Guest lecture: Homology and its computation

Erin Wolf Chambers
erin.chambers@gmail.com

(with thanks to David Letscher, whose lecture on this inspired much of the content here)

Homology

- Homology is a coarser equivalence that homotopy, which you may have seen: any two loops that are homotopic are also homologous (but the reverse is not true).

- The definitions are a bit less intuitive, but the homology groups are easier to compute and still give a lot of topological information about the underlying space.

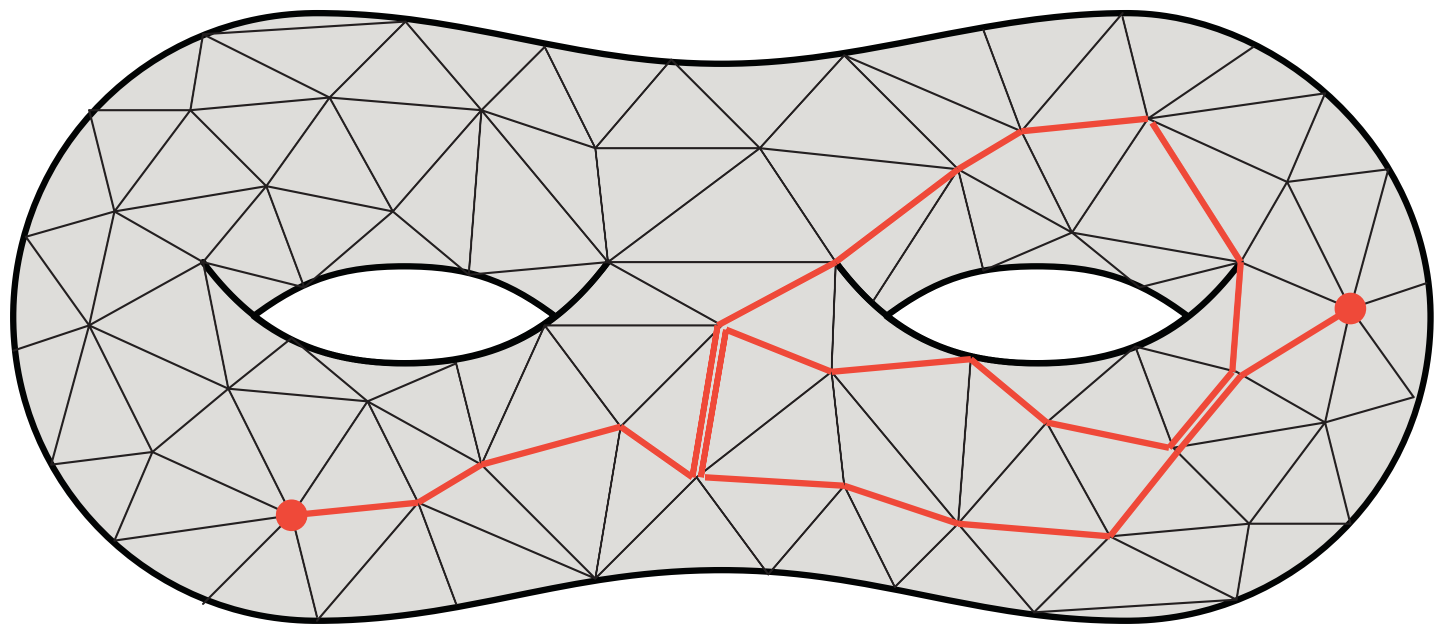
    - This is why they are so popular, and so useful.

k-chains

- Given any simplicial complex, we define the following:
  - The set of k-chains is just the vector space of linear combinations of k-simplices, with coefficients from some group or field.
  - Coefficients in this vector space can be from any group or field.
  - So k-chains are just formal sums of simplices, which form a linear algebra structure.
    - (Computer scientists are usually comfortable with this, since a 1-chain is just a walk in a graph.)

Example:

- Triangulation of a 2-manifold:
  - If we use F_2 coefficients on a 2-complex, we get sums of edges, where repeated edges cancel.
  - So the set of 1-chains with coefficients from F_2 on a surface are the same as the set of even subgraphs.



Boundary maps

- The boundary map δ_k on k-chains simply takes any simplex to the sum of the (k-1)-chains that bound it.

- If coefficients are from F_2 then there is no real idea of direction or orientation: an edge is either present or it is not.

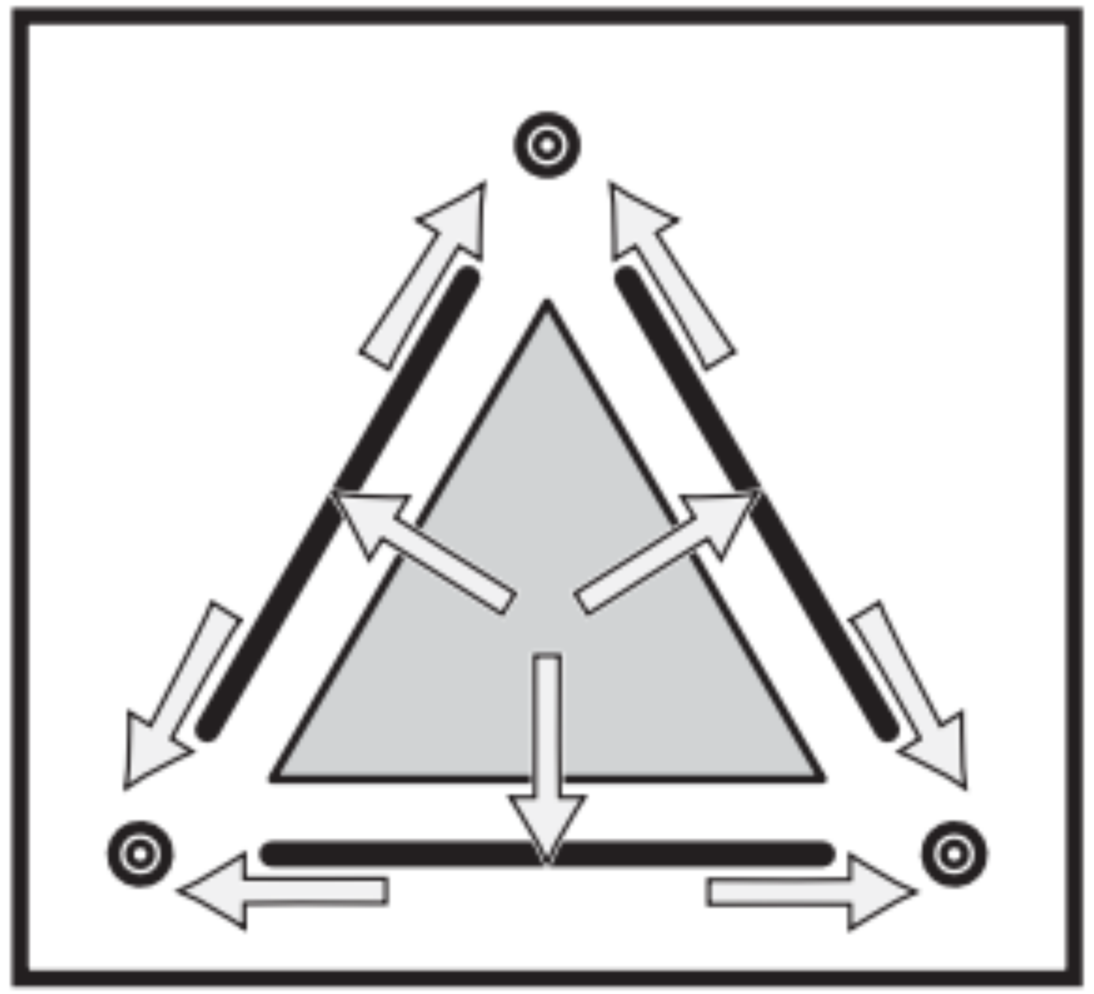
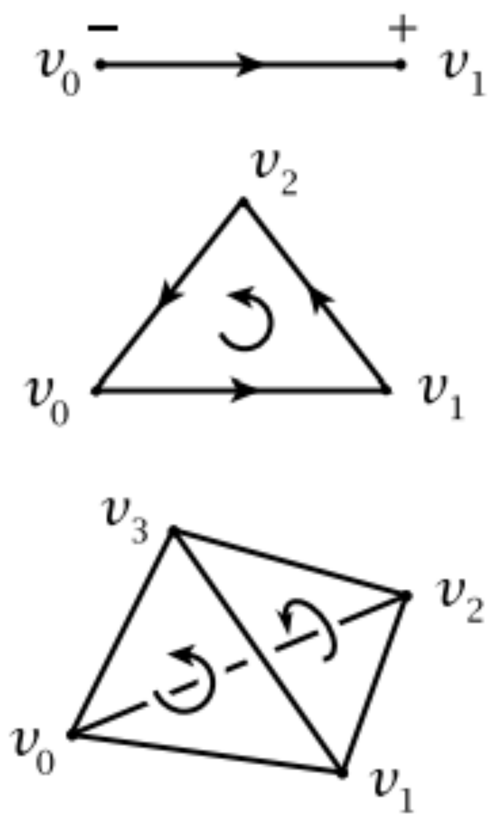


Image from
[Ghrist 2014]

Non-F_2 coefficients

- If linear combinations don't come from F_2 then things get a bit more complex:



$$\partial[v_0, v_1] = [v_1] - [v_0]$$

$$\partial[v_0, v_1, v_2] = [v_1, v_2] - [v_0, v_2] + [v_0, v_1]$$

$$\partial[v_0, v_1, v_2, v_3] = [v_1, v_2, v_3] - [v_0, v_2, v_3] + [v_0, v_1, v_3] - [v_0, v_1, v_2]$$

Image from Hatcher 2001

- Still - at its heart, this is just linear algebra! (Plus something like the right hand rule, if you remember physics.)

Boundary maps

- We can think of the boundary maps on the chain complexes in series: $C_d \rightarrow C_{d-1} \rightarrow \ldots \rightarrow C_1 \rightarrow C_0$
    - This is called the chain complex of M.
- These maps and how they treat the chain complexes are the key to homology.
    - Aside: It's worth noting here that the boundary of a boundary is empty.

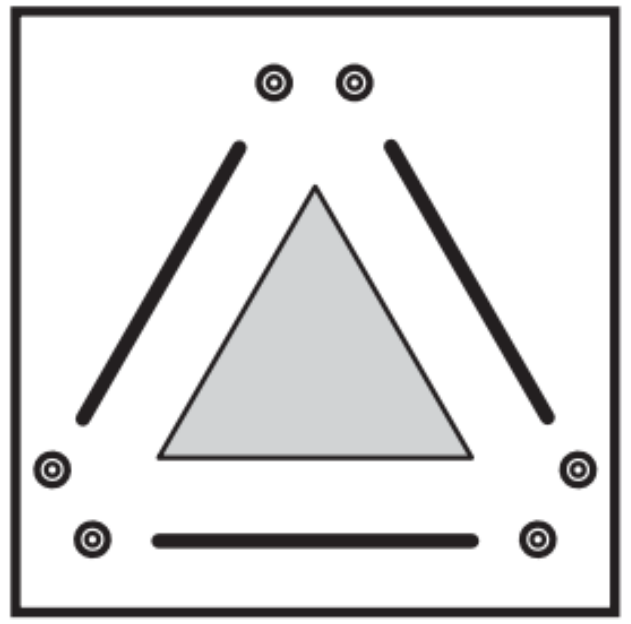


Image from [Ghrist 2014]

Boundaries and cycles

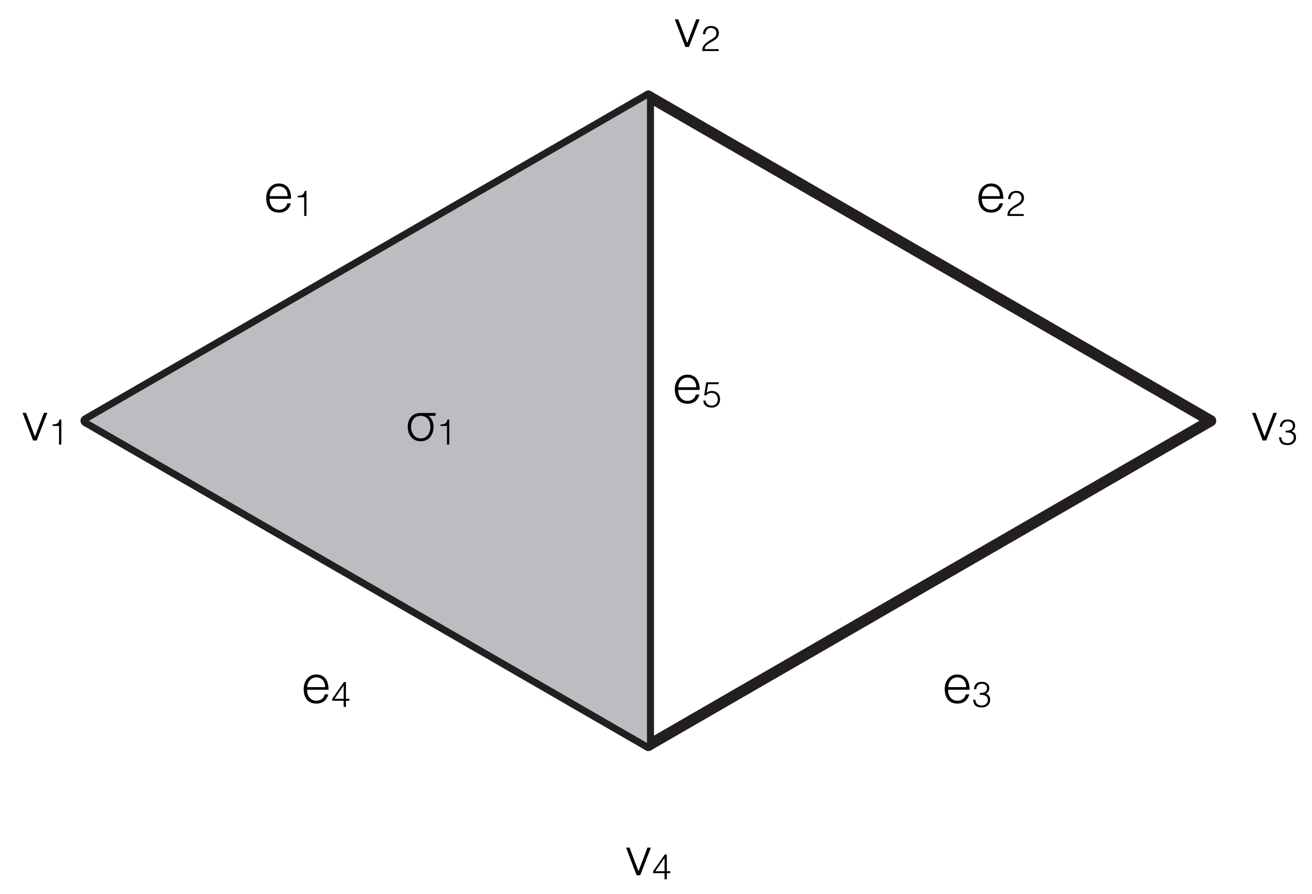
- The image of an element of C_{k-1} when the map δ_k: $C_k \rightarrow C_{k-1}$ is applied is called a boundary
  - B_k = im δ_{k+1} is called the boundary group - or $B_k(C,F_2)$
- A cycle is any element with empty boundary; these are in the kernel of δ_k since the are sent to nothing under the map
  - Think cycles in graph theory - these are precisely 1-chains that have no boundary
  - Similarly, a union of triangles that form a sphere is a 2-chain that has no boundary, but a disk would have boundary
  - Z_k = ker δ_k is the set of these cycles

Homology

- So: we have a chain complex:
    - $C_d \rightarrow C_{d-1} \rightarrow \ldots \rightarrow C_1 \rightarrow C_0$
- And two subgroups inside each C_k, Z_k = ker δ_k and B_k = im δ_{k+1}
- We say two chains in Z_k are homologous if they differ by a boundary
    - So $H_k = Z_k/B_k$
- Let's look at what this means…

A first example

- A very simple complex:

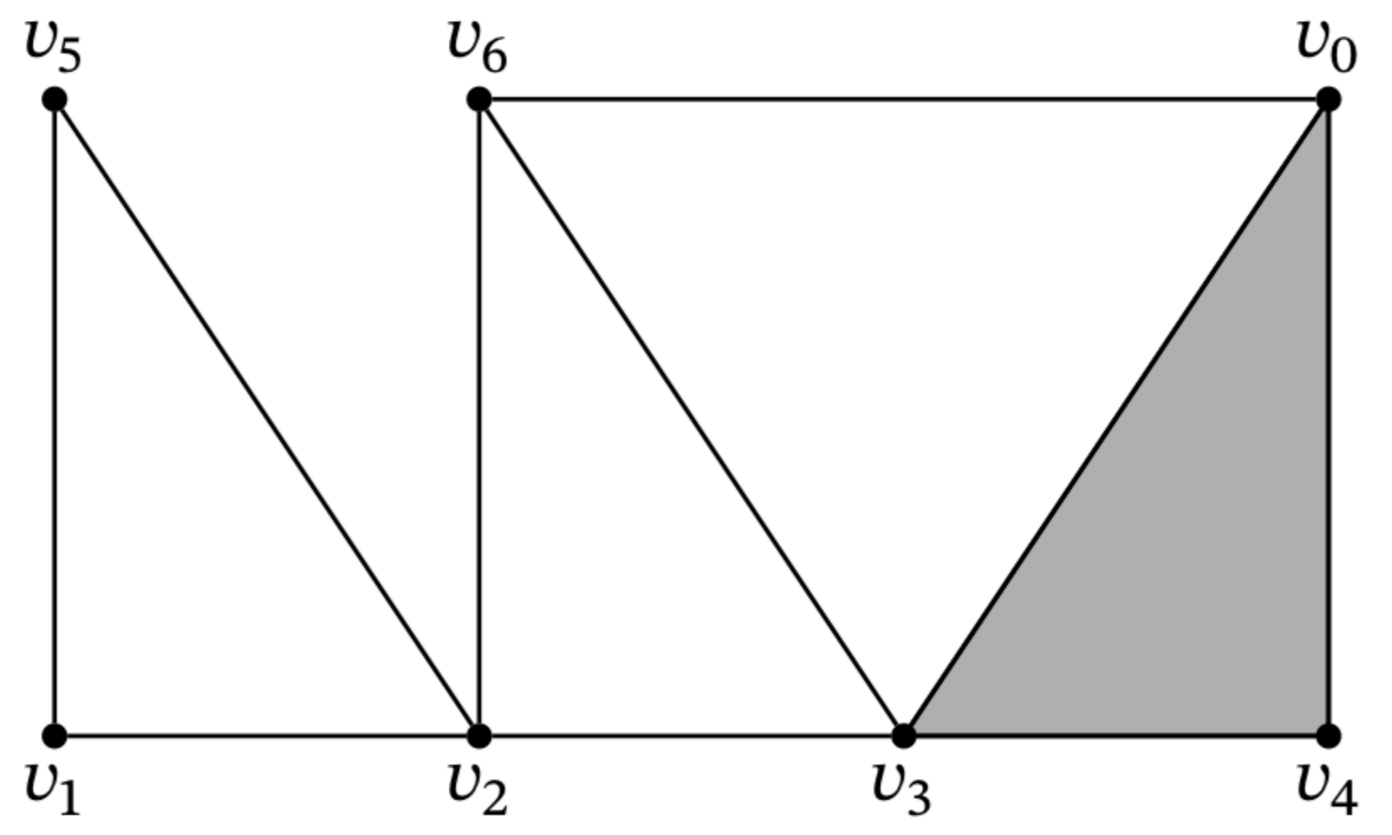


- Here, $e_1 + e_4 + e_5$ is in image δ_2, since it is the boundary of σ_1

- So $e_2 + e_3 + e_5$ is homologous to $e_1 + e_2 + e_3 + e_4$

Slightly more complex

- Another complex:



Example from
[Scoville 2019]

- Here:

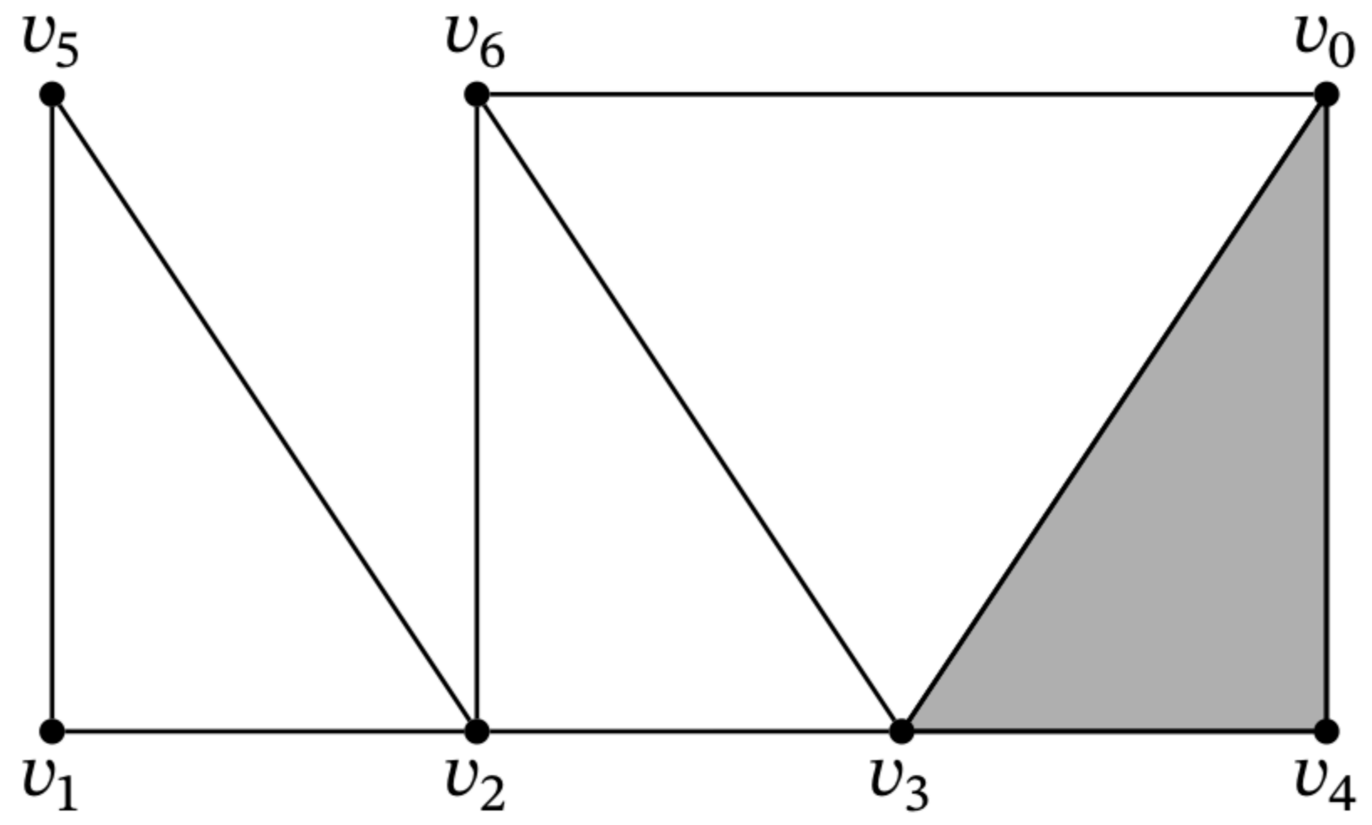
$$
\begin{aligned}
K &= \{v_0, v_1, v_2, v_3, v_4, v_5, v_6, v_0v_3, v_0v_4, v_0v_6, v_1v_2, v_2v_3, v_3v_4, v_1v_5, \\
&\quad v_2v_5, v_2v_6, v_3v_6, , v_0v_3v_4\}.
\end{aligned}
$$

- And:

$$
\begin{aligned}
\partial_1(v_{12}) &= v_2 + v_1, \\
\partial_1(v_{13}) &= v_3 + v_1, \\
\partial_1(v_{24}) &= v_4 + v_2, \\
\partial_1(v_{34}) &= v_4 + v_3, \\
\partial_1(v_{25}) &= v_5 + v_2, \\
\partial_1(v_{45}) &= v_5 + v_4, \\
\partial_1(v_{56}) &= v_6 + v_5,
\end{aligned}
$$

Boundary Matrices

- We often represent this boundary relationship in matrix form.



Example from [Scoville 2019]

- For example:

- We get: $$\partial_1 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\partial_2 = \begin{array}{c} \\ v_0v_3 \\ v_0v_4 \\ v_0v_6 \\ v_1v_2 \\ v_2v_3 \\ v_3v_4 \\ v_1v_5 \\ v_2v_5 \\ v_2v_6 \\ v_3v_6 \end{array} \begin{array}{c} v_0v_3v_4 \\ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \end{array}.$$

Back to homology

- So for chain complexes here, we have maps from $K_2 \rightarrow K_1 \rightarrow K_0$ given by

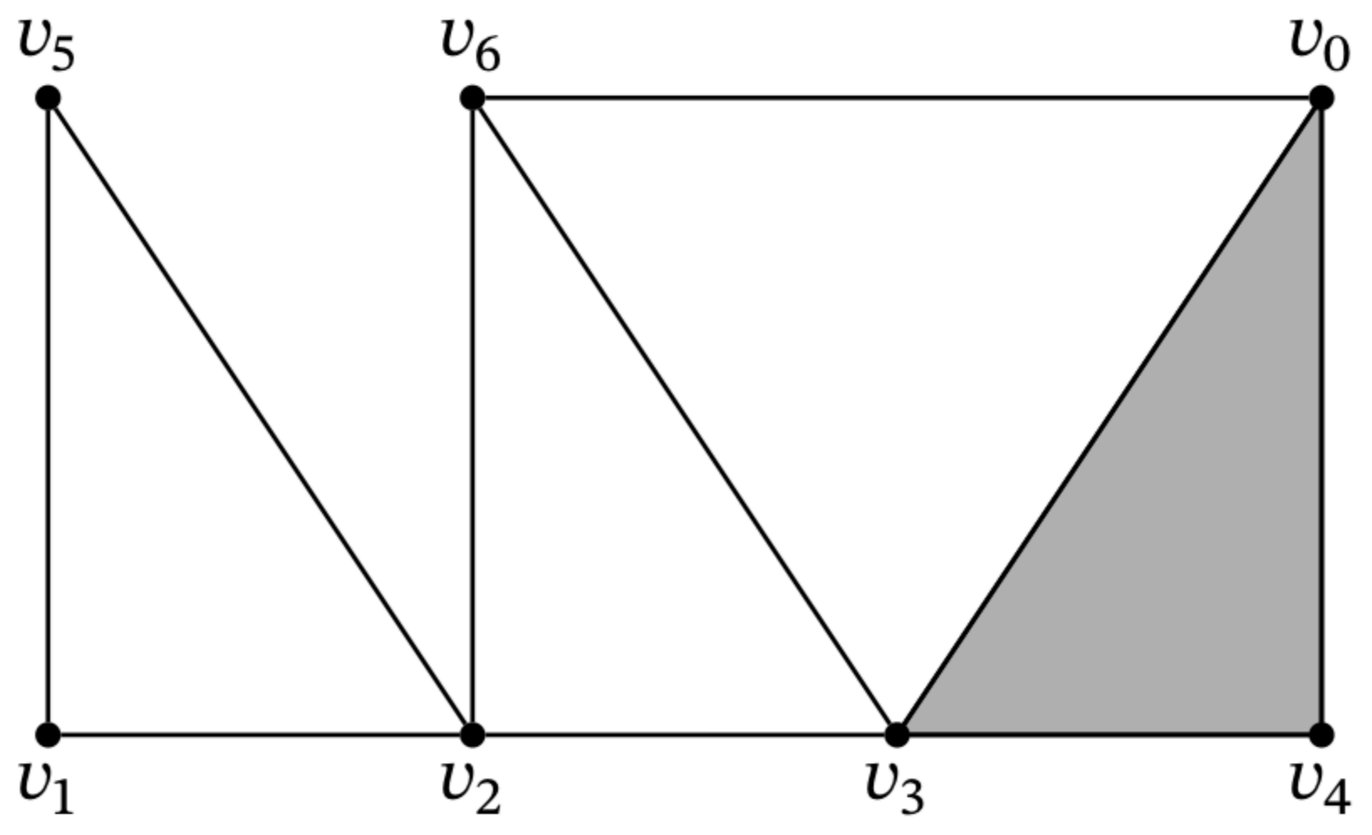
$$\partial_1 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\partial_2 = \begin{array}{c c} & v_0v_3v_4 \\ \begin{array}{c} v_0v_3 \\ v_0v_4 \\ v_0v_6 \\ v_1v_2 \\ v_2v_3 \\ v_3v_4 \\ v_1v_5 \\ v_2v_5 \\ v_2v_6 \\ v_3v_6 \end{array} & \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \end{array}.$$

- Recall that $H_k = Z_k/B_k$. Here that is only interesting for H_0, H_1 and H_2…

Homology example:

- Since $H_k = Z_k/B_k = \ker \delta_k / \mathrm{im}\ \delta_{k+1,}$ here:
  - $\mathrm{rank}(\delta_2) = 1$, $\mathrm{rank}(\delta_1) = 6$ (math!), $\mathrm{rank}(\delta_0) = 0$ (trivially)
  - $\mathrm{null}(\delta_2) = \dim(Z_k) = 0$, $\mathrm{null}(\delta_1) = 4$, $\mathrm{null}(\delta_0) = 7$ (trivially)
- Geometrically:



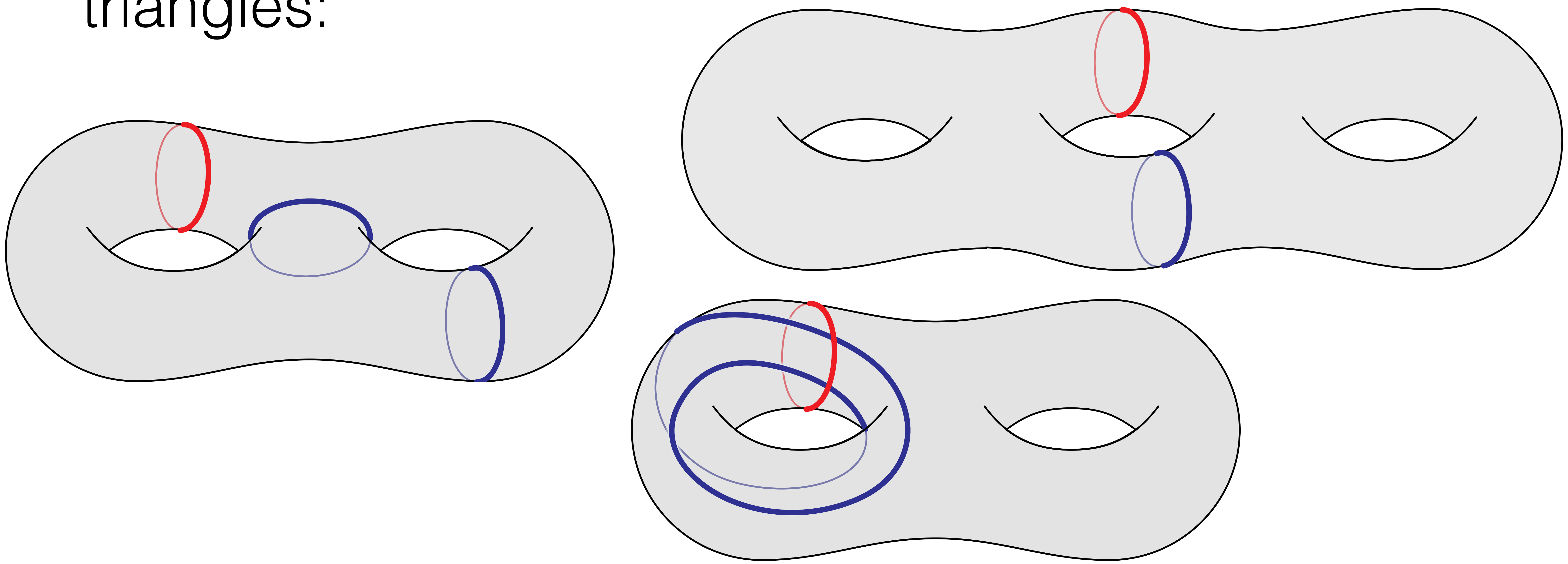
Another example of note: graphs

- Consider a graph G - a compact 1d cell complex.
  - $\delta_0 = 0$, so all vertices are 0-cycles: $Z_0 = C_0$
  - Boundaries B_0 are endpoints of (unions of) paths, so two vertices are homologous if they can be connected
  - Therefore, dimension of H_0 is the number of components of the graph.
  - For H_1, we note that there are no 2-cells, so there are no boundaries B_1, and Z_1 is just all possible cycles or unions of cycles
    - This means H_1 is the same as the cycle basis from graph theory.

Surfaces

- If we have a surface, then we still have triangles, edges, and vertices: $C_2 \to C_1 \to C_0$

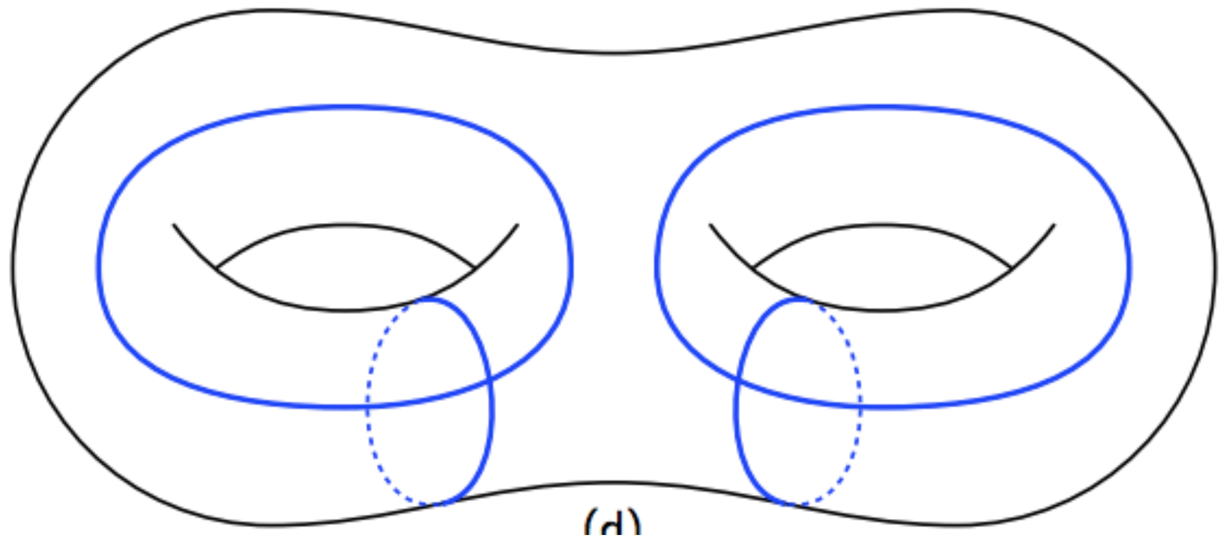
- Here, two cycles of edges in C_1 are homologous if they differ by a boundary of some 2d region built from triangles:



Surfaces continued

- In the end, we get nonzero homology only for H_i with $i<3$, since all cells are of dimension two or less:

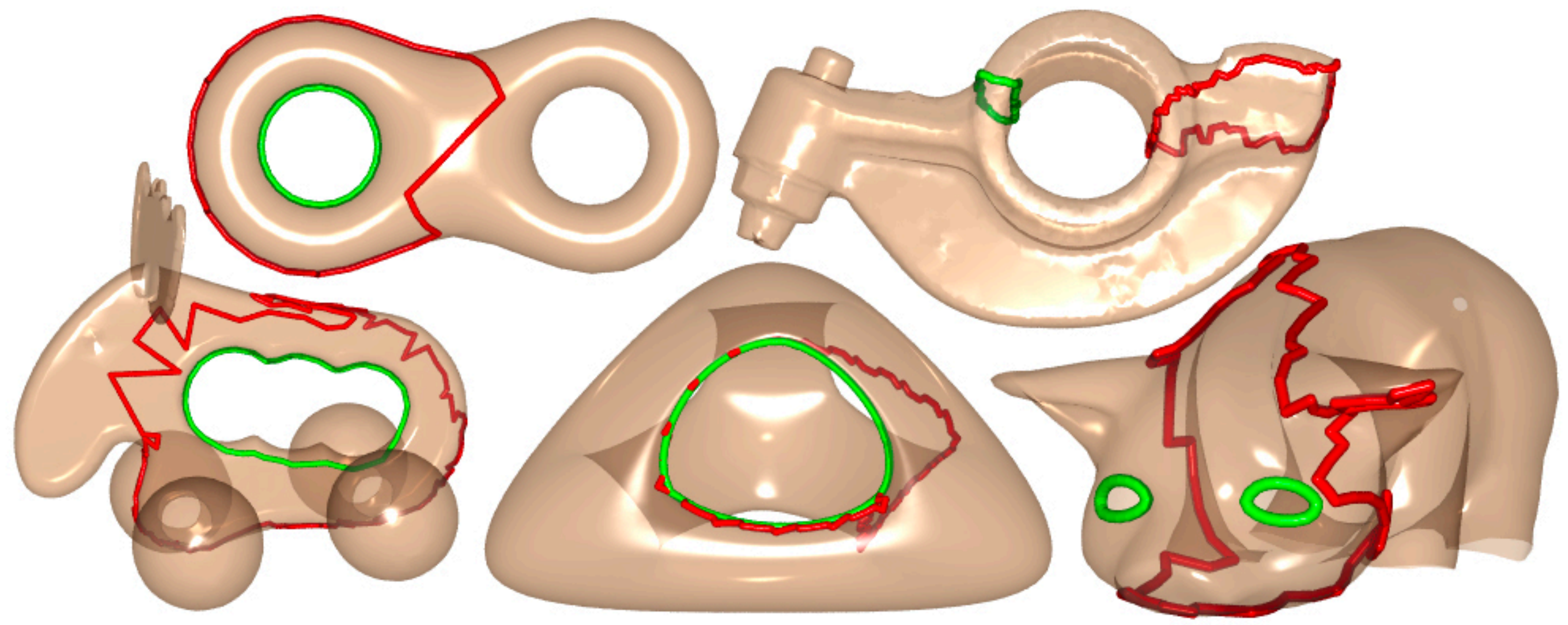
$$\dim H_k(S_g) = \begin{cases} 1 & : \quad k=0 \\ 2g & : \quad k=1 \\ 1 & : \quad k=2 \\ 0 & : \quad k>2 \end{cases}.$$



Erickson-Whittlesey 2005

- The 1 in dimension 2 is for the “inside” of the shape, the only void. The 2g is for 2 loops per handle.

Using homology



[Dey et al, 2010]

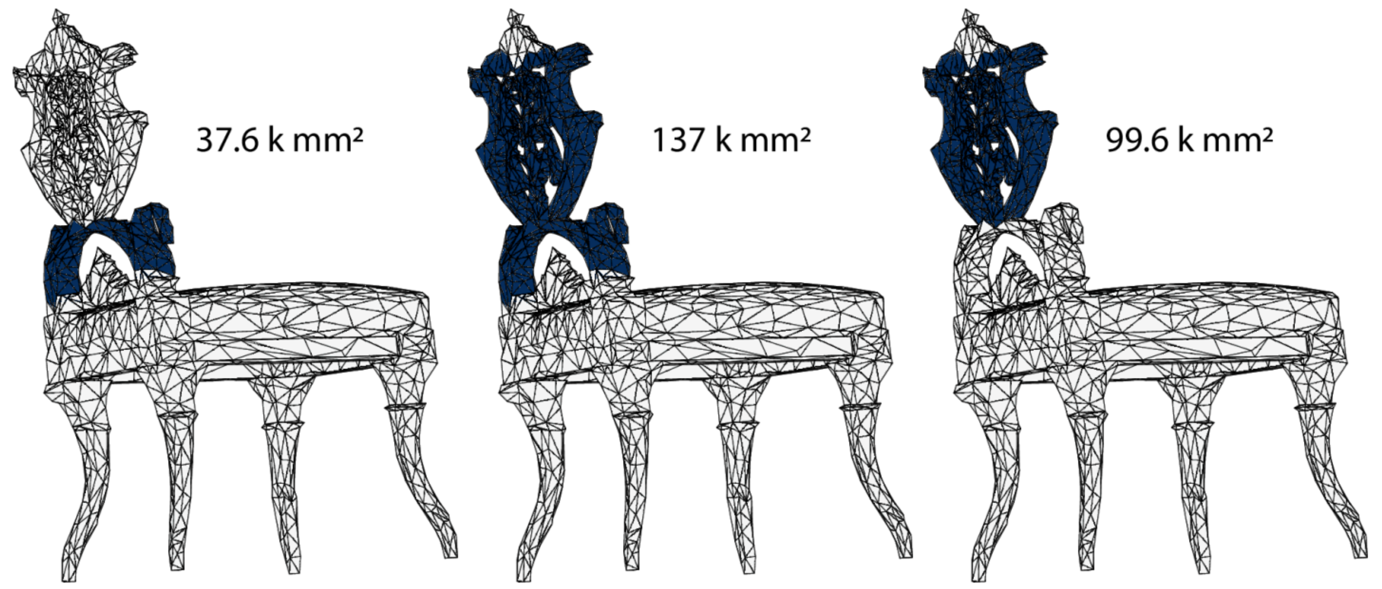


Figure 6: Minimum homology distances between cycles on a Smithsonian scanned model of a renaissance chair. Notice how these bounding chains exemplify the absence of connecting homotopies described in Figure 1.

[Chambers and Vejdemo-Johansson 2014]

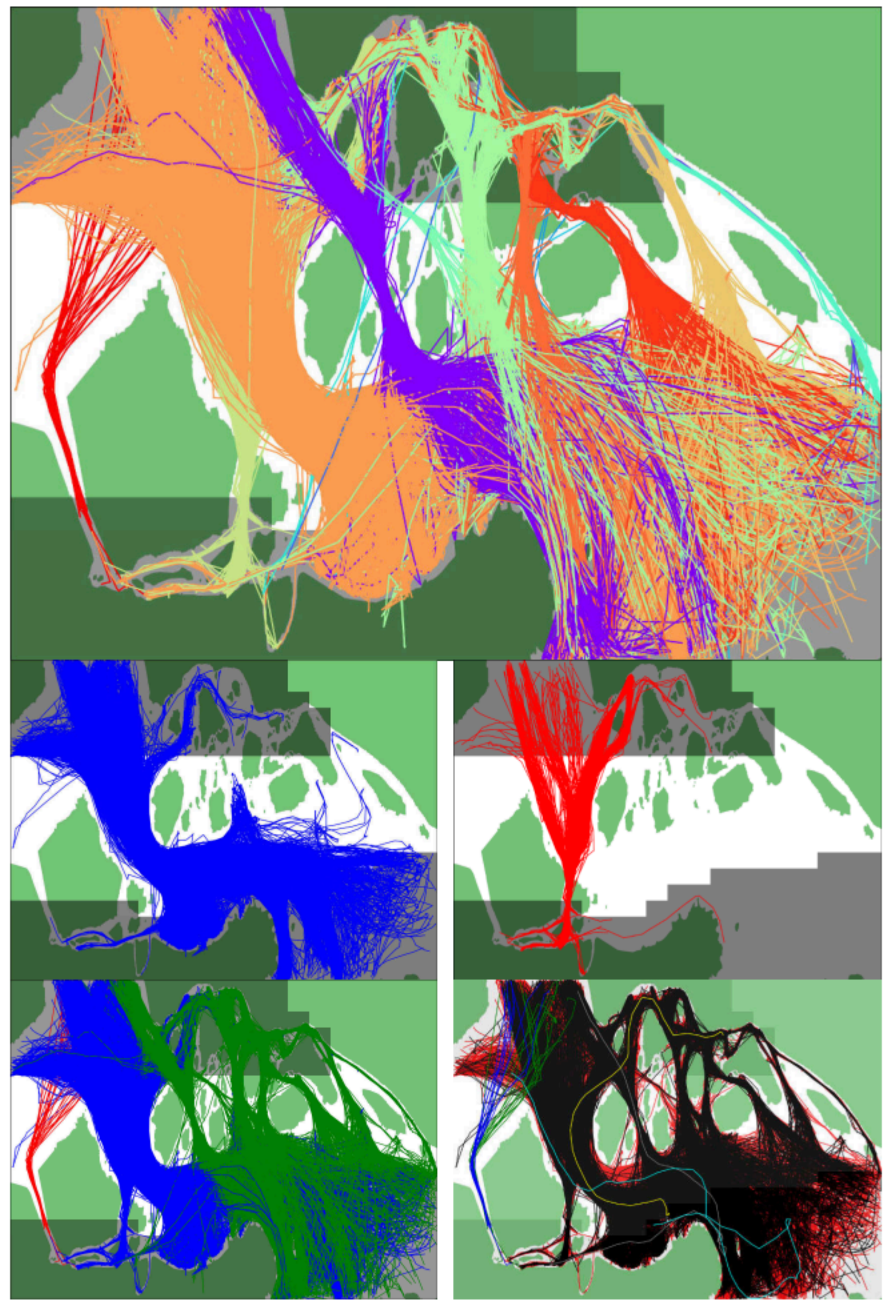


Fig. 4. The top figure displays all 15 found clusters at $r = 0.0002$, while the middle row illustrates two classes in isolation. Note how much variation these classes exhibit - yet they are distinguishable by the 'hard environment constraint' posed by the small southern island that they pass in a distinct manner. The bottom right displays a single linkage clustering by discrete Fréchet distance at distance 0.055, yielding 7 completely environment agnostic clusters. The bottom right figure shows the classification using our method at a higher filtration value of $r = 0.01$, where the smaller islands have been covered by simplices, resulting in only 3 trajectory classes at that filtration level.

[Pokorny et al, 2015]

Ranks of homology groups

- As noted earlier, more often than not, people are as interested in computing the ranks of the homology groups
  - The maximum number of linearly independent generators of the group
- Reason: While not complete invariants, they give a lot of information:
  - Rank of H_0 is the number of connected components
  - Rank of H_1 is the number of “handles” (or genus) in an orientable 2-manifold
  - Rank of H_3 captures the number of “voids” in a 3d-complex

Computing Betti numbers

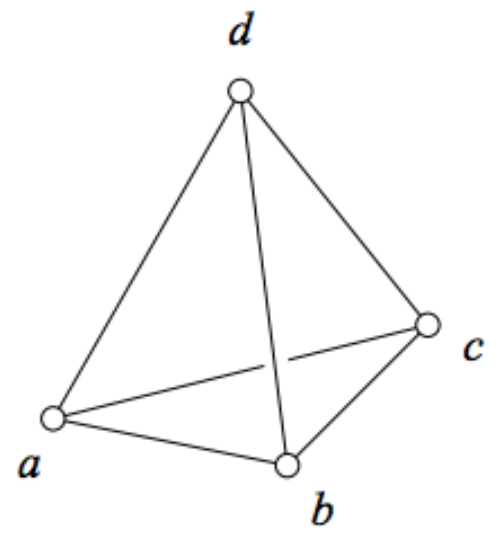
- The ranks of H_k are called Betti numbers:
- To calculate, we go back to the boundary operator in matrix form:

$$\mathsf{C}_p = \{\alpha_1, \alpha_2, \ldots, \alpha_{n_p}\}$$
$$\mathsf{C}_{p-1} = \{\tau_1, \tau_2, \ldots, \tau_{n_{p-1}}\}$$

$$M_p = \begin{array}{c} \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_{n_{p-1}} \end{array} \begin{array}{c} \begin{array}{cccc} \alpha_1 & \alpha_2 & \cdots & \alpha_{n_p} \end{array} \\ \begin{pmatrix} a_1^1 & a_1^2 & \cdots & a_1^{n_p} \\ a_2^1 & a_2^2 & \cdots & a_2^{n_p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_{p-1}}^1 & a_{n_{p-1}}^2 & \cdots & a_{n_{p-1}}^{n_p} \end{pmatrix} \end{array}$$

Example:

- Consider the 2 faces of a tetrahedra abcd:



$$M_2 = \begin{array}{c} \\ ab \\ ac \\ ad \\ bc \\ bd \\ cd \end{array} \begin{array}{c} \begin{array}{cccc} abc & abd & acd & bcd \end{array} \\ \begin{pmatrix} 1 & 1 & & \\ 1 & & 1 & \\ & 1 & 1 & \\ 1 & & & 1 \\ & 1 & & 1 \\ & & 1 & 1 \end{pmatrix} \end{array}$$

- The boundary matrix M_2 (for F_2 coefficients) just indicates which edges bound which triangles.

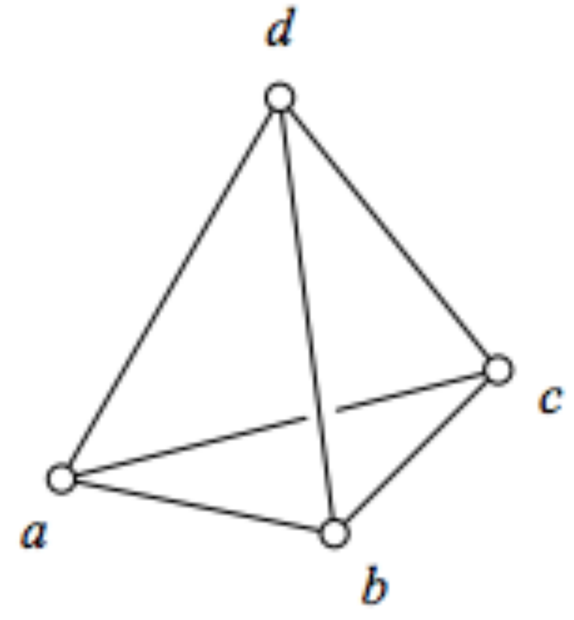
- If we want to find the boundary of a collection of 2-faces, we can multiply the vector of 2-faces by this matrix to find the edges that bound it.

- So these matrices turn δ_k into a simple linear map between vector spaces

Betti numbers

- The rows and columns of this matrix form a basis for C_{k-1} and C_k, respectively.
- Now, remember that $H_k = Z_k/B_k$ (kernel of δ_k mod image of δ_{k-1}).
  - By some fancy linear algebra (the rank nullity theorem - the rank and nullity add up to number of columns in a matrix), we also have that $C_k \simeq Z_k \oplus B_{k-1}$.
- We can reduce the matrix to Smith Normal Form, and separate the dimension of Z_k and B_{k-1} by which rows have a 1 in the diagonal.
  - So we can find the Betti number by doing this for two matrices and subtracting.

Back to our example:



	ab	ac	ad	bc	bd	cd
a	1	1	1			
b	1			1	1	
c		1		1		1
d			1		1	1

	ab	ab ac	ac ad	ab ac bc	ab ad bd	ac ad cd
b a	1					
c b		1				
d c			1			
d						

rank B_0 = 3
rank Z_1 = 3

	abc	abd	acd	bcd
ab	1	1		
ac	1		1	
ad		1	1	
bc	1			1
bd		1		1
cd			1	1

	abc	abc abd	abc abd acd	abc abd acd bcd
bc ac ab	1			
cd bc ad ac		1		
cd bd bc			1	
ad				
bd				
cd				

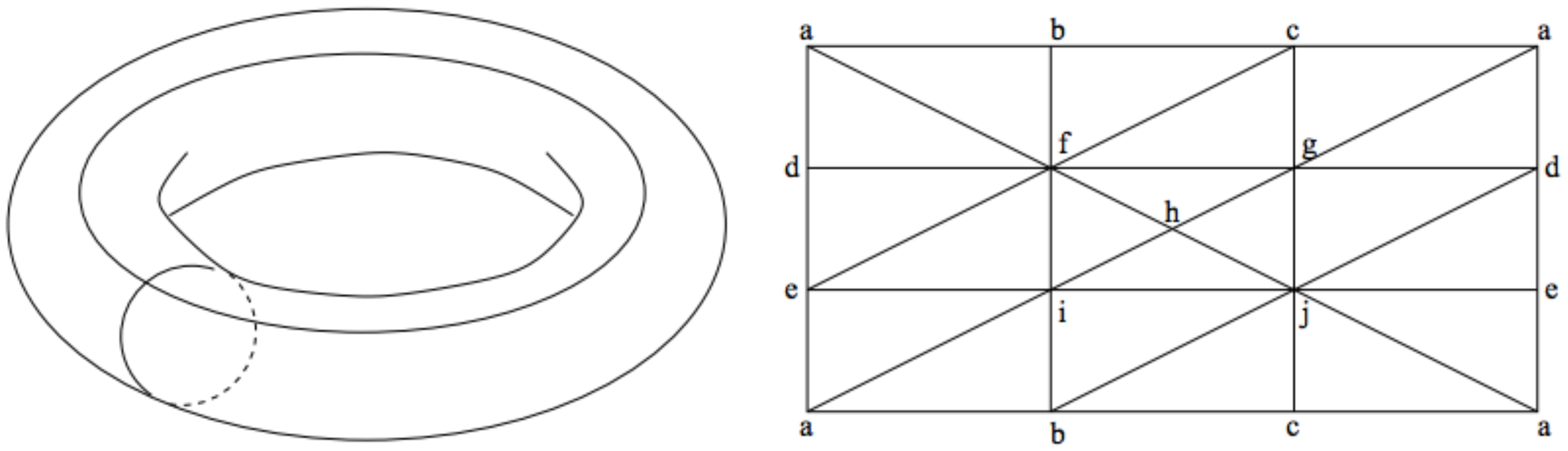
rank B_1 = 3
rank Z_2 = 1

End result:
rank H_0 =
#vertices - rank B_0
= 4-3 = 1
rank H_1 = rank Z_1 - rank B_1
= 3-3 = 0
rank H_2 = rank Z_2 - rank B_2
= 1-0 = 1

Another example

- Consider the torus:

Image from Vanessa Robin's thesis



- This triangulation will lead to a dim 2 boundary matrix that is 20 by 30, and a dim 1 that is 30 by 10.
    - However, it will simplify down by a LOT. Since we know rank H_1 is 2, rank Z_1 - rank B_1 must be 2!

End result

- Betti numbers (and homology more generally) are popular tools.

- Just import your favorite linear algebra library that can reduce matrices.

  - Or (better yet) use one of the many that already exist: mapper, dionysus, javaplex, CTL,…

- Again, homology is not as strong a classifier as homotopy, but it is much more computationally feasible.

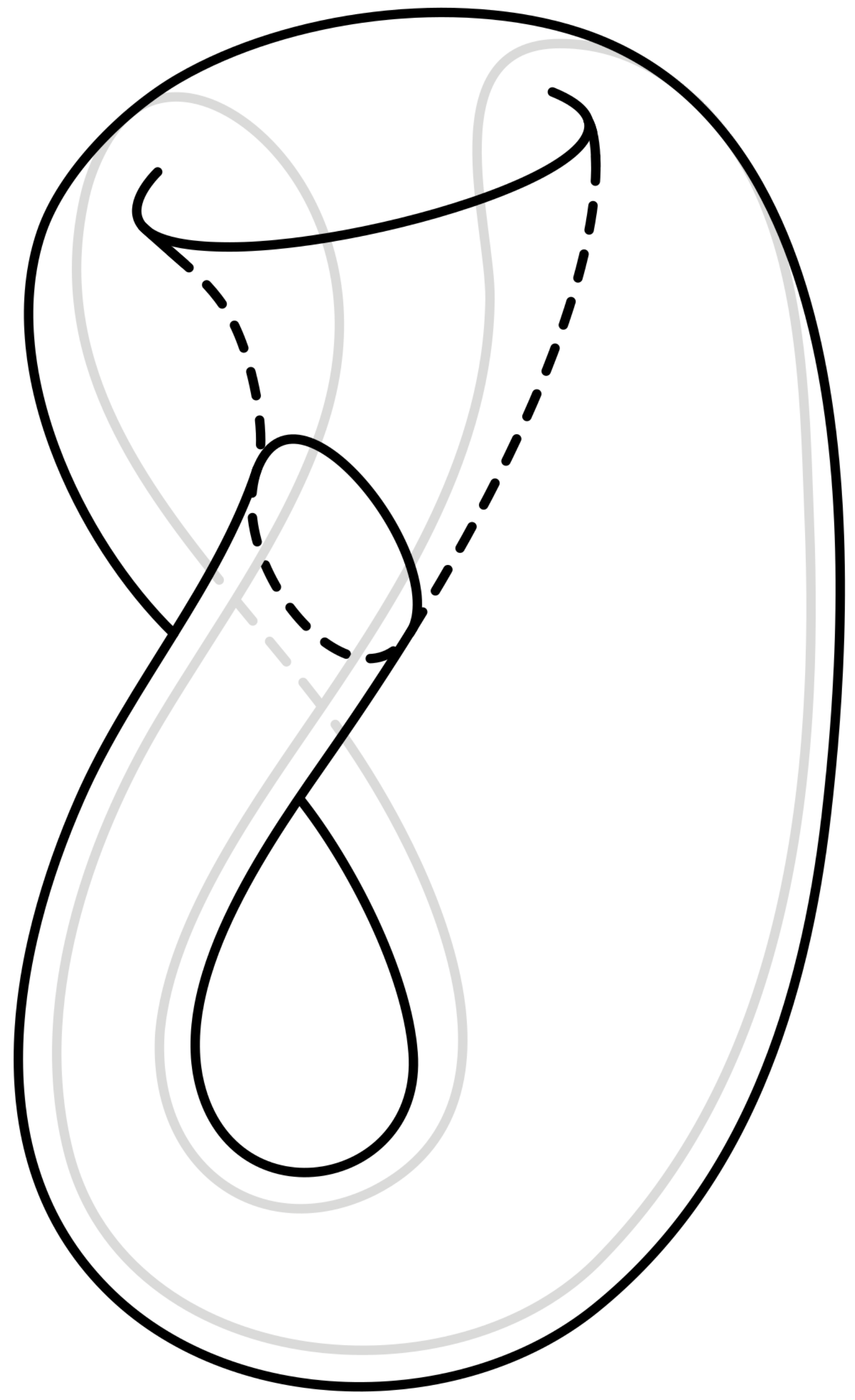
- This is part of where the idea of persistent homology came from.

Non-F_2 coefficients

- Note that I'm really using the fact that we are using 0,1 coefficients for most of these facts! Homology with different coefficients can behave very differently.

- It's also harder to compute if you're not in a field - Smith normal form computation can be prohibitive.

- Consider the Klein bottle:



$$H_k(K^2;\mathbb{F}_2) = \begin{cases} \mathbb{F}_2 & : \quad k=0 \\ \mathbb{F}_2 \oplus \mathbb{F}_2 & : \quad k=1 \\ \mathbb{F}_2 & : \quad k=2 \\ 0 & : \quad k>2 \end{cases} \quad ; \quad H_k(K^2;\mathbb{Z}) = \begin{cases} \mathbb{Z} & : \quad k=0 \\ \mathbb{Z} \oplus \mathbb{Z}_2 & : \quad k=1 \\ 0 & : \quad k>1 \end{cases}$$