

PERSISTENT HOMOLOGY

Luke Crossley

University of California Santa Barbara

Introduction to Homology

Homology is, fundamentally, the study and identification of holes in topological spaces. Each of these spaces can be created out of different dimensional objects called **simplicial complexes**. These complexes are constructed out of a union of points, line segments, planes, and higher dimensional polytopes, which are call n -simplices depending on the dimension of the object. Every simplicial complex, denoted as K , is a collection of subsets of the vertices of the object of interest. As discussed in [1], we will study collections of formal sums of n -simplices called **n -chains**. For simplicity, we will study these collections as vector spaces with coefficients in \mathbb{F}_2 , the field with two elements $\{0, 1\}$. Let $C_n(X)$ denote the vector space with basis given by the n -simplices of X . We can then define the **boundary map** for $n = 1, 2, \dots$ as

$$\partial_n : C_n(X) \rightarrow C_{n-1}(X)$$

which sends each n -simplex to its boundary. We also see that the boundary of a boundary is empty; therefore, the composition of any two consecutive boundary maps is the zero map. From here, we can define what is known as a **chain complex**

$$0 \xrightarrow{\partial_{n+1}} C_n(X) \xrightarrow{\partial_n} C_{n-1}(X) \xrightarrow{\partial_{n-1}} \dots \xrightarrow{\partial_2} C_1(X) \xrightarrow{\partial_1} C_0(X) \xrightarrow{\partial_0} 0$$

Our main quantity of interest is the homology vector space $H_n(X)$ which can compute as the quotient

$$H_n(X) = \frac{\ker(\partial_n)}{\text{im}(\partial_{n+1})}$$

To explain this space in simplest terms, it is the collection of holes (n -chains with trivial boundary) where we take the quotient to ensure that each hole is counted exactly once. The dimension of $H_n(X)$ gives the number of n -dimensional holes present, called the n th **Betti number** of X .

Filtered Simplicial Complexes and Persistent Homology

Persistent Homology attempts to study the homology of a space over some notion of time given a metric to determine the distance or dissimilarity between any two points in the space. To study this further, we introduce the concept of the filtered simplicial complex. Given a finite simplicial complex X and finite sequence of subcomplexes $X_1 \subset X_2 \subset \dots \subset X_k \subset X$, we call X a filtered simplicial complex. While we can compute the homology for each dimension present within each of the subcomplexes, we will discuss an algorithm that provides a more efficient computation and visualization of the significant features present in the simplicial complex. The filtered simplicial complex is created, in most cases, from a point cloud data set using something known as the **Vietoris-Rips complex** [2]. The VR complex defines some $\epsilon > 0$ and adds a 1-simplex between any two points that have a pairwise distance of less than 2ϵ based on the defined metric. Higher dimensional simplices are added as the distances between closed loops of n -simplices are less than 2ϵ . A simple way of viewing this is assuming that each point has some ϵ -neighbourhood defined around it and when the neighbourhoods of any two vertices intersects, a 1-simplex is added connecting the points. This concept be seen in the following figure, although the metric used does not follow strictly from an ϵ -neighbourhood.

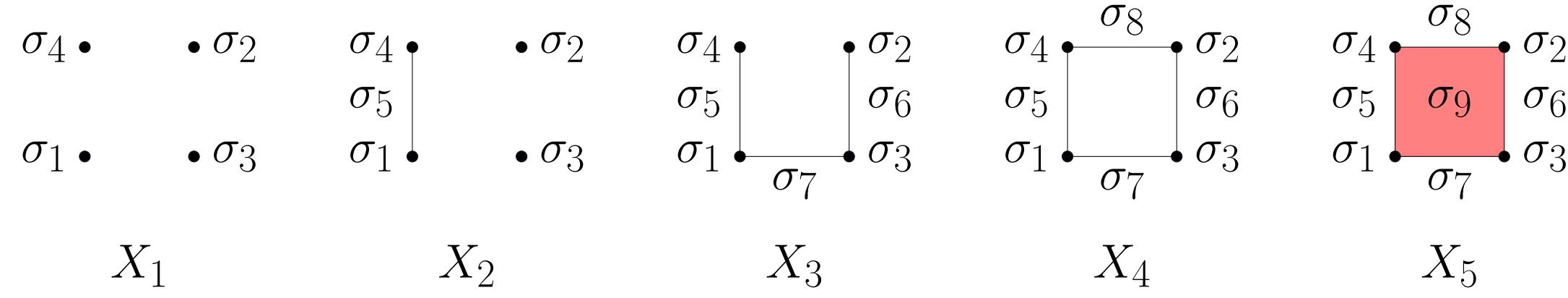


Figure 1: Visual Example of a Filtered Simplicial Complex

The Standard Algorithm and Barcode Plots

To create a graphic that displays birth and death of different features, we seek to create the barcode plot. First, we define the **boundary matrix**. Let n be the number of simplices in the simplicial complex X and denote the simplices as $\sigma_1, \dots, \sigma_n$ ordered in this manner. We construct the $n \times n$ boundary matrix $B = \{b_{i,j}\}$ where each element $b_{i,j}$ is either a 1 if σ_i is a subset of the vertices of σ_j or a 0. In [2], they discuss what they call the **Standard Algorithm** to reduce the boundary matrix. First, they define $\text{low}(j)$ to be the largest value of i such that $b_{i,j} \neq 0$ where $i, j \in \{1, \dots, n\}$. If the j th column contains only values of 0, then $\text{low}(j)$ is undefined. Now, we can discuss the standard algorithm:

```
for j = 1 to n do
    while there exists i < j and low(i) = low(j) do
        | add column i to column j
    end
end
```

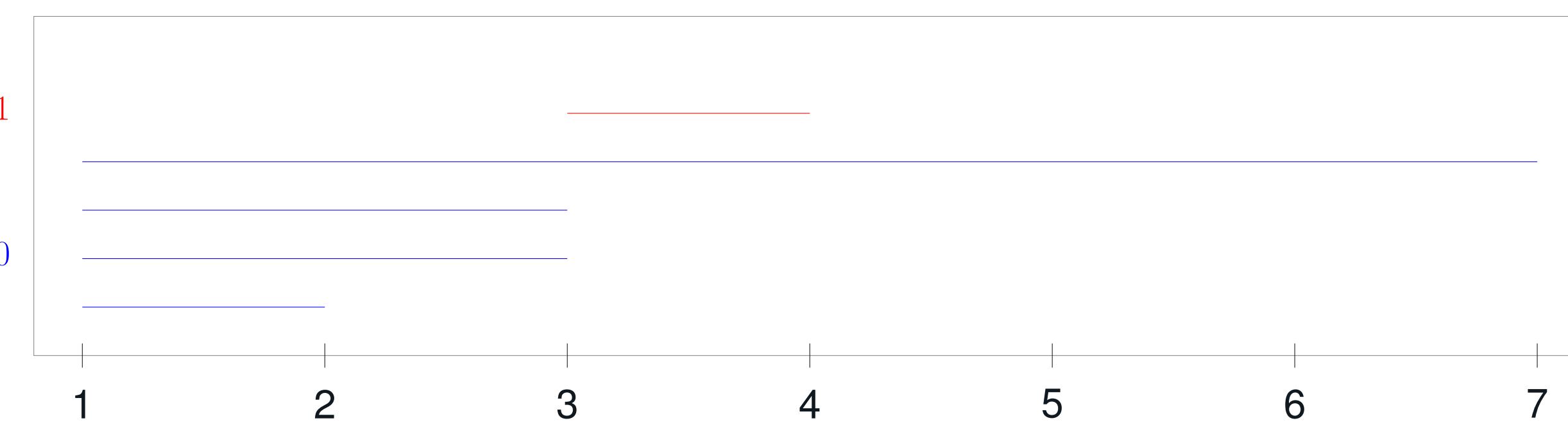
Algorithm 1: The Standard Algorithm for Reducing the Boundary Matrix

With a reduced boundary matrix B' , we can determine the **birth and death** of the features that are present throughout the filtration steps of the filtered complex. If $\text{low}(j) = i$, then we say σ_j is paired with σ_i meaning that a feature is born when σ_i enters the filtration and dies or disappears when σ_j enters into the filtration. If $\text{low}(j)$ is undefined, then σ_j entering the filtration causes a feature to be born. If there exists an l such that $\text{low}(l) = j$, then the feature born due to the entrance of σ_j dies with the entrance of σ_l into the filtration. With no such l , then the feature generated by σ_j 's entrance remains throughout the entire filtration. Given the filtered complex laid out in Figure 1, we can create the boundary matrix and follow the standard algorithm to reduce the boundary matrix. Thus, we have the boundary matrix, B , and the reduced boundary matrix, B' , from the standard algorithm (in \mathbb{F}_2):

$$B = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad B' = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Thus, we get the intervals $[1, \infty)$ for σ_1 unpaired, $[1, 3]$ for σ_2 , which is paired with σ_7 , $[1, 2]$ for σ_4 , which is paired with σ_5 , $[4, 5]$ for σ_8 paired with σ_9 , and $[1, 3]$ for σ_3 which is paired with σ_6 . These were found following the process defined in [2].

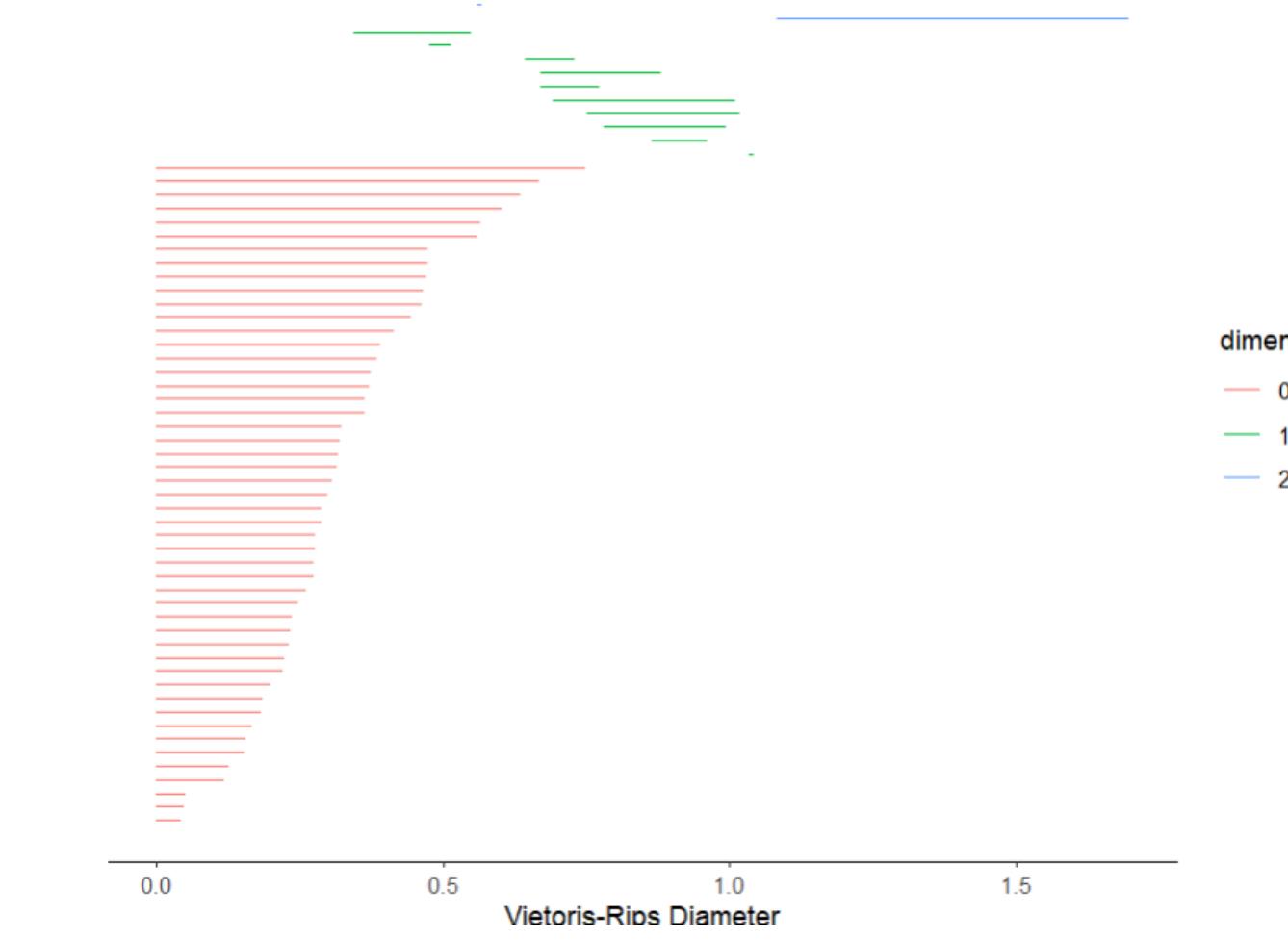
From Figure 1, we can create the **barcode plot** by reading from the intervals of the reduced boundary matrix.



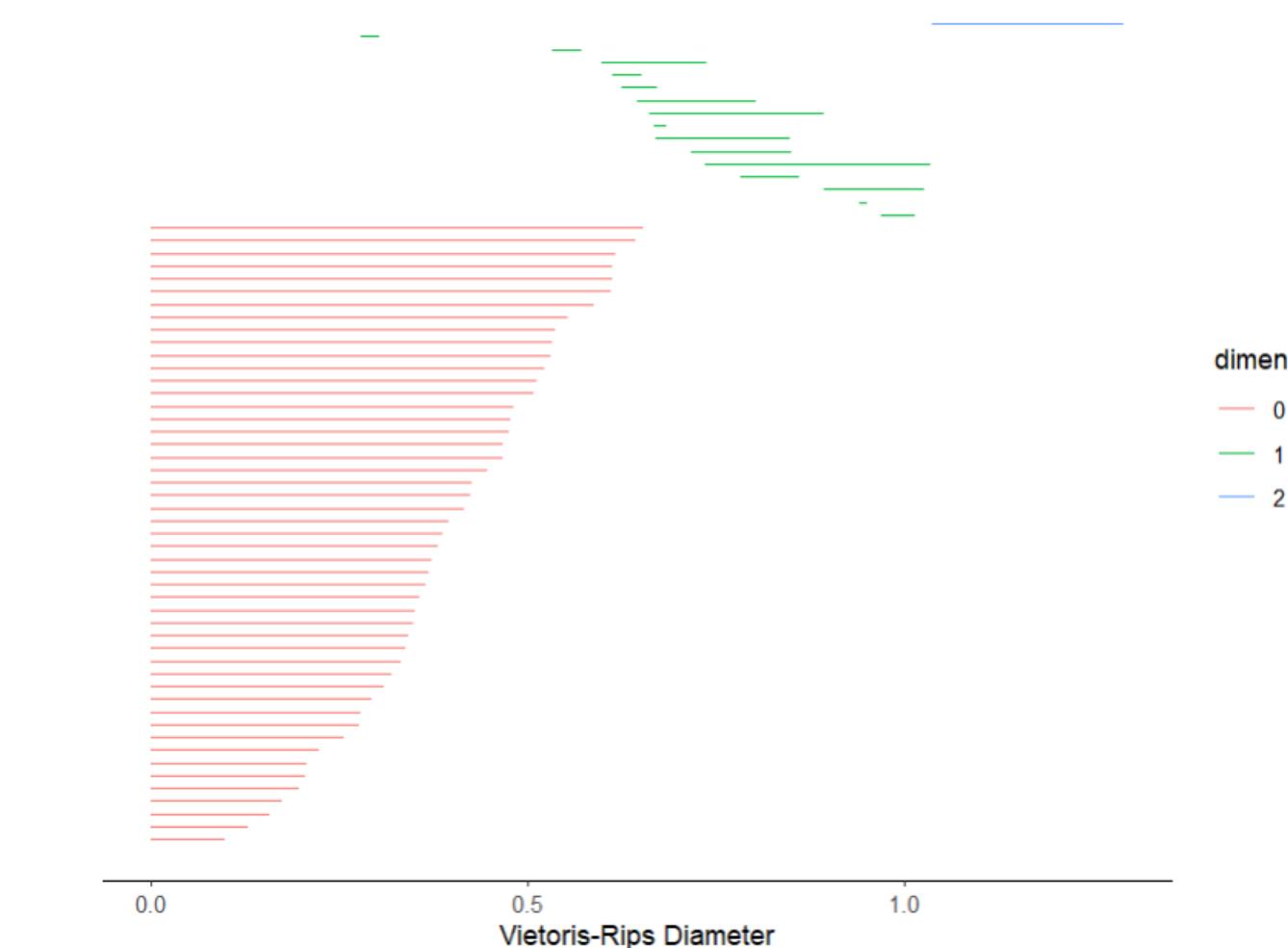
The above barcode for Figure 1 shows the presence of 0 and 1-dimensional holes in the filtered simplicial complex as well as how long these holes persist. The 1-dimensional hole persists only between 3 and 4, indicating that it is not a particularly significant feature of the space X . However, barcode plots are not dependent on which stage of the filtered complex each simplex appears at but rather the chosen metric, i.e., the value of ϵ . Thus, it is possible that ϵ increased a great deal more between 3 and 4 than between 1 and 2. Since we did not choose a specific metric, we displayed the barcode based on the subcomplexes.

Barcode Plots of Different Spheres

We now examine the differences between two different barcode plots based on two slightly differing filtered complexes. Both complexes are created by generating 50 points along the surface of S^2 . In Case 1, the points lie exactly along the surface of S^2 , while in Case 2 each point is slightly perturbed, only roughly being along the surface of S^2 . These plots were generated using the TDAstats package in R. First, we look at the barcode plot for Case 1.



We see that the longest bar in dimension 2 suggests that there is a persistent generator of H_2 . Since there are no other significant generators in either dimension 1 or 0, we can conclude that the space we are in is homologically S^2 . We can view this as pointing to the existence of the hollow inside of S^2 before the interior is filled in by a 3-simplex. Next, we look at the barcode plot of Case 2.



We see slight differences mainly in a greater dispersal of the H_1 generators with two of those generators persisting long enough that they seem like significant features that define the space we are in. While we have a persistent H_2 generator similar to that of the previous barcode plot, it is about as persistent as the H_1 generators. Thus, the perturbation of the points suggests the space we are in is no longer homologically S^2 .

Acknowledgements

I thank Andrew Ortegaray for his guidance as well as the UCSB Directed Reading Program for the opportunity to work on this project.

References

- [1] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2001.
- [2] Nina Otter et al. "A roadmap for the computation of persistent homology". In: *EPJ Data Science* (2017).