Approved for Public Release; Distribution is Unlimited

DARPA Final Report
August 2006–March 2010

Grant #HR0011-06-1-0039

Technical Area: Defense Sciences

Title: Topological Analysis of Partially Ordered Data

Lead Organization: University of California/San Diego

Research Areas: topology, statistics, combinatorics

Technical PoC: David Meyer
Department of Mathematics
University of California/San Diego
La Jolla, CA 92093-0112
dmeyer@math.ucsd.edu
(858)534-5524; fax (858)534-5273

Administrative PoC: Joyce Cheng
Department of Mathematics
University of California/San Diego
La Jolla, CA 92093-0112
jccheng@math.ucsd.edu
(858)534-40231; fax (858)534-5273

20100811063

1. Executive Summary

The DARPA Topological Data Analysis program has supported the development of topological methods for analyzing large sets of points in metric spaces [1–5]. The goal of this project is to develop similar methods for data sets on which the natural structure is not a metric, but rather a *partial order*. These methods are being developed for application to Boolean data, *i.e.*, sets of bit strings, and to geometrical data, *e.g.*, points in Lorentzian manifolds.

2. Research Summary

During the first quarter of Year 1 of the grant I began formulating a framework for topological analysis of points in Lorentzian manifolds, partially ordered by causality. This formulation was the basis for two talks I gave at the European Marie Curie Research Training Network on Random Geometry conference *Causal Sets 2006*: "Dimension of causal sets" and "Algebra and topology of partial orders".

During the second quarter I developed definitions for filtered topologies on antichains in partial orders based upon volume and height.

During the third quarter I developed code for sampling uniformly according to the volume form in compact subsets of two dimensional Minkowski space, and in cylinders of bounded height with flat Lorentzian metrics (Minkowski cylinders).

During the fourth quarter I used PLEX to compute the persistent homology of antichains in posets sampled from Minkowski cylinders and showed that they have persistent first Betti number over a long range of scales.

During the first quarter of Year 2 of the grant I investigated the possibility of estimating another persistent topological property, dimension, for specific *metric* datasets relevant to political science. I described preliminary results of this analysis in a talk, "Long range dependencies in violent events timeseries", in the Temporal Quantitative Analysis Panel session at the *American Political Science Association Annual Meeting*, and in a talk "Long-range temporal and spatial correlations in conflict" in the session on Alternative Conceptions of State Strength: the Role of State Capacity for Development and Peace at the *Sixth Pan-European International Relations Conference*.

During the second quarter I applied the geometrical methods developed for Lorentzian data to probability measures on finite sample spaces, related by the Bayesian order [6]. In this case the homological structure is delicate, and is only persistent for coarse measurements [7].

During the third quarter we began investigating the possibility of analyzing the *order complex* of a set of partially ordered data within the framework of persistence homology.

During the fourth quarter we concluded that for geometrical posets there was little useful information that could be obtained from persistence calculations about the order complex. We also began investigating the possibility of analyzing sets of symmetric or Hermitian matrices, partially ordered by positivity or majorization.

During the first quarter of Year 3 of the grant we developed methods for simulating Boolean test data.

During the second quarter we demonstrated that simulated Boolean test data could be generated with persistent nonzero first and second Betti numbers. As with the geometrical data, the calculations were implemented with PLEX.

During the third quarter we sampled points uniformly from a singular Minkowski space ("pants") with spatial topology changing from S^1 to $S^1 \times S^1$, both intrinsically, and also from a $2+1$ dimensional Minkowski space in which the pants were embedded. We showed that persistent homology calculations on antichains at increasing heights in the poset capture the topology change [8].

During the fourth quarter we began formulating our partially ordered data persistence constructions in a category-theoretic framework.

During the first quarter of Year 4 of the grant we continued to frame our results category-theoretically.

During the second quarter we began drafting a paper on our results from this project.

During the third and final quarter we continued drafting the paper describing our results from this project. The current version of this draft is attached [9].

3. Management Summary

3.1. Personnel

This grant supported the work of the PI, David Meyer; his collaborator, James Pommersheim (Reed College); and his graduate students at UCSD, Durdu Güney and Ben Wilson.

4. Publications and references

[1] H. Edelsbrunner, D. Letscher and A. Zomorodian, "Topological persistence and simplification", in *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, 12–14 November 2000, Redondo Beach, CA (Los Alamitos, CA: IEEE Computer Society Press 2000) 454–463;
Discrete and Computational Geometry **28** (2002) 511–533.

[2] A. Zomorodian and G. Carlsson, "Computing persistent homology", *Discrete and Computational Geometry* **33** (2005) 249–274.

[3] V. de Silva and G. Carlsson, "Topological estimation using witness complexes", in M.

Alexa, M. Gross, H. Pfister and S. Rusinkiewicz, eds., *Symposium on Point-Based Graphics 2004* (Eurographics 2004) 157–166.

[4] `http://math.stanford.edu/comptop/programs/plex/`.

[5] A. Collins, A. Zomorodian, G. Carlsson and L. Guibas, "A barcode shape descriptor for curve point cloud data", *Computers & Graphics* **28** (2004) 881–894.

[6] B. Coecke and K. Martin, "A partial order on classical and quantum states", Oxford University Programming Research Group Report PRG-RR-02-07 (2002).

*[7] D. A. Meyer, "Topological analysis of geometrical poset data", DARPA *Topological Data Analysis Annual Meeting*, 10–11 December 2007, Coronado, CA (2007).

*[8] D. A. Meyer, "Topological analysis of partially ordered data", DARPA *Topological Data Analysis Annual Meeting*, 21–23 January 2009, Santa Barbara, CA (2009).

*[9] D. A. Meyer and B. Wilson, "Topological analysis of partially ordered data", incomplete draft (2010).

* indicates material prepared for this project and appended to this report.

o olo ical anal sis of eo etrical oset ata

David A. Meyer

Department of Mathematics
University of California/San Diego

dmeyer@math.ucsd.edu

Topological Data Analysis Program Review
Coronado, CA, 11 December 2007

Outline

Introduction: metric *versus* partially ordered data

Background: partially ordered sets (posets)

Topologies on (sub)posets: from open sets to simplicial complexes

Geometrical examples: Lorentzian manifolds, probability distributions

Conclusion: directions

Introduction

Metric and non-metric data

Topological analysis has been applied to data sets consisting of large numbers of points in—typically—high dimensional spaces.

Construction of a Rips (or Čech, or Delauney, or α-shape) complex from the data depends on this being a metric space.

But some data sets are not naturally samples from a metric space.

Since this is final exam week at UCSD ...

Example: test results

Consider a test consisting of a set of questions T, and a set of students S who take the test.

Let $C_i \subseteq T$ be the subset of questions answered correctly by student $i \in S$.

With no additional information about the relative importance of the test questions, the only certain comparisons we can make are:

$$\text{student } i \text{ performed no better on the test than did student } j$$
$$\Updownarrow$$
$$C_i \subseteq C_j$$

Thus the data points C_i are most naturally not samples from a (high dimensional) metric space, but are rather samples from the partially ordered set of subsets of T, ordered by inclusion.

Goal: topological analysis of partially ordered data

In the example, $\{C_i \mid i \in S\}$ is a 'point cloud' in a partially ordered set.

The simplest partially ordered sets are totally ordered. These have no interesting topology, being effectively one-dimensional (and simply connected).

In general, however, partially ordered sets are analogous to multi-dimensional spaces.

Thus the goal is to extend the ideas of topological data analysis to provide statistical analysis of partially ordered data.

ac roun

Partially ordered sets

A partially ordered set (poset) is a set with a binary relation $\preceq$ that is:

$$\begin{aligned} \text{reflexive:} \quad & x \preceq x; \\ \text{antisymmetric:} \quad & x \preceq y \wedge y \preceq x \;\Rightarrow\; x = y; \\ \text{transitive:} \quad & x \preceq y \wedge y \preceq z \;\Rightarrow\; x \preceq z. \end{aligned}$$

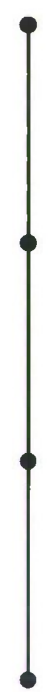
If $x \preceq y$ and $x \neq y$ then $x \prec y$.

Examples

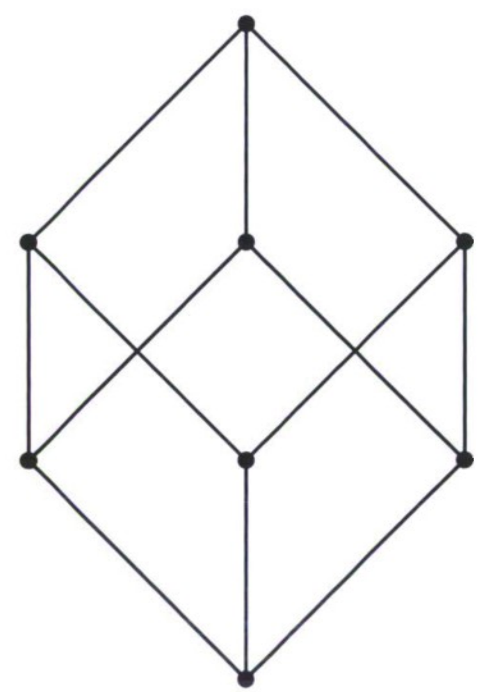
Let $[n] = \{x \in \mathbb{N} \mid x \leq n \in \mathbb{N}\}$.

n is the set $[n]$ with the relation $1 \prec 2 \prec \cdots \prec n$.

B_n is the set $2^{[n]}$ ordered by inclusion.



4



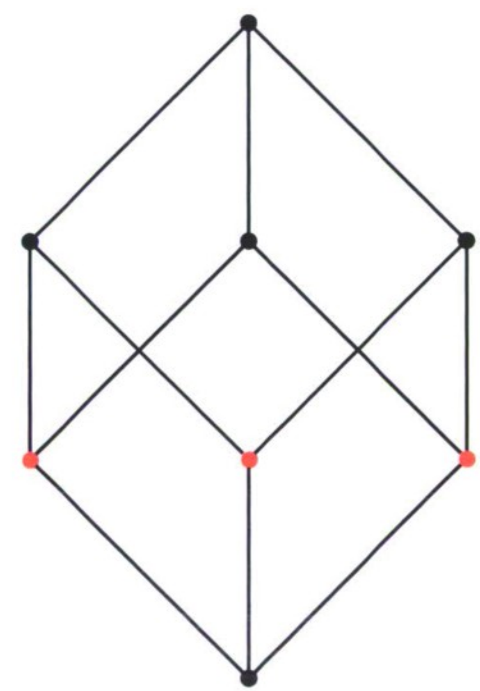
B_3

Subposets

Q is a subposet of a poset P if $Q \subseteq P$ and $x \preceq y$ in Q iff $x \preceq y$ in P.

An antichain is a subposet $A \subseteq P$ such that for all $x \neq y \in A$, $x \npreceq y$ and $y \npreceq x$.

$\mathbf{4}$

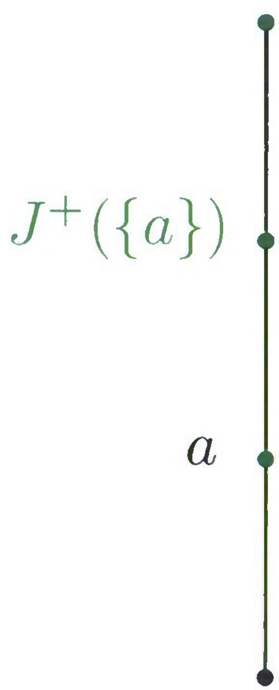


B_3

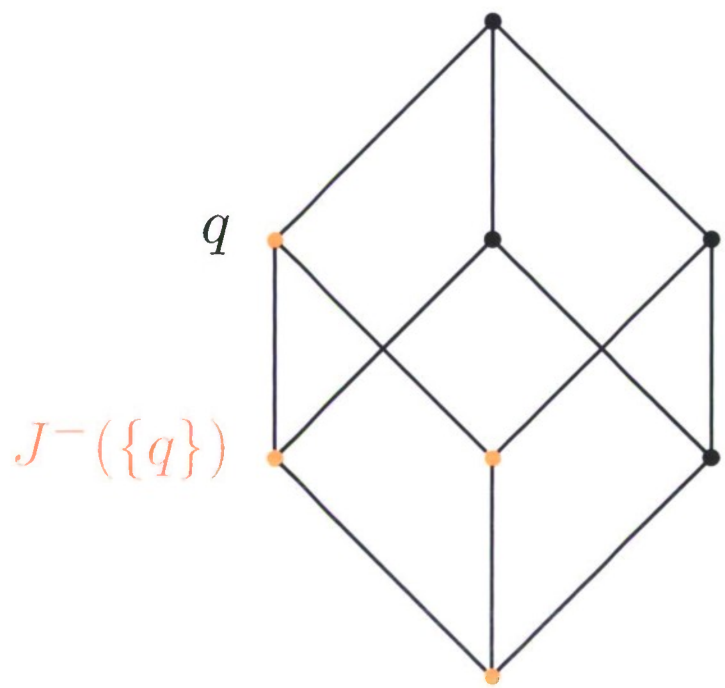
Up- and down-sets

For $Q \subseteq P$, the up-set of Q, $J^+(Q) = \{x \in P \mid \exists q \in Q,\, q \preceq x\}$.

Similarly, the down-set of Q, $J^-(Q) = \{x \in P \mid \exists q \in Q,\, x \preceq q\}$.



4



B_3

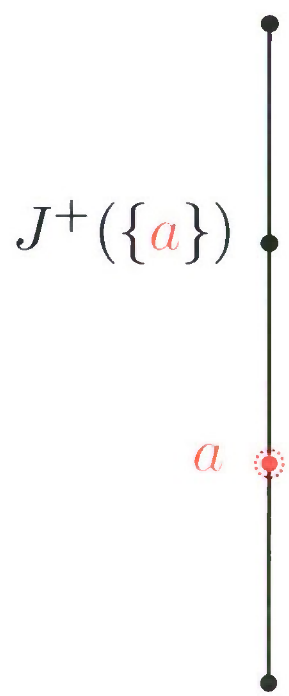
Topologies on [illegible]su[illegible]osets

Open covers of antichains

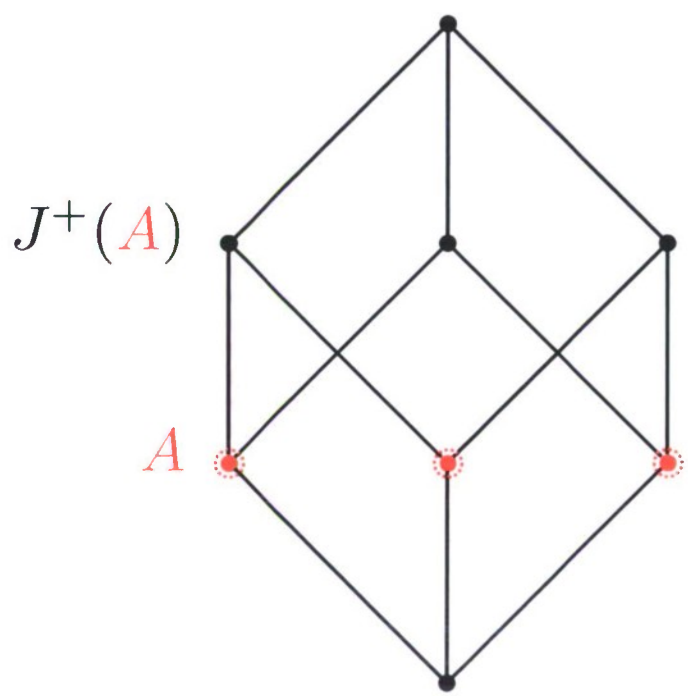
Let $A \subseteq P$ be an antichain.

For $x \in J^+(A)$, let $U_x = \{a \in A \mid a \preceq x\}$.

Any $A \subseteq Y \subseteq J^+(A)$ defines an open cover of A: $\mathcal{U}_Y = \{U_y \mid y \in Y\}$.



4



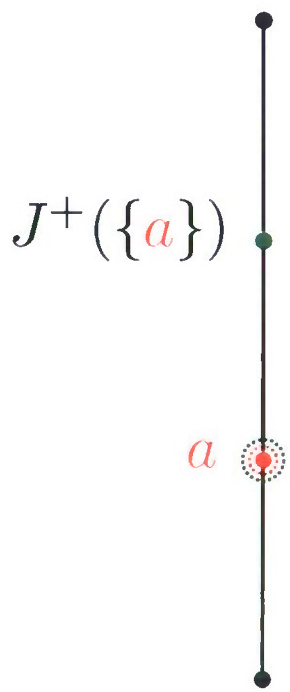
B_3

Open covers of antichains

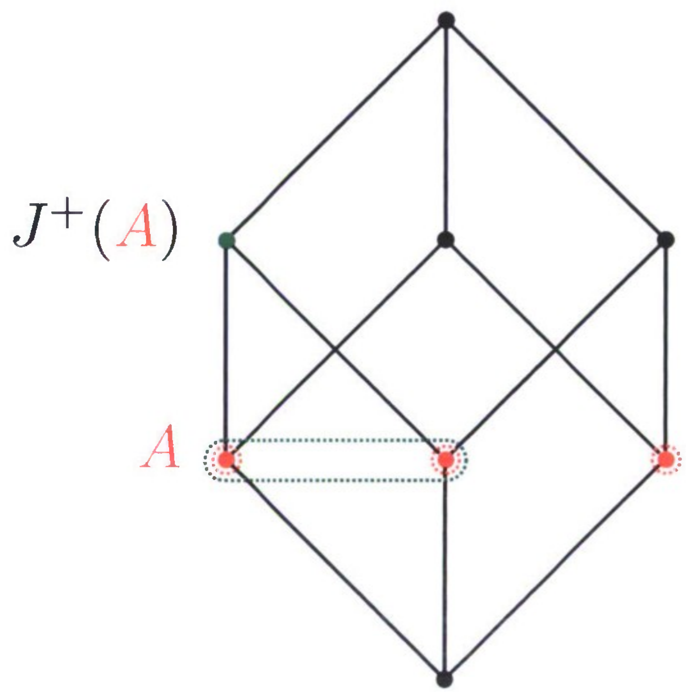
Let $A \subseteq P$ be an antichain.

For $x \in J^+(A)$, let $U_x = \{a \in A \mid a \preceq x\}$.

Any $A \subseteq Y \subseteq J^+(A)$ defines an open cover of A: $\mathcal{U}_Y = \{U_y \mid y \in Y\}$.



4



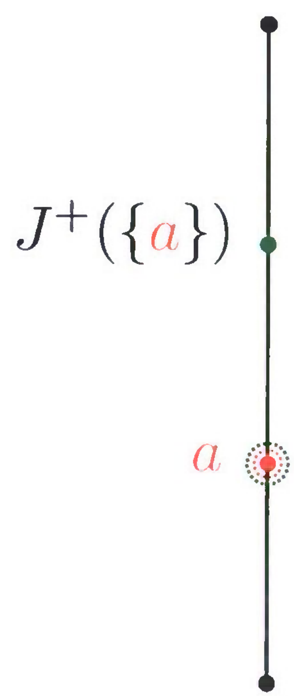
B_3

Open covers of antichains

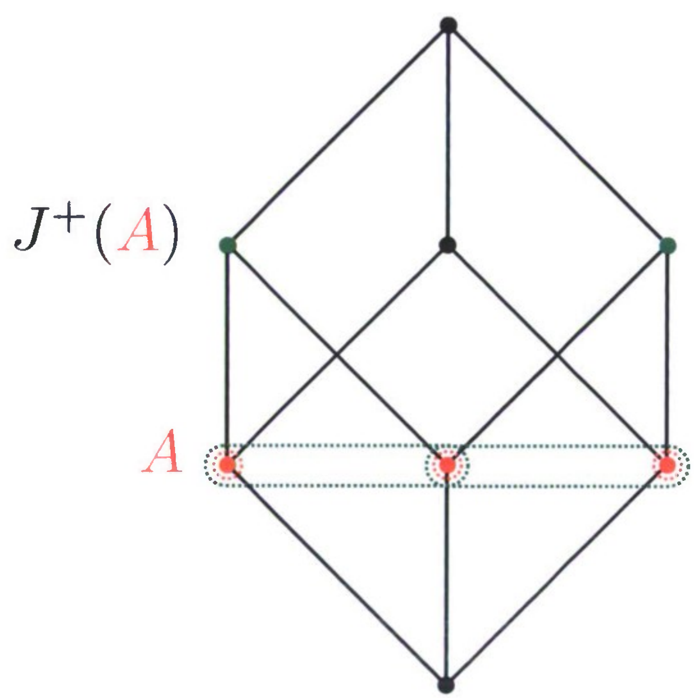
Let $A \subseteq P$ be an antichain.

For $x \in J^+(A)$, let $U_x = \{a \in A \mid a \preceq x\}$.

Any $A \subseteq Y \subseteq J^+(A)$ defines an open cover of A: $\mathcal{U}_Y = \{U_y \mid y \in Y\}$.



4



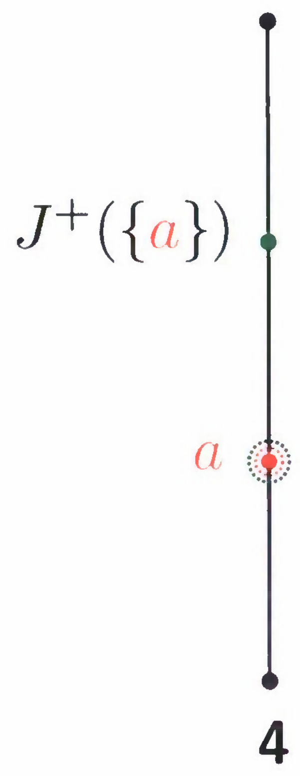
B_3

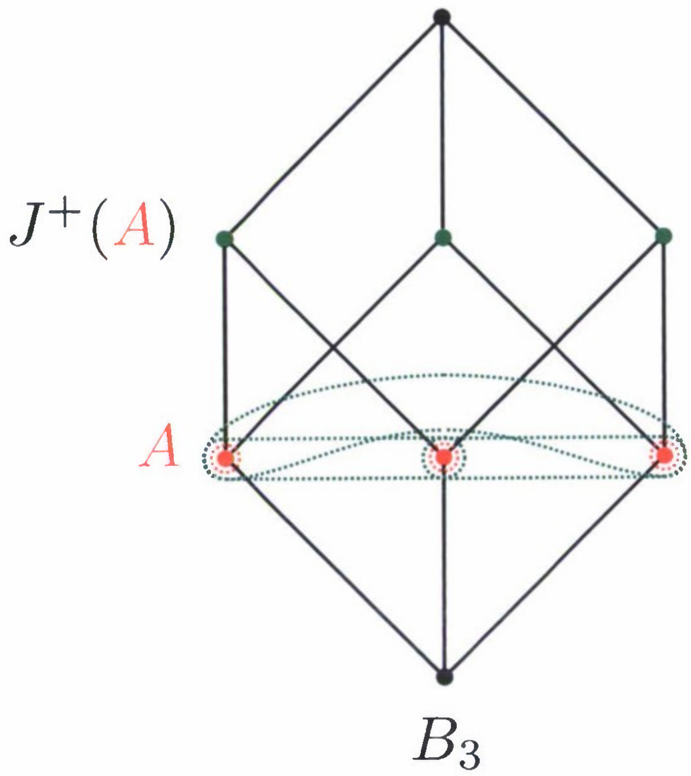
Open covers of antichains

Let $A \subseteq P$ be an antichain.

For $x \in J^+(A)$, let $U_x = \{a \in A \mid a \preceq x\}$.

Any $A \subseteq Y \subseteq J^+(A)$ defines an open cover of A: $\mathcal{U}_Y = \{U_y \mid y \in Y\}$.



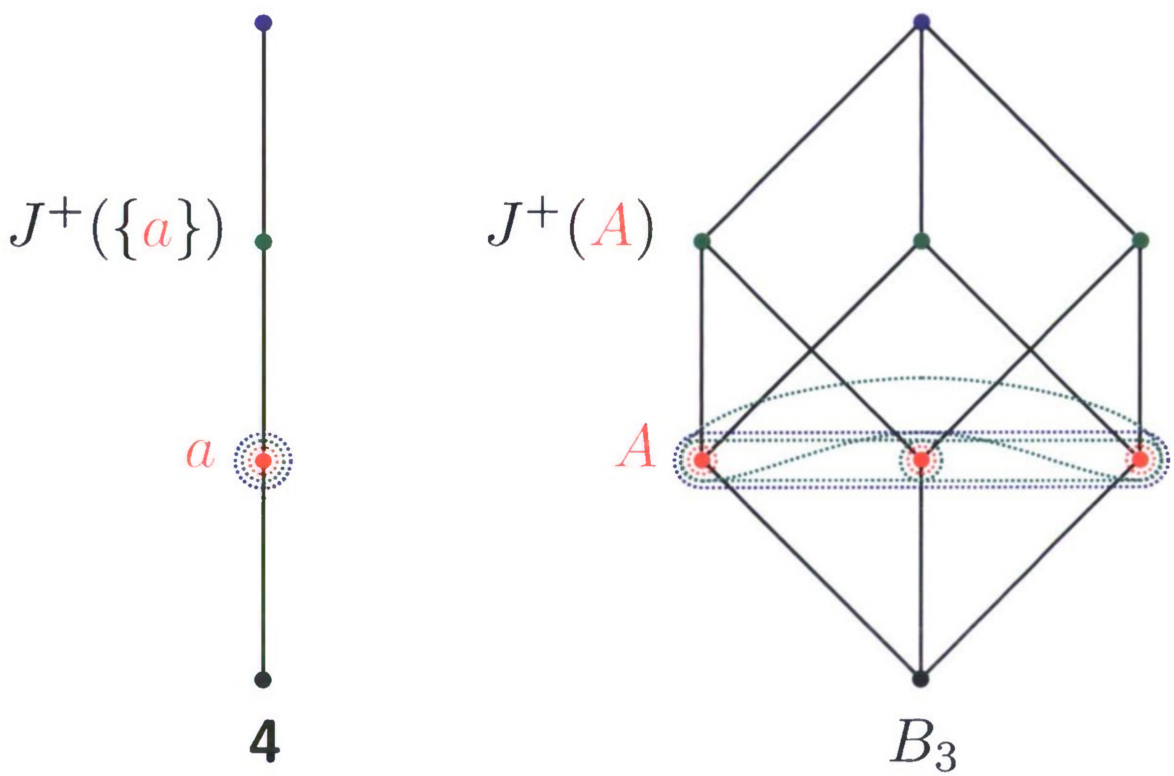


Open covers of antichains

Let $A \subseteq P$ be an antichain.

For $x \in J^+(A)$, let $U_x = \{a \in A \mid a \preceq x\}$.

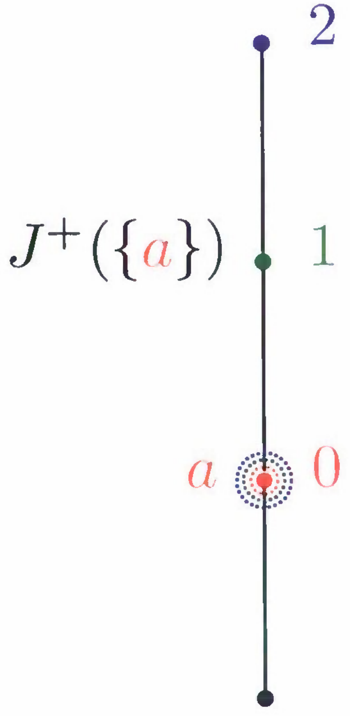
Any $A \subseteq Y \subseteq J^+(A)$ defines an open cover of A: $\mathcal{U}_Y = \{U_y \mid y \in Y\}$.



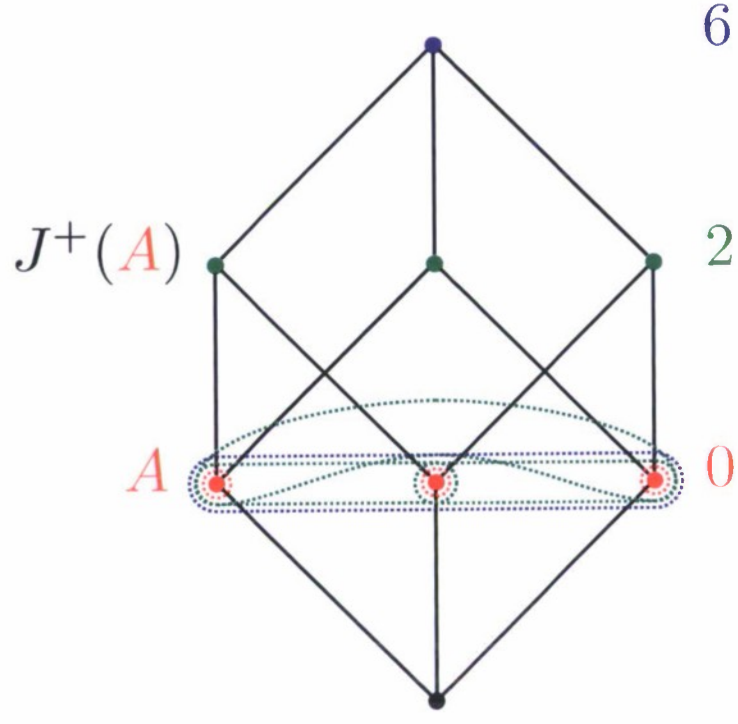
Nerves of open covers

The nerve of $\mathcal{U}_Y$ is the simplicial (Čech) complex, $N(\mathcal{U}_Y)$, with k-simplices the subsets $\{y_0, \ldots, y_k\} \subseteq Y$, for which $\bigcap_{i=0}^{k} U_{y_i} \neq \emptyset$.

Let $N^v(A)$ be the nerve of the open cover $\mathcal{U}_Y$ of A defined by $A \subseteq Y = \{y \in J^+(A) \mid |J^-(\{y\}) \cap J^+(A)| \leq v + 1\} \subseteq J^+(A)$.



4



B_3

Persistence

Let $0 = v_0 < v_1 < \cdots < v_m = \max_{y \in J^+(A)} |J^-(\{y\}) \cap J^+(A)| - 1$.

$N^{v_0}(A) \subseteq \cdots \subseteq N^{v_m}(A) = N(\mathcal{U}_{J^+(A)})$ is a filtration of $N(\mathcal{U}_{J^+(A)})$.

As in the metric case, the filtered complex $N(\mathcal{U}_{J^+(A)})$, with inclusion maps between simplices, gives a persistence complex, with persistent homology.

Also as in the metric situation, it is more convenient for calculations to define the filtered Rips complex, with $R^v(A)$ having k-simplices the subsets $\{a_0, \ldots, a_k\} \subseteq A$ for which there exists $y \in J^+(A)$ with $|J^-(\{y\}) \cap J^+(A)| < v$ and $\{a_0, \ldots, a_k\} \subseteq J^-(\{y\})$.

Unlike the metric situation, the Rips complex is also more convenient for calculations because it has vertex set $A \subseteq J^+(A)$, the vertex set for the nerve complex.

Geometrical examples

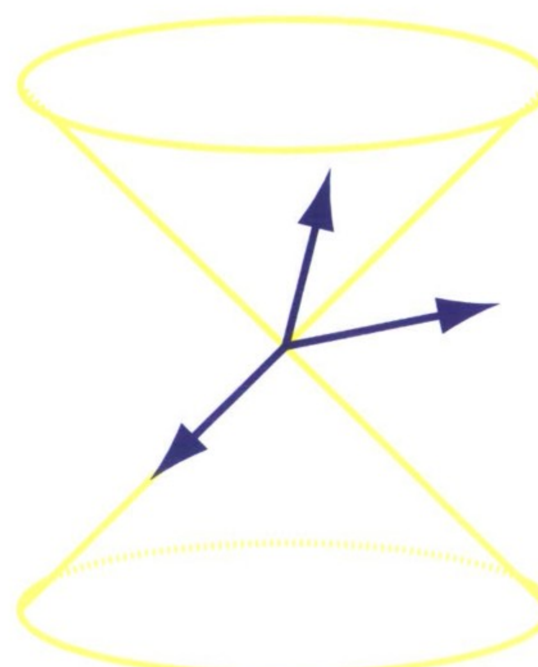
Geometrical posets

Although posets arise in many contexts, to understand the topology being computed it is useful to consider some geometrical examples.

The motivating examples come from Lorentzian manifolds.

A Lorentzian manifold is a smooth manifold M with a metric tensor $g : T_p(M) \times T_p(M) \to \mathbb{R}$ of signature $(-, +, \ldots, +)$.

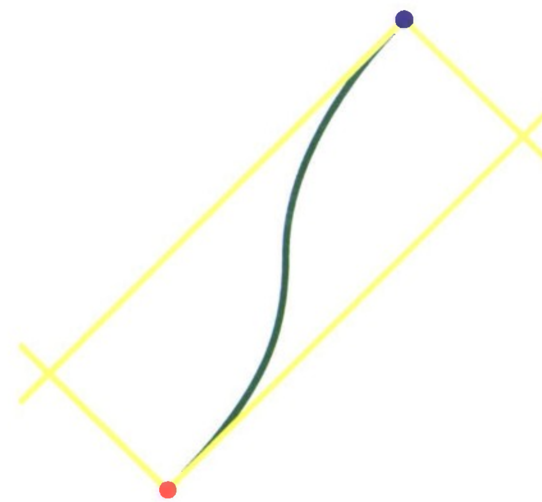
A tangent vector $v \in T_p(M)$ is timelike, null, or spacelike if $g(v, v)$ is negative, zero, or positive, respectively.



For v, w timelike, let $v \sim w$ if $g(v, w) < 0$; this is an equivalence relation with two classes. One is designated future-directed; this designation extends to null vectors in the closure of the set of future-directed timelike vectors.

Posets from Lorentzian manifolds

A C^0, piecewise C^1 curve $\gamma : I \to M$ is causal if every tangent vector to γ is timelike or null, and future-directed.



For $x, y \in M$, x causally precedes y, $x \preceq y$, if there is a causal curve such that $\gamma(0) = x$ and $\gamma(1) = y$.

'Causally precedes' is a partial order on M.

The metric tensor defines a volume form $\sqrt{|\mathrm{det}(g)|}\mathrm{d}x^0 \wedge \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^d$ on M, with respect to which stochastic point processes are defined.

So consider poset data obtained by random sampling from a Lorentzian manifold, retaining only the causal ordering.

A Lorentzian cylinder

Let $M = I \times S^1$. Choose coordinates $0 \leq t \leq 1$, $0 \leq \theta < 2\pi$, in which the metric has the form

$$\begin{pmatrix} -h^2 & 0 \\ 0 & 1 \end{pmatrix};$$

so the volume of M is $2\pi h$.

Sample n points uniformly at random relative to the volume form.

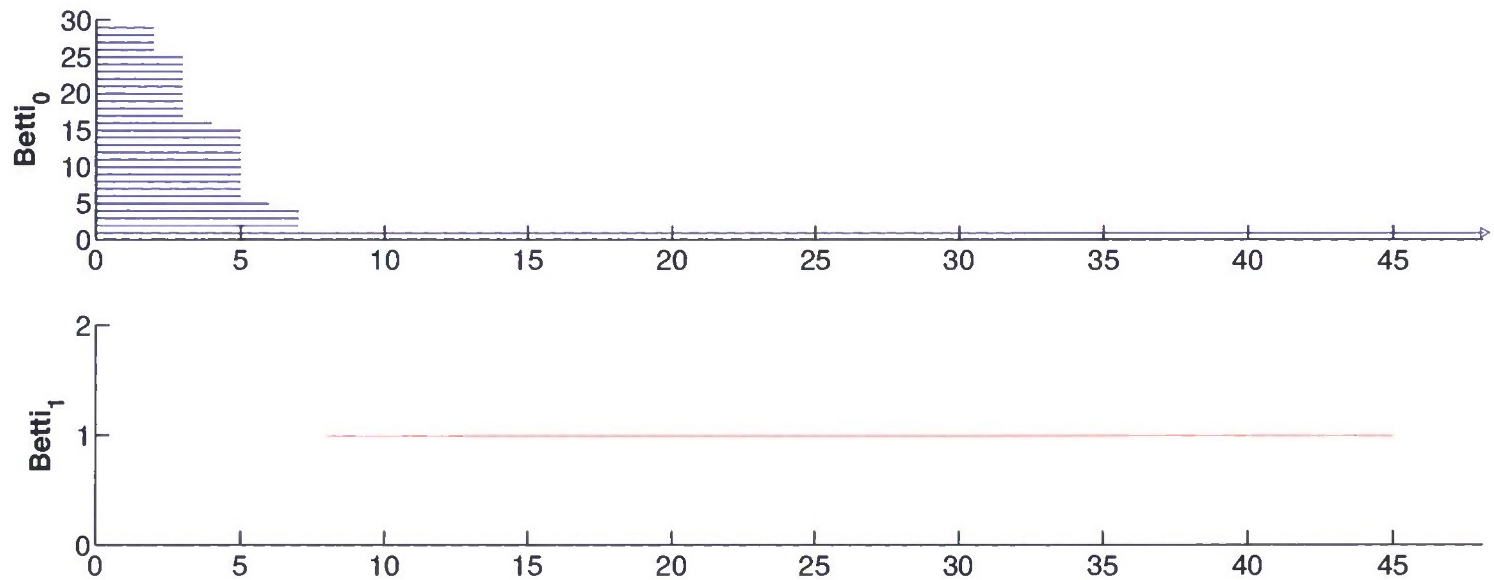
The causal order on these points defines a poset P of size n

Let $A = \{a \in P \mid J^-(\{a\}) = 1\}$ be the antichain of minimal elements in P; A lies in a spacelike surface in M.

Compute persistent homology for the filtered Rips complex $R(\mathcal{U}_{J^+(A)})$.

Persistent homology for Lorentzian cylinder poset data

Some straightforward MATLAB code generates a filtered graph which is fed to PLEX for calculation of persistent homology.



$$n = 1000$$

A 1-cycle is created, and then persists until it gets filled in as past lightcones cover a little more than $1/3$ of the initial spacelike S^1.

Finite probability measures

Let $[d]$ be a finite sample space with a probability measure $p : [d] \to \mathbb{R}$, so that $p_i \geq 0$ and $\sum p_i = 1$.

Probability measures are totally ordered by entropy:

$$S(p) = -\sum_{i=1}^{d} p_i \log p_i.$$

The d atomic measures $p_i = \delta_{ki}$ for $k \in [d]$ each have the minimum entropy, 0, while the uniform measure $p_i = 1/d$ has the maximum entropy, $\log d$.

But there are also natural partial orders on the manifold of probability measures. Majorization is one, but there is also . . .

The Bayesian order [Coecke & Martin 2002]

Suppose p is a prior measure on $[d]$ and some experiment reveals that $i \neq k \in [d]$. Then the posterior measure is

$$\Pi_{\hat{k}} p = \frac{1}{1 - p_k}(p_1, \ldots, p_{k-1}, 0, p_{k+1}, \ldots, p_d).$$

The Bayesian order is defined recursively by

$$p \preceq q \iff (\forall k \in [d])(p_k, q_k < 1 \Rightarrow \Pi_{\hat{k}} p \preceq \Pi_{\hat{k}} q).$$

"p is less uncertain than q about i before i is revealed not to be k iff p is less uncertain than q about i afterwards."

Since this definition is recursive, it is necessary to define

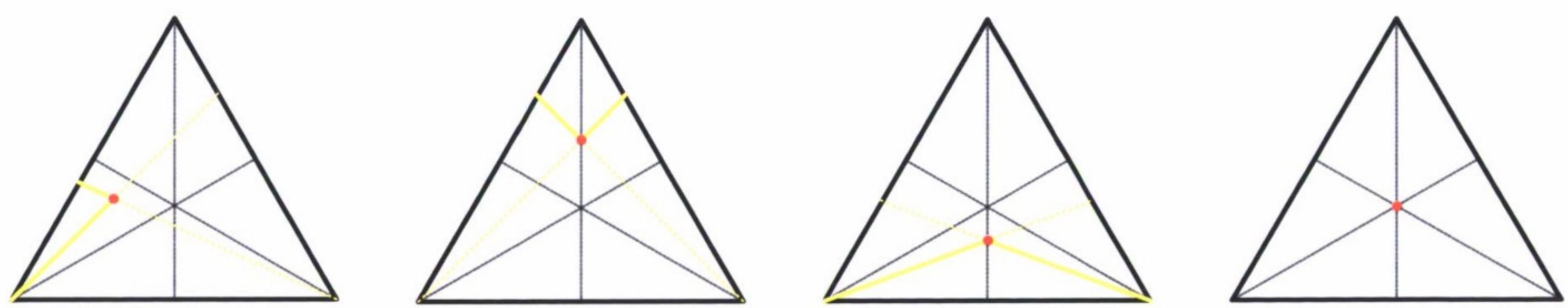
$$(\exists i \neq j)(p_i = 1 - p_j \wedge q_i = 1 - q_j) \implies p \preceq q \Leftrightarrow S(p) \leq S(q).$$

The Bayesian order [Coecke & Martin 2002]

A probability measure p on $[d]$ is decreasing if $(\forall i \in [d-1])(p_i \geq p_{i+1})$.

$p \preceq q$ in the Bayesian order iff there is a permutation $\sigma \in S_d$ such that $p \circ \sigma$ and $q \circ \sigma$ are each decreasing, and for all $i \in [d-1]$

$$(p \circ \sigma)_{i+1}(q \circ \sigma)_i \leq (p \circ \sigma)_i(q \circ \sigma)_{i+1}.$$



The down-sets for various p when $d = 3$.

Partially ordered measure data

Sample n points uniformly at random on the unit $d-1$ dimensional Euclidean simplex.

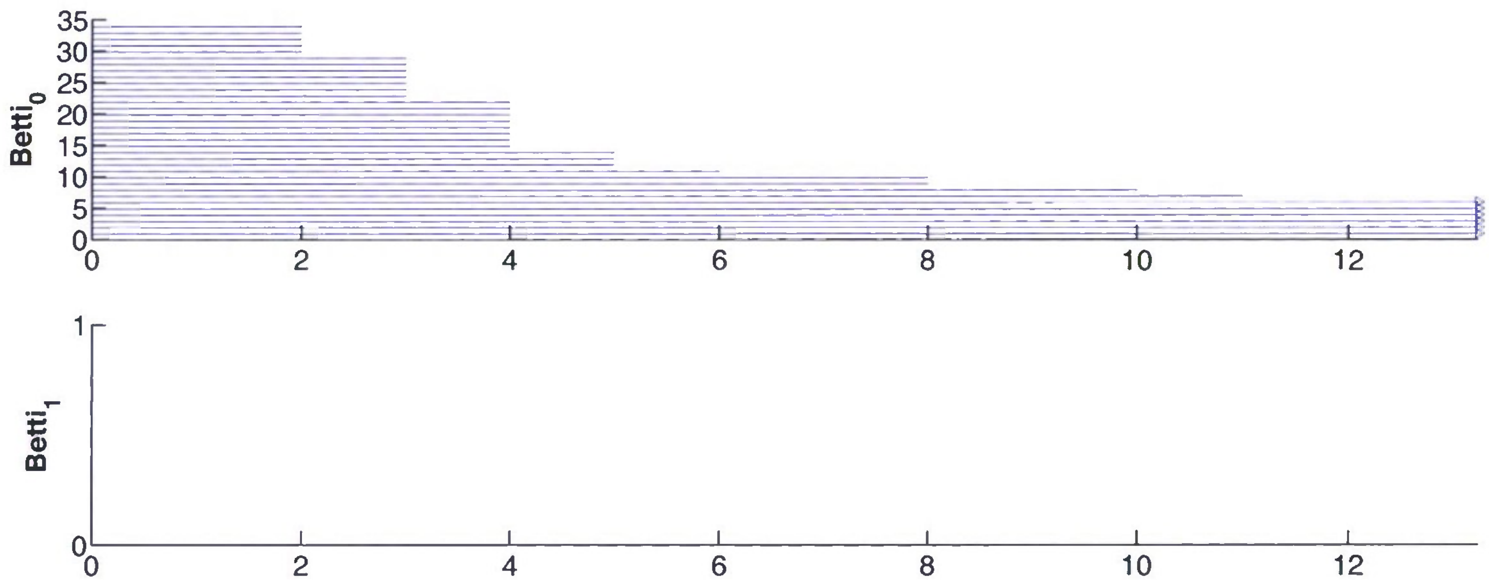
The Bayesian order on these points defines a poset P of size n

Let $A = \{a \in P \mid J^{-}(\{a\}) = 1\}$ be the antichain of minimal elements in P.

Compute persistent homology for the filtered Rips complex $R(\mathcal{U}_{J+(A)})$.

Persistent homology for Bayesian measure poset data

More straightforward MATLAB code generates a filtered graph which is fed to PLEX for calculation of persistent homology.



$n = 1000,\ d = 3$

Only six 0-cycles persist; these correspond to the 2-cells of the barycentric subdivision of the 2-simplex.

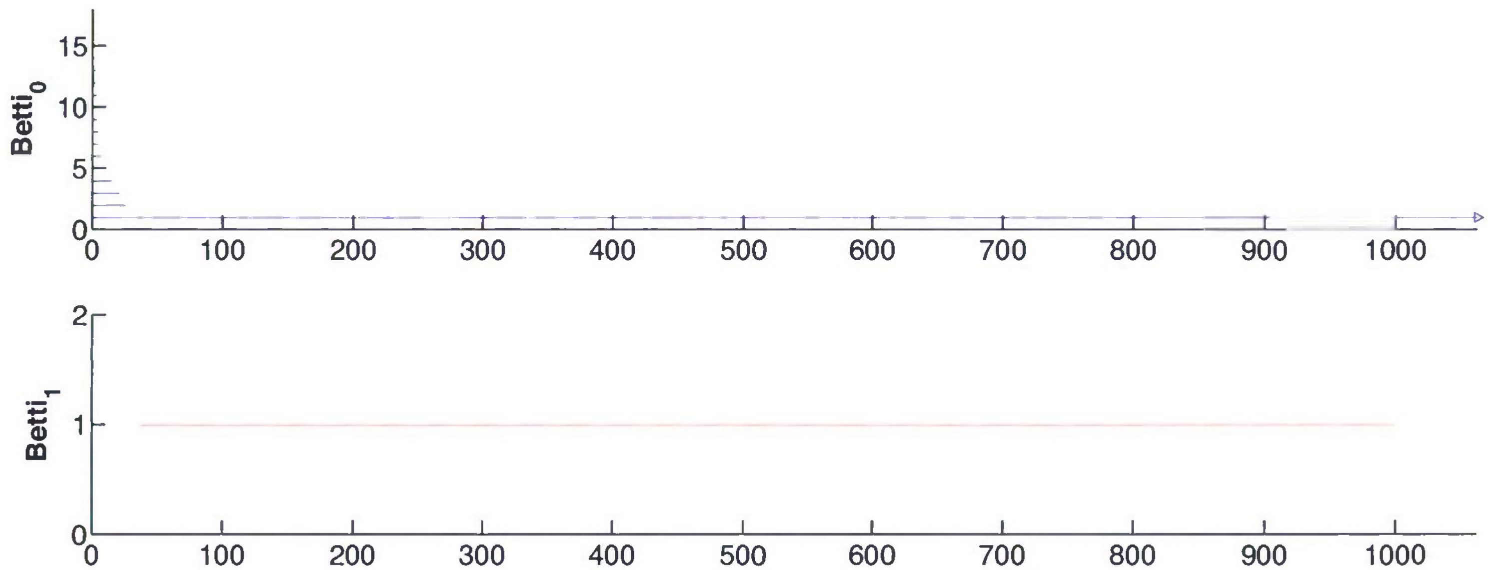
Persistent homology for coarse Bayesian measure poset data

To capture the topology induced by non-generic/singular probability measures, *i.e.*, those with $p_{i_1} = p_{i_2} = \cdots = p_{i_k}$ for some k distinct $i_j \in [d]$, with a finite sample, the sample must be coarse.

Scale each random sample by $c \cdot d!$ for some $c \in \mathbb{N}$ and round off to the nearest $x \in \mathbb{N}^d$ with $\sum x_i = c \cdot d!$.

The $d!$ factor in the renormalization puts integer points on each of the singular sets/cells in the barycentric subdivision, so there is positive probability of sampling them.

Persistent homology for coarse Bayesian measure poset data



$$n = 1000,\ d = 3,\ c = 10$$

Now five of the six 0-cycles persist only for volumes up to a few tens; then the 1-cycle is formed and persists almost until the end; since the maximum point was one of the samples, it eventually kills the 1-cycle.

Conclusion

Summary

Persistent topology depends on a filtered family of open sets. These open sets need not come from increasing a distance in a metric space; they can also be constructed for partially ordered data.

Certain data sets are naturally partially ordered.

For the geometrical partial ordered sets considered, the persistent topology of the minimal antichain is correct.

Coarsening sampled data helped capture the delicate topology created by non-generic points.

Notes

The same construction gives a persistence complex for any subposet, not just the antichain of minimal elements.

This construction can be modified to use a different filtration, *e.g.*, the length k of a longest chain, $c_0 \prec c_1 \prec \cdots \prec c_k$, rather than the volume.

It can also be modified to use intervals, $[x, y] = \{z \in P \mid x \prec z \prec y\}$, rather than down-sets to define the open sets.

Directions

Construction of spatial topology for causally ordered sets is of interest in discrete models for gravity.

In this context, and others, it is also important to estimate the dimension of the partially ordered data.

Analysis of the Bayes order on finite probability measures is a step towards analyzing symmetric or Hermitian matrix data, partially ordered by eigenvalues.

Other, non-geometrical data sets like test results, should be analyzed.

There are other simplicial complexes associated with posets—the order complex and its generalizations—which should be analyzed.

o olo ical anal sis of artiall or ere ata

David A. Meyer

(with **Jon Grice** and **Ben Wilson**)

Department of Mathematics
University of California/San Diego

dmeyer@math.ucsd.edu

Topological Data Analysis Program Review
Santa Barbara, CA, 21 January 2009

Outline

Introduction: metric *versus* partially ordered data

Constructions: partially ordered set (poset) topologies

Binomial examples: test results

Geometric examples: Lorentzian manifolds

Algebraic examples: quantum states [Friday]

Conclusion: directions [Friday]; summary

Introduction

Metric and non-metric data

Topological analysis has been applied to data sets consisting of large numbers of points in—typically—high dimensional spaces.

Construction of a Rips (or Čech, or Delauney, or α-shape) complex from the data depends on this being a metric space.

But some data sets are not naturally samples from a metric space.

Example: test results

Consider a test consisting of a set of questions T, and a set of students S who take the test.

Let $R_s \subseteq T$ be the subset of questions answered correctly by student $s \in S$.

With no additional information about the relative importance of the test questions, the only certain comparisons we can make are:

student s performed no better on the test than did student s'

$$\Updownarrow$$

$$R_s \subseteq R_{s'}$$

Thus the data points R_s are most naturally not samples from a (high dimensional) metric space, but are rather samples from the partially ordered set of subsets of T, ordered by inclusion.

Goal: topological analysis of partially ordered data

In the example, $\{R_s \mid s \in S\}$ is a 'point cloud' in a partially ordered set.

The simplest partially ordered sets are totally ordered. These have no interesting topology, being effectively one-dimensional (and simply connected).

In general, however, partially ordered sets are analogous to multi-dimensional spaces.

Thus the goal is to extend the ideas of topological data analysis to provide statistical analysis of partially ordered data.

onstructions

Partially ordered sets

A partially ordered set (poset) is a set with a binary relation $\preceq$ that is:

$$\begin{aligned} \text{reflexive:} \quad & x \preceq x; \\ \text{antisymmetric:} \quad & x \preceq y \wedge y \preceq x \;\Rightarrow\; x = y; \\ \text{transitive:} \quad & x \preceq y \wedge y \preceq z \;\Rightarrow\; x \preceq z. \end{aligned}$$

If $x \preceq y$ and $x \neq y$ then $x \prec y$.

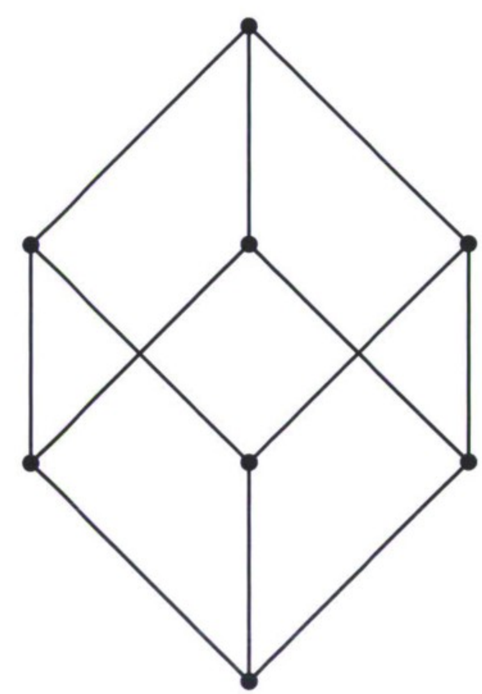
Examples

Let $[n] = \{x \in \mathbb{N} \mid x \leq n \in \mathbb{N}\}$.

n is the set $[n]$ with the relation $1 \prec 2 \prec \cdots \prec n$.

B_n is the set $2^{[n]}$ ordered by inclusion.

4



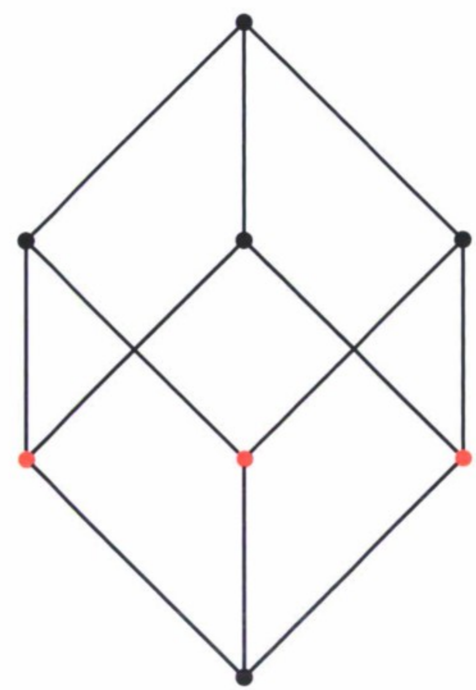
B_3

Subposets

Q is a subposet of a poset P if $Q \subseteq P$ and $x \preceq y$ in Q iff $x \preceq y$ in P.

An antichain is a subposet $A \subseteq P$ such that for all $x \neq y \in A$, $x \npreceq y$ and $y \npreceq x$.

4

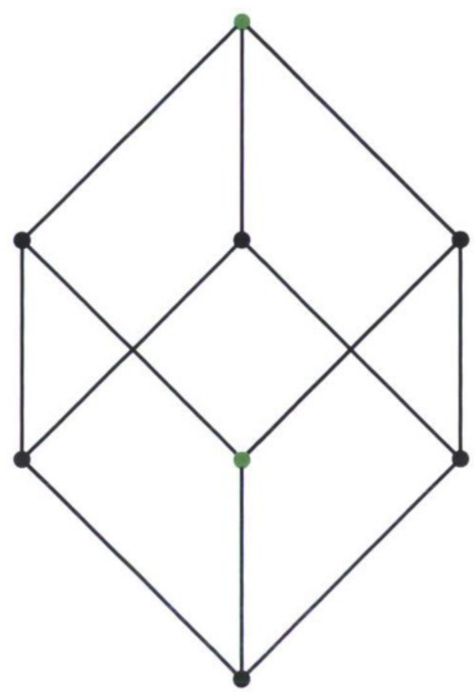


B_3

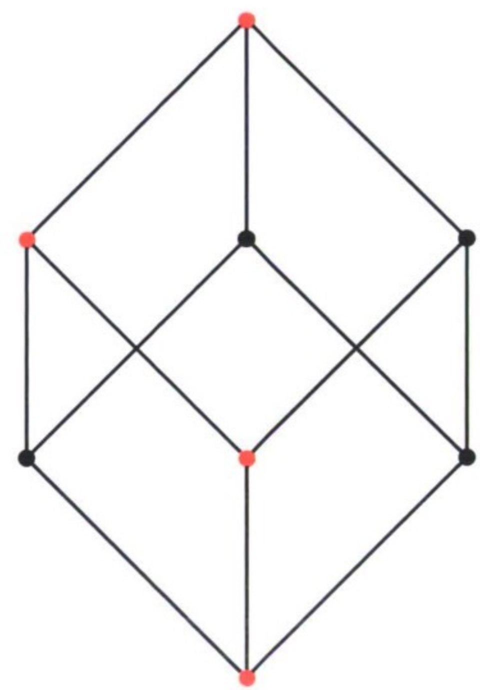
Chains

A subposet $C \subseteq P$ is a chain of length k if $C = \{x_1, \ldots, x_k\}$ and $x_1 \prec x_2 \prec \cdots \prec x_k$ in P.

A chain $C \subseteq P$ is maximal if for all $y \in P \setminus C$, the subposet $C \cup \{y\} \subseteq P$ is not a chain.



C

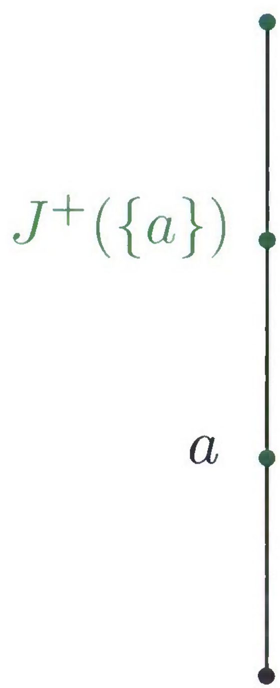


C

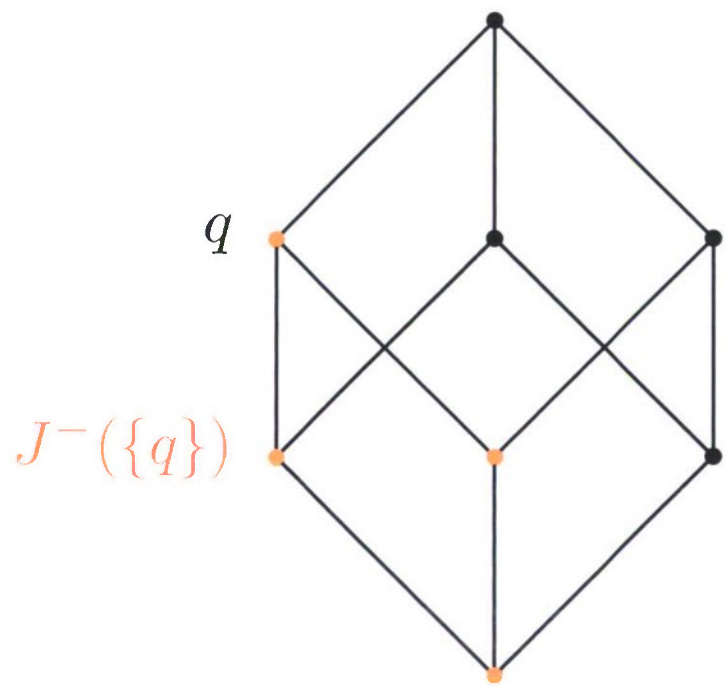
Up- and down-sets

For $Q \subseteq P$, the up-set of Q, $J^+(Q) = \{x \in P \mid \exists q \in Q,\, q \preceq x\}$.

Similarly, the down-set of Q, $J^-(Q) = \{x \in P \mid \exists q \in Q,\, x \preceq q\}$.



4



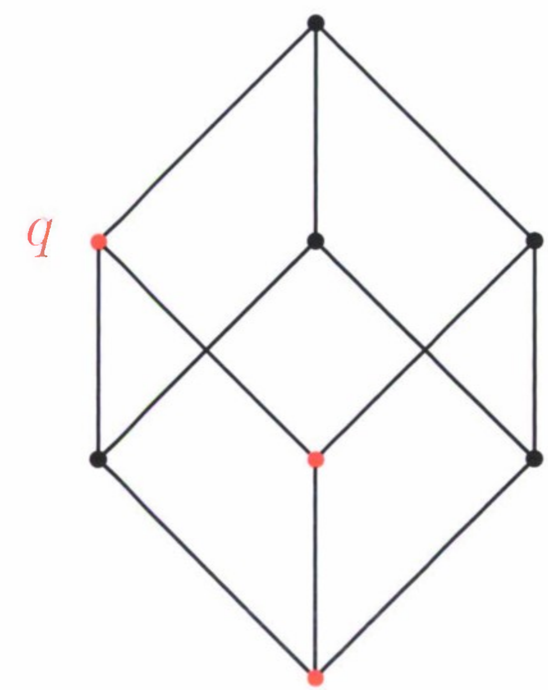
B_3

Height

For $p \in P$, the height of p, $h(p)$, is the length of a longest (maximal) chain in $J^{-}(\{p\})$.

(If all the maximal chains in a poset P have the same length, k, P is a graded poset of length k.)

$h(a) = 2$



$h(q) = 3$

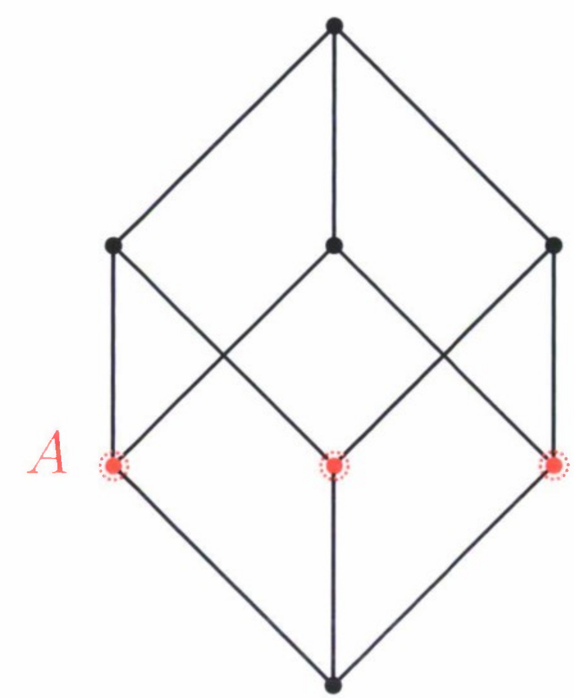
Open covers of antichains

Let $A \subseteq P$ be an antichain.

For $y \in P$, let $U_y = \{a \in A \mid a \preceq y \vee y \preceq a\}$.

Any $Y \subseteq P$ defines an open cover of A: $\mathcal{U}_Y = \{U_y \mid y \in Y\}$.

4



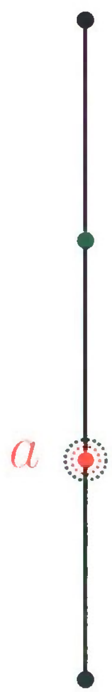
B_3

Open covers of antichains

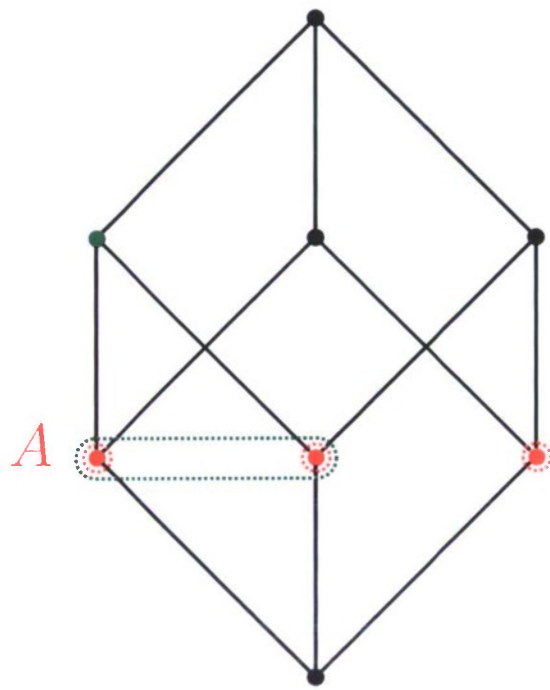
Let $A \subseteq P$ be an antichain.

For $y \in P$, let $U_y = \{a \in A \mid a \preceq y \vee y \preceq a\}$.

Any $Y \subseteq P$ defines an open cover of A: $\mathcal{U}_Y = \{U_y \mid y \in Y\}$.



4



B_3

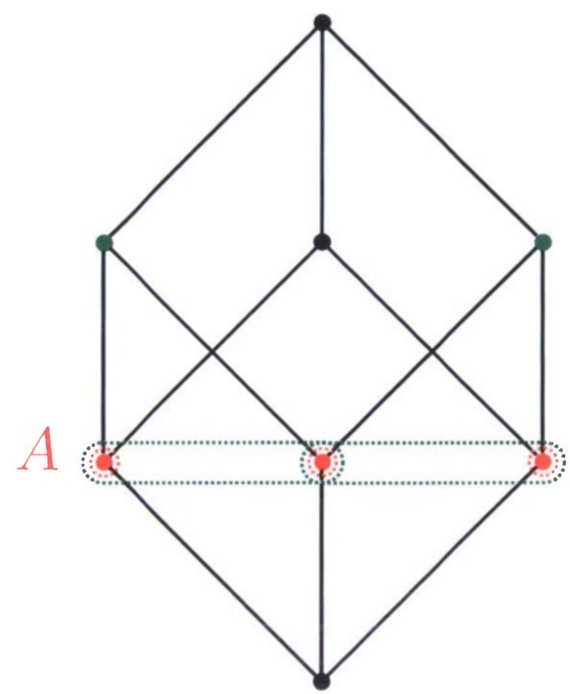
Open covers of antichains

Let $A \subseteq P$ be an antichain.

For $y \in P$, let $U_y = \{a \in A \mid a \preceq y \vee y \preceq a\}$.

Any $Y \subseteq P$ defines an open cover of A: $\mathcal{U}_Y = \{U_y \mid y \in Y\}$.

4



B_3

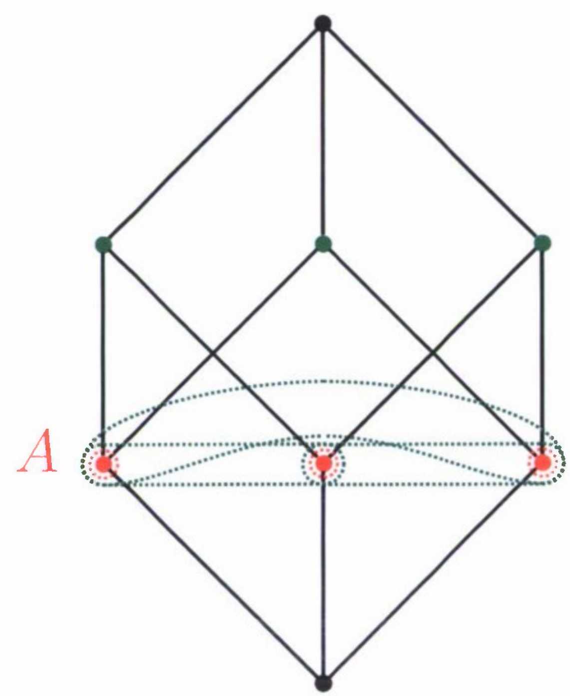
Open covers of antichains

Let $A \subseteq P$ be an antichain.

For $y \in P$, let $U_y = \{a \in A \mid a \preceq y \vee y \preceq a\}$.

Any $Y \subseteq P$ defines an open cover of A: $\mathcal{U}_Y = \{U_y \mid y \in Y\}$.

4



B_3

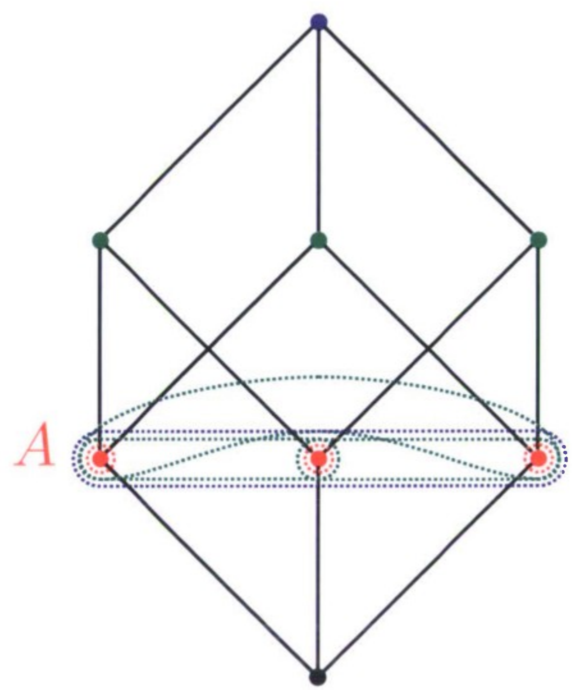
Open covers of antichains

Let $A \subseteq P$ be an antichain.

For $y \in P$, let $U_y = \{a \in A \mid a \preceq y \vee y \preceq a\}$.

Any $Y \subseteq P$ defines an open cover of A: $\mathcal{U}_Y = \{U_y \mid y \in Y\}$.

4



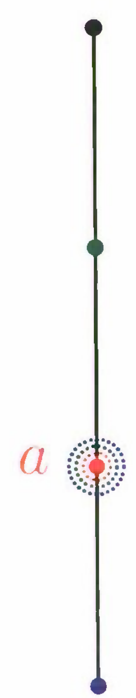
B_3

Open covers of antichains

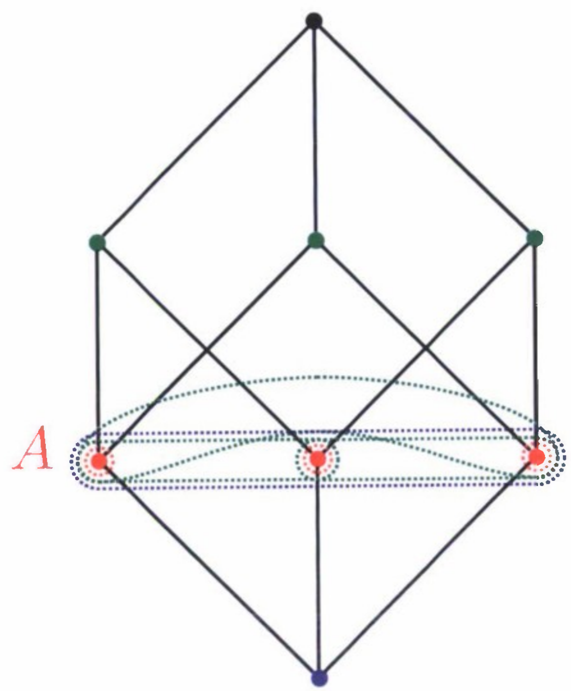
Let $A \subseteq P$ be an antichain.

For $y \in P$, let $U_y = \{a \in A \mid a \preceq y \vee y \preceq a\}$.

Any $Y \subseteq P$ defines an open cover of A: $\mathcal{U}_Y = \{U_y \mid y \in Y\}$.



4

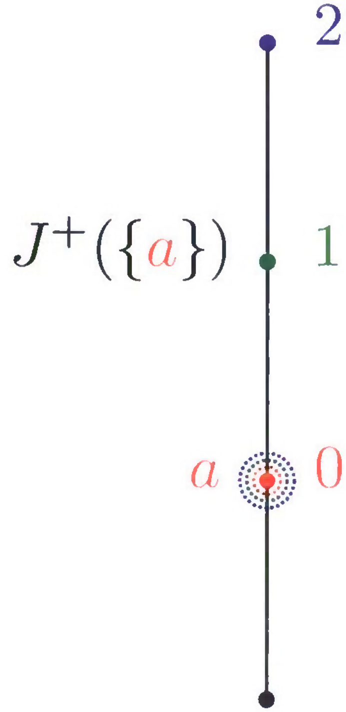


B_3

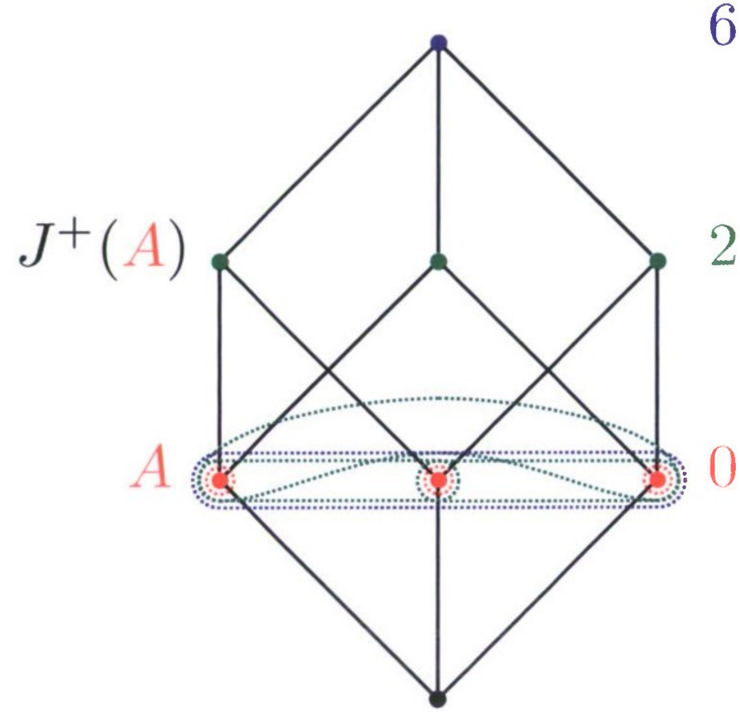
Nerves of open covers

The nerve of $\mathcal{U}_Y$ is the simplicial (Čech) complex, $N(\mathcal{U}_Y)$, with k-simplices the subsets $\{y_0, \ldots, y_k\} \subseteq Y$, for which $\bigcap_{i=0}^{k} U_{y_i} \neq \emptyset$.

Let $N_V^v(A)$ be the nerve of the open cover $\mathcal{U}_Y$ of A defined by $A \subseteq Y = \left\{y \in J^+(A) \mid |J^-(\{y\}) \cap J^+(A)| \leq v+1\right\} \subseteq J^+(A)$.



4

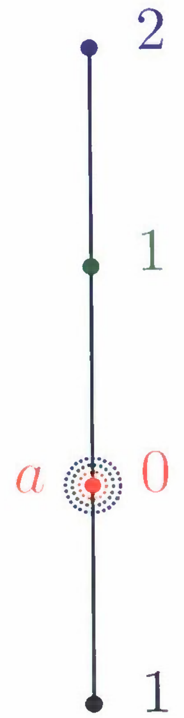


B_3

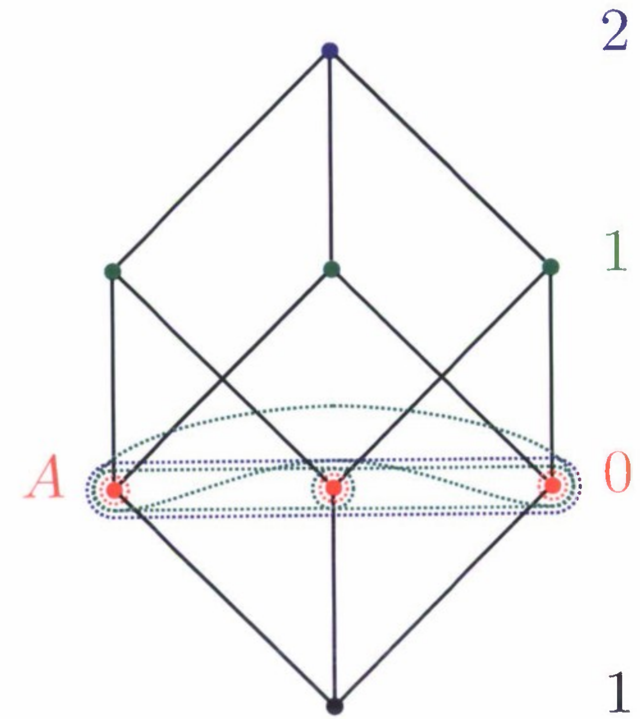
Nerves of open covers

The nerve of $\mathcal{U}_Y$ is the simplicial (Čech) complex, $N(\mathcal{U}_Y)$, with k-simplices the subsets $\{y_0, \ldots, y_k\} \subseteq Y$, for which $\bigcap_{i=0}^{k} U_{y_i} \neq \emptyset$.

When $h(a) = h_0$ for all $a \in A$, let $N_H^h(A)$ be the nerve of the open cover $\mathcal{U}_Y$ of A defined by $A \subseteq Y = \{y \in P \mid |h(y) - h_0| \leq h\} \subseteq P$.



4



B_3

Filtrations

Let $0 = v_0 < v_1 < \cdots < v_m = \max_{y \in J^+(A)} |J^-(\{y\}) \cap J^+(A)| - 1$.

$N_V^{v_0}(A) \subseteq \cdots \subseteq N_V^{v_m}(A) = N(\mathcal{U}_{J^+(A)})$ is a filtration of $N(\mathcal{U}_{J^+(A)})$.

Similarly, if A is a constant height h_0 antichain, let

$$m = \max\{h_0, \max_{y \in P} h(y) - h_0\}.$$

$N_H^0(A) \subseteq \cdots \subseteq N_H^m(A) = N(\mathcal{U}_P)$ is a filtration of $N(\mathcal{U}_P)$.

Persistence

As in the metric case, the filtered complexes $N(\mathcal{U}_{J^+(A)})$ and $N(\mathcal{U}_P)$, with inclusion maps between simplices, give persistence complexes, with persistent homology.

Also as in the metric situation, it is more convenient for calculations to define the filtered Rips complexes:

> $R^v_V(A)$ with k-simplices $\{a_0, \ldots, a_k\} \subseteq A$ for which $\exists y \in J^+(A)$ with $|J^-(\{y\}) \cap J^+(A)| \leq v + 1$ and $\{a_0, \ldots, a_k\} \subseteq J^-(\{y\})$;
>
> $R^h_H(A)$ with k-simplices $\{a_0, \ldots, a_k\} \subseteq A$ for which $\exists y \in P$ with $|h(p) - h_0| \leq h$ and $\{a_0, \ldots, a_k\} \subseteq J^-(\{y\}) \cup J^+(\{y\})$.

Unlike the metric situation, the nerve complexes are less convenient for calculations because they have vertex set $J^+(A)$ or P, both larger than A, the vertex set for the Rips complexes.

Binomial examples

Test results

Consider a set of concepts C, understanding of which a set of test questions T is intended to measure. Specifically, suppose each test question is designed to test understanding of a particular concept, so there is a map $c : T \to C$.

Let us make the simplifying assumption that each student $s \in S$ taking the test either understands or does not understand each concept, so there is a map $u : S \to \mathcal{P}(C)$.

Test results

Finally, suppose that when a student understands a concept s/he is likely to get each of the test questions that test that concept right; and when s/he does not understand a concept, s/he is unlikely to get each of the test questions that test that concept right. More precisely, let

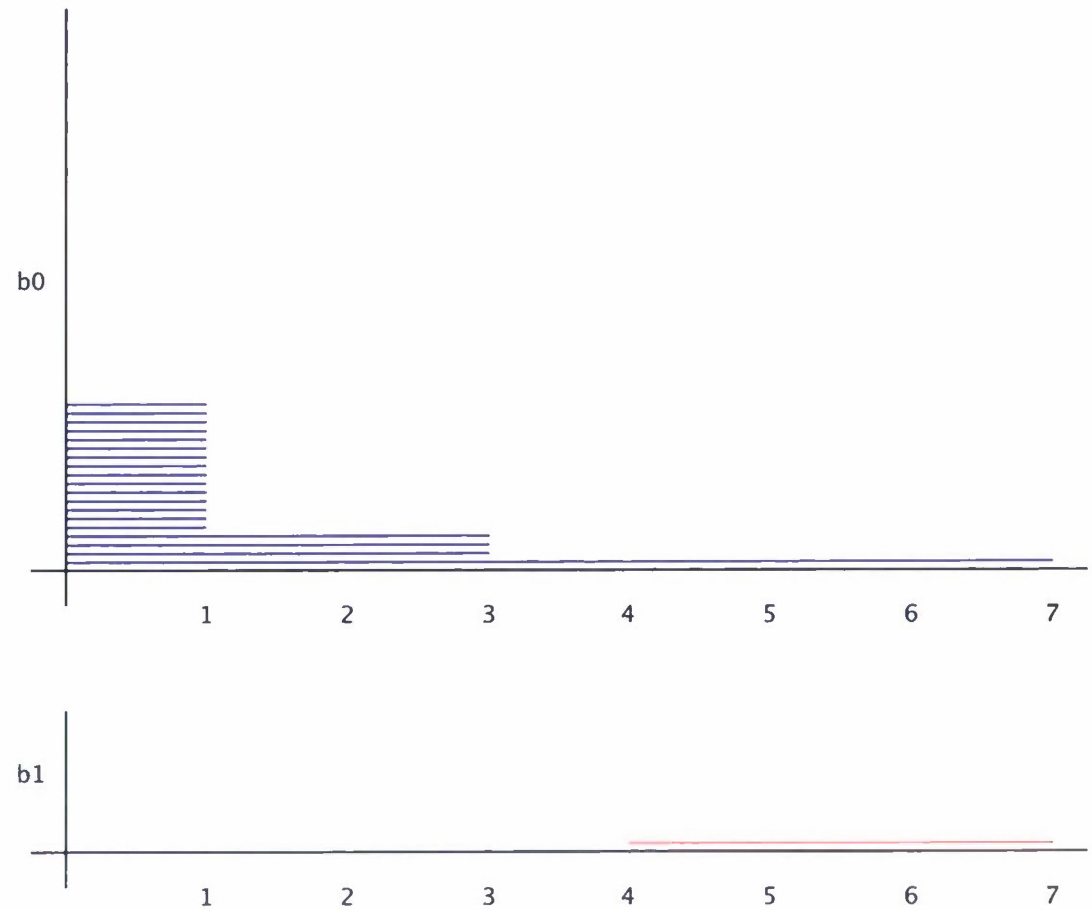
$$\Pr(t \in R_s) = \begin{cases} p & \text{if } c(t) \in u(s); \\ q & \text{otherwise,} \end{cases} \tag{$*$}$$

where p is relatively large, say 0.9, and q is relatively small, say 0.2 if we are imagining a multiple choice test.

For any population of students, *i.e.*, a sample $\{u(s) \mid s \in S\}$ from some probability distribution over $\mathcal{P}(C)$, we can simulate test results using (*), and obtain a multiset $\left(\{\hat{R}_s \mid s \in S\}, n\right)$, where n maps $\{\hat{R}_s \mid s \in S\} \to \mathbb{N}$ and $n(R) = |\{s \in S \mid \hat{R}_s = R\}|$.

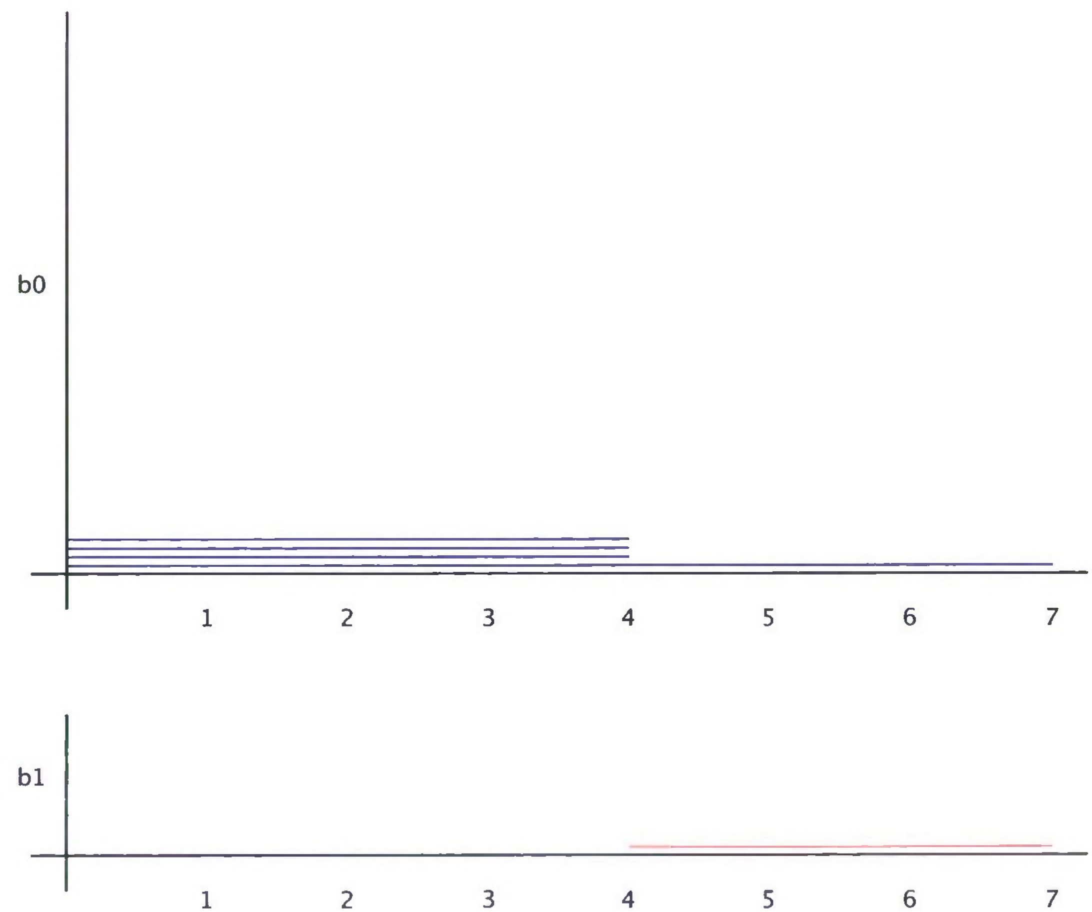
20 questions testing 4 concepts

We simulate $|S| = 100,000$ and filter the resulting 'point cloud' to include only multiplicities at least 20, *i.e.*, approximately discarding results with probability less than 0.0002. Height 1 antichain:



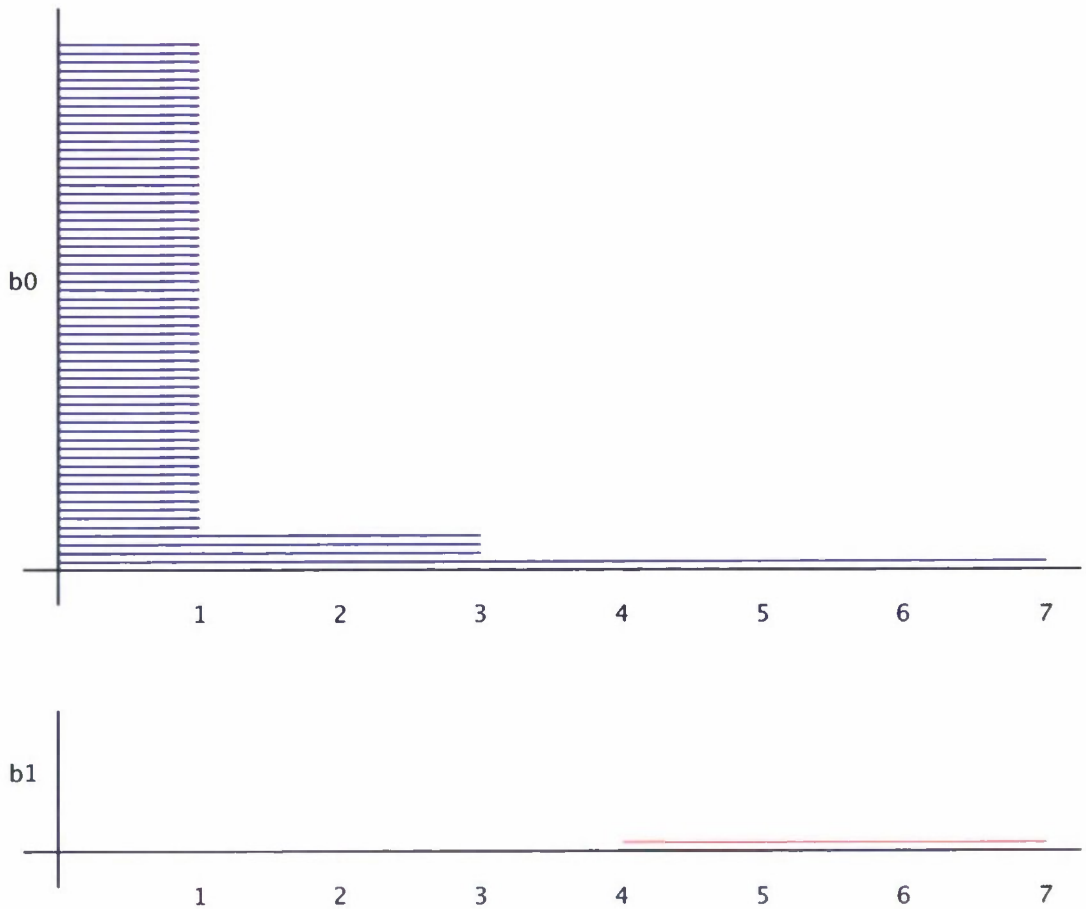
20 questions testing 4 concepts

We simulate $|S| = 100,000$ and filter the resulting 'point cloud' to include only multiplicities at least 20, *i.e.*, approximately discarding results with probability less than 0.0002. Height 2 antichain:



20 questions testing 4 concepts

We simulate $|S| = 100,000$ and filter the resulting 'point cloud' to include only multiplicities at least 20, *i.e.*, approximately discarding results with probability less than 0.0002. Height 3 antichain:



20 questions testing 4 concepts

At each height there is a persistent 1-cycle.

This reflects the structure of the test, namely the map $c : T \to C = \{0, 1, 2, 3\}$, for which we chose $|c^{-1}(i)| = 5$, $i \in C$,

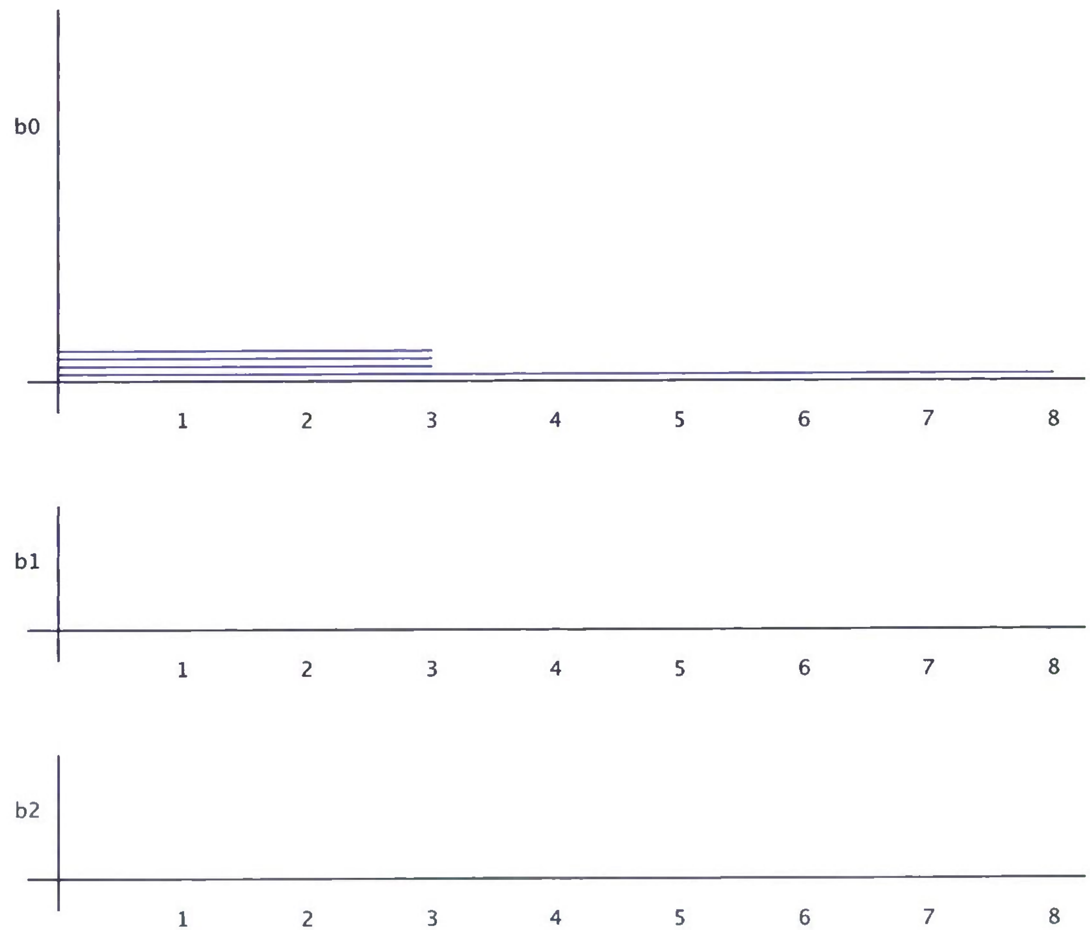
and the probability distribution over $\mathcal{P}(C)$ from which we sampled, students who have learned either one concept i or two concepts $(i, i+1) \pmod 4$; all with equal probability.

20 questions testing 4 concepts

For the same test, we can consider other probability distributions over $\mathcal{P}(C)$, for example, students who have learned either one concept, or two concepts, or three concepts; all with equal probability.

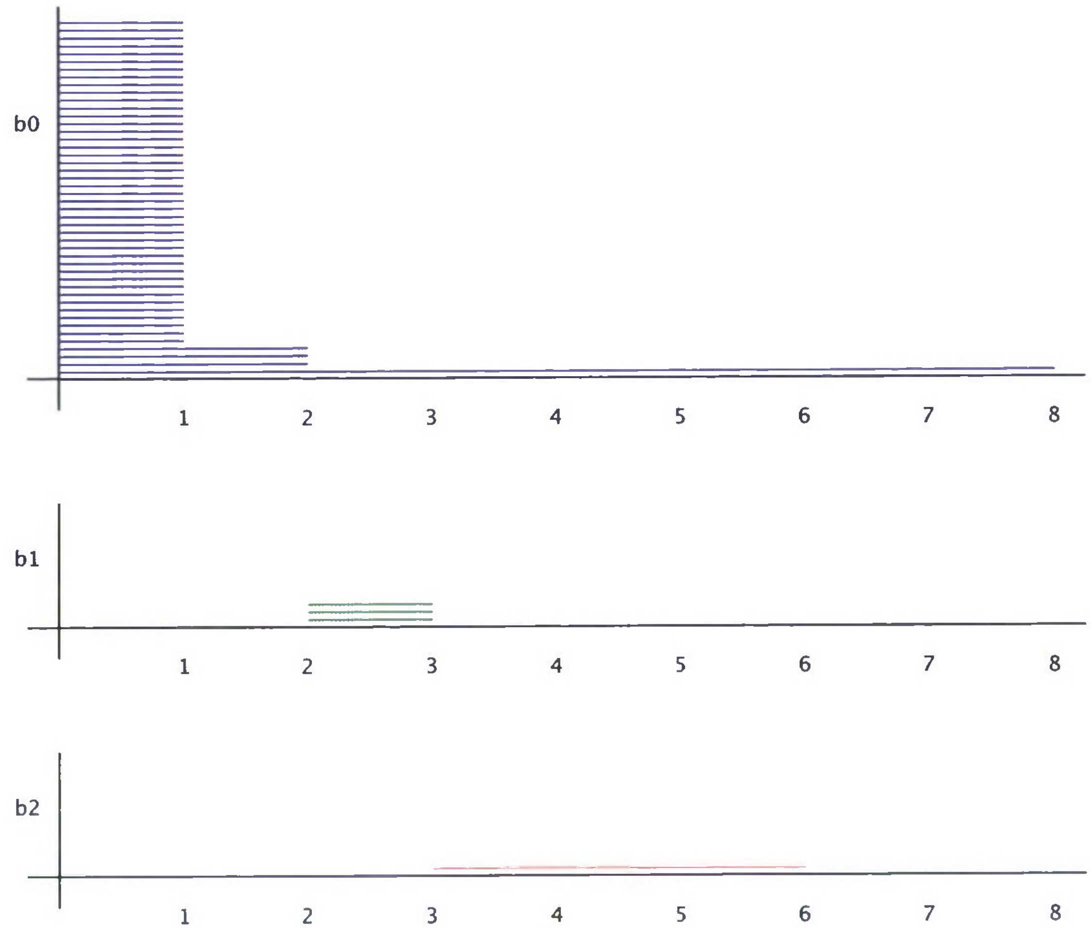
20 questions testing 4 concepts

We simulate $|S| = 100,000$ and filter the resulting 'point cloud' to include only multiplicities at least 30, *i.e.*, approximately discarding results with probability less than 0.0003. Height 1 antichain:



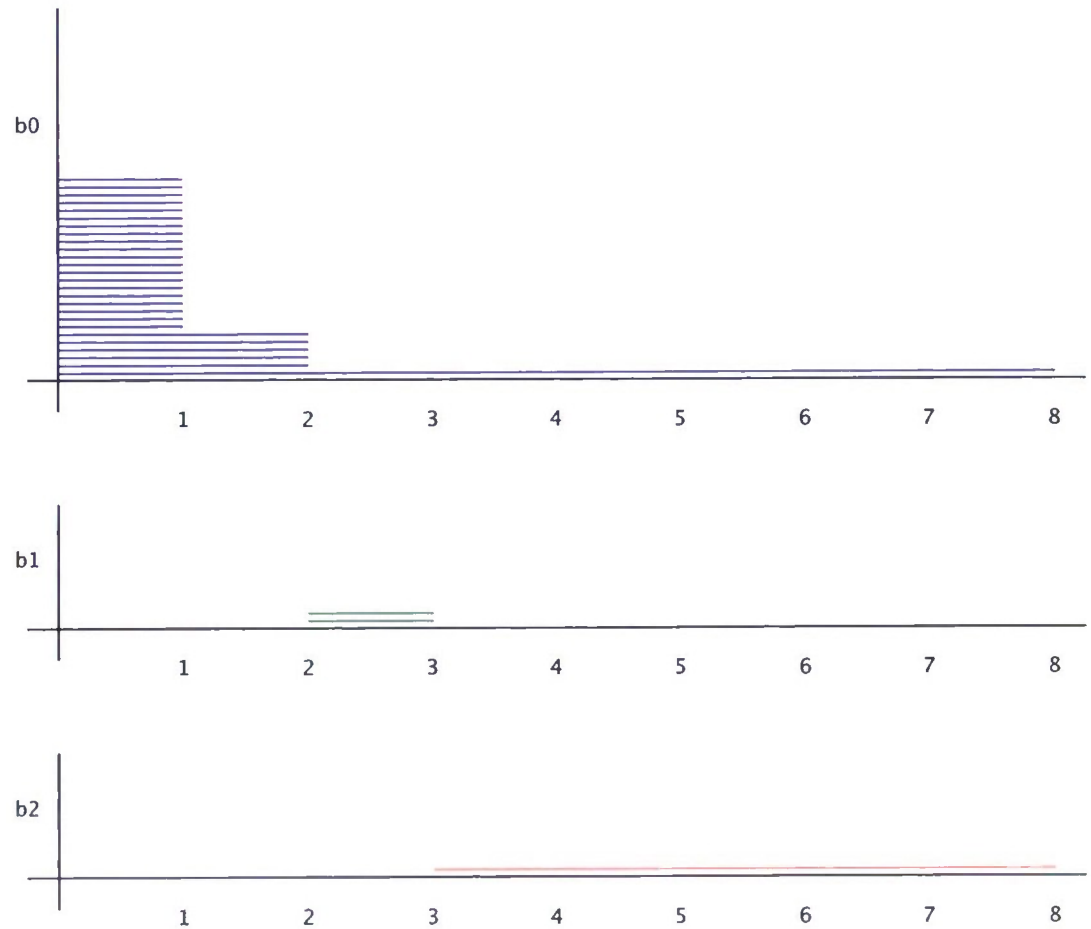
20 questions testing 4 concepts

We simulate $|S| = 100,000$ and filter the resulting 'point cloud' to include only multiplicities at least 30, *i.e.*, approximately discarding results with probability less than 0.0003. Height 2 antichain:



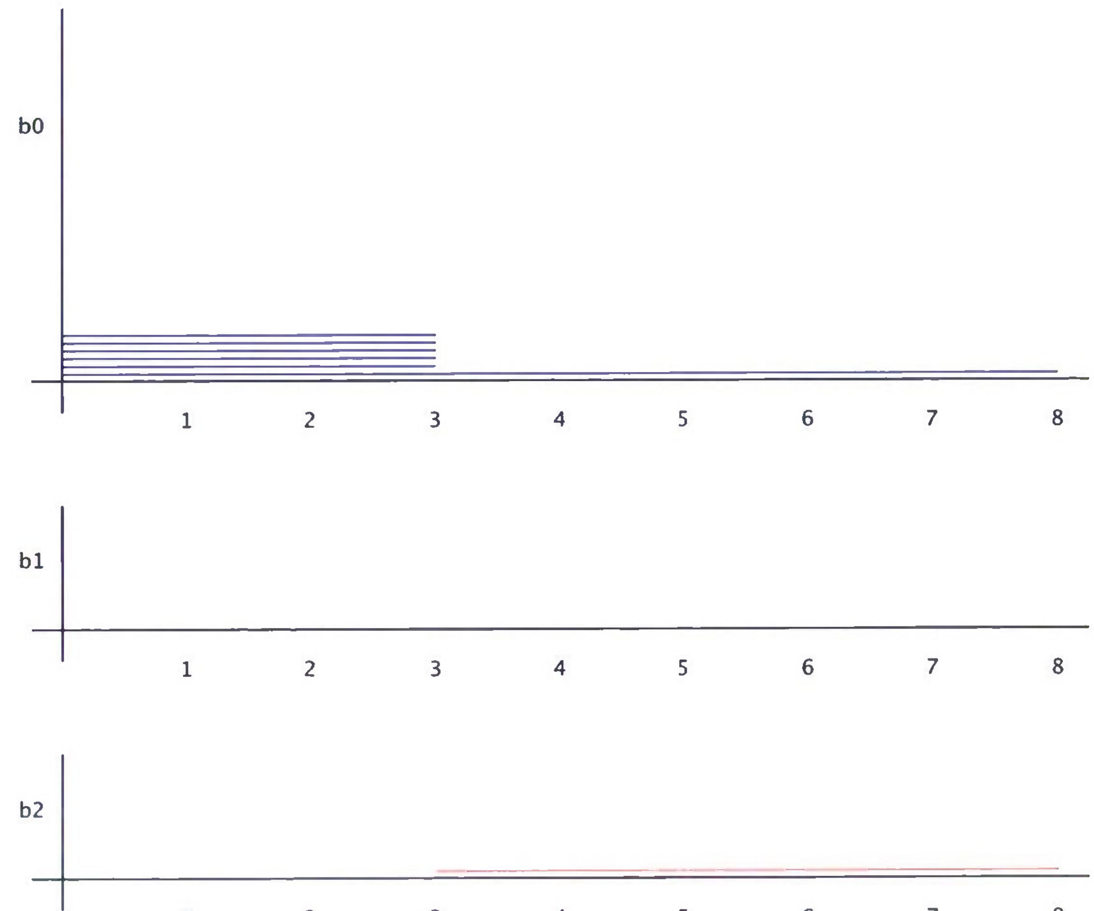
20 questions testing 4 concepts

We simulate $|S| = 100,000$ and filter the resulting 'point cloud' to include only multiplicities at least 30, *i.e.*, approximately discarding results with probability less than 0.0003. Height 3 antichain:



20 questions testing 4 concepts

We simulate $|S| = 100,000$ and filter the resulting 'point cloud' to include only multiplicities at least 30, *i.e.*, approximately discarding results with probability less than 0.0003. Height 4 antichain:



20 questions testing 4 concepts

At heights above 1 there is a persistent 2-cycle, but no 1-cycles, *i.e.*, a 2-sphere.

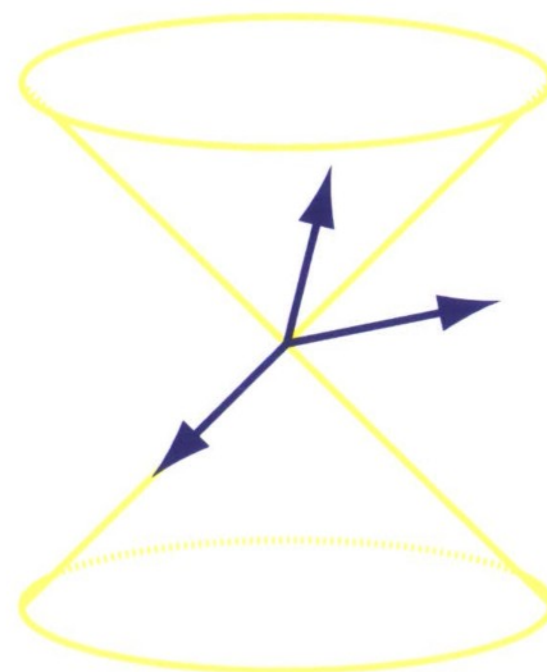
Again, this is determined by the probability distribution over $\mathcal{P}(C)$ from which we sampled, students who have learned either one concept, or two concepts, or three concepts, all with equal probability.

Geometric examples

Lorentzian manifolds

A Lorentzian manifold is a smooth manifold M with a metric tensor $g : T_p(M) \times T_p(M) \to \mathbb{R}$ of signature $(-,+,\ldots,+)$.

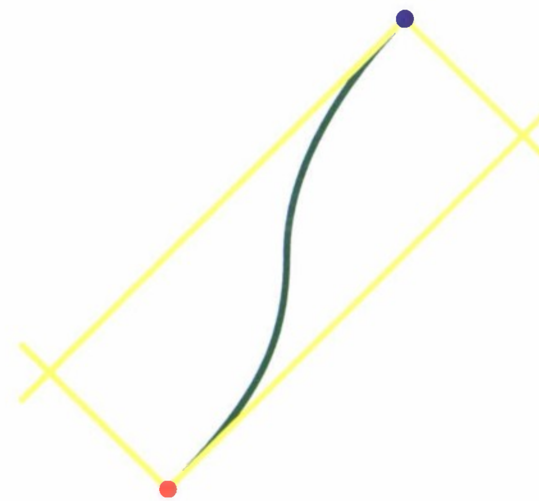
A tangent vector $v \in T_p(M)$ is timelike, null, or spacelike if $g(v,v)$ is negative, zero, or positive, respectively.



For v, w timelike, let $v \sim w$ if $g(v,w) < 0$; this is an equivalence relation with two classes. One is designated future-directed; this designation extends to null vectors in the closure of the set of future-directed timelike vectors.

Posets from Lorentzian manifolds

A C^0, piecewise C^1 curve $\gamma : I \to M$ is causal if every tangent vector to γ is timelike or null, and future-directed.



For $x, y \in M$, x causally precedes y, $x \preceq y$, if there is a causal curve such that $\gamma(0) = x$ and $\gamma(1) = y$.

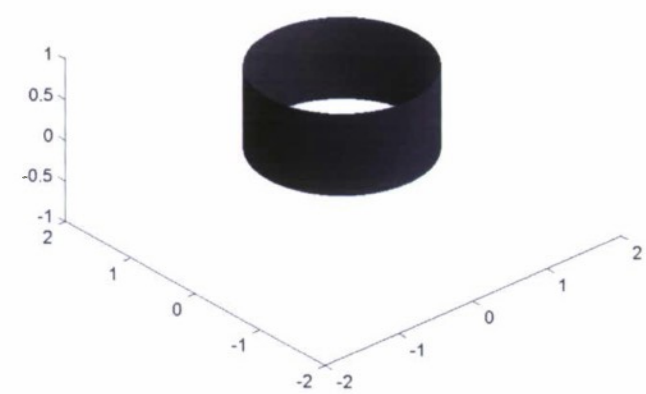
'Causally precedes' is a partial order on M.

The metric tensor defines a volume form $\sqrt{|\det(g)|}\,\mathrm{d}x^0 \wedge \mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^d$ on M, with respect to which stochastic point processes are defined.

So consider poset data obtained by random sampling from a Lorentzian manifold, retaining only the causal ordering.

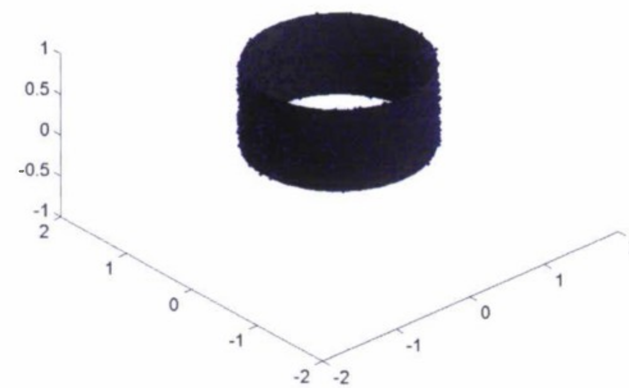
A Lorentzian cylinder

Let $M = I \times S^1$, embedded in $2+1$ dimensional Minkowski space.



A Lorentzian cylinder

Let $M = I \times S^1$, embedded in $2+1$ dimensional Minkowski space.

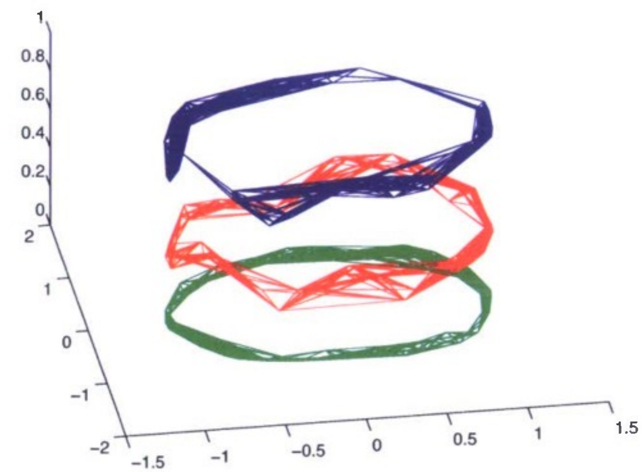


Sample points uniformly at random relative to the volume form; keep $n = 2000$ points which are close to the cylinder.

The causal order on these points defines a poset P of size n.

A Lorentzian cylinder

Let $M = I \times S^1$, embedded in $2+1$ dimensional Minkowski space.



Sample points uniformly at random relative to the volume form; keep $n = 2000$ points which are close to the cylinder.

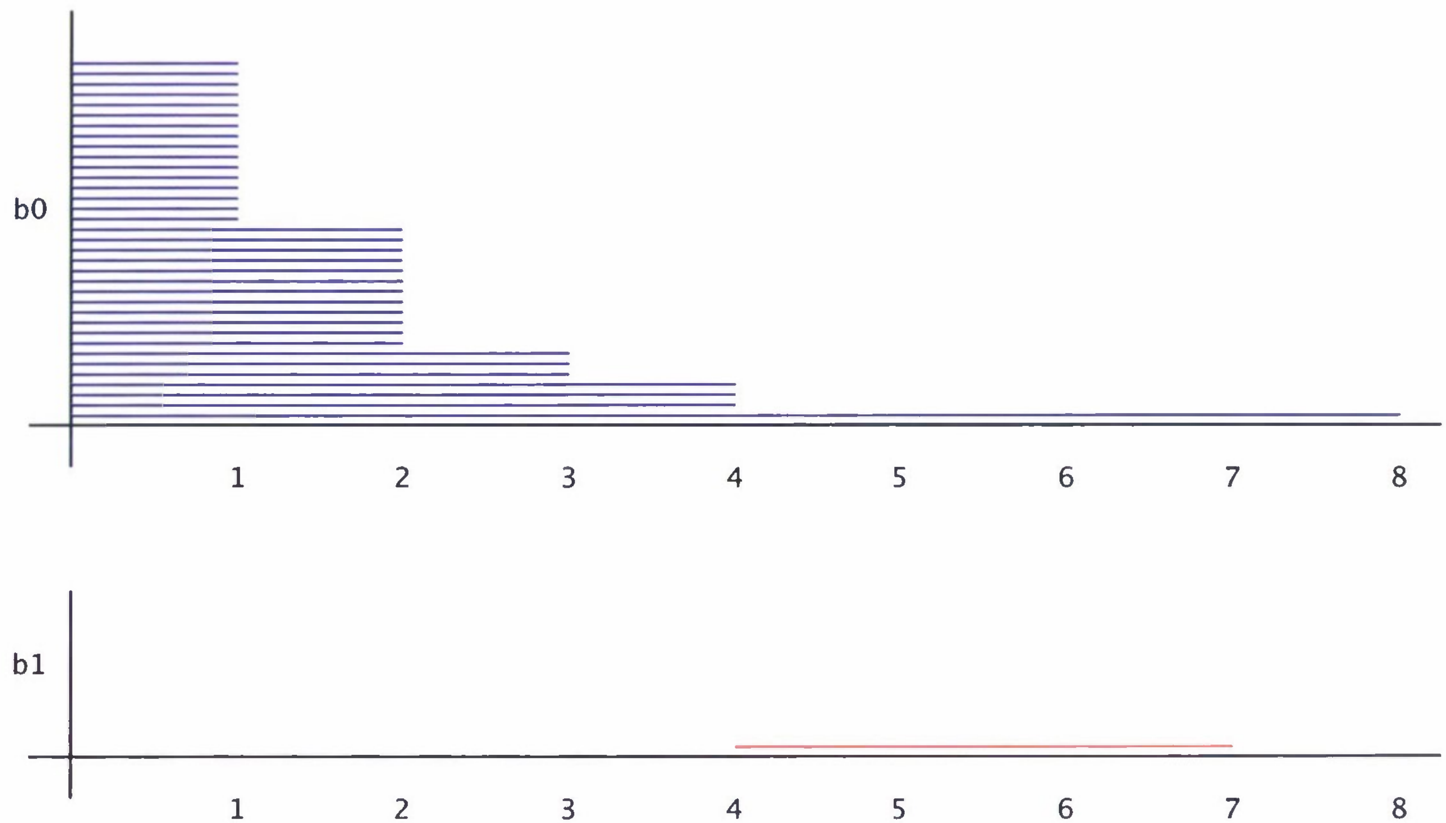
The causal order on these points defines a poset P of size n.

Let $A_h = \{a \in P \mid h(a) = h\}$ be the antichain of height h elements in P; A approximates a spacelike curve in M.

Compute persistent homology for the filtered Rips complex $R(\mathcal{U}_P)$ on A_h.

Persistent homology for Lorentzian cylinder poset data

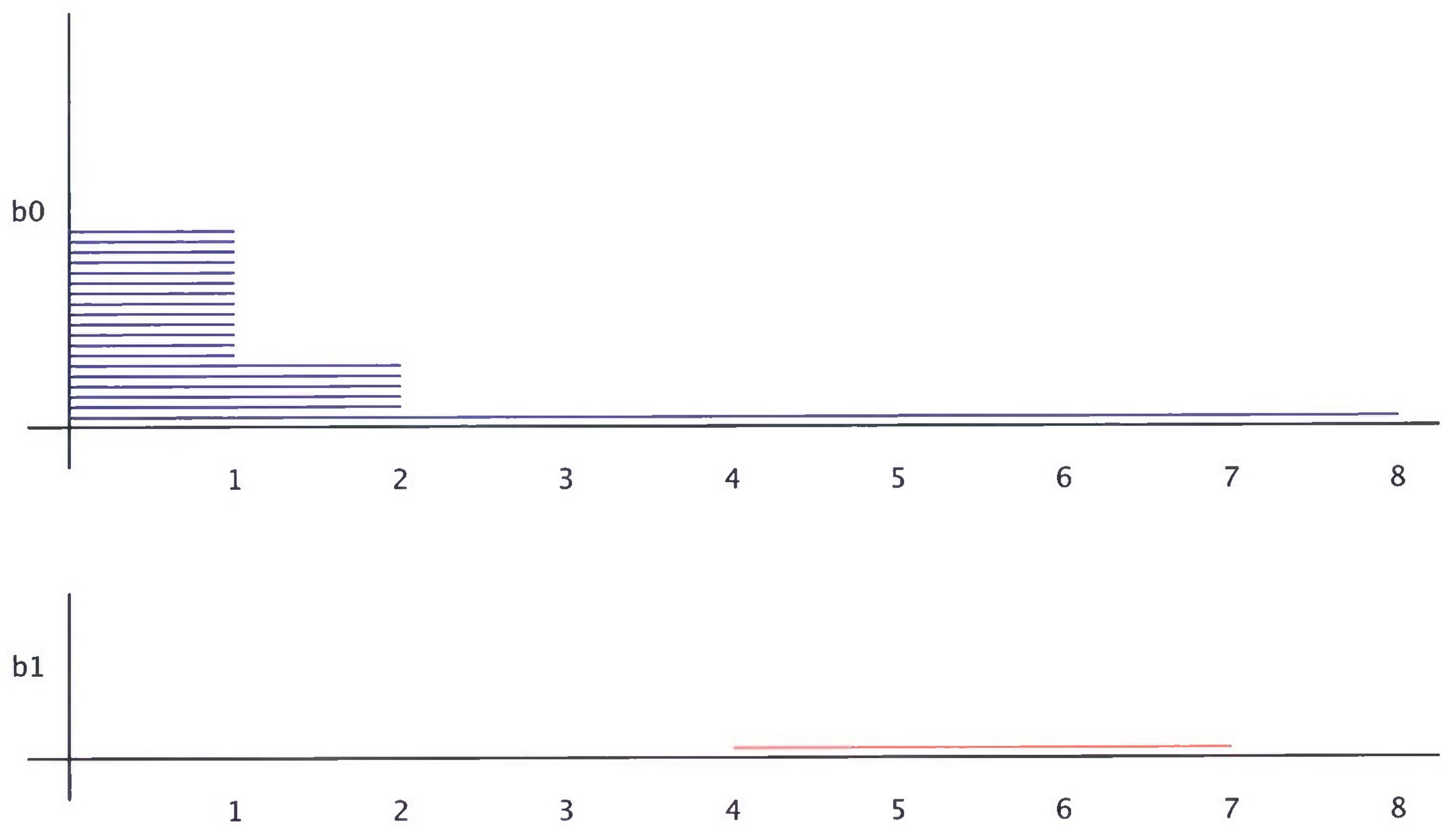
The antichain at height 1:



A 1-cycle is created, and then persists until it gets filled in as past lightcones cover a little more than $1/3$ of the spacelike S^1.

Persistent homology for Lorentzian cylinder poset data

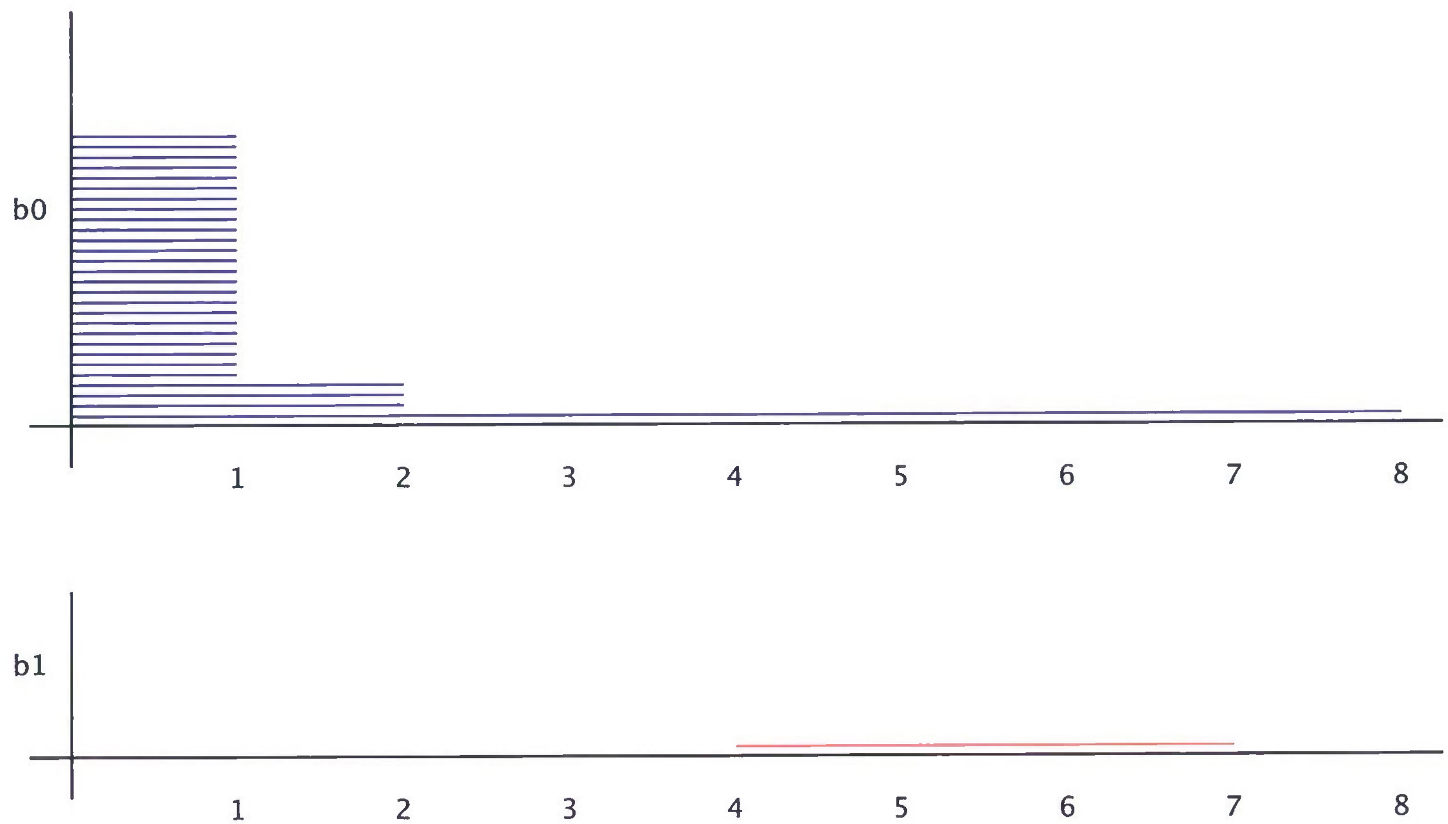
The antichain at height 20:



A 1-cycle is created, and then persists until it gets filled in as past or future lightcones cover a little more than $1/3$ of the spacelike S^1.

Persistent homology for Lorentzian cylinder poset data

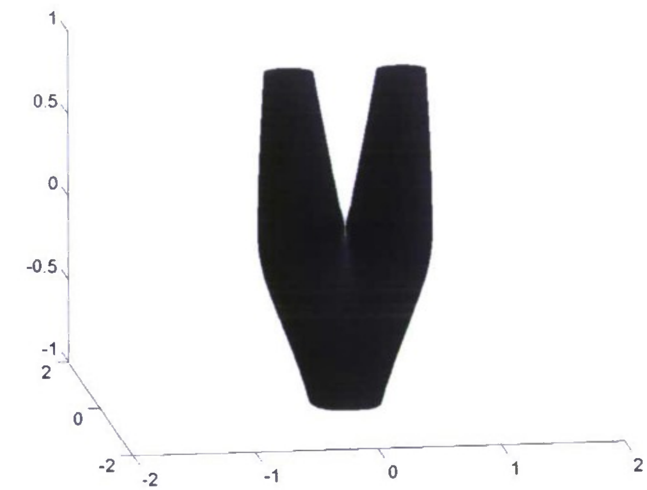
The antichain at height 50:



A 1-cycle is created, and then persists until it gets filled in as future lightcones cover a little more than $1/3$ of the spacelike S^1.

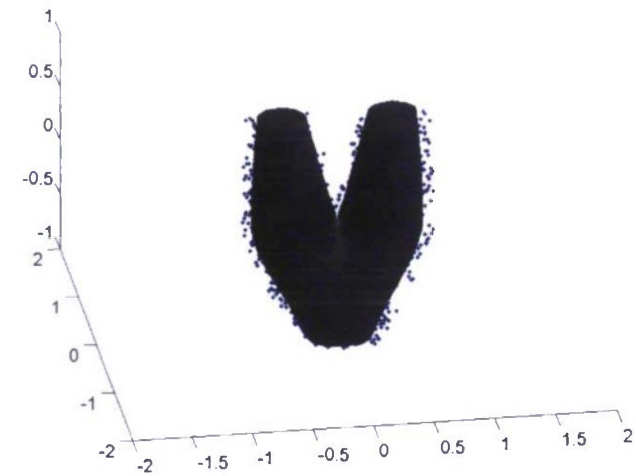
A Lorentzian pair of pants

Let M be a 3-punctured sphere embedded in $2+1$ dimensional Minkowski space so that the induced metric has signature $(-,+)$ at all but one point.



A Lorentzian pair of pants

Let M be a 3-punctured sphere embedded in $2+1$ dimensional Minkowski space so that the induced metric has signature $(-,+)$ at all but one point.

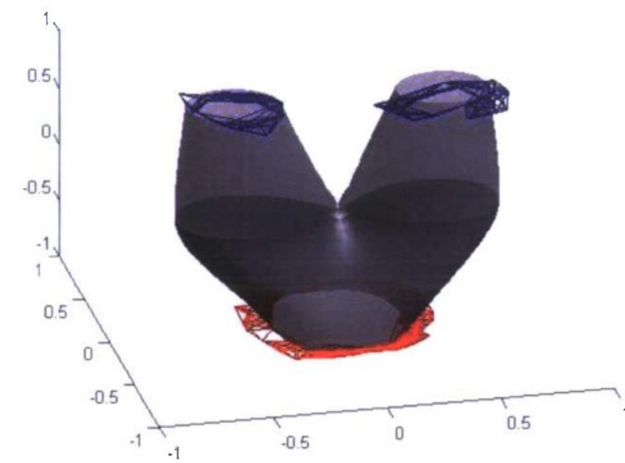


Sample points uniformly at random relative to the volume form; keep $n = 5000$ points which are close to the pants.

The causal order on these points defines a poset P of size n.

A Lorentzian pair of pants

Let M be a 3-punctured sphere embedded in $2+1$ dimensional Minkowski space so that the induced metric has signature $(-,+)$ at all but one point.



Sample points uniformly at random relative to the volume form; keep $n = 5000$ points which are close to the pants.

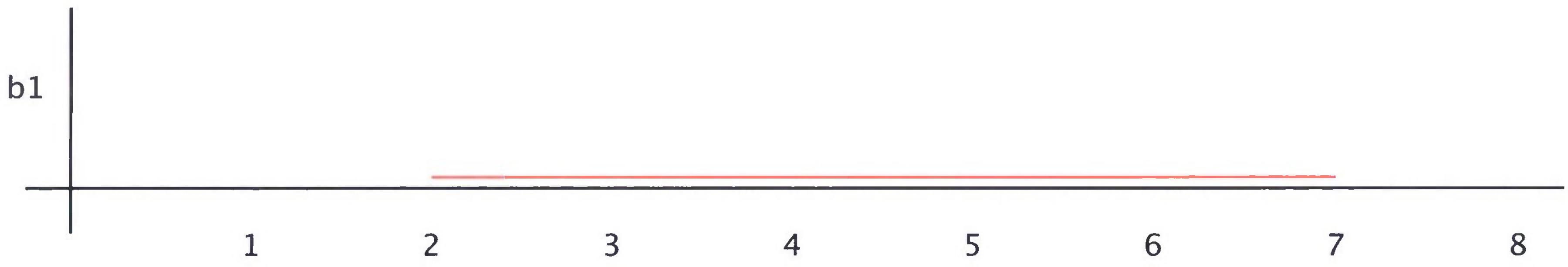
The causal order on these points defines a poset P of size n.

Let $A_h = \{a \in P \mid h(a) = h\}$ be the antichain of height h elements in P; A approximates a spacelike curve in M.

Compute persistent homology for the filtered Rips complex $R(\mathcal{U}_P)$ on A_h.

Persistent homology for Lorentzian pants poset data

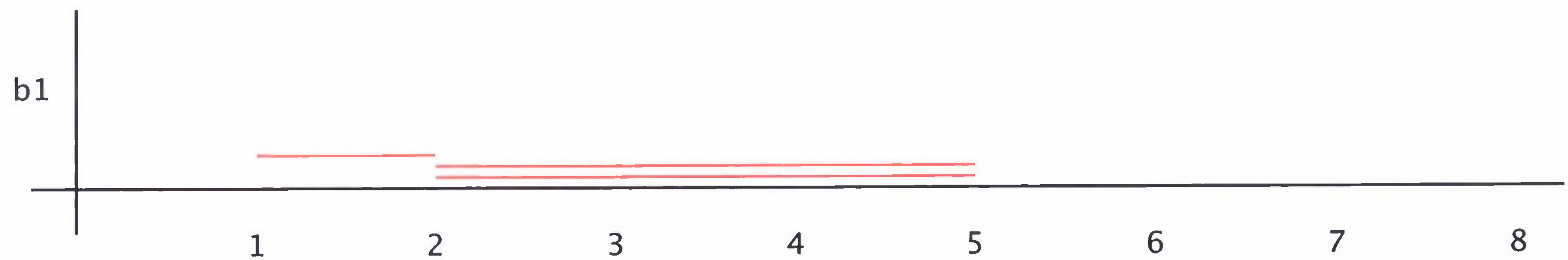
The antichain at height 1:



A 1-cycle is created, and then persists until it gets filled in as past or future lightcones cover a little more than $1/3$ of the spacelike S^1.

Persistent homology for Lorentzian pants poset data

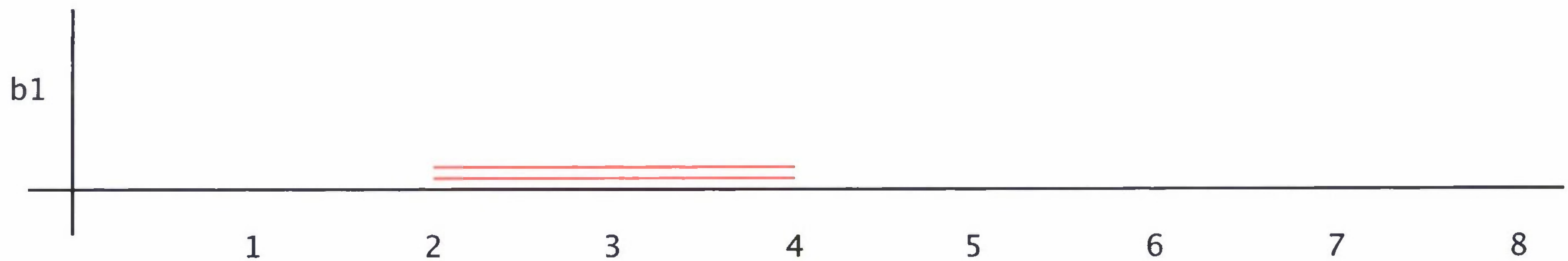
The antichain at height 20:



Two 1-cycles are created, and then persist until they are filled in as past or future lightcones cover a little more than $1/3$ of each spacelike S^1.

Persistent homology for Lorentzian pants poset data

The antichain at height 40:



Two 1-cycles are created, and then persist until they are filled in as past or future lightcones cover a little more than $1/3$ of each spacelike S^1.

onclusion

Summary

Persistent topology depends on a filtered family of open sets. These open sets need not come from increasing a distance in a metric space; they can also be constructed for partially ordered data.

There are multiple constructions giving filtered simplicial complexes for posets.

Certain data sets are naturally partially ordered.

For the binomial partially ordered data considered, the persistent topology of the constant height antichains was correct.

For the geometric partially ordered data considered, the persistent topology of the constant height antichains was correct, even when it changed with height.

Directions

Application to real data sets.

There are other simplicial complexes associated with posets—the order complex and its generalizations—which should be analyzed.

For all of these, developing the poset analogues of landmarks and witness complexes would make computations more tractable.

There are other naturally partially ordered datasets, *e.g.*, symmetric or Hermitian matrix data, partially ordered by eigenvalues; these should be explored.

Abstract

In this paper, we study the topological properties of posets using the tool of persistent homology. We introduce a construction, given a poset, which results in a filtered simplicial complex. We investigate this construction on posets sampled from Lorentzian manifolds.

1 Introduction

2 Preliminaries

We now remind the reader of some elementary facts about posets. Let P be a finite set throughout.

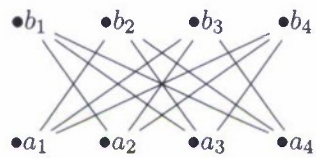
Definition 2.1. A *partially ordered set* or *poset* $(P, \preceq)$ is a set P with a partial order $\preceq$. If two elements $a, b \in P$ are not related, we write $a \not\sim b$.

A partial order is a binary relation satisfying reflexivity, antisymmetry, and transitivity. Throughout this paper we make use of special subsets of posets in which no two elements are related.

Definition 2.2. An *antichain* A is a subset of a poset $(P, \preceq)$ such that the following holds:

- For any two $a, b \in A$, $a \not\sim b$.

Example: Let $n \in \mathbb{N}$. Let S_n denote the following poset. The elements of S_n are the a_i and b_i for $1 \leq i \leq n$ with the relations $a_i \leq b_j$ for $i \neq j$.



The sets of $\{a_1, \ldots, a_n\}$ and $\{b_1, \ldots, b_n\}$ are two antichains in S_n.

When discussing posets, it is often easiest to describe immediate relations and allow all other relations to be implied by transitivity.

Definition 2.3. Given a finite binary relation $(B, \preceq_B)$, the *transitive closure* $(B, \preceq_P)$ is the minimal transitive order which extends $\preceq_B$. That is for $a, b \in B$, $a \preceq_P b$ if there exists $c_1, \ldots, c_k \in B$ such that $a \preceq_B c_1 \preceq_B c_2 \preceq_B \cdots \preceq_B c_k \preceq_B b$.

The transitive closure of a partially ordered set is the partially ordered set itself. Given an antichain A in a poset P, we can define a height function which measures the distance to the given antichain. We will denote a path from p to q by elements $p_0, \ldots, p_n \in P$ where $p = p_0 \preceq p_1 \preceq \cdots \preceq p_n = q$. The length of the path, $\ell(p_0, \ldots, p_n)$, is n. Let

$$\tilde{d}(p, q) = \min_{\text{paths } (p=p_0,\ldots,p_n=q) \in TP} \ell(p_0, \ldots, p_n)$$

If there are no paths from p to q, this distance is ∞. Finally we define our distance function as

$$d(p, q) = \min(\tilde{d}(p, q), \tilde{d}(q, p))$$

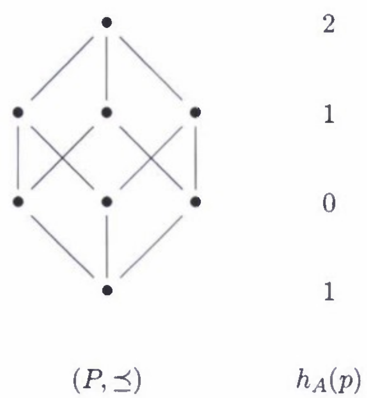
Now we can define our height function. Let A be an inextendable antichain. Define $h_A : P \to \mathbb{Z}$ as follows:

$$h_A(p) = \min\{d(a, p) : a \in A\}$$

This function is well defined because A is inextendable. Suppose $h_A(p) = \infty$ for some $p \in P$. Then there are no paths starting in A and ending at p, nor paths starting at p and ending at A. Thus p is incomparable with all elements of A. So A can be extended by p. Two noted inextendable antichains are the sets A_m and A_M of minimal and maximal elements of the poset respectively. Let h_m and h_M denote the corresponding height functions.

Proposition 2.4. *$A_j = \{p \in P | h_m(p) = j\} \cup \{p \in P | h_m(p) \leq j$ and p is maximal $\}$ is an inextendable antichain in the poset for any $j \in \mathbb{Z}$.*

Example:



3 Construction

We now discuss the construction of a simplicial complex given a poset and antichain pair. Let *FinSimp* denote the category of simplicial complexes and simplicial maps. Let *PA* denote the category of pairs (P, A) of finite posets P and antichains $A \subset P$. For objects $(P, A), (R, B) \in PA$, $Hom((P, A), (R, B))$ is the set of order preserving maps $f : P \to R$ with $f(A) \subset B$. Given an object (P, A) of PA, with $A = \{a_1, \ldots, a_N\}$, one can construct an abstract simplicial complex. The simplicial complex will have vertices $a_1, \ldots, a_N$. What we need is a rule dictating whether or not to assigning a k-simplex to $k + 1$ points. This assignment will be determined by other points in the poset.

Definition 3.1. Let $p \in P$. Define $I(p) = \{q \in P | p \preceq q$ or $q \preceq p\}$. Define $A_p = I(p) \cap A$.

The set $I(p)$ is the set of elements related to p in the poset. Each A_p is the set of elements in A related to p, which gives us information about the relative closeness of elements in the antichain. That is, two elements of A which appear together in some A_p are closer together than two elements of A which do not. We use this information to construct a functor $\Delta : PA \to FinSimp$. Given a poset antichain pair (P, A), the vertices of the simplicial complex $\Delta(P, A)$ are $\{a_1, \ldots, a_N\}$, the elements of the antichain. The vertices $a_{i_0}, \ldots, a_{i_k}$ form a k-simplex if there is a $p \in P$ with $a_{i_0}, \ldots, a_{i_k} \preceq p$. Alternatively, if $A_p = \{a_{i_0}, \ldots, a_{i_k}\}$, we add the k-simplex $[a_{i_0}, \ldots, a_{i_k}]$ to the abstract simplicial complex $\Delta(P, A)$. This creates a well defined simplicial complex as the condition will include all subsimplices of a given simplex by definition.

Example: For the poset S_n and antichain $A = \{a_1, \ldots, a_n\}$, $\Delta(S_n, A)$ is the boundary of the standard $n - 1$ simplex. This simplicial complex is homotopy equivalent to S^{n-2}.

Lemma 3.2. *This construction only depends on the minimal and maximal elements of P.*

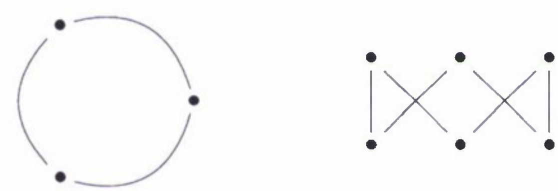
Proof. If $p \in P$ is neither minimal nor maximal, then there is a $q \in P$ with $p < q$ or $q < p$. Assume $p < q$. Then $A_p \subset A_q$. Thus the simplex associated to A_p is a face of A_q. ☐

If P has a unique minimal or maximal element, the resulting complex $\Delta(P, A)$ is a $|A|$-simplex for any antichain A in P.

It is enough to define a map $f : K \to L$ of simplicial complexes by giving a map of vertices so long as the following holds. If $[v_1, \ldots, v_k]$ is a simplex in K, then $[f(v_1), \ldots, f(v_k)]$ must be a simplex in L. We now show this holds for our constructed complex. Let $f : (P, A) \to (R, B)$ be a morphism in PA, that is an order preserving map with $f(A) \subset B$. Define $\Delta(f) : \Delta(P, A) \to \Delta(R, B)$ in the following way. Let $A = \{a_1, \ldots, a_N\}$. Set $\Delta(f)(a_i) = f(a_i)$. This defines a map on the vertices of $\Delta(P, A)$. We need only check that simplices are mapped to simplices. Suppose $[a_{i_1}, \ldots, a_{i_k}]$ is a simplex in $\Delta(P, A)$. Then there is a $p \in P$ such that $\{a_{i_1}, \ldots, a_{i_k}\} \subset A_p$. Then $f(a_i) \in B_{f(p)}$ for all i, by order preservation. So the simplex $[f(a_{i_1}), \ldots, f(a_{i_k})]$ is in $\Delta(R, B)$ and the map is well defined.

There is a functor $\mathcal{F}$:*FinSimp*$\to$*PA* going the other direction, similar to Δ. Let K be a finite simplicial complex. $\mathcal{F}(K)$ is a poset P given by the following construction: Let K_i denote the set of i-dimensional simplices of K. The constructed poset will have $\sum |K_i|$ elements. By inducting on dimension, we can create a poset by defining immediate relations and taking the transitive closure. For $i = 0$, let $K_0 = \{v_0, \ldots, v_N\}$ denote the set of vertices of K. Add an element to P for each element of K_0. For $i = 1$, add an element to P for each element of K_1. Then, for each element $e_k = [v_{k_1}, v_{k_2}]$ of K_1, add relations $v_{k_1} \preceq e_k$ and $v_{k_2} \preceq e_k$. That is, all of the subsimplices of e_k are less than e_k. We then proceed inductively. For each $i \in \mathbb{N}$, we add an element to P for each element in K_i, with relations to all of its subsimplices. We set A to be the set of minimal elements of the constructed P.

Example: The circle with three vertices and three edges is mapped to the crown poset with six elements of height 2 and antichain the set of minimal elements under $\mathcal{F}$.



Proposition 3.3. $\Delta \circ \mathcal{F}$:FinSimp $\to$ FinSimp $= Id_{\text{FinSimp}}$

4 Persistence

The following definition of persistence comes from [Carlsson08]

Definition 4.1. Let $\mathcal{C}$ be a category and P be a poset. We can regard P as a category $\mathcal{P}$ in the usual way, *i.e.*, with object set P and a unique morphism from x to y whenever $x \preceq y$. Then by a *P-Persistent Object* we mean a functor $\Phi : \mathcal{P} \to \mathcal{C}$.

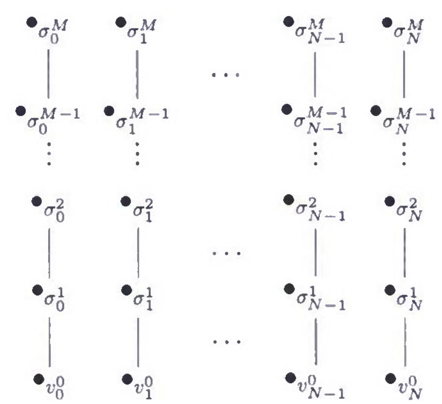
The paper by Carlsson gives examples of $\mathbb{N}$ and $\mathbb{R}$ persistent objects. A natural category of study is the category of topological spaces. Morse theory describes a continuous filtration of a topological space given by a Morse function. This filtration of the topological space produces an $\mathbb{R}$-persistent object in the category of topological spaces.

Definition 4.2. An $\mathbb{N}$-*filtration* of a space X is a sequence $X^0 \subset X^1 \subset X^2 \subset \ldots$ of subspaces of X whose union is X.

Given a poset P and antichain A, we will now describe an $\mathbb{N}$-filtration on the associated simplicial complex $\Delta(P, A)$. Suppose $P^0 \subset \cdots \subset P^M = P$ is an increasing sequence of sub-posets of P. Then we can define a filtration on $\Delta(P, A)$ by $\Delta(P,A)^j = \Delta(P^j, A \cap P^j)$ for each $j \in \mathbb{N}$. That is to say, the jth step in the filtration is the simplicial complex obtained from the construction associated to the poset P^j with antichain $P^j \cap A$. Given a height function h as described above, we get a natural sequence of sub-posets. We can set $P^j = \{p \in P : h(p) \leq j\}$, filtering the poset by its distance from the antichain.

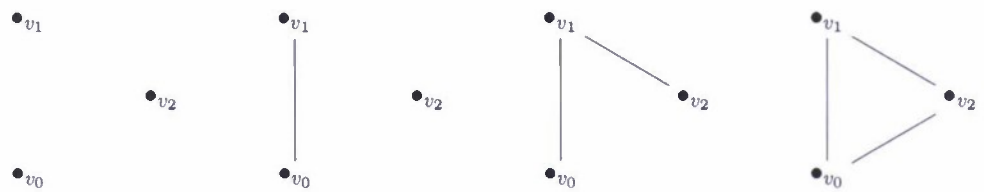
Theorem 4.3 (Realization). *Let K be a finite simplicial complex. Let $\{F^i\}$ be an $\mathbb{N}$-filtration of the simplicial complex with $K^{(0)} = F^0$. There is a poset P and an antichain $A \subset P$ such that $\Delta(P, A) = K$. Moreover, the shortest path height function h provides the same filtration as given by $\{F^i\}$.*

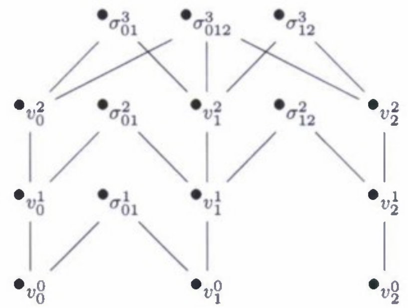
Proof. Suppose $K^{(0)} = \{v_0, \dots, v_N\}$ and let $\tilde{P}$ be an empty poset. Let $M \in \mathbb{N}$ be such that $F^i = F^M$ for all $i \geq M$. For each simplex $\sigma_{i_1 \dots i_k} = [v_{i_1}, \dots, v_{i_k}]$ in F^1, add an element $\sigma^1_{i_1 \dots i_k}$ to $\tilde{P}$. Then add the relations $v^0_{i_j} \preceq \sigma^1_{i_1 \dots i_k}$ for $j = 1 \dots k$. For each simplex $\sigma_{i_1 \dots i_k} = [v_{i_1}, \dots, v_{i_k}]$ in F^2, add an element $\sigma^2_{i_1 \dots i_k}$ to $\tilde{P}$. Then add the relations $\sigma^1_{i_j} \preceq \sigma^2_{i_1 \dots i_k}$ for $j = 1 \dots k$. Proceed inductively. For a simplex $\sigma_{i_1 \dots i_k} = [v_{i_1}, \dots, v_{i_k}]$ in F^i, add an element $\sigma^i_{i_1 \dots i_k}$ to $\tilde{P}$. Then add the relations $\sigma^{i-1}_{i_j} \preceq \sigma^i_{i_1 \dots i_k}$ for $j = 1 \dots k$. Finally, take P to be the transitive closure of $\tilde{P}$. This construction will result in chains $v^0_i \preceq \sigma^1_i \preceq \cdots \preceq \sigma^{M-1}_i \preceq \sigma^M_i$ for each vertex present in K, with additional elements and relations in P for each simplex present in K.



Now to show this construction has the desired properties. Let $A = \{p \in P : \nexists q \in P \text{ with } q \leq q\}$, the set of minimal elements of P. Let $h : P \to \mathbb{N}$ be the length of the shortest path to A function. Let $P^j = \{p \in P : h(p) \leq j\}$ denote the filtration of P given by this height function. By construction, $\Delta(P^0, P^0) = K^{(0)} = F^0$. Now for a general $j \in \mathbb{N}$ with $j \leq M$. Let $\sigma = [v_{i_1}, \dots, v_{i_k}]$ be a simplex in F^j. Then $\sigma^j_{i_1 \dots i_k} \in P$. Moreover, $h(\sigma^j_{i_1 \dots i_k}) \leq j$ because the chain of relations $v^0_i \preceq \sigma^1_i \preceq \cdots \preceq \sigma^{j-1}_i \preceq \sigma^j_{i_1 \dots i_k}$ exists for each $i = 1 \dots k$. So the shortest path to A has length at most j and $\sigma^j_{i_1 \dots i_k} \in P^j$. Thus σ is a simplex in $\Delta(P^j, P^j \cap A)$ and we have that $F^j \subset \Delta(P^j, P^j \cap A)$. Now let σ be a simplex in $\Delta(P^j, P^j \cap A)$. The constructed simplicial complex $\Delta(P^j, P^j \cap A)$ only depends on the maximal elements of P^j. So σ is constructed due to an element $\sigma^j_{i_1 \dots i_k} \in P^j$. This element was added in correspondence with the simplex $[v_{i_1}, \dots, v_{i_k}] \in F^j$. By construction, the element $\sigma^j_{i_1 \dots i_k} \in P^j$ only adds $[v_{i_1}, \dots, v_{i_k}]$ and its sub-simplices to $\Delta(P^j, P^j \cap A)$. So σ, must be $[v_{i_1}, \dots, v_{i_k}]$ or one of its sub-simplices. Thus $\Delta(P^j, P^j \cap A) = F^j$. $\square$

Example: Given the 2-simplex filtered in the following way, the above construction produces the following poset.





This construction is not unique. There are many other posets that will create the same filtered simplicial complex.

5 Category

Let $P = \{p_1, \dots, p_N\}$ be a finite poset and A an antichain. Let B_N denote the boolean algebra on N elements. The category, $\mathcal{B}_N$, has a partial order given by inclusion. There is a functor from $\mathcal{B}_N$ to $\mathcal{PA}$ which takes an element $\{i_1, \dots, i_k\}$ of $\mathcal{B}_N$ to the subposet $\{p_{i_1}, \dots, p_{i_k}\}$ and which takes morphisms in $\mathcal{B}_N$ to inclusions of posets. The category $\mathcal{B}_N$ is a directed set.

Proposition 5.1. $\varinjlim_{I \in \mathcal{B}_N} I = \{1, \dots, N\}$

Proof. $\{1, \dots, N\}$ is cofinal in $\mathcal{B}_N$ because $\{1, \dots, N\}$ is the unique maximal element of B_N. $\square$

For a fixed antichain, we can define a category $\mathcal{C}_P(A)$. The objects of this category are $\Delta(S, S \cap A)$ for each $S \subset P$, denoted Δ_S. That is, this category is the image of the above functor. The morphisms are defined as

$$\mathrm{Hom}_{\mathcal{C}_P(A)}(\Delta_{S_1}, \Delta_{S_2}) = \begin{cases} \iota : \Delta_{S_1} \to \Delta_{S_2} & \text{if } S_1 \subset S_2, \\ \emptyset & \text{otherwise.} \end{cases}$$

This category is a directed set. The object $\Delta_P \in Obj(\mathcal{C}_P(A))$ is cofinal.

Proposition 5.2. $\varinjlim \Delta_{P_i} = \Delta_P$

With definition 4.1 in mind, define $\Phi : \mathcal{B}_N \to \mathcal{C}_P(A)$ to be $\Phi(P_1) = \Delta_{P_1}$. Thus the whole category $\mathcal{C}_P(A)$ is a $\mathcal{B}_N$-persistent object.

Proposition 5.3. *Let P be a poset and $\Phi : \mathcal{P} \to \mathcal{C}$ be a P-persistent object. That is, we have a family of objects $\{c_x\}_{x \in P} \subset Obj(\mathcal{C})$ together with morphisms $\phi_{xy} : c_x \to c_y$ whenever $x \preceq y$ such that $\phi_{yz} \circ \phi_{xy} = \phi_{xz}$ whenever $x \preceq y \preceq z$. Suppose $p_0 \preceq p_1 \preceq \cdots \preceq p_N$ is a chain in P. Then there is an $\mathbb{N}$-persistent object in $\mathcal{C}$ given by the functor*

$$\Phi_{p_0p_1\dots p_N}(i) = \begin{cases} p_i & \text{if } 0 \leq i < N \\ p_N & \text{if } i \geq N \end{cases}.$$

Proposition 5.4. *If $\mathcal{F} : \mathcal{C} \to \mathcal{D}$ is a functor and $\Phi : \mathcal{P} \to \mathcal{C}$ is a P-persistent object, then $\mathcal{F} \circ \Phi : \mathcal{P} \to \mathcal{D}$ is a P-persistent object.*

6 Simplicial Homology

Once we have a simplicial complex built, a natural set of invariant to compute are the Betti numbers. We give a brief description of the Betti numbers here. For a more detailed reference, see [Munkres].

Definition 6.1. The *nth chain group* of an abstract simplicial complex K, denoted $C_n(K)$, is the free abelian group generated by the oriented n-simplices of K. That is $[\sigma] = [\tau]$ if $\sigma = \tau$ and σ and τ are oppositely oriented.

Definition 6.2. The *simplicial chain complex* associated to K is the chain complex $\dots \to C_2(K) \xrightarrow{\partial_2} C_1(K) \xrightarrow{\partial_1} C_0(K) \xrightarrow{\partial_0} 0 \to \dots$ with boundary operator

$$\partial_n([v_{i_0}, \dots, v_{i_n}]) = \sum_{j=0}^{n} (-1)^j [v_{i_0}, \dots, \hat{v_{i_j}}, \dots, v_{i_n}]$$

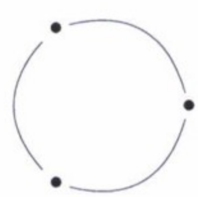
where $[v_{i_0}, \dots, \hat{v_{i_j}}, \dots, v_{i_n}]$ denotes the $n-1$ simplex formed by omitting v_{i_j}.

It is a standard exercise to show that $\partial_i \circ \partial_{i+1} = 0$ for each $i \in \mathbb{N}$. This makes it possible to define our homology groups.

Definition 6.3. The *nth homology group*, $H_n(K)$, is given by $H_n(K) = \ker \partial_n / \operatorname{im} \partial_{n+1}$.

Definition 6.4. The *nth Betti number*, $\beta_n(K)$, is defined as $\beta_n(K) = rank_{\mathbb{Z}}(\mathrm{Free}(H_n(K)))$

Example: The circle



$$H_n(S^1) = \begin{cases} \mathbb{Z} & \text{if } n = 0, 1 \\ 0 & \text{otherwise} \end{cases} \quad \beta_n(S^1) = \begin{cases} 1 & \text{if } n = 0, 1 \\ 0 & \text{otherwise} \end{cases}$$

7 Dimensionality

Let (P, A) be a finite poset antichain pair. Let $\Delta(P, A)$ be the associated simplicial complex.

Definition 7.1. The *dimension* of a poset P is the minimum number of total orders whose intersections give the partial order on P

Example: The poset S_n has dimension n.

Proposition 7.2. *If $Q \subset P$ is an induced subposet, then* $\dim Q \leq \dim P$.

Let X be a finite simplicial complex. Let N denote the number of vertices present in X. Then X is isomorphic to a subcomplex of Δ^N, the standard N-simplex. So the polytope of X, $|X|$, embeds in $\mathbb{R}^{N+1}$.

Definition 7.3. Let σ be a simplex in Δ^N. Let $V = \{e_{i_1}, \ldots, e_{i_k}\}$ be the vertices of σ. Let $\pi_\sigma : \mathbb{R}^{N+1} \to \mathbb{R}^{N+1}$ be the projection map onto the coordinates of σ. Let $\pi_\sigma^c = \mathrm{Id} - \pi_\sigma$. That is projection onto the coordinates not in σ.

If $r \in \mathbb{R}_+$, Define $U(\sigma, r) = \{x : x \in \mathbb{R}^N \text{and} \sum \pi_\sigma(x)_i > 1 - r\}$.

Define $V(\sigma, r) = U(\sigma, r) \cap |X|$.

$V_{\sigma,r}$ is open in X since the topology on X is the same as the induced topology from $\mathbb{R}^{N+1}$, and $U(\sigma, r)$ is open in $\mathbb{R}^{N+1}$.

Note: If $r < 1$, $V(\sigma, r)$ can't contain points in a simplex which does not intersect σ. If $r < \frac{1}{2}$, then $V(\sigma, r) \cap V(\tau, r) = \emptyset$ whenever $\sigma \cap \tau = \emptyset$.

Lemma 7.4. *If $\sigma \in X$ and $r < 1$, $V(\sigma, r)$ is homotopy equivalent to σ.*

Proof. Define $H(x,t) : V(\sigma, r) \to V(\sigma, r)$ by

$$H(x,t) = (1-t)x + t\frac{\pi_\sigma(x)}{\sum \pi_\sigma(x)_i}$$

Note this is well defined since $\pi_\sigma(x)_i \geq 1 - r \neq 0$.

$$\begin{aligned} \sum H(x,t)_i &= \sum (1-t)x_i + t\frac{\pi_\sigma(x)_i}{\sum \pi_\sigma(x)_i} \\ &= (1-t)\sum x_i + \frac{t}{\sum \pi_\sigma(x)_i}\sum \pi_\sigma(x)_i \\ &= (1-t) + t \\ &= 1 \end{aligned}$$

So the homotopy remains in Δ^N. The sum of the coordinates is always 1, and if $x_i = 0$ for any i, then $H(x,t)_i = 0$ for all t. Thus, if $x \in \tau$ for some $\tau \in X$, then $H(x,t) \in \tau$ for all $t \in [0,1]$. The homotopy fixes points in σ. So σ is a deformation retract of $V(\sigma, r)$. □

Since σ is contractible in $|X|$, $V(\sigma, r)$ is contractible in $|X|$.

Lemma 7.5. *If $\sigma, \tau \in X$, then $V(\sigma, r_1) \cap V(\tau, r_2) \subset V(\sigma \cap \tau, r_1 + r_2)$.*

Proof. Let $x \in V(\sigma, r_1) \cap V(\tau, r_2)$. Then

$$\sum \pi_\sigma(x)_i > 1 - r_1, \sum \pi_\tau(x)_i > 1 - r_2$$

and

$$\sum x_i = 1$$

Since the coordinates sum to 1, we also have that

$$\sum \pi_\sigma^c(x)_i < r_1, \sum \pi_\tau^c(x)_i < r_2$$

So

$$\sum \pi_{\sigma\setminus\tau}(x)_i < r_1, \sum \pi_{\tau\setminus\sigma}(x)_i < r_2$$

$$\sum \pi^c_{\sigma\cap\tau}(x) = \sum \pi^c_\sigma(x) + \sum \pi^c_\tau(x) - \sum \pi^c_{\sigma\cup\tau}(x)$$
$$\sum \pi^c_{\sigma\cap\tau}(x) < r_1 + r_2 - \sum \pi^c_{\sigma\cup\tau}(x)$$
$$\sum \pi^c_{\sigma\cap\tau}(x) < r_1 + r_2$$

□

If $r_1, r_2 < 1$ and $\sigma \cap \tau = \emptyset$, then $V(\sigma, r_1) \cap V(\tau, r_2) = \emptyset$. If $r_1 + r_2 < 1$, then $V(\sigma, r_1) \cap V(\tau, r_2) = \{x \in |X| : \sum \pi_\sigma(x)_i > 1 - r_1 \text{ and } \sum \pi_\tau(x)_i > 1 - r_2\}$ is contractible in $|X|$. The above homotopy gives the retraction onto $\sigma \cap \tau$, which is contractible. In fact, if $\sigma_1 \cap \cdots \cap \sigma_n \neq \emptyset$ and $r_1 + r_2 + \cdots + r_n < 1$, then $V(\sigma_1, r_1) \cap V(\sigma_2, r_2) \cap \cdots \cap V(\sigma_n, r_n)$ is contractible.

Lemma 7.6. *If X is a simplicial complex with at most K facets, then $H_n(X) = 0$ for $n \geq K - 2$.*

Proof. Embed $|X|$ in $\Delta^N \subset \mathbb{R}^{N+1}$ as above. Let $M = \{\sigma \in X : \sigma \not\subset \tau \text{ for } \tau \neq \sigma \in X\}$, the facets of X. By assumption $|M| \leq K$. Let $\mathcal{O} = \{V(\sigma, \frac{1}{K}) : \sigma \in M\}$. Then $\mathcal{O}$ is an open cover of X since every $x \in X$ is in some facet of X. For any $\sigma \in M$, $V(\sigma, \frac{1}{K})$ is contractible. By the above lemma, and the fact that the intersection of two simplices is a face of both simplices, we see that the intersection of two such open sets, if nonempty, is also contractible. For $\sigma_1, \ldots, \sigma_k \in M$,

$$V(\sigma_1, \frac{1}{K}) \cap V(\sigma_2, \frac{1}{K}) \cap \cdots \cap V(\sigma_k, \frac{1}{K}) \subset V(\tau, \frac{k}{K})$$

where $\tau = \sigma_1 \cap \cdots \cap \sigma_k$. Since $k \leq K$, $\frac{k}{K} < 1$ and $V(\tau, \frac{k}{K})$ is contractible in X. Since $\mathcal{O}$ is a good cover of $|X|$, the nerve of $\mathcal{O}$, $N(\mathcal{O})$ has the same homology as $|X|$. Since this cover has only K elements, all $K + 1$ intersections are empty and there is only one K intersection to consider. If all of the facets of X intersect in a common simplex, then the K-wise intersection is nonempty. Otherwise, the nonempty K wise intersection results in a single $K - 1$ dimensional simplex in the nerve. The boundary of this simplicial complex is nonempty. In either case, the $K - 2$th homology of the nerve simplicial complex is trivial. For dimension reasons, all higher homology groups are trivial as well. □

Theorem 7.7. *Suppose $H_n(\Delta(P, A)) \neq 0$. Then $\dim P \geq \lceil (n + 2)/2 \rceil$. If A is a minimal or maximal antichain, then $\dim P \geq (n + 2)$*

Proof. By the above lemma, $\Delta(P, A)$ must have at least $n + 2$ facets. Let $V = \{v_1, \ldots v_k\}$ be the vertices present in $\Delta(P, A)$ and $f_1, \ldots, f_l$ be the facets present in S. By construction, each of these facets is present due to an element $p \in P$. Let $I(V)$ be the past and future of V. To show at least $n + 2$ distinct elements of M are needed to create S. Suppose $M = \{m_1, \ldots, m_K\}$ generates all of the faces of S. Then S can be thought of as a nonzero element of the nth homology group of $\Delta(M \cup A, A)$. The boundary of S is still empty. If S was not the boundary of an $n + 1$ chain in $\Delta(P, A)$, then it will not be $\Delta(M \cup A, A)$ either, since there are fewer such chains. By the above lemma, $K \geq n + 2$.

We now list total orders on $\{m_i\} \cup \{v_i\}$ which must be present in any realizer of this partial order.

This subposet consists of $v_1, \ldots, v_{l+2}$ which are unrelated and $m_1, \ldots, m_{n+2}$. For each i, m_i is related a unique set of vertices, and either preceeds all of them, or follows all of them. Split the m_i's into two groups. One in which the m_i's preceed vertices, call it M_{lower} and one where the m_i's follow vertices M_{upper}.

For each $m_i \in M_{\text{lower}}$ there must be a total order $<_i$ on V with $m_i <_i v_j$ for all $v_j \in A_{m_i} \cap V$ and $v_j <_i m_i$ for all $v_j \notin A_{m_i} \cap V$. Since m_i has a distinct set A_{m_i} of vertices, we must have a total order for each $m_i \in M_{\text{lower}}$.

Similarly, we need a total ordering which gives the relations for each $m_i \in M_{\text{upper}}$. These upper and lower m_i's can be related, so the total orderings of one may give the total orderings of the other. So we need at least max $|M_{\text{lower}}|, |M_{\text{upper}}|$ total orderings in any realizer to induce the proper relations. By the pigeon hole principle, this max is at least $\lceil |M|/2 \rceil = \lceil (n+2)/2 \rceil$. Thus we get that $dimP \geq \lceil (n+2)/2 \rceil$. If A is a minimal or maximal antichain, then the above max is $|M|$ and we have that $dimP \geq (n+2)$.

□

This bound is tight. For the poset S_{n+2} and antichain $A = \{a_1, \ldots, a_{n+2}\}$, $\Delta(S_{n+2}, A) \simeq S^n$. This poset has dimension $n+2$ and $\beta_n(S^n) = 1$.

8 Persistent Homology

We have described, for a given poset P, antichain A and height function h, a filtered simplicial complex $\Delta(P, A)$. One can compute the homology of the simplicial complex $\Delta(P, A)$ at any given step of the filtration to obtain some information about the original poset. A more useful approach is to compute the homology at every step of the filtration as a summary of information about the original poset.

Let T be a totally ordered set. Let R be a ring. Now to describe a special type of persistent object, one whose associated poset is totally ordered and target category is the category *Rmod.*

Definition 8.1. A persistence module parameterized by T over R is a family of R-modules $\{M_t\}_{t \in T}$ with R-module homomorphism $\{\varphi_{t,t'}\}_{(t,t') \in T \times T}$ such that $\varphi_{t,t'} \circ \varphi_{t',t''} = \varphi_{t,t''}$ whenever $t \preceq t' \preceq t''$.

When working over a field F, finite type persistence modules parameterized by $\mathbb{Z}$ have a nice decomposition, as formulated in [CZ04]. We now describe this decomposition. Let $I = (m, n) \subset \mathbb{N} \cup \{\infty\}$ be an interval. Define the persistence module $Q(I)$ parameterized by $\mathbb{N}$ over F to be

$$Q(I)_t = \begin{cases} F & \text{if } t \in I, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 8.2. A persistence module parameterized by $\mathbb{N}$ over F, $\{M_n\}_{n \in \mathbb{N}}$ is said to be of *finite type* if

1. M_n is a finite dimensional F vector space for each $n \in \mathbb{N}$
2. There exists an N such that for all $t \geq N$, $\varphi_{N,t}$ is an isomorphism.

Proposition 8.3. *A persistence module parameterized by $\mathbb{N}$ over F of finite type is isomorphic to one of the following type*

$$\bigoplus_{s=1}^{N} Q(I_s)$$

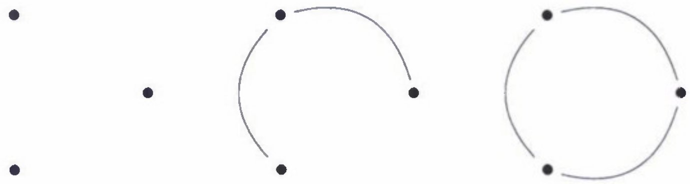
where $I_s \subset \mathbb{N} \cup \{\infty\}$ is an interval for each $s \in \mathbb{N}$.

In short, to encode the information to describe such a persistence module, we only need to list the relevant intervals. Hence the definition

Definition 8.4. A *barcode* is a collection of intervals $\{I_s\}_{s=1}^{N}$, where each interval I_s is bounded below.

By the definition of filtration, we have an inclusion map $\iota : \Delta(P, A)^j \hookrightarrow \Delta(P, A)^k$ for all $j \leq k$. So we have a $\mathbb{N}$-persistent object in the category of simplicial complex whenever we have an $\mathbb{N}$ filtration. By the functoriality of $H_n(-)$, we then get a persistence module parameterized by $\mathbb{N}$ over $\mathbb{Z}$.

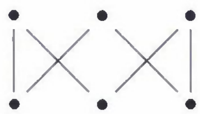
Example: The circle filtered in the following way produces the following persistence modules and barcodes.



$H_0 = \mathbb{Z}^3$	$\mathbb{Z}$	$\mathbb{Z}$
$H_1 = 0$	0	$\mathbb{Z}$

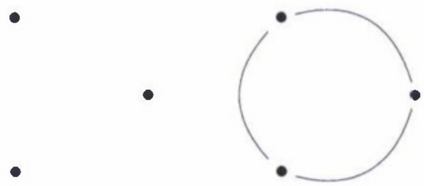
The circle filter in this manner has Betti 0 barcode
and Betti 1 barcode INSERT BARCODE PICTURE HERE

Example: Let P be the crown poset of height 2 on 3 elements. Let A be the antichain of minimal elements of P. Let h be the height function such that the minimal elements have height 0 and the maximal elements have height 1.



The poset filtered by height is then

This filtration on the poset results in the filtered simplicial complex



9 Sampling from a Lorentzian Manifold

One can take a finite sample of a Lorentzian manifold and create a poset using the causal structure of the manifold. This poset should contain some of the topological information about the manifold. By picking antichains in the created poset (analogous to picking space-like hypersurfaces of the original manifold) we can investigate topological properties of the poset.

10 Computed Examples

Let (M, g) be a Lorentzian manifold. Let $X \subset M$ be a finite collection of points. The set X can be given a partial order $\preceq_{(M,g)}$ using the causal structure of the manifold. For two points $x, y \in X$, we say that $x \preceq_{(M,g)} y$ if there is a future-directed time-like curve from x to y in M.

Cylinder In this example, we attempt to recover the space-like topology of the Lorentzian cylinder. The cylinder is described as an identification space of the unit square, and is given the flat metric. 1000 points were sampled the unit square. The set was given the partial order based on the (1+1) Lorentzian metric on the cylinder. The elements were assigned a height based on their shortest distance to the minimal elements of the poset. antichains in the poset were chosen to be sets of elements at the same height.

Example: Height 0, the minimal elements of the poset. The filtered construction resulted in the following barcodes

Pair of Pants In this example, we attempt to recover the space-like topology of the Lorentzian pair pants. The pair of pants were created using an identification space of the unit square.

The pair of pants minus the singularity was given the flat metric. 2000 points were sampled from the space and given the partial order based on the Lorentzian metric. Elements of the poset were assigned a height based on their shortest distance to the minimal elements of the poset. antichains in the poset were chosen as sets of elements of the same height. The resulting barcodes were dependent on the height of the antichain. antichains taken below the singularity resulted in single persistent first homology elements, whereas antichains taken above the singularity resulted in a pair of persistent first homology elements.

INSERT BARCODE PICTURES HERE

References

[1] Collins, A., Zomorodian, A., Carlsson, G, and Guibas, L., *A Barcode Shape Descriptor for Curve Point Cloud Data*, Eurographics Symposium on Point-Based Graphics ,2004.

[2] G. Carlsson, *Topology and Data*, Bulletin of the American Mathematical Society, vol 46, Number 2, pp.255-308, 2009.

[3] Aho, A.; Garey, M. R.; and Ullman, J. D., *The Transitive Reduction of a Directed Graph*, SIAM J. Comput. 1, 131-137, 1972.

[4] Munkres, James R., **Elements of Algebraic Topology**, Perseus Publishing, MA, 1984.