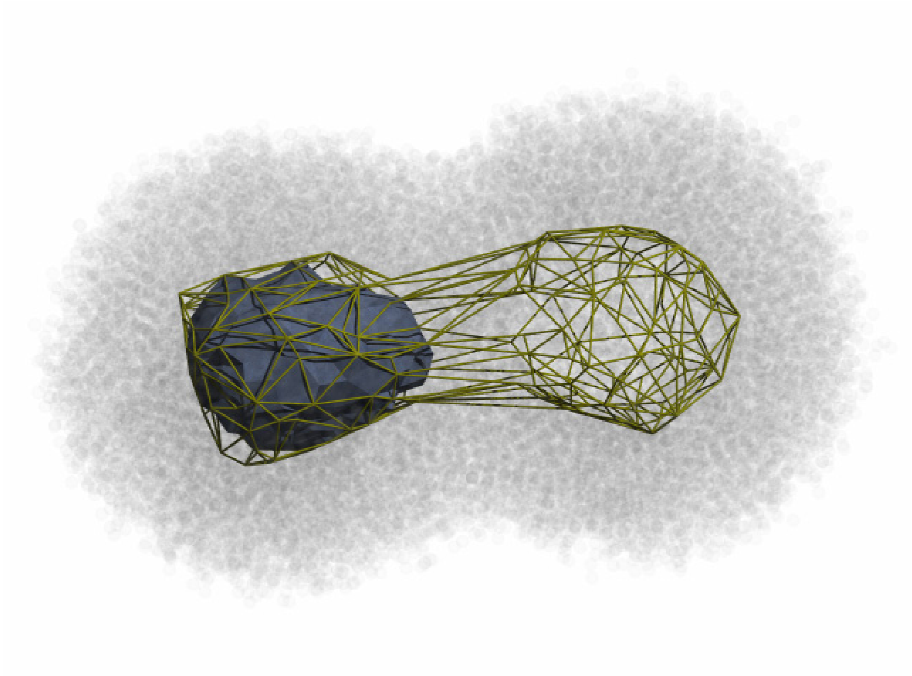
Topological Data Analysis - I

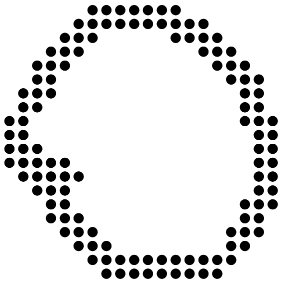


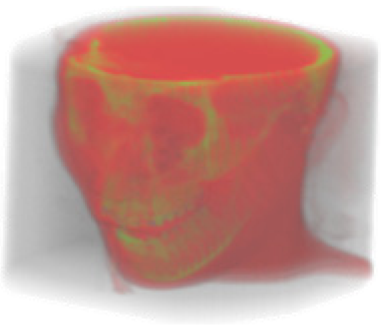
Afra Zomorodian
Department of Computer Science
Dartmouth College
September 3, 2007

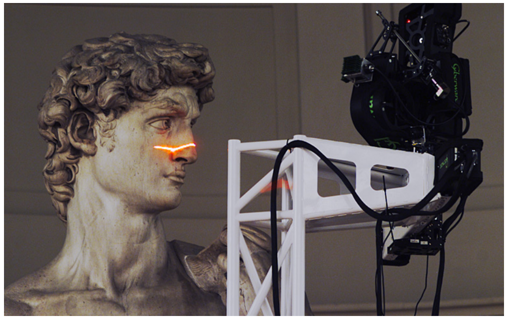
ADFOCS 2007

Acquisition

- Vision: Images (2D)
- GIS: Terrains (3D)
- Graphics: Surfaces (3D)
- Medicine: MRI (Volumetric 3D)

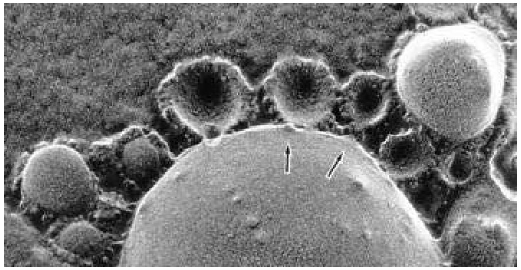


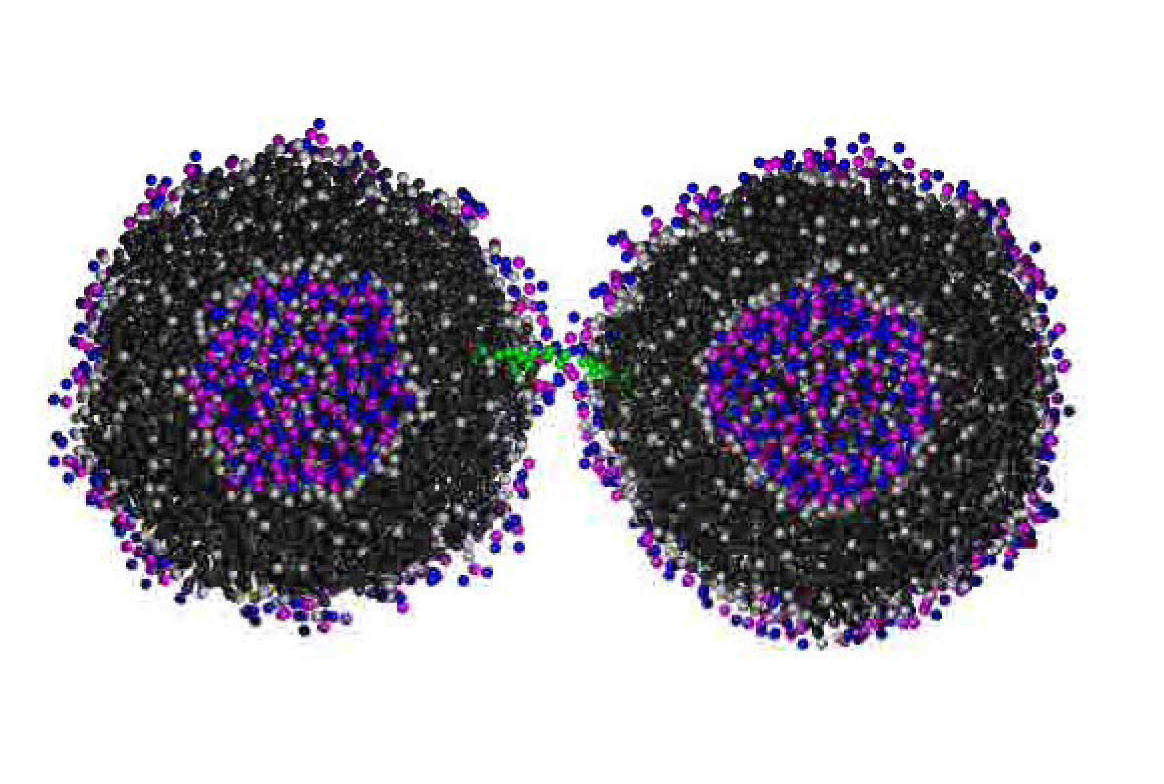




Simulation

- Folding @ Home
  - ~1M CPUs, ~200K active
  - ~200 Tflops sustained performance
  - [Kasson et al. '06]

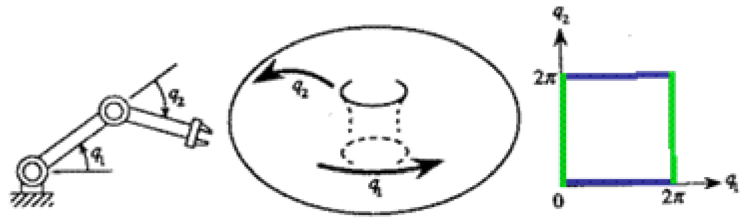




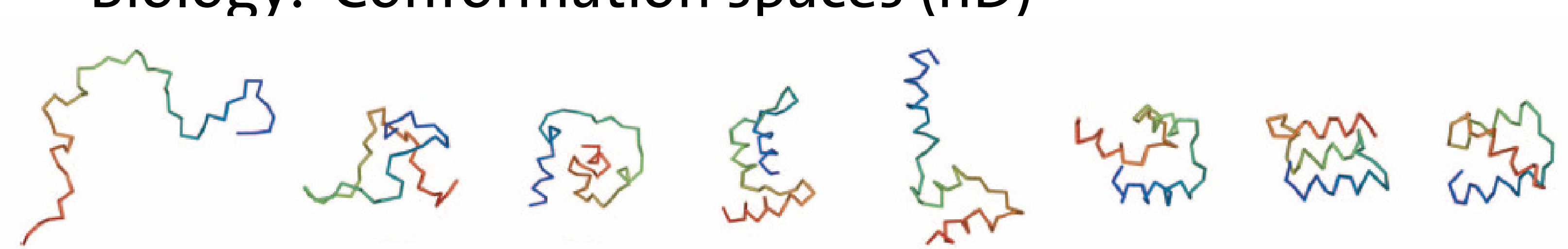
Abstract Spaces

- Spaces with motion
- Each point in abstract space is a snapshot

- Robotics: Configuration spaces (nD)



- Biology: Conformation spaces (nD)

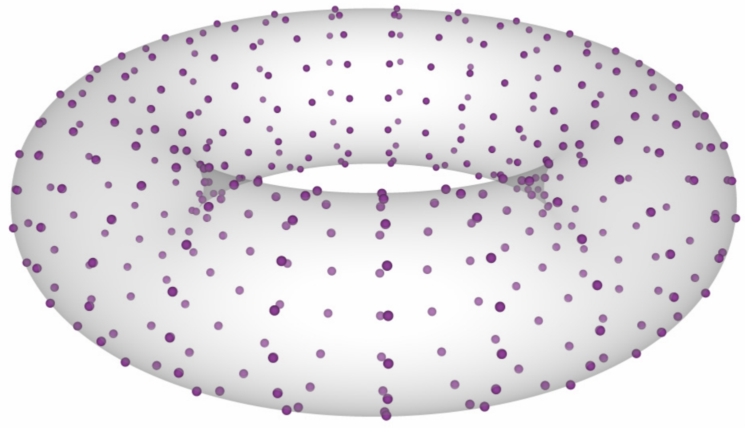


A Thought Exercise

- Example: 1 x 10^6 points in 100 dimensions
- How to compress?
  - Gzip?
  - Zip?
  - Better?
- Arbitrary compression not possible
- Knowledge: Points are on a circle
  - Fit a circle, parameterize it
  - Store angles ($\approx$ 100x compression)
  - Run Gzip
- Insight: Knowledge of structure allows compression
- Topology deals with structure

Computational Topology

- My view
- Input: Point Cloud Data
  - Massive
  - Discrete
  - Nonuniformly Sampled
  - Noisy
  - Embedded in $\mathbb{R}^d$, sometimes d >> 3



- Mission: What is its shape?

Plan

- Today:
  - ☺ Motivation
  - Topology
  - Simplicial Complexes
  - Invariants
  - Homology
  - Algebraic Complexes
- Tomorrow
  - Geometric Complexes
  - Persistent Homology
  - The Persistence Algorithm
  - Application to Natural Images

Outline

☺ Motivation

- Topology
  - Topological Space
  - Manifolds
  - Erlanger Programm
  - Classification
- Simplicial Complexes
- Invariants
- Homology
- Algebraic Complexes

Topological Space

- X: set of points

- Open set: subset of X
- Topology: set of open sets $T \subseteq 2^X$ such that
  1. If $S_1, S_2 \in T$, then $S_1 \cap S_2 \in T$
  2. If $\{S_J \mid j \in T\}$, then $\cup_{j \in J} S_j \in T$
  3. $\emptyset, X \in T$

- $\mathbb{X} = (X, T)$ is a topological space
- Note: different topologies possible
- Metric space: open sets defined by metric

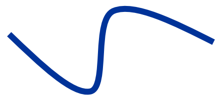
Homeomorphism

- Topological spaces $\mathbb{X}$, $\mathbb{Y}$
- Map f : $\mathbb{X} \rightarrow \mathbb{Y}$
- f is continuous, 1-1, onto (bijective)
- f^{-1} also continuous

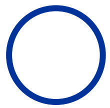
- f is a homeomorphism
- $\mathbb{X}$ is homeomorphic to $\mathbb{Y}$
- $\mathbb{X} \approx \mathbb{Y}$
- $\mathbb{X}$ and $\mathbb{Y}$ have same topological type

Examples

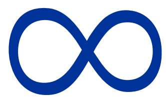
- Closed interval



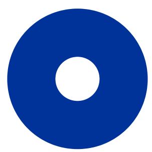
- Circle $\mathbb{S}^1$



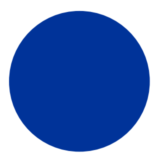
- Figure 8



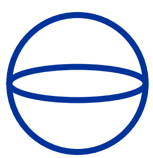
- Annulus



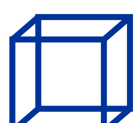
- Ball $\mathbb{B}^2$



- Sphere $\mathbb{S}^2$



- Cube



- interval $\not\approx \mathbb{S}^1$
- $\mathbb{S}^1 \not\approx$ Figure 8
- $\mathbb{S}^1 \not\approx$ Annulus
- Annulus $\not\approx \mathbb{B}^2$
- $\mathbb{S}^2 \approx$ Cube
- Captures
  - boundary
  - junctions
  - holes
  - dimension

- Continuous $\Rightarrow$ no gluing
- Continuous$^{-1} \Rightarrow$ no tearing
- Stretching allowed!

Erlanger Programm 1872

- Christian Felix Klein (1849-1925)
- Unifying definition:
  1. *Transform* space in a fixed way
  2. *Observe* properties that do not change

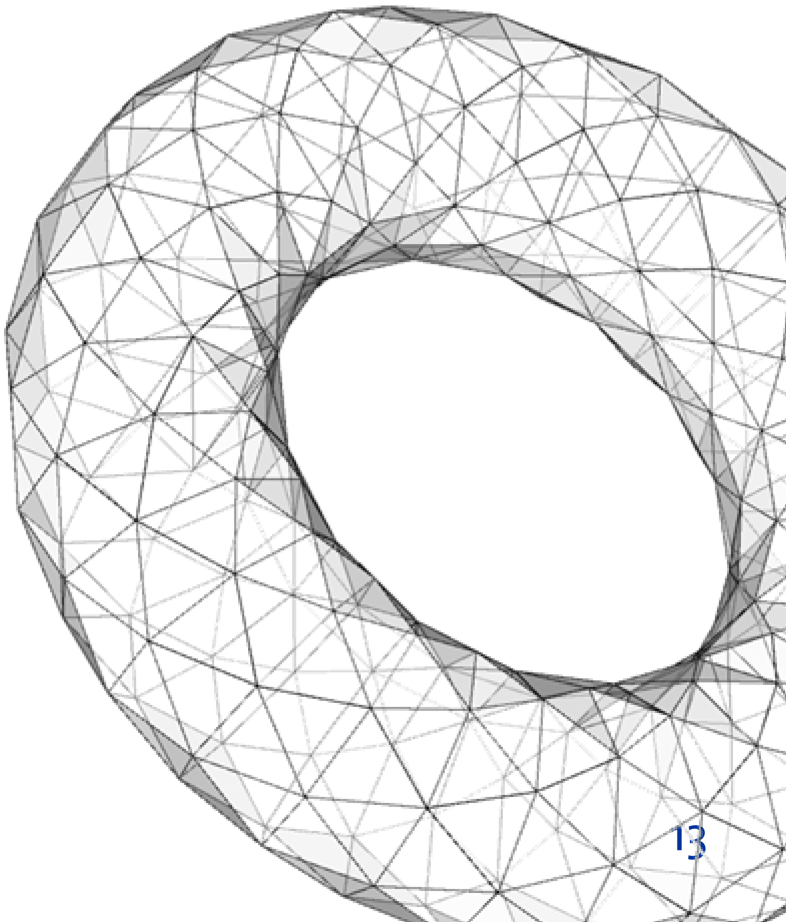
- Transformations
  - Rigid motions: translations & rotations
  - Homeomorphism: stretch, but do not tear or sew

- Rigid motions $\Rightarrow$ Euclidean Geometry
- Homeomorphisms $\Rightarrow$ Topology

Geometry vs. Topology

- Euclidean geometry
  - What does a space look like?
  - Quantitative
  - Local
  - Low-level
  - Fine

- Topology
  - How is a space connected?
  - Qualitative
  - Global
  - High-level
  - Coarse



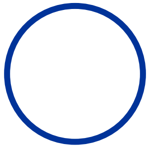
The Homeomorphism Problem

- Given: topological spaces $\mathbb{X}$ and $\mathbb{Y}$
- Question: Are they homeomorphic?

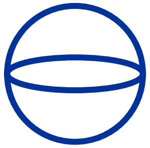
- Much coarser than geometry
  - Cannot capture singular points (edges, corners)
  - Cannot capture size
  - Classification system

Manifolds

- Given $\mathbb{X}$
- Every point $x \in \mathbb{X}$ has neighborhood $\approx \mathbb{R}^d$
- ($\mathbb{X}$ is separable and Hausdorff)
- $\mathbb{X}$ is a d-manifold (d-dimensional)

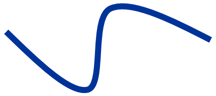


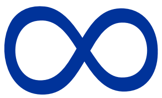
$\mathbb{S}^1$

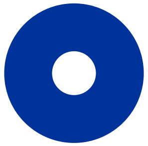


$\mathbb{S}^2$

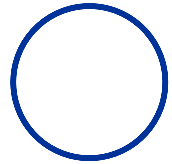
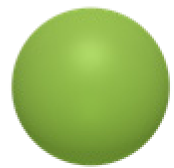
- $\mathbb{X}$ has some points with nbhd $\approx \mathbb{H}^d = \{x \in \mathbb{R}^d \mid x_1 \geq 0\}$
- $\mathbb{X}$ is a d-manifold with boundary
- Boundary $\partial\mathbb{X}$ are those points

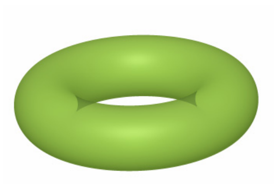




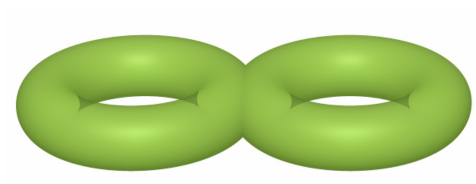


Compact 2-Manifolds

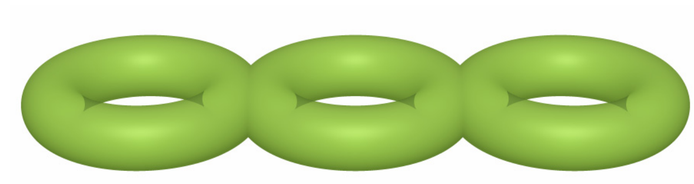
- d = 1: one manifold $\mathbb{S}^1$ 
- d = 2: orientable $\mathbb{S}^2$ 



Torus



Double Torus



Triple Torus

. . .

- d = 2: non-orientable

Projective Plane $\mathbb{P}^2$

Klein Bottle

. . .

Manifold Classification

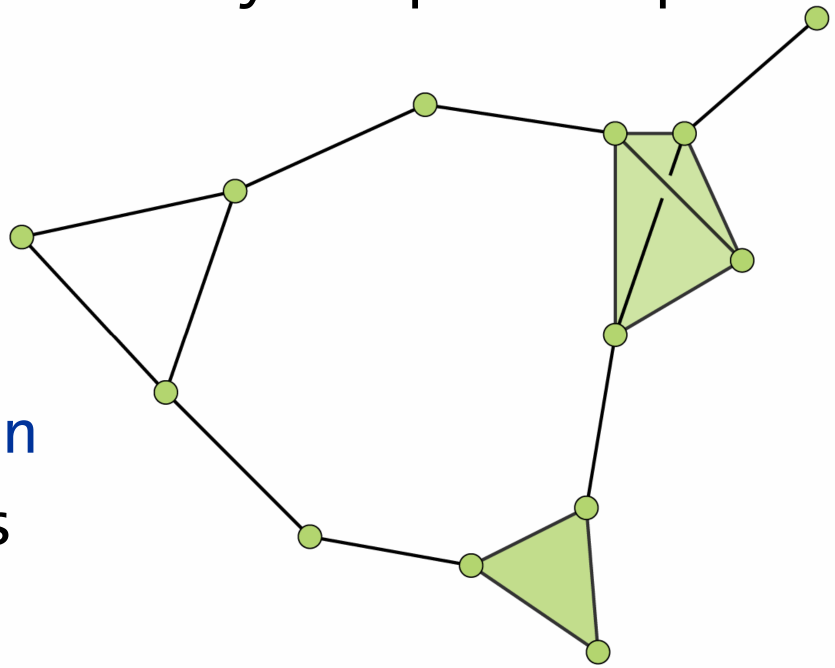
- Compact manifolds
  - closed
  - bounded
- d = 1: Easy
- d = 2: Done [Late 1800's]
- d $\geq$ 4: *Undecidable* [Markov 1958]
  - Dehn's Word Problem 1912
  - [Adyan 1955]
- d = 3: Very hard
  - The Poincaré Conjecture 1904
  - Thurston's Geometrization Program 1982: piece-wise uniform geometry
  - Ricci flow with surgery [Perelman '03]

Outline

- ☺ Motivation
- ☺ Topology
- Simplicial Complexes
  - Geometric Definition
  - Combinatorial Definition
- Invariants
- Homology
- Algebraic Complexes

Simplices

- Simplex: convex hull of affinely independent points



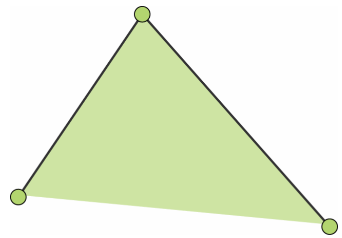
- 0-simplex: vertex
- 1-simplex: edge
- 2-simplex: triangle
- 3-simplex: tetrahedron
- k-simplex: $k + 1$ points

- face of simplex σ: defined by subset of vertices
- Simplicial complex: glue simplices along shared faces

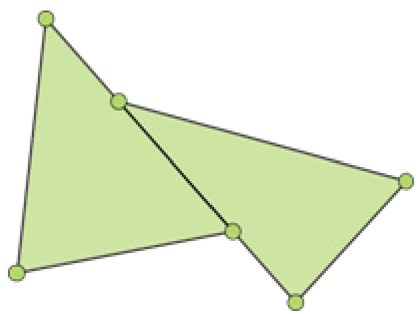
Simplicial Complex

- Every face of a simplex in a complex is in the complex

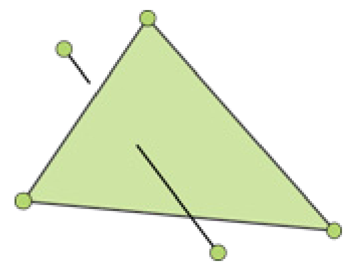
Edge is missing



- Non-empty intersection of two simplices is a face of each of them



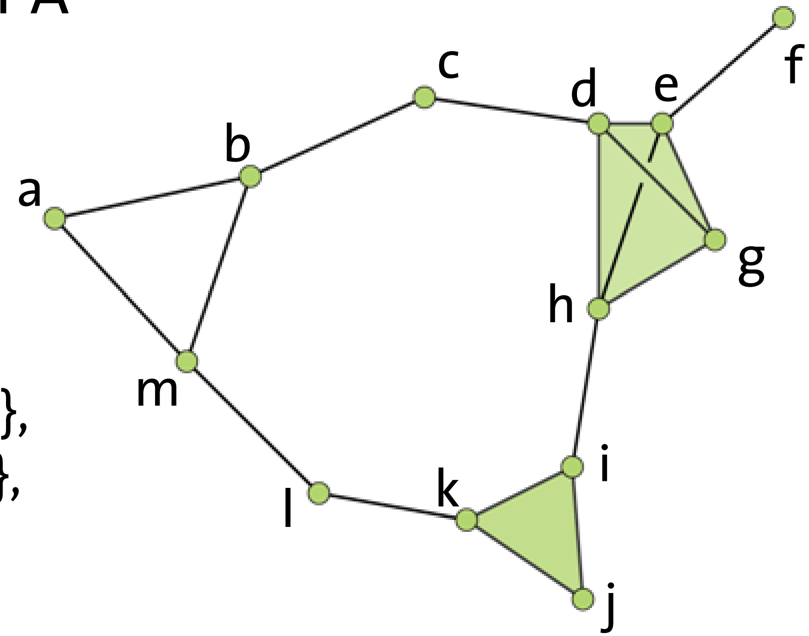
Sharing half an edge



Intersection not a vertex

Abstract Simplicial Complex

- Set of sets $\mathcal{S}$ such that if A $\in \mathcal{S}$, so is every subset of A
- $\mathcal{S}$ = {∅,
  {a}, {b}, {a, b}, {c}, {b, c},
  {d}, {c, d}, {e}, {d, e}, {f},
  {e, f}, {g}, {d, g}, {e, g},
  {d, e, g}, {h}, {d, h}, {e, h},
  {g, h}, {d, g, h}, {d, e, h}, {e, g, h},
  {d, e, g, h}, {i}, {h, i}, {j}, {i, j}, {k},
  {i, k}, {j, k}, {i, j, k}, {l}, {k, l},
  {m}, {a, m}, {b, m}, {l, m}
  }



Abstract → Geometric Visualization → Geometric

Geometric → Vertex Scheme → Abstract

Outline

☺ Motivation

☺ Topology

☺ Simplicial Complexes

- Invariants
  - Definition
  - The Euler Characteristic
  - Homotopy
- Homology
- Algebraic Complexes

Invariants

- The Homeomorphism problem is hard
- How about a partial answer?

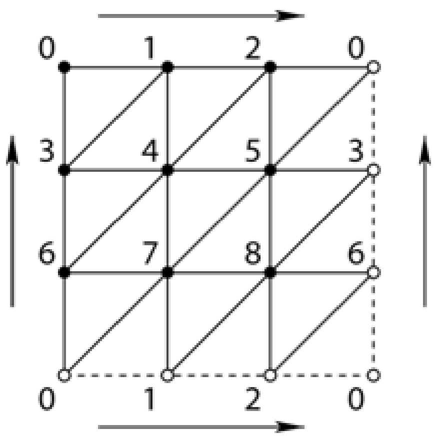
- Topological invariant: a map f that assigns the *same* object to spaces of the *same* topological type
  - $\mathbb{X} \approx \mathbb{Y} \Rightarrow f(\mathbb{X}) = f(\mathbb{Y})$
  - $f(\mathbb{X}) \neq f(\mathbb{Y}) \Rightarrow \mathbb{X} \not\approx \mathbb{Y}$ (contrapositive)
  - $f(\mathbb{X}) = f(\mathbb{Y}) \Rightarrow$ nothing

- Spectrum
  - trivial: $f(\mathbb{X})$ = one object, for all $\mathbb{X}$
  - complete: $f(\mathbb{X}) = f(\mathbb{Y}) \Rightarrow \mathbb{X} \approx \mathbb{Y}$

The Euler Characteristic

- Given: (abstract) simplicial complex K
- s_i: # of i-simplices in K
- Euler characteristic $\xi(K)$:

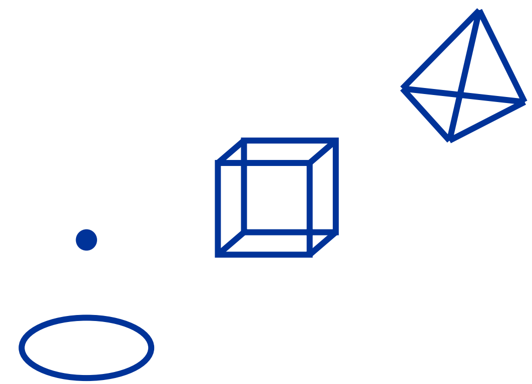
$$\xi(K) = \sum_{i=0}^{\dim K} (-1)^i s_i$$



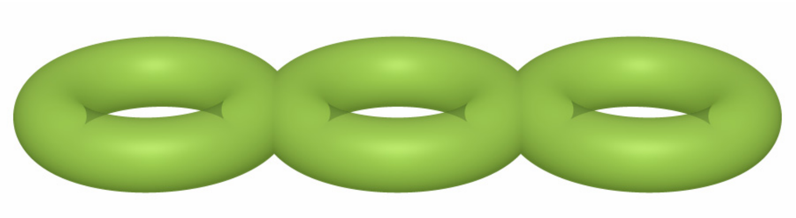
ξ(torus) = 9 – 27 + 18 = 0

The Euler Characteristic

- Invariant, so complex does not matter
- ξ(sphere) = 2
  - ξ(tetrahedron) = 4 – 6 + 4 = 2
  - ξ(cube) = 8 – 12 + 6 = 2
  - ξ(disk $\cup$ point) = 1 – 0 + 1 = 2

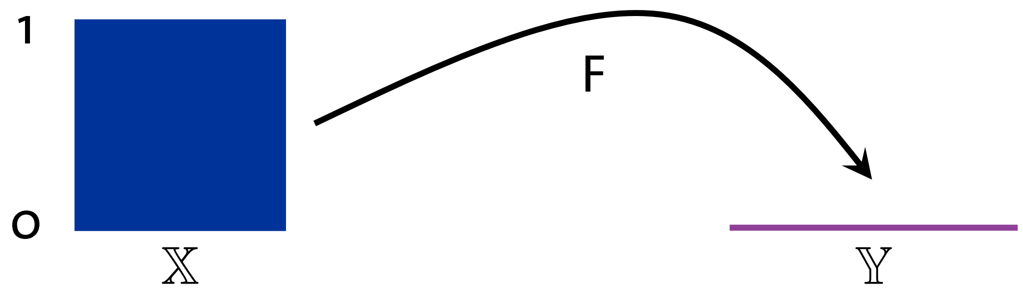


- ξ(g-torus) = 2 – 2g, genus g
- ξ(g$\mathbb{P}^2$) = 2 – g



Homotopy

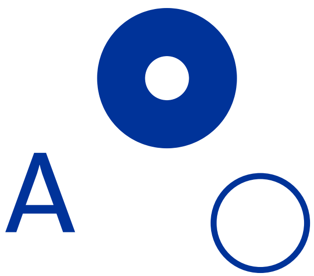
- Given: Family of maps $f_t : \mathbb{X} \to \mathbb{Y}$, $t \in [0,1]$
- Define $F : \mathbb{X} \times [0,1] \to \mathbb{Y}$, $F(x,t) = f_t(x)$



- If F is continuous, f_t is a homotopy
- $f_0, f_1 : \mathbb{X} \to \mathbb{Y}$ are homotopic via f_t
- $f_0 \simeq f_1$

Homotopy Equivalence

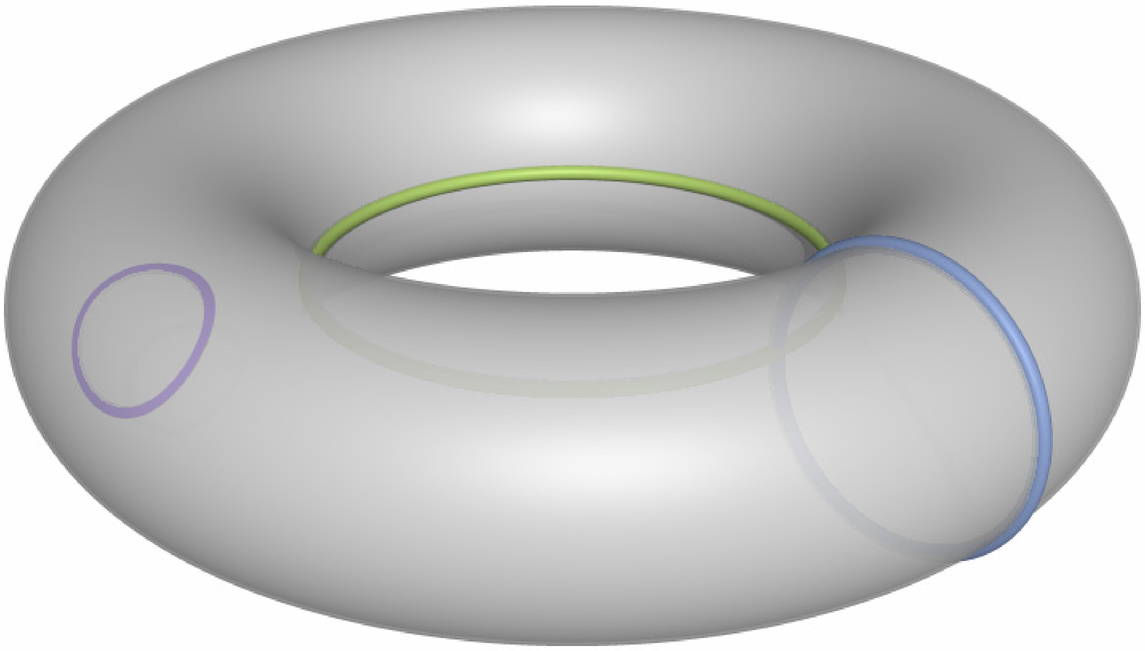
- Given: $f : \mathbb{X} \to \mathbb{Y}$
- Suppose $\exists g : \mathbb{Y} \to \mathbb{X}$ such that
  - $f \circ g \simeq 1_{\mathbb{Y}}$
  - $g \circ f \simeq 1_{\mathbb{X}}$
- f is a homotopy equivalence
- $\mathbb{X}$ and $\mathbb{Y}$ are homotopy equivalent $\mathbb{X} \simeq \mathbb{Y}$
- Comparison
  - Homeomorphism: $g \circ f = 1_{\mathbb{X}}$ $f \circ g = 1_{\mathbb{Y}}$
  - Homotopy: $g \circ f \simeq 1_{\mathbb{X}}$ $f \circ g \simeq 1_{\mathbb{Y}}$
- **(Theorem)** $\mathbb{X} \approx \mathbb{Y} \Rightarrow \mathbb{X} \simeq \mathbb{Y}$
- Contractible: homotopy equivalent to a point



Outline

- ☺ Motivation
- ☺ Topology
- ☺ Simplicial Complexes
- ☺ Invariants
- Homology
  - Intuition
  - Homology Groups
  - Computation
  - Euler-Poincaré
- Algebraic Complexes

Intuition



Overview

- Algebraic topology: algebraic images of topological spaces
- Homology
  - How cells of dimension n attach to cells of dimension $n - 1$
  - Images are groups, modules, and vector spaces
- Simplicial homology: cells are simplices

- Plan:
  - chains: like paths, maybe disconnected
  - cycles: like loops, but a loop can have multiple components
  - boundary: a cycle that bounds

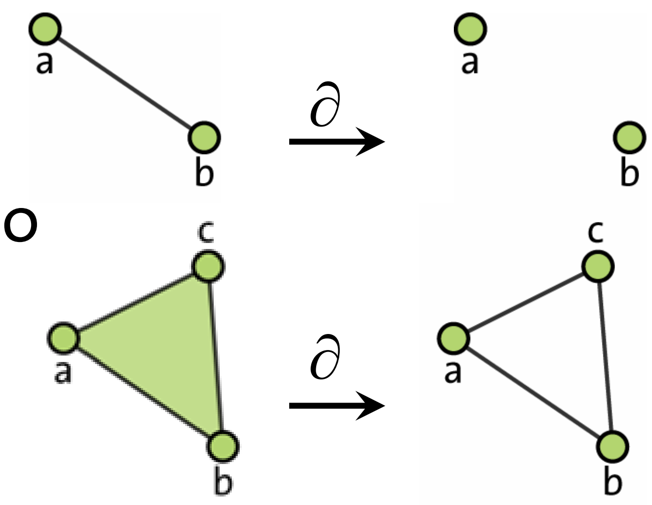
Chains

- Given: Simplicial complex K
- k-chain:
  - list of k-simplices in K
  - formal sum $\sum_i n_i \sigma_i$, where $n_i \in \{0, 1\}$ and $\sigma_i \in K$
- Field $\mathbb{Z}_2$
  - $0 + 0 = 0$
  - $0 + 1 = 1 + 0 = 1$
  - $1 + 1 = 0$
- Chain vector space C_k: vector space spanned by k-simplices in K
- rank $C_k = s_k$, number of k-simplices in K

Boundary Operator

- $\partial_k : C_k \rightarrow C_{k-1}$
- homomorphism (linear)
- $\sigma = [v_0, \ldots, v_k]$
- $\partial_k \sigma = \sum_i [v_0, \ldots, v_i', \ldots, v_k]$,
  where v_i' indicates that v_i is deleted from the sequence
- $\partial_1 ab = a + b$
- $\partial_2 abc = ab + bc + ac$
- $\partial_1 \partial_2 abc = a + b + b + c + a + c = 0$

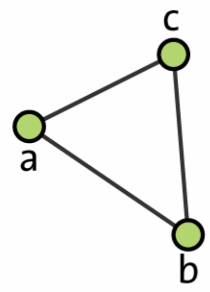
- **(Theorem)** $\partial_{k-1} \partial_k = 0$ for all k



Cycles

- Let c be a k-chain
- If c has no boundary, it is a k-cycle
- $\partial_k c = 0$, so $c \in \ker \partial_k$
- $Z_k = \ker \partial_k$ is a subspace of C_k

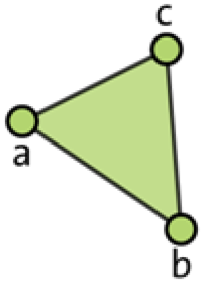
- $\partial_1(ab + bc + ac) =$
  $a + b + b + c + a + c = 0$,
  so 1-chain $ab + bc + ac$ is a 1-cycle



Boundaries

- Let b be a k-chain
- If b bounds something, it is a k-boundary
- $\exists d \in C_{k+1}$ such that $b = \partial_{k+1} d$
- $B_k = \text{im } \partial_{k+1}$ is a subspace of C_k

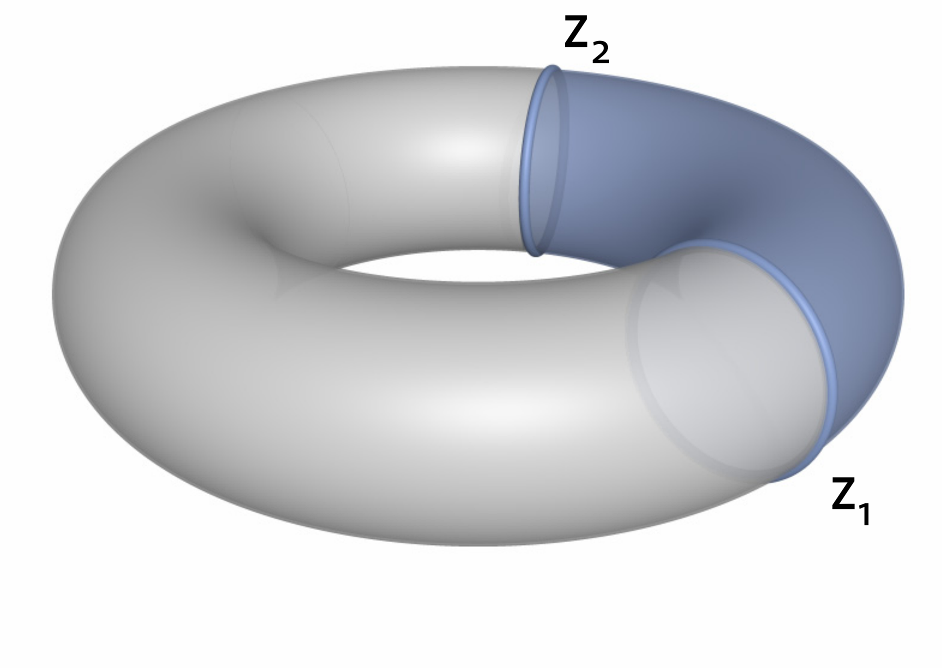
- $\partial_2(abc) = ab + bc + ac$,
  so ab + bc + ac is a 1-boundary



- $\partial_k b = \partial_k \partial_{k+1} d = 0$, so b is also a k-cycle!
- All boundaries are cycles
- $B_k \subseteq Z_k \subseteq C_k$

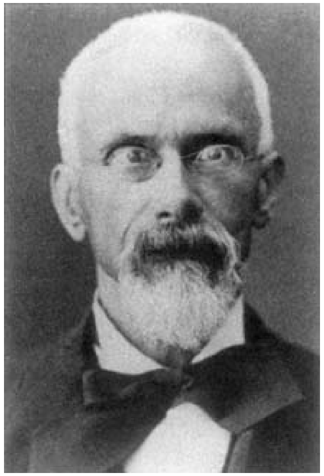
Homology Group

- The kth homology vector space (group) is $H_k = Z_k / B_k = \ker \partial_k / \operatorname{im} \partial_{k+1}$
- (Theorem) $\mathbb{X} \simeq \mathbb{Y} \Rightarrow H_k(\mathbb{X}) \cong H_k(\mathbb{Y})$
- If $z_1 = z_2 + b$, where $b \in B_k$, z_1 and z_2 are homologous, $z_1 \sim z_2$

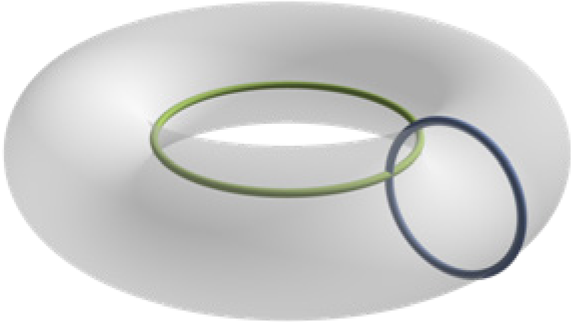


Betti Numbers

- H_k is a vector space
- kth Betti number β_k = rank H_k
  = rank Z_k – rank B_k



- Enrico Betti (1823 – 1892)
- Geometric interpretation in R^3
  - β_0 is number of components
  - β_1 is rank of a basis for tunnels
  - β_2 is number of voids



1, 2, 1

Computation

- ∂_k is linear, so it has a matrix M_k in terms of bases for C_k and C_{k-1}
- $Z_k = \ker \partial_k$, so compute dim(null(M_k))
- $B_k = \operatorname{im} \partial_{k+1}$, so compute dim(range($M_{k+1}$))

- Two Gaussian eliminations, so $O(m^3)$, $m = |K|$
- Same running time for any field

- Over $\mathbb{Z}$, reduction algorithm and matrix entries can get large
- Common source of misunderstanding

Euler-Poincaré

- Recall $\xi(K) = \sum_i (-1)^i s_i$
- s_i = # k-simplices in K
- s_i = rank C_i
- Rewrite: $\xi(K) = \sum_i (-1)^i \text{ rank } C_i$

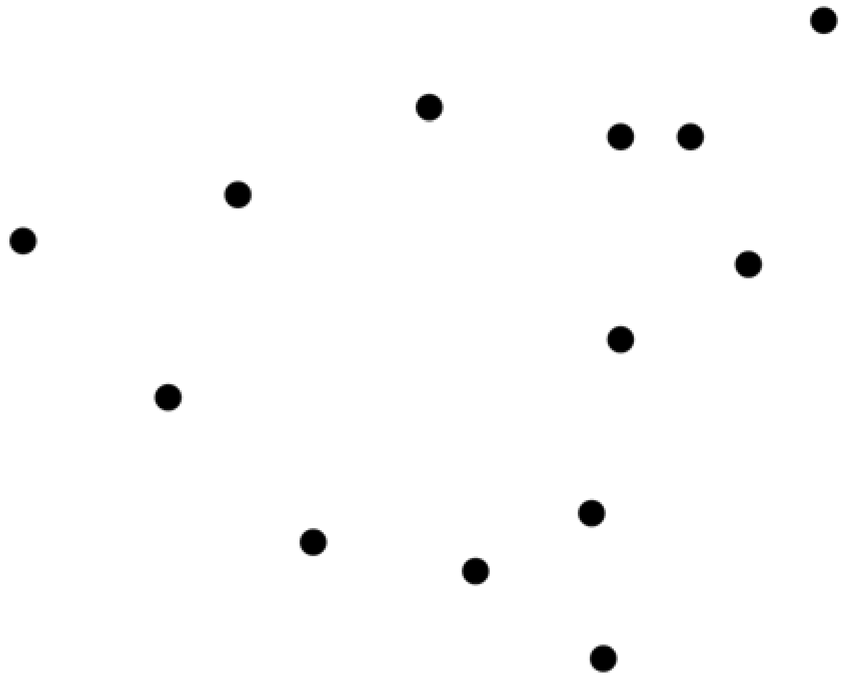
- **(Theorem)** $\xi(K) = \sum_i (-1)^i \text{ rank } H_i = \sum_i (-1)^i \beta_i$

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

Outline

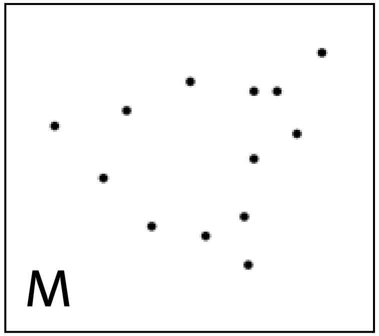
- ☺ Motivation
- ☺ Topology
- ☺ Simplicial Complexes
- ☺ Invariants
- ☺ Homology
- Algebraic Complexes
  - Coverings
  - The Nerve
  - Cech complex
  - Vietoris-Rips Complex

Topology of Points



Topology of Points

- Topological space $\mathbb{X}$
- Underlying space

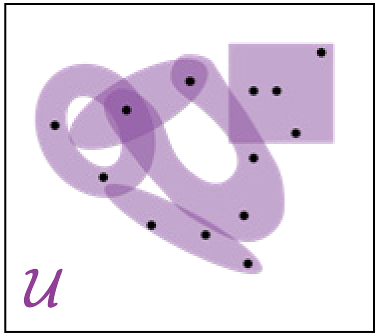


- Given: set of sample points M from $\mathbb{X}$
- Question: How can we recover the topology of $\mathbb{X}$ from M?
- Problem: M has no interesting topology.

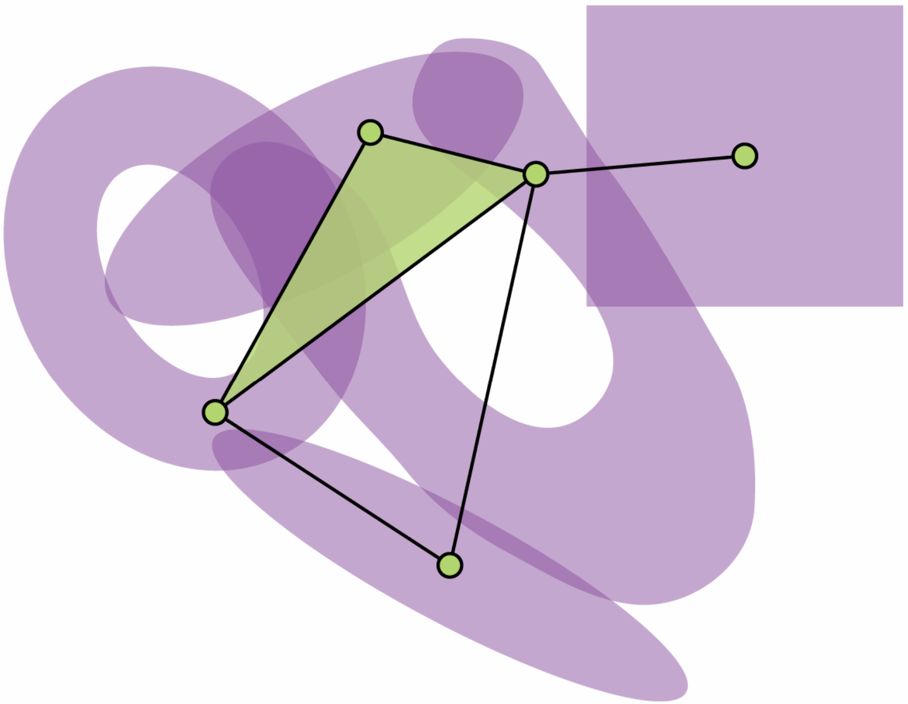
Open Covering

Open Covering

- Cover $\mathcal{U} = \{U_i\}_{i \in I}$
  - U_i, open
  - $M \subseteq \bigcup_{i \in I} U_i$
- Idea: The cover approximates the underlying space $\mathbb{X}$
- Question′: What is the topology of $\mathcal{U}$?
- Problem: $\mathcal{U}$ is an infinite point set

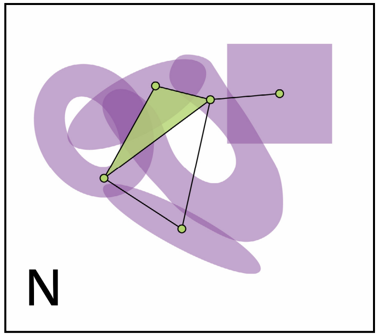


The Nerve



The Nerve

- $\mathbb{X}$: topological space
- $\mathcal{U} = \cup_{i \in I} U_i$: open cover of $\mathbb{X}$

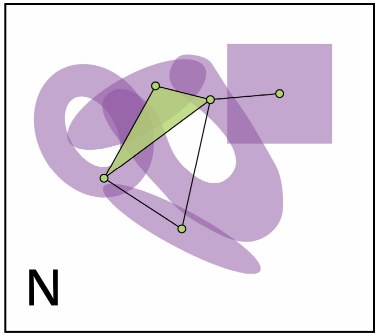


- The nerve N of $\mathcal{U}$ is
  - $\emptyset \in N$
  - If $\cap_{j \in j} U_j \neq \emptyset$ for $J \subseteq I$, then $J \in N$

- Dual structure
- (Abstract) Simplicial complex

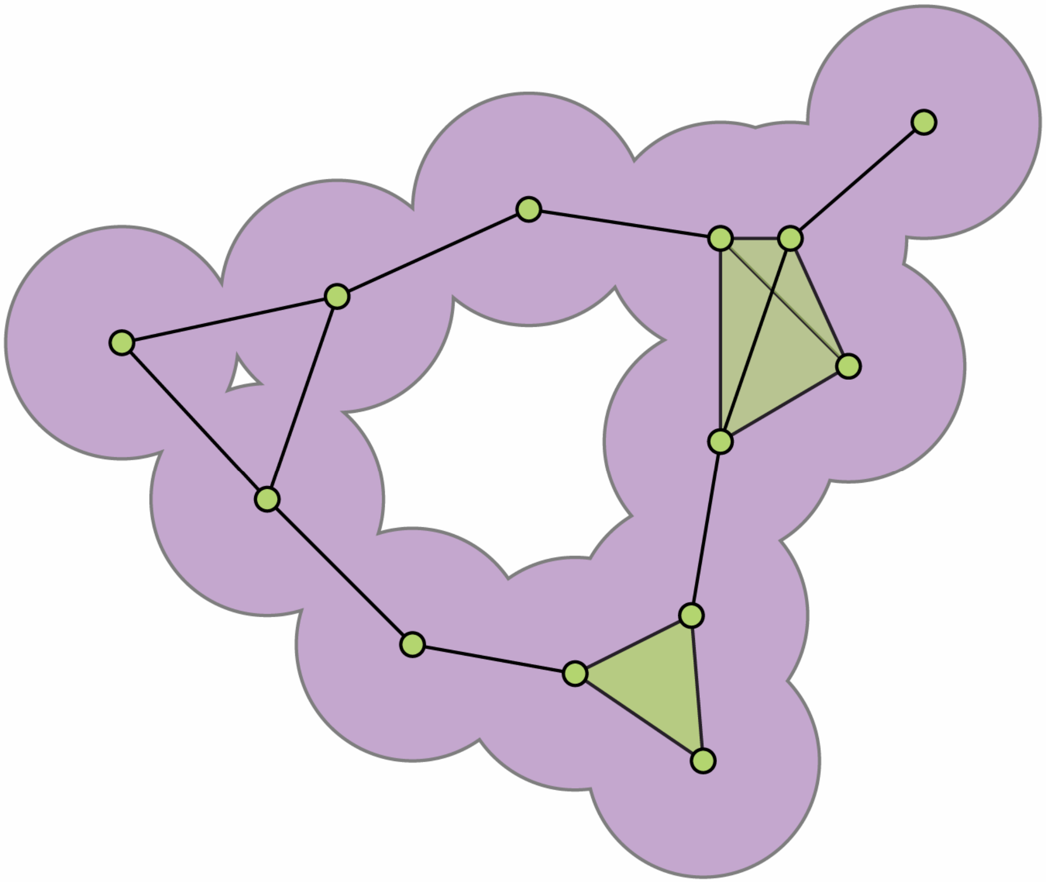
The Nerve Lemma

- **(Lemma [Leray])**
  If sets in the cover are contractible, and their finite unions are contractible, then N $\simeq \mathcal{U}$.

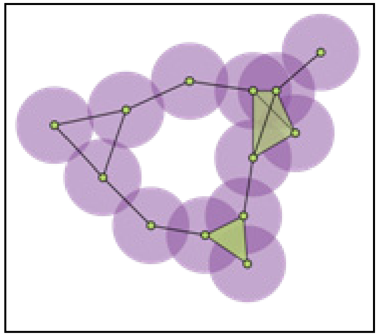


- *The cover should not introduce or eliminate topological structure*
- Idea: Use “nice” sets for covering
  - contractible
  - convex
- Dual (abstract) simplicial complex will be our representation

Cech Complex



Cech Complex

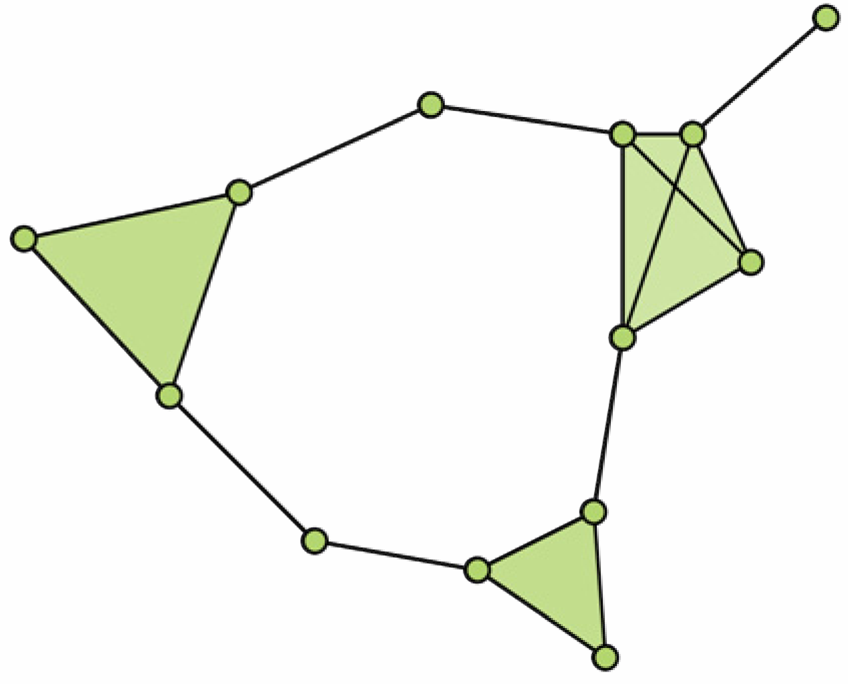


- Set: Ball of radius ε
  $B_\varepsilon(x) = \{ y \mid d(x, y) < \varepsilon\}$
- Cover: B_ε at every point in M
- Cech complex is nerve of the union of ε-balls

$$C_\epsilon(M) = \left\{ \text{conv}\, T \mid T \subseteq M, \bigcup_{m \in T} B_\epsilon(m) \neq \emptyset \right\}$$

- Cover satisfies Nerve Lemma
- Eduard Cech (1893 – 1960)

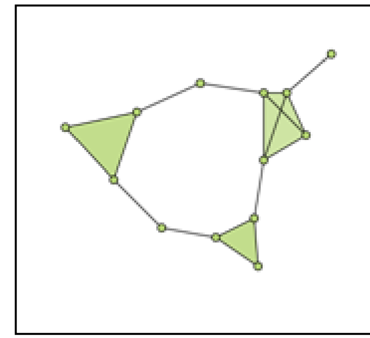
Vietoris-Rips Complex



Vietoris-Rips Complex

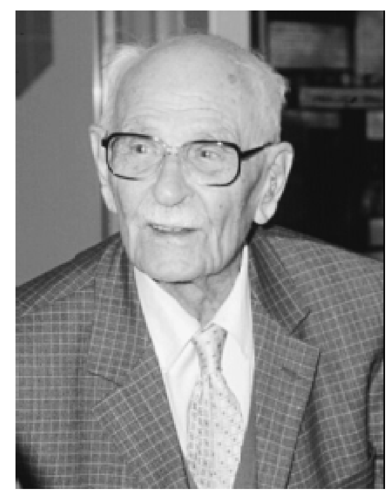
1. Construct ε-graph
2. Expand by add a simplex whenever all its faces are in the complex

- Note: We expand by dimension



$$V_\epsilon(M) = \{\mathsf{conv}\, T \mid T \subseteq M, \mathsf{d}(x,y) < \epsilon, \forall x,y \in T\}$$

- $V_{2\varepsilon}(M) \supseteq C_\varepsilon(M)$
- Not homotopic to union of balls
- Leopold Vietoris (1891 – 2002)
- Eliyahu Rips (1948 –)



Plan

- ☺ Today:
  - Motivation
  - Topology
  - Simplicial Complexes
  - Invariants
  - Homology
  - Algebraic Complexes
- Tomorrow
  - Geometric Complexes
  - Persistent Homology
  - The Persistence Algorithm
  - Application to Natural Images