# Homology and Persistence

Only the survivors win

Roderic Guigó Corominas

August 15th 2019

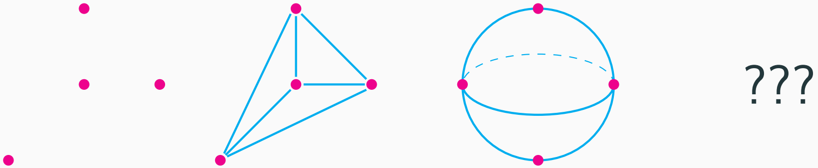
SMTB 2019

# Recall

- Simplicial complexes are the way to think of topological spaces computationally.
- Topological invariants: Betti numbers, Euler characteristic. These two are related by

$$\chi = b_0 - b_1 + b_2 - \cdots = \sum_{i=0} (-1)^i b_i$$

# Flow Chart




$$\underbrace{\text{Data} \quad \Rightarrow \quad \text{Topology}}_{\text{Today}} \quad \Rightarrow \quad \text{Properties} \quad \Rightarrow \quad \text{Meaning}$$

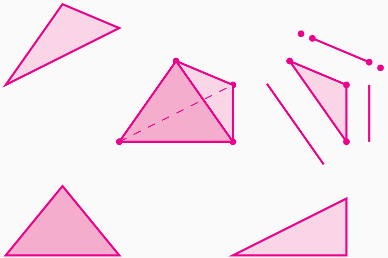
# Homology

# Homology

The true way of counting holes

# Homology: Boundaries

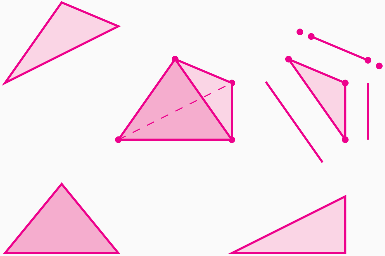
Each standard simplex has smaller simplices as boundaries



How many boundaries does an $n$-simplex have?

# Homology: Boundaries

Each standard simplex has smaller simplices as boundaries



How many boundaries does an $n$-simplex have? n+1

# Homology: Formal Sums

Fix a simplicial complex. Define $C_k$ to be the set of combinations of simpices of dimension $k$. For example in $C_1$

- $(1,2)+(2,3)$ is a valid combination but not $(1,2)+(1,2,3)$.
- $(1,2)+(1,2)=2(1,2)$
- $(1,2)-(1,2)=0$

In $C_3$ we could have

$$(1,4,5)-3(2,3,4)+2(1,2,5).$$

# Homology: Boundaries

There is an operator called boundary

$$\partial \colon C_k \to C_{k-1}$$

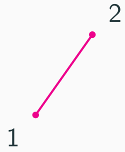
defined by

$$\partial(v_0, v_1, v_2, \ldots, v_k) = (v_1, \ldots, v_k) - (v_0, v_2, \ldots, v_k) + \ldots$$

Clearly the left hand side is a $k$-simplex, and the right hand side is a combination of $k-1$ simplices.

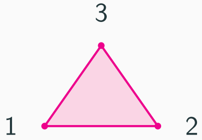
# Homology

Boundary of an interval (1-simplex):




$$\partial(1,2) = (2) - (1)$$

Boundary of a 2-simplex:



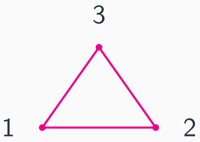

$$\partial(1,2,3) = (2,3) - (1,3) + (1,2)$$

# Homology

Two points (1) and (2) are connected if (2) − (1) is the boundary of an interval. (3) is not connected to either (1) or (2):

# Homology

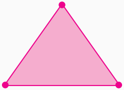
A loop is something that has no boundary...




$$\partial((1,2)+(2,3)-(1,3)) = (2)-(1)+(3)-(2)-(3)+(1) = 0.$$

# Homology

.. and that is not a boundary!!! If there isn't a face inside:



In other words, a combination of simplices without boundary defines a loop if it is not a boundary of a higher simplex!

# Homology

Generalising this to all $k$-simplicies:

- Potential holes are combinations of simplices without boundary.
- If they are filled, they are not truly holes.
- What are the $k$-holes? They are things in $C_k$ that have 0 boundary (cycles) and that are not boundaries of anything in $C_{k+1}$ themselves.

# Example: Sphere

- 2-holes:

$$\partial(L) = a + b - c, \partial(U) = a + b - c$$

  Therefore $L - U$ has no boundary, since $\partial(L) - \partial(U) = 0$. Since there isn't a 3-simplex $L - U$ is a hole.

- 1-holes

$$\partial(a) = v - w, \partial(b) = u - v, \partial(c) = u - w$$

  Therefore $a + b - c$ has no boundary, but it is the boundary of $a + b - c$, so no holes.

- Components.

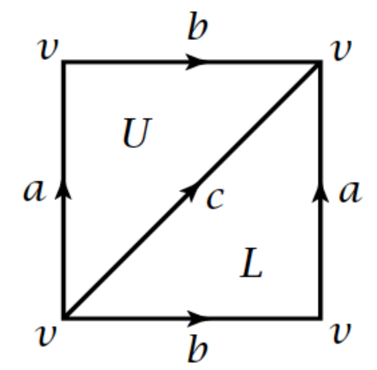
$$\partial(u) = \partial(v) = \partial(w) = 0.$$

  But $v - w$ and $u - v$ are boundaries, so only one component.

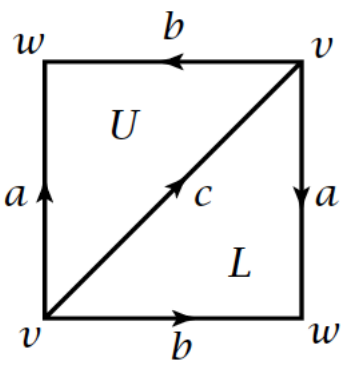
# Example: Sphere

- 2-holes: $L - U$, so $b_2 = 1$.
- 1-holes: None, so $b_1 = 0$.
- Components: $u$, so $b_0 = 1$.

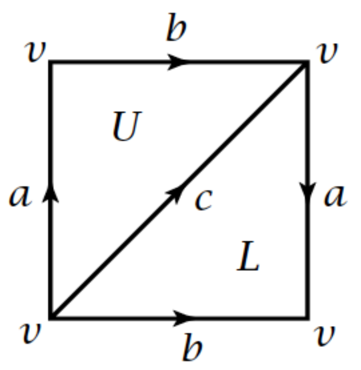
# Example: Torus



# Example: Projective Plane



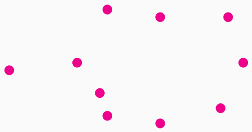
# Example: Klein bottle



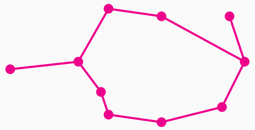
# Persistence

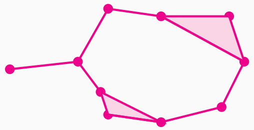
# From data to topology

How do we get a shape from a set of point cloud data?



Connect the dots. But what simplex should we assign?





# Geometry

Insights:

- Data points belong to $\mathbb{R}^n$.
- There is a notion of distance in $\mathbb{R}^n$ (Pitagoras):

$$d((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$$

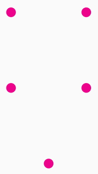
- Two dots should be connected if they are close enough.
- Vary distance and study how it changes.

# Geometry: Cech Complex

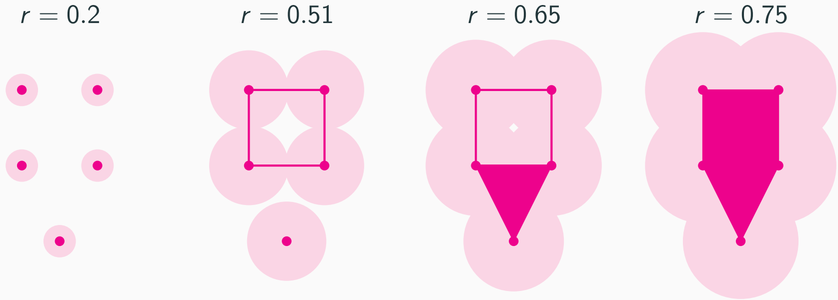
Choose a distance $r$ and draw around each point a circle of radius $r$, and connect points accordingly.

- If the two circles intersect, then connect the corresponding vertices with an edge.
- If three circles intersect, add a face.
- If fourse circles intersect, add a 3-simplex.
- $\vdots$

For example, consider points



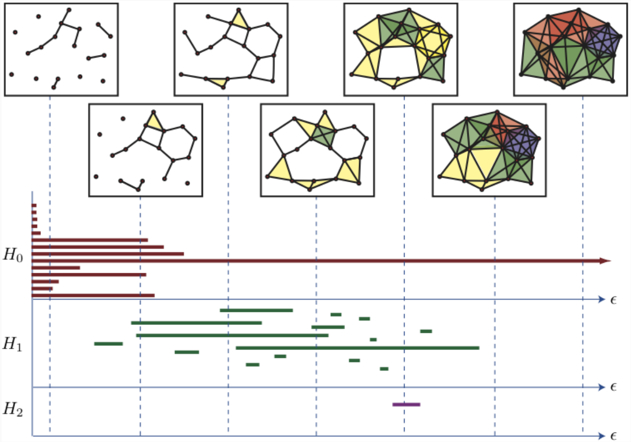
# Geometry: Cech Complex



# Persistent Homology

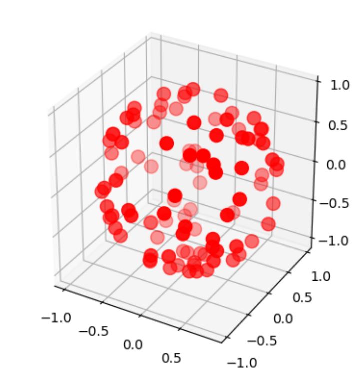
- For $r = 0$ everything is disconnected.
- For $r$ large enough, everything is filled.
- All holes created in the process die, but, for how long do they persist?
- Persistence in time should be related to the relevance of the holes.

# Barcode

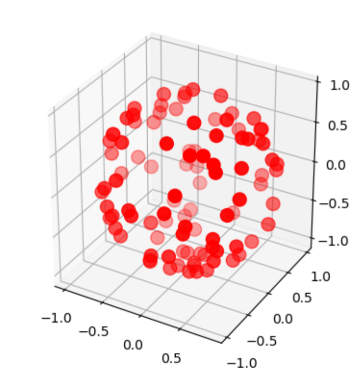


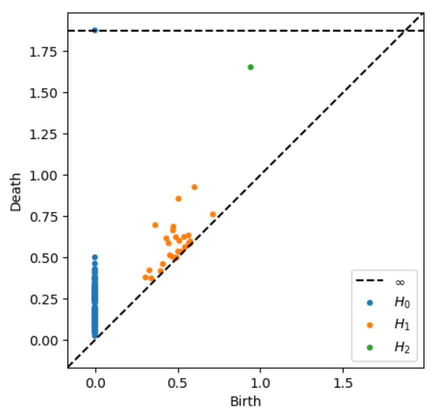

https://www.youtube.com/watch?v=CKfUzmznd9g

# Sampling a sphere

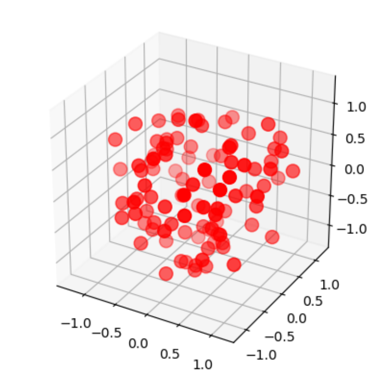


# Sampling a sphere



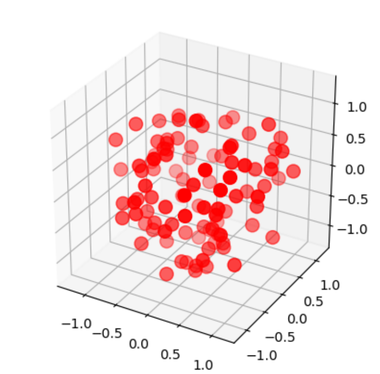



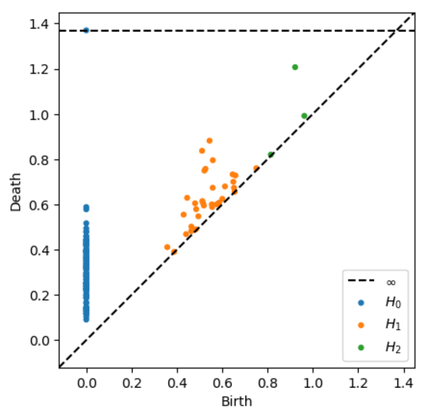
# Sampling a sphere with noise




A little noise doesn't affect too much the results.

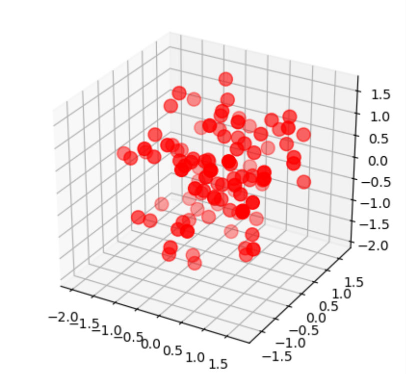
# Sampling a sphere with noise



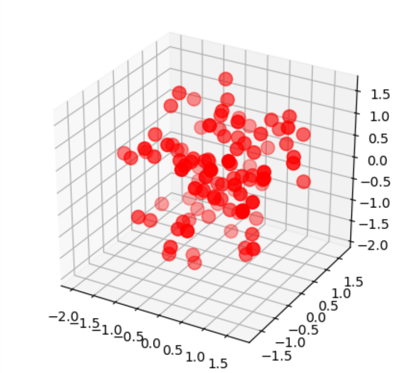

A little noise doesn't affect too much the results.

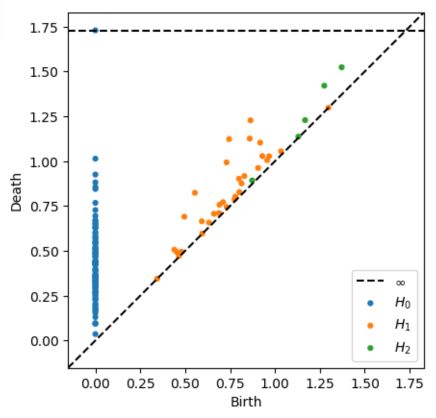
# Sampling a sphere with noise




If there is too much noise, no holes persist.

# Sampling a sphere with noise




If there is too much noise, no holes persist.

# Next Time: Python

**Run some Python code!**

To run the code below:

1. Click on the cell to select it.
2. Press SHIFT+ENTER on your keyboard or press the play button (▶) in the toolbar above.

A full tutorial for using the notebook interface is available here.

In [1]:

```python
%matplotlib inline

import pandas as pd
import numpy as np
import matplotlib

from matplotlib import pyplot as plt
import seaborn as sns

ts = pd.Series(np.random.randn(1000), index=pd.date_range('1/1/2000', periods=1000))
ts = ts.cumsum()

df = pd.DataFrame(np.random.randn(1000, 4), index=ts.index,
                  columns=['A', 'B', 'C', 'D'])
df = df.cumsum()
plt.figure(); df.plot(); plt.legend(loc='best')
```

Out[1]: <matplotlib.legend.Legend at 0x7fb27b72fcc0>

<matplotlib.figure.Figure at 0x7fb283672b70>

**Questions?**