Topological data analysis so far:

Topological data analysis so far:

$$\textit{data} \longrightarrow \textit{distance matrix}$$

Topological data analysis so far:

$$\textit{data} \longrightarrow \textit{distance matrix} \longrightarrow \textit{simplicial complex}$$

Topological data analysis so far:

$$\textit{data} \longrightarrow \textit{distance matrix} \longrightarrow \textit{simplicial complex}$$

using the **clique complex**

Topological data analysis so far:

$$\textit{data} \longrightarrow \textit{distance matrix} \longrightarrow \textit{simplicial complex}$$

using the **clique complex** *or the* **nerve of a cover**.

Topological data analysis so far:

$$\textit{data} \longrightarrow \textit{distance matrix} \longrightarrow \textit{simplicial complex}$$

using the **clique complex** *or the* **nerve of a cover**.

The **graph** *of the simplicial complex can be visualized*

Topological data analysis so far:

$$\textit{data} \longrightarrow \textit{distance matrix} \longrightarrow \textit{simplicial complex}$$

using the **clique complex** *or the* **nerve of a cover**.

The **graph** *of the simplicial complex can be visualized, and used to compute* **connected components/clusters**

Topological data analysis so far:

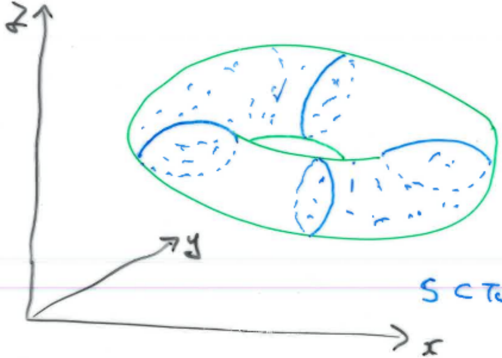
$$\textit{data} \longrightarrow \textit{distance matrix} \longrightarrow \textit{simplicial complex}$$

using the **clique complex** *or the* **nerve of a cover**.

The **graph** *of the simplicial complex can be visualized, and used to compute* **connected components/clusters**

We use **barcodes** *to visualize the persistence of clusters.*

Example Consider a sample S of 750 points selected at random from two "quarter segments" of a torus.



$S \subset \text{Torus} \subset \mathbb{R}^3$

Example 200 people were asked to visit Galway harbour at their convenience once during a 2-week period. They were asked to record the height of the water on their arrival, and then again 2 hours later, and then again 4 hours after their arrival. Each person returns their recordings $(h_0, h_2, h_4) \in \mathbb{R}^3$. From the set $S = \{x_i = (h_{i_0}, h_{i_2}, h_{i_4}) : 1 \le i \le 200\}$ a data analyst can construct

$D = (d_{ij})$ with

$d_{ij} = \| x_i - x_j \|$ Euclidean metric,

The analyst could view the graph of the simplicial complex K_ε for various values of ε.

What do we mean by a "circle" or "hole" in a simplicial complex K ?

Choose a field (such as $\mathbb{R}$, $\mathbb{Q}$ or

$\mathbb{Z}_2$) and let

$$C_n K$$

denote the vector space (over the given field) with one basis element

$$e_\sigma$$

for each n-simplex

$$\sigma = \{v_0, v_1, \dots, v_n\} \in K.$$

$$C_n K$$

denote the vector space (over the given field) with one basis element

$$e_\sigma$$

for each n-simplex

$$\sigma = \{v_0, v_1, \dots, v_n\} \in K.$$

Here we assume the vertices are ordered.

Let

$$d_n : C_n k \longrightarrow C_{n-1} k$$

be the linear homomorphism

defined on basis elements by

$$d_n e_\sigma = \sum_{i=0}^{n} (-1)^i \, e_{\sigma \setminus \{v_i\}}$$

Defn The degree n **Betti number** of K is defined as

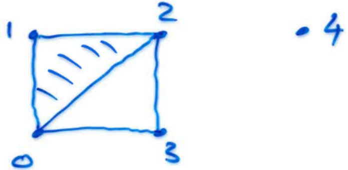
$$\beta_n = \mathrm{Dim}(\ker d_n) - \mathrm{Dim}(\mathrm{Image}\, d_{n+1})$$

Informally, we say that K has "β_n n-dimensional holes" over the given field.

Also, a "0-dimensional hole" is a connected component.

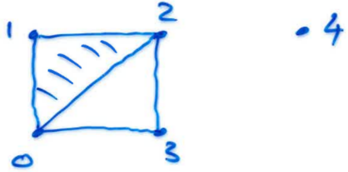
Example

K



Example

K



$$
\begin{aligned}
K \quad = \quad \{ \quad & \{0\}, \{1\}, \{2\}, \{3\}, \{4\}, \\
& \{0,1\}, \{0,2\}, \{0,3\}, \{1,2\}, \{2,3\}, \\
& \{0,1,2\} \ \}
\end{aligned}
$$

Let's work over $\mathbb{R}$.

$$C_2K \xrightarrow{d_2} C_1K \xrightarrow{d_1} C_0K \xrightarrow{d_0} 0$$

$$\mathbb{R} \qquad\qquad \mathbb{R}^5 \qquad\qquad \mathbb{R}^5$$

Let's work over $\mathbb{R}$.

$$C_2 K \xrightarrow{d_2} C_1 K \xrightarrow{d_1} C_0 K \xrightarrow{d_0} 0$$

$$\mathbb{R} \qquad\qquad \mathbb{R}^5 \qquad\qquad \mathbb{R}^5$$

$$d_2 e_{\{0,1,2\}} = e_{\{1,2\}} - e_{\{0,2\}} + e_{\{0,1\}}$$

$$
\begin{aligned}
d_1 : e_{\{0,1\}} &\longmapsto e_{\{1\}} - e_{\{0\}} \\
e_{\{0\,2\}} &\longmapsto e_{\{2\}} - e_{\{0\}} \\
e_{\{0,3\}} &\longmapsto e_{\{3\}} - e_{\{0\}} \\
e_{\{1,2\}} &\longmapsto e_{\{2\}} - e_{\{1\}} \\
e_{\{2,3\}} &\longmapsto e_{\{3\}} - e_{\{2\}}
\end{aligned}
$$

d_2, d_1 have matrices

$$D_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

$$D_1 = \begin{pmatrix} -1 & -1 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\beta_1 = \mathrm{Dim}(\ker d_1) - \mathrm{Dim}(\mathrm{Image}\ d_2)$$

$$= (5 - \mathrm{rank}\, D_1) \quad - \mathrm{rank}\,(D_2)$$

$$= (5 - 3) - 1$$

$$= 1$$

$$\beta_1 = \mathrm{Dim}(\ker d_1) - \mathrm{Dim}(\mathrm{Image}\ d_2)$$

$$= (5 - \mathrm{rank}\, D_1) \quad - \mathrm{rank}\,(D_2)$$

$$= (5 - 3) - 1 .$$

$$= 1$$

i.e. there is one 1-dimensional hole in K.

$$
\begin{aligned}
\beta_0 &= \mathrm{Dim}(\mathrm{Ker}\, d_0) - \mathrm{Dim}(\mathrm{Image}\, d_1) \\
&= 5 - \mathrm{rank}(D_1) \\
&= 5 - 3 \\
&= 2
\end{aligned}
$$

$$\beta_0 = \mathrm{Dim}(\mathrm{Ker}\, d_0) - \mathrm{Dim}(\mathrm{Image}\, d_1)$$

$$= 5 - \mathrm{rank}(D_1)$$

$$= 5 - 3$$

$$= 2$$

i.e. K has two connected components (or 0-dimensional holes)