Persistent Homology

David Gu

Yau Mathematics Science Center
Tsinghua University
Computer Science Department
Stony Brook University

gu@cs.stonybrook.edu

August 28, 2020

Persistent Homology

Cech Complex

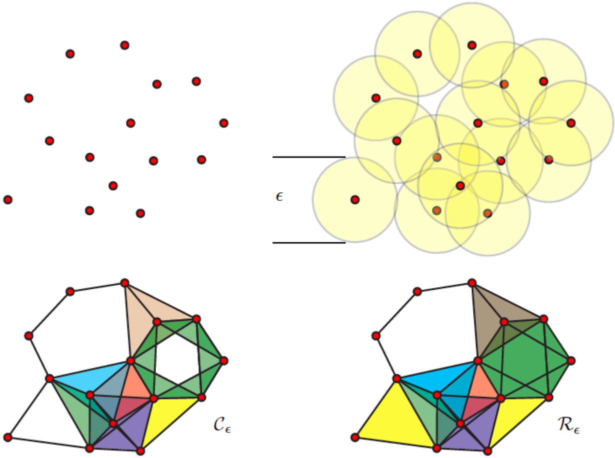


Figure: Cech complex.

Cech Complex

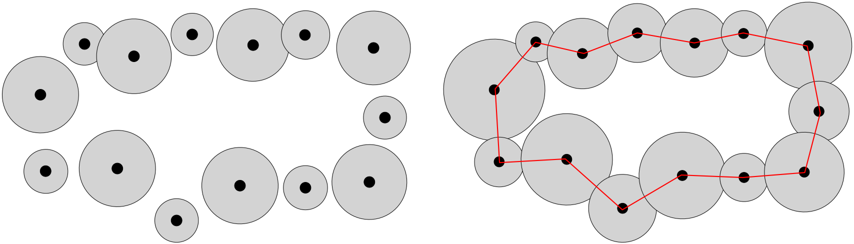
Definition (Cecg Complex)

Given a set of points $\{x_\alpha\}$ in Euclidean space $\mathbb{R}^n$, the Cech complex (also known as the nerve), $\mathcal{C}_\epsilon$, is the abstract simplicial complex where a set of $k+1$ vertices spans a k-simplex whenever the $k+1$ corresponding closed $\epsilon/2$-ball neighborhoods have nonempty intersection.

Definition (Vietoris-Rips Complex)

Given a set of points $\{x_\alpha\}$ in Euclidean space $\mathbb{R}^n$, the Vietoris-Rips complex, $\mathcal{R}_\epsilon$, is the abstract simplicial complex where a set S of $k+1$ vertices spans a k-simplex whenever the distance between any pair of points in S is at most ϵ.

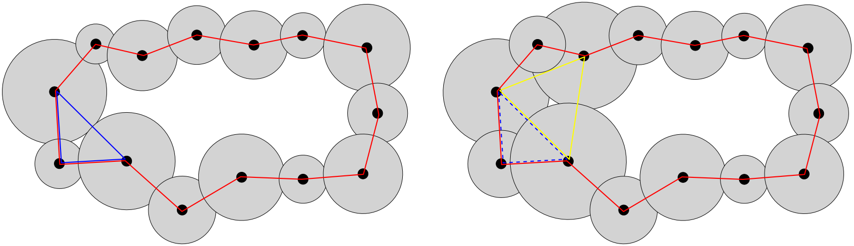
Cech Complex



$\alpha = \alpha_2$

Figure: Cech complex.

Cech Complex



$\alpha = \alpha_3$ $\alpha = \alpha_4$

Figure: Cech complex.

Cech Complex

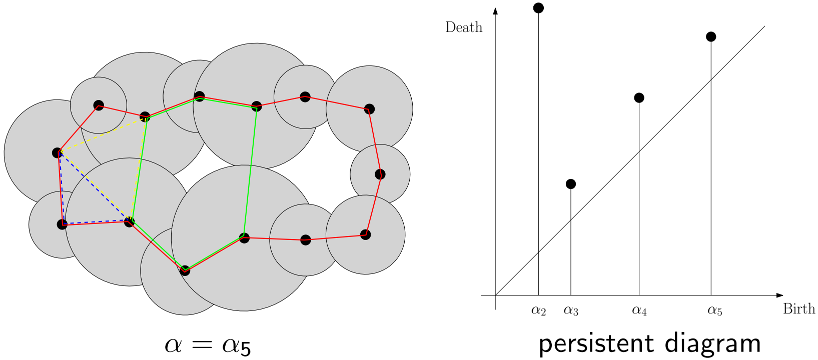


Figure: Cech complex.

Filtration

Definition (filtration)

A filtration of a simplicial complex $\mathbb{K}$ is a nested sequence of complexes,

$$\emptyset = \mathbb{K}_{-1} \subset \mathbb{K}_0 \subset \mathbb{K}_1 \subset \cdots \subset \mathbb{K}_n = \mathbb{K}.$$

Example

Suppose $\mathbb{K}$ is a simplicial complex, we sort all the simplices in a sequence

$$\sigma_1^0, \sigma_2^0, \cdots, \sigma_{n_0}^0, \sigma_1^1, \sigma_2^1, \cdots, \sigma_{n_1}^1, \sigma_1^2, \sigma_2^2, \cdots, \sigma_{n_2}^2.$$

where σ_i^k is the i-th k-simplex in $\mathbb{K}$. Then we relabel all the simplices as

$$\sigma^0, \sigma^1, \sigma^2, \cdots,$$

We define $\mathbb{K}_i$ as the union of $\sigma^0, \sigma^1, \ldots, \sigma^i$.

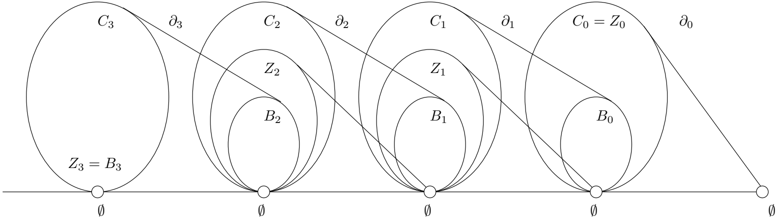


Figure: Chain, cycle, boundary groups and their images under the boundary operators.

$$H_k(\mathbb{K}, \mathbb{Z}_2) = \frac{\text{Ker}\partial_k}{\text{Img}\partial_{k+1}} = \frac{Z_k}{B_k}.$$

Persistent Homology

The inclusion map $f : \mathbb{K}_{i-1} \hookrightarrow \mathbb{K}_i$ defined by $f(x) = x$ induces a homomorphism $f_* : H_p(\mathbb{K}_{i-1}) \to H_p(\mathbb{K}_i)$. The nested sequence of complexes corresponds to a sequence of homology groups connected by the induced maps,

$$0 = H_p(\mathbb{K}_{-1}) \to H_p(\mathbb{K}_0) \to \cdots \to H_p(\mathbb{K}_n) = H_p(\mathbb{K})$$

Persistent homology studies how the homology groups change over the filtration.

Generator and Killer

Definition (positive simplex)

Given a filtration of $\mathbb{K}$, suppose $\mathbb{K}_i - \mathbb{K}_{i-1} = \sigma_i$, where σ_i is a $(k+1)$-simplex. We call σ_i is positive if it belongs to a $(k+1)$-cycle in $\mathbb{K}_i$ and negative otherwise.

A positive simplex is also called a generator, a negative simplex a killer.

Generator and Killer

Definition (Betti Number)

Given a complex K, the i-th Betti number β_i is the rank of $H_i(K)$,

$$\beta_i = \text{Rank} H_i(K, \mathbb{Z}_2)$$

Suppose the number of positive k-simplexes is pos_k, and the number of negative k-simplexes is neg_k, then

$$\boxed{\beta_k = \text{pos}_k - \text{neg}_{k+1}}$$

Persistent Homology

Definition (Persistent Homology)

Define Z_k^l, B_k^l be the K-th cycle group and k-th boundary group respectively, of the l-complex K^l in a filtration. The p-persistent k-th homology group K^l is

$$H_k^{l,p} := \frac{Z_k^l}{B_k^{l+p} \bigcap Z_k^l}.$$

The p-persistent k-th Betti number $\beta_k^{l,p}$ of K^l is the rank of $H_k^{l,p}$.

Lemma

Consider the homomorphism $\eta_k^{l,p} : H_k^l \to H_k^{l+p}$, then

$$img\ \eta_k^{l,p} \cong H_k^{l,p}$$

Generator

Lemma

For each positive k-simplex σ^i, there exists a non-exact k-cycle c^i, c^i contains σ^i but no other positive k-simplices.

Proof.

Start with an arbitrary a k-cycle that contains σ^i and remove other positive k-simplices by adding their corresponding k-cycles. This method succeeds because each added cycle contains only one positive k-simplex by inductive assumption. □

We use σ^i to represent c^i, and in turn the homologous class $[c^i] = c^i + B_k$.

$$\sigma^i \to c^i \to [c^i] = c^i + B_k. \quad \sigma^i \sim c^i$$

We add $[c^i]$ to the basis of $H_k(\mathbb{K}^i)$.

Generator

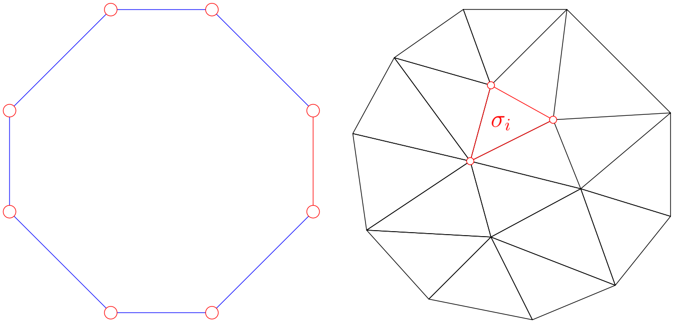


Figure: Generator, positive simplex.

Killer

For each negative $(k+1)$-simplex σ^j, its boundary $d = \partial_{k+1}\sigma^j$ is a k-cycle, and can be represented as the linear combination of the basis of $H_k(\mathbb{K}_{j-1})$,

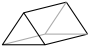
$$[d] = \sum_g [c^g], \{c^g\} \text{ basis } H_k(\mathbb{K}_{j-1}),$$

each $[c^g]$ is represented by a positive k-simplex σ^g, $g < j$, that is not yet paired. The collection of positive non-paired k-simplices is denoted as $\Gamma = \Gamma(d)$,

$$\Gamma(d) := \left\{ \sigma^g : [d] = \sum_g [c^g], \quad \sigma^g \sim c^g \right\}$$

Suppose the youngest positive simplex in $\Gamma(\partial_{k+1}\sigma^j)$ is σ^i, then we form the pair (σ^i, σ^j), and remove $[c^i]$ from $H_k(\mathbb{K}_j)$.
$[c^i]$ is created by σ^i and killed by σ^j, the persistence life of the k-cycle $[c^i]$ is $j - i - 1$.

Example Filtration



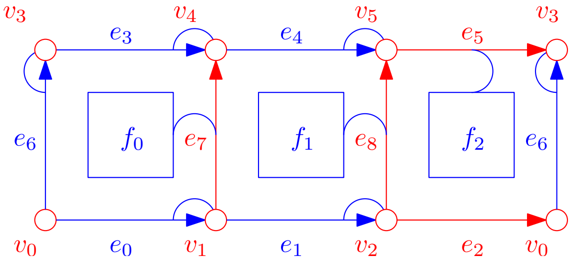


Figure: Generators and killers.

Filtration

$$v_0, v_1, v_2, v_3, v_4, v_5, e_0, e_1, e_2, e_3, e_4, e_5, f_0, f_1, f_2$$

Relabel them as

$$\sigma^0, \sigma^1, \sigma^2, \sigma^3, \sigma^4, \sigma^5, \sigma^6, \sigma^7, \sigma^8, \sigma^9, \sigma^{10}, \sigma^{11}, \sigma^{12}, \sigma^{13}, \sigma^{14}, \sigma^{15}$$

Example Generators

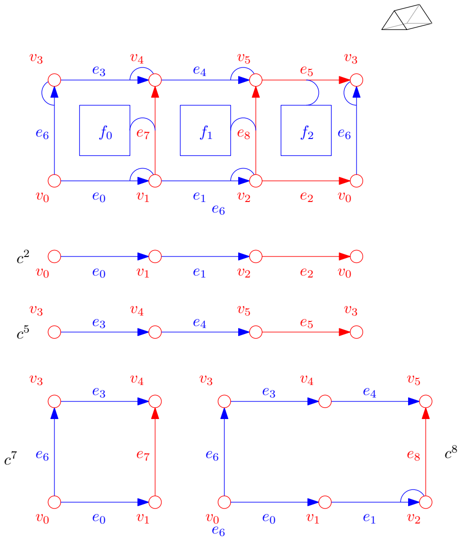


Figure: Generators.

Example Killers

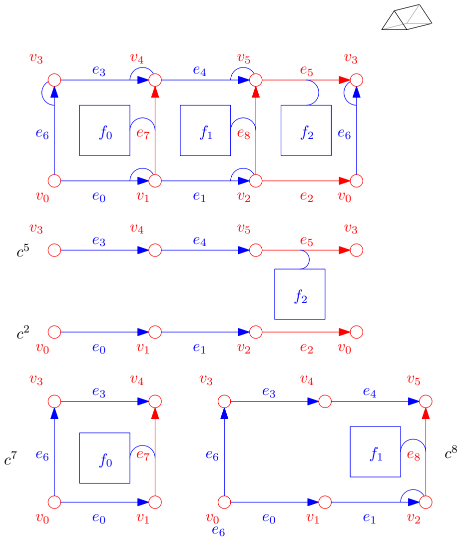
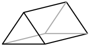


Figure: Killers.

Example Pairing



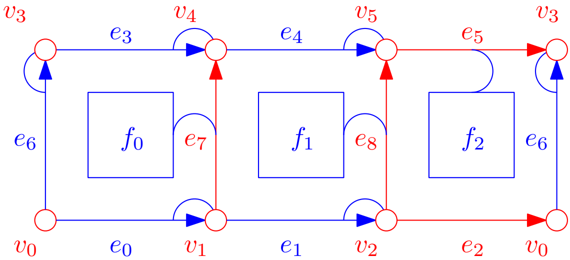


Figure: Generators and killers.

$$\partial_2 f_2 = e_2 + e_5 + e_6 + e_8 = (e_5 + 2e_4 + 2e_3) + (e_2 + 2e_1 + 2e_0) + e_6 + e_8$$
$$= (e_5 + e_4 + e_3) + (e_2 + e_1 + e_0) + \partial_2(f_0 + f_1)$$

Key Lemma

Definition (Collision Free Cycle)

A collision free cycle is one where the youngest positive simplex has not been paired (killed).

Lemma (Collision)

Given a filtration, $\mathbb{K}_j - \mathbb{K}_{j-1} = \sigma^j$, σ^i *is the youngest positive simplex in* $\Gamma(\partial_{k+1}\sigma^j)$. *Let* e *be a collision free* k*-cycle in* $\mathbb{K}_{j-1}$ *homologous to* $\partial_{k+1}\sigma^j$. *Suppose the youngest positive simplex in* e *is* σ^g, *then*

$$\sigma^i = \sigma^g.$$

$$\max \Gamma(\partial_{k+1}\sigma^j) = \max(e) \quad \forall e \text{ collision free}, [e] = [\partial\sigma^j].$$

Key Lemma

Proof.

Let f be the sum of the basis cycles, homologous to $d = \partial_{k+1}\sigma^j$. By definition, f's youngest positive simplex is σ^i, namely the youngest simplex in $\Gamma(\partial_{k+1}\sigma^j)$,

$$\sigma^i = \max \Gamma(\partial_{k+1}\sigma^j).$$

This implies that there are no cycles homologous to d in $\mathbb{K}_{i-1}$ or earlier complexes. Let σ^g be the youngest positive simplex in e. $[e] = [d]$, therefore $g \geq i$.
If $g > i$, then $e = f + c$, where c bounds in $\mathbb{K}^{j-1}$. $\sigma^g \notin f$,implies $\sigma^g \in c$, and as σ^g is the youngest in e, it is also the youngest in c. □

Key Lemma

continued.

Since e is collision free, the cycle created by σ^g, denoted as c^g, is still a non-boundary cycle in $\mathbb{K}_{j-1}$. Hence c^g can't be c, and can't be homologous to c when c becomes a boundary. Namely, when c is killed, σ^g is not paired yet.
It follows that the negative $(k+1)$-simplex that kills c must pair a positive k-simplex in c, which is younger than σ^g, a contradiction.

This lemma shows, when σ^j is added to $\mathbb{K}_{j-1}$, we need to find any collision free cycle e homologous to $\partial_{k+1}\sigma^j$, and pair σ^j with the youngest positive simplex of e.

Pair Algorithm

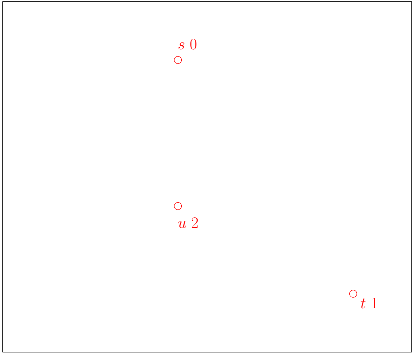
Pair(σ)

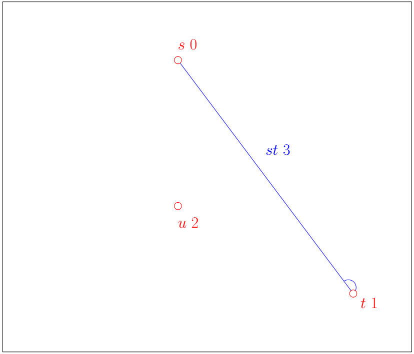
1. $c = \partial_p \sigma$
2. τ is the youngest positive $(p-1)$-simplex in c.
3. **while** σ is paired and c is not empty **do**
4. find (τ, d), d is the p-simplex paired with τ;
5. $c \leftarrow \partial_p d + c$
6. Update τ to be the youngest positive $(p-1)$-simplex in c
7. **end while**
8. **if** c is not empty **then**
9. σ is negative p-simplex and paired with τ
10. **else**
11. σ is a positive p-simplex
12. **endif**

Handle Loop and Tunnel Loop

1. The simplices on the surface M are added into the filtration in any arbitrary order. Since $H_1(M)$ is of rank $2g$, the algorithm Pair generates $2g$ number of unpaired positive edges.
2. The simplices up to dimension 2 in I are added into the filtration. Since $H_1(I)$ of rank g, half of $2g$ positive edges generated in step 1 get paired with the negative triangles in I. Each pair correponds to a killed loop, these g loops are handle loops.
3. Or the simplices up to dimension 2 in O are added into the filtration. Since $H_1(O)$ of rank g, half of $2g$ positive edges genrated in step 2 get paried with the negative triangles in O. Each pair corresponds to a killed loop, these g loops are tunnel loops.

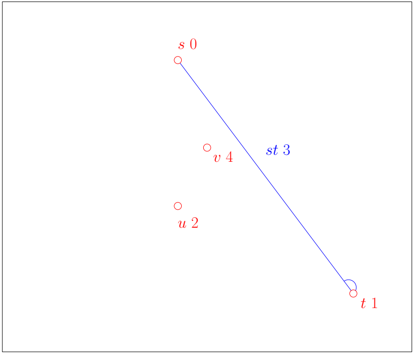
Example

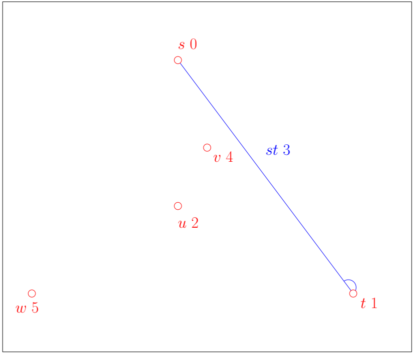




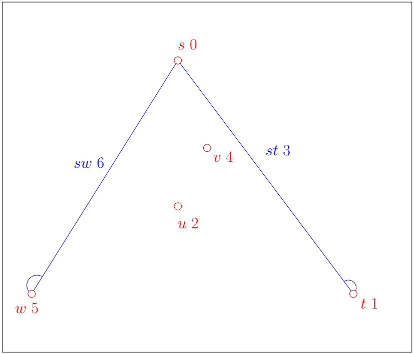
3. $\partial st = s + t$, (t_1, st_3)

Example

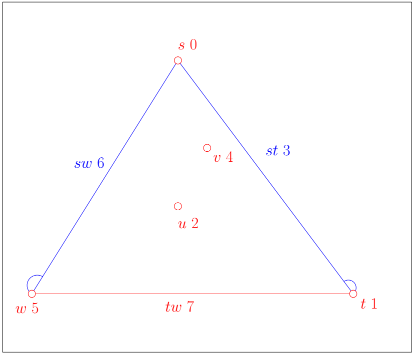




Example



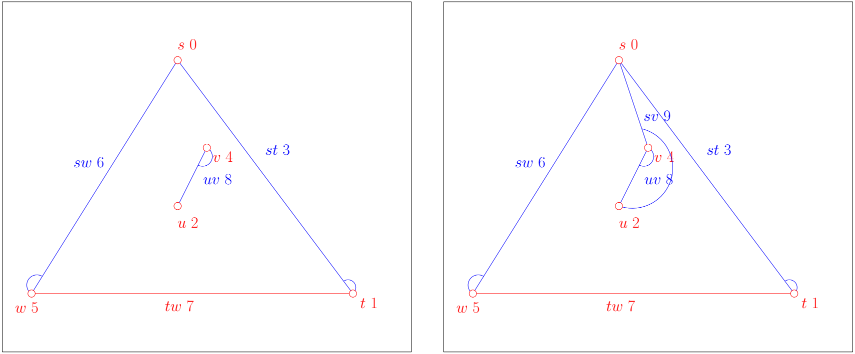
6. $\partial sw = s + w$, (w_5, sw_6)



7. tw

$$
\begin{aligned}
7 : \partial tw = w + t &= w + t + \partial st + \partial sw \\
&= w + t + (s + t) + (s + w) \\
&= 0.
\end{aligned}
$$

Example

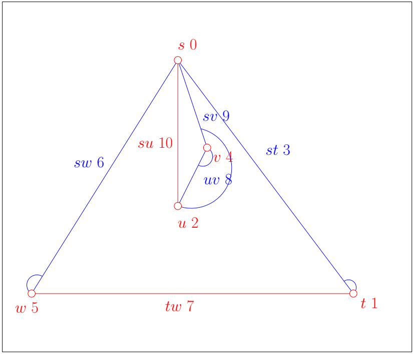


8. $\partial uv = u + v$, (v_4, uv_8)

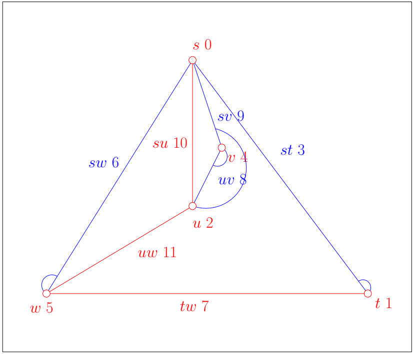
9. (u_2, sv_9)

$$
\begin{aligned}
9.\ \partial sv = s + v &= s + v + \partial uv \\
&= s + v + (u + v) \\
&= s + u
\end{aligned}
$$

.


s 0
sv 9
su 10
st 3
sw 6
v 4
uv 8
u 2
w 5
tw 7
t 1

10. *su*


s 0
sv 9
su 10
st 3
sw 6
v 4
uv 8
u 2
uw 11
w 5
tw 7
t 1

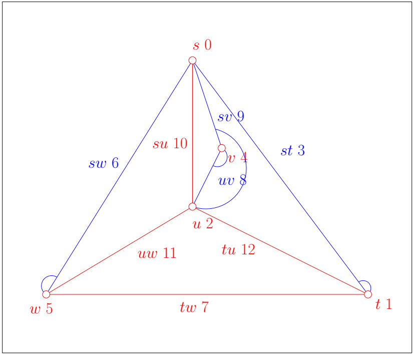
11. *uw*

Example

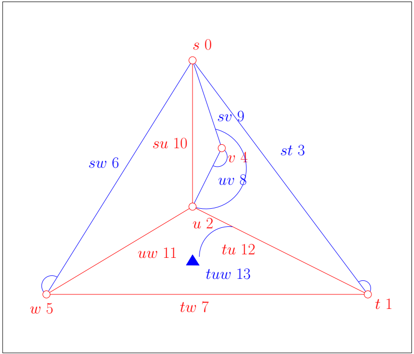
$$
\begin{aligned}
10.\ \partial su &= s + u = s + u + \partial sv \\
&= s + u + (s + v) \\
&= u + v = u + v + \partial uv \\
&= 0.
\end{aligned}
$$

$$
\begin{aligned}
11.\ \partial uw &= u + w = u + w + \partial sw \\
&= u + w + (s + w) \\
&= s + u = s + u + \partial sv \\
&= s + u = s + u + (s + v) \\
&= u + v = u + v + \partial uv \\
&= u + v + (u + v) \\
&= 0.
\end{aligned}
$$

Example



12. *tu*



13. *tuw*,(tu_{12}, tuw_{13})

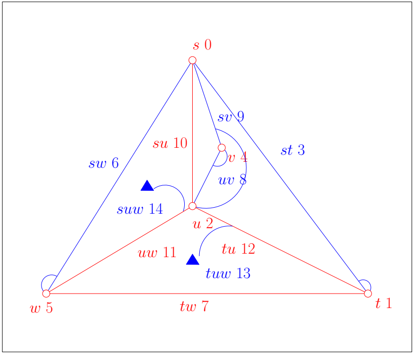
Example

$$
\begin{aligned}
12.\ \partial tu &= t + u = t + u + \partial sv \\
&= t + u + (s + v) \\
&= t + u + s + v + \partial uv \\
&= t + u + s + v + (u + v) \\
&= t + s \\
&= s + t + \partial st \\
&= s + t + (t + s) \\
&= 0.
\end{aligned}
$$

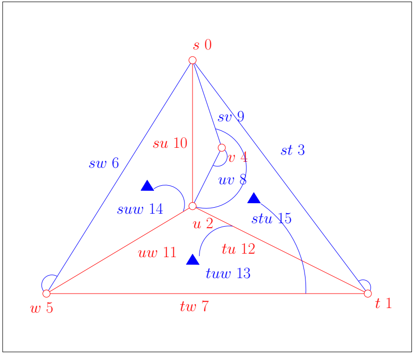
$$
13.\ \partial tuw = tu + uw + wt
$$

$$
(tuw, tu)
$$

Example



14. $\partial suw = uw + su + sw$
(uw_{11}, suw_{14})



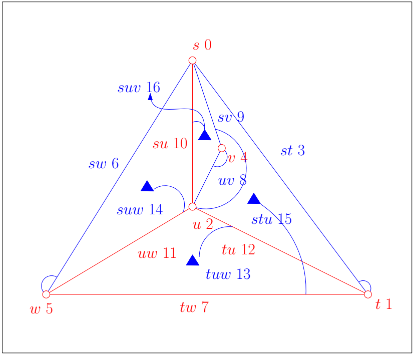
15. $stu, (tw_7, stu_{15})$

Example

$$
\begin{aligned}
15.\ \partial stu &= su + tu + st \\
&= su + st + tu + \partial tuw \\
&= su + st + tu + (tu + uw + tw) \\
&= su_{10} + st + uw_{11} + tw_7 \\
&= su_{10} + st + uw_{11} + tw_7 + \partial suw \\
&= su_{10} + st + uw_{11} + tw_7 + (sw + su_{10} + uw_{11}) \\
&= st + tw_7 + sw
\end{aligned}
$$

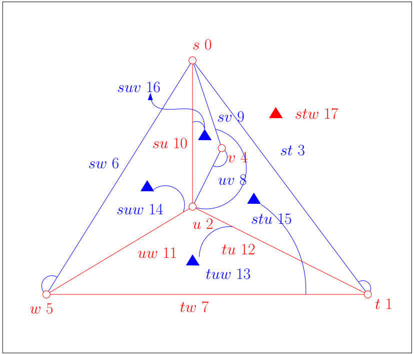
Hence we obtain the pair (stu, tw).

Example



16. $\partial suv = su + uv + sv$

(su_{10}, suv_{16})



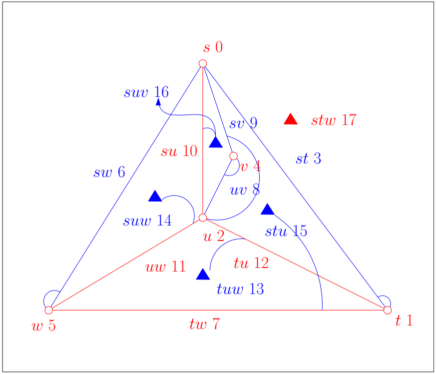
17. stw

Example

17. $\partial stw = tw + sw + st$

$$
\begin{aligned}
&= sw + st + tw + \partial stu \\
&= sw + st + tw_7 + (st + tu_{12} + us_{10}) \\
&= sw + st + tw_7 + (st + tu_{12} + us_{10}) + \partial tuw \\
&= sw + st + tw_7 + (st + tu_{12} + us_{10}) + (tu_{12} + uw_{11} + wt_7) \\
&= sw + us_{10} + uw_{11} \\
&= sw + us_{10} + uw_{11} + \partial suw \\
&= sw + us_{10} + uw_{11} + (su_{10} + uw_{11} + ws) \\
&= 0.
\end{aligned}
$$

Example



Creater	*Killer*
t_1	st_3
u_2	sv_9
v_4	uv_8
w_5	sw_6
tw_7	stu_{15}
su_{10}	suv_{16}
uw_{11}	suw_{14}
tu_{12}	tuw_{13}

Incidence Matrix

Assuming an ordering of the $(p-1)$ simplices and of the p-simplices, the boundary of a p-chian can be obtained by multiplication of the corresponding vector with the incidence matrix,

$$\partial(c_p) = D_p c_p.$$

The incidence matrix is defined as

$$D_p[i,j] = \begin{cases} 1 & \sigma_i^{p-1} \in \sigma_j^p \\ 0 & \sigma_i^{p-1} \notin \sigma_j^p \end{cases}$$

Incidence Matrix and Betti Number

A classic algorithm computes the Betti numbers of K by reducing its incidence matrices to Smith normal form. It uses row and column operations to zero out all entries except along an initial portion of the diagonal.

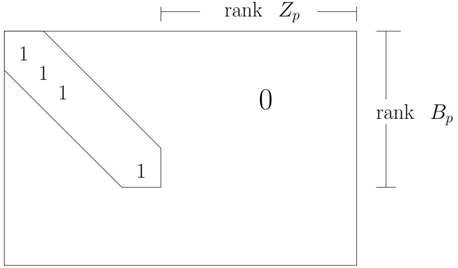


Figure: Smith norm of incidence matrix in $\mathbb{Z}_2$.

The Betti number

$$\boxed{\beta_p = \text{rank}Z_p - \text{rank}B_p}.$$

Pairing Algorithm

Definition (Monotonous Filtering)

A filtering is monotonous, if in the ordering of K, any simplex σ is proceeded by its faces.

An algorithm computes the persistence diagrams by pairing the simplices, which uses column operator to reduce D and another $0-1$ matrix R. Let $\text{low}_R(j)$ be the row index of the last 1 in column j of R, and (undefined if the column is zero).

Definition (Reduced Matrix and Pairing)

We call R reduced and low_R a pairing function, if

$$\text{low}_R(j) \neq \text{low}_R(j'),$$

whenever $j \neq j'$ specify two non-zero columns.

Pairing Algorithm

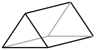
Algorithm: Incidence matrix reduction

1. $R \leftarrow D$
2. **for** $j = 1$ **to** n **do**
3. **while** $\exists j' < j$ **with** $\text{low}_R(j') = \text{low}_R(j)$ **do**
4. add column j' to column j
5. **endwhile**
6. **endfor**.

The pairing is given by

$$\boxed{(\sigma_i, \sigma_j) \iff i = \text{low}_R(j)}.$$

σ_i is positive, it generates a homology class; σ_j is negative, it kills a homolog class.



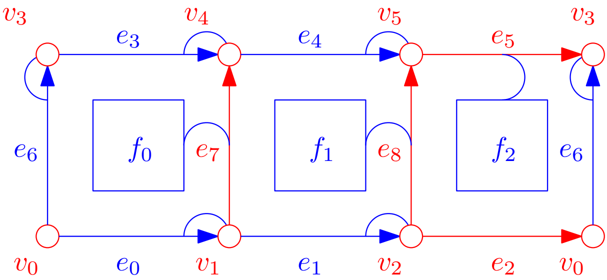


Figure: Generators and killers.

Pairing by matrix induction

Boundary operator ∂_1, incidence matrix D_1,

	e_0	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8
v_0	1	0	1	0	0	0	1	0	0
v_1	1	1	0	0	0	0	0	1	0
v_2	0	1	1	0	0	0	0	0	1
v_3	0	0	0	1	0	1	1	0	0
v_4	0	0	0	1	1	0	0	1	0
v_5	0	0	0	0	1	1	0	0	1

Pairing by matrix induction

$1+2,\ 4+5,\ 3+7,\ 4+8$

	e_0	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8
v_0	1	0	1	0	0	0	1	0	0
v_1	1	1	1	0	0	0	0	1	0
v_2	0	1	0	0	0	0	0	0	1
v_3	0	0	0	1	0	1	1	1	0
v_4	0	0	0	1	1	1	0	0	1
v_5	0	0	0	0	1	0	0	0	0

Pairing by matrix induction

$1+2$, $3+5$, $3+8$

	e_0	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8
v_0	1	0	0	0	0	0	1	0	0
v_1	1	1	0	0	0	0	0	1	0
v_2	0	1	0	0	0	0	0	0	1
v_3	0	0	0	1	0	0	1	1	1
v_4	0	0	0	1	1	0	0	0	0
v_5	0	0	0	0	1	0	0	0	0

Pairing by matrix induction

$6+7$, $6+8$

	e_0	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8
v_0	1	0	0	0	0	0	1	1	1
v_1	1	1	0	0	0	0	0	1	0
v_2	0	1	0	0	0	0	0	0	1
v_3	0	0	0	1	0	0	1	0	0
v_4	0	0	0	1	1	0	0	0	0
v_5	0	0	0	0	1	0	0	0	0

Pairing by matrix induction

$0 + 7, 1 + 8,\ 0 + 8$

	e_0	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8
v_0	1	0	0	0	0	0	1	0	0
v_1	1	1	0	0	0	0	0	0	0
v_2	0	1	0	0	0	0	0	0	0
v_3	0	0	0	1	0	0	1	0	0
v_4	0	0	0	1	1	0	0	0	0
v_5	0	0	0	0	1	0	0	0	0

Generators e_2, e_5, e_7, e_8, corresponding to 0 columns. Killers corresponds to non-zero columns. Pairing

$$(e_0, v_1), (e_1, v_2), (e_3, v_4), (e_4, v_5), (e_6, v_3)$$

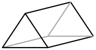
Pairing by matrix induction

$$
\begin{array}{|c|ccc|}
\hline
 & f_0 & f_1 & f_2 \\
\hline
e_0 & 1 & 0 & 0 \\
e_1 & 0 & 1 & 0 \\
e_2 & 0 & 0 & 1 \\
e_3 & 1 & 0 & 0 \\
e_4 & 0 & 1 & 0 \\
e_5 & 0 & 0 & 1 \\
e_6 & 1 & 0 & 1 \\
e_7 & 1 & 1 & 0 \\
e_8 & 0 & 1 & 1 \\
\hline
\end{array}
=
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
1 & 0 & 1 \\
1 & 1 & 0 \\
0 & 1 & 1
\end{bmatrix}
\to^{2+3}
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 1 \\
0 & 0 & 1 \\
1 & 0 & 0 \\
0 & 1 & 1 \\
0 & 0 & 1 \\
1 & 0 & 1 \\
1 & 1 & 1 \\
0 & 1 & 0
\end{bmatrix}
\to^{1+3}
\begin{bmatrix}
1 & 0 & 1 \\
0 & 1 & 1 \\
0 & 0 & 1 \\
1 & 0 & 1 \\
0 & 1 & 1 \\
0 & 0 & 1 \\
1 & 0 & 0 \\
1 & 1 & 0 \\
0 & 1 & 0
\end{bmatrix}
$$

The pairing is

$$(f_0, e_7), (f_1, e_8), (f_2, e_5)$$

Topological Annulus



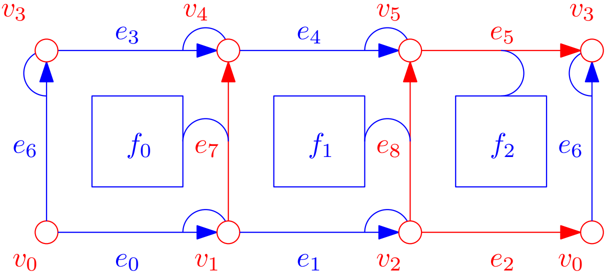
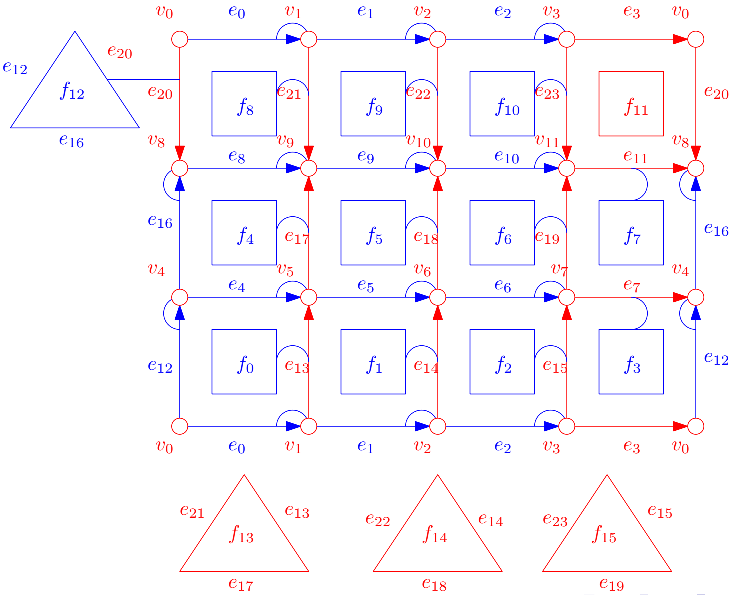
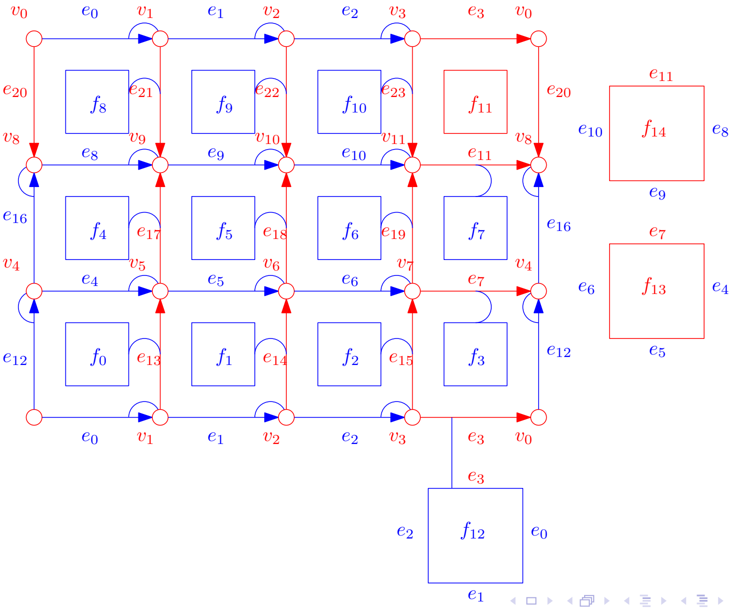


Figure: Topological Annulus.

Topological Torus



Topological Torus



Topological Torus

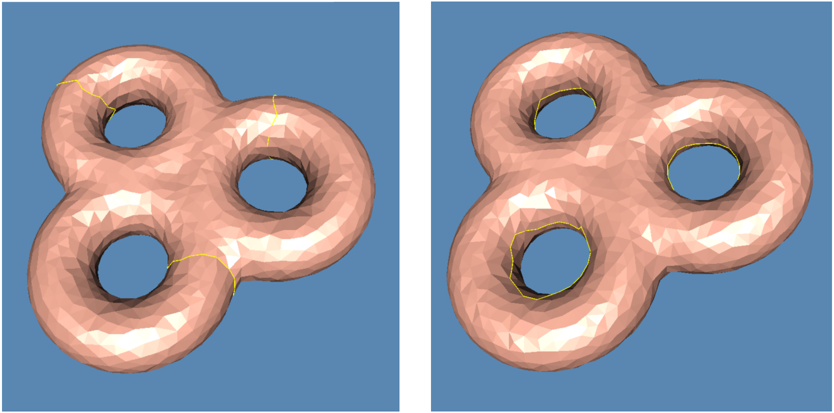


Figure: Handle and tunnel loops.

Topological Torus

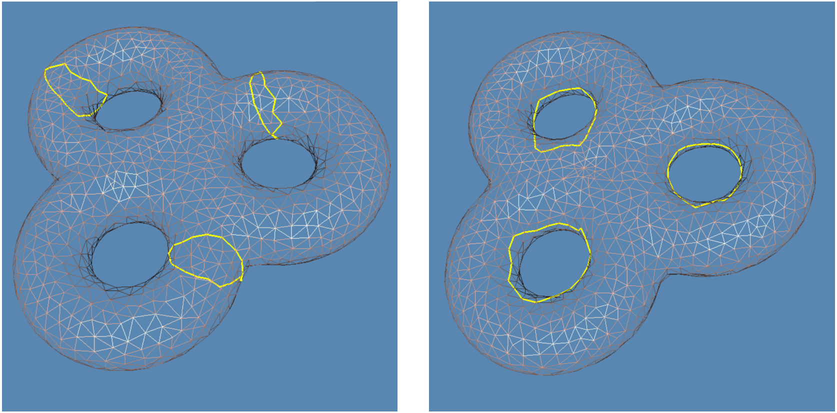


Figure: Handle and tunnel loops.

Topological Torus

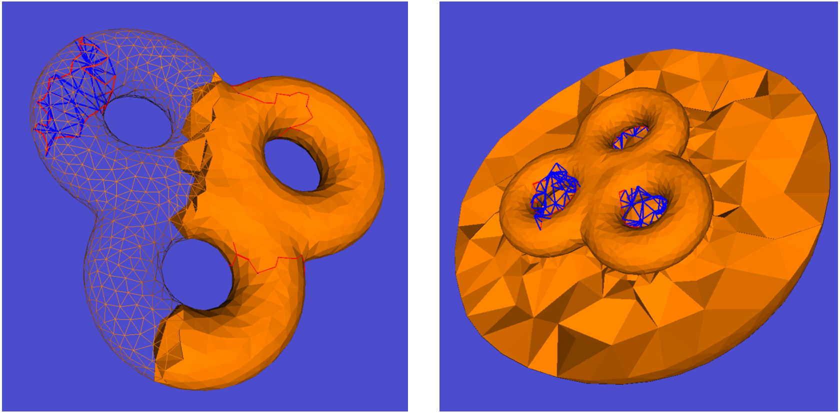


Figure: Interior and exterior volumes.