1 Long Exercises on Topological Data Analysis

1.1 Delivery 2

Find the homology groups with coefficients in Z, Q and F2 of the abstract simplicial complexes K and L whose maximal faces are, respectively.

By computing in Sagemath the following code:

```
K = SimplicialComplex([[1,2,4], [1,2, 5], [1,3,5],
                       [1,3,6], [1,4,6], [2,3,4],
                       [2,3,6], [2,5,6], [4,5,6], [3,4,5]])
K.homology()
```

we get as an output `{0: 0, 1: C2, 2: 0}`, so we have the corresponding homology groups:

$$H_0 = 0, \qquad H_1 = \mathbb{C}^2, \qquad H_2 = 0$$

Additionally, we get that the Euler characterist of K is $\chi(K) = 0$, and therefore, $K \cong \mathbb{T}$, i.e. the corresponding geometric simplicial complex defines a triangulation of a torus.

Similarly, by computing now the code

```
L = SimplicialComplex([[0,1,4], [0,1,5], [0,2,3], [0,2,7], [0,3,5], [0,4,7],
                       [1,2,6], [1,2,8], [1,4,8], [1,5,6], [2,3,6], [2,7,8],
                       [3,4,6], [3,4,8], [3,5,8], [4,6,7], [5,6,7], [5,7,8]])
L.homology()
```

we get the homology groups `{0: 0, 1: Z x Z, 2: Z}`, and thus

$$H_0 = 0, \qquad H_1 = \mathbb{Z} \times \mathbb{Z}, \qquad H_2 = \mathbb{Z}$$

By computing the Euler's characterist we have $\chi(L) = 0$ as well as in the previous case.

Note now that the previous studies are conducted over the ring of coefficients given by $\mathbb{Z}$. Now consider the same calculations but with coefficients in $\mathbb{Q}$. In order to define the abstract simplicial complexes over those fields we consider:

```
K.homology(base_ring = QQ)
```

getting as an output `{0: Vector space of dimension 0 over Rational Field, 1: Vector space of dimension 1 over Rational Field, 0: Vector space of dimension 0 over Rational Field}`, this is

$$H_0 = 0, \qquad H_1 = \mathbb{Q}, \qquad H_2 = 0$$

Similarly, by computing:

```
L.homology(base_ring = QQ)
```

we get `{0: Vector space of dimension 0 over Rational Field, 1: Vector space of dimension 2 over Rational Field, 2: Vector space of dimension 1 over Rational Field}`:

$$H_0 = 0, \qquad H_1 = \mathbb{Q} \times \mathbb{Q}, \qquad H_2 = \mathbb{Q}$$

Finally, consider the finite field of size 2 given as $\mathbb{F}_2 = GF(2) = \mathbb{Z}/2\mathbb{Z}$, also called the Galois field of size 2. Then, we can consider the corresponding homology groups over this field by implementing:

```
K.homology(base_ring=GF(2))
L.homology(base_ring=GF(2))
```

obtaining the homology groups:

$$H_0(K) = 0 \qquad H_1(K) = \mathbb{F}_2 \qquad H_2(K) = \mathbb{F}_2$$
$$H_0(L) = 0 \qquad H_1(L) = \mathbb{F}_2 \times \mathbb{F}_2 \qquad H_2(L) = \mathbb{F}_2$$

1.2 Delivery 3

Prove that, for $a < b < c$, there is an exact sequence of persistence modules $0 \to F[b,c) \to \mathbb{F}[a,c) \to \mathbb{F}[a,b) \to 0$.

Note that using Exercise 3 from Delivery 3, the first statement is direct to see, since we have:

- There exists a morphism $f_1 : \mathbb{F}[b,c) \to \mathbb{F}[a,c)$, since $a < b < c \Rightarrow a < b, b < c \leq c$.
- There exists a morphism $f_2 : \mathbb{F}[a,c) \to \mathbb{F}[a,b)$, since $a < b < c \Rightarrow a = c, a < b < c$.

Clearly we can consider the inclusion and the 0 morphism for the first and last parts of the sequence. Thus, we have the existence of the corresponding sequence of persistence modules.

1.3 Delivery 4

Draw landscape functions and a persistence silhouette for persistence module of the following point cloud in $\mathbb{R}$, R can be used.

The point cloud given is shown on Figure 1

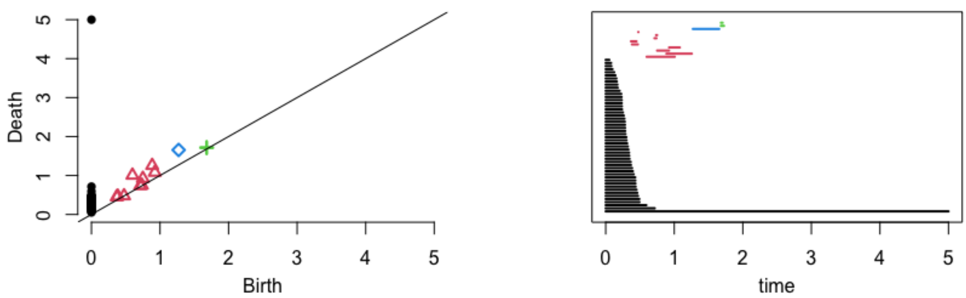

Figure 1: Persistence diagram and barcode of the Vietoris-Rips filtration of point cloud X, p.

Implementing the following code, the persistence silhouette and the corresponding landscape for different orders are conducted, as shown on the subsequent figures.

```
Diag <- ripsDiag(X1, maxdimension = 3, maxscale = 5, library = 'GUDHI')

maxscale <- 3
tseq <- seq(0, maxscale, length = 1000)   #domain
Land <- landscape(Diag[["diagram"]], dimension = 1, KK = 1, tseq)
Sil <- silhouette(Diag[["diagram"]], p = 1, dimension = 1, tseq)
par(mfrow = c(1,2))
```

```
plot(tseq, Land, type = "l")
plot(tseq, Sil, type = "l")

par(mfrow = c(2,2))
for (i in 1:4){
  Land1 = landscape(Diag[['diagram']], dimension = 1, KK=i,tseq)
  plot(tseq, Land1, type= 'l')
}
```

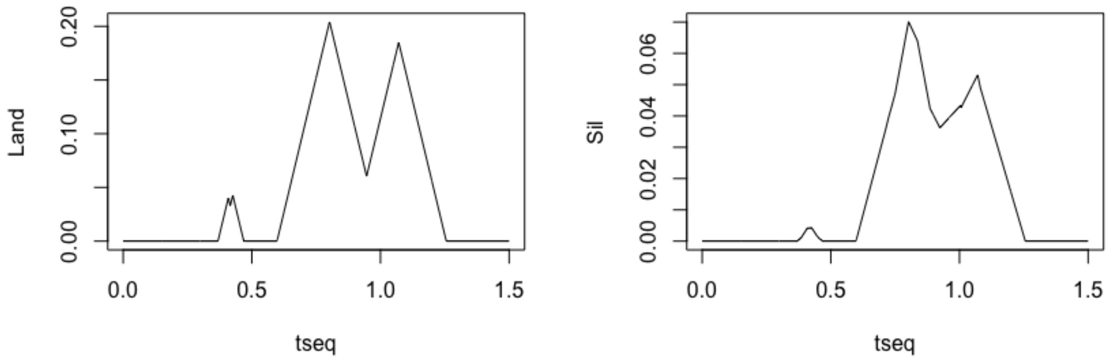

Figure 2: Landscape for order 1 (first landscape function), and silhouette of the Vietoris-Rips filtration of X

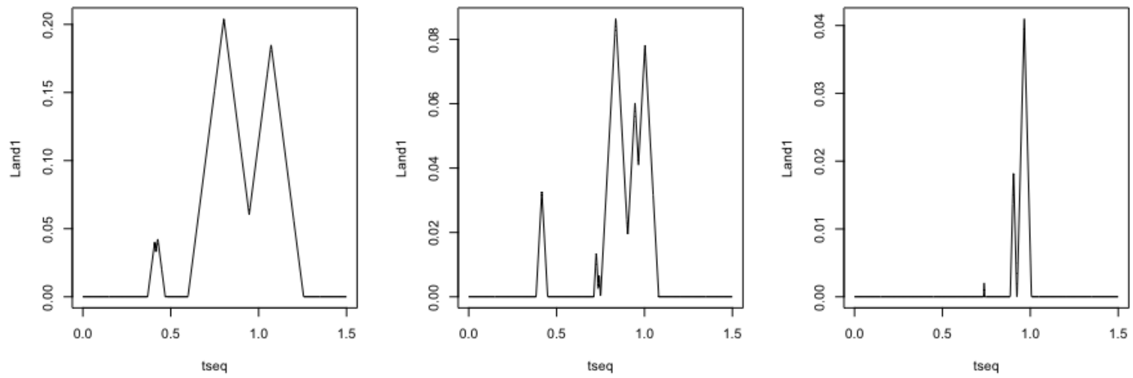

Figure 3: Landscape for orders 1 to 3 of the Vietoris-Rips filtration of X. The rest orders are constantly 0.

1.4 Delivery 5

1. Find out whether an additional set of points belongs to the same source as X_1 or to that of X_2

In order to check whether the set of points belong to one or the other point cloud we may consider the persistence diagrams of X_1 and X_2 with and without this additional set of data and compare distances. Therefore, the corresponding diagrams are presented on the subsequent figures:

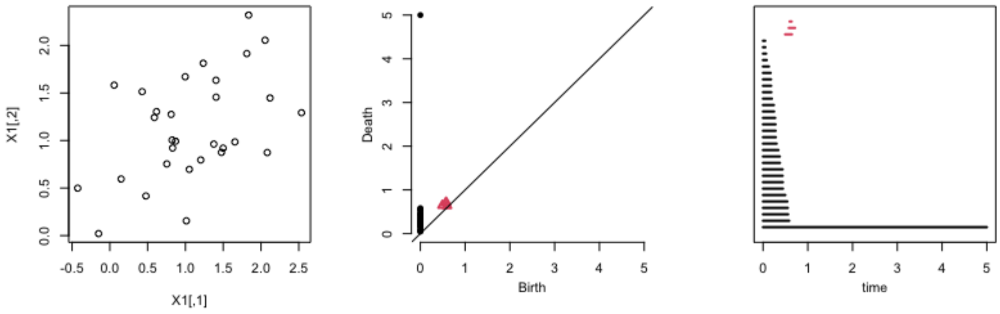

Figure 4: Point cloud X_1 original. Persistence diagram and barcode.

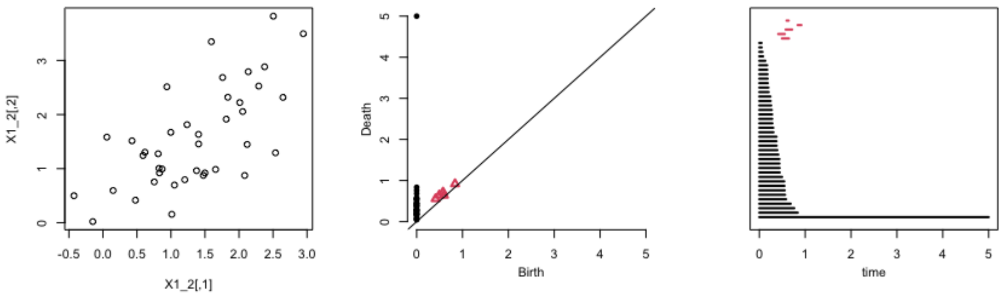

Figure 5: Point cloud X_1 with the additional set of points. Persistence diagram and barcode.

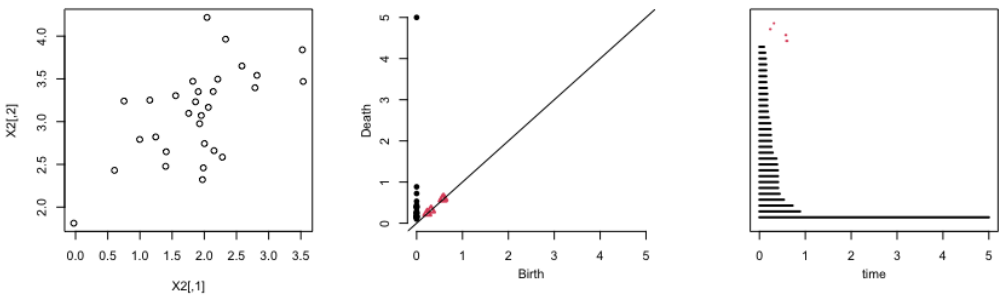

Figure 6: Point cloud X_2 original. Persistence diagram and barcode.

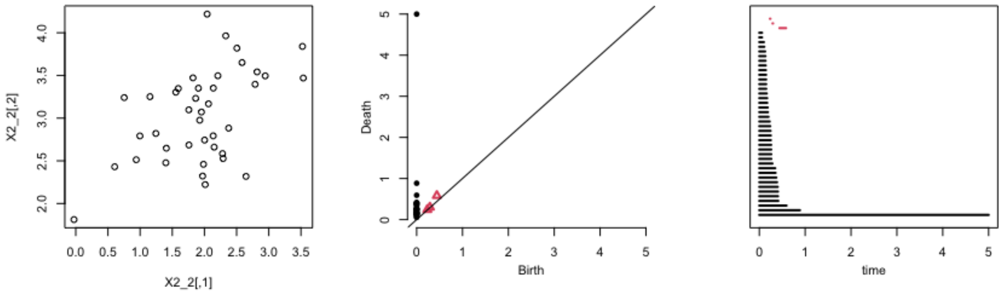

Figure 7: Point cloud X_2 with the additional set of points. Persistence diagram and barcode.

Consider now the computation of the corresponding Hausdorff distances between the sets with and without the points. Remember that, by definition, the Hausdorff distance is given by

$$d_H(X,Y) = max\{sup_{x\in X} inf_{y\in Y} d(x,y), sup_{y\in Y} inf_{x\in X} d(x,y)\}$$

Therefore, for both sets we get:

$$d_H(X_1, X_1^{(2)}) = 1.641944, \qquad d_H(X_2, X_2^{(2)}) = 0.4530203$$

Thus, we have that the set of points are most likely to belong to the point cloud X_2.