UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de São José do Rio Preto

Guilherme Vituri Fernandes Pinto

Motivic constructions on graphs and networks with stability results

São José do Rio Preto
2020

Guilherme Vituri Fernandes Pinto

Motivic constructions on graphs and networks with stability results

Tese apresentada como parte dos requisitos para obtenção do título de Doutor em Matemática, junto ao Programa de Pós-Graduação em Matemática, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: CAPES

Orientador: Prof. Dr. Thiago de Melo
Co-orientador: Prof. Dr. Facundo Mémoli

São José do Rio Preto
2020

P659m

Pinto, Guilherme Vituri Fernandes
Motivic constructions on graphs and networks with stability results / Guilherme Vituri Fernandes Pinto. -- São José do Rio Preto, 2020
122 p.

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto
Orientador: Thiago de Melo
Coorientador: Facundo Mémoli

1. Análise topológica de dados. 2. Grafos. 3. Networks. 4. Clustering. I. Título.

Guilherme Vituri Fernandes Pinto

Motivic constructions on graphs and networks with stability results

Tese apresentada como parte dos requisitos para obtenção do título de Doutor em Matemática, junto ao Programa de Pós-Graduação em Matemática, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto.

Financiadora: CAPES

Comissão Examinadora

Prof. Dr. Facundo Mémoli
The Ohio State University (USA)
Co-orientador

Profa. Dra. Alice Kimie Miwa Libardi
Departamento de Matemática - Unesp Rio Claro

Prof. Dr. Edivaldo Lopes dos Santos
Departamento de Matemática - Universidade Federal de São Carlos

---

Prof. Dr. Washington Mio
Departamento de Matemática - Florida State University (USA)

Prof. Dr. Jamil Viana Pereira
Departamento de Matemática - Unesp Rio Claro

Rio Claro
19 de março de 2020

To the Moon.

ACKNOWLEDGMENTS

This thesis is the product of some years of work and thought, of tons of draft sheet and ghostly ideas that haunted me for many nights.

It would not be possible without financial support from CAPES, and the help of my two advisors, who I can't thank enough here. *Muchas gracias*, Facundo, for receiving me in Ohio, that land so beautiful, for the dinners and the conversations, and for teaching me so much, making so strange connections between many objects that we have studied, linking them together as if, from the very beginning, we could have seen the entire work as one elegant piece; and *muito obrigado* Thiago for accepting to join this adventure called Topological Data Analysis, which none of us knew before, and for continuing to guide me through the mysterious lands of Mathematics for so many years. You first met me when I was a young man on my eighteen, fresh from high school, and continued to supervise me until now, when I became a... well, I am not sure of what I am now.

Being the work of many years of my youth, it has, between the lines, many ideas of my youth. Like the plants in my garden, many ideas died without any apparent cause, but some others flourished and gave fruits.

Being the work of my brain, it is important that I tell one of its small problems: excerpts of books I read or music I hear can assault me suddenly and keep repeating in my mind, until this small excerpt create to itself an entire world of sensations and feelings and meaning. Sometimes it is a music I heard months before, that sounds as if I was in a concert hall. Other times, it is a phrase of a book: that's way I insert so many quotations to say so simple things. While I was in Ohio, I have read many Kurt Vonnegut's books. Vonnegut's "The Sirens of Titan", in particular, impressed me profoundly. The next quotation is thought by Constant, a man who is trapped in the moon of Saturnus called Titan with a robot, a woman and his son (who thinks he is a bird), while looking to the sky:

It was all so sad. But it was all so beautiful, too.

It instantly makes me think about the winter nights in Columbus, and feel them, the sounds of my boots on the snow, the glacial wind, the darkness and white surrounding everything. . .

I miss Ohio. Sometimes I dream about it; in the dreams I am often walking by North High Street, right where I lived, looking to the sky, to the trees so green, with such a vivid green as I never saw before. It made me fell alive as if it was springtime for me, too. Maybe it was.

When I went there, I left my life in Brazil a bit upside down. I went there wanting to be a stranger, as in Schubert's "Der Wanderer" song (here translated by Paul Hindemith, originally written by Georg Philipp Schmidt von Lübeck):

The sun seems so cold to me here,
The flowers faded, the life old,
And what they say has an empty sound;
I am a stranger everywhere.

Where are you, my dear land?
Sought and brought to mind, yet never known,
That land, so hopefully green,
That land, where my roses bloom,

Where my friends wander
Where my dead ones rise from the dead,
That land where they speak my language,
Oh land, where are you?

And a stranger I was, and maybe still am.

However, I always knew that my life there was just temporary. This certainty about the end of my new life made me wonder, and feel as wrote Sir Walter Scott (quoted by Vonnegut in his "Mother Night"):

Breathes there the man, with soul so dead,
Who never to himself hath said,

> ‘This is my own, my native land!’
> Whose heart hath ne’er within him burn’d
> As home his footsteps he hath turn’d
> From wandering on a foreign strand?

And here I am, indeed, just six months after my arrival, after just three planes...

When Constant was going back to Earth in a spaceship, Salo (the robot) hypnotized him:

> Constant was already in a nearly hypnotic state, staring out at the Cosmos through a porthole. Salo came up behind him and spoke to him soothingly.
>
> “You are tired, so very tired, Space Wanderer,” said Salo.
>
> “Stare at the faintest star, Earthling, and think how heavy your limbs are growing.”

How can I relate my return to Brazil to a man in a spaceship being hypnotized by a robot? Not even I know.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, à qual agradeço.

Oh minha lua cheia, oh minha doce amiga!
Possas tu não mais ver em tão cruel fadiga
o homem que tanta vez dos céus hás contemplado
a desoras velando, em livros engolfado.
Melancólica amante! *a claridade tua*
achou-me sempre a ler. Se hoje um teu raio, ó lua,
me levasse a pairar nos cumes apartados,
a borboletear nos antros frequentados
dos espíritos só, a saltitar liberto
da científica névoa, em fundo de um deserto,
à luz crepuscular que tácita derramas
aos selvosos desvãos, por entre as móveis ramas!
Que refrigério d'alma um banho nesse rócio
não dera, amada lua, às febres do teu sócio!
Johann Wolfgang von Goethe, "Fausto". Traduzido por Agostinho de Ornelas.

O full and splendid Moon, whom I
Have, from this desk, seen climb the sky
So many a midnight,—would thy glow
For the last time beheld my woe!
Ever thine eye, most mournful friend,
O'er books and papers saw me bend;
But would that I, on mountains grand,
Amid thy blessed light could stand,
With spirits through mountain-caverns hover,
Float in thy twilight the meadows over,
And, freed from the fumes of lore that swathe me,
To health in thy dewy fountains bathe me!
Johann Wolfgang von Goethe, "Faust". Translated by Bayard Taylor.

RESUMO

Neste trabalho estudamos certos funtores sobre grafos, chamados de representáveis ou motívicos. Esses funtores não mudam os vértices de um grafo, mas apenas suas setas (as arestas direcionadas). Quaisquer tais funtores podem ser estendidos para *networks* (uma generalização de espaços métricos). Funtores de *clustering* sobre grafos dão origem a funtores de *hierarchical clustering* sobre *networks*. Mais ainda, podemos modificar a definição de funtor representável para criar filtrações de complexos simpliciais, que tem como caso particular os complexos de Vietoris-Rips e Čech. Isso faz com que possamos aplicar o funtor de homologia simplicial e obter um diagrama de persistência, como usual em Análise Topológica de Dados. Obtivemos resultados de estabilidade com respeito à distância *bottleneck* e à distância *network*, quando uma certa condição é imposta nos motivos de um funtor representável. Algumas operações sobre grafos (e.g., produtos e suspensão) também podem ser estendidas para *networks*, e três fórmulas de Künneth foram obtidas. Finalmente, alguns algoritmos e códigos para casos especiais são fornecidos com exemplos.

Palavras-chave: Análise topológica de dados, Grafos, Networks, Clustering.

ABSTRACT

In this work we study certain functors on graphs, called representable or motivic. These functors do not change the vertices of a graph, but only its arrows (the directed edges). Any such functor can be extended to networks (a generalization of metric spaces). Clustering functors on graphs give rise to hierarchical clustering functors on networks. Moreover, we can further modify the definition of a representable functor on graphs to create simplicial complex filtrations on networks, which have as particular cases the Vietoris-Rips and the Čech simplicial complexes. This allows us to apply the simplicial homology functor and obtain a persistent diagram, as in the usual pipeline of Topological Data Analysis. We obtained some stability results regarding the bottleneck distance and the network distance, when a certain condition is imposed in the motives of a representable functor. Some operations on graphs can also be extended to networks (products and suspension), and three Künneth formulas were obtained. Finally, some algorithms and codes for special cases are provided with examples.

Keywords: *Topological data analysis, Graphs, Networks, Clustering.*

List of Figures

List of Symbols

The next list describes several symbols that will be later used within the body of the document.

General

GRAPHS

C_n Cycle graph with n vertices . . . . . . (see p. 45)

$D_{(G)}$ description . . . . . . (see p. 44)

D_n Discrete graph with n vertices . . . . . . (see p. 44)

$K_{(G)}$ description . . . . . . (see p. 44)

K_n Complete graph with n vertices . . . . . . (see p. 44)

L_n Line graph with n vertices . . . . . . (see p. 44)

T_n Transitive line graph with n vertices . . . . . . (see p. 45)

SETS

Bar The set of barcodes . . . . . . (see p. 30)

Dgm Set of all finite multisets . . . . . . (see p. 31)

Pers The set of persistent diagrams . . . . . . (see p. 30)

Contents

Introduction

Clustering datasets is useful to identify subgroups that exhibit some kind of proximity or similarity. The practice of clustering metric spaces, with its many algorithms, is well developed [3]. From a theoretical perspective, in [4, 5] the authors used the functoriality to define desirable properties of maps that send a metric space to a hierarchical clustering of its vertices and proved that a unique method, *single linkage* hierarchical clustering, satisfies these properties. When restricting to the category of metric spaces and *injective* maps, the same authors found an infinite family of standard and hierarchical clustering methods satisfying these same properties. This is a counterpart to a result of Kleinberg [6], which states that there is no method of standard (as opposed to hierarchical) clustering satisfying certain natural conditions.

However, when a dataset can no longer be represented as a metric space, the interpretation of a clustering can be more difficult. In [7], the authors extended their previous work and studied hierarchical clustering of *dissimilarity networks*: pairs (X, A) where $A\colon X \times X \to \mathbb{R}^+$ satisfies $A(x, x') = 0$ if, and only if, $x = x'$. Under reasonable conditions, the methods of hierarchical clustering that they identified were well behaved and many results, such as stability with respect to a suitable notion of distance between networks, were proved.

We further generalize these dissimilarity networks and study pairs (X, w), where $w\colon X \times X \to \mathbb{R} \cup \{+\infty\}$ is any function. These objects, called *extended networks* (and whose category is denoted by $\mathcal{N}_{\text{ext}}$) can be regarded as a filtration of (directed) graphs. By studying endofunctors $\mathfrak{F}\colon \mathcal{G} \to \mathcal{G}$ (where $\mathcal{G}$ is the category of graphs) we are able to create many different clustering functors on $\mathcal{G}$. These endofunctors naturally give rise to endofunctors on $\mathcal{N}_{\text{ext}}$ whose output is a generalization of an ultrametric space (or, equivalently, a dendrogram). This approach, although not as general as directly dealing with general endofunctors on $\mathcal{N}_{\text{ext}}$, turns out to be very useful and simplifies many proofs. For example, the notion of quasi-clustering from [7] can be obtained as in Definition 3.15, whenever G is a transitive graph. Besides this, the study of endofunctors on $\mathcal{G}$ is interesting on its own.

We borrow the concept of representable methods from [8] and adapt it to the context of endofunctors on $\mathcal{G}$. Given a set of graphs Ω (the *representers* or *motifs*), we can define $\mathfrak{F}^{\Omega}\colon \mathcal{G} \to \mathcal{G}$ as a functor that captures "interesting shapes" based on Ω.

The definition and properties of endofunctors $\mathfrak{F}$ that carry a notion of density (as in [4]) can be done with a more general construction using simplicial complexes.

Chapter 1 introduces the machinery of TDA: simplicial complexes, persistent vector spaces, persistent diagrams and the bottleneck distance.

In Chapter 2 we present some endofunctors on graphs, and present the main tool of this work: representable endofunctors. The images of these functors $\mathfrak{F}^{\Omega}$ are made when we try to "fit" certain graphs (the elements of Ω) in a given graph, via graph maps. Section 2.4 presents the notion of *pointed* representable functor, which tries to remove the symmetry in the definition of usual representable functors. A composition rule is obtained, and it turns out that every endofunctor is pointed representable. In Section 2.5 we completely characterize the relation between sets of representers Ω_1 and Ω_2 so that the functors represented by them are equal, and show how to "simplify" a given family of representers.

In Chapter 3 we show how an endofunctor between graphs induces a functor between networks (a generalization of a metric space). Moreover, when the endofunctor $\mathfrak{F}$ is no trivial, we show a stability result: the distance between $\widehat{\mathfrak{F}}(\mathbf{X})$ and $\widehat{\mathfrak{F}}(\mathbf{Y})$ is bounded by the distance between $\mathbf{X}$ and $\mathbf{Y}$, for any networks $\mathbf{X}$ and $\mathbf{Y}$.

In Chapter 4 we introduce the notion of representable (or motivic) homology and prove its stability with respect to the bottleneck distance, when the set of representers satisfy just one property (the "repetition closed" property), having the Vietoris-Rips and the Dowker homologies as special cases.

In Chapter 5 we study some operations on graphs and networks, like the product, join and suspension, and obtain Künneth formulas.

In Chapter 6 we see some examples of clustering and motivic homologies and discuss the complexity of such algorithms.

1 Preliminaries

> *"What is the purpose of all this?" he asked politely. "Everything must have a purpose?" asked God. "Certainly," said man. "Then I leave it to you to think of one for all this," said God. And He went away.*
>
> KURT VONNEGUT, *CAT'S CRADLE*

The aim of this chapter is to provide an overview of simplicial complexes and simplicial homology in order to understand the construction of many simplicial complex filtrations over metric spaces and networks. We also present the machinery of Topological Data Analysis: persistent vector spaces, persistent diagrams and the bottleneck distance.

The next section will follow closely the excellent book [9].

1.1 Simplicial complexes and simplicial homology

A simplicial complex is a combinatorial object widely used in Topological Data Analysis (TDA) to "undiscretize" finite data (metric spaces or networks) in order to use tools from topology, like homotopy and homology. Atkin says in his book [10]: "In order to capture the geometric essence of any natural system N, we must choose an appropriate formal geometric structure into which the observables of N can be encoded. It turns out to be useful to employ what is termed a simplicial complex as our formal mathematical framework... A simplicial complex is a natural generalization of the intuitive idea of a Euclidean space, and is formed by interconnecting a number of pieces of varying dimension. The mathematical apparatus, which has its roots in algebraic topology, gives us a systematic procedure for keeping track of how the pieces fit together to generate the entire object, and how they each contribute to the geometrical representation of N."

As we will see in Chapter 4, one of the main constructions in TDA is to associate a filtration of simplicial complexes to a given data and then obtain a barcode that summarizes the way the "holes" appear and disappear in this filtration.

Definition 1.1 ([9, §3])**.** An *(abstract) simplicial complex* K is a collection of finite non-empty sets such that if $\sigma \in K$ then any non-empty subset of σ is also in K.

Each $\sigma = \{v_0, \ldots, v_p\} \in K$ is called a *simplex* or, more precisely, a p-simplex, where $\dim(\sigma) := p$ is the *dimension* of σ. Any non-empty subset τ of σ is called a *face* of σ. It is a *proper face* if $\tau \neq \sigma$. The dimension of K, denoted $\dim(K)$, is the maximal dimension of its simplexes, that is: $\dim(K) = \max\{\dim(\sigma), \sigma \in K\}$. When there is no largest dimension, $\dim(K) = \infty$. The vertex set V of K is the union of its 0-dimensional simplexes. A subcollection of K that is itself a complex is called a *subcomplex* of K. The p-skeleton of K is the subcomplex of K consisting of all the simplexes with dimension less than or equal to p.

Given two simplicial complexes K and L with vertex set V and V', respectively, we say that $f\colon V \to V'$ is a *simplicial map* if $\{v_0, \dots, v_p\} \in K$ implies $\{f(v_0), \dots, f(v_p)\} \in L$, and denote it by $f\colon K \to L$. Then K and L are *isomorphic* if there is a bijective simplicial map $f\colon K \to L$ whose inverse is also a simplicial map. This is equivalent to say that $\{v_0, \dots, v_p\} \in K \Leftrightarrow \{f(v_0), \dots, f(v_p)\} \in L$.

Definition 1.2 ([9]). Let σ be a simplex of K. Define two orderings of its vertex set to be equivalent if they differ from one another by an even permutation. When $\dim(\sigma) > 0$, the orderings of the vertices of σ then fall into two equivalence classes. Each of these classes is called an *orientation* of σ. An *oriented simplex* is a simplex σ together with an orientation of σ. If $\sigma = \{v_0, \dots, v_p\}$, we will denote its oriented version by $[v_0, \dots, v_p]$. When the context is clear, we will use the same symbol σ for the simplex or the oriented simplex.

In what follows, we will be interested only in chains taking values in a field $\mathbb{K}$, because in this case the homology groups will be vector spaces. Theorem 1.19 then assures that we can get a barcode (or persistent diagram) that describes our data.

Definition 1.3 ([9]). Let K be a simplicial complex. A p-chain on K is a function c from the set of oriented p-simplexes of K to a field $\mathbb{K}$ such that:

- $c(\sigma) = -c(\sigma')$ if σ and σ' are opposite orientations of the same simplex.
- $c(\sigma) = 0$ for all but finitely many oriented p-simplexes σ.

Denote by $C_p(K)$ the set of all p-chains of K. We have a natural algebra in this set: for all oriented p-simplexes σ, define

- $(c_1 + c_2)(\sigma) = c_1(\sigma) + c_2(\sigma)$,
- $(\lambda c)(\sigma) = \lambda c(\sigma), \forall \lambda \in \mathbb{K}$.

Thus, $C_p(K)$ is a vector space, called *the vector space of (oriented) p-chains of K*. If $p < 0$ or $p > \dim(K)$, let $C_p(K)$ be the trivial vector space.

If σ is an oriented simplex, the *elementary chain c* corresponding to σ is the function defined by $c(\sigma) = 1$, $c(\sigma') = -1$ if σ' is the opposite orientation of σ, and $c(\tau) = 0$ for all other oriented simplexes τ, where 1 is the identity element of $\mathbb{K}$.

By abuse of notation, we will also use symbol σ to denote the elementary chain corresponding to σ. With this convention, if σ and σ' are opposite orientations of the same simplex, then we have $\sigma = -\sigma'$, because this is true when σ and σ' are interpreted as elementary chains.

The next lemma shows the utility of the elementary chains.

Lemma 1.4 ([9]). The set K_p of p-simplexes of K is a basis for the vector space $C_p(K)$.

Now we will define some algebra in $C_p(K)$ to be able to detect "holes" in the simplicial complex K.

Definition 1.5 ([9])**.** Define the *boundary operator* $\partial_p \colon C_p(K) \to C_{p-1}(K)$ as follows: for any $\sigma = [v_0, \dots, v_p]$ with $p > 0$,

$$\partial_p(\sigma) = \sum_{i=0}^{p} (-1)^i [v_0, \dots, \widehat{v_i}, \dots, v_p],$$

where $\widehat{v_i}$ means that the vertex v_i is being removed from the array. This is a well-defined linear map.

We will often omit the subscript p in ∂_p.

Lemma 1.6 ([9])**.** $\partial_{p-1} \circ \partial_p = 0$.

Lemma 1.6 guarantees that the sequence $C_\bullet(K)$ given by

$$\cdots \xrightarrow{\partial_{p+1}} C_p(K) \xrightarrow{\partial_p} C_{p-1}(K) \xrightarrow{\partial_{p-1}} \cdots$$

is a *semi exact sequence*, that is, denoting $B_p(K) = \operatorname{Im}(\partial_{p+1})$ (the space of *p-boundaries*) and $Z_p(K) = \ker(\partial_p)$ (the space of *p-cycles*), we have $B_p(K) \subseteq Z_p(K)$. Any such sequence of vector spaces and linear maps that form a exact sequence is called a *chain complex*.

We can define the vector space

$$H_p(K) = Z_p(K)/B_p(K),$$

called the p-th *homology group* of K.

Roughly speaking, when $p > 0$, the dimension of $H_p(K)$ measures how many "holes of dimension p" there are in K, while for $H_0(K)$ it measures how many connected components there are in K.

It can be shown [11, Theorem 2.3.3] that different orderings of the same complex yield isomorphic homology groups.

Example 1.7. In Figure 1.1 we have a simplicial complex

$$K_1 = \{[a], [b], [c], [a, b], [a, c], [b, c]\}$$

on the left, which can be seen as the 3 edges of a triangle, and a "filled" triangle on the right, say K_2, obtained from K_1 by adding the 2-simplex $[a, b, c]$. The cycle $[a, b] + [b, c] + [c, a]$ generates $Z_1(K_1)$ and $Z_1(K_2)$, but in K_2 it is also the boundary of $-[a, b, c]$, that is, it gets "killed" in $H_1(K_2)$. Thus, $H_0(K_i) \simeq \mathbb{K}$, $H_1(K_1) \simeq \mathbb{K}$, $H_1(K_2) = 0$ and $H_n(K_i) = 0$ for $n > 1$ and $i = 1, 2$.

The quotient in the definition of H_p ensures that cycles that are also a boundary (that is, cycles that are "filled") will be identified to zero. This is why we say that H_p detects "holes". See [9] for a complete introduction.

Definition 1.8 ([9])**.** Given a simplicial map $f \colon K \to L$, define a homomorphism

$$(f_p)_\# \colon C_p(K) \to C_p(L)$$

for every $p \geq 0$ by

$$(f_p)_\#([v_0, \dots, v_p]) = \begin{cases} [f(v_0), \dots, f(v_p)], & \text{if } f(v_0), \dots, f(v_p) \text{ are distinct,} \\ 0, & \text{otherwise,} \end{cases}$$

Figure 1.1: On the left: the simplicial complex K_1 has a 1-cycle $[a, b] + [b, c] + [c, a]$. Taking K_2 as the union of $[a, b, c]$ with K_1, the same cycle $[a, b] + [b, c] + [c, a]$ is now the boundary of $-[a, b, c]$.

and then extend it linearly. This is a well-defined linear map. The family of homomorphisms $\{(f_p)_\#, p \in \mathbb{Z}\}$ is called the *chain map induced by the simplicial map* f. We will often omit the subscript p of $(f_p)_\#$. We will also use the same symbol ∂ to the boundary operators in K and L; if necessary, we can use the notations ∂_K and ∂_L.

Lemma 1.9 ([9])**.** The homomorphism $f_\#$ commutes with ∂, that is, the squares in the following diagram are commutative:

$$\begin{tikzcd}
\cdots \arrow[r, "\partial"] & C_{p+1}(K) \arrow[r, "\partial"] \arrow[d, "f_\#"'] & C_p(K) \arrow[r, "\partial"] \arrow[d, "f_\#"] & \cdots \\
\cdots \arrow[r, "\partial"] & C_{p+1}(L) \arrow[r, "\partial"] & C_p(K) \arrow[r, "\partial"] & \cdots
\end{tikzcd}$$

This lemma guarantees that $f_\#$ induces a linear map $f_*\colon H_p(K) \to H_p(L)$.

Theorem 1.10 ([9])**.** 1. Let $i\colon K \to K$ be the identity simplicial map. Then the homomorphism $i_*\colon H_p(K) \to H_p(K)$ is the identity map.

2. Let $f\colon K \to L$ and $g\colon L \to M$ be simplicial maps. Then $(g \circ f)_* = g_* \circ f_*$, that is, the following diagram commutes:

$$\begin{tikzcd}
H_p(K) \arrow[rr, "(g\circ f)_*"] \arrow[dr, "f_*"'] & & H_p(M) \\
& H_p(L) \arrow[ur, "g_*"'] &
\end{tikzcd}$$

The readers who already saw a bit of category theory can guess what this theorem means: that the homology map *is a functor* from the category of the simplicial complexes to the category of vector spaces.

The language of category theory may look too abstract, but can really summarize what is happening in many ocasions, like in this one above. This is the right time to introduce it. As MacLane says in the introduction of [12], "*category theory starts with the observation that many properties of mathematical systems can be unified and simplified by a presentation with diagrams of arrows.*"

Definition 1.11 ([9, §28])**.** A *category* C consists of three things:

- A class of objects, $\mathrm{Obj}(C)$.
- For every ordered pair (X, Y) of objects, a set $\mathrm{Hom}_C(X, Y)$ of *morphisms*.

- A function, called *composition of morphisms*,

$$\mathrm{Hom}_{\mathcal{C}}(X,Y) \times \mathrm{Hom}_{\mathcal{C}}(Y,Z) \to \mathrm{Hom}_{\mathcal{C}}(X,Z)$$

  which is defined for every triple (X, Y, Z) of objects.

The image of the pair (f, g) under the composition operation is denoted by $g \circ f$. The following two properties must be satisfied:

Axiom 1 (Associativity) If $f \in \mathrm{Hom}_{\mathcal{C}}(W,X)$, $g \in \mathrm{Hom}_{\mathcal{C}}(X,Y)$ and $h \in \mathrm{Hom}_{\mathcal{C}}(Y,Z)$ then $h \circ (g \circ f) = (h \circ g) \circ f$.

$$\begin{tikzcd} W \arrow[r, "f"] \arrow[rr, bend left, "g\circ f"] & X \arrow[r, "g"] \arrow[rr, bend right, "h\circ g"'] & Y \arrow[r, "h"] & Z \end{tikzcd}$$

Axiom 2 (Existence of identities) If X is an object, there is an element $1_X \in \mathrm{Hom}_{\mathcal{C}}(X,X)$ such that

$$1_X \circ f = f \quad \text{and} \quad g \circ 1_X = g,$$

for every $f \in \mathrm{Hom}_{\mathcal{C}}(W,X)$ and every $g \in \mathrm{Hom}_{\mathcal{C}}(X,Y)$, where W and Y are arbitrary objects.

$$\begin{tikzcd} W \arrow[r, "f"] \arrow[rr, bend left, "f"] & X \arrow[r, "1_X"] \arrow[rr, bend right, "g"'] & X \arrow[r, "g"] & Z \end{tikzcd}$$

In general, we write $f : X \to Y$ to mean $f \in \mathrm{Hom}_{\mathcal{C}}(X,Y)$, and we call X the *domain (object)* of f, and Y, the *range (object)* of f.

Definition 1.12 ([9]). A (covariant) *functor* F from a category $\mathcal{C}$ to a category $\mathcal{D}$ is a function assigning to each object X of $\mathcal{C}$ and object $F(X)$ of $\mathcal{D}$, and to each morphism $f : X \to Y$ of $\mathcal{C}$, a morphism $F(f): F(X) \to F(Y)$ of $\mathcal{D}$. The following two conditions must be satisfied:

$$F(1_X) = 1_{F(X)} \text{ for all } X,$$
$$F(g \circ f) = F(g) \circ F(f).$$

That is: a functor must preserve composition and identities.

As examples of categories, we have:

- The category **Set** of sets and functions.
- The category **Simp** of simplicial complexes and simplicial maps.
- The category **Vec** of vector spaces and linear maps.
- The category **Top** of topological spaces and continuous maps.

Thus, for every $p \in \mathbb{Z}$, we can define the *p-dimensional homology functor*

$$H_p : \mathbf{Simp} \to \mathbf{Vec}$$

given by $K \mapsto H_p(K)$ on objects and $f \mapsto f_*$ on morphisms.

1.2 Persistent homology

The goal of TDA is to summarize data on all possible scales, and analyze which features *persist* the most. So, instead of associating one simplicial complex to a given metric space, we can associate a *filtration* of simplicial complexes. In this section we will present some *persistent* objects.

Definition 1.13 ([13]). A *filtration of simplicial complexes* (or *simplicial filtration*) is a collection $\mathcal{K} = \{K_\epsilon\}_{\epsilon\in\mathbb{R}}$ where K_ϵ is a subcomplex of $K_{\epsilon'}$ for any $\epsilon \leq \epsilon'$. There is a number $\epsilon_{\mathcal{K}}$ such that $K_\epsilon = K_{\epsilon_{\mathcal{K}}}$ for any $\epsilon \geq \epsilon_{\mathcal{K}}$. Thus, $\mathcal{K}$ contains only a finite number of distinct simplicial complexes.

Considering $\mathbb{R}$ as a category with one morphism from ϵ to ϵ' if and only if $\epsilon \leq \epsilon'$, we can see a simplicial filtration as a functor $S\colon \mathbb{R} \to \mathbf{Simp}$, where the map $K_\epsilon = S(\epsilon) \to S(\epsilon') = K_{\epsilon'}$ is the inclusion map.

Applying homology to all complexes in a given filtration $\mathcal{K}$, we obtain a family of vector spaces and linear maps: a particular case of a *persistent vector space*. This algebraic object contains information about the birth and death of “holes” in $\mathcal{K}$. Persistent vector spaces play an important role in TDA, as a link between the geometry of $\mathcal{K}$ and a multiscale summary of it.

Definition 1.14 ([13]). A *persistent vector space* (or *persistent module)* is a collection $\mathbb{V}$ of vector spaces $\{V_\epsilon\}_{\epsilon\in\mathbb{R}}$ and linear maps $\{v_\epsilon^{\epsilon'}\colon V_\epsilon \to V_{\epsilon'}\}_{\epsilon\leq\epsilon'}$ satisfying the following properties:

1. v_ϵ^ϵ is the identity map, for all $\epsilon \in \mathbb{R}$.
2. $v_{\epsilon'}^{\epsilon''} \circ v_\epsilon^{\epsilon'} = v_\epsilon^{\epsilon''}$, for all $\epsilon \leq \epsilon' \leq \epsilon''$.

To write the vector spaces of $\mathbb{V}$ together with its maps in a single notation, we often write $\mathbb{V} = \left\{V_\epsilon \xrightarrow{v_\epsilon^{\epsilon'}} V_{\epsilon'}\right\}_{\epsilon\leq\epsilon'}$.

We say that $\mathbb{V}$ is *finite dimensional* if every V_ϵ has finite dimension.

Again, we can regard $\mathbb{V}$ as a functor $\mathbb{R} \to \mathbf{Vec}$ sending ϵ to V_ϵ and a morphism $\epsilon \to \epsilon'$ (that is, $\epsilon \leq \epsilon'$) to $v_\epsilon^{\epsilon'}$. Properties 1 and 2 are consequences of the functoriality.

The “basic blocks” of persistent vector spaces are defined as follows:

Definition 1.15. Given $b, d \in \mathbb{R}$ with $b \leq d$, an *interval module* $\mathbb{I}_{[b,d)} = \left\{I_\epsilon \xrightarrow{i_\epsilon^{\epsilon'}} I_{\epsilon'}\right\}_{\epsilon\leq\epsilon'}$ is a persistence vector space such that

1. $I_\epsilon = \mathbb{K}$ for $\epsilon \in [b, d)$ and 0 otherwise,
2. $i_\epsilon^{\epsilon'}\colon I_\epsilon \to I_{\epsilon'}$ is the identity map if $\epsilon, \epsilon' \in [b, d)$, and the trivial map otherwise.

Interval modules can be interpreted as features that are *born* at parameter b and *die* at parameter d; they are crucial when we consider the decomposition of modules as direct sums.

Definition 1.16. Let $\mathbb{V} = \left\{V_\epsilon \xrightarrow{v_\epsilon^{\epsilon'}} V_{\epsilon'}\right\}_{\epsilon\leq\epsilon'}$ and $\mathbb{W} = \left\{W_\epsilon \xrightarrow{w_\epsilon^{\epsilon'}} W_{\epsilon'}\right\}_{\epsilon\leq\epsilon'}$ be two persistent vector spaces. The *direct sum* of $\mathbb{V}$ and $\mathbb{W}$ is the persistent vector space given by

$$\mathbb{V} \oplus \mathbb{W} = \left\{V_\epsilon \oplus W_\epsilon \xrightarrow{v_\epsilon^{\epsilon'} \oplus w_\epsilon^{\epsilon'}} V_{\epsilon'} \oplus W_{\epsilon'}\right\}_{\epsilon\leq\epsilon'}$$

A persistent vector space $\mathbb{U}$ is *decomposable* if it can be written as a sum $\mathbb{U} \simeq \mathbb{V} \oplus \mathbb{W}$, where $\mathbb{V}$ and $\mathbb{W}$ are non-zero persistent vector spaces (a persistent vector space $\mathbb{V}$ is non-zero if V_ϵ is a non-trivial vector space, for some $\epsilon \in \mathbb{R}$).

We can consider the category **PVec** whose objects are persistent vector spaces and morphisms $\phi\colon \mathbb{V} \to \mathbb{W}$ are families of linear maps $\phi = \{\phi_\epsilon\colon V_\epsilon \to W_\epsilon\}_{\epsilon\in\mathbb{R}}$ such that the following diagram commutes:

$$\begin{tikzcd} V_\epsilon \arrow[r, "v_\epsilon^{\epsilon'}"] \arrow[d, "\phi_\epsilon"'] & V_{\epsilon'} \arrow[d, "\phi_{\epsilon'}"] \\ W_\epsilon \arrow[r, "w_\epsilon^{\epsilon'}"] & W_{\epsilon'} \end{tikzcd}$$

Two persistent vector spaces are *isomorphic* if ϕ_ϵ is an isomorphism for every $\epsilon \in \mathbb{R}$.

It is possible to define a pseudometric in **PVec**, called the *interleaving distance*:

Definition 1.17 ([13])**.** A homomorphism of degree δ between $\mathbb{V} = \left\{V_\epsilon \xrightarrow{v_\epsilon^{\epsilon'}} V_{\epsilon'}\right\}_{\epsilon\leq\epsilon'}$ and $\mathbb{W} = \left\{W_\epsilon \xrightarrow{w_\epsilon^{\epsilon'}} W_{\epsilon'}\right\}_{\epsilon\leq\epsilon'}$ is a family of maps $\{f_\epsilon\colon \mathbb{V}_\epsilon \to \mathbb{W}_{\epsilon+\delta}\}$ such that the following diagram commutes

$$\begin{tikzcd} V_\epsilon \arrow[r, "v_\epsilon^{\epsilon'}"] \arrow[dr, "f_\epsilon"'] & V_{\epsilon'} \arrow[dr, "f_{\epsilon'}"] & \\ & W_\epsilon \arrow[r, "w_\epsilon^{\epsilon'}"'] & W_{\epsilon'} \end{tikzcd}$$

for all $\epsilon \leq \epsilon'$.

We say that $\mathbb{V}$ and $\mathbb{W}$ are *δ-interleaved* if there are two homomorphisms of degree δ, say $f = \{f_\epsilon\colon \mathbb{V}_\epsilon \to \mathbb{W}_{\epsilon+\delta}\}$ and $g = \{g_\epsilon\colon \mathbb{W}_\epsilon \to \mathbb{V}_{\epsilon+\delta}\}$, such that $f_\epsilon \circ g_{\epsilon+\delta}$ is the identity on V_ϵ and $g_\epsilon \circ f_{\epsilon+\delta}$ is the identity on W_ϵ.

$$\begin{tikzcd} V_\epsilon \arrow[rr, "1_{V_\epsilon}"] \arrow[dr, "f_\epsilon"'] & & V_{\epsilon+2\delta} \\ & W_{\epsilon+\delta} \arrow[ur, "g_{\epsilon+\delta}"'] & \end{tikzcd} \qquad \begin{tikzcd} & V_{\epsilon+\delta} \arrow[dr, "f_{\epsilon+\delta}"] & \\ W_\epsilon \arrow[ur, "g_\epsilon"] \arrow[rr, "1_{W_\epsilon}"'] & & W_{\epsilon+2\delta} \end{tikzcd}$$

Define the *interleaving distance* between $\mathbb{V}$ and $\mathbb{W}$ by

$$d_I(\mathbb{V}, \mathbb{W}) = \inf\{\delta \geq 0 \mid \mathbb{V} \text{ and } \mathbb{W} \text{ are } \delta\text{-interleaved}\}.$$

Notice that if $\mathbb{V}$ and $\mathbb{W}$ are isomorphic then they are 0-interleaved. Informally, d_I is measuring "how far away from being isomorphic" two persistent vector spaces are.

Definition 1.18. Let (X, d_X) be a pair where X is a set (not necessarily finite) and $d_X : X \times X \to \mathbb{R} \cup \{+\infty\}$ is a function. We say that d_X is an *extended pseudometric* if d_X satisfies, for all $x, x', x'' \in X$,

1. $d_X(x, x') \geq 0$.
2. $d_X(x, x') = d_X(x', x)$.
3. $d_X(x, x'') \leq d_X(x, x') + d_X(x', x'')$.

If, moreover, d_X satisfies $d_X(x, x') = 0 \Leftrightarrow x = x'$, then d_X is an extended metric. If, in addition, the codomain of d_X is $\mathbb{R}$ instead of $\mathbb{R} \cup \{+\infty\}$, then d_X is called a *metric*.

With this nomenclature, it can be shown that the interleaving distance d_I defined above is an extended pseudometric on **PVec**. We can always turn it to an extended metric by identifying the elements $\mathbb{V}, \mathbb{W}$ such that $d_I(\mathbb{V}, \mathbb{W}) = 0$.

The next theorem states a simple condition under which $\mathbb{V}$ is decomposable as a sum of interval modules. See [14] for a generalization of this result for a broader class of persistent vector spaces, called *q-tame*.

Theorem 1.19 ([14])**.** Let $\mathbb{V} = \left\{ V_\epsilon \xrightarrow{v_\epsilon^{\epsilon'}} V_{\epsilon'} \right\}_{\epsilon \leq \epsilon'}$ be a finite-dimensional persistent vector space. Then $\mathbb{V}$ is decomposable as a direct sum of interval modules.

This direct sum of interval modules can have repetitions, for example, $\mathbb{V} \simeq \mathbb{I}_{[b,d)} \oplus \mathbb{I}_{[b,d)}$. This motivates the following definition:

Definition 1.20. A *multiset* is a pair $A = (X, m)$ where X is a set and $m : X \to \mathbb{N}^*$ is the *multiplicity* function.

We can regard a multiset as a "set with repetitions", where the number of repetitions of a given element $x \in X$ is given by $m(x)$.

Denote the extended real line $\mathbb{R} \cup \{+\infty\}$ by $\overline{\mathbb{R}}$.

Definition 1.21. Theorem 1.19 states that $\mathbb{V}$ is isomorphic to the sum $\bigoplus_{(b,d) \in A} I_{[b,d)}$ for some multiset A. This multiset is called the *persistence diagram* of $\mathbb{V}$, and denoted by $\mathrm{dgm}(\mathbb{V})$. It can be depicted as a set of points above the diagonal in $\overline{\mathbb{R}}^2$, since every $(b, d) \in \mathrm{dgm}(\mathbb{V})$ satisfies $b \leq d$. An equivalent way to see $\mathrm{dgm}(\mathbb{V})$ is to consider it as a multiset of intervals of $\overline{\mathbb{R}}$, called the *barcode* of $\mathbb{V}$, and denoted by $\mathrm{bar}(\mathbb{V})$: $[b, d) \in \mathrm{bar}(\mathbb{V}) \Leftrightarrow (b, d) \in \mathrm{dgm}(\mathbb{V})$.

When $V_\epsilon \xrightarrow{v_\epsilon^{\epsilon'}} V_{\epsilon'}$ is obtained applying the p-dimensional homology functor to $K_\epsilon \hookrightarrow K_{\epsilon'}$, for some simplicial filtration $\mathcal{K} = \{K_\epsilon\}_{\epsilon \in \mathbb{R}}$, we write $\mathrm{dgm}(\mathbb{V})$ as $\mathrm{dgm}_p(\mathcal{K})$ and $\mathrm{bar}_p(\mathbb{V})$ as $\mathrm{bar}_p(\mathcal{K})$. See Figure 1.2.

The set of barcodes is denoted by **Bar** and the set of persistent diagrams is denoted by **Pers**.

A long bar $[b, d)$ in $\mathrm{bar}_p(\mathcal{K})$ (equivalently, a point (b, d) far away from the diagonal in $\mathrm{dgm}_p(\mathcal{K})$) can be interpreted as a feature that *persisted* along the filtrations: it represents a generator of $H_p(K_b)$ only became a boundary at $H_p(K_d)$. Small bars (points close to the

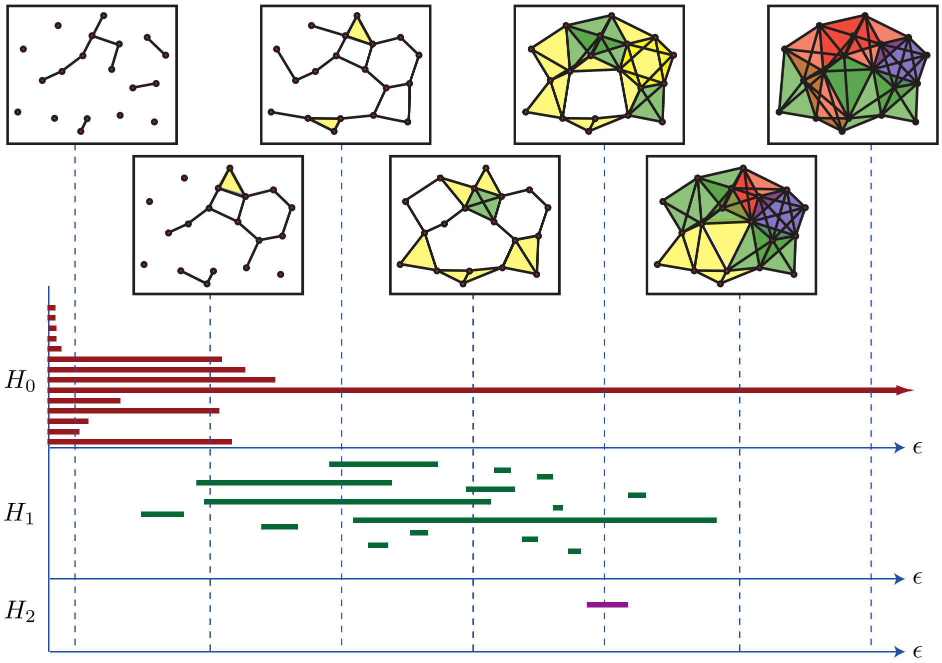


Figure 1.2: On top: the simplicial complex filtration in several different parameters. Below: the corresponding barcode of zero, one and two-dimensional homology. Source of the image: [15].

diagonal in $\mathrm{dgm}_p(\mathcal{K})$), on the other hand, can be seen as *noise*: meaningless information. See Figure 1.3 for an illustration when $\mathcal{K}$ is built over a sample of points in the unit circle.

Denote by **Dgm** the set of all finite multisets (X, m) of $\overline{\mathbb{R}}^2$ such that X is above the diagonal (that is: any $(b, d) \in X$ satisfies $b \leq d$). There is a pseudometric defined in **Dgm** called the *bottleneck distance* and denoted by d_b. For details about it, see [14].

This metric is related to the interleaving distance in the following way:

Theorem 1.22 ([16, Theorem 3.1])**.** Let $\mathbb{V}$ and $\mathbb{W}$ be two finite dimensional persistent vector spaces. Then

$$d_b(\mathrm{dgm}(\mathbb{V}), \mathrm{dgm}(\mathbb{W})) = d_I(\mathbb{V}, \mathbb{W}).$$

This theorem is very useful when we want to calculate an upper bound for the bottleneck distance $d_b(\mathrm{dgm}(\mathbb{V}), \mathrm{dgm}(\mathbb{W}))$: it is enough to find a δ-interleaving between $\mathbb{V}$ and $\mathbb{W}$. For more details about the proof in a more general case, check the very nice book [16].

Now we will define two of the most famous filtrations in TDA: the Vietoris-Rips and the Čech filtrations. In Chapter 4 we will see how these definitions can be extended to networks (see Chapter 3) using the framework of graph filtrations.

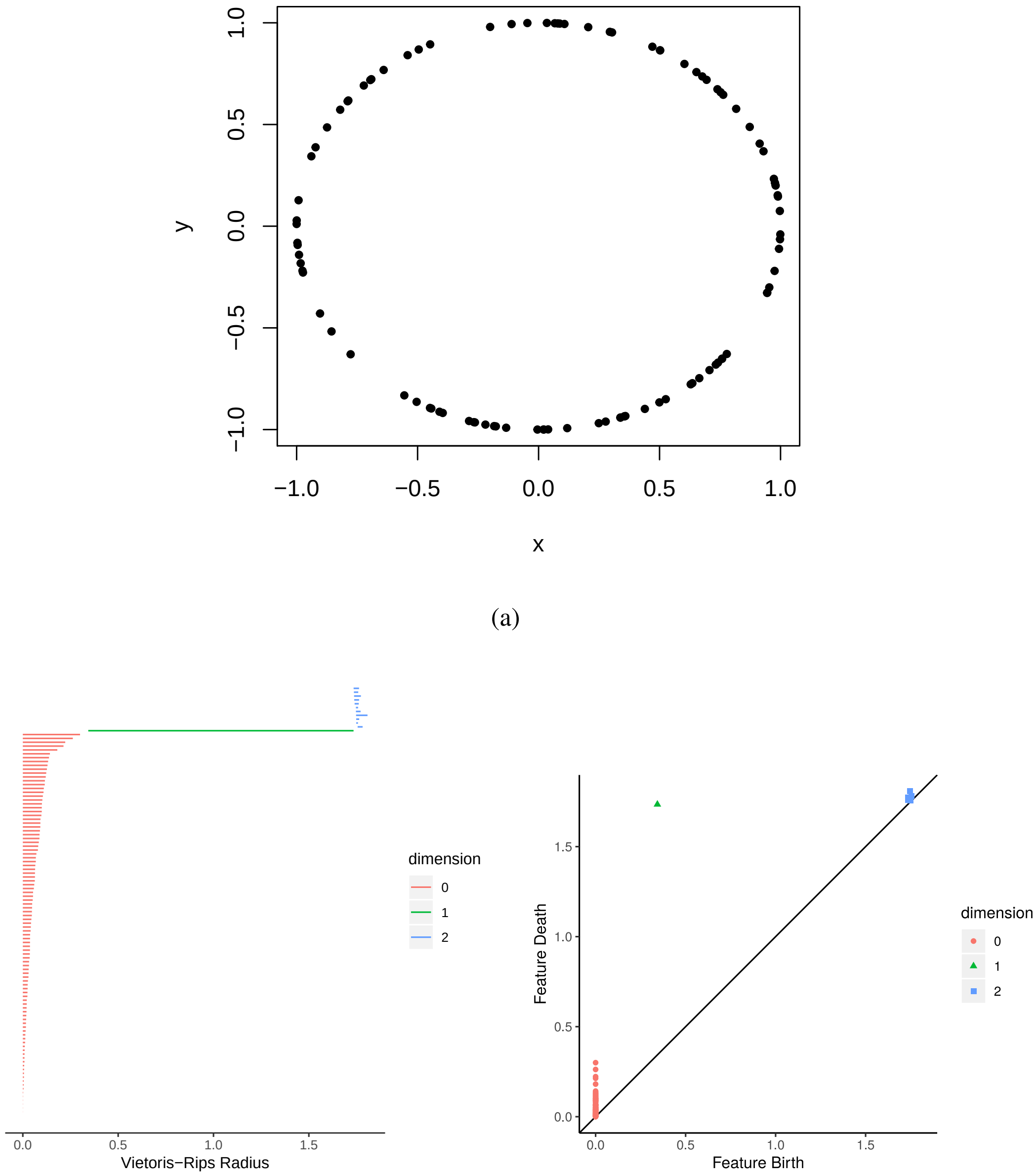
For the next two definitions, let $\mathbf{X} = (X, d)$ be a metric space.

|

Definition 1.23 ([13])**.** Given $\epsilon \in \mathbb{R}$, the *Vietoris-Rips complex of* $\mathbf{X}$ *with radius* ϵ is the simplicial complex $\mathrm{Rips}_\epsilon(\mathbf{X})$ given by

$$\sigma = [x_0, \ldots, x_p] \in \mathrm{Rips}_\epsilon(\mathbf{X}) \Leftrightarrow d(x_i, x_j) \leq \epsilon, \ i, j = 0, \ldots, p.$$

The *Vietoris-Rips filtration* is the simplicial filtration $\mathrm{Rips}(\mathbf{X}) := \{\mathrm{Rips}_\epsilon(\mathbf{X})\}_{\epsilon \in \mathbb{R}}$. See Figure 1.3.



(a)

(b)

(c)

Figure 1.3: (a): $\mathbf{X}$ is a set of 100 random points in a unit circle with the euclidean distance. (b): the barcode of $\mathbf{X}$ via Vietoris-Rips. (c): the persistent diagram. Notice a long bar (and a corresponding point far away from the diagonal) in dimension 1. Image generated in R with the package TDAstats [17].

Definition 1.24. The *Čech complex* of $\mathbf{X}$ with radius ϵ, $\check{\mathrm{C}}\mathrm{ech}_\epsilon(\mathbf{X})$ is given by

$$\sigma = [x_0, \ldots, x_p] \in \check{\mathrm{C}}\mathrm{ech}\,\epsilon(\mathbf{X}) \Leftrightarrow \exists \widetilde{x} \in X \text{ such that } d(x_i, \widetilde{x}) \leq \epsilon,\ i, j = 0, \ldots, p,$$

that is: the balls with center in x_i and radius ϵ have a non-empty intersection.

The *Čech filtration* is the simplicial filtration $\check{\mathrm{C}}\mathrm{ech}(\mathbf{X}) := \{\check{\mathrm{C}}\mathrm{ech}_\epsilon(\mathbf{X})\}_{\epsilon\in\mathbb{R}}$.

It is worth noting that the authors of [18] extended the above definitions to the case of networks and proved stability results. These will be a corollary of Theorem 4.19.

We can summarize the main pipeline in TDA as follows:

1. Consider a set X with some notion of distance or dissimilarity (usually, a metric space).

2. Construct a family of simplicial complexes $\mathcal{K} = \{K_\epsilon\}_{\epsilon\in\mathbb{R}}$ over X.

3. Apply homology to each K_ϵ and obtain a persistent vector space $\mathbb{V}$.

4. Calculate its persistent diagram $\mathrm{dgm}(\mathbb{V})$.

5. Based on $\mathrm{dgm}(\mathbb{V})$, infer some properties of X.

In the present work we will try to extend step 1 to more general objects (called *extended networks*, see Chapter 3) and construct many families of simplicial filtrations in step 2 in a "motivic" manner (see Chapter 4).

1.3 Why TDA?

Gunnar Carlsson summarised the main motivation of TDA in one phrase: "data has shape and shape has meaning". Understanding the geometry of datasets (even the high dimensional ones) can give us insights about them, and enhance the analysis together with some other techniques, such as statistics and machine learning.

The basic tools of Topological Data Analysis were presented in the last two sections. Now let's see how it was born and why it is useful. For this section, let's follow a bit of Perea's "A brief history of persistence" [19].

Early days

TDA's origin can be traced back to [20], where Frosini introduced the concept of "size function": a tool, he says, "to describe and compare 'shapes of objects'". These global shape descriptors were used in [21], where the authors created an efficient content-based image retrieval method. The basic problem was the following: "patent officers have to routinely search large databases of images (the actual number of registered trademarks in the world is enormous and rapidly growing) namely each time a new trademark is submitted for registration, to guarantee that it is sufficiently distinct from existing marks". They associated size functions to each trademark, and then compared these size functions. The results "obtained on a benchmark database of over 10,000 real trademark images supplied by the United Kingdom Patent Office, show that our method can operate in the presence of noise, is computationally efficient, and actually outperforms other tested existing whole-image matching techniques".

Let's take a better look on what they did.

Let S be a topological space and $\phi\colon S \to \mathbb{R}$ a continuous function. We say that the pair (S, ϕ) is a *size pair*, and that ϕ is a *measuring function*. For any $y \in \mathbb{R}$, define the sublevel set

$$S_y = \phi^{-1}((-\infty, y]),$$

and the *(reduced) size function*

$$l(S, \phi)\colon \{(x, y) \in \mathbb{R}^2 \mid x < y\} \to \mathbb{N}$$

by the number of connected components of S_y containing at least one point of S_x. See Figure 1.4.

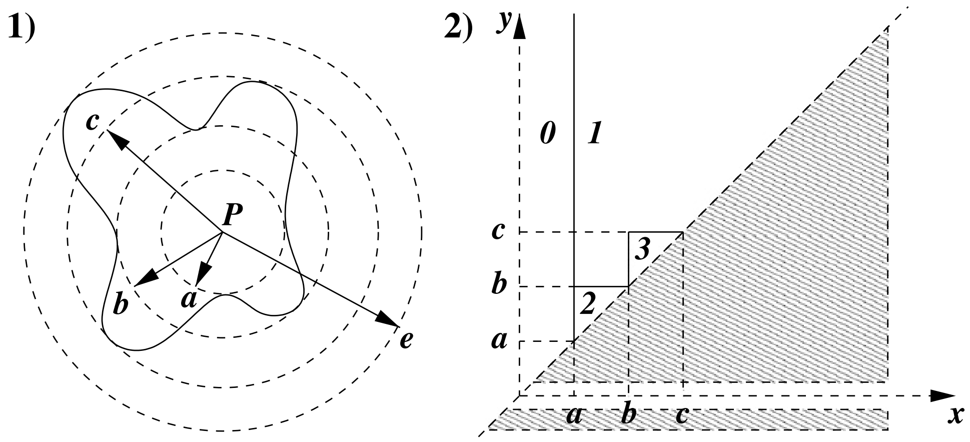


Figure 1.4: 1) A size pair (S, ϕ), where S is a curve in $\mathbb{R}^2$ and $\phi(x)$ is the distance from x to P. 2) For each (x, y) with $x < y$, $l(S, \phi)(x, y)$ is the number of connected components of S_y containing at least one point of S_x. Source of image: [21].

In a more modern language, we can consider the filtration $\{S_x \hookrightarrow S_y\}_{x \le y}$ of topological spaces and its corresponding persistent vector space $\{H_0(S_x) \to H_0(S_y)\}_{y \in \mathbb{R}}$, where H_0 here is the 0-dimensional simplicial homology functor and the linear maps are induced by the inclusions $S_x \hookrightarrow S_y$. For the example in Figure 1.4, the barcode of the corresponding persistent vector space is

$$[a, +\infty),\ [a, b),\ [b, c),\ [b, c).$$

The authors then proceeded by defining 25 functions to be used as measuring functions, and computing, for each trademark, its size functions. Finally, they compared these size functions using a matching distance (or, in our terminology, compared the barcodes using the bottleneck distance).

This is a pipeline that appears many times in TDA: In informal terms, given objects $X_1, \ldots, X_n$ which are hard to compare (images, 3d objects, metric spaces, networks, and so on), we associate to them more tractable objects (persistent diagrams, for example), say $F(X_1), \ldots, F(X_n)$. We then compare these last objects using a distance which is easier to calculate than the distance between the original objects $X_1, \ldots, X_n$. We can't expect the association $X_i \mapsto F(X_i)$ to be an isometry, since $F(X_i)$ is, in some sense, a "simpler" object than X_i. But we can expect that the distance between $F(X_i)$ and $F(X_j)$ is smaller than (or equal to) the distance between X_i and X_j. This kind of upper bound is called a *stability result*, and it is important when applying TDA. Let's make it more precise while we study the content of another very nice paper.

Images, shapes and the importance of stability

In [22], the authors developed some theoretical guarantees regarding stability results for metric spaces and measure metric spaces. As an application, they considered the publicly available dataset obtained in [23]. See Figure 1.5.

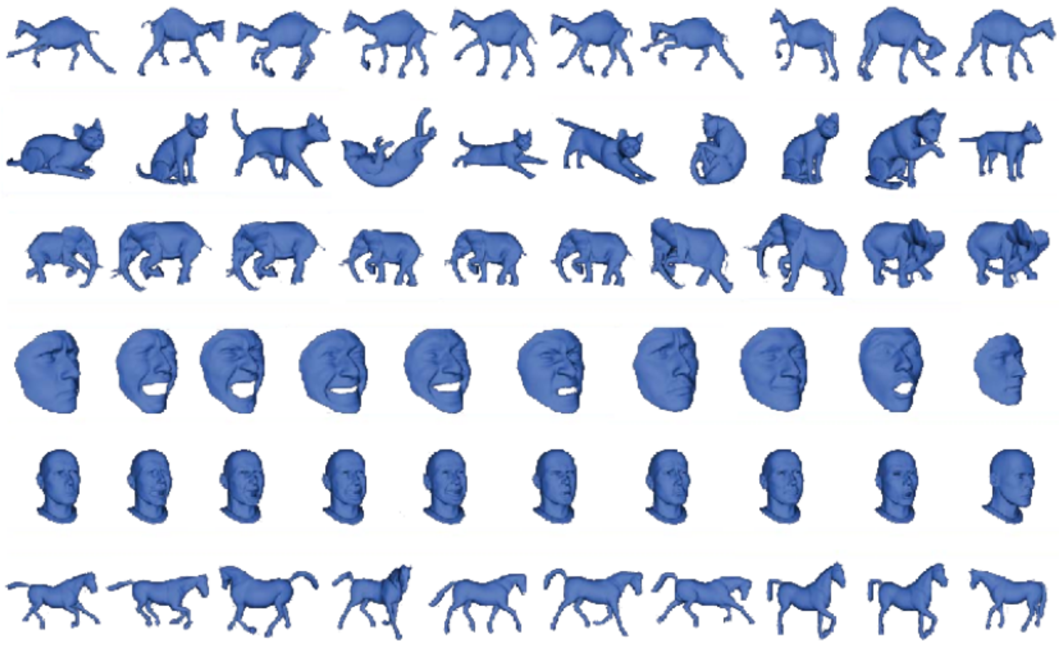


Figure 1.5: A dataset of 3d shapes, in different poses.

This dataset consists of 62 shapes of six different classes: camel, cat, elephant, face, head and horse. Each shape is a list of vertices (in $\mathbb{R}^3$), line segments (edges) and triangles (basically, a 2-dimensional simplicial complex). Shapes of the same class are represented in different poses.

The problem: given a chosen shape among these 62, how can a computer tell to which class this shape belongs, knowing the class of all other shapes?

We can consider each of these shapes as a metric spaces with the euclidean distance, and compute the Gromov-Hausdorff distance (a particular case of the *network distance*, see Chapter 3) between then. This is not useful for two reasons: computing the Gromov-Hausdorff is a NP hard problem (see Chapter 6), which means that the computational time to calculate the distance grows exponentially with the number of points of the metric spaces involved. The second problem is that two shapes of the same class, say two cats in different poses, can have a big Gromov-Hausdorff distance, when both are given the euclidean distance. This is why **the choice of the metric is important**.

Consider, instead, each shape X equipped with the *geodesic distance* d_g^X: it is defined, for each $x, x' \in X$, as the minimum amount of vertices on paths that contain both x and x'. In informal terms, we can think of it as the distance that we have to travel from x to x' while walking on the edges of the shape X. (If X did not have edges, we could create them using some neighborhood graph). For an ant walking on your body right now, the distance from your right hand to your left hand is pretty much the same, whether you have your hands up or down (it is a different story, however, if your hands are touching your body; in this case the ant could follow a shortcut). So, we can expect that the Gromov-Hausdorff distance between the spaces (X, d_g^X) is small when both are of the same class.

The authors then calculated the persistent diagrams in dimensions 1 and 2 of the Vietoris-Rips complex of a simplified version of (X, d_g^X) (equipped with a measure), and used the bottleneck distance as a lower-bound to the Gromov-Wasserstein distance (a generalization of the Gromov-Hausdorff distance for measure metric space. See [22, Section 7] for more details). That is, writing d_{GW} as the Gromov-Wasserstein distance, we have

$$d_b(C(X, d_g^X), C(Y, d_g^Y)) \leq d_{GW}((X, d_g^X), (Y, d_g^Y)),$$

where d_b is the bottleneck distance and $C(X, d_g^X)$ is a simplicial filtration built over (X, d_g^X) (it is a modification of the Vietoris-Rips filtration to the case of measure metric spaces). This *stability* let us use d_b as a lower bound to d_{GW}, and when d_b is big, we are sure that d_{GW} is also big. Thus, we can tell apart shapes that are very different. If, however, d_b is small, we can't assure that d_{GW} is also small; thus, it is important to choose a family of simplicial complexes that have a good "discriminative" power. See more about the creation of simplicial complex families in Chapter 4.

In the end, the authors considered each shape and the closest shape to it (with respect to d_b). A *misclassification* occurs when a shape is of one certain class and its closest shape is of another class. See Figure 1.6

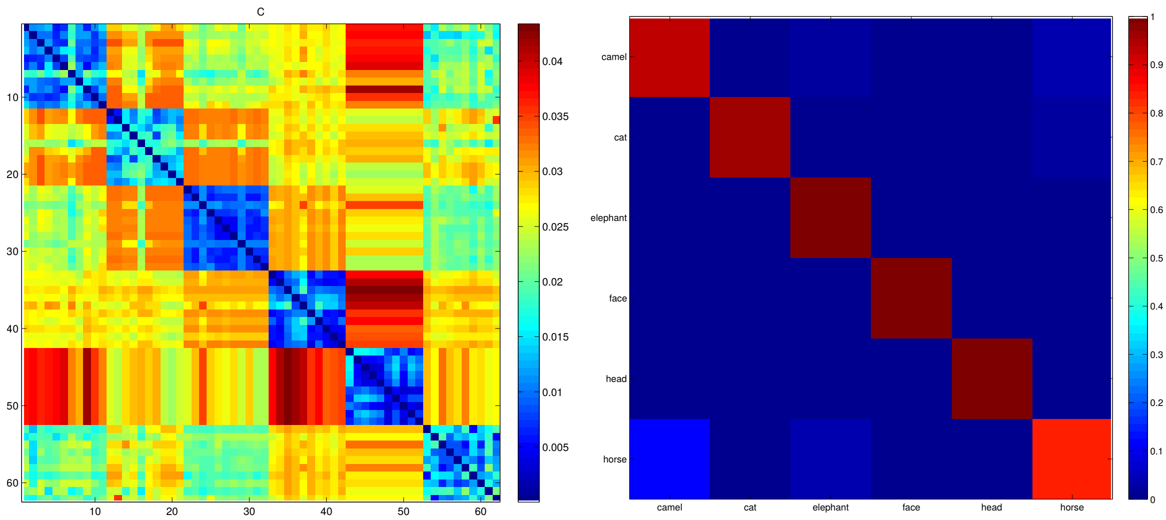


Figure 1.6: On the left: the estimated Gromov-Wasserstein distance between two shapes. On the right: the *confusion matrix*, which shows the probability that a shape from that class is closest to a shape of another class. Source of the image: [22].

As we can see, the classification was really good, even with all the simplifications done before calculating the simplicial complexes (since some of the shapes had more than 20000 points, which is still intractable to calculate today in a usual computer, the authors needed to choose a sample of the vertices). These barcodes can be seen as "signatures" that enable us to identify and classify shapes.

On the same topic: in [24], the authors use ideas from persistent sub-level filtrations to compare 1d signals and denoise them; in [25], the geometry of the space of high-contrast image patches is calculated using persistent homology (surprisingly, a Klein bottle appears!); in [26, 27], persistent diagrams are used to segment images, in order to divide them into several "pieces".

Mapper

The Mapper algorithm is a very distinct approach to TDA, since in its core it makes no use of simplicial filtrations or homology. It can be seen as a discrete version of the *Reeb space*. Given a topological space S and a continuous function $f : S \to \mathbb{R}$ (called a *filter*), the Reeb space (see [28]) is defined as the quotient $S/{\sim_f}$, where $\sim_f$ is the equivalence relation on S given by

$$x \sim_f y \Leftrightarrow [f(x) = f(y) \text{ and } x, y \text{ belong to the same connected component of } f^{-1}(f(x)) = f^{-1}(f(y))].$$

See Figure 1.7 for an example.

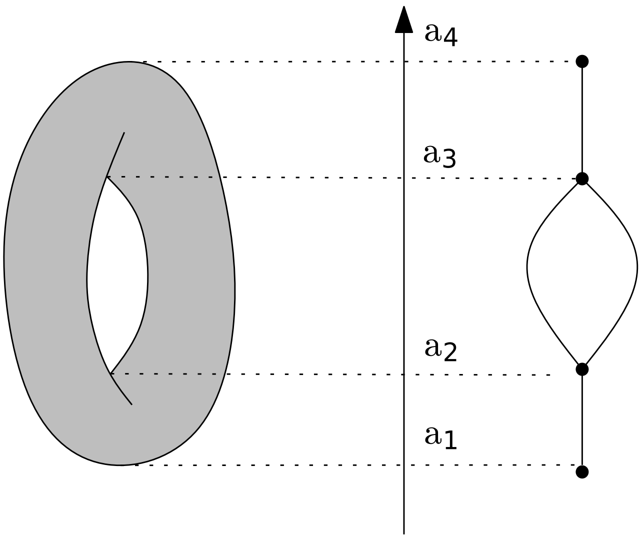


Figure 1.7: The Reeb space of a torus, where f is the projection. Source of the image: [28].

The filter function can be seen as a "lens" through which we see our space (usually high-dimensional). Different filters can give us different informations about the geometry of the space S, and under some conditions, the Reeb space is actually a graph. The Mapper algorithm [29] is a discrete version of the Reeb graph, and to define it the authors had to: 1) change the "connected component" of the Reeb space definition with something that makes more sense with finite data, and 2) change the $f(x) = f(y)$ condition, which rarely occur in practice. Let's make it more precise.

Let (X, d) be a finite metric space and $f\colon X \to \mathbb{R}$ be a function. Consider a covering of $f(X)$ by open intervals, say $I = \{I_1, \ldots, I_n\}$. For each $i = 1, \ldots, n$, consider $J_i = f^{-1}(I_i)$.

Apply a clustering algorithm (usually, the single linkage clustering, with some conditions on the inner distances, see [29] for more details) for each J_i to obtain $\{C^i_1, \dots, C^i_{n_1}\}$. Each C^i_k become a vertex of the Mapper graph, and we draw an edge between two vertices A and B if $A \cap B \neq \emptyset$. See Figure 1.8

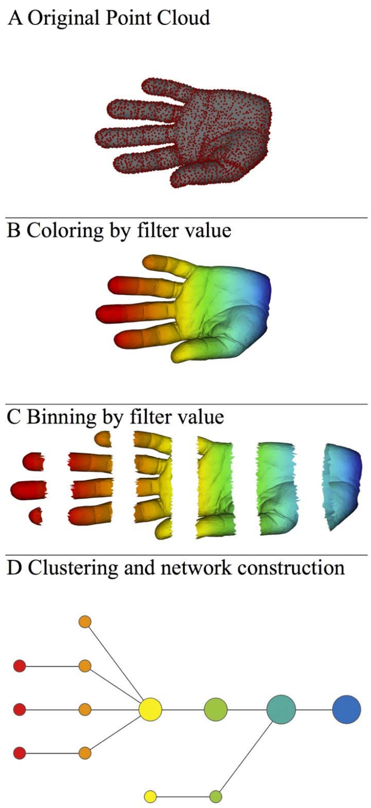


Figure 1.8: The Mapper algorithm. The projection function was used as filter. Source of the image: [30].

In [29], the authors applied the Mapper algorithm to simplify the dataset of shapes used in [22] (the output graph is a sort of "skeleton" of the corresponding shape). See Figure 1.9.

The Mapper algorithm was succesfully used in many cases (see, for example, the medical applications in [30–32]) to detect interesting subgroups of data which could not be discovered using traditional clustering techniques. A generalization with a more "persistent" flavour was developed in [33] and theoretical guarantees were developed in [28, 34–36]. A graphical user interface for the Mapper, made in R language [37] is available at github.com/vituri. No knowledge of programming is needed to use it.

Machine learning

We now go back to the world of persistent diagrams (or, equivalently, barcodes). Suppose, for example, that we take samples 100 samples of points of a torus, say $X_1, \dots, X_{100}$, with 1000 points each, and calculate the corresponding persistent diagrams of the Vietoris-Rips filtration in dimensions 0, 1 and 2. How can we know what is the "mean" persistent diagram in each dimension? Being multisets of the half upper plane, the persistent diagrams do not readily allow us to calculate statistics and means.

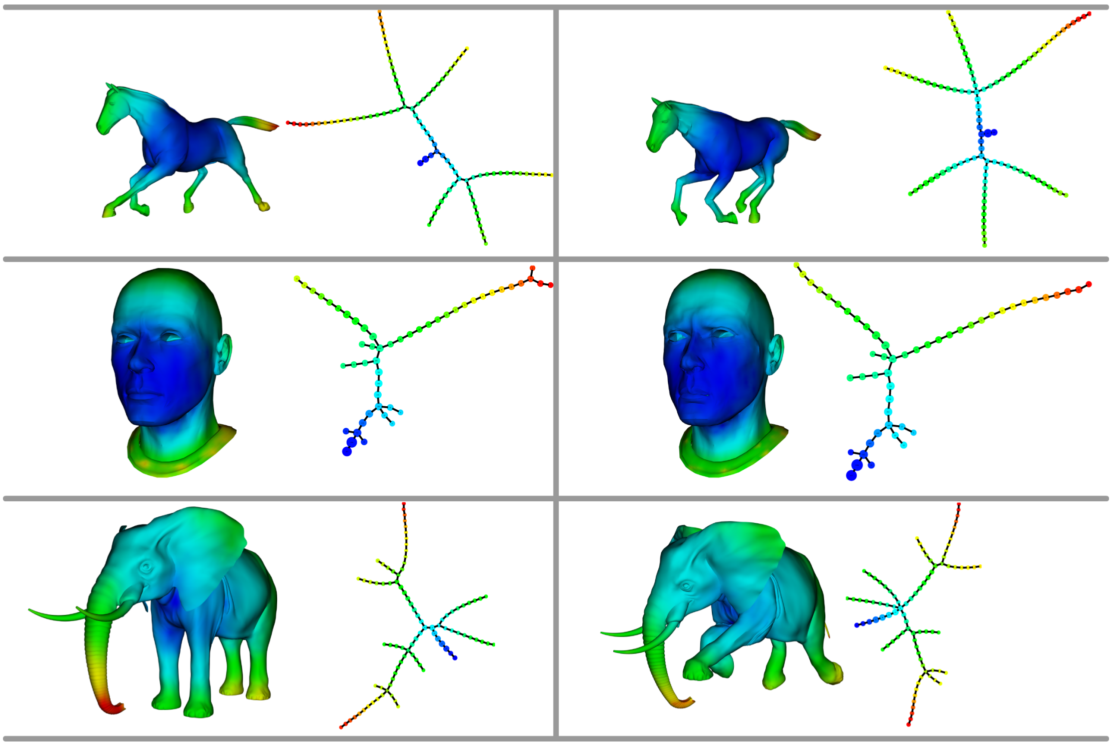


Figure 1.9: The Mapper graph of some 3d shapes (the same dataset used in [22]). The eccentricity function was used as filter. Roughly speaking, this filter measures how far from the center of your metric space a certain point is. The values of the filter function go from blue (lowest) to red (highest). The shapes are also colored according to the filter values. Source of the image: [29].

In [38], Bubenik developed the *persistence landscape*, which is a more tractable object than the persistent diagram. The main idea is to associate a family of simple functions to each persistent diagram. More precisely, let B be a barcode, and define $\beta^{b,d}$ as the number of intervals in B that contains $[b, d)$. The *persistence landscape* of B is a sequence of functions $\lambda_k : \mathbb{R} \longrightarrow \mathbb{R} \cup \{+\infty\}$, where

$$\lambda_k(t) = \sup\{m \geq 0 \mid \beta^{t-m,t+m} \geq k\}.$$

See Figure 1.10 for some graphical representations of the family λ_k for $k = 1, 2, 3$.

Roughly speaking, the persistence landscape is made creating isosceles triangles where two vertices are on the x axis and the third vertex is a point in the (rescaled) persistent diagram.

When the barcode B doesn't have points at infinity, the persistence landscape is a function in a separable Banach (see [38] for more details). This allow us to calculate norms and means. See Figure 1.11.

Even more, we can use persistence landscapes as input in many machine learning methods (see [39, 40], for example). As the authors of [41] point out, "as a persistence landscape is a function, for computational purposes we can convert these functions to matrices by subsampling the persistence landscape in a chosen range of the domain. We may think of this representation as a restriction of the persistence diagram functions to a subsets of the domain, so the addition can still be done pointwisely. In this way, we can simply treat the persistence

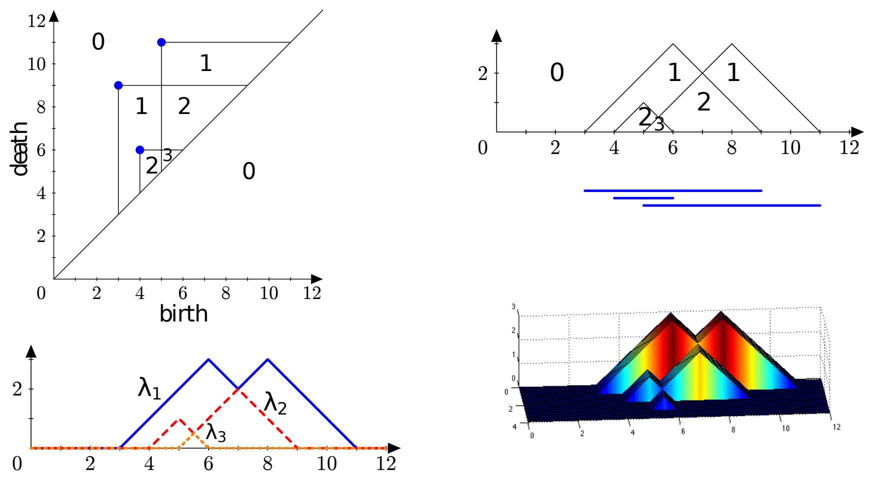


Figure 1.10: Top left: a persistent diagram. Top right: a change of coordinates. Bottom left: the persistence landscapes. Bottom right: a 3d visualization of the persistence landscape, where λ_1 is at the back of the image, and in front of it we see λ_2 and λ_3. Source of the image: [38].

landscape as a finite-size, two-dimensional feature map that can be easily processed by a subsequent convolutional layer in a CNN (convolutional neural network) architecture". This "vectorization" is an important step, since most machine learning methods need vectors (or matrices) as inputs, each of the same size.

Other kinds of vectorization can be made, as in [42], where numbers extracted from the persistent diagrams in dimensions 1 and 2 were used in a linear regression algorithm to predict protein compressibility. More about it can be found in the thesis [43].

A clustering approach using TDA was done in [26], where the authors developed a clustering method called ToMATo defined in terms of persistent diagrams, and applied it to the clustering of proteins that share some similarities.

Where does this thesis fit in the general theory of TDA?

This thesis present a general framework to build graph maps and simplicial complexes determined by basic blocks (graphs) called "motives". These blocks can be seen as simple shapes which we can try to fit in our data (see Chapter 4). Different motives enable us to detect different patterns in data. We then can create many new functors and simplicial complex filtrations by choosing these motives, with some theoretical guarantees (such as stability, see Theorem 4.19). Moreover, our methods work on very general data: sets with a weight function (called *networks*; see Definition 3.1), which generalize the concept of metric space.

TDA was already used in the study of the human brain. In [44], the authors analyzed the activity patterns in the primary visual cortex using the Vietoris-Rips and the Witness complexes and, in their words, "found that the topological structure of activity patterns when the cortex is spontaneously active is similar to those evoked by natural image stimulation and consistent with the topology of a two sphere"; in [45], the authors used persistence landscape to distinguish between pre-seizures and seizures patterns; in [46], persistent diagrams showed

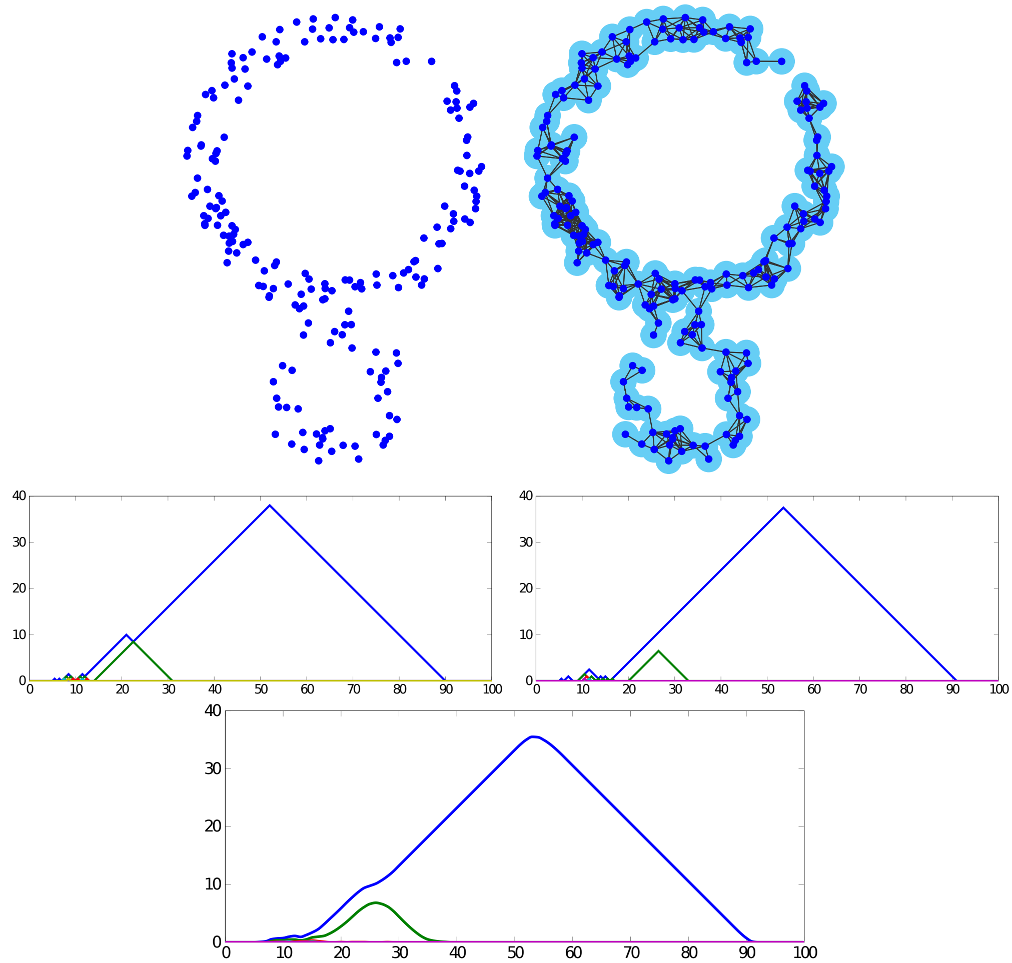


Figure 1.11: Top: 200 points were sample from a pair of annuli, and the Čech simplicial filtration was calculated. This was repeated 100 times. Middle: two persistence landscapes of the 100 obtained above, in dimension 1. Bottom: the mean persistence landscape in dimension 1. Notice that a big triangle and a smaller one indicates two 1-dimensional holes in the dataset, one bigger than the other. Source of the image: [38].

a difference between autistic objects and a control group in the analysis of 2d cortical thickness data; in [47, 48], barcodes were used to distinguish the effects of treatments.

More recent work has dealt with the assymetry of the brain activity. In [49], the authors used the directed clique complex (see Section 4.2) to study how the brain processes information, and to understand "[the] link between neural network structure and its emergent function. [...] We have now come closer to describing such a link by taking the direction of synaptic transmission into account, constructing graphs of a network that reflect the direction of information flow, and analyzing these directed graphs using algebraic topology." Such assimetries were also dealt with in [50], where the authors calculated the path homology (see Section 4.3) of deep-forward networks. Such methods that create ordered simplicial complexes are studied in Chapter 4, with stability guarantees. We also developed motivic clusterings of networks in Chapter 3.

Some more applications

This small list is partially present in [16]:

Genetics A evolutionary tree is not always a *tree* in the graph theory sense, since it can contain lateral gene transfer. In [51], the authors characterized the frequency and scale of such lateral gene transfers in pathogenic bacteria by analyzing the cycles of evolutionaries trees via persistent homology; in [52] a similar approach was used to identify horizontal evolution of viruses.

Time series It is possible to study the geometry of time series with mappings from these series to higher dimensional spaces, and then apply homology to detect circles which represent periodicity. This was studied deeply in [53–55].

Spatial networks Spatial networks are graphs with, possibly, weighted arrows. These are the main object of study in this thesis, and can model many real world problems. In [56], the authors considered cliques (the Vietoris-Rips complexes of Chapter 4) to study the barcode of random, modular and non-modular scale-free networks and networks with exponential connectivity distribution; in [57], persistent homology was used to analyze social and spatial networks, including neurons, genes, online messages, air passengers, Twitter, face-to-face contact and co-authorship; in [58, 59], to study the coverage and hole detection in wireless sensor fields; and in [60] to infer spatial properties of unknown environments via biobots.

And many more See, for example, how TDA was used in the study of plant root systems [61, §IX.4]; the study of the cosmic web and its filamentary structure [62, 63]; the analysis of force networks in granular matter [64] and the analysis of regimes in dynamical systems [65].

2 Graphs and functors

Intellect, peering beyond the star, discovered no Star Maker, but only darkness; no Love, no Power even, but only Nothing. And yet the heart praised.

OLAF STAPLEDON, *THE STAR MAKER*

Graphs are very versatile objects: they can describe structures that have some notion of simmilarity or directionality between pairs. Graphs can be used in applied mathematics to describe, for example, the connections of neurons in the human brain [66] and the compressibility of proteins [42], or in pure mathematics, as category theory [12] and quiver theory [16]. In this chapter we will study functors on graphs that preserve the vertex set, that is, only the arrow set can change under them. These functors will yield clusterings of graphs and reveal the underlying "motivic" geometry of it.

The material of this chapter is mostly a reproduction of the paper [67], co-authored with Facundo Mémoli.

2.1 Background and notation

A (directed) graph is a pair $G = (V, E)$ where V is a finite set and $E \subseteq V \times V$ is such that $E \supseteq \Delta(V) := \{(v, v),\ v \in V\}$. The elements of V are called *vertices*, the elements of E are called *arrows* (or *edges*), and $|G|$ denotes the cardinality of V. Notice that with this definition directed graphs have all self loops. We denote an arrow $(v, v') \in E \setminus \Delta(E)$ by $v \xrightarrow{G} v'$ or, when the context is clear, simply by $v \to v'$. Also, $v \nrightarrow v'$ means $(v, v') \notin E$ and $v \leftrightarrow v'$ means both $v \to v'$ and $v' \to v$. We denote the fact that $v \xrightarrow{G} v'$ or $v = v'$ by $v \overset{G}{\underset{=}{\to}} v'$, or, more simply, by $v \underset{=}{\to} v'$. In all illustrations below we will omit depicting self loops.

To denote that v is a vertex of G, we can write $v \in G$ or $v \in V$.

The category $\mathcal{G}$ of graphs has as objects all graphs and the morphisms are given by

$$\mathrm{Mor}_{\mathcal{G}}(G, G') := \big\{\phi \colon V \to V' \mid (\phi \times \phi)(E) \subseteq E'\big\},$$

that is: $v \xrightarrow{G} v'$ implies $\phi(v) \overset{G'}{\underset{=}{\to}} \phi(v')$, for graphs $G = (V, E), G' = (V', E')$. We call any such map a *graph map*, and denote an element $\phi \in \mathrm{Mor}_{\mathcal{G}}(G, G')$ by $\phi \colon G \to G'$.

Given a graph map $\phi \colon G \to G'$, whenever we want to emphasize that v_1', v_2' are in $\phi(G)$, we will write $\phi \colon G \to (G', v_1', v_2')$. If, even more, we write $\phi \colon (G, v_1, v_2) \to (G', v_1', v_2')$, this will mean that $\phi(v_1) = v_1'$ and $\phi(v_2) = v_2'$.

The *disjoint union of* G *and* G', denoted by $G \sqcup G'$, is the graph with vertex set $V \sqcup V'$ and arrow set $E \sqcup E'$.

Two graphs G and G' are *isomorphic* if there are graph maps $\phi\colon G \to G'$ and $\phi'\colon G' \to G$ such that $\phi \circ \phi'$ and $\phi' \circ \phi$ are the identity maps on G' and G, respectively. Any such ϕ is called an *isomorphism* between G and G'. Thus, in this case, G is obtained from G' by a relabelling of the vertices. Whenever G and G' are isomorphic we may write $G \cong G'$.

When $G = (V, E)$ and $G' = (V', E')$ are graphs with $V \subseteq V'$ and the inclusion map $i\colon G \to G'$ given by $i(v) = v$, $\forall v \in V$, is a graph map, we will denote this simply by $G \hookrightarrow G'$. In this setting, denote by $G' \cap V$ the graph $(V, E' \cap (V \times V))$.

Consider some interesting subcategories of $\mathcal{G}$:

- $\mathcal{G}^{\mathrm{sym}}$, whose objects are *symmetric* graphs (that is: $v \xrightarrow{G} v'$ implies $v' \xrightarrow{G} v$).
- $\mathcal{G}^{\mathrm{trans}}$, whose objects are *transitive* graphs (that is: $v \xrightarrow{G} v'$ and $v' \xrightarrow{G} v''$ implies $v \overset{G}{\underset{=}{\to}} v''$).
- $\mathcal{G}^{\mathrm{clust}} = \mathcal{G}^{\mathrm{sym}} \cap \mathcal{G}^{\mathrm{trans}}$, whose objects are symmetric and transitive graphs, which later we will regard as encoding a *clustering* of their sets of vertices.

Some standard notions of connectivity on graphs are the ones that follow. A pair of vertices (v, v') on a graph $G = (V, E)$ is called:

- *strongly connected* if there is a sequence $v_1, \ldots, v_k \in V$, such that $v = v_1$, $v' = v_k$, and $v_i \to v_{i+1}$ for each i. We denote this by $v \rightsquigarrow v'$ in G. If, moreover, $v_k \xrightarrow{G} v_1$, the sequence $v_1, \ldots, v_k, v_1$ is a *cycle* of size k.
- *weakly connected* if there is a sequence $v_1, \ldots, v_k \in V$, such that $v = v_1$, $v' = v_k$, and $v_i \to v_{i+1}$ or $v_{i+1} \to v_i$, for each i.

Given two sets $A, B \subseteq V \times V$, define

$$A \otimes B := \{(v, v') \in V \times V \mid \exists v_1 \in V \text{s.t.} (v, v_1) \in A \text{ and } (v_1, v') \in B).$$

Now, given a graph $G = (V, E)$, one defines $E^{(2)} := E \otimes E$ and, in general, for $m \in \mathbb{N}$,

$$E^{(m+1)} := E^{(m)} \otimes E.$$

If N is such that $E^{(m)} = E^{(N)}$ for all $m > N$, define $E^{(\infty)} := E^{(N)}$. Notice that $(v, v') \in E^{(\infty)} \Leftrightarrow v \rightsquigarrow v'$.

Here are some important graphs that will appear several times in the text:

- K_n is the *complete graph* with n vertices $a_1, \ldots, a_n$ and all possible arrows.
- D_n is the *discrete graph* (or *totally disconnected graph*) with n vertices $a_1, \ldots, a_n$ and no arrows.
- For a given graph $G = (V, E)$, we will denote by $K_{(G)}$ and $D_{(G)}$ the complete graph and the totally disconnected graph with vertex set V, respectively.
- L_n is the *line graph* with n vertices $a_1, \ldots, a_n$ and arrows $a_i \to a_{i+1}$, $i = 1, \ldots, n - 1$.

- T_n is the *transitive line graph* with n vertices $a_1, \ldots, a_n$ and arrows $a_i \to a_j$, for any $1 \le i < j \le n$.
- C_n is the *cycle graph* obtained by adding the arrow $a_n \to a_1$ to L_n.

For any of the above graphs, its vertices will be called $a_1, \ldots, a_n$ unless stated otherwise.

2.2 Endofunctors

Definition 2.1. A functor $\mathfrak{F}\colon \mathcal{G} \to \mathcal{G}$ is called *vertex preserving* if for any graph $G = (V, E) \in \mathcal{G}$, the graph $\mathfrak{F}(G)$ has vertex set V and, if given any graph map $\phi\colon G \to G'$, we have $\mathfrak{F}(\phi) = \phi$. We will henceforth denote by $\mathrm{Funct}(\mathcal{G}, \mathcal{G})$ the collection of all such functors. *All functors in this work are assumed to be vertex preserving.* Whenever we say that $\mathfrak{F}$ is an endofunctor, we mean that $\mathfrak{F} \in \mathrm{Funct}(\mathcal{G}, \mathcal{G})$.

That $\mathfrak{F}$ is a functor means that for every $G, G' \in \mathcal{G}$ and every graph map $\phi\colon G \to G'$, we have graphs $\mathfrak{F}(G)$ and $\mathfrak{F}(G')$, and the map $\mathfrak{F}(\phi)\colon \mathfrak{F}(G) \to \mathfrak{F}(G')$ is a graph map.

$$\begin{tikzcd} G \arrow[r, "\phi"] \arrow[d, Rightarrow, "\mathfrak{F}"] & G' \arrow[d, Rightarrow, "\mathfrak{F}"] \\ \mathfrak{F}(G) \arrow[r, "\mathfrak{F}(\phi)"] & \mathfrak{F}(G') \end{tikzcd}$$

We will regard two endofunctors $\mathfrak{F}_1, \mathfrak{F}_2$ as equal when $\mathfrak{F}_1(G) = \mathfrak{F}_2(G)$, for all $G \in \mathcal{G}$.

We say that $\mathfrak{F} \in \mathrm{Funct}(\mathcal{G}, \mathcal{G})$ is *symmetric* if $\mathfrak{F}(\mathcal{G}) \subseteq \mathcal{G}^{\mathrm{sym}}$, and that $\mathfrak{F}$ is *transitive* (resp. *clustering*) if $\mathfrak{F}(\mathcal{G}) \subseteq \mathcal{G}^{\mathrm{trans}}$ (resp. $\mathfrak{F}(\mathcal{G}) \subseteq \mathcal{G}^{\mathrm{clust}}$).

Definition 2.2. Given two endofunctors $\mathfrak{F}_1$ and $\mathfrak{F}_2$, define $\mathfrak{F}_1 \cup \mathfrak{F}_2(G) = (V, E_1 \cup E_2)$, where $G = (V, E)$, $\mathfrak{F}_1(G) = (V, E_1)$ and $\mathfrak{F}_2(G) = (V, E_2)$.

Example 2.3. Here are some endofunctors that will be used in the sequel:

- **Full disconnection**: $\mathfrak{F}^{\mathrm{disc}}$ taking $G = (V, E)$ to the totally disconnected graph $D_{(G)}$, that is, $\mathfrak{F}(G) = (V, \Delta(V))$.
- **Connected component**: $\mathfrak{F}^{\mathrm{conn}}$, where $v \xrightarrow{\mathfrak{F}^{\mathrm{conn}}(G)} v'$ if v and v are weakly connected.
- **Full completion**: $\mathfrak{F}^{\mathrm{comp}}$ taking $G = (V, E)$ to complete graph $K_{(G)} = (V, V \times V)$.
- **Reversion**: $\mathfrak{F}^{\mathrm{rev}}$ taking (V, E) to (V, E^{rev}), where $E^{\mathrm{rev}} = \{(v', v) \mid (v, v') \in E\}$.
- **Lower symmetrization**: $\mathfrak{F}^{\mathrm{ls}}$ taking (V, E) to $(V, E \cap E^{\mathrm{rev}})$.
- **Identity**: $\mathfrak{F}^{\mathrm{id}}$ the identity endofunctor.
- **Upper symmetrization**: $\mathfrak{F}^{\mathrm{us}}$ taking (V, E) to $(V, E \cup E^{\mathrm{rev}})$.
- **m-Power**: for $m \in \mathbb{N}$, $\mathfrak{F}^{[m]}$ taking (V, E) to $(V, E^{(m)})$.
- **Transitive closure**: $\mathfrak{F}^{\mathrm{tc}}$ taking $G = (V, E)$ to $(V, E^{(\infty)})$, that is: $v \xrightarrow{\mathfrak{F}^{\mathrm{tc}}(G)} v'$ if $v \rightsquigarrow v'$ in G.

Remark 2.4. The "inversion map" given by $(V, E) \mapsto (V, E^{\text{inv}})$, where $E^{\text{inv}} = \Delta(V) \cup (V \times V \backslash E)$, is not a functor. To see why it fails, just consider the inclusion $D_2 \hookrightarrow K_2$.

Definition 2.5. Define the following partial order on $\text{Funct}(\mathcal{G}, \mathcal{G})$: $\mathfrak{F}_1 \leq \mathfrak{F}_2 \Leftrightarrow \mathfrak{F}_1(G) \hookrightarrow \mathfrak{F}_2(G)$ for all $G \in \mathcal{G}$.

Definition 2.6. An endofunctor $\mathfrak{F}$ is called *arrow increasing* if for any $G \in \mathcal{G}$, $v \xrightarrow{G} v'$ implies $v \xrightarrow{\mathfrak{F}(G)} v'$, that is, $G \hookrightarrow \mathfrak{F}(G)$. According to Definition 2.5, this is equivalent to $\mathfrak{F}^{\text{id}} \leq \mathfrak{F}$. Analogously, we say that $\mathfrak{F}$ is called *arrow decreasing* if $\mathfrak{F} \leq \mathfrak{F}^{\text{id}}$.

Remark 2.7. It is clear that $\mathfrak{F}^{\text{ls}} \leq \mathfrak{F}^{\text{id}} \leq \mathfrak{F}^{\text{us}}$. Notice that if $\mathfrak{F}$ is arrow increasing, then $\mathfrak{F}(L_2) \in \{L_2, K_2\}$. This condition is also sufficient, as we prove next.

Proposition 2.8. Let $\mathfrak{F}$ be an endofunctor. Then, $\mathfrak{F}$ is arrow increasing $\Leftrightarrow \mathfrak{F}(L_2) \in \{L_2, K_2\}$.

Proof. Let G be a graph and suppose $v \xrightarrow{G} v'$. The $\phi\colon (L_2, a_1, a_2) \to (G, v, v')$ is a graph map. By functoriality, $\phi\colon (\mathfrak{F}(L_2), a_1, a_2) \to (\mathfrak{F}(G), v, v')$ is a graph map. If $\mathfrak{F}(L_2) = L_2$ or K_2, then $v \xrightarrow{\mathfrak{F}(G)} v'$. Thus, $G \hookrightarrow \mathfrak{F}(G)$. □

Remark 2.9. Even when $\mathfrak{F}(L_2) = K_2$, we cannot ensure that $\mathfrak{F}$ is symmetric. Indeed, let $\mathfrak{F} = \mathfrak{F}^{\text{us}} \cup \mathfrak{F}^{\text{tc}}$. Then $\mathfrak{F}(L_2) = K_2$ but $\mathfrak{F}(L_3)$ is not symmetric. See Figure 2.1.

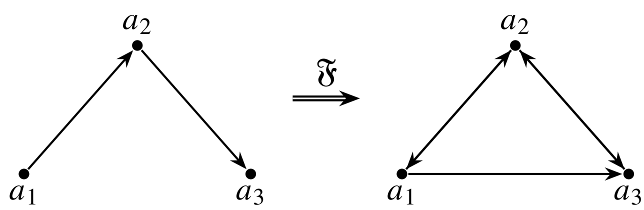


Figure 2.1: An example where $\mathfrak{F}(L_2) = K_2$ but $\mathfrak{F}$ is not symmetric. See Remark 2.9

Similarly to Proposition 2.8, we can obtain some information about $\mathfrak{F}$ by applying it to graphs with just two vertices, as described in the following proposition.

Proposition 2.10. Let $\mathfrak{F}$ be an endofunctor. Then:

1. $\mathfrak{F}(D_2) \neq D_2 \Leftrightarrow \mathfrak{F} = \mathfrak{F}^{\text{comp}}$.
2. $\mathfrak{F}(K_2) \neq K_2 \Leftrightarrow \mathfrak{F} = \mathfrak{F}^{\text{disc}}$.

Proof. (1) First notice that if $\mathfrak{F}(D_2) \neq D_2$, then $\mathfrak{F}(D_2) = K_2$. Indeed, suppose $\mathfrak{F}(D_2)$ has just one arrow. Let p be the graph map $p\colon (D_2, a_1, a_2) \to (D_2, a_2, a_1)$. By functoriality, $\mathfrak{F}(D_2)$ must have both arrows $a_1 \to a_2$ and $a_2 \to a_1$.

Now let $G = (V, E) \in \mathcal{G}$ be a graph with $|G| \geq 2$, and let $v, v' \in G$. Consider the graph map $\phi\colon (D_2, a_1, a_2) \to (G, v, v')$. Applying $\mathfrak{F}$, we obtain $\phi\colon (\mathfrak{F}(D_2), a_1, a_2) \to (\mathfrak{F}(G), v, v')$. Since $\mathfrak{F}(D_2) = K_2$, we have $v \xleftrightarrow{\mathfrak{F}(G)} v'$. Hence, $\mathfrak{F}(G) = K_{(G)}$.

(2) With the same argument used in the previous item, we can show that if $\mathfrak{F}(K_2) \neq K_2$, then $\mathfrak{F}(K_2) = D_2$. Now suppose there is $G \in \mathcal{G}$ such that $\mathfrak{F}(G) \neq D_{(G)}$. Let $v, v' \in G$ such that $v \xrightarrow{\mathfrak{F}(G)} v'$. Consider the graph map $\phi\colon (G, v, v') \to (K_2, a_1, a_2)$ given by $\phi(v) = a_1$ and $\phi(x) = a_2$, for any $x \neq v$. By functoriality, we have a graph map $\phi\colon (\mathfrak{F}(G), v, v') \to (D_2, a_1, a_2)$. But then we cannot have $v \xrightarrow{\mathfrak{F}(G)} v'$. This contradiction finishes the proof. □

Corollary 2.11. If $\mathfrak{F} \neq \mathfrak{F}^{\mathrm{disc}}$, then $\mathfrak{F}(K_n) = K_n$, for any n.

Proof. Given any $a_i, a_j \in K_n$, we can consider the graph map $\phi\colon (K_2, a_1, a_2) \to (K_n, a_i, a_j)$. By functoriality, $\phi\colon (\mathfrak{F}(K_2), a_1, a_2) \to (\mathfrak{F}(K_n), a_i, a_j)$ is a graph map. Since $\mathfrak{F}(K_2) = K_2$ by Proposition 2.10, we have $a_i \xleftrightarrow{\mathfrak{F}(K_n)} a_j$. □

The next proposition is a simple characterization of the transitive closure functor.

Proposition 2.12. Let $\mathfrak{F}\colon \mathcal{G} \to \mathcal{G}^{\mathrm{trans}}$ be a functor such that $\mathfrak{F}(D_2) = D_2$ and $\mathfrak{F}(L_2) = L_2$. Then, $\mathfrak{F} = \mathfrak{F}^{\mathrm{tc}}$.

Lemma 2.13. Let $G = (V, E) \in \mathcal{G}^{\mathrm{trans}}$. Suppose there exists a pair $(v, v') \notin E$. Then there exists a partition $\{C, C'\}$ of V into two non-empty sets with $v \in C$ and $v' \in C'$ such that $(c, c') \notin E$ for all $c \in C$ and $c' \in C'$.

Proof of Proposition 2.12. Pick any graph $G = (V, E)$. Notice that by Theorem 2.8, $\mathfrak{F}$ is arrow increasing.

Assume $v \xrightarrow{\mathfrak{F}^{\mathrm{tc}}(G)} v'$. Then $v \rightsquigarrow v'$ in G, and since $\mathfrak{F}$ is arrow increasing, $v \rightsquigarrow v'$ in $\mathfrak{F}(G)$. Thus, $\mathfrak{F}^{\mathrm{tc}} \leq \mathfrak{F}$.

Now assume that $v \nrightarrow v'$ in $\mathfrak{F}^{\mathrm{tc}}(G)$. By Lemma 2.13 we obtain a partition $\{C, C'\}$ of V with $v \in C$, $v' \in C'$ and the property that $c \nrightarrow c'$ in $\mathfrak{F}^{\mathrm{tc}}(G)$ for all $c \in C$ and $c' \in C$, which implies that $c \nrightarrow c'$ in G, since $G \hookrightarrow \mathfrak{F}^{\mathrm{tc}}(G)$.

Consider the graph map $\phi\colon (G, v, v') \to (L_2, a_2, a_1)$ such that $\phi(C) = a_2$ and $\phi(C') = a_1$. Applying $\mathfrak{F}$, we obtain the graph map $\phi\colon (\mathfrak{F}(G), v, v') \to (L_2, a_2, a_1)$. Thus, $v \nrightarrow v'$ in $\mathfrak{F}(G)$. Hence, $\mathfrak{F}(G) = \mathfrak{F}^{\mathrm{tc}}(G)$. □

Proof of Lemma 2.13. Assume the claim is not true. Then, for any partition $\{C, C'\}$ of V with $v \in C$, $v' \in C'$ there exists some $c \in C$ and $c' \in C'$ with $c \xrightarrow{G} c'$.

Consider first $C_1 = \{v\}$ and $C_1' = V \backslash C_1$. Let $v_1 \in C_1'$ be such that $v \xrightarrow{G} v_1$.

Now, consider $C_2 = \{v, v_1\}$ and $C_2' = V \backslash C_2$. One obtains $v_2 \in C_2'$ such that either $v \xrightarrow{G} v_2$ or $v_1 \xrightarrow{G} v_2$. Since $v \xrightarrow{G} v_1$ and G is transitive, in either case $v \xrightarrow{G} v_2$.

Recursively define $C_j = \{v, v_1, \ldots, v_{j-1}\}$ and $C_j' = V \backslash C_j$ for $j \geq 1$. At each step we obtain $v_j \in C_j'$ such that $v \xrightarrow{G} v_j$. If $v_j = v'$ for some j we have a contradiction.

Furthermore, since at each step $v_j \notin C_j$, the process must end when C_j' contains only one element. Thus, at some step in the process we must have $v_j = v'$. □

Corollary 2.14. If $\mathfrak{F}$ is transitive and arrow increasing, then $\mathfrak{F}^{\mathrm{tc}} \leq \mathfrak{F}$.

Proof. Given $G \in \mathcal{G}$,

$$v \xrightarrow{\mathfrak{F}^{\mathrm{tc}}(G)} v' \Leftrightarrow v \rightsquigarrow v' \text{ in } G \;\Rightarrow v \rightsquigarrow v' \text{ in } \mathfrak{F}(G) \Rightarrow v \xrightarrow{\mathfrak{F}(G)} v'. \qquad \square$$

Remark 2.15. Not all functors $\mathfrak{F}$ satisfy $\mathfrak{F}(\mathcal{G}^{\mathrm{trans}}) \subseteq \mathcal{G}^{\mathrm{trans}}$. Take, for example, $\mathfrak{F}^{\mathrm{us}}$ and $G = \bullet \leftarrow \bullet \rightarrow \bullet$. Then $\mathfrak{F}^{\mathrm{us}}(G) = \bullet \leftrightarrow \bullet \leftrightarrow \bullet$ which is not transitive.

It turns out that in the case of $\mathcal{G}^{\mathrm{sym}}$ we indeed have $\mathfrak{F}(\mathcal{G}^{\mathrm{sym}}) \subset \mathcal{G}^{\mathrm{sym}}$ for any $\mathfrak{F} \in \mathrm{Funct}(\mathcal{G}, \mathcal{G})$. To prove this, we need the following lemma.

Lemma 2.16. For any $G = (V, E) \in \mathcal{G}^{\mathrm{sym}}$ and $v, v' \in G$, there is a graph map $\phi\colon (G, v, v') \to (G, v', v)$.

Proof. Suppose G has connected components $C_1, \dots, C_k$ (recall that for symmetric graphs both notions of connectivity coincides).

If v and v' are in different connected components, say $v \in C_1$ and $v' \in C_2$, define $\phi|_{C_1} \equiv v'$ and $\phi|_{C_i} \equiv v$, $i \neq 1$, and we are done.

If v and v' are in the same component, say $v, v' \in C_1$, we define $\phi|_{C_i} \equiv v$, $i \neq 1$, and the problem reduces to defining ϕ on C_1. Hence, we can suppose that G is connected.

Let $H = (V', E')$ be a connected subgraph of G containing v and v', with the minimum number of vertices possible. It is clear that H is isomorphic to $\mathfrak{F}^{\mathrm{us}}(L_{n+1})$ for some $n \geq 1$. Let $V' = \{x_0, \dots, x_n\}$ with $x_0 = v$, $x_n = v'$ and $x_i \overset{H}{\longleftrightarrow} x_{i+1}$, $i = 0, \dots, n-1$.

For any $x, y \in V$, define $d(x, y)$ as the number of arrows in the shortest path connecting x to y. Let $r\colon G \to H$ be defined by $r(x) = x_k$ where $k = \min\{d(x, v), n\}$ (notice that $n = d(v, v')$). We claim that r is a graph map. Indeed, let $x, y \in V$ such that $x \xrightarrow{G} y$. Suppose $m = d(x, v) \leq d(y, v) = m'$. If $m = m'$, then $r(x) = r(y)$. If $m' = m + 1 \leq n$, then $r(x) = x_m \overset{H}{\longleftrightarrow} x_{m+1} = r(y)$. If $m \geq n$, then $r(x) = r(y) = r_n$. This proves the claim.

Finally, let $\phi = \iota \circ f \circ r\colon G \to G$, where $f\colon H \to H$ is the graph map given by $f(x_i) = x_{n-i}$ and $\iota\colon H \to G$ is the inclusion. Thus, ϕ is a graph map satisfying $\phi(v) = v'$ and $\phi(v') = v$. □

Theorem 2.17. Let $\mathfrak{F}$ be any endofunctor. Then, $\mathfrak{F}(\mathcal{G}^{\mathrm{sym}}) \subset \mathcal{G}^{\mathrm{sym}}$.

Proof. Let $G \in \mathcal{G}^{\mathrm{sym}}$ and $v \xrightarrow{\mathfrak{F}(G)} v'$. Let $\phi\colon (G, v, v') \to (G, v', v)$ be the graph map from Lemma 2.16. Applying $\mathfrak{F}$, we obtain the graph map $\phi\colon (\mathfrak{F}(G), v, v') \to (\mathfrak{F}(G), v', v)$, which implies $v' \xrightarrow{\mathfrak{F}(G)} v$. □

Definition 2.18. An endofunctor $\mathfrak{F}$ is *additive* if

$$\mathfrak{F}\left(\bigsqcup_{i \in I} G_i\right) = \bigsqcup_{i \in I} \mathfrak{F}(G_i)$$

for all finite collections $\{G_i, i \in I\}$ of graphs.

Proposition 2.19. All endofunctors $\mathfrak{F} \in \mathrm{Funct}(\mathcal{G}, \mathcal{G})$ except $\mathfrak{F}^{\mathrm{comp}}$ are additive.

Proof. First consider the case $|I| = 2$. Write $G_i = (V_i, E_i)$ and $\mathfrak{F}(G_i) = (V, E_i^{\mathfrak{F}})$ for $i = 1, 2$. Also, let $V = V_1 \sqcup V_2$, $E = E_1 \sqcup E_2$, $G = G_1 \sqcup G_2 = (V, E)$ and write $\mathfrak{F}(G) = (V, E^{\mathfrak{F}})$. We will prove that $E^{\mathfrak{F}} = E_1^{\mathfrak{F}} \sqcup E_2^{\mathfrak{F}}$.

Let $\phi\colon G \to D_2$ be the graph map given by $\phi(v_1) = a_1$ for all $v_1 \in V_1$ and $\phi(v_2) = a_2$, for all $v_2 \in V_2$. Since $\mathfrak{F}(D_2) = D_2$ (because $\mathfrak{F} \neq \mathfrak{F}^{\mathrm{comp}}$), we cannot have $v_1 \xrightarrow{\mathfrak{F}(G)} v_2$ with $v_1 \in V_1$ and $v_2 \in V_2$. □

2.3 Representable endofunctors

Definition 2.20. Given a family Ω of graphs, we consider the functor $\mathfrak{F}^{\Omega}\colon \mathcal{G} \to \mathcal{G}$ defined as follows: given $G = (V, E)$, $\mathfrak{F}^{\Omega}(G) = (V, E^{\Omega})$, where $(v, v') \in E^{\Omega}$ if and only if there exist $\omega \in \Omega$ and a graph map $\phi\colon \omega \to (G, v, v')$ (this means that $v, v' \in \phi(\omega)$, as defined in Section 2.1). See Figure 2.2. Also, set $\mathfrak{F}^{\Omega}(\phi) = \phi$ for all graph maps $\phi\colon G \to G'$.

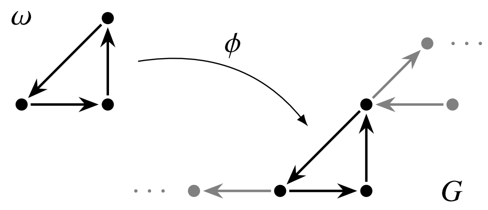


Figure 2.2: A graph map $\phi\colon \omega \to G$.

Definition 2.21. We say that an endofunctor $\mathfrak{F}$ is *representable* (or *motivic*) whenever there exists a family Ω of graphs such that $\mathfrak{F} = \mathfrak{F}^{\Omega}$. In this case we say that $\mathfrak{F}$ is *represented* by Ω, the set of *representers* (or *motifs*) of $\mathfrak{F}$.

The next proposition is immediate.

Proposition 2.22. Assume $\mathfrak{F}$ is represented by a family Ω. Then, $\mathfrak{F}^{\Omega} = \bigcup_{\omega\in\Omega} \mathfrak{F}^{\{\omega\}}$. Thus, if $\Omega_1 \subseteq \Omega_2$ are two families of graphs, we have $\mathfrak{F}^{\Omega_1} \leq \mathfrak{F}^{\Omega_2}$.

Example 2.23. There are several interesting examples of representable functors:

- Let $\Omega = \{D_1\}$, then $\mathfrak{F}^{\Omega} = \mathfrak{F}^{\mathrm{disc}}$.
- Let Ω be the set of all graphs with vertex set $\{x_1, \ldots, x_n\}$ and such that for each $i = 1, \ldots, n-1$ it holds exactly one of the following conditions: $x_i \to x_{i+1}$ or $x_{i+1} \to x_i$. These graphs are called *zig-zag graphs*. Then $\mathfrak{F}^{\Omega} = \mathfrak{F}^{\mathrm{conn}}$.
- If $\Omega = \{D_2\}$, then $\mathfrak{F}^{\Omega} = \mathfrak{F}^{\mathrm{comp}}$.
- If $\Omega = \{K_2\}$, then $\mathfrak{F}^{\Omega} = \mathfrak{F}^{\mathrm{ls}}$.
- If $\Omega = \{L_2\}$, then $\mathfrak{F}^{\Omega} = \mathfrak{F}^{\mathrm{us}}$.
- If $\Omega = \{\bullet \leftarrow \bullet \to \bullet\}$ or $\Omega = \{\bullet \to \bullet \leftarrow \bullet\}$, $\mathfrak{F}^{\Omega}$ is reminiscent of hub or authority type of connections: a connection between two vertices exists if they are both connected to some other vertex (in some direction). In [18] the authors use this kind of connection to construct simplicial complexes over networks.
- If $\Omega = \{L_{m+1}\}$, then $\mathfrak{F}^{\Omega} = \mathfrak{F}^{\mathrm{us}} \circ \mathfrak{F}^{[m]}$.
- If $\Omega = \{C_n\}$, then $v \xrightarrow[=]{\mathfrak{F}^{\Omega}(G)} v'$ when v and v' are in a cycle of size at most n. We can also consider $\Omega = \{C_n\}_{n\in\mathbb{N}}$: then $v \xrightarrow[=]{\mathfrak{F}^{\Omega}(G)} v'$ when v and v' are in a cycle (of any length).

The next proposition follows readily from Definition 2.21.

Proposition 2.24. Every representable functor $\mathfrak{F}^{\Omega}$ is symmetric.

Remark 2.25. Not all symmetric endofunctors are representable. Consider, for example, $\mathfrak{F} = \mathfrak{F}^{\mathrm{ls}} \circ \mathfrak{F}^{[2]}$. Then $\mathfrak{F}(C_4)$ has arrows $a_1 \leftrightarrow a_3$ and $a_2 \leftrightarrow a_4$, that is, $\mathfrak{F}(C_4) = K_2 \sqcup K_2$. See Figure 2.3. Suppose $\mathfrak{F} = \mathfrak{F}^{\Omega}$, for some family of representers Ω. Then we have a graph map $\phi\colon \omega \to (C_4, a_1, a_3)$ for some $\omega \in \Omega$. But, since there are no more arrows in $\mathfrak{F}(C_4)$, we must have $\phi(\omega) = C_4 \cap \{a_1, a_3\}$, which is a graph with two vertices and no arrows. This implies that $\mathfrak{F}(D_2) = K_2$. By Proposition 2.10, we should have $\mathfrak{F} = \mathfrak{F}^{\mathrm{comp}}$, a contradiction.

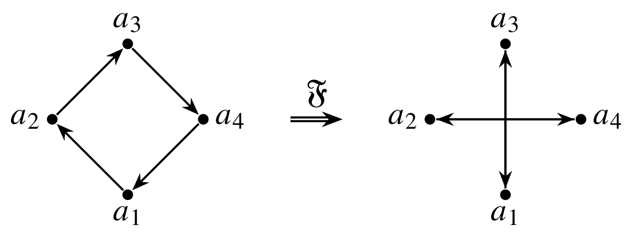


Figure 2.3: The image of C_4 via $\mathfrak{F} = \mathfrak{F}^{\mathrm{ls}} \circ \mathfrak{F}^{[2]}$. See Remark 2.25.

Definition 2.26. Given an endofunctor $\mathfrak{F}$, we define the *set of completions of* $\mathfrak{F}$ by

$$\mathrm{Comp}(\mathfrak{F}) := \{\omega \in \mathcal{G} \mid \mathfrak{F}(\omega) = K_{(\omega)}\},$$

which is non-empty since $D_1 \in \mathrm{Comp}(\mathfrak{F})$.

Proposition 2.27. Let $\mathfrak{F}\colon \mathcal{G} \to \mathcal{G}$ be any endofunctor and $A = \mathrm{Comp}(\mathfrak{F})$. Then $\mathfrak{F}^{A} \leq \mathfrak{F}^{\mathrm{ls}} \circ \mathfrak{F} \leq \mathfrak{F}$.

Proof. The second inequality is clear: $\mathfrak{F}^{\mathrm{ls}} \leq \mathfrak{F}^{\mathrm{id}}$ implies $\mathfrak{F}^{\mathrm{ls}} \circ \mathfrak{F} \leq \mathfrak{F}^{\mathrm{id}} \circ \mathfrak{F} = \mathfrak{F}$.

To prove the first inequality, let $G \in \mathcal{G}$ and suppose that $v \xrightarrow{\mathfrak{F}^{A}(G)} v'$. Then there are $\omega \in A$ and $\phi\colon \omega \to (G, v, v')$. By functoriality, $\phi\colon \mathfrak{F}(\omega) \to (\mathfrak{F}(G), v, v')$ is also a graph map. Since $\mathfrak{F}(\omega) = K_{(\omega)}$, we have $v \xleftrightarrow{\mathfrak{F}(G)} v'$. Thus, $v \xrightarrow{\mathfrak{F}^{\mathrm{ls}} \circ \mathfrak{F}(G))} v'$. □

Corollary 2.28. If $\mathfrak{F}^{\Omega}$ is a representable functor, for some family Ω, then $\mathfrak{F}^{\Omega} = \mathfrak{F}^{A}$.

Proof. We already have $\mathfrak{F} \leq \mathfrak{F}^{\Omega}$. For the other inclusion, just notice that $\Omega \subseteq A$. □

The endofunctors $\mathfrak{F}^{\mathrm{comp}}$ and $\mathfrak{F}^{\mathrm{disc}}$ are, respectively, the maximal and minimal elements of the partial order $\leq$, because their outputs are "all or nothing" graphs (complete or discrete), that is: $\mathfrak{F}^{\mathrm{disc}} \leq \mathfrak{F} \leq \mathfrak{F}^{\mathrm{comp}}$. For other $\mathfrak{F}$, we have two more constraints:

Proposition 2.29. Let $\mathfrak{F} \neq \mathfrak{F}^{\mathrm{disc}}, \mathfrak{F}^{\mathrm{comp}}$ be an endofunctor. Then

$$\mathfrak{F}^{\mathrm{disc}} \leq \mathfrak{F}^{\mathrm{ls}} \leq \mathfrak{F} \leq \mathfrak{F}^{\mathrm{conn}} \leq \mathfrak{F}^{\mathrm{comp}}.$$

Proof. Indeed, if $\mathfrak{F} \neq \mathfrak{F}^{\mathrm{disc}}$, then by Proposition 2.10 we have $\mathfrak{F}(K_2) = K_2$. Thus, by Proposition 2.27, $\mathfrak{F}^{\{K_2\}} = \mathfrak{F}^{\mathrm{ls}} \leq \mathfrak{F}^{\mathrm{Comp}(\mathfrak{F})} \leq \mathfrak{F}$.

Now suppose that we don't have $\mathfrak{F} \leq \mathfrak{F}^{\mathrm{conn}}$. Then, for some $G \in \mathcal{G}$, we have $v \nrightarrow v'$ in $\mathfrak{F}^{\mathrm{conn}}(G)$ and $v \xrightarrow{\mathfrak{F}(G)} v'$. Then v and v' are in different weak connected components of G, say C and C'. Thus, the surjective map $\phi\colon G \to D_2$ given by $\phi(C) = v$ and $\phi(G \setminus C) = v'$ is a graph map and, by functoriality, $\mathfrak{F}(D_2) \neq D_2$. Thus, $\mathfrak{F} = \mathfrak{F}^{\mathrm{comp}}$. □

2.4 Pointed representable functors

For a representable functor $\mathfrak{F}^{\Omega}$, we have $v \xrightarrow{\mathfrak{F}^{\Omega}(G)} v'$ when there are $\omega \in \Omega$ and a graph map $\phi\colon \omega \to (G, v, v')$, that is, $v, v' \in \phi(\omega)$. Now we will adapt this definition in order to obtain, in some cases, non-symmetric endofunctors. In order to do this, we will distinguish which vertices $z, \widehat{z}$ of ω satisfy $\phi(z) = v$ and $\phi(\widehat{z}) = v'$.

Definition 2.30. Let $\mathcal{G}^*$ be the category whose objects are all triples $(\omega, z, \widehat{z})$, called *pointed graphs*, where ω is a graph with $z, \widehat{z} \in \omega$. The morphisms in $\mathcal{G}^*$ from $(\omega_1, z_1, \widehat{z}_1)$ to $(\omega_2, z_2, \widehat{z}_2)$ are all graph maps $\phi\colon (\omega_1, z_1, \widehat{z}_1) \to (\omega_2, z_2, \widehat{z}_2)$.

Given a set Ω^* of pointed graphs, consider $\mathfrak{F}^{\Omega^*} \in \mathrm{Funct}(\mathcal{G}, \mathcal{G})$, where $\mathfrak{F}^{\Omega^*}(G)$ satisfies, for any $G \in \mathcal{G}$,

$$v \xrightarrow{\mathfrak{F}^{\Omega^*}(G)} v' \Leftrightarrow \exists (\omega, z, \widehat{z}) \in \Omega^*, \exists \phi\colon (\omega, z, \widehat{z}) \to (G, v, v').$$

When $\mathfrak{F} = \mathfrak{F}^{\Omega^*}$ for some $\Omega^* \subset \mathcal{G}^*$, we say that $\mathfrak{F}$ is *pointed representable*, and is *represented* by Ω^*.

Example 2.31. If Ω^* consists exactly of:

- (ω, z, z), then $\mathfrak{F}^{\Omega^*} = \mathfrak{F}^{\mathrm{disc}}$, for any $\omega \in \mathcal{G}$ and $z \in \omega$.
- (D_2, a_1, a_2), then $\mathfrak{F}^{\Omega^*} = \mathfrak{F}^{\mathrm{comp}}$.
- (L_2, a_1, a_2), then $\mathfrak{F}^{\Omega^*} = \mathfrak{F}^{\mathrm{id}}$, the identity functor. Remember that L_2 has vertex set $\{a_1, a_2\}$ and an arrow $a_1 \to a_2$.
- (L_2, a_2, a_1), then $\mathfrak{F}^{\Omega^*} = \mathfrak{F}^{\mathrm{rev}}$.
- (K_2, a_1, a_2), then $\mathfrak{F}^{\Omega^*} = \mathfrak{F}^{\mathrm{ls}}$.
- $\{(L_n, a_1, a_n), n \leq m\}$, for a given $m \in \mathbb{N}$, then $\mathfrak{F}^{\Omega^*} = \mathfrak{F}^{[m]}$.
- $\{(L_n, a_1, a_n), n \in \mathbb{N}\}$, then $\mathfrak{F}^{\Omega^*} = \mathfrak{F}^{\mathrm{tc}}$.

Any representable functor $\mathfrak{F}^{\Omega}$ is equal to a pointed representable functor, just taking as representer the set of all possible pointed graphs induced by Ω. More precisely:

Proposition 2.32. Let Ω be a family of graphs. Consider the family of pointed graphs induced by Ω, $P^*(\Omega) := \{(\omega, z, \widehat{z}), z, \widehat{z} \in \omega \in \Omega\}$. Then $\mathfrak{F}^{\Omega} = \mathfrak{F}^{P^*(\Omega)}$.

Proof. It is clear that $\mathfrak{F}^{P^*(\Omega)} \leq \mathfrak{F}^{\Omega}$. Now we prove the remaining inequality. Given $G \in \mathcal{G}$, by definition, $v \xrightarrow{\mathfrak{F}(G)} v' \Leftrightarrow \exists \omega \in \Omega, \exists \phi\colon \omega \to (G, v, v')$. This is equivalent to the condition that there is $z, \widehat{z} \in \omega$ such that $\phi\colon (\omega, z, \widehat{z}) \to (G, v, v')$. Since $(\omega, z, \widehat{z}) \in \Omega^*$, we have $v \xrightarrow{\mathfrak{F}^{P^*(\Omega)}(G)} v'$. Thus $\mathfrak{F}^{\Omega} \leq \mathfrak{F}^{P^*(\Omega)}$. □

Let's consider some standard notation that will ease the reading of what follows: for any graph $G = (V, E)$, the maps $s, t\colon E \to V$ are defined by $s(e) = v$ and $t(e) = v'$, for any

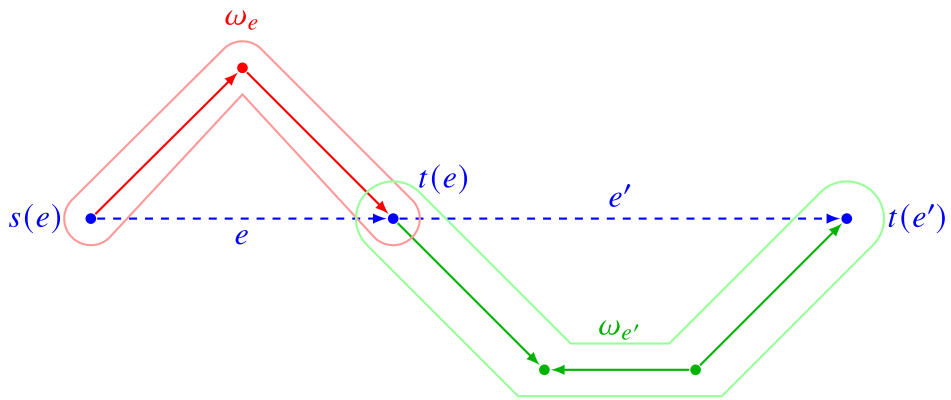


Figure 2.4: In blue: the graph ω, with dashed arrows. In red: the graph ω_e. In green: the graph $\omega_{e'}$. The graph Z is obtained excluding from the picture the blue dashed arrows. See Definition 2.33.

$e = (v, v') \in E$. These functions are called the *source* and *target*, respectively. Thus, for any $e \in E$ we have $s(e) \overset{G}{\underset{=}{\rightarrow}} t(e)$.

When computing the composition $\mathfrak{F}^{\Omega_2^*} \circ \mathfrak{F}^{\Omega_1^*}$, for two families of pointed graphs Ω_1^* and Ω_2^*, we encounter the following construction. Let $(\omega, z, \widehat{z}) \in \Omega_2^*$. For every arrow e in ω, one "attaches" to ω an arbitrary graph $(\omega_e, z_e, \widehat{z}_e) \in \Omega_1^*$ by identifying z_e with $s(e)$ and $\widehat{z}_e$ with $t(e)$. After doing this to all the arrows in ω, we delete the original arrows of ω and only keep the "attached" ones, and thus obtain another graph, let's say Z. See Figure 2.4 for an illustration.

Definition 2.33. Let $(\omega, z, \widehat{z}) \in \mathcal{G}^*$ with vertex set V_ω and arrow set E_ω, and let Ω_1^* be a set of pointed graphs. Let $F\colon E_\omega \longrightarrow \Omega_1^*$ be any function assigning edges from E_ω to pointed graphs in Ω_1^*. For each $e \in E_\omega$ denote by $(\omega_e, z_e, \widehat{z}_e)$ the element $F(e)$ and denote its vertex set by V_{ω_e}. Define a graph ω_F as follows:

1. The vertex set of ω_F is the quotient of the disjoint union
$$\left(V_\omega \bigsqcup_{e \in E_\omega} V_{\omega_e}\right)/\!\sim$$
where for every $e \in E_\omega$ we identify $s(e) \sim z_e$ and $t(e) \sim \widehat{z}_e$.

2. There is an arrow from x to y in $\omega_F \Leftrightarrow x$ and y are both contained in some ω_e and $x \overset{\omega_e}{\underset{=}{\longrightarrow}} y$.

Since $z, \widehat{z} \in \omega_F$, we can consider the pointed graph $(\omega_F, z, \widehat{z})$. Notice that for all $e \in E_\omega$ we have natural inclusions $\iota_e\colon (\omega_e, z_e, \widehat{z}_e) \hookrightarrow (\omega_F, s(e), t(e))$.

Define the *composition* of the single pointed graph $(\omega, z, \widehat{z})$ with the whole set Ω_1^* as
$$(\omega, z, \widehat{z}) \diamond \Omega_1^* := \{(\omega_F, z, \widehat{z}) \mid F\colon E_\omega \longrightarrow \Omega_1^* \text{ is a function}\}.$$

Finally, define
$$\Omega_2^* \diamond \Omega_1^* := \bigcup_{(\omega,z,\widehat{z}) \in \Omega_2^*} (\omega, z, \widehat{z}) \diamond \Omega_1^*,$$
which we will refer to as *the composition of* Ω_2^* *with* Ω_1^*.

Theorem 2.34. Composition of representable functors is also representable. More precisely, given Ω_1^* and Ω_2^* two families of pointed graphs, we have

$$\mathfrak{F}^{\Omega_2^*} \circ \mathfrak{F}^{\Omega_1^*} = \mathfrak{F}^{\Omega_2^* \diamond \Omega_1^*}.$$

Proof. Let $\mathfrak{F}_1 = \mathfrak{F}^{\Omega_1^*}$, $\mathfrak{F}_2 = \mathfrak{F}^{\Omega_2^*}$, $\mathfrak{F} = \mathfrak{F}^{\Omega_2^* \diamond \Omega_1^*}$ and $G = (V, E) \in \mathcal{G}$.

Suppose $v \xrightarrow{\mathfrak{F}_2(\mathfrak{F}_1(G))} v'$. By definition, there are $(\omega, z, \widehat{z}) \in \Omega_2^*$ and a graph map $\phi\colon (\omega, z, \widehat{z}) \to (\mathfrak{F}_1(G), v, v')$. Denote the vertex set of ω by V_ω and the arrow set by E_ω. Since ϕ is a graph map, for each $e \in E_\omega$ we have $\phi(s(e)) \xrightarrow[=]{\mathfrak{F}_1(G)} \phi(t(e))$. Then, for every $e \in E_\omega$ there are $(\omega_e, z_e, \widehat{z}_e) \in \Omega_1^*$ and a graph map $\phi_e\colon (\omega_e, z_e, \widehat{z}_e) \to (G, \phi(s(e)), \phi(t(e)))$. Let $F\colon E_\omega \to \Omega_1^*$ be defined by $F(e) = (\omega_e, z_e, \widehat{z}_e)$.

The graph maps ϕ and ϕ_e induce a graph map $\psi\colon (\omega_F, z, \widehat{z}) \to (G, v, v')$ (notice that if $x \in \omega_e$ and $y \in \omega_{e'}$ with $e \neq e'$, then there is no arrow between x and y in ω_F). This implies that $v \xrightarrow{\mathfrak{F}(G)} v'$, and then $\mathfrak{F}_2 \circ \mathfrak{F}_1 \leq \mathfrak{F}$.

Reciprocally, if $v \xrightarrow{\mathfrak{F}(G)} v'$, there is a graph map $\psi\colon (\omega_F, z, \widehat{z}) \to (G, v, v')$, for some $(\omega, z, \widehat{z}) \in \Omega_2^*$ with vertex set V_ω and arrow set E_ω, and for some $F\colon E_\omega \to \Omega_1^*$, where $F(e) = (\omega_e, z_e, \widehat{z}_e)$.

Consider the graph maps $\phi_e\colon (\omega_e, z_e, \widehat{z}_e) \to (G, \psi(s(e)), \psi(t(e)))$ obtained by restricting the domain of ψ, that is: $\phi_e = \psi|_{\omega_e}$. Thus, $\phi_e(s(e)) \xrightarrow{\mathfrak{F}_1(G)} \phi_e(t(e))$. Consider $\phi\colon (\omega, z, \widehat{z}) \to (\mathfrak{F}_1(G), v, v')$ defined by $\phi(x) = \psi(x)$. This is indeed a graph map since for any $e \in E_\omega$ we have

$$\phi(s(e)) = \phi_e(s(e)) \xrightarrow{\mathfrak{F}_1(G)} \phi_e(t(e)) = \phi(t(e)).$$

Thus, $v \xrightarrow{\mathfrak{F}_2 \circ \mathfrak{F}_1(G)} v'$, that is: $\mathfrak{F} \leq \mathfrak{F}_2 \circ \mathfrak{F}_1$. □

Example 2.35. In Remark 2.25 we have seen that $\mathfrak{F} = \mathfrak{F}^{\mathrm{ls}} \circ \mathfrak{F}^{[2]}$ is not representable. However, this endofunctor is pointed representable: $\mathfrak{F}^{[2]}$ is pointed represented by a line graph with 3 vertices, $\Omega_1^* = \{(L_3, a_1, a_3)\}$, and $\mathfrak{F}^{\mathrm{ls}}$ is pointed represented by the complete graph with two vertices, $\Omega_2^* = \{(\omega, z, \widehat{z})\}$. By the Theorem 2.34, $\mathfrak{F}$ is pointed represented by $\Omega_2^* \diamond \Omega_1^* = \{(Z, z, \widehat{z})\}$, with $(Z, z, \widehat{z})$ isomorphic to (C_4, a_1, a_3). See Figure 2.5.

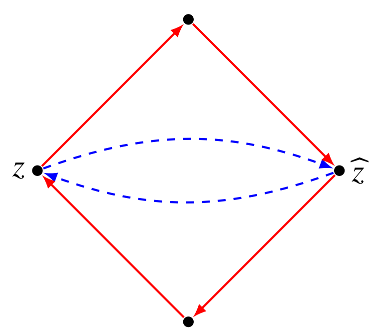


Figure 2.5: In blue: the graph $(\omega, z, \widehat{z})$. In red: two copies of L_3. The graph Z is obtained excluding from the picture the blue arrows. See Example 2.35.

The interesting part of Theorem 2.34 is the formula for the composite, since it turns out that *every* endofunctor is pointed representable.

For any endofunctor $\mathfrak{F}$, let

$$\Omega^*_{\mathfrak{F}} \coloneqq \{(\omega, z, \widehat{z}) \in \mathcal{G}^* \mid z \xrightarrow[=]{\mathfrak{F}(\omega)} \widehat{z}\}.$$

Remark 2.36 (Interpretation of $\Omega^*_{\mathfrak{F}}$)**.** The set $\Omega^*_{\mathfrak{F}}$ is composed of all graphs $(\omega, z, \widehat{z})$ such that $\mathfrak{F}(\omega)$ has at least one arrow, together with all arrows $z \to \widehat{z}$ generated by $\mathfrak{F}$. Intuitively, $\Omega^*_{\mathfrak{F}}$ is the collection of all possible arrows that $\mathfrak{F}$ can create, thus it is not surprising that all the information of $\mathfrak{F}$ is contained in $\Omega^*_{\mathfrak{F}}$ as stated by the following theorem.

Theorem 2.37. For every endofunctor $\mathfrak{F}$, it holds $\mathfrak{F} = \mathfrak{F}^{\Omega^*_{\mathfrak{F}}}$.

Proof. Let G be a graph. If $v \xrightarrow{\mathfrak{F}^{\Omega^*_{\mathfrak{F}}}(G)} v'$, then, by definition, there are $(\omega, z, \widehat{z}) \in \Omega^*_{\mathfrak{F}}$ and a graph map $\phi\colon (\omega, z, \widehat{z}) \to (G, v, v')$. Applying $\mathfrak{F}$ to this diagram, we obtain the graph map $\phi\colon (\mathfrak{F}(\omega), z, \widehat{z}) \to (\mathfrak{F}(G), v, v')$. By definition of $\Omega^*_{\mathfrak{F}}$, we have that $z \xrightarrow{\mathfrak{F}(\omega)} \widehat{z}$, and since ϕ is a graph map we also have $v \xrightarrow{\mathfrak{F}(G)} v'$.

Reciprocally, if $v \xrightarrow{\mathfrak{F}(G)} v'$, then $(G, v, v') \in \Omega^*_{\mathfrak{F}}$ and since the identity $\mathrm{id}_G\colon (G, v, v') \to (G, v, v')$ is a graph map, it follows that $v \xrightarrow{\mathfrak{F}^{\Omega^*_{\mathfrak{F}}}(G)} v'$. □

The set $\Omega^*_{\mathfrak{F}}$ is often unnecessarily large. For example, if $\mathfrak{F} = \mathfrak{F}^{\mathrm{comp}}$, then $\Omega^*_{\mathfrak{F}}$ is the set of all pointed graphs. Reducing its size is the content of the next section.

2.5 Simplification and clustering

Given a set Ω^* of pointed graphs, can one find a simpler (in some suitable sense) Ω^*_0 such that $\mathfrak{F}^{\Omega^*} = \mathfrak{F}^{\Omega^*_0}$? A necessary and sufficient condition for this is the following.

Definition 2.38. We say that $(\omega_2, z_2, \widehat{z}_2)$ *covers* $(\omega_1, z_1, \widehat{z}_1)$ if there is a graph map

$$p\colon (\omega_2, z_2, \widehat{z}_2) \to (\omega_1, z_1, \widehat{z}_1).$$

Given two families of graphs Ω^*_1 and Ω^*_2, we say that Ω^*_2 *covers* Ω^*_1 if for any $(\omega_1, z_1, \widehat{z}_1) \in \Omega^*_1$, there is $(\omega_2, z_2, \widehat{z}_2) \in \Omega^*_2$ that covers $(\omega_1, z_1, \widehat{z}_1)$. We denote this by $\Omega^*_1 \preccurlyeq \Omega^*_2$.

Notice that $\Omega^*_1 \subseteq \Omega^*_2$ implies $\Omega^*_1 \preccurlyeq \Omega^*_2$.

For the next two results, let Ω^*_1 and Ω^*_2 be two families of pointed graphs, and let the endofunctors they represented be denoted by $\mathfrak{F}_1 = \mathfrak{F}^{\Omega^*_1}$ and $\mathfrak{F}_2 = \mathfrak{F}^{\Omega^*_2}$, respectively.

Theorem 2.39. $\mathfrak{F}_1 \leq \mathfrak{F}_2 \Leftrightarrow \Omega^*_1 \preccurlyeq \Omega^*_2$.

Proof. Let's prove the ($\Leftarrow$) implication first. Suppose $v \xrightarrow{\mathfrak{F}_1(G)} v'$. Then there are $(\omega_1, z_1, \widehat{z}_1) \in \Omega^*_1$ and a graph map $\phi\colon (\omega_1, z_1, \widehat{z}_1) \to (G, v, v')$. Since $\Omega^*_1 \preccurlyeq \Omega^*_2$, there are $(\omega_2, z_2, \widehat{z}_2) \in \Omega^*_2$ and a graph map $p\colon (\omega_2, z_2, \widehat{z}_2) \to (\omega_1, z_1, \widehat{z}_1)$.

$$\begin{tikzcd}
(\omega_2, z_2, \widehat{z}_2) \arrow[d, "p"] \arrow[dr, "\phi\circ p"] & \\
(\omega_1, z_1, \widehat{z}_1) \arrow[r, "\phi"] & (G, v, v').
\end{tikzcd}$$

Then $\phi \circ p\colon (\omega_2, z_2, \widehat{z}_2) \to (G, v, v')$ is a graph map, which implies $v \xrightarrow{\mathfrak{F}_2(G)} v'$. Hence, $\mathfrak{F}_1 \leq \mathfrak{F}_2$.

Now for the ($\Rightarrow$) implication, let $(\omega_1, z_1, \widehat{z}_1) \in \Omega_1^*$. It is clear that $z_1 \xrightarrow{\mathfrak{F}_1(\omega_1)} \widehat{z}_1$, and $z_1 \xrightarrow{\mathfrak{F}_2(\omega_1)} \widehat{z}_1$ follows from the fact that $\mathfrak{F}_1 \leq \mathfrak{F}_2$. By definition, there are $(\omega_2, z_2, \widehat{z}_2) \in \Omega_2^*$ and a graph map $p\colon (\omega_2, z_2, \widehat{z}_2) \to (\omega_1, z_1, \widehat{z}_1)$. This proves that $\Omega_1^* \preccurlyeq \Omega_2^*$. □

Corollary 2.40. We have $\mathfrak{F}_1 = \mathfrak{F}_2 \Leftrightarrow \Omega_1^* \preccurlyeq \Omega_2^*$ and $\Omega_2^* \preccurlyeq \Omega_1^*$.

Via Proposition 2.32 we can obtain a theorem similar to Theorem 2.39 which applies in the case of standard (non pointed) representable functors.

Definition 2.41. Let Ω_1 and Ω_2 be any two families of graphs. We say that Ω_2 *covers* Ω_1 if for any $\omega_1 \in \Omega_1$ and $z_1, \widehat{z}_1 \in \omega_1$, there are $\omega_2 \in \Omega_2$ and a graph map $p\colon \omega_2 \to (\omega_1, z_1, \widehat{z}_1)$. We denote this by $\Omega_1 \preccurlyeq \Omega_2$. Notice that $\Omega_1 \preccurlyeq \Omega_2$ is equivalent to $P^*(\Omega_1) \preccurlyeq P^*(\Omega_2)$, where P^* is as in Proposition 2.32.

Corollary 2.42. Let Ω_1 and Ω_2 be two families of graphs. Then $\mathfrak{F}^{\Omega_1} \leq \mathfrak{F}^{\Omega_2} \Leftrightarrow \Omega_1 \preccurlyeq \Omega_2$.

Proof. We already know that $\Omega_1 \preccurlyeq \Omega_2 \Leftrightarrow P^*(\Omega_1) \preccurlyeq P^*(\Omega_2)$. By Proposition 2.32, $\mathfrak{F}^{\Omega_1} = \mathfrak{F}^{P^*(\Omega_1)}$ and $\mathfrak{F}^{\Omega_2} = \mathfrak{F}^{P^*(\Omega_2)}$. Then,

$$\mathfrak{F}^{\Omega_1} \leq \mathfrak{F}^{\Omega_2} \Leftrightarrow \mathfrak{F}^{P^*(\Omega_1)} \leq \mathfrak{F}^{P^*(\Omega_2)} \Leftrightarrow P^*(\Omega_1) \preccurlyeq P^*(\Omega_2) \Leftrightarrow \Omega_1 \preccurlyeq \Omega_2,$$

where the second equivalence follows by Theorem 2.39. □

Example 2.43. Given $\Omega^* \subset \mathcal{G}^*$ and $(\omega_0, z_0, \widehat{z}_0) \in \Omega^*$, let

$$A_0 := \{(\omega, z, \widehat{z}) \in \Omega^* \mid (\omega_0, z_0, \widehat{z}_0) \text{ covers } (\omega, z, \widehat{z})\},$$

and $\Omega_0^* = (\Omega^* \setminus A_0) \cup \{(\omega, z, \widehat{z})\}$. Then $\Omega_0^* \subseteq \Omega^*$ implies $\Omega_0^* \preccurlyeq \Omega^*$, and, moreover, we have $\Omega^* \preccurlyeq \Omega_0^*$. Thus, $\mathfrak{F}^{\Omega^*} = \mathfrak{F}^{\Omega_0^*}$. We can regard Ω_0^* as a simplification of Ω^* which yields the same pointed representable functor.

Consider, for example, $\Omega^* = \{(L_n, a_1, a_n), n \leq m\}$, for a given $m \in \mathbb{N}$. Then, choosing (L_m, a_1, a_m) as the pointed graph $(\omega_0, z_0, \widehat{z}_0)$ above, we obtain $\Omega_0^* = \{(L_m, a_1, a_m)\}$ and $\mathfrak{F}^{[m]} = \mathfrak{F}^{\Omega^*} = \mathfrak{F}^{\Omega_0^*}$, that is: we can "throw away" every (L_n, a_1, a_n) with $n < m$ and still obtain the same functor. When $\Omega^* = \{(L_n, a_1, a_n), n \in \mathbb{N}\}$, however, we cannot obtain a finite family Ω_0^* such that $\mathfrak{F}^{\Omega^*} = \mathfrak{F}^{\Omega_0^*}$.

Another example: take Ω^* as the set of all pointed graphs, and $\Omega_0^* = \{D_2\}$. We have $\Omega_0^* \preccurlyeq \Omega^*$ because $\Omega_0^* \subset \Omega^*$, and $\Omega^* \preccurlyeq \Omega_0^*$, since for any $(\omega, z, \widehat{z}) \in \Omega^*$ the map $\phi\colon (D_2, a_1, a_2) \to (\omega, z, \widehat{z})$ is a graph map. Thus, $\mathfrak{F}^{\Omega^*} = \mathfrak{F}^{\Omega_0^*} = \mathfrak{F}^{\mathrm{comp}}$.

2.5.1 Clustering functors

We already know (Proposition 2.24) that any representable functor $\mathfrak{F}^\Omega$ is symmetric. We now establish conditions under which $\mathfrak{F}^\Omega$ turns out to be a clustering functor (that is, $\mathfrak{F}^\Omega$ is also transitive).

Definition 2.44. For any $G_1, G_2 \in \mathcal{G}$, $v_1 \in G_1$ and $v_2 \in G_2$, define the *wedge product* of (G_1, v_1) and (G_2, v_2) by

$$(G_1, v_1) \vee (G_2, v_2) := \frac{G_1 \sqcup G_2}{v_1 \sim v_2},$$

that is: we identify v_1 and v_2 in the disjoint union $G_1 \sqcup G_2$.

Given two families of graphs Ω_1 and Ω_2, define

$$\Omega_1 \vee \Omega_2 := \bigcup_{\substack{\omega_i \in \Omega_i,\ z_i \in \omega_i \\ i=1,2}} (\omega_1, z_1) \vee (\omega_2, z_2).$$

We say that Ω is *wedge covered* whenever $\Omega \vee \Omega \preccurlyeq \Omega$.

Theorem 2.45. Ω is a wedge covered family of graphs $\Leftrightarrow$ $\mathfrak{F}^\Omega$ is a clustering functor.

Proof. Let's prove the ($\Rightarrow$) implication first. Let G be a graph. Suppose that $v_1 \xrightarrow{\mathfrak{F}^\Omega(G)} v_2$ and $v_2 \xrightarrow{\mathfrak{F}^\Omega(G)} v_3$. Then, for some $\omega_1, \omega_2 \in \Omega$, there exist graph maps $\phi_1 \colon (\omega_1, z_1, \widehat{z}_1) \to (G, v_1, v_2)$ and $\phi_2 \colon (\omega_2, z_2, \widehat{z}_2) \to (G, v_2, v_3)$. Let $\omega = (\omega_1, \widehat{z}_1) \vee (\omega_2, z_2)$. Then the map $\phi \colon (\omega, z_1, \widehat{z}_2) \to (G, v_1, v_3)$ given by

$$\phi(v) = \begin{cases} \phi_1(v), & \text{if } v \in \omega_1, \\ \phi_2(v), & \text{if } v \in \omega_2, \end{cases}$$

is a well defined graph map. Since Ω is wedge covered, there are $\tilde{\omega} \in \Omega$ and a graph map $p \colon (\tilde{\omega}, z, \widehat{z}) \to (\omega, z_1, \widehat{z}_2)$. The composite $\phi \circ p \colon (\tilde{\omega}, z, \widehat{z}) \to (G, v_1, v_3)$ is a graph map, which implies $v_1 \xrightarrow{\mathfrak{F}^\Omega(G)} v_3$, that is, $\mathfrak{F}^\Omega$ is transitive. That $\mathfrak{F}^\Omega$ is symmetric follows from it being representable.

Now let's prove the ($\Leftarrow$) implication. Let $z_1 \in \omega_1 \in \Omega$ and $z_2 \in \omega_2 \in \Omega$. Consider $G = (\omega_1, z_1) \vee (\omega_2, z_2)$. Let $v, v' \in G$. Consider the inclusions $i_1 \colon \omega_1 \hookrightarrow G$ and $i_2 \colon \omega_2 \hookrightarrow G$. Suppose $v \in i_1(\omega_1)$ and $v' \in i_2(\omega_2)$. Then we have $v \xrightarrow{\mathfrak{F}(G)} i_1(z_1)$ and $i_2(z_2) \xrightarrow{\mathfrak{F}(G)} v'$. Since $\mathfrak{F}(G)$ is transitive and $i_1(z_1) = i_2(z_2)$, we have $v \xrightarrow{\mathfrak{F}(G)} v'$. By definition of wedge covered, there are $\omega \in \Omega$ and a graph map $p \colon \omega \to (G, v, v')$. The cases when both $v, v' \in i_1(\omega_1)$ or $v, v' \in i_2(\omega_2)$ are immediate. This proves that Ω is wedge covered. □

Example 2.46. Two important examples of wedge covered families are:

1. the set of arbitrary size *reciprocal* line graphs: $\Omega = \{\mathfrak{F}^{us}(L_n), n \in \mathbb{N}\}$. $\mathfrak{F}^\Omega$ is called *the reciprocal clustering functor*, denoted by $\mathfrak{F}^{rec}$. Notice that $\mathfrak{F}^{us}(L_n)$ has vertex set $\{a_1, \ldots, a_n\}$ and arrows $a_i \leftrightarrow a_{i+1}$, for $i = 1, \ldots, n-1$.

2. the set of arbitrary size cycle graphs $\Omega = \{C_n, n \in \mathbb{N}\}$. The functor $\mathfrak{F}^\Omega$ is called *the non-reciprocal clustering functor*, denoted by $\mathfrak{F}^{nrec}$.

3 Networks and clustering

> *Open your window, Octave; do you not see the infinite? You try to form some idea of a thing that has no limits, you who were born yesterday and who will die tomorrow? [...] All the people of the earth have stretched out their hands to that immensity and have longed to plunge into it. The fool wishes to possess heaven; the sage admires it, kneels before it, but does not desire it. Perfection, my friend, is no more made for us than infinity.*
>
> ALFRED DE MUSSET, *THE CONFESSIONS OF A CHILD OF THE CENTURY (TRANSLATED BY KENDALL WARREN)*

Studying clustering methods of *dissimilarity networks* (a generalization of a metric space) was done in great details in [7,8], following previous works on clustering of metric spaces [4, 5]. The theory of networks in its more general form appeared in [68,69] with a great amount of work done in the persistent homology setting [18,70].

Our main contribution is the relation between graph filtrations and networks, and how functors on $\mathcal{G}$ induce functors on the category of extended networks, being able to define "motivic" (clustering) functors. Moreover, we show a stability result.

The material of this chapter is mostly a reproduction of the paper [67], co-authored with Facundo Mémoli.

3.1 Definition and properties

A network is a complete graph together with a weight function defined on its arrows. More precisely:

Definition 3.1. An *extended network* is a pair $\mathbf{X} = (X, w_X)$, where X is a finite set, and $w_X\colon X \times X \to \mathbb{R} \cup \{+\infty\}$, with $w_X(x,x) < +\infty$, for any $x \in X$, is a function (called *weight function*). A *network map* $\phi\colon \mathbf{X} \to \mathbf{Y}$, where $\mathbf{Y} = (Y, w_Y)$, is a map $\phi\colon X \to Y$ such that $w_Y(\phi(x), \phi(x')) \le w_X(x,x')$ for any $x, x' \in X$, with the convention that $a < +\infty$, for all $a \in \mathbb{R}$.

Denote by $\mathcal{N}_{\mathrm{ext}}$ the category of extended networks and network maps. Dropping the prefix "extended" means that the codomain of the weight function is $\mathbb{R}$, instead of $\mathbb{R} \cup \{+\infty\}$. Thus, the category $\mathcal{N}$ of networks (see [68]) is a subcategory of $\mathcal{N}_{\mathrm{ext}}$.

For any $\mathbf{X} = (X, w_X) \in \mathcal{N}_{\mathrm{ext}}$ and $\epsilon \in \mathbb{R}$, define the graph $\mathbf{X}_\epsilon$ with vertex set

$$V(\mathbf{X}_\epsilon) := \{x \in X \mid w_X(x,x) \le \epsilon\}$$

and arrow set

$$E(\mathbf{X}_\epsilon) := \{(x, x') \in V(\mathbf{X}_\epsilon) \times V(\mathbf{X}_\epsilon) \mid w_X(x, x') \le \epsilon\}.$$

Any functor $\mathfrak{F}: \mathcal{G} \to \mathcal{G}$ induces a functor $\widehat{\mathfrak{F}}: \mathcal{N}_{\text{ext}} \to \mathcal{N}_{\text{ext}}$ defined on the objects of $\mathcal{N}_{\text{ext}}$ by $\widehat{\mathfrak{F}}(\mathbf{X}) = (X, w_X^{\mathfrak{F}})$, where

$$w_X^{\mathfrak{F}}(x, x') := \min\left\{\epsilon \in \mathbb{R} \mid x \xrightarrow[=]{\mathfrak{F}(\mathbf{X}_\epsilon)} x'\right\},$$

and we consider the minimum over the empty set to be $+\infty$. This is the same as saying that $\widehat{\mathfrak{F}}(\mathbf{X})_\epsilon = \mathfrak{F}(\mathbf{X}_\epsilon)$. Now let $\phi: \mathbf{X} \to \mathbf{Y}$ be a network map, with $\mathbf{Y} = (Y, w_Y)$. On morphisms, define $\widehat{\mathfrak{F}}(\phi) := \phi$. Let's prove that $\phi: \widehat{\mathfrak{F}}(\mathbf{X}) \to \widehat{\mathfrak{F}}(\mathbf{Y})$ is a network map.

It is clear that ϕ induces graph maps $\phi_\epsilon: \mathbf{X}_\epsilon \to \mathbf{Y}_\epsilon$ for all $\epsilon \in \mathbb{R}$, because

$$x \xrightarrow[=]{\mathbf{X}_\epsilon} x' \Leftrightarrow w_X(x, x') \le \epsilon \Rightarrow w_Y(\phi(x), \phi(x')) \le \epsilon \Leftrightarrow \phi(x) \xrightarrow[=]{\mathbf{Y}_\epsilon} \phi(x').$$

The functoriality of $\mathfrak{F}$ guarantees that $\phi_\epsilon: \mathfrak{F}(\mathbf{X}_\epsilon) \to \mathfrak{F}(\mathbf{Y}_\epsilon)$ is also a graph map, and the following

$$w_X^{\mathfrak{F}}(x, x') \le \epsilon \Leftrightarrow x \xrightarrow[=]{\mathfrak{F}(\mathbf{X}_\epsilon)} x' \Rightarrow \phi(x) \xrightarrow[=]{\mathfrak{F}(\mathbf{Y}_\epsilon)} \phi(x') \Leftrightarrow w_Y^{\mathfrak{F}}(\phi(x), \phi(x')) \le \epsilon$$

implies that $\phi: \widehat{\mathfrak{F}}(\mathbf{X}) \to \widehat{\mathfrak{F}}(\mathbf{Y})$ is a network map.

Definition 3.2. An *extended ultranetwork* is an extended network $\mathbf{X} = (X, u_X)$ that satisfies the *strong triangle inequality*: $u_X(x, x'') \le \max\{u_X(x, x'), u_X(x', x'')\}$ for all $x, x', x'' \in X$. This implies that $\mathbf{X}_\epsilon \in \mathcal{G}^{\text{trans}}$ for all ϵ. Denote by $\mathcal{U}_{\text{ext}}$ the subcategory of $\mathcal{N}_{\text{ext}}$ whose objects are the extended ultranetworks, and the obvious analogues adding the modifiers "symmetric" or "extended": $\mathcal{U}$, $\mathcal{U}_{\text{ext}}^{\text{sym}}$ and $\mathcal{U}^{\text{sym}}$.

Definition 3.3. A *graph filtration* is a set $\mathbf{G} = \{G_\epsilon\}_{\epsilon\in\mathbb{R}}$, with $G_\epsilon = (V_\epsilon, E_\epsilon) \in \mathcal{G}$, satisfying:

1. (Finiteness) There are $\epsilon_0, \epsilon_1 \in \mathbb{R}$ such that $G_\epsilon = \emptyset$, for all $\epsilon < \epsilon_0$, and $G_\epsilon = G_{\epsilon_1}$, for all $\epsilon \ge \epsilon_1$.

2. (Hierarchy) For every $\epsilon \le \epsilon'$, we have $G_\epsilon \hookrightarrow G_{\epsilon'}$ (here we consider $\emptyset \hookrightarrow G$ for any graph G), that is, $V_\epsilon \subseteq V_{\epsilon'}$ and $E_\epsilon \subseteq E_{\epsilon'}$. In particular, every G_ϵ is a subgraph of G_{ϵ_1}.

3. (Right continuity) For every $\epsilon \in \mathbb{R}$, there is $\delta \in \mathbb{R}$ such that $G_\epsilon = G_{\epsilon+r}$, for all $r \in [0, \delta]$.

Equivalently, $\mathbf{G}$ can be seen as a functor $\mathbb{R} \to \mathcal{G}$ mapping ϵ to G_ϵ and $\epsilon \le \epsilon'$ to the inclusion $G_\epsilon \hookrightarrow G_{\epsilon'}$.

Definition 3.4. Denote by **GFilt** the set of all graph filtrations. Define the subsets $\mathbf{GFilt}^{\text{sym}}$, $\mathbf{GFilt}^{\text{trans}}$ and $\mathbf{GFilt}^{\text{clust}}$ of **GFilt** as the sets of graph filtrations $\mathbf{G} = \{G_\epsilon\}_{\epsilon\in\mathbb{R}}$ such that every G_ϵ is in $\mathcal{G}^{\text{sym}}$, $\mathcal{G}^{\text{trans}}$ and $\mathcal{G}^{\text{clust}}$, respectively.

Define $\Phi: \mathcal{N}_{\text{ext}} \to \mathbf{GFilt}$ by $\Phi(\mathbf{X}) := \{\mathbf{X}_\epsilon\}_{\epsilon\in\mathbb{R}}$. Let $\mathbf{G} = \{G_\epsilon\}_{\epsilon\in\mathbb{R}} \in \mathbf{GFilt}$ be any graph filtration and write $X = V_{\epsilon_1}$ as in the Definition 3.3. Consider

$$w_X(x, x') = \min\left\{\epsilon \mid x \xrightarrow[=]{G_\epsilon} x'\right\},$$

where we consider the minimum over the empty set to be $+\infty$. Finally, set $\mathbf{X_G} = (X, w_X)$. Define the map $\Psi\colon \mathbf{GFilt} \to \mathcal{N}_{\mathrm{ext}}$ by $\Psi(\mathbf{G}) := \mathbf{X_G}$.

Notice that the image of Ψ is the subset of $\mathcal{N}_{\mathrm{ext}}$ consisting of extended networks $\mathbf{X} = (X, w_X)$ such that $w_X(x,x) \le \max\{w_X(x,x'), w_X(x',x)\}$, for every $x, x' \in X$. Call such objects *graph-like networks*, and consider its subcategory, denoted by $\mathcal{N}_{\mathcal{G}}$.

Define the map $M\colon \mathcal{N}_{\mathrm{ext}} \to \mathcal{N}_{\mathcal{G}}$ by $M(\mathbf{X}) = (X, \widetilde{w}_X)$, where $\mathbf{X} = (X, w_X)$ and $\widetilde{w}_X(x,x') := \max\{w_X(x,x), w_X(x,x'), w_X(x',x')\}$, which is a "diagonal majoration" (since in the matricial form the elements $w_X(x,x)$ and $w_X(x',x')$ are on the diagonal of the matrix w_X).

It is easy to see that $\Phi(\mathbf{X}) = \Phi(M(\mathbf{X}))$. Also, notice that M restricted to $\mathcal{N}_{\mathcal{G}}$ is the identity. The next theorem follows readily from the definitions.

Theorem 3.5. $\Phi \circ \Psi\colon \mathbf{GFilt} \to \mathbf{GFilt}$ is the identity map, and $\Psi \circ \Phi\colon \mathcal{N}_{\mathrm{ext}} \to \mathcal{N}_{\mathrm{ext}}$ is equal to M. Thus, restricting Φ to $\mathcal{N}_{\mathcal{G}}$, we have that $\Psi \circ \Phi\colon \mathcal{N}_{\mathcal{G}} \to \mathcal{N}_{\mathcal{G}}$ is the identity. □

Example 3.6. Let $\mathbf{X} = (\{x_1, x_2\}, w_X)$ be a network with $w_X = \begin{pmatrix} 3 & 1 \\ 2 & 3 \end{pmatrix}$. Then

$$\mathbf{X}_\epsilon = \begin{cases} \emptyset, & 0 \le \epsilon < 3, \\ \bullet \leftrightarrow \bullet, & \epsilon \ge 3. \end{cases}$$

Now let $\mathbf{G} = \Phi(\mathbf{X}) = \{\mathbf{X}_\epsilon\}_{\epsilon\in\mathbb{R}}$. Then $\Psi(\mathbf{G}) = (\{x_1, x_2\}, \widetilde{w}_X)$, where $\widetilde{w}_X \equiv 3$, and $\Phi(\Psi(\mathbf{X})) = \Phi(\mathbf{G}) = M(\mathbf{X})$.

Notice that $\mathcal{U}_{\mathrm{ext}}$ is a subcategory of $\mathcal{N}_{\mathcal{G}}$. Indeed, given $\mathbf{X} = (X, u_X) \in \mathcal{U}_{\mathrm{ext}}$, we have

$$u_X(x,x'') \le \max\{u_X(x,x'), u_X(x',x'')\}$$

for all $x, x', x'' \in X$. In particular, for $x = x''$ we obtain $u_X(x,x) \le \max\{u_X(x,x'), u_X(x',x)\}$.

Proposition 3.7. We have:

1. $\mathbf{X} \in \mathcal{N}_{\mathrm{ext}}^{\mathrm{sym}} \Rightarrow \Phi(\mathbf{X}) \in \mathbf{GFilt}^{\mathrm{sym}}$ and $\mathbf{G} \in \mathbf{GFilt}^{\mathrm{sym}} \Leftrightarrow \Psi(\mathbf{G}) \in \mathcal{N}_{\mathcal{G}}^{\mathrm{sym}}$.
2. $\mathbf{X} \in \mathcal{U}_{\mathrm{ext}} \Leftrightarrow \Phi(\mathbf{X}) \in \mathbf{GFilt}^{\mathrm{trans}}$.
3. $\mathbf{X} \in \mathcal{U}_{\mathrm{ext}}^{\mathrm{sym}} \Leftrightarrow \Phi(\mathbf{X}) \in \mathbf{GFilt}^{\mathrm{clust}}$. □

Now let's see how each $\mathbf{X} = (X, u_X) \in \mathcal{U}_{\mathrm{ext}}^{\mathrm{sym}}$ can be biunivocally associated to a function $T_X\colon \mathbb{R} \to \mathrm{SubPart}(X)$, where $\mathrm{SubPart}(X)$ is the set of partitions of subsets of X. Consider $\Phi(\mathbf{X}) = \{\mathbf{X}_\epsilon\}_{\epsilon\in\mathbb{R}}$, which is in $\mathbf{GFilt}^{\mathrm{clust}}$ by Proposition 3.7. Write $\mathbf{X}_\epsilon = (V_\epsilon, E_\epsilon)$. For each $\epsilon \in \mathbb{R}$, consider the following equivalence relation in V_ϵ:

$$x \sim_\epsilon x' \Leftrightarrow x \xrightarrow[=]{\mathbf{X}_\epsilon} x'.$$

Define $T_X(\epsilon)$ as the set of equivalence classes of V_ϵ under this relation, that is, $T_X(\epsilon)$ is the set of connected components of $\mathbf{X}_\epsilon$. The map T_X, called a *treegram*, satisfies the following properties:

1. (Boundary conditions) $T_X(\epsilon) = \emptyset$, for $\epsilon < \epsilon_0 = \min_{x,x' \in X} u_X(x,x')$; and $T_X(\epsilon) = T_X(\epsilon_1)$ for $\epsilon \geq \epsilon_1 = \max\{u_X(X \times X) \setminus \{+\infty\}\}$.

2. (Hierarchy) For any $\epsilon \leq \epsilon'$, $x \sim_\epsilon x'$ implies $x \sim_{\epsilon'} x'$.

3. (Right continuity) For any ϵ, there exists $\delta > 0$ such that $T_X(\epsilon) = T_X(\epsilon + r)$, for any $r \in [0, \delta]$.

The term "treegram" is taken from [69], but here it is a bit more general since we don't need condition 2 in [69, Definition 1]: that $T_X(\epsilon) = \{X\}$ for some ϵ. We depict a treegram as in Figure 3.1. Reading it from left to right, a vertex x appears in the tree at the parameter $u_X(x,x)$ and merges with x' at parameter $u_X(x,x')$. See Example 3.8.

Example 3.8. Let $\mathbf{X} = (X, u_X) \in \mathcal{U}^{\mathrm{sym}}_{\mathrm{ext}}$ be as follows: X is the set $\{x_1, \ldots, x_4\}$ and u_X is given by the following matrix:

$$u_X = \begin{pmatrix} 0 & 3 & 4 & +\infty \\ 3 & 1 & 4 & +\infty \\ 4 & 4 & 2 & +\infty \\ +\infty & +\infty & +\infty & 0 \end{pmatrix},$$

where the entry (i,j) of the matrix is equal to $u_X(x_i, x_j)$. We depict $\mathbf{X}$ as in Figure 3.1.

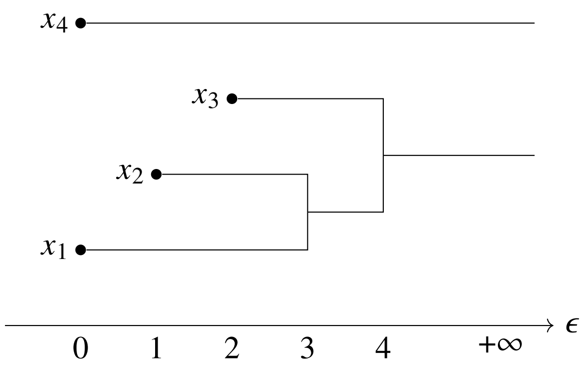


Figure 3.1: A graphical representation of the treegram associated to $\mathbf{X}$ (see Example 3.8). Notice that two branches may never merge.

Theorem 3.9. There exists a bijection between $\mathcal{U}^{\mathrm{sym}}_{\mathrm{ext}}$ and the set of treegrams.

Proof. By Proposition 3.7, there is a bijection between $\mathcal{U}^{\mathrm{sym}}_{\mathrm{ext}}$ and $\mathbf{GFilt}^{\mathrm{clust}}$. In the definition of treegram, we already constructed a map that associates an element of $\mathbf{GFilt}^{\mathrm{clust}}$ to a treegram. Let's define its inverse.

Let $T_X : \mathbb{R} \to \mathrm{SubPart}(X)$ be a treegram and let ϵ be any real number. Write $T_X(\epsilon) = \{X_1, \ldots, X_{n_\epsilon}\}$. Consider the graph filtration $\mathbf{G} = \{G_\epsilon\}_{\epsilon \in \mathbb{R}}$ where G_ϵ is the union of the complete graphs $K_{(X_i)}$. Properties 1–3 of the definitions of graph filtration and treegram ensure that this is a well-define map.

It is a straightforward verification that this is the desired bijection. □

3.2 Hierarchical clustering of extended networks

In Chapter 2 we studied some clustering functors on graphs. Now we will impose some conditions on these functors so that they induce "well-behaved" hierarchical clustering functors on the category $\mathcal{N}_{\text{ext}}$.

Remark 3.10. Let $\mathfrak{F}_1, \mathfrak{F}_2 \colon \mathcal{G} \to \mathcal{G}$ with $\mathfrak{F}_1 \leq \mathfrak{F}_2$, and $\mathbf{X} = (X, w_X) \in \mathcal{N}_{\text{ext}}$. Let $\mathfrak{F}_1(\mathbf{X}) = (X, w_1^{\mathfrak{F}})$ and $\mathfrak{F}_2(\mathbf{X}) = (X, w_2^{\mathfrak{F}})$. For any $\epsilon \in \mathbb{R}$, we have $\mathfrak{F}_1(\mathbf{X}_\epsilon) \hookrightarrow \mathfrak{F}_2(\mathbf{X}_\epsilon)$. Then for any $x, x' \in X$, we have $w_2^{\mathfrak{F}}(x, x') \leq w_1^{\mathfrak{F}}(x, x')$.

Definition 3.11 (Axiom of value)**.** We say that the endofunctor $\mathfrak{F} \colon \mathcal{G} \to \mathcal{G}$ satisfies the *axiom of value (for graphs)*, or simply **A1**, if $\mathfrak{F}(L_2) = D_2$ and $\mathfrak{F}(K_2) = K_2$. This definition is analogous to the one in [7].

Notice that the condition $\mathfrak{F}(L_2) = D_2$ implies $\mathfrak{F}(D_2) = D_2$.

Remark 3.12. In [7], the authors are concerned with functors

$$\mathfrak{H} \colon \mathcal{N}_{\text{dis}} \to \mathcal{N}_{\text{dis}},$$

where $\mathcal{N}_{\text{dis}}$ is the subcategory of $\mathcal{N}$ whose objects are *dissimilarity networks:* pairs (X, w_X) where $w_X(x, x') \geq 0$ for all $x, x' \in X$, and $w_X(x, x') = 0 \Leftrightarrow x = x'$. Such functors are required to satisfy that $\mathbf{Y} := \mathfrak{H}(\mathbf{X})$ be a symmetric ultranetwork for all $\mathbf{X} \in \mathcal{N}_{\text{dis}}$. The *axiom of value (for networks)* of [7] states that if for some ϵ one has $\mathbf{X}_\epsilon = D_2$ or L_2 then $\mathbf{Y}_\epsilon = D_2$, and if $\mathbf{X}_\epsilon = K_2$ then $\mathbf{Y}_\epsilon = K_2$. This definition in turn inspired Definition 3.11 . It is true that if $\mathfrak{F} \colon \mathcal{G} \to \mathcal{G}$ satisfies the axiom of value (for graphs) then $\widehat{\mathfrak{F}}$ satisfies the axiom of value (for networks). However, not all functors $\mathfrak{H}$ are equal to some $\widehat{\mathfrak{F}}$, as shown by Example 3.13 below.

Example 3.13. Let $\mathfrak{H} : \mathcal{N}_{\text{dis}} \to \mathcal{N}_{\text{dis}}$ be the *grafting functor* of [7, Proposition 3]. It is defined, for a given $\beta > 0$, as follows:

$$\big(\mathfrak{H}(\mathbf{X})\big)_\epsilon := \begin{cases} \mathfrak{F}^{\text{nrec}}(\mathbf{X}_\epsilon) \cap \mathfrak{F}^{\text{nrec}}(\mathbf{X}_\beta), & \epsilon \leq \beta, \\ \mathfrak{F}^{\text{rec}}(\mathbf{X}_\epsilon), & \epsilon > \beta, \end{cases}$$

for any $\mathbf{X} \in \mathcal{N}_{\text{dis}}$.

Take $\beta = 2$ and $\mathbf{X} = (X, w_X)$ where $X = \{x_1, x_2, x_3\}$ and the weight map w_X in matrix form is (see Figure 3.2)

$$w_X := \begin{pmatrix} 0 & 2 & 4 \\ 4 & 0 & 2 \\ 2 & 4 & 0 \end{pmatrix}.$$

Thus, $X_\epsilon = \emptyset$ for $\epsilon < 0$, $\mathbf{X}_\beta \cong C_3$, and

$$\mathbf{X}_\epsilon \cong \begin{cases} D_3, & 0 \leq \epsilon < 2, \\ C_3, & 2 \leq \epsilon < 4, \\ K_3, & \epsilon \geq 4. \end{cases}$$

If $\mathfrak{H}$ were equal to $\widehat{\mathfrak{F}}$, for some endofunctor $\mathfrak{F}$, we should have, for $\epsilon = 2$,

$$\mathfrak{F}(C_3) \cong \mathfrak{F}(\mathbf{X}_\epsilon) = \big(\mathfrak{H}(\mathbf{X})\big)_\epsilon = \mathfrak{F}^{\text{nrec}}(\mathbf{X}_\epsilon) \cap \mathfrak{F}^{\text{rec}}(\mathbf{X}_\beta) \cong K_3 \cap D_3 = D_3.$$

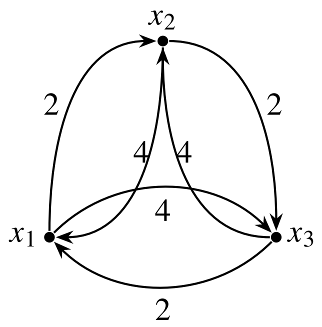


Figure 3.2: Dissimilarity network $\mathbf{X}$ from Example 3.13

Now considering $\mathbf{Y} = (Y, w_Y)$ where $Y = X$ and w_Y is obtained dividing w_X by 2, we obtain, for $\epsilon = 1$,

$$\mathfrak{F}(C_3) \cong \mathfrak{F}(\mathbf{Y}_\epsilon) = \big(\mathfrak{H}(\mathbf{Y})\big)_\epsilon = \mathfrak{F}^{\mathrm{nrec}}(\mathbf{Y}_\epsilon) \cap \mathfrak{F}^{\mathrm{rec}}(\mathbf{Y}_\beta) \cong K_3 \cap K_3 = K_3.$$

Thus, $\mathfrak{F}(C_3) = D_3$ and $\mathfrak{F}_{(}C_3) = K_3$, a contradiction.

Example 3.13 above shows that not all functors $\mathfrak{H}\colon \mathcal{N}_{\mathrm{dis}} \to \mathcal{N}_{\mathrm{dis}}$ arise as some $\widehat{\mathfrak{F}}$. On the other hand, the functors $\widehat{\mathfrak{F}}$ have as domain the category of extended networks, which contains the category of dissimilarity networks. Most functors $\mathfrak{H}\colon \mathcal{N}_{\mathrm{dis}} \to \mathcal{N}_{\mathrm{dis}}$ in [7] are equal to some $\widehat{\mathfrak{F}}$: the reciprocal, non-reciprocal, semi-reciprocal, directed single linkage and unilateral functors. Another advantage of considering functors $\widehat{\mathfrak{F}}$ is that they are simpler in the sense that we can work at the graph level instead of the network level: in order to determine $\widehat{\mathfrak{F}}(\mathbf{X})_\epsilon$ all we need to consider is $\mathbf{X}_\epsilon$, in contrast to the representable functors of [4], which implicitly assume knowledge of the network at (some) parameters greater than ϵ. As a consequence, the proofs are much shorter, including the proof of stability (see Theorem 3.28), which has as a corollary the stability of all functors mentioned in the previous sentence.

Remark 3.14. Property **A1** is not related to symmetry nor transitivity, as the following examples show:

- $\mathfrak{F}$ transitive $\not\Rightarrow$ $\mathfrak{F}$ satisfies **A1**. To see this, consider $\mathfrak{F}^{\mathrm{tc}}$.
- $\mathfrak{F}$ satisfies **A1** $\not\Rightarrow$ $\mathfrak{F}$ is transitive. Consider $\mathfrak{F}^{\mathrm{ls}}$.
- $\mathfrak{F}$ symmetric $\not\Rightarrow$ $\mathfrak{F}$ satisfies **A1**. Consider $\mathfrak{F}^{\mathrm{us}}$.
- $\mathfrak{F}$ satisfies **A1** $\not\Rightarrow$ $\mathfrak{F}$ is symmetric. Consider $\mathfrak{F}^{\Omega^*}$ where $\Omega^* = \{(C_3, a_2, a_3)\}$. Then $\mathfrak{F}^{\Omega^*}(C_3) = \mathfrak{F}^{\mathrm{rev}}(C_3) \notin \mathcal{G}^{\mathrm{sym}}$.

Definition 3.15. Let $G = (V, E) \in \mathcal{G}$ and P a partition of V. Consider the map $\pi\colon V \to P$ which assigns to each x the unique block $B_x \in P$ containing x. Define $G_P \coloneqq (P, E_P)$, where $B \xrightarrow{G_P} B'$ if there are $x \in B$ and $x' \in B'$ such that $x \xrightarrow{G} x'$.

Now given $\mathfrak{F}\colon \mathcal{G} \to \mathcal{G}^{\mathrm{clust}}$, consider the set $P_{\mathfrak{F}}(G)$ of connected components of $\mathfrak{F}(G)$. This set is a partition of V which can be given as the following equivalence relation: $x \sim y$ $\Leftrightarrow x \xrightarrow[=]{\mathfrak{F}(G)} y$. Define $G_{\mathfrak{F}} \coloneqq G_P$, where $P = P_{\mathfrak{F}}(G)$.

Notice that G_P is the graph with vertex set P that has the least possible number of arrows such that $\pi\colon G \to G_P$ is a graph map; that is: if G' is another such graph, then $G_P \hookrightarrow G'$.

Proposition 3.16. Let $\mathfrak{F}^\Omega$ be a representable functor with $D_1 \notin \Omega$. Then $\mathfrak{F}^\Omega$ satisfies **A1** $\Leftrightarrow$ for any $\omega \in \Omega$ and any partition $P = \{A, B\}$ of V_ω (the vertex set of ω) into two blocks, we have $\omega_P = K_2$ (as in Definition 3.15).

Proof. The claim follows from the following equivalences:

- $\mathfrak{F}^\Omega(D_2) = D_2 \Leftrightarrow$ every $\omega \in \Omega$ is connected $\Leftrightarrow$ for any partition $P = \{A, B\}$ of V_ω into two blocks, we have $\omega_P \neq D_2$.
- $\mathfrak{F}^\Omega(L_2) = D_2 \Leftrightarrow$ for any $\omega \in \Omega$, there is no surjective graph map $\omega \to L_2 \Leftrightarrow$ for any $\omega \in \Omega$ and any partition $P = \{A, B\}$ of V_ω into two blocks, we have $\omega_P = K_2$.
- $\mathfrak{F}^\Omega(K_2) = K_2 \Leftrightarrow$ some $\omega \in \Omega$ has more than one vertex. □

Remark 3.17. Theorem 3.23 below states that $\mathfrak{F}^{\mathrm{rec}} \leq \mathfrak{F} \leq \mathfrak{F}^{\mathrm{nrec}}$ for all clustering functors $\mathfrak{F}$ satisfying **A1**. In order to prove this, we will adapt the ideas that led to the proof of [7, Theorem 4] to the setting of functors $\widehat{\mathfrak{F}}$ arising from functors $\mathfrak{F}\colon \mathcal{G} \to \mathcal{G}^{\mathrm{clust}}$.

Definition 3.18 (Extended axiom of value)**.** We say that $\mathfrak{F}\colon \mathcal{G} \to \mathcal{G}^{\mathrm{clust}}$ satisfies the extended axiom of value (or simply **A1′**) if $\mathfrak{F}(T_n) = D_n$ and $\mathfrak{F}(K_n) = K_n$, $\forall n \in \mathbb{N}$.

Theorem 3.19. Let $\mathfrak{F}\colon \mathcal{G} \to \mathcal{G}^{\mathrm{clust}}$. Then, $\mathfrak{F}$ satisfies **A1** $\Leftrightarrow$ $\mathfrak{F}$ satisfies **A1′**.

Proof. Let's prove the ($\Rightarrow$) implication. Pick $a_i, a_j \in T_n$ with $i < j$. Consider the partition $A = \{a_k \mid k < j\}$, $B = \{a_k \mid k \geq j\}$ and the graph map $\phi\colon T_n \to L_2$ given by $\phi(A) = b_1$ and $\phi(B) = b_2$, where $b_1 \to b_2$ is the only arrow in L_2. By functoriality of $\mathfrak{F}$, $\phi\colon \mathfrak{F}(T_n) \to \mathfrak{F}(L_2) = D_2$ is a graph map. Thus, we cannot have $a_i \xrightarrow{\mathfrak{F}(T_n)} a_j$. Since $\mathfrak{F}(T_n)$ is symmetric and a_i and a_j were arbitrary, we have $\mathfrak{F}(T_n) = D_n$. That $\mathfrak{F}(K_n) = K_n$ follows from Proposition 2.11, since $\mathfrak{F}$ satisfies **A1**, which implies $\mathfrak{F} \neq \mathfrak{F}^{\mathrm{disc}}$.

Now let's prove the ($\Leftarrow$) implication. Taking $n = 2$ gives us $T_2 = L_2$. This implies that $\mathfrak{F}(L_2) = D_2$ and $\mathfrak{F}(K_2) = K_2$. □

Definition 3.20. We say that a graph G *has no cycles* if there exists no sequence of points $v_1, \ldots, v_m \in V$ with $v_1 \xrightarrow{G} v_2 \xrightarrow{G} \cdots \xrightarrow{G} v_m \xrightarrow{G} v_1$. This is equivalent to imposing that $\mathfrak{F}^{\{C_n\}}(G) = D_{(G)}$ for any $n \geq 2$, where C_n is the cycle graph.

Lemma 3.21. If $G = (V, E) \in \mathcal{G}$ has no cycles, then there is a bijective map $\phi\colon V \to A_{T_n}$, where $n = |V|$ and $A_{T_n} = \{a_1, \ldots, a_n\}$ is the vertex set of the graph T_n. Moreover, ϕ is a graph map.

Proof. For each $v \in G$, let $P(v) = \{v' \mid v' \xrightarrow{G} v\}$.

Claim 3.21.1. There exists $v_1^* \in G$ such that $P(v_1^*) = \emptyset$.

Proof. If the claim were false, given $v_0 \in G$, there would exist $v_1, \ldots, v_n \in G$ such that $v_0 \xleftarrow{G} v_1 \xleftarrow{G} \cdots \xleftarrow{G} v_n$. Since $|V| = n$, we must have $v_i = v_j$ for some i, j, which would contradict the assumption that G has no cycles. ■

Now let $G_1 = G \setminus \{v_1^*\}$. Let $P_1(v) = \{v' \mid v' \xrightarrow{G_1} v\}$. By the above argument, there is $v_2^* \in G_1$ such that $P_1(v_2^*) = \emptyset$, which implies $P(v_2^*) \subseteq \{v_1^*\}$.

Let $G_2 = G_1 \setminus \{v_2^*\} = G \setminus \{v_1^*, v_2^*\}$ and let $P_2(v) = \{v' \mid v' \xrightarrow{G_2} v\}$. We then find $v_3^* \in G_2$ such that $P_2(v_3^*) = \emptyset$, which implies $P(v_3^*) \subseteq \{v_1^*, v_2^*\}$.

By the same argument, for $v_3^*, \dots, v_n^*$ we will obtain $P(v_i^*) \subseteq \{v_1^*, \dots, v_{i-1}^*\}$. Then, for any $i < j$, we have $v_j^* \not\twoheadrightarrow v_i^*$ in G. Thus, the map $\phi\colon G \to T_n$ given by $\phi(v_i^*) = a_i$, $i = 1, \dots, n$, is a bijective graph map. □

Theorem 3.22. Suppose $G = (V, E) \in \mathcal{G}$ has no cycles. If $\mathfrak{F}\colon \mathcal{G} \to \mathcal{G}^{\mathrm{clust}}$ satisfies **A1**, then $\mathfrak{F}(G) = D_{(G)}$.

Proof. Let $\phi\colon G \to T_n$ be the graph map of Lemma 3.21, where $n = |V|$. By functoriality of $\mathfrak{F}$, $\phi\colon \mathfrak{F}(G) \to \mathfrak{F}(T_n)$ is a graph map. By Theorem 3.19, $\mathfrak{F}(T_n) = D_n$. Thus, $\mathfrak{F}(G) = D_{(G)}$. □

Theorem 3.23. If $\mathfrak{F}\colon \mathcal{G} \to \mathcal{G}^{\mathrm{clust}}$ satisfies **A1**, then

$$\mathfrak{F}^{\mathrm{rec}} \le \mathfrak{F} \le \mathfrak{F}^{\mathrm{nrec}}.$$

Proof. Let $G = (V, E) \in \mathcal{G}$.

For the second inequality, let $\tilde{G} = G_{\mathfrak{F}^{\mathrm{nrec}}} = (\tilde{V}, \tilde{E})$ as in Definition 3.15.

Claim 3.23.1. $\tilde{G}$ has no cycles.

Proof. Indeed, if we had a cycle $x_1 \to x_2 \to \cdots \to x_n \to x_1$ in $\tilde{G}$, this would imply that there are $v_i, \tilde{v}_i \in x_i$ with

$$v_1 \rightsquigarrow \tilde{v}_1 \to v_2 \rightsquigarrow \tilde{v}_2 \to \cdots \to v_n \rightsquigarrow \tilde{v}_n \to v_1 \text{ in } G,$$

hence $x_1 = x_2 = \cdots = x_n$, a contradiction. ■

Consider the graph map given by the projection $\pi\colon G \to \tilde{G}$, $\pi(v) = [v]$. Notice that $[v] \ne [v']$ is equivalent to $v \not\twoheadrightarrow v'$ in $\mathfrak{F}^{\mathrm{nrec}}(G)$. Applying $\mathfrak{F}$, we obtain $\pi\colon \mathfrak{F}(G) \to \mathfrak{F}(\tilde{G}) = D_{(\tilde{G})}$, by Theorem 3.22. This implies that if $v, v' \in G$ satisfy $[v] \ne [v']$, then $v \not\twoheadrightarrow v'$ in $\mathfrak{F}(G)$, or, by the contrapositive, if $v \xrightarrow{\mathfrak{F}(G)} v'$ then $v \xrightarrow{\mathfrak{F}^{\mathrm{nrec}}(G)} v'$, that is, $\mathfrak{F}(G) \le \mathfrak{F}^{\mathrm{nrec}}(G)$.

Now for the leftmost inequality: if $v \xrightarrow{\mathfrak{F}^{\mathrm{rec}}(G)} v'$ then there are $v = v_1, v_2, \dots, v_n = v'$ such that $v_1 \leftrightarrow v_2 \leftrightarrow \cdots \leftrightarrow v_n$ in G. Consider the graph maps $\phi_i\colon K_2 \to (G, v_i, v_{i+1})$ for $i = 1, \dots, n$. Then we also have graph maps $\phi_i\colon \mathfrak{F}(K_2) \to (\mathfrak{F}(G), v_i, v_{i+1})$. Since $\mathfrak{F}$ satisfies **A1**, $\mathfrak{F}(K_2) = K_2$, which implies $v_i \xrightarrow{\mathfrak{F}(G)} v_{i+1}$. By the transitivity of $\mathfrak{F}(G)$, we obtain $v \xrightarrow{\mathfrak{F}(G)} v'$. Thus, $\mathfrak{F}^{\mathrm{rec}} \le \mathfrak{F}$. □

Corollary 3.24. Let $\mathfrak{F}\colon \mathcal{G} \to \mathcal{G}^{\mathrm{clust}}$ be a functor satisfying **A1**. For any extended network $\mathbf{X} = (X, w_X)$, denoting $\widehat{\mathfrak{F}}(\mathbf{X}) = (X, u_X^{\mathfrak{F}})$, $\widehat{\mathfrak{F}}^{\mathrm{nrec}}(\mathbf{X}) = (X, u_X^{\mathrm{nrec}})$ and $\widehat{\mathfrak{F}}^{\mathrm{rec}}(\mathbf{X}) = (X, u_X^{\mathrm{rec}})$, we have

$$u_X^{\mathrm{nrec}} \le u_X^{\mathfrak{F}} \le u_X^{\mathrm{rec}}.$$

Proof. Follows from Theorem 3.23 and Remark 3.10. □

Remark 3.25. We can define, as in [7], an alternative axiom of value: we say that $\mathfrak{F}\colon \mathcal{G} \to \mathcal{G}^{\text{trans}}$ satisfies axiom **A1″** if $\mathfrak{F}(D_2) = D_2$ and $\mathfrak{F}(L_2) = L_2$ (which implies $\mathfrak{F}(K_2) = K_2$). By Proposition 2.12, $\mathfrak{F}^{\text{tc}}$ is the unique endofunctor on $\mathcal{G}$ that satisfies this axiom.

Restricted to $\mathcal{U}^{\text{sym}}$, the map $\widehat{\mathfrak{F}}^{\text{tc}}$ is called *network single linkage hierarchical clustering functor* (see [69]).

When $\mathbf{X} = (X, w_X)$ is a dissimilarity network and we write $\mathfrak{F}^{\text{tc}}(\mathbf{X}) = (X, u^{\text{tc}})$, the value $u^{\text{tc}}(x, x')$ is called the *directed minimum chain cost of* (x, x') in [7], and $\widehat{\mathfrak{F}}^{\text{tc}}$ is the *directed single linkage*. Our result is in agreement with [7, Theorem 7], which states that directed single linkage is the unique functor satisfying **A1″** (for dissimilarity networks).

Remark 3.26. By Theorem 2.17, $\mathfrak{F}(\mathcal{G}^{\text{sym}}) \subset \mathcal{G}^{\text{sym}}$, so, given any two functors $\mathfrak{F}_1\colon \mathcal{G} \to \mathcal{G}^{\text{sym}}$ and $\mathfrak{F}_2\colon \mathcal{G} \to \mathcal{G}^{\text{trans}}$, their composite $\mathfrak{F}_2 \circ \mathfrak{F}_1$ is a clustering functor. Examples are: the reciprocal clustering $\mathfrak{F}^{\text{rec}} = \mathfrak{F}^{\text{tc}} \circ \mathfrak{F}^{\text{ls}}$; the non-reciprocal clustering $\mathfrak{F}^{\text{nrec}} = \mathfrak{F}^{\text{tc}} \circ \mathfrak{F}^{\Omega}$, where $\Omega = \{C_n\}_{n\in\mathbb{N}}$; the semi-reciprocal with size t, $\mathfrak{F}^{\text{semi-}t} = \mathfrak{F}^{\text{tc}} \circ \mathfrak{F}^{\text{ls}} \circ \mathfrak{F}^{[t]}$.

This gives us a simple recipe to construct clustering functors on (extended) networks: just take $\widehat{\mathfrak{F}}$ for $\mathfrak{F} = \mathfrak{F}_2 \circ \mathfrak{F}_1$ as above.

3.2.1 Stability

In [8] the authors defined a distance on the collection of all dissimilarity networks, and in [68] this distance, denoted by $d_{\mathcal{N}}$, was extended to the collection of all networks. Using this distance, we can then ask whether a map $\widehat{\mathfrak{F}} : \mathcal{N} \to \mathcal{N}$, for some endofunctor $\mathfrak{F}$, is *stable* with respect to $d_{\mathcal{N}}$. More precisely: given two networks $\mathbf{X}$ and $\mathbf{Y}$, is it true that $d_{\mathcal{N}}\left(\widehat{\mathfrak{F}}(\mathbf{X}), \widehat{\mathfrak{F}}(\mathbf{Y})\right) \leq d_{\mathcal{N}}(\mathbf{X}, \mathbf{Y})$?

Definition 3.27 ([68])**.** Let $\mathbf{X} = (X, w_X), \mathbf{Y} = (Y, w_Y)$ be two networks. A *correspondence* between X and Y is a subset $R \subseteq X \times Y$ such that the projections $\pi_X\colon R \to X$ and $\pi_Y\colon R \to Y$ are surjective. Denote by $C(X, Y)$ the set of all correspondences between X and Y. The *distortion* of R (with respect to $\mathbf{X}$ and $\mathbf{Y}$) is the quantity

$$\operatorname{dis}(R; \mathbf{X}, \mathbf{Y}) := \max_{(x,y),(x',y')\in R} \left| w_X(x, x') - w_Y(y, y') \right|.$$

The *network distance* between $\mathbf{X}$ and $\mathbf{Y}$ is defined as

$$d_{\mathcal{N}}(\mathbf{X}, \mathbf{Y}) := \frac{1}{2} \inf_{R\in C(X,Y)} \operatorname{dis}(R; \mathbf{X}, \mathbf{Y}).$$

We refer the reader to [68] to see why this is a generalization of the Gromov-Hausdorff distance and the proof that it is actually a pseudometric, among other properties.

Theorem 3.28 (Stability)**.** Let $\mathfrak{F}\colon \mathcal{G} \to \mathcal{G}$ be any endofunctor such that $\mathfrak{F} \neq \mathfrak{F}^{\text{disc}}$. Let $\mathbf{X} = (X, w_X)$ and $\mathbf{Y} = (Y, w_Y)$ be any networks. Then,

$$d_{\mathcal{N}}\left(\widehat{\mathfrak{F}}(\mathbf{X}), \widehat{\mathfrak{F}}(\mathbf{Y})\right) \leq d_{\mathcal{N}}(\mathbf{X}, \mathbf{Y}).$$

Proof. Let $R \in C(X, Y)$ be such that $\eta := \operatorname{dis}(R; \mathbf{X}, \mathbf{Y}) = 2d_{\mathcal{N}}(\mathbf{X}, \mathbf{Y})$. Write $\widehat{\mathfrak{F}}(\mathbf{X}) = (X, w_X^{\mathfrak{F}})$ and $\widehat{\mathfrak{F}}(\mathbf{Y}) = (Y, w_Y^{\mathfrak{F}})$.

Let $(x, y), (x', y') \in R$. Let $\phi\colon X \to Y$ be given by $\phi(x) = y$, $\phi(x') = y'$ and $\phi(a) = b$, for any choice of $b \in Y$ such that $(a, b) \in R$, $a \neq x, x'$.

Claim 3.28.1. The map ϕ induces a graph map $\phi_\epsilon\colon \mathbf{X}_\epsilon \longrightarrow \mathbf{Y}_{\epsilon+\eta}$ given by $\phi_\epsilon(a) = \phi(a)\ \forall\epsilon$.

Proof. Indeed, suppose $a \xrightarrow[=]{\mathbf{X}_\epsilon} a'$, that is, $w_X(a,a') \le \epsilon$. Since $(a,\phi(a)),(a',\phi(a')) \in R$, we have

$$|w_X(a,a') - w_Y(\phi(a),\phi(a'))| \le \eta \Rightarrow w_Y(\phi(a),\phi(a')) \le w_X(a,a') + \eta \le \epsilon + \eta,$$

i.e., $\phi(a) \xrightarrow[=]{\mathbf{Y}_{\epsilon+\eta}} \phi(a')$. ■

Now let $\epsilon := w_X^{\mathfrak{F}}(x,x')$, and so $x \xrightarrow[=]{\mathfrak{F}(\mathbf{X}_\epsilon)} x'$. Observe that ϵ is finite since $\mathfrak{F} \neq \mathfrak{F}^{\mathrm{disc}}$ and for some $\delta \in \mathbb{R}$ we will have that $\mathbf{X}_\delta$ and $\mathfrak{F}(\mathbf{X}_\delta)$ are complete graphs.

Further, notice that the map ϕ_ϵ satisfies $\phi_\epsilon\colon (\mathbf{X}_\epsilon, x, x') \longrightarrow (\mathbf{Y}_{\epsilon+\eta}, y, y')$. The functoriality of $\mathfrak{F}$ implies that $\phi_\epsilon\colon (\mathfrak{F}(\mathbf{X}_\epsilon), x, x') \longrightarrow (\mathfrak{F}(\mathbf{Y}_{\epsilon+\eta}), y, y')$ is a graph map. Then $y \xrightarrow[=]{\mathfrak{F}(\mathbf{Y}_{\epsilon+\eta})} y'$, i.e., $w_Y^{\mathfrak{F}}(y,y') \le \epsilon + \eta$. Thus,

$$w_Y^{\mathfrak{F}}(y,y') - w_X^{\mathfrak{F}}(x,x') \le \eta.$$

Let $\psi\colon Y \longrightarrow X$ be such that $\psi(y) = x$, $\psi(y') = x'$ and $(\psi(b), b) \in R$, for all $b \in Y$, and obtain

$$w_X^{\mathfrak{F}}(x,x') - w_Y^{\mathfrak{F}}(y,y') \le \eta.$$

Both inequalities above lead to

$$\left|w_X^{\mathfrak{F}}(x,x') - w_Y^{\mathfrak{F}}(y,y')\right| \le \eta.$$

But, since $(x,y),(x',y') \in R$ were chosen arbitrarily, we have

$$\operatorname{dis}\left(R; \widehat{\mathfrak{F}}(\mathbf{X}), \widehat{\mathfrak{F}}(\mathbf{Y})\right) \le \eta = \operatorname{dis}(R; \mathbf{X}, \mathbf{Y}),$$

which implies $d_{\mathcal{N}}\left(\widehat{\mathfrak{F}}(\mathbf{X}), \widehat{\mathfrak{F}}(\mathbf{Y})\right) \le d_{\mathcal{N}}(\mathbf{X}, \mathbf{Y})$. □

Corollary 3.29. The reciprocal $\mathfrak{H}^{\mathrm{rec}}$, non-reciprocal $\mathfrak{H}^{\mathrm{nrec}}$, unilateral $\mathfrak{H}^{\mathrm{u}}$ and directed single linkage $\mathfrak{H}^{\mathrm{dsl}}$ functors from [7] are stable.

Proof. Indeed, $\mathfrak{H}^{\mathrm{nrec}} = \widehat{\mathfrak{F}}^{\mathrm{nrec}}$, $\mathfrak{H}^{\mathrm{rec}} = \widehat{\mathfrak{F}}^{\mathrm{rec}}$, $\mathfrak{H}^{\mathrm{u}} = \widehat{\mathfrak{F}}$ where $\mathfrak{F} = \mathfrak{F}^{\mathrm{tc}} \circ \mathfrak{F}^{\mathrm{us}}$, and $\mathfrak{H}^{\mathrm{dsl}} = \widehat{\mathfrak{F}}^{\mathrm{tc}}$. □

3.3 Hierarchical clustering of a graph

An extended network can be seen as a sequence of nested graphs. The method studied in Section 3.2 arose by applying an endofunctor to this sequence.

An alternative approach is to fix one graph and consider many "nested" endofunctors, as in the following definition.

Definition 3.30. Let $\Omega_1, \Omega_2, \ldots, \Omega_n$ be families of graphs with $\Omega_1 \preccurlyeq \Omega_2 \preccurlyeq \ldots \preccurlyeq \Omega_n$. Let $\mathfrak{F}_i := \mathfrak{F}^{\mathrm{tc}} \circ \mathfrak{F}^{\Omega_i}$, for all $i = 1, \ldots, n$. Since $\mathfrak{F}^{\mathrm{tc}}$ is arrow increasing and $\mathfrak{F}^{\Omega_i} \le \mathfrak{F}^{\Omega_{i+1}}$, $i = 1, \ldots, n-1$, we have n clustering functors $\mathfrak{F}_1 \le \mathfrak{F}_2 \le \cdots \le \mathfrak{F}_n$. Take $\mathfrak{F}_0 := \mathfrak{F}^{\mathrm{disc}}$ and $\mathfrak{F}_{n+1} := \mathfrak{F}^{\mathrm{comp}}$. For a given graph $G = (V,E)$, define $u\colon V \times V \longrightarrow \mathbb{R}$ by

$$u(v,v') := \min\left\{i \mid v \xrightarrow{\mathfrak{F}_i(G)} v'\right\}.$$

We call (V,u) the *(symmetric extended) ultranetwork of G obtained from $\Omega_1, \ldots, \Omega_n$*.

Note that it is not interesting to have $\mathfrak{F}_i = \mathfrak{F}_j$ for $i \neq j$. The sets $\Omega_i = \{K_{i+1}\}$ are an example of this, since $\mathfrak{F}_i = \mathfrak{F}^{\mathrm{tc}} \circ \mathfrak{F}^{\mathrm{us}}$, for any $i \geq 1$.

Example 3.31. For $k = 1, \ldots, 4$, let $\Omega_k = \{C_{k+1}\}$. Let G be the graph of Figure 3.3 (left). The ultranetwork of G obtained from $\Omega_1, \ldots, \Omega_4$ is depicted in Figure 3.3 (right).

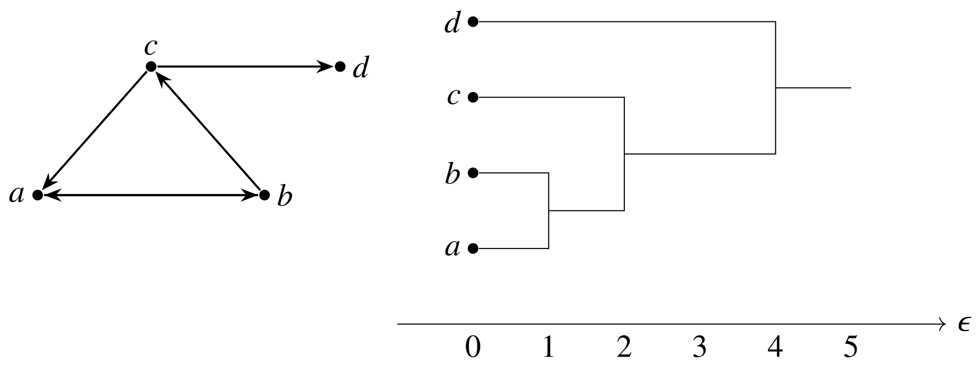


Figure 3.3: On the left: a graph G. On the right: the ultranetwork of G obtained from $\Omega_1, \ldots, \Omega_4$. See Example 3.31

The following proposition states when Ω_i satisfies $\mathfrak{F}_i(L_2) = K_2$, for some i, then all $\mathfrak{F}_j$ with $j > i$ are redundant.

Proposition 3.32. If Ω is such that $\mathfrak{F}^{\Omega}(L_2) = K_2$, then $\mathfrak{F}^{\mathrm{tc}} \circ \mathfrak{F}^{\Omega} = \mathfrak{F}^{\mathrm{conn}}$ or $\mathfrak{F}^{\mathrm{comp}}$.

Proof. Let $G \in \mathcal{G}$. Then $\mathfrak{F}^{\Omega}(L_2) = K_2$ implies $\mathfrak{F}^{\mathrm{us}} \leq \mathfrak{F}$. Thus,

$$\mathfrak{F}^{\mathrm{conn}} = \mathfrak{F}^{\mathrm{tc}} \circ \mathfrak{F}^{\mathrm{us}} \leq \mathfrak{F}^{\mathrm{tc}} \circ \mathfrak{F}.$$

By Proposition 2.29, $\mathfrak{F}^{\mathrm{tc}} \circ \mathfrak{F} = \mathfrak{F}^{\mathrm{conn}}$ or $\mathfrak{F}^{\mathrm{comp}}$. $\square$

4 Homology theories for graphs and networks

There is no beginning, no middle, no end, no suspense, no moral, no causes, no effects. What we love in our books are the depths of many marvelous moments seen all at one time.

KURT VONNEGUT, *SLAUGHTERHOUSE FIVE*

In this chapter we will study different ways to define homology theories on graphs and, by extension, on networks. Our main contribution is Section 4.1, where we adapt the definition of a representable functor $\mathfrak{F}^{\Omega}$ to create (ordered) simplicial complexes instead of graphs.

Many approaches have already been made to define the homology of a graph: in [71,72], the authors study how to map a graph into a simplicial complex by considering all its cliques; a singular homology for graphs was defined in [71,73]; a homotopy theory for undirected graphs was studied in [74,75], which is compatible with the more general homotopy for directed graphs of [76]; there is also a cohomology theory by [77,78] which was further developed in [79]. As an application, the directed clique complex was used in [80] to study the human brain.

Here we will study the *path homology* defined in [81] and how it relates to other homology theories in some special cases.

4.1 Motivic homology

Definition 4.1. Given a graph $G = (V, E)$ and a family of graphs Ω, define the Ω*-complex of* G by

$$C^{\Omega}(G) = \{\sigma \subseteq V \mid \exists \omega \in \Omega,\ \exists \phi \colon \omega \to G \text{ graph map s.t. } \sigma \subseteq \operatorname{Im}(\phi)\}.$$

The homology group $H_p(C^{\Omega}(G))$ is denoted by $H_p^{\Omega}(G)$ and is called the p*-dimensional* Ω*-homology of* G.

Remark 4.2. If we regard a symmetric graph as a 1-dimensional simplicial complex (where we identify the arrows $v \to v'$ and $v' \to v$ with $\{v, v'\}$), it is immediate that $\mathfrak{F}^{\Omega}(G)$ is the 1-skeleton (the set of 1-dimensional simplexes) of $C^{\Omega}(G)$. Thus, we can think of the Ω-complex $C^{\Omega}(G)$ as an extension of the previous theory of representable functors on graphs. We can now study the homology of these new higher dimensional objects.

The next proposition states when two families of graphs generate the same simplicial complex.

Proposition 4.3. Let Ω_1 and Ω_2 be two families of graphs. Then $C^{\Omega_1}(G) \subseteq C^{\Omega_2}(G)$ for every $G \in \mathcal{G}$ if and only if for every $\omega \in \Omega_1$ there is $\widetilde{\omega} \in \Omega_2$ and a surjective graph map $p\colon \widetilde{\omega} \to \omega$.

Proof. For implication ($\Rightarrow$), let $\omega \in \Omega_1$ be any graph, and set $G = \omega$ and $n = |\omega|$. Since $C^{\Omega_1}(\omega) = \Delta_{n-1}$ is a $(n-1)$-simplex, $C^{\Omega_1}(\omega) \subseteq C^{\Omega_2}(\omega)$ and both $C^{\Omega_1}(\omega)$ and $C^{\Omega_2}(\omega)$ have the same vertex set (0-skeleton), we must have $C^{\Omega_2}(\omega) = \Delta_{n-1}$. Thus, there are $\widetilde{\omega} \in \Omega_2$ and a graph map $p\colon \widetilde{\omega} \to \omega$ such that $\omega \subseteq p(\widetilde{\omega})$, that is, p is surjective.

For the reciprocal implication ($\Leftarrow$), let G be any graph and let $\sigma \in C^{\Omega_1}(G)$. By definition, there are $\omega \in \Omega_1$ and a graph map $\phi\colon \omega \to G$ such that $\sigma \subseteq \phi(\omega)$. From our assumption, there are $\widetilde{\omega} \in \Omega_2$ and a surjective graph map $p\colon \widetilde{\omega} \to \omega$.

$$\begin{tikzcd}
\widetilde{\omega} \arrow[d, "p"'] \arrow[dr, "\phi\circ p"] & \\
\omega \arrow[r, "\phi"] & G
\end{tikzcd}$$

Thus, the graph map $\phi \circ p\colon \widetilde{\omega} \to G$ satisfies $\sigma \subseteq \phi \circ p(\widetilde{\omega})$, that is, $\sigma \in C^{\Omega_2}(G)$. $\square$

Corollary 4.4. $C^{\Omega_1}(G) = C^{\Omega_2}(G)$ for every graph G if and only if for every $\omega_i \in \Omega_i$ there are $\widetilde{\omega}_j \in \Omega_j$ and a surjective graph map $p_j\colon \widetilde{\omega}_j \to \omega_i$, for $i \neq j$ in $\{1, 2\}$.

It is worth comparing Proposition 4.3 with Corollary 2.42.

Definition 4.5. For any two graphs $G = (V, E)$ and $G' = (V', E')$, define the *join* $G * G'$ as the graph $G \sqcup G'$ with the additional arrows $v \to v'$, for all $v \in V$ and $v' \in V'$. More explicitly, $G * G' = (V \sqcup V', E \sqcup E' \sqcup (V \times V'))$. See Figure 4.1.

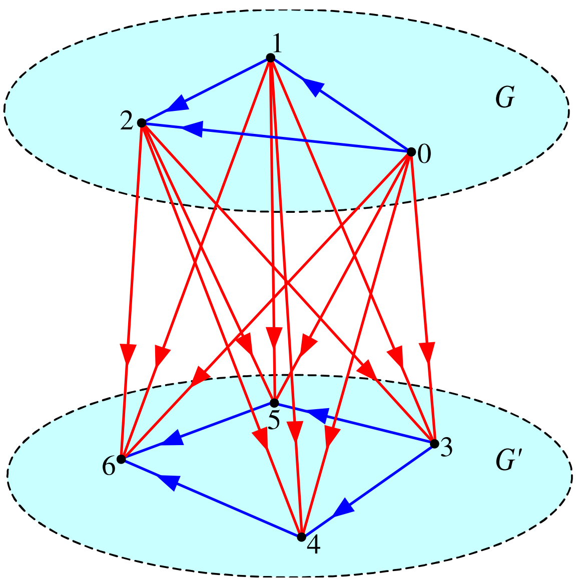


Figure 4.1: The join $G * G'$ of two graphs G and G'. Source of the image: [82].

Example 4.6. 1. Consider $\Omega^{\text{Rips}} := \{K_n\}_{n\geq 1}$. Then $C_p^{\Omega^{\text{Rips}}}(G)$ is called *the Rips complex of G*.

2. Consider $\Omega_{\text{si}}^{\text{Dow}} := \{S_n\}_{n\geq 2}$, where $S_n := D_{n-1} * D_1$. Then $C_p^{\Omega_{\text{si}}^{\text{Dow}}}(G)$ is called *the Dowker sink complex of G*. The *Dowker source complex of G* is given by $C^{\Omega_{\text{so}}^{\text{Dow}}}(G)$, where $\Omega_{\text{so}}^{\text{Dow}} = \{D_1 * D_{n-1}\}_{n\geq 2}$. By [18, Corollary 19], both sink and source complexes yield the same homology, so we can talk about *the* Dowker homology of G, and define $\Omega^{\text{Dow}} := \Omega_{\text{si}}^{\text{Dow}}$.

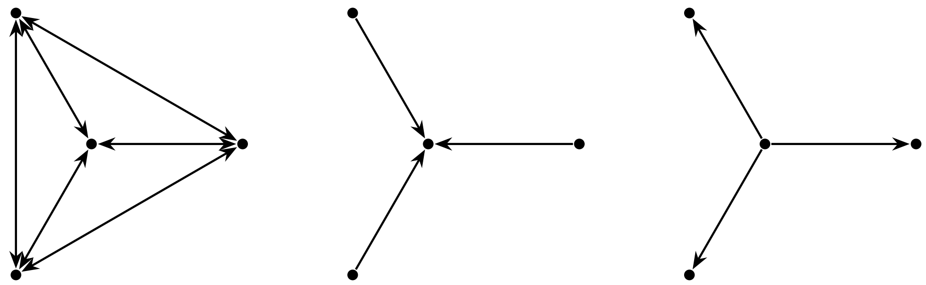


Figure 4.2: (a): The complete graph with 4 vertices, K_4. (b): The join $D_3 * D_1$ which composes $\Omega_{\text{si}}^{\text{Dow}}$. (c): The join $D_1 * D_3$ which composes $\Omega_{\text{so}}^{\text{Dow}}$.

Now let $\mathbf{X} = (X, d)$ be a metric space. Then $\{C^{\Omega^{\text{Rips}}}(\mathbf{X}_\epsilon)\}_{\epsilon\in\mathbb{R}}$ and $\{C^{\Omega^{\text{Dow}}}(\mathbf{X}_\epsilon)\}_{\epsilon\in\mathbb{R}}$ are the usual Vietoris-Rips and Čech simplicial filtrations as in Definitions 1.23 and 1.24. If $\mathbf{X} = (X, w_X)$ is a network, then these filtrations are the Rips and Dowker sink filtrations of [18].

The homologies on Sections 4.2 and 4.3 are built over *ordered* simplicial complexes. We will adapt Definition 4.1 to create a generalization of an ordered simplicial complexes.

Definition 4.7. An *ordered graph* is a pair (ω, l) where $\omega = (V_\omega, E_\omega) \in \mathcal{G}$ and

$$l\colon \{0, \dots, |\omega| - 1\} \to V_\omega$$

is a bijection. Write $l(j) = z_j$. We can omit l and simply state that ω has a total order on its vertex set $\{z_0, \dots, z_{N_\omega}\}$, given by $z_i \leq z_j \Leftrightarrow i \leq j$, where $N_\omega := |\omega| - 1$.

Let Ω^{ord} be a set of ordered graphs. Define the Ω^{ord}*-set of G*, denoted by $OC^{\Omega^{\text{ord}}}(G)$, as the union for $p \geq 0$ of the sets

$$OC_p^{\Omega^{\text{ord}}}(G) = \\ \{(v_0, \dots, v_p) \mid \exists \omega \in \Omega^{\text{ord}},\ \exists \phi\colon \omega \to G \text{ graph map s.t. } \phi(z_i) = v_i,\ i = 0, \dots, p\}.$$

We denote the fact that $\phi(z_i) = v_i$ for $i = 0, \dots, p$ by $(v_0, \dots, v_n) \in [\phi(\omega)]$. Notice that we do allow repetitions in the elements of $OC_p^{\Omega^{\text{ord}}}(G)$, for example, (v_0, v_0, v_0).

Remark 4.8. Let Ω^* be a family of pointed graphs and let $G = (V, E)$ be a graph. Write $\mathfrak{F} = \mathfrak{F}^{\Omega^*}$. Consider $\mathfrak{F}(G)$ as a set of sequences of V given by $V \cup E^{\mathfrak{F}}$, where $\mathfrak{F}(G) = (V, E^{\mathfrak{F}})$. Let Ω^{ord} be the union of all graphs $\omega = (V_\omega, E_\omega)$ such that $(\omega, z, \widehat{z}) \in \Omega^*$, where the first two elements of the order in V_ω are z and $\widehat{z}$, and the remaining are chosen arbitrarily: $V_\omega = \{z, \widehat{z}, \dots, z_{N_\omega}\}$. Then $\mathfrak{F}(G)$ consists of all elements of $OC^{\Omega^{\text{ord}}}(G)$ with length one or two. Thus, we can think of $OC^{\Omega^{\text{ord}}}(G)$ as an extension of the graph $\mathfrak{F}(G)$.

A pointed graph $(\omega, z, \widehat{z})$ is a graph with two distinguished points, and to define $\mathfrak{F}^{\Omega^*}$ it was fundamental to keep track of to which points z and $\widehat{z}$ were mapped.

The notion of ordered graph distinguishes all points of ω in a similar way.

Definition 4.9. Let K be a set of finite sequences of elements of V (shortly, a SoS of V). Consider the two following properties:

(P1) $(v_0, \dots, v_n) \in K \Rightarrow (v_0, \dots, v_i, v_i, \dots, v_n) \in K$, $i = 0, \dots, n$.

(P2) $(v_0, \dots, v_n) \in K \Rightarrow (v_0, \dots, \widehat{v_i}, \dots, v_n) \in K$, $i = 0, \dots, n$, where $\widehat{v_i}$ denotes the exclusion of v_i from the sequence.

If K satisfies property (P2), we say that it is an *ordered simplicial complex* on the vertex set V.

Now, let's explore conditions under which $OC^{\Omega^{\mathrm{ord}}}(G)$ satisfies properties (P1) and (P2).

Proposition 4.10. Suppose that given any $\omega = (V_\omega, E_\omega) \in \Omega^{\mathrm{ord}}$, with ordered vertex set $V_\omega = \{z_0, \dots, z_{N_\omega}\}$, given any $n \leq N_\omega$, and any $i = 0, \dots, n$, there exist $\widehat{\omega} \in \Omega^{\mathrm{ord}}$ and a graph map $p\colon \widehat{\omega} \to \omega$ such that $(z_0, \dots, z_i, z_i, \dots, z_n) \in [p(\widehat{\omega})]$. Then $OC^{\Omega^{\mathrm{ord}}}(G)$ satisfies property (P1), for any $G \in \mathcal{G}$.

Proof. Let $(v_0, \dots, v_n) \in OC^{\Omega^{\mathrm{ord}}}(G)$. By definition, there are $\omega \in \Omega^{\mathrm{ord}}$ and $\phi\colon \omega \to G$ a graph map such that $(v_0, \dots, v_n) \in [\phi(\omega)]$. By assumption, there is $p\colon \widehat{\omega} \to \omega$ with $(z_0, \dots, z_i, z_i, \dots, z_n) \in [p(\widehat{\omega})]$. This implies that $(v_0, \dots, v_i, v_i, \dots, v_n) \in [\phi \circ p(\widehat{\omega})]$, that is, $(v_0, \dots, v_i, v_i, \dots, v_n) \in OC^{\Omega^{\mathrm{ord}}}(G)$. □

Similarly, one can prove the following:

Proposition 4.11. Suppose that given any $\omega = (V_\omega, E_\omega) \in \Omega^{\mathrm{ord}}$, with $V_\omega = \{z_0, \dots, z_{N_\omega}\}$ and given any $i \leq n \leq N_\omega$, there exists $\widehat{\omega} \in \Omega^{\mathrm{ord}}$ and a graph map $p\colon \widehat{\omega} \to \omega$ such that $(z_0, \dots, \widehat{z_i}, \dots, z_n) \in [p(\widehat{\omega})]$. Then $OC^{\Omega^{\mathrm{ord}}}(G)$ satisfies property (P2), for any $G \in \mathcal{G}$. □

The advantage of dealing with ordered simplicial complexes is that we can define a homology theory for it just like in the (standard) simplicial complex case.

Definition 4.12. Let K be a SoS of V satisfying (P2) (that is, an ordered simplicial complex). We can define its *ordered simplicial homology* as usual (see [11]), since in this case the boundary operator is well defined. Explicitly, let $K = \cup_p K_p$, where

$$K_p = \{\alpha \in K \mid \alpha \text{ has length } p + 1\}.$$

Define $A_p^K := \langle K_p \rangle$, the vector space over the field $\mathbb{K}$ generated by K_p. Define the boundary operator $\partial\colon A_p^K \to A_{p-1}^K$, the cycles and the boundaries as usual. Denote by $A_\bullet^K$ the chain complex $\left\{A_p^K \xrightarrow{\partial} A_{p-1}^K\right\}_{p \geq 1}$, and denote its homology by $OH_p(K)$, where the prefix O is to remember that we are dealing with an ordered simplicial complex.

Even when K does not satisfy (P2), we can define a homology for it.

Definition 4.13. Let K be a SoS of V, with $K = \cup_p K_p$ as in Definition 4.12. For every $p \geq 0$, consider $A_p^K = \langle K_p \rangle$ the vector space over the field $\mathbb{K}$ generated by K_p, as before. We would like to define a chain complex

$$\cdots \longrightarrow A_p^K \xrightarrow{\partial} A_{p-1}^K \longrightarrow \cdots$$

with the boundary operator ∂ as the usual one.

However, it can be the case that for some $a \in A_p^K$ it happens $\partial(a) \notin A_{p-1}^K$, since K lacks property (P2). To overcome this, consider the following subspace of A_p^K:

$$\Theta_p^K = \{a \in A_p^K \mid \partial(a) \in A_{p-1}^K\}.$$

Then, clearly,

$$\Theta_\bullet^K: \quad \cdots \longrightarrow \Theta_p^K \xrightarrow{\partial} \Theta_{p-1}^K \longrightarrow \cdots$$

is a chain complex. Thus, we can define its ordered simplicial homology in dimension p as usual and denote it by $OH_p(K)$.

Notice that if K satisfies (P2), then $\Theta_\bullet^K = A_\bullet^K$.

Definition 4.14. Let Ω be a family of graphs. Consider

$$P^{\mathrm{ord}}(\Omega) := \{(\omega, l) \mid \omega = (V_\omega, E_\omega) \in \Omega,\ l\colon \{0, \ldots, |\omega| - 1\} \longrightarrow V_\omega \text{ is a bijection}\},$$

that is, the set of all graphs $\omega \in \Omega$ with all possible orderings. Then it is clear that, for any $G \in \mathcal{G}$, the SoS $K = OC^{P^{\mathrm{ord}}(\Omega)}(G)$ satisfies (P2) (see the proof of Theorem 4.16). One extra condition on Ω is needed to make sure that K also satisfies (P1), which is a necessary condition in Theorem 4.19:

Definition 4.15. The family of graphs Ω is said to be *repetition closed* if for any $\omega \in \Omega$ and $z \in \omega$, there exist $\widehat{\omega} \in \Omega$ and $p\colon \widehat{\omega} \longrightarrow \omega$ a surjective graph map such that $p^{-1}(z)$ has at least two points, that is, $|p^{-1}(z)| \geq 2$.

The next theorem resembles Proposition 2.32 at the homological level, when we regard $C^{\Omega^{\mathrm{ord}}}$ as an extension of $\mathfrak{F}^\Omega$ (see Remark 4.2) and $OC^{\Omega^{\mathrm{ord}}}$ as an extension of $\mathfrak{F}^{\Omega^*}$ (see Remark 4.8).

Theorem 4.16. Suppose that Ω is a repetition closed set of graphs and let $\Omega^{\mathrm{ord}} = P^{\mathrm{ord}}(\Omega)$, as in Definition 4.14. Then, $K = OC^{\Omega^{\mathrm{ord}}}(G)$ satisfies (P1) and (P2) for all $G \in \mathcal{G}$. Moreover, for any $p \geq 0$,

$$H_p^\Omega(G) \simeq OH_p^{\Omega^{\mathrm{ord}}}(G).$$

Proof. First, let's prove the following claim.

Claim 4.16.1. K satisfies properties (P1) and (P2).

Proof. Let $\omega = (V_\omega, E_\omega) \in \Omega^{\mathrm{ord}}$, with ordered vertex set $\{z_0, \ldots, z_{N_\omega}\}$.

Choose $n \leq N_\omega$ and $k \in \{0, \ldots, n\}$. Since Ω is repetition closed, there are $\widehat{\omega} \in \Omega$ and a surjective graph map $p\colon \widehat{\omega} \longrightarrow \omega$ with $|p^{-1}(z_k)| \geq 2$. Consider the following order on the vertex set of $\widehat{\omega}$, say $\{\widehat{z}_0, \ldots, \widehat{z}_M\}$, such that $p(\widehat{z}_i) = z_i$ for $i = 0, \ldots, k$, and $p(\widehat{z}_i) = z_{i-1}$, $i = k, \ldots, n$, as shown in the following diagram.

$$\begin{tikzcd}
\widehat{z}_0 \arrow[d] & \widehat{z}_1 \arrow[d] & \cdots & \widehat{z}_k \arrow[d] & \widehat{z}_{k+1} \arrow[dl] & \widehat{z}_{k+2} \arrow[dl] & \cdots \arrow[dl] \\
z_0 & z_1 & \cdots & z_k & z_{k+1} & z_{k+2} & \cdots
\end{tikzcd}$$

Then Ω^{ord} satisfies the condition of Proposition 4.10, which implies that $OC^{\Omega^{\mathrm{ord}}}(G)$ satisfies (P1) for any $G \in \mathcal{G}$.

For property (P2), consider $\omega = (V_\omega, E_\omega) \in \Omega^{\mathrm{ord}}$, and let n, k and N_ω be as above. Consider ω' differing from ω just by the ordering of the vertex set, which is equal to $\{z_0, \ldots, z_{k-1}, z_{k+1}, \ldots, z_{N_\omega}, z_k\}$, that is: we just put z_k as the last (biggest, in the order) element of ω'. Then for the identity map $i\colon \omega' \to \omega$ we have $(z_0, \ldots, z_{k-1}, z_{k+1}, \ldots, z_n) \in [i(\omega')]$. This is exactly the condition of Proposition 4.11. ■

Now let G be any graph. Notice that

$$(v_0, \ldots, v_n) \in OC^{\Omega^{\mathrm{ord}}}(G) \Leftrightarrow [v_0, \ldots, v_n] \in C^{\Omega}(G)$$

(and also that the elements of $OC^{\Omega^{\mathrm{ord}}}(G)$ can have repetitions like $(v_0, v_0, v_0, v_1, \ldots)$). So, the elements on the right hand side of the equivalence above are sequences without repetitions.

Indeed, implication $(\Rightarrow)$ is clear.

For the converse, let $(v_0, \ldots, v_n)$ such that $[v_0, \ldots, v_n] \in C^{\Omega}(G)$. Then there are $\omega \in \Omega$ and a graph map $\phi\colon \omega \to G$ with $\{v_0, \ldots, v_n\} \subseteq \mathrm{Im}(\phi)$. Using (maybe several times) the fact that the family Ω is repetition closed, we can find $\widehat{\omega} \in \Omega^{\mathrm{ord}}$ and $\psi\colon \widehat{\omega} \to G$ such that $(v_0, \ldots, v_n) \in [\psi(\widehat{\omega})]$.

By [11, Theorem 3.5.4], the map $j\colon OC^{\Omega^{\mathrm{ord}}}(G) \to C^{\Omega}(G)$ given by $(v_0, \ldots, v_p) \mapsto [v_0, \ldots, v_p]$ if all v_i are distinct, and 0 otherwise, induces an isomorphism at the homology level, for all $p \geq 0$:

$$j\colon H_p^{\Omega}(G) \simeq OH_p^{\Omega^{\mathrm{ord}}}(G). \qquad \square$$

Definition 4.17 (Adapted from [83])**.** Let K and $\widetilde{K}$ be two SoS of V and $\widetilde{V}$, respectively. We say that $\widetilde{K}$ is an *expansion* of K if there are functions $\pi\colon \widetilde{V} \to V$ and $i\colon V \to \widetilde{V}$ such that $\pi \circ i$ is the identity of V and

$$(v_0, \ldots, v_n) \in \widetilde{K} \Leftrightarrow (\pi(v_0), \ldots, \pi(v_n)) \in K.$$

Now given two filtrations $\mathcal{K} = \{K_\epsilon\}_{\epsilon\in\mathbb{R}}$ and $\widetilde{\mathcal{K}} = \{\widetilde{K}_\epsilon\}_{\epsilon\in\mathbb{R}}$ of SoS, we say that $\widetilde{\mathcal{K}}$ is an expansion of $\mathcal{K}$ if $\widetilde{K}_\epsilon$ is an expansion of K_ϵ, for every $\epsilon \in \mathbb{R}$.

Lemma 4.18 (Adapted from [83])**.** Let $\mathcal{K} = \{K_\epsilon\}_{\epsilon\in\mathbb{R}}$ and $\widetilde{\mathcal{K}} = \{\widetilde{K}_\epsilon\}_{\epsilon\in\mathbb{R}}$ be two filtrations of SoS of V and $\widetilde{V}$, respectively, satisfying (P1), such that $\widetilde{\mathcal{K}}$ is an expansion of $\mathcal{K}$.

Then their persistent diagrams in dimension $p \geq 0$ obtained from the persistent vector spaces $\{OH_p(K_\epsilon)\}_{\epsilon\in\mathbb{R}}$ and $\{OH_p(\widetilde{K}_\epsilon)\}_{\epsilon\in\mathbb{R}}$, namely $\mathrm{dgm}_p(\mathcal{K})$ and $\mathrm{dgm}_p(\widetilde{\mathcal{K}})$, are the same.

Proof. To ease the notation, set $A_p^\epsilon = A_p^{K_\epsilon}$, $\widetilde{A}_p^\epsilon = A_p^{\widetilde{K}_\epsilon}$, $\Theta_p^\epsilon = \Theta_p^{K_\epsilon}$ and $\widetilde{\Theta}_p^\epsilon = \Theta_p^{\widetilde{K}_\epsilon}$ (see Definition 4.13). Also, identify each $v \in V$ with its image $i(v) \in \widetilde{V}$, where $i\colon V \to \widetilde{V}$ is the map such that $\pi \circ i$ is the identity of V (see Definition 4.17), so that $\pi \circ \pi = \pi$.

Consider the induced maps $\pi_\#\colon \widetilde{\Theta}^\epsilon_p \longrightarrow \Theta^\epsilon_p$ given by $\pi_\#(v_0,\dots,v_p) = (\pi(v_0),\dots,\pi(v_p))$, and $i_\#\colon \Theta^\epsilon_p \longrightarrow \widetilde{\Theta}^\epsilon_p$ given by $i_\#(v_0,\dots,v_p) = (v_0,\dots,v_p)$. It is easy to see that these maps are well defined chain maps and $\pi_\# \circ i_\#$ is the identity on Θ^ϵ_p, since $\pi \circ i$ is the identity map.

By the Definition 4.17, we have

$$\begin{aligned}(v_0,\dots,v_n) \in \widetilde{K}_\epsilon &\Leftrightarrow (v_0,\dots,v_i,v_i,\dots,v_n) \in \widetilde{K}_\epsilon\\ &\Leftrightarrow (\pi(v_0),\dots,\pi(v_i),\pi(v_i),\dots,\pi(v_n)) \in K_\epsilon\\ &\Leftrightarrow (\pi(v_0),\dots,\pi(v_i),\pi(\pi(v_i)),\dots,\pi(\pi(v_n)) \in K_\epsilon\\ &\Leftrightarrow (v_0,\dots,v_i,\pi(v_i),\dots,\pi(v_p)) \in \widetilde{K}_\epsilon.\end{aligned}$$

For each p, define the prism operator $D^p\colon \widetilde{\Theta}^\epsilon_p \longrightarrow \widetilde{A}^\epsilon_{p+1}$ by

$$D^p(v_0,\dots,v_p) = \sum_{j=0}^{p}(-1)^j(v_0,\dots,v_j,\pi(v_j),\dots,\pi(v_p)),$$

and extend it linearly. We drop the superscript p when the context is clear.

Given $v = (v_0,\dots,v_p)$ and $j = 0,\dots,p$, write

$$D_j(v) = (-1)^j(v_0,\dots,v_j,\pi(v_j),\dots,\pi(v_p)).$$

Then $D(v) = \sum_j(-1)^jD_j(v)$. Similarly, set $\partial_k(v) = (-1)^k(v_0,\dots,\widehat{v}_k,\dots,v_p)$ so $\partial(a) = \sum_k(-1)^k\partial_k(a)$.

Claim 4.18.1. Let $v \in \widetilde{\Theta}^\epsilon_p$. Then

$$\partial \circ D(v) + D \circ \partial(v) = \pi_\# \circ i_\#(v) - v.$$

Proof. First, observe that

$$\begin{aligned}\partial \circ D(v) &= \sum_{k=0}^{p+1}\sum_{j=0}^{p}(-1)^{j+k}\partial_k \circ D_j(v)\\ &= \sum_{\substack{j\le k-2\\ 2\le k\le p+1}}(-1)^{j+k}\partial_k \circ D_j(v) + \sum_{\substack{j=k-1,k\\ 1\le k\le p}}(-1)^{j+k}\partial_k \circ D_j(v)\\ &\quad+ \sum_{\substack{j\ge k+1\\ 0\le k\le p-1}}(-1)^{j+k}\partial_k \circ D_j(v) + \sum_{\substack{j=0\\ k=0}}(-1)^{j+k}\partial_k \circ D_j(v)\\ &\quad+ \sum_{\substack{j=p\\ k=p+1}}(-1)^{j+k}\partial_k \circ D_j(v).\end{aligned} \tag{4.1}$$

The following is a mere checking of equalities:

$$\begin{aligned}j \le k-2 &\Rightarrow (-1)^{j+k}\partial_k \circ D_j(v) = (-1)^{j+k}D_j \circ \partial_{k-1};\\ j = k \ge 1 &\Rightarrow (-1)^{j+k}\partial_k \circ D_j(v) = (-1)^{j+k}\partial_k \circ D_{j-1}(v);\\ j \ge k+1 &\Rightarrow (-1)^{j+k}\partial_k \circ D_j(v) = (-1)^{j+k}D_{j-1} \circ \partial_k(v);\\ j = k = 0 &\Rightarrow (-1)^{j+k}\partial_k \circ D_j(v) = i_\# \circ \pi_\#(v);\\ j = p, k = p+1 &\Rightarrow (-1)^{j+k}\partial_k \circ D_j(v) = -v.\end{aligned}$$

For example, if $j = 0$ and $k = 2$, consider $v = (v_0, v_1, v_2)$. Then

$$D_j(v) = (v_0, \pi(v_0), \pi(v_1), \pi(v_2)) \Rightarrow \partial_k(D_j(v)) = (v_0, \pi(v_0), \pi(v_2))$$

while

$$\partial_{k-1}(v) = (v_0, v_2) \Rightarrow D_j(\partial_{k-1}(v)) = (v_0, \pi(v_0), \pi(v_2)).$$

The other cases are checked similarly.

Notice that

$$\sum_{\substack{j=k\\k>0}} (-1)^{j+k}\partial_k \circ D_j(v) = \sum_{\substack{j=k\\k>0}} (-1)^{j+k}\partial_k \circ D_{j-1}(v) = -\sum_{\substack{j=k-1\\k>0}} (-1)^{j+k}\partial_k \circ D_j(v),$$

which implies

$$\sum_{\substack{j=k-1,k\\k>0}} (-1)^{j+k}\partial_k \circ D_j(v) = 0.$$

Then, equation (4.1) can be rewritten as

$$\partial \circ D(v) = \sum_{j\le k-2} (-1)^{j+k} D_j \circ \partial_{k-1}(v) + \sum_{j\ge k+1} (-1)^{j+k} D_{j-1} \circ \partial_k(v) + i_\# \circ \pi_\#(v) - v.$$

Setting $l = k-1$ and $m = j$, we obtain

$$\sum_{j\le k-2} (-1)^{j+k} D_j \circ \partial_{k-1}(v) = \sum_{k=2}^{p+1}\sum_{j=0}^{k-2} (-1)^{j+k} D_j \circ \partial_{k-1}(v) = \sum_{l=1}^{p}\sum_{m=0}^{l-1} (-1)^{m+l+1} D_m \circ \partial_l(v).$$

Similarly, setting $l = k$ and $m = j-1$, we obtain

$$\sum_{j\ge k+1} (-1)^{j+k} D_{j-1} \circ \partial_k(v) = \sum_{m=0}^{p-1}\sum_{l=0}^{m} (-1)^{l+m+1} D_m \circ \partial_l(v).$$

Thus, (4.1) is equal to

$$\begin{aligned}
\partial \circ D(v) &= \sum_{l=1}^{p}\sum_{m=0}^{l-1} (-1)^{m+l+1} D_m \circ \partial_l(v) + \sum_{m=0}^{p-1}\sum_{l=0}^{m} (-1)^{l+m+1} D_m \circ \partial_l(v) + i_\# \circ \pi_\#(v) - v\\
&= \sum_{m=0}^{p-1}\sum_{l=0}^{p} -(-1)^{l+m} D_m \circ \partial_l(v) + i_\# \circ \pi_\#(v) - v\\
&= -D \circ \partial(v) + i_\# \circ \pi_\#(v) - v.
\end{aligned}$$

This proves Claim 4.18.1. ∎

Claim 4.18.2. The image of D^p is contained in $\widetilde{\Theta}^{\epsilon}_{p+1}$.

Proof. Indeed, for $p = 0$ it is clear. Then, given $v \in \widetilde{\Theta}^{\epsilon}_1$,

$$\partial \circ D^1(v) = -D^0 \circ \partial(v) + \pi_\# \circ i_\#(v) - v \in \widetilde{\Theta}^{\epsilon}_0 \subseteq \widetilde{A}^{\epsilon}_0.$$

By the definition of $\widetilde{\Theta}^{\epsilon}_2$, we have $D^1(v) \in \widetilde{\Theta}^{\epsilon}_2$. The same argument can be done to D^p. ∎

We have shown that D is a *chain homotopy* between $\pi_{\#} \circ i_{\#}$ and the identity of $\widetilde{\Theta}^{\epsilon}_{\bullet}$ (see [9] for more details on homological algebra), which implies that its induced in homology $\pi_* \circ i_* \colon OH_p(\widetilde{\Theta}^{\epsilon}_{\bullet}) \to OH_p(\widetilde{\Theta}^{\epsilon}_{\bullet})$ is the identity map, and the same for $i_* \circ \pi_* \colon OH_p(\Theta^{\epsilon}_{\bullet}) \to OH_p(\Theta^{\epsilon}_{\bullet})$. Thus, π_* and i_* are isomorphisms, inverses of each other.

Since the diagram below

$$\begin{tikzcd} \Theta^{\epsilon}_p \arrow[r, hook] \arrow[d, "i_{\#}"'] & \Theta^{\epsilon'}_p \arrow[d, "i_{\#}"] \\ \widetilde{\Theta}^{\epsilon}_p \arrow[r, hook] & \widetilde{\Theta}^{\epsilon'}_p \end{tikzcd}$$

is clearly commutative for every $\epsilon \le \epsilon'$, the following diagram is also commutative:

$$\begin{tikzcd} OH_p(\Theta^{\epsilon}_{\bullet}) \arrow[r] \arrow[d, "i_*"'] & OH_p(\Theta^{\epsilon'}_{\bullet}) \arrow[d, "i_*"] \\ OH_p(\widetilde{\Theta}^{\epsilon}_{\bullet}) \arrow[r] & OH_p(\widetilde{\Theta}^{\epsilon'}_{\bullet}). \end{tikzcd}$$

This proves that $\mathrm{dgm}_p(\mathcal{K})$ and $\mathrm{dgm}_p(\widetilde{\mathcal{K}})$ are equal. □

The following theorem is an adaptation of [83, Theorem 21]:

Theorem 4.19 (Stability)**.** Let Ω^{ord} be a set of ordered graphs such that $OC^{\Omega^{\mathrm{ord}}}(G)$ satisfies (P1), for every $G \in \mathcal{G}$. Let $\mathbf{X} = (X, w_X)$, $\mathbf{Y} = (Y, w_Y) \in \mathcal{N}_{\mathrm{ext}}$ and $p \ge 0$. Let $\mathrm{dgm}_p^{\Omega^{\mathrm{ord}}}(\mathbf{X})$ and $\mathrm{dgm}_p^{\Omega^{\mathrm{ord}}}(\mathbf{Y})$ be the p-dimensional persistent diagrams obtained from the filtrations $\mathcal{K} = \{K_\epsilon = OC^{\Omega^{\mathrm{ord}}}(\mathbf{X}_\epsilon)\}_{\epsilon\in\mathbb{R}}$ and $\mathcal{L} = \{L_\epsilon = OC^{\Omega^{\mathrm{ord}}}(\mathbf{Y}_\epsilon)\}_{\epsilon\in\mathbb{R}}$. Then

$$d_b\Big(\mathrm{dgm}_p^{\Omega^{\mathrm{ord}}}(\mathbf{X}), \mathrm{dgm}_p^{\Omega^{\mathrm{ord}}}(\mathbf{Y})\Big) \le 2 d_{\mathcal{N}_{\mathrm{ext}}}(\mathbf{X}, \mathbf{Y}),$$

where d_b is the bottleneck distance and $d_{\mathcal{N}_{\mathrm{ext}}}$ is the extended network distance. In other words, the map $\mathrm{dgm}_p^{\Omega^{\mathrm{ord}}} \colon \mathcal{N}_{\mathrm{ext}} \to \mathbf{Pers}$ is 2-Lipschitz.

Proof. If $d_{\mathcal{N}_{\mathrm{ext}}}(\mathbf{X}, \mathbf{Y}) = +\infty$, there is nothing to prove.

Let $M \subseteq X \times Y$ be a correspondence such that $2\eta = d_{\mathcal{N}_{\mathrm{ext}}}(\mathbf{X}, \mathbf{Y})$, where $\eta = \mathrm{dis}(M; \mathbf{X}, \mathbf{Y})$ is the distortion of M with respect to $\mathbf{X}$ and $\mathbf{Y}$.

Define the extended networks $\mathbf{M}^{\mathbf{X}} = (M, f)$ where $f(x, y, x', y') = w_X(x, x')$, and $\mathbf{M}^{\mathbf{Y}} = (M, g)$, where $g(x, y, x', y') = w_Y(y, y')$.

Consider $\widetilde{\mathcal{K}} = \{\widetilde{K}_\epsilon = OC^{\Omega^{\mathrm{ord}}}(\mathbf{M}^{\mathbf{X}}_\epsilon)\}_{\epsilon\in\mathbb{R}}$ and $\widetilde{\mathcal{L}} = \{\widetilde{L}_\epsilon = OC^{\Omega^{\mathrm{ord}}}(\mathbf{M}^{\mathbf{Y}}_\epsilon)\}_{\epsilon\in\mathbb{R}}$.

Claim 4.19.1. $\widetilde{\mathcal{K}}$ is an expansion of $\mathcal{K}$ and $\widetilde{\mathcal{L}}$ is an expansion of $\mathcal{L}$.

Proof. Let $\pi_X \colon M \to X$ be the projection, and $i_X \colon X \to M$ be the inclusion $i(x) = (x, y)$, where $y \in Y$ is chosen such that $(x, y) \in M$. Then clearly $i \circ \pi$ is the identity of X. We need to prove that for all $\epsilon \in \mathbb{R}$, we have

$$(v_0, \ldots, v_n) \in \widetilde{K}_\epsilon \Leftrightarrow (\pi(v_0), \ldots, \pi(v_n)) \in K_\epsilon.$$

To prove $(\Rightarrow)$, suppose there is $\phi \colon \omega \to \mathbf{M}^{\mathbf{X}}_\epsilon$ with $\omega \in \Omega^{\mathrm{ord}}$ and $(v_0, \ldots, v_n) \in [\phi(\omega)]$. Then $(\pi(v_0), \ldots, \pi(v_n)) \in [\pi \circ \phi(\omega)]$, since π is clearly a graph map. Thus, $(\pi(v_0), \ldots, \pi(v_n)) \in K_\epsilon$.

For the converse, suppose $(x_0, \ldots, x_n) \in K_\epsilon$. By definition, there are $\omega = (V_\omega, E_\omega) \in \Omega^{\mathrm{ord}}$ with ordered vertex set $V_\omega = \{z_0, \ldots, z_{N_\omega}\}$ and a graph map $\phi\colon \omega \longrightarrow \mathbf{X}_\epsilon$ with $(x_0, \ldots, x_n) \in [\phi(\omega)]$. For each i, choose $v_i \in M$ such that $\pi(v_i) = x_i$. Define $\phi'\colon \omega \longrightarrow \mathbf{M}_\epsilon^{\mathbf{X}}$ by $\phi'(z_i) = v_i$. This is a graph map: indeed, notice that

$$z_i \xrightarrow{\omega} z_j \Rightarrow \phi(z_i) = x_i \xrightarrow[=]{\mathbf{X}_\epsilon} x_j = \phi(z_j) \Rightarrow \phi'(z_i) = v_i \xrightarrow[=]{\mathbf{M}_\epsilon^{\mathbf{X}}} v_j = \phi'(z_j),$$

where the first implication holds since ϕ is a graph map, and the second one holds since

$$x_i \xrightarrow[=]{\mathbf{X}_\epsilon} x_j \Leftrightarrow w_X(x_i, x_j) \leq \epsilon \Leftrightarrow f(v_i, v_j) \leq \epsilon \Leftrightarrow v_i \xrightarrow[=]{\mathbf{M}_\epsilon^{\mathbf{X}}} v_j.$$

The proof for $\mathcal{L}$ and $\widetilde{\mathcal{L}}$ is completely analogous. ■

Claim 4.19.2. There are inclusions $\iota\colon \widetilde{K}_\epsilon \hookrightarrow \widetilde{L}_{\epsilon+\eta}$ and $\widetilde{L}_\epsilon \hookrightarrow \widetilde{K}_{\epsilon+\eta}$.

Proof. We will just show the existence of ι, since the other one is similar.

Let $(v_0, \ldots, v_n) \in \widetilde{K}_\epsilon$, with $v_i = (x_i, y_i)$. Then there are $\omega \in \Omega^{\mathrm{ord}}$ with ordered vertex set $V_\omega = \{z_0, \ldots, z_{N_\omega}\}$ and a graph map $\phi\colon \omega \longrightarrow \mathbf{M}_\epsilon^{\mathbf{X}}$, with $(v_0, \ldots, v_n) \in [\phi(\omega)]$, that is, $\phi(z_i) = v_i$. Since ϕ is a graph map,

$$\begin{aligned} z_i \xrightarrow{\omega} z_j \Rightarrow \phi(z_i) = v_i \xrightarrow[=]{\mathbf{M}_\epsilon^{\mathbf{X}}} v_j = \phi(z_j) \Leftrightarrow f(v_i, v_j) = w_X(x_i, x_j) \leq \epsilon \\ \Rightarrow w_Y(y_i, y_j) = g(v_i, v_j) \leq \epsilon + \eta \Leftrightarrow \phi(z_i) = v_i \xrightarrow[=]{\mathbf{M}_\epsilon^{\mathbf{Y}}} v_j = \phi(z_j). \end{aligned}$$

This proves that $(v_0, \ldots, v_n) \in \widetilde{L}_{\epsilon+\eta}$. ■

Thus, ι induces an inclusion $\Theta_p^{\widetilde{K}_\epsilon} \to \Theta_p^{\widetilde{L}_{\epsilon+\eta}}$, and the following diagram is commutative

$$\begin{tikzcd} \Theta_p^{\widetilde{K}_\epsilon} \arrow[rr, hook] \arrow[dr, hook] & & \Theta_p^{\widetilde{K}_{\epsilon+2\eta}} \\ & \Theta_p^{\widetilde{L}_{\epsilon+\eta}} \arrow[ur, hook] & \end{tikzcd}$$

which in turn gives us the commutativity at the homological level.

By Lemma 4.18, the persistent diagram obtained from $\{\Theta^{\widetilde{K}_\epsilon}\}_{\epsilon\in\mathbb{R}}$ and $\{\Theta^{\widetilde{L}_\epsilon}\}_{\epsilon\in\mathbb{R}}$ by applying the ordered homology functor are the same as $\mathrm{dgm}_p^{\Omega^{\mathrm{ord}}}(\mathbf{X})$ and $\mathrm{dgm}_p^{\Omega^{\mathrm{ord}}}(\mathbf{Y})$. We can conclude that $\{OH_p(\Theta^{K_\epsilon})\}_{\epsilon\in\mathbb{R}}$ and $\{OH_p(\Theta^{L_\epsilon})\}_{\epsilon\in\mathbb{R}}$ are 2η-interleaved, which implies, by Theorem 1.22, that

$$d_b\Big(\mathrm{dgm}_p^{\Omega^{\mathrm{ord}}}(\mathbf{X}), \mathrm{dgm}_p^{\Omega^{\mathrm{ord}}}(\mathbf{Y})\Big) \leq 2\eta = 2\,\mathrm{dis}(M) = d_{\mathcal{N}_{\mathrm{ext}}}(\mathbf{X}, \mathbf{Y}). \qquad \square$$

Corollary 4.20 (Stability for non-ordered complexes)**.** Suppose that Ω is a repetition closed set of graphs. Let $\mathbf{X}, \mathbf{Y} \in \mathcal{N}_{\mathrm{ext}}$ and $p \geq 0$. Let $\mathrm{dgm}_p^\Omega(\mathbf{X})$ and $\mathrm{dgm}_p^\Omega(\mathbf{Y})$ be the p-dimensional persistent diagrams obtained from the filtrations $\{C^\Omega(\mathbf{X}_\epsilon)\}_{\epsilon\in\mathbb{R}}$ and $\{C^\Omega(\mathbf{Y}_\epsilon)\}_{\epsilon\in\mathbb{R}}$. Then

$$d_b\Big(\mathrm{dgm}_p^\Omega(\mathbf{X}), \mathrm{dgm}_p^\Omega(\mathbf{Y})\Big) \leq 2 d_{\mathcal{N}_{\mathrm{ext}}}(\mathbf{X}, \mathbf{Y}).$$

Proof. Let $\Omega^{\text{ord}} = P^{\text{ord}}(\Omega)$. By the proof of Theorem 4.16 and [11, Theorem 3.5.4], the map $j\colon OH_p^{\Omega^{\text{ord}}}(\mathbf{X}_\epsilon) \to H_p^{\Omega}(\mathbf{X}_\epsilon)$ induced by

$$(v_0, \dots, v_p) \mapsto \begin{cases} [v_0, \dots, v_p], & \text{if all } v_i \text{ are distinct,} \\ 0, & \text{otherwise,} \end{cases}$$

is an isomorphism at the homology level, which commutes the diagram

$$\begin{tikzcd} H_p^{\Omega}(\mathbf{X}_\epsilon) \arrow[r] \arrow[d, "j"'] & H_p^{\Omega}(\mathbf{X}_{\epsilon'}) \arrow[d, "j"] \\ OH_p^{\Omega^{\text{ord}}}(\mathbf{X}_\epsilon) \arrow[r] & OH_p^{\Omega^{\text{ord}}}(\mathbf{X}_{\epsilon'}) \end{tikzcd}$$

for all $\epsilon \le \epsilon'$.

Then, the persistent module of $\{H_p^{\Omega}(\mathbf{X}_\epsilon)\}_{\epsilon\in\mathbb{R}}$ is the same as the persistent module of $\{OH_p^{\Omega^{\text{ord}}}(\mathbf{X}_\epsilon)\}_{\epsilon\in\mathbb{R}}$. The analogue is true for $\{H_p^{\Omega}(\mathbf{Y}_\epsilon)\}_{\epsilon\in\mathbb{R}}$ and $\{OH_p^{\Omega^{\text{ord}}}(\mathbf{Y}_\epsilon)\}_{\epsilon\in\mathbb{R}}$. By the Theorem 4.19, the result follows. □

By Corollary 4.20, the maps $\mathrm{dgm}_p^{\text{Rips}}, \mathrm{dgm}_p^{\text{Dow}}\colon \mathcal{N} \to \mathbf{Pers}$ are stable, i.e. 2-Lipschitz, for any $p \ge 0$. These results were obtained in [18, Propositions 12 and 15] in a different manner.

Remark 4.21 (About the necessity of property (P1) in the stability)**.** We don't know if property (P1) is *necessary* to obtain Theorem 4.19. We also don't know if it is necessary for a family Ω to be repetition closed in order to satisfy Corollary 4.20. Any *finite* family Ω is not repetition closed: indeed, given $\omega_1 \in \Omega$ and $z \in \omega_1$, by Definition 4.15 there are $\omega_2 \in \Omega$ and a surjective map $p\colon \omega_2 \to \omega_1$ with $|p^{-1}(z)| \ge 2$. Thus, $|\omega_2| \ge |\omega_1| + 1$. Using the same argument for ω_2, there is $\omega_3 \in \Omega$ with $|\omega_3| \ge |\omega_2| + 1 \ge |\omega_1| + 2$, and so on.

Moreover, we have the following proposition:

Proposition 4.22. Let Ω be a finite family of graphs, with $p = \max\{|\omega|,\ \omega \in \Omega\}$. Then the map $\mathrm{dgm}_p^{\Omega}\colon \mathcal{N}_{\text{ext}} \to \mathbf{Pers}$ is not Lipschitz.

Proof. Let $\mathbf{X} = (X, w_X)$ be a network where X is a set with $p+1$ points and $w_X(x, x') = 0$, $\forall x, x' \in X$. Thus,

$$\mathbf{X}_\epsilon = \begin{cases} \emptyset, & \epsilon < 0; \\ K_{p+1}, & \epsilon \ge 0, \end{cases}$$

which implies

$$C^{\Omega}(\mathbf{X}_\epsilon) = \begin{cases} \emptyset, & \epsilon < 0; \\ \widetilde{\Delta}_p, & \epsilon \ge 0, \end{cases}$$

where $\widetilde{\Delta}_p$ is the set of all proper faces of the p-simplex Δ_p. Thus, $H_{p-1}(\widetilde{\Delta}_p) \simeq \mathbb{K}$. Let $\mathbf{Y} = (Y, w_Y)$ with $Y = \{y\}$ and $w_Y(y, y) = 0$. Then we have $\mathrm{dgm}_{p-1}^{\Omega}(\mathbf{X}) = [0, +\infty)$ and $\mathrm{dgm}_{p-1}^{\Omega}(\mathbf{Y}) = 0$, which implies

$$+\infty = d_b\Big(\mathrm{dgm}_{p-1}^{\Omega}(\mathbf{X}), \mathrm{dgm}_{p-1}^{\Omega}(\mathbf{Y})\Big) > d_{\mathcal{N}}(\mathbf{X}, \mathbf{Y}) = 0. \qquad □$$

Example 4.23. As an example for Proposition 4.22, consider $\Omega = \{K_2\}$. Then, $\mathfrak{F}^\Omega = \mathfrak{F}^{\mathrm{ls}}$. Let $\mathbf{X} = (X, w_X)$ with $X = \{x_1, x_2, x_3\}$ and $w_X(x, x') = 0$, $\forall x, x' \in X$. Then, $C^\Omega(\mathbf{X}_\epsilon) = \{[x_1, x_2], [x_2, x_3], [x_1, x_3]\}$ for $\epsilon \geq 0$ and $\emptyset$ otherwise. Thus, $\mathrm{dgm}_1^\Omega(\mathbf{X}) = [0, +\infty)$.

Under certain conditions, some families Ω generate trivial persistent diagrams. One such condition is the following:

Proposition 4.24. Let Ω be a wedge covered family of graphs (which is equivalent to say that $\mathfrak{F}^\Omega$ is a clustering functor, by Theorem 2.45). For any $\mathbf{X} = (X, w_X) \in \mathcal{N}_{\mathrm{ext}}$ and any $\epsilon \in \mathbb{R}$, $C^\Omega(\mathbf{X}_\epsilon)$ is empty or is a disjoint union of simplexes. Thus, $\mathrm{dgm}_p^\Omega(\mathbf{X})$ is trivial, for any $p > 0$.

Proof. Call $G = \mathbf{X}_\epsilon$. Let $G_1, \ldots, G_n$ be complete graphs with vertex set $V_1, \ldots, V_n$, respectively, such that

$$\mathfrak{F}^\Omega(G) = \bigsqcup_{i=1}^{n} G_i.$$

Let's prove that $C^\Omega(G) = \cup_{i=1}^n \sigma_i$, where σ_i is the simplex over the set V_i.

First note that if $v \in V_i$ and $v' \in V_j$ with $i \neq j$, then there is no graph map $\phi\colon \omega \to (G, v, v')$, for any $\omega \in \Omega$ (otherwise, we would have $v \leftrightarrow v'$ in $\mathfrak{F}^\Omega(G)$).

Now write $V_1 = \{v_1, \ldots, v_k\}$ and let's prove that there is a graph map $\phi : \omega \to G$ such that $V_1 \subseteq \mathrm{Im}(\phi)$. The argument is analogue for $V_2, \ldots, V_n$.

By definition, there are graphs $\omega_1, \omega_2 \in \Omega$ and graph maps $\phi_1\colon (\omega_1, z_1, \widehat{z_1}) \to (G, v_1, v_2)$ and $\phi_2\colon (\omega_2, z_2, \widehat{z_2}) \to (G, v_2, v_3)$. Since Ω is wedge covered, there are $\widetilde{\omega} \in \Omega$ and a surjective graph map $\widetilde{\phi}\colon \widetilde{\omega} \to (\omega_1, z_2) \vee (\omega_2, z_2')$. The maps $\widetilde{\phi}, \phi_1$ and ϕ_2 define a graph map (see the proof of Theorem 2.45) from $\widetilde{\omega}$ to G such that $\{v_1, v_2, v_3\}$ is in its image. In resume: given graph maps that contain $\{v_1, v_2\}$ and $\{v_2, v_3\}$ in its image, it is possible to define a graph map that contains $\{v_1, v_2, v_3\}$ in its image, and all these maps have elements of Ω as the domain.

Since there are, also, ω_4 and a graph map $\phi_4\colon \omega_4 \to (G, v_3, v_4)$, with the same argument as above we find a graph map from an element of Ω to G that contains $\{v_1, \ldots, v_4\}$. Continuing this process we find the desired ω and the graph map $\phi\colon \omega \to G$ such that $V_1 \subseteq \mathrm{Im}(\phi)$. This finishes the proof. □

4.2 Clique homology

Definition 4.25 ([80])**.** Let $G = (V, E)$ be a graph. Consider the ordered simplicial complex $K^{\mathrm{dc}}(G) = \cup_p K_p^{\mathrm{dc}}(G)$, called the *directed clique complex* of G, where

$$K_p^{\mathrm{dc}}(G) = \{(v_0, \ldots, v_p) \mid v_i \xrightarrow{G} v_j, \text{ for all } i < j\},$$

for each $p \geq 0$. Notice that $K^{\mathrm{dc}}(G)$ satisfies property (P2) of Definition 4.9. Its p-dimensional homology (as in Definition 4.12) is called the *directed clique homology of* G, denoted by $H_p^{\mathrm{dc}}(G)$. See Figure 4.3 for an example.

Remark 4.26. Definition 4.25 can be seen as a modification of Definition 4.13. Consider $\Omega^{\mathrm{ord}} = \{T_n\}_{n\geq 1}$, where T_n has ordered vertex set $\{a_0, \ldots, a_{n-1}\}$ and arrows $a_i \to a_j$ for

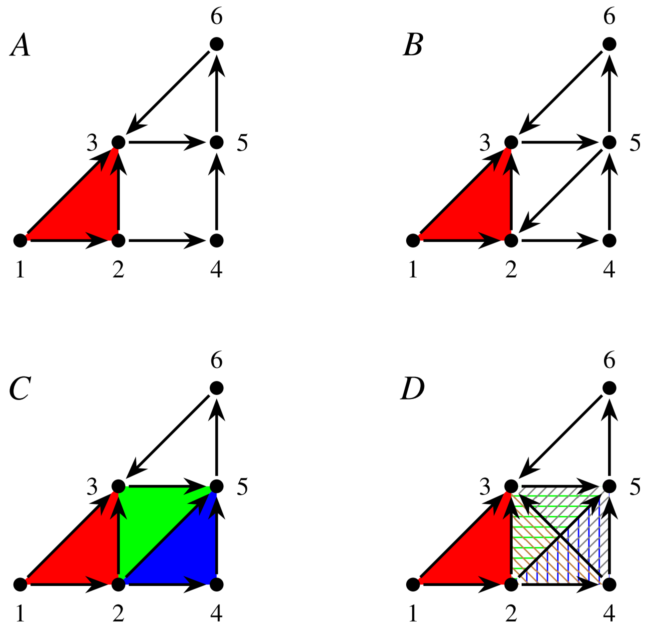


Figure 4.3: Four graphs A, B, C and D, together with its directed clique complexes. The 0-simplexes are represented as the vertices, 1-simplexes as arrows and 2-simplex are colored. The graph D is obtained from C by adding the arrow $(4, 3)$. Notice that in D we have the 3-simplex $(2, 4, 3, 5)$. Adapted from [80].

$i < j$ (equivalently, $T_n \simeq \mathfrak{F}^{\mathrm{tc}}(L_n)$). The complex $K = OC^{\Omega^{\mathrm{ord}}}(G)$ is called the *directed clique complex with repetitions*, and its homology is called *directed clique homology with repetitions*. It follows easily that

$$K_p^{\mathrm{dc}}(G) = \{(v_0, \ldots, v_p) \mid \exists \phi \colon T_{p+1} \longrightarrow G \text{ injective graph map with } \phi(a_i) = v_i,\ i = 0, \ldots, p\}.$$

Thus, the definition of $K^{\mathrm{dc}}(G)$ can be seen as restriction of Definition 4.7 which only allows *injective* graph maps $\phi \colon \omega \longrightarrow G$.

As K. Turner points out in [83], the directed clique homology *is not stable* in the sense of Theorem 4.19. To see this, let $\mathbf{X} = (X, w_X)$ be a network with $X = \{x_1, x_2\}$ and w_X the zero map. Then

$$K^{\mathrm{dc}}(\mathbf{X}_\epsilon) = \begin{cases} \emptyset, & \epsilon < 0, \\ \{(x_1), (x_2), (x_1, x_2), (x_2, x_1)\}, & \epsilon \geq 0. \end{cases}$$

For any $\epsilon \geq 0$, $\partial((x_1, x_2) + \partial(x_2, x_1)) = 0$, and this cycle is not a boundary since $K_2(\mathbf{X}_\epsilon) = \emptyset$. Thus, $H_1(\mathbf{X}_\epsilon)$ is non-trivial for any $\epsilon \geq 0$. Notice that this same cycle is a boundary when considering the directed clique homology with repetitions:

$$\begin{aligned} &\partial((x_1, x_1, x_1) + (x_1, x_2, x_1)) = \\ &\qquad (x_1, x_1) - (x_1, x_1) + (x_1, x_1) + (x_2, x_1) - (x_1, x_1) + (x_1, x_2) = (x_1, x_2) + (x_2, x_1). \end{aligned}$$

Now let $\mathbf{Y} = (Y, w_Y)$ with $Y = \{y\}$ and w_Y the zero map. Then $H_1^{\mathrm{dc}}(\mathbf{Y}_\epsilon) = 0$, for any $\epsilon \in \mathbb{R}$. Thus,

$$+\infty = d_b\big(\mathrm{dgm}_1^{\mathrm{dc}}(\mathbf{X}), \mathrm{dgm}_1^{\mathrm{dc}}(\mathbf{Y})\big) > 0 = d_\mathcal{N}(\mathbf{X}, \mathbf{Y}),$$

where $\mathrm{dgm}_1^{\mathrm{dc}}(\mathbf{X})$ and $\mathrm{dgm}_1^{\mathrm{dc}}(\mathbf{Y})$ are the 1-dimensional persistent diagrams of $\mathbf{X}$ and $\mathbf{Y}$ obtained from the directed clique homology.

On the other hand, the directed clique homology with repetition *is stable*, because Ω^{ord} satisfies Proposition 4.11.

4.3 Path homology

Definition 4.27 (Path homology)**.** Let V be a finite set, and let $p \geq 0$. Denote by $R(V^{p+1})$ the set of *regular p-paths of V*:

$$R(V^{p+1}) \coloneqq \{(v_0, \dots, v_p) \in V^{p+1} \mid v_i \neq v_{i+1},\ i = 0, \dots, p-1\}.$$

Let $I(V^{p+1}) \coloneqq V^{p+1} \setminus R(V^{p+1})$, the set of *irregular p-paths of V*. Write $\Lambda_p(V) = \langle V^{p+1} \rangle$, the vector space over the field $\mathbb{K}$ generated by V^{p+1}. Call $\mathcal{R}_p(V) = \langle R(V^{p+1}) \rangle$ and $\mathcal{I}_p(V) = \langle I(V^{p+1}) \rangle$, and notice that $\Lambda_p(V) = \mathcal{R}_p(V) \oplus \mathcal{I}_p(V)$.

Define the boundary operator $\partial \colon \Lambda_p(V) \to \Lambda_{p-1}(V)$ as usual:

$$\partial(v_0, \dots, v_p) = \sum_{i=0}^{p} (-1)^i (v_0, \dots, \widehat{v_i}, \dots, v_p),$$

where the $\widehat{v_i}$ means the exclusion of v_i from the sequence, and extend it linearly.

Define $\widetilde{\mathcal{R}}_p(V) \coloneqq \Lambda_p(V)/\mathcal{I}_p(V)$. Consider the quotient map $\pi_p \colon \Lambda_p(V) \to \widetilde{\mathcal{R}}_p(V)$. It is clear that $\mathcal{R}_p(V)$ is isomorphic to $\widetilde{\mathcal{R}}_p(V)$ via π_p.

It can be proved (see [82]) that $\partial(\mathcal{I}_p(V)) \subseteq \mathcal{I}_{p-1}(V)$, which implies that the boundary operator is well-defined in $\widetilde{\mathcal{R}}_p(V)$. Denote it by ∂^{reg}. Then the following diagram is commutative:

$$\begin{tikzcd}
\mathcal{R}_p(V) \arrow[r, "\partial"] & \Lambda_{p-1}(V) \arrow[d, "\pi_p"] \\
\widetilde{\mathcal{R}}_p(V) \arrow[u, "\pi_p^{-1}"] \arrow[r, "\partial^{\mathrm{reg}}"] & \widetilde{\mathcal{R}}_{p-1}(V).
\end{tikzcd}$$

Let Ω^{ord} be the set of ordered line graphs $L_n = (V_{L_n}, E_{L_n})$, $n \geq 1$, where the ordered vertex set V_{L_n} is $\{a_0, \dots, a_{n-1}\}$, with $a_i \xrightarrow{L_n} a_{i+1}$, $i = 0, \dots, n-2$.

Now let $G = (V, E)$ be a graph. Consider $K = OC^{\Omega^{\mathrm{ord}}}(G)$ and $K_p = OC_p^{\Omega^{\mathrm{ord}}}(G)$, which is the set of sequences $(v_0, \dots, v_p)$ with $v_i \overset{G}{\underset{=}{\to}} v_{i+1}$, $i = 0, \dots, p-1$. Define $A_p^K = \langle K_p \rangle$ and $\mathcal{A}_p^K = \pi_p(A_p^K)$, called the set of *allowed p-paths on G*

We would like to define the chain complex

$$\cdots \to \mathcal{A}_p^K \xrightarrow{\partial^{\mathrm{reg}}} \mathcal{A}_{p-1}^K \to \cdots$$

but for some $a \in \mathcal{A}_p^K$ it can happen that $\partial^{\mathrm{reg}}(a) \notin \mathcal{A}_{p-1}^K$. To fix this, consider

$$\Xi_p^K = \{a \in \mathcal{A}_p^K \mid \partial^{\mathrm{reg}}(a) \in \mathcal{A}_{p-1}^K\}.$$

Thus,

$$\Xi_\bullet^K \colon \quad \cdots \to \Xi_p^K \xrightarrow{\partial^{\mathrm{reg}}} \Xi_{p-1}^K \to \cdots$$

is a chain complex.

The *p-dimensional path homology of G* is the vector space $H_p^{\Xi}(G) \coloneqq H_p(\Xi_\bullet^K)$.

Remark 4.28. Definition 4.27 can be seen as a modification of Definition 4.13. Let Ω^{ord} be the set of line graphs as in 4.27. Let $G = (V, E)$ be a graph and consider $K = OC^{\Omega^{\mathrm{ord}}}(G)$, with $K_p = OC_p^{\Omega^{\mathrm{ord}}}(G)$ as before. In both definitions we have $A_p^K = \langle K_p \rangle$. In 4.13, we are interested in the chain complex given by

$$\Theta_p^K = \{a \in A_p^K \mid \partial(a) \in A_{p-1}^K\},$$

whose homology we will refer to as the *path homology with repetitions*, while in 4.27 we set

$$\Xi_p^K = \{a \in \pi_p(A_p^K) \mid \partial^{\mathrm{reg}}(a) \in \pi_p(A_{p-1}^K)\},$$

where $\pi_p \colon \Lambda_p(V) \to \Lambda_p(V)/\mathcal{I}_p(V)$ is the projection map, as in 4.27. This map π_p acts as a "regularization" in the sequences, in the sense that it identifies to 0 the sequences with adjacent repetitions (i.e. irregular sequences), for example, (v_0, v_1, v_1).

Notice that $K = OC^{\Omega^{\mathrm{ord}}}(G)$ satisfies property (P1) of Definition 4.9, and thus the map $\mathcal{N}_{\mathrm{ext}} \to \mathbf{Pers}$ given by $\mathbf{X} \mapsto \mathrm{dgm}_p^{\Omega^{\mathrm{ord}}}(\mathbf{X})$ is stable (i.e. 2-Lipschitz), by Theorem 4.19.

Path homology theory has many remarkable properties which resemble the usual simplicial and singular homology theories from topology: it has cone-like constructions, a suspension isomorphism, graphs that behave as spheres at the homological level (with respect to suspensions), and a Künnet formula (see [82]); it has a homotopy theory that agrees with the topology (see [76]): retractions, homotopy groups and a relation between the first homotopy group and the first (path) homology group; and it also has a path cohomology (see [79]). But what makes it so special so as to possess many desirable properties? Is the "regularization" step of Remark 4.28 essential?

The next proposition shows a relation between the path homology with repetitions and the usual path homology.

Proposition 4.29. Let $G = (V, E)$ be a graph, and $\Theta_\bullet^K$ and $\Xi_\bullet^K$ be the chain complexes of the path homology with repetitions and the usual path homology as in Remark 4.28. Then the maps $\pi_p \colon \Lambda_p(V) \to \Lambda_p(V)/\mathcal{I}_p(V)$ induce a map $\pi_* \colon OH_p(\Theta_\bullet^K) \to OH_p(\Xi_\bullet^K)$, for every p. Moreover, it is an isomorphism for $p = 0, 1$.

Proof. Let's prove that $\pi \colon \Theta_\bullet^K \to \Xi_\bullet^K$ given by $\{\pi_p\}$ is a chain map. Consider the diagram:

$$\begin{tikzcd}
\cdots \arrow[r, "\partial^\Theta"] & \Theta_p^K \arrow[r, "\partial^\Theta"] \arrow[d, "\pi_p"] & \Theta_{p-1}^K \arrow[r, "\partial^\Theta"] \arrow[d, "\pi_{p-1}"] & \cdots \\
\cdots \arrow[r, "\partial^\Xi"] & \Xi_p^K \arrow[r, "\partial^\Xi"] & \Xi_{p-1}^K \arrow[r, "\partial^\Xi"] & \cdots
\end{tikzcd}$$

We just need to show that the squares are commutative.

Let $a + b \in \Theta_p^K$, with $a \in \mathcal{R}_p(V)$ a sum of regular sequences and $b \in \mathcal{I}_p(V)$ a sum of irregular sequences. Then $\partial^\Xi(\pi_p(a + b)) = \partial^\Xi(a) \in \Xi_{p-1}$. On the other hand, $\partial^\Theta(a + b) = \partial^\Theta(a) + \partial^\Theta(b)$ and since $\partial^\Theta(\mathcal{I}_p(V)) \subseteq \mathcal{I}_{p-1}(V)$ by [82, Lemma 2.9], $\pi_{p-1}(\partial^\Theta(a) + \partial^\Theta(b)) = \pi_{p-1}(\partial^\Theta(a))$ which is equal to $\partial^\Xi(a)$ having "deleted" the irregular sequences in the summation. Thus, the square commutes.

This proves that we have maps $\pi_* \colon OH_p(\Theta_\bullet^K) \to OH_p(\Xi_\bullet^K)$ induced by the maps π_p.

For the isomorphisms:

The case $p = 0$ follows readily.

Now let $p = 1$. Let $\sigma = a + b \in \ker(\partial_1^{\Theta})$, where $a = \sum \alpha_i(x_i, y_i)$ is a sum of regular sequences and $b = \sum \beta_i(z_i, z_i)$ is a sum of irregular sequences. Suppose $\pi_*[\sigma] = 0$. Then there is $\tau \in \Xi_2$ such that $\partial_2^{\Xi}(\tau) = \pi_1(a + b) = a$. Now, regarding τ as an element of Θ_2, we have $\partial_2^{\Theta}(\tau) = a + c$ where c is a summation of irregular elements (that possible arise when applying the boundary in elements of the form (a, b, a)). Write $c = \sum \gamma_i(w_i, w_i)$. Let $\tau' = \sum \beta_i(z_i, z_i, z_i) - \sum \gamma_i(w_i, w_i, w_i)$. Then $\partial^{\Theta}(\tau + \tau') = (a + c) + b - c = a + b$. Thus, $[\sigma] = 0$. This proves that π_* is injective. The surjective is readily. □

Remark 4.30. The strategy used to prove the isomorphism of Proposition 4.29 is, in short: suppose $\pi_k(\sigma)$ is the boundary of an element $\tau \in \Xi_{k+1}$. Then consider τ as an element of Θ_{k+1} and try to find τ' such that $\partial_{k+1}^{\Theta}(\tau + \tau') = \sigma$.

The difficulties arise when we consider the *semi-regulars* elements, i.e., sequences $(v_0, \ldots, v_p)$ where $v_i = v_{i+2}$ for some i. These sequences have irregular elements in the boundary, and these irregular sequences will be quotiented to zero in Ξ but not in Θ.

Some progress could be made if we were able to prove the following: every element in H_p^{Ξ} has a representer σ that is a sum of regular sequences which are not semi-regular. We don't know if this is true or not.

5 Operations on graphs and networks

I now hurried from star to star, a lost dog looking for its master.

OLAF STAPLEDON, *STAR MAKER*

Operations are useful to construct new objects that somehow relate to the objects that originated it: given two topological spaces A and B, for example, one can define the product $A \times B$, or the suspension ΣA (see [9]). In this chapter we will seek for analogues of these two operations and investigate how they behave with respect to homology.

5.1 Products and Künneth formulas

A Künneth formula gives a relation between the homology of a product of objects (e.g., topological spaces, chain complexes) and the homology of its factors (see [9]).

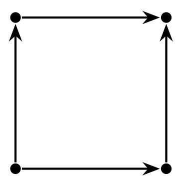
The path homology, as we will see next, behaves specially well with respect to some constructions which are analogous to some operations in classical topology: the join (see [81] for more details), the suspension, and the square product. Later, we will study a different (and somewhat more "natural") product, called "max product", and we will obtain Künneth formulas for the Vietoris-Rips and Dowker constructions, inspired by the work [84].

5.1.1 The path homology case

Definition 5.1. Given two graphs $G = (V, E)$ and $G' = (V', E')$, the *square product* $G \square G'$ is the graph with vertex set $V \times V'$ and arrows

$$(v_1, v_1') \xrightarrow{G \square G'} (v_2, v_2') \Leftrightarrow [v_1 = v_2 \text{ and } v_1' \xrightarrow{G'} v_2'] \text{ or } [v_1 \xrightarrow{G} v_2 \text{ and } v_1' = v_2'].$$

For example, $L_2 \square L_2$ is the following graph:



Theorem 5.2 (Künneth formula [82, Theorem 7.6]). For any graphs G and G' and any integer $r \geq 0$, we have

$$H_r^{\Xi}(G \,\square\, G') \simeq \bigoplus_{\substack{p,q\geq 0\\ p+q=r}} \Big(H_p^{\Xi}(G) \otimes H_q^{\Xi}(G') \Big),$$

where the symbol $\otimes$ denotes the tensor product of two vector spaces (over the field $\mathbb{K}$).

A useful way to think about the tensor product of two vector spaces V and W is the following. Let $\dim(V) = m$ and $\dim(W) = n$. Consider bases $\{v_1, \dots, v_m\}$ and $\{w_1, \dots, w_n\}$ for V and W, respectively. Then $V \otimes W$ is a vector space (over the field $\mathbb{K}$) with $\dim(V \otimes W) = m \times n$ and with a basis given by $\{v_i \otimes w_j;\ i = 1, \dots, m,\ j = 1, \dots, n\}$, for certain vectors $v_i \otimes w_j$. Here is a good example: consider V the vector space of real polynomials on the variable x, with degree at most 2: $V = \langle\{1, x, x^2\}\rangle$, and consider W the analogue of V with $W = \langle\{1, y\}\rangle$. Then $V \otimes W$ is isomorphic to $\langle\{1, x, x^2, y, xy, x^2y\}\rangle$, which has dimension 6. Compare $V \otimes W$ with $V \oplus W = \langle\{1, x, x^2, y\}\rangle$ which has dimension 4. For more precise definition and details about the tensor product, see [85, Chapter 1] or [9, §50]. We will use the fact that there is a natural isomorphism $V \otimes \mathbb{K} \simeq V$ for any vector space V.

Definition 5.3. Given $\mathbf{X} = (X, w_X), \mathbf{Y} = (Y, w_Y) \in \mathcal{N}_{\text{ext}}$, define the *square product* $\mathbf{X} \,\square\, \mathbf{Y} := \Psi(\{X_\epsilon \square Y_\epsilon\}_{\epsilon\in\mathbb{R}})$, where Ψ is the map from Definition 3.4. More explicitly, $\mathbf{X} \,\square\, \mathbf{Y} := (X \times Y, w_\square)$, where

$$w_\square((x, y), (x', y')) := \begin{cases} \max\{w_X(x, x'), w_Y(y, y')\}, & \text{if } x = x' \text{ or } y = y', \\ +\infty, & \text{otherwise.} \end{cases}$$

It follows from the definition of the map Ψ that, for all $\epsilon \in \mathbb{R}$, we have

$$(\mathbf{X} \,\square\, \mathbf{Y})_\epsilon = \mathbf{X}_\epsilon \,\square\, \mathbf{Y}_\epsilon.$$

Corollary 5.4 (Static Künneth formula). For all $\mathbf{X}, \mathbf{Y} \in \mathcal{N}_{\text{ext}}$, integer $r \geq 0$ and real ϵ, we have

$$H_r^{\Xi}((\mathbf{X} \,\square\, \mathbf{Y})_\epsilon) \simeq \bigoplus_{\substack{p,q\geq 0\\ p+q=r}} \Big(H_p^{\Xi}(\mathbf{X}_\epsilon) \otimes H_q^{\Xi}(\mathbf{Y}_\epsilon) \Big).$$

Proof. It follows from Theorem 5.2 and Definition 5.3. □

Corollary 5.4 is called “static” because it relates the path homology of $\mathbf{X} \,\square\, \mathbf{Y}$ with that of $\mathbf{X}$ and $\mathbf{Y}$ at a fixed parameter ϵ. We would like a persistent path homology analogue of it, that is, an equation like

$$\mathrm{dgm}_r^{\Xi}(\mathbf{X} \,\square\, \mathbf{Y}) = \bigoplus_{\substack{p,q\geq 0\\ p+q=r}} \Big(\mathrm{dgm}_p^{\Xi}(\mathbf{X}) \otimes \mathrm{dgm}_q^{\Xi}(\mathbf{Y}) \Big).$$

To do it, we need to give a meaning to the symbols $\oplus$ and $\otimes$ used for two persistent diagrams. This motivates the following definition and results.

Definition 5.5. Let $\mathbb{V} = \left\{ V_\epsilon \xrightarrow{v_\epsilon^{\epsilon'}} V_{\epsilon'} \right\}_{\epsilon\leq\epsilon'}$ and $\mathbb{W} = \left\{ W_\epsilon \xrightarrow{w_\epsilon^{\epsilon'}} W_{\epsilon'} \right\}_{\epsilon\leq\epsilon'}$ be two persistent vector spaces over the same field $\mathbb{K}$. Define the *tensor product of* $\mathbb{V}$ *and* $\mathbb{W}$ by

$$\mathbb{V} \otimes \mathbb{W} := \left\{ V_\epsilon \otimes W_\epsilon \xrightarrow{v_\epsilon^{\epsilon'} \otimes w_\epsilon^{\epsilon'}} V_{\epsilon'} \otimes W_{\epsilon'} \right\}_{\epsilon\leq\epsilon'}.$$

Proposition 5.6. Suppose that $\mathbb{V}$ and $\mathbb{W}$ are (finite) interval decomposable:

$$\mathbb{V} \simeq \bigoplus_{i=1}^{m} \mathbb{I}[b_i, d_i), \quad \mathbb{W} \simeq \bigoplus_{j=1}^{n} \mathbb{I}[\widehat{b}_j, \widehat{d}_j).$$

Then $\mathbb{V} \otimes \mathbb{W}$ is interval decomposable and

$$\mathbb{V} \otimes \mathbb{W} \simeq \bigoplus \mathbb{I}([b_i, d_i) \cap [\widehat{b}_j, \widehat{d}_j)) = \bigoplus \mathbb{I}[\max\{b_i, \widehat{b}_j\}, \min\{d_i, \widehat{d}_j\}),$$

for indices (i, j) such that $\max\{b_i, \widehat{b}_j\} \leq \min(d_i, \widehat{d}_j)$. This last condition is equivalent to $[b_i, d_i) \cap [\widehat{b}_j, \widehat{d}_j) \neq \emptyset$.

Proof. Let

$$\bigoplus_{i=1}^{m} \mathbb{I}[b_i, d_i) = \left\{\widetilde{V}_\epsilon \xrightarrow{\widetilde{v}_{\epsilon,\epsilon'}} \widetilde{V}_{\epsilon'}\right\}_{\epsilon \leq \epsilon'} \text{ and } \bigoplus_{j=1}^{n} \mathbb{I}[\widehat{b}_j, \widehat{d}_j) = \left\{\widetilde{W}_\epsilon \xrightarrow{\widetilde{w}_{\epsilon,\epsilon'}} \widetilde{W}_{\epsilon'}\right\}_{\epsilon \leq \epsilon'}.$$

Since $\mathbb{V} \simeq \bigoplus_{i=1}^{m} \mathbb{I}[b_i, d_i)$ and $\mathbb{W} \simeq \bigoplus_{j=1}^{n} \mathbb{I}[\widetilde{b}_j, \widetilde{d}_j)$, we have two families of isomorphisms $\{\phi\colon V_\epsilon \to \widetilde{V}_\epsilon\}$ and $\{\psi\colon W_\epsilon \to \widetilde{W}_\epsilon\}$ such that, for each $\epsilon \leq \epsilon'$, the diagrams below commute:

$$\begin{tikzcd} V_\epsilon \arrow[r, "v_\epsilon^{\epsilon'}"] \arrow[d, "\phi_\epsilon"] & V_{\epsilon'} \arrow[d, "\phi_{\epsilon'}"] \\ \widetilde{V}_\epsilon \arrow[r, "\widetilde{v}_\epsilon^{\epsilon'}"] & \widetilde{V}_{\epsilon'} \end{tikzcd} \qquad \begin{tikzcd} W_\epsilon \arrow[r, "w_\epsilon^{\epsilon'}"] \arrow[d, "\psi_\epsilon"] & W_{\epsilon'} \arrow[d, "\psi_{\epsilon'}"] \\ \widetilde{W}_\epsilon \arrow[r, "\widetilde{w}_\epsilon^{\epsilon'}"] & \widetilde{W}_{\epsilon'} \end{tikzcd}$$

Then, the following diagram is also commutative:

$$\begin{tikzcd} V_\epsilon \otimes W_\epsilon \arrow[r, "v_\epsilon^{\epsilon'} \otimes w_\epsilon^{\epsilon'}"] \arrow[d, "\phi_\epsilon \otimes \psi_\epsilon"] & V_{\epsilon'} \otimes W_{\epsilon'} \arrow[d, "\phi_{\epsilon'} \otimes \psi_{\epsilon'}"] \\ \widetilde{V}_\epsilon \otimes \widetilde{W}_\epsilon \arrow[r, "\widetilde{v}_\epsilon^{\epsilon'} \otimes \widetilde{v}_\epsilon^{\epsilon'}"] & \widetilde{V}_{\epsilon'} \otimes \widetilde{V}_{\epsilon'} \end{tikzcd}$$

Thus $\phi_\epsilon \otimes \psi_\epsilon$ is an isomorphism by [85, Chapter 1, Corollary 7.8], since ϕ_ϵ and ψ_ϵ are isomorphisms. The same holds for $\phi_{\epsilon'} \otimes \psi_{\epsilon'}$.

Since

$$\widetilde{V}_\epsilon = \bigoplus_{i=1}^{m} \mathbb{I}[b_i, d_i)_\epsilon \quad \text{and} \quad \widetilde{V}_{\epsilon'} = \bigoplus_{i=1}^{m} \mathbb{I}[b_i, d_i)_{\epsilon'},$$

the map $\widetilde{v}_\epsilon^{\epsilon'}$ can be decomposed as a sum $\oplus_{i=1}^{m} v_i$, where

$$v_i = \begin{cases} \text{the identity map } \mathbb{K} \to \mathbb{K}, & \text{if } [\epsilon, \epsilon'] \subseteq [b_i, d_i), \\ 0, & \text{otherwise.} \end{cases}$$

The same holds for $\widetilde{w}_\epsilon^{\epsilon'} = \oplus_{j=1}^{n} w_j$. Then $v_i \otimes w_j$ will be the identity map of $\mathbb{K} \otimes \mathbb{K} \simeq \mathbb{K}$ exactly when $[\epsilon, \epsilon'] \subseteq [b_i, d_i) \cap [\widehat{b}_j, \widehat{d}_j)$, and will be the zero map otherwise.

Hence, we can conclude that

$$\mathbb{V} \otimes \mathbb{W} \simeq \bigoplus \mathbb{I}[\max\{b_i, \widehat{b}_j\}, \min\{d_i, \widehat{d}_j\})$$

for the indices (i, j) such that $\max\{b_i, \widehat{b}_j\} \leq \min\{d_i, \widehat{d}_j\}$. □

Recall that **Bar** is the set of all finite barcodes and every element of **Bar** is a finite union of closed real intervals, with possible repetitions. Intervals of the type $[a,a)$ are regarded as an empty interval.

Definition 5.7. Given barcodes $A, B \in \mathbf{Bar}$, with $A = \bigcup_{i=1}^{m} [b_i, d_i)$, $B = \bigcup_{j=1}^{n} [\widehat{b}_j, \widehat{d}_j)$, define

$$A \oplus B \coloneqq A \cup B = \left(\bigcup_{i=1}^{m} [b_i, d_i)\right) \cup \left(\bigcup_{j=1}^{n} [\widehat{b}_j, \widehat{d}_j)\right).$$

Also, define

$$A \otimes B \coloneqq \bigcup_{i,j} \Big([b_i, d_i) \cap [\widehat{b}_j, \widehat{d}_j) \Big).$$

If $\mathbb{V}$ and $\mathbb{W}$ are such that $A = \operatorname{bar}(\mathbb{V})$ and $B = \operatorname{bar}(\mathbb{W})$, then it follows readily from Definition 1.21 that

$$\operatorname{bar}(\mathbb{V} \oplus \mathbb{W}) = \operatorname{bar}(\mathbb{V}) \oplus \operatorname{bar}(\mathbb{W}),$$

and, by Proposition 5.6,

$$\operatorname{bar}(\mathbb{V} \otimes \mathbb{W}) = \operatorname{bar}(\mathbb{V}) \otimes \operatorname{bar}(\mathbb{W}).$$

Define the operations $\oplus$ and $\otimes$ in **Pers** using the identification between **Pers** and **Bar** stated in Definition 1.21.

Now we are ready to prove the main theorem of this section.

Theorem 5.8 (Persistent Künneth formula)**.** For all $\mathbf{X}, \mathbf{Y} \in \mathcal{N}_{\text{ext}}$ and integer $r \geq 0$, we have:

$$\operatorname{dgm}_r^{\Xi}(\mathbf{X} \,\square\, \mathbf{Y}) = \bigoplus_{\substack{p,q \geq 0 \\ p+q=r}} \Big(\operatorname{dgm}_p^{\Xi}(\mathbf{X}) \otimes \operatorname{dgm}_q^{\Xi}(\mathbf{Y}) \Big).$$

Proof. Given $\epsilon \in \mathbb{R}$ and $n \geq 0$, write $X_n^{\epsilon} = H_n(\mathbf{X}_{\epsilon})$ and $Y_n^{\epsilon} = H_n(\mathbf{X}_{\epsilon})$. Also, write

$$\mathbb{V}_p = \left\{ X_p^{\epsilon} \xrightarrow{v_{\epsilon}^{\epsilon'}} X_p^{\epsilon'} \right\}_{\epsilon \leq \epsilon'}, \quad \mathbb{W}_q = \left\{ Y_q^{\epsilon} \xrightarrow{w_{\epsilon}^{\epsilon'}} Y_q^{\epsilon'} \right\}_{\epsilon \leq \epsilon'},$$

and

$$\mathbb{U}_r = \left\{ \bigoplus_{\substack{p,q \geq 0 \\ p+q=r}} (X_p^{\epsilon} \otimes Y_q^{\epsilon}) \xrightarrow{\oplus (v_{\epsilon}^{\epsilon'} \otimes w_{\epsilon}^{\epsilon'})} \bigoplus_{\substack{p,q \geq 0 \\ p+q=r}} (X_p^{\epsilon'} \otimes Y_q^{\epsilon'}) \right\}_{\epsilon \leq \epsilon'}.$$

Then

$$\operatorname{dgm}_p^{\Xi}(\mathbf{X}) = \operatorname{dgm}(\mathbb{V}_p), \ \operatorname{dgm}_q^{\Xi}(\mathbf{Y}) = \operatorname{dgm}(\mathbb{W}_q), \text{ and } \mathbb{U}_r = \bigoplus_{\substack{p,q \geq 0 \\ p+q=r}} (\mathbb{V}_p \otimes \mathbb{W}_q).$$

By [82, Theorem 4.7], the induced of the cross product map $\times$ (given by the linear extension of $a \otimes b \mapsto a \times b$) is an isomorphism, and then the diagram

$$\begin{tikzcd}
\bigoplus_{\substack{p,q \geq 0 \\ p+q=r}} (X_p^{\epsilon} \otimes Y_q^{\epsilon}) \arrow[r] \arrow[d, "\times"] & \bigoplus_{\substack{p,q \geq 0 \\ p+q=r}} (X_p^{\epsilon'} \otimes Y_q^{\epsilon'}) \arrow[d, "\times"] \\
H_r^{\Xi}((\mathbf{X} \,\square\, \mathbf{Y})_{\epsilon}) \arrow[r] & H_r^{\Xi}((\mathbf{X} \,\square\, \mathbf{Y})_{\epsilon'})
\end{tikzcd}$$

commutes, since the horizontal maps are induced by the inclusions.

Therefore, the persistent module

$$\{H_r^{\Xi}((\mathbf{X} \square \mathbf{Y})_\epsilon) \to \{H_r^{\Xi}((\mathbf{X} \square \mathbf{Y})_{\epsilon'})\}_{\epsilon \le \epsilon'}$$

giving rise to $\mathrm{dgm}_r^{\Xi}(\mathbf{X} \square \mathbf{Y})$ is isomorphic to $\mathbb{U}_r$.

Finally,

$$\begin{aligned} \mathrm{dgm}_r^{\Xi}(\mathbf{X} \square \mathbf{Y}) &= \mathrm{bar}(\mathbb{U}_r) = \mathrm{bar}\left(\bigoplus_{\substack{p,q \ge 0 \\ p+q=r}} (\mathbb{V}_p \otimes \mathbb{W}_q) \right) = \bigoplus_{\substack{p,q \ge 0 \\ p+q=r}} \mathrm{bar}(\mathbb{V}_p \otimes \mathbb{W}_q) \\ &= \bigoplus_{\substack{p,q \ge 0 \\ p+q=r}} (\mathrm{bar}(\mathbb{V}_p) \otimes \mathrm{bar}(\mathbb{W}_q)) = \bigoplus_{\substack{p,q \ge 0 \\ p+q=r}} \left(\mathrm{dgm}_p^{\Xi}(\mathbf{X}) \otimes \mathrm{dgm}_q^{\Xi}(\mathbf{Y}) \right). \end{aligned}$$ □

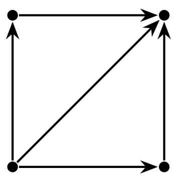
5.1.2 The case of Vietoris-Rips and Dowker homologies

Definition 5.9. Given two graphs $G = (V, E)$ and $G' = (V', E')$, consider now a different product called *max product*, denoted by $G \rtimes G'$. It has vertex set $V \times V'$, and arrows

$$(v_1, v_1') \xrightarrow[=]{G \rtimes G'} (v_2, v_2') \Leftrightarrow v_1 \overset{G}{\underset{=}{\Rightarrow}} v_2 \text{ and } v_1' \overset{G'}{\underset{=}{\Rightarrow}} v_2'.$$

We call *diagonals* the arrows $(v_1, v_1') \xrightarrow{G \rtimes G'} (v_2, v_2')$ with $v_1 \neq v_2$ and $v_1' \neq v_2'$.

For example, $L_2 \rtimes L_2$ is the following graph:



The name "max product" comes from the following definition:

Definition 5.10. Given $\mathbf{X}, \mathbf{Y} \in \mathcal{N}$, define their *max product* by $\mathbf{X} \rtimes \mathbf{Y} := \Psi(\{\mathbf{X}_\epsilon \rtimes \mathbf{Y}_\epsilon\}_{\epsilon \in \mathbb{R}})$, as in Definition 3.4. It follows from the definition of Ψ that

$$(\mathbf{X} \rtimes \mathbf{Y})_\epsilon = \mathbf{X}_\epsilon \rtimes \mathbf{Y}_\epsilon,$$

for all $\epsilon \in \mathbb{R}$. More explicitly, $\mathbf{X} \rtimes \mathbf{Y} = (X \times Y, w_\rtimes)$, where

$$w_\rtimes((x, y), (x', y')) = \max\{\widetilde{w}_X(x, x'), \widetilde{w}_Y(y, y')\},$$

$M(\mathbf{X}) = (X, \widetilde{w}_X)$, $M(\mathbf{Y}) = (Y, \widetilde{w}_Y)$ and M is the diagonal majoration of Definition 3.4:

$$\tilde{w}_X(x, x') = \max\{w_X(x, x), w_X(x, x'), w_X(x', x')\},$$

and the same for $\tilde{w}_Y$.

Notice that when $\mathbf{X}$ and $\mathbf{Y}$ are metric spaces, $w_{\rtimes}$ is the usual “max product”, since M is the identity on $\mathcal{N}_{\mathcal{G}}$, which contains the set of metric spaces.

By definition, $G \rtimes G'$ is obtained from $G \,\square\, G'$ by adding to the last the diagonal arrows. Altought $G \rtimes G'$ has (in general) more arrows than $G \,\square\, G'$, their first path homology groups are isomorphic. More precisely:

Theorem 5.11. The map $\iota\colon H_p^{\Xi}(G \,\square\, G') \to H_p^{\Xi}(G \rtimes G')$ induced by the inclusion $G \,\square\, G' \hookrightarrow G \rtimes G'$ is an isomorphism, for $p = 0, 1$.

Proof. Write $G_{\square} = G \,\square\, G'$ and $G_{\rtimes} = G \rtimes G'$. Let $\pi\colon G_{\rtimes} \to G$ and $\pi'\colon G_{\rtimes} \to G'$ be the projections.

The case $p = 0$ follows readily: just notice that the connected components of $G \,\square\, G'$ and $G \rtimes G'$ are equal.

Consider $p = 1$.

Claim 5.11.1. The map ι is surjective.

Proof. Let

$$\Xi_{\bullet}^{\square}\colon \qquad \cdots \to \Xi_2^{\square} \xrightarrow{\partial^{\square}} \Xi_1^{\square} \to \cdots$$

and

$$\Xi_{\bullet}^{\rtimes}\colon \qquad \cdots \to \Xi_2^{\rtimes} \xrightarrow{\partial^{\rtimes}} \Xi_1^{\rtimes} \to \cdots$$

be the chain complexes that yield the path homology of $G_{\square}$ and $G_{\rtimes}$, respectively (see Definition 4.27).

Let $\beta = \sum_i \lambda_i (a_i, c_i) \in \Xi_1^{\rtimes}$. For all k such that the arrow (a_k, c_k) is a diagonal, consider the 2-path $t_k = (a_k, b_k, c_k)$, where $b_k = (\pi(c_k), \pi'(a_k))$, like in the following figure.

$$\begin{tikzcd} & c_k \\ a_k \arrow[ur] \arrow[r, dashed] & b_k \arrow[u, dashed] \end{tikzcd}$$

It is clear that $t_k \in \Xi_2^{\rtimes}$. Then $\beta + \partial(\lambda_k t_k)$ has no diagonals (a_k, c_k) anymore, and at the homology level we have $[\beta] = [\beta + \partial(\lambda_k t_k)]$. Taking $\alpha = \beta + \partial(\lambda_k t_k)$, we have $\alpha \in \Xi_1^{\square}$ and $\iota[\alpha] = [\alpha] = [\beta]$. This proves the surjectivity. ■

Since $H_1(G_{\square})$ and $H_1(G_{\rtimes})$ are vector spaces, we conclude that

$$\dim H_1^{\Xi}(G_{\square}) \geq \dim H_1^{\Xi}(G_{\rtimes}).$$

Claim 5.11.2. There is a surjection $T\colon H_1^{\Xi}(G_{\rtimes}) \to H_1^{\Xi}(G_{\square})$.

Proof. First, suppose that G and G' are connected. By the Künneth formula,

$$H_1^{\Xi}(G_{\square}) = H_1^{\Xi}(G \,\square\, G') \simeq \left(H_1^{\Xi}(G) \otimes H_0^{\Xi}(G')\right) \oplus \left(H_0^{\Xi}(G) \otimes H_1^{\Xi}(G')\right) \simeq H_1^{\Xi}(G) \oplus H_1^{\Xi}(G'),$$

since $H_0^{\Xi}(G) \simeq \mathbb{K} \simeq H_0^{\Xi}(G')$.

Notice that the projections π and π' are graph maps: given $a \xrightarrow{G_{\rtimes}} b$, it follows from Definition 5.9 that $\pi(a) \overset{G}{\underset{=}{\rightarrow}} \pi(b)$ and $\pi'(a) \overset{G'}{\underset{=}{\rightarrow}} \pi'(b)$.

Consider then the induced maps $\pi_*\colon H_1^{\Xi}(G_{\rtimes}) \to H_1^{\Xi}(G)$ and $\pi'_*\colon H_1^{\Xi}(G_{\rtimes}) \to H_1^{\Xi}(G')$. Remember that at chain level, for any 1-path $(a,b) \in \Xi_1^{\rtimes}$ we have

$$\pi(a,b) = (\pi(a), \pi(b)), \quad \pi'(a,b) = (\pi'(a), \pi'(b)),$$

where $(a,b) = 0$ if $a = b$.

Define $T\colon H_1^{\Xi}(G_{\rtimes}) \to H_1^{\Xi}(G) \oplus H_1^{\Xi}(G')$ by $T[w] = (\pi_*[w], \pi'_*[w])$. Fix $v_0 \in G$ and $v'_0 \in G'$. Let $[\alpha] \in H_1^{\Xi}(G)$ with $\alpha = \sum_i \lambda_i(c_i, d_i)$ and $[\beta] \in H_1^{\Xi}(G')$ with $\beta = \sum_j \lambda'_j(c'_j, d'_j)$. Define $w_i = ((c_i, v'_0), (d_i, v'_0))$, $w'_j = ((v_0, c'_j), (v_0, d'_j))$, $w = \sum_i \lambda_i w_i$ and $w' = \sum_j \lambda'_j w'_j$. Then w and w' are cycles (that is, $\partial^{\rtimes}(w) = 0$ and $\partial^{\rtimes}(w') = 0$), $\pi(w') = \pi'(w) = 0$, $\pi(w) = \alpha$ and $\pi'(w') = \beta$. Thus,

$$T[w + w'] = (\pi_*[w + w'], \pi'_*[w + w']) = (\pi_*[w], \pi'_*[w']) = ([\alpha], [\beta]),$$

which shows that the map $T\colon H_1^{\Xi}(G_{\rtimes}) \to H_1^{\Xi}(G) \oplus H_1^{\Xi}(G') \simeq H_1^{\Xi}(G_{\Box})$ is surjective.

The case when G or G' is not connected is done in a similar way considering each connected component of $G_{\rtimes}$. ■

Thus, we have

$$\dim H_1^{\Xi}(G_{\Box}) \le \dim H_1^{\Xi}(G_{\rtimes}),$$

which implies $\dim H_1^{\Xi}(G_{\Box}) = \dim H_1^{\Xi}(G_{\rtimes})$. Since ι is surjective it must also be injective, hence an isomorphism. □

Corollary 5.12. For any two weakly connected graphs G and G', we have

$$H_1^{\Xi}(G \rtimes G') \simeq H_1^{\Xi}(G \,\Box\, G') \simeq H_1^{\Xi}(G) \oplus H_1^{\Xi}(G').$$

Proof. The first isomorphism comes from Theorem 5.11 above and the second one from Theorem 5.2. □

Given $\mathbf{X}, \mathbf{Y} \in \mathcal{N}_{\text{ext}}$, the square product $\mathbf{X} \,\Box\, \mathbf{Y}$ has an undesirable property: even if $\mathbf{X}$ and $\mathbf{Y}$ are metric spaces, in general $\mathbf{X} \,\Box\, \mathbf{Y}$ is *not* one, because some weights will be $+\infty$. This does not occur with $\mathbf{X} \rtimes \mathbf{Y}$, since it is the usual “max product” of metric spaces.

Corollary 5.13. Given $\mathbf{X}, \mathbf{Y} \in \mathcal{N}_{\text{ext}}$, we have

$$\mathrm{dgm}_p^{\Xi}(\mathbf{X} \rtimes \mathbf{Y}) = \mathrm{dgm}_p^{\Xi}(\mathbf{X} \,\Box\, \mathbf{Y}), \quad \text{for } p = 0, 1.$$

Proof. Remember that

$$\mathrm{dgm}_p^{\Xi}(\mathbf{X} \,\Box\, \mathbf{Y}) = \mathrm{dgm}(\mathbb{V}) \quad \text{and} \quad \mathrm{dgm}_p^{\Xi}(\mathbf{X} \rtimes \mathbf{Y}) = \mathrm{dgm}(\mathbb{W}),$$

where

$$\mathbb{V} = \{H_p^{\Xi}((\mathbf{X} \,\Box\, \mathbf{Y})_{\epsilon}) \to H_p^{\Xi}((\mathbf{X} \,\Box\, \mathbf{Y})_{\epsilon'})\}_{\epsilon \le \epsilon'} \text{ and}$$
$$\mathbb{W} = \{H_p^{\Xi}((\mathbf{X} \rtimes \mathbf{Y})_{\epsilon}) \to H_p^{\Xi}((\mathbf{X} \rtimes \mathbf{Y})_{\epsilon'})\}_{\epsilon \le \epsilon'},$$

and the maps are induced by the inclusions. By Theorem 5.11, the map $\iota\colon H_p^{\Xi}((\mathbf{X}\,\square\,\mathbf{Y})_\epsilon) \to H_p^{\Xi}((\mathbf{X}\rtimes\mathbf{Y})_\epsilon)$ induced by the inclusion $(\mathbf{X}\,\square\,\mathbf{Y})_\epsilon \hookrightarrow (\mathbf{X}\rtimes\mathbf{Y})_\epsilon$ is an isomorphism, and the following diagram

$$\begin{tikzcd} H_p^{\Xi}((\mathbf{X}\,\square\,\mathbf{Y})_\epsilon) \arrow[r] \arrow[d, "i"'] & H_p^{\Xi}((\mathbf{X}\,\square\,\mathbf{Y})_{\epsilon'}) \arrow[d, "i"] \\ H_p^{\Xi}((\mathbf{X}\rtimes\mathbf{Y})_\epsilon) \arrow[r] & H_p^{\Xi}((\mathbf{X}\rtimes\mathbf{Y})_{\epsilon'}) \end{tikzcd}$$

clearly commutes. This proves that $\mathbb{U} \simeq \mathbb{V}$, and consequently

$$\mathrm{dgm}_p^{\Xi}(\mathbf{X}\,\square\,\mathbf{Y}) = \mathrm{bar}(\mathbb{U}) = \mathrm{bar}(\mathbb{V}) = \mathrm{dgm}_p^{\Xi}(\mathbf{X}\rtimes\mathbf{Y}). \qquad \square$$

Definition 5.14 ([86, Definition 4.25])**.** Given simplicial complexes K and L with vertex set V_K and V_L, respectively, their (categorical) cartesian product is the simplicial complex $K \times L$ with vertex set $V_K \times V_L$ and with simplexes given as follows:

$$\sigma \in K \times L \Leftrightarrow \sigma \subseteq \sigma_1 \times \sigma_2, \text{ for some } \sigma_1 \in K,\ \sigma_2 \in L.$$

Now given two simplicial complex filtrations $\mathcal{K} = \{K_\epsilon\}_{\epsilon\in\mathbb{R}}$ and $\mathcal{L} = \{L_\epsilon\}_{\epsilon\in\mathbb{R}}$, define

$$\mathcal{K} \times \mathcal{L} \coloneqq \{K_\epsilon \times L_\epsilon\}_{\epsilon\in\mathbb{R}}.$$

Theorem 5.15. For any graphs G and G', we have

$$C^{\mathrm{Rips}}(G \rtimes G') = C^{\mathrm{Rips}}(G) \times C^{\mathrm{Rips}}(G'),$$

and

$$C^{\mathrm{Dow}}(G \rtimes G') = C^{\mathrm{Dow}}(G) \times C^{\mathrm{Dow}}(G').$$

Proof. We have

$$\begin{aligned} &\{(v_0, v_0'), \ldots, (v_n, v_n')\} \in C^{\mathrm{Rips}}(G \rtimes G') \Leftrightarrow \\ &\qquad (v_i, v_i') \xrightarrow[=]{G\rtimes G'} (v_j, v_j'),\ \forall i, j \Leftrightarrow v_i \xrightarrow[=]{G} v_j \text{ and } v_i' \xrightarrow[=]{G'} v_j',\ \forall i, j \Leftrightarrow \\ &\qquad\qquad \{v_0, \ldots, v_n\} \in C^{\mathrm{Rips}}(G) \text{ and } \{v_0', \ldots, v_n'\} \in C^{\mathrm{Rips}}(G'), \end{aligned}$$

that is, $C^{\mathrm{Rips}}(G \rtimes G') = C^{\mathrm{Rips}}(G) \times C^{\mathrm{Rips}}(G')$.

For the Dowker complex,

$$\begin{aligned} &\{(v_0, v_0'), \ldots, (v_n, v_n')\} \in C^{\mathrm{Dow}}(G \rtimes G') \Leftrightarrow \\ &\qquad \exists\ (\widetilde{v}, \widetilde{v}') \in G \rtimes G' \text{ such that } (v_i, v_i') \xrightarrow[=]{G\rtimes G'} (\widetilde{v}, \widetilde{v}'),\ \forall i \Rightarrow \\ &\qquad\qquad v_i \xrightarrow[=]{G} \widetilde{v} \text{ and } v_i' \xrightarrow[=]{G'} \widetilde{v}',\ \forall i \Rightarrow \\ &\qquad\qquad \{v_0, \ldots, v_n\} \in C^{\mathrm{Dow}}(G) \text{ and } \{v_0', \ldots, v_n'\} \in C^{\mathrm{Dow}}(G'),\ \forall i, j. \end{aligned}$$

Similarly, if $\{(v_0, v_0'), \ldots, (v_n, v_n')\}$ is such that $\{v_0, \ldots, v_n\} \in C^{\mathrm{Dow}}(G)$ and $\{v_0', \ldots, v_n'\} \in C^{\mathrm{Dow}}(G')$, then $\{(v_0, v_0'), \ldots, (v_n, v_n')\} \in C^{\mathrm{Dow}}(G \rtimes G')$, which implies $C^{\mathrm{Dow}}(G \rtimes G') = C^{\mathrm{Dow}}(G) \times C^{\mathrm{Dow}}(G')$. $\square$

In a recent paper [84], the authors explore many persistent Künneth formulas. In particular, they obtain a Künneth formula for the max product of two metric spaces. We will use their results to obtain our persistent version of it in the context of networks.

Theorem 5.16. Let $\mathcal{K} = \{K_\epsilon\}_{\epsilon\in\mathbb{R}}$ and $\mathcal{L} = \{L_\epsilon\}_{\epsilon\in\mathbb{R}}$ be two simplicial complex filtrations. Then, for any integer $r \geq 0$,

$$\mathrm{dgm}_r(\mathcal{K}\times\mathcal{L}) = \bigoplus_{\substack{p,q\geq 0\\ p+q=r}} (\mathrm{dgm}_p(\mathcal{K}) \otimes \mathrm{dgm}_q(\mathcal{L})).$$

Proof. This is a special case of [84, Theorem 4.4]. See Definitions 5.7 and 5.14. □

Theorem 5.17. Let $\mathbf{X}, \mathbf{Y} \in \mathcal{N}_{\mathrm{ext}}$. Then, for any integer $r \geq 0$, we have

$$\mathrm{dgm}_r^{\mathrm{Rips}}(\mathbf{X} \rtimes \mathbf{Y}) = \bigoplus_{\substack{p,q\geq 0\\ p+q=r}} \left(\mathrm{dgm}_p^{\mathrm{Rips}}(\mathbf{X}) \otimes \mathrm{dgm}_q^{\mathrm{Rips}}(\mathbf{Y})\right),$$

and

$$\mathrm{dgm}_r^{\mathrm{Dow}}(\mathbf{X} \rtimes \mathbf{Y}) = \bigoplus_{\substack{p,q\geq 0\\ p+q=r}} \left(\mathrm{dgm}_p^{\mathrm{Dow}}(\mathbf{X}) \otimes \mathrm{dgm}_q^{\mathrm{Dow}}(\mathbf{Y})\right).$$

Proof. Follows directly from Proposition 5.15 and Theorem 5.16. □

5.2 Suspension

Definition 5.18. Given a graph $G = (V, E)$, its *suspension* is the graph $\Sigma G := G * D_2$ where the vertex set of D_2 is $\{S, N\}$ (for South and North). See Figure 5.1. More explicitly,

$$v \xrightarrow[=]{\Sigma G} v' \Leftrightarrow v \xrightarrow[=]{G} v' \text{ or } (v, v') \in V \times \{S, N\}.$$

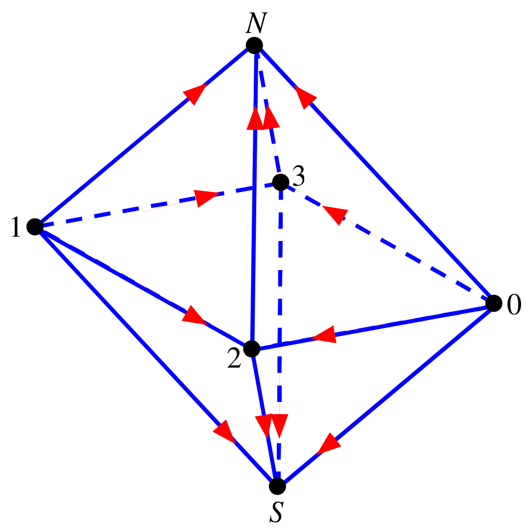


Figure 5.1: A three-dimensional representation of the suspension of the graph with vertex set $\{0, 1, 2, 3\}$. Source of the image: [82].

The *reduced path homology* of G is obtained from the chain complex

$$\widetilde{\Xi}_\bullet: \qquad \cdots \to \Xi_2^K \xrightarrow{\partial^{\mathrm{reg}}} \Xi_1^K \xrightarrow{\partial^{\mathrm{reg}}} \Xi_0^K \xrightarrow{\partial^{\mathrm{reg}}} \mathbb{K} \xrightarrow{\partial^{\mathrm{reg}}} 0,$$

where Ξ_p^K is as in 4.27. That is: $\Xi_{-1}^K = \mathbb{K}$ instead of 0. We denote the p-dimensional homology of this chain complex as $\widetilde{H}_p^{\Xi}(G)$. Notice that $\widetilde{H}_p^{\Xi}(G) = H_p^{\Xi}(G)$ for any $p > 0$.

Theorem 5.19 ([81, Proposition 5.10])**.** For any graph G and any integer $p \geq 0$, we have

$$\widetilde{H}_p^{\Xi}(G) \simeq \widetilde{H}_{p+1}^{\Xi}(\Sigma G).$$

Proof. According to the proof of [81, Proposition 5.10], the isomorphism $\widetilde{H}_p^{\Xi}(G) \simeq \widetilde{H}_{p+1}^{\Xi}(\Sigma G)$ is given by the map $a \mapsto aN - aS$, where

$$aN := \sum_i \lambda_i(v_0, \ldots, v_n, N)$$

when $a = \sum_i \lambda_i(v_0, \ldots, v_n)$, and analogue for aS. □

Definition 5.20. Given an extended network $\mathbf{X} = (X, w_X)$, define its *suspension* by

$$\Sigma\mathbf{X} := \Psi(\{\Sigma(\mathbf{X}_\epsilon)\}_{\epsilon \in \mathbb{R}}),$$

where Ψ is the map from Definition 3.4. Thus,

$$(\Sigma\mathbf{X})_\epsilon = \Sigma(\mathbf{X}_\epsilon),$$

for all $\epsilon \in \mathbb{R}$. More explicitly, $\Sigma\mathbf{X} = (\Sigma X, \Sigma w_X)$ where $\Sigma X := X \sqcup \{S, N\}$ and

$$\begin{aligned} \Sigma w_X(x, x') &= w_X(x, x'), \ x, x' \in X, \\ \Sigma w_X(x, S) &= \Sigma w_X(x, N) = w_X(x, x), \\ \Sigma w_X(S, S) &= \Sigma w_X(N, N) = \min(w_X), \end{aligned}$$

and $+\infty$ otherwise.

Corollary 5.21 (Persistent suspension isomorphism)**.** Given an extended network $\mathbf{X}$ and integer $p \geq 0$, we have

$$\widetilde{\mathrm{dgm}}_p^{\Xi}(\mathbf{X}) = \widetilde{\mathrm{dgm}}_{p+1}^{\Xi}(\Sigma\mathbf{X}),$$

where $\widetilde{\mathrm{dgm}}_i^{\Xi}$ is the persistent diagram calculated over the reduced homology. So, for $p > 0$, $\widetilde{\mathrm{dgm}}_p^{\Xi}(G) = \mathrm{dgm}_p^{\Xi}(G)$.

Proof. Given $\epsilon \leq \epsilon'$, the following diagram clearly commutes

$$\begin{array}{ccc} \widetilde{H}_p^{\Xi}(\mathbf{X}_\epsilon) & \longrightarrow & \widetilde{H}_p^{\Xi}(\mathbf{X}_{\epsilon'}) \\ \downarrow & & \downarrow \\ \widetilde{H}_{p+1}^{\Xi}((\Sigma\mathbf{X})_\epsilon) & \longrightarrow & \widetilde{H}_{p+1}^{\Xi}((\Sigma\mathbf{X})_{\epsilon'}), \end{array}$$

where the horizontal maps are induced by inclusions and the vertical maps are the isomorphism of Theorem 5.19. The result follows. □

5.3 Relations between path homology and Dowker homology

As mentioned in Example 4.6, the p-dimensional Dowker sink and source homologies of a graph G are isomorphic, for any $p \geq 0$. Moreover, we have the following result:

Proposition 5.22 ([18, Corollary 20])**.** For any network $\mathbf{X} = (X, w_X)$ and any $p \geq 0$, the persistent diagrams of $\mathbf{X}$ arising from the p-dimensional Dowker sink homology and Dowker source homology are equal.

Thus, we can refer to *the* persistent diagram arising from the Dowker (sink) homology, denoted by $\mathrm{dgm}_p^{\mathrm{Dow}}(\mathbf{X})$.

Two theorems of [70] relate the path homology with the Dowker homology. In the proofs, the authors used a lemma which requires the field $\mathbb{K}$ of coefficients to be *finite*. We will now remove this restriction.

Definition 5.23 ([70])**.** Let $G = (V, E)$ be a graph.

- A three point subset $\{a, b, c\}$ of V is a *triangle* if $a \longrightarrow b$, $b \longrightarrow c$ and $b \longrightarrow c$ in G.
- A four point subset $\{a, b, c, d\}$ of V is a *short square* if $a \longrightarrow b$, $a \longrightarrow d$, $c \longrightarrow b$ and $c \longrightarrow d$ in G, but $a \nrightarrow c$, $c \nrightarrow a$, $b \nrightarrow d$ and $d \nrightarrow b$.
- A four point subset $\{a, b, c, d\}$ of V is a *long square* if $a \longrightarrow b$, $b \longrightarrow c$, $a \longrightarrow d$, and $d \longrightarrow c$ in G but $a \nrightarrow c$.

We say that G is *square-free* if it doesn't contain a short square nor a long square. See Figure 5.2.

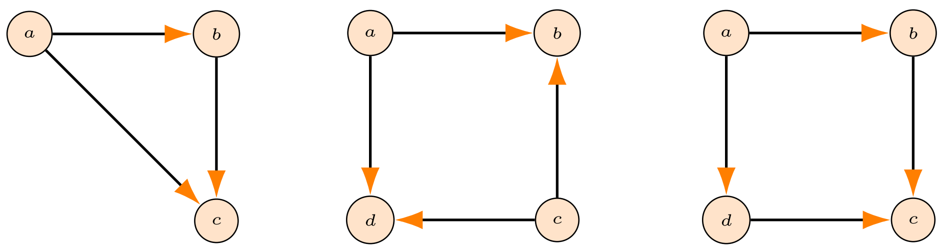


Figure 5.2: A triangle, a short square and a long square. Source of the image: [70].

We say that a network $\mathbf{X}$ is *square-free* if $\mathbf{X}_\epsilon$ is square-free for all $\epsilon \in \mathbb{R}$.

Definition 5.24. Given a graph G, we say that a simplex $[v_0, \ldots, v_p] \in C_p^{\mathrm{Dow}}(G)$ is of *type I* if for some $k \in \{0, \ldots, p\}$ we have $v_i \overset{G}{\underset{=}{\to}} v_k$, for all i. The vertex v_k is called *the sink* of $[v_0, \ldots, v_p]$. Otherwise, we say that $[v_0, \ldots, v_p]$ is of *type II*.

For the next two theorems, we will follow the corresponding proofs of Theorem 4.2 and Theorem 4.3 of [70] and omit the parts that do not assume that the field $\mathbb{K}$ (over which the homology is calculated) is finite. The reader must be careful with the difference in notation.

Theorem 5.25. Let $\mathbf{X} \in \mathcal{N}$ be a square-free network. Then

$$\mathrm{dgm}_1^{\Xi}(\mathbf{X}) = \mathrm{dgm}_1^{\mathrm{Dow}}(\mathbf{X}).$$

Proof. Fix $\epsilon \in \mathbb{R}$ and write $G = \mathbf{X}_\epsilon$.

Recall that $H_1^{\Xi}(G)$ is obtained from the sequence

$$\cdots \xrightarrow{\partial_3^{\Xi}} \Xi_2(G) \xrightarrow{\partial_2^{\Xi}} \Xi_1(G) \xrightarrow{\partial_1^{\Xi}} \Xi_0(G) \to 0,$$

and $H_1^{\mathrm{Dow}}(G)$ is obtained from

$$\cdots \xrightarrow{\partial_3^{\mathrm{Dow}}} C_2^{\mathrm{Dow}}(G) \xrightarrow{\partial_2^{\mathrm{Dow}}} C_1^{\mathrm{Dow}}(G) \xrightarrow{\partial_1^{\mathrm{Dow}}} C_0^{\mathrm{Dow}}(G) \to 0.$$

Define $\phi\colon \Xi_1(G) \to C_1^{\mathrm{Dow}}(G)$ by $\phi(a,b) = [a,b]$. This map is well-defined, as well as its induced in homology: $\phi_*\colon H_1^{\Xi}(G) \to H_1^{\mathrm{Dow}}(G)$. Let's prove that it is injective.

For this, let $[v] \in H_1^{\Xi}(G)$ and suppose that $\phi_*[v] = 0$. This is equivalent to say that there is $w = \sum_i \lambda_i \tau_i \in C_2^{\mathrm{Dow}}(G)$ such that $\partial_2^{\mathrm{Dow}}(w) = \phi(v)$.

Claim 5.25.1. There is $w' = \sum_j \alpha_j \tau'_j$ homologous to w, where each τ'_j is a triangle, that is: $\tau'_j = [x,y,z]$ with $x \to y$, $y \to z$ and $x \to z$

Proof. Indeed, if $\tau_i = [x,y,z]$ is of type II, then τ_i is homologous to a sum of type I simplexes. Suppose then that each τ_i is of type I.

Suppose now that $\tau_1 = [x,y,z_1]$ with sink z_1 is not a triangle. Then $x \to z_1$ and $y \to z_1$ but $x \nrightarrow y$ and $y \nrightarrow x$ (otherwise it would be a triangle). Thus, (x,y) is not in $\phi(v)$, so all the terms $[x,y]$ appearing in $\partial_2^{\mathrm{Dow}}(w)$ must cancellate each other. Let $\Lambda = \{i \mid [x,y] \text{ is a face of } \tau_i\}$. Suppose, for easiness, that $\Lambda = \{1,\ldots,n\}$ and $\tau_i = [x,y,z_i]$. To cancellate the terms $[x,y]$, we must have $\sum_{i=1}^n \lambda_i = 0$.

Consider τ_2. Since G is square-free, we must have $z_1 \to z_2$ or $z_2 \to z_1$.

$$\begin{tikzcd} & z_2 & \\ x \arrow[ur] \arrow[dr] & & y \arrow[ul] \arrow[dl] \\ & z_1 \arrow[uu] & \end{tikzcd}$$

In any case, $b_2 = [x,y,z_1,z_2]$ is a simplex with sink z_1 or z_2. Its boundary is

$$[y,z_1,z_2] - [x,z_1,z_2] + \tau_2 - \tau_1.$$

Notice that $[y,z_1,z_2]$ and $[x,z_1,z_2]$ are triangles.

Then

$$\lambda_2\tau_2 + \partial_3^{\mathrm{Dow}}(-\lambda_2 b_2) = \lambda_2\tau_1 + \text{triangles}.$$

In a similar manner, consider τ_i and define $b_i = [x,y,z_1,z_i]$, for $i = 3,\ldots,n$. Then

$$\lambda_i\tau_i + \partial_3^{\mathrm{Dow}}(-\lambda_i b_i) = \lambda_i\tau_1 + \text{triangles}.$$

Thus, $\sum_{i=1}^n \lambda_i\tau_i$ is homologous to

$$\sum_i \lambda_i\tau_i + \partial_3^{\mathrm{Dow}}\left(\sum_{i=2}^n -\lambda_i b_i\right) = \lambda_1\tau_1 + \sum_{i=2}^n \lambda_i\tau_1 + \text{triangles} = \left(\sum_i \lambda_i\right)\tau_1 + \text{triangles}$$

and $\sum_i \lambda_i = 0$. ■

Now to finish the proof of the injectivity: we have $\phi(v) = w' = \sum_j \alpha_j \tau'_j$ where each τ'_j is a triangle of the form $\tau'_j = [x_j, y_j, z_j]$ with $x_j \to y_j$, $y_j \to z_j$ and $x_j \to z_j$. Then

$$\phi(v) = \partial_2^{\mathrm{Dow}}(w') = \sum_j \alpha_j([y_j, z_j] - [x_j, z_j] + [x_j, y_j]) =$$
$$\sum_j \alpha_j(\phi(y_j, z_j) - \phi(x_j, z_j) + \phi(x_j, y_j)) = \sum_j \alpha_j \phi((y_j, z_j) - (x_j, z_j) + (x_j, y_j)) =$$
$$\sum_j \alpha_j(\phi \circ \partial_2^{\Xi}(x_j, y_j, z_j)) = \phi\left(\partial_2^{\Xi}\left(\sum_j \alpha_j(x_j, y_j, z_j)\right)\right).$$

Call $T = \partial_2^{\Xi}\left(\sum_j \alpha_j(x_j, y_j, z_j)\right)$. By the above, $v - T \in \ker\phi$.

Claim 5.25.2. $\ker\phi = \langle B\rangle$, that is, is generated by

$$B = \{(a, b) + (b, a) \mid (a, b), (b, a) \in \Xi_1(G)\}.$$

Furthermore, $\langle B\rangle \subset \partial_2^{\Xi}(\Xi_2(G))$.

Proof. We have $\phi((a, b) + (b, a)) = [a, b] + [b, a] = [a, b] - [b, a] = 0$, so the vector space generated by B is in $\ker\phi$. For the other inclusion, consider $k = \sum \alpha_i v_i$ not in $\langle B\rangle$. Call $v_1 = (a, b)$. If $b \nrightarrow a$, then $\phi(k)$ has the term $\lambda_i[a, b]$ which doesn't cancelate, thus $\phi(k) \neq 0$; now if $v_j = (b, a)$ for some j, then

$$\phi(k) = 0 \Rightarrow (\lambda_1 - \lambda_j)[a, b] = 0 \Rightarrow \lambda_1 = \lambda_j.$$

Then $\lambda_1 v_1 + \lambda_j v_j \in \langle B\rangle$. Proceed with the same reasoning for every v_i to conclude that $k \in \langle B\rangle$.

Further, the generators $(a, b) + (b, a)$ of B are the boundary of (a, b, a). This proves claim 2. □

Finally, since $v - T \in \ker\phi$, then $v = T + k$ with $k \in \langle B\rangle$. This shows that $v \in \partial_2^{\Xi}(\Xi_2(G))$, that is: $[v] = 0$. Hence, ϕ is injective. ■

The remaining of the proof doesn't use the hypothesis of the field being finite. □

Theorem 5.26. Let $\mathbf{X} \in \mathcal{N}$ be a symmetric network. Then

$$\mathrm{dgm}_1^{\Xi}(\mathbf{X}) = \mathrm{dgm}_1^{\mathrm{Dow}}(\mathbf{X}).$$

Proof. Using the same notation as in the previous proof, let's prove that ϕ_* is injective.

Let $[v] \in H_1^{\Xi}(G)$ and suppose that $\phi_*[v] = 0$. Then there is $w = \sum_i \lambda_i \tau_i \in C_2^{\mathrm{Dow}}(G)$ such that $\partial_2^{\mathrm{Dow}}(w) = \phi(v)$.

We will proceed like before and show that $v - b \in \ker\phi$, with $b \in \partial_2^{\Xi}(\Xi_2(G))$. This will imply $v \in \partial_2^{\Xi}(\Xi_2(G))$.

If $\tau_1 = [x, y, z_1]$ with sink z_1 is a triangle, we are ok. If it is not, then $x \nrightarrow y$ and $y \nrightarrow x$. Call $\Lambda = \{i \mid [x, y]$ is a face of $\tau_i\}$. Suppose $\Lambda = \{1, \dots, n\}$ and $\tau_i = [x, y, z_i]$, $i = 1, \dots, n$, with z_i as a sink. Then the terms $[x, y]$ must cancelate each other in $\partial_2^{\Xi}(w)$, and that implies $\sum_{i=1}^n \lambda_i = 0$.

By the symmetry, we also have the arrows $z_j \to x$ and $z_j \to y$, $j = 1, 2$.

$$\begin{tikzcd}
& z_2 \arrow[dl, leftrightarrow] \arrow[dr, leftrightarrow] & \\
x \arrow[dr, leftrightarrow] & & y \arrow[dl, leftrightarrow] \\
& z_1 &
\end{tikzcd}$$

Define $a_i = (x, z_1, y) - (x, z_i, y) \in \Xi_2(G)$, for $i = 2, \ldots, n$. Notice that $\phi \circ \partial_2^{\Xi}(-a_i) = \partial_2^{\mathrm{Dow}}(\tau_i) - \partial_2^{\mathrm{Dow}}(\tau_1)$. Thus,

$$\sum_{i=1}^{n} \lambda_i \tau_i - \sum_{i=2}^{n} \lambda_i(\tau_i - \tau_1) = \left(\sum_{i=1}^{n} \lambda_i\right) \tau_1 = 0 \Rightarrow \sum_{i=1}^{n} \lambda_i \tau_i = \sum_{i=2}^{n} \lambda_i(\tau_i - \tau_1).$$

Then,

$$\partial_2^{\mathrm{Dow}}\left(\sum_{i=1}^{n} \lambda_i \tau_i\right) = \partial_2^{\mathrm{Dow}}\left(\sum_{i=2}^{n} \lambda_i(\tau_i - \tau_1)\right) = \sum_{i=2}^{n} \lambda_i \phi(\partial_2^{\Xi}(a_i)) = \phi\left(\partial_2^{\Xi}\left(\sum_{i=2}^{n} \lambda_i a_i\right)\right).$$

Proceed like above to deal with all τ_i.

This proves that $\phi(v) = \partial_2^{\mathrm{Dow}}(w) = \phi(\partial_2^{\Xi}(T + S))$, where T is a sum of triangles and S is a sum of squares like a_i above. Thus, $v \in \partial_2^{\Xi}(\Xi_2(G))$ and $[v] = 0$. This proves de injectivity.

The rest of the proof assume the field $\mathbb{K}$ to be finite. $\square$

Corollary 5.27. Let $\mathbf{X}, \mathbf{Y} \in \mathcal{N}$ be symmetric networks. Then

$$\begin{aligned}\mathrm{dgm}_1^{\mathrm{Dow}}(\mathbf{X} \rtimes\!\!\!\ltimes \mathbf{Y}) = \mathrm{dgm}_1^{\Xi}(\mathbf{X} \rtimes\!\!\!\ltimes \mathbf{Y}) = \\ \bigoplus_{\substack{p,q \geq 0 \\ p+q=1}} (\mathrm{dgm}_p^{\Xi}(\mathbf{X}) \otimes \mathrm{dgm}_q^{\Xi}(\mathbf{Y})) = \bigoplus_{\substack{p,q \geq 0 \\ p+q=1}} (\mathrm{dgm}_p^{\mathrm{Dow}}(\mathbf{X}) \otimes \mathrm{dgm}_q^{\mathrm{Dow}}(\mathbf{Y})).\end{aligned}$$

Example 5.28. Let G be the following graph:

$$\begin{tikzcd}
a \arrow[r, leftrightarrow] \arrow[d, leftrightarrow] & b \arrow[d, leftrightarrow] \\
d \arrow[r, leftrightarrow] & c
\end{tikzcd}$$

that is, G is isomorphic to $K_2 \square K_2$. Then $H_p^{\Xi}(G) = 0$ for $p \geq 1$, because G is *contractible* (see the notion of "contractible graph" in [76]). Notice that G is symmetric,

$$C_2^{\mathrm{Dow}}(G) = \{[a, b, c], [a, b, d], [a, c, d], [b, c, d]\}$$

and $C_3^{\mathrm{Dow}}(G) = \emptyset$, which implies that $H_2^{\mathrm{Dow}}(G) \simeq \mathbb{K} \neq H_2^{\Xi}(G) = 0$.

The above example also shows us that we can't expect a Künneth formula for the square product in Dowker homology, because

$$(H_2^{\mathrm{Dow}}(H) \otimes H_0^{\mathrm{Dow}}(H)) \oplus (H_1^{\mathrm{Dow}}(H) \otimes H_1^{\mathrm{Dow}}(H)) \oplus (H_0^{\mathrm{Dow}}(H) \otimes H_2^{\mathrm{Dow}}(H)) \simeq \mathbb{K} \oplus 0 \oplus \mathbb{K},$$

since $H_1^{\mathrm{Dow}}(H) = 0$ and $H_0^{\mathrm{Dow}}(H) \simeq \mathbb{K}$, while $H_2^{\mathrm{Dow}}(H \square H) = \mathbb{K}$.

6 Implementation and algorithms

There is no trap so deadly as the trap you set for yourself.

RAYMOND CHANDLER, *THE LONG GOODBYE*

In this chapter we will discuss algorithmic details and the complexity of some computations in the preceding work.

The ultranetwork of a graph g obtained from the sets of graphs $\Omega_1, \ldots, \Omega_n$ of Section 3.3 can be calculated in a simple way. The following code in R language [37] will demonstrate this. Comments are preceded by the # symbol. We assume the reader has a basic understanding of R and knows how to install packages (if some of them are not already installed).

First, we load some libraries to deal with graphs and plottings.

```
library(igraph); library(ggraph); library(ggdendro); library(patchwork)
```

The `graph_plot` function, as suggested by its name, plots a graph in a nice way.

```
graph_plot = function(g, l = "layout_with_fr"){
  if (is.null(V(g)$name)) label = V(g) else label = V(g)$name
  V(g)$name = letters[1:vcount(g)]
  ggraph(g, layout = do.call(what = l, args = list(g))) +
    geom_edge_link(arrow = arrow(length = unit(4, 'mm')),
                   end_cap = circle(3, 'mm')) +
    geom_node_point(color = "black", size = 6) +
    geom_node_text(label = label, size = 4, col = "white") +
    theme_void()
}
```

To compute the cited ultranetwork, we don't need to check if there are graph maps from some Ω_i to g. We can compute the "walking distance" d of g, as in Lemma 2.16: $d(v, v')$ is the minimum number of edges taken from all possible paths between v and v', that is,

$$d(v, v') = \min\{n \mid \exists \phi \colon L_n \to (g, v, v') \text{ graph map}\},$$

where the min is $+\infty$ if the set above is empty.

Thus, there is a cycle C_n containing v and v' when $n = d(v, v') + d(v', v)$.

```
f_omega_cycle = function(g){
  n = vcount(g) # the number of vertices of g
  d = distances(g, mode = "out") # the distance d
  M = (d + t(d)) - 1 # sum d with its transpose t(d) and subtract 1
```

```
  M = ifelse(test = (M == Inf), yes = n, no = M) # if the distance is infinity, we change it for n
  h = hclust(d = as.dist(d), method = "single") # create the dendrogram
  ggdendrogram(h, hang = -1, theme_dendro = TRUE, # plot it
  rotate = TRUE, segments = TRUE)
}
```

We can input the graph from Example 3.31 as follows:

```
g = graph_from_literal(a +-+ b -+ c -+ a, c -+ d)
```

where `a +-+ b` means $a \leftrightarrow b$, `b -+ c` means $b \longrightarrow c$ and so on. We create the plots

```
p1 = graph_plot(g)
p2 = f_omega_cycle(g)
```

and use the command

```
p1 | p2
```

to show them side by side, as in Figure 6.1.

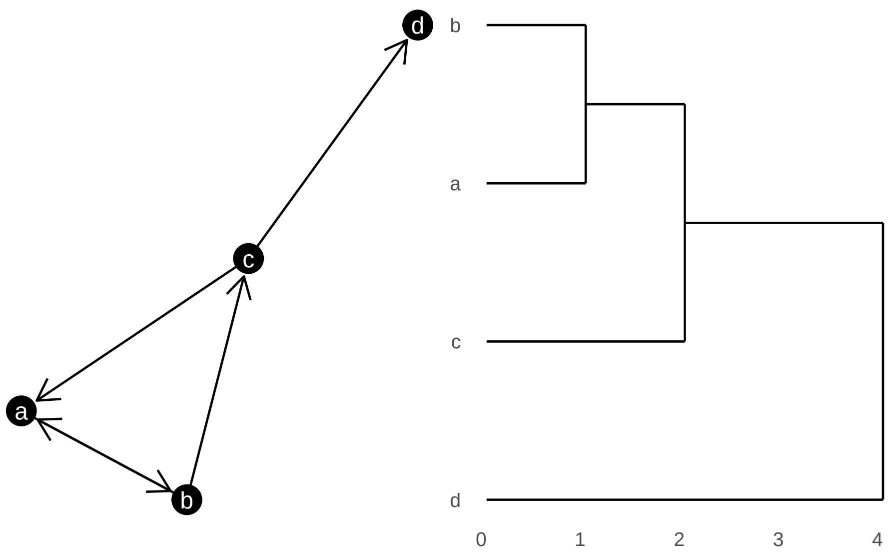


Figure 6.1: The graph from Example 3.31 made in R, with its associated ultranetwork.

We can also create a random Erdos-Renyi graph with the command `random.graph.game` from the excellent package igraph [87] and plot it as in Figure 6.2.

```
g = random.graph.game(n = 12,           # the number of vertices
                      p.or.m = 0.2,     # the probability of having an arrow between two vertices
                      type = "gnp",     # the type of the graph
                      directed = TRUE) # directed or symmetric
p1 = graph_plot(g)
p2 = f_omega_cycle(g)

p1 + p2
```

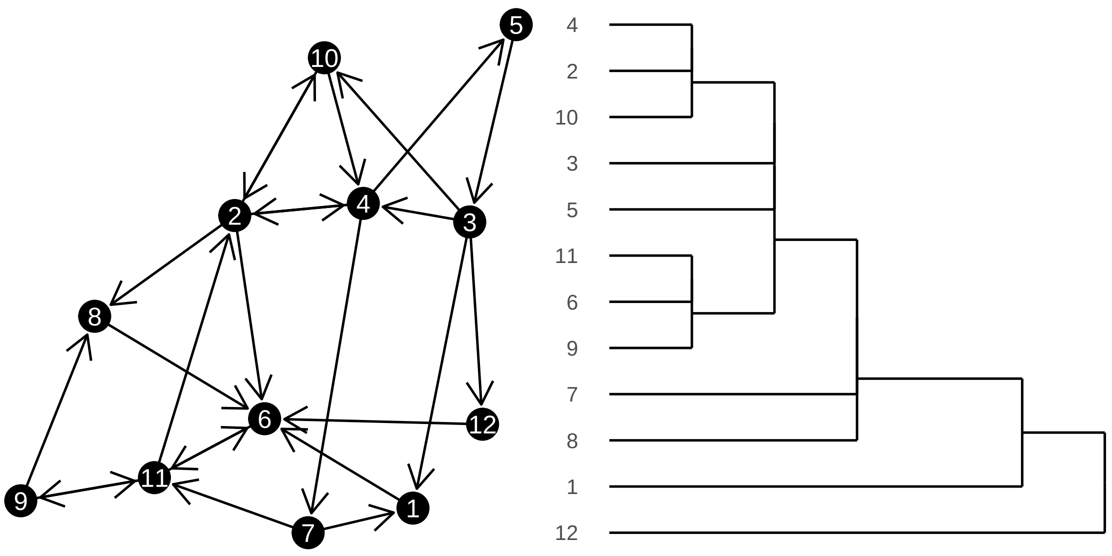


Figure 6.2: A random graph and its associated ultranetwork. Points that cluster earlier are more inter-connected.

For the general case where the sets Ω_i are arbitrary, we need to check if a given graph $\omega \in \Omega$ can be "fitted" (through a graph map) to g. This is a more general case than the "subgraph isomorphism problem", which is known to be a NP-complete problem [88]. These problems are hard to solve (see [89]) in the sense that no polynomial-time algorithm for them is known. The time spent to solve the graph isomorphism grows exponentially as the number of nodes and arrows increase.

In what follows, we will consider a "brute force" approach to compute the motivic homology of a graph.

```
library(magrittr) # library to use the pipe operator %>%
set.seed(1)       # to allow reproducibility
# generate a random graph
g = random.graph.game(n = 10,
                      p.or.m = 0.2,
                      type = "gnp",
                      directed = TRUE)
graph_plot(g)         # plot it
omega = graph.star(3) # set omega to be S_3
# check all possible injective graph maps from omega to g
l = subgraph_isomorphisms(pattern = omega, target = g)
# list the vertices of the 2-simplexes of C_2^Dow(G)
l %>% unlist %>% matrix(ncol = 3, byrow = TRUE)
```

The result is the following plot and a matrix, where each line corresponds to a 2-simplex of $C_2^{\mathrm{Dow}}(g)$:

```
     [,1] [,2] [,3]
[1,]    4    1    8
[2,]    1    3    6
[3,]    9    4    7
[4,]    9    4    8
```

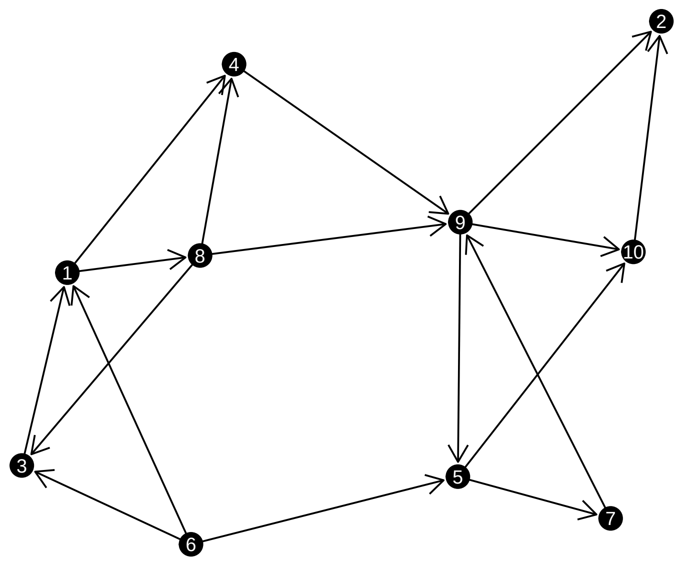


Figure 6.3: A random graph g and the computation of $C_2^{\mathrm{Dow}}(g)$.

```
 [5,]   10    5    9
 [6,]    1    6    3
 [7,]    3    6    8
 [8,]    5    6    9
 [9,]    9    7    8
[10,]    9    7    4
[11,]    9    8    7
[12,]    9    8    4
[13,]    3    8    6
[14,]    4    8    1
[15,]    5    9    6
[16,]   10    9    5
[17,]    2    9   10
[18,]    2   10    9
```

Notice that, for example, $\{4, 1, 8\}$ is indeed a 2-simplex with sink 4.

We can use any graphs `omega` and `g` as parameters to the `subgraph_isomorphisms` function to obtain a list of simplexes. We can store and input these lists into the Perseus software ([90]) directly from R to calculate its homology (and persistent homology).

It is important now to distinguish between two types of families Ω, which we will call *local* and *global* (as suggested by Samir Chowdhury in one of our numerous discussions).

Definition 6.1. We say that a family Ω of graphs is *local* if for any graph $G = (V, E)$, in order to check whether $\sigma = \{v_0, \ldots, v_n\}$ is a simplex of $C^{\Omega}(G)$ (see Definition 4.1), it is enough to consider the subgraph $G \cap \sigma = (\sigma, E \cap (\sigma \times \sigma))$.

To rephrase the definition above more precisely, consider

$$L^{\Omega}(\sigma; G) = \min\{||\omega| - |\sigma|| \mid \phi\colon \omega \to G \text{ is a graph map s.t. } \sigma \subseteq \mathrm{Im}(\phi), \omega \in \Omega\},$$

that is, we compute the minimum number of points needed to cover σ by a graph map

$\phi\colon \omega \to G$. Next, define

$$L(\Omega) = \sup\{L^{\Omega}(\sigma; G);\ G \in \mathcal{G},\ \sigma \in C^{\Omega}(G)\}.$$

Thus, Ω is local (or 0-global) if $L(\Omega) = 0$, and we say that Ω is $L(\Omega)$-global otherwise.

Examples

1. The set Ω^{Rips} is clearly local, since in order to check whether $\sigma = \{v_0, \ldots, v_n\} \in C^{\Omega^{\text{Rips}}}(G)$, we only check if $v_i \xrightarrow[=]{G} v_j$, for all i, j.

2. The set Ω^{Dow} is 1-global. Indeed, to check if $\sigma = \{v_0, \ldots, v_n\} \in C^{\Omega^{\text{Dow}}}(G)$, we only need to obtain one more point $\tilde{v}$ such that $v_i \xrightarrow[=]{G} \tilde{v}$, for all i.

3. The set $\Omega = \{L_n\}_{n\geq 1}$ is ∞-global. To see this, let $n \geq 2$ be any integer and consider $G = L_n$ with vertex set $V = \{v_1, \ldots, v_n\}$. Then clearly $\{v_1, v_n\} \in C^{\Omega}(G)$, but any map from $\omega \in \Omega$ to G containing $\{v_1, v_n\}$ in its image, also contains V, that is, $L^{\Omega}(\sigma; G) \geq n - 2$. Since n is arbitrary, then $L(\Omega) = \infty$.

Remark 6.2. The number $L(\Omega)$ measures, roughly speaking, how far from being local a family Ω is. Knowing that a family Ω is n-local restricts the graphs $\omega \in \Omega$ that we need to insert in the `pattern` argument of the function `subgraph_isomorphisms`. For example, to compute the 1-dimensional homology of $C^{\Omega^{\text{Dow}}}(G)$ we need to know $C_i^{\Omega^{\text{Dow}}}(G)$ for $i = 0, 1, 2$. To do this, we have to check all graph maps from S_i to G, for $i = 2, 3, 4$, where $S_i = D_{i-1} * D_1$ (see Example 4.6). In general, if Ω is n-global, to compute $C_i^{\Omega}(G)$ up to $i = p$, we need to compute all possible graph maps from ω to G, where $|\omega| \leq p + n + 1$.

As an example, let's compute the Dowker homology of the graph of Figure 6.2, in dimensions 0, 1 and 2.

First, create a folder called `example` and download the Perseus software from http://people.maths.ox.ac.uk/nanda/perseus into this folder. We warn the reader that some commands may vary according to the operational system. We are using Ubuntu 18.04.

Start a new Rscript and write in it the previous `library` commands and the `graph_plot` function:

```
library(igraph); library(ggraph); library(ggdendro);
library(patchwork); library(magrittr)

graph_plot = function(g, l = "layout_with_fr") {
  if (is.null(V(g)$name)) label = V(g) else label = V(g)$name
  V(g)$name = letters[1:vcount(g)]
  ggraph(g, layout = do.call(what = l, args = list(g))) +
    geom_edge_link(arrow = arrow(length = unit(4, 'mm')),
                   end_cap = circle(3, 'mm')) +
    geom_node_point(color = "black", size = 6) +
    geom_node_text(label = label, size = 4, col = "white") +
    theme_void()
}
```

Save this script in the `example` folder as, let's say, `calculate_graph_homology.R`. Continue editing it, and add the following lines

```r
# set the file path to be the working directory
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))

set.seed(1)
g = random.graph.game(n = 10, p.or.m = 0.2, type = "gnp", directed = TRUE)
```

to change the working directory and to obtain again the graph g above. Since we want to compute the Dowker homology up to dimension 2, we need to compute $C_i^\Omega(g)$ for $i = 0, \ldots, 3$. To do this, first note that $C_0^\Omega(g)$ is the set of vertices. The Dowker family $\{S_n\}_{n\geq 2}$ (see Example 4.6) is 1-global, so we only need to compute the graph maps from S_i to g for $i = 2, \ldots, 5$. Let's store these graphs in a list:

```r
omega = list()
for (i in 2:5) {
  omega[[i]] = graph.star(i)
}
```

and check that indeed these are the graphs we want:

```r
omega[[2]] %>% graph_plot()
omega[[3]] %>% graph_plot()
omega[[4]] %>% graph_plot()
omega[[5]] %>% graph_plot()
```

Now, let's store the simplexes as a list. Remember that R indices start at 1 (not 0), so $C_0^\Omega(g)$ is `C[[1]]`, and so on.

```r
C = list()

# C[[1]] is just the vertex set of g
C[[1]] = V(g) %>% unlist %>% matrix(ncol = 1)

for (i in 2:5) {
  l = subgraph_isomorphisms(pattern = omega[[i]], target = g)
  if (length(l) > 0) {
    C[[i]] = l %>% unlist %>% matrix(ncol = i, byrow = TRUE)
  }
}
```

Before continue, let's make some comments on Perseus input file: it is a text file where the first line is just the number 1, which indicates that we are inputting natural numbers to denote the vertices of the filtration. The next lines are of the form

$$n \; v_0 \; v_1 \; \ldots \; v_n \; t$$

where n is the dimension of the simplex, $v_0, \ldots, v_n$ are the vertices and t is the time that this simplex appears in the simplicial filtration. Since we are computing the homology of a single simplicial complex, we can set t equal to 1 everywhere.

The following code will create this text file as required by Perseus:

```r
f = tempfile(pattern = "tempfile") # create a temp file
write(x = 1, file = f)             # write 1 in the first line

# write the vertices as required
M = list()
for (i in 1:length(C)) {
  if (length(C[[i]]) > 0) {
    M[[i]] = cbind(i - 1, C[[i]], 1, deparse.level = 0)
    write(x = t(M[[i]]), file = f,
          append = TRUE, ncolumns = ncol(M[[i]]))
  }
}
```

If you have `gedit` installed, you can visualize this temp file with the command

```r
system2(command = "gedit", args = f)
```

or go to the directory indicated by the variable `f` and open the text file with any other text editor. The beginning of the text file (single column) should be like in Figure 6.4.

```
1
0 1 1
0 2 1
0 3 1
0 4 1
0 5 1
0 6 1
0 7 1
0 8 1
0 9 1
0 10 1
```

```
1 8 1 1
1 4 1 1
1 1 3 1
1 9 4 1
1 10 5 1
1 7 5 1
1 1 6 1
1 3 6 1
1 5 6 1
1 9 7 1
1 9 8 1
```

```
1 3 8 1
1 4 8 1
1 5 9 1
1 10 9 1
1 2 9 1
1 2 10 1
2 4 1 8 1
```

Figure 6.4: The Perseus input.

Next, we execute Perseus to compute the homology:

```r
system2(command = "./perseusLin",
        args = paste("nmfsimtop",  f, "homology"))
```

If everything went well, you shall see the message in Figure 6.5.

Now, in the `example` folder there are some text files whose names start with “homology”. Opening the `homology_betti.txt` file, there will be a single line with the numbers

$$1\ 1\ 3$$

which means that at time 1 (the parameter where all simplexes appeared in the filtration), the 0-dimensional homology has 1 generator and the 1-dimensional homology has 3 generators, that is, $H_0^\Omega(g) = \mathbb{K}$, $H_1^\Omega(g) = \mathbb{K}^3$ and $H_2^\Omega(g) = 0$, where $\mathbb{K} = \mathbb{Z}_2$, the field with 2 elements. Perseus only computes homology with $\mathbb{K} = \mathbb{Z}_2$.

To compute the persistent diagram $\mathrm{dgm}_i^\Omega(\mathbf{X})$ of a network $\mathbf{X}$, we proceed as above to obtain $C^\Omega(\mathbf{X}_\epsilon)$ for all ϵ, and then insert this filtration into Perseus, with the corresponding

```
Read 37 top simplices from Input File
Writting Cell Complex From Non-manifold simplicial Complex
Done! Complex stored with 42 cells!
+++coreductions:  42 --> 4, fraction removed 0.904762 at height 1
+++reductions:    4 --> 4, fraction removed 0 at height 2
+++coreductions:  4 --> 4, fraction removed 0 at height 3
Computing Persistence Intervals!
Linearly ordered 4 cells...

Done!!! Please consult [homology*.txt] for results.
```

Figure 6.5: The Perseus output.

time of appearance of the simplexes. The full code with some examples will be available at github.com/vituri, but the brave reader shall be able to compute it using the above examples.

7 Future work

Among the things Billy Pilgrim could not change were the past, the present, and the future.

KURT VONNEGUT, *SLAUGHTERHOUSE FIVE*

While writing this thesis, we had much more questions than answers. Some of them are listed below, to guide future works.

Open questions concerning Chapter 2

1. Let $\mathfrak{F}$ be an symmetric endofunctor such that $\mathfrak{F} \circ \mathfrak{F} = \mathfrak{F}$ (such functors are called *idempotent)*. Is $\mathfrak{F}$ representable? A related question was answered in [8].

Open questions concerning Chapter 4

1. Are the usual path homology and the path homology with repetitions isomorphic in all dimensions?

2. Let Ω^{ord} be such that $K = OC^{\Omega^{\text{ord}}}(G)$ satisfies (P1) for every $G \in \mathcal{G}$. Is it possible to remove the "adjacent repetitions" like the path homology does (removing the irregular sequences), and still obtain something stable? What changes?

3. The original ordered clique complex was useful to study neural networks. What are other interesting sets Ω^{ord} that capture the structure of such networks?

4. What is so special about path homology that makes it have a homotopy theory compatible with its homology? Is is it possible that every family Ω^{ord} that satisfies (P1) gives rise to a homotopy theory, after some "tricks" (like deleting the irregular elements) at the algebraic level?

5. How far is the number

$$\sup_{\Omega \text{ family of graphs}} d_b(\text{dgm}_p^{\Omega}(\mathbf{X}), \text{dgm}_p^{\Omega}(\mathbf{Y}))$$

   from $2d_{\mathcal{N}_{\text{ext}}}(\mathbf{X}, \mathbf{Y})$?

   Putting it differently: given $\mathbf{X}, \mathbf{Y} \in \mathcal{N}_{\text{ext}}$, can we find Ω such that the bottleneck distance between its induced persistent diagrams is as close as possible from the

Gromov-Hausdorff distance between $\mathbf{X}$ and $\mathbf{Y}$? A version of this problem was studied in [91, Proposition 3.1], where the authors proved that, under some conditions, the bottleneck distance is strictly smaller than the network distance.

6. If $H_\bullet^{\Omega_1}(G) \simeq H_\bullet^{\Omega_2}(G)$ for every $G \in \mathcal{G}$, what can we say about Ω_1 and Ω_2? In the case of Dowker sink and Dowker source, the generators of one were the generators of the other after reversing all the arrows.

7. Given a family of ordered graphs Ω^{ord}, how can we create a related family Ω_0^{ord} of ordered graphs such that $OC^{\Omega_0^{\mathrm{ord}}}(G)$ satisfies property (P1)? That is: what do we need to do with Ω^{ord} in order to turn it into a family of graphs that satisfy the Stability Theorem 4.19?

8. Given an endofunctor $\mathfrak{F}$, a graph G and a family Ω of graphs, how $C^{\Omega}(\mathfrak{F}(G))$ and $C^{\mathfrak{F}(\Omega)}(G)$ are related, where $\mathfrak{F}(\Omega) = \{\mathfrak{F}(\omega) \mid \omega \in \Omega\}$?

Open questions concerning Chapter 5

1. Is it true that $H_p^{\Xi}(G \,\square\, G') \simeq H_p^{\Xi}(G \simeq G')$ for $p \geq 2$ and all graphs G?

2. What conditions do we need to impose on a family Ω of graphs in order to have

$$C^{\Omega}(G \rtimes G') = C^{\Omega}(G) \times C^{\Omega}(G'),$$

for any graphs G and G', as in Definition 5.14? For any family with this property, a Künneth formula like the one in Theorem 5.17 will hold.

Open questions concerning Chapter 6

1. For an arbitrary family Ω_1 of graphs, we saw that calculating $C^{\Omega_1}(G)$ is hard. Can we find Ω_2 such that $C^{\Omega_2}(G)$ is easier to calculate, and such that $d_b(\mathrm{dgm}_p^{\Omega_2}(\mathbf{X}), \mathrm{dgm}_p^{\Omega_2}(\mathbf{Y})$ is close to $d_b(\mathrm{dgm}_p^{\Omega_1}(\mathbf{X}), \mathrm{dgm}_p^{\Omega_1}(\mathbf{Y}))$, for arbitrary networks $\mathbf{X}$ and $\mathbf{Y}$? More generally: can we approximate, in some sense, the computation of $C^{\Omega_1}(G)$ in a faster way?

Bibliography

[1] J. W. von Goethe, *Fausto*. Livraria Clássica Editora, 1919.

[2] ——, *Faust*. Dover Children's, 2013.

[3] R.ui Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, May 2005.

[4] G. Carlsson and F. Mémoli, "Classifying Clustering Schemes," *arXiv e-prints*, p. arXiv:1011.5270, Nov 2010.

[5] G. Carlsson and F. Mémoli, "Characterization, stability and convergence of hierarchical clustering methods." *Journal of Machine Learning Research*, vol. 11, pp. 1425–1470, Apr04 2010.

[6] J. M. Kleinberg, "An impossibility theorem for clustering," in *Advances in Neural Information Processing Systems 15*, S. Becker, S. Thrun, and K. Obermayer, Eds. MIT Press, 2003, pp. 463–470. [Online]. Available: http://papers.nips.cc/paper/2340-an-impossibility-theorem-for-clustering.pdf

[7] G. Carlsson, F. Mémoli, A. Ribeiro, and S. Segarra, "Axiomatic Construction of Hierarchical Clustering in Asymmetric Networks," *arXiv e-prints*, p. arXiv:1301.7724, Jan 2013.

[8] G. E. Carlsson, F. Mémoli, A. R. Ribeiro, and S. Segarra, "Excisive hierarchical clustering methods for network data," *ArXiv*, vol. abs/1607.06339, 2016.

[9] J. R. Munkres, *Elements of Algebraic Topology*. Addison Wesley Publishing Company, 1984. [Online]. Available: http://www.worldcat.org/isbn/0201045869

[10] R. Atkin, *Multidimensional man*, ser. Penguin books. Penguin Books, 1981. [Online]. Available: https://books.google.com.br/books?id=ywfXAAAAMAAJ

[11] P. J. Hilton and S. Wylie, *Homology Theory: An Introduction to Algebraic Topology*. Cambridge University Press, 1967. [Online]. Available: https://books.google.com.br/books?id=Y-w3AAAAIAAJ

[12] S. MacLane, *Categories for the Working Mathematician*. New York: Springer-Verlag, 1971, Graduate Texts in Mathematics.

[13] F. Chazal and B. Michel, "An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists," *arXiv e-prints*, p. arXiv:1710.04019, Oct 2017.

[14] F. Chazal, S. Y. Oudot, M. Glisse, and V. De Silva, *The Structure and Stability of Persistence Modules*, ser. SpringerBriefs in Mathematics. Springer Verlag, 2016. [Online]. Available: https://hal.inria.fr/hal-01330678

[15] R. Ghrist, "Barcodes: the persistent topology of data," *Bulletin of the American Mathematical Society*, vol. 45, no. 1, pp. 61–75, 2008.

[16] S. Y. Oudot, *Persistence Theory: From Quiver Representations to Data Analysis*, ser. Mathematical Surveys and Monographs. American Mathematical Society, 2015, no. 209. [Online]. Available: https://hal.inria.fr/hal-01247501

[17] R. R. Wadhwa, D. F. K. Williamson, A. Dhawan, and J. G. Scott, "TDAstats: R pipeline for computing persistent homology in topological data analysis," *Journal of Open Source Software*, vol. 3, no. 28, p. 860, 2018. [Online]. Available: https://doi.org/10.21105/joss.00860

[18] S. Chowdhury and F. Mémoli, "A functorial Dowker theorem and persistent homology of asymmetric networks," *Journal of Applied and Computational Topology*, vol. 2, no. 1, pp. 115–175, Oct 2018.

[19] J. A. Perea, "A brief history of persistence," *arXiv preprint arXiv:1809.03624*, 2018.

[20] P. Frosini, "Measuring shapes by size functions," in *Intelligent Robots and Computer Vision X: Algorithms and Techniques*, vol. 1607. International Society for Optics and Photonics, 1992, pp. 122–133.

[21] A. Cerri, M. Ferri, and D. Giorgi, "Retrieval of trademark images by means of size functions," *Graphical Models*, vol. 68, no. 5-6, pp. 451–471, 2006.

[22] F. Chazal, D. Cohen-Steiner, L. J. Guibas, F. Mémoli, and S. Y. Oudot, "Gromov-Hausdorff stable signatures for shapes using persistence," in *Computer Graphics Forum*, vol. 28, no. 5. Wiley Online Library, 2009, pp. 1393–1403.

[23] R. W. Sumner and J. Popović, "Deformation transfer for triangle meshes," *ACM Transactions on graphics (TOG)*, vol. 23, no. 3, pp. 399–405, 2004.

[24] G. Plonka and Y. Zheng, "Relation between total variation and persistence distance and its application in signal processing," *Advances in Computational Mathematics*, vol. 42, no. 3, pp. 651–674, 2016.

[25] G. Carlsson, T. Ishkhanov, V. De Silva, and A. Zomorodian, "On the local behavior of spaces of natural images," *International journal of computer vision*, vol. 76, no. 1, pp. 1–12, 2008.

[26] F. Chazal, L. J. Guibas, S. Y. Oudot, and P. Skraba, "Persistence-based clustering in riemannian manifolds," *Journal of the ACM (JACM)*, vol. 60, no. 6, p. 41, 2013.

[27] S. Paris and F. Durand, “A topological approach to hierarchical segmentation using mean shift,” in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8.

[28] M. Carriere and S. Oudot, “Structure and stability of the one-dimensional mapper,” *Foundations of Computational Mathematics*, vol. 18, no. 6, pp. 1333–1396, 2018.

[29] G. Singh, F. Mémoli, and G. E. Carlsson, “Topological methods for the analysis of high dimensional data sets and 3d object recognition.” in *SPBG*, 2007, pp. 91–100.

[30] P. Y. Lum, G. Singh, A. Lehman, T. Ishkanov, M. Vejdemo-Johansson, M. Alagappan, J. Carlsson, and G. Carlsson, “Extracting insights from the shape of complex data using topology,” *Scientific reports*, vol. 3, p. 1236, 2013.

[31] J. L. Nielson, J. Paquette, A. W. Liu, C. F. Guandique, C. A. Tovar, T. Inoue, K.-A. Irvine, J. C. Gensel, J. Kloke, T. C. Petrossian *et al.*, “Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury,” *Nature communications*, vol. 6, no. 1, pp. 1–12, 2015.

[32] M. Nicolau, A. J. Levine, and G. Carlsson, “Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 17, pp. 7265–7270, 2011.

[33] T. K. Dey, F. Mémoli, and Y. Wang, “Multiscale mapper: topological summarization via codomain covers,” in *Proceedings of the twenty-seventh annual acm-siam symposium on discrete algorithms*. SIAM, 2016, pp. 997–1013.

[34] ——, “Topological analysis of nerves, reeb spaces, mappers, and multiscale mappers,” *arXiv preprint arXiv:1703.07387*, 2017.

[35] E. Munch and B. Wang, “Convergence between categorical representations of reeb space and mapper,” *arXiv preprint arXiv:1512.04108*, 2015.

[36] V. De Silva, E. Munch, and A. Patel, “Categorified reeb graphs,” *Discrete & Computational Geometry*, vol. 55, no. 4, pp. 854–906, 2016.

[37] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013. [Online]. Available: http://www.R-project.org/

[38] P. Bubenik, “Statistical topological data analysis using persistence landscapes,” *The Journal of Machine Learning Research*, vol. 16, no. 1, pp. 77–102, 2015.

[39] V. Kovacev-Nikolic, P. Bubenik, D. Nikolić, and G. Heo, “Using persistent homology and dynamical distances to analyze protein binding,” *Statistical applications in genetics and molecular biology*, vol. 15, no. 1, pp. 19–38, 2016.

[40] A. Garin and G. Tauzin, “A topological" reading" lesson: Classification of mnist using tda,” *arXiv preprint arXiv:1910.08345*, 2019.

[41] J.-Y. Liu, S.-K. Jeng, and Y.-H. Yang, “Applying topological persistence in convolutional neural network for music audio signals,” *arXiv preprint arXiv:1608.07373*, 2016.

[42] M. Gameiro, Y. Hiraoka, S. Izumi, M. Kramar, K. Mischaikow, and V. Nanda, “A topological measurement of protein compressibility,” *Japan Journal of Industrial and Applied Mathematics*, vol. 32, no. 1, pp. 1–17, 2015.

[43] S. G. S. Calcina, “Topological data analysis: applications in machine learning,” Ph.D. dissertation, Universidade de São Paulo, 2018.

[44] G. Singh, F. Mémoli, T. Ishkhanov, G. Sapiro, G. Carlsson, and D. L. Ringach, “Topological analysis of population activity in visual cortex,” *Journal of vision*, vol. 8, no. 8, pp. 11–11, 2008.

[45] Y. Wang, H. Ombao, M. K. Chung *et al.*, “Persistence landscape of functional signal and its application to epileptic electroencaphalogram data,” *ENAR Distinguished Student Paper Award*, 2014.

[46] M. K. Chung, P. Bubenik, and P. T. Kim, “Persistence diagrams of cortical surface data,” in *International Conference on Information Processing in Medical Imaging*. Springer, 2009, pp. 386–397.

[47] J. Gamble and G. Heo, “Exploring uses of persistent homology for statistical analysis of landmark-based shape data,” *Journal of Multivariate Analysis*, vol. 101, no. 9, pp. 2184–2199, 2010.

[48] G. Heo, J. Gamble, and P. T. Kim, “Topological analysis of variance and the maxillary complex,” *Journal of the American Statistical Aassociation*, vol. 107, no. 498, pp. 477–492, 2012.

[49] M. W. Reimann, M. Nolte, M. Scolamiero, K. Turner, R. Perin, G. Chindemi, P. Dłotko, R. Levi, K. Hess, and H. Markram, “Cliques of neurons bound into cavities provide a missing link between structure and function,” *Frontiers in computational neuroscience*, vol. 11, p. 48, 2017.

[50] S. Chowdhury, T. Gebhart, S. Huntsman, and M. Yutin, “Path homologies of deep feedforward networks,” *arXiv preprint arXiv:1910.07617*, 2019.

[51] K. J. Emmett and R. Rabadan, “Characterizing scales of genetic recombination and antibiotic resistance in pathogenic bacteria using topological data analysis,” in *International Conference on Brain Informatics and Health*. Springer, 2014, pp. 540–551.

[52] J. M. Chan, G. Carlsson, and R. Rabadan, “Topology of viral evolution,” *Proceedings of the National Academy of Sciences*, vol. 110, no. 46, pp. 18 566–18 571, 2013.

[53] J. A. Perea and J. Harer, “Sliding windows and persistence: An application of topological methods to signal analysis,” *Foundations of Computational Mathematics*, vol. 15, no. 3, pp. 799–838, 2015.

[54] J. A. Perea, “Persistent homology of toroidal sliding window embeddings,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 6435–6439.

[55] J. Perea, “Topological time series analysis,” *Notices of the American Mathematical SocietyOTICES OF THE AMERICAN MATHEMATICAL SOCIETY*, vol. 66, no. 5, 2019.

[56] D. Horak, S. Maletić, and M. Rajković, “Persistent homology of complex networks,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 03, p. P03034, 2009.

[57] G. Petri, M. Scolamiero, I. Donato, and F. Vaccarino, “Topological strata of weighted complex networks,” *PloS one*, vol. 8, no. 6, p. e66506, 2013.

[58] V. De Silva and R. Ghrist, “Homological sensor networks,” *Notices of the American Mmathematical Ssociety*, vol. 54, no. 1, 2007.

[59] J. Gao, L. J. Guibas, S. Y. Oudot, and Y. Wang, “Geodesic Delaunay triangulation and witness complex in the plane,” in *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2008, pp. 571–580.

[60] A. Dirafzoon, A. Bozkurt, and E. Lobaton, “Geometric learning and topological inference with biobotic networks,” *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 1, pp. 200–215, 2016.

[61] H. Edelsbrunner and J. Harer, *Computational topology: an introduction*. American Mathematical Soc., 2010.

[62] T. Sousbie, “The persistent cosmic web and its filamentary structure–Ii. Theory and implementation,” *Monthly Notices of the Royal Astronomical Society*, vol. 414, no. 1, pp. 350–383, 2011.

[63] T. Sousbie, C. Pichon, and H. Kawahara, “The persistent cosmic web and its filamentary structure–II. Illustrations,” *Monthly Notices of the Royal Astronomical Society*, vol. 414, no. 1, pp. 384–403, 2011.

[64] M. Kramar, A. Goullet, L. Kondic, and K. Mischaikow, “Persistence of force networks in compressed granular media,” *Physical Review E*, vol. 87, no. 4, p. 042207, 2013.

[65] J. Berwald, M. Gidea, and M. Vejdemo-Johansson, “Automatic recognition and tagging of topologically different regimes in dynamical systems,” *arXiv preprint arXiv:1312.2482*, 2013.

[66] P. Dotko, K. Hess, R. Levi, M. Nolte, M. Reimann, M. Scolamiero, K. Turner, E. Muller, and H. Markram, “Topological analysis of the connectome of digital reconstructions of neural microcircuits,” *arXiv preprint arXiv:1601.01580*, 2016.

[67] G. V. F. Pinto and F. Mémoli, "Motivic clustering schemes for directed graphs," *arXiv e-prints*, p. arXiv:2001.00278, Jan 2020.

[68] S. Chowdhury and F. Mémoli, "Distances and isomorphism between networks and the stability of network invariants," *CoRR*, vol. abs/1708.04727, 2017. [Online]. Available: http://arxiv.org/abs/1708.04727

[69] Z. Smith, S. Chowdhury, and F. Mémoli, "Hierarchical representations of network data with optimal distortion bounds," in *2016 50th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2016, pp. 1834–1838.

[70] S. Chowdhury and F. Mémoli, "Persistent path homology of directed networks," in *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, ser. SODA '18. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2018, pp. 1152–1169. [Online]. Available: http://dl.acm.org/citation.cfm?id=3174304.3175345

[71] A. V. Ivashchenko, "Contractible transformations do not change the homology groups of graphs," *Discrete Mathematics*, vol. 126, no. 1-3, pp. 159–170, 1994.

[72] B. Chen, S.-T. Yau, and Y.-N. Yeh, "Graph homotopy and Graham homotopy," *Discrete Mathematics*, vol. 241, no. 1-3, pp. 153–170, 2001.

[73] M. E. Talbi and D. Benayat, "Homology theory of graphs," *Mediterranean Journal of Mathematics*, vol. 11, no. 2, pp. 813–828, 2014.

[74] E. Babson, H. Barcelo, M. de Longueville, and R. Laubenbacher, "Homotopy theory of graphs," *Journal of Algebraic Combinatorics*, vol. 24, no. 1, pp. 31–44, Aug 2006. [Online]. Available: https://doi.org/10.1007/s10801-006-9100-0

[75] H. Barcelo, X. Kramer, R. Laubenbacher, and C. Weaver, "Foundations of a connectivity theory for simplicial complexes," *Advances in Applied Mathematics*, vol. 26, no. 2, pp. 97 – 128, 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0196885800907103

[76] A. Grigor'yan, Y. Lin, Y. Muranov, and S.-T. Yau, "Homotopy theory for digraphs," *Pure and Applied Mathematics Quarterly*, vol. 10, 07 2014.

[77] A. Dimakis and F. Muller-Hoissen, "Differential calculus and gauge theory on finite sets," *Journal of Physics A: Mathematical and General*, vol. 27, no. 9, p. 3159, 1994.

[78] A. Dimakis and F. Müller-Hoissen, "Discrete differential calculus: Graphs, topologies, and gauge theory," *Journal of Mathematical Physics*, vol. 35, no. 12, pp. 6703–6735, 1994.

[79] A. Grigor'yan, Y. Lin, Y. Muranov, and S.-T. Yau, "Cohomology of digraphs and (undirected) graphs," *Asian Journal of Mathematics*, vol. 19, pp. 887–932, Nov11 2015.

[80] P. Masulli and A. E. P. Villa, “The topology of the directed clique complex as a network invariant,” *SpringerPlus*, vol. 5, no. 1, p. 388, Mar 2016. [Online]. Available: https://doi.org/10.1186/s40064-016-2022-y

[81] A. A. Grigor’yan, Y. Lin, Y. V. Muranov, and S.-T. Yau, “Path complexes and their homologies,” *Fundamentalnaya i Prikladnaya Matematika*, vol. 21, no. 5, pp. 79–128, 2016.

[82] A. Grigor’yan, Y. Lin, Y. Muranov, and S.-T. Yau, “Homologies of path complexes and digraphs,” *arXiv e-prints*, p. arXiv:1207.2834, Jul 2012.

[83] K. Turner, “Rips filtrations for quasi-metric spaces and asymmetric functions with stability results,” *arXiv e-prints*, p. arXiv:1608.00365, Aug 2016.

[84] H. Gakhar and J. A. Perea, “Künneth formulae in persistent homology,” *arXiv preprint arXiv:1910.05656*, 2019.

[85] S.-T. Hu, *Introduction to homological algebra*. Holden-day, 1968.

[86] D. Kozlov, *Combinatorial algebraic topology*. Springer Science & Business Media, 2007, vol. 21.

[87] G. Csardi and T. Nepusz, “The igraph software package for complex network research,” *InterJournal*, vol. Complex Systems, p. 1695, 2006. [Online]. Available: http://igraph.org

[88] S. A. Cook, “The complexity of theorem-proving procedures,” in *Proceedings of the third annual ACM symposium on Theory of computing*, 1971, pp. 151–158.

[89] M. R. Garey and D. S. Johnson, *Computers and intractability*. freeman San Francisco, 1979, vol. 174.

[90] K. Mischaikow and V. Nanda, “Morse theory for filtrations and efficient computation of persistent homology,” *Discrete & Computational Geometry*, vol. 50, no. 2, pp. 330–353, 2013.

[91] S. Chowdhury, N. Clause, F. Mémoli, J. A. Sanchez, and Z. Wellner, “New families of simplicial filtration functors,” *arXiv preprint arXiv:1712.00696*, 2017.

Index