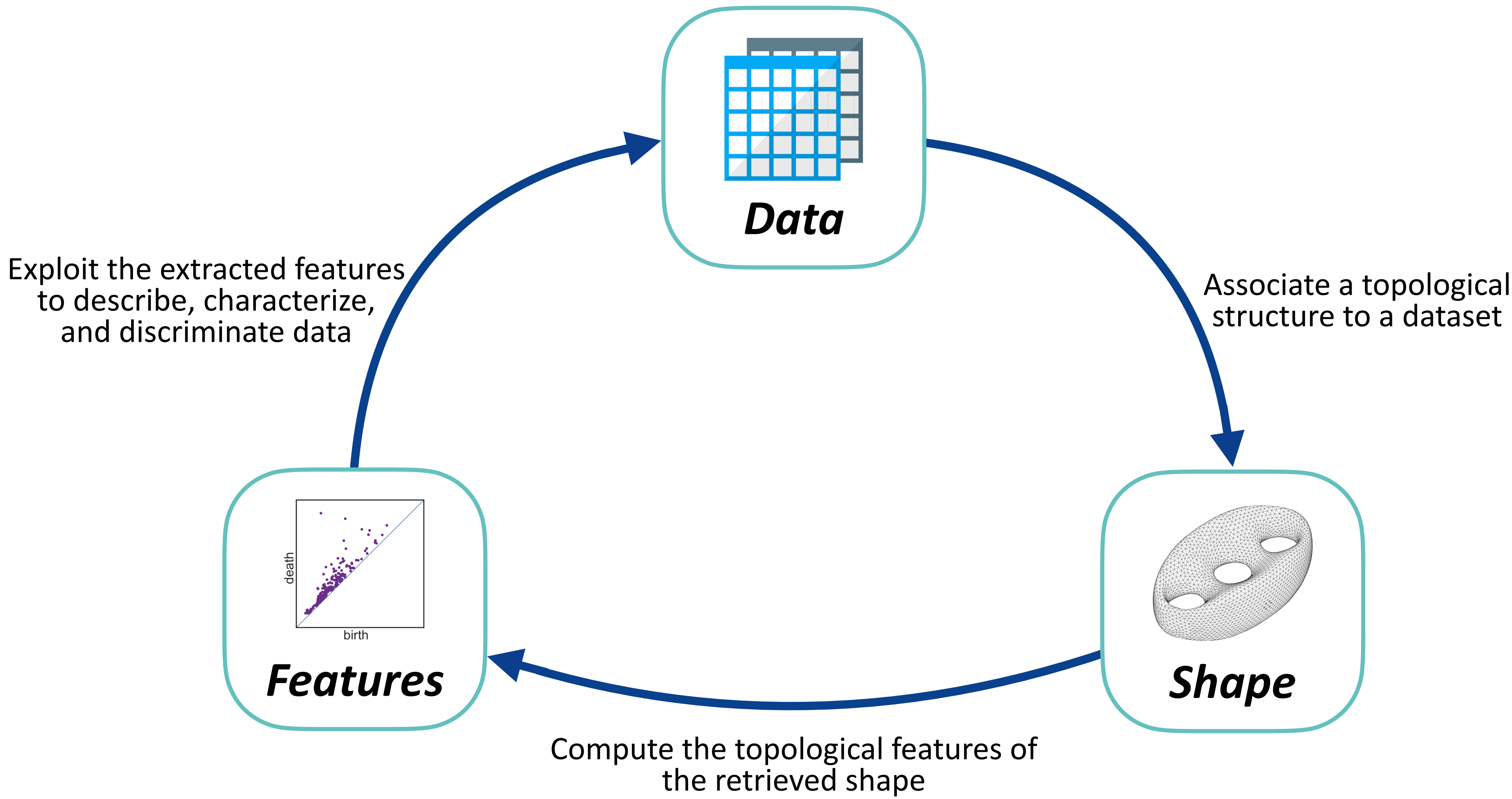
Topological Data Analysis

Complexes & Data

Ulderico Fugacci

CNR - IMATI

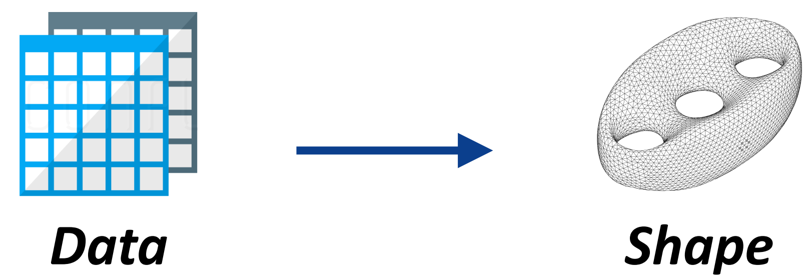
Topological Data Analysis



Complexes & Data

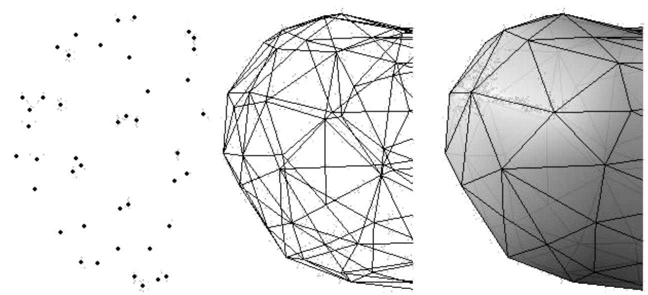
Goal:

We want to associate a topological structure to a given dataset



Due to the nature of data and to our computational ambitions, datasets will be represented by ***"discrete"*** structures

Among various possibilities, ***simplicial complexes*** represent the most suitable choice



In fact, simplicial complexes are able to deal with data:

- of ***large size*** (e.g. consisting of a huge number of samples)
- of ***high dimension*** (e.g. involving a large number of variables or parameters)
- ***unorganized*** (e.g. not arranged in a regular grid)

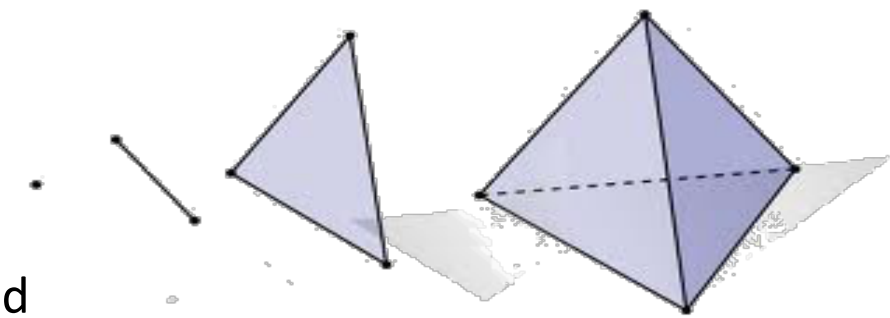
Complexes & Data

- *Simplicial Complexes (and other discrete representations)*
- *From Data to Complexes*

Complexes & Data

- ***Simplicial Complexes** (and other discrete representations)*
- *From Data to Complexes*

Simplicial Complexes



Definitions:

A set $V := \{ v_0, v_1, ..., v_k \}$ of points in $\mathbb{R}^n$ is called ***geometrically independent*** if vectors $v_1 - v_0, ..., v_k - v_0$ are ***linearly independent*** over $\mathbb{R}$

E.g. two distinct points, three non-collinear points, four non-coplanar points

The ***k-simplex*** $\sigma = v_0 v_1 \dots v_k$ spanned by a geometrically independent set $V = \{v_0, v_1, ..., v_k\}$ of in $\mathbb{R}^n$ is the ***convex hull*** of V, i.e. the set of all points $x \in \mathbb{R}^n$ such that

$$x = \sum_{i=0}^{k} t_i v_i \quad \text{where} \quad \sum_{i=0}^{k} t_i = 1 \quad \text{and } t_i \geq 0 \text{ for all } i$$

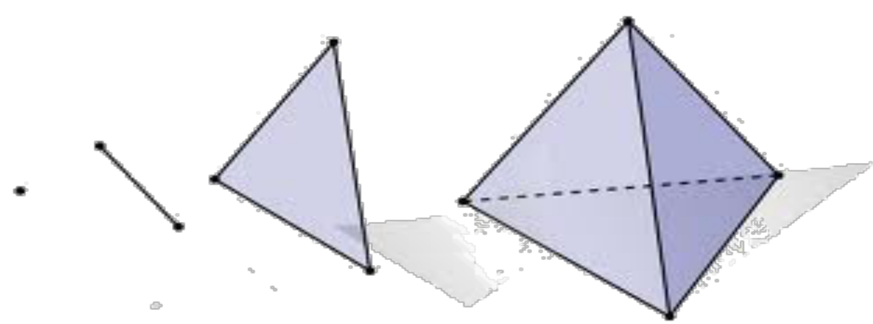
The numbers t_i are uniquely determined by x and are called ***barycentric coordinates*** of x

E.g. a 0-simplex is a vertex, a 1-simplex is an edge, a 2-simplex is a triangle, a 3-simplex is a tetrahedron

Simplicial Complexes

Definitions:

- The points $v_0, v_1, \ldots, v_k$ spanning a k-simplex $\boldsymbol{\sigma}$ are called the ***vertices*** of $\boldsymbol{\sigma}$
- k is called the ***dimension*** of $\boldsymbol{\sigma}$ and denoted as $\dim(\boldsymbol{\sigma})$
- Any simplex $\boldsymbol{\tau}$ spanned by a non-empty subset of V is called a ***face*** of $\boldsymbol{\sigma}$
- Conversely, $\boldsymbol{\sigma}$ is called a ***coface*** of $\boldsymbol{\tau}$

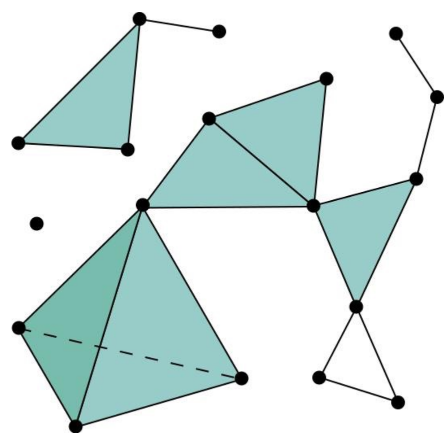


Simplicial Complexes

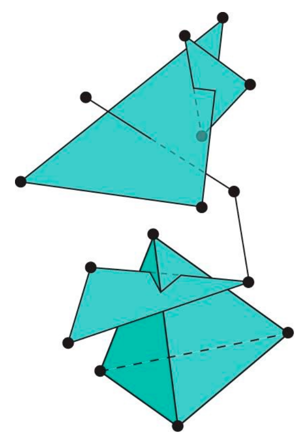
Definition:

A *(geometric)* ***simplicial complex*** K in $\mathbb{R}^n$ is a collection of simplices in $\mathbb{R}^n$ such that

- *Every face of a simplex of K is in K*
- *The non-empty intersection of any two simplices of K is a face of each of them*



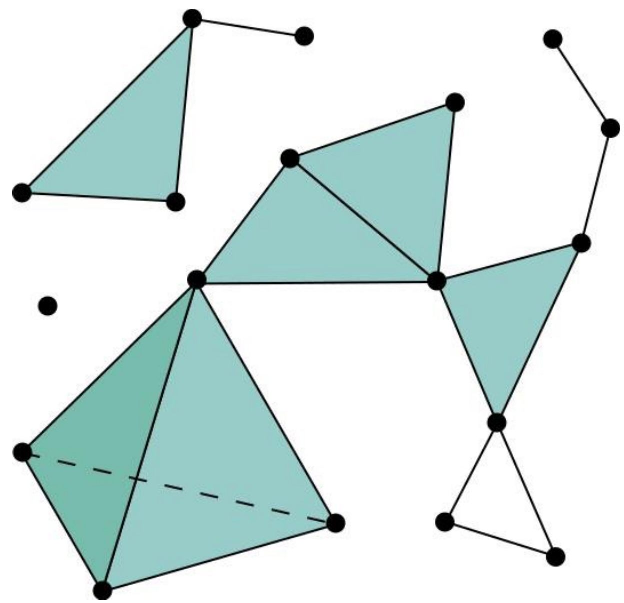
simplicial complex



non-simplicial complex

Simplicial Complexes

Definitions:



Given a (geometric) simplicial complex K in $\mathbb{R}^n$,

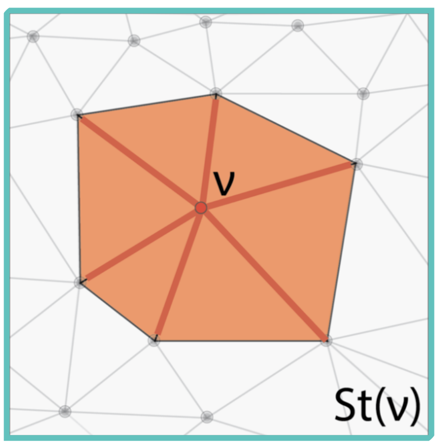
- The ***dimension*** of a simplicial complex K in $\mathbb{R}^n$, denoted as dim(K), is the supremum of the dimensions of the simplices of K
- A simplex σ of K such that dim(σ) = dim(K) is called ***maximal***
- A simplex σ of K which is not a proper face of any simplex of K is called ***top***
- A subcollection of K that is itself a simplicial complex is called a ***subcomplex*** of K

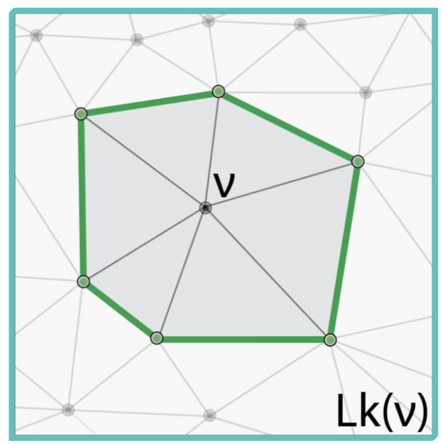
Simplicial Complexes

Definitions:

Given a simplex $\boldsymbol{\sigma}$ of a (geometric) simplicial complex K in $\mathbb{R}^n$,

- The ***star*** of $\boldsymbol{\sigma}$ is the set ***St(σ)*** of the cofaces of $\boldsymbol{\sigma}$
- The ***link*** of $\boldsymbol{\sigma}$ is the set ***Lk(σ)*** of the faces of the simplices in St($\boldsymbol{\sigma}$) such that do not intersect $\boldsymbol{\sigma}$



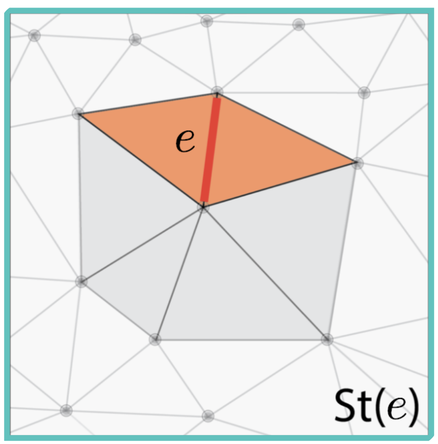


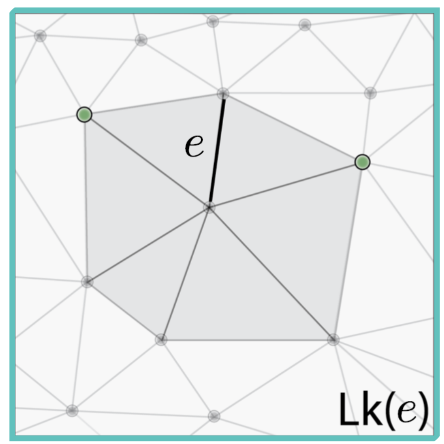
Simplicial Complexes

Definitions:

Given a simplex $\boldsymbol{\sigma}$ of a (geometric) simplicial complex K in $\mathbb{R}^n$,

- The ***star*** of $\boldsymbol{\sigma}$ is the set *St(σ)* of the cofaces of $\boldsymbol{\sigma}$
- The ***link*** of $\boldsymbol{\sigma}$ is the set *Lk(σ)* of the faces of the simplices in St($\boldsymbol{\sigma}$) such that do not intersect $\boldsymbol{\sigma}$





Simplicial Complexes

Given a (geometric) simplicial complex K in $\mathbb{R}^n$,
its ***polytope*** $|K|$ is the subset of $\mathbb{R}^n$ defined as the union of the simplices of K

The polytope $|K|$ can be endowed with ***two possible topologies*** T_1 and T_2:

- ***T_1***: A subset F of $|K|$ is a closed set of $(|K|, T_1)$ if and only if $F \cap \boldsymbol{\sigma}$ is a closed set of $(\boldsymbol{\sigma}, T_{\sigma})$ for each $\boldsymbol{\sigma}$ in K where T_{σ} is the subspace topology induced on $\boldsymbol{\sigma}$ by $\mathbb{E}^n$
- ***T_2***: The subspace topology induced on $|K|$ by $\mathbb{E}^n$

In general, the two topologies T_1, T_2 are ***different***, but

Proposition: *If K is a **finite** simplicial complex, $T_1 = T_2$*

From now on, if not differently specified, we consider only *finite* simplicial complexes

Simplicial Complexes

Proposition:

Given a simplicial complex K and a topological space (X, T), a function f from (|K|, T_1) to (X, T) is ***continuous*** *if and only if $f|_{\sigma}$ is continuous for each $\boldsymbol{\sigma} \in K$*

Definition:

Given two simplicial complexes K and K',

- A function f: K $\longrightarrow$ K' is called a ***simplicial map*** if for every simplex $\boldsymbol{\sigma} = v_0 v_1 \dots v_k$ in K, $f(\boldsymbol{\sigma}) = f(v_0)f(v_1)\dots f(v_k)$ is a simplex in K'
- The restriction f_V of f to the set of vertices V of K is called the ***vertex map*** of f

Simplicial Complexes

Definition:

An ***abstract simplicial complex*** K on a set V is a collection of finite non-empty subsets of V, called ***simplices***, such that if $\boldsymbol{\sigma} \in$ K and $\boldsymbol{\tau} \subseteq \boldsymbol{\sigma}$, then $\boldsymbol{\tau} \in$ K

Analogously to the case of a geometric simplicial complex,

- The elements of V are called ***vertices*** of K
- The ***dimension*** of a simplex $\boldsymbol{\sigma}$ is one less than the number of its elements
- The supremum of the dimensions of the simplices in K is called ***dimension*** of K
- Each non-empty subset $\boldsymbol{\tau}$ of a simplex $\boldsymbol{\sigma} \in$ K is called a ***face*** of $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}$ is called a ***coface*** of $\boldsymbol{\tau}$

The notions of geometric simplicial complex and abstract simplicial complex are equivalent. *More properly, it is always possible,*

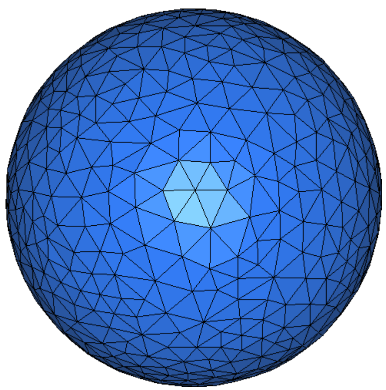
- *Given an abstract simplicial complex, to endow it with a* ***geometric realization***
- *Given a geometric simplicial complex, to* ***forget its geometry*** *thus obtaining an abstract simplicial complex*

Simplicial Complexes

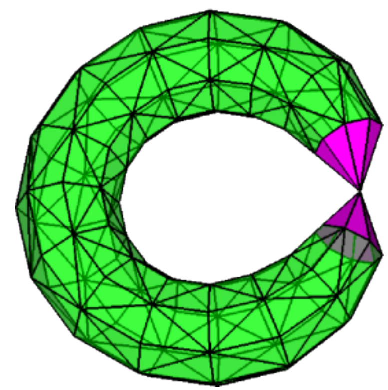
Definition: A simplicial complex K is called

- ***n-manifold [with boundary]*** if its polytope |K| is a (topological) n-manifold [with boundary]
- ***Combinatorial n-manifold [with boundary]*** if, for every vertex v, the link Lk(v) is homeomorphic to the (n − 1)-sphere S^{n-1} [or to the (n − 1)-disk $D^{n-1} := \{x \in \mathbb{R}^{n-1} : |x| \leq 1\}$]

combinatorial manifold



non-combinatorial manifold



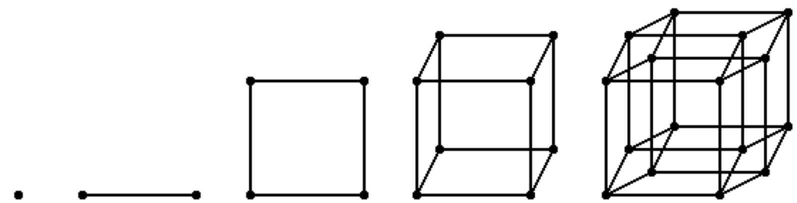
Proposition:

If K is a combinatorial n-manifold [with boundary], then K is a n-manifold [with boundary]

The converse is: ***True for n ≤ 3*** ***Open for n = 4*** ***False for n > 4***

Regular Grids

Hyper-Cube:

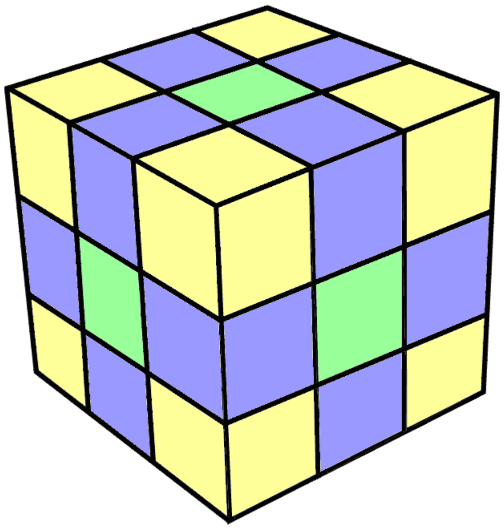


A ***k*-hyper-cube *η*** is the ***Cartesian product of k closed intervals*** of equal length

Regular Grids:

A ***regular grid H*** is a (finite) collection of hyper-cubes such that:

- *Each face of a hyper-cube of H is in H*
- *Each non-empty intersection of two hyper-cubes in H is a face of both*
- *The domain of H is a hyper-cube*

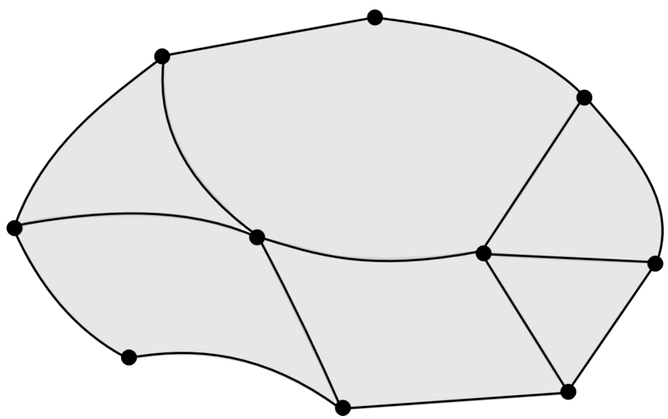


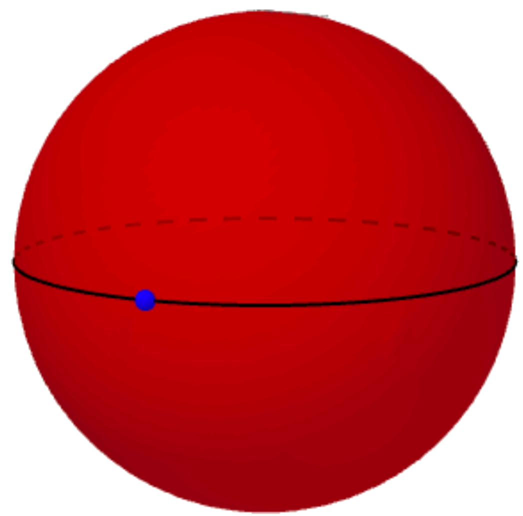
Cell Complexes

Intuitively:

Similarly to simplicial complexes and regular grids,

A ***cell complex*** Γ is a collection of cells ***"suitably glued together"***





Where a ***k-cell*** is a topological space homeomorphic to the ***k-dimensional open disk $i(D^k)$***

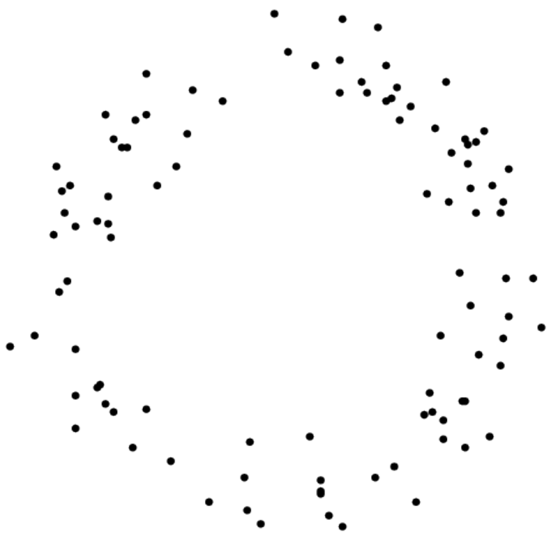
Complexes & Data

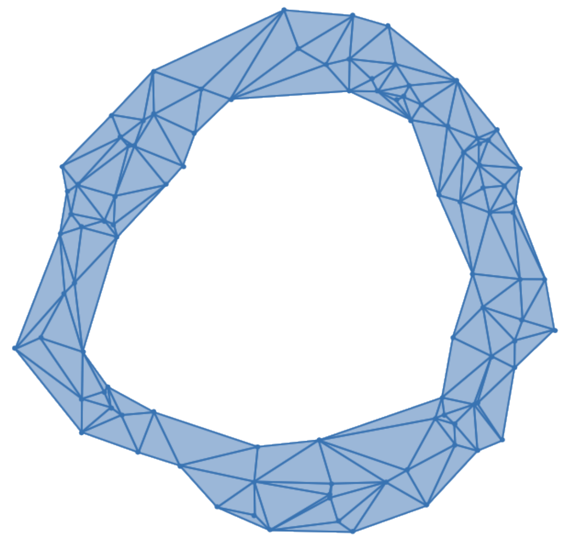
- *Simplicial Complexes (and other discrete representations)*
- ***From Data to Complexes***

From Data to Complexes

Let us consider a dataset represented by a ***finite point cloud V in*** $\mathbb{R}^n$

*Studying the shape of V just by considering the space consisting of its **points does not provide any relevant topological information***





*The **"real" shape** of the dataset can be captured by properly constructing a **complex connecting together close points through simplices***

From Data to Complexes

Standard Constructions:

A number of possible choices have been introduced in the literature:

- ***Delaunay triangulations***
  - ***Voronoi*** *diagrams*
- ***Čech complexes***
- ***Vietoris-Rips complexes***
- ***Alpha-shapes***
- ***Witness complexes***

Most of the above constructions are based on the notion of ***Nerve complex***

From Data to Complexes

A First Classification: Given a finite point cloud V in $\mathbb{R}^n$,

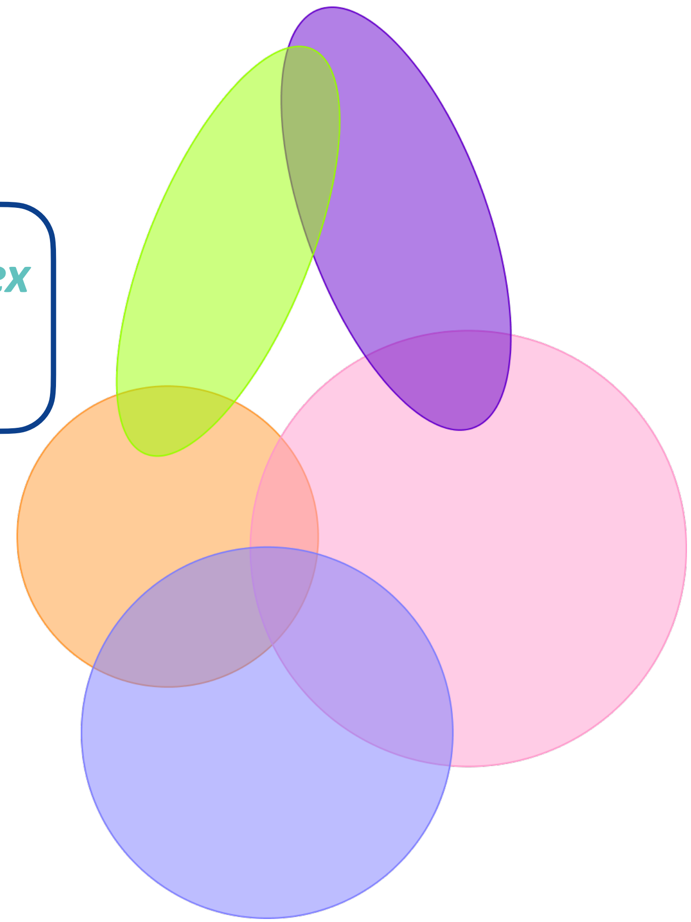
	Output Complex	Dimension	Dependence on a Parameter
Delaunay triangulation	***Geometric***	***n***	✗
Čech complex	***Abstract***	***Arbitrary (up to |V| - 1)***	✓
Vietoris-Rips complex	***Abstract***	***Arbitrary (up to |V| - 1)***	✓
Alpha-shapes	***Geometric***	***n***	✓
Witness complexes	***Abstract***	***Arbitrary (up to |V| - 1)***	✓

Nerve Complexes

Definition:

Given a finite collection S of sets in $\mathbb{R}^n$,

The ***nerve Nrv(S)*** of S is the ***abstract simplicial complex*** generated by the ***non-empty common intersections***



Formally,

$$Nrv(S) := \{\sigma \subseteq S \mid \bigcap_{s \in \sigma} s \neq \emptyset\}$$

Nerve Complexes

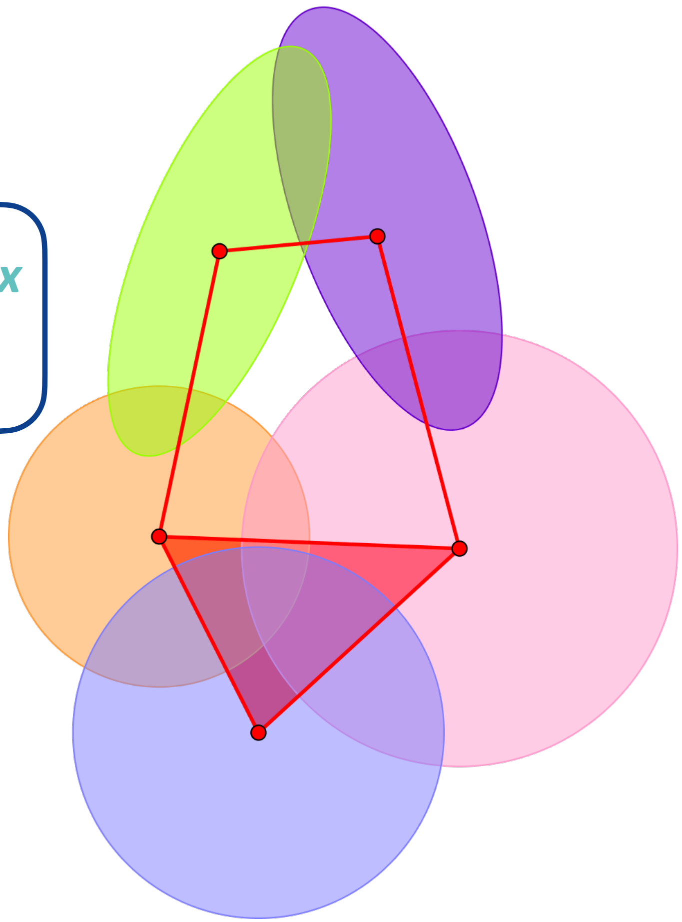
Definition:

Given a finite collection S of sets in $\mathbb{R}^n$,

The ***nerve Nrv(S)*** of S is the ***abstract simplicial complex*** generated by the ***non-empty common intersections***

Formally,

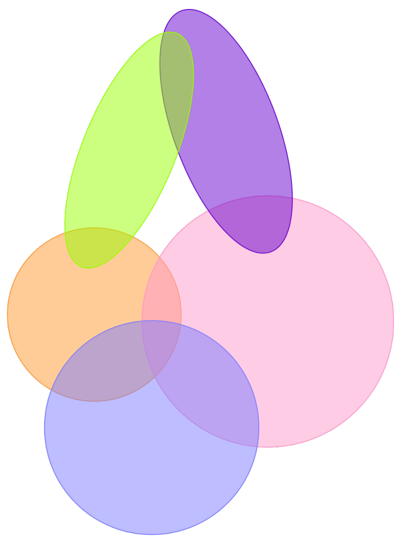
$$Nrv(S) := \{\sigma \subseteq S \mid \bigcap_{s \in \sigma} s \neq \emptyset\}$$



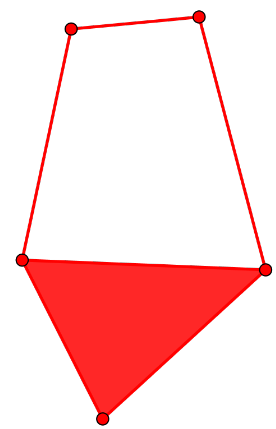
Nerve Complexes

Nerve Theorem:

If S is a finite collection of ***convex*** *sets in* $\mathbb{R}^n$*, then the* ***nerve of S*** *and the* ***union of the sets in S*** *are* ***homotopy equivalent*** *(and so they have the same homology)*



$\simeq$



Nerve Complexes

Nerve Theorem can be ***generalized*** by replacing the ***convexity*** of sets in S with the request that all non-empty common intersections are ***contractible***

(i.e. that can be continuously shrunk to a point)

Original Nerve Theorem:

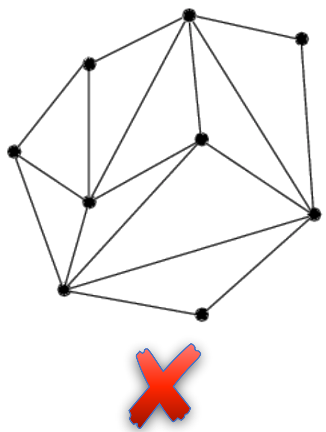
*If S is an open cover of a (para)**compact** space X such that*
*every non-empty intersection of finitely many sets in S is **contractible**,*
*then **X** is **homotopy equivalent** to the nerve **Nrv(S)***

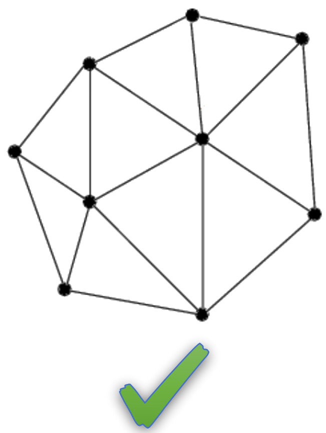
Delaunay Triangulations

Given a finite point cloud V in $\mathbb{R}^n$,

The ***Delaunay triangulation*** of V is a classic notion in Computational Geometry:

- *Producing a "nice" **triangulation** of V*
  - *free of long and skinny triangles*
- *Named after Boris Delaunay for his work on this topic from 1934*
- *Originally defined for sets of points in $\mathbb{R}^2$ but generalizable to arbitrary dimensions*





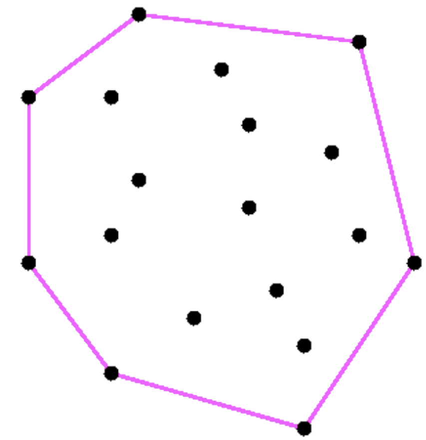
Images from [De Floriani 2003]

Delaunay Triangulations

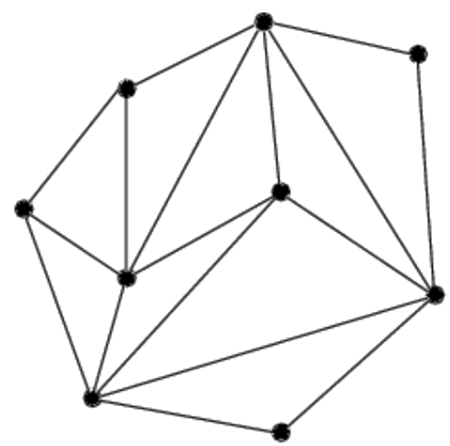
Definitions:

Given a finite point cloud V in $\mathbb{R}^2$,

- The ***convex hull*** of V is the ***smallest convex*** subset ***CH(V)*** of $\mathbb{R}^2$ containing all the points of V



- A ***triangulation*** of V is A ***2-dimensional simplicial complex K*** such that:
  - The domain of K is CH(V)
  - The 0-simplices of K are the points in V



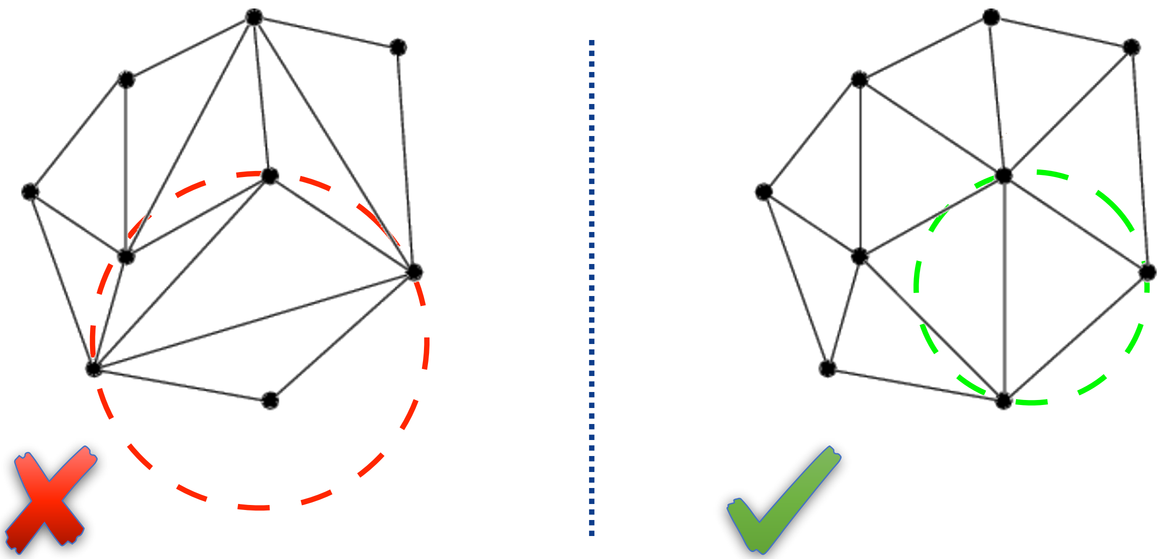
Images from [De Floriani 2003]

Delaunay Triangulations

Definition:

A ***Delaunay triangulation*** is a triangulation ***Del(V)*** of V such that:

the ***circumcircle of any triangle*** does ***not contain any point*** of V in its interior



Images from [De Floriani 2003]

Delaunay Triangulations

Definition:

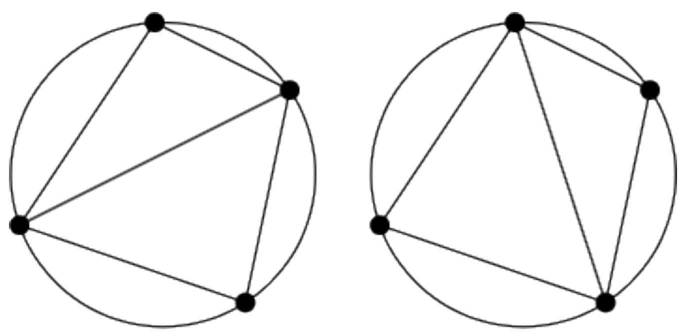
A finite set of points V in $\mathbb{R}^n$ is ***in general position*** if no n + 2 of the points lie on a common (n − 1)-sphere

*E.g. , **for n = 2**,*

V in general position ⇔ *if and only if* ⇔ *No four or more points are co-circular*

Theorem:

*If V is in general position, then Del(V) is **unique***



Images from [De Floriani 2003]

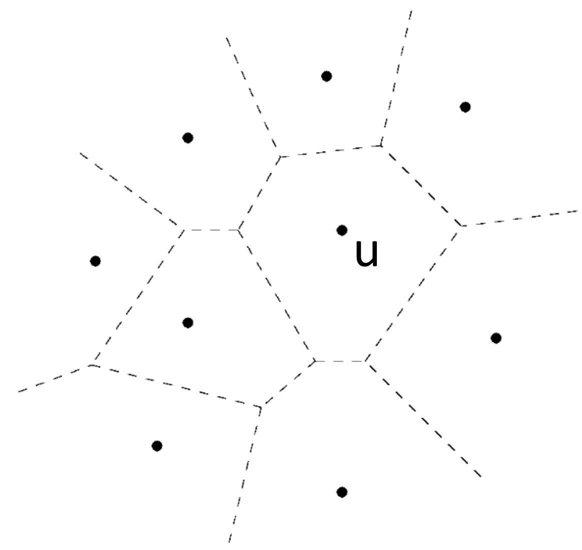
Delaunay Triangulations

Definitions:

The ***Voronoi region*** of u in V is the set of points of $\mathbb{R}^2$ for which u is the closest

$$R_V(u) := \{x \in \mathbb{R}^2 \,|\, \forall v \in V, d(x,u) \leq d(x,v)\}$$

- *Any Voronoi region is a **convex** closed subset of* $\mathbb{R}^2$
- *A Voronoi region is **not necessarily bounded***



The ***Voronoi diagram*** is the collection ***Vor(V)*** of the Voronoi regions of the points of V

Images from [De Floriani 2003]

Delaunay Triangulations

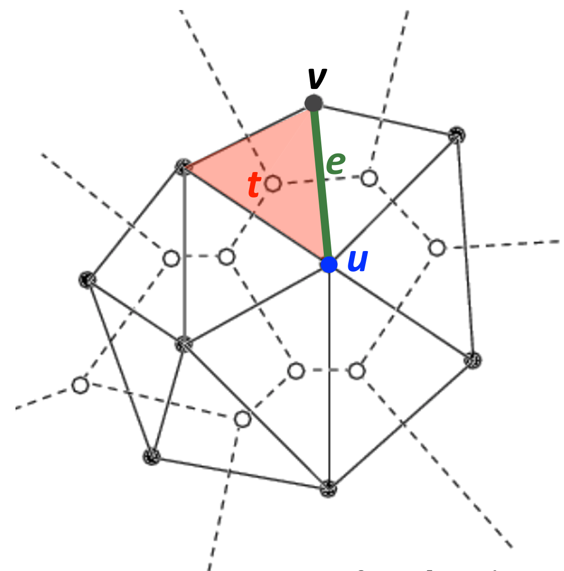
Duality Property:

If V is in general position, then

*the **Delaunay triangulation coincides** with the **nerve of the Voronoi diagram***

$$Del(V) = \{\sigma \subseteq V \mid \bigcap_{u \in \sigma} R_V(u) \neq \emptyset\}$$

- *Each **point u** of V corresponds to a Voronoi region $R_V(u)$*
- *Each **triangle t** of Del(V) correspond to a vertex in Vor(V)*
- *Each **edge e=(u,v)** in Del(V) corresponds to an edge shared by the two Voronoi regions $R_V(u)$ and $R_V(v)$*



Images from [De Floriani 2003]

Delaunay Triangulations

Algorithms:

- ***Two-step algorithms***:
  - *Computation of an arbitrary triangulation K’*
  - *Optimization of K’ to produce a Delaunay triangulation*
- ***Incremental algorithms*** *[Guibas, Stolfi 1983; Watson 1981]:*
  - *Modification of an existing Delaunay triangulation while adding a new vertex at a time*
- ***Divide-and-conquer algorithms*** *[Shamos 1978; Lee, Schacter 1980]:*
  - *Recursive partition of the point set into two halves*
  - *Merging of the computed partial solutions*
- ***Sweep-line algorithms*** *[Fortune 1989]:*
  - *Step-wise construction of a Delaunay triangulation while moving a sweep-line in the plane*

Delaunay Triangulations

Watson's Algorithm:

A Delaunay triangulation is computed by **incrementally adding a single point** to an existing Delaunay triangulation

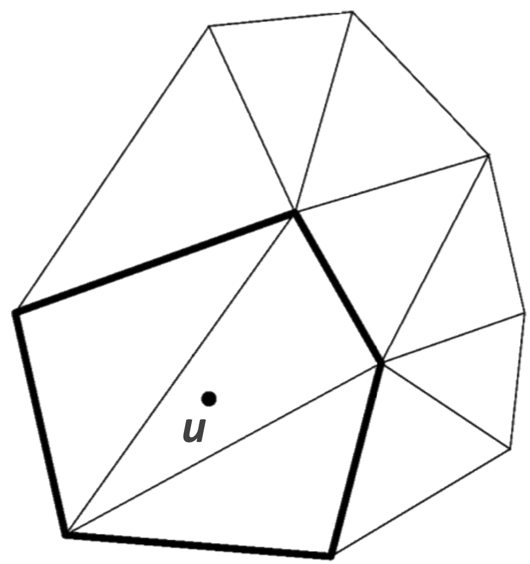
Let V_i be a subset of V and let u be a point in $V \setminus V_i$,

Input:

Del(V_i), a Delaunay triangulation of V_i

Output:

Del(V_{i+1}), a Delaunay triangulation of **$V_{i+1} := V_i \cup \{u\}$**



Images from [De Floriani 2003]

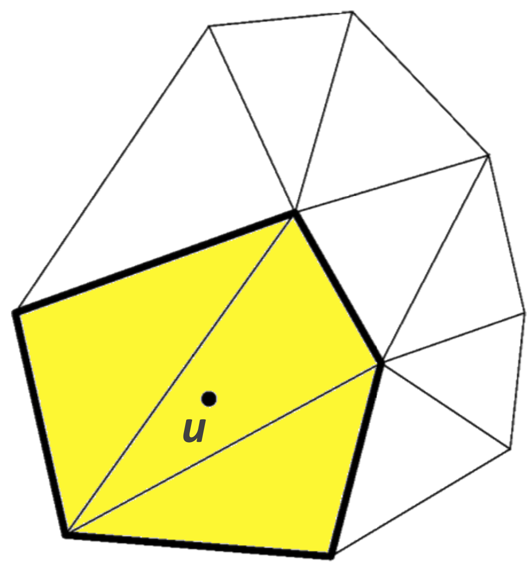
Delaunay Triangulations

Watson's Algorithm:

Given a Delaunay triangulation Del(V_i) of V_i and a point u in $V \setminus V_i$,

- *The **influence region R_u** of a point u is the region in the plane formed by the union of the triangles in Del(V_i) whose circumcircle contains u in its interior*

- *The **influence polygon P_u** of u is the polygon formed by the edges of the triangles of Del(V_i) which bound R_u*

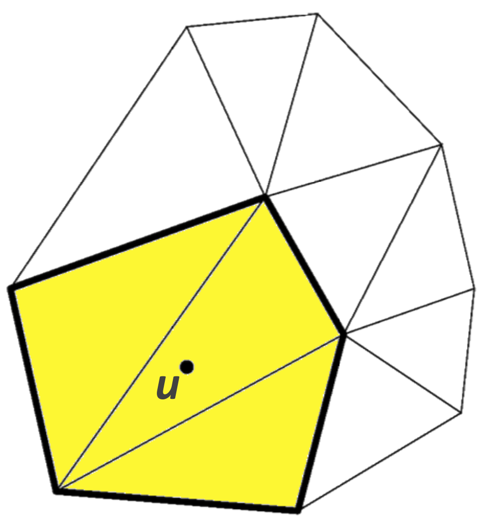


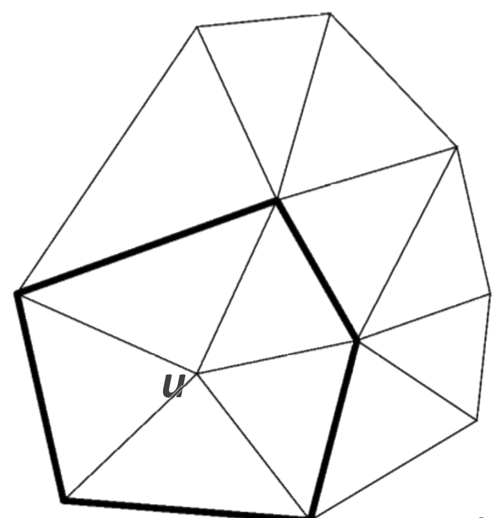
Images from [De Floriani 2003]

Delaunay Triangulations

Watson's Algorithm:

- ***Step 1:***
  Deletion of the triangles of $Del(V_i)$ forming the ***influence region R_u***
- ***Step 2:***
  Re-triangulation of R_u by joining u to the vertices of the influence polygon P_u





Images from [De Floriani 2003]

Delaunay Triangulations

Watson's Algorithm:

Let $N_i = |V_i|$

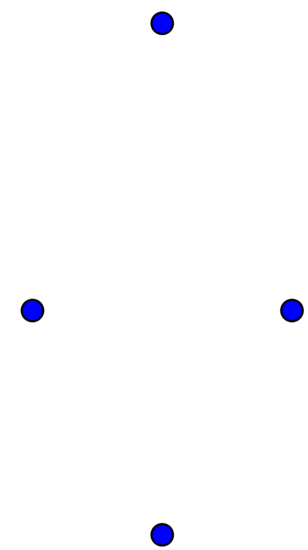
- *Detection of a triangle of $Del(V_i)$ containing the new point u: $O(N_i)$ in the worst case*
- *Detection of the triangles forming the region of influence through a breadth-first search: $O(|R_u|)$*
- *Re-triangulation of P_u is in $O(|P_u|)$*

- ***Inserting a point*** *u in a triangulation with N_i vertices:* ***$O(N_i)$*** *in the worst case*
- ***Inserting all points*** *of V:* ***$O(N^2)$*** *in the worst case, where $N = |V|$*

Čech Complexes

Definition:

Given a finite set of points V in $\mathbb{R}^n$, let us consider:

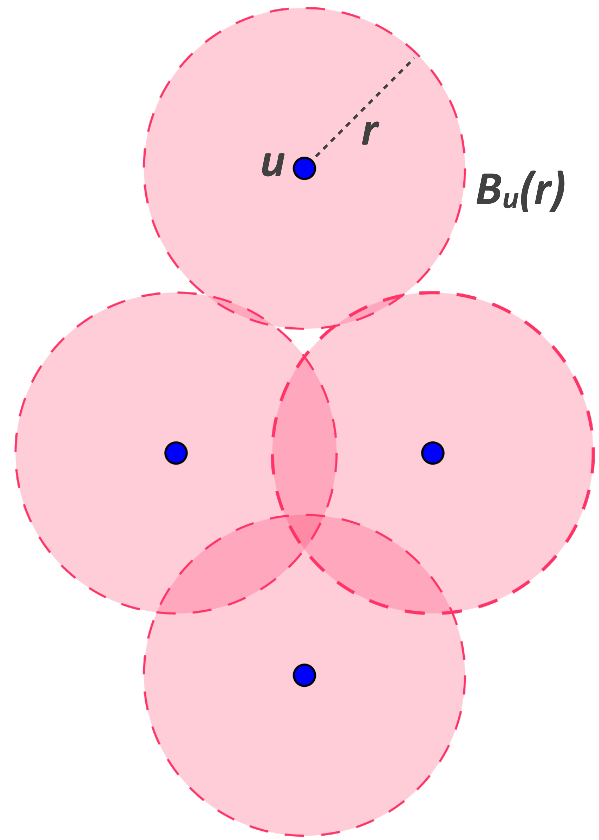


Čech Complexes

Definition:

Given a finite set of points V in $\mathbb{R}^n$, let us consider:

- ***$B_u(r)$**, the **closed ball** with **center u** $\in$ V and **radius r***
- ***S**, the collection of these balls*



Čech Complexes

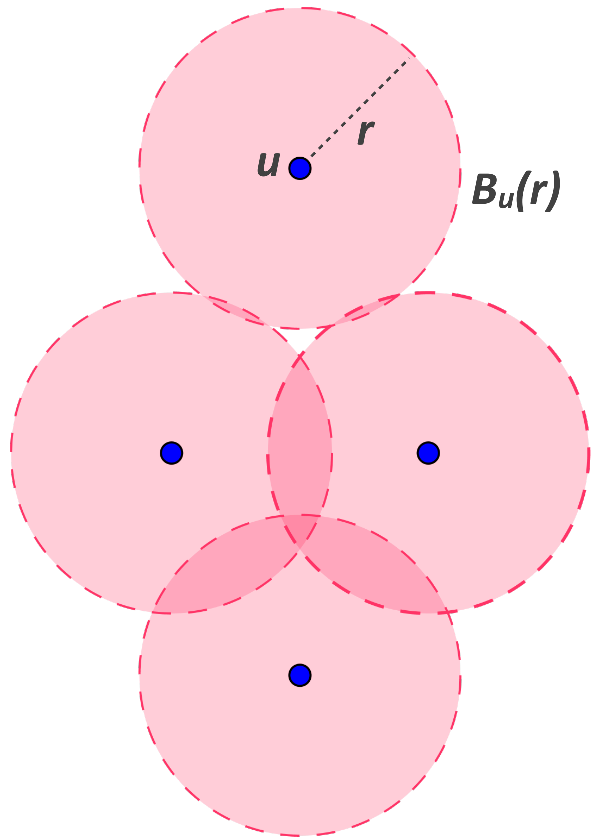
Definition:

Given a finite set of points V in $\mathbb{R}^n$, let us consider:

- ***$B_u(r)$***, *the* ***closed ball*** *with* ***center u*** $\in$ *V and* ***radius r***
- ***S***, *the collection of these balls*

The ***Čech complex Čech(r)*** of V of radius r is the ***nerve of S***

$$\check{C}ech(r) := \{\sigma \subseteq V \mid \bigcap_{u \in \sigma} B_u(r) \neq \emptyset\}$$



Čech Complexes

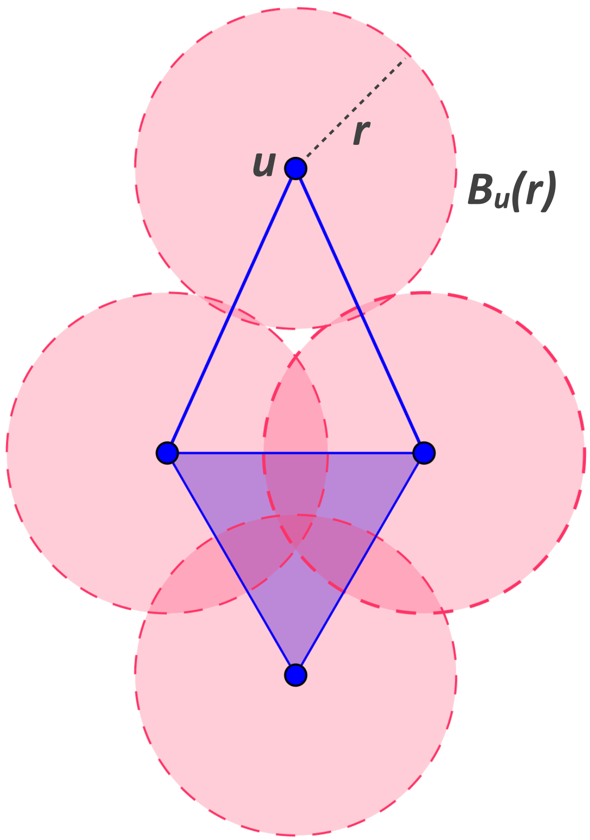
Definition:

Given a finite set of points V in $\mathbb{R}^n$, let us consider:

- ***$B_u(r)$****, the* ***closed ball*** *with* ***center u*** *$\in$ V and* ***radius r***
- ***S****, the collection of these balls*

The ***Čech complex Čech(r)*** of V of radius r is the ***nerve of S***

$$\check{C}ech(r) := \{\sigma \subseteq V \mid \bigcap_{u \in \sigma} B_u(r) \neq \emptyset\}$$



Čech Complexes

Definition:

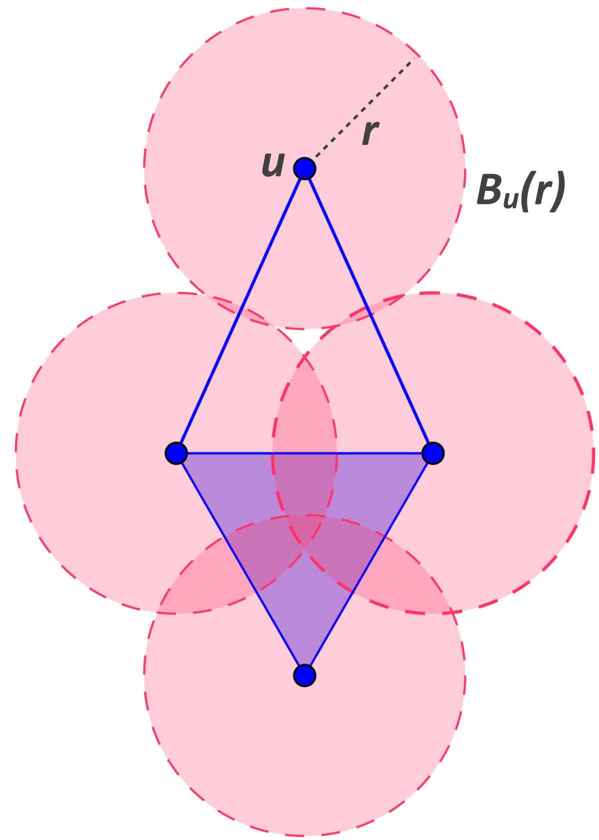
Given a finite set of points V in $\mathbb{R}^n$, let us consider:

- ***$B_u(r)$**, the **closed ball** with **center u** ∈ V and **radius r***
- ***S**, the collection of these balls*

The ***Čech complex Čech(r)*** of V of radius r is the ***nerve of S***

$$\check{C}ech(r) := \{\sigma \subseteq V \mid \bigcap_{u \in \sigma} B_u(r) \neq \emptyset\}$$

In practice, **infeasible construction**

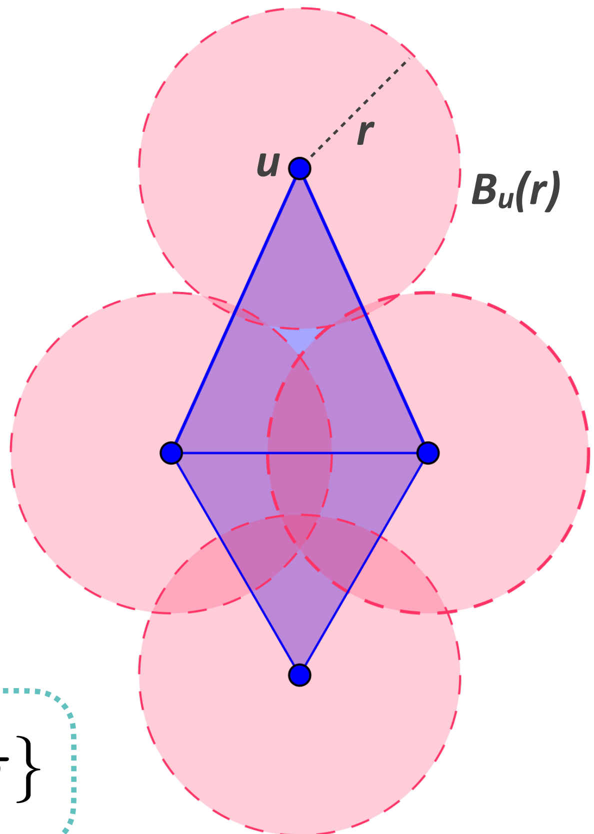


Vietoris-Rips Complexes

Definition:

Given a finite set of points V in $\mathbb{R}^n$,

The ***Vietoris-Rips complex VR(r)*** of V and r is the ***abstract simplicial complex*** consisting of all ***subsets of diameter at most 2r***



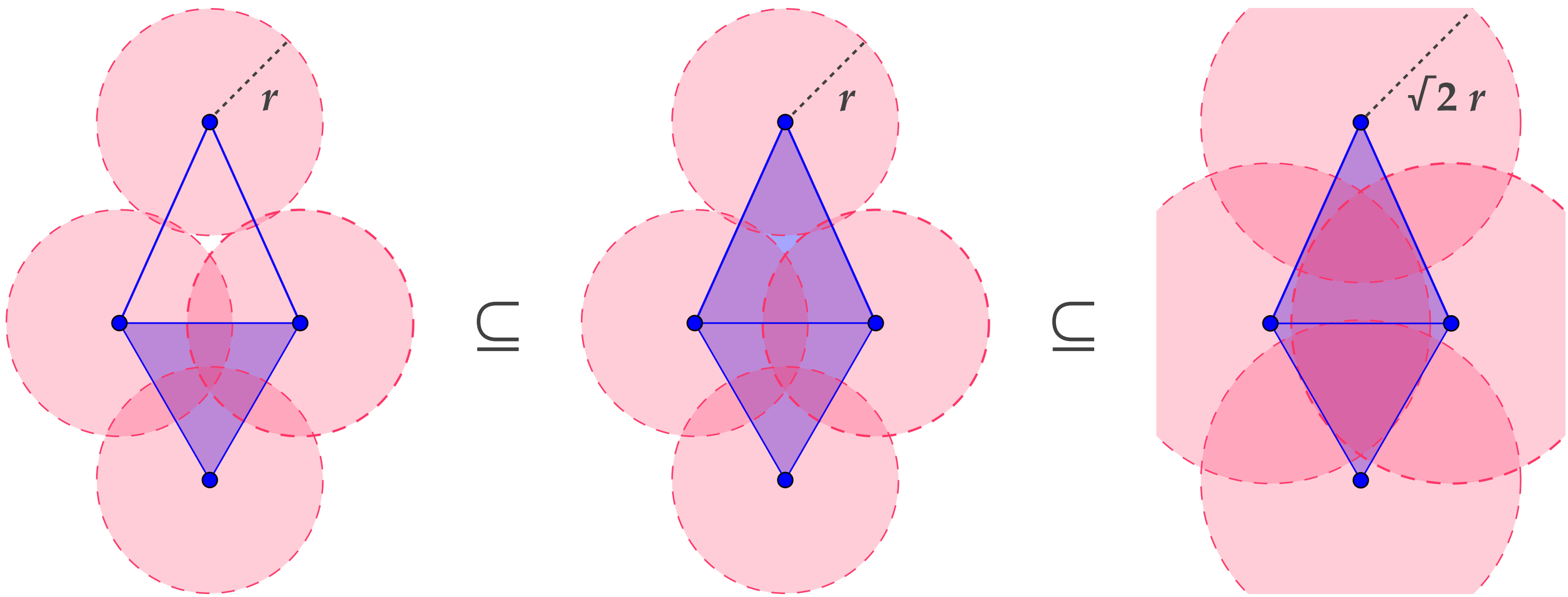
Formally,

$$VR(r) := \{\sigma \subseteq V \mid d(u,v) \leq 2r, \forall\, u,v \in \sigma\}$$

Vietoris-Rips Complexes

Properties:

- $\check{C}ech(r) \subseteq VR(r) \subseteq \check{C}ech(\sqrt{2}r)$

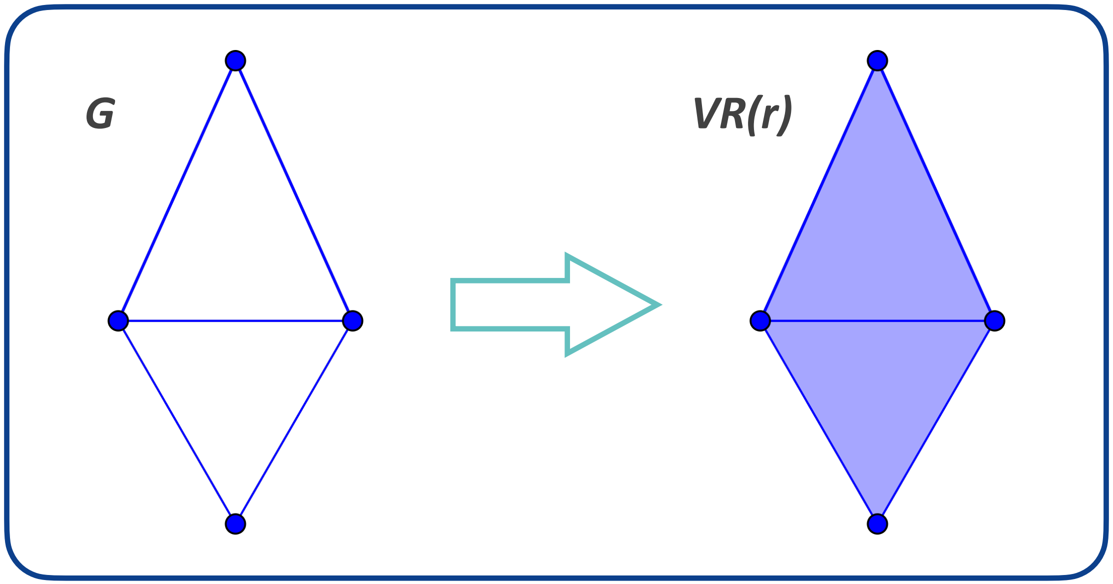


$\check{C}ech(r)$ $VR(r)$ $\check{C}ech(\sqrt{2}r)$

Vietoris-Rips Complexes

Properties:

- $\check{C}ech(r) \subseteq VR(r) \subseteq \check{C}ech(\sqrt{2}r)$
- ***VR(r)*** is completely determined by its 1-skeleton
  - I.e. the graph ***G*** of its vertices and its edges



Vietoris-Rips Complexes

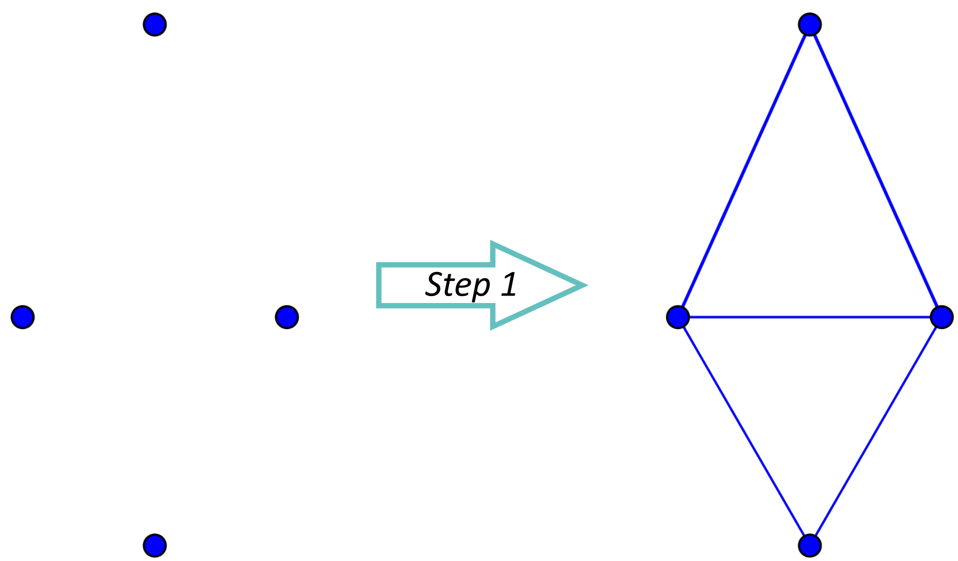
Algorithms:

<u>*Input:*</u> A finite set of points V in $\mathbb{R}^n$ and a real positive number r

<u>*Output:*</u> The Vietoris-Rips complex VR(r)

A ***two-step*** approach is typically adopted:

- ***Step 1 - Skeleton Computation:***
  - *Exact ($O(|V|^2)$ time complexity)*
  - *Approximate*
  - *Randomized*
  - *Landmarking*
- ***Step 2 - Vietoris-Rips Expansion:***
  - *Inductive*
  - *Incremental*
  - *Maximal*



Vietoris-Rips Complexes

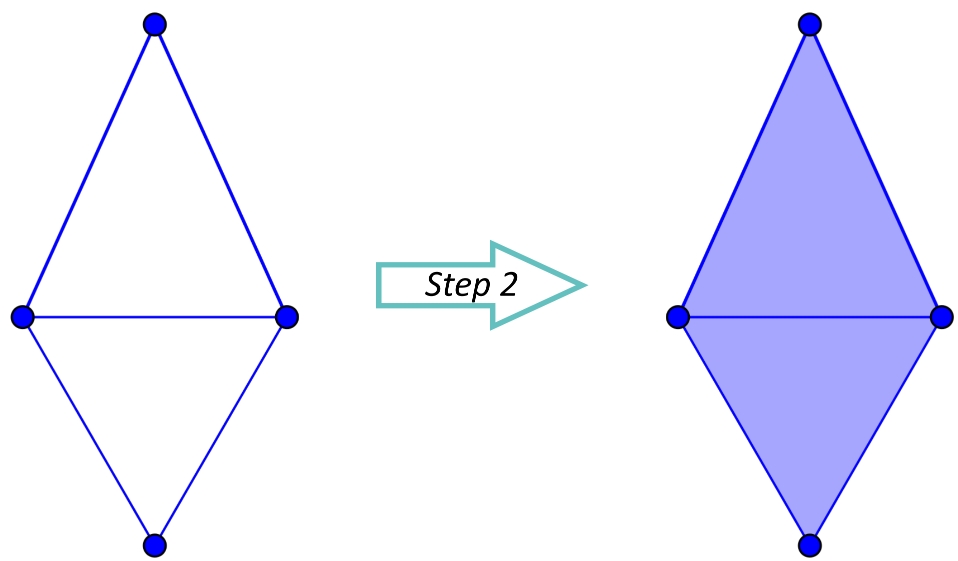
Algorithms:

Input: A finite set of points V in $\mathbb{R}^n$ and a real positive number r

Output: The Vietoris-Rips complex VR(r)

A ***two-step*** approach is typically adopted:

- ***Step 1 - Skeleton Computation:***
  - *Exact ($O(|V|^2)$ time complexity)*
  - *Approximate*
  - *Randomized*
  - *Landmarking*
- ***Step 2 - Vietoris-Rips Expansion:***
  - *Inductive*
  - *Incremental*
  - *Maximal*



Vietoris-Rips Complexes

Inductive VR expansion:

Input: The 1-skeleton $G = (V, E)$ of $VR(r)$

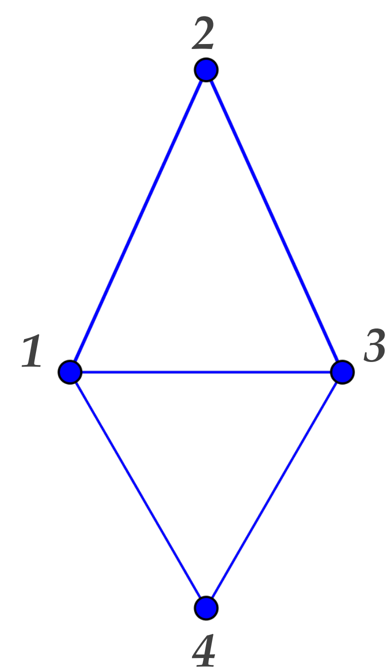
Output: The k-skeleton K of the Vietoris-Rips complex $VR(r)$

INDUCTIVE-VR(G, k)

$K = V \cup E$
for $i = 1$ **to** k
  foreach i-simplex $\sigma \in K$
    $N = \cap_{u \in \sigma}$ LOWER-NBRS(G, u)
      foreach $v \in N$
        $K = K \cup \{ \sigma \cup \{v\}\}$
  return K

LOWER-NBRS(G, u)

  return $\{v \in V \mid v < u, (u, v) \in E\}$



Vietoris-Rips Complexes

Inductive VR expansion:

Input: The 1-skeleton G = (V, E) of VR(r)

Output: The k-skeleton K of the Vietoris-Rips complex VR(r)

INDUCTIVE-VR(G, k)

$K = V \cup E$

for $i = 1$ ***to*** k

foreach *i-simplex* $\sigma \in K$

$N = \cap_{u\in\sigma}$ LOWER-NBRS(G, u)

foreach $v \in N$

$K = K \cup \{\sigma \cup \{v\}\}$

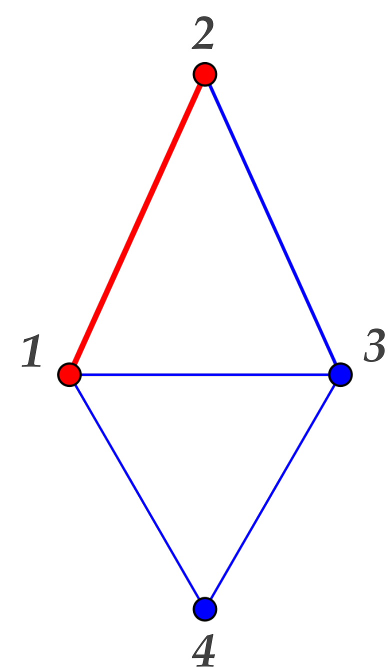
return K

LOWER-NBRS(G, u)

return $\{v \in V \mid v < u, (u, v) \in E\}$

$\sigma = (1, 2)$

$N = \{\ \}$



Vietoris-Rips Complexes

Inductive VR expansion:

Input: The 1-skeleton G = (V, E) of VR(r)

Output: The k-skeleton K of the Vietoris-Rips complex VR(r)

INDUCTIVE-VR(G, k)

$K = V \cup E$

for $i = 1$ **to** k

foreach i-simplex $\sigma \in K$

$N = \cap_{u \in \sigma}$ LOWER-NBRS(G, u)

foreach $v \in N$

$K = K \cup \{ \sigma \cup \{v\}\}$

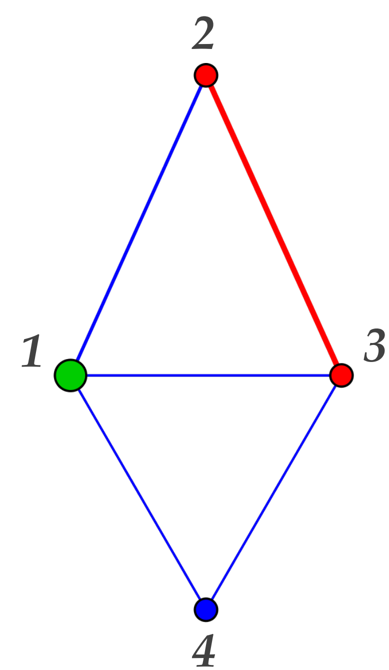
return K

LOWER-NBRS(G, u)

return $\{v \in V \mid v < u , (u, v) \in E\}$

$\sigma = (2, 3)$

$N = \{1\}$



Vietoris-Rips Complexes

Inductive VR expansion:

Input: The 1-skeleton G = (V, E) of VR(r)

Output: The k-skeleton K of the Vietoris-Rips complex VR(r)

INDUCTIVE-VR(G, k)

$K = V \cup E$

for $i = 1$ **to** k

foreach i-simplex $\sigma \in K$

$N = \cap_{u \in \sigma}$ LOWER-NBRS(G, u)

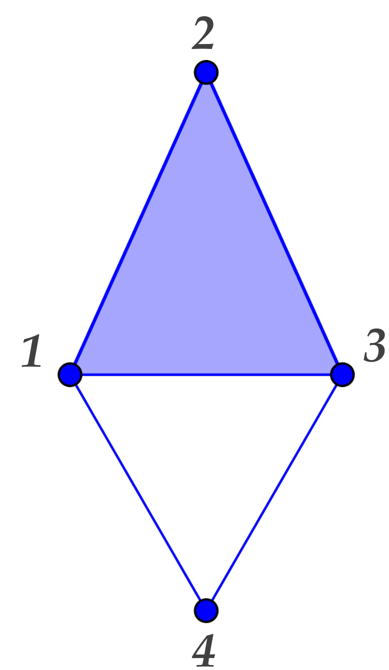
foreach $v \in N$

$K = K \cup \{ \sigma \cup \{v\}\}$

return K

LOWER-NBRS(G, u)

return $\{v \in V \mid v < u , (u, v) \in E\}$



Vietoris-Rips Complexes

Inductive VR expansion:

Input: The 1-skeleton G = (V, E) of VR(r)

Output: The k-skeleton K of the Vietoris-Rips complex VR(r)

INDUCTIVE-VR(G, k)

$K = V \cup E$

for $i = 1$ ***to*** k

foreach *i-simplex* $\sigma \in K$

$N = \cap_{u \in \sigma}$ *LOWER-NBRS(G, u)*

foreach $v \in N$

$K = K \cup \{ \sigma \cup \{v\}\}$

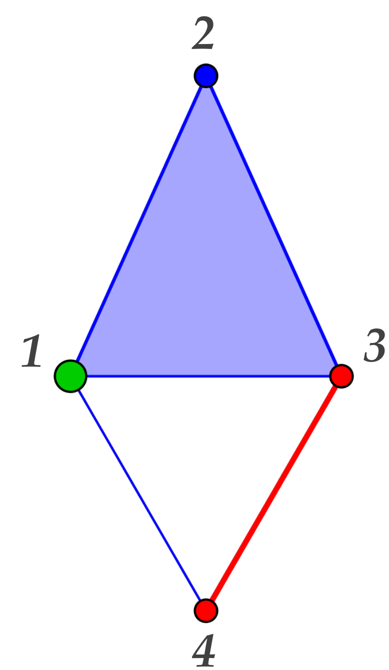
return K

LOWER-NBRS(G, u)

return $\{v \in V \mid v < u, (u, v) \in E\}$

$\sigma = (3, 4)$

$N = \{1\}$



Vietoris-Rips Complexes

Inductive VR expansion:

Input: The 1-skeleton G = (V, E) of VR(r)

Output: The k-skeleton K of the Vietoris-Rips complex VR(r)

INDUCTIVE-VR(G, k)

$K = V \cup E$

for $i = 1$ **to** k

foreach i-simplex $\sigma \in K$

$N = \cap_{u \in \sigma}$ LOWER-NBRS(G, u)

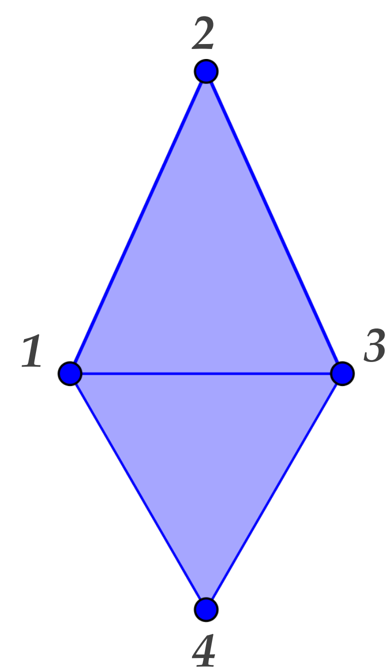
foreach $v \in N$

$K = K \cup \{ \sigma \cup \{v\}\}$

return K

LOWER-NBRS(G, u)

return $\{v \in V \mid v < u , (u, v) \in E\}$



From Data to Complexes

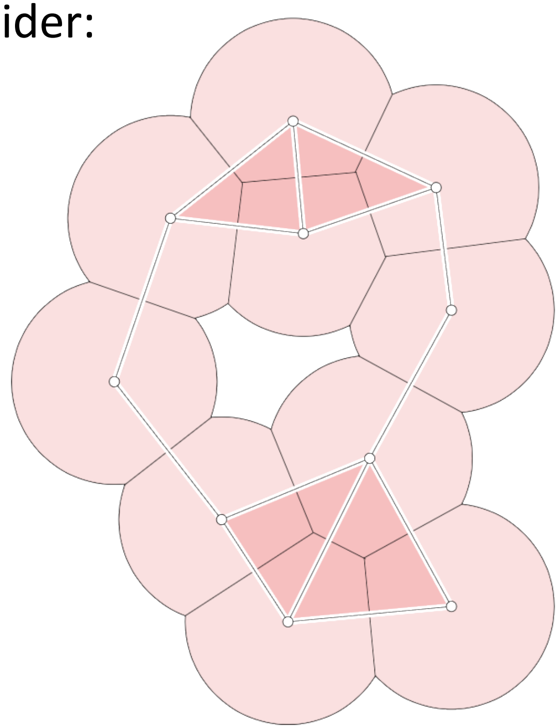
		
Delaunay triangulation	Bounded Dimension	Trivial Homology
Čech/VR complex	“Real” Homology	High Dimension Large Size

Alpha-Shapes

Definition:

Given a finite set of points V in general position of $\mathbb{R}^n$, let us consider:

- ***$A_u(r) := B_u(r) \cap R_V(u)$**, the **intersection** of the **closed ball** with **center u** $\in$ V and **radius r** and the **Voronoi region** of u*
- ***S**, the collection of these convex sets*



The ***alpha-shape Alpha(r)*** of V of radius r is the ***nerve of S***

Formally,

$$Alpha(r) := \{\sigma \subseteq V \mid \bigcap_{u \in \sigma} A_u(r) \neq \emptyset\}$$

$$A_u(r) \subseteq B_u(r) \implies Alpha(r) \subseteq \check{C}ech(r)$$

Image from [Edelsbrunner, Harer 2010]

Witness Complexes

Motivation:

The “shape” of a point cloud can be captured ***without considering all the input points***

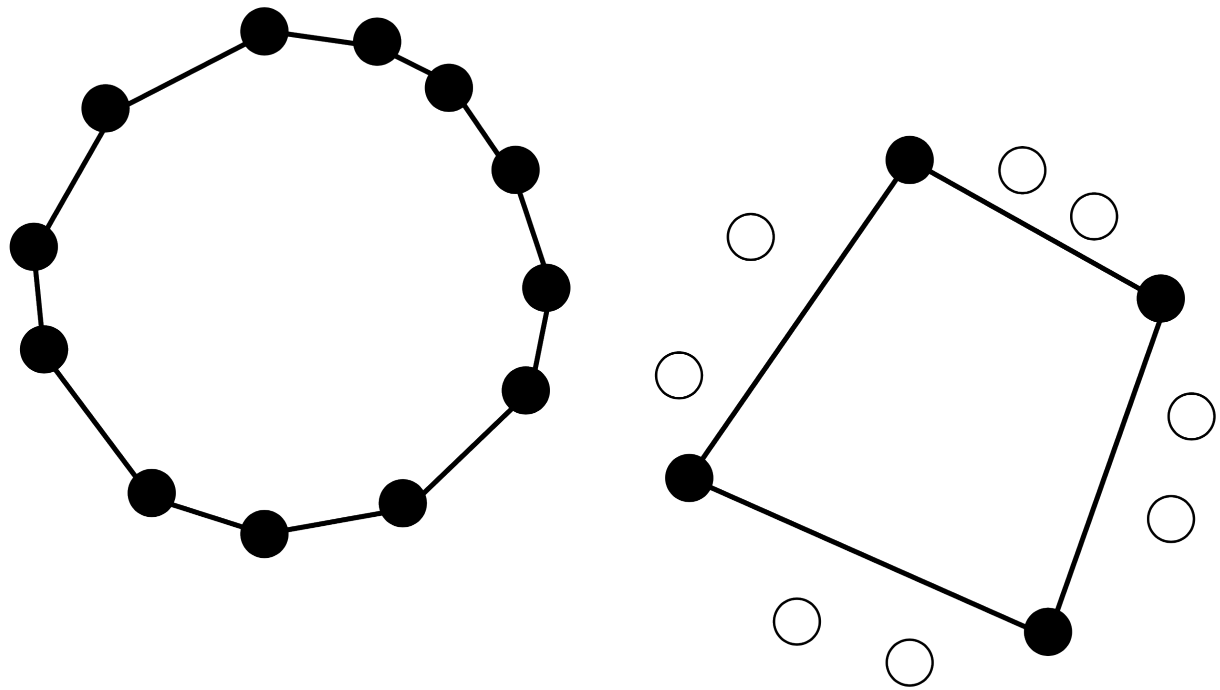
Definitions:

- ***Landmarks:***

  Selected points

- ***Witnesses:***

  Remaining points



Images from [de Silva, Carlsson 2004]

Witness Complexes

Definition:

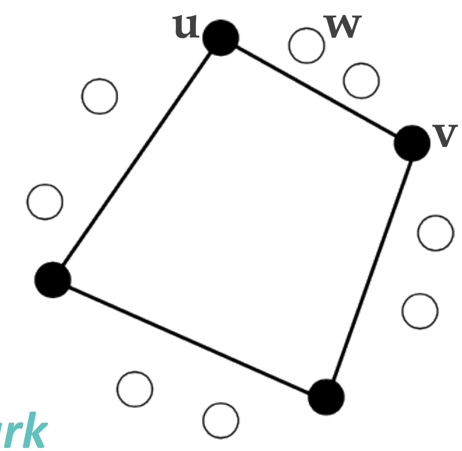
The ***witness complex W(r)*** of radius *r* is defined by:

- *u is in W(r) if u is a landmark*
- *(u, v) is in W(r) if there exists a witness w such that*

$$max\{d(u,w), d(v,w)\} \leq m_w + r$$

where m_w : = the distance of w from the 2nd closest landmark

- *the i-simplex σ is in W(r) if all its edges belong to W(r)*



$W_0(r)$ is defined by setting m_w = 0 for any witness w

$$W_0(r) \subseteq VR(r) \subseteq W_0(2r)$$

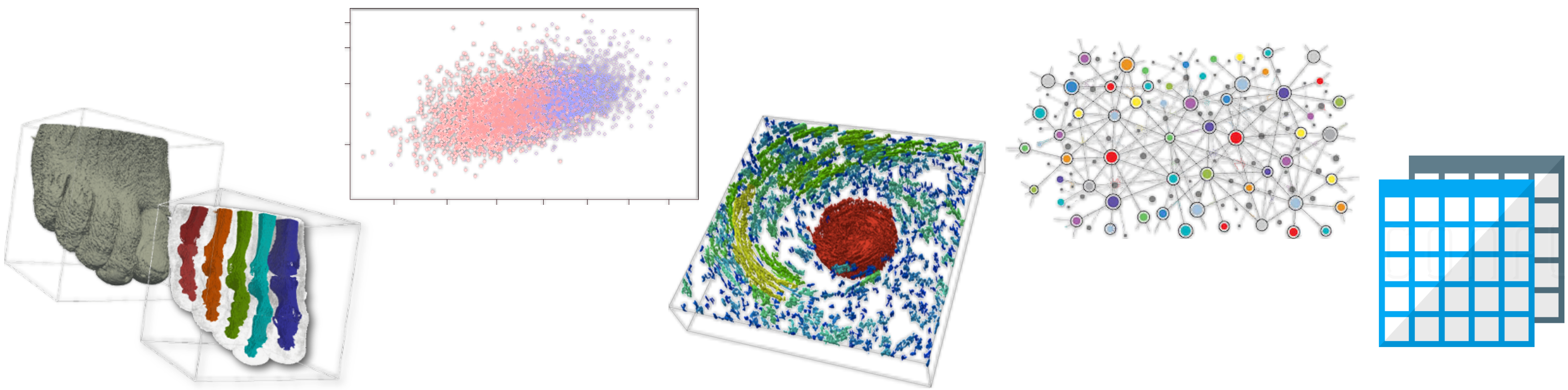
From Data to Complexes

Most of the presented constructions can be ***generalized/adapted*** to the case of

a finite collection of elements endowed with a notion of proximity*

enabling to cover a **wide plethora of datasets**

E.g. complex networks, weighted graphs, multivariate functions, matrices, tensors, etc.



More properly, a* ***semi-metric**, i.e. a distance not necessarily satisfying the triangle inequality*

From Data to Complexes

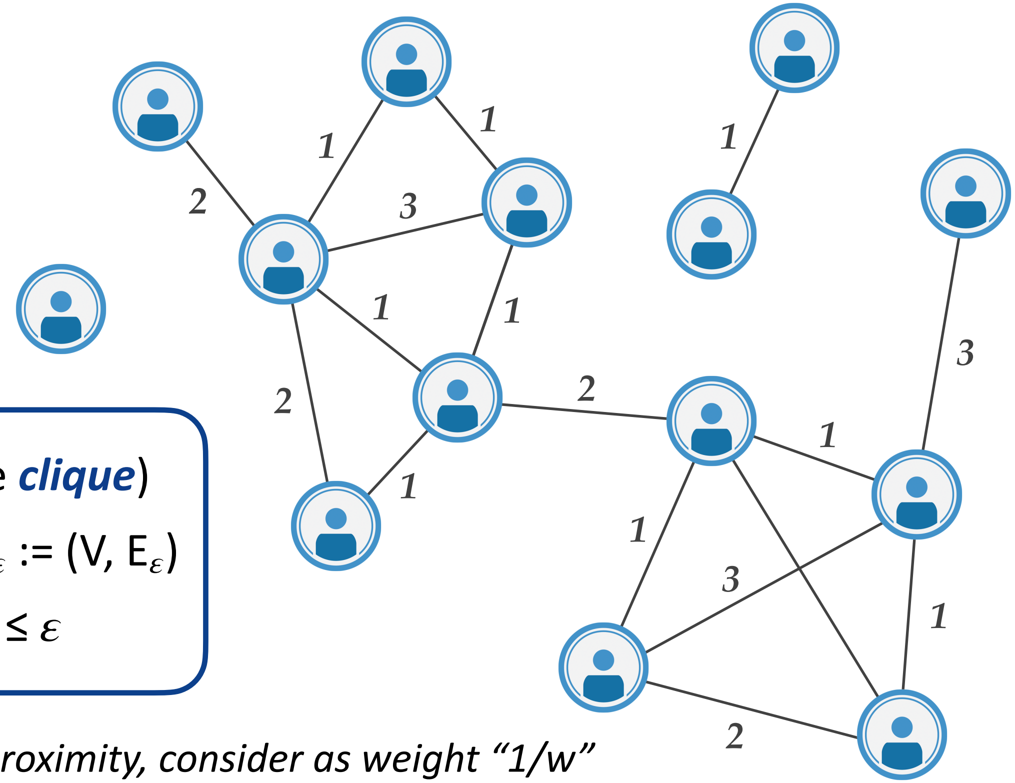
Flag Complex of a Weighted Network:

Let G := (V, E, w: E $\longrightarrow \mathbb{R}$) be a ***weighted undirected graph*** representing a ***network***:

Nodes of V ⟺ *Entities or individuals*

Arcs of E ⟺ *Ties between entities*

Weights of w ⟺ *Proximity of entities**

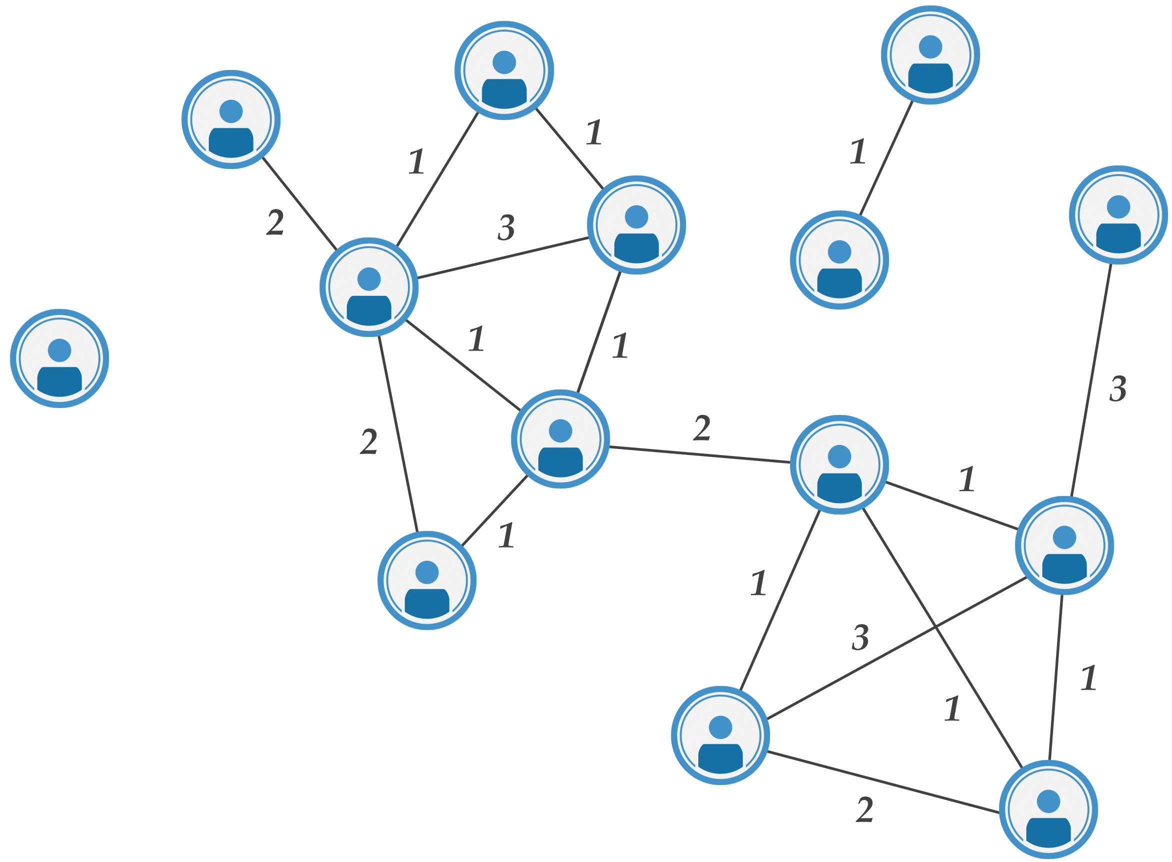


Fixed a ***weight threshold*** ε, the ***flag*** (or the ***clique***) ***complex*** is the ***VR expansion*** of the graph G_ε := (V, E_ε) where E_ε are the arcs of E with weight ≤ ε

**If w represents tie strengths rather than node proximity, consider as weight "1/w"*

From Data to Complexes

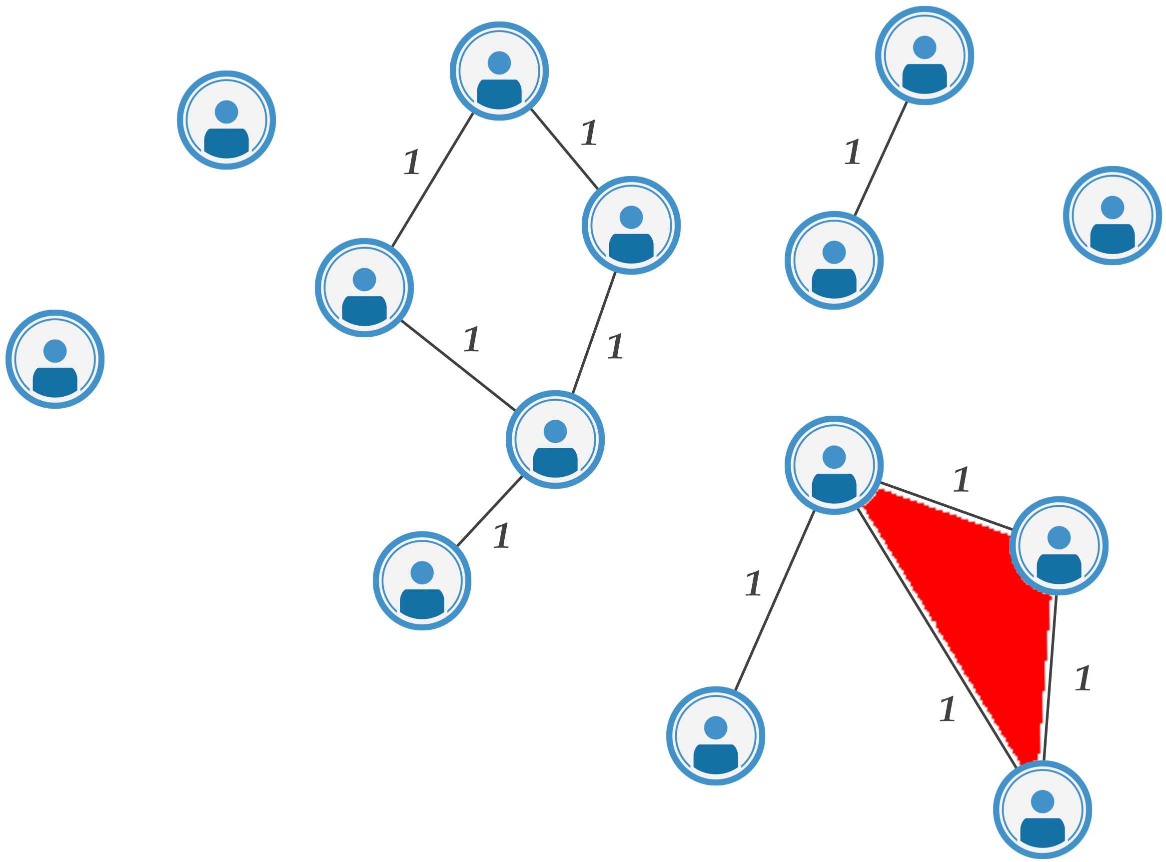
Flag Complex of a Weighted Network:



From Data to Complexes

Flag Complex of a Weighted Network:

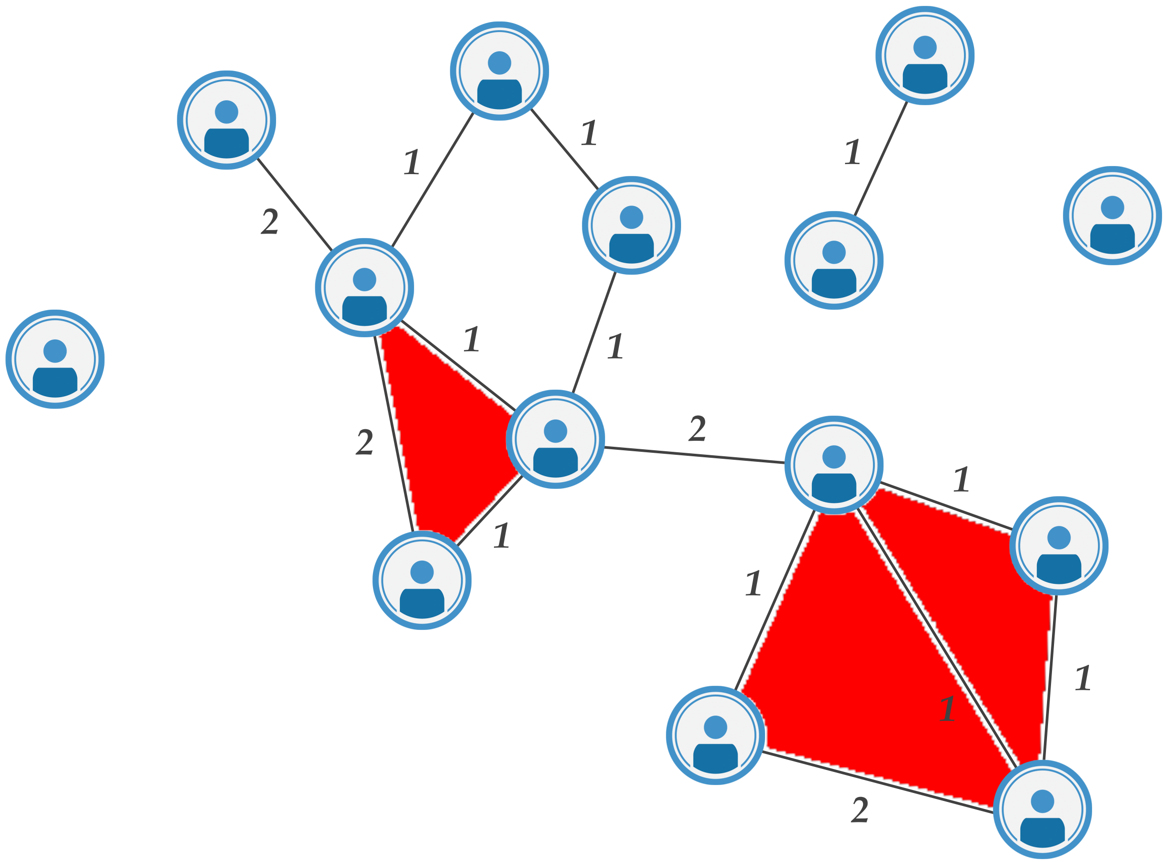
$\varepsilon = 1$



From Data to Complexes

Flag Complex of a Weighted Network:

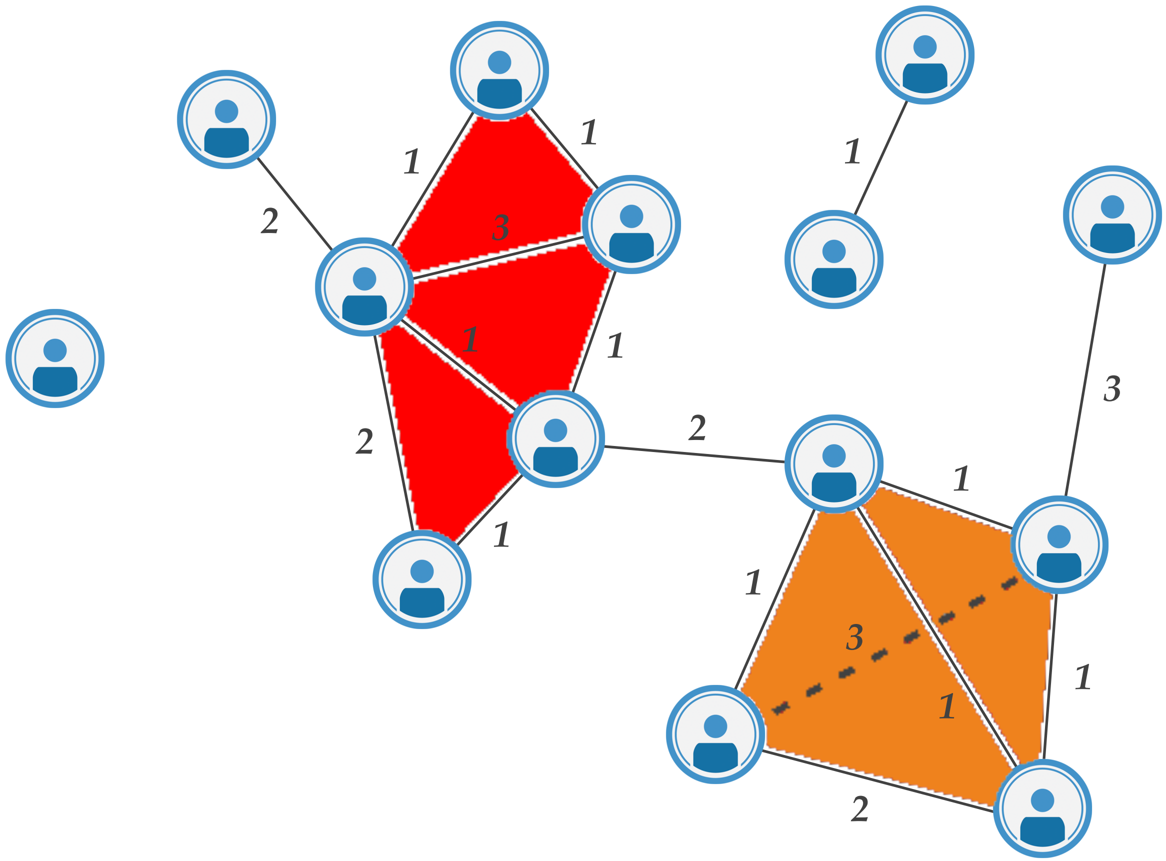
$\varepsilon = 2$



From Data to Complexes

Flag Complex of a Weighted Network:

$\varepsilon = 3$



Bibliography

General References:

- ***Books on TDA:***
  - H. Edelsbrunner, J. Harer. ***Computational topology: an introduction***. American Mathematical Soc., 2010.
  - R. W. Ghrist. ***Elementary applied topology***. Seattle: Createspace, 2014.
- ***Papers on TDA:***
  - G. Carlsson. ***Topology and data***. Bulletin of the American Mathematical Society 46.2, pages 255-308, 2009.

Today's References:

- ***Simplicial Complexes:***
  - J. R. Munkres. ***Elements of algebraic topology***. CRC Press, 2018.
- ***From Data to Complexes:***
  - H. Edelsbrunner, ***Geometry and Topology for Mesh Generation***. Cambridge University Press, 2001.
  - V. de Silva, G. Carlsson. ***Topological estimation using witness complexes***. *SPBG* 4, pages 157-166, 2004.
  - A. Zomorodian, ***Fast construction of the Vietoris-Rips complex***. Computers & Graphics 34.3, pages 263-271, 2010.
  - H. Edelsbrunner. ***Algorithms in Combinatorial Geometry***. Springer Science & Business Media, 2012.