

GABRIEL PENIDE CALVO

**HOMOLOGÍA PERSISTENTE
DE REDES COMPLEJAS**

132b

2017

Publicaciones
del
Departamento
de Geometría y Topología

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

GABRIEL PENIDE CALVO

**HOMOLOGÍA PERSISTENTE
DE REDES COMPLEJAS**

132b

2017

Publicaciones
del
Departamento
de Geometría y Topología

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

© Universidade de Santiago de Compostela, 2017



Esta obra atópase baixo unha licenza internacional Creative Commons BY-NC-ND 4.0. Calquera forma de reprodución, distribución, comunicación pública ou transformación desta obra non incluída na licenza Creative Commons BY-NC-ND 4.0 só pode ser realizada coa autorización expresa dos titulares, salvo excepción prevista pola lei. Pode acceder Vde. ao texto completo da licenza nesta ligazón: <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.gl>



Esta obra se encuentra bajo una licencia internacional Creative Commons BY-NC-ND 4.0. Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra no incluida en la licencia Creative Commons BY-NC-ND 4.0 solo puede ser realizada con la autorización expresa de los titulares, salvo excepción prevista por la ley. Puede Vd. acceder al texto completo de la licencia en este enlace: <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



This work is licensed under a Creative Commons BY NC ND 4.0 international license. Any form of reproduction, distribution, public communication or transformation of this work not included under the Creative Commons BY-NC-ND 4.0 license can only be carried out with the express authorization of the proprietors, save where otherwise provided by the law. You can access the full text of the license at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>

MÁSTER EN MATEMÁTICAS

Traballo Fin de Máster

Homología persistente de redes complejas

Gabriel Penide Calvo

Xullo 2016

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

May 17, 2011

To Whom it May Inspire,

I, like many of you artists out there, constantly shift between two states. The first (and far more preferable of the two) is white-hot, “in the zone” seat-of-the-pants, firing on all cylinders creative mode. This is when you lay your pen down and the ideas pour out like wine from a royal chalice! This happens about 3% of the time.

The other 97% of the time I am in the frustrated, struggling, office-corner-full-of-crumpled-up-paper mode. The important thing is to slog diligently through this quagmire of discouragement and despair. Put on some audio commentary and listen to the stories of professionals who have been making films for decades going through the same slings and arrows of outrageous production problems.

In a word: PERSIST.

PERSIST on telling your story. PERSIST on reaching your audience. PERSIST on staying true to your vision. Remember what Peter Jackson said, “Pain is temporary. Film is forever.” And he of all people should know. So next time you hit writer’s block, or your computer crashes and you lose an entire night’s work because you didn’t hit save (always hit save), just remember: you’re never far from that next burst of divine creativity. Work through that 97% of murky abyssmal mediocrity to get to that 3% which everyone will remember you for!

I guarantee you, the art will be well worth the work!

Your friend and mine,

Austin Madison

“ADVENTURE IS OUT THERE!”

Índice general

| | |
|--|-----------|
| Resumen | 7 |
| Introducción | 9 |
| 1. Homología simplicial | 13 |
| 1.1. Complejos simpliciales | 13 |
| 1.2. Complejos simpliciales abstractos | 15 |
| 1.3. Homología simplicial orientada | 17 |
| 1.4. Invariancia homotópica | 21 |
| 1.5. Computabilidad | 22 |
| 1.5.1. Forma normal de Smith | 22 |
| 1.5.2. Algoritmo para las matrices de incidencia | 24 |
| 2. Construcción de complejos simpliciales | 35 |
| 2.1. Complejo de Čech | 35 |
| 2.2. Complejo de Vietoris-Rips | 38 |
| 2.3. Complejo de Delaunay | 40 |
| 2.4. Otros complejos simpliciales | 42 |
| 3. Homología persistente | 43 |
| 3.1. Filtraciones | 43 |
| 3.1.1. Filtraciones con índices naturales | 44 |
| 3.1.2. Filtraciones con índices reales | 46 |
| 3.2. Códigos de barras | 48 |
| 3.2.1. Intervalos de persistencia | 49 |
| 3.2.2. Filtraciones de Čech y Vietoris-Rips | 50 |
| 3.3. Estructuras algebraicas | 51 |
| 3.3.1. Módulos y complejos de persistencia | 52 |
| 3.3.2. Clasificación sobre cuerpos | 53 |
| 3.3.3. Algoritmo de cálculo sobre cuerpos | 56 |
| 3.4. Notas históricas | 64 |

| | |
|--|-----------|
| 4. Análisis de redes complejas | 67 |
| 4.1. Teoría de grafos | 67 |
| 4.1.1. Conceptos básicos | 68 |
| 4.1.2. Complejo clique | 69 |
| 4.1.3. Distribuciones de grado | 70 |
| 4.2. Herramientas de computación | 70 |
| 4.3. Redes aleatorias de Erdős-Rényi | 71 |
| 4.4. Red de distribución exponencial | 73 |
| 4.5. Red de colaboración científica | 75 |
| 4.6. Conclusiones | 79 |
| A. Código para implementar los ejemplos | 81 |
| A.1. Código en R | 81 |
| A.2. Código en MATLAB | 82 |
| Bibliografía | 85 |

Resumen

El análisis topológico de datos y de estructuras complejas con un alto número de unidades interdependientes se ha convertido en una de las ramas más activas de las matemáticas. Las enormes cantidades de datos que se manejan en la actualidad y el descubrimiento de nuevos tipos de redes en biología, informática y ciencias sociales han obligado a desarrollar técnicas novedosas de procesamiento que permitan revelar las estructuras topológicas subyacentes. La homología persistente es una de ellas y sirve para identificar las propiedades topológicas relevantes en una nube de datos o para codificar un grafo y descartar aquellas características que son simplemente ruido o no sobreviven a un análisis más fino. En este trabajo presentamos las definiciones básicas de homología persistente, su codificación mediante unos diagramas, conocidos como códigos de barras, que muestran visualmente aquellas características topológicas que perduran a lo largo del tiempo, y damos algunas aplicaciones para el estudio de diversos tipos de grafos importantes en el análisis de redes complejas.

Abstract

Topological data and complex structure analysis with a high number of interdependent units have become one of the most active branches of mathematics. Nowadays, huge quantities of data are handled, and so the discovery of new types of networks in biology, computing and social science has led to the development of newfangled processing techniques so as to reveal the underlying topological structures. Persistent homology is one of these techniques and allows the identification of relevant topological properties within a data cloud or the codification of a graph in order to dismiss noisy features or the ones that do not survive to a more refined analysis. Basic definitions are introduced to deal with persistent homology, as well as certain diagrams, known as barcodes, which help on the visualization of those topological features that persist through time, and some applications are given to study different types of graphs in complex network analysis.

Introducción

El análisis topológico de datos reúne técnicas que posibilitan el estudio y tratamiento de grandes cantidades de datos desde el punto de vista de la topología. Estos métodos deben afrontar algunos problemas básicos. En primer lugar, los conjuntos de datos son normalmente vectores muy largos, de modo que se hace muy difícil su visualización al pertenecer a espacios de dimensión grande. Por otra parte, es habitual que existan pérdidas de información debido a una falta de precisión en la obtención de los datos o que la muestra no se haya tomado uniformemente. Finalmente, un tercer problema es la aparición de ruido, que puede desvirtuar la interpretación global de los datos.

La homología persistente surge como una técnica del campo de la topología computacional que estudia conjuntos de datos a distintas resoluciones o escalas. Supongamos que tenemos una nube de datos procedente de un muestreo de puntos correspondientes a un espacio topológico que modele una situación experimental. El objetivo perseguido es conseguir discernir la topología del espacio subyacente utilizando únicamente esos puntos como referencia. Pero si queremos usar herramientas matemáticas debemos dotar de una mayor estructura a los puntos, sabiendo que algunos de ellos se originaron a partir de datos erróneos o ruido topológico. Los complejos simpliciales son construcciones que codifican el espacio en una estructura combinatoria y dejan al descubierto las propiedades topológicas de interés mediante el cálculo de su homología simplicial.

Los complejos simpliciales que manejaremos dependen de un parámetro. La homología persistente es la homología simplicial calculada a través de un rango de valores de ese parámetro. Esto da una visión más dinámica de la situación, ya que se calcula la homología de una filtración o sucesión creciente de complejos simpliciales que en cada etapa representan aproximaciones más finas del espacio. Por otra parte, se elimina la necesidad de buscar un parámetro óptimo para el que la homología simplicial se ajuste a la realidad de los datos. Este detalle es importante, pues a menudo el parámetro óptimo no es fácil de determinar y existe un parámetro óptimo por cada característica diferente.

Asimismo, la homología persistente es una estructura algebraica capaz de almacenar y codificar las propiedades topológicas escondidas en un muestreo de puntos visto a través de una filtración. La codificación de la información se hace a través de unos diagramas llamados códigos de barras. Estos diagramas constituyen un invariante y permiten apreciar visualmente el tiempo de vida de las características topológicas. Aquellas propiedades intrínsecas perduran en las filtraciones, mientras que el ruido aparece y rápidamente des-

aparece, por lo que un vistazo rápido al código de barras proporciona una estimación fiable de la topología [7].

La existencia de este invariante, junto con algoritmos eficientes de cálculo, ha propiciado la aparición de numerosas aplicaciones prácticas de la teoría. Entre ellas, la descripción de formas [11], la eliminación de ruido en datos de densidad volumétricos [24], la detección de agujeros en redes sensoriales [15], el análisis de la actividad neuronal del córtex visual [39], el análisis de la estructura de imágenes naturales [6, 22], el reconocimiento de la forma de caminar humana [31] o el análisis de redes complejas [9, 27] que veremos al final de la memoria.

La memoria se ha estructurado de forma que el capítulo 1 empiece tratando los complejos simpliciales y los complejos simpliciales abstractos. Se pasa después a exponer la homología simplicial y las propiedades que se necesitan para presentar y desarrollar correctamente la teoría de homología persistente en un capítulo posterior. Se dedica un apartado a tratar la invariancia homotópica, resultado de gran importancia en la teoría. Finalmente, el capítulo describe cómo se utilizan la forma normal de Smith de una matriz y el teorema de clasificación de módulos finitamente generados sobre un dominio de ideales principales para conseguir el algoritmo clásico de cálculo de la homología simplicial. Se trata de un algoritmo simple que constituye la base de otros algoritmos modernos más eficientes.

El segundo capítulo se dedica a introducir, entre otros, el complejo de Čech y el complejo de Vietoris-Rips. Estas son las dos construcciones clásicas de complejos simpliciales a partir de nubes de puntos. Seguidamente, se repasan algunas propiedades que relacionan estos dos complejos simpliciales entre sí y con el espacio topológico que representan y, para terminar, se mencionan variantes de los mismos.

A continuación, el capítulo 3 comienza tratando las filtraciones de complejos simpliciales para dar paso, por fin, a definir la homología persistente. Se consideran dos tipos de filtraciones, dependiendo de si el parámetro de la filtración toma valores naturales o reales. Los códigos de barras que después se presentan resumen visualmente los cambios que sufre el complejo simplicial a lo largo de una filtración. Por otra parte, el capítulo también define los complejos de persistencia y los módulos de persistencia y establece un teorema de clasificación de los módulos de persistencia sobre cuerpos para filtraciones con índices naturales. Este teorema de clasificación permite afirmar que los códigos de barras son invariantes de la homología persistente. Posteriormente, se desarrolla y describe un algoritmo de cálculo de la homología persistente basado únicamente en álgebra lineal con coeficientes en un anillo de polinomios. Para acabar, mediante un ejemplo se ilustran los pasos del algoritmo y se incluyen unas breves notas históricas sobre la persistencia.

El último capítulo se inicia con una revisión básica de la teoría de grafos. Brevemente, se hace una exposición de las herramientas computacionales utilizadas para los análisis posteriores. El bloque central del capítulo es la exposición de una aplicación práctica de la homología persistente en el estudio de redes complejas. En particular, se consideran dos redes aleatorias y dos redes con datos reales: una red de correos electrónicos que sigue una distribución exponencial y una red de colaboración científica. Los análisis sirven para

observar que las propiedades persistentes tienen relación con la robustez de las redes e, incluso, permiten distinguir las topologías de redes de construcción similar.

Se incluye un apéndice con los códigos desarrollados en los softwares R y MATLAB que sirven para recrear los ejemplos de redes aleatorias analizados en el capítulo 4.

La memoria finaliza con la bibliografía, que incluye las referencias citadas a lo largo del texto. Cada capítulo tiene una referencia principal indicada al comienzo, aunque secciones concretas mencionan puntualmente otras fuentes consultadas.

Capítulo 1

Homología simplicial

Uno de los problemas fundamentales de la topología es conseguir determinar si dos espacios topológicos son o no homeomorfos o, más generalmente, si son homotópicamente equivalentes. Dar una respuesta satisfactoria a esta pregunta puede resultar muy difícil la mayoría de las veces. Estas dificultades derivaron en el surgimiento de la topología algebraica y en el estudio de propiedades que permaneciesen invariantes bajo determinadas transformaciones.

La homología simplicial surgió durante esta búsqueda de invariantes topológicos. La homología simplicial formaliza la idea de encontrar los agujeros de una determinada dimensión que tiene un cierto espacio. Sin embargo, la homología simplicial necesita una estructura poliédrica subyacente, lo que no la hace válida para espacios que no sean suficientemente «buenos». De esta manera, esta teoría dio paso al estudio de la conocida como homología singular, que sí permitía desarrollar la teoría en espacios topológicos arbitrarios. Además, ambas homologías coinciden para aquellos espacios donde la primera se puede calcular. A pesar de esto, la homología singular resulta de muy difícil cálculo en muchos casos.

La irrupción y desarrollo de la computación a finales del siglo XX provocó que los complejos simpliciales cobraran de nuevo importancia en la topología algebraica debido a las posibilidades que brindan los ordenadores para su tratamiento. Este hecho y sus aplicaciones en muchos ámbitos cotidianos provocaron el regreso al estudio de la homología simplicial.

La exposición del capítulo sigue en gran parte [2], incorporando también elementos de [34].

1.1. Complejos simpliciales

La idea básica para entender el origen de los complejos simpliciales es verlos como una generalización de las triangulaciones de las superficies. Decimos que los $n + 1$ puntos v_0, v_1, \dots, v_n de \mathbb{R}^m son *afínmente independientes* si los vectores $v_1 - v_0, \dots, v_n - v_0$ son linealmente independientes. Dos puntos distintos en \mathbb{R}^m son afínmente independientes, así como tres puntos no alineados, cuatro puntos no coplanares, y así sucesivamente. La

noción de independencia sirve para definir la estructura básica utilizada para construir la homología simplicial.

Definición 1.1. Un *símplice de orden n* o *n -símplice* es la envoltura convexa de $n + 1$ puntos afínmente independientes. En otras palabras, dados v_0, v_1, \dots, v_n puntos afínmente independientes de \mathbb{R}^m se llama n -símplice al conjunto

$$\sigma = \left\{ x = \sum_{i=0}^n \lambda_i v_i \in \mathbb{R}^m : \sum_{i=0}^n \lambda_i = 1, \lambda_i \geq 0 \text{ para todo } i \right\}.$$

Los coeficientes λ_i se conocen como *coordenadas baricéntricas* y los puntos v_i son los *vértices* del símplice, con $i = 0, \dots, n$.

Habitualmente, denotaremos un n -símplice mediante un conjunto de $n + 1$ elementos formado por sus vértices, $\sigma = \{v_0, \dots, v_n\}$, o mediante una yuxtaposición de los mismos, $\sigma = v_0 \cdots v_n$.

Definición 1.2. Sean σ y τ dos símplices. Se dice que τ es una *cara* de σ , y lo denotamos por $\tau \leq \sigma$, si todos los vértices de τ son vértices de σ . Si $\tau \neq \sigma$ y $\tau \leq \sigma$, diremos que τ es una *cara propia* de σ y escribiremos $\tau < \sigma$.

Los complejos simpliciales son familias de símplices que cumplen dos propiedades que recuerdan a las que verifican las triangulaciones.

Definición 1.3. Un *complejo simplicial finito* K es una colección finita de símplices verificando las siguientes condiciones:

- (I) Si $\sigma_1, \sigma_2 \in K$, entonces la intersección $\sigma_1 \cap \sigma_2$ o bien es vacía o bien es una cara común a σ_1 y a σ_2 .
- (II) Si $\sigma \in K$ y $\tau \leq \sigma$, entonces $\tau \in K$.

Un *subcomplejo* de K es una subcolección de símplices de K que sigue siendo un complejo. La *dimensión* de K es el número $\dim K = \max\{\dim \sigma : \sigma \in K\}$.

Observación 1.4. Todo n -símplice σ determina un complejo simplicial finito de dimensión n formado por él mismo y cada una de sus caras.

Sea K un complejo simplicial en \mathbb{R}^m . De considerar todos los puntos de los símplices de K surge un subespacio de \mathbb{R}^m , que llamaremos *poliedro subyacente a K* y que podemos expresar como sigue:

$$|K| = \bigcup_{\sigma \in K} \sigma.$$

Más adelante dotaremos de una topología al poliedro subyacente y veremos su importancia, pues la homología de un espacio X dependerá únicamente del tipo de homotopía de $|K|$, para un complejo simplicial K asociado a X adecuado.

1.2. Complejos simpliciales abstractos

Pese a su sencillez, la gran rigidez que poseen los complejos simpliciales definidos en la sección anterior obligó a buscar una definición un poco más abstracta que permitiese una mayor facilidad de manipulación.

Definición 1.5. Sea V un conjunto. Un *complejo simplicial abstracto* \mathcal{K} es una colección no vacía de partes finitas de V verificando las siguientes condiciones:

- (I) $\{v\} \in \mathcal{K}$ para todo $v \in V$.
- (II) Dado $\sigma \in \mathcal{K}$, todo subconjunto de σ pertenece a \mathcal{K} .

Los elementos del conjunto V se llaman *vértices* de \mathcal{K} y los elementos de \mathcal{K} son los *símplices* de \mathcal{K} . Además, cualquier subcolección de \mathcal{K} que siga siendo un complejo se denomina *subcomplejo* de \mathcal{K} .

La dimensión de un símplex $\sigma \in \mathcal{K}$ es el cardinal de σ como conjunto menos una unidad. La dimensión de un complejo \mathcal{K} se define como el supremo de las dimensiones de sus símplexes (puede no ser finita).

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

Es claro que a todo complejo simplicial K se le puede asociar un complejo simplicial abstracto \mathcal{K} . En primer lugar, se toman como vértices de \mathcal{K} los vértices de K etiquetados de alguna forma. Finalmente, los símplexes de \mathcal{K} son los subconjuntos formados por aquellas etiquetas que se corresponden con vértices de K que están situados en un mismo símplex de K .

Recíprocamente, todo complejo simplicial abstracto finito \mathcal{K} admite una realización geométrica como complejo simplicial finito en algún \mathbb{R}^m . Para construirla, basta con embeber el conjunto de vértices de \mathcal{K} como un subconjunto de puntos afínmente independientes en \mathbb{R}^m , para una dimensión m suficientemente grande.

Cada símplex de \mathcal{K} se puede identificar entonces con el símplex geométrico en \mathbb{R}^m generado por los correspondientes vértices embebidos. La *realización geométrica* de \mathcal{K} , que denotaremos por $|\mathcal{K}|$, es la unión de todos estos símplexes geométricos. Esta realización geométrica es trivialmente el poliedro subyacente a algún complejo simplicial finito, de ahí la notación.

En caso de que el complejo simplicial abstracto \mathcal{K} no sea finito el razonamiento anterior no es válido. Pese a ello, también es posible construir la realización geométrica de \mathcal{K} . Los detalles para el caso general pueden encontrarse en [34]. Sin embargo, en lo que sigue únicamente trabajaremos con complejos simpliciales abstractos finitos, por lo que no se hace necesaria esa demostración.

Ejemplo 1.6. A continuación se muestra un ejemplo sencillo de un complejo simplicial abstracto que representa una triangulación de la banda de Möbius.

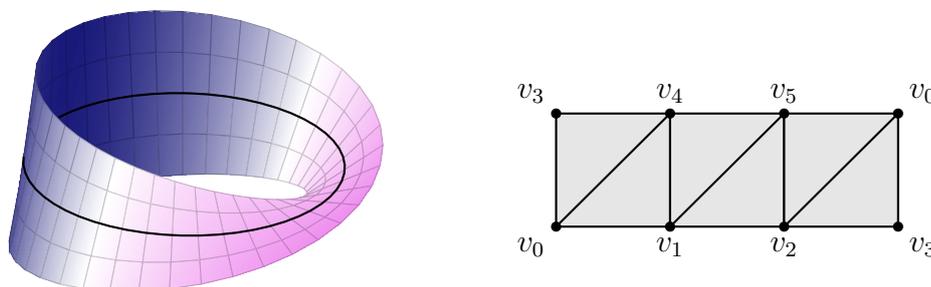


Figura 1.1: Banda de Möbius junto a una triangulación. La banda está tomada de [28].

En este ejemplo, el conjunto de vértices viene dado por

$$V = \{v_0, v_1, v_2, v_3, v_4, v_5\},$$

mientras que el complejo simplicial abstracto se puede expresar como

$$\begin{aligned} \mathcal{K} = \{ & v_0, v_1, v_2, v_3, v_4, v_5, \\ & v_0v_1, v_0v_2, v_0v_3, v_0v_4, v_0v_5, v_1v_2, v_1v_4, v_1v_5, v_2v_3, v_2v_5, v_3v_4, v_4v_5, \\ & v_0v_1v_4, v_0v_2v_3, v_0v_2v_5, v_0v_3v_4, v_1v_2v_5, v_1v_4v_5\}. \end{aligned}$$

Para expresar los elementos de \mathcal{K} se ha dado de alguna forma un «sentido de recorrido». Por ejemplo, el 2-símplice $v_0v_2v_3$ podríamos haberlo expresado como $v_0v_3v_2$, $v_2v_0v_3$, $v_2v_3v_0$ o cualquier otra permutación de los vértices. Es claro que todas sirven para expresar el mismo triángulo como figura geométrica. No obstante, la *orientación* en los símlices será importante para introducir una estructura de módulo a partir de los elementos de \mathcal{K} y poder hablar de la homología del complejo simplicial abstracto.

Este ejemplo también pone de manifiesto la facilidad con la que los ordenadores pueden manipular los complejos simpliciales abstractos: el complejo queda determinado conociendo el conjunto de sus vértices y una lista con las conexiones establecidas entre ellos.

En cuanto se define un nuevo concepto matemático dotado de una cierta estructura el siguiente paso es considerar aquellas aplicaciones entre ellos que respeten esta estructura.

Definición 1.7. Sean \mathcal{K}_1 y \mathcal{K}_2 dos complejos simpliciales abstractos y sea $\varphi: V_1 \rightarrow V_2$ una aplicación definida entre sus conjuntos de vértices. Se dice que φ es una *aplicación simplicial* si dado un símlice $\{v_0, \dots, v_q\} \in \mathcal{K}_1$ se verifica que los vértices $\varphi(v_0), \dots, \varphi(v_q)$ forman un símlice de \mathcal{K}_2 .

Cometiendo un abuso de notación, escribiremos $\varphi: \mathcal{K}_1 \rightarrow \mathcal{K}_2$ para referirnos a una aplicación simplicial entre los vértices de \mathcal{K}_1 y \mathcal{K}_2 .

También es preciso el concepto de «igualdad» o «equivalencia» de complejos simpliciales abstractos una vez tenemos definidas las aplicaciones entre ellos.

Definición 1.8. Dos complejos simpliciales abstractos \mathcal{K}_1 y \mathcal{K}_2 se dicen *isomorfos* si existe una biyección $\varphi: V_1 \rightarrow V_2$ entre los conjuntos de vértices de ambos complejos de forma que $\{v_0, \dots, v_q\} \in \mathcal{K}_1$ si, y solo si, $\{\varphi(v_0), \dots, \varphi(v_q)\} \in \mathcal{K}_2$.

En lo que sigue, a menudo obviaremos los adjetivos «abstracto» y «finito» y hablaremos simplemente de un «complejo simplicial».

1.3. Homología simplicial orientada

Antes de definir la homología simplicial, es necesario introducir una estructura algebraica en los complejos simpliciales. 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. Una combinación lineal de símplices será una cadena simplicial. Finalmente, una aplicación lineal entre los símplices de una cierta dimensión y los símplices de una dimensión una unidad menor permitirá poner en marcha toda la maquinaria del álgebra homológica.

Sea \mathcal{K} un complejo simplicial abstracto. Dado un símplice $\sigma = \{v_0, \dots, v_q\}$, con $q > 0$, se define la siguiente relación de equivalencia sobre el conjunto de las ordenaciones de los 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. Fijamos un anillo unitario y conmutativo R y consideramos las dos posibles orientaciones σ^1 y σ^2 de un q -símplice σ .

Definición 1.9. Se define el *R -módulo de q -cadenas simpliciales orientadas*, y se denota por $C_q(\mathcal{K}; R)$, como el cociente del R -módulo libre generado por todos los q -símplices σ de \mathcal{K} por el submódulo generado por los $\sigma^1 + \sigma^2$. Los elementos de $C_q(\mathcal{K}; R)$ se llaman *q -cadenas simpliciales orientadas*.

Resulta que los elementos de $C_q(\mathcal{K}; R)$ son combinaciones lineales de q -símplices orientados con coeficientes en el anillo R . Además, $C_q(\mathcal{K}; R)$ es un R -módulo libre con una base dada por cada q -símplice $\sigma \in \mathcal{K}$ tomado con una orientación, identificando la otra orientación con el elemento $-\sigma$ del R -módulo.

Por otra parte, como los 0-símplices son los vértices y estos solo tienen una orientación posible, $C_0(\mathcal{K}; R)$ es el R -módulo libre generado por los vértices de \mathcal{K} . También se tiene que estos R -módulos únicamente están definidos para $q \geq 0$, pues no existen símplices de dimensión negativa. Asimismo, se entiende que $C_q(\emptyset; R) = 0$ para todo q y que $C_q(\mathcal{K}; R) = 0$ si $q > \dim(\mathcal{K})$.

Definición 1.10. Llamamos *operador borde de orden q* al homomorfismo de R -módulos

$$\partial_q: C_q(\mathcal{K}; R) \longrightarrow C_{q-1}(\mathcal{K}; R)$$

dado por la extensión lineal de

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

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

Antes de proseguir, tenemos que comprobar que ∂_q está bien definido y que $\partial_q(-\sigma) = -\partial_q(\sigma)$ para un q -símplice σ . Para ello, basta con ver que el lado derecho de la igualdad en la definición anterior cambia de signo al permutar dos vértices adyacentes en $[v_0, \dots, v_q]$. Comparemos, pues, las expresiones de

$$\partial_q([v_0, \dots, v_j, v_{j+1}, \dots, v_q]) \quad \text{y} \quad \partial_q([v_0, \dots, v_{j+1}, v_j, \dots, v_q]).$$

Si el índice de la suma $i \notin \{j, j+1\}$, entonces el resto de sumandos de ambas expresiones difieren precisamente en el signo: los términos son iguales, salvo la permutación de los vértices v_j y v_{j+1} que provoca el cambio de signo.

Veamos ahora los sumandos cuando el índice de la suma es $i = j$ e $i = j+1$. En la primera expresión, se tiene

$$(-1)^j [v_0, \dots, v_{j-1}, v_{j+1}, v_{j+2}, \dots, v_q] + (-1)^{j+1} [v_0, \dots, v_{j-1}, v_j, v_{j+2}, \dots, v_q].$$

En la segunda expresión, se tiene

$$(-1)^j [v_0, \dots, v_{j-1}, v_j, v_{j+2}, \dots, v_q] + (-1)^{j+1} [v_0, \dots, v_{j-1}, v_{j+1}, v_{j+2}, \dots, v_q].$$

Efectivamente, estas dos expresiones difieren en un signo.

Ejemplo 1.11. Sea \mathcal{K} un complejo simplicial de dimensión 2. Supongamos que $C_0(\mathcal{K}; \mathbb{Z})$ está generado por el conjunto de vértices $\{v_0, v_1, v_2, v_3\}$, que $C_1(\mathcal{K}; \mathbb{Z})$ tiene por base a $\{[v_0, v_1], [v_1, v_2], [v_0, v_2], [v_2, v_3]\}$ y que la base de $C_2(\mathcal{K}; \mathbb{Z})$ es $\{[v_0, v_1, v_2]\}$.

El operador borde $\partial_2: C_2(\mathcal{K}; \mathbb{Z}) \rightarrow C_1(\mathcal{K}; \mathbb{Z})$ actúa así:

$$\partial_2([v_0, v_1, v_2]) = [v_1, v_2] - [v_0, v_2] + [v_0, v_1].$$

Por su parte, el operador borde de orden 1 hace lo siguiente sobre una combinación lineal de elementos de la base de $C_1(\mathcal{K}; \mathbb{Z})$ como, por ejemplo, la siguiente:

$$\partial_1(2[v_2, v_3] + 5[v_0, v_1]) = 2(v_3 - v_2) + 5(v_1 - v_0).$$

Finalmente, se tiene que $\partial_0: C_0(\mathcal{K}; \mathbb{Z}) \rightarrow C_{-1}(\mathcal{K}; \mathbb{Z})$ es necesariamente el homomorfismo nulo ya que $C_{-1}(\mathcal{K}; \mathbb{Z}) = 0$.

Un cálculo rutinario sirve para comprobar el siguiente resultado.

Proposición 1.12. *El operador borde verifica que $\partial_q \circ \partial_{q+1} = 0$ para todo q .*

Demostración. En efecto, al hacer el cálculo

$$\begin{aligned} (\partial_q \circ \partial_{q+1})[v_0, \dots, v_{q+1}] &= \partial_q \left(\sum_{i=0}^{q+1} (-1)^i [v_0, \dots, \hat{v}_i, \dots, v_{q+1}] \right) \\ &= \sum_{i=0}^{q+1} (-1)^i \left(\sum_{j<i} (-1)^j [v_0, \dots, \hat{v}_j, \dots, \hat{v}_i, \dots, v_{q+1}] \right. \\ &\quad \left. + \sum_{j>i} (-1)^{j-1} [v_0, \dots, \hat{v}_i, \dots, \hat{v}_j, \dots, v_{q+1}] \right) \end{aligned}$$

los términos de ambas sumas se cancelan por parejas. \square

La proposición anterior es crucial porque convierte a $(C_\bullet(\mathcal{K}; R), \partial_\bullet)$ en un complejo de cadenas. La propiedad $\partial_q \circ \partial_{q+1} = 0$ implica que $\text{Im}(\partial_{q+1}) \subset \ker(\partial_q)$ para todo q . La imagen y el núcleo del operador borde definen dos submódulos de $C_q(\mathcal{K}; R)$ que reciben nombre propio debido a su importancia.

Definición 1.13. Llamamos q -ésimo R -módulo de ciclos a $Z_q(\mathcal{K}; R) = \ker(\partial_q)$. Llamamos q -ésimo R -módulo de bordes a $B_q(\mathcal{K}; R) = \text{Im}(\partial_{q+1})$.

Definición 1.14. Se define el q -ésimo R -módulo de homología simplicial de \mathcal{K} como el siguiente cociente:

$$H_q(\mathcal{K}; R) = \frac{Z_q(\mathcal{K}; R)}{B_q(\mathcal{K}; R)}.$$

Se define también el q -ésimo número de Betti de \mathcal{K} , denotado por β_q , como el rango de $H_q(\mathcal{K}; R)$.

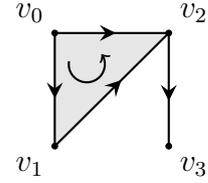


Figura 1.2: Realización geométrica de \mathcal{K} .

Proposición 1.15. Sean \mathcal{K}_1 y \mathcal{K}_2 dos complejos simpliciales y $\varphi: \mathcal{K}_1 \rightarrow \mathcal{K}_2$ una aplicación simplicial. Entonces, para todo $q \geq 0$ la aplicación φ induce un homomorfismo de complejos de cadenas

$$C_q(\varphi): C_q(\mathcal{K}_1; R) \longrightarrow C_q(\mathcal{K}_2; R)$$

definido por la extensión lineal de

$$C_q(\varphi)([v_0, \dots, v_q]) = \begin{cases} [\varphi(v_0), \dots, \varphi(v_q)] & \text{si no aparecen vértices repetidos,} \\ 0 & \text{en otro caso.} \end{cases}$$

En particular, $C_\bullet(\text{id}) = \text{id}$ y si $\psi: \mathcal{K}_2 \rightarrow \mathcal{K}_3$ es otra aplicación simplicial se tiene que $C_\bullet(\psi \circ \varphi) = C_\bullet(\psi) \circ C_\bullet(\varphi)$.

Demostración. Es obvio que la definición de $C_\bullet(\varphi)$ no depende de la permutación que da la orientación. Para comprobar que define un homomorfismo de complejos de cadenas, tenemos que demostrar que $\partial_q^2 \circ C_q(\varphi) = C_{q-1}(\varphi) \circ \partial_q^1$ para $q > 0$ (el caso $q = 0$ es inmediato). Basta comprobarlo en los símlices orientados, ya que estos forman una base. Sea $\sigma = [v_0, \dots, v_q] \in C_q(\mathcal{K}_1; R)$.

Si al aplicar φ no se repite ningún vértice, entonces

$$\begin{aligned} (C_{q-1}(\varphi) \circ \partial_q^1)([v_0, \dots, v_q]) &= C_{q-1}(\varphi) \left(\sum_{i=0}^q (-1)^i [v_0, \dots, \hat{v}_i, \dots, v_q] \right) \\ &= \sum_{i=0}^q (-1)^i [\varphi(v_0), \dots, \widehat{\varphi(v_i)}, \dots, \varphi(v_q)] \\ &= (\partial_q^2 \circ C_q(\varphi))([v_0, \dots, v_q]). \end{aligned}$$

Si al aplicar φ se repite algún vértice, pongamos $\varphi(v_j) = \varphi(v_k)$, por una parte se tiene que $(\partial_q^2 \circ C_q(\varphi))([v_0, \dots, v_q]) = 0$ y, por otra parte,

$$\begin{aligned} (C_{q-1}(\varphi) \circ \partial_q^1)([v_0, \dots, v_q]) &= C_{q-1}(\varphi) \left(\sum_{i=0}^q (-1)^i [v_0, \dots, \hat{v}_i, \dots, v_q] \right) \\ &= \sum_{i=0}^q (-1)^i C_{q-1}(\varphi)([v_0, \dots, \hat{v}_i, \dots, v_q]) \\ &= 0 \end{aligned}$$

pues si $i \notin \{j, k\}$ hay un vértice repetido y $C_{q-1}(\varphi)([v_0, \dots, \hat{v}_i, \dots, v_q]) = 0$ y, suponiendo que $j < k$, se verifica que

$$\begin{aligned} &(-1)^j [\varphi(v_0), \dots, \widehat{\varphi(v_j)}, \dots, \varphi(v_k), \dots, \varphi(v_q)] \\ &\quad + (-1)^k [\varphi(v_0), \dots, \varphi(v_j), \dots, \widehat{\varphi(v_k)}, \dots, \varphi(v_q)] \\ &= ((-1)^j (-1)^{k-j-1} + (-1)^k) [\varphi(v_0), \dots, \widehat{\varphi(v_j)}, \dots, \varphi(v_k), \dots, \varphi(v_q)] \\ &= ((-1)^{k-1} + (-1)^k) [\varphi(v_0), \dots, \widehat{\varphi(v_j)}, \dots, \varphi(v_k), \dots, \varphi(v_q)] \\ &= 0 \end{aligned}$$

ya que son necesarias $k - j - 1$ trasposiciones para pasar de un símplice al otro (contando desde 0, el vértice $\varphi(v_k)$ ocupa la posición $k - 1$ en el primer símplice).

Finalmente, las igualdades $C_\bullet(\text{id}) = \text{id}$ y $C_\bullet(\psi \circ \varphi) = C_\bullet(\psi) \circ C_\bullet(\varphi)$ resultan inmediatas a partir de la definición del homomorfismo en cada grado. \square

Una vez probada la existencia del homomorfismo anterior, es un resultado general de homología que un homomorfismo de complejos de cadenas induce un homomorfismo en homología.

Corolario 1.16. *Toda aplicación simplicial $\varphi: \mathcal{K}_1 \rightarrow \mathcal{K}_2$ induce un homomorfismo*

$$H_q(\varphi): H_q(\mathcal{K}_1; R) \longrightarrow H_q(\mathcal{K}_2; R).$$

Además, $H_\bullet(\text{id}) = \text{id}$ y $H_\bullet(\psi \circ \varphi) = H_\bullet(\psi) \circ H_\bullet(\varphi)$ para $\psi: \mathcal{K}_2 \rightarrow \mathcal{K}_3$ simplicial.

Este homomorfismo viene dado de la siguiente forma: si $[z] \in H_q(\mathcal{K}_1; R)$ es la clase de un q -ciclo, entonces $H_q(\varphi)([z]) = [C_q(\varphi)(z)]$.

1.4. Invariancia homotópica

En la sección anterior hemos descrito un proceso para asociar un R -módulo graduado a un complejo simplicial \mathcal{K} . La presente sección pretende recoger un resultado de gran importancia que afirma que estos módulos de homología solo dependen del tipo de homotopía del poliedro subyacente $|\mathcal{K}|$. Como los objetivos de este trabajo son otros, no se dará una demostración de este hecho. La prueba puede encontrarse en [2] o en [34], que dedican un capítulo entero cada uno a tratar este asunto.

Empezamos dando una topología al poliedro $|\mathcal{K}|$. Para ello, en un primer momento parece natural plantearse dar la topología relativa de \mathbb{R}^m , ya que $|\mathcal{K}| \subset \mathbb{R}^m$. También existe la posibilidad de dar al poliedro la topología débil de los símplices: la mayor topología que hace que las inclusiones $\sigma \hookrightarrow |\mathcal{K}|$ sean continuas. Sin más que tener en cuenta que los símplices son subespacios compactos de \mathbb{R}^m , es inmediata la siguiente afirmación.

Proposición 1.17. *La topología relativa y la topología débil de $|\mathcal{K}| \subset \mathbb{R}^m$ coinciden. Aún más, esta topología hace a $|\mathcal{K}|$ compacto.*

La proposición anterior no es cierta si \mathcal{K} no es finito. En general, la topología débil de $|\mathcal{K}|$ es más fina que la topología relativa que hereda de \mathbb{R}^m .

De esta manera es también inmediato que $A \subset |\mathcal{K}|$ es cerrado (abierto) si, y solo si, $A \cap \sigma$ es cerrado (abierto) en σ para todo $\sigma \in \mathcal{K}$.

Una vez introducida una topología en $|\mathcal{K}|$ tiene sentido la siguiente definición.

Definición 1.18. Sea X un espacio topológico. Se dice que X es *triangulable* si existen un poliedro $|\mathcal{K}|$ y un homeomorfismo $h: |\mathcal{K}| \rightarrow X$. En ese caso, se dice que el par (\mathcal{K}, h) es una *estructura simplicial* o *triangulación* de X .

Estamos ya en condiciones de enunciar el principal resultado que recogerá esta sección.

Teorema 1.19. *Si \mathcal{K}_1 y \mathcal{K}_2 son dos complejos simpliciales tales que sus poliedros subyacentes $|\mathcal{K}_1|$ y $|\mathcal{K}_2|$ tienen el mismo tipo de homotopía, entonces los módulos de homología $H_q(\mathcal{K}_1; R)$ y $H_q(\mathcal{K}_2; R)$ son isomorfos para todo q .*

En particular, los módulos de homología son invariantes topológicos para cualquier poliedro $|\mathcal{K}|$. Esto permite dar la siguiente definición.

Definición 1.20. Sea X un espacio topológico triangulable. Se define el q -ésimo R -módulo de homología de X como $H_q(X; R) := H_q(\mathcal{K}; R)$, donde (\mathcal{K}, h) es una triangulación de X .

Para un espacio triangulable puede probarse que la homología simplicial coincide con su homología singular [34, Capítulo 4, §34], por lo que se trata de una buena definición. Obviamente, la definición anterior no depende de la triangulación escogida ya que los poliedros que dan las dos triangulaciones son homeomorfos por definición (y, por tanto, del mismo tipo de homotopía).

1.5. Computabilidad

Como ya se ha comentado previamente, la homología simplicial es extremadamente útil debido a su computabilidad. Ello permite calcular toda la información relevante en tiempos de ejecución razonables. La clave del proceso reside en las matrices asociadas a los operadores borde ∂_q , que posibilitan el desarrollo de un algoritmo para calcular los módulos de homología en cualquier grado y los respectivos números de Betti.

1.5.1. Forma normal de Smith

El algoritmo parte del teorema de diagonalización de matrices con coeficientes sobre un dominio de ideales principales.

Teorema 1.21 (de diagonalización). *Sea R un dominio de ideales principales. Supongamos que $A \in \mathcal{M}_{m \times n}(R)$. Entonces, A es equivalente a una matriz diagonal*

$$\text{diag}(a_1, \dots, a_k, 0, \dots, 0) = \left[\begin{array}{ccc|c} a_1 & & & 0 \\ & \ddots & & \\ & & a_k & \\ \hline & & & 0 \end{array} \right],$$

donde $a_i \neq 0$ para $1 \leq i \leq k$ y se verifica que $a_1 | \dots | a_k$, esto es, a_i divide a a_{i+1} para $i \in \{1, \dots, k-1\}$.

La matriz diagonal resultante de aplicar el teorema de diagonalización se denomina *forma normal de Smith* y se consigue mediante sucesivas transformaciones por filas y

columnas de la matriz original. La demostración que proporciona el algoritmo de reducción tanto para dominios euclídeos como para el caso general de dominios de ideales principales puede encontrarse en [25, Capítulo 7].

El segundo enunciado crucial para nuestros propósitos es el teorema de estructura para módulos finitamente generados sobre dominios de ideales principales, un resultado muy importante y ampliamente conocido.

Teorema 1.22 (de estructura). *Sea R un dominio de ideales principales. Supongamos que M es un R -módulo finitamente generado. Entonces*

$$M \cong \left(R \oplus \cdots \oplus R \right) \oplus \left(\frac{R}{(a_1)} \oplus \cdots \oplus \frac{R}{(a_k)} \right),$$

donde los r primeros sumandos se corresponden con la parte libre de M , denotada por L_M , y los k sumandos restantes con la parte de torsión de M , denotada por T_M . Además, se tiene que los elementos $a_1, \dots, a_k \in R$, denominados coeficientes o elementos de torsión, son no nulos, únicos y verifican que $a_1 | \cdots | a_k$.

Los detalles de la demostración se pueden consultar en [32, Capítulo III, §7]. En líneas generales, la demostración interpretada en términos matriciales sigue los pasos descritos a continuación. Por ser M finitamente generado podemos tomar una familia finita de generadores $\{x_1, \dots, x_n\}$ de M junto con una presentación $\varepsilon: F \rightarrow M$, donde F es un R -módulo libre de rango n . Al aplicar el primer teorema de isomorfía a la aplicación sobreyectiva que da la presentación obtenemos que

$$\frac{F}{\ker(\varepsilon)} \cong M.$$

Si suponemos que una base de F es $\{e_1, \dots, e_n\}$, entonces se verifica que $\varepsilon(e_i) = x_i$, con $1 \leq i \leq n$. Como $\ker(\varepsilon)$ es un submódulo de F , se tiene que $\ker(\varepsilon)$ es libre de rango menor o igual que n . De esta forma, la sucesión exacta corta $\ker(\varepsilon) \rightarrow F \rightarrow M$ permite escribir una familia finita de generadores $\{f_1, \dots, f_m\}$ de $\ker(\varepsilon)$ en función de los generadores de F como

$$\begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix},$$

donde la matriz $A := (a_{ij}) \in \mathcal{M}_{m \times n}(R)$ tiene los coeficientes en el dominio de ideales principales R . El teorema de diagonalización nos permite reducir la matriz A a su forma normal de Smith, pongamos $\text{diag}(a_1, \dots, a_k, 0, \dots, 0)$. Esta matriz es equivalente a A , por lo que podemos escribir

$$X^{-1}AY = \left[\begin{array}{ccc|c} a_1 & & & 0 \\ & \ddots & & \\ & & a_k & \\ \hline & & & 0 \end{array} \right],$$

siendo $X \in \mathcal{M}_{m \times m}(R)$ e $Y \in \mathcal{M}_{n \times n}(R)$ matrices inversibles. Las matrices X e Y determinan nuevas bases $\{f'_1, \dots, f'_m\}$ y $\{e'_1, \dots, e'_n\}$ de $\ker(\varepsilon)$ y F , respectivamente.

Ahora bien, como $\text{diag}(a_1, \dots, a_k, 0, \dots, 0)$ es la matriz de $\{f'_1, \dots, f'_m\}$ con respecto a $\{e'_1, \dots, e'_n\}$, podemos expresar la nueva base de $\ker(\varepsilon)$ de la siguiente forma:

$$f'_i = a_i e'_i, \quad 1 \leq i \leq k.$$

Por tanto, la forma normal de Smith de A proporciona la descomposición de M dada por el teorema de estructura. Las k entradas no nulas de la diagonal son los coeficientes de torsión de M . Por su parte, las $r = n - k$ columnas de ceros indican el número de copias del anillo R que conforman la parte libre de M .

1.5.2. Algoritmo para las matrices de incidencia

En lo que sigue supondremos que R es un dominio de ideales principales, aunque no se haga mención explícita. Sean M y M' dos R -módulos libres y finitamente generados. Fijemos bases $\{e_1, \dots, e_n\}$ y $\{e'_1, \dots, e'_m\}$ para cada uno, respectivamente. Si $f: M \rightarrow M'$ es un homomorfismo de R -módulos, entonces podemos escribir

$$f(e_j) = \sum_{i=1}^m a_{ij} e'_i, \quad 1 \leq j \leq n,$$

para ciertos $a_{ij} \in R$. La matriz (a_{ij}) se llama *matriz asociada a f* con respecto a las bases de M y M' . El teorema de diagonalización nos permite reducir esta matriz a su forma normal de Smith. A continuación, vamos a plantear esta misma situación en el contexto homológico en el que estamos trabajando.

Dado un complejo simplicial \mathcal{K} , para cada grado q tenemos definido el operador borde $\partial_q: C_q(\mathcal{K}; R) \rightarrow C_{q-1}(\mathcal{K}; R)$, que es un homomorfismo de R -módulos. Fijamos bases en $C_q(\mathcal{K}; R)$ y $C_{q-1}(\mathcal{K}; R)$ orientando los símlices y denotamos por $M(\partial_q)$ a la matriz asociada al operador borde ∂_q . Las entradas de estas matrices tienen sus coeficientes en el conjunto $\{-1, 0, 1\} \subset R$, por lo que se pueden ver como matrices de incidencia.

Las formas normales de Smith de las matrices $M(\partial_q)$ no solo son importantes para poder simplificarlas al máximo, sino que probaremos que contienen toda la información necesaria para determinar los módulos de homología de \mathcal{K} .

Antes de demostrar la validez de esta afirmación introduciremos nueva terminología y probaremos unos resultados previos. La estructura simplicial no interviene en ningún momento. Sean $Z_q = \ker(\partial_q)$ y $B_q = \text{Im}(\partial_{q+1})$ los submódulos de ciclos y bordes, respectivamente.

Definición 1.23. Llamamos *q -ésimo R -módulo de bordes débiles a*

$$W_q = \{c_q \in C_q: \exists \lambda \in R, \lambda \neq 0, \text{ con } \lambda c_q \in B_q\}.$$

Los bordes débiles constituyen un R -módulo intermedio entre el de bordes y el de ciclos.

Lema 1.24. *En las condiciones anteriores, los R -módulos de bordes, bordes débiles y ciclos verifican la siguiente cadena de inclusiones para todo q :*

$$B_q \subset W_q \subset Z_q \subset C_q.$$

Demostración. La primera inclusión es obvia, pues basta tomar $\lambda = 1 \in R$. La tercera inclusión es cierta por construcción.

La segunda inclusión se deduce de que C_q es libre y, por tanto, libre de torsión. La igualdad $\lambda c_q = \partial_{q+1}(d_{q+1})$, con $\lambda \neq 0$, implica que

$$\partial_q(\lambda c_q) = \lambda \partial_q(c_q) = 0,$$

de donde se sigue inmediatamente la relación de contenido. \square

El siguiente teorema proporciona un resultado análogo al de bases para espacios vectoriales, pero adaptado al contexto de R -módulos y complejos de cadenas. Se trata del paso previo al cálculo efectivo de los módulos de homología de cualquier complejo simplicial.

Teorema 1.25. *Sean R un dominio de ideales principales y $(C_\bullet, \partial_\bullet)$ un complejo de cadenas finito. Supongamos que cada R -módulo C_q es libre y finitamente generado. Entonces, para cada q existen submódulos $U_q, V_q \subset C_q$ tales que*

$$Z_q = V_q \oplus W_q \quad y \quad C_q = U_q \oplus V_q \oplus W_q,$$

donde Z_q es el R -módulo de ciclos y W_q es el R -módulo de bordes débiles. Además, se tiene que

$$L_{H_q} \cong \frac{Z_q}{W_q} \quad y \quad T_{H_q} \cong \frac{W_q}{B_q},$$

siendo B_q es el R -módulo de bordes.

Demostración. Sean B_q, W_q y Z_q los submódulos de bordes, bordes débiles y ciclos ya conocidos. Sean también L_{H_q} y T_{H_q} la parte libre y la parte de torsión de H_q , respectivamente.

En primer lugar, veremos la segunda parte del enunciado. Tenemos que probar que $Z_q = V_q \oplus W_q$. Veamos entonces que el módulo de bordes débiles W_q es un sumando directo de Z_q . Para ello, consideramos la proyecciones naturales sobre los cocientes

$$\pi_1: Z_q \longrightarrow H_q, \quad \pi_2: H_q \longrightarrow \frac{H_q}{T_{H_q}}$$

y el isomorfismo que nos da el teorema de estructura

$$h: \frac{H_q}{T_{H_q}} \xrightarrow{\cong} L_{H_q}.$$

A continuación, probaremos que el núcleo de la aplicación $g = h \circ \pi_2 \circ \pi_1$ es exactamente W_q . En efecto, si $z \in Z_q$ es un elemento del núcleo de g , es decir, es tal que $g(z) =$

$h([z] + T_{H_q}) = 0$, entonces $[z]$ es un elemento de torsión de H_q . Luego existe un $\lambda \in R$, $\lambda \neq 0$, de manera que $\lambda[z] = [\lambda z] = 0$. Así, $\lambda z \in B_q$ y, por tanto, $z \in W_q$ por definición. Recíprocamente, si $z \in W_q$ y $\lambda z \in B_q$ para un $\lambda \in R$, $\lambda \neq 0$, se tiene que $\pi_1(\lambda z) = [\lambda z] = 0$ y, en consecuencia,

$$\lambda g(z) = g(\lambda z) = h([\lambda z] + T_{H_q}) = h(0 + T_{H_q}) = 0 \in L_{H_q},$$

lo que implica que $g(z) = 0$ al no haber torsión en L_{H_q} y ser $\lambda \neq 0$.

De esta forma, el primer teorema de isomorfía permite escribir

$$\frac{Z_q}{W_q} = \frac{Z_q}{\ker g} \cong \text{Im } g = L_{H_q},$$

ya que g es sobreyectiva al serlo también las proyecciones sobre los cocientes π_1 y π_2 . En particular, esto nos dice que $\frac{Z_q}{W_q}$ es un módulo libre.

Sea $\{c_1 + W_q, \dots, c_k + W_q\}$ una base de $\frac{Z_q}{W_q}$ y sea $\{d_1, \dots, d_t\}$ una base de W_q (que es libre al ser, por el Lema 1.24, un submódulo de un módulo libre). Se comprueba que el conjunto $\{c_1, \dots, c_k, d_1, \dots, d_t\}$ es una base de Z_q . Una vez probado esto, si llamamos V_q al submódulo libre generado por los elementos $\{c_1, \dots, c_k\}$ habremos conseguido la descomposición en suma directa buscada para los ciclos:

$$Z_q = V_q \oplus W_q.$$

Veamos entonces que $\{c_1, \dots, c_k, d_1, \dots, d_t\}$ es una base de Z_q . Probamos primero que es un sistema de generadores. Sea $z \in Z_q$. Si consideramos z en el cociente $\frac{Z_q}{W_q}$, existen $\mu_1, \dots, \mu_k \in R$ de forma que

$$z + W_q = \sum_{i=1}^k \mu_i c_i + W_q.$$

Luego $z - \sum_{i=1}^k \mu_i c_i \in W_q$. Esto permite escribir

$$z - \sum_{i=1}^k \mu_i c_i = \sum_{j=1}^t \nu_j d_j$$

para ciertos $\nu_1, \dots, \nu_t \in R$. Por tanto, $z = \sum_{i=1}^k \mu_i c_i + \sum_{j=1}^t \nu_j d_j$ y, así, concluimos que $\{c_1, \dots, c_k, d_1, \dots, d_t\}$ genera Z_q .

Para ver que los elementos son linealmente independientes, supongamos que existen $\mu_1, \dots, \mu_k, \nu_1, \dots, \nu_t \in R$ tales que

$$0 = \sum_{i=1}^k \mu_i c_i + \sum_{j=1}^t \nu_j d_j.$$

Entonces, en el cociente $\frac{Z_q}{W_q}$ se tiene que $0 = \sum_{i=1}^k \mu_i c_i$; de aquí deducimos que $c_i = 0$ para $i = 1, \dots, k$ y, finalmente, que $d_j = 0$ para $j = 1, \dots, t$.

Para acabar la demostración del teorema tenemos que probar la existencia de la descomposición en suma directa del módulo C_q que se cita en el enunciado. Sabemos que B_{q-1} es un módulo libre (Lema 1.24). Entonces, la siguiente sucesión exacta corta escinde:

$$0 \longrightarrow Z_q \hookrightarrow C_q \xrightarrow{\partial_q} B_{q-1} \longrightarrow 0.$$

ρ_q

Esto permite deducir que $C_q = Z_q \oplus \text{Im}(\rho_q)$.

Basta tomar $U_q := \text{Im}(\rho_q)$ para obtener la descomposición deseada

$$C_q = U_q \oplus Z_q = U_q \oplus V_q \oplus W_q,$$

que es cierta para todo grado q .

Finalmente, se tiene que

$$H_q = \frac{Z_q}{B_q} \cong \frac{V_q \oplus W_q}{B_q} = V_q \oplus \frac{W_q}{B_q} \cong \frac{Z_q}{W_q} \oplus \frac{W_q}{B_q}.$$

□

Observación 1.26. Nótese que los submódulos W_q y $Z_q = V_q \oplus W_q$ están unívocamente determinados. Sin embargo, los submódulos U_q y V_q no lo están.

Estamos ya en condiciones de desarrollar el algoritmo anunciado para calcular la homología de manera efectiva. El cálculo se reduce a computar $\frac{Z_q}{W_q}$ y $\frac{W_q}{B_q}$.

Orientemos los símlices del complejo simplicial \mathcal{K} . Una vez orientados, elijamos una base $\{e_1, \dots, e_n\}$ de C_q y una base $\{e'_1, \dots, e'_m\}$ de C_{q-1} respecto de las cuales la matriz $M(\partial_q)$ asociada al homomorfismo $\partial_q: C_q \rightarrow C_{q-1}$ se exprese en su forma normal de Smith

$$\begin{array}{cccc} & e_1 & \cdots & e_k & e_{k+1} & \cdots & e_n \\ \begin{array}{c} e'_1 \\ \vdots \\ e'_k \\ e'_{k+1} \\ \vdots \\ e'_m \end{array} & \left[\begin{array}{ccc|ccc} a_1 & & & & & \\ & \ddots & & & & 0 \\ & & a_k & & & \\ \hline & & & & & \\ & & & 0 & & \\ & & & & & 0 \end{array} \right], \end{array}$$

donde los coeficientes de la diagonal cumplen que $a_1 | \cdots | a_k$.

Lema 1.27. *Con las notaciones anteriores, se verifica que:*

- (1) $\{e_{k+1}, \dots, e_n\}$ es una base de Z_q .

(2) $\{e'_1, \dots, e'_k\}$ es una base de W_{q-1} .

(3) $\{a_1e'_1, \dots, a_ke'_k\}$ es una base de B_{q-1} .

Demostración. Sea $c_q \in C_q$ una q -cadena arbitraria, que escribimos como

$$c_q = \sum_{i=1}^n \gamma_i e_i.$$

Gracias a la matriz en forma normal podemos expresar su borde como

$$\partial_q(c_q) = \sum_{i=1}^k \gamma_i a_i e'_i.$$

Para demostrar (1), dado que $a_i \neq 0$, la q -cadena c_q es un ciclo si, y solo si, $\gamma_i = 0$ para $i = 1, \dots, k$.

Como consecuencia de lo anterior, para probar la afirmación (3) basta notar que todo $(q-1)$ -borde $\partial_q(c_q)$ se encuentra en el módulo generado por el conjunto $\{a_1e'_1, \dots, a_ke'_k\}$. Como $a_i \neq 0$, los elementos del conjunto anterior son independientes.

Por último, veamos (2). Dado que $a_i e'_i = \partial_q(e_i)$, $i = 1, \dots, k$, se tiene que $e'_1, \dots, e'_k \in W_{q-1}$. Recíprocamente, sea

$$c_{q-1} = \sum_{i=1}^m \eta_i e'_i$$

una $(q-1)$ -cadena y supongamos que $c_{q-1} \in W_{q-1}$. Entonces, c_{q-1} satisface una ecuación de la forma

$$\lambda c_{q-1} = \partial_q(c_p) = \sum_{i=1}^k \gamma_i a_i e'_i$$

para algún $\lambda \in R$, con $\lambda \neq 0$. Igualando los coeficientes de esta expresión junto con los correspondientes del c_{q-1} arbitrario (multiplicado por λ) deducimos que $\lambda \eta_i = 0$ para $i > k$. Luego, $\eta_i = 0$ para $i > k$. Por tanto, $\{e'_1, \dots, e'_k\}$ es una base de W_{q-1} . \square

El Lema 1.27 es la pieza final para conseguir toda la información buscada. A continuación se recogen explícitamente sus consecuencias.

Corolario 1.28. *Con las notaciones anteriores, se verifican los siguientes hechos referentes a la forma normal de Smith de $M(\partial_q)$:*

(1') *El rango de Z_q son las columnas de ceros.*

(2') *El rango de W_{q-1} son las filas no nulas.*

(3') *Existe un isomorfismo que permite expresar la parte de torsión de H_{q-1} :*

$$\frac{W_{q-1}}{B_{q-1}} \cong \frac{R}{(a_1)} \oplus \dots \oplus \frac{R}{(a_k)}.$$

Demostración. Los puntos (1') y (2') son consecuencias inmediatas de los puntos (1) y (2) del Lema 1.27, respectivamente.

Por su parte, el isomorfismo mencionado en el punto (3') se deduce de los puntos (2) y (3) del Lema 1.27. \square

De aquí se sigue que la forma normal de $M(\partial_q)$ proporciona los coeficientes de torsión del módulo de homología de \mathcal{K} en grado $q-1$: son las entradas no nulas y distintas de 1 de la forma normal. Esta forma normal también permite calcular el rango de Z_q . Análogamente, de $M(\partial_{q+1})$ se obtiene el rango de W_q . Esto permite calcular el q -ésimo número de Betti, pues

$$\beta_q = \text{rango}(Z_q) - \text{rango}(W_q).$$

Por tanto, el q -ésimo módulo de homología H_q queda completamente determinado por las matrices $M(\partial_q)$ y $M(\partial_{q+1})$, ya que de ellas se extraen β_q y los coeficientes de torsión, que conforman un sistema completo de invariantes.

En consecuencia, hemos probado el siguiente teorema.

Teorema 1.29. *Los módulos de homología de un complejo simplicial \mathcal{K} sobre un dominio de ideales principales R son algorítmicamente computables.*

La potencia de este resultado se comprueba fácilmente mediante un ejemplo sencillo. En particular, se pone de manifiesto que las únicas herramientas necesarias pertenecen al álgebra lineal.

Ejemplo 1.30. Consideremos el complejo simplicial \mathcal{K} representado en la Figura 1.3. Tenemos que \mathcal{K} es de dimensión 2. Esto nos dice directamente que $H_q(\mathcal{K}; R) = 0$ para $q \geq 3$. Resta por calcular la homología en grados 0, 1 y 2.

En primer lugar, escogemos bases para $C_0(\mathcal{K}; R)$, $C_1(\mathcal{K}; R)$ y $C_2(\mathcal{K}; R)$ orientando los simplices de todas las dimensiones.

El complejo simplicial \mathcal{K} únicamente cuenta con el 2-símplice orientado $\sigma = [v_0, v_1, v_2] \in C_2(\mathcal{K}; R)$. Por su parte, una base para $C_1(\mathcal{K}; R)$ está formada por los 6 simplices orientados denotados por a, b, \dots, f . Finalmente, los 5 vértices v_0, v_1, \dots, v_4 generan $C_0(\mathcal{K}; R)$.

Trabajaremos con coeficientes en el anillo $R = \mathbb{Z}$. Procedemos a expresar en forma matricial el homomorfismo $\partial_2: C_2(\mathcal{K}; \mathbb{Z}) \rightarrow C_1(\mathcal{K}; \mathbb{Z})$. Para ello, calculamos

$$\begin{aligned} \partial_2(\sigma) &= [v_1, v_2] - [v_0, v_2] + [v_0, v_1] \\ &= -b - c + a. \end{aligned}$$

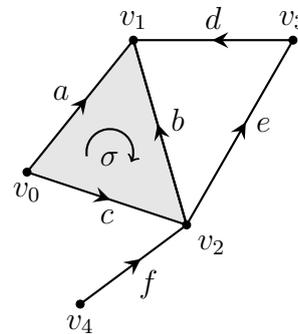


Figura 1.3: El complejo simplicial \mathcal{K} .

Como $\{\sigma\}$ es una base de $C_2(\mathcal{K}; \mathbb{Z})$ y $\{a, b, c, d, e, f\}$ es una base de $C_1(\mathcal{K}; \mathbb{Z})$, representamos matricialmente el homomorfismo como la matriz de la izquierda, mientras que a la derecha está su correspondiente forma normal:

$$\begin{array}{c} \sigma \\ a \\ b \\ c \\ d \\ e \\ f \end{array} \begin{bmatrix} 1 \\ -1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{array}{c} \sigma \\ a - b - c \\ b \\ c \\ d \\ e \\ f \end{array} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

De aquí deducimos que $\text{rango}(Z_2) = 0$ y $\text{rango}(W_1) = 1$. Nótese que el algoritmo utilizado para calcular la forma normal de Smith también permite obtener las nuevas bases, incluidas alrededor de la matriz.

En el caso de $M(\partial_1)$, esta matriz viene dada por

$$\begin{array}{c} a \quad b \quad c \quad d \quad e \quad f \\ v_0 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{array} \begin{bmatrix} -1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix},$$

que en forma normal se expresa como

$$\begin{array}{c} a \quad b \quad f \quad z_1 \quad z_2 \quad z_3 \\ v_1 - v_0 \\ v_1 - v_2 \\ v_2 - v_4 \\ v_4 - v_3 \\ v_4 \end{array} \left[\begin{array}{cccc|cc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right],$$

donde $z_1 = d - b - f$, $z_2 = e + d - b$ y $z_3 = c + b - a$.

Así, $\text{rango}(Z_1) = 2$ y $\text{rango}(W_0) = 4$. Finalmente, $\partial_0 = 0$ ya que $C_{-1}(\mathcal{K}; \mathbb{Z}) = 0$. Esto nos dice que $\text{rango}(Z_0) = \text{rango}(C_0(\mathcal{K}; \mathbb{Z})) = 5$. Teniendo en cuenta que $W_2 = 0$ al tenerse $C_3(\mathcal{K}; \mathbb{Z}) = 0$, la fórmula para calcular los números de Betti vista anteriormente arroja los siguientes resultados:

$$\beta_2 = 0 - 0 = 0, \quad \beta_1 = 2 - 1 = 1, \quad \beta_0 = 5 - 4 = 1.$$

Y puesto que todas las formas normales tienen unos en las diagonales, esto provoca que no haya coeficientes de torsión y, en consecuencia, que los módulos de homología no tengan parte de torsión. Por tanto:

$$H_2(\mathcal{K}; \mathbb{Z}) = 0, \quad H_1(\mathcal{K}; \mathbb{Z}) = \mathbb{Z}, \quad H_0(\mathcal{K}; \mathbb{Z}) = \mathbb{Z}.$$

En particular, los módulos de homología de grado 2 con coeficientes en \mathbb{Z} y en \mathbb{Z}_2 ponen de manifiesto que el plano proyectivo real es \mathbb{Z}_2 -orientable (en general, toda variedad lo es), pese a ser una superficie no orientable. Además, efectivamente desaparece la torsión en grado 1.

Trabajar con coeficientes en \mathbb{Z}_2 hace también los cálculos más sencillos, al no tener que preocuparse por los signos. Esto lleva consigo detrás el hecho de olvidarse de la orientación en los símlices.

Este último ejemplo también evidencia el problema que acarrea la discretización de espacios. Aunque esto hace posible su manipulación por máquinas, al trabajar en un espacio relativamente sencillo como el plano proyectivo real ya hay que manipular matrices de un tamaño considerable, con los costes computacionales que ello conlleva.

Capítulo 2

Construcción de complejos simpliciales

La enorme cantidad de datos de todo tipo que se produce cada día hace necesaria la aparición de técnicas de análisis para su posterior tratamiento e interpretación. Los avances tecnológicos han permitido que recabar datos sea una tarea relativamente sencilla en todos los ámbitos de la ciencia moderna. La obtención masiva de datos se hace en campos tan dispares como la genética, el reconocimiento de imágenes o las finanzas. De esta manera, el desarrollo de métodos de obtención de datos debe traer consigo el nacimiento de nuevas técnicas que extraigan la información relevante que contienen los datos.

Por ello, este capítulo expone varias formas de asociar un complejo simplicial a una nube de puntos que represente los datos a estudiar. Estas construcciones básicas permiten utilizar las técnicas homológicas presentadas previamente. Además, sirven para desarrollar las nuevas herramientas de análisis topológico de datos que se exponen en el próximo capítulo.

La estructura de este capítulo sigue el trabajo [5] de G. Carlsson y las definiciones allí expuestas. También se incluyen algunas explicaciones y observaciones recogidas en [37, 38].

2.1. Complejo de Čech

El primer objetivo planteado es conseguir asociar un complejo simplicial (abstracto) a la nube de puntos de interés que queramos estudiar. Aunque empezaremos considerando espacios topológicos arbitrarios, enseguida nos restringiremos a espacios métricos. Esto es razonable porque la propia naturaleza de los datos hace que lleven consigo alguna métrica detrás. Además, no es una gran restricción puesto que las nubes de datos están habitualmente en espacios que son, de hecho, euclídeos.

Para empezar, damos la definición de un tipo especial de recubrimiento que usaremos para construir un primer complejo simplicial más adelante.

Definición 2.1. Sean X un espacio topológico y $\mathcal{U} = \{U_i\}_{i \in I}$ un recubrimiento abierto de X . Decimos que \mathcal{U} es un *buen recubrimiento* de X si todos los abiertos $U_i \in \mathcal{U}$ y todas las intersecciones finitas y no vacías $U_{i_1} \cap \cdots \cap U_{i_q}$ son contráctiles.



Figura 2.1: Dos recubrimientos distintos para \mathbb{S}^1 , con los abiertos representados en el espacio ambiente. A la izquierda, un recubrimiento por abiertos cuya intersección tiene dos componentes conexas. A la derecha, un buen recubrimiento.

La condición de contractibilidad hace que posiblemente sean necesarios más abiertos para recubrir el espacio. Con esto se gana un estudio local más detallado, con lo que se captura de forma más fiel y precisa la forma del espacio.

El complejo de Čech pretende reconstruir un espacio mediante un recubrimiento abierto del mismo, pero de forma que sea computacionalmente más fácil trabajar con él. Previamente hacemos una construcción más general, válida para cualquier recubrimiento (no necesariamente abierto ni bueno) de un espacio topológico.

Definición 2.2. Sean X un espacio topológico y $\mathcal{U} = \{U_i\}_{i \in I}$ 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, $U_{i_0} \cap \dots \cap U_{i_q} \neq \emptyset$.

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.

El complejo de Čech es un tipo especial de nervio.

Definición 2.3. Sean X un espacio topológico y $\mathcal{U} = \{U_i\}_{i \in I}$ un buen recubrimiento de X . Llamamos *complejo de Čech* al nervio de \mathcal{U} .

Aunque resulta evidente que la realización geométrica del complejo de Čech asociado al buen recubrimiento de un espacio no preserva la forma ni el tamaño de ese espacio, sí se tiene que respeta el tipo de homotopía. Como hemos establecido en el Teorema 1.19, esto último es suficiente al trabajar en homología. Existen varios enunciados del próximo resultado, pero nos interesa el siguiente por su interpretación en el lenguaje de complejos de Čech.

Teorema 2.4 (del nervio). *Sea \mathcal{U} un buen recubrimiento finito de un espacio topológico X . Entonces, la realización geométrica de su nervio, $|\mathcal{N}(\mathcal{U})|$, tiene el mismo tipo de homotopía que X .*



Figura 2.2: En rojo están los respectivos nervios de los recubrimientos anteriores de \mathbb{S}^1 . Como el recubrimiento de la derecha es bueno, decimos que su nervio es un complejo de Čech.

La demostración puede encontrarse en [30, Teorema 15.21].

Existe un tipo especial de complejo de Čech, que es el que se suele usar cuando el espacio topológico es en realidad un espacio métrico (X, d) . Esta es la situación con la que nos encontraremos al tratar con datos: la nube de puntos de interés posiblemente se halla en algún espacio euclídeo \mathbb{R}^m .

Dado un $\varepsilon > 0$, para cada $x \in X$ tenemos la bola abierta de centro el punto x y radio ε :

$$B(x, \varepsilon) = \{y \in X : d(x, y) < \varepsilon\}.$$

Entonces, un recubrimiento abierto de X viene dado por la familia $\mathcal{B}(X, \varepsilon) = \{B(x, \varepsilon)\}_{x \in X}$ para cada $\varepsilon > 0$. Más generalmente, si $Y \subset X$ es un subconjunto tal que $X = \bigcup_{y \in Y} B(y, \varepsilon)$, podemos construir el nervio del recubrimiento $\{B(y, \varepsilon)\}_{y \in Y}$. Denotamos esta construcción por $\check{C}(Y, \varepsilon)$ y la llamamos *complejo de Čech asociado a Y con parámetro ε* .

En general, estos recubrimientos no son necesariamente buenos. Sin embargo, el siguiente resultado establece condiciones para la existencia de un buen recubrimiento por bolas abiertas. La demostración de la primera parte se debe a Jean-Claude Hausmann y puede consultarse en [26], donde se deja abierta la pregunta sobre la veracidad de la segunda afirmación. En [33], Janko Latschev da una respuesta afirmativa probando un resultado más fuerte.

Teorema 2.5. *Sea M una variedad riemanniana compacta. Entonces, existe un $\varepsilon_0 > 0$ de forma que $\mathcal{B}(M, \varepsilon)$ es un buen recubrimiento de M y $\check{C}(M, \varepsilon)$ tiene el mismo tipo de homotopía que M para todo $\varepsilon \leq \varepsilon_0$. Además, para cada $\varepsilon \leq \varepsilon_0$ existe un subconjunto finito $Y \subset M$ tal que el subcomplejo $\check{C}(Y, \varepsilon) \subset \check{C}(M, \varepsilon)$ también tiene el mismo tipo de homotopía que M .*

Dicho de otro modo, si el espacio es suficientemente bueno, existe un conjunto finito de puntos que tiene asociado un recubrimiento (también finito) cuyo complejo de Čech captura toda la información relevante sobre la estructura topológica del espacio original, para parámetros pequeños. Este conjunto finito de puntos es el que interesa en la práctica para que el complejo de Čech sea finito.

2.2. Complejo de Vietoris-Rips

Pese a sus buenas propiedades teóricas, en la práctica ocurre que el complejo de 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 en almacenamiento.

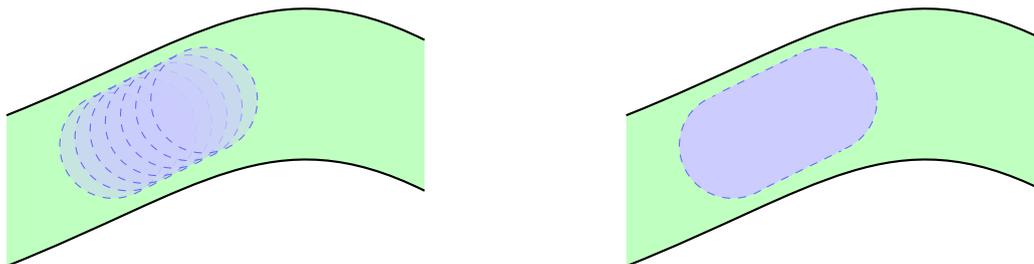


Figura 2.3: En el dibujo está representado un pequeño trozo de algún espacio métrico sin agujeros. A la izquierda, las bolas forman parte de un recubrimiento abierto del espacio. El elevado número de intersecciones provocará que el complejo de Čech tenga un gran número de símlices de distintas dimensiones. 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 $\mathcal{B}(X, \varepsilon)$, donde (X, d) es el espacio métrico en el que estamos trabajando. El complejo de Vietoris-Rips es una variante del complejo de Čech que implementa esta solución.

Definición 2.6. Sea (X, d) un espacio métrico. Dado un $\varepsilon > 0$, llamamos *complejo de Vietoris-Rips asociado a X con parámetro ε* , y lo denotamos por $VR(X, \varepsilon)$, al complejo simplicial cuyos vértices son los puntos de X y en el que los puntos $\{x_0, \dots, x_q\}$ forman un q -símlice 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 de Čech, es obvio que el complejo de Vietoris-Rips es un complejo simplicial abstracto, pues $d(x, x) = 0 \leq \varepsilon$ para todo $x \in X$ y la distancia entre dos puntos de cualquier subconjunto de $\{x_0, \dots, x_q\}$ sigue siendo ε o menos.

La primera observación que debe hacerse es que los conjuntos de vértices para los complejos de Čech y Vietoris-Rips son idénticos. Las diferencias empiezan a aflorar al considerar distintos valores del parámetro ε . Por ejemplo, la Figura 2.4 muestra cinco puntos, cada uno con una bola de radio ε centrada en él, para un determinado $\varepsilon > 0$.

En este caso, el complejo de Čech estaría formado por cinco 0-símlices y tres 1-símlices. Para el complejo de Vietoris-Rips solo dos puntos están lo suficientemente cerca entre sí como para que aparezca un 1-símlice, luego tenemos cinco 0-símlices y un único 1-símlice.

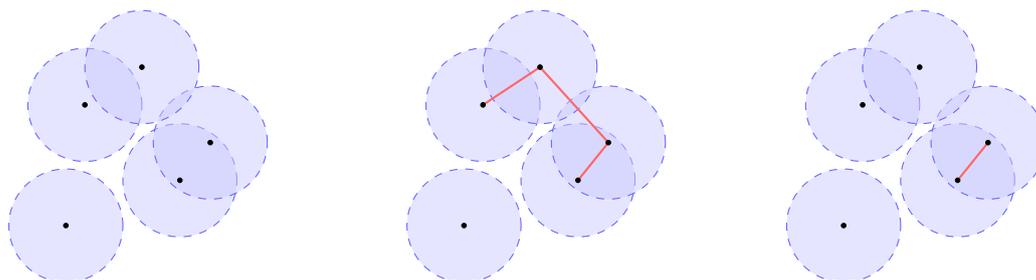


Figura 2.4: A la izquierda, una nube de cinco puntos con bolas abiertas del mismo radio centradas en ellos. En el centro, el complejo de Čech para el parámetro elegido. A la derecha, el complejo de Vietoris-Rips para el parámetro elegido.

Consideremos de nuevo los cinco puntos anteriores, pero incrementemos el parámetro ε un poco para ver cómo afecta a los complejos. La nueva situación está reflejada en la Figura 2.5.

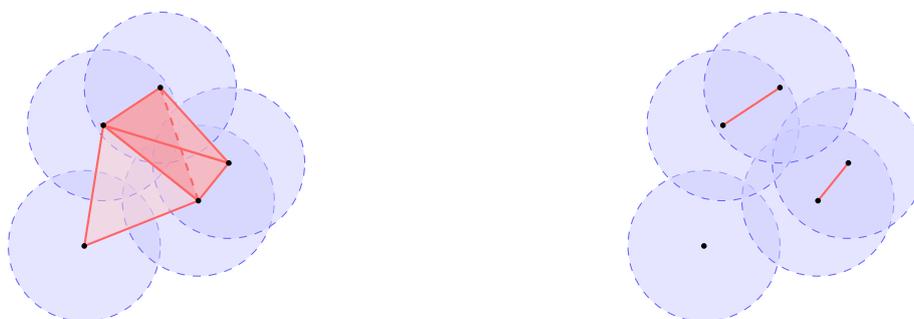


Figura 2.5: A la izquierda, el complejo de Čech con el nuevo parámetro. A la derecha, el complejo de Vietoris-Rips con el nuevo parámetro.

Con el nuevo parámetro, el complejo de Čech es mucho más grande: cinco 0-símplices, ocho 1-símplices, cinco 2-símplices y un 3-símplice. Por su parte, el complejo de Vietoris-Rips tan solo añade un 1-símplice más con respecto a la situación anterior, pues los puntos no están lo suficientemente cerca como para que se generen más símplices.

Estas consideraciones deberían poner aún más de manifiesto la distinta naturaleza de los dos complejos para un mismo parámetro. El complejo de Čech se fundamenta en el radio de las bolas y sus intersecciones, mientras que el de Vietoris-Rips queda determinado por sus vértices y 1-símplices, pues un símplice de dimensión mayor aparece si, y solo si, todas sus caras pertenecen al complejo.

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, y el ejemplo que acabamos de ver evidencia que los complejos pueden tener distintas estructuras topológicas. Aún así, el parámetro ε es útil como herramienta para comparar ambos complejos.

Proposición 2.7. *Sea $\varepsilon > 0$. Se verifican las siguientes inclusiones:*

$$\check{C}\left(X, \frac{\varepsilon}{2}\right) \subset VR(X, \varepsilon) \subset \check{C}(X, \varepsilon).$$

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. \square

La proposición anterior nos dice entonces que el complejo de Vietoris-Rips con parámetro ε capturará bien la topología de los datos subyacentes si así lo hacen los respectivos complejos de Čech con parámetros $\frac{\varepsilon}{2}$ y ε . Con un poco más de trabajo, las inclusiones se pueden afinar todavía más en los espacios euclídeos [14, 19].

2.3. Complejo de Delaunay

En general, tal y como indica el ejemplo y corrobora la Proposición 2.7, 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.

A grandes rasgos, la solución que se propone 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. Además, es también una forma de discretizar espacios continuos cuando se pretende calcular un complejo de Vietoris-Rips del espacio. Por ejemplo, para calcular el de \mathbb{S}^1 haría falta una infinidad de 0-símplices para representar todos los puntos de la circunferencia.

Por otra parte, a veces sucede que hay aglomeraciones de puntos y no interesa ir tanto al detalle, sino detectar estas acumulaciones. Este razonamiento lleva a plantearse si las bolas métricas utilizadas hasta ahora son entornos demasiado restrictivos cuando se producen agrupaciones de puntos.

Continuamos trabajando en un espacio métrico (X, d) y consideramos un subconjunto no vacío $\mathcal{L} \subset X$ de *puntos de referencia*.

Definición 2.8. Dado un punto de referencia $\lambda \in \mathcal{L}$, se define la *celda de Voronoi asociada a λ* como

$$V_\lambda = \{x \in X : d(x, \lambda) \leq d(x, \lambda') \text{ para todo } \lambda' \in \mathcal{L}\}.$$

Las celdas de Voronoi forman un recubrimiento de X , ya que para cada $x \in X$ siempre es posible comparar su distancia a cualquier elemento de \mathcal{L} . Las celdas de Voronoi descomponen el espacio métrico en distintas regiones. Nótese que el recubrimiento que forman las celdas de Voronoi no es abierto.

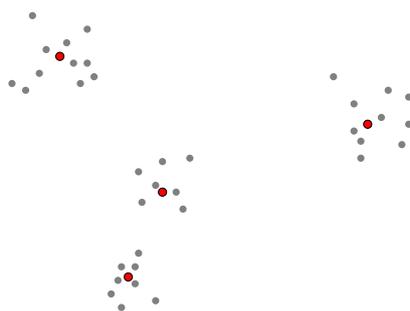


Figura 2.6: Una nube de puntos en el plano con cuatro acumulaciones (en inglés, «cluster») representadas por los puntos de referencia marcados en rojo.

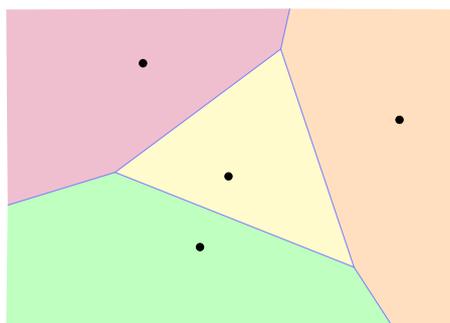


Figura 2.7: Los cuatro puntos de referencia con sus celdas de Voronoi en distintos colores.

Definición 2.9. Llamamos *complejo de Delaunay asociado a \mathcal{L}* al nervio del recubrimiento dado por las celdas de Voronoi.

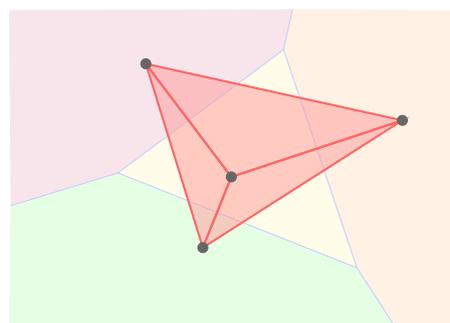


Figura 2.8: El complejo de Delaunay resultante del diagrama anterior. Los puntos de referencia son los cuatro 0-símplices, mientras que las intersecciones entre las regiones crean seis 1-símplices y tres 2-símplices.

Los puntos de referencia se pueden ver como representantes de un cúmulo de puntos que se encuentra en la celda de Voronoi correspondiente. Es decir, todos los puntos de la celda se sustituyen por el de referencia. Además, los complejos de Delaunay son complejos simpliciales de dimensiones más pequeñas que los de Čech y Vietoris-Rips.

2.4. Otros complejos simpliciales

Las nociones geométricas que hay detrás del complejo de Delaunay tienen su origen en las triangulaciones de Delaunay y sirven para refinar los complejos de Čech y Vietoris-Rips, que son las dos construcciones básicas. H. Edelsbrunner introdujo en [17] los complejos α combinando la descomposición de Voronoi con los complejos de Čech, de forma que estos conservasen la buena propiedad homotópica del teorema del nervio. Por su parte, la idea de elegir un conjunto de puntos de referencia les sirvió a V. de Silva y G. Carlsson para definir los complejos testigo en [13], más fácilmente computables que el de Vietoris-Rips aunque nuevamente sin las propiedades teóricas del complejo de Čech. Los complejos testigo también son una solución a los complejos de Delaunay degenerados (discretos) que resultan cuando el espacio métrico es finito.

Capítulo 3

Homología persistente

Dada una nube de puntos, ya sabemos asociar distintos complejos simpliciales de los que podemos calcular su homología. Pero, conociendo un poco mejor el complejo simplicial, conseguiremos más información y más sofisticada. Al variar el parámetro de los complejos simpliciales tenemos una filtración. Las filtraciones aparecen de forma natural al estudiar muchos procesos, como al refinar un análisis de datos. Mediante una filtración del complejo simplicial lograremos sucesivas aproximaciones que van mejorando la representación del espacio topológico subyacente y, con ella, podremos calcular la homología persistente. Esto significa que identificaremos aquellas clases de homología que aparecen en un instante de la filtración y perduran a lo largo de la misma, así como aquellas que nacen y desaparecen en algún momento.

La información que aportan los módulos de homología persistente queda recogida en unos diagramas de persistencia conocidos como códigos de barras. Estas representaciones gráficas hacen que sea más sencillo seguir la evolución de los generadores de la homología a lo largo de la filtración y sirven para distinguir visualmente las características que aparecen eventualmente de aquellas características que persisten durante toda la filtración del complejo.

Gracias al artículo [41] de A. Zomorodian y G. Carlsson, el cálculo de los códigos de barras se reduce nuevamente al álgebra lineal sobre un dominio de ideales principales. La mayor contribución de su artículo es el establecimiento de una correspondencia que describe por completo la estructura de la homología persistente sobre cualquier cuerpo.

3.1. Filtraciones

La construcción de algunos complejos simpliciales a partir de una nube de puntos requiere la elección de un parámetro ε , como hemos visto. Para valores pequeños del parámetro ε el complejo es un conjunto discreto de puntos; por otra parte, si el parámetro ε es muy grande el complejo se reduce a un único símplice de gran dimensión y todas sus caras. Surge la necesidad de encontrar un valor óptimo de ε que consiga que el complejo simplicial capture la topología del espacio subyacente al muestreo de datos.

Sin embargo, en lugar de encontrar ese valor óptimo de ε , se consideran distintos valores crecientes del parámetro y se estudia cómo cambia el complejo simplicial. Esto da lugar al concepto de filtración.

3.1.1. Filtraciones con índices naturales

En primer lugar, con vistas a que la exposición sea más clara, definiremos el concepto de filtración de un complejo simplicial con índices enteros no negativos. Posteriormente, generalizaremos la definición a filtraciones de complejos de cadenas con índices en la recta real ampliada.

Definición 3.1. Una *filtración* de un complejo simplicial \mathcal{K} es una sucesión de subcomplejos simpliciales encajados que empieza en el complejo vacío y termina en el complejo completo:

$$\emptyset = \mathcal{K}^0 \subset \mathcal{K}^1 \subset \dots \subset \mathcal{K}^n = \mathcal{K}.$$

Un complejo simplicial con una filtración se llama *complejo simplicial filtrado*.

Una filtración puede pensarse como una construcción en la que se van añadiendo colecciones de símlices en cada etapa, con la condición de que ningún símlice entre en la secuencia antes que todas sus caras.

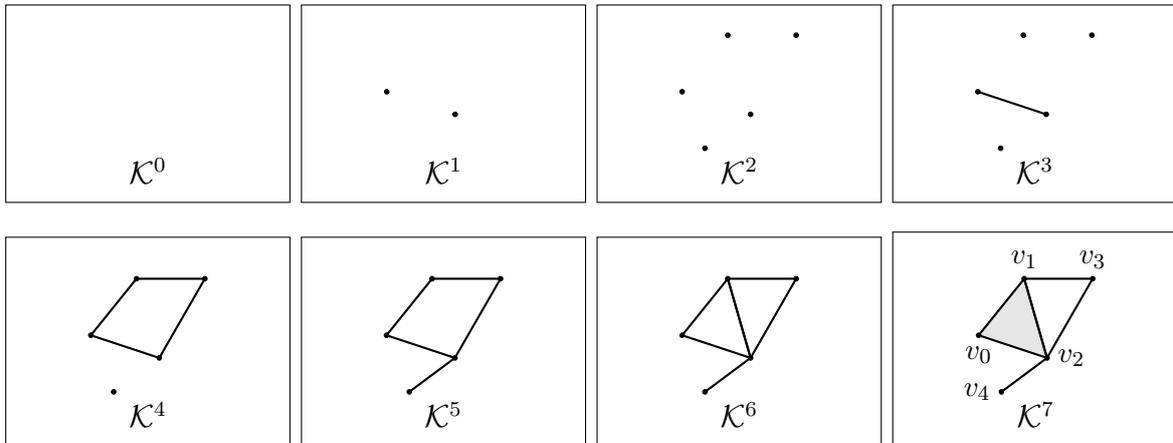


Figura 3.1: Una filtración del complejo simplicial de la Figura 1.3.

Fijado un anillo conmutativo y unitario R , es claro que cada subcomplejo simplicial \mathcal{K}^i de una filtración da lugar a su propio complejo de cadenas asociado $(C_\bullet(\mathcal{K}^i; R), \partial_\bullet^i)$. Así, para cada índice i de la filtración, que llamaremos *grado filtrante*, tenemos definidos en cualquier grado los R -módulos de cadenas, ciclos, bordes y homología, que por comodidad y simplicidad de notación denotamos respectivamente por

$$C_\bullet^i = C_\bullet(\mathcal{K}^i; R), \quad Z_\bullet^i = Z_\bullet(\mathcal{K}^i; R), \quad B_\bullet^i = B_\bullet(\mathcal{K}^i; R), \quad H_\bullet^i = H_\bullet(\mathcal{K}^i; R).$$

Por otra parte, la cadena de contenidos de una filtración proporciona inclusiones $\mathcal{K}^i \hookrightarrow \mathcal{K}^{i+j}$ para cada natural $j \geq 0$. Estas inclusiones son aplicaciones simpliciales, luego por el Corolario 1.16 inducen homomorfismos en homología:

$$\eta_{\bullet}^{i,j} : H_{\bullet}^i \longrightarrow H_{\bullet}^{i+j}.$$

Definición 3.2. Se define el q -ésimo módulo de j -persistencia de \mathcal{K}^i como la imagen del homomorfismo inducido por la inclusión:

$$H_q^{i,j} := \text{Im } \eta_q^{i,j}.$$

En el caso $j = 0$, se tendría que $H_q^{i,0} = H_q^i$ ya que $\eta_q^{i,0} = \text{id}$.

Según el Corolario 1.16, el homomorfismo inducido en homología $\eta_q^{i,j}$ consiste en llevar una clase de homología de \mathcal{K}^i , $[\alpha] \in H_q^i$, en aquella clase de homología de \mathcal{K}^{i+j} que la contiene (y que seguimos denotando por $[\alpha] \in H_q^{i+j}$). Estos homomorfismos inducidos en homología codifican los cambios topológicos locales que se producen durante el crecimiento del complejo simplicial en el transcurso de la filtración. La homología persistente intenta medir los cambios globales.

Además, dos ciclos homólogos en \mathcal{K}^i siguen existiendo y siendo homólogos en \mathcal{K}^{i+j} . Teniendo esto en cuenta, los módulos de homología persistente están formados por las clases de homología de \mathcal{K}^i que todavía viven en \mathcal{K}^{i+j} . En este sentido, esta interpretación permite reescribir los módulos de persistencia en la forma dada por la siguiente proposición.

Proposición 3.3. *Se verifica que:*

$$H_q^{i,j} \cong \frac{Z_q^i}{B_q^{i+j} \cap Z_q^i}.$$

Demostración. El módulo cociente está bien definido porque $\mathcal{K}^i \subset \mathcal{K}^{i+j}$, lo que en particular induce las inclusiones $Z_q^i \subset C_q^i \subset C_q^{i+j}$. Así, los módulos del denominador son ambos submódulos de C_q^{i+j} , luego su intersección es de nuevo un módulo (y es obviamente submódulo del numerador).

De esta forma, el isomorfismo se sigue inmediatamente a partir de las consideraciones previas a la proposición. En otras palabras, un ciclo de C_q^i seguirá siendo un ciclo en C_q^{i+j} , esto es, $Z_q^i \subset Z_q^{i+j}$. Sin embargo, su clase con respecto a B_q^i puede ser no nula, y ser nula o no con respecto a B_q^{i+j} . \square

De forma natural aparecen los números de Betti, en analogía con los de homología simplicial.

Definición 3.4. Se define el q -ésimo número de Betti de j -persistencia asociado al complejo simplicial \mathcal{K}^i de la filtración como

$$\beta_q^{i,j} := \text{rango } H_q^{i,j}.$$

Por tanto, el número de Betti persistente $\beta_q^{i,j}$ cuenta las clases de homología del complejo \mathcal{K}^{i+j} creadas durante la filtración, en el complejo \mathcal{K}^i o antes.

3.1.2. Filtraciones con índices reales

Como anunciamos previamente, a continuación generalizaremos la definición de filtración a complejos de cadenas e índices en un continuo de valores. La definición y los lemas posteriores están tomados de [4].

Denotemos por $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ el conjunto totalmente ordenado que denominamos recta real ampliada.

Definición 3.5. Llamamos $\overline{\mathbb{R}}$ -filtración ascendente de un complejo de cadenas $(C_\bullet, \partial_\bullet)$ a una colección de complejos de cadenas $\{(\mathcal{F}^r C_\bullet, \partial_\bullet)\}_{r \in \overline{\mathbb{R}}}$ tal que:

- (I) $(\mathcal{F}^r C_\bullet, \partial_\bullet)$ es un subcomplejo de cadenas de $(C_\bullet, \partial_\bullet)$ para todo $r \in \overline{\mathbb{R}}$.
- (II) $(\mathcal{F}^r C_\bullet, \partial_\bullet) \subset (\mathcal{F}^{r'} C_\bullet, \partial_\bullet)$ para $r \leq r'$, con $r, r' \in \overline{\mathbb{R}}$.
- (III) $\bigcap_{r \in \overline{\mathbb{R}}} (\mathcal{F}^r C_\bullet, \partial_\bullet) = \emptyset$.
- (IV) $\bigcup_{r \in \overline{\mathbb{R}}} (\mathcal{F}^r C_\bullet, \partial_\bullet) = (C_\bullet, \partial_\bullet)$.

Un complejo de cadenas junto con una $\overline{\mathbb{R}}$ -filtración se denomina *complejo de cadenas $\overline{\mathbb{R}}$ -filtrado*.

De forma equivalente, las propiedades (III) y (IV) de la definición se reescriben, respectivamente, como

$$(\mathcal{F}^{-\infty} C_\bullet, \partial_\bullet) = \emptyset \quad \text{y} \quad (\mathcal{F}^{+\infty} C_\bullet, \partial_\bullet) = (C_\bullet, \partial_\bullet).$$

Nótese que estamos denotando por ∂_q la diferencial en grado q de los subcomplejos de cadenas de la filtración, que se corresponden con las respectivas restricciones del operador borde del complejo de cadenas. Esto significa que la diferencial conserva la filtración, en el sentido de que $\partial_q(\mathcal{F}^r C_q) \subset \mathcal{F}^r C_{q-1}$.

Resulta evidente que en la definición de $\overline{\mathbb{R}}$ -filtración se puede sustituir $\overline{\mathbb{R}}$ por cualquier otro conjunto totalmente ordenado en el que todo subconjunto tenga supremo e ínfimo. En este caso, la recta real ampliada es conveniente para trabajar con las filtraciones que proporcionan los complejos de Čech y de Vietoris-Rips, cuyos parámetros toman valores reales positivos.

Lema 3.6. *Sea $(C_\bullet, \partial_\bullet)$ un complejo de cadenas $\overline{\mathbb{R}}$ -filtrado. Para toda q -cadena $\alpha \in C_q$ existe un $r_0 \in \overline{\mathbb{R}}$ tal que*

$$\sup\{r' : \alpha \notin \mathcal{F}^{r'} C_q\} = r_0 = \inf\{r'' : \alpha \in \mathcal{F}^{r''} C_q\}.$$

Demostración. Es consecuencia de las propiedades que cumple una $\overline{\mathbb{R}}$ -filtración por definición y del orden de $\overline{\mathbb{R}}$.

En primer lugar, tenemos garantizada la existencia de supremos e ínfimos en $\overline{\mathbb{R}}$. Llamando

$$r_0 = \sup\{r' : \alpha \notin \mathcal{F}^{r'} C_q\} \quad \text{y} \quad s_0 = \inf\{r'' : \alpha \in \mathcal{F}^{r''} C_q\},$$

es claro que $r_0 \leq s_0$, pues una vez aparece $\alpha \in C_q$ en la filtración no puede desaparecer por la propiedad (II) de la definición. Finalmente, por ser supremo del conjunto, $r_0 + \varepsilon \geq s_0$ para todo $\varepsilon > 0$; luego $r_0 \geq s_0$. Por tanto, $r_0 = s_0$. \square

Usemos nuevamente la notación condensada C_\bullet^r , Z_\bullet^r , B_\bullet^r y H_\bullet^r para representar los módulos de cadenas, ciclos, bordes y homología para el r -ésimo elemento de una $\overline{\mathbb{R}}$ -filtración. En analogía con los complejos simpliciales filtrados por índices naturales, un complejo de cadenas $\overline{\mathbb{R}}$ -filtrado también proporciona inclusiones $\mathcal{F}^r C_\bullet \hookrightarrow \mathcal{F}^{r+l} C_\bullet$ para cada $l \geq 0$, que inducen los homomorfismos en homología

$$\eta_\bullet^{r,l}: H_\bullet^r \longrightarrow H_\bullet^{r+l}.$$

Definición 3.7. Se define el q -ésimo módulo de l -persistencia de $\mathcal{F}^r C_q$ como la imagen del homomorfismo inducido por la inclusión:

$$H_q^{r,l} := \text{Im } \eta_q^{r,l}.$$

La definición de los números de Betti persistentes es completamente análoga, así como la reinterpretación dada por la Proposición 3.3 en términos de Z_q^r y B_q^{r+l} . En este sentido, $H_q^{r,l}$ caracteriza aquellos q -ciclos de $\mathcal{F}^r C_q$ que no son borde de ninguna $(q+1)$ -cadena del complejo $\mathcal{F}^{r+l} C_q$, que es más grande. De esta forma, recordemos que si $\alpha \in Z_q^r$, entonces α representa una clase de homología $[\alpha] \in H_q^r$. Como $Z_q^r \subset Z_q^{r'}$ para $r' \geq r$, α también representa una clase de homología en $H_q^{r'}$ (que seguimos denotando por $[\alpha]$). Pero puede ocurrir que $[\alpha] \neq 0$ en H_q^r y que $[\alpha] = 0$ en $H_q^{r'}$.

Veamos cómo se comportan los índices de una $\overline{\mathbb{R}}$ -filtración con el nacimiento y muerte de clases de homología.

Lema 3.8. Sea $(C_\bullet, \partial_\bullet)$ un complejo de cadenas $\overline{\mathbb{R}}$ -filtrado. Para todo q -ciclo $\alpha \in Z_q$, el conjunto

$$I_\alpha := \{r \in \overline{\mathbb{R}}: 0 \neq [\alpha] \in H_q^r\}$$

o bien es vacío o bien es un intervalo.

Demostración. Sean $\alpha \in Z_q$ un ciclo y r_1 el valor dado por el Lema 3.6. Tenemos dos posibilidades: que el ciclo α sea también un borde o que no lo sea.

Si $\alpha \in B_q$, entonces existe un $\beta \in C_{q+1}$ tal que $\partial_{q+1}(\beta) = \alpha$. Sea r_2 el correspondiente valor dado por el Lema 3.6 para β . Como $\beta \in \mathcal{F}^r C_{q+1}$, para algún $r \geq r_2$, esto implica que $\partial_{q+1}(\beta) = \alpha \in \mathcal{F}^r C_q$. Por tanto, necesariamente debe ocurrir que $r_2 \geq r_1$. De esta forma, α representa una clase no nula de homología exactamente en el intervalo (posiblemente vacío) que empieza en r_1 y termina en r_2 . Este intervalo contiene el extremo inicial r_1 si, y solo si, $\alpha \in \mathcal{F}^{r_1} C_q$, y contiene el extremo final r_2 si, y solo si, $\beta \notin \mathcal{F}^{r_2} C_{q+1}$.

Si α no es un q -borde proveniente de una $(q+1)$ -cadena, entonces α representa una clase no trivial de homología en $\mathcal{F}^r C_q$ exactamente cuando r está en el intervalo $\{s: s \geq r_1\}$ o el intervalo $\{s: s > r_1\}$, que empieza en r_1 y termina en el infinito. Como antes, r_1 pertenece al intervalo si, y solo si, $\alpha \in \mathcal{F}^{r_1} C_q$. \square

Ejemplo 3.9 (Filtración asociada al complejo de Čech). Supongamos que el conjunto $X = \{x_1, \dots, x_n\}$ es un muestreo de puntos de algún espacio métrico (M, d) . Sea $\mathcal{K} = \check{C}(X, +\infty)$ el mayor complejo simplicial que se puede elaborar con el conjunto de vértices X .

La colección de complejos de Čech $\{\check{C}(X, \varepsilon)\}_{\varepsilon \geq 0}$ con distintos valores no negativos del parámetro proporciona una $\overline{\mathbb{R}}$ -filtración, que pasamos a describir. Cada uno de los elementos de la familia anterior es un complejo simplicial, que por brevedad denotaremos por \mathcal{K}^ε . Fijado un anillo R , para cada uno de estos complejos simpliciales, basta tomar el correspondiente complejo de cadenas $(C_\bullet(\mathcal{K}^\varepsilon; R), \partial_\bullet^\varepsilon)$ para obtener la filtración. En efecto, $(C_\bullet(\mathcal{K}^\varepsilon; R), \partial_\bullet^\varepsilon)$ es un subcomplejo de cadenas de $(C_\bullet(\mathcal{K}; R), \partial_\bullet)$, puesto que \mathcal{K}^ε es un subcomplejo simplicial de \mathcal{K} para todo $\varepsilon \geq 0$, y

$$(C_\bullet(\mathcal{K}^\varepsilon; R), \partial_\bullet^\varepsilon) \subset (C_\bullet(\mathcal{K}^{\varepsilon'}; R), \partial_\bullet^{\varepsilon'})$$

para $\varepsilon \leq \varepsilon'$, ya que $\check{C}(X, \varepsilon) \subset \check{C}(X, \varepsilon')$. Además, $\mathcal{K}^\varepsilon = \emptyset$ para todo $\varepsilon \leq 0$ y, si δ es el diámetro del conjunto X , entonces $\mathcal{K}^\varepsilon = \mathcal{K}$ para $\varepsilon > \frac{\delta}{2}$. De esta manera, la colección

$$\{(C_\bullet(\mathcal{K}^\varepsilon; R), \partial_\bullet^\varepsilon)\}_{\varepsilon \geq 0}$$

es una $\overline{\mathbb{R}}$ -filtración del complejo de cadenas $(C_\bullet(\mathcal{K}; R), \partial_\bullet)$.

Ejemplo 3.10 (Filtración asociada al complejo de Vietoris-Rips). Supongamos que $X = \{x_1, \dots, x_n\}$ es de nuevo un muestreo de puntos de algún espacio métrico (M, d) . La colección de complejos de Vietoris-Rips $\{VR(X, \varepsilon)\}_{\varepsilon \geq 0}$ con distintos valores no negativos del parámetro proporciona otra $\overline{\mathbb{R}}$ -filtración. Nuevamente, por simplicidad, denotamos por \mathcal{K}^ε cada uno de los complejos simpliciales de la familia anterior y tomamos $\mathcal{K} = VR(X, +\infty)$, el mayor complejo simplicial que se puede elaborar con el conjunto de vértices X .

Para un anillo R fijado, construimos el complejo de cadenas asociado a cada complejo simplicial, $(C_\bullet(\mathcal{K}^\varepsilon; R), \partial_\bullet^\varepsilon)$, y conseguimos así una nueva $\overline{\mathbb{R}}$ -filtración del complejo de cadenas $(C_\bullet(\mathcal{K}; R), \partial_\bullet)$, que esta vez vendrá dada de la siguiente forma:

$$\{(C_\bullet(\mathcal{K}^\varepsilon; R), \partial_\bullet^\varepsilon)\}_{\varepsilon \geq 0}.$$

Las comprobaciones de que verifica la definición son inmediatas y análogas a las del complejo de Čech (en este caso, $\mathcal{K}^\varepsilon = \mathcal{K}$ para $\varepsilon > \delta$, siendo δ el diámetro del conjunto X).

3.2. Códigos de barras

Para visualizar fácilmente la homología persistente se usan los llamados códigos de barras. Estas representaciones en el plano incluyen los índices ascendentes de la filtración en el eje horizontal y los distintos módulos de homología en el eje vertical. La homología persistente se codifica mediante ciertos intervalos horizontales, que son representantes arbitrarios de cada clase. Los extremos de estos intervalos indican los momentos de la filtración en los que nace y muere cada clase de homología.

Siguiendo el artículo de P. Bubenik y P. T. Kim [4], damos una definición más formal de un código de barras y un resultado sobre los códigos de barras asociados a las filtraciones de Čech y Vietoris-Rips.

3.2.1. Intervalos de persistencia

Los códigos de barras están formados por los intervalos durante los que persisten o viven las clases no triviales de homología a través de una filtración. Los extremos de los intervalos indican las etapas de la filtración en las que nacen y mueren las clases de homología.

Definición 3.11. Sean $(C_\bullet, \partial_\bullet)$ un complejo de cadenas $\overline{\mathbb{R}}$ -filtrado y $\alpha \in Z_q$ un ciclo. Se define el *intervalo de persistencia* representado por α como el intervalo I_α dado por el Lema 3.8.

Obviamente, de la definición de I_α se sigue que si $\alpha' = \alpha + \partial_{q+1}(\beta)$ para algún $\beta \in C_{q+1}$, entonces $I_{\alpha'} = I_\alpha$ ya que $[\alpha'] = [\alpha]$ y estas serán clases no nulas para los mismos parámetros de la $\overline{\mathbb{R}}$ -filtración. Recíprocamente, puede ocurrir que $[\alpha] \neq [\alpha']$ y, sin embargo, $I_\alpha = I_{\alpha'}$. Por tanto, tiene sentido la siguiente definición.

Definición 3.12. Para un complejo de cadenas $\overline{\mathbb{R}}$ -filtrado $(C_\bullet, \partial_\bullet)$ se define su *código de barras* asociado como el multiconjunto $\{J_\alpha\}_{\alpha \in S}$, donde $S \subset Z_\bullet$ es un cierto subconjunto de ciclos, tal que:

- (I) J_α es un subintervalo de I_α .
- (II) Para todo $r \in \overline{\mathbb{R}}$, el conjunto $\{[\alpha] \in H_\bullet^r : \alpha \in S, r \in J_\alpha\}$ está formado por los generadores de H_\bullet^r .

La definición anterior es bastante oscura en una primera lectura. Para desentrañarla, vamos a detenernos un momento a analizar su significado.

En primer lugar, el subconjunto S de los ciclos tiene como función evitar los intervalos de persistencia redundantes que habría para ciclos homólogos, tal y como indica el comentario previo a la definición. En este sentido, el hecho de que un código de barras se defina como un multiconjunto es porque pueden nacer y morir simultáneamente clases de homología distintas que generen el mismo intervalo, provocando que este deba aparecer con multiplicidad mayor que uno.

Por otra parte, la selección de los subintervalos $J_\alpha \subset I_\alpha$ se hace necesaria porque al avanzar en la filtración se ganan nuevas clases de homología y se pierden otras al volverse triviales o fusionarse con una clase de homología más antigua. Por ejemplo, supongamos que las clases de homología $[\sigma]$ y $[\tau]$ nacen, respectivamente, en $\mathcal{F}^r C_q$ y $\mathcal{F}^{r'} C_q$, con $r \leq r'$. Ambas clases se fusionan si existen $l, l' \geq 0$ tales que

$$\eta_q^{r,l}([\sigma]) = \eta_q^{r',l'}([\tau]).$$

En ese caso, como la clase $[\sigma]$ es más antigua, el código de barras refleja únicamente un subintervalo $J_\tau \subset I_\tau$ (desde que $[\tau]$ nace hasta que se fusiona con $[\sigma]$) en vez del intervalo de persistencia I_τ al completo; en otras palabras, mientras es un generador de la homología. Por tanto, los intervalos de persistencia que aparecen íntegros en el código de barras corresponden a las clases de homología primigenias, que son las que no se fusionan con otras.

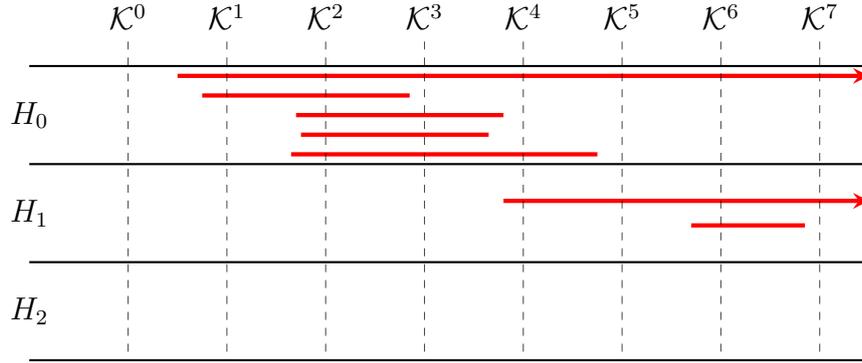


Figura 3.2: Código de barras correspondiente a la filtración de la Figura 3.1.

Ejemplo 3.13. La Figura 3.2 muestra el código de barras asociado a la filtración de la Figura 3.1. Los generadores de H_0 reflejan las componentes conexas del complejo simplicial. Las dos barras superiores del código de barras representan las componentes conexas asociadas a los dos 0-símplices que aparecen en la filtración antes de \mathcal{K}^1 . Antes de \mathcal{K}^3 aparece un 1-símplice que une estos vértices, así que las dos clases de homología se fusionan y permanece la primera que apareció en la filtración. De forma similar, otras tres clases aparecen en distintos puntos de la filtración y acaban fusionándose con la primera, que es la que perdura a lo largo del tiempo.

Para H_1 , un primer generador nace cuando antes de \mathcal{K}^4 se cierra un 1-ciclo, creando un agujero. Más adelante, aparece una nueva arista y este agujero se divide en dos. En este caso, el nuevo generador se vuelve trivial cuando se rellena la cara con un 2-símplice.

En vista del código de barras, podemos constatar que existen dos características topológicas persistentes en la filtración. Una significa que el complejo es conexo, la otra detecta el ciclo formado en \mathcal{K}^4 que nunca se rellena. El resto es ruido topológico.

3.2.2. Filtraciones de Čech y Vietoris-Rips

En un primer momento, tal y como se establece en la demostración del Lema 3.8, el intervalo de persistencia de un ciclo puede contener sus dos extremos, solo uno de ellos o ninguno. Sin embargo, las construcciones específicas de las filtraciones de Čech y Vietoris-Rips permiten concretar los tipos de intervalos de persistencia asociados a sus ciclos.

El siguiente resultado proporciona el valor exacto del número $r_0 \in \overline{\mathbb{R}}$ del Lema 3.6 para las filtraciones de Čech y Vietoris-Rips.

Lema 3.14. Sean (M, d) un espacio métrico y $X = \{x_1, \dots, x_n\}$ un muestreo de puntos de M . Sea $\{(\mathcal{F}^r C_\bullet, \partial_\bullet)\}_{r \in \overline{\mathbb{R}}}$ una filtración de Čech o de Vietoris-Rips y sea $\alpha = \sum_{i=0}^m \alpha_i \sigma_i$ una q -cadena, donde $\sigma_i = [x_{i_0}, \dots, x_{i_q}]$ es un q -símplice.

(1) Para la filtración de Čech sea

$$r_0 = \max_{0 \leq i \leq m} \inf \{ \varepsilon_i : \text{existe un } x \in M \text{ tal que } x_{i_0}, \dots, x_{i_q} \in B(x, \varepsilon_i) \}.$$

Entonces, $\alpha \notin \mathcal{F}^{r'} C_q$ para todo $r' \leq r_0$ y $\alpha \in \mathcal{F}^{r''} C_q$ para todo $r'' > r_0$.

(2) Para la filtración de Vietoris-Rips sea

$$r_0 = \max_{0 \leq i \leq m} \max_{0 \leq j, k \leq q} d(x_{i_j}, x_{i_k}).$$

Entonces, $\alpha \notin \mathcal{F}^{r'} C_q$ para todo $r' < r_0$ y $\alpha \in \mathcal{F}^{r''} C_q$ para todo $r'' \geq r_0$.

Demostración. El Lema 3.6 proporciona la existencia de un $r_0 \in \overline{\mathbb{R}}$ que verifica la tesis. Veamos que, en efecto, este valor es el que se indica en el enunciado para cada filtración. En primer lugar, nótese que para construir la q -cadena $\alpha = \sum_{i=0}^m \alpha_i \sigma_i$ en un complejo de la filtración es necesario y suficiente que todos los elementos de la familia compuesta por los m símlices orientados de dimensión q dada por $\{[x_{i_0}, \dots, x_{i_q}]\}_{0 \leq i \leq m}$ estén presentes en el complejo $\mathcal{F}^{r_0} C_q$ de la filtración.

Para la filtración de Čech, los puntos x_{i_0}, \dots, x_{i_q} forman un q -símplice si las correspondientes bolas de radio r_0 centradas en los puntos tienen intersección común no vacía. Si tomamos un punto $x \in M$ en esa intersección, $B(x, r_0)$ deberá contener también los vértices del símplice. Es claro que debemos tomar el radio más pequeño posible para cada grupo de q puntos (es un ínfimo por ser las bolas abiertas) y después tomar el máximo de todos ellos, pues tiene que ser el más grande posible entre los mismos (en otro caso, algún grupo de q vértices no formaría un símplice).

Para la filtración de Vietoris-Rips, los puntos x_{i_0}, \dots, x_{i_q} deben estar todos a distancia r_0 o menos de los demás para formar símplice. Tomando el primer máximo nos aseguramos de que los puntos definan un símplice. Finalmente, el segundo máximo es necesario para que todos los grupos de q vértices formen un símplice. \square

Proposición 3.15. Sea $\{(\mathcal{F}^r C_\bullet, \partial_\bullet)\}_{r \in \overline{\mathbb{R}}}$ una filtración de Čech o de Vietoris-Rips. Dado un q -ciclo $\alpha \in Z_q$, el intervalo de persistencia de α o bien es vacío o bien es de la forma

(1) (r_1, r_2) o $(r_1, +\infty]$ para la filtración de Čech.

(2) $[r_1, r_2)$ o $[r_1, +\infty)$ para la filtración de Vietoris-Rips.

Demostración. La misma demostración vale en ambos casos. Para determinar si un extremo pertenece o no al intervalo de persistencia, tenemos en cuenta las observaciones hechas en la demostración del Lema 3.8.

Si $\alpha \in Z_q$ es un q -ciclo, por el Lema 3.6 existe un intervalo de persistencia (posiblemente vacío) para α . Si α no es un borde, basta aplicar el Lema 3.14 a α . Si $\alpha \in B_q$ es un borde, existe un $\beta \in C_{q+1}$ tal que $\partial_{q+1}(\beta) = \alpha$; basta entonces aplicar el Lema 3.14 a β . \square

3.3. Estructuras algebraicas

Los módulos de homología persistente se pueden englobar en los conocidos como módulos de persistencia. Además, veremos que los módulos de persistencia están relacionados con otra estructura que llamaremos complejo de persistencia.

El estudio de la homología persistente desde este punto de vista permite dar algunos resultados interesantes que se relacionarán con los códigos de barras. Este hecho es importante, pues ya hemos visto que los códigos de barras recogen la información relevante de la homología persistente. Como ocurría en la homología simplicial, el teorema de estructura vuelve a ser clave para relacionar conceptos y establecer resultados.

3.3.1. Módulos y complejos de persistencia

Las siguientes definiciones son generalizaciones de las presentadas en [41], donde se trabaja directamente con filtraciones con índices naturales. En la presente sección, empezamos desarrollando la teoría para conjuntos de índices arbitrarios con una relación de orden total y, posteriormente, nos restringiremos a \mathbb{N} . Como en capítulos previos, R es un anillo conmutativo y unitario.

Definición 3.16. Sea (I, \leq) un conjunto totalmente ordenado. Un R -módulo de persistencia parametrizado por I es una familia de R -módulos $\{M^i\}_{i \in I}$ junto con homomorfismos de R -módulos $\{\varphi^{i \rightarrow j} : M^i \rightarrow M^j\}_{i \leq j}$ tales que:

- (I) $\varphi^{i \rightarrow i} = \text{id}_{M^i}$ para todo $i \in I$.
- (II) $\varphi^{j \rightarrow k} \circ \varphi^{i \rightarrow j} = \varphi^{i \rightarrow k}$ para todo $i \leq j \leq k$.

Una definición completamente análoga a la anterior permite establecer lo que es un complejo de persistencia.

Definición 3.17. Sea (I, \leq) un conjunto totalmente ordenado. Un complejo de persistencia parametrizado por I es una familia de complejos de cadenas $\{(C_\bullet^i, \partial_\bullet^i)\}_{i \in I}$ sobre R junto con aplicaciones de cadenas $\{f_\bullet^{i \rightarrow j} : C_\bullet^i \rightarrow C_\bullet^j\}_{i \leq j}$ tales que:

- (I) $f_\bullet^{i \rightarrow i} = \text{id}_{C_\bullet^i}$ para todo $i \in I$.
- (II) $f_\bullet^{j \rightarrow k} \circ f_\bullet^{i \rightarrow j} = f_\bullet^{i \rightarrow k}$ para todo $i \leq j \leq k$.

Notación 3.18. Anteriormente, hemos usado otra notación para los superíndices de las aplicaciones cuando el conjunto de índices era \mathbb{N} o $\overline{\mathbb{R}}$. En esos casos particulares, como hay definida una operación suma y existe un elemento neutro para la misma, los superíndices separados por comas indicaban el índice filtrante de inicio y el número de pasos que se avanza en la filtración. Para conjuntos de índices arbitrarios, al carecer en general de una operación suma, la notación mediante las flechas indica los índices filtrantes de inicio y final.

El siguiente diagrama muestra cómo podría ser un pequeño trozo de un complejo de persistencia, fijados los grados filtrantes. Cada columna es un complejo de cadenas, mientras que las aplicaciones de cadenas conectan por filas las etapas fijadas de la filtración.

$$\begin{array}{ccccccc}
& & \vdots & & \vdots & & \\
& & \downarrow \partial_{q+1}^i & & \downarrow \partial_{q+1}^j & & \\
\cdots & \xrightarrow{f_q^{\cdot \rightarrow i}} & C_q^i & \xrightarrow{f_q^{i \rightarrow j}} & C_q^j & \xrightarrow{f_q^{j \rightarrow \cdot}} & \cdots \\
& & \downarrow \partial_q^i & & \downarrow \partial_q^j & & \\
\cdots & \xrightarrow{f_{q-1}^{\cdot \rightarrow i}} & C_{q-1}^i & \xrightarrow{f_{q-1}^{i \rightarrow j}} & C_{q-1}^j & \xrightarrow{f_{q-1}^{j \rightarrow \cdot}} & \cdots \\
& & \downarrow \partial_{q-1}^i & & \downarrow \partial_{q-1}^j & & \\
& & \vdots & & \vdots & &
\end{array}$$

Ejemplo 3.19. Un complejo simplicial I -filtrado junto con las inclusiones de simplices forman un complejo de persistencia parametrizado por I . Por su parte, la homologıa de un complejo de persistencia da lugar a un modulo de persistencia parametrizado por I , siendo las aplicaciones $\eta_{\bullet}^{i \rightarrow j}$ definidas previamente los homomorfismos requeridos en cada grado homologico.

3.3.2. Clasificacion sobre cuerpos

El siguiente paso es clasificar, salvo isomorfismo, los modulos de persistencia sobre cuerpos. Los resultados de este apartado, de gran importancia, se deben a la extension que A. Zomorodian y G. Carlsson [41] hicieron de un algoritmo para calcular la persistencia. El algoritmo original estaba definido unicamente para complejos simpliciales tridimensionales y coeficientes en \mathbb{Z}_2 . La extension de este algoritmo se hace viendo un sistema inductivo de espacios vectoriales como una unica entidad: un modulo de persistencia. Esto permite generalizar la clasificacion de sistemas inductivos de espacios vectoriales tridimensionales sobre \mathbb{Z}_2 a dimensiones arbitrarias y cualquier cuerpo. Ademas, veremos que no existe una clasificacion ası de simple para sistemas inductivos de modulos sobre un anillo arbitrario.

Estos teoremas de clasificacion exigen una condicion de finitud en las estructuras que manejamos.

Definicion 3.20. Un modulo de persistencia $\mathcal{M} = \{M^i, \varphi^{i \rightarrow j}\}_{i \leq j}$ parametrizado por I sobre un anillo R se dice de *tipo finito* si cada R -modulo M^i es finitamente generado como R -modulo y si existe un $n \in I$ tal que para todo $j \geq n$ las aplicaciones $\varphi^{n \rightarrow j}$ son isomorfismos.

Definicion 3.21. Un complejo de persistencia $\mathcal{C} = \{C_{\bullet}^i, f_{\bullet}^{i \rightarrow j}\}_{i \leq j}$ parametrizado por I sobre un anillo R se dice de *tipo finito* si cada complejo de cadenas C_q^i es finitamente generado como R -modulo y si existe un $n \in I$ tal que para todo $j \geq n$ las aplicaciones $f_{\bullet}^{n \rightarrow j}$ son isomorfismos.

Observacion 3.22. Como los complejos simpliciales considerados en este trabajo son finitos, cualquier filtracion genera un complejo de persistencia de tipo finito cuya homologıa es un modulo de persistencia de tipo finito.

En lo que sigue, nos centraremos en módulos y complejos de persistencia parametrizados por \mathbb{N} y, más adelante, comentaremos los parametrizados por $\overline{\mathbb{R}}$, pues son los casos de interés que estamos manejando.

En primer lugar, damos una estructura de $R[t]$ -módulo a los módulos de persistencia parametrizados por \mathbb{N} . Para ello, equipamos $R[t]$ de una estructura de anillo graduado tomando la descomposición

$$R[t] = \bigoplus_{i \in \mathbb{N}} t^i \cdot R,$$

donde $t^i \cdot R = \{rt^i : r \in R\}$. El producto de polinomios cumple la condición de graduación:

$$(t^i \cdot R)(t^j \cdot R) \subset t^{i+j} \cdot R.$$

Esto permite dotar de una estructura de módulo graduado sobre el anillo de polinomios $R[t]$ a cualquier módulo de persistencia $\mathcal{M} = \{M^i, \varphi^{i,j}\}$ parametrizado por \mathbb{N} como

$$M = \bigoplus_{i \in \mathbb{N}} M^i,$$

donde la acción de t sobre M viene dada por la extensión lineal de

$$t \cdot m^i = \varphi^{i,1}(m^i)$$

para todo $m^i \in M^i$. Así,

$$(t^i \cdot R)M^j \subset M^{i+j}.$$

En otras palabras, la acción de t aumenta el grado en una unidad.

En el caso de un complejo de persistencia que derive de la filtración de un complejo simplicial, la acción del anillo de polinomios conecta la homología a través de los distintos complejos de la filtración. Es por ello que las propiedades algebraicas de este objeto contienen la información de los módulos de homología persistente de la filtración.

Para poder aplicar el Teorema 1.22, consideraremos que el anillo de coeficientes R es un cuerpo \mathbb{F} . De esta forma, los R -módulos se vuelven \mathbb{F} -espacios vectoriales y el anillo de polinomios $\mathbb{F}[t]$ es un dominio de ideales principales.

Teorema 3.23. *Sea $\mathcal{M} = \{M^i, \varphi^{i,j}\}$ un módulo de persistencia de tipo finito parametrizado por \mathbb{N} sobre $\mathbb{F}[t]$. Entonces*

$$M \cong \left(\bigoplus_{i=1}^n t^{a_i} \cdot \mathbb{F}[t] \right) \oplus \left(\bigoplus_{j=1}^m t^{b_j} \cdot \frac{\mathbb{F}[t]}{(t^{c_j})} \right),$$

donde $n, m \geq 1$ y a_i, b_j y c_j son potencias enteras no negativas de t .

Demostración. El módulo $M = \bigoplus_i M^i$ es un $\mathbb{F}[t]$ -módulo finitamente generado por ser \mathcal{M} un módulo de persistencia de tipo finito. Puesto que \mathbb{F} es un cuerpo, el anillo de polinomios $\mathbb{F}[t]$ es un dominio de ideales principales. Por tanto, $M = \bigoplus_i M^i$ es un módulo graduado

finitamente generado sobre un dominio de ideales principales graduado y el Teorema 1.22 permite descomponerlo en suma directa de su parte libre y su parte de torsión.

La componente libre está formada entonces por una suma directa de un cierto número de copias del anillo graduado $\mathbb{F}[t]$. La parte de torsión será una suma directa de un cierto número de cocientes del anillo graduado $\mathbb{F}[t]$ por ideales de este anillo, que son todos homogéneos de la forma (t^s) .

Por último, el cambio en la graduación tanto en la parte libre como en la parte de torsión se debe a que el módulo M es graduado y, entonces, el isomorfismo de módulos graduados debe ser un isomorfismo de grado 0. Así, si z es un generador homogéneo de grado $\alpha > 0$ de la parte libre de M , entonces se hace necesario aumentar en α el grado de $\mathbb{F}[t]$ (mediante la acción de t^α) para que el isomorfismo sea de grado 0. Lo mismo ocurre con la parte de torsión de M . Por tanto, el Teorema 1.22 adopta la forma del enunciado en este contexto. \square

En particular, la homología persistente en un grado fijado q de un complejo simplicial \mathbb{N} -filtrado con coeficientes en un cuerpo \mathbb{F} es un módulo de persistencia de tipo finito parametrizado por \mathbb{N} sobre $\mathbb{F}[t]$, luego admite una descomposición como la anterior. Este resultado tiene una interpretación inmediata en términos de los códigos de barras. Las potencias de la variable t capturan la aparición y desaparición de características topológicas a lo largo de la filtración.

- En primer lugar, la parte libre se encuentra en correspondencia biyectiva con los generadores de la homología que nacen en el grado filtrante a_i y que sobreviven hasta el final de la filtración.
- Por su parte, los elementos de torsión se corresponden con los generadores de la homología que surgen en el grado filtrante b_j y que se vuelven triviales al alcanzar el grado filtrante $b_j + c_j$.

Observación 3.24. Mediante la teoría de Artin-Rees de álgebra conmutativa se puede probar [41] que existe una correspondencia biyectiva entre la categoría de módulos de persistencia de tipo finito parametrizados por \mathbb{N} sobre R y la categoría de módulos graduados (con graduación no negativa) finitamente generados sobre $R[t]$. Esto también indica que una clasificación simple no es posible si el anillo de coeficientes considerado no es un cuerpo. Por ejemplo, la clasificación de módulos sobre $\mathbb{Z}[t]$ es muy complicada, aunque se pueden asignar invariantes de interés a los $\mathbb{Z}[t]$ -módulos.

Anteriormente hemos trabajado con filtraciones con índices en $\overline{\mathbb{R}}$, pues son las que surgen de modo natural al considerar las filtraciones dadas por los complejos de Čech y Vietoris-Rips, por ejemplo. También es posible hacer una clasificación de los módulos de persistencia de tipo finito como $\mathbb{F}[t]$ -módulos cuando están parametrizados por $\overline{\mathbb{R}}$.

En este caso, se introducen los conceptos de módulos de intervalos y el módulo de persistencia se caracteriza mediante ellos, como en [8]. Al igual que con el Teorema 3.23, mediante esta descripción se relaciona el módulo de persistencia con su código de barras asociado, por lo que quedan completamente determinados de ambas formas.

Sin embargo, según describe G. Carlsson en [5], podemos utilizar cualquier aplicación $\mathbb{N} \rightarrow \mathbb{R}$ que preserve el orden para obtener un módulo de persistencia parametrizado por \mathbb{N} a partir de otro parametrizado por $\overline{\mathbb{R}}$.

Una primera opción es fijar $\varepsilon \in \mathbb{R}$ y considerar la aplicación $f_\varepsilon: \mathbb{N} \rightarrow \mathbb{R}$ dada por $f_\varepsilon(n) = n\varepsilon$. La monotonía permite tener un único antecedente para cada índice real de la filtración de la forma $n\varepsilon$, el cual se corresponde con el índice n de la \mathbb{N} -filtración. Nótese que cuanto más pequeño (en valor absoluto) sea el valor de ε , más fina será la \mathbb{N} -filtración que se obtenga mediante este método.

Otra de las posibilidades hace uso de que el complejo simplicial es finito y, por tanto, únicamente habrá un número finito de valores reales en los que se produzca una transición en los complejos intermedios debido a la adición de un nuevo símplex. Si estos valores de transición forman el conjunto $\{r_1, \dots, r_N\}$, definimos la aplicación $g: \mathbb{N} \rightarrow \mathbb{R}$ dada por $g(n) = r_n$ para $n \leq N$ y $g(n) = r_N$ para $n \geq N$. La ventaja de este método es que está adaptado al complejo en cuestión, pues observa precisamente los momentos en los que la estructura topológica del complejo simplicial sufre modificaciones.

3.3.3. Algoritmo de cálculo sobre cuerpos

A. Zomorodian y G. Carlson introducen en [41] un algoritmo de cálculo de la homología persistente de un complejo simplicial que aprovecha la estructura de $\mathbb{F}[t]$ -módulo que esta tiene como módulo de persistencia parametrizado por \mathbb{N} . Las principales características del algoritmo son que deriva directamente del algoritmo de reducción de matrices a su forma escalonada, permite obtener el código de barras y, mediante un refinamiento que presentan en el artículo, no se hace necesario calcular explícitamente el módulo de persistencia.

Para desarrollar el algoritmo tendremos que utilizar bases homogéneas, pues veremos que respetan propiedades fundamentales. Las bases homogéneas son aquellas bases formadas por elementos homogéneos de la estructura graduada.

Definición 3.25. Sea $R = \bigoplus_i R^i$ un anillo graduado. Llamamos *elemento homogéneo de grado i* a un elemento que pertenezca a un único R^i de la descomposición. Además, decimos que una base de R es una *base homogénea* si está formada por elementos homogéneos de R .

Denotamos por $\deg(r_i) = i$ el grado de un elemento homogéneo $r_i \in R^i$ del anillo $R = \bigoplus_i R^i$. Las representaciones matriciales del operador borde con respecto a bases homogéneas verifican una igualdad básica que usaremos más adelante para calcular la homología persistente de la \mathbb{N} -filtración de un complejo simplicial.

Proposición 3.26. Sea \mathcal{C} un complejo de persistencia de tipo finito parametrizado por \mathbb{N} sobre un cuerpo \mathbb{F} . Supongamos que $\{e_j\}$ y $\{e'_i\}$ son bases homogéneas para C_q y C_{q-1} , respectivamente. Los elementos de la matriz $M(\partial_q)$ del operador borde con respecto a esas bases son homogéneos y verifican que

$$\deg(e_j) = \deg(M_{ij}) + \deg(e'_i),$$

donde M_{ij} denota el elemento que ocupa la posición (i, j) en $M(\partial_q)$.

Demostración. Como $\partial_q: C_q \rightarrow C_{q-1}$ es un homomorfismo de R -módulos, al considerar la \mathbb{N} -filtración en el complejo de persistencia \mathcal{C} , se tiene que es un homomorfismo de grado 0. Por tanto, fijadas las bases homogéneas $\{e_j\}$ y $\{e'_i\}$, el elemento M_{ij} es también homogéneo al estar obligado a ser una potencia de t que aumente el grado de e'_i exactamente lo que le falta para alcanzar el grado de e_j , para que así ∂_q sea un homomorfismo de grado 0. \square

El algoritmo de cálculo se basará en la eliminación gaussiana por columnas e involucrará dos de sus tres operaciones elementales: intercambiar dos columnas entre sí (tipo 1), multiplicar una columna por una unidad del anillo (tipo 2) y sumar a una columna otra columna multiplicada por un elemento homogéneo (tipo 3). Conviene resaltar que la operación de tipo 3 se hace multiplicando por un elemento homogéneo del anillo. Ello es clave para que el algoritmo aproveche la invarianza en la homogeneidad de los elementos de la matriz y de las bases mediante estas operaciones de tipo 3. Enunciamos y probamos el resultado para su posterior uso.

Corolario 3.27. *Sea \mathcal{C} un complejo de persistencia de tipo finito parametrizado por \mathbb{N} sobre un cuerpo \mathbb{F} . Supongamos que $\{e_j\}$ y $\{e'_i\}$ son bases homogéneas para C_q y C_{q-1} , respectivamente. Sea $M(\partial_q)$ la representación matricial con respecto a esas bases. Entonces, se tiene que:*

1. *La homogeneidad de la base $\{e_j\}$ y de los elementos de la matriz $M(\partial_q)$ no varía mediante operaciones elementales de tipo 3 por columnas.*
2. *La homogeneidad de la base $\{e'_i\}$ no varía mediante operaciones elementales de tipo 3 por filas.*

Demostración. Probamos primero el resultado por columnas. Supongamos que $\lambda \in \mathbb{F}[t]$ es un elemento homogéneo del anillo tal que al sustituir la j -ésima columna de $M(\partial_q)$ por

$$(\text{columna } j) + \lambda(\text{columna } i)$$

se anula el elemento M_{kj} con ayuda del pivote M_{ki} . Esta operación cambia el elemento de la base de las columnas e_j por $e_j + \lambda e_i$. Además, como tenemos $M_{kj} + \lambda M_{ki} = 0$ y λ lo hemos tomado homogéneo, necesariamente

$$\deg(M_{kj}) = \deg(\lambda M_{ki}) = \deg(\lambda) + \deg(M_{ki}).$$

Luego, $\deg(\lambda) = \deg(M_{kj}) - \deg(M_{ki})$.

Ahora bien,

$$\begin{aligned} \deg(\lambda e_i) &= \deg(\lambda) + \deg(e_i) \\ &= (\deg(M_{kj}) - \deg(M_{ki})) + \deg(e_i) \\ &= (\deg(e_j) - \deg(e'_k)) - (\deg(e_i) - \deg(e'_k)) + \deg(e_i) \\ &= \deg(e_j) - \deg(e'_k) - \deg(e_i) + \deg(e'_k) + \deg(e_i) \\ &= \deg(e_j), \end{aligned}$$

donde se ha usado la Proposición 3.26 en la segunda igualdad. Por tanto, $e_j + \lambda e_i$ es un elemento homogéneo de la nueva base por columnas y, entonces, sigue siendo una base homogénea.

Por otra parte, los nuevos elementos $M_{lj} + \lambda M_{li}$ de la j -ésima columna siguen siendo todos homogéneos tras la operación elemental de tipo 3. En efecto,

$$\begin{aligned}
 \deg(\lambda M_{li}) &= \deg(\lambda) + \deg(M_{li}) \\
 &= (\deg(M_{kj}) - \deg(M_{ki})) + \deg(M_{li}) \\
 &= (\deg(e_j) - \deg(e'_k)) - (\deg(e_i) - \deg(e'_k)) \\
 &\quad + (\deg(e_i) - \deg(e'_l)) \\
 &= \deg(e_j) - \deg(e'_l) \\
 &= \deg(M_{lj}),
 \end{aligned}$$

donde se ha usado la Proposición 3.26 en la segunda igualdad y en la cuarta. Por tanto, $M_{lj} + \lambda M_{li}$ es un elemento homogéneo para todo l y, entonces, los elementos de la j -ésima columna siguen siendo homogéneos tras una operación elemental de tipo 3.

La prueba para el caso por filas es similar. En esta situación, pongamos nuevamente que utilizamos un $\lambda \in \mathbb{F}[t]$ homogéneo y sustituimos la j -ésima fila por

$$(\text{fila } j) + \lambda(\text{fila } i)$$

para eliminar el elemento M_{jk} con ayuda del elemento pivote M_{ik} . Esta operación cambia el elemento de la base por filas e'_j por $e'_i - \lambda e'_j$. Además, como tenemos que $M_{jk} + \lambda M_{ik} = 0$ y λ es homogéneo, $\deg(M_{jk}) = \deg(\lambda) + \deg(M_{ik})$. De esta forma,

$$\begin{aligned}
 \deg(\lambda e'_j) &= \deg(\lambda) + \deg(e'_j) \\
 &= (\deg(M_{jk}) - \deg(M_{ik})) + \deg(e'_j) \\
 &= \deg(e_k) - \deg(M_{ik}) \\
 &= \deg(e_k) - (\deg(e_k) - \deg(e'_i)) \\
 &= \deg(e'_i),
 \end{aligned}$$

donde se ha usado la Proposición 3.26 en la segunda igualdad y en la tercera. Por tanto, $e'_i - \lambda e'_j$ es un elemento homogéneo de la nueva base por filas y, entonces, sigue siendo una base homogénea. \square

Como anunciamos previamente, que se conserve la homogeneidad tanto en las bases como en la matriz resulta ser fundamental para poder calcular inductivamente la homología persistente. Así que nos interesan representaciones matriciales en bases homogéneas.

Es obvio que la base del módulo de cadenas libre C_q que viene dada por la colección de q -símplices orientados es homogénea, pues cada q -símplice entra en un determinado grado filtrante de cualquier \mathbb{N} -filtración del complejo simplicial.

Definición 3.28. Sean \mathcal{K} un complejo simplicial \mathbb{N} -filtrado y \mathcal{C} su correspondiente complejo de persistencia asociado. Llamamos *base estándar* de C_q a la base homogénea dada por la colección de q -símplices orientados de \mathcal{K} .

Las bases estándar de C_q y C_{q-1} sirven para escribir inicialmente la matriz $M(\partial_q)$ y, a partir de esta representación, elaborar un algoritmo que parte de las mismas ideas que el algoritmo de cálculo de la homología simplicial. Sin embargo, veremos que no es necesario calcular la forma normal de Smith de $M(\partial_q)$ como hacíamos en ese caso. En su lugar, bastará con calcular la forma escalonada por columnas de la matriz.

Como primera aproximación, vamos a reinterpretar, cuando el anillo de coeficientes es un dominio de ideales principales graduado, el algoritmo que describimos para calcular la homología simplicial en el apartado 1.5.2, olvidando momentáneamente la estructura de complejo de persistencia y volviendo a los complejos de cadenas.

En primer lugar, representamos la matriz $M(\partial_q)$ con respecto a la base estándar de C_q y una base homogénea de Z_{q-1} (en vez de la base estándar homogénea de C_{q-1} , pues $\text{Im}(\partial_q) = B_{q-1} \subset Z_{q-1}$). A continuación, reducimos esta matriz a su forma normal de Smith $\tilde{M}(\partial_q)$ y suponemos que $\{e'_i\}$ es la nueva base de Z_{q-1} que proporciona esta representación matricial en forma normal.

Proposición 3.29. *Sea $(C_\bullet, \partial_\bullet)$ un complejo de cadenas finito sobre un dominio de ideales principales graduado R . Supongamos que $\tilde{M}(\partial_q)$ es la forma normal de Smith de $M(\partial_q)$ con respecto a una base homogénea $\{e'_i\}$ de Z_{q-1} . Se verifica que:*

- (a) *Si la fila i de $\tilde{M}(\partial_q)$ es de ceros, entonces contribuye a H_{q-1} con un término libre que sufre un aumento de $\deg(e'_i)$ en su graduación.*
- (b) *Si la fila i de $\tilde{M}(\partial_q)$ tiene un término en la diagonal, entonces contribuye a H_{q-1} con un término de torsión homogéneo (él mismo) que sufre un aumento de $\deg(e'_i)$ en su graduación.*

Demostración. El Lema 1.27 adaptado al caso de estructuras graduadas permite afirmar los resultados, sin más que reinterpretar la descripción de H_{q-1} dada por el Teorema 1.22 (en términos de un dominio de ideales principales graduado, de forma similar a lo hecho en la demostración del Teorema 3.23) usando la nueva base $\{e'_i\}$ de Z_{q-1} . \square

Describimos a continuación cómo actúa el nuevo algoritmo.

Teorema 3.30. *Sea \mathcal{C} un complejo de persistencia de tipo finito parametrizado por \mathbb{N} sobre un cuerpo \mathbb{F} . Entonces, el homomorfismo $\partial_q: C_q \rightarrow C_{q-1}$ se puede representar en forma matricial homogénea con respecto a la base estándar de C_q y una base homogénea de Z_{q-1} .*

Demostración. La prueba se hace por inducción en q . El caso inicial es trivial. Como $\partial_0 = 0$, se tiene que $Z_0 = C_0$ y la base estándar dada por los vértices (0-símplices) es homogénea para Z_0 y se puede usar para representar ∂_1 junto con la base de C_1 dada por los 1-símplices orientados.

Supongamos ahora que tenemos una representación matricial de $M(\partial_q)$ con respecto a la base estándar de C_q y una base homogénea $\{e'_i\}$ de Z_{q-1} . Tenemos que encontrar una base homogénea de Z_q y representar la matriz $M(\partial_{q+1})$ en esa base y la estándar de C_{q+1} para completar la inducción.

En primer lugar, ordenamos los elementos de la base $\{e'_i\}$ en orden decreciente de grados. A continuación, mediante eliminación gaussiana por columnas se convierte la matriz $M(\partial_q)$ en su forma escalonada por columnas, usando únicamente operaciones elementales de tipos 1 y 3. El resultado es una matriz con escalones de distintas alturas, pivotes en columnas consecutivas y en la que únicamente hay elementos no nulos bajo los pivotes.

$$\begin{bmatrix} \boxed{*} & 0 & & \cdots & 0 \\ & \boxed{*} & 0 & & \cdots \\ * & * & 0 & & \vdots \\ & * & \boxed{*} & 0 & \cdots \\ * & & & 0 & \cdots & 0 \end{bmatrix}$$

Figura 3.3: Ejemplo de una matriz escalonada por columnas. Los asteriscos indican elementos no nulos y los pivotes están dentro de un recuadro.

Empezamos por la primera columna a la izquierda. Si el primer elemento de la columna es cero, intercambiamos la columna con la primera que tenga un elemento no nulo mediante una operación de tipo 1. La fila se convierte entonces en una fila pivote y se pueden eliminar las entradas no nulas de esta fila pivote con operaciones de tipo 3. El algoritmo continúa actuando de la misma forma en la submatriz resultante de eliminar la primera fila y la primera columna, bajando en cada paso una fila y avanzando por las sucesivas columnas hacia la derecha hasta alcanzar o bien una columna de ceros o bien una columna con un elemento que vuelva a servir de pivote para eliminar los elementos de su fila. Finalmente, si hiciera falta, mediante intercambios de columnas se recolocan las columnas de forma adecuada.

Por el Corolario 3.27, la operación elemental de tipo 3, consistente en sumar a una columna otra columna multiplicada por un elemento homogéneo, no varía la homogeneidad de los elementos de la base de C_q y tampoco hace que se pierda la homogeneidad de los elementos de la matriz.

Además, por álgebra lineal sabemos que el número de pivotes de la forma escalonada es igual a

$$\text{rango}(M(\partial_q)) = \text{rango}(\text{Im}(\partial_q)) = \text{rango}(B_{q-1}).$$

Afirmamos que los elementos de la base correspondientes a las columnas sin pivote de la forma escalonada $\tilde{M}(\partial_q)$ conforman la base homogénea de Z_q que buscábamos para completar la inducción. Los próximos dos lemas sirven para probar esta última afirmación, que finaliza la demostración del teorema, y sirven para acabar la descripción del algoritmo de cálculo. \square

Lema 3.31. *Supongamos que $\{e_j\}$ y $\{e'_i\}$ son bases homogéneas para C_q y C_{q-1} , respectivamente. Entonces, los pivotes de la forma escalonada por columnas de la matriz $M(\partial_q)$ coinciden con los elementos de la diagonal en la forma normal de Smith. Además, el grado de los elementos de la base en las filas con pivote coincide en ambos casos.*

Demostración. El orden decreciente de grados que tomamos para la base por filas $\{e'_i\}$ en la demostración del teorema anterior provoca que el grado de los elementos de esta base decrezca monótonamente según se desciende en las filas de la matriz. Además, dentro de la columna j , el grado $\deg(e_j)$ es una constante c . Así, fijado j , la Proposición 3.26 nos dice que, para todo i ,

$$\deg(M_{ij}) = c - \deg(e'_i).$$

De esta forma, el grado de los elementos de la matriz dentro de cada columna crece monótonamente al bajar en las filas. Esto permite eliminar los elementos no nulos que están por debajo de los pivotes mediante operaciones de tipo 3 por filas que no cambian los pivotes ni los grados de los elementos de la base por filas, por el Corolario 3.27. Finalmente, mediante intercambios de filas y columnas se recoloca la matriz en su forma diagonal si fuese necesario. \square

Lo que resta de la prueba del Teorema 3.30 se sigue de que $\partial_q \circ \partial_{q+1} = 0$ y, por tanto, $M(\partial_q) \cdot M(\partial_{q+1}) = 0$. Esta relación no cambia mediante operaciones elementales, que únicamente generan matrices intermedias de cambio de base. Además, como el dominio de ∂_q es el codominio de ∂_{q+1} , las operaciones elementales hechas por columnas en $M(\partial_q)$ para conseguir la forma escalonada por columnas $\tilde{M}(\partial_q)$ se corresponden con operaciones hechas por filas en $M(\partial_{q+1})$. Estas operaciones por filas anulan aquellas filas de $M(\partial_{q+1})$ que se corresponden con columnas pivote no nulas de $\tilde{M}(\partial_q)$ y proporcionan el cambio de base que da la representación de ∂_{q+1} que buscábamos para completar la inducción.

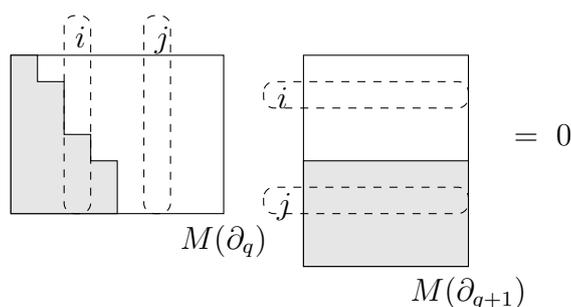


Figura 3.4: La relación $M(\partial_q) \cdot M(\partial_{q+1}) = 0$ queda inalterada por operaciones elementales. Al reducir $M(\partial_q)$ a su forma escalonada por columnas $\tilde{M}(\partial_q)$, las respectivas operaciones por filas en $M(\partial_{q+1})$ anulan las filas que se corresponden con columnas pivote de $\tilde{M}(\partial_q)$.

Lema 3.32. *Para representar el homomorfismo ∂_{q+1} con respecto a la base estándar de C_{q+1} y la base calculada en el Teorema 3.30 para Z_q basta con eliminar las filas de $M(\partial_{q+1})$ que se corresponden con columnas pivote de $\tilde{M}(\partial_q)$.*

Demostración. Únicamente hemos usado operaciones elementales de tipo 1 y tipo 3. De ellas, las de tipo 3 son las únicas que cambian valores en la matriz. Supongamos que en $M(\partial_q)$ sustituimos la j -ésima columna por

$$(\text{columna } j) + \lambda(\text{columna } i),$$

con $\lambda \in \mathbb{F}[t]$ homogéneo, para eliminar un elemento en la fila pivote i . Esta operación cambia el elemento de la base de las columnas e_j por $e_j + \lambda e_i$ en $M(\partial_q)$. Para conseguir el mismo cambio de base en las filas de $M(\partial_{q+1})$ debemos sustituir la i -ésima fila por

$$(\text{fila } i) - \lambda(\text{fila } j).$$

Pero las sucesivas operaciones elementales por columnas que convierten la i -ésima columna de $\tilde{M}(\partial_q)$ en columna pivote acaban provocando que la i -ésima fila de $\tilde{M}(\partial_{q+1})$ sea nula y, por otra parte, la j -ésima fila no sufre cambios por estas operaciones en ningún momento.

Por tanto, no necesitamos realizar operaciones por filas. Es suficiente con eliminar las filas correspondientes a columnas pivote de una dimensión inferior para obtener la representación de ∂_{q+1} en términos de la base estándar de C_{q+1} y la base homogénea de Z_q . \square

La validez del algoritmo queda probada con el último resultado, que completa la inducción del Teorema 3.30.

El Lema 3.31 establece que si lo único que nos interesa son los grados de los elementos de la base, podemos obtenerlos directamente de la forma escalonada por columnas. Además, permite reformular el Teorema 3.23 en términos de la matriz escalonada por columnas. En consecuencia, esta matriz es suficiente para calcular los códigos de barras.

Corolario 3.33. *Sea \mathcal{C} un complejo de persistencia de tipo finito parametrizado por \mathbb{N} sobre un cuerpo \mathbb{F} . Sea $\tilde{M}(\partial_q)$ la matriz en forma escalonada por columnas correspondiente al homomorfismo ∂_q con respecto a la base estándar $\{e_j\}$ de C_q y la base homogénea $\{e'_i\}$ de Z_{q-1} . Se verifica que:*

- (a) *Si la fila i de $\tilde{M}(\partial_q)$ tiene un pivote $\tilde{M}_{ij} = t^s$, entonces contribuye a H_{q-1} con un término*

$$t^{\deg(e'_i)} \cdot \frac{\mathbb{F}[t]}{(t^s)}.$$

- (b) *Si la fila i de $\tilde{M}(\partial_q)$ no tiene un pivote, entonces contribuye a H_{q-1} con un término*

$$t^{\deg(e'_i)} \cdot \mathbb{F}[t].$$

Equivalentemente, en el código de barras asociado a H_{q-1} obtenemos un intervalo con extremos $\deg(e'_i)$ y $\deg(e'_i) + s$ o un intervalo no acotado con extremo inferior $\deg(e'_i)$, respectivamente.

Terminamos la sección con un ejemplo de cálculo.

Ejemplo 3.34. Calculemos la homología persistente de la \mathbb{N} -filtración de la Figura 3.1 mediante el algoritmo que acabamos de describir. Usaremos coeficientes sobre el cuerpo \mathbb{R} , luego el módulo de persistencia será un $\mathbb{R}[t]$ -módulo.

Los simplices de la filtración tienen los siguientes grados:

| v_0 | v_2 | v_1 | v_3 | v_4 | v_0v_2 | v_0v_1 | v_3v_1 | v_2v_3 | v_4v_2 | v_2v_1 | $v_0v_1v_2$ |
|-------|-------|-------|-------|-------|----------|----------|----------|----------|----------|----------|-------------|
| 1 | 1 | 2 | 2 | 2 | 3 | 4 | 4 | 4 | 5 | 6 | 7 |

La representación matricial del operador borde $\partial_1: C_1 \rightarrow C_0$, con la base de C_0 en orden decreciente de grados, viene dada por

$$M(\partial_1) = \begin{bmatrix} & v_0v_2 & v_0v_1 & v_3v_1 & v_2v_3 & v_4v_2 & v_2v_1 \\ v_4 & 0 & 0 & 0 & 0 & t^3 & 0 \\ v_3 & 0 & 0 & t^2 & t^2 & 0 & 0 \\ v_1 & 0 & t^2 & t^2 & 0 & 0 & t^4 \\ v_2 & t^2 & 0 & 0 & t^3 & t^4 & t^5 \\ v_0 & t^2 & t^3 & 0 & 0 & 0 & 0 \end{bmatrix},$$

mientras que su forma escalonada por columnas adopta la forma

$$\tilde{M}(\partial_1) = \begin{bmatrix} & v_4v_2 & v_3v_1 & v_0v_1 & v_0v_2 & z_1 & z_2 \\ v_4 & \boxed{t^3} & 0 & 0 & 0 & 0 & 0 \\ v_3 & 0 & \boxed{t^2} & 0 & 0 & 0 & 0 \\ v_1 & 0 & t^2 & \boxed{t^2} & 0 & 0 & 0 \\ v_2 & t^4 & 0 & 0 & \boxed{t^2} & 0 & 0 \\ v_0 & 0 & 0 & t^3 & t^2 & 0 & 0 \end{bmatrix},$$

donde

$$z_1 = v_2v_3 - v_3v_1 + v_0v_1 - t \cdot v_0v_2$$

y

$$z_2 = v_2v_1 - t^2 \cdot v_0v_1 - t^3 \cdot v_0v_2$$

forman una base homogénea de Z_1 .

Con esta información, el Corolario 3.33 permite concluir que el código de barras correspondiente a H_0 incluye los siguientes intervalos:

$$\begin{aligned} [\deg(v_4), \deg(v_4) + 3] &= [2, 5), \\ [\deg(v_3), \deg(v_3) + 2] &= [2, 4), \\ [\deg(v_1), \deg(v_1) + 2] &= [2, 4), \\ [\deg(v_2), \deg(v_2) + 2] &= [1, 3), \\ [\deg(v_0), +\infty] &= [1, +\infty). \end{aligned}$$

Por su parte, la matriz asociada a $\partial_2: C_2 \rightarrow C_1$ y su forma escalonada por columnas vienen ambas dadas por

$$M(\partial_2) = \left[\begin{array}{c|c} & v_0v_1v_2 \\ \hline v_2v_1 & t \\ v_4v_2 & 0 \\ v_2v_3 & 0 \\ v_3v_1 & 0 \\ v_0v_1 & t^3 \\ v_0v_2 & t^4 \end{array} \right].$$

Para obtener la representación en términos de la base estándar de C_2 y la base ordenada $\{z_1, z_2\}$ de Z_1 que obtuvimos previamente, basta con eliminar la segunda y las tres últimas filas, que son las asociadas a los pivotes de $\tilde{M}(\partial_1)$. Luego tenemos

$$\tilde{M}(\partial_2) = \left[\begin{array}{c|c} & v_0v_1v_2 \\ \hline z_2 & t \\ z_1 & 0 \end{array} \right],$$

donde también hemos sustituido los elementos v_2v_3 y v_2v_1 de la base por

$$z_1 = v_2v_3 - v_3v_1 + v_0v_1 - t \cdot v_0v_2$$

y

$$z_2 = v_2v_1 - t^2 \cdot v_0v_1 - t^3 \cdot v_0v_2.$$

De esta forma, obtenemos los dos siguientes intervalos de persistencia para el código de barras asociado a H_1 :

$$\begin{aligned} [\deg(z_2), \deg(z_2) + 1] &= [6, 7), \\ [\deg(z_1), +\infty) &= [4, +\infty). \end{aligned}$$

Los cálculos de los intervalos de persistencia anteriores confirman que el código de barras que mostramos intuitivamente en la Figura 3.2 para la filtración de la Figura 3.1 es correcto.

3.4. Notas históricas

Como cuentan H. Edelsbrunner y J. Harer en [18], la homología persistente, como muchos otros conceptos matemáticos, surgió de forma independiente en tres trabajos distintos. Los primeros descubrimientos fueron de forma casi simultánea en los últimos años del siglo XX. En primer lugar, algunos trabajos de P. Frosini y M. Ferri contienen la persistencia de la homología en grado cero. Por otra parte, la tesis doctoral de V. Robins incluye una definición de homología persistente para estudiar conjuntos fractales con formas alfa. Por último, H. Edelsbrunner introdujo, entre otras, una de las herramientas básicas en la teoría de persistencia: las filtraciones en complejos simpliciales.

La corta existencia del concepto de homología persistente provoca que la definición aquí presentada no se corresponda con las iniciales. Por ello, es fácil encontrar literatura sobre homología persistente en la que los conceptos no coinciden exactamente debido al escaso desarrollo de la teoría y las necesidades de refinarla.

En cuanto a los métodos de cálculo, existe un algoritmo primigenio de emparejamiento persistente de símplexes. A. Zomorodian y G. Carlsson son los primeros en dar el algoritmo de cálculo efectivo de intervalos de persistencia descrito anteriormente, e incluso consiguen un algoritmo que no usa las matrices. Ello se debe a observaciones hechas a partir de los dos lemas probados en la sección precedente y la no necesidad de realizar operaciones por filas para el cálculo inductivo de las matrices. De esta manera, el algoritmo se ejecuta directamente sobre el cuerpo \mathbb{F} y se obtiene directamente el código de barras sin calcular explícitamente el $\mathbb{F}[t]$ -módulo de persistencia.

Capítulo 4

Análisis de redes complejas

Un gran número de sistemas en ciencias naturales y sociales se modela gracias a redes. La complejidad de estas redes se pone de manifiesto tanto en su estructura y características topológicas como en su dinámica evolutiva. La visión clásica de las redes se hacía a través de la teoría de grafos. No obstante, este punto de vista tiene limitaciones en cuanto a sus aplicaciones en modelos reales. La aparición de nuevos sistemas de redes, más allá de los grafos regulares y aleatorios, ha propiciado, junto a algunos métodos estadísticos, la utilización de técnicas novedosas como la homología persistente.

Siguiendo algunos ejemplos dados en el artículo [27] de Danijela Horak, Slobodan Matetić y Milan Rajković y el principal ejemplo del artículo [9] de C. J. Carstens y K. J. Horadam, el presente capítulo muestra algunas aplicaciones prácticas de la homología persistente al análisis de redes complejas. El capítulo comienza con unos preliminares básicos que introducen la terminología de teoría de grafos que utilizaremos. Posteriormente se explica el uso que se hace de la homología persistente para examinar distintos tipos de redes.

Con el fin de estudiar en profundidad una red es habitual observar su dinámica evolutiva para conocer el comportamiento que muestra a lo largo del tiempo. Esto motiva la búsqueda de aquellas características topológicas persistentes en la red. Uno de los objetivos primordiales es identificar las propiedades que hacen a las redes menos vulnerables frente a los ataques y distinguirlas de las propiedades que debilitan la red. En este sentido, los análisis mediante homología persistente de redes complejas con distintas distribuciones de grado reflejan la robustez de la red.

4.1. Teoría de grafos

En esta sección revisamos los conceptos básicos de teoría de grafos que necesitaremos en el resto del capítulo. En primer lugar, se repasan rápidamente las definiciones elementales. A continuación, se introducen algunos fundamentos de las distribuciones de grado, donde se combina la teoría de grafos con la de probabilidades. Ello es necesario para el posterior

estudio que realizaremos mediante homología persistente. Las definiciones están tomadas en su mayoría de [20].

4.1.1. Conceptos básicos

Informalmente, un grafo es un conjunto de puntos unidos de cierta forma mediante líneas entre ellos que crean conexiones. A continuación, detallamos una definición más formal.

Definición 4.1. Un *grafo simple* G es un par ordenado (V, E) , donde V es un conjunto no vacío y numerable y $E \subset \{\{u, v\} : u, v \in V\}$.

V se denomina conjunto de *vértices* o *nodos* y E se llama conjunto de *aristas*, *conexiones* o *enlaces*. Una arista se representa como un par no ordenado $\{u, v\}$ de vértices, con $u, v \in V$. En ese caso, los vértices se dicen *vecinos* o *adyacentes* y la arista se dice *incidente* con los vértices.

Más adelante, usaremos a menudo el término *red* para denominar los grafos simples. También obviaremos el adjetivo «simple», ya que no trataremos el caso de grafos con aristas múltiples.

Definición 4.2. Un *grafo con pesos* es un grafo $G = (V, E)$ junto con una función $\omega : E \rightarrow \mathbb{R}$ que asigna un valor real a cada arista.

Definición 4.3. Sea $G = (V, E)$ un grafo. Se dice que $G' = (V', E')$ es un *subgrafo* de G si se verifica:

- (I) $V' \subset V$.
- (II) $E' \subset \{\{u, v\} : u, v \in V'\} \cap E$.

Definición 4.4. Para cada $n \geq 1$, llamamos *grafo completo* a un grafo K_n formado por n vértices y en el que dos nodos distintos siempre están unidos por una arista.

Para trabajar cómodamente con grafos cuyos vértices no están etiquetados, resulta útil asignar unas etiquetas genéricas. Por ejemplo, un grafo con n vértices podemos etiquetarlo asignando a cada nodo un elemento del conjunto $\{1, 2, \dots, n\}$. Mediante este etiquetado, es posible recoger fácilmente la información básica de un grafo en una matriz.

Definición 4.5. Sea $G = (V, E)$ un grafo simple, con $V = \{1, 2, \dots, n\}$. Definimos la *matriz de adyacencia* de G como la matriz cuadrada $A = (a_{ij})$, donde para $1 \leq i, j \leq n$ hacemos

$$a_{ij} = \begin{cases} 1, & \{i, j\} \in E \\ 0, & \{i, j\} \notin E. \end{cases}$$

Definición 4.6. Sea $G = (V, E)$ un grafo. Un *camino* es una sucesión finita de aristas (no necesariamente distintas)

$$\{u_1, v_1\}, \{u_2, v_2\}, \dots, \{u_k, v_k\}$$

para las que $v_i = u_{i+1}$, con $i = 1, \dots, k - 1$. Si $v_k = u_1$, el camino se dice *cerrado* o un *lazo*. Se llama *longitud* del camino al número de aristas k . Si los vértices de un camino son distintos dos a dos, se dice que el camino es *simple*. Si las aristas de un camino son distintas dos a dos, el camino se llama *recorrido*. Un recorrido cerrado es un *círculo*.

Los caminos permiten establecer el concepto de conexión en un grafo.

Definición 4.7. Sea $G = (V, E)$ un grafo. Decimos que G es *conexo* si cada par de vértices de G se puede unir mediante un camino de aristas.

La conectividad está relacionada con la *robustez* o tolerancia del grafo a mantener intactas sus propiedades topológicas ante la adición o eliminación de vértices.

4.1.2. Complejo clique

El siguiente paso es asociar un complejo simplicial al grafo para poder estudiar su homología persistente a través de una filtración. Aunque podríamos considerar los grafos como complejos simpliciales de dimensión uno y trabajar directamente con una filtración del grafo, esto solo proporcionaría números de Betti no triviales de dimensiones cero y uno. Gracias a los complejos simpliciales podemos codificar información topológica de los grafos en otras dimensiones; en nuestro caso, utilizaremos una construcción que se denomina complejo clique.

Definición 4.8. Sea $G = (V, E)$ un grafo. Un *q-clique* es un subgrafo completo de G con q vértices.

Definición 4.9. Sea $G = (V, E)$ un grafo. El *complejo clique* asociado a G se construye introduciendo un q -símplice en el complejo si $q + 1$ vértices forman un $(q + 1)$ -clique.

Así, un 3-clique se convierte en un triángulo o 2-símplice y un 4-clique en un tetraedro sólido o 3-símplice, por ejemplo.

Pero el complejo clique asociado a un grafo no es nada nuevo, ya que resulta ser un cierto complejo de Vietoris-Rips de parámetro $\varepsilon = 1$. Por concretar, los puntos del espacio métrico son los vértices del grafo y la función distancia está dada por la longitud del camino más corto entre cada par de vértices.

Una propiedad del complejo clique es que los cliques se corresponden con agrupamientos de nodos altamente conectados que representan comunidades. Por ejemplo, el primer número de Betti del complejo clique representa la cantidad de ciclos del complejo. En el grafo original, cada triángulo es un ciclo que aumenta β_1 en una unidad. 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 tienen 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.

4.1.3. Distribuciones de grado

El estudio de la distribución de grado sirve para conocer el comportamiento de las redes complejas. En primer lugar, establecemos lo que entendemos por el grado de un vértice.

Definición 4.10. Sea $G = (V, E)$ un grafo. El *grado* de un nodo $v \in V$, escrito $\deg(v)$, se define como el número de aristas que son incidentes con v .

La idea es ver cómo se distribuyen los grados de los nodos a lo largo de la red. Basta con inspeccionar la matriz de adyacencia para comprobar las diferencias existentes entre los grados de los distintos nodos de la red.

Definición 4.11. Sea $G = (V, E)$ un grafo de n vértices. Supongamos que $n(k)$ es el número de nodos de grado k que tiene G . Se define la *distribución de grado* de la red G como

$$P(k) := \frac{n(k)}{n}.$$

Así, para cada k , el número $P(k)$ representa la probabilidad de que un vértice seleccionado al azar tenga grado k . Y si vemos $P(k)$ como una función de k estamos ante una distribución de probabilidades sobre el grafo.

4.2. Herramientas de computación

Existe bastante software que permite calcular la homología persistente. El paquete javaPlex [40], escrito en Java y preparado para usarse con la interfaz de MATLAB, cuenta con una amplia documentación que incluye numerosos ejemplos. Fue desarrollado por el grupo de topología computacional de la Universidad de Stanford, está basado en la librería Plex y sustituye al antiguo paquete jPlex. Además, es capaz de trabajar en los cuerpos finitos \mathbb{Z}_p (e incluso en \mathbb{Q}) y producir imágenes con los códigos de barras. El artículo [36] detalla los pormenores de estos y otros softwares disponibles, comparando las funciones de cada uno e indicando sus distintos objetivos o especializaciones. Estas diferencias están provocadas en parte por uno de los problemas fundamentales que deben afrontar estos programas: las representaciones fieles de los espacios en forma de complejos simpliciales tienden a ser muy costosas computacionalmente.

En el artículo [27] que seguiremos se hace uso de la librería Plex en MATLAB para los cálculos relacionados con el análisis de conjuntos de datos convertidos en complejos simpliciales y del paquete (no oficial) Simplicial Homology [16] del software de álgebra discreta computacional GAP [21] para trabajar con la homología simplicial. Por su parte, en el otro artículo de referencia [9] se utiliza el software Gephi [3] para manipular y dibujar los grafos, código en Java para tratar los datos y javaPlex para el cálculo de los intervalos de persistencia.

En nuestro caso, los intentos de reproducir algunos de los ejemplos expuestos en [27] se han realizado gracias al paquete igraph [12] del software estadístico R para generar los grafos, junto con el paquete javaPlex ejecutado en MATLAB para el cálculo de la

homología persistente y la obtención de los correspondientes códigos de barras. Todo el código utilizado para generar los ejemplos propios puede consultarse en el Apéndice A de la memoria.

4.3. Redes aleatorias de Erdős-Rényi

El primer ejemplo que se estudia en el artículo [27] es el de redes aleatorias. Existen varias formas de construir un grafo aleatorio, pero todas parten de la idea de describirlos mediante una distribución de probabilidades que genere la red. Aunque no es el único, el artículo utiliza el modelo de Erdős-Rényi, uno de los más conocidos, para crear la red aleatoria de estudio.

Definición 4.12. Para construir una *red aleatoria de Erdős-Rényi*, denotada por $G(n, p)$, se considera un número fijo n de vértices aislados y entre cada par de nodos se incluye una arista con probabilidad $0 < p < 1$.

En la práctica, según [20], se fija un parámetro p para construir la red. De esta forma, para cada par de nodos se genera uniformemente un número aleatorio r en el intervalo $(0, 1)$ y se incluye una arista entre los nodos seleccionados si $p > r$.

En este caso, la construcción de la red $G(n, p)$ hace que sea claro que la existencia de cada posible arista siga una distribución de Bernoulli con probabilidad p . Por tanto, la distribución de grado para un vértice v tomado al azar toma la forma de una binomial de parámetros $n - 1$ y p (pues v se puede conectar con $n - 1$ nodos con probabilidad p), luego

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}.$$

Un resultado clásico de estadística [10] establece que cuando n va aumentando y el producto $(n - 1)p = \lambda$ se mantiene constante, la distribución binomial converge hacia una distribución de Poisson de parámetro λ . Esto es,

$$P(k) \rightarrow \frac{\lambda^k e^{-\lambda}}{k!}.$$

La primera red aleatoria estudiada en [27] se generó a partir de 2000 nodos, con una probabilidad de enlace entre ellos de $p = 0,005$. Mediante el software estadístico R generamos un grafo aleatorio con los parámetros del artículo. A continuación, utilizamos MATLAB para construir el complejo clique del grafo, filtrarlo y calcular su homología persistente. Para obtener la \mathbb{N} -filtración, tomamos como i -ésimo complejo el i -esqueleto del complejo clique, es decir, el conjunto de símlices de dimensión menor o igual que i . Debido a la dispersión de la red, truncamos la construcción del complejo clique en el 3-esqueleto. La Figura 4.1 muestra el código de barras correspondiente a la red aleatoria generada y la filtración descrita.

Un rasgo característico que presenta el código de barras superior que vamos a analizar es que cada generador pervive o bien durante una única etapa de la filtración o bien para

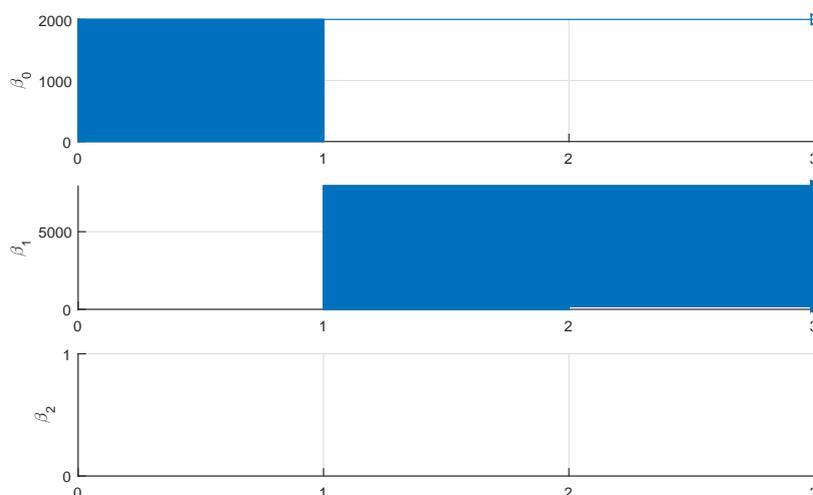


Figura 4.1: Código de barras para una red aleatoria $G(n, p)$ con $n = 2000$ y $p = 0,005$.

siempre. Esto se debe a la elección de la filtración mediante los sucesivos esqueletos del complejo, puesto que la homología simplicial en grado j es nula para todo i -esqueleto con $i < j$ y, como vimos en el primer capítulo de la memoria, esta solo depende de los símlices de dimensiones j y $j + 1$ del complejo (esto es, el $(j + 1)$ -esqueleto). En particular, esto significa que para calcular la homología persistente en grado 0 para esta filtración basta con estudiar el 1-esqueleto (el propio grafo) porque rellenar los triángulos del 2-esqueleto no modifica el número de componentes conexas. Análogamente, ocurre lo mismo en dimensiones superiores.

En grado 0, el diagrama permite apreciar que la filtración empieza con todos los nodos sin conexiones en la etapa 0, por eso se muestran 2000 barras. Sin embargo, a partir de la etapa 1 se incluyen todas las aristas del grafo y el código de barras indica que hay un único generador persistente correspondiente a una línea que perdura a lo largo de toda la filtración. Puesto que H_0 indica las componentes conexas del grafo, significa que la red aleatoria es conexa.

Por otra parte, del diagrama se concluye que $\beta_2 = 0$ y mediante comandos de MATLAB obtenemos que $\beta_1 = 7750$ a partir de la etapa 2 de la filtración. Esto quiere decir que solo una mínima parte de los ciclos eran triángulos que se rellenan en la siguiente etapa de la filtración y que el resto están necesariamente formados por cuatro o más vértices. En resumen, la homología no trivial se encuentra íntegramente recogida en H_0 y H_1 .

Los resultados obtenidos en este ejemplo mediante cálculos propios concuerdan con los descritos en [27]: el grafo que generan los autores del artículo es conexo, no hay H_2 y una pequeña parte de los ciclos se rellena tras la primera etapa, mientras que el resto perdura para siempre como generadores de H_1 .

A continuación, se estudia un nuevo ejemplo de grafo aleatorio construido también mediante el modelo de Erdős-Rényi, pero con otro parámetro p . Al incrementar la pro-

bilidad manteniendo el número de nodos aparecen eventualmente homologías de grado superior, como en el caso de una red $G(n, p)$ con $n = 2000$ y $p = 0,02$ que muestra la Figura 4.2. Sin embargo, H_2 aparece únicamente como ruido topológico durante un grado filtrante. H_0 se comporta de la misma forma que antes y H_1 pierde casi un tercio de sus generadores tras una etapa, quedando finalmente $\beta_1 = 27285$ (en la etapa 2 hay un total de 38052 generadores para H_1).

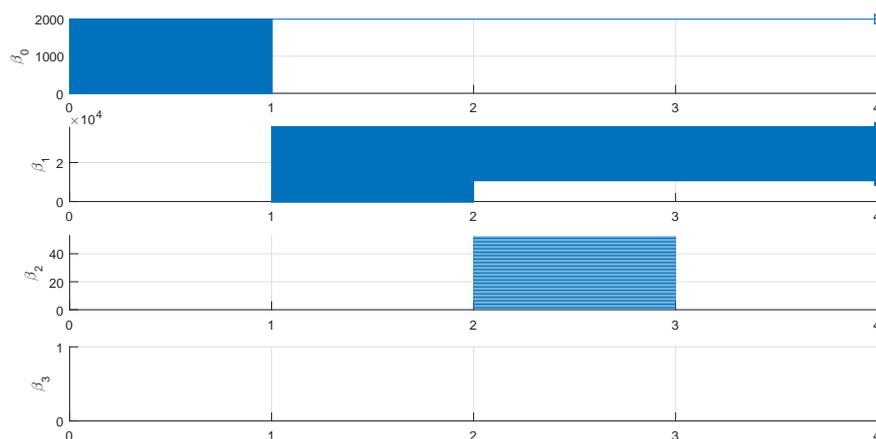


Figura 4.2: Código de barras para una red aleatoria $G(n, p)$ con $n = 2000$ y $p = 0,02$.

Aparte de la observación sobre la breve aparición de H_2 , cabe mencionar que el código de barras es coherente con resultados expuestos en [29] que afirman que si α es tal que $n = p^\alpha$ y α no está en el intervalo $(\frac{-1}{q}, \frac{-1}{2q+1})$, se tiene que H_q es casi siempre nulo. Aquí, H_2 vive por un corto período de tiempo, mientras que H_0 y H_1 perduran.

Como antes, los resultados arrojados por nuestra simulación también están en consonancia con los resultados comentados en el artículo de referencia [27].

4.4. Red de distribución exponencial

Aunque durante los primeros años después de su aparición se creía que el modelo de Erdős-Rényi servía para modelar redes del mundo real, finalmente se comprobó empíricamente que algunas suelen tener un comportamiento de distribución exponencial, en vez de poissoniana.

Una forma natural en que surgen estas redes es como sigue. Supongamos que empezamos con un único nodo inicial. A partir de ese momento, los nuevos nodos se van añadiendo de uno en uno y se van conectando con los nodos ya existentes aleatoriamente. Cuanto más tarde se haya añadido un nodo, menos grado se espera que tenga. La distribución de grado toma la forma

$$P(k) = Ae^{-\alpha k},$$

con A y α constantes.

Una propiedad que tienen estas distribuciones exponenciales es que poseen colas a la derecha, lo que significa que la mayoría de los nodos tienen un grado bajo, pero que un pequeño número de nodos, conocidos como *hubs* o *nodos centrales*, cuentan con grados muy altos.

En el ejemplo a estudiar se considera una red de correos electrónicos perteneciente a la Universitat Rovira i Virgili de Tarragona, con datos tomados del artículo [23]. Los nodos de la red, 1669 en total, son direcciones de correo electrónico y las aristas representan comunicaciones bidireccionales entre ellas. Una vez desechados los correos masivos (aquellos difundidos desde una única dirección a más de 50 remitentes), la red es no dirigida y presenta una distribución de grado exponencial

$$P(k) \propto e^{-\frac{k}{9.2}}$$

para $k \geq 2$. Según apuntan los autores en [23], del total de 1669 nodos, 1133 pertenecen a una gran componente conexa y el resto están aislados o conectados por pares. En vista del código de barras correspondiente a H_0 de la Figura 4.3, los autores de [27] han tratado solo los datos de esta gran componente.

Al observar la Figura 4.3, se aprecia que únicamente H_0 , H_1 y H_2 presentan características persistentes. El código de barras indica que los demás módulos de homología tienen generadores que perviven en intervalos que apenas duran una etapa de la filtración.

Conviene notar que esta red tiene un número de nodos comparable con los de las redes aleatorias estudiadas en la sección precedente. No obstante, en este caso aparecen bastantes módulos de homología de dimensiones superiores debido a la organización interna que rige el envío de correos electrónicos. El sistema de comunicaciones hace más viable la formación de subgrafos completos que generen ciclos no triviales, debido al intercambio de mensajes en comunidades de usuarios. Esta situación no se produce en el caso de las redes aleatorias, donde los 1-ciclos son los que dominan el complejo.

Por otra parte, la breve vida de los módulos de homología de dimensiones superiores a 2 sugiere que las comunidades de más de 4 usuarios se producen, en realidad, dentro de una comunidad mayor, dado que el clique debe formar parte del borde de otro de dimensión superior que lo anule en homología en la siguiente etapa de la filtración. Los generadores persistentes para H_1 y H_2 manifiestan que los correspondientes usuarios cruzan mensajes sin formar parte de una comunidad mayor (sin completar el clique), pues su «triángulo» o «tetraedro» no se rellena al avanzar un paso en la filtración. En otras palabras, no todas las personas del grupo intercambian mensajes con el resto de personas del grupo.

Sin embargo, esta red muestra un comportamiento más robusto que las dos redes aleatorias estudiadas en la sección anterior. Mientras que en aquel caso los únicos módulos de homología que presentaban persistencia eran H_0 y H_1 , en el caso de la red de correos electrónicos se tiene que H_2 también muestra características persistentes. En consecuencia, la persistencia da importancia a la resistencia que posee la red frente a la adición o eliminación de nodos; al fin y al cabo, se trata de la homología calculada a lo largo de una filtración.

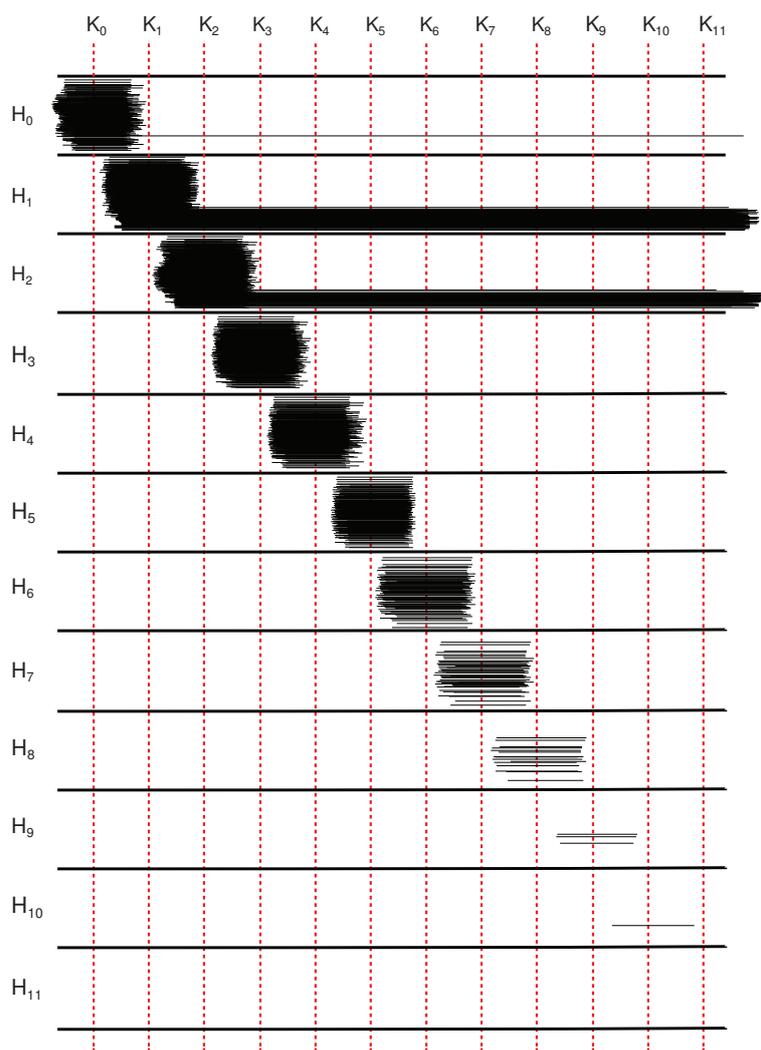


Figura 4.3: Código de barras para una red de correos electrónicos con distribución de grado exponencial con 1669 nodos. La imagen está tomada directamente del artículo [27].

© SISSA Medialab Srl. Reproduced by permission of IOP Publishing. All rights reserved.

4.5. Red de colaboración científica

Para el tercer ejemplo de estudio pasamos a detallar los resultados descritos en el artículo [9] de C. J. Carstens y K. J. Horadam. En él, se calcula la homología persistente de cuatro redes de colaboración científica reales y esta se utiliza para tratar de distinguirlas de redes aleatorias similares. El análisis que detallaremos aquí es el de una red de colaboración entre científicos tomada de [35].

Las aristas de la red se consideran con pesos, que indican conexiones de mayor o menor fuerza entre científicos. Se puede transformar un grafo con pesos en otro sin pesos manteniendo solamente aquellas aristas que superen un peso umbral fijado ω^* . Para un grafo con pesos G , denotamos por $G(\omega^*)$ este subgrafo umbral. Con esto, una sucesión de pesos

$\omega_1 > \omega_2 > \dots > \omega_k$ proporciona una filtración del grafo

$$\emptyset \subset G(\omega_1) \subset G(\omega_2) \subset \dots \subset G(\omega_k) \subset G.$$

Además, si construimos el complejo clique de G , la misma sucesión de pesos proporciona también una filtración del complejo clique, ya que el i -esqueleto del complejo clique de cualquier subgrafo $G(\omega_j)$ es necesariamente un subcomplejo del i -esqueleto del complejo clique de G .

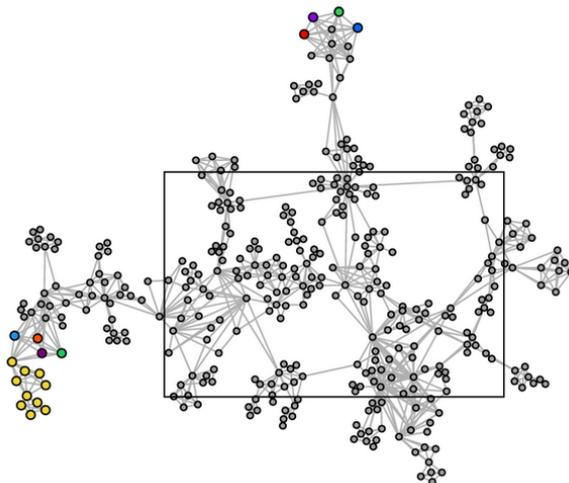


Figura 4.4: La red de colaboración científica estudiada. La componente conexa más grande se encuentra enmarcada en un recuadro. La imagen está tomada de [9].

El artículo [9] solo estudia la persistencia de β_0 y β_1 , luego les basta con considerar la filtración del grafo y del 2-esqueleto. La red estudiada tiene a los autores de artículos por nodos y dos científicos están conectados por una arista si son coautores de, al menos, un artículo. Además, por cada artículo en el que colaboren se suma un peso de

$$\frac{1}{n-1}$$

a la correspondiente arista, siendo n el número de autores del artículo. Por tanto, las conexiones más fuertes se corresponden con la gente que colabora a menudo y en grupos pequeños, mientras que las débiles forman parte de una comunidad (tanto más grande cuanto más débil).

Al igual que en la red de correos electrónicos anterior, los autores de [9] reducen el estudio de la red a la mayor componente conexa, que consta de 379 vértices y 914 aristas, y en la que el rango de pesos oscila entre 0,125 y 4,75. La red se muestran en la Figura 4.4.

Para grado 0, empiezan la filtración con $\omega^* = 5$ y van disminuyendo este valor. Así, en un primer momento aparecen los 379 vértices pero ninguna arista. Las primeras aristas surgen al sobrepasar el peso umbral $\omega^* = 4,75$. A medida que ω^* se hace más pequeño, el grafo se vuelve conexo y se tiene, finalmente, $\beta_0 = 1$.

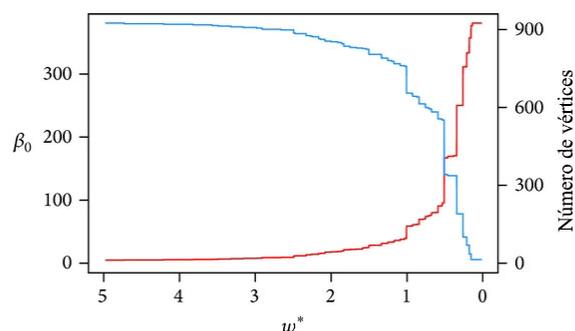


Figura 4.5: En azul, la evolución de β_0 al avanzar en la filtración; en rojo, la evolución del número de vértices al avanzar en la filtración. La imagen está tomada de [9].

La Figura 4.5 presenta la evolución de β_0 en azul y la del número de vértices en rojo, a medida que se avanza en la filtración del grafo. Los autores de [9] han optado por una gráfica de este tipo en lugar del código de barras correspondiente a H_0 para comparar la evolución de este β_0 con respecto a la evolución del β_0 correspondiente a un total de 1000 grafos aleatorios de Erdős-Rényi con el mismo número de nodos y aristas y los mismos pesos. Al hacerlo, se encontraron diferencias significativas durante las etapas intermedias de la filtración que permiten distinguir la red científica de los grafos aleatorios.

En particular, la Figura 4.6 sirve para constatar visualmente las diferentes evoluciones de β_0 para la red científica (en azul) y diez redes aleatorias (todas superpuestas en rojo). A pesar de que en las etapas iniciales y finales las gráficas son similares, durante las etapas intermedias de la filtración surgen comportamientos diferenciadores. Por otra parte, los autores afirman que también compararon los correspondientes códigos de barras mediante la distancia de cuello de botella (definida, por ejemplo, en [18]) y sus cálculos arrojaron diferencias suficientemente significativas para permitir discernir las topologías de la red de colaboración y las redes aleatorias.

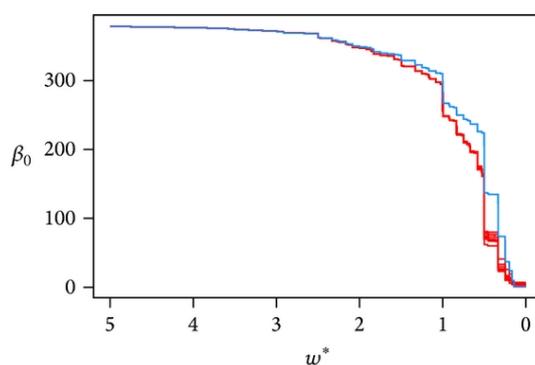


Figura 4.6: En azul, la evolución de β_0 para la red científica; en rojo, superpuestas, las evoluciones de β_0 para diez redes aleatorias. La imagen está tomada de [9].

Finalmente, el artículo analiza el primer número de Betti del complejo clique asociado a la red. La Figura 4.7 muestra el código de barras correspondiente a H_1 en las etapas finales de la filtración. Los colores sirven para identificar los ciclos del complejo que se ven reflejados mediante barras en el diagrama inferior. Los autores notan que los ciclos con más aristas se crean para los pesos umbrales más pequeños, aunque muchas de esas conexiones tienen pesos grandes.

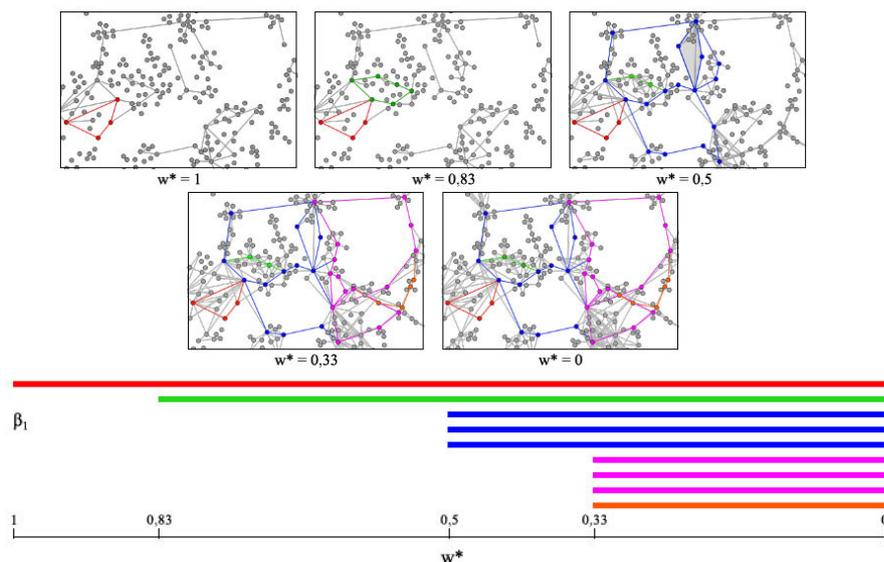


Figura 4.7: Código de barras para H_1 correspondiente a cinco etapas finales de la filtración. En la parte superior se señalan los ciclos generadores en distintos colores para identificarlos fácilmente en el diagrama inferior. En la etapa correspondiente a $\omega^* = 0,5$ hay dos triángulos sombreados para indicar que ahí no hay ningún ciclo. La imagen está tomada de [9].

Nuevamente, tratan de usar la información dada por β_1 para distinguir esta red de otras aleatorias. Como antes, generan 1000 redes aleatorias y calculan una media de 520,65 intervalos (con una desviación típica de 4,39) para el código de barras de H_1 , cuando la red científica tiene 9 intervalos como se observa en la Figura 4.7. Este comportamiento ya había quedado patente cuando analizamos nuestras redes aleatorias: las redes de Erdős-Rényi no tienen tendencia a formar agrupaciones, sino que los 1-ciclos dominan la red.

Por tanto, β_1 es suficiente para comparar y distinguir redes aleatorias de otras más estructuradas. Esto les sugiere que un análisis más profundo de la homología persistente a lo largo de toda la filtración permita detectar diferencias más sutiles en la topología de otras redes que posean una estructura mucho más parecida.

Para acabar, un detalle menor que comentan es que todos los ciclos de H_1 que nacen viven hasta el final. Esto no sería así si, por ejemplo, los cuatro autores del ciclo rojo que aparece en $\omega^* = 1$ hubieran colaborado en un artículo. En ese supuesto, aparecería una arista diagonal en $\omega^* = 0,33$ que mataría el ciclo.

4.6. Conclusiones

Los análisis de redes complejas mediante homología persistente permiten establecer una nueva medida de la robustez de una red. La aparición de generadores persistentes en los módulos de homología de dimensión superior guarda relación con la resistencia de la red frente a ataques que añadan o eliminen nodos. En particular, la red de correos electrónicos con distribución exponencial muestra características persistentes en H_2 , al contrario que las redes aleatorias o la red de colaboración, que serán más débiles en este sentido.

Como segunda aplicación interesante, la homología persistente, incluso en las dimensiones más pequeñas, es una herramienta para distinguir la red de colaboración científica estudiada de redes aleatorias creadas con características muy parecidas.

Apéndice A

Código para implementar los ejemplos

El presente apéndice recoge los códigos en R y MATLAB desarrollados para crear el primer grafo aleatorio estudiado en el capítulo anterior y analizar su homología persistente. Pequeñas modificaciones de los mismos permiten recrear el segundo ejemplo.

Se ha optado por definir una métrica sobre el grafo y, mediante esta, calcular los cliques que conforman el complejo clique, sabiendo que el paquete `igraph` tiene una función que devuelve directamente los cliques de un grafo (y que, posiblemente, haría la implementación del código más eficiente).

A.1. Código en R

El siguiente código en R crea un grafo aleatorio $G(2000, 0,005)$ y guarda en sendos ficheros de texto la información sobre los vértices conectados y sobre el número de nodos y aristas.

```
library(igraph)
n <- 2000
p <- 0.005
g <- sample_gnp(n, p, directed = FALSE)

v <- get.data.frame(g, 'vertices')
e <- get.data.frame(g, 'edges')
num_aristas <- nrow(e)

write.table(e, "grafo_aleatorio.txt", sep="\t", ...
            col.names = FALSE, row.names = FALSE)
write(n, "variables.txt", sep="\n", append = FALSE)
write(num_aristas, "variables.txt", sep="\n", append = TRUE)
```

A.2. Código en MATLAB

El siguiente código en MATLAB lee la información sobre el grafo creado previamente en R, crea el complejo clique de dimensión 3 asociado al grafo y lo filtra mediante sus sucesivos esqueletos. Después, calcula la homología persistente del grafo para esa filtración y dibuja el código de barras asociado. Finalmente, imprime por pantalla los números de Betti persistentes.

```
% En primer lugar, situados en el directorio donde esté
% descargado javaPlex hay que ejecutar el comando
% load_javaplex
% para cargar el paquete y poder trabajar con el código inferior.

% Incorporamos las variables que necesitamos del script de R.
fileID = fopen('variables.txt', 'r');
variables = fscanf(fileID, '%d');
fclose(fileID);
n = variables(1);
num_aristas = variables(2);

% Leemos los datos sobre las aristas
% conectadas del archivo creado con R.
fileID = fopen('grafo_aleatorio.txt', 'r');
aleatorio = fscanf(fileID, '%d %d', [2, num_aristas]);
aleatorio = aleatorio';
fclose(fileID);

% Definimos un espacio métrico sobre  $G(n,p)$ 
% dotando a cada arista de la métrica que la
% haga isométrica al intervalo  $[0,1]$ .
distancias = zeros(n);
for i = 1:num_aristas
    distancias(aleatorio(i,1),aleatorio(i,2)) = 1;
    distancias(aleatorio(i,2),aleatorio(i,1)) = 1;
end

% Inicializamos el complejo simplicial.
complejo = api.Plex4.createExplicitSimplexStream();

% Añadimos los vértices en la etapa 0 de la filtración.
for i = 1:n
    complejo.addVertex(i, 0);
end
```

```
% Añadimos las aristas en la etapa 1 de la filtración.
for i = 1:num_aristas
    complejo.addElement([aleatorio(i,1), aleatorio(i,2)], 1);
end

% Añadimos los 2-símplices en la etapa 2 de la filtración.
for i = 1:n-2
    for j = i+1:n-1
        if (distancias(i,j) == 1)
            for k = j+1:n
                if (distancias(i,k) == 1 && distancias(j,k) == 1)
                    complejo.addElement([i, j, k], 2);
                end
            end
        end
    end
end

% Añadimos los 3-símplices en la etapa 3 de la filtración.
for i = 1:n-3
    for j = i+1:n-2
        if (distancias(i,j) == 1)
            for k = j+1:n-1
                if (distancias(i,k) == 1 && distancias(j,k) == 1)
                    for l = k+1:n
                        if (distancias(i,l) == 1 && distancias(j,l) == 1 ...
                            && distancias(k,l) == 1)
                            complejo.addElement([i, j, k, l], 3);
                        end
                    end
                end
            end
        end
    end
end

% Finalizamos el complejo.
complejo.finalizeStream();

% Comprobamos que el código anterior crea,
% efectivamente, un complejo simplicial.
complejo.validateVerbose();
```

```
% Calculamos la homología persistente del complejo hasta la
% de dimensión 2 (calculamos H_0, H_1 y H_2) sobre el cuerpo Q.
persistencia = api.Plex4.getRationalSimplicialAlgorithm(3);

% Dibujamos el código de barras.
dibujar_intervalos = persistencia.computeIntervals(complejo);

% Guardamos el código de barras en un archivo, indicando el
% número de etapas de la filtración que se debe considerar
% y la dimensión homológica máxima a mostrar.
options.filename = 'codigo_de_barras';
options.max_filtration_value = 3;
options.max_dimension = 2;
plot_barcodes(dibujar_intervalos, options);

% Imprimimos por pantalla los números de Betti
% persistentes, correspondientes a los intervalos
% no acotados del código de barras.
intervalos = persistencia.computeAnnotatedIntervals(complejo);
intervalos_infinitos = intervalos.getInfiniteIntervals();
intervalos_infinitos.getBettiNumbers()
```

Bibliografía

- [1] Aaronson, S., Scoville, N. A. Lusternik–Schnirelmann category for simplicial complexes. *Illinois J. Math.*, **57**(3):743–753, 2013.
- [2] Ayala, R., Domínguez, E., Quintero, A. *Elementos de la teoría de homología clásica*, **65**. Universidad de Sevilla, 2002.
- [3] Bastian, M., Heymann, S., Jacomy, M. Gephi: An open source software for exploring and manipulating networks, 2009. Software disponible en <http://www.gephi.org>.
- [4] Bubenik, P., Kim, P. T. A statistical approach to persistent homology. *Homology, Homotopy Appl.*, **9**(2):337–362, 2007.
- [5] Carlsson, G. Topology and data. *Bull. Am. Math. Soc.*, **46**(2):255–308, 2009.
- [6] Carlsson, G., Ishkhanov, T., De Silva, V., Zomorodian, A. On the local behavior of spaces of natural images. *IJCV*, **76**(1):1–12, 2008.
- [7] Carlsson, G., Singh, G., Zomorodian, A. Computing multidimensional persistence. En *Algorithms and computation*, 730–739. Springer, 2009.
- [8] Carlsson, G., Zomorodian, A., Collins, A., Guibas, L. J. Persistence barcodes for shapes. *Int. J. Shaping Model.*, **11**(02):149–187, 2005.
- [9] Carstens, C. J., Horadam, K. J. Persistent homology of collaboration networks. *Math. Probl. Eng.*, **2013**, 2013.
- [10] Casella, G., Berger, R. L. *Statistical inference*. Duxbury Pacific Grove, 2002.
- [11] Collins, A., Zomorodian, A., Carlsson, G., Guibas, L. J. A barcode shape descriptor for curve point cloud data. *Computers & Graphics*, **28**(6):881–894, 2004.
- [12] Csardi, G., Nepusz, T. The igraph software package for complex network research. *InterJournal*, Complex Systems 1695, 2006. <http://igraph.org>.
- [13] De Silva, V., Carlsson, G. Topological estimation using witness complexes. *Proc. Sympos. Point-Based Graphics*, 157–166, 2004.

- [14] De Silva, V., Ghrist, R. Coverage in sensor networks via persistent homology. *Algebr. Geom. Topol.*, **7**(1):339–358, 2007.
- [15] De Silva, V., Ghrist, R., Muhammad, A. Blind swarms for coverage in 2-D. En *Robotics: Science and Systems*, 335–342, 2005.
- [16] Dumas, J.-G., Heckenbach, F., Saunders, B. D., V., W. Simplicial Homology, a (proposed) GAP share package. <http://ljk.imag.fr/membres/Jean-Guillaume.Dumas/Homology/>.
- [17] Edelsbrunner, H. The union of balls and its dual shape. En *Proceedings of the ninth annual symposium on Computational geometry*, 218–231. ACM, 1993.
- [18] Edelsbrunner, H., Harer, J. Persistent homology – a survey. *Contemporary mathematics*, **453**:257–282, 2008.
- [19] Edelsbrunner, H., Harer, J. *Computational topology: an introduction*. American Mathematical Society, 2010.
- [20] Estrada, E., Knight, P. *A first course in network theory*. Oxford University Press, 2015.
- [21] The GAP Group. *GAP – Groups, Algorithms, and Programming, Version 4.4.10*, 2007. <http://www.gap-system.org>.
- [22] Ghrist, R. Barcodes: the persistent topology of data. *Bull. Am. Math. Soc.*, **45**(1):61–75, 2008.
- [23] Guimerà, R., Danon, L., Díaz-Guilera, A., Giralt, F., Arenas, A. Self-similar community structure in a network of human interactions. *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.*, **68**(6):065103, 2003.
- [24] Gyulassy, A., Natarajan, V. Topology-based simplification for feature extraction from 3D scalar fields. En *Visualization, 2005. VIS 05. IEEE*, 535–542. IEEE, 2005.
- [25] Hartley, B., Hawkes, T. O. *Rings, modules and linear algebra*. Chapman & Hall/CRC, 1970.
- [26] Hausmann, J.-C. On the Vietoris-Rips complexes and a cohomology theory for metric spaces. *Ann. Math. Studies*, **138**:175–188, 1995.
- [27] Horak, D., Maletić, S., Rajković, M. Persistent homology of complex networks. *J. Stat. Mech. Theor. Exp.*, **2009**(03):P03034, 2009.
- [28] Jake (<http://tex.stackexchange.com/users/2552/jake>). Moebius strip using TikZ. Stack Exchange, 10 de junio de 2013. <http://tex.stackexchange.com/questions/118563/moebius-strip-using-tikz>, visitado el 3 de marzo de 2016.

-
- [29] Kahle, M. Topology of random clique complexes. *Discrete Math.*, **309**(6):1658–1671, 2009.
- [30] Kozlov, D. *Combinatorial algebraic topology*. Springer Science & Business Media, 2007.
- [31] Lamar León, J., García Reyes, E. B., González Díaz, R. Human gait identification using persistent homology. En *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, 244–251. Springer, 2012.
- [32] Lang, S. *Algebra*. Addison-Wesley, 1993.
- [33] Latschev, J. Vietoris-Rips complexes of metric spaces near a closed Riemannian manifold. *Arch. Math. (Basel)*, **77**(6):522–528, 2001.
- [34] Munkres, J. R. *Elements of algebraic topology*. Addison-Wesley Reading, 1984.
- [35] Newman, M. E. J., 2006. *Phys. Rev. E* **74**, 036104.
- [36] Otter, N., Porter, M. A., Tillmann, U., Grindrod, P., Harrington, H. A. A roadmap for the computation of persistent homology. *arXiv preprint arXiv:1506.08903*, 2015.
- [37] Salvatore, J. Applying Topology to Data, Part 1: A Brief Introduction to Abstract Simplicial and Čech Complexes. DyingLoveGrape. http://www.dyinglovegrape.com/math/topology_data_1.php, visitado el 11 de marzo de 2016.
- [38] Salvatore, J. Applying Topology to Data, Part 2: More Čech Complexes, the Vietoris-Rips Complexes, and Clustering. DyingLoveGrape. http://www.dyinglovegrape.com/math/topology_data_2.php, visitado el 11 de marzo de 2016.
- [39] Singh, G., Memoli, F., Ishkhanov, T., Sapiro, G., Carlsson, G., Ringach, D. L. Topological analysis of population activity in visual cortex. *J. Vis.*, **8**(8):1–18, 2008.
- [40] Tausz, A., Vejdemo-Johansson, M., Adams, H. JavaPlex: A research software package for persistent (co)homology. En *Proceedings of ICMS 2014*, Lecture Notes in Computer Science 8592, 129–136, 2014. Software disponible en <http://appliedtopology.github.io/javaplex/>.
- [41] Zomorodian, A., Carlsson, G. Computing persistent homology. *Discrete Comput. Geom.*, **33**(2):249–274, 2005.

