USP — 24/11/22

Análise Topológica de Dados e suas aplicações

https://raphaeltinarrage.github.io

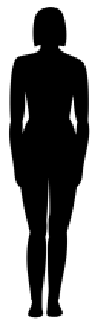
Last update: November 24, 2022

Geometria em dados

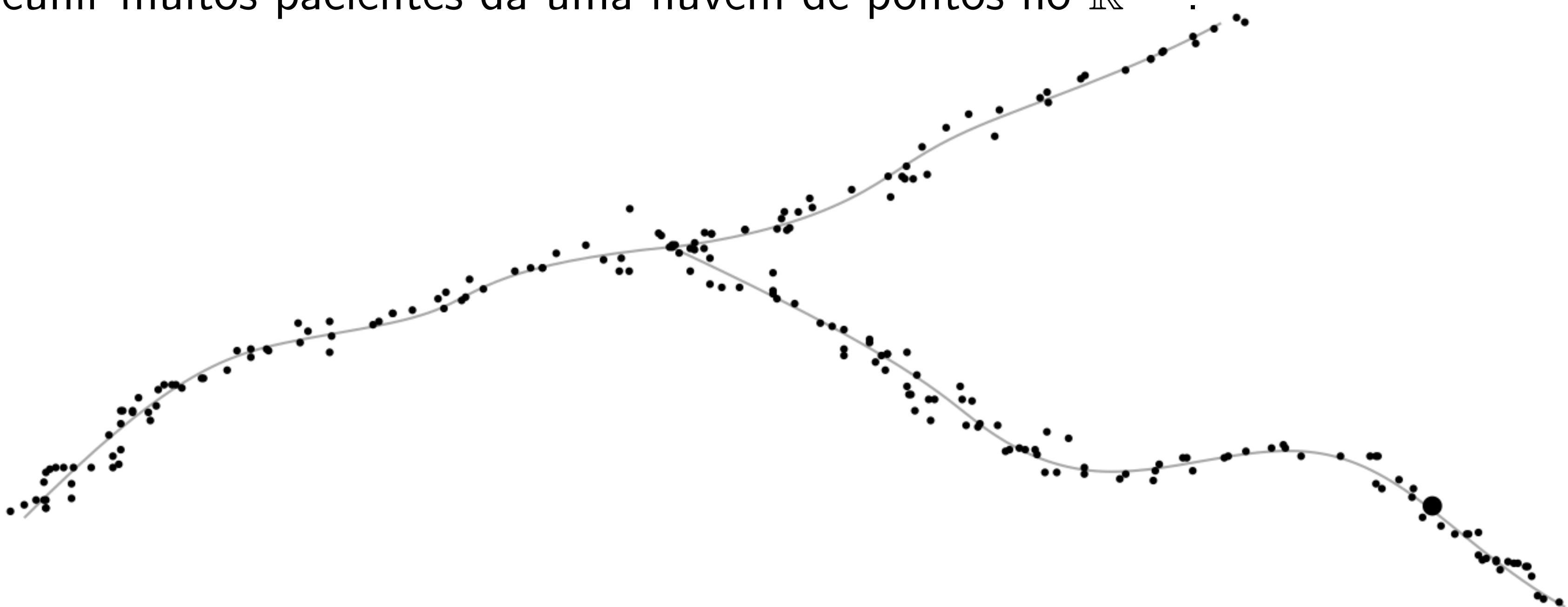
[**Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival**, Monica Nicolau, Arnold J Levine, and Gunnar Carlsson, *Proceedings of the National Academy of Sciences*, 2011]

Os autores estudam os tecidos de pacientes infectados com câncer de mama. Eles obtêm 262 variáveis genômicas por paciente.

$$(x_1, x_2, ..., x_{262}) \longleftarrow$$



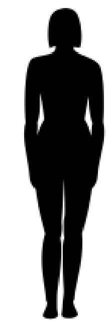
Reunir muitos pacientes dá uma nuvem de pontos no $\mathbb{R}^{262}$.



[**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]

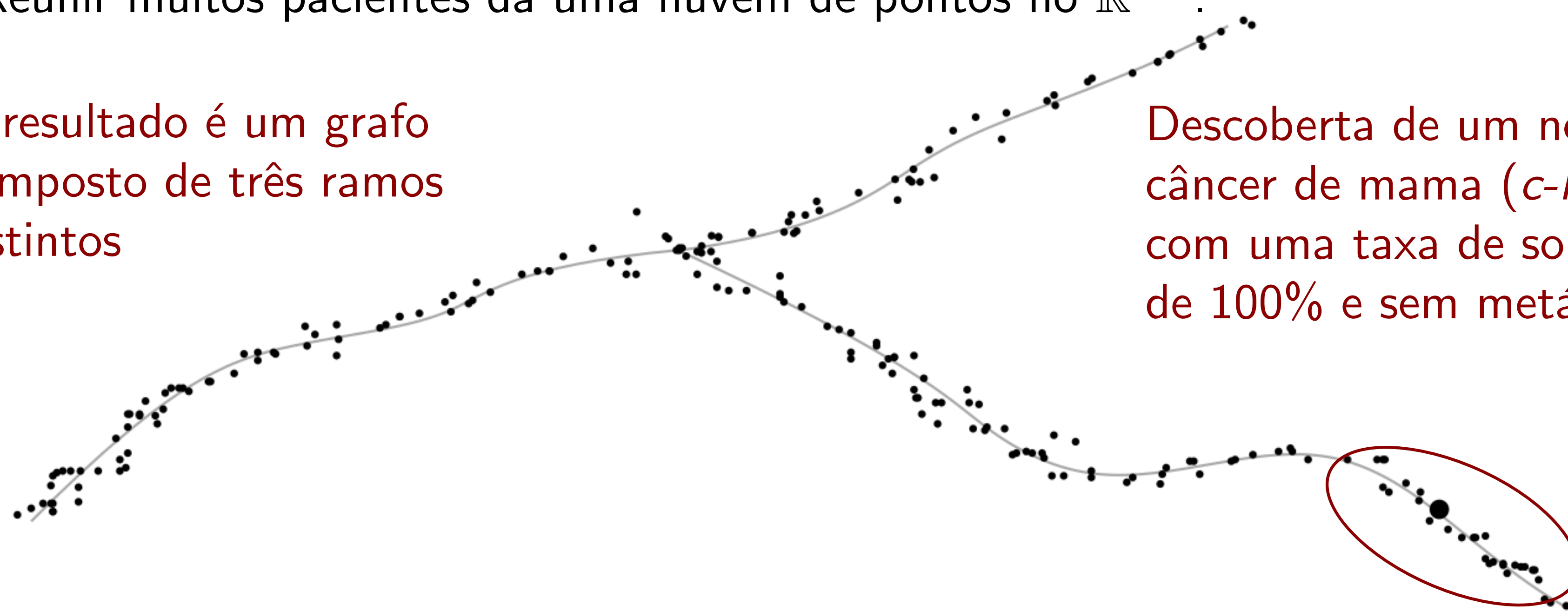
Os autores estudam os tecidos de pacientes infectados com câncer de mama. Eles obtêm 262 variáveis genômicas por paciente.

$$(x_1, x_2, ..., x_{262}) \longleftarrow$$



Reunir muitos pacientes dá uma nuvem de pontos no $\mathbb{R}^{262}$.

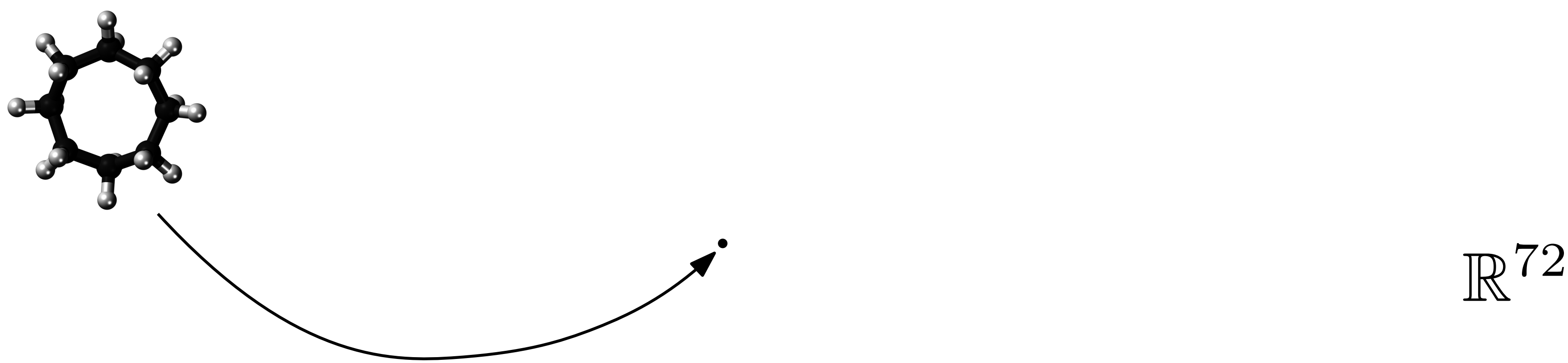
O resultado é um grafo composto de três ramos distintos



Descoberta de um novo tipo de câncer de mama (*c-MYB*$^+$) com uma taxa de sobrevivência de 100% e sem metástases

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

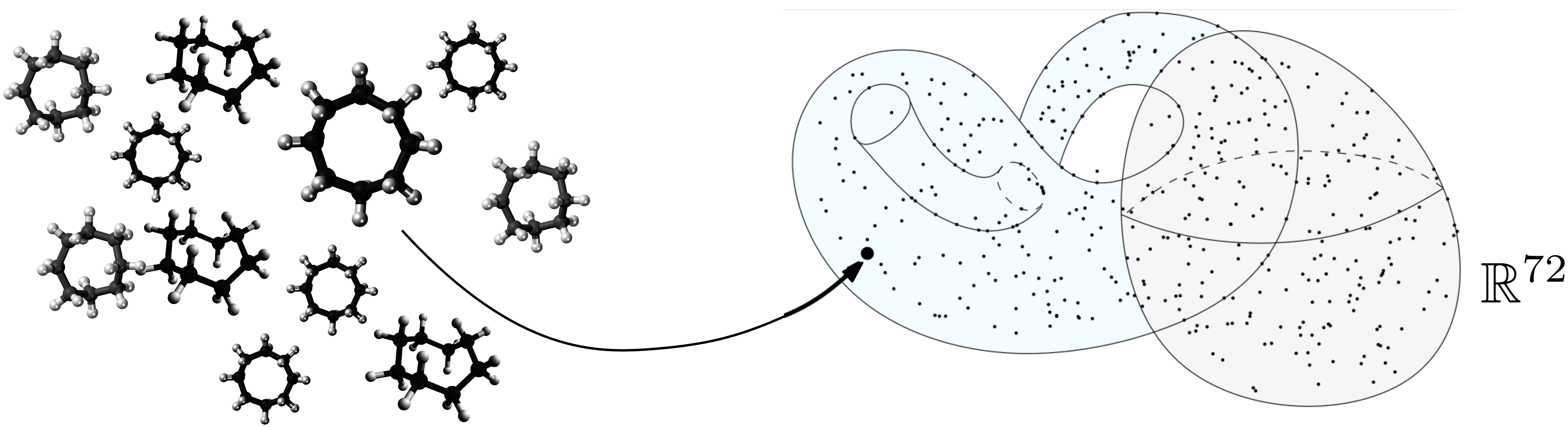
A molécula de ciclo-octano $\mathrm{C_8H_{16}}$ contém 24 átomos.
Cada átomos tem 3 coordenadas espaciais.
Assim, a conformação de uma molécula pode ser resumida por **um ponto** no $\mathbb{R}^{72}$ $(3 \times 24 = 72)$.



Geometria em dados

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

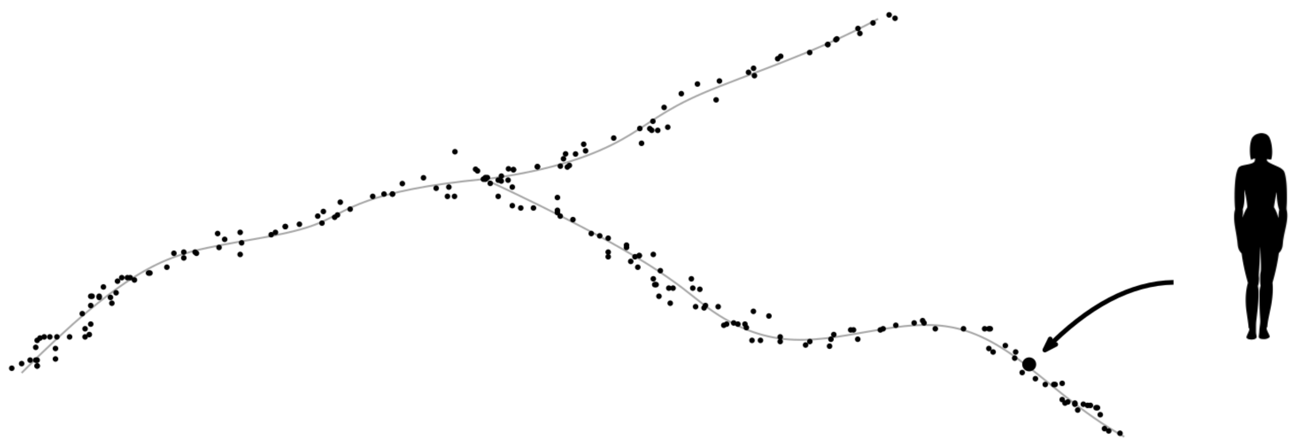
A molécula de ciclo-octano $\mathrm{C_8H_{16}}$ contém 24 átomos.
Cada átomos tem 3 coordenadas espaciais.
Assim, a conformação de uma molécula pode ser resumida por **um ponto** no $\mathbb{R}^{72}$ $(3 \times 24 = 72)$.

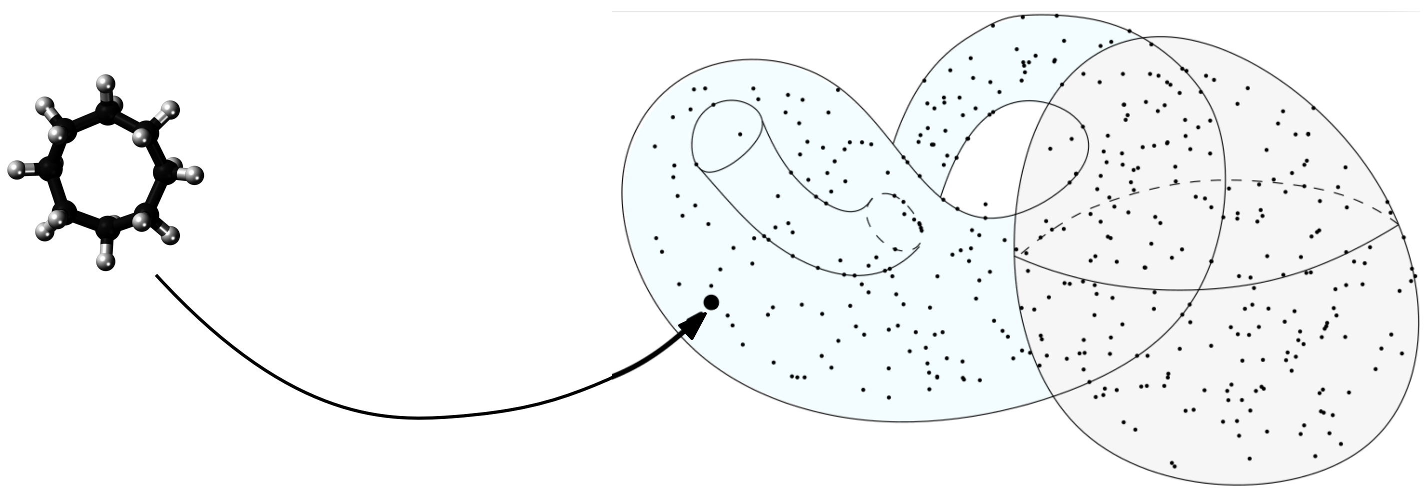


Analisando muitas dessas moléculas, obtemos uma **nuvem de pontos** no $\mathbb{R}^{72}$.

Os autores mostram que esta nuvem de pontos está próxima a um objeto de pequena dimensão: **a união de uma esfera e de uma garrafa de Klein**.

Geometria em dados





Pergunta: Qual é a forma dos dados?

Protagonistas em topologia

Em topologia, estudamos **espaços topológicos**.

Definição: um espaço topológico é um conjunto X dotado de uma coleção de **conjuntos abertos** $\{O_\alpha, \alpha \in A\}$, com $O_\alpha \subset X$, tal que

- $\emptyset$ e X são conjuntos abertos,
- uma união infinita de conjuntos abertos é um conjunto aberto,
- uma interseção finita de conjuntos abertos é um conjunto aberto.

Definição: Dados dois espaços topológicos X e Y, uma função $f\colon X \to Y$ é **contínua** se para cada conjunto aberto $O \subset Y$, ta pré-imagem $f^{-1}(O)$ é um conjunto aberto de X.

$$X \longrightarrow Y$$

Protagonistas em topologia

Em topologia, estudamos **espaços topológicos**.

Definição: um espaço topológico é um conjunto X dotado de uma coleção de **conjuntos abertos** $\{O_\alpha, \alpha \in A\}$, com $O_\alpha \subset X$, tal que

- $\emptyset$ e X são conjuntos abertos,
- uma união infinita de conjuntos abertos é um conjunto aberto,
- uma interseção finita de conjuntos abertos é um conjunto aberto.

Definição: Dados dois espaços topológicos X e Y, uma função $f\colon X \to Y$ é **contínua** se para cada conjunto aberto $O \subset Y$, ta pré-imagem $f^{-1}(O)$ é um conjunto aberto de X.

$$X \longrightarrow Y$$

tradução em
cálculo ϵ-δ

Podemos pensar em **subconjunto** $X \subset \mathbb{R}^n$ e $Y \subset \mathbb{R}^m$,

e funções $f\colon X \to Y$ **contínuas** no seguinte sentido:

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

Exemplos de espaços topológicos

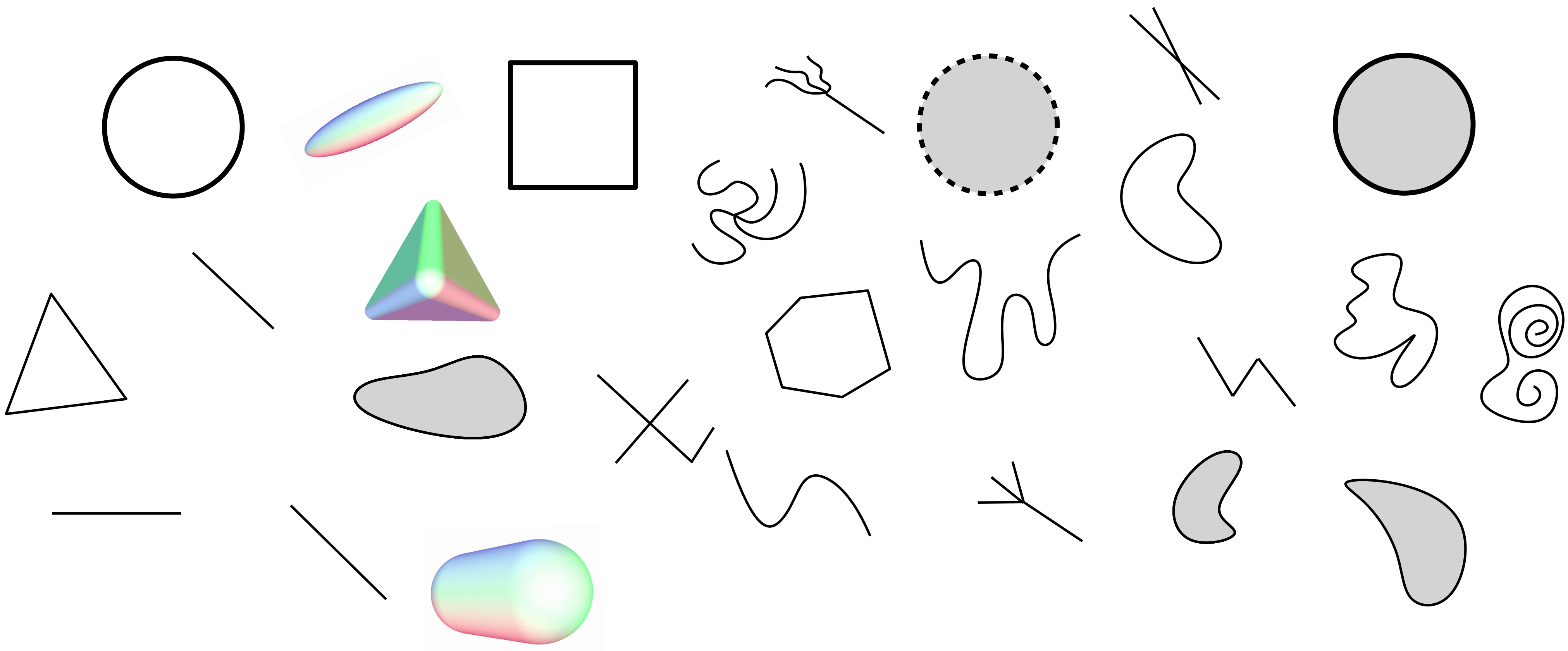
No $\mathbb{R}^n$, podemos definir:

- a esfera de raio unitário $\mathbb{S}_{n-1} = \{x \in \mathbb{R}^n, \|x\| = 1\}$
- o cubo $\mathcal{C}_{n-1} = \{x = (x_1, ..., x_n) \in \mathbb{R}^n, \max(|x_1|, ..., |x_n|) = 1\}$
- as bolas abertas $\mathcal{B}(x, r) = \{y \in \mathbb{R}^n, \|x - y\| < r\}$
- as bolas fechadas $\overline{\mathcal{B}}(x, r) = \{y \in \mathbb{R}^n, \|x - y\| \leq r\}$

Exemplos de espaços topológicos

No $\mathbb{R}^n$, podemos definir:

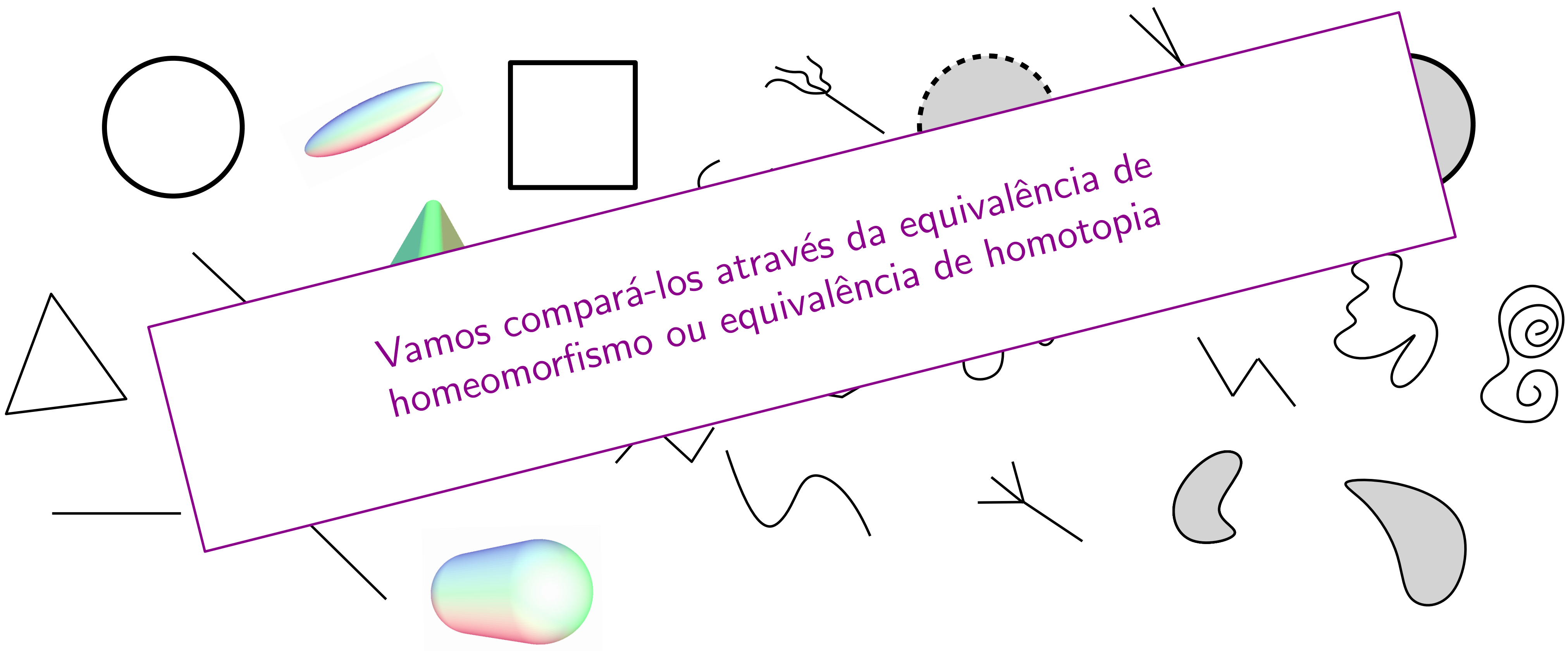
- a esfera de raio unitário $\mathbb{S}_{n-1} = \{x \in \mathbb{R}^n, \|x\| = 1\}$
- o cubo $\mathcal{C}_{n-1} = \{x = (x_1, ..., x_n) \in \mathbb{R}^n, \max(|x_1|, ..., |x_n|) = 1\}$
- as bolas abertas $\mathcal{B}(x, r) = \{y \in \mathbb{R}^n, \|x - y\| < r\}$
- as bolas fechadas $\overline{\mathcal{B}}(x, r) = \{y \in \mathbb{R}^n, \|x - y\| \leq r\}$



Geralmente, não temos uma linda definição algébrica...

No $\mathbb{R}^n$, podemos definir:

- a esfera de raio unitário $\mathbb{S}_{n-1} = \{x \in \mathbb{R}^n, \|x\| = 1\}$
- o cubo $\mathcal{C}_{n-1} = \{x = (x_1, ..., x_n) \in \mathbb{R}^n, \max(|x_1|, ..., |x_n|) = 1\}$
- as bolas abertas $\mathcal{B}(x, r) = \{y \in \mathbb{R}^n, \|x - y\| < r\}$
- as bolas fechadas $\overline{\mathcal{B}}(x, r) = \{y \in \mathbb{R}^n, \|x - y\| \leq r\}$



Vamos compará-los através da equivalência de homeomorfismo ou equivalência de homotopia

Geralmente, não temos uma linda definição algébrica...

I - Comparar espaços topológicos

1 - Equivalência de homeomorfia

2 - Equivalência de homotopia

II - Invariantes topológicos

1 - Número de componentes conexas

2 - Característica de Euler

3 - Números de Betti

4 - Homologia (simplicial sobre $\mathbb{Z}/2\mathbb{Z}$)

III - Inferência topológica

1 - Estimação do parâmetro t

2 - Nervos

IV - Homologia persistente

1 - Módulos de persistência

2 - Decomposição

3 - Estabilidade

V - Aplicações

Homeomorfismos

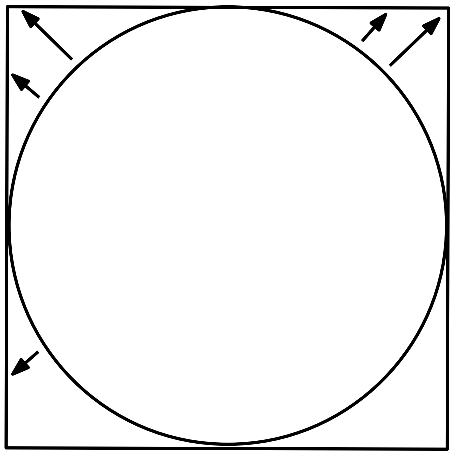
Definition: Sejam X e Y dois espaços topológicos, e $f\colon X \to Y$ uma função. Dizemos que f é um **homeomorfismo** se

- f é uma bijeção,
- $f\colon X \to Y$ é contínua,
- $f^{-1}\colon Y \to X$ é contínua.

Se existe tal homeomorfismo, dizemos que os dois espaços topológicos são **homeomórficos**.

Exemplo: O círculo e o quadrado são homeomórficos via

$$\begin{aligned} f\colon \mathbb{S}_1 &\longrightarrow \mathcal{C} \\ (x_1, x_2) &\longmapsto \frac{1}{\max(|x_1|, |x_2|)}(x_1, x_2) \end{aligned}$$



Interpretação: Os homeomorfismos permitem 'deformações contínuas'.

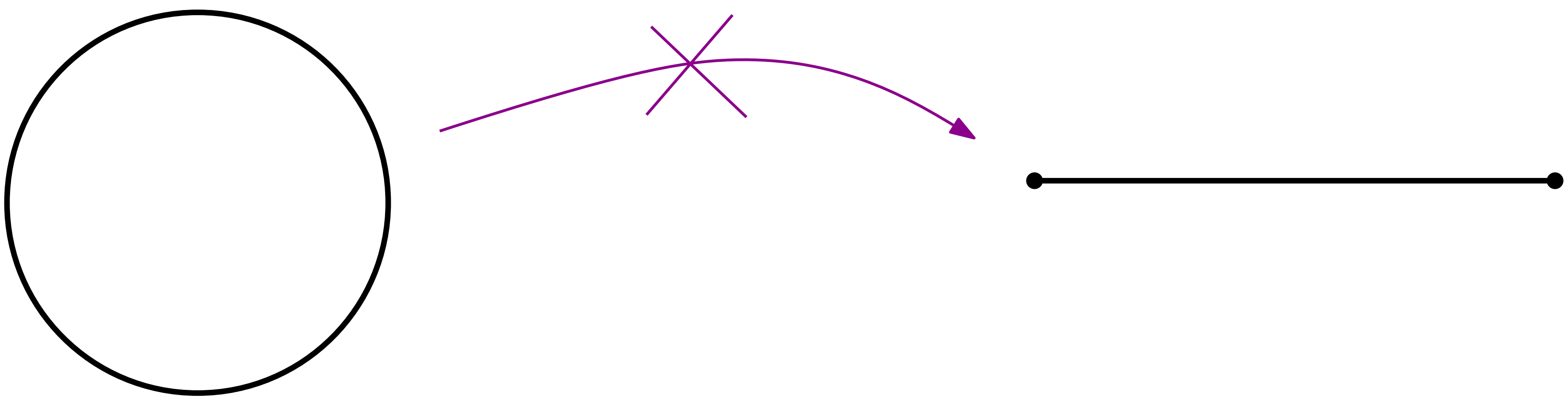
Homeomorfismos

Definition: Sejam X e Y dois espaços topológicos, e $f\colon X \to Y$ uma função. Dizemos que f é um **homeomorfismo** se

- f é uma bijeção,
- $f\colon X \to Y$ é contínua,
- $f^{-1}\colon Y \to X$ é contínua.

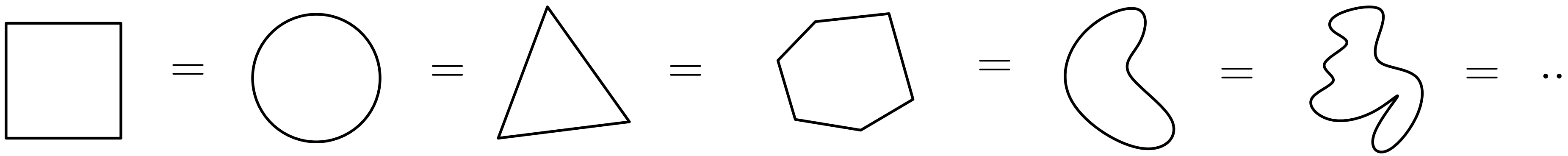
Se existe tal homeomorfismo, dizemos que os dois espaços topológicos são **homeomórficos**.

Exemplo: O círculo e o intervalo $[0,1]$ não são homeomórficos.



Classes de homeomorfismo

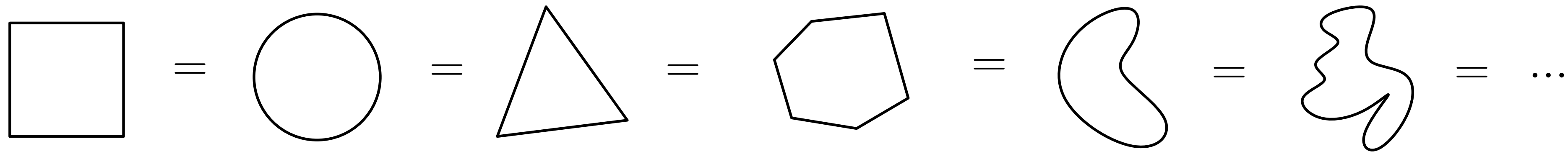
Podemos reunir espaços topológicos que são homeomórficos



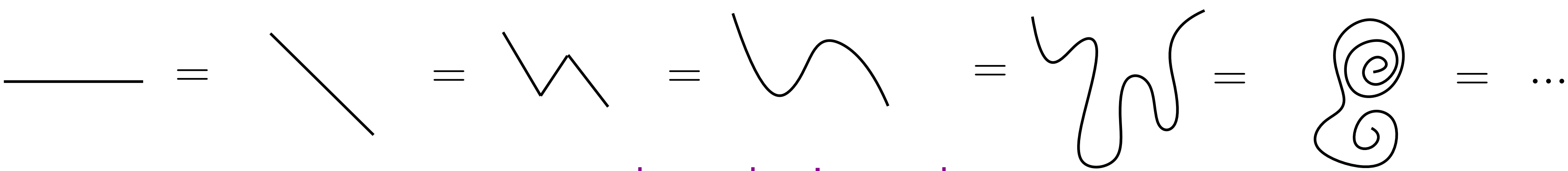
a classe dos círculos

Classes de homeomorfismo

Podemos reunir espaços topológicos que são homeomórficos



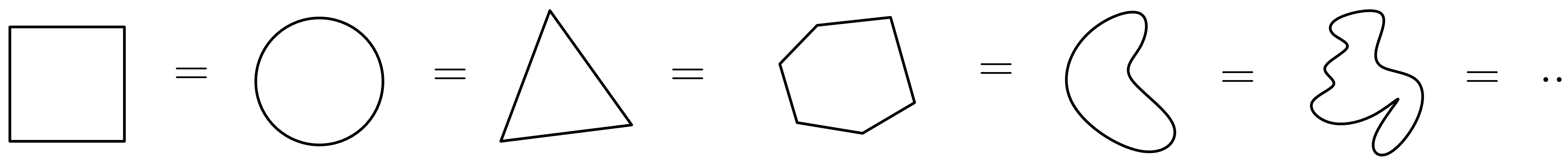
a classe dos círculos



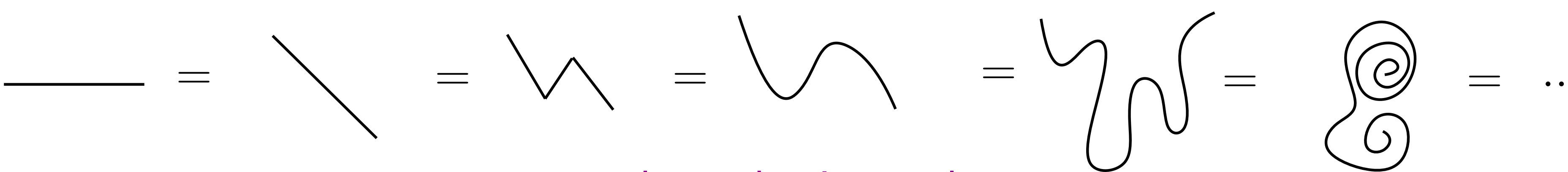
a classe dos intervalos

Classes de homeomorfismo

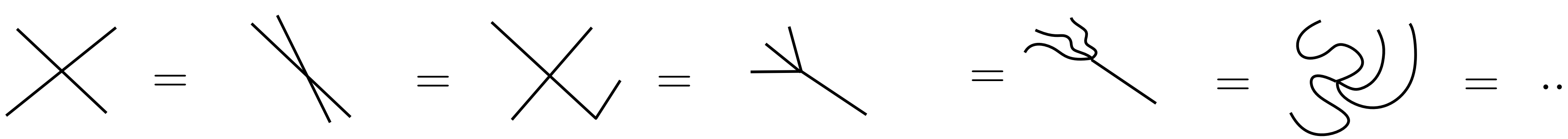
Podemos reunir espaços topológicos que são homeomórficos



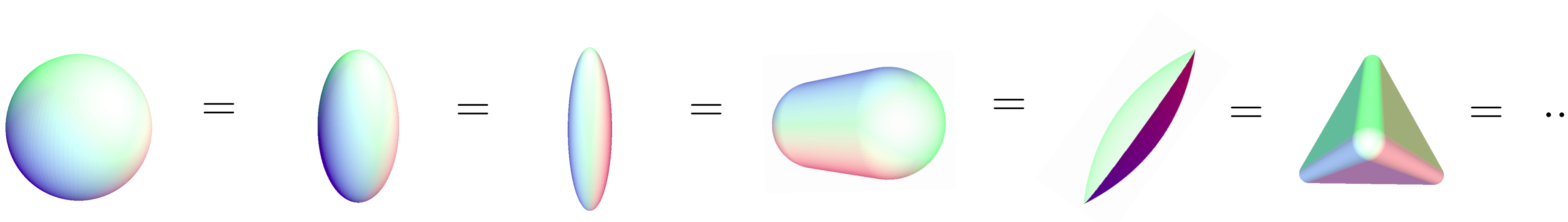
a classe dos círculos



a classe dos intervalos

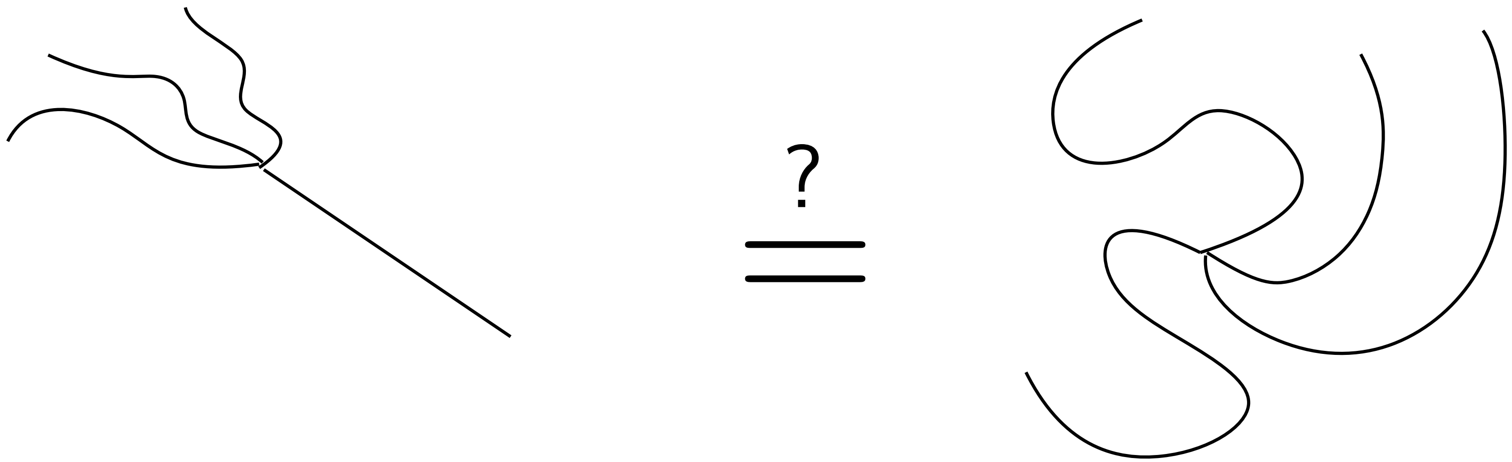


a classe das cruzes



a classe das esferas

Em geral, pode ser complicado determinar se dois espaços são homeomórficos.



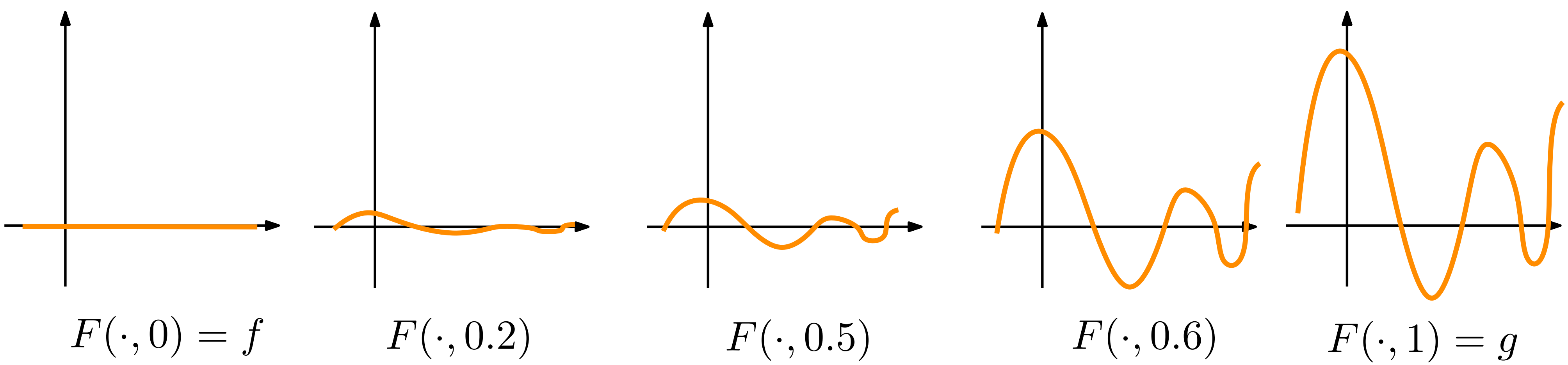
Para responder a este problema, usaremos a noção de **invariantes**.

Definição: Sejam X, Y dois espaços topológicos, and $f, g\colon X \to Y$ dois funções contínuas. Uma **homotopia** entre f e g é uma colecção $(f_t\colon X \to Y, t \in [0,1])$ tal que:

- f_0 é igual a f,
- f_1 é igual a g,
- a função $F\colon X \times [0,1] \to Y$ definida como $F(x,t) = f_t(x)$ é contínua.

Se tal homotopia existe, dizemos que as funções f e g são **homotópicas**.

Exemplo: Homotopia entre $f\colon \mathbb{R} \to \mathbb{R}$ e $g\colon \mathbb{R} \to \mathbb{R}$.

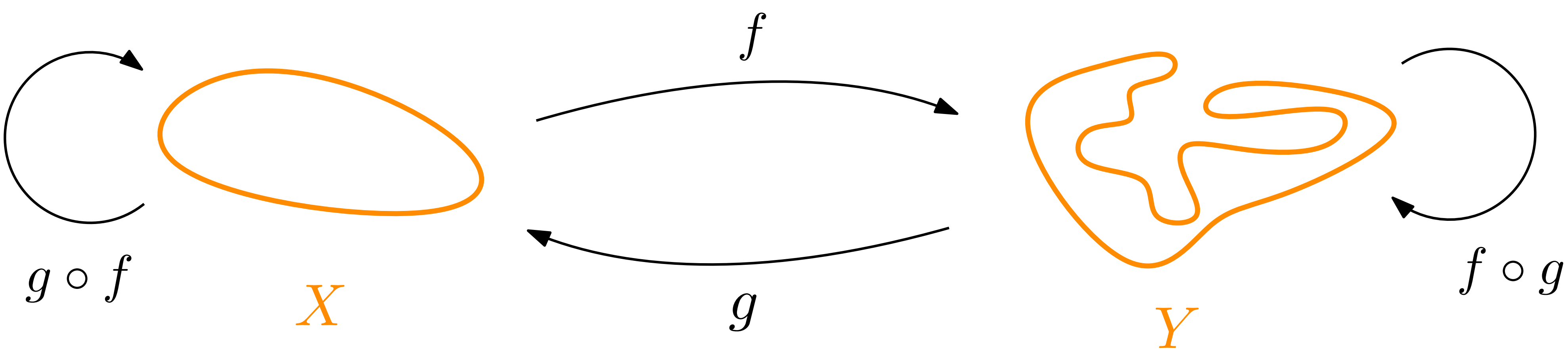


Equivalência de homotopia

Defintion: Sejam X, Y dois espaços topológicos. Uma **equivalência de homotopia** entre X e Y é um par de funções $f\colon X \to Y$ e $g\colon Y \to X$ tais que:

- $g \circ f\colon X \to X$ é homotópica à função identidade $\mathrm{id}\colon X \to X$,
- $f \circ g\colon Y \to Y$ é homotópica à função identidade $\mathrm{id}\colon Y \to Y$.

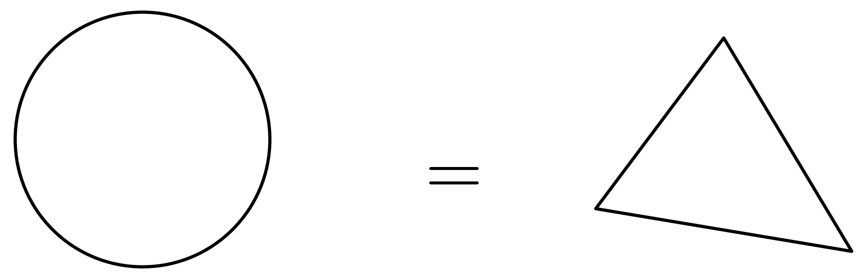
Se tal equivalência de homotopia existe, dizemos que X e Y são **homotopicamente equivalentes**.



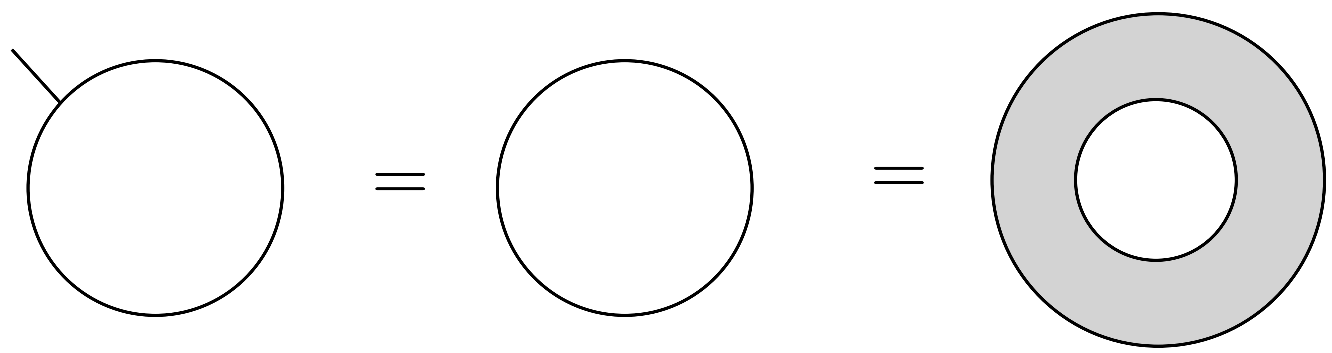
Proposição: Se dois espaços topológicos são homeomórficos, então eles são homotopicamente equivalentes.

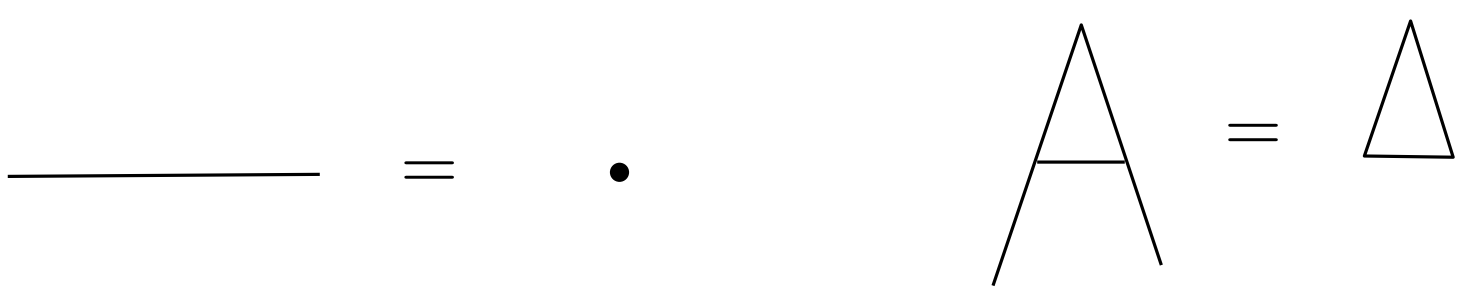
Equivalência de homotopia

A equivalência da homotopia permite **deformar** continuamente o espaço



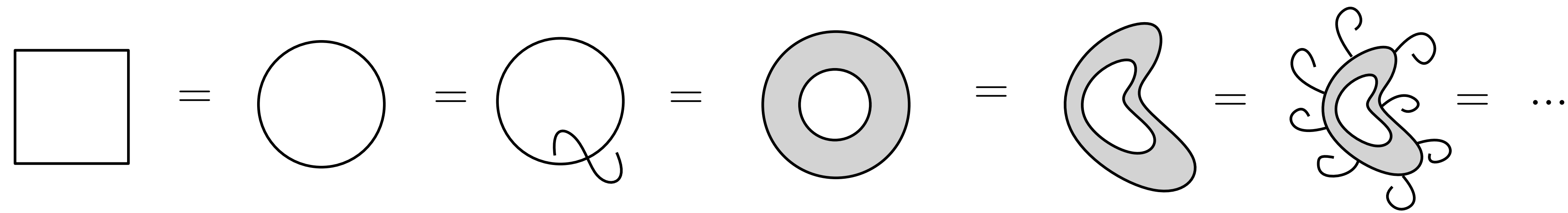
e **retrair** ele.



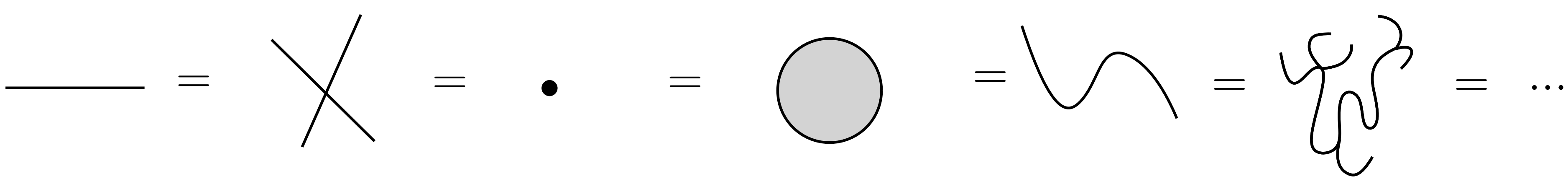


Classes de homotopia

Assim como antes, podemos classificar os espaços topológicos de acordo com esta relação, e obter **classes de equivalência de homotopia**.



a classe dos círculos



a classe dos pontos

a classe das esferas, a classe dos toros, a classe das garrafas de Klein, ...

Classes de homotopia

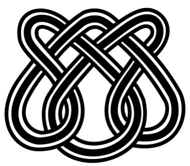
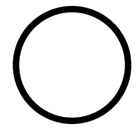
Exemplo: Classificar os seguintes espaços em classes de homotopia.

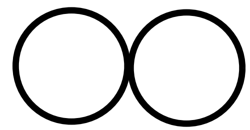
A B C D E F

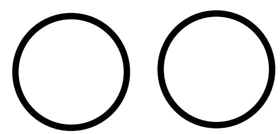
Classes de homotopia

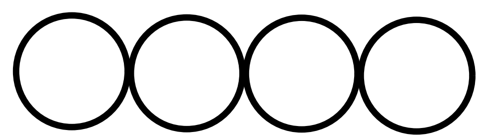
Exemplo: Classificar os seguintes espaços em classes de homotopia.

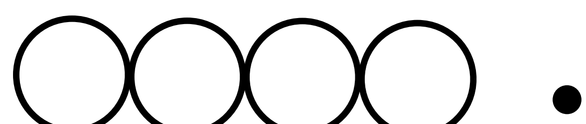
C E F $\approx \bullet$

A D  $\approx$ 

B $\approx$ 

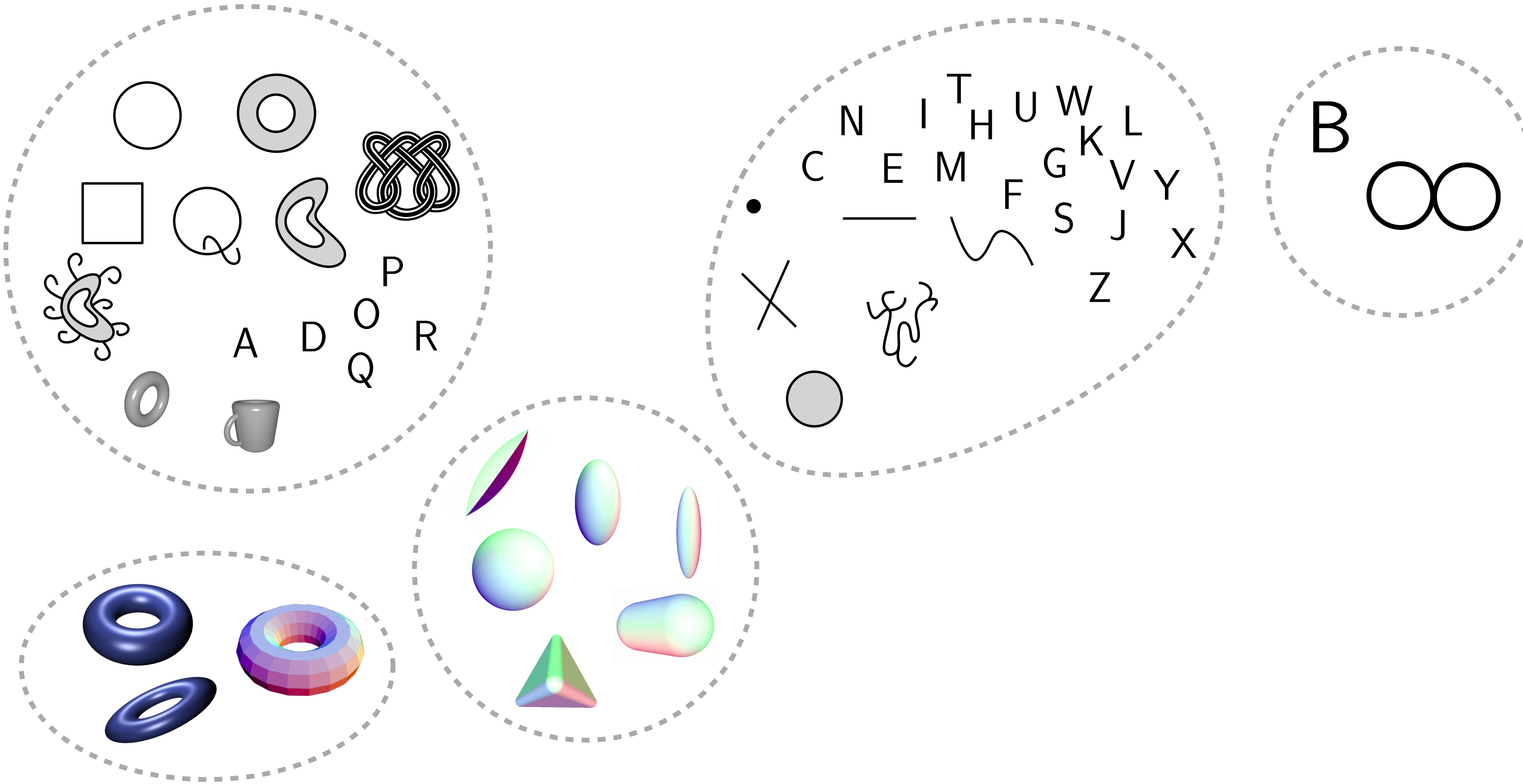
$\approx$ 

$\approx$ 

SPFC $\approx$ 

Invariantes

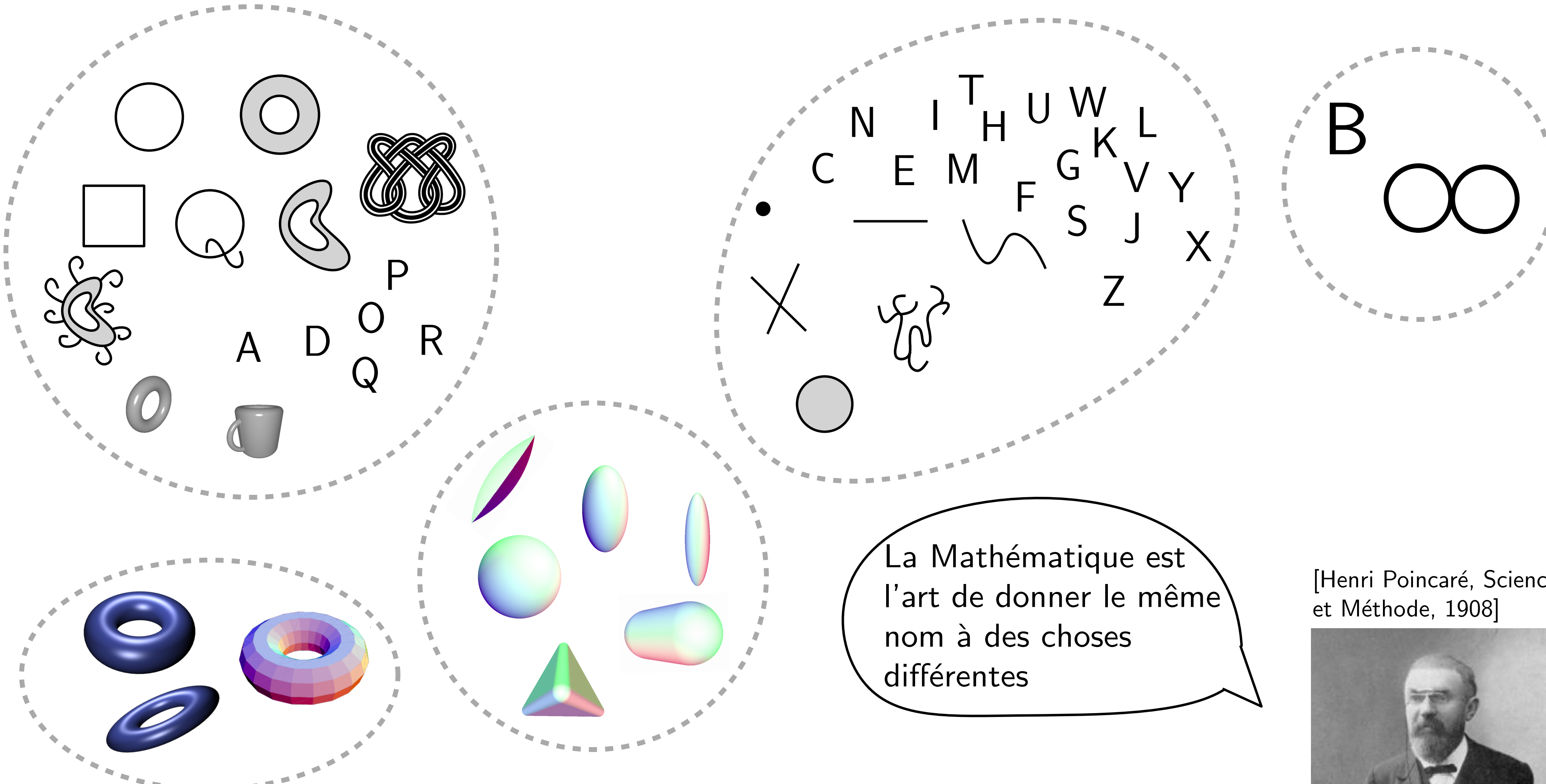
Reunimos os espaços topológicos em classes de homotopia.



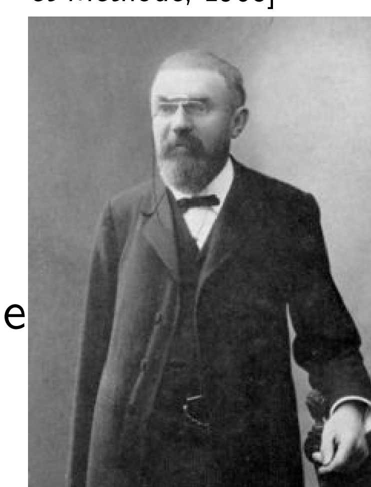
- Dado um espaço topológico de X, como reconhecer em que classe ele pertence?
- Quais são as **características comuns** dos espaços numa mesma classe?

Invariantes

Reunimos os espaços topológicos em classes de homotopia.



[Henri Poincaré, Science et Méthode, 1908]



- Dado um espaço topológico de X, como reconhecer em que classe ele pe
- Quais são as **características comuns** dos espaços numa mesma classe?

I - Comparar espaços topológicos

1 - Equivalência de homeomorfia

2 - Equivalência de homotopia

II - Invariantes topológicos

1 - Número de componentes conexas

2 - Característica de Euler

3 - Números de Betti

4 - Homologia (simplicial sobre $\mathbb{Z}/2\mathbb{Z}$)

III - Inferência topológica

1 - Estimação do parâmetro t

2 - Nervos

IV - Homologia persistente

1 - Módulos de persistência

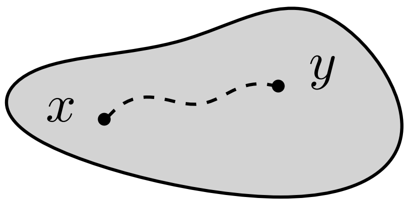
2 - Decomposição

3 - Estabilidade

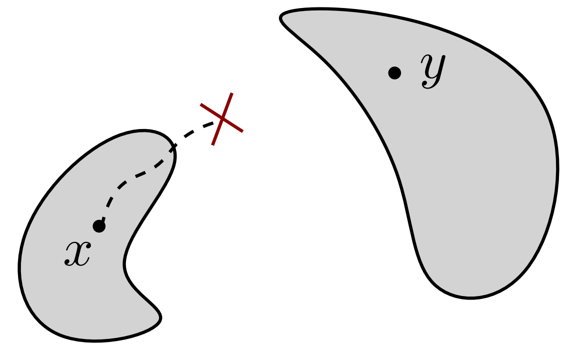
V - Aplicações

Componentes conexas

Definição: Um subconjunto $X \subset \mathbb{R}^n$ é **conexo** (por arcos) se quaisquer dois dos seus pontos estão ligados por um caminho, ou seja, por tudos $x, y \in X$, existe uma função $f\colon [0,1] \to X$ contínua tal que $f(0) = x$ e $f(1) = y$.



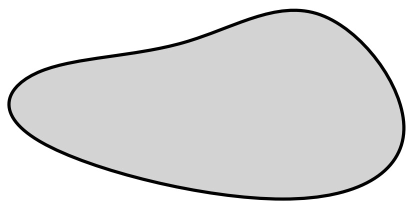
espaço conexo



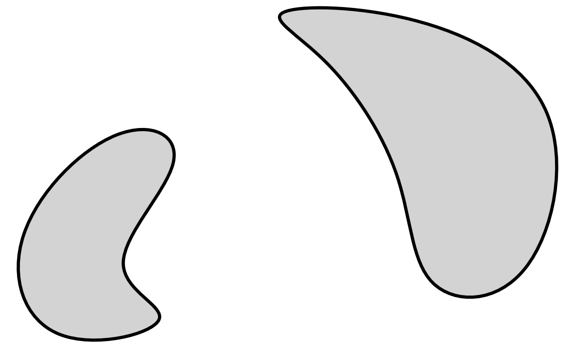
espaço não conexo

Componentes conexas

Definição: Um subconjunto $X \subset \mathbb{R}^n$ é **conexo** (por arcos) se quaisquer dois dos seus pontos estão ligados por um caminho, ou seja, por tudos $x, y \in X$, existe uma função $f\colon [0,1] \to X$ contínua tal que $f(0) = x$ e $f(1) = y$.

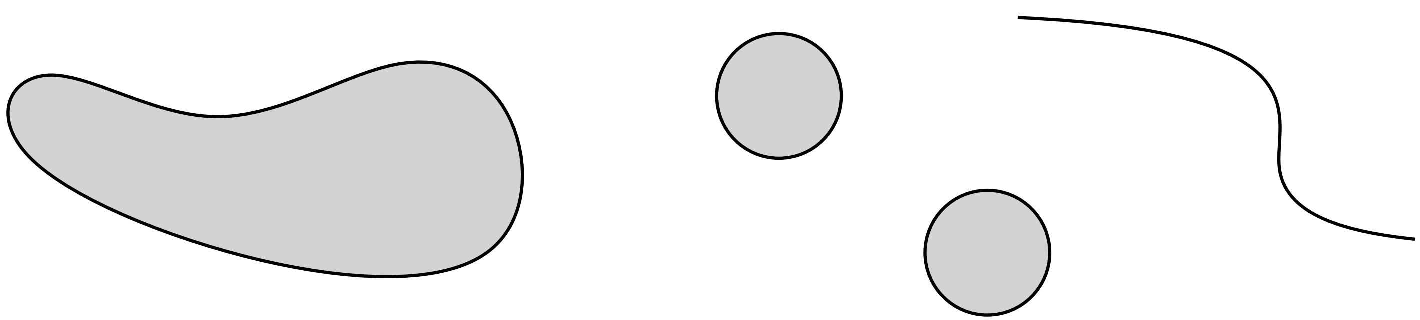


espaço conexo



espaço não conexo

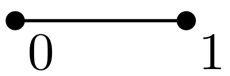
De modo geral, todo espaço topológico X pode ser dividido em **componentes conexas**.



Proposição: Se dois espaços X e Y são homotopicamente equivalentes, então eles têm o mesmo número de componentes conexas.

Consequência: Se dois espaços X e Y são homeomórficos, então eles têm o mesmo número de componentes conexas.

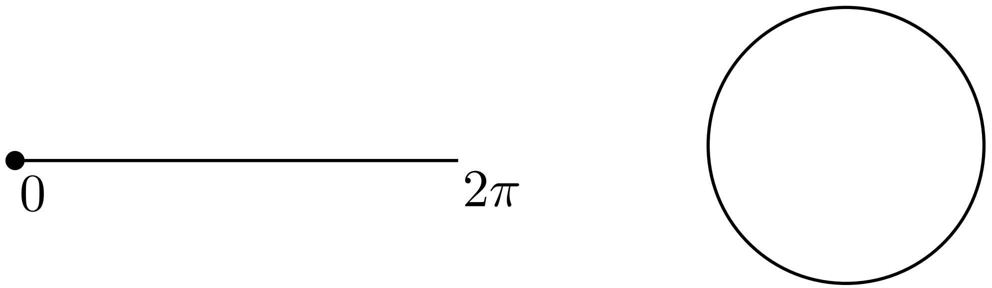
Exemplo: Os subcojuntos $[0,1]$ e $[0,1] \cup [2,3]$ do $\mathbb{R}$ eles não são homeomórficos, nem homotopicamente equivalentes.
De fato, o primeiro tem uma componente conexa, e o segundo duas.



Proposição: Se dois espaços X e Y são homotopicamente equivalentes, então eles têm o mesmo número de componentes conexas.

Exemplo: O intervalo $[0, 2\pi)$ e o círculo $\mathbb{S}_1 \subset \mathbb{R}^2$ não são homeomórficos.

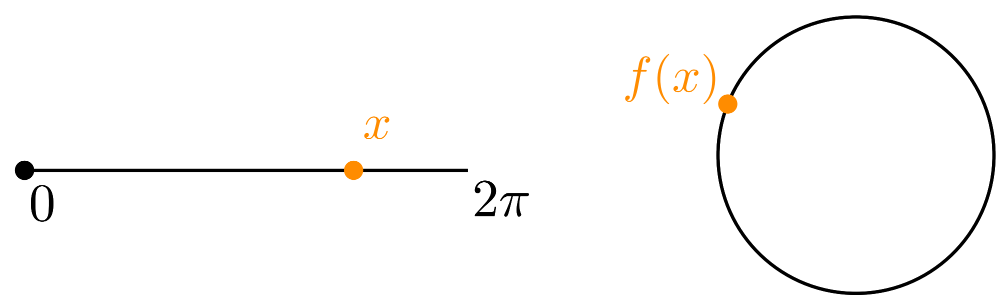
Provaremos isto por contradição. Suponhamos que sejam homeomórficos. Por definição, isto significa que existe uma função $f\colon [0, 2\pi) \to \mathbb{S}_1$ que é contínua, invertível, e com inversa contínua.



Proposição: Se dois espaços X e Y são homotopicamente equivalentes, então eles têm o mesmo número de componentes conexas.

Exemplo: O intervalo $[0, 2\pi)$ e o círculo $\mathbb{S}_1 \subset \mathbb{R}^2$ não são homeomórficos.

Provaremos isto por contradição. Suponhamos que sejam homeomórficos. Por definição, isto significa que existe uma função $f\colon [0, 2\pi) \to \mathbb{S}_1$ que é contínua, invertível, e com inversa contínua.



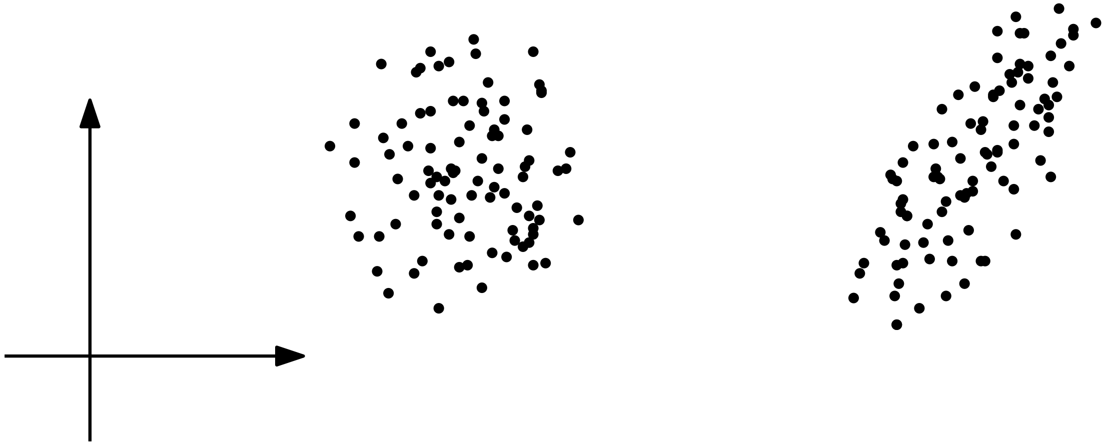
Seja $x \in [0, 2\pi)$ tal que $x \neq 0$. Considere os subconjuntos $[0, 2\pi) \setminus \{x\} \subset [0, 2\pi)$ e $\mathbb{S}_1 \setminus \{f(x)\} \subset \mathbb{S}_1$, e a função induzida

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

A função g é um homeomorfismo.

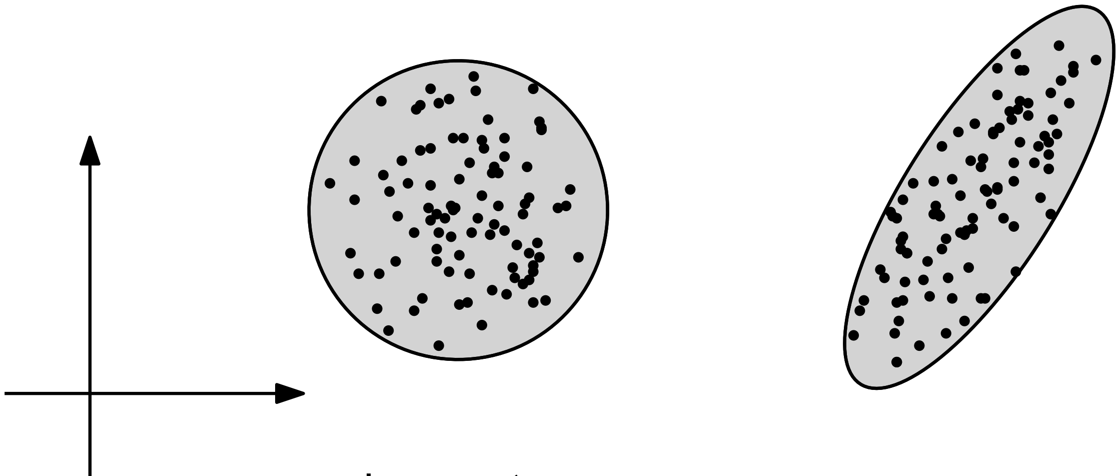
Porém, $[0, 2\pi) \setminus \{x\}$ tem duas componentes conexas, a $t\mathbb{S}_1 \setminus \{f(x)\}$ apenas uma. Isto é um absurdo.

Nas aplicações, encontrar componentes conexas corresponde a uma tarefa **classificação**.

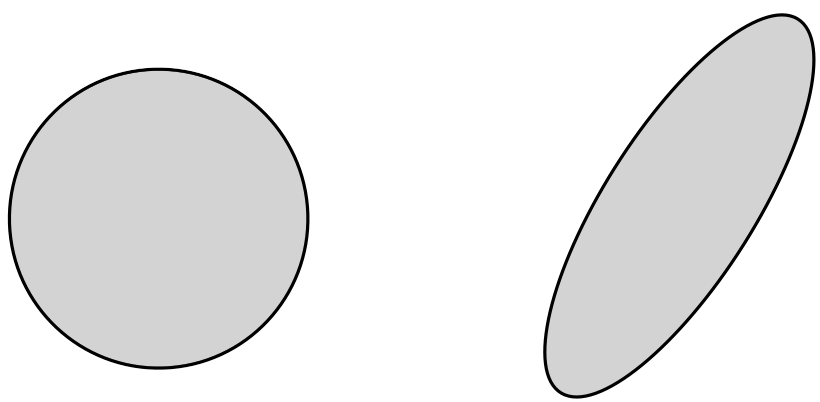


Propriedade de invariância - aplicações

Nas aplicações, encontrar componentes conexas corresponde a uma tarefa **classificação**.



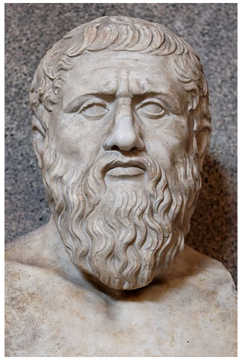
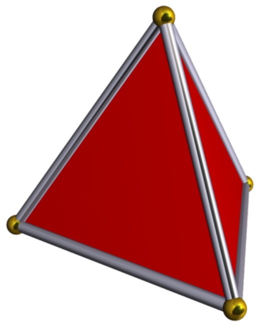
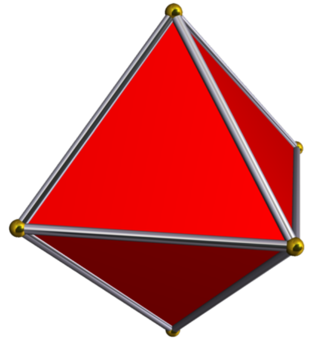
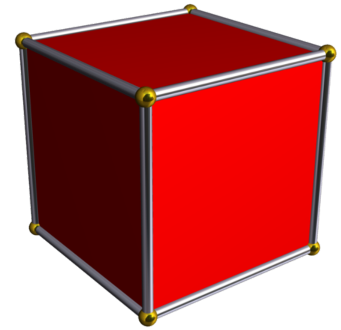
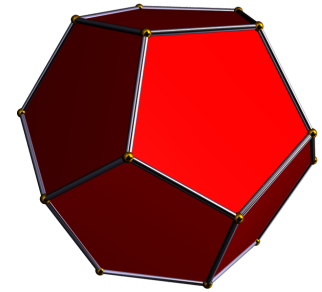
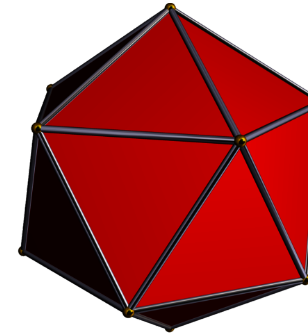
cluster 1 cluster 2



componente conexa 1 componente conexa 2

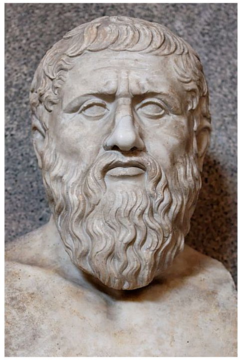
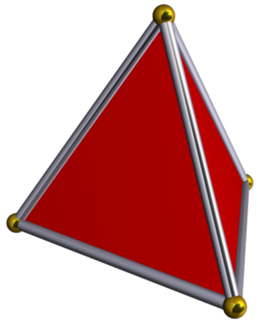
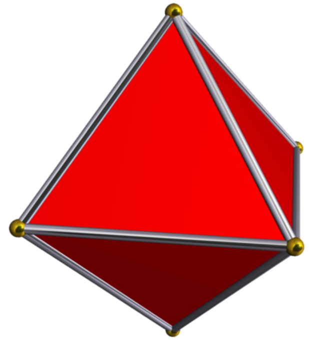
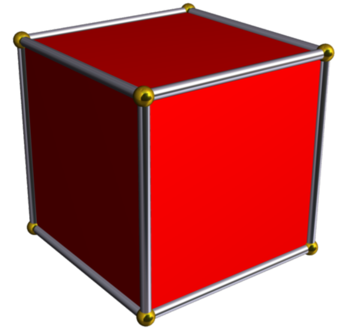
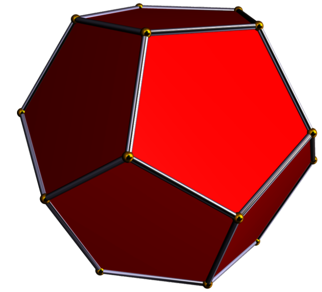
Podemos pensar nestes conjuntos como um **espaço topológico subjacente**, sobre os quais os pontos são amostrados.

Característica de Euler para poliedros

número de faces	4	8	6	12	20
número de arestas	6	12	12	30	30
número de vértices	4	6	8	20	12

Característica de Euler para poliedros


número de faces	4	8	6	12	20
número de arestas	6	12	12	30	30
número de vértices	4	6	8	20	12
χ	2	2	2	2	2

Proposição [Euler, 1758]: Em qualquer poliedro convexo, temos número de faces $-$ número de arestas $+$ número de vértices $= 2$

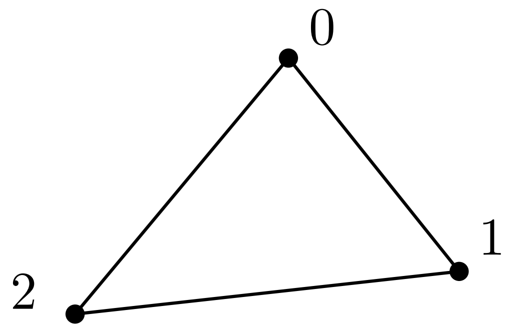
Definição: Seja V um conjunto (chamado o conjunto de *vértices*). Um **complexo simplicial** é uma colecção K de subconjuntos de V (chamados os *simplexos*) tais que, por todo $\sigma \in K$ e todo $\tau \subset \sigma$, temos $\tau \in K$.

A **dimensão** de um simplexo $\sigma \in K$ é definida como $|\sigma| - 1$.

Exemplo: Sejam $V = \{0, 1, 2\}$ e

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

É um complexo simplicial.



Contém três simplexos de dimensão 0 ($[0]$, $[1]$ e $[2]$) e três simplexos de dimensão 1 ($[0, 1]$, $[1, 2]$ e $[0, 2]$).

Característica de Euler para complexos

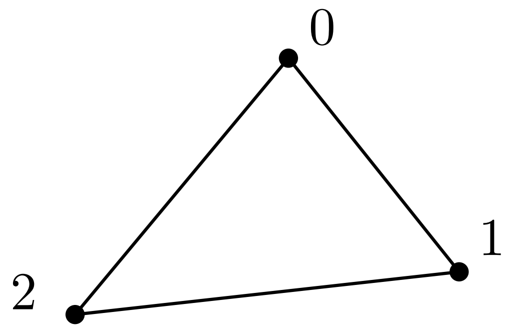
Definição: Seja V um conjunto (chamado o conjunto de *vértices*). Um **complexo simplicial** é uma colecção K de subconjuntos de V (chamados os *simplexos*) tais que, por todo $\sigma \in K$ e todo $\tau \subset \sigma$, temos $\tau \in K$.

A **dimensão** de um simplexo $\sigma \in K$ é definida como $|\sigma| - 1$.

Exemplo: Sejam $V = \{0, 1, 2\}$ e

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

É um complexo simplicial.



(é um círculo)

Contém três simplexos de dimensão 0 ($[0]$, $[1]$ e $[2]$) e três simplexos de dimensão 1 ($[0, 1]$, $[1, 2]$ e $[0, 2]$).

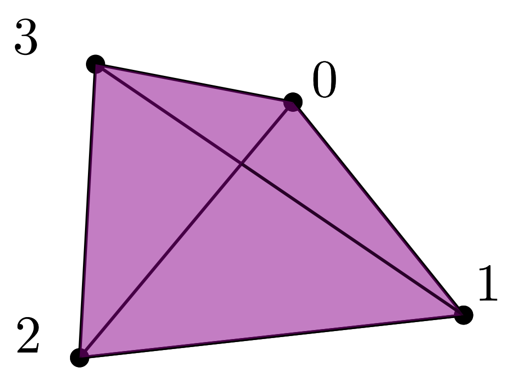
Definição: Seja V um conjunto (chamado o conjunto de *vértices*). Um **complexo simplicial** é uma colecção K de subconjuntos de V (chamados os *simplexos*) tais que, por todo $\sigma \in K$ e todo $\tau \subset \sigma$, temos $\tau \in K$.

A **dimensão** de um simplexo $\sigma \in K$ é definida como $|\sigma| - 1$.

Exemplo: Sejam $V = \{0, 1, 2, 3\}$ e

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

É um complexo simplicial.



Contém quatro simplexos de dimensão 0 ($[0]$, $[1]$, $[2]$ e $[3]$), seis simplexos de dimensão 1 ($[0,1]$, $[1,2]$, $[2,3]$, $[3,0]$,$[0,2]$ e $[1,3]$) e quatro simplexos de dimensão 2 ($[0,1,2]$, $[0,1,3]$, $[0,2,3]$ e $[1,2,3]$).

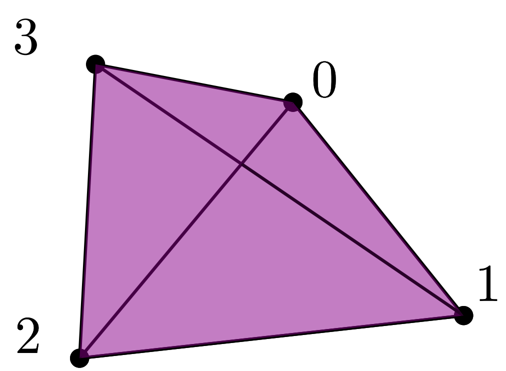
Definição: Seja V um conjunto (chamado o conjunto de *vértices*). Um **complexo simplicial** é uma colecção K de subconjuntos de V (chamados os *simplexos*) tais que, por todo $\sigma \in K$ e todo $\tau \subset \sigma$, temos $\tau \in K$.

A **dimensão** de um simplexo $\sigma \in K$ é definida como $|\sigma| - 1$.

Exemplo: Sejam $V = \{0, 1, 2, 3\}$ e

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

É um complexo simplicial.



(é uma esfera)

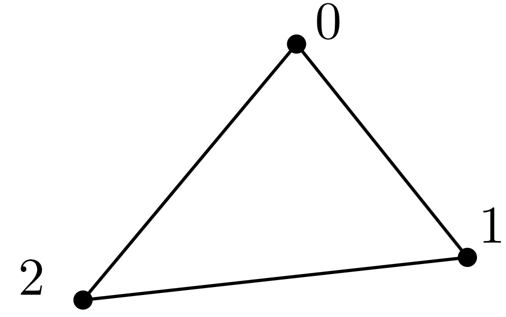
Contém quatro simplexos de dimensão 0 ($[0]$, $[1]$, $[2]$ e $[3]$), seis simplexos de dimensão 1 ($[0, 1]$, $[1, 2]$, $[2, 3]$, $[3, 0]$,$[0, 2]$ e $[1, 3]$) e quatro simplexos de dimensão 2 ($[0, 1, 2]$, $[0, 1, 3]$, $[0, 2, 3]$ e $[1, 2, 3]$).

Definição: Seja K um complexo simplicial de dimenção n. A sua *característica de Euler* é o número inteiro

$$\chi(K) = \sum_{0\leq i\leq n} (-1)^i \cdot (\text{número de simplexos de dimensão } i).$$

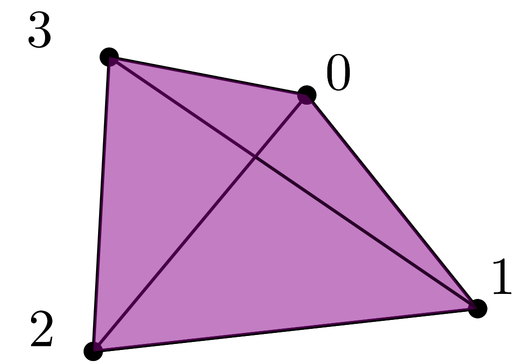
Exemple: O complexo simplicial $K = \{[0], [1], [2], [0,1], [1,2], [2,0]\}$ tem característica de Euler

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



Exemplo: O complexo simplicial
$K = \{[0], [1], [2], [3], [0,1], [1,2], [2,3], [3,0], [0,2], [1,3], [0,1,2], [0,1,3], [0,2,3], [1,2,3]\}$
tem característica de Euler

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

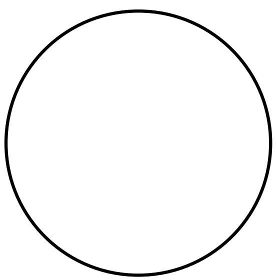


Característica de Euler para espaços

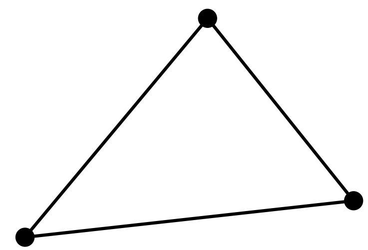
Os complexos simpliciais também são espaços topológicos.

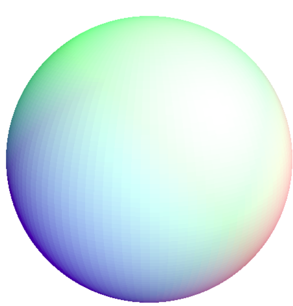
Definição: Uma **triangulação** de um espaço topológico X é um complexo simplicial K homeomórfico a X.

Definição: Seja X um espaço topológico. A sua **característica de Euler** é definida como a característica de Euler de uma de suas triangulações.

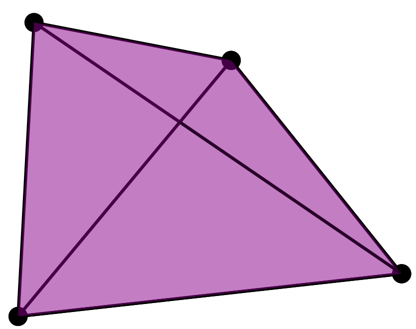


$\chi(\mathbb{S}_1) = 0$





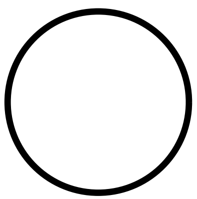
$\chi(\mathbb{S}_2) = 2$



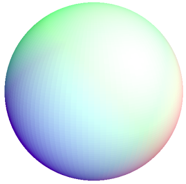
Proposição: Se dois espaços X e Y são homotopicamente equivalentes, então $\chi(X) = \chi(Y)$.

Portanto, a característica de Euler é um **invariante** de classes de equivalência de homotopia.
Podemos usar esta informação para provar que dois espaços não são homotopicamente equivalentes.

Exemplo: O círculo tem característica de Euler 0, e a esfera 2.
Portanto, eles não são equivalentes.



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



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

Proposição: Se dois espaços X e Y são homotopicamente equivalentes, então $\chi(X) = \chi(Y)$.

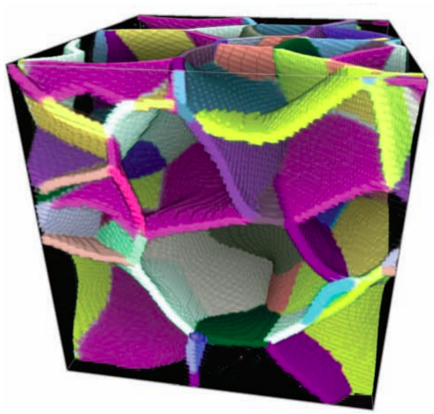
Portanto, a característica de Euler é um **invariante** de classes de equivalência de homotopia.
Podemos usar esta informação para provar que dois espaços não são homotopicamente equivalentes.

Exemplo: Classificação das superfícies.
As classes de homeomorfismo das superfícies *conexes e compactas* são classificadas por sua característica de Euler.

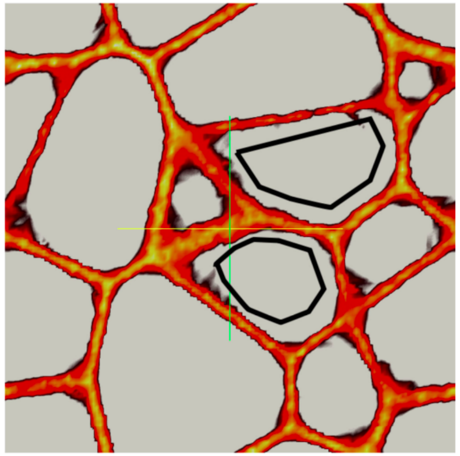
χ $\quad 2 \quad 0 \quad -2 \quad -4 \quad 2 - 2 \times$ género

Propriedade de invariância - aplicações

[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]



A característica Euler 'conta' o número de buracos

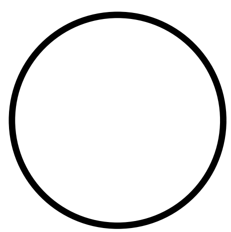
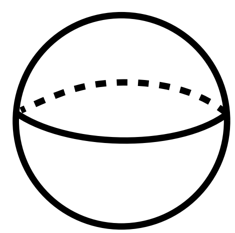
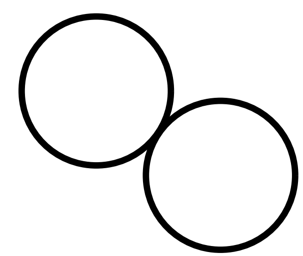
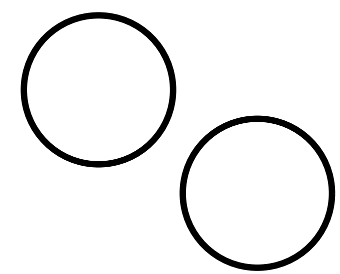
Números de Betti

Para todo espaço topológico X, podemos definir uma sequência de números inteiros

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

chamados de **números de Betti**.

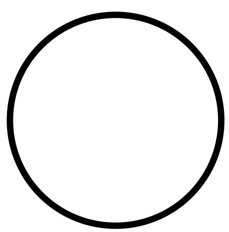
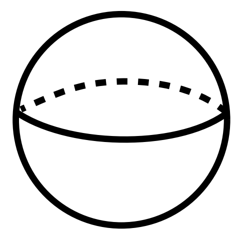
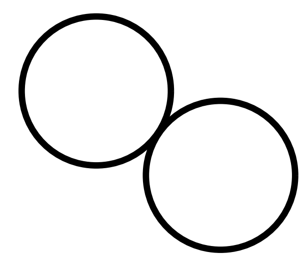
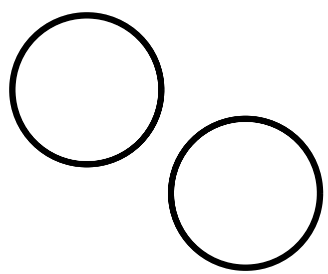
Construção dos números Betti: baseada na teoria da homologia.

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

Números de Betti

Interpretação:

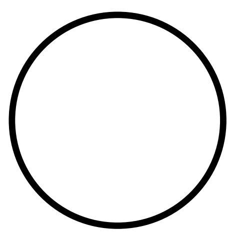
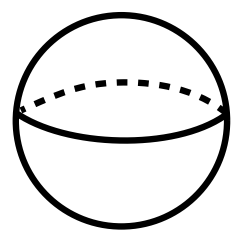
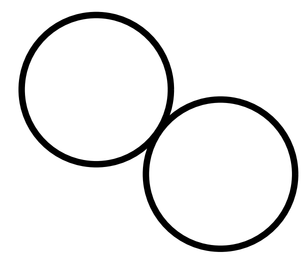
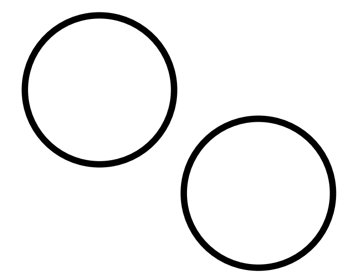
- $\beta_0(X)$ é o número de componentes conexas de X
- $\beta_1(X)$ é o número de 'buracos' em X
- $\beta_2(X)$ é o número de 'vazios' em X
- $\ldots$

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

Números de Betti

Interpretação:

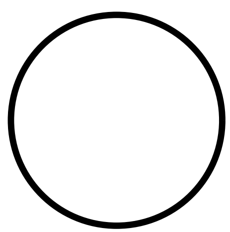
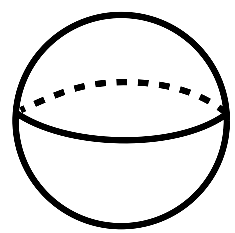
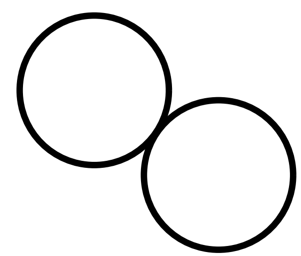
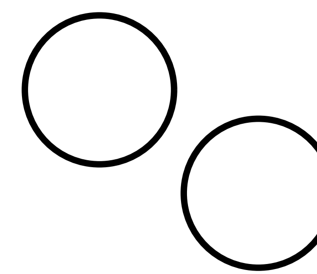
- $\beta_0(X)$ é o número de componentes conexas de X
- $\beta_1(X)$ é o número de 'buracos' em X
- $\beta_2(X)$ é o número de 'vazios' em X
- $\ldots$

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

Números de Betti

Interpretação:

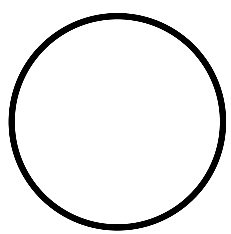
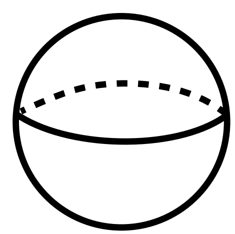
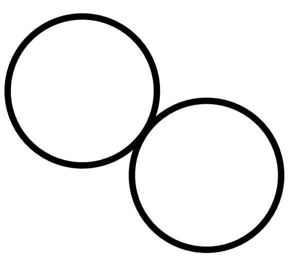
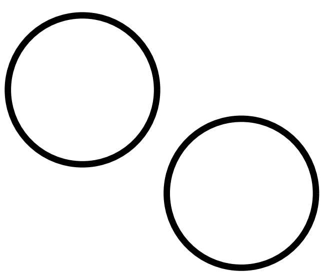
- $\beta_0(X)$ é o número de componentes conexas de X
- $\beta_1(X)$ é o número de 'buracos' em X
- $\beta_2(X)$ é o número de 'vazios' em X
- $\ldots$

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

Números de Betti

Interpretação:

- $\beta_0(X)$ é o número de componentes conexas de X
- $\beta_1(X)$ é o número de 'buracos' em X
- $\beta_2(X)$ é o número de 'vazios' em X
- $\ldots$

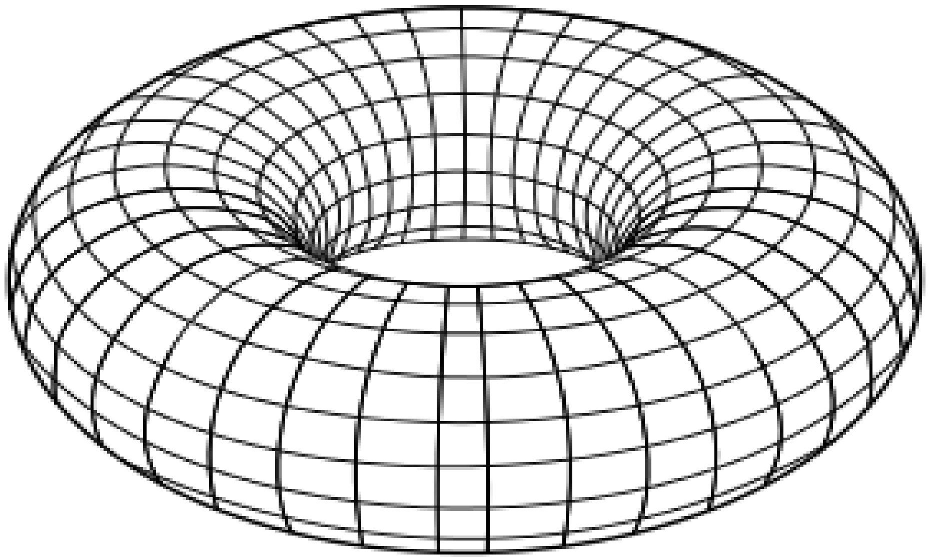
X					
$\beta_0(X)$	1	1	1	1	2
$\beta_1(X)$	0	1	0	2	2
$\beta_2(X)$	0	0	1	0	0

Interpretação:

- $\beta_0(X)$ é o número de componentes conexas de X
- $\beta_1(X)$ é o número de 'buracos' em X
- $\beta_2(X)$ é o número de 'vazios' em X
- $\ldots$

Exemplo: Números de Betti do toro:

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



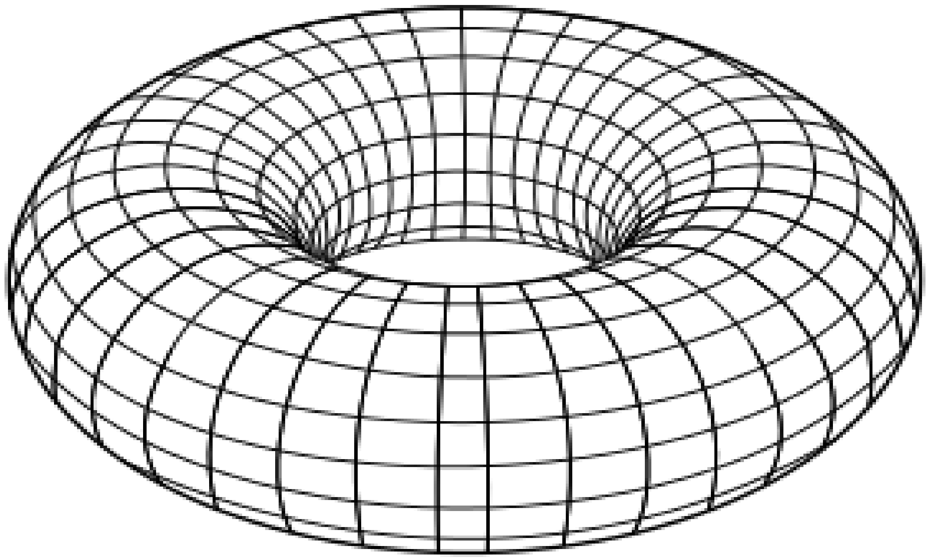
Números de Betti

Interpretação:

- $\beta_0(X)$ é o número de componentes conexas de X
- $\beta_1(X)$ é o número de 'buracos' em X
- $\beta_2(X)$ é o número de 'vazios' em X
- $\ldots$

Exemplo: Números de Betti do toro:

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



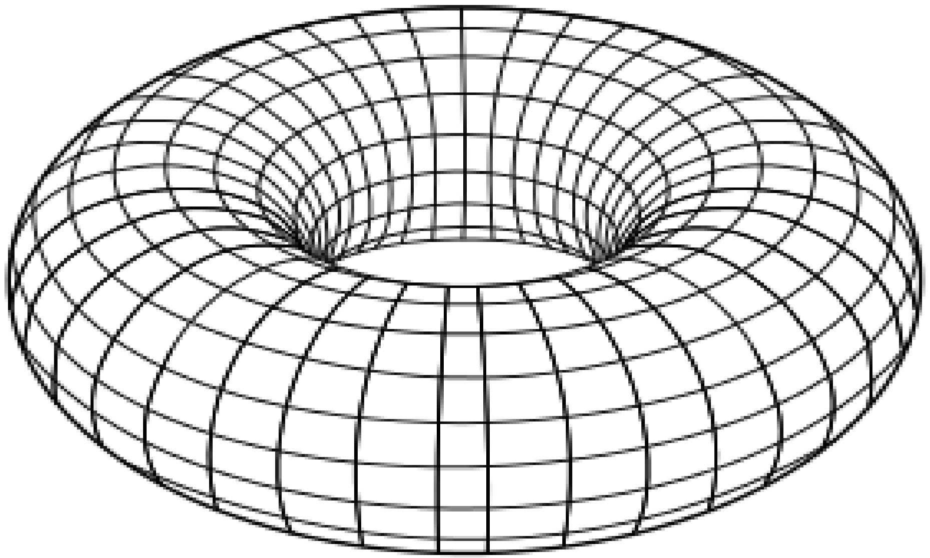
Números de Betti

Interpretação:

- $\beta_0(X)$ é o número de componentes conexas de X
- $\beta_1(X)$ é o número de 'buracos' em X
- $\beta_2(X)$ é o número de 'vazios' em X
- $\ldots$

Exemplo: Números de Betti do toro:

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



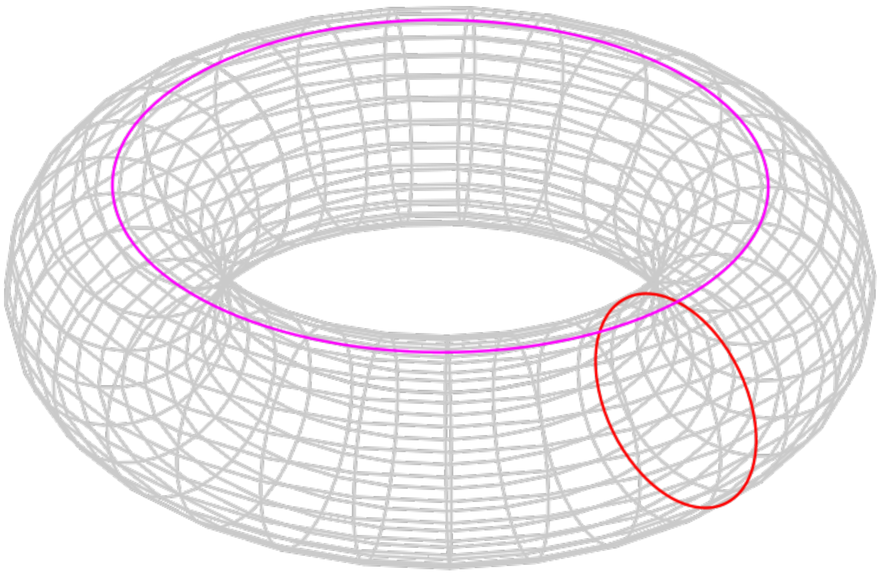
Números de Betti

Interpretação:

- $\beta_0(X)$ é o número de componentes conexas de X
- $\beta_1(X)$ é o número de 'buracos' em X
- $\beta_2(X)$ é o número de 'vazios' em X
- $\ldots$

Exemplo: Números de Betti do toro:

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



Propriedade de invariância - na teoria

Proposição: Se dois espaços X e Y são homotopicamente equivalentes, então eles têm os mesmos números de Betti.

Consequentemente, dois espaços com números de Betti diferentes não podem ser equivalentes à homotopia.

Exemplo: A esfera $\mathbb{S}_n \subset \mathbb{R}^{n+1}$ tem números de Betti

$$\begin{aligned} \beta_i(X) &= 1 \quad \text{if} \;\; i = 0 \text{ or } n, \\ \beta_i(X) &= 0 \quad \text{else.} \end{aligned}$$

Portanto, se $n \neq m$, então $\mathbb{S}_n$ e $\mathbb{S}_m$ não são equivalentes.

Proposição: Se dois espaços X e Y são homotopicamente equivalentes, então eles têm os mesmos números de Betti.

Consequentemente, dois espaços com números de Betti diferentes não podem ser equivalentes à homotopia.

Exemplo: Invariância de domínio de Brouwer.

Vamos mostrar que $\mathbb{R}^n$ e $\mathbb{R}^m$, com $n \neq m$, não são homeomórficos.

Seja $h\colon \mathbb{R}^n \to \mathbb{R}^m$ um homeomorfismo.
Escolha qualquer $x \in \mathbb{R}^n$ e considera a restrição

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

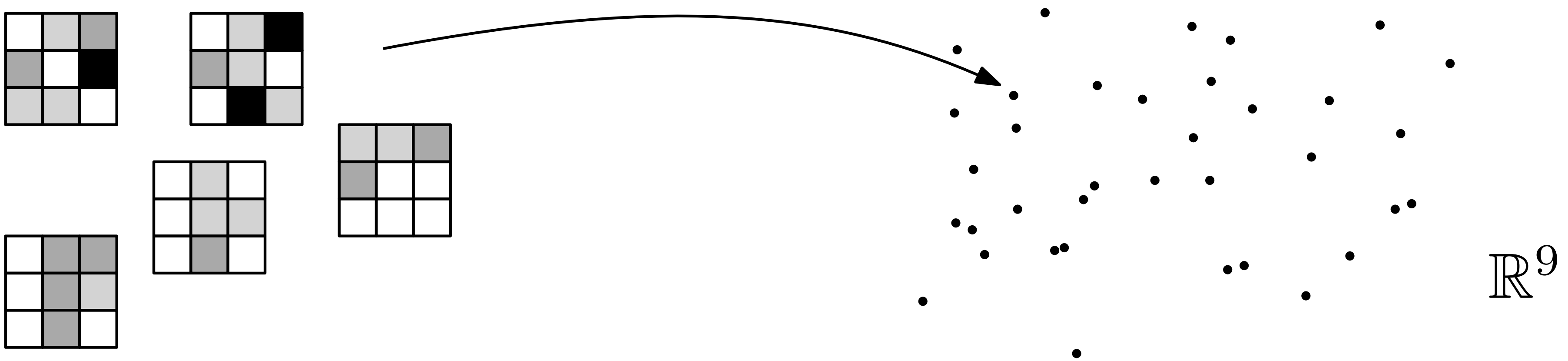
Ainda é um homeomorfismo.

Mas $\mathbb{R}^n \setminus \{x\}$ é homotópico à esfera $\mathbb{S}_{n-1}$, e $\mathbb{R}^m \setminus \{x\}$ is é homotópico à esfera $\mathbb{S}_{m-1}$

Já vimos que $\mathbb{S}_{n-1}$ e $\mathbb{S}_{m-1}$ são homotópicos se e somente se $m = n$. Isto é uma contradição.

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

A partir de uma grande coleção de imagens naturais, os autores extraem patches 3×3. Como cada patch é composto de 9 pixels, ele pode ser visto como um vetor 9-dimensional, e o conjunto inteiro como uma nuvem de pontos em $\mathbb{R}^9$.

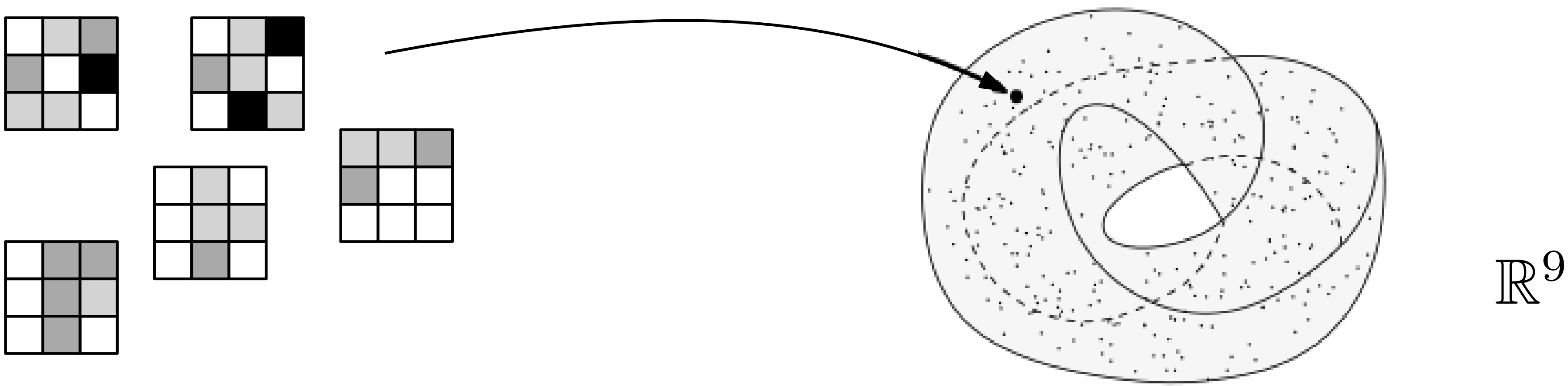


Observam que a nuvem de pontos está próxima de uma forma cujos número de Betti são

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

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

A partir de uma grande coleção de imagens naturais, os autores extraem patches 3×3. Como cada patch é composto de 9 pixels, ele pode ser visto como um vetor 9-dimensional, e o conjunto inteiro como uma nuvem de pontos em $\mathbb{R}^9$.



Observam que a nuvem de pontos está próxima de uma forma cujos número de Betti são

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

São os números Betti de uma garrafa de Klein!

(e os autores realmente mostram que o conjunto de dados se concentra perto de uma garrafa de Klein imersa no $\mathbb{R}^9$.)

Homologia simplicial

Para todo complexo simplicial K, podemos definir uma sequência de espaços vetoriais sobre o corpo finito $\mathbb{Z}/2\mathbb{Z}$:

$$H_0(K), \quad H_1(K), \quad H_2(K), \quad H_3(K), \quad \ldots$$

chamados de **grupos de homologia simplicial**.

Observação: O corpo $\mathbb{Z}/2\mathbb{Z}$ é o conjunto $\{0, 1\}$ dotado das operações

$$\begin{aligned} 0+0&=0 \\ 0+1&=1 \\ 1+0&=1 \\ 1+1&=0 \end{aligned} \qquad\qquad \begin{aligned} 0\times 0&=0 \\ 0\times 1&=0 \\ 1\times 0&=0 \\ 1\times 1&=1 \end{aligned}$$

Um espaço vetorial sobre $\mathbb{Z}/2\mathbb{Z}$ de dimensão finita pode ser escrito como $(\mathbb{Z}/2\mathbb{Z})^k$.

Se X é um espaço topológico, definimos seus grupos de homologia $H_i(X)$ como os de uma de suas triangulações.

Seja $n \geq 0$. As n-**cadeias** de K é o conjunto $C_n(K)$ cujos elementos são as somas formais

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

Podemos definir sobre $C_n(K)$ uma **estrutura de grupo** com

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

Sejam $n \geq 1$ e $\sigma = [x_0, ..., x_n] \in K_{(n)}$ um simplexo de dimensão n. O seu **bordo** é o seguinte element de $C_{n-1}(K)$: $\partial_n \sigma = \sum_{\substack{\tau \subset \sigma \\ |\tau| = |\sigma| - 1}} \tau$. Podemos estender o operador ∂_n como uma transformação linear $\partial_n : C_n(K) \to C_{n-1}(K)$.

Seja $n \geq 0$. Temos uma sequência de espaços vetoriais

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

As aplicações ∂_{n+1} e ∂_n são transformações lineares, e podemos considerar seus núcleos e suas imagens. Definimos os n-**ciclos** $Z_n(K) = \mathrm{Ker}(\partial_n)$ e os n-**bordos** $B_n(K) = \mathrm{Im}(\partial_{n+1})$.

Definição: O n-**ésimo grupo de homologia (simplicial)** de K é o espaço vetorial quociente

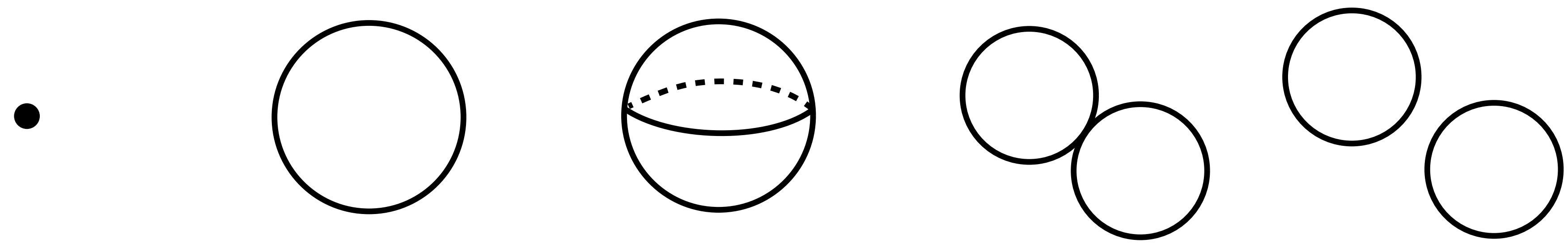
$$H_n(K) = Z_n(K)/B_n(K).$$

Interpretação dos grupos de homologia

Um espaço vetorial sobre $\mathbb{Z}/2\mathbb{Z}$ de dimensão finita pode ser escrito como $(\mathbb{Z}/2\mathbb{Z})^k$.

Definição: Sejam K um complexo simplicial e $n \geq 0$. Seu n-ésimo **número de Betti** é definido como $\beta_n(K) = \dim H_n(K)$.

X



$H_0(X)$	$\mathbb{Z}/2\mathbb{Z}$	$\mathbb{Z}/2\mathbb{Z}$	$\mathbb{Z}/2\mathbb{Z}$	$\mathbb{Z}/2\mathbb{Z}$	$(\mathbb{Z}/2\mathbb{Z})^2$
$\beta_0(X)$	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

Homologia é um functor

Temos visto que a homologia transforma **espaços topológicos** em **espaços vetoriais**

$$\begin{aligned} H_i \colon \mathrm{Top} &\longrightarrow \mathrm{Vect} \\ X &\longmapsto H_i(X) \end{aligned}$$

Na verdade, ele também transforma **aplicações contínuas** em **transfomações lineares**.

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

Esta operação preserva os **diagramas comutativos**:

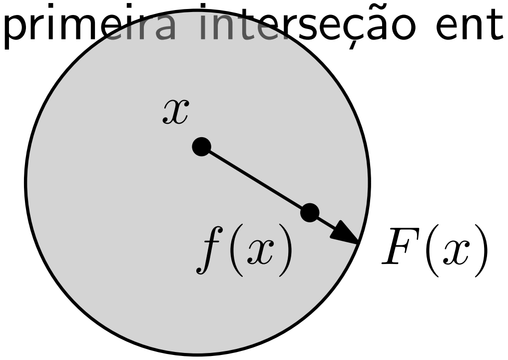
$$\begin{tikzcd} X \arrow[r, "f"'] \arrow[rr, bend left, "g\circ f"] & Y \arrow[r, "g"'] & Z, \end{tikzcd} \qquad \begin{tikzcd} H_n(X) \arrow[r, "H_n(f)"'] \arrow[rr, bend left, "H_n(g\circ f)"] & H_n(Y) \arrow[r, "H_n(g)"'] & H_n(Z). \end{tikzcd}$$

$$\boxed{H_n(g \circ f) = H_n(g) \circ H_n(f)}$$

Aplicação: Teorema do ponto fixo de Brouwer

Seja $f\colon \mathcal{B} \to \mathcal{B}$ uma aplicação contínua, one $\mathcal{B}$ é a bola fechada unidade de $\mathbb{R}^n$. Vamos mostrar que f tem um ponto fixo ($f(x) = x$).

Caso contrário, podemos definir uma aplicação $F\colon \mathcal{B} \to \partial\mathcal{B}$ tal que F restrita to $\partial\mathcal{B}$ é a identidade.
Para isso, define $F(x)$ como a primeira interseção entre a meia linha $[x, f(x))$ e $\partial\mathcal{B}$.



Chame a inclusão $i\colon \partial\mathcal{B} \to \mathcal{B}$. Então $F \circ i\colon \partial\mathcal{B} \to \partial\mathcal{B}$ é a identidade.
Por functorialidade, temos diagramas comutativos

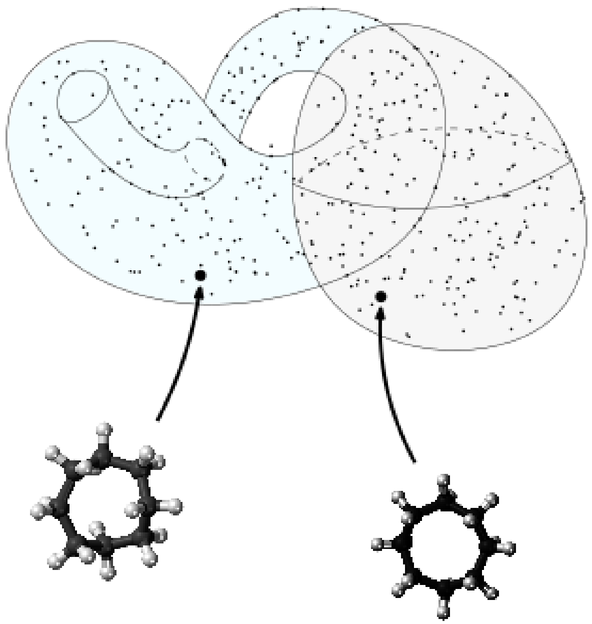
$$\begin{tikzcd} \partial\mathcal{B} \arrow[r, "i"'] \arrow[rr, bend left, "\mathrm{id}"] & \mathcal{B} \arrow[r, "F"'] & \partial\mathcal{B}, \end{tikzcd} \qquad \begin{tikzcd} H_i(\partial\mathcal{B}) \arrow[r, "H_i(i)"'] \arrow[rr, bend left, "H_i(\mathrm{id})"] & H_i(\mathcal{B}) \arrow[r, "H_i(F)"'] & H_i(\partial\mathcal{B}). \end{tikzcd}$$

Mas para $i = n - 1$, é absurdo:

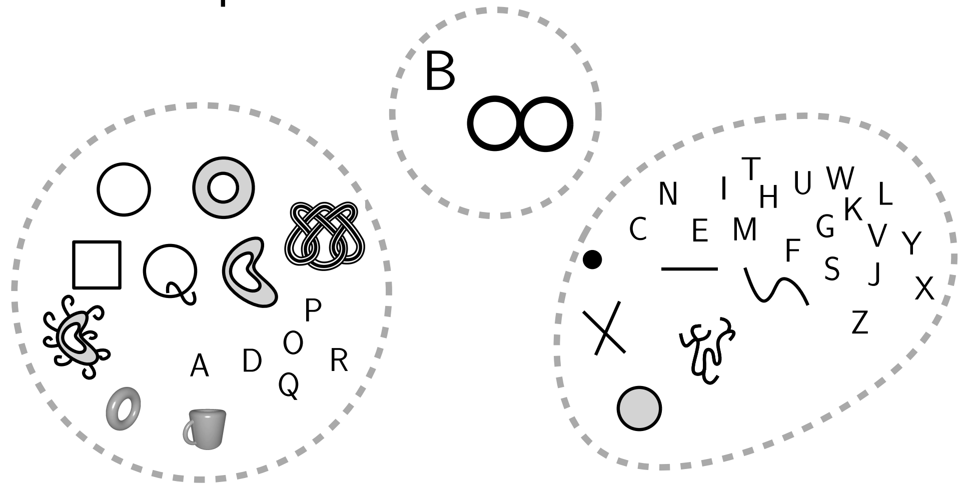
$$\begin{tikzcd} \mathbb{Z}/2\mathbb{Z} \arrow[r] \arrow[rr, bend left, "\mathrm{id}"] & 0 \arrow[r] & \mathbb{Z}/2\mathbb{Z}. \end{tikzcd}$$

Conclusão temporária

Alguns conjuntos de dados contêm topologia



Os espaços topológicos podem ser classificados em **classes de equivalência** de homotopia



Invariantes de classes de homotopia permitem descrever e compreender os espaços topológicos

- Número de componentes conexas
- Característica de Euler χ
- Número de Betti $\beta_0, \beta_1, \beta_2, \ldots$
- Grupos de homologia $H_0, H_1, H_2, \ldots$

I - Comparar espaços topológicos

1 - Equivalência de homeomorfia

2 - Equivalência de homotopia

II - Invariantes topológicos

1 - Número de componentes conexas

2 - Característica de Euler

3 - Números de Betti

4 - Homologia (simplicial sobre $\mathbb{Z}/2\mathbb{Z}$)

III - Inferência topológica

1 - Estimação do parâmetro t

2 - Nervos

IV - Homologia persistente

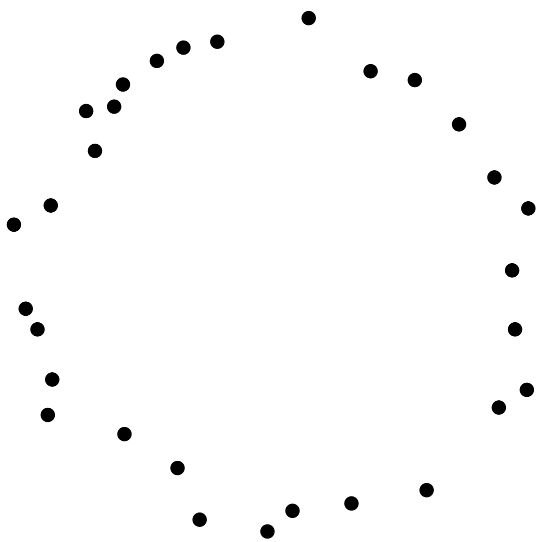
1 - Módulos de persistência

2 - Decomposição

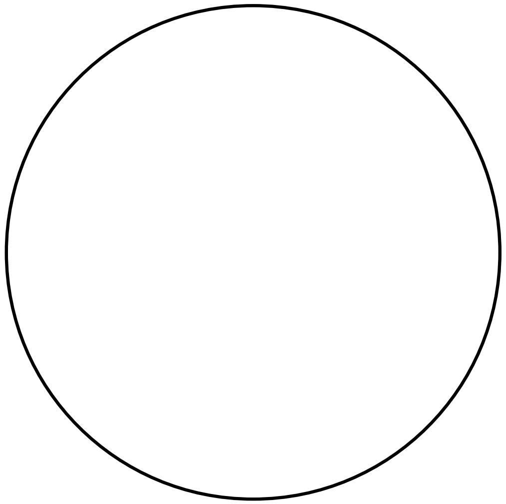
3 - Estabilidade

V - Aplicações

Seja $\mathcal{M} \subset \mathbb{R}^n$ um subconjunto.
Suponha que tenhamos uma amostra finita $X \subset \mathcal{M}$.
Problema: Estimar os grupos de homologia de $\mathcal{M}$ a partir de X.



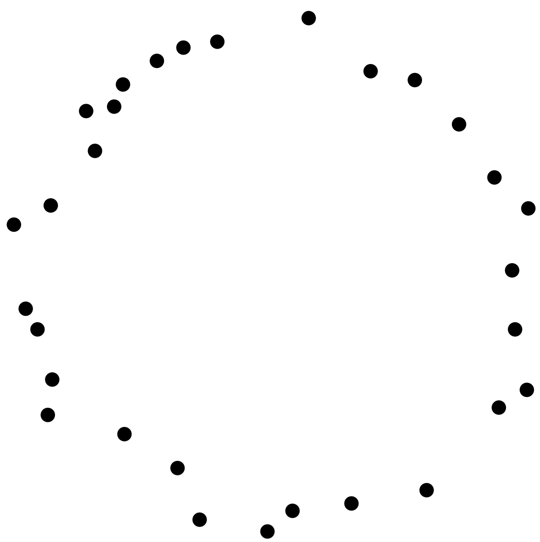
X

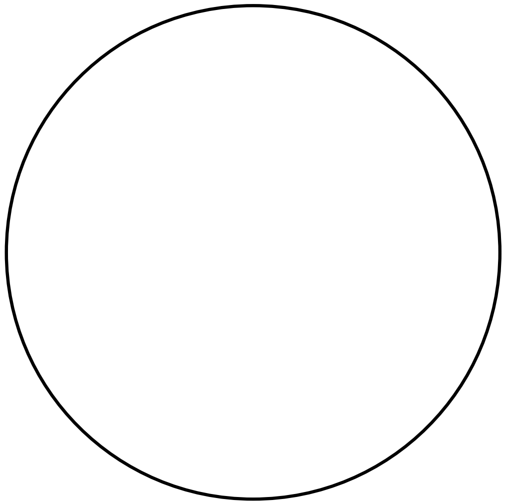


$\mathcal{M}$

O problema da inferência homológica

Seja $\mathcal{M} \subset \mathbb{R}^n$ um subconjunto.
Suponha que tenhamos uma amostra finita $X \subset \mathcal{M}$.
Problema: Estimar os grupos de homologia de $\mathcal{M}$ a partir de X.

X 

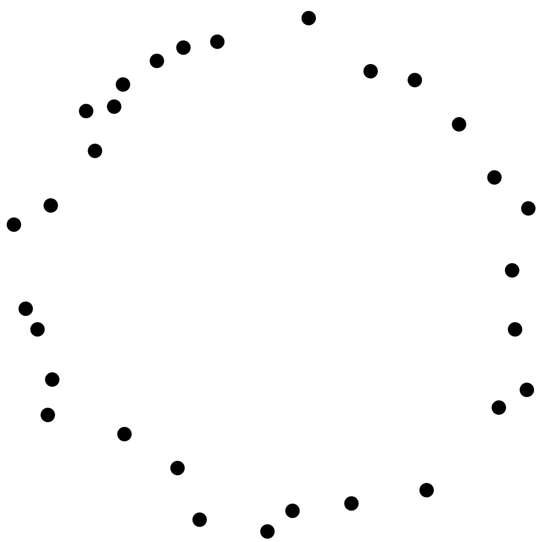
 $\mathcal{M}$

Não podemos usar X diretamente. Sua homologia é decepcionante:

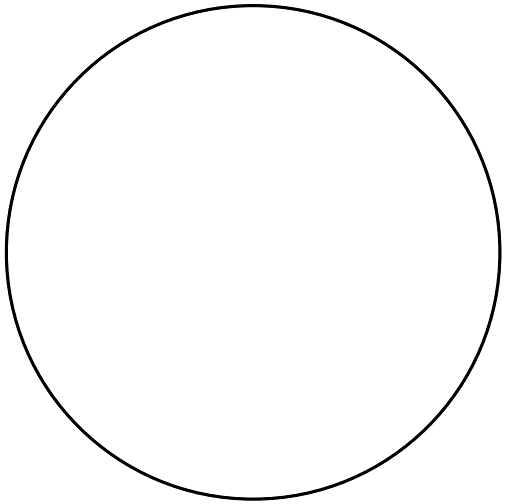
$$\beta_0(X) = 30 \quad \text{e} \quad \beta_i(X) = 0 \ \text{ para } \ i \geq 1$$

número de componentes conexas
$=$ número de pontos de X

Seja $\mathcal{M} \subset \mathbb{R}^n$ um subconjunto.
Suponha que tenhamos uma amostra finita $X \subset \mathcal{M}$.
Problema: Estimar os grupos de homologia de $\mathcal{M}$ a partir de X.



X



$\mathcal{M}$

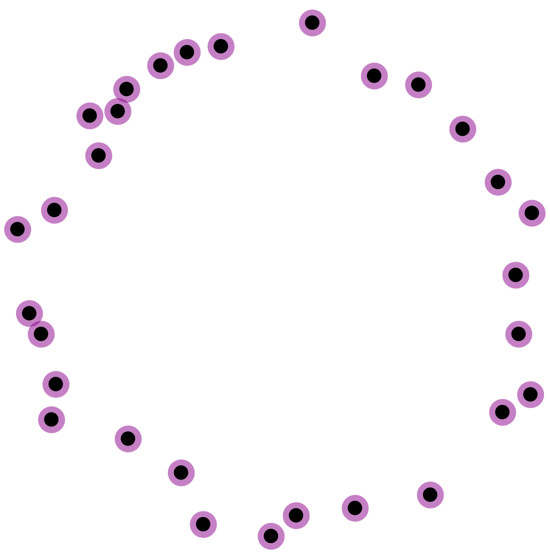
Não podemos usar X diretamente.

Idéia: Espessar X.

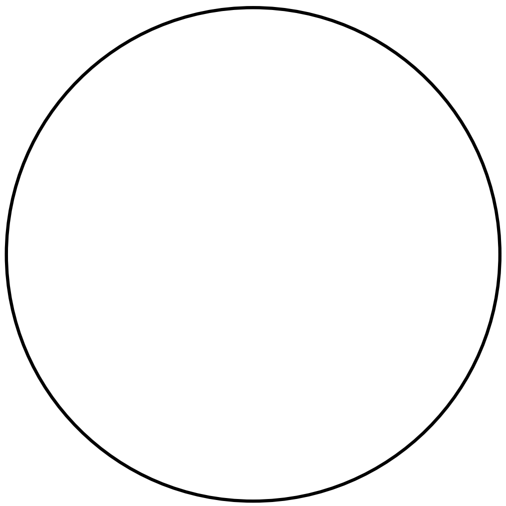
Definição: Por todo $t \geq 0$, o t-**espessamento** do conjunto X, denotado X^t, é o conjunto de pontos do espaço ambiente a uma distância máxima de t de X:

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

Seja $\mathcal{M} \subset \mathbb{R}^n$ um subconjunto.
Suponha que tenhamos uma amostra finita $X \subset \mathcal{M}$.
Problema: Estimar os grupos de homologia de $\mathcal{M}$ a partir de X.



$X^{0.05}$



$\mathcal{M}$

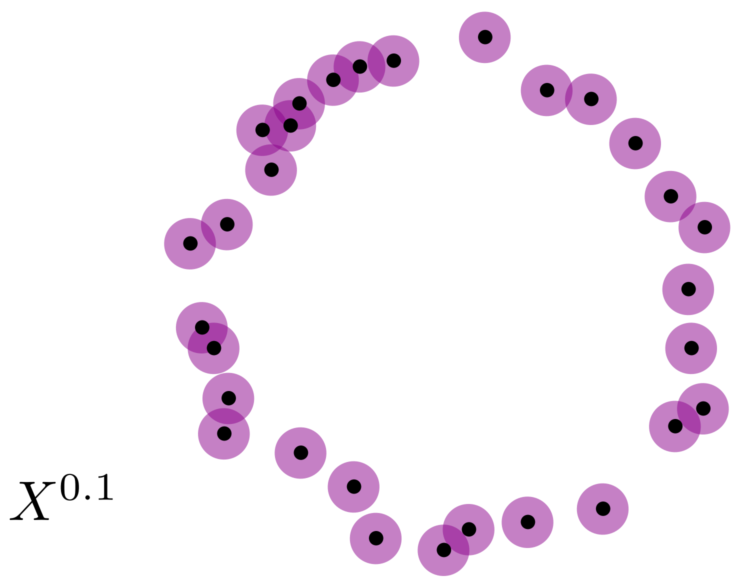
Não podemos usar X diretamente.

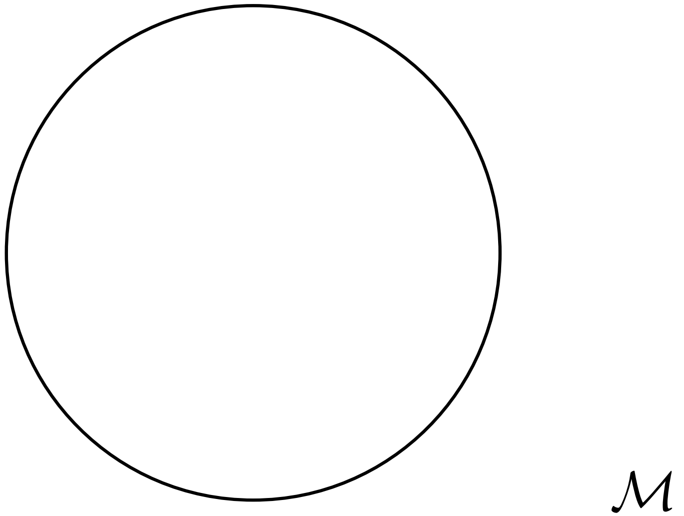
Idéia: Espessar X.

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

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

Seja $\mathcal{M} \subset \mathbb{R}^n$ um subconjunto.
Suponha que tenhamos uma amostra finita $X \subset \mathcal{M}$.
Problema: Estimar os grupos de homologia de $\mathcal{M}$ a partir de X.





Não podemos usar X diretamente.

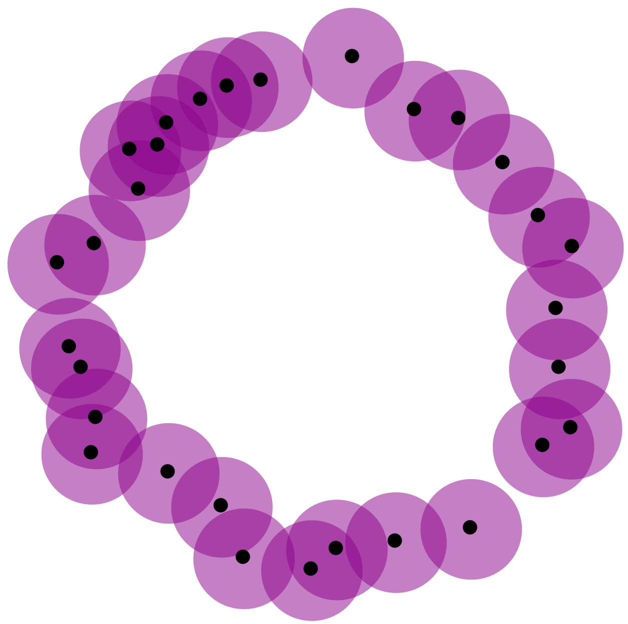
Idéia: Espessar X.

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

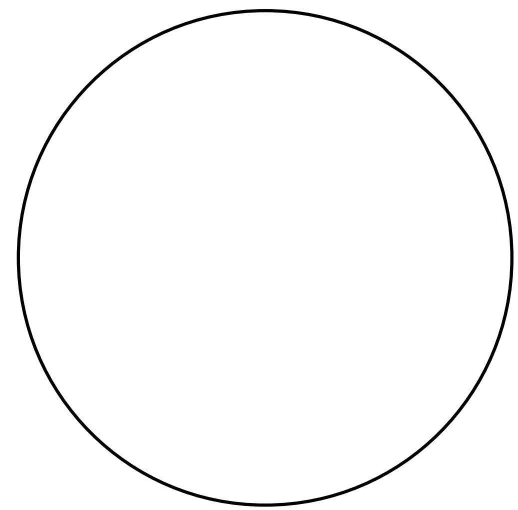
$$X^t = \{y \in \mathbb{R}^n, \exists x \in X, \|x - y\| \leq t\}.$$

O problema da inferência homológica

Seja $\mathcal{M} \subset \mathbb{R}^n$ um subconjunto.
Suponha que tenhamos uma amostra finita $X \subset \mathcal{M}$.
Problema: Estimar os grupos de homologia de $\mathcal{M}$ a partir de X.



$X^{0.2}$



$\mathcal{M}$

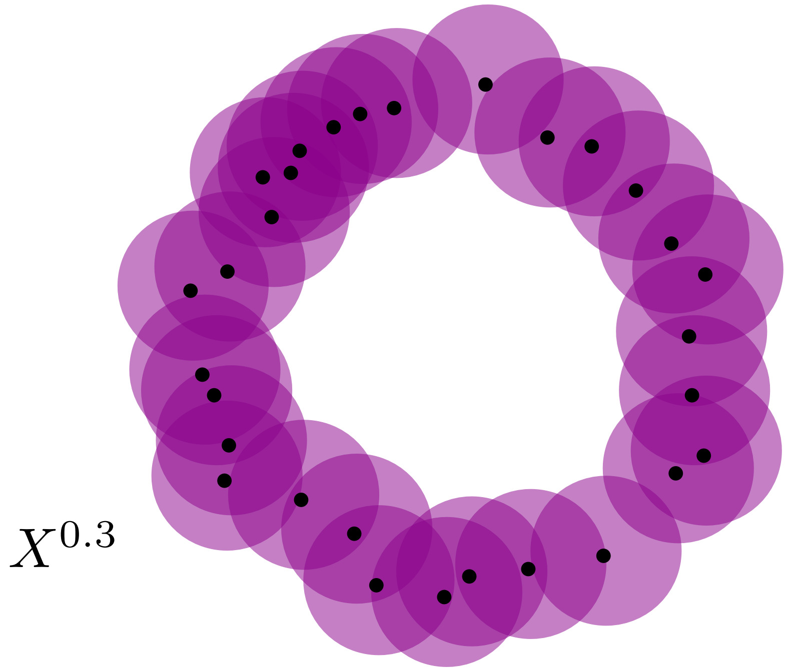
Não podemos usar X diretamente.

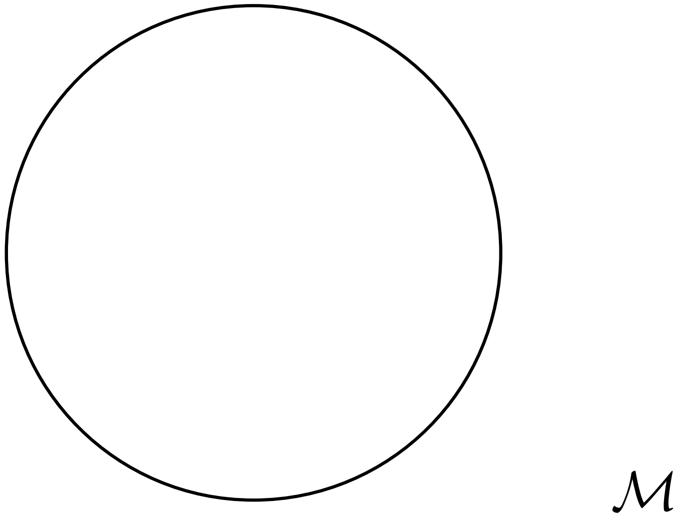
Idéia: Espessar X.

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

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

Seja $\mathcal{M} \subset \mathbb{R}^n$ um subconjunto.
Suponha que tenhamos uma amostra finita $X \subset \mathcal{M}$.
Problema: Estimar os grupos de homologia de $\mathcal{M}$ a partir de X.





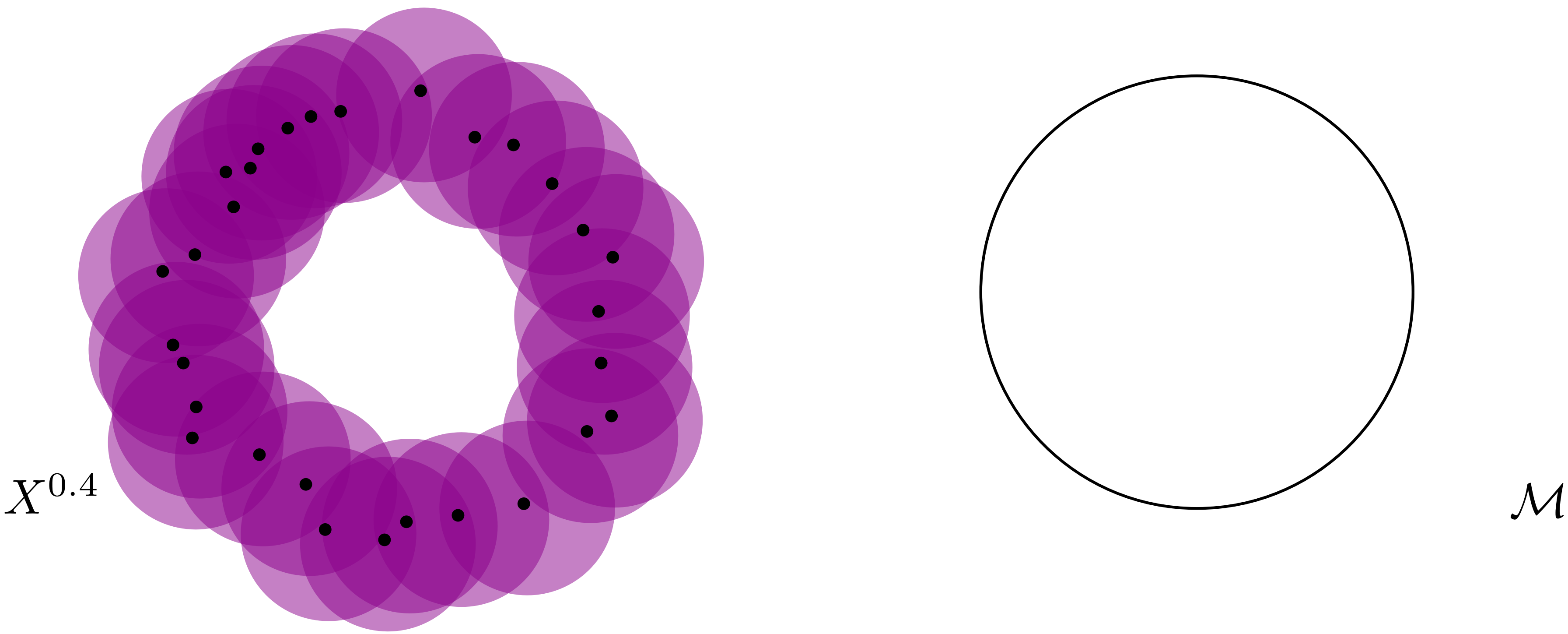
Não podemos usar X diretamente.

Idéia: Espessar X.

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

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

Seja $\mathcal{M} \subset \mathbb{R}^n$ um subconjunto.
Suponha que tenhamos uma amostra finita $X \subset \mathcal{M}$.
Problema: Estimar os grupos de homologia de $\mathcal{M}$ a partir de X.



Não podemos usar X diretamente.

Idéia: Espessar X.

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

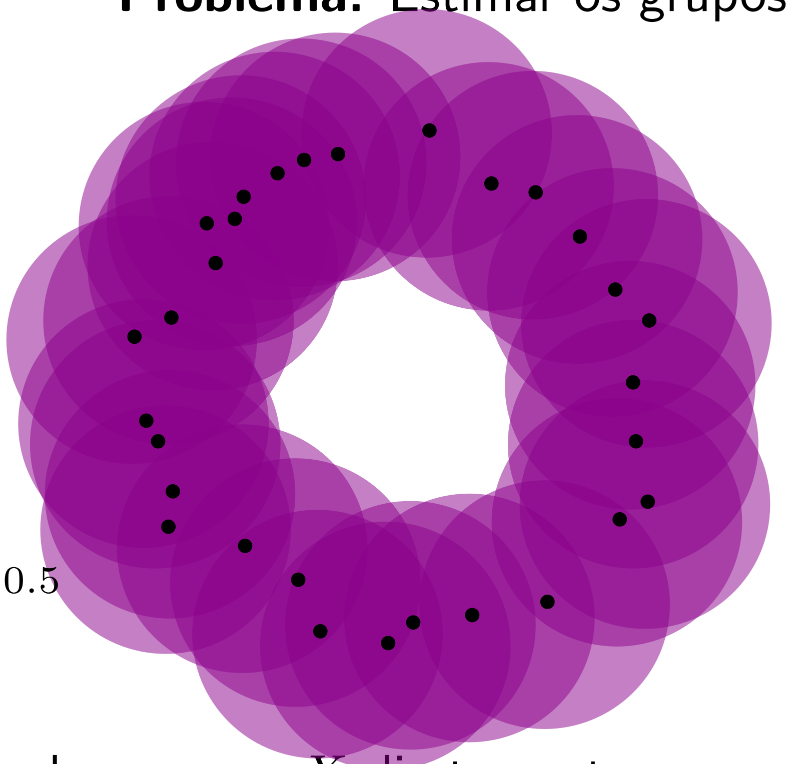
$$X^t = \{y \in \mathbb{R}^n, \exists x \in X, \|x - y\| \leq t\} .$$

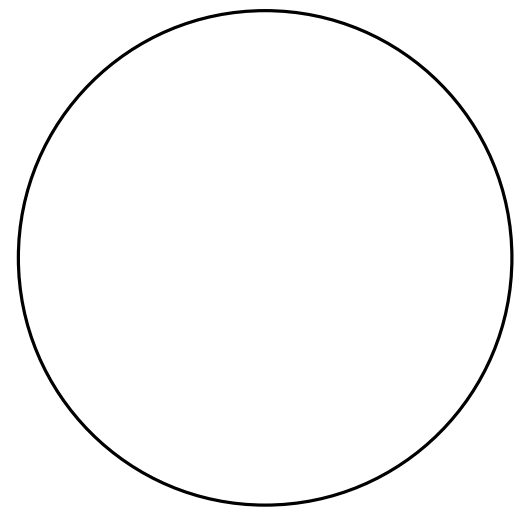
O problema da inferência homológica

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

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

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





Não podemos usar X diretamente.

Idéia: Espessar X.

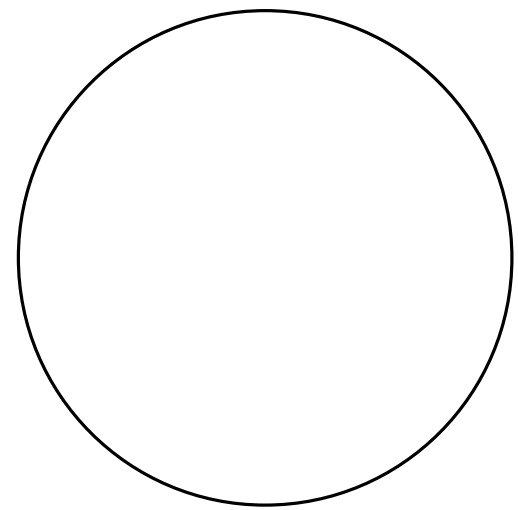
Definição: Por todo $t \geq 0$, o t-**espessamento** do conjunto X, denotado X^t, é o conjunto de pontos do espaço ambiente a uma distância máxima de t de X:

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

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

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

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



$\mathcal{M}$

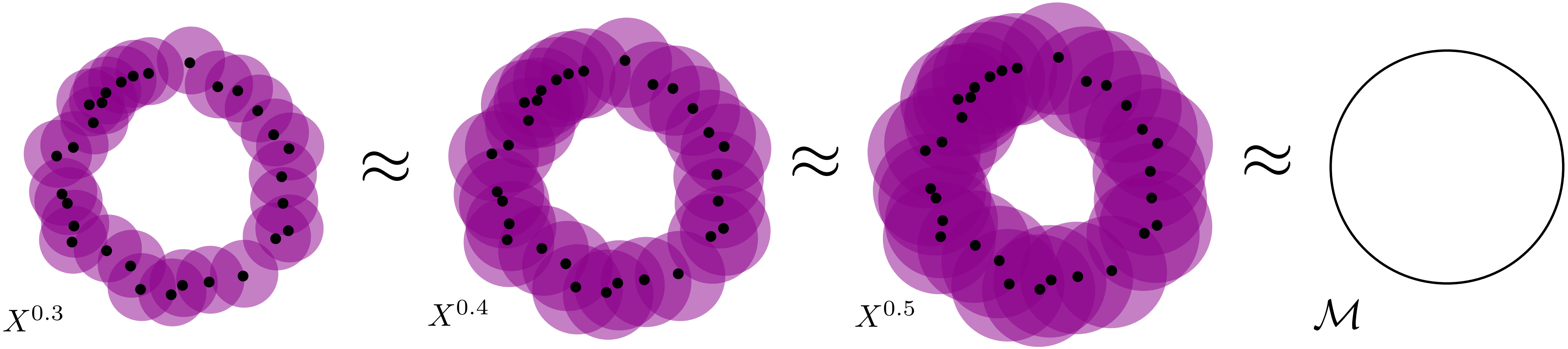
Não podemos usar X diretamente.

Idéia: Espessar X.

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

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

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

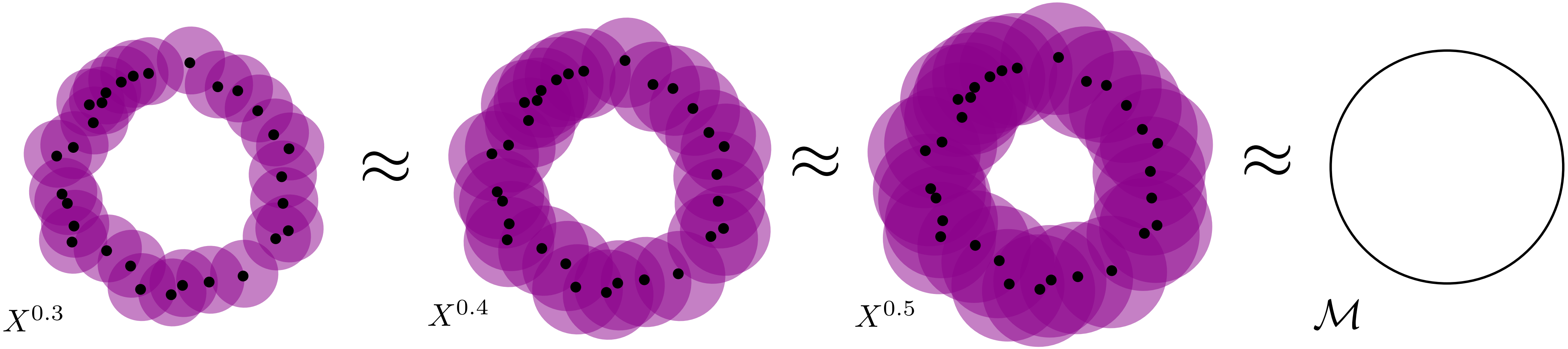


Portanto, podemos encontrar a homologia de $\mathcal{M}$:

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

O problema da inferência homológica

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



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

$$\begin{gathered}\beta_0(\mathcal{M}) = \beta_0(X^{0.3}) \\ \beta_1(\mathcal{M}) = \beta_1(X^{0.3}) \\ \beta_2(\mathcal{M}) = \beta_2(X^{0.3}) \\ \cdots\end{gathered}$$

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

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

I - Comparar espaços topológicos

1 - Equivalência de homeomorfia

2 - Equivalência de homotopia

II - Invariantes topológicos

1 - Número de componentes conexas

2 - Característica de Euler

3 - Números de Betti

4 - Homologia (simplicial sobre $\mathbb{Z}/2\mathbb{Z}$)

III - Inferência topológica

1 - Estimação do parâmetro t

2 - Nervos

IV - Homologia persistente

1 - Módulos de persistência

2 - Decomposição

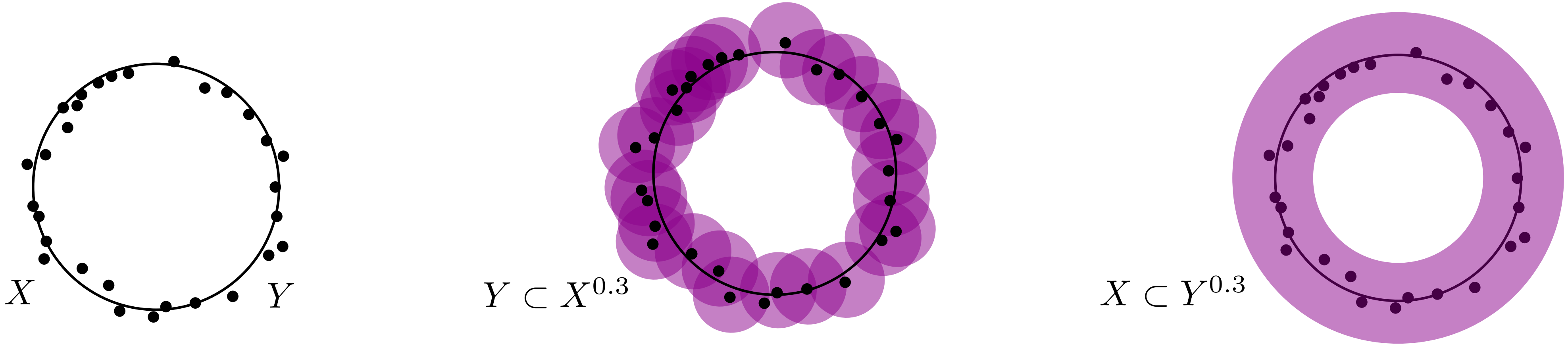
3 - Estabilidade

V - Aplicações

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

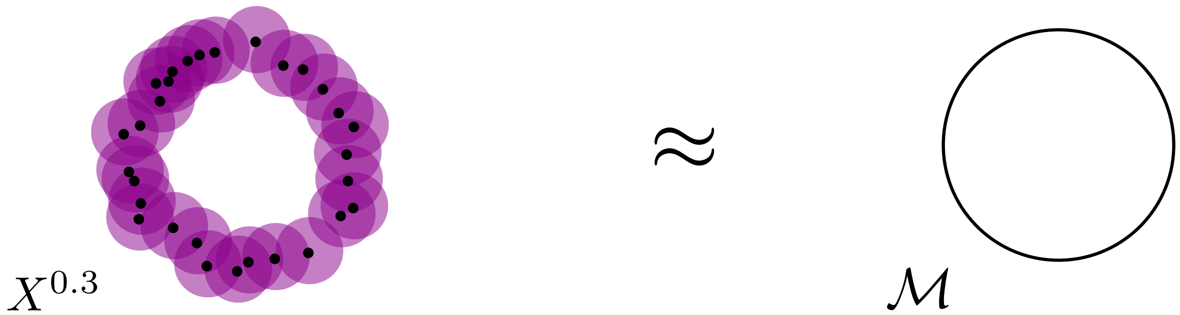
$$\mathrm{d_H}(X, Y) = \max\left\{ \sup_{y \in Y} \inf_{x \in X} \|x - y\|,\ \sup_{x \in X} \inf_{y \in Y} \|x - y\| \right\}.$$

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



Seleção do parâmetro t

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

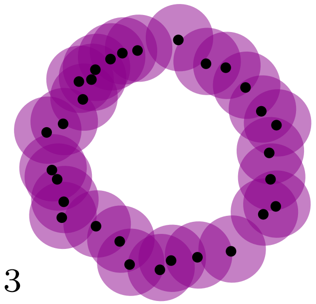


Teorema (Frédéric Chazal, David Cohen-Steiner et André Lieutier, 2009):
Sejam X e $\mathcal{M}$ dois subconjuntos de $\mathbb{R}^n$. Suponha que $\mathcal{M}$ tenha um *reach* estritamente positivo, e que $\mathrm{d_H}\left(X, \mathcal{M}\right) \leq \frac{1}{17}\text{reach}(\mathcal{M})$.
Então X^t e $\mathcal{M}$ são homotopicamente equivalentes, desde que

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

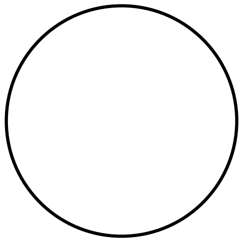
Seleção do parâmetro t

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



$X^{0.3}$

$\approx$



$\mathcal{M}$

Teorema (Frédéric Chazal, David Cohen-Steiner et André Lieutier, 2009):
Sejam X e $\mathcal{M}$ dois subconjuntos de $\mathbb{R}^n$. Suponha que $\mathcal{M}$ tenha um *reach* estritamente positivo, e que $\mathrm{d_H}(X, \mathcal{M}) \leq \frac{1}{17}\mathrm{reach}(\mathcal{M})$.
Então X^t e $\mathcal{M}$ são homotopicamente equivalentes, desde que

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

Teorema (Partha Niyogi, Stephen Smale et Shmuel Weinberger, 2008):
Sejam X e $\mathcal{M}$ dois subconjuntos de $\mathbb{R}^n$, com $\mathcal{M}$ uma subvariedade, e X um subconjunto finito de $\mathcal{M}$. Suponha que $\mathcal{M}$ tenha um *reach* estritamente positivo.
Então X^t e $\mathcal{M}$ são homotopicamente equivalentes, desde que

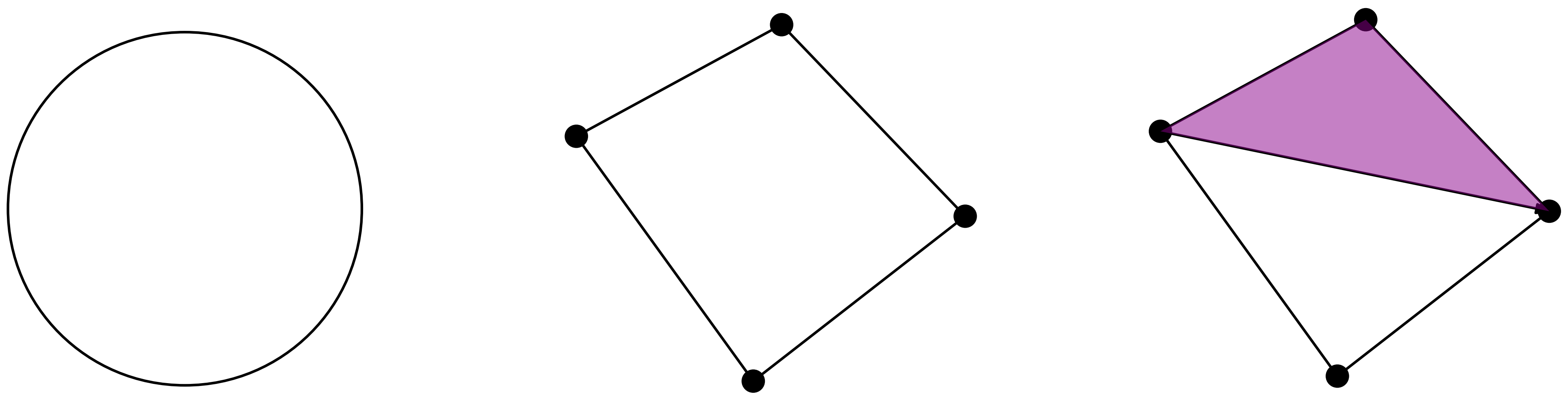
$$t \in \left[2\mathrm{d_H}(X, \mathcal{M}), \sqrt{\frac{3}{5}}\mathrm{reach}(\mathcal{M})\right).$$

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

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

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

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

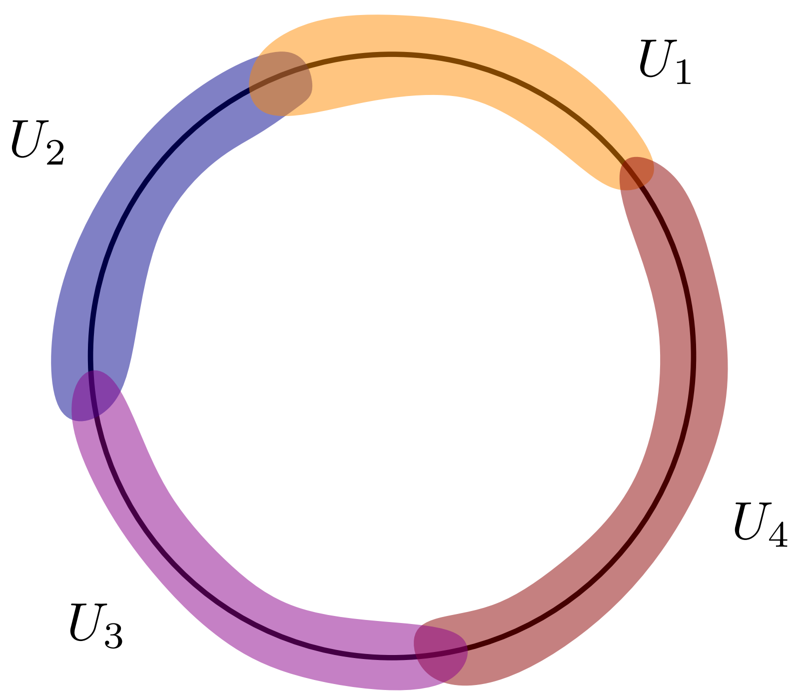


Nervos

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

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

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

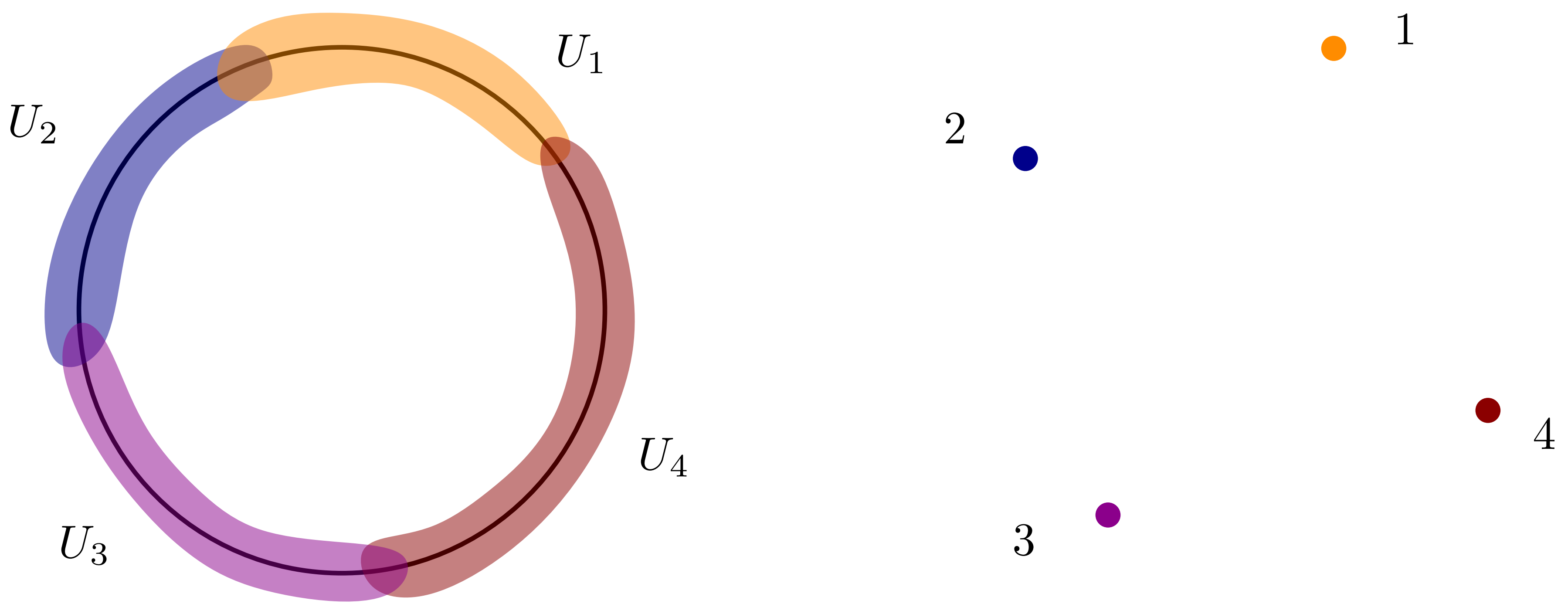


Nervos

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

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

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

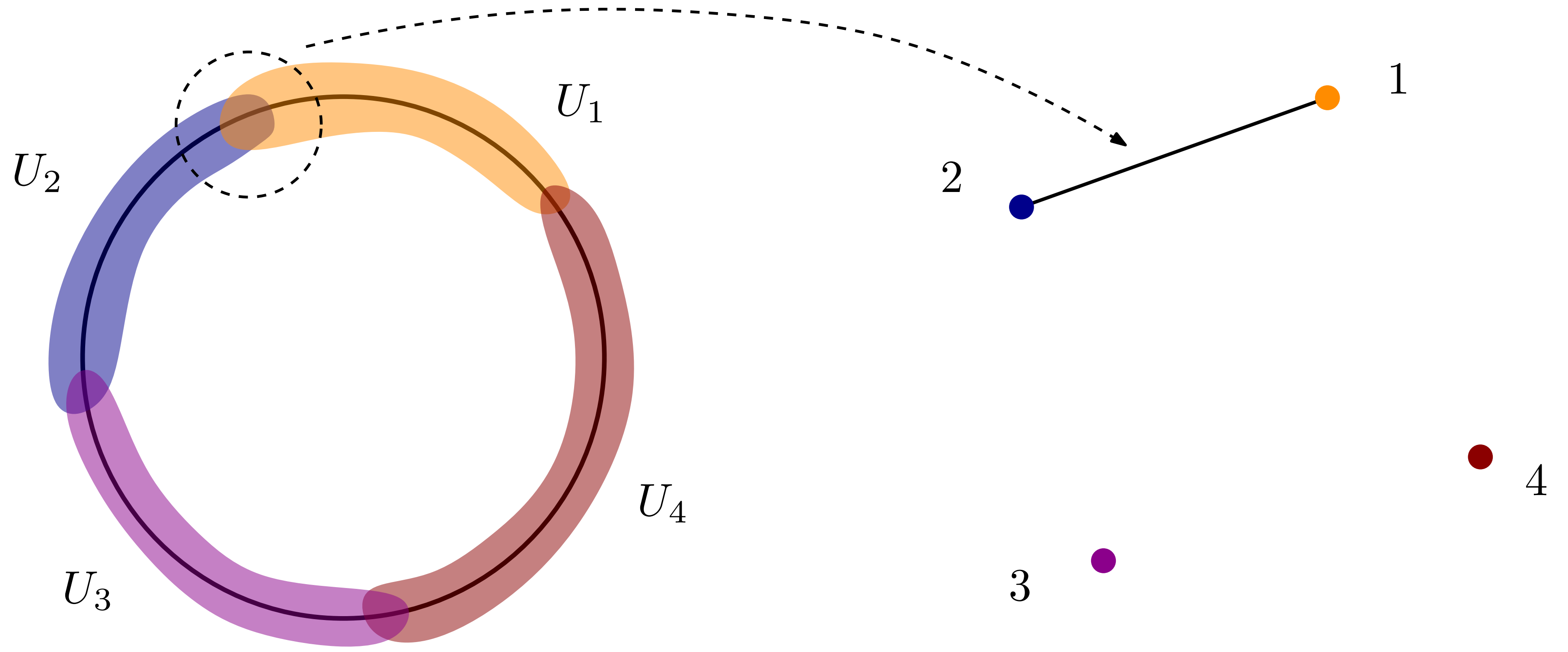


Nervos

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

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

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

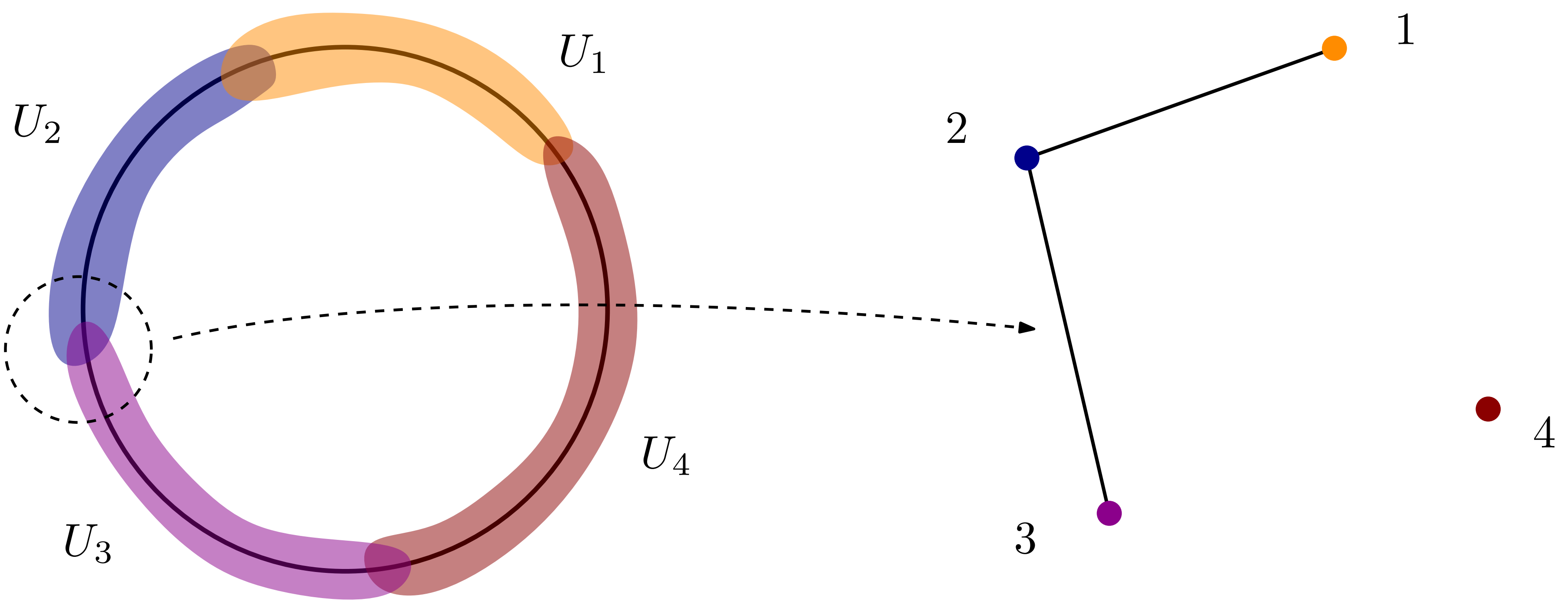


Nervos

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

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

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

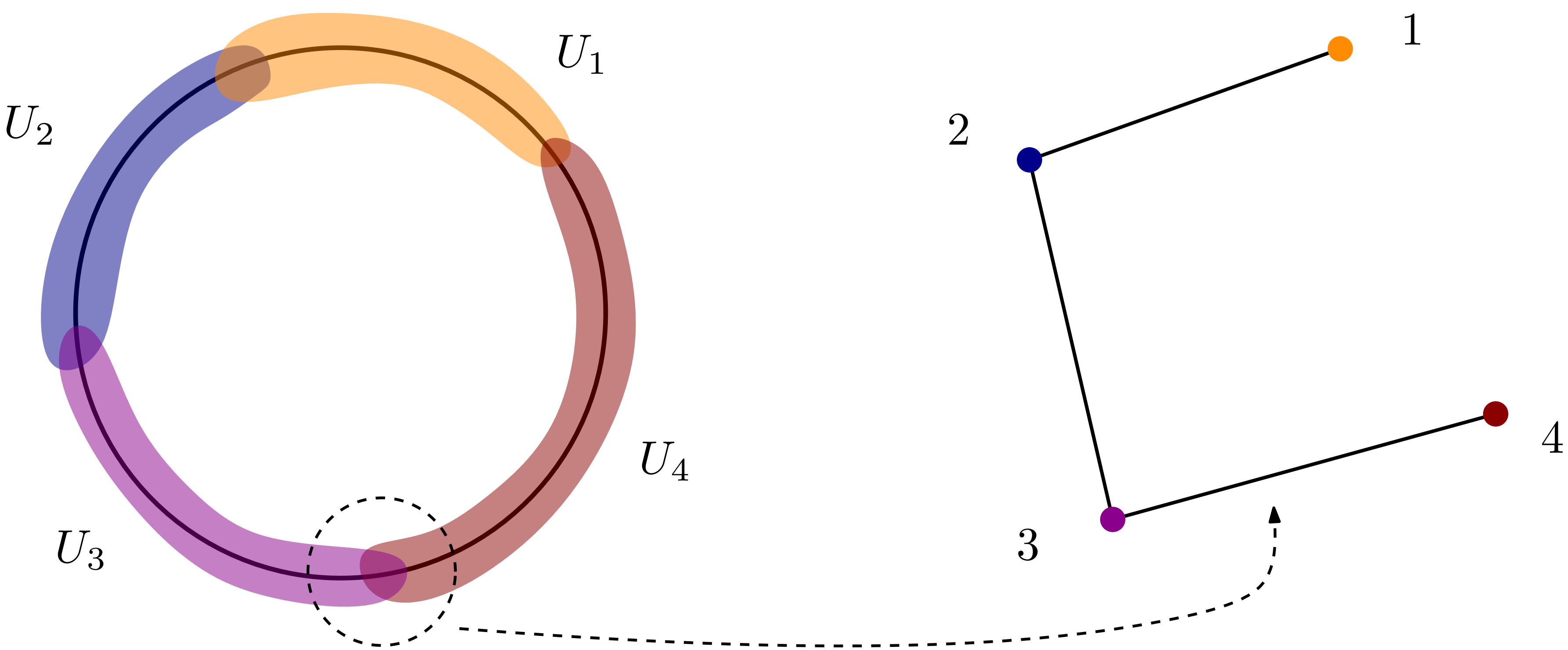


Nervos

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

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

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

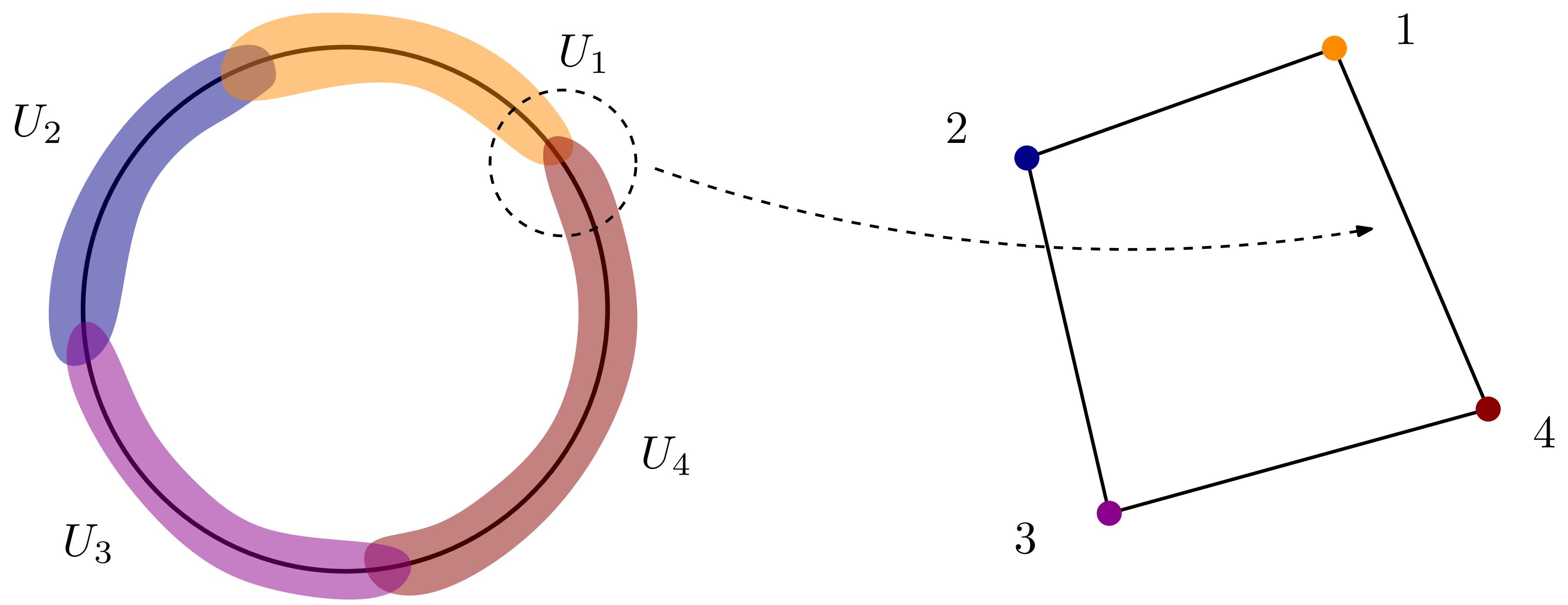


Nervos

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

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

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

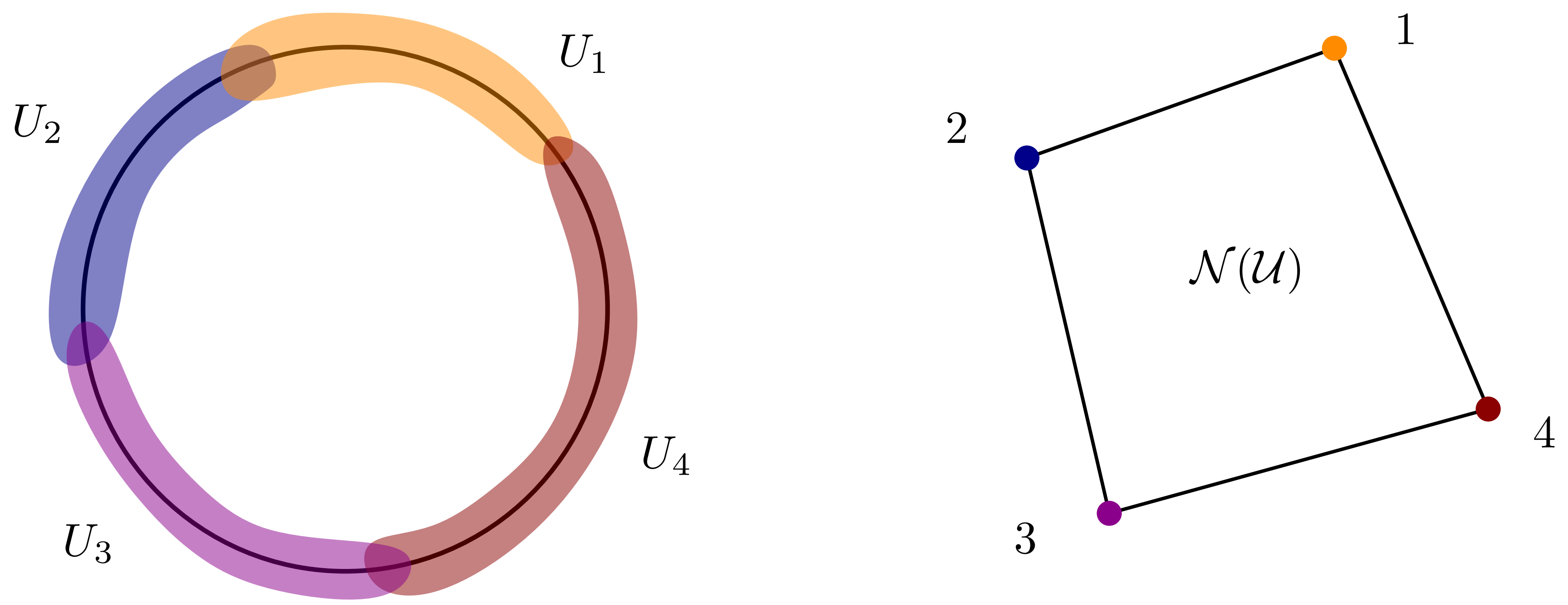


Nervos

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

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

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

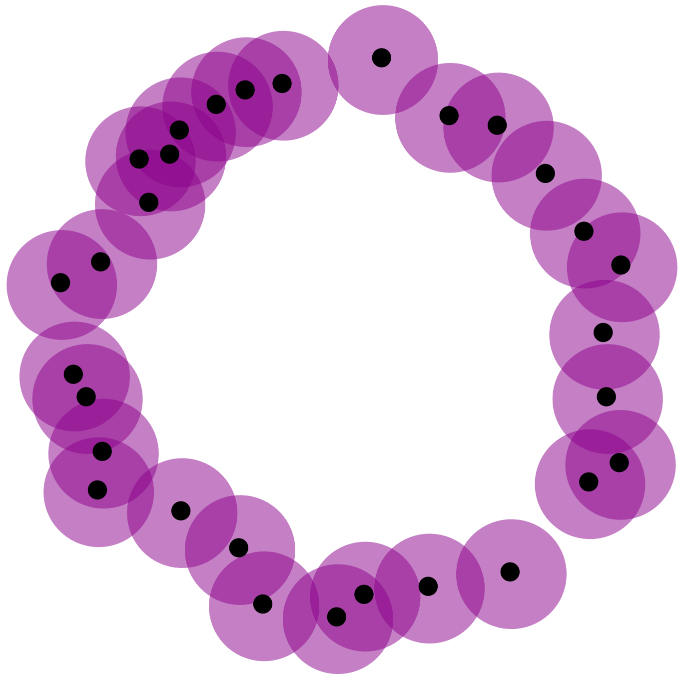


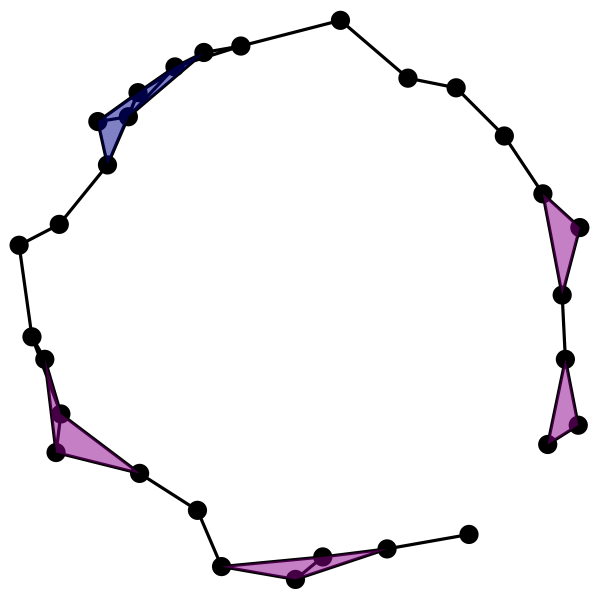
Nervos

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

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

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





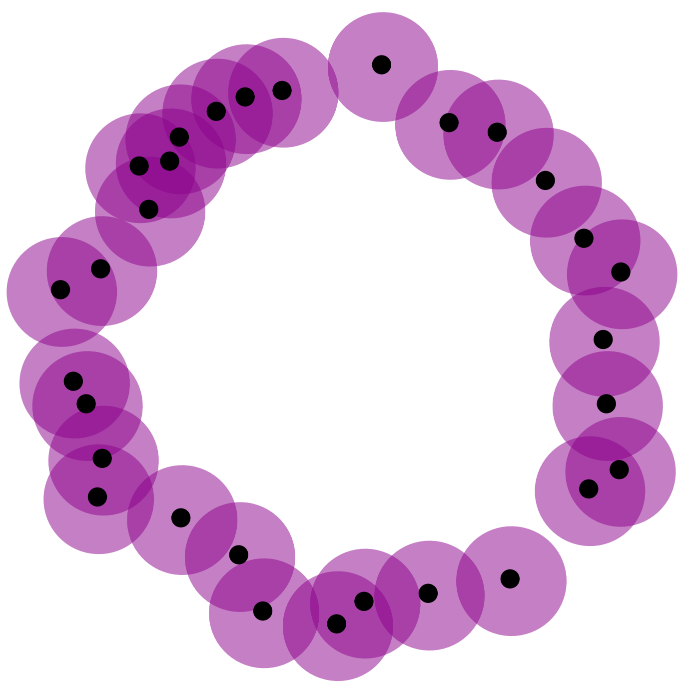
$X^{0.2} = \bigcup_{x \in X} \overline{\mathcal{B}}(x, 0.2)$ é coberto por $\mathcal{U} = \left\{ \overline{\mathcal{B}}(x, 0.2), x \in X \right\}$

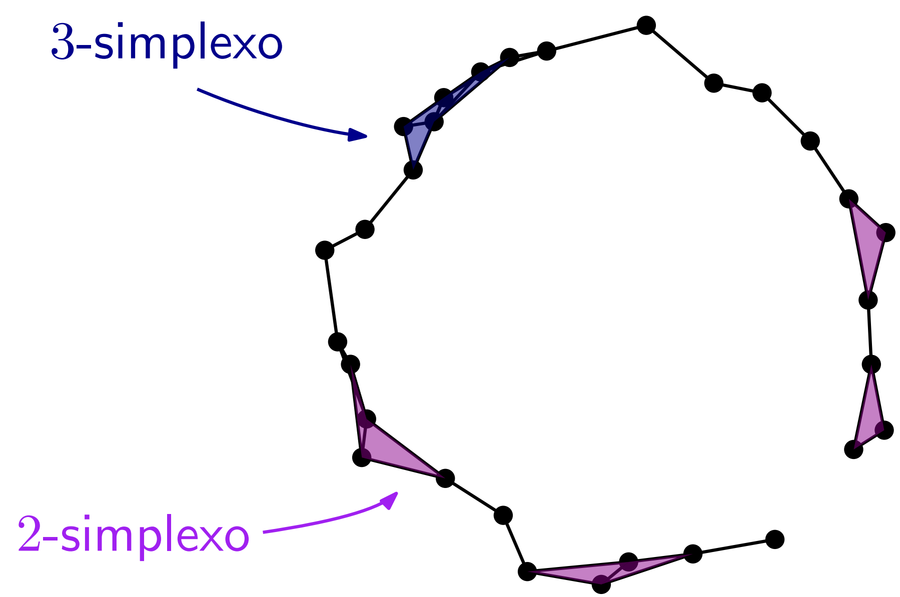
Nervos

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

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

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





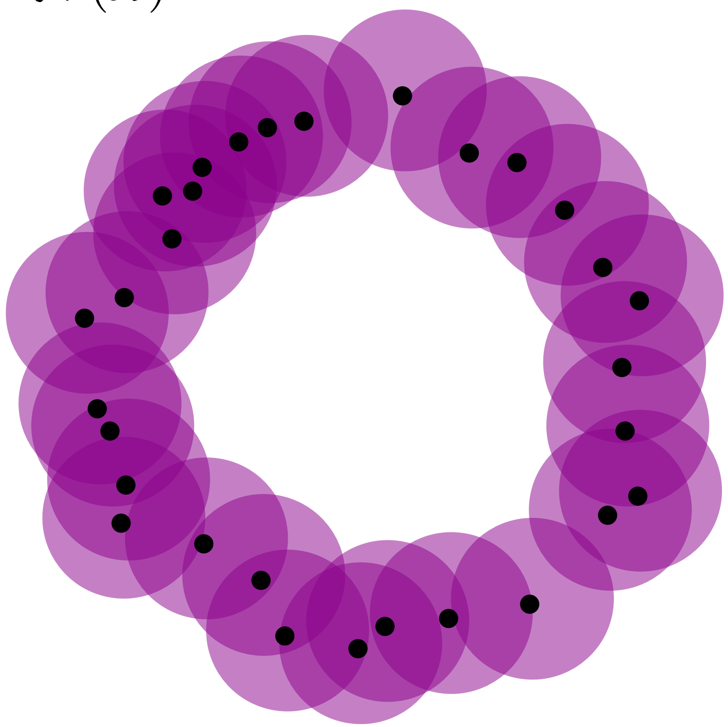
$X^{0.2} = \bigcup_{x \in X} \overline{\mathcal{B}}(x, 0.2)$ é coberto por $\mathcal{U} = \left\{\overline{\mathcal{B}}(x, 0.2), x \in X\right\}$

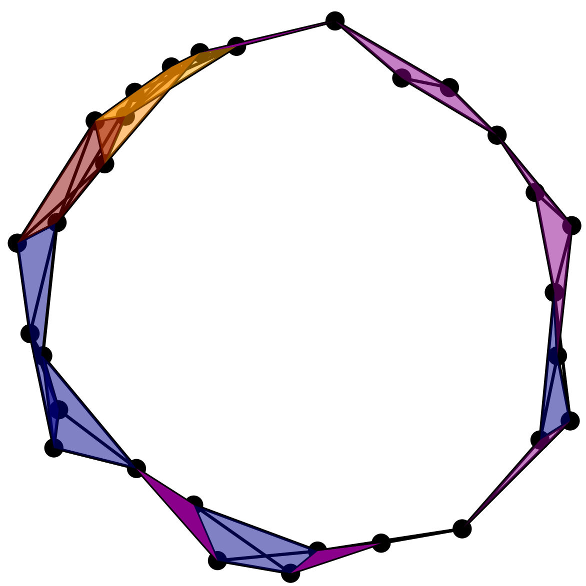
Nervos

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

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

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





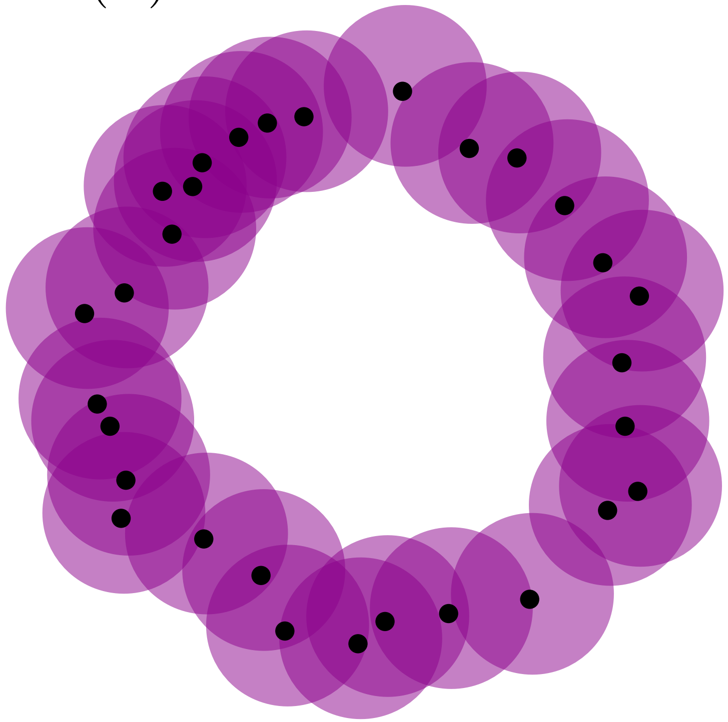
$X^{0.3} = \bigcup_{x \in X} \overline{\mathcal{B}}(x, 0.3)$ é coberto por $\mathcal{U} = \{\overline{\mathcal{B}}(x, 0.3), x \in X\}$

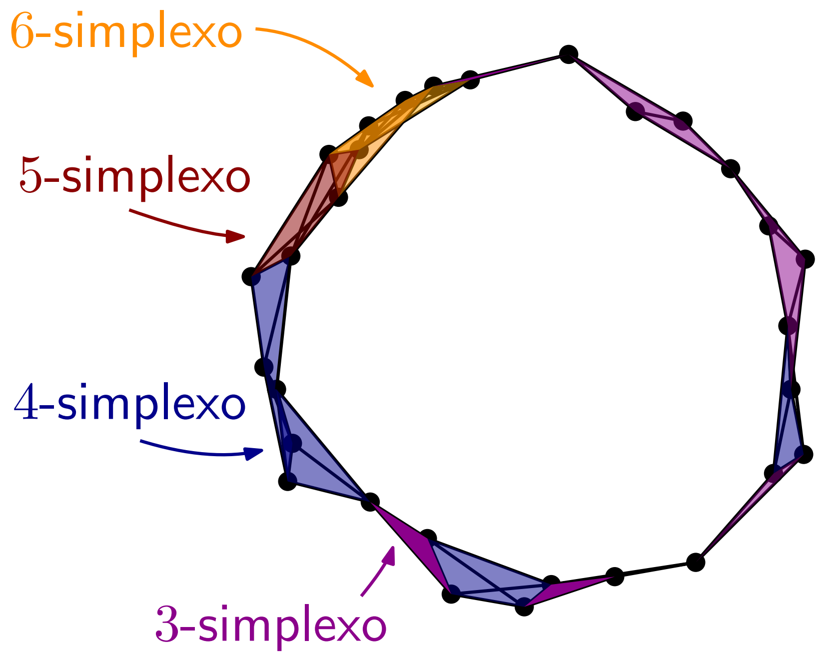
Nervos

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

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

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





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

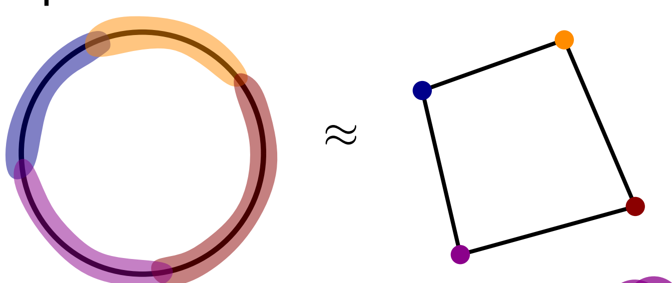
Nervos

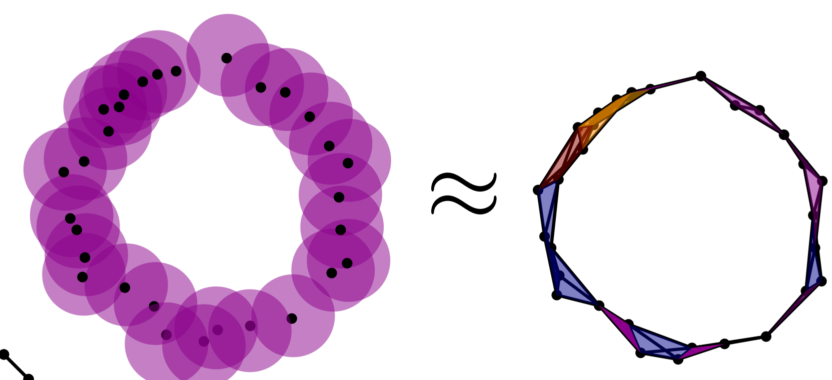
Definição: Seja X um espaço topológico, e $\mathcal{U} = \{U_i\}_{1 \leq i \leq N}$ uma *cobertura* de X, isto é, uma coleção de subconjuntos $U_i \subset X$ tal que

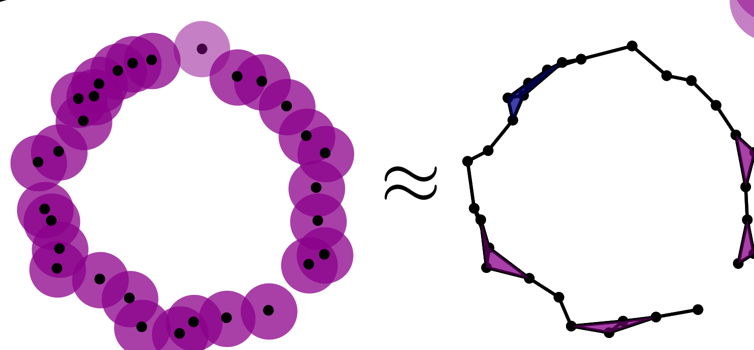
$$\bigcup_{1 \leq i \leq N} U_i = X.$$

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

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







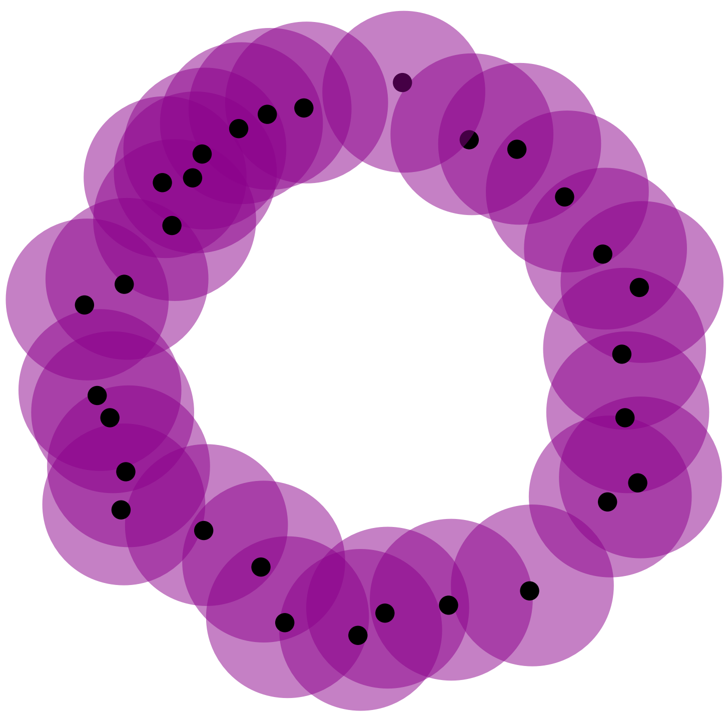
Complexo de Čech

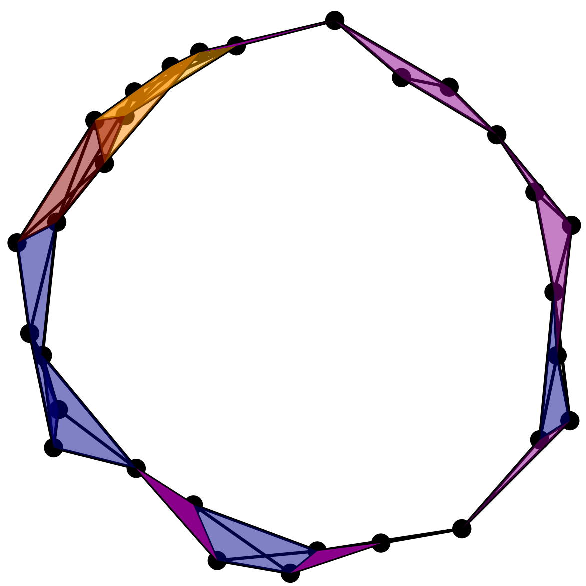
Seja X un subconjunto finito de $\mathbb{R}^n$, et $t \geq 0$. Seja a coleção

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

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

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



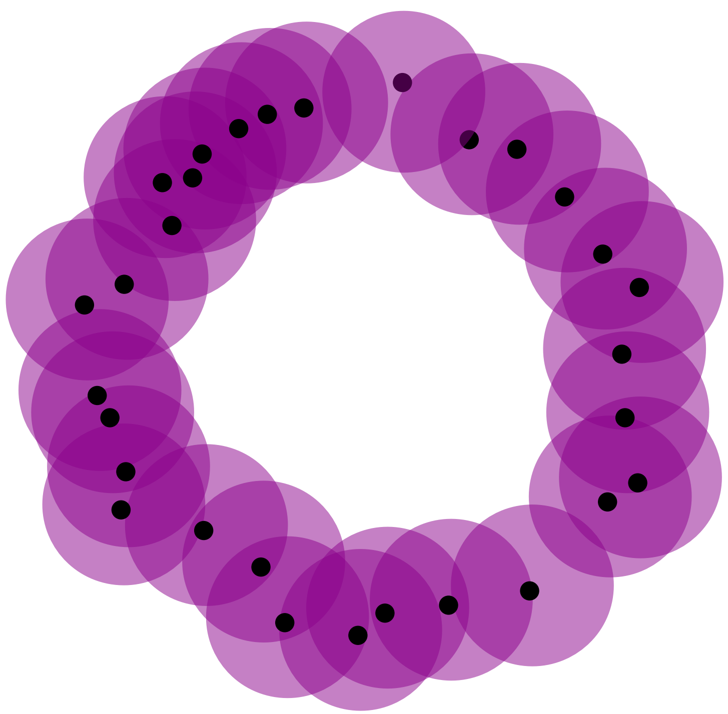


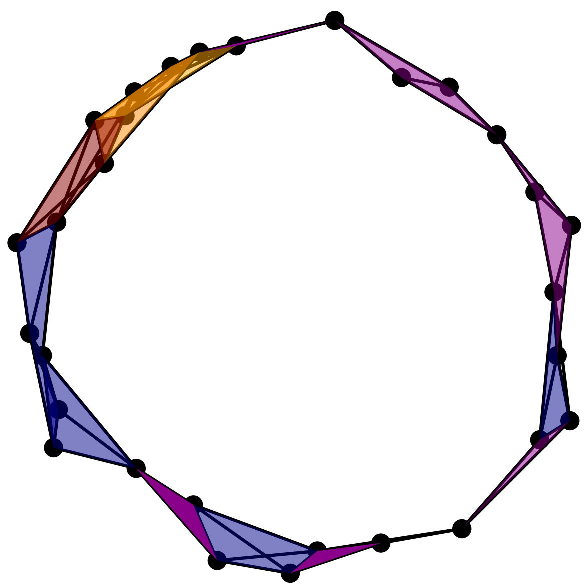
Seja X un subconjunto finito de $\mathbb{R}^n$, et $t \geq 0$. Seja a coleção

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

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

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





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

O problema da escala

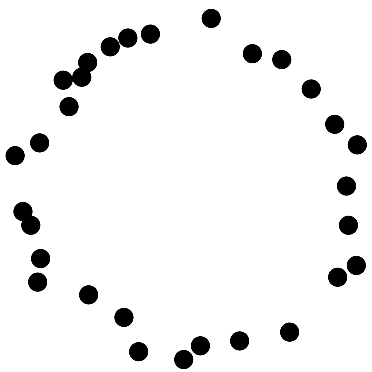
Questão 1 : Como selecionar um t tal que $X^t \approx \mathcal{M}$?

Teorema (Frédéric Chazal, David Cohen-Steiner et André Lieutier, 2009):
Sejam X e $\mathcal{M}$ dois subconjuntos de $\mathbb{R}^n$. Suponha que $\mathcal{M}$ tenha um *reach* estritamente positivo, e que $\mathrm{d_H}(X,\mathcal{M}) \leq \frac{1}{17}\mathrm{reach}(\mathcal{M})$.
Então X^t e $\mathcal{M}$ são homotopicamente equivalentes, desde que

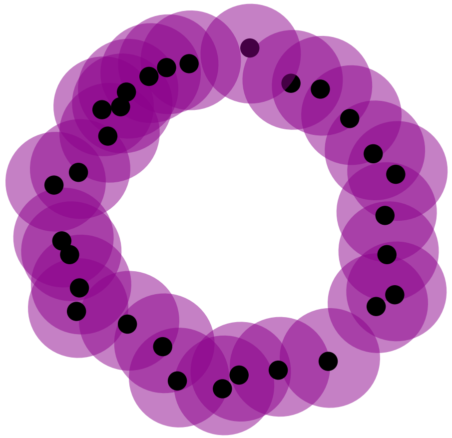
$$t \in \left[4\mathrm{d_H}(X,\mathcal{M}), \mathrm{reach}(\mathcal{M}) - 3\mathrm{d_H}(X,\mathcal{M})\right).$$

Teorema (Partha Niyogi, Stephen Smale et Shmuel Weinberger, 2008):
Sejam X e $\mathcal{M}$ dois subconjuntos de $\mathbb{R}^n$, com $\mathcal{M}$ uma subvariedade, e X um subconjunto finito de $\mathcal{M}$. Suponha que $\mathcal{M}$ tenha um *reach* estritamente positivo.
Então X^t e $\mathcal{M}$ são homotopicamente equivalentes, desde que

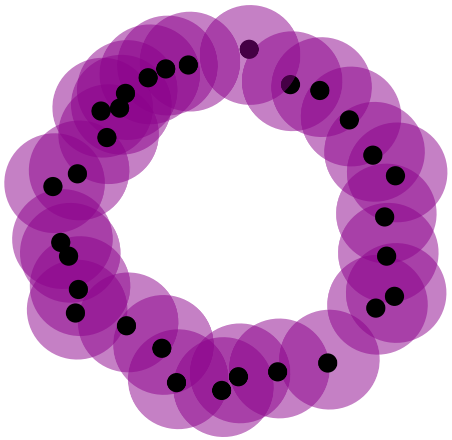
$$t \in \left[2\mathrm{d_H}(X,\mathcal{M}), \sqrt{\frac{3}{5}}\mathrm{reach}(\mathcal{M})\right).$$



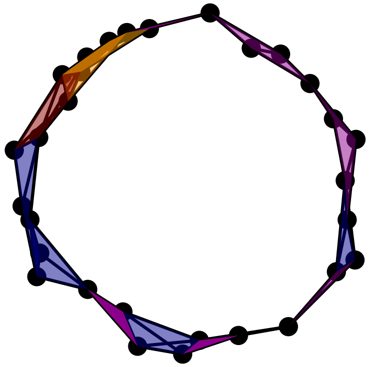
estimar t.



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



calcular o nervo



O problema da escala

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

Teorema (Frédéric Chazal, David Cohen-Steiner et André Lieutier, 2009):
Sejam X e $\mathcal{M}$ dois subconjuntos de $\mathbb{R}^n$. Suponha que $\mathcal{M}$ tenha um *reach* estritamente positivo, e que $\mathrm{d_H}(X, \mathcal{M}) \leq \frac{1}{17}\text{reach}(\mathcal{M})$.
Então X^t e $\mathcal{M}$ são homotopicamente equivalentes, desde que

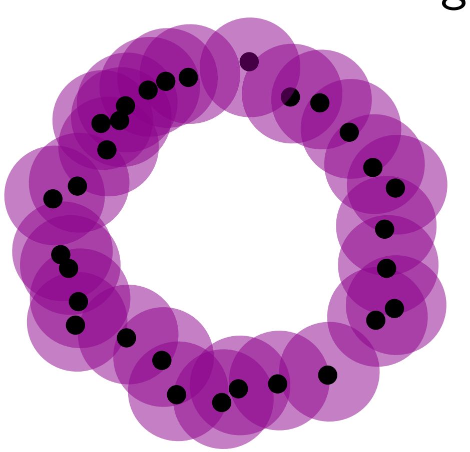
$$\boxed{t \in \left[4\mathrm{d_H}(X, \mathcal{M}), \text{reach}(\mathcal{M}) - 3\mathrm{d_H}(X, \mathcal{M})\right).}$$

Teorema (Partha Niyogi, Stephen Smale et Shmuel Weinberger, 2008):
Sejam X e $\mathcal{M}$ dois subconjuntos de $\mathbb{R}^n$, com $\mathcal{M}$ uma subvariedade, e X um subconjunto finito de $\mathcal{M}$. Suponha que $\mathcal{M}$ tenha um *reach* estritamente positivo.
Então X^t e $\mathcal{M}$ são homotopicamente equivalentes, desde que

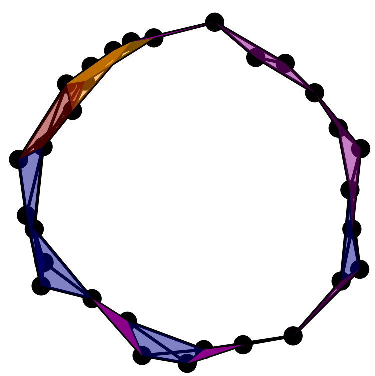
$$\boxed{t \in \left[2\mathrm{d_H}(X, \mathcal{M}), \sqrt{\frac{3}{5}}\text{reach}(\mathcal{M})\right).}$$

não sabemos estas quantidades!

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



calcular o nervo



O problema da escala

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

Teorema (Frédéric Chazal, David Cohen-Steiner et André Lieutier, 2009):
Sejam X e $\mathcal{M}$ dois subconjuntos de $\mathbb{R}^n$. Suponha que $\mathcal{M}$ tenha um *reach* estritamente positivo, e que $\mathrm{d_H}(X,\mathcal{M}) \leq \frac{1}{17}\mathrm{reach}(\mathcal{M})$.
Então X^t e $\mathcal{M}$ são homotopicamente equivalentes, desde que

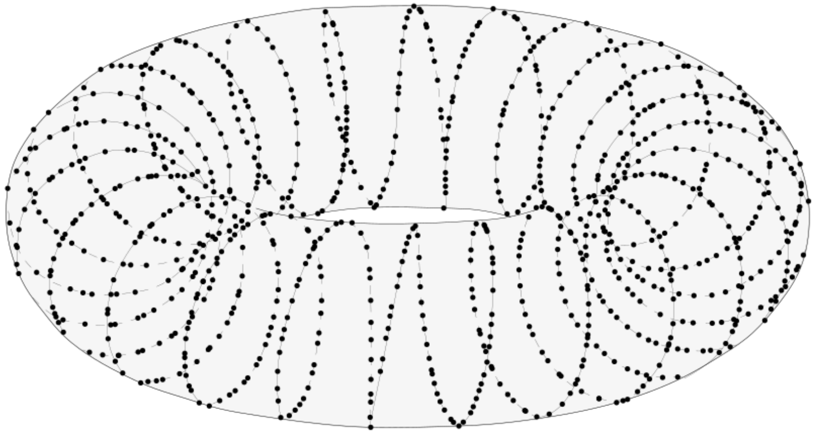
$$\boxed{t \in [4\mathrm{d_H}(X,\mathcal{M}), \mathrm{reach}(\mathcal{M}) - 3\mathrm{d_H}(X,\mathcal{M})).}$$

Teorema (Partha Niyogi, Stephen Smale et Shmuel Weinberger, 2008):
Sejam X e $\mathcal{M}$ dois subconjuntos de $\mathbb{R}^n$, com $\mathcal{M}$ uma subvariedade, e X um subconjunto finito de $\mathcal{M}$. Suponha que $\mathcal{M}$ tenha um *reach* estritamente positivo.
Então X^t e $\mathcal{M}$ são homotopicamente equivalentes, desde que

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

não sabemos estas quantidades!

Este objeto é de dimensão 1 ou 2?

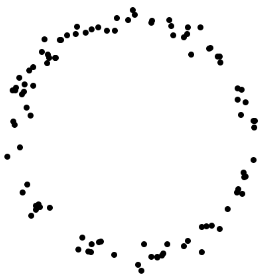
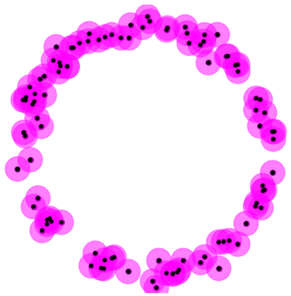
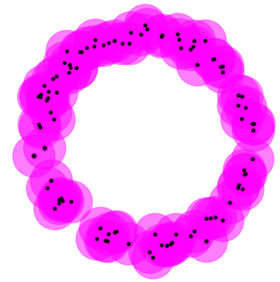
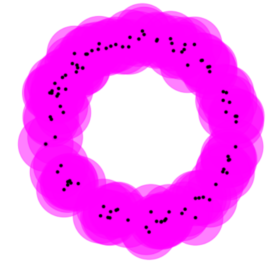
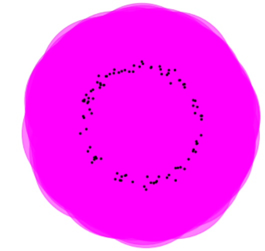


Idéia (análise multiescala): Em vez de escolher um valor por t, vamos escolher todos eles.

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

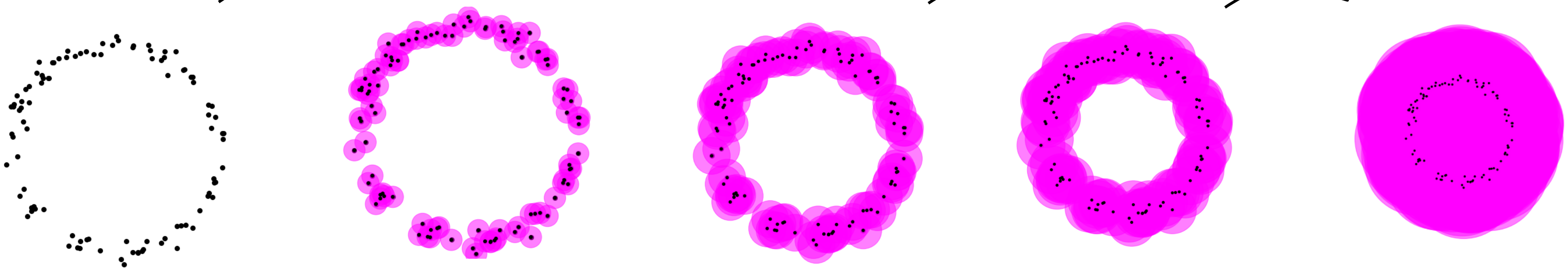
Homologia da filtração de Cech

Vamos calcular a homologia de cada espessamento:

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

Vamos calcular a homologia de cada espessamento:

inclusões i_s^t: $X^0 \xrightarrow{i_0^{0.1}} X^{0.1} \xrightarrow{i_{0.1}^{0.2}} X^{0.2} \xrightarrow{i_{0.2}^{0.3}} X^{0.3} \xrightarrow{i_{0.3}^{1}} X^1$



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

$$0 \xrightarrow{(i_0^{0.1})_*} 0 \xrightarrow{(i_{0.1}^{0.2})_*} \mathbb{Z}/2\mathbb{Z} \xrightarrow{(i_{0.2}^{0.3})_*} \mathbb{Z}/2\mathbb{Z} \xrightarrow{(i_{0.3}^{1})_*} 0$$

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

Módulos de persistência

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

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

$$\begin{tikzcd} V^r \arrow[r, "v_r^s"] \arrow[rr, bend right, "v_r^t"'] & V^s \arrow[r, "v_s^t"] & V^t \end{tikzcd}$$

Módulo de persistência associado à filtração de Cech :

$$\begin{tikzcd} \cdots \arrow[r, dashed] & X^{t_1} \arrow[r, hook, "i_{t_1}^{t_2}"] & X^{t_2} \arrow[r, hook, "i_{t_2}^{t_3}"] & X^{t_3} \arrow[r, hook, "i_{t_3}^{t_4}"] & X^{t_4} & \cdots \arrow[l, hook', dashed] \end{tikzcd}$$

$$\begin{tikzcd} \cdots \arrow[r, dashed] & H_i(X^{t_1}) \arrow[r, "(i_{t_1}^{t_2})_*"] & H_i(X^{t_2}) \arrow[r, "(i_{t_2}^{t_3})_*"] & H_i(X^{t_3}) \arrow[r, "(i_{t_3}^{t_4})_*"] & H_i(X^{t_4}) \arrow[r, dash, dashed] & \cdots \end{tikzcd}$$

Monitorando a evolução dos buracos

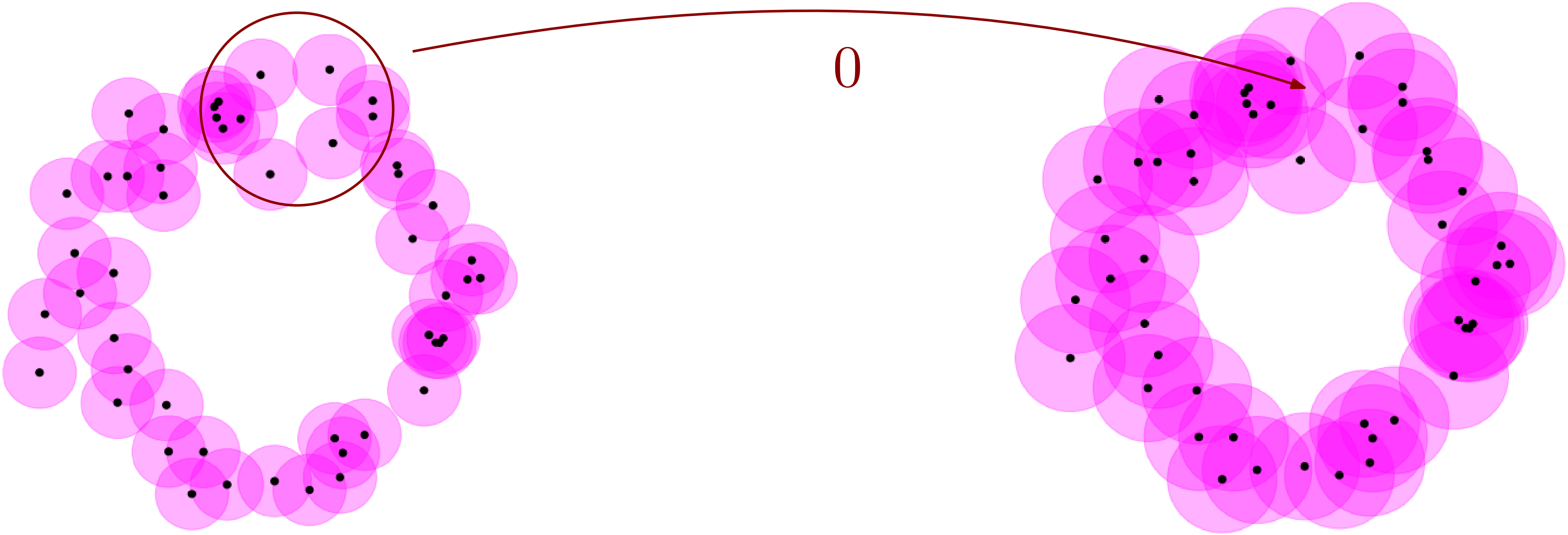
$$\dashrightarrow H_i(X^{t_1}) \xrightarrow{\left(i_{t_1}^{t_2}\right)_*} H_i(X^{t_2}) \xrightarrow{\left(i_{t_2}^{t_3}\right)_*} H_i(X^{t_3}) \xrightarrow{\left(i_{t_2}^{t_3}\right)_*} H_i(X^{t_4}) \dashrightarrow$$

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

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

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

Sua **persistância** é a diferença.



Monitorando a evolução dos buracos

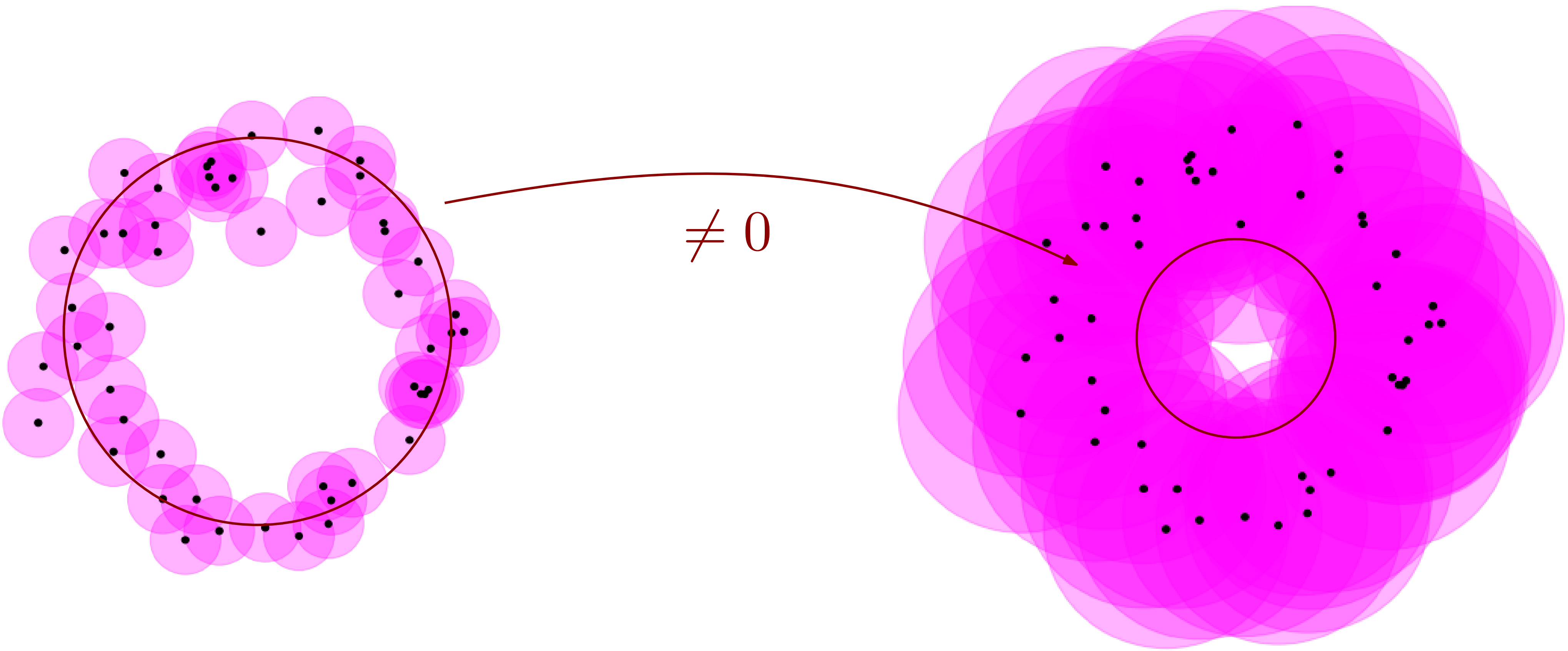
$$\dashrightarrow H_i(X^{t_1}) \xrightarrow{\left(i_{t_1}^{t_2}\right)_*} H_i(X^{t_2}) \xrightarrow{\left(i_{t_2}^{t_3}\right)_*} H_i(X^{t_3}) \xrightarrow{\left(i_{t_2}^{t_3}\right)_*} H_i(X^{t_4}) \dashrightarrow$$

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

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

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

Sua **persistância** é a diferença.



Monitorando a evolução dos buracos

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

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

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

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

Sua **persistância** é a diferença.

Interpretação:

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

Códigos de barras

Teorema (Crawley-Boevey, 2015):

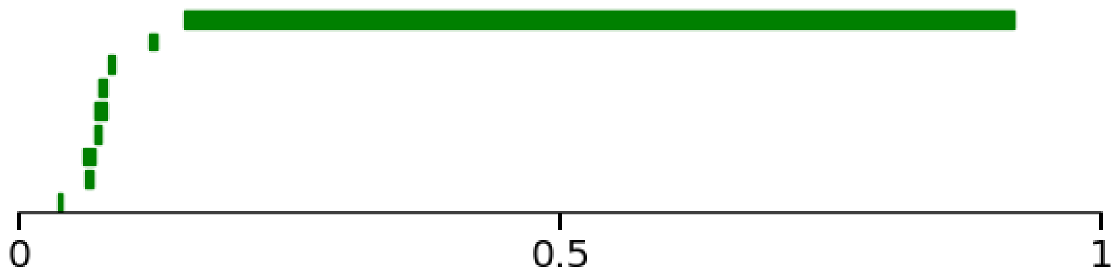
Um módulo de persistência (regular) pode ser decomposto em módulos-intervalos.

Este conjunto múltiplo de intervalos é chamado de **código de barras**. Ele descreve completamente o módulo de persistência.

Módulo de persistência : $\mathbb{V}$

Código de barras: $\{$ [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) $\}$

Representação gráfica:



Códigos de barras

Código de barras do módulo de persistência associado à filtração de Čech: H_0 em vermelho e H_1 em verde.

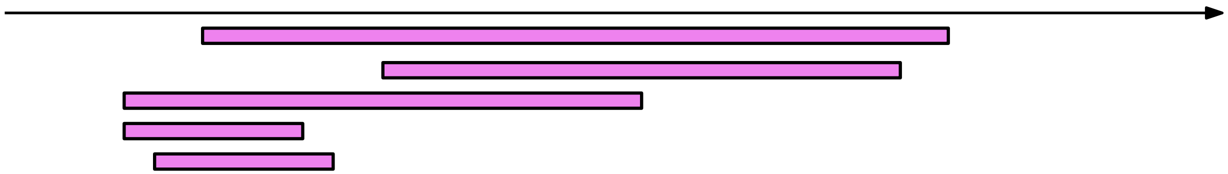
Códigos de barras

Código de barras do módulo de persistência associado à filtração de Čech: H_0 em vermelho e H_1 em verde.

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

Diagramas de persistência

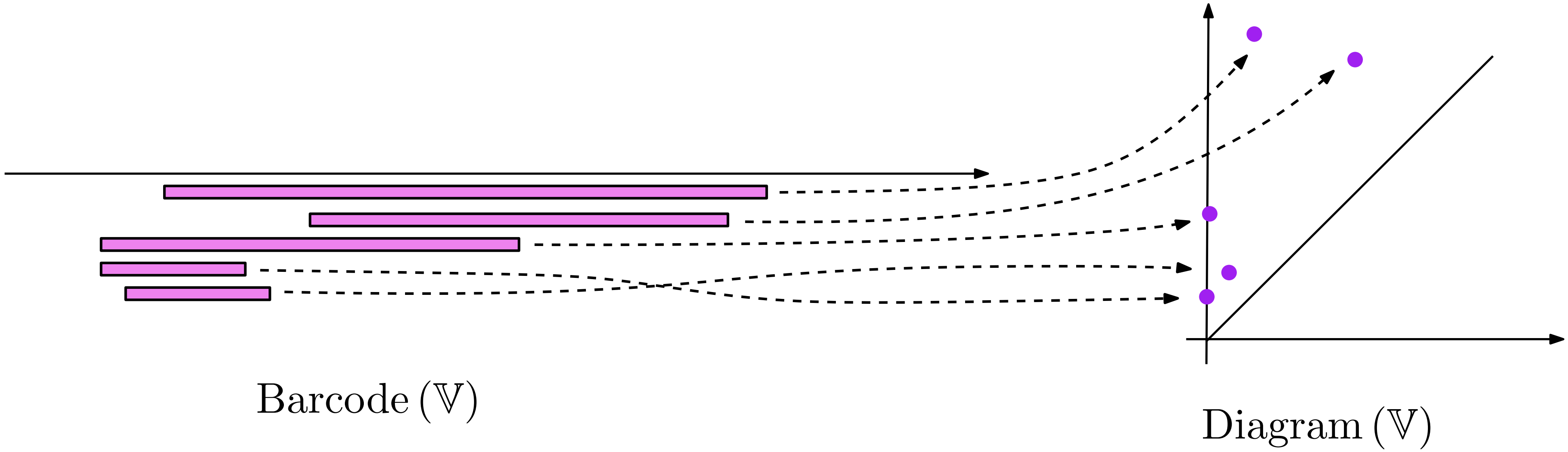
Cada módulo de persistência de $\mathbb{V}$ está associado a um código de barras



$\mathrm{Barcode}\,(\mathbb{V})$

Diagramas de persistência

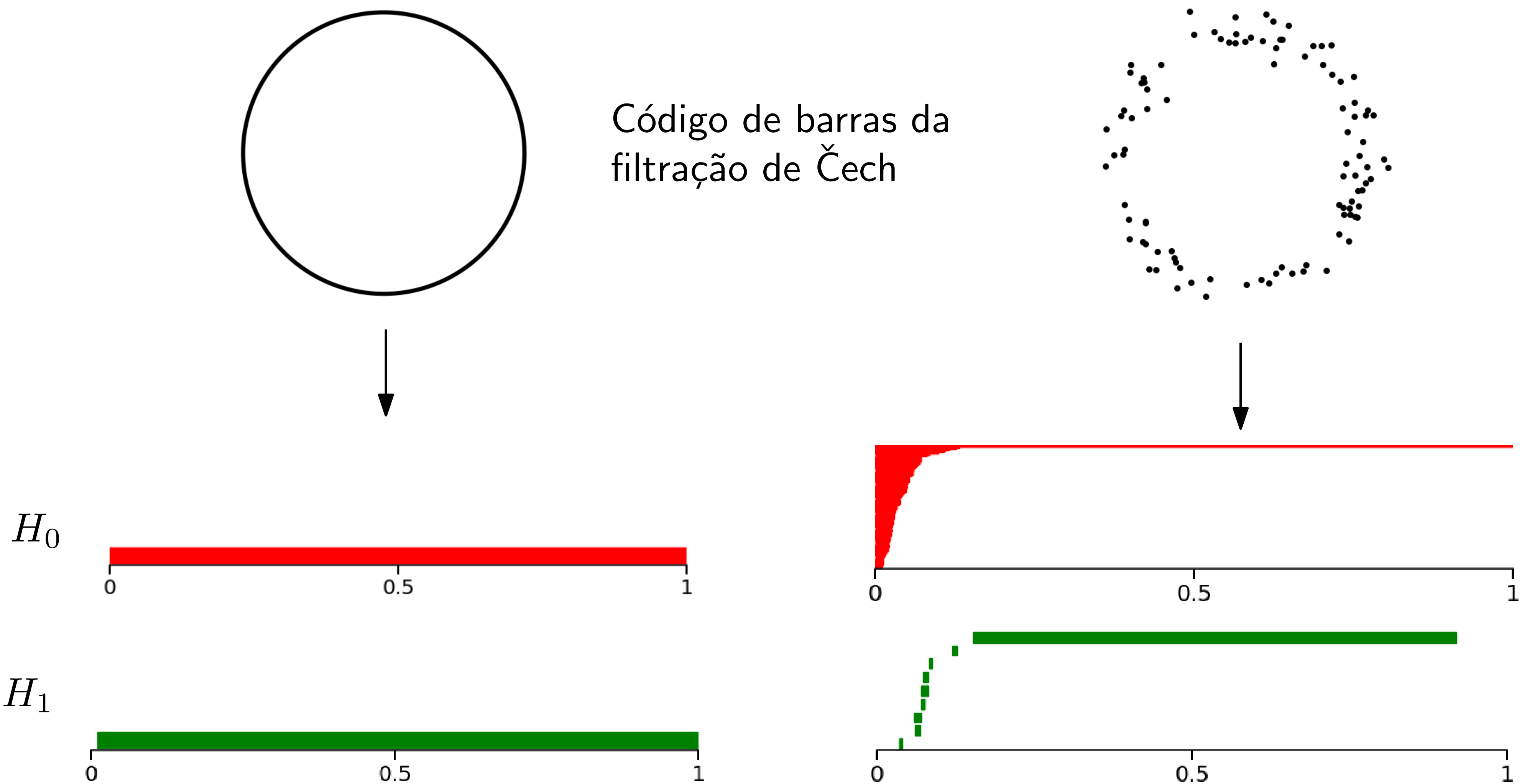
Cada módulo de persistência de $\mathbb{V}$ está associado a um código de barras



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

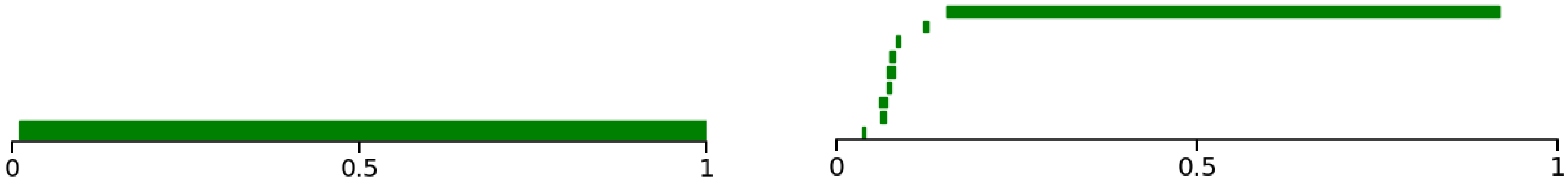
Estabilidade dos módulos de persistência

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



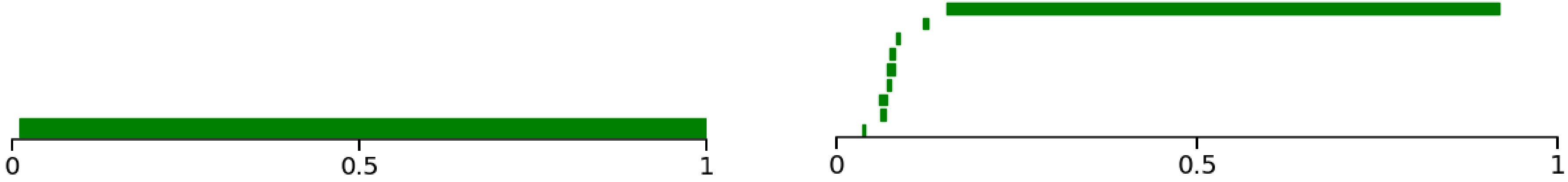
Distância bottleneck

Sejam dois códigos de barras P e Q.



Distância bottleneck

Sejam dois códigos de barras P e Q.



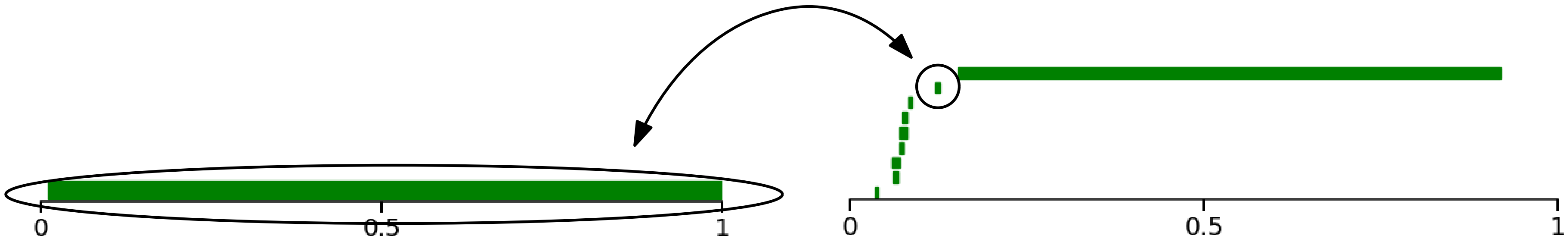
Uma **correspondência parcial** entre P e Q e um subconjunto $M \subset P \times Q$ tal como

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

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

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

Sejam dois códigos de barras P e Q.



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

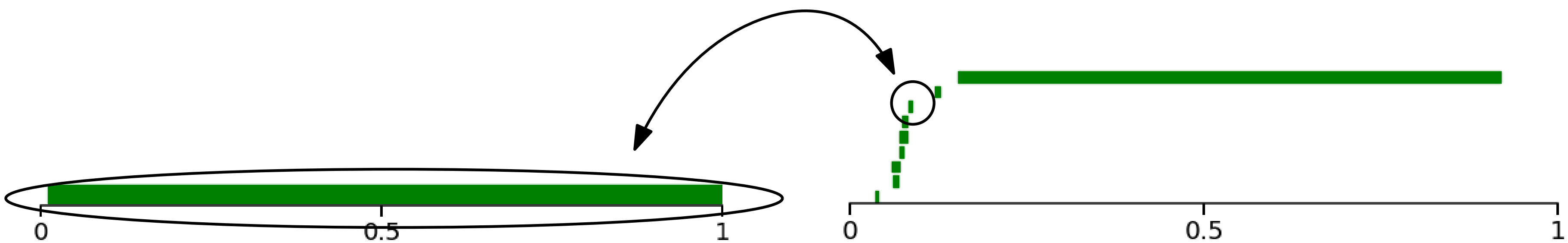
- por todo $p \in P$, existe no máximo um $q \in Q$ tal como $(p, q) \in M$,
- por todo $q \in Q$, existe no máximo um $p \in P$ tal como $(p, q) \in M$.

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

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

Distância bottleneck

Sejam dois códigos de barras P e Q.



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

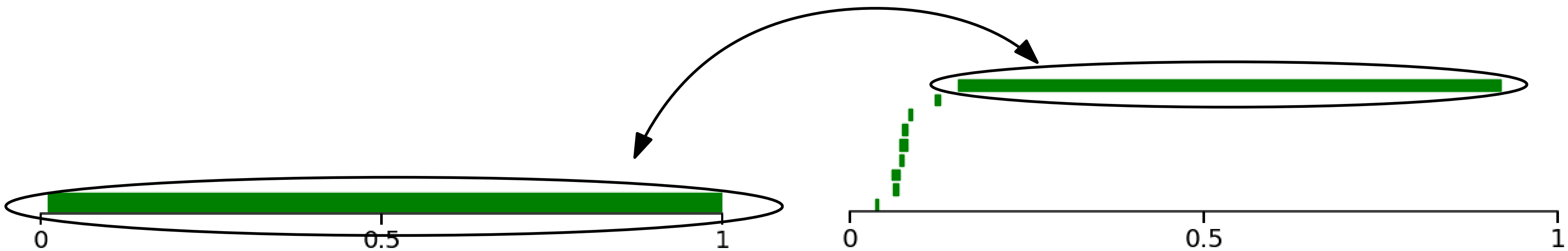
- por todo $p \in P$, existe no máximo um $q \in Q$ tal como $(p, q) \in M$,
- por todo $q \in Q$, existe no máximo um $p \in P$ tal como $(p, q) \in M$.

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

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

Distância bottleneck

Sejam dois códigos de barras P e Q.



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

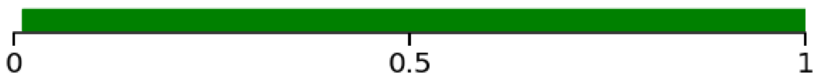
- por todo $p \in P$, existe no máximo um $q \in Q$ tal como $(p, q) \in M$,
- por todo $q \in Q$, existe no máximo um $p \in P$ tal como $(p, q) \in M$.

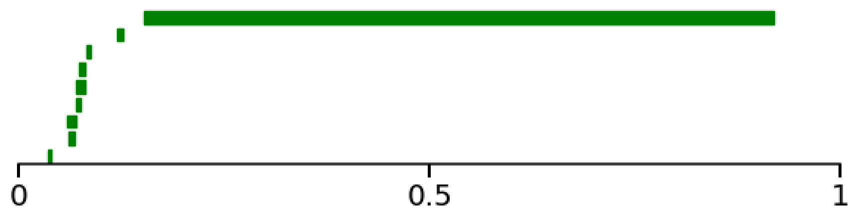
As barras $p \in P$ (resp. $q \in Q$) tal como existe $q \in Q$ (resp. $p \in P$) com $(p, q) \in M$ são referidas como **correspondentes** por M.

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

Distância bottleneck

Sejam dois códigos de barras P e Q.





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

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

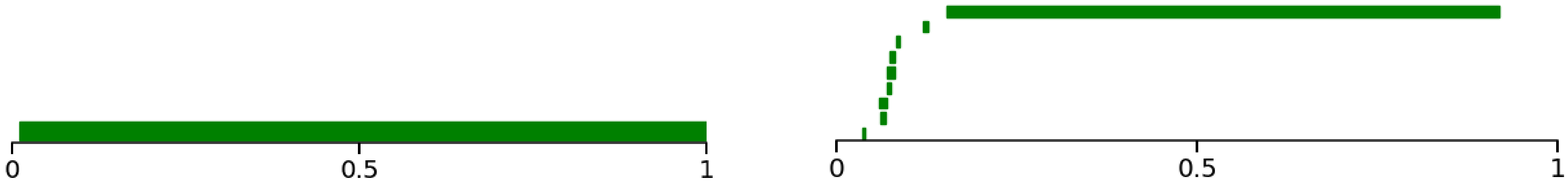
As barras $p \in P$ (resp. $q \in Q$) tal como existe $q \in Q$ (resp. $p \in P$) com $(p, q) \in M$ são referidas como **correspondentes** por M.

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

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

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

Sejam dois códigos de barras P e Q.

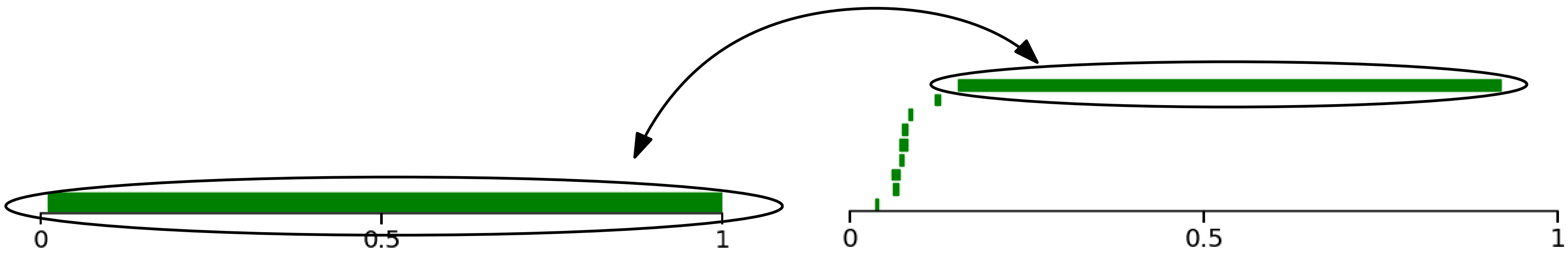


Definição : A **distância bottleck** entre P e Q é definida como o menor custo de qualquer correspondência:

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

Distância bottleneck

Sejam dois códigos de barras P e Q.



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

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

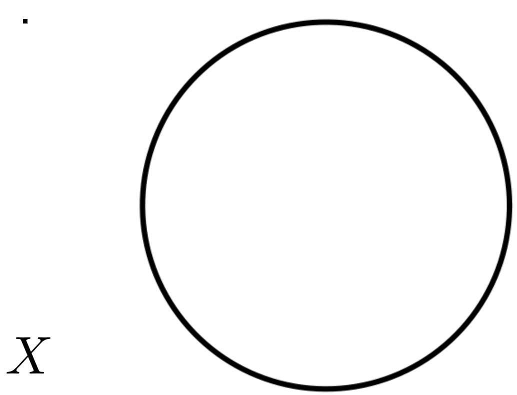
Teorema da estabilidade

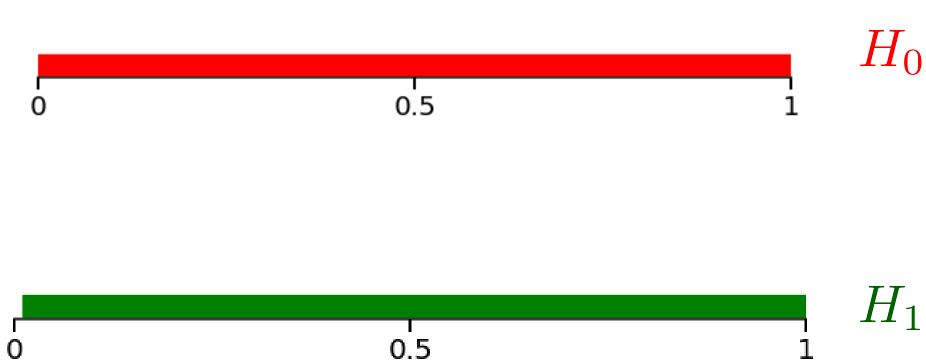
Teorema (Cohen-Steiner, Edelsbrunner et Harer, 2005):
Sejam X e Y dois subconjuntos do $\mathbb{R}^n$.
Considerem suas filtrações de Čech, e seus módulos de persistência de i-ésima homologia, notados $\mathbb{U}$ e $\mathbb{V}$. Então

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

distância bottleneck

distância de Hausdorff





Teorema da estabilidade

Teorema (Cohen-Steiner, Edelsbrunner et Harer, 2005):
Sejam X e Y dois subconjuntos do $\mathbb{R}^n$.
Considerem suas filtrações de Čech, e seus módulos de persistência de i-ésima homologia, notados $\mathbb{U}$ e $\mathbb{V}$. Então

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

.

distância bottleneck

distância de Hausdorff

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

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

onde $\mathrm{d_i}$ é a **distância de entrelaçamento**.

I - Comparar espaços topológicos

1 - Equivalência de homeomorfia

2 - Equivalência de homotopia

II - Invariantes topológicos

1 - Número de componentes conexas

2 - Característica de Euler

3 - Números de Betti

4 - Homologia (simplicial sobre $\mathbb{Z}/2\mathbb{Z}$)

III - Inferência topológica

1 - Estimação do parâmetro t

2 - Nervos

IV - Homologia persistente

1 - Módulos de persistência

2 - Decomposição

3 - Estabilidade

V - Aplicações

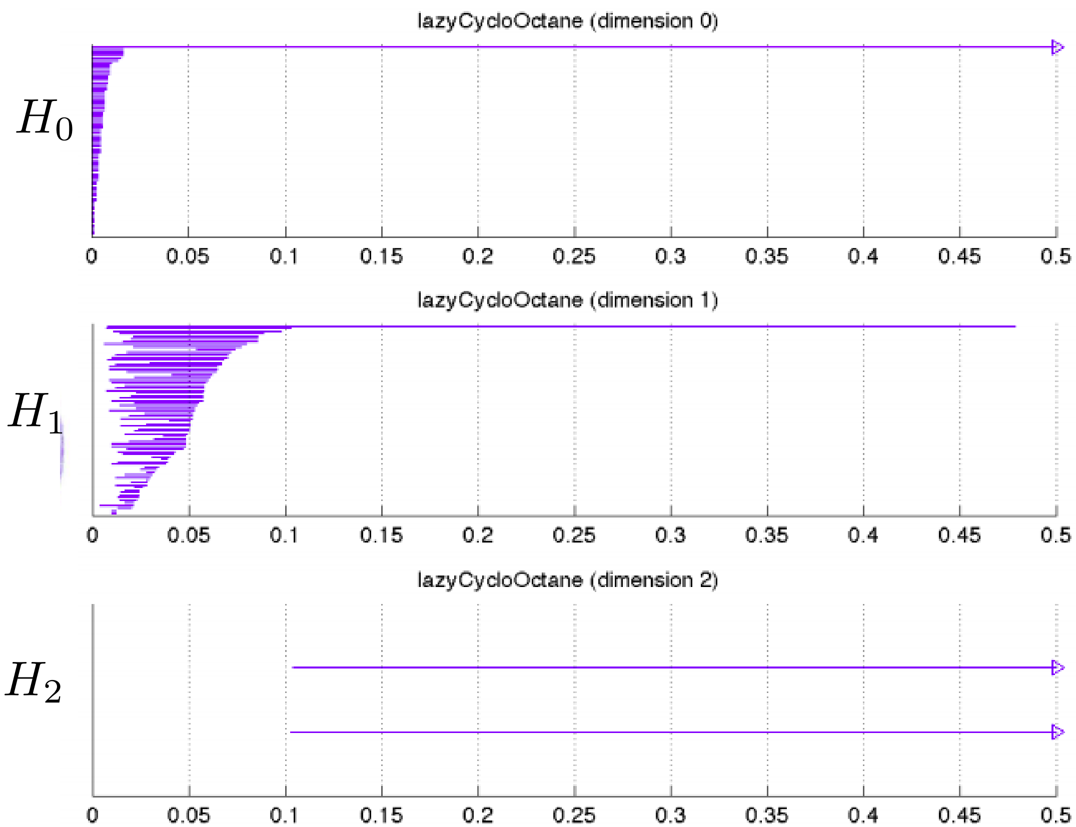
Inferência topológica I

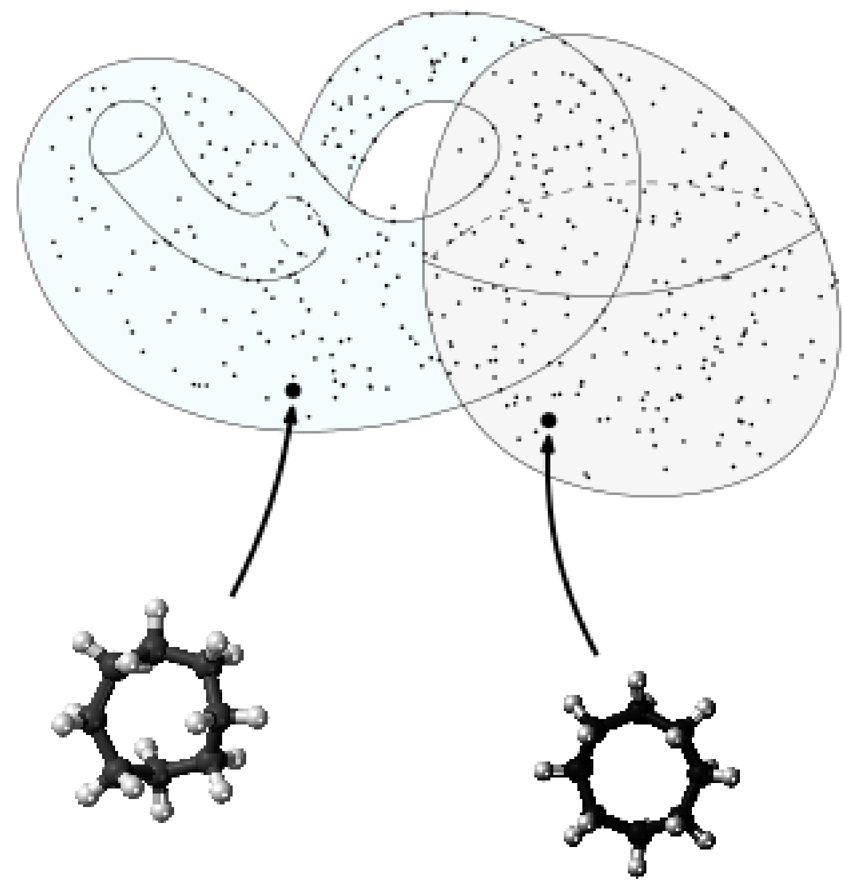
[S. Martin, A. Thompson, E. A. Coutsias, and J-P. Watson, Topology of cyclo-octane energy landscape, 2010]

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

A molécula de ciclo-octano $\mathrm{C_8H_{16}}$ contém 24 átomos.
Ao gerar muitas moléculas, obtemos uma nuvem de pontos em $\mathbb{R}^{72}$ $(3 \times 24 = 72)$.

Os códigos de barras são calculados:



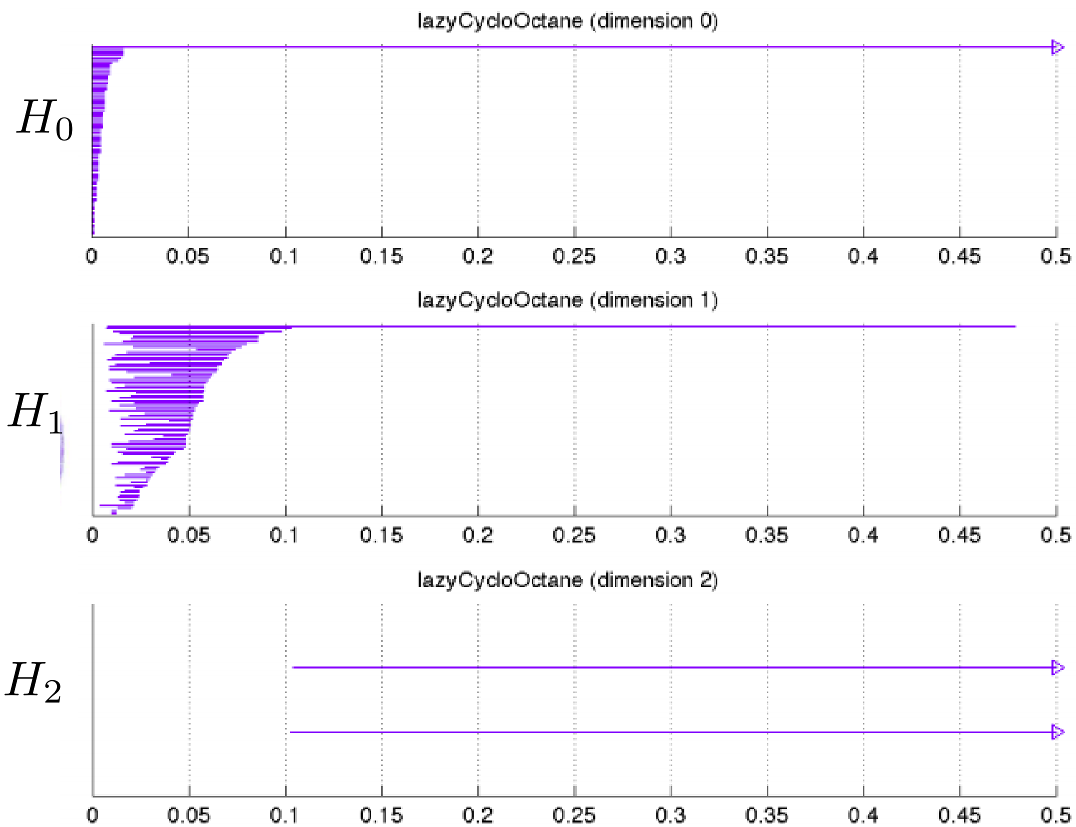


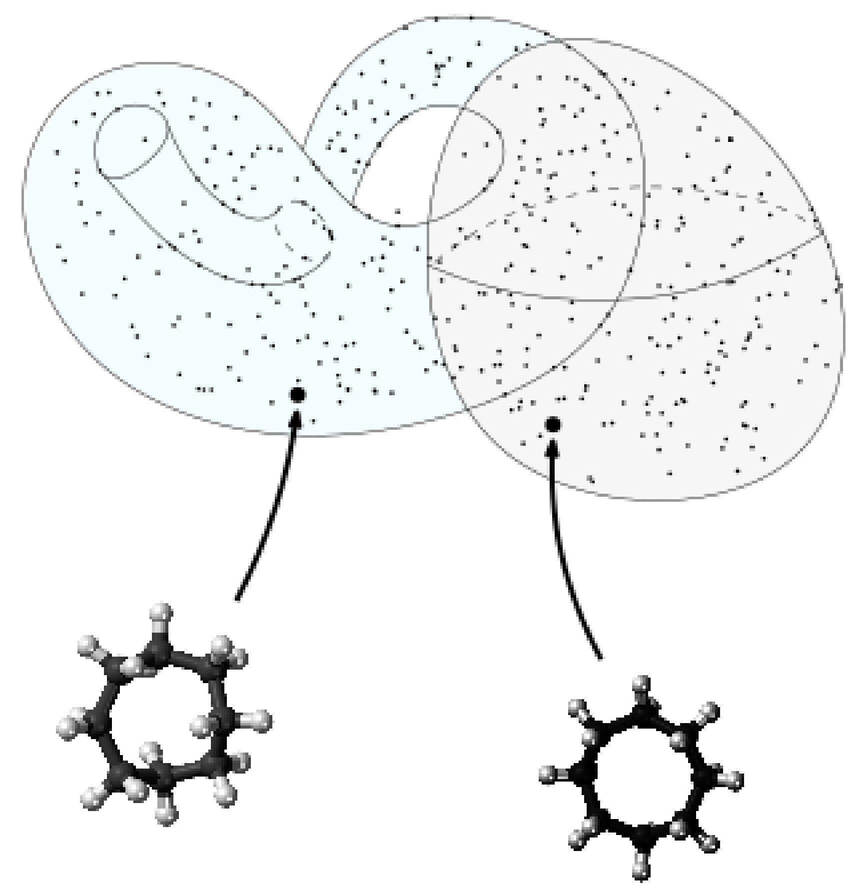
[S. Martin, A. Thompson, E. A. Coutsias, and J-P. Watson, Topology of cyclo-octane energy landscape, 2010]

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

A molécula de ciclo-octano $\mathrm{C_8H_{16}}$ contém 24 átomos.
Ao gerar muitas moléculas, obtemos uma nuvem de pontos em $\mathbb{R}^{72}$ $(3 \times 24 = 72)$.

Os códigos de barras são calculados:



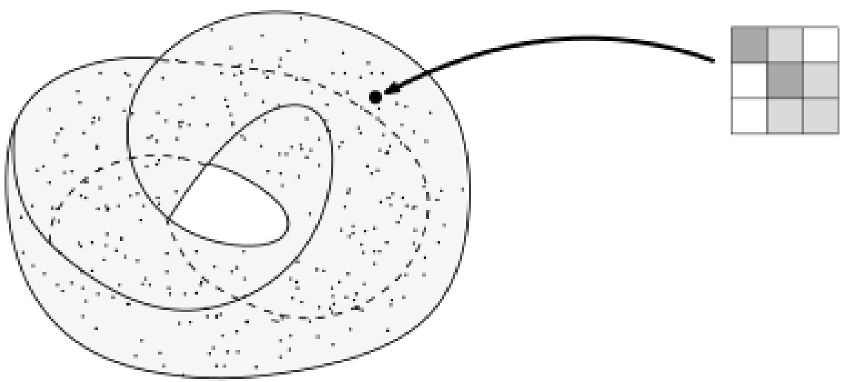


Deduzimos: $H_0 = \mathbb{Z}/2\mathbb{Z}, \quad H_1 = \mathbb{Z}/2\mathbb{Z}, \quad H_2 = (\mathbb{Z}/2\mathbb{Z})^2$

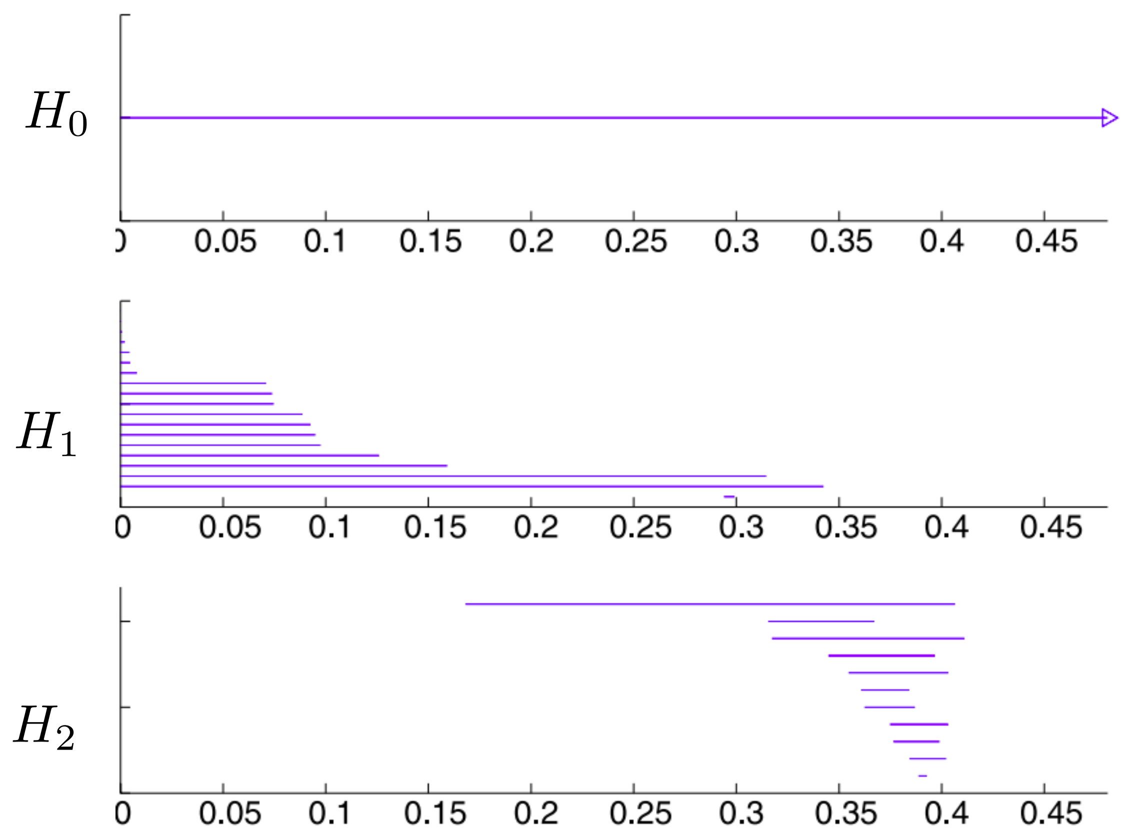
[G. Carlsson, T. Ishkhanov, V. de Silva, and A. Zomorodian, On the Local Behavior of Spaces of Natural Images, 2008]

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

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



Os códigos de barras são obtidos:

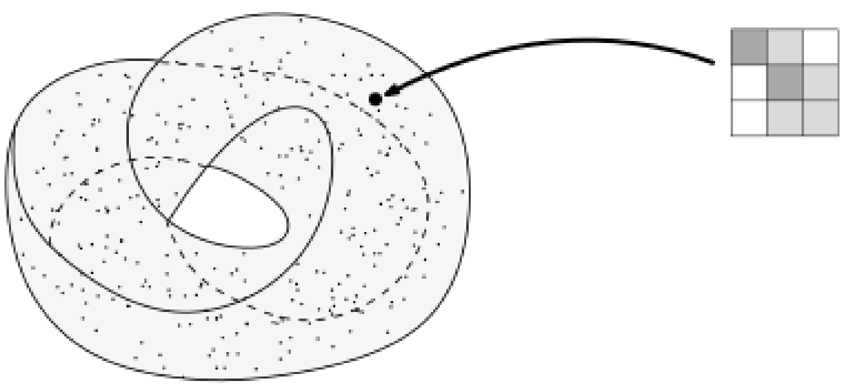


Inferência topológica II

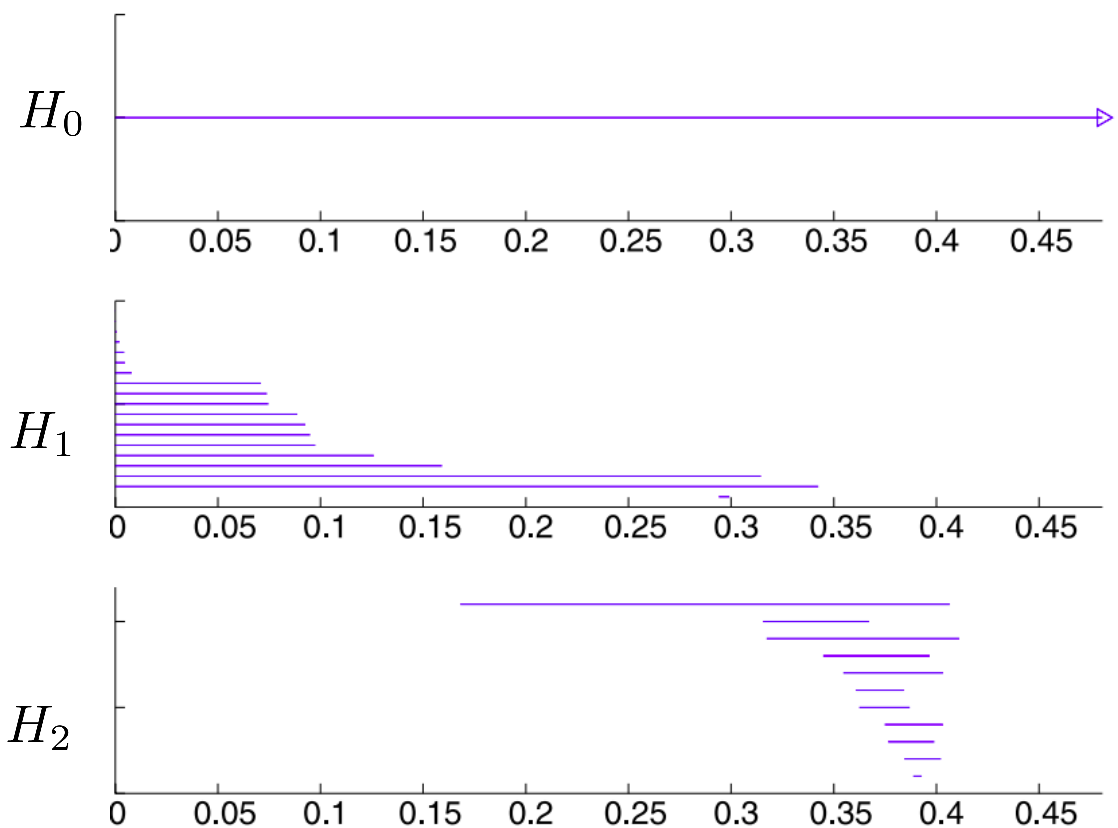
[G. Carlsson, T. Ishkhanov, V. de Silva, and A. Zomorodian, On the Local Behavior of Spaces of Natural Images, 2008]

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

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



Os códigos de barras são obtidos:



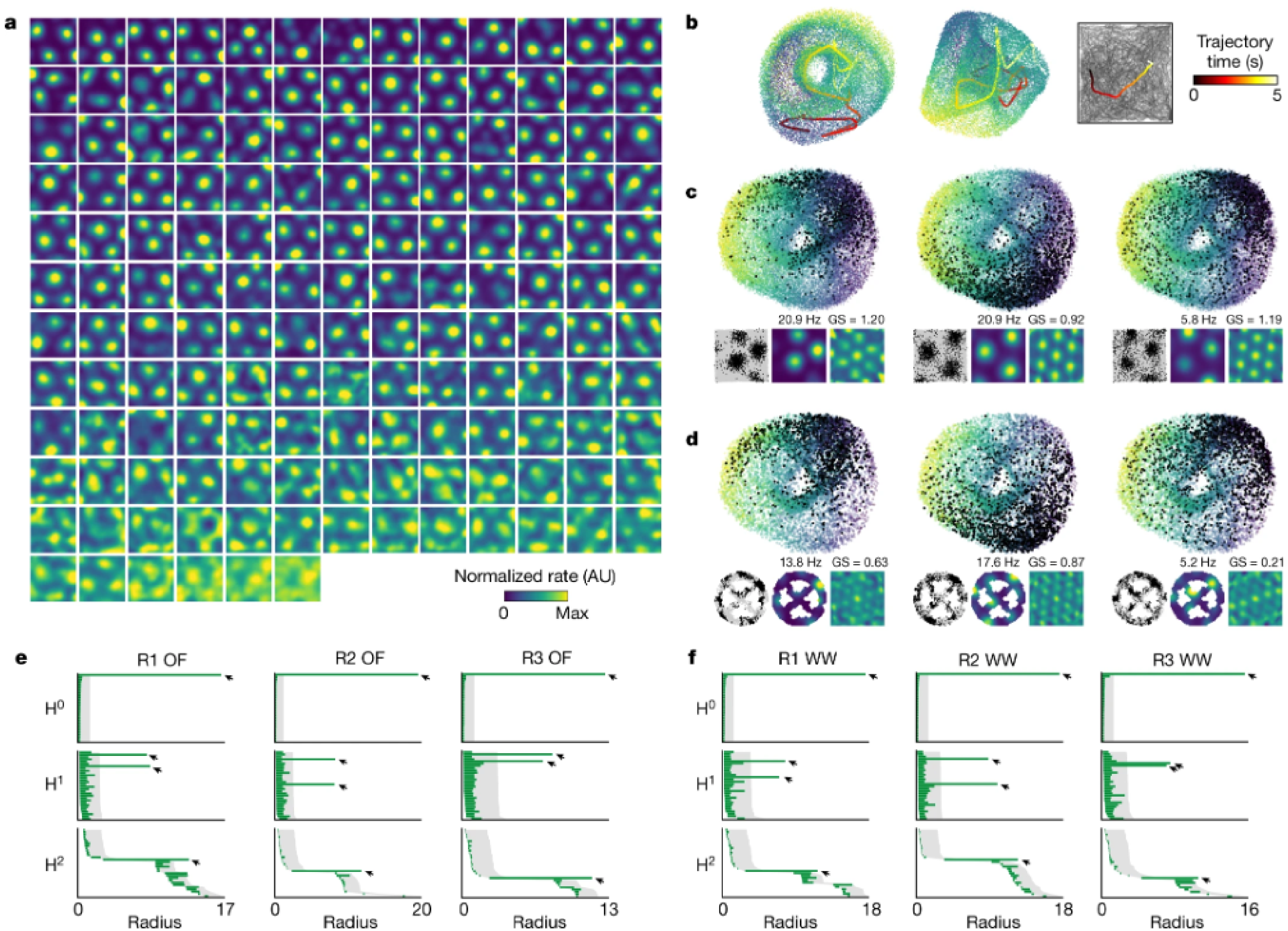
Deduzimos:

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

Inferência topológica III

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

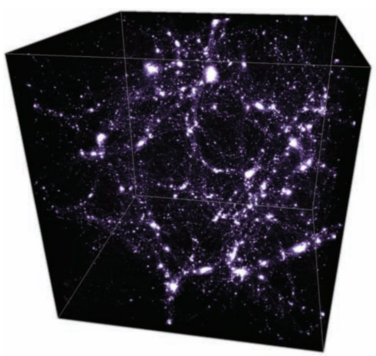
Os autores registraram spikes de grid-cells de ratos, e aplicaram redução de dimensionalidade na matriz de firing. Ao aplicar a homologia persistente a esta nuvem de pontos, observamos a homologia de um toro.



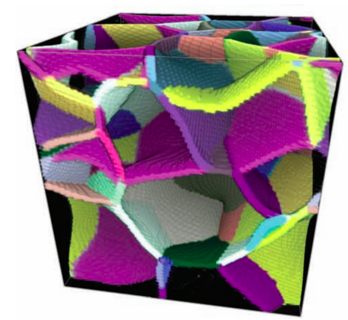
Análise multiescala I

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

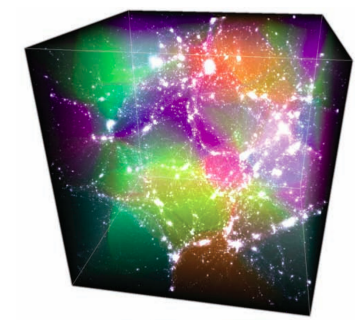
https://www.giss.nasa.gov/staff/mway/cluster/sousbie2011mnras.pdf



visto como um objeto de dimensão 1



de dimensão 2



de dimensão 3

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

https://arxiv.org/pdf/1608.04519.pdff

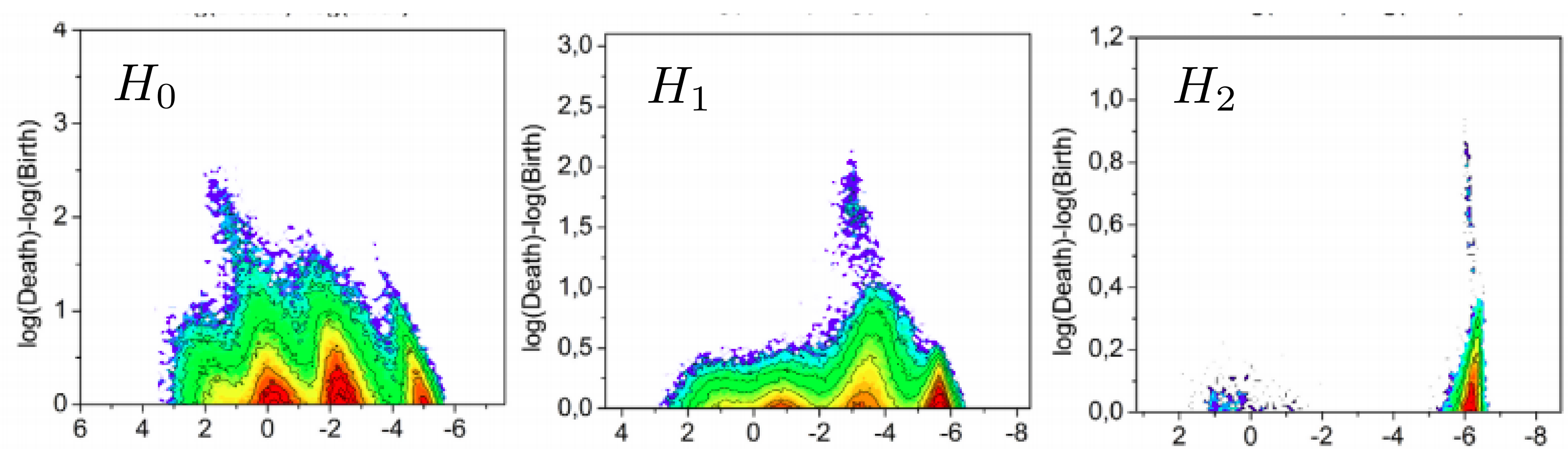


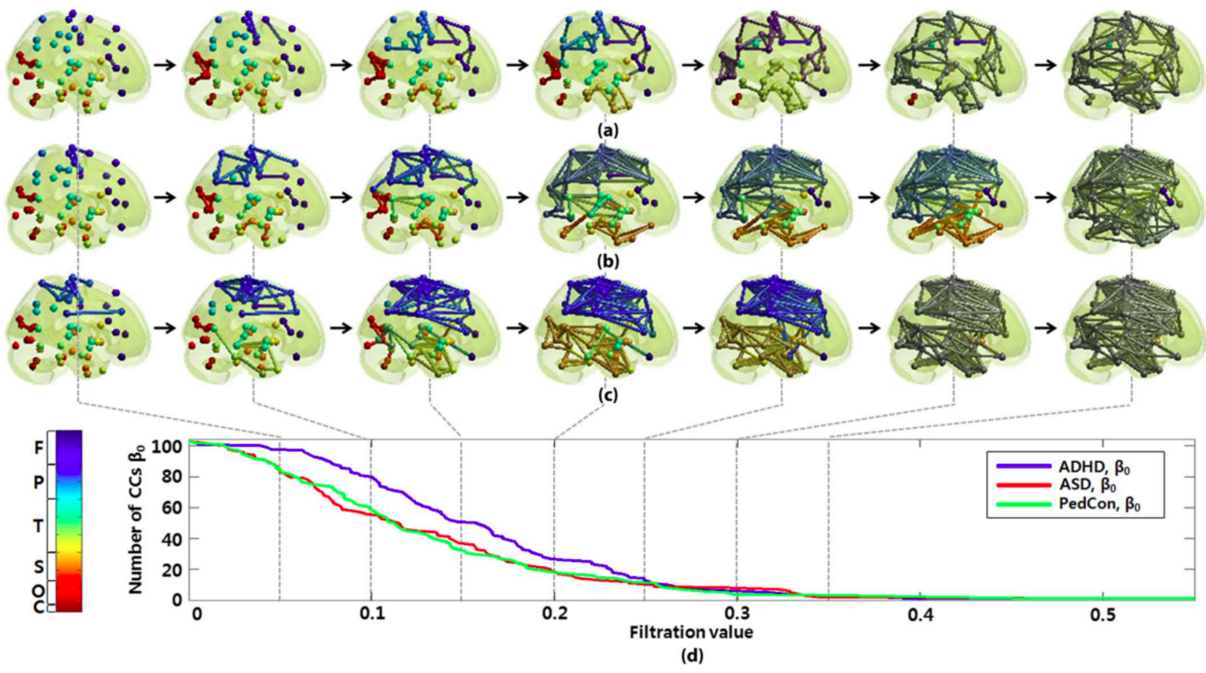
Diagrama de persistência média para um modelo Voronoï (escala log)

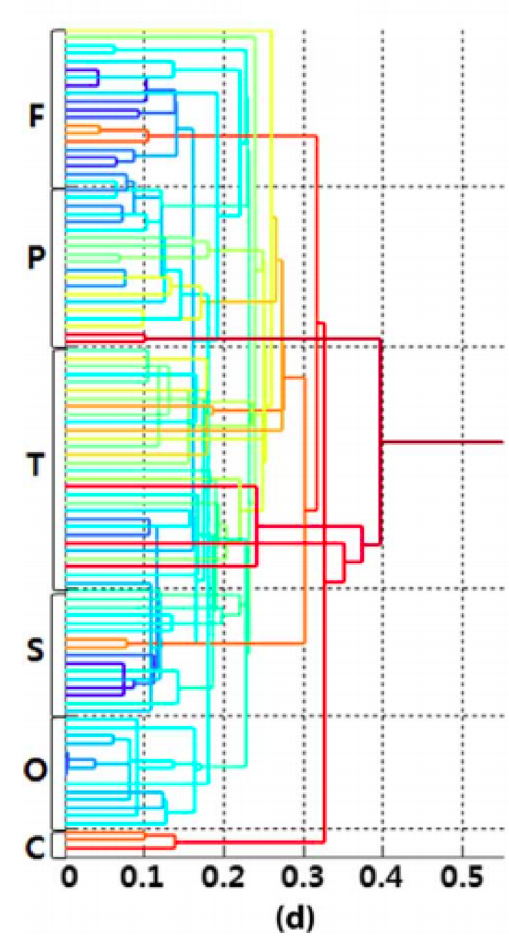
Análise multiescala II

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

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

→ A H_0-homologia persistente induz uma classificação hierárquica





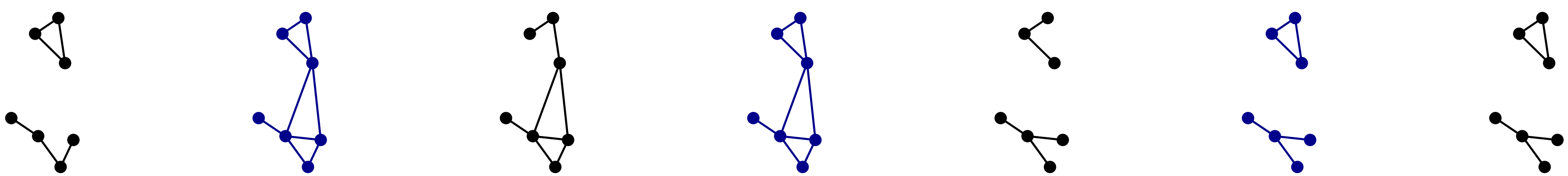
Análise de grafos temporais

Em colaboração com Cláudio Linhares e Jean Ponciano.

Um grafo temporal consiste nos dados de um grafo G e uma coleção de pares (v, t), onde v é uma aresta de G e t um número real em $[0, T]$.

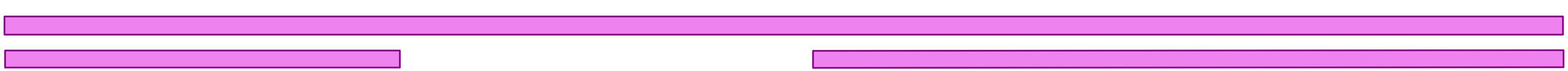
Nesse contexto, temos vários espaços topológicos. Desejamos entender a **evolução** dos clusters.

Podemos construir a *filtração zig-zag* associada.



$$G_0 \to G_0 \cup G_1 \leftarrow G_1 \to G_1 \cup G_2 \leftarrow G_2 \to G_2 \cup G_3 \leftarrow G_3$$

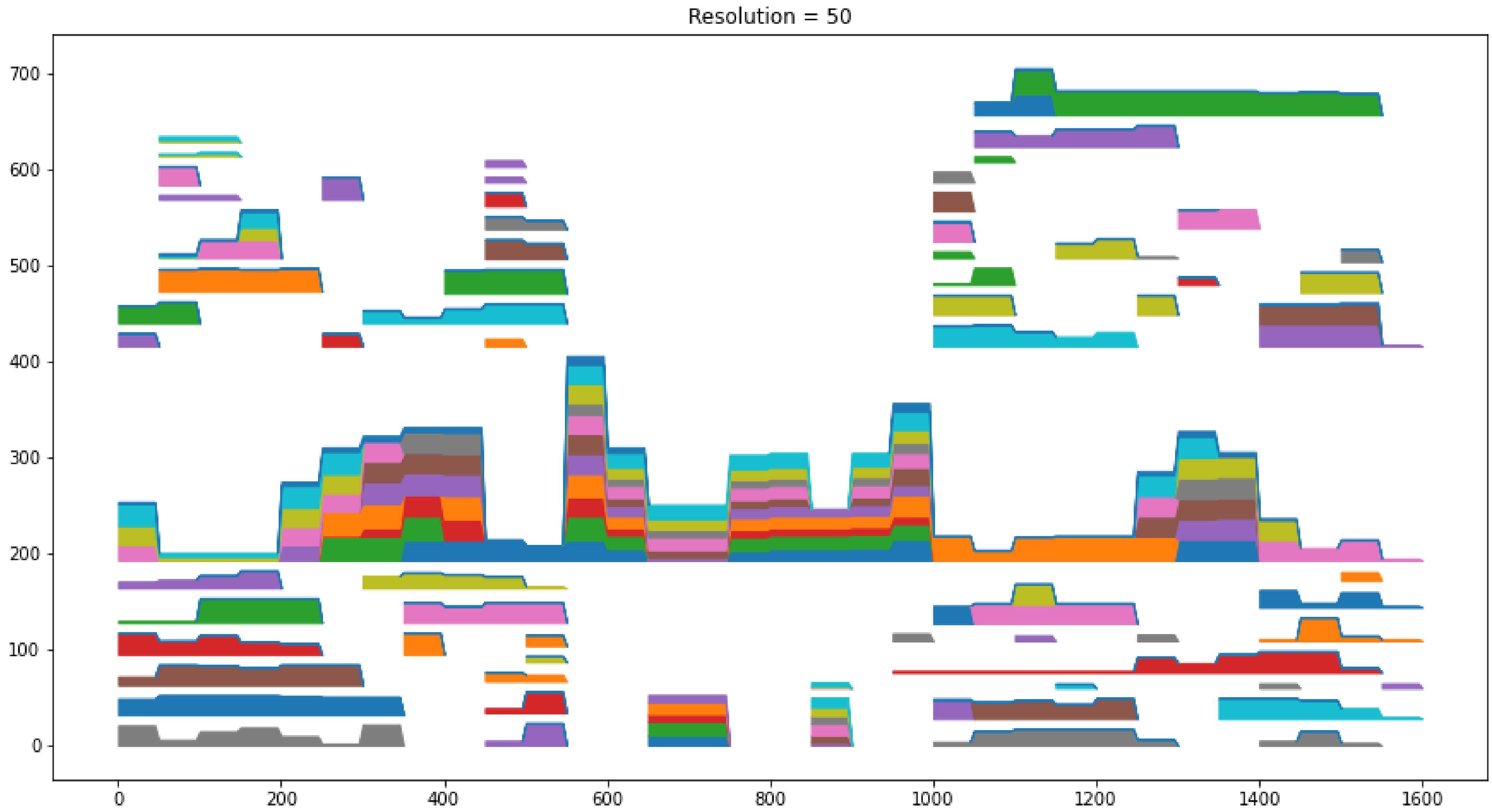
A teoria da homologia persistente zigzag também oferece um código de barras.



Análise de grafos temporais

Em colaboração com Cláudio Linhares e Jean Ponciano.

Um grafo temporal consiste nos dados de um grafo G e uma coleção de pares (v, t), onde v é uma aresta de G e t um número real em $[0, T]$.



No machine learning I

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

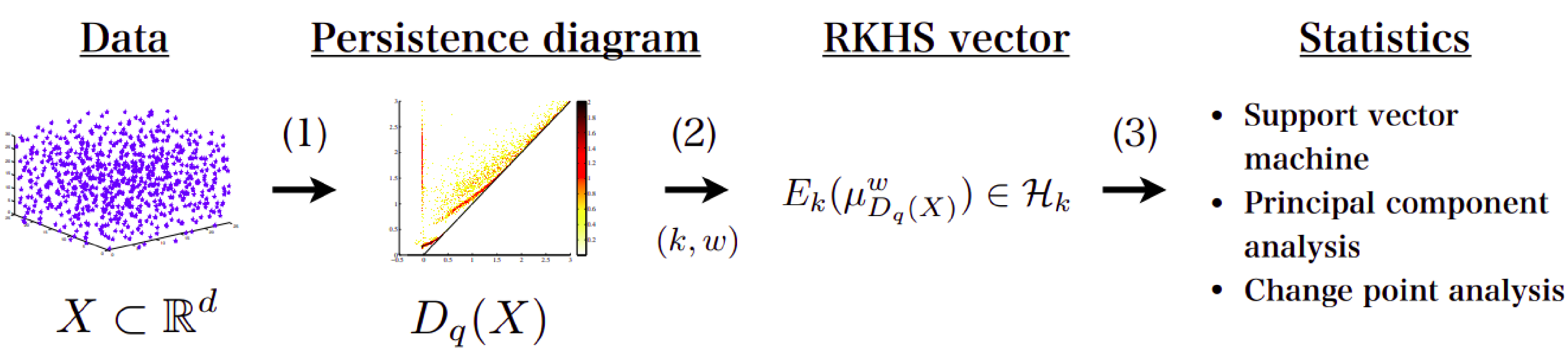
https://arxiv.org/abs/1706.03358

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

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

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

⟶ podemos usar o **kernel trick**



No machine learning II

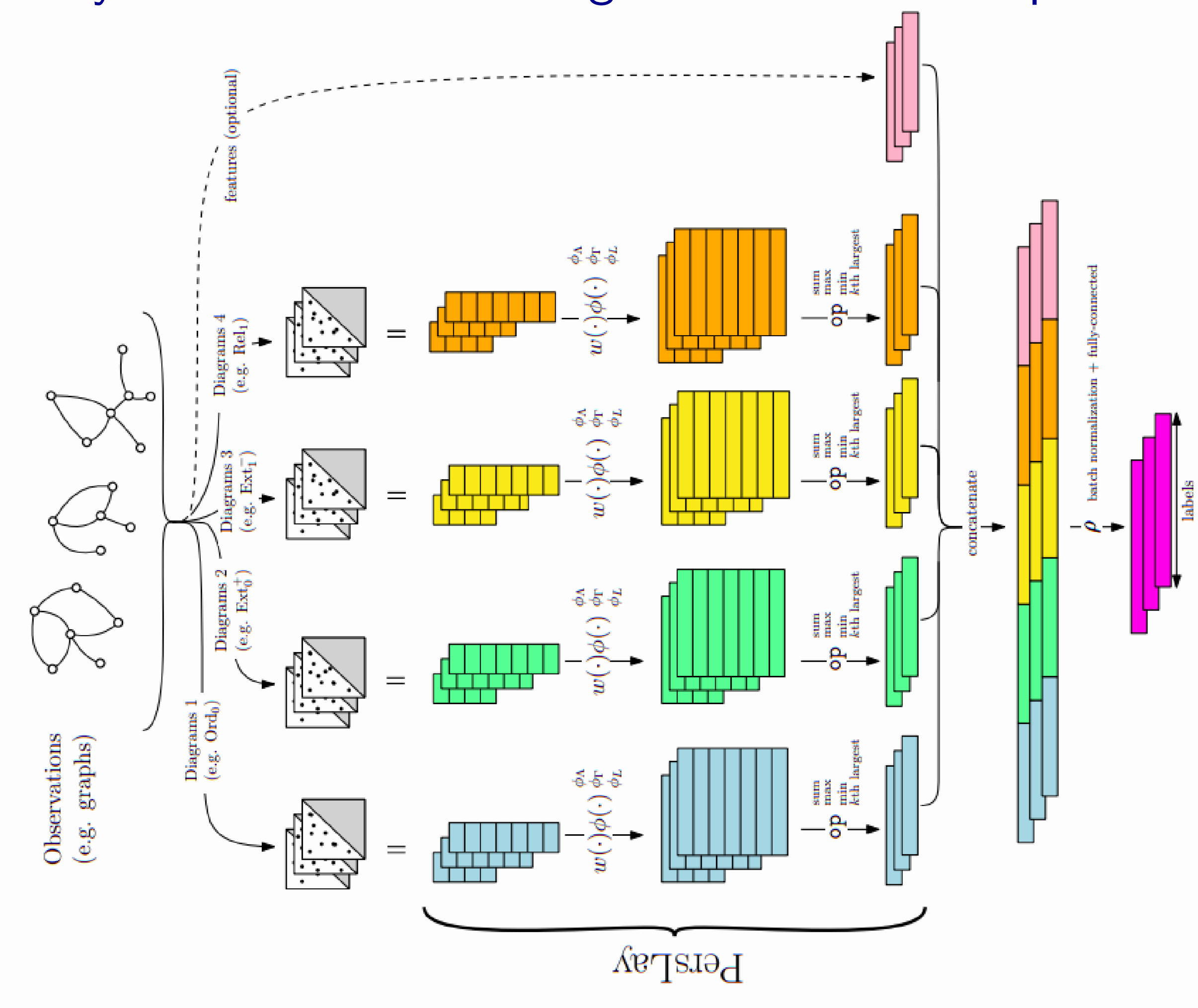
Rickard Brüel-Gabrielsson, Bradley J. Nelson, Anjan Dwaraknath, Primoz Skraba, Leonidas J. Guibas, Gunnar Carlsson, A Topology Layer for Machine Learning, 2019

https://arxiv.org/abs/1905.12200

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

https://arxiv.org/abs/1904.09378

Os diagramas de persistência podem ser integrados com uma rede neural

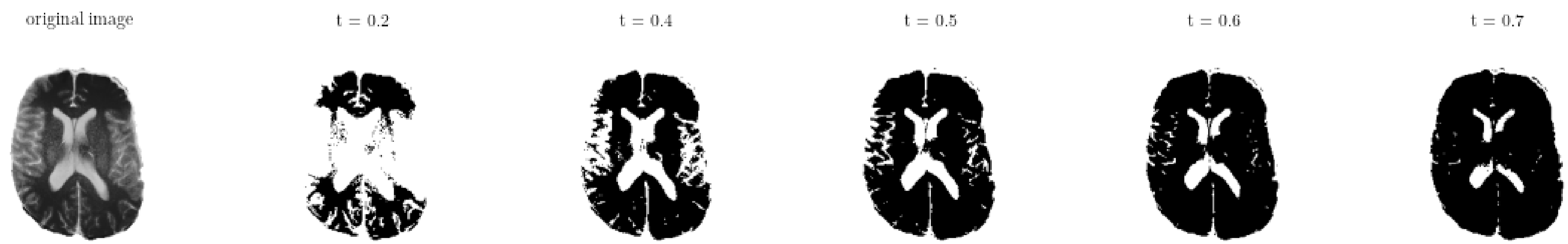


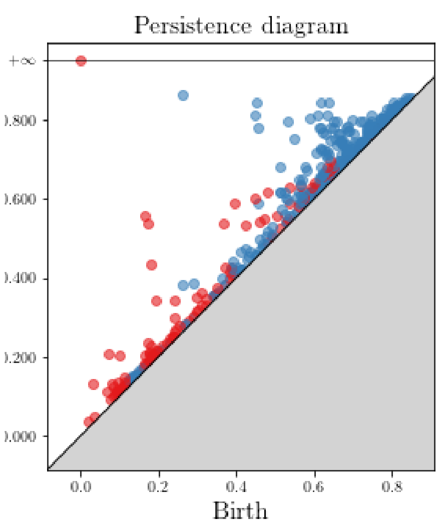
Em colaboração com Anton François.

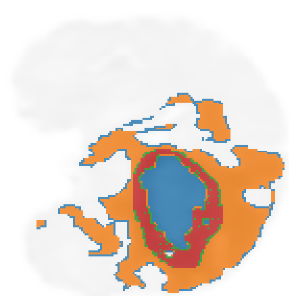
O **glioblastoma** é o tumor cerebral mais comum, difuso, de grau variável de agressividade, e cujo prognóstico médico é difícil de estabelecer.

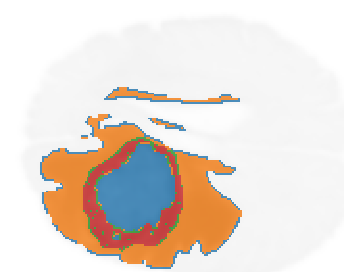
Neste contexto, o problema da **segmentação** consiste na demarcação automática das três regiões que formam o tumor (edema, núcleo necrótico e enhancing tumor).

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









Classificação

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

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

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

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

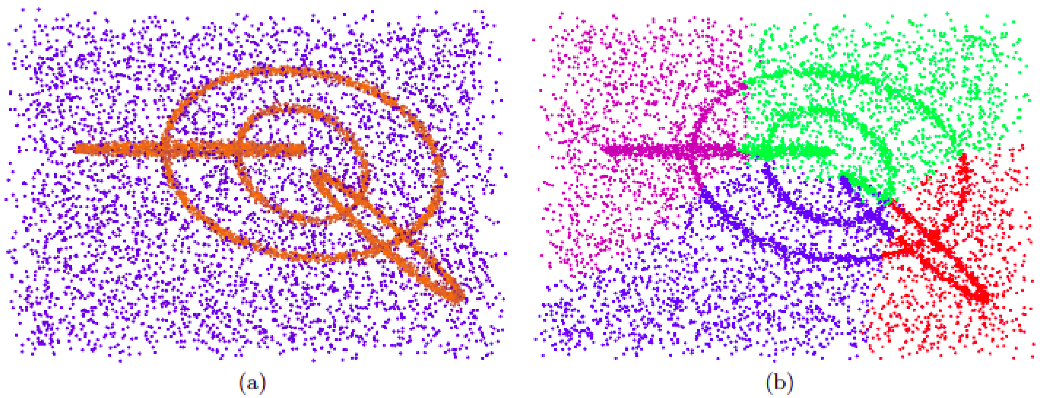
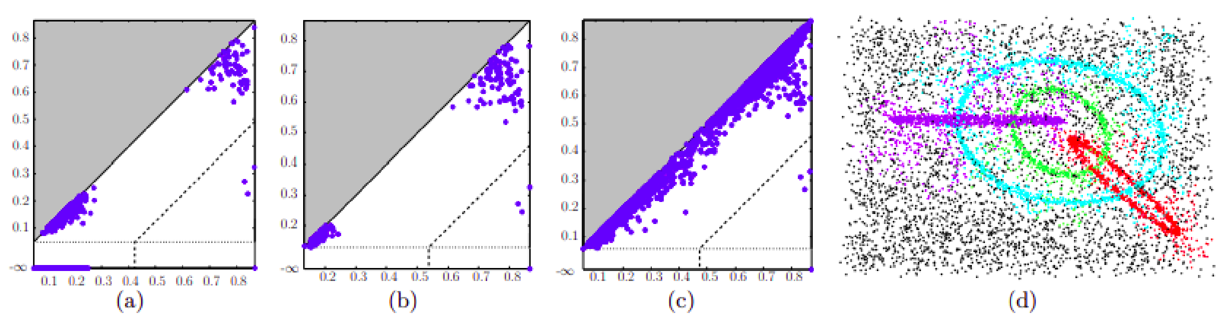
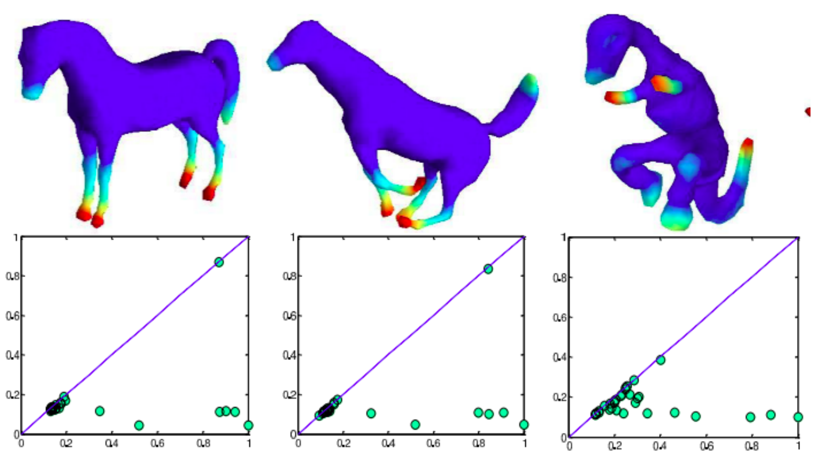


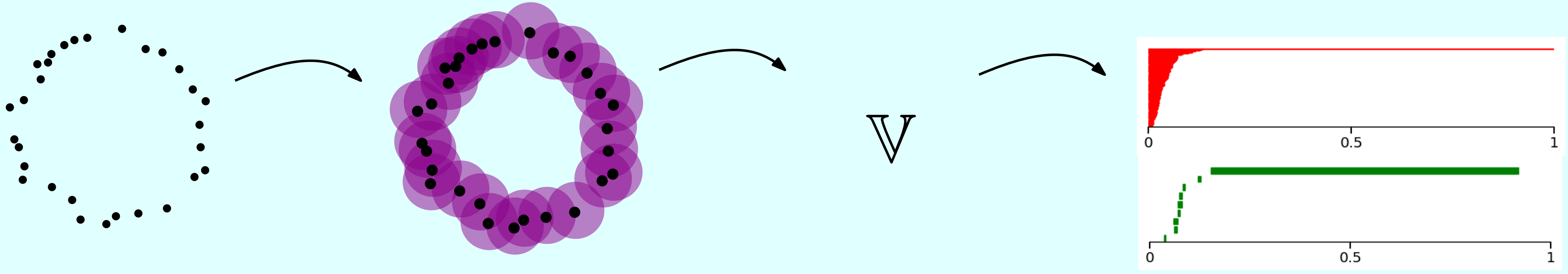
Figure 7: (a) The rings data set with the estimated density function. (b) The result obtained using spectral clustering.





Conclusão

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



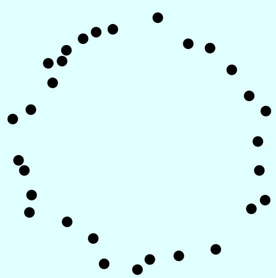
nuvem de pontos　filtração de Čech　módulo de persistência　código de barras

Permite analisar os dados a partir de uma nova perspectiva.

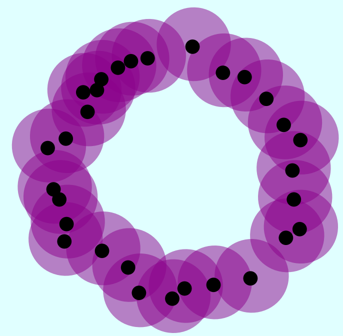
Um curso de TDA: `https://raphaeltinarrage.github.io/EMAp.html`

Conclusão

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



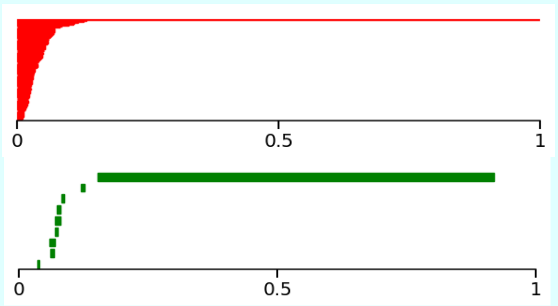
nuvem de pontos



filtração de Čech

$\mathbb{V}$

módulo de persistência



código de barras

Permite analisar os dados a partir de uma nova perspectiva.

Um curso de TDA: `https://raphaeltinarrage.github.io/EMAp.html`

Obrigado!