The Many Forms of Merge Trees

Justin M. Curry

June 5, 2018

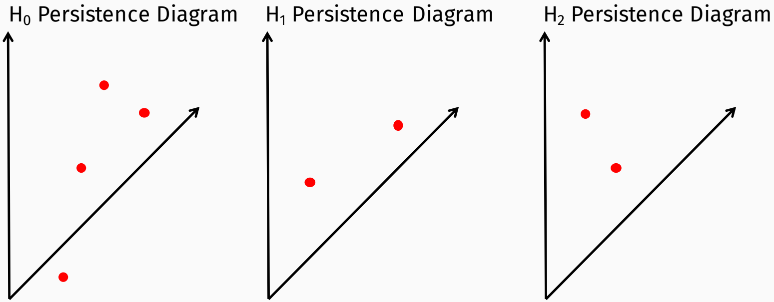
University at Albany SUNY

Motivating Problem

Characterizing the Persistence Map

Persistence Pipeline

$$f : X \to \mathbb{R} \quad \rightsquigarrow \quad F : t \mapsto X_{\leq t} \mapsto H_i(X_{\leq t}) \quad \rightsquigarrow \quad \mathrm{PD}(f)_i$$



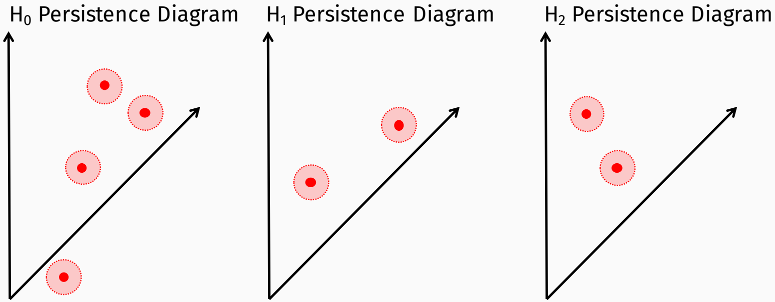
Open Question

What arrangements of points are allowed? In other words, what is the image of the persistence map and what is its fiber?

Characterizing the Persistence Map

Persistence Pipeline

$$f : X \to \mathbb{R} \quad \rightsquigarrow \quad F : t \mapsto X_{\leq t} \mapsto H_i(X_{\leq t}) \quad \rightsquigarrow \quad \mathrm{PD}(f)_i$$



Active Question

Describe explicitly the pushforward measure.

Mental Models for Merge Trees

Mental Models for Derivatives

Bill Thurston *On Proof and Progress in Mathematics*

(1) Infinitesimal: the ratio of the infinitesimal change in the value of a function to the infinitesimal change in a function.
(2) Symbolic: the derivative of x^n is nx^{n-1}, the derivative of $\sin(x)$ is $\cos(x)$, the derivative of $f \circ g$ is $f' \circ g * g'$, *etc.*
(3) Logical: $f'(x) = d$ if and only if for every ϵ there is a δ such that when $0 < |\Delta x| < \delta$,

$$\left| \frac{f(x+\Delta x) - f(x)}{\Delta x} - d \right| < \delta.$$

Mental Models for Derivatives

Bill Thurston *On Proof and Progress in Mathematics*

(4) Geometric: the derivative is the slope of a line tangent to the graph of the function, if the graph has a tangent.
(5) Rate: the instantaneous speed of $f(t)$, when t is time.
(6) Approximation: The derivative of a function is the best linear approximation to the function near a point.
(7) Microscopic: The derivative of a function is the limit of what you get by looking at it under a microscope of higher and higher power.

Mental Models for Derivatives

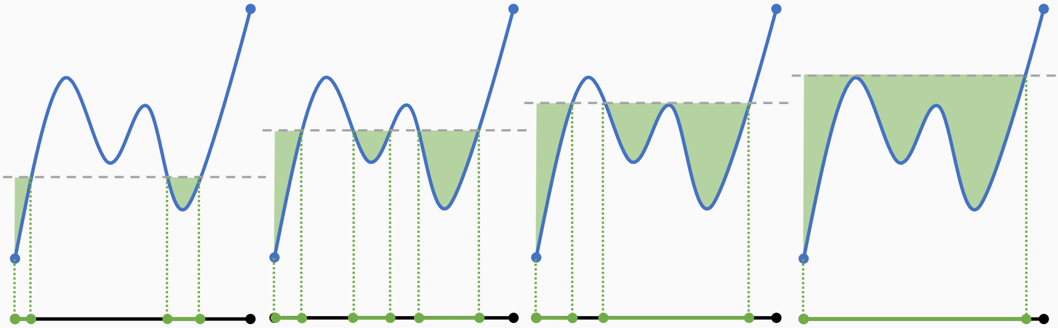
Bill Thurston *On Proof and Progress in Mathematics*

37. The derivative of a real-valued function f in a domain D is the Lagrangian section of the cotangent bundle $T^*(D)$ that gives the connection form for the unique flat connection on the trivial $\mathbf{R}$-bundle $D \times \mathbf{R}$ for which the graph of f is parallel.

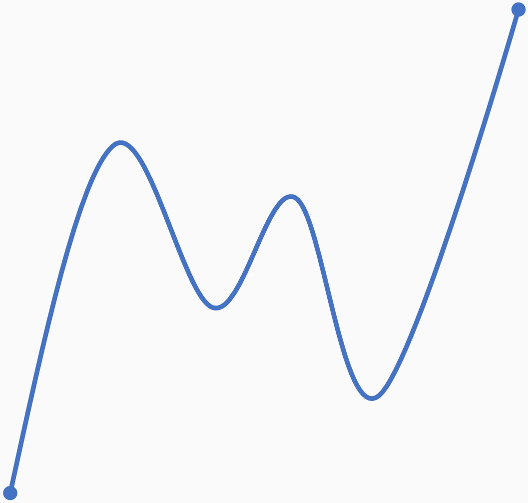
Mental Models for Merge Trees

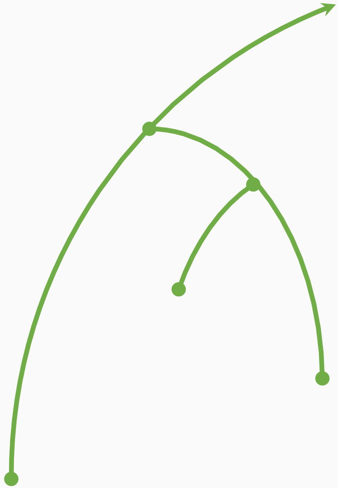
What about Merge Trees?

Merge Trees: Primal Sources



Merge Trees: Primal Sources





Mental Models for Merge Trees

1. **Physical:** The merge tree tracks how water fills the portion above the graph of a function.
2. **Statistical:** The merge tree tracks clusters across multiple scales.
3. **Morse-Theoretic:** The merge tree of a Morse function $f : M \to \mathbb{R}$ is a graph whose leaf nodes correspond to index 0 critical points and whose internal nodes correspond to index 1 critical points.
4. **Computational:** The merge tree is a disjoint set system, where sets and union operations are indexed by a height parameter.
5. **Quotient Space:** The merge tree of $f : X \to \mathbb{R}$ is the quotient space associated to the epigraph of f

$$\pi : \Gamma_f^+ \to \mathbb{R} \qquad \text{where} \qquad \Gamma_f^+ = \{(x,t) \mid f(x) \leq t\}$$

   where $(x,t) \sim (x',t')$ iff $t = t'$ and $[x] = [x'] \in \pi_0(\pi^{-1}(t))$.

Mental Models for Merge Trees

Mental Models 6 through 38 were lost in a cab in Philly.

Mental Models for Merge Trees

39. **Topological-Axiomatic:** A connected, locally-finite, contractible, one-dimensional cell complex T with a distinguished edge e_∞ making $T - e_\infty$ a compact rooted tree, with root v_∞, and a map $\pi : T \to \mathbb{R}$ that is orientation-preserving in a precise sense
40. **Functorial:** A merge tree is a **Morse set**, which is a functor $S : (\mathbb{R}, \leq) \to \mathbf{Set}$ satisfying certain tameness properties:

$$\varphi_{tr} : S(r) \to S(t) \qquad \text{with} \qquad \varphi_{tr} = \varphi_{ts} \circ \varphi_{sr} \qquad \forall s \in [t, r]$$

41. **(Co)Sheaf Theoretic:** The merge tree is the étalé (display) space associated to a constructible (co)sheaf of sets on $\mathbb{R}$, equipped with the up-set (down-set) Alexandrov topology.
42. **Section of the Containment Poset** The merge tree is a barcode (collection of half-open intervals) equipped with a choice of attaching maps.

The Elder Rule

The Elder Rule 1.0

From Edelsbrunner and Harer's *Computational Topology*

> ELDER RULE. At a juncture, the older of the two merging paths continues and the younger path ends.

Letting $a \leq b$ be two thresholds, we let $\beta(a, b)$ be the number of components in $\mathbb{X}_b$ that have a non-empty intersection with $\mathbb{X}_a$. In terms of the merge tree, this is the number of subtrees with topmost points at value b that reach down to level a or below. Each such subtree has a unique path, its longest, that spans the entire interval between a and b. It follows that $\beta(a, b)$ is also the number of paths in the path decomposition of $G(f)$ that span $[a, b]$. We note that any path decomposition that is not generated using the Elder Rule does not have this property. In particular, if f is Morse then the Elder Rule generates a unique path decomposition, which is therefore the only one for which the number of paths spanning $[a, b]$ is equal to $\beta(a, b)$ for all values of $a \leq b$.

Persistent Homology in Degree 0

If I take a nice enough **persistent set** $S : (\mathbb{R}, \leq) \to \textbf{Set}$ and I consider the associated freely generated persistence module F_S, then...

Persistent Homology in Degree 0

If I take a nice enough **persistent set** $S : (\mathbb{R}, \leq) \to \textbf{Set}$ and I consider the associated freely generated persistence module F_S, then...

GABRIEL TRUMPETS HIS HORN AND DECLARES

Persistent Homology in Degree 0

If I take a nice enough **persistent set** $S : (\mathbb{R}, \leq) \to \mathbf{Set}$ and I consider the associated freely generated persistence module F_S, then...

GABRIEL TRUMPETS HIS HORN AND DECLARES

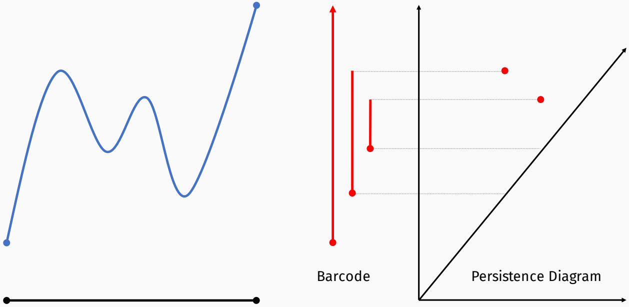
"There exists a consistent change of basis that expresses this module as a direct sum of interval modules!"

Persistent Homology in Degree 0

If I take a nice enough **persistent set** $S : (\mathbb{R}, \leq) \to$ **Set** and I consider the associated freely generated persistence module F_S, then...

GABRIEL TRUMPETS HIS HORN AND DECLARES

"There exists a consistent change of basis that expresses this module as a direct sum of interval modules!"



East Coast vs. West Coast?

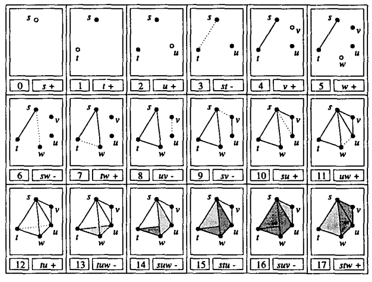
East Coast vs. West Coast?

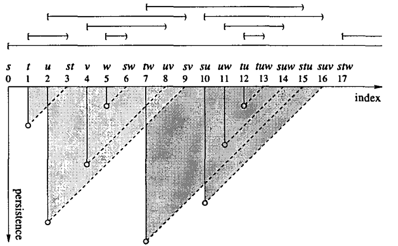
Topological Persistence and Simplification *

Herbert Edelsbrunner
Department of Computer Science
Duke University, Durham
and Raindrop Geomagic, RTP
North Carolina

David Letscher
Department of Mathematics
Oklahoma State University
Stillwater, Oklahoma

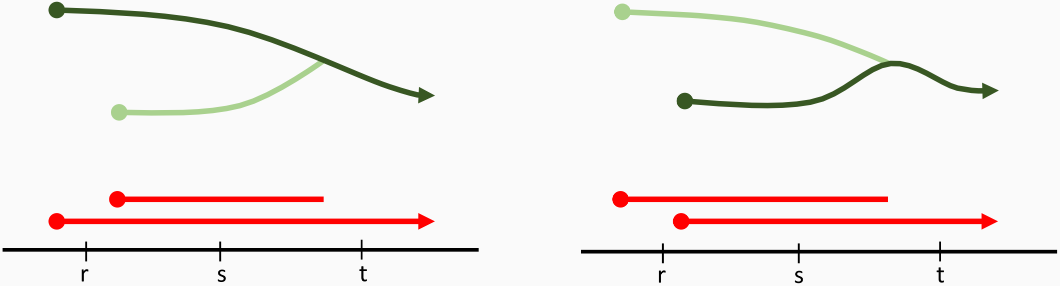
Afra Zomorodian
Department of Computer Science
University of Illinois
Urbana, Illinois





The Elder Rule 2.0

For persistent H_0 the barcode/PD is completely determined by the merge tree via the **Elder Rule**.

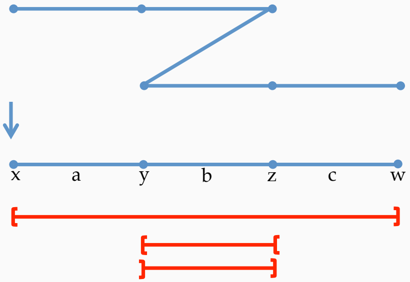


The rank of the map $\varphi_{tr} : H_0(F(r)) \to H_0(F(t))$ is equal to the number of bars born before r and that die after t.

Why Does this Work?

Note that the Elder Rule says that every bar in a barcode determines a local section of the map $\pi : T \to \mathbb{R}$. *N.B. Here I'm using mental model 39 of a merge tree.*

This is *false* for Reeb graphs



Moreover, it frustrates the idea that there is any connection between a branch decomposition of a graph and the associated indecomposables.

The Elder Rule 3.0

A persistent set $S : (\mathbb{R}, \leq) \to \mathbf{Set}$ has an associated poset

$$P_S := \bigsqcup_{t \in \mathbb{R}} S(t) = \cup_{t \in \mathbb{R}} S(t) \times \{t\} \quad \text{where} \quad (x, r) \leq (y, t) \Leftrightarrow \varphi_{tr}(x) = y,$$

which has an order-preserving, i.e. continuous in the up-set topology, map

$$\begin{array}{c} P_S \\ \pi \big\downarrow \\ \mathbb{R} \end{array}$$

N.B. This is exactly the étalé space construction. So this is mental model 41 version A.

In general, the étalé space of a sheaf on a poset is a poset...consider multi-parameter clustering and merge sheets over $\mathbb{R}^2$

The Elder Rule 3.0

To guarantee that the étalé space is a nice tree, we introduce the tameness assumptions of mental model 40:

Definition

A **Morse set** is a functor $S : (\mathbb{R}, \leq) \to \textbf{set}$ which admits a minimal sequence of times $\tau = \{\tau_1 < \cdots < \tau_n\}$ so that

1. $S(t) = \varnothing$ for $t < \tau_1$,
2. $S|_{[\tau_i, \tau_{i+1})}$ is naturally isomorphic to the constant functor with value $S(\tau_i)$ for $i = 1, \ldots, n-1$,
3. $S(t) = \{*\}$ for $t \geq \tau_n$,
4. for all $x \in S(\tau_{i+1})$ the fiber $\varphi^{-1}_{i+1,i}(x)$ has cardinality 0,1, or 2, and
5. each element $y \in S(t)$ is contained in a unique, oldest maximal chain C_y in the associated poset P_S.

The Elder Rule 3.0

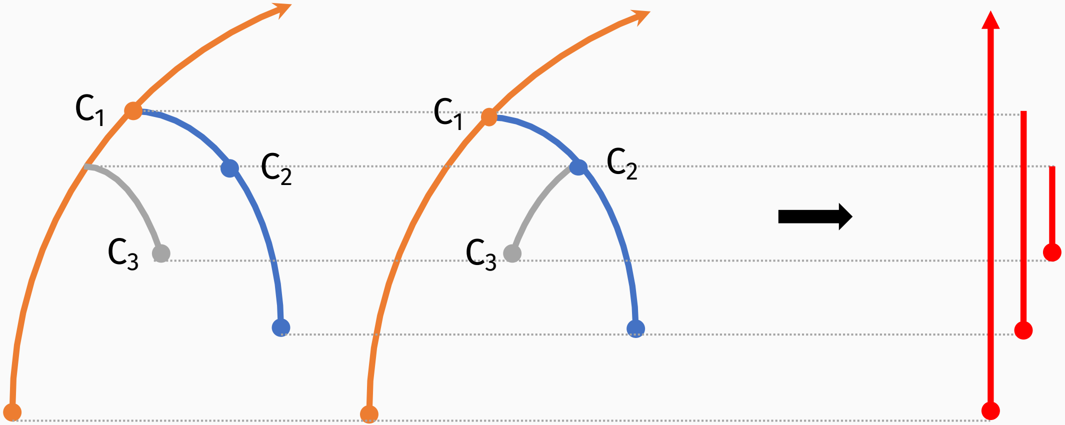
Definition

Let S be a Morse set. The **Elder Rule** gives the following inductive chain decomposition of the poset P_S

- Let P_1 be the poset P_S and let C_1 be the unique, oldest maximal chain containing the element $\star \in P_1(\tau_n)$.
- Let $P_{i+1} = P_i - C_i$, i.e. the poset P_i with the chain C_i removed.
- By the fourth and fifth hypotheses of a Morse set, P_{i+1} has a unique element $\star_{n-i} \in P_{i+1}(\tau_{n-i})$ that is contained in a unique, oldest chain C_{i+1}.

Let $B = \{\pi(C_i)\}$ to be the set of intervals associated to the chains C_i via projection along π. This defines the barcode associated to S by the Elder Rule.

Two Examples, Same Barcode



The Elder Rule 3.0

Note that given any collection of intervals B, there is an associated persistence module V_B, all of whose shift maps are diagonal matrices.

Fact (Others???, C. '17)

Let S be a Morse set and let B be the set of intervals generated by the Elder rule.

There is an explicit natural isomorphism (change of basis)

$$\beta : V_B \to F_S.$$

Consequently, the Elder Rule recovers Gabriel's decomposition.

Question

Question/Claim

The chain decomposition perspective should work for arbitrary functors

$$F : (\mathbb{R}, \leq) \to \mathbf{set}$$

and in particular indecomposables of all interval types should be obtained.

Counting Merge Trees

Mental Model No.42

Definitions

Let B be a collection of intervals $\{I_j\}_{j=1}^N$ where $I_1 = [b_1, \infty)$ and $I_j = [b_j, d_j)$ with $d_{j+1} < d_j$. Let $B_{>1} = B \setminus I_1$.

The **containment poset** is the poset where $I_j \leq I_k$ iff $I_j \subseteq I_k$.

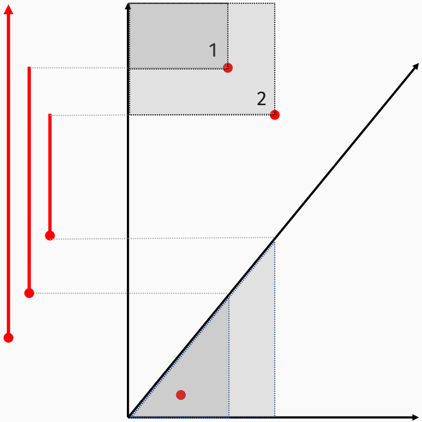
The **strict up-set** over I_j is $U_j = \{I_k \mid I_j \subset I_k\}$.

Merge Trees are Sections of the Containment Poset

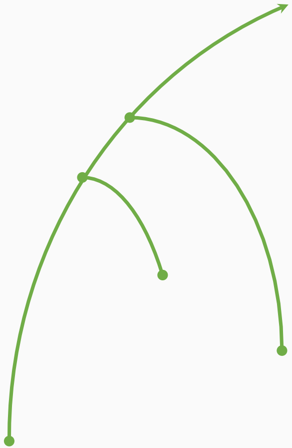
A merge tree with barcode B is a section of the following map

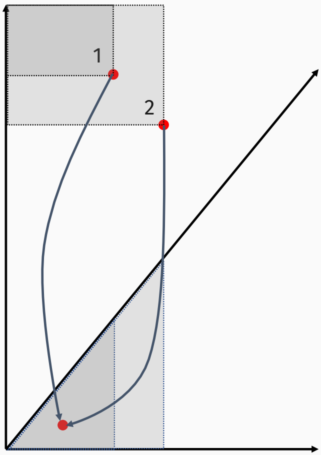
$$\begin{array}{ccccc} \bigsqcup_{j>1} U_j & \qquad & \{(I_j, I_k) \mid I_j \subsetneq I_k\} & \qquad & I_{s(j)} \\ \pi \big\downarrow \big\uparrow s & & \big\downarrow & & \big\uparrow \\ B_{>1} & & \{I_j\}_{j=2}^N & & I_j \end{array}$$

Containment Poset

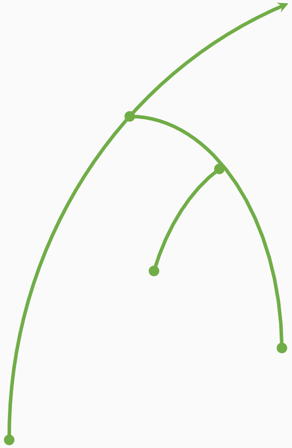


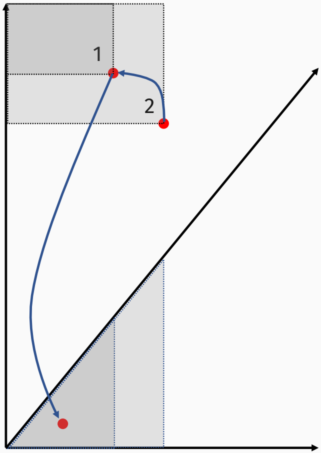
Sections and Associated Merge Trees





Sections and Associated Merge Trees





Counting Merge Trees

Theorem (C. '17)

Fix a barcode B like before. Set $\mu_B(I_j) = |U_j|$, which is the number of points up and to the left of the point I_j.

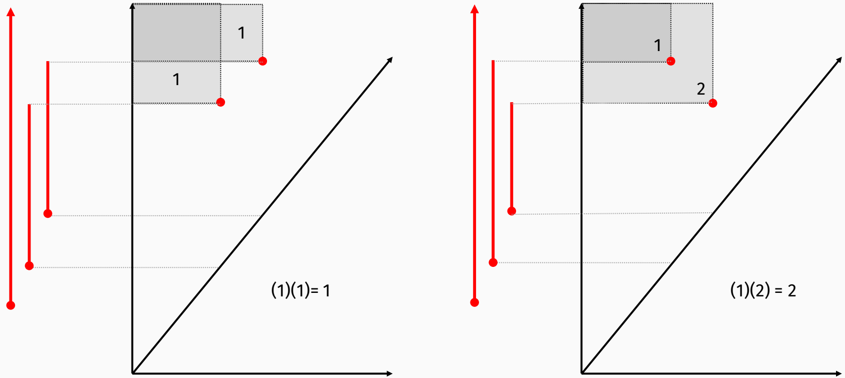
The number of non-isomorphic merge trees with barcode B is

$$C(B) := \prod_{j=2}^{N} \mu_B(I_j)$$

Induced Stratification of the Configuration Space

We can think of the Elder rule as a map $\Xi : \mathcal{T} \to \mathcal{B}$ from the space of merge trees to the space of barcodes/persistence diagrams, which is a stratified by space of possible containment posets.

The counting function C is the pushforward of the indicator function or, said differently, is the constructible function associated to $\Xi_* \Bbbk_{\mathcal{T}}$ by K_0.

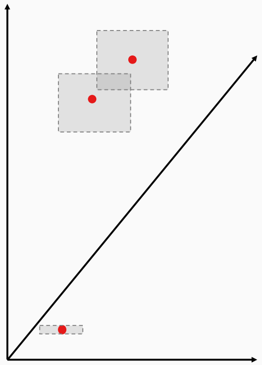


Understanding the Pushforward Measure

The integral

$$\Xi_* \Xi^* \mu(V) = \int_{B \in V} C(B) d\mu$$

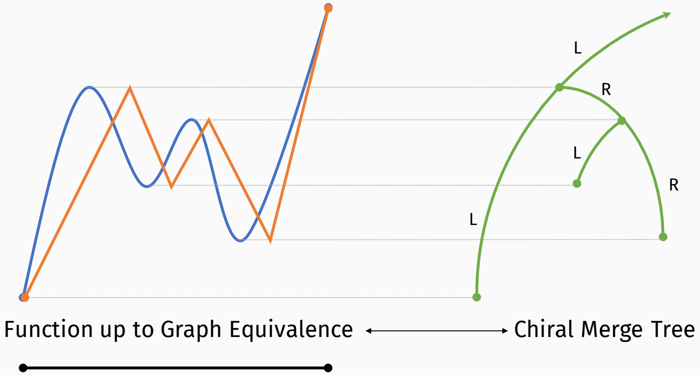
provides a proxy to the pushforward measure on persistence diagrams, promising a more accurate understanding of statistics on persistence diagrams in degree 0.



Further Work

Chiral Merge Trees

For time-series analysis, we can consider **chiral merge trees** to distinguish up-trending versus down-trending behavior.



Cor. (C. '17)

If we fix a barcode B with one bar of the form $[b_1, \infty)$ and $N-1$ bars of the form $[b_j, d_j)$, then the graph-equivalence classes of functions $f : [0,1] \to \mathbb{R}$ attaining minima on the boundary is

$$2^{N-1} \prod_{j=2}^{N} \mu_B(I_j).$$

Open Question

Work in Progress (IMA Crowd Source)

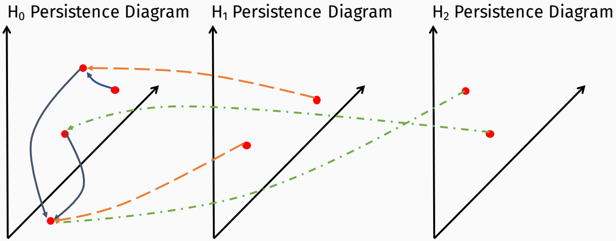
For $X = M_g$ some genus g surface, the space of graph equivalence classes of functions with the same persistence in all degrees is discrete and precisely enumerable.

Decorated Merge Trees (w/ Rachel Levanger)

Persistent homology is naturally indexed over the merge tree:

$$\begin{array}{ccc} \Gamma_f^+ & \xrightarrow{q} & T \\ \downarrow{\scriptstyle \pi} & \swarrow{\scriptstyle \pi} & \\ \mathbb{R} & & \end{array}$$

We lift bars in a barcode to obtain a **Decorated Merge Trees**. We refine the bottleneck distance to penalize matchings that "jump" components.



Thank You!