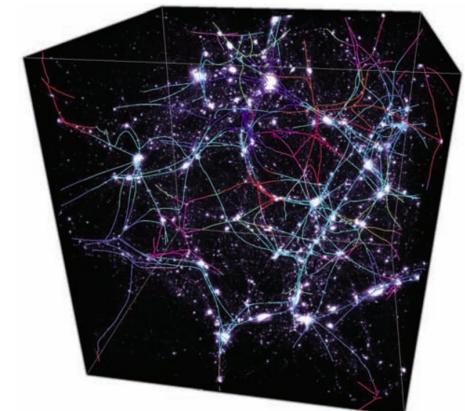
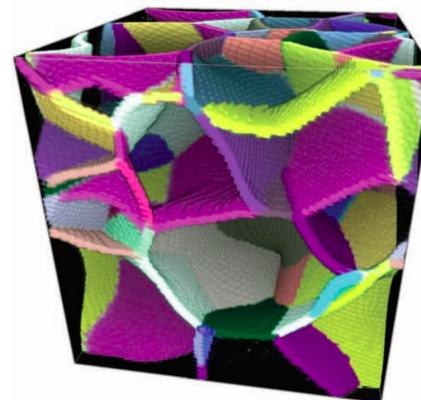
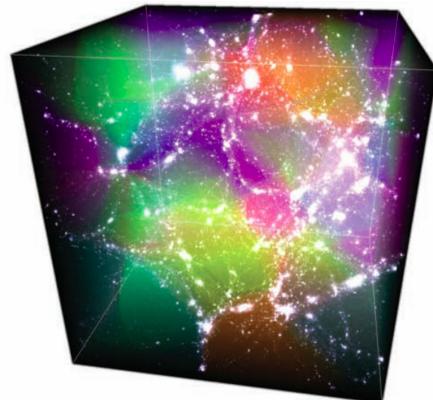


# Invariance - dans les applications

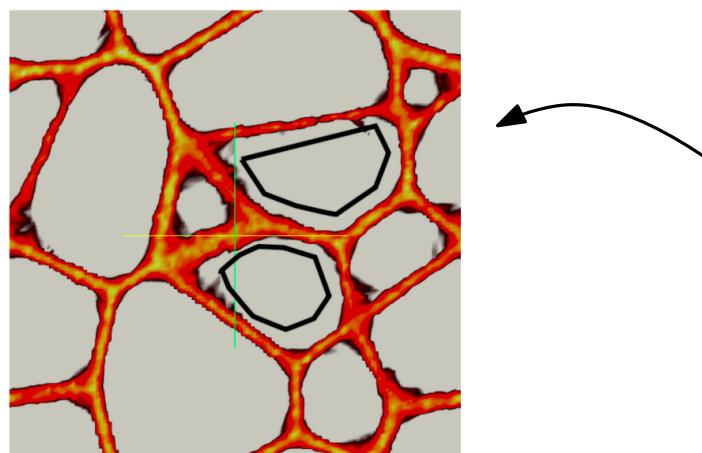
25/30 (2/2)

La caractéristique d'Euler contient des informations à propos de la classe d'homéomorphisme (et d'homotopie) de l'espace considéré.

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



[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]



La caractéristique d'Euler "compte" le nombre de trous

## I - Comparer les espaces topologiques

- 1 - Équivalence d'homéomorphie
- 2 - Équivalence d'homotopie

## II - Invariants topologiques

- 1 - Nombre de composantes connexes
- 2 - Caractéristique d'Euler
- 3 - Nombres de Betti

(Prochaine séance - Homologie persistante)

Pour tout espace topologique  $X$ , on peut définir une suite d'entier

$$\beta_0(X), \quad \beta_1(X), \quad \beta_2(X), \quad \beta_3(X), \quad \dots$$

appelés les **nombres de Betti**.

Construction des nombres de Betti :

Pour tout espace topologique  $X$ , on peut définir une suite d'entier

$$\beta_0(X), \quad \beta_1(X), \quad \beta_2(X), \quad \beta_3(X), \quad \dots$$

appelés les **nombres de Betti**.

Construction des nombres de Betti :

rendez-vous à la prochaine séance !  
(basé sur la théorie de l'homologie)

# Nombres de Betti

27/30 (3/11)

Pour tout espace topologique  $X$ , on peut définir une suite d'entier

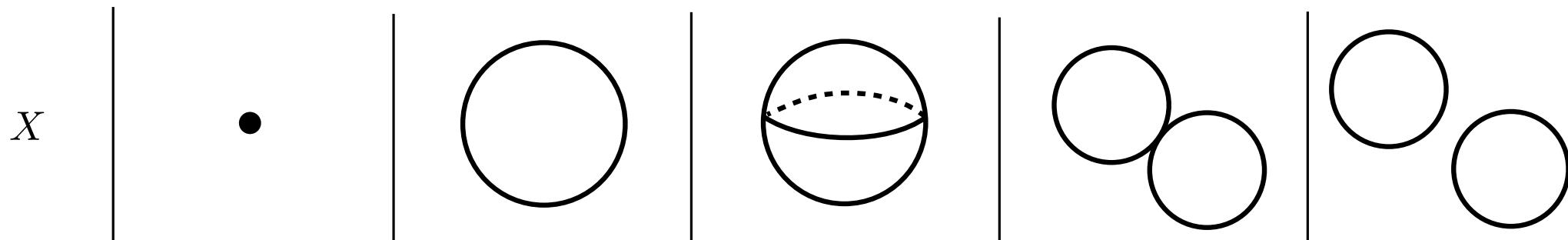
$$\beta_0(X), \quad \beta_1(X), \quad \beta_2(X), \quad \beta_3(X), \quad \dots$$

appelés les **nombres de Betti**.

Construction des nombres de Betti :

rendez-vous à la prochaine séance !  
(basé sur la théorie de l'homologie)

Exemples :

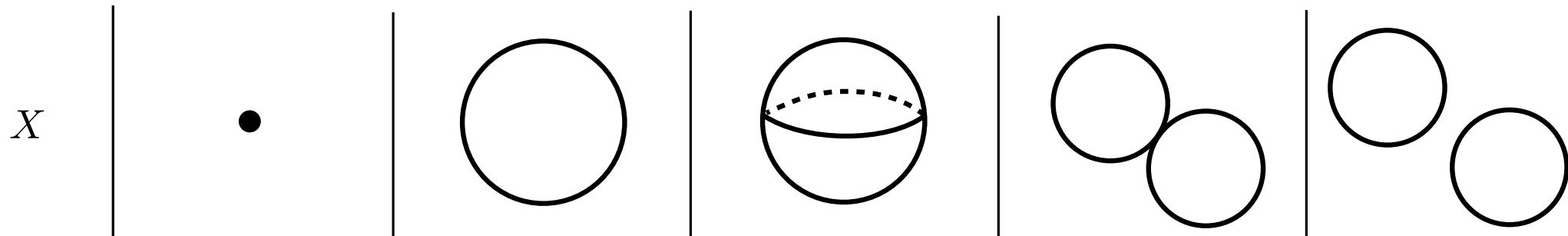


$X$	•	circle	circle with dashed diameter	two overlapping circles	two separate circles
$\beta_0(X)$	1	1	1	1	2
$\beta_1(X)$	0	1	0	2	2
$\beta_2(X)$	0	0	1	0	0

Interprétation :

- $\beta_0(X)$  est le nombre de composantes connexes de  $X$
- $\beta_1(X)$  est le nombre de "trous" dans  $X$
- $\beta_2(X)$  est le nombre de "cavités"  $X$
- ...

Exemples :

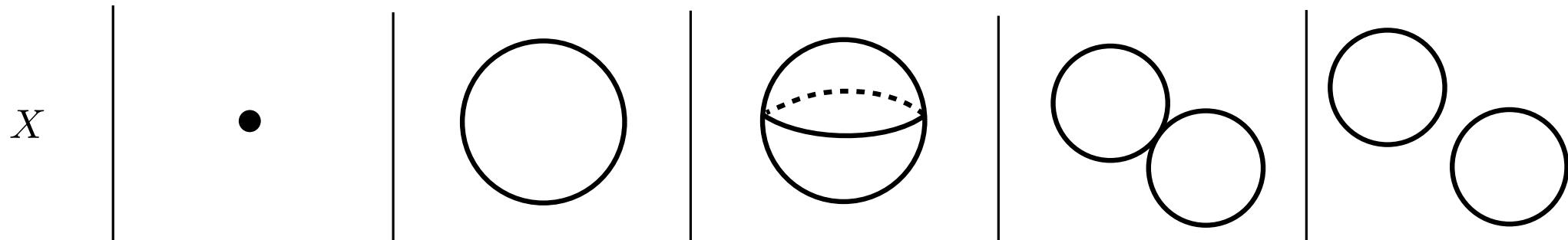


$X$					
$\beta_0(X)$	1	1	1	1	2
$\beta_1(X)$	0	1	0	2	2
$\beta_2(X)$	0	0	1	0	0

Interprétation :

- $\beta_0(X)$  est le nombre de composantes connexes de  $X$
- $\beta_1(X)$  est le nombre de "trous" dans  $X$
- $\beta_2(X)$  est le nombre de "cavités"  $X$
- ...

Exemples :

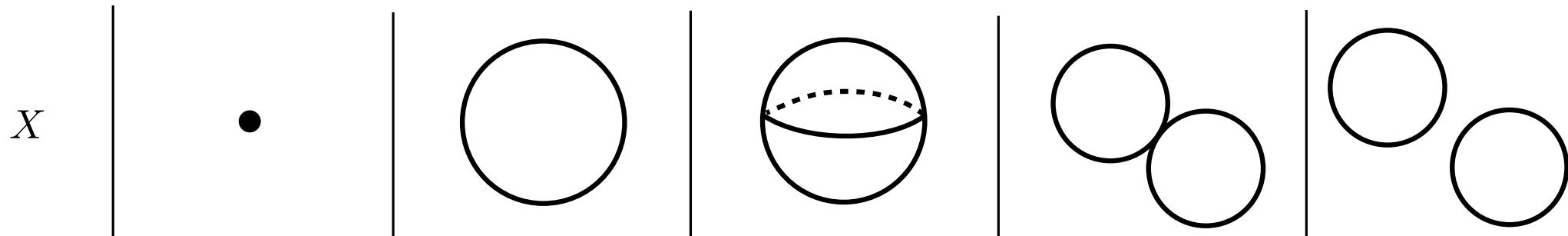


$\beta_0(X)$	1	1	1	1	2
$\beta_1(X)$	0	1	0	2	2
$\beta_2(X)$	0	0	1	0	0

Interprétation :

- $\beta_0(X)$  est le nombre de composantes connexes de  $X$
- $\beta_1(X)$  est le nombre de "trous" dans  $X$
- $\beta_2(X)$  est le nombre de "cavités"  $X$
- ...

Exemples :



$\beta_0(X)$	1	1	1	1	2
$\beta_1(X)$	0	1	0	2	2
$\beta_2(X)$	0	0	1	0	0

Interprétation :

- $\beta_0(X)$  est le nombre de composantes connexes de  $X$
- $\beta_1(X)$  est le nombre de "trous" dans  $X$
- $\beta_2(X)$  est le nombre de "cavités"  $X$
- ...

Exemples :

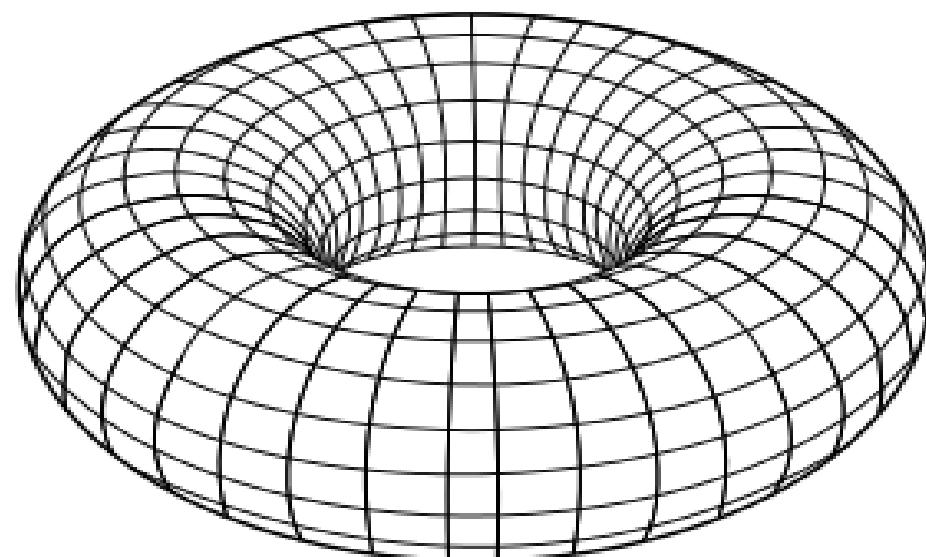
$X$					
$\beta_0(X)$	1	1	1	1	2
$\beta_1(X)$	0	1	0	2	2
$\beta_2(X)$	0	0	1	0	0

Interprétation :

- $\beta_0(X)$  est le nombre de composantes connexes de  $X$
- $\beta_1(X)$  est le nombre de "trous" dans  $X$
- $\beta_2(X)$  est le nombre de "cavités"  $X$
- ...

Exemple: Nombres de Betti du tore :

$$\beta_0(X) = 1, \quad \beta_1(X) = 2, \quad \beta_2(X) = 1, \quad \beta_3(X) = 0, \quad \dots$$

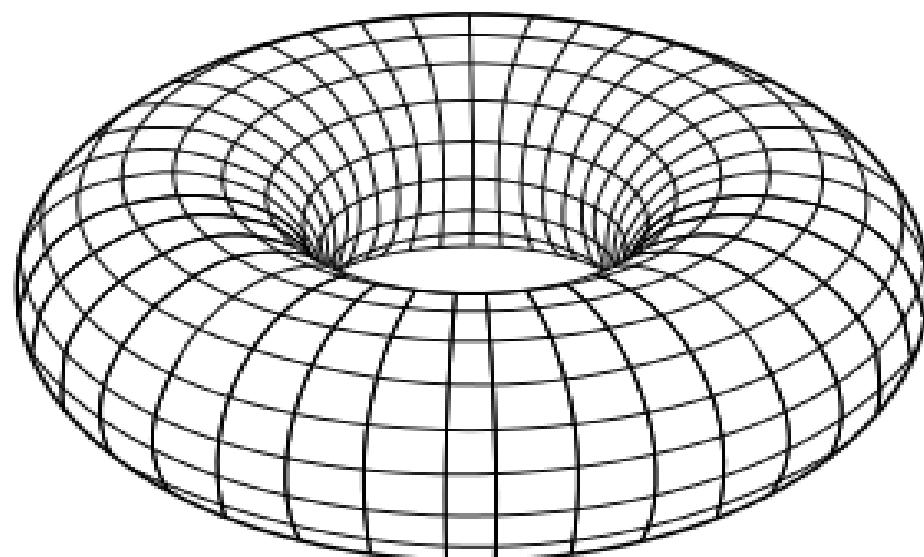


Interprétation :

- $\beta_0(X)$  est le nombre de composantes connexes de  $X$
- $\beta_1(X)$  est le nombre de "trous" dans  $X$
- $\beta_2(X)$  est le nombre de "cavités"  $X$
- ...

Exemple: Nombres de Betti du tore :

$$\beta_0(X) = 1, \quad \beta_1(X) = 2, \quad \beta_2(X) = 1, \quad \beta_3(X) = 0, \quad \dots$$

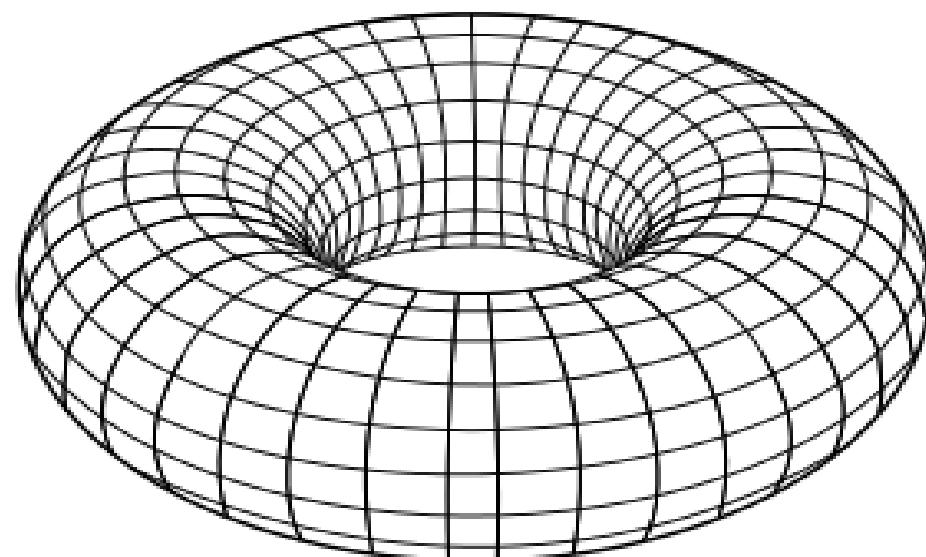


Interprétation :

- $\beta_0(X)$  est le nombre de composantes connexes de  $X$
- $\beta_1(X)$  est le nombre de "trous" dans  $X$
- $\beta_2(X)$  est le nombre de "cavités"  $X$
- ...

Exemple: Nombres de Betti du tore :

$$\beta_0(X) = 1, \quad \beta_1(X) = 2, \quad \boxed{\beta_2(X) = 1}, \quad \beta_3(X) = 0, \quad \dots$$

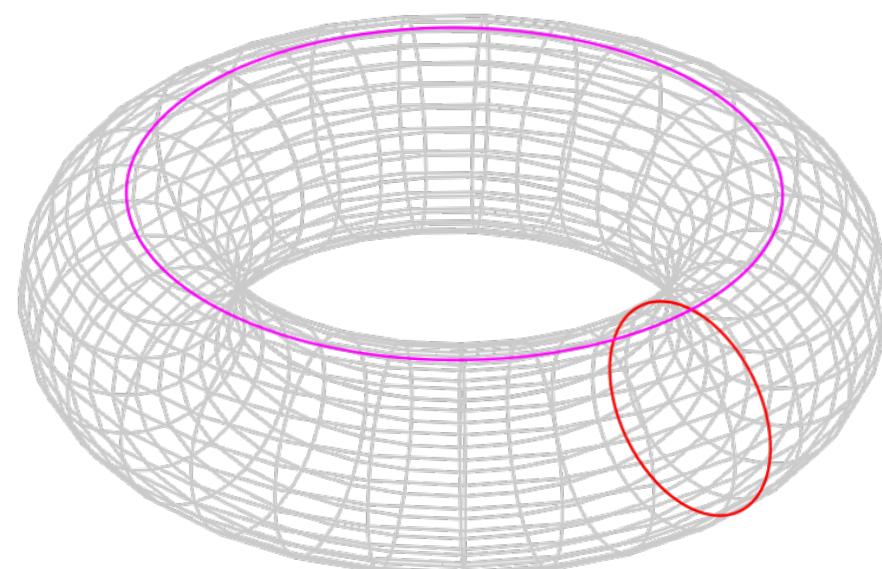


Interprétation :

- $\beta_0(X)$  est le nombre de composantes connexes de  $X$
- $\beta_1(X)$  est le nombre de "trous" dans  $X$
- $\beta_2(X)$  est le nombre de "cavités"  $X$
- ...

Exemple: Nombres de Betti du tore :

$$\beta_0(X) = 1, \quad \boxed{\beta_1(X) = 2,} \quad \beta_2(X) = 1, \quad \beta_3(X) = 0, \quad \dots$$



**Proposition :** Si deux espaces topologiques  $X$  et  $Y$  sont homotopiquement équivalents, alors ils ont les même nombres de Betti.

Ainsi, deux espaces avec les même nombres de Betti ne peuvent pas être homotopiquement équivalents.

**Exemple :** La sphère de dimension  $n$ ,  $\mathbb{S}_n \subset \mathbb{R}^{n+1}$ , a pour nombres de Betti

$$\begin{aligned}\beta_i(X) &= 1 \quad \text{si } i = 0 \text{ ou } n, \\ \beta_i(X) &= 0 \quad \text{sinon.}\end{aligned}$$

Par suite, si  $n \neq m$ , alors  $\mathbb{S}_n$  et  $\mathbb{S}_m$  ne sont pas homotopiquement équivalents.

**Proposition :** Si deux espaces topologiques  $X$  et  $Y$  sont homotopiquement équivalents, alors ils ont les même nombres de Betti.

Ainsi, deux espaces avec les même nombres de Betti ne peuvent pas être homotopiquement équivalents.

**Exemple :** Théorème de l'invariance du domaine de Brouwer.

Montrons que  $\mathbb{R}^n$  et  $\mathbb{R}^m$ , avec  $n \neq m$ , ne sont pas homéomorphes.

Soit  $h: \mathbb{R}^n \rightarrow \mathbb{R}^m$  un homéomorphisme.

Soit  $x \in \mathbb{R}^n$ , et considérons l'application induite

$$h: \mathbb{R}^n \setminus \{x\} \longrightarrow \mathbb{R}^m \setminus \{h(x)\}$$

C'est encore un homéomorphisme.

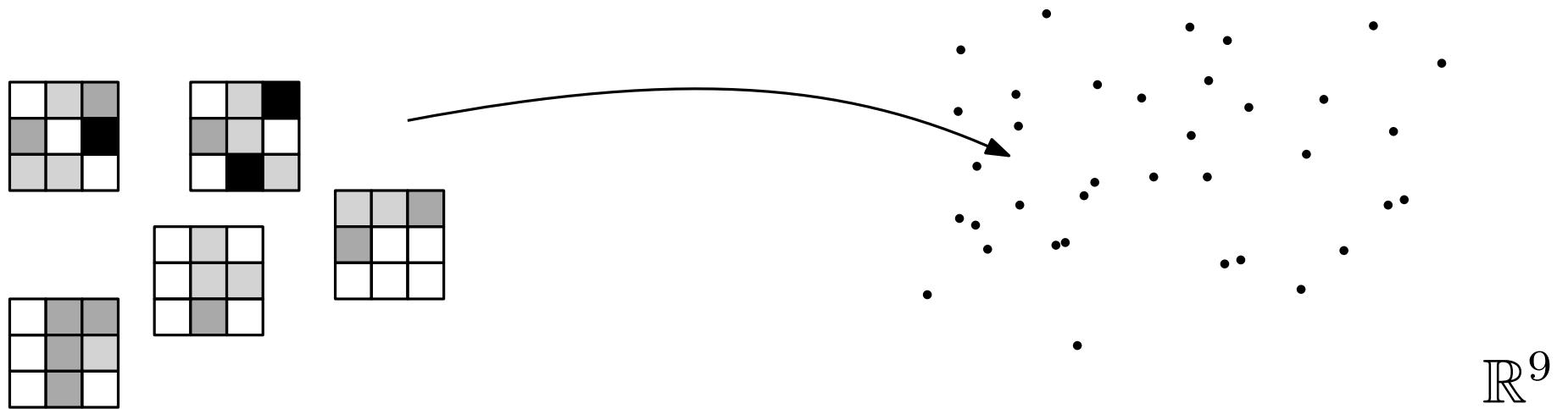
Mais  $\mathbb{R}^n \setminus \{x\}$  est homotopiquement équivalent à la sphère  $\mathbb{S}_{n-1}$ , et  $\mathbb{R}^m \setminus \{x\}$  à la sphère  $\mathbb{S}_{m-1}$

Nous avons vu que  $\mathbb{S}_{n-1}$  et  $\mathbb{S}_{m-1}$  ne sont homotopiquement équivalentes que si  $m = n$ . Nous avons une contradiction.

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

À partir d'une grande collection d'images, les auteurs extraient des imagettes de taille  $3 \times 3$ .

Chacune de ces imagettes contient 9 pixels, et peut donc être vue comme un vecteur dans  $\mathbb{R}^9$ . L'ensemble des imagettes peut être vu comme un nuage de point dans  $\mathbb{R}^9$ .



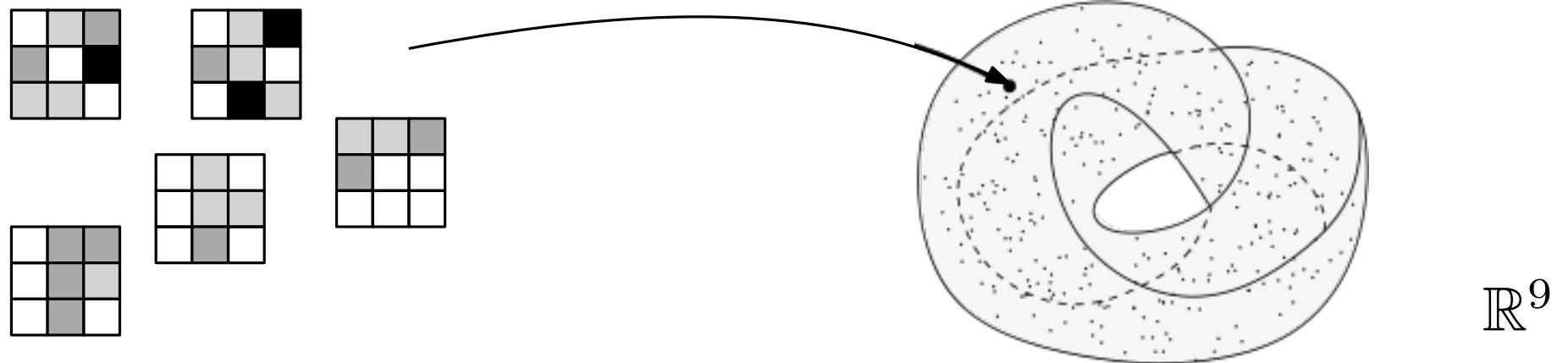
Ils observent que le nuage de point est proche d'une forme dont les nombres de Betti sont :

$$\beta_0(X) = 1, \quad \beta_1(X) = 2, \quad \beta_2(X) = 1, \quad \beta_3(X) = 0$$

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

À partir d'une grande collection d'images, les auteurs extraient des imagettes de taille  $3 \times 3$ .

Chacune de ces imagettes contient 9 pixels, et peut donc être vue comme un vecteur dans  $\mathbb{R}^9$ . L'ensemble des imagettes peut être vu comme un nuage de point dans  $\mathbb{R}^9$ .



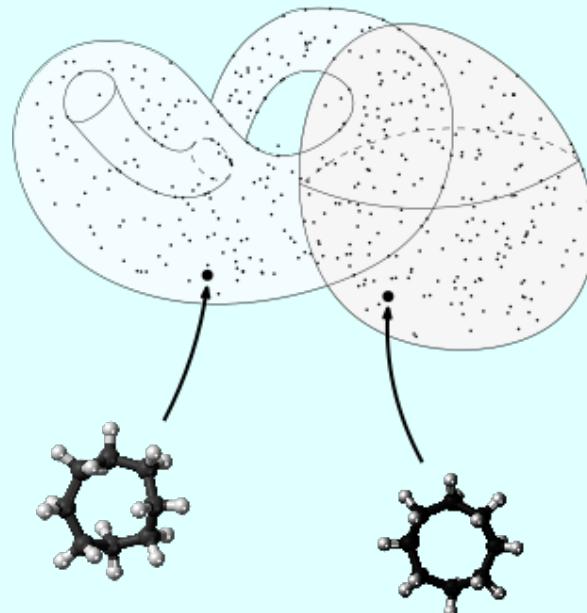
Ils observent que le nuage de point est proche d'une forme dont les nombres de Betti sont :

$$\beta_0(X) = 1, \quad \beta_1(X) = 2, \quad \beta_2(X) = 1, \quad \beta_3(X) = 0$$

Ce sont les nombres de Betti de la bouteille de Klein !

# Conclusion

On peut espérer trouver de la topologie intéressante dans les jeux de données.



Les **invariants** permettent de les décrire et les comprendre.

Nombre de composantes connexes = 1

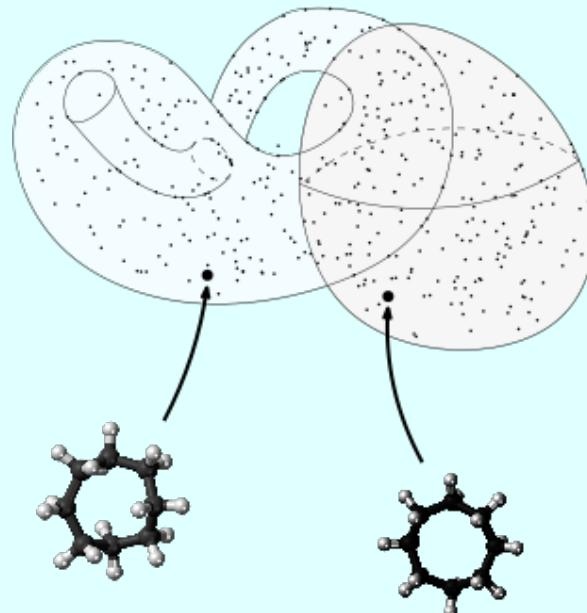
$$\chi = 2$$

$$\beta_0(X) = 1, \quad \beta_1(X) = 1, \quad \beta_2(X) = 2$$

Prochaine semaine : comment calculer ces invariants en pratique ?

# Conclusion

On peut espérer trouver de la topologie intéressante dans les jeux de données.



Les **invariants** permettent de les décrire et les comprendre.

Nombre de composantes connexes = 1

$$\chi = 2$$

$$\beta_0(X) = 1, \quad \beta_1(X) = 1, \quad \beta_2(X) = 2$$

Prochaine semaine : comment calculer ces invariants en pratique ?

Merci !