Persistent diagrams as diagrams

Ulrich Bauer

TUM

June 7, 2018

Abel Symposium 2018 Topological Data Analysis, Geiranger

Joint work with Michael Lesnick (Princeton/Albany)

Persistence diagrams: multiset of points $(b,d) \in \overline{\mathbb{R}}^2 : b \leq d$ (Edelsbrunner et al. 2000, 2007)

- Persistence diagrams: multiset of points $(b, d) \in \overline{\mathbb{R}}^2 : b \leq d$ (Edelsbrunner et al. 2000, 2007)
- Persistence barcodes: multiset of intervals, decomposition structure $M \cong \bigoplus_{I \in B(M)} \mathbb{K}(I)$ (Edelsbrunner et al. 2000, Carlsson et al. 2004, 2005, Crawley-Boewey 2015)

- Persistence diagrams: multiset of points $(b, d) \in \overline{\mathbb{R}}^2 : b \leq d$ (Edelsbrunner et al. 2000, 2007)
- Persistence barcodes: multiset of intervals, decomposition structure $M \cong \bigoplus_{I \in B(M)} \mathbb{K}(I)$ (Edelsbrunner et al. 2000, Carlsson et al. 2004, 2005, Crawley-Boewey 2015)
- Persistence measures: for all $a < b \le c < d$, count multiplicity of $0 \to \mathbb{K} \to \mathbb{K} \to 0$ as summand of $M_a \to M_b \to M_c \to M_d$ (Chazal et al. 2015)

- Persistence diagrams: multiset of points $(b, d) \in \overline{\mathbb{R}}^2 : b \leq d$ (Edelsbrunner et al. 2000, 2007)
- Persistence barcodes: multiset of intervals, decomposition structure $M \cong \bigoplus_{I \in B(M)} \mathbb{K}(I)$ (Edelsbrunner et al. 2000, Carlsson et al. 2004, 2005, Crawley-Boewey 2015)
- Persistence measures: for all $a < b \le c < d$, count multiplicity of $0 \to \mathbb{K} \to \mathbb{K} \to 0$ as summand of $M_a \to M_b \to M_c \to M_d$ (Chazal et al. 2015)
- ▶ Rank invariant (rank function): $(s, t) \mapsto \operatorname{rank} M_{s,t}$ (for $s \le t$ or s < t) (Carlsson at el. 2009)

- Persistence diagrams: multiset of points $(b, d) \in \overline{\mathbb{R}}^2 : b \leq d$ (Edelsbrunner et al. 2000, 2007)
- Persistence barcodes: multiset of intervals, decomposition structure $M \cong \bigoplus_{I \in B(M)} \mathbb{K}(I)$ (Edelsbrunner et al. 2000, Carlsson et al. 2004, 2005, Crawley-Boewey 2015)
- Persistence measures: for all $a < b \le c < d$, count multiplicity of $0 \to \mathbb{K} \to \mathbb{K} \to 0$ as summand of $M_a \to M_b \to M_c \to M_d$ (Chazal et al. 2015)
- ▶ Rank invariant (rank function): $(s, t) \mapsto \operatorname{rank} M_{s,t}$ (for $s \le t$ or s < t) (Carlsson at el. 2009)
- Matching diagrams: sequence of partial bijections (Edelsbrunner et al. 2014)

Inerval decompositions and persistence modules

Theorem (Crawley-Boewey 2015)

Any pointwise finite-dimensional (pfd) persistence module (a diagam $M : \mathbb{R} \to \mathbf{vect}$) has an essentially unique decomposition as a direct sum of indecomposable interval modules, isomorphic to

$$0 \to \cdots \to 0 \to \underbrace{\mathbb{K} \to \cdots \to \mathbb{K}}_{\text{supported by an interval } I \subseteq \mathbb{R}} \to 0 \to \cdots$$

► The corresponding collection (multiset) of intervals is the *persistence barcode* of *M*.

Inerval decompositions and persistence modules

Theorem (Crawley-Boewey 2015)

Any pointwise finite-dimensional (pfd) persistence module (a diagam $M : \mathbb{R} \to \mathbf{vect}$) has an essentially unique decomposition as a direct sum of indecomposable interval modules, isomorphic to

$$0 \to \cdots \to 0 \to \underbrace{\mathbb{K} \to \cdots \to \mathbb{K}}_{supported \ by \ an \ interval \ I \subseteq \mathbb{R}}_{} \to 0 \to \cdots$$

- ► The corresponding collection (multiset) of intervals is the *persistence barcode* of *M*.
- The points in the persistence diagram are the endpoints of the intervals in the barcode.

Inerval decompositions and persistence modules

Theorem (Crawley-Boewey 2015)

Any pointwise finite-dimensional (pfd) persistence module (a diagam $M : \mathbb{R} \to \mathbf{vect}$) has an essentially unique decomposition as a direct sum of indecomposable interval modules, isomorphic to

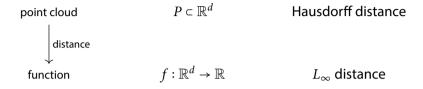
$$0 \to \cdots \to 0 \to \underbrace{\mathbb{K} \to \cdots \to \mathbb{K}}_{supported \ by \ an \ interval \ I \subseteq \mathbb{R}}_{} \to 0 \to \cdots$$

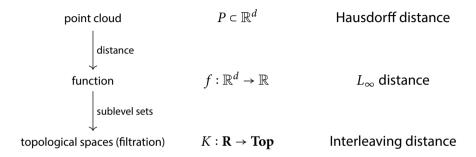
- ► The corresponding collection (multiset) of intervals is the persistence barcode of M.
- ► The points in the *persistence diagram* are the endpoints of the intervals in the barcode.
- This is not a diagram in the sense of category theory (functor)!

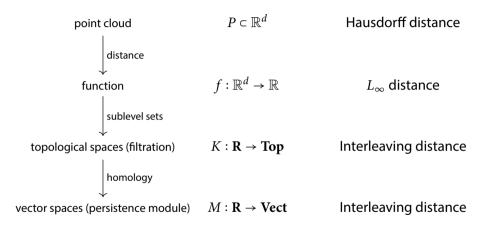
point cloud

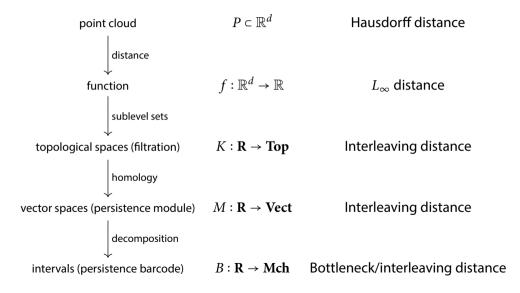
 $P \subset \mathbb{R}^d$

Hausdorff distance







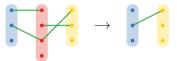


The category of matchings

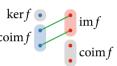
Consider the category Mch (a subcategory of the category Rel of sets and relations) with

- objects: sets,
- morphisms: matchings (partial bijections).

Composition:



(Co)kernel/image:

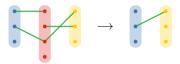


The category of matchings

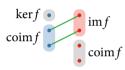
Consider the category Mch (a subcategory of the category Rel of sets and relations) with

- objects: sets,
- morphisms: matchings (partial bijections).

Composition:



(Co)kernel/image:



Mch is *Puppe-exact* (*p-exact*):

- ▶ it has a zero object (∅)
- ▶ it has all (co)kernels
- every mono (epi) is (co)kernel
- every morphism $f: A \to B$ has an epi-mono factorization $A \twoheadrightarrow \operatorname{im} f \hookrightarrow B$

but not additive:

▶ it does not have all (co)products

▶ A barcode (collection of intervals) can be read as a diagram $\mathbb{R} \to \mathbf{Mch}$:



 $t\mapsto \{\text{intervals in barcode containing }t\} \quad (s\le t)\mapsto \{\text{intervals containing both }s,t\}$

▶ A barcode (collection of intervals) can be read as a diagram $\mathbb{R} \to \mathbf{Mch}$:

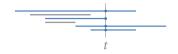


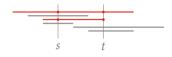
$$t\mapsto \{\text{intervals in barcode containing }t\} \quad (s\le t)\mapsto \{\text{intervals containing both }s,t\}$$

A matching diagram defines a barcode:



▶ A barcode (collection of intervals) can be read as a diagram $\mathbb{R} \to \mathbf{Mch}$:





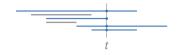
 $t \mapsto \{\text{intervals in barcode containing } t\} \quad (s \le t) \mapsto \{\text{intervals containing both } s, t\}$

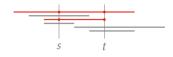
A matching diagram defines a barcode:



• equivalence classes $\mathcal{E}(D) := \left(\bigcup_{t \in \mathbb{R}} \{t\} \times D_t\right) / \sim$, where $(s, x) \sim (t, y)$ for all $s \leq t, \ x \in D_s, \ y \in D_t$

▶ A barcode (collection of intervals) can be read as a diagram $\mathbb{R} \to \mathbf{Mch}$:





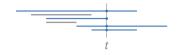
 $t \mapsto \{\text{intervals in barcode containing } t\} \quad (s \le t) \mapsto \{\text{intervals containing both } s, t\}$

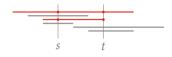
A matching diagram defines a barcode:



- equivalence classes $\mathcal{E}(D) := \left(\bigcup_{t \in \mathbb{R}} \{t\} \times D_t\right) / \sim$, where $(s, x) \sim (t, y)$ for all $s \leq t, x \in D_s, y \in D_t$
- project to first component: supporting interval

▶ A barcode (collection of intervals) can be read as a diagram $\mathbb{R} \to \mathbf{Mch}$:





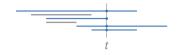
 $t \mapsto \{\text{intervals in barcode containing } t\} \quad (s \le t) \mapsto \{\text{intervals containing both } s, t\}$

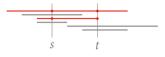
A matching diagram defines a barcode:



- equivalence classes $\mathcal{E}(D) := \left(\bigcup_{t \in \mathbb{R}} \{t\} \times D_t\right) / \sim$, where $(s, x) \sim (t, y)$ for all $s \leq t, x \in D_s, y \in D_t$
- project to first component: supporting interval

▶ A barcode (collection of intervals) can be read as a diagram $\mathbb{R} \to \mathbf{Mch}$:





 $t \mapsto \{\text{intervals in barcode containing } t\} \quad (s \le t) \mapsto \{\text{intervals containing both } s, t\}$

A matching diagram defines a barcode:



- equivalence classes $\mathcal{E}(D) := \left(\bigcup_{t \in \mathbb{R}} \{t\} \times D_t\right) / \sim$, where $(s, x) \sim (t, y)$ for all $s \leq t, x \in D_s, y \in D_t$
- project to first component: supporting interval

Turn this into an equivalence of categories $\mathbf{Barc} \simeq \mathbf{Mch}^{\mathbb{R}}$

A category of barcodes

Proposition

The functor category is equivalent to Barc, the category with

- objects: barcodes (as a disjoint union of intervals),
- morphisms: overlap matchings X → Y: if I ∈ U is matched to J ∈ V, then I overlaps J to the right:
 - ▶ I bounds J above (every $s \in J$ is bounded above by some $t \in I$),
 - J bounds I below,
 - $I \cap J = \emptyset$.

A category of barcodes

Proposition

The functor category is equivalent to **Barc**, the category with

- objects: barcodes (as a disjoint union of intervals),
- morphisms: overlap matchings X → Y: if I ∈ U is matched to J ∈ V, then I overlaps J to the right:
 - ▶ I bounds J above (every $s \in J$ is bounded above by some $t \in I$),
 - J bounds I below,
 - $I \cap I = \emptyset.$
- composition: $\tau \bullet \sigma = \{(I, K) \in \tau \circ \sigma \mid I \text{ overlaps } K \text{ above}\}.$

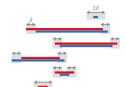
$$(I,K) \in \tau \bullet \sigma \text{ (overlap)} \qquad (I,K) \notin \tau \bullet \sigma \text{ (no overlap)}$$

Bottleneck distance as an interleaving distance

 δ -matching between barcodes U, V:

- if *I* is matched to *J*, then endpoints are δ -close
- unmatched intervals are 2δ -trivial (shorter than 2δ)

Bottleneck distance: $d_B(U, V) = \inf\{\delta \mid \exists \delta\text{-matching } U \nrightarrow V\}$

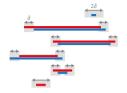


Bottleneck distance as an interleaving distance

 δ -matching between barcodes U, V:

- if *I* is matched to *J*, then endpoints are δ -close
- unmatched intervals are 2δ -trivial (shorter than 2δ)

Bottleneck distance:
$$d_B(U, V) = \inf\{\delta \mid \exists \delta \text{-matching } U \nrightarrow V\}$$



 δ -interleaving between diagrams X, Y indexed over \mathbb{R} (in any category): natural transformations $f_t: X_t \to Y_{t+\delta}, g_t: Y_t \to X_{t+\delta}$ yielding commutative diagrams

$$X_{t-\delta} \longrightarrow X_t \longrightarrow X_{t+\delta}$$

$$X_{t-\delta} \longrightarrow Y_t \longrightarrow Y_{t+\delta}$$

$$\forall t \in \mathbb{R}.$$

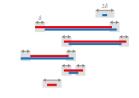
Interleaving distance: $d_B(U, V) = \inf\{\delta \mid \exists \delta \text{-interleaving } X \leftrightarrow Y\}$

Bottleneck distance as an interleaving distance

 δ -matching between barcodes U, V:

- if *I* is matched to *J*, then endpoints are δ -close
- unmatched intervals are 2δ -trivial (shorter than 2δ)

Bottleneck distance:
$$d_B(U, V) = \inf\{\delta \mid \exists \delta \text{-matching } U \nrightarrow V\}$$



 δ -interleaving between diagrams X, Y indexed over \mathbb{R} (in any category): natural transformations $f_t: X_t \to Y_{t+\delta}, g_t: Y_t \to X_{t+\delta}$ yielding commutative diagrams

$$X_{t-\delta} \longrightarrow X_t \longrightarrow X_{t+\delta}$$

$$X_{t-\delta} \longrightarrow Y_t \longrightarrow Y_{t+\delta}$$

$$\forall t \in \mathbb{R}.$$

Interleaving distance: $d_B(U, V) = \inf\{\delta \mid \exists \delta \text{-interleaving } X \leftrightarrow Y\}$

Proposition

 $d_I = d_B$ (using the equivalence **Barc** \simeq **Mch**^{\mathbb{R}}).

Non-functoriality of persistence barcodes

Can a pfd persistence module $M : \mathbf{vect}^{\mathbb{R}}$ be turned into its barcode $B(M) : \mathbf{Mch}^{\mathbb{R}}$ by a functor $B : \mathbf{vect} \to \mathbf{Mch}$ (or $\mathbf{vect}^{\mathbb{R}} \to \mathbf{Mch}^{\mathbb{R}}$)?

• This would preserve δ -interleavings, and thus yield stability of persistence barcodes.

Non-functoriality of persistence barcodes

Can a pfd persistence module $M : \mathbf{vect}^{\mathbb{R}}$ be turned into its barcode $B(M) : \mathbf{Mch}^{\mathbb{R}}$ by a functor $B : \mathbf{vect} \to \mathbf{Mch}$ (or $\mathbf{vect}^{\mathbb{R}} \to \mathbf{Mch}^{\mathbb{R}}$)?

• This would preserve δ -interleavings, and thus yield stability of persistence barcodes.

Theorem

There is no functor $\mathbf{vect} \to \mathbf{Mch}$ sending every vector space V to a set of cardinality $\dim V$ (equivalently, a linear map f to a matching of cardinality $\operatorname{rank} f$).

But:

• there is such a functor for subcategories of monos/epis of persistence modules $\mathbf{vect}^{\mathbb{R}}$:

Structure of persistence sub-/quotient modules

Proposition

Let N be a quotient module of a persistence module M (for M woheadrightarrow N an epimorphism).

Then there is an injective map between the barcodes $B(N) \hookrightarrow B(M)$.

If *J* is mapped to *I*, then

- I and J are aligned below, and
- I bounds J above.

This construction is functorial. There is a dual result for submodules.



Structure of persistence sub-/quotient modules

Proposition

Let N be a quotient module of a persistence module M (for M woheadrightarrow N an epimorphism).

Then there is an injective map between the barcodes $B(N) \hookrightarrow B(M)$.

If J is mapped to I, then

- I and J are aligned below, and
- I bounds J above.

This construction is functorial. There is a dual result for submodules.

Rephrased for $\mathbf{Mch}^{\mathbb{R}}$:

Proposition

There is a functor from epimorphisms of persistence modules to epimorphisms of matching diagrams.
(Dually, there is a functor from monos to monos.)



Induced matchings

Theorem

For $f: M \to N$ a morphism of pfd persistence modules, the epi-mono factorization $M \twoheadrightarrow \operatorname{im} f \hookrightarrow N$ gives an induced matching $\chi(f)$ between their barcodes. If I is matched to J, then

- (i) I overlaps J above.
- (ii) If ker f is δ -trivial, then
 - (a) I bounds $I(\delta)$ above, and
 - (b) any unmatched interval of B(M) is δ -trivial.
- (iii) If coker f is δ -trivial, then
 - (a) $I(\delta)$ bounds J below, and
 - (b) any unmatched interval of B(N) is δ -trivial.



Induced matchings

Theorem

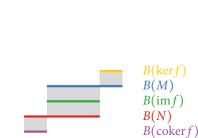
For $f: M \to N$ a morphism of pfd persistence modules, the epi-mono factorization $M \twoheadrightarrow \operatorname{im} f \hookrightarrow N$ gives an induced matching $\chi(f)$ between their barcodes. If I is matched to J, then

- (i) I overlaps J above.
- (ii) If ker f is δ -trivial, then
 - (a) I bounds $I(\delta)$ above, and
 - (b) any unmatched interval of B(M) is δ -trivial.
- (iii) If coker f is δ -trivial, then
 - (a) $I(\delta)$ bounds J below, and
 - (b) any unmatched interval of B(N) is δ -trivial.

Rephrased in $\mathbf{Mch}^{\mathbb{R}}$:

Theorem

If $f: M \to N$ has δ -trivial (co)kernel, then so does $\chi(f)$.

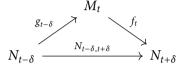


 $B(\operatorname{im} f)$

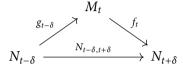
 $B(M(\delta))$

B(N)

Consider interleaving $f_t: M_t \to N_{t+\delta}, g_t: N_t \to M_{t+\delta} \ (\forall t \in \mathbb{R})$:

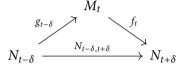


Consider interleaving $f_t: M_t \to N_{t+\delta}, g_t: N_t \to M_{t+\delta} \ (\forall t \in \mathbb{R})$:



- $im N_{t-\delta,t+\delta} \hookrightarrow im f_t \hookrightarrow N_{t+\delta},$
 - $M_t \rightarrow \inf_t$.

Consider interleaving $f_t: M_t \to N_{t+\delta}$, $g_t: N_t \to M_{t+\delta}$ ($\forall t \in \mathbb{R}$):



- $im N_{t-\delta,t+\delta} \hookrightarrow im f_t \hookrightarrow N_{t+\delta},$
- $M_t \rightarrow \inf_t$.

Consider interleaving $f_t: M_t \to N_{t+\delta}, g_t: N_t \to M_{t+\delta} \ (\forall t \in \mathbb{R})$:

$$N_{t-\delta} \xrightarrow{N_{t-\delta,t+\delta}} N_{t+\delta}$$

- $im N_{t-\delta,t+\delta} \hookrightarrow im f_t \hookrightarrow N_{t+\delta},$
 - $M_t \rightarrow \inf_t$.

Consider interleaving $f_t: M_t \to N_{t+\delta}, g_t: N_t \to M_{t+\delta} \ (\forall t \in \mathbb{R})$:

$$N_{t-\delta} \xrightarrow{S_{t-\delta,t+\delta}} N_{t+\delta}$$

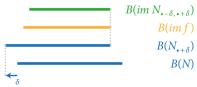
- $\qquad \text{im} \, N_{t-\delta,t+\delta} \hookrightarrow \text{im} f_t \hookrightarrow N_{t+\delta},$
- $M_t \rightarrow \inf f_t$.



Consider interleaving $f_t: M_t \to N_{t+\delta}$, $g_t: N_t \to M_{t+\delta}$ ($\forall t \in \mathbb{R}$):

$$N_{t-\delta} \xrightarrow{N_{t-\delta,t+\delta}} N_{t+\delta}$$

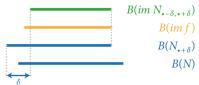
- $im N_{t-\delta,t+\delta} \hookrightarrow im f_t \hookrightarrow N_{t+\delta},$
- $M_t \rightarrow \inf_t$.



Consider interleaving $f_t: M_t \to N_{t+\delta}, g_t: N_t \to M_{t+\delta} \ (\forall t \in \mathbb{R})$:

$$N_{t-\delta} \xrightarrow{N_{t-\delta,t+\delta}} N_{t+\delta}$$

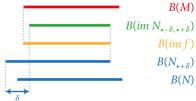
- $\qquad \text{im } N_{t-\delta,t+\delta} \hookrightarrow \text{im } f_t \hookrightarrow N_{t+\delta},$
- $M_t \rightarrow \text{im } f_t$.



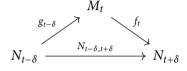
Consider interleaving $f_t: M_t \to N_{t+\delta}, g_t: N_t \to M_{t+\delta} \ (\forall t \in \mathbb{R})$:

$$N_{t-\delta} \xrightarrow{N_{t-\delta,t+\delta}} N_{t+\delta}$$

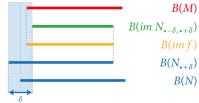
- $im N_{t-\delta,t+\delta} \hookrightarrow im f_t \hookrightarrow N_{t+\delta},$
- $M_t \rightarrow \inf f_t$.



Consider interleaving $f_t: M_t \to N_{t+\delta}, g_t: N_t \to M_{t+\delta} \ (\forall t \in \mathbb{R})$:



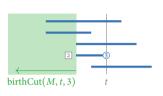
- $\qquad \text{im } N_{t-\delta,t+\delta} \hookrightarrow \text{im } f_t \hookrightarrow N_{t+\delta},$
- $M_t \rightarrow \inf f_t$.



Matching diagrams from persistence modules

Let $M : \mathbf{vect}^{\mathbb{R}}$. For $t \in \mathbb{R}$, $i \in \mathbb{N}$, define

$$\begin{aligned} & \text{birthCut}(M,t,i) = \{s < t \mid \text{rank}\,M_{s,t} < i\}, \\ & \text{birthOrd}(M,t,i) = \min\{i - \text{rank}\,M_{s,t} > 0 \mid s < t\}, \\ & \text{birthId}(M,t,i) = (\text{birthCut}(M,t,i), \text{birthOrd}(M,t,i)). \end{aligned}$$



Construct a matching diagram $B(M) : \mathbb{R} \to \mathbf{Mch}$: for all $t \le u \in \mathbb{R}$, define

$$B(M)_t = \{i \in \mathbb{N} \mid i \leq \dim M_t\}$$

$$B(M)_{t,u} = \{(i,j) \mid \text{birthId}(M,t,i) = \text{birthId}(M,u,i).$$

Yields a barcode without using interval decomposition!



More about matching diagrams

Proposition

- $im B(M)_{t,u} = \{j \in \mathbb{N} \mid j \leq \operatorname{rank} M_{t,u} \}.$
- ► If $i \le j$, then $birthCut(M, t, i) \subseteq birthCut(M, t, j)$ (bars with smaller label i at parameter value t are born earlier)
- ► If $(i,j) \in B(M)_{t,u}$, then $i \ge j$ (decreasing numbers along each bar): $i-j = \dim(\operatorname{im} M_{s,t} \cap \ker M_{t,u})$ for $s \in \operatorname{argmax}_s \{\operatorname{rank} M_{s,t} < i\}$
- ▶ If (i, j), $(k, l) \in B(M)_{t,u}$, then $i \ge k \Leftrightarrow j \ge l$
- Thus, bars are partially ordered; extends to lexicographic order
 - earlier birth, and (for same birth)
 - later death

Applies even to q-tame persistence modules (rank $M_{t,u} < \infty \ \forall t < u$)!

Induced matchings for matching diagrams

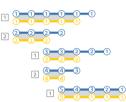
Let N be a quotient module of a persistence module M (M woheadrightarrow N an epimorphism). Define

$$\chi(M \hookrightarrow N)_t = \{(i,j) \mid \text{birthId}(M,t,i) = \text{birthId}(N,t,i)\}.$$

Theorem

B and χ form a functor from epimorphisms of persistence modules to epimorphisms of matching diagrams.

(Dually, there is a functor from monos to monos.)



- ▶ This is the structure theorem for sub-/quotient modules in terms of matching diagrams.
- Using an epi-mono factorization, this yields induced matchings and algebraic stability for q-tame persistence modules.
- Can be used to guide the construction of a decomposition for pdf modules.

Thanks for your attention!

