



UNIVERSITAT DE
BARCELONA

Facultat de Matemàtiques
i Informàtica

GRAU DE MATEMÀTIQUES

Treball final de grau

Homologia persistent d'hipergrafs

Autor: Cinta Isern i Vizoso

Director: Dr. Carles Casacuberta Vergés

Realitzat a: Departament de Matemàtiques i Informàtica

Barcelona, 10 de juny de 2025

Abstract

This work explores persistent homology applied to hypergraphs equipped with a filtration. Hypergraphs are structures that model higher-order interactions beyond those captured by graphs, by including hyperedges, that can contain any finite set of vertices. We focus on *embedded homology*, a construction proposed by Bressan, Li, Ren, and Wu in 2019, which retains the explicit information of a hypergraph without resorting to its simplicial closure, i.e., the simplicial complex obtained by adding all subsets of its hyperedges. We implement an algorithm due to Liu, Feng, Wu, and Xia (2024) to compute persistent homology of filtered hypergraphs, and apply it to mathematical coauthorship networks. Results are compared with those obtained using the simplicial closure of hypergraphs, highlighting conceptual and descriptive differences between the two approaches. Overall, this project aims to contribute to the dissemination, understanding, and application of a new framework for the topological analysis of data with complex, higher-order interactions.

Resum

Aquest treball estudia l'homologia persistent aplicada a hipergrafs amb filtracions. Els hipergrafs són estructures que modelen interaccions d'ordre superior més enllà dels grafs, mitjançant hiperarestes, que poden contenir qualsevol nombre de vèrtexs. Ens centrem en l'homologia restringida (*embedded homology*), una construcció proposada per Bressan, Li, Ren i Wu el 2019, que preserva la informació explícita dels hipergrafs sense recórrer a la clausura simplicial, que és el complex simplicial obtingut afegint-hi tots els subconjunts de les hiperarestes. Hem implementat un algorisme de Liu, Feng, Wu i Xia (2024) per calcular homologia persistent d'hipergrafs amb filtracions, i l'hem aplicat a xarxes de coautoria d'articles de matemàtiques. Comparem els resultats amb els que s'obtenen mitjançant la clausura simplicial, destacant les diferències conceptuais i descriptives entre ambdós enfocaments. Aquest treball pretén contribuir a la difusió, comprensió i aplicació d'un nou enfocament per a l'anàlisi topològica de dades amb interaccions complexes i d'ordre superior.

Agraïments

Vull expressar el meu més sincer agraïment al meu tutor, en Carles Casacuberta, per haver-me transmès tants coneixements i, sobretot, per la seva dedicació i suport constant al llarg de tot aquest treball. Li agraeixo profundament la confiança que ha dipositat en mi, la seva guia durant tot el treball i la seva disponibilitat per ajudar-me sempre que ho he necessitat. També vull donar les gràcies als professors de l'assignatura de topologia global de corbes i superfícies, ja que va ser gràcies a aquella assignatura que vaig descobrir l'interès per aquesta branca de la matemàtica i em vaig animar a endinsar-m'hi a través d'aquest treball. No puc deixar d'agrair el suport incondicional de la meua família, especialment al meu pare, la meua mare i la meua germana bessona, per haver estat sempre al meu costat, animant-me i recolzant-me en tot moment, especialment durant el procés d'elaboració d'aquest treball. També vull fer una menció especial als meus dos gossos, que amb la seva companyia i alegria diària han fet el camí molt més agradable. Finalment, vull agrair als meus amics del grau de matemàtiques per tot el suport i l'ajuda que m'han donat, i a aquells companys que ja havien cursat assignatures relacionades, com topologia algebraica, i que m'han ajudat a entendre conceptes fonamentals per al desenvolupament d'aquest treball. Gràcies a tots per fer aquest camí més enriquidor i compartit.

Índex

Introducció	2
1 Homologia	5
1.1 Complexos de cadenes	5
1.2 Homologia simplicial	5
2 Complex de cadenes ínfim i suprem	8
2.1 Definicions	8
3 Homologia d'hipergrafs	12
3.1 Clausura simplicial d'un hipergraf	12
3.2 Homologia restringida	14
4 Homologia persistent	18
4.1 Homologia persistent clàssica	18
4.2 Descriptors numèrics de persistència	22
4.3 Homologia persistent d'hipergrafs	23
5 Algorisme per a l'homologia persistent d'hipergrafs	25
5.1 Algorisme clàssic d'homologia persistent per a complexos de cadenes . . .	25
5.2 Algorisme d'homologia persistent d'hipergrafs	27
5.2.1 Construcció del complex de cadenes suprem	27
6 Exemple pràctic de l'homologia persistent d'hipergrafs	30
6.1 Resultats obtinguts	31
6.1.1 Topologia algebraica	31
6.1.2 Topologia general	35
Conclusions	39

Introducció

El projecte

Actualment, la gran quantitat i complexitat de dades generades en àmbits com la biologia, la informàtica o les xarxes socials fa necessari el desenvolupament d'eines que permetin entendre no només els elements que componen les dades, sinó també les relacions entre elles. Aquestes relacions no són sempre binàries, sinó que poden involucrar múltiples elements simultàniament. És en aquest context on els hipergrafs emergeixen com una eina matemàtica per poder modelar relacions d'ordre superior. A diferència dels grafs clàssics, on cada aresta connecta exactament dos vèrtexs, un hipergraf permet que una hiperaresta connecti un nombre arbitrari de vèrtexs, fet que en fa una generalització natural per representar interaccions complexes entre múltiples elements.

A principis del segle XX, matemàtics com Sperner, Ramsey, Hall i König van ser precursors en l'estudi de propietats combinatòries de famílies de conjunts, tot i que encara no existia una teoria unificada. Va ser a partir de la dècada dels seixanta quan la teoria dels hipergrafs es va començar a consolidar com a disciplina pròpia. En aquest procés, destaca la figura de Claude Berge, qui va formalitzar per primera vegada el concepte d'hipergraf cap al 1960, tot generalitzant la teoria de grafs clàssica per permetre la representació de relacions que involucren més de dos elements. Des de llavors, els hipergrafs han esdevingut una eina fonamental tant en el desenvolupament teòric com en aplicacions pràctiques.

Com a representació topològica general, els hipergrafs han estat utilitzats en àrees diverses com la química, la biologia, la informàtica o l'anàlisi de dades complexes [10, 15]. Paral·lelament, l'anàlisi topològica de dades (TDA) ha sorgit com un camp que aplica eines de la topologia algebraica, com l'homologia, per extreure informació rellevant de grans conjunts de dades. Una de les eines centrals dins del TDA és l'*homologia persistent*, que permet estudiar com les característiques topològiques d'un objecte (com components connexes o cicles) apareixen i desapareixen al llarg d'una filtració. Aquesta tècnica és especialment útil per analitzar la forma de dades, i ha estat àmpliament aplicada en biologia, medicina, visió per computador, i ciències socials [21, 23].

Tot i la seva utilitat, l'homologia persistent tradicional es basa en complexos simplicials derivats de dades puntuals o grafs, la qual cosa limita l'anàlisi a interaccions binàries. Això pot resultar problemàtic en contextos on les dades naturals presenten interaccions explícitament d'ordre superior. Per superar aquesta limitació, s'ha proposat recentment estendre l'homologia persistent als hipergrafs, que generalitzen tant els grafs com els complexos simplicials i permeten capturar de manera directa aquestes relacions complexes. En aquest context, s'ha desenvolupat la construcció coneguda com a *embedded homology*, que permet definir l'homologia d'un hipergraf a partir de la seva estructura original, sense recórrer a la clausura simplicial. En el marc d'aquest treball, ens referirem a aquesta

construcció com a *homologia restringida*. Aquesta metodologia ha estat desenvolupada per Bressan, Li, Ren i Wu [5], inspirant-se en el model de *path complexes* proposat per Yau, Muranov, Lin i Grigor'yan l'any 2012 a [16]. L'homologia restringida permet obtenir una homologia fidel a la informació explícita de l'hipergraf, i a partir d'aquesta, Liu, Feng, Wu i Xia han proposat un algorisme per calcular-ne la versió persistent. Aquesta es troba a l'article [18], que adapta les eines de la topologia computacional a aquest nou marc.

Tot i així, no existeix una definició única per a l'homologia d'hipergrafs, i diversos autors han desenvolupat construccions alternatives per estudiar la seva topologia. Un dels mètodes més coneguts és la *construcció bipartida* de Dowker, basada en la relació d'incidència entre vèrtexs i hiperarestes per generar complexos filtrats [14]. D'altra banda, el treball de Gasparovic, Purvine, Sazdanovic, Wang, Wang i Ziegelmeier [10] ofereix una visió general i comparativa de les aproximacions recents a l'homologia d'hipergrafs, incloent-hi la construcció de la clausura simplicial i la definició de l'homologia restringida.

L'objectiu d'aquest treball és entendre, presentar i aplicar l'homologia restringida. Per una banda, s'estudia en profunditat la definició i propietats d'aquesta i per l'altra, s'implementa i s'aplica l'algorisme per al càlcul de la seva homologia persistent proposat per Liu, Feng, Wu i Xia a [18].

Per tal d'il·lustrar-ne el funcionament, es presenta un exemple concret basat en una xarxa de coautoria matemàtica, on les col·laboracions entre autors es modelen com a hiperarestes. En aquest cas, es calcula l'homologia persistent dels hipergrafs filtrats per l'any de publicació dels articles, mitjançant l'homologia restringida, i es compara amb la de la seva clausura simplicial. Cal remarcar que l'objectiu principal del treball no és la comparació entre àrees matemàtiques, sinó posar de manifest les diferències conceptuals i pràctiques entre dues maneres d'obtenir l'homologia d'un hipergraf: una fidel a la informació explícita i una altra basada en la clausura simplicial. Aquesta comparació permet evidenciar com el model d'homologia restringida pot aportar informació diferenciada i valuosa en contextos reals.

L'homologia persistent per a hipergrafs no només generalitza la dels grafs, sinó també la dels complexos simplicials. Aquesta observació justifica que la comparació natural es faci amb aquests últims, ja que la major part de tècniques de TDA es fonamenten en construccions simplicials. Entendre com es comporta l'homologia persistent en hipergrafs en relació amb aquests models pot ajudar a determinar-ne els avantatges i les limitacions.

Tot i que el cas pràctic que s'analitza té un caràcter acadèmic, l'ús de l'homologia persistent d'hipergrafs ja ha estat explorat amb èxit en contextos biomèdics, com el disseny de fàrmacs mitjançant intel·ligència artificial. En [17], per exemple, es proposa una representació molecular basada en hipergrafs i es construeixen descriptors topològics a partir de la cohomologia persistent restringida per alimentar models d'aprenentatge automàtic. Aquesta metodologia ha demostrat un rendiment superior en la predicció de l'afinitat d'unió proteïna-l·ligand en comparació amb descriptors tradicionals, fet que subratlla el potencial real d'aquestes eines en problemes de gran rellevància en biomedicina.

Amb tot, aquest treball pretén contribuir a la difusió i comprensió d'un nou enfocament per a l'anàlisi topològica de dades amb interaccions d'ordre superior.

Estructura de la memòria

El treball està estructurat en sis capítols, seguits d'un apartat de conclusions que en recull les idees principals. Al llarg de la memòria s'hi combinen tres grans blocs temàtics, una

primera part d'àlgebra, una segona centrada en topologia, i una tercera amb enfocament més computacional.

Al capítol 1, s'introdueixen els conceptes fonamentals de la teoria de l'homologia. Es defineixen els complexos de cadenes i l'operador vora, que permeten construir grups d'homologia a partir d'estructures algebraiques. A continuació, s'estudia el cas particular de l'homologia simplicial, dins del marc clàssic de la topologia algebraica, establint una base conceptual per comprendre les construccions d'homologia aplicades a estructures com els hipergrafs.

El capítol 2 es dedica a les construccions anomenades complex de cadenes ínfim i complex de cadenes suprem, associades a un hipergraf. S'exposen les seves definicions i propietats. Aquests complexos seran claus per definir l'homologia restringida.

En el capítol 3, es presenta l'homologia per a hipergrafs. En primer lloc, s'explica com es pot aplicar l'homologia clàssica a través de la clausura simplicial d'un hipergraf. Posteriorment, es desenvolupa la construcció coneguda com a *homologia restringida* (o *embedded homology*), que permet analitzar la topologia de l'hipergraf mantenint-se fidel a la seva estructura original.

El capítol 4 introdueix la teoria de l'homologia persistent. Primerament, s'exposen les definicions clàssiques aplicades a complexos filtrats i s'expliquen els conceptes de persistència i descriptors numèrics. En la segona part del capítol, s'estén la teoria a hipergrafs, adaptant la construcció persistent a partir de l'homologia restringida.

Al capítol 5, el focus es centra en els aspectes computacionals. Es revisa l'algorisme tradicional per calcular l'homologia persistent a través de complexos de cadenes filtrats, i s'explica amb detall l'algorisme adaptat a hipergrafs desenvolupat per Liu, Feng, Wu i Xia. Es dedica especial atenció a la construcció efectiva dels complexos ínfim i suprem al llarg de la filtració, que són la base per obtenir els resultats computables.

El capítol 6 presenta un exemple pràctic que posa en funcionament l'algorisme anterior. Es treballa amb una xarxa de coautoria en matemàtiques, on les col·laboracions entre autors es modelen com a hiperarestes. S'estudia l'homologia persistent dels hipergrafs filtrats per any de publicació dels articles, tant a través de la clausura simplicial com mitjançant l'homologia restringida. Aquesta comparació té com a objectiu exemplificar l'ús pràctic de l'homologia d'hipergrafs.

Finalment, l'apartat de conclusions recull les aportacions principals del treball, posa en relleu les diferències observades entre els diversos models, i suggereix possibles línies de recerca futures.

Capítol 1

Homologia

Per tal d'estudiar la forma dels espais topològics, la topologia algebraica introdueix eines que associen a cada espai certs invariants algebraics. Un dels mètodes fonamentals per aconseguir-ho és a través de la teoria d'homologia. Aquesta permet capturar propietats com la connexió, els “forats” o l'orientabilitat. Abans d'abordar el cas concret dels hipergrafs, es presenten els conceptes bàsics d'aquesta teoria, començant pels complexos de cadenes i la seva homologia. Totes les definicions utilitzades en aquest capítol es basen en els llibres [12] i [20].

1.1 Complexos de cadenes

Un *complex de cadenes* és una successió de grups abelians C_n , $n \geq 0$, i d'homomorfismes $\partial_n : C_n \rightarrow C_{n-1}$ (anomenats *aplicacions vora*) tals que $\partial_n \circ \partial_{n+1} = 0$ per a tot n . Denotarem un complex per $\{C_*, \partial_*\}$ al llarg d'aquest treball. Cada C_n del complex de cadenes s'anomena *grup de cadenes de dimensió n* . Si considerem una successió de subgrups $\{D_n\}_{n \geq 0}$ amb $D_n \subseteq C_n$ tal que, per a tot $n \geq 1$, $\partial_n(D_n) \subseteq D_{n-1}$, aleshores direm que $\{D_*, \partial_*\} := \{D_*, \partial_*|_{D_*}\}$ és un *subcomplex de cadenes* de $\{C_*, \partial_*\}$.

La *n -homologia* H_n d'un complex de cadenes $\{C_*, \partial_*\}$ es defineix com el grup quocient

$$H_n(\{C_*, \partial_*\}) = \text{Ker}(\partial_n) / \text{Im}(\partial_{n+1}).$$

1.2 Homologia simplicial

Donat un conjunt V , es defineix el seu conjunt de parts $\mathcal{P}(V)$ com la col·lecció de tots els subconjunts no buits de V . Un *complex simplicial (abstracte)* K és una parella (V_k, K) on V_k és un conjunt i K és un subconjunt de $\mathcal{P}(V_k)$ que satisfà les condicions següents:

- Per a qualsevol $v \in V_k$, el conjunt v pertany a K .
- Per a qualsevol $\sigma \in K$ i qualsevol subconjunt no buit $\tau \subseteq \sigma$, τ pertany a K .

Els elements de V_k s'anomenen *vèrtexs* i els elements de K s'anomenen *símplexs*. Un símplex format per $n+1$ elements de V_k s'anomena *n -símplex*. Es dirà K_n a la col·lecció de tots els n -símplexs de K . Una $(n-1)$ -cara d'un n -símplex és un $(n-1)$ -símplex obtingut

eliminant un vèrtex. La *dimensió* d'un complex simplicial K és el màxim enter n tal que K_n és no buit.

Donat un complex simplicial $K = (V_K, \mathcal{K})$, un *subcomplex* de K és un complex simplicial $K' = (V', \mathcal{K}')$ tal que $V' \subseteq V_K$ i $\mathcal{K}' \subseteq \mathcal{K}$.

Sigui G un grup abelià. Donat un conjunt finit no buit S , a partir d'ara s'usarà $G(S)$ per denotar la col·lecció de totes les combinacions lineals dels elements de S amb coeficients a G .

Sigui K un complex simplicial amb vèrtexs $V_k = \{v_1, \dots, v_n\}$. Es diu que K és un *complex simplicial ordenat* si s'ha fixat un ordre total en el conjunt V_k , és a dir, si $v_1 < \dots < v_n$ on $<$ és l'ordre total de V_k .

Donat un n -símplex ordenat $\Delta^n = \{v_0, \dots, v_n\}$, on $v_0, \dots, v_n$ són vèrtexs de V_k , considerem les seves $(n-1)$ -cares definides com $\Delta_i^{n-1} = \{v_0, \dots, \hat{v}_i, \dots, v_n\}$ per a tot $0 \leq i \leq n$. Aleshores, l'aplicació vora

$$\partial_n : G(\Delta^n) \longrightarrow G(\Delta_0^{n-1}, \dots, \Delta_n^{n-1})$$

es defineix com l'homomorfisme

$$\partial_n = \sum_{i=0}^n (-1)^i d_i$$

on d_i envia Δ^n a la cara Δ_i^{n-1} .

A partir d'un complex simplicial ordenat K , es pot construir de manera natural un complex de cadenes associat. Fixat un grup abelià G , es defineix per a cada $n \geq 0$ el grup $C_n(K; G) := G(K_n)$, on K_n són els n -símplexs de K . Amb les aplicacions vora ∂_n definides anteriorment, s'obté un complex de cadenes $(C_*(K; G), \partial_*)$.

Així doncs, l'homologia d'un complex simplicial és un cas particular de l'homologia de complexos de cadenes. Si K és un complex simplicial, per a tot $p \geq 0$ el grup d'homologia p -èsim de K es defineix com el grup d'homologia del seu complex de cadenes associat

$$H_p(K) := \text{Ker}(\partial_p) / \text{Im}(\partial_{p+1}).$$

En particular, el grup $H_0(K)$ proporciona informació sobre la connexió del complex. Aquest grup és lliure de rang igual al nombre de components connexes de la realització geomètrica $|K|$.¹

Des del punt de vista algebraic, considerant el complex de cadenes simplicial amb coeficients en $\mathbb{Z}$, es té que

$$H_0(K) = \text{Ker}(\partial_0) / \text{Im}(\partial_1).$$

Com que $\partial_0 : C_0 \rightarrow 0$ és l'aplicació nul·la, resulta que $\text{Ker}(\partial_0) = C_0$, i per tant

$$H_0(K) = C_0 / \text{Im}(\partial_1).$$

El grup C_0 és lliurement generat pels vèrtexs del complex, mentre que $\text{Im}(\partial_1)$ conté relacions del tipus $[v_j] - [v_i]$, provinents de les arestes $[v_i, v_j] \in K_1$. Aquestes relacions identifiquen vèrtexs connectats mitjançant cadenes de 1-símplexs, i per tant el quocient reflecteix la partició dels vèrtexs en components connexes.

¹La realització geomètrica $|K|$ d'un complex simplicial K és un espai topològic construït unint els símplexs segons la seva estructura combinatoria. Per a una definició formal, vegeu [20].

Així, $H_0(K) \cong \mathbb{Z}^r$, on r és el nombre de components connexes de $|K|$.

Aquest resultat es pot justificar a partir del teorema 2.2 de [20], on es demostra que si $|K|$ és connex, llavors $H_0(K) \cong \mathbb{Z}$. A partir d'aquest cas particular, es dedueix que en el cas general, el grup $H_0(K)$ és isomorf a $\mathbb{Z}^r$, amb r igual al nombre de components connexes.

El grup $H_1(K)$ detecta informació topològica que no és visible des de la connexió de vèrtexs, i té a veure amb la presència de cicles o forats de dimensió 1. A nivell algebraic, $H_1(K)$ es defineix com

$$H_1(K) = \text{Ker}(\partial_1) / \text{Im}(\partial_2).$$

Els elements de $\text{Ker}(\partial_1)$ són cadenes de 1-símplexs (arestes) que tenen vora nul·la. Això vol dir que la suma de les cares (vèrtexs) de la cadena es cancel·la, i per tant aquestes cadenes formen camins tancats anomenats *cicles*.

D'altra banda, $\text{Im}(\partial_2)$ està format per les vores de les cadenes de 2-símplexs (triangles). Aquestes vores són precisament aquells cicles que es poden “omplir” amb un conjunt de 2-símplexs dins del complex.

Per tant, el quocient $\text{Ker}(\partial_1) / \text{Im}(\partial_2)$ identifica els cicles que no poden ser escrits com a vores de cap combinació de 2-símplexs del complex. Aquests cicles representen forats de dimensió 1, i formen la base del grup $H_1(K)$. En conseqüència, $H_1(K)$ capta la presència de cicles dins la topologia del complex.

De manera més general, per a qualsevol $n \geq 1$, el grup d'homologia $H_n(K)$ mesura la presència de “forats” de dimensió n a la realització geomètrica $|K|$. És a dir, $H_n(K)$ està format per classes de cicles n -dimensionals que no són frontera de cap combinació de $(n + 1)$ -símplexs dins del complex, i per tant no es poden omplir dins del complex. Aquesta interpretació geomètrica i la seva construcció algebraica es troben detallades a la secció 2.1 de [20].

Capítol 2

Complex de cadenes ínfim i suprem

En aquest capítol es presenten les definicions de complex de cadenes ínfim i complex de cadenes suprem. Aquestes construccions permeten, a partir d'un subgrup graduat donat, construir el subcomplex més gran contingut o el més petit que el conté, tot preservant l'estructura de complex de cadenes. Aquestes eines són essencials per definir, en la secció següent, l'homologia d'hipergrafs. Les definicions i les proposicions que apareixen en aquest capítol s'han extret de [5].

2.1 Definicions

Donat un complex de cadenes $\{C_*, \partial_*\}$, un grup abelià graduat $\{D_n\}_{n \geq 0}$ s'anomena *subordinat* a $\{C_*, \partial_*\}$ si, per a tot $n \geq 0$, el grup D_n és un subgrup de C_n .

Definició 2.1.1. Donat un grup abelià graduat $D = \{D_n\}_{n \geq 0}$ subordinat a un complex de cadenes $\{C_*, \partial_*\}$, es defineix el *complex de cadenes ínfim* $\text{Inf}_*(D_*, \{C_*, \partial_*\})$ com

$$\text{Inf}_n(D_*, \{C_*, \partial_*\}) = \sum \{C'_n \mid C'_n \subseteq D_n \text{ i } \{C'_*, \partial_*\} \text{ és un subcomplex de } \{C_*, \partial_*\}\}.$$

En aquesta expressió, la suma es refereix a suma de subgrups; és a dir, és el més petit subgrup de D_n que conté tots els subgrups C'_n . Aquest complex $\text{Inf}_n(D_*, \{C_*, \partial_*\})$ és el major subcomplex de cadenes $\{C'_*, \partial_*\}$ de $\{C_*, \partial_*\}$ tal que cada grup de cadenes C'_n està contingut en D_n .

D'aquesta definició es dedueix que, si $\{D_*, \partial_*\}$ és un subcomplex de $\{C_*, \partial_*\}$, aleshores $\text{Inf}_*(D_*, \{C_*, \partial_*\}) = \{D_*, \partial_*\}$.

Proposició 2.1.2. *Segui $\{C_*, \partial_*\}$ un complex de cadenes i $\{D_n\}_{n \geq 0}$ un grup abelià graduat subordinat a $\{C_*, \partial_*\}$. El grup de cadenes de dimensió n del complex de cadenes ínfim ve donat per*

$$\text{Inf}_n(D_*, \{C_*, \partial_*\}) = D_n \cap \partial_n^{-1}(D_{n-1}).$$

Demostració. Es vol provar que els grups $I_n := D_n \cap \partial_n^{-1}(D_{n-1})$ defineixen el complex de cadenes ínfim tal com s'ha donat a la definició 2.1.1, és a dir, com el màxim subcomplex $\{C'_*, \partial_*\}$ de $\{C_*, \partial_*\}$ tal que $C'_n \subseteq D_n$ per a tot $n \geq 0$.

Primer cal provar que $\{I_*, \partial_*\}$ forma un subcomplex, és a dir, que per a tot $n \geq 1$,

$$\partial_n(I_n) \subseteq I_{n-1}.$$

Sigui $x \in I_n$. Per definició, $x \in D_n$ i $\partial_n(x) \in D_{n-1}$. Com que $\{C_*, \partial_*\}$ és un complex de cadenes, es compleix que

$$\partial_{n-1}(\partial_n(x)) = 0 \in D_{n-2}.$$

Així, $\partial_n(x) \in \text{Ker}(\partial_{n-1}) \subseteq \partial_{n-1}^{-1}(D_{n-2})$, i com que $\partial_n(x) \in D_{n-1}$, aleshores

$$\partial_n(x) \in D_{n-1} \cap \partial_{n-1}^{-1}(D_{n-2}) = I_{n-1}.$$

Així doncs, $\partial_n(I_n) \subseteq I_{n-1}$, i per tant $\{I_*, \partial_*\}$ és un subcomplex de $\{C_*, \partial_*\}$. Resta veure que aquest subcomplex és el més gran possible tal que $I_n \subseteq D_n$ per a tot $n \geq 0$.

Sigui $\{C'_*, \partial_*\}$ un subcomplex de $\{C_*, \partial_*\}$ tal que $C'_n \subseteq D_n$ per a tot $n \geq 0$. Aleshores

$$\partial_n(C'_n) \subseteq C'_{n-1} \subseteq D_{n-1} \Rightarrow C'_n \subseteq \partial_n^{-1}(D_{n-1})$$

i com que $C'_n \subseteq D_n$, tenim que

$$C'_n \subseteq D_n \cap \partial_n^{-1}(D_{n-1}) = I_n.$$

Per tant, $\{C'_n\}_{n \geq 0} \subseteq \{I_n\}_{n \geq 0}$. Els grups $I_n = D_n \cap \partial_n^{-1}(D_{n-1})$ defineixen el màxim subcomplex de $\{C_*, \partial_*\}$ tal que els seus grups de cadenes es troben continguts a $\{D_n\}_{n \geq 0}$, tal com calia demostrar. $\square$

Definició 2.1.3. Donat un grup abelià graduat $\{D_n\}_{n \geq 0}$ subordinat a un complex de cadenes $\{C_*, \partial_*\}$, es defineix el *complex de cadenes suprem* $\text{Sup}_*(D_*, \{C_*, \partial_*\})$ com

$$\text{Sup}_n(D_*, \{C_*, \partial_*\}) = \bigcap \{C'_n \mid D_n \subseteq C'_n \text{ i } \{C'_*, \partial_*\} \text{ és un subcomplex de } \{C_*, \partial_*\}\}.$$

Aquest subcomplex és el menor subcomplex de cadenes $\{C'_*, \partial_*\}$ de $\{C_*, \partial_*\}$ tal que cada grup de cadenes C'_n conté D_n .

Si $\{D_*, \partial_*\}$ és un subcomplex de $\{C_*, \partial_*\}$, aleshores $\text{Sup}_*(D_*, \{C_*, \partial_*\}) = \{D_*, \partial_*\}$.

Proposició 2.1.4. Sigui $\{C_*, \partial_*\}$ un complex de cadenes i $\{D_n\}_{n \geq 0}$ un grup abelià graduat subordinat a $\{C_*, \partial_*\}$. El grup de cadenes de dimensió n del complex de cadenes suprem ve donat per

$$\text{Sup}_n(D_*, \{C_*, \partial_*\}) = D_n + \partial_{n+1}(D_{n+1}).$$

Demostració. Es vol provar que els grups

$$S_n := D_n + \partial_{n+1}(D_{n+1})$$

defineixen el complex de cadenes suprem tal com s'ha donat a la definició 2.1.3, és a dir, com el mínim subcomplex de $\{C_*, \partial_*\}$ que conté $\{D_n\}_{n \geq 0}$, és a dir, tal que $D_n \subseteq C'_n$ per a tot $n \geq 0$.

Per tal de veure que $\{S_*, \partial_*\}$ és un subcomplex de $\{C_*, \partial_*\}$, cal comprovar que per a tot $n \geq 0$:

$$\partial_n(S_n) \subseteq S_{n-1}.$$

Sigui $x \in D_n$ i $y \in D_{n+1}$. Com que $S_n = D_n + \partial_{n+1}(D_{n+1})$, qualsevol element de S_n és de la forma $x + \partial_{n+1}(y)$, i per tant

$$\partial_n(x + \partial_{n+1}(y)) = \partial_n(x) + \partial_n(\partial_{n+1}(y)).$$

Atès que $\{C_*, \partial_*\}$ és un complex de cadenes, es té que $\partial_n \circ \partial_{n+1} = 0$, i en conseqüència

$$\partial_n(x + \partial_{n+1}(y)) = \partial_n(x).$$

Ara bé, com que $x \in D_n$, se'n dedueix que $\partial_n(x) \in \partial_n(D_n) \subseteq S_{n-1}$. Així, $\partial_n(x + \partial_{n+1}(y)) \in S_{n-1}$, i per tant

$$\partial_n(S_n) \subseteq S_{n-1}.$$

Així doncs, $\{S_*, \partial_*\}$ és un subcomplex de $\{C_*, \partial_*\}$. Resta veure que és el mínim amb aquesta propietat. Si $\{C'_*, \partial_*\}$ és un subcomplex tal que $D_n \subseteq C'_n$ per a tot $n \geq 0$, aleshores també $\partial_{n+1}(D_{n+1}) \subseteq \partial_{n+1}(C'_{n+1}) \subseteq C'_n$, i per tant

$$D_n + \partial_{n+1}(D_{n+1}) \subseteq C'_n.$$

Aleshores, per a tot $n \geq 0$, es compleix que $S_n \subseteq C'_n$, i així $\{S_*, \partial_*\}$ és el menor subcomplex de $\{C_*, \partial_*\}$ que conté $\{D_n\}_{n \geq 0}$. $\square$

Proposició 2.1.5. *Sigui $\{C_*, \partial_*\}$ un complex de cadenes i $\{D_n\}_{n \geq 0}$ un grup abelià graduat subordinat a $\{C_*, \partial_*\}$. Sigui $\text{Inf}_* = \text{Inf}_n(D_*, \{C_*, \partial_*\})$ i $\text{Sup}_* = \text{Sup}_n(D_*, \{C_*, \partial_*\})$ els complexos de cadenes ínfim i suprem definits a 2.1.1 i 2.1.3 respectivament. Aleshores, per a tot $n \geq 0$, es compleix que*

$$H_n(\text{Inf}_*) \cong H_n(\text{Sup}_*).$$

Demostració. Segons les proposicions 2.1.2 i 2.1.4, els grups de cadenes del complex ínfim i del complex suprem es defineixen com $\text{Inf}_n = D_n \cap \partial_n^{-1}(D_{n-1})$ i $\text{Sup}_n = D_n + \partial_{n+1}(D_{n+1})$ respectivament.

L'objectiu és demostrar que els grups d'homologia associats a aquests dos complexos són isomorfs. Per definició, el grup d'homologia del complex ínfim de grau n és

$$H_n(\text{Inf}_*) := \frac{\text{Ker}(\partial_n|_{\text{Inf}_n})}{\text{Im}(\partial_{n+1}|_{\text{Inf}_{n+1}})}.$$

Com que $\text{Inf}_n \subseteq D_n$ per a tot $n \geq 0$ aleshores per qualsevol element $x \in \text{Inf}_n$ tal que $\partial_n(x) = 0$ pertany automàticament a D_n , s'obté que $\text{Ker}(\partial_n|_{\text{Inf}_n}) \subseteq \text{Ker}(\partial_n|_{D_n})$. D'altra banda, veiem que $\text{Ker}(\partial_n|_{D_n}) \subseteq \text{Ker}(\partial_n|_{\text{Inf}_n})$. Si $x \in \text{Ker}(\partial_n|_{D_n})$, llavors $x \in D_n$ i $\partial_n(x) = 0 \in D_{n-1}$. Això implica que $x \in \partial_n^{-1}(D_{n-1})$, i com que $x \in D_n$ s'obté que $x \in D_n \cap \partial_n^{-1}(D_{n-1}) = \text{Inf}_n$. Aleshores $x \in \text{Ker}(\partial_n|_{\text{Inf}_n})$, i $\text{Ker}(\partial_n|_{D_n}) \subseteq \text{Ker}(\partial_n|_{\text{Inf}_n})$. D'aquestes dues inclusions es conclou que $\text{Ker}(\partial_n|_{D_n}) = \text{Ker}(\partial_n|_{\text{Inf}_n})$.

Ara es vol veure que $\text{Im}(\partial_{n+1}|_{\text{Inf}_{n+1}}) = D_n \cap \partial_{n+1}(D_{n+1})$. Sigui $z \in D_n \cap \partial_{n+1}(D_{n+1})$, aleshores existeix $y \in D_{n+1}$ tal que $\partial_{n+1}(y) = z$, i com que $z \in D_n$, es dedueix que $\partial_{n+1}(y) \in D_n$. Per tant, $y \in \partial_{n+1}^{-1}(D_n)$, i com que $y \in D_{n+1}$, llavors $y \in D_{n+1} \cap \partial_{n+1}^{-1}(D_n) = \text{Inf}_{n+1}$. Així $z = \partial_{n+1} y \in \text{Im}(\partial_{n+1}|_{\text{Inf}_{n+1}})$ i $D_n \cap \partial_{n+1}(D_{n+1}) \subseteq \text{Im}(\partial_{n+1}|_{\text{Inf}_{n+1}})$. D'altra banda, si $y \in \text{Inf}_{n+1}$, llavors $y \in D_{n+1}$ i $\partial_{n+1}(y) \in D_n$, per tant $\partial_{n+1}(y) = D_n \cap \partial_{n+1}(D_{n+1})$, aleshores $\text{Im}(\partial_{n+1}|_{\text{Inf}_{n+1}}) \subseteq D_n \cap \partial_{n+1}(D_{n+1})$. D'aquestes dues inclusions s'obté que $\text{Im}(\partial_{n+1}|_{\text{Inf}_{n+1}}) = D_n \cap \partial_{n+1}(D_{n+1})$.

Com a conseqüència de les dues igualtats anteriors, s'obté que

$$H_n(\text{Inf}_*) = \frac{\text{Ker}(\partial_n|_{D_n})}{D_n \cap \partial_{n+1}(D_{n+1})}.$$

El grup d'homologia de dimensió n del complex suprem és

$$H_n(\text{Sup}_*) := \frac{\text{Ker}(\partial_n|_{\text{Sup}_n})}{\text{Im}(\partial_{n+1}|_{\text{Sup}_{n+1}})}.$$

Ara es vol veure que $\text{Ker}(\partial_n|_{\text{Sup}_n}) = \text{Ker}(\partial_n|_{D_n}) + \partial_{n+1}(D_{n+1})$ i que $\text{Im}(\partial_{n+1}|_{\text{Sup}_{n+1}}) = \partial_{n+1}(D_{n+1})$.

Veiem la primera igualtat. Sigui $z \in \text{Sup}_n$, aleshores es pot escriure com $z = x + \partial_{n+1}(y)$ amb $x \in D_n$ i $y \in D_{n+1}$. Aleshores $\partial_n(z) = \partial_n(x + \partial_{n+1}(y)) = \partial_n(x) + \partial_n(\partial_{n+1}(y)) = \partial_n(x)$, ja que $\partial_n \circ \partial_{n+1} = 0$ per la propietat del complex. Per tant, si $z \in \text{Ker}(\partial_n|_{\text{Sup}_n})$, aleshores $\partial_n(z) = 0$ i per tant $\partial_n(x) = 0$, és a dir, $x \in \text{Ker}(\partial_n|_{D_n})$. Això mostra que $\text{Ker}(\partial_n|_{\text{Sup}_n}) \subseteq \text{Ker}(\partial_n|_{D_n}) + \partial_{n+1}(D_{n+1})$. Per veure la inclusió inversa, sigui $x \in \text{Ker}(\partial_n|_{D_n})$ i $y \in D_{n+1}$. Llavors $z = x + \partial_{n+1}(y) \in \text{Sup}_n$ i $\partial_n(z) = \partial_n(x) + \partial_n(\partial_{n+1}(y)) = 0 + 0 = 0$, i per tant $z \in \text{Ker}(\partial_n|_{\text{Sup}_n})$. Així, qualsevol element de $\text{Ker}(\partial_n|_{D_n}) + \partial_{n+1}(D_{n+1})$ pertany a $\text{Ker}(\partial_n|_{\text{Sup}_n})$, i d'aquí se'n dedueix la igualtat $\text{Ker}(\partial_n|_{\text{Sup}_n}) = \text{Ker}(\partial_n|_{D_n}) + \partial_{n+1}(D_{n+1})$.

Pel que fa a la imatge, com que $\text{Sup}_{n+1} = D_{n+1} + \partial_{n+2}(D_{n+2})$, qualsevol element $z \in \text{Sup}_{n+1}$ es pot escriure com $z = x + \partial_{n+2}(w)$ amb $x \in D_{n+1}$ i $w \in D_{n+2}$. Llavors, $\partial_{n+1}(z) = \partial_{n+1}(x + \partial_{n+2}(w)) = \partial_{n+1}(x) + \partial_{n+1}(\partial_{n+2}(w)) = \partial_{n+1}(x)$, ja que $\partial_{n+1} \circ \partial_{n+2} = 0$. Això mostra que $\partial_{n+1}(z) \in \partial_{n+1}(D_{n+1})$, i per tant $\text{Im}(\partial_{n+1}|_{\text{Sup}_{n+1}}) \subseteq \partial_{n+1}(D_{n+1})$. Per a la inclusió inversa, si $x \in D_{n+1}$, com que $x \in \text{Sup}_{n+1}$ per definició, es té que $\partial_{n+1}(x) \in \text{Im}(\partial_{n+1}|_{\text{Sup}_{n+1}})$, i així $\partial_{n+1}(D_{n+1}) \subseteq \text{Im}(\partial_{n+1}|_{\text{Sup}_{n+1}})$. D'aquestes dues inclusions es conclou que

$$\text{Im}(\partial_{n+1}|_{\text{Sup}_{n+1}}) = \partial_{n+1}(D_{n+1}).$$

Per completar la demostració, es recorda que el segon teorema d'isomorfia estableix que donat un grup abelià G , i H i N dos subgrups de G , aleshores

$$\frac{H + N}{N} \cong \frac{H}{H \cap N}.$$

Aquest resultat s'aplica al cas $H = \text{Ker}(\partial_n|_{D_n})$ i $N = \partial_{n+1}(D_{n+1})$, dins el grup abelià C_n . Així, l'homologia del complex suprem es pot escriure com

$$H_n(\text{Sup}_*(D_*, C_*)) = \frac{\text{Ker}(\partial_n|_{D_n}) + \partial_{n+1}(D_{n+1})}{\partial_{n+1}(D_{n+1})},$$

i, aplicant el teorema d'isomorfia,

$$\frac{\text{Ker}(\partial_n|_{D_n}) + \partial_{n+1}(D_{n+1})}{\partial_{n+1}(D_{n+1})} \cong \frac{\text{Ker}(\partial_n|_{D_n})}{\text{Ker}(\partial_n|_{D_n}) \cap \partial_{n+1}(D_{n+1})}.$$

D'altra banda, s'ha vist anteriorment que l'homologia del complex ínfim és

$$H_n(\text{Inf}_*(D_*, C_*)) = \frac{\text{Ker}(\partial_n|_{D_n})}{D_n \cap \partial_{n+1}(D_{n+1})},$$

i atès que per a tot $x \in D_n \cap \partial_{n+1}(D_{n+1})$, existeix $y \in D_{n+1}$ tal que $x = \partial_{n+1}(y)$, i per la propietat del complex de cadenes es té que $\partial_n(x) = \partial_n(\partial_{n+1}(y)) = 0$, es dedueix que $x \in \text{Ker}(\partial_n|_{D_n})$. Així, $D_n \cap \partial_{n+1}(D_{n+1}) \subseteq \text{Ker}(\partial_n|_{D_n}) \cap \partial_{n+1}(D_{n+1})$. L'altra inclusió és immediata, ja que $\text{Ker}(\partial_n|_{D_n}) \subseteq D_n$. Per tant, $\text{Ker}(\partial_n|_{D_n}) \cap \partial_{n+1}(D_{n+1}) = D_n \cap \partial_{n+1}(D_{n+1})$. Per tant, es conclou que

$$H_n(\text{Sup}_*) \cong H_n(\text{Inf}_*) \quad \text{per a tot } n \geq 0,$$

que acaba la demostració. $\square$

Capítol 3

Homologia d'hipergrafs

En aquest capítol es descriuen dues construccions diferents d'homologia associades a un hipergraf, basades en definicions i resultats introduïts per diversos autors. Concretament, es segueix el treball desenvolupat per Bressan, Li, Ren i Wu [5], així com la noció de complex simplicial associat a un hipergraf proposada per Parks i Lipscomb [22], construcció que en aquest treball s'anomenarà clausura simplicial.

3.1 Clausura simplicial d'un hipergraf

El concepte d'hipergraf generalitza el de graf. Mentre que en un graf cada aresta uneix exactament dos vèrtexs, en un hipergraf una hiperaresta pot reunir qualsevol nombre de vèrtexs.

Definició 3.1.1. Un *hipergraf* és una parella $(V_{\mathcal{H}}, \mathcal{H})$ on $V_{\mathcal{H}}$ és un conjunt i $\mathcal{H}$ és un subconjunt de $\mathcal{P}(V_{\mathcal{H}})$. Els elements de $V_{\mathcal{H}}$ s'anomenen *vèrtexs* i els de $\mathcal{H}$, *hiperarestes*.

Tal com s'ha introduït a la secció 1.2, el conjunt de parts $\mathcal{P}(V_{\mathcal{H}})$ representa la col·lecció de tots els subconjunts no buits del conjunt que conté tots dels vèrtexs.

Si v és un vèrtex, el conjunt $\{v\}$ no és necessàriament una hiperaresta. Això contrasta amb el cas dels complexos simplicials, on per definició $\{v\}$ sempre forma part del complex.

Donat un hipergraf $(V_{\mathcal{H}}, \mathcal{H})$, un *subhipergraf* de $(V_{\mathcal{H}}, \mathcal{H})$ és un hipergraf $(V'_{\mathcal{H}}, \mathcal{H}')$ tal que $V'_{\mathcal{H}} \subseteq V_{\mathcal{H}}$, $\mathcal{H}' \subseteq \mathcal{H}$, i per a tot $\sigma \in \mathcal{H}'$, es compleix que $\sigma \subseteq V'_{\mathcal{H}}$.

Per a $n \geq 0$, s'anomenarà *n-hiperaresta* a tota hiperaresta formada per $n + 1$ vèrtexs. Es denota per $\mathcal{H}_n$ la col·lecció de totes les *n-hiperarestes* d'un hipergraf $\mathcal{H}$. Si, donat un hipergraf $\mathcal{H}$, se satisfà que, si per a qualsevol hiperaresta $\sigma \in \mathcal{H}$, qualsevol subconjunt no buit de σ també és una hiperaresta de $\mathcal{H}$, llavors $\mathcal{H}$ compleix les condicions de complex simplicial donades a la secció 1.2. Per tant, $\mathcal{H}$ és un complex simplicial i les hiperarestes de $\mathcal{H}$ són símplexs. En aquest cas, ja se sap com calcular els grups d'homologia, ja que en complir les condicions de complex simplicial es aplicar directament la definició d'homologia simplicial exposada a la secció 1.2. Cal destacar que tot complex simplicial és, per definició, un cas particular d'hipergraf. Això és degut al fet que un complex simplicial és un hipergraf que satisfà la condició addicional de ser tancat per subconjunts.

Si, donat un hipergraf $\mathcal{H}$, no se satisfà la condició anterior, és a dir, que no és un complex simplicial, aleshores podem associar-li un complex simplicial. La *clausura simplicial* de $\mathcal{H}$,

denotada per $K_{\mathcal{H}}$, és el menor complex simplicial tal que les hiperarestes de $\mathcal{H}$ són simplexs de $K_{\mathcal{H}}$. Més precisament, l'agrupació de tots els simplexs de $K_{\mathcal{H}}$ consisteix en tots els subconjunts no buits $\tau \subseteq \sigma$, per a tot $\sigma \in \mathcal{H}$. Les primeres definicions d'homologia per a hipergrafs es van basar en considerar la clausura simplicial de l'hipergraf.

Es defineix l'homologia de la clausura simplicial d'un hipergraf $\mathcal{H}$ com

$$H_n^K(\mathcal{H}) := H_n(K_{\mathcal{H}}).$$

Exemple 3.1.2. Donat l'hipergraf $\mathcal{H} = \{\{v_0\}, \{v_1\}, \{v_2\}, \{v_0, v_1\}, \{v_0, v_1, v_2\}\}$, $K_{\mathcal{H}}$ és el menor complex simplicial que conté totes les hiperarestes de $\mathcal{H}$ com a simplexs. Per tant, per obtenir $K_{\mathcal{H}}$ cal afegir-hi tots els subconjunts no buits de cada hiperaresta. Els que cal afegir-hi són $\{v_0, v_2\}$ i $\{v_1, v_2\}$. Per tant,

$$K_{\mathcal{H}} = \{\{v_0\}, \{v_1\}, \{v_2\}, \{v_0, v_1\}, \{v_0, v_2\}, \{v_1, v_2\}, \{v_0, v_1, v_2\}\}.$$

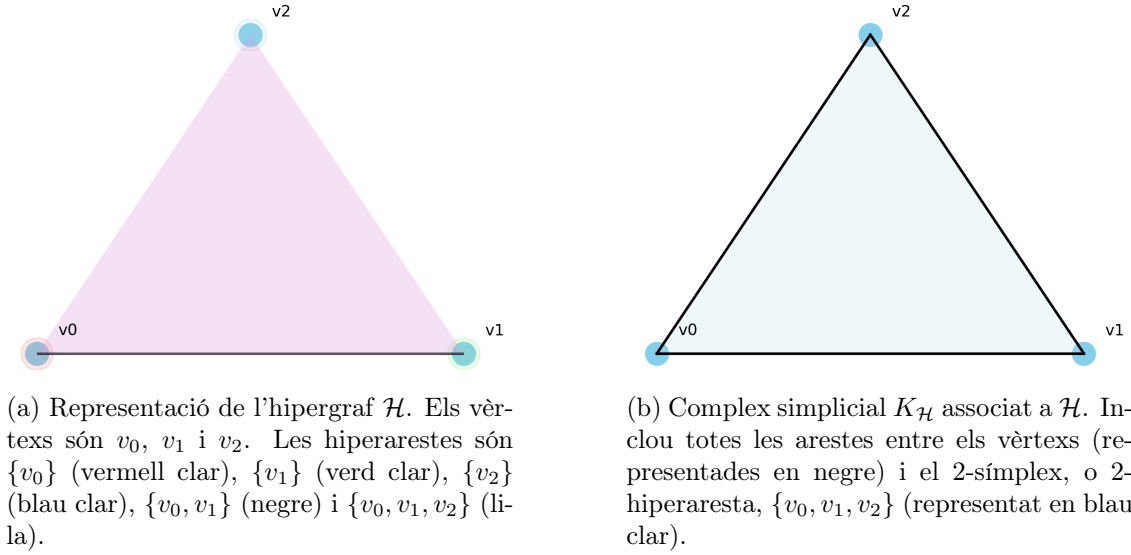


Figura 3.1: A l'esquerra es troba representat l'hipergraf $\mathcal{H}$ i a la dreta el seu complex simplicial associat $K_{\mathcal{H}}$.

Tot seguit es calcula el grup d'homologia n -èsim de la clausura simplicial de l'hipergraf.

Seguint la notació i les definicions de la secció 1.2, s'obtenen els següents grups de cadenes i aplicacions vora per a la clausura simplicial $K_{\mathcal{H}}$:

- $C_0(K_{\mathcal{H}}) = \mathbb{Z}(v_0) \oplus \mathbb{Z}(v_1) \oplus \mathbb{Z}(v_2)$. És el grup generat pels 0-símplexs $\{v_0\}$, $\{v_1\}$ i $\{v_2\}$.
- $C_1(K_{\mathcal{H}}) = \mathbb{Z}(v_0v_1) \oplus \mathbb{Z}(v_0v_2) \oplus \mathbb{Z}(v_1v_2)$. És el grup generat pels 1-símplexs $\{v_0, v_1\}$, $\{v_0, v_2\}$ i $\{v_1, v_2\}$.
- $C_2(K_{\mathcal{H}}) = \mathbb{Z}(v_0v_1v_2)$. És el grup generat pel 2-símplex $\{v_0, v_1, v_2\}$.

Les aplicacions vora són:

- $\partial_0 : C_0 \longrightarrow 0$ tal que $\partial_0(\{v_i\}) = 0$ amb $i = 0, 1, 2$. Per tant, $\text{Ker}(\partial_0) = C_0$ i la imatge és $\text{Im}(\partial_0) = 0$.

- $\partial_1 : C_1 \rightarrow C_0$ amb $\partial_1(\{v_0, v_1\}) = \{v_1\} - \{v_0\}$, $\partial_1(\{v_0, v_2\}) = \{v_2\} - \{v_0\}$ i $\partial_1(\{v_1, v_2\}) = \{v_1\} - \{v_0\}$. Així doncs la imatge és $\text{Im}(\partial_1) = \langle \{v_1\} - \{v_0\}, \{v_2\} - \{v_0\}, \{v_1\} - \{v_0\} \rangle$ i $\text{Ker}(\partial_1) = \langle \{v_0, v_1\} - \{v_0, v_2\} + \{v_1, v_2\} \rangle$.
- $\partial_2 : C_2 \rightarrow C_1$ amb $\partial_2(\{v_0, v_1, v_2\}) = \{v_0, v_1\} - \{v_0, v_2\} + \{v_1, v_2\}$. Per tant, $\text{Im}(\partial_2) = \langle \{v_0, v_1\} - \{v_0, v_2\} + \{v_1, v_2\} \rangle$ i $\text{Ker}(\partial_2) = 0$.

Ara ja es poden calcular els grups d'homologia H_0^K , H_1^K i H_2^K :

$$H_0^K = \text{Ker}(\partial_0)/\text{Im}(\partial_1) = \frac{C_0}{\langle v_1 - v_0, v_2 - v_0 \rangle} \cong \mathbb{Z}.$$

$$H_1^K = \text{Ker}(\partial_1)/\text{Im}(\partial_2) = \frac{\langle \{v_0, v_1\} - \{v_0, v_2\} + \{v_1, v_2\} \rangle}{\langle \{v_0, v_1\} - \{v_0, v_2\} + \{v_1, v_2\} \rangle} = 0.$$

$$H_2^K = \text{Ker}(\partial_2)/\text{Im}(\partial_3) = \text{Ker}(\partial_2) = 0.$$

Sigui $\mathcal{H} = (V_{\mathcal{H}}, \mathcal{H})$ un hipergraf. Un *camí* de longitud $k \geq 0$ entre dos vèrtexs $v_0, v_k \in V_{\mathcal{H}}$ és una successió $v_0, E_1, v_1, E_2, \dots, E_k, v_k$, on $E_i \in \mathcal{H}$ per a $i = 1, \dots, k$ i $\{v_{i-1}, v_i\} \subseteq E_i$. Dos vèrtexs $u, v \in V_{\mathcal{H}}$ es diuen *connectats* si existeix un camí que els uneix (el cas $k = 0$ dóna el camí trivial (u)).

Aquesta relació de connexió entre vèrtexs mitjançant camins alternats compleix les propietats de reflexivitat, simetria i transitivitat, i per tant defineix una relació d'equivalència sobre el conjunt de vèrtexs. Les seves classes d'equivalència corresponen a subconjunts de vèrtexs que estan connectats entre si, i reben el nom de *components connexes* de l'hipergraf. Denotem per $c(\mathcal{H})$ el nombre de components connexes.

Lema 3.1.3. *Sigui $K_{\mathcal{H}}$ la clausura simplicial d'un hipergraf $\mathcal{H}$. Dos vèrtexs estan connectats en $\mathcal{H}$ si i només si ho estan a $K_{\mathcal{H}}$. En conseqüència,*

$$c(\mathcal{H}) = c(K_{\mathcal{H}}) \quad i \quad \text{rank } H_0^K(K_{\mathcal{H}}) = c(\mathcal{H}).$$

Demostració. Direcció ($\Rightarrow$). Un camí $v_0, E_1, v_1, \dots, E_k, v_k$ en $\mathcal{H}$ satisfà $\{v_{i-1}, v_i\} \subseteq E_i \in \mathcal{H}$. Com que $\{v_{i-1}, v_i\}$ és un subconjunt no buit d'una hiperaresta, és un 1-símplex de $K_{\mathcal{H}}$; la mateixa successió de vèrtexs és, per tant, camí a $K_{\mathcal{H}}$.

Direcció ($\Leftarrow$). Si $v_0, \dots, v_k$ és un camí de $K_{\mathcal{H}}$, cada aresta $\{v_{i-1}, v_i\}$ està continguda en alguna $E_i \in \mathcal{H}$. Per tant, $v_0, E_1, v_1, \dots, E_k, v_k$ és un camí a $\mathcal{H}$.

Com que la relació de connexió coincideix, els dos objectes tenen el mateix nombre de components, i el rang de H_0 de qualsevol complex simplicial és precisament el nombre de components connexes. $\square$

3.2 Homologia restringida

Tot seguit es presenta l'homologia restringida, que constitueix el marc principal d'estudi en aquest treball. Aquesta permet analitzar la topologia d'un hipergraf de manera directa, sense necessitat de passar per la seva clausura simplicial. Seguin la construcció que

proposen Bressan, Li, Ren i Wu [5], es pot definir l'homologia restringida d'un hipergraf $\mathcal{H}$ a partir dels complexos de cadenes ínfim i suprem

$$\begin{aligned}\text{Inf}_*(\mathcal{H}) &= \text{Inf}(G(\mathcal{H}_*), G((K_{\mathcal{H}})_*)), \\ \text{Sup}_*(\mathcal{H}) &= \text{Sup}(G(\mathcal{H}_*), G((K_{\mathcal{H}})_*)),\end{aligned}$$

on $G(\mathcal{H}_*)$ és el grup graduat generat per les hiperarestes de $\mathcal{H}$ i $G((K_{\mathcal{H}})_*)$ és el complex de cadenes de la clausura simplicial associat $\mathcal{H}$.

Definició 3.2.1. Donat un hipergraf $\mathcal{H}$ i un grup abelià G , és defineix el *grup d'homologia restringida n -èsim* de $\mathcal{H}$ amb coeficients a G com

$$H_n(\mathcal{H}) = H_n(\text{Inf}_*(\mathcal{H})).$$

Cal destacar que l'homologia del complex de cadenes ínfim i la del complex suprem són isomorfes (vist a la proposició 2.1.5). Això garanteix que l'homologia restringida d'un hipergraf es pot calcular indistintament a partir del complex ínfim o del suprem.

Si $\mathcal{H}$ és un complex simplicial, aleshores l'homologia restringida de $H_*(\mathcal{H})$ coincideix amb l'homologia simplicial de la clausura simplicial de $\mathcal{H}$.

A continuació es calcula l'homologia de l'hipergraf utilitzat per l'exemple 3.1.2. Aquest càlcul serveix tant per il·lustrar com es determina l'homologia d'un hipergraf amb un exemple concret, com per poder comparar el resultat amb l'homologia de la clausura simplicial que s'ha obtingut a la secció anterior.

Exemple 3.2.2. Sigui $\mathcal{H}$ l'hipergraf $\mathcal{H} = \{\{v_0\}, \{v_1\}, \{v_2\}, \{v_0, v_1\}, \{v_0, v_1, v_2\}\}$. Per calcular l'homologia restringida de $\mathcal{H}$, és necessari trobar el complex de cadenes ínfim $\text{Inf}(G(\mathcal{H}_*), G((K_{\mathcal{H}})_*))$. En primer lloc, s'ha de determinar explícitament el grup graduat $G(\mathcal{H}_*)$ i el complex de cadenes $G((K_{\mathcal{H}})_*)$:

- Grup graduat $G(\mathcal{H}_*)$:
 - $G_0(\mathcal{H}) := G(\mathcal{H}_0) = \mathbb{Z}(v_0) \oplus \mathbb{Z}(v_1) \oplus \mathbb{Z}(v_2)$. És el grup abelià generat per les 0-hiperarestes $\{v_0\}, \{v_1\}$ i $\{v_2\}$.
 - $G_1(\mathcal{H}) := G(\mathcal{H}_1) = \mathbb{Z}(v_0v_1)$. És el grup abelià generat per les 1-hiperarestes que en aquest cas és només una, $\{v_0, v_1\}$.
 - $G_2(\mathcal{H}) := G(\mathcal{H}_2) = \mathbb{Z}(v_0v_1v_2)$. És el grup abelià generat per les 2-hiperarestes que en aquest cas torna a ser només una, $\{v_0, v_1, v_2\}$.
- Complex de cadenes $G((K_{\mathcal{H}})_*)$. Aquest és el complex de cadenes associat a $K_{\mathcal{H}}$. $G_n(K_{\mathcal{H}}) := G((K_{\mathcal{H}})_n) = C_n(K_{\mathcal{H}})$ per $n = 0, 1, 2$, on C_n són els grups abelians definits a l'exemple 3.1.2 i amb les aplicacions vora ∂_i també definides a l'exemple 3.1.2.

$$G((K_{\mathcal{H}})_*) : C_2(K_{\mathcal{H}}) \xrightarrow{\partial_2} C_1(K_{\mathcal{H}}) \xrightarrow{\partial_1} C_0(K_{\mathcal{H}}) \xrightarrow{\partial_0} 0.$$

Utilitzant la proposició 2.1.2 ja es pot trobar el complex de cadenes ínfim associat a l'hipergraf $\mathcal{H}$ a partir del grup graduat $G(\mathcal{H}_*)$ i el complex de cadenes $G((K_{\mathcal{H}})_*)$.

$$\text{Inf}_0(G(\mathcal{H}_*), G(\mathcal{H}_*)) = G_0(\mathcal{H}) \cap \partial_0^{-1}(G_{-1}(\mathcal{H})) = G_0(\mathcal{H}) = \mathbb{Z}(v_0) \oplus \mathbb{Z}(v_1) \oplus \mathbb{Z}(v_2),$$

on $G_{-1}(\mathcal{H}) = 0$. Per tant, com ∂_0 ho envia tot al zero, tot $G_0(\mathcal{H})$ hi queda inclòs.

$$\text{Inf}_1(G(\mathcal{H}_*), G(\mathcal{H}_*)) = G_1(\mathcal{H}) \cap \partial_1^{-1}(G_0(\mathcal{H})) = G_1(\mathcal{H}) = \mathbb{Z}(v_0v_1),$$

on la penúltima igualtat és conseqüència de $\partial_1(v_0v_1) = (v_1) - (v_0) \in G_0(\mathcal{H})$.

$$\text{Inf}_2(G(\mathcal{H})_*, G(\mathcal{H}_*)) = G_2(\mathcal{H}) \cap \partial_2^{-1}(G_1(\mathcal{H})) = 0.$$

Aquí tenim que $\partial_2^{-1}(G_1(\mathcal{H})) = 0$, ja que $\partial(v_0v_1v_2) = (v_1v_2) - (v_0v_2) + (v_0v_1)$ i només $(v_0v_1) \in G_1(\mathcal{H})$. Així doncs, s'obté el complex de cadenes ínfim següent

$$\text{Inf}_*(\mathcal{H}) : 0 \longrightarrow G_1(\mathcal{H}) \xrightarrow{\partial_1} G_0(\mathcal{H}) \xrightarrow{\partial_0} 0,$$

i amb aquest ja es poden calcular els grups d'homologia de l'hipergraf:

$$H_0(\mathcal{H}) = \text{Ker}(\partial_0)/\text{Im}(\partial_1) = \frac{G_0(\mathcal{H})}{\langle (v_1) - (v_0) \rangle} = \mathbb{Z}([v_0]) \oplus \mathbb{Z}([v_2]) \cong \mathbb{Z}^2.$$

$$H_1(\mathcal{H}) = \text{Ker}(\partial_1)/\text{Im}(\partial_2) = \text{Ker}(\partial_1) = 0.$$

$$H_2(\mathcal{H}) = 0.$$

Tal com es pot observar comparant els resultats obtinguts a l'exemple 3.1.2 i a l'exemple 3.2.2, els grups d'homologia de l'hipergraf i els de la seva clausura simplicial no coincideixen. Per exemple, el grup H_0 en el segon cas és isomorf a $\mathbb{Z} \oplus \mathbb{Z}$, mentre que en el primer és isomorf a $\mathbb{Z}$. Aquest fet mostra que, en general, no es pot esperar que les dues homologies coincideixin.

En el model d'homologia d'un hipergraf, el grup $H_0(\mathcal{H})$ s'obté a partir del complex de cadenes ínfim definit com $\text{Inf}_n = G(\mathcal{H}_n) \cap \partial_n^{-1}(G(\mathcal{H}_{n-1}))$, on només s'hi consideren les hiperarestes explícites del hipergraf. El grup $G(\mathcal{H}_0)$ conté els vèrtexs que han estat declarats com a 0-hiperarestes, i només aquests es tenen en compte dins del càlcul. A partir d'aquesta construcció, es dedueix que el grup $H_0(\mathcal{H})$ detecta quins vèrtexs estan connectats entre si mitjançant hiperarestes de dimensió 1. Només les 1-hiperarestes (com $\{u, v\}$) introdueixen relacions entre vèrtexs dins del càlcul. Aquestes connexions només s'apliquen si tant u com v apareixen com a 0-hiperarestes.

Les hiperarestes de dimensió superior a 1 no afegeixen cap connexió nova entre vèrtexs a H_0 , llevat que els vèrtexs d'aquesta hiperaresta apareguin com a 1-hiperarestes dins de $\mathcal{H}_1$. Això implica que, si un hipergraf només conté hiperarestes de grau superior, pot passar que cap vèrtex estigui considerat com a connectat amb cap altre. A més, si no hi ha cap 0-hiperaresta, és a dir, si $G(\mathcal{H}_0) = 0$, aleshores $\text{Inf}_0 = 0$, i per tant $H_0(\mathcal{H}) = 0$. En aquest cas, l'homologia és trivial, ja que no hi ha cap vèrtex considerat dins del càlcul.

En resum, $H_0(\mathcal{H})$ compta el nombre de components connexes de vèrtexs connectats mitjançant hiperarestes de dimensió 1, sempre que aquests vèrtexs formin una 0-hiperaresta. Si dues parts del hipergraf no estan unides per aquest tipus d'hiperaresta, queden en components diferents des del punt de vista d'aquesta homologia. De fet, aquest comportament ja es reflecteix explícitament a la Proposició 3.5 de l'article [5], on s'indica que si tots els vèrtexs són 0-hiperarestes, llavors $\mathcal{H}_0 \cup \mathcal{H}_1$ forma un complex simplicial, i el grup $H_0(\mathcal{H})$ és isomorf a $\mathbb{Z}^{\oplus k}$, on k és el nombre de components connexes d'aquesta unió.

El grup $H_1(\mathcal{H})$, en el context de l'homologia restringida d'un hipergraf, mesura la presència de cicles o forats de dimensió 1 formats per 1-hiperarestes (és a dir, hiperarestes amb dos vèrtexs) que no són "tancats" per cap conjunt de 2-hiperarestes dins del mateix hipergraf. Aquesta detecció de cicles és possible encara que no hi hagi cap 0-hiperaresta o només n'hi hagi algunes. La condició de cicle només exigeix que la frontera algebraica global sigui nul·la respecte a les 0-hiperarestes existents.

Així, els vèrtexs implicats en un cicle no necessiten aparèixer tots com a 0-hiperarestes, n'hi ha prou amb una combinació d'arestes que tanqui algebraicament, i que aquesta vora nul·la caigui dins del grup generat per les 0-hiperarestes. D'altra banda, aquest cicle només queda “omplert” i, per tant, no apareix a l'homologia, si existeix un conjunt de 2-hiperarestes explícites la combinació de les quals tingui com a vora exactament aquell cicle. Això és rellevant especialment per a cicles amb més de tres arestes, ja que cap 2-hiperaresta única els pot contenir completament, cal una combinació suficient de 2-hiperarestes amb les cares adequades. Aquest fet contrasta amb el comportament de la clausura simplicial, on tota hiperaresta de dimensió superior genera automàticament totes les seves cares inferiors. Per tant, una 3-hiperaresta com (a, b, c, d) indueix la presència del 2-símplex (a, b, c) , que pot matar cicles formats per 1-hiperarestes com (a, b) , (b, c) , (a, c) . En canvi, en l'homologia incrustada, si el cicle no és literalment la vora d'un conjunt de 2-hiperarestes existents, no desapareix encara que tots els seus vèrtexs formin part d'una hiperaresta de grau superior.

Aquest mateix principi s'estén a dimensions més altes, el grup $H_n(H)$ recull classes de cicles d'ordre n (formats per n -hiperarestes amb frontera nul·la) que no són frontera de cap combinació de $(n+1)$ -hiperarestes explícites, i per tant representen forats de dimensió n dins de l'estructura del hipergraf.

Exemple 3.2.3. Sigui l'hipergraf $\mathcal{H} = \{(v_1v_2), (v_2v_3), (v_1v_3)\}$, on totes les hiperarestes són de dimensió 1, i no hi ha cap 0-hiperaresta ni cap hiperaresta de dimensió superior. És a dir, el conjunt $\mathcal{H}_0$ és buit, $\mathcal{H}_1 = \{(v_1v_2), (v_2v_3), (v_1v_3)\}$, i $\mathcal{H}_n = \emptyset$ per a $n \neq 1$.

A partir d'això, es poden construir els diferents grups implicats en el càlcul de la homologia restringida:

- Grup graduat: $G(\mathcal{H}_1) = \mathbb{Z}(v_1v_2) \oplus \mathbb{Z}(v_2v_3) \oplus \mathbb{Z}(v_1v_3)$.
- Com que $H_0 = \emptyset$, tenim que $G(\mathcal{H}_0) = 0$.

L'única combinació lineal de les 1-hiperarestes que té vora nul·la (tenint en compte que la vora de $(v_i v_j)$ és $(v_j) - (v_i)$) és $(v_1v_2) + (v_2v_3) - (v_1v_3)$, ja que la seva vora és

$$\partial((v_1v_2)) + \partial((v_2v_3)) - \partial((v_1v_3)) = (v_2 - v_1) + (v_3 - v_2) - (v_3 - v_1) = 0.$$

Així doncs,

$$\text{Inf}_1 = G(\mathcal{H}_1) \cap \partial_1^{-1}(G(\mathcal{H}_0)) = \langle (v_1v_2) + (v_2v_3) - (v_1v_3) \rangle \cong \mathbb{Z}.$$

Com que no hi ha hiperarestes de dimensió 2, es dedueix que

$$\text{Inf}_2 = 0 \quad \text{i per tant} \quad \text{Im}(\partial_2) = 0.$$

Finalment, el complex de cadenes ínfim i els grups d'homologia són

$$\text{Inf}_*(H) : \quad 0 \longrightarrow \langle c \rangle \xrightarrow{\partial_1} 0,$$

on $c = (v_1v_2) + (v_2v_3) - (v_1v_3)$, i

$$H_0(\mathcal{H}) = 0, \quad H_1(\mathcal{H}) = \text{Ker}(\partial_1) / \text{Im}(\partial_2) = \langle c \rangle / 0 \cong \mathbb{Z}.$$

Aquest exemple mostra clarament que és possible detectar cicles de dimensió 1 en homologia restringida tot i no tenir cap 0-hiperaresta. El grup $H_1(\mathcal{H})$ recull cicles formats per 1-hiperarestes la frontera algebraica dels quals es cancel·la, encara que aquestes no estiguin explícitament presents a l'hipergraf.

Capítol 4

Homologia persistent

L'homologia persistent és una eina matemàtica fonamental en l'anàlisi topològica de dades, ja que permet mesurar la presència i la persistència d'estructures topològiques, com components connexes o forats. Aquesta metodologia és especialment útil per distingir les estructures més importants i robustes dins de les dades. L'homologia persistent s'aplica en diversos camps com l'anàlisi de xarxes socials, la biologia (com l'estudi de proteïnes i gens) i el reconeixement d'imatges. Tots els conceptes i definicions esmentats en aquest capítol han estat extrets dels treballs i llibres [6, 7, 8, 11].

4.1 Homologia persistent clàssica

Donat un complex simplicial K , una *funció de filtració* és una aplicació $f : K \rightarrow \mathbb{R}$ que assigna un valor real a cada símplex de K , de manera que compleix la següent condició:

- Donats dos símplexs τ i σ de K tals que $\tau \subseteq \sigma$, aleshores $f(\tau) \leq f(\sigma)$.

Aquesta propietat assegura que per a qualsevol valor $a \in \mathbb{R}$, el conjunt

$$K(a) := \{\sigma \in K \mid f(\sigma) \leq a\}$$

forma un subcomplex de K . Sigui $a_1 < a_2 < \dots < a_n$ valors de la funció f sobre els símplexs de K , i sigui $a_0 = -\infty$. La funció de filtració dóna lloc a una successió ordenada de subcomplexos

$$\emptyset = K_0 \subseteq K_1 \subseteq K_2 \subseteq \dots \subseteq K_n = K$$

on $K_i := K(a_i)$.

Per a tot $i \leq j$, tenim una aplicació d'inclusió de l'espai K_i a l'espai K_j , i per tant, un homomorfisme induït $f_{i,j}^p : H_p(K_i) \rightarrow H_p(K_j)$ per a cada dimensió p . La filtració dóna lloc a una seqüència de grups d'homologia, un per a cada pas de filtració, connectats mitjançant aquests homomorfismes induïts per inclusió. Així, per a cada dimensió p s'obté una cadena de la forma

$$0 = H_p(K_0) \xrightarrow{f_{0,1}^p} H_p(K_1) \xrightarrow{f_{1,2}^p} \dots \xrightarrow{f_{n-1,n}^p} H_p(K_n) = H_p(K).$$

A mesura que la filtració avança d'un subcomplex K_{i-1} a K_i , poden aparèixer noves classes d'homologia, mentre que d'altres poden desaparèixer. Aquest procés permet agrupar les classes d'homologia segons el moment en què neixen (apareixen) i el moment en què moren (deixen de ser representatives). Les classes que perduren durant més passos de la filtració són més significatives.

Els grups d'homologia persistent de dimensió p , denotats per $H_p^{i,j}$, són les imatges dels homomorfismes induïts per les inclusions $f_{i,j}^p : H_p(K_i) \rightarrow H_p(K_j)$ amb $0 \leq i \leq j \leq n$. És a dir,

$$H_p^{i,j} = \text{Im} f_{i,j}^p = \text{Im}(H_p(K_i) \rightarrow H_p(K_j)).$$

El p -èsim nombre de Betti persistent $\beta_p^{i,j}$, és el rang de $H_p^{i,j}$, $\beta_p^{i,j} = \text{rank} H_p^{i,j}$.

Una filtració $\{K_i\}_{i=0}^n$ indueix per a cada i , un complex de cadenes $C_*^{(i)} = \{C_n^{(i)}, \partial_n^{(i)}\}_{n \geq 0}$, on $C_n^{(i)} := C_n(K_i)$ i

$$\partial_n^{(i)} : C_n^{(i)} \rightarrow C_{n-1}^{(i)}$$

és l'aplicació vora descrita a la secció 1.2 i definida sobre els símplexs de K_i . La família de complexos de cadenes $\{C_*^{(i)}, \partial_*^{(i)}\}_{i \geq 0}$ s'anomena *complex de cadenes filtrat*, i aquest juntament amb una col·lecció d'homomorfismes de complexos de cadenes

$$f_*^{(i)} : C_*^{(i)} \rightarrow C_*^{(i+1)}$$

que commuten amb les vores, és a dir

$$\partial_n^{(i+1)} \circ f_n^{(i)} = f_{n-1}^{(i)} \circ \partial_n^{(i)},$$

s'anomena *complex de persistència*. Aquesta estructura de complex de persistència es pot representar en un diagrama com el següent

$$\begin{array}{ccccccc} \partial_3^{(0)} \downarrow & & \partial_3^{(1)} \downarrow & & \partial_3^{(2)} \downarrow & & \\ C_2^{(0)} & \xrightarrow{f_2^{(0)}} & C_2^{(1)} & \xrightarrow{f_2^{(1)}} & C_2^{(2)} & \xrightarrow{f_2^{(2)}} & \dots \\ \partial_2^{(0)} \downarrow & & \partial_2^{(1)} \downarrow & & \partial_2^{(2)} \downarrow & & \\ C_1^{(0)} & \xrightarrow{f_1^{(0)}} & C_1^{(1)} & \xrightarrow{f_1^{(1)}} & C_1^{(2)} & \xrightarrow{f_1^{(2)}} & \dots \\ \partial_1^{(0)} \downarrow & & \partial_1^{(1)} \downarrow & & \partial_1^{(2)} \downarrow & & \\ C_0^{(0)} & \xrightarrow{f_0^{(0)}} & C_0^{(1)} & \xrightarrow{f_0^{(1)}} & C_0^{(2)} & \xrightarrow{f_0^{(2)}} & \dots \end{array}$$

A continuació es presenten dos exemples habituals de funcions de filtració sobre complexos simplicials construïts a partir d'un conjunt de punts en $\mathbb{R}^n$. Són dos dels més utilitzats per a l'homologia persistent, especialment en aplicacions computacionals: el complex de Čech i el complex de Vietoris–Rips.

Exemple 4.1.1. *Complex de Čech com a funció de filtració.*

Sigui $X = \{x_1, x_2, \dots, x_N\} \subset \mathbb{R}^n$ un conjunt finit de punts. Un símplex σ és qualsevol subconjunt finit no buit de X , $\sigma \subseteq X$. Sigui K un complex simplicial format pels símplexs de X , definim una funció de filtració $f_{\check{C}ech} : K \rightarrow \mathbb{R}$ com

$$f_{\check{C}ech}(\sigma) := \inf \left\{ \epsilon \geq 0 \mid \bigcap_{x \in \sigma} B(x, \frac{1}{2}\epsilon) \neq \emptyset \right\}.$$

Aquest valor representa el mínim radi ϵ tal que les boles centrades en els punts de σ tenen intersecció comuna. Per a cada $\alpha \in \mathbb{R}$, definim el subcomplex filtrat com

$$K(\alpha) := \{\sigma \subseteq K \mid f_{\check{C}ech}(\sigma) \leq \alpha\},$$

que coincideix amb el complex de Čech de paràmetre α . Aquesta funció compleix la condició de filtració, si $\tau \subseteq \sigma$, aleshores $f(\tau) \leq f(\sigma)$. Si és necessari un radi ϵ per tal que les boles de σ intersequin, aleshores les boles d'un subconjunt més petit τ ($\tau \subseteq \sigma$) segurament intersecaran abans o com a molt en el mateix moment.

Exemple 4.1.2. *Complex de Vietoris–Rips com a funció de filtració.*

Amb el mateix conjunt de punts $X = \{x_1, x_2, \dots, x_N\} \subset \mathbb{R}^n$, definim per a cada símplex $\sigma \subseteq X$ la funció de filtració

$$f_{\text{Rips}}(\sigma) := \max\{d(x, x') \mid x, x' \in \sigma\}.$$

Aquest valor correspon al diàmetre del conjunt de punts de σ . De nou, si $\tau \subseteq \sigma$, es té $f(\tau) \leq f(\sigma)$. Els subcomplexos filtrats queden definits com

$$K(\alpha) := \{\sigma \subseteq X \mid f_{\text{Rips}}(\sigma) \leq \alpha\},$$

i corresponen als complexos de Vietoris–Rips de paràmetre α . Si $\tau \subseteq \sigma$ amb τ i σ símplexs formats amb subconjunts de punts de X , aleshores les parelles $(x, x') \in \tau \times \tau$ són un subconjunt de les parelles de σ . Per tant, el màxim de les distàncies dins de τ serà menor o igual que el màxim dins de σ i, en conseqüència, $f_{\text{Rips}}(\tau) \leq f_{\text{Rips}}(\sigma)$.

A la figura 4.1 es mostra com, a partir d'un mateix conjunt de punts i un mateix valor de paràmetre ϵ , es poden construir dos complexos filtrats diferents: el complex de Čech $\check{C}_\epsilon$ i el complex de Vietoris–Rips R_ϵ .

En el cas de Čech, un símplex es forma quan les boles de radi $\epsilon/2$ centrades als punts tenen intersecció comuna. En canvi, al complex de Rips, només cal que tots els punts del símplex estiguin a distància $\leq \epsilon$ entre ells.

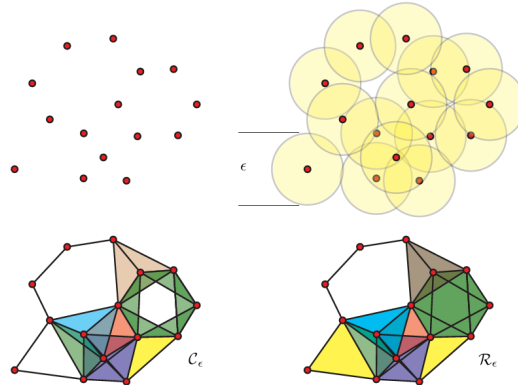


Figura 4.1: Visualització d'un mateix conjunt de punts (a dalt a l'esquerra), amb la construcció del complex de Čech $\check{C}_\epsilon$ (a baix a la dreta) i del complex de Vietoris–Rips R_ϵ (a baix a l'esquerra), per un valor fix de ϵ . A dalt a la dreta es mostren les boles de radi $\epsilon/2$ utilitzades per a la construcció del complex de Čech. Extret de [11].

Diagrames de persistència i codi de barres

Els diagrames de persistència són una representació que permet codificar la informació dels nombres de Betti persistents en forma de punts en el pla $\mathbb{R}^2$. Cada punt (a_i, a_j) amb $i < j$ indica que una classe d'homologia neix al pas K_i de la filtració i mor entrant a K_j . Per a $i < j$, la multiplicitat $\mu_p^{i,j}$ dels punts $(a_i, a_j) \in \text{Dgm}_p(f)$ es defineix com

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

El conjunt de punts (a_i, a_j) amb multiplicitat $\mu_p^{i,j} > 0$ constitueixen el diagrama de persistència. Tots els punts es troben estrictament per sobre de la diagonal $a_i = a_j$ ja que una classe no pot morir abans de néixer. Els punts més a dalt i a l'esquerra del gràfic corresponen a classes d'homologia persistents que han existit durant un llarg període de temps i, per tant, són més importants des del punt de vista topològic.

Donada una filtració de complexos simplicials $\{K_i\}_{i=0}^n$, i una dimensió $p \geq 0$ un *codi de barres de persistència de dimensió p* és una col·lecció d'interval del tipus $[b, d) \subseteq \mathbb{R} \cup \{\infty\}$, on cada interval representa una classe d'homologia p -dimensional que neix al pas K_b i mor entrant a K_d , o bé persisteix indefinidament si $d = \infty$. El codi de barres de persistència s'expressa com un conjunt d'interval

$$\text{Bar}_p(f) = \{[b_i, d_i)\}_i.$$

Aquests intervals es representen en un diagrama on l'eix horitzontal correspon als índexs de la filtració i l'eix vertical ordena les classes d'homologia. Aquesta representació rep el nom de *codi de barres de persistència*. El nombre d'interval $[b, d)$ amb $b \leq i \leq j \leq d$ és el *nombre de Betti persistent* $\beta_p^{i,j}$.

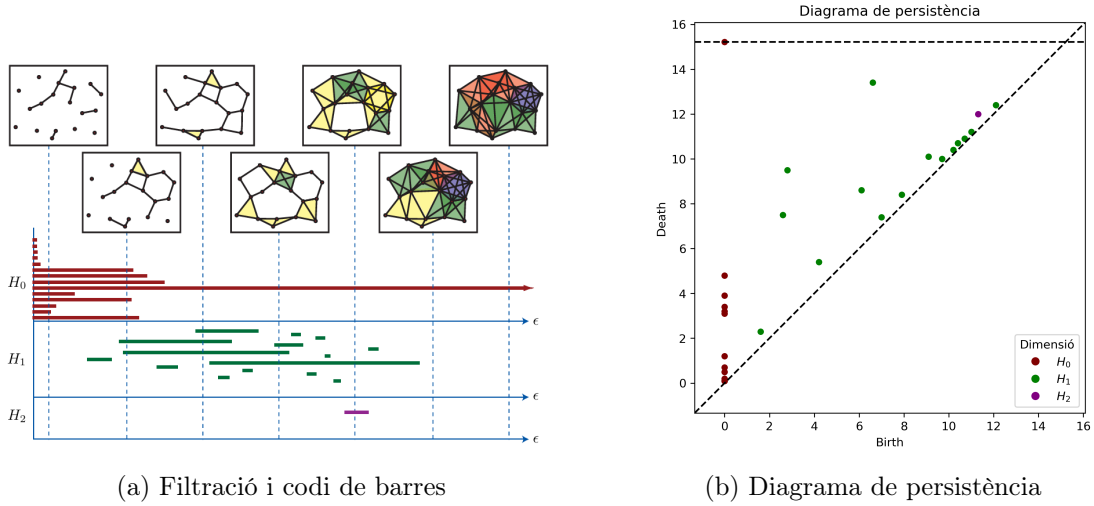


Figura 4.2: A l'esquerra es troba el codi de barres per a les homologies H_0 , H_1 , H_2 . A la dreta s'observa el diagrama de persistència corresponent, amb punts (naixement, mort). Exemple basat en la figura 4 de [11].

A la figura 4.2 es mostra de forma visual com es construeixen els codis de barres i el diagrama de persistència a partir d'una filtració de complexos de Vietoris-Rips. Aquest exemple s'ha generat a partir de la funció de filtració definida anteriorment, $f_{\text{Rips}}(\sigma) :=$

$\max\{d(x, x') \mid x, x' \in \sigma\}$ i els subcomplexos filtrats corresponents, $K(\epsilon) := \{\sigma \subseteq X \mid f_{\text{Rips}}(\sigma) \leq \epsilon\}$. A mesura que el paràmetre ϵ creix, es van afegint més símplexs al complex. La successió de complexos filtrats genera una evolució de la topologia que es pot registrar mitjançant els codis de barres.

A dalt a l'esquerra de la figura 4.2 es visualitzen diferents etapes del complex de Rips per a aquests valors de ϵ , juntament amb els codis de barres per H_0 , H_1 i H_2 . A la dreta, es mostra el diagrama de persistència corresponent, on cada barra s'ha representat com un punt amb coordenades (ϵ_i, ϵ_j) , indicant el valor de naixement i mort. Les classes de H_0 apareixen típicament en $\epsilon = 0$ i moren quan les components es connecten. Les classes de H_1 corresponen a cicles que apareixen més endavant i tenen persistències diverses. La classe de H_2 , més breu, representa una cavitat tridimensional.

4.2 Descriptors numèrics de persistència

Tot i que els diagrames de persistència proporcionen una representació visual rica de les característiques topològiques de les dades, sovint cal transformar aquesta informació en formats numèrics per facilitar-ne l'anàlisi. Per això, s'utilitzen descriptors numèrics de l'homologia persistent, que resumeixen la informació del diagrama en valors quantitius. En aquesta secció es presenten dos d'aquests descriptors: la persistència total i l'entropia de persistència. Les definicions d'aquesta secció s'han extret de [2, 3, 9].

Persistència total

La *persistència total* és un descriptor numèric que mesura la suma de la durada de totes les característiques topològiques d'una dimensió determinada dins d'un diagrama de persistència. S'expressa com

$$P_k = \sum_{i \in I_k} (d_i^k - b_i^k),$$

on I_k és el conjunt de característiques topològiques de dimensió k , i b_i^k i d_i^k corresponen al moment de naixement i de mort de la i -èsima característica. Aquest valor resumeix numèricament la presència i persistència de patrons topològics, com poden ser components connexes, cicles o cavitats, en funció de la seva dimensió.

Entropia de persistència

L'*entropia de persistència* és un descriptor numèric inspirat en l'entropia d'informació de Shannon, i s'utilitza per quantificar la dispersió de les persistències dins d'un diagrama de persistència. És a dir, mesura com de repartides estan les longituds de les barres en un codi de barres.

Per calcular-la, primer es defineix una distribució de probabilitat a partir de la persistència de cada característica dins d'una dimensió d'homologia concreta. La persistència es calcula com la diferència entre la seva mort i el seu naixement. Així doncs, fixada una dimensió k , la probabilitat associada a cada característica es defineix com

$$p_i := \frac{d_i - b_i}{\sum_{j \in I_k} (d_j - b_j)},$$

on b_i i d_i són el naixement i la mort de la característica i , i I_k és el conjunt de totes les barres del codi de barres de dimensió k . A partir d'aquesta probabilitat, l'entropia de persistència per a la dimensió k s'expressa com

$$\text{Entropia de persistència en } H_k := - \sum_{i \in I_k} p_i \log_2 p_i.$$

Aquest valor mesura fins a quin punt les barres del codi de barres tenen longituds similars o desiguals. L'entropia persistent assoleix el seu màxim quan totes les barres tenen la mateixa longitud, ja que la informació topològica està distribuïda uniformement. En aquest cas, l'entropia és pròxima a $\log(n)$, on n és el nombre total de barres. Per contra, si hi ha barres amb longituds molt desiguals, per exemple, unes poques molt llargues i la resta molt curtes, l'entropia disminueix, indicant que la persistència està concentrada en poques característiques dominants.

Així, l'entropia de persistència permet obtenir un valor escalar que resumeix el grau de regularitat o desequilibri del conjunt de barres d'un codi de barres, i pot ser útil per comparar estructures topològiques entre diferents dades o sota diferents condicions.

4.3 Homologia persistent d'hipergrafs

S'ha vist a la secció anterior que un procés de filtració és essencial per a l'homologia persistent. Donat un hipergraf $\mathcal{H}$, una funció de filtració sobre $\mathcal{H}$ és una aplicació $f : \mathcal{H} \rightarrow \mathbb{R}$, que assigna un valor real a cada hiperaresta i compleix que donades dues hiperarestes $\sigma, \tau \in \mathcal{H}$ tals que $\sigma \subseteq \tau$ aleshores $f(\sigma) \leq f(\tau)$. Aquesta última condició és anàloga a la del cas d'un complex simplicial i assegura que per a cada $a \in \mathbb{R}$, el conjunt

$$\mathcal{H}(a) = \{\sigma \in \mathcal{H} \mid f(\sigma) \leq a\}$$

forma un subhipergraf $\mathcal{H}(a) \subseteq \mathcal{H}$.

Escollint una filtració f , uns valors d'imatge $-\infty = a_0 < a_1 < \dots < a_n$ i agafant $\mathcal{H}_i = \mathcal{H}(a_i)$ s'obté una filtració d'hipergrafs

$$H_0 \xrightarrow{f_1} H_1 \xrightarrow{f_2} \dots \xrightarrow{f_n} H_n = H.$$

Donat un grup abelià G , la successió anterior induïx dos complexos de persistència

$$\text{Inf}_*(\mathcal{H}_0) \xrightarrow{(f_1)_*} \text{Inf}_*(\mathcal{H}_1) \xrightarrow{(f_2)_*} \dots \xrightarrow{(f_3)_*} \text{Inf}_*(\mathcal{H}_n), \quad (4.1)$$

$$\text{Sup}_*(\mathcal{H}_0) \xrightarrow{(\hat{f}_1)_*} \text{Sup}_*(\mathcal{H}_1) \xrightarrow{(\hat{f}_2)_*} \dots \xrightarrow{(\hat{f}_n)_*} \text{Sup}_*(\mathcal{H}_n), \quad (4.2)$$

on $\text{Inf}_*(\mathcal{H}_i)$ i $\text{Sup}_*(\mathcal{H}_i)$ són els complexos de cadenes ínfim i suprem de $\mathcal{H}_i$. Les inclusions entre hipergrafs també induïxen aplicacions entre els complexos de cadenes. Cada $f_i : \mathcal{H}_{i-1} \rightarrow \mathcal{H}_i$ induïx una aplicació entre les clausures simplicials, de $K_{\mathcal{H}_{i-1}}$ a $K_{\mathcal{H}_i}$, la qual induïx una aplicació entre el complex de cadenes associat a $K_{\mathcal{H}_{i-1}}$ i el complex de cadenes associat a $K_{\mathcal{H}_i}$. Restringint aquestes aplicacions als complexos ínfim i suprem, obtenim els complexos de persistència anteriors (4.1) i (4.2). Les dues successions induïxen la mateixa successió de grups d'homologia de dimensió p

$$H_p(\mathcal{H}_0) \rightarrow H_p(\mathcal{H}_1) \rightarrow \dots \rightarrow H_p(\mathcal{H}_n).$$

Aquesta successió de grups d'homologia de dimensió p s'anomena l'*homologia persistent de dimensió p de l'hipergraf filtrat*. Cal notar que els complexos de persistència construïts a partir dels complexos de cadenes ínfim i suprem poden ser diferents.

Exemple 4.3.1. Donada una col·lecció ordenada d'hiperarestes $\sigma_1, \sigma_2, \dots, \sigma_n$, es pot construir una successió d'hipergrafs

$$\mathcal{H}_0 \subseteq \mathcal{H}_1 \subseteq \dots \subseteq \mathcal{H}_n$$

on cada $\mathcal{H}_i$ es defineix com el conjunt de les i primeres hiperarestes de la successió $\mathcal{H}_i = \{\sigma_1, \sigma_2, \dots, \sigma_n\}$. Per tant, aquesta construcció consisteix a afegir les hiperarestes una a una seguint l'ordre donat, i dona lloc a una filtració d'hipergrafs, ja que cada hipergraf $\mathcal{H}_i$ és un subhipergraf de $\mathcal{H}_{i+1}$. Cada nova hiperaresta pot modificar l'homologia de l'hipergraf. Per exemple, l'addició d'una 0-hiperaresta introdueix una nova classe d'homologia de dimensió 0. En canvi, si s'afegeix una 1-hiperaresta que connecta dos vèrtexs que formen una 0-hiperaresta cadascun, pot fer desaparèixer una classe d'homologia 0.

Capítol 5

Algorisme per a l'homologia persistent d'hipergrafs

En aquest capítol es presenta l'algorisme per calcular l'homologia persistent d'hipergrafs. Per tal de descriure-ho, cal introduir primer l'algorisme d'homologia persistent per a complexos de cadenes, ja que el mètode utilitzat per a hipergrafs es basa en una adaptació directa d'aquest. El desenvolupament dels algorismes d'aquest capítol segueix l'article [18] escrit per Liu, Feng, Wu i Xia.

5.1 Algorisme clàssic d'homologia persistent per a complexos de cadenes

Donada una família de complexos de cadenes $\{C_*^{(i)}, \partial_*^{(i)}\}_{i \geq 0}$ i una col·lecció d'homomorfismes $f_*^i : \{C_*^{(i)}, \partial_*^{(i)}\} \longrightarrow \{C_*^{(i+1)}, \partial_*^{(i+1)}\}$, s'ha vist a la secció 4.1 que aquesta estructura defineix un complex de persistència que es pot representar en el següent diagrama

$$\begin{array}{ccccccc} \partial_3^{(0)} \downarrow & & \partial_3^{(1)} \downarrow & & \partial_3^{(2)} \downarrow & & \\ C_2^{(1)} & \xrightarrow{f_2^{(0)}} & C_2^{(2)} & \xrightarrow{f_2^{(1)}} & C_2^{(3)} & \xrightarrow{f_2^{(2)}} & \dots \\ \partial_2^{(0)} \downarrow & & \partial_2^{(1)} \downarrow & & \partial_2^{(2)} \downarrow & & \\ C_1^{(1)} & \xrightarrow{f_1^{(0)}} & C_1^{(2)} & \xrightarrow{f_1^{(1)}} & C_1^{(3)} & \xrightarrow{f_1^{(2)}} & \dots \\ \partial_1^{(0)} \downarrow & & \partial_1^{(1)} \downarrow & & \partial_1^{(2)} \downarrow & & \\ C_0^{(1)} & \xrightarrow{f_0^{(0)}} & C_0^{(2)} & \xrightarrow{f_0^{(1)}} & C_0^{(3)} & \xrightarrow{f_0^{(2)}} & \dots \end{array}$$

Per tal de calcular l'homologia d'un complex de cadenes filtrat, és fonamental disposar d'una representació algebraica que reculli totes les relacions de vora entre els simplexs del complex. Aquesta representació es fa mitjançant la matriu vora global.

Suposem que el complex de cadenes filtrat té una base compatible. Això vol dir que podem representar el complex de cadenes filtrat com una seqüència ordenada de bases $a_1, a_2, \dots, a_n$, de manera que, per a cada pas k , els elements $(a_1, a_2, \dots, a_k)$ formen una base del complex de cadenes $C_*^{(k)}$ corresponent al k -èsim nivell de la filtració, i les imatges

dels operadors vora ∂ sobre aquestes bases són conegudes. Per tant, la successió

$$\{C_*^{(1)}, \partial_*^{(1)}\} \xrightarrow{f_*^1} \{C_*^{(2)}, \partial_*^{(2)}\} \xrightarrow{f_*^2} \dots \xrightarrow{f_*^{n-1}} \{C_*^{(n)}, \partial_*^{(n)}\}$$

es pot escriure com $\{C_*^{(1)}, \partial_*^{(1)}\} \subseteq \{C_*^{(2)}, \partial_*^{(2)}\} \subseteq \dots \subseteq \{C_*^{(n)}, \partial_*^{(n)}\}$ amb $a_1, a_2, \dots, a_k$ formant la base de $\{C_*^{(k)}, \partial_*^{(k)}\}$.

Com que es coneix com actua l'operador vora ∂ sobre cada element a_j , aleshores es pot construir una matriu $n \times n$ anomenada *matriu vora*. Aquesta matriu conté a la columna j els coeficients que expressen $\partial(a_j)$ com una combinació lineal d'elements anteriors de la base, és a dir $\partial(a_j) = \sum_i c_j^i a_i$ amb $i < j$, i per tant el valor de la fila i i columna j de la matriu és $\partial(i, j) = c_j^i$.

Per analitzar l'evolució de les característiques topològiques al llarg de la filtració (com ara la creació o destrucció de components connexes), s'hi aplica el que es coneix com a l'*algorisme clàssic de reducció de matrius* a la matriu vora. Aquest algorisme modifica les columnes de la matriu vora mitjançant combinacions lineals per tal que cada columna, al final de la reducció, tingui com a màxim un únic valor no nul, situat a la fila amb l'índex més alt possible dins d'aquella columna (és a dir, la fila més avall). Aquesta posició s'anomena $low(j)$, i ha de ser única entre columnes diferents, és a dir, no poden existir dues columnes amb el mateix valor de low .

Aquest algorisme permet obtenir una versió reduïda de la matriu vora, a partir de la qual es pot extreure el codi de barres de persistència de manera directa. Aquest codi de barres indica en quin moment apareixen i desapareixen les diferents característiques topològiques de l'espai al llarg de la filtració. Les columnes que són completament zero representen característiques topològiques que neixen però no moren mai, i es tradueixen en barres infinites. Les columnes amb un valor de $low(j) = i$ indiquen que una característica neix al l'etapa i i desapareix a l'etapa j , per tant, formen barres infinites $[i, j]$. Definint B com el conjunt d'índexs de columnes zero, P com el conjunt de columnes no nul·les i $T = \{low(p) \mid p \in P\}$, aleshores es tenen les següents barres:

- Les barres infinites són $\{[b, \infty] \mid b \in B - T\}$.
- Les barres finites són $\{[b, d] \mid d \in P, low(d) = b\}$.

Algorithm 1 Algorisme clàssic de reducció de matrius

Entrada: Matriu vora $n \times n$ del complex de cadenes filtrat

Sortida: Matriu vora reduïda

```

1: Defineix  $L = [0, 0, \dots, 0]$ 
2: for  $j = 1$  to  $n$  do
3:   while  $low(j) \neq 0$  and  $L[low(j)] \neq 0$  do
4:     Suma la columna  $low(j)$  a la columna  $j$ 
5:   end while
6:   if  $L[low(j)] == 0$  then
7:     Assigna  $L[low(j)] = j$ 
8:   end if
9: end for
```

L'algorisme clàssic de reducció de matrius es pot veure a 1. Aquest algorisme recorre la matriu de manera iterativa, processant cada columna j seguint l'ordre de filtració. Per

a cada columna, es determina la fila més baixa amb un valor no nul, que s'anomena $low(j)$. Si una altra columna anterior ja té aquest mateix valor de low , es considera un conflicte, dues columnes estarien indicant la mort d'una mateixa característica, cosa que no està permesa. Per resoldre-ho, es modifica la columna actual sumant-hi la columna anterior conflictiva. Aquesta operació es repeteix fins que $low(j)$ sigui únic. El vector L que s'utilitza a l'algorisme serveix per registrar, per a cada fila i , si aquesta ja ha estat assignada com a $low(j)$ d'alguna columna. D'aquesta manera s'evita que dues columnes comparteixin la mateixa posició baixa no nul·la, i s'assegura que cada mort topològica s'aparella de manera unívoca amb el seu naixement corresponent.

5.2 Algorisme d'homologia persistent d'hipergrafs

Per aplicar l'algorisme clàssic de reducció de matrius (algorisme 1) a hipergrafs, és necessari transformar l'hipergraf filtrat en un complex de cadenes filtrat. Aquesta conversió es pot fer de dues formes, construint el complex suprem associat a l'hipergraf, o bé construint el complex ínfim.

En aquesta secció es descriu amb detall la construcció del complex suprem. Tot i que existeixen dues construccions possibles (suprem i ínfim) per a l'homologia d'hipergrafs, ens centrarem en el complex suprem, ja que en el capítol posterior només s'utilitzarà aquest. Aquesta elecció es basa en els resultats presentats a l'article [18], on a la secció 4.2 es demostra que el complex suprem és computacionalment més eficient. A la secció 3.2.2 d'aquest mateix article es detalla la construcció del complex ínfim.

En concret, el complex suprem resulta més eficient perquè la seva construcció només requereix resoldre dos sistemes d'equacions lineals per cada hiperaresta afegida a la filtració (com es veurà més endavant), mentre que el complex ínfim implica passos addicionals per extreure la base i més comprovacions. Això fa que, tot i tenir la mateixa complexitat teòrica, $\mathcal{O}(n^4)$ en temps i $\mathcal{O}(n^2)$ en espai, amb n el nombre d'hiperarestes, l'algorisme del complex suprem sigui més ràpid i senzill d'implementar en la pràctica, tal com remarquen els autors a la secció 4.2 de [18].

Sense pèrdua de generalitat, es considera, per simplicitat, una filtració de l'hipergraf que consisteix en una seqüència creixent d'hipergrafs

$$\emptyset = \mathcal{H}_0 \subset \mathcal{H}_1 \subset \cdots \subset \mathcal{H}_n = \mathcal{H},$$

on cada $\mathcal{H}_i$ conté les primeres i hiperarestes $\sigma_1, \sigma_2, \dots, \sigma_i$ de l'hipergraf $\mathcal{H}$. Aquesta forma de filtració facilita la descripció dels algorismes, però també és aplicable el mètode quan diverses hiperarestes s'incorporen en un mateix pas, sempre que es pugui establir un ordre compatible. A continuació, es donen els algorismes per transformar un hipergraf filtrat en el seu complex de cadenes suprem filtrat i en el seu complex de cadenes ínfim filtrat sobre el cos de coeficients.

5.2.1 Construcció del complex de cadenes suprem

S'anomena D_k^i al grup de cadenes de k -hiperarestes de l'hipergraf $\mathcal{H}_i$, és a dir, al grup abelià lliure generat per les k -hiperarestes de $\mathcal{H}_i$. A la secció 3.2 s'ha definit el complex suprem d'un hipergraf com $\text{Sup}_*(\mathcal{H}) = \text{Sup}_*(G(H_*), G((K_{\mathcal{H}})_*))$. D'altra banda, tal com s'expressa a la proposició 2.1.4, el grup de cadenes suprem de dimensió k es pot calcular

com $\text{Sup}_k^i = D_k^i + \partial(D_{k+1}^i)$ on ∂ és l'aplicació vora. En conseqüència, qualsevol element de la base del grup Sup_k^i pot aparèixer en dues formes: com una hiperaresta σ_i , o bé com la vora $\partial(\sigma_i)$ d'una hiperaresta. El punt clau del procediment consisteix a identificar els nous elements de la base que es generen quan s'afegeix una hiperaresta durant el procés de filtració.

Suposem que s'afegeix una hiperaresta de dimensió k , σ_i . Aquesta pot generar nous elements de base tant en Sup_k^i com en Sup_{k-1}^i . En el primer cas, el candidat és la pròpia hiperaresta σ_i , en el segon cas, el candidat és la seva vora $\partial(\sigma_i)$. Per decidir si realment cal incorporar aquests elements a la base, cal comprovar si ja són generables a partir dels elements del pas anterior Sup_*^{i-1} . Aquesta verificació es duu a terme resolent sistemes d'equacions lineals.

Per poder decidir si una hiperaresta σ_i o la seva vora $\partial(\sigma_i)$ són elements nous de la base, cal representar-los com a vectors en un espai on es puguin comparar amb els elements ja existents. Per tal de representar vectorialment els possibles elements de la base i comprovar que aporten nova informació al complex suprem, cal treballar dins d'un espai on totes les vores estiguin ben definides. Per això es considera la clausura simplicial $K_{\mathcal{H}_n}$ d'un hipergraf $\mathcal{H}_n$. Aquesta estructura permet, en particular, definir la vora de qualsevol hiperaresta de manera anàloga a com es fa en el cas dels símplexs.

A partir d'aquest complex $K_{\mathcal{H}_n}$ es construeix una *pre-base*, que servirà com a base de referència per expressar tots els vectors candidats a formar part del complex suprem. Aquesta *pre-base* es construeix de la manera següent.

En primer lloc, s'inclouen totes les hiperarestes σ_i de $\mathcal{H}_n$ per a $1 \leq i \leq n$. A continuació, per a cada σ_i es calcula la seva vora $\partial(\sigma_i)$, que s'expressa com a combinació lineal de subconjunts del complex $K_{\mathcal{H}_n}$, és a dir $\partial(\sigma_i) = \sum_j c_j^i \alpha_j$ amb $\alpha_j \in K_{\mathcal{H}_n}$. Si $c_j^i \neq 0$ per a algun $\alpha_j \notin \mathcal{H}_n$ aleshores també s'afegeix α_j a la *pre-base*.

Aquest procediment es repeteix per a totes les hiperarestes de $\mathcal{H}_n$ i així s'obté una *pre-base* completa que conté totes les hiperarestes originals i totes les parts de les seves vores que manquen dins l'hipergraf. Així, com que per definició $\text{Sup}_k^i = D_k^i + \partial(D_{k+1}^i)$, tots els elements de la base dels grups de cadenes suprema poden expressar-se com a combinacions lineals d'elements de la *pre-base*. Per tant, cada element de la base del complex suprem pot ser representat com un vector respecte a aquesta base auxiliar. Aquest fet permet implementar una representació matricial del conjunt d'elements de la base existents, i anar-hi afegint nous vectors cada vegada que es detecti un nou element independent. La matriu serveix com a eina central per construir el complex suprem filtrat.

Algorithm 2 Obtenció del complex de cadenes suprem filtrat

Entrada: Hipergraf filtrat: $\sigma_1, \sigma_2, \dots, \sigma_n$

Sortida: Base del complex de cadenes suprem filtrat

```

1: for  $i = 1$  to  $n$  do
2:   Suposa que  $\sigma_i$  és una  $k$ -hiperaresta
3:   if  $\text{Sup}_k^{i-1} \cdot X = \sigma_i$  no té solució then
4:     Afegeix  $\sigma_i$  a la base de  $\text{Sup}_k^i$ 
5:   end if
6:   if  $\text{Sup}_{k-1}^{i-1} \cdot X = \partial(\sigma_i)$  no té solució then
7:     Afegeix  $\partial(\sigma_i)$  a la base de  $\text{Sup}_{k-1}^i$ 
8:   end if
9: end for

```

L'algorisme 2 s'utilitza per construir el complex de cadenes suprem filtrat. A efectes pràctics es denota per Sup_k^i la matriu que representa la base del grup de cadenes suprema de dimensió k en el pas i de la filtració. L'algorisme recorre les hiperarestes de l'hipergraf una a una, seguint l'ordre imposat per la filtració. Quan es processa una nova hiperaresta σ_i de dimensió k , cal determinar si aquesta i la seva frontera $\partial(\sigma_i)$ són realment nous elements de la base, o si ja poden generar-se a partir dels elements que es tenien en el pas anterior Sup_*^{i-1} com s'ha mencionat abans. Aquesta decisió es pren resolent dos sistemes d'equacions lineals:

- Si el sistema $\text{Sup}_k^{i-1}X = \sigma_i$ no té solució, vol dir que σ_i no es pot generar a partir dels vectors existents, i per tant s'afegeix com a nou element a la base de Sup_k^i .
- Si el sistema $\text{Sup}_k^{i-1}X = \partial(\sigma_i)$ no té solució, aleshores la vora $\partial(\sigma_i)$ també s'incorpora com a nou element de la base de Sup_{k-1}^i .

Si σ_i és pot generar, és a dir, $\text{Sup}_k^{i-1}X = \sigma_i$ té solució, aleshores $\sigma_i \in \text{Sup}_k^{i-1}$, i per tant, $\partial(\sigma_i) \in \text{Sup}_{k-1}^{i-1}$. Això vol dir que $\text{Sup}_{k-1}^{i-1}X = \partial(\sigma_i)$ té solució i $\partial(\sigma_i)$ és pot generar. En aquest cas, cap dels dos elements s'afegeix a la base. Per tant, només es considera el sistema $\text{Sup}_{k-1}^{i-1}X = \partial(\sigma_i)$ si el sistema $\text{Sup}_k^{i-1}X = \sigma_i$ no és pot resoldre.

Després d'executar aquest algorisme, el complex de cadenes suprem filtrat queda representat mitjançant una base compatible amb la filtració. Aquesta base es pot utilitzar directament com a entrada per a l'algorisme de càlcul d'homologia persistent que s'ha descrit a la secció 5.1 (algorisme 1).

Capítol 6

Exemple pràctic de l'homologia persistent d'hipergrafs

Per tal d'il·lustrar l'ús pràctic de l'algorisme explicat al capítol 5 i la seva aplicació en el càlcul d'homologia persistent, es presenta un exemple basat en dades reals extretes del repositori *arXiv*. En concret, s'ha construït un hipergraf a partir de col·laboracions d'autors en articles de les àrees de topologia algebraica (**math.AT**) i topologia general (**math.GN**), durant el període comprès entre els anys 2004 i 2024.

Tot el codi utilitzat per obtenir les dades i per generar els resultats d'aquest capítol (la construcció de l'hipergraf i el càlcul d'homologia persistent) està disponible al següent enllaç: <https://github.com/cintaa05/TFGCinta/blob/main/TFGCintaIsern.ipynb>.

En aquest context, cada hiperaresta de l'hipergraf representa un conjunt d'autors que han col·laborat en un mateix article, i els vèrtex són els diferents autors. Aquest hipergraf servirà com a entrada per a la comparació entre l'homologia restringida de l'hipergraf i l'homologia de la seva clausura simplicial, mitjançant l'homologia persistent.

Tal com s'ha exposat en la secció 4.3, una filtració sobre un hipergraf $\mathcal{H}$ es defineix com una funció $f : \mathcal{H} \rightarrow \mathbb{R}$ que assigna un valor real a cada hiperaresta i compleix la condició d'ordre: si $\sigma \subseteq \tau$, aleshores $f(\sigma) \leq f(\tau)$.

En l'exemple considerat, s'ha emprat una filtració temporal, on la funció f assigna a cada hiperaresta l'any més antic en què el corresponent grup d'autors ha col·laborat. Això permet ordenar cronològicament les hiperarestes i construir una filtració

$$\mathcal{H}(a_0), \mathcal{H}(a_1), \dots, \mathcal{H}(a_n),$$

on cada $\mathcal{H}(a_i)$ conté totes les hiperarestes (articles) amb any de publicació menor o igual a a_i .

Els passos concrets de la construcció són els següents:

- Cada hiperaresta es forma a partir d'un conjunt d'autors i un any de publicació.
- Es determinen els anys mínims de col·laboració i s'ordenen les hiperarestes segons aquest valor.
- Per a cada any a_i , es defineix el subhipergraf $\mathcal{H}(a_i)$ amb totes les hiperarestes corresponents a anys $\leq a_i$.

- La successió de subhipergrafs obtinguda defineix una filtració creixent sobre la qual es pot aplicar l'homologia persistent.

Aquest procés permet construir dues famílies de complexos de cadenes

$$\text{Inf}_*(\mathcal{H}_0) \rightarrow \text{Inf}_*(\mathcal{H}_1) \rightarrow \cdots \rightarrow \text{Inf}_*(\mathcal{H}_n) \quad \text{i} \quad \text{Sup}_*(\mathcal{H}_0) \rightarrow \text{Sup}_*(\mathcal{H}_1) \rightarrow \cdots \rightarrow \text{Sup}_*(\mathcal{H}_n),$$

les quals corresponen, respectivament, als complexos ínfim i suprem associats als subhipergrafs de la filtració.

Tot i que l'algorisme presentat al capítol 5 està pensat per calcular l'homologia restringida d'un hipergraf, s'ha aprofitat aquest mateix procediment per calcular també l'homologia persistent de la clausura simplicial a cada pas de la filtració. En aquest cas, per a cada subhipergraf $\mathcal{H}(a_i)$, es construeix la seva clausura simplicial, que inclou tots els subconjunts no buits de cada hiperaresta, assegurant que el complex resultant sigui tancat sota inclusió de cares. En el context d'aquest exemple, això significa que si un conjunt d'autors ha col·laborat junts en un article, la clausura inclou també totes les possibles col·laboracions parcials dins d'aquest grup. Això permet utilitzar exactament el mateix algorisme, ja que com s'ha exposat a la secció 3.2 l'homologia restringida coincideix amb l'homologia simplicial quan l'hipergraf és un complex simplicial.

6.1 Resultats obtinguts

Com ja s'ha exposat, les dades utilitzades s'han extret del repositori *arXiv*. Per tal de garantir una mostra equilibrada i tractable computacionalment, s'ha seleccionat un màxim de 200 articles per any per a cadascuna de les dues àrees considerades, topologia algebraica (*math.AT*) i topologia general (*math.GN*), durant el període 2004-2024.

A partir d'aquestes dades i el codi de l'algorisme del capítol anterior s'han generat els diagrames de persistència i els codis de barres dels grups d'homologia H_0 , H_1 i H_2 , tant per al model d'homologia restringida com per al de clausura i per a cadascuna de les àrees. També s'han generat les taules de valors corresponents a la persistència total i a l'entropia de persistència, que proporcionen una mesura quantitativa de la complexitat topològica detectada al llarg de la filtració.

Tot i que en aquest capítol només s'analitzen detalladament les àrees de topologia algebraica i topologia general, a l'Apèndix es recullen els diagrames i mesures corresponents a altres àrees de les matemàtiques del repositori *arXiv*, que també han estat processades mitjançant el mateix codi.

6.1.1 Topologia algebraica

Homologia persistent restringida

A la figura 6.1 es mostren, respectivament, el diagrama de persistència (a l'esquerra) i el codi de barres (a la dreta) obtinguts per a l'àrea de topologia algebraica. Els resultats s'han generat a partir del complex suprem associat al hipergraf (com s'explica a la secció 5.2.1). Cal destacar, com ja s'ha mencionat, que els grups d'homologia induïts pel complex de persistència ínfim donen els mateixos resultats. A més, a les taules 6.1 i 6.2 es presenten diversos descriptors numèrics corresponents als codis de barres obtinguts. La taula 6.1

recull la persistència total, l'entropia de persistència i el nombre total de barres en la dimensió 0, mentre que la taula 6.2 mostra els mateixos valors per a la dimensió 1.

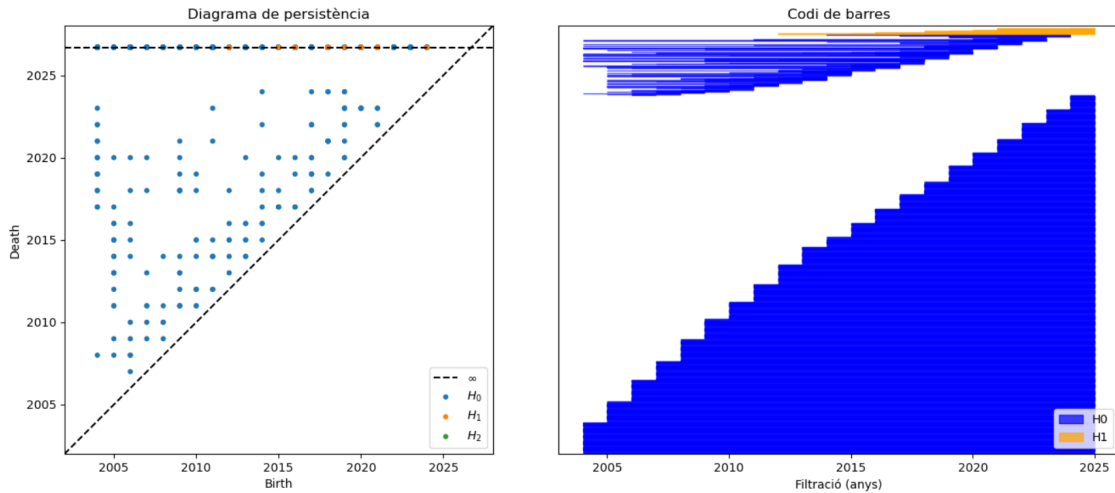


Figura 6.1: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir del complex suprem dels hipergrafs de la filtració corresponent a l'àrea de topologia algebraica.

Pel que fa a la interpretació dels resultats, es pot observar que la major part de l'estructura detectada es concentra en dimensió 0. Les barres blaves del codi de barres corresponen a classes d'homologia H_0 , que, en el cas de l'homologia restringida, identifiquen vèrtexs (autors) que entren a la xarxa mitjançant una hiperaresta de dimensió 0, és a dir, publicant un article individualment. Aquestes classes persisteixen mentre l'autor continua sense connectar-se a cap altre mitjançant una 1-hiperaresta (un article signat exactament per dos autors). Per tant, perquè dues classes de H_0 es fusionin (i una desaparegui), cal que dos autors que ja hagin publicat individualment signin conjuntament un article de coautoría exclusiva. Cal destacar que els autors que només col·laboren en articles amb tres o més signants no generen cap classe pròpia en H_0 , ja que no hi ha cap hiperaresta de dimensió 0 que representi la seva entrada individual a la xarxa. En aquest model, la seva presència queda "oculta" a la dimensió 0, i només pot reflectir-se en dimensions superiors si participa en estructures més complexes.

Aquest comportament explica la presència de barres de curta durada (autors que entren a la xarxa amb un article sol i aviat col·laboren bilateralment), així com d'altres de més llargues o fins i tot persistents fins al final, corresponents a autors que no han col·laborat mai de manera exclusiva amb cap altre autor que també ha publicat un article sol.

Pel que fa a la dimensió H_1 , el codi de barres reflecteix l'aparició i persistència de cicles de col·laboracions entre autors. En aquest context, aquests cicles s'interpreten com conjunts de col·laboracions per parelles entre autors que tanquen un cicle (per exemple, col·laboracions entre (a, b) , (b, c) , i (c, a)) sense que hi hagi una hiperaresta de dimensió 2 que els englobi tots tres simultàniament. Com que s'està utilitzant l'homologia restringida, cal destacar que la presència d'una hiperaresta de dimensió superior com (a, b, c, d) no omple automàticament el cicle (a, b) , (b, c) , (c, a) . Això significa que aquests cicles poden seguir existint com a classes no trivials a H_1 encara que apareguin hiperarestes més grans que incloguin els mateixos autors.

Les barres que apareixen en el codi de barres de H_1 representen, per tant, configuracions de col·laboració cíclica entre parelles d'autors, les quals es mantenen en el temps fins que es trenquen explícitament per la incorporació d'una hiperaresta que les faci col·lapsar.

En la segona dimensió, les classes d'homologia H_2 representarien cavitats tridimensionals dins la xarxa de col·laboracions, generades per combinacions de col·laboracions entre grups de tres autors, és a dir, per 2-hiperarestes que delimiten un volum buit. En el cas de l'homologia restringida, aquestes cavitats només poden desaparèixer si s'hi incorporen explícitament 3-hiperarestes que cobreixin completament les seves cares. En canvi, la presència de 4-hiperarestes que incloguin els mateixos autors no elimina la cavitat, ja que les hiperarestes de dimensió superior no generen automàticament els subconjunts inferiors necessaris per tancar-la. En els resultats de la figura 6.1, no s'observa cap classe en H_2 , fet que indica que no es formen cavitats d'aquest tipus al llarg de la filtració.

Àrea	Hom. (H_0)	PT(fin.)	PT(inf.)	EP(fin.)	EP(inf.)	#Bar(fin.)	#Bar(inf.)
Topo. Alg.	restringida	872	10143	6.8390	9.5443	154	904
	clausura	1459	11780	7.8318	9.7365	300	1028
Topo. Gen.	restringida	284	6540	5.4278	9.0534	56	647
	clausura	685	9677	6.6318	9.6442	132	982

Taula 6.1: Per a al grup d'homologia de dimensió 0, persistència total (PT) i entropia de persistència (EP), separades per barres finites (fin.) i infinites (inf.), per a dues àrees i dues variants de càlcul de l'homologia. #Bar indica el nombre total de barres de cada tipus.

Àrea	Hom. (H_1)	PT(fin.)	PT(inf.)	EP(fin.)	EP(inf.)	#Bar(fin.)	#Bar(inf.)
Topo. Alg.	restringida	0	84	0	3.6696	0	15
	clausura	3	1025	0	7.1477	1	185
Topo. Gen.	restringida	0	150	0	4.2247	0	23
	clausura	2	569	0	6.1137	1	91

Taula 6.2: Per a al grup d'homologia de dimensió 1, persistència total (PT) i entropia de persistència (EP), separades per barres finites (fin.) i infinites (inf.), per a dues àrees i dues variants de càlcul de l'homologia. #Bar indica el nombre total de barres de cada tipus.

Pel que fa al diagrama de persistència, s'hi poden identificar diversos patrons. En primer lloc, una part important dels punts blaus (H_0) es troben relativament a prop de la diagonal, indicant la presència de característiques de curta durada. Això reflecteix autors que entren a la xarxa mitjançant una publicació individual (hiperarestes de dimensió 0) i que aviat estableixen col·laboracions bilaterals (1-hiperarestes). D'altra banda, la presència de diversos punts a l'esquerra del diagrama, especialment al voltant de l'any 2004, indica que un nombre significatiu d'autors van entrar a la xarxa publicant articles individuals als anys 2004 i 2005.

Per a la dimensió 1, totes les classes detectades (H_1) es troben a la part superior del diagrama de persistència, alineades amb la recta de persistència infinita. Això indica que, un cop apareixen, aquestes estructures no són mai eliminades al llarg de la filtració. Aquesta observació es veu reforçada pels valors de la taula 6.2, on la persistència total per a les barres finites a H_1 és nul·la fet que confirma que totes les persistències són infinites.

Finalment, no es detecta cap punt verd al diagrama i per tant cap classe de H_2 .

Dels valors de les taules 6.1 i 6.2 s'observa que el nombre de barres finites és molt més baix que el nombre de barres infinites, tant per dimensió 0 com per dimensió 1. Aquest fet en dimensió 0 indica que hi ha un nombre força elevat d'autors que han publicat sols i mai han publicat amb un autor que també ha publicat un article individual. En dimensió 1, indica que es formen diversos cicles de col·laboració entre autors, però molt pocs arriben a desaparèixer. Això es deu al fet que, perquè un cicle es pugui destruir caldria que els autors implicats haguessin col·laborat en diferents grups de tres que cobreixin tot el cicle, cosa poc habitual si el grup és gran.

Els valors d'entropia en dimensió 0 són 6.8390 i 9.5443 per les barres finites i infinites respectivament. Com tenim un total de 154 i 904 barres i $\log_2(154) \approx 7.27$, $\log_2(904) \approx 9.82$ i els valors 6.8390 i 9.5443 s'aproximen força a aquests, aleshores la distribució de les persistències és força equilibrada, és a dir, la majoria de barres tenen longituds similars, i no hi ha unes poques dominants. Per a la dimensió 1, l'entropia de les barres finites és 0 ja que no n'hi ha, mentre que per a les barres infinites és 3.6696. Com que hi ha 15 barres infinites i el valor màxim teòric és $\log_2(15) \approx 3.9069$, aquest resultat indica que la persistència està molt repartida entre les barres, sense que cap d'elles domini clarament.

Homologia persistent de la clausura simplicial

A continuació, s'analitzen els resultats obtinguts aplicant homologia persistent sobre la clausura simplicial dels hipergrafs corresponents a l'àrea de topologia algebraica. A diferència de l'homologia restringida, aquest enfocament considera automàticament tots els subconjunts de vèrtexs de cada hiperaresta. Això implica que qualsevol col·laboració entre més de dos autors incorpora també totes les col·laboracions per parelles i individuals entre els seus subconjunts, fet que pot tenir un impacte significatiu en l'estructura topològica resultant.

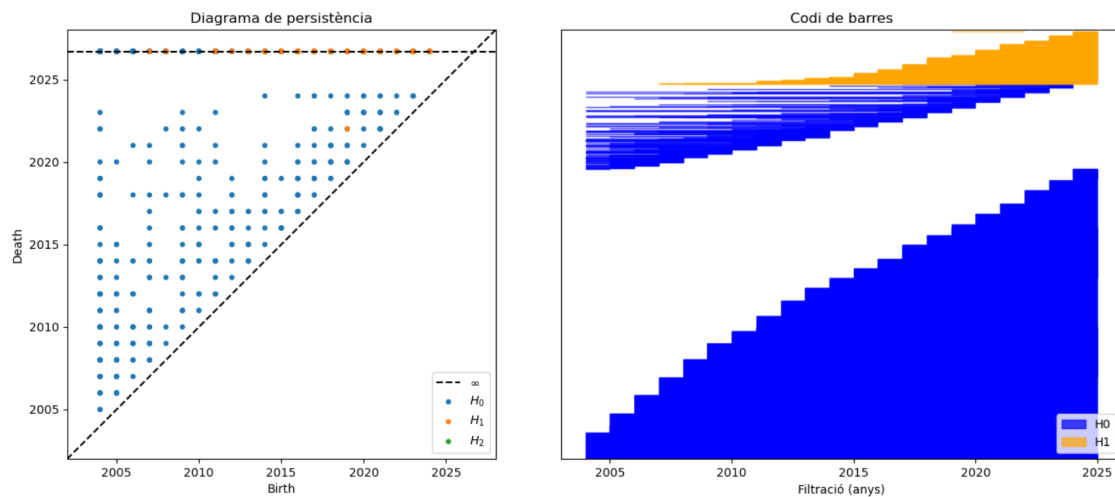


Figura 6.2: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de les clausures simplicials dels hipergrafs de la filtració corresponent a l'àrea de topologia algebraica.

Els diagrames de persistència i codis de barres corresponents mostren una variació important respecte al cas de l'homologia restringida. Com es pot veure a la figura 6.2 i

a la taula 6.1, el nombre de classes d'homologia en dimensió H_0 és considerablement més gran, i apareixen noves classes en H_1 , fet que reflecteix la presència de cicles completats automàticament per la clausura. Aquest fenomen es veu també reflectit en els valors de persistència total i del nombre total de barres (taula 6.1 i 6.2), especialment per a H_0 en barres finites i H_1 en barres infinites, on s'observa un augment notable.

Tot i que la clausura simplicial pot introduir informació que no està explícitament present en l'hipergraf original, aquesta pot resultar útil en determinats casos. En particular, el lema 3.1.3 garanteix que el nombre de components connexes de l'hipergraf coincideix amb el rang del grup H_0 de la seva clausura simplicial. Això permet analitzar l'evolució de la connexitat de la xarxa de col·laboracions a través de l'homologia persistent de la clausura simplicial, amb la garantia que aquesta mesura reflecteix la connexió real entre vèrtexs de l'hipergraf.

Tanmateix, en dimensions superiors ($H_1, H_2, \dots$), la interpretació es torna molt més delicada. En passar a la clausura simplicial, qualsevol subconjunt no buit d'una hiperaresta es converteix automàticament en un símplex del complex, tot i que aquests subconjunts no hagin estat observats com a hiperarestes reals en el hipergraf. Això pot donar lloc a l'aparició de cicles ficticis, que no reflecteixen col·laboracions reals.

Aquest comportament pot inflar la quantitat de classes detectades a H_1 i H_2 , mostrant estructures topològiques que no es corresponen necessàriament amb patrons de col·laboració reals. Per això, mentre que en H_0 la clausura és útil per obtenir informació fidel sobre la connectivitat, en dimensions superiors cal interpretar els resultats amb molta cautela i contrastar-los, si és possible, amb l'homologia restringida.

En el diagrama de persistència de la figura 6.2 s'observa un punt groc finit, que representa una classe finita d'homologia en dimensió H_1 . Aquest fet també es recull a la taula 6.2, on s'indica que el nombre total de barres finites en H_1 per a la clausura simplicial és 1. Això suggereix l'existència d'un únic cicle de col·laboracions per parelles que ha estat eliminat. Aquesta desaparició s'explica pel fet que diversos dels autors implicats han col·laborat en grups de tres. Com que treballem amb la clausura simplicial, aquestes col·laboracions no necessàriament corresponen a articles signats exactament per tres autors, sinó que poden provenir d'articles amb més signants, dels quals la clausura genera automàticament tots els subconjunts. El fet que no apareguin més cicles finits reforça la idea que, en general, els autors que formaven els cicles detectats en l'homologia restringida no han col·laborat en suficients grups de tres dins de col·laboracions més grans. Si hagués estat així, la clausura hauria generat prou 2-símplexs per tancar i eliminar aquests cicles.

6.1.2 Topologia general

Homologia persistent restringida

En aquest apartat es comenten breument els resultats obtinguts per a l'àrea de topologia general, utilitzant l'homologia persistent restringida. En la figura 32 s'observen el diagrama de persistència i el codi de barres obtinguts.

Tot i que l'estructura global dels resultats obtinguts per a l'àrea de topologia general (figura 32) és força similar a la de Topologia Algebraica, s'hi poden identificar algunes diferències. En dimensió 0, es pot observar una menor densitat aparent de punts al diagrama de persistència. Tot i que això podria suggerir una reducció del nombre de classes, cal recordar que diversos punts poden representar múltiples barres amb la mateixa per-

sistència. Aquest fenomen apunta a una menor variabilitat en les durades de les classes d'homologia H_0 , cosa que es reflecteix en l'entropia persistent per a barres finites. En aquest cas, el valor de l'entropia és 5.4278, calculat sobre un total de 56 barres finites. El màxim teòric d'entropia per a 56 barres (amb una distribució completament uniforme) seria $\log_2(56) \approx 5.8074$. Per tant, com que s'ha obtingut un valor d'entropia força proper al màxim, la distribució de les persistències de les barres és bastant equilibrada. Tanmateix, si comparem aquest resultat amb el de topologia algebraica, on l'entropia observada és 6.8390 amb un màxim teòric de 7.27, veiem que la topologia algebraica s'apropa lleugerament més al seu màxim teòric que la topologia general. Això indica que, tot i que en ambdós casos la diversitat en les llargades de les barres és alta, la distribució de les persistències és una mica més uniforme en topologia algebraica. En conseqüència, en el cas de topologia general hi ha lleugerament més variabilitat entre les llargades de les barres, mentre que en topologia algebraica la persistència està més repartida de manera equilibrada entre les diferents classes.

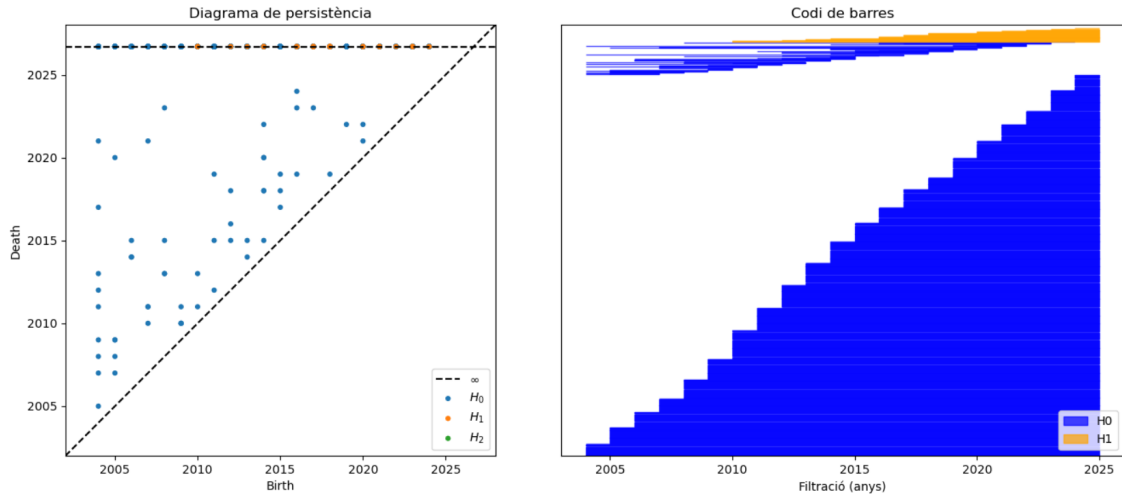


Figura 6.3: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia restringida dels hipergrafs de la filtració corresponent a l'àrea de topologia general.

El nombre total de barres finites i infinites de H_0 és significativament més baix (56 i 647) que en el cas de topologia algebraica (154 i 904). Aquesta diferència no implica necessàriament una xarxa col·laborativa més cohesionada, sinó que reflecteix que, en aquesta àrea, hi ha menys publicacions totalment individuals o entre només dues persones, i una proporció més gran d'articles que ja des del principi són escrits per grups més amplis. Els valors d'entropia obtinguts són 5.4278 per a 56 barres i 9.0534 per a 647 barres. Pel que fa a l'entropia en dimensió zero, en ambdós casos, l'entropia persistent s'aproxima molt al màxim teòric ($\log_2(56) \approx 5.8074$ i $\log_2(647) \approx 9.3391$). Això indica que la distribució de les llargades de les barres és molt equilibrada i que la diversitat de persistències dins la xarxa és molt alta, especialment en el cas amb més barres, on la distribució s'acosta gairebé al màxim d'uniformitat possible.

Pel que fa a la dimensió 1, observant la taula 6.2 veiem que tampoc hi ha cap barra finita però aquest cop n'hi ha més d'infinites que per a la topologia algebraica. Això indica que en topologia general es formen més cicles entre parelles d'autors que en topologia algebraica, i per tant podria reflectir una estructura col·laborativa més fragmentada. En

aquest context, una major fragmentació vol dir que la xarxa està formada per molts petits grups o cicles que es mantenen separats: els autors col·laboren dins d'un cert grup estable, però aquests grups no s'uneixen fàcilment amb altres grups. Això pot indicar que, tot i que hi ha molta activitat col·laborativa, aquesta es produeix en cercles relativament tancats o que els membres dels grups canvien sovint, creant nous cicles però sense consolidar grans components connectats. Finalment, els valors d'entropia per a la dimensió 1 són 5.4278 (per barres finites) i 9.0534 (per 647 barres infinites). Altre cop, l'entropia persistent s'aproxima molt al màxim teòric per al nombre de barres corresponent ($\log_2(56) \approx 5.8074$ i $\log_2(647) \approx 9.3391$). Tant en topologia algebraica com en topologia general, les entropies observades a les dimensions 0 i 1 s'apropen molt als seus màxims teòrics respectius, cosa que indica una distribució de persistències molt equilibrada en ambdós casos.

Homologia persistent de la clausura simplicial

A la figura 6.4, s'observen els resultats obtinguts aplicant homologia persistent sobre la clausura simplicial dels hipergrafs corresponents a l'àrea de topologia general. En comparació amb els resultats obtinguts usant l'homologia restringida, les figures mostren una estructura força similar, especialment en la dimensió H_0 , tot i que s'observen algunes diferències rellevants. En particular, apareixen barres finites addicionals en H_1 , que poden correspondre a cicles "fictícies" introduïts pel procés de clausura. Tal com s'ha discutit anteriorment, aquests cicles poden no reflectir col·laboracions reals observades al hipergraf original, sinó estructures inferides automàticament a partir de la inclusió sistemàtica de subconjunts dins les hiperarestes.

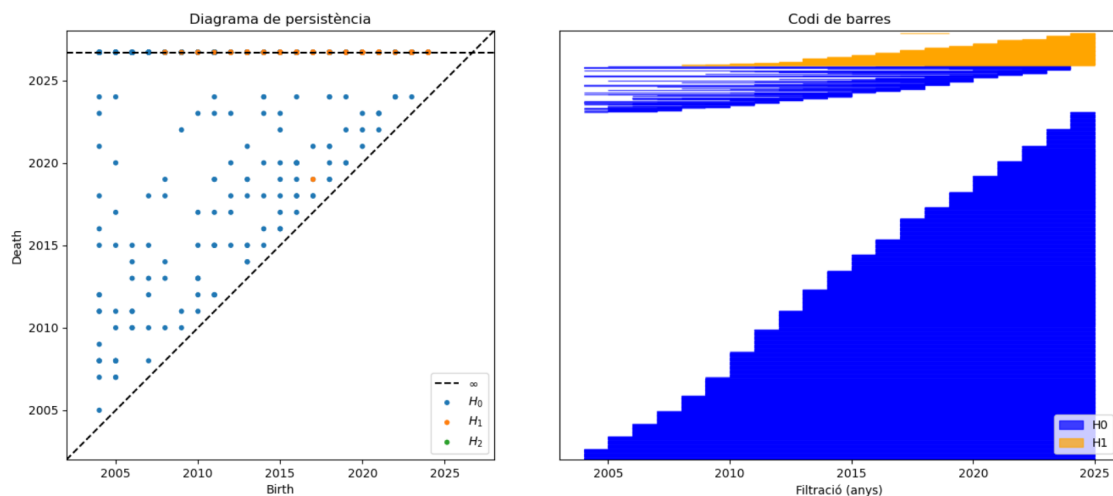


Figura 6.4: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia restringida dels hipergrafs de la filtració corresponent a l'àrea de topologia general

A partir dels resultats d'homologia H_0 de la clausura, que mesura la connectivitat del graf, podem comparar la fragmentació de la xarxa en topologia general i algebraica utilitzant les dades de la taula 6.2. En topologia general, hi ha 132 barres finites (persistència total 685) i 982 barres infinites (persistència total 9677), mentre que en topologia algebraica aquests valors són encara més elevats, 300 barres finites (persistència total 1459) i 1028 barres infinites (persistència total 11780). Això indica que, en ambdós casos, la

xarxa està formada per moltes components connexes que es mantenen separades durant tota la filtració (barres infinites), però en topologia algebraica la fragmentació és encara més gran, ja que hi ha més barres infinites. Tot i així, també hi ha més petits grups inicials que s'acaben unint a altres components (barres finites). Aquesta diferència reflecteix una estructura col·laborativa més compartimentada i dinàmica en topologia algebraica, on és més freqüent que petits grups es connectin progressivament, però sense arribar a consolidar una xarxa globalment connectada.

Cal destacar que el càlcul de l'homologia persistent sobre la clausura simplicial ha resultat significativament més costós computacionalment, a causa de l'augment del nombre de simplexs que es generen automàticament durant el procés.

Conclusions

Aquest treball ha tingut com a objectiu principal comprendre i presentar l'homologia persistent aplicada a hipergrafs, aprofundint en la teoria existent i mostrant-ne el funcionament mitjançant un exemple pràctic. S'ha buscat entendre les construccions fonamentals i il·lustrar-ne l'aplicabilitat. A partir d'aquest marc, s'han extret diverses conclusions.

En primer lloc, l'ús de la clausura simplicial per definir l'homologia d'un hipergraf constitueix un enfocament clàssic i intuïtiu. Aquest mètode permet aplicar directament els instruments de la topologia algebraica tradicional, i en particular en dimensió zero reflecteix fidelment la connexió entre vèrtexs, tal com mostra el lema 3.1.3, on es demostra que el nombre de components connexes d'un hipergraf coincideix amb el de la seva clausura simplicial. Tanmateix, aquest enfocament presenta limitacions rellevants. El procés de clausura simplicial introdueix estructures addicionals (com arestes i altres símplexs) que no formen part de l'hipergraf original, cosa que pot alterar la topologia percebuda i afegir informació fictícia. Aquest fenomen es tradueix en la possible aparició de cicles o cavitats inexistents, i suposa, a més, una càrrega computacional considerable per al càlcul de l'homologia persistent.

Davant d'aquestes limitacions, el model d'homologia restringida, que ha estat el nucli central d'aquest treball, permet analitzar la topologia dels hipergrafs de manera directa i fidel a la informació explícita. Només es consideren les hiperarestes definides originalment, sense afegits artificials. Aquesta construcció es fonamenta en una base algebraicotopològica sòlida, com demostra la coincidència entre les homologies obtingudes a partir dels complexos ínfim i suprem.

Pel que fa a la interpretació geomètrica, l'homologia restringida ofereix una lectura clara en dimensió zero: dos vèrtexs només es consideren connectats si són 0-hiperarestes i hi ha una 1-hiperaresta que els uneix explícitament. Aquesta propietat evita inferències implícites, però també implica que la homologia pot ser trivial ($H_0 = 0$) en estructures aparentment riques si no hi ha vèrtexs declarats explícitament. Aquest comportament reflecteix amb precisió la informació que ha estat representada de forma explícita al model, sense afegir connexions no especificades. Això també vol dir que, si es busca una mesura més general de la connexió global (com ara el nombre total de components connexes), pot ser més adequat utilitzar l'homologia de la clausura simplicial. Per tant, segons l'objectiu de l'anàlisi, pot ser convenient escollir una o altra aproximació per al cas de H_0 . A més, s'ha vist amb un exemple que l'homologia restringida no coincideix sempre amb la de la clausura simplicial, la qual cosa reforça la idea que cadascun dels enfocaments capta aspectes diferents de l'estructura del hipergraf.

Un altre aspecte rellevant observat és que, en homologia restringida, els cicles, tant en dimensió 1 com en superiors, poden aparèixer de manera natural i tenir una interpretació geomètrica clara. A diferència de la clausura simplicial, on s'afegit automàtic de totes les

cares de les hiperarestes pot eliminar cicles prematurament, en el model restringit aquests cicles poden persistir si les hiperarestes de dimensió 1 no proporcionen suficient informació de vora per “tapar-los”. Això fa que, en alguns casos, un cicle es mantingui en la versió restringida però desaparegui en la clausura simplicial. Per exemple, les hiperarestes (a, b) , (b, c) , (a, c) i la 3-hiperaresta (a, b, c, d) , la clausura simplicial genera (a, b, c) , que mata el cicle, mentre que en el model restringit aquest cicle pot continuar viu si no hi ha una hiperaresta que realment compleixi el paper de relació de frontera.

Cal destacar també que, en cap dels casos analitzats al capítol 6, no s’han detectat classes d’homologia en dimensió 2 (H_2). Aquest fet es pot explicar perquè la xarxa no és prou densa per generar cavitats bidimensionals persistents. A més, en construir la clausura simplicial, l’afegit automàtic de tots els subconjunts de cada hiperaresta fa molt improbable la formació de cavitats.

Des del punt de vista computacional, s’ha implementat i analitzat un algorisme per calcular l’homologia persistent d’hipergrafs, adaptant els mètodes clàssics als nous tipus de complexos definits. S’ha constatat que l’estructura combinatòria dels hipergrafs pot provocar un creixement considerable de la complexitat computacional, fet que reforça la necessitat d’optimització futura.

Aquest treball obre la porta a diverses línies de recerca, com l’aplicació directa d’aquesta homologia a dades reals estructurades com a hipergrafs (per exemple, en xarxes socials o sistemes biològics), la millora de l’algorisme desenvolupat, i el tractament de filtracions més complexes. A més, tot i que l’objectiu no ha estat comparar la topologia algebraica amb altres branques com la topologia general, el present estudi demostra com els conceptes clàssics poden ser adaptats i estesos a estructures de dades no convencionals. En definitiva, l’homologia persistent d’hipergrafs representa una eina poderosa per a la representació i anàlisi de relacions d’ordre superior en dades complexes.

Bibliografia

- [1] M. E. Aktas, T. Nguyen, R. Riza, M. I. Islam i E. Akbas. *Hypergraph classification via persistent homology*. arXiv preprint, arXiv:2306.11484, 2023.
- [2] N. Atienza, R. Gonzalez-Diaz i M. Soriano-Trigueros, *On the stability of persistent entropy and new summary functions for Topological Data Analysis*, Pattern Recognition, vol. 107, 107509, 2020.
- [3] N. Atienza, R. Gonzalez-Diaz i M. Rucco. *Persistent Entropy for Separating Topological Features from Noise in Vietoris-Rips Complexes*, Journal of Intelligent Information Systems, vol. 52, núm. 3, pp. 637–655, 2019.
- [4] M. A. Bahmanian i M. Šajna. *Hypergraphs: connection and separation*, arXiv preprint, arXiv:1504.04274, University of Ottawa, 2015.
- [5] S. Bressan, J. Li, S. Ren, i J. Wu. *The embedded homology of hypergraphs and applications*. Asian Journal of Mathematics, 23(3):479–500, 2019.
- [6] T. K. Dey i Y. Wang, *Computational Topology for Data Analysis*. Cambridge University Press, 2022.
- [7] H. Edelsbrunner i J. Harer. *Computational Topology: An Introduction*. American Mathematical Society, 2010.
- [8] H. Edelsbrunner, D. Letscher i A. Zomorodian. *Computing Persistent Homology*, Discrete and Computational Geometry, vol. 28, núm. 4, pp. 511–533, 2002.
- [9] A. El Yaagoubi Bourakna, M. K. Chung i H. Ombao. *Modeling and Simulating Dependence in Networks Using Topological Data Analysis*, arXiv preprint, arXiv:2209.10416, 2022.
- [10] E. Gasparovic, E. Purvine, R. Sazdanovic, B. Wang, Y. Wang i L. Ziegelmeier. arXiv preprint, arXiv:2409.18310, 2024.
- [11] R. Ghrist. *Barcodes: the persistent topology of data*. University of Pennsylvania, 2008.
- [12] R. E. Green, C. A. Joslyn, A. Myers, M. G. Rawson i M. Robinson. *The Topological Structures of the Orders of Hypergraphs*. arXiv preprint, arXiv:2504.11760, 2025.
- [13] A. Grigor’yan, Y. Lin, Y. Muranov i S.T. Yau, *Homologies of path complexes and digraphs*, arXiv preprint, arXiv:1207.2834, 2012.
- [14] A. Hatcher. *Algebraic topology*, Cambridge University Press, 2002.

- [15] S. Klamt, U. U. Haus, F. Theis, *Hypergraphs and cellular networks*, PLoS Computational Biology, 5, e1000385, 2009.
- [16] E. V. Konstantinova i V. A. Skorobogatov, *Application of hypergraph theory in chemistry*, Discrete Mathematics, 235, 365-383, 2001.
- [17] X. Liu, X. Wang, J. Wu i K. Xia. *Hypergraph-based persistent cohomology (HPC) for molecular representations in drug design*, Briefings in Bioinformatics, 22(5):bbaa411, 2021.
- [18] X. Liu, H. Feng, J. Wu i K. Xia. *Computing Hypergraph Homology*, Chern Institute of Mathematics, Nankai University, 2024.
- [19] J. Liu, D. Chen, J. Li i J. Wu. *Neighborhood hypergraph model for topological data analysis*, Computational and Mathematical Biophysics, 10:262–280, 2022.
- [20] V. Navarro i P. Pascual. *Topologia algebraica*. Departament d'Àlgebra i Geometria Universitat de Barcelona i Departament de Matemàtica Aplicada I, Universitat Politècnica de Catalunya, 1999.
- [21] V. B. Pacheco i M. Giannopoulos. *Persistent Homology in Medical Image Processing: A Literature Review*, medRxiv preprint, 2025.
- [22] A.D. Parks i S.L. Lipscomb. *Homology and hypergraph acyclicity: a combinatorial invariant for hypergraphs*, Naval Surface Warfare Center, 1991.
- [23] K. Xia i G. W. Wei. *Persistent homology analysis of protein structure, flexibility and folding*, International Journal for Numerical Methods in Biomedical Engineering, vol. 30, núm. 8, pp. 814–844, 2014.

Apèndix. Anàlisi per a diverses àrees

En aquest apèndix es presenten els codis de barres i els diagrames de persistència per a altres àrees de les matemàtiques, juntament amb dues taules 3 i 4 per als descriptors numèrics calculats en cada cas.

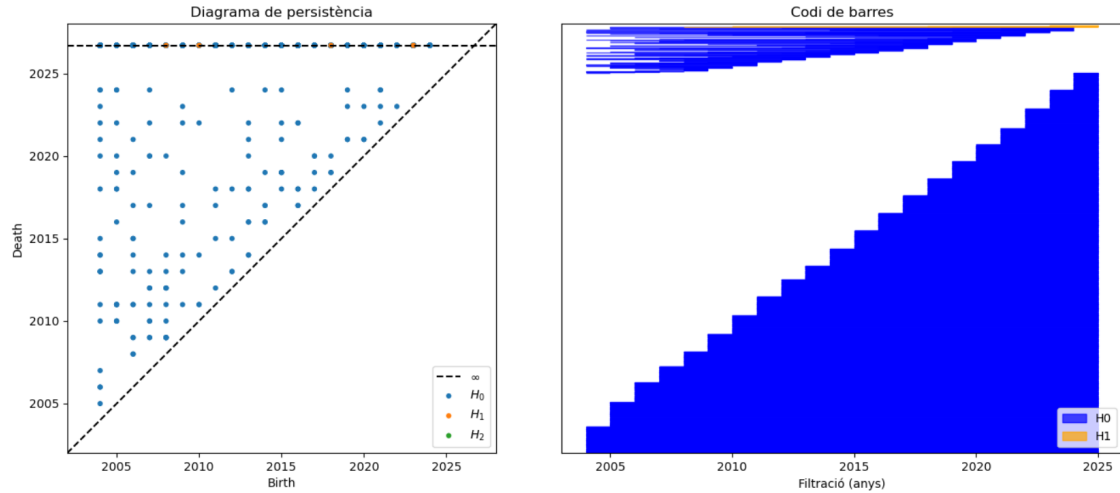


Figura 5: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia restringida dels hipergrafs de la filtració corresponent a l'àrea de geometria algebraica.

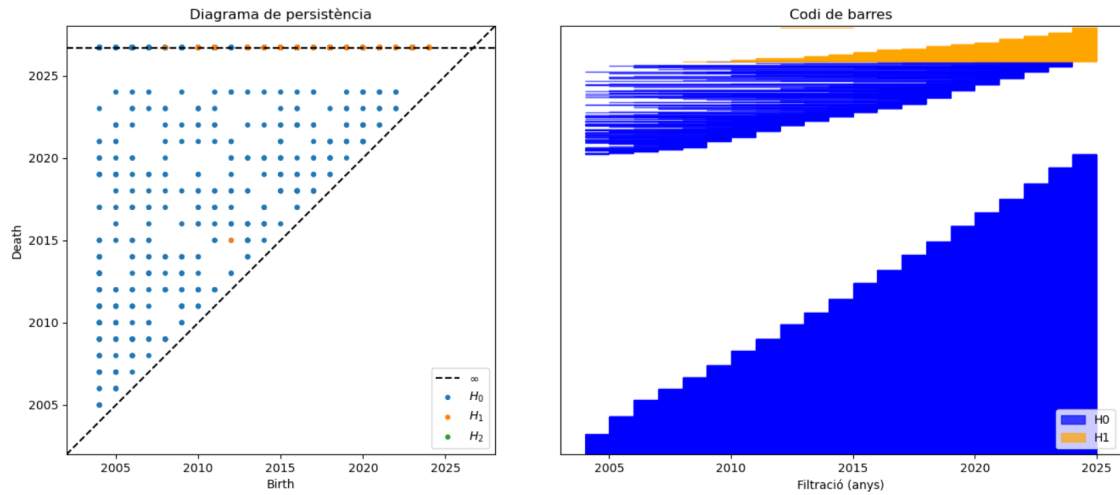


Figura 6: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia de la clausura simplicial dels hipergrafs de la filtració corresponent a l'àrea de geometria algebraica.

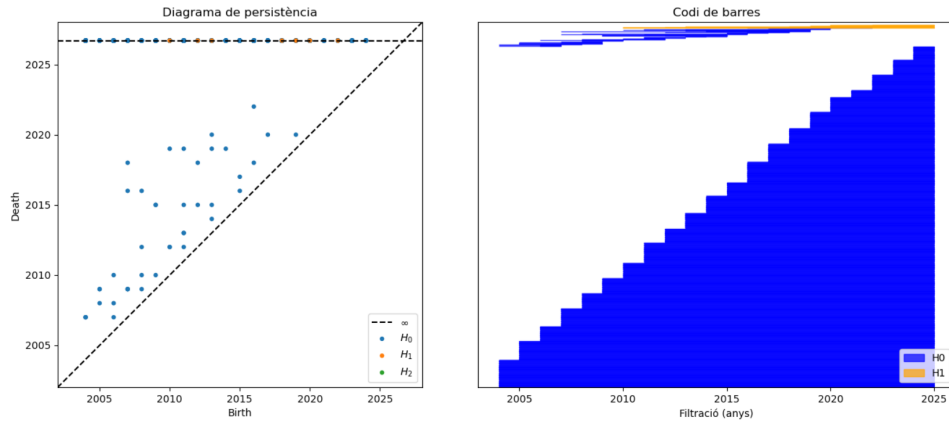


Figura 7: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia restringida dels hipergrafs de la filtració corresponent a l'àrea d'anàlisi de PDES.

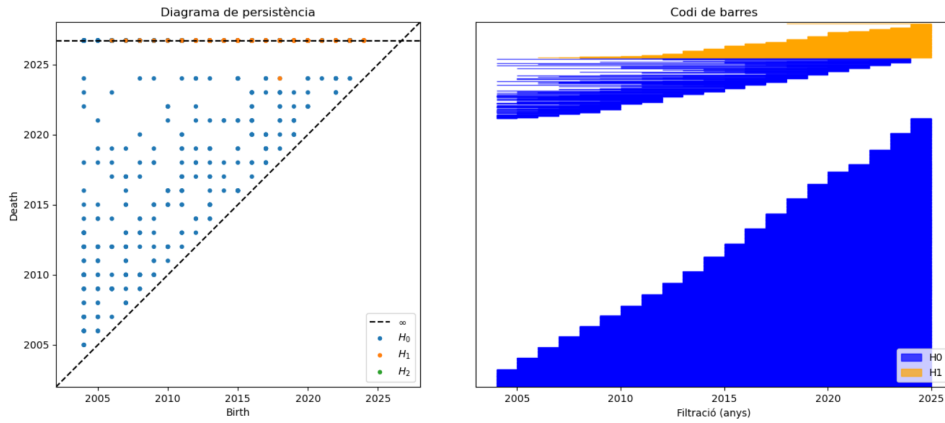


Figura 8: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia de la clausura simplicial dels hipergrafs de la filtració corresponent a l'àrea d'anàlisi de PDES.

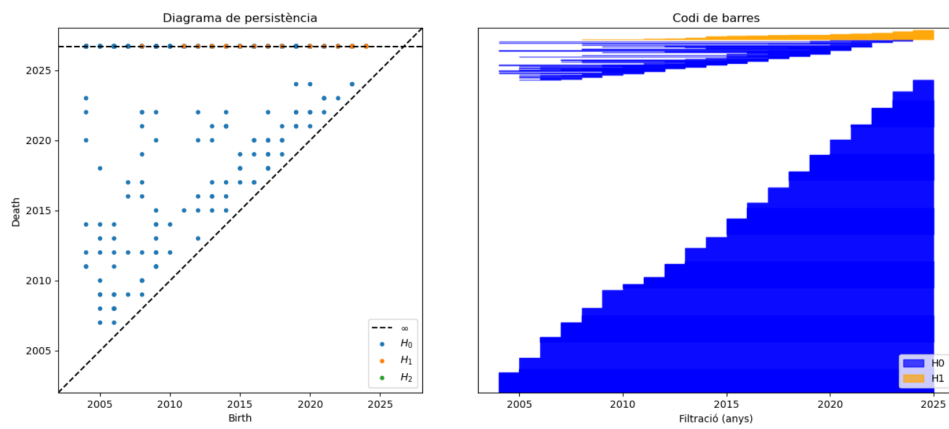


Figura 9: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia restringida dels hipergrafs de la filtració corresponent a l'àrea de teoria de categories.

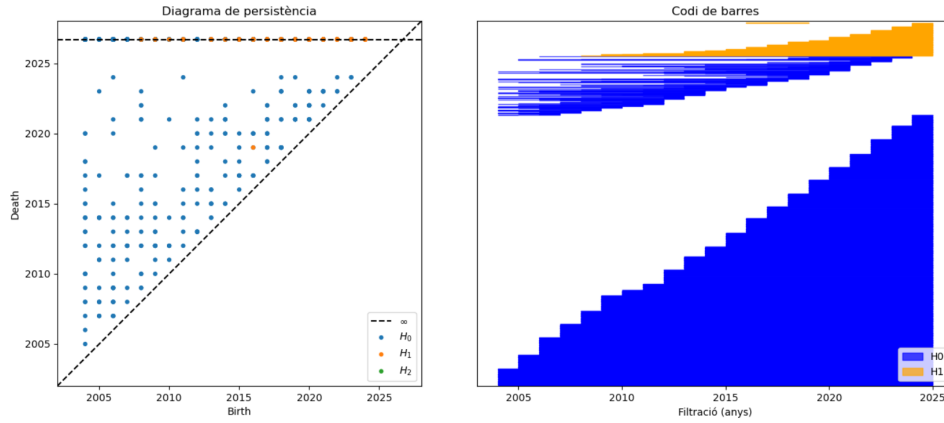


Figura 10: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia de la clausura simplicial dels hipergrafs de la filtració corresponent a l'àrea de teoria de categories.

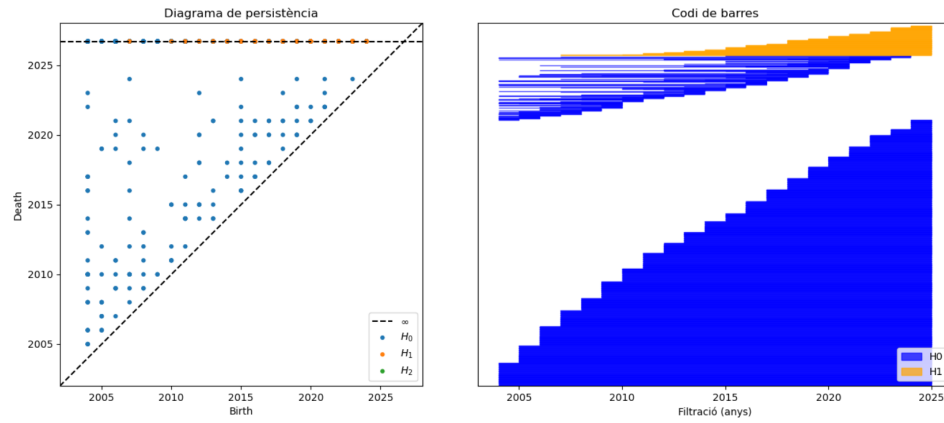


Figura 11: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia restringida dels hipergrafs de la filtració corresponent a l'àrea d'àlgebra commutativa.

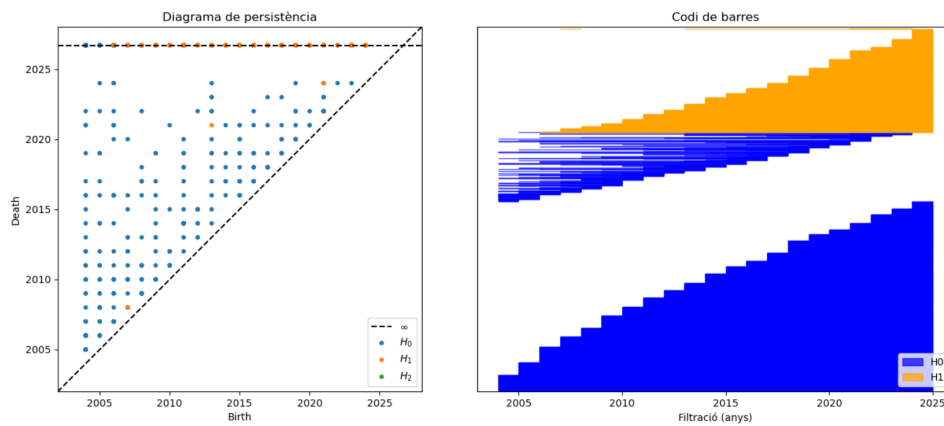


Figura 12: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia de la clausura dels hipergrafs de la filtració corresponent a l'àrea d'àlgebra commutativa.

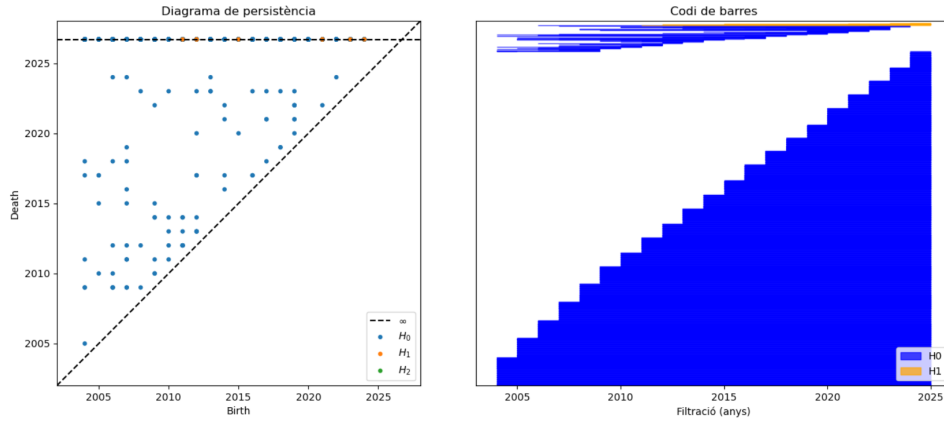


Figura 13: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia restringida dels hipergrafs de la filtració corresponent a l'àrea de geometria diferencial.

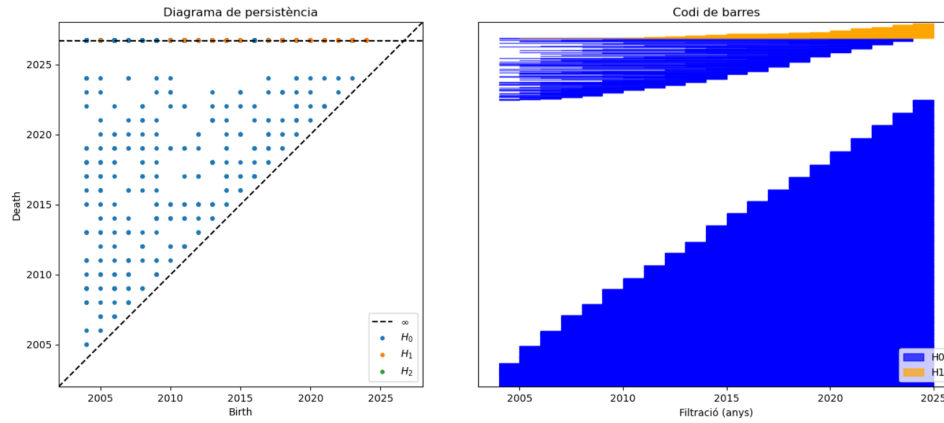


Figura 14: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia de la clausura simplicial dels hipergrafs de la filtració corresponent a l'àrea de geometria diferencial.

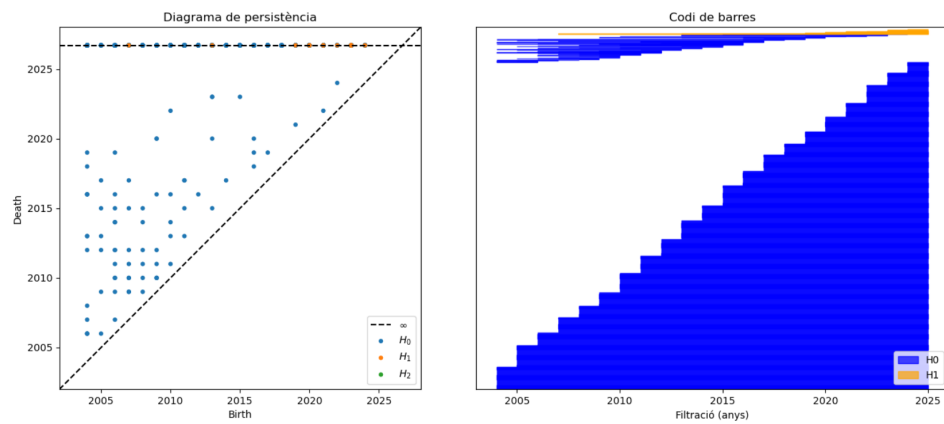


Figura 15: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia restringida dels hipergrafs de la filtració corresponent a l'àrea de sistemes dinàmics.

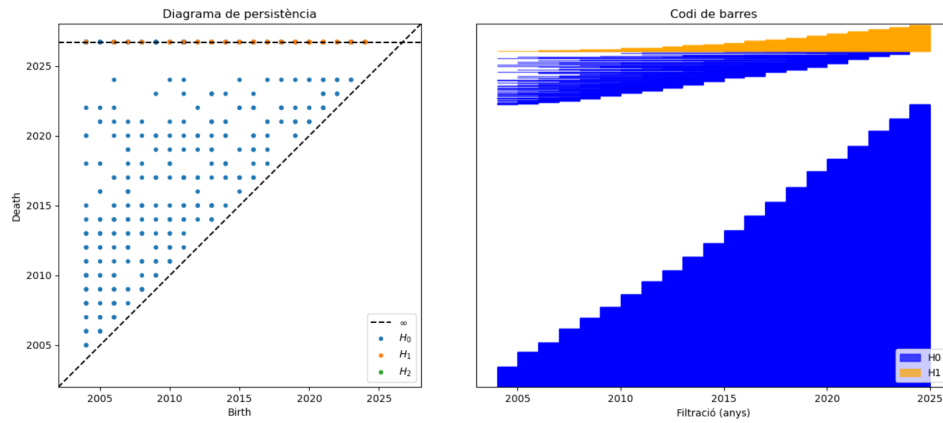


Figura 16: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia de la clausura dels hipergrafs de la filtració corresponent a l'àrea de sistemes dinàmics.

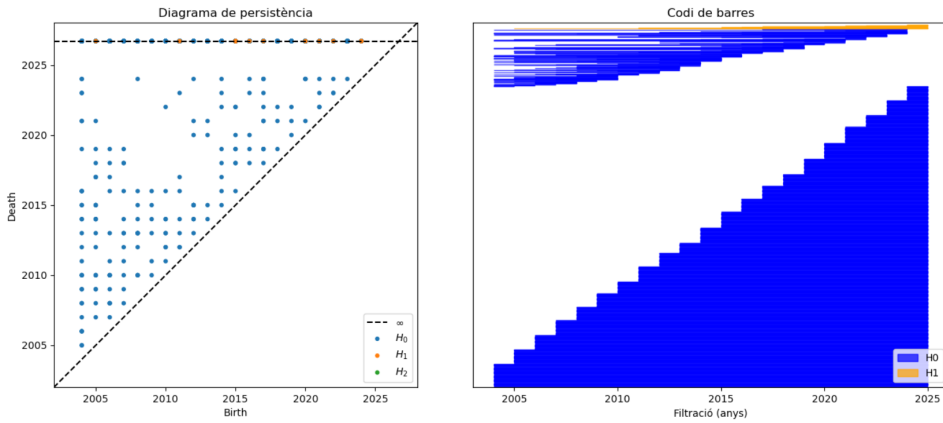


Figura 17: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia restringida dels hipergrafs de la filtració corresponent a l'àrea de topologia geomètrica.

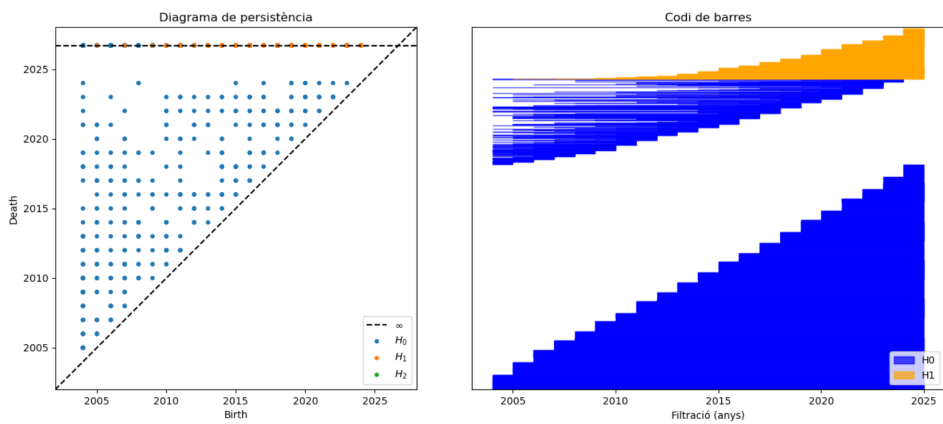


Figura 18: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia de la clausura simplicial dels hipergrafs de la filtració corresponent a l'àrea de topologia geomètrica.

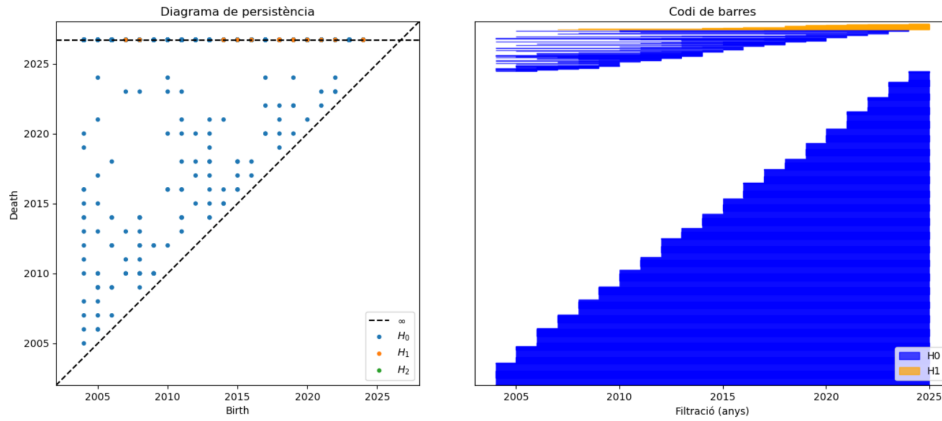


Figura 19: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia restringida dels hipergrafs de la filtració corresponent a l'àrea de teoria de grups.

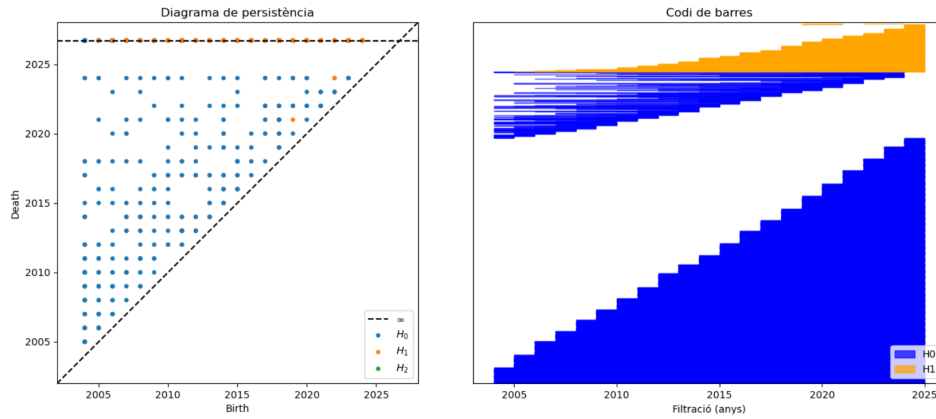


Figura 20: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia de la clausura simplicial dels hipergrafs de la filtració corresponent a l'àrea de teoria de grups.

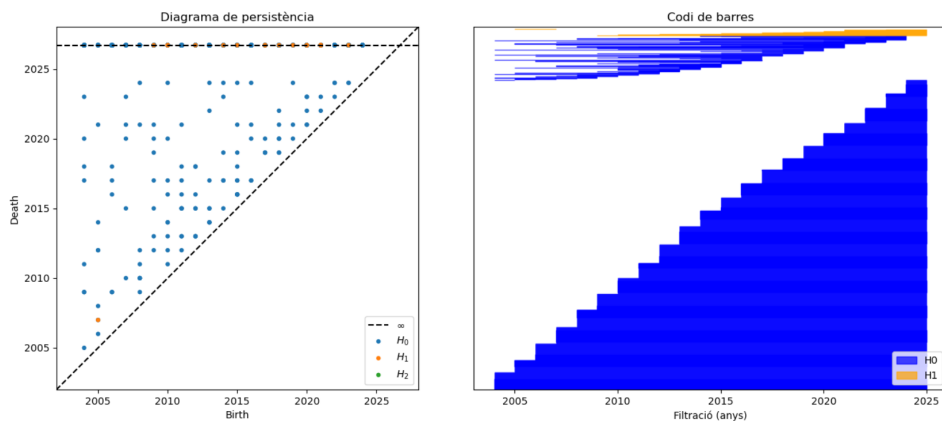


Figura 21: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia restringida dels hipergrafs de la filtració corresponent a l'àrea de teoria K i homologia.

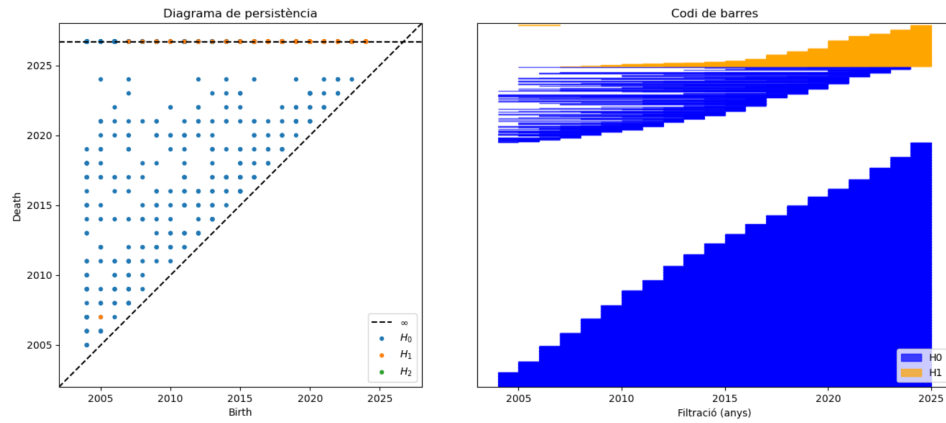


Figura 22: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia de la clausura simplicial dels hipergrafs de la filtració corresponent a l'àrea de teoria K i homologia.

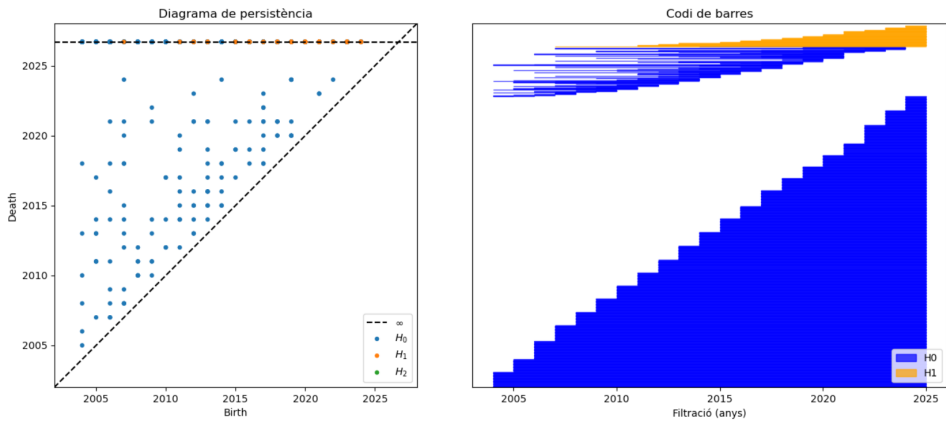


Figura 23: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia restringida dels hipergrafs de la filtració corresponent a l'àrea de lògica.

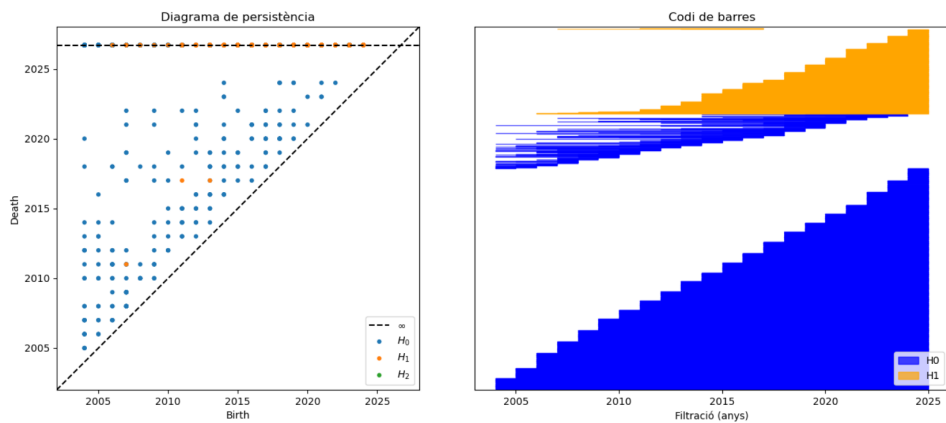


Figura 24: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia de la clausura simplicial dels hipergrafs de la filtració corresponent a l'àrea de lògica.

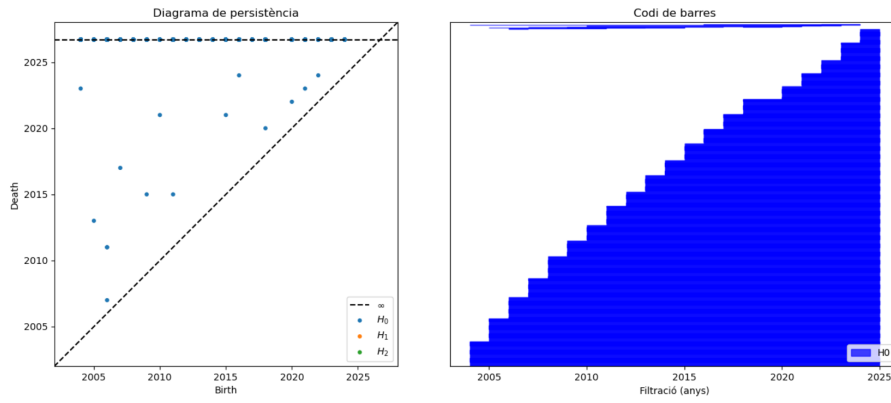


Figura 25: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia restringida dels hipergrafs de la filtració corresponent a l'àrea de física matemàtica.

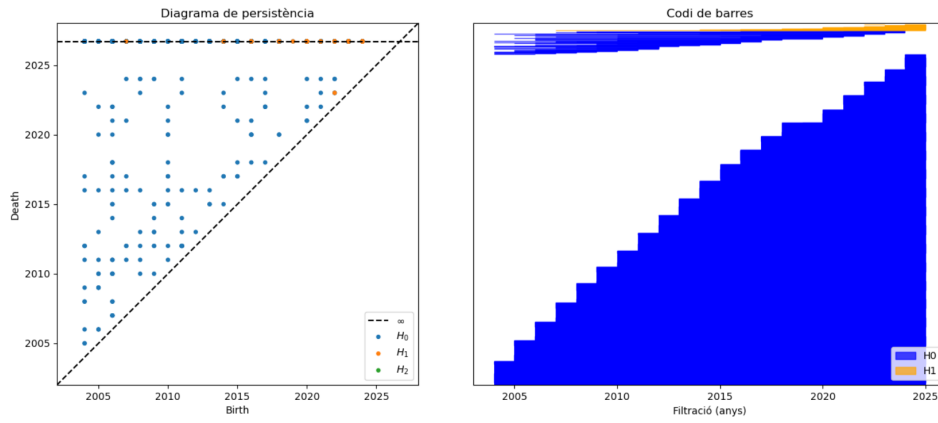


Figura 26: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia de la clausura simplicial dels hipergrafs de la filtració corresponent a l'àrea de física matemàtica.

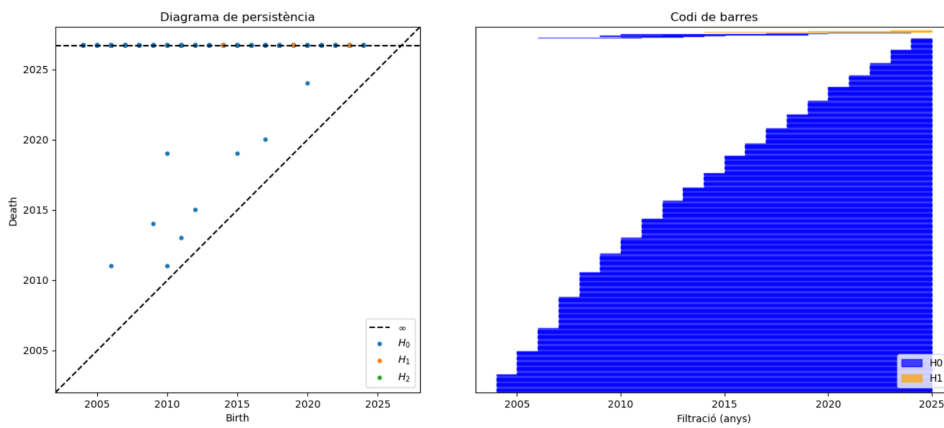


Figura 27: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia restringida dels hipergrafs de la filtració corresponent a l'àrea d'anàlisi numèrica.

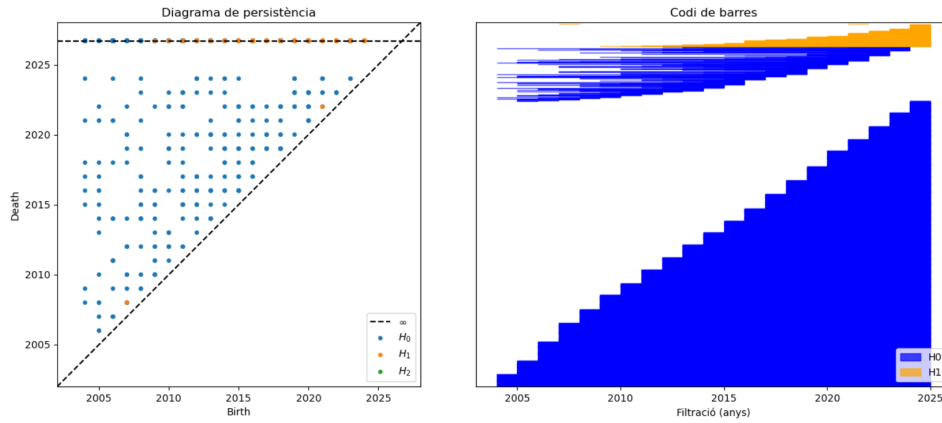


Figura 28: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia de la clausura simplicial dels hipergrafs de la filtració corresponent a l'àrea d'anàlisi numèrica.

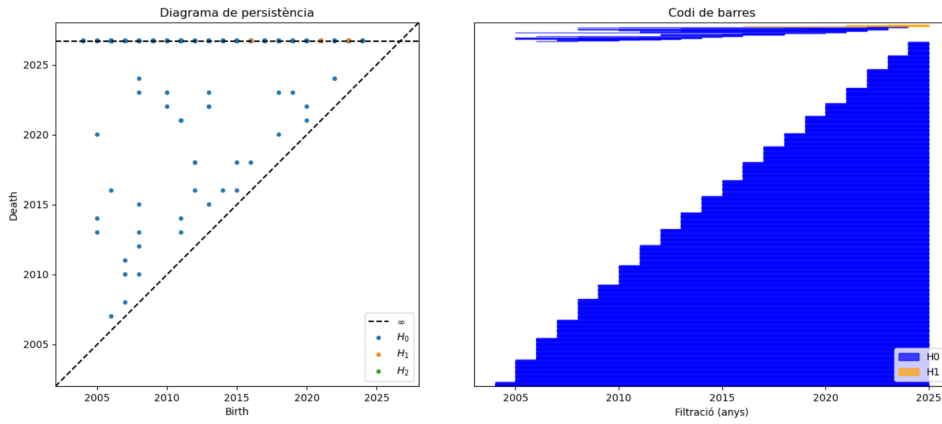


Figura 29: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia restringida dels hipergrafs de la filtració corresponent a l'àrea de probabilitat.

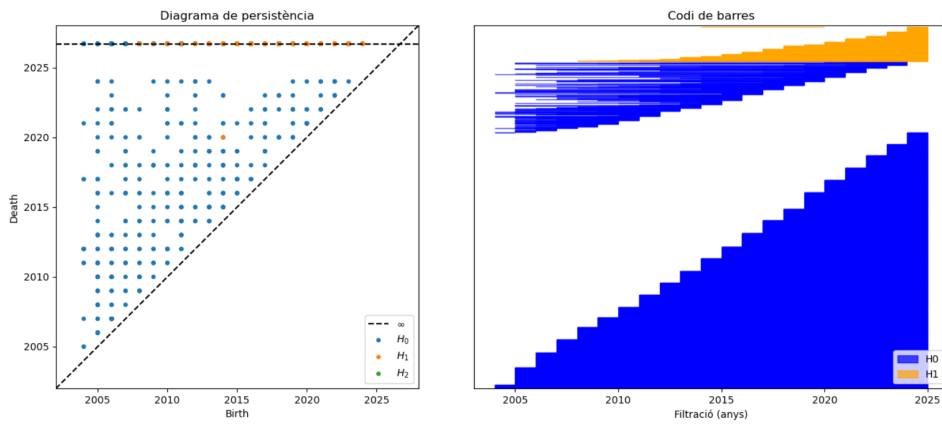


Figura 30: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia de la clausura simplicial dels hipergrafs de la filtració corresponent a l'àrea de probabilitat.

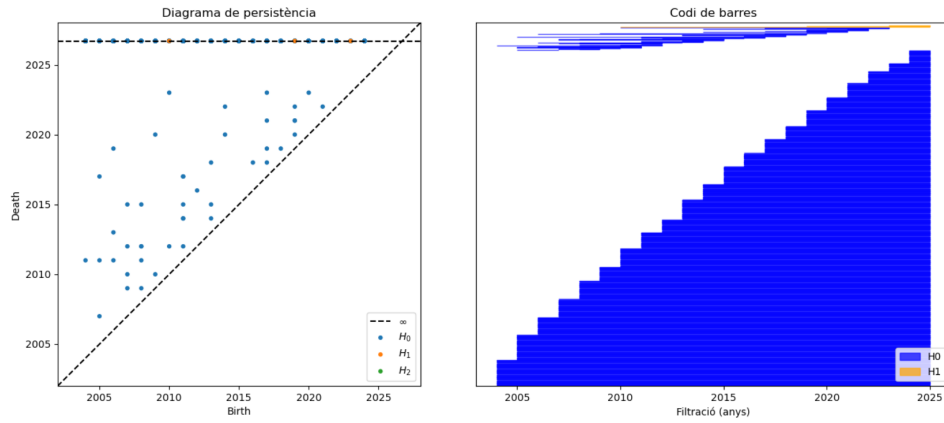


Figura 31: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia restringida dels hipergrafs de la filtració corresponent a l'àrea d'estadística.

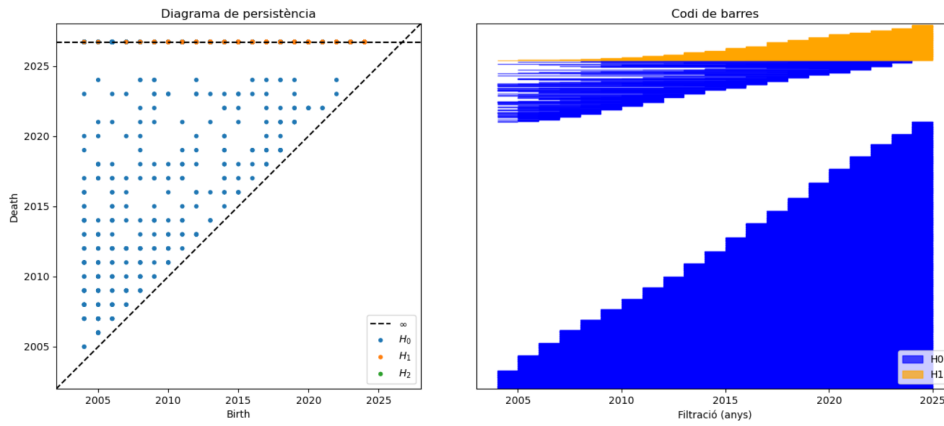


Figura 32: Diagrama de persistència (esquerra) i codi de barres (dreta) obtinguts a partir de l'homologia de la clausura simplicial dels hipergrafs de la filtració corresponent a l'àrea d'estadística.

Àrea	Hom. (H_0)	PT(fin.)	PT(inf.)	EP(fin.)	EP(inf.)	#Bar(fin.)	#Bar(inf.)
Geom. Alg.	restringida	883	12130	6.6994	9.8741	141	1169
	clausura	2304	12572	8.2009	9.9496	383	1239
An. PDEs	restringida	156	8274	5.0774	9.3126	42	773
	clausura	1479	12203	7.8147	9.9585	286	1266
Teo. Cat.	restringida	527	8696	6.3792	9.4281	111	861
	clausura	1071	10352	7.3855	9.6650	222	1013
Àlg. Comm.	restringida	689	7049	6.7331	9.0076	148	613
	clausura	1199	8313	7.5357	9.2372	263	722
Geom. Dif.	restringida	437	11660	5.9332	9.7730	84	107
	clausura	1784	14415	7.7804	10.0977	292	1356
Sist. Din.	restringida	424	9335	5.9199	9.5043	76	880
	clausura	1508	14703	7.7489	10.1970	276	1474
Topo Geom.	restringida	1021	9372	7.0741	9.5128	175	912
	clausura	1989	9533	8.1160	9.5466	360	947
Teo. Grups	restringida	611	9386	6.4785	9.5070	119	895
	clausura	1480	10955	7.7575	9.7575	290	1072
K -teo i Hom.	restringida	599	8934	6.4378	9.4427	122	850
	clausura	1558	9917	7.7533	9.5538	290	932
Lògica	restringida	577	7158	6.5431	9.1387	121	706
	clausura	881	8232	7.3016	9.3269	200	811
Fis. Mat.	restringida	91	11594	3.5383	9.7346	15	1032
	clausura	804	20831	6.6591	10.5938	135	1865
An. Num.	restringida	36	6354	2.9683	8.8737	9	557
	clausura	1343	13467	7.5427	10.0475	249	1307
Proba.	restringida	232	9069	4.9003	9.4977	39	882
	clausura	1844	12231	8.0249	9.9660	346	1242
Estad.	restringida	193	7206	5.0995	9.0694	44	642
	clausura	1579	12461	7.7575	9.9467	282	1226

Taula 3: Per al grup d'homologia de dimensió 0, persistència total (PT) i entropia de persistència (EP), separades per barres finites (fin.) i infinites (inf.), per a diverses àrees i dues variants de càlcul de l'homologia. #Bar indica el nombre total de barres de cada tipus.

Àrea	Hom. (H_1)	PT(fin.)	PT(inf.)	EP(fin.)	EP(inf.)	#Bar(fin.)	#Bar(inf.)
Geom. Alg.	restringida	0	37	0.0000	1.6199	0	4
	clausura	3	747	0.0000	6.5831	1	139
An. PDEs	restringida	0	54	0.0000	2.5790	0	7
	clausura	6	1026	0.0000	6.9345	1	160
Teo. Cat.	restringida	0	152	0.0000	4.1127	0	26
	clausura	3	636	0.0000	6.4425	1	123
Àlg. Comm.	restringida	0	408	0.00	5.7039	0	69
	clausura	12	2746	1.1887	8.1918	3	391
Geom. Dif.	restringida	0	53	0.0000	2.4792	0	8
	clausura	0	355	0.0000	5.5663	0	69
Sist. Din.	restringida	0	53	0.0000	2.9608	0	13
	clausura	0	903	0.0000	6.6563	0	137
Topo Geom.	restringida	0	83	0.0000	3.0882	0	12
	clausura	0	1067	0.0000	7.2263	0	212
Teo. Grups	restringida	0	113	0.0000	9.5070	0	17
	clausura	4	1351	1.0000	7.2645	2	207
K -teo i Hom.	restringida	2	120	0.0000	3.8117	1	17
	clausura	2	765	0.0000	6.8330	1	157
Lògica	restringida	0	273	0.0000	5.2025	0	50
	clausura	14	1819	1.5567	7.8824	3	308
Fis. Mat.	restringida	0	0	0.0000	0.0000	0	0
	clausura	1	128	0.0000	4.4096	1	33
An. Num.	restringida	0	16	0.0000	1.1982	0	3
	clausura	2	448	1.0000	6.0389	2	102
Proba.	restringida	0	15	0.0000	9.4977	0	4
	clausura	6	834	0.0000	6.8533	1	166
Estad.	restringida	0	20	0.0000	1.0763	0	3
	clausura	0	1088	0.0000	6.9543	0	162

Taula 4: Per al grup d'homologia de dimensió 1, persistència total (PT) i entropia de persistència (EP), separades per barres finites (fin.) i infinites (inf.), per a diverses àrees i dues variants de càlcul de l'homologia. #Bar indica el nombre total de barres de cada tipus.