

Introducció a l'anàlisi topològica de dades i aplicacions

ESTHER IBÁÑEZ-MARCELO

Reserveu-vos el dret a pensar, perquè fins i tot pensar malament és millor que no pensar gens.

Hipàtia d'Alexandria (350–370 dC - 415 dC)

Resum: L'anàlisi topològica de dades (*Topological Data Analysis*, TDA) ha tingut un gran auge els darrers deu anys. En aquest article en presentem una introducció a partir dels conceptes i les definicions fonamentals que ens permeten descriure les eines principals d'aquesta nova branca en la qual es dona més importància a la forma que a la mètrica. Veurem en què consisteix l'homologia persistent i com pot ser aplicada en un ampli rang de camps científics, com, per exemple, la neurociència, el reconeixement d'imatges i la genètica, entre d'altres.

Paraules clau: homologia persistent, diagrama de persistència, codi de barres, *Mapper*, anàlisi de dades.

Classificació MSC2010: 55N31, 62R40, 68T09.

1 Introducció

En els últims vint anys s'han desenvolupat un conjunt de mètodes basats en la topologia algebraica com a noves eines per a l'anàlisi de dades. És el que s'anomena *anàlisi topològica de dades* (*topological data analysis*; d'ara endavant, per simplificar-ho, usarem la sigla TDA per a referir-nos a aquests mètodes). A grans trets, la TDA deixa de banda les mètriques i se centra a extreure la *forma* de les dades. És a dir, dona importància a la topologia i no a les mètriques.

La TDA ens permet extreure informació de dades amb soroll i d'alta dimensionalitat, fet que normalment és un gran repte amb les metodologies basades en distàncies entre punts donades per una mètrica.

Dins de la TDA, l'eina principal és l'anomenada *homologia persistent* [12, 42], la qual ens permet obtenir un resum multiescala d'un conjunt de dades. A més a més, aquesta aproximació té garanties de robustesa respecte de les perturbacions de les dades [29, 22, 26].

La topologia tracta la informació geomètrica des del punt de vista qualitatiu. És a dir, estudia com estan relacionats els conjunts de punts dins d'un espai. Aquesta manera d'estudiar les propietats geomètriques depèn molt menys del sistema mètric (o de coordenades) triat i ens és molt més útil quan només volem entendre l'estructura de les dades, sense entrar detalladament en les mètriques. A més, amb l'homologia persistent veurem que podem definir l'escala a la qual volem estudiar la topologia de les dades. Per tant, a part de ser un mètode a primera vista més qualitatiu, també és quantificable, tal com veurem més endavant.

En els darrers anys, la computació de l'homologia persistent i la representació dels seus resultats han evolucionat molt ràpidament gràcies també a la tecnologia i a l'increment del potencial de les màquines actuals, fet que ha permès que un càlcul que fa cinc anys no era possible de fer degut al volum de dades, avui dia sí que ho sigui. A més, al llarg dels anys, han anat sortint diverses formes de representar els resultats de l'homologia persistent, de manera que fins i tot aquestes característiques homològiques es poden usar com a dades d'entrada d'algoritmes d'aprenentatge automàtic (*machine learning*, ML). D'altra banda, l'increment de programari, principalment com a llibreria de llenguatges tan usats com Python, ha permès que molts més usuaris no experts en la teoria d'homologia persistent poguessin utilitzar aquesta tècnica i, per tant, que l'ús dels mètodes de TDA es difongués d'una manera més ràpida.

D'altra banda, en relació amb les tècniques de TDA, també hi podem incloure l'anomenat *algoritme Mapper* [89], que divideix les dades en clústers a partir d'un filtre de clusterització i extreu l'estructura de les dades a nivell de graf. Les mateixes dades es poden estudiar a partir de filtres diversos, d'algoritmes diferents d'anàlisi de clústers i també a escales distintes.

1.1 Fem una mica d'història

La idea principal de les eines de TDA rau en abandonar les mètriques estàndard entre punts i centrar-se en la forma de les dades al llarg d'un rang d'escales. Es tracta de donar una visió global de les nostres dades o espai topològic, en comptes de focalitzar-nos en el seu comportament a nivell mètric. A més a més, l'homologia persistent garanteix robustesa davant de perturbacions de les dades; en aquest sentit, en els darrers anys se n'ha millorat d'una manera significativa la computació per a conjunts de dades cada cop més grans. Aquestes eines també poden ser molt útils en espais i dades en què és difícil definir una mètrica.

Si anem enrere en el temps, podem dir que l'homologia persistent va néixer a partir dels treballs sobre *funcions de mida* (*size functions*) de Patrizio Frosini, als anys noranta [47], i de Vanessa Robins [86], a finals dels noranta, en els quals s'inferia la topologia dels atractors en sistemes dinàmics a partir de dades experimentals.

Cal dir que després hem d'esperar fins als anys 2000, en què apareixen els treballs en àmbit computacional i geomètric d'Edelsbrunner, Letscher, Zomorodian i Delfinado [36, 43], els quals permeten calcular d'una manera eficient

els nombres de Betti persistents i, des del punt de vista algebraic, tenim els resultats de Carlsson [99].

Podem dir que durant la primera dècada del 2000 se'n desenvolupa la teoria, mentre que a partir del 2010 s'amplia l'ús d'aquestes eines en problemes reals. Els resultats d'estabilitat en homologia persistent a partir del 2007 ([29]) van permetre confiar en aquesta tècnica i estendre'n l'ús en àmbits més aplicats. Gràcies a millores en el càlcul teòric i pràctic de l'homologia [5], trobem aplicacions que van des de la neurociència amb tractaments a nivells diferents [33, 60, 67, 7, 52] fins al reconeixement d'imatges [71], passant per la genètica [81, 21], alguns problemes de cobertura en xarxes de comunicació [35], l'estudi de sistemes espacials [46] i la classificació de xarxes [90, 62].

Podeu trobar una bona revisió sobre la història de l'homologia persistent a [80] i sobre el càlcul de l'homologia persistent, a [41]. D'altra banda, en l'àmbit més computacional, a [79] es presenta un bon resum de llibreries i tècniques per a calcular l'homologia persistent.

2 Definicions bàsiques: cadenes, símplex i homologia

A continuació presentarem d'una manera més tècnica alguns conceptes bàsics per a poder entendre i definir l'homologia, i, tot seguit, veurem l'homologia persistent. Se'n pot trobar més informació en les referències clàssiques de topologia algebraica, com ara els llibres de Hatcher [56] i de Munkres [77], i si es vol aprofundir en conceptes com filtracions, homologia simplicial i homologia persistent, es poden consultar [43, 12, 42, 99].

2.1 Símplex, complex i cadena simplicial, i aplicació vora

De manera intuïtiva, podem definir un *complex simplicial* \mathcal{K} a \mathbb{R}^N (o, també, la seva anomenada *realització geomètrica*) com un espai topològic construït per símplexs. Un *símplex* de dimensió n a \mathbb{R}^N , o *n-símplex*, $\sigma_n = \{p_0, \dots, p_{n-1}\}$, es defineix pels seus vèrtexs p_i i es pot pensar com l'envolupant convexa d'aquest conjunt de punts. Si considerem els símplexs immersos en l'espai, aleshores els podem visualitzar com a punts o vèrtexs (en dimensió 0), arestes o segments (en dimensió 1), triangles (en dimensió 2), i així successivament. Cal dir que, des del punt de vista de grafs, els símplexs són les anomenades *cliques* (o *n-cliques*); és a dir, subgrafs complets: nodes (en dimensió 0), arestes (en dimensió 1), triangles plens (en dimensió 2), tetraedres (en dimensió 3), etc. (vegeu la figura 1).

Ara bé, de manera formal, un *complex simplicial abstracte* és una estructura combinatòria que consisteix en una parella (X, \mathcal{K}) , on X és un conjunt i \mathcal{K} és una família de subconjunts no buits de X tal que, si $\sigma \in \mathcal{K}$ i $\tau \subseteq \sigma$, aleshores $\tau \in \mathcal{K}$.

A partir d'això, donat un $\sigma \in \mathcal{K}$, si $\tau \subseteq \sigma$, aleshores diem que τ és una *cara* de σ .

Donat un n -símplex $\sigma_n = \{p_0, \dots, p_{n-1}\}$ i una permutació ρ , considerem el símplex $\sigma' = \{p_{\rho(0)}, \dots, p_{\rho(n-1)}\}$. Si ρ té signe positiu (resp., negatiu), diem que σ i σ' tenen (resp., no tenen) la mateixa orientació.

Donat un complex simplicial, podem crear-ne la realització topològica, és a dir, l'espai topològic del qual el complex simplicial seria una triangulació. Així, d'ara endavant, doncs, quan parlem d'espai topològic ens referirem a l'espai topològic induït per un complex simplicial.

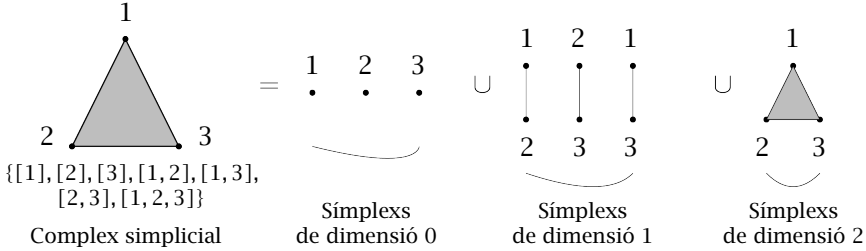


FIGURA 1: Exemple de complex simplicial i les seves parts. Un triangle ple està compost per tres 0-símplexs (punts), tres 1-símplexs (segments) i un 2-símplex (el triangle ple $[1, 2, 3]$).

Una *cadena simplicial* és la combinació entera formal de n -símplexs orientats, i el conjunt de cadenes simplicials ve donat per:

$$C_n(X) = \{r_1\sigma_1 + r_2\sigma_2 + \dots + |r_i \in \mathbb{Z}, \sigma_i \in X_n\},$$

on X_n es refereix al conjunt de símplexs de dimensió n , i en aquesta combinació, el símplex d'orientació oposada a σ_i s'identifica amb $-\sigma_i$. És a dir, a nivell de cadena simplicial, $[p_0, \dots, p_{n-1}] = \text{sgn}(\rho)[p_{\rho(0), \dots, \rho(n-1)}]$, on ρ és una permutació.

A continuació podem definir l'aplicació de les cadenes de dimensió n , $C_n(X)$, a les cadenes de dimensió $n - 1$, $C_{n-1}(X)$, que es correspon amb el concepte intuïtiu de vora. Definim l'*aplicació vora* com

$$\begin{aligned} \partial_n: C_n(X) &\rightarrow C_{n-1}(X) \\ [v_1, \dots, v_n] &\mapsto \sum_{i=0}^n (-1)^i [v_0, \dots, \hat{v}_i, \dots, v_n]. \end{aligned}$$

Observem que se satisfà $\partial_n \circ \partial_{n+1} = 0$ per a tota n ; en podem veure un exemple a la figura 2.

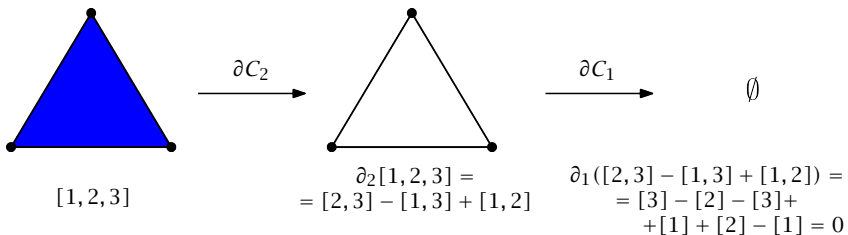


FIGURA 2: Exemple d'ús de l'aplicació vora sobre el 2-símplex $[1, 2, 3]$.

2.2 L'homologia simplicial i els nombres de Betti

L'homologia ens proporciona invariants topològics que permeten diferenciar un espai topològic en funció de les seves característiques *de forma*. De manera grollera, podem dir que l'homologia compta el nombre de forats per a cada dimensió. Per exemple, les components connexes d'un espai topològic ens determinarien l'homologia de dimensió 0, els cicles serien els elements de l'homologia de dimensió 1, les cavitats correspondrien a l'homologia de dimensió 2, i així successivament.

Donat un complex, l'homologia simplicial formalitza la idea de comptar el nombre de forats de dimensió n .

Un complex simplicial X induïx un complex de cadenes simplicials, $C_n(X)$ (per simplificar-ho usarem només la notació C_n),

$$\dots \xrightarrow{\partial_{n+2}} C_{n+1} \xrightarrow{\partial_{n+1}} C_n \xrightarrow{\partial_n} C_{n-1} \xrightarrow{\partial_{n-1}} \dots$$

que s'acostuma a escriure com a $(C_\bullet, \partial_\bullet)$. Les cadenes que tenen vora 0 s'anomenen *cicles* i formen un subespai, Z_n , de C_n , mentre que el subespai B_n de C_n està format per les cadenes que pertanyen a la imatge de ∂_{n+1} . Donat que $\partial_n \circ \partial_{n+1} = 0$, obtenim la inclusió següent:

$$B_n \subseteq Z_n \subseteq C_n.$$

A la figura 3 es mostren les relacions de subgrup entre els grups de vores, cicles i cadenes entre dimensions, mitjançant l'aplicació vora.

La n -homologia d'aquest complex es defineix com el quocient de dos \mathbb{Z} -mòduls (és a dir, grups abelians), el nucli de l'aplicació ∂_n i la imatge de l'aplicació vora en una dimensió superior, ∂_{n+1} ,

$$H_n(X) = \ker \partial_n / \text{im } \partial_{n+1} = Z_n / B_n.$$

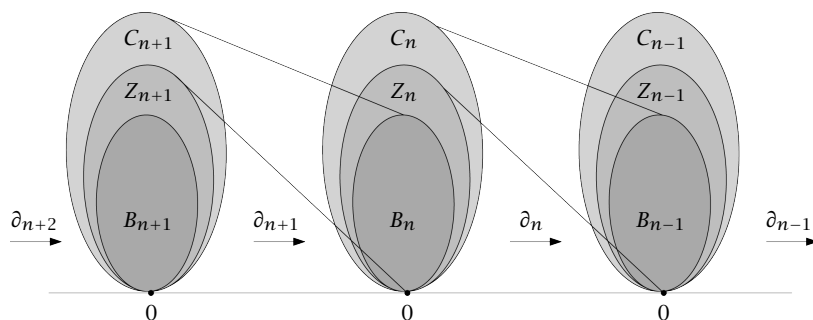


FIGURA 3: Complex de cadenes consistent en una successió lineal de grups de cadenes, cicles i vores connectats mitjançant homomorfismes.

Per a calcular l'homologia d'un espai topològic, usarem els complexos simplicials, fet pel qual rep el nom de *homologia simplicial*.

Els *nombres de Betti*, β_n , corresponen a les dimensions de l'homologia, és a dir, al rang¹ dels grups $H_n(X)$. Per tant, $\beta_n = \text{rang } Z_n - \text{rang } B_n$.

En altres paraules, el k -èsim nombre de Betti, β_k , es correspon amb el nombre de forats k -dimensionals en un espai topològic. Per exemple,

- β_0 és el nombre de components connexes.
- β_1 és el nombre de forats circulars (o 1-dimensionals).
- β_2 és el nombre de cavitats o forats 2-dimensionals.

Podem veure un exemple senzill d'un tor a la figura 4.

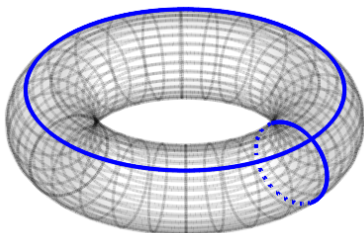


FIGURA 4: Exemple gràfic dels nombres de Betti corresponents a un tor: $\beta_0 = 1$ perquè tenim una única component connexa; $\beta_1 = 2$ perquè tenim dos cicles 1-dimensionals ressaltats a la figura; $\beta_2 = 1$ a causa de la cavitat interna que defineix la superfície del tor; finalment, $\beta_i = 0$ per a tota $i > 2$.

3 Homologia persistent i filtracions

Abans de passar a definir l'homologia persistent, ens cal introduir el concepte de *filtració*, el qual ens permet definir la idea de *persistència*.

3.1 Filtracions

Donat un complex simplicial X , una *filtració simplicial* és una col·lecció jeràrquica de complexos simplicials X^i de la forma:

$$\emptyset = X^0 \subseteq X^1 \subseteq \dots \subseteq X^m = X.$$

D'altra banda, també es poden definir filtracions sobre espais topològics (en particular, sobre varietats) a través de funcions. Donada una varietat \mathbb{M} i una funció, $f: \mathbb{M} \rightarrow \mathbb{R}$, definim la família de subespais (subvarietats) $\mathbb{M}^t = f^{-1}((-\infty, t])$, $t \in \mathbb{R}$, que compleixen $\mathbb{M}^t \subseteq \mathbb{M}^s$, si $t < s$. Aquesta inclusió ens defineix una filtració i la funció f és l'anomenada *funció de filtració*. En aquest cas, l'homologia persistent ens dona informació sobre la forma de les subvarietats.

¹ Ens referim a rang de $H_n(X)$ com el nombre màxim de generadors linealment independents; és a dir, el nombre de generadors de la part lliure de torsió de $H_n(X)$.

3.2 Homologia persistent

Un cop definides l'homologia simplicial i les filtracions, podem descriure l'homologia persistent, la qual és una versió parametritzada dels nombres de Betti, però amb l'avantatge que podem obtenir una representació gràfica, anomenada *diagrama persistent* (o *de persistència*), o bé una representació equivalent anomenada *codi de barres* (*barcode*); ambdues representacions contenen molta més informació que els nombres de Betti (vegeu la subsecció 3.4). De fet, les representacions gràfiques de l'homologia persistent en codifiquen fidedignament tota la informació i això ens permet, a més, tenir la informació homològica des del punt de vista quantitatiu.

Donada una filtració $f: X \rightarrow \mathbb{R}$ sobre un espai topològic X , $X^0 \subseteq X^1 \subseteq \dots \subseteq X^m$, es pot veure que la inclusió d'aquests espais topològics indueix un morfisme de grups de les respectives homologies $X^j \subset X^{j+1}$; és a dir, la inclusió $X^j \hookrightarrow X^{j+1}$ indueix el morfisme $H_k(X^j) \rightarrow H_k(X^{j+1})$.

L' i -èsim complex simplicial X^i en una filtració dona lloc al seu propi complex de cadenes (C_n^i, ∂_n^i) i als n -èsims mòduls de cadenes, cicles, vores i homologia se'ls denota amb C_n^i , Z_n^i , B_n^i i H_n^i , respectivament. Observem que l'ús de superíndexs serveix per a indicar el pas a la filtració, però no té a veure amb la cohomologia.

D'aquesta manera, podem definir la n -èsima homologia p -persistent de X^i , per a $p \in \mathbb{N}$, com

$$H_n^{i,p}(X) = Z_n^i / (B_n^{i+p} \cap Z_n^i).$$

$H_n^{i,p}(X)$ caracteritza els forats de dimensió n a X^{i+p} creats pel subcomplex X^i . Aquests forats existeixen per a tots els complexos X^j a la filtració amb índex $i \leq j \leq i + p$. Els elements no nuls de $H_n^{i,p}(X)$ són classes de n -cicles en el complex X^i que no són vora de cap $(n + 1)$ -cadena del complex més gran X^{i+p} .

Diem que una classe homològica $\alpha \in H_n(X^i)$ neix al pas i si $\alpha \notin \text{Im}(H_n(X^{i-1}) \rightarrow H_n(X^i))$. Anàlogament, diem que $\alpha \in H_n(X^j)$ mor al pas j si $\alpha \neq 0$ però la seva imatge sota l'aplicació $H_n(X^j) \rightarrow H_n(X^{j+1})$ és zero. Si α neix al pas i i mor al pas j , definim la seva *persistència* com el número $p = j - i$. Cal notar que $H_n^{i,p}(X)$ representa les classes d'homologia que han nascut abans del pas i i que moren després del pas $i + p$. Per això es diu que tenen persistència p .

Volem remarcar que, tot i que per ara hem parlat de la \mathbb{Z} -homologia, a la pràctica es treballa sempre sobre un cos \mathbb{F} , habitualment $\mathbb{Z}/2\mathbb{Z}$. Treballar sobre un cos és computacionalment més senzill (no ens hem de preocupar de la torsió) i, a més, ens assegura que el diagrama de persistència conté la mateixa informació que la continguda en $\{H_n^{i,p}(X)\}_{i,p,n}$; vegeu [99] per a més detalls sobre aquesta correspondència.

3.2.1 Homologia d -persistent De la mateixa manera que triem el paràmetre o la mètrica per a fer la filtració, podem triar dos paràmetres diferents i obtenir dues filtracions diferents. Això va donar lloc a l'estudi de l'*homologia d -persistent*, també coneguda com a *multiparamètrica* [49]. Per al càlcul de

l'homologia d -persistent, ens cal tenir filtracions indexades per més d'un paràmetre; per exemple, donat un conjunt de dades $X \subset \mathbb{R}^d$ ens podem interessar, posem per cas, per una banda en les regions més denses [14] i, per l'altra, en aquelles on la curvatura és més gran [17]. Això ens permet definir de manera natural els complexos \mathbb{Z}^n -filtrats, $\{K_u\}_{u \in \mathbb{Z}^n}$, $u = (u_1, \dots, u_n)$, on $K_u \subset K_v$ quan $u \leq v$; és a dir, $(u_1 \leq v_1, \dots, u_n \leq v_n)$. En aquest punt, els càlculs es compliquen més, i no entrarem en els detalls, però van obrir les portes a treballs com els de Carlsson i Knudson [16, 64] i a d'altres més recents [55, 87, 91].

3.3 Filtracions associades a un conjunt de dades

Un cop definida l'homologia persistent des del punt de vista teòric, cal que fem un petit resum de com es genera una filtració de complexos simplicials a partir d'un conjunt de dades.

Donat un conjunt de punts en un espai mètric, obtenim una filtració usant *filtracions via complexos de Čech o Rips-Vietoris* [26, 53, 57]; aquesta última és la filtració més comuna, per exemple, quan es treballa amb dades biomèdiques o d'imatge (vegeu la figura 5).

A continuació, donem una pinzellada de com funcionarien aquestes filtracions per als complexos més coneguts.

Filtració de Čech Donat un conjunt de punts en un espai mètric, afegim una bola de radi r a cada punt. Per definir els símplexs de cada pas, considerem la intersecció mútua entre boles i, aleshores, definim un $(k - 1)$ -símplex quan la intersecció de k boles de radi r és no buida. A mesura que anem incrementant el radi r tindrem més símplexs de grau superior fins que tots estiguin interconnectats. El radi r de les boles és el paràmetre de la filtració; en podeu veure un petit exemple a la figura 5.

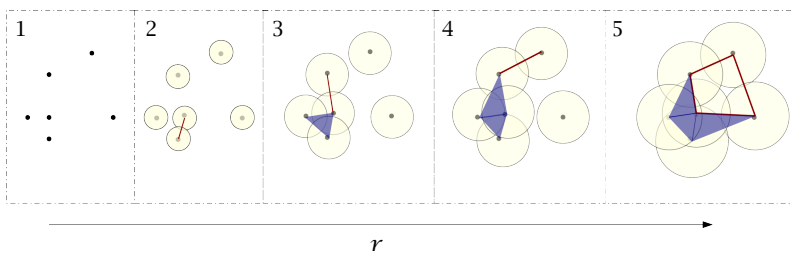


FIGURA 5: Filtració mitjançant el complex de Čech. Els passos a la filtració són determinats per la *distància llindar* entre punts, que seria el radi de la bola considerada a l'entorn de cada punt.

La *filtració de Rips-Vietoris* és definida d'una manera similar a la filtració via els complexos de Čech. La diferència rau en el fet que, en aquest cas, considerem que tenim un $(k - 1)$ -símplex si hi ha k boles de radi r que tenen intersecció no buida dues a dues. Aquest fet redueix el cost computacional ja que només ens cal saber les distàncies entre punts dos a dos.

De la mateixa manera, a partir d'un graf amb pesos es pot generar una filtració assumint que els pesos de cada aresta representen la distància entre nodes; és a dir, la funció de filtració seria la funció de pesos a les arestes. Aquests pesos seran modificats o transformats segons ens convingui per tal de generar la filtració volguda, i en el sentit que ens convingui més, ja que poden ser correlacions o altres mesures que ens interessi tenir en compte per a analitzar les dades. En podem veure un exemple a la figura 6. Observem que fer la filtració de Rips-Vietoris en un graf seria equivalent a considerar les cliques del graf en funció dels seus pesos. Això ens determinaria la mateixa filtració. Cal remarcar que, en aquest cas, estem usant la definició de complex simplicial abstracte ja que els nodes no tenen una realització a \mathcal{R}^2 .

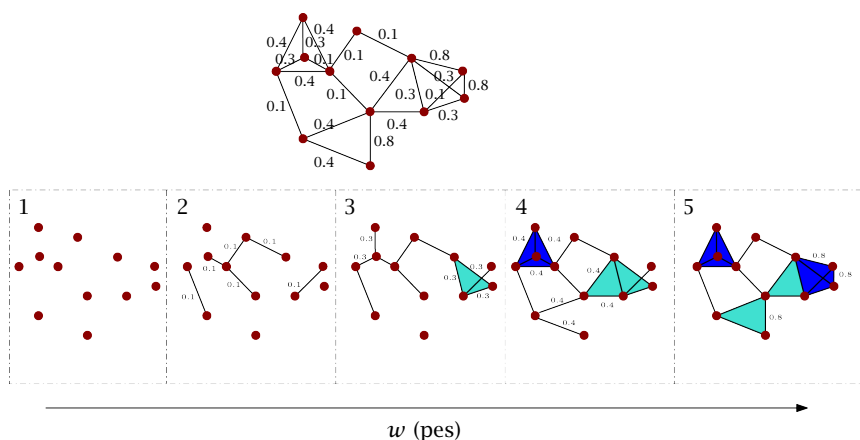


FIGURA 6: Filtració via un complex de cliques donada una xarxa amb pesos. Els passos a la filtració són determinats pels pesos a les arestes del graf donat.

Tot i que les filtracions a partir dels complexos de Čech i Rips-Vietoris són les més naturals, a mesura que el nombre de punts d'entrada augmenta, la complexitat en el càlcul d'aquests complexos es dispara. Normalment es prefereix usar el complex de Rips-Vietoris, en comptes del de Čech, ja que el seu càlcul té un cost computacional més baix.²

Ara bé, hi ha maneres més eficients de generar un complex simplicial, com ara via els complexos α i els *complexos testimoni* (*witness complexes*) [34]. També, Zomorodian ([98]) va introduir els conjunts *endreçats*, que consisteix a crear un conjunt simplicial minimal capaç de capturar la topologia del complex simplicial original. Una altra aproximació dispersa (*sparse*), que redueix la complexitat d'una filtració d'un complex simplicial en un espai mètric n -dimensional, és la proposada per Sheehy [88] i Dey *et al.* [37].

² Cal notar que, fixat un radi, la cardinalitat del complex de Rips-Vietoris és major que la de Čech, però, tot i això, usar el complex de Rips-Vietoris és computacionalment més econòmic ja que només cal calcular les interseccions dos a dos.

3.4 Invariants topològics: *barcodes* i *diagrames de persistència*

El càlcul de l'homologia persistent sobre una filtració proporciona invariants topològics, multiconjunts de punts a \mathbb{R}^2 , que poden ser representats segons el que ens convingui més. Parlarem dels diagrames de persistència i dels codis de barres (*barcodes*), unes representacions completament equivalents.

L'homologia persistent, tal com està definida, se centra en els cicles de dimensió n que viuen al llarg dels intervals definits per la filtració i assigna un valor de rellevància proporcional a la llargada d'aquests intervals. És a dir, el rang de la imatge de l'aplicació $H_n(X^i) \rightarrow H_n(X^j)$ és el nombre de classes d'homologia de dimensió n independents que neixen al complex X^i o abans d'aquest, i encara *viuen* (existeixen) al complex X^j . Això inclou les classes anomenades *essencials*, que són aquelles que no moren al final de la filtració. D'aquesta manera, podem definir tant el codi de barres [50] com el conjunt d'intervals que representa cadascuna de les classes d'homologia persistent a través d'un interval $(i, i + p)$, on i indica el pas de la filtració en què aquest generador ha aparegut per primer cop (diem que hi ha *nascut*), i $i + p$, el pas en què ha deixat d'existir (diem que hi ha *mort*); p indicaria la seva persistència en termes absoluts al llarg de la filtració. Per a més detalls, es pot consultar l'article de Zomorodian i Carlsson [99], en el qual es defineix de manera més precisa què representen aquests intervals (o *barres*), que correspondrien a les classes d'homologia graduada [61]. Gràficament, tal com es pot veure a la figura 7, aquests intervals es representen com a barres horitzontals amb coordenades a l'eix horitzontal $(i, i + p)$ i altura lliure; és a dir, a l'eix vertical es representen, en ordre arbitrari, els generadors homològics.

Una altra manera de representar la mateixa informació és amb un *diagrama de persistència* (PD) (ens referirem a $PD(X)$ com el diagrama de persistència del complex simplicial X), que consisteix en un multiconjunt de parelles d'índexs referents a la filtració que representen el *naixement* i la *mort* per a cadascuna de les classes d'homologia persistent representades al pla xy (on l'eix horitzontal acostuma a representar el punt de naixement, mentre que l'eix vertical representa el punt de mort, tal com es pot veure a la figura 8).

Notem que, normalment, si una classe essencial neix a X^i , aleshores es representa amb el punt (i, ∞) en el diagrama de persistència o en el codi de barres.

Cal remarcar, també, que, fixat un pas i a la filtració, si fem un tall en aquell pas tant en un PD com en un codi de barres, el nombre de punts o intervals que talleu és igual al corresponent nombre de Betti al pas i de la filtració. D'aquesta manera, podem dir que el paràmetre de filtració ens proporciona una versió parametritzada dels nombres de Betti.

Tant els diagrames de persistència com els codis de barres es poden definir per a cada dimensió d'homologia que vulguem estudiar. És a dir, podem tenir un PD per als grups d'homologia de dimensió 0 (l'evolució de les components connexes), de dimensió 1 (l'evolució dels cicles), de dimensió 2 (l'evolució de les cavitats), i, anàlogament, per a dimensions superiors.

Observem que els PD i els codis de barres són equivalents: només cal convertir l'interval (i, j) dels codis de barres al corresponent punt en \mathbb{R}^2 amb

coordenades (i, j) . Podeu veure un exemple tant de codi de barres com de diagrama de persistència a les figures 7 i 8, respectivament.

Intuïtivament, podem dir que aquells n -cicles que *viuen o sobreviuen* durant més temps tenen més significació i caracteritzen millor l'espai topològic inicial que aquells que *viuen* durant intervals de temps curts, que, sovint, només representen soroll a les dades.

Codis de barres

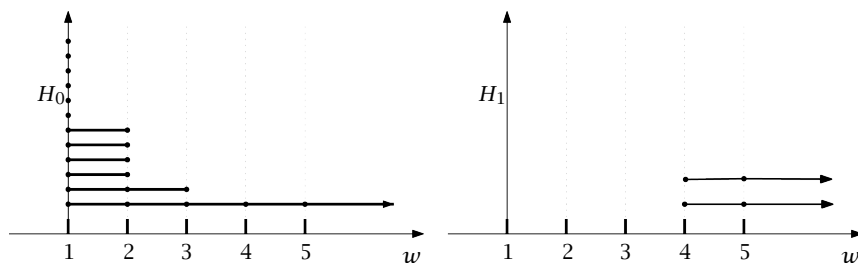


FIGURA 7: Codis de barres per a H_0 i H_1 produïts a partir de la filtració de la figura 6. L'eix horitzontal indica el pas a la filtració, que, en aquest cas, ve donat pels pesos a les arestes del graf inicial.

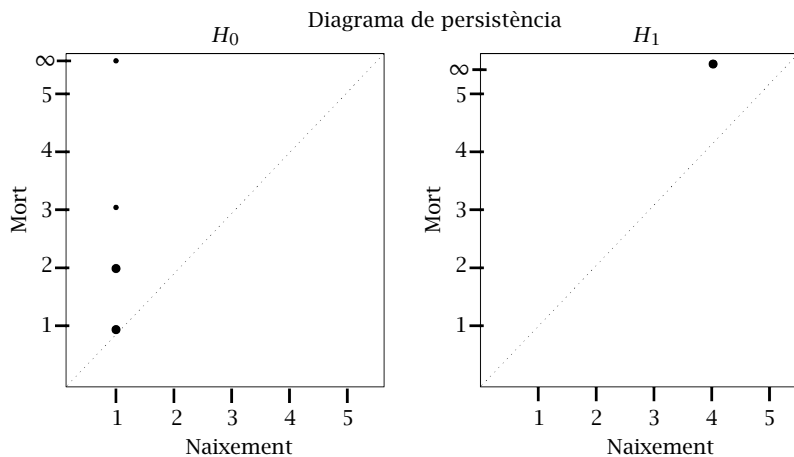


FIGURA 8: Codis de barres per a H_0 i H_1 produïts a partir de la filtració de la figura 6. L'eix horitzontal indica el pas a la filtració en què s'ha creat l'invariant topològic, mentre que l'eix vertical indica quan deixa d'existir. Els punts més grans intenten representar la multiplicitat en aquell punt, ja que, per exemple, per a H_0 , hi ha moltes components connexes que *neixen i moren* al mateix moment. En aquest cas, l'infinit representa que perdura durant tot el temps de la filtració.

OBSERVACIÓ. Un cop determinada l'homologia persistent, seria molt interessant trobar els generadors del grup n -homològic. És possible trobar un representant de cada classe d'homologia però el problema rau en identificar el generador a les dades originals. En altres paraules, seria molt interessant,

un cop trobat el grup d'homologia de dimensió n , ser capaços de dir on es troba aquell cicle de dimensió n a les dades, però només podem optar per usar un representant del grup d'homologia. Una solució seria trobar tota la classe d'homologia, però a nivell computacional no seria factible per poc que el conjunt de dades fos una mica gran. Respecte aquesta qüestió, hi ha treballs on es troben i s'usen els representants dels generadors del grup d'homologia per a localitzar les característiques topològiques principals (cicles), en particular en l'àmbit de la neurociència [83].

3.5 Comparació entre diagrames de persistència i estabilitat

Ara bé, un cop sabem com podem representar l'homologia persistent mitjançant els invariants topològics, una pregunta natural és: com els podem comparar? Aquesta pregunta ha rebut molta atenció en la darrera dècada atès que, així com els diagrames de persistència són els intermediaris (*proxies*) topològics d'un conjunt de dades, tenim que les distàncies entre diagrames serveixen com a intermediari de similaritat entre conjunts de dades. A més, aquesta informació topològica pot ser usada com a intermediari per a algorismes d'aprenentatge automàtic [75] i classificació [15]. A la figura 9, podem trobar un exemple que il·lustra la utilitat de la comparació de diagrames de persistència.

I això ens porta a un problema d'aparellament, és a dir, com podem comparar dos multiconjunts de punts. Per a fer-ho, es defineixen les distàncies següents entre dos diagrames de persistència, una de les quals s'esmentarà a la prova d'estabilitat dels diagrames de persistència (vegeu la subsecció 3.6).

Donats dos diagrames de persistència $PD(X)$ i $PD(Y)$, la *distància q de Wasserstein* (q -Wasserstein d'ara endavant) és definida per [8, 30]:

$$d_q(PD(X), PD(Y)) = \inf_{\nu: PD(X) \rightarrow PD(Y)} \left[\sum_{x \in PD(X)} \|x - \nu(x)\|_q^q \right]^{1/q},$$

on ν varia entre totes les bijeccions.³ La *distància de coll d'ampolla* (*bottleneck distance*), d_∞ o també anomenada d_B , és definida per [29, 45]

$$d_\infty(PD(X), PD(Y)) = \inf_{\nu: PD(X) \rightarrow PD(Y)} \sup_{x \in PD(X)} \|x - \nu(x)\|_\infty.$$

Observem que la distància de coll d'ampolla és un cas especial de la distància q -Wasserstein, (d_q , on $q \rightarrow \infty$). Tot i que siguin distàncies ben definides, tenim el problema del seu càlcul real atès que ens trobem davant d'un problema d'aparellament, on hem de tenir en compte un nombre no trivial de bijeccions possibles entre $PD(X)$ i $PD(Y)$. Aquest càlcul, malgrat que els algorismes d'aparellament acostumen a reduir l'espai de cerca, sol ser feixuc a mesura que creix la dimensió del multiconjunt (és a dir, el nombre de punts que forma el diagrama de persistència). Al treball de Kerber, Morozov *et al.* podem trobar alguns exemples d'implementacions [63].

³ Si els cardinals de $PD(X)$ i de $PD(Y)$ són diferents, s'igualen afegint punts al diagrama amb cardinal menor.

D'altra banda, tot i ser distàncies ben definides (el conjunt de codis de barres genera de forma natural la distància de Wasserstein), desafortunadament, l'espai resultant no és complet (és només un espai mètric), fet pel qual no és adient per a fer inferència estadística (no tenim una mitjana única) [73]. Per això, s'han introduït diferents transformacions dels diagrames de persistència o dels codis de barres per a poder usar-los com a característiques d'entrada d'altres algoritmes, o bé, per a extreure'n informació estadística.

A continuació, presentem algunes d'aquestes transformacions i donem referències sobre d'altres que han anat derivant de les propostes inicials.

3.5.1 Distància via nuclis Per una banda, tenim mètodes que ens permeten treballar en espais de Hilbert via *funcions de nucli* [84, 18, 66, 24, 65]. Això és molt útil per a poder aplicar algoritmes com ara màquines de vector de suport (*support vector machines*, SVM) o anàlisi de components principals (*principal component analysis*, PCA) directament.

Uns dels primers a proposar un nucli estable com a solució per a treballar en un espai de Hilbert van ser Reininghaus *et al.* [84].

Recordem que, donat un conjunt X , una funció $k: X \times X \rightarrow \mathbb{R}$ és un *nucli* (*kernel*) si existeix un espai de Hilbert \mathcal{H} i una aplicació $\phi: X \rightarrow \mathcal{H}$, anomenada *aplicació característica* (*feature map*), tal que

$$k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}} \quad \forall x, y \in X.$$

Equivalentment, k és un nucli si es pot representar matricialment com una matriu simètrica i definida positiva. Un dels principals avantatges dels nuclis és que permeten anar a espais de grans dimensions, \mathcal{H} , a dividir les dades (i després classificar-les) sense conèixer explícitament ϕ . A més, els nuclis permeten passar d'espais *estranys* (grafs, diagrames de persistència, cadenes...) a espais de Hilbert on podem treballar molt més fàcilment i dividir les dades i classificar-les. Més en concret, tenim el nucli proposat per Reininghaus *et al.* el 2015, tal que, donats dos conjunts de punts corresponents a dos diagrames de persistència $PD(X)$, $PD(Y)$, definim:

$$K_{\sigma}(PD(X), PD(Y)) = \frac{1}{8\pi\sigma} \sum_{p \in PD(X), q \in PD(Y)} e^{-\frac{\|p-q\|^2}{8\sigma}} - e^{-\frac{\|p-\hat{q}\|^2}{8\sigma}},$$

on $\hat{q} = (q_2, q_1)$ si $q = (q_1, q_2)$, σ és el paràmetre d'escala relacionat amb la mida dels passos a la filtració.

Aquest nucli ens indica que, a mesura que K_{σ} tendeix a zero, la distància entre els diagrames de persistència va augmentant. Una idea intuïtiva de com funciona consisteix en imaginar-se una distribució gaussiana sobre cada punt del diagrama de persistència i *restar* un diagrama de l'altre, on el paràmetre σ actua com a resolució per a definir quan dos punts són iguals o no.

Altres treballs on es defineixen i es treballen diferents nuclis es troben a [18, 66, 24, 65]. De la mateixa manera que havíem parlat de l'homologia d -persistent (vegeu la subsecció 3.2.1), també hi ha el multinucli per a aquesta homologia multiparamètrica [32].

3.6 Estabilitat dels diagrames de persistència

En les seccions anteriors hem vist que l'homologia persistent descriu, mitjançant invariants topològics, les característiques multiescala d'un espai topològic. També hem vist que els diagrames de persistència, un exemple de com es pot codificar la informació de l'homologia persistent, són un multiconjunt de punts a \mathbb{R}^2 . Així mateix, hem descrit diverses formes i mètriques per a comparar una parella de diagrames de persistència. Ara bé, si pertorbem una mica l'espai inicial (ja sigui donat per un conjunt de punts a \mathbb{R}^n com per un graf amb pesos), canvia això el resultat de l'homologia persistent?

Per tal que els diagrames de persistència puguin ser ben interpretats i també usats com a característica d'un espai topològic, cal que siguin robustos a pertorbacions petites. Hem de pensar que, en l'aplicació de l'homologia persistent, les nostres dades poden tenir soroll o potser, no volem que depenguin de la mostra que agafem. Per això, els resultats sobre l'estabilitat dels diagrames de persistència són molt rellevants. La primera prova d'estabilitat va ser donada per Cohen-Steiner i els seus col·laboradors [29].

La prova (no entrarem en el detall) es basa a mostrar l'estabilitat dels diagrames de persistència d'una funció que indueix una filtració sobre un espai topològic X . A grans trets, la prova mostra que, donades dues funcions (*tame*), f i g , tals que $f, g: X \rightarrow \mathbb{R}$, on X és un espai topològic i les dues funcions són filtracions de X , aleshores, si la distància L_∞ entre elles és petita, la distància (usant la distància de coll d'ampolla d_B) entre els corresponents diagrames de persistència $\text{PD}(f, X)$ i $\text{PD}(g, X)$, també ho serà. Aquí $\text{PD}(f, X)$ i $\text{PD}(g, X)$ fan referència al diagrama de persistència obtingut usant les respectives funcions de filtració sobre l'espai topològic X . És a dir, per a cada dimensió p , tenim que el resultat principal ve donat per

$$d_B(\text{PD}(f, X), \text{PD}(g, X)) \leq \|f - g\|_\infty. \quad (1)$$

De la mateixa manera, obtenim una intuïció geomètrica d'estabilitat en diagrames de persistència gràcies al resultat següent [23], on es pot veure que, donats dos subconjunts de punts X i Y en un espai mètric (M, d_M) , aleshores els diagrames de persistència⁴ corresponents a X i Y satisfan:

$$d_B(\text{PD}(X), \text{PD}(Y)) \leq d_H(X, Y), \quad (2)$$

on d_H és la distància de Hausdorff definida sobre dos conjunts de punts X i Y ; és a dir,

$$d_H(X, Y) = \max\{\sup_{x \in X} \inf_{y \in Y} d_M(x, y), \sup_{y \in Y} \inf_{x \in X} d_M(x, y)\}.$$

Podem trobar un exemple d'aquest tipus d'estabilitat a la figura 9.

Existeixen altres treballs que mostren l'estabilitat dels diagrames de persistència, com els de Chatzal *et al.* i Frosini [26, 22, 48].

⁴ Observem que les cardinalitats de $\text{PD}(X)$, $\text{PD}(Y)$ són iguals considerant el conjunt de punts de la diagonal amb multiplicitat infinita.

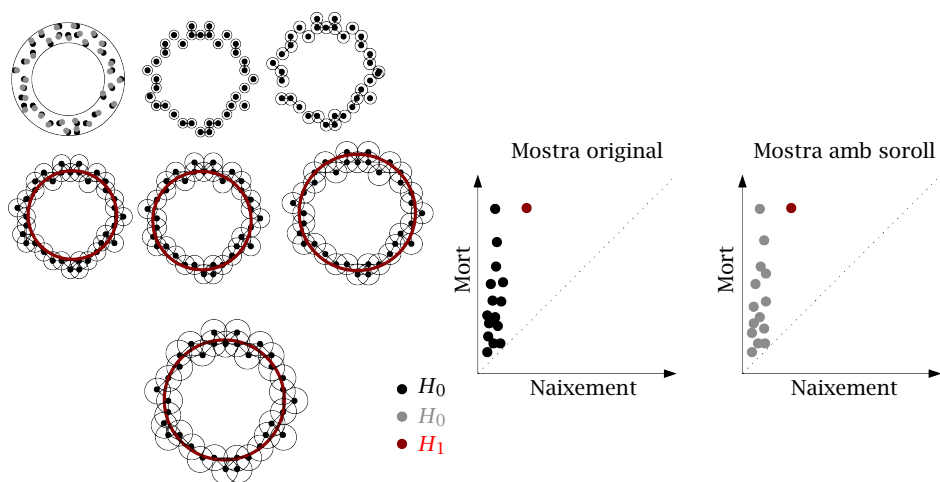


FIGURA 9: Il·lustració d'estabilitat: dades originals (punts negres), X , dades amb soroll (punts blaus), Y . Dues mostres diferents generen dos conjunts de punts diferents, (X, Y) , a partir dels quals es calcula l'homologia persistent via una filtració de Čech i obtenim els diagrames de persistència mostrats a la figura central ($PD(X)$) i dreta ($PD(Y)$). La possibilitat de poder comparar diagrames de persistència permet dir que els dos conjunts de dades representen el mateix espai topològic. En aquesta il·lustració, a més, podem veure que les formes principals (en aquest cas, un cercle) determinants del nostre espai són presents al llarg de més passos de la filtració. En aquest cas, es mostren els diagrames de persistència per a H_0 i H_1 en el mateix gràfic. Al diagrama de l'esquerra s'observa que hi ha una component connexa principal i un cicle (de dimensió 1) principal sobre el mostreig de les dades originals, X , mentre que el diagrama de persistència de la dreta correspon a un mostreig de les dades amb una petita pertorbació, Y . Tant la mostra original com la mostra amb soroll produeixen diagrames de persistència similars, és a dir, tindrien una distància entre ells baixa, o bé una similitud alta.

3.7 Altres transformacions sobre els resultats en homologia persistent

3.7.1 Paisatges persistents (*persistent landscapes*, PL) Una de les transformacions més primerenques va ser la proposada per Bubenik el 2012 [11, 10]. Els *paisatges persistents* es desenvolupen per a construir funcions lineals a trossos a partir dels diagrames de persistència (PD) per a poder-hi fer anàlisi de dades *a posteriori* (vegeu-ne un exemple a la figura 10). El paisatge persistent és el conjunt de funcions $\{\lambda_k: \mathbb{R} \rightarrow \mathbb{R}\}_{k \in \mathbb{N}}$, on definim $\lambda_k(t)$ com el k -èsim valor més gran del conjunt de números

$$\{\max(\min(t - a_i, b_i - t), 0)\}_{i=1}^n,$$

on (a_i, b_i) són punts del PD, per a $i \in \{1, \dots, n\}$. Això ens permet estendre una funció de \mathbb{R}^2 a \mathbb{R} , de manera que obtenim un espai de funcions separable

(en aquest cas, un espai de Banach), i així podem fer inferència estadística dins d'aquest espai. Definint la p -distància entre dos paisatges persistents, λ i λ' , com $\|\lambda - \lambda'\|$, es pot demostrar que és estable respecte a la norma del suprem [10].

A la figura 10 il·lustrem les transformacions que experimenta un diagrama persistent fins a esdevenir un paisatge persistent. Aquestes representacions van ser també estudiades per Chazal [25].

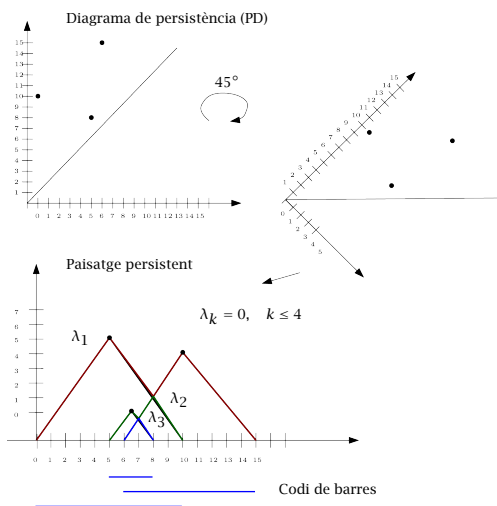


FIGURA 10: Exemple de transformació d'un diagrama de persistència (PD) en un paisatge persistent (PL). Aquesta transformació ens permet passar d'un multiconjunt a un conjunt de funcions lineals a trossos on es pot treballar estadísticament.

3.7.2 Vectoritzacions Consisteixen a vectoritzar directament el diagrama de persistència en un espai euclidià, tal com es fa, per exemple, amb les *imatges persistents* (*persistent images*, PI) [1]. En aquest cas, la vectorització es duu a terme de la forma següent: es construeix una representació en \mathbb{R}^n integrant la densitat estimada d'un diagrama de persistència en una superfície. En altres paraules, creen l'anomenada *superfície persistent* a partir dels diagrames de persistència prenent la suma ponderada de gaussians centrades a cada punt del diagrama de persistència. Es creen uns vectors, anomenats *imatges persistents*, integrant la superfície persistent sobre una graella. Això permet aplicar les tècniques d'aprenentatge automàtic per a espais vectorials finits. Es demostra que les imatges persistents són estables respecte a la distància 1-Wasserstein entre diagrames de persistència [1], i que es poden concatenar diagrames de persistència de dimensions diverses en un únic vector, de manera que permet tenir en compte diferents dimensions d'homologia de cop. Les imatges persistents s'han analitzat des del punt de vista estadístic en un treball de Chen *et al.* [28].

Troblem altres exemples de vectoritzacions a [3], on es modelen els diagrames de persistència com una funció de probabilitat bidimensional a l'espai vectorial tangent d'una varietat riemanniana, així com una proposta per a ser usada en xarxes neurals profundes [59] o una transformació dels diagrames de persistència en polinomis complexos [40], fet que permet treballar directament a l'espai de funcions dels polinomis.

3.7.3 Vinyars D'altra banda, Cohen-Steiner i els seus col·laboradors ([31]) van definir el concepte de *vinyar* (*vineyard*), que, en poques paraules, són diagrames de persistència que varien en el temps. Atès que podem tenir problemes allà on el nostre espai topològic i el complex simplicial variïn en el temps, i potser volem saber com varien, es van definir els vinyars, que consisteixen a seguir les trajectòries dels punts d'un diagrama de persistència inicial al llarg del temps.

Donat un conjunt de punts dinàmic a l'espai $\mathbb{X}(t) = \{x_1(t), \dots, x_n(t)\}$, obtenim el corresponent conjunt de diagrames de persistència $PD(\mathbb{X}(t))$ per a cada temps t . Aquest paràmetre *temps* és l'anomenat *vinyar*; cada punt $p \in PD(\mathbb{X}(t))$ per sobre de la diagonal es mou al llarg del temps traçant una corba, que s'anomena *vinya* (*vine*).

3.8 Programari per al càlcul de l'homologia persistent

De la mateixa manera que hem mencionat les dificultats de generar un complex simplicial a gran escala i la filtració corresponent, un altre àmbit en el qual s'ha treballat intensament els darrers anys és el del càlcul de l'homologia persistent des del punt de vista pràctic. L'algoritme clàssic del càlcul de l'homologia persistent es basa en la reducció de la matriu de l'operador vora [44]. Podem trobar propostes de millora en l'algoritme en els treballs de Milosavljević [74], Chen [27], Lewis [70] i Boissonnat [9], per posar-ne uns quants exemples; a més dels treballs de Bauer i els seus col·laboradors [4, 6, 5], que proposen un algoritme paral·lelitzable en què divideixen el complex simplicial sobre el qual es calcula l'homologia persistent. A [79] podem trobar una anàlisi prou recent i més completa dels diferents algorismes i programari que existeixen.

Un dels programes més coneguts i usats és l'anomenat *Ripser* [94], un dels més antics és *Eirene* [58] i un dels més actuals és *Giotto* [92]. Segurament, ens n'estem deixant uns quants més, ja que en els últims cinc anys hi ha hagut molts avenços pel que fa a la implementació de l'algoritme per al càlcul de l'homologia persistent.

D'altra banda, per a l'homologia d -persistent, el 2015 va aparèixer *RIVET* [93] per al càlcul i la visualització dels diagrames de persistència bidimensionals (vegeu la subsecció 3.4), gràcies al treball de Lesnick i Wright [69].

4 L'algoritme Mapper

D'altra banda, trobem un altre tipus d'algoritme molt diferent dels que calculen l'homologia persistent que té en compte la topologia de les dades, però d'una altra manera que expliquem a continuació. Es tracta de l'algoritme anomenat

Mapper, que va ser introduït per Singh, Mémoli i Carlsson el 2007 [89] i popularitzat i implementat els cinc anys següents.⁵ A grans trets, donada una funció de clusterització, l'algoritme genera un graf esquelet on es codifica la interacció entre els clústers detectats.

Els complexos de Čech són un cas particular de famílies de complexos associades a recobriments oberts en el cas d'espais topològics. Donat un recobriment, $\mathcal{U} = (U_i)_{i \in I}$ d'un espai topològic \mathbb{M} , és a dir, una família de conjunts U_i tal que $\mathbb{M} = \cup_{i \in I} U_i$, el *nervi* de \mathcal{U} és el complex simplicial abstracte $C(\mathcal{U})$ que té per vèrtexs els conjunts U_i i tal que

$$\sigma = [U_{i_0}, \dots, U_{i_k}] \text{ és un símplex de } C(\mathcal{U}) \text{ si i només si } \bigcap_{j=0}^k U_{i_j} \neq \emptyset.$$

Donat un recobriment, cada element del recobriment pot representar un clúster local o un grup de punts que comparteixen alguna característica comuna. El nervi ens dona una descripció compacta i global de la relació entre aquests conjunts U_i i els seus patrons d'intersecció; en podem veure un exemple a la figura 11. Usar els recobriments de nervi com una manera de compactar informació i visualitzar-la va ser la idea inicial de l'algoritme *Mapper*.

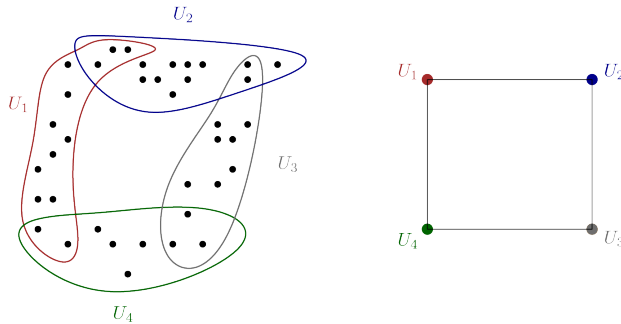


FIGURA 11: Nervi d'un recobriment sobre un conjunt de punts en el pla.

Sigui $f: X \rightarrow \mathbb{R}^d$, $d \geq 1$, una funció contínua amb valors reals, i sigui $\mathcal{U} = (U_i)_{i \in I}$ un recobriment de \mathbb{R}^d . El recobriment a partir del *pull back* de X induït per (f, \mathcal{U}) és la col·lecció de conjunts oberts $(f^{-1}(U_i))_{i \in I}$. El *pull back* refinat és la col·lecció de components connexes dels conjunts oberts $f^{-1}(U_i)$, $i \in I$.

La idea de l'algoritme *Mapper* és, donat un conjunt de dades \mathbb{X} i una funció *ben triada* sobre els valors reals $f: \mathbb{X} \rightarrow \mathbb{R}^d$, resumir \mathbb{X} a través del nervi del *pull back* refinat d'un recobriment de \mathcal{U} de $f(\mathbb{X})$. Per a recobriments ben triats, el nervi correspon a un graf, una manera fàcil i útil de visualitzar les dades resumidament. Vegeu-ne una il·lustració a la figura 12.

⁵ En aquest mateix període, els autors de l'algoritme van crear una empresa per a explotar-lo a nivell comercial.

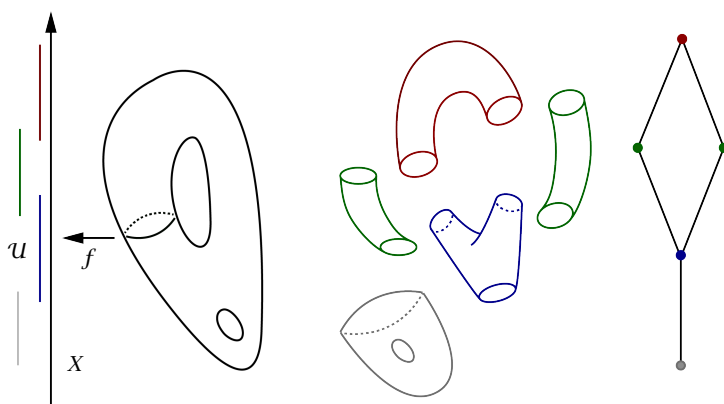


FIGURA 12: Recobriment refinat del *pull back* sobre la funció altura a una superfície a \mathbb{R}^3 i el nervi corresponent en forma de graf.

D'aquesta manera, la part més important és escollir la funció f , que serà la nostra funció de filtratge o clusterització i que triarem en funció de l'aspecte de les dades \mathbb{X} que vulguem inspeccionar. D'altra banda, cal escollir bé el recobriment \mathcal{U} de $f(\mathbb{X})$ i, després, fer una bona clusterització de les dades $f^{-1}(U)$ en clústers, C_{u,i_U} , a partir dels quals calcularem el nervi. Normalment s'usa un paràmetre de resolució per a determinar quan s'intersequen dos elements del recobriment.

Podem trobar anàlisis d'estabilitat del *Mapper* en els treballs [19, 20], mentre que la inestabilitat ha estat estudiada per Dey *et al.* [38, 39]. De la mateixa manera que trobem programari o llibreries en Python per a usar de manera pràctica l'homologia persistent, també existeixen implementacions del *Mapper* en Python [95]. Remarquem, també, que l'algoritme *Mapper* va esdevenir una companyia i l'algoritme va ser semiprivat durant un temps.

5 Aplicacions i evolució de la TDA versus el ML

Al llarg de les seccions precedents, d'una banda, hem introduït l'homologia persistent, les seves propietats i la manera de codificar-ne la informació resultant mitjançant els diagrames de persistència o els diagrames de barres; de l'altra, hem vist que l'espai dels diagrames de persistència no sempre és l'adient si volem usar-ne la informació com a paràmetres d'entrada, per exemple, en algorismes de classificació d'aprenentatge automàtic. A la subsecció 3.5 hem vist que existeix un conjunt de transformacions que ens permeten passar d'un espai amb *mal comportament* a d'altres amb els quals podem *fer estadística* (vegeu la figura 13). A continuació, es mostren diversos exemples, en àmbits molt variats, en què s'ha aplicat l'homologia persistent com una eina per a extreure informació d'un conjunt de dades, normalment sorolloses.

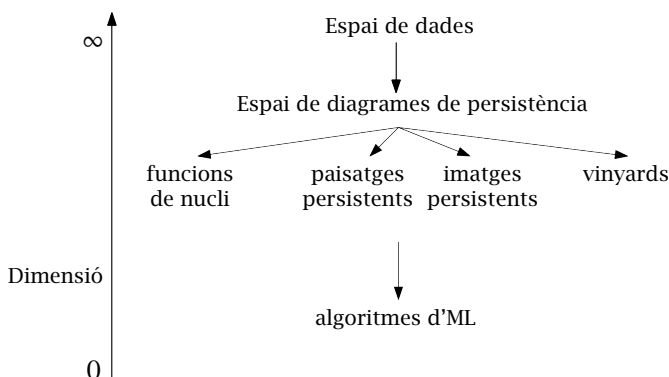


FIGURA 13: Esquema de la possible evolució d'una anàlisi de dades via l'homologia persistent. Les dades inicials acostumen a viure en un espai de dimensió elevada. L'homologia persistent permet comprimir la informació de les dades a \mathbb{R}^2 i, mitjançant algunes transformacions, es poden usar els resultats de l'homologia persistent com a dades d'entrada d'algoritmes d'aprenentatge automàtic.

Una de les primeres aplicacions de l'homologia persistent com una eina per a classificar imatges la trobem a [71]. En aquest cas, s'arriba a una millora de la classificació d'imatges gràcies a la combinació de l'ús dels diagrames de persistència i la tècnica de *BoF* (*bag-of-features*) per a la classificació de dades.

D'altra banda, Yoo *et al.* ([97]) van utilitzar els vinyards per a determinar propietats dinàmiques temporals del cervell sense imposar cap líndar a les dades i d'una manera robusta. Més concretament, determinen tasques específiques de connectivitats funcionals a partir de dades d'electroencefalogrames; comparen els resultats obtinguts amb altres metodologies, com la PCA, i altres mesures de xarxes. Els vinyards són la tècnica més robusta, sense que calgui imposar un líndar temporal a les xarxes d'entrada.

En trobem altres aplicacions en el camp de l'anàlisi de les sèries temporals [82], ja sigui en l'àmbit de la genètica [81, 21] o utilitzant els paisatges persistents per a analitzar sèries temporals en economia [51]. També n'existeixen en l'àmbit de la ciència dels materials [68]. En el camp de les imatges mèdiques, s'han tractat problemes de classificació emprant o bé la distància de coll d'ampolla [2] o bé imatges persistents [1], i s'ha pogut estudiar l'arbre que formen les artèries cerebrals [7].

La neurociència —i, més concretament, la classificació del tipus de xarxes (o grafs)— és un dels camps on més s'ha anat introduint l'ús de la topologia persistent [33, 90]. Amb l'ús dels nuclis, trobem aplicacions tant en la classificació d'imatges [84] (on es millora la taxa de classificació en comptes d'usar els paisatges persistents), com en la neurociència [60].

Entre el 2015 i el 2020, el nombre de treballs en els quals s'usa l'homologia persistent com una eina per a l'ús i la classificació de dades ha augmentat d'una

manera significativa i, a banda de les referències citades en aquest article, se'n poden trobar moltes més.

Així mateix l'homologia persistent ha contribuït a millorar o entendre més bé les xarxes neuronals convolucionals (*convolutional neural networks*, CNN). Per exemple, els treballs de Guss i Salakhutdinov [54] i de Rieck *et al.* [85] proposen una mesura de complexitat per a les arquitectures de xarxes neuronals basades en TDA, i en un nou treball de Carlsson i Gabrielsson [13] es pot veure que, aplicant eines topològiques a les xarxes convolucionals profundes (*deep convolutional networks*), milloren els càlculs d'aquestes xarxes. D'altra banda, Naitzat *et al.* ([78]) revelen intuïcions sobre com es comporten les xarxes respecte a la topologia de les dades d'entrada en funció de la seva profunditat. Podeu trobar més informació sobre homologia persistent i aprenentatge profund a [76].

Pel que fa a l'algoritme *Mapper*, en trobem algunes aplicacions a [72, 96, 89] en diferents camps, des de la medicina fins a la política i els esports.

6 Conclusions i remarques

Segurament aquesta revisió es deixa moltes referències, però, si més no, esperem que a partir de les que es faciliten el lector pugui trobar aquelles que més li interessin, tant des d'un punt de vista teòric com pràctic. Així mateix, esperem que aquest resum de la TDA contribueixi a la difusió d'aquestes tècniques *recents* d'anàlisi de dades sense demanar al lector gaire coneixements previs d'àlgebra o de teoria d'homologia. La intenció ha estat escriure un treball per a tots els públics, de manera que qualsevol matemàtic de qualsevol àmbit —tant si és dins la recerca com si no— pugui seguir i entendre les idees principals de l'homologia persistent i l'algoritme *Mapper*, així com la importància que té com una eina nova que es pot emprar en camps com l'aprenentatge automàtic i la intel·ligència artificial.

Agraïments

Vull agrair a Miguel Teixidó el seu suport, els seus suggeriments i la seva dedicació com a revisor *personal*. També cal dir que aquest treball ha estat escrit per tres persones —una, jo mateixa, i dues que, al llarg de l'escriptura d'aquest manuscrit, han estat (i estan) creixent dins meu. Finalment, vull donar les gràcies a totes les persones de la Societat Catalana de Matemàtiques que any rere any treballen per fer possible aquesta publicació.

En l'àmbit institucional, vull agrair el suport d'Intesa Sanpaolo Innovation Center (Torí, Itàlia) i de la ISI Foundation (Torí, Itàlia).

Referències

- [1] ADAMS, H.; EMERSON, T.; KIRBY, M.; NEVILLE, R.; PETERSON, C.; SHIPMAN, P.; CHEPUSHTANOVA, S.; HANSON, E.; MOTTA, F.; ZIEGELMEIER, L. «Persistence

- images: a stable vector representation of persistent homology». *J. Mach. Learn. Res.*, 18 (2017), article núm. 8, 35 p.
- [2] ADCOCK, A.; RUBIN, D.; CARLSSON, G. «Classification of hepatic lesions using the matching metric». *Comput. Vision Image Understanding*, 121 (2014), 36–42.
- [3] ANIRUDH, R.; VENKATARAMAN, V.; NATESAN RAMAMURTHY, K.; TURAGA, P. «A Riemannian framework for statistical analysis of topological persistence diagrams». A: *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2016, 1023–1031.
- [4] BAUER, U.; KERBER, M.; REININGHAUS, J. «Distributed computation of persistent homology». A: *Proceedings of the Meeting on Algorithm Engineering & Experiments*. Filadèlfia, Pa.: Society for Industrial and Applied Mathematics, 2014, 31–38.
- [5] BAUER, U.; KERBER, M.; REININGHAUS, J. «Clear and compress: Computing persistent homology in chunks». A: BREMER, P. T.; HOTZ, I.; PASCUCCI, V.; PEIKERT, R. (ed.). *Topological Methods in Data Analysis and Visualization III*. Cham: Springer, 2014, 103–117.
- [6] BAUER, U.; KERBER, M.; REININGHAUS, J.; WAGNER, H. «Phat—persistent homology algorithms toolbox». *J. Symbolic Comput.*, 78 (2017), 76–90.
- [7] BENDICH, P.; MARRON, J. S.; MILLER, E.; PIELOCH, A.; SKWERER, S. «Persistent homology analysis of brain artery trees». *Ann. Appl. Stat.*, 10 (1) (2016), 198–218.
- [8] BERTSEKAS, D. P. «The auction algorithm: a distributed relaxation method for the assignment problem». *Ann. Oper. Res.*, 14 (1–4) (1988), 105–123.
- [9] BOISSONNAT, J.-D.; DEY, T. K.; MARIA, C. «The compressed annotation matrix: an efficient data structure for computing persistent cohomology». *Algorithmica*, 73 (3) (2015), 607–619.
- [10] BUBENIK, P. «Statistical topological data analysis using persistence landscapes». *J. Mach. Learn. Res.*, 16 (2015), 77–102.
- [11] BUBENIK, P.; DŁOTKO, P. «A persistence landscapes toolbox for topological statistics». *J. Symbolic Comput.*, 78 (2017), 91–114.
- [12] CARLSSON, G. «Topology and data». *Bull. Amer. Math. Soc. (N.S.)*, 46 (2) (2009), 255–308.
- [13] CARLSSON, G.; GABRIELSSON, R. B. «Topological approaches to deep learning». A: BAAS N. A.; CARLSSON, G. E.; QUICK, G.; SZYMIK, M.; THAULE, M. (ed.). *Topological Data Analysis*. Cham: Springer, 2020, 119–146. (Abel Symposia)
- [14] CARLSSON, G.; ISHKHANOV, T.; DE SILVA, V.; ZOMORODIAN, A. «On the local behavior of spaces of natural images». *Int. J. Comput. Vis.*, 76 (1) (2008), 1–12.

- [15] CARLSSON, G.; JARDINE, R.; FEICHTNER-KOZLOV, D.; MOROZOV, D. [et al.]. «Topological data analysis and machine learning theory». Tech. Rep. 12w5081, Banff Inter. Res. Station Math. Innov. Disc (2012).
- [16] CARLSSON, G.; SINGH, G.; ZOMORODIAN, A. «Computing multidimensional persistence». A: *Algorithms and Computation*. Berlin: Springer, 2009. 730–739. (Lecture Notes in Comput. Sci.; 5878)
- [17] CARLSSON, G.; ZOMORODIAN, A.; COLLINS, A.; GUIBAS, L. J. «Persistence barcodes for shapes». *Int. J. Shape Model.*, 11 (2) (2005), 149–187.
- [18] CARRIÈRE, M.; CUTURI, M.; OUDOT, S. «Sliced Wasserstein kernel for persistence diagrams». A: *Proceedings of the 34th International Conference on Machine Learning*. Vol. 17. Sydney, Austràlia, 2017, 664–673.
- [19] CARRIÈRE, M.; MICHEL, B.; OUDOT, S. «Statistical analysis and parameter selection for Mapper». *J. Mach. Learn. Res.*, 19 (2018), article núm. 12, 39 p.
- [20] CARRIÈRE, M.; OUDOT, S. «Structure and stability of the one-dimensional mapper». *Found. Comput. Math.*, 18 (6) (2018), 1333–1396.
- [21] CHAN, J. M.; CARLSSON, G.; RABADAN, R. «Topology of viral evolution». *Proc. Natl. Acad. Sci. USA*, 110 (46) (2013), 18566–18571.
- [22] CHAZAL, F.; DE SILVA, V.; GLISSE, M.; OUDOT, S. *The Structure and Stability of Persistence Modules*. Cham: Springer, 2016. (SpringerBriefs in Mathematics)
- [23] CHAZAL, F.; DE SILVA, V.; OUDOT, S. «Persistence stability for geometric complexes». *Geom. Dedicata*, 173 (2014), 193–214.
- [24] CHAZAL, F.; FASY, B.; LECCI, F.; MICHEL, B.; RINALDO, A.; WASSERMAN, L. «Robust topological inference: distance to a measure and kernel distance». *J. Mach. Learn. Res.*, 18 (2017), article núm. 159, 40 p.
- [25] CHAZAL, F.; FASY, B. T.; LECCI, F.; RINALDO, A.; WASSERMAN, L. «Stochastic convergence of persistence landscapes and silhouettes». A: *Computational Geometry (SoCG'14)*. Nova York: ACM, 2014, 474–483.
- [26] CHAZAL, F.; OUDOT, S. Y. «Towards persistence-based reconstruction in Euclidean spaces». A: *Computational Geometry (SCG'08)*. Nova York: ACM, 2008, 232–241.
- [27] CHEN, C.; KERBER, M. «An output-sensitive algorithm for persistent homology». *Comput. Geom.*, 46 (4) (2013), 435–447.
- [28] CHEN, Y.-C.; WANG, D.; RINALDO, A.; WASSERMAN, L. «Statistical analysis of persistence intensity functions». Preprint, 2015. [Disponible en línia a: <https://arxiv.org/abs/1510.02502>]
- [29] COHEN-STEINER, D.; EDELSBRUNNER, H.; HARER, J. «Stability of persistence diagrams». *Discrete Comput. Geom.*, 37 (1) (2007), 103–120.
- [30] COHEN-STEINER, D.; EDELSBRUNNER, H.; HARER, J.; MILEYKO, Y. «Lipschitz functions have L_p -stable persistence». *Found. Comput. Math.*, 10 (2) (2010), 127–139.

- [31] COHEN-STEINER, D.; EDELSBRUNNER, H.; MOROZOV, D. «Vines and vineyards by updating persistence in linear time». A: *Computational Geometry (SCG'06)*. Nova York: ACM, 2006, 119-126.
- [32] CORBET, R.; FUGACCI, U.; KERBER, M.; LANDI, C.; WANG, B. «A kernel for multi-parameter persistent homology». *Comput. Graph.: X*, 2 (2019), 100005.
- [33] CURTO, C. «What can topology tell us about the neural code?». *Bull. Amer. Math. Soc. (N.S.)*, 54 (1) (2017), 63-78.
- [34] DE SILVA, V.; CARLSSON, G. «Topological estimation using witness complexes». A: *SPBG'04: Proceedings of the First Eurographics conference on Point-Based Graphics*. The Eurographics Association, 2004, 157-166.
- [35] DE SILVA, V.; GHRIST, R. «Coverage in sensor networks via persistent homology». *Algebr. Geom. Topol.*, 7 (2007), 339-358.
- [36] DELFINADO, C. J. A.; EDELSBRUNNER, H. «An incremental algorithm for Betti numbers of simplicial complexes on the 3-sphere». *Grid generation, finite elements, and geometric design. Comput. Aided Geom. Design*, 12 (7) (1995), 771-784.
- [37] DEY, T. K.; FAN, F.; WANG, Y. «Computing topological persistence for simplicial maps [extended abstract]». A: *Computational Geometry (SoCG'14)*. Nova York: ACM, 2014, 345-354.
- [38] DEY, T. K.; MÉMOLI, F.; WANG, Y. «Multiscale mapper: topological summarization via codomain covers». A: *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*. Nova York: ACM, 2016, 997-1013.
- [39] DEY, T. K.; MÉMOLI, F.; WANG, Y. «Topological analysis of nerves, Reeb spaces, mappers, and multiscale mappers». Preprint, 2017. [Disponible en línea a: <https://arxiv.org/abs/1703.07387>]
- [40] DI FABIO, B.; FERRI, M. «Comparing persistence diagrams through complex vectors». A: *Image Analysis and Processing—ICIAP 2015. Part I*. Cham: Springer, 2015, 294-305. (Lecture Notes in Comput. Sci.; 9279)
- [41] EDELSBRUNNER, H.; HARER, J. «Persistent homology—a survey». A: *Surveys on Discrete and Computational Geometry*. Providence, RI: Amer. Math. Soc., 2008, 257-282. (Contemp. Math.; 453)
- [42] EDELSBRUNNER, H.; HARER, J. L. *Computational Topology. An Introduction*. Providence, RI: American Mathematical Society, 2010.
- [43] EDELSBRUNNER, H.; LETSCHER, D.; ZOMORODIAN, A. «Topological persistence and simplification». *Discrete and computational geometry and graph drawing (Columbia, SC, 2001)*. *Discrete Comput. Geom.*, 28 (4) (2002), 511-533.
- [44] EDELSBRUNNER, H.; PARSA, S. «On the computational complexity of Betti numbers: reductions from matrix rank». A: *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*. Nova York: ACM, 2014, 152-160.

- [45] EFRAT, A.; ITAI, A.; KATZ, M. J. «Geometry helps in bottleneck matching and related problems». *Algorithmica*, 31 (1) (2001), 1-28.
- [46] FENG, M.; HICKOK, A.; PORTER, M. A. «Topological data analysis of spatial systems». Preprint, 2014. [Disponible en línia a: <https://arxiv.org/abs/2104.00720>]
- [47] FROSINI, P. «Measuring shapes by size functions». A: *Proceedings: Intelligent Robots and Computer Vision X: Algorithms and Techniques*. Vol. 1607. Boston, MA, 1992, 122-133.
- [48] FROSINI, P.; LANDI, C. «Stability of multidimensional persistent homology with respect to domain perturbations». Preprint, 2010. [Disponible en línia a: <https://arxiv.org/abs/1001.1078>]
- [49] FROSINI, P.; MULAZZANI, M. «Size homotopy groups for computation of natural size distances». *Bull. Belg. Math. Soc. Simon Stevin*, 6 (3) (1999), 455-464.
- [50] GHRIST, R. «Barcodes: the persistent topology of data». *Bull. Amer. Math. Soc. (N.S.)*, 45 (1) (2008), 61-75.
- [51] GIDEA, M.; KATZ, Y. «Topological data analysis of financial time series: landscapes of crashes». *Phys. A*, 491 (2018), 820-834.
- [52] GIUSTI, C.; GHRIST, R.; BASSETT, D. S. «Two's company, three (or more) is a simplex». *J. Comput. Neurosci.*, 41 (1) (2016), 1-14.
- [53] GREENBERG, M. J. *Algebraic Topology: A first Course*. CRC Press, 2018.
- [54] GUSS, W. H.; SALAKHUTDINOV, R. «On characterizing the capacity of neural networks using algebraic topology». Preprint, 2018. [Disponible en línia a: <https://arxiv.org/abs/1802.04443>]
- [55] HARRINGTON, H. A.; OTTER, N.; SCHENCK, H.; TILLMANN, U. «Stratifying multiparameter persistent homology». *SIAM J. Appl. Algebra Geom.*, 3 (3) (2019), 439-471.
- [56] HATCHER, A. *Algebraic Topology*. Cambridge: Cambridge University Press, 2002.
- [57] HAUSMANN, J.-C. «On the Vietoris-Rips complexes and a cohomology theory for metric spaces». A: *Prospects in Topology*. Princeton, NJ: Princeton University Press, 1995, 175-188. (Ann. of Math. Stud.; 138)
- [58] HENSELMAN, G.; GHRIST, R. «Matroid filtrations and computational persistent homology». Preprint, 2016. [Disponible en línia a: <https://arxiv.org/abs/1606.00199>]
- [59] HOFER, C.; KWITT, R.; NIETHAMMER, M.; UHL, A. «Deep learning with topological signatures». Preprint, 2017. [Disponible en línia a: <https://arxiv.org/abs/1707.04041>]
- [60] IBÁÑEZ-MARCELO, E.; CAMPIONI, L.; PHINYOMARK, A.; PETRI, G.; SANTAR-CANGELO, E. L. «Topology highlights mesoscopic functional equivalence between imagery and perception: The case of hypnotizability». *NeuroImage*, 200 (2019), 437-449.

- [61] JACOBSON, N. *Lie Algebras*. Nova York: Dover, p. 163, 1979.
- [62] KANNAN, H.; SAUCAN, E.; ROY, I.; SAMAL, A. «Persistent homology of unweighted complex networks via discrete Morse theory». *Sci. Rep.*, 9, article núm. 13817 (2019).
- [63] KERBER, M.; MOROZOV, D.; NIGMETOV, A. «Geometry helps to compare persistence diagrams». *ACM J. Exp. Algorithmics*, 22 (2017), art. 1.4, 20 p.
- [64] KNUDSON, K. P. «A refinement of multi-dimensional persistence». *Homology Homotopy Appl.*, 10 (1) (2008), 259–281.
- [65] KUSANO, G.; FUKUMIZU, K.; HIRAOKA, Y. «Persistence weighted Gaussian kernel for topological data analysis». A: *ICML'16: Proceedings of the 33rd International Conference on International Conference on Machine Learning*. Vol. 48. Nova York: JMLR.org, 2016, 2004–2013.
- [66] KWITT, R.; HUBER, S.; NIETHAMMER, M.; LIN, W.; BAUER, U. «Statistical topological data analysis - A kernel perspective». A: *Advances in Neural Information Processing Systems*. Vol. 28. Curran Associates, Inc., 2015, 3070–3078.
- [67] KYEONG, S.; PARK, S.; CHEON, K.-A.; KIM, J.-J.; SONG, D.-H.; KIM, E. «A new approach to investigate the association between brain functional connectivity and disease characteristics of attention-deficit/hyperactivity disorder: Topological neuroimaging data analysis». *PLoS One*, 10 (9) (2015), 1–15.
- [68] LEE, Y.; BARTHEL, S. D.; DŁOTKO, P.; MOHAMAD MOOSAVI, S.; HESS, K.; SMIT, B. «Quantifying similarity of pore-geometry in nanoporous materials». *Nat. Commun.*, 8, article núm. 15396 (2017).
- [69] LESNICK, M.; WRIGHT, M. «Interactive visualization of 2-D persistence modules». Preprint, 2015. [Disponible en línia a: <https://arxiv.org/abs/1512.00180>]
- [70] LEWIS, R.; MOROZOV, D. «Parallel computation of persistent homology using the blowup complex». A: *Proceedings of the 27th ACM Symposium on Parallelism in Algorithms and Architectures, SPAA '15*. Nova York: ACM, 2015, 323–331.
- [71] LI, C.; OVSJANIKOV, M.; CHAZAL, F. «Persistence-based structural recognition». A: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2014, 2003–2010.
- [72] LUM, P. Y.; SINGH, G.; LEHMAN, A.; ISHKANOV, T.; VEJDEMO-JOHANSSON, M.; ALAGAPPAN, M.; CARLSSON, J.; CARLSSON, G. «Extracting insights from the shape of complex data using topology». *Sci. Rep.*, 3, article núm. 1236 (2013).
- [73] MILEYKO, Y.; MUKHERJEE, S.; HARER, J. «Probability measures on the space of persistence diagrams». *Inverse Problems*, 27 (12) (2011), 124007, 22 p.
- [74] MILOSAVLJEVIĆ, N.; MOROZOV, D.; ŠKRABA, P. «Zigzag persistent homology in matrix multiplication time». A: *Computational Geometry (SCG'11)*. Nova York: ACM, 2011, 216–225.

- [75] MITCHELL, T. M. *Machine Learning*. Burr Ridge, IL: McGraw Hill, 45 (1997), 37.
- [76] MORONI, D.; PASCALI, M. A. «Learning topology: Bridging computational topology and machine learning». A: *Pattern Recognition. ICPR International Workshops and Challenges*. Springer International Publishing, 2021, 211–226. (Lecture Notes in Computer Science)
- [77] MUNKRES, J. R. *Elements of Algebraic Topology*. Menlo Park, CA: Addison-Wesley Publishing Company, 1984.
- [78] NAITZAT, G.; ZHITNIKOV, A.; LIM, L.-H. «Topology of deep neural networks». *J. Mach. Learn. Res.*, 21 (2020), article núm. 184, 40 p.
- [79] OTTER, N.; PORTER, M. A.; TILLMANN, U.; GRINDROD, P.; HARRINGTON, H. A. «A roadmap for the computation of persistent homology». *EPJ Data Sci.*, 6, article núm. 17 (2017).
- [80] PEREA, J. A. «A brief history of persistence». Preprint, 2018. [Disponible en línia a: <https://arxiv.org/abs/1809.03624>]
- [81] PEREA, J. A.; DECKARD, A.; HAASE, S. B.; HARER, J. «SW1PerS: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data». *BMC Bioinf.*, 16, article núm. 257 (2015).
- [82] PEREA, J. A.; HARER, J. «Sliding windows and persistence: an application of topological methods to signal analysis». *Found. Comput. Math.*, 15 (3) (2015), 799–838.
- [83] PETRI, G.; EXPERT, P.; TURKHEIMER, F.; CARHART-HARRIS, R.; NUTT, D.; HELLYER, P. J.; VACCARINO, F. «Homological scaffolds of brain functional networks». *J. R. Soc. Interface*, 11 (1) (2014), 20140873.
- [84] REININGHAUS, J.; HUBER, S.; BAUER, U.; KWITT, R. «A stable multi-scale kernel for topological machine learning». A: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2015, 4741–4748.
- [85] RIECK, B.; TOGNINALLI, M.; BOCK, C.; MOOR, M.; HORN, M.; GUMBSCH, T.; BORGWARDT, K. «Neural persistence: A complexity measure for deep neural networks using algebraic topology». Preprint, 2018. [Disponible en línia a: <https://arxiv.org/abs/1812.09764>]
- [86] ROBINS, V. «Towards computing homology from finite approximations». A: *Proceedings of the 14th Summer Conference on General Topology and its Applications. Topology Proc.*, 24 (1999), Summer, 503–532 (2001).
- [87] SCOLAMIERO, M.; CHACHÓLSKI, W.; LUNDMAN, A.; RAMANUJAM, R.; ÖBERG, S. «Multidimensional persistence and noise». *Found. Comput. Math.*, 17 (6) (2017), 1367–1406.
- [88] SHEEHY, D. R. «Linear-size approximations to the Vietoris-Rips filtration». *Discrete Comput. Geom.*, 49 (4) (2013), 778–796.
- [89] SINGH, G.; MÉMOLI, F.; CARLSSON, G. «Topological methods for the analysis of high dimensional data sets and 3D object recognition». A: *Eurographics*

- Symposium on Point-Based Graphics*. The Eurographics Association, 2007, 91-100.
- [90] SIZEMORE, A.; GIUSTI, C.; BASSETT, D. S. «Classification of weighted networks through mesoscale homological features». *J. Complex Netw.*, 5 (2) (2017), 245-273.
- [91] SKRYZALIN, J.; CARLSSON, G. «Numeric invariants from multidimensional persistence». *J. Appl. Comput. Topol.*, 1 (1) (2017), 89-119.
- [92] TAUZIN, G.; LUPO, U.; TUNSTALL, L.; BURELLA PÉREZ, J.; CAORSI, M.; MEDINA-MARDONES, A. M.; DASSATTI, A.; HESS, K. «giotto-tda: a topological data analysis toolkit for machine learning and data exploration». *J. Mach. Learn. Res.*, 22 (2021), article núm. 39, 6 p.
- [93] THE RIVET DEVELOPERS. «Rivet». [Disponible en línea a: <https://github.com/rivetTDA/rivet/>]
- [94] TRALIE, C.; SAUL, N.; BAR-ON, R. «Ripser.py: A lean persistent homology library for Python». *The Journal of Open Source Software*, 3 (29) (2018), 925.
- [95] VAN VEEN, H.; SAUL, N.; EARGLE, D.; MANGHAM, S. «Kepler Mapper: A flexible Python implementation of the Mapper algorithm». *The Journal of Open Source Software*, 4 (42) (2019), 1315.
- [96] YAO, Y.; SUN, J.; HUANG, X.; BOWMAN, G. R.; SINGH, G.; LESNICK, M.; GUIBAS, L. J.; PANDE, V. S.; CARLSSON, G. «Topological methods for exploring low-density states in biomolecular folding pathways» *J. Chem. Phys.*, 130 (14) (2009), 144115.
- [97] YOO, J.; KIM, E. Y.; AHN, Y. M.; YE, J. C. «Topological persistence vineyard for dynamic functional brain connectivity during resting and gaming stages». *J. Neurosci. Methods*, 267 (2016), 1-13.
- [98] ZOMORODIAN, A. «The tidy set: a minimal simplicial set for computing homology of clique complexes [extended abstract]».A: *Computational Geometry (SCG'10)*. Nova York: ACM, 2010, 257-266.
- [99] ZOMORODIAN, A.; CARLSSON, G. «Computing persistent homology». *Discrete Comput. Geom.*, 33 (2) (2005), 249-274.

ISI FOUNDATION
VIA CHISOLA 5, 10126, TORÍ, ITÀLIA
esther.ibanez@isi.it