

HOMOLOGÍA PERSISTENTE

Esneider Sierra

Resumen

La homología persistente intenta capturar la evolución que tienen las conexiones de una nube de puntos. La idea es partir de una colección finita de datos (puntos en \mathbb{R}^n) e ir generando conexiones, esto es, generar una colección de complejos simpliciales de manera que el siguiente contenga al anterior (una sucesión creciente cuya relación es la contención de subcomplejos); capturar la evolución consiste en calcular los grupos de homología de cada complejo simplicial y guardar esa información en un diagrama de persistencia, hacer esto permite conocer fracturas en los datos y con ello eliminar ruido o elementos que, según el contexto en el que se trabaje, no son relevantes.

Palabras clave: nube de puntos, homología, diagrama de persistencia.

1. Teorema del Nervio

1.1. Un ejemplo como motivación

Algo que es común en el análisis de datos es considerar grandes conjuntos que almacenan información que no necesariamente se puede visualizar. Podemos considerar información que está contenida en la configuración de las uniones de proteínas (para más detalles consultar [3] y [5]), inicialmente no tenemos idea de cuán dispersos están los datos o si albergan conexiones. Para generar conexiones se puede aplicar una filtración de Vietoris-Rips y se usa una distancia dinámica ($D_{i,j} = 1 - |C_{i,j}|$ donde $C_{i,j}$ es la componente (i, j) -ésima de una matriz de correlación). Al hacer esto podemos calcular grupos de homología de cada complejo simplicial en la filtración y generar un diagrama de persistencia que contiene la información de cuándo se generan "huecos" d -dimensionales y cuándo desaparecen.

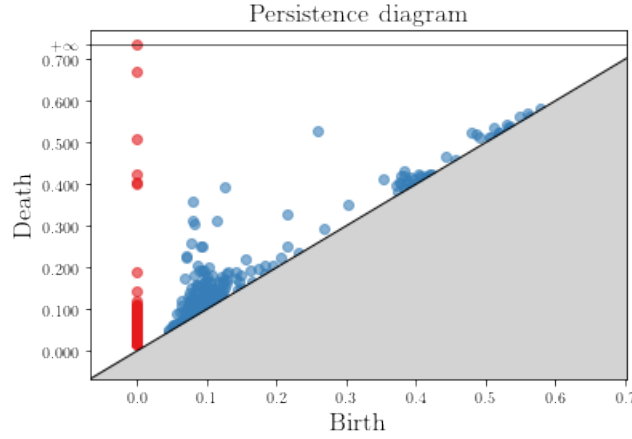


Figura 1: Diagrama de persistencia tomado de [3]

En este diagrama tenemos que los huecos de dimensión 0 (las componentes conexas) se representan en rojo y los de dimensión 1 en azul. Los huecos que mueren muy rápido, puede ser porque se llena el espacio al formarse un símplex de dimensión 2 en la filtración, están cerca de la diagonal y quienes sobreviven por más tiempo se encuentran lejos de ella.

1.2. Nervio

Para empezar a generar la sucesión de complejos simpliciales que se relacionan mediante la contención de subcomplejos simpliciales, podemos empezar con la definición de *nervio*:

Definición 1. Considere $F = \{F_1, F_2, \dots, F_n\}$ una colección finita de conjuntos. Definimos el *nervio* de F como todas las subcolecciones no vacías cuyos conjuntos tienen intersección común no vacía, esto es:

$$Nrv(F) = \{X \subseteq F : \bigcap X \neq \emptyset\}$$

Nota: Si $\sigma = \{F_i\}_{i \in I}$ está en $Nrv(F)$ y $\tau = \{F_j\}_{j \in J}$ es tal que $\tau \subseteq \sigma$ entonces $\bigcap_{j \in J} F_j \subseteq \bigcap_{i \in I} F_i \neq \emptyset$ de modo que $\bigcap_{j \in J} F_j \neq \emptyset$ pues $\tau \neq \emptyset$, y así $\tau \in Nrv(F)$. Esto nos dice que $Nrv(F)$ es un complejo simplicial.

1.3. Teorema del Nervio

En la Figura 2, la unión de los conjuntos no es homotópicamente equivalente al nervio del conjunto, además los conjuntos F_i no son convexos. Para que podamos generar conexiones en nuestra nube de

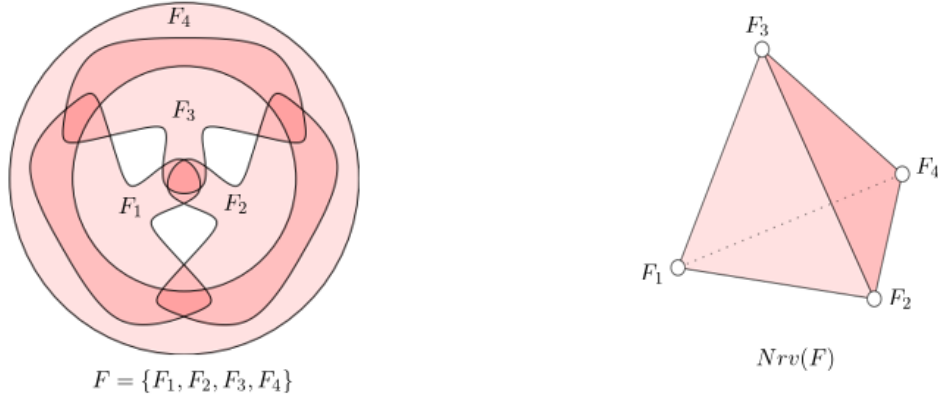


Figura 2: Note que $\bigcup F$ es homotópico a un disco con tres huecos, y $Nrv(F) \simeq S^2$.

puntos y que la homotopía se preserve con uniones necesitaremos el siguiente teorema:

Teorema 1. (*Teorema del Nervio*) Si F es una colección de conjuntos cerrados y convexos en un espacio métrico. Entonces el nervio de F y la unión de los conjuntos en F tienen el mismo tipo de homotopía.

2. Filtraciones

2.1. Filtración de Čech

Como queremos generar conexiones entre puntos, y de algún modo capturar los huecos que pueden darse en esas conexiones, podemos considerar un radio r y generar bolas cerradas por cada punto en la nube de puntos. Luego, para obtener un complejo simplicial de esa construcción y asegurarnos de conservar la homotopía, nos queda aplicar el nervio a la colección de esas bolas.

Definición 2. Sea S un conjunto finito de puntos en \mathbb{R}^n , el **complejo de Čech** de S y r está dado por el nervio de la colección:

$$F = \{B_r(x) : x \in S\}$$

lo que es equivalente a considerar

$$\check{Cech}(r) = \{\sigma \subseteq S : \bigcap_{x \in \sigma} B_r(x) \neq \emptyset\}.$$

Nota: dado un conjunto finito de puntos en \mathbb{R}^n , si $r_0 \leq r_1$ entonces $\check{Cech}(r_0)$ es un subcomplejo de $\check{Cech}(r_1)$, de modo que si tenemos una sucesión finita de radios $\{r_i\}$ tendremos una filtración de Čech:

$$\check{Cech}(r_0) \subseteq \check{Cech}(r_1) \subseteq \cdots \subseteq \check{Cech}(r_t).$$

2.2. Filtración de Vietoris-Rips

Para nuestros propósitos, podemos considerar otro tipo de filtraciones. La elección de cuál filtración se debería elegir no es algo sencillo de determinar, dependerá del contexto en el que estemos trabajando o de la naturaleza de los datos que exploremos. En nuestro ejemplos usaremos filtraciones de *Vietoris-Rips*, que definiremos en seguida.

Definición 3. Sea S un conjunto finito de puntos en \mathbb{R}^n , el **complejo de Vietoris-Rips** de S y r está dado por

$$VR(r) = \{\sigma \subseteq S : \text{diam}(\sigma) \leq 2r\}.$$

Una manera de pensar estos complejos, es considerar una colección finita de conjuntos (bolas en \mathbb{R}^n centradas en los puntos de nuestra nube de puntos) pero en lugar de subcolecciones con intersección común no vacía, podemos considerar que las intersecciones entre parejas de nuestra colección tenga intersección no vacía y cuando eso suceda agregar un complejo simplicial de dimensión dos o más, según corresponda. Esto nos dice que, para un conjunto finito de puntos S tenemos que $\check{Cech}(r) \subseteq VR(r)$. La Figura 3 ilustra la diferencia entre la filtración de Čech y Vietoris-Rips.

Nota: si tenemos una colección de radios $\{r_i\}$ tendremos una filtración de Vietoris-Rips:

$$VR(r_0) \subseteq VR(r_1) \subseteq \cdots \subseteq VR(r_t).$$

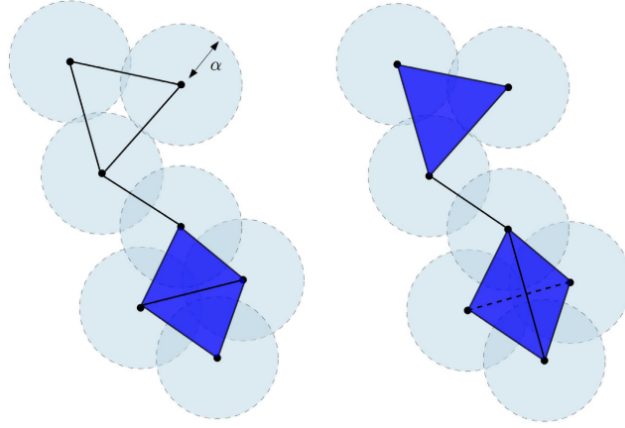


Figura 3: A la izquierda la filtración de Čech y a la derecha la filtración de Vietoris-Rips

3. Grupos de Homología Persistente

3.1. Objetos de Persistencia

Hasta el momento, para un conjunto finito de puntos S y una colección finita y creciente de radios $\{r_i\}$, podemos construir una filtración (ya sea con complejos de Čech o de Vietoris-Rips):

$$F_0K \subseteq F_1K \subseteq \dots \subseteq F_nK \subseteq \dots \subseteq F_mK \subseteq \dots \subseteq F_tK$$

y considerando las inclusiones $i^n : F_nK \rightarrow F_{n+1}K$ tenemos la sucesión inducida por el funtor de complejos de cadenas simpliciales:

$$C(F_0K) \xrightarrow{C(i^0)} C(F_1K) \xrightarrow{C(i^1)} C(F_2K) \xrightarrow{C(i^2)} \dots \xrightarrow{C(i^{t-1})} C(F_tK) \quad (\alpha)$$

de modo que podemos aplicar el funtor de homología y obtener, en cada grado $k \in \mathbb{N}$, la sucesión de R -módulos:

$$H_k(F_0K) \xrightarrow{H_k(i^0)} H_k(F_1K) \xrightarrow{H_k(i^1)} H_k(F_2K) \xrightarrow{H_k(i^2)} \dots \xrightarrow{H_k(i^{t-1})} H_k(F_tK) \quad (\beta)$$

y podemos notar que $H(i^n)$. Ahora bien, lo que nos interesa en este punto es capturar la evolución de las conexiones que generamos con la filtración y para esto consideremos las siguientes definiciones:

Definición 4. *Un complejo de cadenas de persistencia (C^\bullet, f^\bullet) es una sucesión*

$$C^0 \xrightarrow{f^0} C^1 \xrightarrow{f^1} C^2 \xrightarrow{f^2} \dots$$

donde cada C^n es un complejo de cadenas.

Un módulo de persistencia (H^\bullet, f^\bullet) es una sucesión de R -módulos y mapas de la forma

$$H^0 \xrightarrow{f^0} H^1 \xrightarrow{f^1} H^2 \xrightarrow{f^2} \dots$$

En (α) y (β) tenemos ejemplos de un complejo de cadenas de persistencia y un de módulo de persistencia. Para fines prácticos, la función $f^{n \rightarrow m} : C^n \rightarrow C^m$ se define como

$$f^{n \rightarrow m} := f^{m-1} \circ \dots \circ f^{n+1} \circ f_n.$$

3.2. Grupos de Homología Persistente

Dada una filtración $(F_\bullet K, f^\bullet)$ se obtiene un complejo de cadenas de persistencia $(C^\bullet, C(f^\bullet))$, y fijado $k \in \mathbb{N}$ se aplica el funtor de homología en el complejo de cadenas de persistencia para obtener un módulo de persistencia $(H^\bullet, H_k(C(f^\bullet)))$.

Combinando las anteriores definiciones y con el propósito de responder la pregunta de cómo evolucionan las clases en homología a medida que avanzamos en la filtración, añadimos la siguiente definición:

Definición 5. *Para cada par de enteros $n \leq m$, el grupo de homología persistente $PH_k^{n \rightarrow m}(C^\bullet, f^\bullet)$ de grado $k \in \mathbb{N}$ asociado al complejo de cadenas de persistencia (C^\bullet, f^\bullet) es el submódulo de $H_k(C^m)$ dado por*

$$\begin{aligned} PH_k^{n \rightarrow m}(C^\bullet, f^\bullet) &:= \frac{f_k^{n \rightarrow m}(\ker \partial_k^n)}{\text{Im}(\partial_{k+1}^m) \cap f_k^{n \rightarrow m}(\ker \partial_k^n)} \\ &= f_k^{n \rightarrow m}(H_k(C^n)) \\ &\subseteq H_k(C^m) \end{aligned}$$

donde $H_k(C^n) = \ker(\partial_k^n) / \text{Im}(\partial_{k+1}^n)$.

Para tener una mejor idea de cómo funciona la definición cuando el complejo de cadenas persistente está dado por un filtración $(F_\bullet K, f^\bullet)$, considere el siguiente diagrama:

$$\begin{array}{ccccccc}
 \vdots & & \vdots & & \vdots & & \\
 \partial_{k+1}^{n-1} \downarrow & & \partial_{k+1}^n \downarrow & & \partial_{k+1}^m \downarrow & & \\
 \cdots \longrightarrow & C_k(F_{n-1}K) & \xrightarrow{f_k^{n-1}} & C_k(F_nK) & \xrightarrow{f_k^{n \rightarrow m}} & C_k(F_mK) & \longrightarrow \cdots \\
 \partial_k^{n-1} \downarrow & & \partial_k^n \downarrow & & \partial_k^m \downarrow & & \\
 \cdots \longrightarrow & C_{k-1}(F_{n-1}K) & \xrightarrow{f_{k-1}^{n-1}} & C_{k-1}(F_nK) & \xrightarrow{f_{k-1}^{n \rightarrow m}} & C_{k-1}(F_mK) & \longrightarrow \cdots \\
 \partial_{k-1}^{n-1} \downarrow & & \partial_{k-1}^n \downarrow & & \partial_{k-1}^m \downarrow & & \\
 \vdots & & \vdots & & \vdots & &
 \end{array}$$

Y al igual que en homología simplicial, podemos definir números de Betti en homología persistente:

Definición 6. El número de Betti (n, m) -persistente de (C^\bullet, f^\bullet) de grado k se define como:

$$\beta_k^{n,m}(C^\bullet, f^\bullet) := \text{rank}(PH_k^{n \rightarrow m}(C^\bullet, f^\bullet)).$$

Si tenemos la filtración de *tipo finito* (después de un cierto N los complejos simpliciales son iguales) dada por una nube de puntos

$$F_0K \subseteq F_1K \subseteq \cdots \subseteq F_nK \subseteq \cdots \subseteq F_mK \subseteq \cdots$$

tendremos la iteración en el complejo de cadenas de persistencia

$$\cdots \rightarrow C_k(F_nK) \xrightarrow{C_k(i^{n \rightarrow m})} C_k(F_mK) \rightarrow \cdots$$

donde $i^{n \rightarrow m} : F_nK \hookrightarrow F_mK$ y simplificaremos la notación

$$PH_k^{n \rightarrow m}(C(F_\bullet K), C(i^\bullet)) = PH_k^{n \rightarrow m}$$

y en este caso $PH_k^{n \rightarrow m} = \text{Im}(H_k i^{n \rightarrow m}) \subseteq H_k(F_mK)$.

Además, si $n' \leq n$ entonces $PH_k^{n \rightarrow m} \subseteq PH_k^{n' \rightarrow m}$.

3.3. Persistencia

Definición 7. Si γ es una clase en $H_k(F_n K) = PH_k^{n \rightarrow n}$, decimos que **nace** en $F_n K$ si $\gamma \notin PH_k^{n-1 \rightarrow n}$.

Si γ nace en $F_n K$ entonces **muere entrando** en $F_m K$ si se fusiona con una clase anterior yendo de $F_{m-1} K$ a $F_m K$, es decir, si $i_k^{n \rightarrow m-1}(\gamma) \notin PH_k^{n-1 \rightarrow m-1}$ pero $i_k^{n \rightarrow m-1}(\gamma) \in PH_k^{n-1 \rightarrow m}$.

Para tener una mejor idea de cómo funciona la definición anterior, considere el siguiente diagrama:

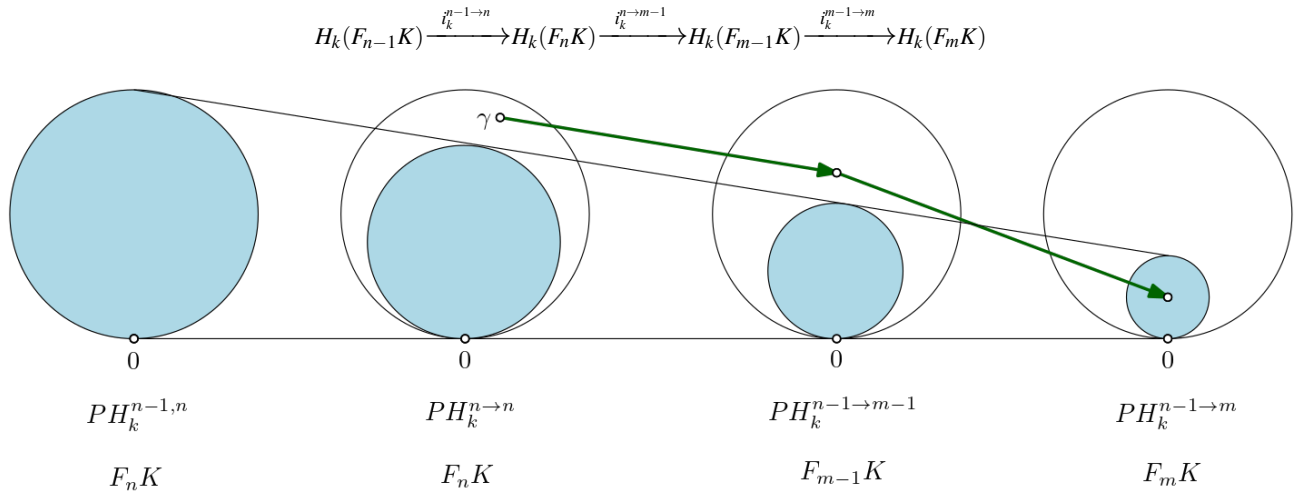


Figura 4: $\gamma \in H_k(F_n K)$ nace en $F_n K$ y muere entrando en $F_m K$

Lo que nos interesa en este punto es determinar cuánto tiempo de "vida útil" tiene una clase, de ahí el porqué de la palabra persistencia.

Definición 8. Si una clase γ nace en $F_n K$ y muere entrando en $F_m K$ entonces su **índice de persistencia** es $m - n$.

4. Diagramas de Persistencia y Ejemplos

Los diagramas de persistencia pueden ser definidos formal y rigurosamente, pero en este caso daremos una manera aproximada de hacerlo: nos interesa calcular ejemplos y almacenar la información de los índices de persistencia en cada clase. Con esto en mente iniciemos con un ejemplo de un diagrama de barras que se transforma en un diagrama de persistencia:

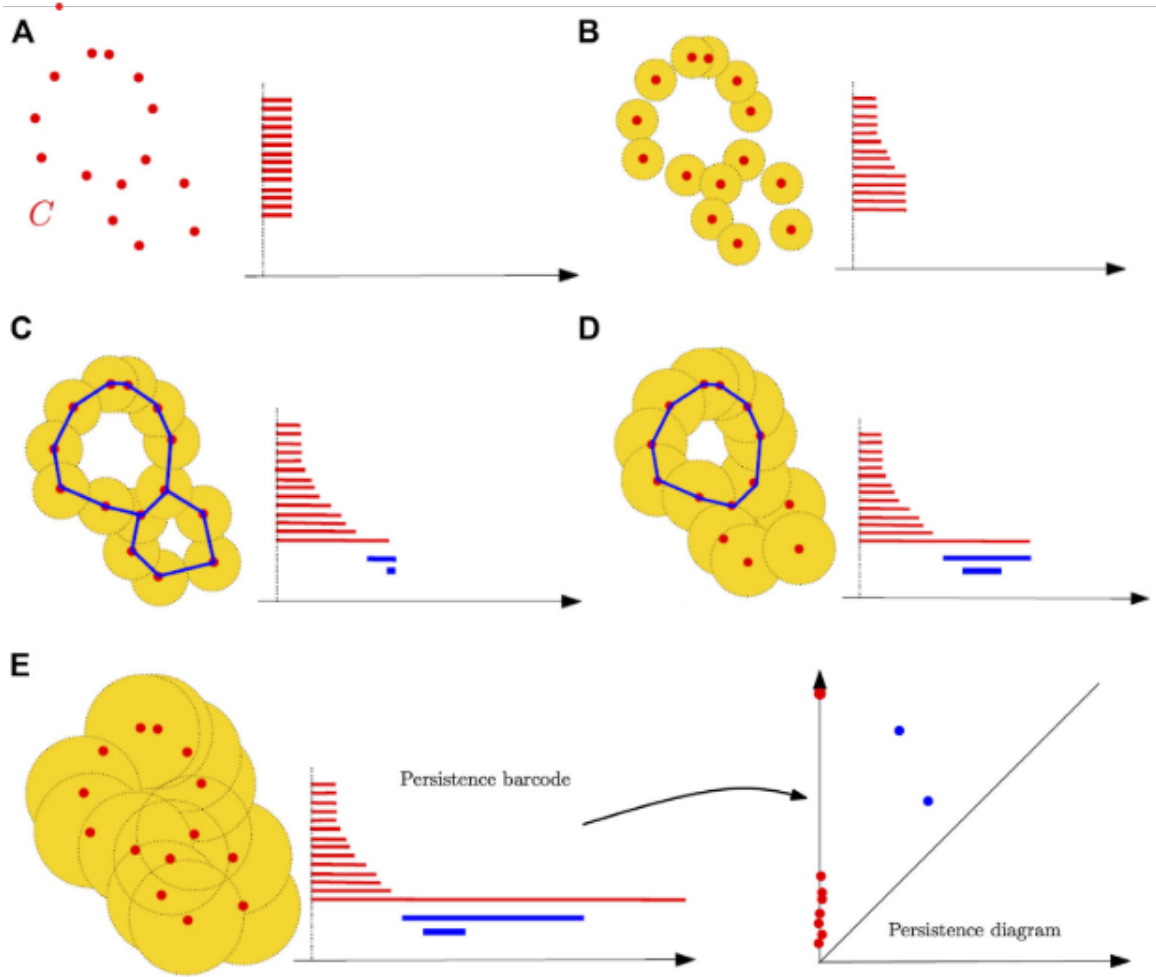


Figura 5: Este ejemplo fue tomado de [3]. Las barras rojas indican que se está calculando $PH_0^{i \rightarrow j}$, las barras azules $PH_1^{i' \rightarrow j'}$ y si hay verdes corresponderán a dimensión 2 en homología. En la parte **A** tenemos una barra roja por cada componente conexa, en este caso el radio empieza en $r = 0$. En la parte **B** tenemos que el radio aumentó al punto de que dos o más bolas tienen intersección no vacía y la primera de esas intersecciones representará a todas las componentes conexas que empiecen a unirse a ella, de ahí que la regla que se siga sea la de los ancianos (Elder Rule), en este sentido hay barras que dejaron de crecer, pues murieron en el momento en que se unieron a una componente conexa anciana. Note que en la parte **C** solo hay una componente conexa y las barras rojas han dejado de crecer, pues las demás componentes conexas terminaron muriendo cuando se fueron uniendo a las demás; además se puede notar que en los grupos de homología de dimensión 1 tendremos dos generadores, luego surgirán dos barras azules y podemos notar, en la parte **D**, un 1-símplex pasó a ser un 2-símplex con el mismo tipo de homotopía de D^1 . En la parte **E** solo sobrevivió la componente conexa representada por la barra roja. Luego pasamos al diagrama de persistencia, en donde el eje x representa el nacimiento de las clases y la distancia de la diagonal a los puntos de color representa el tiempo de vida, o el índice de persistencia, de las clases en grupos de homología.

Referencias

- [1] H. Edelsbrunner and J. Harer, *Computational Topology: An Introduction*, Duke University Press, 2008.
- [2] Vidit Nanda, *Computational Algebraic Topology - Lecture Notes*.
- [3] F. Chazal and B. Michel, *An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists*, Front. Artif. Intell., 2021.
- [4] Clara Löh, *Applied Algebraic Topology*, Universität Regensburg, 2023.
- [5] GUDHI Project, *TDA Tutorial*, Disponible en: <https://github.com/GUDHI/TDA-tutorial>, 2024.

Para los ejemplos de nuestros espacios topológicos favoritos puede consultar el cuaderno de pluto:
<https://github.com/SneikF/PersistenceHomologyExamples.git>