MAPPER CLUSTERING

For an unkown population space X choose:

MAPPER CLUSTERING

For an unkown population space X choose:

- finite sample $S \subset X$

MAPPER CLUSTERING

For an unkown population space X choose:

- finite sample $S \subset X$
- metric $d_X(x, y)$ (evaluated only for $x, y \in S$)

MAPPER CLUSTERING

For an unkown population space X choose:

- finite sample $S \subset X$
- metric $d_X(x, y)$ (evaluated only for $x, y \in S$)
- continuous map $f : X \to Z$ (evaluated only for $x \in S$)

MAPPER CLUSTERING

For an unkown population space X choose:

- finite sample $S \subset X$
- metric $d_X(x, y)$ (evaluated only for $x, y \in S$)
- continuous map $f\colon X \to Z$ (evaluated only for $x \in S$)
- open cover $\{V_\alpha\}_{\alpha \in A}$ of Z

MAPPER CLUSTERING

For an unkown population space X choose:

- finite sample $S \subset X$
- metric $d_X(x, y)$ (evaluated only for $x, y \in S$)
- continuous map $f\colon X \to Z$ (evaluated only for $x \in S$)
- open cover $\{V_\alpha\}_{\alpha \in A}$ of Z

Cluster each $f^{-1}(V_\alpha) \cap S \;=\; S_{\alpha,1} \sqcup S_{\alpha,2} \sqcup \cdots \sqcup S_{\alpha,n_\alpha}$

MAPPER CLUSTERING

For an unkown population space X choose:

- finite sample $S \subset X$
- metric $d_X(x, y)$ (evaluated only for $x, y \in S$)
- continuous map $f \colon X \to Z$ (evaluated only for $x \in S$)
- open cover $\{V_\alpha\}_{\alpha \in A}$ of Z

Cluster each $f^{-1}(V_\alpha) \cap S \;=\; S_{\alpha,1} \sqcup S_{\alpha,2} \sqcup \cdots \sqcup S_{\alpha,n_\alpha}$

Output

$$K = \mathrm{Nerve}(\ \{S_{\alpha,j}\}_{\alpha \in A, 1 \leq j \leq n_\alpha}\ )$$

Example 1 : $\mathbf{S} = \{\mathbf{v}_1, \ldots, \mathbf{v}_{200}\} \subset \mathbf{X} \subset \mathbb{R}^2$

Example 1 : $\mathbf{S} = \{\mathbf{v}_1, \ldots, \mathbf{v}_{200}\} \subset \mathbf{X} \subset \mathbb{R}^2$

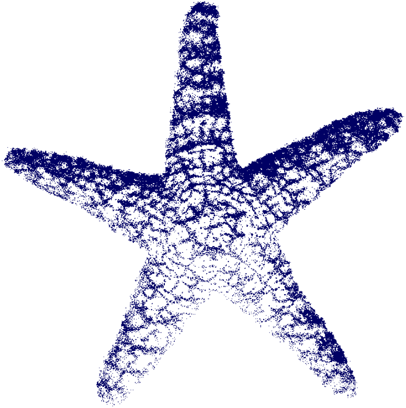
$\{V_\alpha\}_{\alpha \in A}$ an open cover of $Z = [0, \infty)$ with no triple overlaps

Example 1 : $\mathbf{S} = \{\mathbf{v_1}, \ldots, \mathbf{v_{200}}\} \subset \mathbf{X} \subset \mathbb{R}^2$

$\{V_\alpha\}_{\alpha \in A}$ an open cover of $Z = [0, \infty)$ with no triple overlaps

$$f : X \to [0, \infty), x \mapsto d_X(v_1, x)$$

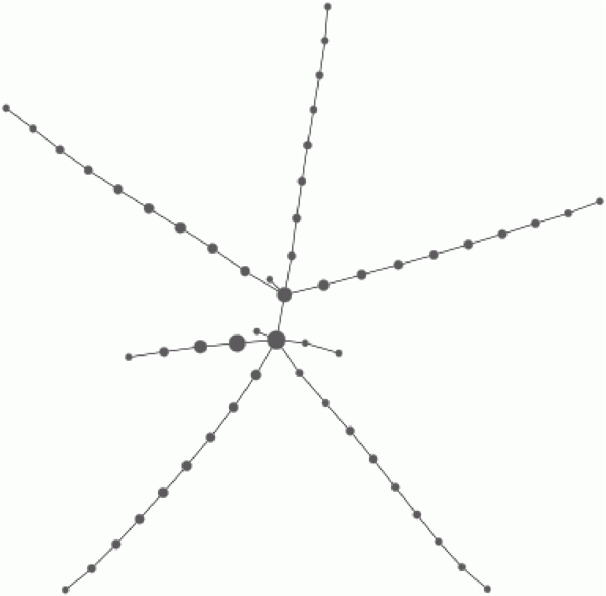
Example 1 : $\mathbf{S} = \{\mathbf{v_1}, \ldots, \mathbf{v_{200}}\} \subset \mathbf{X} \subset \mathbb{R}^2$

$X =$ 

$\{V_\alpha\}_{\alpha \in A}$ an open cover of $Z = [0, \infty)$ with no triple overlaps

$$f : X \to [0, \infty), x \mapsto d_X(v_1, x)$$

Mapper output for starfish sample



Example 2 : $\mathbf{S} = \{\mathbf{v_1}, \ldots, \mathbf{v_{310}}\} \subset \mathbf{X} \subset$ all breast tissue samples

Example 2 : $\mathbf{S} = \{\mathbf{v_1}, \ldots, \mathbf{v_{310}}\} \subset \mathbf{X} \subset$ all breast tissue samples

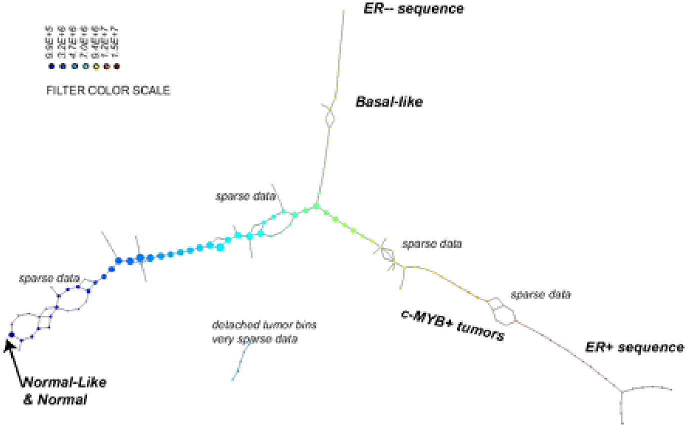
$\{V_\alpha\}_{\alpha \in A}$ an open cover of $Z = [0, \infty)$ with no triple overlaps

Example 2 : $\mathbf{S} = \{\mathbf{v_1}, \ldots, \mathbf{v_{310}}\} \subset \mathbf{X} \subset \text{all breast tissue samples}$

$\{V_\alpha\}_{\alpha \in A}$ an open cover of $Z = [0, \infty)$ with no triple overlaps

$$f : X \to [0, \infty), x \mapsto d_X(v_1, x) \quad \text{with } v_1 \text{ healthy}$$

Example 2 : $\mathbf{S} = \{\mathbf{v}_1, \ldots, \mathbf{v}_{310}\} \subset \mathbf{X} \subset$ all breast tissue samples



$\{V_\alpha\}_{\alpha \in A}$ an open cover of $Z = [0, \infty)$ with no triple overlaps

$$f : X \to [0, \infty), x \mapsto d_X(v_1, x) \quad \text{with } v_1 \text{ healthy}$$

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

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

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

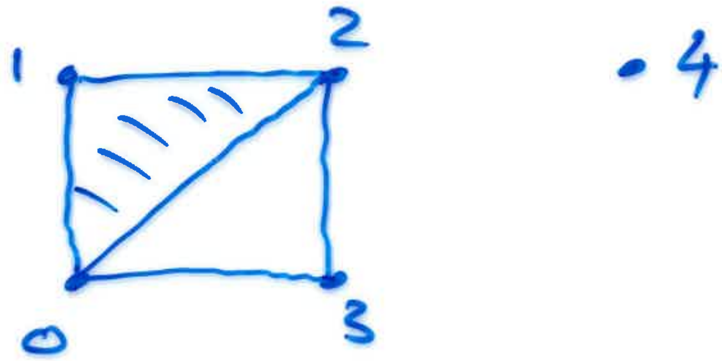
$$\beta_n = \mathrm{Dim}(\ker d_n) - \mathrm{Dim}(\mathrm{Image}\, d_{n+1})$$

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



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

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

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

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

$$d_1 : e_{\{0,1\}} \longmapsto e_{\{1\}} - e_{\{0\}}$$

$$e_{\{0\,2\}} \longmapsto e_{\{2\}} - e_{\{0\}}$$

$$e_{\{0,3\}} \longmapsto e_{\{3\}} - e_{\{0\}}$$

$$e_{\{1,2\}} \longmapsto e_{\{2\}} - e_{\{1\}}$$

$$e_{\{2,3\}} \longmapsto e_{\{3\}} - e_{\{2\}}$$

d_2, d_1 have matrices

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

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

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

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

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

$$= 1$$

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

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

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

$$= 5 - 3$$

$$= 2$$

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