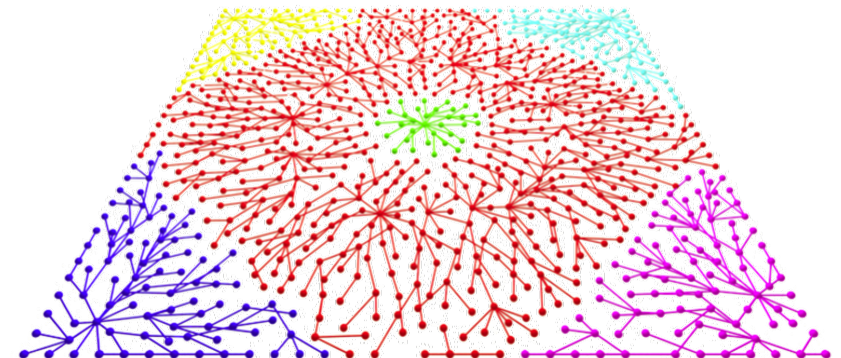


# CS233, CME251: Geometric and Topological Data Analysis

Leonidas Guibas  
Computer Science Department  
Stanford University



Lecture 7  
18 April 2022



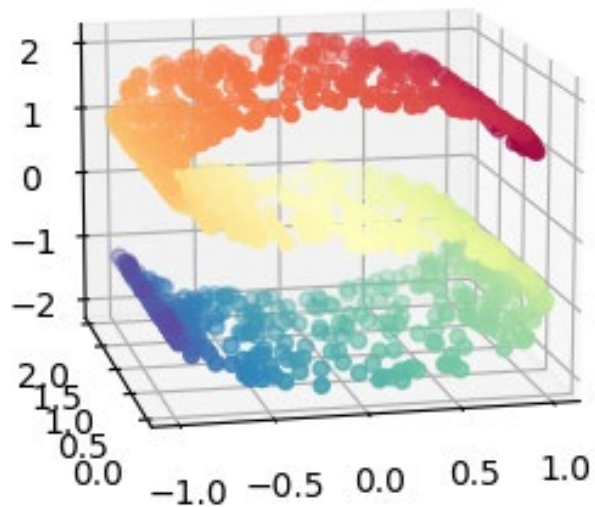
# Last Time: Non-Linear Dimensionality Reduction

Isomap, Laplacian Eigenmaps, Locally Linear Embeddings, t-SNE, Autoencoders

# Dimensionality Reduction

**Dimensionality Reduction** is the process of **transforming data** from a **high-dimensional space** into a **low-dimensional space** so that the **low-dimensional representation** retains some **meaningful properties of the original data**.

*Source: Wikipedia on Dimensionality Reduction*



**S-curve**

## Key Assumptions:

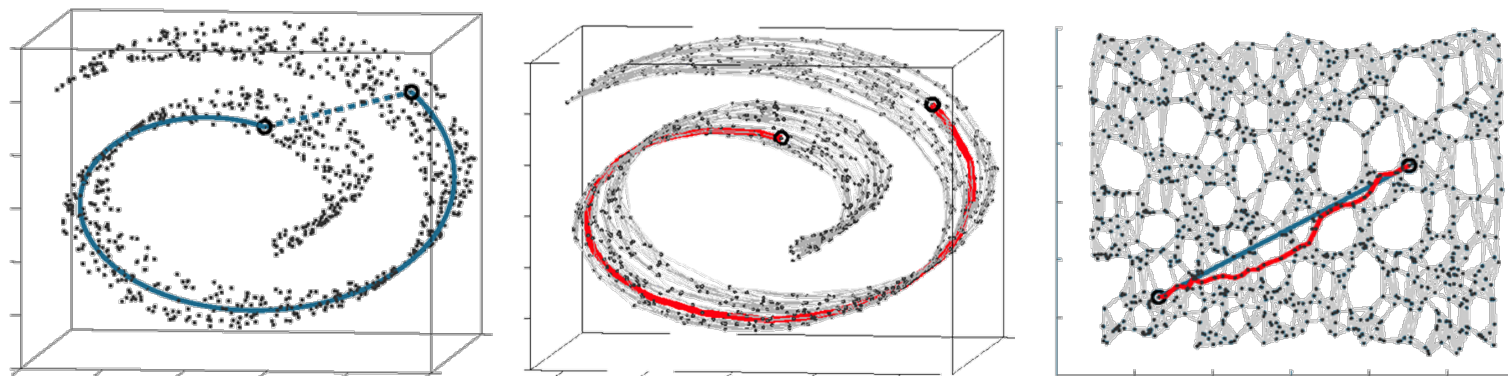
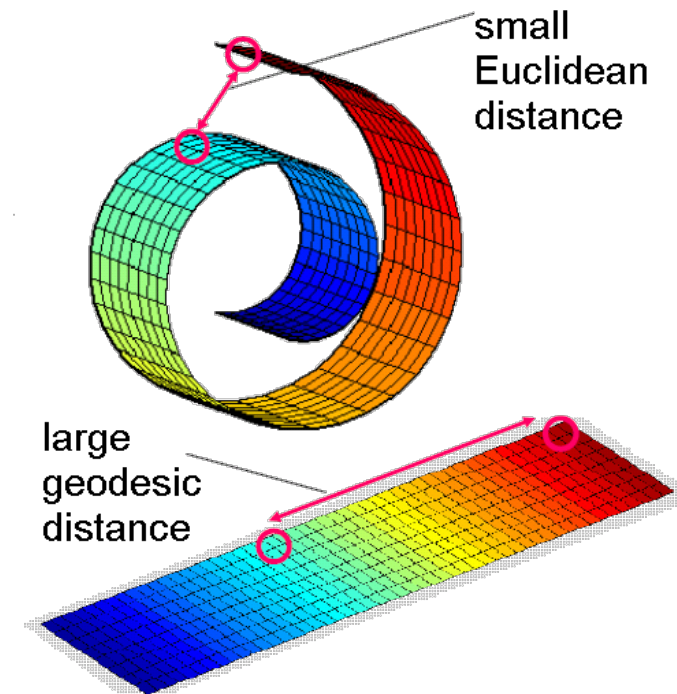
- Represent each data point by a point in a lower dimensional space.
- Choose **the low-dimensional points so that they can represent some properties of the data points in the original space**, such as the pairwise distances between points, namely two close points in the original high-dimensional space should also be close in the low-dimensional space.

# Methods for Non Linear Dimensionality Reduction

- **Sammon's mapping, 1969**
- Self-organizing map (SOM, aka Kohonen map, based on neural networks), 1982
- Principal curves and manifolds, 1984
- **Autoencoders (some neural networks), 19xx**
- Generative topographic map (GTM, probabilistic version of SOM), 1996
- Curvilinear component/distance analysis (CCA, CDA), 1997
- Kernel PCA (kPCA), 1998
- **ISOMAP, 2000**
- **Locally-linear embedding (LLE), 2000**
- **Laplacian Eigenmaps, 2001**
- Hessian LLE, 2003
- Gaussian process latent variable models (GPLVM), 2004
- Maximum variance unfolding (MVA, aka semidefinite embedding), 2004
- Maximum variance unfolding (MVA, aka semidefinite embedding), 2004
- Relational perspective map, 2004
- Nonlinear PCA (based on neural networks), 2005
- Local tangent space alignment (LTSA), 2005
- Modified LLE, 2006
- Diffusion maps, 2006
- Local multidimensional scaling, 2006
- Manifold alignment, 2008
- Manifold sculpting, 2008
- **t-distributed stochastic neighbor embedding (t-SNE), 2008**
- Diffeomorphic Dimensionality Reduction (Diffeomap), 2009
- Rank visu, 2009
- Topologically Constrained Isometric Embedding (TCIE), 2010

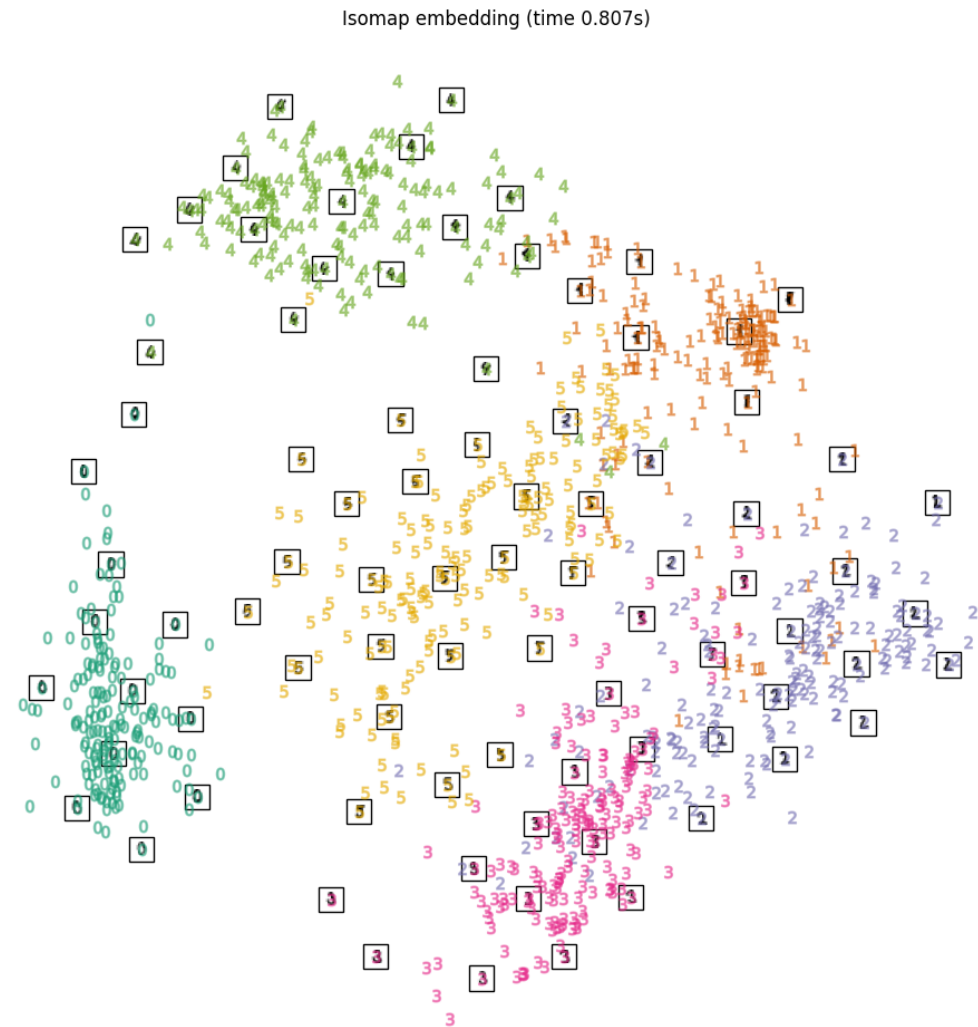
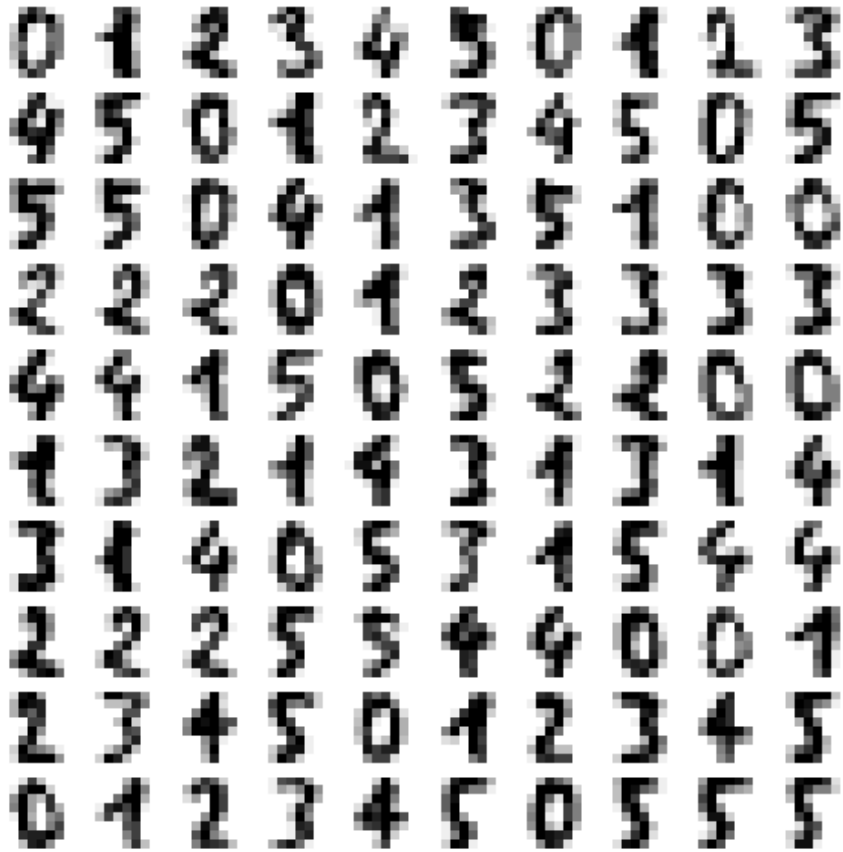
# ISOMAP (J. B. Tenenbaum, V. de Silva, and J. C. Langford)

- Example of non-linear structure (Swiss roll)
  - Only the geodesic distances reflect the true low-dimensional geometry of the manifold
- ISOMAP (Isometric Feature Mapping)
  - Uses the geodesic manifold distances between all pairs
  - Preserves the intrinsic geometry of the data -- Preserves the geodesic distances
  - Estimate geodesic distances between point pairs
  - Use MDS for an embedding



# Isomap - MNIST

A selection from the 64-dimensional digits dataset

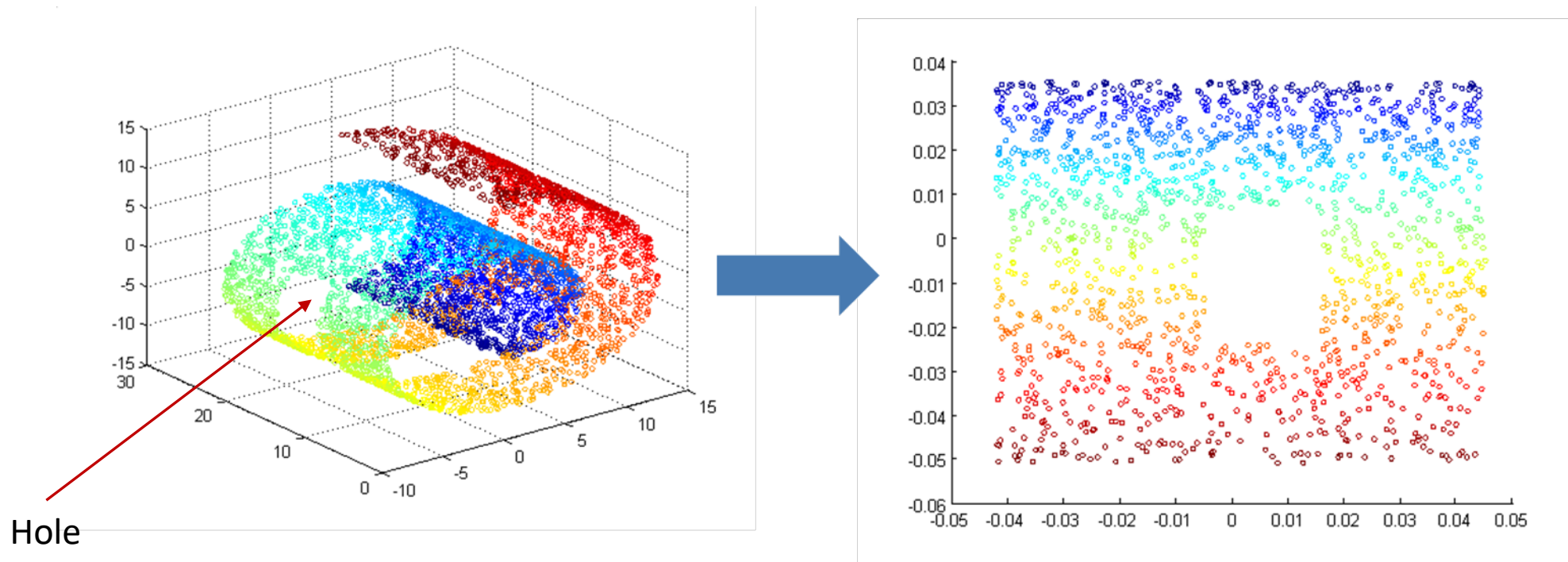


*Image Generated running the scikit-learn demo for Manifold Learning on MNIST:*

[https://scikit-learn.org/stable/auto\\_examples/manifold/plot\\_lle\\_digits.html#sphx-glr-auto-examples-manifold-plot-lle-digits-py](https://scikit-learn.org/stable/auto_examples/manifold/plot_lle_digits.html#sphx-glr-auto-examples-manifold-plot-lle-digits-py)

# Laplacian Eigenmaps (M. Belkin and P. Niyogi)

- Start same as Isomap, but use a spectral embedding in lieu of MDS



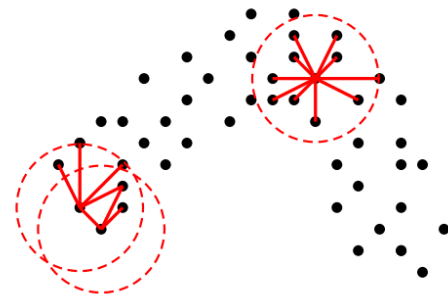
Hole distorts long geodesic distances, but affects less diffusion distances

# Laplacian Eigenmaps – Algorithmic Overview

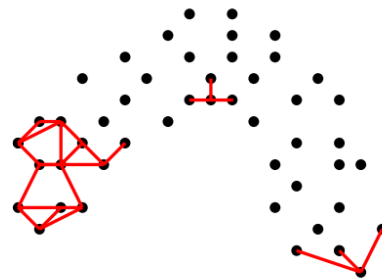
Given a set of points  $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_K\}$  in  $\mathbb{R}^D$ , construct a weighted graph with  $K$  nodes, one for each point, and a set of edges connecting neighboring points. The **embedding map can be provided by computing the eigenvectors of the graph Laplacian.**

**Neighborhood Graph Construction:** The first step of Laplacian Eigenmaps is to build a neighborhood graph  $G$  from the given data  $\mathcal{X}$  by **connecting only “nearby” points, where nearby is defined in one of the following ways:**

- **$\epsilon$ -ball approach:** Two points  $\mathbf{x}_i, \mathbf{x}_j$  are nearby if  $\|\mathbf{x}_i - \mathbf{x}_j\|^2 \leq \epsilon$ ,
- **kNN approach:** Two points  $\mathbf{x}_i, \mathbf{x}_j$  are nearby if one is among the  $k$  nearest neighbors of the other.



$\epsilon$ -ball graph



$k$ NN graph ( $k = 3$ )

# Laplacian Eigenmaps - MNIST

A selection from the 64-dimensional digits dataset

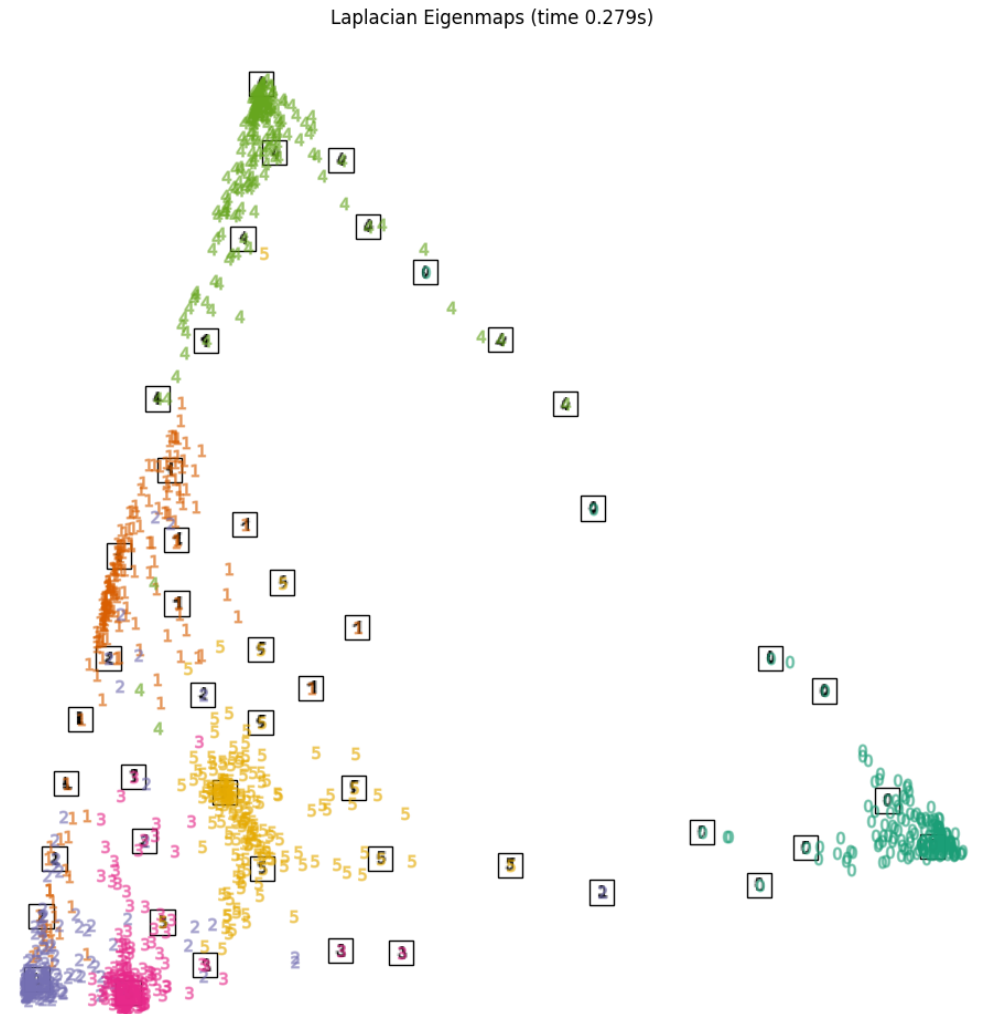
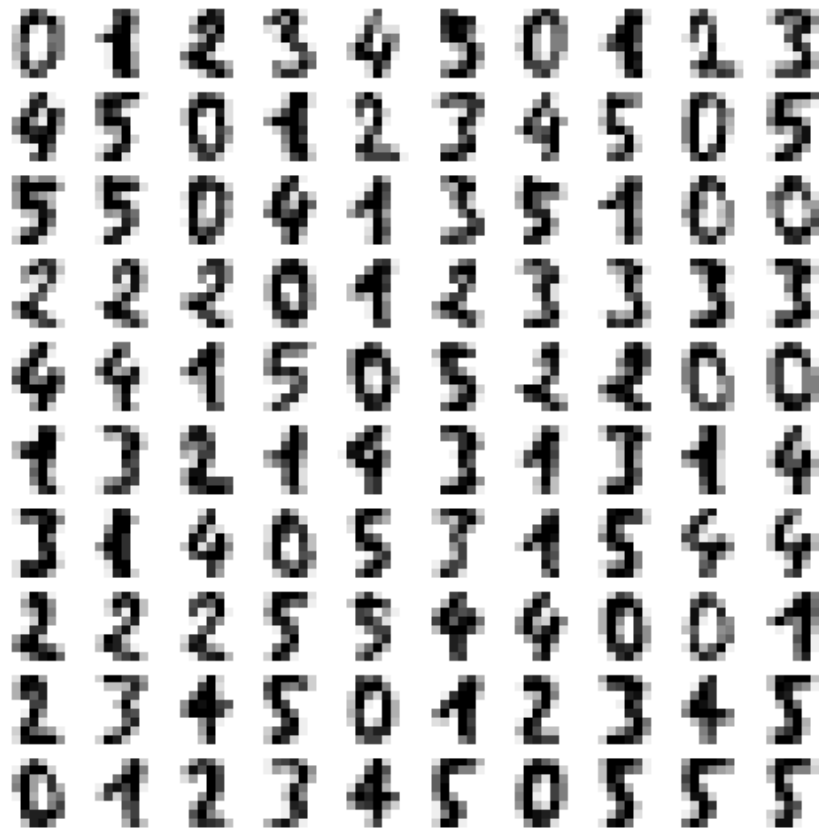


Image Generated running the scikit-learn demo for Manifold Learning on MNIST:

[https://scikit-learn.org/stable/auto\\_examples/manifold/plot\\_lle\\_digits.html#sphx-glr-auto-examples-manifold-plot-lle-digits-py](https://scikit-learn.org/stable/auto_examples/manifold/plot_lle_digits.html#sphx-glr-auto-examples-manifold-plot-lle-digits-py)

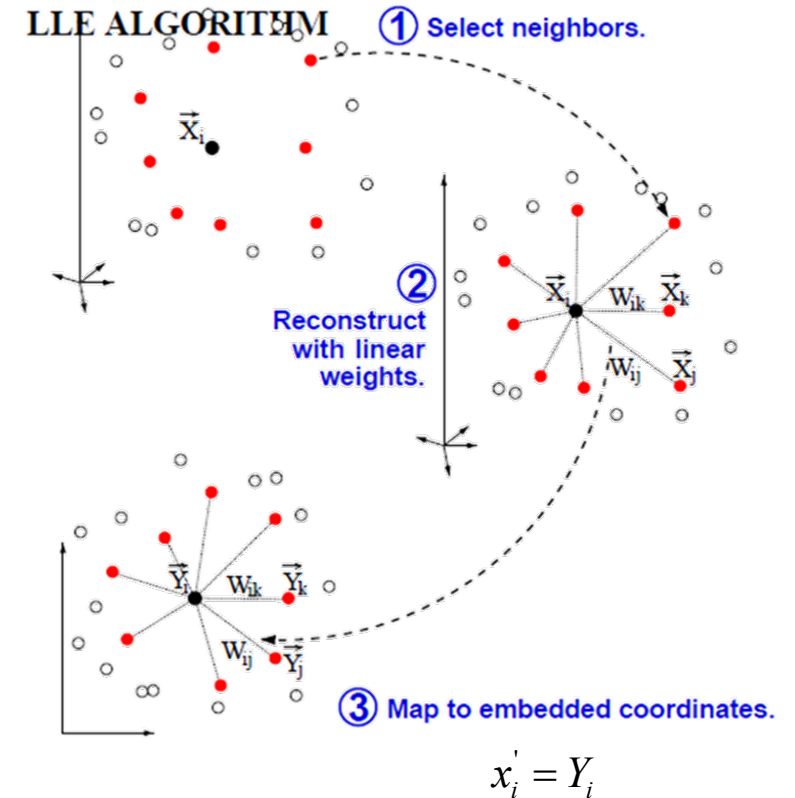
# Locally Linear Embeddings (LLE) (S. T. Roweis and L. K. Saul)

- Define neighborhood relations between points (build NN graph)
  - $k$  nearest neighbors
  - $\varepsilon$ -balls
- Find weights that reconstruct each data point from its neighbors:

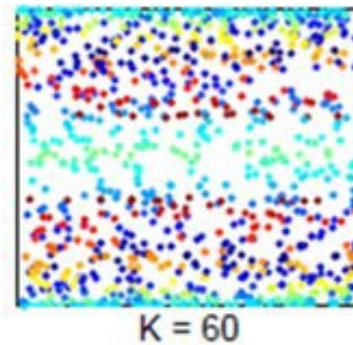
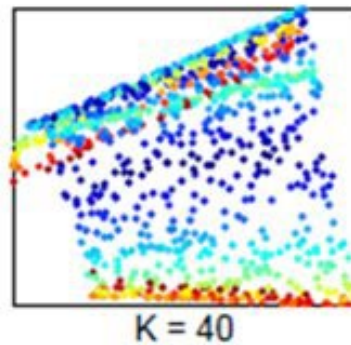
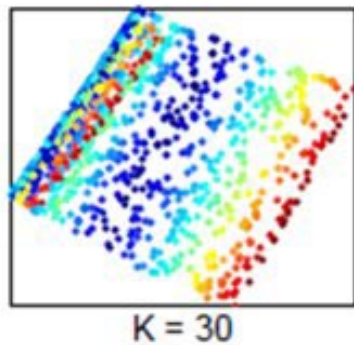
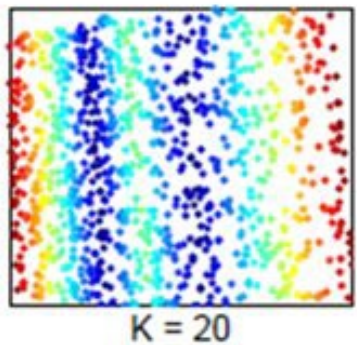
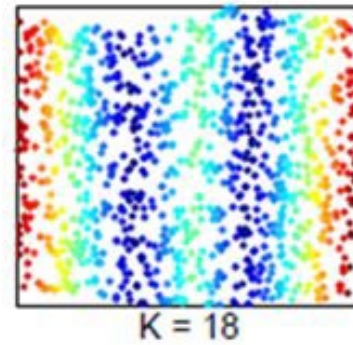
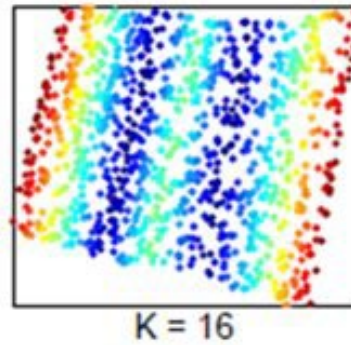
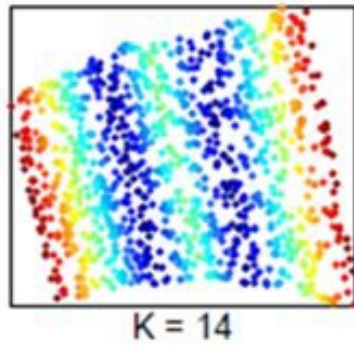
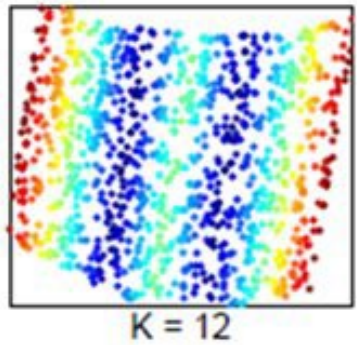
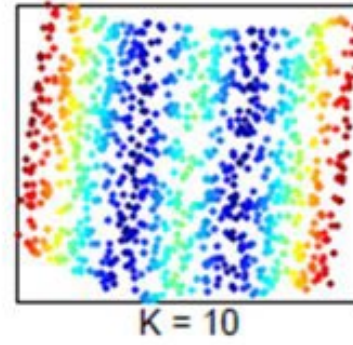
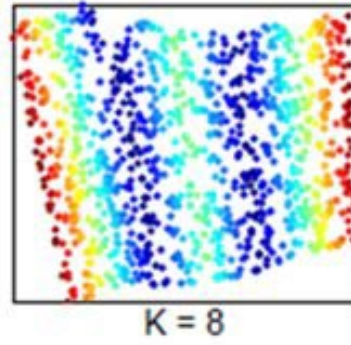
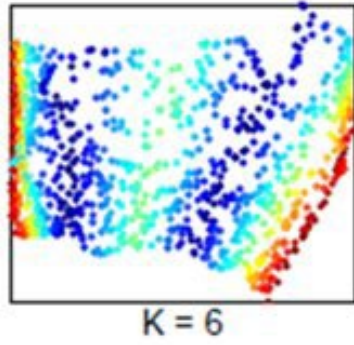
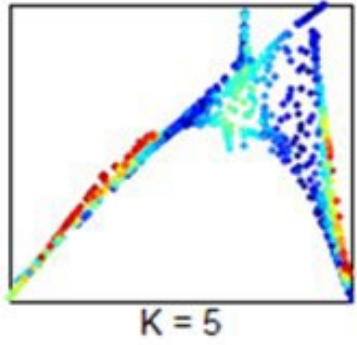
$$\min_{\sum_j w_{ij}=1} \left\| \mathbf{x}_i - \sum_{j \in N(i)} w_{ij} \mathbf{x}_j \right\|^2$$

- Find low-dimensional coordinates so that the same weights hold:  $\mathbf{x}'_1, \dots, \mathbf{x}'_n \in R^d$

$$\min_{\mathbf{x}'_1, \dots, \mathbf{x}'_n} \sum_i \left\| \mathbf{x}'_i - \sum_{j \in N(i)} w_{ij} \mathbf{x}'_j \right\|^2$$

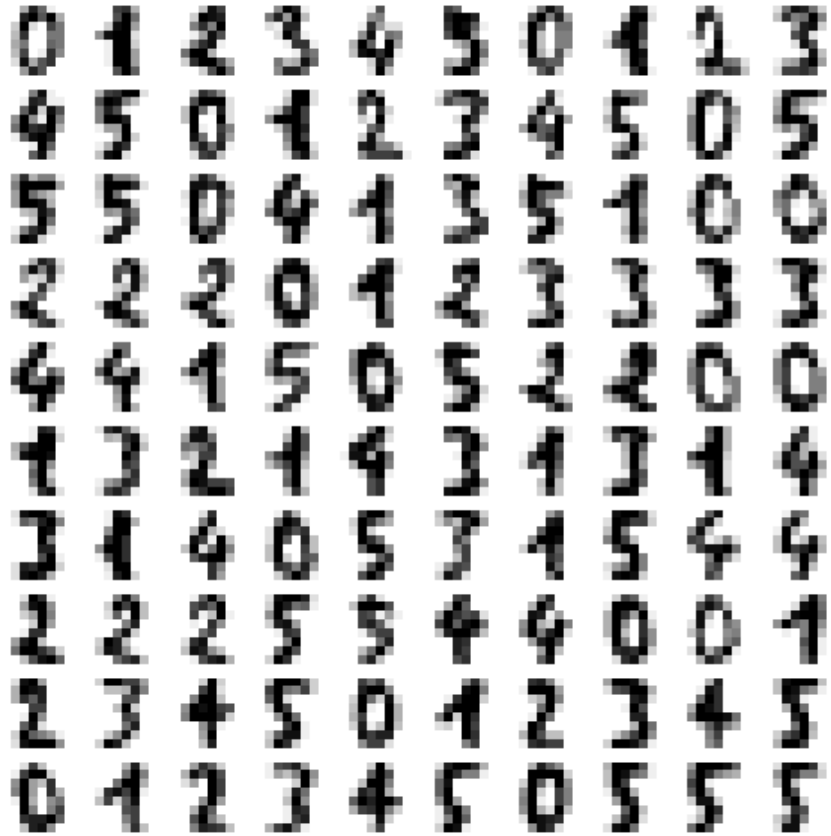


# Locally Linear Embeddings – Neighborhood size



# Locally Linear Embeddings - MNIST

A selection from the 64-dimensional digits dataset



Modified LLE embedding (time 0.532s)

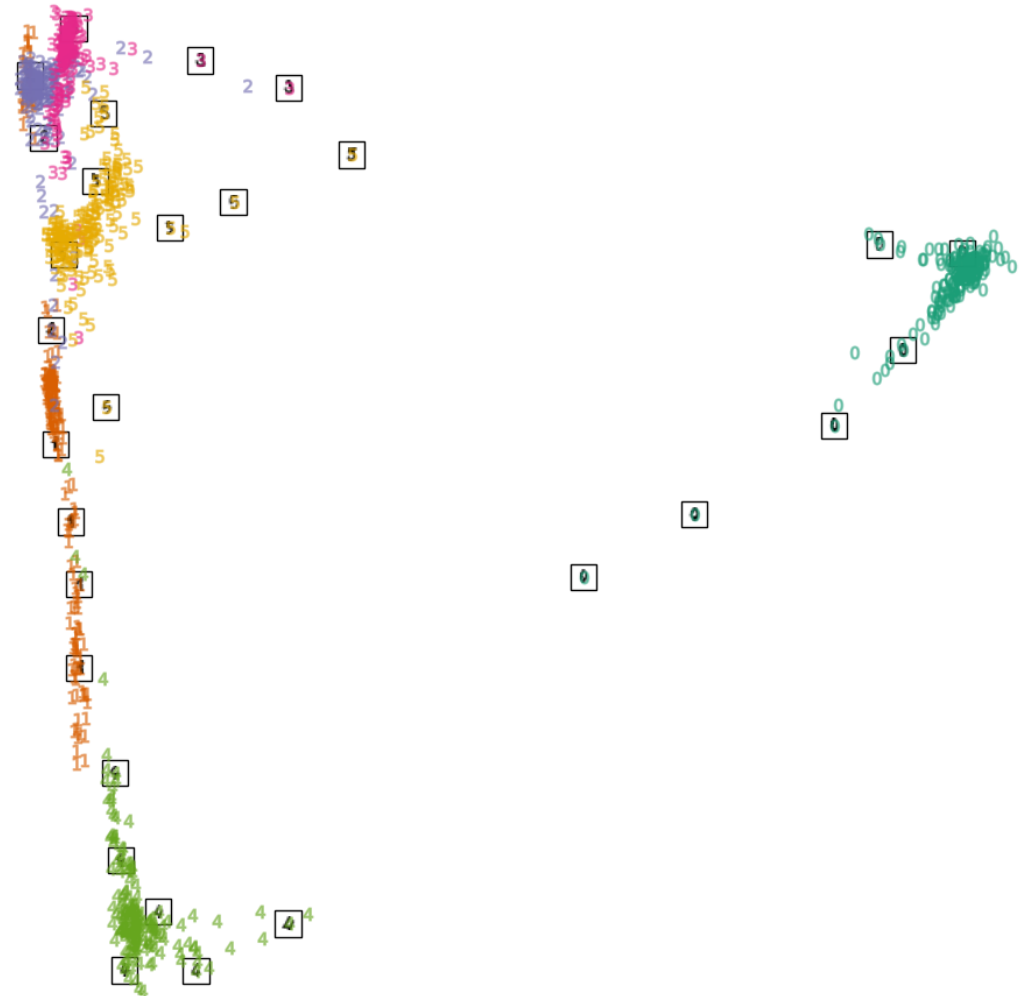


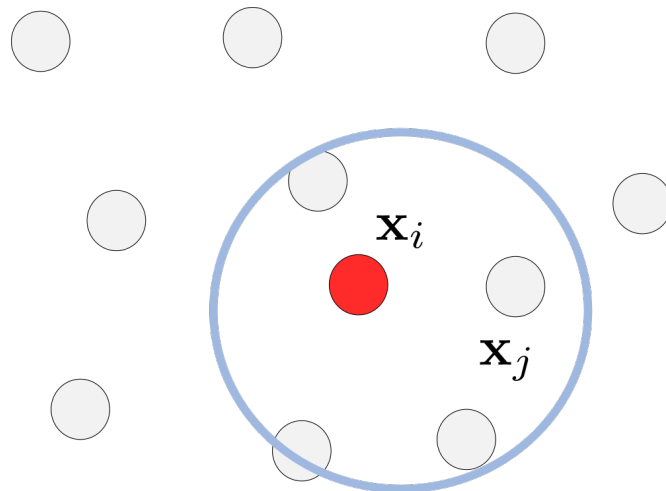
Image Generated running the scikit-learn demo for Manifold Learning on MNIST:

[https://scikit-learn.org/stable/auto\\_examples/manifold/plot\\_lle\\_digits.html#sphx-glr-auto-examples-manifold-plot-lle-digits-py](https://scikit-learn.org/stable/auto_examples/manifold/plot_lle_digits.html#sphx-glr-auto-examples-manifold-plot-lle-digits-py)

# Stochastic Neighborhood Embedding

## Neighborhood Embedding Basics:

- Consider the neighborhood around an input data point  $\mathbf{x}_i \in \mathbb{R}^d$
- Let us assume that we have a Gaussian distribution centered around  $\mathbf{x}_i$
- Then the probability that  $\mathbf{x}_i$  chooses some other datapoint  $\mathbf{x}_j$  as its neighbor is **proportional to the density under this Gaussian**
- A point closer to  $\mathbf{x}_i$  will be more likely than one further away



# Stochastic Neighborhood Embedding

We want to convert high-dimensional data into conditional probabilities that represent similarities:

- **Similarities of points in high-dimensional space:**

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma_i^2)},$$

- **Similarities of points in low-dimensional space:**

$$q_{j|i} = \frac{\exp(-\|y_i - y_j\|^2)}{\sum_{k \neq i} \exp(-\|y_i - y_k\|^2)}$$

- **Cost function** to be minimized using gradient descent (KL-Divergence):

$$KL(P \parallel Q) = \sum_i \sum_j p_{j|i} \log\left(\frac{p_{j|i}}{q_{j|i}}\right)$$

# Symmetric Stochastic Neighborhood Embedding

We want to minimize the KL-divergence between **joint probability distributions**:

- **Similarities of points in high-dimensional space:**

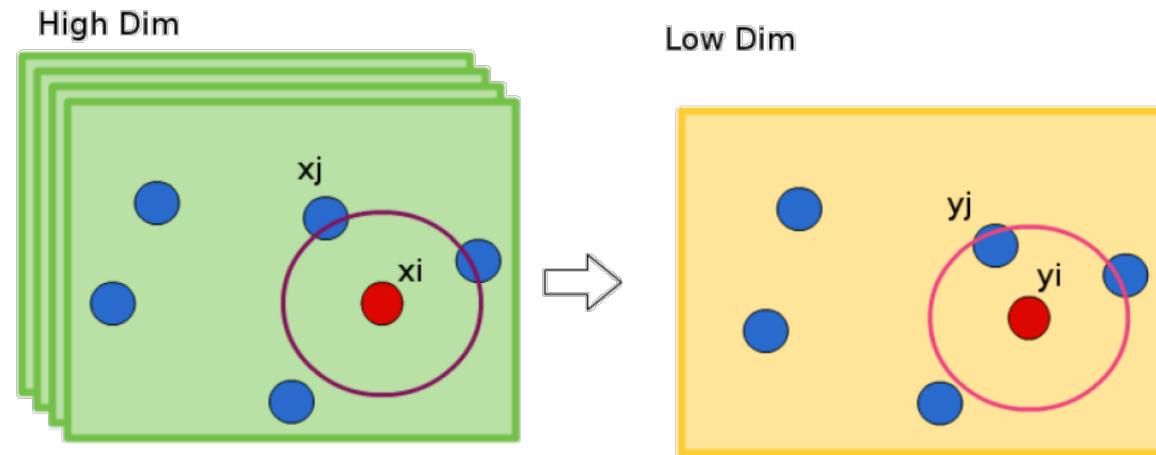
$$p_{ij} = \frac{\exp(-\| \mathbf{x}_i - \mathbf{x}_j \|^2 / 2\sigma^2)}{\sum_{k \neq i} \exp(-\| \mathbf{x}_i - \mathbf{x}_k \|^2 / 2\sigma^2)}$$

We can **symmetrize** the similarity as follows:

$$p_{ij} = \frac{p_{i|j} + p_{j|i}}{2N}$$

# t-Distributed Stochastic Neighbor Embedding

Measure pairwise similarities between high-dimensional and low-dimensional objects



- Similarity of datapoints in High Dimension

$$p_{ij} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma^2)}$$

- Similarity of datapoints in Low Dimension

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq i} (1 + \|y_i - y_k\|^2)^{-1}}$$

Heavy tail Student t-distribution  
William Sealy Gosset

$$C = \sum_i KL(P_i \| Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}}$$

# t-SNE - Impact of Perplexity

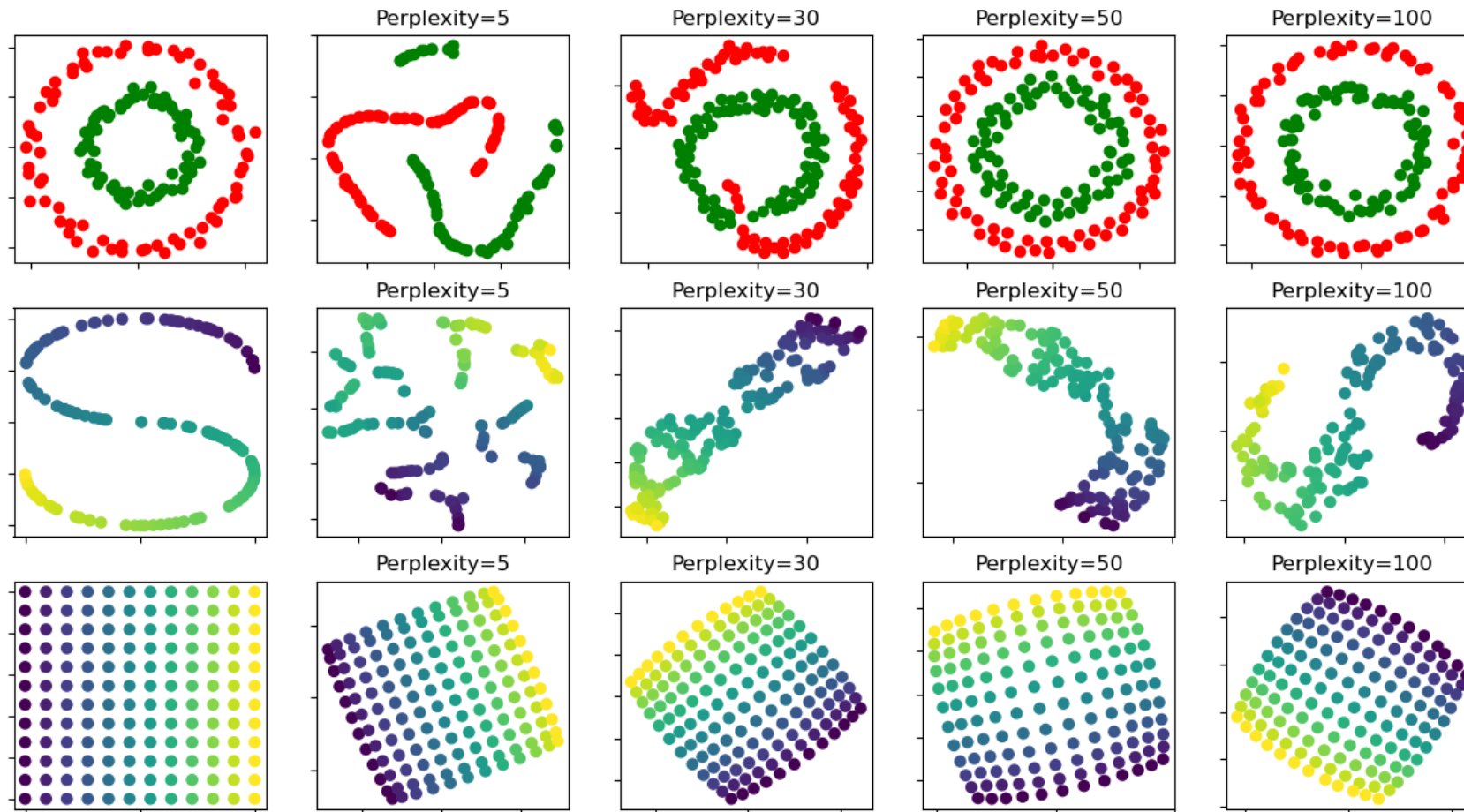
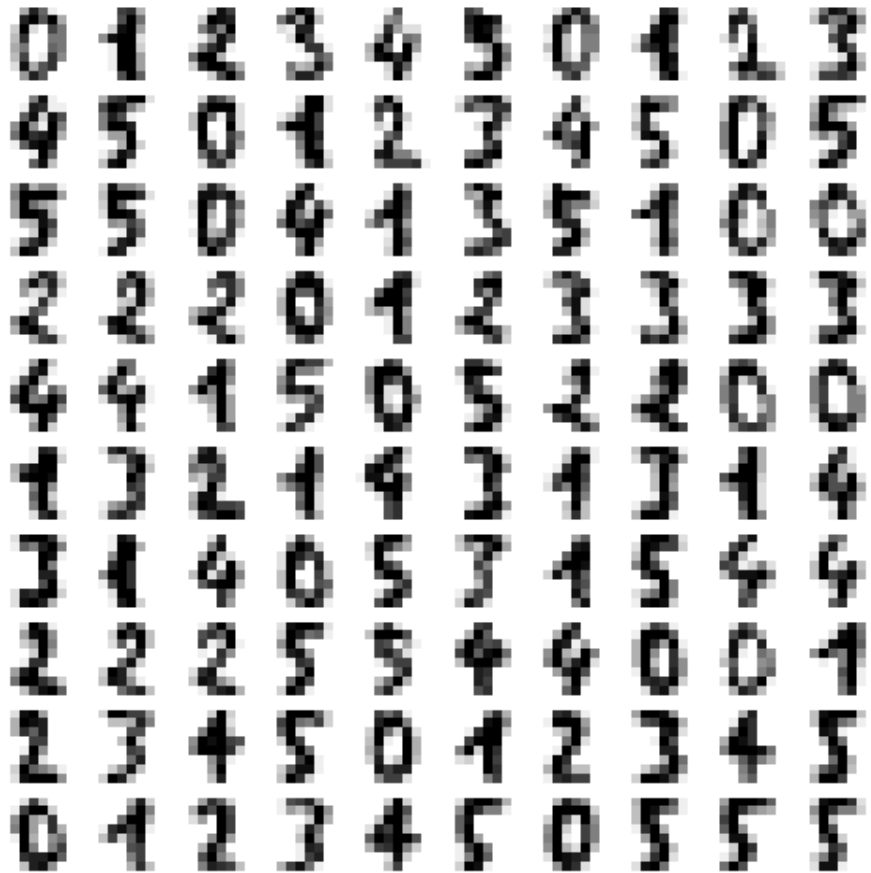


Image Generated running the code from:

[https://scikit-learn.org/stable/auto\\_examples/manifold/plot\\_t\\_sne\\_perplexity.html](https://scikit-learn.org/stable/auto_examples/manifold/plot_t_sne_perplexity.html)

# t-SNE - MNIST

A selection from the 64-dimensional digits dataset



t-SNE embedding (time 1.483s)

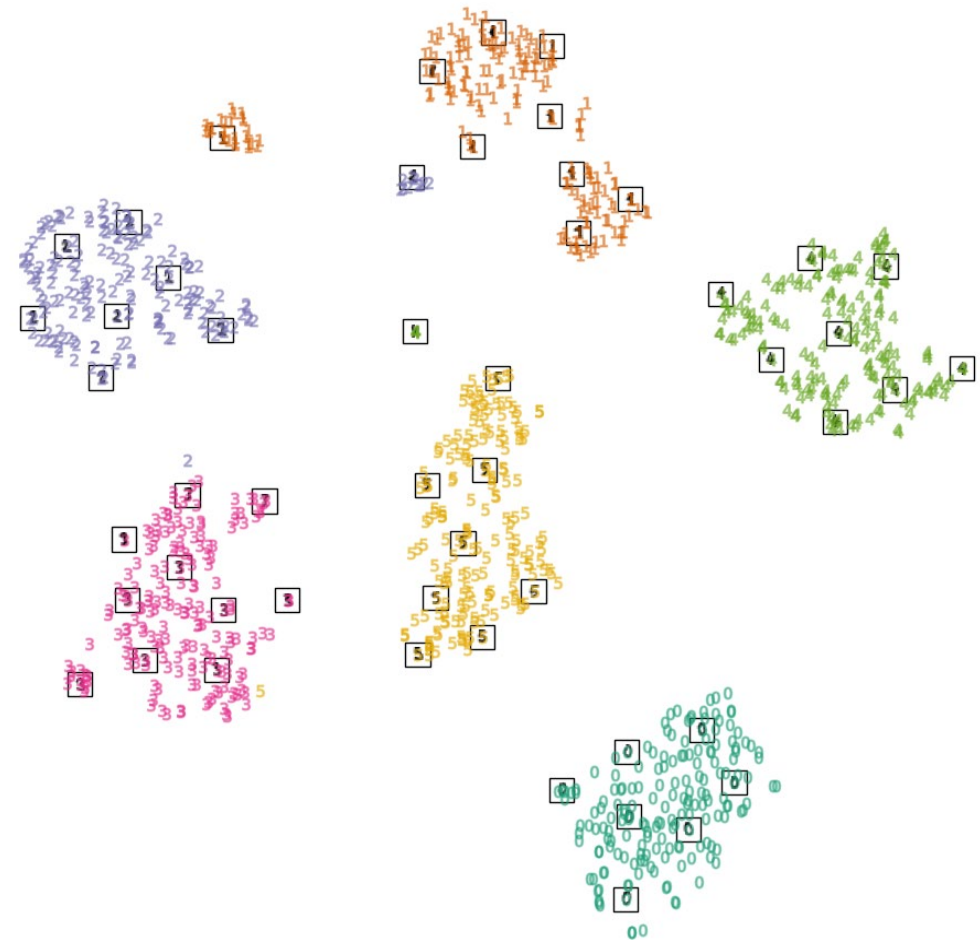
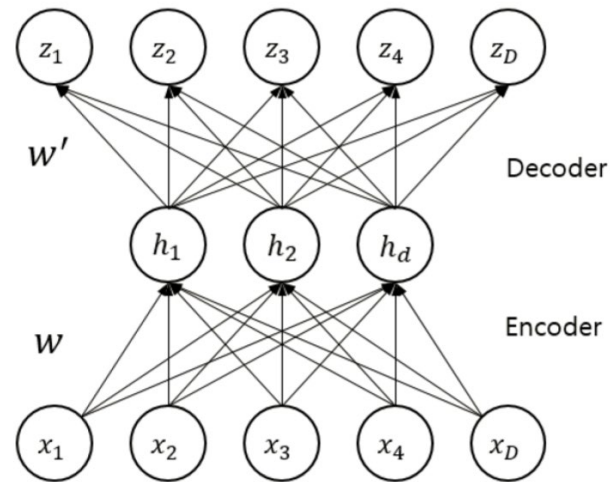


Image Generated running the scikit-learn demo for Manifold Learning on MNIST:

[https://scikit-learn.org/stable/auto\\_examples/manifold/plot\\_lle\\_digits.html#sphx-glr-auto-examples-manifold-plot-lle-digits-py](https://scikit-learn.org/stable/auto_examples/manifold/plot_lle_digits.html#sphx-glr-auto-examples-manifold-plot-lle-digits-py)

# Autoencoders

- An autoencoder is a feedforward neural network trained to reproduce its input at the output layer



- Encoder:

$$\mathbf{h} = f_{\theta}(\mathbf{x}) = a(\mathbf{W}\mathbf{x} + \mathbf{b})$$

- Decoder:

$$\mathbf{z} = g_{\theta'}(\mathbf{h}) = a(\mathbf{W}'\mathbf{h} + \mathbf{b}')$$

$$\theta, \theta' : \{\mathbf{W}, \mathbf{b}\}, \{\mathbf{W}', \mathbf{b}'\}$$

- Parameter estimation
- Back-propagation

$$\theta^*, \theta'^* = \operatorname{argmin}_{\theta, \theta'} \frac{1}{n} \sum_{i=1}^n L(\mathbf{x}^{(i)}, \mathbf{z}^{(i)})$$

$$= \operatorname{argmin}_{\theta, \theta'} \frac{1}{n} \sum_{i=1}^n L(\mathbf{x}^{(i)}, g_{\theta'}(f_{\theta}(\mathbf{x}^{(i)})))$$

# Today: Topological Data Analysis (TDA)

Slides ack: Afra Zomorodian, Ryan Lewis,  
Gunnar Carlsson, Samir Chowdhury

# Topology and Data Analysis

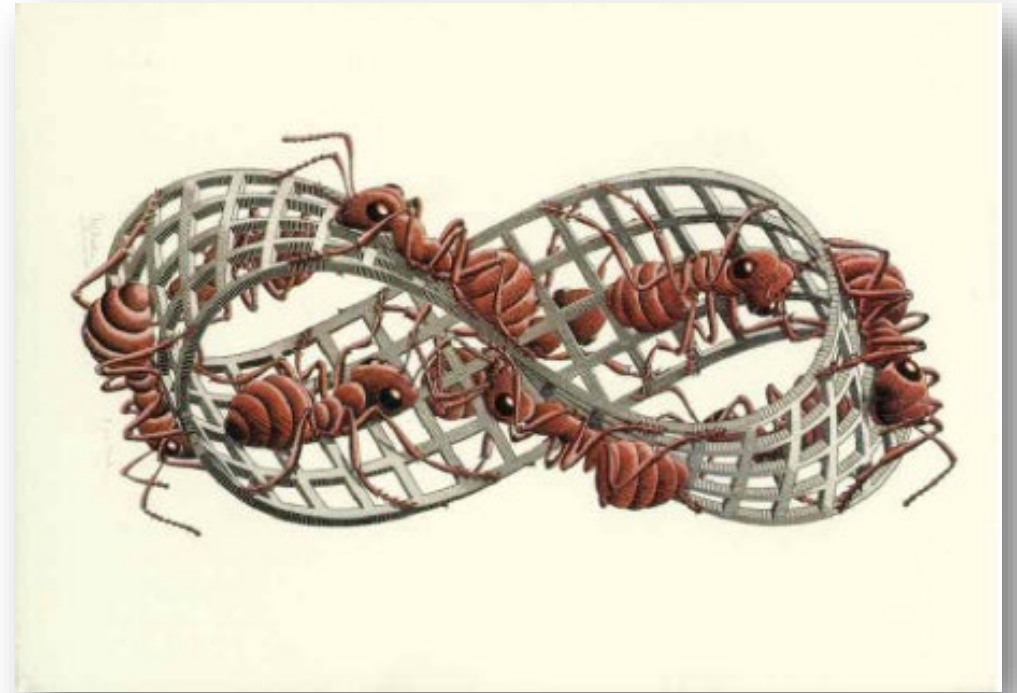
Topology — Computational Topology



Homology — **Persistent Homology**



Topological Data Analysis Applications



# Computational Topology and Data Analysis



Herbert Edelsbrunner



Gunnar Carlsson

1990 -- 2022



Afra Zomorodian

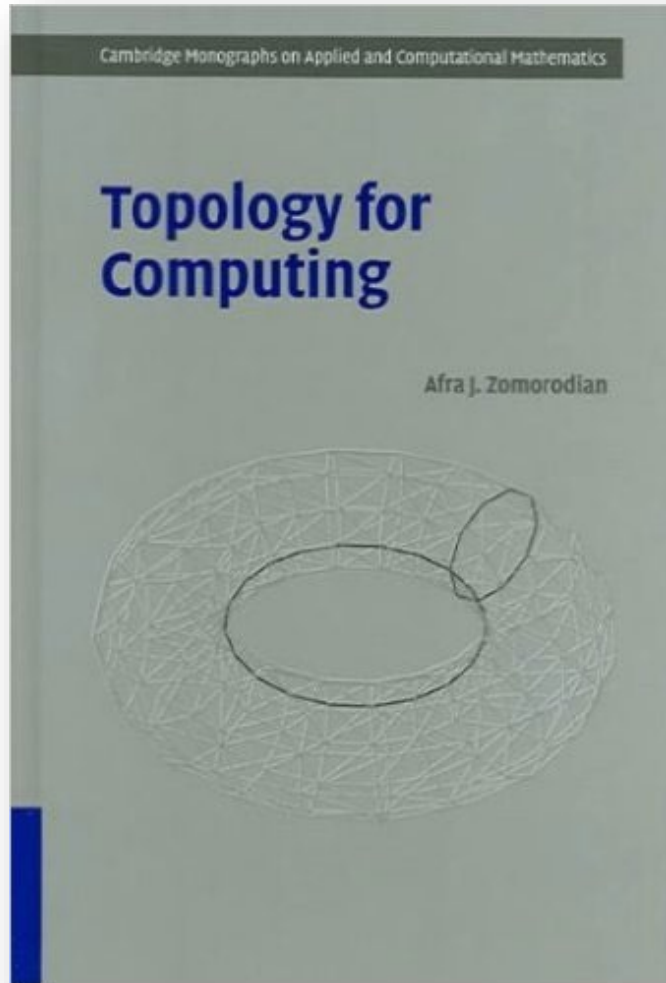


Robert Ghrist

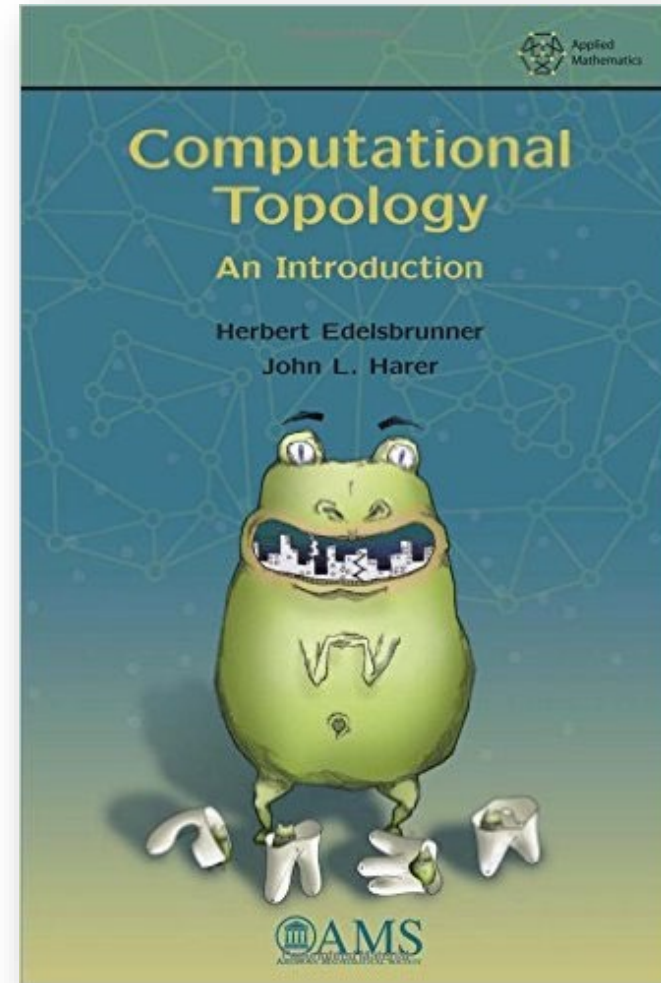


Frederic Chazal

# Some Textbooks

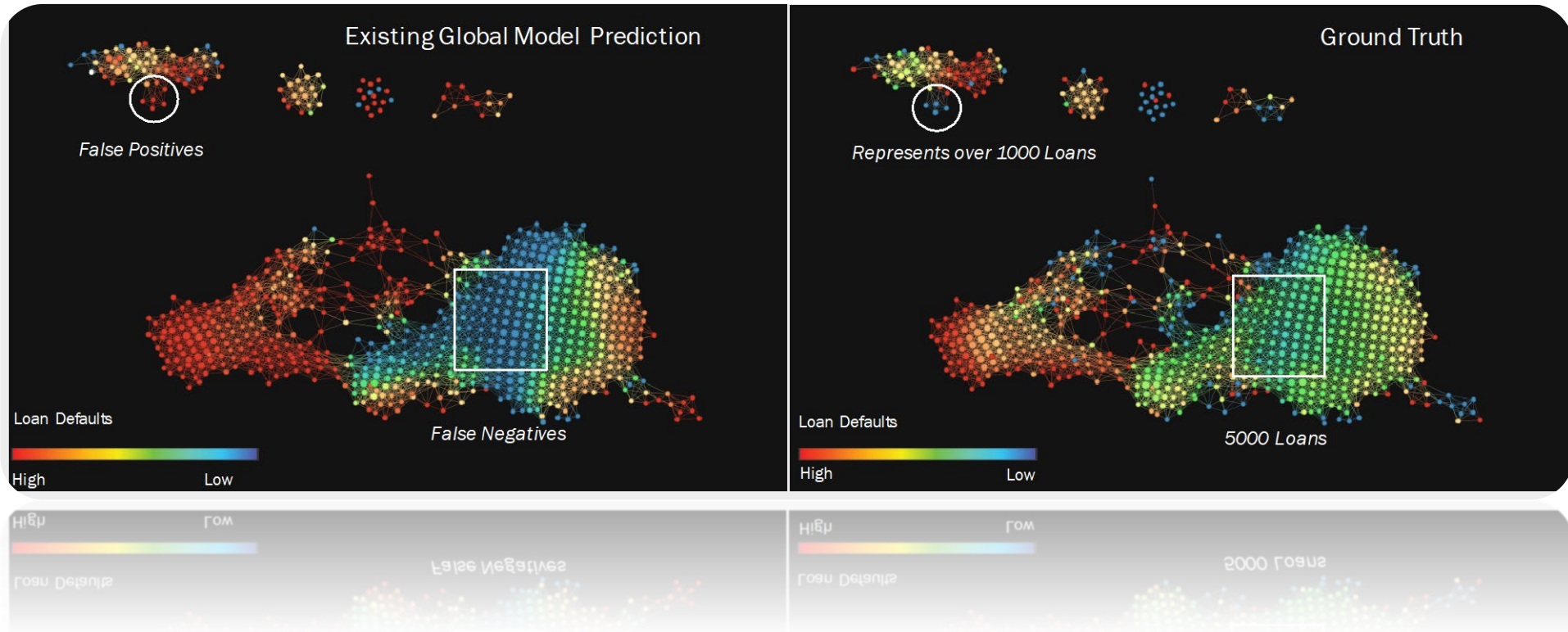


2001



2010

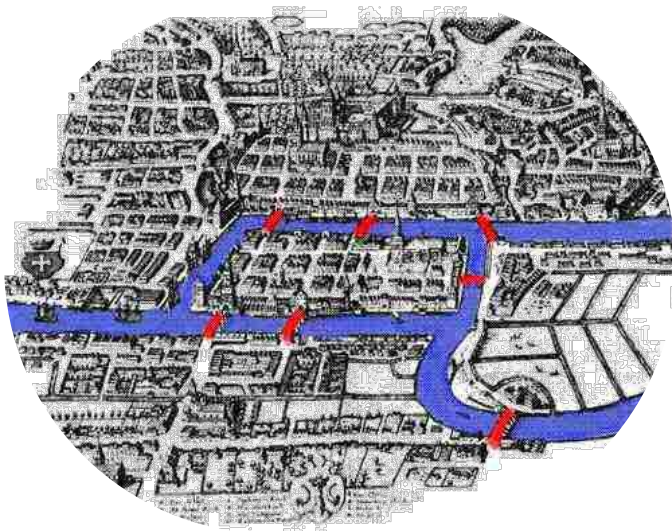
# Some Companies



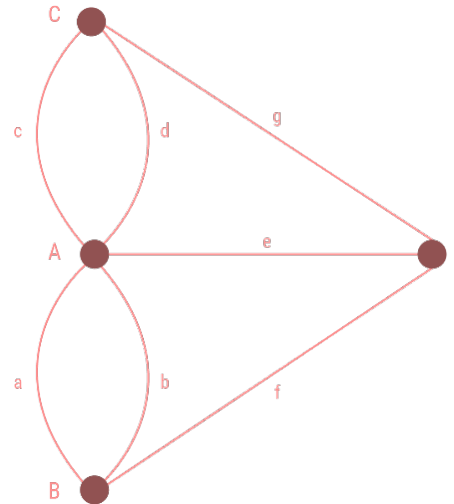
Ayasdi (Now part of SymphonyAI)

# Topology Studies Connectivity

... and obstructions to connectivity

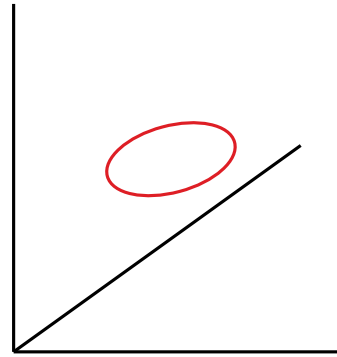
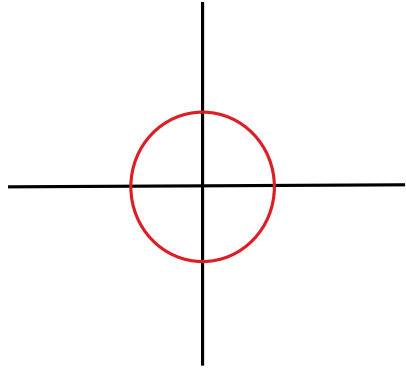


The bridges of  
Königsberg (Kalingrad)



Leonhard Euler  
1736

# Topological (Connectivity) Invariances

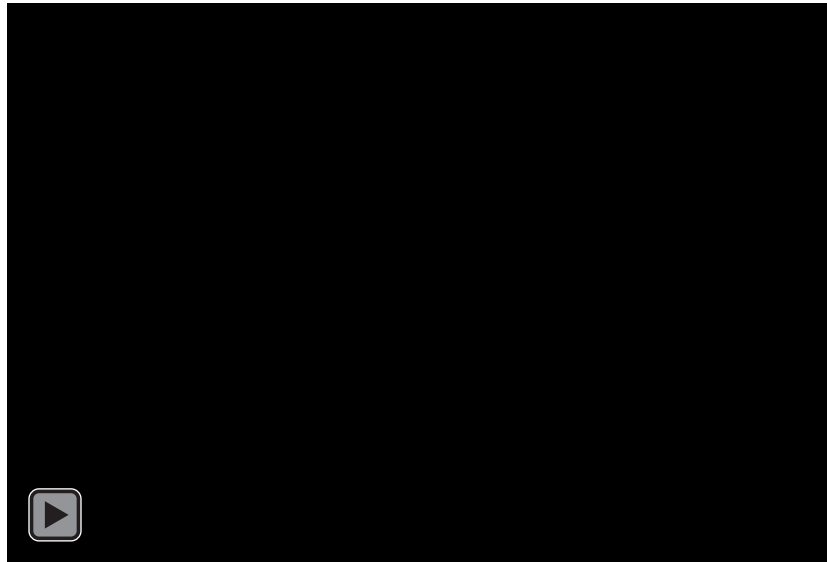


A

*A*

Coordinate-free

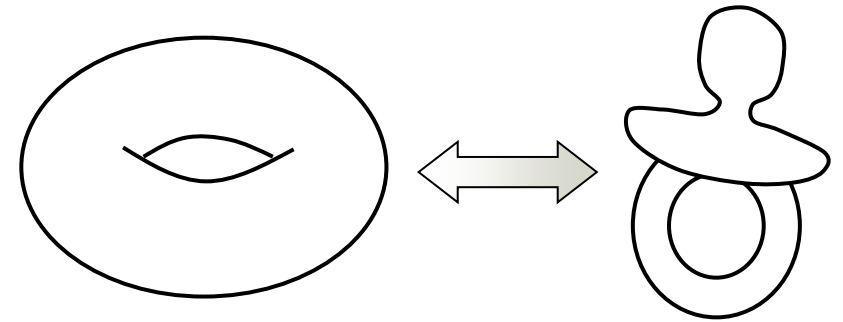
Deformation-invariant



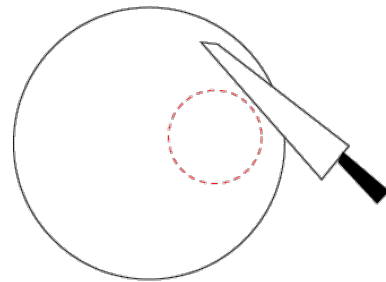
Gunnar Carlsson: Topology does not take distances too seriously

# 2-Manifolds: Ordinary Surfaces

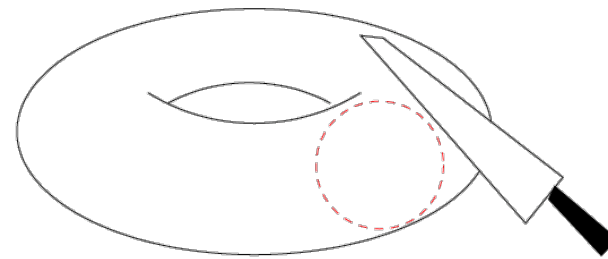
- We are allowed to stretch and shrink
  - Homeomorphism: 1-1, onto, bi-continuous



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



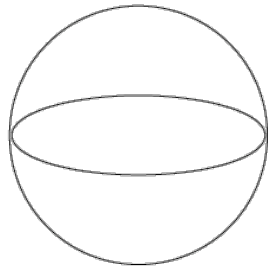
(a) No matter where we cut the sphere, we get two pieces



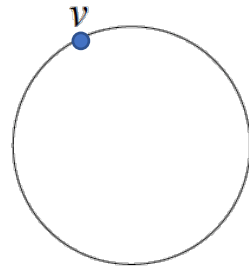
(b) If we're careful, we can cut the torus and still leave it in one piece.

- Note: connectivity information is indexed by dimension

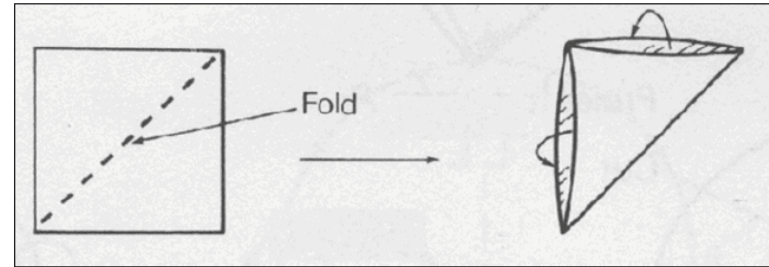
# 2-Manifold Zoo



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

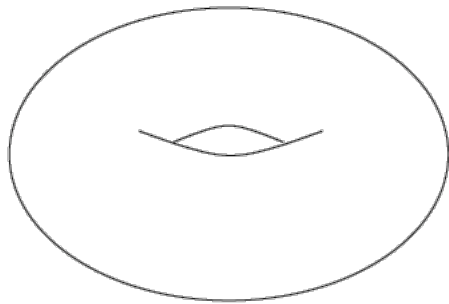


(b) Identify boundary to  $v$

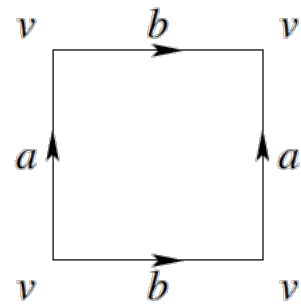


(c) Instructions for a flat sphere

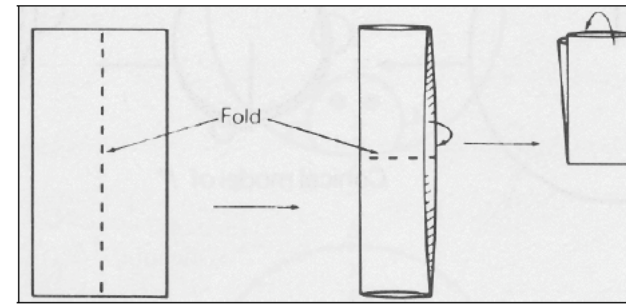
Sphere



(a) Donut surface



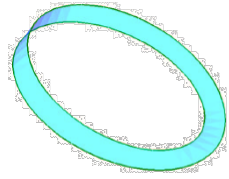
(b) Diagram



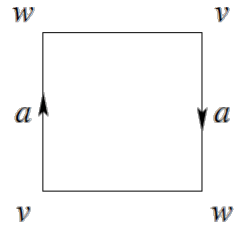
(c) Instructions for a flat torus

Torus

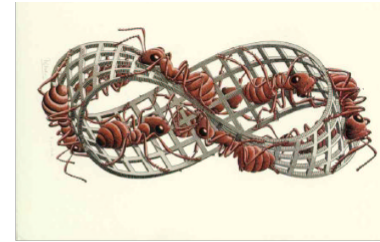
# More Exotic Animals



(a) Embedded

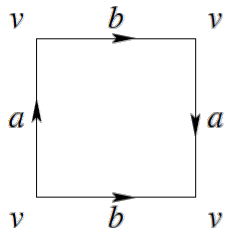


(b) Diagram

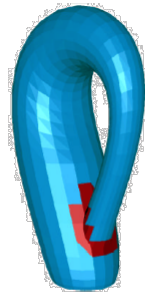


(c) Escher's *Möbius Strip II* (on its side)

Möbius strip



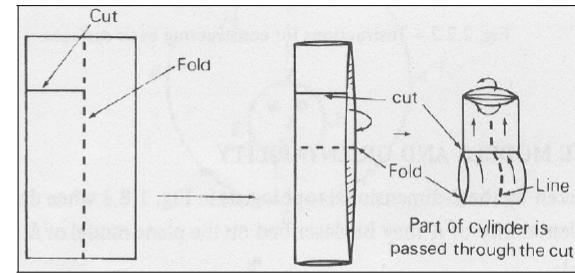
(a) Diagram  
Cross cap + disk



(b) An immersion

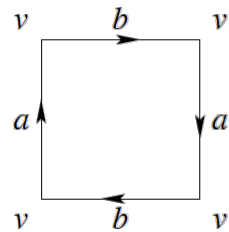
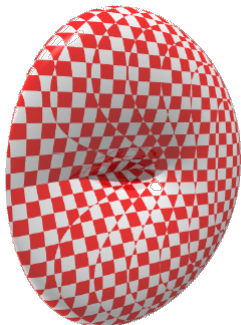


(c) Cut in half (a Möbius strip)

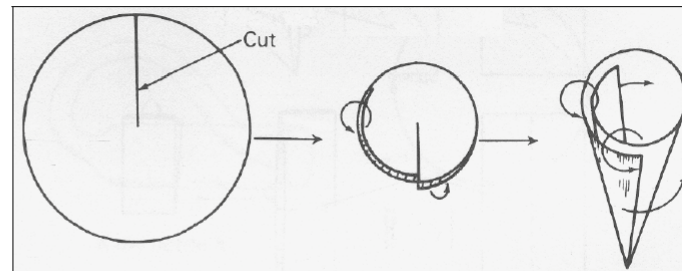


(d) Instructions for a flat  $\mathbb{K}^2$

Klein bottle



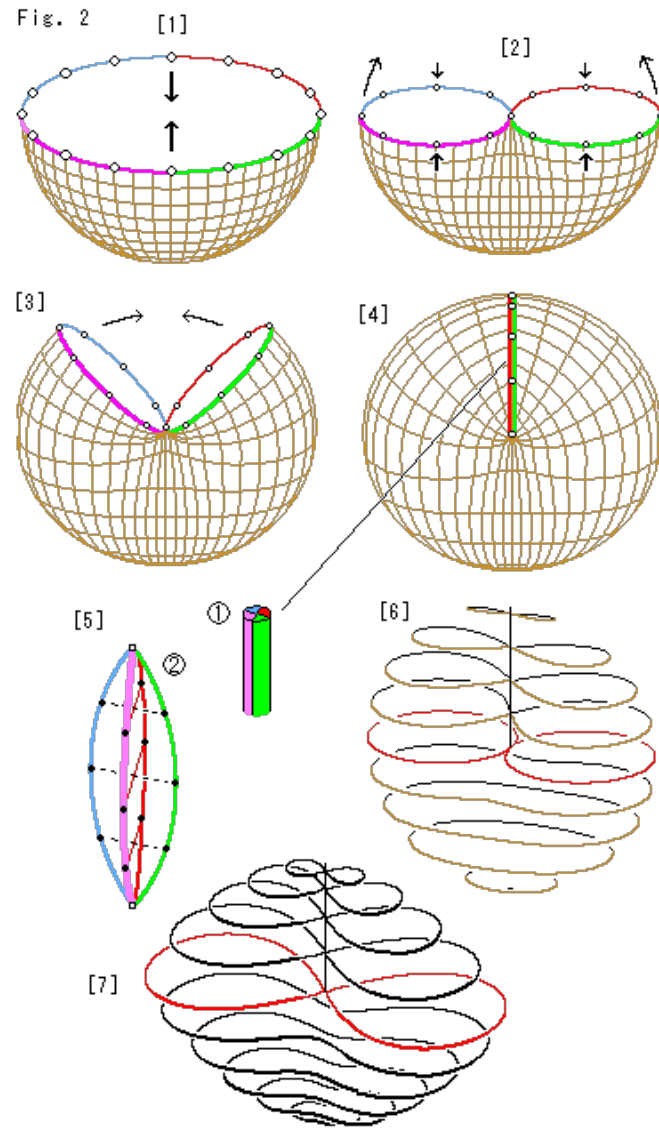
(a) Diagram



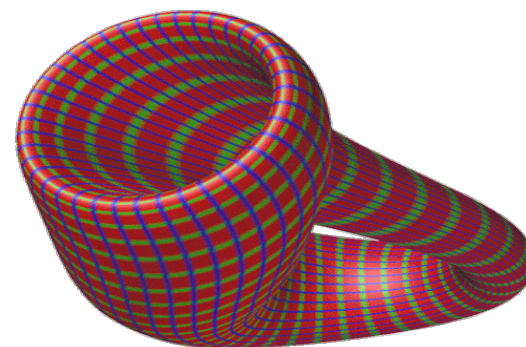
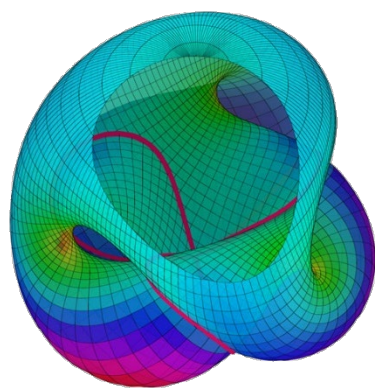
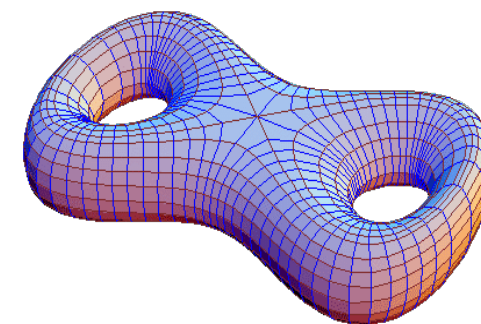
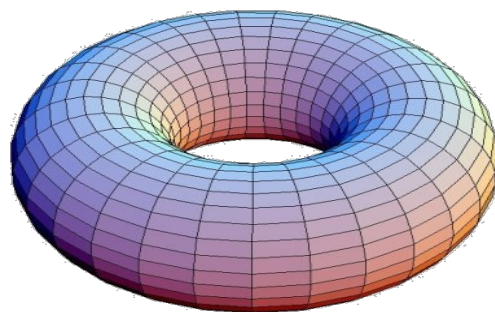
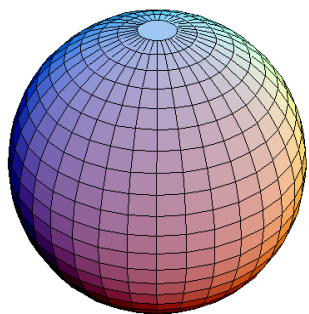
(b) Instructions for a flat  $\mathbb{R}P^2$

Projective plane

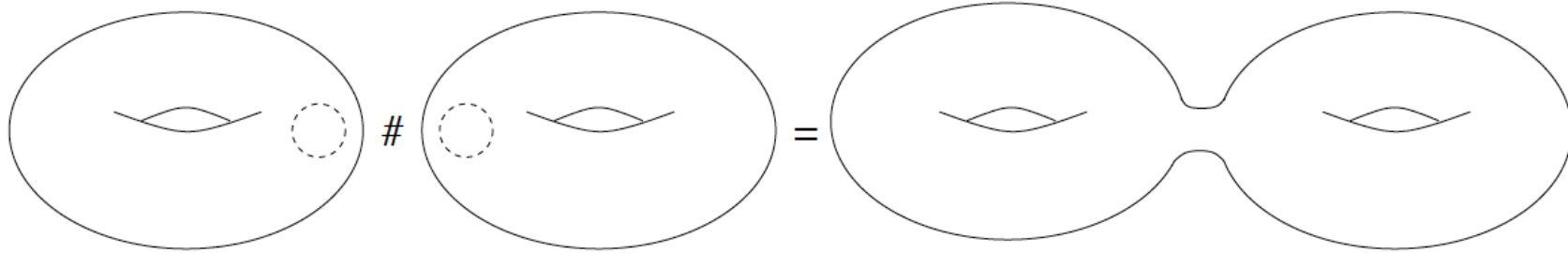
# Projective Plane Detail



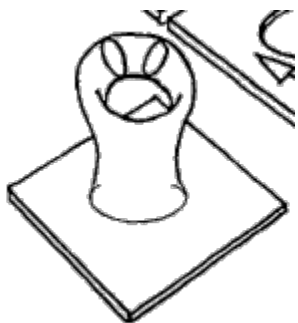
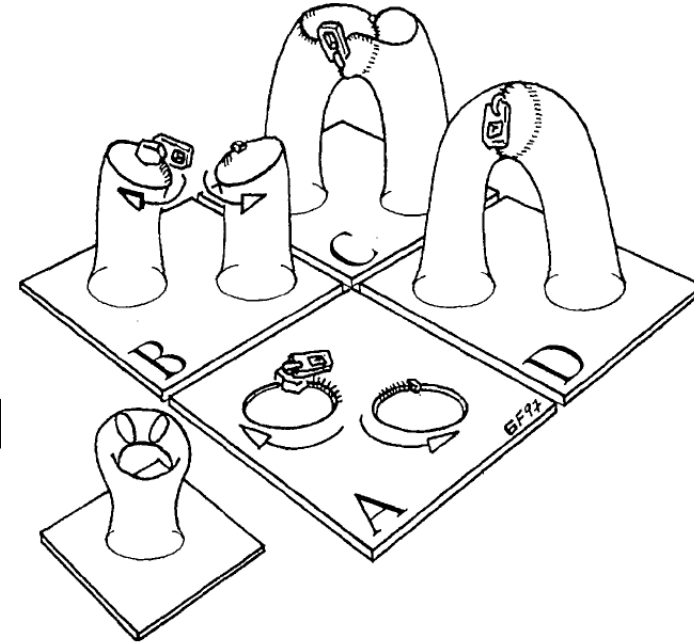
# Topology Aims to Classify



# Connected Sums

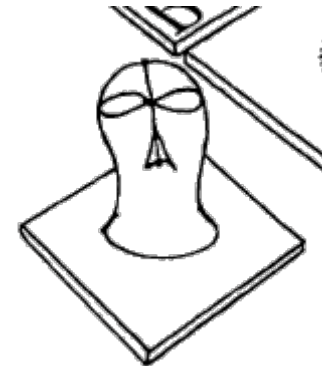


- **Classification Theorem of 2-Manifolds (1860):** Every closed connected compact surface is a connected sum of a sphere with a number of tori and projective planes (sphere + handles + cross cups)



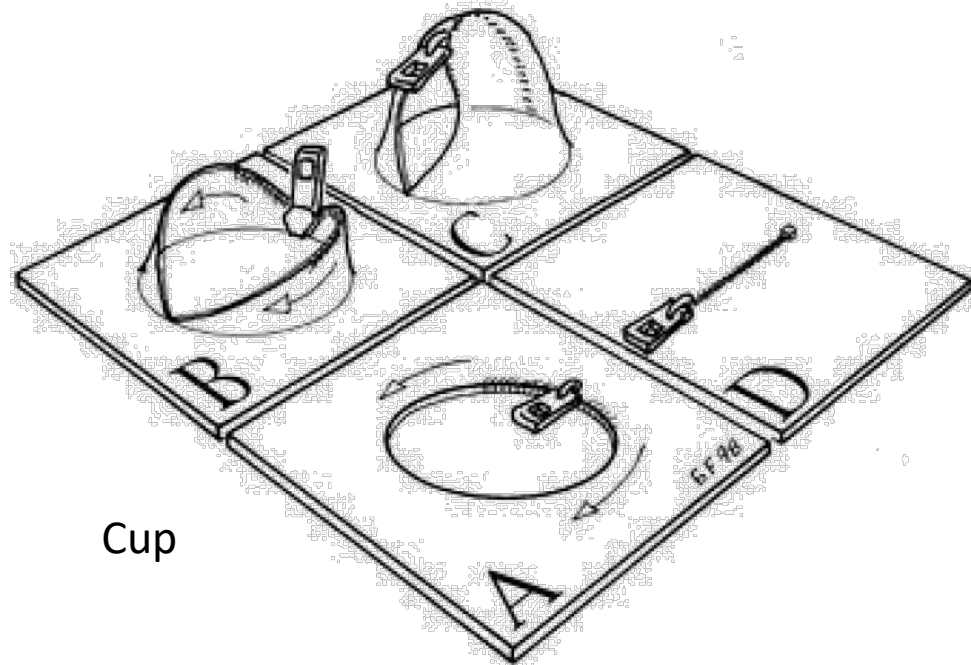
Handle

Klein bottle = sphere + 2 cross cups  
= sphere + cross handle

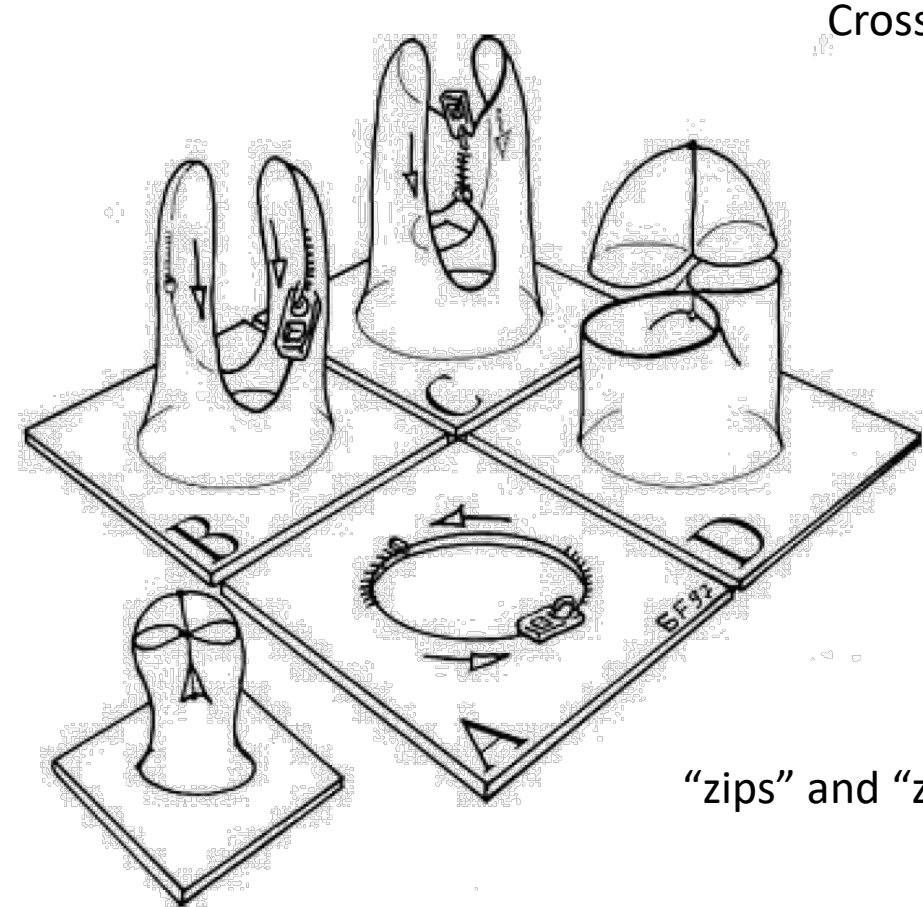


Cross cup

# J. Conway's 1992 ZIP Proof (Zero Irrelevancy Proof)



Cup



Cross cup

"zips" and "zipper"

[Francis and Weeks, AMM, 1999

<https://www.maths.ed.ac.uk/~v1ranick/papers/francisweeks.pdf>]

# Algebraic Topology

Many other branches: Point set topology, differential topology, combinatorial topology



Luitzen Egbertus Jan Brouwer  
1881-1966



Emmy Noether  
1882-1935



Henri Poincaré  
1854-1912



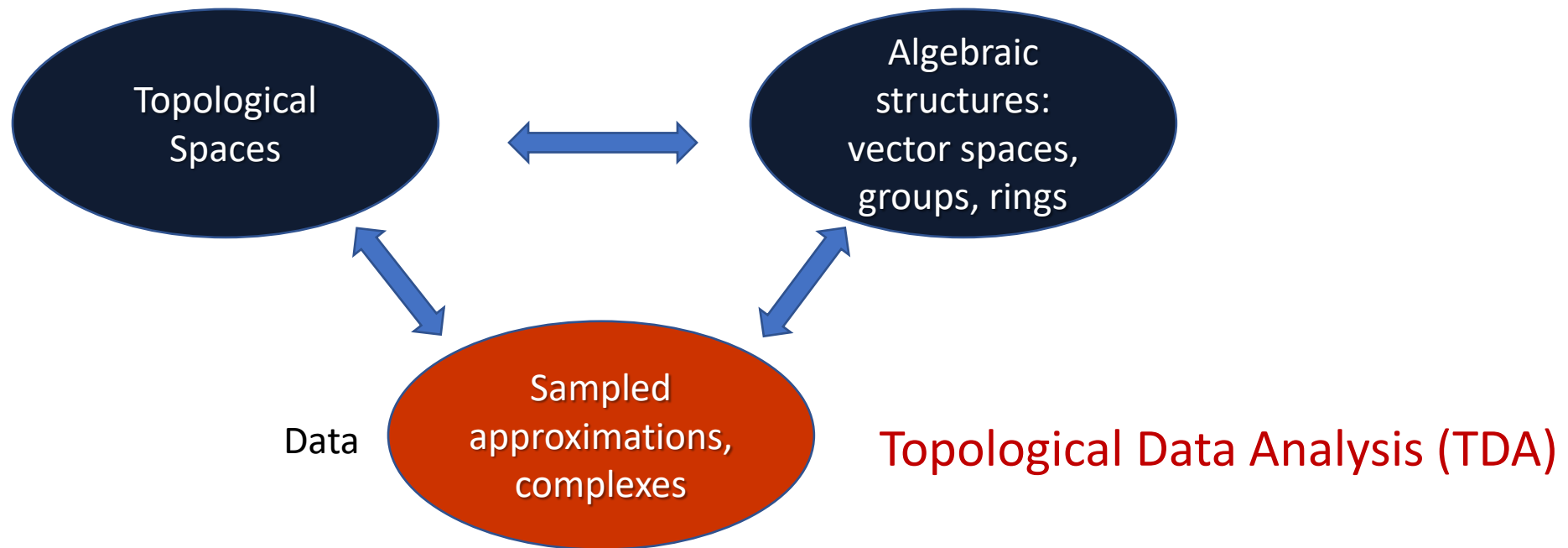
Heinz Hopf  
1894-1971



Leopold Vietoris  
1892-2002

# Algebraic Topology and Topology Inference

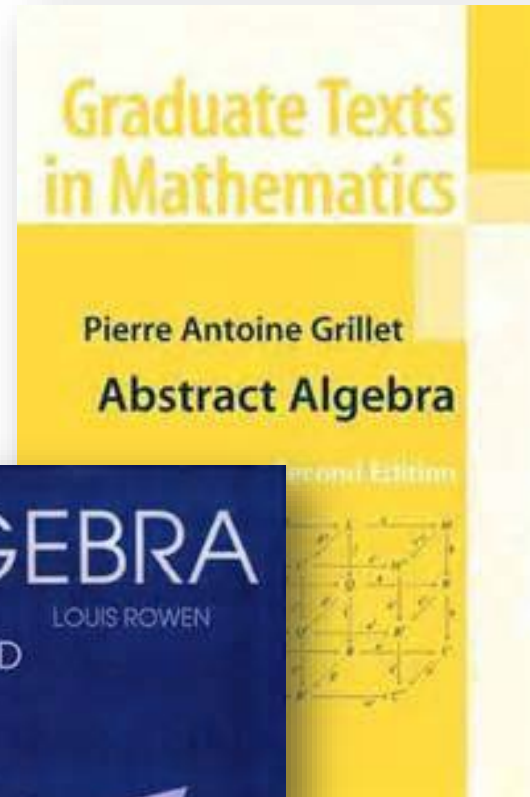
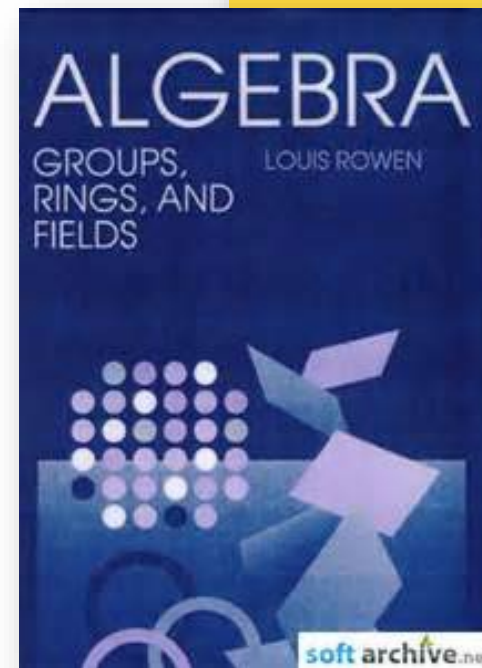
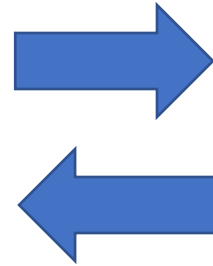
- Unlike geometry, topology studies the **global** structure of spaces
- Getting to this structure via **algebra** algebraic topology



# From Data to Algebraic Objects



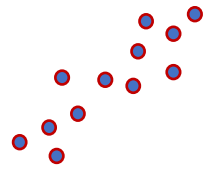
Algebraic entities as  
data descriptors



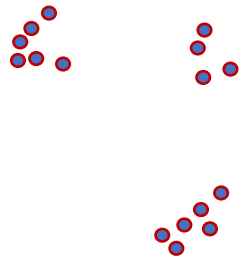
# Studying Sampled Spaces

# The “Shape of Data”

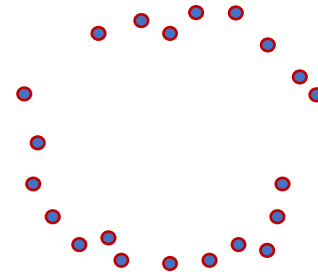
- TDA Studies Sampled Spaces



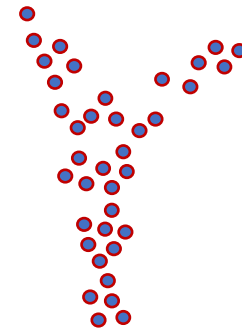
Regression



Cluster

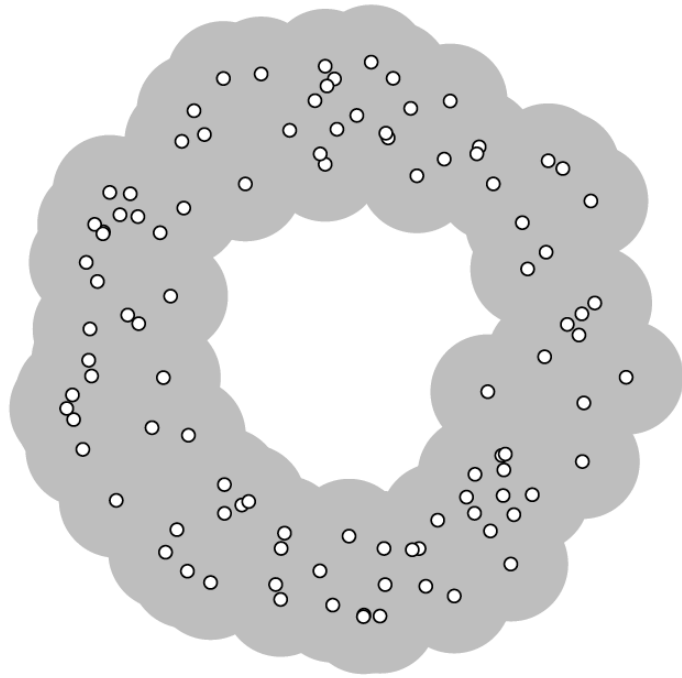


Loop



Flared

# Recovering “Shape” from Sampled Data



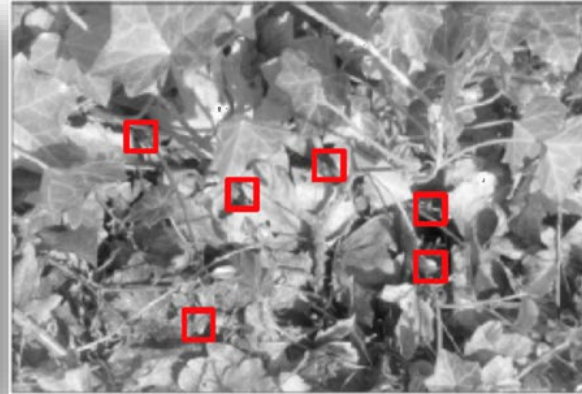
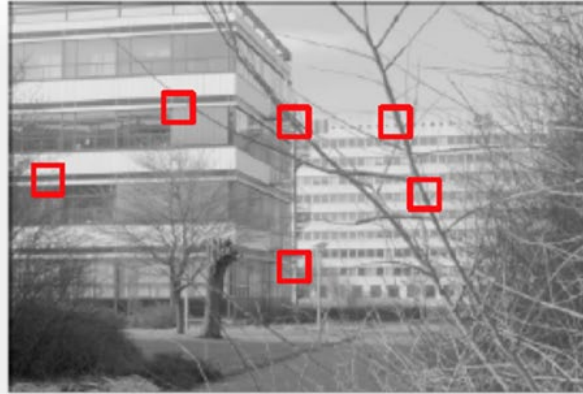
1. Set of points in  $\mathbb{R}^2$
2. Looks like an annulus.

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

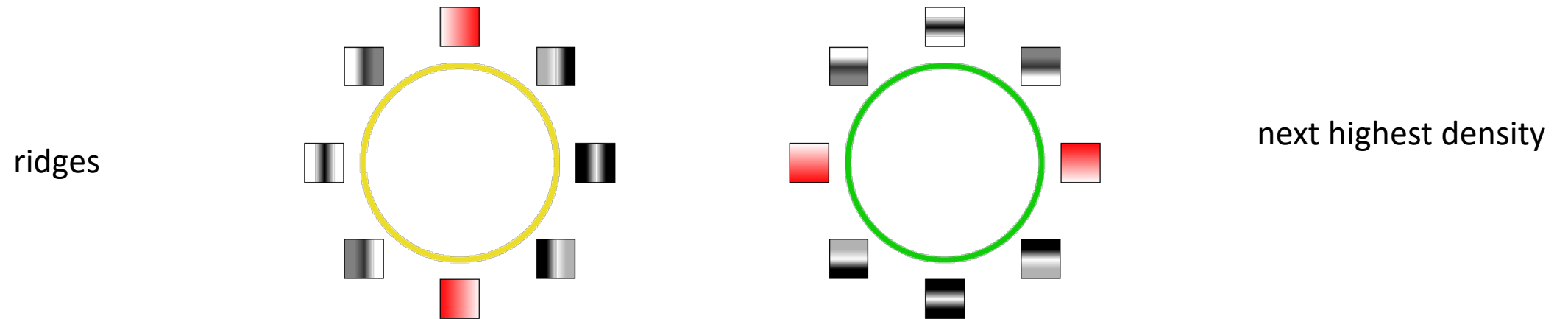
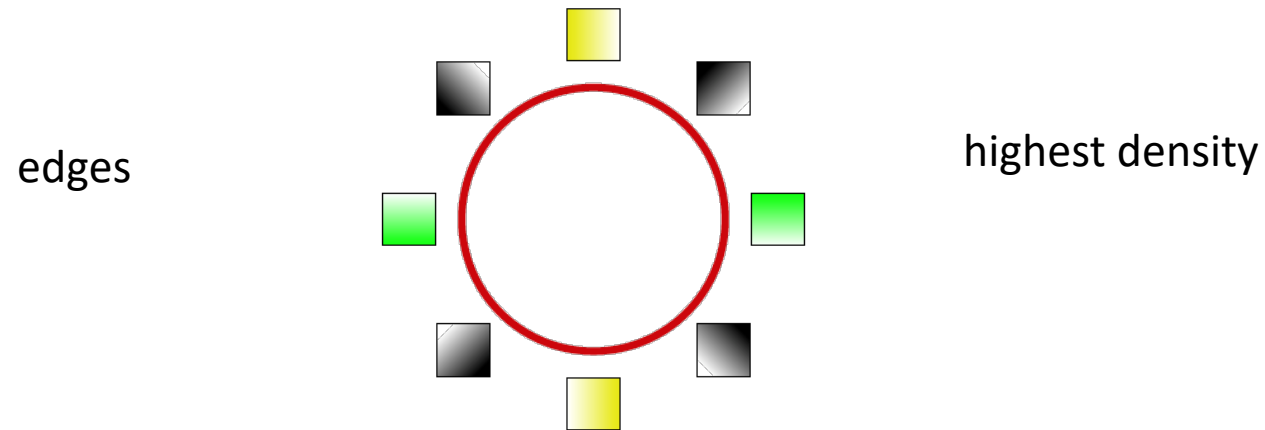
(Carlsson, Ishkanov, de Silva, Zomorodian IJCV 2008)

- Lee-Mumford-Pedersen (2003) 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

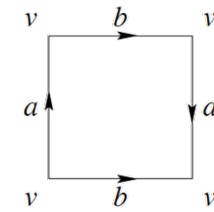
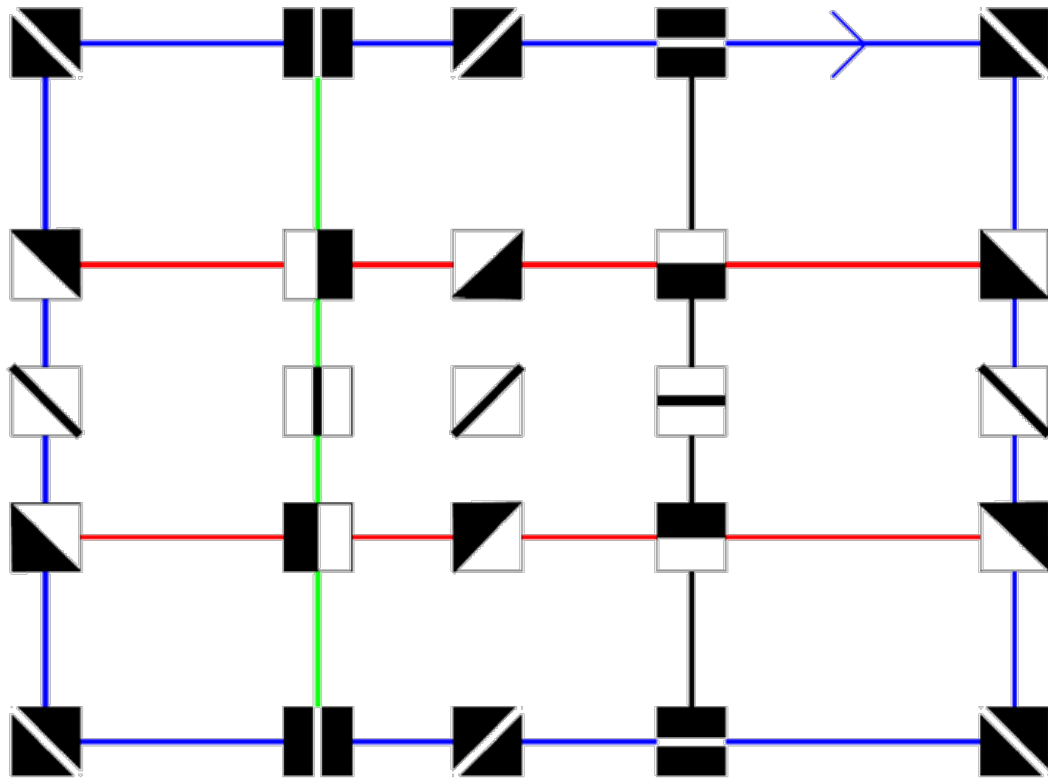
# Natural 3 x 3 Patches

- 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 any real 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



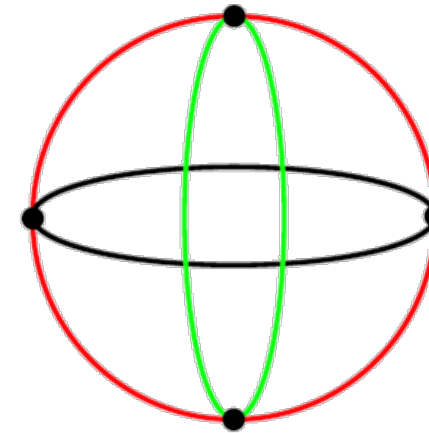
# Klein Bottle Structure



(a) Diagram



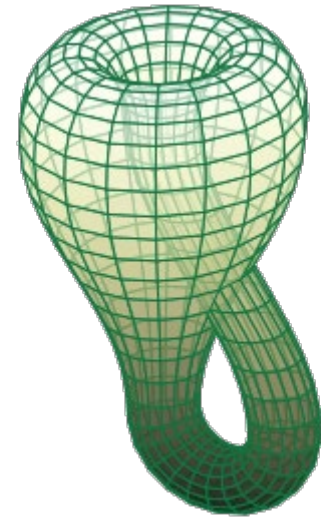
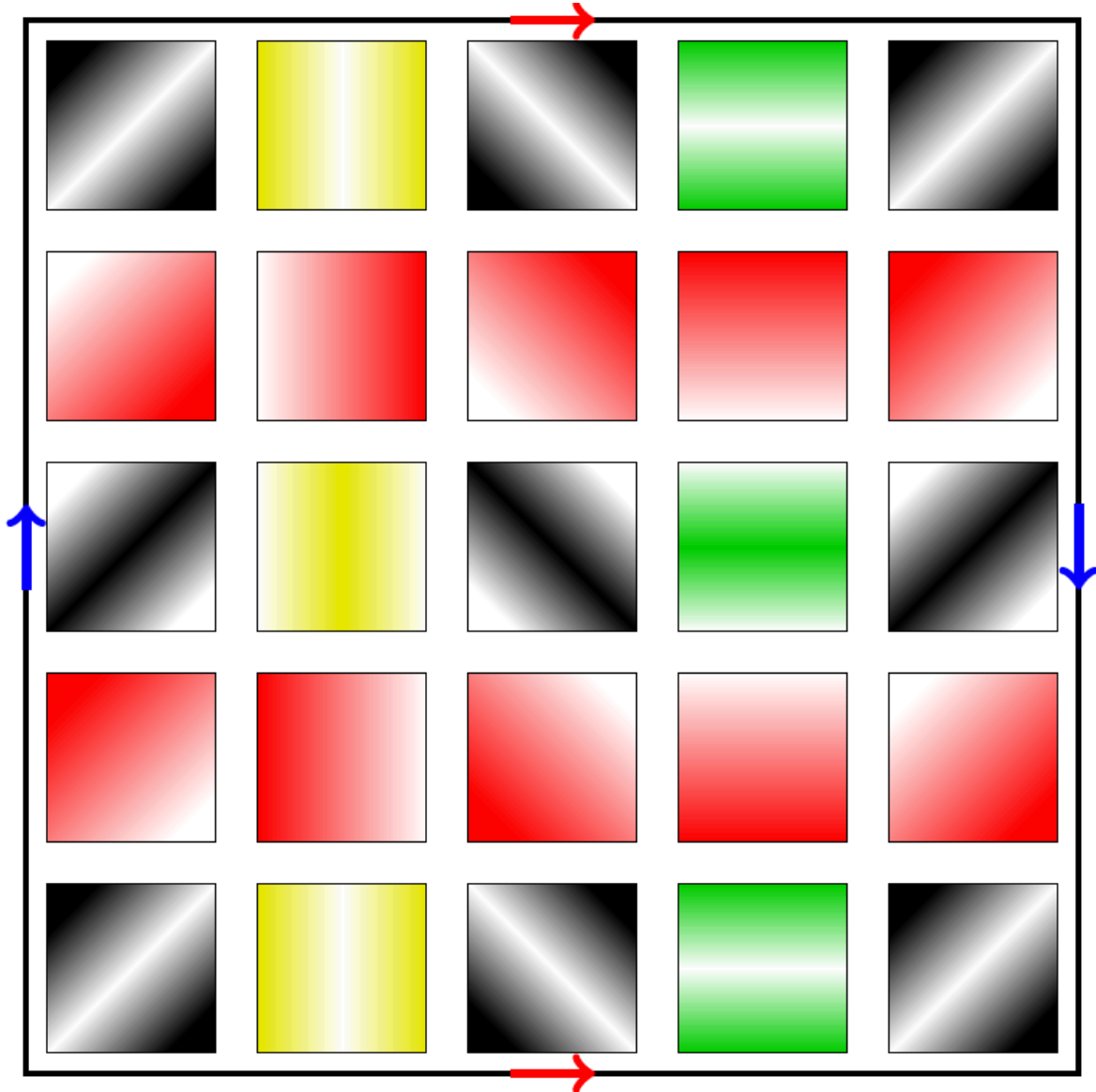
(b) An immersion



$$(\beta_1 = 5)$$

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

# Klein Bottle of Pixel Patches

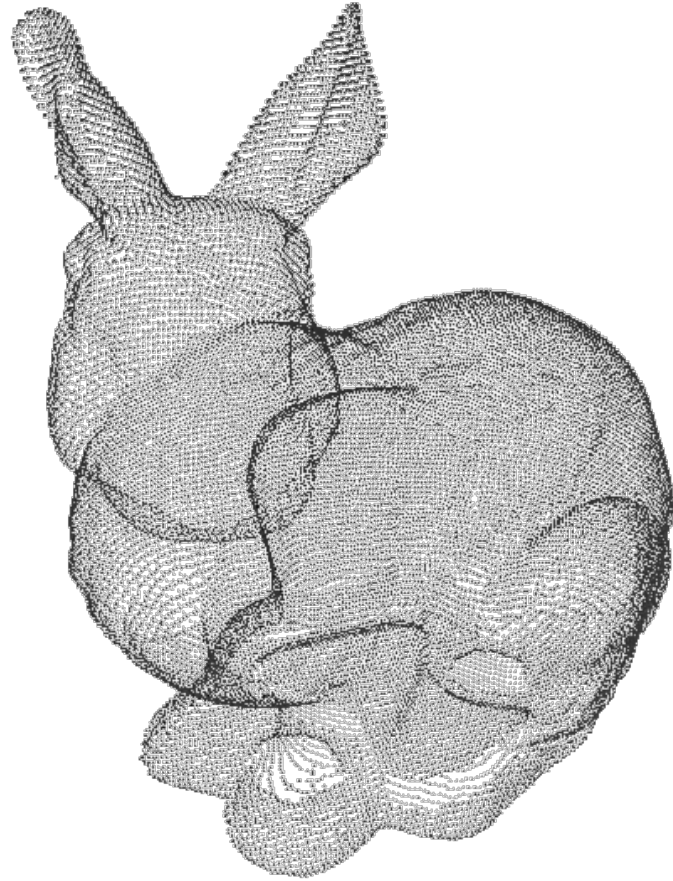


# Applications of the Analysis

- 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 of image convnets

# Simplicial Complexes: Combinatorial Topology

# Point Clouds Have No Higher-D Topology

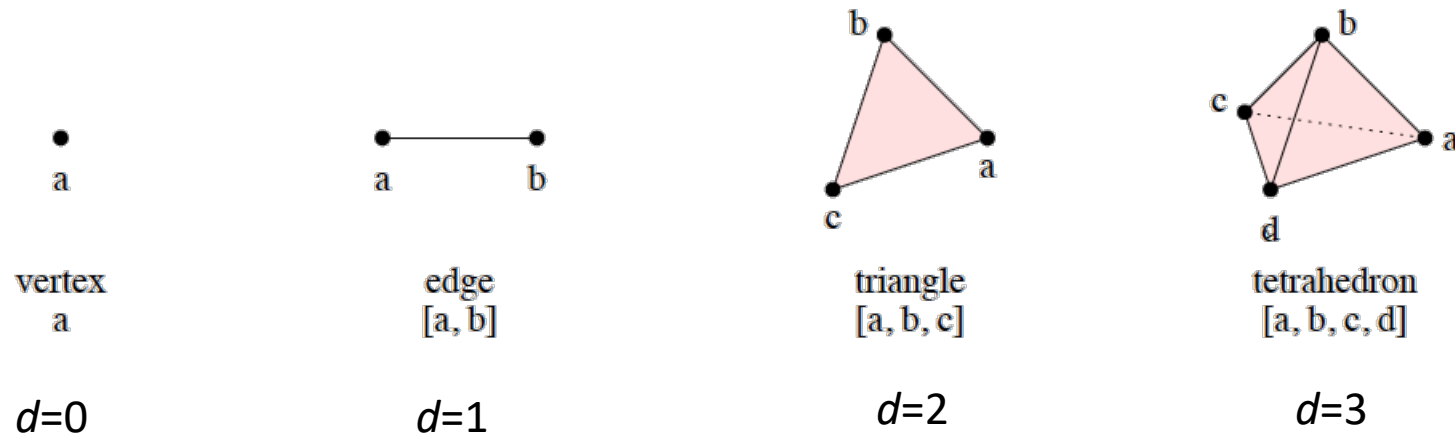


Connect the points into simplicial complexes

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

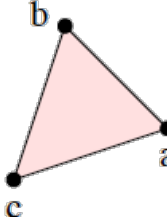
- $\sigma$ : a  $k$ -simplex defined by  $S = \{v_0, v_1, \dots, v_k\}$ .
- $\tau$  defined by  $T \subseteq S$  is a **face** of  $\sigma$
- $\sigma$  is its **coface**.
- $\sigma \geq \tau$  and  $\tau \leq \sigma$ .
- $\sigma \leq \sigma$  and  $\sigma \geq \sigma$ .



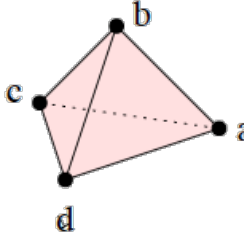
vertex  
a



edge  
[a, b]



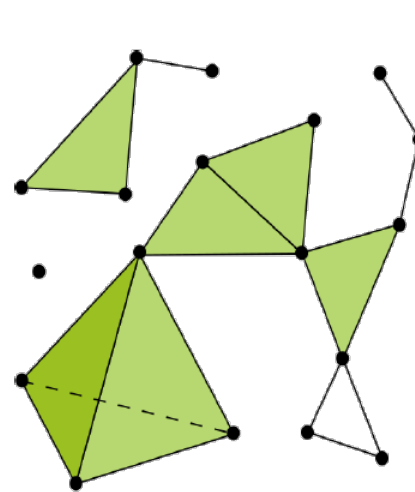
triangle  
[a, b, c]



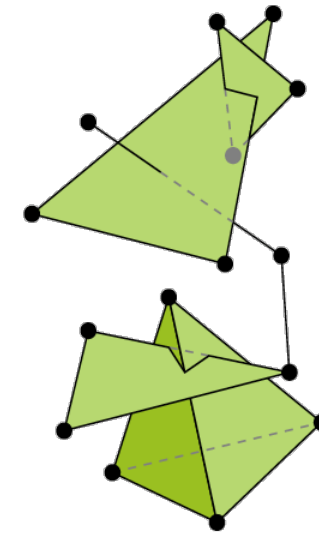
tetrahedron  
[a, b, c, d]

# 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  $K$ .

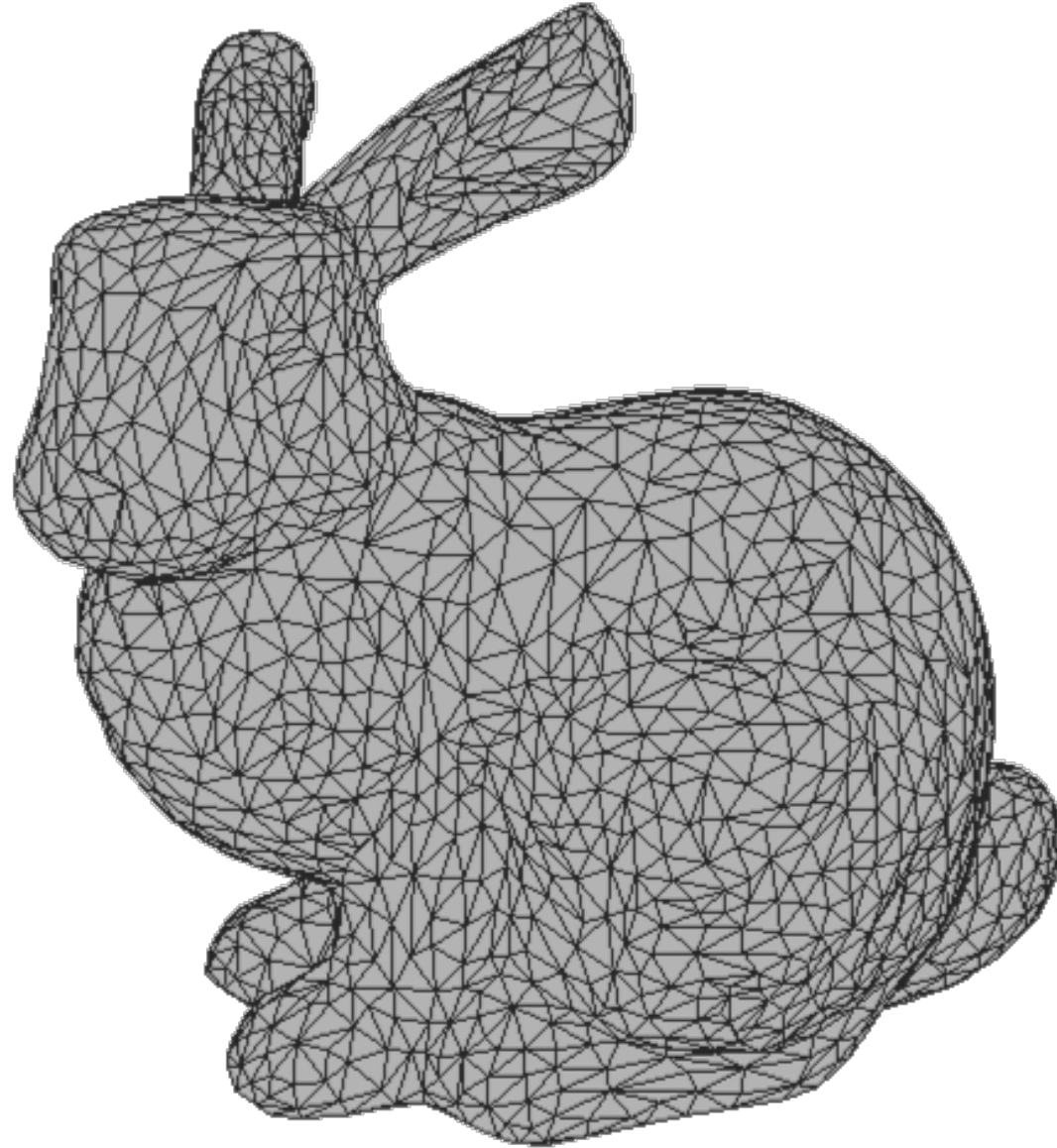


(left) an example



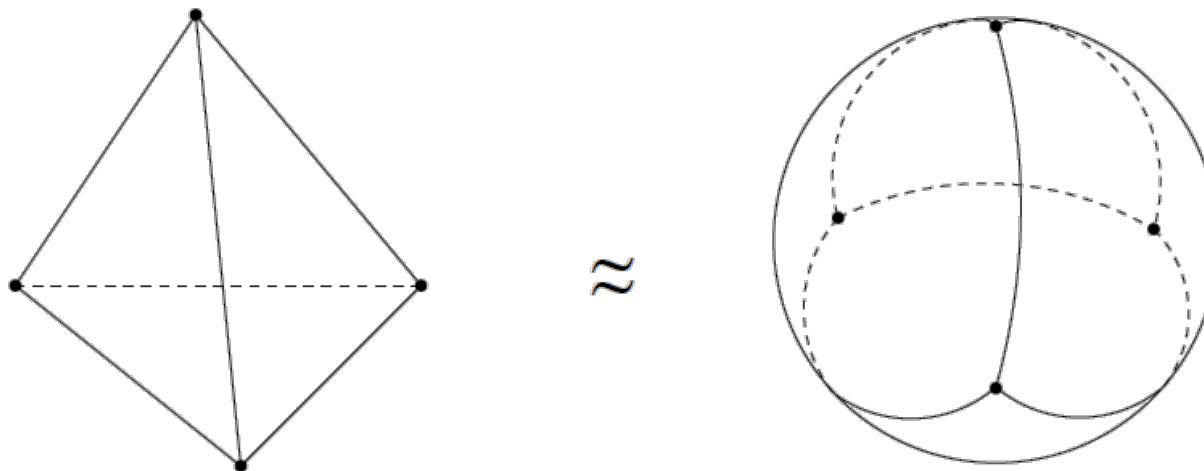
(right) a non example

# Meshes are 2-D Complexes



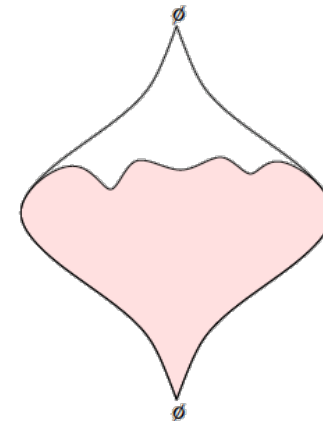
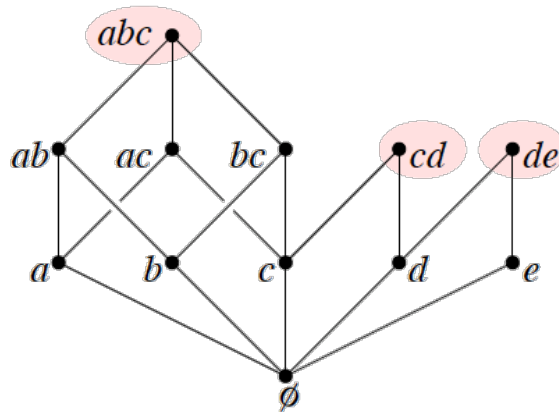
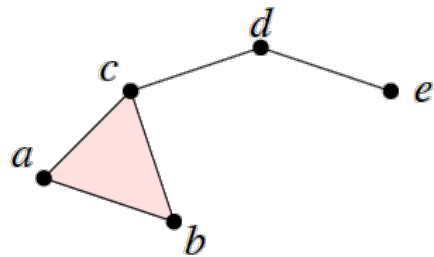
# Continuous to Discrete Link: Triangulations

- 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  $|K| \approx \mathbb{X}$ .



# Abstract Simplicial Complexes

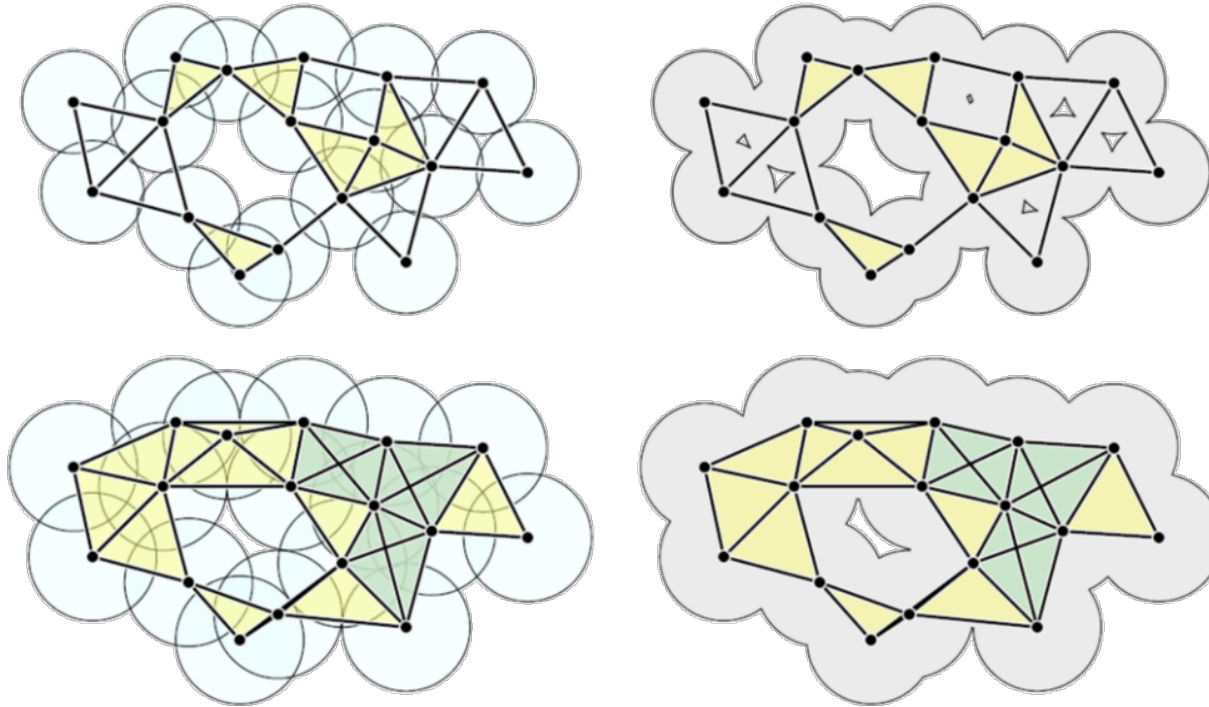
- An **abstract simplicial complex** is a set  $K$ , together with a collection  $\mathcal{S}$  of subsets of  $K$  called **(abstract) simplices** such that:
  1. For all  $v \in K$ ,  $\{v\} \in \mathcal{S}$ . We call the sets  $\{v\}$  the **vertices** of  $K$ .
  2. If  $\tau \subseteq \sigma \in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



Natural partial order structure

# The Nerve of a Finite Cover

Cover a topological space  
by open sets



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

$$\sigma = [U_{i_0}, \dots, U_{i_k}] \in K(\mathcal{U}) \iff \bigcap_{i=0}^k U_{i_j} \neq \emptyset$$

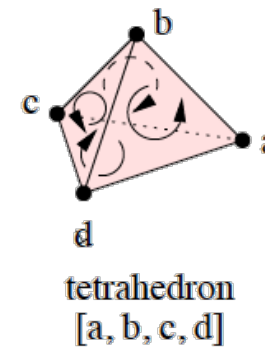
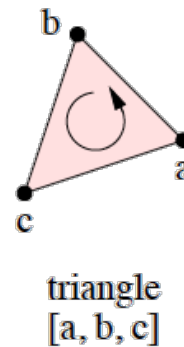
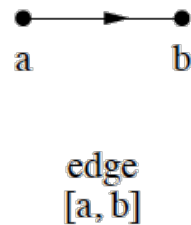
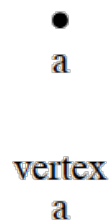
# Orientability

- 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  $\sigma$ , where

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

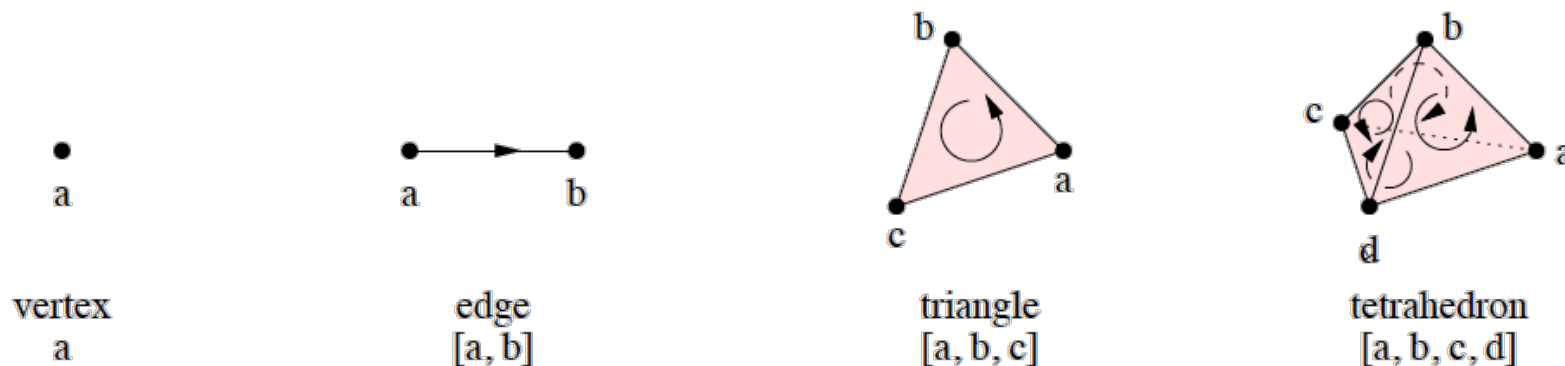
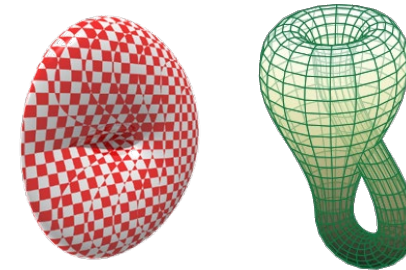
are equivalent orderings if the parity of the permutation  $\tau$  is even.

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



# Orientability

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



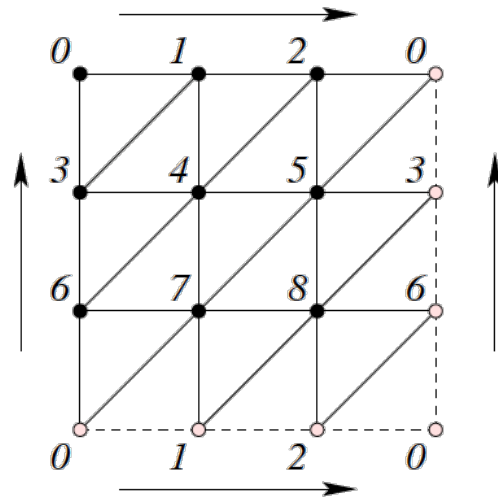
# Euler Characteristic: A Topological Invariant

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

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

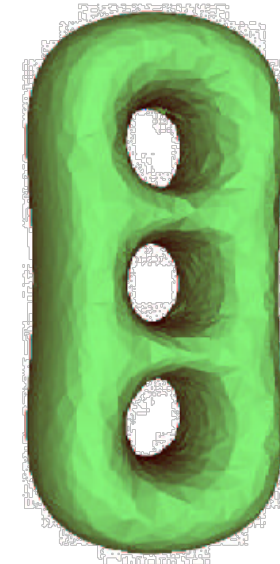
- $v - e + f = 1$  (Graph Theory)
- Invariant for  $|K|$
- **Any** triangulation gives the same answer!
- Intrinsic property

# More on Euler



2-Manifold	$\chi$
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**  $g$ .



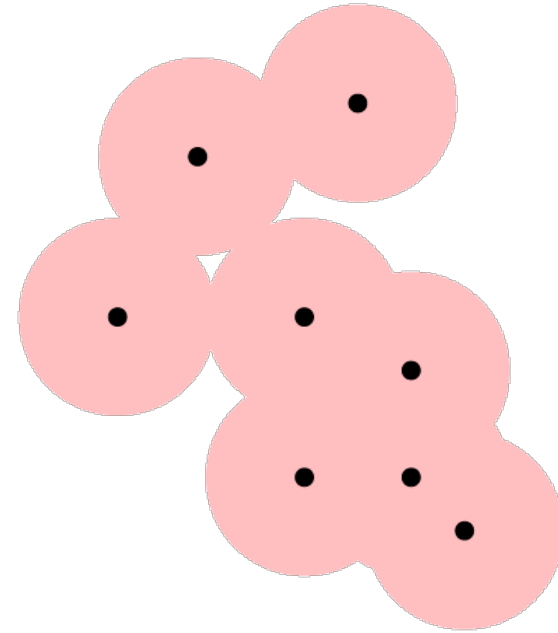
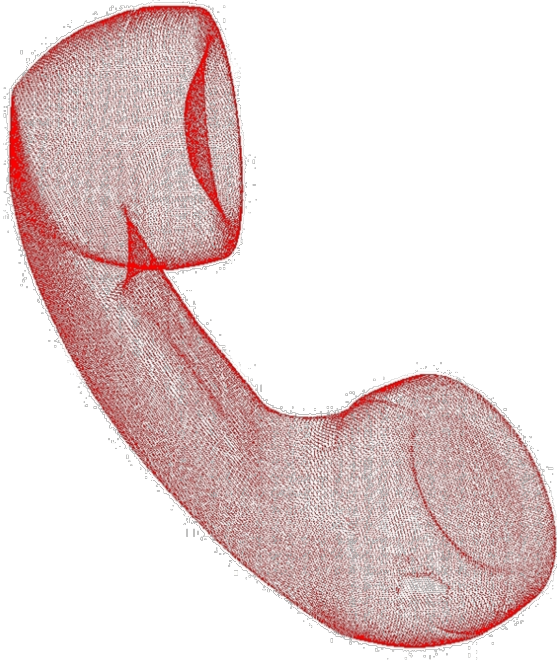
# denotes connected sum

# 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?

# Useful Complexes on Point Clouds

# $\epsilon$ -Balls



- $\epsilon$ -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

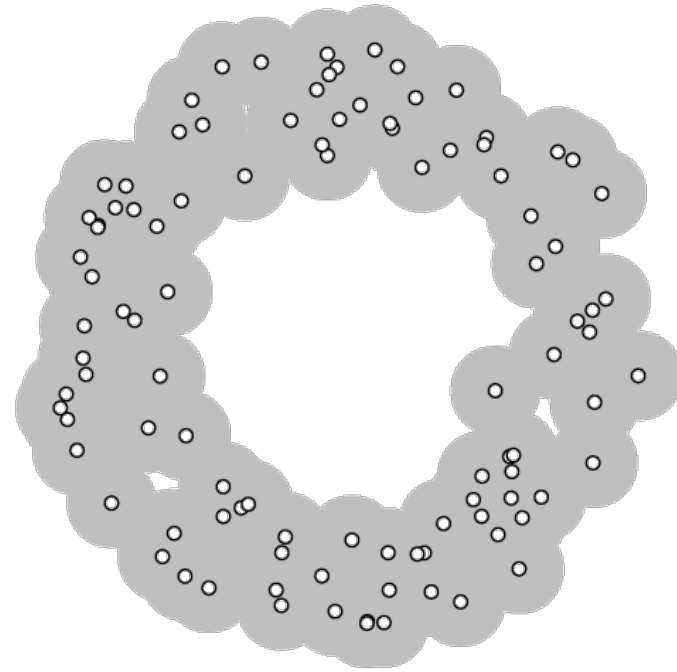
For a dataset  $X$  we study the topology of the *union of balls*

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

**Two Issues:**

**Scale:** No natural choice of  $\epsilon$ !

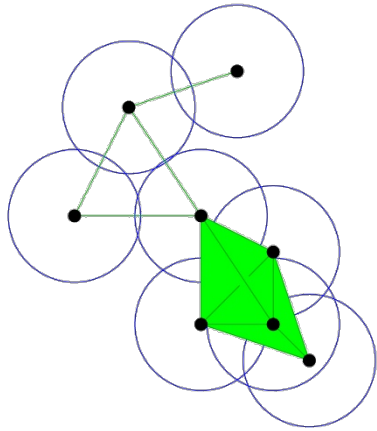
**Conception:** How to encode  $M_\epsilon$  on computer?



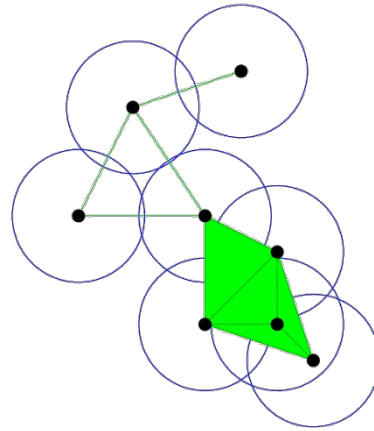
# Complex Zoo

Must choose which simplices to introduce

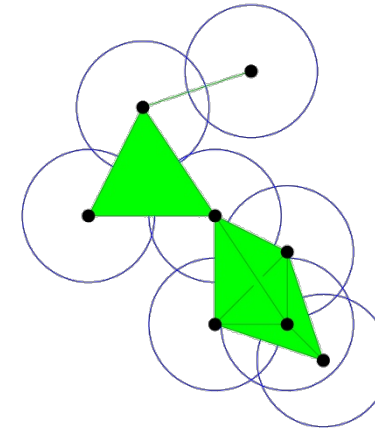
Čech



Alpha

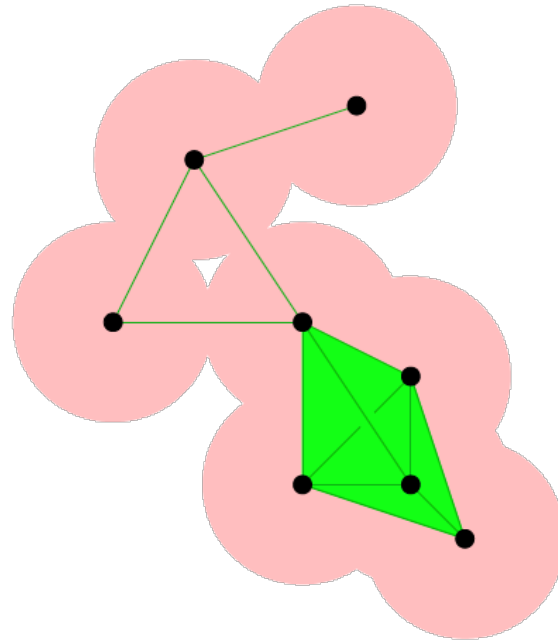


Rips



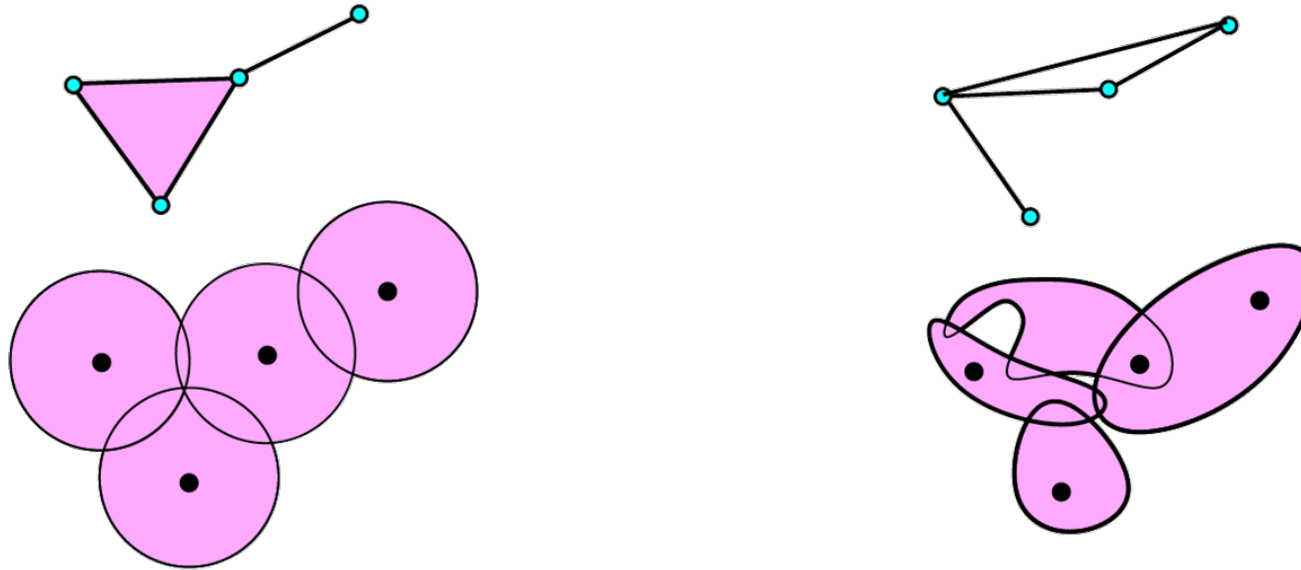
Combinatorial complexes provide discrete representations of the underlying space

# Čech Complex



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

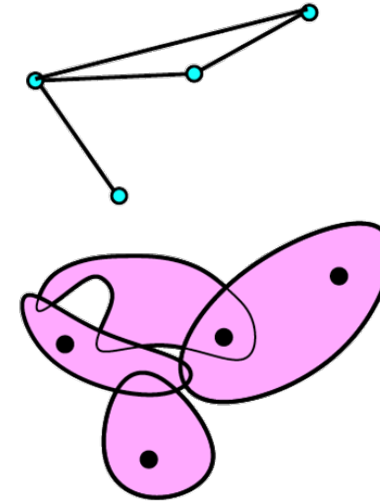
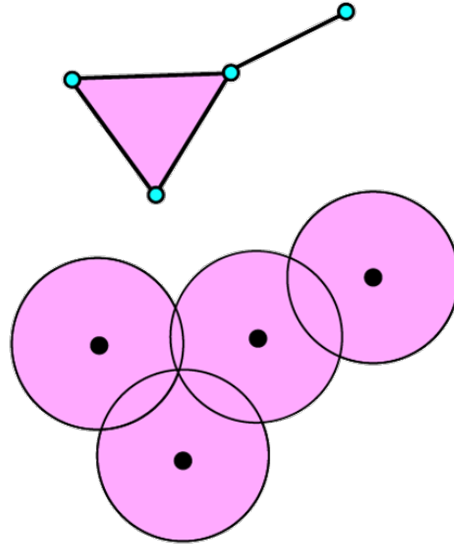
# General Čech Complex = Nerve



- 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

Lemma (Nerve Lemma, Leray '45)  
 $\check{C}_\epsilon$  is topologically equivalent to  $M_\epsilon$ .

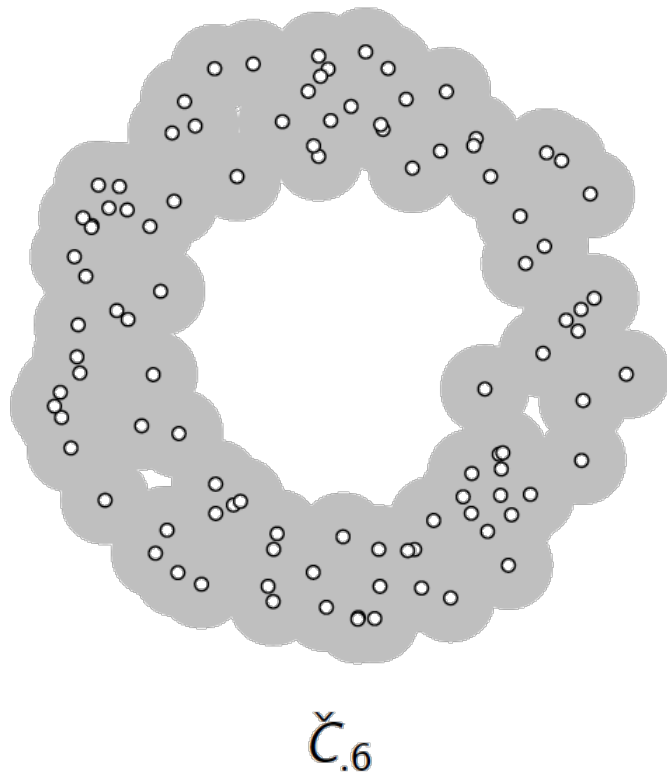


**Nerve theorem (Leray):** If all the intersections between opens in  $\mathcal{U}$  are either empty or contractible then  $C(\mathcal{U})$  and  $X = \cup_{i \in I} U_i$  are homotopy equivalent.

$\Rightarrow$  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!

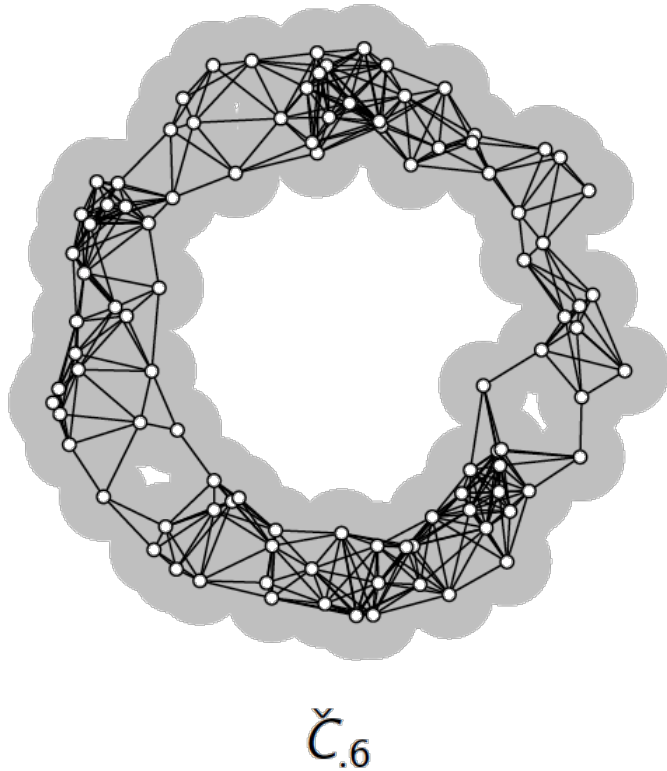
# Čech Complex



The Čech complex  $\check{C}_\epsilon$  encodes the intersection pattern of  $M_\epsilon$ : Encode:

Points as *vertices*  
(0-cells)

# Čech Complex

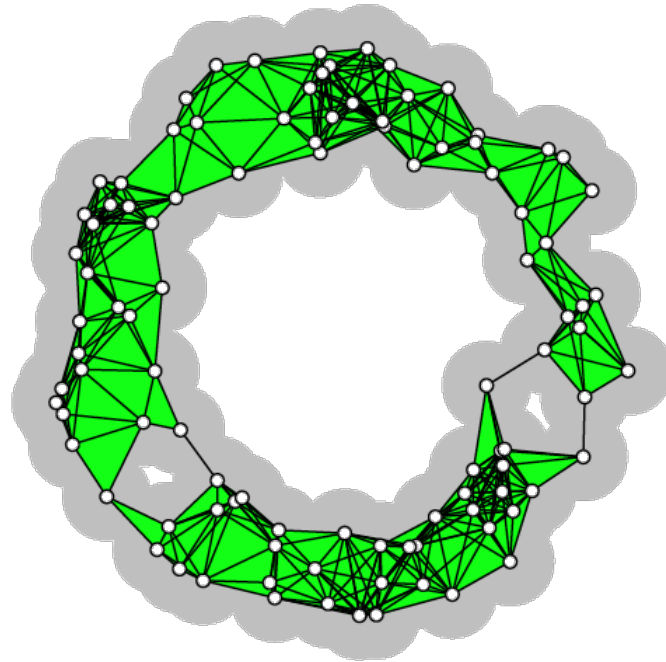


The Čech complex  $\check{C}_{\epsilon}$  encodes the intersection pattern of  $M_{\epsilon}$ : Encode:

Points as *vertices*  
(0-cells)

Pairwise intersections  
as *edges* (1-cells)

# Čech Complex



Č.6

The Čech complex  $\check{C}_\epsilon$  encodes the intersection pattern of  $M_\epsilon$ : Encode:

Points as *vertices*  
(0-cells)

Pairwise intersections  
as *edges* (1-cells)

Threeway intersections  
as *triangles* (2-cells)

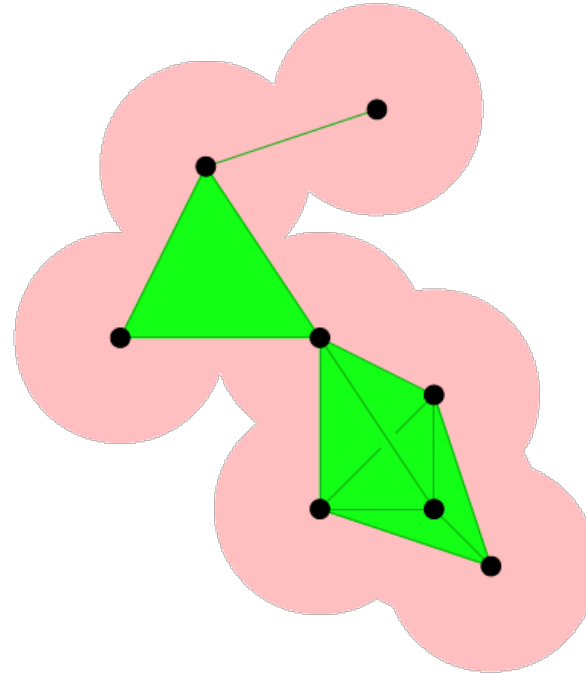
$k$ -way intersections as  
( $k+1$ )-cells

Can be hard to compute ...

# Rips-Vietoris Complex

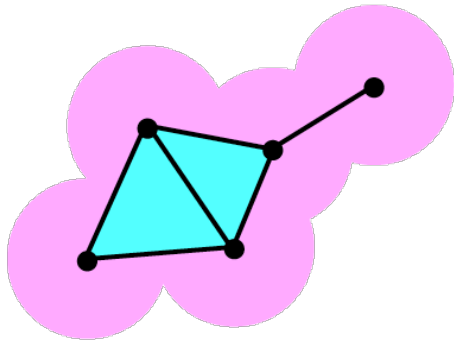
The “poor man’s” alternative  
to the Čech

This is a common complex  
for computations

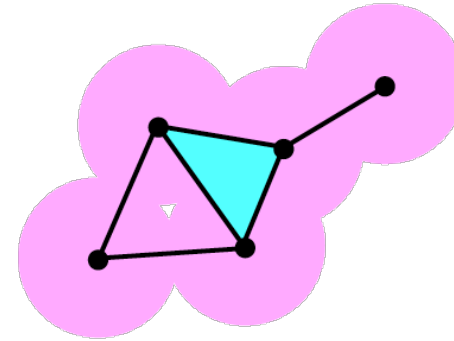


- $R_\epsilon(M) = \{\text{conv } T \mid T \subseteq M, d(m_i, m_j) < \epsilon, m_i, m_j \in T\}.$
- Still  $O\left(\binom{m}{k}\right)$  for the  $k$ th skeleton
- Need  $(k + 1)$ st skeleton for computing  $H_k$

# 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  $p_0, \dots, p_k \in L$ ,

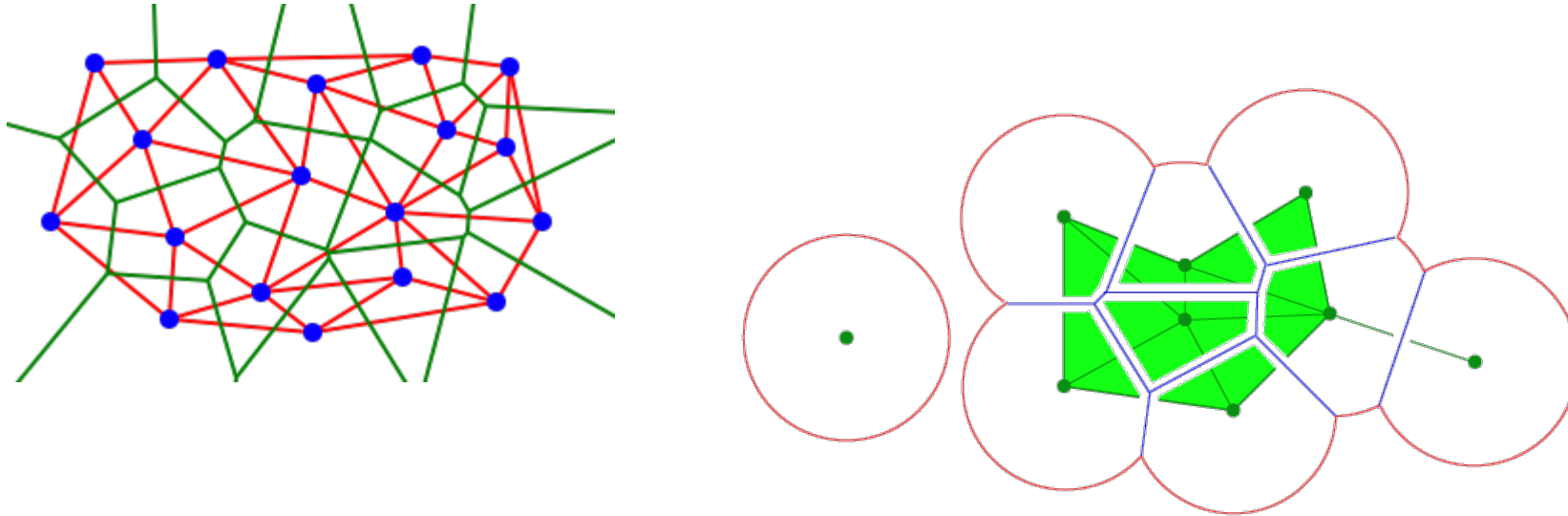
$$\sigma = [p_0 p_1 \dots p_k] \in \mathcal{R}^\alpha(L) \text{ iff } \forall i, j \in \{0, \dots, k\}, d(p_i, p_j) \leq \alpha$$

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

Note  $\alpha$  difference!

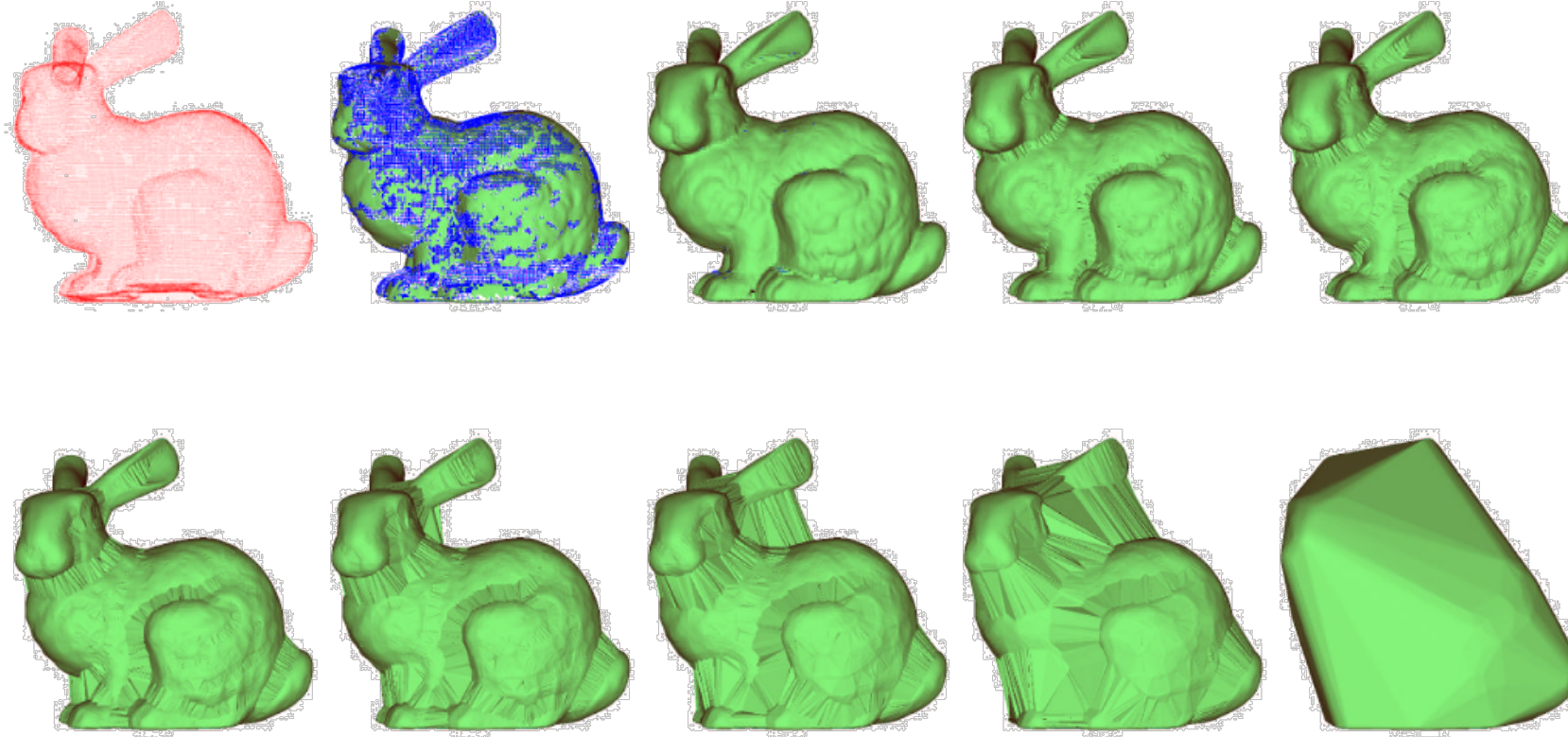
$$\mathcal{C}^{\frac{\alpha}{2}}(L) \subseteq \mathcal{R}^\alpha(L) \subseteq \mathcal{C}^\alpha(L) \subseteq \mathcal{R}^{2\alpha}(L) \subseteq \dots$$

# 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

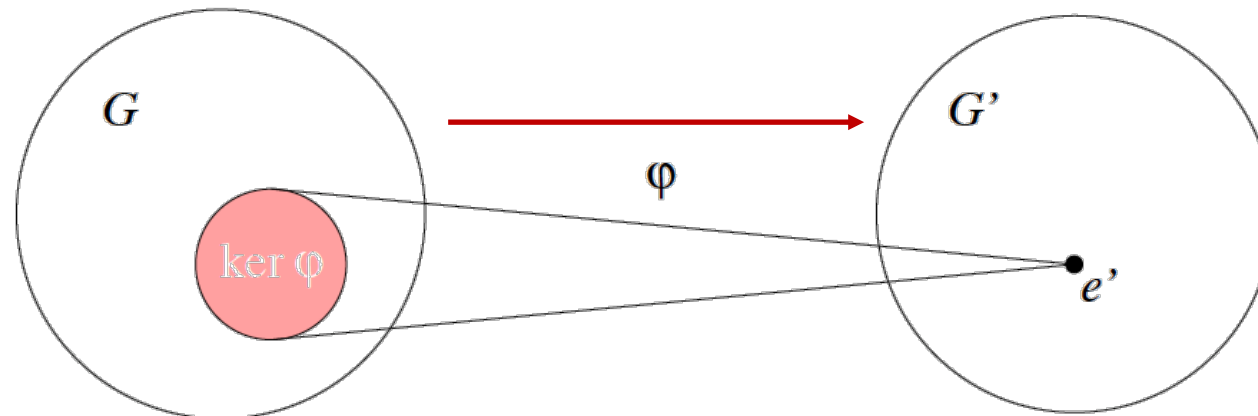
	*		

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

$\mathbb{Z}_m$  integers modulo  $m$

# Homomorphisms

- A map  $\varphi$  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  $\varphi$** .



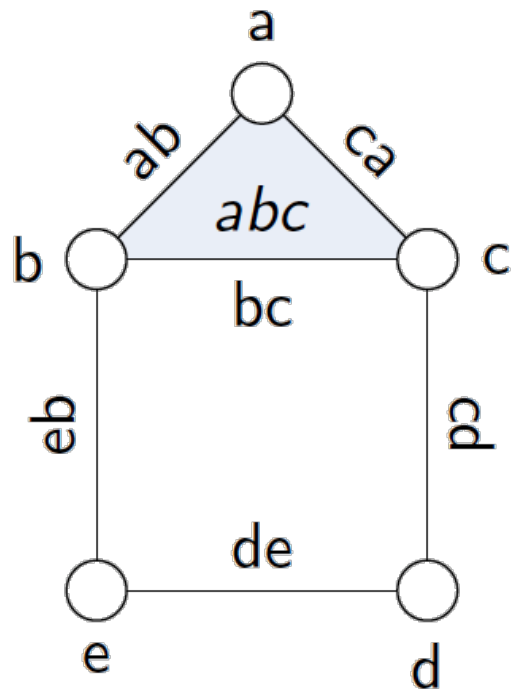
# Finitely Generated Abelian Groups

- 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



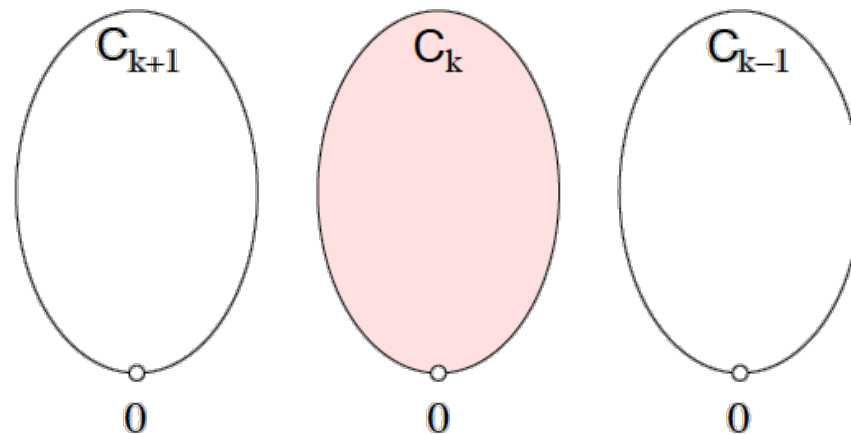
A simplicial complex is a collection of simplices

- ▶ 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  $\sigma$  and  $\tau$  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
- $\text{rank } \mathbf{C}_k = ?$



We take linear combinations of  
simplices of a given dimension  $k$

# 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

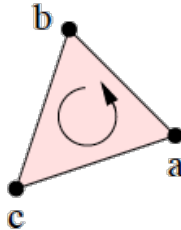
- $\partial_1[a, b] = b - a.$
- $\partial_2[a, b, c] = [b, c] - [a, c] + [a, b] = [b, c] + [c, a] + [a, b].$
- $\partial_3[a, b, c, d] = [b, c, d] - [a, c, d] + [a, b, d] - [a, b, c].$
- $\partial_1\partial_2[a, b, c] = [c] - [b] - [c] + [a] + [b] - [a] = 0.$



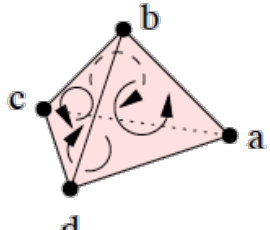
vertex  
a



edge  
[a, b]



triangle  
[a, b, c]



tetrahedron  
[a, b, c, d]

# Boundary Theorem

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

- Proof:

$$\partial_{k-1}\partial_k[v_0, v_1, \dots, v_k] =$$

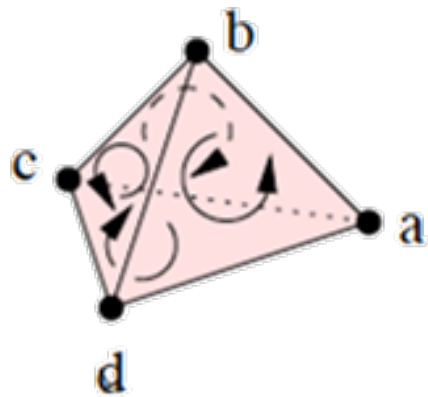
$$= \partial_{k-1} \sum_i (-1)^i [v_0, v_1, \dots, \hat{v}_i, \dots, v_k]$$

$$= \sum_{j < i} (-1)^i (-1)^j [v_0, \dots, \hat{v}_j, \dots, \hat{v}_i, \dots, v_k]$$

$$+ \sum_{j > i} (-1)^i (-1)^{j-1} [v_0, \dots, \hat{v}_i, \dots, \hat{v}_j, \dots, v_k]$$

$$= 0,$$

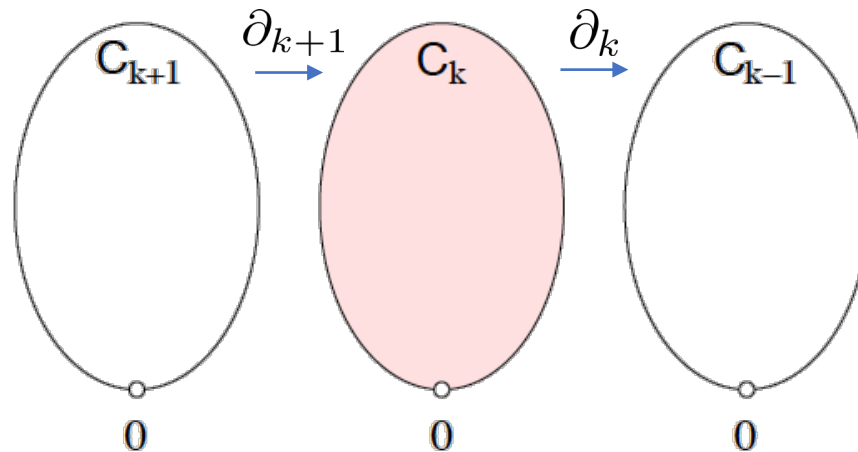
as switching  $i$  and  $j$  in the second sum negates the first sum.



# Chain Complex

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

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

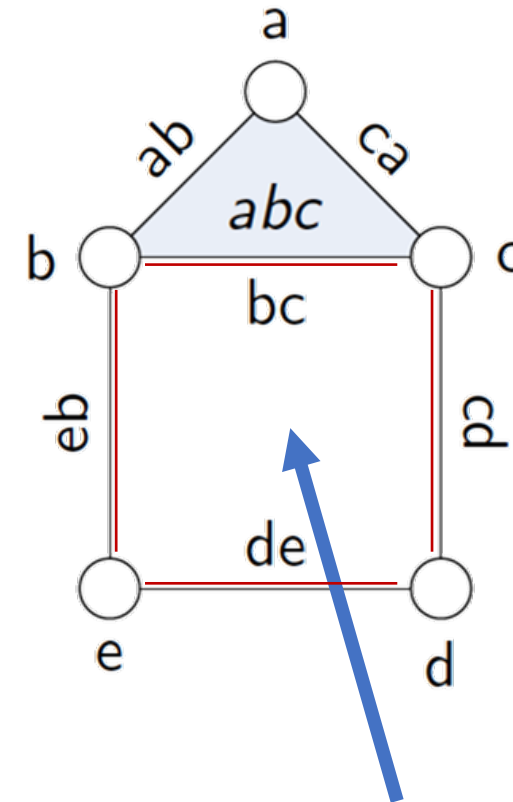
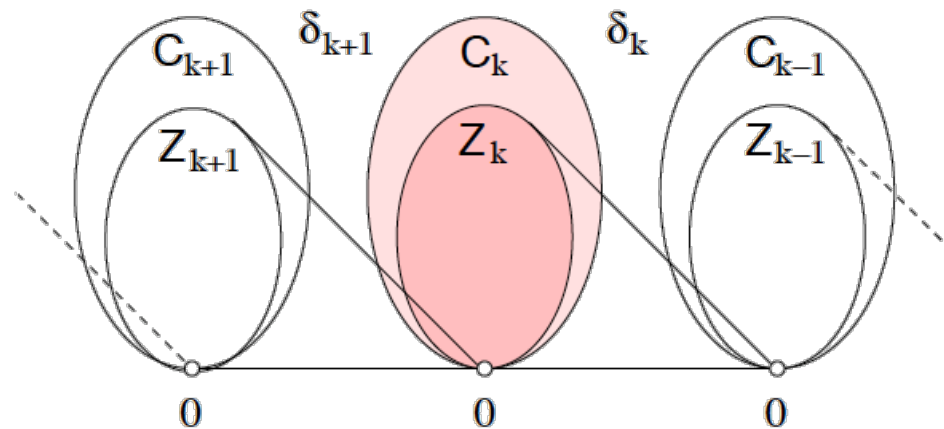


# Cycle Group

- 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

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

$$\partial = \delta$$



$$\begin{aligned} \partial(bc + cd + de + eb) \\ = c - b + d - c + e - d + b - e = 0. \end{aligned}$$



# 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  $b$ ?

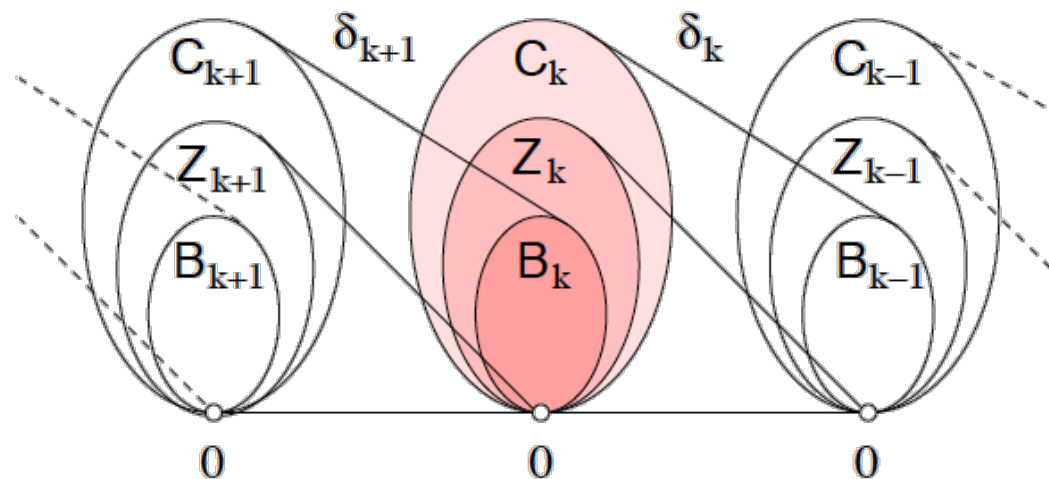
$$\partial_k b = \partial_k \partial_{k+1} c = \emptyset,$$

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

by the boundary theorem.

- That is, every boundary is a cycle!

$$\partial = \delta$$

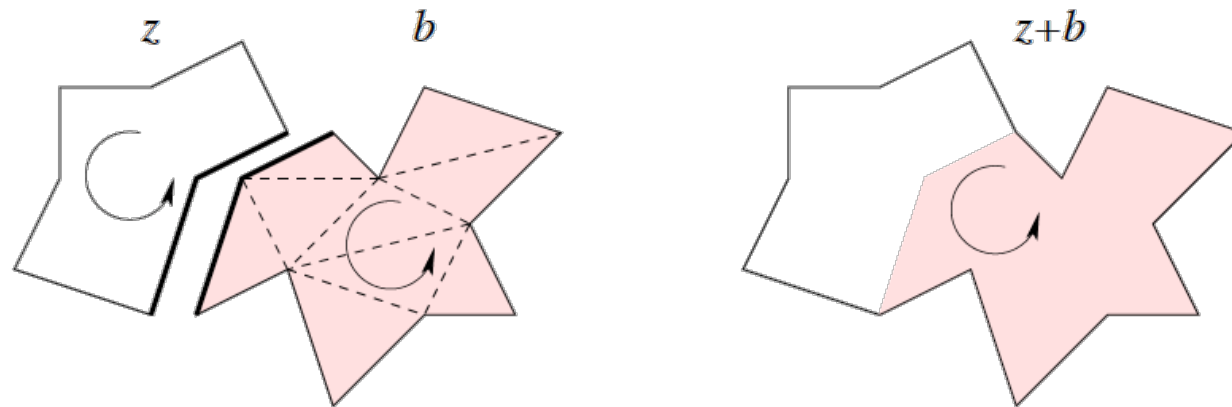


Nesting behavior!

# Equivalent Cycles

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

Cosets!

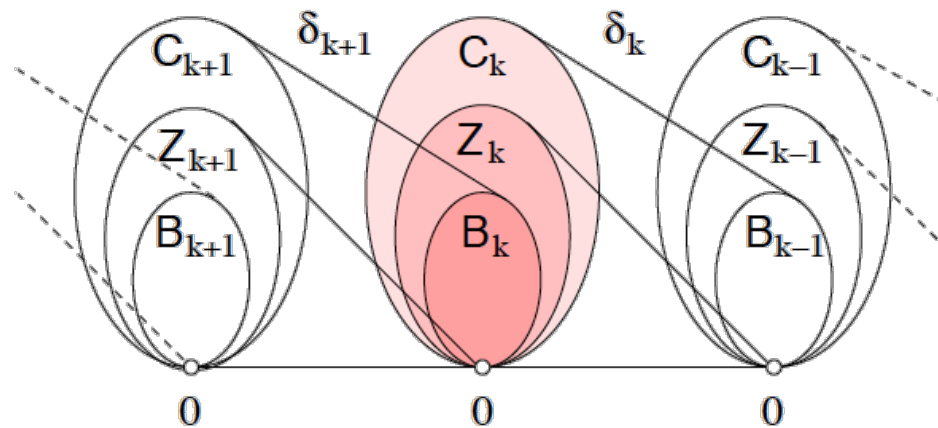


# Simplicial Homology

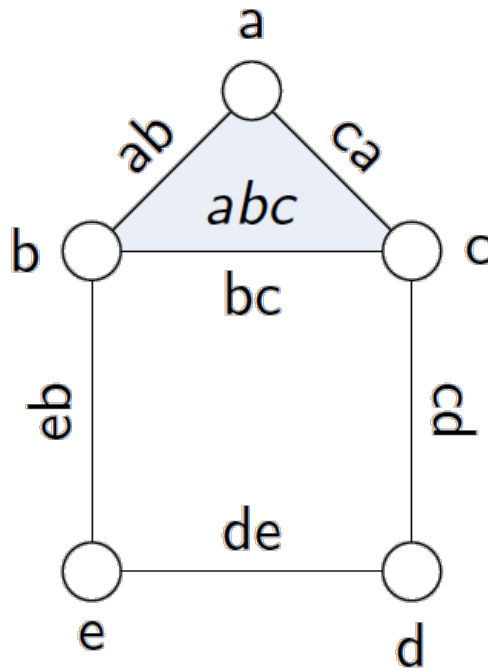
- The  $k$ th homology group is

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

- 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



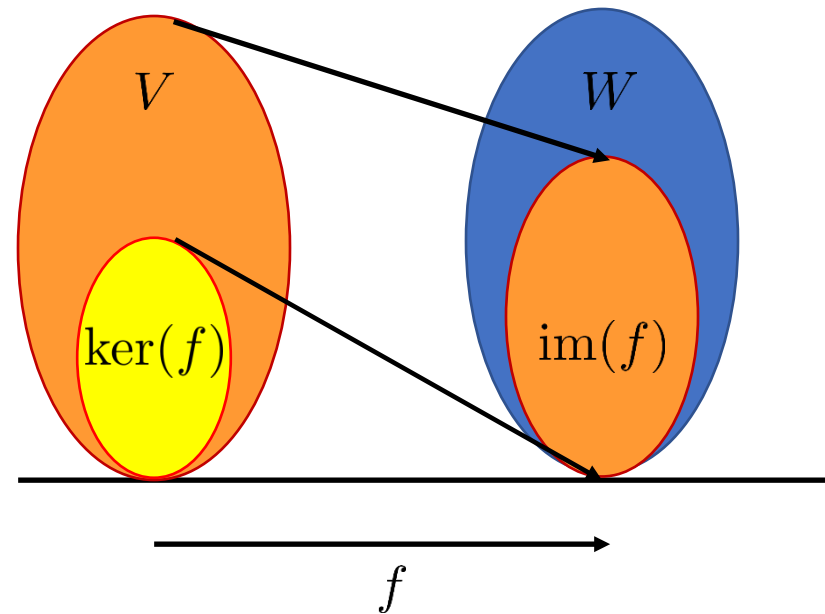
In other words..

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

Homology of a space  $X$  is the quotient:

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

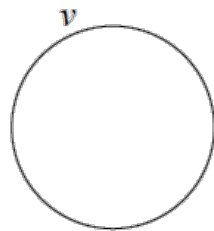
# In Vector Spaces



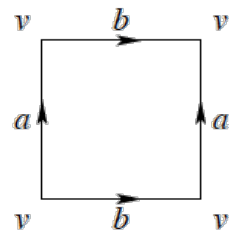
$$V \approx \ker(f) \oplus \text{im}(f)$$

# Homology of 2-Manifolds

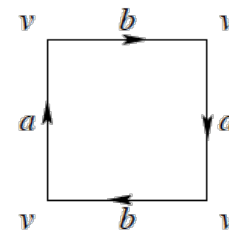
2-manifold	$H_0$	$H_1$	$H_2$
sphere	$\mathbb{Z}$	$\{0\}$	$\mathbb{Z}$
torus	$\mathbb{Z}$	$\mathbb{Z} \times \mathbb{Z}$	$\mathbb{Z}$
projective plane	$\mathbb{Z}$	$\mathbb{Z}_2$	$\{0\}$
Klein bottle	$\mathbb{Z}$	$\mathbb{Z} \times \mathbb{Z}_2$	$\{0\}$



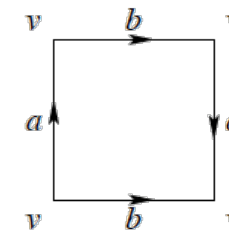
(a) Sphere



(b) Torus

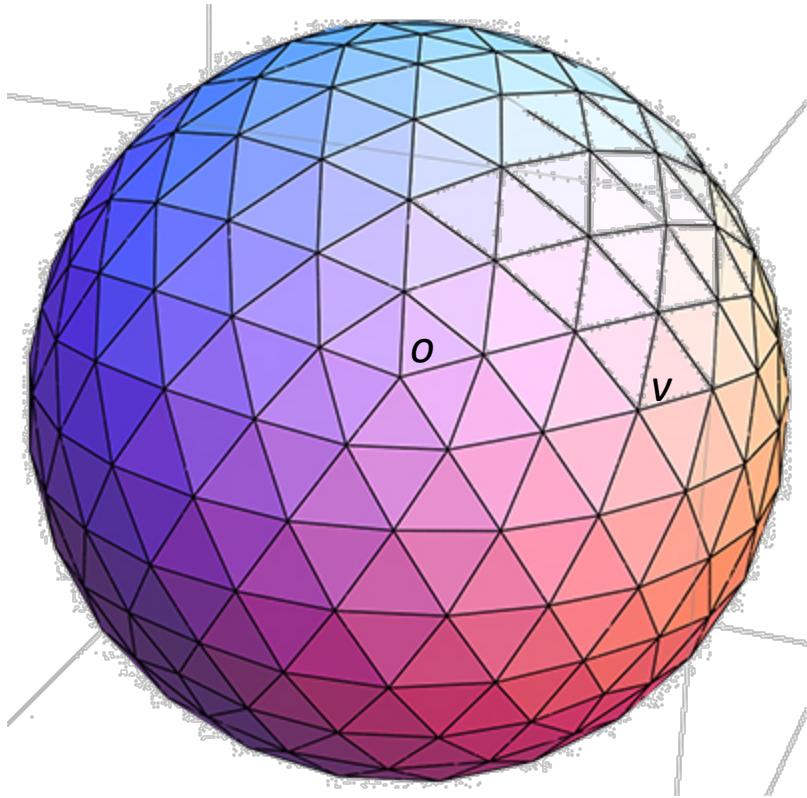


(c) Projective plane



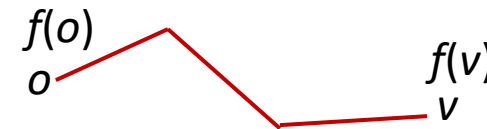
(d) Klein bottle

# 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

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

- The  $k$ th Betti number  $\beta_k$  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:
    - $\beta_0$  measures the number of components of the complex.
    - $\beta_1$  is the rank of a basis for the tunnels.
    - $\beta_2$  counts the number of voids in the complex.

# Betti Numbers $\beta_i$

- Ranks  $\beta_i$  of the free part of homology groups  $H_i$

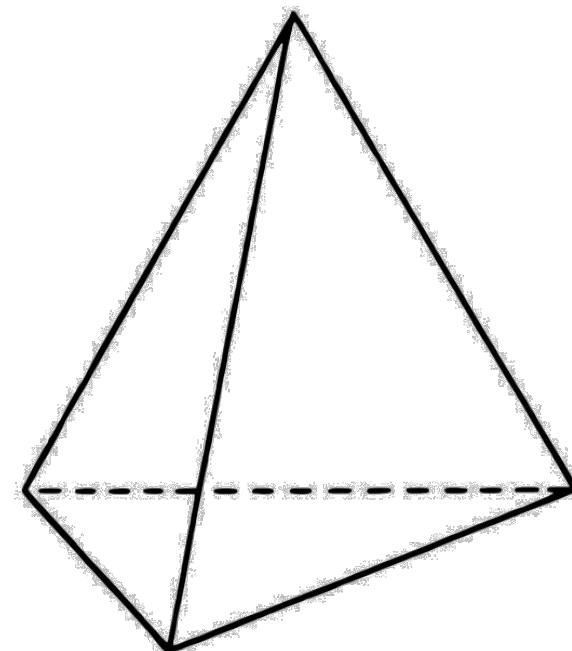
$$\beta_i = z_i - b_i$$

- $\beta_0$  counts the number of connected components

- $\beta_1$  counts the number of independent loops

- $\beta_2$  counts the number of independent voids

- ...



Topology is fundamentally a tool for classification

# Invariance of Homology Groups

- (Hauptvermutung) Any two triangulations of a topological space have a common refinement (Poincaré 1904)
  - True for polyhedra of dimension  $\leq 2$  (Papakyriakopoulos 1943)
  - True for 3-manifolds (Moïse 1953)
  - False in dimensions  $\geq 6$  (Milnor 1961)
  - False for manifolds of dimension  $\geq 5$  (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 Formula

- Homology functors  $\mathbf{H}_*$
- $\mathbf{H}_*(\mathbf{C}_*)$  is a chain complex:

$$\dots \rightarrow \mathbf{H}_{k+1} \xrightarrow{\partial_{k+1}} \mathbf{H}_k \xrightarrow{\partial_k} \mathbf{H}_{k-1} \rightarrow \dots$$

2-manifold	$\mathbf{H}_0$	$\mathbf{H}_1$	$\mathbf{H}_2$
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	$\mathbb{Z}$	$\mathbb{Z} \times \mathbb{Z}_2$	$\{0\}$

- What is its Euler characteristic?  $\sum_i (-1)^i (z_i + b_{i-1}) = \sum_i (-1)^i (z_i - b_i)$
- (Theorem)  $\chi(K) = \chi(\mathbf{C}_*) = \chi(\mathbf{H}_*(\mathbf{C}_*))$ .
- $\sum_i (-1)^i s_i = \sum_i (-1)^i \text{rank}(\mathbf{H}_i) = \sum_i (-1)^i \beta_i$
- Sphere:  $2 = 1 - 0 + 1$
- Torus:  $0 = 1 - 2 + 1$

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

# That's All

