MATH412/COMPSCI434/MATH713
Fall 2025
Topological Data Analysis

Topic 3: Simplicial Homology - Part 3
Topic 4: Introduction to Persistent Homology - Part 1

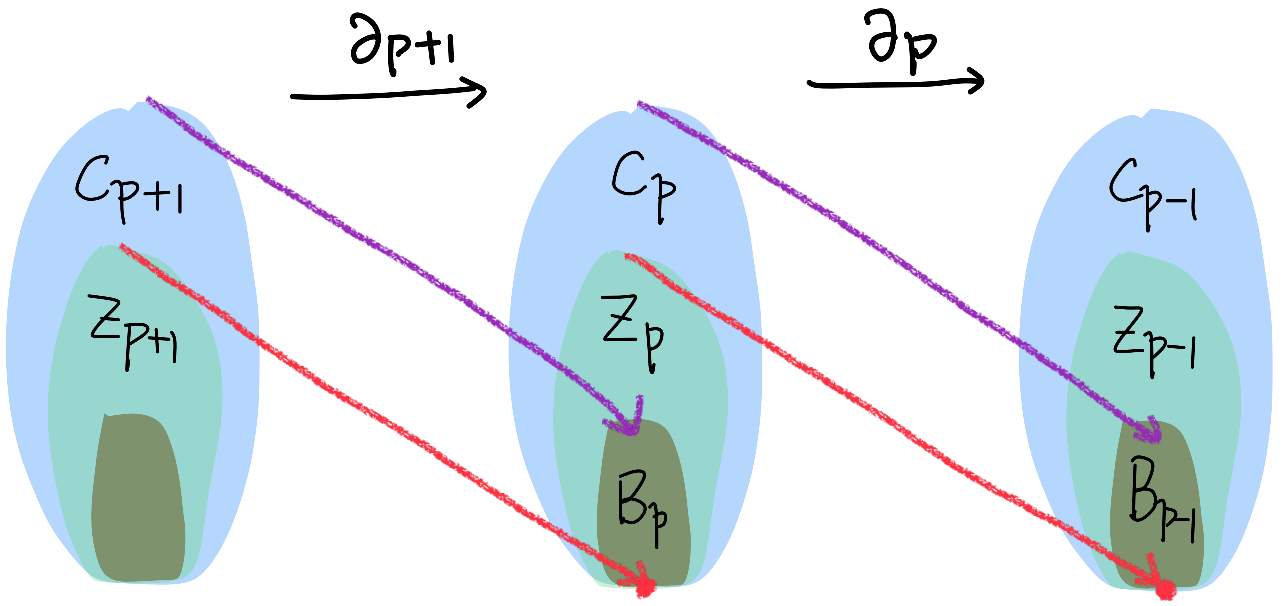
Instructor: Ling Zhou

Review: Cycles and Boundaries

boudaries $\subseteq$ cycles $\subseteq$ chains

Under field (e.g. $\mathbb{Z}_2$) coefficients, B_p, Z_p, C_p are all vector spaces.

$$\partial_p \circ \partial_{p+1} = 0 \Rightarrow B_p \subseteq Z_p \subseteq C_P$$



Review: Homology groups

- p-th **homology group** is $H_p(K) = Z_p / B_p$
  - c_1 is **homologous to** c_2 if $c_1 - c_2 \in B_p$, i.e, $c_1 - c_2$ is a boundary cycle
  - The **homology class** $[c]$ represents the family p-cycles homologous to c
- **Betti number**: $\beta_p(K) = \dim H_p(K) = \dim Z_p - \dim B_p$

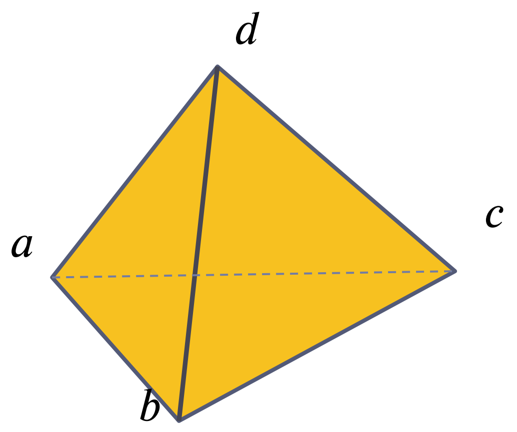
- Theorem: Two homotopy equivalent topological spaces have isomorphic homology groups and thus same Betti numbers.

$$\cdots \xrightarrow{\partial_{p+2}} \mathrm{C}_{p+1} \xrightarrow{\partial_{p+1}} \mathrm{C}_p \xrightarrow{\partial_p} \mathrm{C}_{p-1} \xrightarrow{\partial_{p-1}} \cdots .$$

Matrix view and computation

Boundary Matrix

- $K^p = \left\{ \alpha_1, \ldots, \alpha_{n_p} \right\}$, $K^{p-1} = \left\{ \tau_1, \ldots, \tau_{n_{p-1}} \right\}$
  - K^p forms a basis for p-th chain group C_p
- For simplicity, we **use $\mathbb{Z}_2$ coefficients**, and define the $n_{p-1} \times n_p$ **boundary matrix** A_p
  - $A_p[i][j] = 1$ iff $\tau_i \subseteq \sigma_j$
  - representing $\partial_p : C_p \to C_{p-1}$ w.r.t. ordered bases $\left\{ \alpha_1, \ldots, \alpha_{n_p} \right\}$ and $\left\{ \tau_1, \ldots, \tau_{n_{p-1}} \right\}$



$$A_2 = \begin{array}{c} \\ ab \\ ac \\ ad \\ bc \\ bd \\ cd \end{array} \begin{array}{c} \begin{array}{cccc} abc & abd & acd & bcd \end{array} \\ \begin{pmatrix} 1 & 1 & & \\ 1 & & 1 & \\ & 1 & 1 & \\ 1 & & & 1 \\ & 1 & & 1 \\ & & 1 & 1 \end{pmatrix} \end{array}$$

Boundary Matrix

- Vector representation of a p-chain $c = \sum_{i=1}^{n_p} c_i \alpha_i$:
  - Under basis K^p, vector representation of c is $\vec{c} = \left[c_1,\ c_2, \ldots,\ c_{n_p}\right]^T$
- Boundary $\partial_p c$ is a $(p-1)$-chain with vector representation $A_p \vec{c}$ w.r.t basis K^{p-1}

$$A_p\vec{c} = \begin{bmatrix} a_1^1 & a_1^2 & \cdots & a_1^{n_p} \\ a_2^1 & a_2^2 & \cdots & a_2^{n_p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_{p-1}}^1 & a_{n_{p-1}}^2 & \cdots & a_{n_{p-1}}^{n_p} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_{n_p} \end{bmatrix}$$

Observations

- To compute the cycle space $Z_p = \ker \partial_p$,
  - solve the equation $A_p c = 0$ to find a basis for the kernel
- To compute the boundary space $B_p = \mathrm{Im}\partial_{p+1}$
  - compute the column space of A_{p+1} and find a basis for the image
- We can do both on A_p through **column reduction**

null space

column space

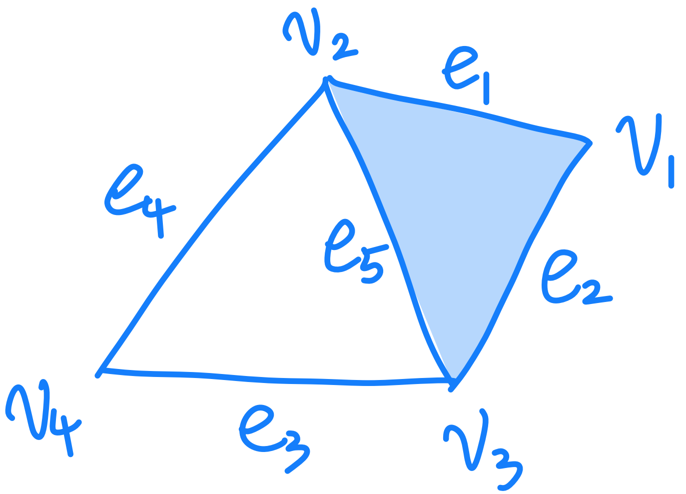
Matrix reduction

- Turn A_p into the **column reduced form**
  - Each non-zero column has a unique lowID/pivot: index of lowest 1-entry
- **Column operations** in Gaussian elimination:
  - scaling (not needed over $\mathbb{Z}_2$)
  - swap (not necessary)
  - add one column to another
- Do **left-to-right column reduction** and get bases of
  - $B_{p-1} = \mathrm{Im}\partial_p$: the reduced columns
  - $Z_p = \ker \partial_p$: the column operations

$$\begin{bmatrix} * & * & * & 0 \\ * & 1 & * & 0 \\ 1 & 0 & * & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

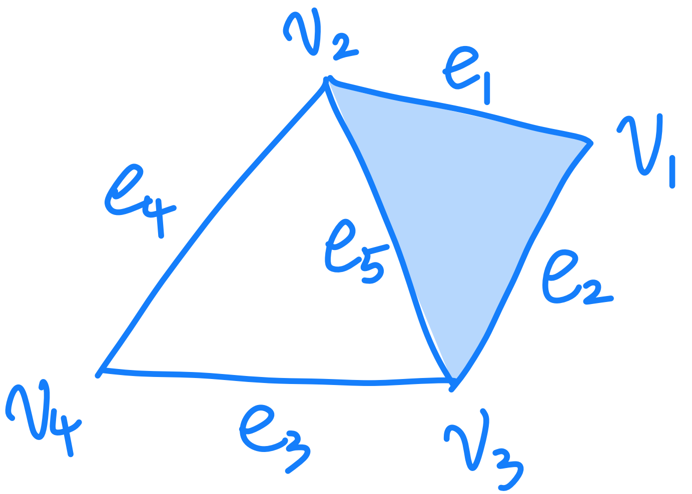
Column reduced form

$low[i] \neq low[j]$



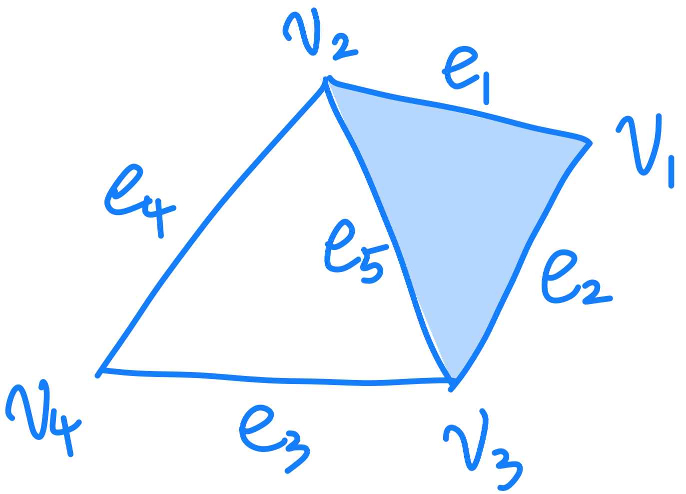
	v_1v_2	v_1v_3	v_3v_4	v_2v_4	v_2v_3
	e1	**e2**	**e3**	**e4**	**e5**
v1	1	1	0	0	0
v2	1	0	0	1	1
v3	0	1	1	0	1
v4	0	0	1	1	0

add e_3 to e_4



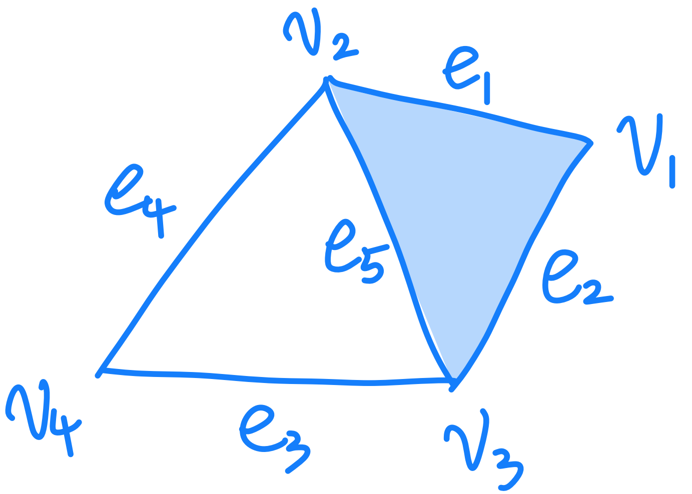
	$v_1 v_2$	$v_1 v_3$	$v_3 v_4$	$v_2 v_4$	$v_2 v_3$
	e1	**e2**	**e3**	**e4+e3**	**e5**
v1	1	1	0	0	0
v2	1	0	0	1	1
v3	0	1	1	1	1
v4	0	0	1	0	0

add e_2 to $e_4 + e_3$



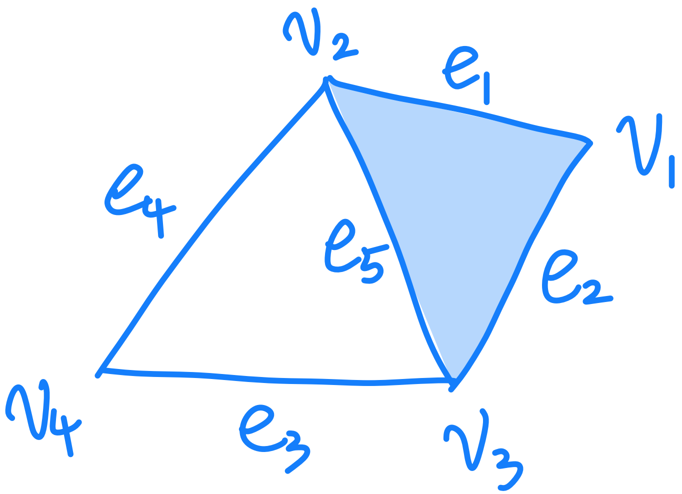
	v_1v_2	v_1v_3	v_3v_4	v_2v_4	v_2v_3
	e1	**e2**	**e3**	**e4+e3+e2**	**e5**
v1	1	1	0	1	0
v2	1	0	0	1	1
v3	0	1	1	0	1
v4	0	0	1	0	0

add e_1 to $e_4+e_3+e_2$



	v_1v_2	v_1v_3	v_3v_4	v_2v_4	v_2v_3
	e1	**e2**	**e3**	**e4+e3+e2 +e1**	**e5**
v1	1	1	0	0	0
v2	1	0	0	0	1
v3	0	1	1	0	1
v4	0	0	1	0	0

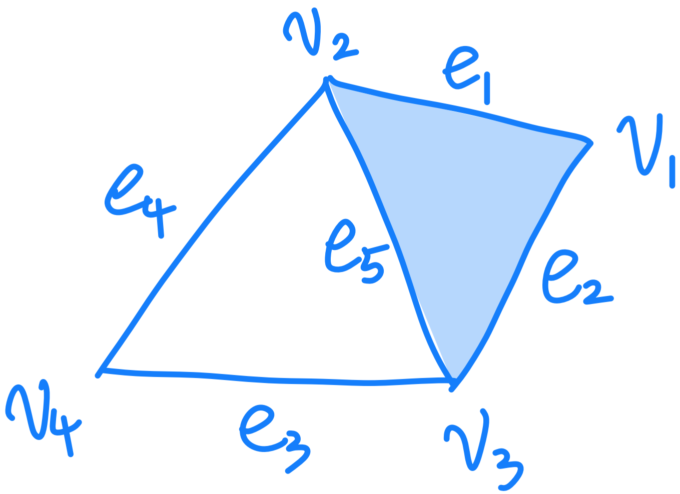
add e_2 to e_5



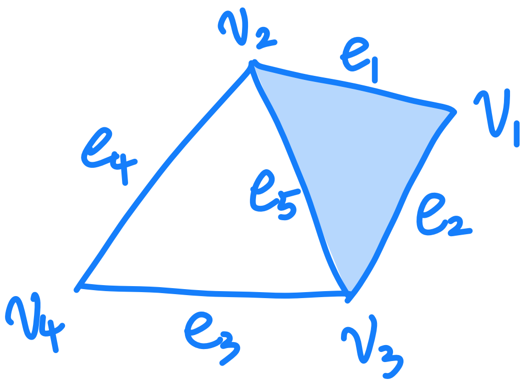
v_1v_2 v_1v_3 v_3v_4 v_2v_4 v_2v_3

	e1	e2	e3	e4+e3+e2 +e1	e5+e2
v1	1	1	0	0	1
v2	1	0	0	0	1
v3	0	1	1	0	0
v4	0	0	1	0	0

add e_1 to e_5+e_2



	v_1v_2	v_1v_3	v_3v_4	v_2v_4	v_2v_3
	e1	e2	e3	e4+e3+e2 +e1	e5+e2+e1
v1	1	1	0	0	0
v2	1	0	0	0	0
v3	0	1	1	0	0
v4	0	0	1	0	0

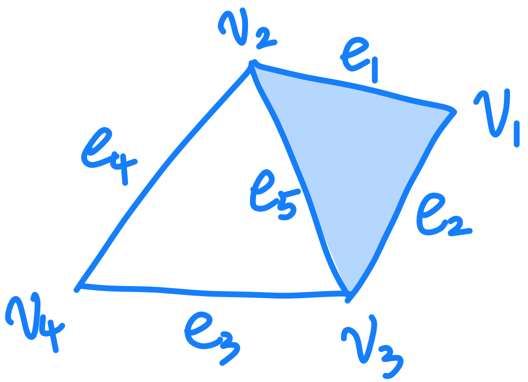


$$A_1 \xrightarrow{\text{reduction}} \tilde{A}_1 =$$

	v_1v_2	v_1v_3	v_3v_4	v_2v_4	v_2v_3
	e1	**e2**	**e3**	**e4+e3+e2+e1**	**e5+e2+e1**
v1	1	1	0	0	0
v2	1	0	0	0	0
v3	0	1	1	0	0
v4	0	0	1	0	0

- $\dim B_0 = 3$, $\quad \dim(B_1) = \operatorname{rank}(A_1) = \#$ of pivots
- $\dim Z_1 = 2$, $\quad \dim(Z_1) = \dim C_1 - \operatorname{rank}(A_1)$

$= \#$ of columns $- \#$ of pivots



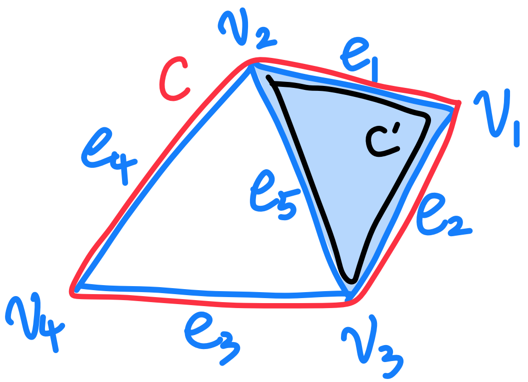
$A_1 \xrightarrow{\text{reduction}} \tilde{A}_1 =$

	v_1v_2	v_1v_3	v_3v_4	v_2v_4	v_2v_3
	e1	**e2**	**e3**	**e4+e3+e2 +e1**	**e5+e2+e1**
v1	1	1	0	0	0
v2	1	0	0	0	0
v3	0	1	1	0	0
v4	0	0	1	0	0

- $\beta_p = \dim Z_p - \dim B_p$

$\beta_0 = \dim Z_0 - \dim B_0 = \dim C_0 - \dim B_0 = \#\text{ of rows} - \#\text{ of pivots}$

$= 4 - 3 = 1$



$$A_1 \xrightarrow{\text{reduction}} \tilde{A}_1 =$$

	v_1v_2	v_1v_3	v_3v_4	v_2v_4	v_2v_3
	e1	e2	e3	e4+e3+e2+e1	e5+e2+e1
v1	1	1	0	0	0
v2	1	0	0	0	0
v3	0	1	1	0	0
v4	0	0	1	0	0
	$= \alpha_1$	$= \alpha_2$	$= \alpha_3$		

Z_1

- Bases for B_0 and Z_1

• $\vec{v} := (v_1, v_2, v_3, v_4) \Rightarrow B_0 = \langle \alpha_1 \cdot \vec{v}, \alpha_2 \cdot \vec{v}, \alpha_3 \cdot \vec{v} \rangle$

• $Z_1 = \langle \underbrace{e_1 + e_2 + e_3 + e_4}_{= C}, \underbrace{e_1 + e_2 + e_5}_{= C'} \rangle$

Left-to-right Column Reduction Algorithm

$R = \partial$; $\quad V = \mathrm{Id}_{n_p \times n_p}$

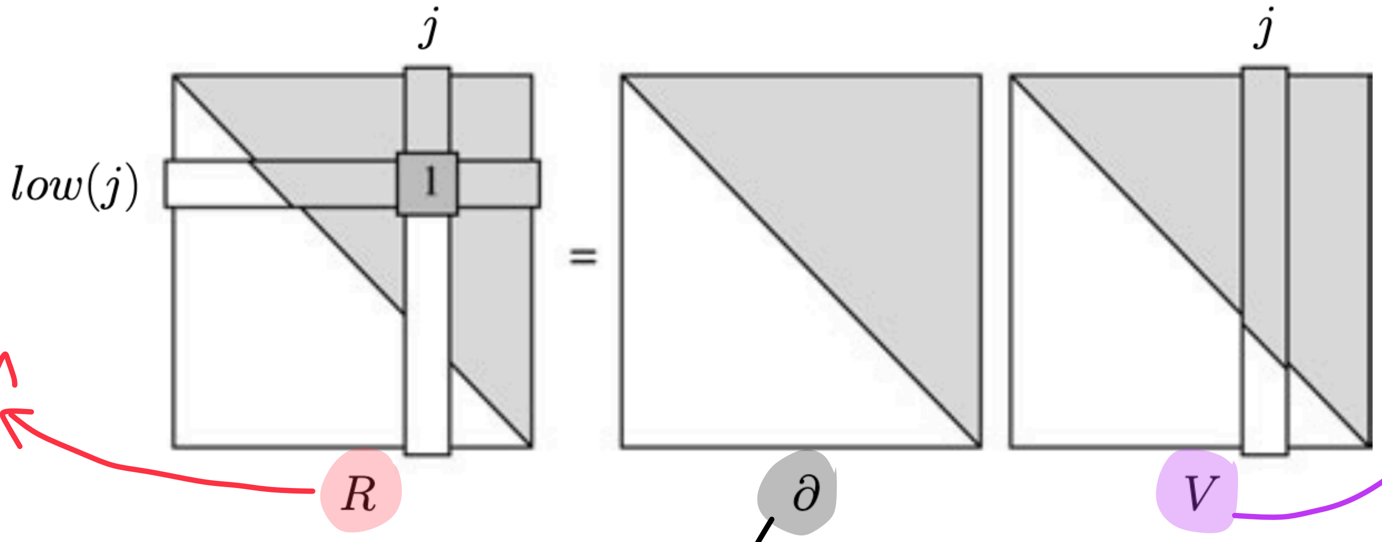
for $j = 1$ **to** m **do**

  while there exists $j_0 < j$ with $low(j_0) = low(j)$ **do**

    add column j_0 to column j ; add V_{j_0} to V_j

  endwhile

endfor.



reduced form

$\Downarrow$

boundaries B_{p-1}

original boundary matrix A_p

column operations

$\Downarrow$

cycles Z_p

Properties

- Let $n_p =$ the number of p-simplices. Let A_j be the j-th column of A.
  - Theorem:
    - The algorithm terminates in $O(n_p n_{p-1}^2)$ time
    - The output matrix R is in column reduced form
    - The set $\{R_j : R_j \neq 0\}$ form a basis for B_{p-1}
    - The set $\{V_j : R_j = 0\}$ form a basis for Z_p

This is not the only reduction algorithm!! Any elimination via row/column additions to convert a matrix into a reduced form works!

Computing a basis for homology

reduce A_{p+1} reduce A_p

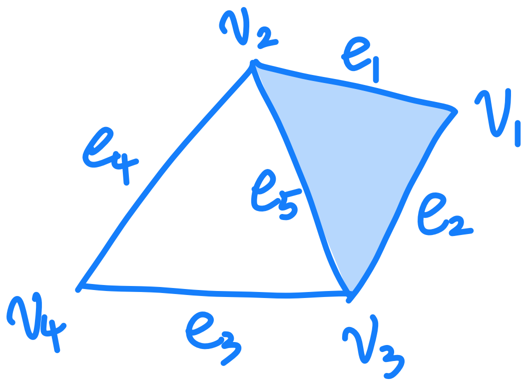
$$\left[\begin{array}{c|c} \text{Basis of } B_p & \text{Basis of } Z_p \end{array}\right]$$

- Left part is already column reduced

the RHS after reduction

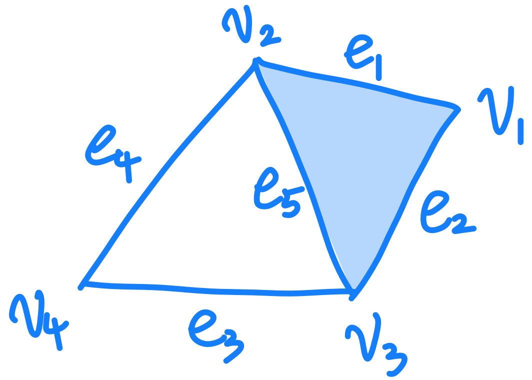
- Apply Reduction to the above matrix to obtain $\boxed{\text{basis of } H_p}$

Computing a basis for homology



	B_1	Z_1	
	e5+e2+e1	e4+e3+e2+e1	e5+e2+e1
E1	1	1	1
E2	1	1	1
E3	0	1	0
E4	0	1	0
E5	1	0	1

Computing a basis for homology



	e5+e2+e1		e4+e3+e2+e1	0
E1	1		1	0
E2	1		1	0
E3	0		1	0
E4	0		1	0
E5	1		0	0

a basis for H_1

$\Downarrow$

$$H_1 = \langle [e_4+e_3+e_2+e_1] \rangle$$

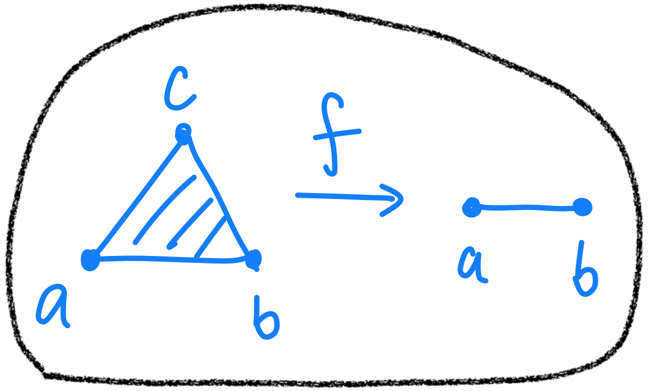
Functoriality of Homology

Functoriality of Simplicial Homology

- Recall: Given simplicial complexes K and L, a function $f : V(K) \to V(L)$ is called a simplicial map if for any $\sigma = \{p_0, \ldots, p_d\} \in \Sigma(K)$, $f(\sigma) = \left\{ f\left(p_0\right), \ \ldots, \ f\left(p_d\right) \right\}$ spans a simplex in L, i.e., $f(\sigma) \in \Sigma(L)$.
- Let $K = (V, \Sigma)$ and $K' = (V', \Sigma')$ and let $f : V \to V'$ be a simplicial map. Then,
  - f induces a linear map on homology groups $f_p : H_p(K) \to H_p(K')$
  - If there exist $K'' = (V'', \Sigma'')$ and another simplicial map $g : V' \to V''$, then
    - $(g \circ f)_p = g_p \circ f_p$

Construction of f_p

- First, f induces linear maps on chain spaces $\bar{f}_p : C_p(K) \to C_p(K')$
  - $\bar{f}_p(\sigma) = \begin{cases} f(\sigma) & \text{if } f(\sigma) \text{ is a p-simplex} \\ 0 & \text{otherwise} \end{cases}$
- Then, $\bar{f}_p : C_p(K) \to C_p(K')$ induces $f_p : H_p(K) \to H_p(K')$
  - $f_p([c]) := [\bar{f}_p(c)]$

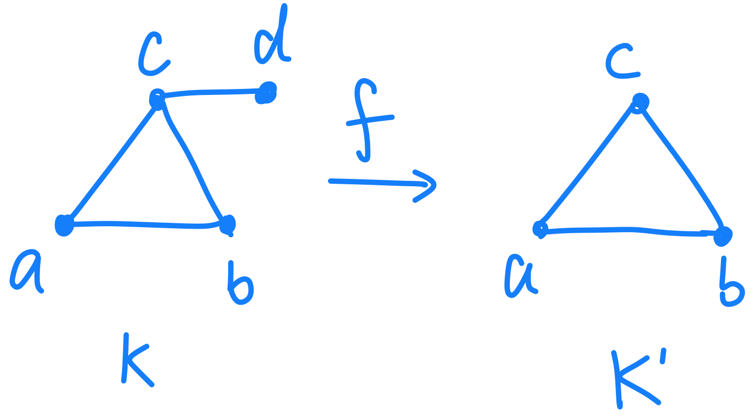


$$\begin{tikzcd}
\cdots \arrow[r] & C_p(K) \arrow[r] \arrow[d, "\bar{f}_p"'] & C_{p-1}(K) \arrow[r] \arrow[d, "\bar{f}_{p-1}"'] & \cdots \\
\cdots \arrow[r] & C_p(K') \arrow[r] & C_{p-1}(K') \arrow[r] & \cdots
\end{tikzcd}$$

$f_p : H_p(K) \to H_p(K')$

for every p

Construction of f_p



$$\begin{tikzcd}
 & \langle ab, ac, bc, cd \rangle \arrow[d, phantom, "\parallel"] & & \\
\cdots \arrow[r] & C_1(K) \arrow[r, "\partial_1"] \arrow[d, "\bar{f}_1"'] & C_0(K) \arrow[r] \arrow[d, "\bar{f}_0"'] & \cdots \\
\cdots \arrow[r] & C_1(K') \arrow[r, "\partial_1"'] \arrow[d, phantom, "\parallel"] & C_0(K) \arrow[r] & \cdots \\
 & \langle ab, ac, bc \rangle & &
\end{tikzcd}$$

$$f_1 : H_1(K) \longrightarrow H_1(K')$$

$$[\alpha] \longmapsto [\bar{f}_1(\alpha)]$$

e.g: for $\alpha = ab + bc + ac$

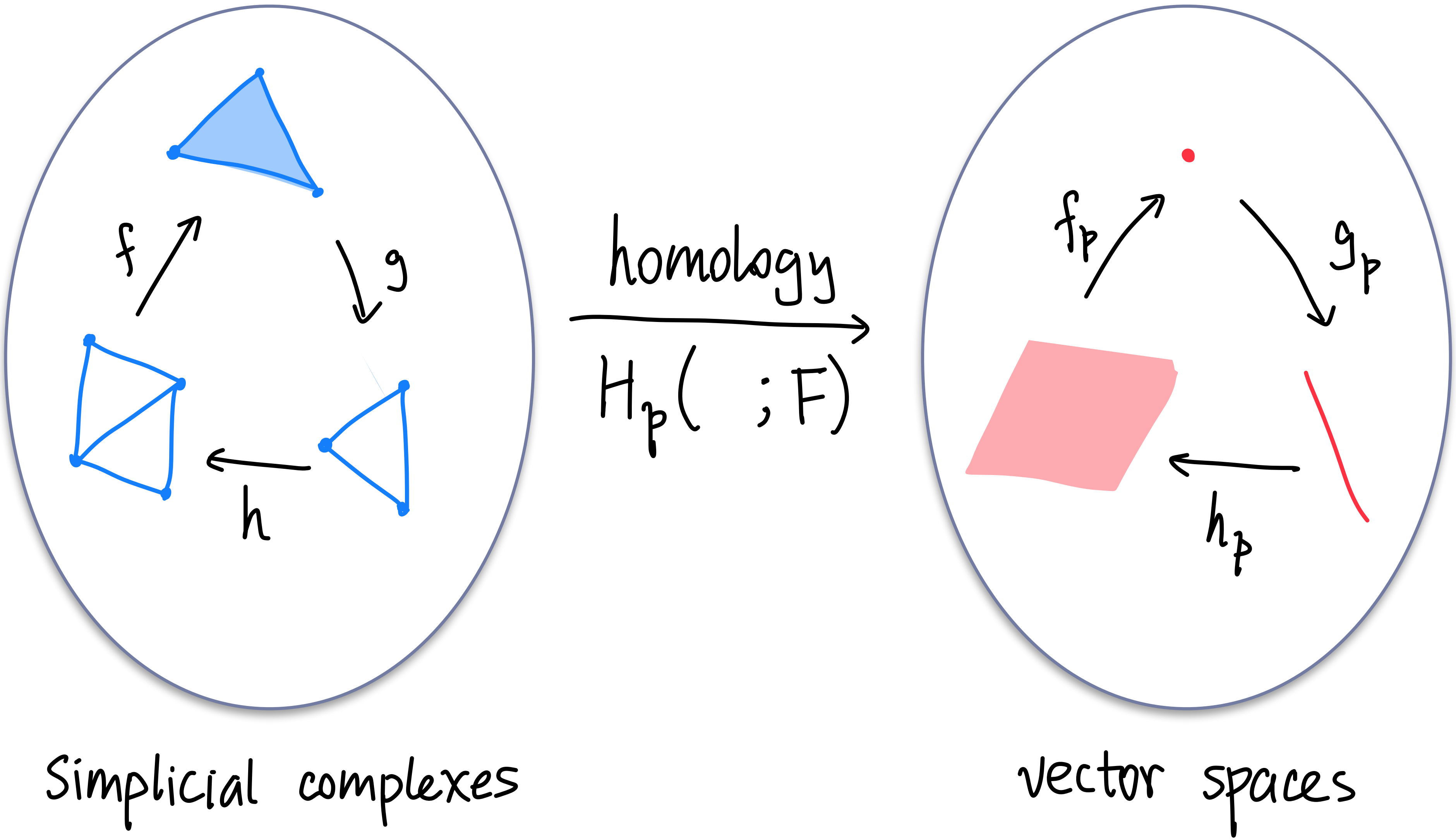
$$\Rightarrow f_1([\alpha]) = [\bar{f}_1(\alpha)] = [ab + bc + ac]$$

$\bar{f}_0 : a \longmapsto a,\ b \longmapsto b,\ c \longmapsto c,\ d \longmapsto c$

$\bar{f}_1 : ab \longmapsto ab,\ ac \longmapsto ac,\ bc \longmapsto bc,$

$cd \longmapsto 0$

Mind picture of functoriality

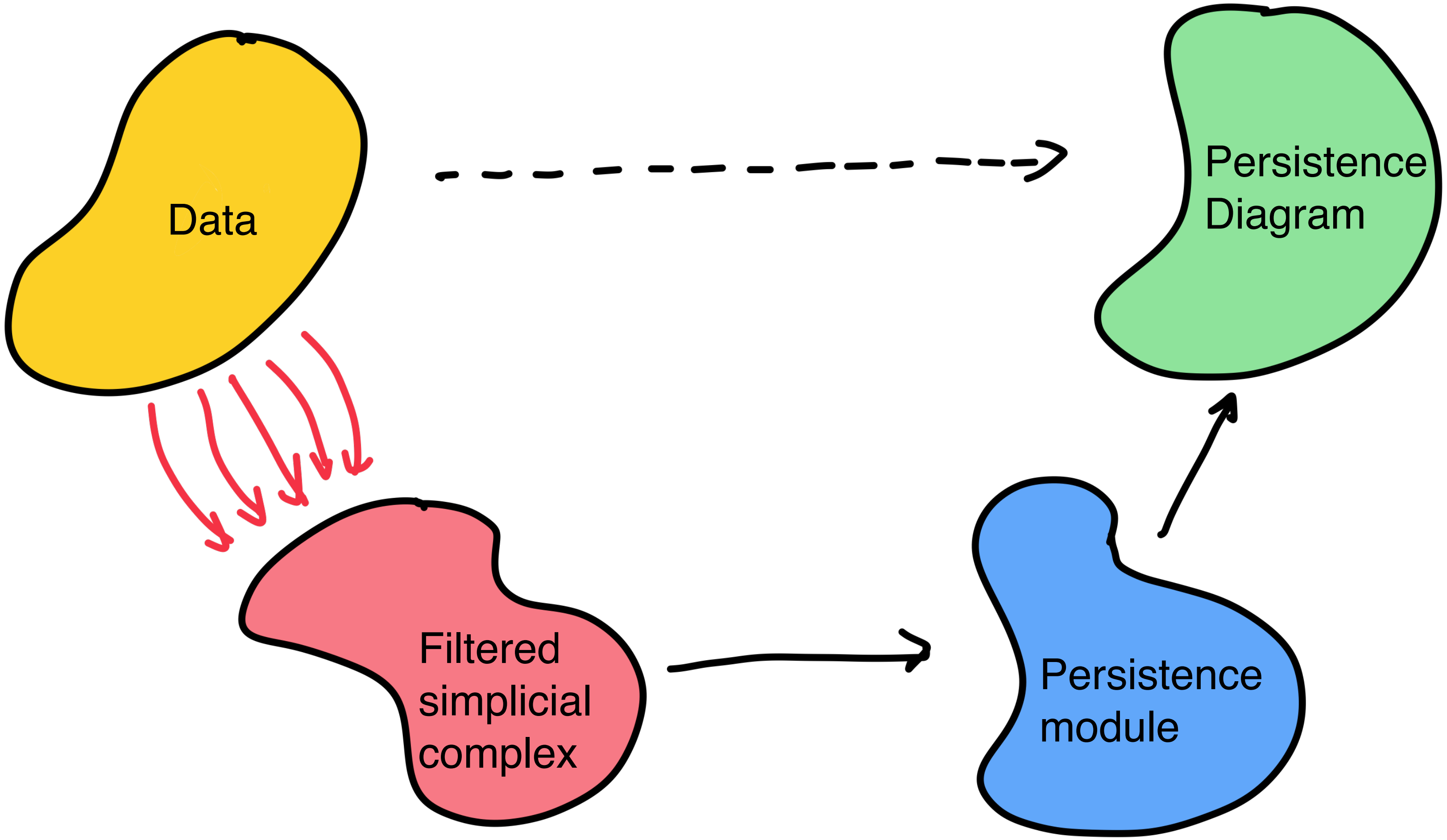


Topic 4: Intro to Persistent Homology

Persistent homology

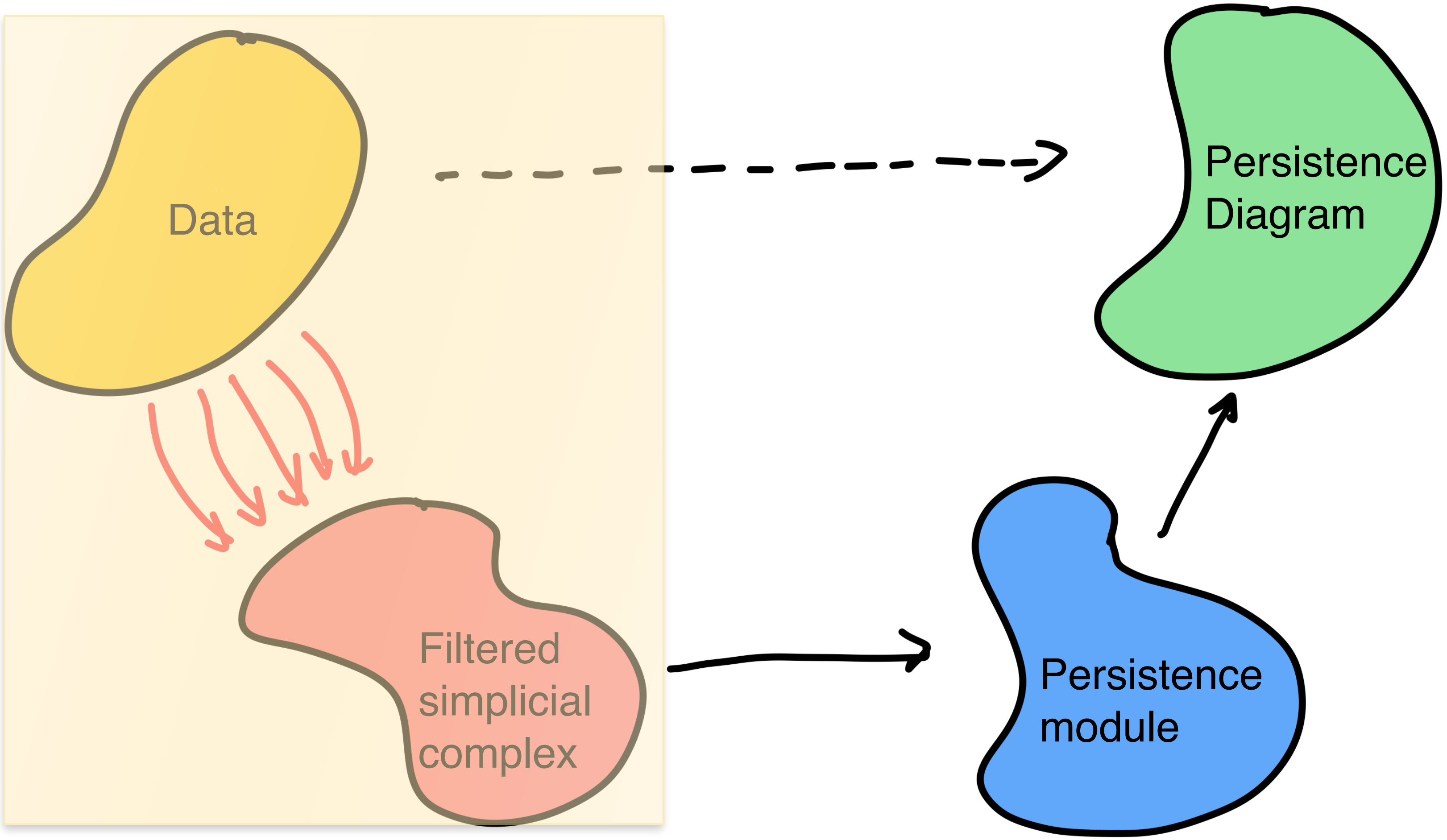
- A modern extension of homology to ``**sequence of spaces**”
  - [Edelsbrunner, Letcher, and Zomorodian, FOCS 2000]
  - Significantly broaden its practical power
- What is persistent homology (PH)
  - Motivation
  - Persistent betti numbers and persistence diagrams
- Algorithm(s) for persistent homology

Mind picture

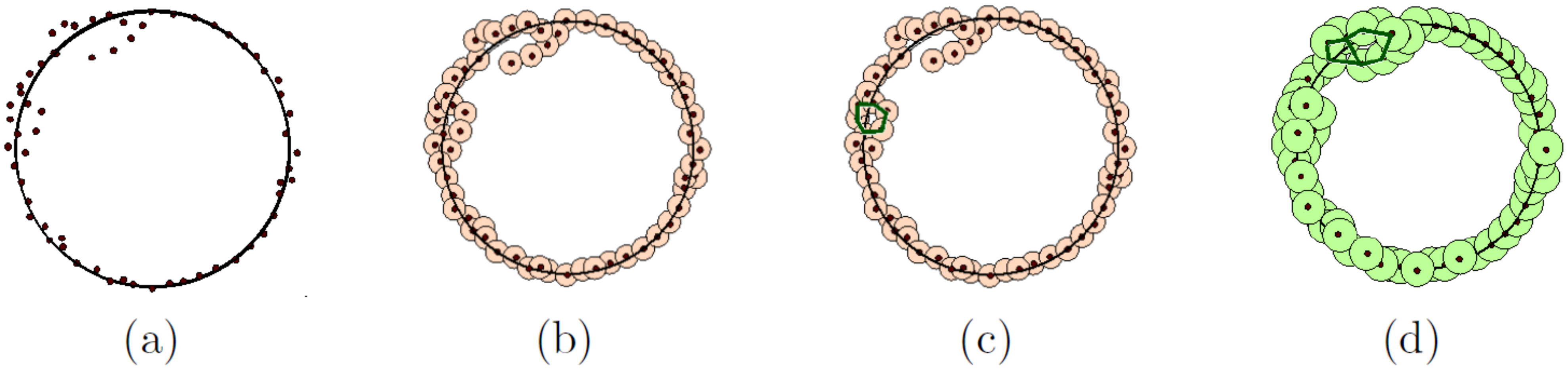


Filtrations

Filtered simplicial complex

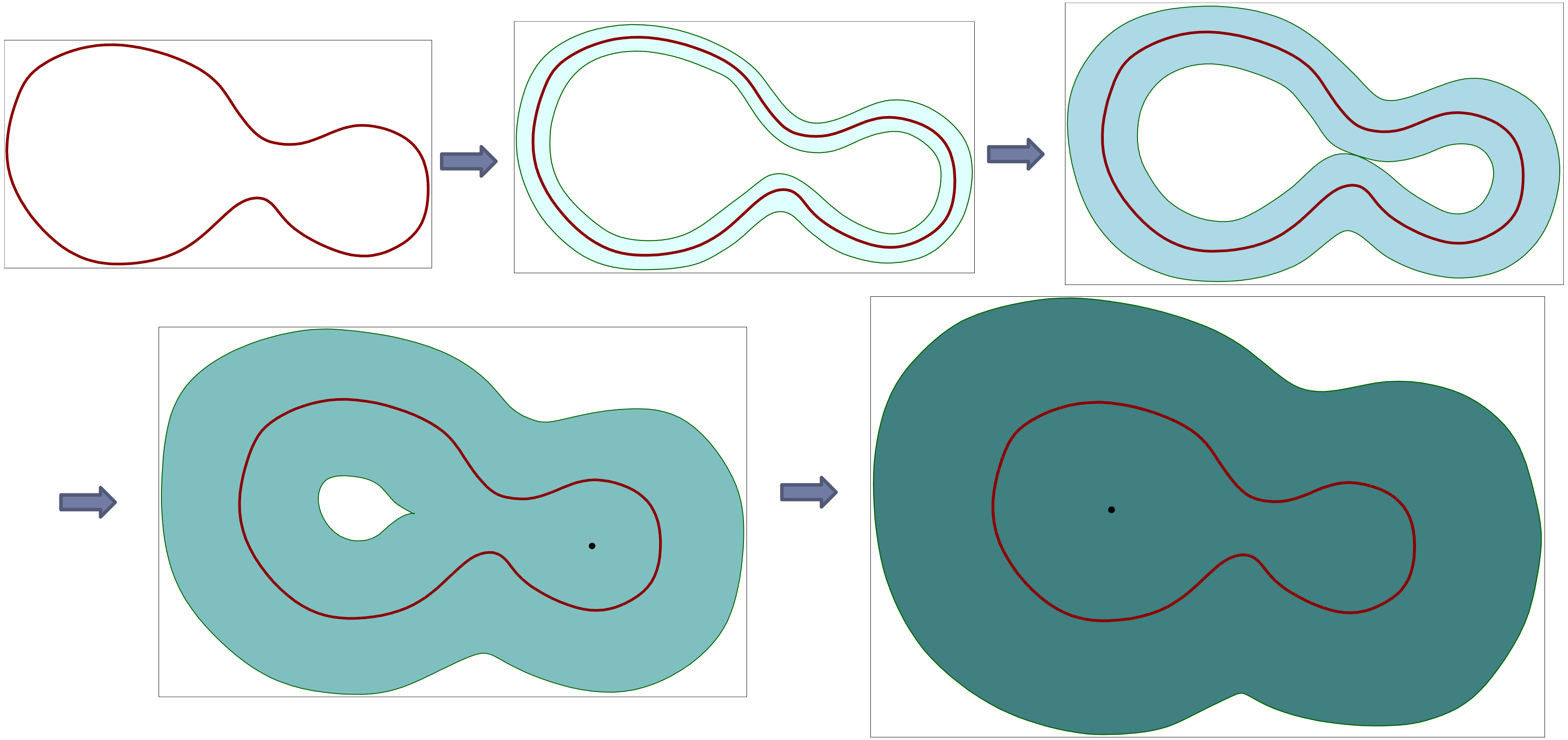


Issue of Scale



- Which scale to take?
- No single good scale!
- All scales?
- Some ``features” persists longer than others

Another Example

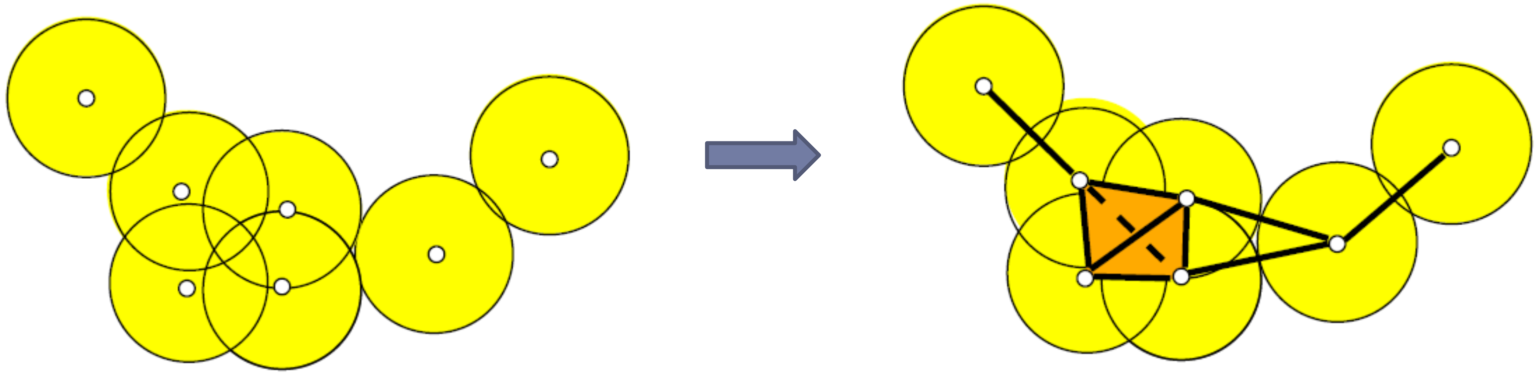


- Want to capture features of different ``sizes”

Recall: Čech Complex

$C^r(P) := Nrv\left(\{B(p_i, r) : p_i \in P\}\right)$

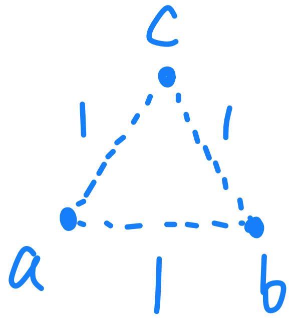
- Given a set of points $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$
- Given a real value $r > 0$, the *Čech complex* $C^r(P)$ is the nerve of the set of closed balls $\{B(p_i, r)\}_{i=1,\ldots,n}$, where $B(p, r) = \{x \in \mathbb{R}^d \mid d(p, x) \leq r\}$
  - i.e, $\sigma = \left\{p_{i_0}, \ldots, p_{i_s}\right\} \in C^r(P)$ iff $\bigcap_{j=0,\ldots,s} B(p_{i_j}, r) \neq \emptyset$
- The definition can be extended to a finite sample P of a metric space (X, d).



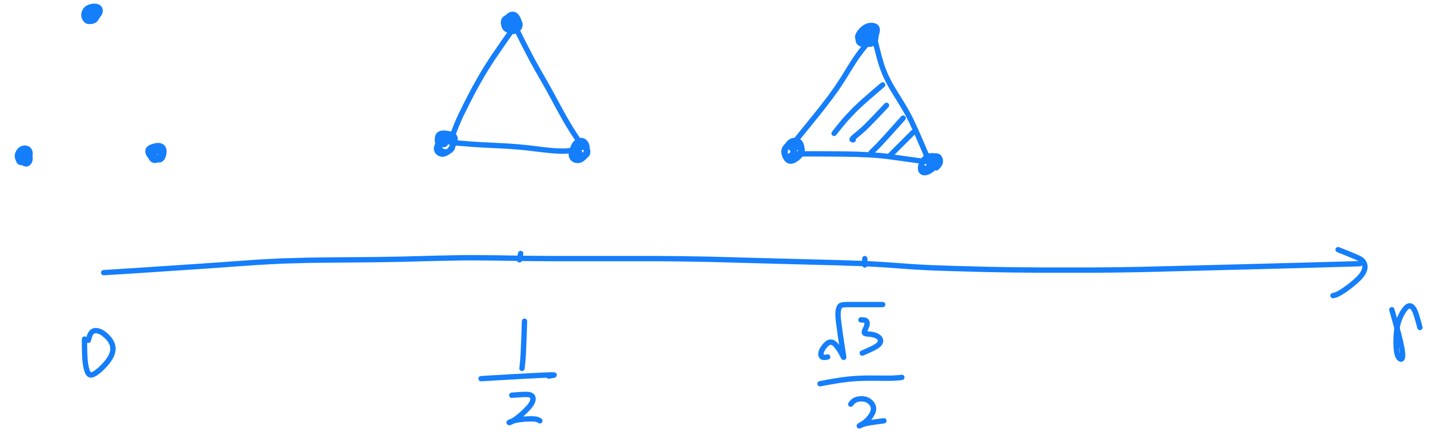
Čech Filtration

$C^r(P) = \{ p_{i_1} \cdots p_{i_k} \mid \bigcap_j B(p_{i_j}, r) \neq \emptyset \}$

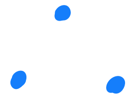
- Given a set of points $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$
- $(C^r(P))_{r \geq 0}$ is called the Čech filtration

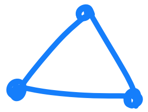


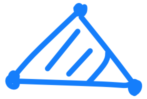
$C^r(\{a,b,c\})$:



$C^r(\{a,b,c\}) =$

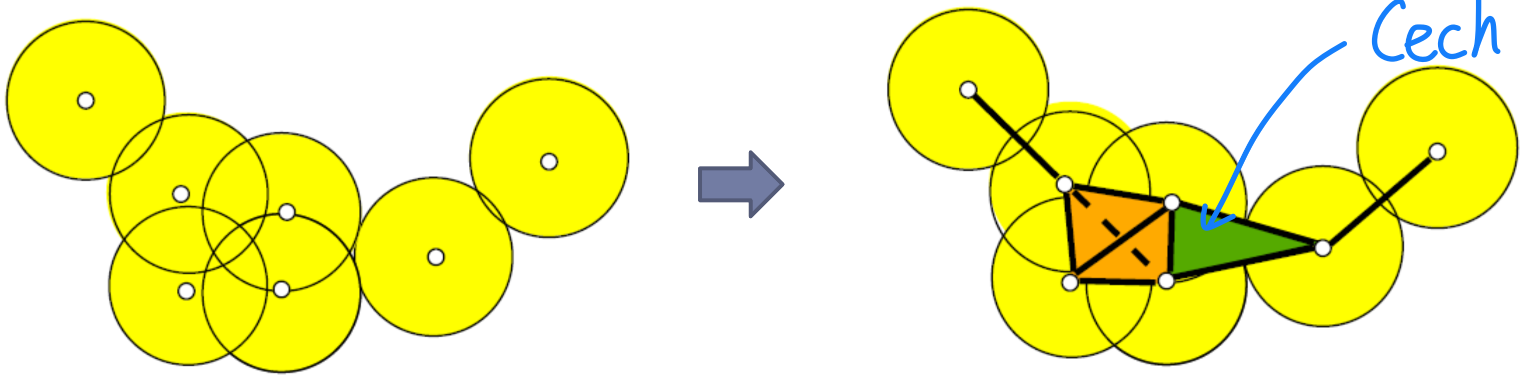

if $r \in [0, \frac{1}{2})$


if $r \in [\frac{1}{2}, \frac{\sqrt{3}}{2})$


if $r \geq \frac{\sqrt{3}}{2}$

Recall: Vietoris-Rips (Rips) Complex

- Given a set of points $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$
- Given a real value $r > 0$, the *Vietoris-Rips (Rips) complex* $Rips^r(P)$ is:
  - $\left\{ (p_{i_0}, \ldots, p_{i_k}) \mid B(p_{i_l}, r) \cap B(p_{i_j}, r) \neq \emptyset, \forall l, j = 0, \ldots, k \right\}.$
- More generally for P in a metric space (X, d):
  - $Rips^r(P) = \left\{ (p_{i_0}, \ldots, p_{i_k}) \mid d(p_{i_l}, p_{i_j}) \leq 2r, \forall l, j = 0, \ldots, k \right\}.$



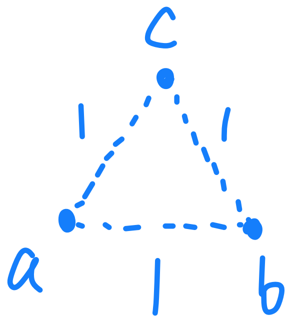
Čech cpx does NOT have this

Vietoris-Rips (Rips) Filtration

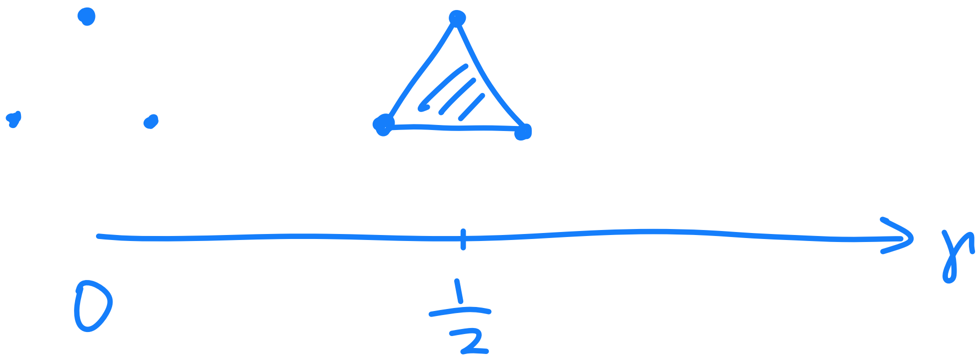
$\mathrm{Rips}^r(P) = \{ p_{i_1} \cdots p_{i_k} \mid d(p_{i_j}, p_{i_\ell}) \leq 2r \}$

Another convention is to use r instead.

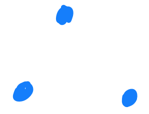
- Given a set of points $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$
- $(Rips^r(P))_{r \geq 0}$ is called the Vietoris-Rips (Rips) Filtration

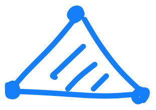


$\mathrm{Rips}^r(\{a,b,c\})$



$\mathrm{Rips}^r(\{a,b,c\}) = \begin{cases} \text{(see drawing 3)} & \text{if } r \in [0, \frac{1}{2}) \\ \text{(see drawing 4)} & \text{if } r \in [\frac{1}{2}, \infty) \end{cases}$





Finitely presented filtration

- A filtration $(K_t)_{t \in [0,\infty)}$ is called **finitely represented** if
  - There exist $0 = t_0 < t_1 < \cdots < t_n$ such that
  - $K_t = K_{t'}, \quad \forall t_i \leq t < t' < t_{i+1}$ and $i = 0,\ldots,n$ $(t_{n+1} := \infty)$

- So $(K_t)_{t \in [0,\infty)}$ is essentially the same as (or can be reconstructed from) $(K_{t_i})_{i=0,\ldots,n}$
- Both Čech and Rips filtrations are finitely represented