Steinhaus Filtration: Movie Recs using Topology!

Bala Krishnamoorthy
Washington State University

with D. Arendt (PNNL), M. Broussard, N. Saul, A. Thrall

SoCG 2025

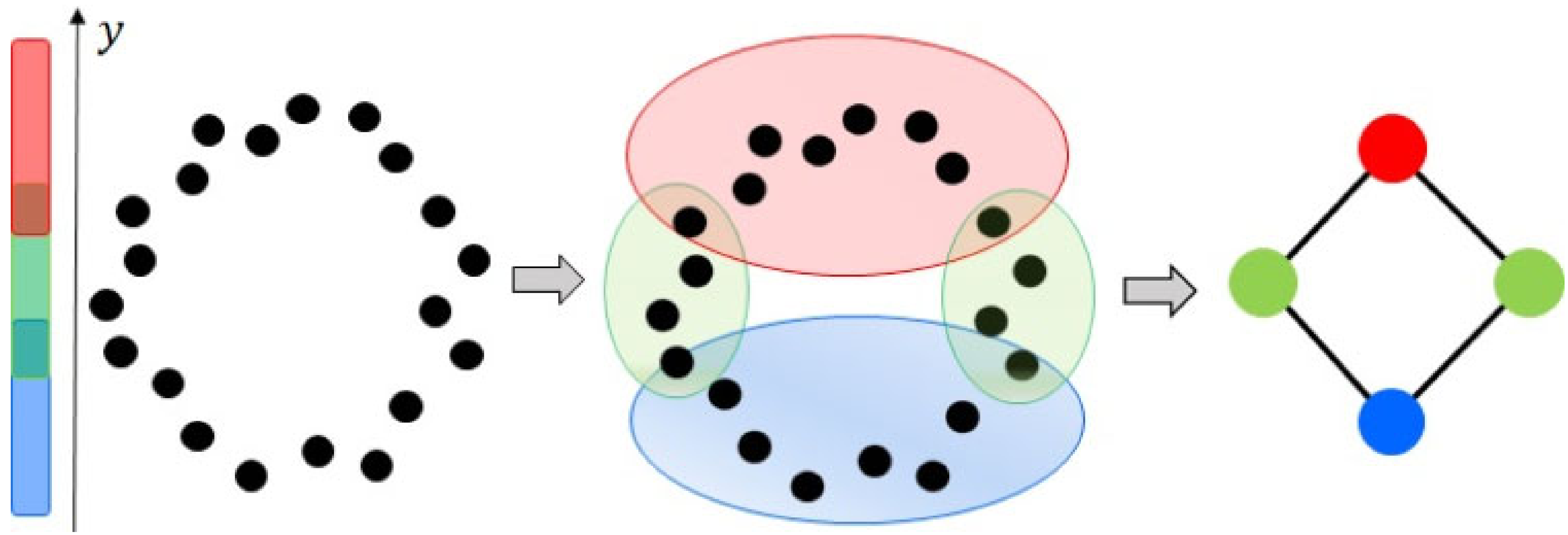
Movie Recommendations

?

(images: Wikipedia)

MAPPER

Singh, Mémoli, Carlsson (2007)



lens → cover → pullback → refine → nerve

MAPPER: APPLICATIONS

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

MAPPER: APPLICATIONS

- → numerous areas in last decade
  biomedicine, criminology, machine learning, ...
- → use 1-skeleton for interpretation
- → 1D 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, Mémoli, Wang (2016): Multiscale mapper
  - tower of covers

→ Carrière and Oudot (2017): Stability of 1D mapper
  - multinerve mapper → extended persistence diagram

MAPPER: STABILITY

→ use framework of persistence
  - stability of persistence modules

→ Dey, Mémoli, Wang (2016): Multiscale mapper
  - tower of covers

→ Carrière and Oudot (2017): Stability of 1D mapper
  - multinerve 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

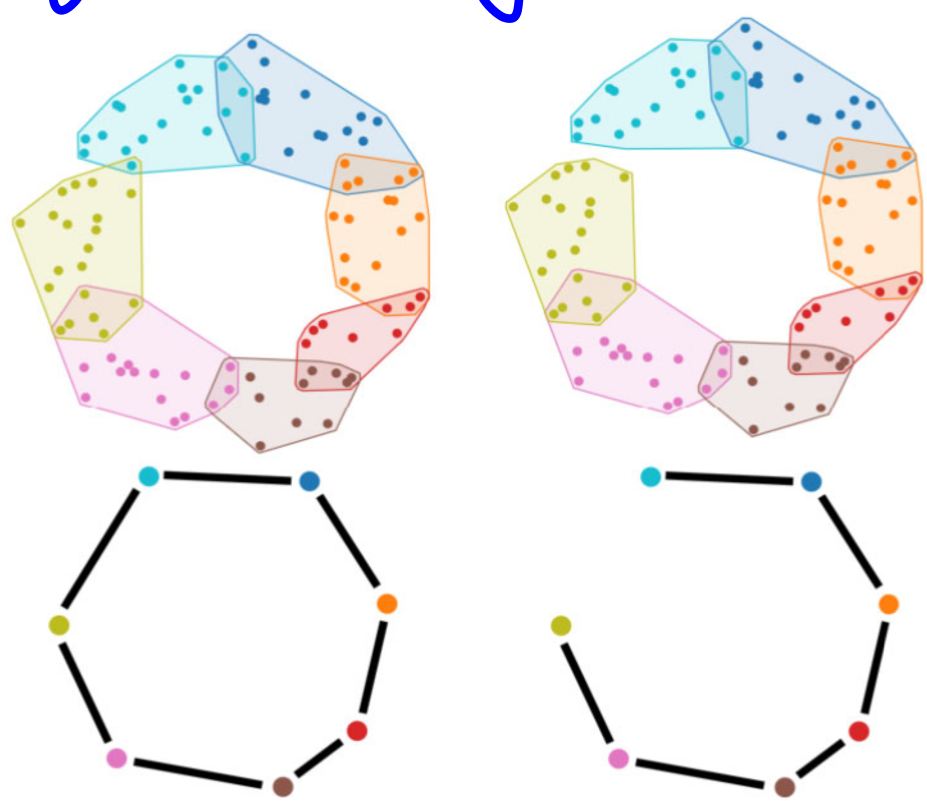
? Define a mapper on abstract covers?

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 Jaccard distance
- stability (for finite case)

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

RESULTS

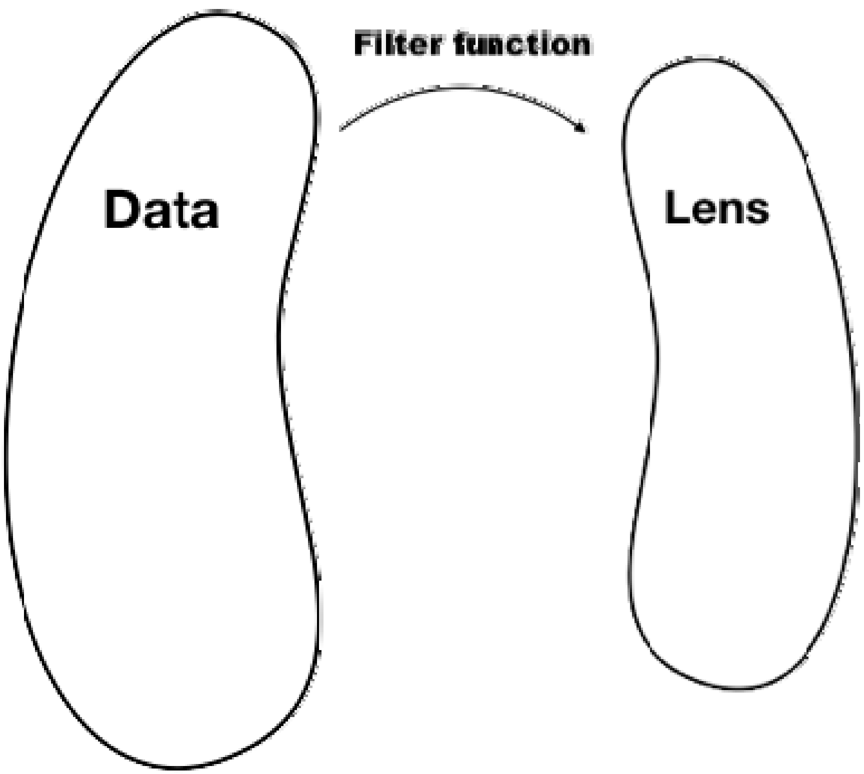
✓ Steinhaus filtration from a single abstract cover
- generalized Jaccard distance
- stability (for finite case)

✓ stable paths in a single mapper
- stability measures strength of overlap

✓ Applications
- movie recommendations
- explainable machine learning

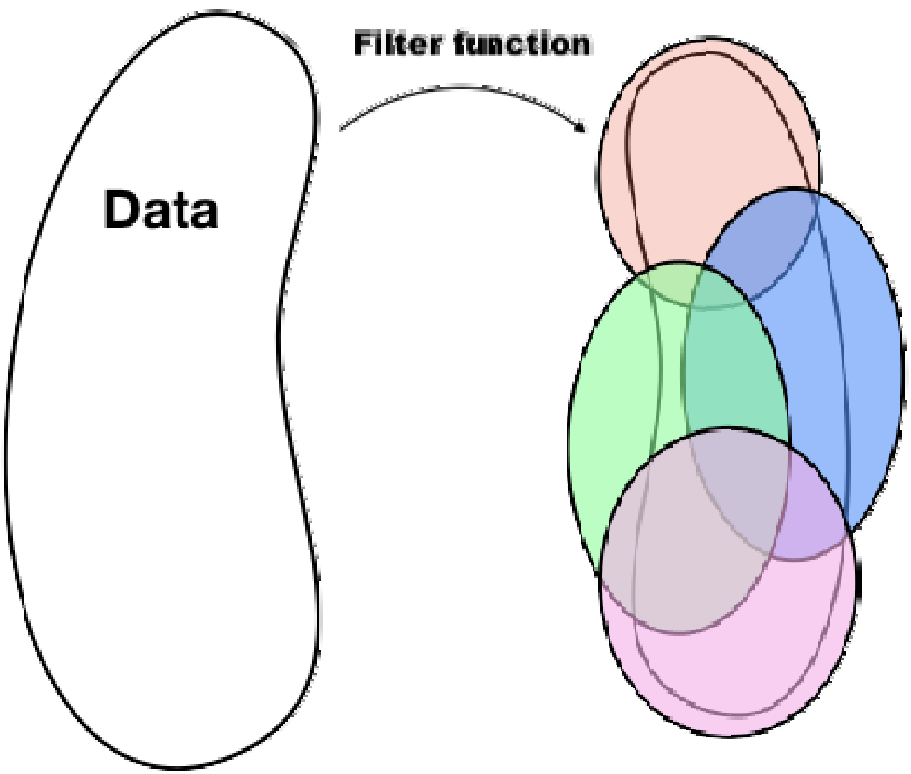
MAPPER

lens $\to$ cover $\to$ pullback $\to$ refine $\to$ nerve



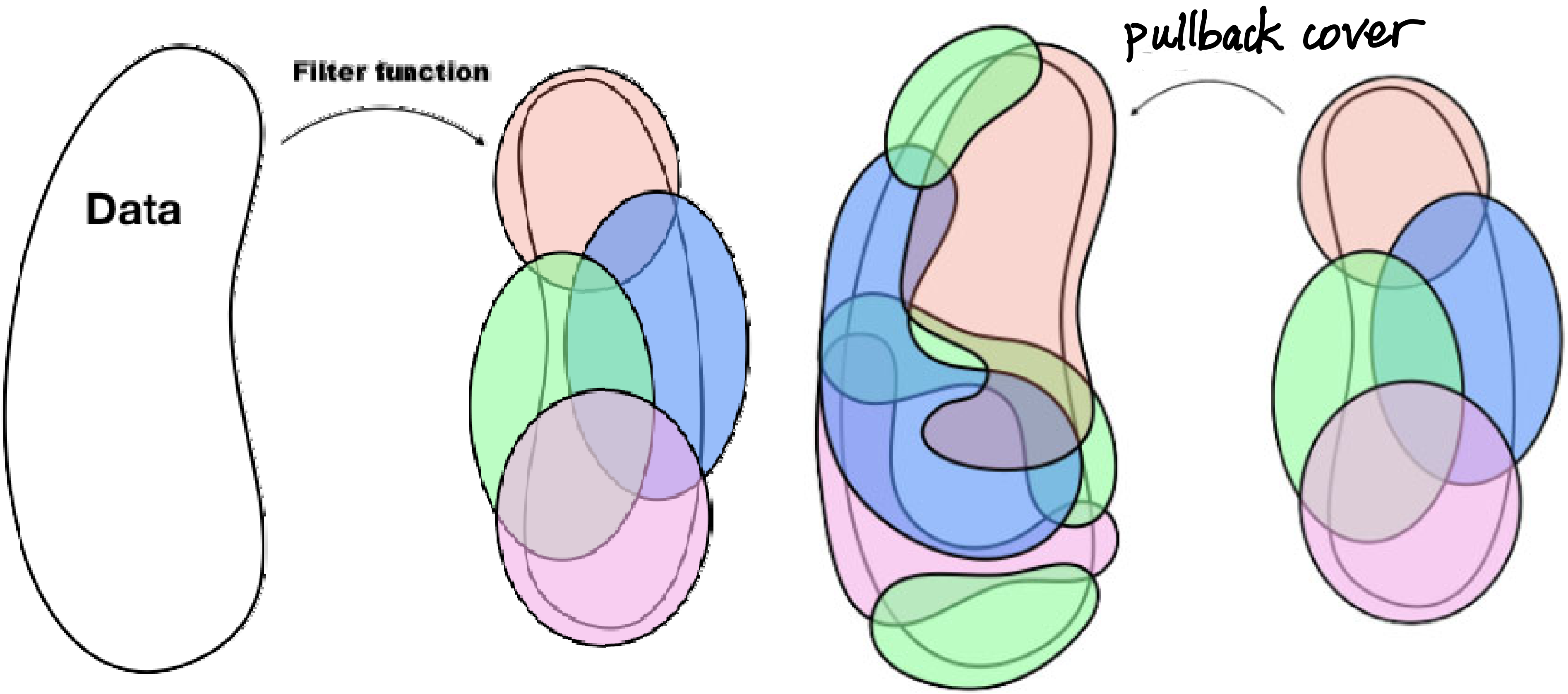
MAPPER

lens → <u>cover</u> → pullback → refine → nerve



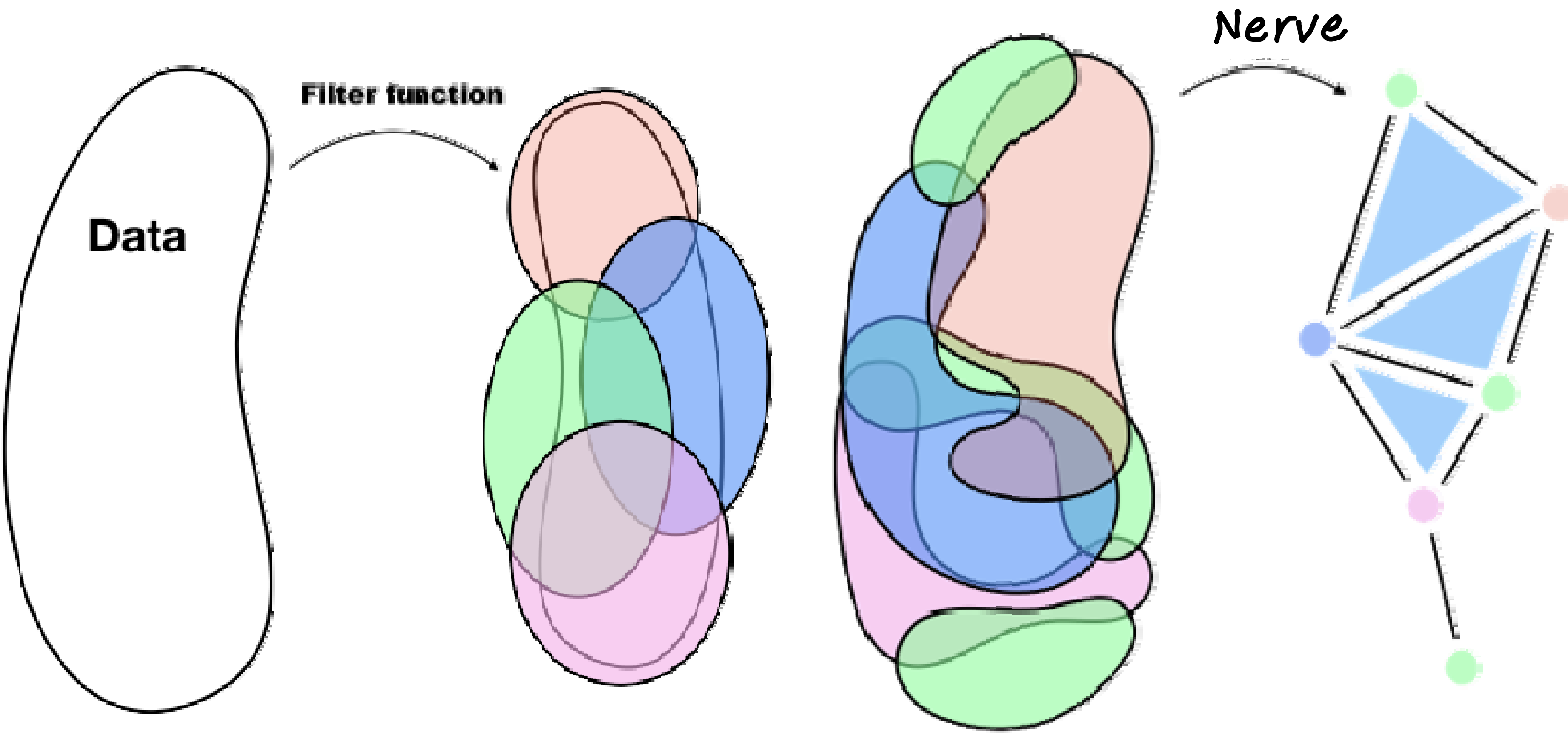
MAPPER

lens $\to$ cover $\to$ pullback $\to$ refine $\to$ nerve

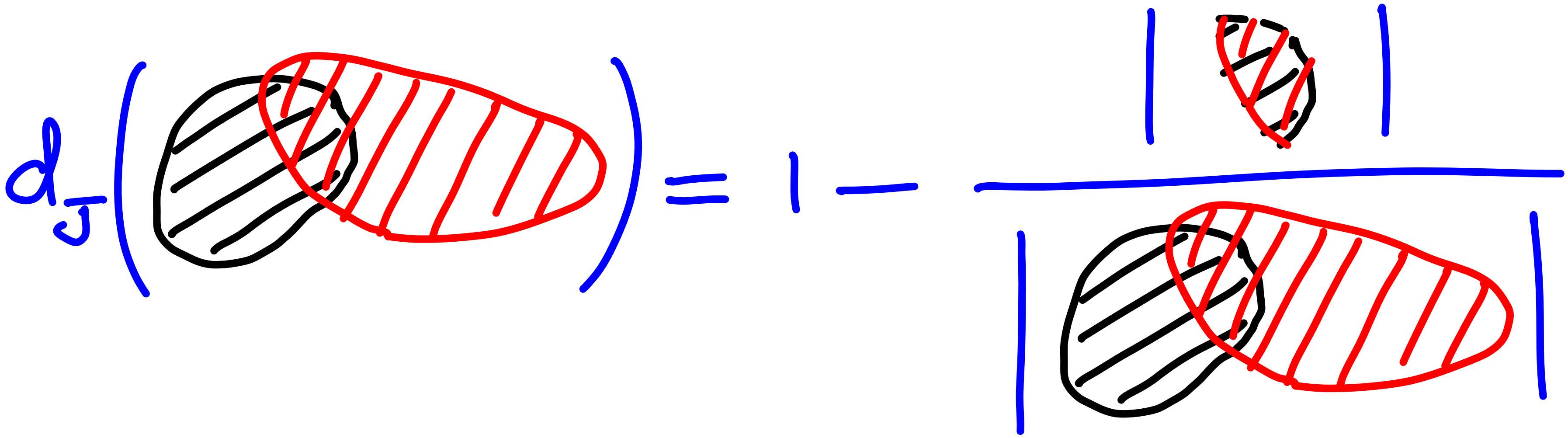


MAPPER

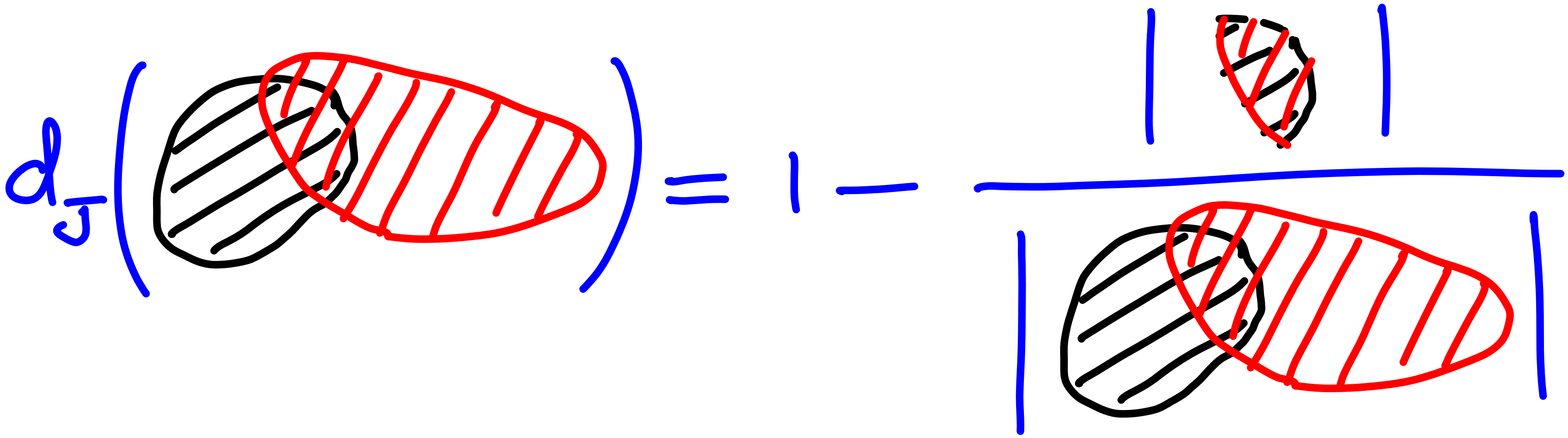
lens $\to$ cover $\to$ pullback $\to$ refine $\to$ <u>nerve</u>



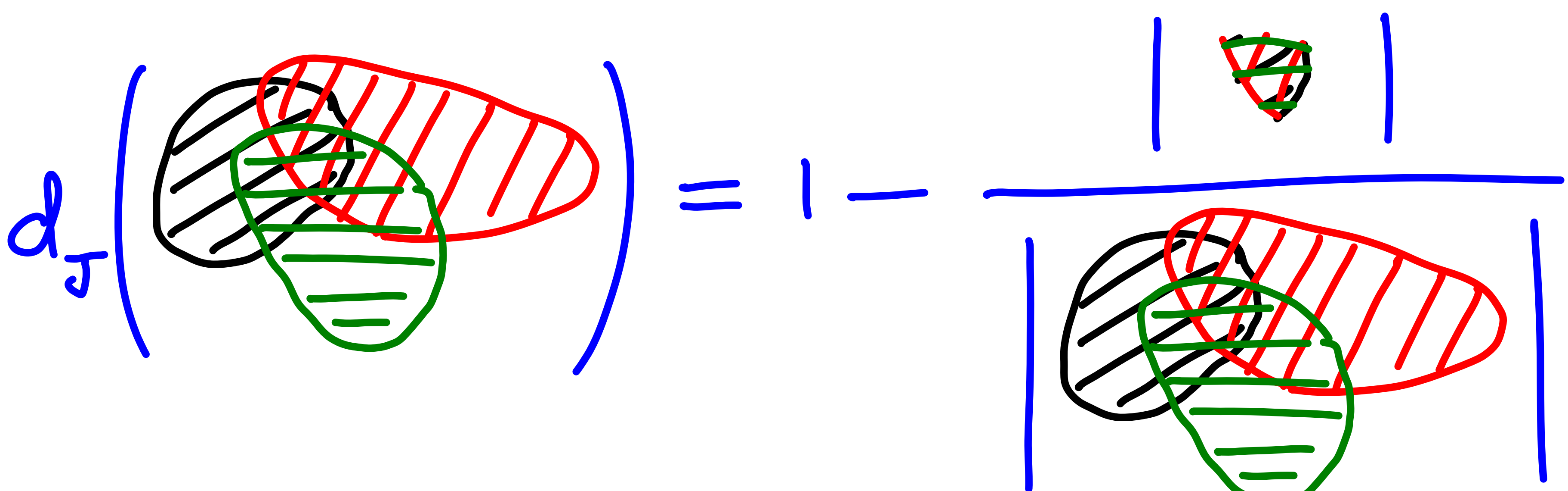
JACCARD DISTANCE



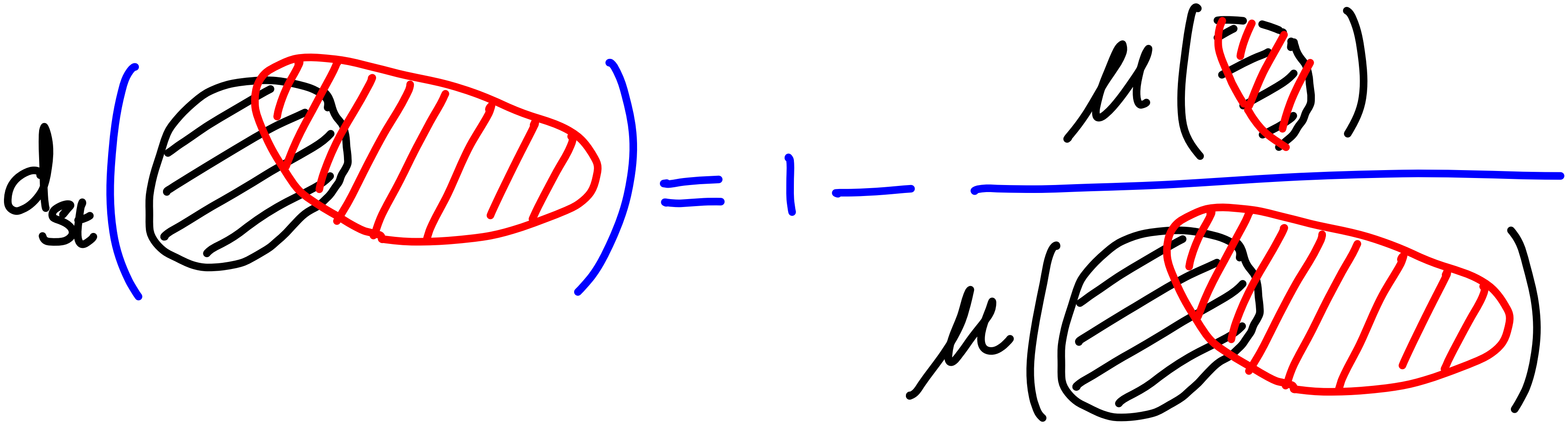
JACCARD DISTANCE



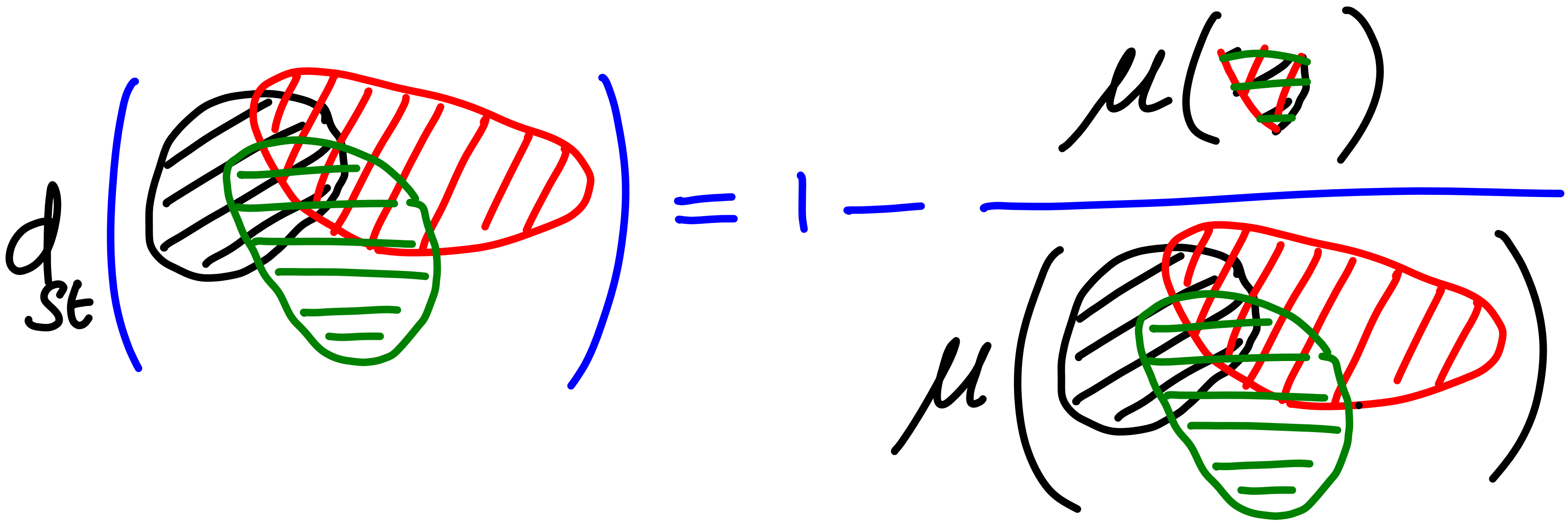
Generalized Jaccard distance:



STEINHAUS DISTANCE



Generalized Steinhaus distance: (μ: measure)



Steinhaus Nerve

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

STEINHAUS NERVE

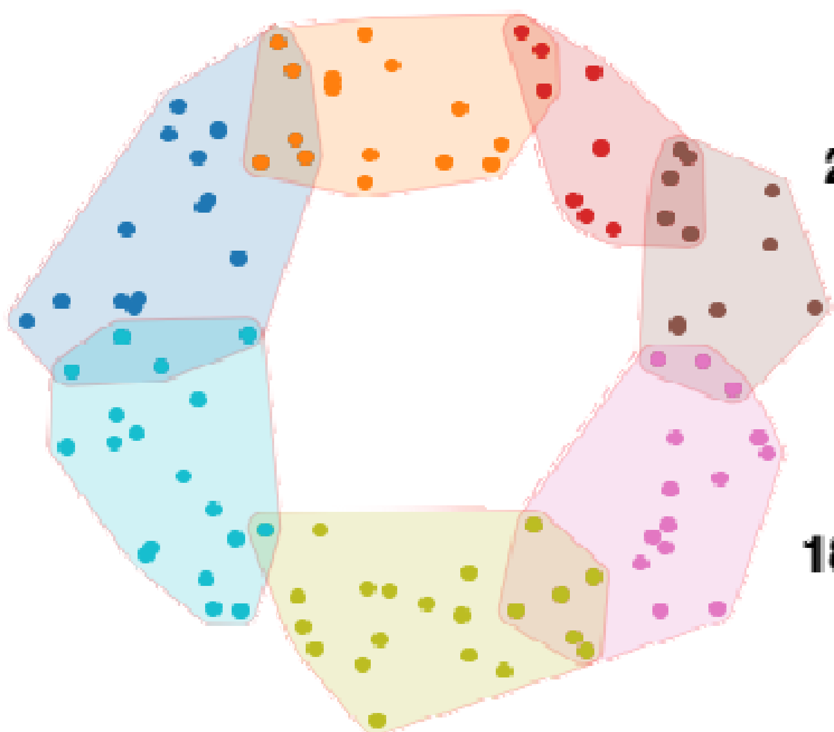
→ $\mathrm{Nrv}_{st}(\mathcal{U})$: nerve of cover $\mathcal{U}$ with each simplex σ given $d_{st}(\sigma)$ as weight or birth time

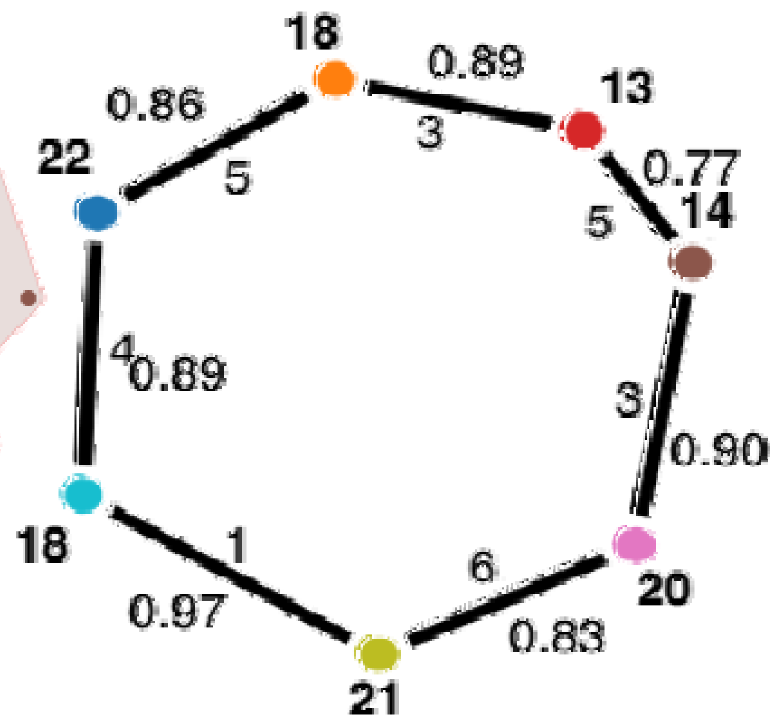
Theorem 1. $\mathrm{Nrv}_{st}(\mathcal{U})$ is a filtered simplicial complex.

STEINHAUS NERVE

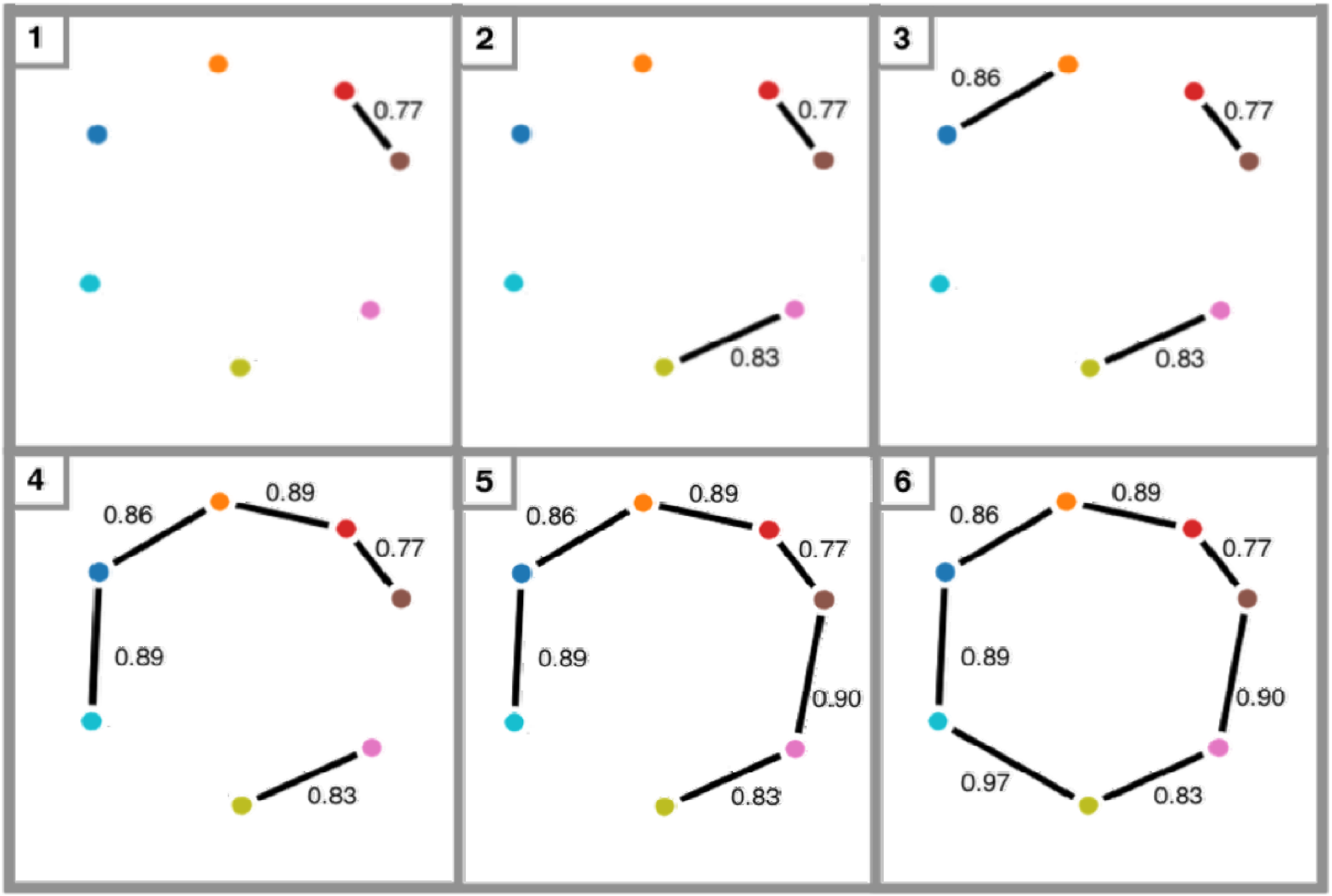
→ $\mathrm{Nrv}_{st}(\mathcal{U})$: nerve of cover $\mathcal{U}$ with each simplex σ given $d_{st}(\sigma)$ as weight or birth time

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

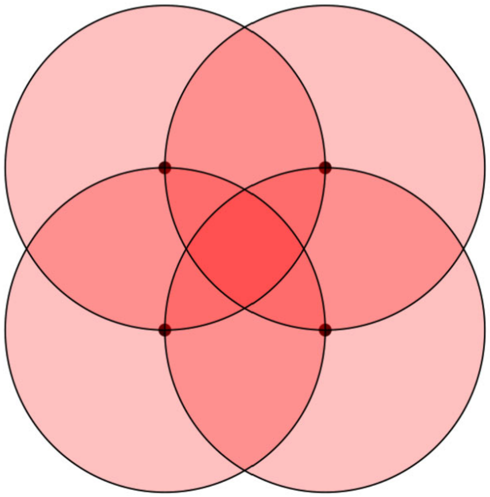




Steinhaus Filtration



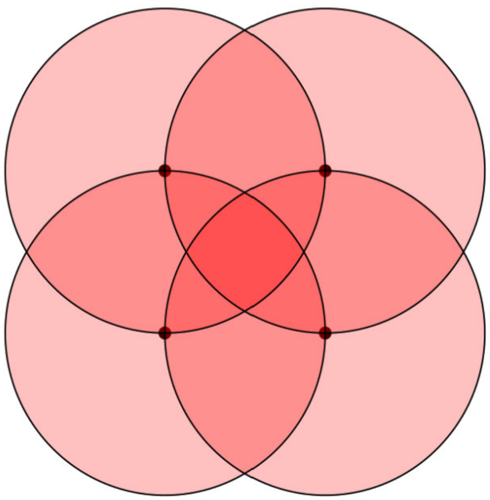
STEINHAUS vs ČECH/VR



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

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

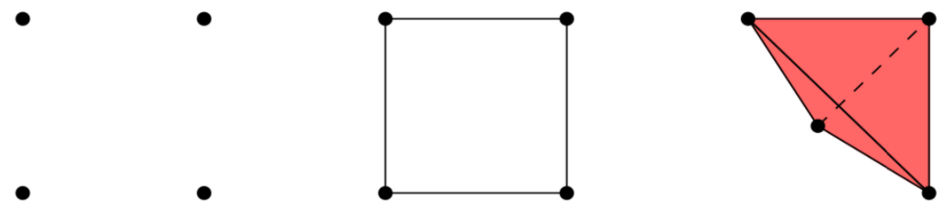
STEINHAUS vs ČECH/VR



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

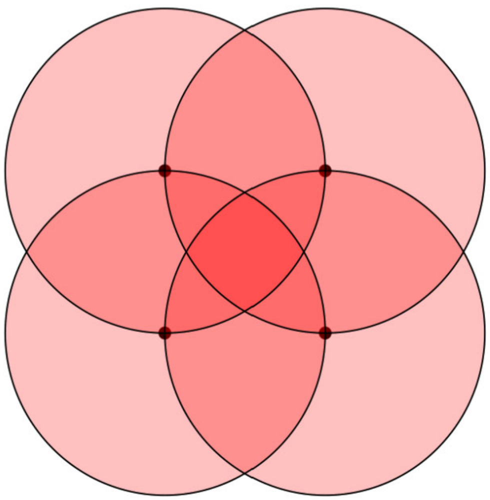
$\mathcal{U} = \{\text{discs of radius } 1 \text{ @ each pt}\}$

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



$r = 0$ $\qquad r = \frac{1}{2}$ $\qquad r = \frac{\sqrt{2}}{2}$

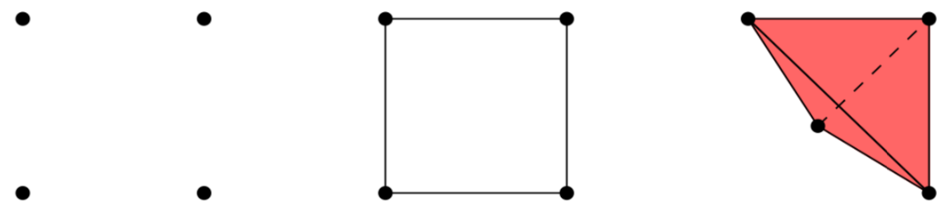
STEINHAUS vs ČECH/VR



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

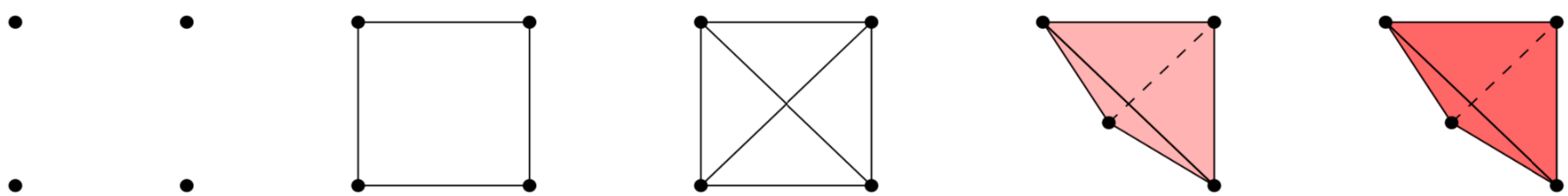
$\mathcal{U} = \{\text{discs of radius } 1 \text{ @ each pt}\}$

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



$r = 0$ $\quad r = \frac{1}{2}$ $\quad r = \frac{\sqrt{2}}{2}$

Steinhaus filtration w/ Lebesgue (vol.) measure:



$d_{St} = 0$ $\quad d_{St} \approx 0.756$ $\quad d_{St} \approx 0.9$ $\quad d_{St} \approx 0.935$ $\quad d_{St} \approx 0.959$

STEINHAUS vs ČECH/VR

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

STEINHAUS vs ČECH/VR

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 $R > diam(X)$ 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 $\mathbb{U}, \mathbb{V}$ are q-tame (i.e., finite rank) persistence modules that are ϵ-interleaved, then $d_B(\mathrm{dgm}(\mathbb{U}), \mathrm{dgm}(\mathbb{V})) \le \epsilon$

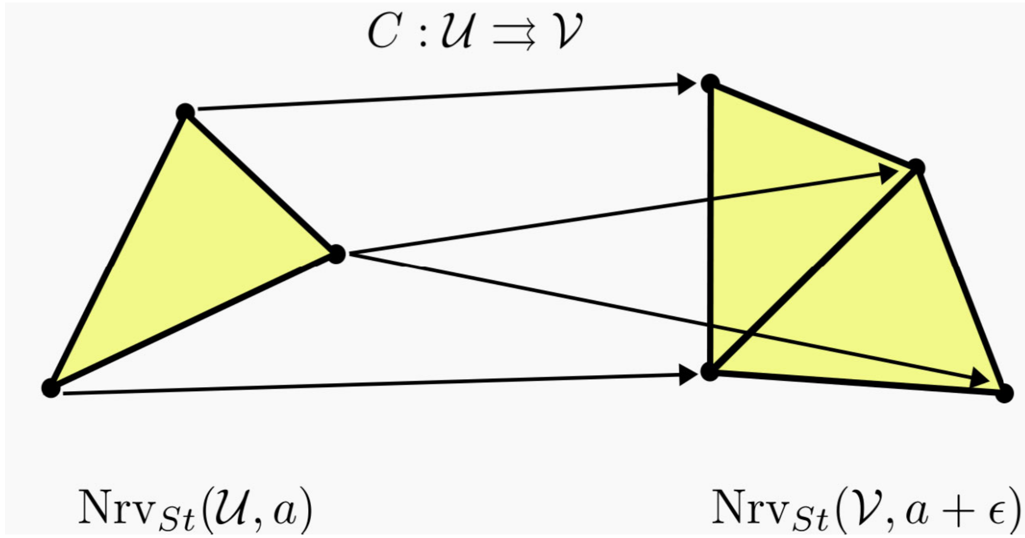
STABILITY

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

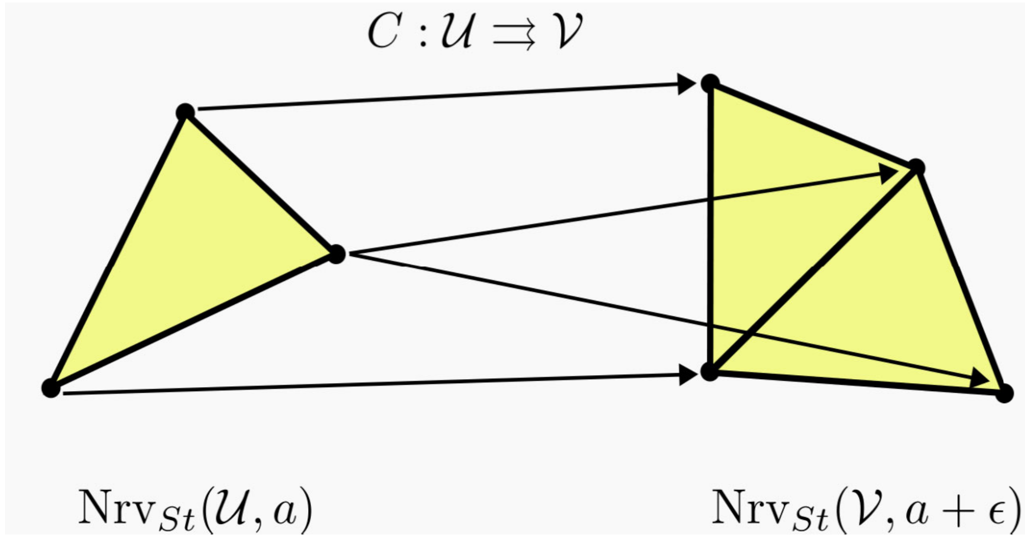
<u>Theorem</u> (Chazal et al., 2009). If $\mathbb{U}, \mathbb{V}$ are <u>q-tame</u> (i.e., finite rank) persistence modules that are <u>ϵ-interleaved</u>, then $d_B(\mathrm{dgm}(\mathbb{U}), \mathrm{dgm}(\mathbb{V})) < \epsilon$

<u>Proposition</u> (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

$\mathrm{dis}(C) =$

$$\sup\{|d_{st}(\{U_i \mid i \in \sigma\}) - d_{st}(\{V_j \mid j \in \tau\})| : (\sigma, \tau) \in C\}$$

to define $d(\mathcal{U}, \mathcal{V})$

ϵ-INTERLEAVING

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

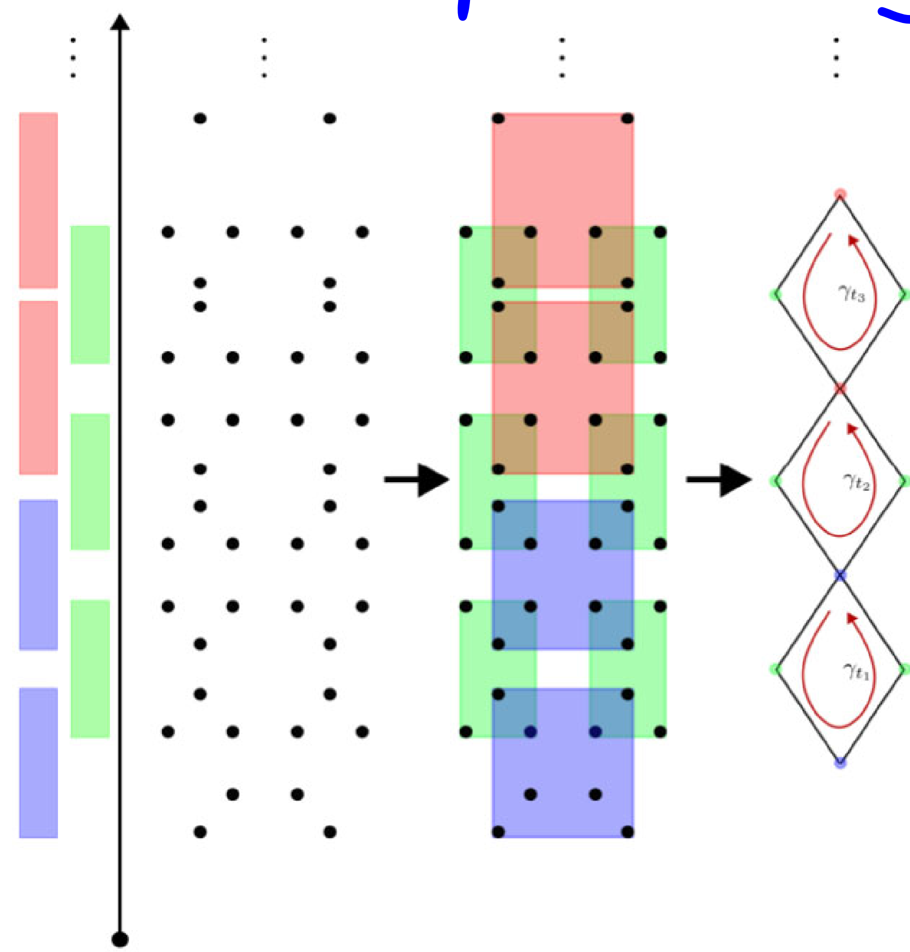
$$\begin{tikzcd}
H^a(\mathcal{U}) \arrow[rr] \arrow[dr, "\phi_a"] & & H^b(\mathcal{U}) \arrow[dr, "\phi_b"] & \\
& H^{a+\epsilon}(\mathcal{V}) \arrow[rr] & & H^{b+\epsilon}(\mathcal{V})
\end{tikzcd}
\qquad
\begin{tikzcd}
H^{a-\epsilon}(\mathcal{U}) \arrow[rr] \arrow[dr, "\phi_{a-\epsilon}"] & & H^{a+\epsilon}(\mathcal{U}) \\
& H^a(\mathcal{V}) \arrow[ur, "\psi_a"] &
\end{tikzcd}$$

$$\begin{tikzcd}
& H^{a+\epsilon}(\mathcal{U}) \arrow[rr] & & H^{b+\epsilon}(\mathcal{U}) \\
H^a(\mathcal{V}) \arrow[rr] \arrow[ur, "\psi_a"] & & H^b(\mathcal{V}) \arrow[ur, "\psi_b"] &
\end{tikzcd}
\qquad
\begin{tikzcd}
& H^a(\mathcal{U}) \arrow[dr, "\phi_a"] & \\
H^{a-\epsilon}(\mathcal{V}) \arrow[rr] \arrow[ur, "\psi_{a-\epsilon}"] & & H^{a+\epsilon}(\mathcal{V})
\end{tikzcd}$$

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

NON q-TAME!

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



STABILITY

<u>Theorem 5.</u> If $H_*(\mathrm{Nrv}_{st}(\mathcal{U}))$ and $H_*(\mathrm{Nrv}_{st}(\mathcal{V}))$ are q-tame, then

$$d_B\big(\mathrm{dgm}(H_*(\mathrm{Nrv}_{st}(\mathcal{U}))), \mathrm{dgm}(H_*(\mathrm{Nrv}_{st}(\mathcal{V})))\big) \leq d(\mathcal{U},\mathcal{V})$$

STABILITY

<u>Theorem 5.</u> If $H_*(\mathrm{Nrv}_{st}(\mathcal{U}))$ and $H_*(\mathrm{Nrv}_{st}(\mathcal{V}))$ are q-tame, then

$$d_B\big(\mathrm{dgm}(H_*(\mathrm{Nrv}_{st}(\mathcal{U}))), \mathrm{dgm}(H_*(\mathrm{Nrv}_{st}(\mathcal{V})))\big) \leq d(\mathcal{U},\mathcal{V})$$

✓ q-tame: holds when $\mathcal{U}, \mathcal{V}$ finite

MOVIE RECOMMENDATIONS

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

→

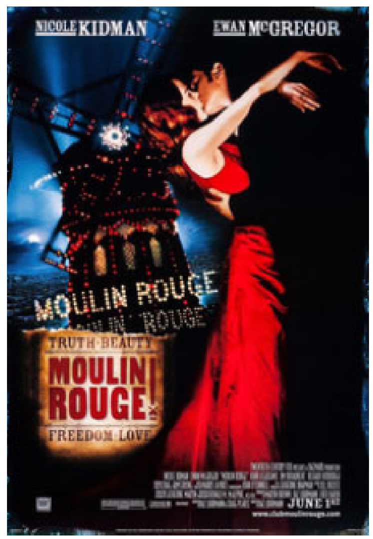
(images: Wikipedia)

Movie Recommendations

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

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

→



(images: Wikipedia)

MOVIE RECOMMENDATIONS

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

→

(images: Wikipedia)

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

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

STABLE PATHS

→ path P is <u>ρ-stable</u> if $\max_{e \in P} \{d_{st}(e)\} \leq \rho$.

STABLE PATHS

→ path P is <u>ρ-stable</u> if $\max_{e \in P} \{d_{st}(e)\} \leq \rho$.

→ most stable s-t path:

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

STABLE PATHS

→ path P is <u>ρ-stable</u> if $\max_{e \in P} \{d_{st}(e)\} \leq \rho$.

→ most stable s-t path:

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

✓ stability $\Rightarrow$ paths are automatically stable!

STABLE PATHS

→ path P is <u>ρ-stable</u> if $\max_{e \in P} \{d_{st}(e)\} \leq \rho$.

→ most stable s-t path:
- ρ-stable s-t path for smallest ρ
- shortest path, if multiple paths $\exists$ at same ρ.

✓ stability $\Rightarrow$ paths are automatically stable!

→ short v/s stable : pareto-frontier

Mulan → Moulin Rouge

(images: WWW)

MULAN → MOULIN ROUGE

shortest path

(images: WWW)

EXPLANATIONS IN ML MODEL

→ Fashion-MNIST (clothing)
- 70,000 images
- 28×28 pixels
- 10 classes

EXPLANATIONS IN ML MODEL

→ Fashion-MNIST (clothing)
- 70,000 images
- 28×28 pixels
- 10 classes

→ logistic regression predictor (93% acc.)

→ 10-dim prob. $\xrightarrow{\text{UMAP}}$ 2d filter

→ Steinhaus filtration on Mapper

→ study stable paths

EXPLANATIONS IN ML MODEL

→ Fashion-MNIST (clothing)
- 70,000 images
- 28×28 pixels
- 10 classes

→ logistic regression predictor (93% acc.)

→ 10-dim prob. $\xrightarrow{\text{UMAP}}$ 2d filter

→ Steinhaus filtration on Mapper

→ study stable paths

EXPLANATIONS IN ML MODEL

- → Fashion-MNIST (clothing)
  - 70,000 images
  - 28×28 pixels
  - 10 classes
- → logistic regression predictor (93% acc.)
- → 10-dim prob. $\xrightarrow{\text{UMAP}}$ 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 rec's with ratings?

OPEN QUESTIONS

? Multiparameter persistence for abstract covers?

? Movie rec's with ratings?

? ρ-stable loops? ρ-stable flares?

OPEN QUESTIONS

? Multiparameter persistence for abstract covers?

? Movie rec's with ratings?

? ρ-stable loops? ρ-stable flares?

Thank You!