The Waisman Laboratory for Brain Imaging and Behavior

University of Wisconsin SCHOOL OF MEDICINE AND PUBLIC HEALTH

Lecture 1 Simplicial homology and persistent homology

Moo K. Chung
Department of Biostatistics and Medical Informatics
University of Wisconsin-Madison

https://github.com/laplcebeltrami/ISBI2023TDA

Acknowledgement

Zijian Chen, Tahmineh Azizi, Vijay Anand D, Sixtus Dakurah, Soumya Das, Tananun Songdechakraiwut, Univ. of Wisconsin-Madison, US

Shih-Gu Huang, Anqi Qiu National University of Singapore, Singapore

Anass El Yaagoubi Bourakna, Hernando Ombao KAUST, Saudi Arabia

Hyekyung Lee, Dong Soo Lee Seoul National University, Korea

Grants: *NIH R01 EB022856, R01 EB028753, NSF DMS-2010778*

Matlab toolbox PH-STAT

Statistical Inference on Persistent Homology

https://github.com/laplcebeltrami/PH-STAT

The codes are used to publish in leading journals and conferences since 2009: IEEE Transactions on Medical Imaging, NeuroImage, Human Brain Mapping, Annals of Applied Statistics, Information Processing in Medical Imaging (IPMI), MICCAI, ISBI

Corresponding functions are color coded.

`WS_cluster.m` `PH_hodge_betti.m`

References

Gunnar Carlsson 2009, A User's Guide to Topological Data Analysis

Herbert Edelsbrunner and John L. Harer Computational Topology: An Introduction 2010, American Mathematical Society

Chung et al. 2020 Review: Toplogical distance and losses in brain networks arXiv:2102.08623

Chung PH-STAT manual https://arxiv.org/pdf/2304.05912.pdf

Topological Data Analysis

- Branch of data science that uses topology
- Study properties of data that remain invariant under continuous transformations
- Identify underlying topologically invariant patterns across subjects

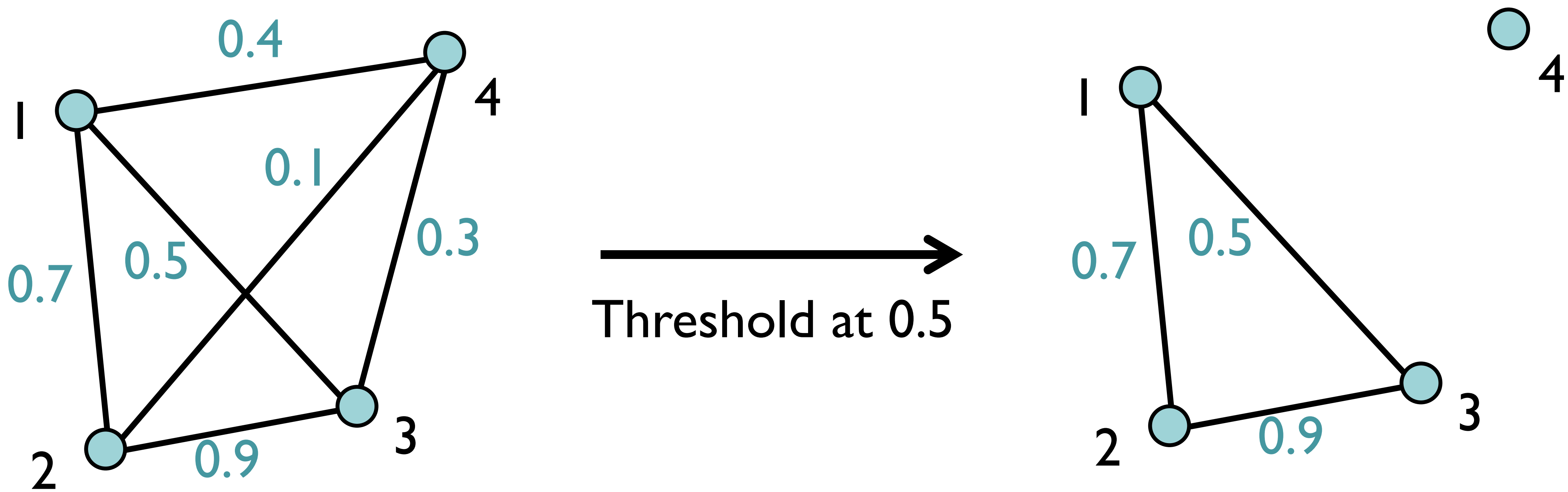
Persistent Homology

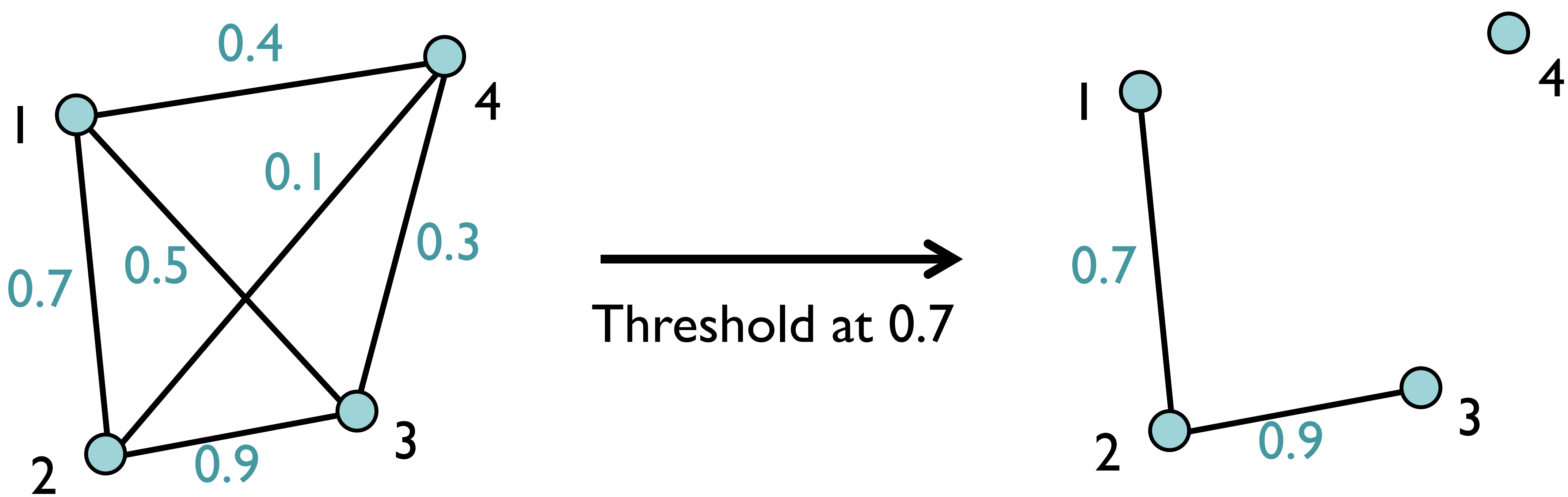
- Captures the topological features of data across different scales
- More persistent topological features → signal
- Less persistent topological features → noise
- Multiscale approach

What is wrong with standard brain network analysis?

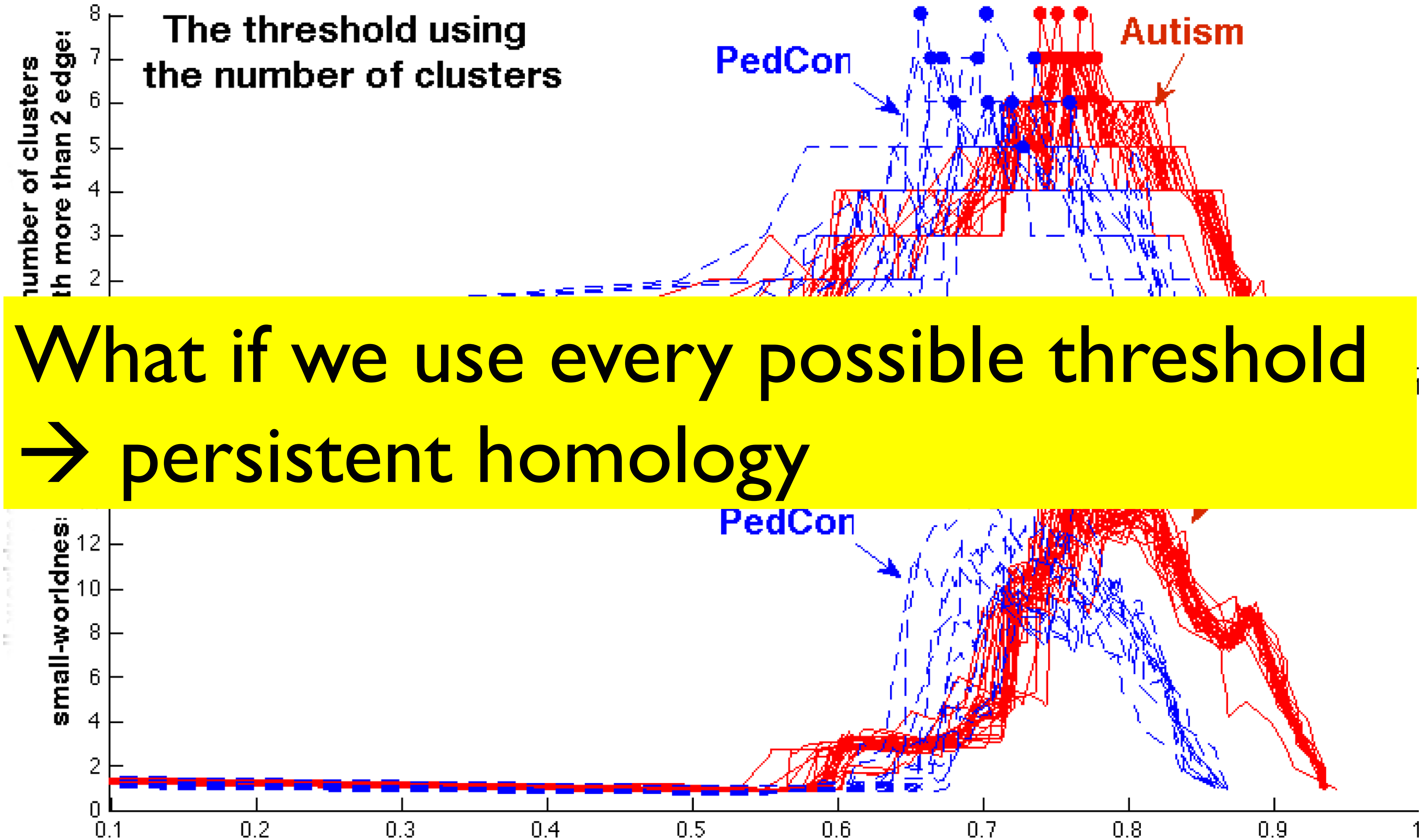
Edge weight ρ_{ij} between node *i* and *j*

→ Connectivity matrix $\rho = (\rho_{ij})$





Single threshold often suboptimal → multiple thresholds

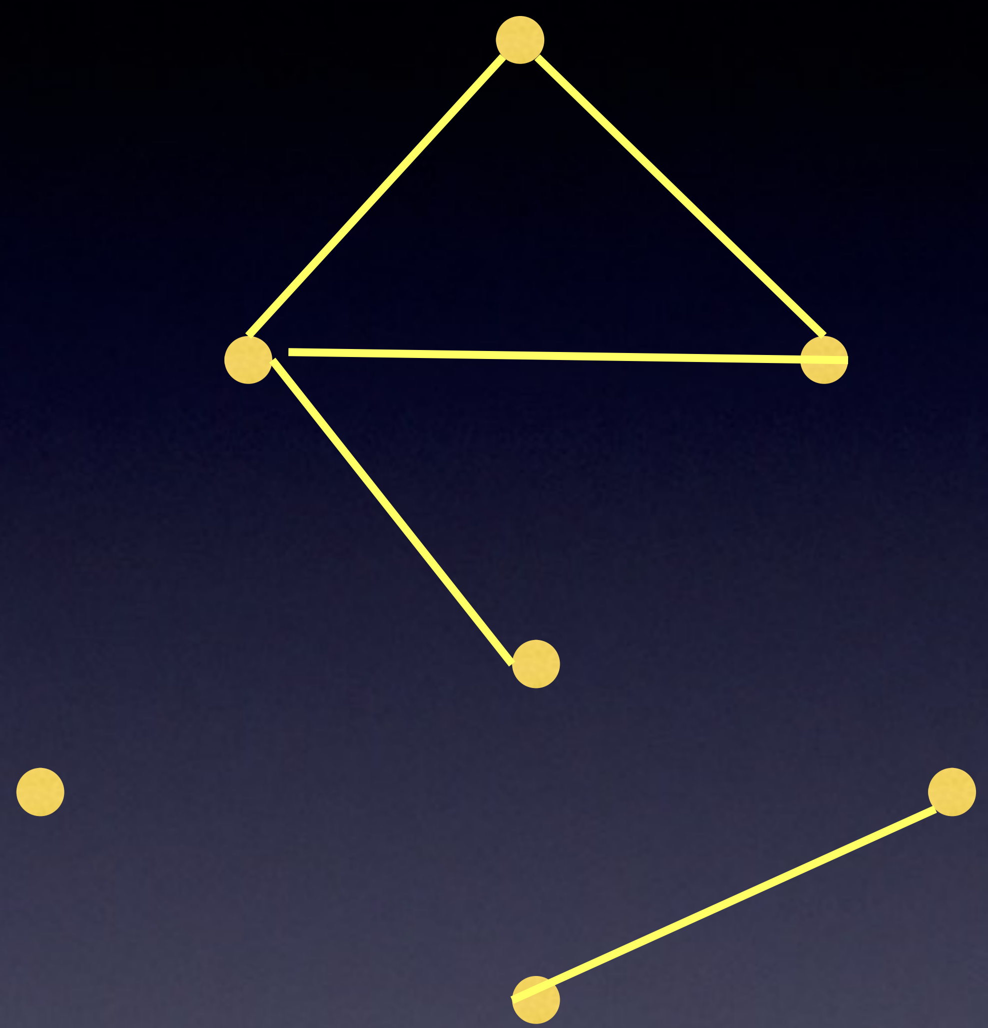


What if we use every possible threshold
→ persistent homology

Lee et al. 2011. IEEE TMI 30:1154-1165

Betti numbers β_i

of i-dimensional holes/loops



β_0 = # of connected components = 3

β_1 = # of cycles = 1

Euler characteristic: $\chi = 3 - 1 = 2$

Betti numbers β_i # of i-dimensional holes/loops

β_0 = # of connected components = 3

β_1 = # of 1D holes = 1

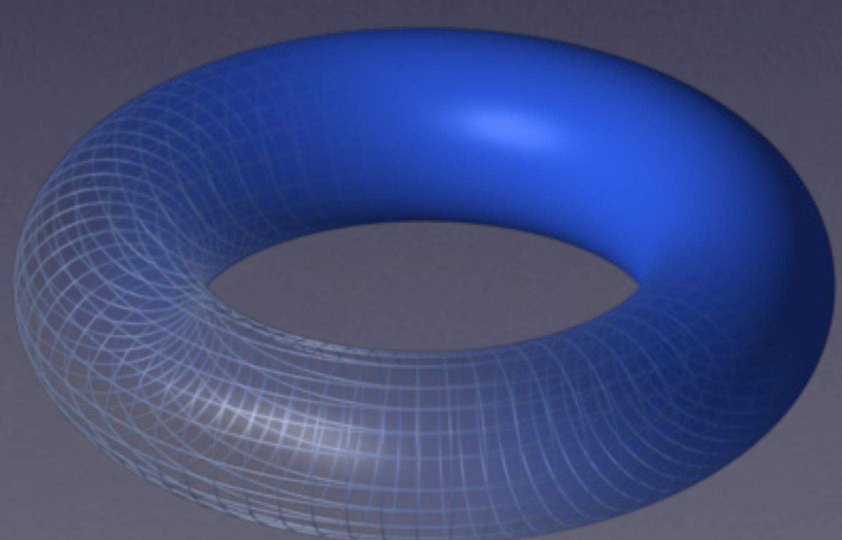
β_2 = # of 2D cavities =0

Betti-number representation:

(3,1,0,0,…)

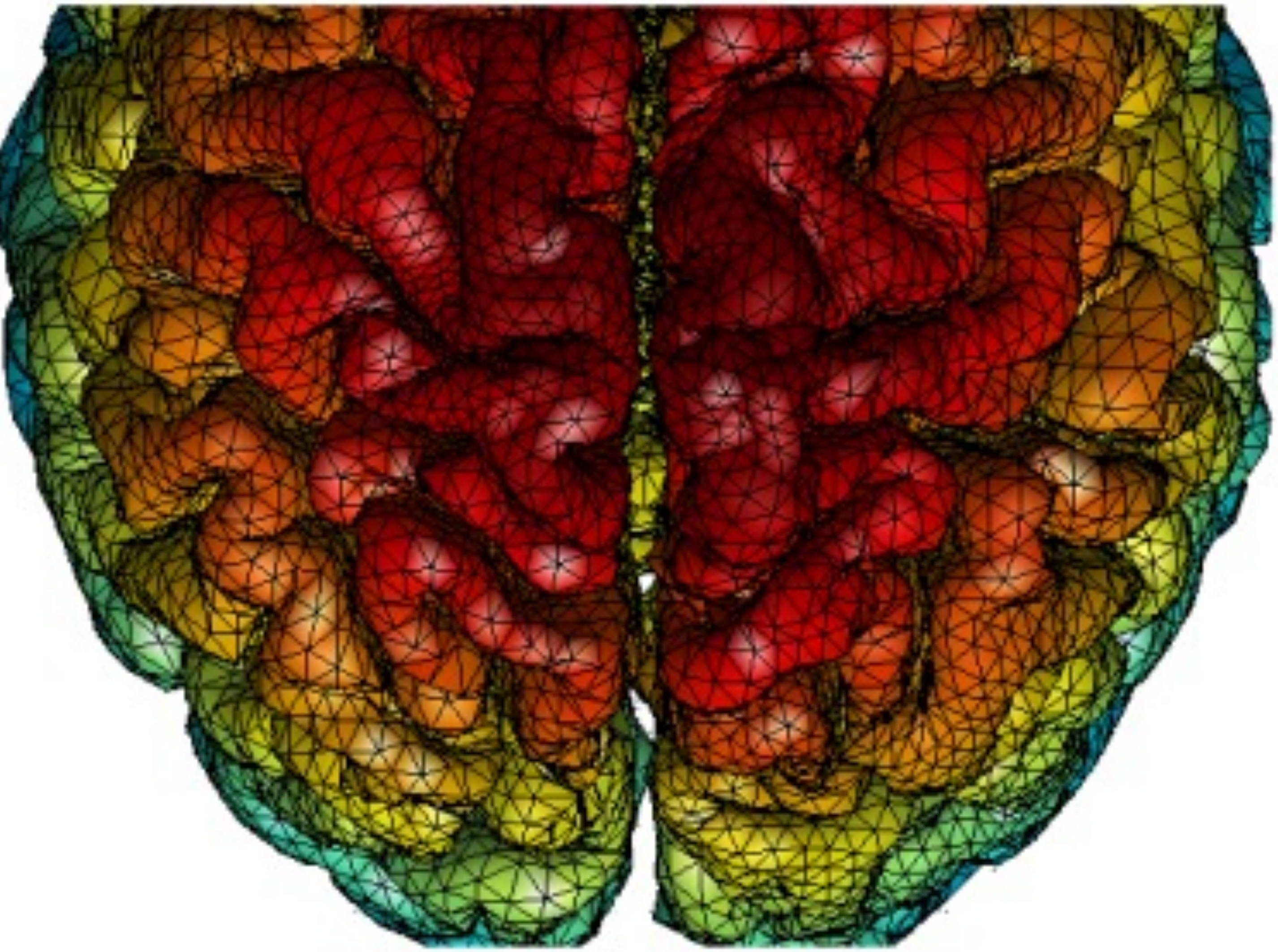
Euler characteristic:

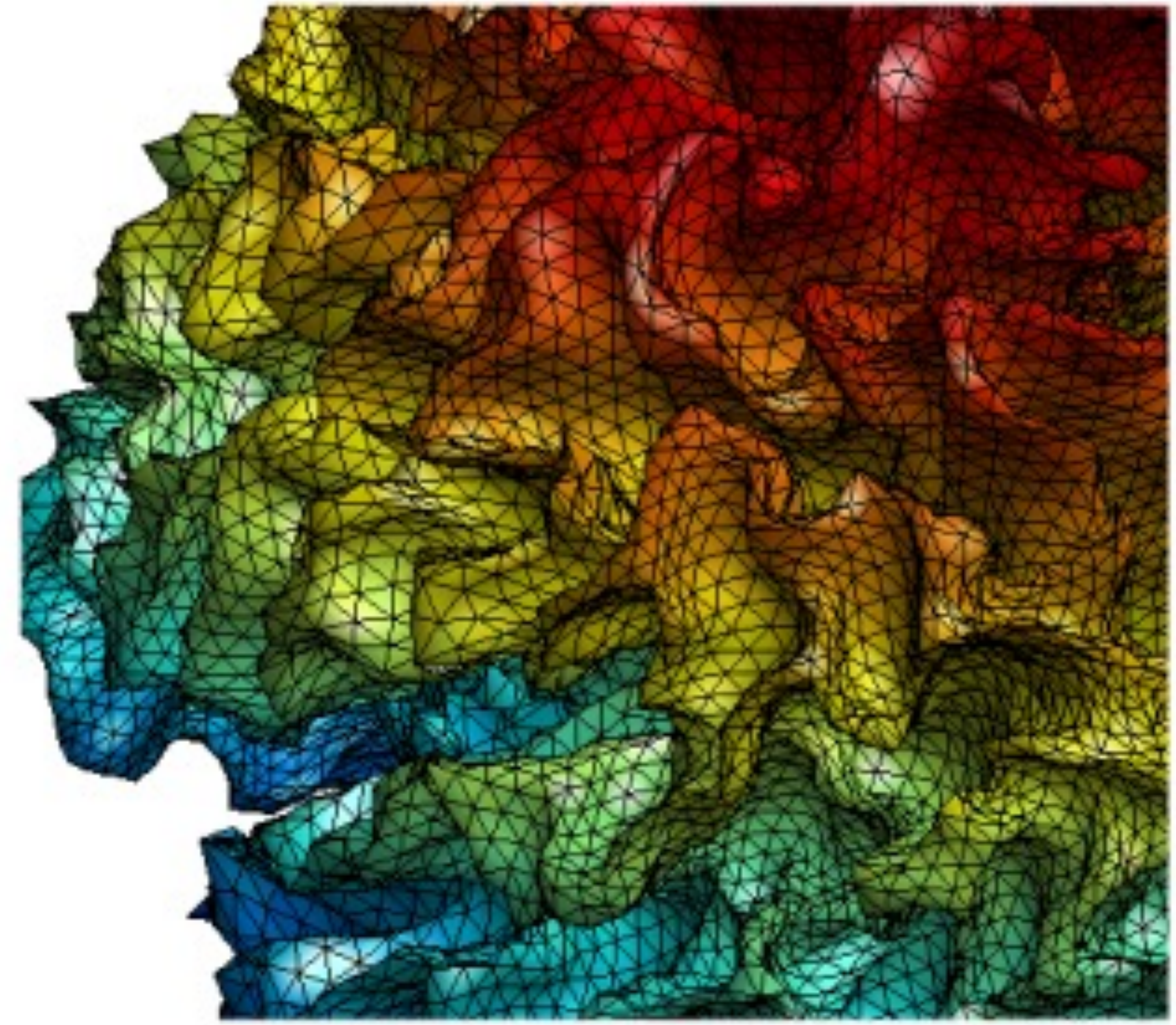
$$\chi = \beta_0 - \beta_1 = 2$$



$$\beta_0 = 1, \beta_1 = 2, \beta_2 = 1$$

(1,2,1,0,0,…)



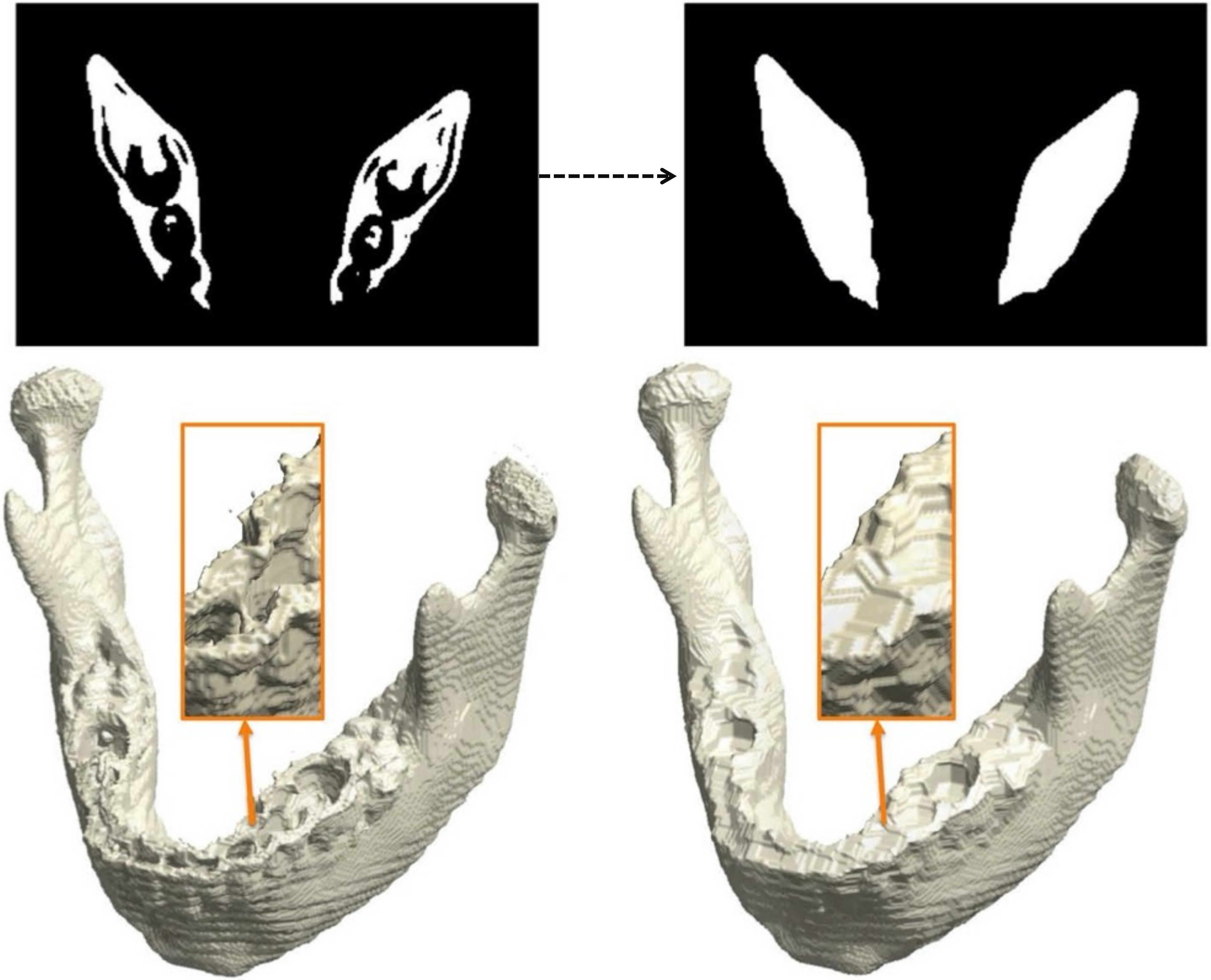


Euler characteristic of a surface mesh from SurfStat

$N - E + F = 2$ for a surface topologically equivalent to a sphere. For each triangle, there are three edges. Since two adjacent triangles share the same edge, the total number of edges is $E = 3F/2$. Hence, we have $F=2N-4$ for a closed surface.

Can be used to check topological artifacts in FreeSurfer

Topology correction in CT segmentation



Hole & handles corrected using Euler characteristic

Chung et al. 2015 Medical Image Analysis. 22:63-76

Keith Worsley's random field theory

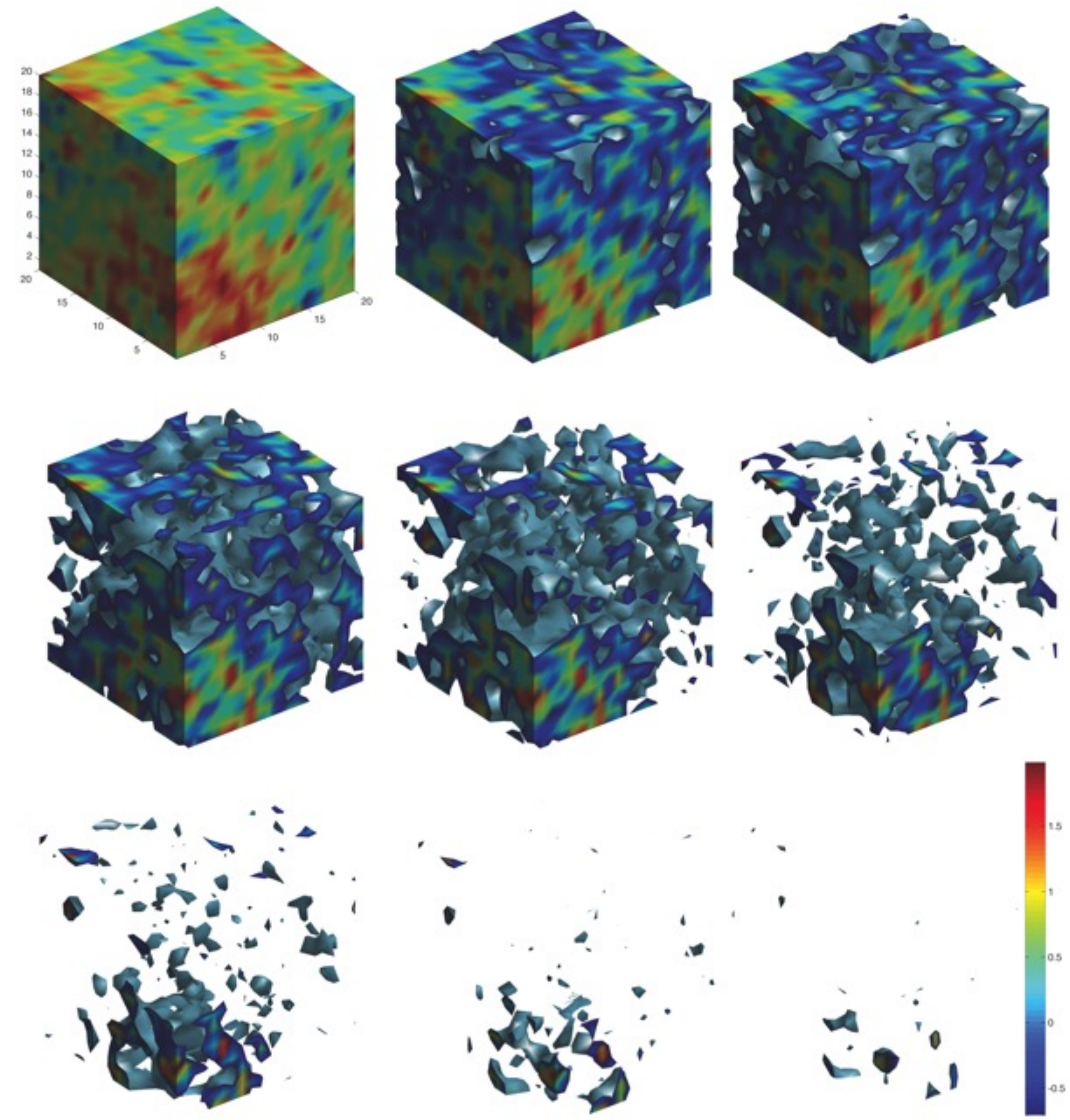
Random field

$$P\Big(\sup_{x\in\mathbb{M}} T(x) > h\Big)$$

$$A_h = \{x \in \mathbb{M} : T(x) > h\}$$

$$P\Big(\sup_{x\in\mathbb{M}} T(x) > h\Big) = \mathbb{E}\chi(A_h)$$

$$\chi(A_h) = \sum_j (-1)^j \beta_j(A_h)$$



Milor 1963 Morse theory

Adler, 1994 The geometry of random fields

Worsley et al., 1996 Human Brain Mapping

Morse Filtration

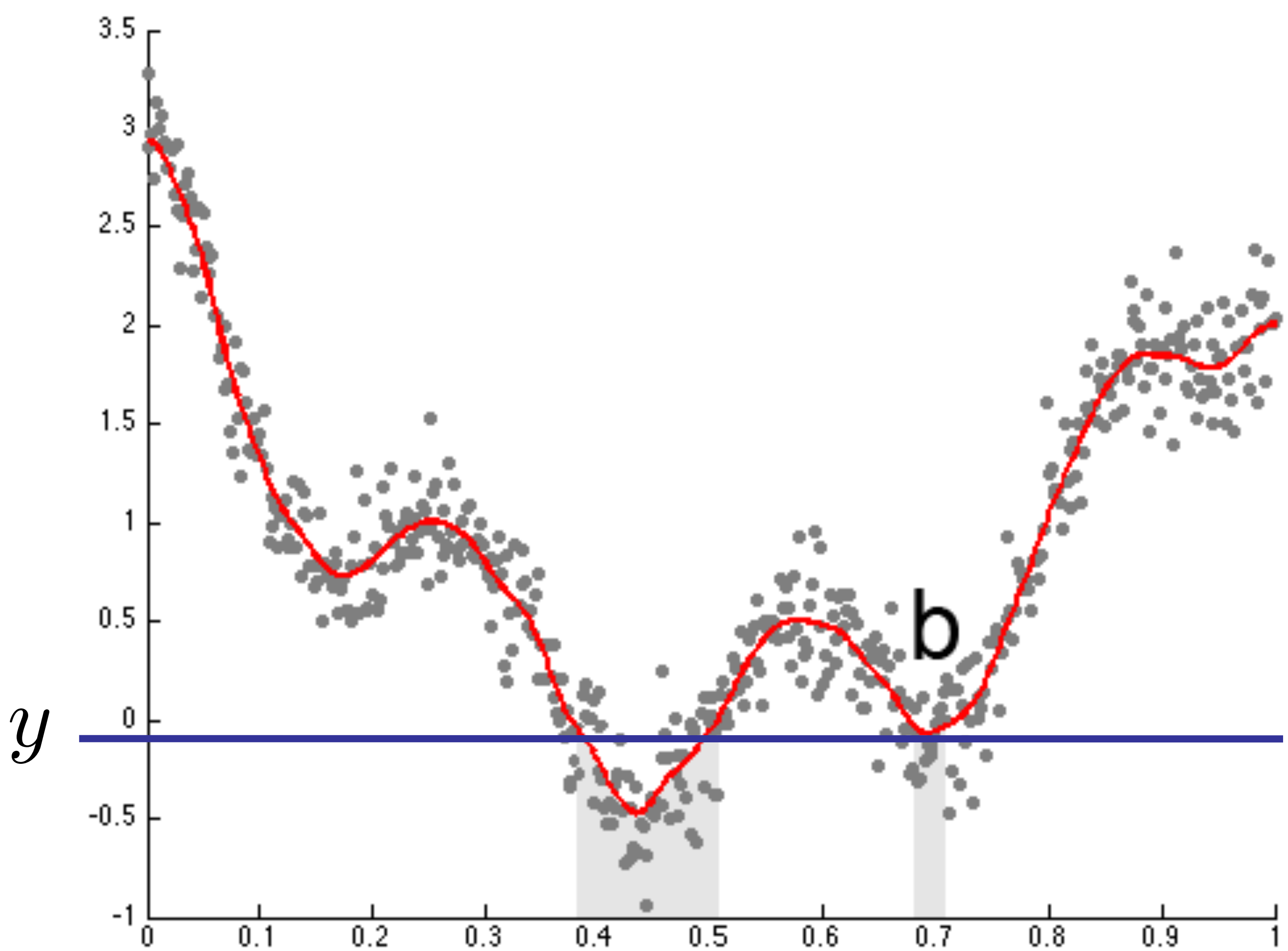
<u>Chung et al., 2009</u> Information Processing in Medical Imaging (IPMI) 5636:386-397.

`PH_morse1D.m`

Morse theory for functional data

$$Y(t) = \mu(t) + \epsilon(t)$$

Unknown signal μ is assumed to be a Morse function: all critical values are unique.



Sublevel set

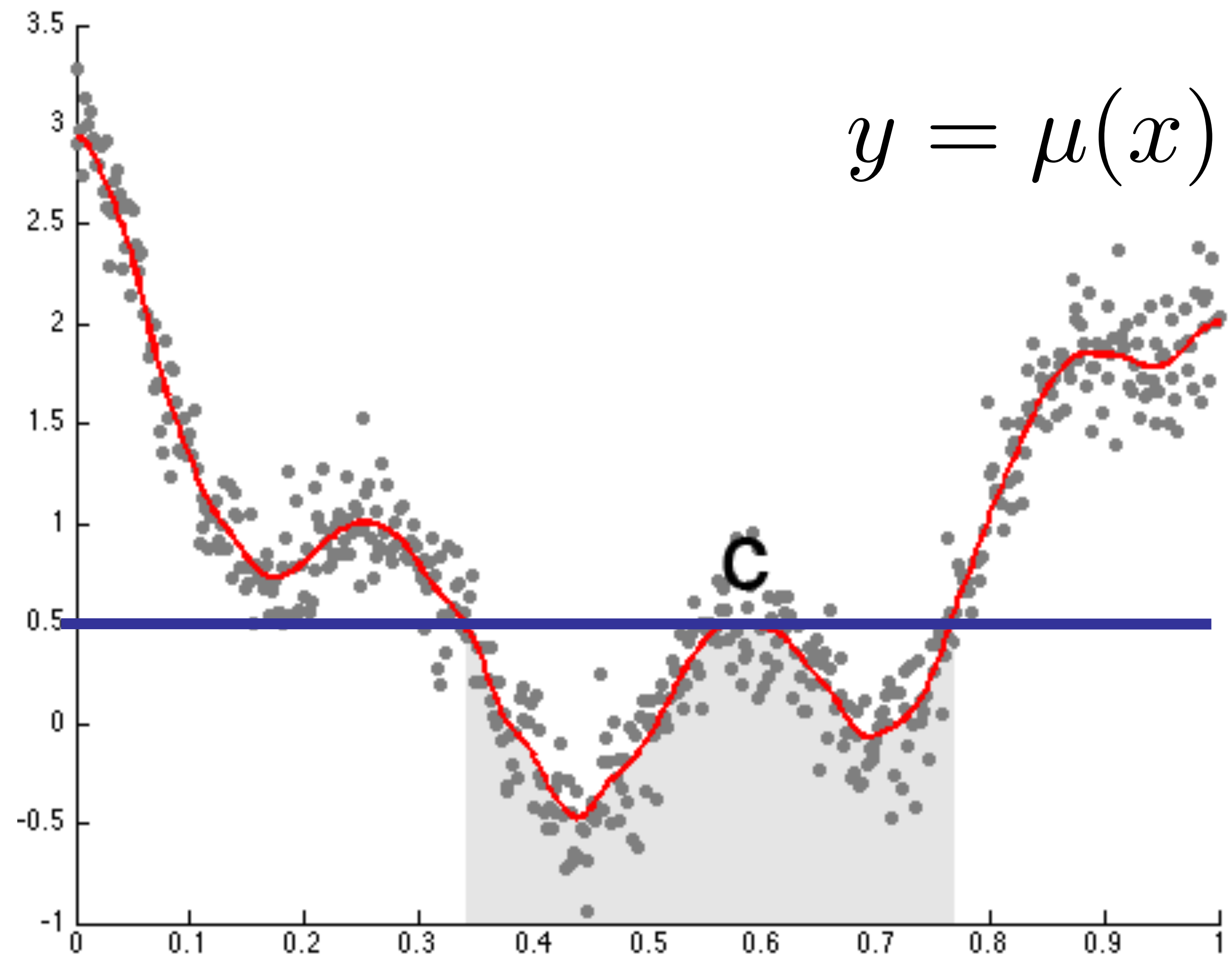
$$R_y = \{t : \mu(t) \leq y\}$$

The topology of sublevel set is characterized by Betti-0 number only

Morse filtration

Sublevel set

$$R_y = \{t : \mu(t) \leq y\}$$



Morse filtration

$$R_b \subset R_c$$

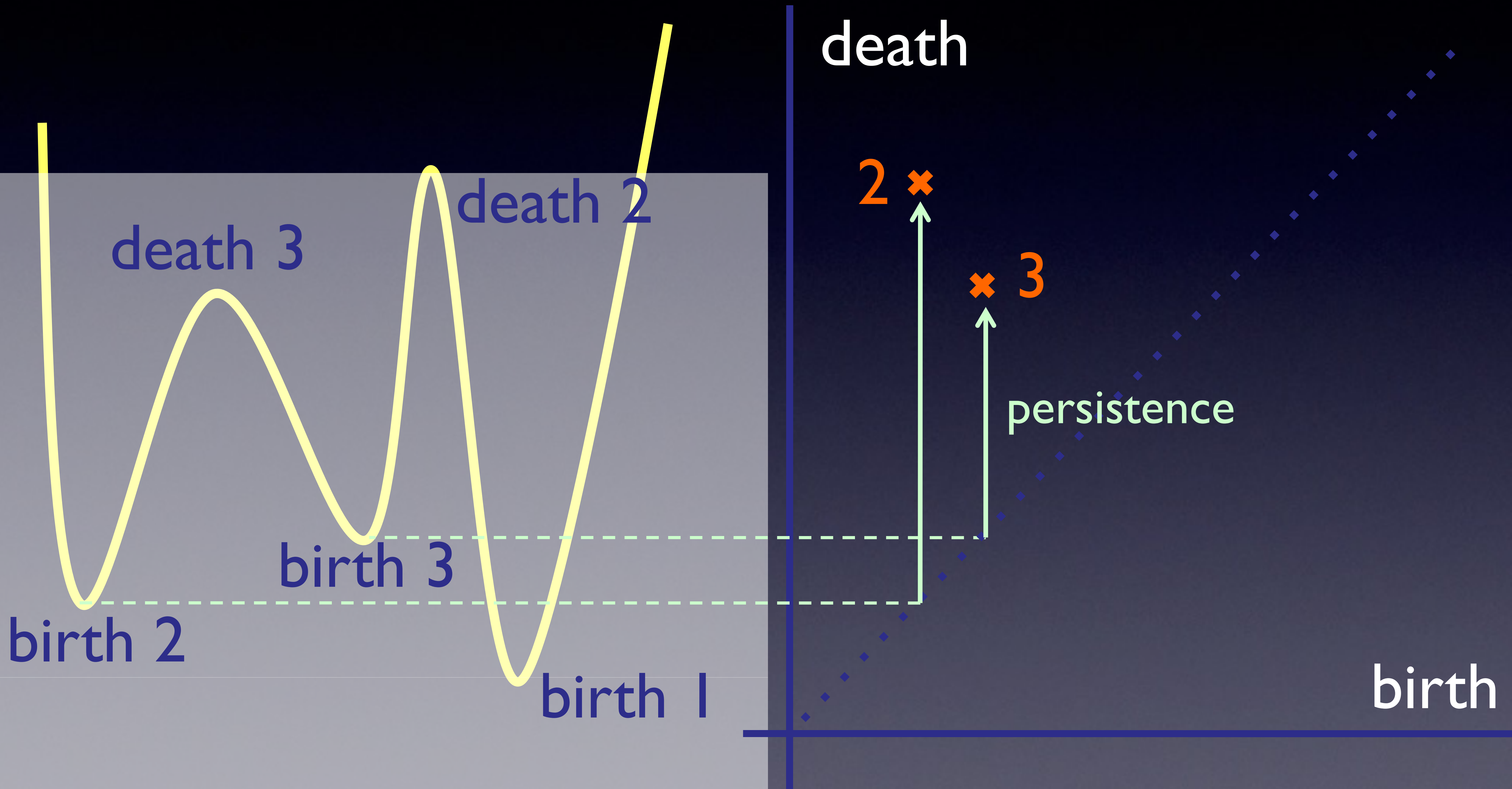
Component dies at c

$$\beta_0(R_c) = \beta_0(R_b) - 1$$

PH_morse1D.m

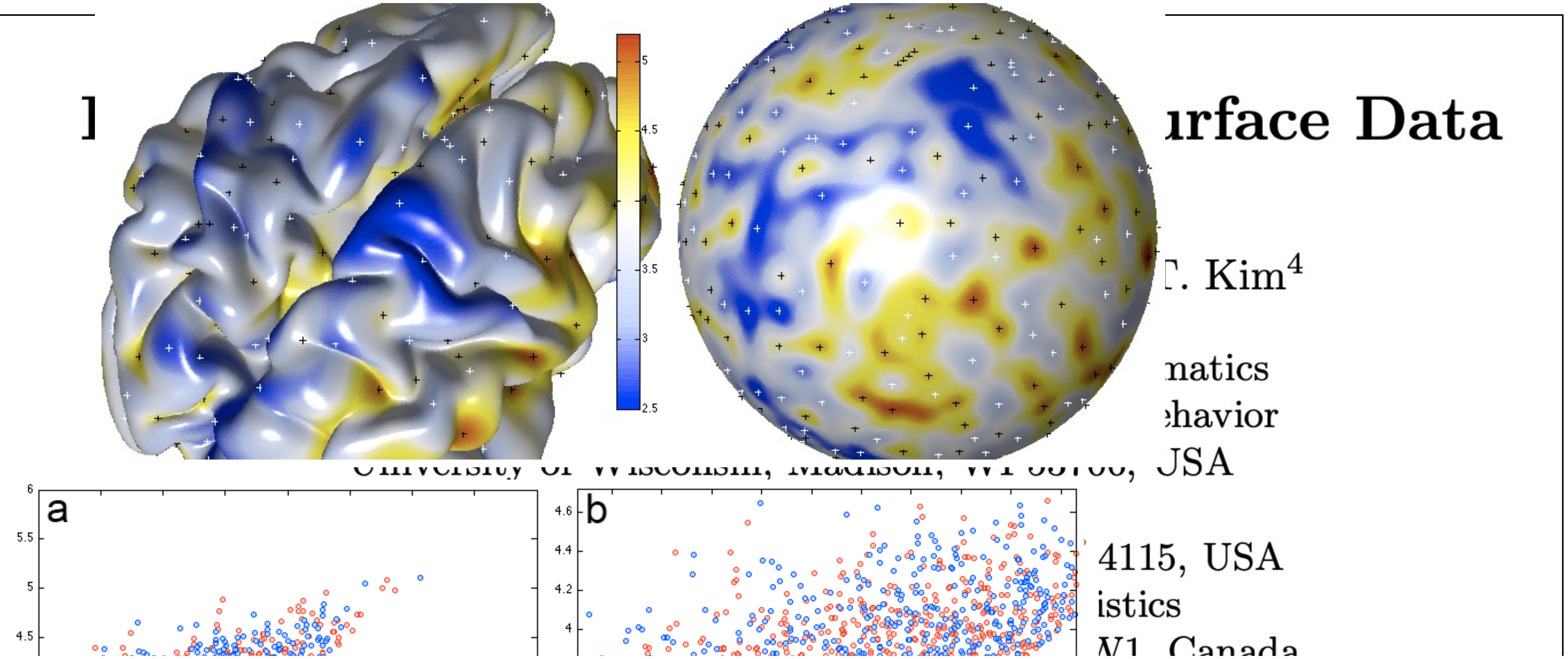
Persistence Diagram (PD)

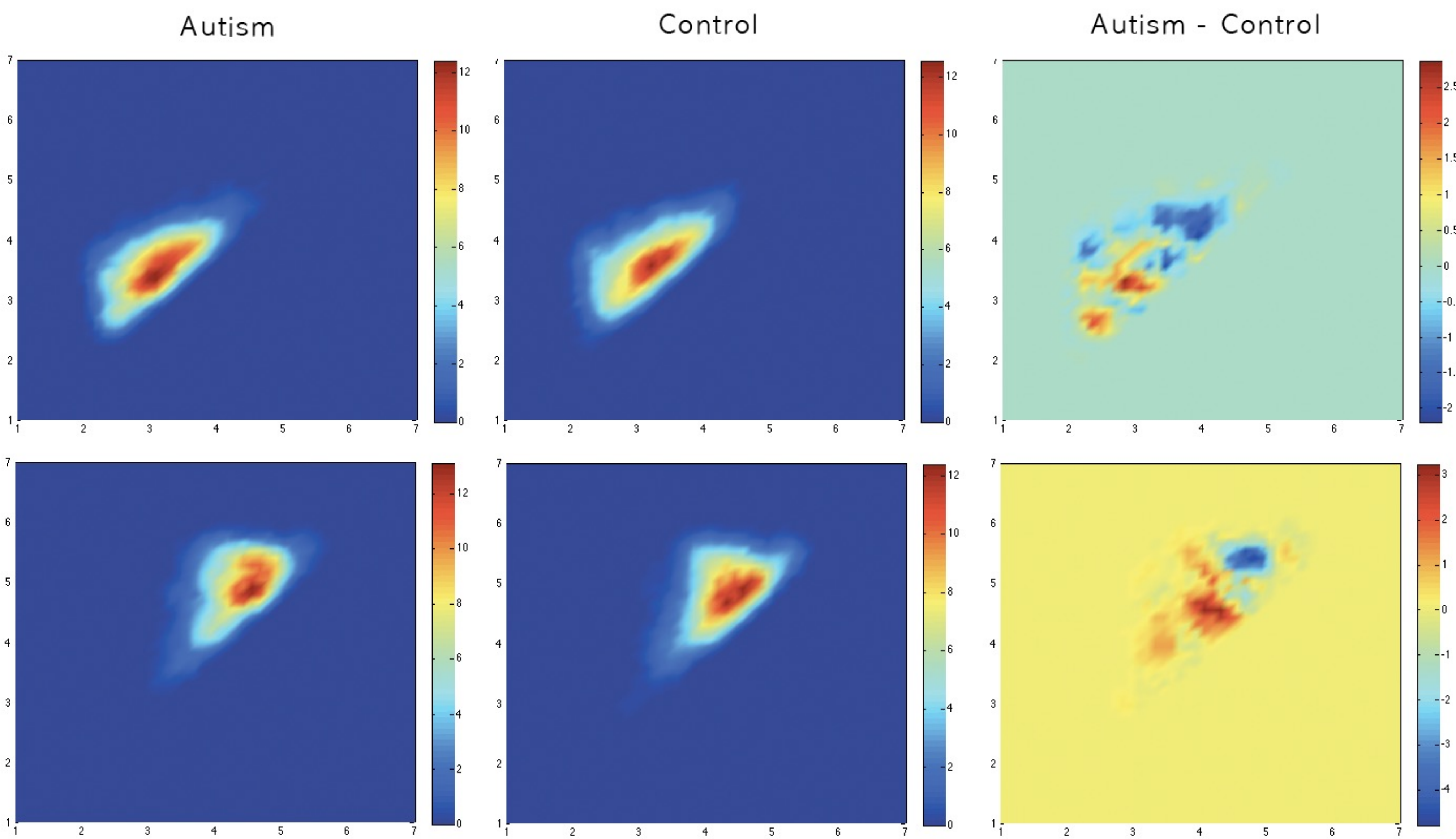
O(n log n)



Elder's rule:

Pair the time of death with the time of the closest earlier birth.





Chung et al., 2009
Information Processing in Medical Imaging (IPMI) 5636:386-397.

First TDA paper in medical imaging

Boundary matrix

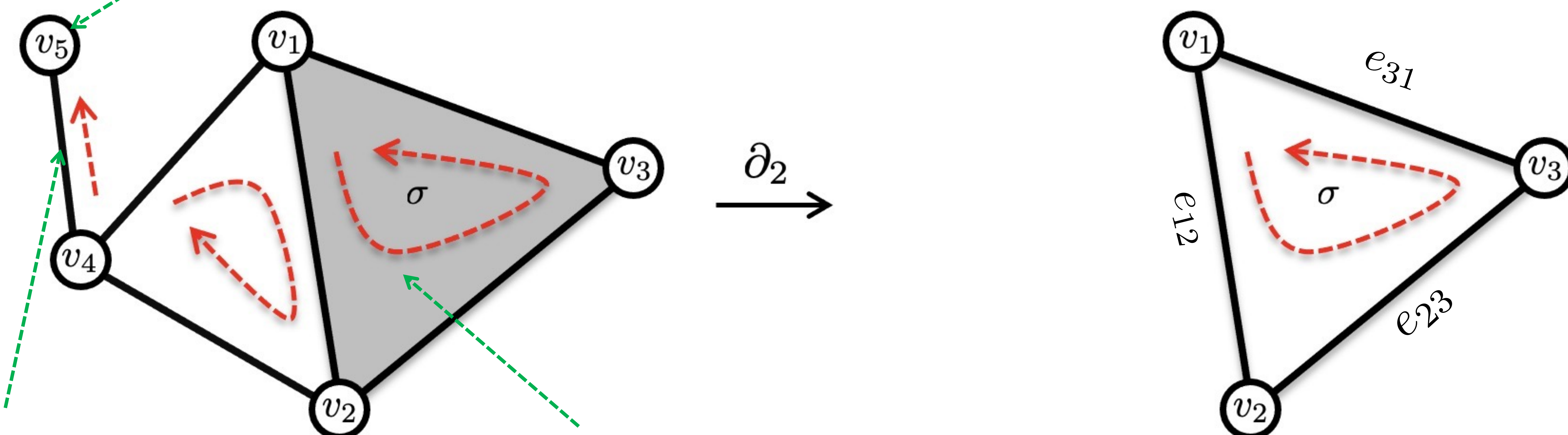
n-simplex

The basic building block of persistent homology

The smallest convex set containing $n+1$ points

$$\sum_{i=0}^{n} x_i = 1, x_i \geq 0$$

0-simplex $[v_5]$



1-simplex

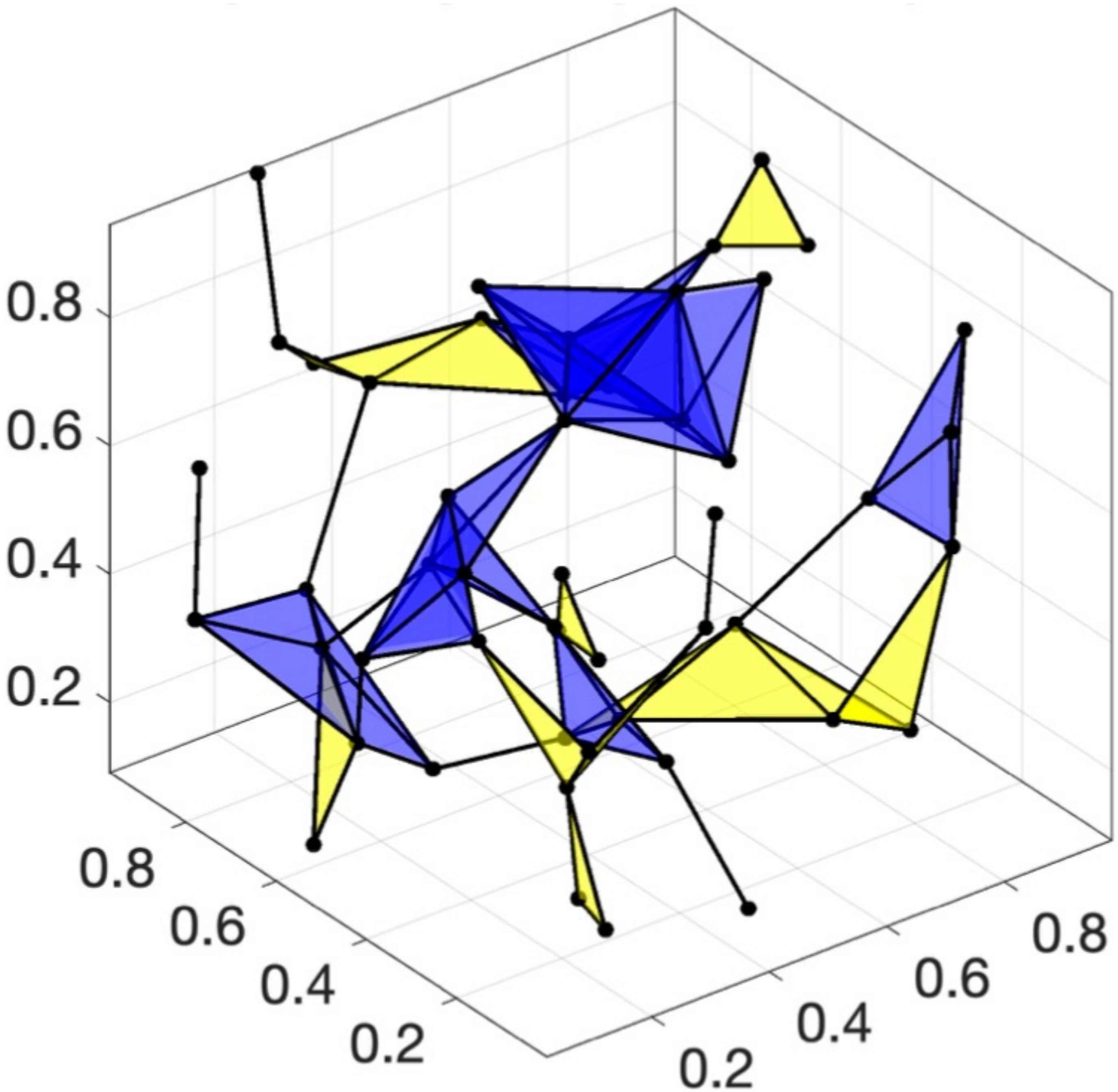
$[v_4, v_5] = -[v_5, v_4]$

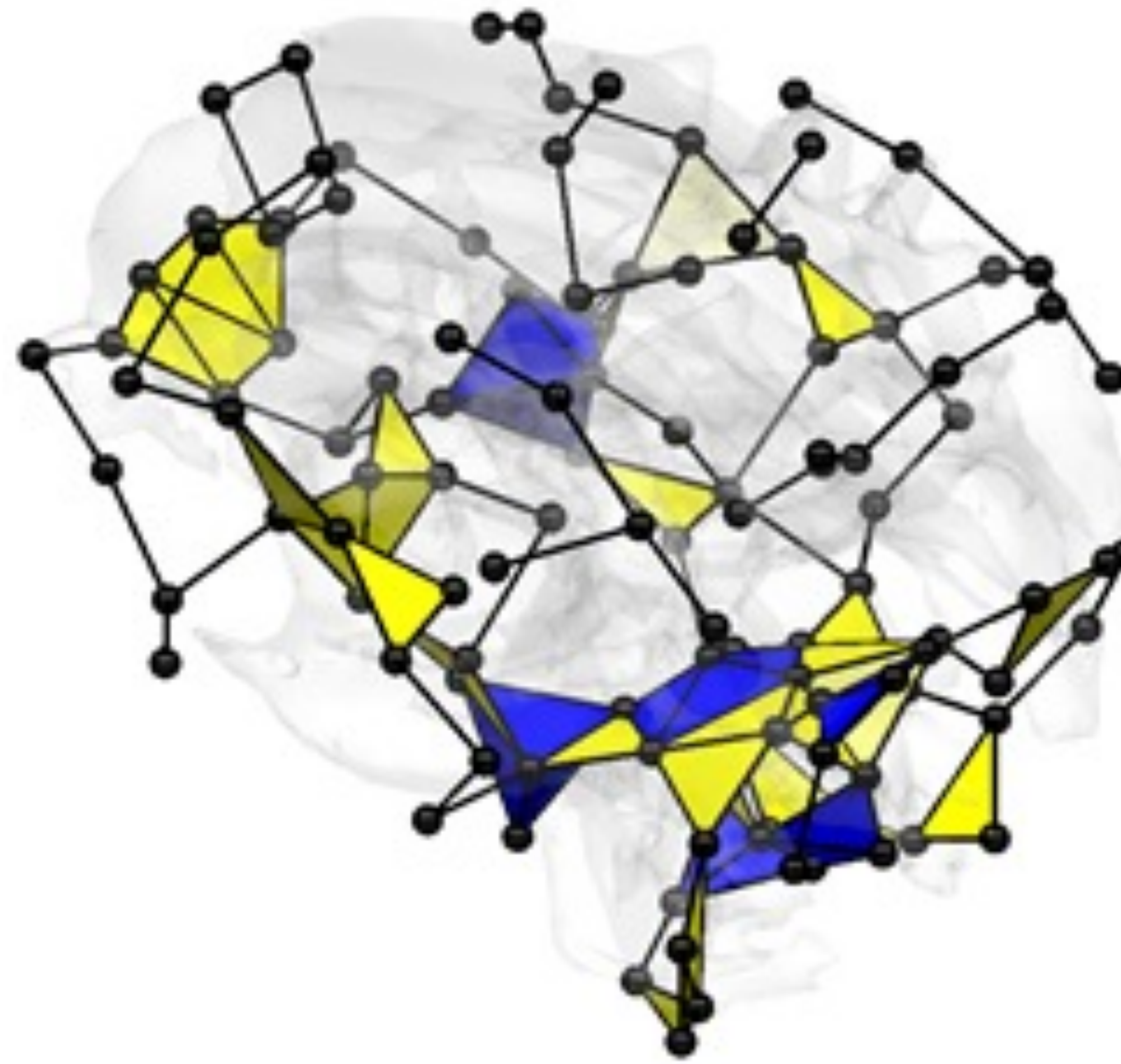
2-simplex

$\sigma = [v_1, v_2, v_3]$

Simplicial complex

A simplicial complex is a set composed of points, line segments, triangles, and their n-dimensional counterparts.



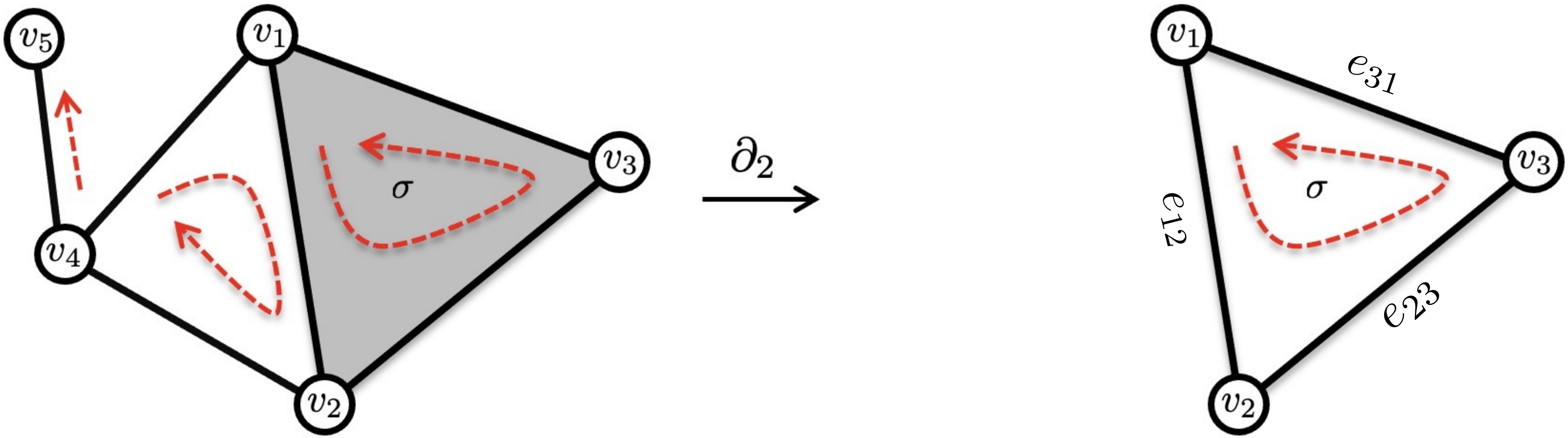


PH_rips.m

Boundary operators ∂_k

∂_k Removes the filled-in interior of *k*-simplexes

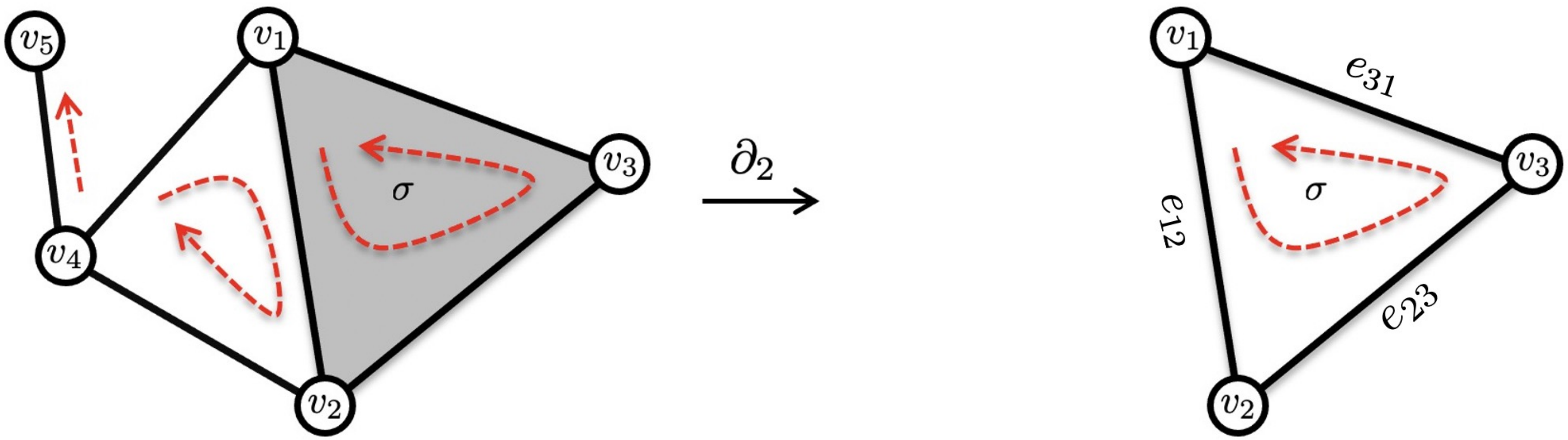
$$\partial_k : C_k \to C_{k-1}$$



$$\partial_2 \sigma = [v_1, v_2] + [v_2, v_3] + [v_3, v_1]$$

$$\partial_1 \partial_2 \sigma = v_2 - v_1 + v_3 - v_2 + v_1 - v_3 = 0$$

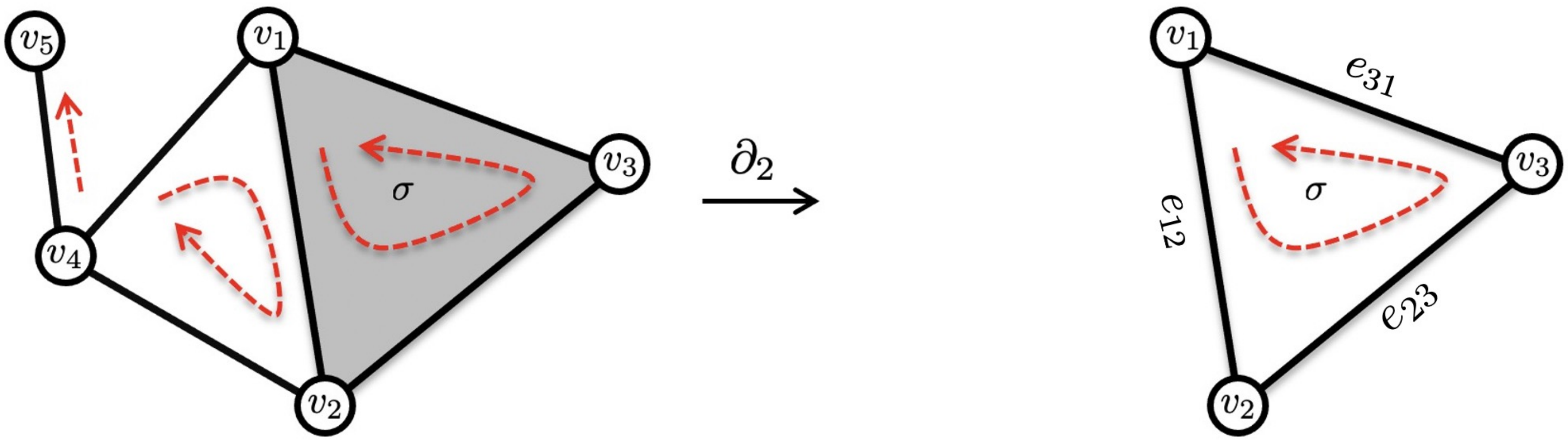
Boundary matrix ∂_0



$$\boldsymbol{\partial}_0 = \begin{array}{c} \\ 0 \end{array} \begin{array}{c} \begin{array}{ccccc} v_1 & v_2 & v_3 & v_4 & v_5 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

PH_boundary.m

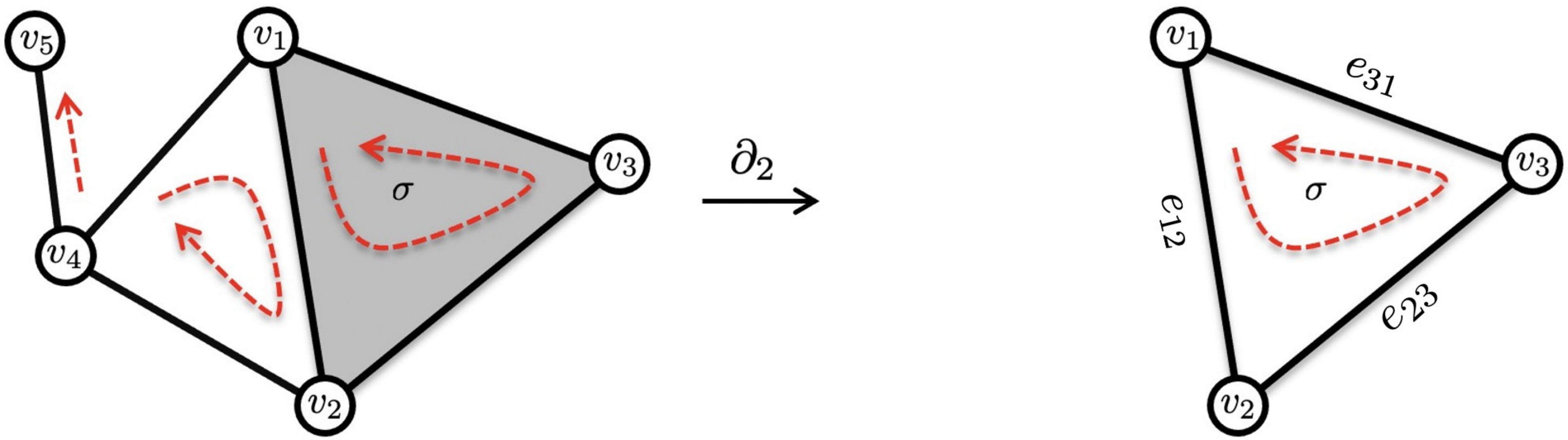
Boundary matrix ∂_1



$$\partial_1 \partial_2 \sigma = v_2 - v_1 + v_3 - v_2 + v_1 - v_3 = 0$$

$$\boldsymbol{\partial}_1 = \begin{array}{c|cccccc} & e_{12} & e_{23} & e_{31} & e_{24} & e_{41} & e_{45} \\ \hline v_1 & -1 & 0 & 1 & 0 & 1 & 0 \\ v_2 & 1 & -1 & 0 & -1 & 0 & 0 \\ v_3 & 0 & 1 & -1 & 0 & 0 & 0 \\ v_4 & 0 & 0 & 0 & 1 & -1 & -1 \\ v_5 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}$$

Boundary matrix ∂_2



$$\partial_2 \sigma = [v_1, v_2] + [v_2, v_3] + [v_3, v_1]$$

$$\boldsymbol{\partial}_2 = \begin{array}{c} \\ e_{12} \\ e_{23} \\ e_{31} \\ e_{24} \\ e_{41} \\ e_{45} \end{array} \begin{array}{c} \sigma \\ \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \end{array}$$

PH_boundary.m

Boundary matrix ∂_k

$$
\begin{array}{cc}
 & \text{# of } k\text{-dimensional simplices } \sigma_j \\
\text{# of } (k\text{-}1)\text{-dimensional simplices } \tau_i &
\left[\begin{array}{ccccccc}
1 & 0 & 1 & \dots & 1 & 0 & 1 \\
\vdots & & & & \vdots & & \\
1 & 1 & & & & 0 & 0 \\
0 & 1 & 0 & \dots & 0 & 1 & -1
\end{array}\right]
\end{array}
$$

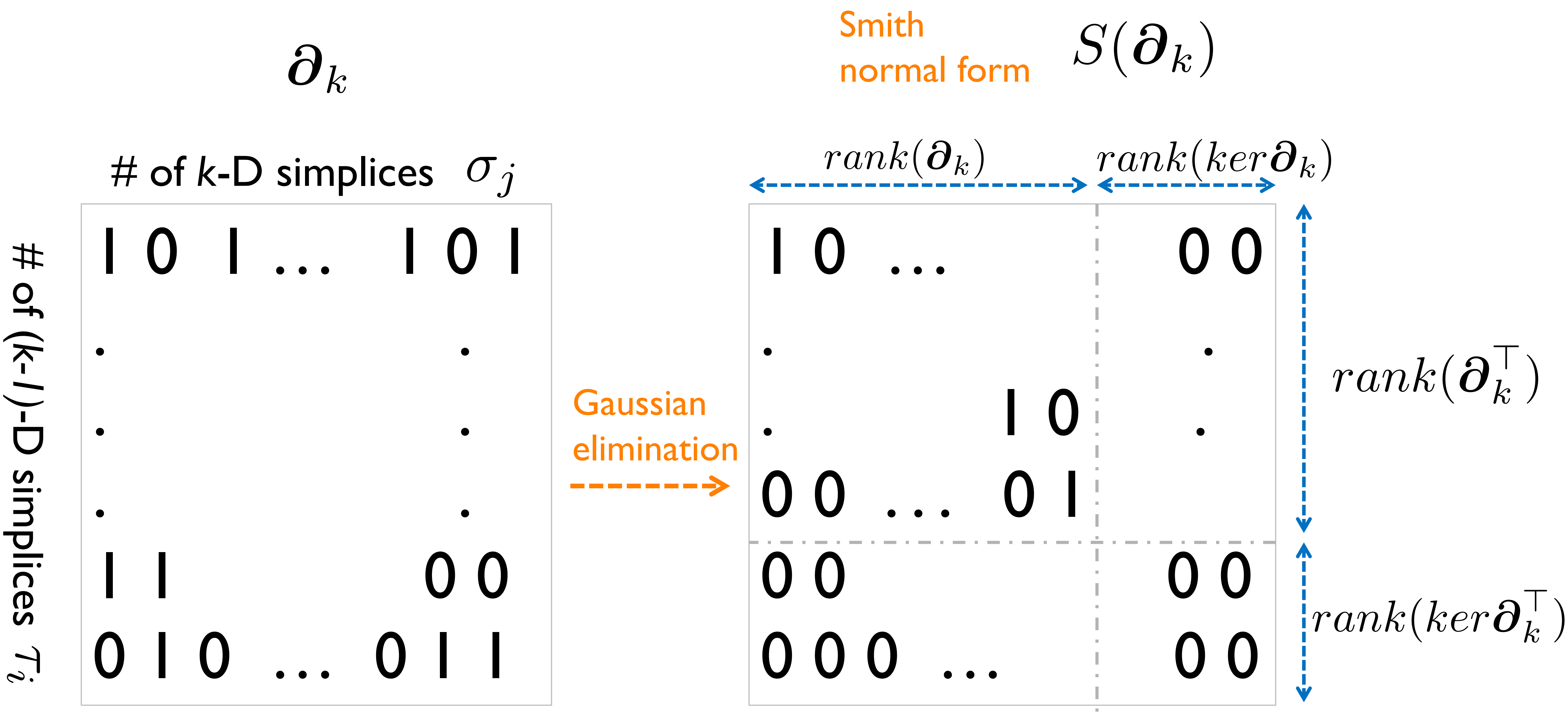
∂_k

(i,j)-th entry = 1 if $\tau_i \subset \sigma_j$

Sign depends on the orientation of τ_i

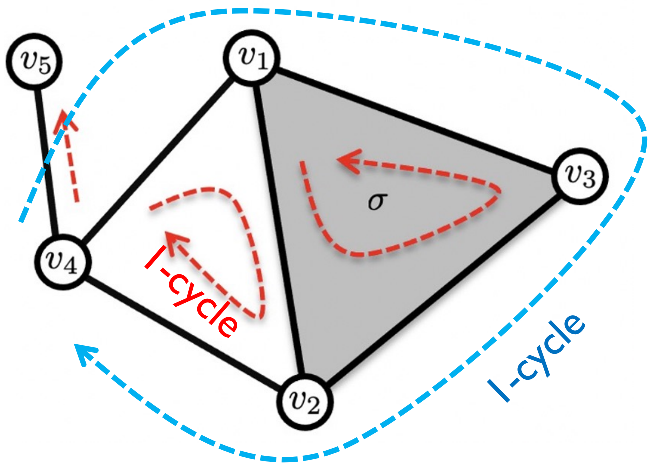
PH_boundary.m

Rank nullity theorem for boundary matrix



$$\beta_k = rank(ker\boldsymbol{\partial}_k) - rank(\boldsymbol{\partial}_{k+1})$$

Computing Betti numbers through boundary matrices



$$\mathcal{S}(\boldsymbol{\partial}_1) = \begin{pmatrix} 1&0&0&0&0&0 \\ 0&1&0&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&0&0&0 \end{pmatrix}, \quad \mathcal{S}(\boldsymbol{\partial}_2) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\beta_0 = rank(ker\boldsymbol{\partial}_0) - rank(\boldsymbol{\partial}_1)$$ = 5-4

$$\beta_1 = rank(ker\boldsymbol{\partial}_1) - rank(\boldsymbol{\partial}_2)$$ = 2-1

PH_boundary_betti.m

k-th Hodge Laplacian

PH_hodge.m

$$\Delta_k = \partial_{k+1}\partial_{k+1}^\top + \partial_k^\top \partial_k$$

0th Hodge Laplacian
Graph Laplacian

# of nodes

$$\Delta_0 = \partial_1 \partial_1^\top$$

# of nodes

1st Hodge Laplacian

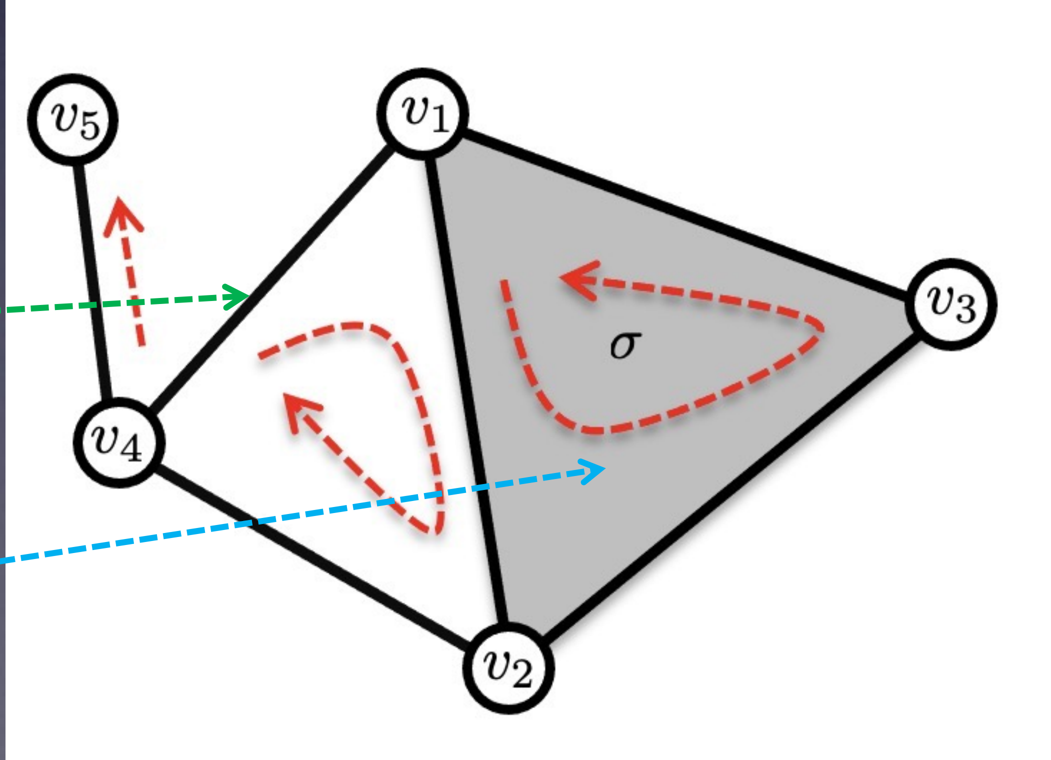
# of edges

# of edges

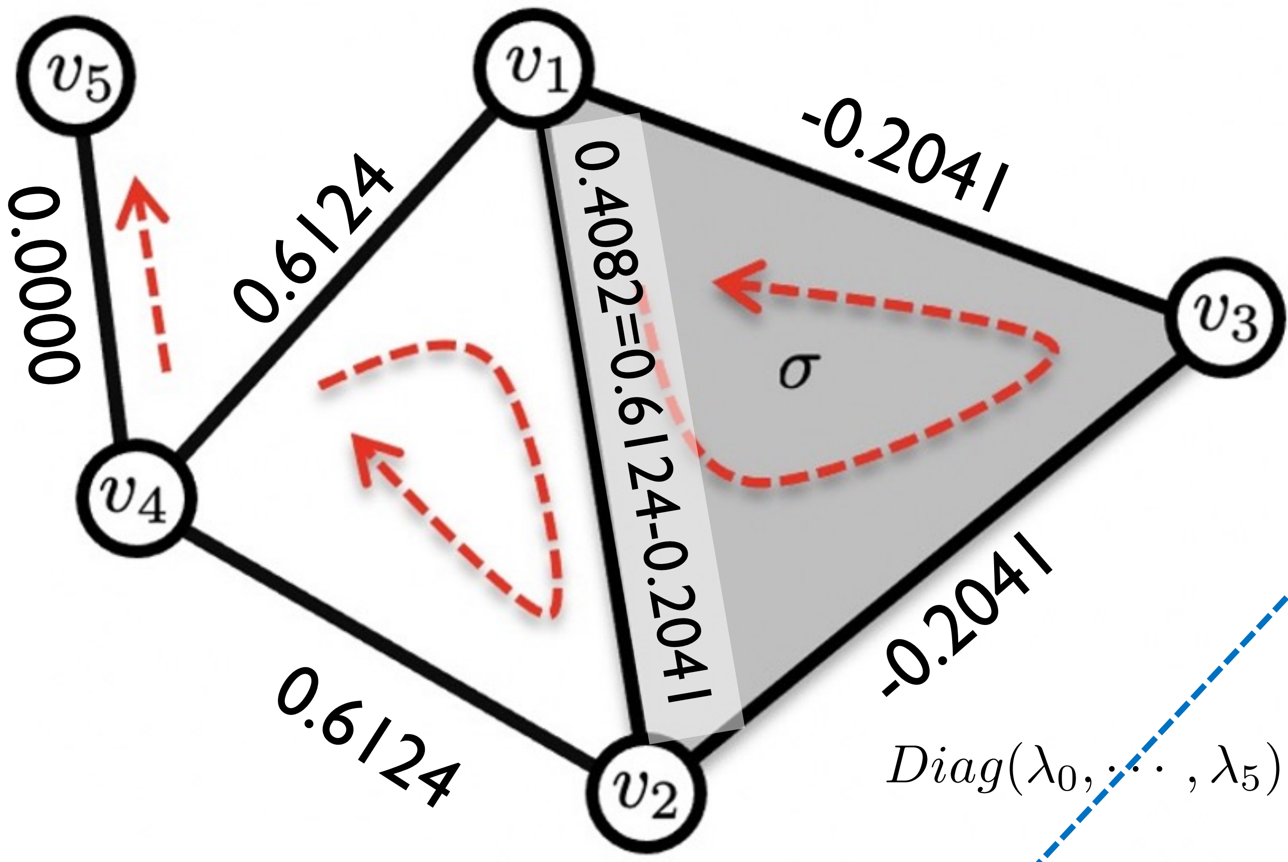
$$\Delta_1 = \partial_2 \partial_2^\top + \partial_1^\top \partial_1$$

# of edges

# of edges



Eigenvectors of Hodge Laplacian



$$\Delta_1 = \boldsymbol{\partial}_2 \boldsymbol{\partial}_2^\top + \boldsymbol{\partial}_1^\top \boldsymbol{\partial}_1$$

$$ker \Delta_1$$

span

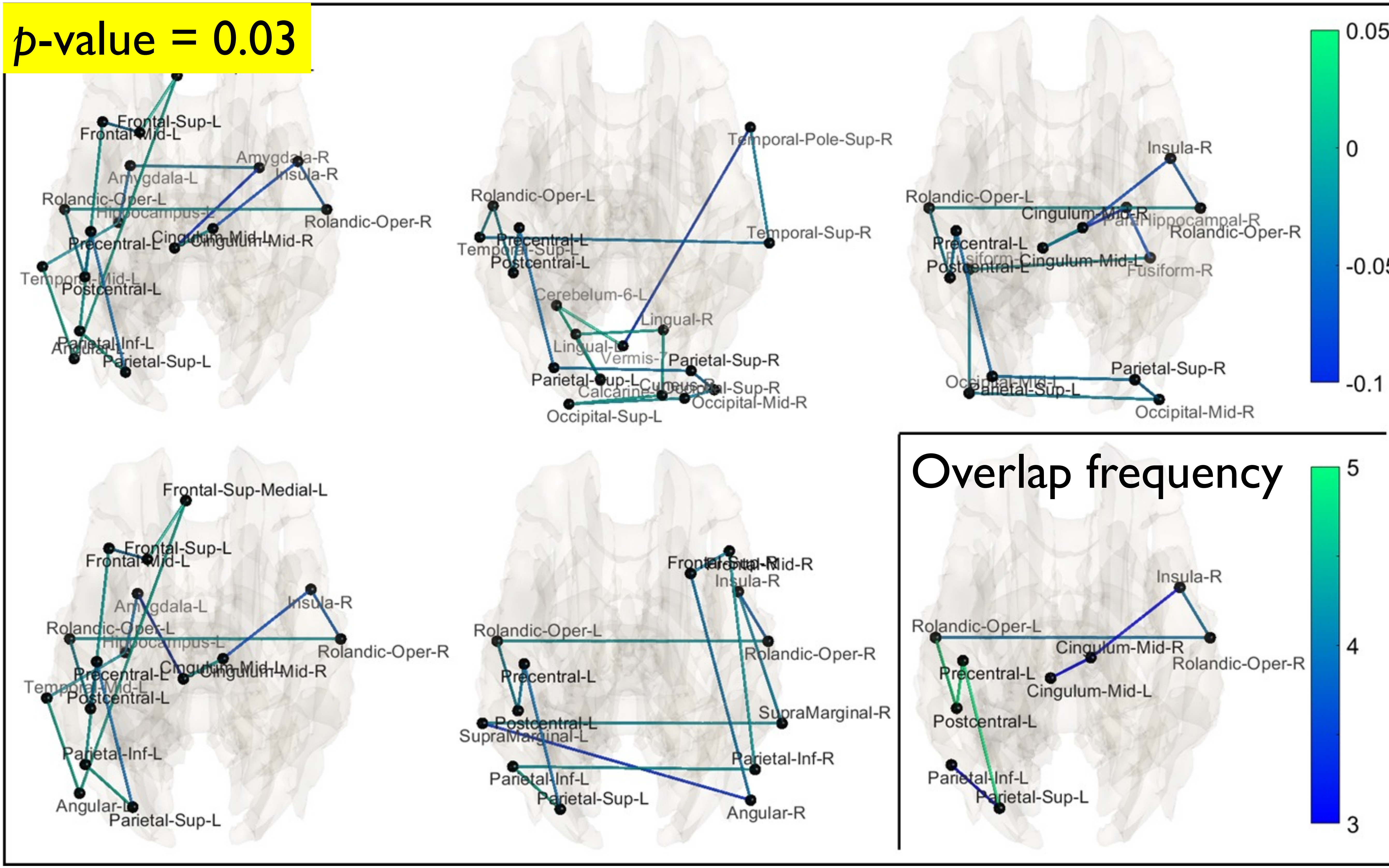
$$Diag(\lambda_0, \cdots, \lambda_5) = \begin{pmatrix} \boxed{0.0000} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.8299 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2.6889 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3.0000 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4.0000 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4.4812 \end{pmatrix}$$

$$(\mathbf{f}_0, \cdots, \mathbf{f}_5) = \begin{pmatrix} 0.4082 & -0.0000 & 0.0000 & 0.5774 & 0.7071 & -0.0000 \\ -0.2041 & 0.1993 & -0.5765 & 0.5774 & -0.3536 & 0.3578 \\ -0.2041 & -0.1993 & 0.5765 & 0.5774 & -0.3536 & -0.3578 \\ 0.6124 & -0.4325 & 0.1793 & -0.0000 & -0.3536 & 0.5299 \\ 0.6124 & 0.4325 & -0.1793 & -0.0000 & -0.3536 & -0.5299 \\ 0.0000 & -0.7392 & -0.5207 & 0.0000 & 0.0000 & -0.4271 \end{pmatrix}$$

Basis vector for cycles

Five biggest cycle differences (male – female) in HCP

p-value = 0.03



Anand et al. 2023, IEEE TMI in press arXiv:2110.14599

Rips filtrations

Rips filtration

PH_rips.m

Metric space

$$\mathcal{X} = (V, w) \quad w = (w_{ij})$$

Node set

Metric

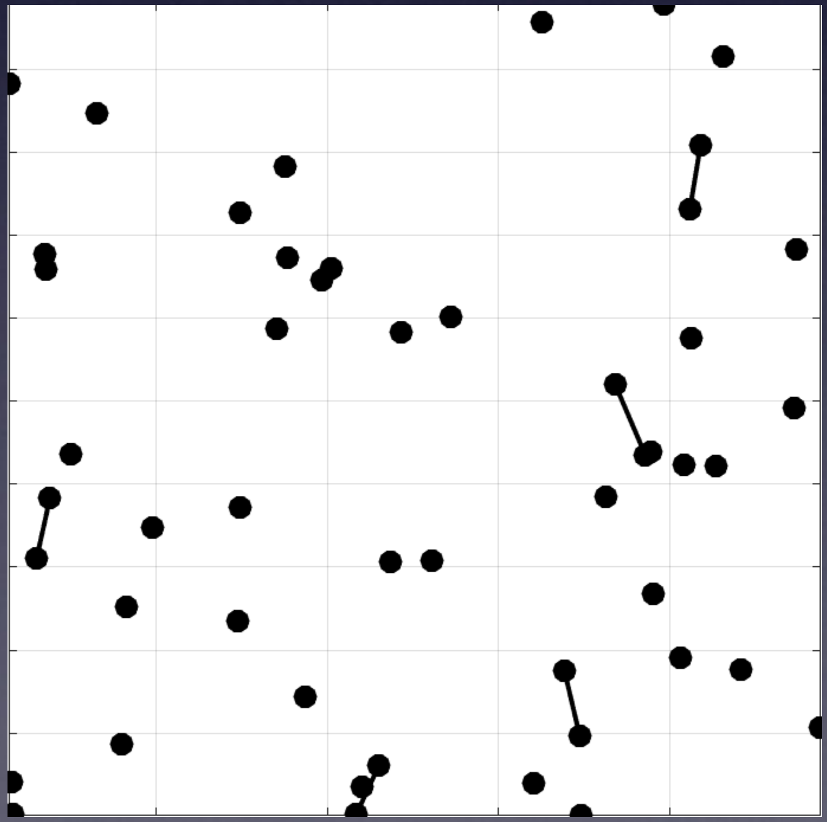
$$w_{ik} < w_{ij} + w_{jk}$$

Rips filtration

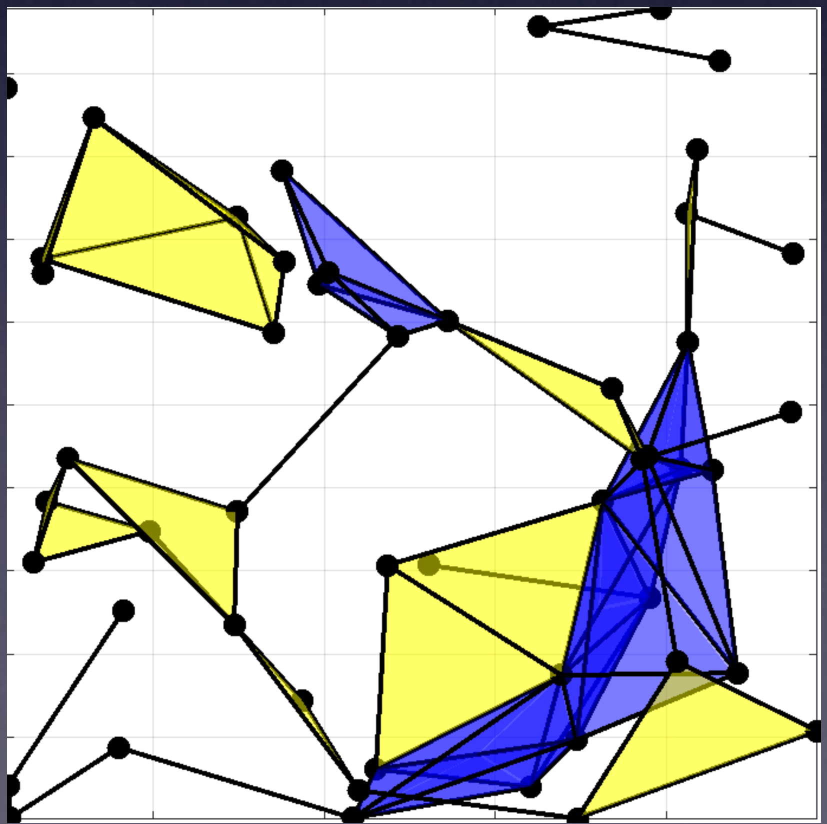
$$\mathcal{X}_{\epsilon_0} \subset \mathcal{X}_{\epsilon_1} \subset \mathcal{X}_{\epsilon_2} \subset \cdots$$

for filtration values

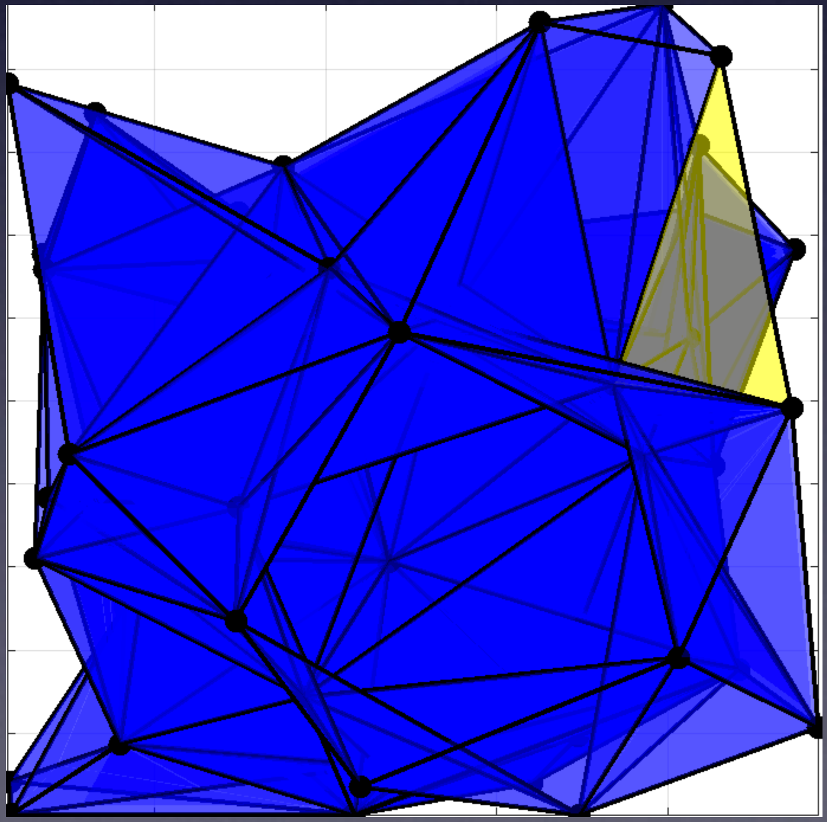
$$\epsilon_0 < \epsilon_1 < \epsilon_2 < \cdots$$



$\epsilon = 0.1$



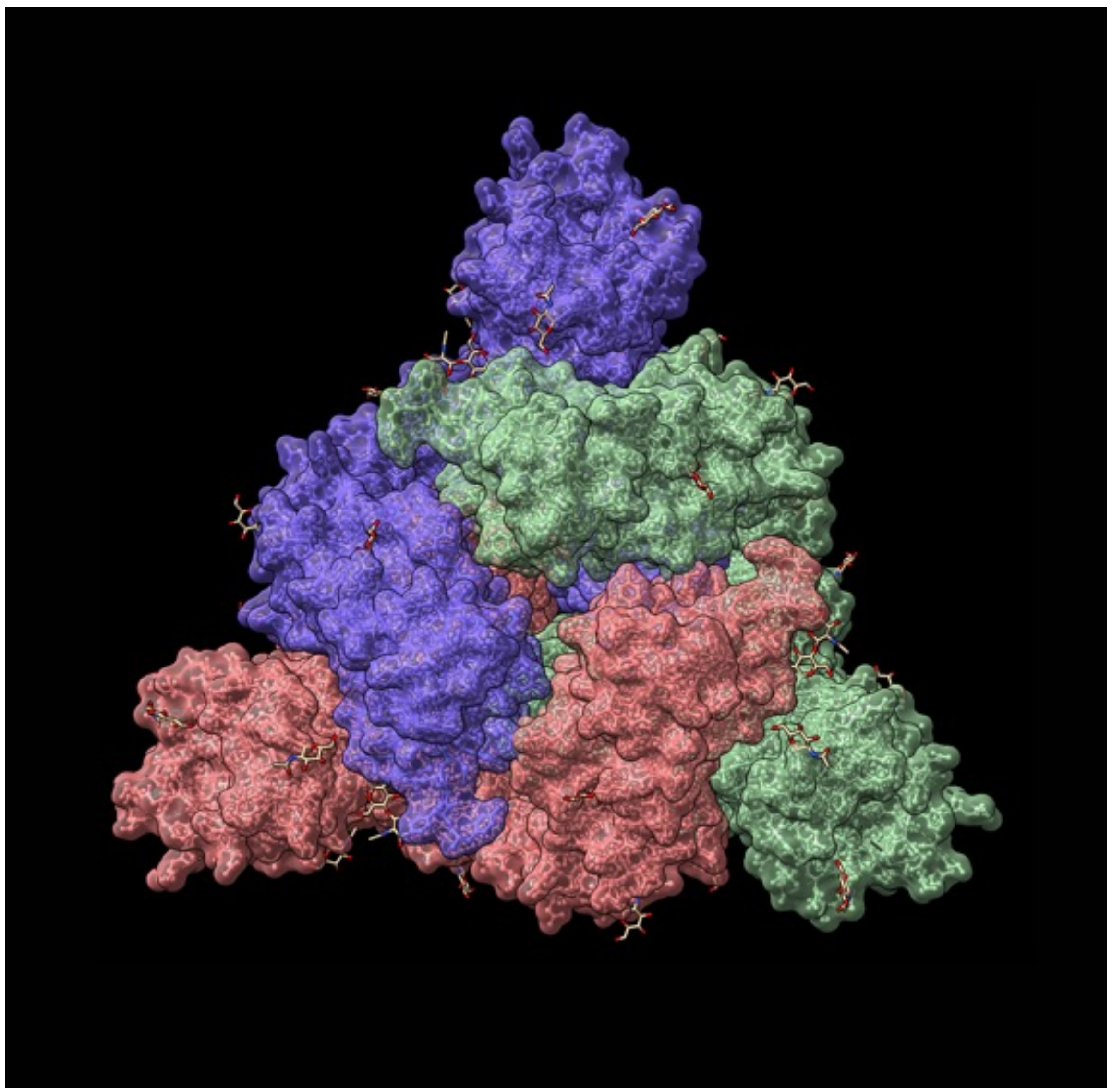
$\epsilon = 0.3$

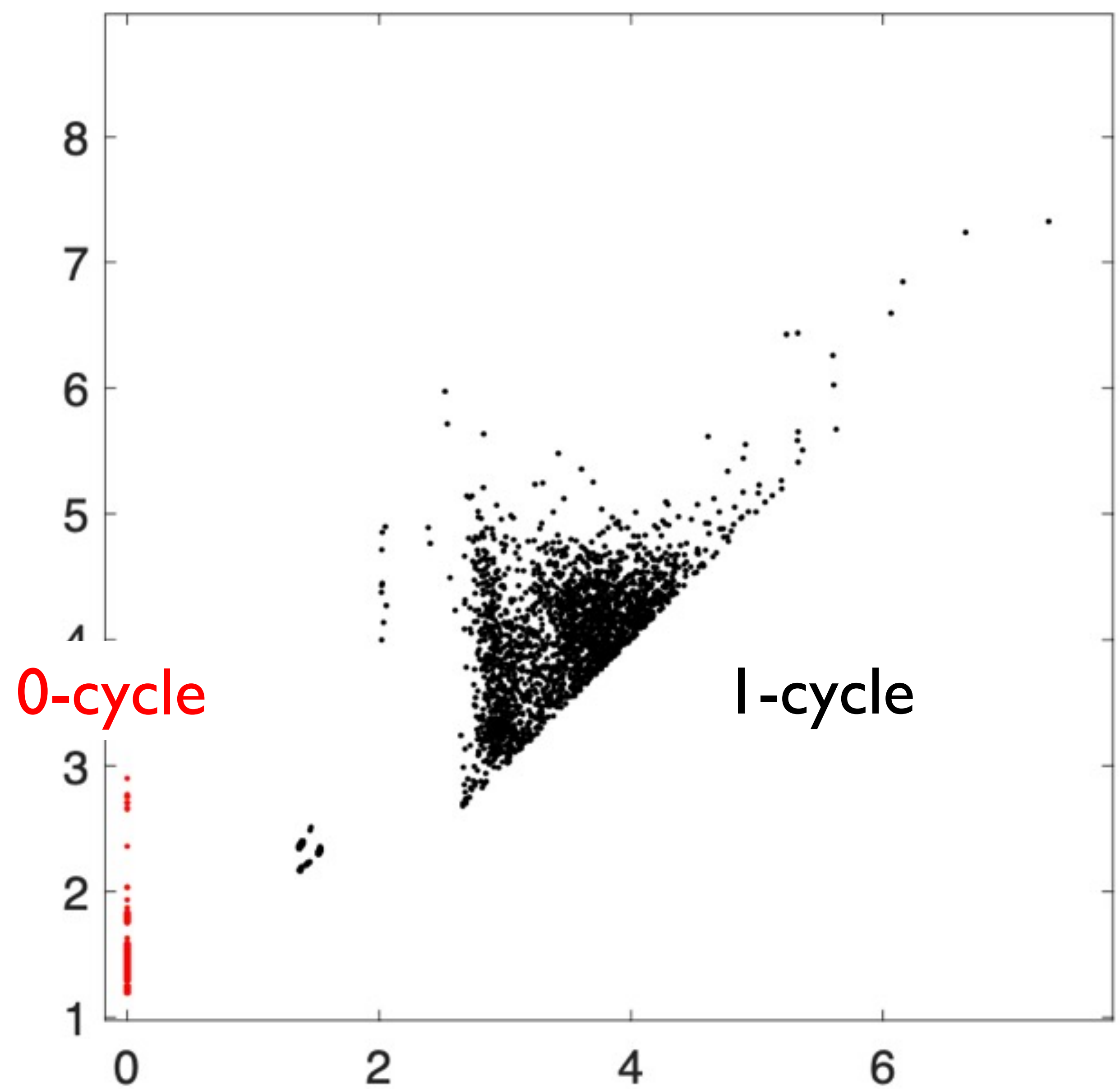


$\epsilon = 0.5$

Persistence Diagram (PD) of a protein molecule

Rips filtration on distance between 8000 atoms

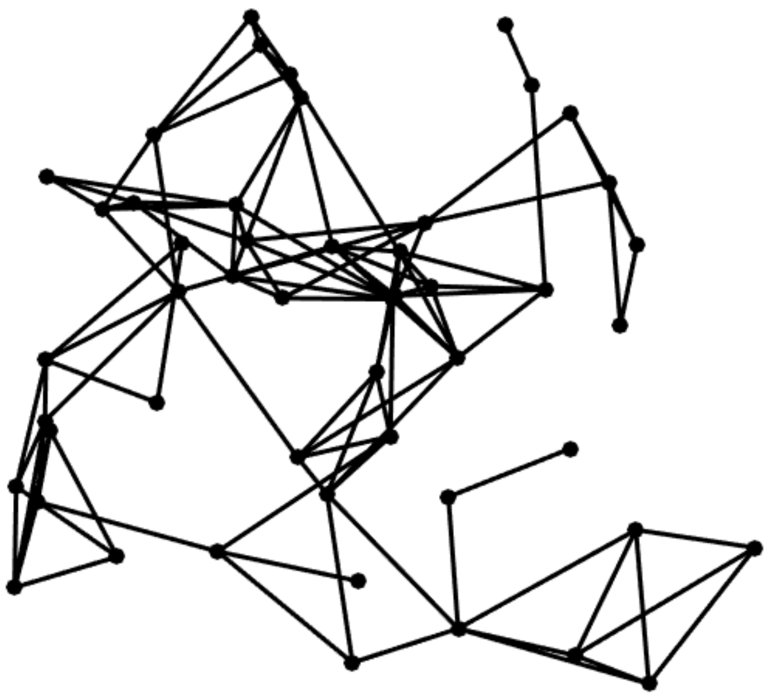




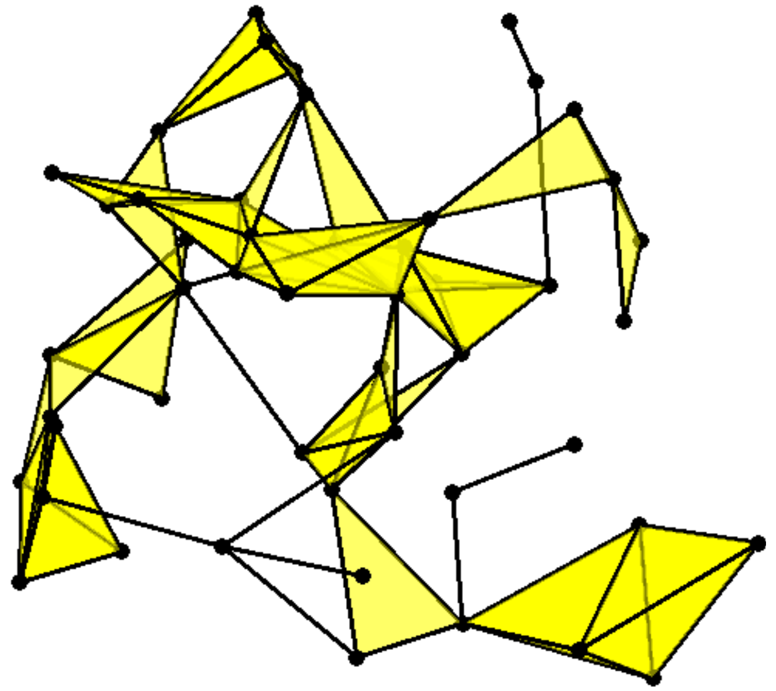
Extremely slow computation → Simply use graph filtration

k-skeleton

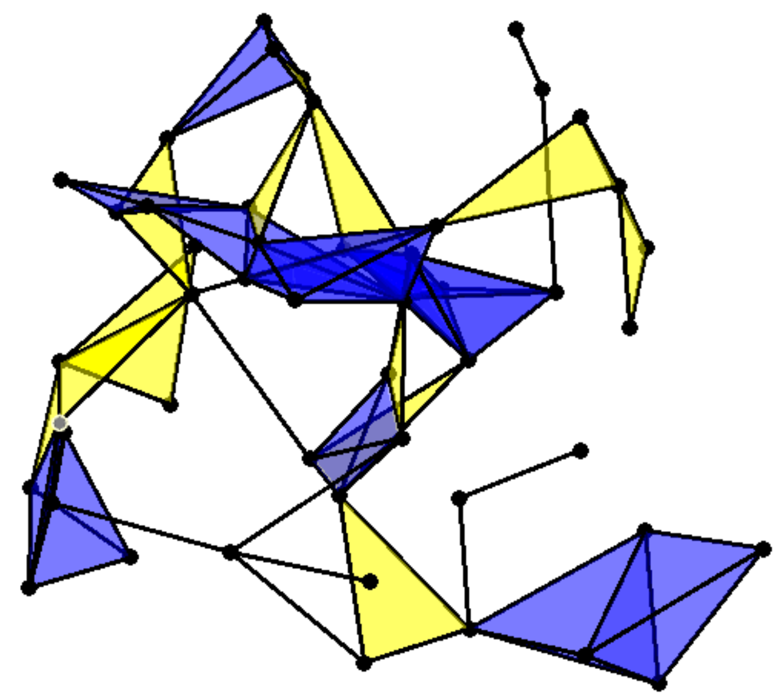
A simplicial complex consisting of up to k-simplices



1-skeleton



2-skeleton



3-skeleton

PH_rips.m

Graph filtrations

Baseline filtration for brain networks introduced in

Lee et al. 2011 ISBI
Lee et al. 2011 MICCAI 302-309
Lee et al. 2012 IEEE Transactions on Medical Imaging 31:2267-2277

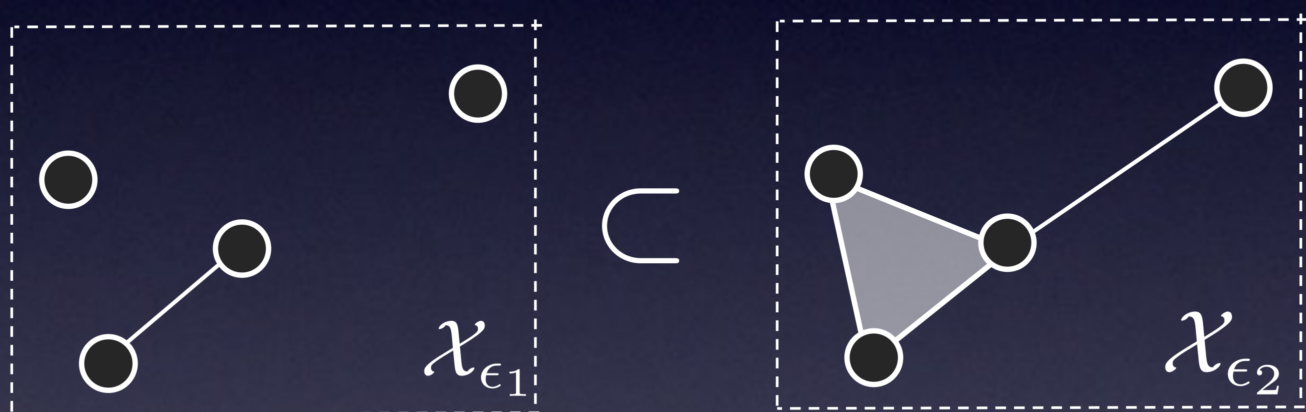
Rips filtration vs. graph filtration

Metric space

$$\mathcal{X} = (V, w) \quad w = (w_{ij})$$

Node set — Metric

Rips complex = Simplicial complex



Rips filtration

$$\mathcal{X}_{\epsilon_0} \subset \mathcal{X}_{\epsilon_1} \subset \mathcal{X}_{\epsilon_2} \subset \cdots$$

for increased radius

$$\epsilon_0 < \epsilon_1 < \epsilon_2 < \cdots$$

Weighted graph

$$\mathcal{X} = (V, w) \quad w = (w_{ij})$$

Node set — Edge weight

Binary graph: 1-skeleton

Graph filtration

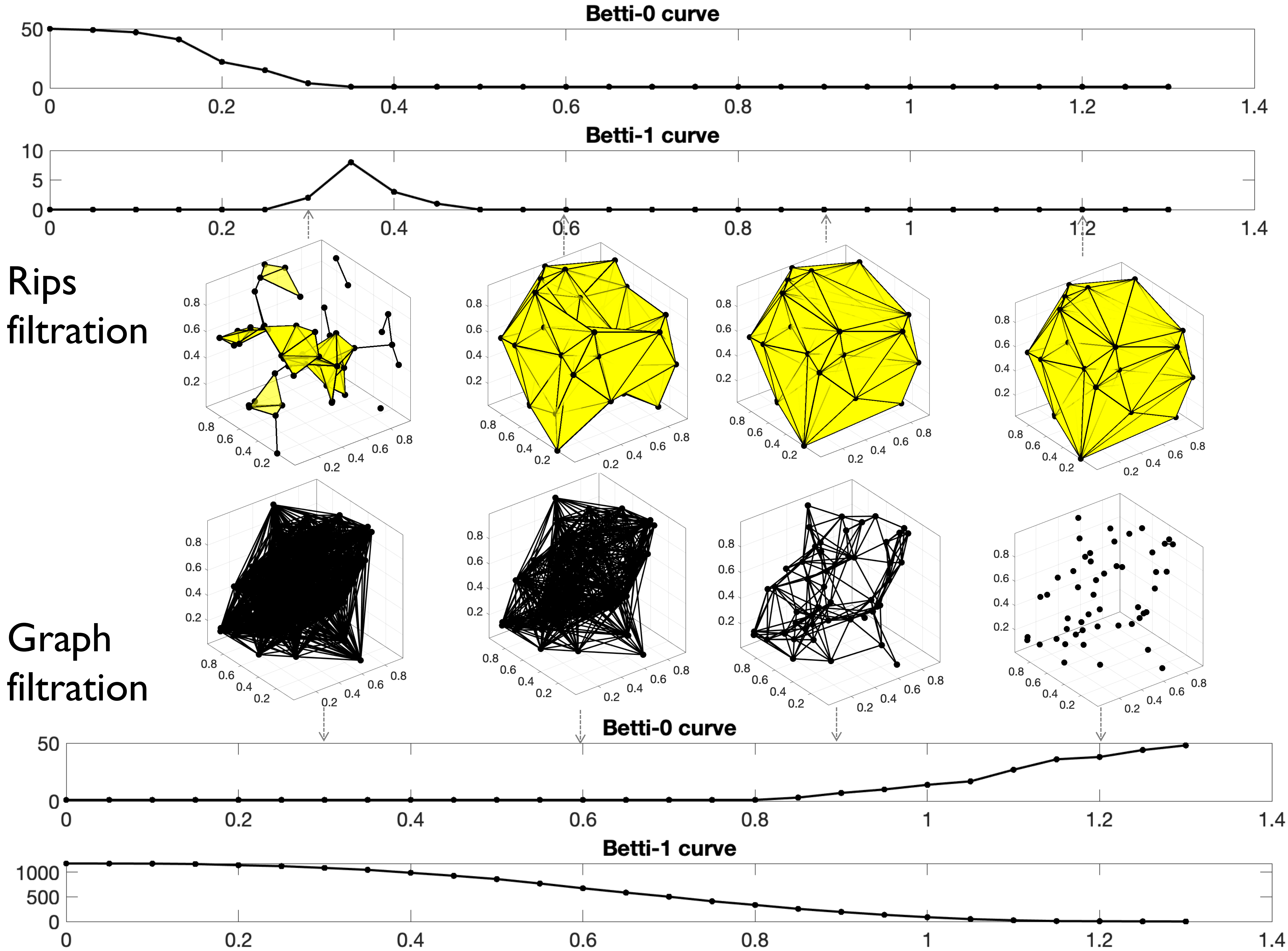
$$\mathcal{X}_{\epsilon_0} \supset \mathcal{X}_{\epsilon_1} \supset \mathcal{X}_{\epsilon_2} \supset \cdots$$

for increased edge weights

$$\epsilon_0 < \epsilon_1 < \epsilon_2 < \cdots$$

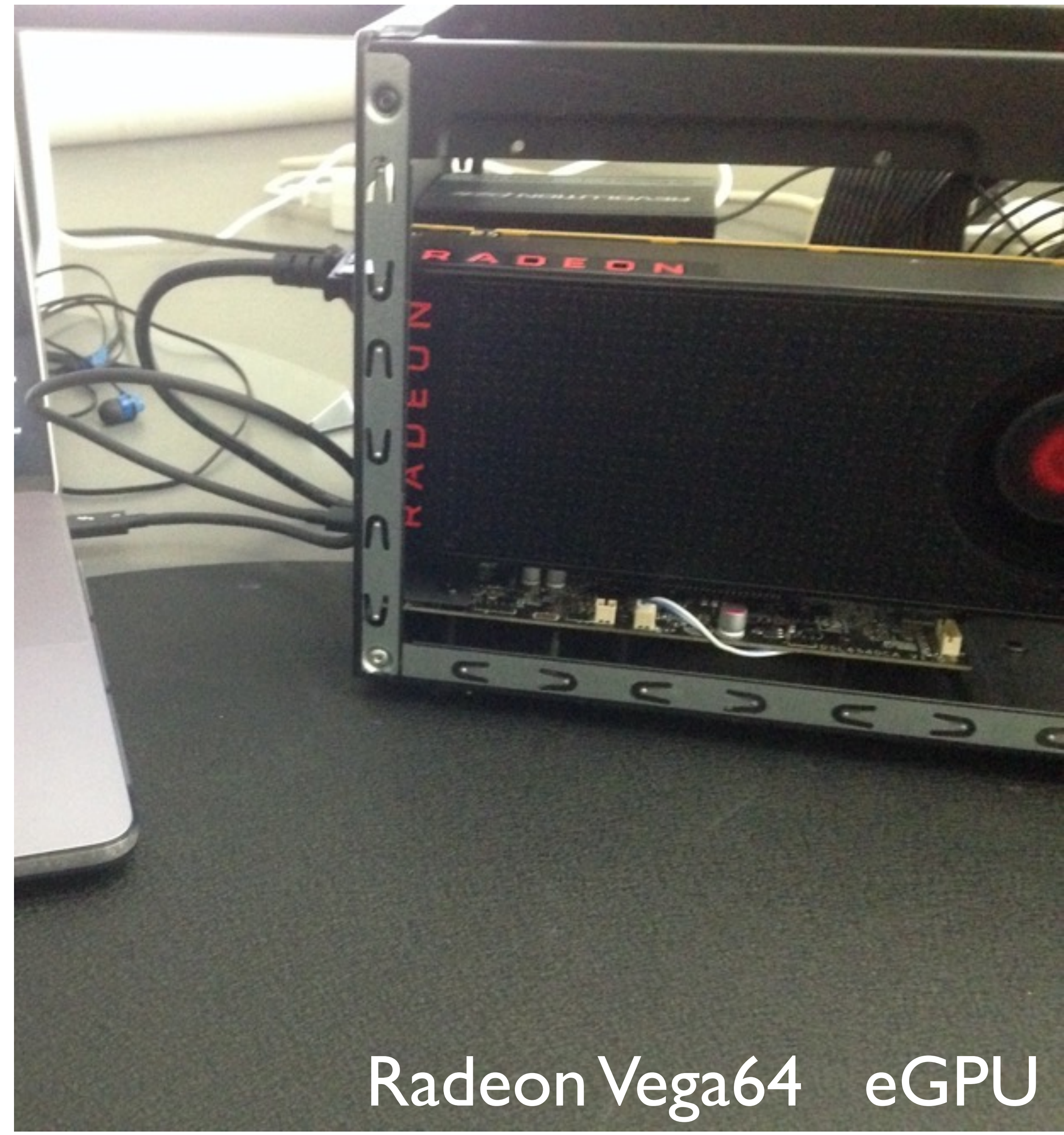
PH_rips.m

PH_graph.m



Betti-0 curve
Betti-1 curve
Rips filtration
Graph filtration
Betti-0 curve
Betti-1 curve

How to compute the number of cycles in big network data?



How many cycles in the network?

<u>*Chung et al. 2019 Network Neuroscience 3:674-*</u>

Fast computation of Betti curves

Computation of β_0: Can use a built-in function in MATLAB.

```
[beta_0, S] = graphconncomp(adj)
```

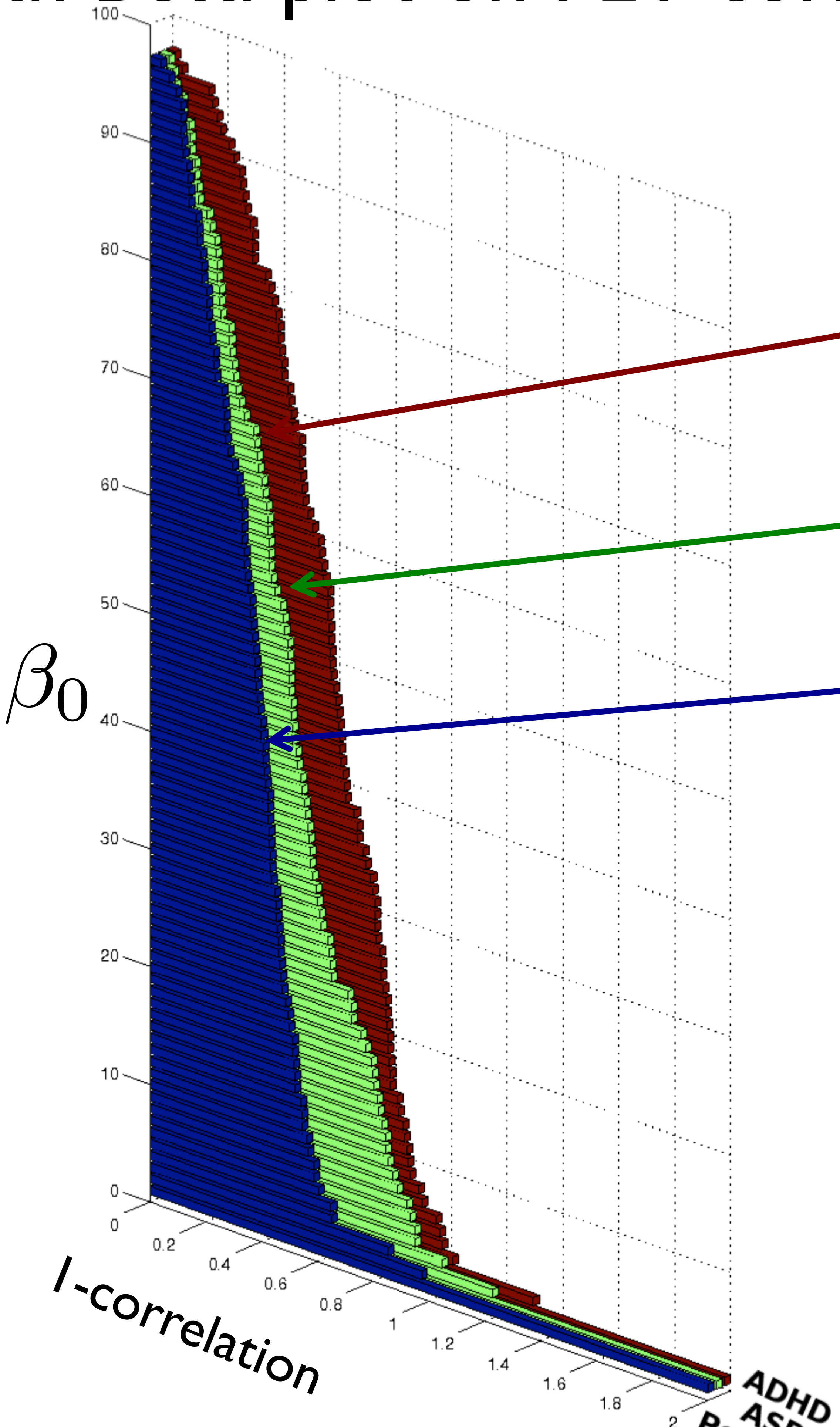
Computation of β_1: As a function of β_0

$$\beta_1 = \beta_0 - p + q$$

```
q=sum(sum(adj))/2;
beta_1 = beta_0 - p + q;
```

PH_betti.m

0-th Betti plot on PET correlation network



24 attention deficit hyperactivity disorder (ADHD) children

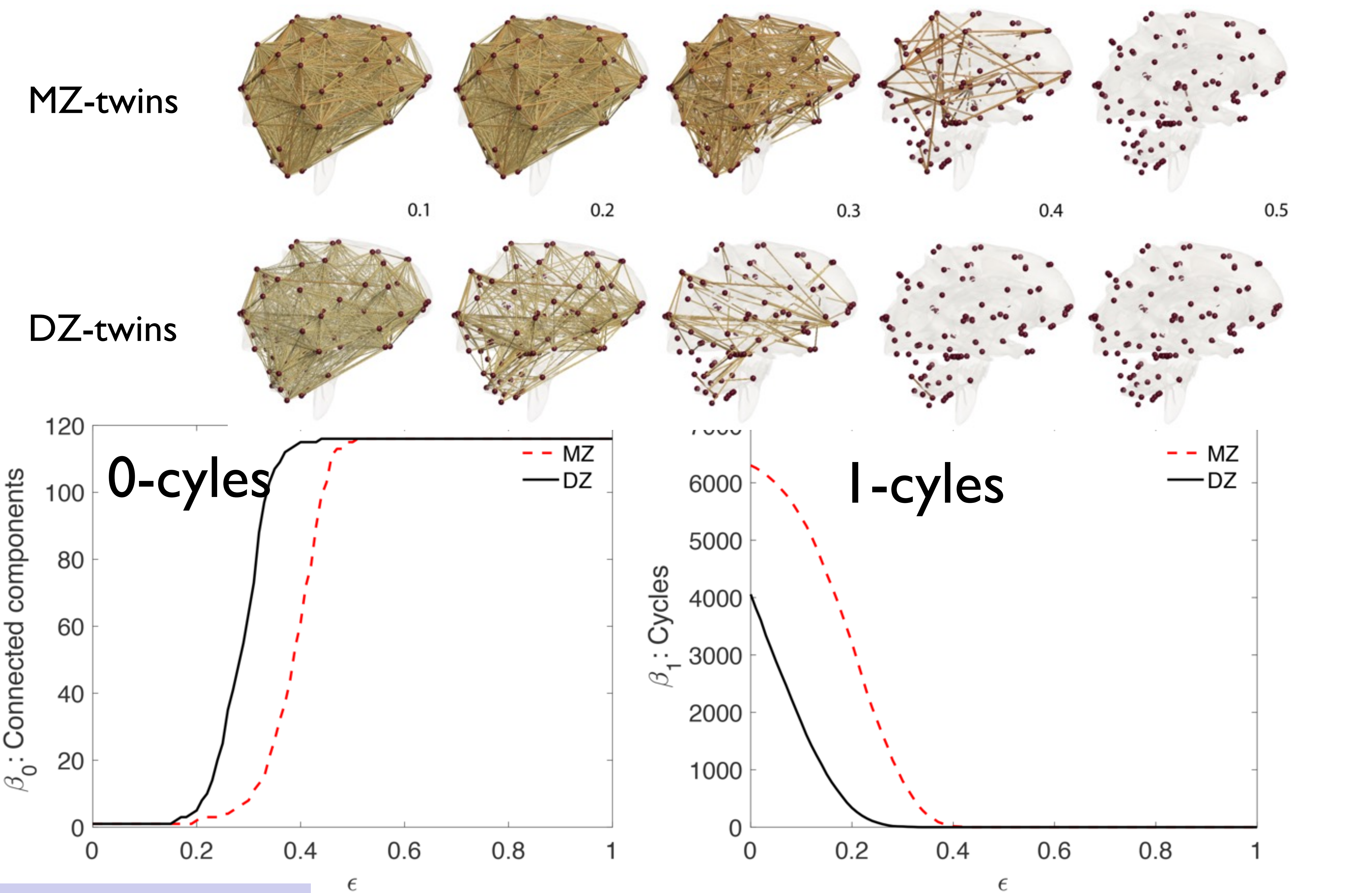
26 autism spectrum disorder (ASD) children

11 pediatric control subjects

PH_betti.m

Lee et al. (2011) ISBI

Genetic effect on Betti curves of rs-fMRI network



PH_betti.m

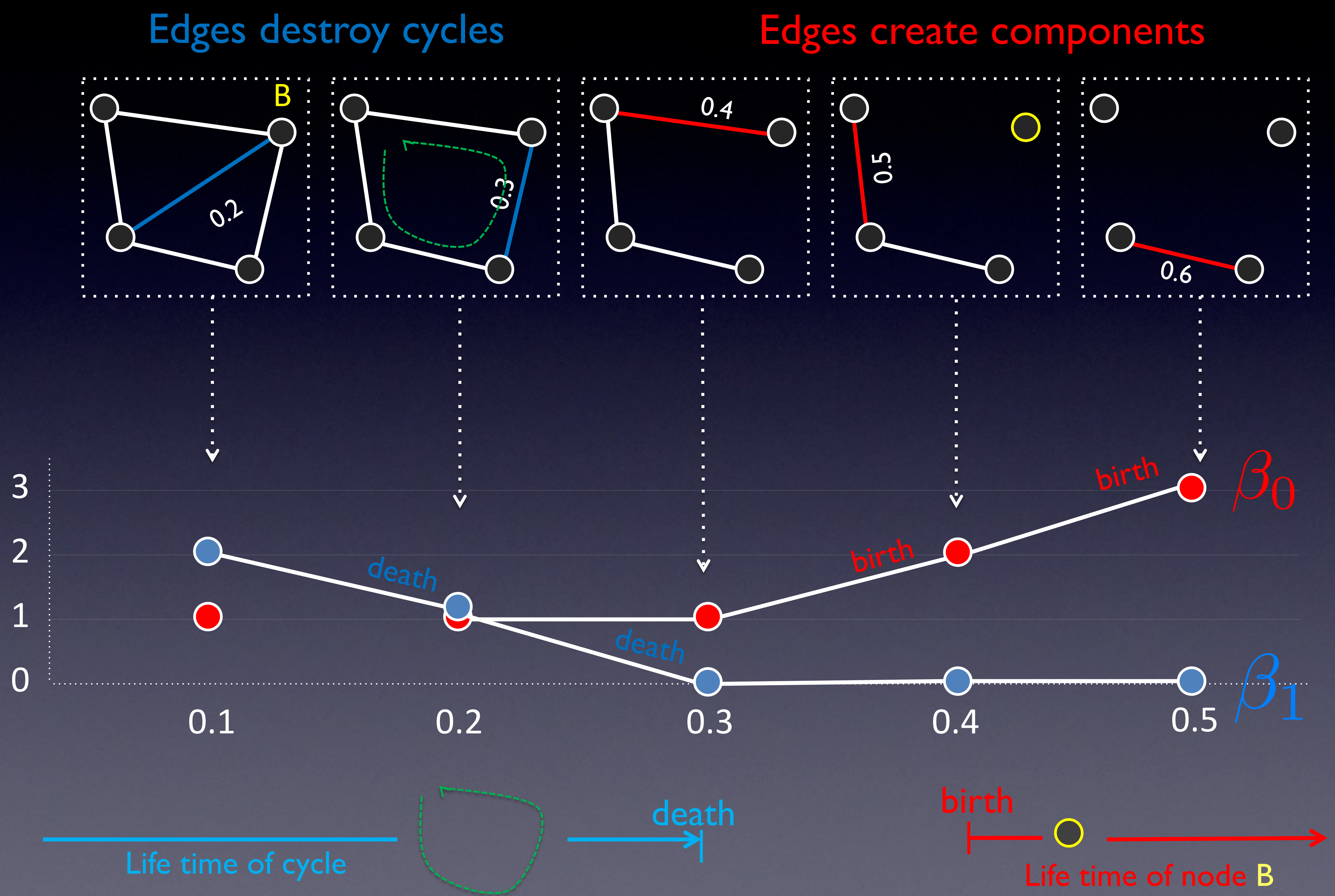
Chung et al. 2019 Network Neuroscience 3:674-694

Birth and death decomposition

Songdechakraiwut et al. 2021 MICCAI 166-176

Songdechakraiwut and Chung. 2023, Annals of Appled Statistics

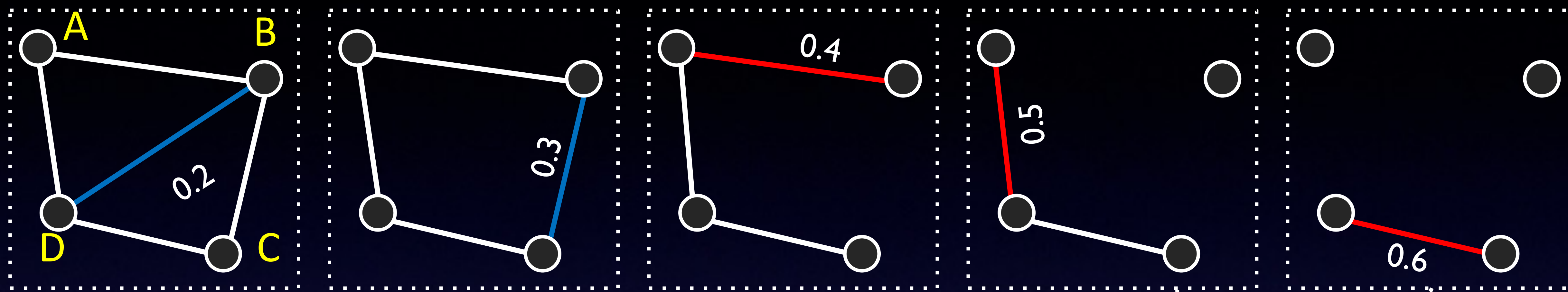
Persistence = Life time (death – birth) of a feature



<u>*Theorem*</u> *Birth & death sets partition the edge set*

E_1 Edges destroy cycles

E_0 Edges create components

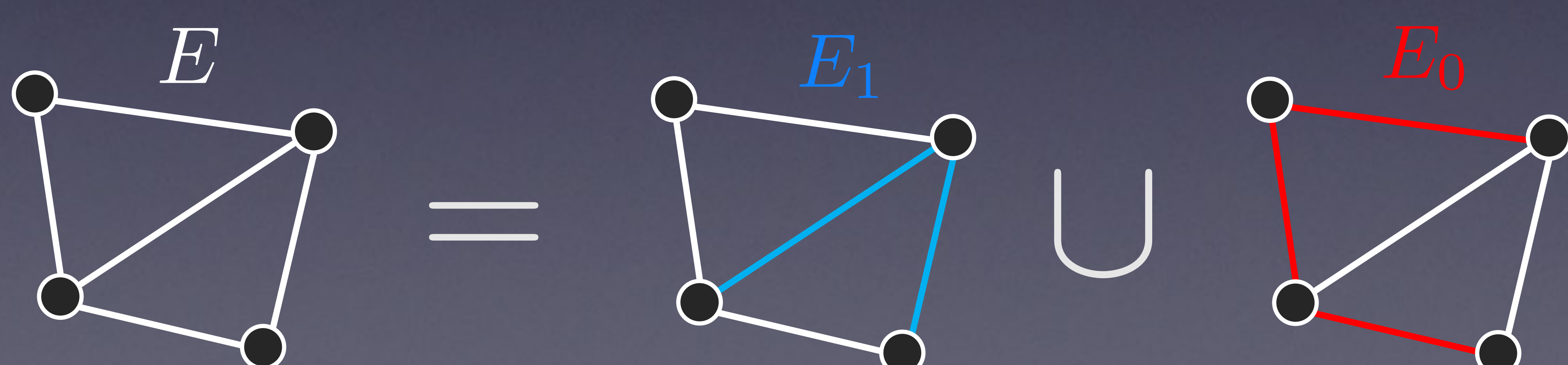


$$\#(E_1) = 1 + \frac{|V|(|V|-3)}{2}$$

$$\#(E_0) = |V| - 1$$

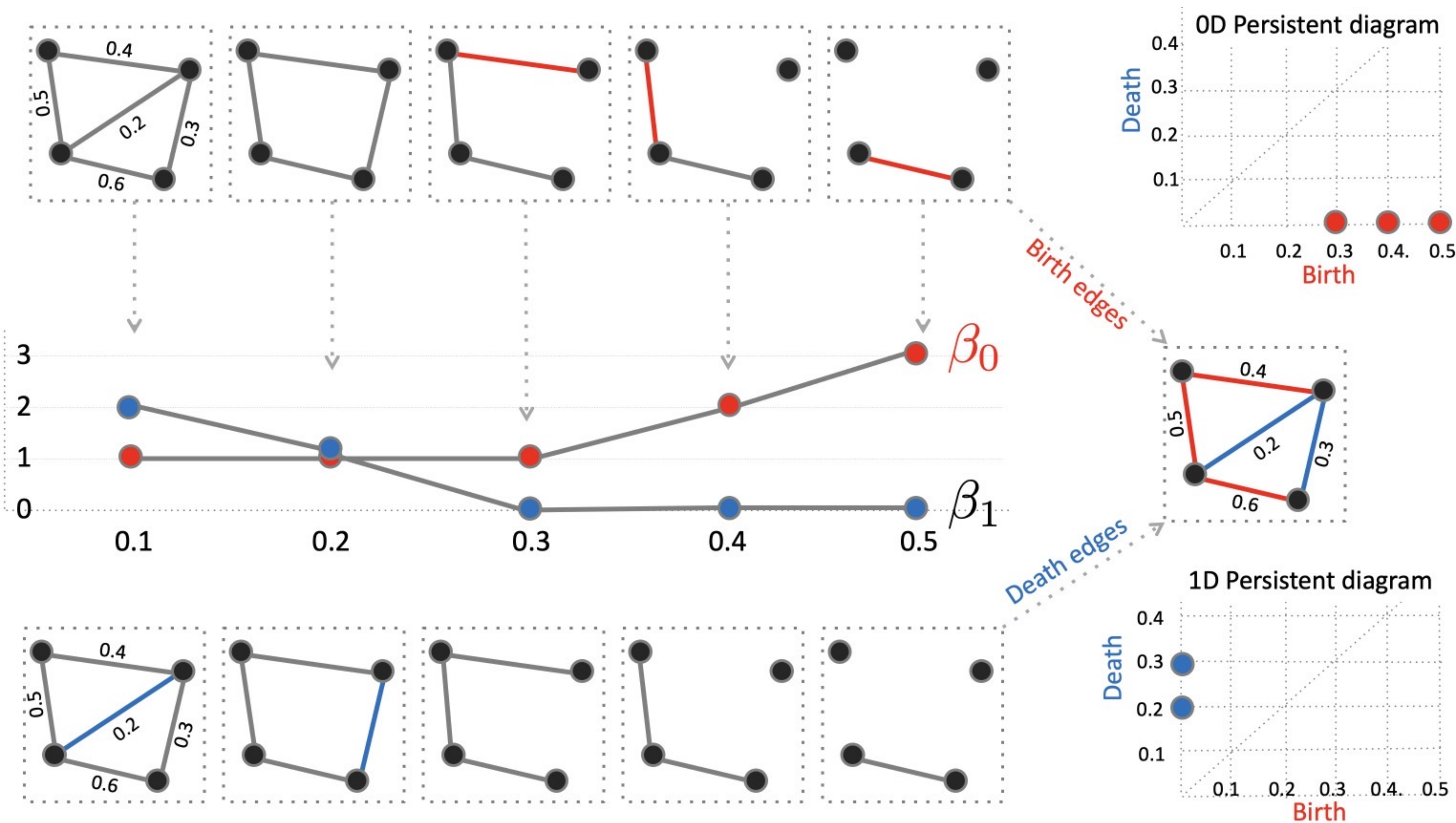
$$\#(E) = \frac{|V|(|V|-1)}{2}$$

Maximum spanning tree



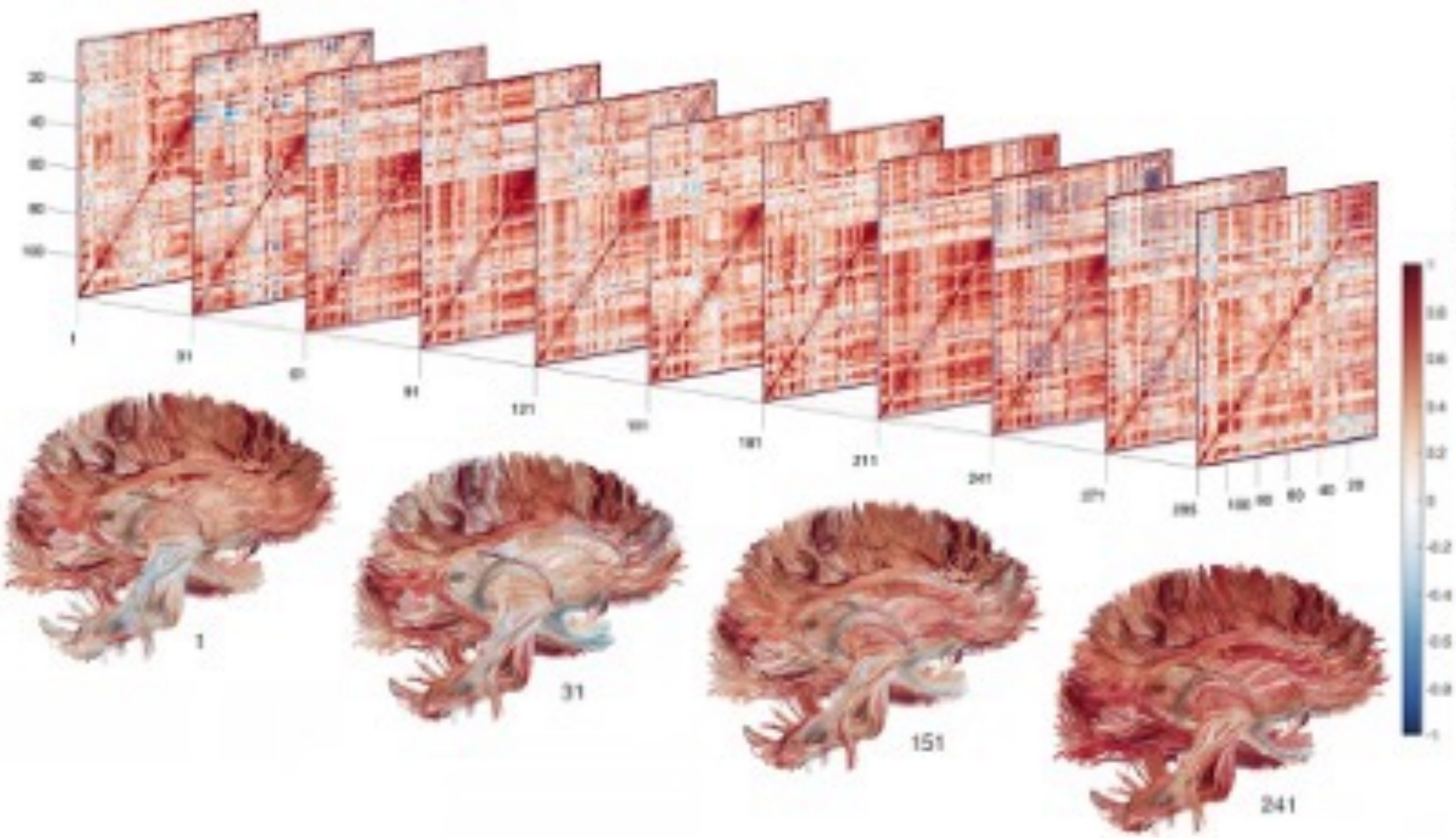
$O(|E| \log |V|)$

Songdechakraiwut & Chung 2020 <u>arXiv: 2012.00675</u>

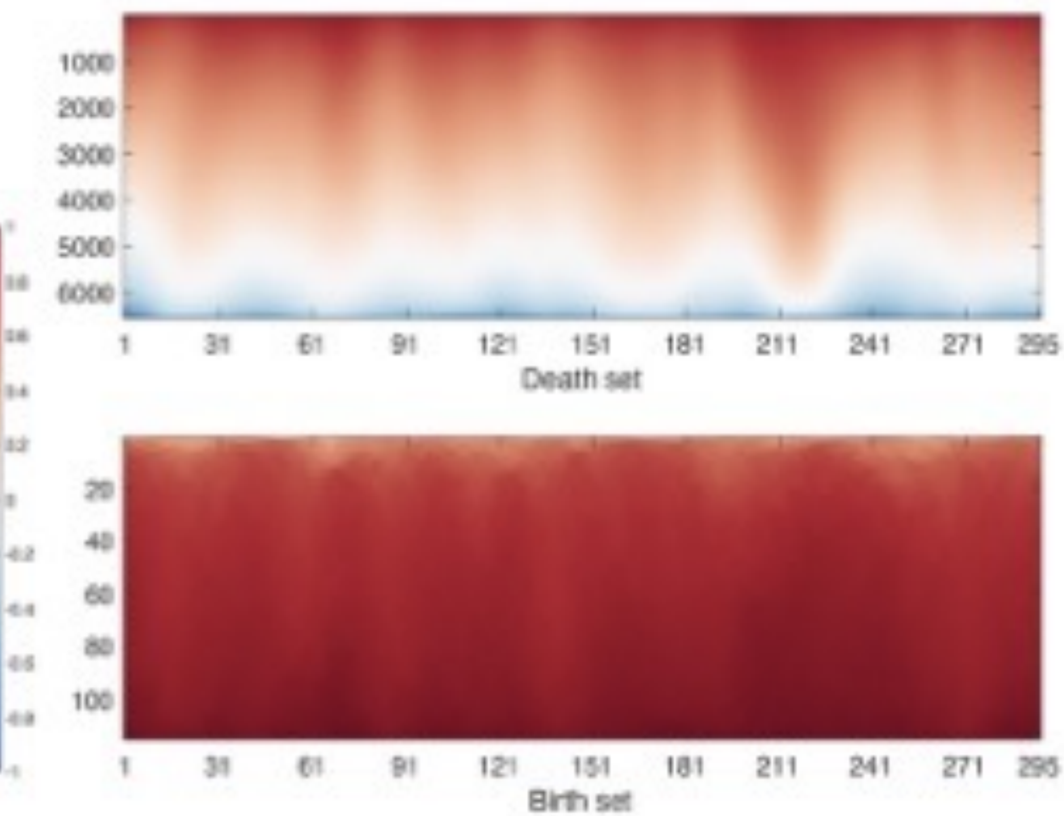


WS_decompose.m

Dynamically changing correlation network from rs-fMRI



Dynamically changing birth-death sets



WS_decompose.m

Topological inference

Songdechakraiwut and Chung. 2023, Annals of Appled Statistics

Chung et al. 2023 under review in NeuroImage arXiv:2302.06673

2-Wasserstein distance between persistent diagrams

Random variables: $X \sim f_1 \quad Y \sim f_2$

2-Wasserstein distance: $\mathcal{D}(X,Y) = \left(\inf \mathbb{E}\|X - Y\|^2\right)^{1/2}$

Persistent diagrams

$$P_1 = \{x_1, \cdots, x_q\} \subset \mathbb{R}^2 \qquad P_2 = \{y_1, \cdots, y_q\} \in \mathbb{R}^2$$

Empirical distributions

$$f_1(x) = \frac{1}{q}\sum_{i=1}^{q}\delta(x - x_i) \qquad f_2(y) = \frac{1}{q}\sum_{i=1}^{q}\delta(y - y_i)$$

$$\mathcal{D}(P_1, P_2) = \inf_{\psi: P_1 \to P_2} \Big(\sum_{x \in P_1} \|x - \psi(x)\|^2\Big)^{1/2}$$

Assignment problem: Hungarian algorithm

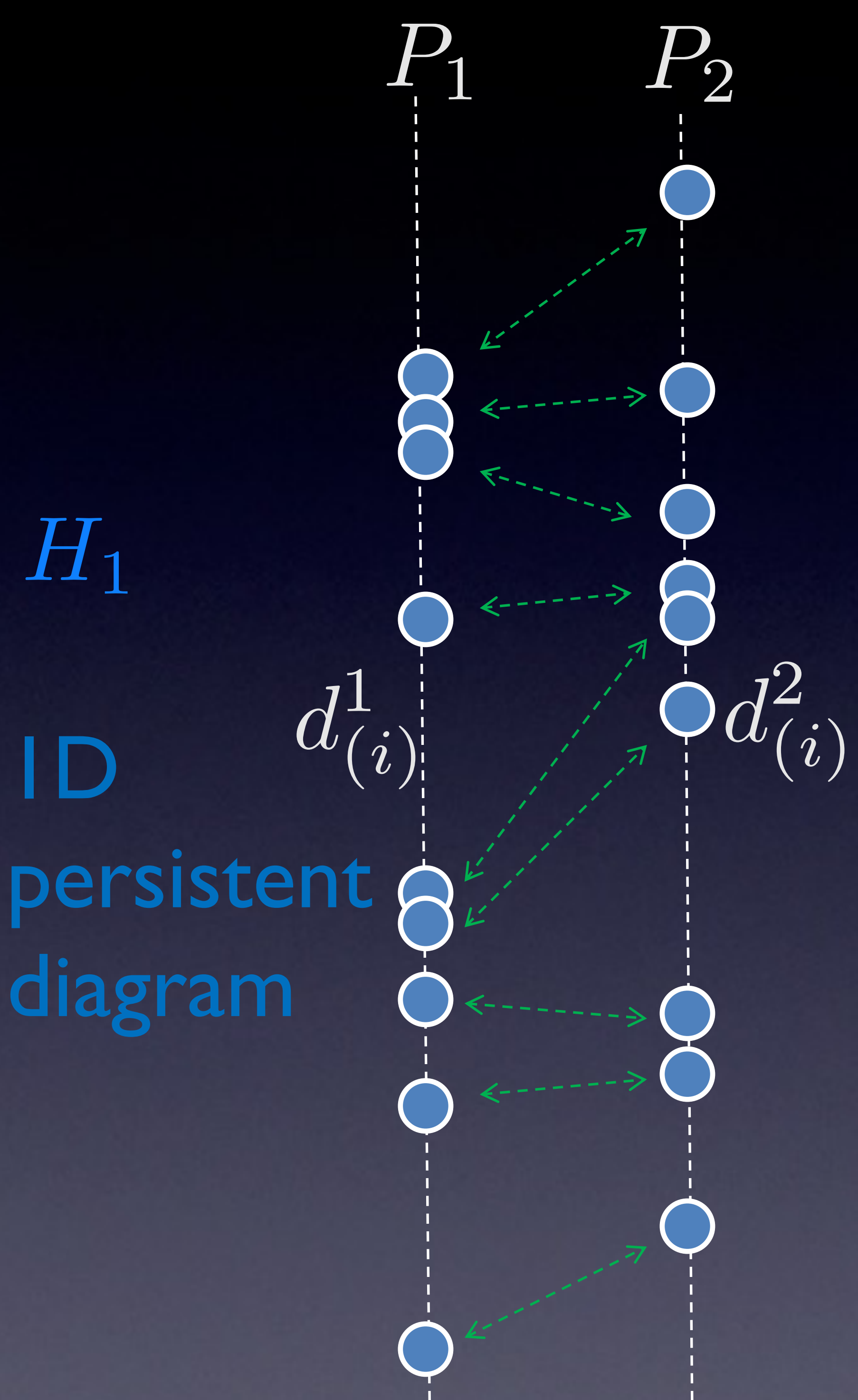
$\mathcal{O}(q^3)$

Graph filtration

$\mathcal{O}(q \log q)$

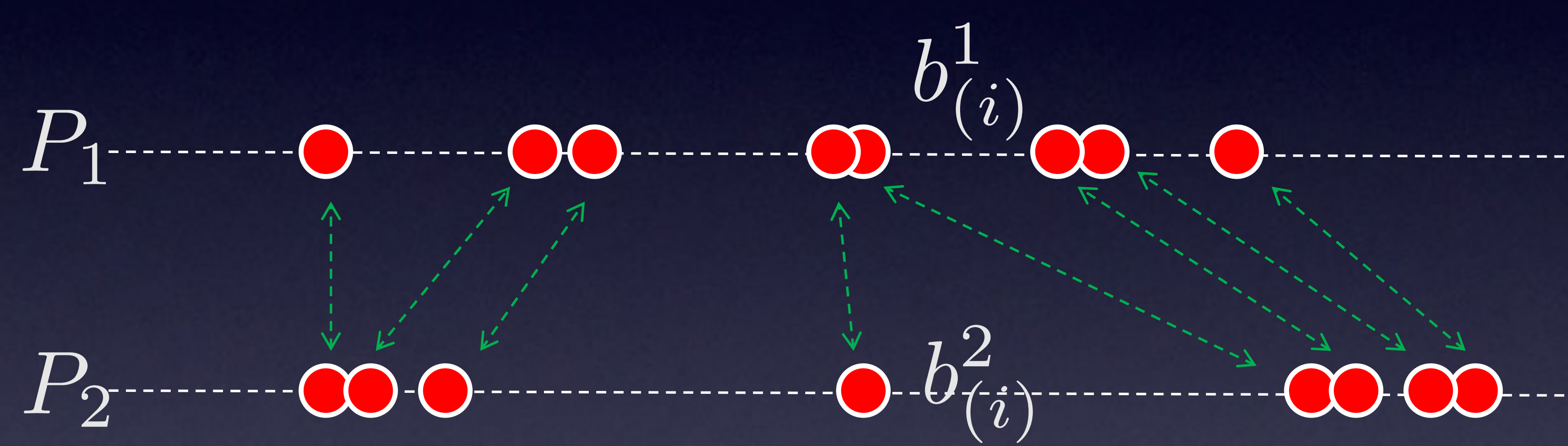
Wasserstein distance for graph filtrations

WS_pdist2.m



$$\mathcal{D}_1(P_1, P_2)^2 = \sum_i (d^1_{(i)} - d^2_{(i)})^2$$

H_0 0D persistent diagram

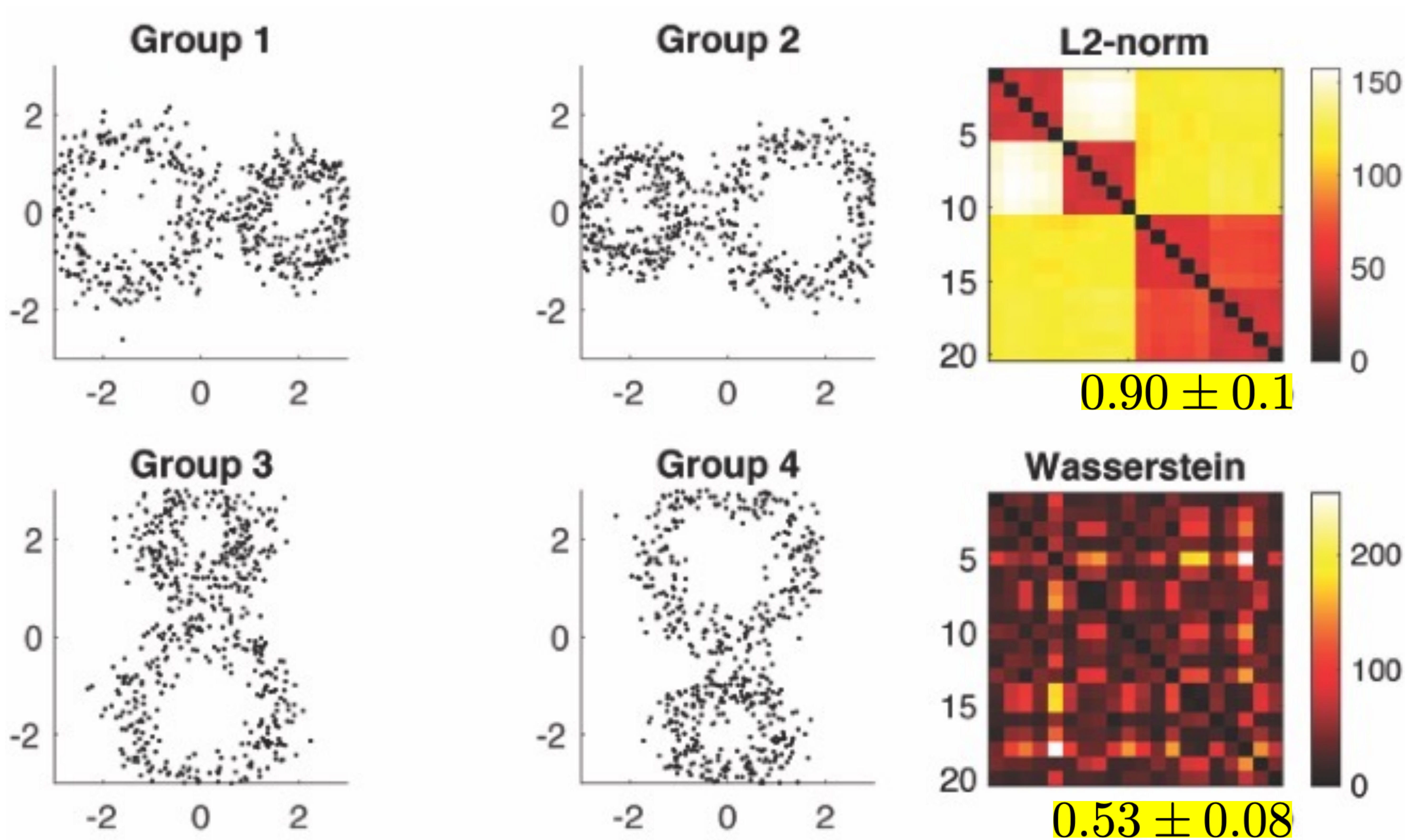


$$\mathcal{D}_0(P_1, P_2)^2 = \sum_i (b^1_{(i)} - b^2_{(i)})^2$$

Songdechakraiwu & Chung 2023

Annals of Applied Statistics 17:403-433

Topologically invariant patterns



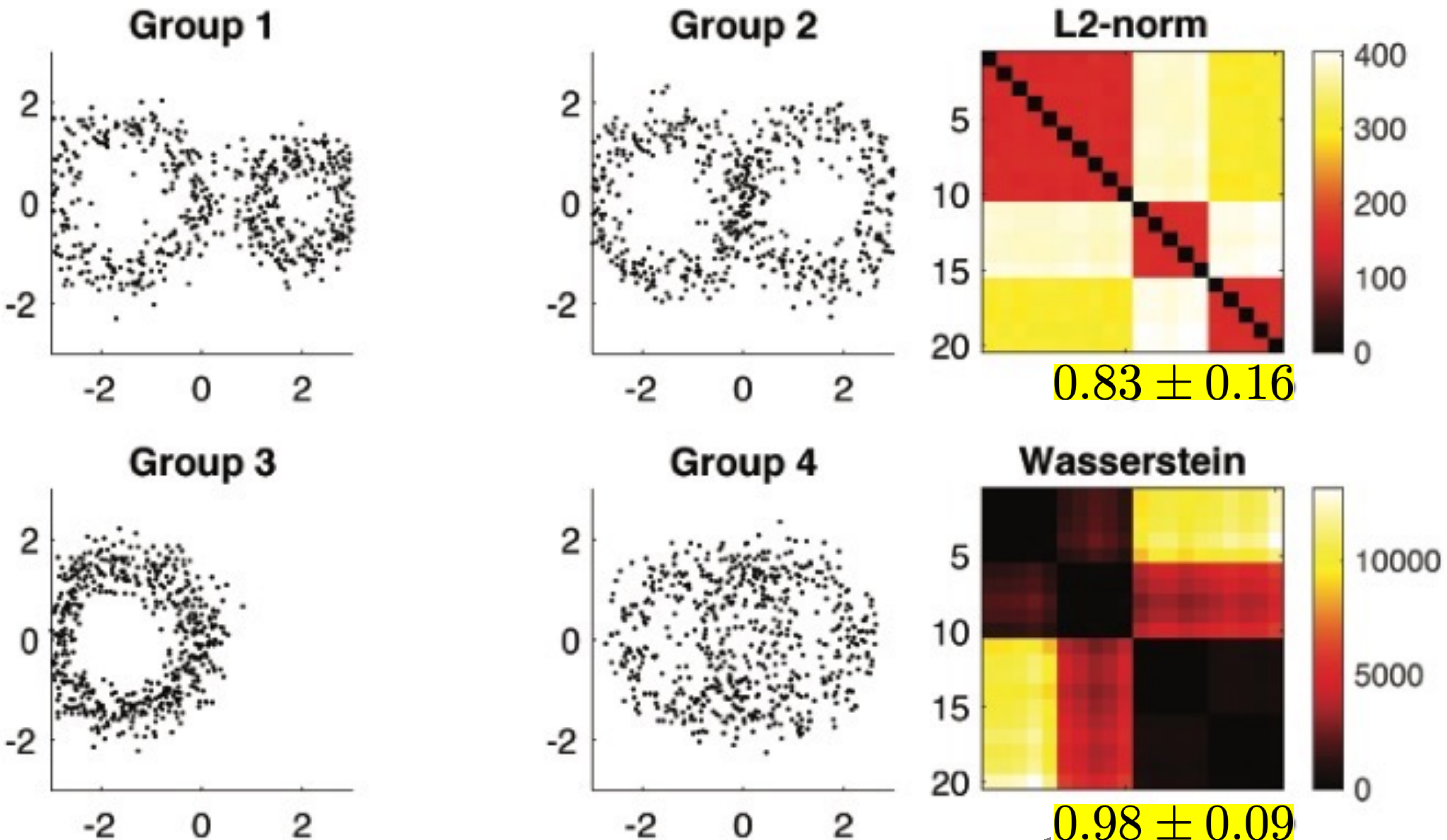
0.90 ± 0.1

0.53 ± 0.08

`WS_cluster.m`

Clustering accuracy

Topologically different patterns

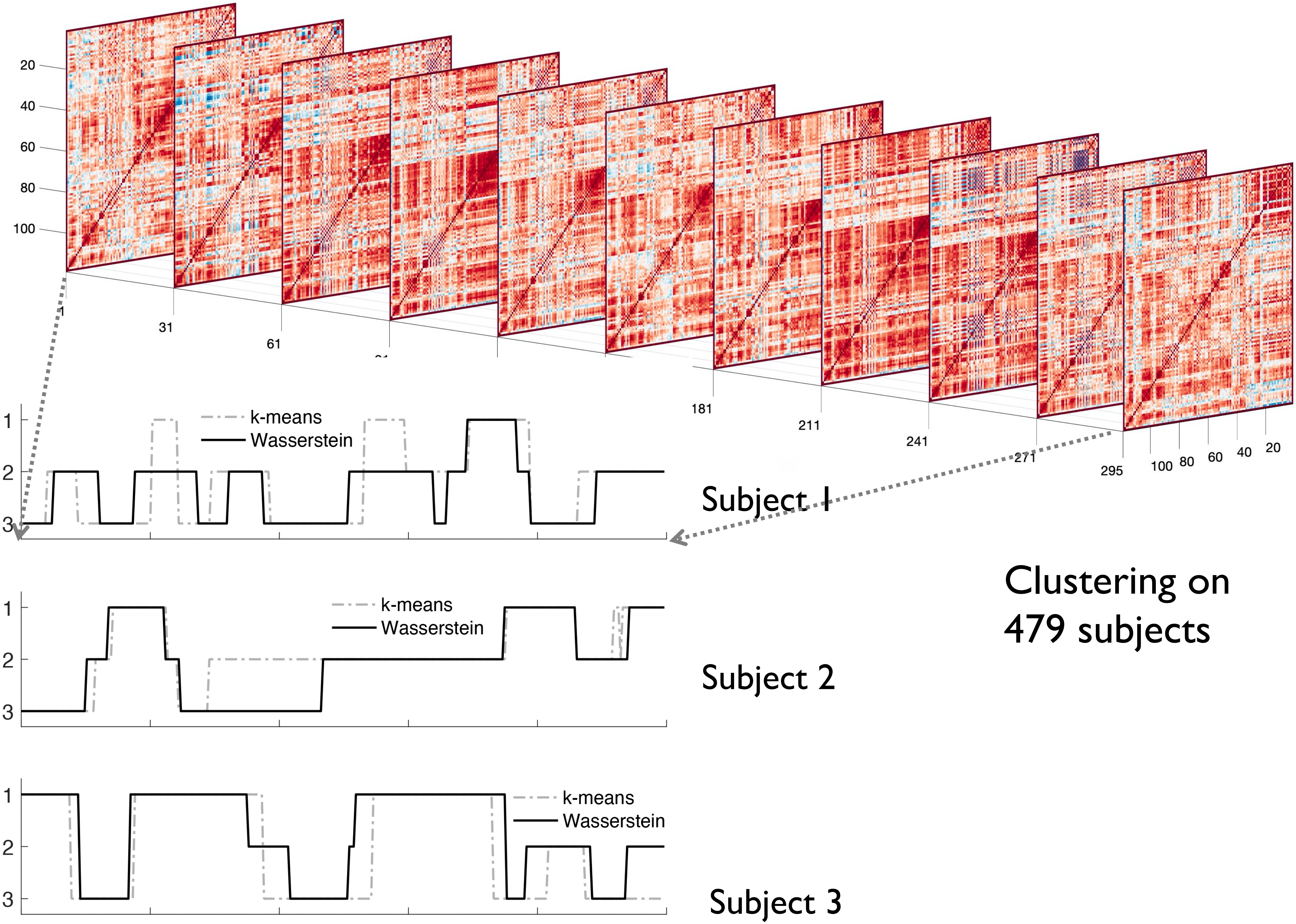


0.83 ± 0.16

0.98 ± 0.09

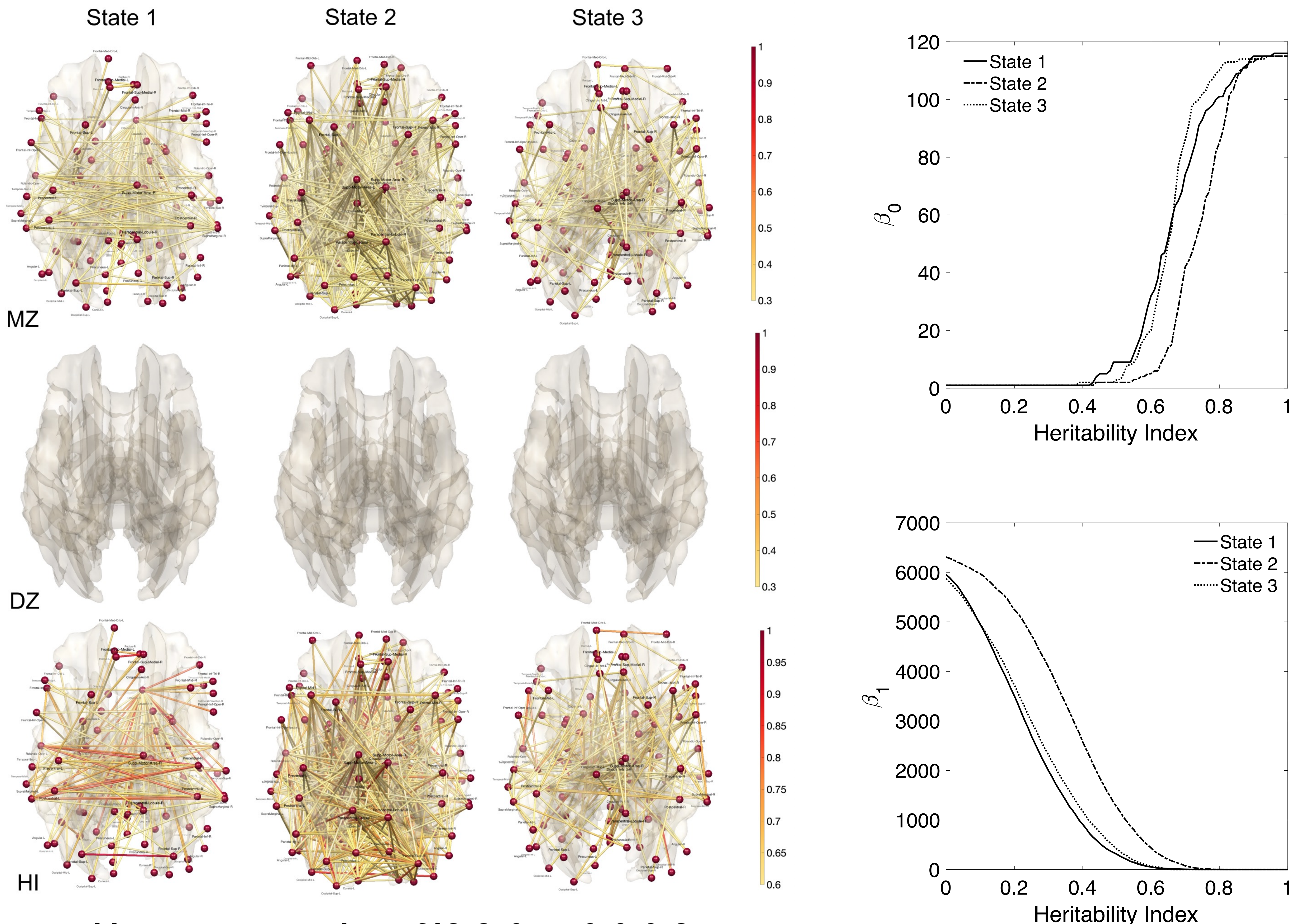
WS_cluster.m

Clustering accuracy



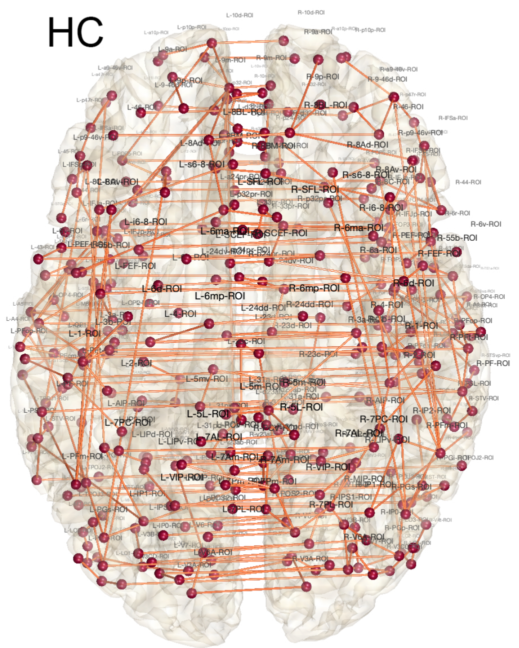
k-means
Wasserstein
Subject 1
Subject 2
Subject 3
Clustering on 479 subjects

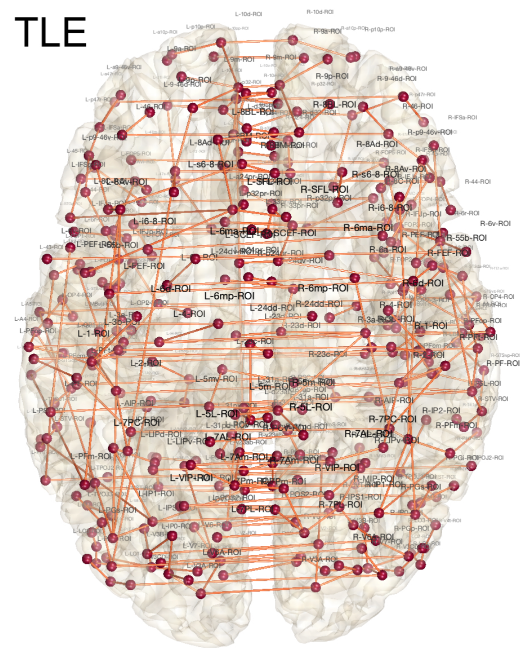
Heritability of state-space of rs-fMRI brain network

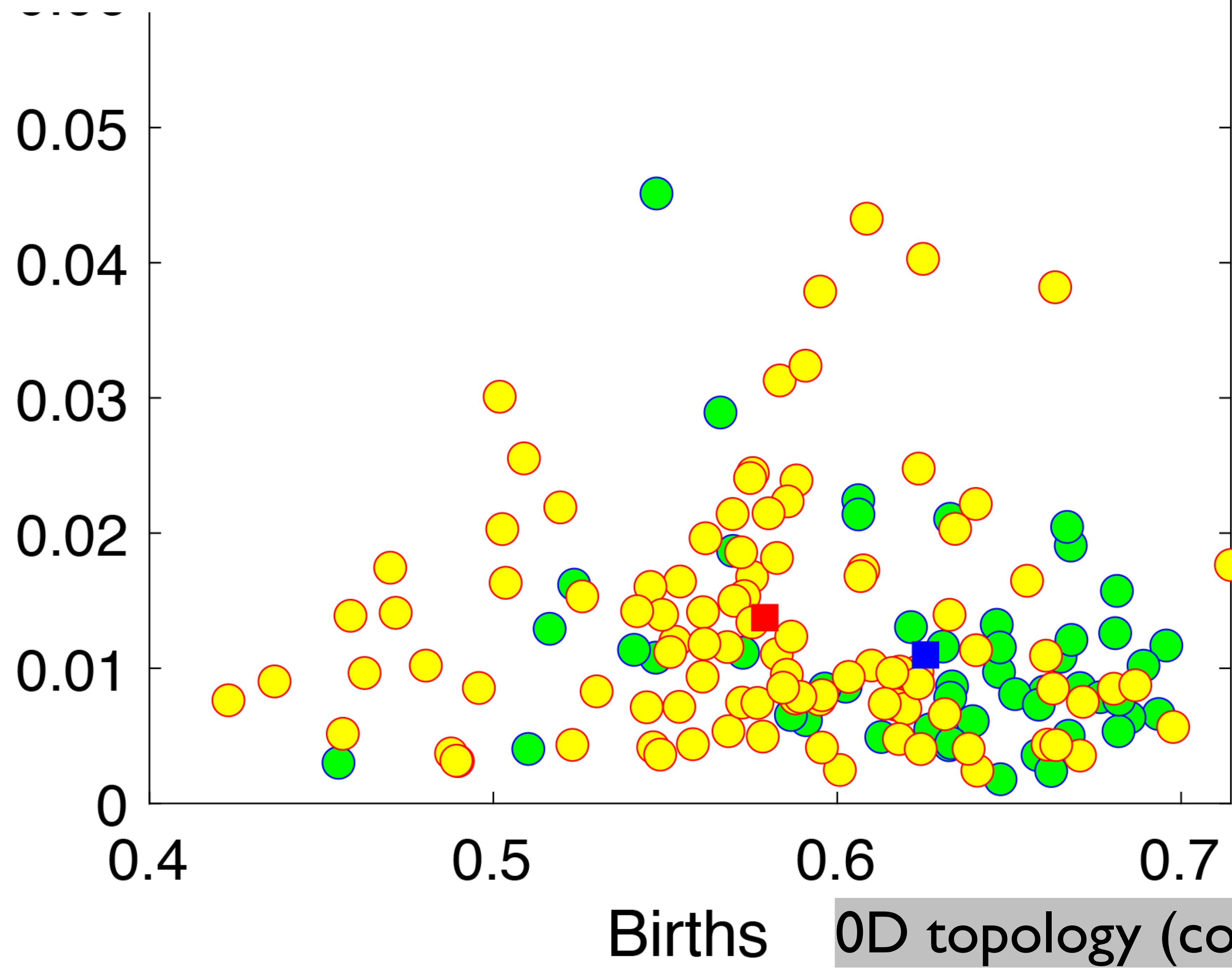


https://arxiv.org/pdf/2201.00087

Orthogonal topological embedding







1D topology (cycles)

0D topology (connected components)

Healthy controls (HC)

Temporal lobe epilepsy (TLE)

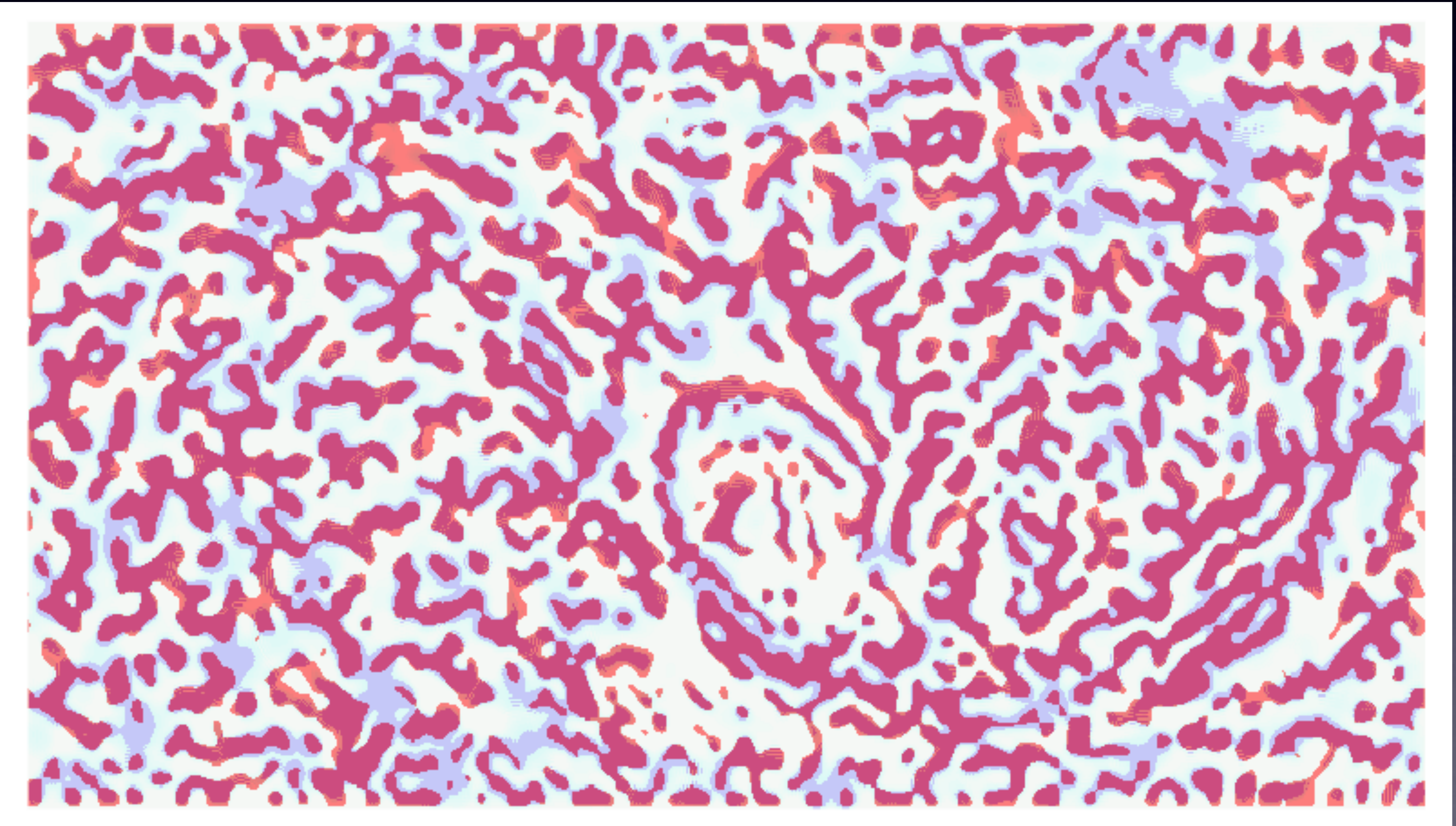
WS_embed.m

ISBI 2023 Friday 4:00-5:40pm

SALÓN BARAHONA 4

Special Session 6:

Wasserstein Distance in Biomedical Imaging



Minisymposium Topological Data Analysis and Machine Learning
August 20-25, 2023 Tokyo, Japan

10th International Congress on Industrial and Applied Mathematics

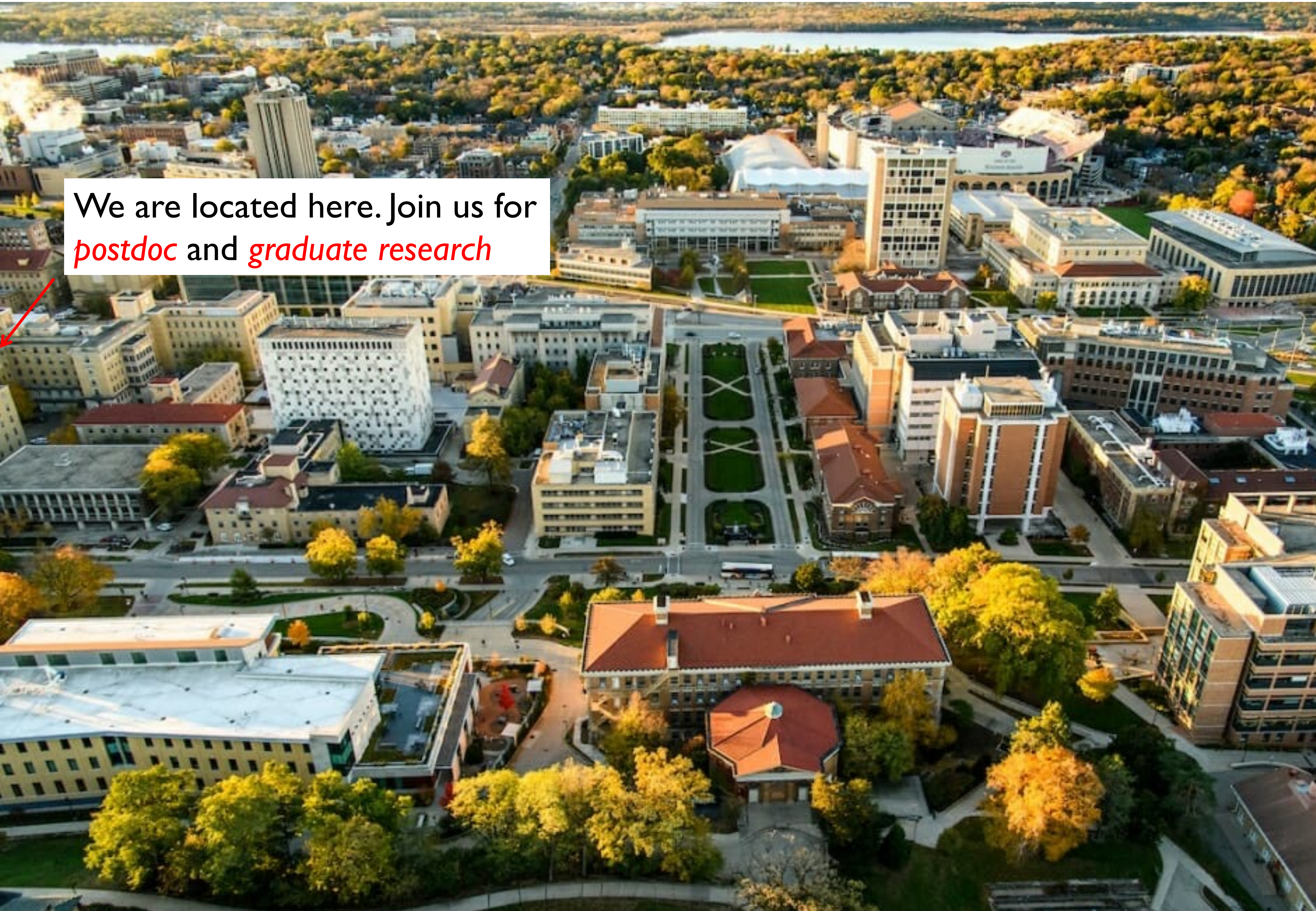
ICIAM 2023 TOKYO

Organizer(s) : Jae-Hun Jung, Shizuo Kaji, Moo K. Chung

Speakers Info :

- Tomoo Yokoyama (Gifu University)
- Jongbaek Song (KIAS)
- Mason Poter (UCLA)
- Keunsu Kim (POSTECH)
- Peter Bubenik (University of Florida)
- Tamal K. Dey (Purdue University)
- Yuan Wang (University of South Carolina)
- Guowei Wei (Michigan State University)
- Alexander Strang (Chicago University)
- Mathieu Carriere (INRIA)
- Heather Harrington (Oxford University)

Thank you.



We are located here. Join us for *postdoc* and *graduate research*