



SAPIENZA
UNIVERSITÀ DI ROMA

FACOLTÀ DI SCIENZE MATEMATICHE FISICHE E NATURALI
Corso di Laurea Triennale in Matematica

OMOLOGIA PERSISTENTE APPLICATA ALLO STUDIO DEL CONNETTOMA

Relatore:
Ch.mo prof.
Marco Manetti

Candidato:
Veronica Paparozzi
matricola 1643491

Sessione Autunnale
Anno Accademico 2018-2019
Dipartimento di Matematica 'Guido Castelnuovo'

A chi c'è sempre stato.

La matematica è l'alfabeto con cui Dio ha scritto l'universo.

-G. Galilei

Indice

Indice	i
Introduzione	iii
1 Omologia simpliciale	1
1.1 Definizioni e risultati preliminari	1
1.2 Complessi simpliciali astratti	5
1.3 Complesso di cricche di un grafo semplice	10
1.3.1 Calcolo dell'omologia	12
2 Omologia persistente	19
2.1 Costruzione dell'omologia persistente	19
2.2 Visualizzazione	22
2.2.1 Barcode	22
2.2.2 Diagramma di persistenza	22
2.3 Stabilità	23
2.4 <i>Point cloud</i> e costruzione del complesso filtrato	25
3 Applicazione al connettoma	29
3.1 Nozioni preliminari sul cervello <i>umano</i>	29
3.2 Costruzione del modello	30
3.3 Studio della struttura	32
3.3.1 Cricche	32
3.3.2 Cavità	34
3.4 Risoluzione delle discordanze tra i due modelli	36
Bibliografia	37

Introduzione

L'omologia persistente è uno *strumento* utilizzato nell'analisi topologica (*Topological Data Analysis*, o TDA) per rilevare le caratteristiche peculiari di una struttura di dati. Nato in origine come strumento utilizzato in teoria della misura (anni '90), è stato successivamente riformulato per la topologia algebrica (inizio 2000), e nel corso degli anni ampliato a strutture algebriche sempre più generali. In particolare, sarà quest'ultimo l'ambito a cui ci restringeremo.

Per avvicinarci a comprenderne il funzionamento e lo scopo, consideriamo il seguente esempio:

Ci troviamo sulla superficie della Terra; ci poniamo il problema di contare le montagne. Cominciamo dal principio: *cos'è una montagna?* Banalmente possiamo identificare ciascuna montagna con il suo picco. Ma a questo punto sorge un problema: non tutti i picchi che vediamo corrispondono a montagne, alcuni possono essere colline, o altipiani. Come possiamo risolvere? Dovremmo individuare i picchi che sono particolarmente alti: fissando un valore di soglia (ad esempio 500 metri), potremmo classificare come montagna (i.e., picco rilevante) tutte quelle vette che superano tale soglia.

Possiamo fare un discorso analogo con quella che è una delle caratteristiche invarianti di un oggetto topologico: l'omologia. Senza entrare nel formalismo (momentaneamente), questo è equivalente a dire che deformando "senza strappi" un oggetto, questo presenterà sempre lo stesso numero di *buchi*.

Abbiamo visto che contare i picchi non ci permette di distinguere una montagna da una collina; allo stesso modo, contare i buchi (i.e., calcolare l'omologia) non ci permette di distinguere un buco *rilevante* da uno *fittizio*. In questo caso, come possiamo fare una distinzione? La risposta a questa domanda è data dall'omologia persistente.

Negli ultimi anni, gli studi su questo campo hanno prodotto sempre più velocemente un gran numero di risultati, ed è in costante crescita il numero di applicazioni di questo strumento (dallo studio della struttura del DNA a quello delle proteine, dal riconoscimento dei vari leucociti e delle cellule tumorali a quello delle lesioni causate da melanomi o da problemi di natura epatica, dallo studio delle spirali dei cicloni a quello delle galassie, dal riconoscimento della scrittura a mano a quello delle impronte digitali). Tale fatto è giustificato sia dalla *consistenza* della rappresentazione (vedi Sezione 2.1 e 2.2) che dalla *stabilità* dei risultati sotto piccole perturbazioni dei dati iniziali (vedi Sezione 2.3).

Lo scopo di questa tesi è introdurre questo *strumento*, focalizzandosi sulle proprietà rilevanti per descrivere una particolare applicazione: lo studio del *connettoma*, ovvero della rete di connessioni nervose all'interno del cervello.

In particolare, nel capitolo 1 vedremo i concetti di partenza: costruiremo la struttura utilizzata per rappresentare il connettoma (i.e., il complesso di cricche associato ad un grafo semplice) partendo dal concetto di complesso simpliciale astratto, ed entreremo nel merito dell'omologia presentando alcuni metodi di calcolo con relativa applicazione.

Il capitolo 2 costituisce il *cervello* della descrizione: qui *costruiremo* l'omologia persistente a partire dai concetti enunciati precedentemente, ne mostreremo i metodi di rappresentazione grafica ed enunceremo il Teorema di stabilità per giustificarne l'attendibilità. Infine, considereremo alcune strutture che serviranno per la costruzione dei modelli del terzo capitolo.

Il capitolo 3 tratta dell'applicazione al connettoma: riporteremo i risultati di uno studio [Siz17] (focalizzandoci sull'analisi di due tipi di strutture: le cricche e le cavità) e ne descriveremo i passaggi salienti utilizzando gli strumenti matematici enunciati nei primi due capitoli.

Le figure non originali utilizzate in questa tesi hanno la fonte indicata nelle rispettive didascalie.

Capitolo 1

Omologia simpliciale

1.1 Definizioni e risultati preliminari

In questa sezione ci occuperemo di definire i concetti di base e i risultati utili per trattare gli argomenti nel seguito.

Le nozioni qui riportate sono tratte da [Spa66], [AM69], [Man14] e [Zom05].

Cominciamo col definire brevemente cos'è una *categoria*.

Una **categoria** \mathbf{C} è una struttura matematica che consiste di tre dati:

- una classe $Ob(\mathbf{C})$ i cui elementi sono detti *oggetti* della categoria (indicati con le lettere A, B, C, \dots);
- una classe $Mor(\mathbf{C})$ i cui elementi sono detti *morfismi*: data una coppia di oggetti A, B in \mathbf{C} , l'insieme $\mathbf{C}(A, B) \subset Mor(\mathbf{C})$ è composto da tutti i *morfismi* da A in B ;
- data una terna di oggetti A, B, C in \mathbf{C} , una *legge di composizione*:

$$\begin{aligned} \mathbf{C}(A, B) \times \mathbf{C}(B, C) &\rightarrow \mathbf{C}(A, C) \\ (f, g) &\mapsto g \circ f, \end{aligned}$$

con $f : A \rightarrow B, g : B \rightarrow C$ e $g \circ f$ la loro composizione, che per semplicità di notazione denoteremo con gf .

Una categoria, affinché risulti tale, deve rispettare tre assiomi:

A1: Due insiemi $\mathbf{C}(A_1, B_1), \mathbf{C}(A_2, B_2)$ si diranno *disgiunti* a meno che $A_1 = A_2, B_1 = B_2$;

A2: Dati $f \in \mathbf{C}(A, B), g \in \mathbf{C}(B, C), h \in \mathbf{C}(C, D)$, allora

$$h(gf) = (hg)f$$

A3: Per ogni $A, B \in Ob(\mathbf{C})$ esiste un morfismo $Id_A \in \mathbf{C}(A, A)$ tale che se $f \in \mathbf{C}(A, B)$, allora $Id_A f = f$ e se $g \in \mathbf{C}(B, A)$ si ha $g Id_A = g$.

In particolare, in **A3**, il morfismo identità se esiste è unico. Possiamo considerare **A2** come la *condizione di associatività* della composizione, e **A3** come la *condizione di esistenza (e unicità) dell'elemento neutro*.

Dunque, in una categoria, utilizziamo i morfismi per porre in relazione differenti oggetti della categoria considerata; allo stesso modo, possiamo stabilire una relazione tra categorie differenti. Tale trasformazione è detta *functore*. Un **functore covariante** [o *functore controvariante*] $F : \mathbf{C} \rightarrow \mathbf{D}$ è un'applicazione che associa

- ad ogni oggetto X di \mathbf{C} un oggetto $F(X)$ di \mathbf{D} ;
- ad ogni morfismo $f \in \mathbf{C}(X, Y)$ un morfismo $F(f) \in \mathbf{D}(F(X), F(Y))$ [o $F(f) \in \mathbf{D}(F(Y), F(X))$] tale che

$$F(fg) = (F(f))(F(g)) \quad [o \quad F(fg) = (F(g))(F(f))] ,$$

$$F(Id_X) = Id_{F(X)} .$$

In questa trattazione utilizzeremo esclusivamente funtori covarianti. In particolare, possiamo definire la categoria dei funtori (covarianti), considerando questi ultimi come gli oggetti della categoria e i loro morfismi dati dalle cosiddette *trasformazioni naturali*.

Dati due funtori $F, G: \mathbf{C} \rightarrow \mathbf{D}$, dare una **trasformazione naturale** $\phi: F \rightarrow G$, significa dare, per ogni oggetto X di \mathbf{A} , un morfismo $\phi_X \in \mathbf{D}(F(X), G(X))$ tale che per ogni morfismo $f \in \mathbf{C}(X, Y)$ si abbia un diagramma commutativo

$$\begin{array}{ccc} F(X) & \xrightarrow{F(f)} & F(Y) \\ \phi_X \downarrow & & \phi_Y \downarrow \\ G(X) & \xrightarrow{G(f)} & G(Y) . \end{array}$$

Diremo che, la trasformazione naturale ϕ è un **isomorfismo di funtori** se ϕ_X è un isomorfismo nella categoria \mathbf{D} per ogni $X \in \mathbf{C}$. In tal caso scriveremo $F \simeq G$.

In particolare, diremo che un funtore $F: \mathbf{C} \rightarrow \mathbf{D}$ è una **equivalenza di categorie** se esiste un funtore $G: \mathbf{D} \rightarrow \mathbf{C}$ tale che le composizioni FG e GF siano isomorfe al funtore identità, ovvero

$$FG \simeq Id_{\mathbf{D}} \quad e \quad GF \simeq Id_{\mathbf{C}} .$$

Due categorie si dicono **equivalenti** se esiste una *equivalenza* fra di loro.

Consideriamo ora un anello R fissato, tale che esso sia un anello commutativo unitario con unità diversa da 0.

Un **modulo differenziale**, o **complesso**, C consiste di un R -modulo C e di un endomorfismo $d: C \rightarrow C$ tale che $dd = 0$. L'endomorfismo d è detto *differenziale*, o *operatore di bordo*, di C . Per un modulo differenziale C definiamo il sottomodulo dei *cicli* $Z(C, R) = \ker d$ e il sottomodulo dei *bordi* $B(C, R) = \text{Im } d$, dove $\ker d$ e $\text{Im } d$ indicano rispettivamente il nucleo e l'immagine di d . Essendo $dd = 0$, abbiamo che $B(C, R) \subseteq Z(C, R)$. Con queste ipotesi è possibile definire il **modulo di omologia** $H(C, R)$ come il modulo quoziente

$$H(C, R) = Z(C, R)/B(C, R) .$$

Gli elementi di $H(C, R)$ sono detti *classi di omologia*. In particolare, se z è un ciclo, la sua classe di omologia in $H(C, R)$ è denotata con $[z]$. Diremo che due cicli z_1 e z_2 sono nella stessa classe di equivalenza, e scriveremo $[z_1] = [z_2]$, se la loro differenza è un bordo.

Sia $\tau: C \rightarrow D$ un omomorfismo tra moduli differenziali che commuta con gli operatori di bordo, allora avremo che τ manda cicli di C in cicli di D . Dunque τ induce un omomorfismo

$$\tau_*: H(C, R) \rightarrow H(D, R)$$

tale che $\tau_*([z]) = [\tau(z)]$, per ogni $z \in Z(C)$. Poiché $(\tau_1\tau_2)_* = \tau_{1*}\tau_{2*}$, esiste un funtore covariante dalla categoria dei gruppi differenziali alla categoria dei gruppi che assegnano al gruppo differenziale C il corrispondente modulo di omologia $H(C, R)$ e all'omomorfismo τ l'omomorfismo indotto τ_* .

Un **anello graduato** è un anello R insieme a una famiglia $\{R_p\}_{p \geq 0}$ di sottogruppi del gruppo additivo di R , tale che $R = \bigoplus_{p=0}^{\infty} R_p$ e $R_m R_n \subseteq R_{m+n}$ per ogni $m, n \geq 0$.

Dato un anello graduato R , un **modulo graduato** su R è un R -modulo M insieme alla famiglia $\{M_p\}_{p \geq 0}$ di sottogruppi di M tale che $M = \bigoplus_{p=0}^{\infty} M_p$ e $R_m M_n \subseteq M_{m+n}$. Un elemento $x \in M$ è detto **omogeneo** di grado p se $x \in M_p$ per qualche p .

Dati due R -moduli graduati M e N , un **omomorfismo graduato** di grado d di R -moduli graduati è una collezione di omomorfismi di R -moduli $\{\tau_p\}_{p \geq 0}$ tale che $\tau_p(M_p) \subseteq N_{p+d}$ per ogni intero d .

Un **modulo differenziale graduato** su R è un modulo graduato avente il differenziale compatibile con la struttura graduata, ovvero diremo che $d_p : M_p \rightarrow M_{p+r}$ è il differenziale di grado r , con r intero.

In particolare, un **complesso di catene** è un modulo differenziale graduato con differenziale di grado $r = -1$. Dunque un complesso di catene C consiste in una sequenza di R -moduli $\{C_p\}$ e omomorfismi $d_p : C_p \rightarrow C_{p-1}$, con $C_p = 0$ se $p < 0$, tali che

$$C_{p+1} \xrightarrow{d_{p+1}} C_p \xrightarrow{d_p} C_{p-1}$$

sia l'omomorfismo banale, ovvero $d_p d_{p+1} = 0$. Il rispettivo omomorfismo graduato di grado $d = 0$ per i moduli differenziali graduati è $\tau : M_p \rightarrow N_p$, che induce l'omomorfismo $\tau_* : H_p(M, R) \rightarrow H_p(N, R)$. In particolare, diremo che τ è un **quasi-isomorfismo** se τ_* è un isomorfismo per qualunque p .

Per un complesso di catene, al variare di p intero, abbiamo i seguenti R -moduli graduati

$$Z(C, R) = \{Z_p(C, R) = \ker d_p\} \quad , \quad B(C, R) = \{B_p(C, R) = \text{Im } d_{p+1}\}$$

da cui il gruppo di omologia dato da

$$H(C, R) = \{H_p(C, R) = Z_p(C, R)/B_p(C, R)\} \quad .$$

Un **sottocomplesso** C' di un complesso di catene C , denotato con $C' \subset C$, è un complesso di catene tale che $C'_p \subset C_p$ e $d'_p = d_p|_{C'_p}$, per ogni p . Inoltre esiste un'inclusione naturale $i : C'_p \hookrightarrow C_p$, per ogni p .

Diremo che un complesso di catene C è **aciclico** se

$$\dots \xrightarrow{d_{p+2}} C_{p+1} \xrightarrow{d_{p+1}} C_p \xrightarrow{d_p} C_{p-1} \xrightarrow{d_{p-1}} \dots$$

è una successione esatta, ovvero $\text{Im } d_{p+1} = \ker d_p$, per ogni p . Nel gruppo di omologia, ciò equivale a dire che $H_p(C, R) = 0$, per ogni p .

Scegliendo $R = \mathbb{Z}$ nelle definizioni suddette, tutti gli R -moduli diventano \mathbb{Z} -moduli e dunque **gruppi abeliani**.

Siano $\tau, \tau' : C \rightarrow D$ omomorfismi di grado $d = 0$ tra moduli differenziali graduati. Un'omotopia tra τ e τ' è una successione di omomorfismi di grado $d = 1$, tale $h = \{h_p : C_p \rightarrow D_{p+1}\}$ al variare di p intero, tale che

$$dh + hd = \tau - \tau' \quad .$$

In tal caso, scriveremo $\tau \sim \tau'$, dove \sim è la relazione di omotopia.

In particolare, diremo che $\tau : C \rightarrow D$ è un'equivalenza omotopica se esiste $\sigma : D \rightarrow C$ tale che $\tau\sigma \sim Id_D$ e $\sigma\tau \sim Id_C$. In questo caso diremo che C e D sono **omotopicamente equivalenti**. In termini di omologia, questo fatto implica un isomorfismo tra i moduli di omologia associati a C e D [HS97, segue dalla Proposizione 3.1].

Un complesso C si dice **contrattile**, o contraibile, se $Id_C \sim 0$. Ovvero, se esiste per ogni p un omomorfismo $h_p : C_p \rightarrow C_{p+1}$ tale che $d_{p+1}h_p + h_{p-1}d_p = Id_C$.

Osservazione 1. In particolare, se C è un complesso contrattile, allora C è aciclico. Per dimostrare ciò è sufficiente far vedere che $\text{Im } d_{p+1} \supseteq \ker d_p$, per ogni p , in quanto il viceversa segue dalla definizione di complesso.

Consideriamo $x \in \ker d_p \subseteq C_p$. Essendo C contrattile si ha che esiste per ogni p l'omomorfismo $h_p : C_p \rightarrow C_{p+1}$ tale che $d_{p+1}h_p + h_{p-1}d_p = \text{Id}_C$, da cui:

$$d_{p+1}(h_p(x)) + h_{p-1}(d_p(x)) = \text{Id}_C(x) \quad ,$$

dove $h_{p-1}(d_p(x)) = h_{p-1}(0) = 0$ e $\text{Id}_C(x) = x$. Risulterà $d_{p+1}(h_p(x)) = x$, da cui $x \in \text{Im } d_{p+1}$.

Riportiamo alcuni risultati che saranno utili nella prossima sezione, riducendoci al caso particolare di *complessi di catene*, sapendo però che vale per complessi generici.

Teorema 1.1.1. *Consideriamo una successione esatta corta di complessi di catene*

$$0 \longrightarrow C \xrightarrow{\alpha} D \xrightarrow{\beta} F \longrightarrow 0 \quad .$$

Allora per ogni p esiste $\sigma_p : H_p(F) \rightarrow H_{p-1}(C)$ tale che la seguente successione

$$\dots \xrightarrow{\sigma_{p+1}} H_p(C) \xrightarrow{\alpha_*} H_p(D) \xrightarrow{\beta_*} H_p(F) \xrightarrow{\sigma_p} H_{p-1}(C) \xrightarrow{\alpha_*} \dots$$

sia esatta.

Per la dimostrazione si fa riferimento a [HS97, Teorema 2.1].

In vista di ciò, diremo che ad ogni successione esatta corta di complessi (di catene) corrisponde una *successione esatta lunga* in omologia.

Segue direttamente il seguente corollario:

Corollario 1.1.2. *Data la successione esatta corta di complessi di catene*

$$0 \longrightarrow C \xrightarrow{\alpha} D \xrightarrow{\beta} F \longrightarrow 0$$

avremo che

- α è un quasi-isomorfismo $\iff F$ è aciclico ;
- β è un quasi-isomorfismo $\iff C$ è aciclico.

Riportiamo ora un risultato di classificazione che sarà utilizzato sia nel primo capitolo, per il calcolo dei gruppi di omologia, sia nel secondo capitolo, per definire la struttura dei moduli dell'omologia di persistenza.

Teorema 1.1.3 (di struttura). *Dato D dominio a ideali principali,*

(1) *ogni D -modulo finitamente generato è isomorfo a*

$$D^\beta \oplus \left[\bigoplus_i (D/a_i D) \right] ,$$

con $\beta \in \mathbb{Z}$ numero di Betti di D e $a_i \in D$ tale che $a_i | a_{i+1}$, $\forall i$.

(2) ogni D -modulo graduato M si decompone unicamente nella forma

$$\left(\bigoplus_{i=1}^n \Sigma^{\alpha_i} D \right) \oplus \left(\bigoplus_{j=1}^m \Sigma^{\gamma_j} D / d_j D \right),$$

dove $d_j \in D$ sono elementi omogenei tali che $d_j | d_{j+1}$, $\alpha_i, \gamma_j \in \mathbb{Z}$, e Σ^α è la α -traslazione verso l'alto nell'indice di gradazione.

Per la dimostrazione si rimanda a []. Il termine a sinistra è detto *sottomodulo libero*, mentre quello a destra è detto *sottomodulo di torsione*. In particolare, il sottomodulo libero comprende tutti i generatori che generano un numero infinito di elementi, mentre il sottomodulo di torsione sono quei generatori che ne generano un numero finito.

1.2 Complessi simpliciali astratti

In questa sezione ci occuperemo di descrivere la categoria dei *complessi simpliciali astratti*, che costituiranno la struttura di partenza del nostro studio, e introdurremo una sua sottocategoria che sarà l'oggetto di interesse principale nel seguito: il complesso di cricche. Costruiremo, inoltre, il complesso di catene associato e, da qui, riporteremo un risultato che garantisce l'indipendenza dell'omologia dall'ordinamento definito sugli elementi di quest'ultimo.

Passiamo, dunque, a definire cos'è un *complesso simpliciale astratto* e come sono definiti i *morfismi* tra questi oggetti.

Definizione 1.1. Un **complesso simpliciale astratto** (K, I) consiste di un insieme I , i cui elementi sono detti *vertici*, e un insieme K , definito come una famiglia di parti finite di I , tali per cui valgano le seguenti condizioni:

- (a) $\langle i \rangle \in K$ per ogni $i \in I$;
- (b) $s \in K$ e $t \subseteq s \Rightarrow t \in K$, per ogni $t \neq \emptyset$.

Nel seguito useremo la notazione $K = (K, I)$ quando non è necessario esplicitare l'insieme dei vertici del complesso simpliciale astratto. Gli elementi di K sono detti **simplessi**: dato un intero $p \geq 0$, diremo che il semplice $s = \langle i_0, \dots, i_p \rangle \in K$ ha dimensione p se esso ha cardinalità $p + 1$, ovvero se s contiene esattamente $p + 1$ vertici distinti: chiameremo un tale s , p -simple. Se t è un q -simple con $q \leq p$ allora $t \subseteq s$, e chiameremo t *faccia* di s (*faccia propria* se $q < p$, e scriveremo $t \subset s$).

La dimensione di K coincide con il massimo delle dimensioni dei suoi simplessi.

Da (a) si può dedurre che esiste una corrispondenza biunivoca tra gli 0-simplessi di K e l'insieme dei vertici I . Da (b) si ha che i complessi simpliciali astratti hanno la proprietà di essere chiusi rispetto all'inclusione, dunque che una faccia di un simple è ancora un simple.

Definizione 1.2. Dati $(K_1, I_1), (K_2, I_2)$ complessi simpliciali astratti, definiamo il **morfismo simpliciale** $\phi : (K_1, I_1) \rightarrow (K_2, I_2)$ come l'applicazione data da $\phi : I_1 \rightarrow I_2$ che manda gli 0-simplessi di K_1 negli 0-simplessi di K_2 tale che per ogni $s \in K_1$ simple si abbia $\phi(s)$ simple di K_2 .

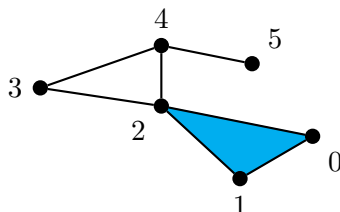
Definizione 1.3. Dato un complesso simpliciale astratto K e un sottoinsieme $L \subseteq K$, diremo che L è un **sottocomplesso** di K se L è un complesso simpliciale astratto.

In particolare, ogni simple di K può essere visto come un sottocomplesso di K . Inoltre, il morfismo simpliciale associato a $L \subseteq K$ diventa un'inclusione.

Definizione 1.4. Definiamo la **chiusura** di L , denotata con \bar{L} , come il più piccolo sottocomplesso di K contenente L , ovvero:

$$\bar{L} = \{t \in K \mid t \subseteq s \in L\}.$$

Esempio 1.2.1. Consideriamo il complesso simpliciale astratto K



Risulterà $K = \{\langle 0, 1, 2 \rangle, \langle 0, 1 \rangle, \langle 1, 2 \rangle, \langle 0, 2 \rangle, \langle 2, 3 \rangle, \langle 3, 4 \rangle, \langle 2, 4 \rangle, \langle 4, 5 \rangle, \langle 0 \rangle, \langle 1 \rangle, \langle 2 \rangle, \langle 3 \rangle, \langle 4 \rangle, \langle 5 \rangle\}$.

Definizione 1.5. Dato un complesso simpliciale astratto K di dimensione n , chiamiamo p -**scheletro** di K , denotato con $K^{(p)}$, l'insieme di tutti i semplici di K di dimensione minore e uguale a p , con $p = 0, \dots, n$.

Proseguendo l'esempio 1.2.1, otteniamo

$$K^{(0)} = \{\langle 0 \rangle, \langle 1 \rangle, \langle 2 \rangle, \langle 3 \rangle, \langle 4 \rangle, \langle 5 \rangle\},$$

$$K^{(1)} = \{\langle 0, 1 \rangle, \langle 1, 2 \rangle, \langle 0, 2 \rangle, \langle 2, 3 \rangle, \langle 3, 4 \rangle, \langle 2, 4 \rangle, \langle 4, 5 \rangle, \langle 0 \rangle, \langle 1 \rangle, \langle 2 \rangle, \langle 3 \rangle, \langle 4 \rangle, \langle 5 \rangle\},$$

$$K^{(2)} = K.$$

Notiamo che $K^{(p)} \setminus K^{(p-1)}$ coincide con l'insieme dei p -simplessi di K . In particolare, $K^{(0)}$ coincide con l'insieme dei vertici di K . Avremo che $K^{(0)} \subset K^{(1)} \subset \dots \subset K^{(p)} \subset \dots \subset K^{(n)} = K$. Dunque se K è finito, anche $K^{(p)}$ è finito per ogni p .

Definiamo ora un particolare tipo di complesso simpliciale astratto:

Definizione 1.6. Sia K un complesso simpliciale astratto. Diremo che K è un **complesso di cricche** se risulta

$$s = \langle i_0, \dots, i_p \rangle \in K \iff \forall j, k = 0, \dots, p \text{ vale } t = \langle i_j, i_k \rangle \in K \quad .$$

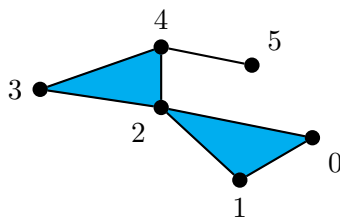
L'implicazione verso destra ci dice che K è un complesso simpliciale astratto. Il viceversa stabilisce che K è univocamente determinato da $K^{(1)}$.

Proponiamo un esempio per chiarire meglio il concetto

Esempio 1.2.2. Notiamo che il complesso simpliciale astratto dell'Esempio 1.2.1 è un complesso simpliciale astratto ma non è un complesso di cricche. Questo perchè si ha:

$$\langle 2, 3 \rangle, \langle 2, 4 \rangle, \langle 3, 4 \rangle \in K \text{ ma } \langle 2, 3, 4 \rangle \notin K \quad .$$

Aggiungendo $\langle 2, 3, 4 \rangle$ a K otteniamo ancora un complesso simpliciale astratto K'



dove, applicando la definizione, si verifica che K' è un complesso di cricche.

Occupiamoci di dare un'*orientazione* al nostro complesso simpliciale astratto K

Definizione 1.7. Dato un complesso simpliciale astratto (K, I) e un intero $p \geq 0$, definiamo $K_p = \{(i_0, \dots, i_p) \in I^{p+1} \mid \langle i_0, \dots, i_p \rangle \in K\}$ l'insieme dei p -**simplessi orientati** di K , dove (i_0, \dots, i_p) è un p -simplesso su cui è fissato un ordinamento dei vertici in cui è ammesso che $(i_0, \dots, i_j, \dots, i_j, \dots, i_p) \in K_p$ per qualche $j = 0, \dots, p$.

Osserviamo che, poiché gli indici possono essere ripetuti, K_p è non vuoto per ogni p , anche nel caso di (K, I) finito con $|I| < p$.

Definizione 1.8. Una p -**catena** di un complesso simpliciale astratto K è una combinazione lineare di p -simplessi orientati a coefficienti nell'anello R

$$\sum_i a_i \cdot s_i \quad , \quad \text{dove } s_i \in K_p, a_i \in R .$$

Possiamo ora costruire l' R -modulo libero sui p -simplessi orientati di K :

$$C_p(K, R) = \bigoplus_{(i_0, \dots, i_p) \in K_p} R \cdot (i_0, \dots, i_p) \quad ,$$

dove $C_p(K, R)$ è l'insieme di tutte le p -catene generate dagli elementi di K_p . Definiamo il *differenziale* $d_p : C_p(K, R) \rightarrow C_{p-1}(K, R)$ sugli elementi della base, ovvero sui p -simplessi orientati:

$$d_p((i_0, \dots, i_p)) = \sum_{k=0}^p (-1)^k (i_0, \dots, \widehat{i}_k, \dots, i_p) \quad ,$$

dove $(i_0, \dots, \widehat{i}_k, \dots, i_p)$ indica la $(p-1)$ -faccia del p -simplesso orientato (i_0, \dots, i_p) , ottenuta togliendo il vertice \widehat{i}_k . Con qualche calcolo è possibile verificare che $d_p d_{p+1} = 0$, $\forall p \geq 0$.

Grazie a quest'ultima considerazione, possiamo definire il seguente complesso di catene

$$C(K, R) : \dots \longrightarrow C_p(K, R) \longrightarrow C_{p-1}(K, R) \longrightarrow \dots \longrightarrow C_1(K, R) \longrightarrow C_0(K, R) \longrightarrow 0 \quad .$$

Consideriamo una relazione di ordine totale qualunque su I , (I, \leq) , e denotiamo con

$$H_p = \{(i_0, \dots, i_p) \in K_p \mid i_j \leq i_l, \forall j < l\}$$

il sottoinsieme dei p -simplessi ordinati e con

$$L_p = \{(i_0, \dots, i_p) \in K_p \mid i_j < i_l, \forall j < l\}$$

il sottoinsieme dei p -simplessi strettamente ordinati.

Indichiamo con E_p l' R -modulo libero generato da H_p e con $D_p \subset E_p$ il sottomodulo libero generato da L_p . Il differenziale è dato da

$$d_E : E_p \rightarrow E_{p-1}, \quad \text{con } d_E(i_0, \dots, i_p) = \sum_{k=0}^p (-1)^k (i_0, \dots, \widehat{i}_k, \dots, i_p) .$$

Essendo D_* sottocomplesso di E_* si ha che il differenziale di D è il differenziale di E ristretto a D , indicato con d_D . Si ha, in particolare, che $E_* \subset C_*$.

Vogliamo dimostrare il seguente risultato

Teorema 1.2.3. *Le inclusioni $D_* \rightarrow E_* \rightarrow C_*$ inducono isomorfismi in omologia:*

$$H_*(D(K, R)) \simeq H_*(E(K, R)) \simeq H_*(C(K, R)) .$$

Per farlo, enunciamo alcuni risultati preliminari.

Dimostriamo innanzitutto che l'inclusione $i : D_* \rightarrow E_*$ è un'equivalenza omotopica. Definiamo la proiezione $\pi : E_* \rightarrow D_*$, tale che

$$\pi(i_0, \dots, i_p) = \begin{cases} (i_0, \dots, i_p) & \text{se } (i_0, \dots, i_p) \in L_p, \\ 0 & \text{se } (i_0, \dots, i_p) \in H_p - L_p. \end{cases}$$

Lemma 1.2.4. *L'inclusione $i : D_* \rightarrow E_*$ è un'equivalenza omotopica.*

Dimostrazione. Per definizione $\pi i = Id$. Vogliamo far vedere che esiste un'omotopia $h : E_* \rightarrow E_{*+1}$ tale che $dh + hd = Id - i\pi$. Dato (i_0, \dots, i_p) semplice ordinato, definiamo

$$h(i_0, \dots, i_p) = \begin{cases} 0 & \text{se } i_j < i_l, \quad \forall j < l, \\ (-1)^k (i_0, \dots, i_k, i_k, i_{k+1}, \dots, i_p) & \text{altrimenti,} \end{cases}$$

dove $0 \leq k < p$ è il più piccolo indice tale che $i_k = i_{k+1}$. Se $(i_0, \dots, i_p) \in L_p$, si ha banalmente che $(dh + hd)(i_0, \dots, i_p) = 0$. Esaminiamo quindi il caso $(i_0, \dots, i_p) \in H_p - L_p$ e sia $0 \leq k < p$ il più piccolo indice tale che $i_k = i_{k+1}$. Otteniamo

$$h(i_0, \dots, i_p) = (-1)^k (i_0, \dots, i_k, i_k, i_{k+1}, \dots, i_p),$$

da cui

$$\begin{aligned} dh(i_0, \dots, i_p) &= (-1)^k \sum_{j=0}^{k-1} (-1)^j (i_0, \dots, \widehat{i_j}, \dots, i_{k-1}, i_k, i_k, i_{k+1}, \dots, i_p) \\ &\quad + (-1)^{2k} (i_0, \dots, i_p) + (-1)^{2k+1} (i_0, \dots, i_p) + (-1)^{2(k+1)} (i_0, \dots, i_p) \\ &\quad + (-1)^k \sum_{j=k+2}^p (-1)^{j+1} (i_0, \dots, i_k, i_k, i_{k+1}, i_{k+2}, \dots, \widehat{i_j}, \dots, i_p). \end{aligned}$$

Calcoliamo ora

$$\begin{aligned} d(i_0, \dots, i_p) &= \sum_{j=0}^{k-1} (-1)^j (i_0, \dots, \widehat{i_j}, \dots, i_{k-1}, i_k, \dots, i_p) \\ &\quad + \left[(-1)^k (i_0, \dots, i_{k-1}, i_{k+1}, \dots, i_p) + (-1)^{k+1} (i_0, \dots, i_k, i_{k+2}, \dots, i_p) \right] \\ &\quad + \sum_{j=k+2}^p (-1)^j (i_0, \dots, i_{k+2}, \dots, \widehat{i_j}, \dots, i_p), \end{aligned}$$

da cui si ha

$$\begin{aligned} hd(i_0, \dots, i_p) &= (-1)^{k-1} \sum_{j=0}^{k-1} (-1)^j (i_0, \dots, \widehat{i_j}, \dots, i_{k-1}, i_k, i_k, i_{k+1}, \dots, i_p) \\ &\quad + h \left[(-1)^k (i_0, \dots, i_{k-1}, i_{k+1}, \dots, i_p) + (-1)^{k+1} (i_0, \dots, i_k, i_{k+2}, \dots, i_p) \right] \quad (I) \\ &\quad + (-1)^k \sum_{j=k+2}^p (-1)^j (i_0, \dots, i_k, i_k, i_{k+1}, i_{k+2}, \dots, \widehat{i_j}, \dots, i_p). \end{aligned}$$

Il termine in (I) vale 0, infatti: se $i_k = i_{k+1}$ era l'unica coppia uguale, per la definizione di h

il termine tra parentesi quadre va a 0, se invece esiste un $k+1 \leq l \leq p$ tale che $i_l = i_{l+1}$ si ha $(-1)^{k+l-1}(i_0, \dots, i_{k-1}, i_{k+1}, \dots, i_l, i_{l+1}, \dots, i_p) + (-1)^{k+l}(i_0, \dots, i_{k-1}, i_k, i_{k+2}, \dots, i_l, i_{l+1}, \dots, i_p) = 0$, essendo $i_k = i_{k+1}$.

Sommando otteniamo

$$(dh + hd)(i_0, \dots, i_p) = (-1)^{2(k+1)}(i_0, \dots, i_p) = (i_0, \dots, i_p).$$

□

Per ogni $k \geq 0$, definiamo

$$H_p^k = \{(i_0, \dots, i_p) \in K_p \mid i_j < i_l, \forall j < l \leq p - k\},$$

con E_p^k modulo libero generato da H_p^k . In particolare, si ha $E_*^k \subset E_*^{k+1}$.

Lemma 1.2.5. *L'inclusione $E_*^k \subset E_*^{k+1}$ è un quasi-isomorfismo per ogni $k \geq 0$.*

Dimostrazione. Per il Corollario 1.1.2, dimostrare il lemma è equivalente a far vedere che E_*^{k+1}/E_*^k è contraibile, ovvero che, fissato $k \geq 0$ e $p \geq 0$ esiste un'omotopia $h : E_p^{k+1} \rightarrow E_{p+1}^{k+1}$ tale che $dh + hd \equiv Id$ modulo E^k . Consideriamo h definita nel modo seguente

$$h(i_0, \dots, i_p) = (-1)^j(i_0, \dots, i_{j-1}, i_{p-k}, i_j, \dots, i_p)$$

dove $j \leq p - k$ è il più piccolo indice tale che $i_{p-k} \leq i_j$. Osserviamo che h preserva il sottocomplesso E^k . Otteniamo il seguente diagramma

$$\begin{array}{ccc} E_p^{k+1} & \xrightarrow{d} & E_{p-1}^{k+1} \\ h \downarrow & & h \downarrow \\ E_{p+1}^{k+1} & \xrightarrow{d} & E_p^{k+1} \end{array}.$$

Fissiamo

$$(i_0, \dots, i_p) \in E_p^{k+1} \quad \text{tale che} \quad i_0 \leq \dots \leq i_{p-k-1},$$

e sia $j \leq p - k$ il più piccolo indice tale che $i_{p-k} \leq i_j$. Si ha

$$h(i_0, \dots, i_p) = (-1)^j(i_0, \dots, i_{j-1}, i_{p-k}, i_j, \dots, i_p)$$

modulo il sottocomplesso E^k si ha:

$$\begin{aligned} dh(i_0, \dots, i_p) &\equiv (-1)^j \sum_{l=0}^{j-1} (-1)^l (i_0, \dots, \widehat{i}_l, \dots, i_{j-1}, i_{p-k}, i_j, \dots, i_p) + (-1)^{2j} (i_0, \dots, i_p) \\ &\quad + (-1)^j \sum_{l=j}^{p-k-1} (i_0, \dots, i_{j-1}, i_{p-k}, i_j, \dots, \widehat{i}_l, \dots, i_{p-k-1}, i_{p-k}, \dots, i_p). \end{aligned}$$

Percorrendo il diagramma dall'altra parte:

$$\begin{aligned} d(i_0, \dots, i_p) &\equiv \sum_{l=0}^{j-1} (-1)^l (i_0, \dots, \widehat{i}_l, \dots, i_{j-1}, \dots, i_p) \\ &\quad + \sum_{l=j}^{p-k-1} (-1)^l (i_0, \dots, i_j, \dots, \widehat{i}_l, \dots, i_{p-k-1}, \dots, i_p), \end{aligned}$$

da cui

$$\begin{aligned} hd(i_0, \dots, i_p) \equiv & (-1)^{j-1} \sum_{l=0}^{j-1} (-1)^l (i_0, \dots, \widehat{i}_l, \dots, i_{j-1}, i_{p-k}, i_j, \dots, i_p) \\ & + (-1)^j \sum_{l=j}^{p-k-1} (-1)^l (i_0, \dots, i_j, i_{p-k}, i_j, \dots, \widehat{i}_l, \dots, i_{p-k-1}, \dots, i_p) . \end{aligned}$$

Sommando si ottiene

$$(dh + hd)(i_0, \dots, i_p) \equiv (i_0, \dots, i_p) .$$

□

Siamo ora pronti a dimostrare il Teorema 1.2.3:

Dimostrazione. Per il lemma 1.2.4 si ha $H_*(D) \simeq H_*(E)$. Per il lemma 1.2.5, $H_*(E^k) \simeq H_*(E^{k+1})$ per ogni $k \geq 0$. Per definizione di H_p^k , risulta $H_p = H_p^0$ per ogni $p \geq 0$, e dunque $E_* = E_*^0$. Otteniamo

$$H_*(E) = H_*(E^0) \simeq H_*(E^1) \simeq \dots \simeq H_*(E^k) \simeq H_*(E^{k+1}) \simeq \dots \simeq H_*(E^p) = H_*(C) ,$$

da cui la tesi.

□

1.3 Complesso di cricche di un grafo semplice

In questa sezione ci occupiamo di definire i grafi, in particolare daremo per nota la definizione di grafo generico e vedremo direttamente la definizione di grafo semplice, per poi mostrare la corrispondenza biunivoca che sussiste tra questi e i complessi di cricche, ottenendo un nuovo oggetto: il complesso di cricche relativo ad un grafo semplice dato, che useremo come il complesso simpliciale astratto su cui calcoleremo i gruppi di omologia del complesso di catene ad esso associato. Sfruttando i concetti enunciati precedentemente utilizziamo la seguente definizione

Definizione 1.9. Un **grafo semplice** $G = (V, E)$ è un complesso simpliciale astratto di dimensione minore o uguale di 1, dove $V = K^{(0)}$ è detto insieme dei vertici ed $E = K^{(1)} - K^{(0)}$ è l'insieme dei lati.

In particolare utilizzeremo la notazione $V(G)$ e $E(G)$ per indicare rispettivamente l'insieme dei vertici e l'insieme dei lati di $G = (V, E)$. Scriveremo $\langle u, v \rangle$ per denotare gli elementi di $E(G)$, dove $u, v \in V(G)$. Un grafo così definito è anche detto **grafo indiretto**, o non ordinato.

Due vertici $u, v \in V(G)$ si dicono *adiacenti* se esiste un elemento $e = \langle u, v \rangle \in E(G)$. Un lato $e \in E(G)$ è *incidente* a $v \in V(G)$ se $v \in e$.

Graficamente possiamo rappresentare ogni vertice con un punto e ogni lato con un segmento che collega due di questi. Notiamo che questa rappresentazione non è univoca; la posizione dei vertici e dei lati non è rilevante ai fini della struttura, serve solo a stabilire le relazioni di *incidenza* tra lati e vertici.

Definizione 1.10. Un grafo H è un **sottografo** di G se

$$V(H) \subseteq V(G) \quad e \quad E(H) \subseteq E(G) .$$

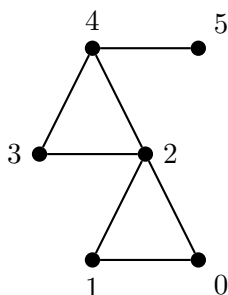
Un grafo è *finito* se $|V(G)| < \infty$ e $|E(G)| < \infty$, *infinito* altrimenti. In particolare, in un grafo semplice è sufficiente richiedere la finitezza di uno solo dei due insiemi, in quanto da essa segue necessariamente la finitezza dell'altro.

Un **cammino** da v_0 a v_k in un grafo $G = (V, E)$ è una sequenza finita $\{v_0, \dots, v_k\}$ dove $v_i \in V$, per $0 \leq i \leq k$.

Definizione 1.11. Un grafo si dice **connesso** se esiste un cammino da u a v , $\forall u, v \in V$.

Nel seguito, quando parleremo di grafi, considereremo sempre grafi semplici, indiretti, finiti e connessi.

Esempio 1.3.1. *Riportiamo un esempio di grafo:*



Definizione 1.12. Dato un grafo $G = (V, E)$, diremo che G è un **grafo completo**, o **cricca**, se

$$\forall u, v \in V(G), \quad e = \langle u, v \rangle \in E(G) \quad .$$

Definizione 1.13. Sia G un grafo. Una **k -cricca** di G , indicata con $\langle v_0, \dots, v_{k-1} \rangle$, è un sottografo completo di G costituito da k elementi di $V(G)$.

In particolare, un sottoinsieme (i.e., una faccia) di una k -cricca è ancora una cricca: se infatti da questa togliamo un vertice v_j , con $j = 0, \dots, k-1$, verranno di conseguenza eliminati tutti i lati incidenti a v_j , mantenendo inalterato il resto della struttura, e in particolare le incidenze tra vertici e lati, ottenendo così una $k-1$ -cricca. Questa proprietà sarà rilevante fra poco, nel momento in cui andremo a relazionare grafi e complessi di cricche. Inoltre, chiameremo **massimale** una cricca che non è faccia di nessun'altra cricca.

Detta \mathbf{K} la categoria dei complessi simpliciali astratti, $\mathbf{G} \subset \mathbf{K}$ la sottocategoria dei grafi e $\mathbf{F} \subset \mathbf{K}$ la sottocategoria dei complessi di cricche, vogliamo dimostrare l'*equivalenza* tra \mathbf{G} ed \mathbf{F} .

Definiamo due funtori:

$$Y : \mathbf{F} \longrightarrow \mathbf{G} \quad \text{e} \quad X : \mathbf{G} \longrightarrow \mathbf{F} .$$

Dalla definizione 1.9 è ben definita l'applicazione Y che ad ogni complesso di cricche K associa il suo 1-scheletro, $K^{(1)}$. Definiamo poi X nel modo seguente:

$$X(\langle v_0, \dots, v_k \rangle) = \langle \phi(v_0), \dots, \phi(v_k) \rangle ,$$

dove $\phi : K^{(1)} \longrightarrow K$ è il morfismo simpliciale dato da $\phi : V(G) \longrightarrow K^{(0)}$. Tale X associa ad ogni $k+1$ -cricca di G il rispettivo k -simpleso indotto da ϕ . Una tale definizione mantiene le proprietà dei complessi simpliciali astratti, essendo il sottoinsieme di una cricca ancora una cricca, così come la faccia di un simpleso risulta essere ancora un simpleso. È possibile verificare che

$$XY \simeq \text{Id}_{\mathbf{F}} \quad \text{e} \quad YX \simeq \text{Id}_{\mathbf{G}} ,$$

da cui si ha che X è un'*equivalenza di categorie*.

Definizione 1.14. Definiamo **complesso di cricche di un grafo G** , indicato con $X(G)$, il complesso simpliciale astratto i cui vertici sono dati da $V(G)$ e i simplessi sono in corrispondenza con le cricche del grafo.

Esempio 1.3.2. Possiamo cogliere i dettagli di questa applicazione considerando, ad esempio, il grafo G dell'esempio 1.3.1, al quale applicandovi X si ottiene $X(G)$ dato dall'esempio 1.2.2.

Essendo $X(G) \in \mathbf{F} \subset \mathbf{K}$, possiamo costruire su di esso un complesso di catene, tale $C(X(G), R)$, in maniera analoga a quanto fatto per i complessi simpliciali astratti. Utilizzeremo questo complesso per il calcolo dei relativi gruppi di omologia.

1.3.1 Calcolo dell'omologia

Intuitivamente, possiamo descrivere il k -esimo gruppo di omologia di un insieme X , dato l'endomorfismo $d : X \rightarrow X$, come un insieme contenente tutti i k -buchi di X , ovvero tutti i k -cicli di X che non sono k -bordi. Dato un complesso simpliciale astratto K , consideriamo il complesso di catene associato $C(K, F)$ con F campo generico, definiamo $\beta_k := \dim H_k(C(K, F))$. Chiameremo tale valore **k -esimo numero di Betti**, che corrisponde intuitivamente al numero di k -buchi del complesso K . In questa sezione ci occuperemo di calcolare gli $H_k(C(X(G), F))$, per $k \geq 0$, dato un grafo G . Per i nostri scopi, ci limiteremo al caso $F = \mathbb{Z}_2$, in modo da facilitare alcuni procedimenti che vedremo a breve.

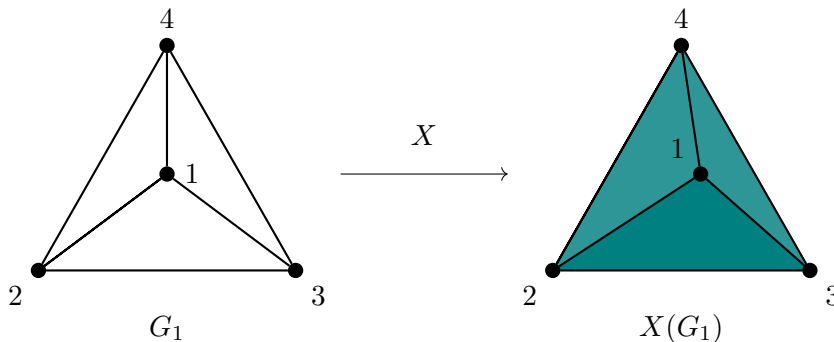
Come preannunciato, per determinare $H_k(C(X(G), R))$, ci viene in contro il Teorema 1.1.3. Nel nostro caso $D = \mathbb{Z}_2$, dunque il relativo sottomodulo di torsione risulta banale essendo $a_i \in \mathbb{Z}_2$. Otteniamo che per il nostro H_k , essendo uno \mathbb{Z}_2 -modulo finitamente generato, dal punto (1) del Teorema 1.1.3 risulta

$$H_k(C(X(G), \mathbb{Z}_2)) = \mathbb{Z}_2^{\beta_k}.$$

Ci è dunque sufficiente calcolare i rispettivi β_k . Al fine di velocizzare i calcoli, possiamo sfruttare alcuni risultati noti per anticipare alcuni valori di β_k . In particolare, per $k = 0$, si ha che il valore β_0 è pari al numero di *componenti connesse* di $X(G)$ (vedi Proposizione 2.7 [Hat01]). Dato che i grafi che consideriamo sono tutti connessi, e dunque anche il complesso di cricche associato, sappiamo a priori che $\beta_0 = 1$.

Vediamo ora qualche esempio di calcolo.

Iniziamo con il complesso di cricche associato al grafo completo G_1 :



Sappiamo già che $H_0(C(X(G_1), \mathbb{Z}_2)) = \mathbb{Z}_2$. Vogliamo dimostrare che $H_p(C(X(G_1), \mathbb{Z}_2)) = 0$ per ogni $p > 0$. Per farlo possiamo estendere il concetto ad un qualunque grafo completo, con il complesso di catene a coefficienti nel solito anello R :

Proposizione 1.3.3. *Sia $X(G)$ complesso di cricche con G grafo completo. Allora risulta*

$$H_p(C(X(G), R)) = 0, \quad \forall p > 0.$$

Dimostrazione. Definiamo l'applicazione lineare $h_p : C_p(X(G), R) \rightarrow C_{p+1}(X(G), R)$ nel modo seguente:

$$h_p((i_0, \dots, i_p)) = (1, i_0, \dots, i_p)$$

e facciamo vedere che $d_{p+1}h_p + h_{p-1}d_p = Id_{C_p}$ per ogni $p > 0$. Dunque, fissiamo $p > 0$ e calcoliamo

$$\begin{aligned} d_{p+1}h_p((i_0, \dots, i_p)) &= d_{p+1}((1, i_0, \dots, i_p)) = \sum_{k=0}^p (-1)^{k+1} (1, i_0, \dots, \hat{i}_k, \dots, i_p) + (i_0, \dots, i_p), \\ h_{p-1}d_p((i_0, \dots, i_p)) &= h_{p-1}\left(\sum_{k=0}^p (-1)^k (i_0, \dots, \hat{i}_k, \dots, i_p)\right) = \sum_{k=0}^p (-1)^k (1, i_0, \dots, \hat{i}_k, \dots, i_p). \end{aligned}$$

Da cui, sostituendo, otteniamo

$$\begin{aligned} (d_{p+1}h_p + h_{p-1}d_p)((i_0, \dots, i_p)) &= \sum_{k=0}^p (-1)^{k+1} (1, i_0, \dots, \hat{i}_k, \dots, i_p) \\ &\quad + \sum_{k=0}^p (-1)^k (1, i_0, \dots, \hat{i}_k, \dots, i_p) + (i_0, \dots, i_p) \\ &= (i_0, \dots, i_p). \end{aligned}$$

Si avrà pertanto

$$H_p(C(X(G), R)) = 0, \quad \forall p > 0.$$

□

Avremo dunque

$$H_0(C(X(G_1), \mathbb{Z}_2)) = \mathbb{Z}_2 \text{ e } H_p(C(X(G_1), \mathbb{Z}_2)) = 0, \quad \forall p > 0.$$

Notiamo che questo ragionamento non ha validità se effettuato su un grafo che non sia completo: non sarebbe infatti garantita l'esistenza dell'1-simplesso avente come vertici di incidenza i_0 e il vertice aggiunto, posto a 1 in quanto consideriamo i vertici del grafo ordinati con i primi numeri naturali (a partire appunto da 1).

Dunque per calcolare il rango dei gruppi di omologia di un complesso di cricche associato ad un grafo che non sia completo dobbiamo rifarci ad altri metodi. In questa trattazione forniremo un risultato che ci sarà utile in alcune situazioni, e vedremo principalmente due metodi che verranno messi a confronto sia per mostrare la validità del risultato sia per sottolineare la differenza di *costo*. Rimandiamo a [ELZ02] per ulteriori metodi di calcolo.

Formula di Eulero - Poincaré. Il primo che andiamo a descrivere non è un vero e proprio metodo, ma sarà un'utile considerazione vista la semplicità delle situazioni che andremo a considerare nell'applicazione del calcolo. Il tutto si basa sul seguente risultato

Teorema 1.3.4 (di Eulero-Poincaré). *Dato un complesso simpliciale astratto K di dimensione n si ha*

$$\sum_{k=0}^n (-1)^k s_k = \sum_{k=0}^n (-1)^k \beta_k,$$

dove s_k è il numero di k -simplessi in K e β_k è il k -esimo numero di Betti.

Dimostrazione. Nella sezione precedente abbiamo definito il complesso D , avente come elementi della base i k -simplessi, con $k = 0, \dots, n$. Dunque si ha $s_k = |D_k|$. Inoltre, per definizione, $\beta_k = \dim H_k(C)$.

Dal Teorema 1.2.3, risulta $\dim H_k(D) = \dim H_k(C)$.

In particolare,

$$\beta_k = \dim H_k(D) = \dim \ker d_{D_k} - \dim \operatorname{Im} d_{D_{k+1}}, \quad \forall k = 0, \dots, n. \quad (1.1)$$

Applicando il *Teorema del rango* al differenziale d_{D_k} del complesso D , si ha

$$s_k = \dim \ker d_{D_k} + \dim \operatorname{Im} d_{D_k},$$

da cui, esplicitando $\dim \ker d_{D_k}$ e sostituendo in (1.1) otteniamo

$$\beta_k = s_k - \dim \operatorname{Im} d_{D_k} - \dim \operatorname{Im} d_{D_{k+1}}.$$

Dunque

$$\begin{aligned} \sum_{k=0}^n (-1)^k \beta_k &= \sum_{k=0}^n (-1)^k (s_k - \dim \operatorname{Im} d_{D_k} - \dim \operatorname{Im} d_{D_{k+1}}) \\ &= s_0 - \dim \operatorname{Im} d_{D_0} - \dim \operatorname{Im} d_{D_1} - s_1 + \dim \operatorname{Im} d_{D_1} + \dim \operatorname{Im} d_{D_2} \\ &\quad + \dots + (-1)^n (s_n - \dim \operatorname{Im} d_{D_n} - \dim \operatorname{Im} d_{D_{n+1}}). \end{aligned}$$

Notiamo che i termini $\dim \operatorname{Im} d_{D_k}$ si cancellano per $k = 1, \dots, n-1$, ottenendo

$$\sum_{k=0}^n (-1)^k \beta_k = \sum_{k=0}^n (-1)^k s_k - \dim \operatorname{Im} d_{D_0} - \dim \operatorname{Im} d_{D_{n+1}},$$

dove $\dim \operatorname{Im} d_{D_0} = \dim \operatorname{Im} d_{D_{n+1}} = 0$, essendo K di dimensione n . □

Ciò si rivelerà utile nel caso in cui sono noti tutti i β_k eccetto uno.

Algoritmo di riduzione (*o di riduzione della matrice associata all'operatore di bordo*).

Il primo metodo che trattiamo, che è anche quello più datato, si rifà ad un noto algoritmo (*Algoritmo di riduzione di Smith-Poincaré*) che prevede la riduzione della matrice associata all'operatore di bordo, detta anche *matrice di riduzione standard*, alla *forma normale di Smith*. Tale procedimento è utilizzato nel caso di coefficienti in anelli a ideali principali, in particolare \mathbb{Z} , ma poichè in questo caso siamo a coefficienti nel campo \mathbb{Z}_2 , la situazione è più semplice, in quanto, come detto sopra, il sottogruppo di torsione del gruppo di omologia risulta essere banale; non è dunque necessario determinare i vari a_i , che nel caso verrebbero a determinarsi prendendo gli unici valore diversi da zero della matrice ridotta, ovvero i primi valori non nulli della diagonale. Trovandoci a coefficienti in \mathbb{Z}_2 , l'operatore di bordo $d_p : C_p(X(G), \mathbb{Z}_2) \rightarrow C_{p-1}(X(G), \mathbb{Z}_2)$ è un'applicazione lineare, possiamo pertanto associarvi una matrice $m_{p-1} \times m_p$ a coefficienti in \mathbb{Z}_2 , detta M_p , con m_{p-1} numero di elementi della base di C_{p-1} e m_p numero di elementi della base di C_p , che possiamo vedere anche come il numero di $p-1$ -simplessi e p -simplessi di $X(G)$, rispettivamente. Il nostro obiettivo è calcolare

$$\dim H_p(C(X(G), \mathbb{Z}_2)) = \dim \left(\frac{\ker d_p}{\operatorname{Im} d_{p+1}} \right) = \dim \ker d_p - \dim \operatorname{Im} d_{p+1}.$$

Per il Teorema del rango si ha che

$$\dim C_p(X(G), \mathbb{Z}_2) = \dim \ker d_p + \dim \operatorname{Im} d_p \quad ,$$

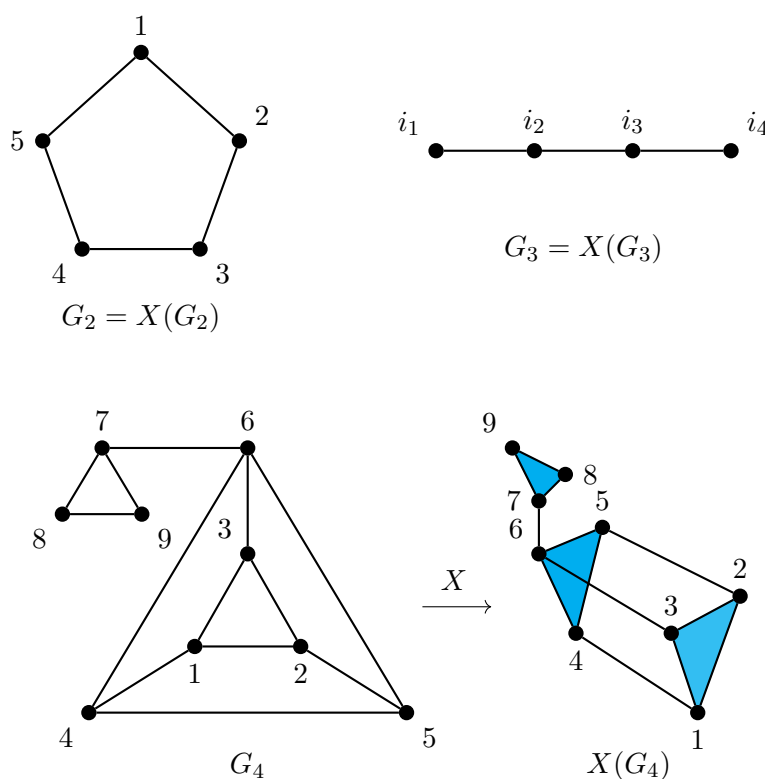
dove $\dim C_p(X(G), \mathbb{Z}_2) = m_p$ e $\dim \operatorname{Im} d_p = \operatorname{rk} M_p$. Da cui $\dim \ker d_p = m_p - \operatorname{rk} M_p$, ottenendo:

$$\beta_p = m_p - \operatorname{rk} M_p - \operatorname{rk} M_{p+1} \quad .$$

Vediamo che alla fine il tutto si riduce al calcolo del rango delle matrici M_p e M_{p+1} , facilmente ricavabile con il metodo di eliminazione di Gauss.

Passiamo all'applicazione nel calcolo.

Consideriamo, ad esempio, i seguenti grafi G_2 , G_3 e G_4 , con il relativo complesso di cricche associato:



Esaminiamo $X(G_2)$ e $X(G_3)$. In questi casi il problema si riduce a calcolare β_0 e β_1 . Essendo per entrambi $\beta_0 = 1$, in quanto pari al numero di componenti connesse, possiamo utilizzare il Teorema 1.3.4 per calcolare β_1 . Applichiamo la formula, esplicitando β_1 , a $X(G_2)$:

$$\beta_1 = -s_0 + s_1 + \beta_0 = -5 + 5 + 1 = 1 \quad ,$$

e ora a $X(G_3)$:

$$\beta_1 = -s_0 + s_1 + \beta_0 = -4 + 3 + 1 = 0 \quad .$$

Otteniamo

$$H_0(C(X(G_2), \mathbb{Z}_2)) = H_0(C(X(G_3), \mathbb{Z}_2)) = \mathbb{Z}_2 \quad ,$$

$$H_1(C(X(G_2), \mathbb{Z}_2)) = \mathbb{Z}_2 \quad e \quad H_1(C(X(G_3), \mathbb{Z}_2)) = 0 \quad .$$

Esaminiamo ora $X(G_4)$. Anche in questo caso si ha $\beta_0 = 1$. Diversamente dagli esempi precedenti però, dobbiamo determinare due valori: β_1 e β_2 . Possiamo procedere nel modo seguente:

dall'*algoritmo di riduzione* si ha che

$$\beta_1 = m_1 - rkM_1 - rkM_2, \quad \beta_2 = m_2 - rkM_2 - rkM_3.$$

Dato che il calcolo del rango di quattro matrici non è un'operazione troppo *economica*, potremmo ridurci a calcolare uno dei due valori di β_i con il metodo di riduzione e ricavare l'altro utilizzando il Teorema 1.3.4. Notiamo che, essendo $d_3 = 0$, si ha $rkM_3 = 0$, dunque converrà calcolare β_2 con il metodo di riduzione, piuttosto che β_1 , il quale richiederebbe il calcolo del rango di due matrici.

Dobbiamo innanzitutto costruire la matrice M_2 : scriviamo gli elementi delle basi di $C_2(X(G_4), \mathbb{Z}_2)$ e $C_1(X(G_4), \mathbb{Z}_2)$, ponendoli in ordine lessicografico:

$$C_2(X(G_4), \mathbb{Z}_2) = \{(1, 2, 3), (4, 5, 6), (7, 8, 9)\},$$

$$C_1(X(G_4), \mathbb{Z}_2) = \{(1, 2), (1, 3), (1, 4), (2, 3), (2, 5), (3, 6), (4, 5), (4, 6), (5, 6), (6, 7), (7, 8), (7, 9), (8, 9)\}.$$

Otteniamo la matrice

$$M_2 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix},$$

dove $rkM_2 = 3$. Risulta dunque

$$\beta_2 = 3 - 3 = 0,$$

e applicando la formula di Eulero-Poincaré:

$$\beta_1 = -s_0 + s_1 - s_2 + \beta_0 + \beta_2 = -6 + 9 - 2 + 1 + 0 = 2.$$

In generale, quest'ultimo metodo ha un costo abbastanza elevato, pari a $O(n^3)$, con n numero di simplessi della base. Data la grandezza delle strutture che nella pratica vengono studiate, è più comodo poter fare riferimento ad un metodo meno *costoso*. Quello che andremo a descrivere necessita dell'introduzione di una nuova struttura.

Insiemi simpliciali

Gli insiemi simpliciali generalizzano i complessi simpliciali astratti rendendoli spazi topologici. Intuitivamente, un insieme simpliciale è come un complesso simpliciale astratto in cui i simplessi possono *collassare* in un punto e due o più vertici distinti identificati in uno solo. Poiché il nostro obiettivo è dare le nozioni sufficienti ad utilizzare il terzo metodo di calcolo, rimandiamo a [Zom10] per la formalizzazione completa.

Definizione 1.15. Un **insieme simpliciale** \mathbf{X} è una collezione di insiemi $\{X_p\}_{p \in \mathbb{Z}_{>0}}$ e morfismi

$$D_j : X_p \longrightarrow X_{p-1} \quad S_j : X_p \longrightarrow X_{p+1},$$

in cui, per $0 \leq i < p$, sono soddisfatte le seguenti uguaglianze:

$$\begin{aligned} D_i D_j &= D_{j-1} D_i, \quad \text{se } i < j, \\ S_i S_j &= S_{j+1} S_i, \quad \text{se } i \leq j, \\ D_i D_j &= \begin{cases} S_{j-1} D_i, & \text{se } i < j, \\ Id_{X_n}, & \text{se } i = j \text{ o } i = j + 1, \\ S_j D_{i-1}, & \text{se } i > j + 1. \end{cases} \end{aligned}$$

Un elemento di X_p è un p -simpleso orientato, D_j è detto j -esimo *operatore di faccia* (che elimina il j -esimo vertice nel semplice) e S_j j -esimo *operatore di degenerazione* (che pone accanto al vertice i_j la sua copia). Formalmente

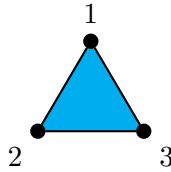
$$\begin{aligned} D_j((i_0, \dots, i_p)) &= (i_0, \dots, \widehat{i_j}, \dots, i_p), \\ S_j((i_0, \dots, i_p)) &= ((i_0, \dots, i_j, i_j, \dots, i_p)). \end{aligned}$$

Possiamo costruire ricorsivamente l'insieme simpliciale X associato ad un dato complesso simpliciale astratto K , in particolare considereremo K complesso di cricche:

$$\begin{aligned} X_0 &= K_0, \\ X_p &= \tilde{K}_p \cup \left(\bigcup_{i=0}^p S_i(X_{p-1}) \right), \quad \text{con } n > 0, \end{aligned}$$

con $\tilde{K}_p = K^{(p)} - K^{(p-1)}$, ovvero l'insieme che contiene tutti gli elementi di K_p tali che ogni elemento abbia vertici tutti distinti, ovvero i p -simplessi.

Esempio 1.3.5. Consideriamo il seguente complesso di cricche K



L'insieme simpliciale associato X è dato da

$$\begin{aligned} X_0 &= \{1, 2, 3\}, \\ X_1 &= \{(1, 2), (1, 3), (2, 3), (1, 1), (2, 2), (3, 3)\}, \\ X_2 &= \{(1, 2, 3), (1, 1, 2), (1, 2, 2), (2, 2, 3), (2, 3, 3), (1, 1, 3), (1, 3, 3), (1, 1, 1), (2, 2, 2), (3, 3, 3)\}. \end{aligned}$$

Il nostro obiettivo è semplificare il complesso di cricche da cui vogliamo partire, in modo da ridurre la mole di calcoli per determinare il gruppo di omologia, lasciando inalterate le proprietà topologiche della struttura. In altre parole, lo rendiamo minimale rispetto a quest'ultime. Queste *operazioni* sono:

- la **TRIM** (o *potatura*), in cui eliminiamo le *foglie*;
- la **THIN** (o *assottigliamento*), in cui per ogni cricca massimale effettuiamo il *collasso*.

Definizione 1.16. Dato X insieme simpliciale, diremo che un semplice $s \in X$ è una **foglia** se per ogni intero k vale

$$\frac{H_k(\bar{s})}{H_k(\bar{s} \cap (X - \bar{s}))} = 0 .$$

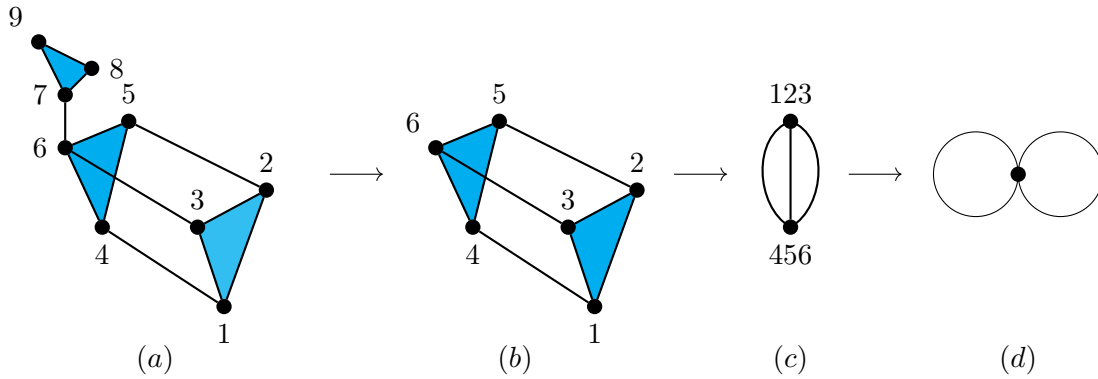
Definizione 1.17. Dato un insieme simpliciale X e $s \in X$ p -simpleso, far **collassare** s significa identificarlo in un singolo punto. Questo ci restituisce un nuovo insieme simpliciale

$$X' = X/s .$$

È possibile dimostrare che l'utilizzo di queste operazioni lascia inalterate le proprietà omologiche del complesso (vedi Teorema 2, 3 in [Zom10]). Al termine di questo lavoro di *riduzione* otteniamo un insieme **tidy**, ovvero un insieme *ripulito*.

Definizione 1.18. Un insieme **tidy** è un complesso simpliciale astratto *potato* e *assottigliato*.

Applichiamo queste operazioni all'insieme simpliciale associato a $X(G_4)$:



Dove in $(a) \rightarrow (b)$ avviene l'eliminazione della foglia $(6, 7, 8, 9)$, in $(b) \rightarrow (c)$ facciamo collassare i semplici (cricche massimali) $(1, 2, 3)$ e $(4, 5, 6)$, ottenendo i vertici (123) e (456) rispettivamente. A questo punto i semplici di dimensione maggiore che sono rimasti sono tre, aventi tutti quanti gli stessi estremi $(123, 456)$, ne collassiamo uno ottenendo (d) , il nostro insieme tidy. Avremo dunque che

$$H_0(C(X(G_4), \mathbb{Z}_2)) = \mathbb{Z}_2 , \quad H_1(C(X(G_4), \mathbb{Z}_2)) = \mathbb{Z}_2^2 \quad e \quad H_2(C(X(G_4), \mathbb{Z}_2)) = 0 ,$$

coerentemente con quanto ottenuto tramite il metodo di riduzione.

Capitolo 2

Omologia persistente

2.1 Costruzione dell'omologia persistente

In questa sezione ci occuperemo di dare le nozioni di partenza per costruire l'omologia persistente definita sui complessi simpliciali astratti. Forniremo, inoltre, una descrizione della struttura utilizzando il Teorema 1.1.3, e da qui una *giustificazione* al metodo di rappresentazione grafica trattato nella sezione successiva.

Le argomentazioni di questa sezione sono tratte da [ZC05], [EH08].

Definizione 2.1. Dato un complesso simpliciale astratto K , una **filtrazione** di K è la sequenza di sottocomplessi

$$\emptyset = K^0 \subseteq K^1 \subseteq \dots \subseteq K^n = K ,$$

dove $K^i \subseteq K^j$, per ogni $i < j$ e $K^i = 0$, per ogni $i \leq 0$ e $i > n$ per qualche intero n . In questo caso, chiameremo K **complesso filtrato**.

Per la proprietà dei complessi simpliciali astratti di chiusura rispetto all'inclusione, si ha che K^i è un complesso simpliciale astratto per ogni i . Dunque, ad esso possiamo associare un operatore di bordo $d_k^i: C_k(K^i, R) \rightarrow C_{k-1}(K^i, R)$ e gli R -moduli

$$C_k^i := C_k(K^i, R), \quad Z_k^i := Z_k(K^i, R), \quad B_k^i := B_k(K^i, R) \quad e \quad H_k^i := H_k(C(K^i, R)), \quad \forall i, k \geq 0 ,$$

dove, come si può notare, l'apice i indica l'indice di filtrazione. Inoltre, all'inclusione $K^i \subseteq K^j$ è associato il morfismo simpliciale $\iota: K^i \rightarrow K^j$, per ogni $i < j$.

Definizione 2.2. Dato K complesso filtrato, il k -esimo modulo di omologia p -persistente di K^i , con $p \geq 0$, è dato da

$$H_k^{i,j} = \frac{Z_k^i}{(B_k^j \cap Z_k^i)},$$

dove $p = j - i$. Il rango di $H_k^{i,j}$ è il k -esimo **numero di Betti p -persistente**, indicato con $\beta_k^{i,j}$. In particolare, considerando $R = F$, con F campo qualunque, i suddetti R -moduli diventano spazi vettoriali, e dunque risulterà $\beta_k^{i,j} = \dim H_k^{i,j}$

Notiamo che tale definizione è ben posta, infatti $Z_k^i \subseteq Z_k^j$ sono sottomoduli di C_k^j , per ogni $j \geq i$, e lo stesso vale per l'intersezione, che risulterà un sottomodulo del denominatore.

Diremo che una classe di omologia e ¹ nasce in K^i se non appartiene all'immagine del morfismo indotto dall'inclusione $K^{i-1} \subset K^i$, e muore in K^j se l'immagine del morfismo

¹Per semplicità notazionale scriveremo e al posto di $[e]$ per indicare un elemento di $H_k^{i,j}$.

indotto da $K^{i-1} \subset K^{j-1}$ non contiene l'immagine di e , che invece è contenuta nell'immagine del morfismo indotto da $K^{i-1} \subset K^j$. In tal caso diremo che la **persistenza** di e è $p = j - i$. Per rimanere coerenti con la terminologia, possiamo considerare la persistenza p come la *durata di vita* della classe $e \in H_*^{i,j}$.

Definizione 2.3. Un **complesso di persistenza** \mathcal{C} è una famiglia di complessi di catene $\{C_*^i\}_{i \geq 0}$ su R , con i morfismi di catene $f_*^i : C_*^i \rightarrow C_*^{i+1}$, tali da avere il seguente diagramma

$$\begin{array}{ccccccc}
 & & \vdots & & \vdots & & \vdots \\
 & & d_{p+2}^i \downarrow & & d_{p+2}^{i+1} \downarrow & & d_{p+2}^{i+2} \downarrow \\
 \dots & \xrightarrow{f_{p+1}^{i-1}} & C_{p+1}^i & \xrightarrow{f_{p+1}^i} & C_{p+1}^{i+1} & \xrightarrow{f_{p+1}^{i+1}} & C_{p+1}^{i+2} \xrightarrow{f_{p+1}^{i+2}} \dots \\
 & & d_{p+1}^i \downarrow & & d_{p+1}^{i+1} \downarrow & & d_{p+1}^{i+2} \downarrow \\
 \dots & \xrightarrow{f_p^{i-1}} & C_p^i & \xrightarrow{f_p^i} & C_p^{i+1} & \xrightarrow{f_p^{i+1}} & C_p^{i+2} \xrightarrow{f_p^{i+2}} \dots \\
 & & d_p^i \downarrow & & d_p^{i+1} \downarrow & & d_p^{i+2} \downarrow \\
 \dots & \xrightarrow{f_{p-1}^{i-1}} & C_{p-1}^i & \xrightarrow{f_{p-1}^i} & C_{p-1}^{i+1} & \xrightarrow{f_{p-1}^{i+1}} & C_{p-1}^{i+2} \xrightarrow{f_{p-1}^{i+2}} \dots \\
 & & d_{p-1}^i \downarrow & & d_{p-1}^{i+1} \downarrow & & d_{p-1}^{i+2} \downarrow \\
 & & \vdots & & \vdots & & \vdots
 \end{array}$$

Denoteremo con \mathcal{K} il complesso di persistenza associato al complesso filtrato K insieme ai morfismi di inclusione $f_*^i : C_*^i(K, R) \rightarrow C_*^{i+1}(K, R)$ indotti da $\iota : K^i \rightarrow K^{i+1}$. In particolare, il morfismo $\iota : K^i \rightarrow K^j$ induce il morfismo $f_*^{i,j} : C_*^i(K, R) \rightarrow C_*^j(K, R)$.

Definizione 2.4. Un **modulo di persistenza** \mathcal{M} è una collezione $\{M^i, \phi^i\}_i$ data dalla famiglia di R -moduli M^i insieme agli omomorfismi $\phi^i : M^i \rightarrow M^{i+1}$.

In particolare, risulta che l'omologia di un complesso di persistenza è un modulo di persistenza, dove gli omomorfismi sono dati da $\phi_*^i : H_*^i \rightarrow H_*^{i+1}$. Indicheremo tale oggetto con $\mathcal{H} = \{H_*^i, \phi_*^i\}$. In particolare definiamo $\phi_*^{i,j} : H_*^i \rightarrow H_*^j$ come l'omomorfismo dato dalla composizione $\phi_*^j \phi_*^{j-1} \dots \phi_*^{i+1} \phi_*^i$.

Definizione 2.5. Diremo che un complesso di persistenza $\{C_*^i, f_*^i\}$ (modulo di persistenza $\{M^i, \phi^i\}$) è di **tipo finito** se ogni componente del complesso (modulo) è un R -modulo finitamente generato, e se i morfismi $f_*^i(\phi^i)$ sono isomorfismi per $i \geq n$, per qualche intero n .

Essendo il nostro complesso simpliciale astratto K finito, segue che anche il complesso di persistenza associato \mathcal{K} è di tipo finito (avendo un numero finito di sottocomplessi nella filtrazione), così come il modulo di persistenza \mathcal{H} . D'ora in poi, quando parleremo di moduli di persistenza, li considereremo sempre di tipo finito.

Prendiamo ora il modulo di persistenza $\mathcal{M} = \{M^i, \phi^i\}$, con M^i R -moduli finitamente generati, e l'anello dei polinomi $R[t]$ graduato non negativamente tramite la graduazione standard, ovvero tale che $(t^n) = Rt^n$, per ogni $n \geq 0$. Vogliamo far vedere che esiste un'equivalenza tra la categoria dei moduli di persistenza di tipo finito su R , indicata con \mathbf{P} , e la categoria dei moduli graduati non negativamente finitamente generati su $R[t]$, indicata con \mathbf{M} .

Definiamo il funtore $\alpha: \mathbf{P} \rightarrow \mathbf{M}$ nel modo seguente:

$$\alpha(\mathcal{M}) = \bigoplus_{i=0}^{\infty} M^i,$$

dove la struttura di R -modulo è data dalla somma delle singole componenti, nulle quasi ovunque, e l'azione di t è definita come

$$t \cdot (m^0, m^1, m^2, \dots) = (0, \phi^0(m^0), \phi^1(m^1), \phi^2(m^2), \dots),$$

ovvero t è la funzione che trasla ciascuna componente nella componente successiva nell'indice di graduazione i . In particolare, $t^n M^m \subseteq M^{n+m}$, per ogni intero $n, m \geq 0$. Tale α è anche detto *corrispondenza*.

Teorema 2.1.1 (di corrispondenza – Zomorodian -Carlsson).

La corrispondenza $\alpha: \mathbf{P} \rightarrow \mathbf{M}$ è un'equivalenza di categorie.

La dimostrazione è la teoria di Artin-Rees nell'algebra commutativa.

In questo modo possiamo vedere il modulo di persistenza \mathcal{H} come un modulo graduato non negativamente su $R[t]$, con gli H_*^i R -moduli finitamente generati. Nel caso di $R = F$ campo, l'anello graduato $F[t]$ è un dominio a ideali principali, i cui ideali sono della forma (t^n) , per ogni $n \geq 0$. Dunque ci troviamo nelle ipotesi del punto (2) del Teorema 1.1.3, per cui, considerando $\mathcal{M} = \mathcal{H}$, risulterà

$$\alpha(\mathcal{H}) = \left(\bigoplus_{i=1}^p (t^{\alpha_i}) \right) \oplus \left(\bigoplus_{j=1}^q (t^{\beta_j}) / (t^{\gamma_j}) \right),$$

dove $\alpha_i, \beta_j, \gamma_j$ sono potenze di interi non negative e $p, q > 0$.

Possiamo vedere i termini del primo addendo come i generatori delle classi di omologia che *nascono* all'indice di filtrazione α_i e permangono per tutta la sua durata (dette *classi essenziali*), mentre gli elementi del secondo addendo sono quei generatori che appaiono al momento β_j e *muoiono* al momento γ_j della filtrazione. Per rendere più rigorosa questa parametrizzazione, definiamo le cosiddette *coppie di persistenza*.

Definizione 2.6. Un \mathcal{P} -intervallo, o **coppia di persistenza**, è una coppia ordinata $(i, j) \in \mathbb{Z} \times \mathbb{Z}^\infty$ dove $0 \leq i < j$ e $\mathbb{Z}^\infty = \mathbb{Z} \cup \{+\infty\}$.

Detto $\mathcal{S} = \{(i_1, j_1), (i_2, j_2), \dots, (i_n, j_n)\}$ il multinsieme² delle coppie di persistenza, possiamo definire una corrispondenza biunivoca $\beta: \mathcal{S} \rightarrow \mathbf{M}$ nel modo seguente:

$$\beta(i, j) = \begin{cases} (t^i)/(t^j) & \text{se } j \in \mathbb{Z}, \\ (t^i) & \text{se } j = +\infty. \end{cases}$$

Otteniamo che

$$\beta(\mathcal{S}) = \bigoplus_{l=1}^n \beta(i_l, j_l).$$

Corollario 2.1.2. *La corrispondenza β definisce una biiezione tra gli insiemi finiti di \mathcal{P} -intervalli e i moduli graduati finitamente generati su $F[t]$. Di conseguenza, le classi di isomorfismo dei moduli di persistenza di tipo finito su F sono in corrispondenza biunivoca con gli insiemi finiti di \mathcal{P} -intervalli.*

² \mathcal{S} è un *multinsieme* in quanto ci possono essere classi di omologia che nascono e muoiono simultaneamente, quindi tali che esista la coppia (i, j) con molteplicità maggiore di 1, per qualche $0 \leq l \leq n$.

Dunque possiamo definire una base del modulo di persistenza \mathcal{H} utilizzando come suoi elementi le coppie di persistenza. Risulterà che alla coppia $(i, j) \in \mathcal{S}$ è associata la classe di omologia $e \in H_*^{i,j}$.

2.2 Visualizzazione

Data la corrispondenza appena enunciata, possiamo rappresentare graficamente l'omologia di persistenza utilizzando i \mathcal{P} -intervalli. Tra i metodi di visualizzazione abbiamo il *barcode*, il *diagramma di persistenza* e il *panorama di persistenza*. Quest'ultima è la rappresentazione sviluppata più di recente, che si può vedere come una fusione delle prime due. In questo scritto ci occuperemo di descrivere il barcode e il diagramma di persistenza. La descrizione del primo tratta da [Ghr08] e del secondo da [EH08].

2.2.1 Barcode

Il **barcode** è una rappresentazione grafica di un modulo di persistenza dato \mathcal{M} in cui le coppie di persistenza (i, j) (i.e., gli elementi della base) sono viste come una collezione di segmenti orizzontali posti in un piano avente per ascissa i parametri della filtrazione e per ordinata i generatori dell'omologia posti in ordine arbitrario. Formalmente

Definizione 2.7. Il **barcode** corrispondente ad un modulo di persistenza dato \mathcal{M} è una collezione di \mathcal{P} -intervalli, che possiamo vedere come sottoinsiemi di $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ ³, definito nel modo seguente

$$Bar(\mathcal{M}) := \{(i, j) \in \overline{\mathbb{R}} \mid (i, j) \in \mathcal{S} \text{ con } \beta(\mathcal{S}) \simeq \alpha(\mathcal{M})\},$$

dove α e β sono le corrispondenze definite nella sezione precedente.

Il barcode si rivela utile per distinguere le classi di omologia rilevanti da quelle che, invece, rappresentano solo un *rumore*. In particolare, le prime saranno date da linee più *lunghe*, che corrispondono a quelle classi di omologia con *durata di vita* (i.e., persistenza) maggiore.

2.2.2 Diagramma di persistenza

Questa è solitamente la rappresentazione più utilizzata. Diversamente dal barcode, qui il piano di rappresentazione ha ascissa data dall'indice di *nascita* i e le ordinate dall'indice di *morte* j . Essendo per definizione $i \leq j$, avremo che le coppie di persistenza associate ad un certo modulo di omologia si trovano tutte al di sopra, al più coincidenti, della bisettrice del I e III quadrante. Abbiamo un diagramma di persistenza per ogni dimensione k dei generatori del modulo di omologia, indicato con $Dgm_k(\mathcal{H})$. In generale

Definizione 2.8. Dato un modulo di persistenza \mathcal{M} definiamo il suo **diagramma di persistenza** come segue

$$Dgm(\mathcal{M}) := \{(i, j) \in \overline{\mathbb{R}}^2 \mid (i, j) \in Bar(\mathcal{M})\}.$$

Fissato un intero $k \geq 0$, dal diagramma di persistenza $Dgm_k(\mathcal{H})$ possiamo determinare $\beta_k^{i,j}$, ovvero il numero di elementi della base di $H_k^{i,j}$. Gli elementi appartenenti a $H_k^{i,j}$ sono tutte quelle classi di omologia nate prima di K^i e ancora vive in K^j . Data la corrispondenza biunivoca tra le classi di omologia e le coppie di persistenza, nel diagramma di persistenza

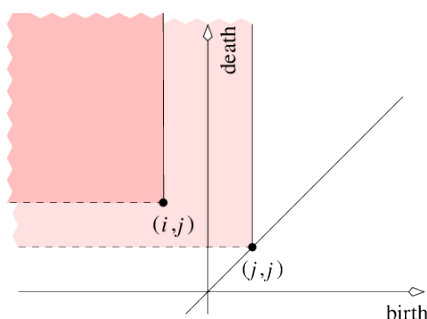


Figura 2.1: Diagramma di persistenza con evidenziato di un colore più scuro il quadrante generato dalla coppia (i, j) . [EH08]

ciò si traduce nel contare tutte le coppie che si trovano nel quadrante $(-\infty, i] \times (j, +\infty)$ (Fig.2.1).

Nel caso $i = j$ otteniamo β_k^j , ovvero il numero di Betti associato al modulo di omologia di K^j . Con questa rappresentazione la *persistenza* di una coppia è data dalla distanza (lungo l'asse verticale) tra la coppia e la bisettrice.

2.3 Stabilità

L'importanza di usare le coppie di persistenza come descrittori dell'omologia è data dalla *stabilità* dei risultati sotto lievi perturbazioni. Tale stabilità è stata dimostrata in relazione a tre tipi di distanza: la distanza *bottleneck* [CSEH05], [CDGO13], [EH08] (nota inizialmente come distanza *matching*), la distanza di *Hausdorff* [CSEH05] e la distanza *Wasserstein* [EM], [EH08], quest'ultima caso particolare della prima. In [CSEH05] e [EH08] la stabilità è dimostrata considerando diagrammi di persistenza relativi a *funzioni definite su spazi topologici X a valori reali*, dove la filtrazione è data dai *sottolivelli* di $f: X \rightarrow \mathbb{R}$ al variare di un parametro $a \in \mathbb{R}$. Mentre in [CDGO13] vengono considerati i diagrammi di persistenza relativi a moduli di persistenza, su cui viene definita una particolare distanza, detta *distanza di interleaving*. I concetti trattati nel seguito sono principalmente estratti da quest'ultimo. In questa sezione riporteremo gli argomenti significativi per enunciare il **Teorema di stabilità** in relazione alla distanza bottleneck.

Iniziamo col definire questo tipo di distanza

Definizione 2.9. Dati \mathcal{M}, \mathcal{N} moduli di persistenza e $\Gamma := \{\gamma: Dgm(\mathcal{M}) \rightarrow Dgm(\mathcal{N}) \mid \gamma \text{ biiezione tra sottoinsiemi di } \mathbb{R}^2\}$, definiamo la **distanza bottleneck** come la distanza fra diagrammi di persistenza data da

$$d_b(Dgm(\mathcal{M}), Dgm(\mathcal{N})) = \inf_{\gamma \in \Gamma} \sup_{x \in Dgm(\mathcal{M})} \|x - \gamma(x)\|_\infty,$$

dove $\|\cdot\|_\infty$ indica la distanza L_∞ tra punti in \mathbb{R}^2 ; dunque ponendo $x := (i, j) \in Dgm(\mathcal{M})$ e $\gamma(x) := (i', j') \in Dgm(\mathcal{N})$, otteniamo $\|x - \gamma(x)\|_\infty := \max\{|i - i'|, |j - j'|\}$.

Tale definizione si può estendere a coppie in $\overline{\mathbb{R}}^2$ dove, nel caso in cui $j, j' = +\infty$ (i.e., x e $\gamma(x)$ sono classi essenziali), poniamo per convenzione $j - j' = 0$.

Osservazione 2. Notiamo che, senza ipotesi aggiuntive, non è detto che esista una biiezione γ tra due diagrammi di persistenza dati, in quanto questi possono differire per il numero di

³Dalla definizione 2.6 è ammesso che $j = +\infty$.

elementi [EH08]. Per ovviare a questo problema, consideriamo come punti del diagramma anche tutti i punti che si trovano sulla diagonale D , dove $D := \{(i, j) \in \mathbb{R}^2 \mid i = j\}$. In questo modo, i punti x che non hanno un corrispettivo tramite γ fuori dalla diagonale, avranno assegnato il punto $\gamma(x) \in D$ tale da rispettare la distanza bottleneck. Ovvero, detti $c := (i, i)$ gli elementi della diagonale, si ha $\gamma(x) = y$, dove $y \in D$ tale che $d(x, y) = \min_{c \in D} \{d(x, c)\}$, con d distanza euclidea (vedi Fig.2.2).

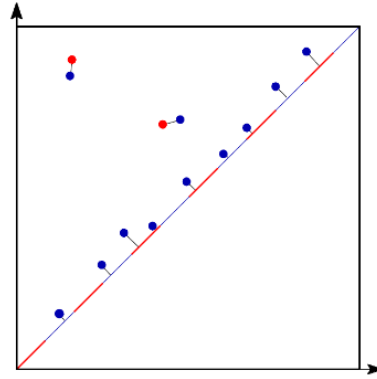


Figura 2.2: Distanza bottleneck dati due diagrammi di persistenza. I punti aventi stesso colore sono elementi dello stesso diagramma. La lunghezza del segmento tra due punti rappresenta il valore della distanza bottleneck, da cui l'associazione tramite γ . Notiamo che i punti blu più vicini alla diagonale sono in corrispondenza con i punti di quest'ultima che gli sono più vicini, in linea con quanto detto nell'Osservazione 2. [God]

Definizione 2.10. Diremo che un modulo di persistenza $\mathcal{M} = \{M^i, \phi^i\}$ è **q-tame** se

$$rk(\phi^{i,j}) < \infty, \quad \forall i < j.$$

In particolare, prendendo come modulo di persistenza \mathcal{H} , questo è equivalente a dire che $\beta_*^{i,j} < \infty$, per ogni $i < j$.

Osservazione 3. Facciamo notare che, se un modulo di persistenza è di tipo finito, risulterà che esso è anche q-tame.

Definizione 2.11. Dati due moduli di persistenza $\mathcal{M} = \{M^i, \phi^i\}$, $\mathcal{N} = \{N^i, \psi^i\}$ e un valore reale $\epsilon > 0$, diremo che \mathcal{M} e \mathcal{N} sono **ϵ -interleaved** se esistono gli omomorfismi di grado ϵ

$$\mu_i: M^i \rightarrow N^{i+\epsilon} \quad e \quad \nu_i: N^i \rightarrow M^{i+\epsilon},$$

tali che i seguenti diagrammi commutino per ogni $i < j$:

$$\begin{array}{ccccc} & & M^i & \xrightarrow{\phi^{i,j}} & M^j \\ & \nearrow \nu_{i-\epsilon} & \downarrow \mu_i & & \downarrow \mu_j \\ N^{i-\epsilon} & \xrightarrow{\psi^{i-\epsilon, i+\epsilon}} & N^{i+\epsilon} & \xrightarrow{\psi^{i+\epsilon, j+\epsilon}} & N^{j+\epsilon} \end{array},$$

$$\begin{array}{ccccc} M^{i-\epsilon} & \xrightarrow{\phi^{i-\epsilon, i+\epsilon}} & M^{i+\epsilon} & \xrightarrow{\phi^{i+\epsilon, j+\epsilon}} & M^{j+\epsilon} \\ & \searrow \mu_{i-\epsilon} \nu_i & \uparrow & & \uparrow \nu_j \\ & & N^i & \xrightarrow{\psi^{i,j}} & N^j \end{array}.$$

Osservazione 4. Presi K, L complessi filtrati ϵ -interleaved, allora si ha che

$$K^{i-\epsilon} \subseteq L^i \subseteq K^{i+\epsilon} .$$

In particolare, anche i moduli di persistenza associati $\mathcal{H}, \mathcal{H}'$ (rispettivamente) saranno ϵ -interleaved, secondo la Definizione 2.11.

Definizione 2.12. Definiamo la **distanza di interleaving** tra due moduli di persistenza \mathcal{M}, \mathcal{N} nel modo seguente

$$d_i(\mathcal{M}, \mathcal{N}) = \inf\{\epsilon \mid \mathcal{M}, \mathcal{N} \text{ sono } \epsilon\text{-interleaved}\} .$$

Osservazione 5. Notiamo che la distanza di interleaving è, in realtà, una *pseudodistanza*, in quanto è verificata sia la simmetria che la disuguaglianza triangolare (vedi [CDGO, Proposizione 4.3]), ma non è necessariamente vero che $\mathcal{M} \simeq \mathcal{N}$ se $d_i(\mathcal{M}, \mathcal{N}) = 0$.

Siamo ora pronti a enunciare il risultato centrale di questa sezione:

Teorema 2.3.1 (di stabilità). *Siano \mathcal{M}, \mathcal{N} moduli di persistenza q -tame, allora*

$$d_i(\mathcal{M}, \mathcal{N}) = d_b(Dgm(\mathcal{M}), Dgm(\mathcal{N})) .$$

Per la dimostrazione si fa riferimento a [CDGO13, Teorema 4.11].

In altre parole, questo teorema ci dice che se due moduli di persistenza differiscono di una quantità ϵ abbastanza piccola, allora anche le coppie di persistenza, prese come punti di $\overline{\mathbb{R}}^2$, differiranno di tale ϵ nella distanza bottleneck. Questo fatto ci sarà utile nel seguito sia per giustificare la validità di questa rappresentazione (soprattutto nel caso pratico, in cui i valori raccolti possono discostarsi lievemente da quelli effettivi) sia per costruire il complesso filtrato su cui definiamo l'omologia persistente; quest'ultimo punto sarà chiarito nella prossima sezione.

2.4 Point cloud e costruzione del complesso filtrato

Abbiamo visto che, data una filtrazione (e.g., di complessi simpliciali astratti), possiamo definire l'omologia persistente relativa al complesso di persistenza associato al complesso filtrato. In questa sezione vedremo come costruire una filtrazione in modo da ottenere un complesso simpliciale astratto. In letteratura, sono riportati più modi per ricavare una filtrazione, noi utilizzeremo quello che vede come insieme di partenza la *point cloud*.

Nell'applicazione, un insieme di dati può essere identificato come un insieme di punti immerso in uno spazio metrico (i.e., *point cloud*). La branca emergente che si occupa dello studio delle caratteristiche topologiche di queste strutture è l'analisi topologica (*Topological Data Analysis*). Lo studio, in generale, segue due passaggi:

- per prima cosa si approssima l'insieme di punti ad una struttura combinatoria, nel nostro caso un complesso simpliciale astratto;
- successivamente si utilizzano strumenti di topologia algebrica per calcolare gli invarianti topologici della struttura ottenuta al passo precedente, che per noi sarà l'omologia di persistenza.

Per costruire un complesso simpliciale a partire da un insieme finito di punti esistono numerose tecniche (in [EH10] ne sono definite varie); in questo scritto ne vedremo due.

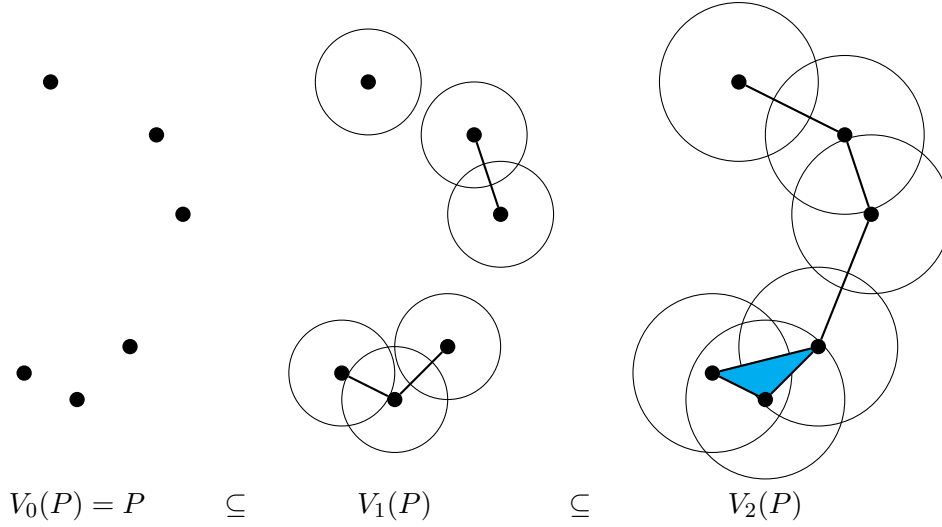
La prima che trattiamo è una delle più utilizzate in TDA, grazie al costo computazionale ridotto rispetto alle altre nel caso in cui vogliamo lavorare su dimensioni maggiori di 3 [Zom].

Definizione 2.13. Dato un insieme finito di punti $P \subset \mathbb{R}^n$ e un valore reale $r > 0$, il **complesso di Vietoris-Rips** $V_r(P)$ è il complesso simpliciale astratto definito nella maniera seguente:

$$V_r(P) = \{s \subseteq P \mid d(x, y) < r, \forall x, y \in s\},$$

dove d è la distanza euclidea.

Esempio 2.4.1. Vediamo un esempio di filtrazione del complesso di Vietoris-Rips a partire dall'insieme di punti P :



In particolare, il complesso $V_r(P)$ è un complesso di cricche. Per questo motivo, il costo di *definizione* (i.e., elenco delle componenti della struttura, ovvero dei suoi semplici) è minore rispetto a quello di un generico complesso simpliciale astratto: essendo il primo univocamente determinato dal suo 1-scheletro è sufficiente elencare gli elementi di quest'ultimo, mentre nel secondo caso dovremmo elencare anche tutti i semplici di dimensione superiore (e.g., come avviene nel complesso di Čech). Notiamo, inoltre, che $V_s(P) \subseteq V_t(P)$ se $s < t$. In questo modo possiamo definire ciascun $V_r(P)$ come elemento di una *filtrazione*, dove chiamiamo $V(P) := \{V_r(P)\}_{r \in \mathbb{R}_{>0}}$ il complesso filtrato.

Per la seconda costruzione che trattiamo dobbiamo fare un piccolo passo indietro e tornare al concetto di grafo.

Definizione 2.14. Dato un grafo G , diremo che G è un **grafo pesato** se esiste un'applicazione $w: E(G) \rightarrow \mathbb{R}_{>0}$ tale che ad ogni lato di G associa un valore reale positivo detto *peso*.

In particolare, $V(G) := P$, ovvero vediamo l'insieme dei vertici del grafo come un insieme di punti in uno spazio metrico. Imponendo che w sia iniettiva⁴ (i.e., i lati hanno tutti pesi distinti), possiamo costruire una filtrazione del complesso $X(G) := \{X(G_i)\}_{i=0}^{|E(G)|}$, dove

$$G_0 := V(G),$$

$$G_1 := G_0 \cup e_1, \text{ con } e_1 := w^{-1}(\max_{e \in E(G)} \{w(e)\})^5,$$

$$G_2 := G_1 \cup e_2, \text{ con } e_2 := w^{-1}(\max_{e \in E(G) \setminus e_1} \{w(e)\}),$$

⁴Grazie alla stabilità è sempre possibile considerare una tale w , senza perdita di generalità.

⁵Avendo preso w iniettiva, tale definizione è ben posta.

⋮

$G_i := G_{i-1} \cup e_i$, con $e_i := w^{-1}(\max_{e \in E(G) \setminus \{e_1, \dots, e_{i-1}\}} \{w(e)\})$.

In altre parole, ogni momento i della filtrazione è scandito dall'aggiunta di uno e un solo lato di G , che coincide con quello di peso massimo (o minimo, in base a cosa vogliamo dare la precedenza) tra quelli in $E(G)$ che sono rimasti da aggiungere.

Segue dunque dalla definizione che $G_i \subseteq G_j$ per ogni $i < j$, da cui otteniamo la filtrazione

$$X(G_0) \subseteq X(G_1) \subseteq \dots \subseteq X(G_i) \subseteq X(G_{|E(G)|}) = X(G) .$$

Notiamo che l'indice i corrisponde al numero di lati aggiunti fino al momento i della filtrazione. In questo modo possiamo scegliere di associare all'indice i la densità dei lati nel sottografo G_i , ovvero ponendo $\rho = \frac{i}{|E(G)|}$. In particolare, data la coppia di persistenza (i, j) consideriamo equivalentemente la coppia $(\rho_{\text{birth}}, \rho_{\text{death}})$, dove $\rho_{\text{birth}} = \frac{i}{|E(G)|}$ e $\rho_{\text{death}} = \frac{j}{|E(G)|}$.

Utilizzando quest'ultima costruzione, possiamo ricondurci al complesso di Vietoris-Rips: ponendo $w: V(G) \times V(G) \rightarrow \mathbb{R}_{>0}$ come $w((x, y)) = d((x, y))$ per ogni $x, y \in V(G)$, dove $d((x, y))$ indica la distanza euclidea tra i due vertici (i.e., elementi della point cloud), otteniamo una filtrazione che aggiunge, a partire dai vertici del grafo, tutti i lati in ordine *crescente* di peso (i.e., distanza), ovvero verranno aggiunti prima i lati che collegano punti più vicini tra loro. Rimane da risolvere il problema dell'iniettività di w : siccome alcune distanze possono coincidere, ci converrà renderle differenti in modo da definire l'ordine di aggiunta dei lati nella filtrazione; il Teorema 2.3.1 ci garantisce che, variando di poco questi valori, le variazioni sui diagrammi di persistenza saranno a loro volta un *rumore*, ovvero non percepite dall'omologia persistente. In questo modo, otteniamo un complesso filtrato che differisce da quello precedentemente definito in 2.13 dal solo fatto che, mentre in quest'ultima costruzione ogni complesso della filtrazione differisce dal successivo per l'aggiunta di uno e un solo lato, nel primo può avvenire che nel momento successivo della filtrazione si aggiungano, ad esempio, più lati o addirittura semplici di dimensione maggiore, così come potrebbe lasciarlo invariato; ai fini dell'omologia persistente queste differenze non hanno rilevanza.

Equivalentemente possiamo considerare

$$w((x, y)) = \frac{1}{d((x, y))} , \quad \forall x, y \in V(G) , \quad (2.1)$$

e aggiungere i lati nella filtrazione in ordine *decescente*.

Capitolo 3

Applicazione al connettoma

In questo capitolo ci occuperemo di applicare i concetti enunciati finora allo studio di una particolare struttura: il *connettoma*. Il termine, coniato nel 2005, si rifà al concetto di genoma: così come quest'ultimo indica la mappatura del DNA, il connettoma indica la mappatura del cervello e, più nello specifico, delle connessioni neurali.

I mattoncini di base di questa struttura complessa sono i *neuroni*, nel loro piccolo così semplici: essi consistono di un *corpo cellulare* e di *prolungamenti* che stabiliscono un *contatto* con gli altri neuroni, in una relazione di *contiguità*.

Al giorno d'oggi l'unico connettoma che è stato mappato in maniera completa, ogni singolo neurone e ogni singola connessione, è quello di un verme nematode lungo circa 1 mm: il *Caenorhabditis elegans* (o *C. elegans*). Questa struttura è stata vista consistere di circa 300 neuroni e 6000 connessioni. Per ottenere tale mappatura ci sono voluti oltre 12 anni di studi.

Nell'uomo è stato stimato un numero di circa 10^{11} neuroni e 10^{15} connessioni. Al momento la mappatura del connettoma umano è stata fatta per alcune piccole zone, fare quella dell'intera struttura è considerata una delle più grandi sfide tecnologiche del nostro tempo.

3.1 Nozioni preliminari sul cervello *umano*

Il cervello è l'organo principale del sistema nervoso centrale (SNC), formato da due regioni: il telencefalo, ovvero la porzione con maggiore estensione che contiene i due emisferi (speculari anatomicamente ma non funzionalmente), e il diencefalo, cioè la parte più interna totalmente rivestita dal telencefalo. Il SNC può essere suddiviso in due parti: l'encefalo, che comprende cervello e tronco encefalico, e il midollo spinale, che consiste in un fascio di neuroni che percorre l'intero canale vertebrale fino, nella maggior parte dei casi, alle prime vertebre lombari. Dal SNC si dirama il sistema nervoso periferico, costituito dai nervi che trasmettono i segnali dal SNC al resto del corpo (vie efferenti o motorie), e viceversa (vie afferenti o sensitive).

Il tessuto nervoso del SNC è costituito da due tipi di sostanze:

- la *sostanza grigia*, che comprende il corpo del neurone (detto soma), le ramificazioni dendritiche ("uscenti" da quest'ultimo) e le terminazioni assoniche ("entranti");
- la *sostanza bianca*, costituita dagli assoni dei neuroni (i.e., fibre nervose), che trasportano l'informazione (i.e., impulso elettrico) tra neuroni. Il nome è dovuto alla colorazione che assume a causa della mielina, la guaina isolante che lo riveste e permette all'impulso nervoso di propagarsi più rapidamente.

Per quello che concerne questo scritto ci limiteremo a descrivere il cervello considerando due macroaree:

- la *corteccia cerebrale*, ovvero lo strato laminare più esterno che riveste il telencefalo, avente uno spessore di 2-4 mm, costituito dalla sola sostanza grigia;
- la *sottocorteccia cerebrale*, che comprende tutta la parte sottostante del cervello.

Dal punto di vista evolutivo, la corteccia è la zona formatasi più di recente, ed è sede delle funzioni mentali cognitive complesse, quali pensiero, consapevolezza, linguaggio, memoria e attenzione.

Data la sua complessità, possiamo considerare la connettività del cervello dotata di una struttura *multiscala*, ognuna corrispondente ad una diversa risoluzione spaziale. Ci sono tre livelli principali di organizzazione:

- la *microscala*, in cui si considerano i singoli neuroni (corrispondente ad una risoluzione spaziale di micrometri);
- la *mesoscala*, dove si considerano gruppi neuronali costituiti da centinaia o migliaia di neuroni (risoluzione di centinaia di micrometri);
- la *macroscala*, corrispondente a più ampi sistemi che possono essere suddivisi in zone, ognuna caratterizzata da una specifica struttura di connettività con le altre (risoluzione di millimetri).

Lo studio riportato in questo capitolo segue quest'ultima risoluzione. In particolare, esistono varie tecniche di acquisizione dei dati non invasive a questo livello di visualizzazione, tutte basate sul principio fisico di risonanza magnetica: l'Imaging a Risonanza Magnetica Funzionale (fMRI) e l'Imaging a tensore di diffusione (DTI) e (DSI). Queste ultime due permettono l'acquisizione delle zone in cui sono presenti fibre nervose sfruttando il tracciamento delle molecole d'acqua: dato il rivestimento mielinico di tali fibre nervose, è altamente improbabile che queste vengano attraversate da molecole d'acqua, che dunque dovranno muoversi tangenzialmente ad esse, permettendo così il tracciamento delle connessioni neuronali (che non è comunque esente da errori). La DSI è, tra quelle indicate, la tecnica di acquisizione più precisa e completa.

3.2 Costruzione del modello

Per procedere con lo studio è essenziale costruire innanzitutto la struttura su cui lavorare. Sulle basi di quanto detto nei capitoli precedenti, l'intenzione è quella di creare il nostro complesso di cricche a partire dalla point cloud. Dato il livello di scala dello studio, si è scelto di suddividere il cervello in 83 zone, secondo la suddivisione di Lausanne (vedi Fig. 3.1), ognuna identificata con un vertice del grafo (i.e., elemento della point cloud). Ciascun nodo è posizionato in corrispondenza del centro di massa della regione anatomica del cervello che va a rappresentare. Da questa suddivisione passiamo alla costruzione del complesso filtrato. Si è optato per l'utilizzo di due modelli che verranno messi a confronto nel corso dello studio, uno *teorico* e uno *empirico*.

Modello minimamente cablato. Questo primo modello, che sarà il nostro *modello nullo*¹, ha una struttura che riprende il complesso di Vietoris-Rips (Definizione 2.13). L'obiettivo alla base della scelta di questo modello è riprodurre la tendenza del cervello nel

¹Termine utilizzato per indicare il modello a cui fare riferimento come modello di comparazione.

conservare il costo di cablaggio. Con questo si intende che nel cervello, tendenzialmente, è stata rilevata una preferenza dell'impulso elettrico nel percorrere quei collegamenti di lunghezza più breve, dunque avranno peso maggiore quei percorsi con distanza minore. L'importanza (i.e., forza di un collegamento) è inversamente proporzionale alla distanza tra i nodi, possiamo quindi sfruttare la costruzione del secondo tipo trattata in 2.4, utilizzando come *funzione di peso* w quella definita in (2.1). In questo modo otteniamo un complesso filtrato scandito dall'aggiunta di lati in ordine decrescente di peso.

Auditory	Banks of superior temporal sulcus (RH, LH)
	Superior temporal (RH, LH)
	Transverse temporal (RH, LH)
Cingulo-Opercular	Pars orbitalis (RH, LH)
	Pars opercularis (RH, LH)
	Rostral middle frontal (RH, LH)
	Rostral anterior cingulate (RH, LH)
	Caudal anterior cingulate (RH, LH)
	Supramarginal (RH, LH)
Default Mode	Superior frontal (RH, LH)
	Posterior cingulate (RH, LH)
	Isthmus cingulate (RH, LH)
	Precuneus (RH, LH)
Dorsal Attention	Superior parietal (RH, LH)
Fronto-Parietal	Frontal pole (RH, LH)
	Medial orbitofrontal (RH, LH)
	Pars triangularis (RH, LH)
	Caudal middle frontal (RH, LH)
	Inferior parietal (RH, LH)
	Insula (RH, LH)
Other	Parahippocampal (RH, LH)
	Entorhinal (RH, LH)
	Temporal pole (RH, LH)
	Inferior temporal (RH, LH)
	Middle temporal (RH, LH)
	Brainstem
Somatosensory	Precentral (RH, LH)
	Paracentral (RH, LH)
	Postcentral (RH, LH)
Subcortical	Thalamus proper (RH, LH)
	Caudate (RH, LH)
	Putamen (RH, LH)
	Pallidum (RH, LH)
	Accumbens area (RH, LH)
	Hippocampus (RH, LH)
Amygdala (RH, LH)	
Ventral Attention	Lateral orbitofrontal (RH, LH)
Visual	Cuneus (RH, LH)
	Pericalcarine (RH, LH)
	Lateral occipital (RH, LH)
	Lingual (RH, LH)
	Fusiform (RH, LH)

Figura 3.1: Elenco delle 83 zone in cui è stato suddiviso il cervello (contate rispettivamente per ciascun emisfero). [Siz17]

Modello acquisito tramite DSI. Il secondo modello che andiamo a descrivere, come il primo si rifà alla seconda costruzione descritta in 2.4, ma si differenzia per la scelta

della funzione w : in questo caso il peso del collegamento sarà direttamente proporzionale alla quantità di *materia bianca* rilevata tramite l'utilizzo della DSI (Diffusion Spectrum Imaging). Tale valore è ottenuto tramite il rilevamento della quantità di *fibre nervose* che collegano ciascuna coppia di zone. Anche qui vengono aggiunti i lati con ordine di peso decrescente.

Notiamo che il primo è un modello generale, ovvero è sostanzialmente lo stesso per ogni individuo confrontabile (i.e., avente stessa struttura cerebrale), mentre il secondo è fortemente individuale, in quanto si occupa di quantificare il numero di connessioni cerebrali proprie di ciascun soggetto. Nel seguito li indicheremo rispettivamente con MW (che sta per Minimally Wired) e DSI (dal nome della tecnica utilizzata per estrapolare i dati).

3.3 Studio della struttura

Lo studio [Siz17] si occupa di analizzare e descrivere due particolari strutture all'interno del cervello: le *cricche* e le *cavità*. L'analisi viene fatta su un campione di 8 individui, a ognuno dei quali corrispondono 3 scansioni tramite DSI, effettuate nell'arco di più giorni in modo da minimizzare gli errori di tracciamento.

3.3.1 Cricche

In 1.13 abbiamo visto il significato matematico di cricca. Vediamo adesso il corrispondente neurologico: possiamo vedere una k -cricca come un insieme di k zone cerebrali capaci di scambiare informazioni rapidamente, operare all'unisono o aventi funzioni simili. In particolare, saremo interessati alle cricche massimali (anche dette unità di processione). Quando parleremo di **grado** di un nodo v , ci riferiremo al numero di lati incidenti a v . Definiamo **grado di partecipazione** di un nodo v in una k -cricca massimale, indicato con $P_k(v)$, come il numero di k -cricche massimali a cui il nodo v appartiene. La **partecipazione totale**, è data da $P(v) = \sum_k P_k(v)$. Chiameremo, invece, **forza** di un nodo v , il valore dato dalla somma del peso dei lati incidenti a v , indicato con $F(v) = \sum_{e \ni v} w(e)$.

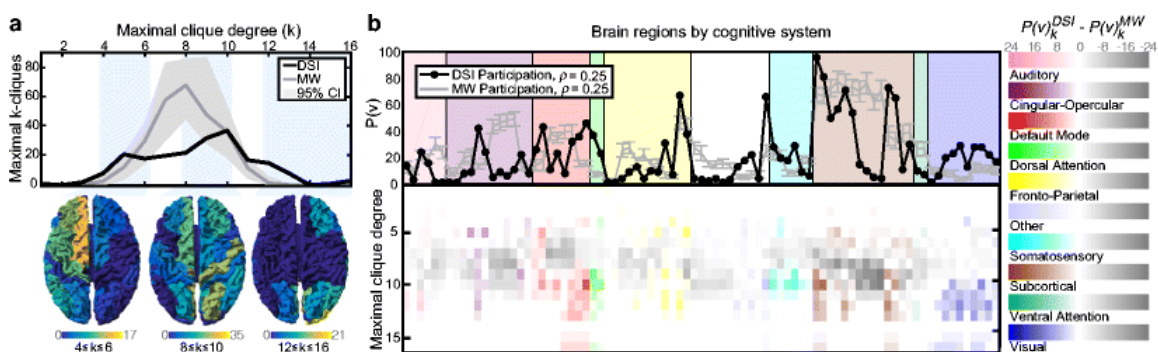


Figura 3.2: Distribuzione spaziale delle cricche massimali in DSI e MW. (a) Distribuzione delle cricche massimali rilevate tramite DSI (tracciato nero) e MW (in grigio), ottenuta a densità di lati $\rho = 0,25$. (b) Valori di partecipazione dei nodi (in alto) e la differenza tra modello reale e modello nullo (in basso) secondo la legenda a destra. Le zone considerate sono quelle elencate in fig.3.1. [Siz17]

Distribuzione anatomica delle cricche massimali

La conta delle cricche massimali nei due complessi filtrati che vengono considerati (DSI e MW) è stata effettuata al livello di filtrazione $\rho = 0,25$ (dove ricordiamo che $\rho =$

(# lati aggiunti)/(# lati totali) corrisponde alla densità dei lati del complesso di cricche al momento di filtrazione ρ , in modo da non considerare i collegamenti fittizi, ovvero quelli con peso troppo basso.

Da un confronto dei risultati ottenuti per ciascun modello, sorgono alcune differenze (Fig.3.2): nel modello DSI risulta una preferenza per le piccole ($k = 4, \dots, 6$ nodi) e grandi ($k = 12, \dots, 16$ nodi) unità di processione rispetto a quelle di media grandezza, mentre nel modello nullo i valori sono pressoché simili, differendo principalmente per quelle cricche di media grandezza ($k = 6, \dots, 10$ nodi) che sono in numero nettamente superiore.

È stata poi analizzata la distribuzione spaziale di queste cricche: si è visto che le cricche massimali più grandi sono formate da nodi localizzati, quasi esclusivamente, nel sistema sottocorticale, dorsale, visivo e default mode, suggerendo che questi sistemi siano altamente interconnessi. In particolare, la partecipazione massimale (i.e., il valore di k) di un nodo varia abbastanza uniformemente percorrendo l'asse antero-posteriore del cervello (Fig. 3.3), dove nella parte anteriore abbiamo una predominanza delle cricche massimali di piccola e media grandezza ($k = 1, \dots, 8$), mentre in quella posteriore vi sono prevalentemente cricche con un maggior numero di nodi ($k = 12, \dots, 16$). Confrontando con il modello nullo, le discrepanze maggiori si hanno principalmente nelle zone cingolo-opercolari e sottocorticali (Fig.3.2b), causate dalla differenza vista in corrispondenza delle cricche con approssimativamente 8 nodi.

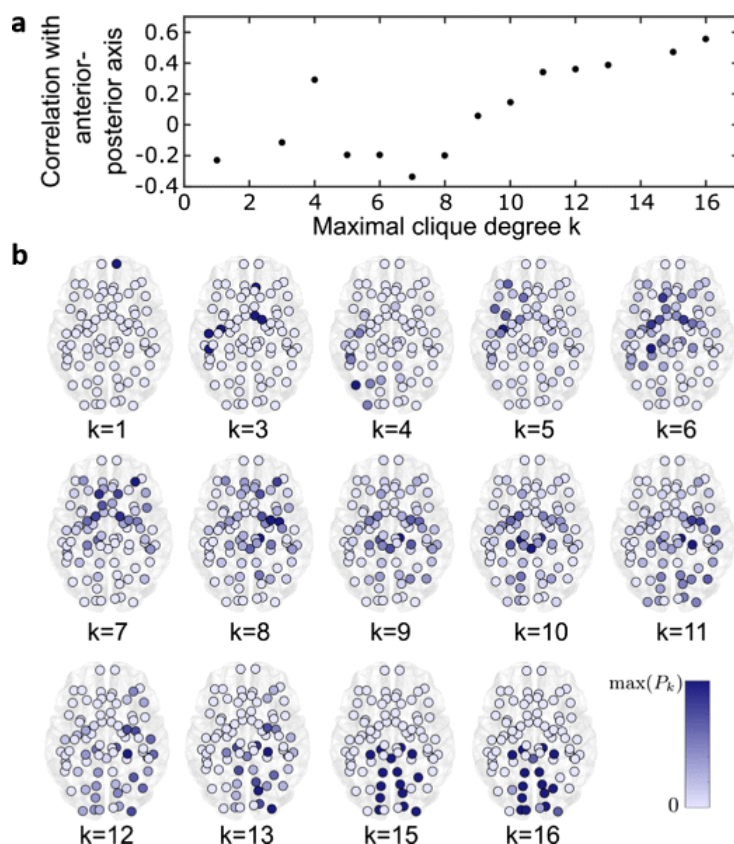


Figura 3.3: Correlazione delle cricche massimali con l'asse antero-posteriore. (a) Coefficiente di correlazione di Pearson di $P_k(v)$ lungo l'asse antero-posteriore. (b) Distribuzione spaziale di $P_k(v)$ per ogni k . Il colore del nodo corrisponde al valore di $P_k(v)$ compreso tra 0 e il massimo della partecipazione di ogni nodo per un dato k . [Siz17]

Correlazione tra partecipazione e forza di un nodo

Si è poi voluta esaminare un'eventuale correlazione tra la partecipazione totale $P(v)$ e la forza $F(v)$ di un nodo v . A questo proposito è stata trovata una correlazione lineare tra questi due parametri, in cui risulta $r = 0.957$, dove r è il coefficiente di correlazione di Pearson².

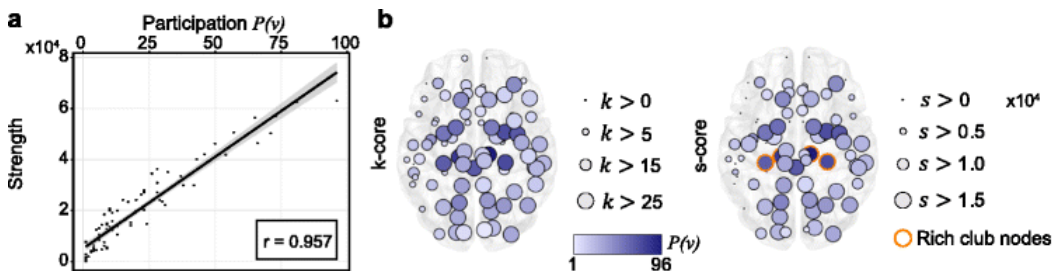


Figura 3.4: Tracciato di partecipazione delle cricche massimali. (a) Grafico di diffusione che relaziona la partecipazione di un nodo con la sua forza. (b) Calcolo della decomposizione del k -core (sinistra) e dell' s -core (destra) confrontando la partecipazione nelle cricche massimali con i nodi che formano rich-club. [Siz17]

Questa correlazione implica che le regioni con forti connessioni col resto del cervello, hanno anche un alto grado di partecipazione in molte cricche massimali. Per verificare tale affermazione, viene misurata l'associazione tra i **rich-club**, ovvero i gruppi di zone cerebrali fortemente connesse tra loro, e le altre zone, introducendo il concetto di *coreness*: un **k-core** è il più grande sottografo di G avente nodi di *grado* maggiore o uguale a k ; un **s-core**, similmente, è il più grande sottografo di G avente lati con *peso* maggiore o uguale a s . Da queste nozioni, si è passati a studiare un eventuale legame tra nodi con alta partecipazione e nodi appartenenti a k - e s -core con k e s elevati (Fig. 3.4b). Si è visto che nodi con partecipazione maggiore appartenevano a k -core e s -core con k e s più alti degli altri. Inoltre, è stata osservata l'appartenenza ai rich-club di nodi con alta partecipazione (Fig. 3.4b destra). Da questi due risultati si è potuto dedurre che le regioni nei rich-club tendono a partecipare alle unità di elaborazione locali sotto forma di cricche.

3.3.2 Cavità

Così come le cricche possono essere viste come zone altamente collegate tra loro, le cavità rappresentano zone in cui vi è un'assenza di connessioni forti. Dal punto di vista topologico, possiamo vedere questi oggetti come classi non banali dei moduli di omologia, dove due *cavità* sono considerate simili se appartengono alla stessa classe di equivalenza. Possiamo dunque utilizzare l'omologia persistente per determinare le cavità rilevanti nel connettoma, utilizzando come complessi filtrati i due modelli descritti in 3.1.

Abbiamo visto che l'*importanza* di una cavità dipende dalla durata di vita p di quest'ultima; in 2.2 abbiamo definito tale valore come la differenza tra l'indice di nascita e quello di morte della classe di omologia considerata, in questo caso però ci converrà considerare anche un altro tipo di relazione: definiamo il rapporto $\pi = \rho_{\text{death}}/\rho_{\text{birth}}$. Calcolando in questo modo la durata di vita di una cavità, viene data più importanza al momento in

²Date due variabili statistiche X e Y , l'indice di correlazione di Pearson è definito come la loro covarianza divisa per il prodotto delle deviazioni standard delle due variabili. Esso assume valori nell'intervallo $[-1, 1]$, dove a 1 corrisponde una correlazione lineare positiva perfetta ($Y = a + bX$, con $b > 0$) e a -1 la correlazione lineare negativa perfetta ($Y = a + bX$, con $b < 0$).

cui questa si forma: se due cavità hanno stessa durata di vita ma una nasce e muore in un momento antecedente della filtrazione rispetto all'altra, nel primo conto risulteranno indistinguibili, mentre nel secondo quella più precoce (i.e., con ρ_{birth} più basso) avrà una rilevanza maggiore dell'altra. Se per esempio consideriamo due cavità $(\frac{5}{20}, \frac{10}{20})$ e $(\frac{10}{20}, \frac{15}{20})$, avremo che $p = 5$ per entrambe, ma $\pi = 2$ nel primo caso e $\pi = 1,5$ nel secondo. Calcolan-

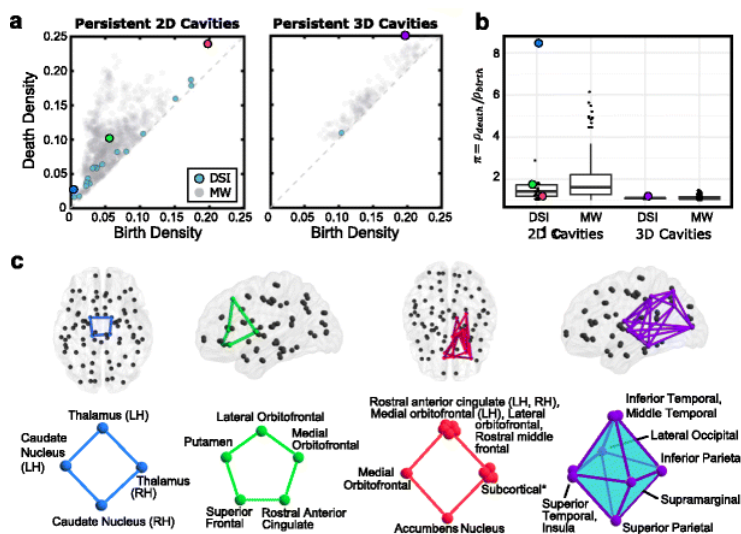


Figura 3.5: Tracciato delle cavità caratteristiche trovate in seguito alla filtrazione. (a) Diagramma di persistenza per il gruppo di controllo con DSI (verde acqua) e il modello nullo (grigio), dove a sinistra abbiamo le cavità di dimensione 2 e a destra quelle a 3 dimensioni. (b) Grafico che esprime il valore di π per ciascuna cavità in entrambi i modelli. Notiamo che le differenze dei valori di π per le cavità tridimensionali tra il modello DSI e MW non sono significative. (c) Cicli rappresentanti le cavità persistenti, mostrate nel cervello (in alto) e schematicamente (in basso). [Siz17]

do l'omologia persistente sui due modelli costruiti, è stata rilevata anche qui una differenza fra le cavità rilevanti trovate nel modello MW e quello ottenuto tramite DSI. Analizzando quest'ultimo, sono state selezionate 4 cavità (Fig. 3.5 c) con valori p e π particolarmente elevati (Fig. 3.5 a-b). Confrontando con MW risulta anche qui la presenza dei cicli blu

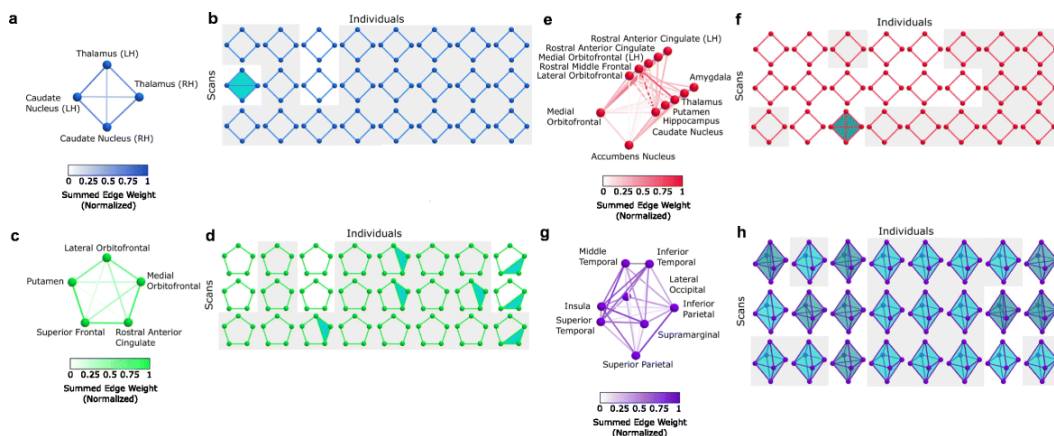


Figura 3.6: Cicli e cavità simili rilevate nel modello acquisito tramite DSI. (a-c-e-g) Cavità con lati colorati in base all'incidenza di rilevamento nelle varie scansioni tra individui. (b-d-f-h) Le colonne corrispondono a scansioni di uno stesso individuo. Notiamo che a e c sono state rilevate in almeno una scansione per ogni individuo, mentre e e g in almeno una scansione in 7 individui su 8. [Siz17]

e rosso, ma non vi è traccia degli altri due. Inoltre, in quest'ultimo modello, risulta un maggior numero di cavità non rilevate nel primo.

Per avvalorare la consistenza dei cicli trovati nel primo caso, è stata controllata la loro presenza nelle scansioni dei vari individui: nella Fig. 3.6 b-d-f-h sono riportate in grigio le zone che corrispondono a cavità simili.

3.4 Risoluzione delle discordanze tra i due modelli

Abbiamo visto che ci sono varie differenze fra i dati ottenuti attraverso i due modelli, sia nel numero di cricche massimali, sia nelle cavità trovate. Queste discordanze possono essere in parte spiegate con il fatto che non in tutte le zone del cervello vale la conservazione del costo di cablaggio. In particolare, provando a rimuovere dall'insieme dei vertici $V(G)$ le zone che corrispondono ad aree della sottocorteccia, si ha una maggiore somiglianza tra i due modelli (DSI^{cort} e MW^{cort}), soprattutto per quanto riguarda le cavità persistenti di 2 dimensioni (Fig. 3.7). Questo fatto suggerisce che la conservazione del costo di cablaggio funziona

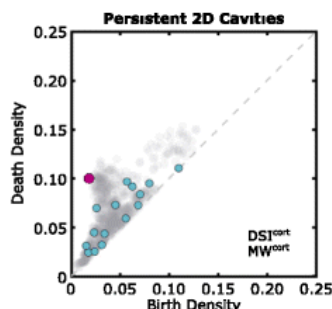


Figura 3.7: Diagramma di persistenza di DSI^{cort} (verde acqua) e MW^{cort} (grigio). [Siz17]

per le zone della corteccia mentre non vale per quelle della sottocorteccia. Inoltre, questa differenza nel numero di cavità, minore nel modello DSI, implica che l'inclusione di queste zone provoca un collasso prematuro di alcune classi di omologia. Consideriamo anche che nella fig. 3.4b si evince l'alta partecipazione di queste zone nelle cricche massimali con un maggior numero di nodi e collegamenti con peso più elevato.

Da tutte queste considerazioni si è concluso che le regioni sottocorticali formano una struttura a cono con le altre zone del cervello: possiamo infatti immaginare alcuni cicli corticali come la base di un cono completato dalle regioni della sottocorteccia (Fig. 3.8).

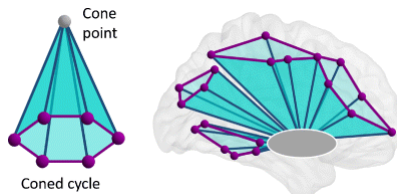


Figura 3.8: Regioni della sottocorteccia come coni nella struttura cerebrale. Un ciclo (rosa, a sinistra) può essere la base del cono, dove il punto cono (grigio) triangola l'interno del ciclo, così che il ciclo risulti essere anche un bordo. Nel cervello, il grande numero di cavità topologiche viste dopo aver rimosso le zone sottocorticali, indica che queste regioni possano agire come punti cono per molti cicli corticali. [Siz17]

Bibliografia

- [AM69] Atiyah M.F., Macdonald I.G., *Introduction to Commutative Algebra*, University of Oxford, Addison-Wesley Publishing Company, 1969.
- [CDGO13] Chazal F., De Silva V., Glisse M., Oudot S., *The structure and stability of persistence modules*, 2013.
- [CSEH05] Cohen-Steiner D., Edelsbrunner H., Harer J., *Stability of Persistence Diagrams*, 2005.
- [EH08] Edelsbrunner H., Harer J., *Persistent Homology - a Survey*, 2008.
- [EH10] Edelsbrunner H., Harer J., *Computational Topology. An introduction*, Duke University, 2010.
- [ELZ02] Edelsbrunner H., Letscher D., Zomorodian A.J., *Topological Persistence and Simplification*, 2002.
- [EM] Edelsbrunner H., Morozov D., *Persistent homology*, Chapter 26, s.d. .
- [Ghr08] Ghrist R., *Barcodes: the persistent topology of data*, 2008.
- [God] Godi F., *Bottleneck distance user manual*, <http://gudhi.gforge.inria.fr/python/latest/bottleneck_distance_user.html>, ultimo agg. 2018.
- [Hat01] Hatcher A., *Algebraic Topology*, 2001.
- [HS97] Hilton P.J., Stammach U., *A Course in Homological Algebra*, Springer, 1997, Second Edition.
- [Man14] Manetti M., *Topologia*, Springer, 2007, 2^a edizione (2014).
- [Siz17] Sizemore A., Giusti C., Khan A., Vettel J.M., Betzel R.F. Bassett D.S., *Cliques and cavities in the human connectome*, 2017.
- [Spa66] Spanier E., *Algebraic Topology*, McGraw-Hill, 1966.
- [ZC05] Zomorodian A. J., Carlsson G., *Computing persistent homology*, Discrete & Computational Geometry, 2005.
- [Zom05] Zomorodian A. J., *Topology for computing*, Cambridge University Press, 2005.
- [Zom10] Zomorodian A. J., *The tidy set: a minimal simplicial set for computing homology of clique complexes*, In Proceedings of the Annual Symposium on Computational Geometry, 2010.
- [Zom] Zomorodian A. J., *Fast Construction of the Vietoris-Rips Complex*, Dartmouth College, s.d. .

Ringraziamenti

Questo scritto è il perfetto connubio tra due campi del sapere che mi hanno accompagnata durante questo percorso di studi. Il primo scelto, il secondo indotto inaspettatamente poco dopo.

I primi ringraziamenti vanno innanzitutto a coloro che hanno contribuito alla formazione di questo scritto: la prima persona che ringrazio è il mio relatore, il Prof. Marco Manetti, per aver scelto di seguirmi durante questa compilazione, per la sua disponibilità e, soprattutto, per la sua grande apertura nei confronti di un ambito *differente*. Successivamente, ringrazio una persona che ha in qualche modo incentivato questo lavoro, la Prof.ssa Ada Francia, la quale mi ha seguita e aiutata in maniera diversa ma certamente non meno importante. Un altro ringraziamento va alla Dott.ssa Martina Scolamiero, per la revisione dello scritto e gli spunti conseguenti. Ringrazio inoltre il Prof. Fulvio Bongiorno e il Prof. Andrea Damiani, per i consigli e il tempo dedicatomi.

Passando alla seconda categoria, vorrei ringraziare tutte quelle persone che hanno contribuito indirettamente, affiancandomi nel corso di questi anni: un primo grande grazie va ai miei genitori, coloro che *ci sono sempre stati*, che incondizionatamente hanno pensato al mio bene nonostante i miei pochi sforzi per meritarlo. A mio fratello Marco, alle nostre chiacchierate e alle nostre discussioni, al nostro essere, in fondo, così *vicini*. A mio nonno Adriano, per farmi sentire ogni giorno la sua *nipotina*. E a tutto il resto della famiglia, per essersi comportata in modo da meritare tale nome. Un grazie speciale a Yasmin, la compagna di una vita, che da troppo tempo sopporta le mie turbe e mi è stata vicina sia durante che oltre i (tanti) momenti di crisi. A Giulia, Federica e Francesca, compagne di cazzeggio e di studi, ma anche di momenti che richiedevano presenza e conforto. Agli amici dell'università, che mi hanno accompagnata in questo percorso; un grazie particolare a coloro che sono andati anche oltre: Carla, Edoardo e Alessandro (e all'inventore della Canasta). Un ultimo grazie a quella persona che mi ha fatto fare il primo passo per permettere l'arrivo di questo momento: la Prof.ssa Filomena Curcio, che per prima mi ha insegnato il rigore e trasmesso la bellezza della matematica.

Infine un messaggio, per la terza persona a cui è dedicata questa tesi: *ad maiora!*