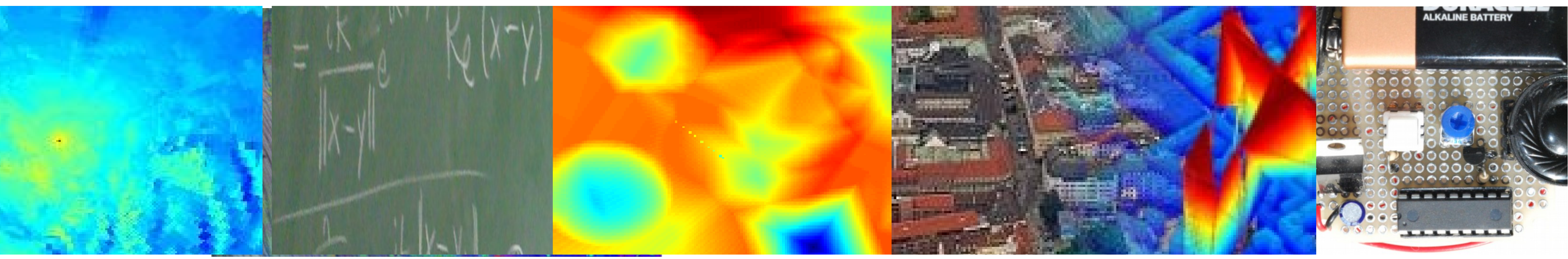
Introduction *to the* Special Session *on* Topological Data Analysis



Harish Chintakunta, Michael Robinson, Hamid Krim

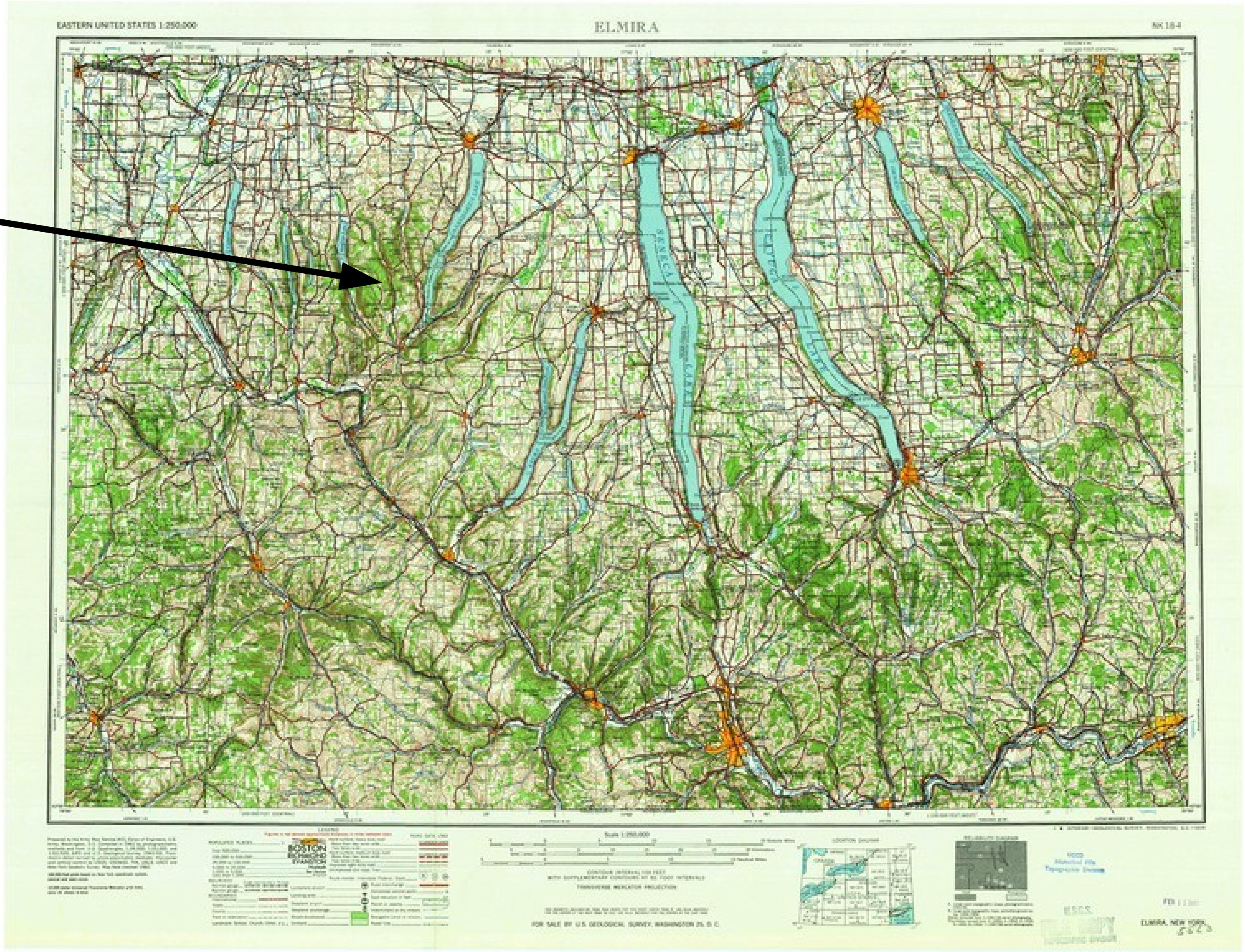
FLORIDA POLYTECHNIC UNIVERSITY

AU

NC STATE UNIVERSITY

What is topology?

Not this!
This is
TopoGRAPHY!



(Thanks USGS!)

Michael Robinson

What is topology?

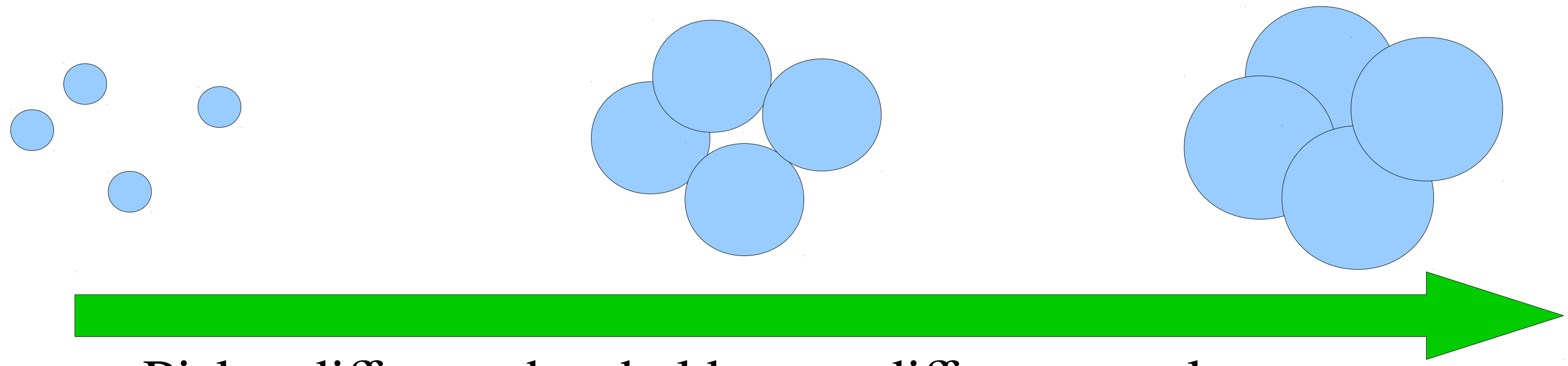
=

Topology is the study of spaces under continuous deformations

Michael Robinson

Topology and the curse of thresholds

- Although topology is very flexible, it **also** seems quite brittle. And in signal processing, that's bad!



Pick a different threshold… get different topology

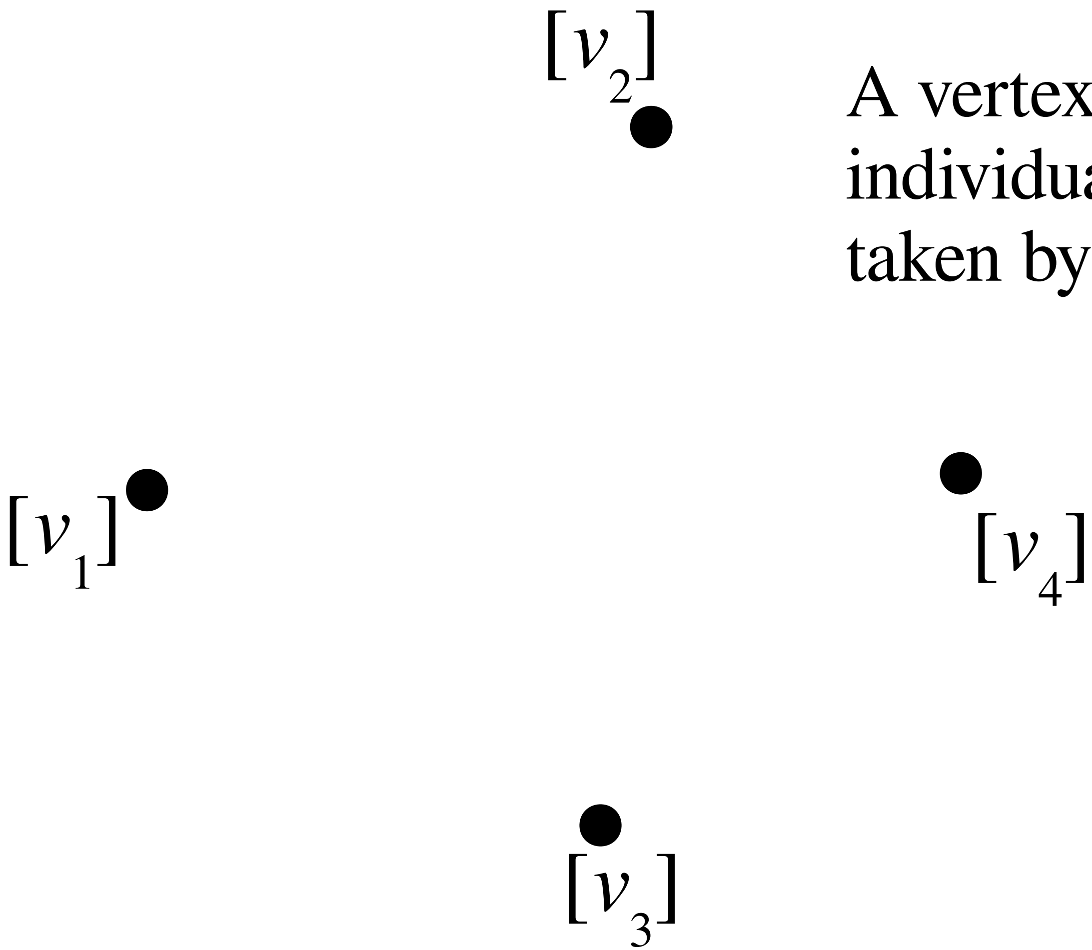
- But a nice idea *persists*… and in the end prevails

Herbert Edelsbrunner, David Letscher, and Afra Zomorodian, Topological persistence and simplification, *Discrete Comput. Geom*. 28 (2002), no. 4, 511–533.

- Rather than selecting one threshold, let's use them **all**!

Simplicial complexes

- A simplicial complex is a collection of *vertices* and ...

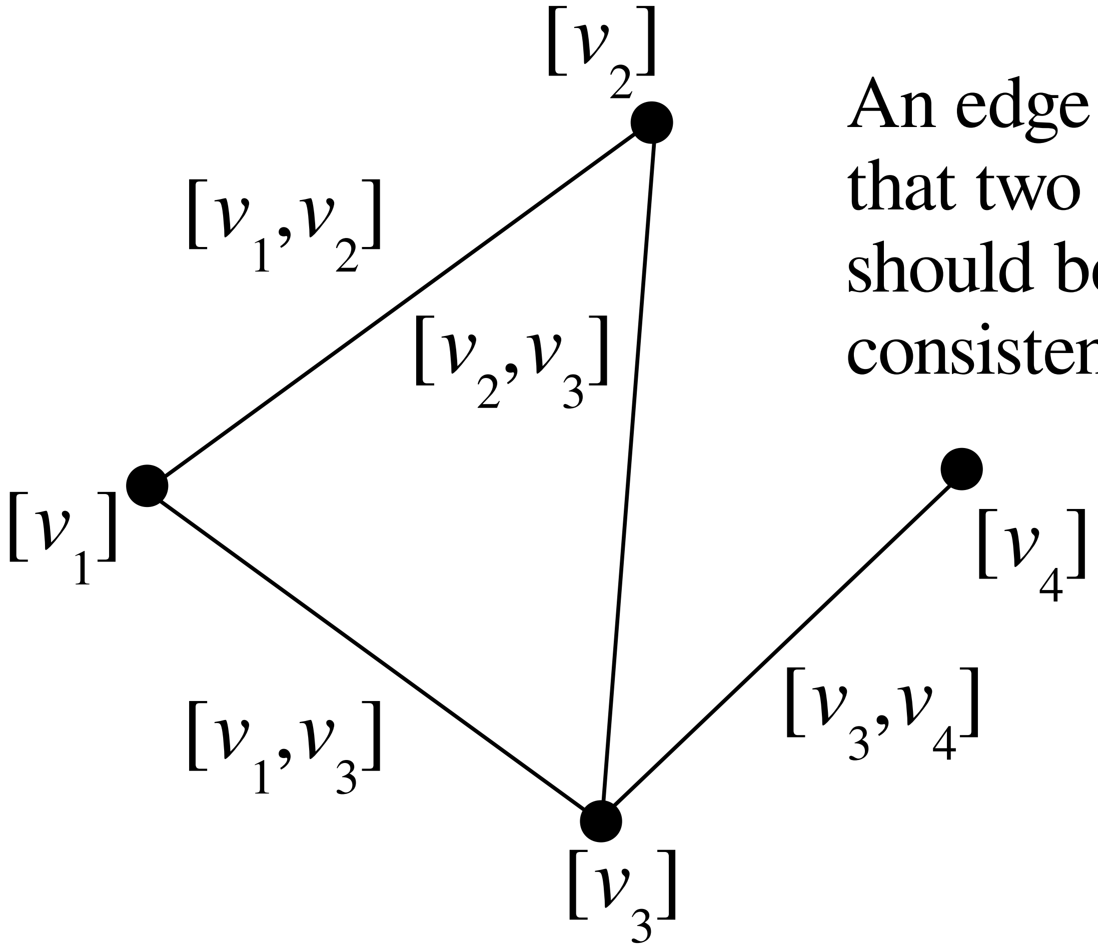


A vertex represents an individual measurement taken by a sensor

Michael Robinson

Simplicial complexes

- … *edges* (pairs of vertices) and ...



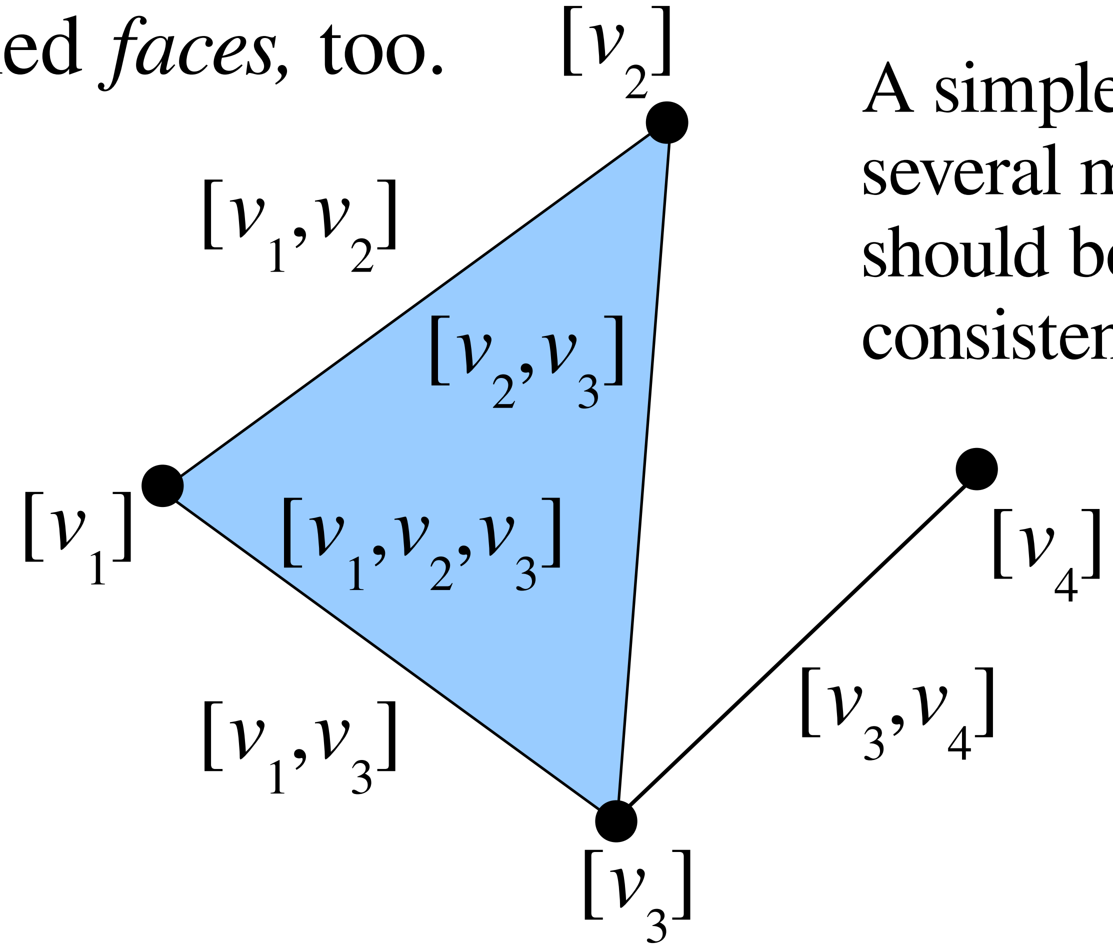
An edge represents the fact that two measurements should be tested for consistency

Usually people call this a *graph*; I will too

Michael Robinson

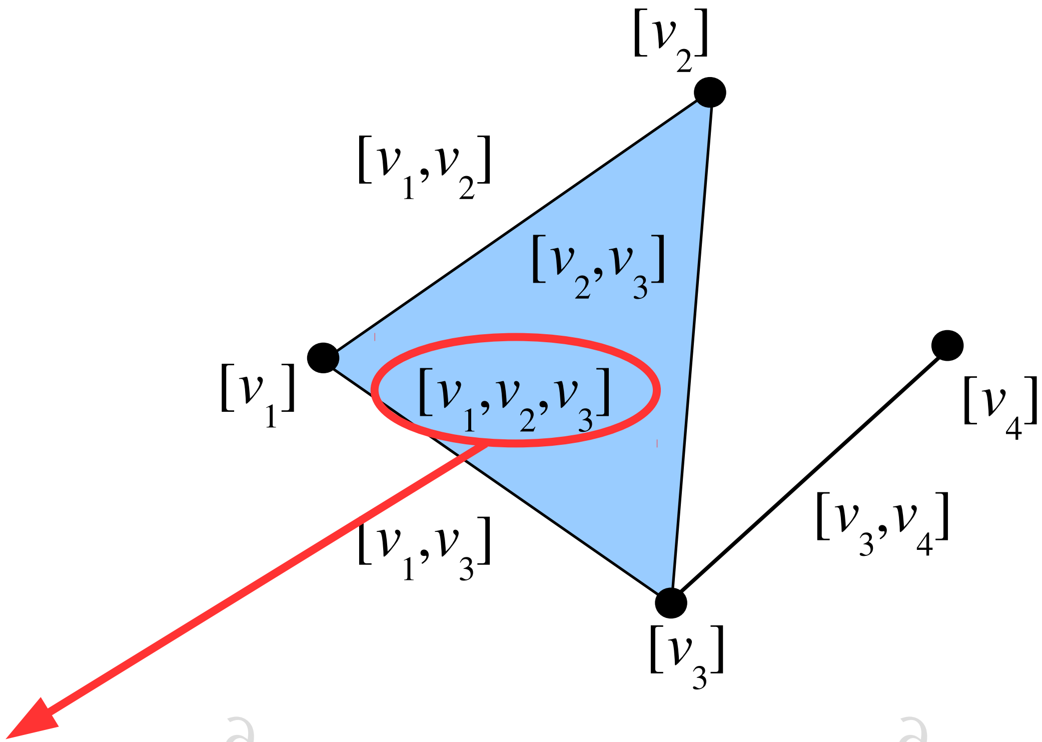
Simplicial complexes

- ... higher dimensional *simplices* (tuples of vertices)
- Whenever you have a simplex, you have all subsets, called *faces*, too.



A simplex represents that several measurements should be tested for consistency

Simplicial chain complex



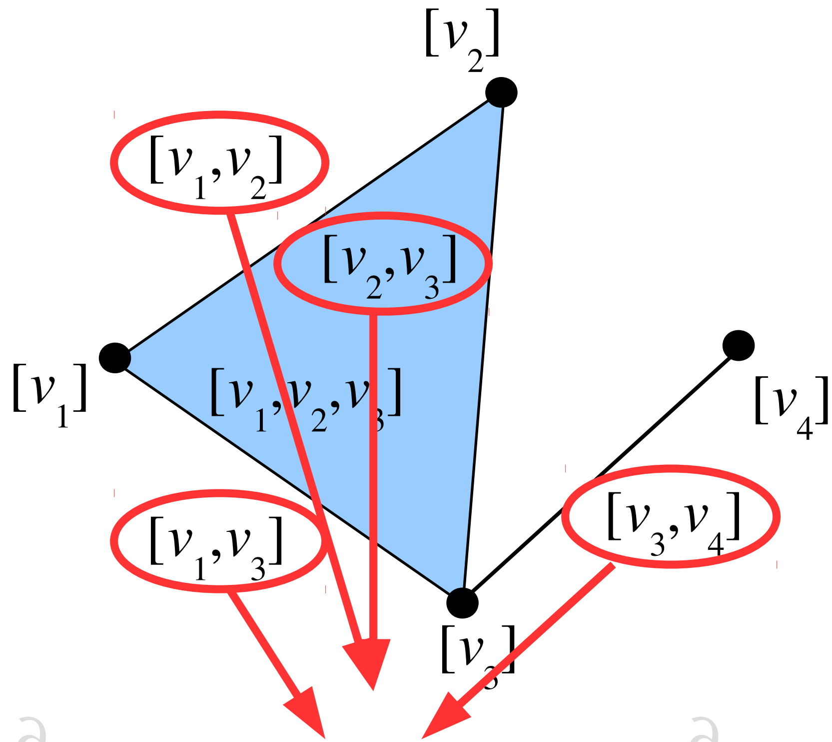
$$\xrightarrow{\partial_3} \mathbb{R} \xrightarrow{\partial_2} \mathbb{R}^4 \xrightarrow{\partial_1} \mathbb{R}^4 \xrightarrow{\partial_0} 0$$

	[v_1,v_2,v_3]
[v_1,v_2]	+1
[v_1,v_3]	-1
[v_2,v_3]	+1
[v_3,v_4]	0

	[v_1,v_2]	[v_1,v_3]	[v_2,v_3]	[v_3,v_4]
[v_1]	-1	-1	0	0
[v_2]	+1	0	-1	0
[v_3]	0	+1	+1	-1
[v_4]	0	0	0	+1

Michael Robinson

Simplicial chain complex

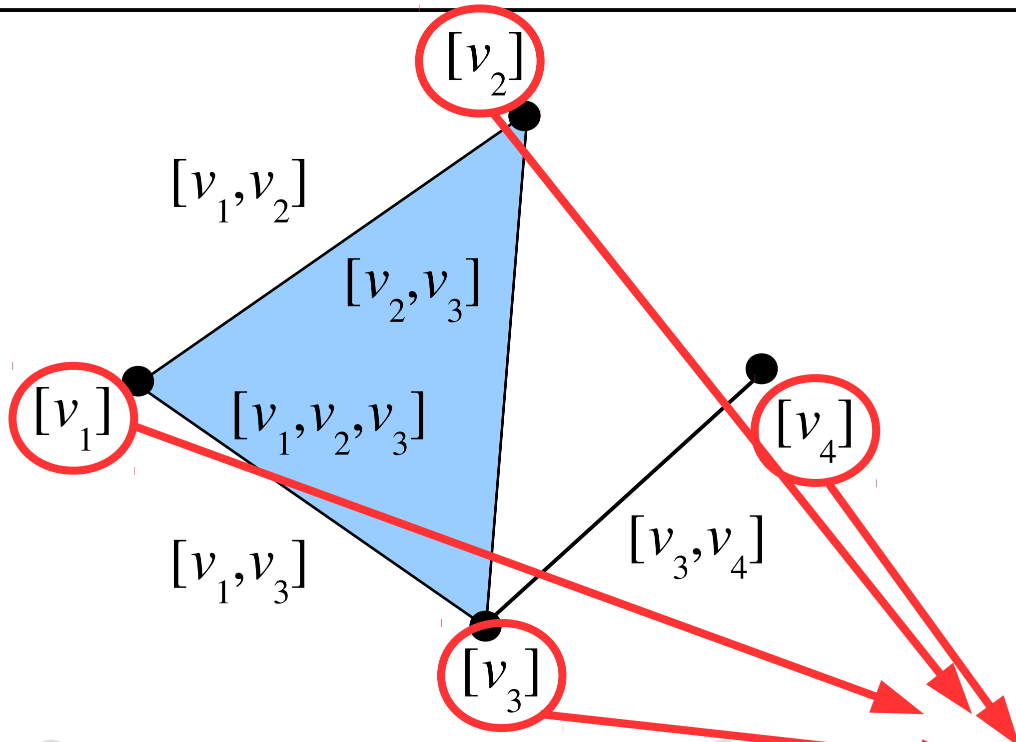


$$\xrightarrow{\partial_3} \mathbb{R} \xrightarrow{\partial_2} \mathbb{R}^4 \xrightarrow{\partial_1} \mathbb{R}^4 \xrightarrow{\partial_0} 0$$

$$\partial_2: \quad \begin{array}{c} \\ [v_1,v_2] \\ [v_1,v_3] \\ [v_2,v_3] \\ [v_3,v_4] \end{array} \begin{array}{c} [v_1,v_2,v_3] \\ \begin{pmatrix} +1 \\ -1 \\ +1 \\ 0 \end{pmatrix} \end{array}$$

$$\partial_1: \quad \begin{array}{c} \\ [v_1] \\ [v_2] \\ [v_3] \\ [v_4] \end{array} \begin{array}{c} \begin{array}{cccc} [v_1,v_2] & [v_1,v_3] & [v_2,v_3] & [v_3,v_4] \end{array} \\ \begin{pmatrix} -1 & -1 & 0 & 0 \\ +1 & 0 & -1 & 0 \\ 0 & +1 & +1 & -1 \\ 0 & 0 & 0 & +1 \end{pmatrix} \end{array}$$

Simplicial chain complex

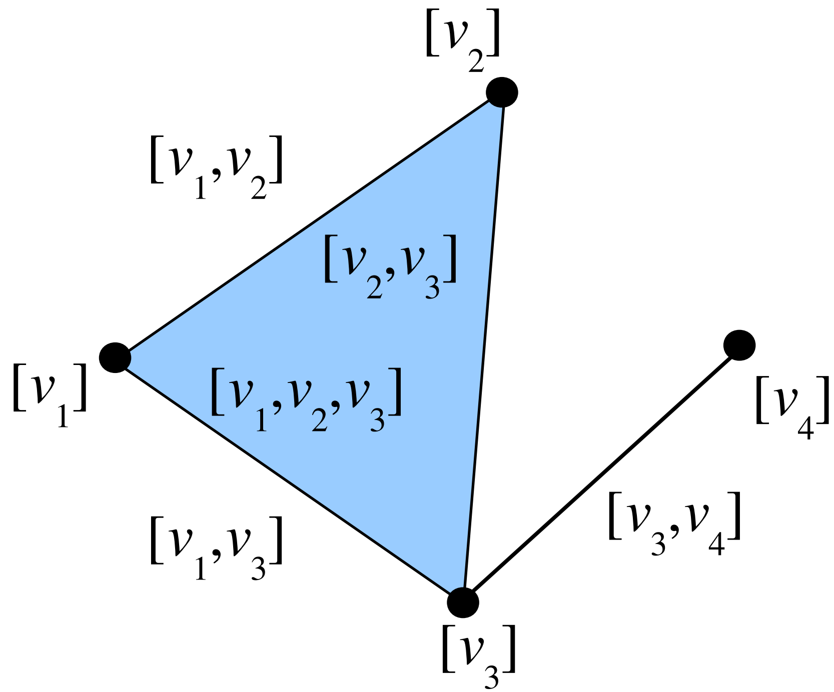


$$\xrightarrow{\partial_3} \mathbb{R} \xrightarrow{\partial_2} \mathbb{R}^4 \xrightarrow{\partial_1} \mathbb{R}^4 \xrightarrow{\partial_0} 0$$

	[v_1,v_2,v_3]
[v_1,v_2]	+1
[v_1,v_3]	-1
[v_2,v_3]	+1
[v_3,v_4]	0

	[v_1,v_2]	[v_1,v_3]	[v_2,v_3]	[v_3,v_4]
[v_1]	-1	-1	0	0
[v_2]	+1	0	-1	0
[v_3]	0	+1	+1	-1
[v_4]	0	0	0	+1

Simplicial chain complex



$$\xrightarrow{\partial_3} \mathbb{R} \xrightarrow{\partial_2} \mathbb{R}^4 \xrightarrow{\partial_1} \mathbb{R}^4 \xrightarrow{\partial_0} 0$$

∂_2:

	$[v_1,v_2,v_3]$
$[v_1,v_2]$	+1
$[v_1,v_3]$	-1
$[v_2,v_3]$	+1
$[v_3,v_4]$	0

∂_1:

	$[v_1,v_2]$	$[v_1,v_3]$	$[v_2,v_3]$	$[v_3,v_4]$
$[v_1]$	-1	-1	0	0
$[v_2]$	+1	0	-1	0
$[v_3]$	0	+1	+1	-1
$[v_4]$	0	0	0	+1

Michael Robinson

Simplicial chain complex

$$\begin{pmatrix} -1 & -1 & 0 & 0 \\ +1 & 0 & -1 & 0 \\ 0 & +1 & +1 & -1 \\ 0 & 0 & 0 & +1 \end{pmatrix} \begin{pmatrix} +1 \\ -1 \\ +1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\xrightarrow{\partial_3} \mathbb{R} \xrightarrow{\partial_2} \mathbb{R}^4 \xrightarrow{\partial_1} \mathbb{R}^4 \xrightarrow{\partial_0} 0$$

∂_2: column $[v_1,v_2,v_3]$; rows $[v_1,v_2]$, $[v_1,v_3]$, $[v_2,v_3]$, $[v_3,v_4]$

$$\begin{pmatrix} +1 \\ -1 \\ +1 \\ 0 \end{pmatrix}$$

∂_1: columns $[v_1,v_2]$, $[v_1,v_3]$, $[v_2,v_3]$, $[v_3,v_4]$; rows $[v_1]$, $[v_2]$, $[v_3]$, $[v_4]$

$$\begin{pmatrix} -1 & -1 & 0 & 0 \\ +1 & 0 & -1 & 0 \\ 0 & +1 & +1 & -1 \\ 0 & 0 & 0 & +1 \end{pmatrix}$$

Michael Robinson

Simplicial chain complex

$$\text{image } \partial_2 = \text{image}\begin{pmatrix}+1\\-1\\+1\\0\end{pmatrix} \subseteq \text{kernel } \partial_1 = \text{kernel}\begin{pmatrix}-1 & -1 & 0 & 0\\+1 & 0 & -1 & 0\\0 & +1 & +1 & -1\\0 & 0 & 0 & +1\end{pmatrix}$$

$$\begin{pmatrix}-1 & -1 & 0 & 0\\+1 & 0 & -1 & 0\\0 & +1 & +1 & -1\\0 & 0 & 0 & +1\end{pmatrix}\begin{pmatrix}+1\\-1\\+1\\0\end{pmatrix} = \begin{pmatrix}0\\0\\0\\0\end{pmatrix}$$

$$\xrightarrow{\partial_3} \mathbb{R} \xrightarrow{\partial_2} \mathbb{R}^4 \xrightarrow{\partial_1} \mathbb{R}^4 \xrightarrow{\partial_0} 0$$

	$[v_1,v_2,v_3]$
$[v_1,v_2]$	+1
$[v_1,v_3]$	-1
$[v_2,v_3]$	+1
$[v_3,v_4]$	0

	$[v_1,v_2]$	$[v_1,v_3]$	$[v_2,v_3]$	$[v_3,v_4]$
$[v_1]$	-1	-1	0	0
$[v_2]$	+1	0	-1	0
$[v_3]$	0	+1	+1	-1
$[v_4]$	0	0	0	+1

Homology of a chain complex

$$\text{image } \partial_2 = \text{image} \begin{pmatrix} +1 \\ -1 \\ +1 \\ 0 \end{pmatrix} \subseteq \text{kernel } \partial_1 = \text{kernel} \begin{pmatrix} -1 & -1 & 0 & 0 \\ +1 & 0 & -1 & 0 \\ 0 & +1 & +1 & -1 \\ 0 & 0 & 0 & +1 \end{pmatrix}$$

$$H_2 = \ker \partial_2 / \text{img } \partial_3 \qquad H_1 = \ker \partial_1 / \text{img } \partial_2 \qquad H_0 = \ker \partial_0 / \text{img } \partial_1$$

$$\xrightarrow{\partial_3} \mathbb{R} \xrightarrow{\partial_2} \mathbb{R}^4 \xrightarrow{\partial_1} \mathbb{R}^4 \xrightarrow{\partial_0} 0$$

$$\partial_2: \begin{array}{c} [v_1,v_2,v_3] \end{array} \quad \begin{array}{c} [v_1,v_2] \\ [v_1,v_3] \\ [v_2,v_3] \\ [v_3,v_4] \end{array} \begin{pmatrix} +1 \\ -1 \\ +1 \\ 0 \end{pmatrix}$$

$$\partial_1: \begin{array}{cccc} [v_1,v_2] & [v_1,v_3] & [v_2,v_3] & [v_3,v_4] \end{array} \quad \begin{array}{c} [v_1] \\ [v_2] \\ [v_3] \\ [v_4] \end{array} \begin{pmatrix} -1 & -1 & 0 & 0 \\ +1 & 0 & -1 & 0 \\ 0 & +1 & +1 & -1 \\ 0 & 0 & 0 & +1 \end{pmatrix}$$

Homology of a chain complex

$$\text{image } \partial_2 = \text{image} \begin{pmatrix} +1 \\ -1 \\ +1 \\ 0 \end{pmatrix} \subseteq \text{kernel } \partial_1 = \text{kernel} \begin{pmatrix} -1 & -1 & 0 & 0 \\ +1 & 0 & -1 & 0 \\ 0 & +1 & +1 & -1 \\ 0 & 0 & 0 & +1 \end{pmatrix}$$

BUBBLES	LOOPS	COMPONENTS
$H_2 = \ker \partial_2 / \operatorname{img} \partial_3$	$H_1 = \ker \partial_1 / \operatorname{img} \partial_2$	$H_0 = \ker \partial_0 / \operatorname{img} \partial_1$

$$\xrightarrow{\partial_3} \mathbb{R} \xrightarrow{\partial_2} \mathbb{R}^4 \xrightarrow{\partial_1} \mathbb{R}^4 \xrightarrow{\partial_0} 0$$

∂_2: column $[v_1,v_2,v_3]$; rows $[v_1,v_2]$, $[v_1,v_3]$, $[v_2,v_3]$, $[v_3,v_4]$

$$\begin{pmatrix} +1 \\ -1 \\ +1 \\ 0 \end{pmatrix}$$

∂_1: columns $[v_1,v_2]$, $[v_1,v_3]$, $[v_2,v_3]$, $[v_3,v_4]$; rows $[v_1]$, $[v_2]$, $[v_3]$, $[v_4]$

$$\begin{pmatrix} -1 & -1 & 0 & 0 \\ +1 & 0 & -1 & 0 \\ 0 & +1 & +1 & -1 \\ 0 & 0 & 0 & +1 \end{pmatrix}$$

Persistent homology

Goal: obtain a filtration of spaces from a finite pseudometric space X

$$X_0 \subseteq X_1 \subseteq X_2 \subseteq \ldots \subseteq X$$

Tactic: Vietoris-Rips complexes

$VR_\varepsilon(X)$ = set of all subsets of X with diameter ε or less

- This is an abstract simplicial complex, and

$$VR_\varepsilon(X) \subseteq VR_\eta(X) \text{ if } \varepsilon \leq \eta$$

Persistent homology

Goal: obtain a filtration of spaces from a finite pseudometric space X

$$H_k(X_0) \rightarrow H_k(X_1) \rightarrow H_k(X_2) \rightarrow \ldots \rightarrow H_k(X)$$

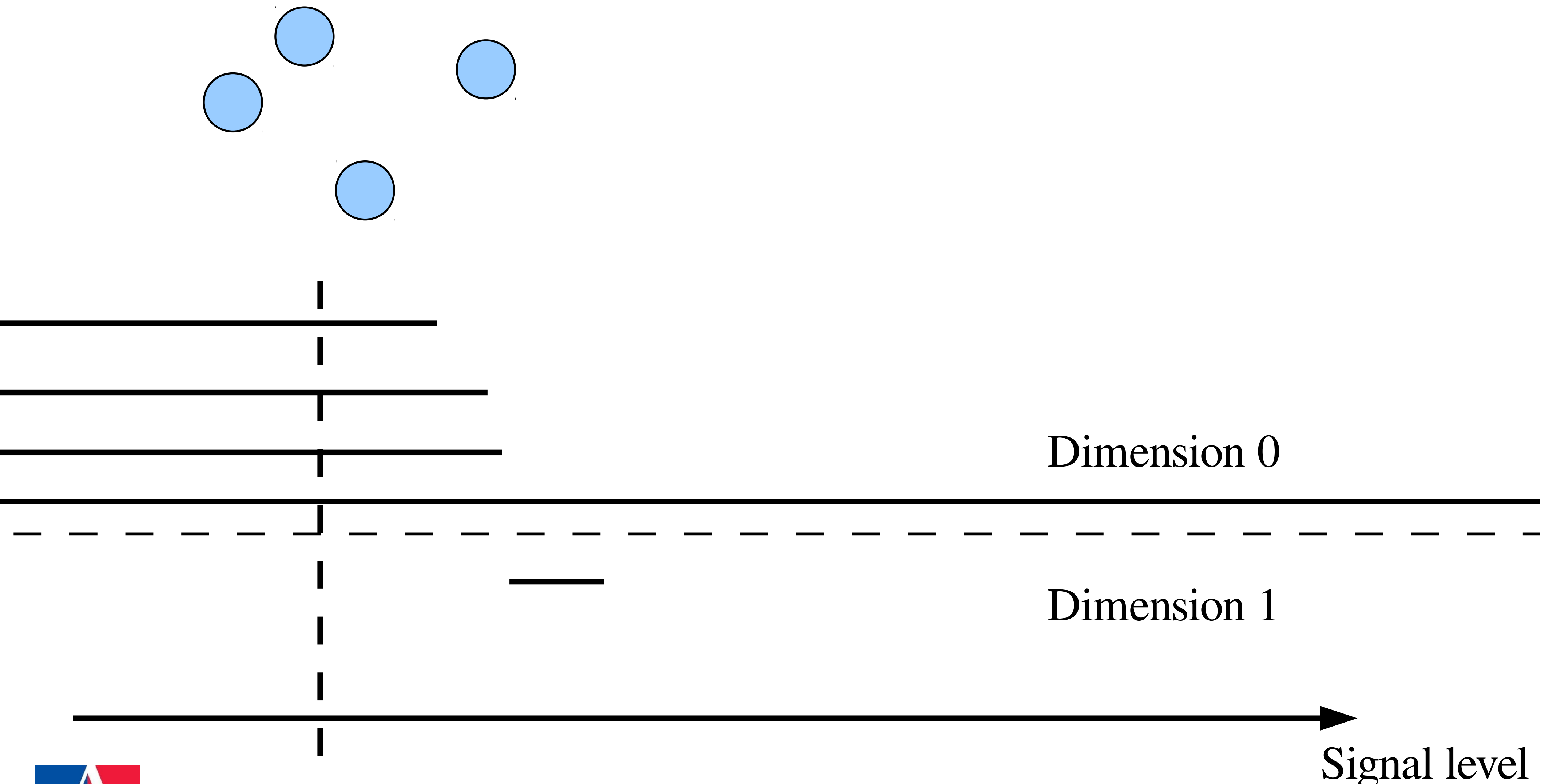
Tactic: Vietoris-Rips complexes

$\mathrm{VR}_\varepsilon(X)$ = set of all subsets of X with diameter ε or less

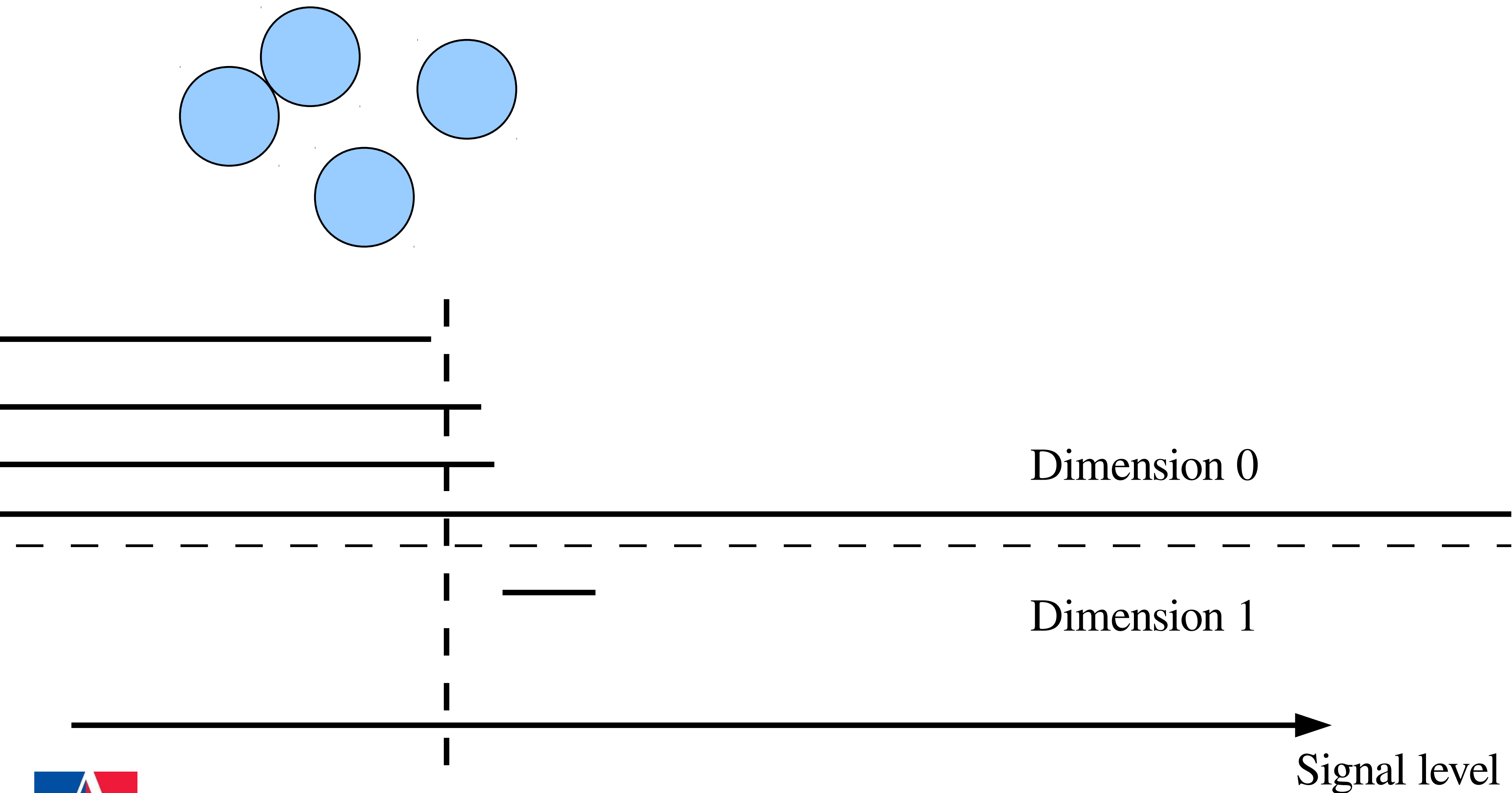
- This is an abstract simplicial complex, and

$$\mathrm{VR}_\varepsilon(X) \subseteq \mathrm{VR}_\eta(X) \text{ if } \varepsilon \leq \eta$$

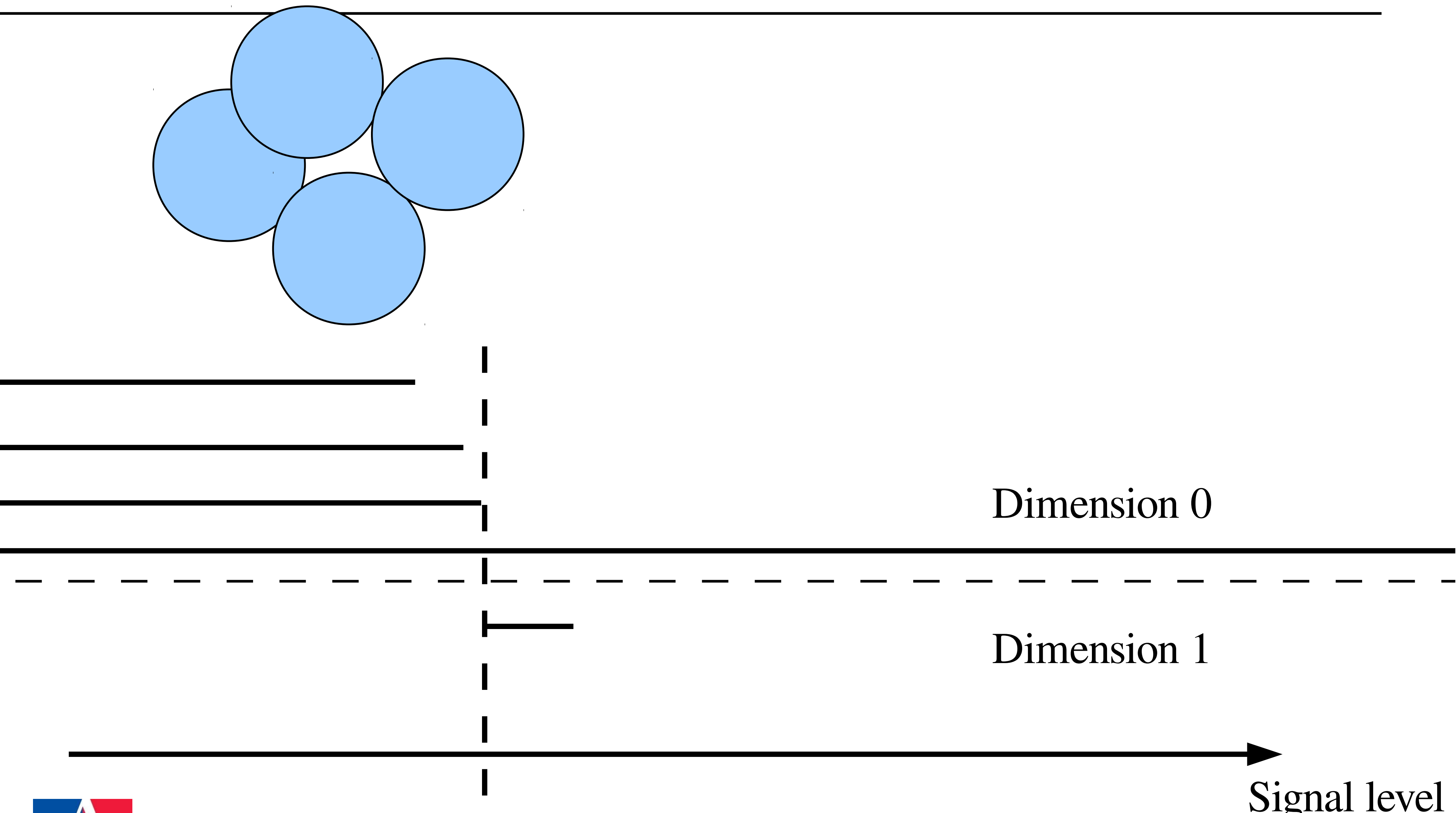
Persistence and model robustness



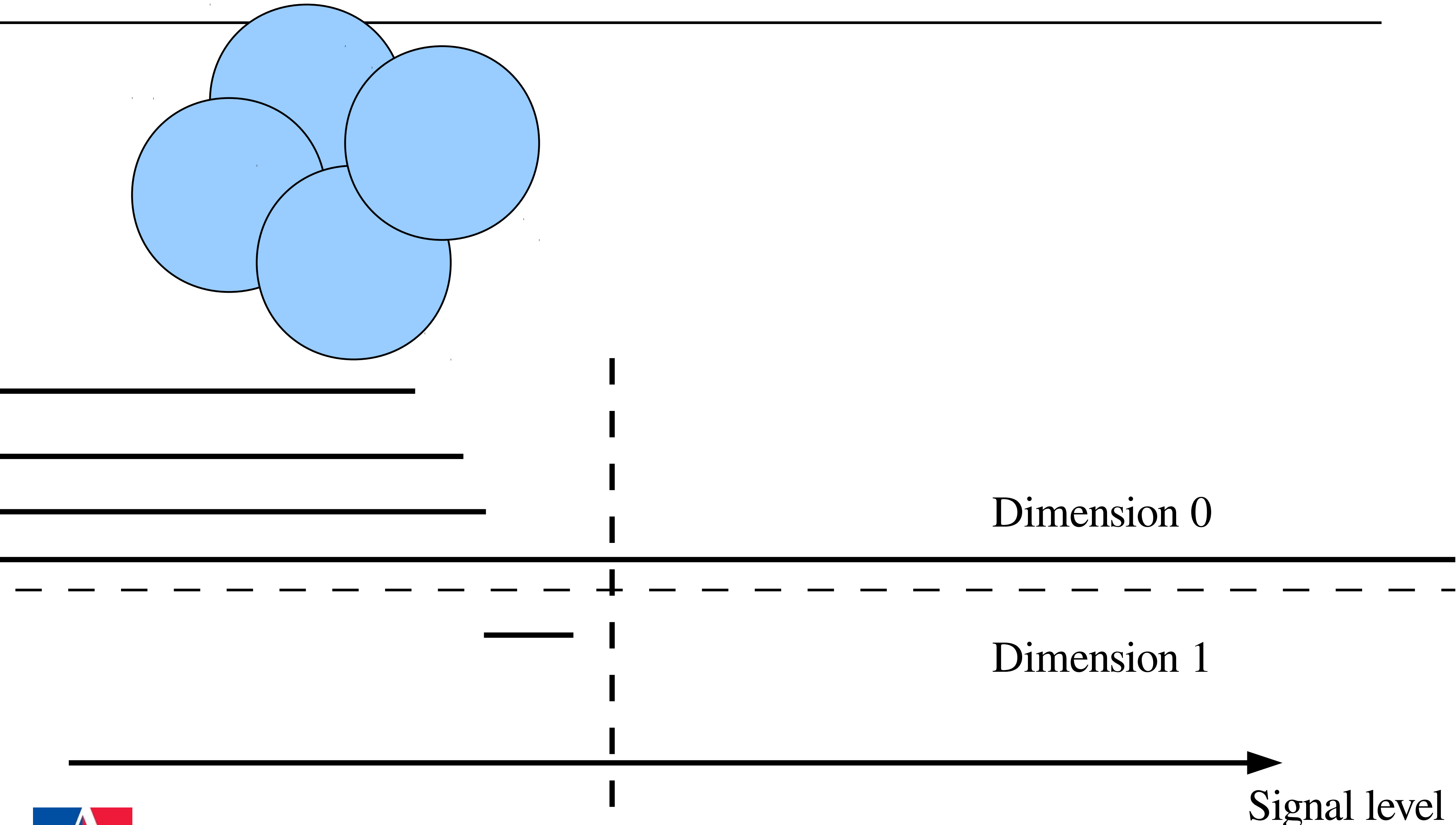
Persistence and model robustness



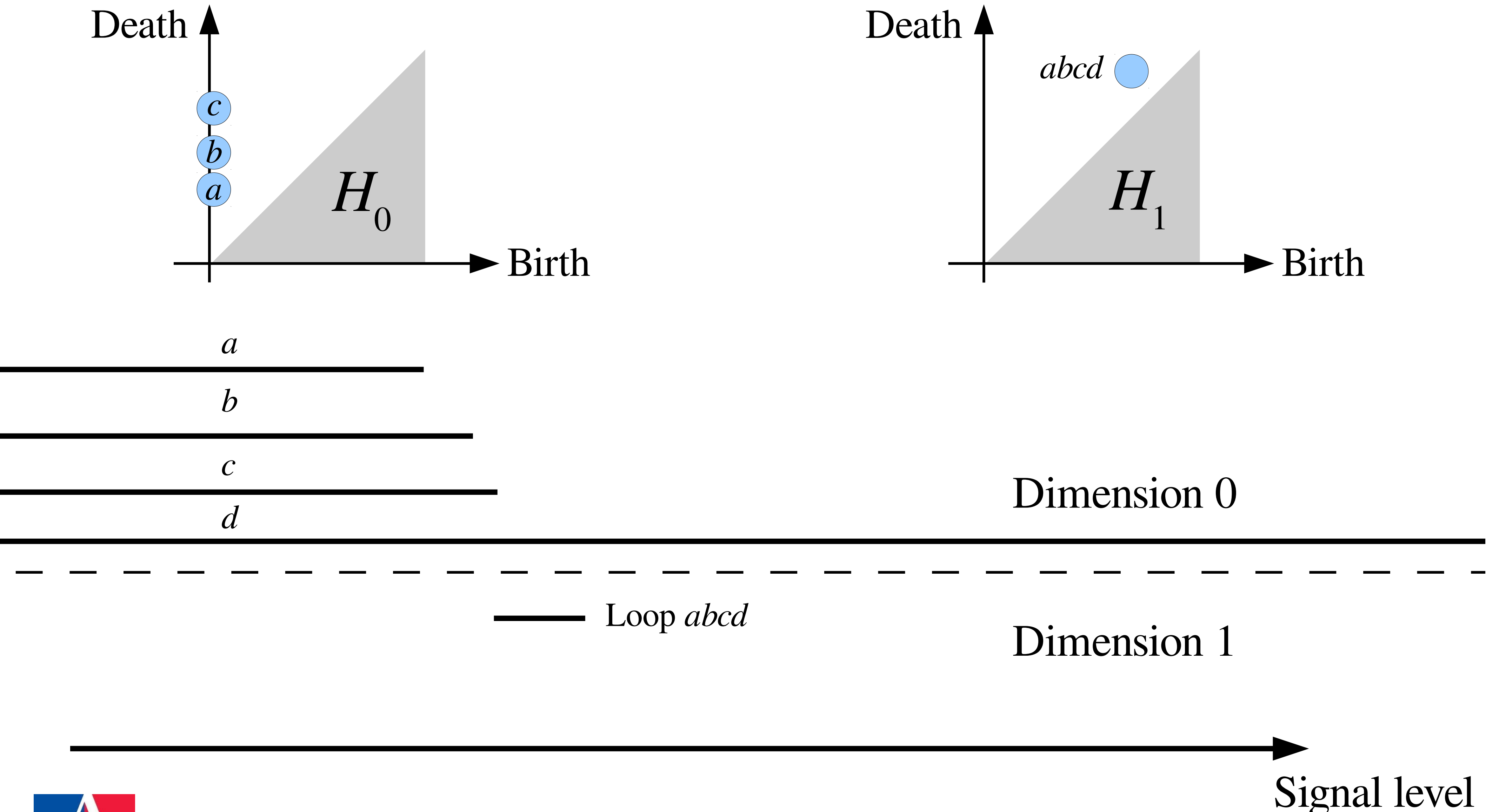
Persistence and model robustness



Persistence and model robustness



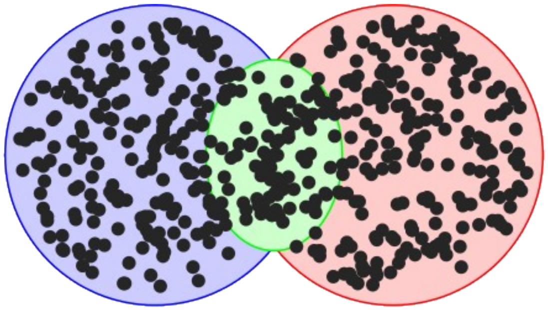
Persistence diagrams



This sessions' talks

Hierarchical Overlapping Clustering

Fernando Gama, Santiago Segarra, Alejandro Ribeiro



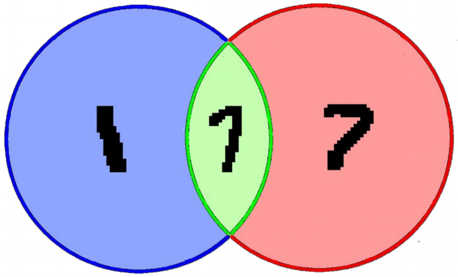
- **Clustering:** Partition dataset. But sets might not admit a partition.
- Where should the green points go?

- **Coverings:** Points can be classified in more than one cluster.

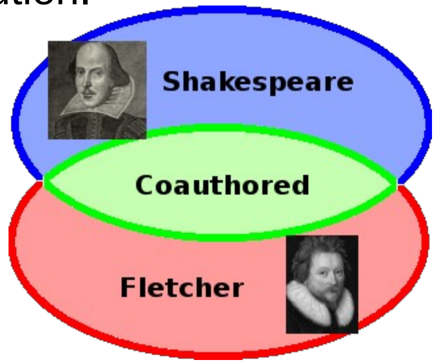
- Hierarchical Overlapping Clustering through Cut Metrics.
- Non-overlapping Clustering: Each point in only one cluster. Partition.
- Hierarchical Clustering: Collection of partitions. Resolution of clusters.
- Ultrametrics: Determines resolution at which nodes are clustered together.
- Non-overlapping Clustering → Hierarchical Clustering → Ultrametrics

- Convex combination of ultrametrics → Cut Metrics
- Cut Metrics: Resolution at which nodes are grouped together. No transitivity.
- Nested coverings: Collection of coverings.
- Covering: Nodes can be in more than one group.
- Cut Metrics → Nested Coverings (Hierarchical) → Covering

- **Results:** MNIST Handwritten Digits. Authorship Attribution.



Identify ambiguous digits
When classifying 1 and 7



Classify plays by author
and co-authored plays

Distances between Directed Networks and Applications

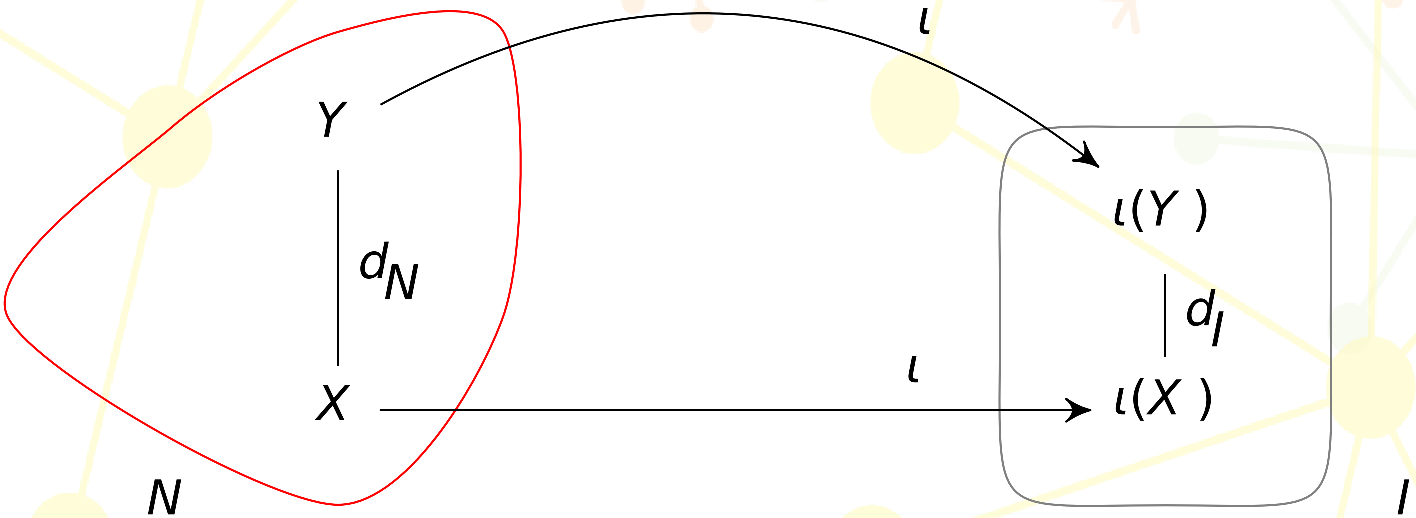
Facundo Mémoli, joint work with Samir Chowdhury

Background:

- Data sets containing asymmetric edge relations can be interpreted as directed, weighted networks.
- A central goal of network analysis is to develop metrics that efficiently compute dissimilarity between networks.

Our Contributions:

- A legitimate metric between directed, weighted networks.
- Easily computable invariants to test for dissimilarity between networks.
- Lower bounds for the network distance, based on these invariants.



Hypergraph signal processing

Sergio Barbarossa and Mikhail Tsitsvero

Idea: Hypergraphs generalize graphs and simplicial complexes

All cliques are simplices → Flag complex → Abstract simplicial complex

Vertices and edges only → Undirected graph → Abstract simplicial complex

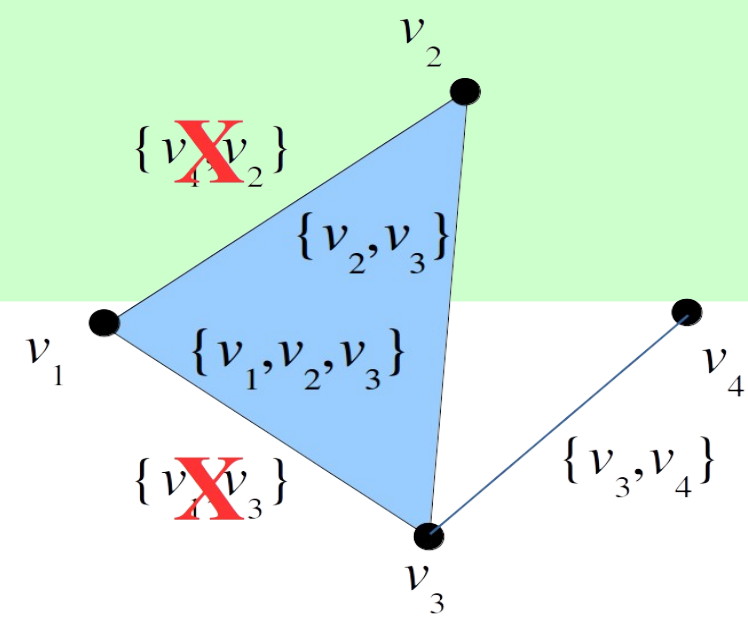
Simplices are defined by their vertices, simplices always have all their faces

Abstract simplicial complex

↓

Undirected hypergraph

Hyperedges defined by vertices, but other than that, anything goes...

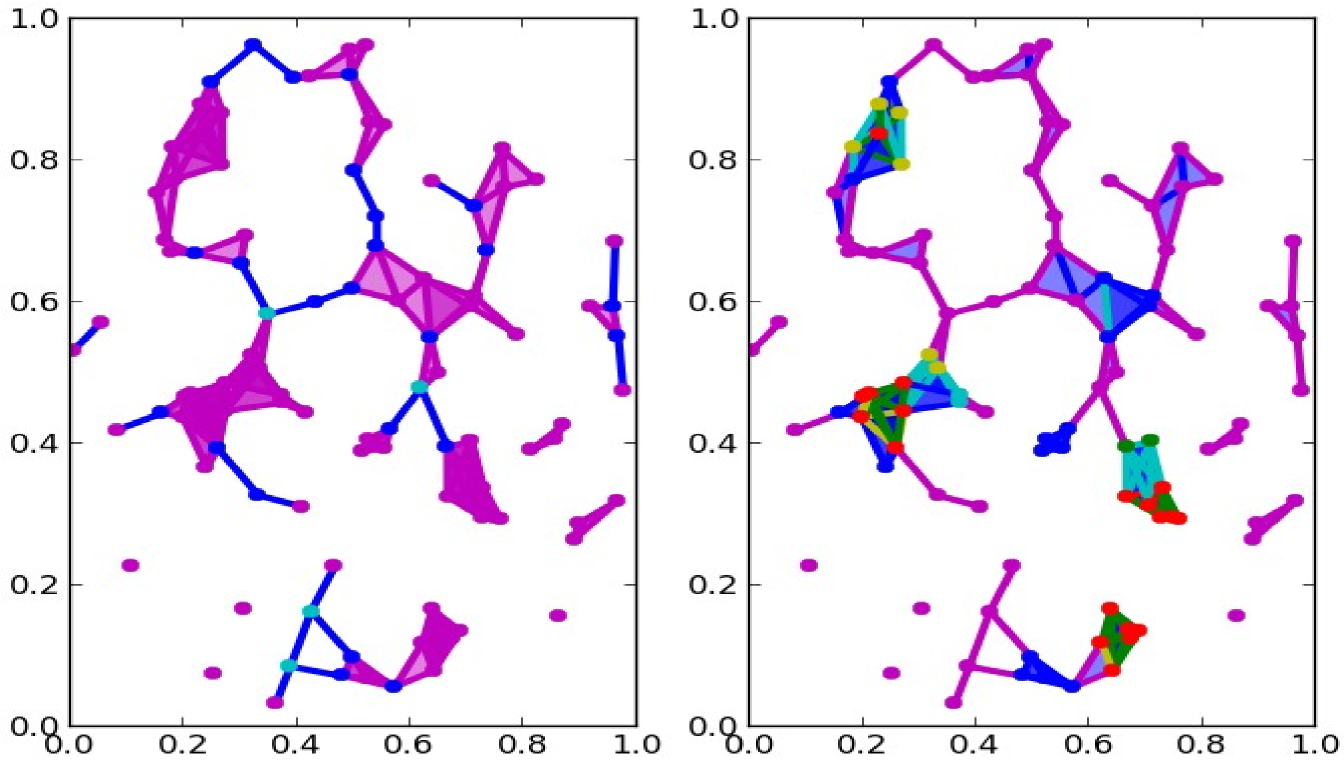


Persistent local homology

Brittany Fasy and Bei Wang

Idea: local homology detects stratifications!

Magenta = 0
Blue = 1
Cyan = 2
Green = 3
Yellow = 4
Red = 5+

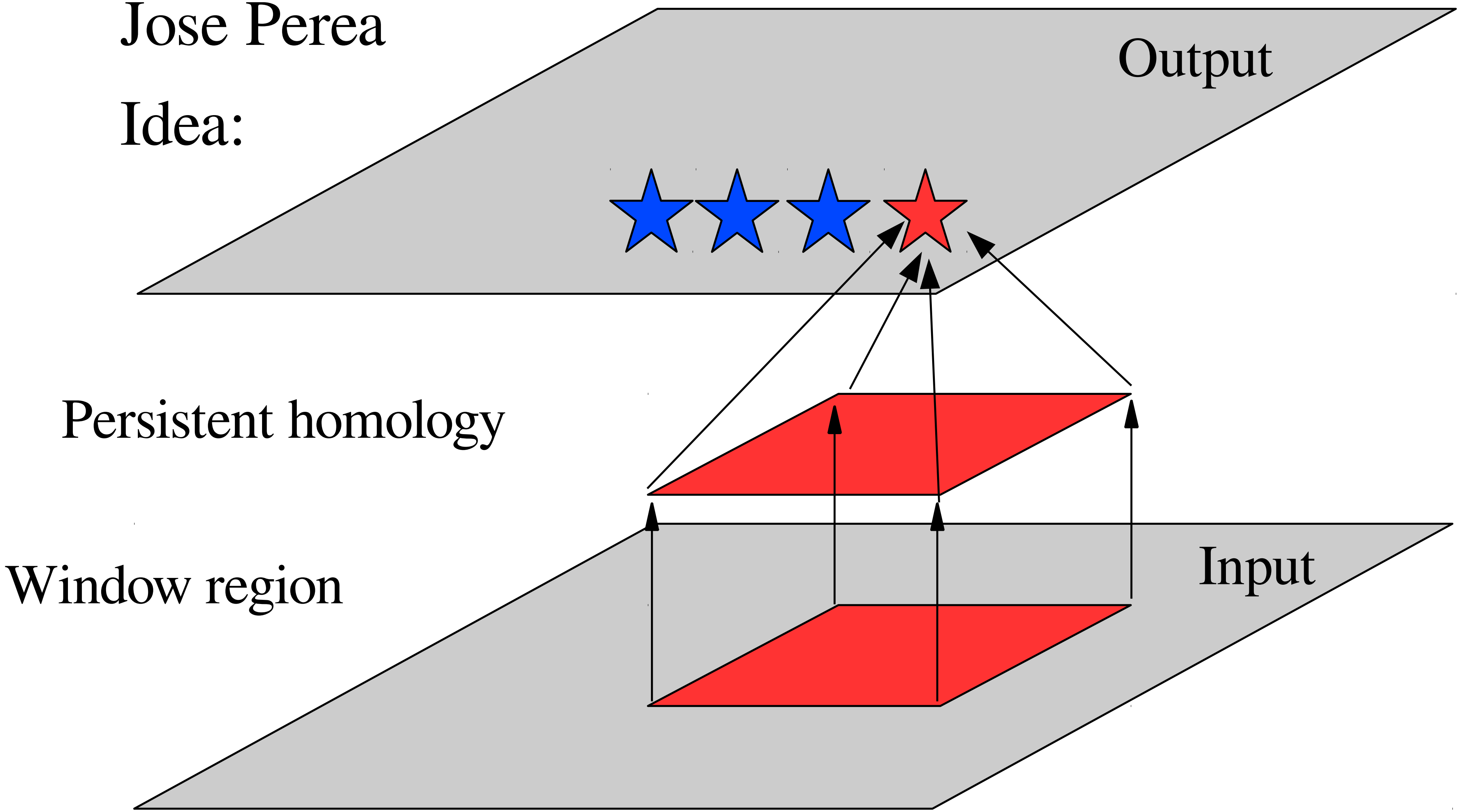


Local H_1 Local H_2

Persistent homology and sliding windows

Jose Perea

Idea:



For more information

...or if you need anything...

Harish Chintakunta, hchintakunta@flpoly.org

Michael Robinson, michaelr@american.edu

Hamid Krim, ahk@ncsu.edu