Homology and Persistent Homology Bootcamp Notes Summer@ICERM 2017

Melissa McGuirl

September 19, 2017

1 Preface

This bootcamp is intended to be a review session for the material covered in Week 1 of Summer@ICERM 2017. My goal is to clarify any confusing topics covered in Week 1 and to present the material in a way that connects all of the material into a full picture. If you have any questions about the material please feel free to email me at melissa_mcguirl@brown.edu. Most of the exercises have been taken directly from the references listed at the end.

2 Introduction to Topological Data Analysis

Topological data analysis (TDA) lies in the intersection of data science and algebraic topology. In data science, one of the main goals is to extract important features from large datasets and then classify or cluster the data based on their features. There are several methods for doing this and the choice of method depends largely on the data being studied.

In algebraic topology, one of the main goals is to classify topological spaces based on their topological features. There are two primary tools for classifying topological spaces: homotopy and homology. While homotopy is easier to define, it is nearly impossible to compute the homotopy type of a given space. In contrast, homology is tricky to define but as we have seen it is "easy" to compute. For this reason, homology is the tool that connects topology to data science.

Datasets often come in the form of point clouds lying in some metric space. Computing the homology of a point cloud itself would be uninformative as the homology groups of dimension $N > 0$ would all be trivial. Rather, we assume the points have been sampled from some underlying manifold and we attempt to extract the homology of that manifold by growing balls around each point in the point cloud. Since a priori it (usually) cannot be known how big to grow the balls, we instead compute the homology of the covered point cloud for a finite sequence of radii. This is known as *persistent homology*.

The persistent homology of a point cloud is typically represented by barcodes or persistence diagrams. These representations will be discussed below. Given barcodes or persistence diagrams, one can begin to classify different point clouds, or datasets, based on their topological features. This method has been successfully applied to a range of fields such as oncology, evolutionary biology, image processing and more.

3 Simplicial Complexes

Simplicial complexes are the main data structure used to represent topological spaces in TDA. We will review the main types of simplicial complexes used in TDA, but first we need to recall a few fundamental definitions. For the following definitions, let $u_0, u_1, \dots u_n$ be points in $\mathbb{R}^d$.

Definition 3.1 (Affine and Convex Combinations) *An affine combination of* $\{u_i\}_{i=0}^n$ *is a point* $x = \sum_{j=1}^n \lambda_i u_i$ *such that* $\sum_{i=1}^n \lambda_i = 1$. *A convex combination of* $\{u_i\}_{i=0}^n$ *is a point* $x = \sum_{j=1}^n \lambda_i u_i$ *such that* $\sum_{i=1}^n \lambda_i = 1$ ***and*** $\lambda_i \geq 0$ *for all* $i = 0, 1, \dots n$.

Definition 3.2 (Affine and Convex Hulls) *The affine hull of* $u_0, u_1, \dots u_n$ *is the set of all possible affine combinations of* $\{u_i\}_{i=0}^n$*:*

$$aff(u_0, u_1, \dots, u_n) \doteq \{x = \sum_{i=1}^n \lambda_i u_i : \sum_{i=1}^n \lambda_i = 1\}.$$

Similarly, the convex hull of $u_0, u_1, \dots u_n$ *is the set of all possible convex combinations of* $\{u_i\}_{i=0}^n$*:*

$$conv(u_0, u_1, \dots, u_n) \doteq \{x = \sum_{j=1}^n \lambda_i u_i : \sum_{i=1}^n \lambda_i = 1 \text{ and } \lambda_i \geq 0 \forall i\}.$$

Example 3.1 *Given two points* $x_1, x_2 \in \mathbb{R}^d$*, the parametrization of the line connecting* x_1 *to* x_2 *is given by*

$$x(t) = (1-t)x_1 + tx_2, \quad 0 \leq t \leq 1$$

Taking $\lambda_1 = (1-t)$ *and* $\lambda_2 = t$ *we see that for all* $0 \leq t \leq 1$

$$\sum_{i=1}^2 \lambda_i = (1-t) + t = 1,$$

and moreover since $0 \leq t \leq 1$, $\lambda_i \geq 0$ *for* $i = 1, 2$. *It is now not hard to see that*

$$conv(x_1, x_2) = \{x = \lambda_1 x_1 + \lambda_2 x_2 : \sum_{i=1}^2 \lambda_i = 1 \text{ and } \lambda_i \geq 0 \forall i\} = \{x = (1-t)x_1 + tx_2 : 0 \leq t \leq 1\}$$

Definition 3.3 (Affinely Independent) *The* $n+1$ *points,* $u_0, u_1, \dots u_n$*, are said to be affinely independent if and only if the n vectors* $u_i - u_0$*, for* $1 \leq i \leq n$*, are linearly independent. Note, in* $\mathbb{R}^d$ *there are at most* $d+1$ *affinely independent points.*

Just as the convex hull of two points is the edge or line segment connecting the points, the convex hull of 3 affinely independent points is a triangle parametrized by those points, and so on. We are now ready to define a simplex.

Definition 3.4 (Simplex) *A k-simplex is the convex hull of* $k+1$ *affinely independent points:*

$$\sigma = conv(u_0, u_1, \dots, u_n), \quad dim(\sigma) = n.$$

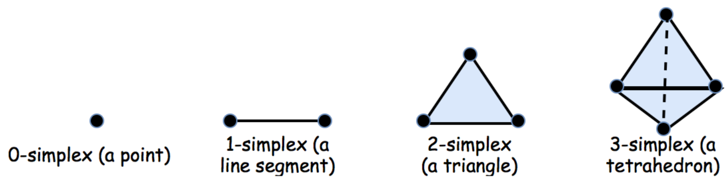


Figure 1: Figure courtesy of `http://outlace.com/TDApart1.html`

A 0-simplex is a vertex, a 1-simplex is an edge, a 2-simplex is a triangle, a 3-simplex is a tetrahedron, and so on.

Definition 3.5 (Face of a Simplex) *Given* $\sigma = conv(u_0, u_1, \dots, u_n)$*, a face* τ *of* σ*, denoted* $\tau \leq \sigma$*, is* $\tau = conv(u_{i_1}, u_{i_2}, \dots, u_{i_m})$*, where* $u_{i_1}, u_{i_2}, \dots, u_{i_m} \subset u_0, u_1, \dots, u_n$*. If* $u_{i_1}, u_{i_2}, \dots, u_{i_m} \subsetneq u_0, u_1, \dots, u_n$*, we say that* τ *is a proper face.*

Simplicies are the building blocks of simplicial complexes. If we glue together simplicies in a "nice enough" way, we get a simplicial complex. More formally, we have the following definition.

Definition 3.6 (Simplicial Complex) *A simplicial complex is a finite collection of simplices K such that*

1. $\sigma \in K$ *and* $\tau \leq \sigma$ *implies* $\tau \in K$ *and*
2. $\sigma_1, \sigma_2 \in K$ *implies either (i)* $\sigma_1 \cap \sigma_2 = \varnothing$ *or (ii)* $\sigma_1 \cap \sigma_2$ *is a face of both* σ_1 *and* σ_2*.*

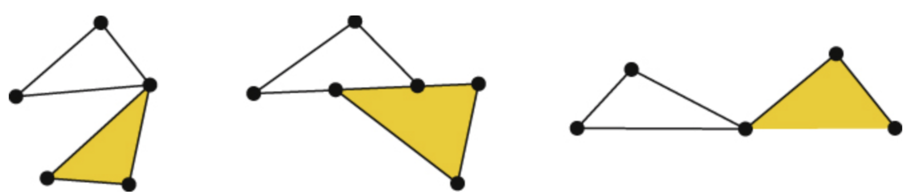


Figure 2: A simplicial complex (left) and a collection of simplices (middle and right) which do not form a simplicial complex. Figure reference: `http://iopscience.iop.org/article/10.1088/1742-5468/2009/03/P03034/fulltext/`

In practice, it is usually easier to construct a simplicial complex abstractly, without having to deal with how to place it into Euclidean space. For this reason, we have the notion of an abstract simplicial complex.

Definition 3.7 (Abstract Simplicial Complex) *An abstract simplicial complex is a finite collection of sets A such that* $\alpha \in A$ *and* $\beta \subset \alpha$ *implies* $\beta \in A$*.*

So why do we care about simplicial complexes? Simplicial complexes are the fundamental object we use to approximate the homology of a point cloud. We do this by first taking a covering of our point cloud.

Definition 3.8 (Cover) *Let X be a topological space. A cover of X is a collection of sets $U = \{U_i\}_{i \in I}$ such that $X \subset \bigcup\limits_{i \in I} U_i$.*

Definition 3.9 (Nerve of a Cover) *Let X be a topological space, and let $U = \{U_i\}_{i \in I}$ be any cover of X. The nerve of U, denoted by $\mathcal{N}(U)$, is the abstract simplicial complex with vertex set I, where a family $\{i_0, ..., i_k\}$ spans a k-simplex if and only if $U_{i_0} \cap \ldots \cap U_{i_k} \neq \varnothing$.*

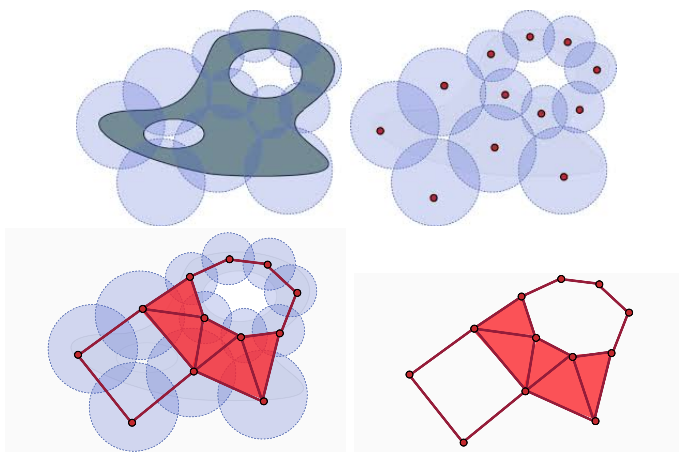


Figure 3: Cover of a finite metric space (top left), vertex set generate by cover (top right), edges and faces generated by non-empty intersections of the cover (bottom left), and the nerve of the cover (bottom right). Figure reference: `http://www.dyinglovegrape.com/math/topology_data_1.php`

A natural question to ask is, how well is the simplicial complex representation of my data set capturing the true topological features of my data? Luckily, we have the following results.

Theorem 3.1 (Nerve Theorem) *Let U be a finite collection of closed, convex sets in Euclidean space. Then, the nerve of U and the union of the sets in U have the same homotopy type.*

Now, let's discuss the common simplicial complex constructions used in topological data analysis.

Definition 3.10 (Cech Complex) *Let X be a finite set of points in $\mathbb{R}^d$. For each $x \in X$, let $B_r(x) = \{y \in \mathbb{R}^d : d(x,y) \leq r\}$ be the closed ball centered at x with radius $r \geq 0$. The Cech complex of X and r is the nerve of $\{B_r(x)\}_{x \in X}$. Namely,*

$$\mathit{Cech}(X,\ r) = \{\sigma \subset X : \bigcap_{x \in \sigma} B_r(x) \neq \varnothing\}$$

Observe, the vertex set of Cech(X, r) is precisely X itself. In Euclidean space, closed balls are closed, convex sets so the Nerve Theorem applies to the Cech Complex. A similar construction is the Vietoris Rips complex.

Definition 3.11 (Vietoris Rips Complex) *Let X be a finite set of points in $\mathbb{R}^d$. For each $x \in X$, let $B_r(x) = \{y \in \mathbb{R}^d : d(x,y) \leq r\}$ be the closed ball centered at x with radius $r \geq 0$. The Vietoris Rips complex of X and r is defined to be*

$$VR(X,r) = \{\sigma \subset X : B_r(x_i) \cap B_r(x_j) \neq \varnothing \textit{ for all } x_i, x_j \in \sigma\}$$

In other words, VR(X, r) consists of all subsets of X whose diameter is no greater than 2r. From the definitions, it is not hard to see that $\mathrm{Cech}(X,r) \subset \mathrm{VR}(X,r)$. You will also prove in the exercises that $\mathrm{VR}(X,r) \subset \mathrm{Cech}(X, \sqrt{2}r)$. Thus, while the Nerve Theorem does not apply to Vietoris-Rips complexes directly, the Vietoris-Rips complex is nested between two Cech complexes so it is reasonable to say that the Vietoris-Rips complex is a fairly good approximation. Moreover, it is computationally easier to construct the Vietoris-Rips complex than the Cech complex.

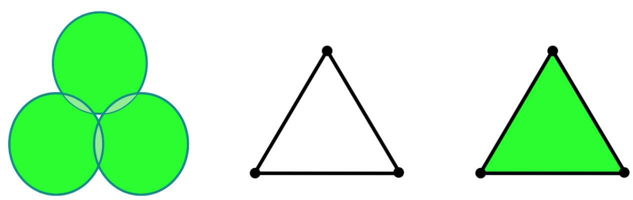


Figure 4: A cover of three points in $\mathbb{R}^2$ (left) with the corresponding Cech (middle) and Vietoris-Rips (right). Figure reference: `http://brown.edu/Research/kalisnik/Notes2.pdf`

One of the main drawbacks of the Cech and Vietoris Rips constructions is that the dimension gets very large, even if $X \subset \mathbb{R}^d$ for d small. The Delaunay and Alpha complexes offer a solution to this issue. First, we must define the Voronoi diagram.

Definition 3.12 (Voronoi Diagram) *Given a finite point set $X \subset \mathbb{R}^d$, we define the Voronoi cell of a point $p \in X$ to be*

$$V_p = \{x \in \mathbb{R}^d : d(x,p) \leq d(x,q) \ \forall \ q \in X\}.$$

The collection of all Voronoi cells is called the Voronoi diagram of X.

Note, the Voronoi diagram actually covers the entire ambient space $\mathbb{R}^d$. We now use the Voronoi diagram and the nerve construction to construct new simplicial complexes.

Definition 3.13 (Delaunay Complex) *The Delaunay complex of a finite point set $X \subset \mathbb{R}^d$ is (isomorphic to) the nerve of the Voronoi diagram. Namely,*

$$Del(X) = \{\sigma \subset X : \bigcap_{p \in \sigma} V_p \neq \varnothing\}$$

Building off of the Delaunay complex, we can use a radius constraint to get a family of subcomplexes of the Delaunay complex.

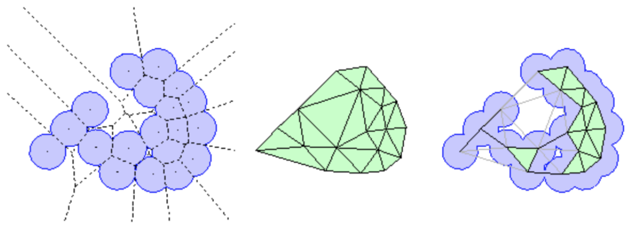


Figure 5: A Voronoi diagram of a finite set of points (left), the covex hull of the points (middle), and the corresponding Delaunay complex. Figure reference: `http://sts.bioe.uic.edu/castp/background.php`

Definition 3.14 (Alpha Complex) *Let $X \subset \mathbb{R}^d$ be a collection of finite points. For each $p \in X$, let $R_r(p) = B_r(p) \cap V_p$ be the intersection of the r-ball around p with its Voronoi cell. Then, the Alpha complex of X and r is given by*

$$Alpha(X, r) = \{\sigma \subset X : \bigcap_{p \in \sigma} R_r(p) \neq \varnothing\}$$

The last simplicial complex construction we will discuss is the Witness Complex. In practice, we have very large datasets and it is usually too expensive to compute a simplicial complex using every point in the dataset. For this reason, we instead take a subset of points, called landmark points, and build a simplicial complex using these points instead.

Definition 3.15 (Witness Complex) *Let $X \subset \mathbb{R}^d$ be any metric space, and let $L \subset X$ be a subset of X called the landmark set. Fix $\epsilon > 0$ and for every point $x \in X$, let $m_x = \min_{l \in L} d(x, l)$. Then, the witness complex $W(X, L, \epsilon)$ is a simplicial complex with vertex set L and $\{l_0, \ldots, l_k\}$ spans a k-simplex in $W(X, L, \epsilon)$ if and only if there is a point $x \in X$, called the witness, such that $d(x, l_i) \leq m_x + \epsilon$ for all $i = 0, 1, \ldots k$.*

3.1 Exercises

1. List the pros and cons of using Cech complexes, Vietoris Rips complexes, Delaunay complexes, Alpha complexes, and Witness complexes.

2. Prove: Let X be a finite set of points in some Euclidean space and let $r \geq 0$. Then,

$$\mathrm{VR}(X, r) \subset \mathrm{Cech}(X, \sqrt{2}r).$$

   Note, this is a tighter inclusion than what you saw in week 1.

3. If K is a p-dimensional simplicial complex and for each k, n_k is the number of k simplices in K, then the **Euler number** of K is given by

$$\chi(K) = n_0 - n_1 + n_2 - \ldots + (-1)^n n_p.$$

   Show that any two triangulations of the circle S^1 have the same Euler number.

4. For each positive integer n, find a simplicial complex with Euler number n. For each positive integer n, find a connected simplicial complex with Euler number n.

5. Which integers can occur as the Euler number of a one-dimensional simplicial complex?

4 Simplicial Homology

We have just seen several ways to build simplicial complexes on a finite metric space. We are interested in extracting topological features from our data. The main tool we will use to do this is called simplicial homology. Throughout the following we will use the working example provided by Sara Kalisnik ([4]) below. Note, here we have 5 0-simplices $\{a, b, c, d, e\}$, 6 1-simplices $\{A, B, C, D, E, F\}$, and 1 2-simplex $\{\tau\}$.

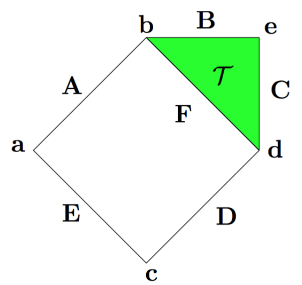


Figure 6: Simplicial complex K example to be used throughout this section. Figure reference: Sara Kalisnik ([4]).

In general, simplicial homology can be computed over any field $\mathbb{F}$. In practice, $\mathbb{F} = \mathbb{Z}/2\mathbb{Z}$.

Definition 4.1 (i-chains) *Let K be a simplicial complex. An i-chain is a formal sum of i-simplices $\sum c_i\sigma_i$, where $c_i \in \mathbb{F}$, and the sum is taken over all possible i-simplices $\sigma_i \in K$. The set of all i-chains is denoted $C_i(K)$.*

$C_i(K)$ forms a vector space over $\mathbb{F}$, and this is called the vector space of i-chains in K. In fact, the i-simplices form a basis for $C_i(K)$ and so the dimension of $C_i(K)$ is equivalent to the number of i-simplices in K. When taking $\mathbb{F} = \mathbb{Z}/2\mathbb{Z}$, the coefficients $c_i \subset \{0, 1\}$ and adding simplices is done with modulo 2 arithmetic.

Example 4.1 *Consider K given in Figure 6. Then,*

$$C_0(K) = \langle a, b, c, d, e\rangle$$
$$C_1(K) = \langle A, B, C, D, E, F\rangle$$
$$C_2(K) = \langle \tau\rangle$$

An example of a $0-chain$ is $a + c + e$. Similarly, and example of a 1-chain is $A + B + D + E$, and so on.

Given any simplex, we can also define its boundary as follow.

Definition 4.2 (Boundary of a Simplex) *Let $\sigma = [u_0, u_1, \dots u_k]$ be a k-simplex. The boundary of σ is a map $\partial_k : C_k(X) \to C_{k-1}(X)$ defined as*

$$\partial_k \sigma = \sum_{i=0}^{k} [u_0, u_1, \dots \hat{u}_j, \dots, u_k],$$

where the $\hat{u}_j$ notation indicates that u_j is omitted.

Example 4.2 *Consider K given in Figure 6.*

$$\partial(\tau) = B + F + C$$
$$\partial(A) = a + b$$

Now we have $C_i(K)$ for each i and we have maps $\partial_i : C_i(K) \to C_{i-1}(K)$. Putting this all together we get a chain complex.

Definition 4.3 (Chain Complex) *A chain complex is a sequence of chain groups connected by boundary maps:*

$$\dots \overset{\partial_{i+2}}{\to} C_{i+1}(K) \overset{\partial_{i+1}}{\to} C_i(K) \overset{\partial_i}{\to} C_{i-1}(K) \overset{\partial_{i-1}}{\to} \dots$$

Example 4.3 *Consider K given in Figure 6. Its chain complex is given by*

$$\varnothing \overset{\partial_3}{\to} \langle \tau \rangle \overset{\partial_2}{\to} \langle A, B, C, D, E, F \rangle \overset{\partial_1}{\to} \langle a, b, c, d, e \rangle \overset{\partial_0}{\to} \varnothing$$

Definition 4.4 (i-cycles) *An i-cycle is an i-chain c such that $\partial_i c = 0$.*

In other words, an i-chain is an i-cycle if it is in the kernel of the boundary map ∂_i. The set of all such i-cycles forms a subspace of $C_i(K)$, which we denote as $Z_i(K) = \ker \partial_i$.

Example 4.4 *Consider K given in Figure 6. Then,*

$$\partial(C + B + F) = d + e + e + b + b + d = (2d + 2e + 2b) mod\ 2 = 0$$

Thus, $(C + B + F)$ is a 1-chain.

Definition 4.5 (i-boundaries) *An an i-chain c is an i-boundary if there exists an $i+1$-chain $d \in C_{i+1}(K)$ such that $c = \partial_{i+1}(d)$.*

In other words, an i-chain if is an i-boundary if it is in the image of the boundary map ∂_{i+1}. The set of all such i-boundaries forms a subspace of $C_i(K)$, which we denote as $B_i(K) = \mathrm{In}\partial_{i+1}$.

The fundamental property that makes homology work is that the boundary of a boundary is necessarily zero.

Lemma 4.1 (Fundamental Lemma of Homology) *$\partial_i \circ \partial_{i+1}(d) = 0$ for all $i \in \mathbb{Z}$ and for all $i+1$-chains d.*

We now have enough machinery to define homology groups.

Definition 4.6 (i-th homology group) *The ith homology group of a simplicial complex K is the quotient group*

$$H_i(K) = Z_i(K)/B_i(K)$$

Despite their definition, the homology groups do not depend on choice of triangulation. In other words, no matter how we triangulate a given topological space, we will always get the same groups!

We are often interested in finding the rank of the homology groups. The rank of the ith homology group is the difference between the rank of the group of i-kernels and the rank of group of i-boundaries.

Definition 4.7 (i-th Betti number) *The ith Betti number, β_i is the rank of the ith homology group.*

4.1 Exercises

1. Prove the fundamental lemma of homology.
2. Construct a simplicial complex with $\beta_2 = 3$, $\beta_1 = 2$, $\beta_0 = 1$. Prove your construction is correct.
3. A tetrahedron T is a surface with 4 vertices, 6 edges, and 4 two-dimensinoal faces (note: T is not filled in). Write down the chain complex associated with T (at all dimensions). Note that all but finitely many of them are 0. The underlying field is $F_1 = \mathbb{Z}/2\mathbb{Z}$.
4. Triangulate the cylinder $S^1 \times [0,1]$ and calculate its mod 2 homology. Try a different triangulation and verify that the triangulation does not affect the homology.
5. Make your own simplicial complex and calculate its homology.

5 Persistent Homology

Many of the simplicial complex constructions depended on a parameter r. How do you know which r best captures the homology of your space? The idea behind persistent homology is that instead of choosing one value of r, we can compute the homology across a sequence of r values and study which topological features persist.

We will not review all of the algebraic machinery or the algorithm for computing persistent homology here, but rather give a general overview of the main ideas.

Say we have point cloud data sampled from an annulus. If we take the nerve of a collection of balls $B_r(x)$ centered around each point x in the point cloud, the homology of the nerve construction is highly dependent on r. For example, if r is very small, there will be little overlap between the balls and therefore the dimension of H_0 will be high and inconsistent with the true homology of an annulus. In contrast, if r is very large we will have to much overlap and the 1-dimensional hole in the center of the annulus will be filled in. Since we usually have no way to determine which r value is best, we take a range of r values to get a sequence of simplicial complexes for a point cloud data set.

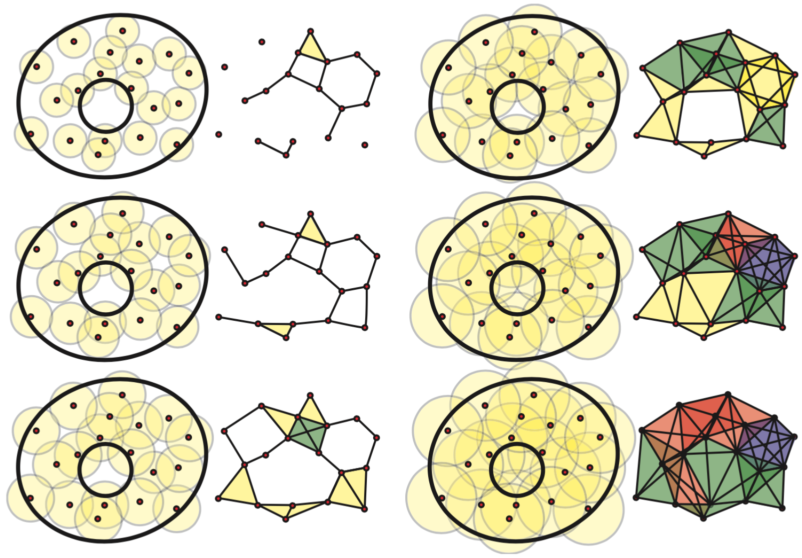


Figure 7: A sequence of Vietoris Rips complexes for a point cloud data set representing an annulus. Upon increasing the radius of the balls, we see holes appear and disappear. Which holes are real and which are noise? Figure reference: `https://www.math.upenn.edu/~ghrist/preprints/barcodes.pdf`.

In persistent homology, we get a parametrized family of topological spaces, say for example $\{VR(X,r)\}_{r\in[0,N]}$, a sequence of Vietoris Rips complexes associated to a fixed point cloud for increasing parameter values $0 \le r \le N$. From here, we get natural inclusion maps

$$VR(X,0) = VR(X,r_0) \hookrightarrow VR(X,r_1) \hookrightarrow VR(X,r_2) \hookrightarrow \ldots \hookrightarrow VR(X,r_M) = VR(X,N)$$

Instead of computing the homology for each $VR(X,r_i)$ separately, we observe that the inclusion maps above induce maps on homology $H_*VR(X,r_i) \to H_*VR(X,r_j)$ for all $i < j$. These maps reveal which features persist. As we saw in Week 1, we can preform a series of matrix reductions to compute the persistent homology of a parametrized family of topological spaces at once. We will not review this method here but you are encouraged to review the notes on Piazza.

By convention, when two features join into one feature, the feature that appeared first in the filtration persists while the other feature dies off. This is known as the Elder rule.

5.1 Exercises

1. Consider a collection of 12 equally-spaced points on a unit circle (think of tick marks on a clock). Remove from them all the points corresponding to the prime numbers $(2, 3, 5, 7, 11)$. Use the remaining points on the circle as the basis of a sequence of Vietoris-Rips complexes based on an increasing sequence $\{\epsilon_i\}$ of distances starting with $\epsilon_0 = 0$. Without worrying about the actual values of the ϵ_i, describe what happens to the sequence of VR complexes, i.e, do they change? When do they change?

Does H_0 ever increase or decrease? What about H_1?

2. Repeat exercise 1 with the Cech construction. Does anything change?

6 Representations of Persistent Homology

In persistent homology you compute the homology of a parametrized family of topological spaces. For each dimension i, you get parameter intervals arising from the basis of the ith homology group. These intervals indicate for which values of the filtration the topological feature exists.

There are two common ways of visualizing the persistent homology of a space. The first method is called a barcode diagram, and the second method is called a persistence diagram.

An example of a barcode diagram is depicted below. Each bar corresponds to a generator of homology of the parametrized family of topological spaces. The left end point of each bar corresponds to the parameter value at which the topological feature first appears and the right end point corresponds to the parameter value at which the topological feature dies off. Note, the y-axis of the barcode diagram is insignificant.

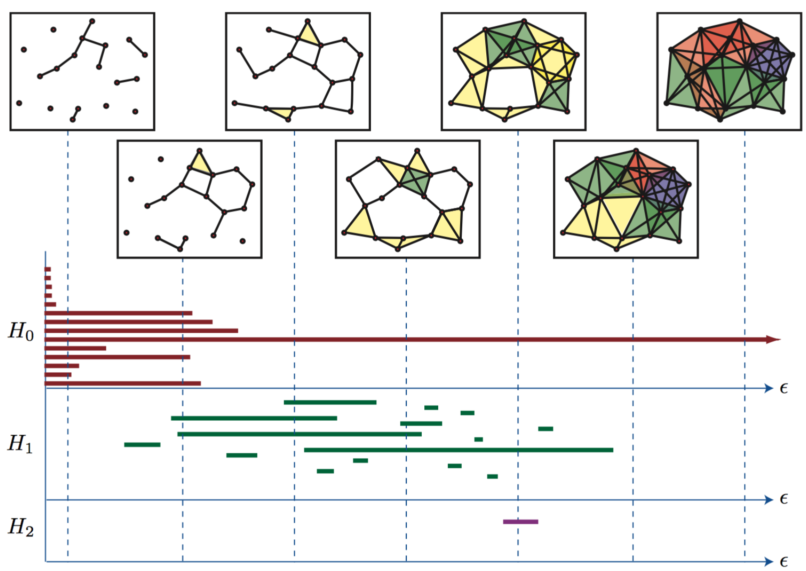


Figure 8: An example of a barcode diagram for dimensions 0, 1, and 2 of a filtered simplicial complex. Figure reference: `https://www.math.upenn.edu/~ghrist/preprints/barcodes.pdf`.

The parameter value at which a feature first appears is called the **birth time**, and the parameter value at which a feature dies off is called the **death time**. The bars in a barcode diagram can thus be represented as a collection of pairs (b_i, d_i) corresponding to the birth and death time of each bar. If we plot these points (b_i, d_i) in $\mathbb{R}^2$ we get a persistence diagram. Note, $b_i \leq d_i$ for all i so all points in a persistence diagram lie on or above the diagonal $y = x$.

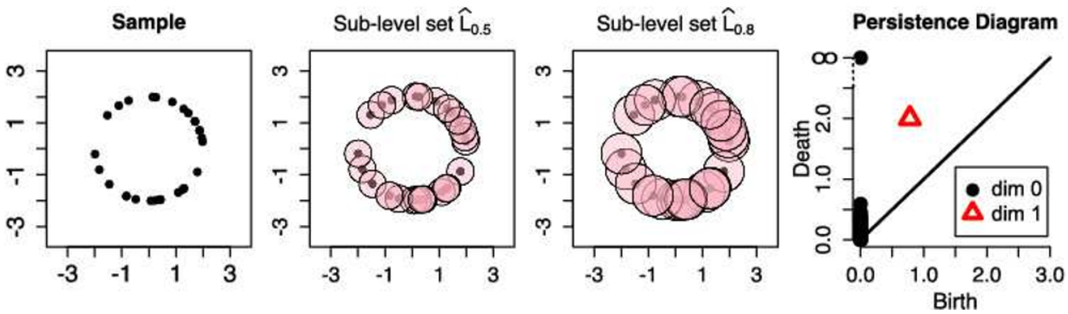
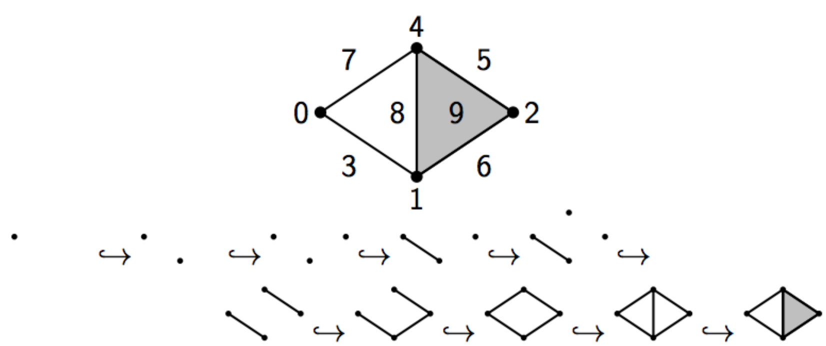


Figure 9: An example of a persistence diagram. Figure reference: `https://www.semanticscholar.org/paper/Confidence-Sets-for-Persistence-Diagrams-Fasy-Lecci/2ec2038f229c40bd37552c26545743f02fe1715d`.

6.1 Exercises

1. Given the following filtration of spaces, compute the Birth-Death pairs for H_0 and H_1. Represent your results in barcodes and the convert it into a persistent diagram.

   Hint: recall that intuitively, H_0 counts the number of connected components, and H_1 counts the number of essential loops.



7 Functional Persistence

In many applications, one might not want to apply persistent homology to a point cloud, but rather to a topological space X with an associated function $f : X \to \mathbb{R}$. For example, let X be a 3-D image of a tumor, and let $f : X \to \mathbb{R}$ be a function that associates to each point in X a measure of its radioactivity. Then, we can filter the image by radioactivity and compute the homology of the corresponding super or sub level sets of $f^{-1}((-\infty, a))$ for each

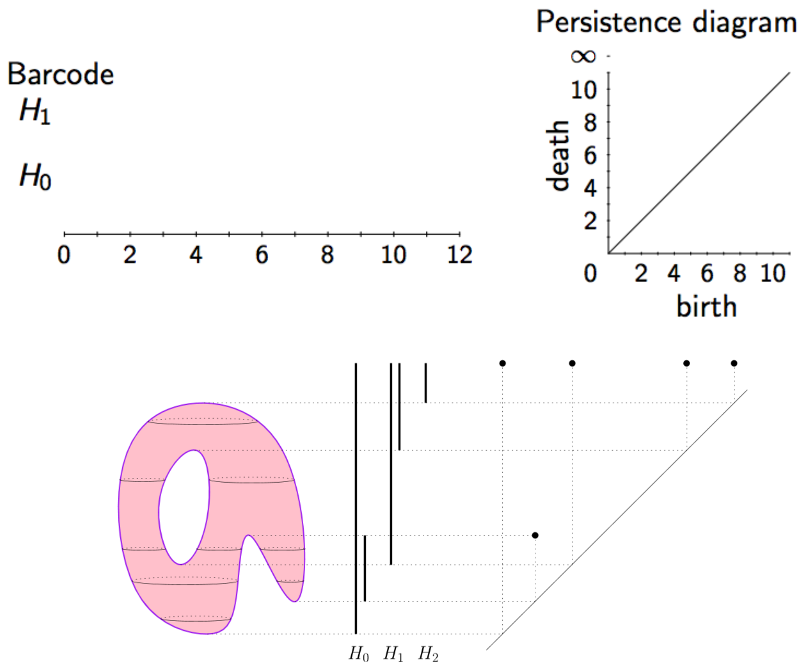


Figure 10: A topological space X filtered by its z-coordinate and its corresponding persistence barcode and persistence diagram. Figure reference: `https://pdfs.semanticscholar.org/678d/47b8392f16250ce47184cb04ebf35bc6e086.pdf?_ga=2.241750406.529015329.1499364066-110733714.1499364066`

$a \in \mathbb{R}$. This is the main idea behind functional persistence.

Formally, let X be a topological space and let $f : X \to \mathbb{R}$ be a function associated to X.

Definition 7.1 (Homological Regular Values and Homological Critical Values) *A real number a is a homological regular value if there exists an $\epsilon > 0$ such that for all $x, y \in (a - \epsilon, a + \epsilon)$ with $x < y$, the inclusion $f^{-1}(-\infty, x) \to f^{-1}(-\infty, y)$ induces isomorphisms on all homological groups. A real number a is is a homological critical value if it is not a homological regular value.*

Definition 7.2 (Tame) *A function $f : X \to \mathbb{R}$ is called tame if it has finitely many homological critical values and the homology groups $H_k(f^{-1}((-\infty, a])$ are finite dimensional for all $k \in \mathbb{N}$, $a \in \mathbb{R}$.*

7.1 Exercises

1. Define a function f on the digits $0, 1, 2, \ldots 9$ such that functional persistence distinguishes between all 10 digits.

2. Repeat the previous exercise with the English alphabet.

8 References

[1] Crossley, Martin D.. Essential Topology. 2004. Springer.

[2] Edelsbrunner, Herbert and John L. Harer. Computational Topology: An Introduction. 2010. American Mathematical Society.

[3] Hatcher, Allen. Algebraic Topology. 2001. Cambridge University Press. (Also available online freely at https://www.math.cornell.edu/~hatcher/AT/AT.pdf)

[4] Course Notes from Sara Kalisnik: http://brown.edu/Research/kalisnik/appliedtop.html

[5] https://www.math.upenn.edu/~ghrist/preprints/HAD.pdf

[6] https://people.clas.ufl.edu/peterbubenik/files/abacus_1.pdf

[7] https://www.math.upenn.edu/~ghrist/preprints/barcodes.pdf