Persistent Homology

(Or how I learn this new data analysis methodology in 2020)

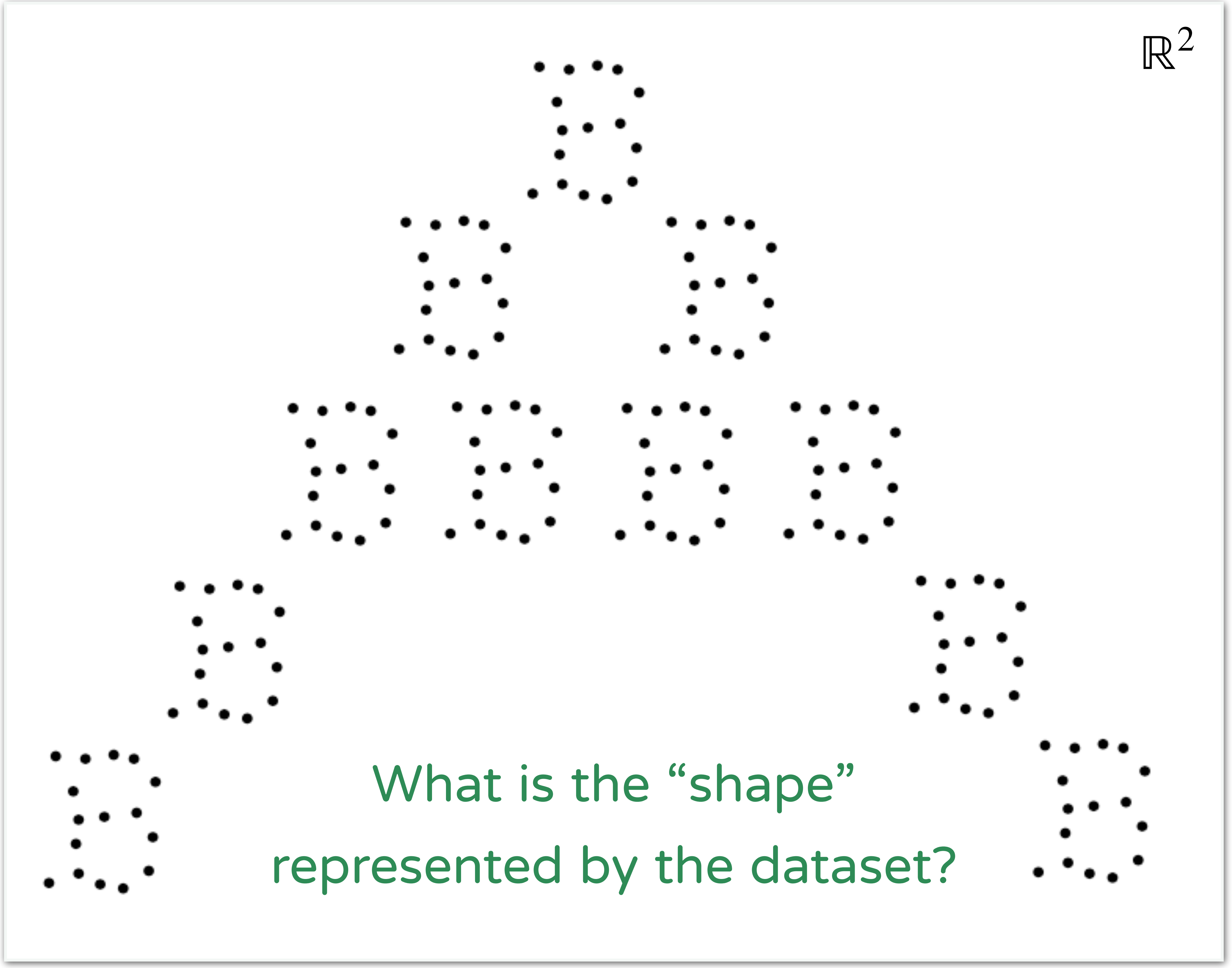
Tzu-Chi Yen

MATH 6220: Introduction to Topology II

Class Presentation

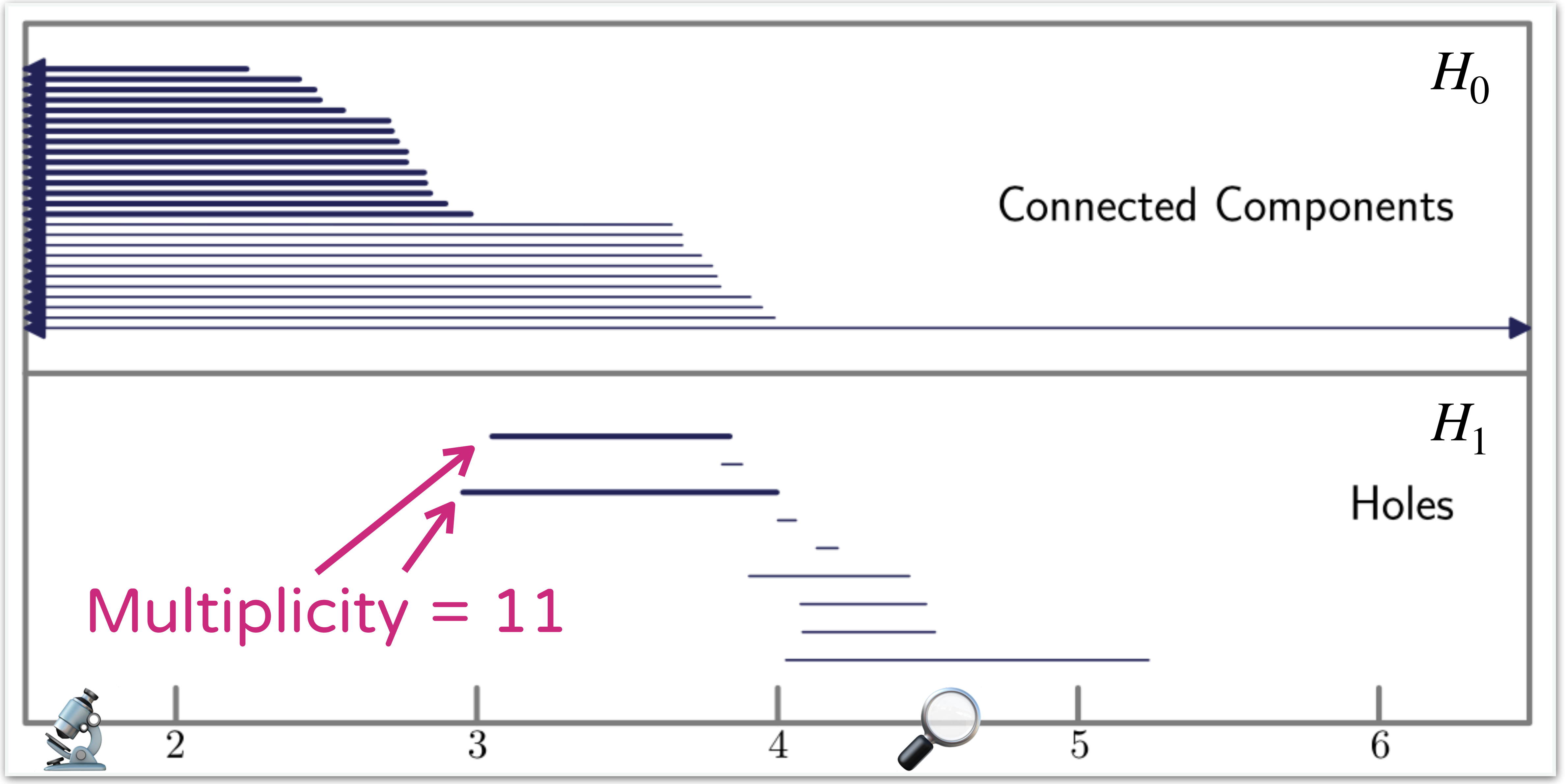
April 17, 2020

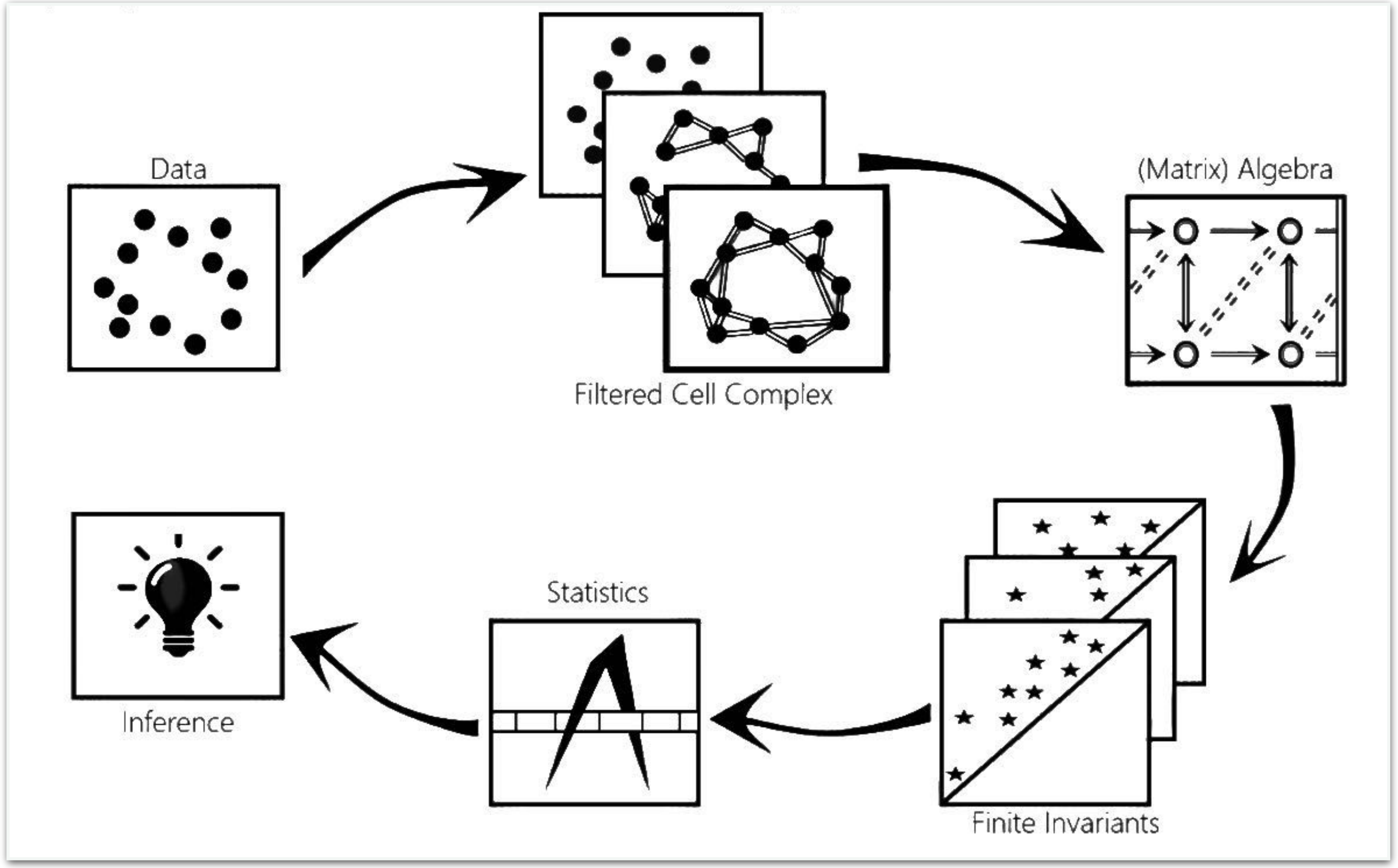
A Motivating Example



- Letter A $\times 1$?
- Letter B $\times 11$?
- Points $\times 176$?

Scale matters!





[Steve Oudot “Persistence Theory.” AMS monograph (2015), Vidit Nanda’s talk on topological data analysis (2017)]

Outline

- ~~A motivating example~~
- Introduction & active research directions

[Gunnar Carlsson "Persistent Homology and Applied Homotopy Theory." arXiv (2020)]

- Theoretical foundations

Stability theorem

  - Topological / algebraic persistence, & stability

Structure theorem

- Computation

[Steve Oudot "Persistence Theory." AMS monograph (2015)]

[Vidit Nanda's notes on computational algebraic topology (2020)]

- (Three interesting applications)

[Ann E. Sizemore et al., "TDA for the network neuroscientist." Network Neuroscience (2019).]

[Tamal. K. Dey et al. "An efficient algorithm for 1-dimensional (persistent) path homology." SoCG (2020).]

[Michelle Feng & Mason A. Porter "Persistent Homology of Geospatial Data: A Case Study with Voting." SIAM Review (2020).]

- Beyond persistence homology

[Han Riess. "Beyond Persistent Homology: A Mathematical Guide." Preprint (2019).]

- Conclusion

Introduction

- Original motivation: Extend the ideas of algebraic topology from the category of spaces X to situations where we only have a sampling of the space X.
    - Only 20 years in history!

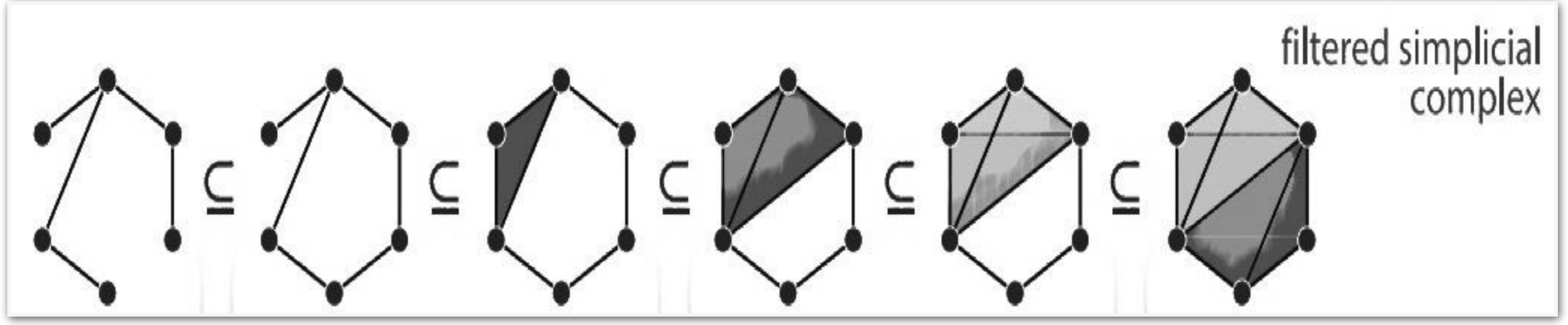
- Active directions include:
    - Coordinatization of barcodes
    - Stability results
    - Coverage and evasion problems
    - Generalized persistence
    - Probabilistic analysis and inference
    - Symplectic geometry

[Gunnar Carlsson "Persistent Homology and Applied Homotopy Theory." arXiv (2020)]

Topological Persistence

- Def: Filtration

  A filtration is a sequence of nested topological spaces $X_0 \subseteq X_1 \subseteq \ldots \subseteq X_n$.



- Note: From topological persistence to algebraic persistence

  Applying the "d-dimensional" homology with coefficient in a field $\mathbb{F}$, i.e., $H_d(\cdot, \mathbb{F})$, to a filtration gives a sequence of vector spaces & linear maps,

$$H_d(K_0;\mathbb{F}) \xrightarrow{H_d \circ f_0} H_d(K_1;\mathbb{F}) \xrightarrow{H_d \circ f_1} \cdots \xrightarrow{H_d \circ f_{n-1}} H_d(K_n;\mathbb{F})$$

where f_* is the inclusion simplicial map sending K_* to the next K_{*+1}.

[Steve Oudot "Persistence Theory." AMS monograph (2015)]

Algebraic Persistence (1 of 2)

- Def: Persistence module

A persistence module is a sequence $\mathbb{V}_*$ of vector spaces and linear maps, i.e.,

$$V_0 \xrightarrow{\phi_0} V_1 \xrightarrow{\phi_1} V_2 \xrightarrow{\phi_2} \dots \xrightarrow{\phi_{n-1}} V_n$$

- Def: Persistence homology groups

The persistent homology groups of $\mathbb{V}_*$ are $H_i^j(\mathbb{V}_*) = \text{Im}\,[\psi_{j-1} \circ \ldots \circ \psi_{i+1} \circ \psi_i]$, assuming $i \leq j$. For convenience, we can write $\psi_i^j : V_i \to V_j$. Then, we have the following interpretation:

An element $\alpha \in V_i$ is BORN at i if $\alpha \notin \text{Im}\,\psi_{i-1}$ and DIES at $j > 1$ if $\psi_i^j(\alpha) = 0$ but $\psi_i^{j-1}(\alpha) \neq 0$.

- Def: Interval module

Given $i < j$, the interval module $\mathbb{I}_*^{[i,j)}$ is:

Nicest possible persistence modules

$$\begin{array}{ccccccccccccccc} \cdots & \longrightarrow & 0 & \longrightarrow & 0 & \longrightarrow & \boxed{\mathbb{F} & \xrightarrow{\text{Id}} & \mathbb{F} & \xrightarrow{\text{Id}} & \cdots & \xrightarrow{\text{Id}} & \mathbb{F}} & \longrightarrow & 0 & \longrightarrow & 0 & \longrightarrow & \cdots \\ & & & & & & \| & & \| & & & & \| & & \| \\ & & & & & & V_i & & V_{i+1} & & & & V_{j-1} & & V_j \end{array}$$

[Vidit Nanda's notes on computational algebraic topology (2020)]

Algebraic Persistence (2 of 2)

- Structure theorem

To each persistence module $\mathbb{V}_*$ of vector spaces over $\mathbb{F}$, one can associate a multi-set of intervals,

$$\mathrm{Bar}(\mathbb{V}_*) = \{[i,j) \mid 0 \leq i < j < \infty\} \; .$$

such that $\mathbb{V}_*$ is isomorphic to a direct sum of interval modules,

$$\mathbb{V}_* \cong \bigoplus_{[i,j) \in \mathrm{Bar}(V)} \mathbb{I}_*^{[i,j)} \; .$$

- "proof"

Every persistence module is an "honest module" in the sense of abstract algebra over the polynomial ring $\mathbb{F}[t]$, where t acts on $\alpha \in V_i$ by pushing it to $\psi_i(\alpha) \in V_{i+1}$. These "finitely generated modules" over $\mathbb{F}[t]$ (a principle ideal domain) decompose as follows.

$\exists \{\alpha_i, i \in I\}$ and $\exists \{\beta_j < \delta_j, j \in J\}$, such that

"torsion part" (quotient by ideal)

$$\mathbb{V}_* \cong \boxed{\bigoplus_{\alpha_i} t^{\alpha_i} \cdot \mathbb{F}[t]} \oplus \boxed{\bigoplus_{\beta_j < \gamma_j} t^{\beta_j} \cdot \mathbb{F}[t]/t^{\gamma_j}} \; .$$

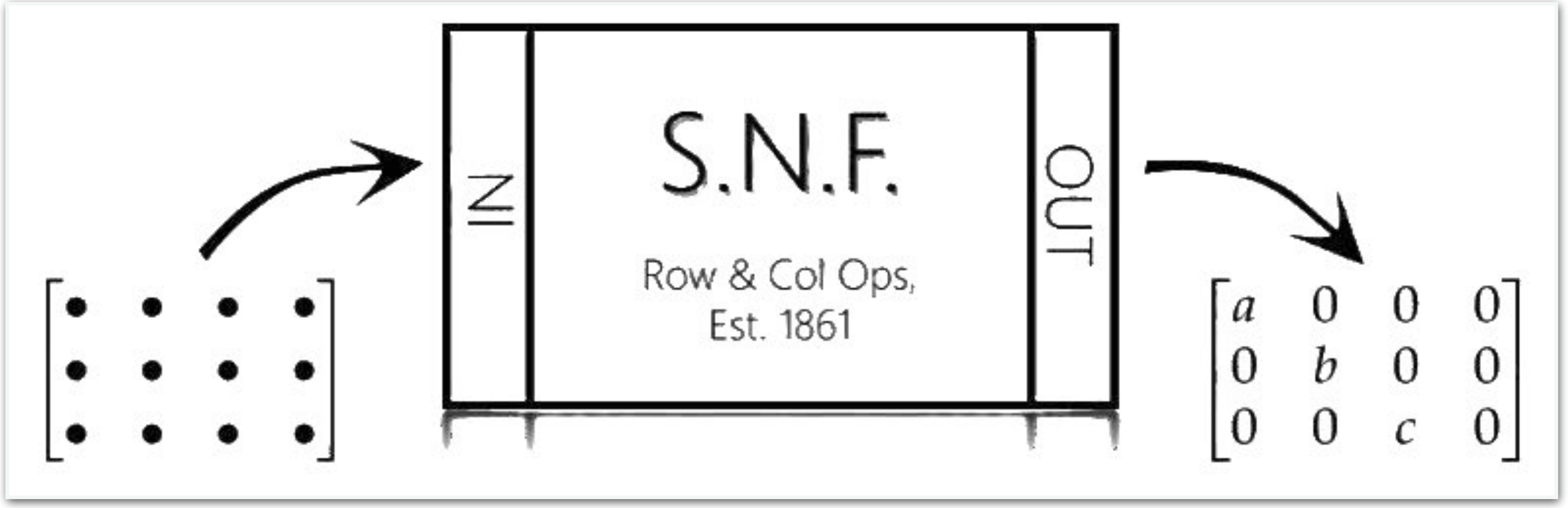
"free part"

And in this case ⋯

$$\mathrm{Bar}(\mathbb{V}_*) = \coprod_{i,j} \{[\alpha_i, \infty)\}\{[\beta_j, \gamma_j)\} \; .$$

[Afra Zomorodian & Gunnar Carlsson "Computing Persistent Homology." Discrete Comput. Geom. (2005)]

Computation



Algorithm 1: Matrix reduction

Input: $m \times m$ binary matrix M

Let $R = M$;
for $j = 1$ *to* m **do**
  while *there exists* $k < j$ *with* $\mathsf{low}(k, R) = \mathsf{low}(j, R) \neq 0$ **do**
    add (modulo 2) column k to column j in R;
  end
end

Output: R

- Caveat: The row & col operations are computationally expensive!

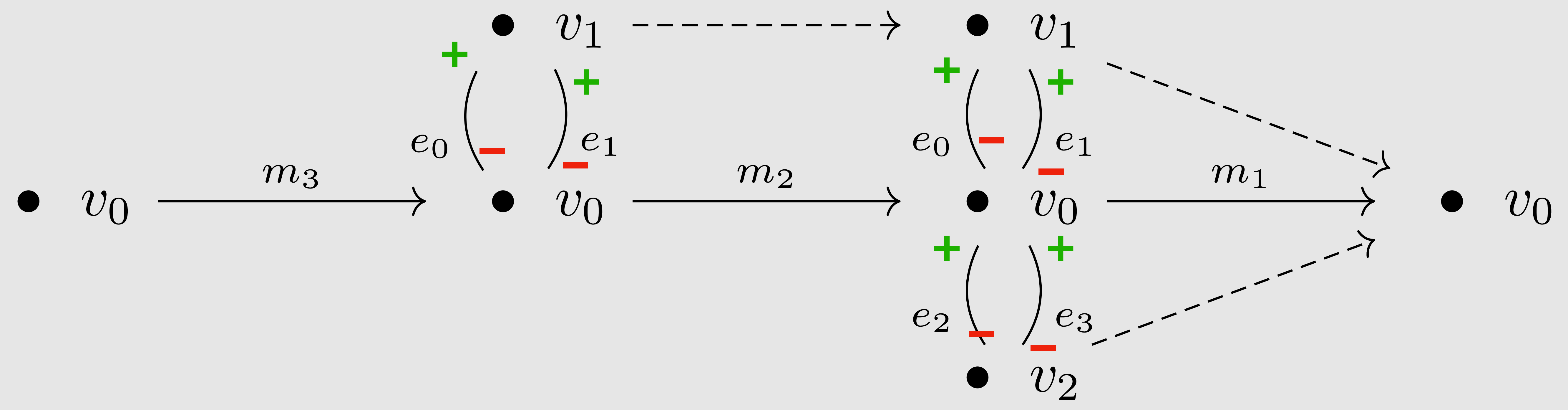
  The time complexity scales **at most cubic** in the number m of simplices of K. But m is **combinatorially** large.

[Steve Oudot's notes, Vidit Nanda's talk on topological data analysis (2017)]

Example Persistent Homology Calculation (outcome: barcodes)

Four (small) cellular complexes:	X_3	X_2	X_1	X_0

Sequence of simplicial maps:



Dimension 1: $0 \qquad \mathbb{R}^2 \qquad \mathbb{R}^4 \qquad 0$

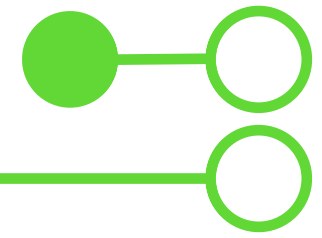
$$\begin{array}{c|cc} & |e_0\rangle & |e_1\rangle \\ \hline \langle v_0| & -1 & -1 \\ \langle v_1| & +1 & +1 \end{array} \qquad \text{Rank=1}$$

$$\begin{array}{c|cccc} & |e_0\rangle & |e_1\rangle & |e_2\rangle & |e_3\rangle \\ \hline \langle v_0| & -1 & -1 & +1 & +1 \\ \langle v_1| & +1 & +1 & 0 & 0 \\ \langle v_2| & 0 & 0 & -1 & -1 \end{array} \qquad \text{Rank=2}$$

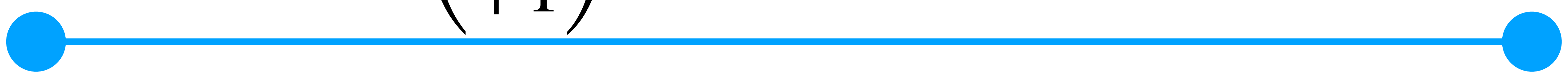
Dimension 0: $\mathbb{R} \qquad \mathbb{R}^2 \qquad \mathbb{R}^3 \qquad \mathbb{R}$

$$X_3:\ 0 \to \mathbb{R} \to 0 \qquad X_2:\ \mathbb{R}^2 \xrightarrow{\begin{pmatrix} -1 & -1 \\ +1 & +1 \end{pmatrix}} \mathbb{R}^2 \to 0 \qquad X_1:\ \mathbb{R}^4 \xrightarrow{\begin{pmatrix} -1 & -1 & +1 & +1 \\ +1 & +1 & 0 & 0 \\ 0 & 0 & -1 & -1 \end{pmatrix}} \mathbb{R}^3 \to 0 \qquad X_0:\ 0 \to \mathbb{R} \to 0$$

$H_1(X_i; \mathbb{R})$: $0 \qquad \mathbb{R}/0 \cong \mathbb{R} \qquad \mathbb{R}^4/\mathbb{R}^2 \cong \mathbb{R}^2 \qquad 0$

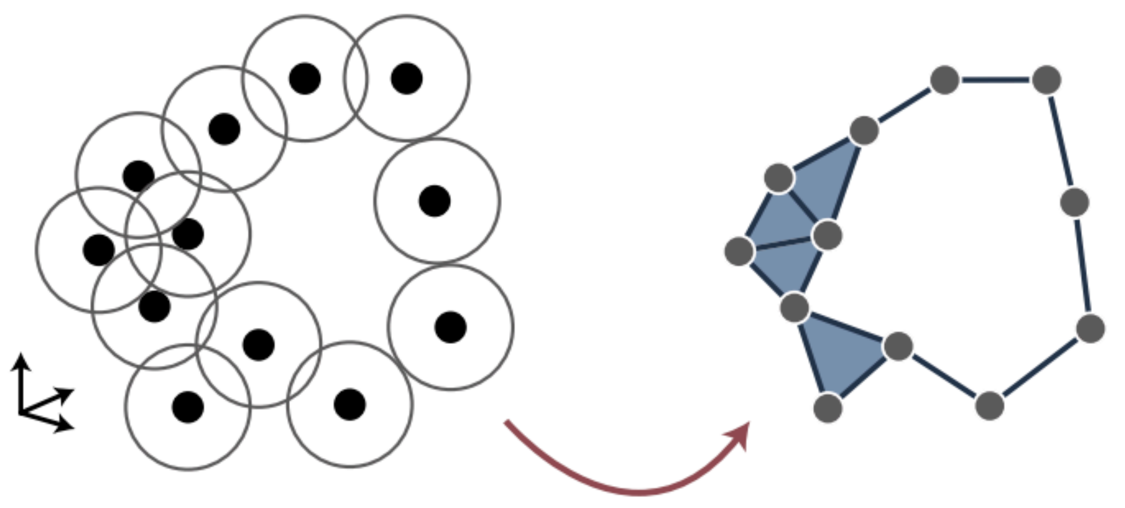


$H_0(X_i; \mathbb{R})$: $\mathbb{R} \qquad \mathbb{R}^2 / \begin{pmatrix} -1 \\ +1 \end{pmatrix} \cong \mathbb{R} \qquad \mathbb{R}^3/\mathbb{R}^2 \cong \mathbb{R} \qquad \mathbb{R}$



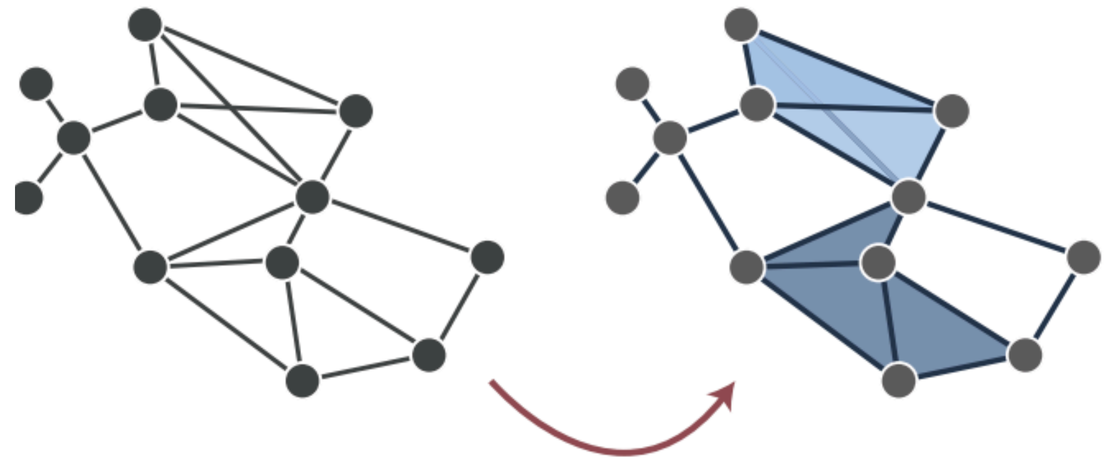
[Michael Robinson, https://www.youtube.com/watch?v=4z9NXnk_jXM (2016)]

From Data to Simplicial Complex



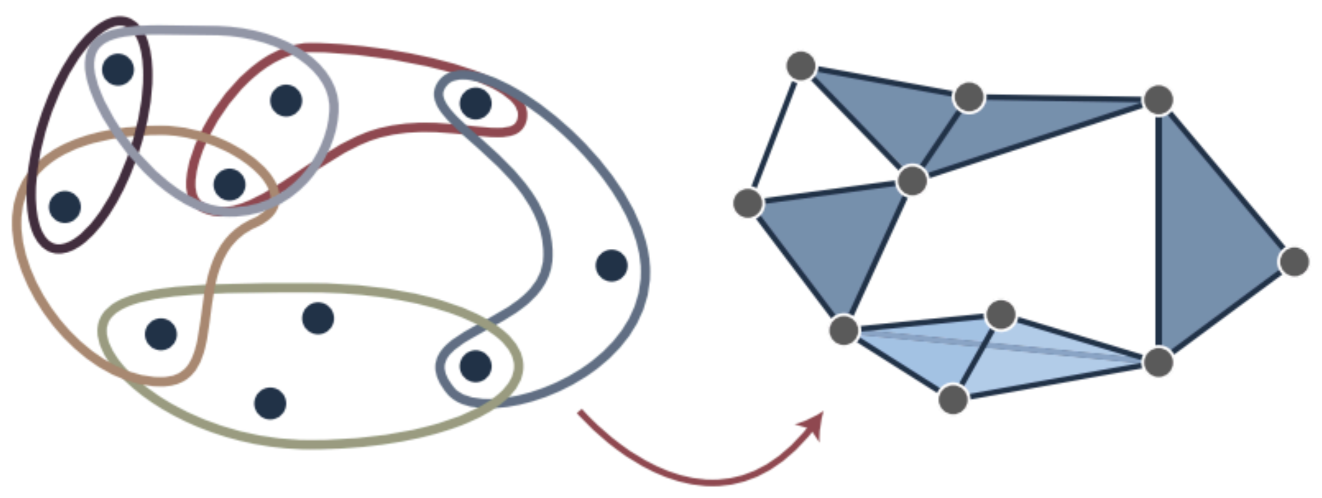
- Vietoris-Rips complex

  - Input format: point cloud



- Clique (or flag) complex

  - Input format: binary graph



- Nerve complex

  - Input format: graph where groups of nodes share a feature

- Other flavors include witness complex, Čech complex, Alpha complex, etc. You can even craft your own! For example, statistically sound filtrations or filtrations that better detect the changes in data.
[Rob Ghrist, "Elementary Applied Topology." (2014)]

- To learn which computational tool to use (e.g., software packages), see Nina Otter et al., "A roadmap for the computation of persistent homology." EPJ Data Science (2017).

[Figures on top: Ann E. Sizemore et al., "TDA for the network neuroscientist." Network Neuroscience (2019).]

Stability (1 of 2; Goal: be able to compare barcodes & persistence modules)

- Def: ϵ-matching

  Let B and B' be two barcodes. A ϵ-matching between them is a bijection $\mu : B_0 \xrightarrow{\simeq} B_0'$ where $B_0 \subseteq B$ and $B_0' \subseteq B'$ satisfying:

  (A) All intervals in $(B - B_0)$ and $(B' - B_0')$ have length $\leq 2\epsilon$.

  (B) If $\mu\left([a,b)\right) = [a',b')$ for some $[a,b) \in B_0$, then $|a - a'| < \epsilon$ and $|b - b'| < \epsilon$.

- Def: The Bottleneck distance between barcodes B and B' is

$$d_{\text{Bottle}}(B,B') = \inf\{\epsilon > 0 \mid \exists \epsilon\text{-matching } B \leftrightarrow B'\} .$$

Stability (2 of 2; Goal: be able to compare barcodes & persistence modules)

- Def: ϵ-interleaving

  Let U_* and V_* be two $\mathbb{R}^+$-indexed persistence modules. A ϵ-interleaving between them is a family of maps $\{\phi_t : U_t \to V_{t+\epsilon} \mid \forall t \in \mathbb{R}^+\}$ and $\{\psi_t : V_t \to U_{t+\epsilon} \mid \forall t \in \mathbb{R}^+\}$ such that all possible triangles and parallelograms commute, e.g.,

$$\begin{tikzcd} U_s \arrow[rr, "f_{s,s+2\epsilon}"] \arrow[dr, "\phi_s"'] & \forall s \leq t & U_{s+2\epsilon} \\ & V_{s+\epsilon} \arrow[ur, "\psi_{s+\epsilon}"'] & \end{tikzcd}
\qquad
\begin{tikzcd} U_s \arrow[rr, "f_{s,t}"] \arrow[dr, "\phi_s"'] & & U_t \arrow[dr, "\psi_t"] & \\ & V_{s+\epsilon} \arrow[rr, "g_{s+\epsilon,t+\epsilon}"'] \arrow[r, phantom, "\forall s"] & & V_{t+\epsilon} \end{tikzcd}$$

Note: $f_{s,t} : U_s \to U_t$ is a linear map for $s \leq t$ such that $f_{s,s} = \mathrm{Id}$ $\forall s$ and $f_{s,u} = f_{t,u} \circ f_{s,t}$ $\forall s \leq t \leq u$. Similarly for $g_{s,t}$ and V_* .

[+ 2 more]

- Def: The Interleaving distance between U_* and V_* is

$$d_{\mathrm{Interl}}(U_*, V_*) = \inf\{\epsilon > 0 \mid \exists \epsilon\text{-interleaving } U_* \leftrightarrow V_*\} .$$

- Stability Theorem

  If U_* and V_* are tame $\mathbb{R}^+$-indexed persistence modules, then there is an isometry:

$$d_{\mathrm{Interl}}(U_*, V_*) = d_{\mathrm{Bottle}(\mathrm{Bar}(U_*),\mathrm{Bar}(V_*))} .$$

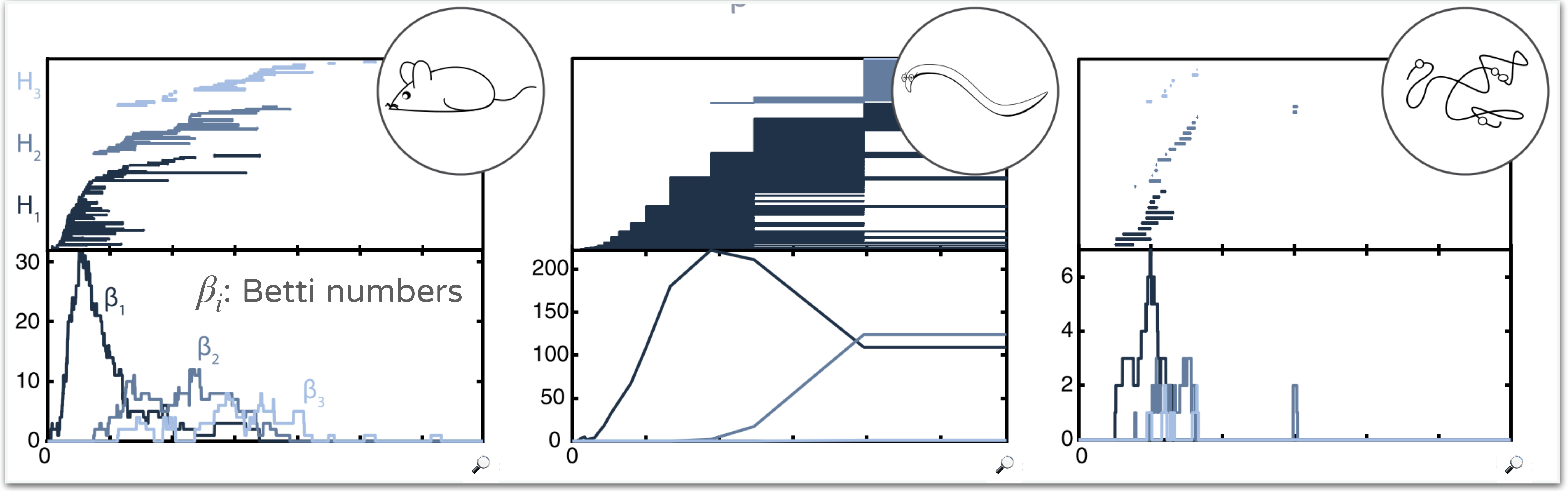
"proof" See Lesnick & Bauer's "induced matchings" (SOCG'14)

Noise in data is no more than noise in Barcode!

(justifies the use of homology in data analysis)

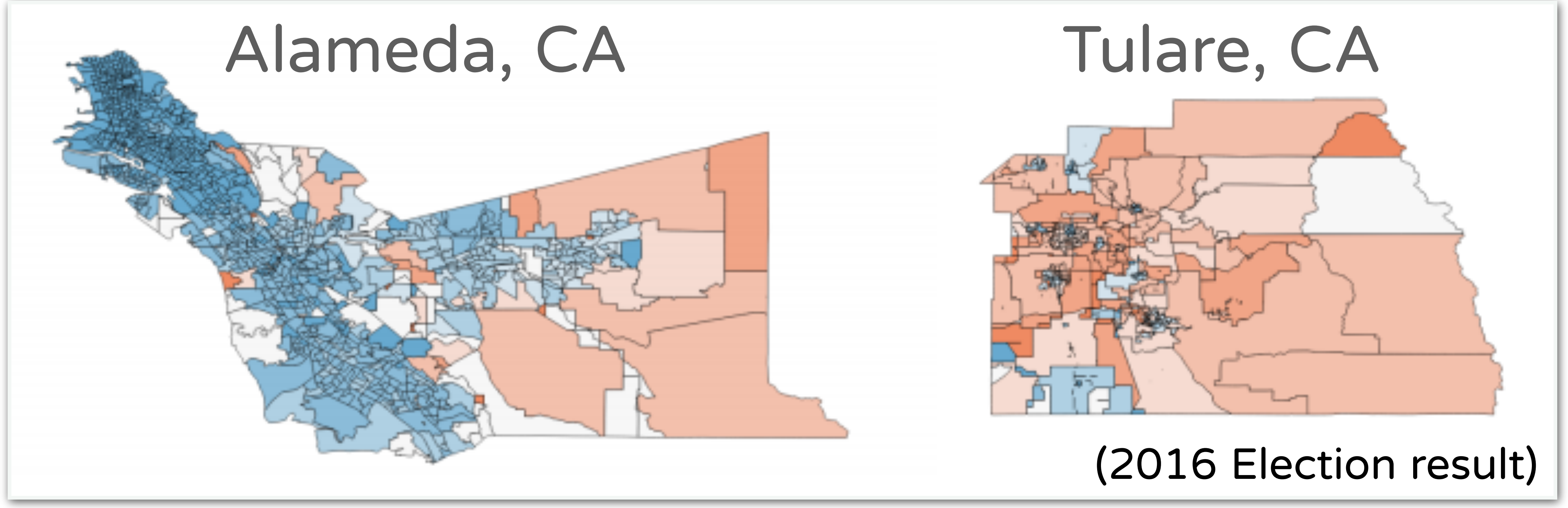
Application 1 of 3: Persistent Homology in Neuroscience

"Find 'holes' in the network of interacting brain regions, across organisms"



Application 2 of 3: Persistent Homology of Geospatial Data

"Find a Blue island in the Red sea ···"



Issue	VR	Alpha	Adjacency	Level set
Scaling	✗	✗	✓	✗
Contiguity	✗	✗	✓	✓

"··· by constructing better complexes for 2D data"

[Ann E. Sizemore et al., "TDA for the network neuroscientist." Network Neuroscience (2019).]
[Michelle Feng & Mason A. Porter "Persistent Homology of Geospatial Data: A Case Study with Voting." SIAM Review (2020).]

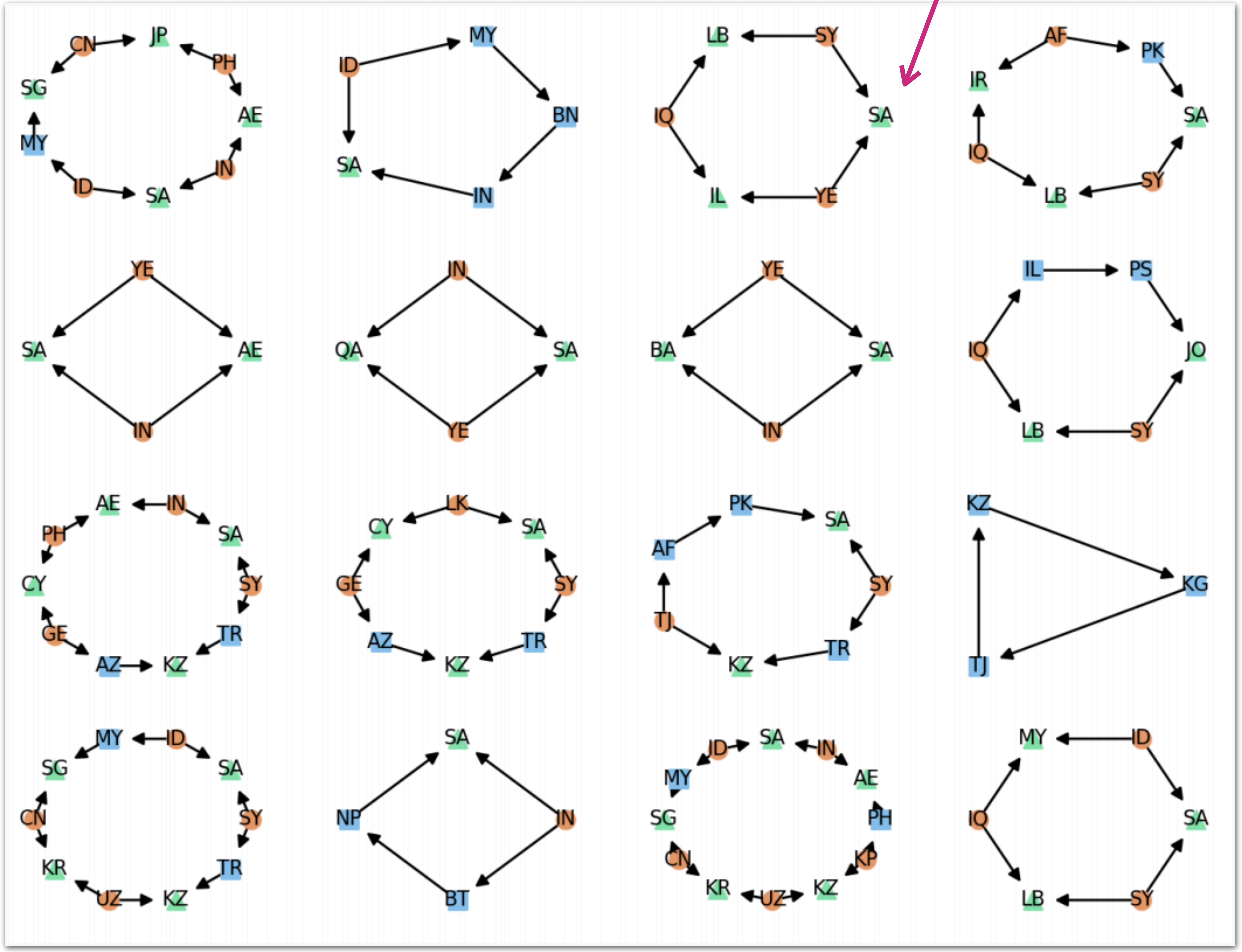
Application 3 of 3: Persistent Path Homology

2-letter country code

- On the right 👉: Minimal cycles for persistent path homology on migration network
- Dataset 💾: UN's net migration network (a weighted directed graph)

Compute "minimal homology basis"

Topological analysis -> Geometric analysis!



[Tamal. K. Dey et al. "An efficient algorithm for 1-dimensional (persistent) path homology." SoCG (2020).]

Beyond Persistent Homology

- Computation & math issues (e.g., do we really need <u>all</u> simplices before we can compute homology?)

Han Riess, "Beyond Persistent Homology: A Mathematical Guide." Preprint (2019).

Justin Curry, Robert Ghrist & Vidit Nanda, "Discrete Morse Theory for Computing Cellular Sheaf Cohomology." FoCM (2016).

Conclusion

- (see the whole slides)

Thanks!

Tzu-Chi Yen <tzuchi.yen@colorado.edu>