Computational Topology for Data Analysis: Notes from Book by

Tamal Krishna Dey
Department of Computer Science
Purdue University
West Lafayette, Indiana, USA 46907

Yusu Wang
Halıcıoğlu Data Science Institute
University of California, San Diego
La Jolla, California, USA 92093

Topic 2: Complexes

This topic introduces a very basic tool on which topological data analysis (TDA) is built. It is called simplicial complexes. Data supplied as a discrete set of points do not have an interesting topology. Usually, we construct a scaffold on top of it which is commonly taken as a simplicial complex. It consists of vertices at the data points, edges connecting them, triangles, tetrahedra and their higher dimensional analogues to establish higher order connectivity. Section 2.1 formalizes this construction. There are different kinds of simplicial complexes. Some are easier to compute, but take more space. The others are more sparse, but takes more time to compute. Sections 2.2 presents an important construction called the *nerve* and a complex called the Čech complex which is defined on this construction. This section also presents a commonly used complex in topological data analysis called the Vietoris-Rips complex that interleaves with the Čech complexes in terms of containment. In section 2.3, we introduce some of the complexes which are sparser in size than the Vietoris-Rips or Čech complexes.

2.1 Simplicial complex

A complex is a collection of some basic elements that satisfy certain properties. In a simplicial complex, these basic elements are simplices.

Definition 1 (Simplex)**.** A *k-simplex* σ in an Euclidean space $\mathbb{R}^m$ is the convex hull[1] of a set P of $k+1$ affinely independent points in $\mathbb{R}^m$. In particular, a 0-simplex is a *vertex*, a 1-simplex is an *edge*, a 2-simplex is a *triangle*, and a 3-simplex is a *tetrahedron*. A k-simplex is said to have *dimension* k. For $0 \leq k' \leq k$, a k'*-face* (or, simply a *face*) of σ is a k'-simplex that is the convex hull of a nonempty subset of P. Faces of σ come in all dimensions from zero (σ's vertices) to k; and σ is a face of σ. A *proper face* of σ is a simplex that is the convex hull of a proper subset of P; i.e. any face except σ. The $(k-1)$-faces of σ are called *facets* of σ; σ has $k+1$ facets.

In Figure 2.1(a), triangle *abc* is a 2-simplex which has three vertices as 0-faces and three edges as 1-faces. These are proper faces out of which edges are its facets. Similarly, a tetrahedron has four 0-faces (vertices), six 1-faces (edges), four 2-faces (triangles), and one 3-face (tetrahedron itself) out of which vertices, edges, triangles are proper. The triangles are facets.

Definition 2 (Geometric simplicial complex)**.** A *geometric simplicial complex* K, also known as a *triangulation*, is a set containing finitely[2] many simplices that satisfies the following two restrictions.

- K contains every face of each simplex in K.
- For any two simplices $\sigma, \tau \in K$, their intersection $\sigma \cap \tau$ is either empty or a face of both σ and τ.

---

[1]Convex hull of a set of given points $p_0, \ldots, p_k$ in $\mathbb{R}^m$ is the set of all points $x \in \mathbb{R}^m$ that are convex combination of the given points, i.e., $x = \Sigma_{i=0}^k \alpha_i p_i$ for $\alpha_i \geq 0$ and $\Sigma \alpha_i = 1$.

[2]Topologists usually define complexes so they have countable cardinality. We restrict complexes to finite cardinality here.

The *dimension* k of K is the maximum dimension of any simplex in K which is why we also refer it as a simplicial k-complex.

The above definition of simplicial complexes is very geometric which is why they are referred as geometric simplicial complexes. Figure 2.1 shows such a geometric simplicial 2-complex in $\mathbb{R}^2$ (left) and another in $\mathbb{R}^3$ (right). There is a parallel notion of simplicial complexes that is devoid of geometry.

Definition 3 (Abstract simplicial complex)**.** A collection K of subsets of a given set $V(K)$ is an *abstract simplicial complex* if every element $\sigma \in K$ has all of its subsets $\sigma' \subseteq \sigma$ also in K. Each such subset σ' with $|\sigma'| = k' + 1$ is called a *k'-face* (or, simply a *face*) of σ and σ with $|\sigma| = k + 1$ is a *k-coface* (or, simply a *coface*) of σ'. Sometimes, σ' is also called a face of σ with *co-dimension* $k - k'$. The elements of $V(K)$ are the vertices of K. Each (sub)set in K is a simplex whose dimension equals its cardinality minus 1.

A geometric simplicial complex K in $\mathbb{R}^m$ is called a *geometric realization* of an abstract simplicial complex K' if and only if there is an embedding $e : V(K') \to \mathbb{R}^m$ that takes every k-simplex $\{v_0, v_1, \ldots, v_k\}$ in K' to a k-simplex in K that is the convex hull of $e(v_0), e(v_1), \ldots, e(v_k)$. For example, the complex drawn in $\mathbb{R}^2$ in Figure 2.1(left) is a geometric realization of the abstract complex with vertices a, b, c, d, e, f, eight 1-simplices $\{a, b\}, \{a, d\}, \{a, f\}, \{b, c\}, \{b.d\}, \{c, d\}, \{d, e\}, \{d, f\}$, and one 2-simplex $\{a, b, d\}$.

An abstract simplicial complex K with m vertices can always be geometrically realized in $\mathbb{R}^{m-1}$ as a subcomplex of a geometric $(m-1)$-simplex. To make the realization canonical, we choose the $(m-1)$-simplex to be in $\mathbb{R}^m$ with a vertex v_i having the ith coordinate to be 1 and all other coordinates 0. We define K's underlying space as the underlying space of this canonical geometric realization.

Definition 4 (Underlying space)**.** The *underlying space* of an abstract simplicial complex K, denoted $|K|$, is the pointwise union of its simplices in its canonical geometric realization; that is, $|K| = \bigcup_{\sigma \in K} |\sigma|$ where $|\sigma|$ is the restriction of this realization on σ. In case K is geometric, its geometric realization can be taken as itself.

Because of the equivalence between geometric and abstract simplicial complexes, we drop the qualifiers "geometric" and "abstract" and call them simply as simplicial complexes when it is clear from the context which one we actually mean. Also, sometimes, we denote a simplex $\sigma = \{v_0, v_1, \cdots, v_k\}$ simply as $v_0 v_1 \cdots v_k$.

Definition 5 (*k*-skeleton)**.** The k-skeleton of a simplicial complex K, denoted by K^k, is the subcomplex formed by all of its k-dimensional simplices and their faces.

In Figure 2.1, the 1-skeleton of the simplicial complex on left consists of six vertices a, b, c, d, e, f and eight edges adjoining them.

Stars and links. Given a simplex $\tau \in K$, its *star* in K is the set of simplices that have τ as a face, denoted by $\mathrm{St}(\tau) = \{\sigma \in K \mid \tau \subseteq \sigma\}$ (recall that $\tau \subseteq \sigma$ means that τ is a face of σ). Generally, the star is not closed under face relation and hence is not a simplicial complex. We can make it so by adding all missing faces. The result is the *closed star*, denoted by $\overline{\mathrm{St}}(\tau) = \bigcup_{\sigma \in \mathrm{St}(\tau)} \{\sigma\} \cup \{\sigma' \in K \mid$

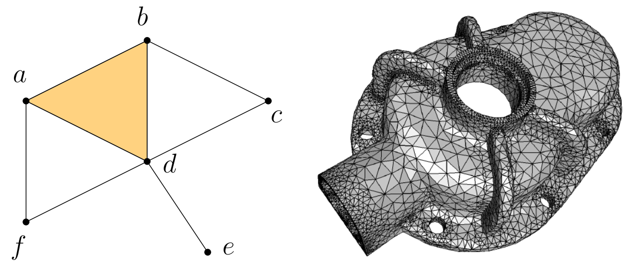

Figure 2.1: (left) A simplicial complex with six vertices, eight edges, and one triangle, (right) A simplicial 2-complex triangulating a 2-manifold in $\mathbb{R}^3$.

$\sigma' \subset \sigma\}$, which is also the smallest subcomplex that contains the star. The *link of* τ consists of the set of simplices in the closed star that are disjoint from τ, that is, $\mathrm{Lk}(\tau) = \{\sigma \in \overline{\mathrm{St}}(\tau) \mid \sigma \cap \tau = \emptyset\}$. Intuitively, we can think of the star (resp. the closed star) of a vertex as an open (resp. closed) neighborhood around it, and the link as the boundary of that neighborhood.

In Figure 2.1(left), we have

- $\mathrm{St}(a) = \{\{a\}, \{a,b\}, \{a,d\}, \{a,f\}, \{a,b,d\}\}$, $\overline{\mathrm{St}}(a) = \mathrm{St}(a) \cup \{\{b\}, \{d\}, \{f\}, \{b,d\}\}$
- $\mathrm{St}(f) = \{\{f\}, \{a,f\}, \{d,f\}\}$, $\overline{\mathrm{St}}(f) = \mathrm{St}(f) \cup \{\{a\}, \{d\}\}$
- $\mathrm{St}(\{a,b\}) = \{\{a,b\}, \{a,b,d\}\}$, $\overline{\mathrm{St}}(\{a,b\}) = \mathrm{St}(\{a,b\}) \cup \{\{a\}, \{b\}, \{d\}, \{a,d\}, \{b,d\}\}$
- $\mathrm{Lk}(a) = \{\{b\}, \{d\}, \{f\}, \{b,d\}\}$, $\mathrm{Lk}(f) = \{\{a\}, \{d\}\}$, $\mathrm{Lk}(\{a,b\}) = \{\{d\}\}$.

Triangulation of a manifold. Given a simplicial complex K and a manifold M, we say that K is a triangulation of M if the underlying space $|K|$ is homeomorphic to M. Note that if M is a k-manifold, the dimension of K is also k. Furthermore, for any vertex $v \in K$, the underlying space $|\mathrm{St}(v)|$ of the star $\mathrm{St}(v)$ is homeomorphic to the open k-ball $\mathbb{B}^k_o$ if v maps to an interior point in M and to the k-dimensional halfspace $\mathbb{H}^k$ if v maps to a point on the boundary of M. The underlying space $|\mathrm{Lk}(v)|$ of the link $\mathrm{Lk}(v)$ is homeomorphic to $(k-1)$-sphere $\mathbb{S}^{k-1}$ if v maps to interior and to a closed $(k-1)$-ball $\overline{\mathbb{B}^{k-1}_o}$ otherwise.

Simplicial map. Corresponding to the continuous functions (maps) between topological spaces, we have a notion called simplicial map between simplicial complexes.

Definition 6 (Simplicial map)**.** A map $f : K_1 \to K_2$ is called *simplicial* if for every simplex $\{v_0, \ldots, v_k\} \in K_1$, we have the simplex $\{f(v_0), \ldots, f(v_k)\}$ in K_2.

A simplicial map is called a *vertex map* if the domain and codomain of f are only vertex sets $V(K_1)$ and $V(K_2)$ respectively. Every simplicial map is associated with a vertex map. However, a vertex map $f : V(K_1) \to V(K_2)$ does not necessarily extend to a simplicial map from K_1 to K_2.

Fact 1. *Every continuous map $f : |K_1| \to |K_2|$ can be approximated arbitrarily closely by simplicial maps on appropriate subdivisions of K_1 and K_2.*

There is also a counterpart of homotopic maps in simplicial setting.

Definition 7 (Contiguous map)**.** Two simplicial maps $f_1 : K_1 \to K_2$, $f_2 : K_1 \to K_2$ are contiguous if for every simplex $\sigma \in K_1$, $f_1(\sigma) \cup f_2(\sigma)$ is a simplex in K_2.

Contiguous maps play an important role in topological analysis. We use a result involving contiguous maps and homology groups. We defer stating it where we introduce homology groups.

2.2 Nerves, Čech and Rips complex

Recall Definition of covers. A cover of a topological space defines a special simplicial complex called its *nerve*. The nerve plays an important role in bridging topological spaces to complexes which we will see below and also later. We first define the nerve in general terms which can be specialised to covers easily.

Definition 8 (Nerve)**.** Given a finite collection of sets $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$, we define the *nerve* of the set $\mathcal{U}$ to be the simplicial complex $N(\mathcal{U})$ whose vertex set is the index set A, and where a subset $\{\alpha_0, \alpha_1, \ldots, \alpha_k\} \subseteq A$ spans a k-simplex in $N(\mathcal{U})$ if and only if $U_{\alpha_0} \cap U_{\alpha_1} \cap \ldots \cap U_{\alpha_k} \neq \emptyset$.

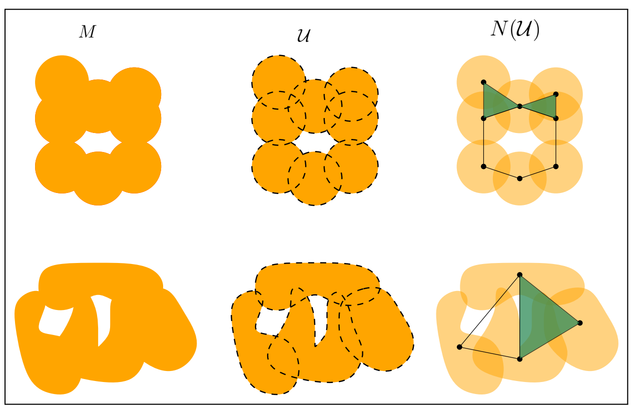

Figure 2.2: Examples of two spaces (left), open covers of them (middle), and their nerves (right). (Top) the intersections of covers are contractible, (bottom) the intersections of covers are not necessarily contractible.

Taking $\mathcal{U}$ to be a cover of a topological space in the above definition, one gets a nerve of a cover. Figure 2.2 shows two topological spaces, their covers, and corresponding nerves.

One important result involving nerves is the so called Nerve Theorem which have different forms that depend on the type of topological spaces and covers. Adapting to our need, we state it for metric spaces which are a special type of topological spaces as we have observed.

Theorem 1 (Nerve Theorem [4, 23])**.** *Given a finite cover $\mathcal{U}$ (open or closed) of a metric space M, the underlying space $|N(\mathcal{U})|$ is homotopy equivalent to M if every non-empty intersection $\cap_{i=0}^{k} U_{\alpha_i}$ of cover elements is homotopy equivalent to a point, that is, contractible.*

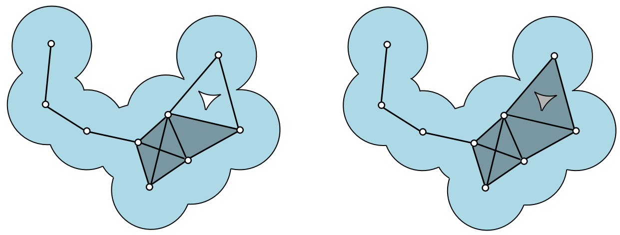

Figure 2.3: Čech complex $\mathbb{C}^r(P)$ and Rips complex $\mathbb{VR}^r(P)$

The cover in the top row of Figure 2.2 satisfy the property of the above theorem and its nerve is homotopy equivalent to M whereas the same is not true for the cover shown in the bottom row.

Given a finite subset P for a metric space (M, d), we can build an abstract simplicial complex called Čech complex with vertices in P using the concept of nerve.

Definition 9 (Čech complex)**.** Let (M, d) be a metric space and P be a finite subset of M. Given a real $r > 0$, the Čech complex $\mathbb{C}^r(P)$ is defined to be the nerve of the set $\{B(p_i, r)\}$ where

$$B(p_i, r) = \{x \in M \mid \mathsf{d}(p_i, x) \leq r\}$$

is the geodesic open ball of radius r centering p_i.

Observe that if M is Euclidean, the balls considered for Čech complex are necessarily convex and hence their intersections are contractible. By Theorem 1, Čech complex in this case is homotopy equivalent to the space of union of the balls. The Čech complex is related to another complex called Vietoris-Rips complex which is often used in topological data analysis.

Definition 10 (Vietoris-Rips complex)**.** Let (P, d) be a finite metric space. Given a real $r > 0$, the Vietoris-Rips (Rips in short) complex is the abstract simplicial complex $\mathbb{VR}^r(P)$ where a simplex $\sigma \in \mathbb{VR}^r(P)$ if and only if $\mathsf{d}(p, q) \leq 2r$ for every pair of vertices of σ.

Notice that the 1-skeleton of $\mathbb{VR}^r(P)$ determines all of its simplices. It is the completion (in terms of simplices) of its 1-skeleton; see Figure 2.3. Also, observe the following fact.

Fact 2. *Let P be a finite subset of a metric space (M, d) where M satisfies the property that, for any real $r > 0$ and two points $p, q \in M$ with $\mathsf{d}(p, q) \leq 2r$, the metric balls $B(p, r)$ and $B(q, r)$ have non-empty intersection. Then, the 1-skeletons of $\mathbb{VR}^r(P)$ and $\mathbb{C}^r(P)$ coincide.*

Notice that if M is Euclidean, it satisfies the condition stated in the above fact and hence for finite point sets in any Euclidean space, Čech and Rips complexes defined with Euclidean balls share the same 1-skeleton. However, for a general finite metric space (P, d), it may happen that for some $p, q \in P$, one has $\mathsf{d}(p, q) \leq 2r$ and $B(p, r) \cap B(q, r) = \emptyset$.

An easy but important observation is that the Rips and Čech complexes interleave.

Proposition 2. *Let P be a finite subset of a metric space (M, d). Then,*

$$\mathbb{C}^r(P) \subseteq \mathbb{VR}^r(P) \subseteq \mathbb{C}^{2r}(P).$$

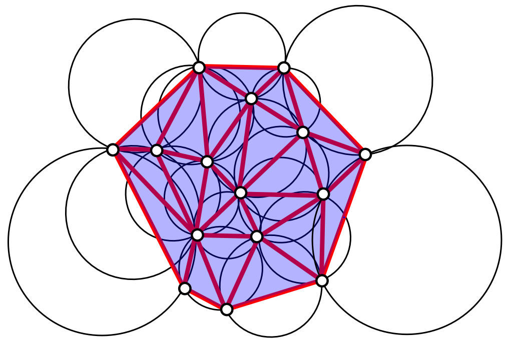

Figure 2.4: Every triangle in a Delaunay complex has an empty open circumdisk.

Proof. The first inclusion is obvious because if there is a point x in the intersection $\cap_{i=1}^{k} B(p_i, r)$, the distances $\mathsf{d}(p_i, p_j)$ for every pair (i, j), $1 \le i, j \le k$, are at most $2r$. It follows that for every simplex $\{p_1, \ldots, p_k\} \in \mathbb{C}^r(P)$ is also in $\mathbb{VR}^r(P)$.

To prove the second inclusion, consider a simplex $\{p_1, \ldots, p_k\} \in \mathbb{VR}^r(P)$. Since by definition of the Rips complex $\mathsf{d}(p_i, p_1) \le 2r$ for every p_i, $i = 1, \ldots, k$, we have $\cap_{i=1}^{k} B(p_i, 2r) \supset p_1 \neq \emptyset$. Then, by definition, $\{p_1, \ldots, p_k\}$ is also a simplex in $\mathbb{C}^{2r}(P)$. □

2.3 Sparse complexes

The Rips and Čech complexes are often too large to handle in practice. For example, the Rips complex with n points in $\mathbb{R}^d$ can have $\Omega(n^d)$ simplices. In practice, they can become large even in dimension as low as three. Just to give a sense of the scale of the problem, we note that the Rips or Čech complex built out of a few thousand points often has triangles in the range of millions. There are other complexes that are much sparser in size because of which they may be preferred sometimes for computations.

2.3.1 Delaunay complex

This is a special complex that can be constructed out of a point set $P \in \mathbb{R}^d$. This complex embeds in $\mathbb{R}^d$ (in the generic setting). Because of its various optimal properties, this complex is used in many applications involving mesh generation, in particular in $\mathbb{R}^2$ and $\mathbb{R}^3$, see [7]. However, computation of Delaunay complexes in high dimensions beyond $\mathbb{R}^3$ can be time intensive, so it is not yet the preferred choice for applications in dimensions beyond $\mathbb{R}^3$.

Definition 11 (Delaunay simplex; Complex)**.** In the context of a finite point set $P \in \mathbb{R}^d$, a k-simplex σ is *Delaunay* if its vertices are in P and there is an open d-ball whose boundary contains its vertices and is *empty*—contains no point in P. Note that any number of points in P can lie on the boundary of this ball. But, for simplicity, we will assume that only the vertices of σ are on the boundary of its empty ball. A *Delaunay complex* of P, denoted Del P, is a (geometric) simplicial complex with vertices in P in which every simplex is Delaunay and $|\mathrm{Del}\, P|$ coincides with the convex hull of P, as illustrated in Figure 2.4.

In $\mathbb{R}^2$, a Delaunay complex of a set of points in general position is made out of Delaunay triangles and all of their lower dimensional faces. Similarly, in $\mathbb{R}^3$, a Delaunay complex is made out of Delaunay tetrahedra and all of their lower dimensional faces.

Fact 3. *Every non-degenerate point set (no $d + 2$ points are co-spherical) admits a unique Delaunay complex.*

Delaunay complexes are dual to the famous Voronoi diagrams defined below.

Definition 12 (Voronoi diagram)**.** Given a finite point set $P \subset \mathbb{R}^d$ in generic position, the Voronoi diagram $\mathrm{Vor}\,(P)$ of P is the tessellation of the embedding space $\mathbb{R}^d$ into convex cells V_p for every $p \in P$ where

$$V_p = \{x \in \mathbb{R}^d \mid d(x, p) \leq d(x, q)\ \forall q \in P\}.$$

A k-face of $\mathrm{Vor}\,(P)$ is the intersection of $(d - k + 1)$ Voronoi cells.

Duality between Delaunay complex and Voronoi diagram is expresed by the duality among their faces. Specifically, a Delaunay k-simplex in $\mathrm{Del}\,(P)$ is dual to a Voronoi $(d - k)$-face in $\mathrm{Vor}\,(P)$. The Voronoi diagram dual to the Delaunay complex in Figure 2.4 is shown in Figure 2.5.

The following optimality properties make Delaunay complexes useful for applications.

Fact 4. *A triangulation of a point set $P \subset \mathbb{R}^d$ is a geometric simplicial complex whose vertex set is P and whose simplices tessellate the convex hull of P. Among all triangulations of a point set $P \subset \mathbb{R}^d$,* $\mathrm{Del}\,P$ *achieves the following optimized criteria:*

1. *In $\mathbb{R}^2$,* $\mathrm{Del}\,P$ *maximizes the minimum angle of triangles in the complex.*
2. *In $\mathbb{R}^2$,* $\mathrm{Del}\,P$ *minimizes the largest circumcircle for triangles in the complex.*
3. *For a simplex in* $\mathrm{Del}\,P$*, let its min-ball be the smallest ball that contains the simplex in it. In all dimensions,* $\mathrm{Del}\,P$ *minimizes the largest min-ball.*

1-skeletons of Delaunay complexes in $\mathbb{R}^2$ are planar graphs and hence they have size $O(n)$ for n points. They can be computed in $\Theta(n \log n)$ time. In $\mathbb{R}^3$, their size grows to $O(n^2)$ and they can be computed in $\Theta(n^2)$ time. In $\mathbb{R}^d$, $d \geq 3$, Delaunay complexes have size $\Theta(n^{\lceil d/2 \rceil})$ and can be computed in optimal time $\Theta(n^{\lceil d/2 \rceil})$ [5].

Alpha complex. Alpha complexes are subcomplexes of the Delaunay complexes which are parameterized by a real $\alpha \geq 0$. For a given point set P and $\alpha \geq 0$, an alpha complex consists of all simplices in $\mathrm{Del}\,(P)$ that have a circumscribing ball of radius at most α. It can also be described alternatively as a nerve. For each point $p \in P$, let B_p^α denote a closed ball of radius α centering p. Consider the closed set D_p defined as follows:

$$D_p^\alpha = \{x \in B_p^\alpha \mid d(x, p) \leq d(x, q)\ \forall q \in P\}$$

The alpha complex $\mathrm{Del}_\alpha(P)$ is the nerve of the closed sets $\{D_p^\alpha\}_{p \in P}$. Another interpretation for alpha complex stems from its relation to the Voronoi diagram of the point set P. Alpha complex contains a k-simplex $\sigma = \{p_0, \ldots, p_k\}$ if and only if $\cup_{p \in P} B_p^\alpha$ meets the intersection of Voronoi cells $V_{p_0} \cap V_{p_1} \cdots \cap V_{p_k}$. Figure 2.5 shows an alpha complex for the point set in Figure 2.4 for an α. The Voronoi diagram is shown with the dotted segments.

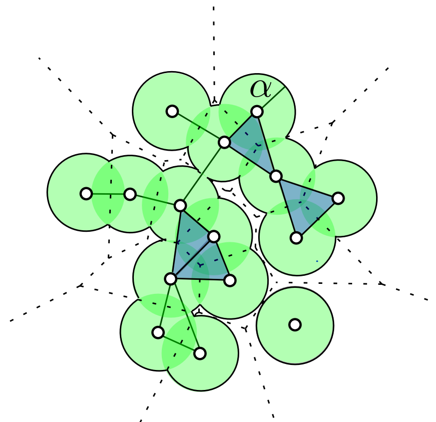

Figure 2.5: Alpha complex of the point set in Figure 2.4 for an α indicated in the figure. The Voronoi diagram of the point set is shown with dotted edges. The traingles and edges in the complex are shown with solid edges which are subset of the Delaunay complex.

2.3.2 Witness complex

The witness complex defined by de Silva and Carlsson [9] sidesteps the size problem by a subsampling strategy. First, we define the witness complex with two finite set, P called the witnesses and Q called the landmarks. The complex is built with vertices in the landmarks where the simplices are defined with a notion of witness from the witness set. Given a finite set P equipped with pairwise distances $\mathsf{d} : P \times P \to \mathbb{R}$, we can build the witness complex on a subsample $Q \subseteq P$.

Definition 13 (Weak witness)**.** Let $Q \subseteq P$ where P is a finite set with a real valued function on pairs $\mathsf{d} : P \times P \to \mathbb{R}$. A simplex $\sigma = \{q_1, \ldots, q_k\}$ with $q_i \in Q$ is weakly witnessed by $x \in P \setminus Q$ if $\mathsf{d}(q, x) \leq \mathsf{d}(p, x)$ for every $q \in \{q_1, \ldots, q_k\}$ and $p \in Q \setminus \{q_1, \ldots, q_k\}$.

We now define the witness complex using the notion of weak witnesses.

Definition 14 (Witness complex)**.** Let P, Q be a finite sets as in Defintion 13. The witness complex $\mathcal{W}(Q, P)$ is defined as the collection of all simplices whose all faces are weakly witnessed by a point in $P \setminus Q$.

Observe that a simplex which is weakly witnessed may not have all its faces weakly witnessed (Exercise 5). This is why the definition above forces the condition to have a simplicial complex.

When P is $\mathbb{R}^d$ and Q is a finite subset of it, we have the notion of strong witness.

Definition 15 (Strong witness)**.** Let $Q \subset \mathbb{R}^d$ be a finite set. A simplex $\sigma = \{q_1, \ldots, q_d\}$ with $q_i \in Q$ is strongly witnessed by $x \in \mathbb{R}^d$ if $\mathsf{d}(q, x) \leq \mathsf{d}(p, x)$ for every $q \in \{q_1, \ldots, q_d\}$ and $p \in Q \setminus \{q_1, \ldots, q_d\}$ and additionally, $\mathsf{d}(q_1, x) = \cdots = \mathsf{d}(q_d, x)$.

When $Q \subset \mathbb{R}^d$ as in the above definition, the following fact holds [8].

Proposition 3. *A simplex σ is strongly witnessed if and only if every face $\tau \leq \sigma$ is weakly witnessed.*

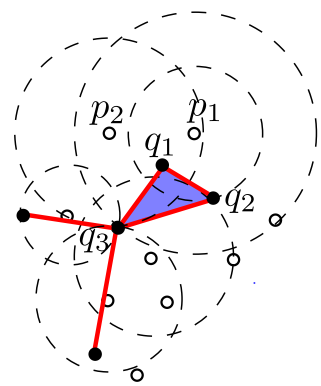

Figure 2.6: A witness complex constructed out of the points in Figure 2.4 where landmarks are the black dots and the witness points are the hollow dots. The witnesses for five red edges and the blue triangle are the centers of the six circles; e.g., the triangle $q_1q_2q_3$ and the edge q_1q_3 are weakly witnessed by the points p_1 and p_2 respectively.

Furthermore, when $Q \subset \mathbb{R}^d$, we have some connections of the witness complex to the Dealunay complex. By definition, we know the following:

Fact 5. *Let Q be a finite subset of $\mathbb{R}^d$. Then a simplex σ is in the Delaunay triangulation* Del Q *if and only if σ is strongly witnessed by points in $\mathbb{R}^d$.*

By combining the above fact and the observation that every simplex in a witness complex is strongly witnessed, we have the following result which was observed by de Silva [8].

Proposition 4. *If P is a finite subset of $\mathbb{R}^d$ and $Q \subseteq P$, then $\mathcal{W}(Q, P) \subseteq \mathrm{Del}\, Q$.*

One important implication of the above observation is that the witness complexes for point samples in an Euclidean space are embedded in that space.

The concept of the witness complex has a parallel in the concept of the restricted Delaunay complex. When the set P in Proposition 4 is not necessarily a finite subset, but only a subset X of $\mathbb{R}^d$, and Q is finite, we can relate $\mathcal{W}(Q, P)$ to the *restricted Delaunay complex* $\mathrm{Del}|_X\, Q$ defined as the collection of Delaunay simplices in Del Q whose Voronoi duals have non-empty intersection with X.

Proposition 5.

1. $\mathcal{W}(Q, \mathbb{R}^d) = \mathrm{Del}|_{\mathbb{R}^d}\, Q := \mathrm{Del}\, Q$ *[8].*
2. $\mathcal{W}(Q, M) = \mathrm{Del}|_M\, Q$ *if $M \subseteq \mathbb{R}^d$ is a smooth* 1- *or* 2-*manifold [2].*
3. $\mathcal{W}(Q, P) = \mathrm{Del}|_M\, Q$ *where P and Q are sufficiently dense sample of a* 1-*manifold M in $\mathbb{R}^2$ and the result does not extend to other cases of submanifolds embedded in Euclidean spaces [17].*

2.3.3 Graph induced complex

The witness complex does not capture the topology of a manifold even if the input sample is dense except for smooth curves in the plane. One can modify them with extra structures such as putting weights on the points and changing the metric to weighted distances to tackle this problem as shown in [3]. But, this becomes clumsy in terms of the 'practicality' of a solution. We study another complex called *graph induced complex*(GIC) [11] which also uses subsampling, but is more powerful in capturing topology and in some case geometry. The adavantage of the GIC over the witness complex is that GIC is not necessarily a subcomplex of the Delaunay complex and hence contains few more simplices which aid topology inference. But, for the same reason, it may not embed in the Euclidean space where its input vertices lie.

In the following definition, 2^Q denotes the set of all subsets of a set Q. For the nearest point map ν in the definition, we have to consider this power set because the nearest point may not be unique.

Definition 16. Let (P, d) be a metric space where P is a finite set and $G(P)$ be a graph with vertices in P. Let $Q \subseteq P$ and let $\nu : P \to 2^Q$ be the map given by $\nu(p) = \operatorname{argmin} \mathsf{d}(p, Q)$. The *graph induced complex* (GIC) $\mathcal{G}(G(P), Q, \mathsf{d})$ is the simplicial complex containing a k-simplex $\sigma = \{q_1, \ldots, q_{k+1}\}$, $q_i \in Q$, if and only if there exists a $(k+1)$-clique $\{p_1, \ldots, p_{k+1}\} \subseteq P$ so that $q_i \in \nu(p_i)$ for each $i \in \{1, 2, \ldots, k+1\}$. To see that it is indeed a simplicial complex, observe that a subset of a clique is also a clique.

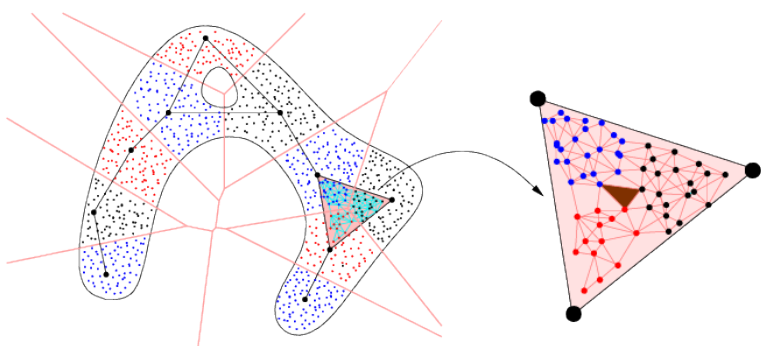

Figure 2.7: A graph induced complex shown with bold vertices, edges, and a shaded triangle on left. The input graph within the shaded triangle is shown on right. The 3-clique with three different colors (shown inside the shaded triangle on the right) cause the shaded triangle in the left to be in the graph induced complex.

Input graph $G(P)$. The input point set P can be a finite sample of a subset X of an Euclidean space, such as a manifold or a compact subset. In this case, we may consider the input graph $G(P)$ to be the neighborhood graph $G^\alpha(P) := (P, E)$ where there is an edge $\{p, q\} \in E$ if and only if $\mathsf{d}(p, q) \leq \alpha$. The intuion is that if P is a sufficiently dense sample of X, then $G^\alpha(P)$ captures the local neighborhoods of the points in X. Figure 2.7 shows a graph induced complex for a point data in the plane with a neighborhood graph where d is the Euclidean metric. To emphasize the dependence on α we use the notation $\mathcal{G}^\alpha(P, Q, \mathsf{d}) := \mathcal{G}(G^\alpha(P), Q, \mathsf{d})$.

Subsample Q. Of course, the ability of capturing the topology of the sampled space after subsampling with Q depends on the quality of Q. We quantify this quality with a parameter $\delta > 0$.

Definition 17. A subset $Q \subseteq P$ is called a *δ-sample* of a metric space (P, d), if the following condition holds:

- $\forall p \in P$, there exists a $q \in Q$, so that $\mathsf{d}(p, q) \leq \delta$.

Q is called *δ-sparse* if the following condition holds:

- $\forall (q, r) \in Q \times Q$ with $q \neq r$, $\mathsf{d}(q, r) \geq \delta$.

The first condition ensures Q to be a good sample of P with respect to the parameter δ and the second condition enforces that the points in Q cannot be too close relative to the distance δ.

Metric d. The metric d assumed in the metric space (P, d) will be of two types in our discussion below; (i) the Euclidean metric denoted d_E, (ii) the graph metric d_G derived from the the input graph $G(P)$ where $\mathsf{d}_G(p, q)$ is the shortest path distance between p and q in the graph $G(P)$ assuming its edges have non-negative weights such as their Euclidean lengths.

Constructing a GIC. One may wonder how to efficiently construct the graph induced complexes in practice. Experiments show that the following procedure runs quite efficiently in practice. It takes advantage of computing nearest neighbors within a range and, more importantly, computing cliques only in a sparsified graph.

Let the ball $B(q, \delta)$ in metric d be called the δ-cover for the point q. A graph induced complex $\mathcal{G}^\alpha(P, Q, \mathsf{d})$ where Q is a δ-sparse δ-sample can be built easily by identifying δ-covers with a rather standard greedy (farthest point) iterative algorithm. Let $Q_i = \{q_1, \ldots, q_i\}$ be the point set sampled so far from P. We maintain the invariants (i) Q_i is δ-sparse and (ii) every point $p \in P$ that are in the union of δ-covers $\bigcup_{q \in Q_i} B(q, \delta)$ have their closest point $\nu(p) = \mathrm{argmin}_{q \in Q_i} \mathsf{d}(p, q)$ in Q_i identified. To augment Q_i to $Q_{i+1} = Q_i \cup \{q_{i+1}\}$, we choose a point $q_{i+1} \in P$ that is outside the δ-covers $\bigcup_{q \in Q_i} B(q, \delta)$. Certainly, q_{i+1} is at least δ units away from all points in Q_i thus satisfying the first invariant. For the second invariant, we check every point p in the δ-cover of q_{i+1} and update $\nu(p)$ to be q_{i+1} if its distance to q_{i+1} is smaller than the distance $\mathsf{d}(p, \nu(p))$. At the end, we obtain a sample $Q \subseteq P$ whose δ-covers cover the entire point set P and thus is a δ-sample of (P, d) which is also δ-sparse due to the invariants maintained. Next, we construct the simplices of $\mathcal{G}^\alpha(P, Q, \mathsf{d})$. This needs identifying cliques in $G^\alpha(P)$ that have vertices with different closest points in Q. We delete every edge pp' from $G^\alpha(P)$ where $\nu(p) = \nu(p')$. Then, we determine every clique $\{p_1, \ldots p_k\}$ in the remaining sparsified graph and include the simplex $\{\nu(p_1), \ldots, \nu(p_k)\}$ in $\mathcal{G}^\alpha(P, Q, \mathsf{d})$. The main saving here is that many cliques of the original graph are removed before it is processed for clique computation.

2.4 Notes and Exercises

Simplicial complexes is a fundamental structure in algebraic topology. A good source for the subject is Munkres [21].

The concept of nerve is credited to Alesandroff [1]. The nerve theorem has different versions. It holds for open covers for topological spaces with some mild conditions [23]. Borsuk proved it for closed covers again with some conditions on the space and covers [4]. The assumptions of both are satisfied by metric spaces and finite covers with which we state the theorem in section 2.2. A version of the theorem is also credited to Leray [20].

Čech and Vietoris-Rips complexes have turned out to be a very effective data structure in topological data analysis. Čech complexes were introduced to define Čech homology. Leonid Vietoris [22] introduced Vietoris complex for extending the homology theory from simplicial complexes to metric spaces. Later, Eliyahu Rips used it in hyperbolic group theory [16]. Jean-Claude Hausmann named it as Vietoris-Rips complex and showed that it is homotopoy equivalent to a compact Riemannian manifold when the vertex set spans all points of the manifold and parameter to build it is sufficiently small [18]. This result was further improved by Latschev [19] who showed that the homotopy equivalence holds even when the vertex set is finite.

Delaunay complex is a very well known and useful data structure for various geometric applications in two and three dimensions. They enjoy various optimal properties. For example, for a given point set $P \subset \mathbb{R}^2$, among all simplicial complexes linearly embedded in $\mathbb{R}^2$ with vertex set P, the Delaunay complex maximizes the minimum angle over all triangles as stated in Fact 4. Many such properties and algorithms for computing Delaunay complexes are described in books by Edelsbrunner [12] and Cheng et al. [6]. Alpha complex was proposed in [14] and further developed in [15]. The first author of this can attest to the historic fact that the development of the persistence algorithm was motivated by the study of alpha complexes and their Betti numbers. The book by Edelsbrunner and Harer [13] confirms this. Witness complexes are proposed by de Silva and Carlsson [9] in an attempt to build a sparser complex out of a dense point sample. The graph induced complex is also another such construction proposed by Dey, Fang and Wang [10].

Exercises

1. Suppose we have a collection of set $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ where there exists an element $U \in \mathcal{U}$ that contains all other elements in $\mathcal{U}$. Show that the nerve complex $N(\mathcal{U})$ is contractible to a point.

2. Given a parameter α and a set of points $P \in \mathbb{R}^d$, show that the alpha complex $\mathrm{Del}_\alpha(P)$ is contained in the intersection of Delauney complex and Čech complex at scale α; that is, $\mathrm{Del}_\alpha(P) \subseteq \mathrm{Del}(P) \cap \mathbb{C}^\alpha(P)$.

3. Let K be a triangulation of a surface without boundary that has genus g. Prove that $\beta_1(K) = 2g$.

4. We state the nerve theorem (Theorem 1) for covers where either all cover elements are closed or all cover elements are open. Show that the theorem does not hold if we mix open and closed elements in the cover.

5. Give an example where a simplex which is weakly witnessed may not have all its faces weakly witnessed. Show that (i) $\mathcal{W}(Q, P') \subseteq \mathcal{W}(Q, P)$ for $P' \subseteq P$, (ii) $\mathcal{W}(Q', P)$ may not be a subcomplex of $\mathcal{W}(Q, P)$ where $Q' \subseteq Q$.

6. Consider Definition 16 for Graph induced complex. Let $\mathbb{VR}(G)$ be the clique complex given by the input graph $G(P)$. Assume that the map $\nu : P \to 2^Q$ sends every point to a singleton under input metric d. Then, $\nu : P \to \nu(P)$ is a well defined vertex map. Prove that the vertex map $\nu : P \to Q$ extends to a simplicial map $\bar{\nu} : \mathbb{VR}(G) \to \mathcal{G}(G(P), Q, \mathsf{d})$. Also, show that every simplicial complex $K(Q)$ with the vertex set Q for which $\bar{\nu} : \mathbb{VR}(G) \to K(Q)$ becomes simplicial must contain $\mathcal{G}(G(P), Q, \mathsf{d})$.

Bibliography

[1] Paul Aleksandroff. Über den allgemeinen dimensionsbegriff und seine beziehungen zur elementaren geometrischen anschauung. *Mathematische Annalen*, 98:617–635, 1928.

[2] Dominique Attali, Herbert Edelsbrunner, and Yuriy Mileyko. Weak witnesses for delaunay triangulations of submanifolds. In *Proc. ACM Sympos. on Solid and Physical Modeling*, pages 143–150, 2007.

[3] Jean-Daniel Boissonnat, Leonidas J. Guibas, and Steve Y. Oudot. Manifold reconstruction in arbitrary dimensions using witness complexes. In *Proc. 23rd ACM Sympos. Comput. Geom.*, pages 194–203, 2007.

[4] Karol Borsuk. On the imbedding of systems of compacta in simplicial complexes. *Fundamenta Mathematicae*, 35:217–234, 1948.

[5] Bernard Chazelle. An optimal convex hull algorithm in any fixed dimension. *Discret. Comput. Geom.*, 10:377–409, 1993.

[6] Siu-Wing Cheng, Tamal K. Dey, and Edgar A. Ramos. Manifold reconstruction from point samples. In *SODA '05: Proc. 16th Ann. ACM-SIAM Sympos. Discrete Algorithms*, pages 1018–1027, 2005.

[7] Siu-Wing Cheng, Tamal K. Dey, and Jonathan R. Shewchuk. *Delaunay Mesh Generation*. CRC Press, 2012.

[8] Vin de Silva. A weak definition of delaunay triangulation, 2003. arXiv:cs.CG/0310031.

[9] Vin de Silva and Gunnar Carlsson. Topological estimation using witness complexes. In *Proc. Sympos. Point-Based Graphics.*, 2004.

[10] Tamal K. Dey, Fengtao Fan, and Yusu Wang. Computing topological persistence for simplicial maps. *CoRR*, abs/1208.5018, 2012.

[11] Tamal K. Dey, Fengtao Fan, and Yusu Wang. Graph induced complex for point data. In *Proc. 29th. Annu. Sympos. Comput. Geom. (SoCG)*, pages 107–116, 2013.

[12] Herbert Edelsbrunner. *Geometry and Topology for Mesh Generation*, volume 7 of *Cambridge monographs on applied and computational mathematics*. Cambridge University Press, 2001.

[13] Herbert Edelsbrunner and John Harer. *Computational Topology: An Introduction*. Applied Mathematics. American Mathematical Society, 2010.

[14] Herbert Edelsbrunner, David G. Kirkpatrick, and Raimund Seidel. On the shape of a set of points in the plane. *IEEE Trans. Inf. Theory*, 29(4):551–558, 1983.

[15] Herbert Edelsbrunner and Ernst P. Mücke. Three-dimensional alpha shapes. *ACM Trans. Graph.*, 13(1):43–72, 1994.

[16] Mikhail Gromov. Hyperbolic groups. In S.M. Gersten, editor, *Essays in Group Theory*, volume 8, pages 75–263. Mathematical Sciences Research Institute Publications, Springer, 1987.

[17] Leonidas J. Guibas and Steve Y. Oudot. Reconstructing using witness complexes. *Discrete. Comput. Geom.*, 30:325–356, 2008.

[18] Jean-Claude Hausmann. On the vietoris–rips complexes and a cohomology theory for metric spaces. *Annals Math. Studies*, 138:175–188, 1995.

[19] Janko Latschev. Vietoris-rips complexes of metric spaces near a closed riemannian manifold. *Archiv der Mathematik*, 77(6):522–528, 2001.

[20] Jean Leray. Sur la forme des espaces topologiques et sur les points fixes des représentations. *J. Math. Pure Appl.*, 24:95–167, 1945.

[21] James R. Munkres. *Elements of Algebraic Topology*. Addison–Wesley Publishing Company, Menlo Park, 1984.

[22] Leopold Vietoris. Über den höheren zusammenhang kompakter räume und eine klasse von zusammenhangstreuen abbildungen. *Mathematische Annalen*, 97:454–472, 1927.

[23] André Weil. Sur les théoréms de de rham. *Commentarii Mathematici Helvetici*, 26:119–145, 1952.