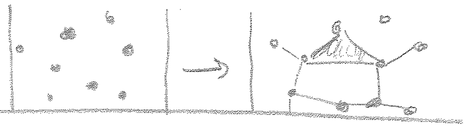
CAT 1
complexes



Goal 0. TOPOLOGICAL DATA ANALYSIS is the study of DATA (usually, finite metric spaces) using invariants from ALGEBRAIC TOPOLOGY (usually, simplicial homology)

We will take few min to talk about finite metric spaces, and then most of the time talking about "simplicial complexes"

Def 1 A FINITE METRIC SPACE (M, d) is a finite set M along with a function $d : M \times M \longrightarrow \mathbb{R}_{\geq 0}$ that satisfies three axioms.

[IDENTITY] $d(x,y) = 0 \iff x = y \quad \forall x, y \in M$

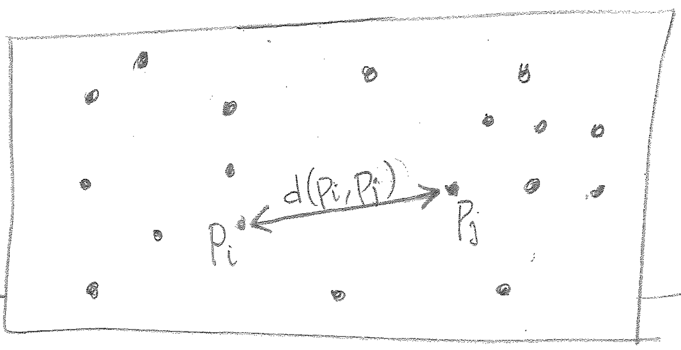
[SYMMETRY] $d(x,y) = d(y,x) \quad \forall x, y \in M$

[Δ-INEQUALITY] $d(x,y) + d(y,z) \geq d(x,z) \quad \forall x, y, z \in M.$

Ex. 2. Nice examples come from choosing

M = {finitely many points $p_1, \dots, p_k \in \mathbb{R}^n$}

d = Euclidean distance.

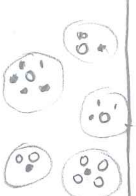


Def 3.

a) A map $f : (M,d) \longrightarrow (M',d')$ of metric spaces is called an ISOMETRIC EMBEDDING if

$$d(x,y) = d'(f(x), f(y)) \quad \forall x, y \in M.$$

b) The HAUSDORFF DISTANCE between two subsets M_0, M_1



of the same metric space (M, d) is given by

$$\boxed{d^M_{Haus}(M_0, M_1) = \underline{\inf}\, \{\varepsilon > 0 \mid M_0^{+\varepsilon} \supseteq M_1 \text{ and } M_1^{+\varepsilon} \supseteq M_0\}}$$

where $M_0^{+\varepsilon} = \{x \in M \mid \exists y \in M_0 \text{ with } d(x,y) \leq \varepsilon\}$

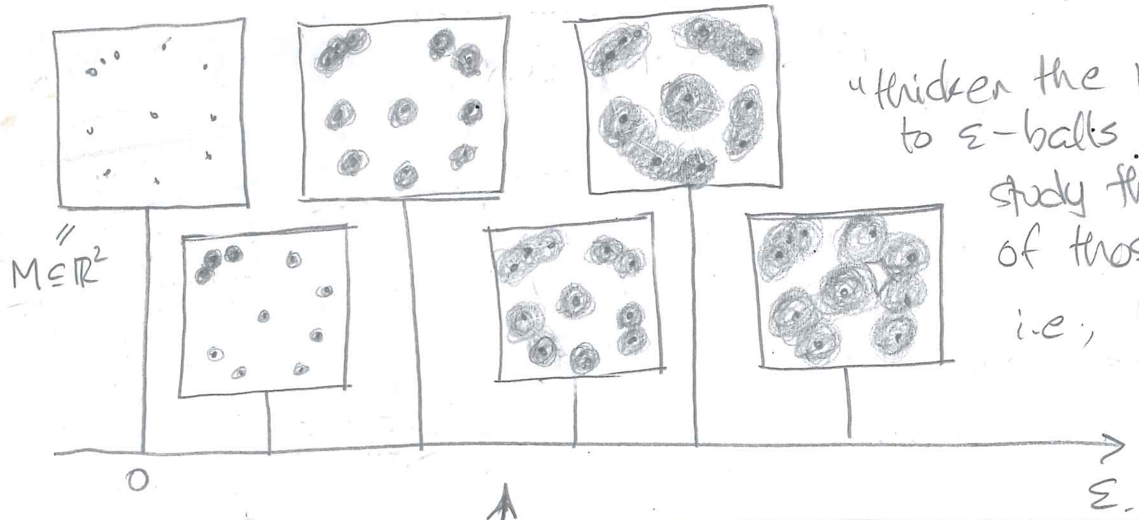
c) The GROMOV-HAUSDORFF DISTANCE between two metric spaces (M,d) and (M',d') is given by

$$\boxed{\inf_{p,p'} \{d^Z_{Haus}(p(M), p'(M'))\}}$$

where $p: M \to Z$ and $p': M' \to Z$ are isometric embeddings into a third metric space Z. [The inf is over ALL choices of Z and p and p', so it is Basically IMPOSSIBLE to compute.] ← (we can still get bounds, though..)

SIMPLICIAL COMPLEXES

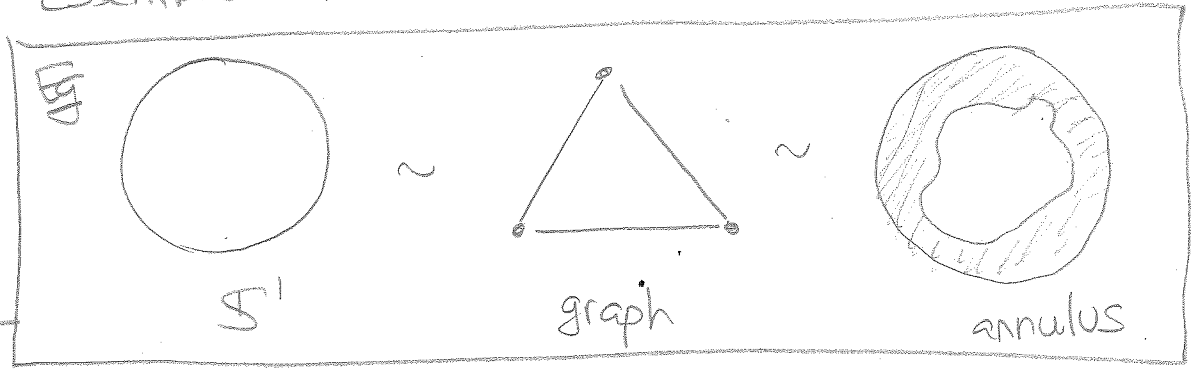
Idea 4 We'd like to study a given finite (M,d) at various scales: eg for points in the plane, we want to see the evolution of geometric features of $M^{+\varepsilon}$ for $\varepsilon > 0$



"thicken the points to ε-balls and study the union of those balls, i.e., $M^{+\varepsilon}$.

Problem 5 The trouble with this strategy that "unions of balls", even in Euclidean space, are often difficult to work with algorithmically. (If I give you M as a list of points $\{p_i\}$, how will you test if $M^{+\varepsilon}$ is connected?)

Among the earliest ideas in computational algebraic topology is to approximate "continuous spaces", eg manifolds, by "discrete representatives" that preserve the essential structure.



Def 6.

a) An ABSTRACT SIMPLICIAL COMPLEX K is a collection of nonempty subsets of a set K_0 of VERTICES which satisfy two conditions:

(i) $\forall v \in K_0$, the set $\{v\}$ lies in K

(ii) $\forall \sigma \in K$ and $\tau \subseteq \sigma$, we have $\tau \in K$ also.

b) Each $\sigma \in K$ is called a SIMPLEX, and each $\tau \subseteq \sigma$ is called a FACE (often written $\tau \leq \sigma$).

The DIMENSION of σ is [# - cardinality]

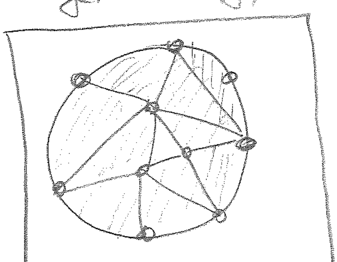
$$\boxed{\dim \sigma = \# \sigma - 1}$$

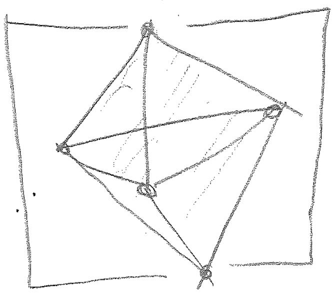
(So the vertices correspond bijectively with 0-dim simplices). Write K_d to indicate the set of d-simplices of K. Now $\boxed{\dim K = \max_{\sigma \in K} \dim \sigma}$

Eg 7

a) Any graph is a simplicial complex of dimension ≤ 1.

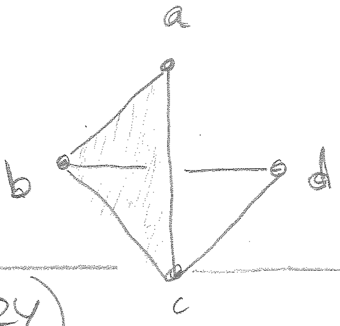
b) More generally, a triangulated disk or sphere:





(These are both 2-dimensional)

c) $K_0 = \{a, b, c, d\}$
$K_1 = \{ab, ac, bc, bd, cd\}$
$K_2 = \{abc\}$



(FROM COMBINATORICS TO GEOMETRY)

Def 8 Given points $\{p_0, \ldots, p_d\} \subseteq \mathbb{R}^n$, the GEOMETRIC SIMPLEX spanned by them is the subset

$$\left\{ \sum_0^d t_i p_i \;\middle|\; t_i \geq 0 \text{ and } \sum_i t_i = 1 \right\} \subseteq \mathbb{R}^n$$

"convex hull"

We call the p_i "AFFINELY INDEPENDENT" if the set $\{p_1 - p_0, \ldots, p_i - p_0\}$ of vectors is linearly independent.

Def 9 Let $\varphi : K_0 \to \mathbb{R}^n$ be a map whose image is affinely independent. Then the GEOMETRIC REALIZATION of K (with respect to φ) is the union

$$\boxed{|K|_\varphi = \bigcup_{\sigma \in K} |\sigma|_\varphi,}$$

where for each $\sigma = (v_0, \ldots, v_i) \in K$ the space $|\sigma|_\varphi$ is the geometric simplex spanned by $\{\varphi(v_0), \ldots, \varphi(v_i)\}$ in $\mathbb{R}^n$

Thm 10 The homeomorphism type of $|K|$ is independent of φ (provided $\varphi(K_0)$ is affinely independent). And moreover, $|K|$ can always be embedded in $\mathbb{R}^{\#K_0}$.

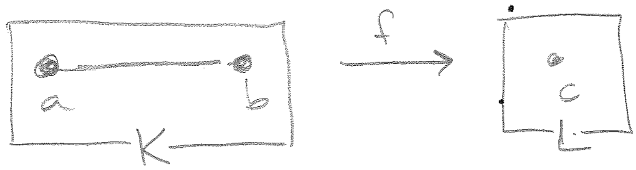
Pf If $\varphi, \varphi' : K_0 \to \mathbb{R}^n$ both have affinely indep images, we can find an invertible matrix A sending each $\varphi(v_i) - \varphi(v_0)$ to $\varphi'(v_i) - \varphi'(v_0)$. Now $A(|K|_\varphi) = |K|_{\varphi'}$. And for the second part, enumerate $K_0 = \{v_1, \ldots, v_n\}$ and let φ send v_i to the i^{th} basis vector of $\mathbb{R}^n$. This is affinely independent

Def 11 A SIMPLICIAL MAP $K \to L$ is an assignment $f: K_0 \to L_0$ of vertices to vertices so that every simplex $\sigma \in K$ is sent to a simplex $f(\sigma)$ of L; i.e.,

$$\sigma = (v_0, \ldots, v_d) \longmapsto f(\sigma) = (f(v_0), \ldots, f(v_d))$$

But note that f need NOT be injective, so in general $\dim f(\sigma) \leq \dim \sigma$. Eg:

Warning about graph morphisms



$f(a) = c = f(b)$
So $f(ab) = c$
has dropped dimension

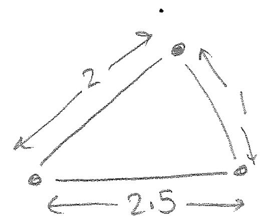
Prop 12 Every simplicial map $f: K \to L$ induces a continuous map $|f|: |K| \to |L|$ of geometric realizations. [Eg: $K \subseteq L$ subcomplex]

FROM DATA TO SIMPLICIAL COMPLEXES

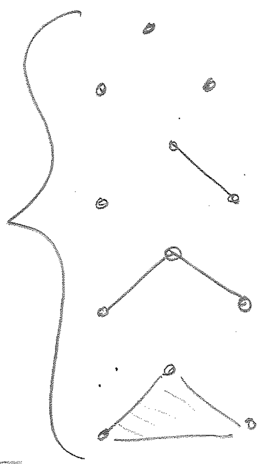
Let (M, d) be a finite metric space and $\varepsilon > 0$ a real number.

Def 13 The VIETORIS-RIPS complex of M at scale ε is the simplicial complex whose k-simplices are all subsets $\{p_0, \ldots, p_k\} \subseteq M$ with $d(p_i, p_j) \leq \varepsilon \ \forall i, j$.

So if $M =$



then $VR_t(M) =$



$t \in [0,1)$

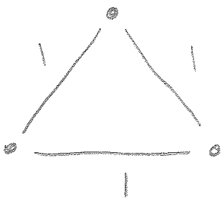
$t \in [1,2)$

$t \in [2, 2.5)$

$t \geq 2.5$

Def 14 Assume (M,d) is isometrically embedded in a larger metric space (Z,d). The <u>Čech complex</u> of M at radius ε is the simplicial complex whose k-simplices are all $\{p_0, \dots, p_k\} \subseteq M$ so that $\exists\, y \in Z$ with $d(y, p_i) \leq \varepsilon \ \forall i$.

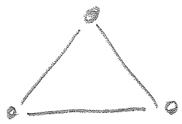
Note! y need not be in M

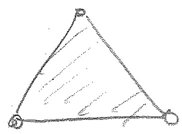
Eg: $M =$ 

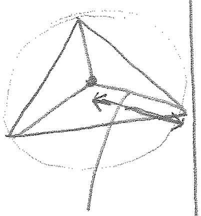
$\subseteq \mathbb{R}^2$

$\check{C}_\varepsilon(M) =$

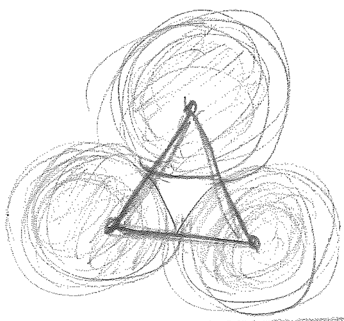

$\varepsilon \in [0, 1/2)$


$\varepsilon \in [1/2, 1/\sqrt{3})$


$\varepsilon \geq 1/\sqrt{3}$


$1/\sqrt{3}$

Note 15 The VR complex is easier to compute, but much larger in general. The Čech complex is harder to compute, but (i) smaller and (ii) more faithful to the underlying union of balls!



The "hole" is visible in $\check{C}_\varepsilon(M)$, <u>not</u> in $VR_\varepsilon(M)$