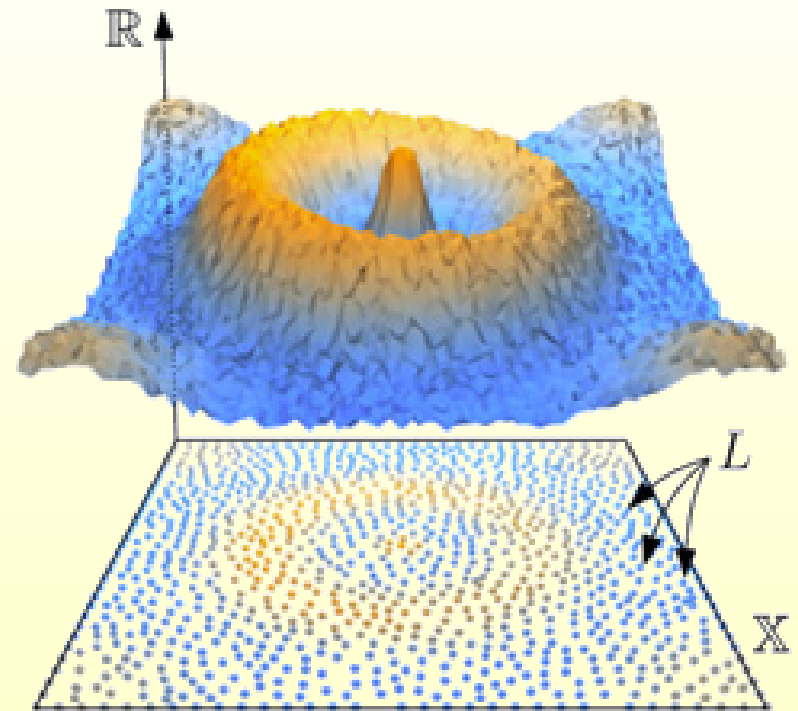


CS233: Geometric and Topological Data Analysis

Topological Data Analysis (TDA),
Computational Topology,
Homology

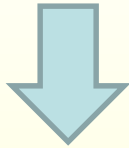
23 April 2018

Slides ack: Afra Zomorodian, Ryan Lewis,
Gunnar Carlsson



Topology and Data Analysis

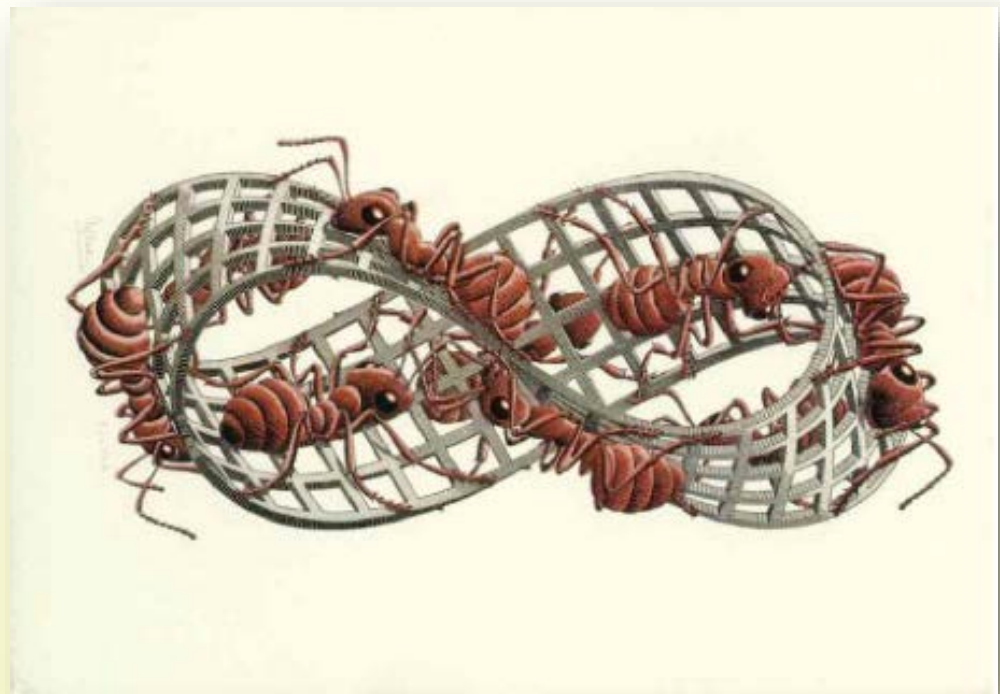
Topology — Computational
Topology



Homology — Persistent
Homology



Topological Data Analysis
Applications



Computational Topology and Data Analysis

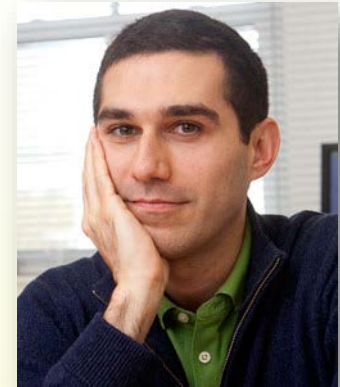


Herbert Edelsbrunner



Gunnar Carlsson

1990 -- 2018



Afra Zomorodian

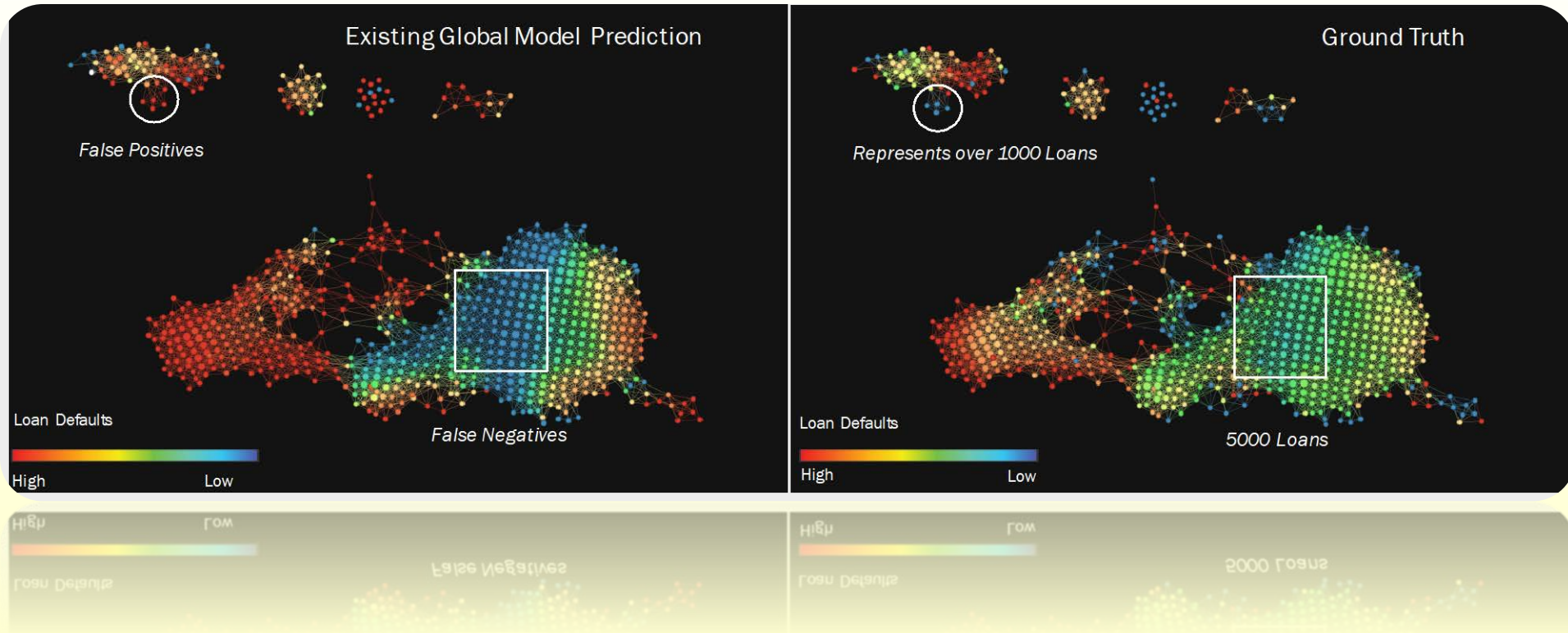


Frederic Chazal



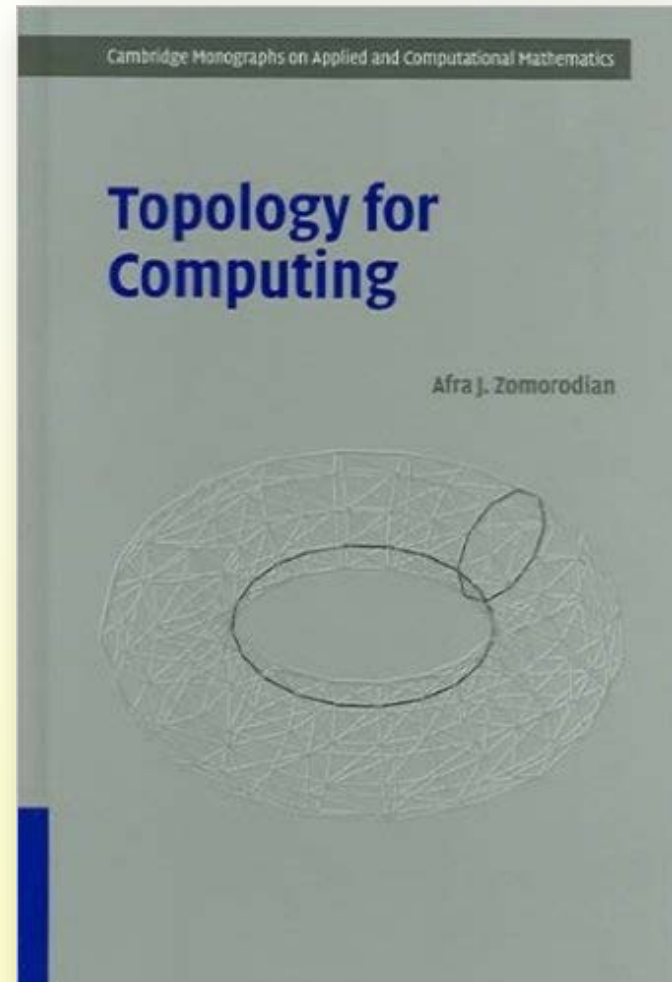
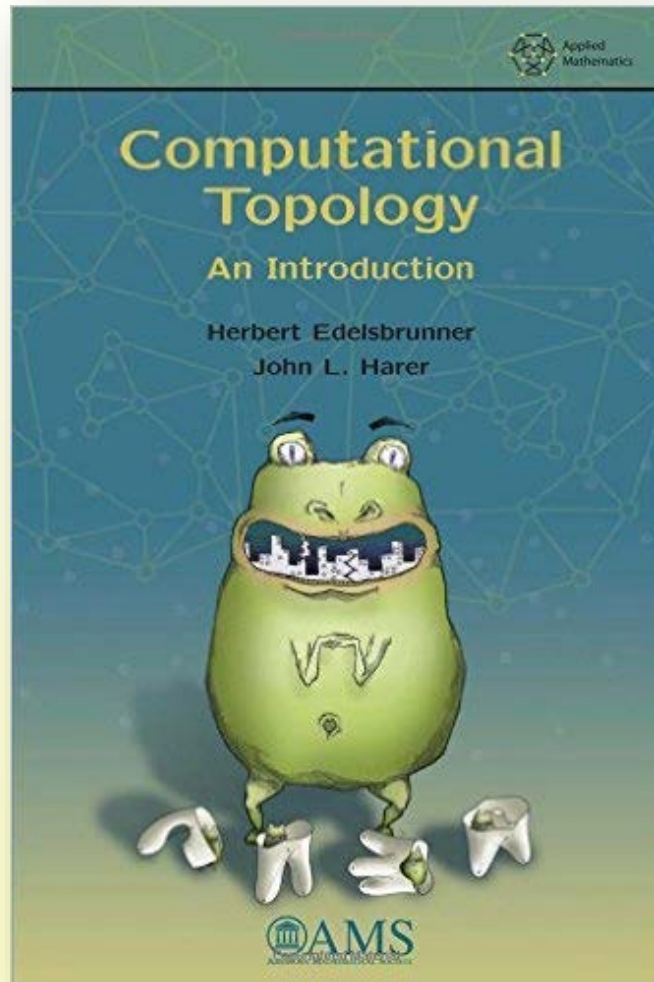
Robert Ghrist

Some Companies



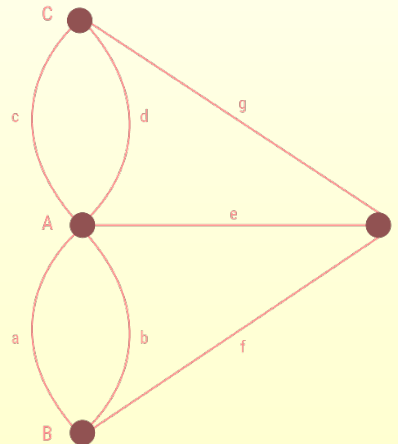
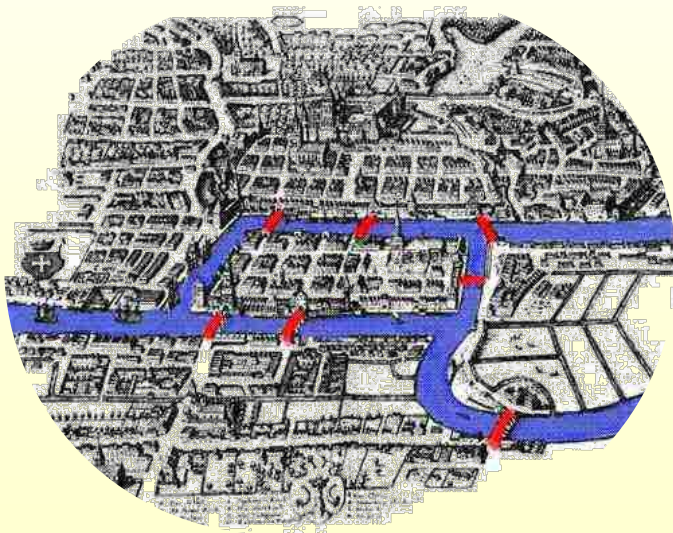
AYASDI

Some Textbooks



Topology Studies Connectivity

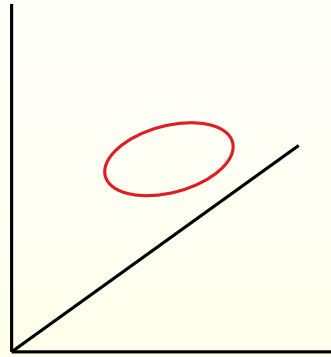
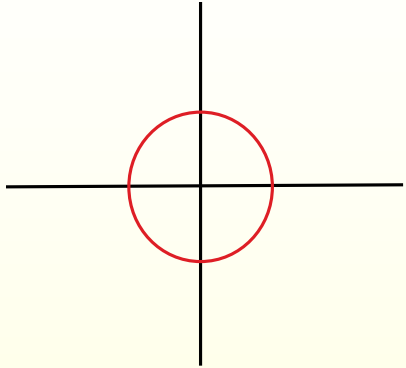
... and obstructions to connectivity



The bridges of
Königsberg

Leonhard Euler

Invariances

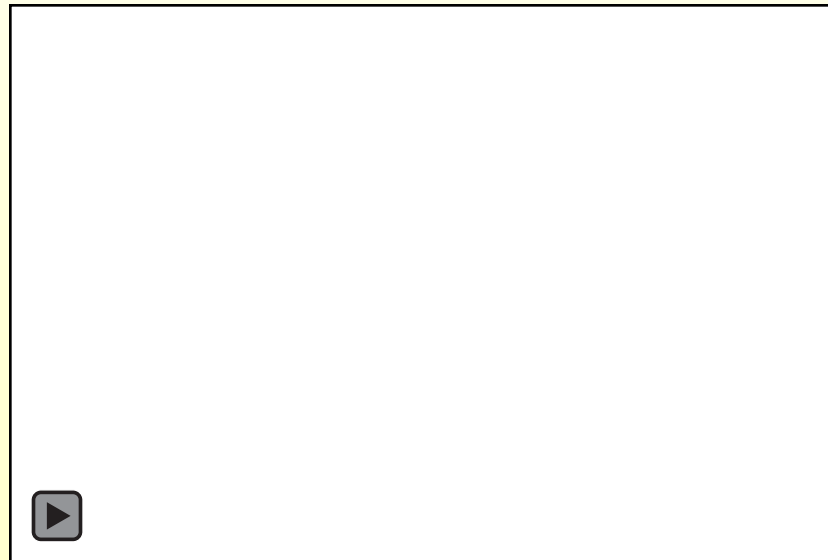


A

A

Coordinate-free

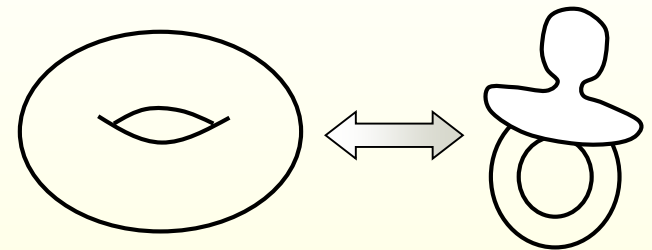
Deformation-invariant



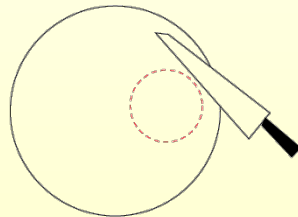
Topology does not take distances too seriously

2-Manifolds: Ordinary Surfaces

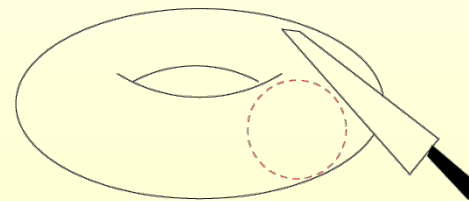
- ◆ We are allowed to stretch and shrink
 - ◆ Homeomorphism: 1-1, onto, bi-continuous



- ◆ But we care about cutting, puncturing, stitching, gluing ...



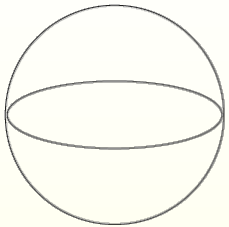
(a) No matter where we cut the sphere, we get two pieces



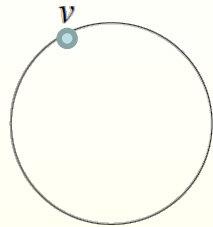
(b) If we're careful, we can cut the torus and still leave it in one piece.

- ◆ Note: connectivity information is indexed by dimension

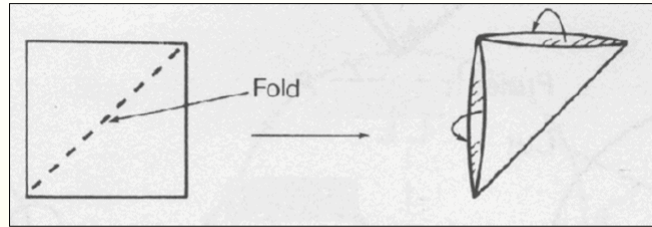
2-Manifold Zoo



(a) $\{x \in \mathbb{R}^3 \mid |x| = 1\}$

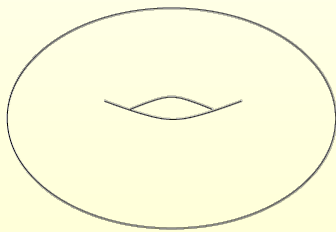


(b) Identify boundary to v

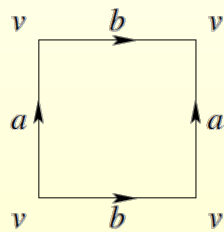


(c) Instructions for a flat sphere

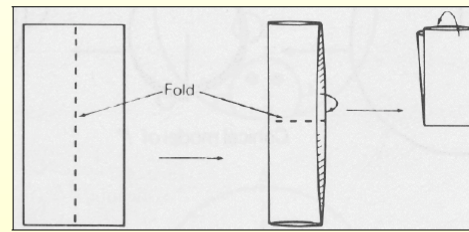
Sphere



(a) Donut surface



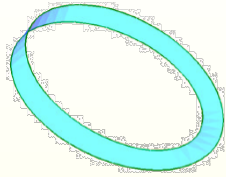
(b) Diagram



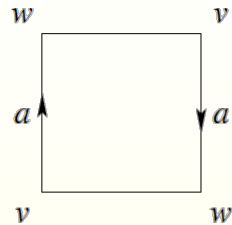
(c) Instructions for a flat torus

Torus

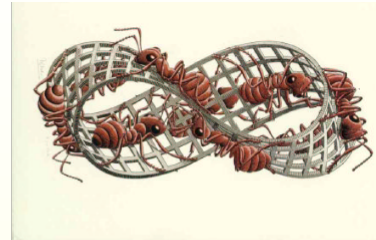
More Exotic Animals



(a) Embedded

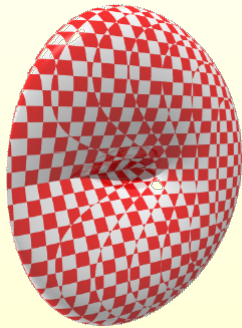


(b) Diagram

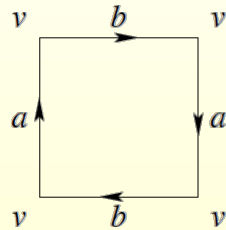


(c) Escher's *Möbius Strip II* (on its side)

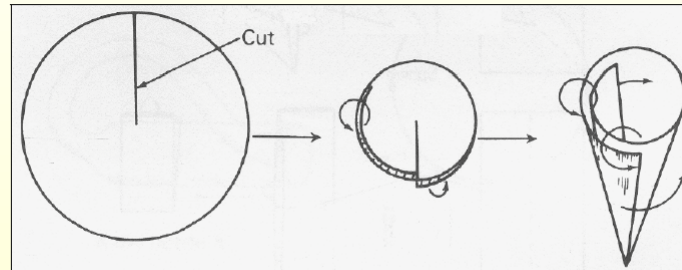
Möbius strip



Cross cap + disk

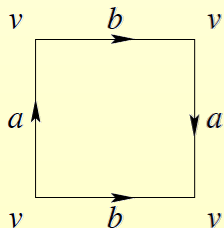


(a) Diagram

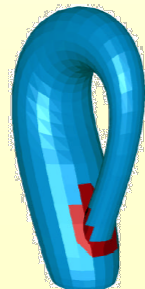


(b) Instructions for a flat \mathbb{RP}^2

Projective plane



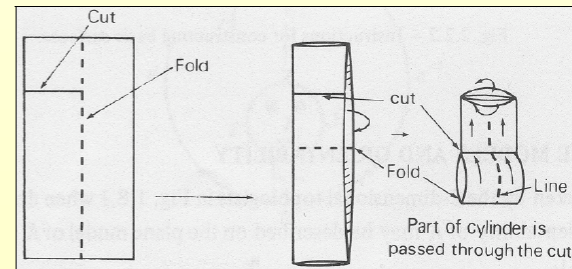
(a) Diagram



(b) An immersion



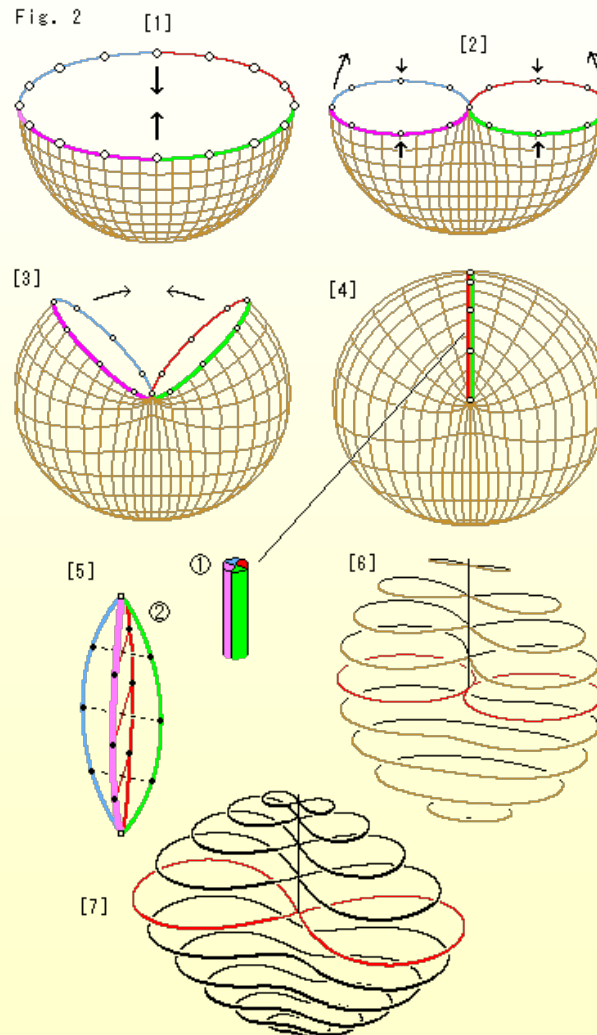
(c) Cut in half (a Möbius strip)



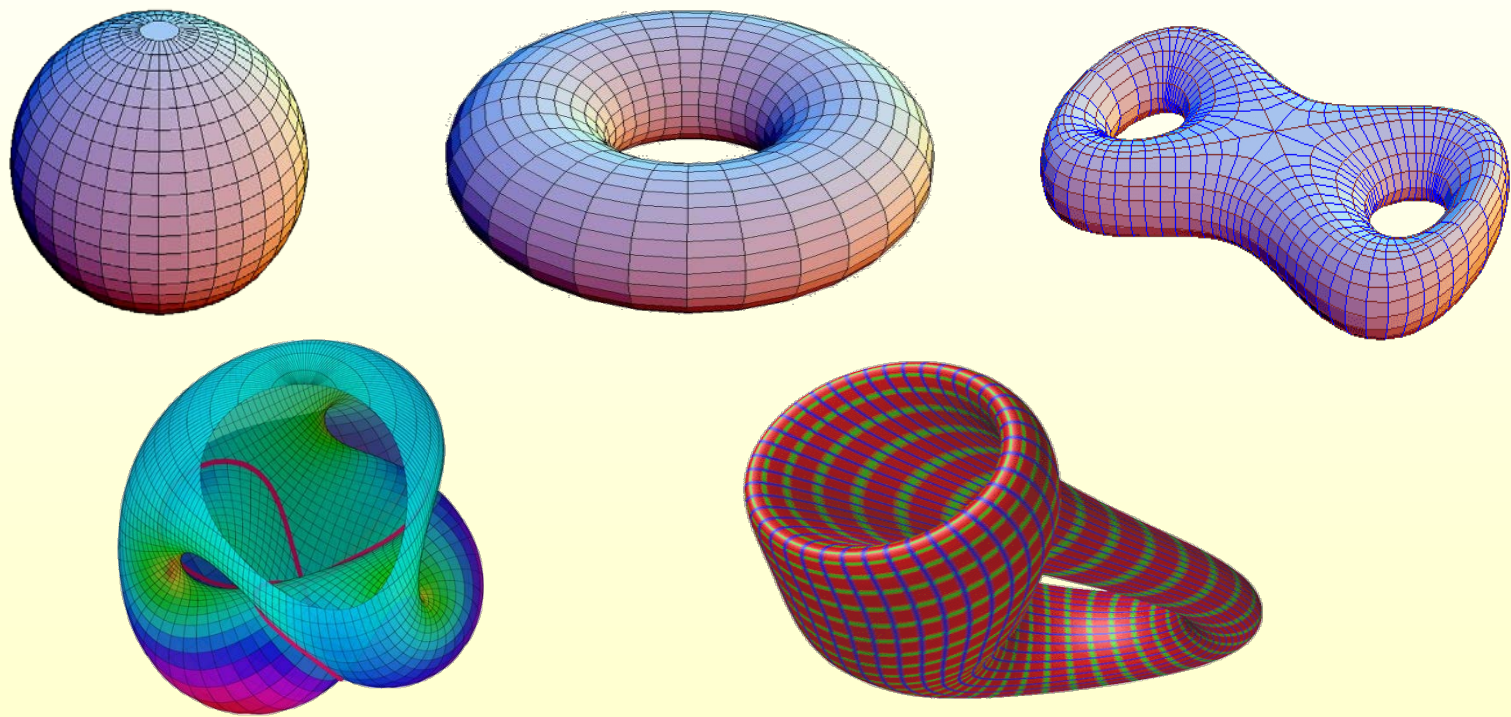
(d) Instructions for a flat \mathbb{K}^2

Klein bottle

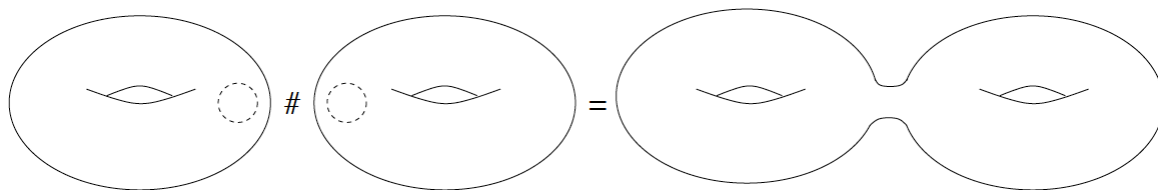
Projective Plane



Topology Aims to Classify



Connected Sums

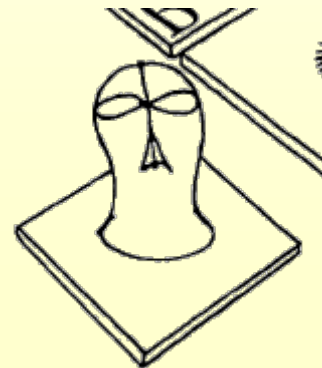


- ◆ **Classification Theorem of 2-Manifolds (1860):** Every closed connected compact surface is a connected sum of a sphere with a number of tori and projective planes (sphere + handles + cross cups)



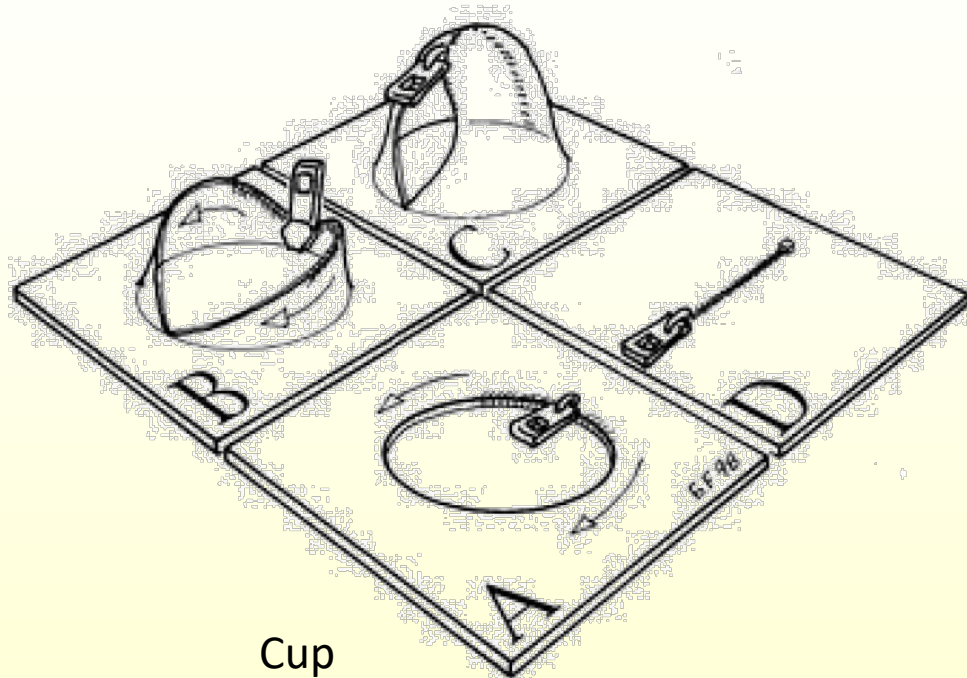
Handle

Klein bottle = sphere + 2 cross cups

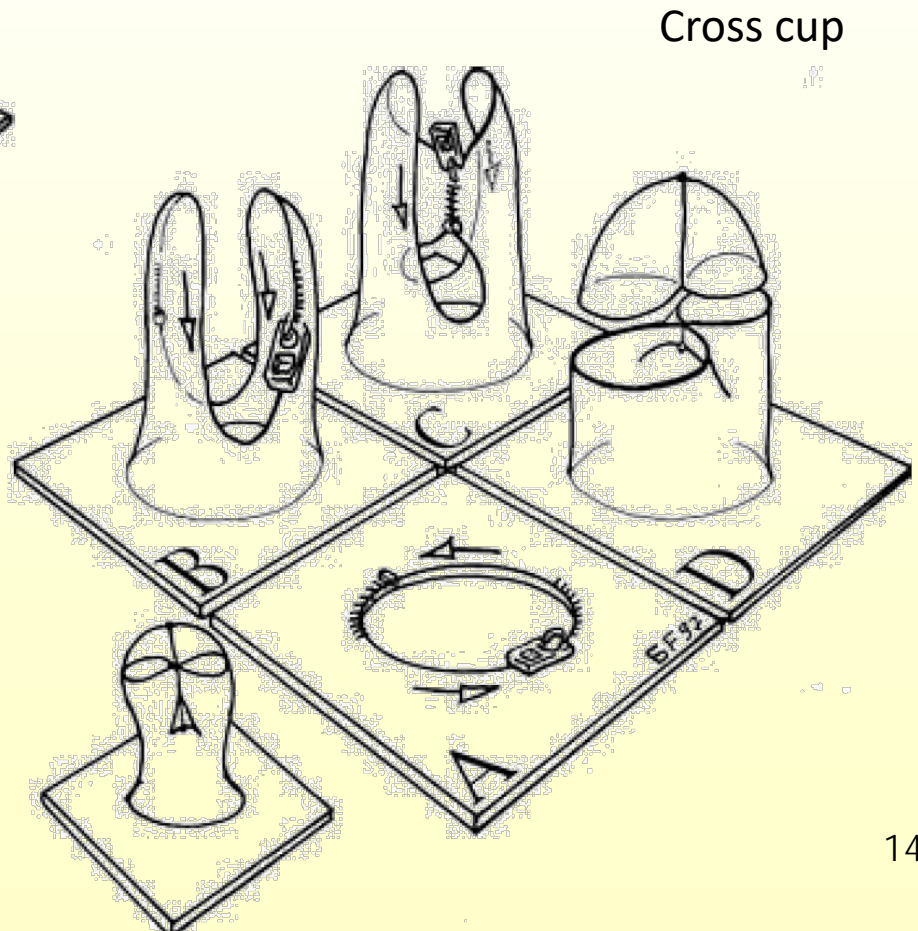


Cross cup

J. Conway's ZIP Proof (Zero Irrelevancy Proof)



Cup



Cross cup

Algebraic Topology



Henri Poincaré

Luitzen Egbertus Jan Brouwer



Emmy Noether



Heinz Hopf



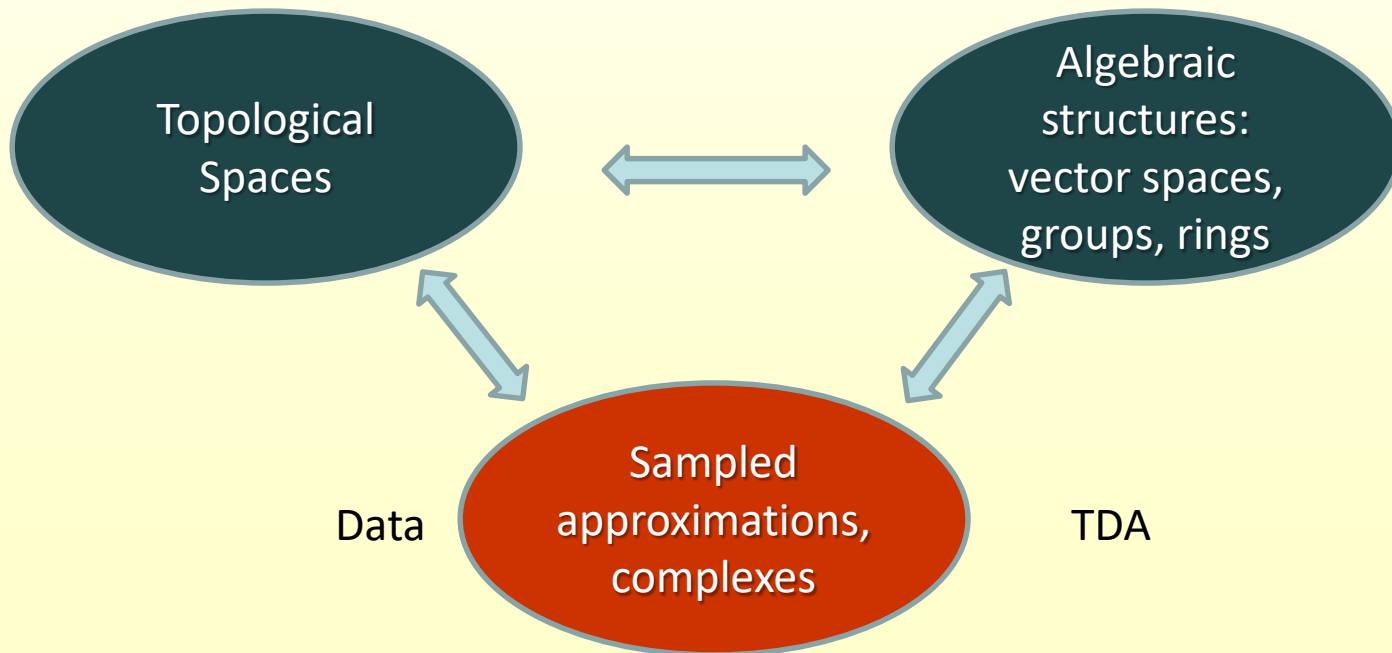
Leopold Vietoris

Topology and Topology Inference

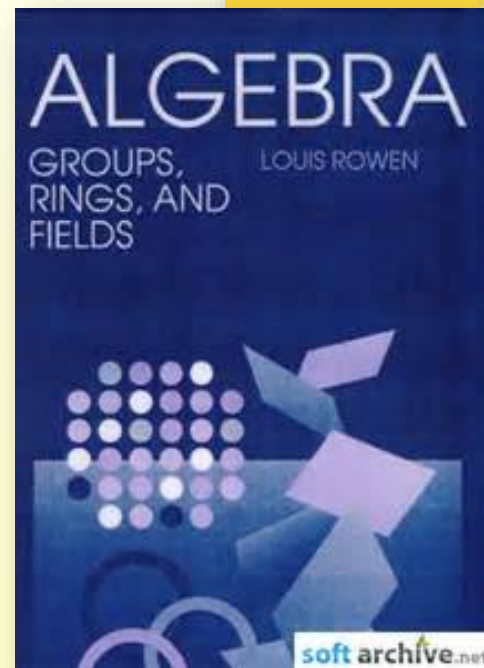
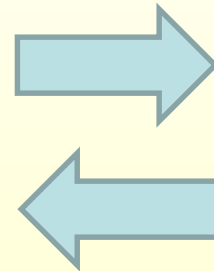
- ◆ Unlike geometry, topology studies **global** structure

- ◆ Getting to topology via **algebra**

algebraic topology



From Data to Algebraic Objects

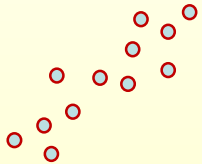


Algebraic objects as
data descriptors

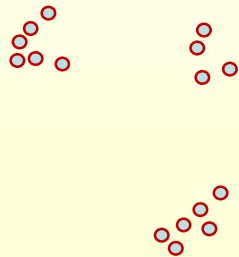
Sampled Spaces

TDA Studies Sampled Spaces

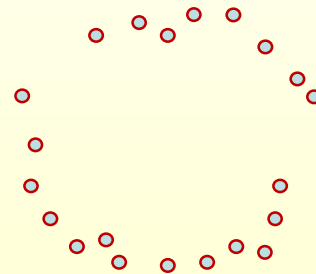
“The Shape of Data”



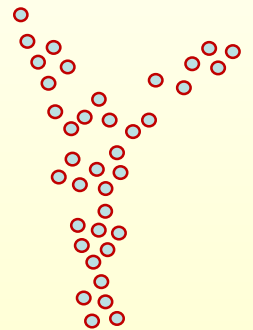
Regression



Cluster

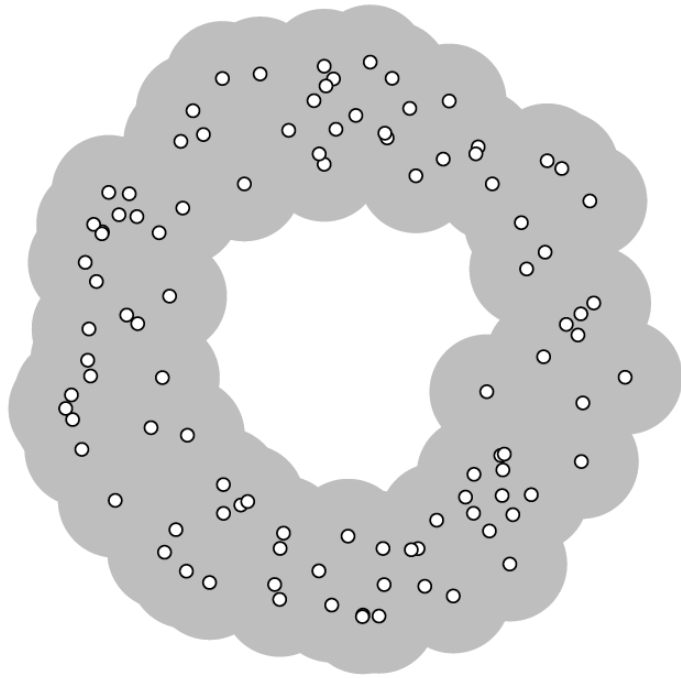


Loop



Flared

Recovering “Shape” from Sampled Data



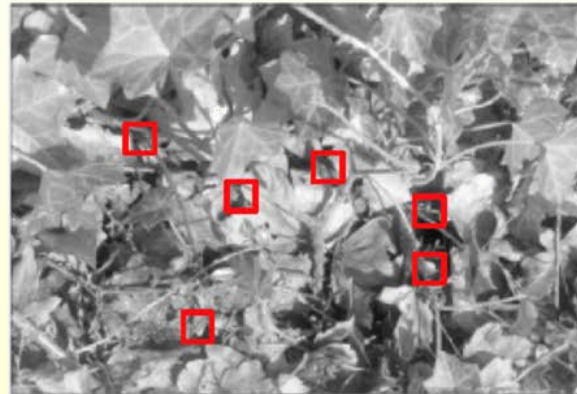
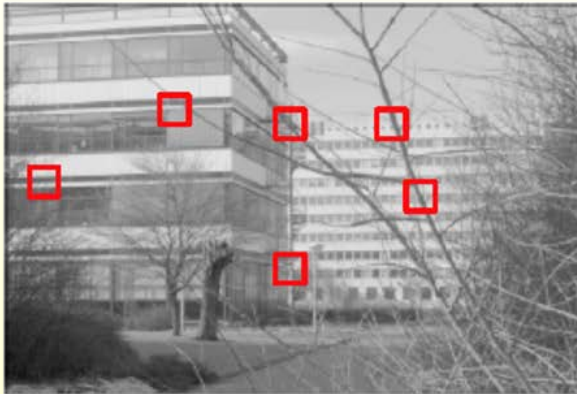
1. Set of points in \mathbb{R}^2
2. Looks like an annulus.

What is this?

What does it look like ?

Aim: recover the topology of the underlying space from which the data was sampled

Example: The Space of Natural Images



Example: The Space of Natural Images

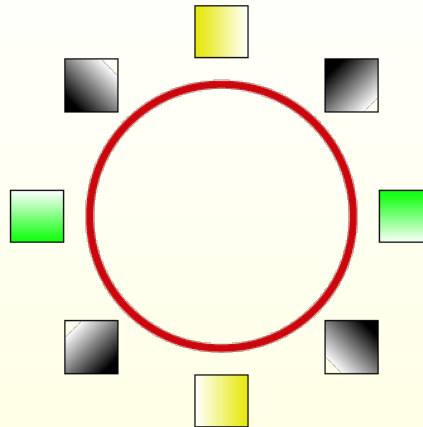
(Carlsson, Ishkanov, de Silva, Zomorodian IJCV 2008)

- ◆ Lee-Mumford-Pedersen investigated whether a statistically significant difference exists between natural and random images
- ◆ Natural images form a “subspace” of all images. Dimension of ambient space e.g. $640 \times 480 = 307\,200$
- ◆ This space of natural images should have:
 - ◆ high dimension: there are many different images
 - ◆ even higher co-dimension: random images look nothing like natural ones
- ◆ Data is a collection of black-and-white images used in cognitive science research

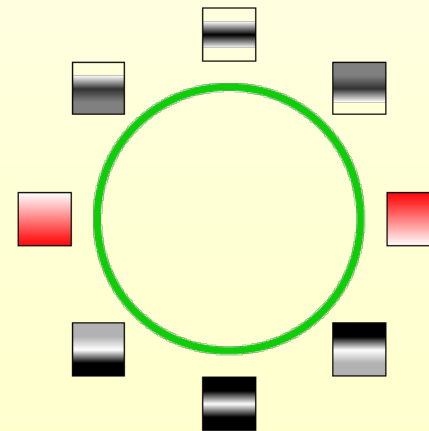
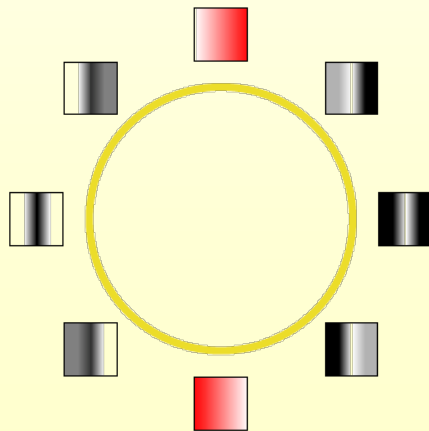
Natural 3 x 3 Patches

- ◆ Instead of studying entire images, we consider the distribution of 3 x 3 pixel patches
- ◆ Most of these are roughly constant in natural images -- they drown out structure
- ◆ L.M.P. chose 8,500,000 patches with high contrast
- ◆ Each 3 x 3-patch is considered a vector in \mathbb{R}^9
- ◆ Normalize brightness: $\mathbb{R}^9 \rightarrow \mathbb{R}^8$
- ◆ Normalize contrast: $\mathbb{R}^8 \rightarrow S^7$

High-Density Areas

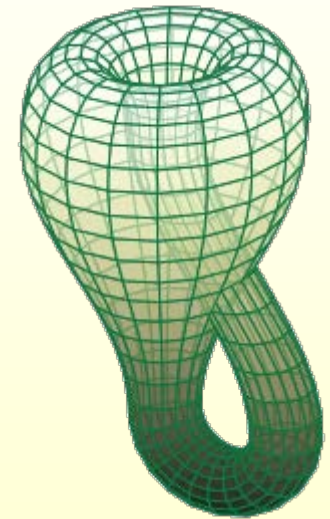
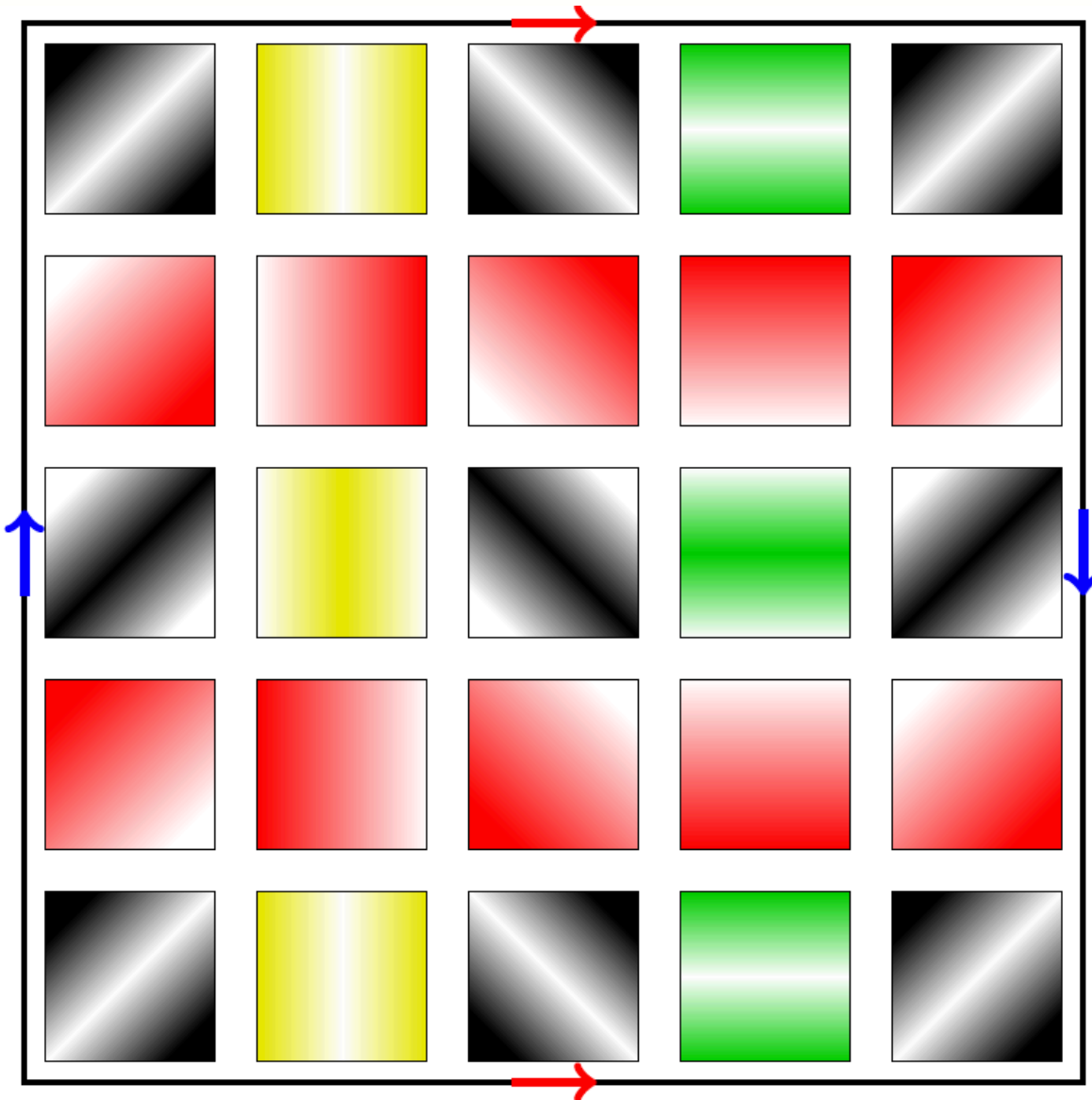


highest density

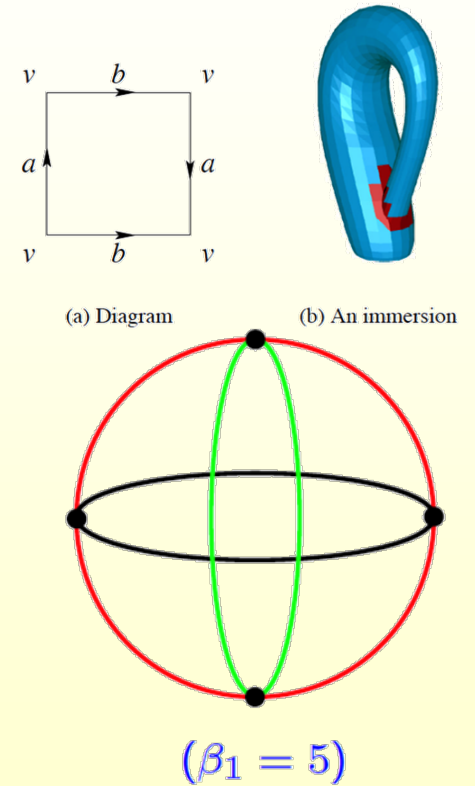
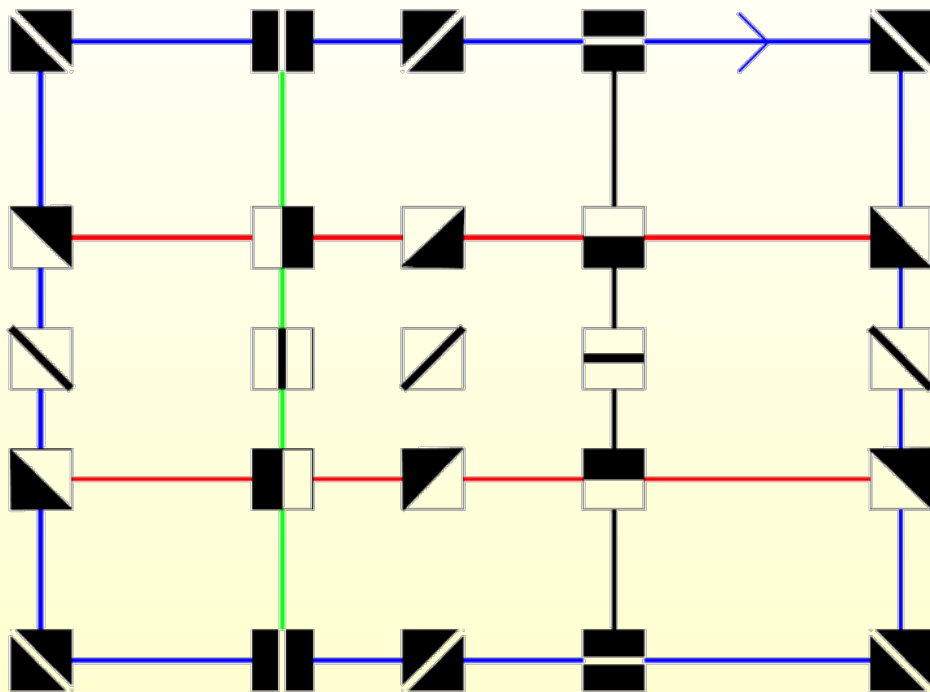


next highest density

Klein Bottle of Pixel Patches



Klein Bottle Structure



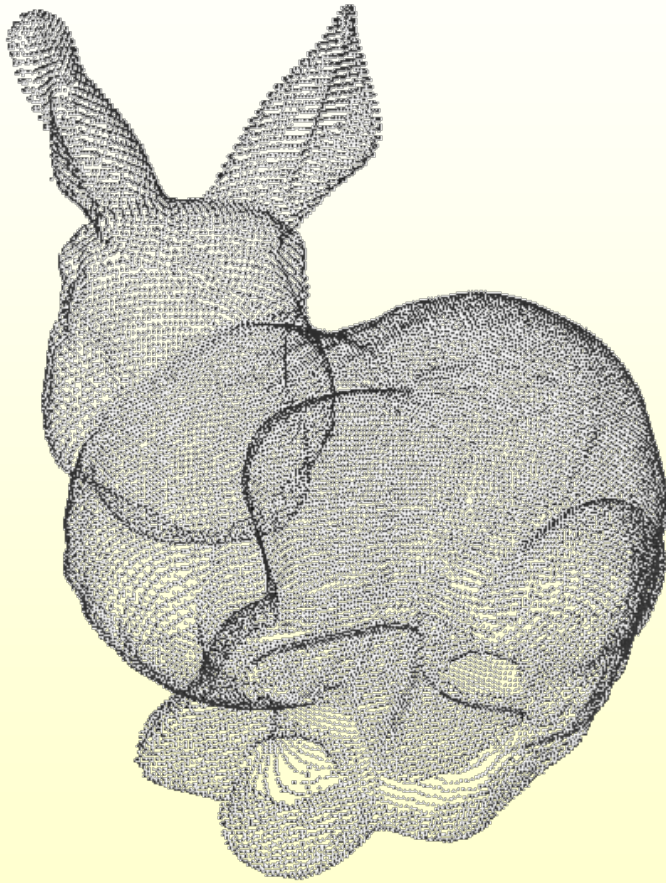
(source: [Carlsson, Ishkhanov, de Silva, Zomorodian 2008])

Applications of the Analysis

- ◆ An efficient way to parametrize image patches.
- ◆ **Image compression:** a 3 x 3-cluster may be described using 4 values:
 - ◆ 2x : Position of its projection onto the Klein bottle
 - ◆ 1x : Original brightness
 - ◆ 1x : Original contrast
- ◆ **Texture analysis:** textures yield distributions of occurring patches on the Klein bottle. Rotating the texture corresponds to translating the distribution.
- ◆ **Deep nets:** regularization of initial layers

Simplicial Complexes: Combinatorial Topology

Point Clouds Have No Higher-d Topology

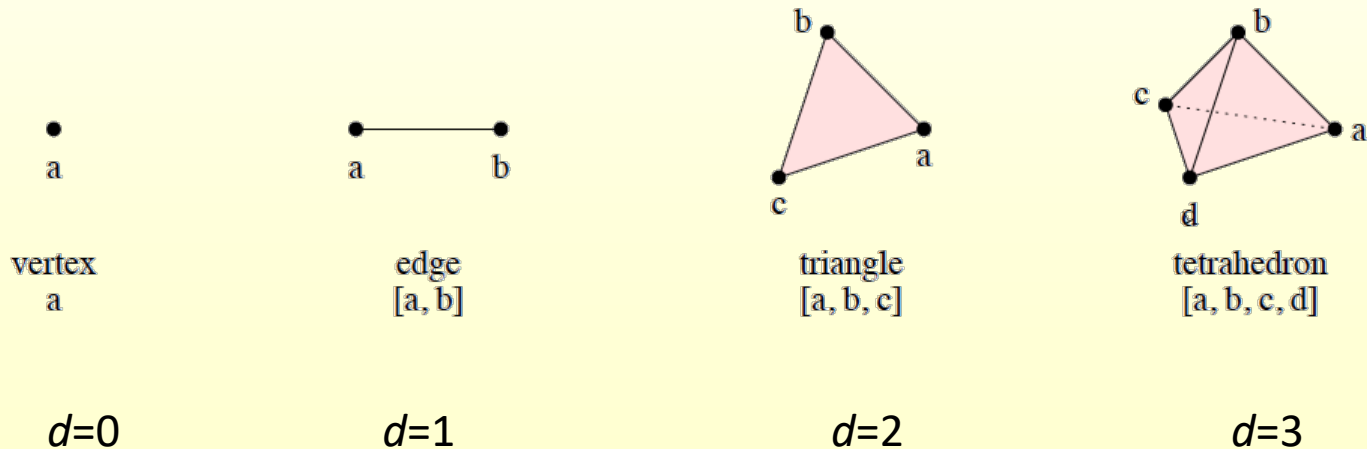


Connect the points into simplicial complexes

Simplicial homology

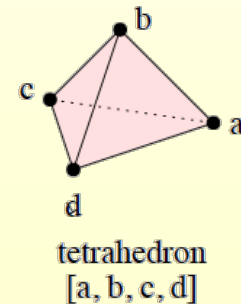
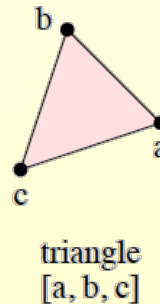
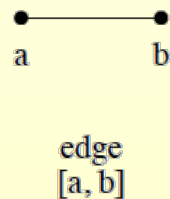
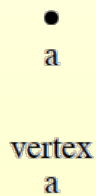
A Simplex

- A **k -simplex** is the convex hull of $k + 1$ affinely independent points $S = \{v_0, v_1, \dots, v_k\}$. The points in S are the **vertices** of the simplex.
- A k -simplex is a k -dimensional subspace of \mathbb{R}^d , $\dim \sigma = k$.



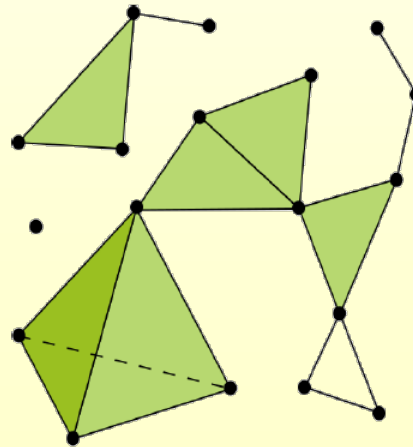
Faces / Subsimplices

- σ : a k -simplex defined by $S = \{v_0, v_1, \dots, v_k\}$.
- τ defined by $T \subseteq S$ is a **face** of σ
- σ is its **coface**.
- $\sigma \geq \tau$ and $\tau \leq \sigma$.
- $\sigma \leq \sigma$ and $\sigma \geq \sigma$.

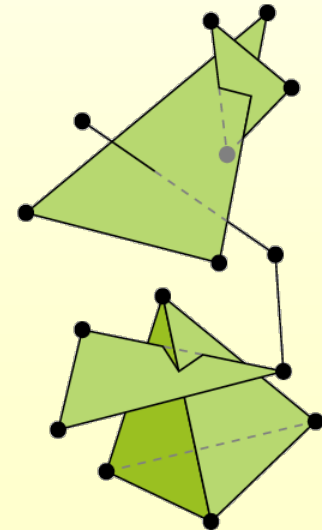


Simplicial Complexes

- A **simplicial complex** K is a finite set of simplices such that
 1. $\sigma \in K, \tau \leq \sigma \Rightarrow \tau \in K$,
 2. $\sigma, \sigma' \in K \Rightarrow \sigma \cap \sigma' \leq \sigma, \sigma'$ or $\sigma \cap \sigma' = \emptyset$.
- The **dimension** of K is $\dim K = \max\{\dim \sigma \mid \sigma \in K\}$.
- The **vertices** of K are the zero-simplices in K .
- A simplex is **principal** if it has no proper coface in K .

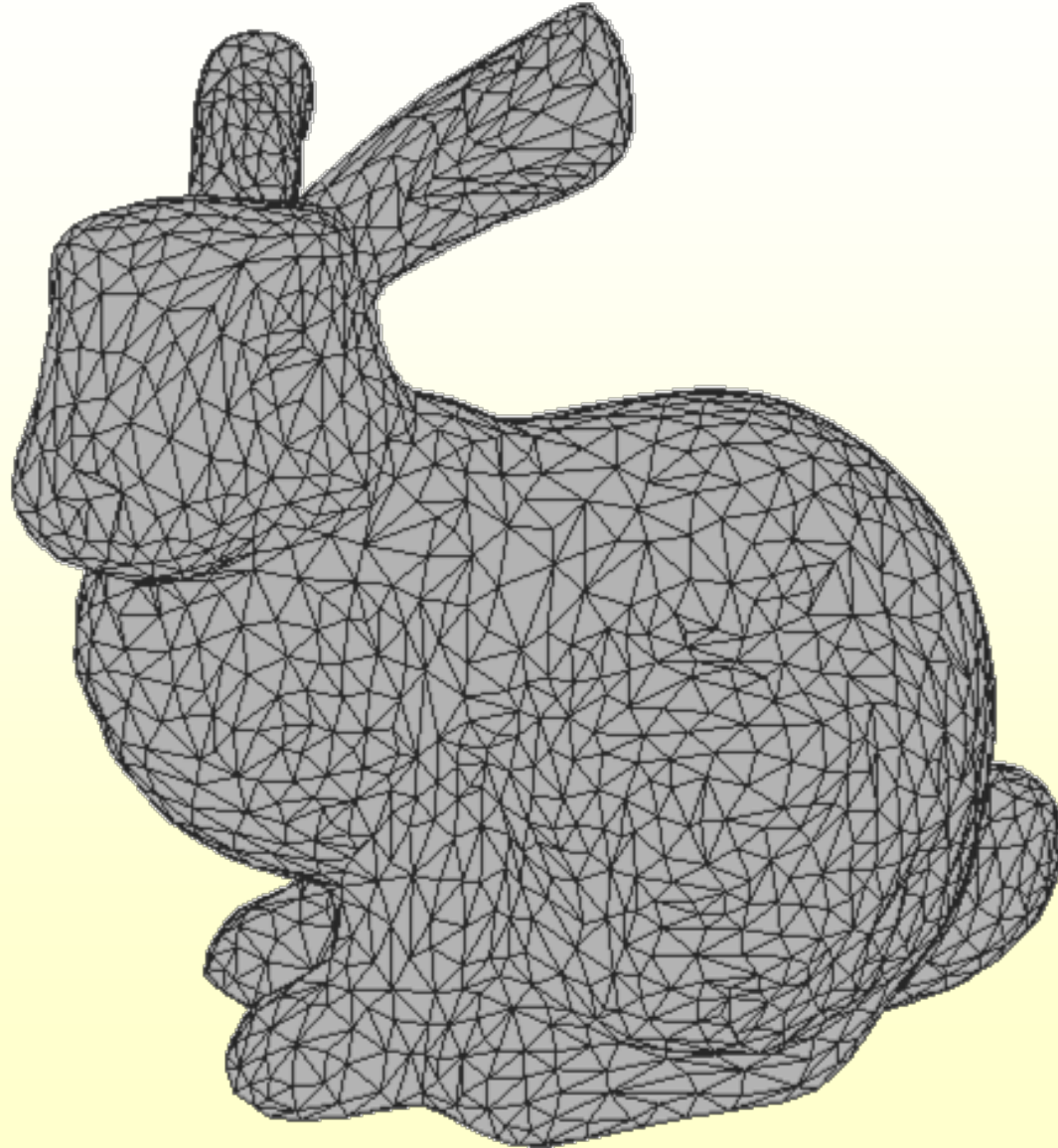


(left) an example

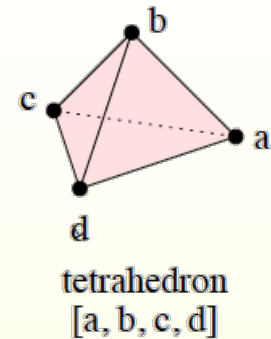
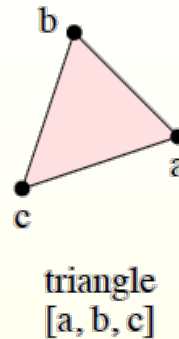
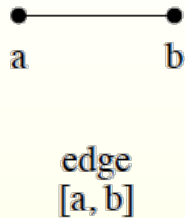
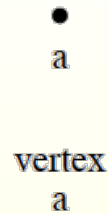


(right) a non example

Meshes are 2-d Complexes



Size of a Simplex



- \emptyset is the (-1) -simplex.
- A k -simplex has $\binom{k+1}{l+1}$ faces of dimension l
- Total size is:

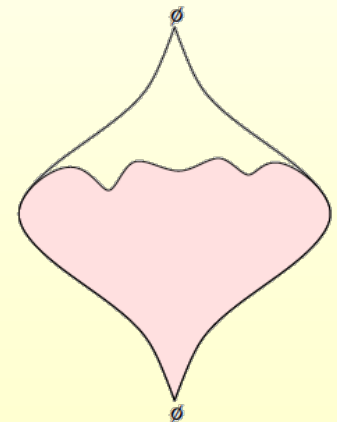
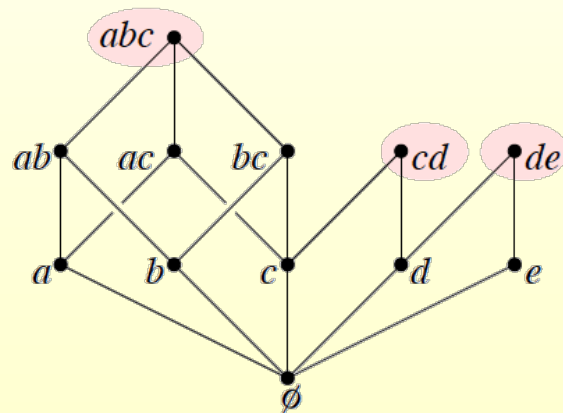
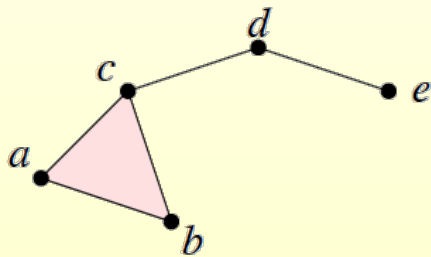
$$\sum_{l=-1}^k \binom{k+1}{l+1} = 2^{k+1}$$

k/l	0	1	2	3
0	1	0	0	0
1	2	1	0	0
2	3	3	1	0
3	4	6	4	1
4	?	?	?	?

Binomial coefficients

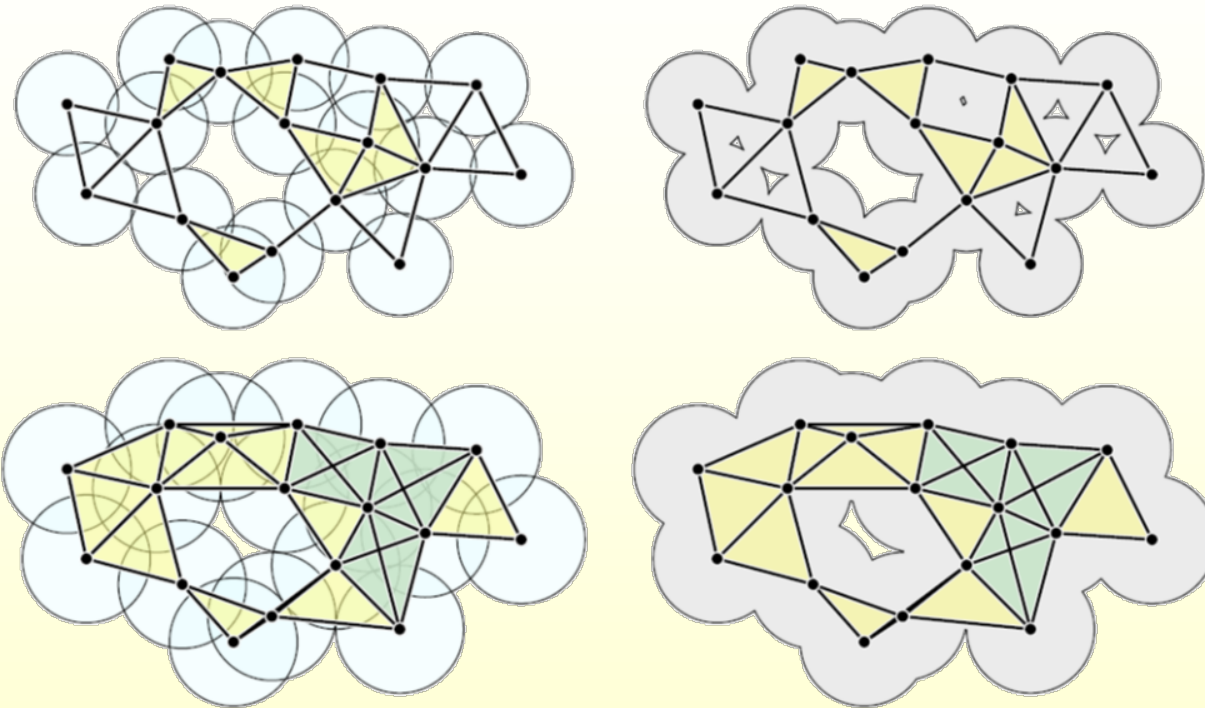
Abstract Simplicial Complexes

- An **abstract simplicial complex** is a set K , together with a collection \mathcal{S} of subsets of K called **(abstract) simplices** such that:
 1. For all $v \in K$, $\{v\} \in \mathcal{S}$. We call the sets $\{v\}$ the **vertices** of K .
 2. If $\tau \subseteq \sigma \in \mathcal{S}$, then $\tau \in \mathcal{S}$.
- We call \mathcal{S} the complex.



Natural partial order structure

The Nerve of a Finite Cover

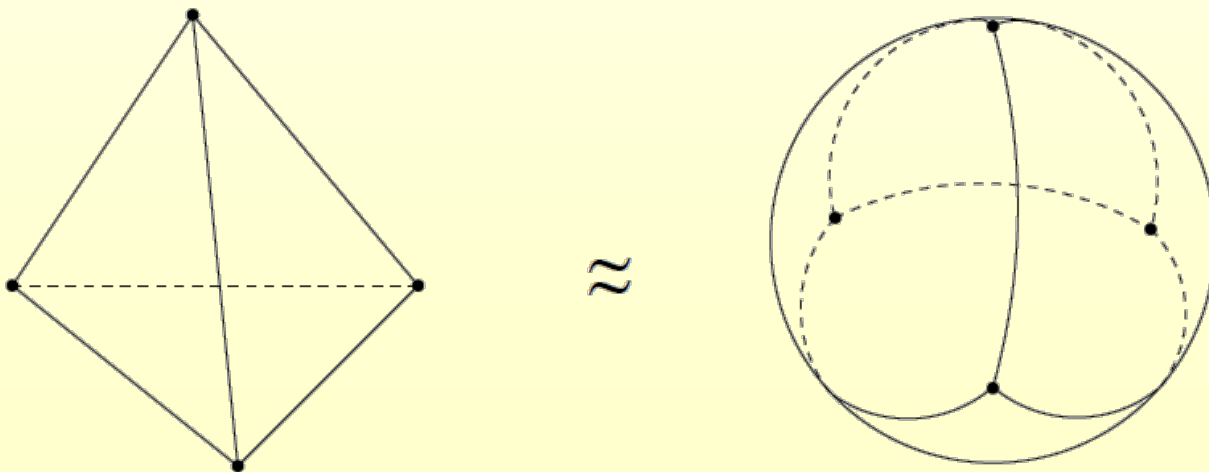


The **nerve** of \mathcal{U} is the simplicial complex $K(\mathcal{U})$ defined by

$$\sigma = [U_{i_0}, \dots, U_{i_k}] \in K(\mathcal{U}) \iff \bigcap_{i=1}^k U_{i_j} \neq \emptyset$$

Continuous to Discrete Link: Triangulations

- The **underlying space** $|K|$ of a simplicial complex K is $|K| = \cup_{\sigma \in K} \sigma$.
- $|K|$ is a topological space.
- A **triangulation** of a topological space \mathbb{X} is a simplicial complex K such that $|K| \approx \mathbb{X}$.



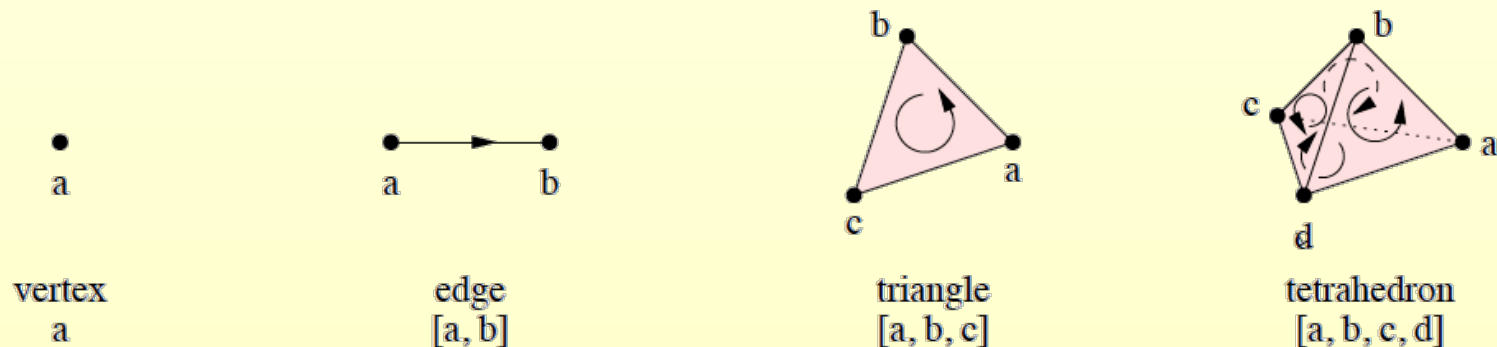
Orientability

- An **orientation** of a k -simplex $\sigma \in K$, $\sigma = \{v_0, v_1, \dots, v_k\}$, $v_i \in K$ is an equivalence class of orderings of the vertices of σ , where

$$(v_0, v_1, \dots, v_k) \sim (v_{\tau(0)}, v_{\tau(1)}, \dots, v_{\tau(k)})$$

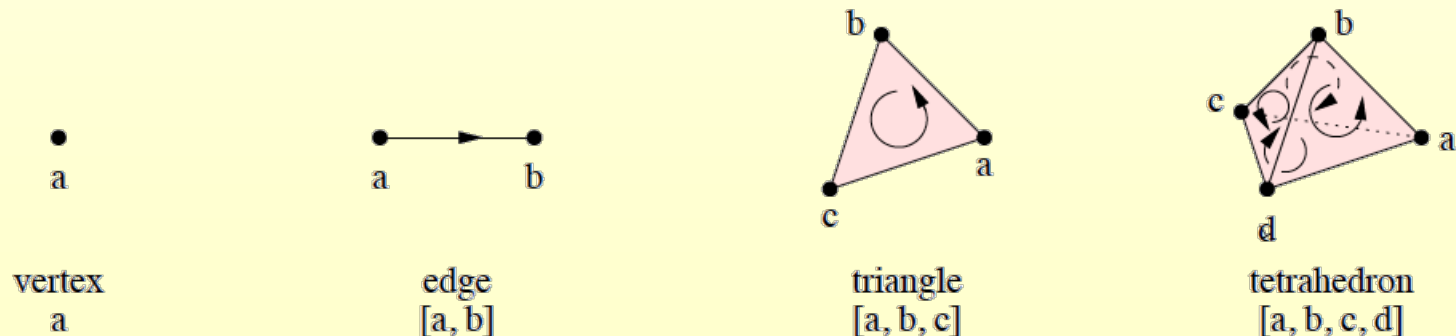
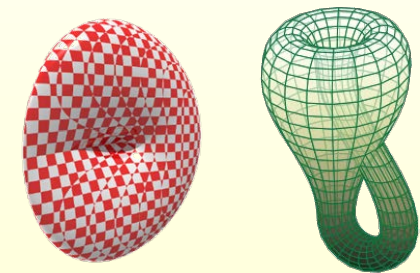
are equivalent orderings if the parity of the permutation τ is even.

- We denote an **oriented simplex**, a simplex with an equivalence class of orderings, by $[\sigma]$.



Orientability

- Two k -simplices sharing a $(k - 1)$ -face σ are **consistently oriented** if they induce different orientations on σ .
- A triangulable d -manifold is **orientable** if all d -simplices can be oriented consistently.
- Otherwise, the d -manifold is **non-orientable**



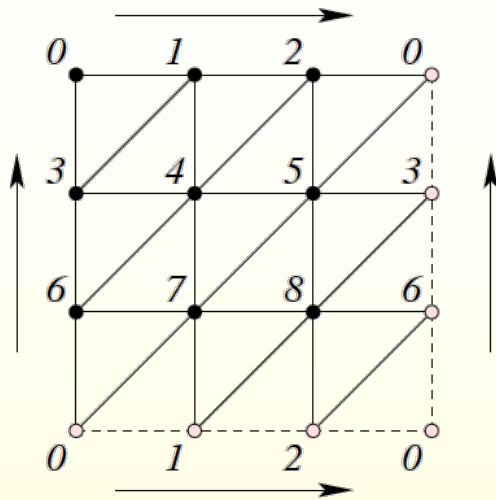
Euler Characteristic: A Topological Invariant

- K a simplicial complex with s_k k -simplices.
- The **Euler characteristic** $\chi(K)$ is

$$\chi(K) = \sum_{i=0}^{\dim K} (-1)^i s_i = \sum_{\sigma \in K - \{\emptyset\}} (-1)^{\dim \sigma}.$$

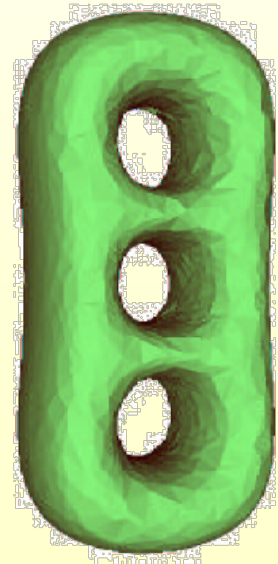
- $v - e + f = 1$ (Graph Theory)
- Invariant for $|K|$
- **Any** triangulation gives the same answer!
- Intrinsic property

More on Euler



2-Manifold	χ
Sphere S^2	2
Torus T^2	0
Klein bottle \mathbb{K}^2	0
Projective plane \mathbb{RP}^2	1

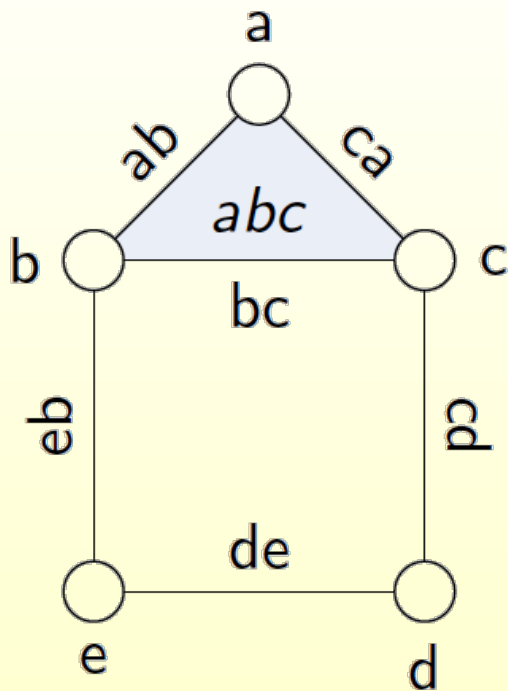
- (Theorem) For compact surfaces M_1, M_2 ,
 $\chi(M_1 \# M_2) = \chi(M_1) + \chi(M_2) - 2$.
- $\chi(gT^2) = 2 - 2g$
- $\chi(g\mathbb{RP}^2) = 2 - g$
- The connected sum of g tori is called a surface with **genus** g .



Topological Classification via Invariants

- (Theorem) Closed compact surfaces M_1 and M_2 are homeomorphic, $M_1 \approx M_2$ iff
 1. $\chi(M_1) = \chi(M_2)$ and
 2. either both surfaces are orientable or both are non-orientable.
- “iff” so full answer. We’re done!
- Higher dimensions?

Summary: Simplicial Complexes

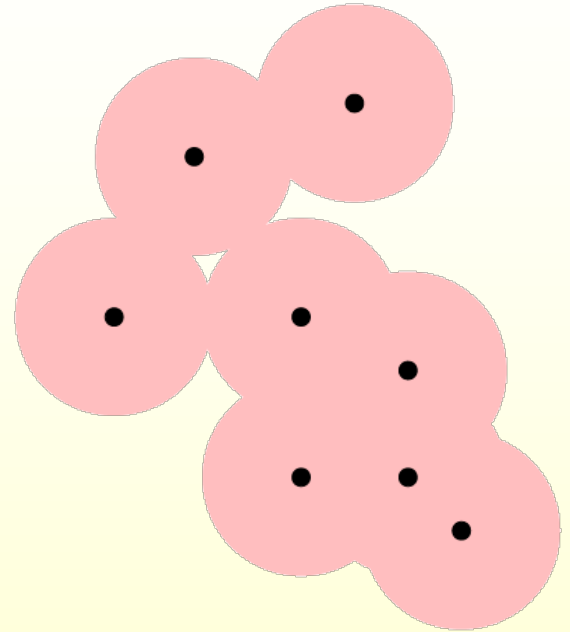
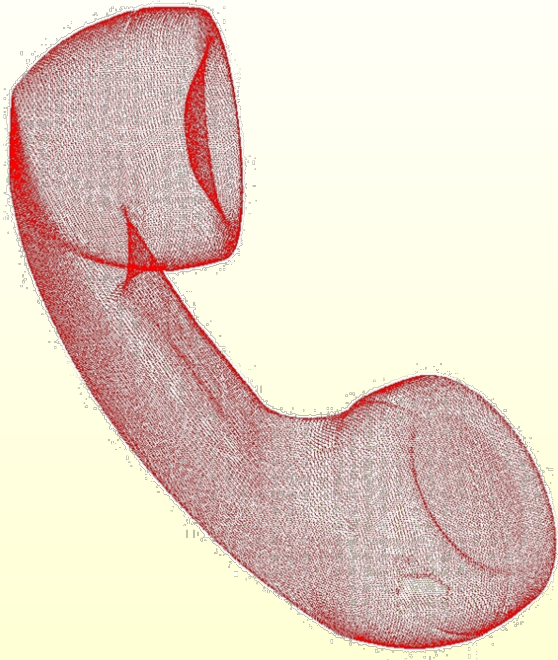


A simplicial complex is a collection of simplices

- ▶ Each simplex has a dimension.
- ▶ Collection is closed under subset relation.
- ▶ Simplices of dimension d have $d + 1$ vertices
- ▶ Each simplex represented by an ordered list of vertices

Useful Complexes on Point Clouds

ϵ -Balls



- ϵ -ball: $B_\epsilon(x) = \{y \mid d(x, y) < \epsilon\}$.
- Open sets and topology
- Manifold is $\tilde{M} = \bigcup_{m_i \in M} B_\epsilon(m_i)$

A Model Space

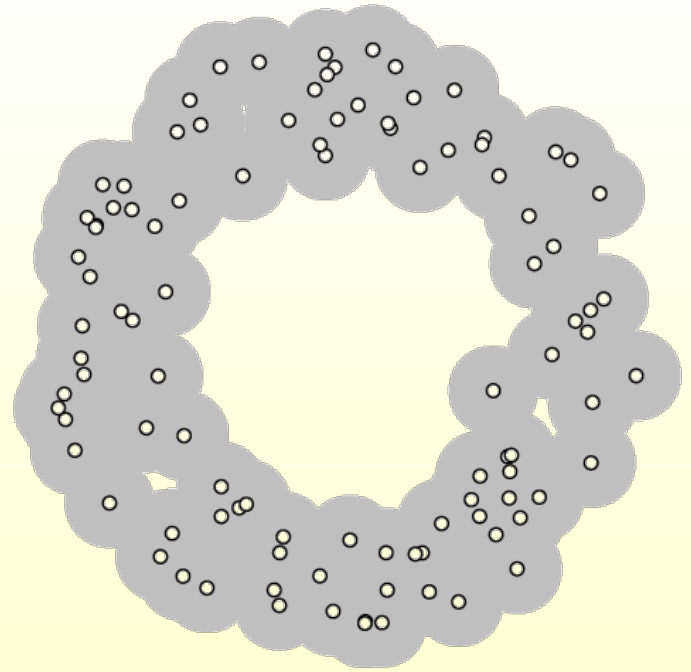
For a dataset X we study the topology of the *union of balls*

$$M_\epsilon = \bigcup_{x \in X} B_\epsilon(x)$$

Two Issues:

Scale: No natural choice of ϵ !

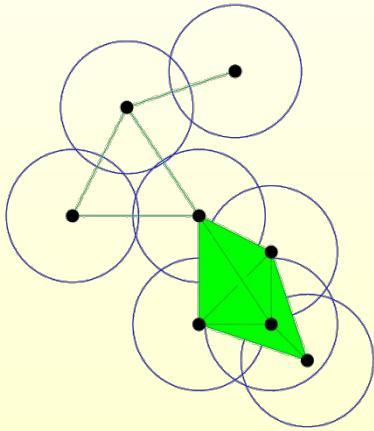
Conception: How to encode M_ϵ on computer?



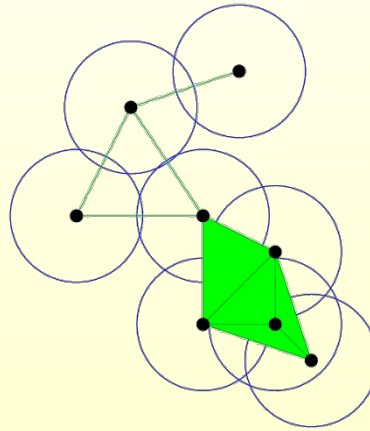
Complex Zoo

Must choose which simplices to introduce

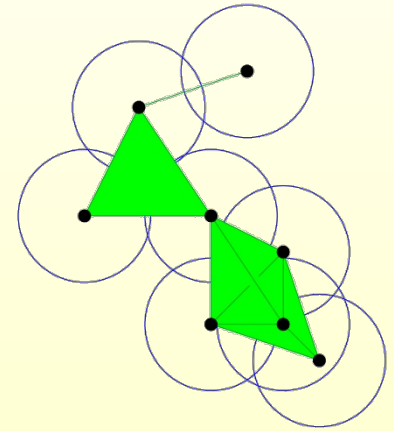
Čech



Alpha

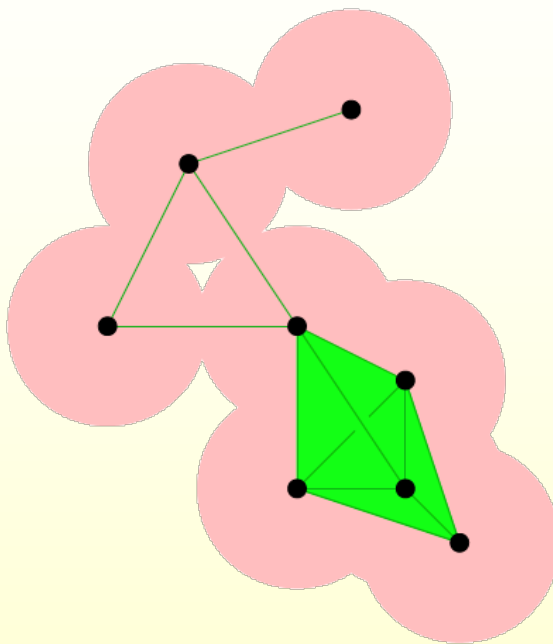


Rips



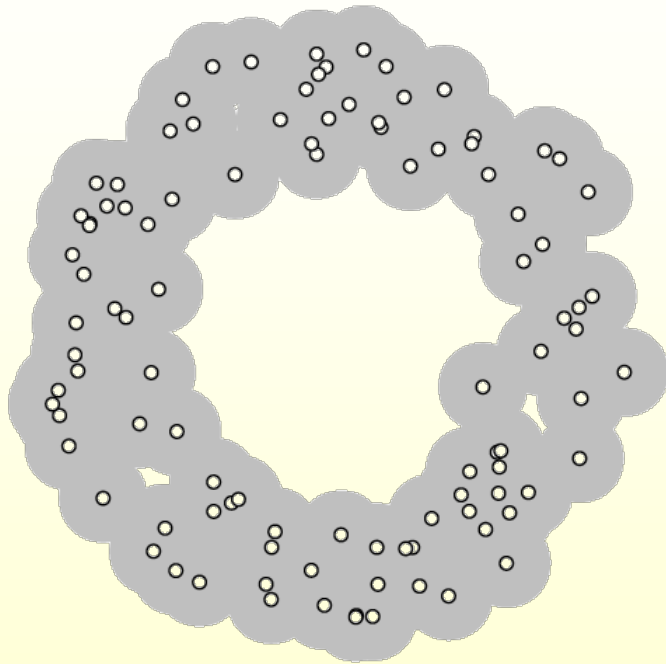
Combinatorial complexes provide discrete representations of the underlying space

Čech Complex



- $C_\epsilon(M) = \{ \text{conv } T \mid T \subseteq M, \bigcap_{m_i \in T} B_\epsilon(m_i) \neq \emptyset \} .$
- $\sum_{k=0}^m \binom{m}{k} = 2^{m+1} - 1$
- $C_\epsilon(M) \simeq \tilde{M}$

Čech Complex



\check{C}_ϵ

The Čech complex \check{C}_ϵ encodes the intersection pattern of M_ϵ : Encode:

Points as *vertices*
(0-cells)

Čech Complex



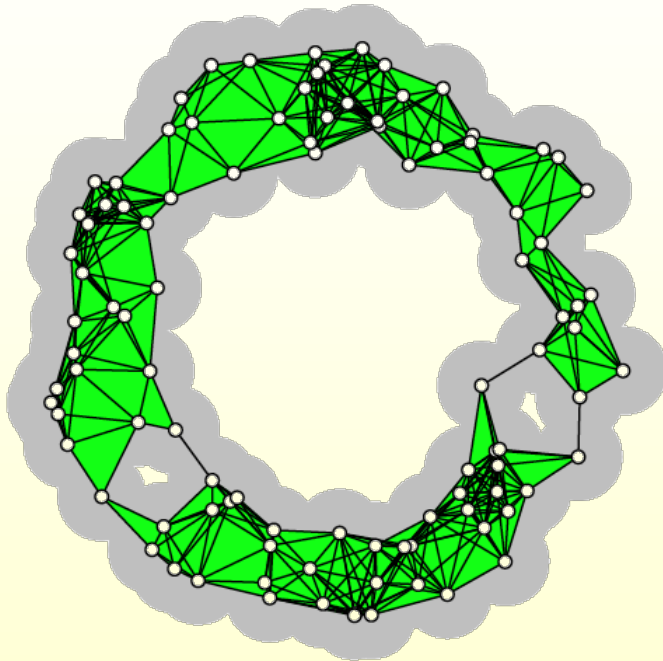
$\check{C}_{.6}$

The Čech complex \check{C}_ϵ encodes the intersection pattern of M_ϵ : Encode:

Points as *vertices*
(0-cells)

Pairwise intersections
as *edges* (1-cells)

Čech Complex



\check{C}_ϵ

The Čech complex \check{C}_ϵ encodes the intersection pattern of M_ϵ : Encode:

Points as *vertices*
(0-cells)

Pairwise intersections
as *edges* (1-cells)

Threeway intersections
as *triangles* (2-cells)

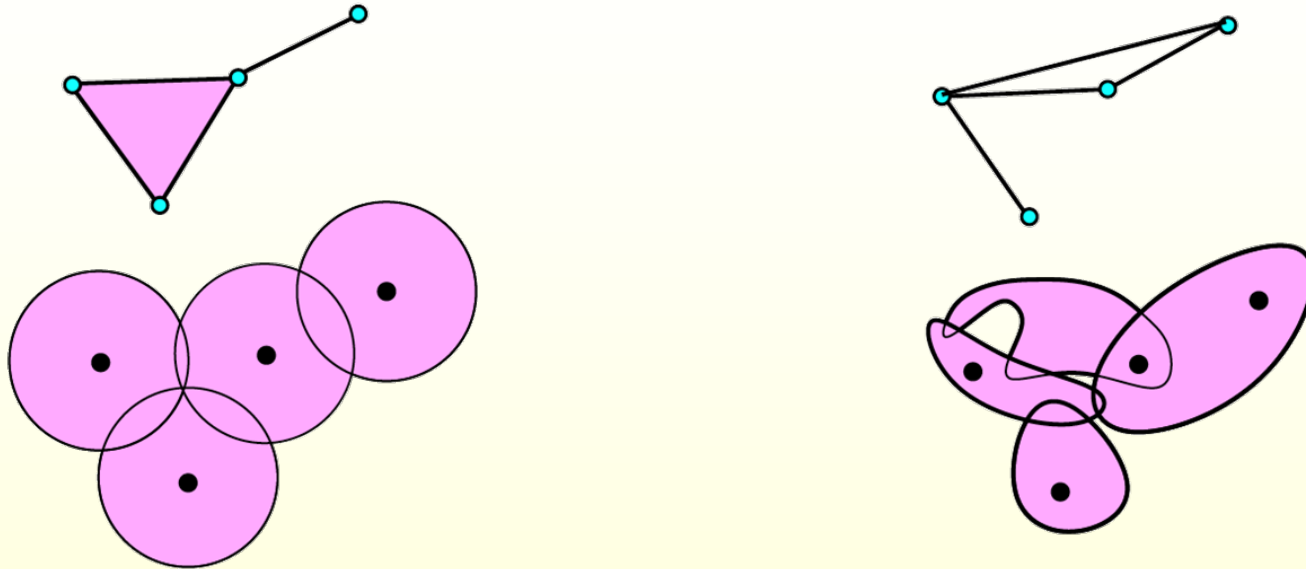
k -way intersections as
($k+1$)-cells

Lemma (Nerve Lemma, Leray '45)

\check{C}_ϵ is topologically equivalent to M_ϵ .

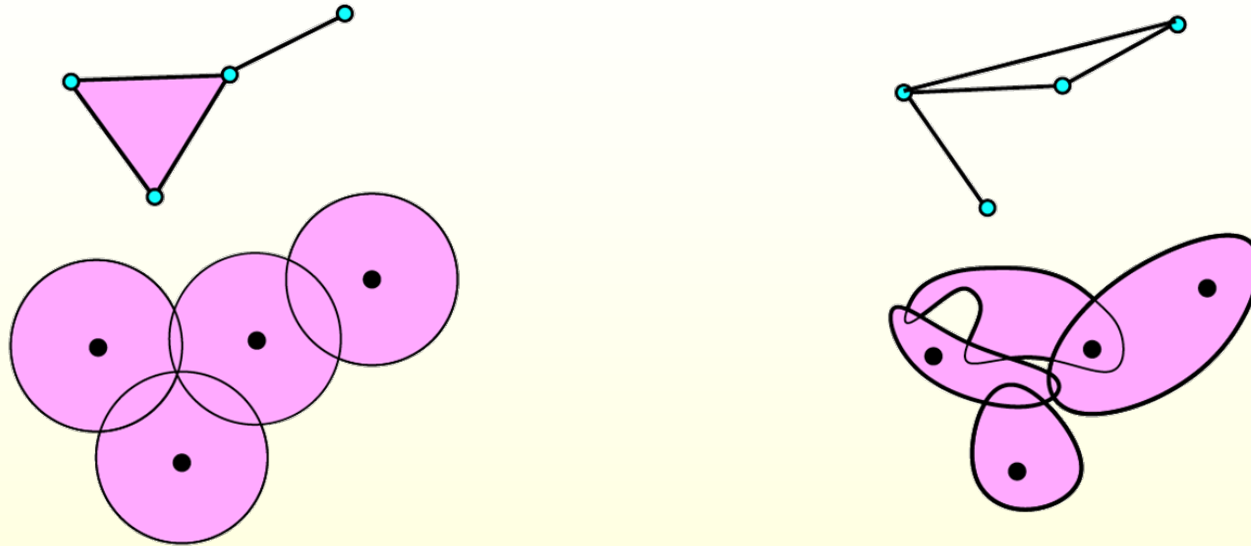
Can be hard to compute ...

General Čech Complex = Nerve



- Let $\mathcal{U} = (U_i)_{i \in I}$ be a covering of a topological space X by open sets:
 $X = \cup_{i \in I} U_i$.
- The Čech complex $C(\mathcal{U})$ associated to the covering \mathcal{U} is the simplicial complex defined by:
 - the vertex set of $C(\mathcal{U})$ is the set of the open sets U_i
 - $[U_{i_0}, \dots, U_{i_k}]$ is a k -simplex in $C(\mathcal{U})$ iff $\cap_{j=0}^k U_{i_j} \neq \emptyset$.

General Čech Complex



Nerve theorem (Leray): If all the intersections between opens in \mathcal{U} are either empty or contractible then $C(\mathcal{U})$ and $X = \cup_{i \in I} U_i$ are homotopy equivalent.

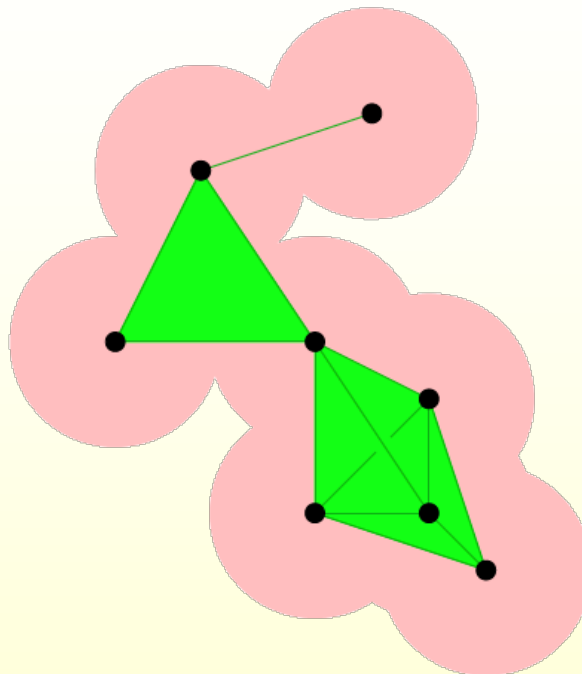
\Rightarrow The combinatorics of the covering (a simplicial complex) carries the topology of the space.

Warning: even when the open sets are euclidean balls, the computation of the Čech complex is a very difficult task!

Rips-Vietoris Complex

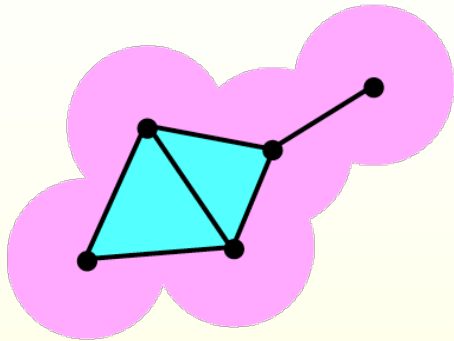
The “poor man’s” alternative
to the Čech

This is a common complex
For computations

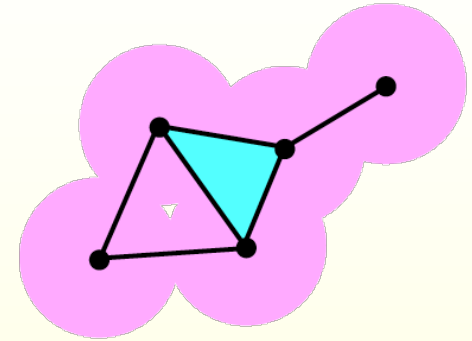


- $R_\epsilon(M) = \{\text{conv } T \mid T \subseteq M, d(m_i, m_j) < \epsilon, m_i, m_j \in T\}.$
- Still $O\left(\binom{m}{k}\right)$ for the k th skeleton
- Need $(k + 1)$ st skeleton for computing H_k

Rips vs. Čech



Rips vs Čech



Let $L = \{p_0, \dots, p_n\}$ be a (finite) point cloud (in a metric space).

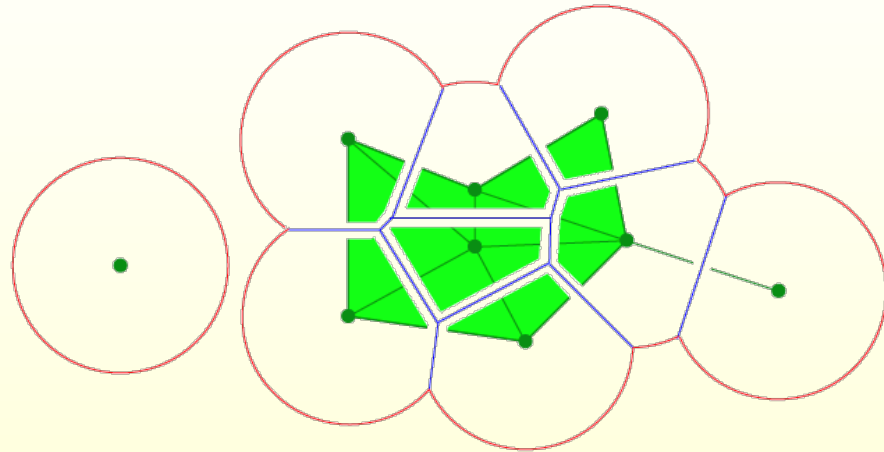
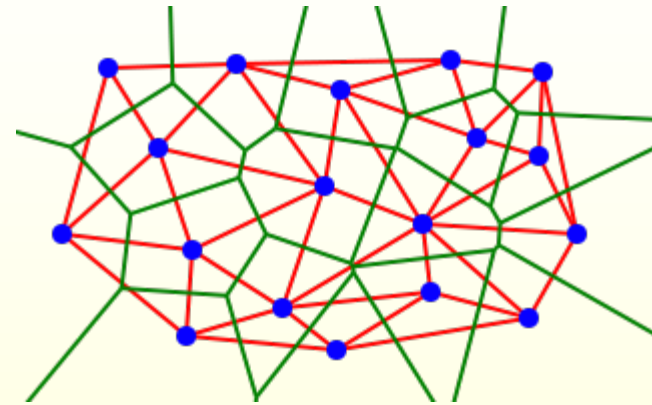
The **Rips complex** $\mathcal{R}^\alpha(L)$: for $p_0, \dots, p_k \in L$,

$$\sigma = [p_0 p_1 \dots p_k] \in \mathcal{R}^\alpha(L) \text{ iff } \forall i, j \in \{0, \dots, k\}, d(p_i, p_j) \leq \alpha$$

- Easy to compute and fully determined by its 1-skeleton
- Rips-Čech interleaving: for any $\alpha > 0$,

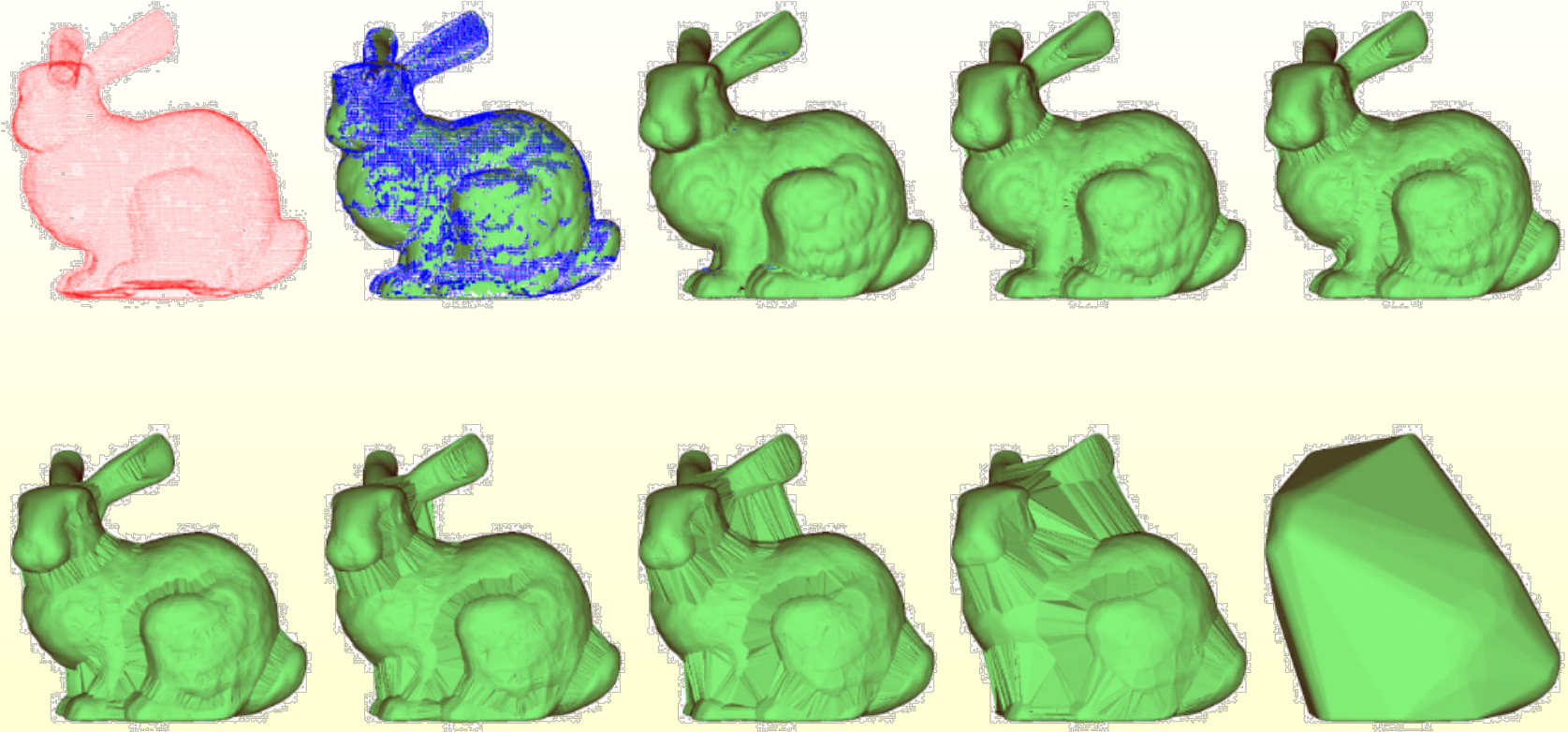
$$\mathcal{C}^{\frac{\alpha}{2}}(L) \subseteq \mathcal{R}^\alpha(L) \subseteq \mathcal{C}^\alpha(L) \subseteq \mathcal{R}^{2\alpha}(L) \subseteq \dots$$

Alpha Complex



- $V(m_i) = \{x \in \mathbb{R}^3 \mid d(x, m_i) \leq d(x, m_j) \forall m_j \in M\}$
- $\hat{V}(m_i) = B_\epsilon(m_i) \cap V(m_i)$
- $A_\epsilon = \left\{ \text{conv} T \mid T \subseteq M, \bigcap_{m_i \in T} \hat{V}(m_i) \neq \emptyset \right\}$
- $A_\epsilon(M) \simeq \tilde{M}$, $A_\epsilon \subseteq D$, the **Delaunay complex**
- $O(n \log n + n^{\lceil d/2 \rceil})$

Alpha Complexes on the Stanford Bunny



- 34,834 points, 1,026,111 complexes

Algebraic Structures: Group Theory

Groups

- A **group** $\langle G, * \rangle$ is a set G , together with a binary operation $*$ on G , such that the following axioms are satisfied:
 - (a) $*$ is associative.
 - (b) G has an **identity** e element for $*$ such that $e * x = x * e = x$ for all $x \in G$.
 - (c) any element a has an **inverse** a' with respect to the operation $*$, i.e. $\forall a \in G, \exists a' \in G$ such that $a' * a = a * a' = e$.
- If G is finite, the **order** of G is $|G|$.
- We often omit the operation and refer to G as the group.
- $\langle \mathbb{Z}, + \rangle, \langle \mathbb{R}, \cdot \rangle, \langle \mathbb{R}, + \rangle$, are all groups.
- A group G is **abelian** if its binary operation $*$ is commutative.

Subgroups

- Let $\langle G, * \rangle$ be a group and $S \subseteq G$. If S is closed under $*$, then $*$ is the **induced operation on S from G** .
- A subset $H \subseteq G$ of group $\langle G, * \rangle$ is a **subgroup of G** if H is a group and is closed under $*$. The subgroup consisting of the identity element of G , $\{e\}$ is the **trivial subgroup** of G . All other subgroups are **nontrivial**.
- (Theorem) $H \subseteq G$ of a group $\langle G, * \rangle$ is a subgroup of G iff:
 1. H is closed under $*$,
 2. the identity e of G is in H ,
 3. for all $a \in H$, $a^{-1} \in H$.
- Example: subgroups of \mathbb{Z}_4

Cosets

- Let H be a subgroup of G . Let the relation \sim_L be defined on G by:
 $a \sim_L b$ iff $a^{-1}b \in H$. Let \sim_R be defined by: $a \sim_R b$ iff $ab^{-1} \in H$.
Then \sim_L and \sim_R are both equivalence relations on G .
- Let H be a subgroup of group G . For $a \in G$, the subset
 $aH = \{ah \mid h \in H\}$ of G is the **left coset** of H containing a , and
 $Ha = \{ha \mid h \in H\}$ is the **right coset** of H containing a .
- If left and right cosets match, the subgroup is **normal**.
- All subgroups H of an abelian group G are normal, as
 $ah = ha, \forall a \in G, h \in H$
- $\{0, 2\}$ is a subgroup of \mathbb{Z}_4 . It is normal. The coset of 1 is
 $1 + \{0, 2\} = \{1, 3\}$. That's all folks!

Factor / Quotient Groups

- Let H be a normal subgroup of group G .
- Left coset multiplication is well-defined by the equation
$$(aH)(bH) = (ab)H$$
- The cosets of H form a group G/H under left multiplication
- G/H is the **factor group** (or **quotient group**) of G modulo H .
- The elements in the same coset of H are **congruent modulo H** .

Example

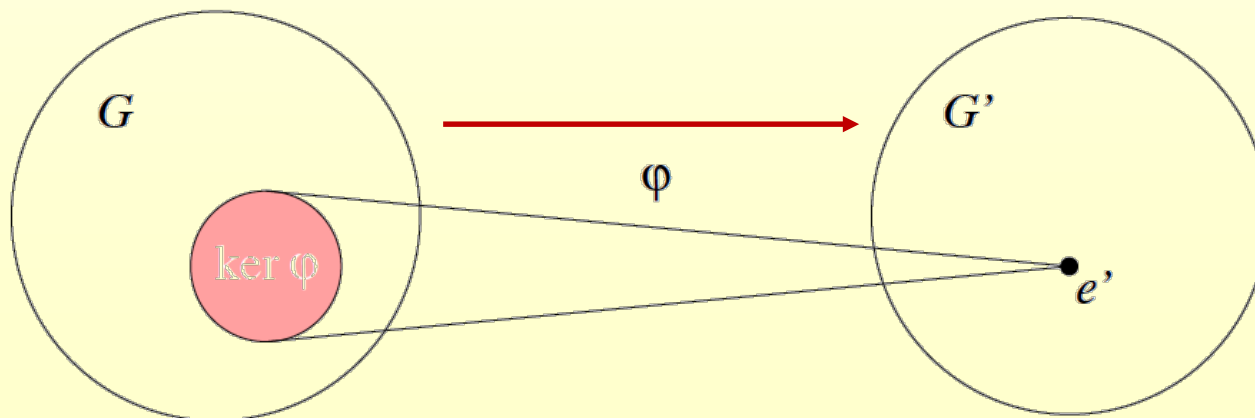
\mathbb{Z}_6	0	2	4	1	3	5
0	0	2	4	1	3	5
2	2	4	0	3	5	1
4	4	0	2	5	1	3
1	1	3	5	2	4	0
3	3	5	1	4	0	2
5	5	1	3	0	2	4

	*		

- $\{0, 2, 4\}$ is a normal subgroup
- Cosets $\{0, 2, 4\}, \{1, 3, 5\}$
- $\mathbb{Z}_6 / \{0, 2, 4\} \cong \mathbb{Z}_2$

Homomorphisms

- A map φ of a group G into a group G' is a *homomorphism* if $\varphi(ab) = \varphi(a)\varphi(b)$ for all $a, b \in G$.
- If e is the identity in G , then $\varphi(e)$ is the identity e' in G' .
- If $a \in G$, then $\varphi(a^{-1}) = \varphi(a)^{-1}$.
- If H is a subgroup of G , then $\varphi(H)$ is a subgroup of G' .
- If K' is a subgroup of G' , then $\varphi^{-1}(K')$ is a subgroup of G .
- The normal subgroup $\ker \varphi = \varphi^{-1}(\{e'\}) \subseteq G$, is the **kernel of φ** .



Decompositions for Finitely Generated Abelian Groups

- Let G_1, G_2, \dots, G_n be groups.
- The set is $\prod_{i=1}^n G_i$ (Cartesian product)
- Binary operation:
 $(a_1, a_2, \dots, a_n) \times (b_1, b_2, \dots, b_n) = (a_1 b_1, a_2 b_2, \dots, a_n b_n)$.
- Then $\langle \prod_{i=1}^n G_i, \times \rangle$ is a group.
- We call it the **direct product** of the groups G_i .
- Sometimes called **direct sum** with \oplus .

- (Theorem) Every finitely generated abelian group is isomorphic to product of cyclic groups of the form

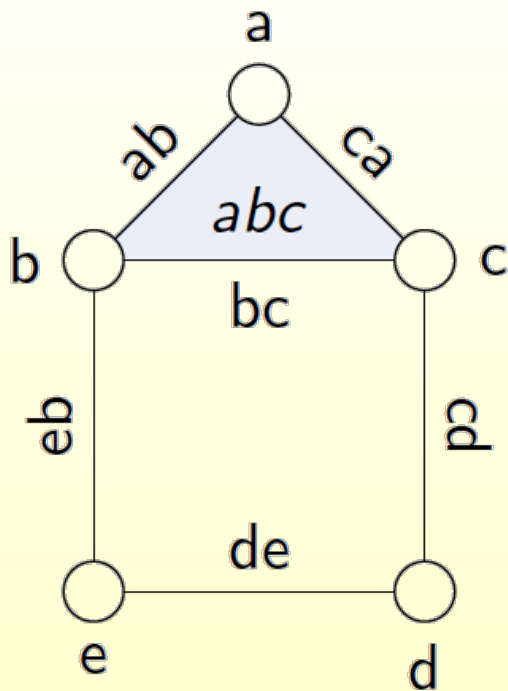
$$\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \dots \times \mathbb{Z}_{m_r} \times \mathbb{Z} \times \mathbb{Z} \times \dots \times \mathbb{Z},$$

where m_i divides m_{i+1} for $i = 1, \dots, r - 1$.

- The direct product is unique: the number of factors of \mathbb{Z} is unique and the cyclic group orders m_i are unique.
- Free: basis, rank, vector space
- Torsion: module

Algebraic Topology: Homology

Topology of Simplicial Complexes



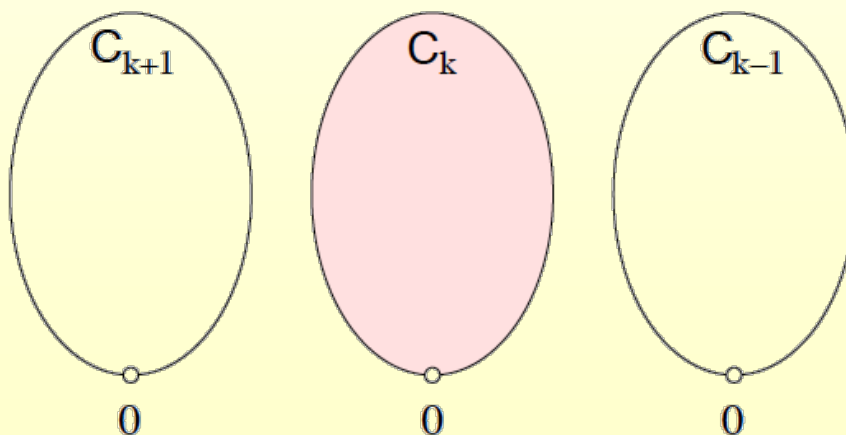
A simplicial complex is a collection of simplices

- ▶ Each simplex has a dimension.
- ▶ Collection is closed under subset relation.
- ▶ Simplices of dimension d have $d + 1$ vertices
- ▶ Each simplex represented by an ordered list of vertices

Chain Groups

Other coefficient
fields/rings also OK

- Simplicial complex K
- **k -chain**: $c = \sum_i n_i [\sigma_i]$, $n_i \in \mathbb{Z}$, $\sigma_i \in K$ (like a path)
- $[\sigma] = -[\tau]$ if $\sigma = \tau$ and σ and τ have different orientations.
- The **k th chain group \mathbf{C}_k** of K is the free abelian group on its set of oriented k -simplices
- $\text{rank } \mathbf{C}_k = ?$



Boundary Operator

- The boundary operator $\partial_k : \mathbf{C}_k \rightarrow \mathbf{C}_{k-1}$ is a homomorphism defined linearly on a chain c by its action on any simplex

$$\sigma = [v_0, v_1, \dots, v_k] \in c,$$

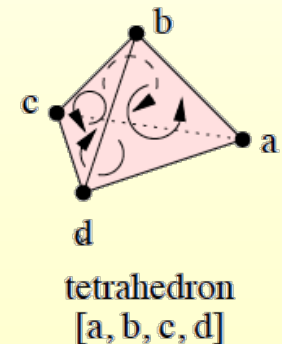
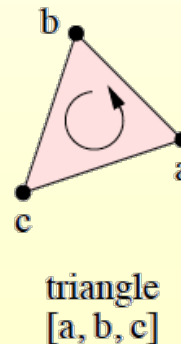
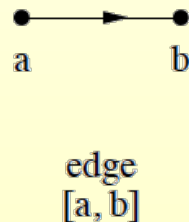
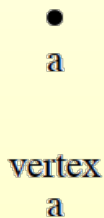
$$\partial_k \sigma = \sum_i (-1)^i [v_0, v_1, \dots, \hat{v}_i, \dots, v_k],$$

where \hat{v}_i indicates that v_i is deleted from the sequence.

- $\partial_1[a, b] = b - a.$
- $\partial_2[a, b, c] = [b, c] - [a, c] + [a, b] = [b, c] + [c, a] + [a, b].$
- $\partial_3[a, b, c, d] = [b, c, d] - [a, c, d] + [a, b, d] - [a, b, c].$

Boundary Examples

- $\partial_1[a, b] = b - a.$
- $\partial_2[a, b, c] = [b, c] - [a, c] + [a, b] = [b, c] + [c, a] + [a, b].$
- $\partial_3[a, b, c, d] = [b, c, d] - [a, c, d] + [a, b, d] - [a, b, c].$
- $\partial_1\partial_2[a, b, c] = [c] - [b] - [c] + [a] + [b] - [a] = 0.$



Boundary Theorem

- (Theorem) $\partial_{k-1}\partial_k = 0$, for all k .

- Proof:

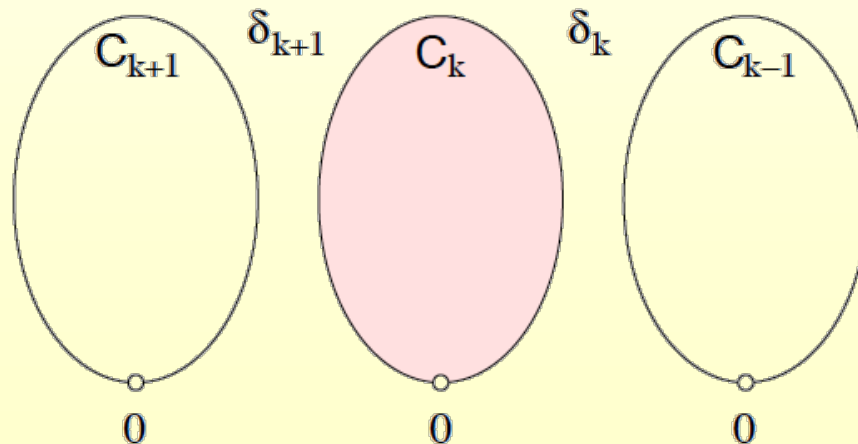
$$\begin{aligned}\partial_{k-1}\partial_k[v_0, v_1, \dots, v_k] &= \\ &= \partial_{k-1} \sum_i (-1)^i [v_0, v_1, \dots, \hat{v}_i, \dots, v_k] \\ &= \sum_{j < i} (-1)^i (-1)^j [v_0, \dots, \hat{v}_j, \dots, \hat{v}_i, \dots, v_k] \\ &\quad + \sum_{j > i} (-1)^i (-1)^{j-1} [v_0, \dots, \hat{v}_i, \dots, \hat{v}_j, \dots, v_k] \\ &= 0,\end{aligned}$$

as switching i and j in the second sum negates the first sum.

Chain Complex

- The boundary operator connects the chain groups into a **chain complex** C_* :

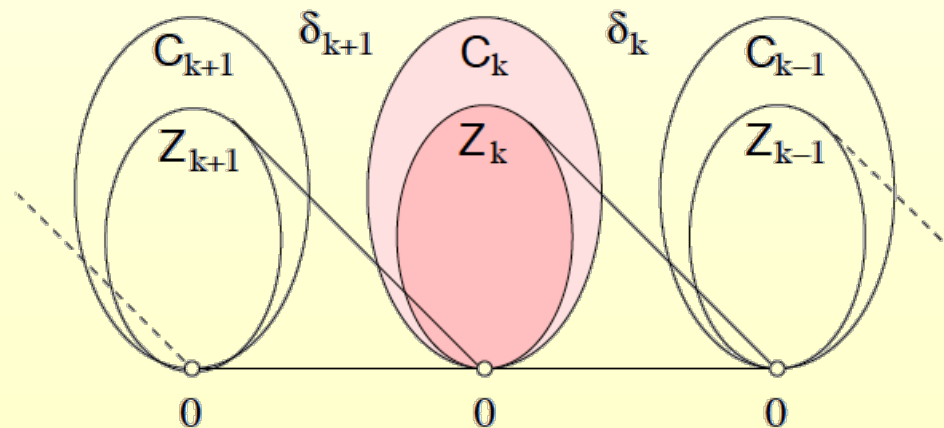
$$\dots \rightarrow C_{k+1} \xrightarrow{\partial_{k+1}} C_k \xrightarrow{\partial_k} C_{k-1} \rightarrow \dots$$



Cycle Group

- Let c be a k -chain
- If it has no boundary, it is a k -cycle (zycle?)
- $\partial_k c = \emptyset$, so $c \in \ker \partial_k$
- The k th cycle group is

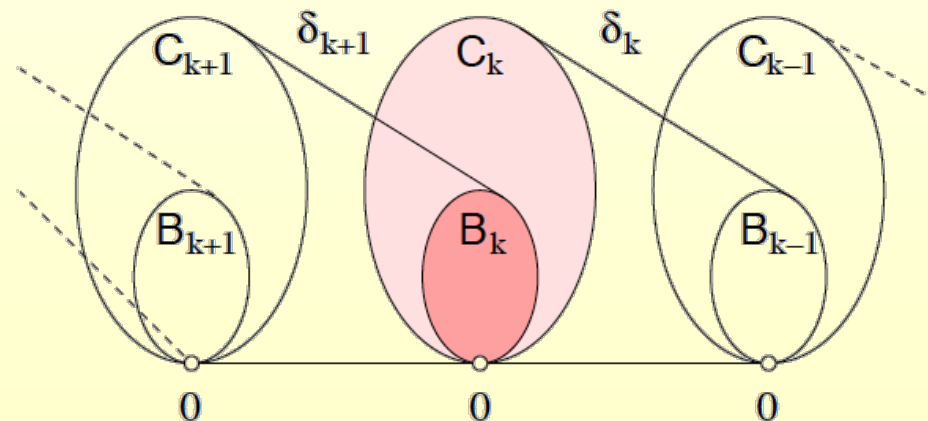
$$Z_k = \ker \partial_k = \{c \in C_k \mid \partial_k c = \emptyset\}.$$



Boundary Group

- Let b be a k -chain
- If b is a boundary of something, it is a k -boundary.
- The k th boundary group is

$$\mathbf{B}_k = \text{im } \partial_{k+1} = \{c \in \mathbf{C}_k \mid \exists d \in \mathbf{C}_{k+1} : c = \partial_{k+1}d\}.$$



Boundaries are Cycles!

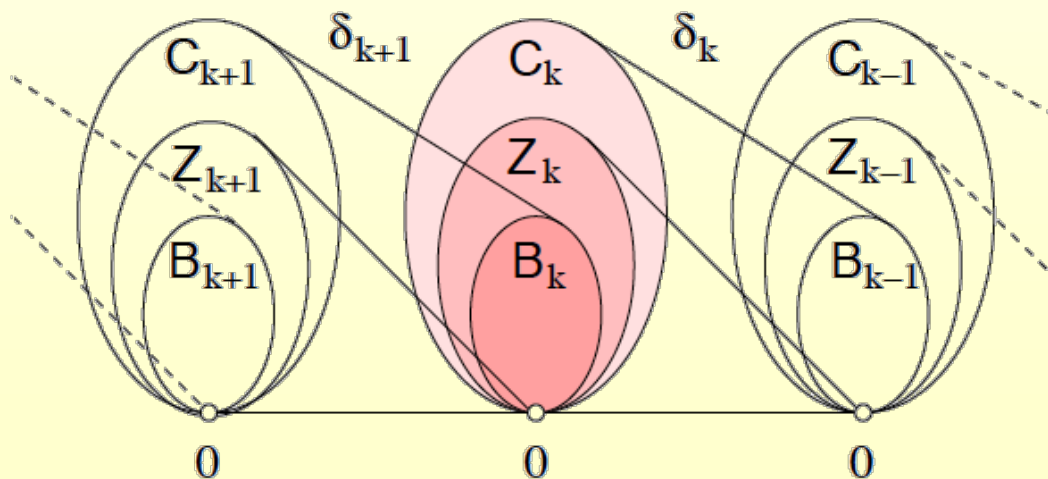
- Let b be a k -boundary.
- Then, $\exists c \in \mathbf{C}_{k+1}$, such that $b = \partial_{k+1}c$.
- What is the boundary of b ?

$$\partial_k b = \partial_k \partial_{k+1} c = \emptyset,$$

$\bullet \mathbf{B}_k \subseteq \mathbf{Z}_k \subseteq \mathbf{C}_k$

by the boundary theorem.

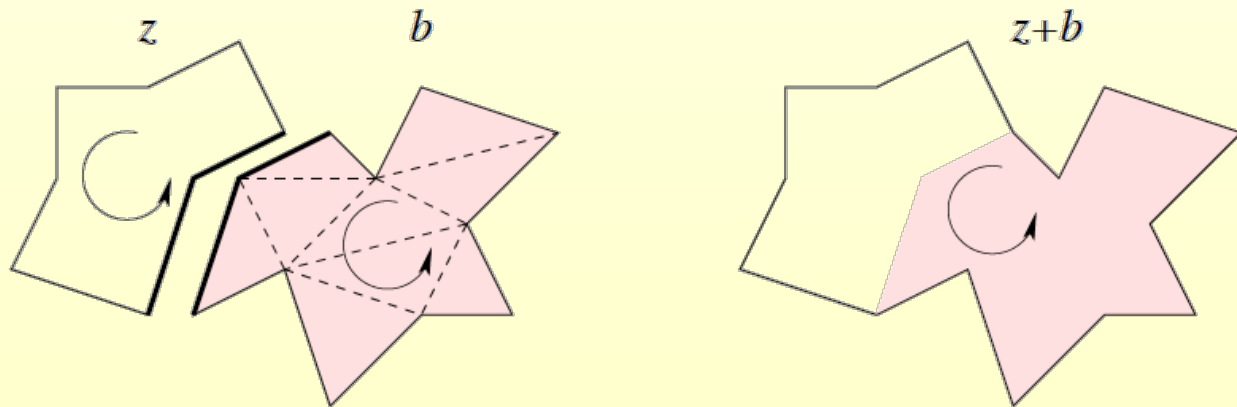
- That is, every boundary is a cycle!



Nesting behavior

Equivalent Cycles

- z is a k -cycle
- b is a k -boundary
- We would like to have $z + b$ be equivalent to z Cosets!
- That is, if $z_1 - z_2 = b$ where b is a boundary, then $z_1 \sim z_2$
- Any boundary would do!

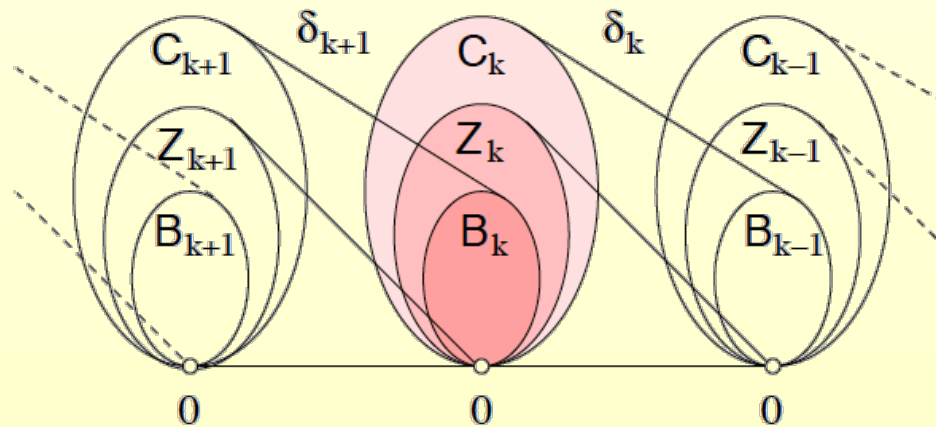


Simplicial Homology

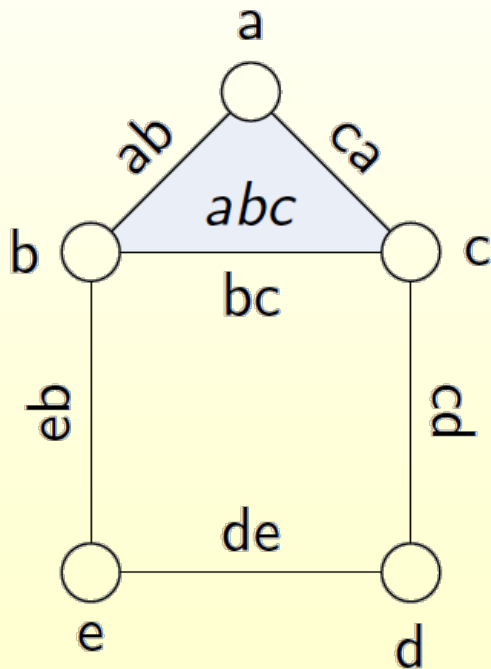
- The k th homology group is

$$H_k = Z_k / B_k = \ker \partial_k / \text{im } \partial_{k+1}.$$

- If $z_1 = z_2 + B_k$, $z_1, z_2 \in Z_k$, we say z_1 and z_2 are **homologous**
- $z_1 \sim z_2$.



To Repeat



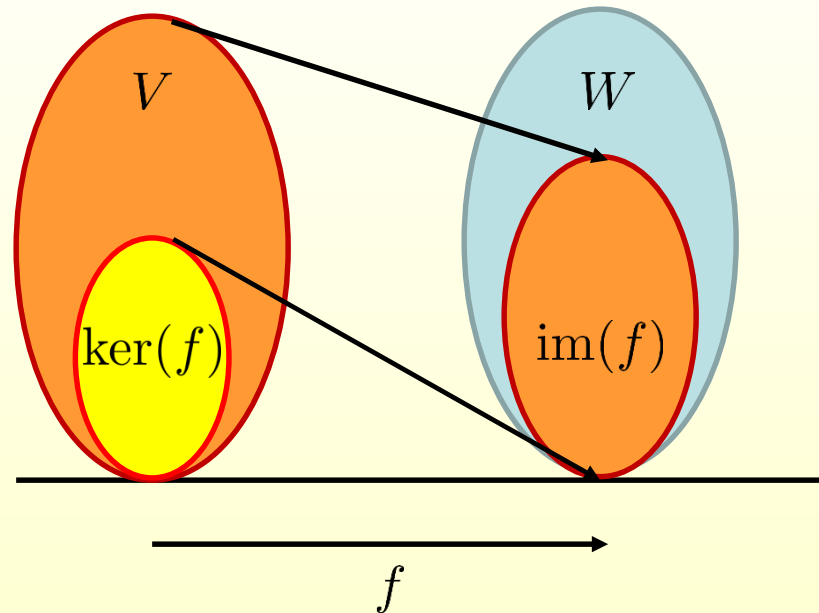
In other words..

- ▶ The kernel (null space) of ∂_k is the vector space of cycles in dimension k .
- ▶ The image of ∂_k is the subspace of boundary cycles in dimension $k - 1$.

Homology of a space X is the quotient:

$$H_k(X) = \ker(\partial_k) / \text{im}(\partial_{k+1})$$

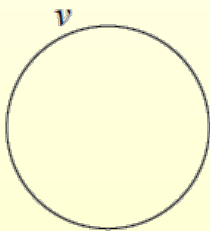
In Vector Spaces



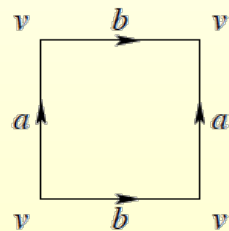
$$V \approx \ker(f) \oplus \text{im}(f)$$

Homology of 2-Manifolds

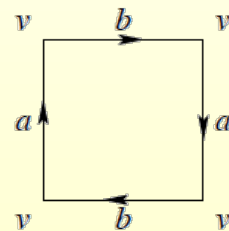
2-manifold	H_0	H_1	H_2
sphere	\mathbb{Z}	$\{0\}$	\mathbb{Z}
torus	\mathbb{Z}	$\mathbb{Z} \times \mathbb{Z}$	\mathbb{Z}
projective plane	\mathbb{Z}	\mathbb{Z}_2	$\{0\}$
Klein bottle	\mathbb{Z}	$\mathbb{Z} \times \mathbb{Z}_2$	$\{0\}$



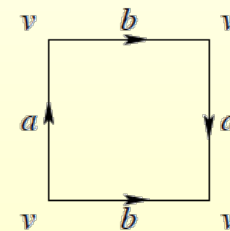
(a) Sphere



(b) Torus

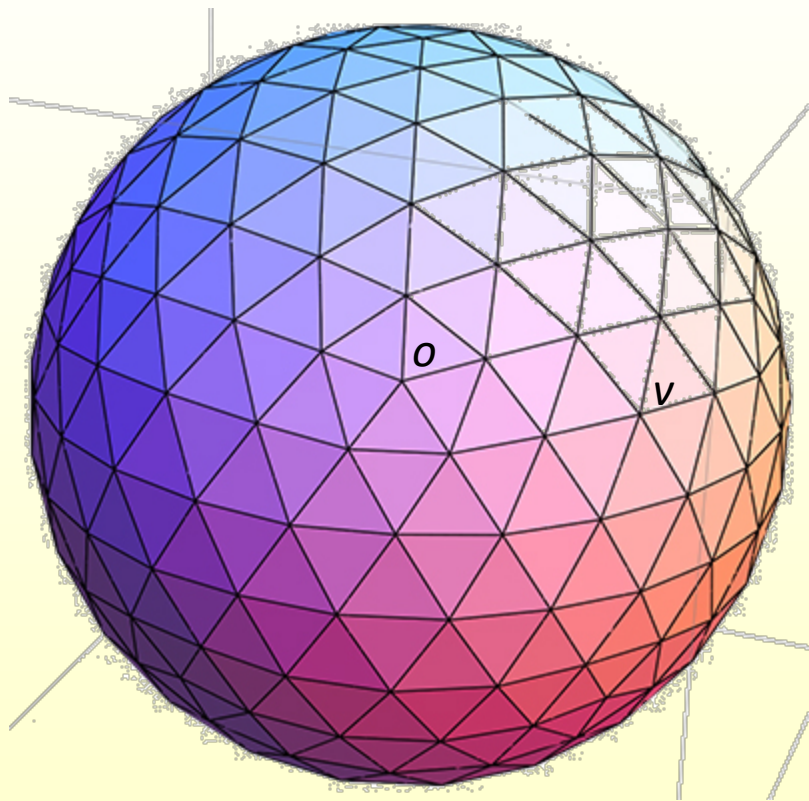


(c) Projective plane



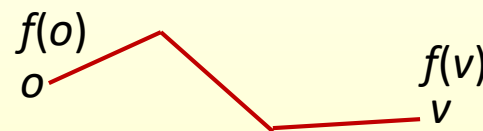
(d) Klein bottle

Why is $H_0(S^2) = \mathbb{Z}$?



A 0-chain is an assignment of integers $f(v)$ to each mesh vertex v .

The mesh is connected, so all pairs of vertices are homologous.



So we can pick a specific vertex o and use paths from o to each other v to “cancel” f at v .



Homology Groups

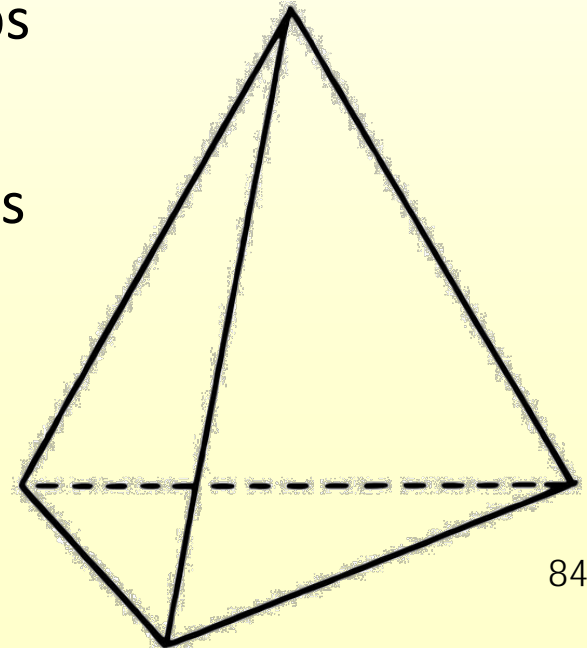
- Homology groups are finitely generated abelian.
- (Theorem) Every finitely generated abelian group is isomorphic to product of cyclic groups of the form

$$\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \dots \times \mathbb{Z}_{m_r} \times \mathbb{Z} \times \mathbb{Z} \times \dots \times \mathbb{Z},$$

- The k th Betti number β_k of a simplicial complex K is $\beta_k = \beta(\mathbf{H}_k)$, the rank of the free part of \mathbf{H}_k .
- Torsion coefficients
 - Alexander Duality:
 - β_0 measures the number of components of the complex.
 - β_1 is the rank of a basis for the tunnels.
 - β_2 counts the number of voids in the complex.

Betti Numbers β_i

- ◆ Ranks of the free part of homology groups H_i
- ◆ β_0 counts the number of connected components
- ◆ β_1 counts the number of independent loops
- ◆ β_2 counts the number of independent voids
- ◆ ...



Invariance of Homology Groups

- (Hauptvermutung) Any two triangulations of a topological space have a common refinement (Poincaré 1904)
 - True for polyhedra of dimension ≤ 2 (Papakyriakopoulos 1943)
 - True for 3-manifolds (Moïse 1953)
 - False in dimensions ≥ 6 (Milnor 1961)
 - False for manifolds of dimension ≥ 5 (Kirby and Siebenmann 1969)
- Singular homology

Euler Revisited

- Let K be a simplicial complex and $s_i = |\{\sigma \in K \mid \dim \sigma = i\}|$. The Euler characteristic $\chi(K)$ is

$$\chi(K) = \sum_{i=0}^{\dim K} (-1)^i s_i = \sum_{\sigma \in K - \{\emptyset\}} (-1)^{\dim \sigma}.$$

- We have new language!
- Let \mathbf{C}_* be the chain complex on K
- $\text{rank}(\mathbf{C}_i) = |\{\sigma \in K \mid \dim \sigma = i\}|$ ($= n_i = z_i + b_{i-1}$)
- $\chi(K) = \chi(\mathbf{C}_*) = \sum_i (-1)^i \text{rank}(\mathbf{C}_i)$.

$$\sum_i (-1)^i (z_i + b_{i-1}) = \sum_i (-1)^i (z_i - b_i)$$

Euler - Poincaré

- Homology functors H_*
- $H_*(C_*)$ is a chain complex:

$$\dots \rightarrow H_{k+1} \xrightarrow{\partial_{k+1}} H_k \xrightarrow{\partial_k} H_{k-1} \rightarrow \dots$$

- What is its Euler characteristic?
- (Theorem) $\chi(K) = \chi(C_*) = \chi(H_*(C_*))$.

- $\sum_i (-1)^i s_i = \sum_i (-1)^i \text{rank}(H_i) = \sum_i (-1)^i \beta_i$

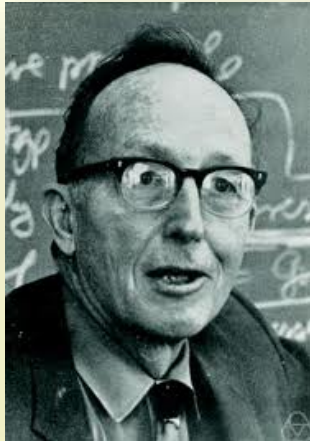
- Sphere: $2 = 1 - 0 + 1$

- Torus: $0 = 1 - 2 + 1$ $\sum_i (-1)^i (z_i + b_{i-1}) = \sum_i (-1)^i (z_i - b_i)$

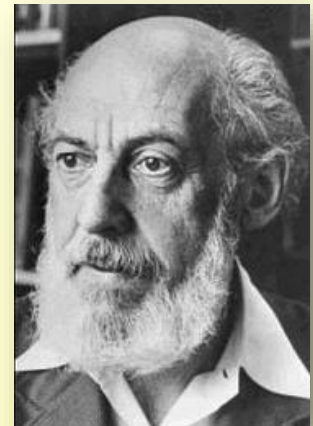
Homological Algebra: Functors and Categories



Henri Cartan



Saunders MacLane



Samuel Eilenberg

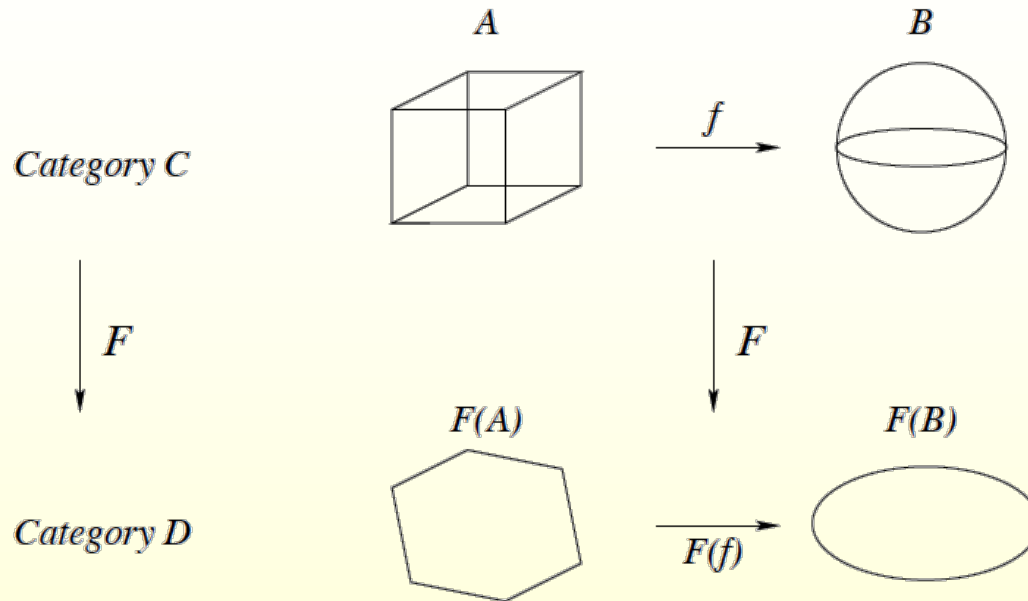
Categories

- A collection $\text{Ob}(\mathcal{C})$ of **objects**
- Sets $\text{Mor}(X, Y)$ of **morphisms** for each pair $X, Y \in \text{Ob}(\mathcal{C})$
- An identity morphism $1 = 1_X \in \text{Mor}(X, X)$ for each X .
- a composition of morphisms function
 - $\circ : \text{Mor}(X, Y) \times \text{Mor}(Y, Z) \rightarrow \text{Mor}(X, Z)$ for each triple $X, Y, Z \in \text{Ob}(\mathcal{C})$, satisfying $f \circ 1 = 1 \circ f = f$, and $(f \circ g) \circ h = f \circ (g \circ h)$.
- A **category** \mathcal{C}

Example Categories

category	morphisms
sets	arbitrary functions
groups	homomorphisms
topological spaces	continuous maps
topological spaces	homotopy classes of maps

Functors



- $X \in \mathcal{C}, F(X) \in \mathcal{D},$
- $f \in \text{Mor}(X, Y), F(f) \in \text{Mor}(F(X), F(Y))$
- $F(1) = 1$ and $F(f \circ g) = F(f) \circ F(g)$
- F is a **(covariant) functor**

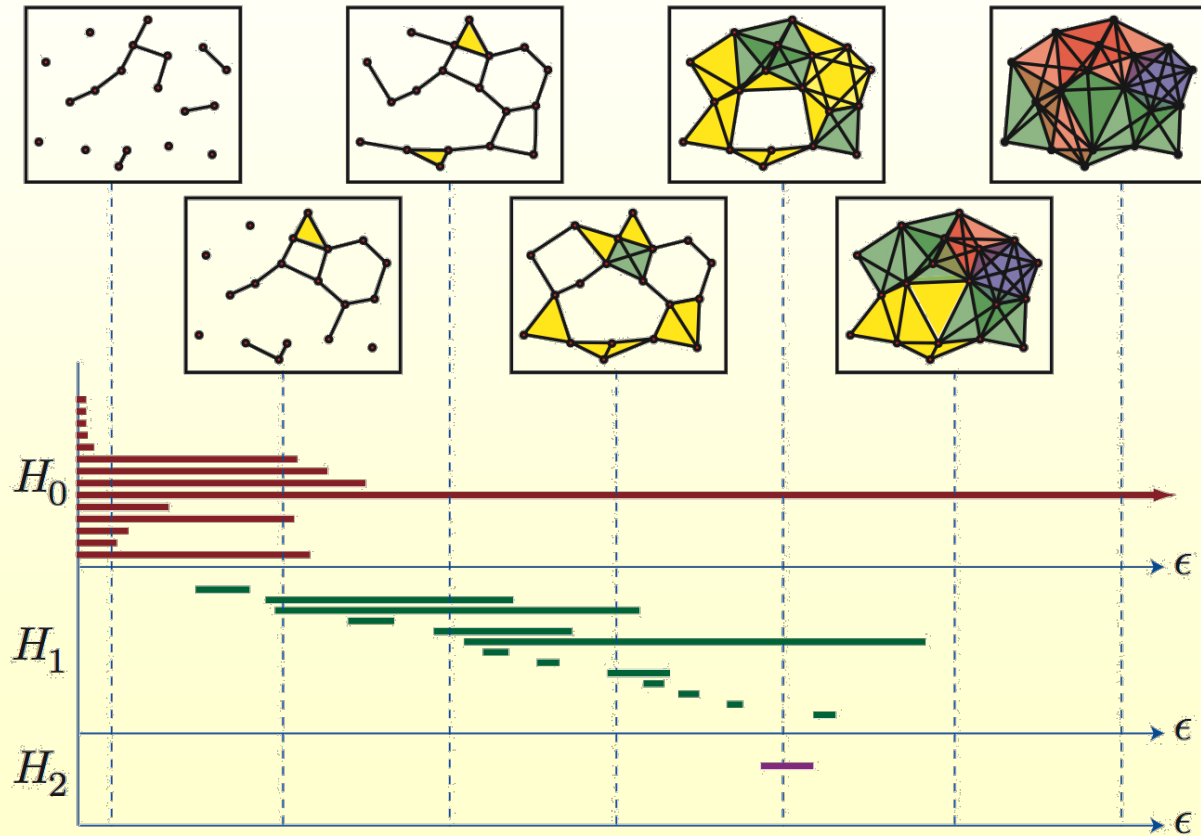
Functoriality

transformation of input \Rightarrow transformation of output
Specifically, this is a commutative diagram:

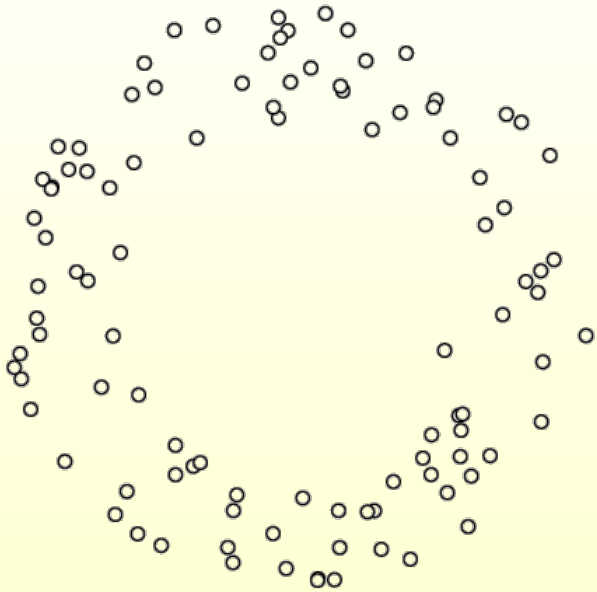
$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ H_* \downarrow & & \downarrow H_* \\ H_*(X) & \xrightarrow{H_*(f)} & H_*(Y) \end{array}$$

Moral: Invariants are not artifacts of arbitrary choices!

Persistent Homology

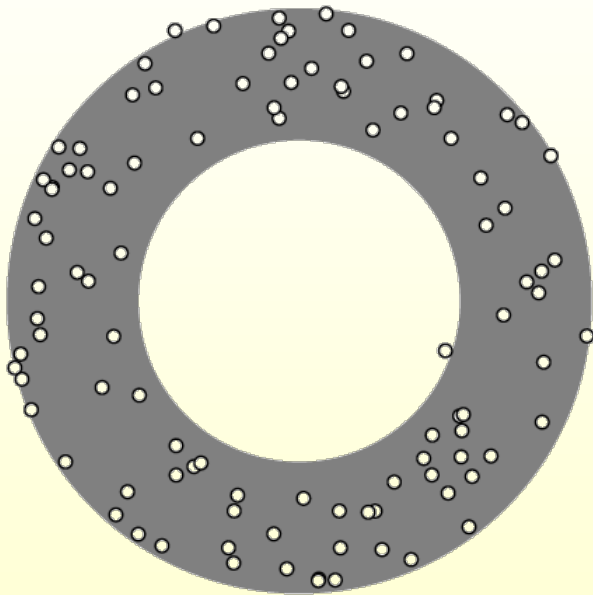


Sampled Data Has “Shape”



2-dimensional
Approximates annulus

Sampled Data Has “Shape”



2-dimensional

Approximates annulus

Topological features of annulus:

1 component ($\beta_0 = 1$)

1 loop ($\beta_1 = 1$)

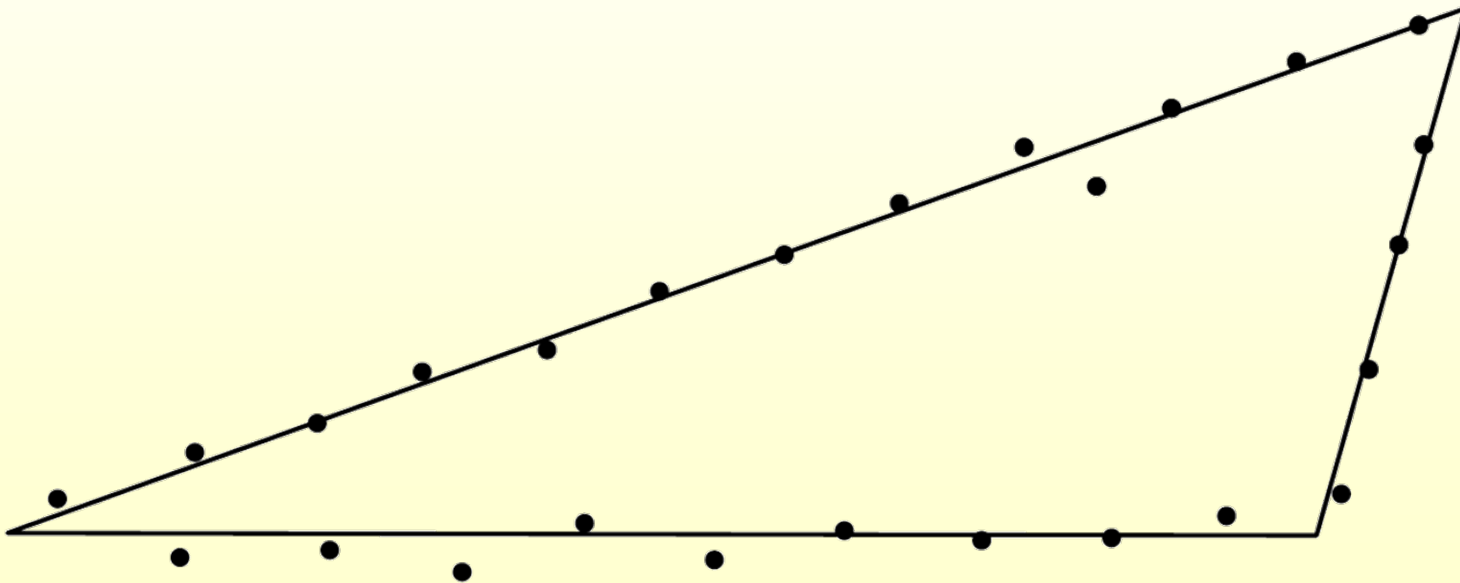
Goal: Recover *topology* of annulus from point cloud

We do so by building various complexes on the point cloud

Filtrations

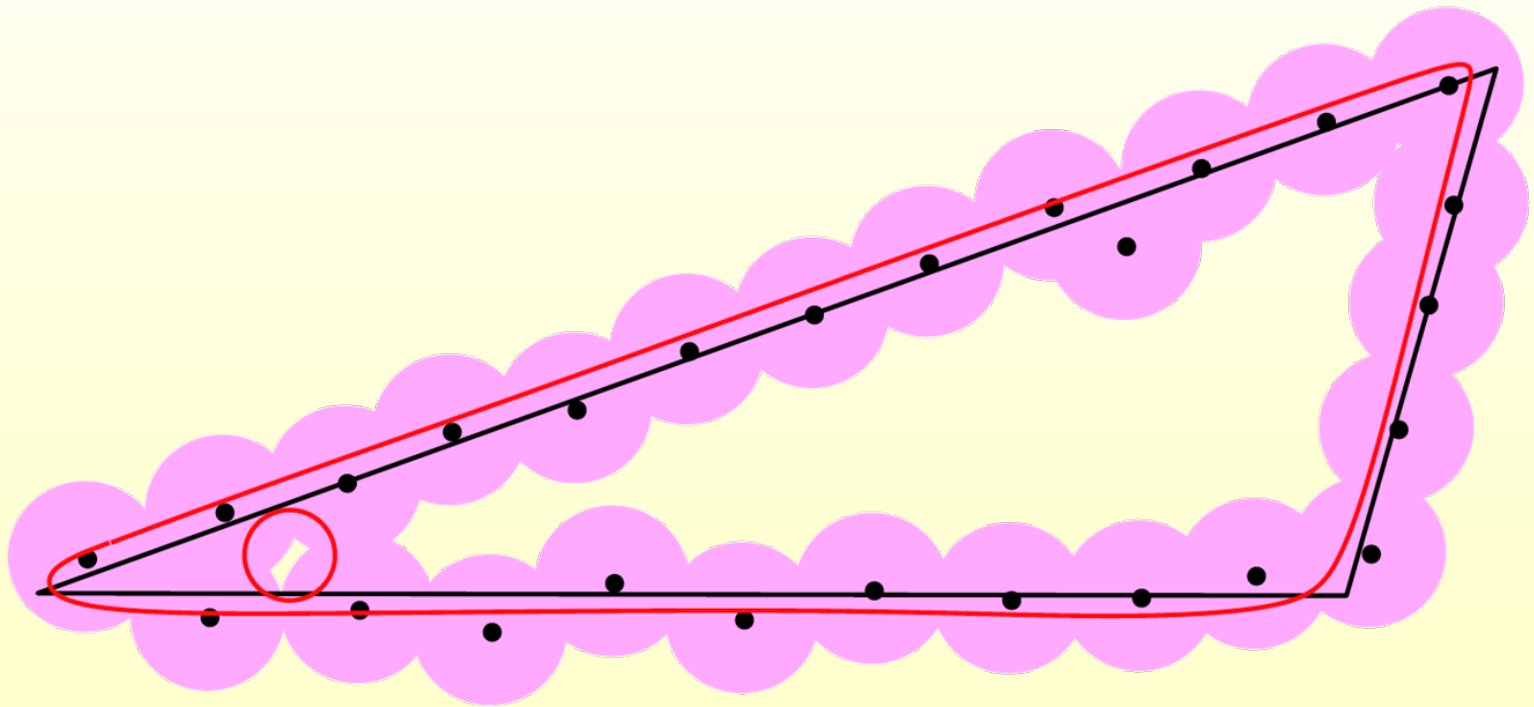
How to Choose ϵ ?

- ◆ How to determine the topology of the underlying space from a point cloud approximation?



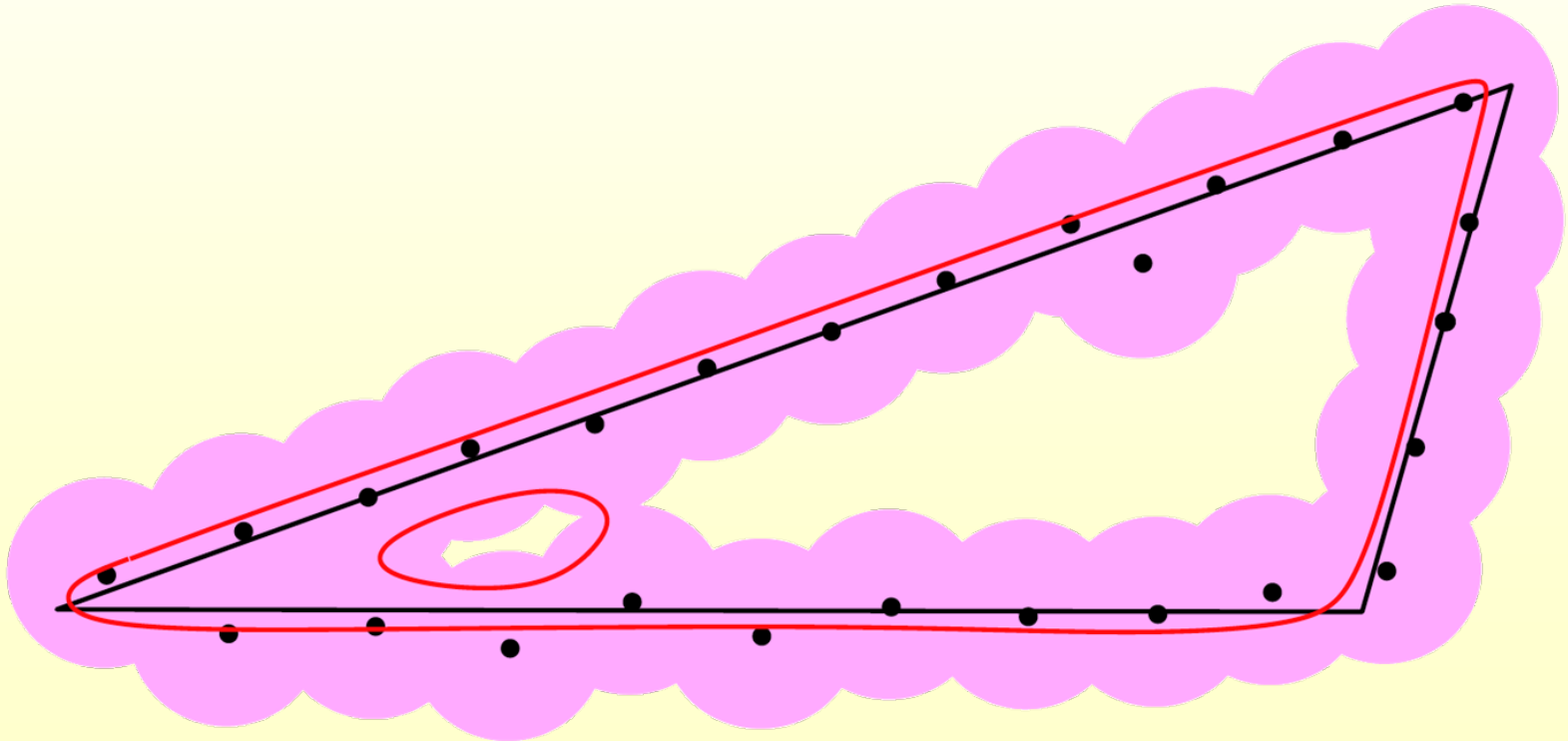
How to Choose ϵ ?

- ◆ How to determine the topology of the underlying space from a point cloud approximation?



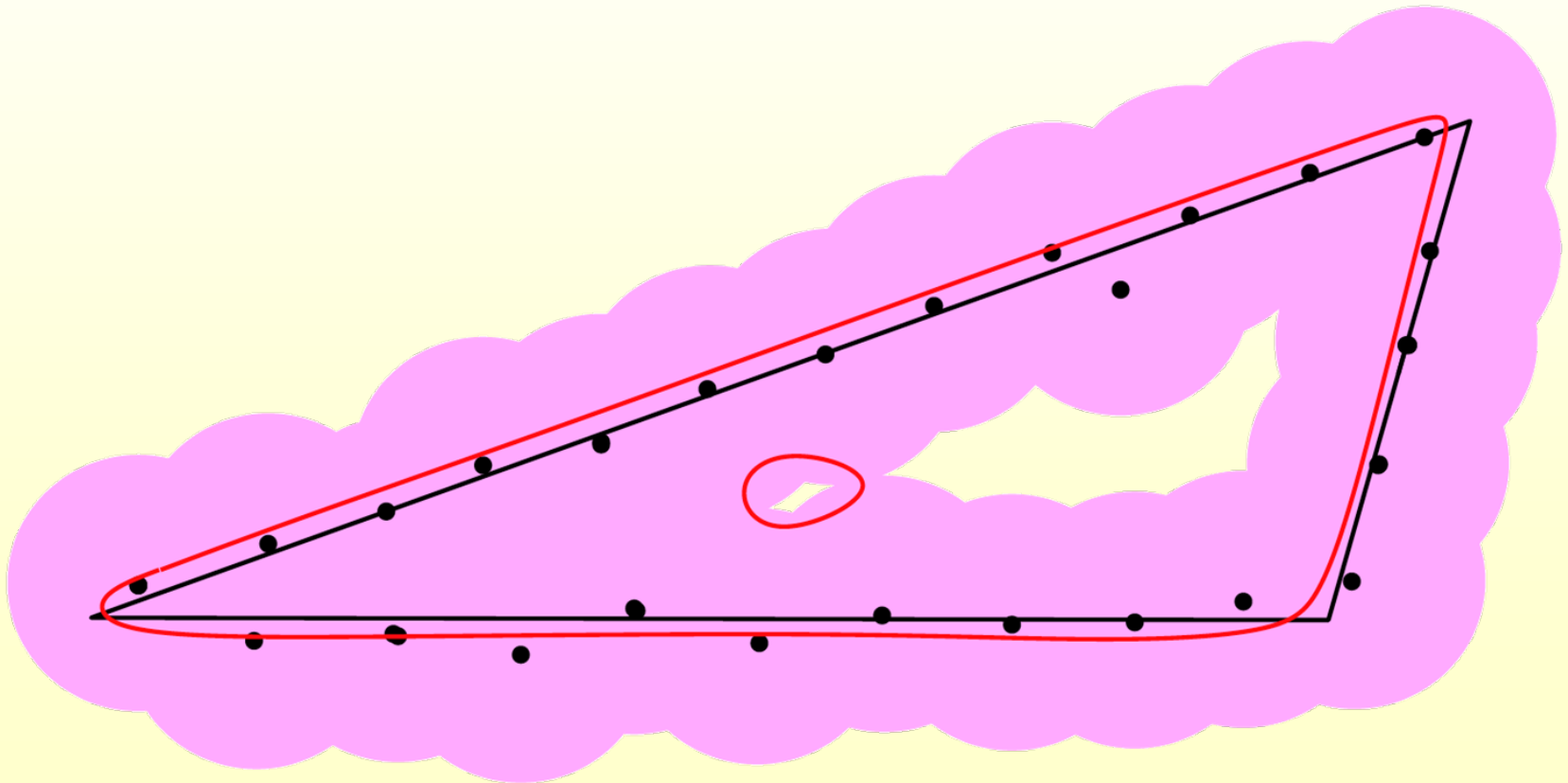
How to Choose ϵ ?

- ◆ How to determine the topology of the underlying space from a point cloud approximation?

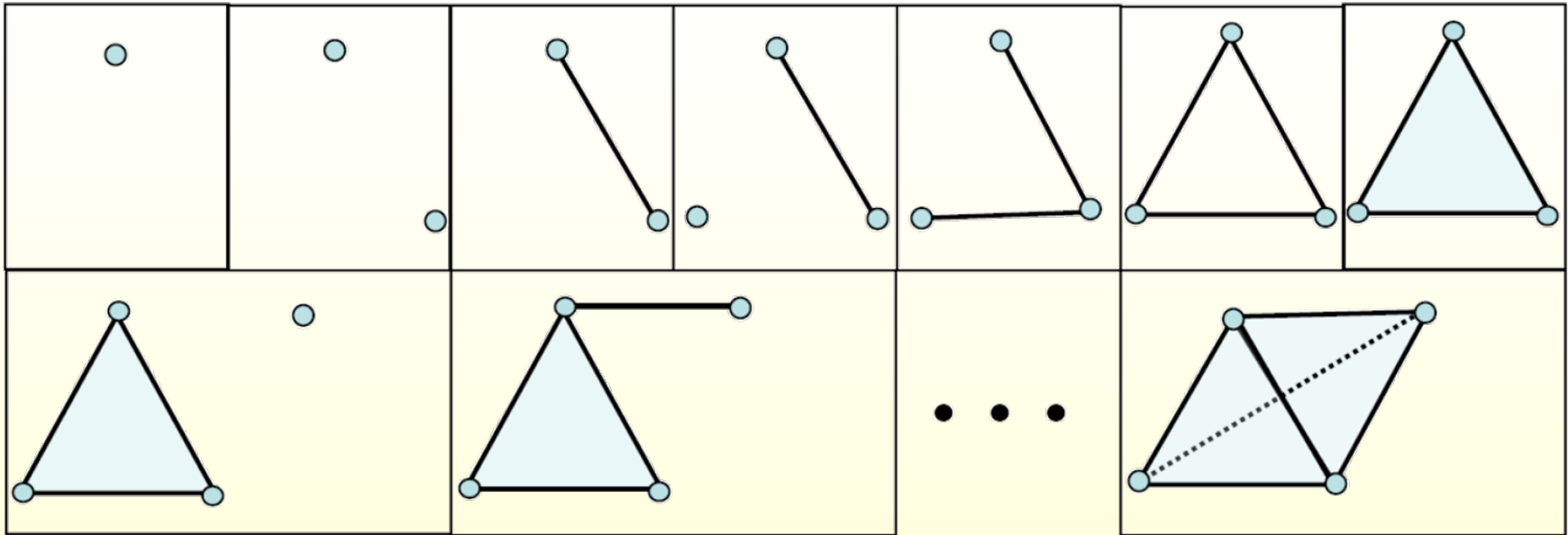


How to Choose ϵ ?

- ◆ How to determine the topology of the underlying space from a point cloud approximation?



Filtrations

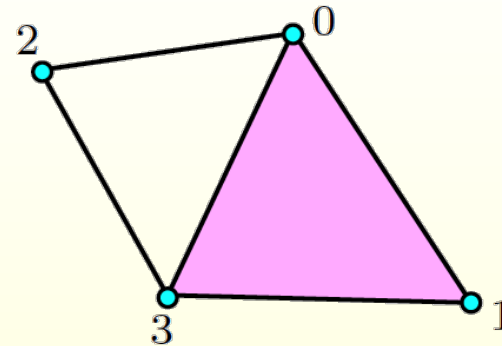
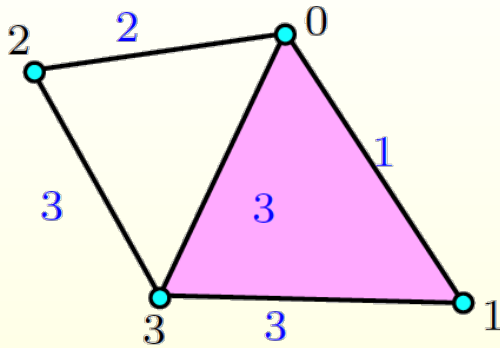


A **filtration** of a (finite) simplicial complex K is a sequence of subcomplexes such that

- i) $\emptyset = K^0 \subset K^1 \subset \dots \subset K^m = K$,
- ii) $K^{i+1} = K^i \cup \sigma^{i+1}$ where σ^{i+1} is a simplex of K .

Sub-simplices of a simplex must be added before the simplex!

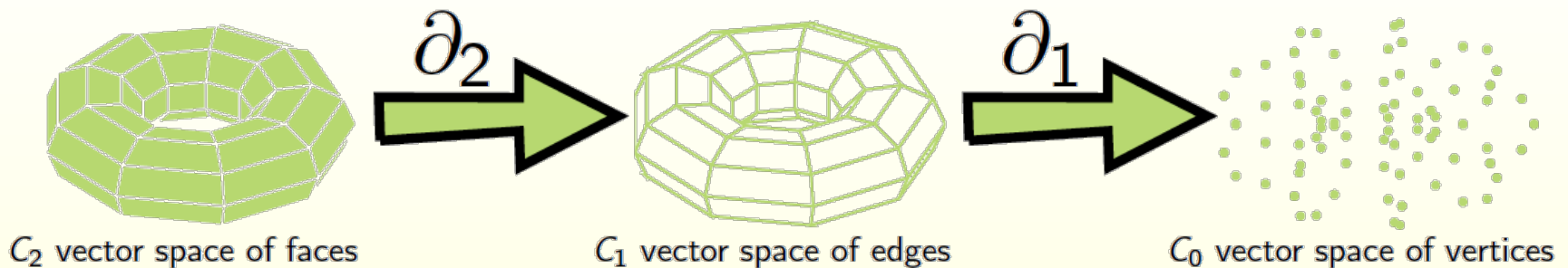
The Sub-Level Set Filtration



- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \dots, v_k] \in K$, $f(\sigma) = \max_{i=0, \dots, k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

Persistent Homology:
Do not choose an ϵ !

Standard Homology



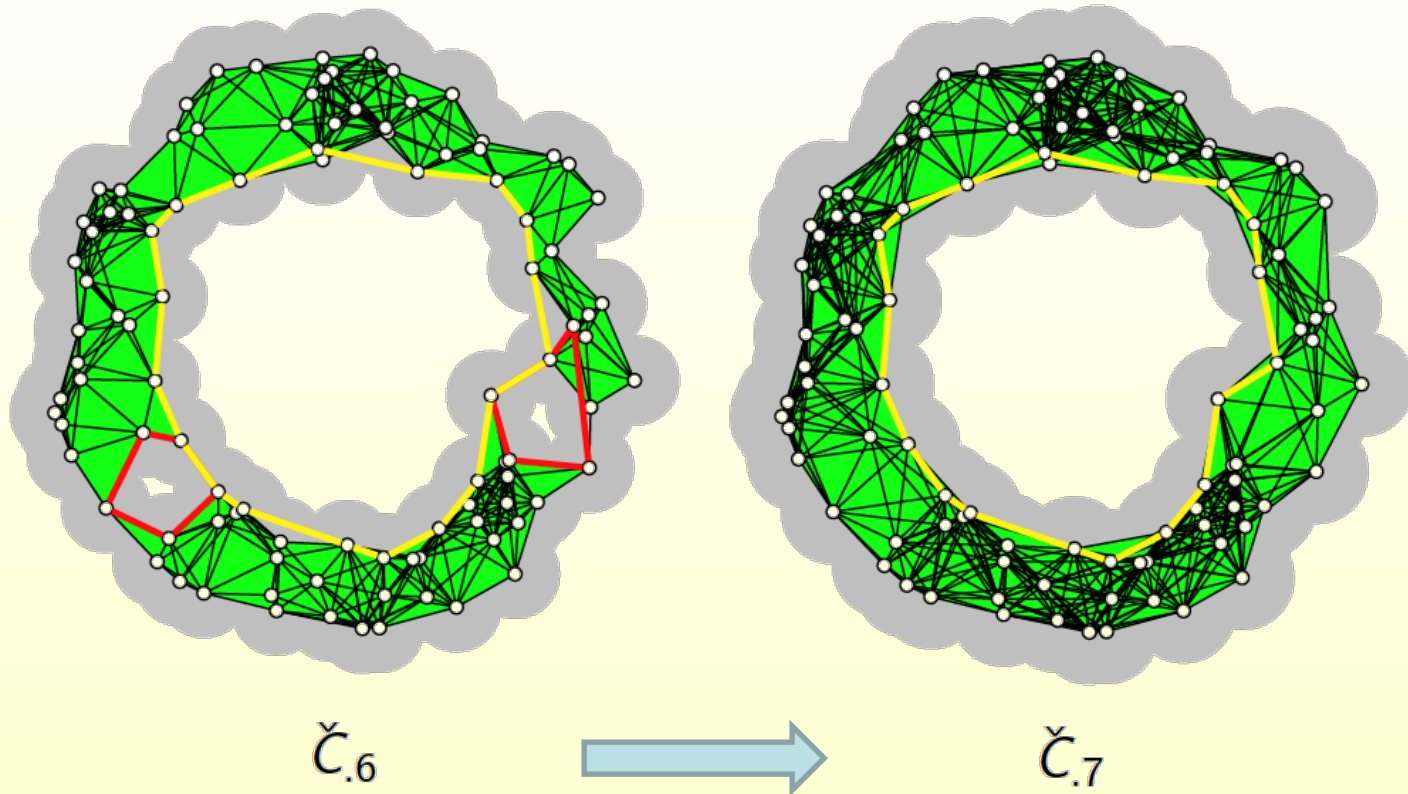
Take the linear extension of the boundary operator:

$$\partial_d([v_0, \dots, v_d]) = \sum_{i=0}^d (-1)^i [v_0, \dots, \hat{v}_i, \dots, v_d]$$

Fact: $\partial_{d-1} \circ \partial_d \equiv 0 \Rightarrow \text{Im } \partial_d \subseteq \text{ker } \partial_{d-1}$

Definition: $H_d(K) = \text{ker } \partial_d / \text{Im } \partial_{d+1}$

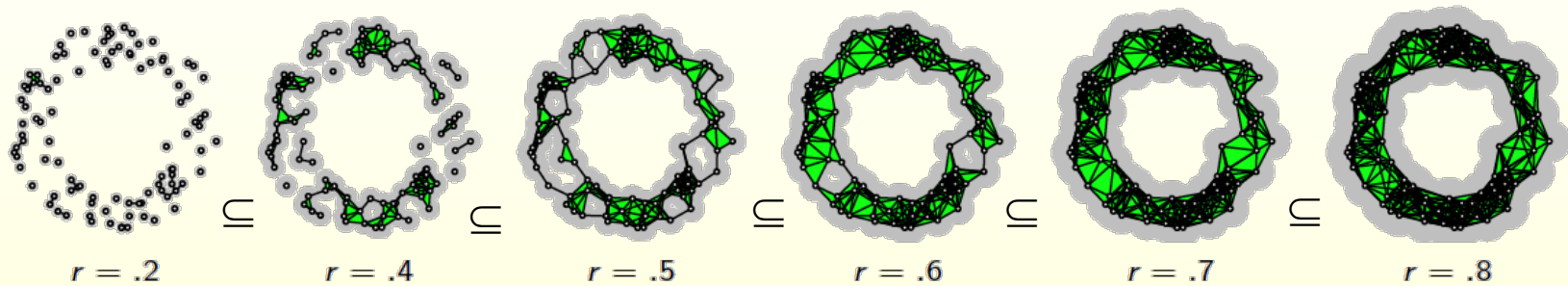
We Can Track Topological Features in a Filtration



Functoriality allows us to systematically track holes over time!

The inclusion map among the complexes translates to a homomorphism between the homology groups

Persistent Homology is Functorial Homology

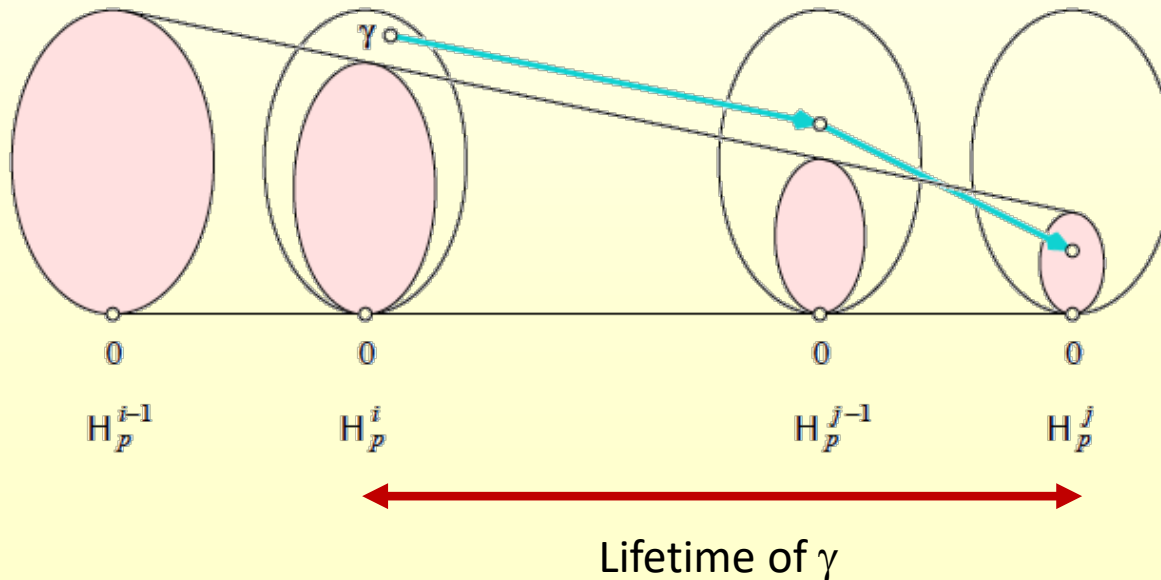
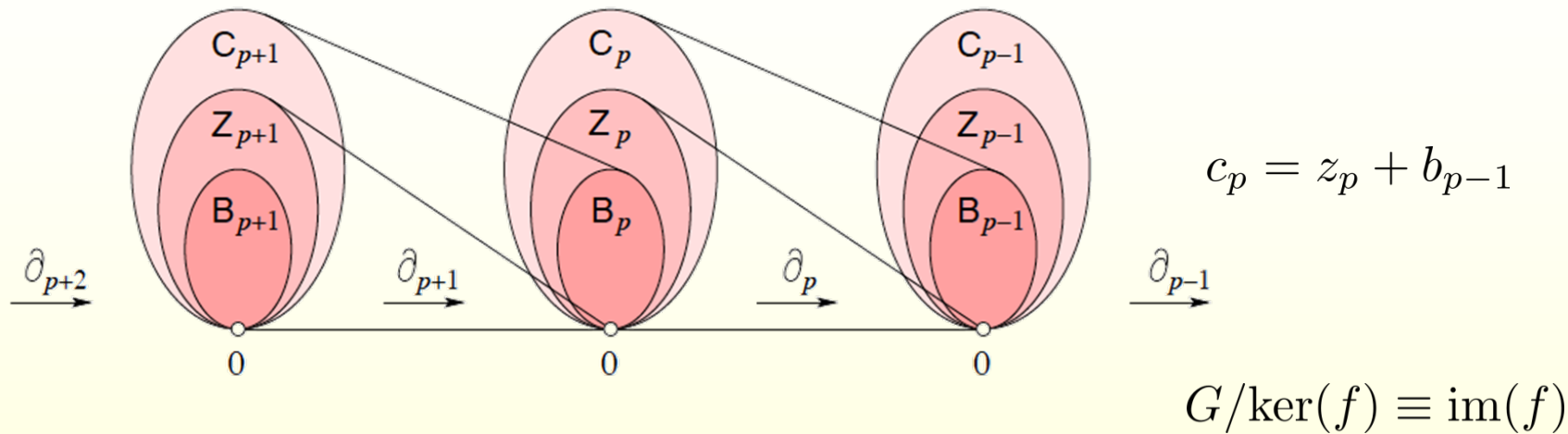


$$H_d(\check{C}_*) = \bigoplus_{\epsilon} H_d(\check{C}_\epsilon)$$

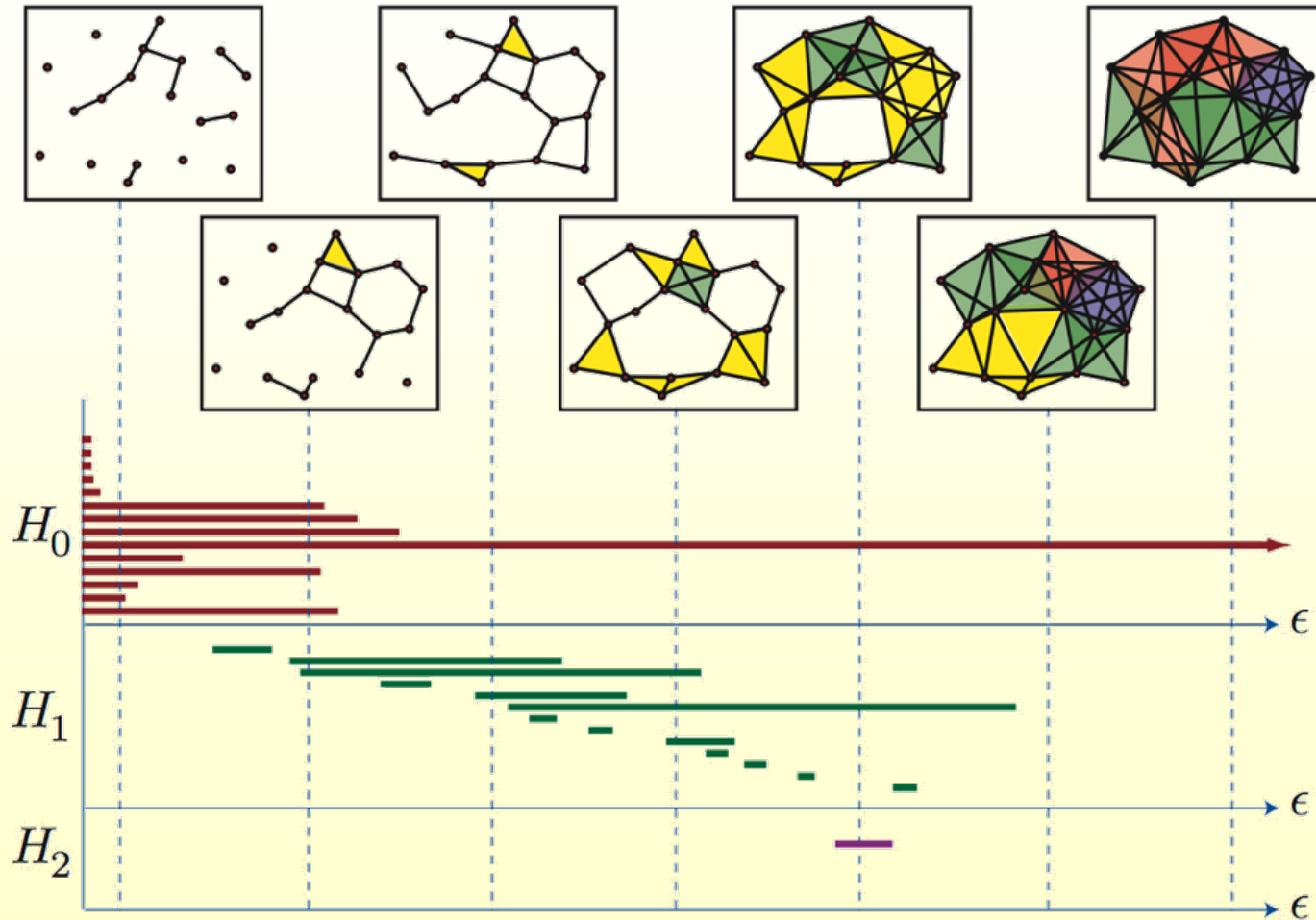
Homology of the entire filtration

Homomorphisms at the homology level allow us
to track homology classes – i.e., topological features

Persistence of Homology Classes

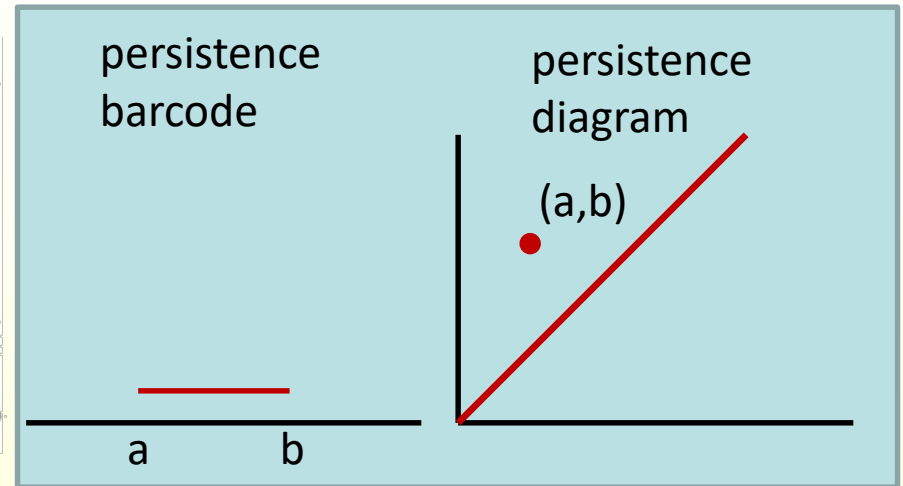
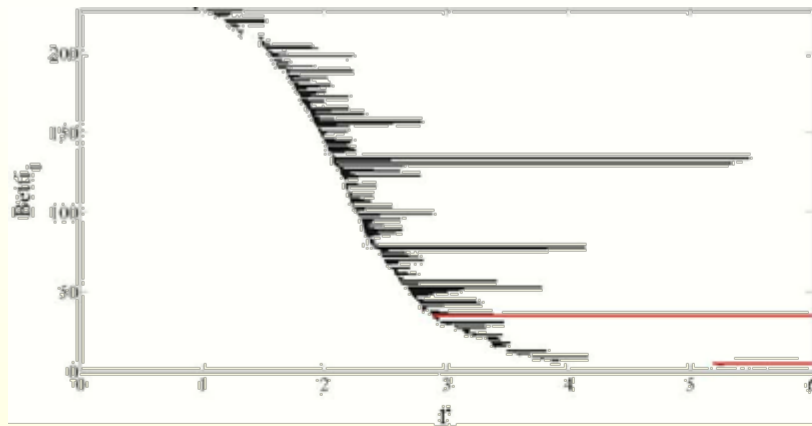


Barcodes are the Lifetimes of Topological Features

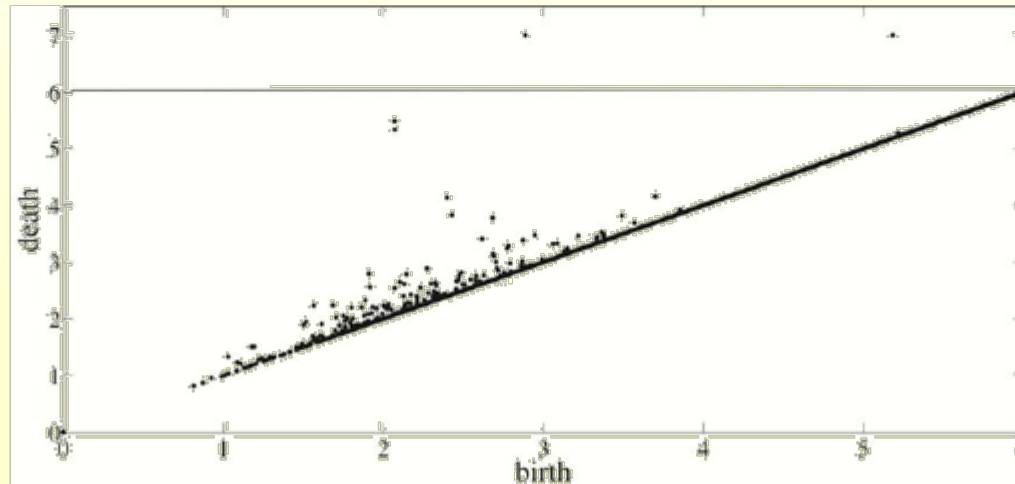


Barcodes are the output of persistent homology

Another View: Persistence Diagrams



long barcodes =
points away from
the diagonal =
robust features

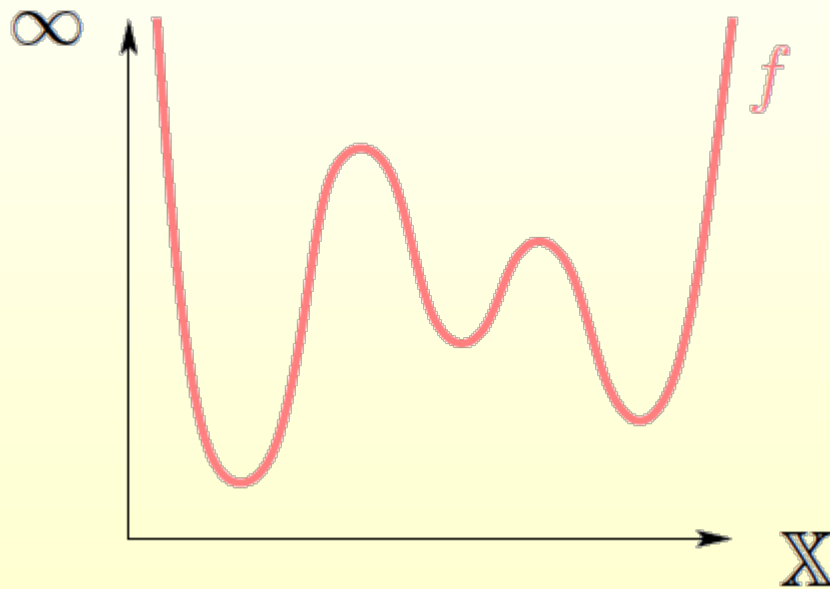


Short barcodes =
points near
the diagonal =
noise

Map 1-D intervals to points in 2-D

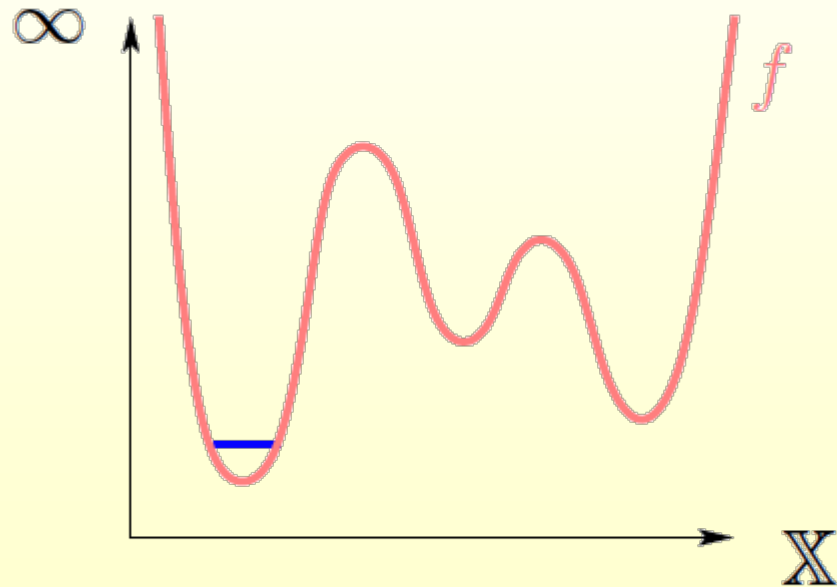
Persistence Provides a Pairing: Birth and Death of a Top Feature

- ◆ Sublevel sets of a function example



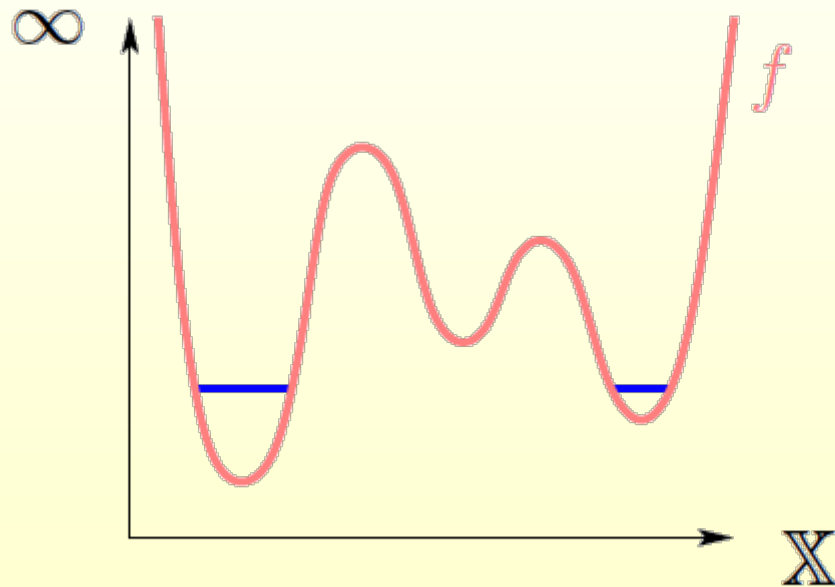
Persistence Provides a Pairing

- ◆ Sublevel sets of a function example



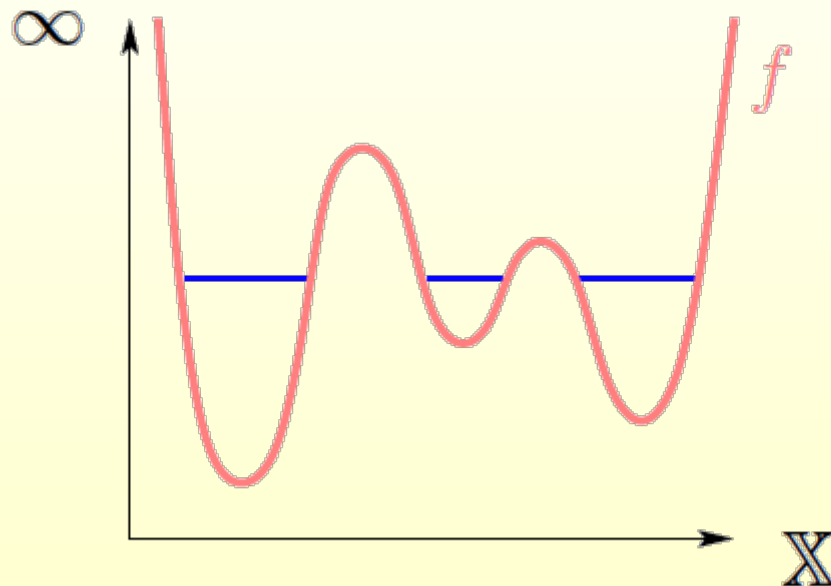
Persistence Provides a Pairing

- ◆ Sublevel sets of a function example



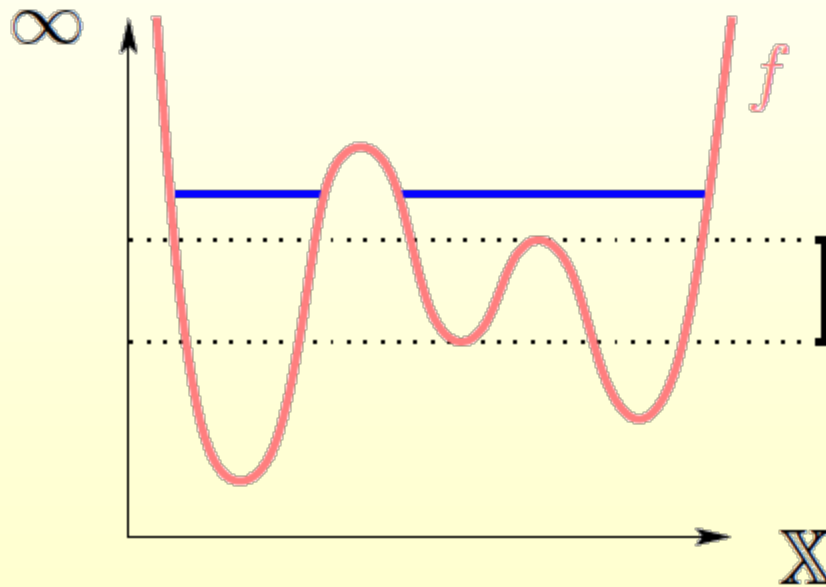
Persistence Provides a Pairing

- ◆ Sublevel sets of a function example



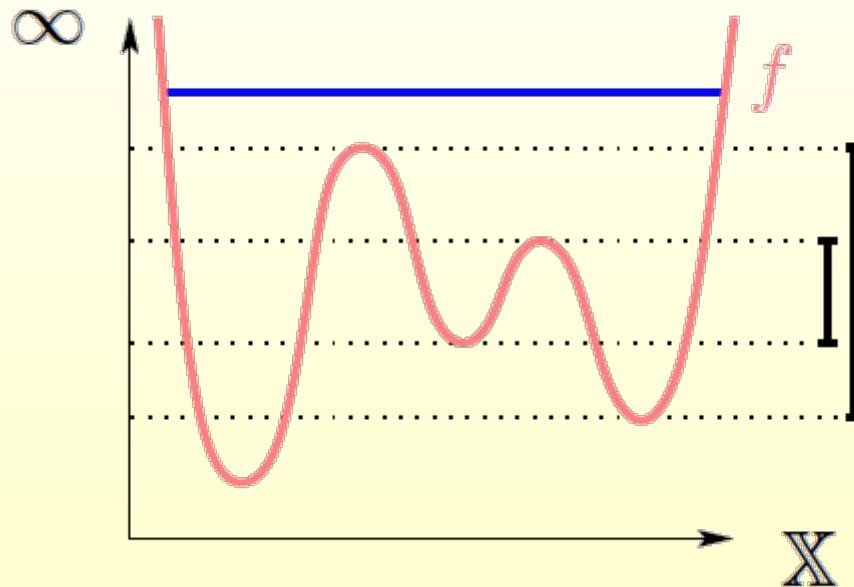
Persistence Provides a Pairing

- ◆ Sublevel sets of a function example



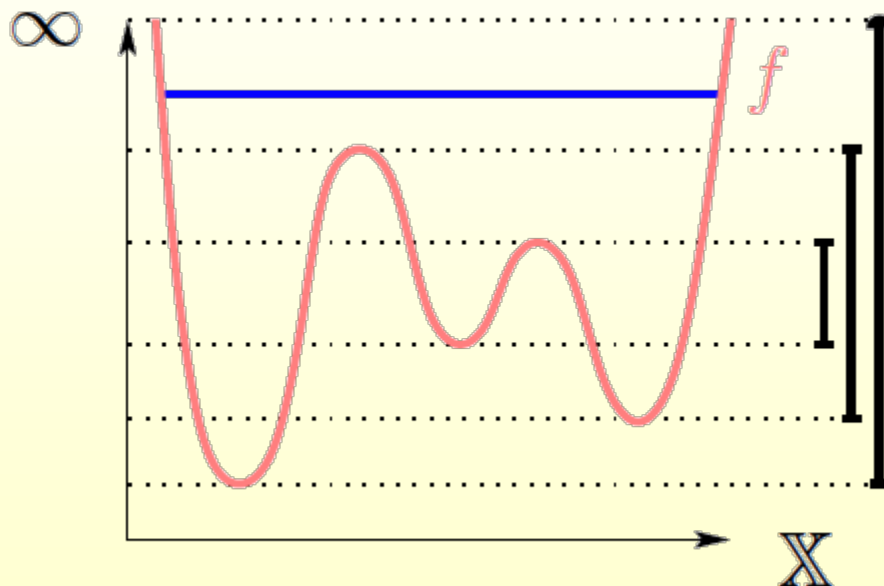
Persistence Provides a Pairing

- ◆ Sublevel sets of a function example



Persistence Provides a Pairing

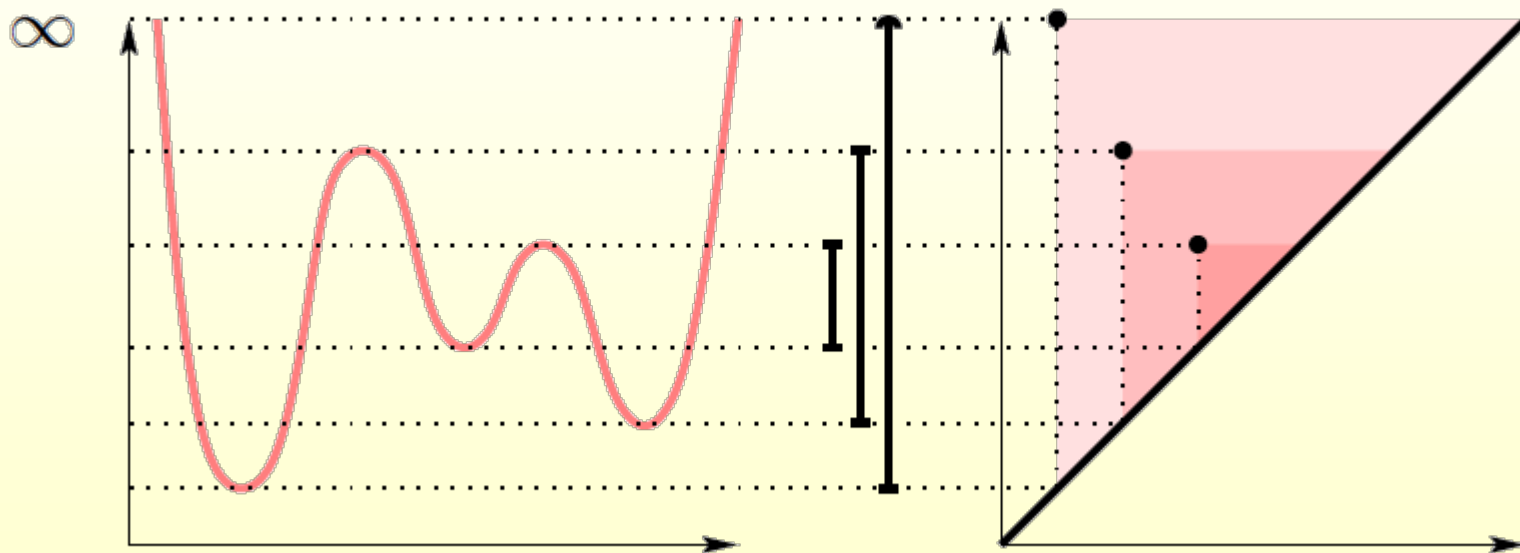
- ◆ Sublevel sets of a function example



- ◆ Pair thresholds that create components with those that destroy them

Persistence Provides a Pairing

- ◆ That pairing is the persistence diagram



- ◆ The diagonal is always included

The End