Persistent Homology Tutorial

Xiaojin Zhu

Department of Computer Sciences
University of Wisconsin-Madison

jerryzhu@cs.wisc.edu

2013

Persistent homology

- A rapidly growing branch of topology

Persistent homology

- A rapidly growing branch of topology
- mathematically defines “holes” in data:

Persistent homology

- A rapidly growing branch of topology
- mathematically defines "holes" in data:
  - 0^{th} order holes: clusters

Persistent homology

- A rapidly growing branch of topology
- mathematically defines “holes” in data:
  - 0^{th} order holes: clusters
  - 1^{st} order holes: holes

Persistent homology

- A rapidly growing branch of topology
- mathematically defines “holes” in data:
  - 0^{th} order holes: clusters
  - 1^{st} order holes: holes
  - 2^{nd} order holes: voids

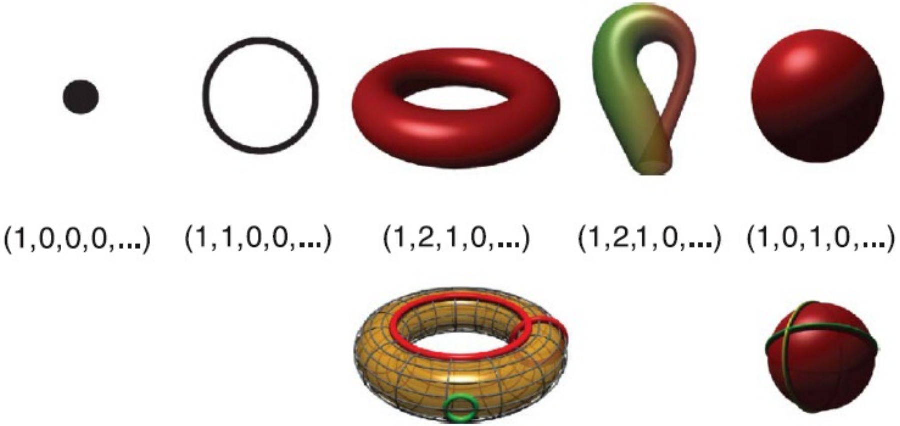
Persistent homology

- A rapidly growing branch of topology
- mathematically defines "holes" in data:
  - 0^{th} order holes: clusters
  - 1^{st} order holes: holes
  - 2^{nd} order holes: voids
  - higher order holes, too

Persistent homology

- A rapidly growing branch of topology
- mathematically defines "holes" in data:
  - 0^{th} order holes: clusters
  - 1^{st} order holes: holes
  - 2^{nd} order holes: voids
  - higher order holes, too
- Betti numbers: the number of k^{th} order holes

Betti number examples



[Reproduced from Singh *et al.* J. Vision 2008]

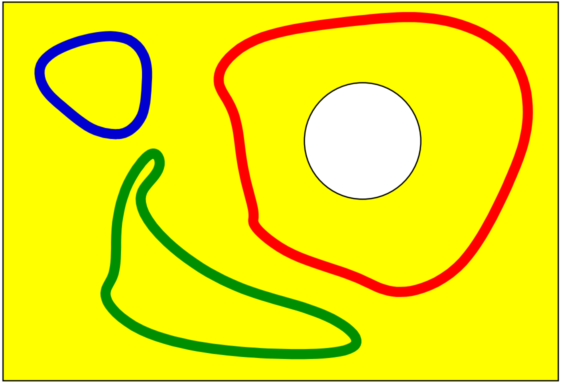
Plan of this talk

- Persistent homology tutorial

Plan of this talk

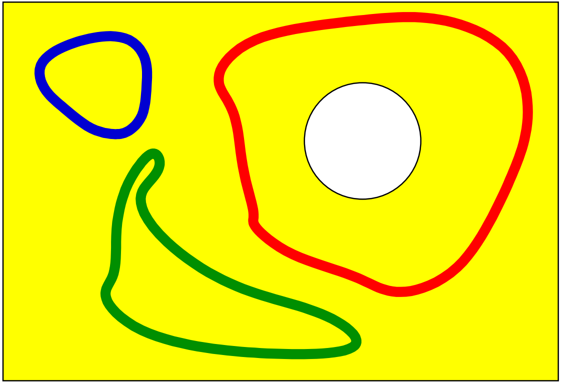
- Persistent homology tutorial
- An application in natural language processing

Holes and equivalent rubber bands



- blue $\sim$ green, not red

Holes and equivalent rubber bands



- blue $\sim$ green, not red
- two equivalent classes $\Leftrightarrow$ one hole.

Group Theory

Definition

A group $\langle G, * \rangle$ *is a set* G *with a binary operation* $*$ *such that*

Group Theory

Definition

A group $\langle G, * \rangle$ is a set G with a binary operation $*$ such that

1. *(associative)* $a * (b * c) = (a * b) * c$ *for all* $a, b, c \in G$.

Group Theory

Definition

A group $\langle G, * \rangle$ *is a set* G *with a binary operation* $*$ *such that*

1. *(associative)* $a * (b * c) = (a * b) * c$ *for all* $a, b, c \in G$.
2. *(identity)* $\exists e \in G$ *so that* $e * a = a * e = a$ *for all* $a \in G$.

Group Theory

Definition

A group $\langle G, * \rangle$ *is a set* G *with a binary operation* $*$ *such that*

1. *(associative)* $a * (b * c) = (a * b) * c$ *for all* $a, b, c \in G$.
2. *(identity)* $\exists e \in G$ *so that* $e * a = a * e = a$ *for all* $a \in G$.
3. *(inverse)* $\forall a \in G$, $\exists a' \in G$ *where* $a * a' = a' * a = e$.

Group Theory

Definition

A group $\langle G, * \rangle$ *is a set* G *with a binary operation* $*$ *such that*

1. *(associative)* $a * (b * c) = (a * b) * c$ *for all* $a, b, c \in G$.
2. *(identity)* $\exists e \in G$ *so that* $e * a = a * e = a$ *for all* $a \in G$.
3. *(inverse)* $\forall a \in G$, $\exists a' \in G$ *where* $a * a' = a' * a = e$.

- Examples: $\langle \mathbb{Z}, + \rangle$, $\langle \mathbb{R}, + \rangle$, $\langle \mathbb{R}_+, \times \rangle$, $\langle \mathbb{R} \backslash \{0\}, \times \rangle$.

Group Theory

Definition

A group $\langle G, * \rangle$ *is a set* G *with a binary operation* $*$ *such that*

1. *(associative)* $a * (b * c) = (a * b) * c$ *for all* $a, b, c \in G$.
2. *(identity)* $\exists e \in G$ *so that* $e * a = a * e = a$ *for all* $a \in G$.
3. *(inverse)* $\forall a \in G$, $\exists a' \in G$ *where* $a * a' = a' * a = e$.

- Examples: $\langle \mathbb{Z}, + \rangle$, $\langle \mathbb{R}, + \rangle$, $\langle \mathbb{R}_+, \times \rangle$, $\langle \mathbb{R} \backslash \{0\}, \times \rangle$.
- $\mathbb{Z}_2$

$+_2$	0	1
0	0	1
1	1	0

Group Theory

Definition

A *group* $\langle G, * \rangle$ *is a set* G *with a binary operation* $*$ *such that*

1. *(associative)* $a * (b * c) = (a * b) * c$ *for all* $a, b, c \in G$.
2. *(identity)* $\exists e \in G$ *so that* $e * a = a * e = a$ *for all* $a \in G$.
3. *(inverse)* $\forall a \in G$, $\exists a' \in G$ *where* $a * a' = a' * a = e$.

- Examples: $\langle \mathbb{Z}, + \rangle$, $\langle \mathbb{R}, + \rangle$, $\langle \mathbb{R}_+, \times \rangle$, $\langle \mathbb{R} \backslash \{0\}, \times \rangle$.
- $\mathbb{Z}_2$

$+_2$	0	1
0	0	1
1	1	0

- All our groups G are abelian: $\forall a, b \in G, a * b = b * a$.

Subgroup

Definition

A subset $H \subseteq G$ *of a group* $\langle G, * \rangle$ *is a* <u>*subgroup*</u> *of* G *if* $\langle H, * \rangle$ *is itself a group.*

Subgroup

Definition

A subset $H \subseteq G$ of a group $\langle G, * \rangle$ is a subgroup of G if $\langle H, * \rangle$ is itself a group.

- $\{e\}$ is the trivial subgroup of any group G

Subgroup

Definition

A subset $H \subseteq G$ of a group $\langle G, * \rangle$ is a *subgroup* *of G if $\langle H, * \rangle$ is itself a group.*

- $\{e\}$ is the trivial subgroup of any group G
- $\langle \mathbb{R}_+, \times \rangle$ is a subgroup of $\langle \mathbb{R} \backslash \{0\}, \times \rangle$

Subgroup

Definition

A subset $H \subseteq G$ of a group $\langle G, * \rangle$ is a subgroup of G if $\langle H, * \rangle$ is itself a group.

- $\{e\}$ is the trivial subgroup of any group G
- $\langle \mathbb{R}_+, \times \rangle$ is a subgroup of $\langle \mathbb{R} \backslash \{0\}, \times \rangle$
- not $\langle \mathbb{R}_-, \times \rangle$

Coset

Definition

Given a subgroup H of an abelian group G, for any $a \in G$, the set $a * H = \{a * h \mid h \in H\}$ is the coset of H represented by a.

Coset

Definition

Given a subgroup H of an abelian group G, for any $a \in G$, the set $a * H = \{a * h \mid h \in H\}$ is the coset of H represented by a.

- $H = \mathbb{R}_+$, $G = \mathbb{R} \backslash \{0\}$

Coset

Definition

Given a subgroup H of an abelian group G, for any $a \in G$, the set $a * H = \{a * h \mid h \in H\}$ is the coset of H represented by a.

- $H = \mathbb{R}_+$, $G = \mathbb{R} \backslash \{0\}$
- $3.14 \times \mathbb{R}_+$ is a coset which is the same as $\mathbb{R}_+$

Coset

Definition

Given a subgroup H of an abelian group G, for any $a \in G$, the set $a * H = \{a * h \mid h \in H\}$ is the coset of H represented by a.

- $H = \mathbb{R}_+$, $G = \mathbb{R} \backslash \{0\}$
- $3.14 \times \mathbb{R}_+$ is a coset which is the same as $\mathbb{R}_+$
- $-1 \times \mathbb{R}_+ = \mathbb{R}_-$ is another coset (not a subgroup)

Coset

Definition

Given a subgroup H of an abelian group G, for any $a \in G$, the set $a * H = \{a * h \mid h \in H\}$ is the coset of H represented by a.

- $H = \mathbb{R}_+$, $G = \mathbb{R} \backslash \{0\}$
- $3.14 \times \mathbb{R}_+$ is a coset which is the same as $\mathbb{R}_+$
- $-1 \times \mathbb{R}_+ = \mathbb{R}_-$ is another coset (not a subgroup)
- cosets have equal sizes and partition G.

Homomorphism

Definition

A map $\phi : G \mapsto G'$ *is a* *homomorphism* *if* $\phi(a * b) = \phi(a) \star \phi(b)$ *for* $\forall a, b \in G$.

Homomorphism

Definition

A map $\phi : G \mapsto G'$ *is a* *homomorphism* *if* $\phi(a * b) = \phi(a) \star \phi(b)$ *for* $\forall a, b \in G$.

- $\langle \mathbb{R}_+, \times \rangle$ to $\langle \mathbb{Z}_2, +_2 \rangle$: trivial homomorphism $\phi(a) = 0, \forall a \in \mathbb{R}_+$

Homomorphism

Definition

A map $\phi : G \mapsto G'$ is a homomorphism if $\phi(a * b) = \phi(a) \star \phi(b)$ for $\forall a, b \in G$.

- $\langle \mathbb{R}_+, \times \rangle$ to $\langle \mathbb{Z}_2, +_2 \rangle$: trivial homomorphism $\phi(a) = 0, \forall a \in \mathbb{R}_+$
- negation in natural language: G_N

$*$	$\sqcup$	not
$\sqcup$	$\sqcup$	not
not	not	$\sqcup$

homomorphism (isomorphism) from G_N to $\mathbb{Z}_2$: $\phi(\sqcup) = 0, \phi(\text{not}) = 1$.

Kernel

Definition

The kernel of a homomorphism $\phi : G \mapsto G'$ *is*
$\textbf{ker}\phi = \{a \in G \mid \phi(a) = e'\}$.

Kernel

Definition

The kernel of a homomorphism $\phi : G \mapsto G'$ *is* $\mathbf{ker}\phi = \{a \in G \mid \phi(a) = e'\}$.

- In the $\phi : G_N \mapsto \mathbb{Z}_2$ example, $\mathbf{ker}\phi = \{\sqcup\}$.

Kernel

Definition

The kernel of a homomorphism $\phi : G \mapsto G'$ *is*
$\mathbf{ker}\phi = \{a \in G \mid \phi(a) = e'\}$.

- In the $\phi : G_N \mapsto \mathbb{Z}_2$ example, $\mathbf{ker}\phi = \{\sqcup\}$.
- Another example: $\phi : \langle \mathbb{R}\backslash\{0\}, \times \rangle \mapsto G_N$ by $\phi(a) = \sqcup$ if $a > 0$ and "not" if $a < 0$. $\mathbf{ker}\phi = \mathbb{R}_+$

Kernel

Definition

The kernel of a homomorphism $\phi : G \mapsto G'$ *is*
$\mathbf{ker}\phi = \{a \in G \mid \phi(a) = e'\}$.

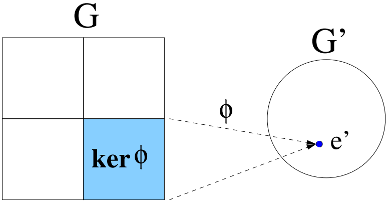
- In the $\phi : G_N \mapsto \mathbb{Z}_2$ example, $\mathbf{ker}\phi = \{\sqcup\}$.
- Another example: $\phi : \langle \mathbb{R} \backslash \{0\}, \times \rangle \mapsto G_N$ by $\phi(a) = \sqcup$ if $a > 0$ and "not" if $a < 0$. $\mathbf{ker}\phi = \mathbb{R}_+$
- For any homomorphism $\phi : G \mapsto G'$, $\mathbf{ker}\phi$ is a subgroup of G.

Kernel

Definition

The *kernel* *of a homomorphism* $\phi : G \mapsto G'$ *is*
$\mathbf{ker}\phi = \{a \in G \mid \phi(a) = e'\}$.

- In the $\phi : G_N \mapsto \mathbb{Z}_2$ example, $\mathbf{ker}\phi = \{\sqcup\}$.
- Another example: $\phi : \langle \mathbb{R}\backslash\{0\}, \times\rangle \mapsto G_N$ by $\phi(a) = \sqcup$ if $a > 0$ and "not" if $a < 0$. $\mathbf{ker}\phi = \mathbb{R}_+$
- For any homomorphism $\phi : G \mapsto G'$, $\mathbf{ker}\phi$ is a subgroup of G.
- Cosets $a * \mathbf{ker}\phi$ partition G



Quotient group

- Let $\langle H, * \rangle$ be a subgroup of an abelian group $\langle G, * \rangle$.

Quotient group

- Let $\langle H, * \rangle$ be a subgroup of an abelian group $\langle G, * \rangle$.
- A new operation on the cosets of H:
  $(a * H) \star (b * H) = (a * b) * H, \forall a, b \in G$.

Quotient group

- Let $\langle H, * \rangle$ be a subgroup of an abelian group $\langle G, * \rangle$.
- A new operation on the cosets of H:
  $(a * H) \star (b * H) = (a * b) * H, \forall a, b \in G$.

Definition

The cosets $\{a * H \mid a \in G\}$ under the operation $\star$ form a group, called the quotient group G/H.

Quotient group

- Let $\langle H, * \rangle$ be a subgroup of an abelian group $\langle G, * \rangle$.
- A new operation on the cosets of H:
  $(a * H) \star (b * H) = (a * b) * H, \forall a, b \in G$.

Definition

The cosets $\{a * H \mid a \in G\}$ under the operation $\star$ form a group, called the quotient group G/H.

- Example: $G = \mathbb{R}\backslash\{0\}$ and $\mathbf{ker}\phi = \mathbb{R}_+$, two cosets: $\mathbb{R}_+$ and $\mathbb{R}_-$.

Quotient group

- Let $\langle H, * \rangle$ be a subgroup of an abelian group $\langle G, * \rangle$.
- A new operation on the cosets of H:
  $(a * H) \star (b * H) = (a * b) * H, \forall a, b \in G$.

Definition

The cosets $\{a * H \mid a \in G\}$ under the operation $\star$ form a group, called the quotient group G/H.

- Example: $G = \mathbb{R} \backslash \{0\}$ and $\mathbf{ker}\phi = \mathbb{R}_+$, two cosets: $\mathbb{R}_+$ and $\mathbb{R}_-$.
- The quotient group $(\mathbb{R} \backslash \{0\})/\mathbb{R}_+$ has the two coset elements.

Quotient group

- Let $\langle H, * \rangle$ be a subgroup of an abelian group $\langle G, * \rangle$.
- A new operation on the cosets of H:
  $(a * H) \star (b * H) = (a * b) * H, \forall a, b \in G$.

Definition

The cosets $\{a * H \mid a \in G\}$ under the operation $\star$ form a group, called the <u>*quotient group*</u> G/H.

- Example: $G = \mathbb{R} \backslash \{0\}$ and $\mathbf{ker}\phi = \mathbb{R}_+$, two cosets: $\mathbb{R}_+$ and $\mathbb{R}_-$.
- The quotient group $(\mathbb{R} \backslash \{0\})/\mathbb{R}_+$ has the two coset elements.
- $R_- \star \mathbb{R}_- = (-1 \times R_+) \star (-1 \times R_+) = (-1 \times -1) \times R_+ = 1 \times R_+ = R_+$.

Quotient group

- Let $\langle H, * \rangle$ be a subgroup of an abelian group $\langle G, * \rangle$.
- A new operation on the cosets of H:
  $(a * H) \star (b * H) = (a * b) * H, \forall a, b \in G$.

Definition

The cosets $\{a * H \mid a \in G\}$ under the operation $\star$ form a group, called the quotient group G/H.

- Example: $G = \mathbb{R} \backslash \{0\}$ and $\mathbf{ker}\phi = \mathbb{R}_+$, two cosets: $\mathbb{R}_+$ and $\mathbb{R}_-$.
- The quotient group $(\mathbb{R} \backslash \{0\})/\mathbb{R}_+$ has the two coset elements.
- $R_- \star \mathbb{R}_- = (-1 \times R_+) \star (-1 \times R_+) = (-1 \times -1) \times R_+ = 1 \times R_+ = R_+$.
- This quotient group $(\mathbb{R} \backslash \{0\})/\mathbb{R}_+$ is isomorphic to $\mathbb{Z}_2$.

Rank

Definition

Let S be a subset of a group G. The subgroup generated by S, $\langle S \rangle$, is the subgroup of all elements of G that can expressed as the finite operation of elements in S and their inverses.

Rank

Definition

Let S be a subset of a group G. The subgroup generated by S, $\langle S \rangle$, is the subgroup of all elements of G that can expressed as the finite operation of elements in S and their inverses.

- $\mathbb{Z}$ is itself the subgroup generated by $\{1\}$

Rank

Definition

Let S be a subset of a group G. The subgroup generated by S, $\langle S \rangle$, is the subgroup of all elements of G that can expressed as the finite operation of elements in S and their inverses.

- $\mathbb{Z}$ is itself the subgroup generated by $\{1\}$
- Even integers is the subgroup generated by $\{2\}$.

Rank

Definition

Let S be a subset of a group G. The subgroup generated by S, $\langle S \rangle$, is the subgroup of all elements of G that can expressed as the finite operation of elements in S and their inverses.

- $\mathbb{Z}$ is itself the subgroup generated by $\{1\}$
- Even integers is the subgroup generated by $\{2\}$.

Definition

The rank of a group G is $\mathrm{rank}(G) = \min\{|S| \mid S \subseteq G, \langle S \rangle = G\}$.

Rank

Definition

Let S be a subset of a group G. The subgroup generated by S, $\langle S \rangle$, is the subgroup of all elements of G that can expressed as the finite operation of elements in S and their inverses.

- $\mathbb{Z}$ is itself the subgroup generated by $\{1\}$
- Even integers is the subgroup generated by $\{2\}$.

Definition

The rank of a group G is $\mathrm{rank}(G) = \min\{|S| \mid S \subseteq G, \langle S \rangle = G\}$.

- $\mathrm{rank}(G)$ is the size of the smallest subset that generates G.

Rank

Definition

Let S be a subset of a group G. The subgroup generated by S, $\langle S \rangle$, is the subgroup of all elements of G that can expressed as the finite operation of elements in S and their inverses.

- $\mathbb{Z}$ is itself the subgroup generated by $\{1\}$
- Even integers is the subgroup generated by $\{2\}$.

Definition

The rank of a group G is $\mathrm{rank}(G) = \min\{|S| \mid S \subseteq G, \langle S \rangle = G\}$.

- $\mathrm{rank}(G)$ is the size of the smallest subset that generates G.
- $\mathrm{rank}(\mathbb{Z}) = 1$ since $\mathbb{Z} = \langle \{1\} \rangle$.

Rank

Definition

Let S be a subset of a group G. The subgroup generated by S, $\langle S \rangle$, is the subgroup of all elements of G that can expressed as the finite operation of elements in S and their inverses.

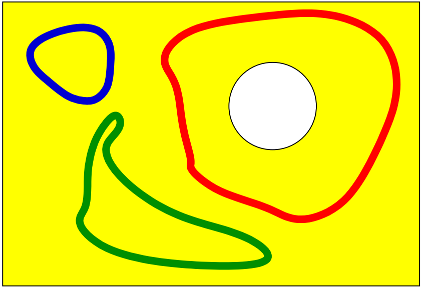
- $\mathbb{Z}$ is itself the subgroup generated by $\{1\}$
- Even integers is the subgroup generated by $\{2\}$.

Definition

The rank of a group G is $\mathrm{rank}(G) = \min\{|S| \mid S \subseteq G, \langle S \rangle = G\}$.

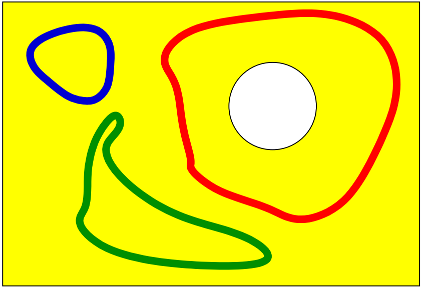
- $\mathrm{rank}(G)$ is the size of the smallest subset that generates G.
- $\mathrm{rank}(\mathbb{Z}) = 1$ since $\mathbb{Z} = \langle \{1\} \rangle$.
- $\mathrm{rank}(\mathbb{Z} \times \mathbb{Z}) = 2$ since $\mathbb{Z} \times \mathbb{Z} = \langle \{(0,1),(1,0)\} \rangle$.

The group of rubber bands



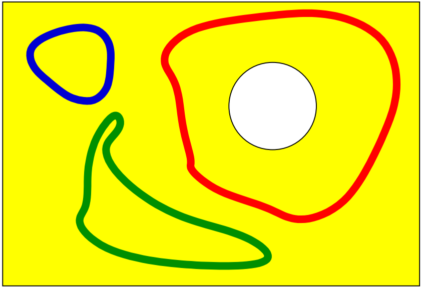
- To count "holes" in homology, consider the group of cycles (the rubber bands)

The group of rubber bands



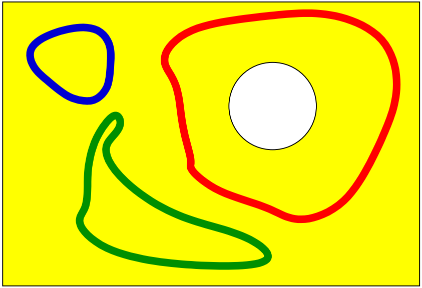
- To count “holes” in homology, consider the group of cycles (the rubber bands)
- The kernel: “uninteresting rubber bands” that do not surround holes

The group of rubber bands



- To count "holes" in homology, consider the group of cycles (the rubber bands)
- The kernel: "uninteresting rubber bands" that do not surround holes
- The quotient group "all rubber bands" / "uninteresting rubber bands" will identify holes.

The group of rubber bands



- To count "holes" in homology, consider the group of cycles (the rubber bands)
- The kernel: "uninteresting rubber bands" that do not surround holes
- The quotient group "all rubber bands" / "uninteresting rubber bands" will identify holes.
- Computation: need discrete rubber bands $\Rightarrow$ simplicial complex

Simplex

Definition

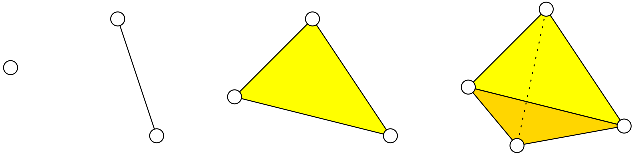
A p*-simplex* σ *is the convex hull of* $p+1$ *affinely independent points* $x_0, x_1, \ldots, x_p \in \mathbb{R}^d$*. We denote* $\sigma = \mathrm{conv}\{x_0, \ldots, x_p\}$*. The dimension of* σ *is* p*.*

Simplex

Definition

A <u>p-simplex</u> σ is the convex hull of $p+1$ affinely independent points $x_0, x_1, \ldots, x_p \in \mathbb{R}^d$. We denote $\sigma = \text{conv}\{x_0, \ldots, x_p\}$. The dimension of σ is p.

- $p = 0, 1, 2, 3$



Simplicial complex

Definition

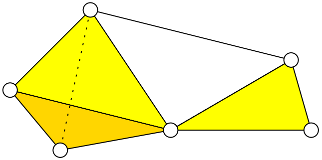
A simplicial complex K *is a finite collection of simplices such that* $\sigma \in K$ *and* τ *being a face of* σ *implies* $\tau \in K$*, and* $\sigma, \sigma' \in K$ *implies* $\sigma \cap \sigma'$ *is either empty or a face of both* σ *and* σ'*.*

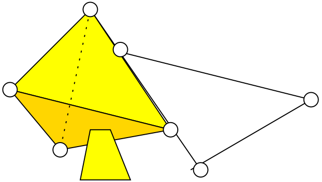
Simplicial complex

Definition

A <u>*simplicial complex*</u> K *is a finite collection of simplices such that* $\sigma \in K$ *and* τ *being a face of* σ *implies* $\tau \in K$*, and* $\sigma, \sigma' \in K$ *implies* $\sigma \cap \sigma'$ *is either empty or a face of both* σ *and* σ'*.*

- Properly aligned



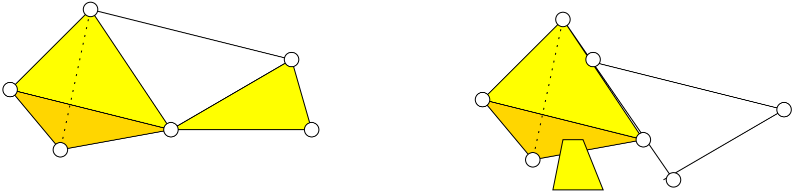


Simplicial complex

Definition

A simplicial complex K is a finite collection of simplices such that $\sigma \in K$ and τ being a face of σ implies $\tau \in K$, and $\sigma, \sigma' \in K$ implies $\sigma \cap \sigma'$ is either empty or a face of both σ and σ'.

- Properly aligned



- Simplicial complex = the yellow space in the rubber band picture

Chain

Definition

A p*-chain is a subset of* p*-simplices in a simplicial complex* K*.*

Chain

Definition

A p-chain is a subset of p-simplices in a simplicial complex K.

- Example: K=tetrahedron.

Chain

Definition

A $\underline{p\text{-}chain}$ *is a subset of* p*-simplices in a simplicial complex* K.

- Example: K=tetrahedron.
- A 2-chain is a subset of the four triangles.

Chain

Definition

A $\underline{p\textit{-chain}}$ *is a subset of* p*-simplices in a simplicial complex* K.

- Example: K=tetrahedron.
- A 2-chain is a subset of the four triangles.
- 2^4 distinct 2-chains.

Chain

Definition

A p-chain is a subset of p-simplices in a simplicial complex K.

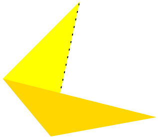
- Example: K=tetrahedron.
- A 2-chain is a subset of the four triangles.
- 2^4 distinct 2-chains.
- 2^6 distinct 1-chains (subsets of edges).

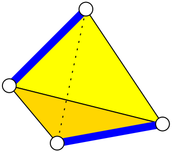
Chain

Definition

A p-chain is a subset of p-simplices in a simplicial complex K.

- Example: K=tetrahedron.
- A 2-chain is a subset of the four triangles.
- 2^4 distinct 2-chains.
- 2^6 distinct 1-chains (subsets of edges).
- Left: a 2-chain, right: a 1-chain



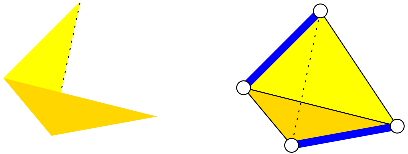


Chain

Definition

A p-chain is a subset of p-simplices in a simplicial complex K.

- Example: K=tetrahedron.
- A 2-chain is a subset of the four triangles.
- 2^4 distinct 2-chains.
- 2^6 distinct 1-chains (subsets of edges).
- Left: a 2-chain, right: a 1-chain



- A p-chain does not have to be connected.

Chain group

Definition

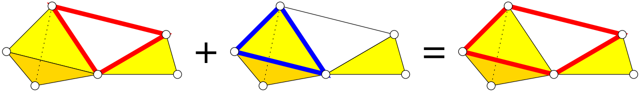
The set of p-chains of a simplicial complex K form a p-*chain group* C_p.

Chain group

Definition

The set of p-chains of a simplicial complex K form a p-chain group C_p.

- Mod-2 addition



Boundary

Definition

The *boundary* *of a* p*-simplex is the set of* $(p-1)$*-simplices faces.*

Boundary

Definition

The boundary of a p-simplex is the set of $(p-1)$-simplices faces.

- boundary of a tetrahedron = the four triangles faces

Boundary

Definition

The boundary of a p-simplex is the set of $(p-1)$-simplices faces.

- boundary of a tetrahedron = the four triangles faces
- boundary of a triangle = the three edges

Boundary

Definition

The <u>boundary</u> of a p-simplex is the set of $(p-1)$-simplices faces.

- boundary of a tetrahedron $=$ the four triangles faces
- boundary of a triangle $=$ the three edges
- boundary of an edge $=$ its two vertices

Boundary of a p-chain

Definition

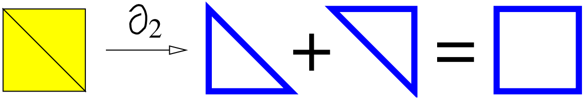
The boundary of a p-chain is the Mod-2 sum of the boundaries of its simplices. Taking the boundary is a group homomorphism ∂_p from C_p to C_{p-1}.

Boundary of a p-chain

Definition

The boundary of a p-chain is the Mod-2 sum of the boundaries of its simplices. Taking the boundary is a group homomorphism ∂_p from C_p to C_{p-1}.

- Faces shared by an even number of p-simplices in the chain will cancel out:



Cycles

Definition

A p-*cycle* c *is a* p*-chain with empty boundary:* $\partial_p c = 0$ *(the identity in* C_{p-1}*).*

Cycles

Definition

A p-cycle c is a p-chain with empty boundary: $\partial_p c = 0$ (the identity in C_{p-1}).

- Discrete p-dimensional "rubber bands"

Cycles

Definition

A p-cycle c is a p-chain with empty boundary: $\partial_p c = 0$ (the identity in C_{p-1}).

- Discrete p-dimensional "rubber bands"
- Left: a 1-cycle; Right: not a cycle

Cycles

Definition

A p-cycle c is a p-chain with empty boundary: $\partial_p c = 0$ (the identity in C_{p-1}).

- Discrete p-dimensional "rubber bands"
- Left: a 1-cycle; Right: not a cycle

- Z_p= all p-cycles (all rubber bands)

Cycles

Definition

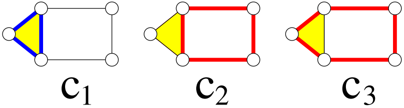
A p-cycle c is a p-chain with empty boundary: $\partial_p c = 0$ (the identity in C_{p-1}).

- Discrete p-dimensional “rubber bands”
- Left: a 1-cycle; Right: not a cycle

- Z_p= all p-cycles (all rubber bands)
- $\partial_p Z_p = 0$: Z_p is the kernel $\mathbf{ker}\partial_p$ and a subgroup of C_p.

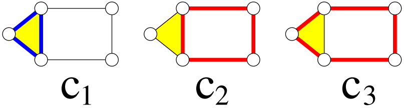
Boundary-Cycle

- The boundary of any $(p+1)$-chain is always a p-cycles



Boundary-Cycle

- The boundary of any $(p+1)$-chain is always a p-cycles

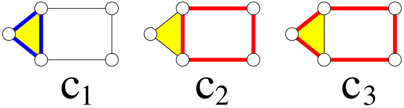


Definition

A p-boundary-cycle is a p-cycle that is also the boundary of some $(p+1)$-chain.

Boundary-Cycle

- The boundary of any $(p+1)$-chain is always a p-cycles



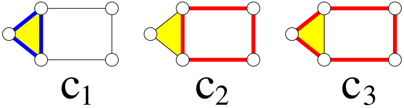
Definition

A p-boundary-cycle is a p-cycle that is also the boundary of some $(p+1)$-chain.

- Let $B_p = \partial_{p+1} C_{p+1}$, the p-boundary-cycles.

Boundary-Cycle

- The boundary of any $(p+1)$-chain is always a p-cycles



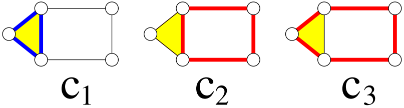
Definition

A p-boundary-cycle is a p-cycle that is also the boundary of some $(p+1)$-chain.

- Let $B_p = \partial_{p+1} C_{p+1}$, the p-boundary-cycles.
- B_p are the uninteresting rubber bands (e.g., $B_1 = \{0, c_1\}$)

Boundary-Cycle

- The boundary of any $(p+1)$-chain is always a p-cycles

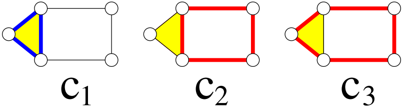


Definition

A p-boundary-cycle is a p-cycle that is also the boundary of some $(p+1)$-chain.

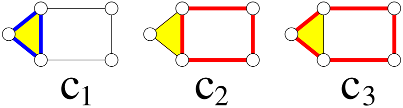
- Let $B_p = \partial_{p+1} C_{p+1}$, the p-boundary-cycles.
- B_p are the uninteresting rubber bands (e.g., $B_1 = \{0, c_1\}$)
- B_p is a subgroup of Z_p (all rubber bands).

Interesting rubber bands



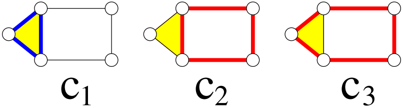
- c_2 and c_3 in Z_1 but not in B_1

Interesting rubber bands



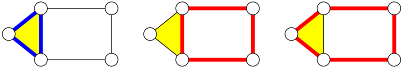
- c_2 and c_3 in Z_1 but not in B_1
- We can drag rubber band c_2 over the yellow triangle to make c_3

Interesting rubber bands



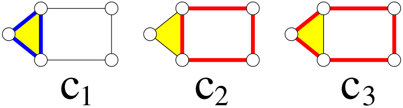
- c_2 and c_3 in Z_1 but not in B_1
- We can drag rubber band c_2 over the yellow triangle to make c_3
- Formally, $c_3 = c_2 + c_1$.

Interesting rubber bands



- c_2 and c_3 in Z_1 but not in B_1
- We can drag rubber band c_2 over the yellow triangle to make c_3
- Formally, $c_3 = c_2 + c_1$.
- c_2 and c_3 are equivalent in the hole they surround.

Interesting rubber bands



- c_2 and c_3 in Z_1 but not in B_1
- We can drag rubber band c_2 over the yellow triangle to make c_3
- Formally, $c_3 = c_2 + c_1$.
- c_2 and c_3 are equivalent in the hole they surround.
- The equivalence class: $c + B_p$

Homology group

Definition

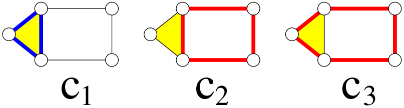
The p-th homology group is the quotient group $H_p = Z_p/B_p$.

Homology group

Definition

The p-th homology group is the quotient group $H_p = Z_p / B_p$.

- Example:

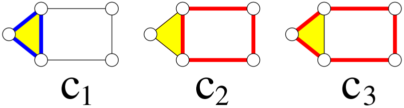


Homology group

Definition

The p-th homology group is the quotient group $H_p = Z_p / B_p$.

- Example:



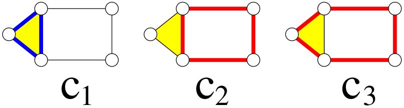
- All the 1-cycles : $Z_1 = \{0, c_1, c_2, c_3\}$.

Homology group

Definition

The p-th homology group is the quotient group $H_p = Z_p/B_p$.

- Example:



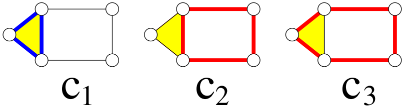
- All the 1-cycles : $Z_1 = \{0, c_1, c_2, c_3\}$.
- The uninteresting 1-cycles: $B_1 = \{0, c_1\}$, a subgroup of Z_1.

Homology group

Definition

The p-th homology group is the quotient group $H_p = Z_p/B_p$.

- Example:



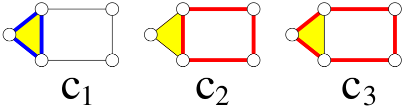
- All the 1-cycles : $Z_1 = \{0, c_1, c_2, c_3\}$.
- The uninteresting 1-cycles: $B_1 = \{0, c_1\}$, a subgroup of Z_1.
- The interesting 1-cycles: $c_2 + B_1 = c_3 + B_1 = \{c2, c3\}$

Homology group

Definition

The p-th homology group is the quotient group $H_p = Z_p/B_p$.

- Example:



- All the 1-cycles : $Z_1 = \{0, c_1, c_2, c_3\}$.
- The uninteresting 1-cycles: $B_1 = \{0, c_1\}$, a subgroup of Z_1.
- The interesting 1-cycles: $c_2 + B_1 = c_3 + B_1 = \{c2, c3\}$
- The homology group $H_1 = Z_1/B_1$ isomorphic to $\mathbb{Z}_2$

Betti number

Definition

The p-th Betti number is the rank of the homology group: $\beta_p = \text{rank}(H_p)$.

Betti number

Definition

The p-th Betti number is the rank of the homology group: $\beta_p = \mathrm{rank}(H_p)$.

- In our example, $\beta_1 = \mathrm{rank}(\mathbb{Z}_2) = 1$ (one 1st-order hole)

Betti number

Definition

The p-th Betti number is the rank of the homology group: $\beta_p = \mathrm{rank}(H_p)$.

- In our example, $\beta_1 = \mathrm{rank}(\mathbb{Z}_2) = 1$ (one 1st-order hole)
- β_p is the number of independent p-th holes.

Betti number

Definition

The p-th Betti number is the rank of the homology group: $\beta_p = \mathrm{rank}(H_p)$.

- In our example, $\beta_1 = \mathrm{rank}(\mathbb{Z}_2) = 1$ (one 1st-order hole)
- β_p is the number of independent p-th holes.
- A tetrahedron has $\beta_0 = 1$ (connected), $\beta_1 = \beta_2 = 0$ (no holes or voids)

Betti number

Definition

The p-th Betti number is the rank of the homology group: $\beta_p = \text{rank}(H_p)$.

- In our example, $\beta_1 = \text{rank}(\mathbb{Z}_2) = 1$ (one 1st-order hole)
- β_p is the number of independent p-th holes.
- A tetrahedron has $\beta_0 = 1$ (connected), $\beta_1 = \beta_2 = 0$ (no holes or voids)
- A hollow tetrahedron has $\beta_0 = 1, \beta_1 = 0, \beta_2 = 1$

Betti number

Definition

The p-th Betti number is the rank of the homology group: $\beta_p = \text{rank}(H_p)$.

- In our example, $\beta_1 = \text{rank}(\mathbb{Z}_2) = 1$ (one 1st-order hole)
- β_p is the number of independent p-th holes.
- A tetrahedron has $\beta_0 = 1$ (connected), $\beta_1 = \beta_2 = 0$ (no holes or voids)
- A hollow tetrahedron has $\beta_0 = 1, \beta_1 = 0, \beta_2 = 1$
- Removing the four triangle faces, the edge skeleton has $\beta_0 = 1$, $\beta_1 = 3$ (one is the sum of the other three), $\beta_2 = 0$ (no more void).

Betti number

Definition

The p-th Betti number is the rank of the homology group: $\beta_p = \text{rank}(H_p)$.

- In our example, $\beta_1 = \text{rank}(\mathbb{Z}_2) = 1$ (one 1st-order hole)
- β_p is the number of independent p-th holes.
- A tetrahedron has $\beta_0 = 1$ (connected), $\beta_1 = \beta_2 = 0$ (no holes or voids)
- A hollow tetrahedron has $\beta_0 = 1, \beta_1 = 0, \beta_2 = 1$
- Removing the four triangle faces, the edge skeleton has $\beta_0 = 1$, $\beta_1 = 3$ (one is the sum of the other three), $\beta_2 = 0$ (no more void).
- Removing the edges, $\beta_0 = 4$ (4 vertices) and $\beta_1 = \beta_2 = 0$.

From data to simplicial complex

- Given data $x_1, \ldots, x_n \in \mathbb{R}^d$.

From data to simplicial complex

- Given data $x_1, \ldots, x_n \in \mathbb{R}^d$.
- If any subset of $p+1$ points are within diameter ϵ, we add a p-simplex generated by those points.

From data to simplicial complex

- Given data $x_1, \ldots, x_n \in \mathbb{R}^d$.
- If any subset of $p+1$ points are within diameter ϵ, we add a p-simplex generated by those points.

Definition

A Vietoris-Rips complex of diameter ϵ is the simplicial complex $VR(\epsilon) = \{\sigma \mid \mathrm{diam}(\sigma) \leq \epsilon\}$.

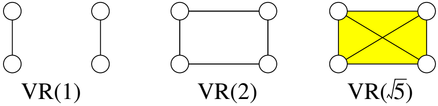
From data to simplicial complex

- Given data $x_1, \ldots, x_n \in \mathbb{R}^d$.
- If any subset of $p+1$ points are within diameter ϵ, we add a p-simplex generated by those points.

Definition

A Vietoris-Rips complex of diameter ϵ is the simplicial complex $VR(\epsilon) = \{\sigma \mid \mathrm{diam}(\sigma) \le \epsilon\}$.

- Example



Filtration

- Which ϵ should we pick?

Filtration

- Which ϵ should we pick?
- Don't pick – look at all ϵ's

Filtration

- Which ϵ should we pick?
- Don't pick – look at all ϵ's

Definition

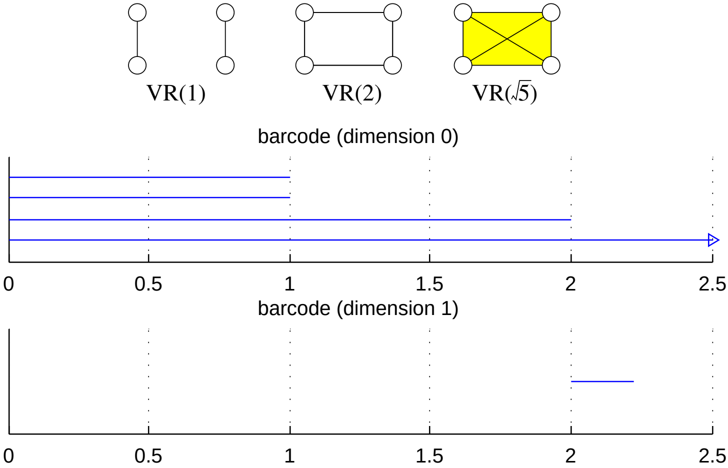
An increasing sequence of ϵ produces a filtration, i.e., a sequence of increasing simplicial complexes $VR(\epsilon_1) \subseteq VR(\epsilon_2) \subseteq \ldots$, with the property that a simplex enters the sequence no earlier than all its faces.

Persistent homology

- In a filtration, at what value of ϵ does a hole appear, and how long does it persist till it is filled in?

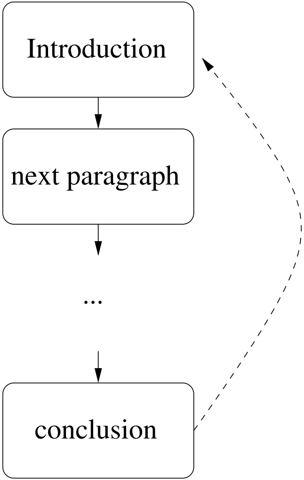
Persistent homology

- In a filtration, at what value of ϵ does a hole appear, and how long does it persist till it is filled in?
- Barcode



Applications to natural language processing

Good articles "tie back."



How can we capture such loopy structure in text documents?

Applications to natural language processing

- Some documents "straight," others "twist and turn"

Applications to natural language processing

- Some documents “straight,” others “twist and turn”
- Divide a document into small units $x_1, \ldots, x_n$ (e.g., sentences, paragraphs).

Applications to natural language processing

- Some documents “straight,” others “twist and turn”
- Divide a document into small units $x_1, \ldots, x_n$ (e.g., sentences, paragraphs).
- Given distance function $D(x_i, x_j) \geq 0$ (e.g., Euclidean, cosine)

Applications to natural language processing

- Some documents "straight," others "twist and turn"
- Divide a document into small units $x_1, \ldots, x_n$ (e.g., sentences, paragraphs).
- Given distance function $D(x_i, x_j) \geq 0$ (e.g., Euclidean, cosine)
- We will focus on the 0-th (clusters) and 1st (holes) order homology classes.

Example: Itsy bitsy spider

The Itsy Bitsy Spider climbed up the water spout
Down came the rain and washed the spider out
Out came the sun and dried up all the rain

And the Itsy Bitsy Spider climbed up the spout again

- bag-of-words

again	all	and	bitsy	came	climbed	down	dried	itsy	out	rain	spider	spout	sun	the	up	washed	water
0	0	0	1	0	1	0	0	1	0	0	1	1	0	2	1	0	1
0	0	1	0	1	0	1	0	0	1	1	1	0	0	2	0	1	0
0	1	1	0	1	0	0	1	0	1	1	0	0	1	2	1	0	0
1	0	1	1	0	1	0	0	1	0	0	1	1	0	2	1	0	0

Example: Itsy bitsy spider

The Itsy Bitsy Spider climbed up the water spout
Down came the rain and washed the spider out
Out came the sun and dried up all the rain

And the Itsy Bitsy Spider climbed up the spout again

- bag-of-words

again	all	and	bitsy	came	climb ed	down	dried	itsy	out	rain	spider	spout	sun	the	up	wash ed	water
0	0	0	1	0	1	0	0	1	0	0	1	1	0	2	1	0	1
0	0	1	0	1	0	1	0	0	1	1	1	0	0	2	0	1	0
0	1	1	0	1	0	0	1	0	1	1	0	0	1	2	1	0	0
1	0	1	1	0	1	0	0	1	0	0	1	1	0	2	1	0	0

- vertices

Similarity Filtration (SIF)

$D_{max} = \max D(x_i, x_j), \forall i, j = 1 \ldots n$
FOR $m = 0, 1, \ldots M$
    Add $VR\left(\frac{m}{M} D_{max}\right)$ to the filtration
END
Compute persistent homology on the filtration

- larger diameter, looser tie-backs

Similarity Filtration (SIF)

$D_{max} = \max D(x_i, x_j), \forall i, j = 1 \ldots n$
FOR $m = 0, 1, \ldots M$
    Add $VR\left(\frac{m}{M} D_{max}\right)$ to the filtration
END
Compute persistent homology on the filtration

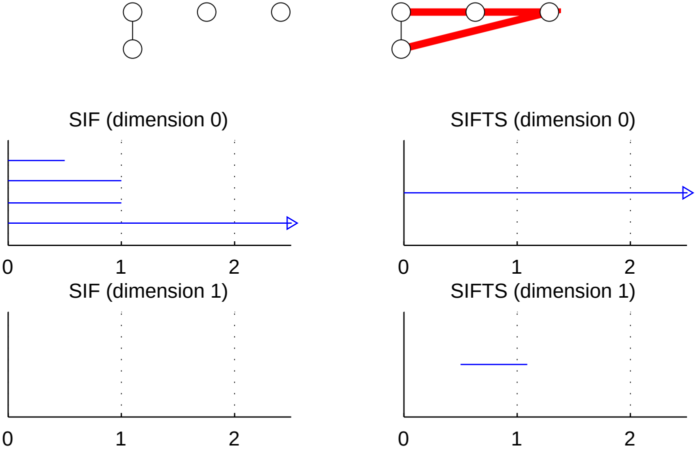
- larger diameter, looser tie-backs
- order of $x_1 \ldots x_n$ ignored

Similarity Filtration with Time Skeleton (SIFTS)

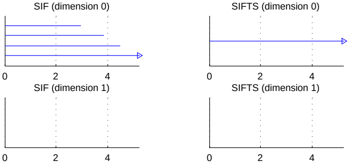
$D(x_i, x_{i+1}) = 0$ for $i = 1, \ldots, n-1$
$D_{max} = \max D(x_i, x_j), \forall i, j = 1 \ldots n$
FOR $m = 0, 1, \ldots M$
    Add $VR\left(\frac{m}{M} D_{max}\right)$ to the filtration
END
Compute persistent homology on the filtration

- time edges allow tie-back in time

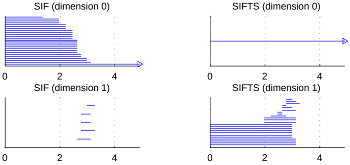
SIF vs. SIFTS on Itsy bitsy spider



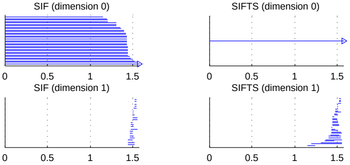
On Nursery Rhymes and Other Stories



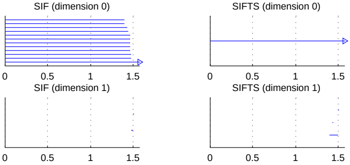
Row Row Row Your Boat



London Bridge



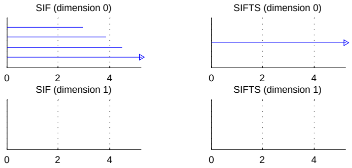
Little Red-Cap



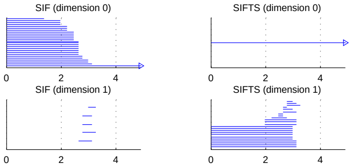
Alice in Wonderland

- London Bridge: "My fair Lady" repeats 12 times.

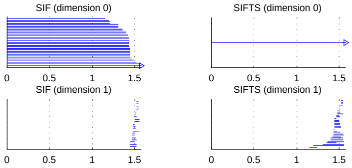
On Nursery Rhymes and Other Stories



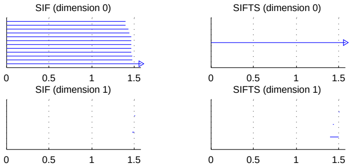
Row Row Row Your Boat



London Bridge



Little Red-Cap



Alice in Wonderland

- London Bridge: "My fair Lady" repeats 12 times.
- Little Red-Cap: "The better to see you with, my dear" and "The better to eat you with!"

On Child and Adolescent Writing

- Older writers have more complex barcodes?

On Child and Adolescent Writing

- Older writers have more complex barcodes?
- LUCY corpus: children (ages 9–12, 150 essays), undergraduates (48 essays)

On Child and Adolescent Writing

- Older writers have more complex barcodes?
- LUCY corpus: children (ages 9–12, 150 essays), undergraduates (48 essays)
- average article length: child=11.6 sentences, adolescent=25.8

On Child and Adolescent Writing

- Older writers have more complex barcodes?
- LUCY corpus: children (ages 9–12, 150 essays), undergraduates (48 essays)
- average article length: child=11.6 sentences, adolescent=25.8
- SIFTS barcode summary statistics:

On Child and Adolescent Writing

- Older writers have more complex barcodes?
- LUCY corpus: children (ages 9–12, 150 essays), undergraduates (48 essays)
- average article length: child=11.6 sentences, adolescent=25.8
- SIFTS barcode summary statistics:
  - holes?: what percentage of articles have H_1 holes

On Child and Adolescent Writing

- Older writers have more complex barcodes?
- LUCY corpus: children (ages 9–12, 150 essays), undergraduates (48 essays)
- average article length: child=11.6 sentences, adolescent=25.8
- SIFTS barcode summary statistics:
  - holes?: what percentage of articles have H_1 holes
  - $|H_1|$: number of holes in the article

On Child and Adolescent Writing

- Older writers have more complex barcodes?
- LUCY corpus: children (ages 9–12, 150 essays), undergraduates (48 essays)
- average article length: child=11.6 sentences, adolescent=25.8
- SIFTS barcode summary statistics:
  - holes?: what percentage of articles have H_1 holes
  - $|H_1|$: number of holes in the article
  - ϵ^*: the smallest ϵ when the first hole in H_1 forms.

	child	adolescent	adol. trunc.
holes?	87%	100%*	98%*
$\|H_1\|$	3.0 ($\pm$0.2)	17.6 ($\pm$0.9)*	3.9 ($\pm$0.2)*
ϵ^*	1.35 ($\pm$.02)	1.27 ($\pm$.02)*	1.38 ($\pm$.01)

*: statistically significantly different from "child"

Is Homology Merely Counting Repeats?

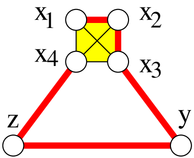
- On $x_1 \leadsto x_2 \leadsto x_3$ where x_1, x_2, x_3 SIFTS will find two holes: $x_1 \rightleftharpoons x_2$, $x_2 \rightleftharpoons x_3$

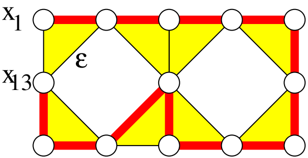
Is Homology Merely Counting Repeats?

- On $x_1 \rightsquigarrow x_2 \rightsquigarrow x_3$ where x_1, x_2, x_3 SIFTS will find two holes: $x_1 \rightleftharpoons x_2$, $x_2 \rightleftharpoons x_3$
- k such repeats of x will generate $k-1$ holes. The Betti number $\beta_1 = k - 1$?

Is Homology Merely Counting Repeats?

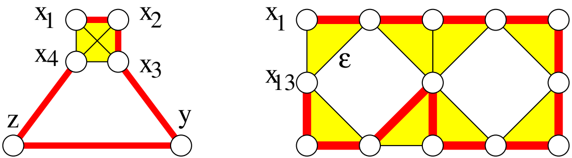
- On $x_1 \rightsquigarrow x_2 \rightsquigarrow x_3$ where x_1, x_2, x_3 SIFTS will find two holes: $x_1 \rightleftharpoons x_2$, $x_2 \rightleftharpoons x_3$
- k such repeats of x will generate $k-1$ holes. The Betti number $\beta_1 = k - 1$?
- No.





Is Homology Merely Counting Repeats?

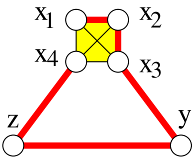
- On $x_1 \rightsquigarrow x_2 \rightsquigarrow x_3$ where x_1, x_2, x_3 SIFTS will find two holes: $x_1 \rightleftharpoons x_2$, $x_2 \rightleftharpoons x_3$
- k such repeats of x will generate $k-1$ holes. The Betti number $\beta_1 = k-1$?
- No.

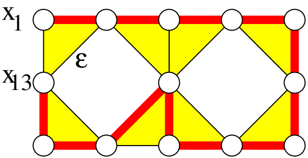


  - Left: $k-1=3$, SIFTS correctly finds $\beta_1 = 1$

Is Homology Merely Counting Repeats?

- On $x_1 \rightsquigarrow x_2 \rightsquigarrow x_3$ where x_1, x_2, x_3 SIFTS will find two holes: $x_1 \rightleftharpoons x_2$, $x_2 \rightleftharpoons x_3$
- k such repeats of x will generate $k-1$ holes. The Betti number $\beta_1 = k-1$?
- No.





  - Left: $k-1=3$, SIFTS correctly finds $\beta_1 = 1$
  - Right: $k-1=12$, merging x 0 holes, SIFTS correctly finds $\beta_1 = 2$

Summary

- Persistent homology may offer new representations for machine learning

To read more, see the references in
Xiaojin Zhu. **Persistent homology: An introduction and a new text representation for natural language processing**. IJCAI, 2013.

Summary

- Persistent homology may offer new representations for machine learning
- How to best use it?

To read more, see the references in
Xiaojin Zhu. **Persistent homology: An introduction and a new text representation for natural language processing**. IJCAI, 2013.