Finding responsible simplices for holes in Persistent Homology

Pia Pachinger

May 14, 2018

Contents

1 Introduction and motivation

1.1 Introduction to Topological Data Analysis and Persistent Homology

1.1.1 Topological Data Analysis

Topological data analysis (TDA) is, as its name says, the analysis of data with the help of the abstract notions of closeness and connectivity of topology. Homeomorphisms are an important part of topology, they describe whether the fundamental features of two different spaces are the same. If we think of the classical description of a homeomorphism forming a cup into a donut, one could suppose that holes form an important part in describing the topology of a space. And this is exactly what Homology does, homology groups describe the holes of different dimensions of a space. Homology uses algebraic methods in order to extract topological characteristics of a space. There is even a connection between homeomorphisms and homology groups. Homeomorphic topological spaces have isomorphic homology groups.

1.1.2 Persistent Homology

The challenge of extracting the essential shape of a discrete set of points is often addressed in the field of Topological Data Analysis. Persistent Homology analyses the topological features of spaces induced by discrete sets of data points. For doing so it makes use of a combination of sequences and homology. Persistent homology makes it possible to extract whether an extracted feature of a topological space (for example a hole) is "robust"by for example using different distance functions for computing the shape of this space. Small irregularities in scientific datasets are often seen as "noise", many times they do not deliver useful information about the data. Nevertheless, the decision of whether extracted information is "noise"or essential information depends on the point of view of the data scientist and the source of the data. Persistent homology is able to define precisely how big the irregularity of a shape is; thus, it makes it easy to detect "noise". This ability of persistent homology of extracting substantial information of data without scaling it makes it useful to be used before using other analytic methods.

1.1.3 A brief history of Persistent Homology

Persistent homology was invented independently from each other in the last fifteen years of the past century by researchers in Bologna, Duke University (North Carolina) and Boulder University in Colorado. I can highly recommend the book "Computational Topology, An Introduction"[3] by Professors Edelsbrunner and Harer for interested students that want to acquire additional knowledge in this field, both of them researched at Duke University at the time Persistent homology arose. I used it for writing chapter 2 of my thesis. In the survey of N. Otter [8] you can find useful information about which software tools to use for computing persistent homology.

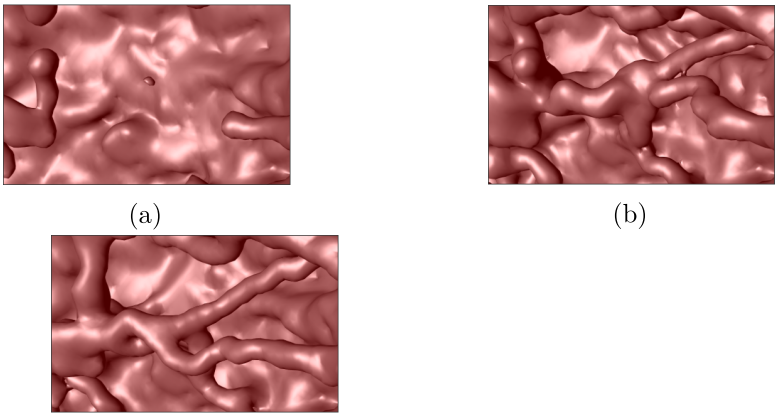
[4], [3], [5], [7], [6]

1.2 Motivation for finding representatives of holes

In a discretized topological space, holes are represented by equivalence classes of cycles. The algorithm of detecting holes does not detect explicit cycles that represent the hole. It defines a hole be expressing how "big"it is. In most cases, this special feature of knowing explicit cycles that represent holes is not needed as persistent homology tries to capture the robust shape of a space and not independent cycles. Nevertheless, there exist examples where the knowledge of single representatives of holes is requested. Cardiac image analysis tries to reconstruct trabeculae, tissue elements, traditional image segmentation methods are not capable of capturing their delicate surface precisely enough. Without going into details, the best way to reconstruct an image of those trabeculae is to use optimal loops representing a hole. [9]

I focused on this exact topic: Given information about a hole, I want to extract a single loop that represents it. I circumscribed the topic of my thesis to the case of 1-cycles and holes of dimension 1, thus, cycles the way we imagine them and "$2D$"holes. The algorithm turns the problem of finding a representative loop into the problem computing the shortest path between two discrete points. This exact topic has also been treated by researchers all over the united states in the past year. Except for the introduction, the article [9] they published has not been taken into account in this document though.

I want to thank Morten Brun for being an excellent supervisor. Moreover, I want to thank Francisco Gomez for receiving me at the National University



(a) (b) (c)

Figure 1: Caption(a) shows the segmentation result of trabeculae with traditional image processing methods and not using topological data analysis. In figure (b) topological data analysis was used before and for optimizing the image processing methods that followed. In caption (c) we see the improvement of the image provided persistent homology and the knowledge of the optimal cycles representing holes. [9]

of Colombia, for introducing me to topological data analysis and for explaining me the main construct of the algorithm that I am going to provide. This thesis is dedicated to my family and friends, and to Mao.

2 The theory behind persistent homology

2.1 Simplicial Complexes

2.1.1 Simplices

Most definitions and theorems are taken out of the great book "Computational Topology: An Introduction" by Professors Edelsbrunner and Harer from Duke University [3].

Definition 2.1. Let $x_1, x_2, \ldots, x_k$ be elements of $\mathbb{R}^n$ and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be in $\mathbb{R}$. Moreover, we require that the sum of all λ_i is 1. We call the sum $\sum_{j=1}^{k} \lambda_j \cdot x_j$ an *affine combination* of the points $x_1, x_2, \ldots, x_k$.

Definition 2.2. Let $x_1, x_2, \ldots, x_k$ be points in $\mathbb{R}^n$. We call the set of all their affine combinations their *affine hull*.

Definition 2.3. Let λ_j and μ_j be elements of $\mathbb{R}$ for $j = 1, ..., k$. We call the points $x_1, x_2, \ldots, x_k \in \mathbb{R}^n$ *affinely independent* iff two points $\sum_{j=1}^{k} \lambda_j \cdot x_j$ and $\sum_{j=1}^{k} \mu_j \cdot x_j$ in $\mathbb{R}^n$ are the same iff all the coefficients accord with each other, thus $\lambda_j = \mu_j$ for $j = 1, \ldots, n$.

Definition 2.4. Let $\sum_{j=1}^{k} \lambda_j x_j$ be an affine combination. Iff the coefficients $\lambda_1, \lambda_2, \ldots, \lambda_k$ are greater than zero, we call it a *convex combination*.

Definition 2.5. Let $x_1, x_2, \ldots, x_k$ be in $\mathbb{R}^n$. We call the set of all convex combinations of $x_1, x_2, \ldots, x_k$ their *convex hull*.

A simplex is the generalization of a line segment, a triangle, a tetrahedron and so forth:

Definition 2.6. Let $x_1, x_2, \ldots, x_{k+1}$ be in affinely independent points in $\mathbb{R}^n$. A *k-simplex* or *simplex of dimension k* σ is the convex hull of $x_1, x_2, \ldots, x_{k+1}$. We say the x_i *span* σ and we denote σ as

$$\sigma = [x_1, x_2, \ldots, x_{k+1}].$$

Definition 2.7. We call a 0-simplex *vertex*, an 1-simplex *edge*, a 2-simplex *triangle*, a 3-simplex *tetrahedron*.

Remark 2.8. If we take a subset of affinely independent points in $\mathbb{R}^n$, then it is again affinely independent, thus, it defines a simplex aswell.

Definition 2.9. Let $x_1, x_2, \ldots, x_{n+1}$ be a set of affinely independent points that span an n-simplex σ. We take a subset consisting of $k+1$ elements of these points. The k-simplex spanned by that subset is called a *k-face* of σ. If k is strictly smaller than n, then we call the face a *proper face* of σ.

Remark 2.10. Let τ be a (proper) face of a simplex σ. We write $\tau \leq \sigma$ $(\tau < \sigma)$.

Definition 2.11. We call the union of all proper faces of a simplex its *boundary*. We write *bd* σ. Moreover, we call $\sigma \setminus (\text{bd } \sigma)$ the *interior* of the simplex.

Remark 2.12. Let the points $x_1, x_2, \ldots, x_k \in \mathbb{R}^n$ be affinely independent, x a point in the simplex σ spanned by $x_1, x_2, \ldots, x_k$. Furthermore, let $\sum_{j=1}^{k} \lambda_j \cdot x_j = x$ be its representation as a linear combination of $x_1, x_2, \ldots, x_k$. Then, x belongs to the boundary bd σ of σ iff one of the λ_is is zero.

2.1.2 Simplicial Complexes

Later on, we will use simplices for describing closeness of the vertices. We will now introduce a set that allows us to unite simplices in a practical way: faces of simplices are always contained in the set. This means that if three vertices are close, two of them are close too.

Definition 2.13. We define a *simplicial complex* as a set of simplices with the following properties:
The intersection of two simlices is either empty or a face of both of them.
All the faces of a simplex in the simplicial complex are included in the simplicial complex aswell.

Definition 2.14. We call the maximum of the dimensions of the simplices included in a simplicial complex K its *dimension*.

Definition 2.15. A simplicial complex that is a subset of another simplicial complex K is called a subcomplex of K.

Theorem 2.16. *Let K be a simplicial complex, $L \subset K$ a subset. Then L is a subcomplex of K iff any face τ of a simplex in L is again in K.*

Later on, we will see that taking only the vertices and edges of a simplicial complex we get a graph. We will generalise this idea of extracting simplicies with dimension smaller than some threshold.

Definition 2.17. We define the *i-skeleton* $K^{(i)}$ of a simplicial complex K as the union of all simplices of a dimension that is lower than i.

Now we define something similar to the "social network" of a simplex:

Definition 2.18. Let K be a simplicial complex, σ a simplex in K. We define the star $St(\sigma) \subseteq K$ of σ as the subset of K containing all simplices that have σ as a face.

This we need for later. If we not only take out a simplex of a simplicial complex but also all the simplices including it, we prevent the remaining complex from losing the property of containing all faces of simplices.

Theorem 2.19. *Let K be a simplicial complex, τ a simplex in K. Then the complement $K \setminus St(\tau)$ of the star of τ in K is a subcomplex of K.*

Proof. We have to show that any face of a simplex in $St(\tau)$ is an element of $St(\tau)$. Suppose that there is a simplex ϵ in $K \setminus St(\tau)$, whose face σ has been deleted. Then σ has τ as a face. But then also ϵ has τ as a face, and ϵ is not an element of $St(\tau)$, a contradiction. □

Definition 2.20. We take the union of the simplices of a simplicial complex K which is a subspace of $\mathbb{R}^n$. The subspace topology of this union inherited from $\mathbb{R}^n$ is called the *underlying space*.

2.1.3 Abstract simplicial complexes

In this section we will note that we can express simplicial complexes with nothing more than subsets of the set of vertices that span the complex, a very practical characteristic that allows us to express closeness in a simple way, similar to topologies.

Definition 2.21. We define an *abstract simplicial complex A* as a collection of sets which is closed under the operation of taking subsets: If $\alpha \in A$ and $\beta \subseteq \alpha$, then $\beta \in A$.

Definition 2.22. We call those sets that are elements of an abstract simplicial complex *simplices*.

Definition 2.23. We define the *dimension* of a simplex α as the cardinality of the simplex subtracted by one: $dim\alpha = |\alpha| - 1$. The *dimension* of an abstract simplicial complex is the maximum dimension of its simplices.

Definition 2.24. A nonempty subset of a simplex is a *face*, a proper subset is called a *proper face*.

Definition 2.25. A *vertex* is an element of a simplex. The *vertex set* of an abstract simplicial complex is the union of all its simplices.

Definition 2.26. Two abstract simplicial complexes are called *isomorphic* if there exists a bijection between their vertex sets.

A maximal simplicial complex is a complex where all the vertices are close and everything is connected.

Definition 2.27. One can construct an abstract simplicial complex of a set of vertices V by taking the powerset of V. We call it the *maximal simplicial complex* of V.

Definition 2.28. We take a simplicial complex K, we can construct its abstract equivalent A by putting the vertices that span its faces into sets, ergo by making "abstract" faces out of them. A is called the *vertex scheme* of K.

Definition 2.29. The geometric equivalent of an abstract simplicial complex is called its *geometric realization*.

This theorem tells us that we can see simplicial complexes and abstract simplicial complexes as the same:

Theorem 2.30 (Geometric Realization Theorem)**.** *There exists a geometric realization in* $\mathbb{R}^{2d+1}$ *for any abstract simplicial complex of dimension* d*.*

2.2 The Vietoris-Rips-Complex

Definition 2.31. Let S be a subset of a metric space (X, d). Then the diameter $diamS$ is the supremum of all the distances between the elements in S: $diamS \coloneqq sup\{d(x, y) : x, y \in S\}$.

Definition 2.32. Let S be a finite subset of a metric space (X, d), $r \in \mathbb{R}_+$. We call the set $\{\sigma \subseteq S \mid diam(\sigma) \leq 2r\}$ the *Vietoris-Rips complex* $Rips(S, r)$ of S and radius r.

Theorem 2.33. *The Vietoris-Rips complex of a set* S *is indeed an abstract simplicial complex.*

Proof. Let V be the Vietoris-Rips complex of a set of points S. To prove that V is an abstract simplicial complex we first must prove that it is closed under the operation of taking subsets. Let σ be a simplex in V. Then $diam(\sigma) \leq 2r$. For any face ρ of σ holds: $diam(\rho) \leq diam(\sigma)$. Therefore, ρ is an element of V. □

Remark 2.34. Let S be a set of points. Then the Vietoris-Rips complex Vietoris-Rips$(0, S)$ of S is S itself.

Remark 2.35. Let S be a set of points. Then the complex Vietoris-Rips(r_1, S) is a subcomplex of Vietoris-Rips(r_2, S) iff $r_1 \leq r_2$.

Remark 2.36. The set of vertices S of a Vietoris-Rips complex $V \coloneqq$ Vietoris-Rips(r, S) is always a subset of V.

2.3 The Chain Group

Definition 2.37. Let K be a simplicial and let A be a ring, a_i elements in A and σ_i p-simplices in K. A formal sum $\sum a_i\sigma_i$ is called a *p-chain.*

Remark 2.38. Computational topology mostly uses $\mathbb{Z}_2$ as the ring that provides the coefficients. So do we.

Definition 2.39. Let $x = \sum a_i\sigma_i$ and $y = \sum b_i\sigma_i$ be two p-chains of a simlicial complex. We define the addition on x and y as follows:

$$x + y = \sum(a_i + b_i)\sigma_i.$$

Remark 2.40. When writing a chain $\sum a_i\sigma_i$, we only have two coefficients 0 and 1, $a_i = 0$ or $a_i = 1$. For being gentle to our eyes, we simply don't write the simplices with coefficient zero. Moreover, we define the zero-chain as $0 = \sum 0 \cdot \sigma_i$.

Remark 2.41. Note that the addition $a_i + b_i$ is $\mod 2$, therefore, if we take the same simplex σ twice we get $\sigma + \sigma = 0$.

Theorem 2.42. *Let p be a dimension. The p-chains of a simplicial complex together with the previously defined addition form an abelian group* $(C_p, +)$.

Proof. Firstly, we will show that the addition $+$ is an operation. For two elements $x = \sum a_i\sigma_i$ and $y = \sum b_i\sigma_i$ in C_p their sum $x + y = \sum(a_i + b_i)\sigma_i$ is clearly an element of C_p. Secondly, we want to proove the associativiy of the addition. Let $x = \sum a_i\sigma_i$, $y = \sum b_i\sigma_i$ and $z = \sum c_i\sigma_i$ be in C_p. Then

$$\begin{aligned}(x + y) + z &= \sum (a_i + b_i)\sigma_i + \sum c_i\sigma_i \\ = \sum (a_i + b_i + c_i)\sigma_i &= \sum a_i\sigma_i + \sum(b_i + c_i)\sigma_i \\ &= x + (y + z)\end{aligned}$$

Thirdly, we need to show that there exists an identity element. We will do so with the element 0 that we defined before as $0 = \sum 0 \cdot \sigma_i$. Let x be an element of C_p. $0 + x = \sum(0 + a_i)\sigma_i = \sum a_i\sigma_i = x$. The other way round we get $x + 0 = 0$. Moreover, we need to show that there exists an inverse element for every element of C_p. Clearly, every element itself is its inverse. And, for coming to an end, C_p is abelian because $\mathbb{Z}_2$ is abelian:

$$x + y = \sum (a_i + b_i)\sigma_i = \sum (b_i + a_i)\sigma_i = y + x.$$

□

Definition 2.43. Let K be a simplicial complex. The group $(C_p, +)$ is called the *group of p-chains* or the *p-th chain group*. Denote $C_p = C_p(K)$.

Remark 2.44. Let d be the dimension of a simplicial complex. For all integers p with $0 \leq p \leq d$ there exists a group of p-chains.

Definition 2.45. We call the formal sum of all $p-1$-faces of a p-simplex σ its **boundary** $\partial_p\sigma$. Or, more formally: Let $v_1, v_2, \ldots, v_{p+1}$ be the points that span σ. Denote $[v_1, v_2, \ldots, \hat{v}_i, \ldots, v_{p+1}]$ as the simplex spanned by the points $v_1, v_2, \ldots, v_{i-1}, v_{i+1}, \ldots, v_{p+1}$. Then we can write the boundary of σ as $\partial_p\sigma = \sum_{i=1}^{p+1}[v_1, v_2, \ldots, \hat{v}_i, \ldots, v_{p+1}]$.

Definition 2.46. We define the *boundary* $\partial_p c$ of a p-chain $c = \sum a_i\sigma_i$ as the boundary of the simplices that create it: $\partial_p\sigma = \sum a_i\partial_p\sigma_i$.

Definition 2.47. The function $c \mapsto \partial_p c$ that assigns any p-chain c its boundary is called the *boundary operator*.

Remark 2.48. Let K be a simplicial complex. The boundary operator ∂_p is a function with the group of p-chains as a domain and the group of $(p-1)$-chains as a codomain: $\partial_p : C_p \to C_{p-1}$.

Remark 2.49. For convenience we will often not write the index of the boundary operator.

Theorem 2.50. *The boundary operator is a homomorphism.*

Proof. Let $c = \sum a_i\sigma_i$ and $d = \sum b_i\sigma_i$ be p-chains. Then

$$\begin{aligned}\partial_p(c+d) &= \partial_p(\sum a_i\sigma_i + \sum b_i\sigma_i) = \partial_p(\sum(a_i+b_i)\sigma_i) \\ &= \sum(a_i+b_i)\partial_p\sigma_i = \sum a_i\partial_p\sigma_i + \sum b_i\partial_p\sigma_i \\ &= \partial_p(\sum a_i\sigma_i) + \partial_p(\sum b_i\sigma_i) = \partial_p c + \partial_p d\end{aligned}$$

. □

Definition 2.51. The boundary operator is also called the *boundary homomorphism.*

Definition 2.52. The sequence $\ldots \xrightarrow{\partial_{p+2}} Cp+1 \xrightarrow{\partial_{p+1}} Cp \xrightarrow{\partial_p} Cp-1\ldots$ of chain groups constructed by the boundary operator is called the *chain complex.*

2.4 Cycles and boundaries

If we walk along a cycle, there is no starting and no ending point, thus, there is no boundary:

Definition 2.53. We call a p-chain whose boundary is 0 a *p-cycle*.

Theorem 2.54. *The set Z_p of all the p-cycles of a group of p-chains forms an abelian subgroup.*

Proof. The elements of Z_p are the ones with empty boundary (the ones that have 0 as a boundary). Therefore, they are exactly the kernel of the boundary homomorphism ∂_p. Due to that, Z_p is a subgroup of C_p. It is abelian because the group of p-chains it lives in is abelian. □

Definition 2.55. Let K be a simplicial complex. We call $Z_p = Z_p(K)$ the *group of p-cycles.*

Remark 2.56. Z_p is abelian because its superset C_p is.

Vertices always work a little different than the rest of the simplicies. As we cannot form circles jumping just from one vertex to another, being a cycle and not being one is the same in the space of only points:

Remark 2.57. The group of 0-chains is the same as the group of 0-cycles of a simplicial complex.

Now, what is the difference between being a cycle and being a boundary? There must exist some chain whose boundary you are:

Definition 2.58. We call a p-chain a *p-boundary* if it is an element of the image of the boundary operator ∂_{p+1}. Sometimes we omit the p.

The properties of B_p and Z_p being subgroups will allow us to work with quotient spaces later on.

Theorem 2.59. *The subset B_p of a chain group C_p consisting of all the boundaries in C_p is an abelian subgroup.*

Proof. Let c and d be two elements in B_p. Being an element of B_p means being the image of some element in C_{p+1}: $\partial_{p+1}\tilde{c} = c$ and $\partial_{p+1}\tilde{d} = d$. But the boundary operator is a homomorphism, therefore we can write $\partial_(\tilde{c} + \tilde{d}) = \partial\tilde{c} + \partial\tilde{d} = c + d$. Thus, $c + d$ is an element of B_p. Finally, B_p is abelian because its superset Z_p is. □

Definition 2.60. Let K be a simplicial complex. We call $B_p = B_p(K)$ the *group of p-boundaries.*

Theorem 2.61. *Let K be a subcomplex of L, then the p-th boundary group of K is a subset of the p-th boundary group of L for $p \leq dim(L)$.*

Proof. Let $\partial\sigma \in K$ be a boundary of a simplex $\sigma \in K$, then both, σ and its boundary are in L. □

Theorem 2.62 (Fundamental Lemma of Homology)**.** *Let c be a $p+1$-chain. The boundary $\partial_p \partial_{p+1} c$ of the boundary of c is always empty.*

Remark 2.63. The group of p-boundaries is a subgroup of the p-cycles, $B_p \subseteq Z_p$.

Theorem 2.64. *Let S be a set of points in a metric space (X, d) and let $v_1, v_2, \ldots, v_n$ be elements of S, $n \geq 2$. Let K be a Vietoris Rips complex, $K \coloneqq$ Vietoris-Rips(r, S). Let $c = \sum_{i=1}^{p} [v_1, v_2, \ldots, \hat{v}_i, \ldots, v_n]$, where $\hat{v}_i$ denotes a missing vertex v_i, be an element of the chain group $C_p(K)$. Then the $p-1$-simplex $\sigma \coloneqq [v_1, v_2, \ldots, v_p]$, whose boundary is c, is an element of K.*

Proof. As c is an element of $C_p(K)$, all the summands $\tau_i \coloneqq [v_1, v_2, \ldots, \hat{v}_i, \ldots, v_n]$ of c are elements of K. Now let's look at the construction of the Vietoris Rips complex K. The τ_i are in K because their diameter is less than r. The union of τ_i and τ_j gives σ for $i \neq j$. Now we know that the distances between all pairs of vertices in σ except $d(v_i, v_j)$ are less than r. As the number of τ_is is bigger than two, looking closely at τ_k, $k \neq i$ and $k \neq j$, we realize that $d(v_i, v_j)$ is also less than r. Therefore, the diameter of σ is less than r and σ is an element of K. □

Definition 2.65. We call the quotient space Z_p / B_p of the group of p-cycles modulo group of the p-boundaries the *p-th homology group H_p*. Moreover, we call the rank of H_p the *pth Betti number β_p*, $\beta_p = |H_p|$.

And now we get to the abstract description of what we consider a hole.

Definition 2.66. An element of Z_p / B_p is called a *homology class.* A homology class is said to be *trivial* if it is homologous to the zero element of H_p.

Thus a hole can be described by different cycles.

Definition 2.67. Any two cycles c and d that are in the same homology group are *homologous*, write $c \sim d$. We write the homology group that c belongs in as $\bar{c} \coloneqq c + B_p$ and call c a representative of $\bar{c}$.

Remark 2.68. The homology group H_p is abelian because the group of cycles Z_p is.

2.5 Filtrations

In this chapter we describe how different simplicial complexes can be connected, we will remain with a increasing (in terms of taking supersets) sequence of subcomplexes. The Vietoris-Rips complex is a good example to explain this procedure: by incresing the radius between the points we turn into simplices, we create a sequence of subcomplexes.

Definition 2.69. A function $f : K \to \mathbb{R}$ with a simplicial complex K as a domain and an image in $\mathbb{R}$ is said to be *monotonic* if the image of a face τ of a simplex σ is less or equal than the image of σ: $f(\tau) \leq f(\sigma)$.

Definition 2.70. Let Y be a metric space, $f : X \to Y$ a function. Furthermore, let c be in Y. We call the set $\{x \in X : f(x) \leq c\}$ of all elements in X whose image is smaller than c a *sublevel set* of f. Similar to that, we call the set $\{x \in X : f(x) \geq c\}$ of all elements in X whose image is greater than c a *superlevel set* of f.

Theorem 2.71. *Let $f : K \to \mathbb{R}$ be a monotonic function with a simplicial complex K as a domain. The sublevel set $K(a) := f^{-1}(-\infty, a]$ is a subcomplex of K for all a in $\mathbb{R}$.*

Proof. We have to show that $K(a)$ is a simplicial complex. Let σ be a simplex in $K(a)$. We want to show that any face μ of σ is an element of $K(a)$. The image $f(\sigma)$ of σ is smaller than a, but as f is monotonic, the image $f(\mu)$ of μ is smaller than $f(\sigma)$. Therefore it is in $K(a)$. □

Definition 2.72. Let K be a simplicial complex with m simplices, $f : K \to \mathbb{R}$ monotonic. Let $A \coloneqq \{a_1, a_2, \ldots, a_m\}$ be the set of images $f(\sigma)$ of the simplices σ in K, $n \leq m$. Choose a subset of A, order them and rename them to end up with $a_1 < a_2 < \cdots < a_n$ for $n \leq m$. Define $a_0 \coloneqq -\infty$. By successively computing the sublevel set $K_i \coloneqq K(a_i)$ of f, we get a sequence

$$\emptyset \coloneqq K_0 \subset K_1 \subset \cdots \subset K_n \eqqcolon K,$$

where K_i is a strict subset of K_{i+1}, called *filtration* of f, of simplicial subcomplexes of K.

Remark 2.73. Let $K_0 \subseteq K_1 \subseteq \cdots \subseteq K_m$ be a filtration of a simplicial complex K. Elements K_i, $i < m$ do not necessarily contain all vertices of K.

Proof. Let us look at the simplicial complex $K \coloneqq \{\{v_1\}, \{v_2\}\}$. Define the monotonic function $f : K \to \mathbb{R}$ by $f(v_1) \coloneqq 1$ and $f(v_2) \coloneqq 2$. Thus, $a_i = i$, and the two different filtrations we can get are: $K_0 \subseteq K_1 = \{\{v_1\}\} \subseteq K_2 = K$ and $K_0 \subseteq K_1 = K$. □

The Vietoris-Rips filtration is, amongst others, one of the most common ways to express closeness with simplices:

Theorem 2.74. *Let S be a finite subset of a metric space. For every $r \in \mathbb{R}_+$ compute the Vietoris-Rips complex Rips(S, r) (definition 2.32). Define K as the maximal complex of S (definition 2.27). Moreover, define $f : K \to \mathbb{R}_+$ as:*

$$f(\sigma) \coloneqq \left\{ \begin{array}{ll} \mathit{min}\{r \in \mathbb{R}^+ : \sigma \in \mathit{Rips}(S, r)\}, & \mathit{for}\ \sigma \notin S \\ 0, & \mathit{for}\ \sigma \in S \end{array} \right\}$$

Then f is a monotonic function.

Definition 2.75. Let S be a finite subset of a metric space. We call the filtration of the function f defined in the above theorem (theorem 2.74) the *Rips-filtration* Rips(S) of S.

As intuitively the distance of a point to itself is zero, we want the data points themselves to be in the filtration from the beginning on:

Remark 2.76. Let $K_0 \subset K_1 \subset \cdots \subset K_n$ be a Rips-filtration of a set S. Then K_1 is equal to S.

Proof. Vietoris-Rips$(S, 0) = S$ (remark 2.34). The codomain of f is a subset of $\mathbb{R}^+$ and its smallest value is $0 = a_1$ (definition 2.72). Thus, $K_1 = f^{-1}(-\infty, 0] = S$. □

2.6 Birth and death of homology classes

As the notion of closeness increases in a filtration, new simplices appear in later elements of the sequence of complexes, we need an expression to say that a chain appears:

Definition 2.77. Let $K_0 \subset K_1 \subset \cdots \subset K_n$ be a filtration of a simplicial complex. A chain c in $C_p(K_j)$ is said to be *born* in $C_p(K_j)$, if it is not an element of $C_p(K_i)$ for $i < j$.

We also want to speak about the notion of a hole being born, and in the case of holes, about death.

Theorem 2.78. *Let K_i and K_j be elements of a filtration K, $i \leq j$. The map $f_p^{i,j} : H_p(K_i) \to H_p(K_j)$ defined by $c + B_p(K_i) \mapsto c + B_p(K_j)$ is a homomorphism.*

Proof. As K_i is a subset of K_j, the chain $c \in K_i$ is an element of K_j and $f_p^{i,j}(c + B_p(K_i)) = c + B_p(K_j)$ is clearly an element of $H_p(K_j)$. Moreover,

$$f_p^{i,j}[(c + B_p(K_i)) + (d + B_p(K_i))] = f_p^{i,j}[(c + d) + B_p(K_i)] = \\ (c + d) + B_p(K_j) = (c + B_p(K_j)) + (d + B_p(K_j))$$

. □

Remark 2.79. By applying $f_p^{i,i+1}$ successively to $H_p(K_i)$, we get the sequence

$$\emptyset =: H_p(K_0) \subset H_p(K_1) \subseteq \cdots \subseteq H_p(K_n) = H_p(K)$$

of the homology groups connected by homomorphisms.

Definition 2.80. Let $K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n$ be a filtration. Let $f_p^{i,j}$ be the homomorphism between the homology groups. We call its image

$$H_p^{i,j} := im f_p^{i,j} \subseteq H_p(K_j)$$

the *p-th persistent homology group* for $0 \leq i \leq j \leq n$.

Remark 2.81. Note that $H_p^{i,i} = im f_p^{i,i}$ is the same as $H_p(K_i)$.

The elements of the p-th persistent homology group $H_p^{i,j}$ are the homology classes of K_i that still exist at K_j we want to proove the existence of an isomorphism that is a demostrative representation of what $H_p^{i,j}$ looks like.

Lemma 2.82. *The kernel of $f_p^{i,j}$ is of the form*

$$ker\ f_p^{i,j} = {}^{B_p(K_j) \cap Z_p(K_i)} \big/ {}_{B_p(K_i)}$$

Proof. The image of a class in $H_p(K_i)$ is trivial iff its representant is a boundary in $Z_p(K_j)$. □

Theorem 2.83. *$H_p^{i,j}$ is isomorphic to* $Z_p(K_i) \big/ (B_p(K_j) \cap Z_p(K_i))$.

Proof. $f_p^{i,j}$ is a homomorphism. Its domain is $H_p(K_i)$. The First isomorphism theorem states that

$$\text{Im } f_p^{i,j} \cong H_p(K_i) \big/ \ker f_p^{i,j}.$$

$$H_p(K_i) = Z_p(K_i) \big/ B_p(K_i) \text{ and } ker f_p^{i,j} = B_p(K_j) \cap Z_p(K_i) \big/ B_p(K_i).$$

Now, $Z_p(K_i)$ is abelian, so, $B_p(K_i)$ and $B_p^j \cap Z_p^i$ are normal subgroups of $Z_p(K_i)$. Applying the Third isomorphism theorem, we get, that

$$\text{Im } f_p^{i,j} \cong Z_p(K_i) \big/ (B_p(K_j) \cap Z_p(K_i)).$$

□

Definition 2.84. A class γ in $H_p(K_i)$ is said to be *born in* K_i if it is not an element of $H_p^{i-1,i}$:

$$\gamma \in H_p(K_i) \text{ and } \gamma \notin H_p^{i-1,i}.$$

We also say that γ is born at (time) a_i.
Moreover, the class γ born in K_i is said to die entering K_j if

$$f_p^{i,j-1}(\gamma) \notin H_p^{i-1,j-1} \text{ but } f_p^{i,j}(\gamma) \in H_p^{i-1,j},$$

which means that γ becomes part of an older class in the transition from K_{j-1} to K_j. We also say that γ dies at (time) a_i. Figure 2.

Definition 2.85. The *persistence* $p(\gamma)$ of a homology class γ born in K_i and dying entering K_j is the difference between the function values a_i and a_j: $p(\gamma) = a_j - a_i$.

Definition 2.86. We define $\mu_p^{i,j}$ as the number of homology classes of dimension p born in K_i and dying entering K_j.

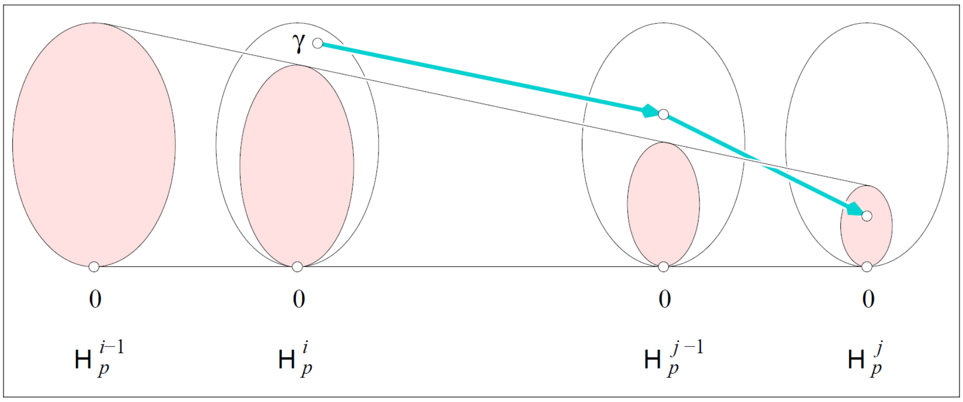


Figure 2: γ is born in K_i and it dies entering K_j. The red fields represent the images of $f_k^{i-1,i}$, $f_k^{i-1,j-1}$,$f_k^{i-1,j}$.

Definition 2.87. We call a collection of objects in which elements can occur many times a *multiset*. The times an element occurs in a multiset is called its *multiplicity*.

Definition 2.88. Let $f : K \to \mathbb{R}$ be monotonic. The *p-th persistence diagram* of the filtration of is the diagram that includes the function values a_i and a_j of f as points (a_i, a_j) and $\mu_p^{i,j}$ as their multiplicity for any p. We denote this diagram as $Dgm_p(f)$.

Definition 2.89. We call the interval (a_i, a_j) that defines the time of birth and death of a homology class its persistence interval / betti interval.

Remark 2.90. We do not distinguish between homology classes that have the same persistence interval.

Theorem 2.91. *Let $f : K \to \mathbb{R}$ be monotonic, K_i and K_j elements of the filtration of f. Moreover, let $\mu_p^{i,j}$ be the number of classes of dimension p that are born in K_i and die entering K_j. Then we can express $\mu_p^{i,j}$ with the p-th persistent betti numbers:*

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

for all p and all $i < j$.

Proof. $\beta_p^{i,j-1}$ describes the number of homology classes that are born before or at K_i and that die entering K_k for $j - 1 \leq k$. The classes included in $\beta_p^{i,j}$ are also born before or in K_i and they die entering K_k for $j \leq k$. The substraction $\beta_p^{i,j-1} - \beta_p^{i,j}$ counts the homology classes that are born before or in K_i and die entering K_j. Likewise does the $\beta_p^{i-1,j-1} - \beta_p^{i-1,j}$ count the homology classes that are born before or in K_i and die entering K_j. $\square$

3 Finding the responsible simplices for a hole

Finally we reached the chapter of exploring the question which is the headline of this article: As we are working with the betti numbers for aquiring knowledge about the holes of a space, we loose track of the circles that actually represent these holes. In some sence, we want to pass through the whole process that we described in all the previous chapters backwards now: Taking a hole defined by its persistence interval (a_i, a_j) (remark 2.90, definition 2.89), can we find a cycle living in the birthplace K_i of γ and representing γ? We will circumscribe our topic to the case of one-dimensional holes, so the "2D"holes. And yes, we can, in most cases. We proceed in two steps: First, we find an edge born in K_i that is a summand of a cycle in $C_1(K_i)$ that represents γ (definition 2.67). After that, we "delete"e from K_i. we percieve the edges of this new complex as edges between the vertices of a graph and we compute the shortest path $s \in C_1(K_i)$ between the two faces of e (as e is not an element of the new complex, the shorest path cannot be e). Finally, the sum $s + e$ represents γ.

3.1 Finding the responsible edge

Now, we consider the following task: Take a Rips-filtration

$$K_0 \subset K_1 \subset \cdots \subset K_m$$

of a set of points. Suppose, that we were given a homology class γ defined by its persistence interval (a_i, a_j). We want to find the last edge e missing in order to create the first cycle(s) that represent γ. The edge e is clearly born in K_i. It is computationally easy to find all the edges born in K_i (with the help of a distance matrix). It is a little harder to find out which edge is the edge we are looking for. Our approach to finding e is deleting one by one every edge born in K_i from K_i and checking wheather γ is born at the same time as it did before. If we delete one edge and a_i, the birth time of γ, does not change, the edge we are looking for is not e. Then we continue by adding the edge again and subtracting a different edge born in K_i. When a_i finally changes, we must have found e.

If we delete an edge e of a simplicial complex K_i, we have to make sure that it is still a simplicial complex. Therefore, we will also delete all the simplices which have e as a face (theorem 2.19). Usually, that shouldn't be many, because e was born in K_i, so all the simplices which have it as a face must have been born in the exact same moment. First, we will make sure, that the new complex $K_{i-\frac{1}{2}}$ is not a subset of an earlier complex K_{i-1}:

Lemma 3.1. *Let $K_0 \subset K_1 \subset \cdots \subset K_m$ be a Rips-filtration of a finite subset of a metric space. Suppose that the edge e is born in K_i. Construct a new simplicial subcomplex $K_{i-\frac{1}{2}} := K_i \backslash St(e)$ by taking the difference of K_i and the star $St(e) \subseteq K_i$ of e (this is possible due to theorem 2.19). Then*

$$K_{i-1} \subsetneq K_{i-\frac{1}{2}} \subset K_i.$$

Proof. We want to show, that K_{i-1} is a proper subset of $K_{i-\frac{1}{2}}$. The edge e is born in K_i, and it is a face of all the other simplices in $\mathrm{St}(e)$. If any simplex of $\mathrm{St}(e)$ was an element of K_{i-1}, then e would have to be in there too, because it is a face. But that is not the case, therefore $K_{i-1} \subsetneq K_{i-\frac{1}{2}}$. $K_{i-\frac{1}{2}}$ is clearly a subset of K_i. And because $\mathrm{St}(e)$ persists of at least e, it has to be a proper subset. □

Secondly, we define a new filtration that includes $K_{i-\frac{1}{2}}$, the complex of which we deleted e. For doing this, we have to define a suitable monotonic function.

Theorem 3.2. *Take the Rips filtration $K_0 \subset K_1 \subset \cdots \subset K_m$ of a set of vertices with the corresponding monotonic function $f : K_m \to \mathbb{R}$ (definition def:RipsFiltration). Let e be an edge born in K_i. Moreover, let $K_{i-1} \neq K_i \backslash St(e)$. Denote the images of f as*

$$f(K_i \backslash K_{i-1}) = r_i,\ 1 < i \leq m.$$

Define an $r_{i-\frac{1}{2}}$ such that

$$r_{i-1} < r_{i-\frac{1}{2}} < r_i.$$

Furthermore, define a function $f^*: K_m \to \mathbb{R}$ by changing the function values *of f of elements in $St(e)$.:*

$$f^*(\sigma) = \left\{ \begin{array}{ll} r_{i+\frac{1}{2}}, & \textit{for } \sigma \in St(e) \\ f(\sigma), & \textit{for } \sigma \notin St(e) \end{array} \right\}$$

Then f^* is a monotonic function.

Definition 3.3. We call the filtration

$$K_0 \subset K_1 \subset \cdots \subset K_{i-1} \subset K_{i-\frac{1}{2}} \subset K_i \subset \cdots \subset K_m$$

defined by the function f^* described in the previous theorem the *Rips* filtration* of V and e.

Remark 3.4. We will denote the Rips* filtration

$$K_0 \subset K_1 \subset \cdots \subset K_{i-1} \subset K_{i-\frac{1}{2}} \subset K_i \subset \cdots \subset K_m$$

as

$$K_0 \subset K_1 \subset \cdots \subset K_{i-1} \subset K_{i-\frac{1}{2}} \subset K_{\tilde{i}} \subset \cdots \subset K_m$$

from now on, in order to avoid confusions with the "normal"Rips-filtration

$$K_0 \subset K_1 \subset \cdots \subset K_{i-1} \subset K_i \subset \cdots \subset K_m.$$

Now, we define the subset of $H_1(K_{\tilde{i}})$ of holes that have e as a summand in all the cycles that represent it. We need this definition, because edges with this property are the ones we can detect.

Definition 3.5. Let $K_0 \subset K_1 \subset \cdots \subset K_m$ be a Rips filtration of a set of vertices and e_1. Let $e_1, e_2, \ldots, e_n$ be all edges contained in K_i. We call the subset of $H_1(K_{\tilde{i}})$ of all the homology classes of dimension one born in K_i and dying in K_j for which all representatives contain e_1 as a summand $H_1^{i,j}(e_1)$:

$$\overline{\sum a_i e_i} \in H_1^{i,j}(e) \Rightarrow a_1 \neq 0.$$

Moreover, we call the subset of $H_1(K_{\tilde{i}})$ of all the homology classes of dimension one born in K_i for which all representatives contain e_1 as a summand $H_1^i(e_1)$

The following remark will help us for using the previous definition better. It only says, that homology classes that have an edge e as summand in all their representative cycles can't be born before e, thus, $K_{\tilde{i}}$

Remark 3.6. Let

$$K_0 \subset K_1 \subset \cdots \subset K_m$$

be a Rips filtration of a set V and let

$$K_0 \subset K_1 \subset \cdots \subset K_{i-1} \subset K_{i-\frac{1}{2}} \subset K_{\tilde{i}} \subset \cdots \subset K_m$$

be a Rips* filtration of an edge e and V. There exists no homology class that is an element of $H_1^{i,j}(e)$ for some j and element of a younger homology group than $H_1(K_{\tilde{i}})$ (a younger homology group is the homology group of a proper subcomplex of $K_{\tilde{i}}$, so for example of $K_{i-\frac{1}{2}}$)).

Proof. By the definition of the Rips* filtration, the statement follows. □

Lemma 3.7. *Let $K_0 \subset K_1 \subset \cdots \subset K_{i-1} \subset K_{i-\frac{1}{2}} \subset K_{\tilde{i}} \subset \cdots \subset K_m$ be a Rips* filtration of a set of vertices and e. Then,*

$$H_1^i(e) = H_1(K_{\tilde{i}}) \setminus H^{i-\frac{1}{2},\tilde{i}}.$$

Proof. By remark 3.6, $H_1^i(e)$ is a subset of $H_1(K_{\tilde{i}})$.
First, we will show, that $H_1^i(e)$ is a subset of $H_1(K_{\tilde{i}}) \setminus H^{i-\frac{1}{2},\tilde{i}}$.
Suppose now, that $H_1^i(e)$ is not a subset of $H_1(K_{\tilde{i}}) \setminus H^{i-\frac{1}{2},\tilde{i}}$. Then, there exists an element γ in $H_1^i(e) \subseteq H_1(K_{\tilde{i}})$, which is not an element of $H_1(K_{\tilde{i}}) \setminus H^{i-\frac{1}{2},\tilde{i}}$. Now, γ is an element of $H_1^{i-\frac{1}{2},\tilde{i}}$, which means, that it is an element of $H_1(K_{i-\frac{1}{2}})$, a contradiction to remark 3.6.
Thus, $H_1^i(e) \subseteq H_1(K_{\tilde{i}}) \setminus H^{i-\frac{1}{2},\tilde{i}}$.
Now, we will show, that $H_1(K_{\tilde{i}}) \setminus H^{i-\frac{1}{2},\tilde{i}}$ is a subset of $H_1^i(e)$.
Suppose, that $H_1^i(e) \subset H_1(K_{\tilde{i}}) \setminus H^{i-\frac{1}{2},\tilde{i}}$. Then, there exists a homology class γ in $H_1(K_{\tilde{i}}) \setminus H^{i-\frac{1}{2},\tilde{i}}$ that is not an element of $H_1^i(e)$. The edge e is the only one born in $K_{\tilde{i}}$. Thus, it must be summand of all the first cycles that represents γ. This is a contradiction to the fact that γ is not an element of $H_1^i(e)$.
So, $H_1(K_{\tilde{i}}) \setminus H^{i-\frac{1}{2},\tilde{i}} \subseteq H_1^i(e)$. □

Theorem 3.8. *The number of holes in $H_1^i(e)$ is equal to the number of holes born in $K_{\tilde{i}}$.*

Proof. See the previous lemma. □

Lemma 3.9. *The number of holes of dimension one born in $K_{\tilde{i}}$ and dying in K_j is equal to $\mu_1^{i,j} - \mu_1^{i-\frac{1}{2},j}$.*

Proof. The holes born in $K_{\tilde{i}}$ are and dying in K_j are the ones that are born in K_i and die in K_j and that are not born in $K_{i-\frac{1}{2}}$ and dying in K_j. □

Theorem 3.10. *Let*

$$K_0 \subset K_1 \subset \cdots \subset K_m$$

be the Rips-filtration of a set of vertices V, moreover, let

$$K_0 \subset K_1 \subset \cdots \subset K_{i-1} \subset K_{i-\frac{1}{2}} \subset K_{\tilde{i}} \subset \cdots \subset K_m$$

be the Rips filtration of V and an edge e born in K_i. Then, the*

$$\textit{number of homology classes in } H_1^{i,j}(e) = \mu_1^{i,j} - \mu_1^{i-\frac{1}{2},j}.$$

Proof. See the previous lemma. □

Now, suppose, that we have calculated all the homology classes of a data set S with the help of a Rips-filtration. We pick the holes of dimension 1 defined by the persistence interval (r_i, r_j), there exist $\mu_1^{i,j}$ of them and we do not distinguish between them. We want to find an edge e with one of the homology classes that we picked in $H_1^{i,j}(e)$. So, we start by computing the Rips*-filtration of S and an edge born in K_i. Most of the holes will probably have the persistence interval $(r_{i-\frac{1}{2}}, r_j)$, this means, that they are not an element of $H_1^{i,j}(e)$.Now, the holes with a persistence interval that has not changed, are the interesting ones. Their representatives were born in $K_{\tilde{i}}$. If we get such an interval, then we know, that e is responsible for one of the holes defined by (r_i, r_j).

3.2 Some Graph theory

Before beginning to talk about shortest paths in simplicial complexes, we repeat some important definitions and results from graph theory [2]

Definition 3.11. A *simple unweighted graph* G is defined as a pair $(V, E) = (V(G), E(G)) = G$ of a set V and a subset $E \subset V \times V$ of the Cartesian product $V \times V$. The se V is called the set of *vertices* and also written as $V(G) = V$. Let v_1 and v_2 be vertices. An element (v_1, v_2) of E is called *edge* of the vertices v_1 and v_2. The set E of edges is also written as $E(G) = E$.

Remark 3.12. We can identify a simple graph G with a 1-skeleton (definition 2.17 $K^{(i)}$ of a simplicial complex K. Identify an edge (v_1, v_2) of G with an edge $[v_1, v_2]$ of $K^{(i)}$.

Definition 3.13. Let $G = (V(G), E(G))$ be a graph. A pair $F = (V(F), E(F))$ of subsets $V(F) \subseteq V(G)$ and $E(F) \subseteq E(G)$ is called a *subraph* of G.

It is very figurative to travel on edges from vertex to vertex:

Definition 3.14. Let G be a graph, v_1 and v_2 two of its vertices. A *path* in G between v_i and v_j is a subgraph of G whose vertices $v_1 =: w_1, w_2, \ldots, w_n \coloneqq v_2$ can be ordered in a sequence $(w_1, w_2, \ldots, w_n)$ such that the only vertices that share edges are neighbours in the sequence and v_i and v_j are at the very edges of the sequence.

Remark 3.15. We can identify the path $(w_1, w_2, \ldots, w_n)$ with the 1-chain $[w_1, w_2] + [w_2, w_3] + \cdots + [w_{n-1}, w_n]$.

Definition 3.16. The *length* of a path p is the number of its edges.

Definition 3.17. A path is called a *trail* if every vertex except the first and last one included in the sequence of vertices only appears once.

Definition 3.18. Let v_i and v_j be vertices of a simple unweighted graph G. The *shortest path* between v_i and v_j is the path $P \subseteq G$ whose set of edges $E(P)$ is the smallest of all paths between v_i and v_j.

Remark 3.19. A shortest path is a trail.

Now we come to an important remark on the possibility of solving the task of finding the shortest path between two vertices. We will see that this task is always solvable:

Definition 3.20. Let G be a simple, unweighted graph, $v_i, v_j \in G$. The *single-pair shortest path problem for a simple, unweighted graph* is the task of finding the shortest path between v_i and v_j.

Theorem 3.21. *The single-pair shortest path problem for a simple, unweighted graph can be solved.*

Proof. Dijkstra-algorithm □

3.3 Finding the responsible simplices

Now, given a hole γ of dimension 1 defined by its persistence interval (a_i, a_j), we can find an edge e that is summand of a representative cycle of γ. Given e, we want to find a representative cycle for γ now.

In order to prepare for the main theorem of this chapter, we see now that the existence of a shortest path between the faces of an edge e and a cycle that contains e is somewhat equivalent:

Lemma 3.22. *Let $K_0 \subseteq K_1 \subseteq \cdots \subseteq K_m$ be a filtration of a monotonic function. Moreover, let the two vertices v_1 and v_2 be elements of K_i (remark 2.73, definition 2.77) and let the edge $e \coloneqq [v_1, v_2]$ be born in K_j for $0 < i < j \leq m$. Then there exists a shortest path between v_2 and v_1 in K_i iff e is a summand of a cycle $e + c$ in $Z_1(K_j)$, where c is in $C_1(K_i)$.*

Proof. Let us see one direction of the prooge, the other one is even easier. Suppose, that there exists a shortest path s between v_2 and v_1 in $C_1(K_i)$ (remark 3.15). First, we want to show that $e + s$ is a cycle in K_j. Let $(v_1 \eqqcolon w_1, w_2, \ldots, w_k \coloneqq v_2)$ define s. As $e = [w_k, w_1]$ is not an element in K_i but s is a chain in K_i, we know that e and s are not equal. Now let us look at the boundary of $e + s \in C_1(K_j)$:

$$\begin{aligned} &\partial(e+s) \\ = \partial[w_k, w_1] + \partial[w_1, w_2] + \partial[w_2, w_3] + \cdots + \partial[w_{k-2}, w_{k-1}] + \partial[w_{k-1}, w_k] \\ = w_1 - w_k + w_2 - w_1 + w_3 - w_2 + \cdots + w_{k-1} - w_{k-2} + w_k - w_{k-1} \\ = 0. \end{aligned}$$

Therefore, there exists a cycle $e + s$ with e as a summand. □

Theorem 3.23. *Let $K_0 \subseteq K_1 \subseteq \cdots \subseteq K_m$ be a Rips-filtration. Let the edge $e \coloneqq [v_1, v_2]$ be born in K_i, $i > 1$ (remark 2.76). Let $H_1^{i,k}(e)$ be nonempty for some $k \geq i$ and let e not be summand of any homology class that is not an element of $H_1^{i,k}(e)$ for some $k \geq i$. If there exists a shortest path s in*

$C_1(K_{i-1})$ between v_1 and v_2, then, $e + s$ is a representative for a hole in $H^{i,j}(e)$ for some $j \geq i$.

Proof. Because s lies in $C_1(K_{i-1})$ and e is born in K_i and because of lemma 3.22, we know that $e + s$ is a cycle. We want to show that $e + s$ is a representative cycle some element in some $H^{j,k}(e)$. To do so, we first have to show that $\overline{e + s}$ is not trivial in $H_1(K_i)$.
Suppose that $\overline{e + s}$ is trivial. Now, because some $H^{i,j}(e)$ is nonempty, there exists a γ in $H^{i,j}(e)$ such that γ is of the form $\gamma = \overline{e + c}$ for some c in $C_1(K_i)$. Note that the boundary of $e+c$ is zero. We want to show now that $\overline{c + s} = \gamma$. We know that

$$\partial s = \partial(e + s) - \partial(e) = v_i - v_j,$$

and that

$$\partial c = \partial(e + c) - \partial e = v_i - v_j.$$

Therefore $\partial(c + s) = 2v_i - 2v_j = 0$. Thus, $c + s$ is in $Z_1(K_i)$.

$$\overline{c + s} = \overline{c + s + e + s} = \overline{c + e} = \gamma.$$

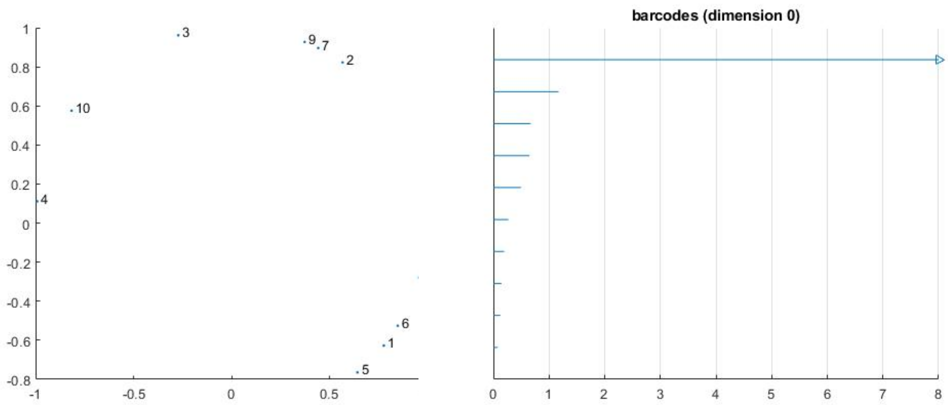
But $c + s$ does not contain e as a summand, which means that γ is not an element of any $H^{i,k}(e)$ for any k, contradiction to the fact that e is not a summand of any representative of any homology class that is not an element of some $H^{i,k}(e)$.
Therefore, we know that $\overline{e + s}$ is not trivial. So, $e + s$ is a representative for a hole in $H^{i,j}(e)$ for some $j \geq i$. □

Now, we can propose an algorithm that detects holes in many cases: Let $K_0 \subset K_1 \subset \cdots \subset K_m$ be a Rips-filtration. Given a persistence interval (r_i, r_j), we want to find as many representative cycles for holes with this persistence as we can. We can find all the edges $e_1, e_2, \ldots, e_k$, $k \leq \mu_1^{i,j}$, whose $H_1^{i,j}(e)$ is not empty. This procedure is described in section 3.1. Then, for every e_l, $1 \leq k$ do the following: Compute the shortest path s_j between the faces of e_j if possible. If e_j is not summand of a homology class that is not element of $H_1^{i,j}(e)$, then $e_k + s_k$ is a representative cycle for a hole defined by (r_i, r_j).

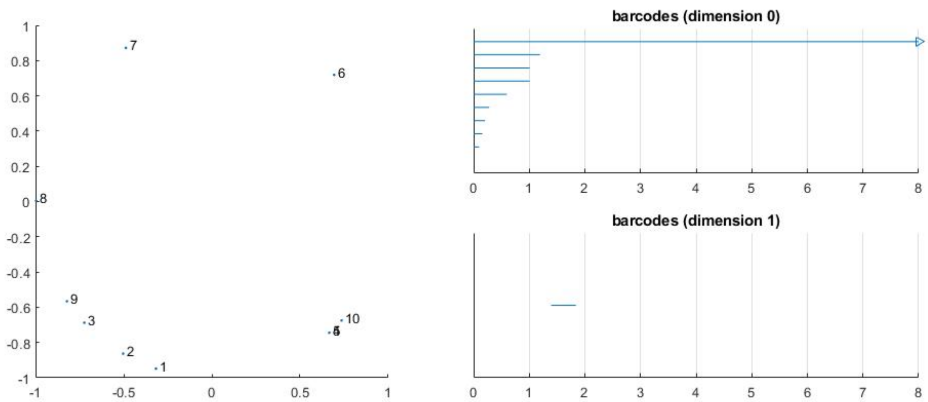
3.4 A field study with Javaplex

Javaplex is an open source software package provided by the University of Stanford, it was built over the past ten years by the research group for computational topology there. It serves to compute persistent homology. It can be used on Java and Matlab, I chose to use it on Matlab. The research group provides an extremely helpful tutorial[1].

In order to test the theories I proved, I created a program that creates a random set of 10 points in the unit circle in $\mathbb{R}^2$. The probablility of getting only one hole of dimension 1 before all the components are connected is pretty high in this constricted setting. We see that the theory works in this restricted setting. The (messily displayed) code can be found in the appendix and the visual outcome, the plots of the small pointclouds and the persistence barcodes can be observed in Figure 3.



(a) A circle without a hole (b)



(c) A circle with a hole (d)

Figure 3: (a) shows a pointcloud that doesn't produce any homology class of a dimenion greater than zero. (b) shows the connected components. In figure (c) some may be able to observe the cycle $(6, 7, 8, 2, 10)$ representing a homology class of dimension 2 which is shown in the barcodes in (d). This cycle has been detected by my program.

References

[1] Henry Adams and Andrew Tausz. Javaplex tutorial. *Google Scholar*, 2011.

[2] Bondi and Murty. *Graph Theory.* Springer, 1994.

[3] Edelsbrunner and Harer. *Computational Topology: An Introduction.* American Mathematical Society, 2010.

[4] Herbert Edelsbrunner and John Harer. Persistent homology-a survey. *Contemporary mathematics*, 453:257–282, 2008.

[5] Ulderico Fugacci, Sara Scaramuccia, Federico Iuricich, and L De Floriani. Persistent homology: a step-by-step introduction for newcomers. In *Proceedings of the Conference on Smart Tools and Applications in Computer Graphics*, pages 1–10. Eurographics Association, 2016.

[6] Wang Kairui Glen. The basic theory of persistent homology, 2012.

[7] Prerna Nadathur. An introduction to homology. *University of Chicago*, 2007.

[8] Nina Otter, Mason A Porter, Ulrike Tillmann, Peter Grindrod, and Heather A Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6(1):17, 2017.

[9] Pengxiang Wu, Chao Chen, Yusu Wang, Shaoting Zhang, Changhe Yuan, Zhen Qian, Dimitris Metaxas, and Leon Axel. Optimal topological cycles and their application in cardiac trabeculae restoration. In *International Conference on Information Processing in Medical Imaging*, pages 80–92. Springer, 2017.

```
clc; clear; close all;
import edu.stanford.math.plex4.*;

% create the set of points
point_cloud = ...
    examples.PointCloudExamples.getRandomSphereProductPoints(10, ...
    1, 1)

% plot the points
scatter(point_cloud(:,1),point_cloud(:,2),'.')
% set the ratio of the chart
daspect([1 1 2])
% name the points
for K = 1 : length(point_cloud)
text(point_cloud(K,1) + 0.02, point_cloud(K,2) + 0.02, ...
    string(K));
end
```

```
%% create a Vietoris-Rips stream
% max_dimension describes the maximum dimension of the ...
    chain group, so we
% compute holes up to dimension n-1
max_dimension = 3;
% the maximum filtration value is the maximum length of ...
    points to be
% connected in the Vietoris-Rips complex
max_filtration_value = 4;
% the number of divisions gives us an idea about how ...
    exactly we are
% calculating
num_divisions = 10000;

% create the Vietoris-Rips stream
stream = api.Plex4.createVietorisRipsStream(point_cloud, ...
    max_dimension, max_filtration_value, num_divisions);

%% get persistence algorithm over Z/2Z
persistence = ...
    api.Plex4.getModularSimplicialAlgorithm(max_dimension, 3);
intervals = persistence.computeAnnotatedIntervals(stream)

%% plot the barcodes
```

```
options.filename = 'barcodes';
options.max_filtration_value = 8;
plot_barcodes(intervals, options);
```

```
% check whether there is a hole of dim 1
intervals_dim1 = ...
    edu.stanford.math.plex4.homology.barcodes.BarcodeUtility.getEndpoints(intervals, .
    1, 0)
if length(intervals_dim1) > 0
    % if there is one, start to search for the constructing ...
        points
    % get the radius where the hole appears
    dist = intervals_dim1(1)

    % calculate the distance matrix
    distArray = pdist(point_cloud)
    distMat = squareform(distArray)

    % we can delete all distances bigger than the distance ...
        where the hole
    % appears, this gives us all the connected points ...
        before the hole
    % closes
    matBeforeHole = distMat.*(distMat<dist)

    % plot the graph of all connected points before the ...
        hole appears
    lenPoints = length(point_cloud)
    names = arrayfun(@string, 1:lenPoints)
    celldata = cellstr(names)
    graphBeforeHole = graph(matBeforeHole,celldata)

    % iterate through all entries of the distance matrix
    for i = 1:lenPoints
        for j = i+1:lenPoints
            el = distMat(j,i)

            % check whether the entry is close enough to ...
                our selected
            % distance (as we can only approximate the ...
                Betti intervals we
            % have to accept approximate values)
            if abs(el-dist)<0.001
```

```
% delete one of the points with distance el
new_cloud = point_cloud
new_cloud(i,:) = []

% create the Vietoris-Rips for the ...
    pointcloud with that one
% point extracted
stream = ...
    api.Plex4.createVietorisRipsStream(new_cloud, ...
    max_dimension, max_filtration_value, ...
    num_divisions);

% create the persistence interval for the ...
    new set of points
persistence = ...
    api.Plex4.getModularSimplicialAlgorithm(max_dimension, ...
    2)
intervals = ...
    persistence.computeIntervals(stream)
% get the holes of dimension one for the ...
    new stream
intervals_dim1 = ...
    edu.stanford.math.plex4.homology.barcodes.BarcodeUtility.getEndpoi
    1, 0)

%% check whether the new Vietoris-Rips ...
    stream gives the
%% same holes at the same distance
% check whether the hole of dimension one ...
    still exists
if length(intervals_dim1) > 0
    % now we look at the new radius at ...
        which the hole of
    % dimension one appears
    newDist = intervals_dim1(1)

    % check whether the hole appears at the ...
        same distance,
    % if not, the points we're looking at ...
        are essential for
    % the hole
    if newDist ≠ dist
        % compute the shortest path between ...
            the two points
```

```
                        % we identified to be essential to ...
                            the hole, that
                        % gives us the rest of the points ...
                            that creat the
                        % hole
                        pointsOfHole = ...
                            shortestpath(graphBeforeHole,i,j)

                    end
                    % in the case that the new persistence ...
                        interval differs
                    % significantly from the original one, ...
                        we don't compute
                    % anything and continue iterating in ...
                        the hope of
                    % finding significant points
                else
                    % if there is no hole of dim 1, this ...
                        also means that
                    % the points are essential and we ...
                        create the shortest
                    % path
                    pointsOfHole = ...
                        shortestpath(graphBeforeHole,i,j)
                    display(pointsOfHole)
                 end

            end

        end
    end
    % those points create the hole
    display(pointsOfHole)

else
    % in the case that the persistence doesn't give us a ...
        hole of dimension
    % one, we want the function to communicate that
    display('There is no hole of dimension 1')
end
```