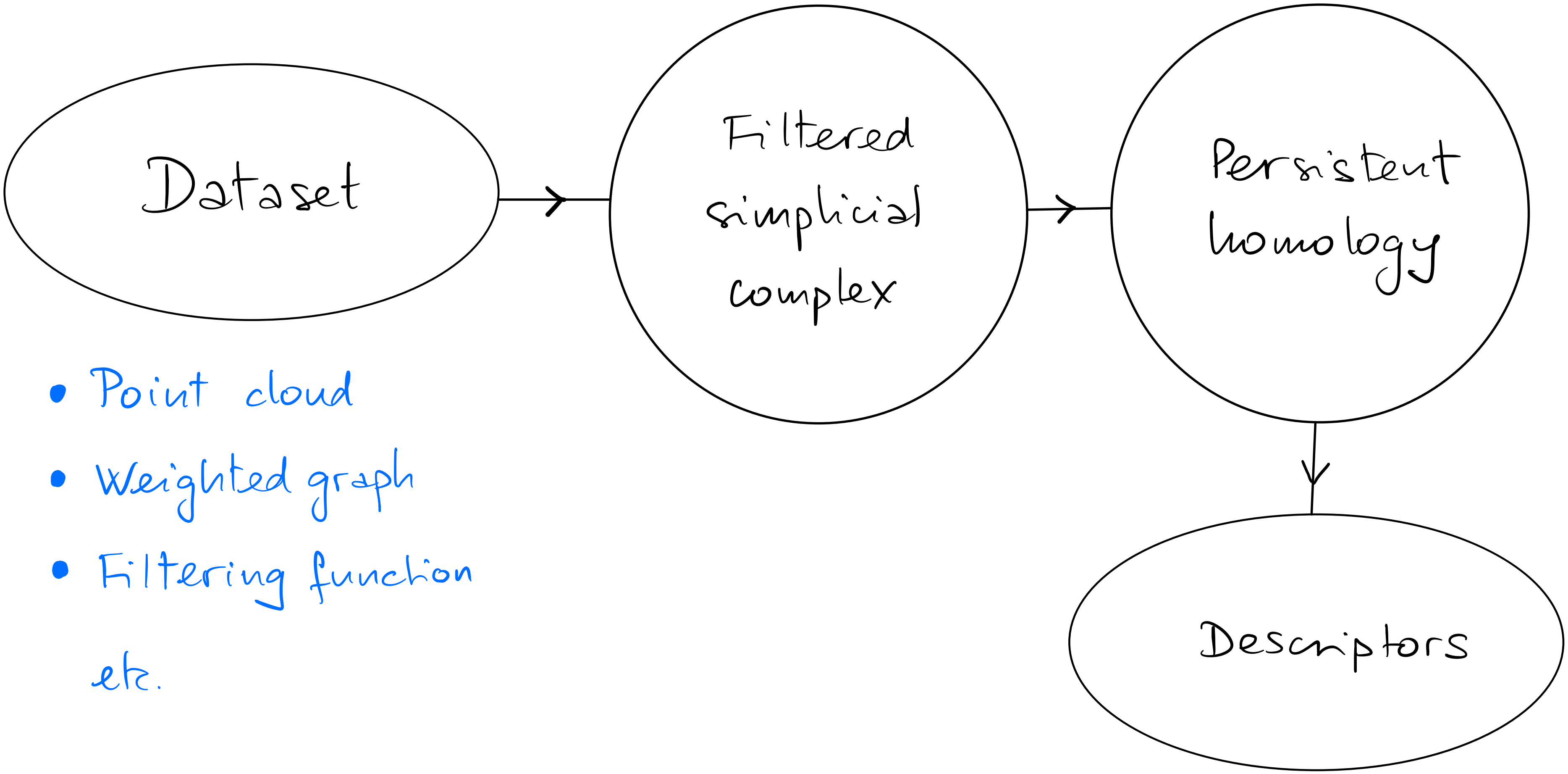
Topological Data Analysis

2022–2023

Lecture 1

Simplicial Complexes from Datasets

3 November 2022


Dataset
Filtered simplicial complex
Persistent homology
Descriptors
• Point cloud
• Weighted graph
• Filtering function
etc.

Abstract simplicial complexes

An *abstract simplicial complex* with vertex set $V = \{v_i\}_{i \in I}$ is a collection K of nonempty finite subsets $\{v_{i_0}, \dots, v_{i_n}\} \subseteq V$ such that

a) $\{v\} \in K \quad \forall\, v \in V$;

b) if $S \in K$ and $S' \subseteq S$ with $S' \neq \emptyset$, then $S' \in K$.

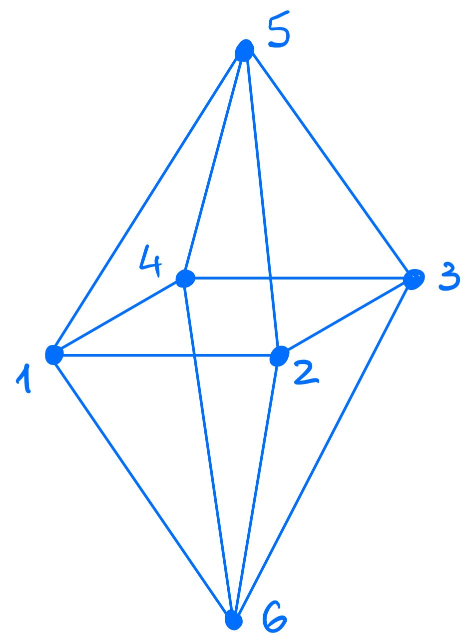
- The elements of K are called *faces*. For $n \geq 0$, a face $\{v_{i_0}, \dots, v_{i_n}\}$ of cardinality $n+1$ is an *n-face*. The set of 0-faces is in bijective correspondence with V and its elements are called *vertices*. The 1-faces are called *edges*.

- The collection of all n-faces of K for $0 \leq n \leq d$ is an abstract simplicial complex for every $d \geq 0$, called the *d-skeleton* of K. The 1-skeleton of K is an undirected graph.

- Every abstract simplicial complex is determined by its maximal faces, i.e., those not contained in any larger face.

- If the index set I of V is equipped with a total order, then K is called <u>ordered</u>. If K is ordered, then we denote its faces as $(i_0 \dots i_n)$ with $i_0 < \dots < i_n$ instead of $\{v_{i_0}, \dots, v_{i_n}\}$.

<u>Example 1:</u> $V = \{1, 2, 3, 4, 5, 6\}$



K:

<u>0-faces</u>	<u>1-faces</u>		<u>2-faces</u>	
(1)	(12)	(26)	(125)	(235)
(2)	(14)	(34)	(126)	(236)
(3)	(15)	(35)	(145)	(345)
(4)	(16)	(36)	(146)	(346)
(5)	(23)	(45)		
(6)	(25)	(46)		

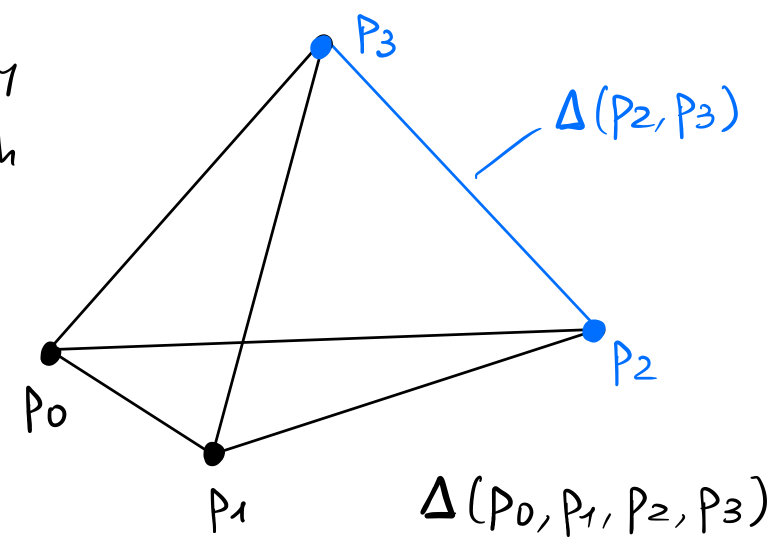
$\underbrace{\hspace{6em}}_{\text{maximal faces}}$

Geometric simplicial complexes

An *n-simplex* of $\mathbb{R}^N$ is the convex hull of $n+1$ affinely independent points $p_0, \dots, p_n$ in $\mathbb{R}^N$.

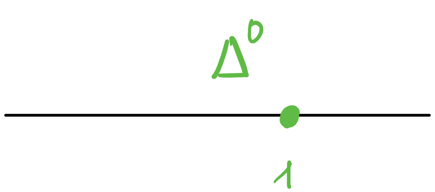
$$\Delta(p_0, \dots, p_n) = \{ x_0 p_0 + \dots + x_n p_n \in \mathbb{R}^N \mid x_0 + \dots + x_n = 1,\ x_i \geq 0\ \forall i \}.$$

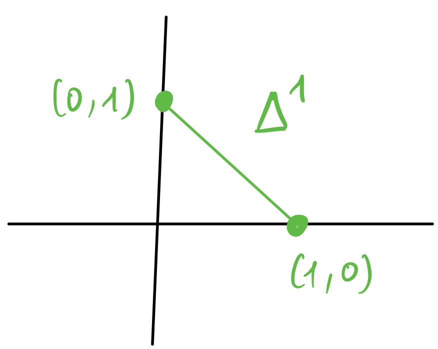
Every subset $\{p_{i_0}, \dots, p_{i_k}\} \subseteq \{p_0, \dots, p_n\}$ spans a k-simplex $\Delta(p_{i_0}, \dots, p_{i_k})$, which is called a *k-face* of $\Delta(p_0, \dots, p_n)$.

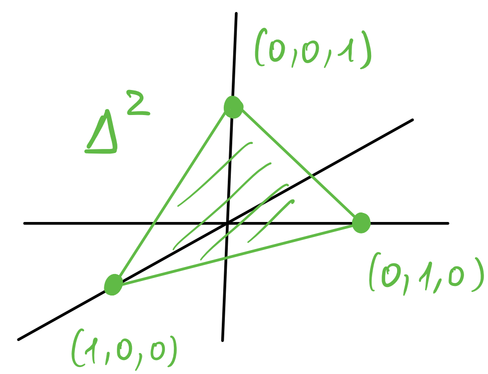


The *standard n-simplex* Δ^n is the convex hull of the coordinate unit points in $\mathbb{R}^{n+1}$:

$$\Delta^n = \Delta(e_1, \dots, e_{n+1}), \quad e_i = (0, \dots, \overset{i}{1}, \dots, 0).$$



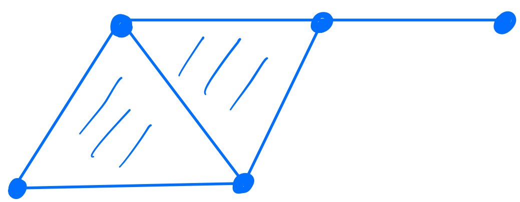


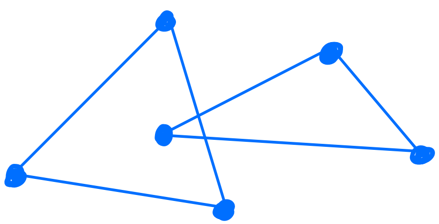


A <u>geometric simplicial complex</u> is a set X of simplices $\sigma \subset \mathbb{R}^N$ for some N such that

a) every face of X is in X;

b) any two simplices in X are either disjoint or intersect in one common face.


YES


NO

The <u>dimension</u> of a finite geometrical simplicial complex is the maximum of the dimensions of its simplices.

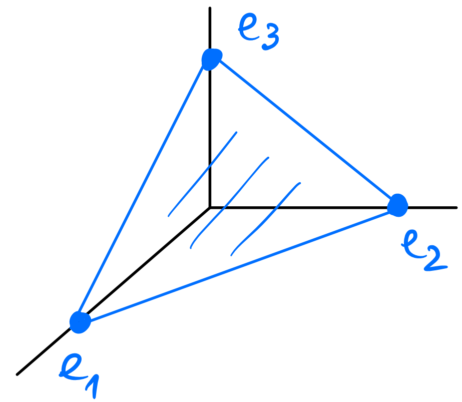
Every geometric simplicial complex X has an <u>underlying</u> topological space, namely $|X| = \bigcup_{\sigma \in X} \sigma$ endowed with the Euclidean topology.

Then $|X|$ is called a <u>polyhedron</u>.

Let K be a finite ordered abstract simplicial complex with vertex set $V=\{v_1,\dots,v_n\}$. The geometric realization of K is the geometric simplicial complex X_K with a k-face $\Delta(e_{i_0},\dots,e_{i_k})$ in $\mathbb{R}^n$ for each k-face $(i_0\dots i_k)$ of K.

Example 2: $V=\{1,2,3\}$

K: (1) (12) (123)
(2) (13)
(3) (23)



$$|X_K| = \Delta(e_1,e_2,e_3) = \Delta^2 \subset \mathbb{R}^3$$

We denote $|K| = |X_K|$, and call K a triangulation of a topological space Y if $|K| \cong Y$.

If K is the abstract simplicial complex from Example 1, then $|K|$ is homeomorphic to the 2-sphere S^2. Hence K is a triangulation of S^2.

Conversely, every geometric simplicial complex X determines an abstract simplicial complex K_X whose set of vertices is the set of 0-faces of X and whose elements are $\{v_{i_0}, \dots, v_{i_n}\}$ for each n-simplex $\Delta(v_{i_0}, \dots, v_{i_n})$ of X. If the 0-faces of X are ordered, then K_X is ordered.

There are face-preserving bijective correspondences

$$X_{K_X} \cong X \qquad \text{and} \qquad K_{X_K} \cong K$$

which induce homeomorphisms $|X_{K_X}| \cong |X|$ and $|K_{X_K}| \cong |K|$.

They can be polyhedra
in different ambient spaces.
For example, if X is an octahedron
in $\mathbb{R}^3$ then $|X_{K_X}| \subset \mathbb{R}^6$.

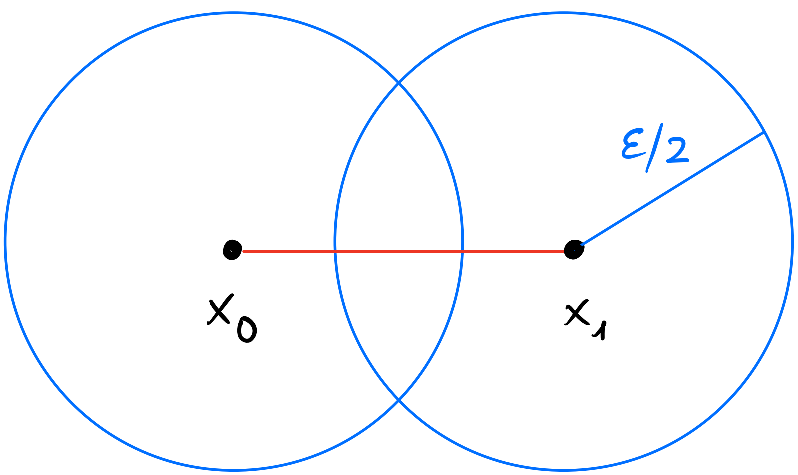
Čech complexes

A <u>point cloud</u> is a finite set of points $X = \{x_i\}_{i \in I}$ in $\mathbb{R}^N$ with $N \geq 2$.

To every point cloud X and every $\varepsilon \geq 0$, one associates an abstract simplicial complex $C_\varepsilon(X)$, called <u>Čech complex</u>, with vertex set X and a k-face for each collection $x_{i_0}, \dots, x_{i_k}$ such that

$$\overline{B}_{\varepsilon/2}(x_{i_0}) \cap \dots \cap \overline{B}_{\varepsilon/2}(x_{i_k}) \neq \emptyset$$

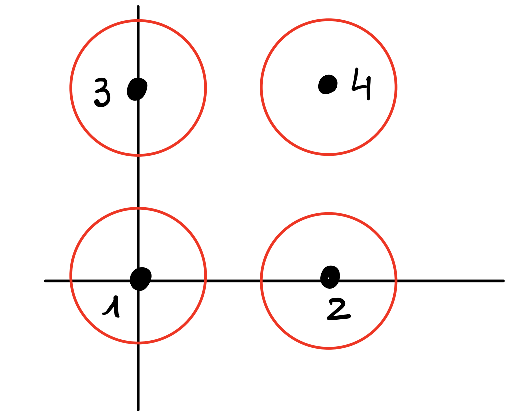
where $\overline{B}_r(p)$ denotes a closed ball of radius r centered at p in $\mathbb{R}^N$.



There is an edge $\{x_0, x_1\}$ in $C_\varepsilon(X)$ if and only if $\|x_0 - x_1\| \leq \varepsilon$

Example: $X = \{(0,0), (1,0), (0,1), (1,1)\}$ in $\mathbb{R}^2$.

$x_1 \quad x_2 \quad x_3 \quad x_4$

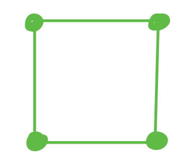


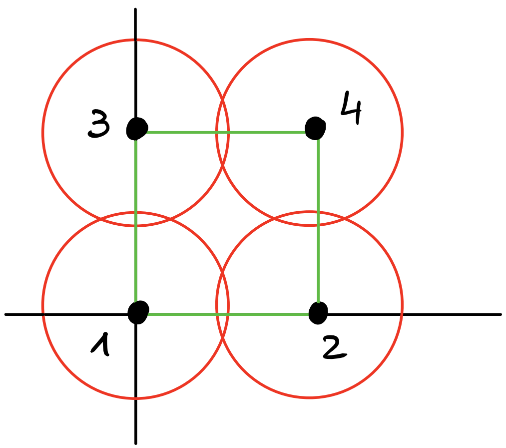
① For $0 \le \varepsilon < 1$, $C_\varepsilon(X)$: $(1)\ (2)\ (3)\ (4)$

maximal faces

② For $1 \le \varepsilon < \sqrt{2}$, $C_\varepsilon(X)$: $(1\,2)\ \ (2\,4)$

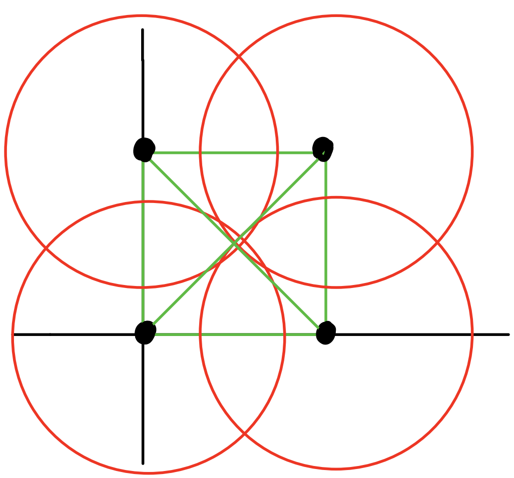
$(1\,3)\ \ (3\,4)$





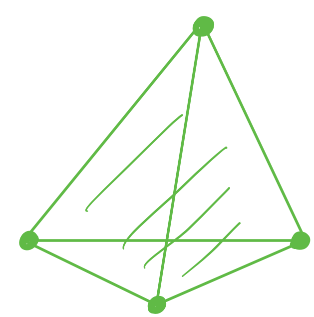
maximal face

③ For $\varepsilon \geq \sqrt{2}$, $C_\varepsilon(X)$: $(1\,2\,3\,4)$



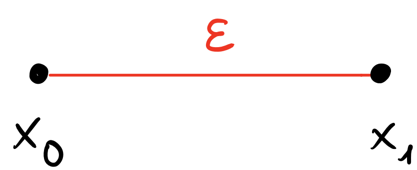
$$|C_\varepsilon(X)| \cong \Delta^3$$

if $\varepsilon \geq \sqrt{2}$



The Vietoris–Rips complex $R_\varepsilon(X)$ is the abstract simplicial complex with vertex set X and a k-face for each collection $x_{i_0}, \dots, x_{i_k}$ such that $\|x_{i_r} - x_{i_s}\| \le \varepsilon \;\; \forall\, r, s \in \{0, \dots, k\}$.

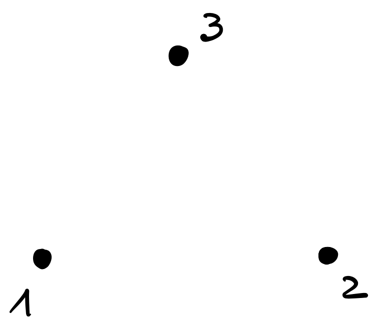
In other words, $R_\varepsilon(X)$ has a k-face $\{x_{i_0}, \dots, x_{i_k}\}$ if and only if $\operatorname{diam}\{x_{i_0}, \dots, x_{i_k}\} \le \varepsilon$, where $\operatorname{diam}(A) = \sup\{\|a-b\| : a, b \in A\}$.



There is an edge $\{x_0, x_1\}$ in $R_\varepsilon(X)$ if and only if $\|x_0 - x_1\| \le \varepsilon$

- Hence, the 1-skeleta of $C_\varepsilon(X)$ and $R_\varepsilon(X)$ coincide.
- There are inclusions $C_\varepsilon(X) \subseteq R_\varepsilon(X) \subseteq C_{\varepsilon\sqrt{2}}(X)$ for all X and $\varepsilon \ge 0$.
- $R_\varepsilon(X)$ is a flag complex, that is, every collection of pairwise adjacent vertices spans a face. A flag complex is determined by its 1-skeleton.
- $R_\varepsilon(X)$ is determined by the distance matrix $(\|x_i - x_j\|)$ for $x_i, x_j \in X$.

Example:



$X = \{x_1, x_2, x_3\}$ in $\mathbb{R}^2$

$\|x_i - x_j\| = 1$ if $i \neq j$

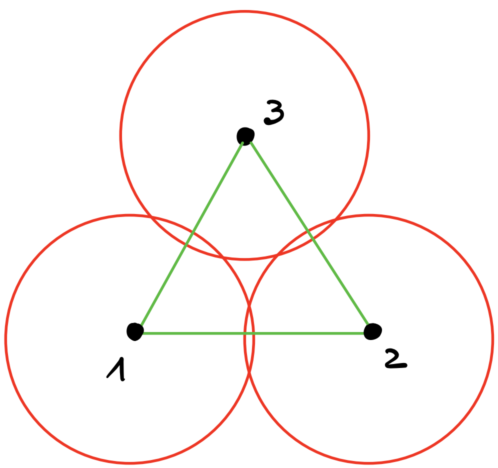
Distance matrix:

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

① For $0 \leq \varepsilon < 1$, $R_\varepsilon(X)$: $(1)\,(2)\,(3)$

② For $\varepsilon \geq 1$, $R_\varepsilon(X)$: $(1\,2\,3)$

However, for $1 \leq \varepsilon < \frac{2\sqrt{3}}{3}$, $C_\varepsilon(X)$: $(12)(13)(23)$.



Hence $C_\varepsilon(X) \neq R_\varepsilon(X)$ if $1 \leq \varepsilon < \frac{2\sqrt{3}}{3}$.

- If $0 \leq \varepsilon < \min\{\|x_i - x_j\| : x_i \neq x_j\}$, then $C_\varepsilon(X) = R_\varepsilon(X) = X$.
- If $\varepsilon \geq \mathrm{diam}(X)$, then $|R_\varepsilon(X)| \cong \Delta^{n-1}$ where $n = \mathrm{card}(X)$. If $\varepsilon \geq \sqrt{2}\,\mathrm{diam}(X)$, then $|C_\varepsilon(X)| \cong \Delta^{n-1}$ as well.

Both $\{C_\varepsilon(X)\}_{\varepsilon \geq 0}$ and $\{R_\varepsilon(X)\}_{\varepsilon \geq 0}$ are called <u>filtered simplicial complexes</u>, and ε is the <u>filtering parameter</u>. We can let ε range over $\mathbb{R}$ by defining $C_\varepsilon(X) = R_\varepsilon(X) = \emptyset$ if $\varepsilon < 0$.

Weighted graphs

In this course, a graph is a 1-dimensional abstract simplicial complex.

A graph is called a clique if is fully connected, i.e., every pair of vertices is joined by an edge.

A weighted graph is a graph with a real number $\omega \geq 0$ attached to each of its edges.

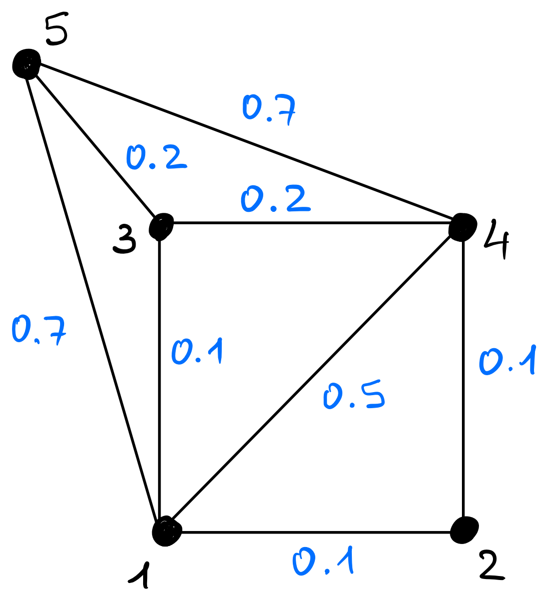
A finite weighted graph G with set of vertices $V = \{v_i\}_{i \in I}$ yields a Vietoris-Rips complex $R_\varepsilon(G)$ for $\varepsilon \geq 0$ with set of vertices V and a k-face $\{x_{i_0}, \dots, x_{i_k}\}$ for $k \geq 1$ if and only if $x_{i_0}, \dots, x_{i_k}$ form a clique and every edge $\{x_{i_r}, x_{i_s}\}$ between them has weight $\omega(x_{i_r}, x_{i_s}) \leq \varepsilon$.

- Every point cloud X gives rise to a fully connected graph with set of vertices X and edge weights $\omega(x_i, x_j) = \|x_i - x_j\|$ for $i \neq j$.

- Weighted graphs arise from complex systems, and from activation correlations in neural networks, etc.

Example:

G



$R_\varepsilon(G)$:

$\varepsilon < 0.1$	(1) (2) (3) (4) (5)
$0.1 \leq \varepsilon < 0.2$	(12) (13) (24)
$0.2 \leq \varepsilon < 0.5$	(12) (13) (24) (34) (35)
$0.5 \leq \varepsilon < 0.7$	(124) (134) (35)
$\varepsilon \geq 0.7$	(124) (1345)

In general, if $\varepsilon \geq w(e)$ for every edge $e \in G$, then $R_\varepsilon(G)$ is equal to the <u>clique complex</u> of G, whose faces are the maximal cliques in G, i.e., those not contained in any larger clique. In the above example, the maximal cliques are (124) and (1345).

Eduard Čech (1893 – 1960)

Leopold Vietoris (1891 – 2002)

Eliyahu Rips (1948 –)

The Čech complex $C_\varepsilon(X)$ is the <u>nerve</u> of the covering $\{\overline{B}_{\varepsilon/2}(x_i)\}_{i\in I}$ of X.

<u>Nerve Theorem:</u>

For every $\varepsilon \geq 0$ and every $X = \{x_i\}_{i\in I}$,

$$|C_\varepsilon(X)| \simeq \bigcup_{i\in I} \overline{B}_{\varepsilon/2}(x_i).$$

homotopy equivalence