



FACULTAD DE MATEMÁTICAS

DEPARTAMENTO DE GEOMETRÍA Y TOPOLOGÍA

GRADO EN MATEMÁTICAS

TRABAJO FIN DE GRADO

**Sobre las diferentes nociones topológicas de complejo
y algunas de sus aplicaciones**

Laura Núñez de Arco Valenzuela

Tutores:
Rafael Ayala y José Antonio Vilches

Índice

1. Introducción	3
2. Nociones básicas sobre topología	5
3. Complejos simpliciales	7
3.1. Complejos clique	12
4. Complejos cúbicos. Complejos celulares	14
5. Nociones básicas sobre homología	17
5.1. Homología persistente	18
6. Nervio de un recubrimiento	23
7. Complejos de Čech	24
8. Complejos de Vietoris-Rips	26
8.1. Comparación con el complejo Čech	27
8.2. Homología persistente con complejos anidados	30
9. Otros complejos simpliciales	33
9.1. Complejos de Delaunay	33
9.2. Complejos Alfa	38
9.3. Complejos Witness	41
9.3.1. Relación entre witness débil y fuerte	44
9.3.2. Generalizaciones	46
9.3.3. Criterio para elegir los puntos de referencia	48

Resumen

Abstract: The main goal of this work is to introduce the different kinds of combinatorial structures used in topological data analysis. It is focused on a detailed description of certain types of simplicial, cubical and cellular complexes, namely flag complexes, Čech complexes, Vietoris-Rips complexes, Witness complexes, Delaunay complexes and Alpha complexes will be considered. Additionally, persistent homology will be introduced as the main tool to extract relevant topological information from data sets via the above mentioned kinds of complexes.

1. Introducción

En este trabajo se pretende exponer de forma muy breve algunos de los conceptos y métodos topológicos empleados en el Análisis Topológico de Datos, disciplina surgida hace unos veinte años con el propósito de complementar las herramientas estadísticas clásicas usadas en el análisis de conjuntos masivos de datos, como por ejemplo, la minería de datos o los diversos métodos de clustering.

Puede decirse que hoy en día generan continuamente enormes cantidades de datos en cualquier ámbito de la actividad humana: finanzas, industria, comercio, comunicaciones, ingeniería, ciencia, medicina, etc. Dichos datos pueden ser muy complejos, no solo por su volumen, sino por el gran número de atributos o características que miden, y que gracias al desarrollo de la informática es posible guardarlos y procesarlos a gran velocidad.

El problema que se plantea consiste esencialmente en determinar qué propiedades de esos datos son de interés para obtener una interpretación adecuada de los mismos, que permitan obtener conclusiones válidas sobre el fenómeno, proceso o hipótesis que los origina.

Si hubiera que resumir en una sola frase qué distingue al Análisis Topológico de Datos del resto de las ramas de la Estadística que se ocupan del análisis de datos, sería su interés por obtener formas o patrones que se puedan detectar en dichos conjuntos. Por ejemplo, suponiendo que se establece un criterio de proximidad entre sus puntos, ¿es posible detectar si parte de ellos se distribuyen a lo largo de una curva, un cierto tipo de grafo o una superficie?. Es en el análisis de estas propiedades geométricas cualitativas en las que, por su propia naturaleza, la Topología desempeña un papel destacado, pues su objetivo es el estudio de las propiedades de las figuras que son independientes de las coordenadas que se usen, y que se mantengan al aplicarle pequeñas deformaciones continuas.

Ahora bien, el Análisis Topológico de Datos no solo hace uso de herramientas de la Topología, sino que se ayuda del Álgebra, la Geometría Computacional o la Probabilidad para extraer información de una nube de datos, y después, por medio de diversas técnicas de computación, se realiza un estudio estadístico de la información obtenida. Esta combinación de métodos de diversas áreas supone un poderoso recurso, pero también hacen del Análisis Topológico de Datos una materia complicada y difícil de estudiar. Por ello en esta memoria nos hemos limitado a una exposición muy superficial de sus elementos básicos, haciendo énfasis en los aspectos topológicos: la descripción de ciertos complejos simpliciales que dependen de un parámetro y que aproximan la forma de los datos a distintos niveles, así como las primeras nociones de la homología persistente de las filtraciones inducidas por dichos complejos. Es esta una torre formada por los espacios vectoriales de homología de cada nivel de la filtración, y que permite detectar qué propiedades se mantienen al ir avanzando en ella, y por tanto son dignas de ser analizadas e interpretadas, y cuáles son las efímeras, y por tanto no relevantes, probablemente debidas a ciertos defectos del conjunto de datos.

La Memoria se ha dividido en secciones con los siguientes contenidos: la Sección 2 recoge los conceptos básicos de topología utilizados, y en las Secciones 3 y 4 se presentan las definiciones y propiedades elementales de los complejos simpliciales y algunas de sus variantes. En la Sección 5 se exponen las definiciones de homología persistente, código de barras y diagrama de persistencia. Las Secciones 7 y 8 tratan los complejos de Čech y de Vietoris-Rips, y por último, en la Sección 9 se introducen los complejos alfa y los complejos testigos (witness), que han sido propuestos para reducir la complejidad de los complejos de Vietoris-Rips, y así facilitar un cálculo eficiente, desde el punto de vista computacional, de la homología simplicial.

2. Nociones básicas sobre topología

Comenzaremos formalizando la noción de espacio topológico. Luego, introduciremos los complejos simpliciales, y la estructura combinatorial que usaremos para representarlos.

Intuitivamente, un espacio topológico es un conjunto de puntos con una cierta estructura de proximidad que permite definir de manera formal a la convergencia, continuidad y conectividad, entre otros conceptos.

Definición 2.1. Una **topología** en un conjunto X es un subconjunto $T \subseteq 2^X$ tal que:

- Si $S_1, S_2 \in T$, entonces $S_1 \cap S_2 \in T$.
- Si $\{S_j | j \in J\} \subseteq T$, entonces, $\cup_{j \in J} S_j \in T$.
- $\emptyset, X \in T$.

El par $X = (X, T)$ se llama **espacio topológico**, y los elementos de la topología T se llaman conjuntos abiertos de dicho espacio topológico.

Definición 2.2. Un conjunto $S \in T$ se dice **cerrado** si su complementario en X es abierto.

Definición 2.3. Un subconjunto $A \subseteq X$ con topología inducida $T_A = \{S \cap A | S \in T\}$ es un **subespacio** \mathbb{A} de X .

Un ejemplo familiar de un espacio topológico es el espacio euclídeo d -dimensional \mathbb{R}^d , donde usaremos la métrica euclídea para medir distancias y definir conjuntos abiertos. Además, podremos convertir cualquier subconjunto del espacio euclídeo en un espacio topológico usando la topología inducida.

Una vez que han sido definidos objetos dotados de una cierta estructura, resulta natural considerar aplicaciones entre ellos compatibles con la estructura de los mismos. En nuestro caso, homeomorfismos entre espacios topológicos.

Definición 2.4. Sean (X, T) e (Y, T') dos espacios topológicos y sea $f : X \rightarrow Y$ una aplicación entre ellos. Diremos que f es **continua** en un punto $a \in X$ si para todo entorno U de $f(a)$, existe un entorno V de a tal que $f(V) \subset U$.

Definición 2.5. Un **homeomorfismo** $f : X \rightarrow Y$ es una biyección tal que ambas f y f^{-1} son continuas. Dado un homeomorfismo $f : X \rightarrow Y$, decimos que X es homeomorfo a Y . La cualidad de ser homeomorfo es una relación de equivalencia para espacios topológicos.

Definición 2.6. Sean X e Y espacios topológicos y $A \subset X$. Si $f, g : X \rightarrow Y$ son aplicaciones continuas, tales que $f|_A = g|_A$, se dice que f y g son homótopas a A , si existe una aplicación continua, $H : X \times [0, 1] \rightarrow Y$, que denominaremos **homotopía**, tal que

$$(I) \quad H(x, 0) = f(x), \text{ para cada } x \in X.$$

$$(II) \quad H(x, 1) = g(x), \text{ para cada } x \in X.$$

Definición 2.7. En las condiciones anteriores, si H además de ser homotopía cumple

$$H(x, t) = f(x) = g(x), \text{ para cada } x \in A \text{ y } t \in [0, 1]$$

se tiene que H es **homotopía relativa**, y que f y g son homótopas relativamente a A .

Definición 2.8. Una aplicación $f : X \rightarrow Y$ es una **equivalencia de homotopía** si existe una aplicación $g : X \rightarrow Y$, tal que $f \circ g \simeq 1_Y$ y $g \circ f \simeq 1_X$, donde $1_X, 1_Y$ son las aplicaciones identidad en los correspondientes espacios.

Dada una equivalencia de homotopía $f : X \rightarrow Y$, decimos que X e Y son homotópicamente equivalentes o que tienen el mismo tipo de homotopía, y se denota por $X \simeq Y$. Se verifica que ser del mismo tipo de homotopía es una relación de equivalencia entre espacios topológicos.

Definición 2.9. Un espacio con el tipo de homotopía de un punto se dice **contráctil**.

3. Complejos simpliciales

Los complejos simpliciales son útiles en el análisis de datos topológicos porque permiten introducir modelos finitos o combinatorios de espacios topológicos adaptados a los conjuntos de datos.

Veamos primero una definición para complejo simplicial de un modo combinatorio, para verlo luego con un enfoque más geométrico.

Definición 3.1. Sea S un conjunto finito. Un **complejo simplicial abstracto** es una colección finita K de subconjuntos de S , cerrados para la relación de inclusión, es decir: para cada σ en K , todos los subconjuntos de σ están también en K , es decir, si $\sigma \in K$ y $\tau \subset \sigma$, entonces $\tau \in K$.

Cada elemento $\sigma \in K$, se llama *símplice*, o **q -símplice** si $\text{card}(\sigma) = q + 1$. Si $\tau \subset \sigma$ se denomina *cara del símplice*, y se denota $\tau \leq \sigma$. Un q -símplice tiene 2^{q+1} caras, contando el vacío y al propio símplice.

Ejemplo 3.1. Si $S = \{s_i\}$, $i \in I$, el subconjunto $\{s_2, s_5, s_7\}$ es un 2-símplice, con:

- tres caras que son, a su vez 1-símplice: $\{\{s_2, s_5\}, \{s_2, s_7\}, \{s_5, s_7\}\}$,
- tres caras que son 0-símplice: $\{\{s_2\}, \{s_5\}, \{s_7\}\}$,
- y por supuesto, la cara del subconjunto vacío: $\{\emptyset\}$.

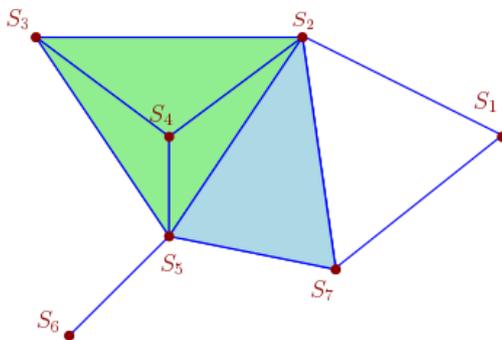


Figura 1. Representación del Ejemplo 3.1

Definición 3.2. Un complejo simplicial abstracto cuyos símplices tienen dimensión menor o igual que uno, se denomina **grafo abstracto**.

De hecho, los complejos simpliciales más familiares son los grafos abstractos; estos, usualmente se presentan como un par (V, E) , donde V es el conjunto no vacío de los vértices (ó 0-símplices) y E es una colección de parejas no ordenadas en V .

V se denomina conjunto de vértices o nodos y E se llama conjunto de aristas o enlaces. Estas aristas son los 1-símplices del grafo. Una arista se presenta como un par no ordenado $\{u, v\}$ de vértices, con $u, v \in V$. En es caso, los vértices se dicen vecinos o adyacentes y la arista se dice incidente con los vértices.

Los grafos abstractos son ideales para expresar relaciones binarias entre objetos. Las relaciones de orden superior nos llevan a los complejos simpliciales.

Ahora veamos un enfoque geométrico en lugar de combinatorio; recordemos la definición usual de símplice y de complejo simplicial.

Definición 3.3. Se dice que los puntos $\{p_0, p_1, \dots, p_r\} \subset \mathbb{R}^n$ son **afínmente independientes** si los vectores $p_1 - p_0, p_2 - p_0, \dots, p_r - p_0$ son linealmente independientes.

Definición 3.4. Un conjunto es convexo si dados dos puntos cualesquiera, el segmento que los une está contenido en el conjunto. Entonces, se define la **envolvente convexa** de un conjunto P como la intersección de todos los convexos que contienen al conjunto P .

Definición 3.5. Si $\{x_0, \dots, x_q\} \subset \mathbb{R}^n$ es un conjunto de $q + 1$ puntos afínmente independientes, a la envolvente convexa de estos puntos se le llama símplice q -dimensional o **q -símplice** y se denota por (x_0, \dots, x_q) .

Así, un q -símplice se llama vértice, arista, triángulo o tetraedro para $q = 0, 1, 2, 3$, respectivamente.

Nota 3.1. El q -símplice estándar Δ^q es:

$$\Delta^q := \{x_i \geq 0, \quad i = 0, \dots, q \quad \sum_{j=0}^q x_j = 1\}$$

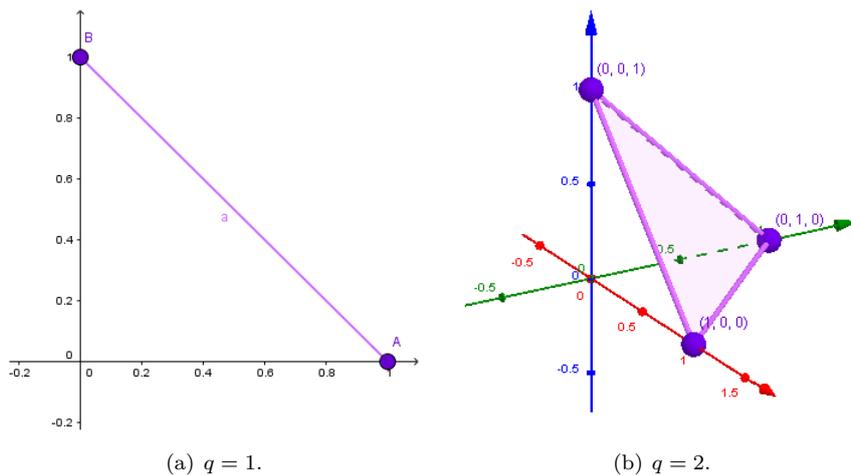


Figura 2. q -símplice estándar.

Se verifica que todo símplice es la imagen por una aplicación afín de un símplice estándar.

Las caras del símlice Δ^q son copias de Δ^j para $j \leq q$ a través de la restricción a \mathbb{R}^{q+1} , haciendo la intersección con subespacios obtenidos anulando $n - q$ coordenadas, con n la dimensión del espacio ambiente. A través de esta restricción, podemos ver cómo estas caras están unidas a lo largo de la frontera de Δ^q .

Se puede considerar que un complejo simplicial K es un espacio producido por una copia de Δ^q para cada q -símlice de K , unidos por algunas de sus caras.

Definición 3.6. Llamamos **complejo simplicial geométrico finito** a una colección finita K de símlices en algún \mathbb{R}^n verificando:

- (I) Si $\sigma_1, \sigma_2 \in K$ entonces $\sigma_1 \cap \sigma_2 = \emptyset$ ó $\sigma_1 \cap \sigma_2$ es una cara común de σ_1 y σ_2 .
- (II) Si $\sigma \in K$ y $\tau \leq \sigma$ entonces $\tau \in K$.

Nota 3.2. Todas las definiciones anteriores se pueden considerar también para complejos simpliciales geométricos con un conjunto infinito de símlices.

En esta definición vemos que los símlices intersecan en caras comunes, y en la anterior, esto se tiene por construcción de los q -símlices, en la que los vamos construyendo a partir de $(q - 1)$ -símlices, identificando caras.

La noción de complejo simplicial abstracto formaliza las propiedades que cumple el conjunto de símlices de un complejo simplicial.

Definición 3.7. La **dimensión** de K es $dim(K) = \max\{dim(\sigma); \sigma \in K\}$.

Llamaremos grafo geométrico a todo complejo simplicial geométrico K tal que $dim(K) = 1$.

Definición 3.8. Diremos que $K' \subseteq K$ es un **subcomplejo** si es un conjunto de símlices de K que es por sí mismo un complejo simplicial.

Dado un grafo G , diremos que G' es un subgrafo de G si es subcomplejo de éste.



Figura 3. K' subcomplejo de K .

Definición 3.9. Se llama **q -esqueleto** de K y se denota K^q al subcomplejo

$$K^q = \{\sigma \in K; dim(\sigma) \leq q\}.$$

Diremos que K^0 es el conjunto de vértices de K .

Como ocurre siempre en Matemáticas, una vez que han sido definidos objetos dotados de una cierta estructura, se pasa a considerar aplicaciones entre ellos compatibles con la estructura de los mismos. En nuestro caso, definimos las aplicaciones simpliciales como aquellas que respetan la estructura simplicial. Es decir, transforma símlices en símlices. Más explícitamente:

Definición 3.10. Sean K_1 y K_2 complejos simpliciales y φ una aplicación definida entre los vértices de K_1 y K_2 . Se dice que φ es **aplicación simplicial** si dado un símlice $\sigma \in K$ con $\sigma = (v_0, \dots, v_n)$, los vértices $\varphi(v_0), \dots, \varphi(v_n)$ están en un mismo símlice de K_2 .

Definición 3.11. Un **isomorfismo simplicial** entre dos complejos simpliciales K_1 y K_2 es una biyección φ entre los vértices tal que (v_0, \dots, v_n) es un símlice de K_1 si y sólo si $(\varphi(v_0), \dots, \varphi(v_n))$ es un símlice de K_2 .

Definición 3.12. El conjunto de los puntos de los símlices de K se denomina **poliedro subyacente** a K , y lo denotaremos por $|K|$, es decir:

$$|K| = \cup\{\sigma; \sigma \in K\} \subseteq \mathbb{R}^n$$

Dotamos de topología al complejo simplicial K como un espacio cociente construido a partir de símlices geométricos, identificando algunos de ellos a través de caras comunes. La topología euclídea sobre $|K|$ coincide con la topología débil asociada a los símlices o a sus esqueletos. Por tanto, un conjunto es abierto si al cortarlo con cada símlice lo es.

Nota 3.3. El q -esqueleto de K , $q \in \mathbb{N}$, también puede verse como el espacio cociente:

$$|K^q| := \left(|K^{q-1}| \cup \coprod_{\sigma: \dim \sigma = q} \Delta^q \right) / \sim$$

Donde \sim es la relación de equivalencia que identifica las caras de Δ^q con las correspondientes caras combinatorias de $\sigma \in |K^j|$ para $j < q$.

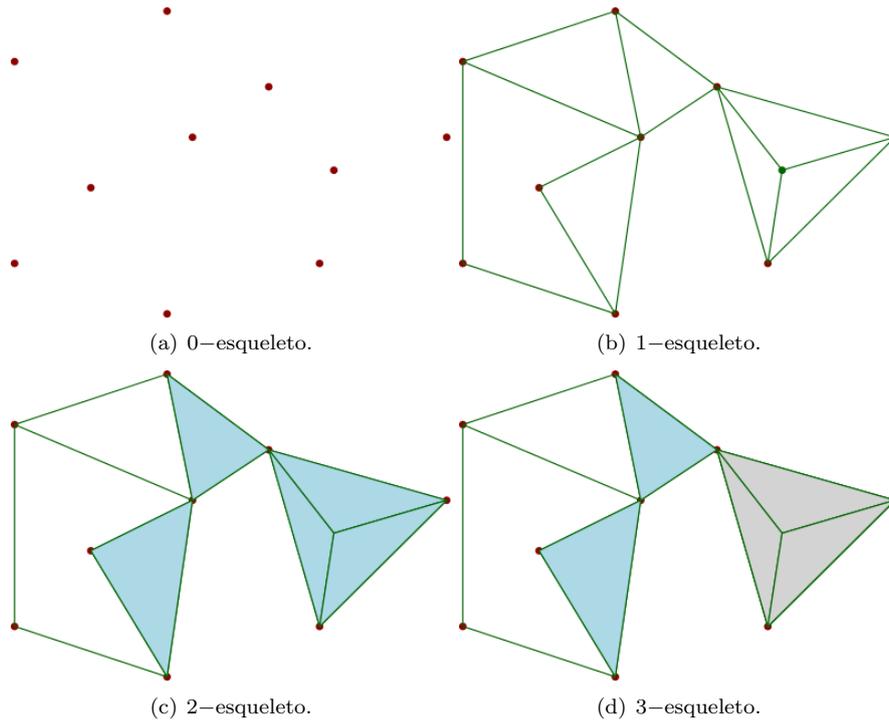


Figura 4. Se representan los 0-simplices como puntos rojos, los 1-simplices como segmentos verdes, los 2-simplices como triángulos celestes, y el único 3-simplice como un tetraedro gris.

Luego, $|K^0| = S$ es un conjunto discreto, y el 1-esqueleto $|K^1|$ se corresponde con un grafo.

Por la construcción, $|K^q| \supseteq |K^{q-1}|$ y por lo tanto podemos identificar $|K|$ con $|K| = \bigcup_{q=0}^{\dim(K)} |K^q|$.

Sabemos que los simplices geométricos quedan determinados por sus vértices y, por lo tanto, se pueden ver como simplices abstractos (simplemente, considerando su conjunto de vértices). Más generalmente, si K es un complejo simplicial geométrico en \mathbb{R}^n , entonces identificando los simplices geométricos (v_0, \dots, v_k) con el conjunto finito $\{v_0, \dots, v_k\}$, se obtiene un complejo simplicial abstracto, que denotaremos $\mathcal{A}(K)$.

Recíprocamente, dado un simple abstracto, puede asociársele un simple geométrico en algún espacio euclídeo, eligiendo puntos afinmente independientes correspondientes a sus vértices y tomando su envolvente convexa. Por supuesto esta elección no es única, pero lo es salvo isomorfismo simplicial.

Definición 3.13. Si K es un complejo simplicial abstracto, entonces determina un complejo simplicial geométrico, denotado por $\mathcal{R}(K)$ y llamado **realización geométrica** del complejo K .

Se tiene el siguiente resultado cuya demostración se puede ver en ([38],3.12).

Teorema 3.1. Un complejo simplicial abstracto K de dimensión n tiene una realización geométrica en \mathbb{R}^{2n+1} .

Nota 3.4. Aunque en las triangulaciones de grafos, la cota $2n + 1$ es óptima, en otras, como las de superficies, se puede mejorar dicha cota.

Es importante remarcar la diferencia entre K y $\mathcal{R}(K)$. El primero es un complejo simplicial (abstracto o geométrico), es decir una colección de símlices. El segundo es la unión de esos símlices (visto como un subespacio topológico de \mathbb{R}^n).

En caso de que el complejo simplicial abstracto K no sea finito el resultado anterior es válido si K es numerable y cada vértice está en un número finito de símlices.

Nota 3.5. Dado K un complejo simplicial abstracto se tiene que:

- (a) K es isomorfo a $\mathcal{A}(\mathcal{R}(K))$.
- (b) Si K y K' son isomorfos, entonces $\mathcal{R}(K)$ y $\mathcal{R}(K')$ son isomorfos.

3.1. Complejos clique

Un primer ejemplo de complejo simplicial abstracto que se puede construir a partir de datos combinatorios es el complejo clique o bandera asociado a un grafo.

Definición 3.14. Para cada $n \geq 1$, llamamos **grafo completo** al grafo K_n formado por n vértices y en el que dos vértices cualesquiera distintos siempre están unidos por una arista.

Definición 3.15. Si $G = (V, E)$ un grafo, el **complejo clique** asociado a G se construye introduciendo un q -símplex en el complejo si $q + 1$ vértices forman un subgrafo completo K_q de G , y se denota por $Cl(G)$.

Nota 3.6. Se verifica que $Cl(G)$ es complejo simplicial, ya que un subconjunto de vértices de un subgrafo completo, genera también un subgrafo completo.

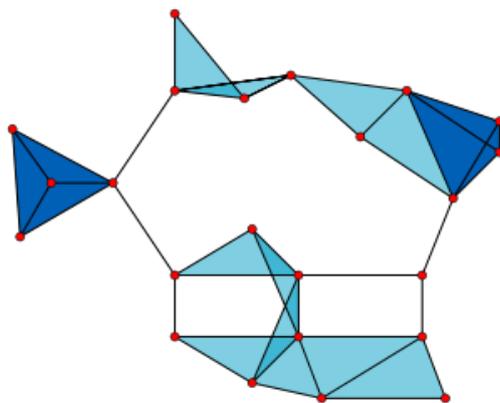


Figura 5. Complejo clique. Se muestran los 0-símlices como puntos rojos, los 1-símlices como segmentos negros, los 2-símlices como triángulos celestes, y los 3-símlices como tetraedros azul oscuro

Una propiedad del complejo clique es que los cliques se corresponden con agrupamientos de nodos altamente conectados. Por ejemplo, en el grafo original, cada triángulo es un ciclo, sin embargo, en el complejo clique los triángulos se rellenan y el ciclo deja de existir. Esto se traduce en que los agujeros detectados en el complejo clique han de tener cuatro o más vértices. Así, el ciclo más simple que se puede crear en el complejo clique consiste en cuatro nodos conectados como un cuadrado sin las conexiones diagonales.

4. Complejos cúbicos. Complejos celulares

Una generalización de los complejos simpliciales son los complejos celulares, que son uniones de conjuntos convexos a través de su frontera.

Definición 4.1. Se llamará **célula convexa** de \mathbb{R}^n a un subconjunto que sea el conjunto de soluciones de un sistema de igualdades y desigualdades de la forma

$$f_i = 0, \text{ para } i = 1, \dots, r$$

$$g_j \geq 0, \text{ para } j = 1, \dots, t$$

donde $f_i, g_j : \mathbb{R}^n \rightarrow \mathbb{R}$ son aplicaciones afines para todo par i, j .

Definición 4.2. Se llaman **cara** de una célula a la obtenida al sustituir alguna de las ecuaciones $g_i \geq 0$ anteriores por $g_i = 0$.

Definición 4.3. Un **complejo celular** en \mathbb{R}^n es un conjunto finito K de células convexas en \mathbb{R}^n tales que:

- (a) Las caras de todo elemento de K pertenecen a K .
- (b) La intersección de dos células de K es una cara de ambas.

Por $|K|$ representaremos la reunión de los elementos de K que es compacto si todas las células que lo forman son compactas.

A las 0-células las llamaremos *vértices* y la dimensión de un complejo celular es el máximo de las dimensiones de sus células.

Nota 4.1. Para el caso particular en el que todas las células de un complejo celular sean simples, tenemos los complejos simpliciales. De hecho, todo complejo celular admite una subdivisión simplicial sin introducir nuevos vértices (Ver [46], 1.4.d).

Otro ejemplo de complejos celulares de interés son los complejos cúbicos, en los que las células o piezas básicas de la estructura son productos de intervalos.

Definición 4.4. Un **cubo** Q es un producto finito de intervalos de \mathbb{R}^n , es decir,

$$Q = I_1 \times I_2 \times \dots \times I_i \subset \mathbb{R}^n, \text{ con } i \in \mathbb{Z}, i < n$$

donde cada $I_i \subset \mathbb{R}$ es un intervalo de longitud uno o cero.

Se dice que el cubo tiene dimensión q , con q el número de intervalos I_i de longitud uno.

Ejemplo 4.1. En la Figura 6 se indican cubos de dimensiones 0, 1 y 2 respectivamente:

- $Q_1 = [1] \times [4] \subset \mathbb{R}^2$; $\dim(Q_1) = 0$.
- $Q_2 = [1, 2] \times [3] \subset \mathbb{R}^2$; $\dim(Q_2) = 1$.
- $Q_3 = [1, 2] \times [1, 2] \subset \mathbb{R}^2$; $\dim(Q_3) = 2$.

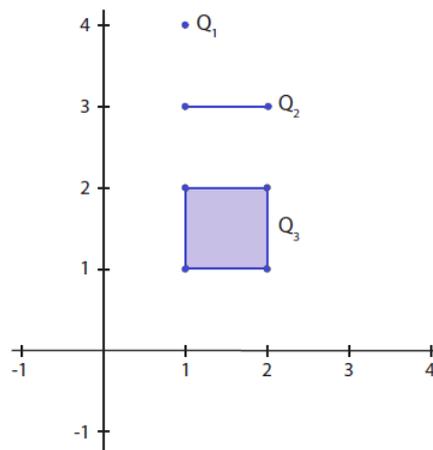


Figura 6. Cubos Q_1, Q_2, Q_3 representados en \mathbb{R}^2 .

Las **caras** de un cubo se obtienen al sustituir un factor de la forma $I_l = [l, l + 1]$ por $[l, l]$ o $[l + 1, l + 1]$, con $l \in \mathbb{Z}$.

Las caras del cubo son obviamente cubos de dimensión más pequeña. La clase resultante de complejos cúbicos son naturales en muchas aplicaciones.

Definición 4.5. Un **complejo cúbico** K será un conjunto de cubos que cumple las condiciones análogas a los complejos simpliciales, es decir:

- Si un cubo está en K , también están todas sus caras.
- Si dos cubos de K se cortan, entonces su intersección es una cara común.

Un cubo de dimensión r se llama vértice, arista, cuadrado o cubo para $0 \leq r \leq 3$, respectivamente.

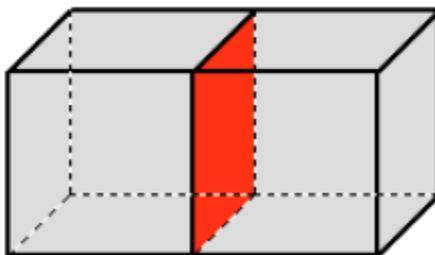


Figura 7. Complejo cúbico compuesto por dos cubos que comparten una cara, destacada en rojo.

Los complejos cúbicos aparecen de forma natural en topología digital, ya que toda imagen (objeto digital) se representa en \mathbb{R}^2 o en \mathbb{R}^3 mediante un conjunto de cuadrados (pixels) o cubos (voxels); es decir, como un subconjunto de la malla plana o tridimensional formada los puntos de coordenadas enteras.

La imagen queda definida al iluminar ciertos pixels o voxels. El uso de cubos en vez de símplexes permite simplificar su descripción combinatoria. Si tuviéramos que usar un complejo simplicial para describirla, tendríamos que tomar dos o seis símplexes para cada cuadrado o cubo, respectivamente.

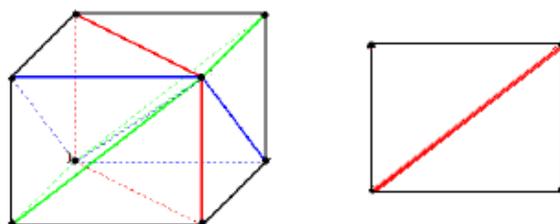


Figura 8. Subdivisiones en símplexes de un cubo y un cuadrado.

Nótese también que para representar un cubo basta indicar el vértice que tiene las coordenadas más pequeñas. Esto puede ser interesante en cuanto a almacenamiento de datos.

5. Nociones básicas sobre homología

De un modo general, la homología estudia el grado de conexión del espacio topológico con herramientas algebraicas. Más precisamente, la homología puede usarse para caracterizar rasgos interesantes de un complejo simplicial (una triangulación del espacio considerado) como racimos (clusters) conectados, túneles, cavidades, etc., lo que podría revelar relaciones subyacentes del comportamiento del conjunto de datos.

Antes de definir la homología simplicial, es necesario dotar a los complejos simpliciales de una estructura algebraica. Para ello, daremos una orientación en cada símplice del complejo mediante una relación de equivalencia. Esta relación genera dos clases de equivalencia para cada símplice, lo que permitirá establecer un signo, es decir, lo que conoceremos como una cadena simplicial y su opuesta. Finalmente, se definirá una aplicación lineal entre los símplices de una cierta dimensión y los símplices de una dimensión una unidad menor, que será la versión algebraica del borde y permitirá poner en marcha toda la maquinaria del álgebra homológica.

Definición 5.1. *Sea K un complejo simplicial. Dado un símplice $\sigma = (v_0, \dots, v_q)$, con $q > 0$, se define la siguiente relación de equivalencia sobre el conjunto de todas las ordenaciones de dichos vértices de σ :*

$$(v_0, \dots, v_q) \sim (v_{\pi(0)}, \dots, v_{\pi(q)})$$

si π es una permutación par de los índices. Esta es una relación de equivalencia y origina dos clases de equivalencia, cada una llamada orientación de σ .

En otras palabras, dos símplices con los mismos vértices poseen la misma orientación si se puede pasar de uno a otro mediante un número par de trasposiciones en el orden de los vértices. Al escoger una de las orientaciones, σ se dice un símplice orientado. Cabe decir que los 0-símplices solo tienen una orientación al estar formados por un único vértice.

En adelante, escribiremos $[v_0, \dots, v_q]$ para denotar el símplice (v_0, \dots, v_q) con la orientación definida por la ordenación de los índices.

A continuación, se introducen las estructuras algebraicas que usaremos en el resto del texto. Sea un cuerpo \mathbb{F} .

Definición 5.2. *Se define el \mathbb{F} -espacio vectorial de q -cadenas simpliciales orientadas, y se denota por $C_q(K; \mathbb{F})$, como el cociente del \mathbb{F} -espacio vectorial generado por todos los q -símplices σ de K por el subespacio generado por los $\sigma^1 + \sigma^2$, donde σ^1 y σ^2 representan las dos posibles orientaciones de un q -símplice. Los elementos de $C_q(K; \mathbb{F})$ se llaman q -cadenas simpliciales orientadas.*

Definición 5.3. Llamamos **operador borde** de orden q a la aplicación lineal:

$$\partial_q : C_q(K; \mathbb{F}) \rightarrow C_{q-1}(K; \mathbb{F})$$

dada por extensión lineal de

$$\partial_q([v_0, \dots, v_q]) = \sum_{i=0}^q (-1)^i [v_0, \dots, v_{i-1}, v_{i+1}, \dots, v_q],$$

donde $[v_0, \dots, v_{i-1}, v_{i+1}, \dots, v_q]$ representa el $(q-1)$ -símplice orientado resultante de eliminar en $[v_0, \dots, v_q]$ el vértice que ocupa la posición i .

Proposición 5.1. El operador borde verifica que $\partial_q \circ \partial_{q+1} = 0$ para todo q .

La proposición anterior es crucial porque convierte a $(C_\bullet(K; \mathbb{F}); \partial_\bullet)$ en un complejo de cadenas. La propiedad $\partial_q \circ \partial_{q+1} = 0$ implica que $\text{Im}(\partial_{q+1}) \subseteq \text{ker}(\partial_q)$ para todo q .

Definición 5.4. Se denominan **q -ciclos** a los elementos $x \in Z_q(K; \mathbb{F}) = \text{ker}(\partial_q)$ y **q -bordes** a los elementos de $B_q(K; \mathbb{F}) = \text{Im}(\partial_{q+1})$. Nótese que $B_q(K; \mathbb{F}) \subseteq Z_q(K; \mathbb{F})$.

Definición 5.5. Se define el **q -ésimo \mathbb{F} -espacio vectorial de homología simplicial** de K como el siguiente cociente:

$$H_q(K; \mathbb{F}) = \frac{Z_q(K; \mathbb{F})}{B_q(K; \mathbb{F})}$$

Se define también el **q -ésimo número de Betti** de K , denotado por β_q , como la dimensión de $H_q(K; \mathbb{F})$.

Por abuso de notación, escribiremos $H_q(K)$ para denotar $H_q(K, \mathbb{F})$.

La homología de un espacio se puede describir con los números de Betti. Para $q = 0, 1, 2$, β_q se puede interpretar intuitivamente: β_0 representa el número de componentes conexas, β_1 es el número de túneles, β_2 el número de cavidades. Por ejemplo, una esfera tiene $\beta_0 = 1, \beta_1 = 0, \beta_2 = 1$, ya que tiene una única componente conexa, ningún túnel, y una sola cavidad.

5.1. Homología persistente

Consideremos la formación de un complejo simplicial usando bolas de diámetro ε centradas en puntos de un conjunto. Para un ε pequeño, el complejo simplicial es sólo un conjunto disjunto de vértices. Pero para ε suficientemente grande, el complejo simplicial se convierte en un gran racimo (cluster) de símlices. Pero ¿cuál es el valor de ε que nos da la estructura “correcta”? La homología persistente nos da una respuesta rigurosa a esta pregunta.

Incrementando ε , obtendremos una secuencia de complejos simpliciales anidados, llamada filtración, que podremos examinar para obtener propiedades de conectividad y su robustez. Las características topológicas pueden aparecer y desaparecer cuando aumenta ε . Aquellas que existan durante un rango corto del parámetro ε se considerarán ruido.

Este planteamiento nos permite verlas como una colección de códigos de barras (barcodes), uno en cada dimensión, con cada característica representada por una barra. Cuanto mayor sea la duración de vida de una característica, mayor será la longitud de su barra. De este modo, la homología persistente puede usarse como medida de escala o resolución de una característica topológica.

Definición 5.6. Sea K un complejo simplicial. Una **filtración** es una sucesión de subcomplejos crecientes de K

$$\emptyset = K_0 \subseteq K_1 \subseteq \dots \subseteq K_n = K.$$

Cada subcomplejo K_i se denomina i -ésimo nivel de la filtración.

Nota 5.1. Una filtración se puede interpretar como una manera de obtener el complejo total K a partir de una familia de subcomplejos atendiendo a un determinado criterio, correspondiendo cada nivel K_i al estado de dicha construcción en el instante de tiempo i .

Ejemplo 5.1. Existen muchas formas de definir una filtración en un complejo. Entre las más habituales podemos destacar:

1. Filtración por esqueletos, tomando $K_0 = \emptyset$ y K_i el $i - 1$ -esqueleto de K .
2. Filtración por conjuntos de nivel de una función. Sea $f : K \rightarrow \mathbb{R}$ una función monótona creciente, es decir, tal que $f(\sigma) \leq f(\tau)$ siempre que $\sigma \leq \tau$.

Esta condición asegura que $K(a) = f^{-1}(-\infty, a]$ es un subcomplejo de K para todo $a \in \mathbb{R}$, por tanto, si $a_1 < \dots < a_n$ y se considera $a_0 = -\infty$, se obtiene la filtración

$$\emptyset = K_0 \subseteq K_1 \subseteq \dots \subseteq K_n = K.$$

donde $K_i = f^{-1}(-\infty, a_i]$.

A continuación se introducirán los conceptos básicos sobre homología persistente a un nivel elemental, de modo que permita su uso en el estudio de los distintos complejos tratados en este trabajo.

Dada una filtración $\{K_i\}_{i=0, \dots, n}$, para cada $i \leq j$, la inclusión canónica $K_i \hookrightarrow K_j$ induce inclusiones de p -ciclos $Z_p(K_i) \hookrightarrow Z_p(K_j)$ e inclusiones de p -bordes $B_p(K_i) \hookrightarrow B_p(K_j)$ y para cada dimensión p , denotaremos por $i_p^{i,j} : H_p(K_i) \rightarrow H_p(K_j)$ a la aplicación inducida en homología por dicha inclusión. En consecuencia, obtenemos el siguiente diagrama de espacios vectoriales:

$$0 = H_p(K_0) \rightarrow H_p(K_1) \rightarrow \dots \rightarrow H_p(K_n) = H_p(K),$$

A continuación, introducimos el concepto de homología persistente asociada a una filtración de un complejo. Su propósito principal es estudiar la evolución de clases de

ciclos a medida que nos movemos en la filtración, desde K_i hasta K_j . En este proceso pueden aparecer nuevas clases de homología así como anularse otras, bien al volverse triviales, o bien al unirse con otras.

Definición 5.7. *Los p -ésimos espacios vectoriales de homología persistente asociados a una filtración de K , se definen como las imágenes de los homomorfismos inducidos por la inclusión, $H_p^{i,j}(K) = \text{Im } \iota^{i,j}$, para $0 \leq i \leq j \leq n$. El correspondiente p -ésimo número de Betti de persistencia es la dimensión de estos espacios vectoriales, $\beta_p^{i,j}(K) = \dim H_p^{i,j}(K)$.*

En este contexto serán significativos aquellos p -ciclos creados en el nivel i y que no son nulos en el nivel j .

De forma similiar, definimos los espacios vectoriales de homología persistente reducidos y los números de Betti de persistencia reducida.

Nótese que para $i = j$ se tiene que $H_p^{i,i} = H_p(K_i)$ y $\beta_p^{i,i} = \beta_p(K_i)$.

Es fácil comprobar que $H_p^{i,j}(K) = Z_p(K_i) / (B_p(K_j) \cap Z_p(K_i))$.

Definición 5.8. *Diremos que una clase γ en $H_p(K_i)$ **nace** en K_i si $\gamma \notin H_p^{i-1,j}(K)$. Por otra parte, si γ nace en K_i , entonces decimos que **muere** entrando en K_j si se une con una clase que proviene de un nivel anterior al movernos de K_{j-1} hasta K_j , esto es,*

$$\iota^{i,j-1} \notin H_p^{i-1,j-1}, \text{ pero } \iota^{i,j}(\gamma) \in H_p^{i-1,j}.$$

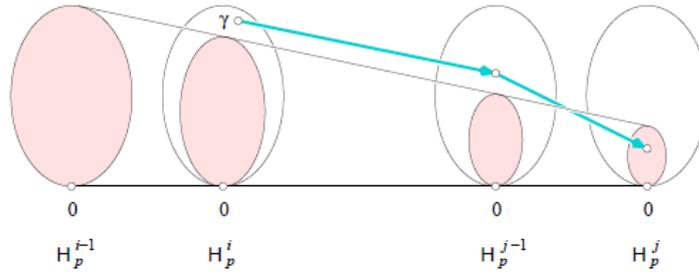


Figura 9. La clase γ nace en K_i ya que no aparece en la imagen sombreada de $H_p^{i-1} = H_p(K_{i-1})$. Además, γ muere entrando en K_j , ya que es la primera vez que su imagen se une con la imagen de $H_p^{i-1} = H_p(K_{i-1})$.

Si γ nace en K_i y muere entrando en K_j , entonces llamamos persistencia a la diferencia entre los valores de la función correspondientes, $\text{pers}(\gamma) = a_j - a_i$, y al intervalo semiabierto $[a_i, a_j)$ se le llama intervalo de persistencia de γ . A veces, es preferible ignorar los valores de la función y considerar la diferencia en los índices, $j - i$, que llamaremos índice de persistencia de la clase. Si γ nace en K_i pero no muere nunca, entonces diremos que tanto su persistencia como su índice de persistencia son infinitos.

Sea $\mu_p^{i,j}$ el número de clases p -dimensionales independientes nacidas en K_i y que mueren entrando en K_j entonces se tiene:

$$\mu_p^{i,j} = \left(\beta_p^{i,j-1} - \beta_p^{i,j} \right) - \left(\beta_p^{i-1,j-1} - \beta_p^{i-1,j} \right),$$

para todo $i < j$ y todo p . Véase el Teorema 114 en [22].

La primera diferencia cuenta las clases que nacen en o antes de K_i y mueren entrando en K_j , mientras que por otro lado, la segunda diferencia cuenta las clases que nacen en o antes de K_{i-1} y mueren entrando en K_j .

Análogamente, $\mu_p^{i,\infty} = \beta_p^{i,n} - \beta_p^{i-1,n}$, cuenta el número de clases independientes que nacen en K_i y no mueren.

Estas nociones de nacimiento, muerte y persistencia se suelen visualizar de dos formas: mediante los llamados códigos de barras o alternativamente a través de diagramas de persistencia. Indicaremos de forma breve e intuitiva en qué consisten.

Definición 5.9. Sea K un complejo simplicial y $\{K_i\}$ una filtración, si $0 \leq p \leq \dim(K)$, el p -ésimo código de barras de la filtración se obtiene del siguiente modo: en el primer cuadrante del plano se marcan en el eje horizontal los niveles de la filtración y en el eje vertical una ordenación arbitraria de intervalos de persistencia $[a_i, a_j)$ o semirrectas $[a_i, \infty)$.

Si consideramos la filtración de la Figura 10, se tienen los siguientes códigos de barras en las Figuras 11 y 12.

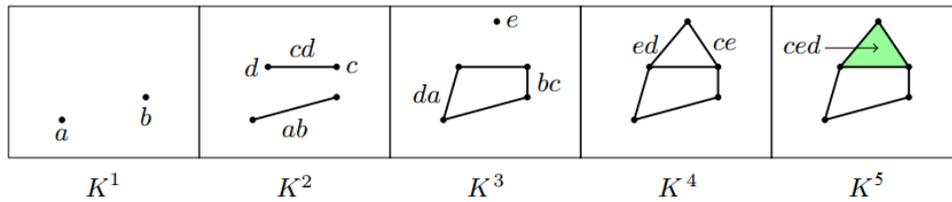


Figura 10.

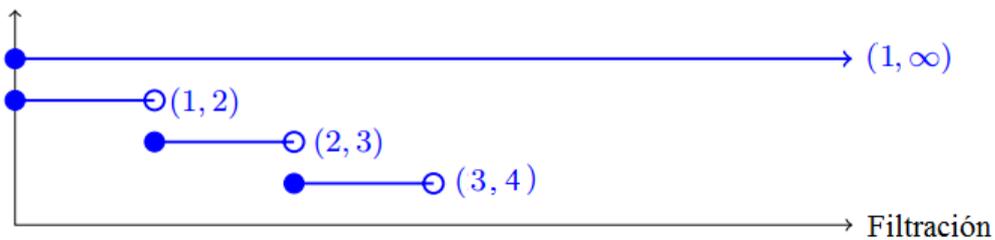


Figura 11. 1-código de barras de la filtración

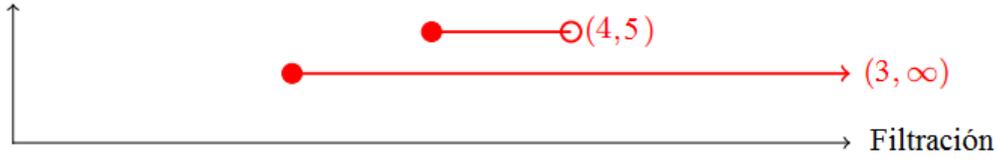


Figura 12. 2-código de barras de la filtración

Nótese que para cada nivel i de la filtración, $\beta_p(K_i)$ es el cardinal de la intersección de la vertical sobre dicho nivel con los intervalos.

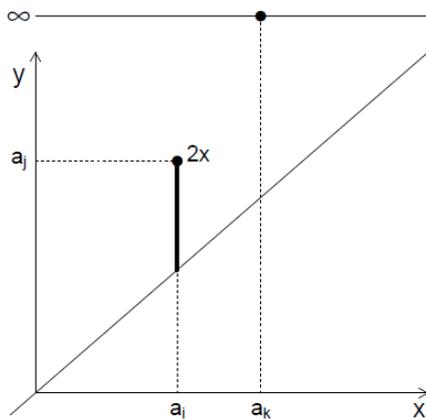
Definición 5.10. Sea K un complejo simplicial y $\{K_i\}$ una filtración, el p -ésimo **diagrama de persistencia** de una filtración de un complejo K , denotado por $Dgm_p(K_i)$, es un conjunto de puntos (x, y) del primer cuadrante del plano ampliado $\overline{\mathbb{R}^2} = \mathbb{R} \times (\mathbb{R} \cup \infty)$, a los que se le asocia una multiplicidad $mul_p(x, y)$, definida como:

$$mul_p(x, y) = \begin{cases} \mu_p^{i,j} & , \text{ si } x = a_i, y = a_j \\ \mu_p^{i,\infty} & , \text{ si } x = a_i, y = \infty \\ 0 & , \text{ en otro caso} \end{cases}$$

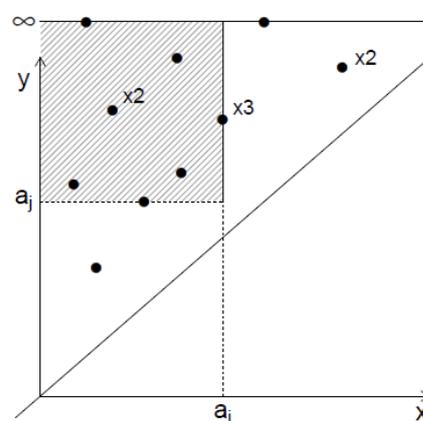
donde (a_i, a_j) son intervalos de persistencia.

Para visualizar el p -ésimo diagrama de persistencia, dibujamos el primer cuadrante, para los puntos con coordenadas finitas, y representamos los puntos con ordenada infinita en una línea horizontal visualmente bien separada de los puntos con coordenada finita. De este modo, representamos cada punto indicando su multiplicidad $mul(x, y)$.

La importancia de los diagramas de persistencia es que permiten obtener los números de Betti de persistencia: $\beta_p^{k,l}$ es el número de puntos, cada uno contado tantas veces como indique su multiplicidad, situados en el rectángulo $\{(x, y) \in \overline{\mathbb{R}^2} \mid x \leq a_i, y > a_j\}$.



(a) Se representaba el punto $x = (a_i, a_j)$ con multiplicidad 2.



(b) Se observa que $\beta_p^{i,j} = 9$.

6. Nervio de un recubrimiento

Supongamos un conjunto de datos finito dado S , en un espacio topológico X . La idea clave en el análisis topológico de datos es aproximar X .

Definición 6.1. *Un recubrimiento abierto de S en X es una familia $\mathcal{U} = \{\mathcal{U}_i\}_{i \in I}$ de subconjuntos abiertos de X tales que $S \subseteq \cup_{i \in I} \mathcal{U}_i$.*

Definición 6.2. *Sean X un espacio métrico y $\mathcal{U} = \{\mathcal{U}_i\}_{i \in I}$ un recubrimiento de X . El **nervio** de \mathcal{U} , denotado por $\mathcal{N}(\mathcal{U})$, es el complejo simplicial abstracto cuyo conjunto de vértices es I y para el que una familia de índices $\{i_0, \dots, i_q\} \subset I$ forma un q -símplice si, y solo si, $\mathcal{U}_{i_0}, \dots, \mathcal{U}_{i_q} \neq \emptyset$.*

Es obvio que los conjuntos unitarios pertenecen a $\mathcal{N}(\mathcal{U})$ (se entiende que los conjuntos de \mathcal{U} son no vacíos) y que si $\{i_0, \dots, i_q\} \in \mathcal{N}(\mathcal{U})$, entonces todo subconjunto de $\{i_0, \dots, i_q\}$ forma también un símplice ya que la intersección de los conjuntos asociados a esos índices es no vacía al haber menos elementos.

Por tanto, el nervio de un recubrimiento es efectivamente un complejo simplicial abstracto.

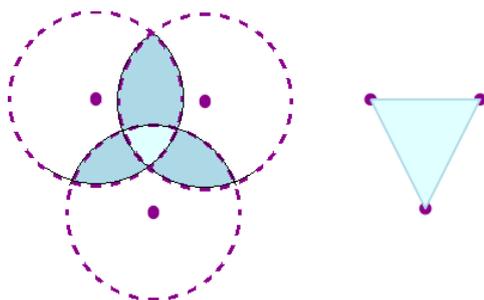


Figura 13. A la derecha se representa el nervio de la unión de las tres bolas abiertas de la izquierda.

Teorema 6.1 (Teorema del Nervio, [28], Corolario 4G.3). *Si \mathcal{U} es una colección finita de subconjuntos abiertos contráctiles de X tal que todas las intersecciones de subconjuntos de \mathcal{U} no vacías son contráctiles, entonces $\mathcal{N}(\mathcal{U})$ tiene el mismo tipo de homotopía que la unión de los \mathcal{U}_i , $i \in I$.*

7. Complejos de Čech

Como primer ejemplo de complejo usado en el análisis de datos, tenemos los complejos de Čech que se construye como el nervio de un recubrimiento formado por bolas abiertas.

Sea S un conjunto finito contenido en un subespacio eúclideo Y , $S \subseteq Y \subseteq \mathbb{R}^n$.

Definición 7.1. Sea $B_\varepsilon(x)$ la bola abierta de radio ε centrada en x :

$$B_\varepsilon(x) = \{y \in Y \mid d(x, y) < \varepsilon\}, \text{ con } \varepsilon \in \mathbb{R}, \text{ y } x \in Y$$

Dado $S \subseteq Y$ y $\varepsilon \in \mathbb{R}$, el conjunto de bolas abiertas de radio ε centradas en cada punto de S es un recubrimiento de S : $U_\varepsilon(S) = \{B_\varepsilon(x) \mid x \in S\}$. Se define entonces el **complejo Čech** C_ε como el nervio de este recubrimiento.

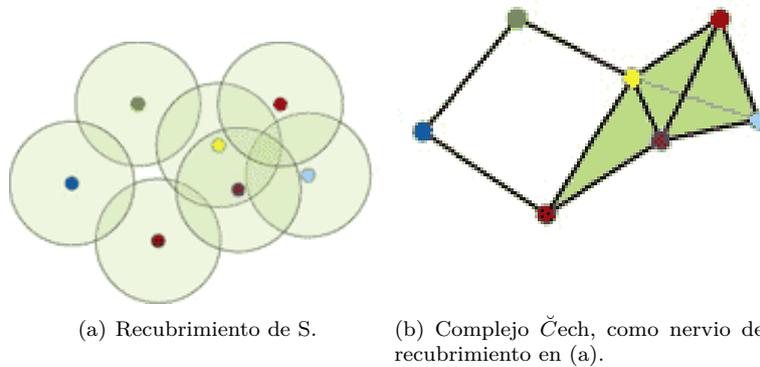


Figura 14. Complejo Čech.

Es decir, un k -símplice de $\check{C}_\varepsilon(S)$ es una colección de $k + 1$ puntos distintos x_i , $i = 1, \dots, k + 1$ de S tal que las bolas de radio ε con centro x_i tienen intersección no vacía.

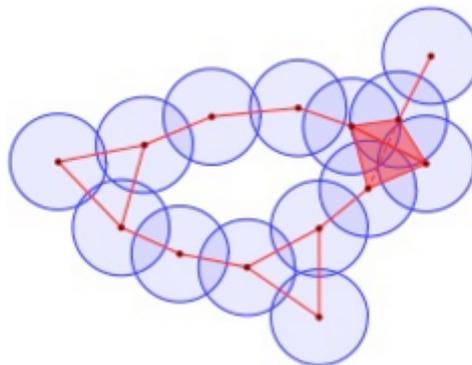


Figura 15. $\check{C}_\varepsilon(S) = \{\sigma \subseteq S : \bigcap_{x \in \sigma} B_\varepsilon(x) \neq \emptyset\}$.

Además, como las bolas son convexas, y por tanto, contráctiles, el nervio captura la topología del recubrimiento, en virtud del Teorema del Nervio. Es decir, tienen el mismo tipo de homotopía.

Podemos construir el complejo Čech para cada valor ε . Claramente, para los casos extremos se tiene: $\check{C}_0(S) = S$ y $\check{C}_\infty(S)$ es un $(|S| - 1)$ -símplice, realmente a partir de un cierto valor r , se tiene que $\check{C}_r(S)$ es un $(|S| - 1)$ -símplice. Por lo tanto, el complejo Čech puede tener dimensión mucho mayor que el espacio Y . Como un n -símplice tiene 2^{n+1} caras, el complejo puede volverse enorme a mayores escalas.

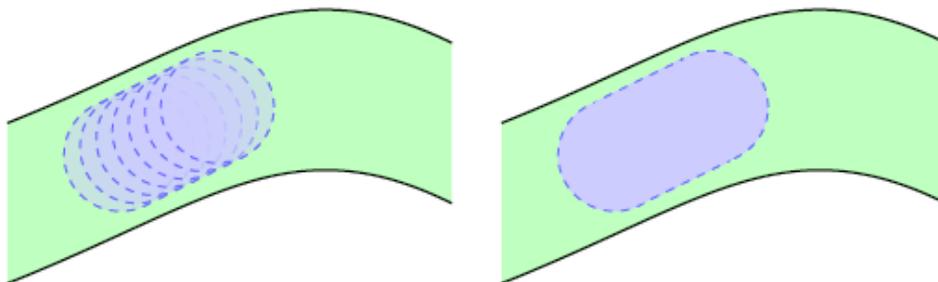
Los inconvenientes de trabajar con un complejo Čech son sus requisitos de memoria (no se puede simplemente almacenar el 1-esqueleto y completar el resto) y su construcción (debemos comprobar muchas intersecciones para construir el complejo completo). Sin embargo, estas desventajas se compensan con la precisión topológica: el complejo $\check{C}_\varepsilon(S)$ tiene siempre el mismo tipo de homotopía a la unión de bolas de radio $\varepsilon \in \mathbb{R}$.

8. Complejos de Vietoris-Rips

Veamos ahora el complejo Vietoris-Rips, un ejemplo de una clase más general de complejos simpliciales, ya que el complejo Čech es un subcomplejo de los Vietoris-Rips, incluso a veces un subcomplejo propio.

Una de las ventajas de su construcción es que es computacionalmente menos costoso de construir que el complejo Čech, en lo que respecta a la entrada y almacenamiento de datos. Para codificar un complejo simplicial típico de forma tratable computacionalmente es necesario enumerar todos los símlices. En contraste, veremos que el complejo Vietoris-Rips está enteramente especificado sabiendo los vértices de S y las aristas (dependiendo de cierto ε). El resto de símlices de mayores dimensiones estarán todos contabilizados por el grafo que determina el complejo.

Pese a sus buenas propiedades teóricas, en la práctica ocurre que el complejo Čech es computacionalmente inmanejable. Por ejemplo, puede ocurrir que haya muchas redundancias en el recubrimiento. Dicho de otro modo, es posible que varios elementos del recubrimiento se solapen, creando símlices innecesarios de distintas dimensiones y que realmente no aportan nada, más allá de gasto de almacenamiento.



En el dibujo está representado un pequeño trozo de algún espacio métrico sin agujeros. En la izquierda, las bolas forman parte de un recubrimiento abierto del espacio. El elevado número de intersecciones provocará que el complejo Čech tenga un gran número de símlices de distintas dimensiones. A la derecha, sin embargo, un único abierto recubre exactamente el mismo pedazo de espacio que las bolas. En ese caso, el abierto solo aporta un 0-símlice al complejo de Čech.

La solución que se propone a este problema es poder recuperar el complejo al completo únicamente por medio de la información sobre la distancia entre vértices. Esto hace que no sea necesario comprobar si existen intersecciones no vacías en todas las subcolecciones de $B_\varepsilon(x)$. El complejo de Vietoris-Rips es una variante del complejo Čech que implementa esta solución.

Consideremos un subconjunto discreto S de un espacio euclídeo \mathbb{R}^n correspondiente a una nube de puntos. Elegimos una constante, $\varepsilon > 0$. El complejo Vietoris-Rips de escala $\varepsilon \in \mathbb{R}$, $VR_\varepsilon(S)$, es el complejo simplicial cuyos símlices son las colecciones finitas de puntos en S cuya distancia dos a dos es menor o igual que ε (que es claramente una familia cerrada bajo la operación de tomar subconjuntos).

Definición 8.1. Sea S un conjunto finito contenido en un subespacio euclídeo X , $S \subseteq X \subseteq \mathbb{R}^n$. Dado un $\varepsilon \geq 0$, llamamos **complejo de Vietoris-Rips** asociado a S con parámetro ε , y lo denotamos por $VR_\varepsilon(S)$, al complejo simplicial cuyos vértices son los puntos de S y en el que los puntos $\{x_0, \dots, x_q\}$ forman un q -símplice si y solo si $d(x_i, x_j) \leq \varepsilon$ para todo $0 \leq i, j \leq q$.

Como en el caso del complejo Čech, es obvio que el complejo Vietoris-Rips es un complejo simplicial abstracto, pues $d(x, x) \leq 0, \forall x \in S$ y la distancia entre dos puntos de cualquier subconjunto de $\{x_0, \dots, x_q\}$ sigue siendo ε o menos.

Lo que esperamos es que $VR_\varepsilon(S)$ nos dé una **buena aproximación a la estructura subyacente** de la nube de puntos. Nótese la dificultad en especificar lo que eso significa: supone elegir correctamente ε para la mejor aproximación, lo que depende en gran medida de la naturaleza del problema cuyos datos se estén manejando.

A pesar de que este complejo está basado en puntos de \mathbb{R}^n , puede haber símplices de dimensión mucho mayor que n si el parámetro ε elegido es suficientemente grande, ya que entonces habrá que considerar muchas aristas.

8.1. Comparación con el complejo Čech

La primera observación que debe hacerse es que los conjuntos de vértices para ambos complejos son idénticos. Las diferencias surgen al comparar distintos valores de ε . Por ejemplo, en la siguiente figura se muestran cinco puntos cada uno con una bola de radio ε , centrada en él.

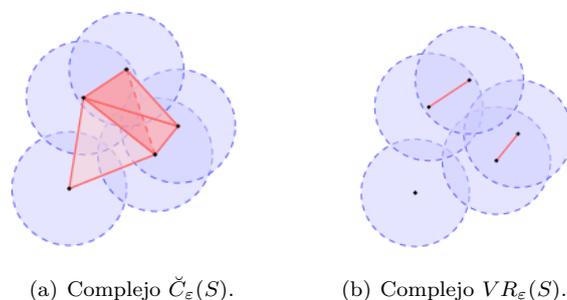


Figura 16. Comparación de los complejos Čech y Vietoris Rips para el mismo parámetro ε .

Podemos observar que el complejo Čech tendría cinco 0-símplices, ocho 1-símplices, cinco 2-símplices y un 3-símplice. Para el complejo Vietoris-Rips sólo dos parejas de puntos están lo suficientemente cerca entre sí como para que aparezca un 1-símplice, luego tenemos cinco 0-símplices y dos 1-símplice.

Consideremos de nuevo los cinco puntos anteriores, pero disminuyamos el parámetro ε un poco para ver cómo afecta a los complejos. La nueva situación se refleja en la Figura 17.

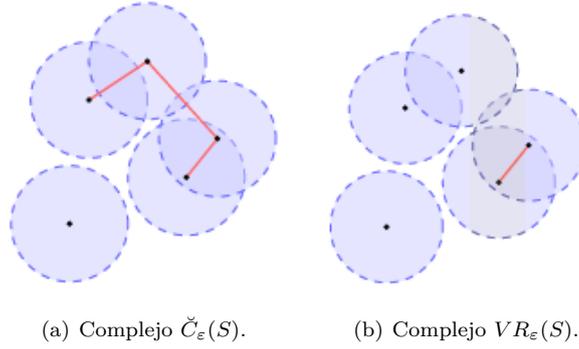


Figura 17. Comparación de los complejos Čech y Vietoris Rips para un nuevo parámetro ε más pequeño.

En este caso, el complejo Čech es mucho más pequeño: estaría formado por cinco 0-símplices y tres 1-símplices. Por su parte, el complejo Vietoris-Rips sólo pierde un 1-símplice: sólo dos puntos están lo suficientemente cerca entre sí como para que aparezca un 1-símplice, luego tenemos cinco 0-símplices y un único 1-símplice.

El grado con que un complejo Vietoris-Rips (para algún ε fijado) captura por completo la topología de la nube de puntos no es obvia. Aunque, el complejo Vietoris-Rips parece capturar la presencia de los agujeros grandes, puede eliminar otros de menor amplitud. Veámoslo en la siguiente figura comparando el complejo Čech con bolas de radio ε , con el complejo Vietoris-Rips con bolas de radio 2ε (al considerar bolas de radio 2ε basta que éstas se corten dos a dos para poder marcar el símplice).

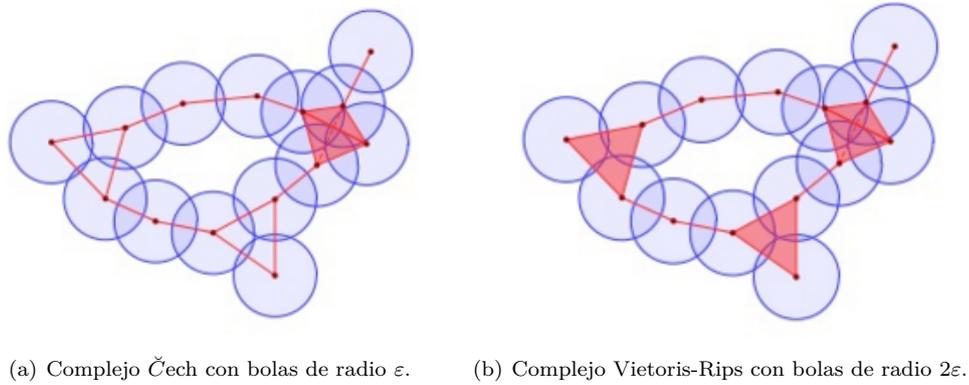


Figura 18. Comparación de los complejos $\check{C}_\varepsilon(S)$ y $VR_\varepsilon(S)$.

Observamos que el complejo Vietoris-Rips captura un agujero con el 2-símplice que se encuentra más a la izquierda, donde el complejo Čech lo dejaba fuera.

Proposición 8.1. Sea $\varepsilon \geq 0$, se verifican las siguientes inclusiones:

$$\check{C}_{\frac{\varepsilon}{2}}(S) \subset VR_{\varepsilon}(S) \subset \check{C}_{\varepsilon}(S).$$

Demostración. La condición que se pide para el complejo de Vietoris-Rips es más restrictiva que la condición para el complejo de Čech. Si los centros de dos bolas se encuentran a una distancia menor o igual que ε , entonces esas bolas de radio ε han de intersectarse. La segunda inclusión es, por tanto, inmediata.

La primera inclusión es consecuencia de la desigualdad triangular. Si q bolas abiertas de radio $\frac{\varepsilon}{2}$ intersecan para formar un q -símplice, entonces la distancia entre dos centros cualesquiera de ellas es menor o igual que $\frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$, ya que los puntos de la intersección distan menos que $\frac{\varepsilon}{2}$ de cualquier centro.

Q.E.D.

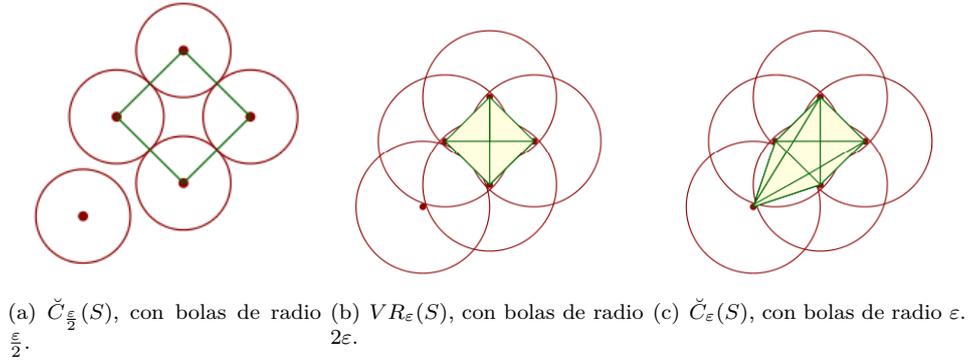


Figura 19. Inclusiones entre complejos Čech y Vietoris-Rips.

La proposición anterior nos dice entonces que el complejo de Vietoris-Rips con parámetro ε está encajado entre complejos de Čech, por tanto capturará bien la topología de los datos subyacentes si así lo hacen los respectivos complejos de Čech con parámetros $\frac{\varepsilon}{2}$ y ε .

De forma análoga, se tiene que el complejo de Čech está encajado entre complejos de Vietoris-Rips, con una cota más fina, véase [15] para la demostración:

$$VR_{\varepsilon}(S) \subset \check{C}_{\varepsilon'}(S) \subset VR_{\varepsilon'}(S), \text{ donde } \frac{\varepsilon'}{\varepsilon} \geq \sqrt{\frac{n+1}{n}}, S \subset \mathbb{R}^n.$$

En general, tal y como indica el ejemplo y corrobora la proposición, el complejo de Vietoris-Rips tiene menos símplices que el complejo de Čech para un mismo parámetro. A pesar de esto, el complejo de Vietoris-Rips sigue siendo computacionalmente caro, porque todavía es preciso guardar el conjunto de vértices del espacio métrico en cuestión. Esto parece irrelevante en las nubes de puntos vistas hasta ahora, pero se vuelve más importante a medida que esa nube de puntos crece o cuando el espacio es un continuo de puntos.

Llegados a este punto, cabe preguntarse si el complejo de Vietoris-Rips conserva la buena propiedad dada por el Teorema del Nervio. La respuesta es negativa, ya que este complejo no verifica las hipótesis del Teorema. Y hemos visto con los ejemplos que, efectivamente los complejos pueden no ser del mismo tipo de homotopía.

8.2. Homología persistente con complejos anidados

Asociar a una nube de puntos un complejo simplicial requiere elegir un parámetro ε , que realmente significa especificar un criterio de proximidad entre los puntos dados. Si consideramos la familia obtenida al variar ε , se obtiene una filtración, y tiene sentido calcular su homología persistente que es más adecuada que la homología ordinaria de cada nivel de la filtración para detectar propiedades geométricas relevantes de dichos puntos.

En efecto, si ε suficientemente pequeño, el complejo es un conjunto discreto; y para ε suficientemente grande, el complejo es un símplice de una gran dimensión. Por tanto, cabe preguntarse cuál es el valor adecuado de ε que captura de la mejor forma posible la topología del conjunto de datos.

Consideremos por ejemplo, la nube de puntos y la secuencia de complejos Vietoris-Rips que se indica en la Figura 20.

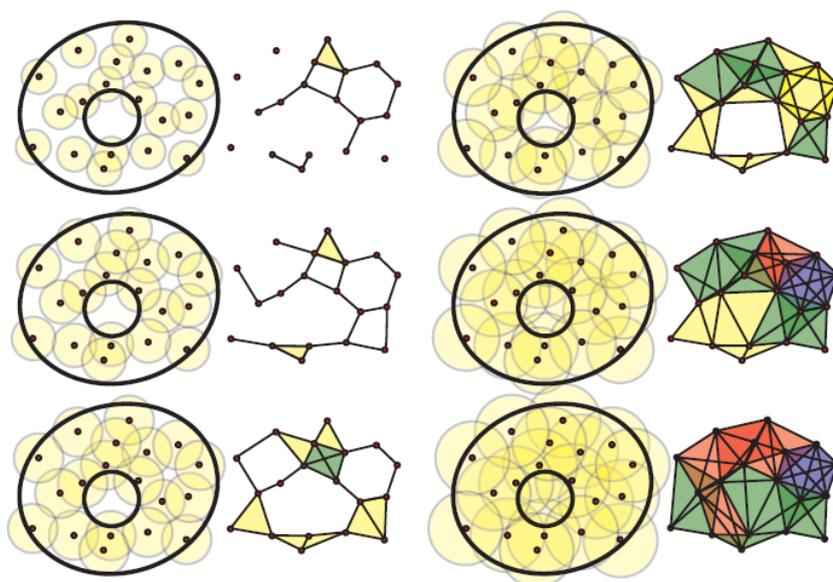


Figura 20. Secuencia de complejos Vietoris-Rips anidados.

La nube de puntos parece estar distribuida de forma que da lugar a una corona circular plana. ¿Puede esto deducirse de la sucesión de complejos obtenida al ir aumentando el radio de las bolas centradas en los puntos? Al observar la figura, parece que no es sencillo indicar el ε adecuado: a medida que aumentamos ε van desapareciendo los agujeros del complejo, incluido el agujero más grande, el de la corona circular.

Sin embargo, al considerar la homología persistente de la filtración representada por los códigos de barras de la figura, observamos que $\beta_0 = 1$ a partir de un cierto instante, lo cual indica que la figura que mejor se adapta a nuestro conjunto de puntos es conexa.

Y luego, que $\beta_1 = 2$ para un intervalo de longitud considerable si se compara con otros segmentos del código de barras para H_1 . Esto parece indicar que considerando un nivel de proximidad de tamaño medio, parece natural admitir que una figura regular que se adapta bien a la nube de puntos es una corona circular.

Nótese cómo el código de barras de H_2 es un intervalo muy pequeño, lo que está de acuerdo con el hecho de que nuestros puntos de partida están en el plano.

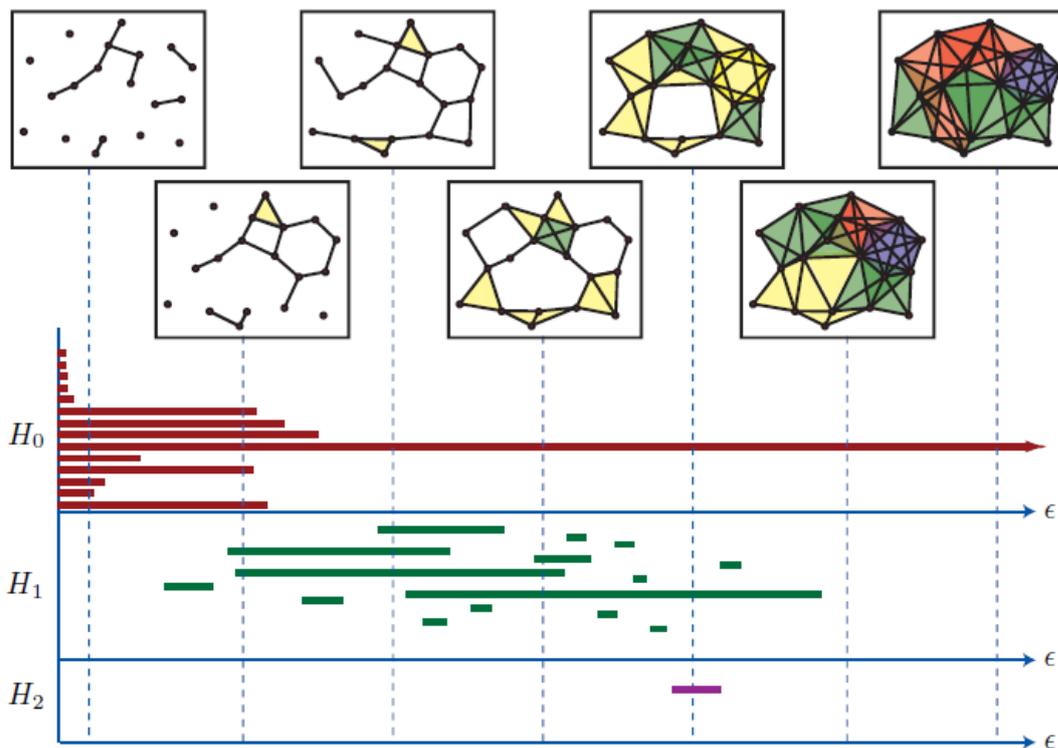


Figura 21. Secuencia de complejos Vietoris-Rips anidados y su diagrama de barras.

Este ejemplo es un caso simple que pone de manifiesto como aplicar la homología persistente para detectar propiedades del conjunto de datos de partida, depende esencialmente de evaluar cuáles son los intervalos de persistencia adecuados. Lo que a su vez, es claro que estará en función de la naturaleza del problema o los datos considerados.

Damos a continuación un ejemplo ilustrativo desarrollado en [5] de cuáles son los pasos que se siguen en el proceso del análisis topológico de datos, cuyo esquema se representa en la Figura 22.

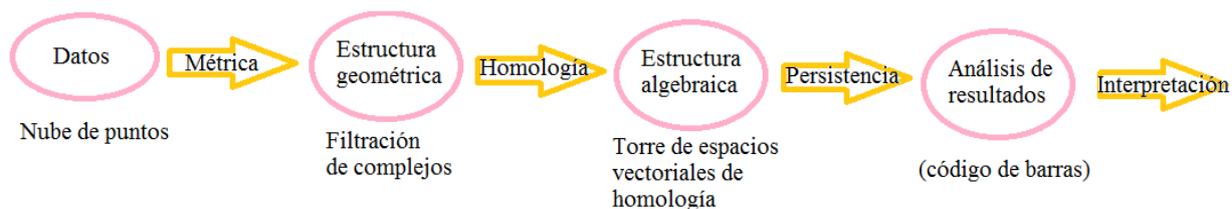


Figura 22. Esquema del proceso seguido en el análisis topológico de datos.

En la Figura 23, se indican dos nubes de puntos, en azul las coordenadas de ciudades en los distritos electorales que han votado salir de la Unión Europea, y en rojo las coordenadas de las que han votado permanecer en ella.

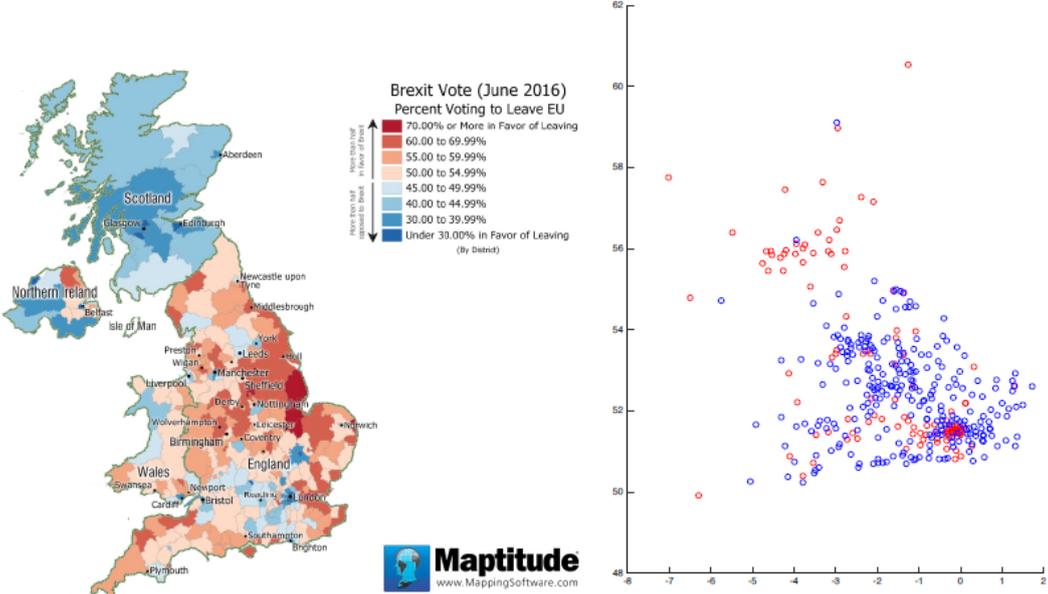


Figura 23.

A continuación, se indican los códigos de barras de dimensiones cero y uno, para las filtraciones de complejos de Vietoris-Rips de ambas nubes de puntos.

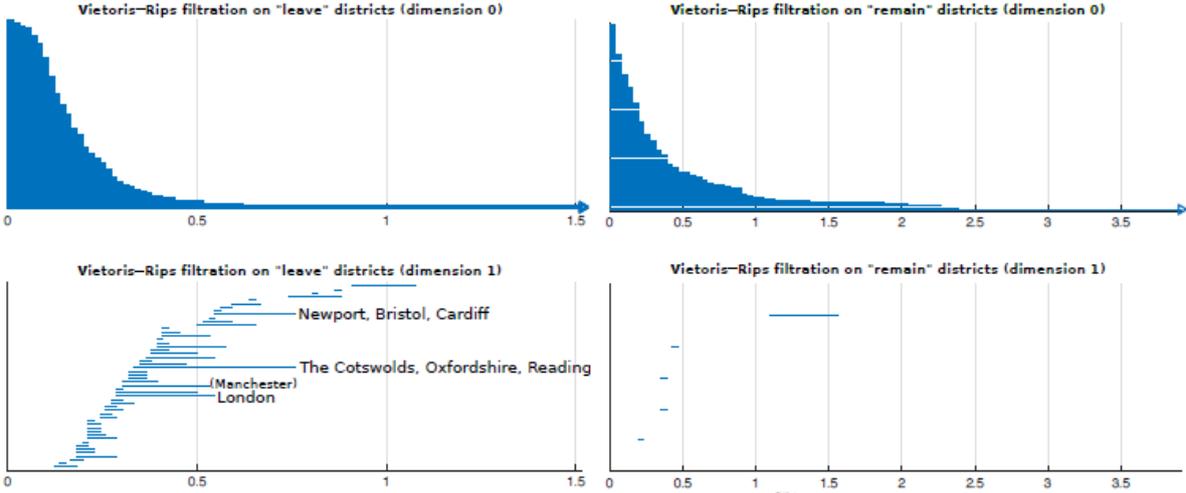


Figura 24.

Como se observa, aparecen muchos más 1–ciclos que persisten en el código de barras de la nube de puntos azul, es decir, de los distritos que votaron a favor de la salida de la Unión Europea, con un nivel notable de persistencia.

En el estudio indicado, se analiza la relación entre el tamaño y la persistencia de los ciclos y la distribución de las zonas en las que se votaron cosas contrarias.

9. Otros complejos simpliciales

9.1. Complejos de Delaunay

A grandes rasgos, una solución que se propone al problema de acumular demasiados datos dificultando la computabilidad, es “juntar” varios puntos próximos en un único punto, para así reducir la carga computacional en almacenamiento, sin que ello afecte a la estructura topológica del objeto geométrico que modela los datos. Además es también una forma de discretizar espacios continuos cuando se pretende calcular un complejo de Vietoris-Rips del espacio.

Por otra parte, a veces sucede que hay aglomeraciones de datos representados por puntos muy próximos entre sí, y no interesa un excesivo detalle, sino que es suficiente detectar estas acumulaciones y considerarlas como un solo punto. Esta situación lleva a plantearse si las bolas métricas usadas en los complejos Čech o Vietoris-Rips, son entornos demasiado restrictivos cuando se producen aglomeraciones de puntos como en la Figura 25.

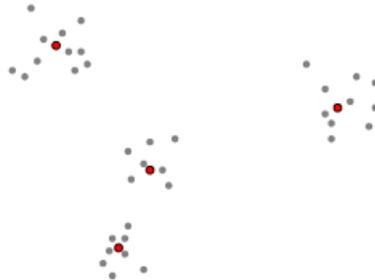


Figura 25.

En este tipo de situaciones es más adecuado identificar esas acumulaciones a un solo punto y considerar células convexas que los contienen y que dan cuenta de la proximidad de los centros de las acumulaciones.

Definición 9.1. Dado un punto de referencia $\lambda \in L \in \mathbb{R}^n$, se define la **célula de Voronoi** asociada a λ como:

$$V_\lambda = \{x \in \mathbb{R}^n : d(x, \lambda) \leq d(x, \lambda'), \forall \lambda' \in L\}.$$

Las células de Voronoi forman un recubrimiento de \mathbb{R}^n , ya que para cada $x \in \mathbb{R}^n$ siempre es posible comparar su distancia a cualquier elemento de L . Las células de Voronoi descomponen el espacio métrico en distintas regiones.

En la Figura 26 se muestran los cuatro puntos de referencia (de acumulación) tomados en la nube de puntos anterior, y las células de Voronoi correspondientes.

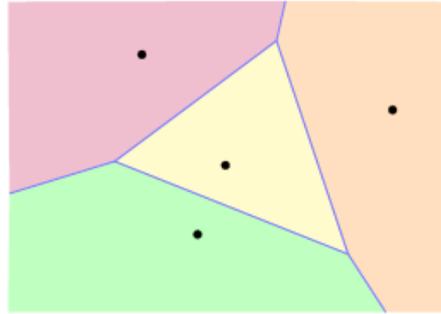


Figura 26. Diagrama de Voronoi.

La colección de células de Voronoi y sus caras constituyen un complejo celular denominado **diagrama de Voronoi** de L y denotado por $Vor(L)$.

Los diagramas de Voronoi son profusamente estudiados y tienen múltiples aplicaciones en diversas áreas. Como un ejemplo de problema importante que fue resuelto gracias a ello, puede citarse que fueron empleados para detener el brote de cólera en Londres en 1854. Dividiendo la ciudad por proximidad a diversos pozos como se observa en la Figura 27, el físico John snow pudo determinar que un pozo particular era el responsable de la mayoría de las muertes.



Figura 27. Diagrama de Voronoi para el mapa de Londres en 1854.

Con un diagrama de Voronoi, podemos descomponer el mapa de los Estados Unidos usando la localización de los aeropuertos como nube de puntos (Figura 28).

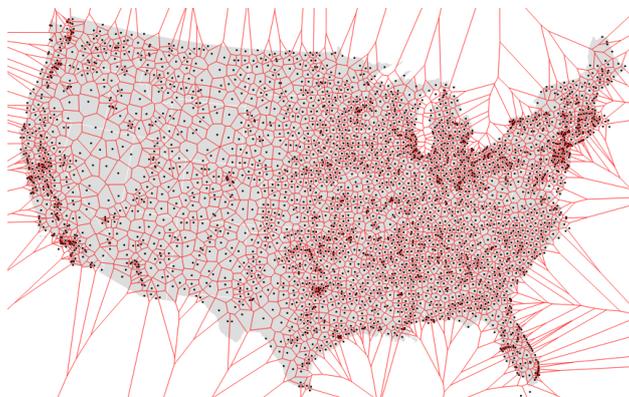


Figura 28.

E incluso podemos aplicar esta misma descomposición para toda la superficie de la Tierra (Figura 29).

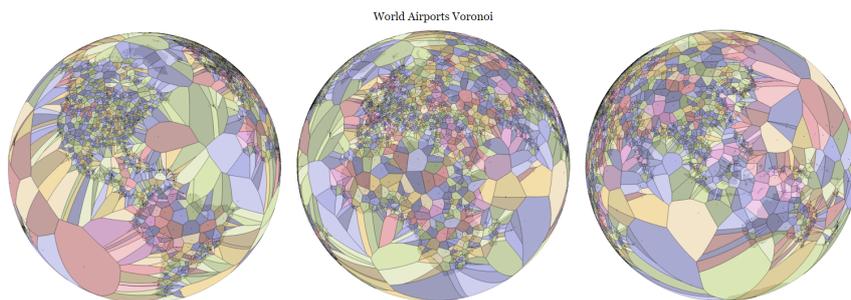


Figura 29.

Mediante los diagramas de Voronoi se construyen los complejos de Delaunay que constituyen una herramienta geométrica para solucionar los problemas computacionales de los complejos de Čech y Vietoris-Rips, ya que evitan tener que utilizar símlices de dimensión excesivamente alta.

Definición 9.2. Llamamos **complejo de Delaunay** asociado a L al nervio del recubrimiento dado por las células de Voronoi.

$$Del(L) = \{\sigma \subset L : \cap_{\lambda \in \sigma} V_\lambda \neq \emptyset\}.$$

Esta definición es equivalente a la siguiente: un q -símlice $\sigma = \langle p_0, \dots, p_q \rangle \in Del(L)$ si existe una bola cerrada \bar{B}^n tal que $\sigma \cap \bar{B}^n \subset S^{n-1}$ y $L \cap int(\bar{B}^n) = \emptyset$.

Para una exposición detallada de las propiedades de los diagramas de Voronoi y los complejos de Delaunay, véase [13] y [11].

Observación 9.1. La idea de la construcción del complejo de Delaunay es similar a lo que se hace en el juego de unir una serie de puntos numerados para obtener una figura: se presenta un conjunto numerado de puntos, y la tarea es empezar en el número uno y dibujar líneas hasta que llegemos al final, cuando aparece una imagen.

En este juego, se unen dos puntos si son consecutivos según los números que se indica, es decir, se une cada punto con los más cercanos según este orden.

Esto funciona porque hemos ordenado la colección de puntos y entendemos que debemos unirlos según una secuencia particular. Sin embargo, al estudiar una serie de datos, solo tenemos una colección de puntos sin etiquetar. ¿Cómo podemos conectarlos para formar una imagen razonable?

Probablemente no querremos conectar los puntos que no estén cerca unos de otros. Pero, ¿hay alguna forma ordenada de decidir qué significa cerca? Para responder a esta pregunta, se introduce el digrama de Voronoi, que proporciona un criterio de proximidad. Y una vez que tenemos esto, construimos el complejo de Delaunay, que conecta los puntos siguiendo ese criterio.

Continuando con el ejemplo de la nube de puntos anterior, y basándonos en las células de Voronoi obtenidas en la Figura 26, tenemos el complejo de Delaunay de la Figura 30, con cuatro 0-símplices dados por los cuatro puntos de referencia tomados, seis 1-símplices y tres 2-símplices formados por las intersecciones entre las regiones o células de Voronoi:

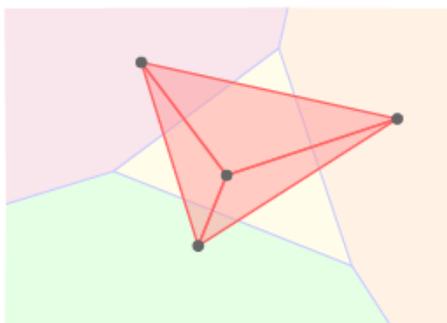


Figura 30. Complejo de Delaunay.

Ejemplo 9.1. Si el mayor número de células de Voronoi con intersección no vacía es $m + 1$, entonces $\dim(\text{Del}(S)) = m$. Por tanto, el complejo de Delaunay no siempre se puede sumergir en \mathbb{R}^n .

Por ejemplo, tomemos cuatro puntos en \mathbb{R}^2 , situados en una circunferencia, entonces las células de Voronoi de dichos puntos, se cortan en el centro de la circunferencia, dando lugar a un tetraedro en el complejo de Delaunay, es decir, un 3-símplice.

Se demuestra que si esta situación no se produce, entonces sí es cierto que el complejo de Delaunay está contenido en \mathbb{R}^n y además es un complejo simplicial:

Teorema 9.1 (Triangulación de Delaunay). (*[6], Teorema 4.3*) *Sea P un conjunto de puntos que satisface la siguiente condición (posición general respecto a esfera):*

ningún subconjunto de $n + 2$ puntos de P están contenidos en una misma S^{n-1} .

Entonces $Del(P)$ es un complejo simplicial contenido en \mathbb{R}^n .

Por tanto, si los puntos están situados adecuadamente, sí tenemos garantizado que $Del(S)$ tiene dimensión más pequeña que los de Čech y Vietoris-Rips.

Si no lo están, existen algoritmos para conseguir que modificando la posición inicial de esos puntos, se sitúen en posición general. Ver [13] y [11].

9.2. Complejos Alfa

Los complejos alfa están relacionados con los complejos Witness (que veremos a continuación) y con los complejos de Delaunay, ya que también buscan reducir el tamaño del complejo, limitar su dimensión usando la geometría del espacio incrustado.

Fueron introducidos en [20], y luego extendidos a conjuntos finitos de \mathbb{R}^3 con el propósito de estudiar estructuras de biomoléculas.

Apoyándonos en las definiciones anteriores de regiones de Voronoi, y complejos de Delaunay, podemos construir un complejo más pequeño computacionalmente de la siguiente forma: consideremos la intersección de la región de Voronoi de un punto $x \in S$ y una bola abierta de radio ε centrada en dicho punto x , denotada por $B_\varepsilon(x)$, se tiene que esta intersección es contráctil.

Luego, por el Teorema del Nervio, podemos intersecar esos dos subespacios y obtener un complejo simplicial realizable en \mathbb{R}^n que guarde el mismo tipo de homotopía. De este modo, las regiones definidas por las intersecciones $R_x(\varepsilon) = V_x \cap B_\varepsilon(x)$, $\forall x \in S$ determinan un recubrimiento de la nube de puntos, ya que como para toda pareja $x, x' \in S$, sus regiones de Voronoi $V_x, V_{x'}$ tienen intersección disjunta o se solapan a lo largo de uno de sus bordes comunes, se tiene que la unión de las $R_\varepsilon(x)$, para todo $x \in S$ recubre a la unión de las bolas abiertas $B_\varepsilon(x)$.

Definición 9.3. Sea $\varepsilon > 0$, se define el **complejo alfa** de un conjunto finito S como el nervio del recubrimiento formado por $R_\varepsilon(x)$, $\forall x \in S$. Explícitamente,

$$A_\varepsilon(S) = \{\sigma \subset S : \cap_{x \in \sigma} R_x(\varepsilon) \neq \emptyset\}.$$

En la siguiente figura podemos ver: a la izquierda el diagrama de Voronoi con el complejo de Delaunay para una cierta nube de puntos; en el centro la intersección de las regiones de Voronoi con las bolas de radio ε centradas en los puntos; y a la derecha, el complejo alfa al que dan lugar dichas intersecciones.

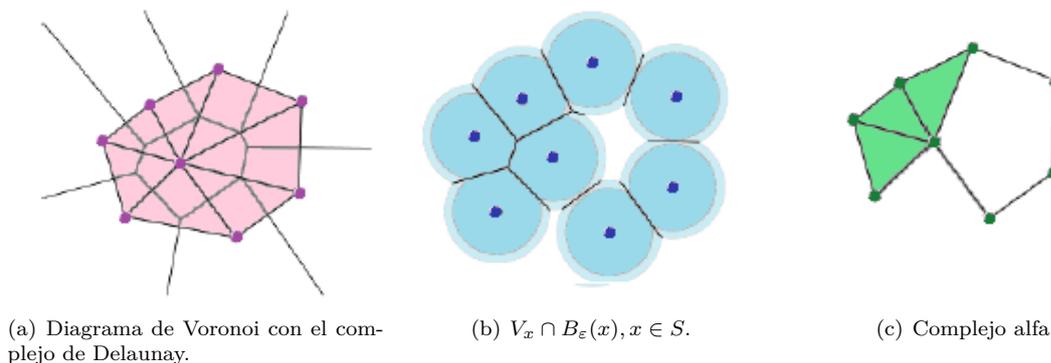


Figura 31. Construcción del complejo alfa.

Construimos los complejos alfa a partir de los complejos Delaunay; para cada s3mplice del complejo Delaunay, determinamos el menor ε para el cual el s3mplice entra dentro del complejo alfa. Luego, clasificamos los s3mplices por su m3nimo ε para tener un orden parcial de los s3mplices. De esta forma podemos determinar el complejo alfa para cualquier ε .

$A_\varepsilon(S)$ es un subcomplejo del complejo Delaunay, para todo ε . En particular, para ε suficientemente grande, la uni3n de las bolas $B_\varepsilon(x)$ contiene a todos los puntos $x \in S$ y todas se cortan entre s3, por lo tanto $A_\varepsilon(S)$ es el complejo $Del(S)$.

Adem3s, $A_\varepsilon(S)$ y $\check{C}_\varepsilon(S)$ tienen el mismo tipo de homotop3a (pues hemos visto que a ambos se les puede aplicar el Teorema del Nervio).

Por definici3n tenemos que $R_\varepsilon(x) \subseteq B_\varepsilon(x)$, lo que implica $A_\varepsilon(S) \subseteq \check{C}_\varepsilon(S)$. De hecho, tenemos $A_\varepsilon(S) \subseteq D(S)$ (con $D(S)$ denotando el complejo Delaunay de S), lo que implica que si los puntos est3n en posici3n general, la dimensi3n del complejo alfa es a lo sumo d . Una dimensi3n m3s peque3a que la del complejo $\check{C}ech$, $2^{Card(S)}$. Esto no ser3a demasiado interesante sin el siguiente teorema, que es consecuencia inmediata del Teorema del Nervio aplicado a los $A_\varepsilon(S)$.

Teorema 9.2. *La realizaci3n geom3trica del complejo alfa $A_\varepsilon(S)$ tiene el mismo tipo de homotop3a que la uni3n de las bolas cerradas $\cup_{x \in S} B_\varepsilon(x)$.*

En la siguiente figura, se puede observar la comparación entre los distintos complejos asociados a una nube de puntos, estudiados hasta ahora.

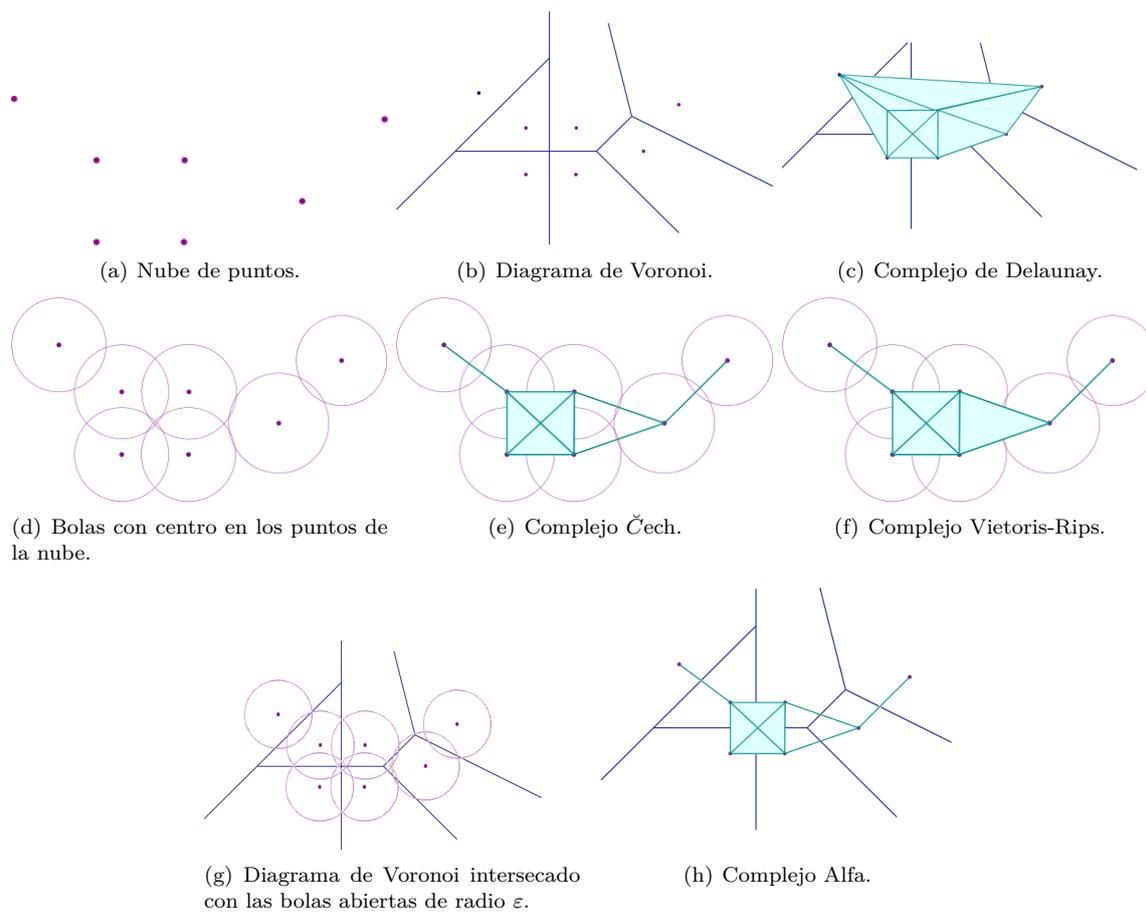


Figura 32. Comparación de los complejos de Delaunay, Čech, Vietoris-Rips y Alfa.

Es evidente que si $\varepsilon \leq \varepsilon'$, se tiene que $A_\varepsilon(L) \subseteq A_{\varepsilon'}(L)$. Por tanto, los complejos alfa dan lugar a una filtración y tiene sentido considerar su espacio vectorial de persistencia asociado.

9.3. Complejos Witness

Aunque la traducción de witness es testigo, mantendremos la notación como witness, por ser la usual.

Los complejos Vietoris-Rips de conjuntos de datos grandes pueden ser demasiado pesados de almacenar y manipular. Para hacer frente a esta preocupación, las construcciones complejas witness de Carlsson y de de Silva reducen en gran medida el tamaño a expensas del potencial de precisión topológica. Esto se ha convertido en algo complicado, con una variedad de definiciones relacionadas construidas para adaptarse a los diferentes conjuntos de datos.

Sin embargo, los enfoques experimentales han demostrado que las construcciones menos rigurosas también dan la homología correcta con un alto nivel de confianza. Otro aspecto a tener en cuenta es que nuestro enfoque es construir una estructura de variedad de nuestros datos y utilizarla luego para inferir propiedades de nuestros datos.

Por esta razón, hemos optado por utilizar la construcción de un complejo basado en puntos de referencia más eficiente, específicamente, el complejo witness. Para ello, seleccionamos ciertos puntos de nuestros datos para ser puntos de referencia, y los tomaremos como nuestros 0-símplices (vértices).

Supongamos como en secciones anteriores, un conjunto de datos finito dado $S \subset \mathbb{R}^n$. Un subconjunto $L \subset S$ de puntos de referencia se escoge en el conjunto de vértices, y los puntos restantes juegan el papel de determinar qué símplices estarán en el complejo. Diremos que un punto es witness de dos puntos de referencia basándonos en un determinado criterio. De este modo, dos puntos de referencia formarán un 1-símplice si y solo si tienen un witness en común.

Definición 9.4. Sea σ un símplice con vértices en L . Diremos que $x \in \mathbb{R}^n$ es un **witness fuerte** para σ respecto de L , si $|x - a| \leq |x - b|$, con $a \in \sigma$, $b \in L$.

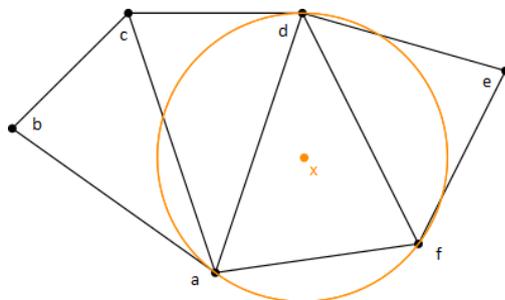
Esto, en particular, implica que $|x - a| = |x - a'|$, si $a, a' \in \sigma$.

Definición 9.5. Sea σ un símplice con vértices en L . Diremos que $x \in \mathbb{R}^n$ es un **witness débil** para σ respecto de L , si $|x - a| \leq |x - b|$, con $a \in \sigma$, $b \in L \setminus \sigma$.

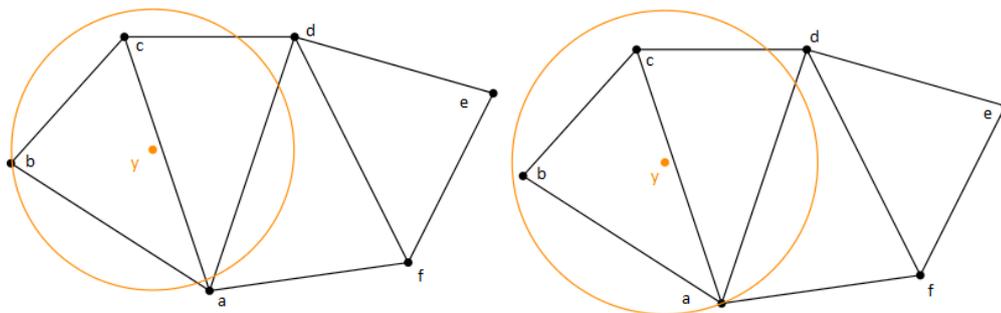
Nota 9.1. Sea $\sigma = [a_0 a_1 \dots a_p]$ un p -símplice con vértices en L . Entonces x es un witness fuerte para σ respecto de L si y solo si es un witness débil y además $|x - a_0| = |x - a_1| = \dots = |x - a_p|$.

Ejemplo 9.2. Las condiciones de witness pueden interpretarse de forma obvia en términos de esferas centradas en x . En la siguiente figura vemos:

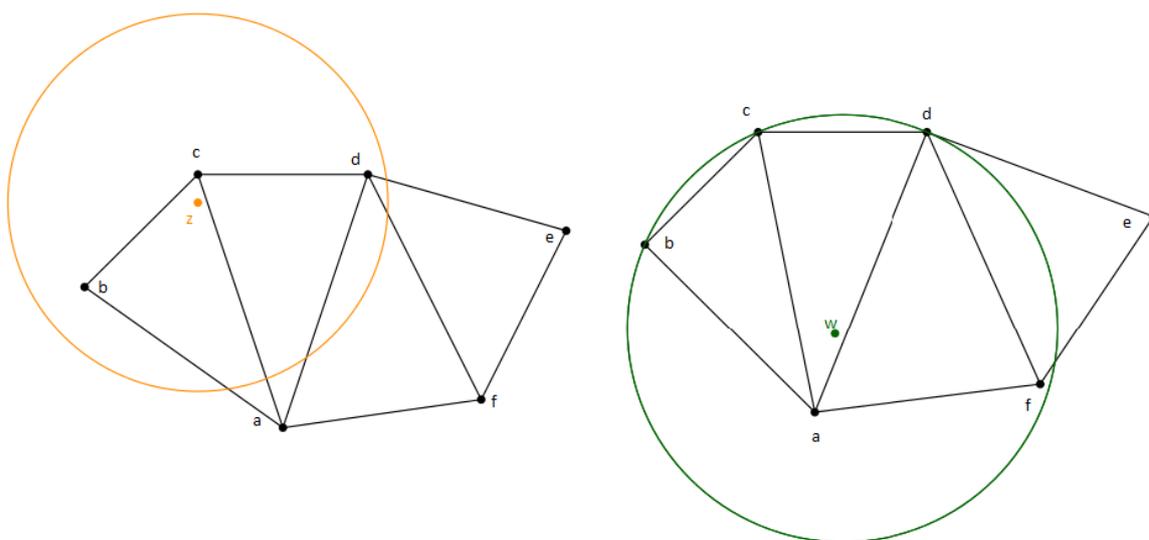
- x es un witness fuerte para el triángulo $[adf]$, ya que es el circuncentro de dicho triángulo, es decir, está a la misma distancia de sus tres vértices:



- y es un witness débil para la arista $[bc]$, y para el triángulo $[abc]$:



- z es un witness débil para $[bcd]$. Además w , el circuncentro de $[bcd]$ está cerca de a , luego $[bcd]$ no tiene un witness fuerte (la distancia de w a los puntos a y f es menor que a los puntos b, c, d , aunque estas tres últimas sean iguales):



Definición 9.6. El **complejo witness fuerte**, $W(L, S)$ es un complejo simplicial abstracto en $L \subset S \subset \mathbb{R}^n$ definido por la siguiente condición:

$$\sigma \in W(L, S) \Leftrightarrow \text{existe algún punto en } S \text{ que es witness fuerte para } \sigma.$$

Si σ tiene un witness fuerte x , entonces x es witness fuerte para toda cara de σ , $\tau \leq \sigma$, con lo que el conjunto que hemos definido, efectivamente tiene estructura de complejo simplicial.

Nota 9.2. Nótese la semejanza entre la construcción de los complejos de Delaunay y los witness fuertes. En efecto, recuérdese que $\sigma \in Del(L)$ si existe una bola cuyo interior no corta a L y cuya frontera contiene a σ . Para que $\sigma \in W(L, S)$ se exige que exista una bola cumpliendo esas condiciones pero centrada en un punto de S . Por tanto, se tiene que $W(L, S) \subset Del(L)$; en particular, $Del(L) = W(L, \mathbb{R}^n)$.

Definición 9.7. El **complejo witness débil**, $W_\infty(L, S)$ es un complejo simplicial abstracto en $L \subset S \subset \mathbb{R}^n$ definido por la siguiente condición:

$$\sigma \in W_\infty(L, S) \Leftrightarrow \text{toda cara } \tau \leq \sigma \text{ tiene un witness débil en } S.$$

Y como trivialmente ser witness fuerte implica ser witness débil, se tiene que x es witness débil para toda cara de σ , $\tau \leq \sigma$. Por lo tanto, tenemos que $W(L, S) \subset W_\infty(L, S)$.

Nota 9.3. En la definición, es necesario exigir que todas las caras tengan un witness débil para que tal y como lo hemos definido, nuestra construcción tenga estructura de complejo simplicial, ya que la existencia de un witness débil para un símplice no asegura que éste sea también witness débil de todas sus caras. Ni siquiera podemos afirmar que otro punto de la muestra sea witness débil de sus caras.

En la siguiente figura, veremos que es posible que un símplice tenga witness débil, pero no todas sus caras:

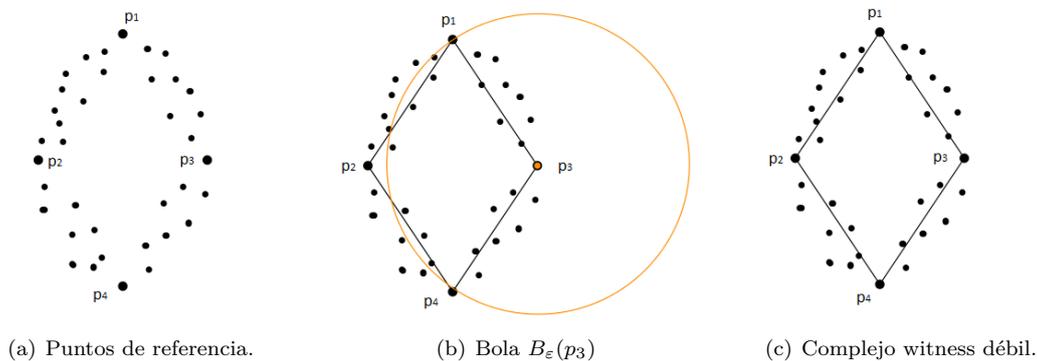


Figura 33. Construcción del complejo witness débil.

(a) Tenemos la nube de puntos, S , y destacados los puntos de referencia escogidos, $L = \{p_1, p_2, p_3, p_4\}$.

- (b) Observamos que $\{p_1, p_3\}$ y $\{p_3, p_4\}$ tienen witness débiles en los puntos de S (análogamente para $\{p_1, p_2\}$, y $\{p_2, p_4\}$).
- (c) Tenemos el complejo witness débil asociado a la nube puntos. En él, vemos que no se encuentran la arista $[p_2, p_3]$ (ni los triángulos $[p_1, p_2, p_3]$, $[p_2, p_3, p_4]$), ya que no existe ningún punto en S que sea witness débil de $\{p_2, p_3\}$, puesto que la distancia entre estos puntos es mayor que la distancia de cada uno de ellos a p_1 o p_4 . Así, aunque haya puntos que sean witness débil de los triángulos $[p_1, p_2, p_3]$, $[p_2, p_3, p_4]$, no hay de la arista $[p_2, p_3]$, por lo que estos triángulos no pertenecen al complejo witness débil.

Existe una versión clique del complejo witness, que es más fácil de obtener mediante los algoritmos usados para los complejos de Vietoris-Rips, y requiere menos memoria para almacenarlo.

Definición 9.8. *Definimos $W_1(L) \supseteq W_\infty(L)$ formalmente como sigue:*

- $W_1(L)$ tiene el mismo 1-esqueleto que $W_\infty(L)$.
- El p -símplice $\sigma = [a_0 a_1 \dots a_p]$ pertenece a $W_1(L)$ si y solo si todas sus aristas pertenecen a $W_1(L)$.

En otras palabras, $W_1(L)$ es el mayor complejo simplicial que tiene los mismos vértices y aristas que $W_\infty(L)$. En la práctica, rara vez usaremos $W_\infty(L)$ en lugar de $W_1(L)$, ya que su construcción es computacionalmente costosa.

9.3.1. Relación entre witness débil y fuerte

Una diferencia fundamental entre los dos conceptos definidos (witness débil y witness fuerte) es que los witness débiles de un símplice generan un conjunto con medida positiva mientras que los witness fuertes forman un conjunto de medida cero. De ello se deduce que la probabilidad de encontrar un testigo fuerte por muestreo es cero.

Esto nos lleva al siguiente resultado, que implica que la existencia de testigos débiles se puede utilizar para deducir la existencia de testigos fuertes.

Teorema 9.3 (de Silva, [14]). *Sea un conjunto finito $L \subset S \subset \mathbb{R}^n$ y un p -símplice $\sigma = (a_0, \dots, a_p) \subset L$. Entonces σ tiene un witness fuerte respecto a L si y solo si σ y todas sus caras tienen witnesses débiles respecto a L .*

Demostración

Supongamos que para toda cara $\tau \leq \sigma$ existe un witness débil $x_\tau \in \mathbb{R}^n$ para τ respecto a L . Veremos que la envolvente convexa de los puntos x_τ contiene un witness fuerte para σ respecto a L .

Usaremos un argumento de inducción en la dimensión de σ , el caso $\dim(\sigma) = 0$ es trivial.

Denotaremos por $R(a, b) = \{x \in \mathbb{R}^n \mid |x - a| \leq |x - b|\}$ para $a, b \in L$; obsérvese que estos conjuntos son convexos.

Supongamos que $\dim(\sigma) \geq 1$, y sean $a_0, a_1 \in \sigma$ distintos. Consideremos el subespacio

$$H = R(a_0, a_1) \cap R(a_1, a_0) = \{x \in \mathbb{R}^n \mid |x - a_0| = |x - a_1|\}$$

y sea $\sigma' = \sigma \setminus \{a_0\}$ y $L' = L \setminus \{a_0\}$.

Para cada $\tau \leq \sigma'$ encontraremos un witness débil $y_\tau \in H$ para τ respecto a L' . Por hipótesis de inducción, podemos encontrar un witness fuerte x' para σ' respecto a L' .

Como H es convexo, por construcción tenemos $x' \in H$. Entonces x' es un witness fuerte para σ respecto a L . Es más, los puntos y_τ están en el recubrimiento convexo del witness débil original x_τ , por lo que es cierto también para x' .

Queda entonces localizar un witness débil y_τ , para cada $\tau \leq \sigma'$. En las Figuras 34, 35 vemos los cuatro casos.

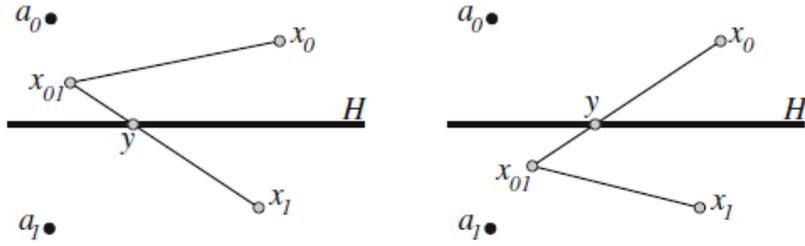


Figura 34. Localizando y cuando $a_1 \in \tau$.

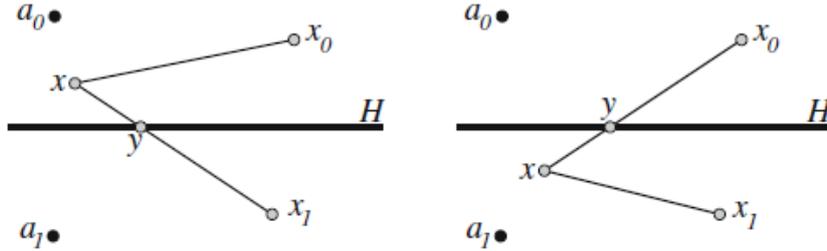


Figura 35. Localizando y cuando $a_1 \notin \tau$.

Supongamos primero que $a_1 \in \tau$, y escribimos $\tau = \nu \cup \{a_1\}$ donde $a_1 \notin \nu$.

Sean x_0, x_1, x_{01} los witness débiles para $\nu \cup \{a_0\}, \nu \cup \{a_1\}, \nu \cup \{a_0, a_1\}$ respecto a L . Por definición, $x_0 \in R(a_0, a_1)$ y $x_1 \in R(a_1, a_0) \Rightarrow x_{01} \in R(a_0, a_1)$ o $x_{01} \in R(a_1, a_0)$.

- Si $x_{01} \in R(a_0, a_1)$, por continuidad se tiene que el segmento $[x_{01}, x_1]$ corta con H en algún punto y . Como $x_{01}, x_1 \in R(a, b)$ para todo $a \in \tau, b \in L' \setminus \tau$, entonces lo mismo es cierto para y (por convexidad), luego y es el testigo débil buscado para τ .

- Si $x_{01} \in R(a_1, a_0)$, entonces el segmento $[x_{01}, x_0]$ debe cortar con H en el mismo punto y . Como $x_{01}, x_0 \in R(a, b)$ para todo $a \in \nu \cup \{a_0\}$, $b \in L' \setminus \tau$, entonces lo mismo es cierto para y . Es más, $|y - a_0| = |y - a_1|$, luego $y \in R(a_1, b)$ para todo $b \in L' \setminus \tau$, y de nuevo, y es un witness débil para τ .

Supongamos ahora que $a_1 \notin \tau$:

Sean x, x_0, x_1 los witness débiles para $\tau, \tau \cup \{a_0\}, \tau \cup \{a_1\}$ respecto a A . Por definición, $x_0 \in R(a_0, a_1)$ y $x_1 \in R(a_1, a_0)$. Luego, $x \in R(a_0, a_1)$ o $x \in R(a_1, a_0)$. Si $x \in R(a_0, a_1)$, entonces el segmento $[x, x_0]$ debe cortar con H en algún punto y .

Como $x, x_0 \in R(a, b)$ para todo $a \in \tau$, $b \in L' \setminus \tau$, entonces lo mismo es cierto para y , y de nuevo y es un witness débil para τ .

Q. E. D.

Corolario 9.1. *Si $S = \mathbb{R}^d$, entonces $W(L, \mathbb{R}^d) = W_\infty(L, \mathbb{R}^d)$.*

9.3.2. Generalizaciones

Al igual que con los complejos ya estudiados, para que los complejos witness permitan obtener información relevante del conjunto de datos dado, es conveniente obtener una filtración a partir de ellos. Ello es posible considerando una familia de tales complejos en función de un parámetro de proximidad, $\varepsilon \geq 0$, que da lugar a una familia anidada de complejos simpliciales adecuadas para la homología persistente.

Cuando hacemos variar ε , la apariencia, el tiempo de vida y la desaparición de las clases de homología, nos dan información topológica detallada más fiable que los números de Betti del complejo para cualquier valor de ε .

9.3.2.1. Complejo de Witness uniparámetro En nuestro caso, veremos entonces que para obtener complejos witness anidados, debemos extender la definición de witness débil a ε -witness.

Definición 9.9. *Sea σ un q -símplice con vértices en L , diremos que $x \in S$ es ε -**witness** para $\sigma = (l_0, \dots, l_q)$ respecto de L , si*

$$d(x, l_i) \leq d(x, L \setminus \sigma) + \varepsilon, \forall i = 0, \dots, q.$$

Definición 9.10. *El **complejo ε -witness**, $W_\infty(L, S, \varepsilon)$ es un complejo simplicial abstracto con vértices en $L \subset S \subset \mathbb{R}^n$ definido por la siguiente condición:*

$$\sigma \in W_\infty(L, S, \varepsilon) \Leftrightarrow \text{toda cara } \tau \leq \sigma \text{ tiene un } \varepsilon\text{-witness en } S.$$

Al igual que antes, se tiene la correspondiente versión clique, que da lugar a la filtración $W_1(L, S, \varepsilon)$.

Cuando ε es pequeño, el complejo tiene muy pocos simplices y por tanto puede que no sea una buena representación de la forma del conjunto de datos. A medida que ε crece, podemos encontrar más puntos witness, y el complejo se va ampliando, hasta que para un ε muy grande, incluso los agujeros de mayor dimensión se van rellenando.

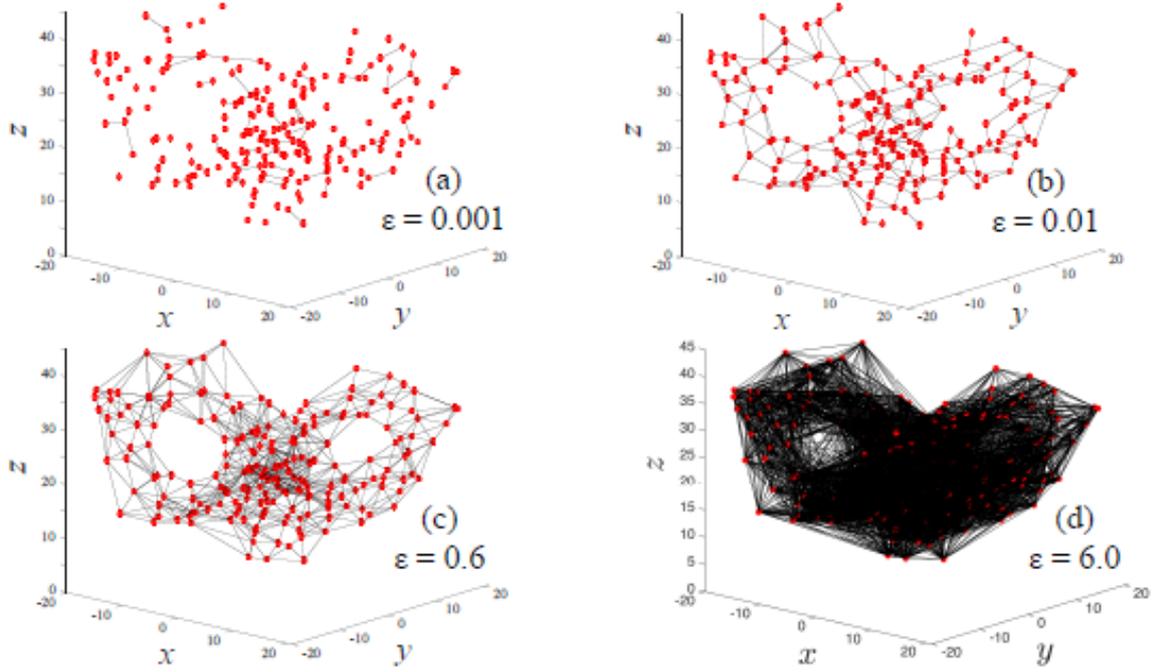


Figura 36. Variación de los complejos ε -witness según el parámetro ε .

9.3.2.2. Complejos de Witness biparamétrico Construiremos la familia de complejos simpliciales $W_1(L, \varepsilon, \nu)$, donde $\varepsilon \in [0, \infty]$ y ν es un entero no negativo, extendiendo la definición de complejo witness. La familia está doblemente anidada en ε y ν , luego si $\varepsilon \leq \varepsilon'$ y $\nu \leq \nu'$, entonces $W(L, \varepsilon, \nu) \subseteq W_1(L, \varepsilon', \nu')$.

Sea $m_\nu(q)$, la distancia entre q y el ν -ésimo punto de referencia más cercano y $m_0(q) = 0, \forall q \in S$.

Definición 9.11. El complejo (ε, ν) -witness $W_1(L, \varepsilon, \nu)$ se define como:

- (i) L es el conjunto de vértices.
- (ii) El simplex de dimensión $p \geq 1$, $\sigma = [a_0 a_1 \dots a_p]$, está en el complejo $W_1(L, \varepsilon, \nu)$ si existe un witness $q \in S$ tal que

$$\max\{d(a_0, q), d(a_1, q), \dots, d(a_p, q)\} \leq \varepsilon + m_\nu(q)$$

Podemos destacar los casos particulares en los que $\nu = 0, 1, 2$.

- $\nu = 0$: La familia de complejos $W_1(L, \varepsilon, 0)$ está íntimamente relacionada con la familia $VR\varepsilon(L)$. Concretamente, se tienen las siguientes inclusiones:

$$W_1(L, \varepsilon, 0) \subseteq VR_{2\varepsilon}(L) \subseteq W_1(L, 2\varepsilon, 0).$$

- $\nu = 1$: En cierto modo, este es el caso más interesante de los tres, ya que puede interpretarse como el resultado de una familia de recubrimientos del espacio X por regiones de Voronoi que rodean a cada punto de interés, y que se superponen cada vez más a medida que ε toma valores muy altos.
- $\nu = 2$: A pesar de que esta familia persistente no es tan relevante como la del caso anterior, tenemos la siguiente igualdad para $\varepsilon = 0$:

$$W_1(L, 0, 2) = W_1(L, S).$$

9.3.3. Criterio para elegir los puntos de referencia

Los puntos de referencias $L \subset S$ se eligen para ser tratados como el conjunto de vértices del complejo y se asume que están bien distribuidos sobre los datos. Los complejos witness se basan en la idea de que los puntos que no se han escogido como puntos de referencia, es decir, $S - L$ pueden usarse para determinar las aristas y los símlices de mayores dimensiones del complejo: la arista $[l_i, l_j]$ entre dos puntos de referencia, se incluye en el complejo si existe un punto cuyos dos vecinos más cercanos en el conjunto de puntos de referencia son l_i , y l_j .

Se suele escoger el conjunto de los puntos de referencia de una de estas dos formas: aleatoriamente, o por el método maxmin. Ambos métodos parecen dar resultados razonables. El método maxmin es el siguiente procedimiento inductivo. Comenzamos seleccionando $l_1 \in Z$ tal que $\{l_1, l_2, \dots, l_{i-1}\}$ sea el conjunto de puntos que maximiza la función

$$z \mapsto \min\{d(z, l_1), d(z, l_2), \dots, d(z, l_{i-1})\},$$

Continuamos hasta que obtengamos el número de puntos de referencia deseado.

Esta forma de maxmin nos da unos puntos de referencia más espaciados, pero también tiende a tomar los más extremos, como se aprecia en la Figura 37. Para más detalles, véase [16].



Figura 37. Puntos de referencia tomados aleatoriamente a la izquierda y por maxmin a la derecha.

Nota 9.4. Observamos que el complejo witness depende de los puntos de referencia escogidos. La elección y el tamaño del conjunto de puntos de referencia es más un arte que una ciencia, por ello, usualmente, el complejo witness es uno de los métodos menos prácticos para el análisis topológico de grandes conjuntos de datos. Para el análisis, lo mejor es empezar eligiendo múltiples conjuntos de puntos de referencia y ver si el resultado es replicable.

Bibliografía

- [1] Ahtziri, J. *Teoría de Morse discreta y gráficas de Reeb aplicadas a ATD*. Centro de Investigación Matemática (CIMAT).
- [2] Alexander, Z., Bradley, E., Meiss, J. D. & Sanderson, N. F. (2015). *Simplicial multivalued maps and the witness complex for dynamical analysis of time series*. SIAM Journal on Applied Dynamical Systems, 14(3), 1278-1307.
- [3] Artamonov, O. (2010) *Topological methods for the representation and analysis of exploration data in oil industry* (Doctoral dissertation, Kaiserslautern, Techn. Univ., Diss., 2010).
- [4] Attali, D., Edelsbrunner, H. & Mileyko, Y. (2007, June). *Weak witnesses for Delaunay triangulations of submanifolds*. En *Proceedings of the 2007 ACM symposium on Solid and physical modeling* (pp. 143-150). ACM.
- [5] Bernadette J. Stolz, Heather A. Harrington & Mason A. Porter. (2016). *The Topological “Shape” of Brexit*. University of Oxford.
- [6] Boissonnat, J., Chazal, F. & Yvinec, M. (2016). *Geometry and Topology Inference*.
- [7] Boissonnat, J. D., Dyer, R., & Ghosh, A. (2015). *A probabilistic approach to reducing the algebraic complexity of computing Delaunay triangulations*. arXiv preprint arXiv:1505.05454.
- [8] Botnan, M. B. (2011). *Three Approaches in Computational Geometry and Topology: Persistent Homology, Discrete Differential Geometry and Discrete Morse Theory*.
- [9] Capitelli, N. & Minian, G. (2009) *Notas de Topología Combinatoria*. Unión Matemática Argentina, Reunión Anual 2009, Mar del Plata.
- [10] Chazal, F. *Statistical properties of topological information inferred from data*.
- [11] Cheng, S.W., Dey, T. K. & Shewchuck, J. (2012). *Delaunay Mesh Generation* CRC Press.
- [12] Cirafici, M. (2015). *Persistent homology and string vacua*. arXiv preprint arXiv:1512.01170.
- [13] De Berg, M., Cheong, O., Van Kreveld, M. & Overmars, M. (2008). *Computational Geometry: Algorithms and Applications*. Springer Science & Business Media.
- [14] De Silva, W. (2008) *A weak characterisation of the Delaunay triangulation*. Geometriae Dedicata, 135(1), 39-64.
- [15] De Silva, V. & Ghrist, R. (2007) *Coverage in sensor networks via persistent homology*. Alg. and Geom. Topology No. 7, 339-358.
- [16] De Silva, & V. Carlsson, G. (2003). *Topological approximation by small simplicial complexes*. preprint.
- [17] De Silva, V. & Carlsson, G. (2004). *Topological estimation using witness complexes*. Proc. Sympos. Point-Based Graphics, 157-166.

- [18] Domínguez, E. & Quintero, A. (2002). *Elementos de la teoría de homología clásica* (Vol. 65). Universidad de Sevilla.
- [19] Edelsbrunner, H. & Harer, J. (2010). *Computational topology: an introduction*. American Mathematical Soc.
- [20] Edelsbrunner, H. (1995) *The union of balls and its dual shape*. *Discrete & Computational Geometry*, 13(3-4), 415-440.
- [21] Erdmann, M. (2009) *On the topology of discrete strategies*. *The International Journal of Robotics Research*.
- [22] Espinosa, M. E. (2015). *Homología persistente*. Centro de Investigación Matemática (CIMAT).
- [23] Gao, J., Guibas, L. J., Oudot, S. Y. & Wang, Y. (2008). *Geodesic Delaunay triangulation and witness complex in the plane*. En *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms* (pp. 571-580). Society for Industrial & Applied Mathematics.
- [24] Ghrist, R. (2008) *Barcodes: the persistent topology of data*. *Bulletin of the American Mathematical Society*, 45(1), 61-75.
- [25] Ghrist, R. (2014) *Elementary applied topology*. Createspace.
- [26] Guibas, L. J. & Oudot, S. Y. (2008). *Reconstruction using witness complexes*. *Discrete & computational geometry*, 40(3), 325-356.
- [27] Guzmán, A. *Análisis topológico de datos: teoría y práctica*.
- [28] Hatcher, A. (2002) *Algebraic Topology*. Cambridge University Press.
- [29] Khoury, M. *Introduction to Simplicial Homology. Topics in Computational Topology: An Algorithmic View*.
- [30] Macho, M. (2006). *Topología Algebraica*. Departamento de Matemáticas, Facultad de Ciencia y Tecnología, Universidad del País Vasco-Euskal Herriko Unibertsitatea.
- [31] Martínez, B. (2015). *Homología Persistente en el Análisis Topológico de Datos*. Tesis dirigida por Esteban, E.. Universidad Nacional Autónoma de México, Facultad de Ciencias.
- [32] Minian, E. G. (2004) *Notas de Topología Algebraica. Curso dictado en Elena II, Vaquerías*. Universidad Nacional de Córdoba, Fa. MAF.
- [33] Oman, E. (2013). *An introduction to Computational Cubical Homology*.
- [34] Otter, N., Porter, M. A., Tillmann, U., Grindrod, P. & Harrington, H. A. (2015) *A roadmap for the computation of persistent homology*. arXiv preprint arXiv:1506.08903.
- [35] Past, J. E. (2015). *Applications of Persistent Homology for Image Analysis*.
- [36] Penide, G. (2016) *Homología persistente de redes complejas*. Trabajo de Fin de Máster, Universidade de Santiago de Compostela.

- [37] Rolón, Z. & Leidy C. (2015). *Sobre perspectiva y homología de Imágenes*.
- [38] Spanier, E. H. (1994). *Algebraic topology* (Vol. 55, No. 1). Springer Science & Business Media.
- [39] Synak, P. (2015) *Diplomová práce*. Masarykova Univerzita, Přírodovědecká Fakulta, Ústav Matematiky a statistiky.
- [40] Tadeo Ibarra, J. *Homología Singular, Construcción, Ejemplos y Aplicaciones*.
- [41] Tamal, K. Dey. *CSE 5339: Computational Topology and Data Analysis. (Topic 2: Complexes on data. (a. Simplicial Complexes))*. The Ohio State University.
- [42] Yin, Q. & Wang, W. (2015) *An Analysis of Spaces of Range Image Small Patches*. Open Cybernetics, Systemics Journal, 9, 275-279.
- [43] Zomorodian, A. J. (Ed.) (2012). *Advances in Applied and Computational Topology: American Mathematical Society Short Course on Computational Topology, January 4-5, 2011, New Orleans, Louisiana* (Vol. 70). American Mathematical Soc.
- [44] Zomorodian, A. (2010). *Fast construction of the Vietoris-Rips complex*. Computers & Graphics, 34(3), 263-271.
- [45] Zhu, X. (2013). *Persistent Homology: An Introduction and a New Text Representation for Natural Language Processing*.
- [46] Hudson, J. F. (1969). *Piecewise linear topology*. Benjamin, New York.