

STEINHAUS FILTRATION: MOVIE RECS using TOPOLOGY!

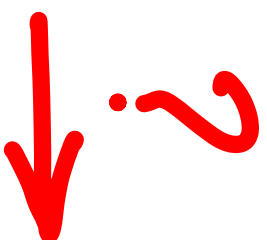
Bala Krishnamoorthy
Washington State University

with D. Arendt (PNVL), M. Broussard, N. Saul, A. Threl

SocG 2025



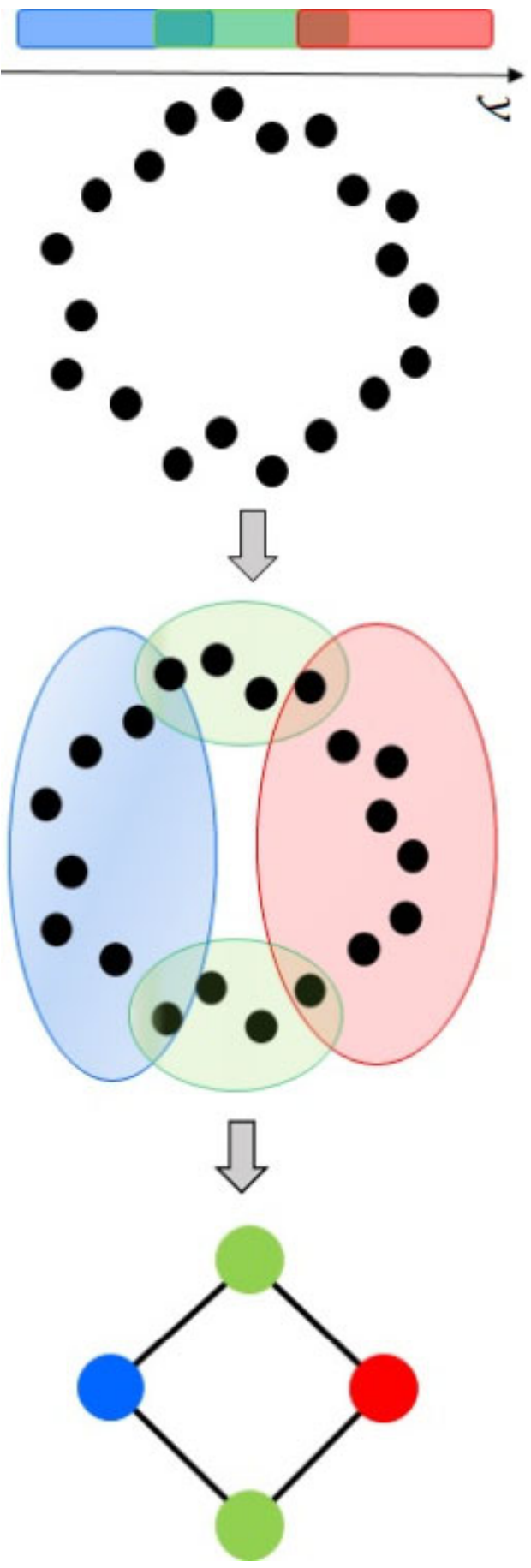
Movie Recommendations



(images: Wikipedia)

MAPPER

Singh, Mémoli, Carlsson (2007)



pers \rightarrow cover \rightarrow pullback \rightarrow refine \rightarrow nerve

MAPPER: APPLICATIONS

→ numerous areas in last decade
biomedicine, criminology, machine learning, ...

MAPPER: APPLICATIONS

- numerous areas in last decade
biomedicine, criminology, machine learning, ...
- use 1-skeleton for interpretation
- ID features: paths, flares, loops, ...

MAPPER: STABILITY

- use framework of persistence
 - stability of persistence modules

MAPPER: STABILITY

- use framework of persistence
 - stability of persistence modules
- Dey, Mémetti, Wang (2016): Multiscale mapper
 - tower of covers
- Carrière and Dudot (2017): Stability of 1D mapper
 - multiverve mapper → extended persistence diagram

MAPPER: STABILITY

- use framework of persistence
 - stability of persistence modules
- Dey, Mémetti, Wang (2016): Multiscale mapper
 - tower of covers
- Carrière and Dudot (2017): Stability of 1D mapper
 - multiverve mapper → extended persistence diagram
- most applications still use a single mapper

ABSTRACT COVERS?

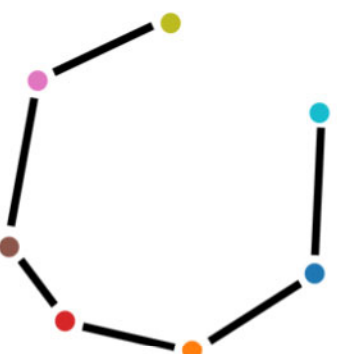
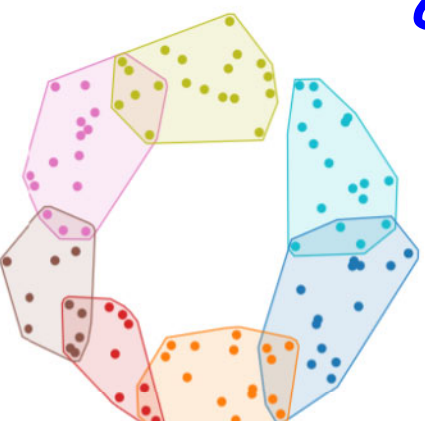
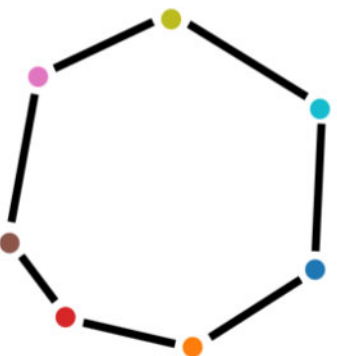
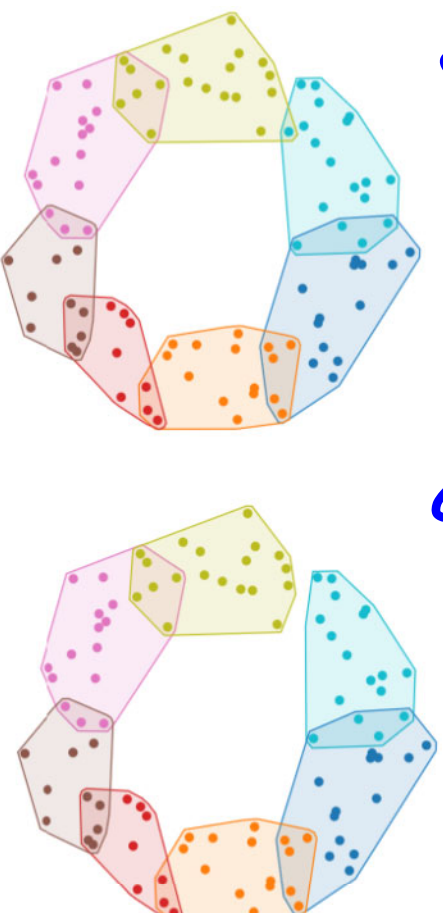
- Netflix: movies/viewers
 - no obvious metric available

ABSTRACT COVERS?

- Netflix: movies/viewers
 - no obvious metric available
- ? Define a mapper on abstract covers?

ABSTRACT COVERS?

- Netflix: movies/viewers
 - no obvious metric available
- ? Define a mapper on abstract covers?
- ? Stability from a single abstract cover?



RESULTS

- ✓ Steinhaus filtration from a single abstract cover
 - generalized Jacard distance
 - stability (for finite case)

RESULTS

- ✓ Steinhaus filtration from a single abstract cover
 - generalized Jacard distance
 - stability (for finite case)
- ✓ stable paths in a single mapper
 - stability measures strength of overlap

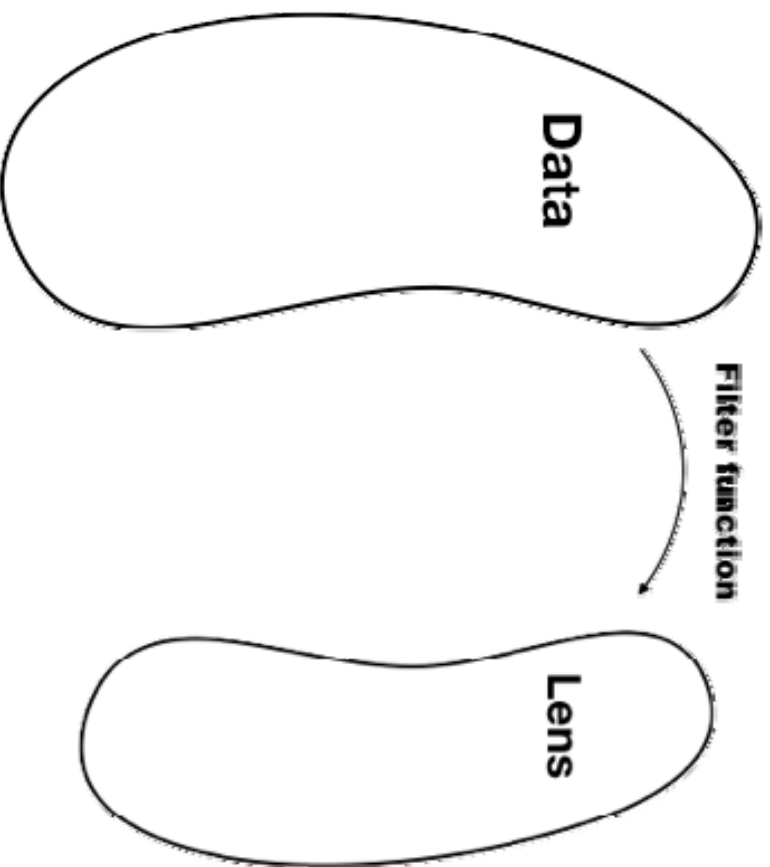
RESULTS

- ✓ Steinhaus filtration from a single abstract cover
 - generalized Jaccard distance
 - stability (for finite case)

- ✓ stable paths in a single mapper
 - stability measures strength of overlap
- ✓ Applications
 - movie recommendations
 - explainable machine learning

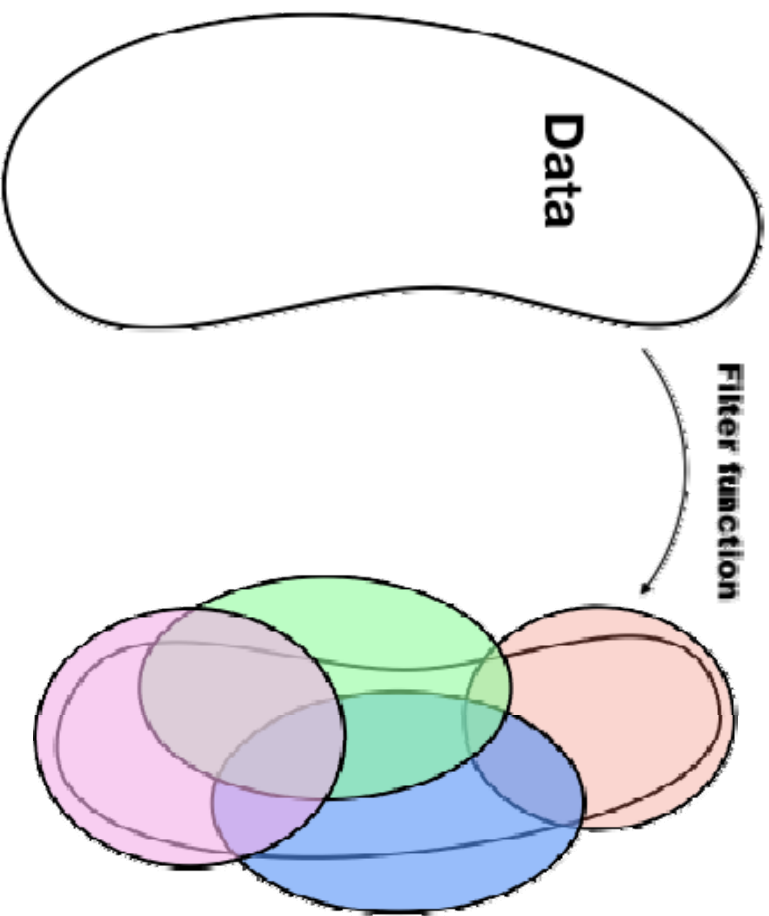
MAPPER

lens → cover → pullback → refine → nerve



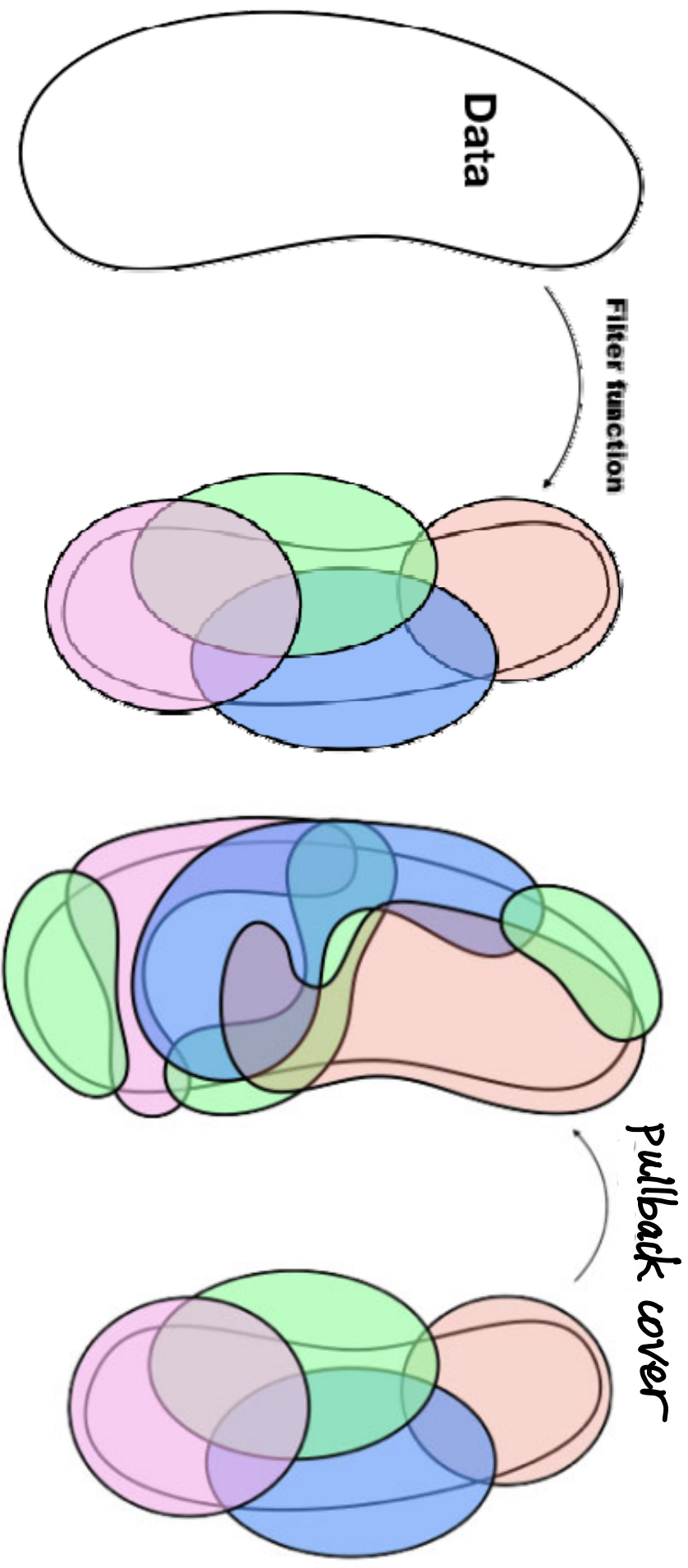
MAPPER

lers → cover → pullback → refine → nerve



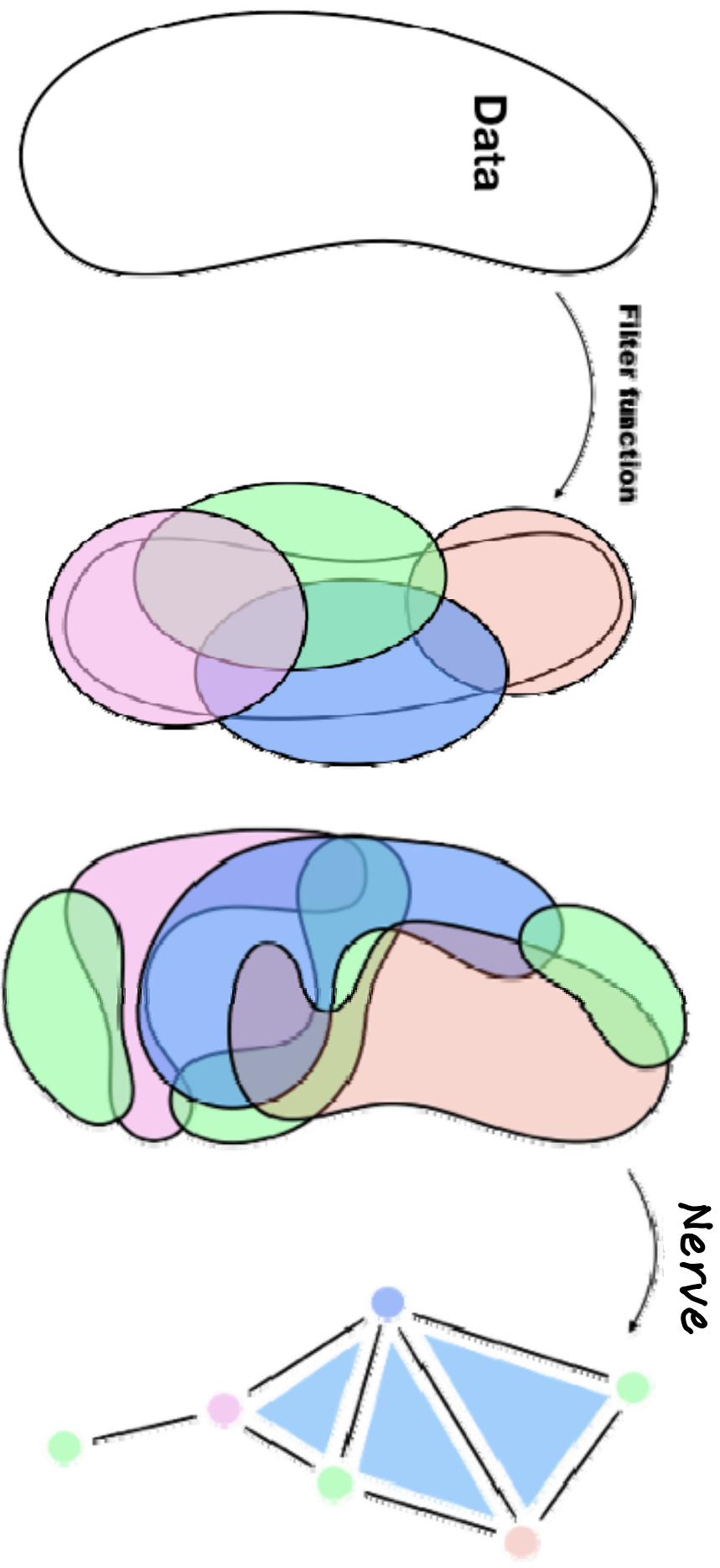
MAPPER

lers \rightarrow cover \rightarrow pullback \rightarrow refine \rightarrow nerve



MAPPER

lers \rightarrow cover \rightarrow pullback \rightarrow refine \rightarrow nerve



JACCARD DISTANCE

$$d_j(\text{Diagram 1}) = 1 - \frac{|\text{Diagram 2}|}{|\text{Diagram 3}|}$$

The diagram illustrates the Jaccard distance calculation using three Venn diagrams. Diagram 1 shows two overlapping sets: a black set with diagonal hatching and a red set with diagonal hatching. Diagram 2 shows the intersection of the two sets, which is shaded with both black and red diagonal hatching. Diagram 3 shows the union of the two sets, with the black set having black hatching and the red set having red hatching.

JACCARD DISTANCE

$$d_j \left(\begin{array}{|c|} \hline \text{Diagram of two overlapping sets with red and black shading} \\ \hline \end{array} \right) = 1 - \frac{\begin{array}{|c|} \hline \text{Diagram of the intersection of the two sets} \\ \hline \end{array}}{\begin{array}{|c|} \hline \text{Diagram of the union of the two sets} \\ \hline \end{array}}$$

Generalized Jaccard distance:

$$d_j \left(\begin{array}{|c|} \hline \text{Diagram of two overlapping sets with red, black, and green shading} \\ \hline \end{array} \right) = 1 - \frac{\begin{array}{|c|} \hline \text{Diagram of the intersection of the two sets} \\ \hline \end{array}}{\begin{array}{|c|} \hline \text{Diagram of the union of the two sets} \\ \hline \end{array}}$$

STEINHAUS DISTANCE

$$d_{st} \left(\begin{array}{c} \text{[Diagram: Two overlapping ellipses, one black with diagonal lines, one red with diagonal lines. The intersection is shaded with both colors.]}\end{array} \right) = 1 - \frac{\mu \left(\begin{array}{c} \text{[Diagram: A small ellipse with diagonal lines, representing the intersection of the two sets above.]}\end{array} \right)}{\mu \left(\begin{array}{c} \text{[Diagram: A single red ellipse with diagonal lines, representing the union of the two sets above.]}\end{array} \right)}$$

Generalized Steinhaus distance: (μ : measure)

$$d_{st} \left(\begin{array}{c} \text{[Diagram: Two overlapping ellipses, one black with diagonal lines, one red with diagonal lines. The intersection is shaded with both colors.]}\end{array} \right) = 1 - \frac{\mu \left(\begin{array}{c} \text{[Diagram: A small ellipse with diagonal lines, representing the intersection of the two sets above.]}\end{array} \right)}{\mu \left(\begin{array}{c} \text{[Diagram: Two overlapping ellipses, one black with diagonal lines, one red with diagonal lines. The intersection is shaded with both colors.]}\end{array} \right)}$$

STEINHAUS NERVE

→ $Nrv_{st}(\mathcal{U})$: nerve of cover \mathcal{U} with each simplex σ given $d_{st}(\tau)$ as weight or birth time

STEINHAUS NERVE

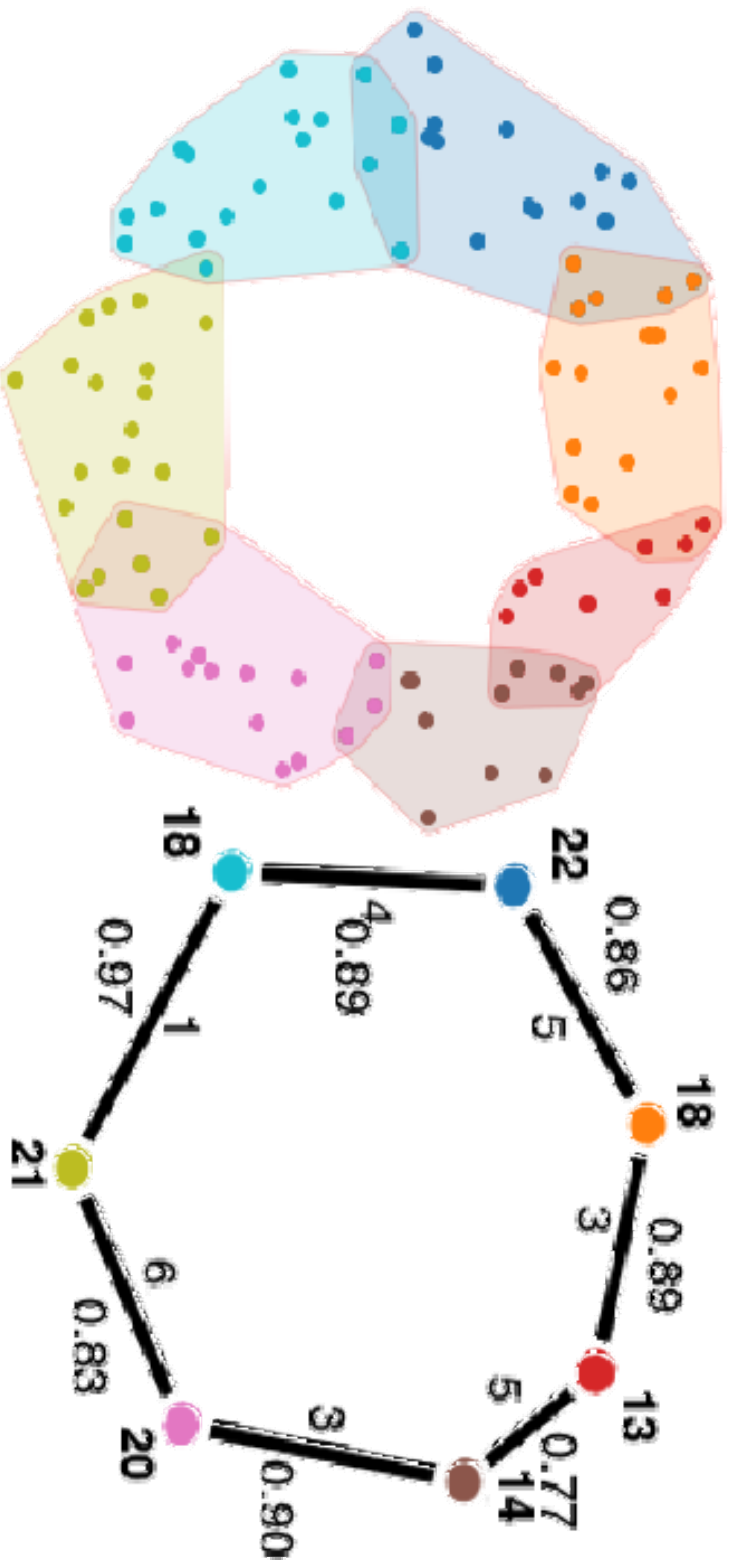
→ $Nrv_{st}(\mathcal{U})$: nerve of cover \mathcal{U} with each simplex σ given $d_{st}(\tau)$ as weight or birth time

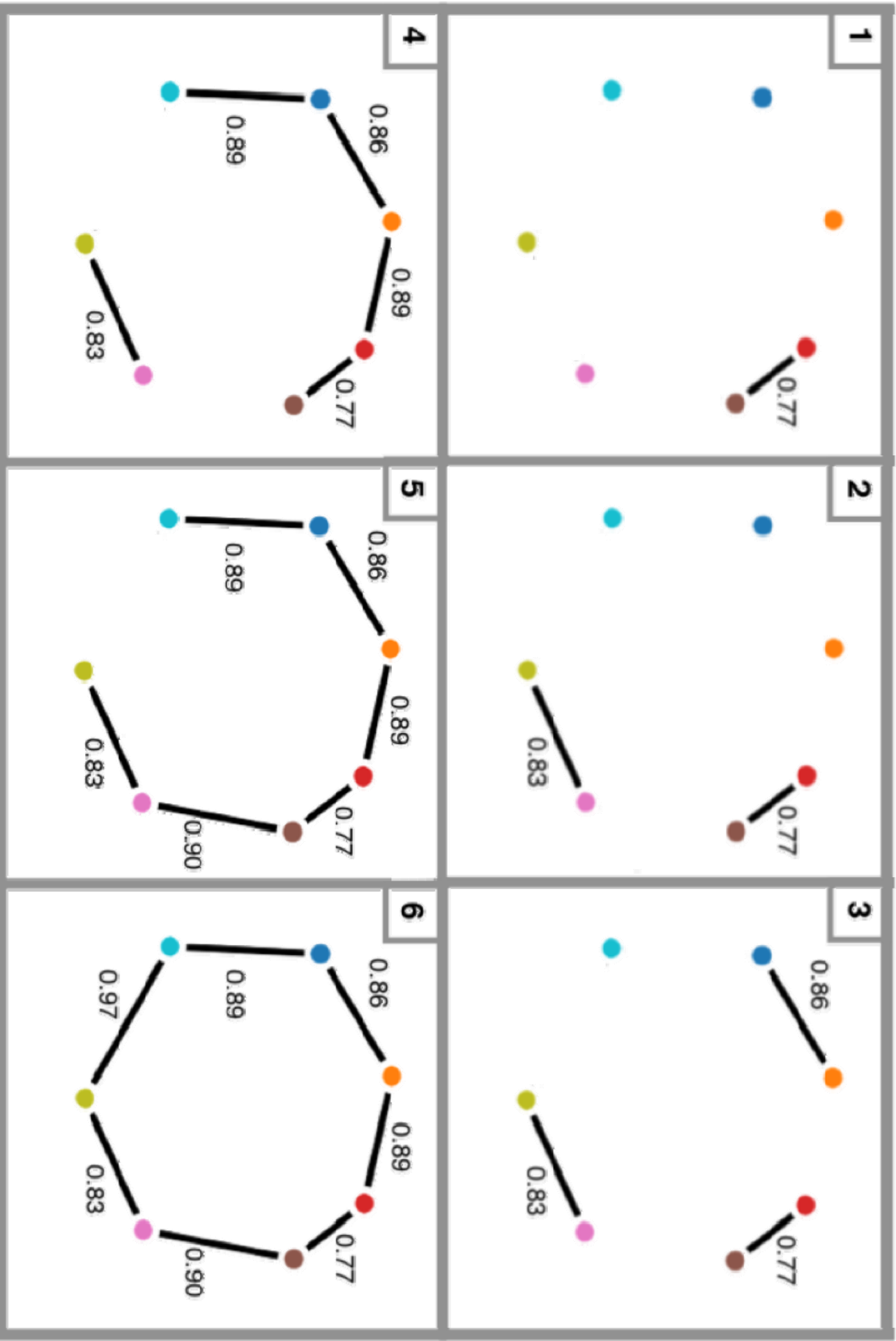
Theorem 1. $Nrv_{st}(\mathcal{U})$ is a filtered simplicial complex.

STEINHAUS NERVE

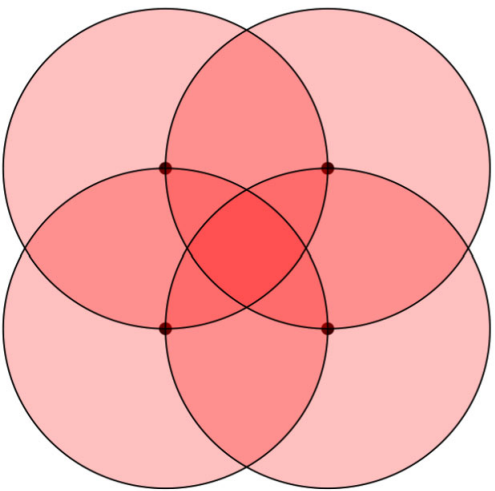
→ $Nrv_{st}(\mathcal{U})$: nerve of cover \mathcal{U} with each simplex σ given $d_{st}(\kappa)$ as weight or birth time

Theorem 1. $Nrv_{st}(\mathcal{U})$ is a filtered simplicial complex.





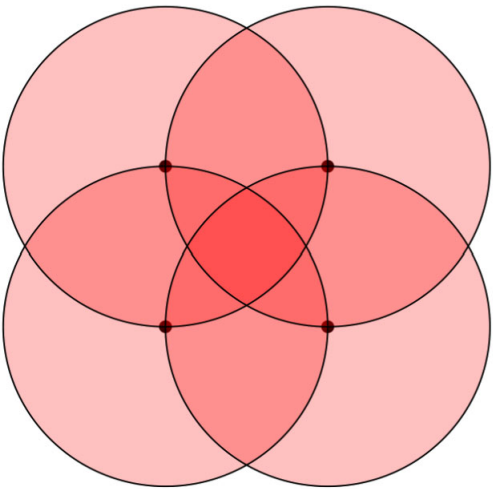
STEINHAUS vs ČECH/VR



$X = \{4 \text{ corners of unit square}\}$

$\mathcal{U} = \{ \text{discs of radius 1 @ each pt} \}$

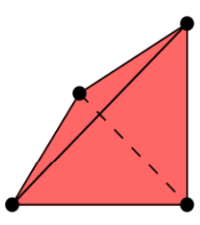
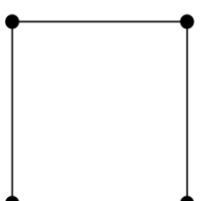
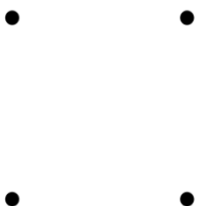
STEINHAUS vs ČECH/VR



$X = \{4 \text{ corners of unit square}\}$

$\mathcal{U} = \{ \text{discs of radius } 1 \text{ @ each pt} \}$

$\check{C}ech(\mathcal{U})$:

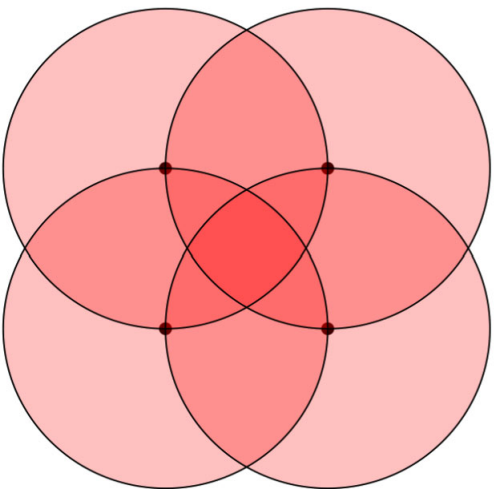


$r = 0$

$r = \frac{1}{2}$

$r = \frac{\sqrt{2}}{2}$

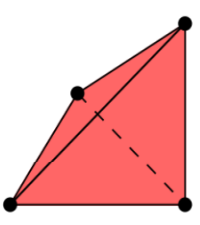
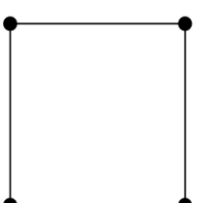
STEINHAUS vs ČECH/VR



$X = \{4 \text{ corners of unit square}\}$

$\mathcal{U} = \{ \text{discs of radius } 1 \text{ @ each pt} \}$

$\check{C}ech(\mathcal{U})$:

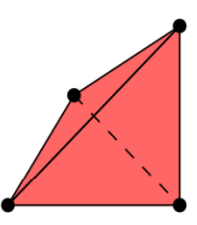
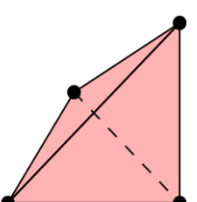
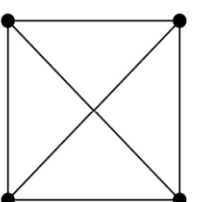
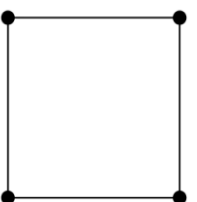
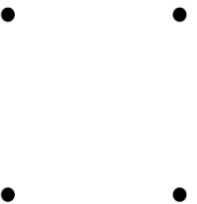


$$r = 0$$

$$r = \frac{1}{2}$$

$$r = \frac{\sqrt{2}}{2}$$

Steinhaus filtration w/ Lebesgue (vol.) measure:



$$d_{St} = 0$$

$$d_{St} \approx 0.756$$

$$d_{St} \approx 0.9$$

$$d_{St} \approx 0.935$$

$$d_{St} \approx 0.959$$

STEINHAUS vs ČECH/ ν_R

Theorem 2. For finite $X \subset \mathbb{R}$ (in 1D), $\check{C}ech(X) \equiv$
Steinhaus filtration on X w/ $\{R\text{-balls centered}$
@ each pt in $X\}$ for $R > \text{diam}(X)$ w/ Lebesgue msr.

STEINHAUS vs ČECH/VR

Theorem 2. For finite $X \subset \mathbb{R}^d$ (in 1D), $\check{C}ech(X) \equiv$
Steinhaus filtration on X w/ $\frac{1}{2}R$ -balls centered
@ each pt in X ? for $R > \text{diam}(X)$ w/ Lebesgue msr.

Lemma 3. VR filtration determines the 1-skeleton of
Steinhaus filtration in arbitrary dimensions.

STABILITY

- stability of q -tame persistence modules
 - Chazal, de Silva, Glisse, Oudot (2016)

STABILITY

→ stability of q -tame persistence modules

— Chazal, de Silva, Glisse, Oudot (2016)

Theorem (Chazal et al, 2009). If U, V are q -tame (i.e, finite rank) persistence modules that are ϵ -interleaved, then $d_B(\text{dgm}(U), \text{dgm}(V)) < \epsilon$

STABILITY

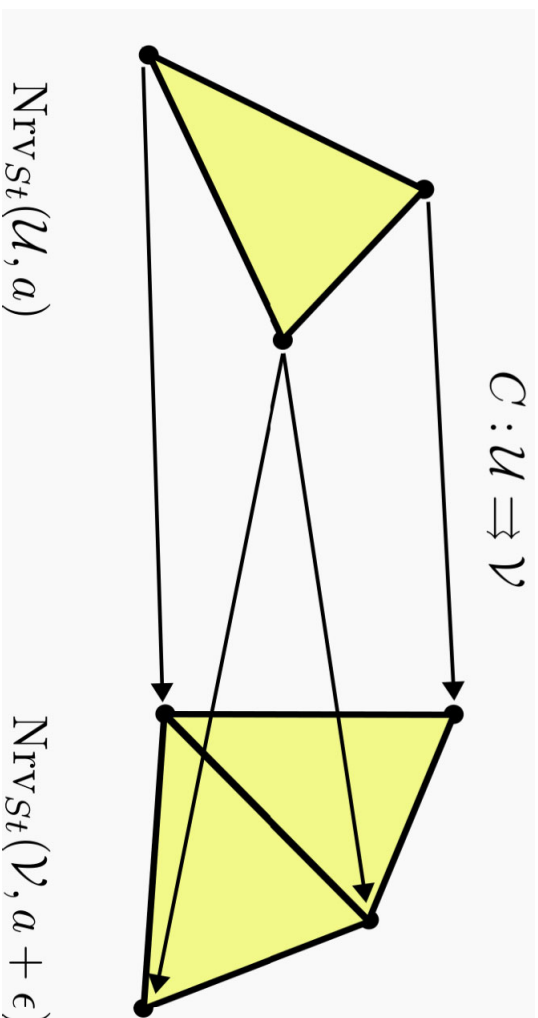
- stability of q -tame persistence modules
 - Chazal, de Silva, Glisse, Oudot (2016)

Theorem (Chazal et al, 2009). If \mathbb{U}, \mathbb{V} are q -tame (i.e. finite rank) persistence modules that are ϵ -interleaved, then $d_B(\text{dgm}(\mathbb{U}), \text{dgm}(\mathbb{V})) < \epsilon$

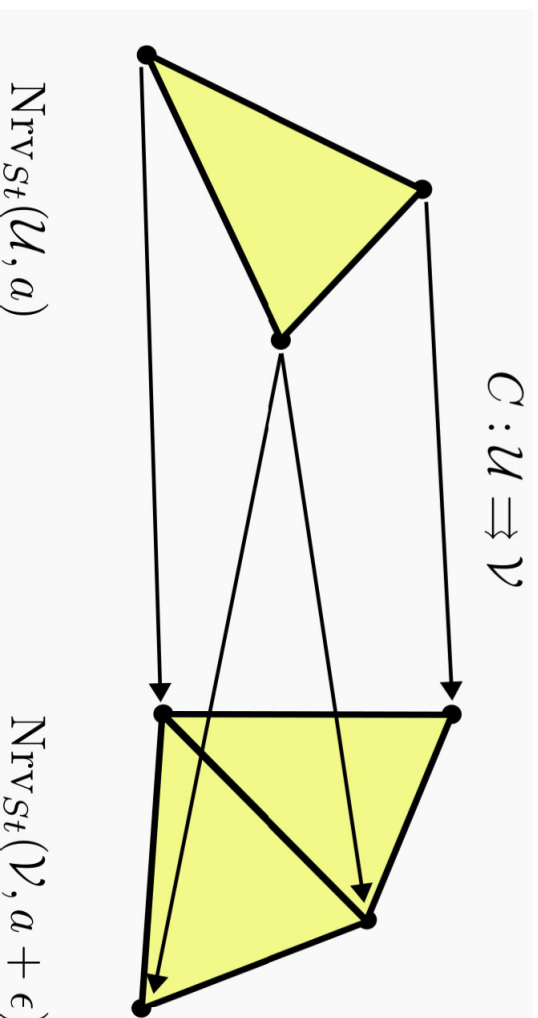
Proposition (Chazal, de Silva, Oudot, 2013).

If $C: X \rightrightarrows Y$ and $C^T: Y \rightrightarrows X$ are ϵ -simplicial multivalued maps, then they induce an ϵ -interleaving between $H_*(\mathbb{U})$ and $H_*(\mathbb{V})$.

OUR PSEUDOMETRIC



OUR PSEUDOMETRIC



→ Find correspondence C w/ smallest distortion

$$\text{dis}(C) = \sup \{ |d_{st}(\{u_i | i \in \sigma\}) - d_{st}(\{v_j | j \in \tau\})| : (v, \tau) \in C \}$$

to define $d(u, v)$

ϵ -INTERLEAVING

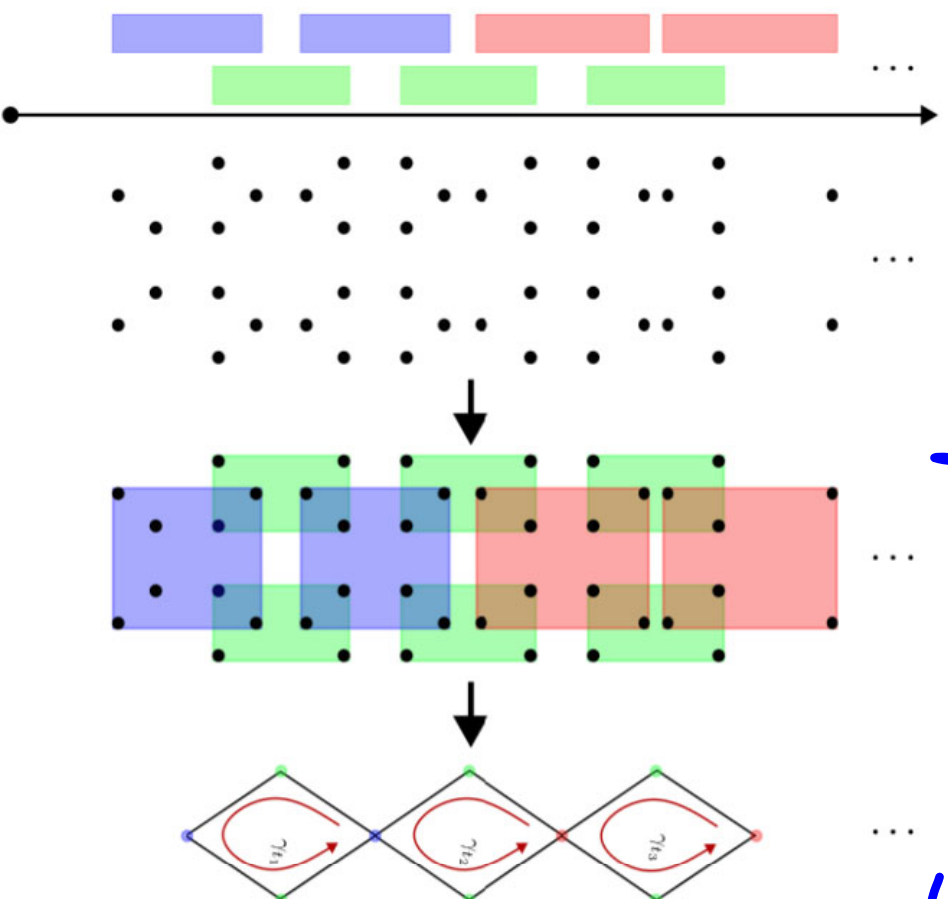
Proposition 4. Let $H^a(\mathcal{U}) := H_*(\text{Nrv}_{\text{St}}(\mathcal{U}, a))$. We can find maps $\phi \in \text{Hom}^\epsilon(H(\mathcal{U}), H(\mathcal{V}))$ & $\psi \in \text{Hom}(H(\mathcal{V}), H(\mathcal{U}))$ s.t.

$$\begin{array}{ccc}
 H^a(\mathcal{U}) & \longrightarrow & H^b(\mathcal{U}) \\
 \searrow \phi_a \swarrow & & \searrow \phi_b \swarrow \\
 H^{a+\epsilon}(\mathcal{V}) & \longrightarrow & H^{b+\epsilon}(\mathcal{V}) \\
 \\
 H^{a+\epsilon}(\mathcal{U}) & \longrightarrow & H^{b+\epsilon}(\mathcal{U}) \\
 \searrow \psi_a \swarrow & & \searrow \psi_b \swarrow \\
 H^a(\mathcal{V}) & \longrightarrow & H^b(\mathcal{V}) \\
 \\
 H^{a-\epsilon}(\mathcal{U}) & \longrightarrow & H^{a+\epsilon}(\mathcal{U}) \\
 \searrow \phi_{a-\epsilon} \swarrow & & \searrow \psi_a \swarrow \\
 H^a(\mathcal{V}) & & \\
 \\
 H^{a-\epsilon}(\mathcal{V}) & \longrightarrow & H^{a+\epsilon}(\mathcal{V}) \\
 \searrow \psi_{a-\epsilon} \swarrow & & \searrow \phi_a \swarrow \\
 H^a(\mathcal{U}) & &
 \end{array}$$

for all d 's $a \leq b$ when $\epsilon > d(\mathcal{U}, \mathcal{V})$.

Non q -TAME!

✓ non- q -tame Steinhaus complex in totally bounded space



STABILITY

Theorem 5. If $H_*(Nv_{St}(u))$ and $H_*(Nv_{St}(v))$ are q -tame, then

$$d_B(d_{gm}(H_*(Nv_{St}(u))), d_{gm}(H_*(Nv_{St}(v)))) \leq d(u, v)$$

STABILITY

Theorem 5. If $H_*(Nrv_{St}(u))$ and $H_*(Nrv_{St}(v))$ are q -tame, then

$$d_B(dgm(H_*(Nrv_{St}(u))), dgm(H_*(Nrv_{St}(v)))) \leq d(u, v)$$

✓ q -tame: holds when u, v finite

MOVIE RECOMMENDATIONS

→ MovieLens-20m: 20mil ratings, 140k users, 27.3k movies



(images: Wikipedia)

MOVIE RECOMMENDATIONS

→ MovieLens-20m: 20mil ratings, 140k users, 27.3k movies



→ Cover movies w/ sets of watchers who rated each movie

(images: Wikipedia)

MOVIE RECOMMENDATIONS

→ MovieLens-20m: 20mil ratings, 140k users, 27.3k movies



→ Cover movies w/ sets of watchers who rated each movie

→ Construct 1-skeleton of Steinhaus filtration w/ counts $C_\mu(v) = 1 \vee 1$

(images: Wikipedia)

STABLE PATHS

→ path P is P -stable if $\max_{e \in P} \{d_{st}(e)\} \leq P$.

STABLE PATHS

→ path P is f -stable if $\max_{e \in P} \{d_{st}(e)\} \leq f$.

→ most stable s - t path:

- f -stable s - t path for smallest f
- shortest path, if multiple paths \exists at same f .

STABLE PATHS

→ path P is f -stable if $\max_{e \in P} \{d_{st}(e)\} \leq f$.

→ most stable s-t path:

- f -stable s-t path for smallest f
- shortest path, if multiple paths \exists at same f .

✓ Stability \Rightarrow paths are automatically stable!

STABLE PATHS

→ path P is P -stable if $\max_{e \in P} \{d_{st}(e)\} \leq P$.

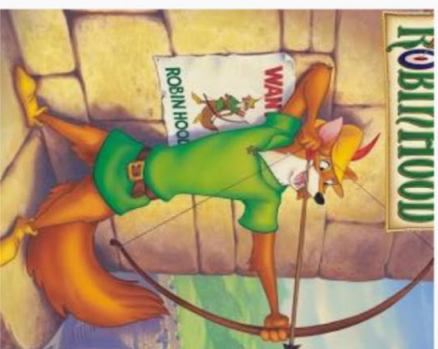
→ most stable s-t path:

- P -stable s-t path for smallest P
- shortest path, if multiple paths \exists at same P .

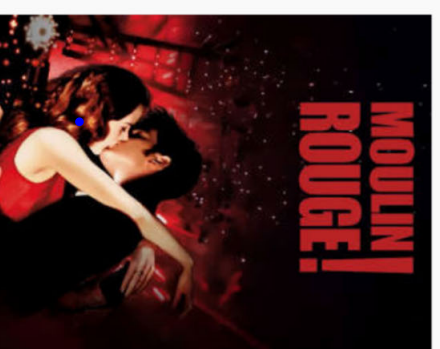
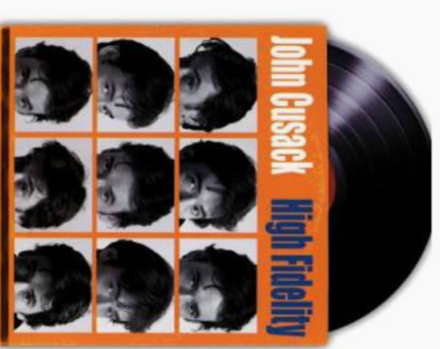
✓ Stability \Rightarrow paths are automatically stable!

→ short v/s stable: pareto-frontier

MULAN → MOULIN ROUGE



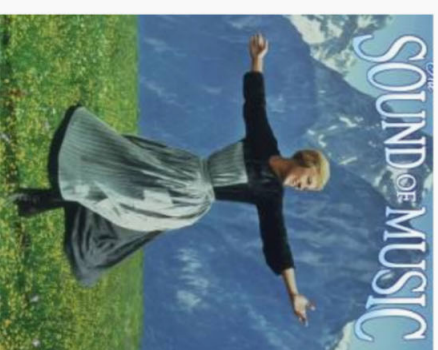
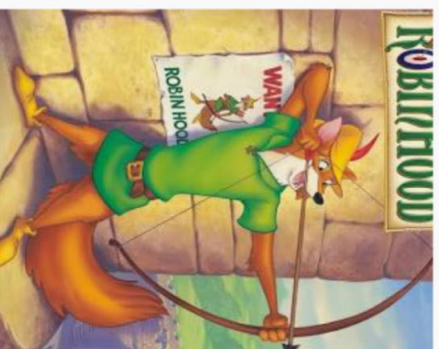
•



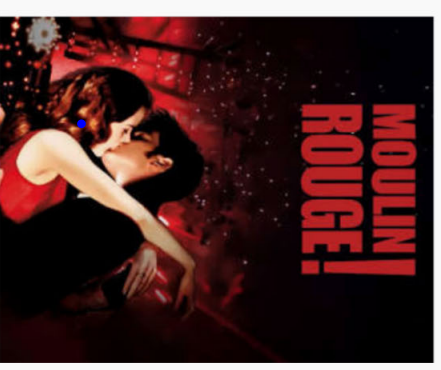
(images: WWWW)

MULAN → MOULIN ROUGE

shortest path



•



(images: WWWW)

EXPLANATIONS IN ML MODEL

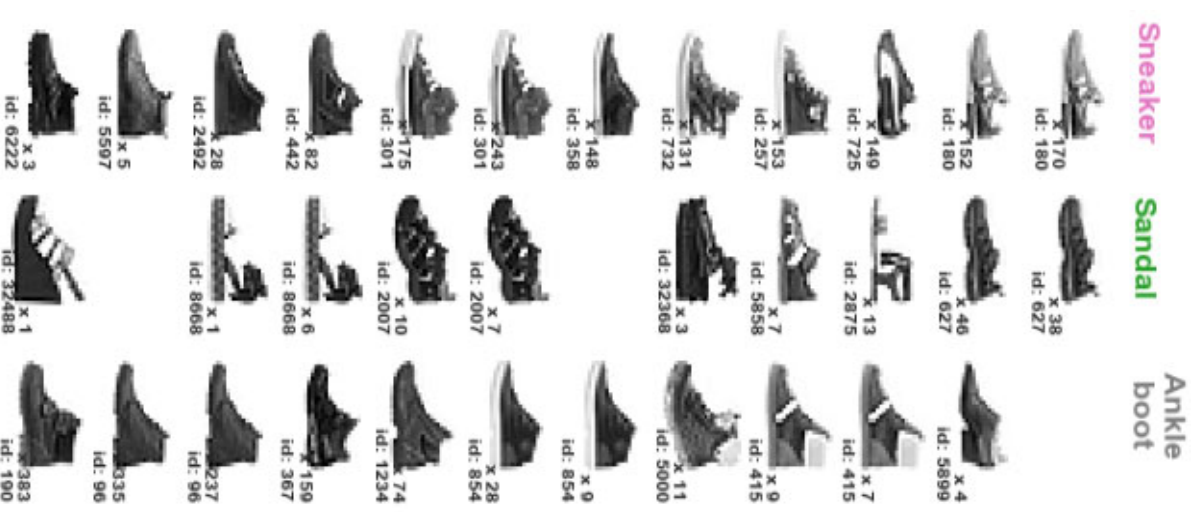
- Fashion-MNIST (clothing)
 - 70000 images
 - 28×28 pixels
 - 10 classes

EXPLANATIONS IN ML MODEL

- Fashion-MNIST (clothing)
 - 7000 images
 - 28×28 pixels
 - 10 classes
- logistic regression predictor (93% acc.)
- 10-dim prob. UMAP → 2d filter
- Steinhaus filtration on Mapper
- study stable paths

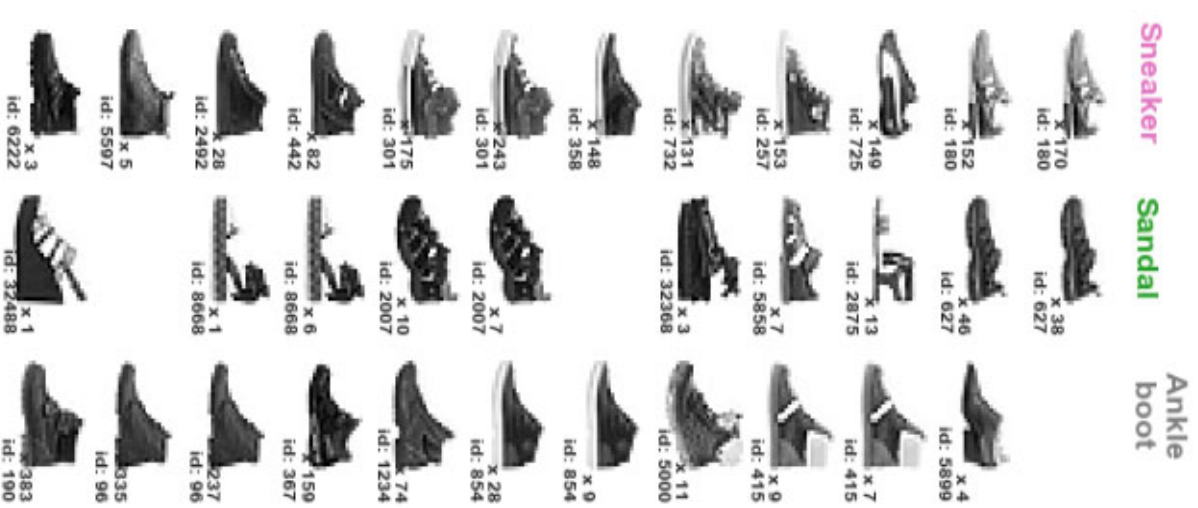
EXPLANATIONS IN ML MODEL

- Fashion-MNIST (clothing)
 - 70000 images
 - 28×28 pixels
 - 10 classes
- logistic regression predictor (93% acc.)
- 10-dim prob. UMAP → 2d filter
- Steinhaus filtration on Mapper
- study stable paths



EXPLANATIONS IN ML MODEL

- Fashion-MNIST (clothing)
 - 70000 images
 - 28×28 pixels
 - 10 classes
- logistic regression predictor (93% acc.)
- 10-dim prob. UMAP → 2d filter
- Steinhaus filtration on Mapper
- study stable paths
- Rathore et al. (2021)



OPEN QUESTIONS

? Multiparameter persistence for abstract covers?

OPEN QUESTIONS

- ? Multiparameter persistence for abstract covers?
- ? Movie recs with ratings?

OPEN QUESTIONS

- ? Multiparameter persistence for abstract covers?
- ? Movie recs with ratings?
- ? p -stable loops? p -stable flares?

OPEN QUESTIONS

- ? Multiparameter persistence for abstract covers?
- ? Movie recs with ratings?
- ? P -stable loops? P -stable flares?

Thank you!