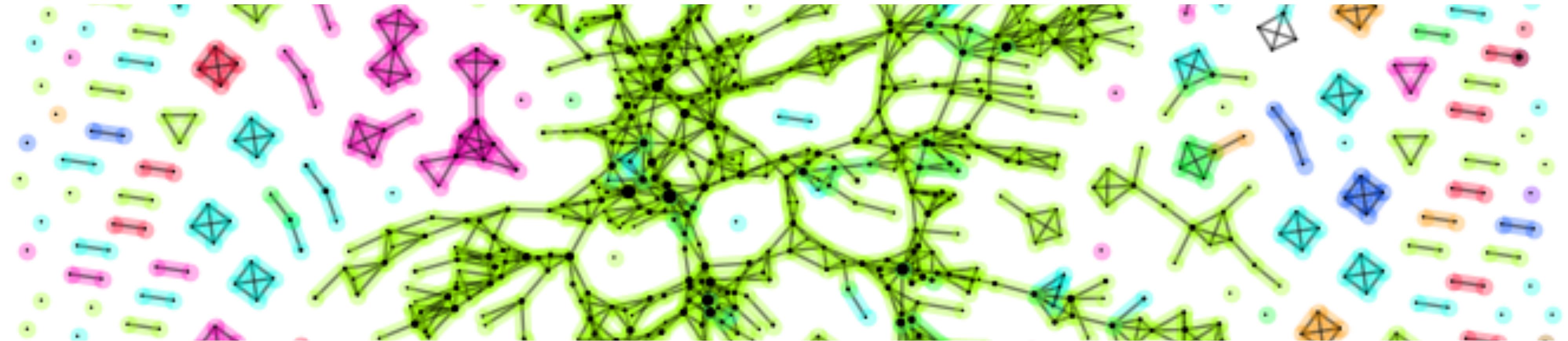


# TDA Lecture #3

## Mapper and persistent homology



**G. Petri/P. Expert**

**ISI** ISI Foundation  
& ISI Global Science  
Foundation

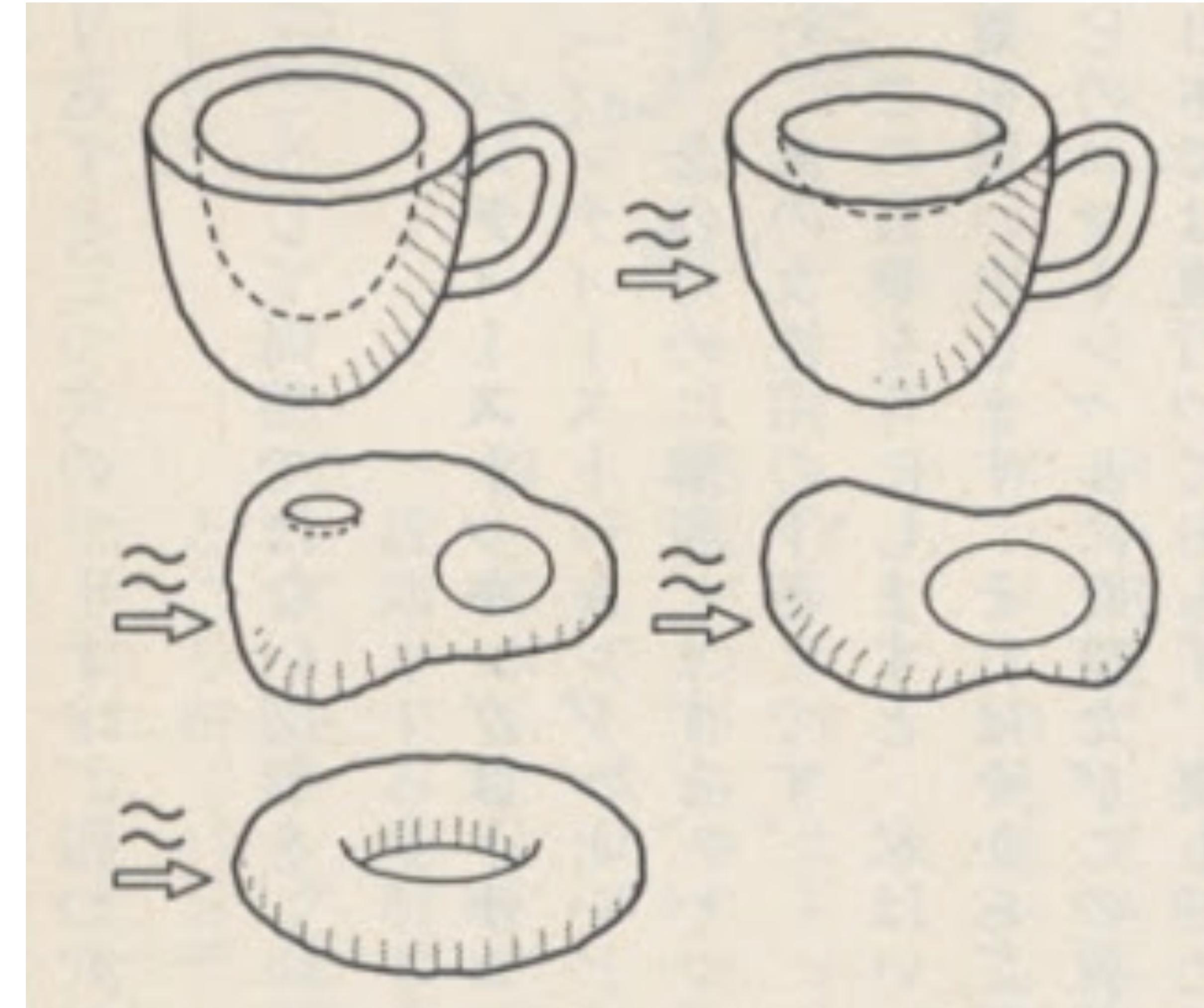
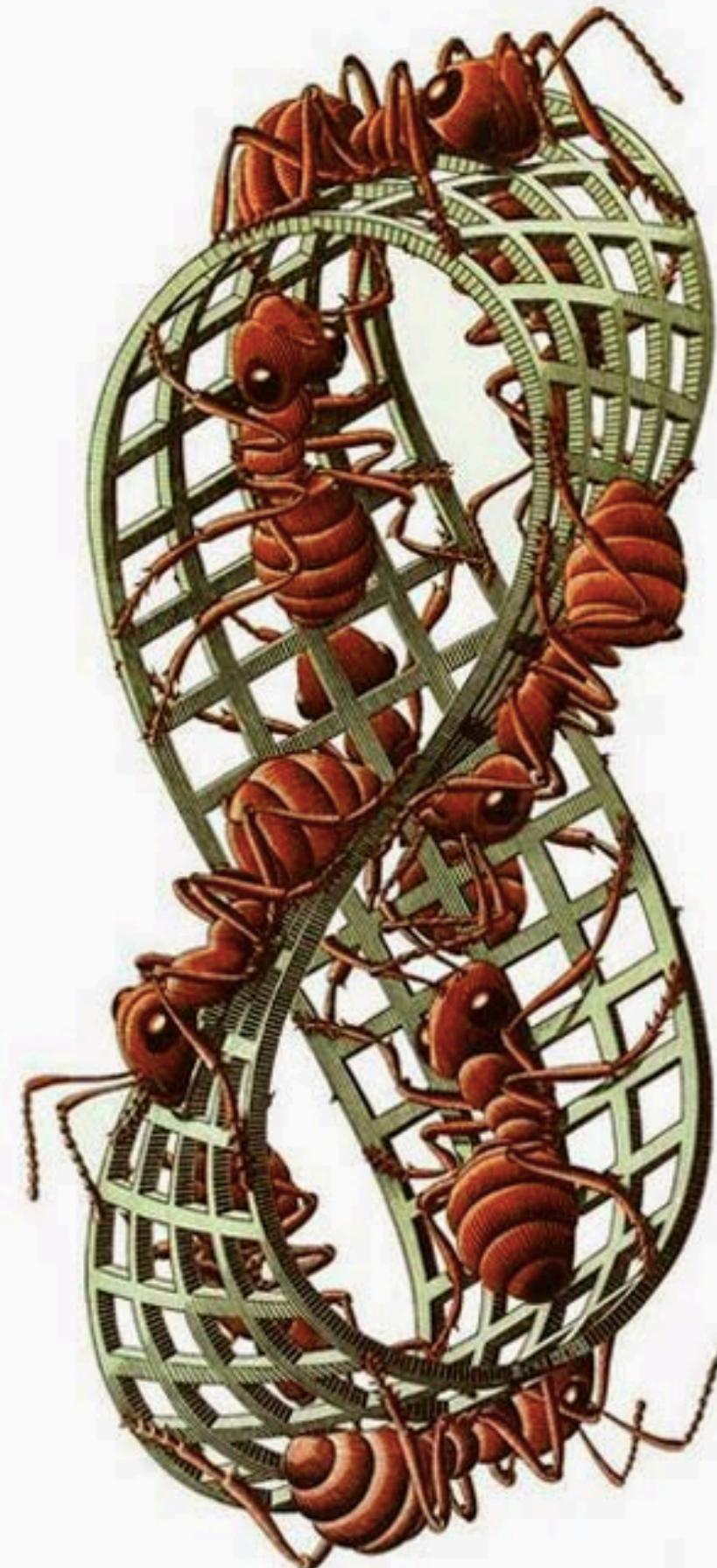
CNTDA2019 , UKM, Kuala Lumpur  
Sept 2019

# What do we mean by topology?

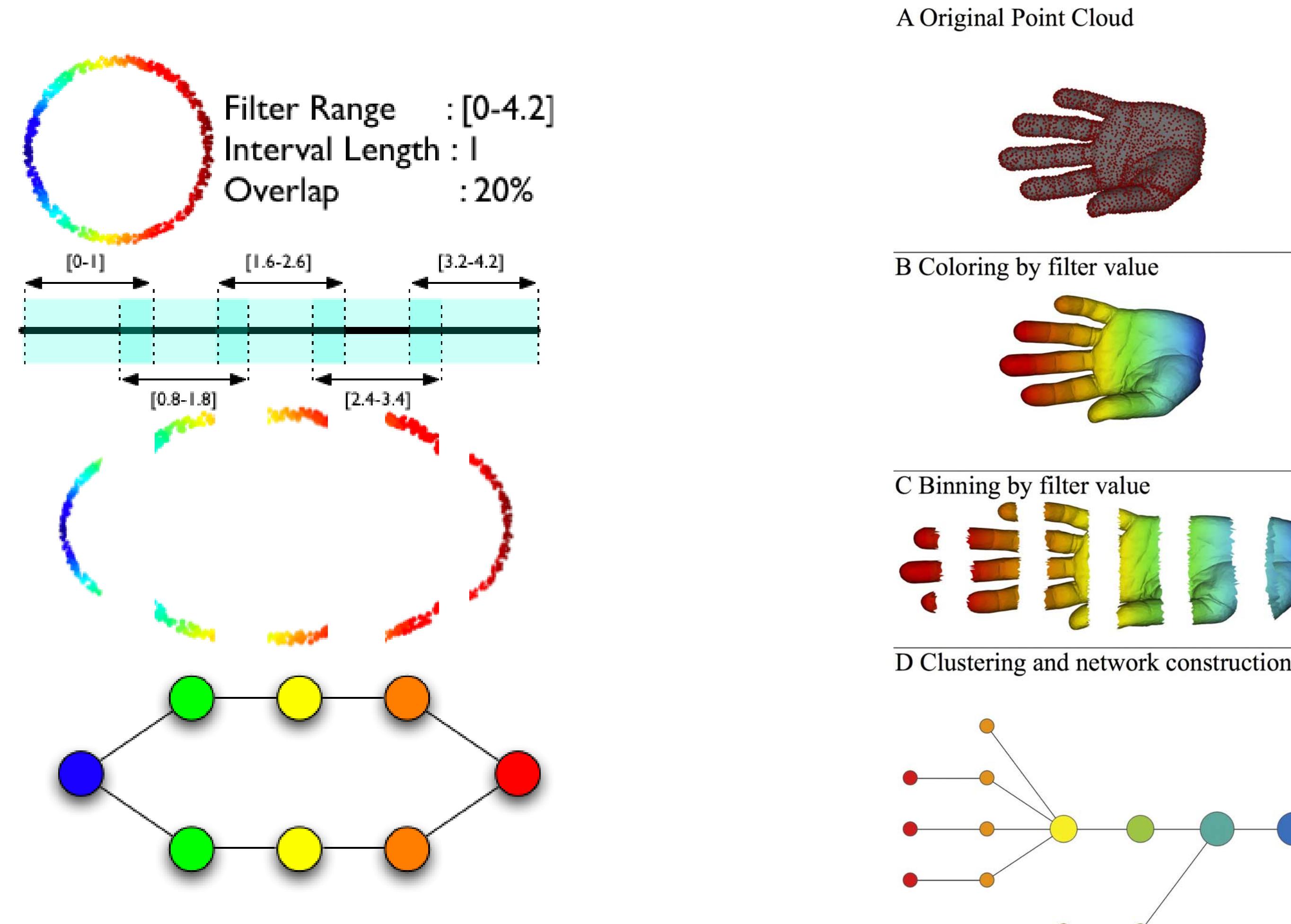
— 102. —  
 auf die Fläche gerichtet, sich nur entlang der Fläche bewegen kann, so kann dasselbe, wenn es einmal an der Außenseite sich befindet, wie es sich auch bewegen mag, niemals an die Innenseite gelangen und umgekehrt. Ebenso kann man entweder die Außenseite oder die Innenseite der Fläche für sich mit Farbe anstreichen. Doch nun kann man den Schlauch noch in ganz anderer Weise zusammenfügen, indem man nämlich das eine Ende nach innen umschläft, das andere dagegen durch die Wandung in das Innere hineinleitet und dann mit dem umgeschlagenen Ende vereinigt. v. Fig. 21. c.  
 Fig. 21.



So med diese Weise haben wir eine durchaus zusammenhängende Doppelfläche gewonnen, bei welcher eine Innere- und Außenseite etwa durch besonderen farbigen Anstrich nicht mehr zu unterscheiden ist. Denken wir uns auf dieser Fläche ein zweidimensionales Wesen, so wird dies, indem es an seinen früheren Ort zurückgelangt, dabei sein eigener Anblick verlieren können, und es muss zweimal herumkricken, ehe es in die Ausgangslage zurück-



# TDA as topological simplification



Topological Methods for the Analysis of High Dimensional  
Data Sets and 3D Object Recognition

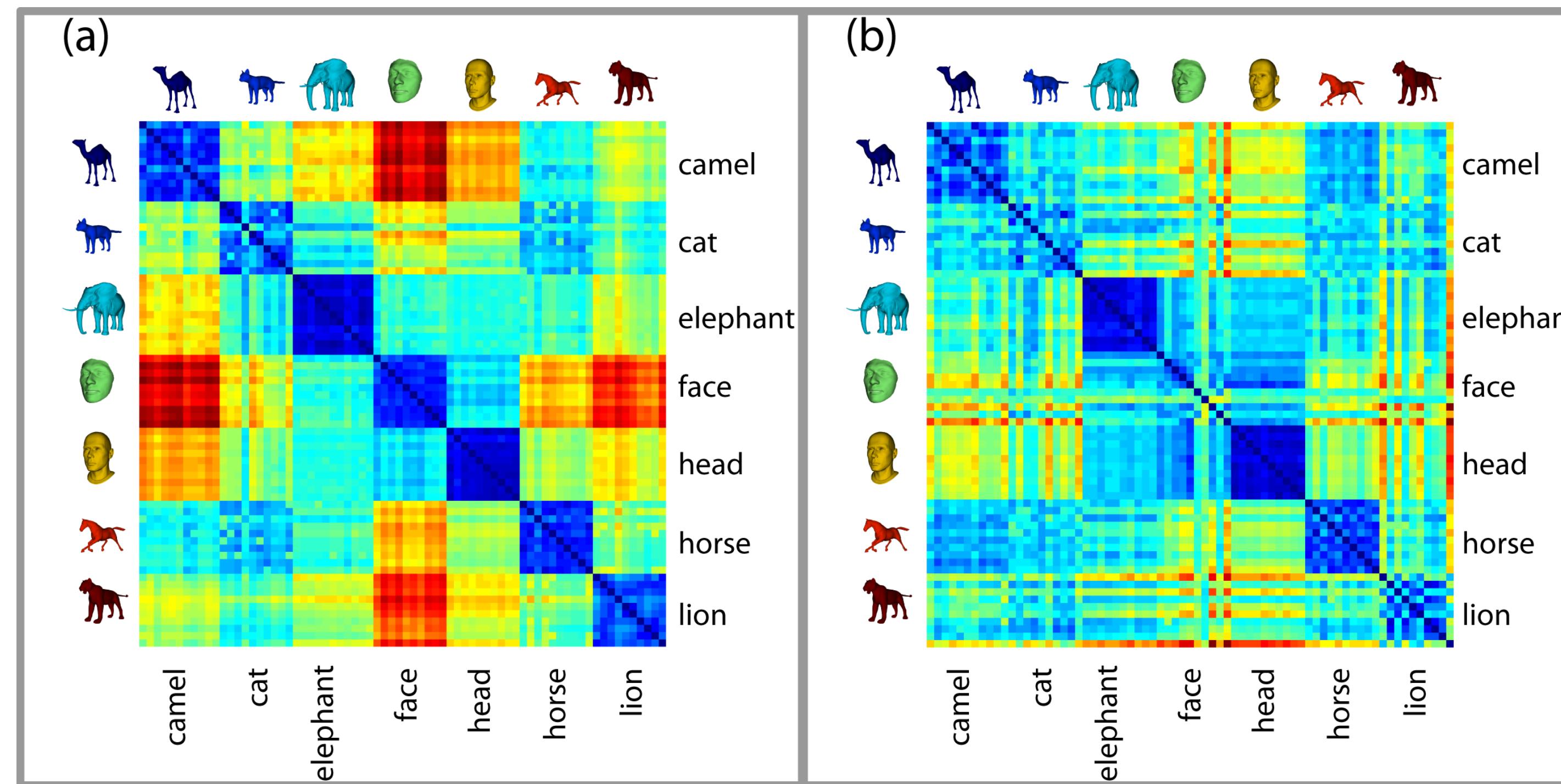
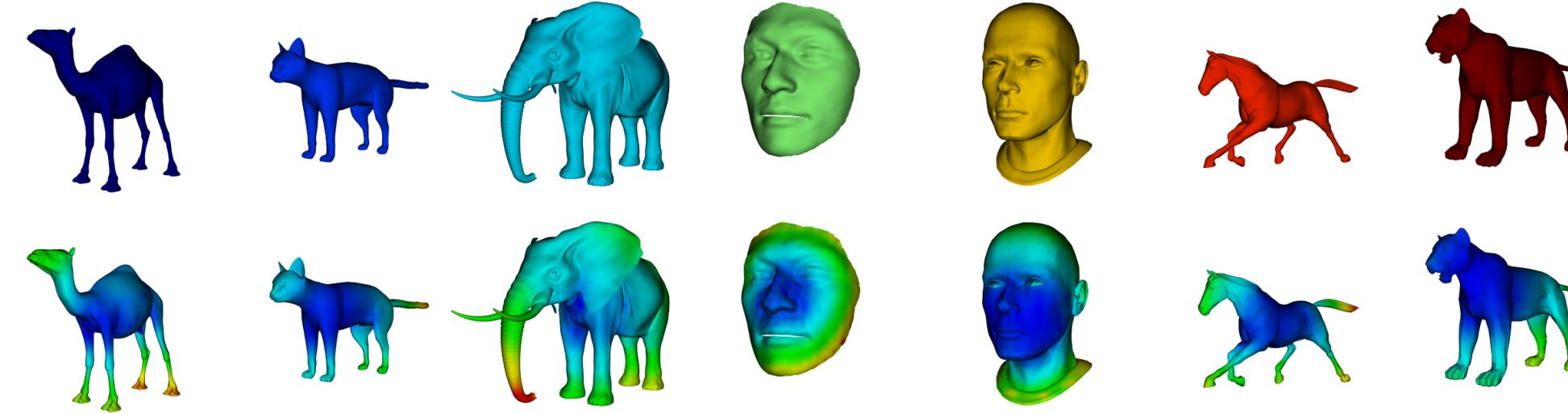
Gurjeet Singh<sup>1</sup>, Facundo Mémoli<sup>2</sup> and Gunnar Carlsson<sup>1,2</sup>



OPEN SUBJECT AREAS:  
APPLIED MATHEMATICS  
COMPUTATIONAL SCIENCE  
SCIENTIFIC DATA  
Extracting insights from the shape of  
complex data using topology

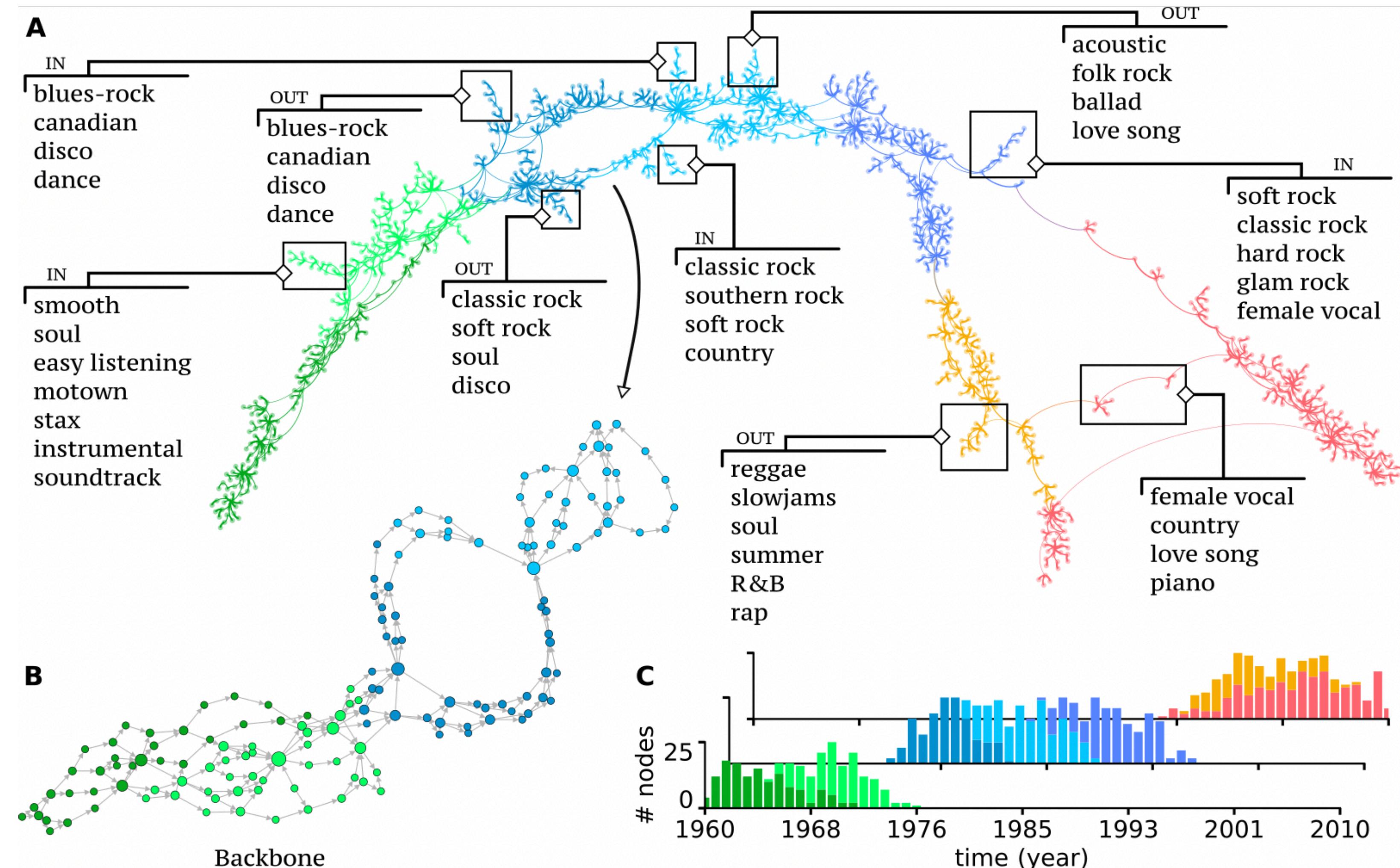
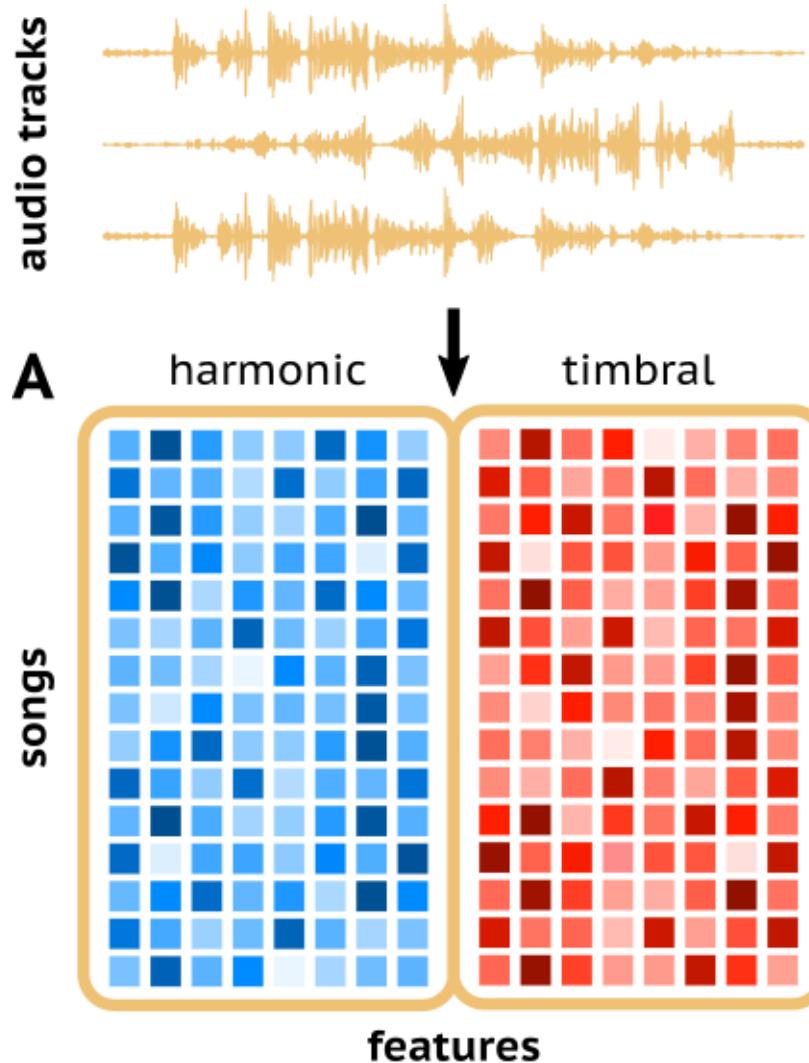
<sup>1</sup>Jayadev Inc., Palo Alto, CA, <sup>2</sup>School of Computer Science, Jack Baskin School of Engineering, University of California Santa Cruz, Santa Cruz, CA, United States, <sup>3</sup>School of Mathematics, University of Minnesota, Minneapolis, MN 55455, USA, <sup>4</sup>Department of Mathematics, Stanford University, Stanford, CA, 94301, USA

# TDA as topological simplification

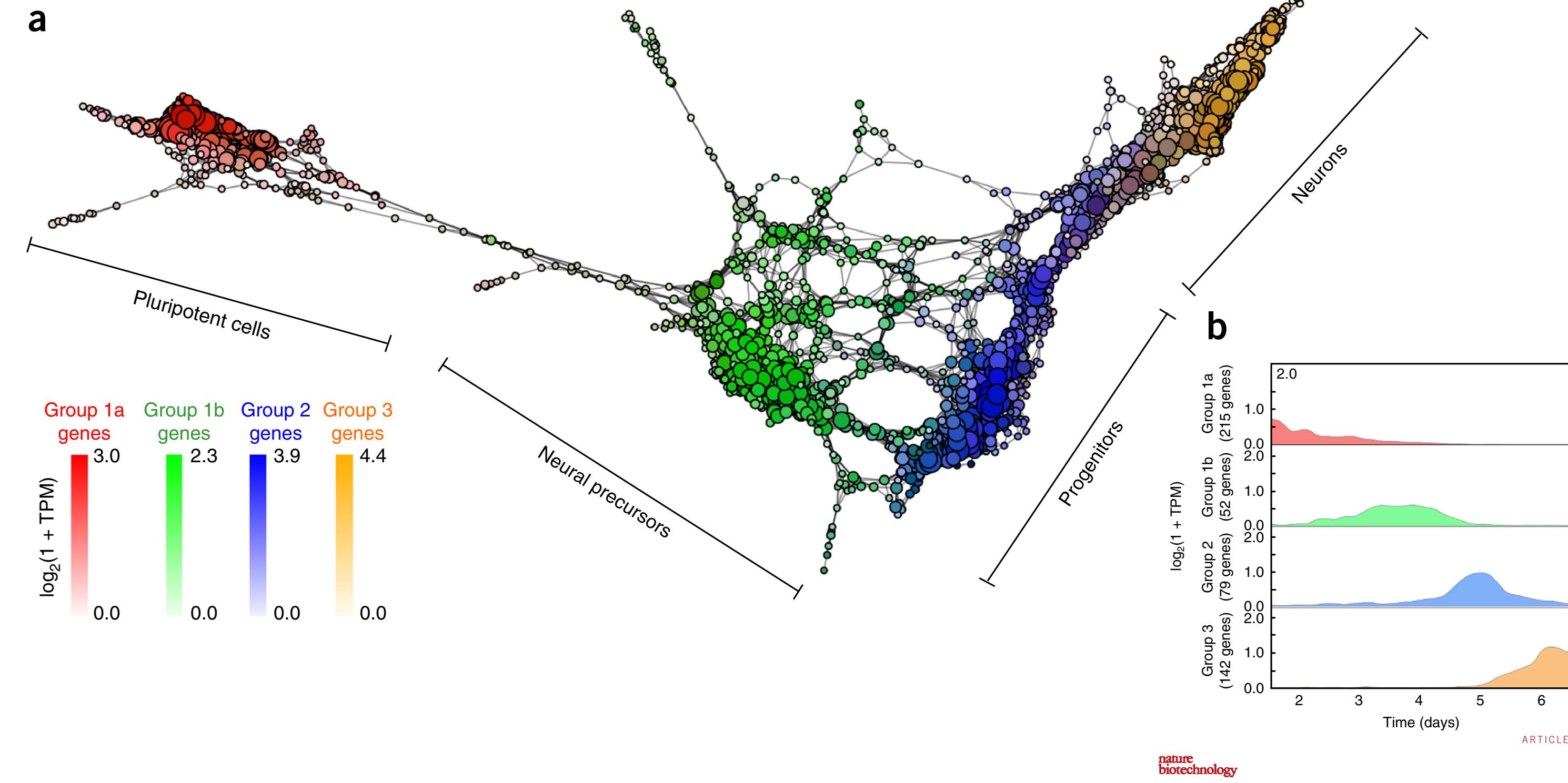
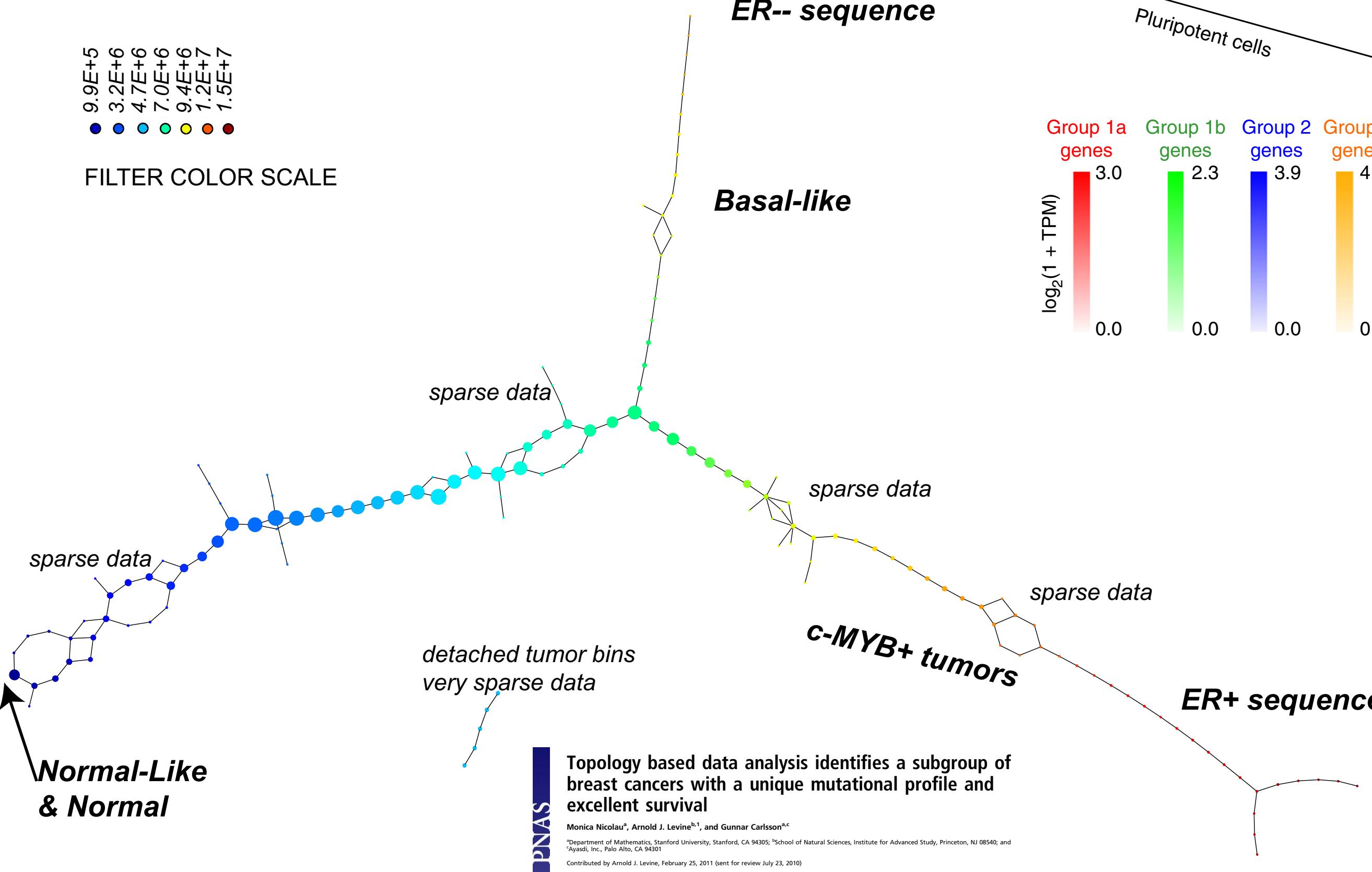


# Notebook 01

# What does it mean in practice?



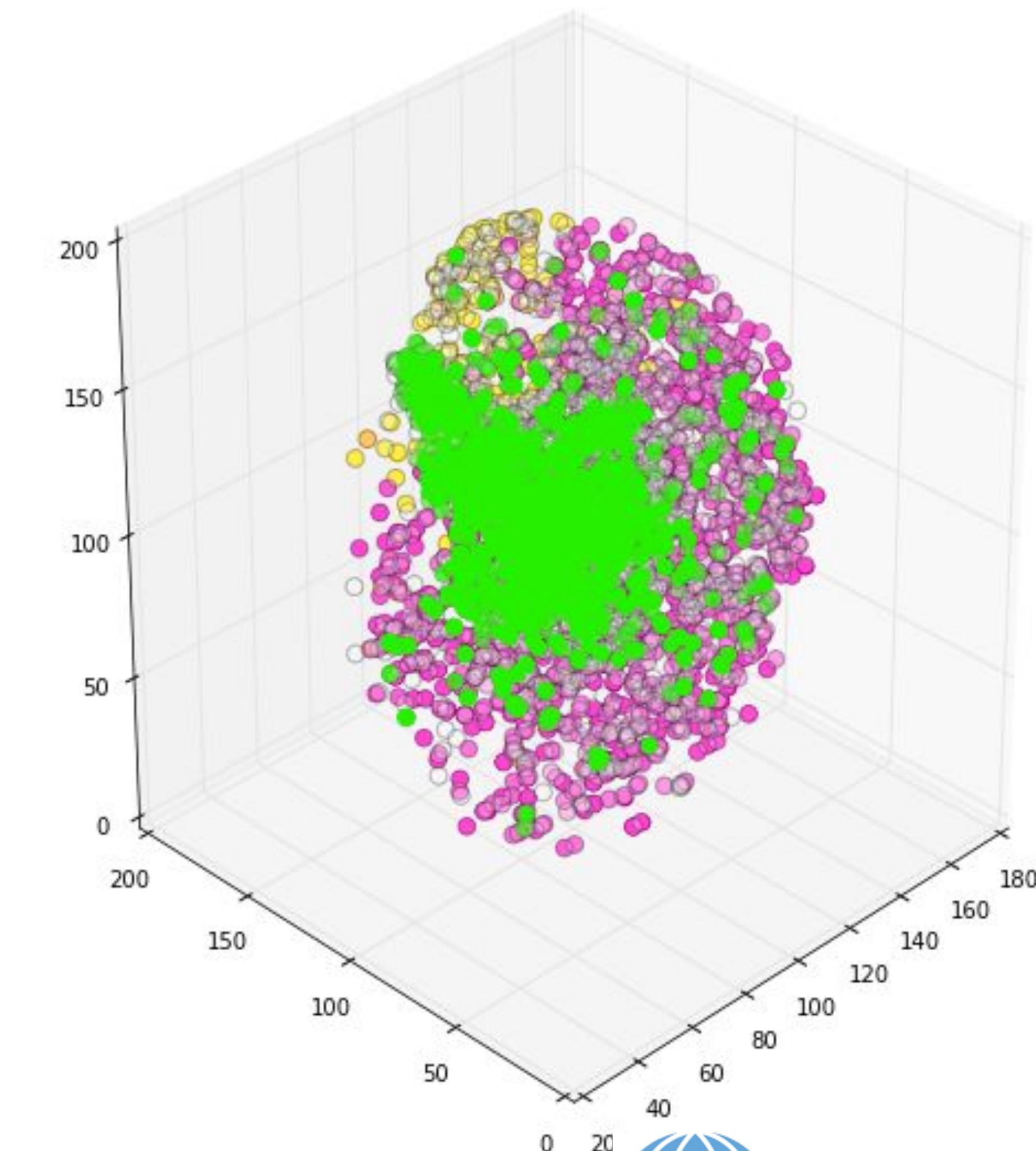
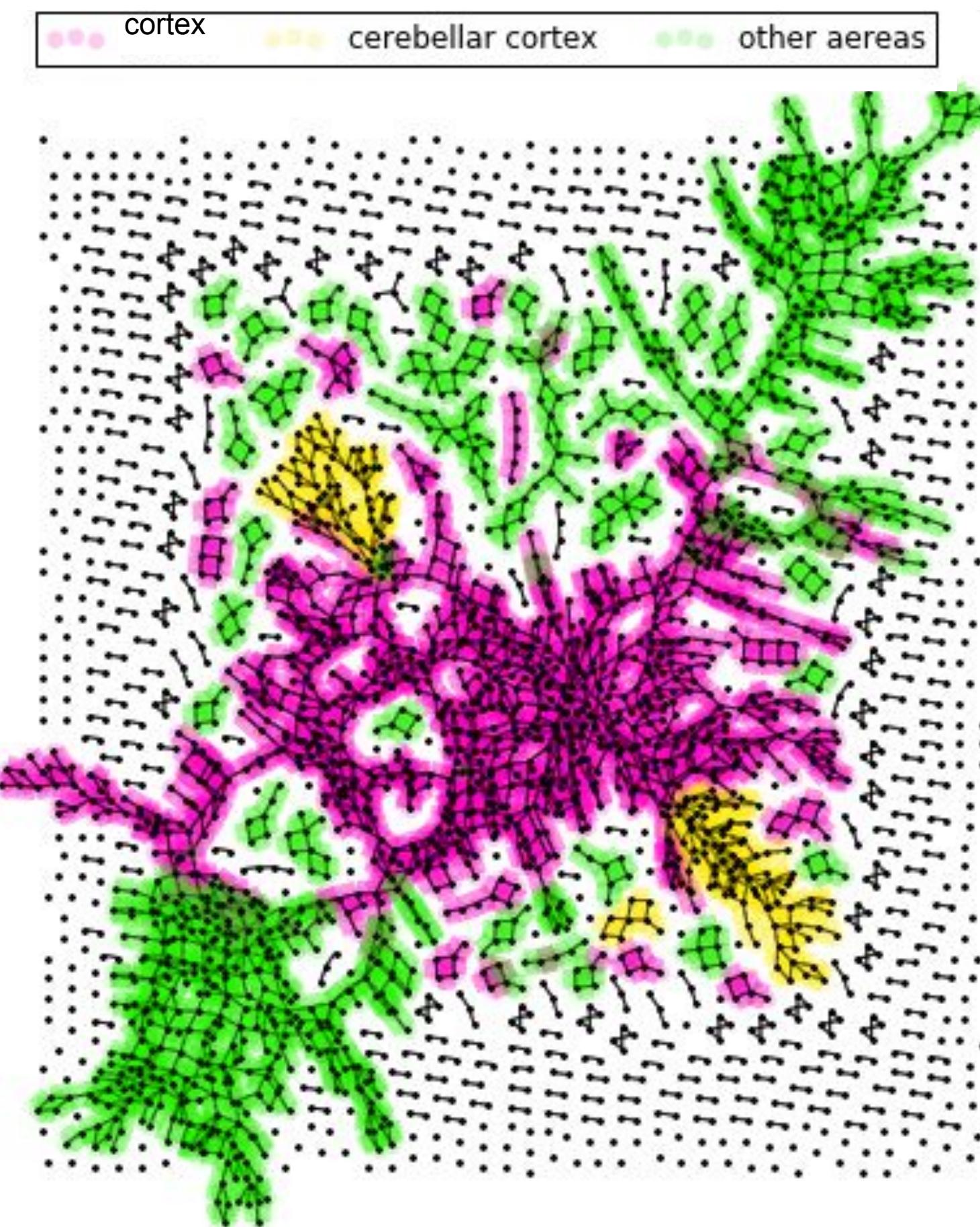
# Do topological gene-backbones carry information?



Single-cell topological RNA-seq analysis reveals insights into cellular differentiation and development  
Abbas H Rizvi<sup>1,2,6</sup>, Pablo G Camara<sup>3,4,6</sup>, Elena K Kandror<sup>1,2</sup>, Thomas J Roberts<sup>1,2,4</sup>, Ira Schieren<sup>2,5</sup>, Tom Maniatis<sup>1,2</sup>& Raul Rabanal<sup>3,4</sup>

# Notebook 02

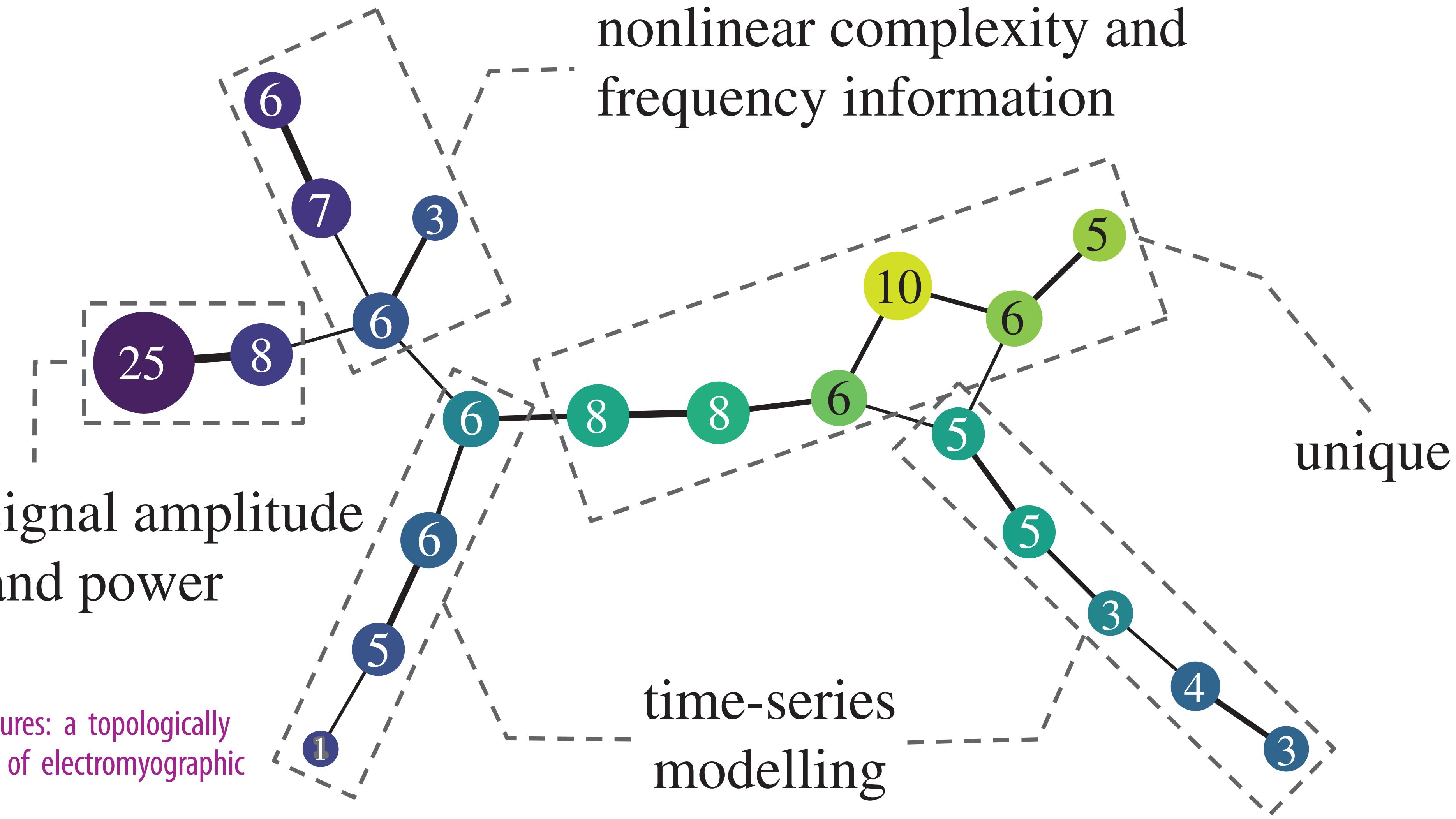
# Do topological gene-backbones carry information?



ALLEN INSTITUTE for  
BRAIN SCIENCE

# Can we build topological feature charts?

nonlinear complexity and frequency information



INTERFACE

[rsif.royalsocietypublishing.org](http://rsif.royalsocietypublishing.org)

Navigating features: a topologically informed chart of electromyographic features space

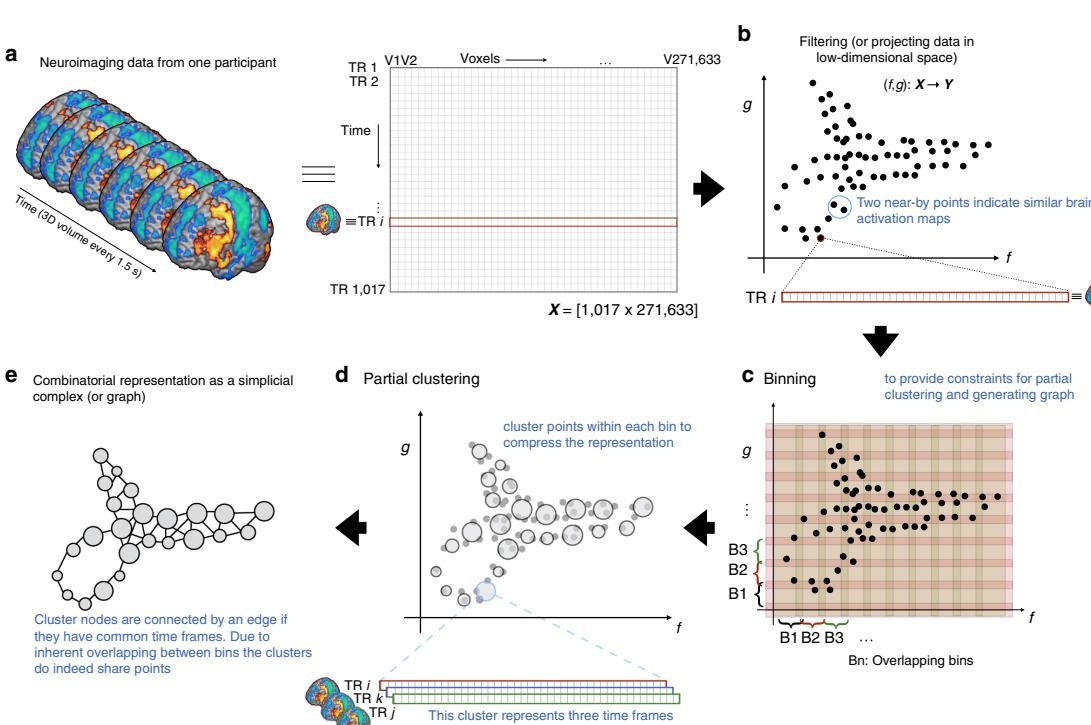
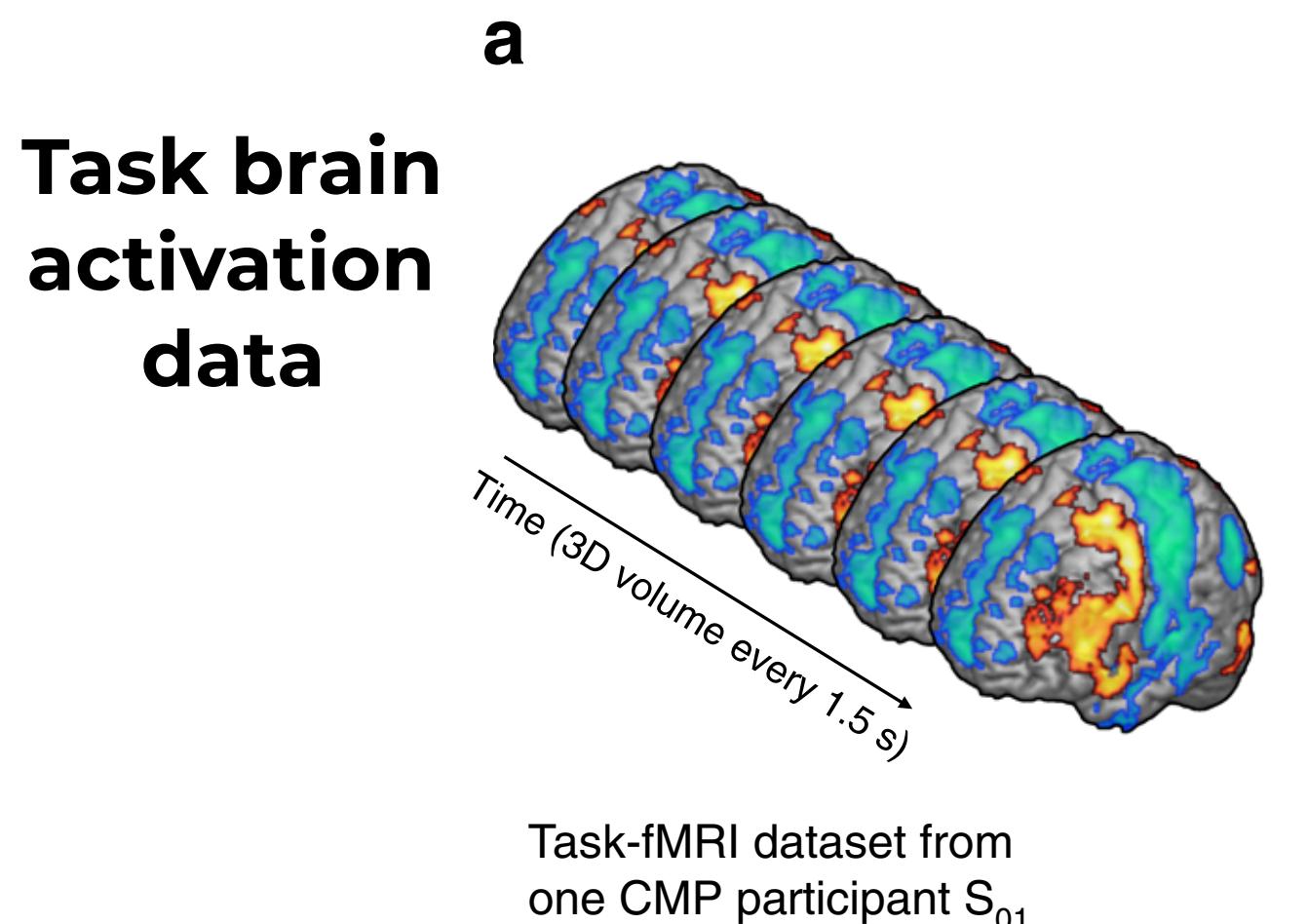
Angkoon Phinyomark<sup>1,2</sup>, Rami N. Khushaba<sup>3</sup>, Esther Ibáñez-Marcelo<sup>1</sup>,  
Alice Patania<sup>4</sup>, Erik Scheme<sup>2</sup> and Giovanni Petri<sup>1</sup>

Research

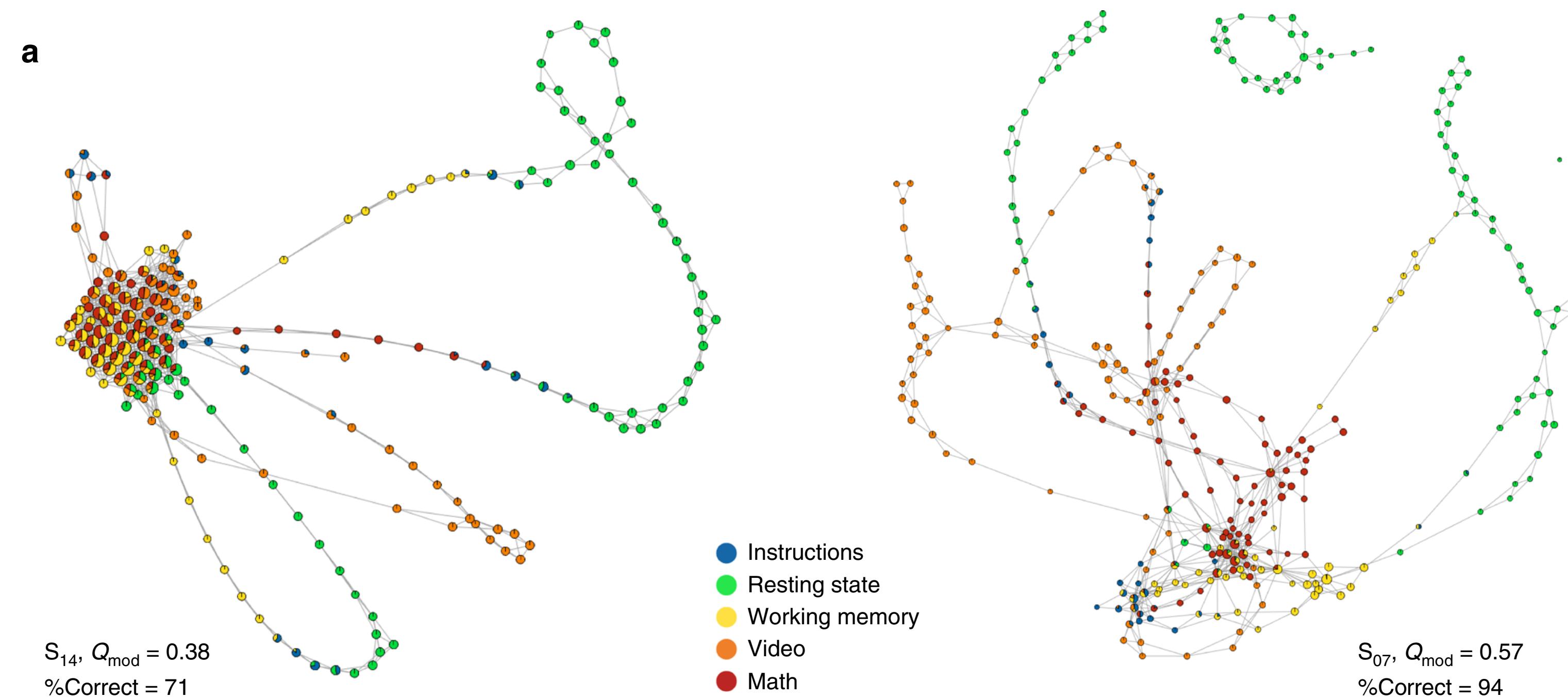


# Notebook 03

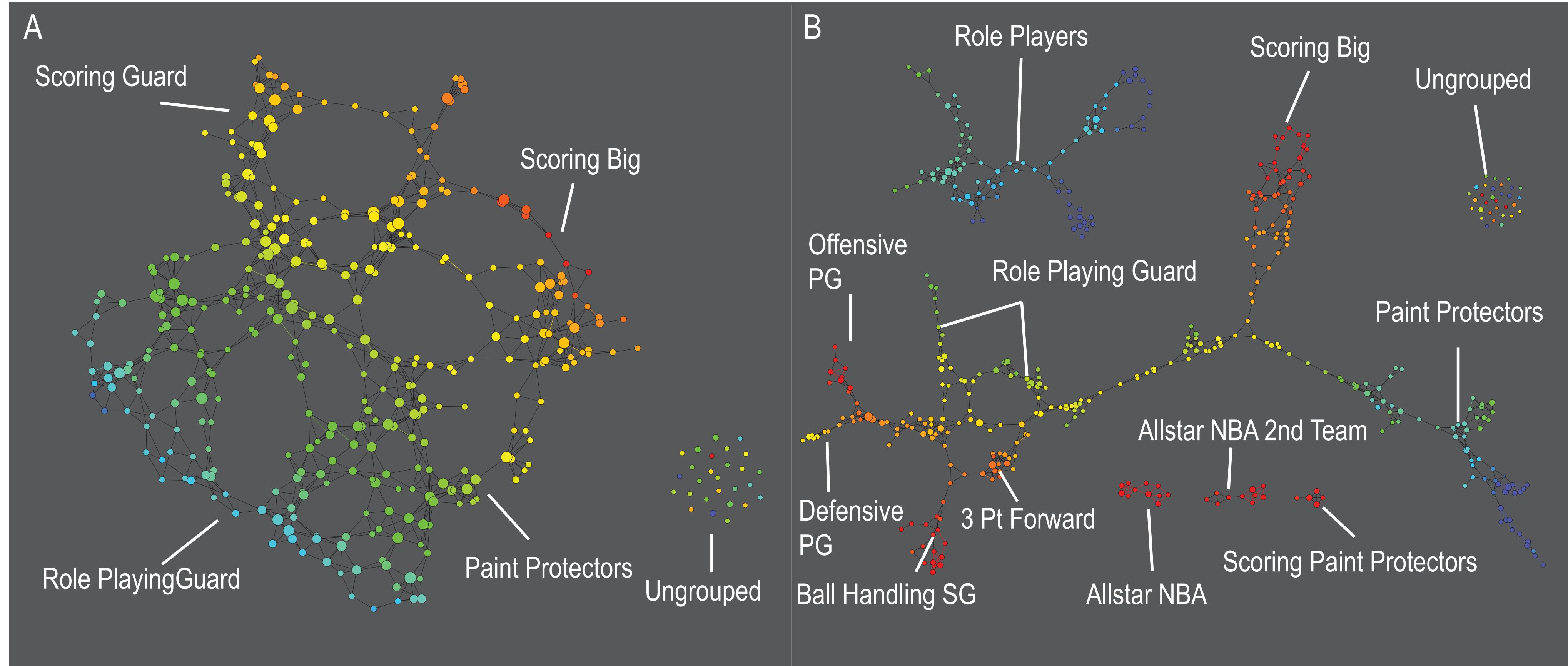
# Approximate activity landscapes using topology



# Approximate activity landscapes using topology



# Multi-resolution analysis?

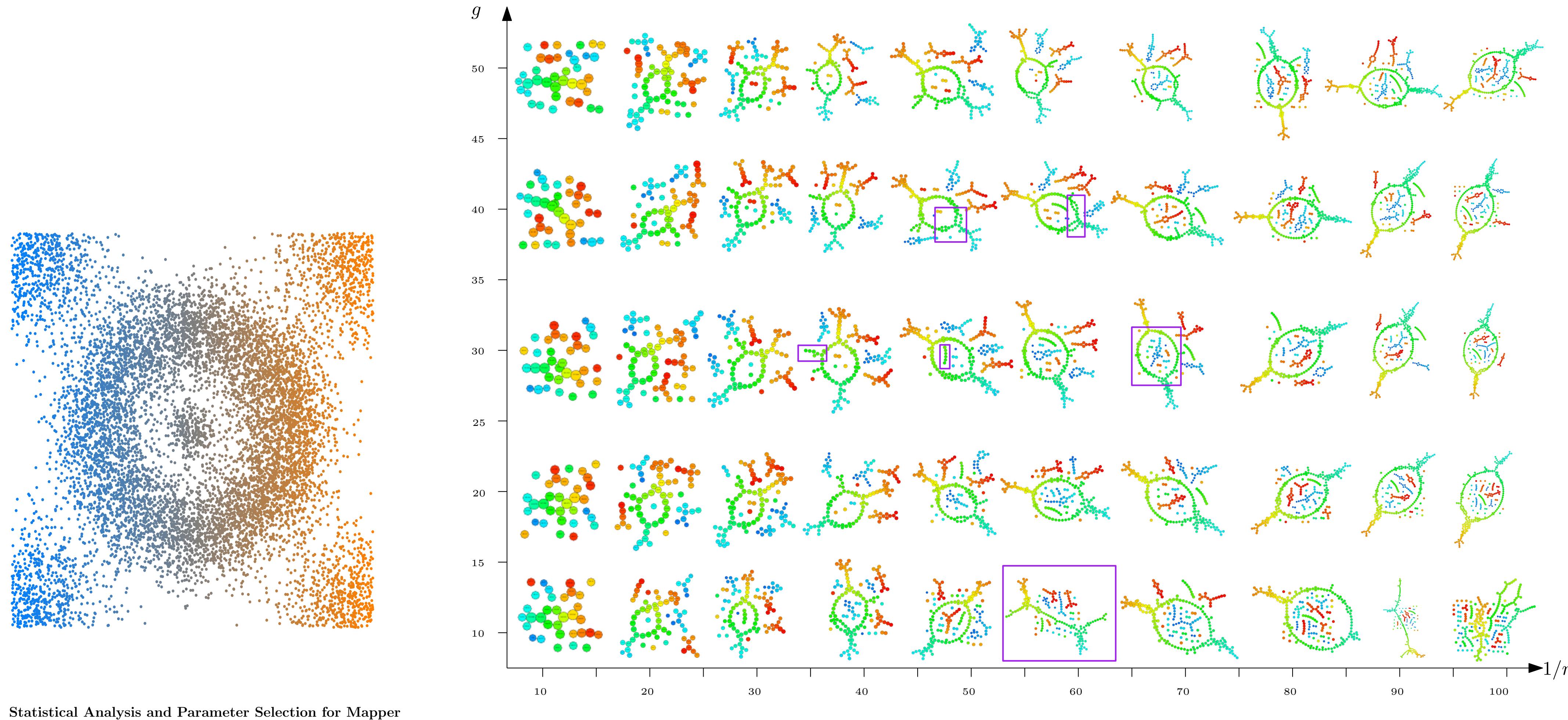


Extracting insights from the shape of  
complex data using topology

P. Y. Lum<sup>1</sup>, G. Singh<sup>1</sup>, A. Lehman<sup>1</sup>, T. Ishkanov<sup>1</sup>, M. Vejdemo-Johansson<sup>2</sup>, M. Alagappan<sup>1</sup>, J. Carlsson<sup>3</sup>  
& G. Carlsson<sup>1,4</sup>

<https://www.wired.com/2012/04/analytics-basketball/>

# Automatic Mapper parameter choice



Mathieu Carrière  
Inria Saclay  
91120 Palaiseau, France

MATHIEU.CARRIERE@INRIA.FR

Bertrand Michel  
LMJL UMR 6629– Ecole Centrale Nantes  
44322 Nantes, France

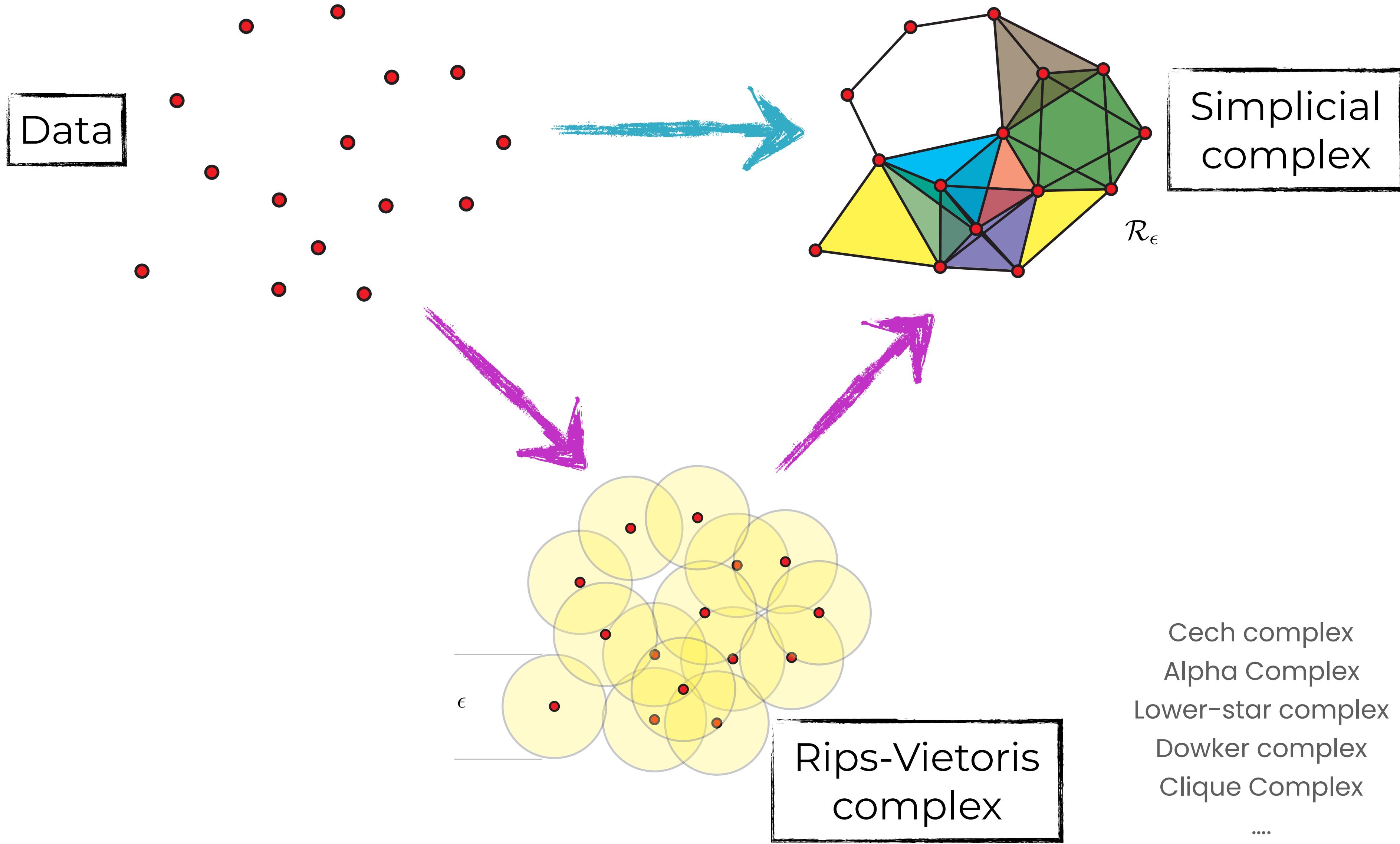
BERTRAND.MICHEL@EC-NANTES.FR

Steve Oudot  
Inria Saclay  
91120 Palaiseau, France

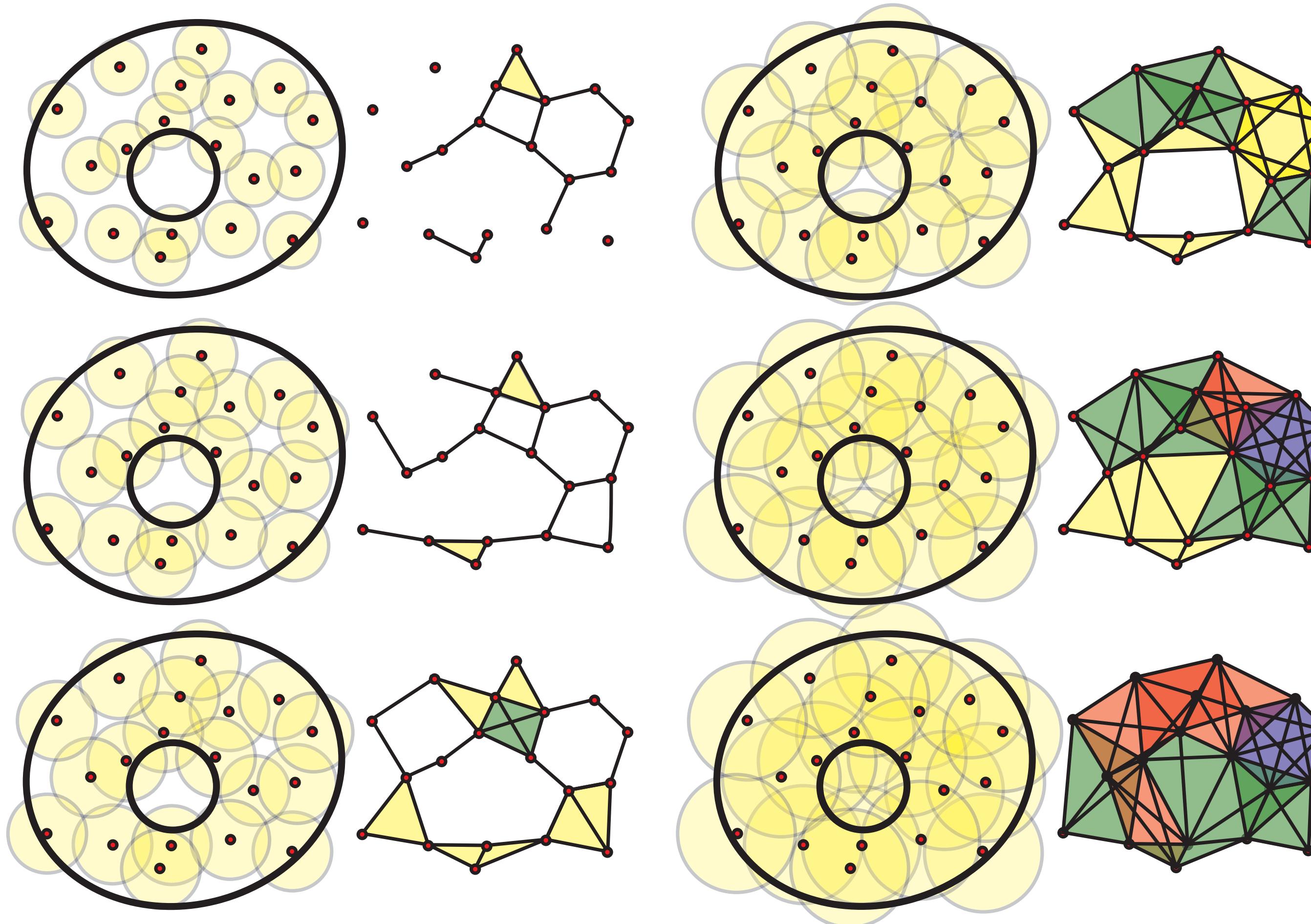
STEVE.OUDOT@INRIA.FR

**Selection based on “extended” persistent homology**

# The shape of data: what simplicial complex?

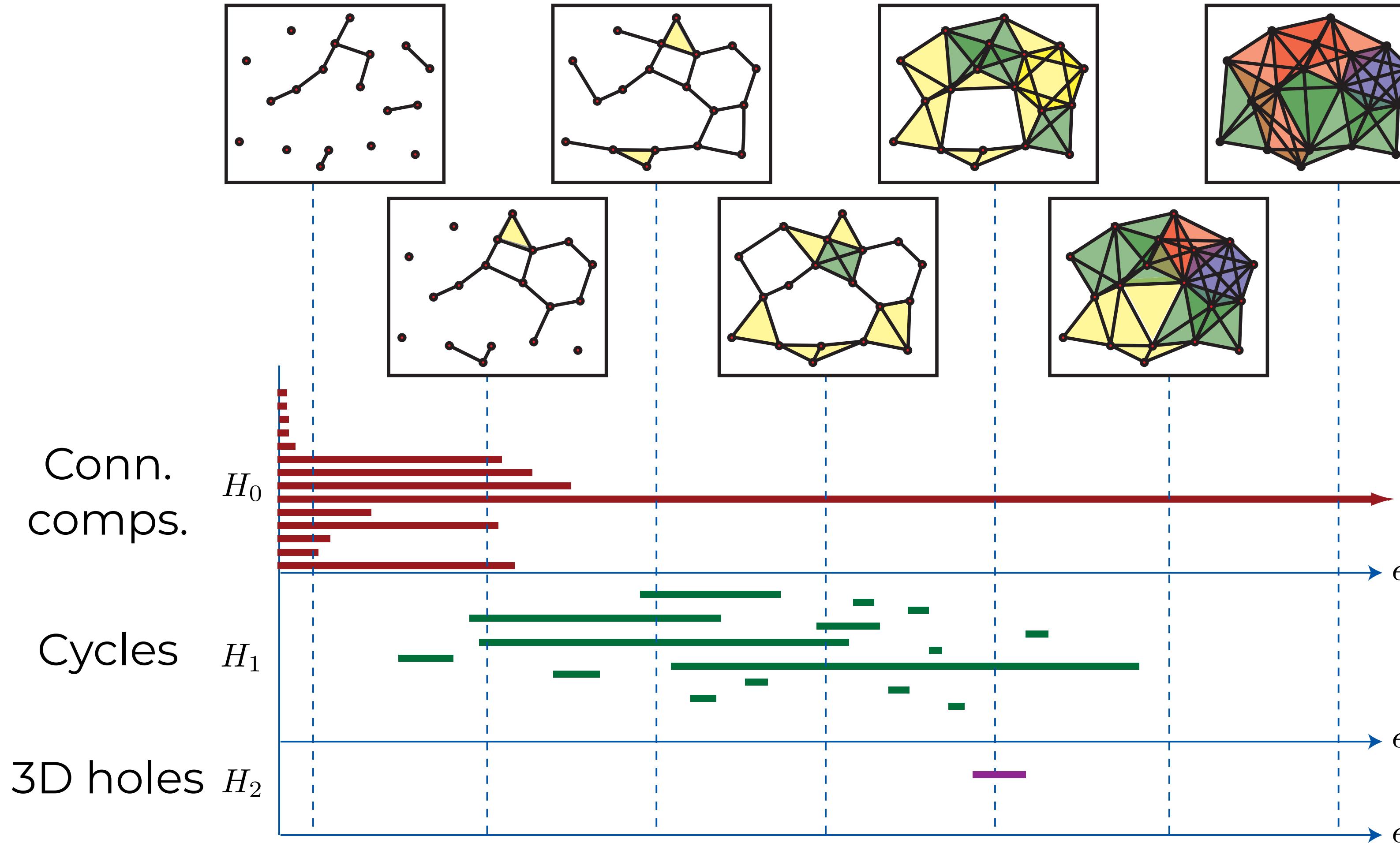


# Persistent homology in two slides.. and a donut!



What is the appropriate **radius** for the balls?

# Persistent homology: Barcode



# Persistent homology in two slides.. and a donut!

The set of vector spaces and linear maps  $\{H_n(X_t), H_n(i_t)\}_{t \in \mathbb{N}}$  can be represented as a finitely generated  $k[x]$ -module:

$$H_n = \bigoplus_{t \in \mathbb{N}} H_n(X_t) \quad \text{with} \quad \cdot x = H_n(X_t) \rightarrow H_n(X_{t+1})$$
$$m \rightarrow H_n(i_t)(m)$$

## Structure theorem

$$M \simeq \bigoplus_{j=1}^n k[x](-a_j) \bigoplus_{i=1}^m k[x](-c_i)/x_i^{d_i}$$

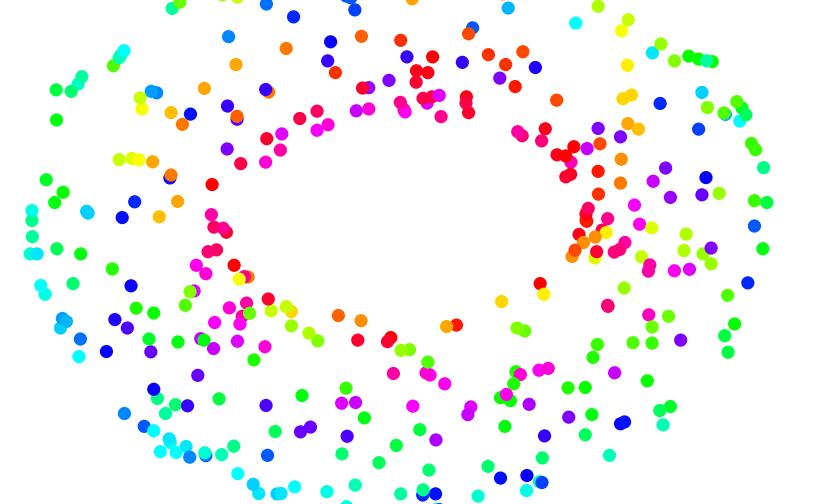
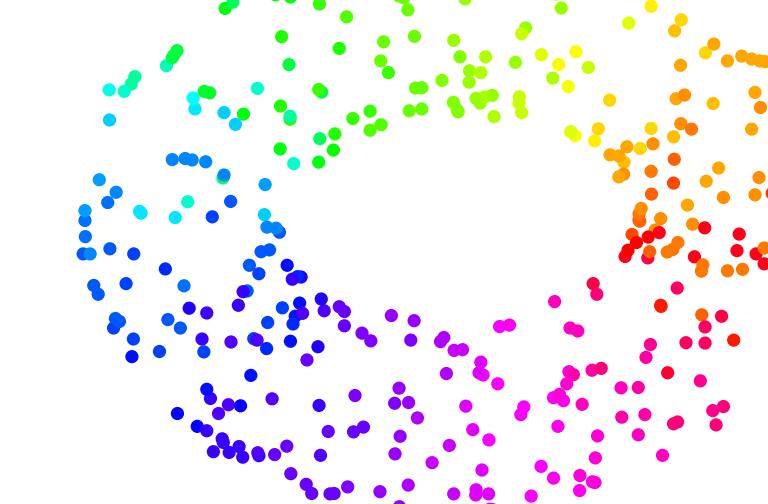
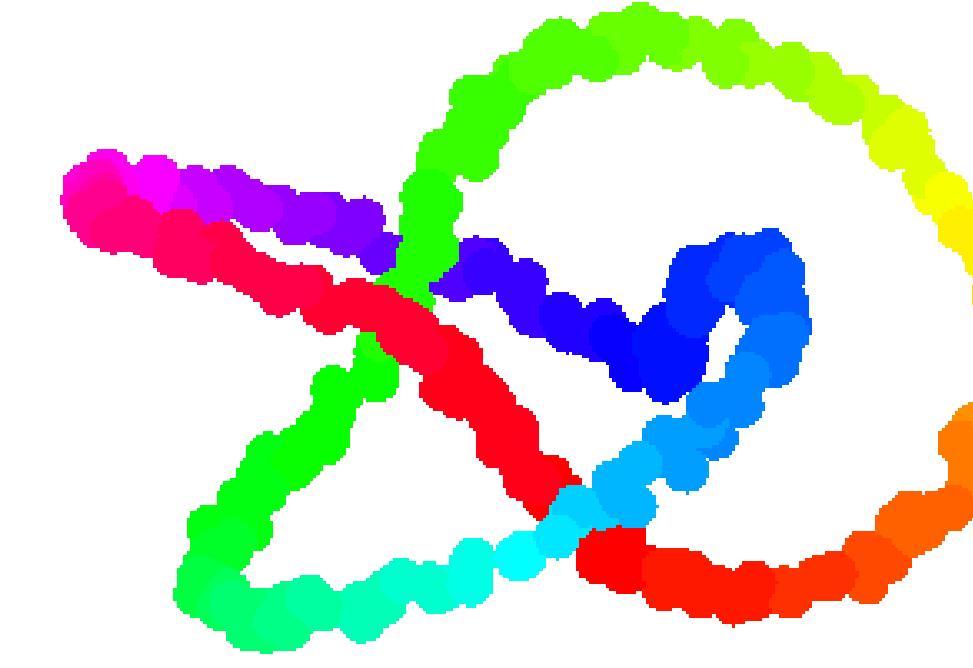
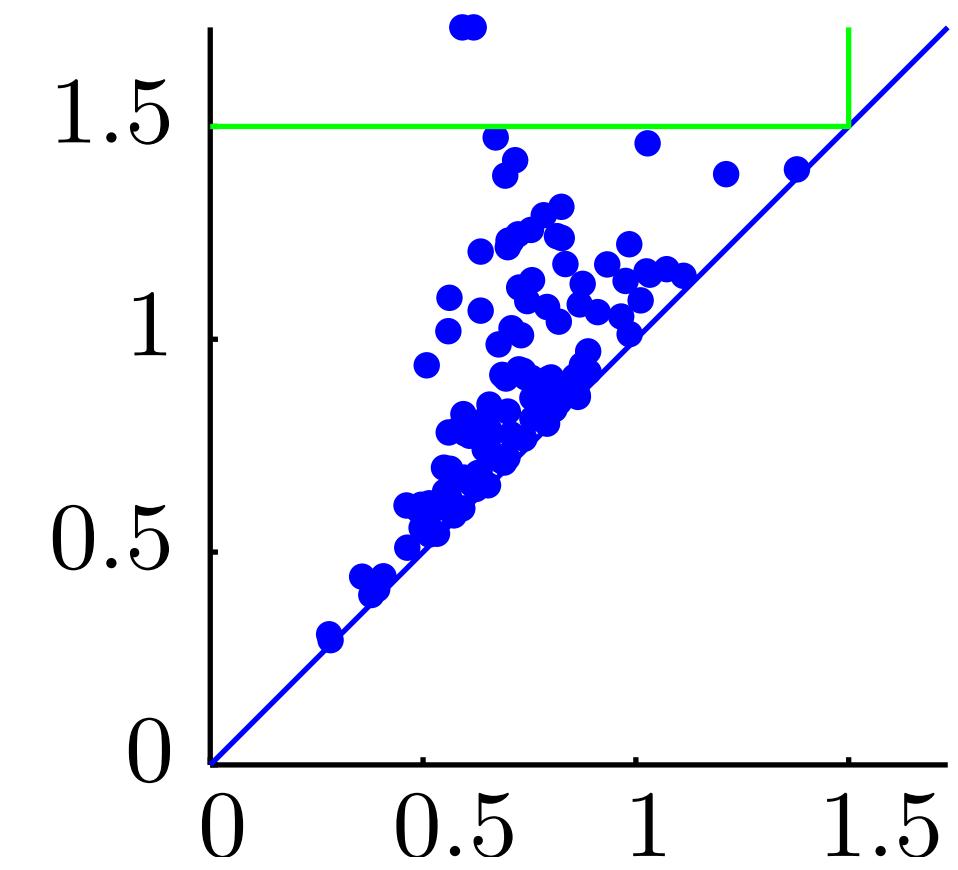
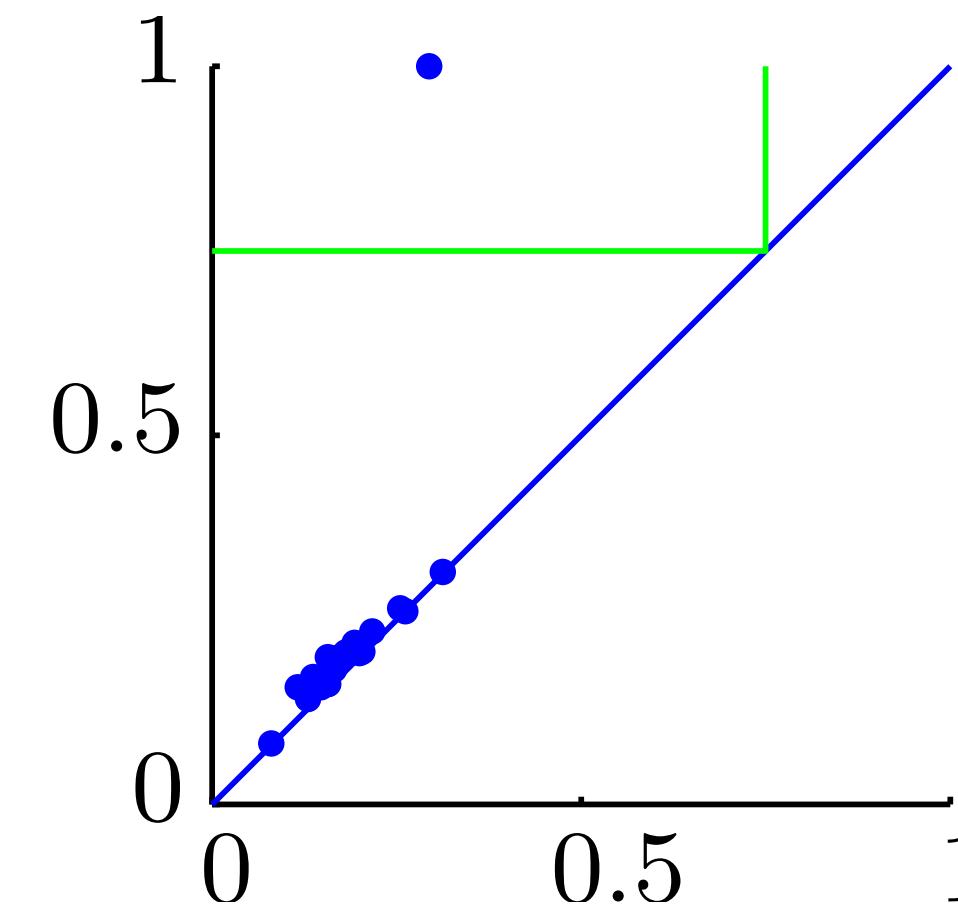
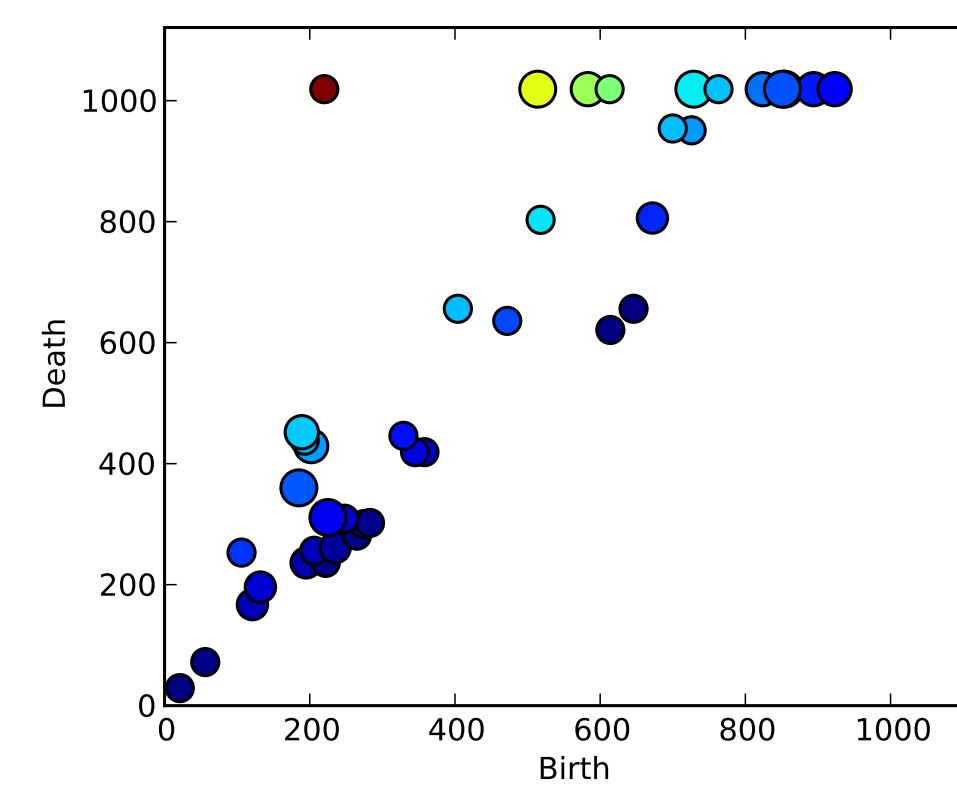
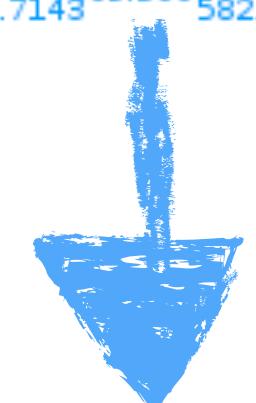
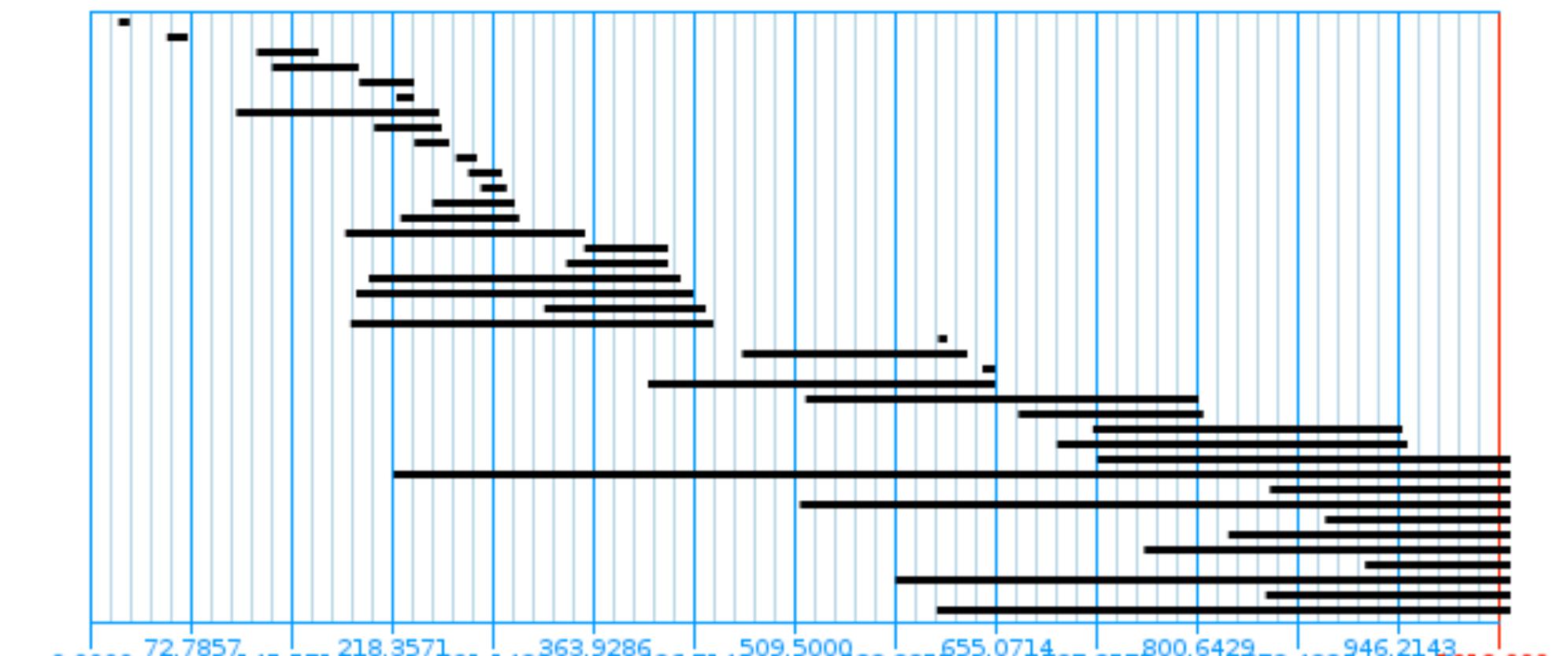
## Remark

*The degrees of the generators and the torsion degrees  $\{d_i\}$  completely determine the module*

## Barcode

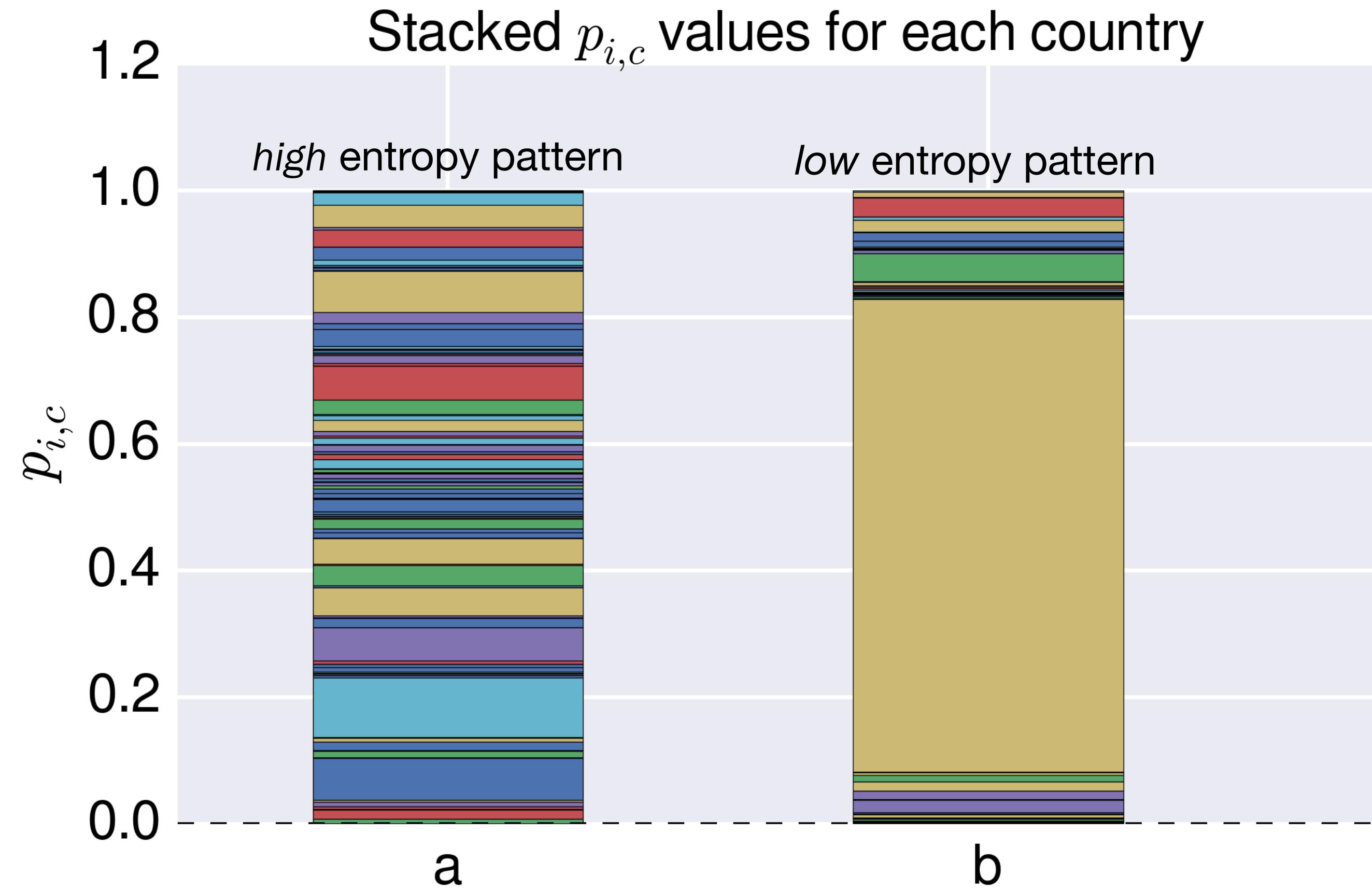
m finite intervals  $[c_i, d_i)$  and n half intervals  $[a_j, \infty)$

# Barcode to Persistence Diagram



# Notebook 04

# Phom app: ENTROPY



a

b

$$A_{i,c}(t) \quad i = \text{cell}, c = \text{country}$$

cell level

$$A_{i,c} = \sum_t A_{i,c}(t)$$

T-slices

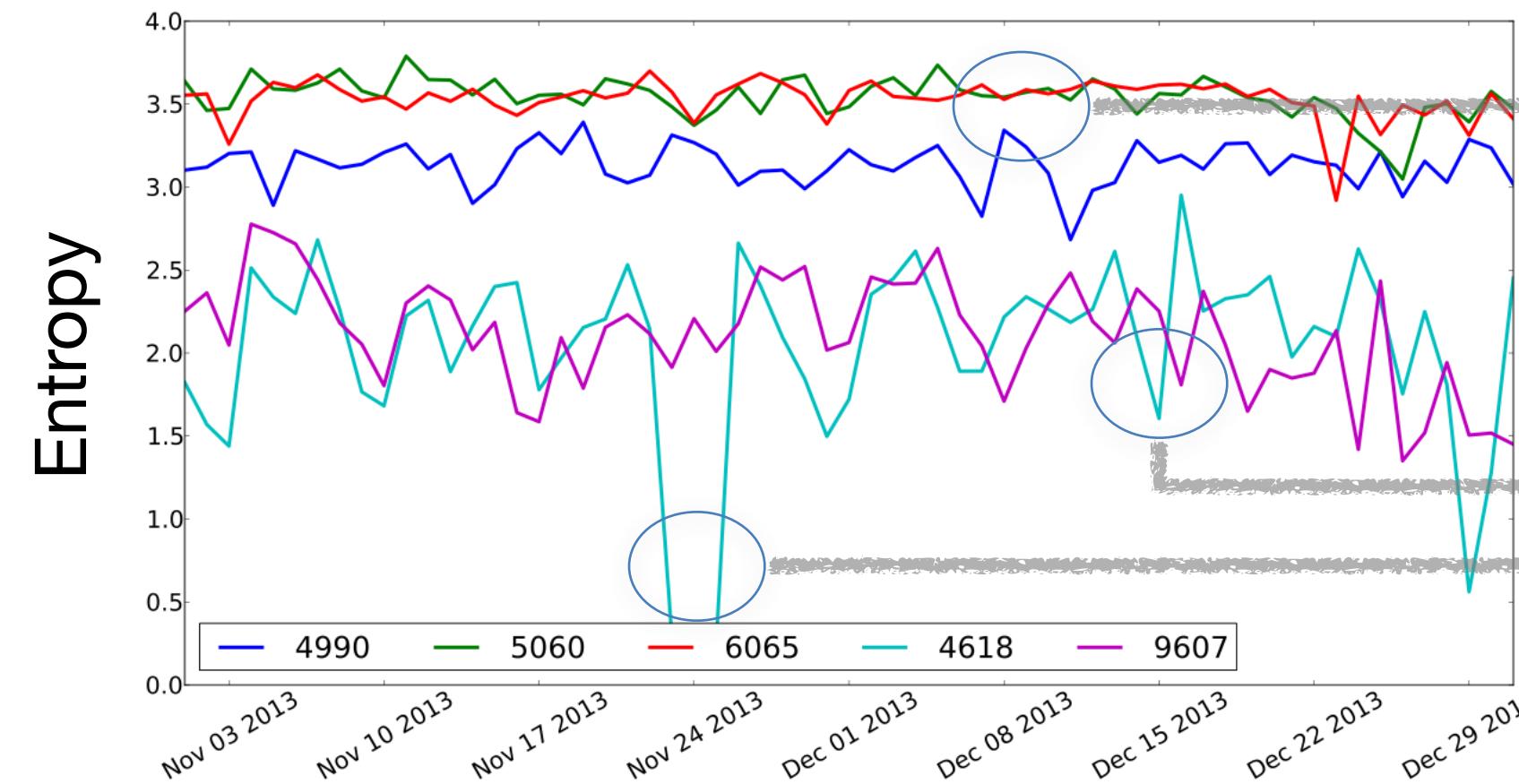
$$A_{i,c}(T) = \sum_{t \in T} A_{i,c}(t)$$

district level

$$A_{D,c} = \sum_t \sum_{i \in D} A_{i,c}(t)$$

$$\left\{ \begin{array}{l} p_{i,c} = A_{i,c} / \sum_c A_{i,c} \\ e_i = \sum_c -p_{i,c} \log(p_{i,c}) \end{array} \right.$$

# ENTROPY: single cells



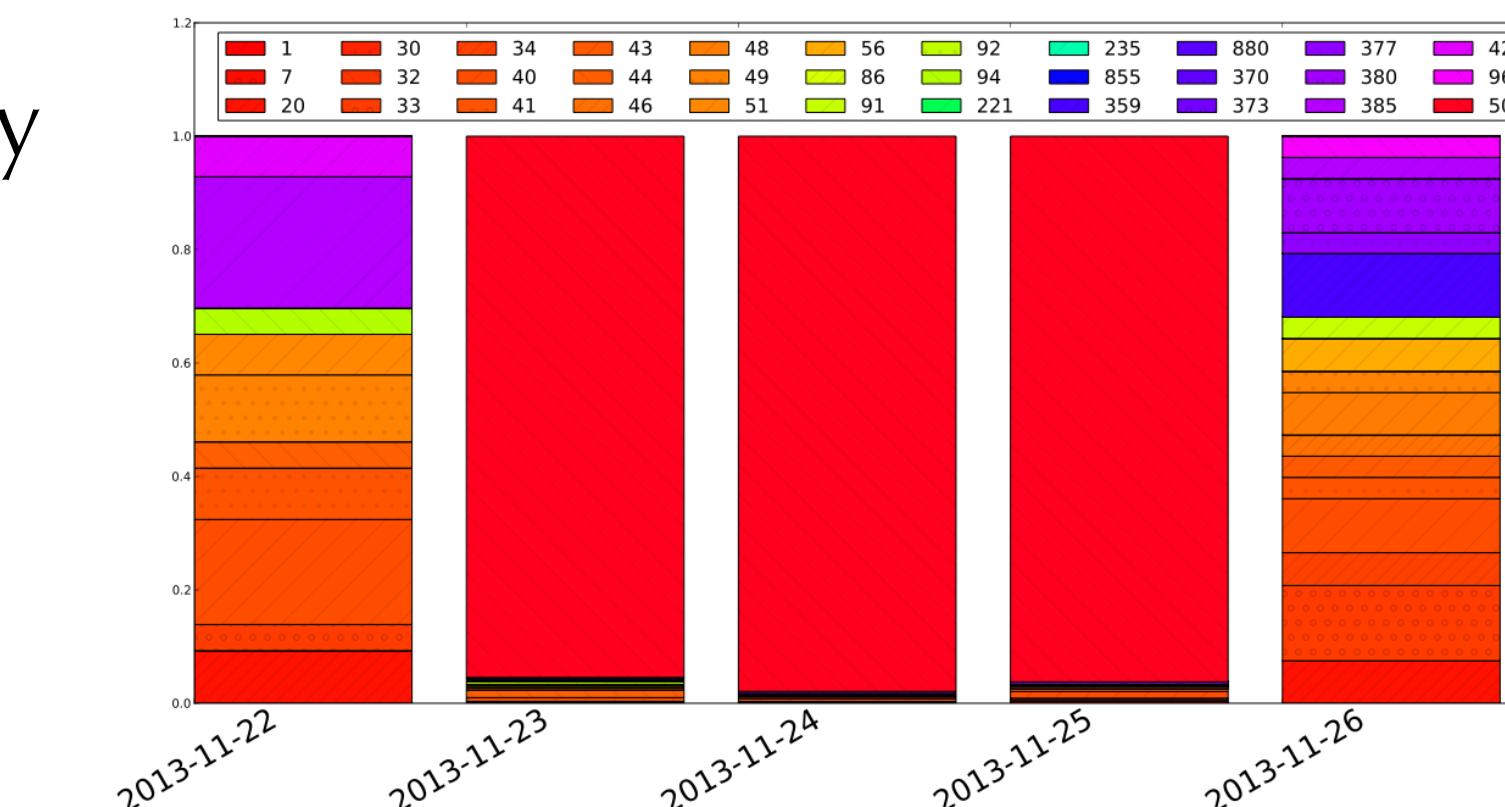
High steady entropy areas (Milano's Dome, Central railway station, Linate airport, Milano's stock exchange, etc..)

Low entropy areas with weekly patterns (University, peripheral areas, etc..)

The cell displays very low entropy and a large Haitian activity.

- Haitian Movie Festival Award?
- Big wedding?

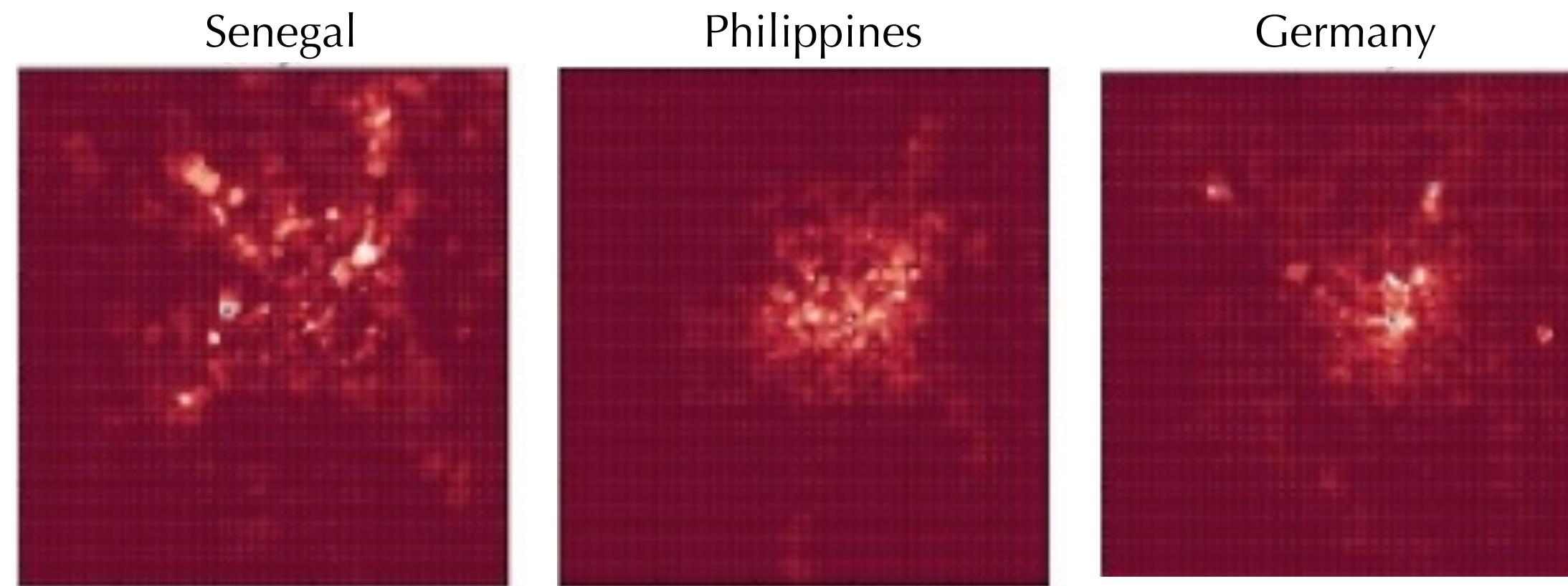
Activity stack chart for cell 4618



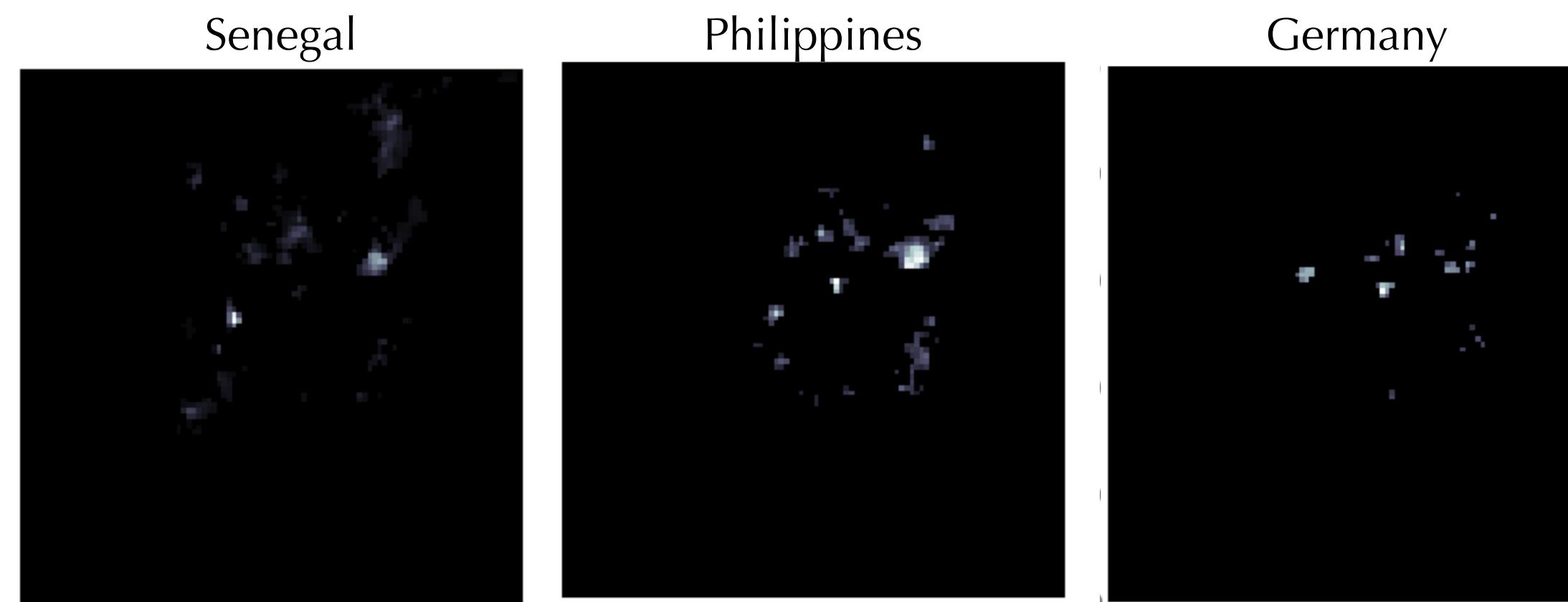
# ENTROPY: country maps

Country-specific patterns should correspond with low entropy and large country-activity.

Activity



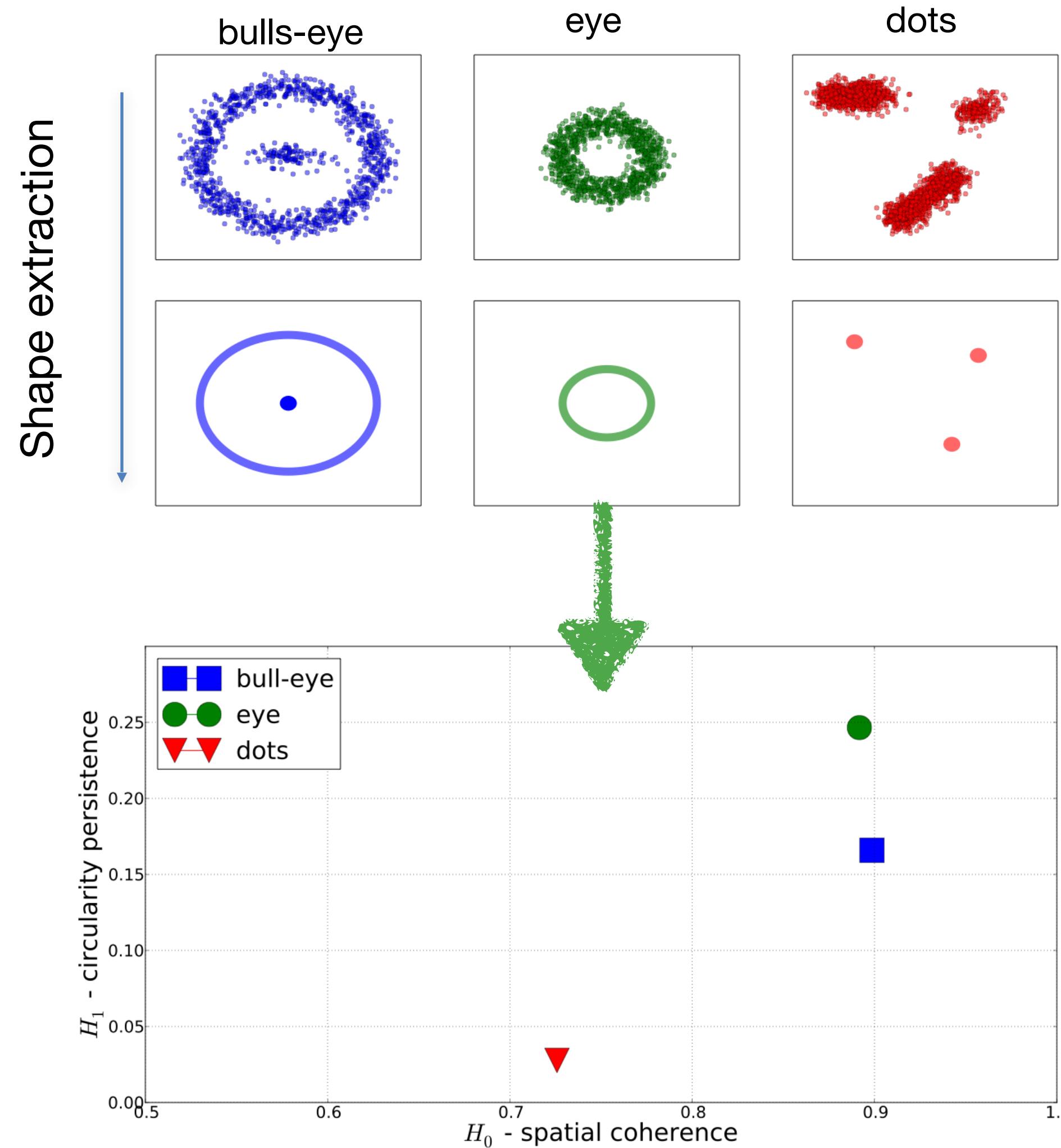
Entropy-filtered  
activity



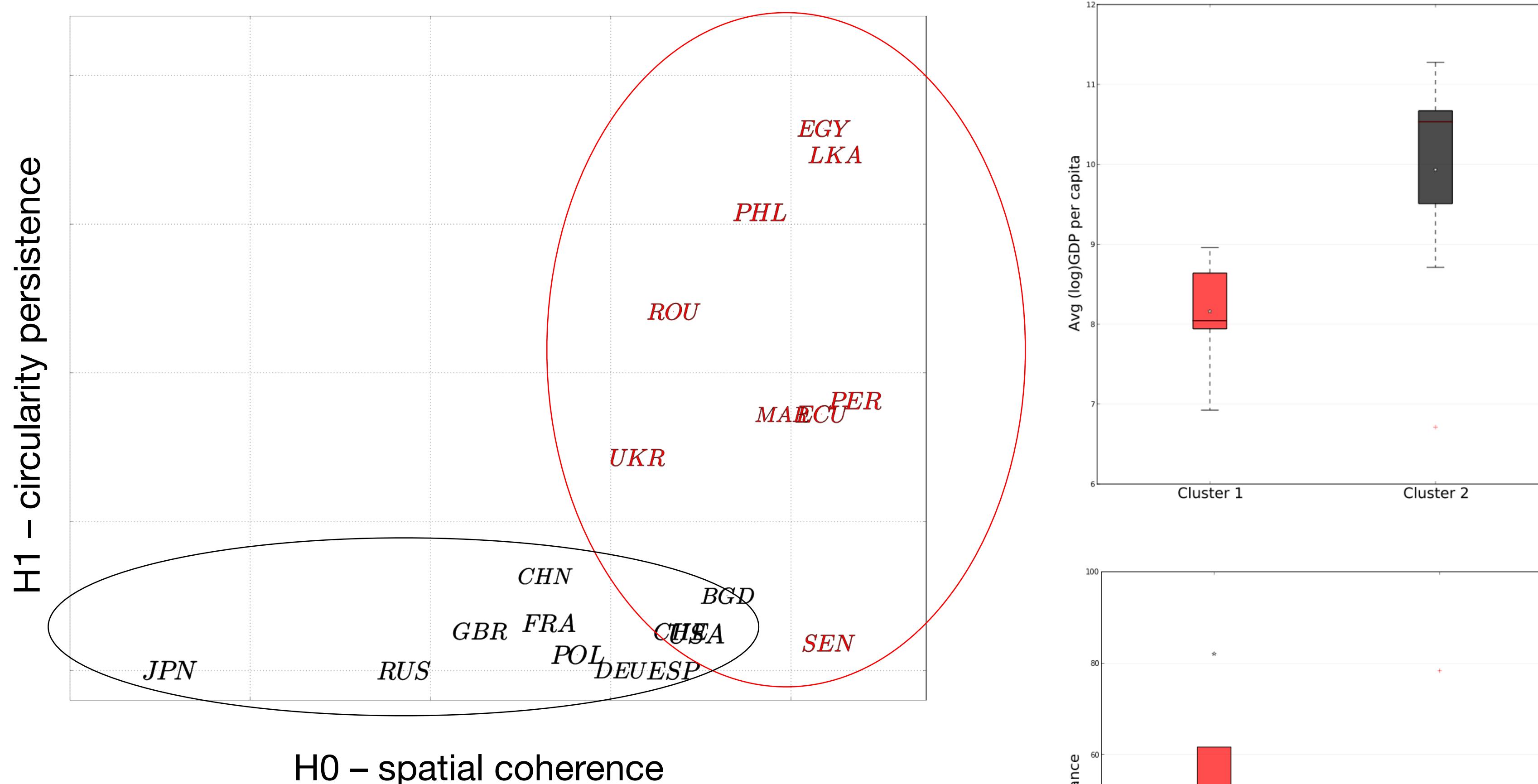
# ENTROPY: country maps

## Persistent homology: classifying spatial patterns of international communities

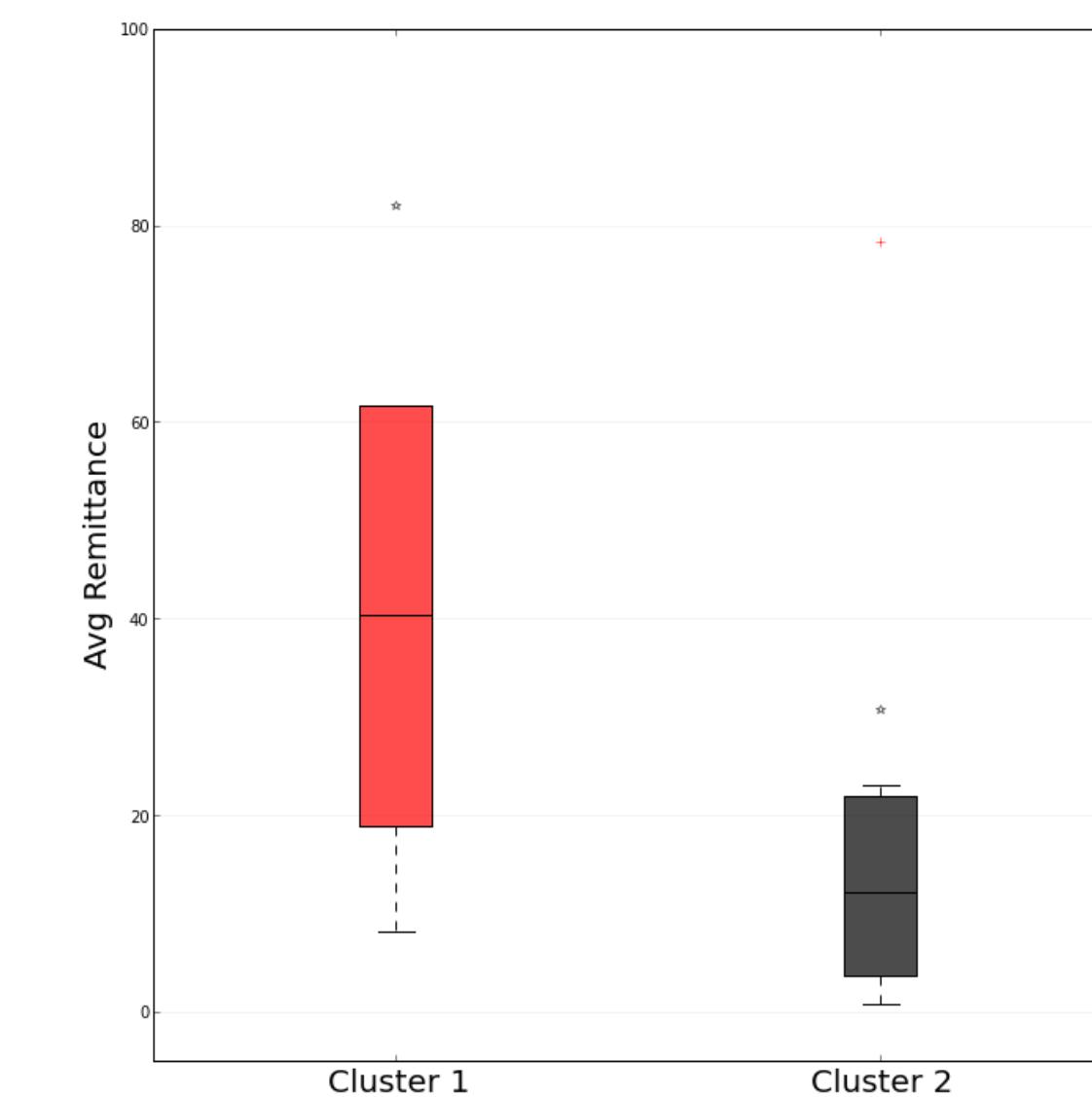
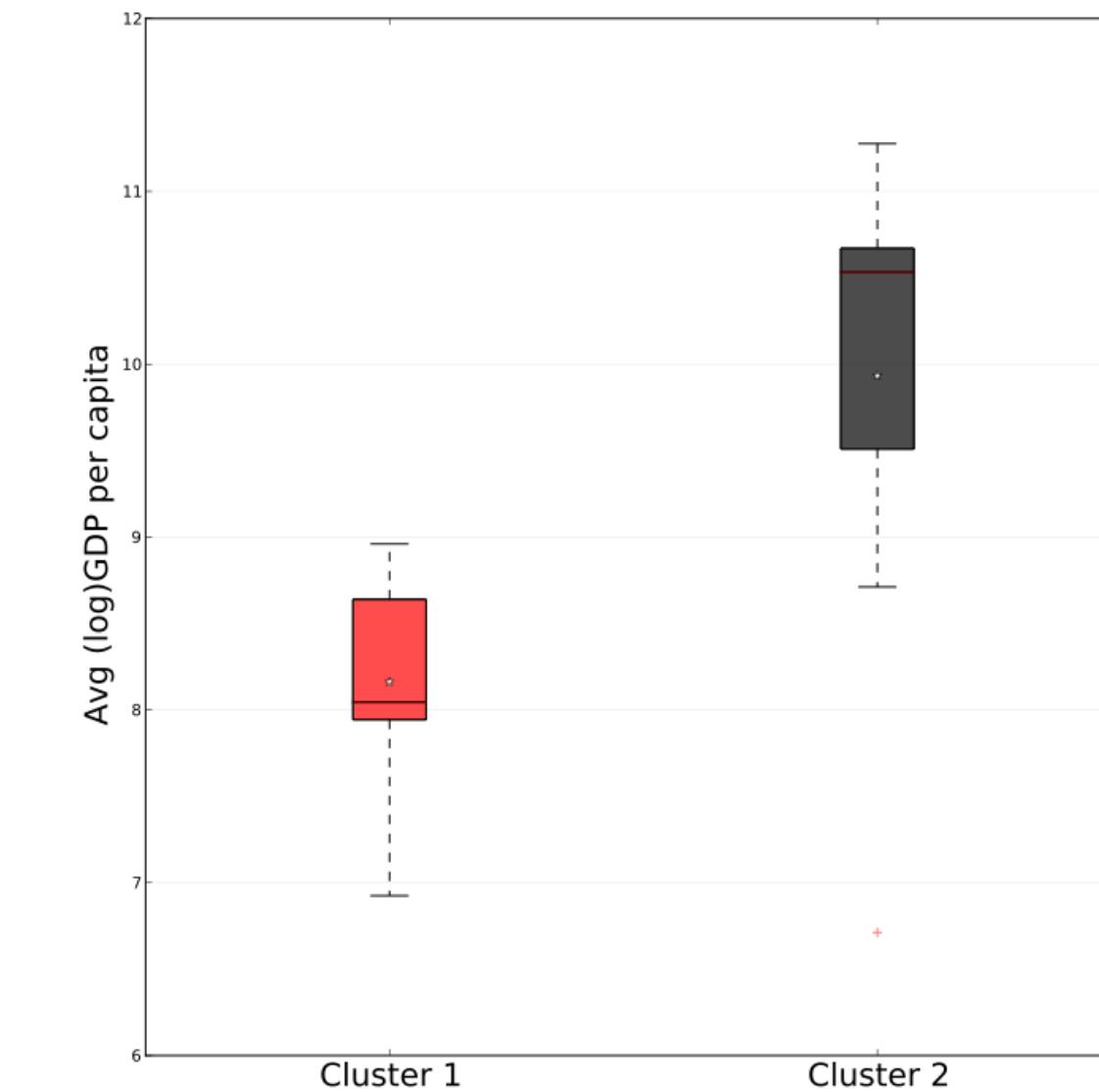
1. Recurring noisy patterns:
  1. disconnected clusters
  2. rings surrounding the city center
2. indicator based on their topological features
  1. robust to stretching and noise
  2. multiscale shape recognition method  
(Ghrist, Bull.AMS 45 (2008)).
  3. focus on the spatial coherence of country clusters:
    1. **connected components** ( $H_0$ ),
    2. **circularity features** ( $H_1$ )



# Persistent homology: entropy-activity maps



1. Clusters in 2D-scatter plot are simple k-means.
2. GDP per capita and remittance of countries in the two clusters clearly differentiate between richer migrant communities with respect people from developing countries.



# Previous work: phylogenetic trees

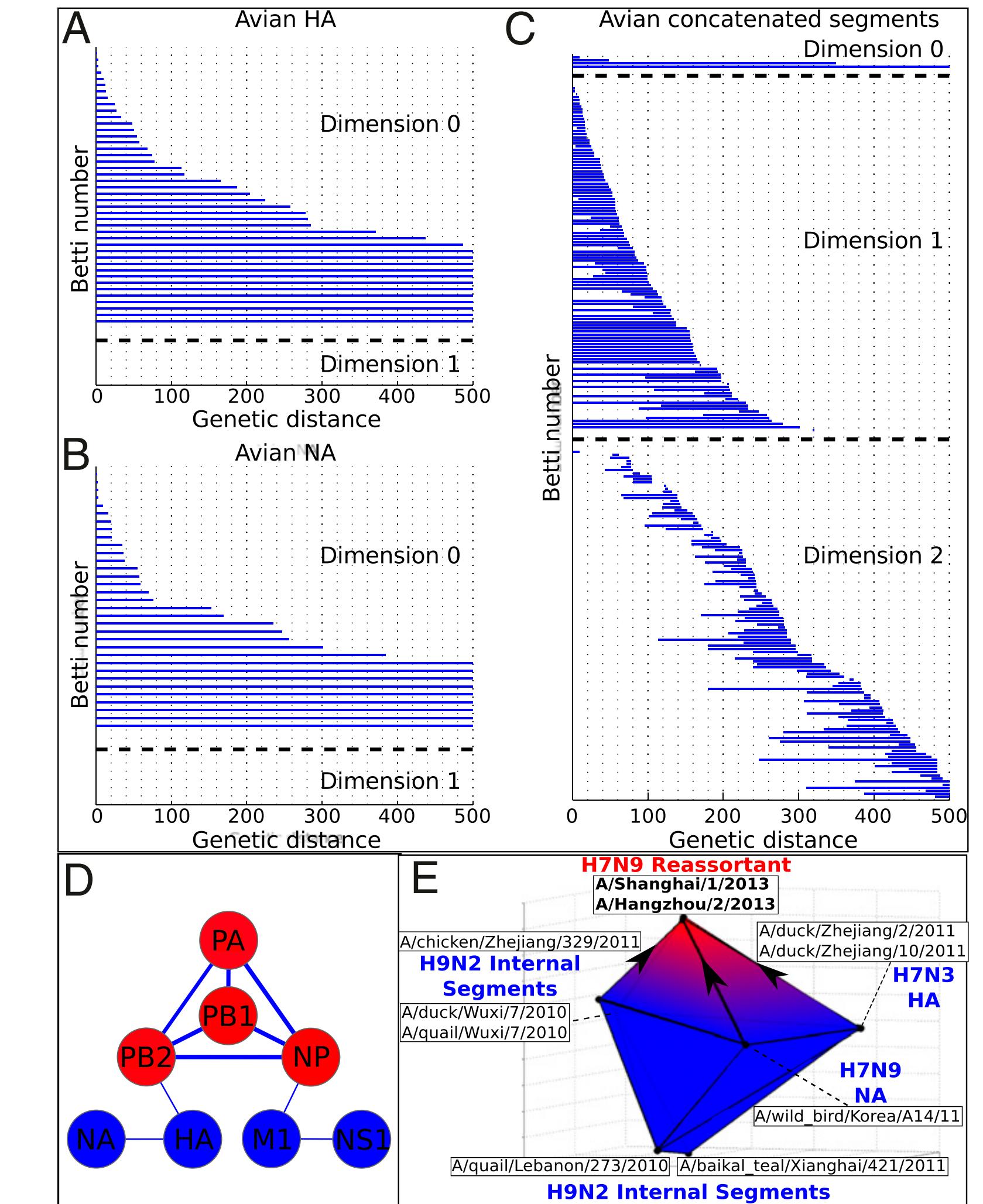
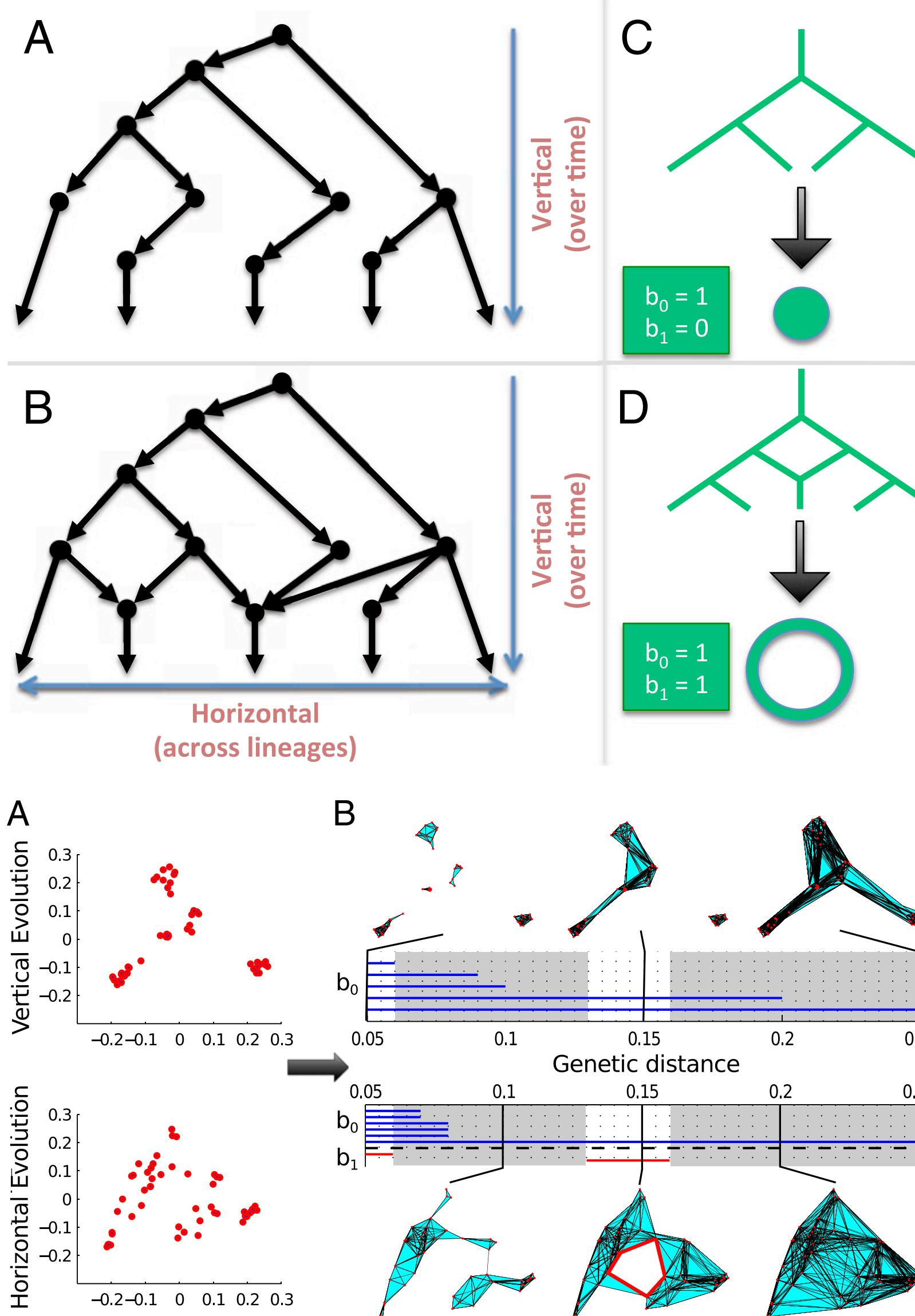
## Topology of viral evolution

Joseph Minhow Chan<sup>a,b</sup>, Gunnar Carlsson<sup>c</sup>, and Raul Rabadan<sup>a,b,d,1</sup>

<sup>a</sup>Center for Computational Biology and Bioinformatics and Departments of <sup>b</sup>Biomedical Informatics and <sup>d</sup>Systems Biology, Columbia University College of Physicians and Surgeons, New York, NY 10032; and <sup>c</sup>Department of Mathematics, Stanford University, Stanford, CA 94305

Edited\* by Arnold J. Levine, Institute for Advanced Study, Princeton, NJ, and approved October 11, 2013 (received for review July 18, 2013)

AS

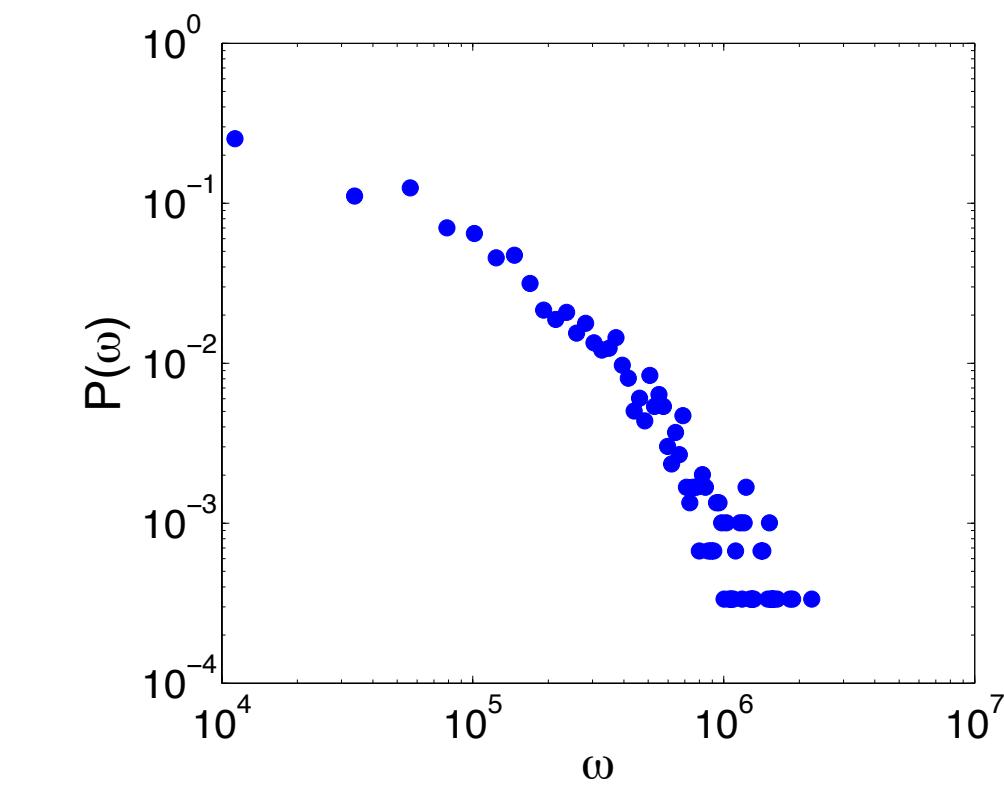
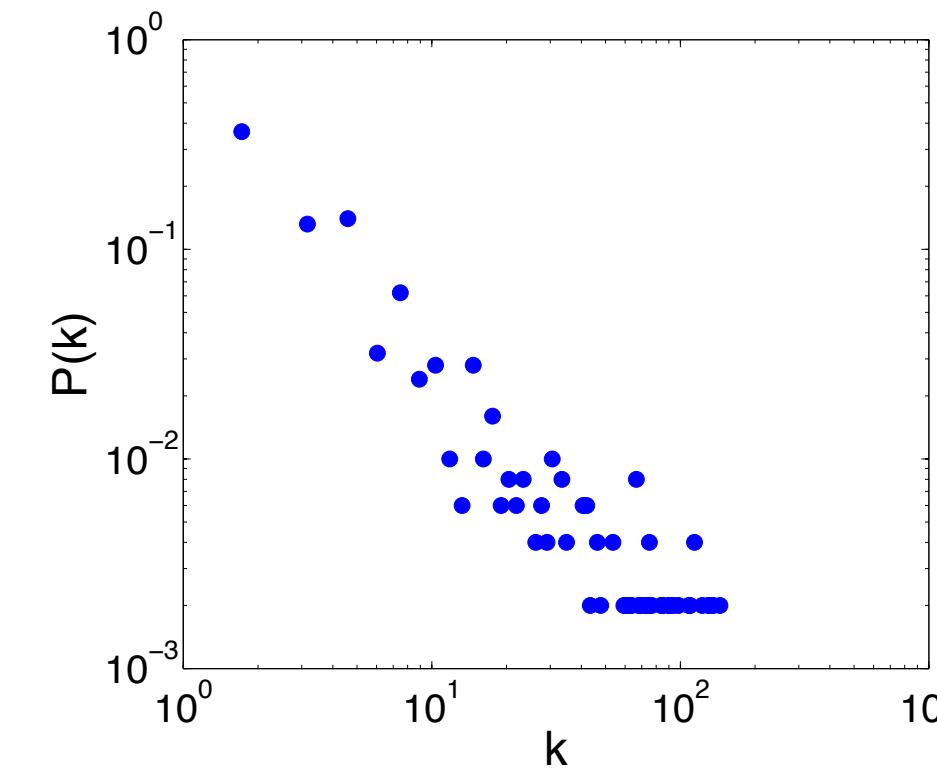
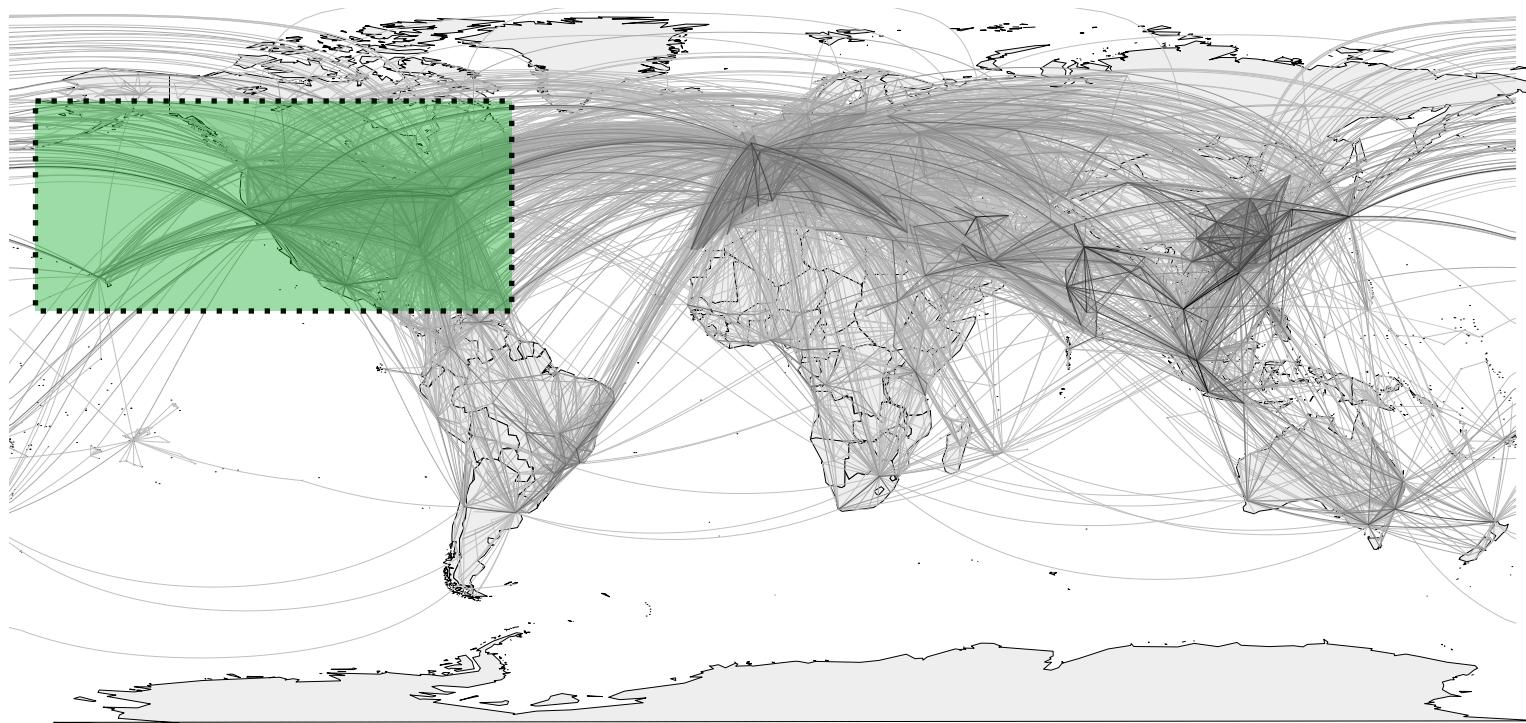


What about  
networks?

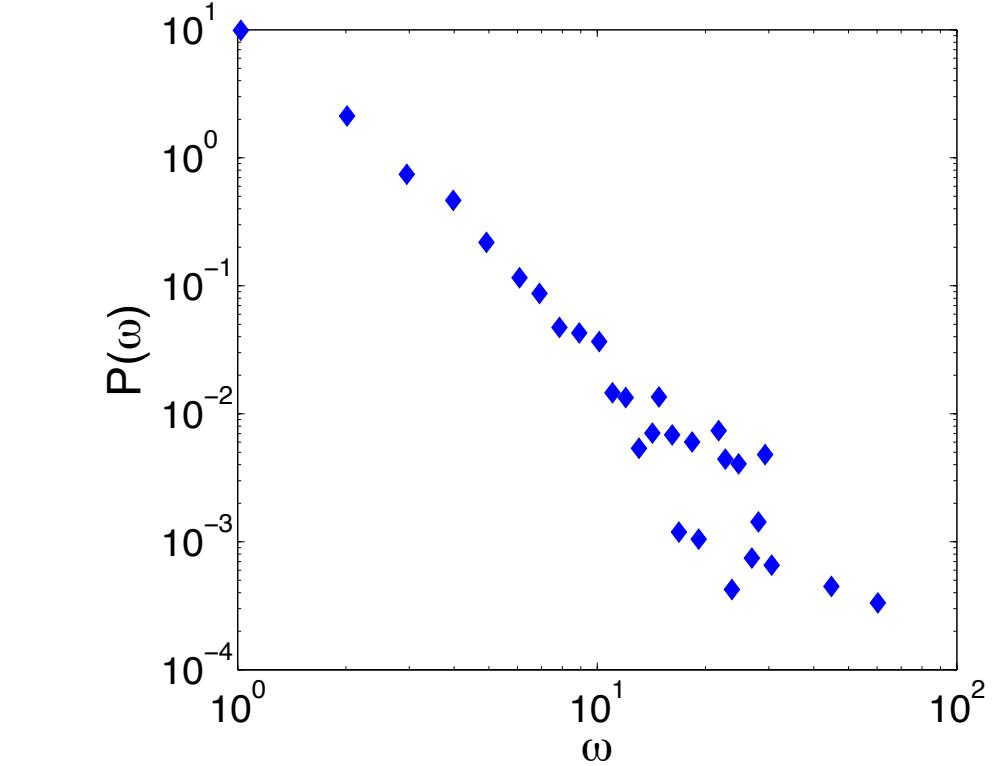
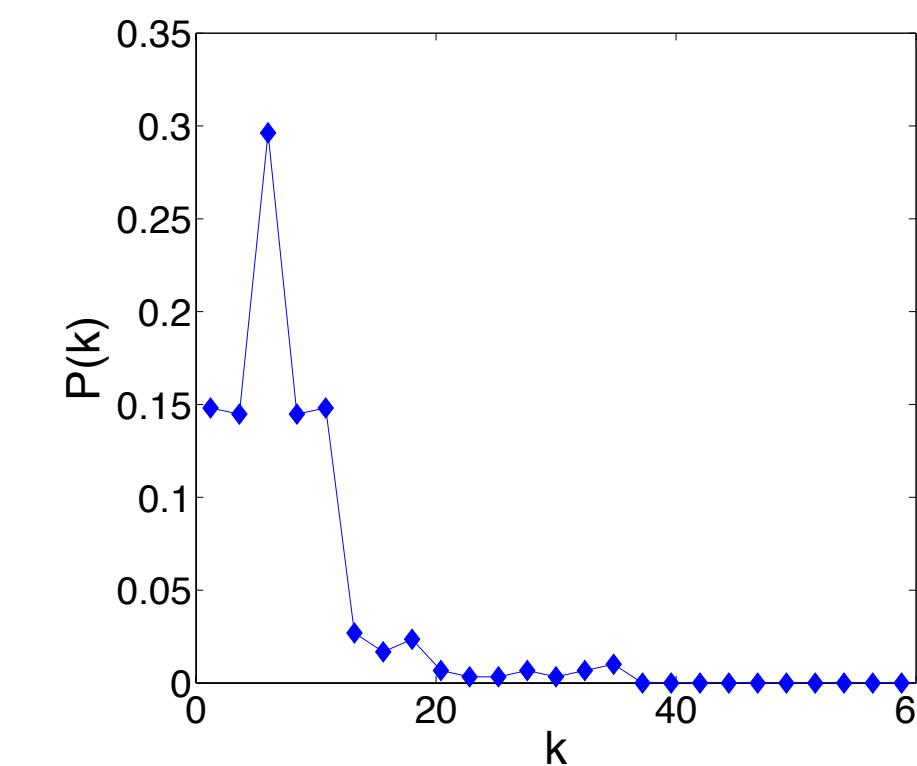
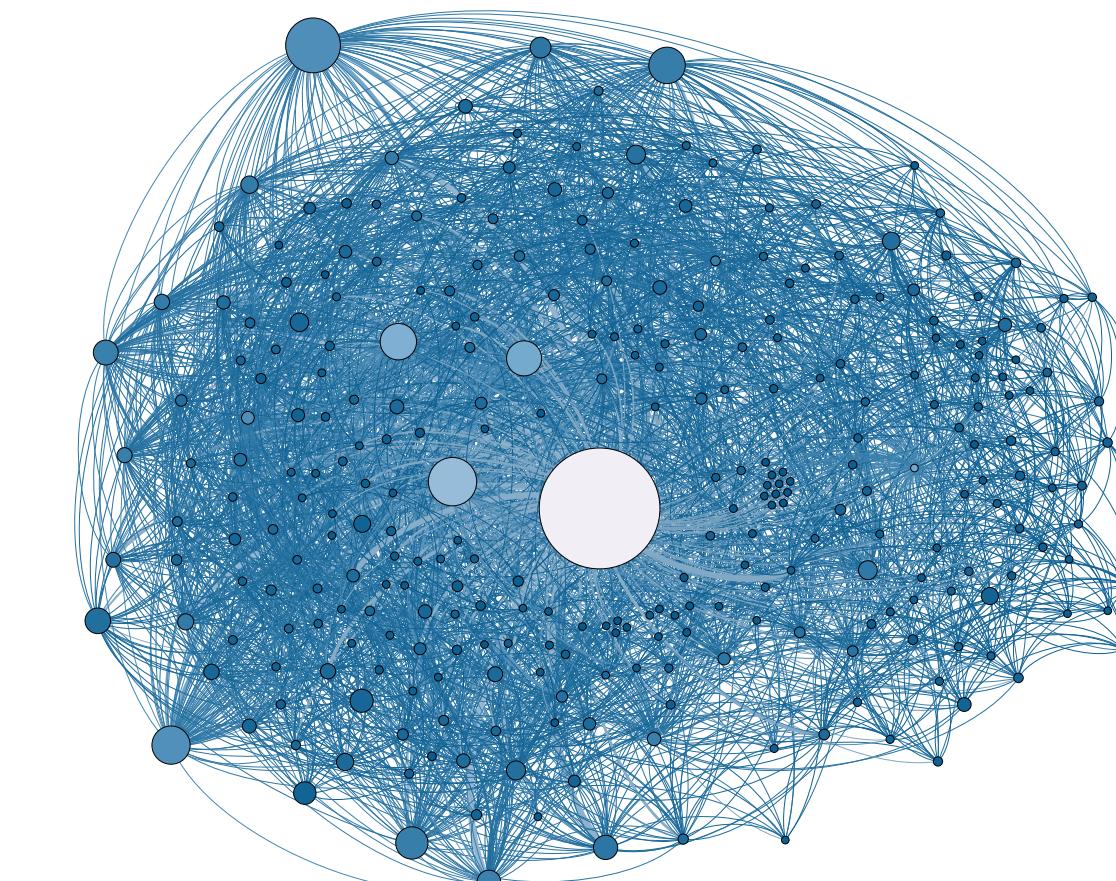
# Metrical persistent homology

No metric space, but one can define a few: shortest path,  
commute time  
other Kernel distances..

US Top-500 airports network

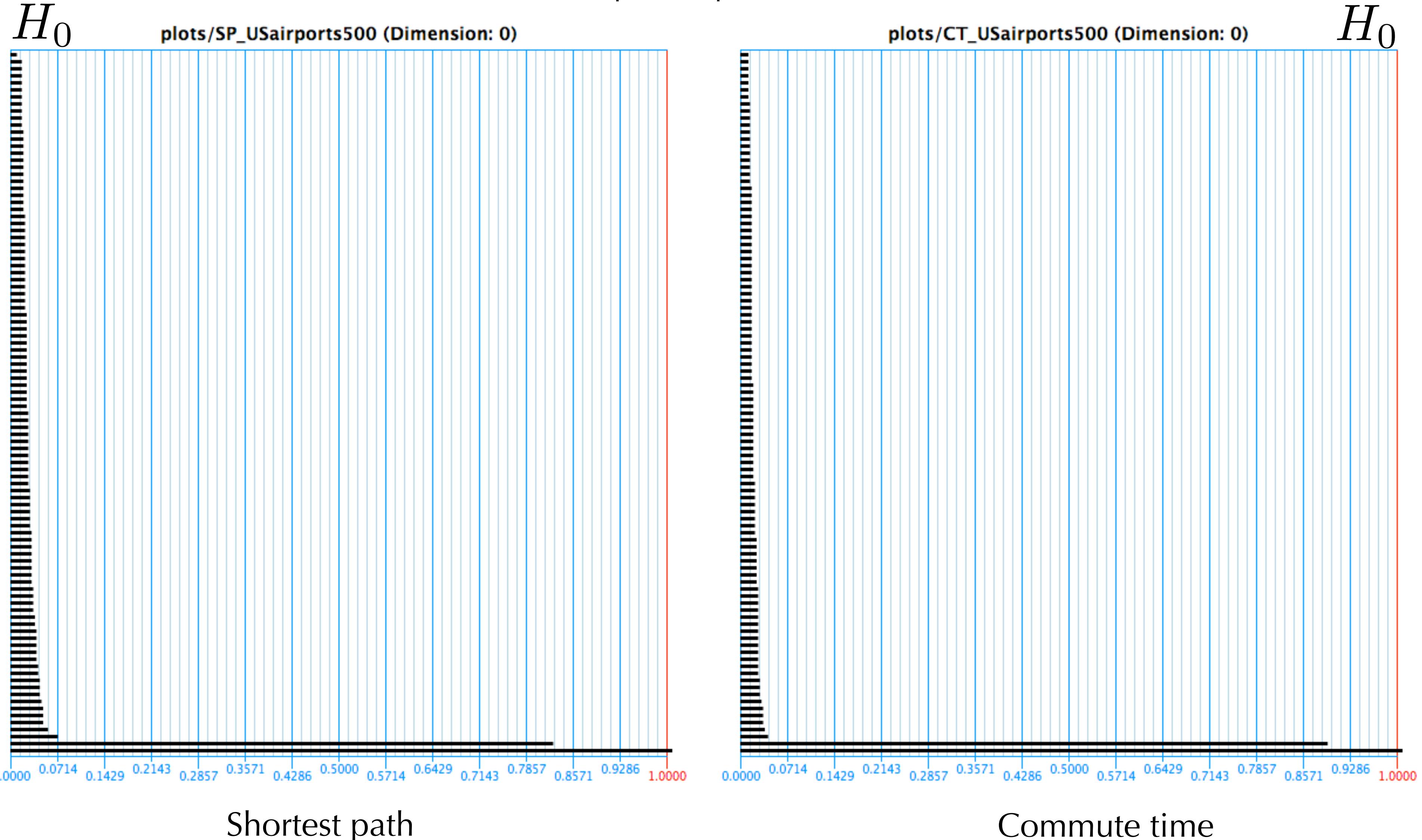


C. Elegans neural network

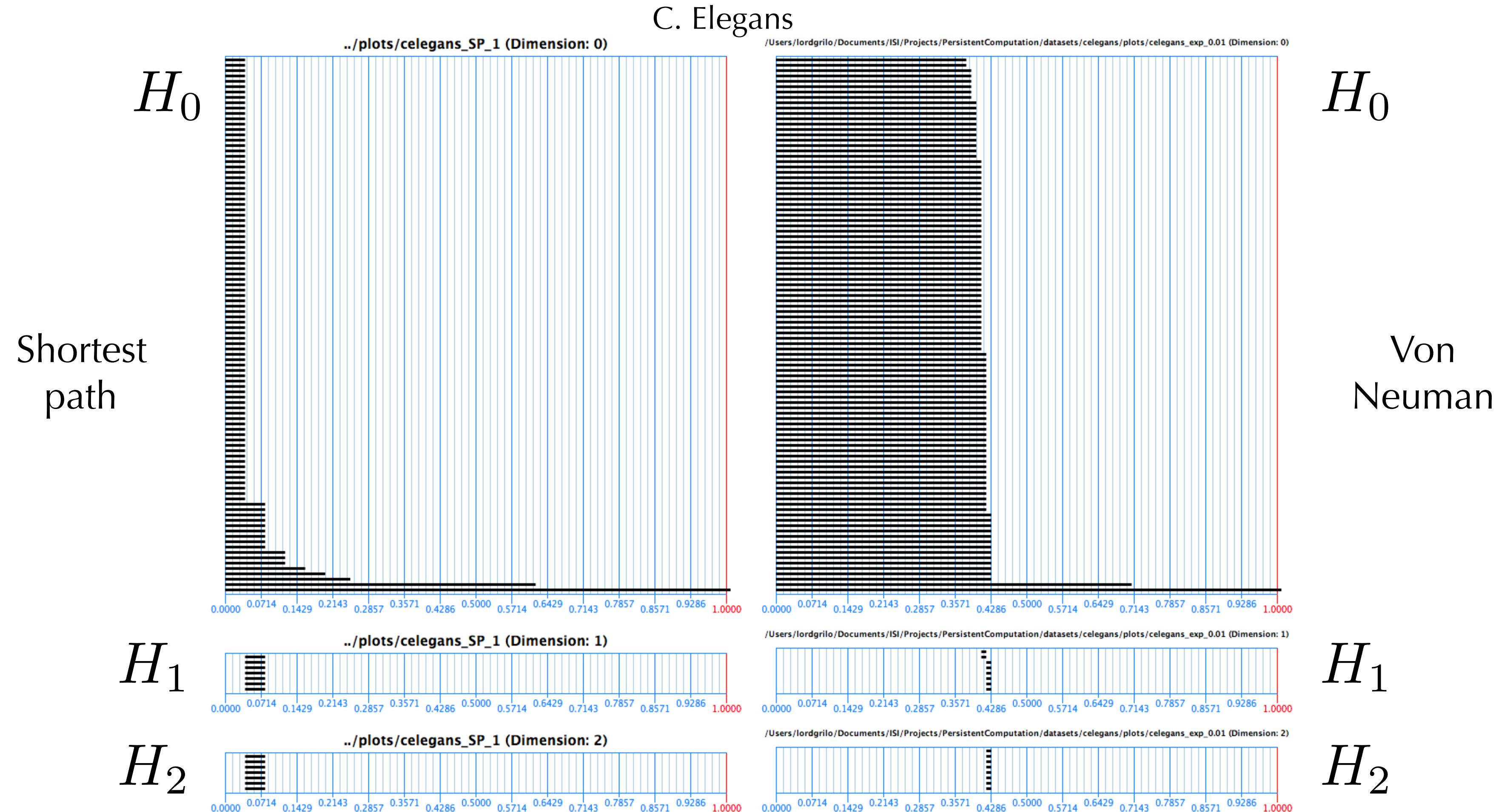


# Metrical persistent homology II

US Top500 airports network



# Metrical persistent homology II<sup>bis</sup>

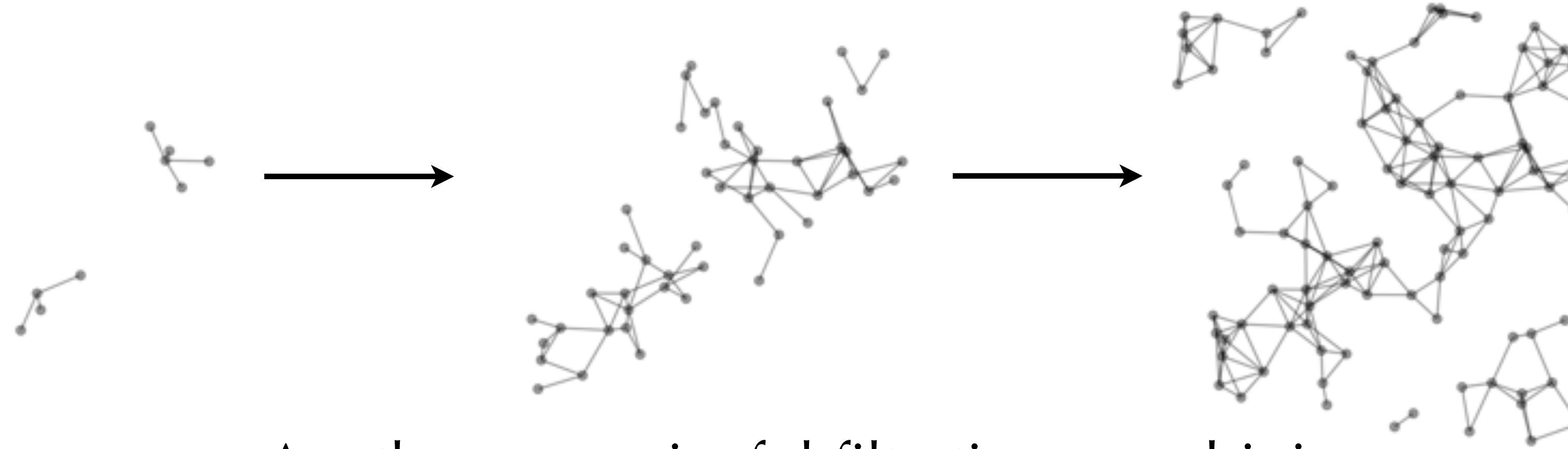


1. Different metrics ~ same information
2. Cycles appear at emergence of GCC
3. No need of homology to do this!!

# Metrical persistent homology III

Metrics do not convey much information.

What now? The important ingredient is the **filtration**.



Are there meaningful filtrations combining:

**Network “linking patterns”**

&

**Weight structure**

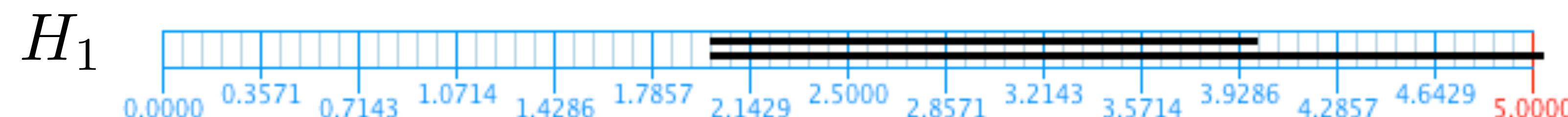
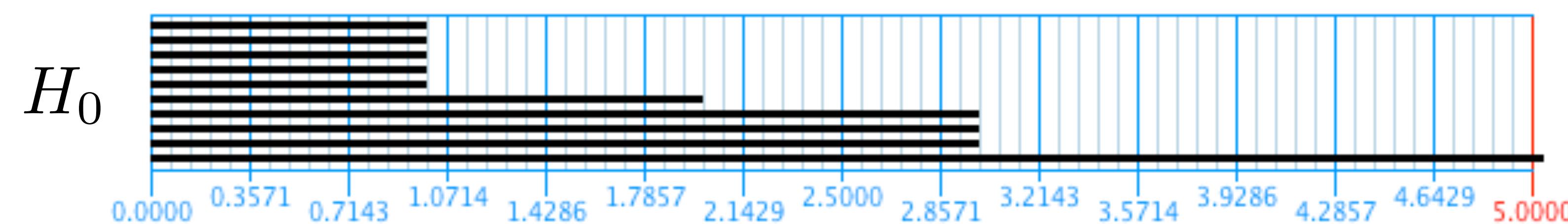
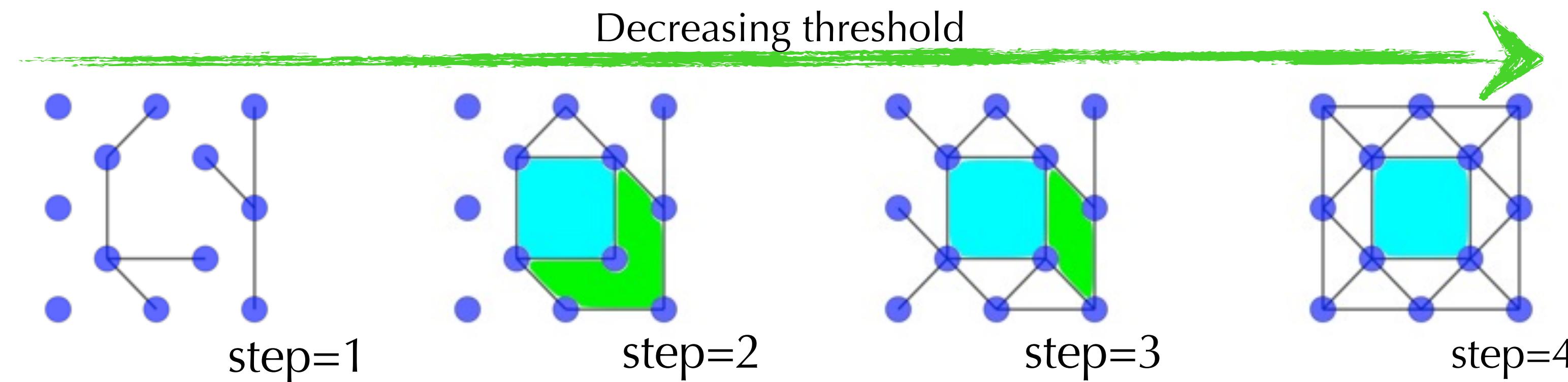
?

## Remark

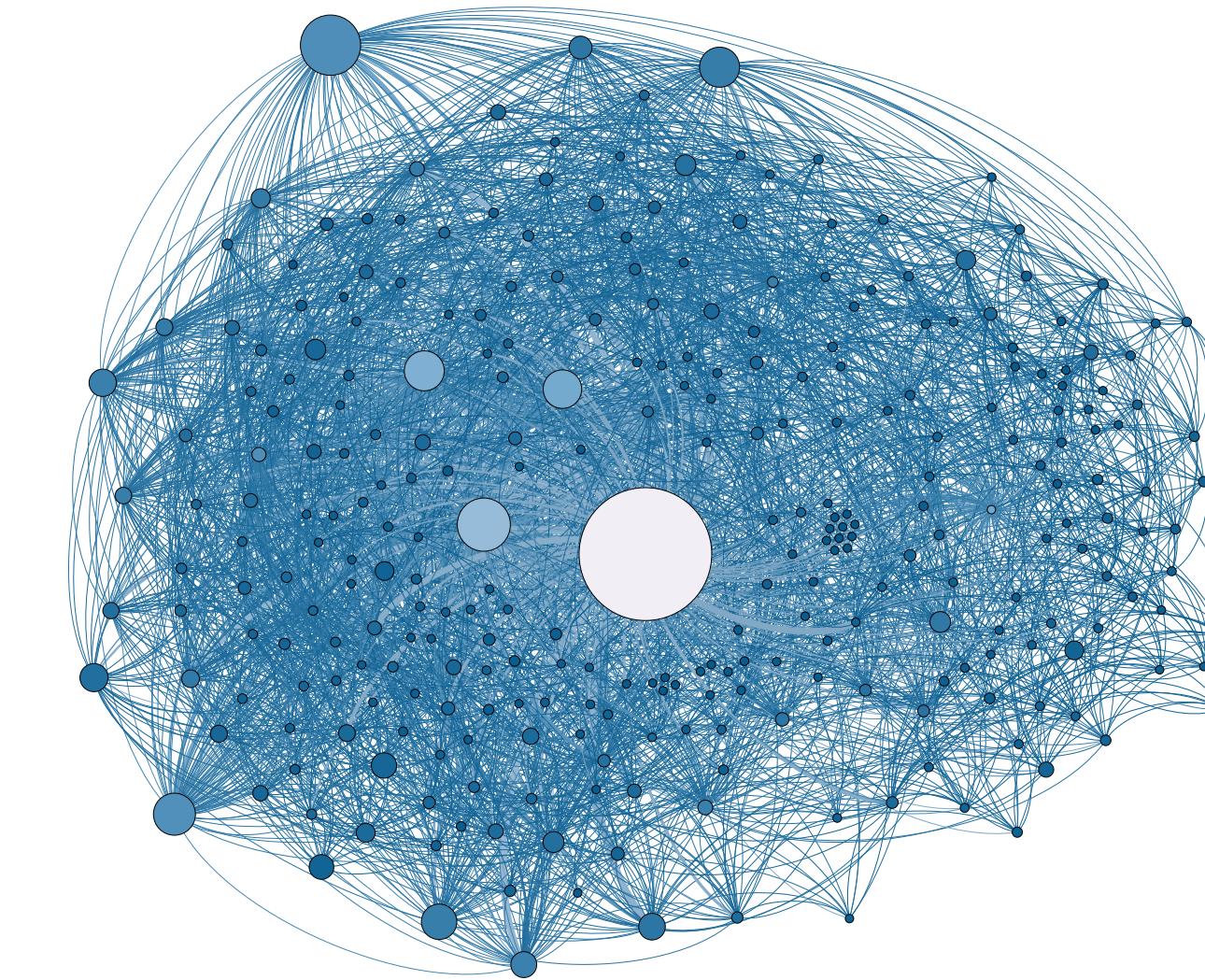
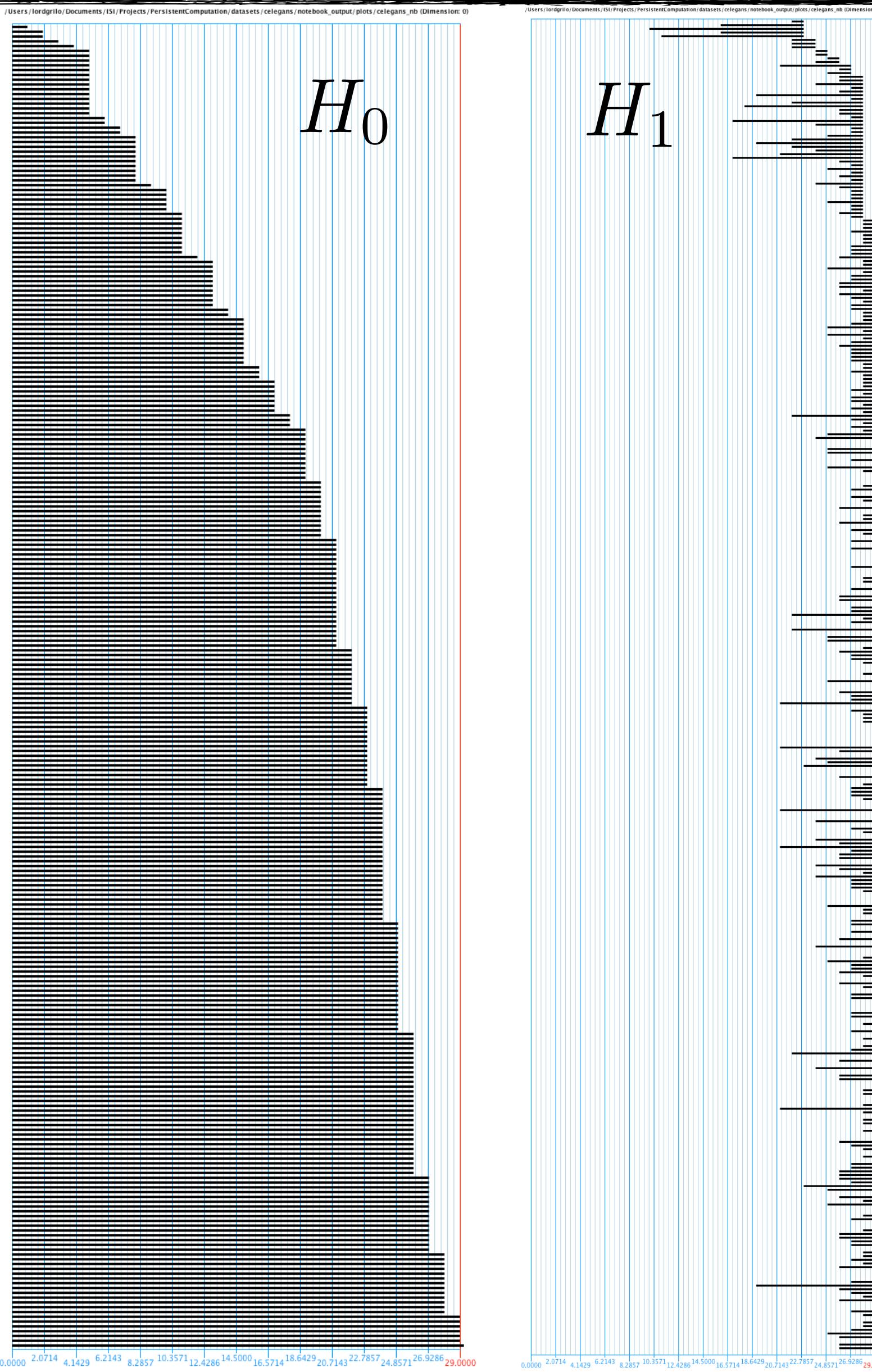
*The Rips complex of points in a metric space is the clique complex of a metric graph.*

# Weighted clique filtration I

- 1) Given a real network  $G=(V,E)$ , consider another one with the same number of nodes  $N=|V|$  and no edges.
- 2) Start adding edges from  $G$ , in decreasing order of weight.
- 3) At each step, calculate the associated clique complex.

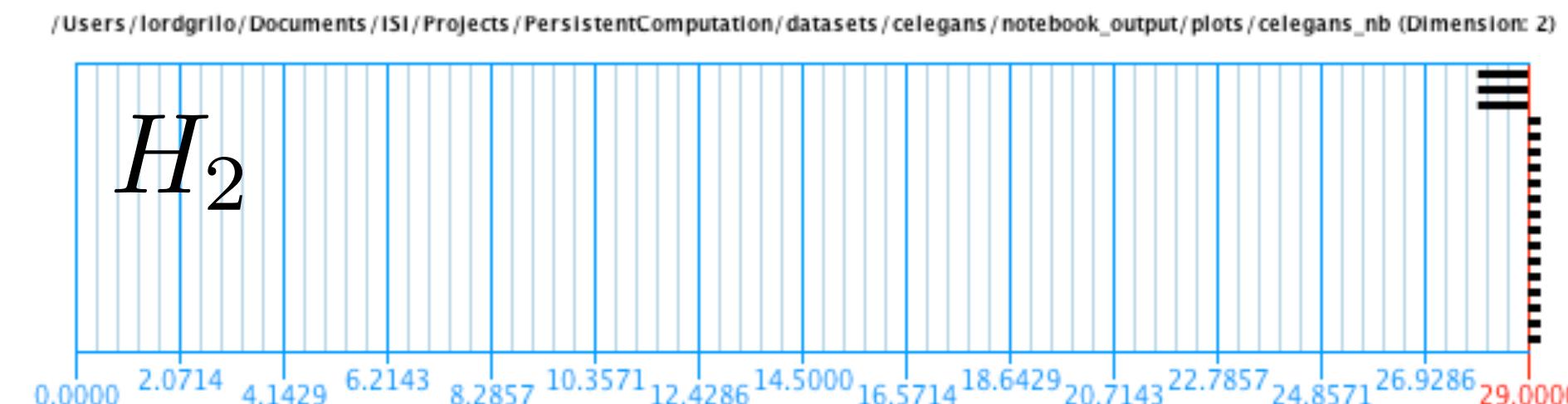


# Weighted clique filtration II<sup>bis</sup>

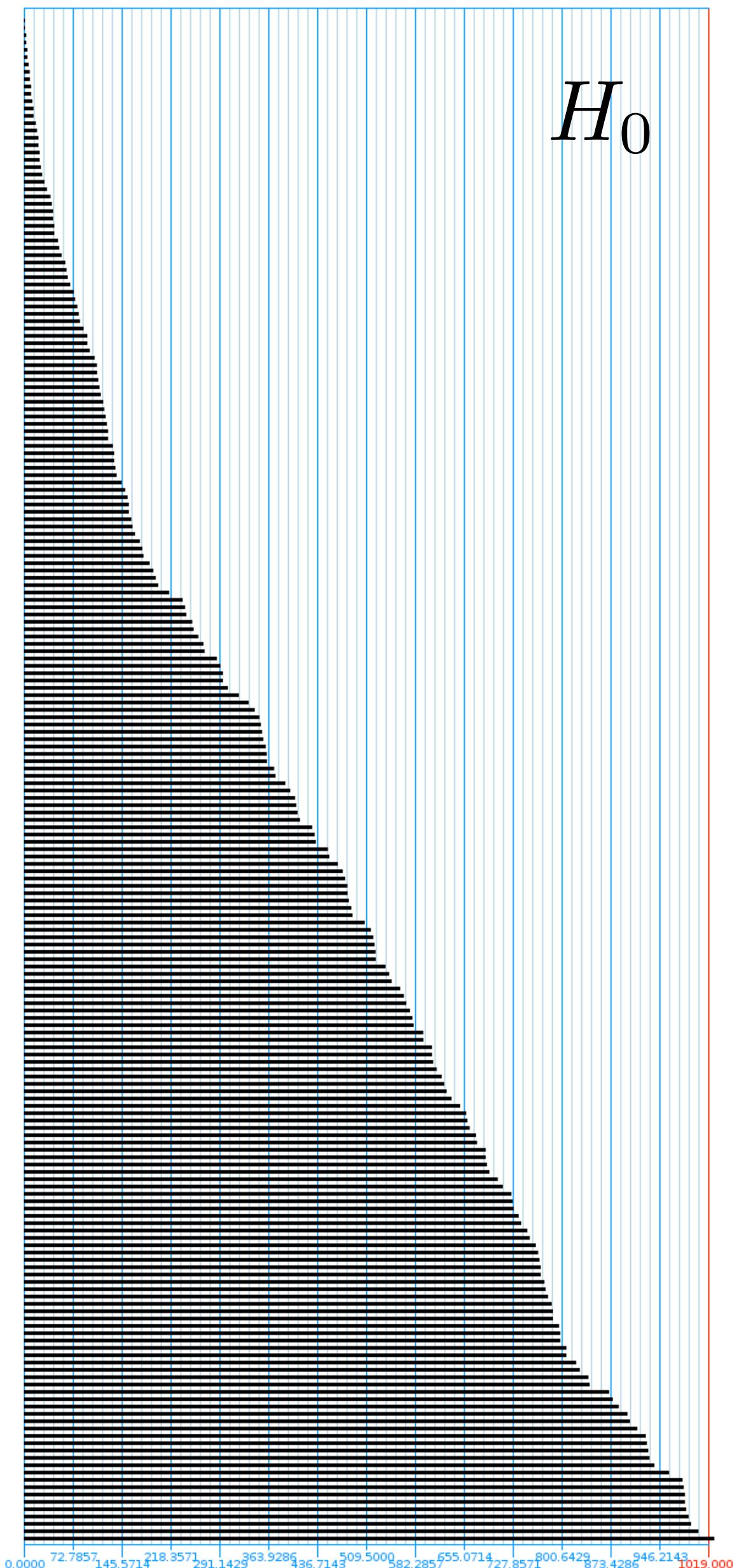


Much more structure appears!

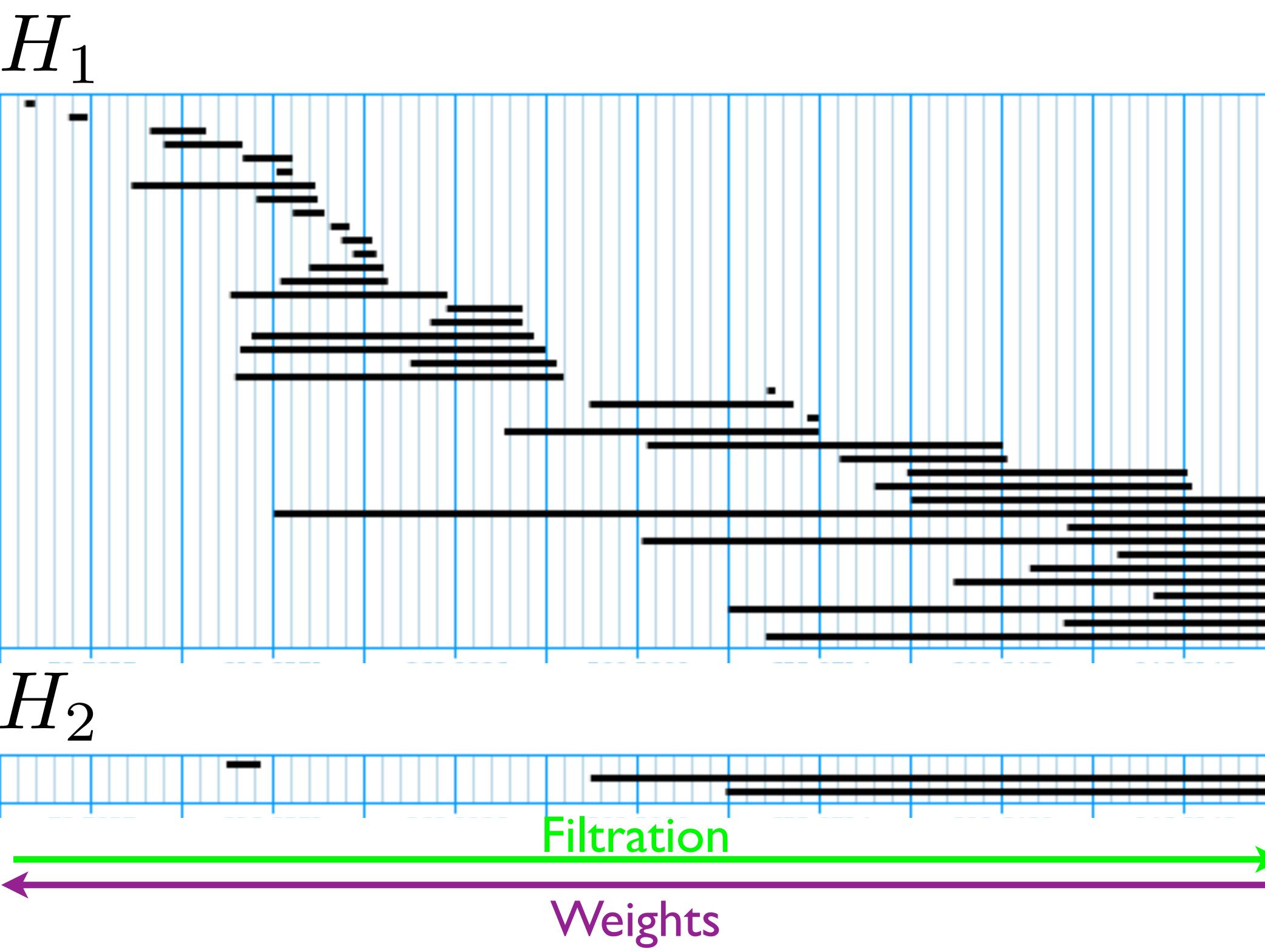
- Conn. comps. smoothly coalesce.
- A large number of loops.
- Even some “3D”-holes for low weights.



# Weighted clique filtration II



- $H_0$  Connected components
- $H_1$  1D Cycles (2-simplexes)
- $H_2$  3D holes (3-simplexes)

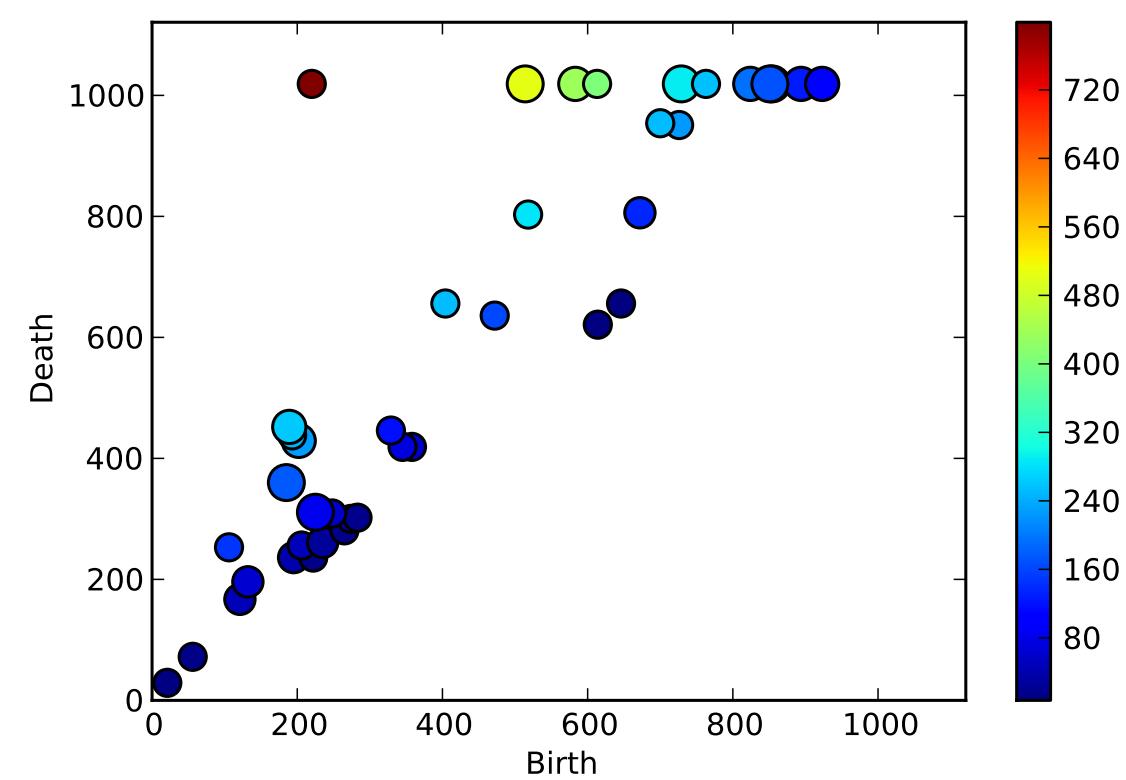


# Datasets I

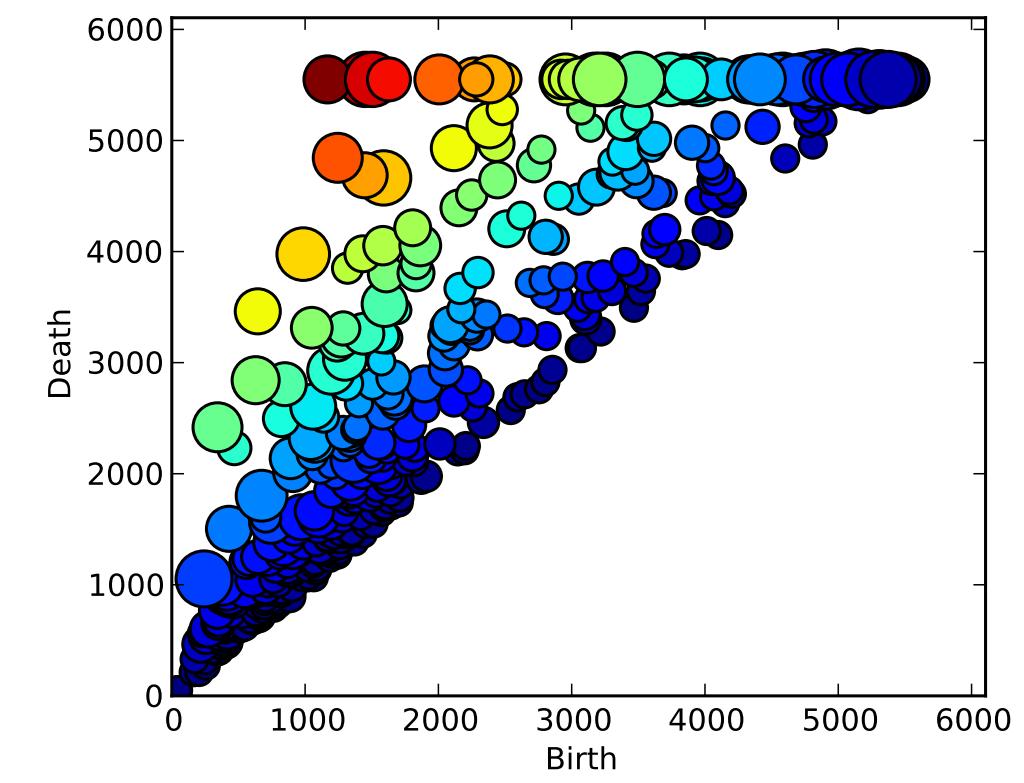
Focus on  $H_1$  map each generator to a point in the plane:

Cycle  (Birth,Death)

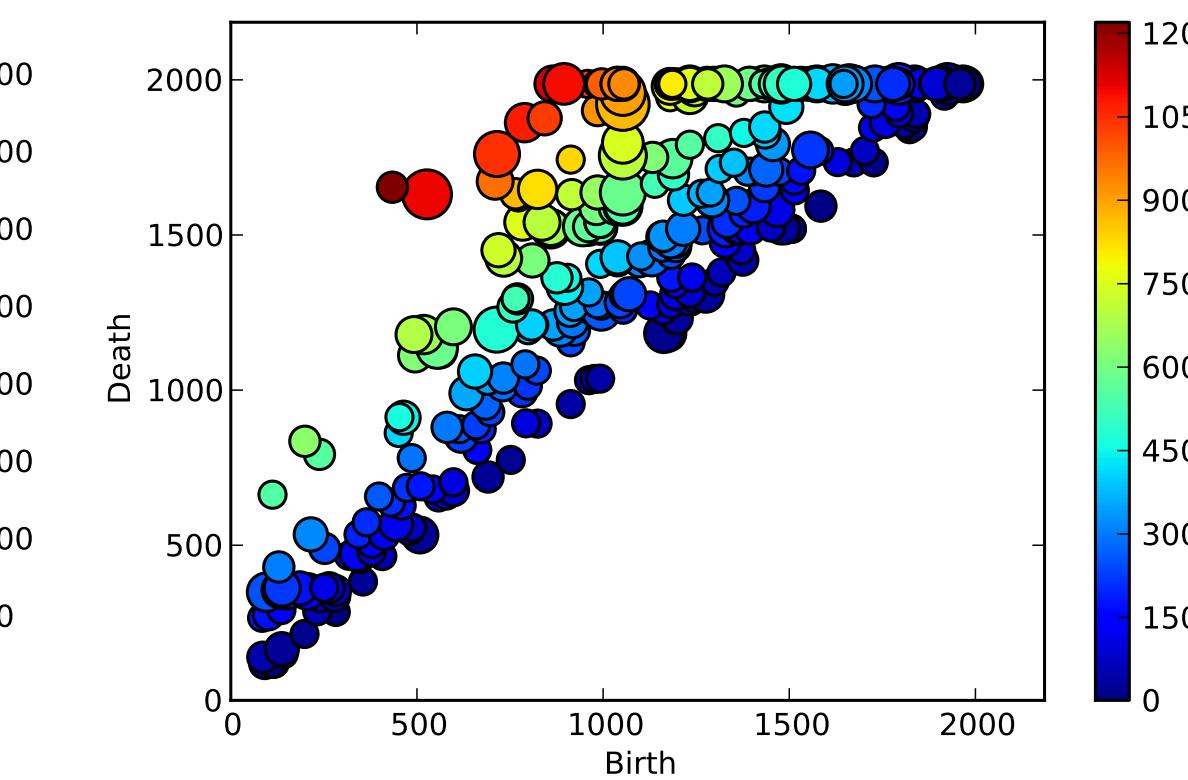
US2000 air passenger network



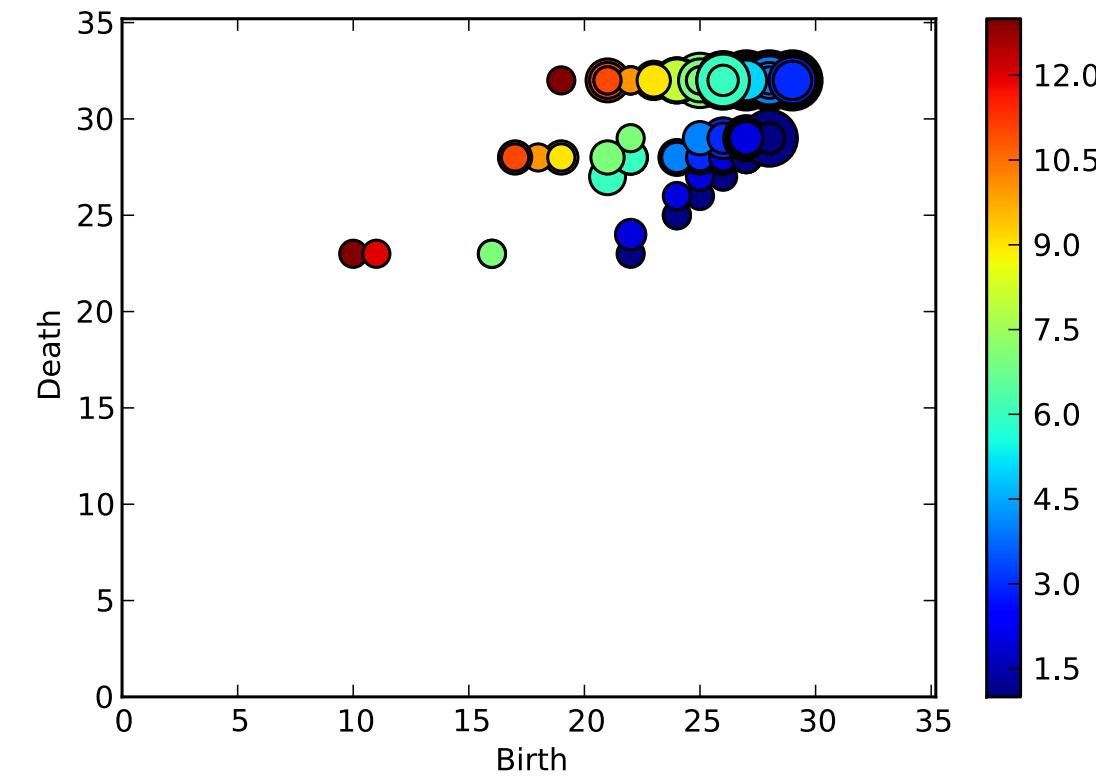
Human gene (sample)



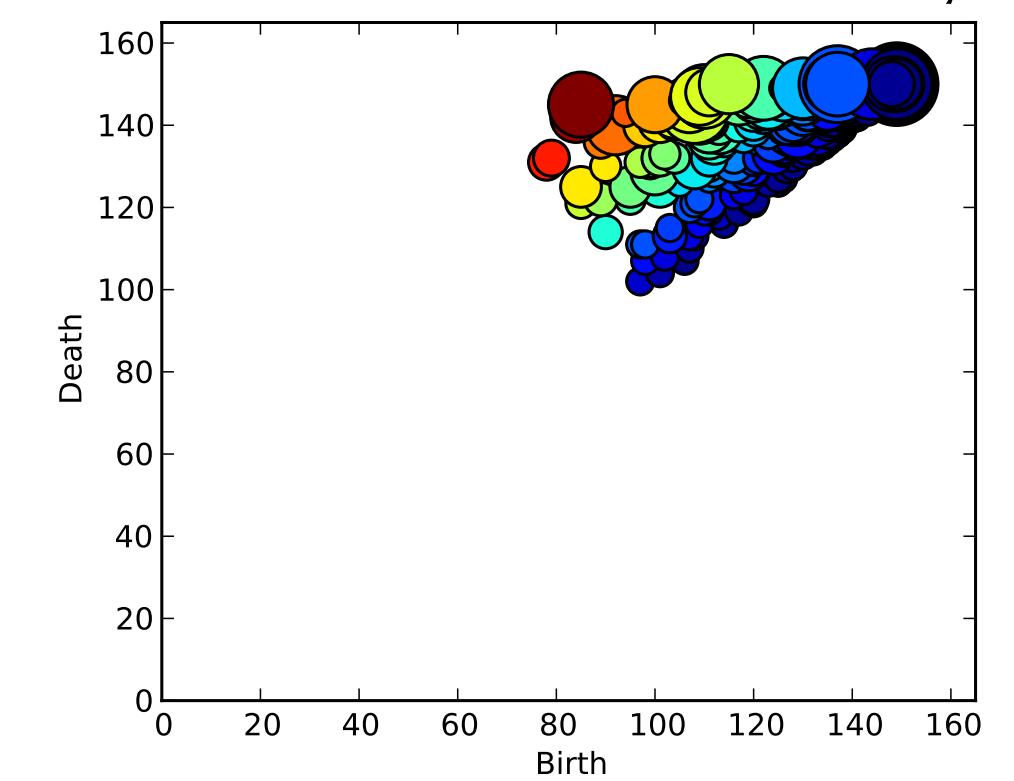
FB-like social network



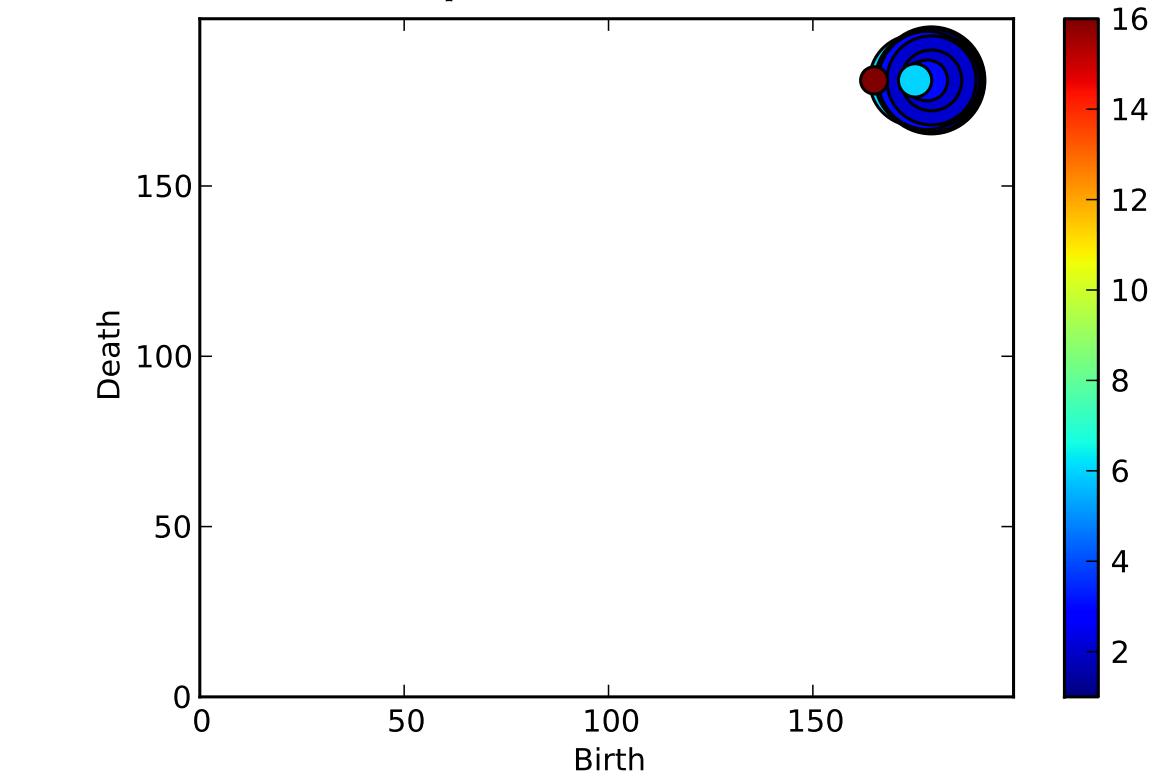
C. elegans



Kids contact duration (day II)



Kumpula et al. model

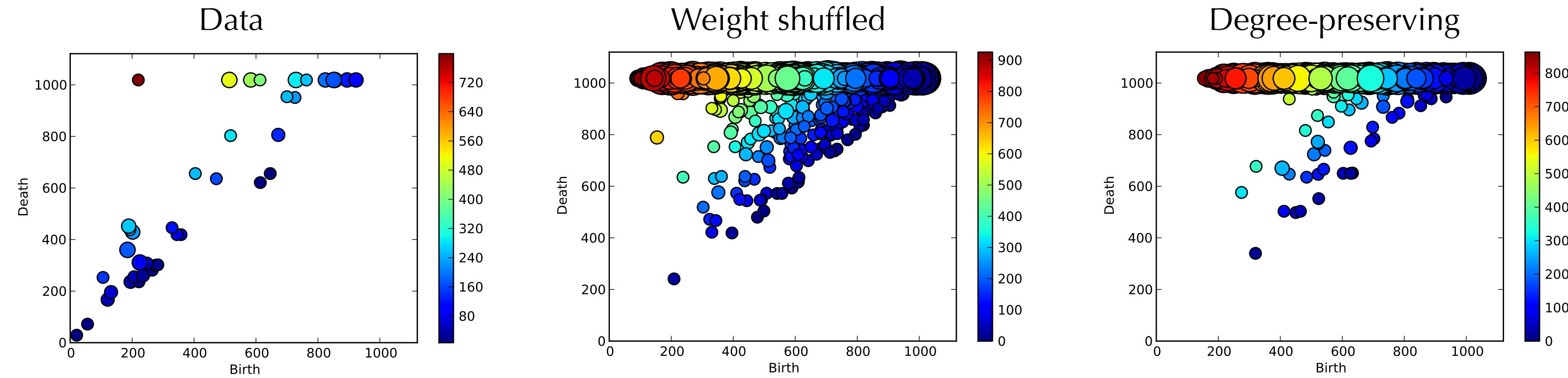


$H_1$

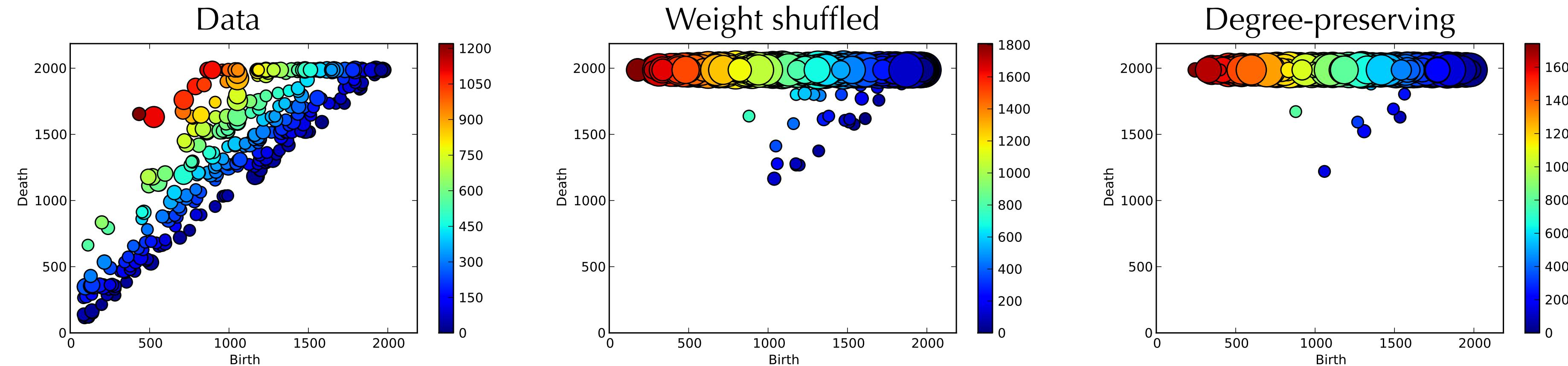
What is significant?

# Datasets II: null models?

US 2000 air passenger network



FB-like social network

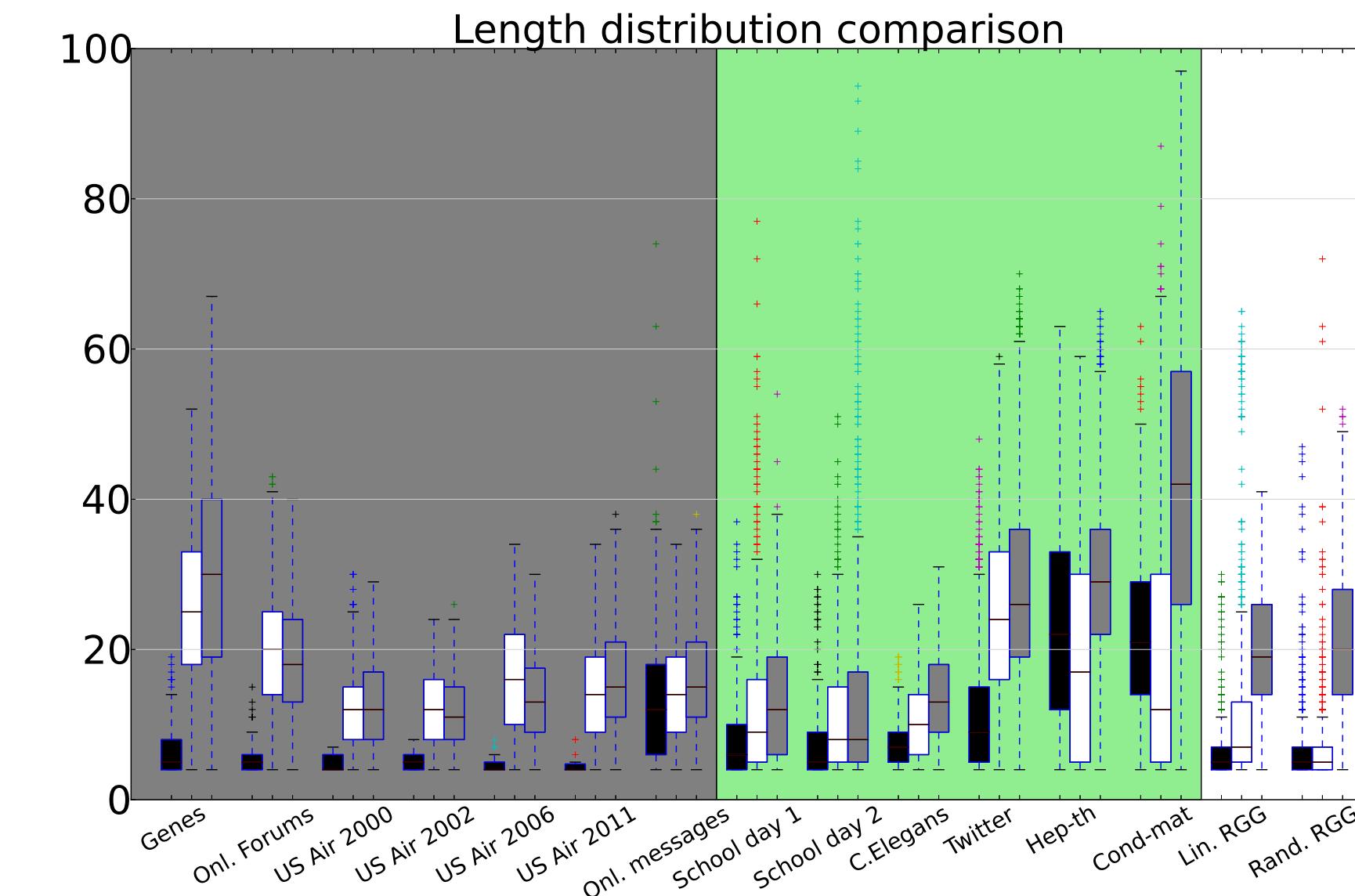
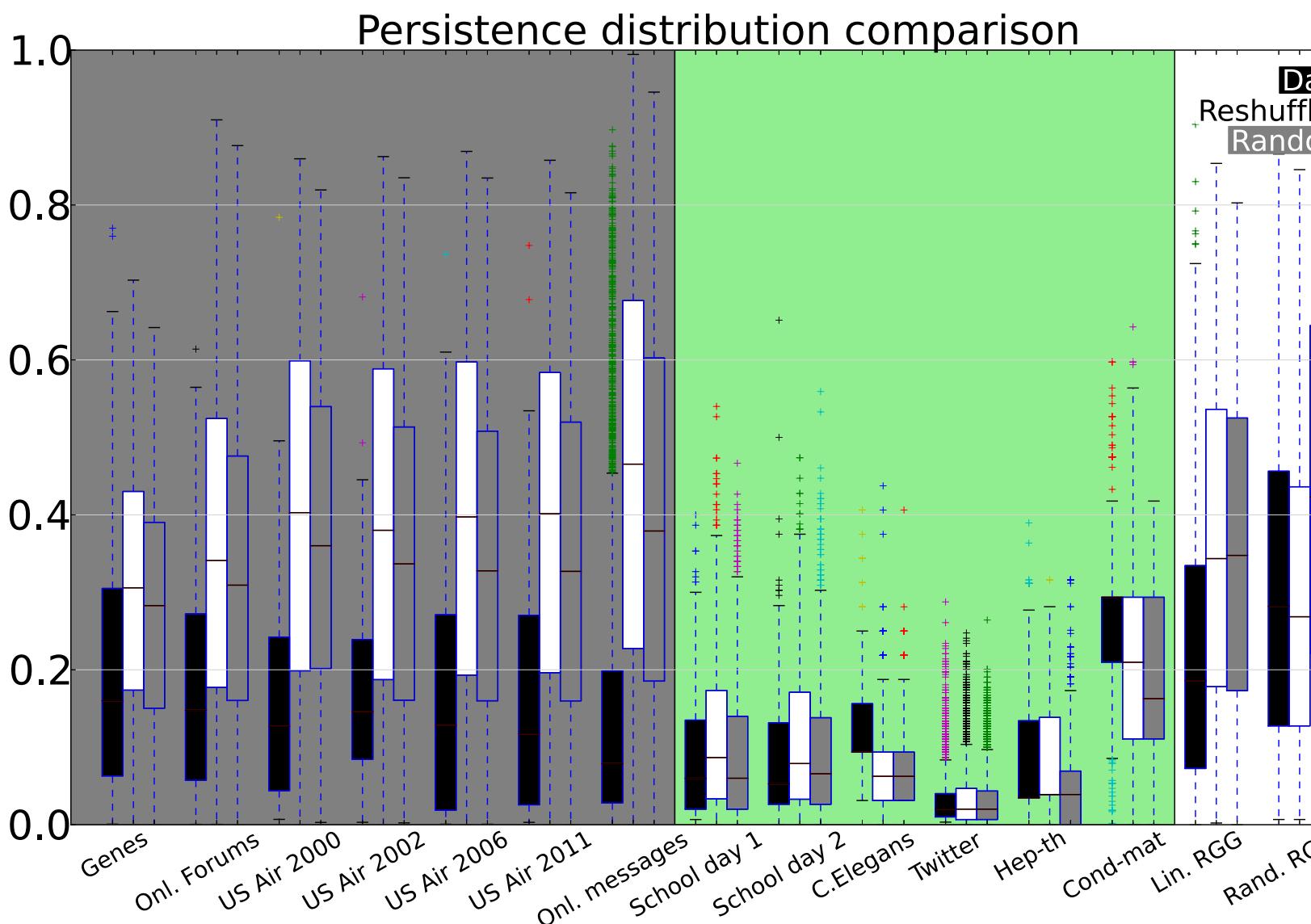
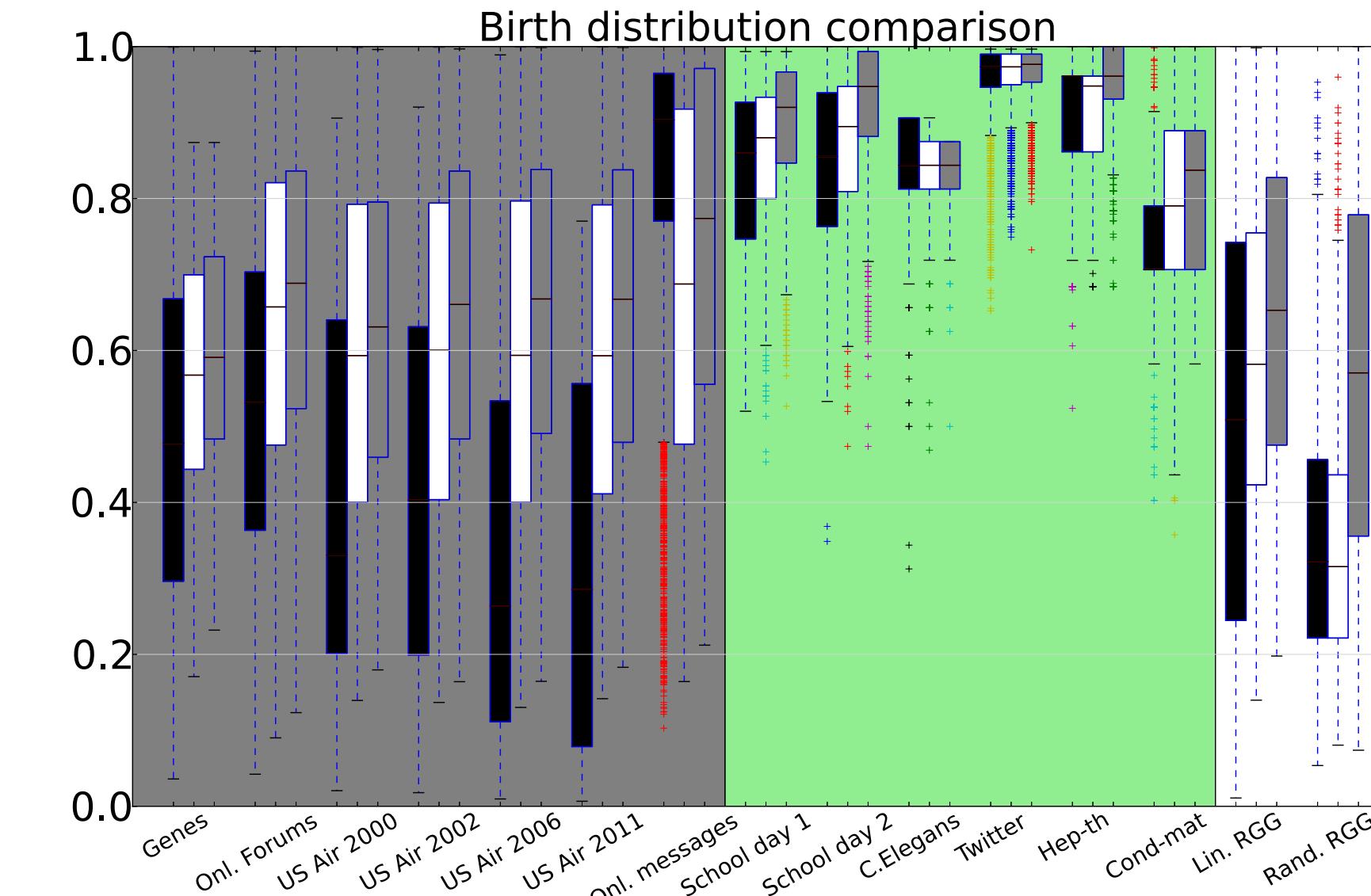


Topological cycles are not simply reproduced by  
statistical network properties

# Datasets III: statistics of generators

Generators are characterized by:

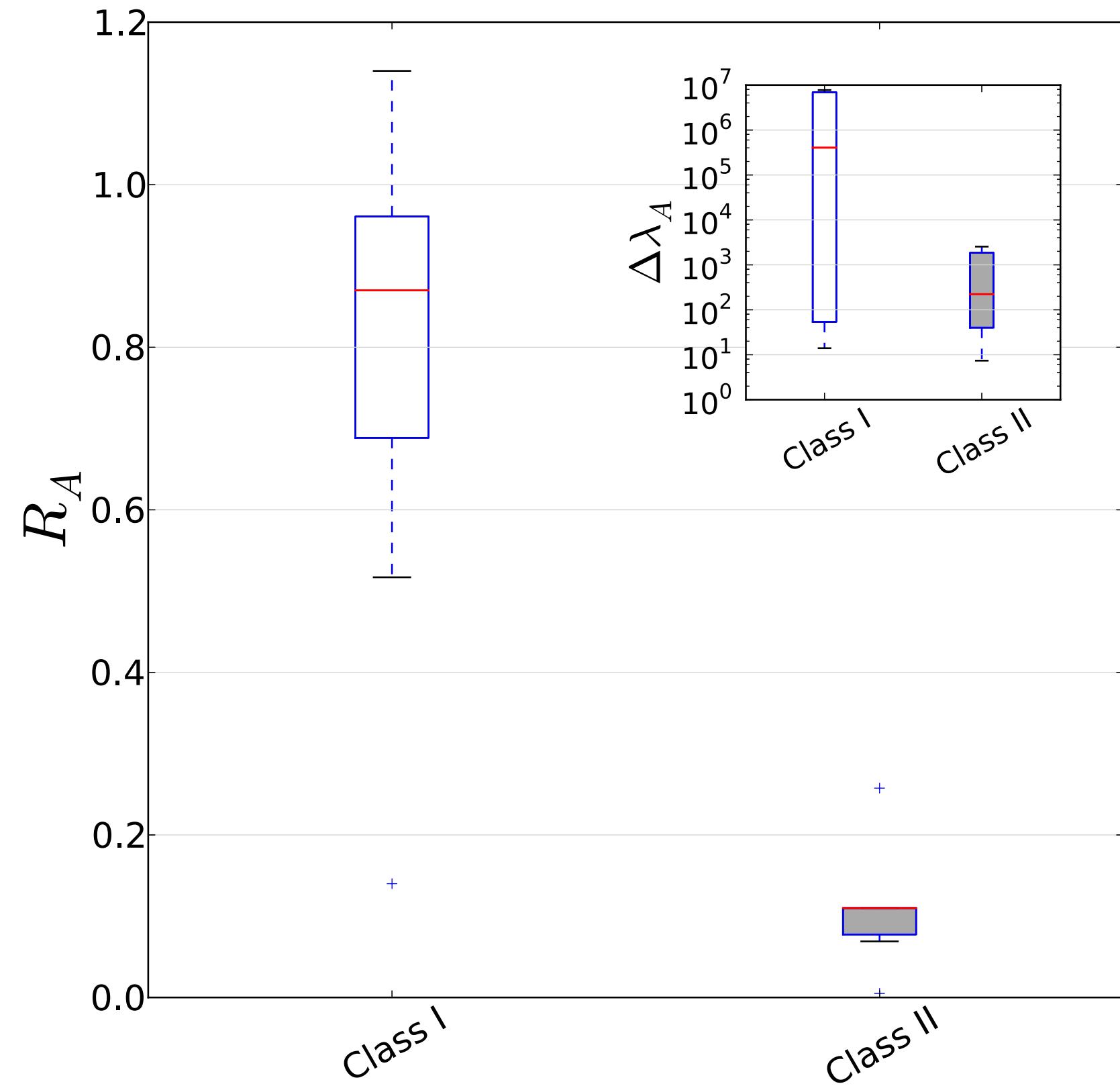
1. Birth
2. Persistence
3. Length (of the cycle)



# Datasets IV: spectral correlates?

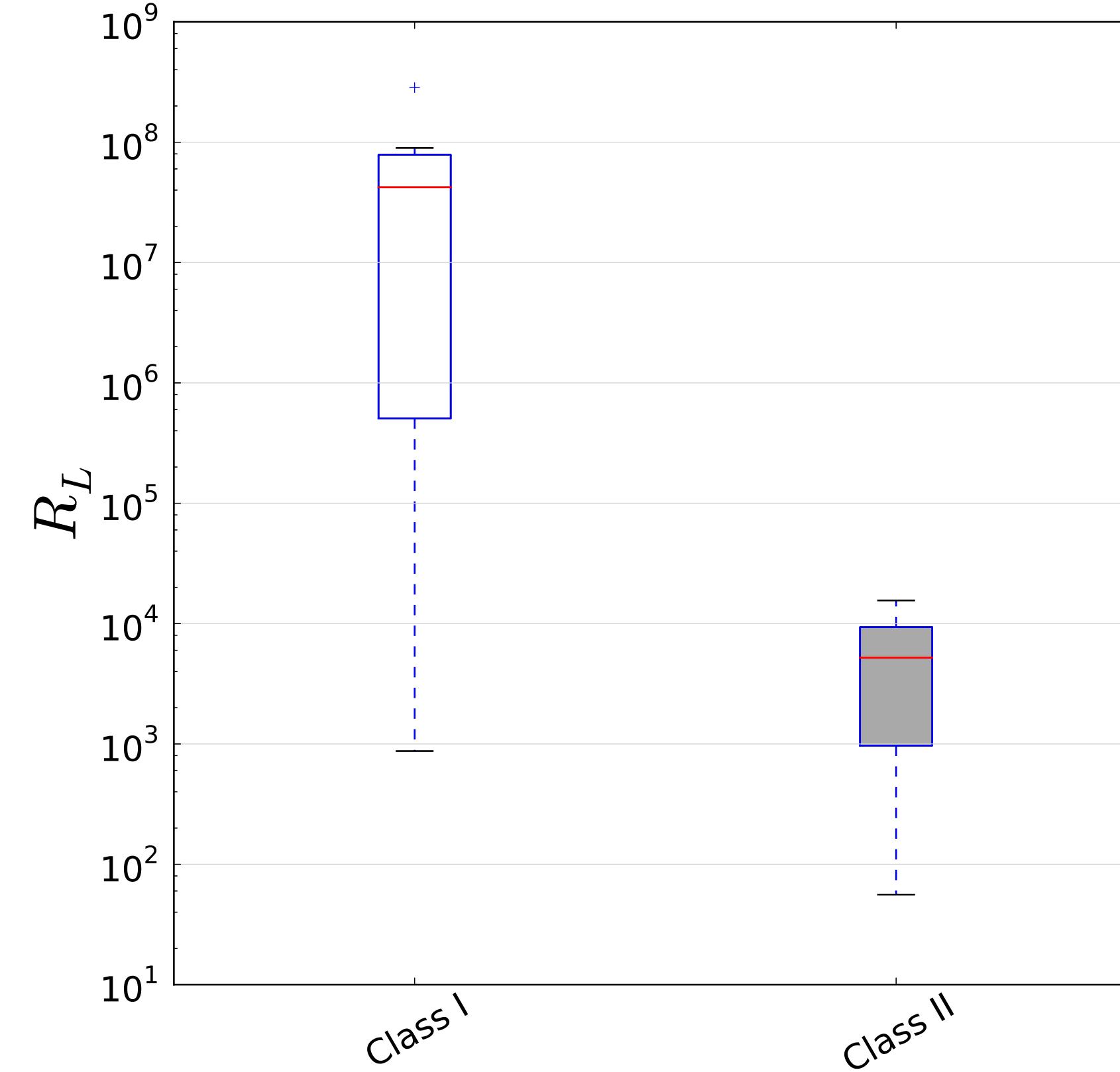
$$R_A = \frac{\lambda_1 - \lambda_2}{\lambda_2 - \lambda_N}$$

Related to graph expansion properties



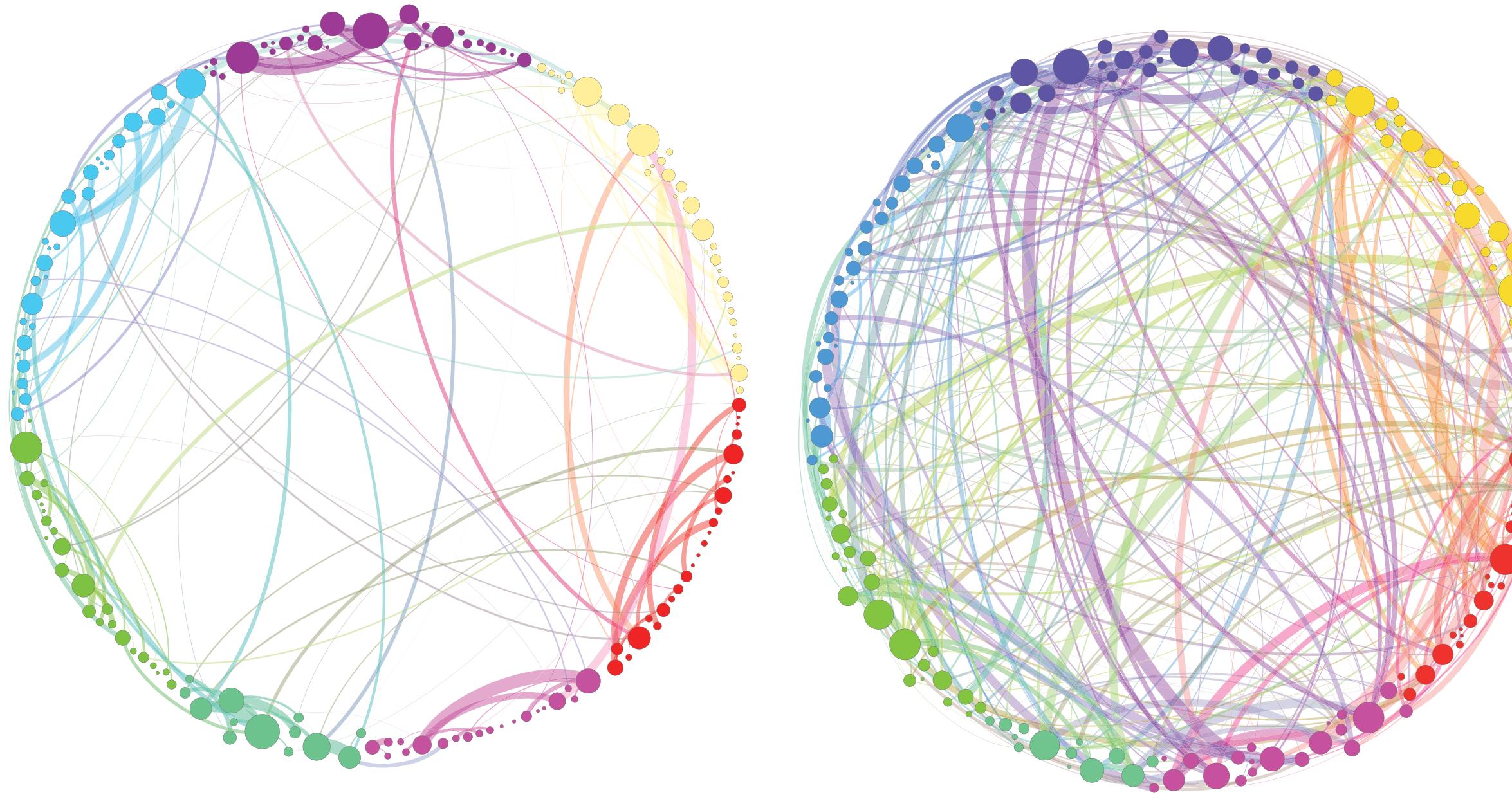
$$R_L = \frac{\lambda_N^L}{\lambda_2^L}$$

Related to synchronisation and dyn. processes

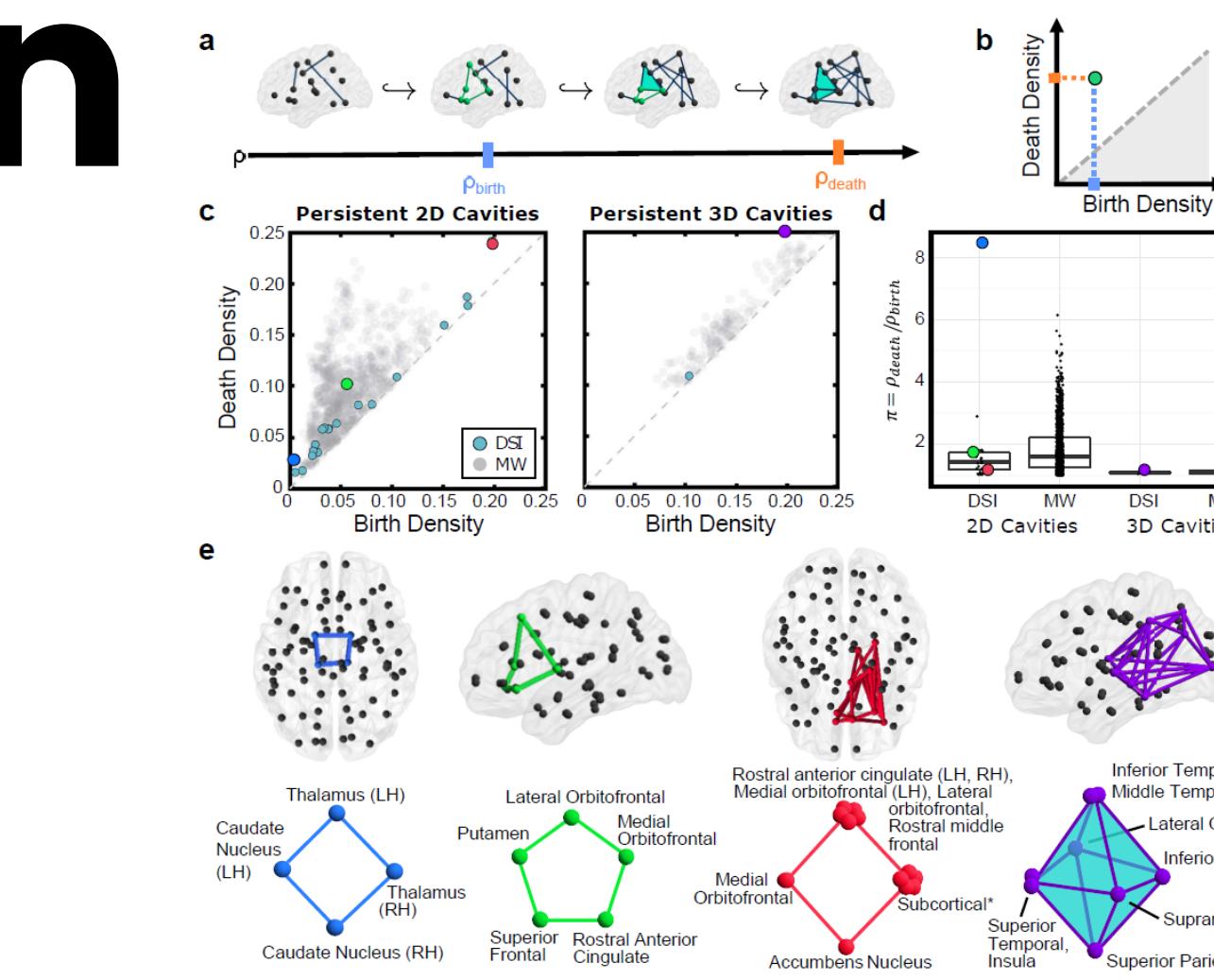


# Do homological mesoscales characterise brain networks?

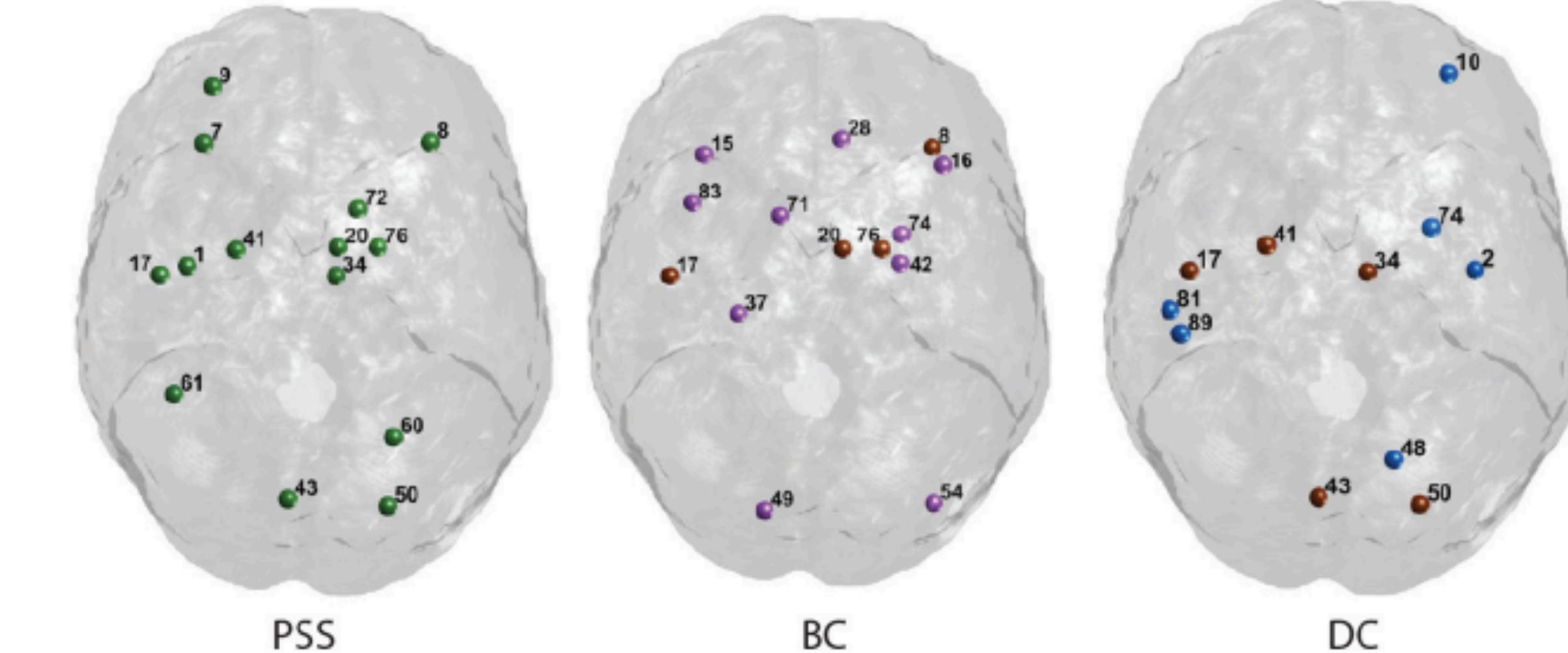
## Graph-representations of 1-dimensional hole structure



Petri, Giovanni, et al. "Homological scaffolds of brain functional networks." *Journal of The Royal Society Interface* 11.101 (2014): 20140873.



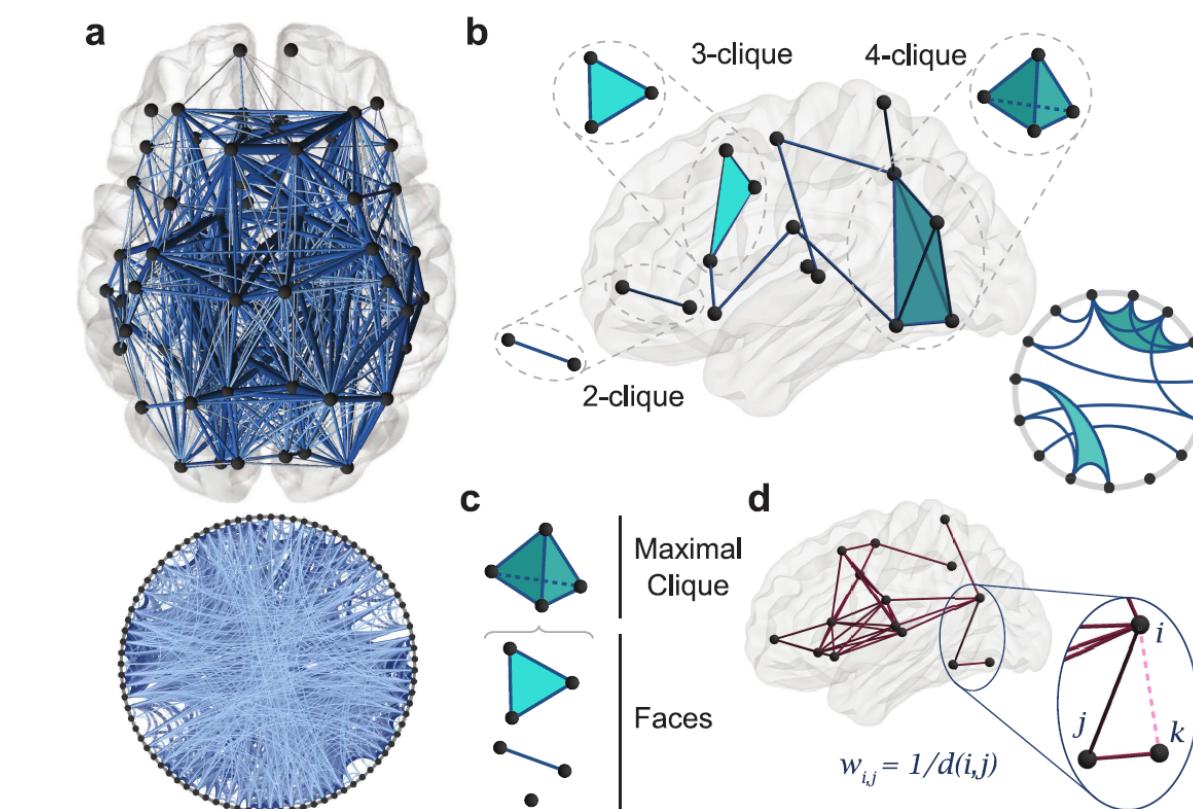
## Homological persistence centrality



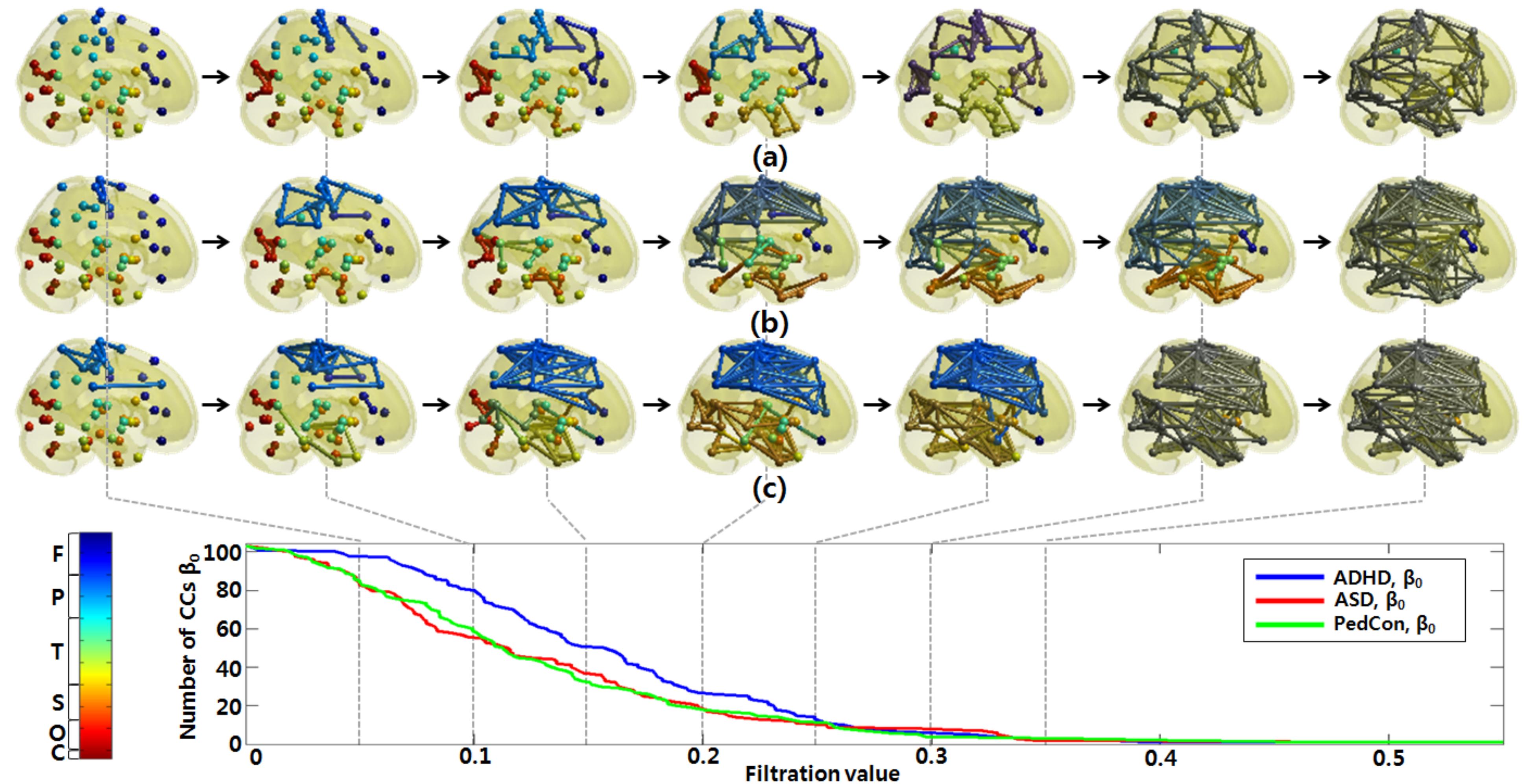
Lord, Louis-David, et al. "Insights into brain architectures from the homological scaffolds of functional connectivity networks." *Frontiers in systems neuroscience* 10 (2016).

## Structural brain cavities

Sizemore, Ann, et al. arXiv:1608.03520 (2016).

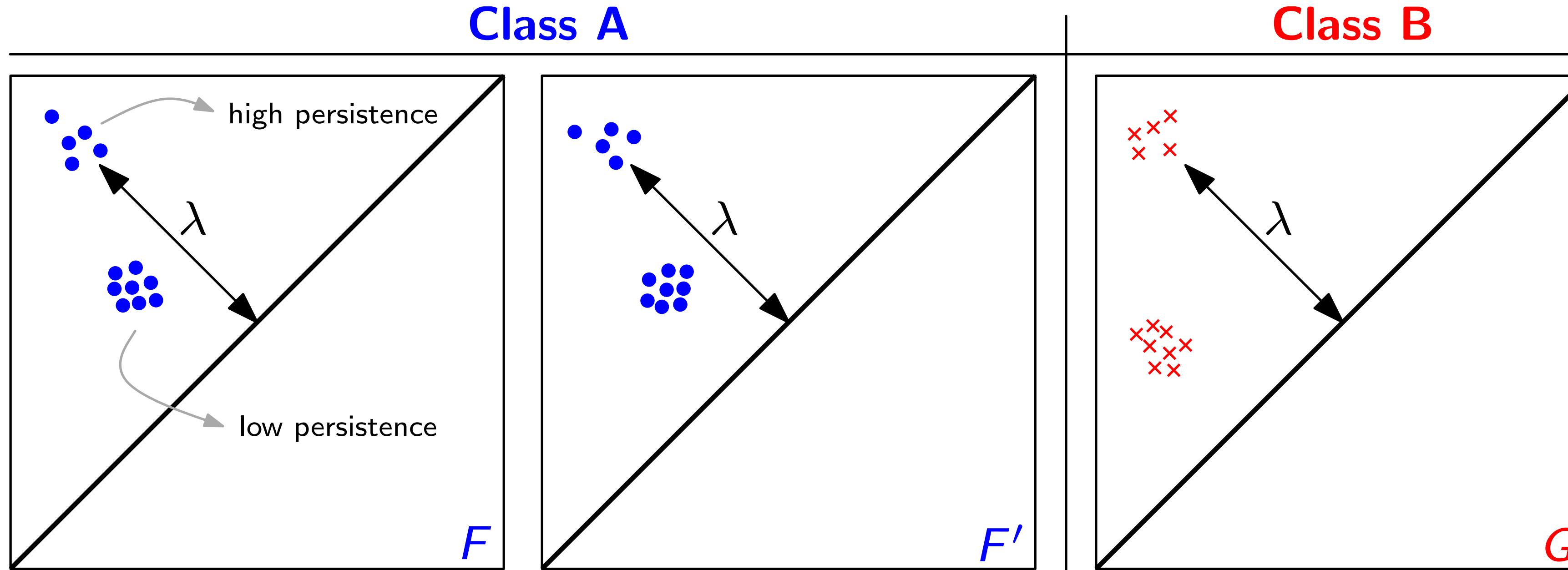


# How do we compare this things?



# Notebook 05

# Distances?



## Distances and Kernels:

- Bottleneck Distance
- (sliced) Wasserstein distance
- Persistence Scale Kernel (Reininghaus et al 2015)
- Weighted Persistence Kernel (Kusano et al, 2016)

# Distances?

## Bottleneck Distance

$$\delta_\infty(D_1, D_2) = \inf_{\gamma} \sup_{z \in D_1} \|z - \gamma(z)\|_\infty$$

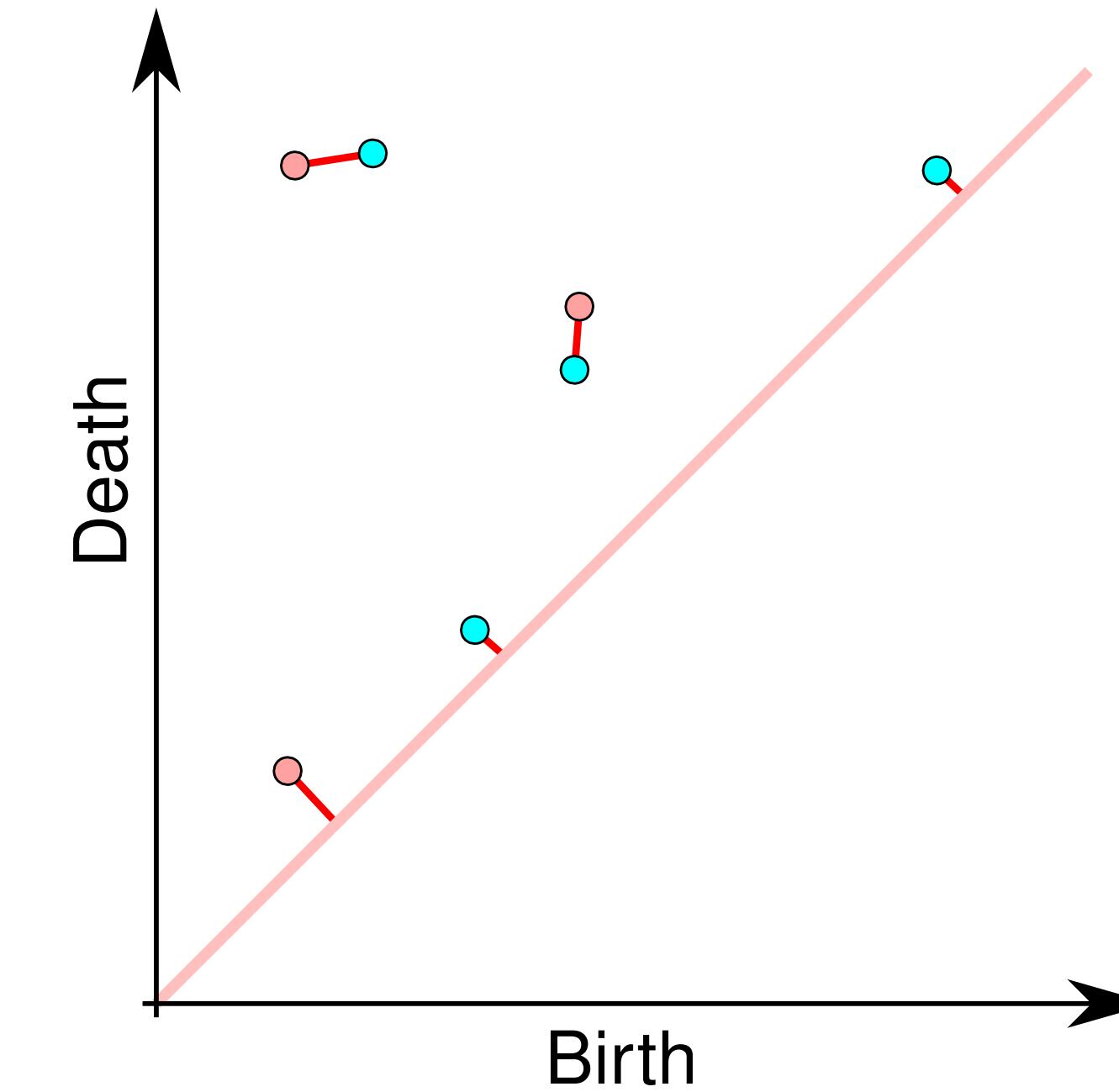
## Wasserstein Distance

$$W_p(D_1, D_2) = \inf_{\gamma} \left( \sum_{u \in D_1} \|u - \gamma(u)\|_\infty^p \right)^{1/p}$$

## Sliced Wasserstein Distance

$$\text{SW}(\text{Dg}_1, \text{Dg}_2) \stackrel{\text{def.}}{=} \frac{1}{2\pi} \int_{\mathbb{S}_1} \mathcal{W}(\mu_1^\theta + \mu_{2\Delta}^\theta, \mu_2^\theta + \mu_{1\Delta}^\theta) d\theta.$$

$$k_{\text{SW}}(\text{Dg}_1, \text{Dg}_2) \stackrel{\text{def.}}{=} \exp \left( - \frac{\text{SW}(\text{Dg}_1, \text{Dg}_2)}{2\sigma^2} \right).$$



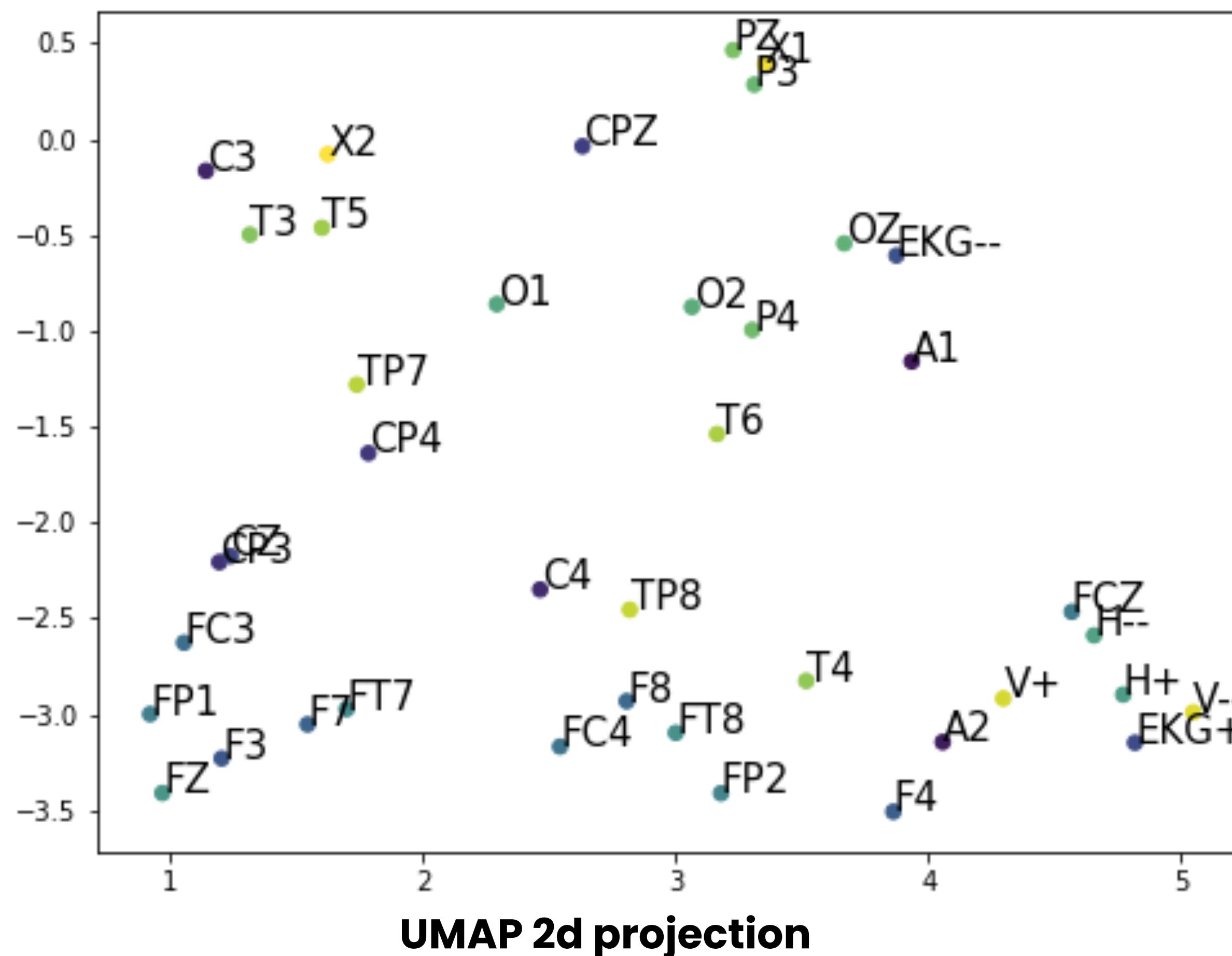
## Persistence Scale Kernel

$$k_\sigma(F, G) = \frac{1}{8\pi\sigma} \sum_{\substack{p \in F \\ q \in G}} e^{-\frac{\|p-q\|^2}{8\sigma}} - e^{-\frac{\|p-\bar{q}\|^2}{8\sigma}}.$$

# Notebook 06

# Do homological shapes capture transitions?

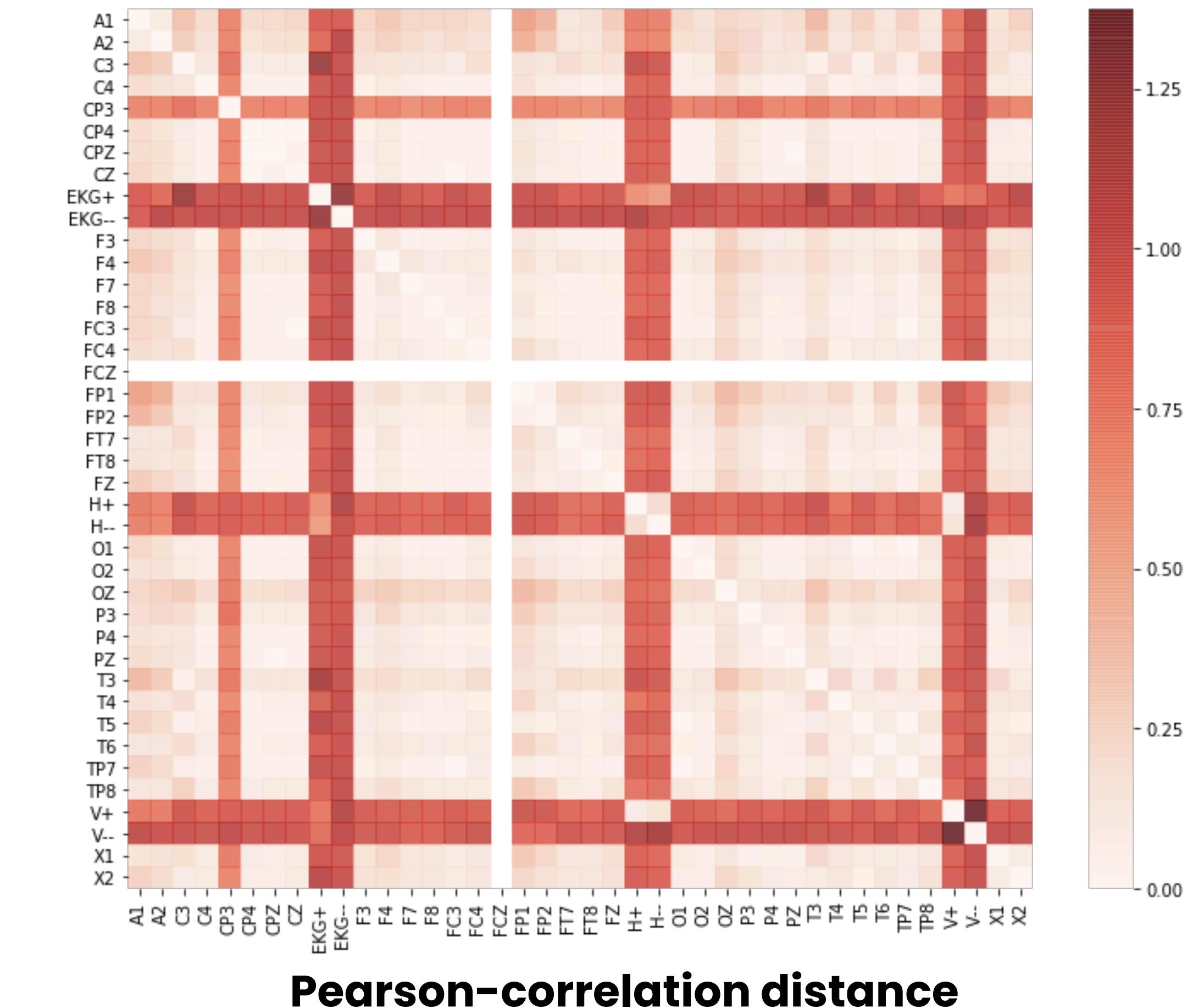
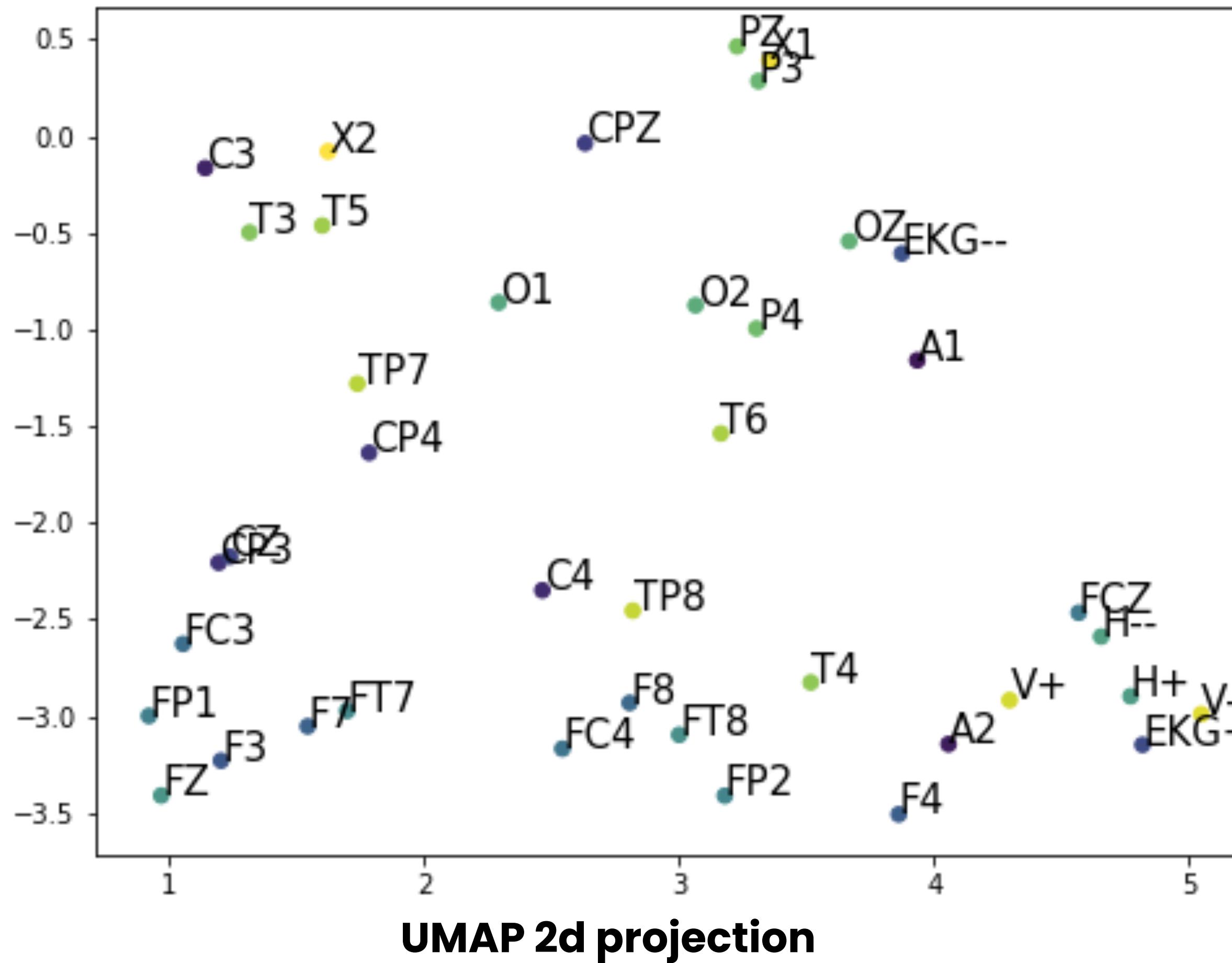
## Correlation space of EEG signals



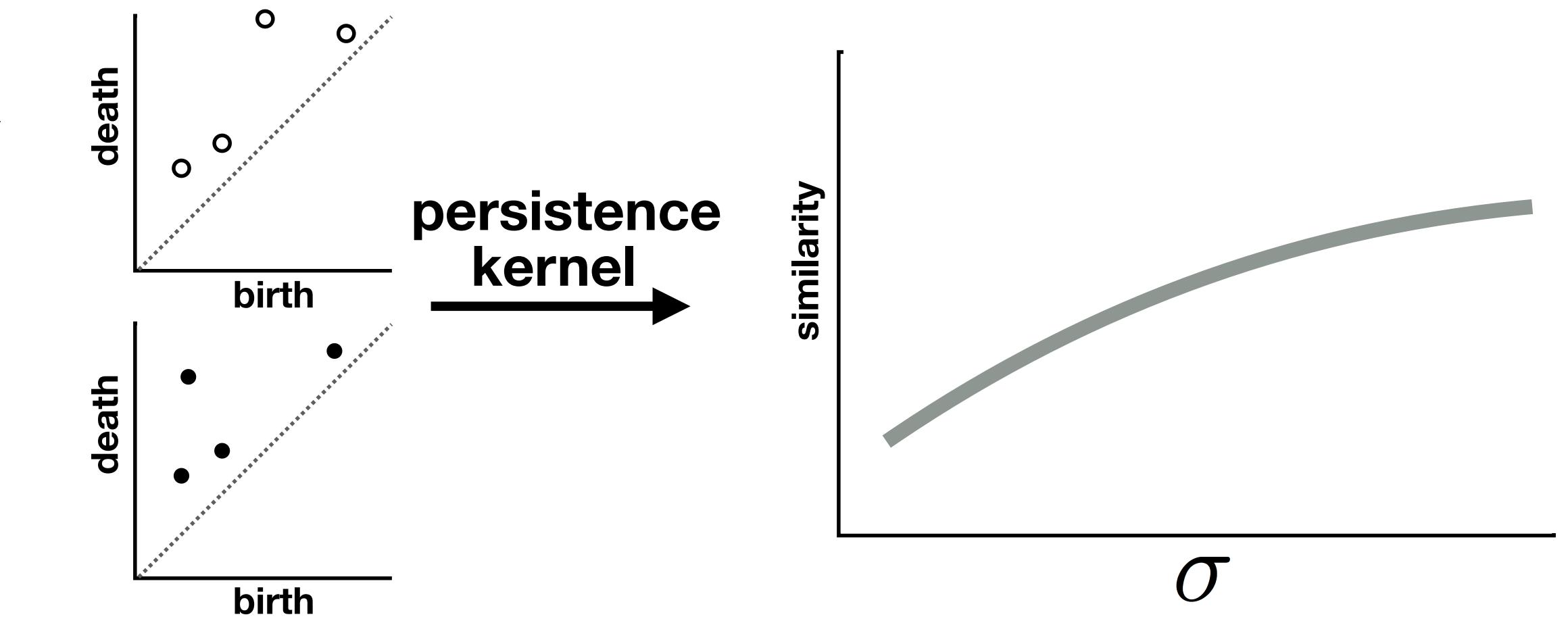
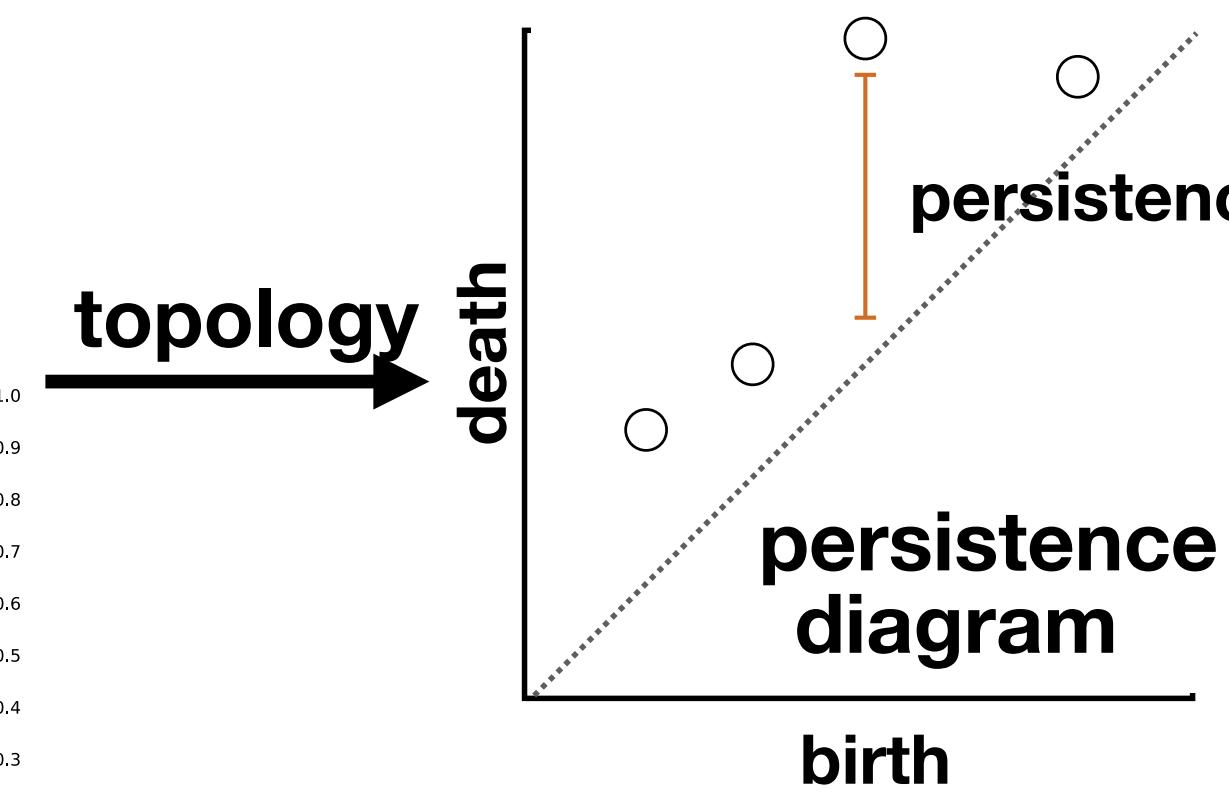
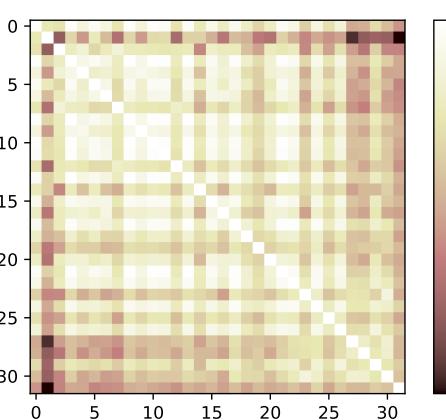
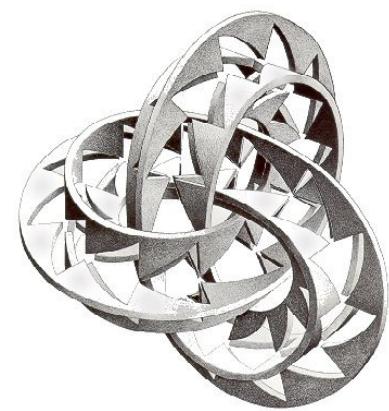
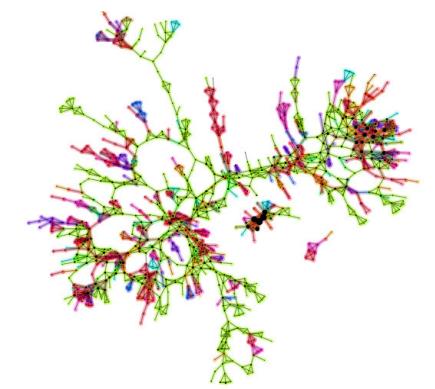
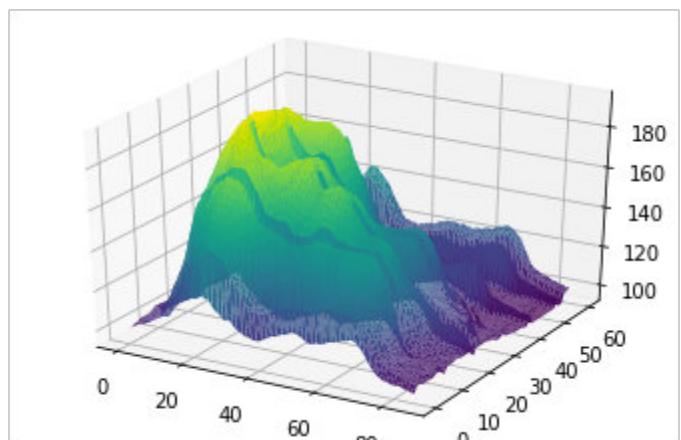
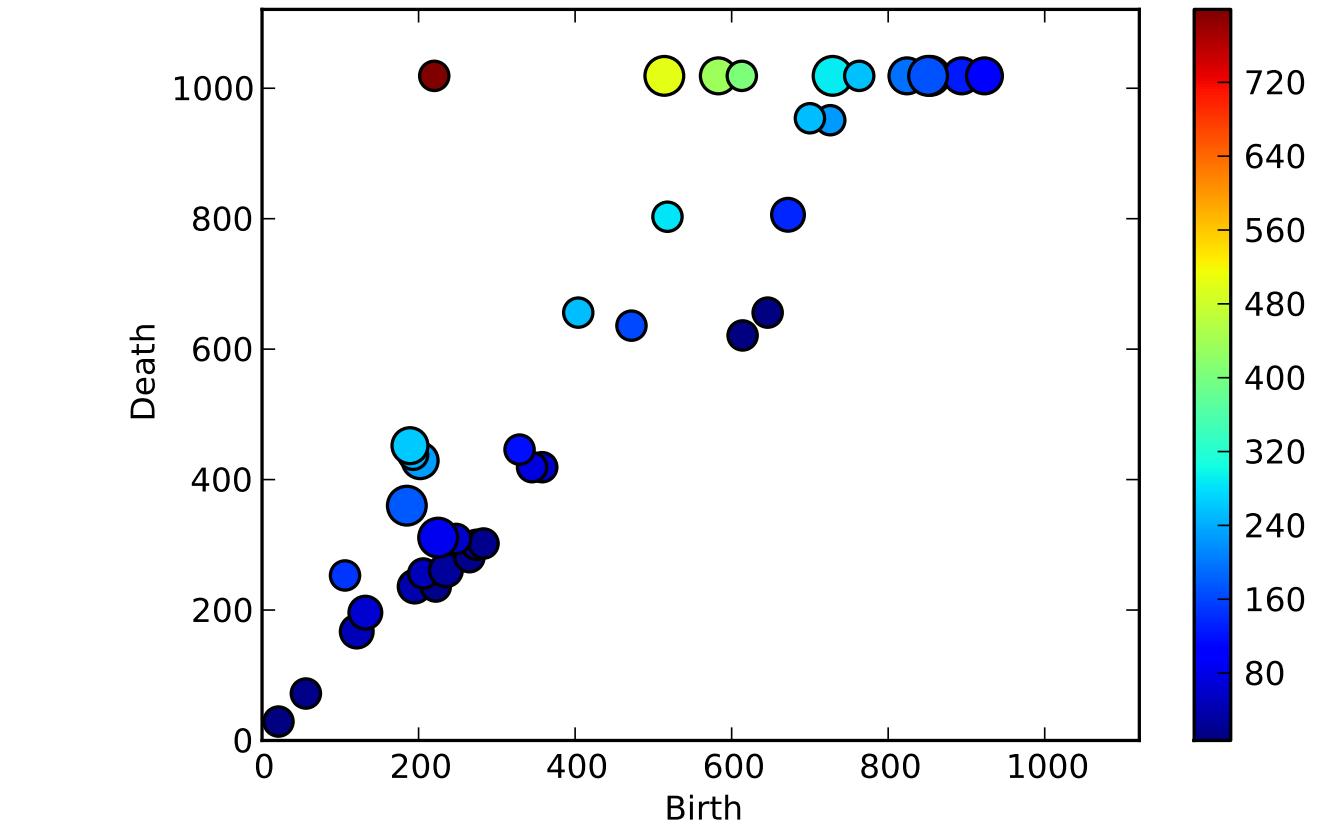
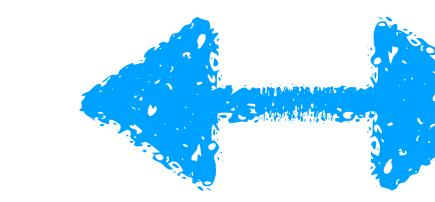
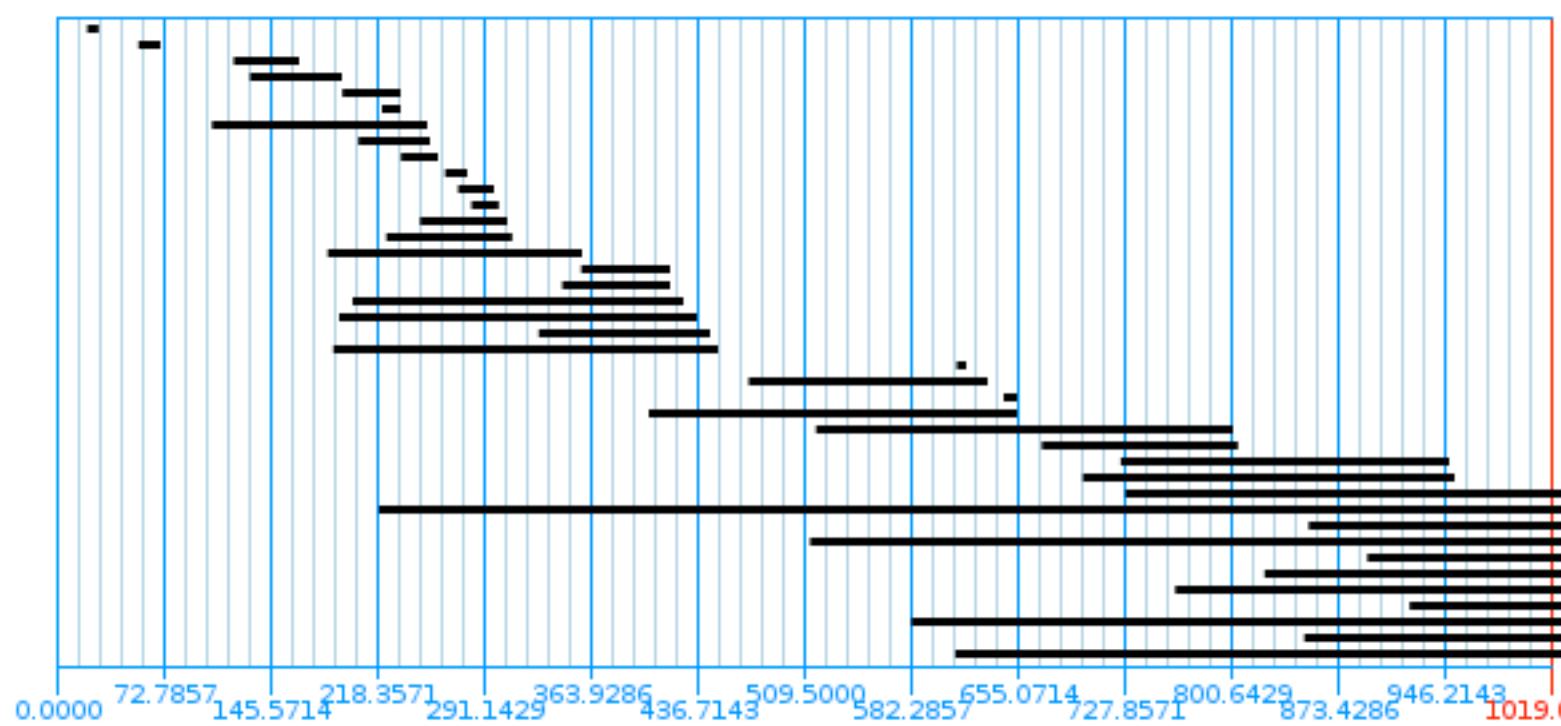
29 EEG channels  
>10<sup>5</sup> timepoints

# Do homological shapes capture transitions?

## Correlation space of EEG signals



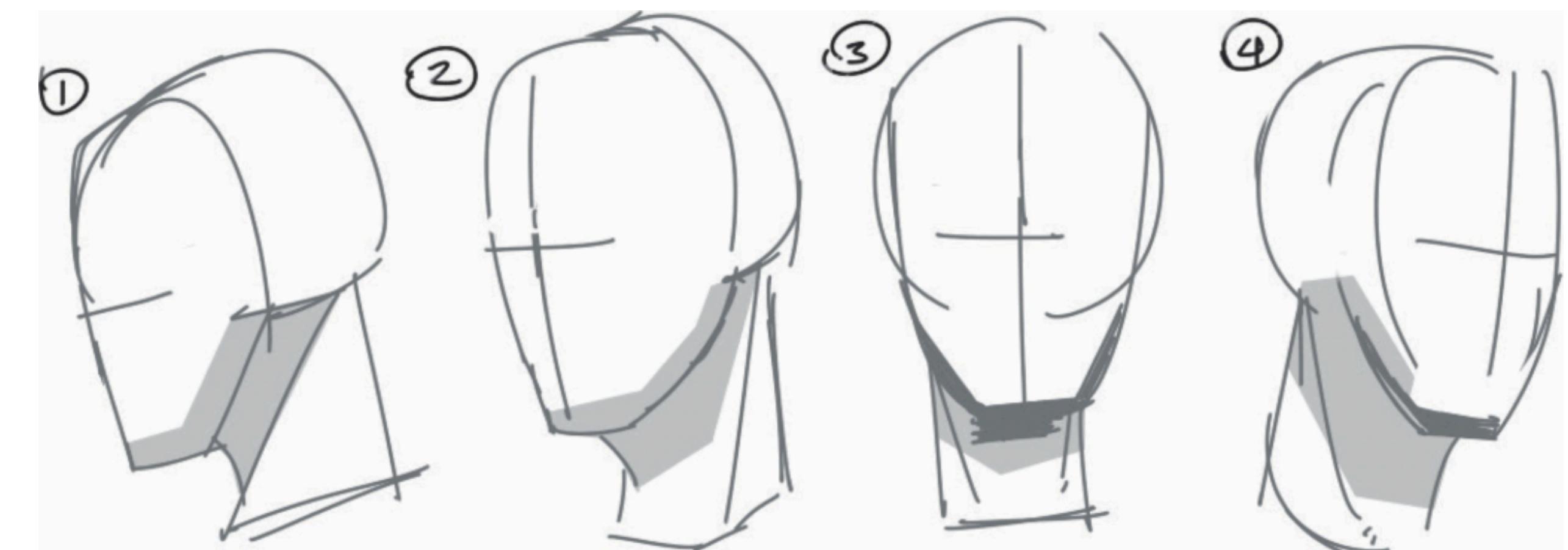
# Do homological shapes capture transitions?



# Do homological shapes capture transitions?

## Dataset

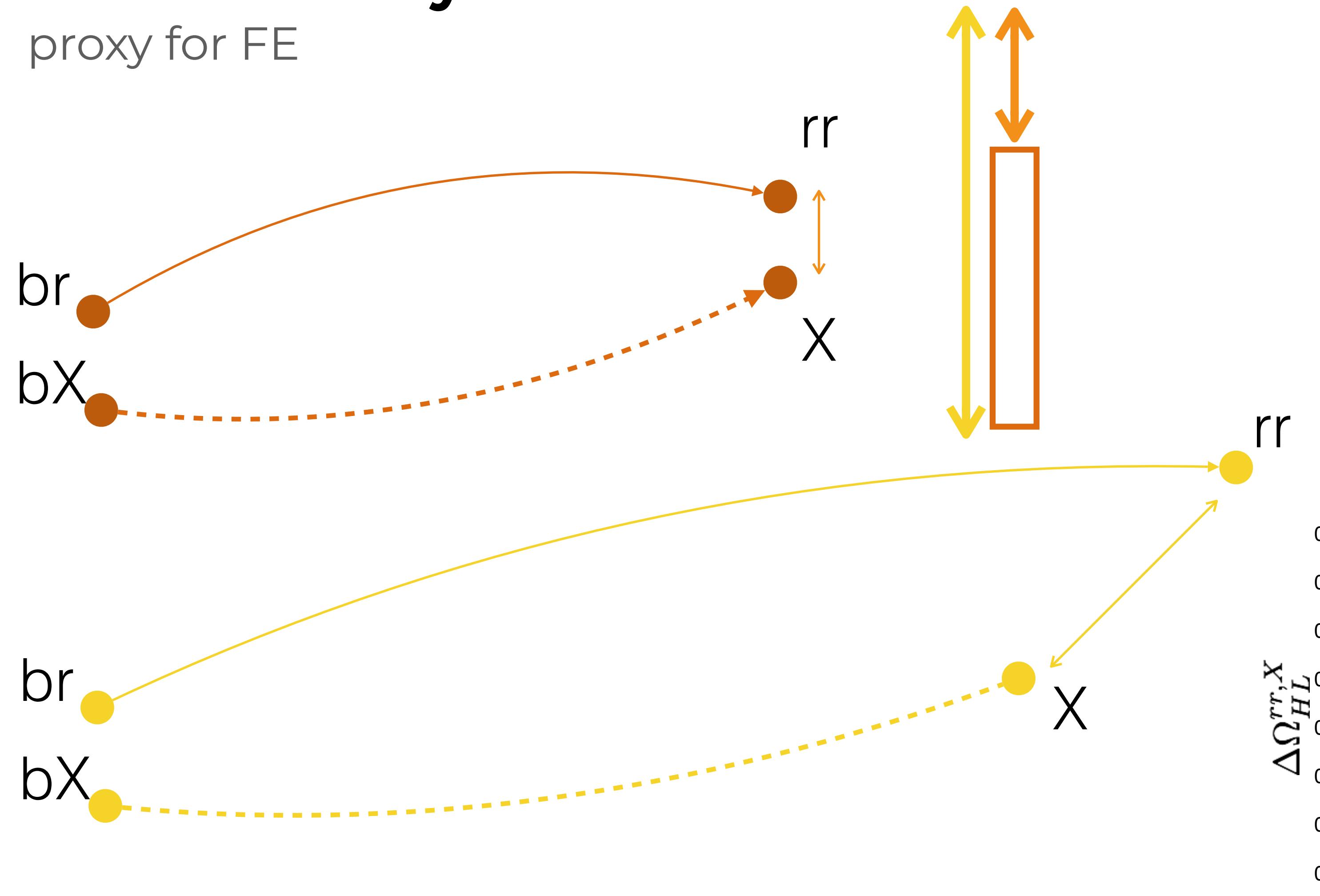
- 18 highly hypnotisable subjects, 19 lowly hypnotisable subjects
- Hypno-score proxy for increased functional equivalence (FE)
- Scalp EEG recording (29 EEG channels)
- Visual + kinetic mental image / Rotation / visual + kinetic mental image
- **Hypothesis:**
  - high FE -> accurate real-task state reproduction -> vivid images
  - high FE -> small changes from basal state -> effortless imagination



# Do homological shapes capture transitions?

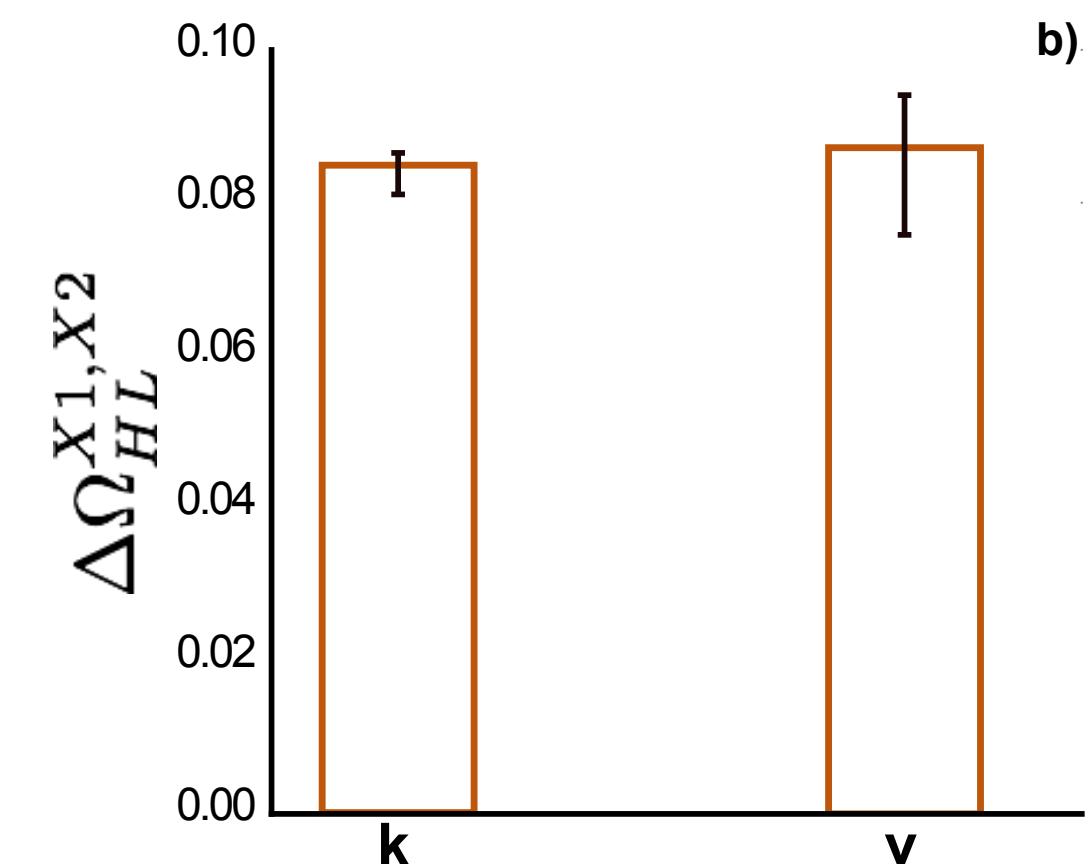
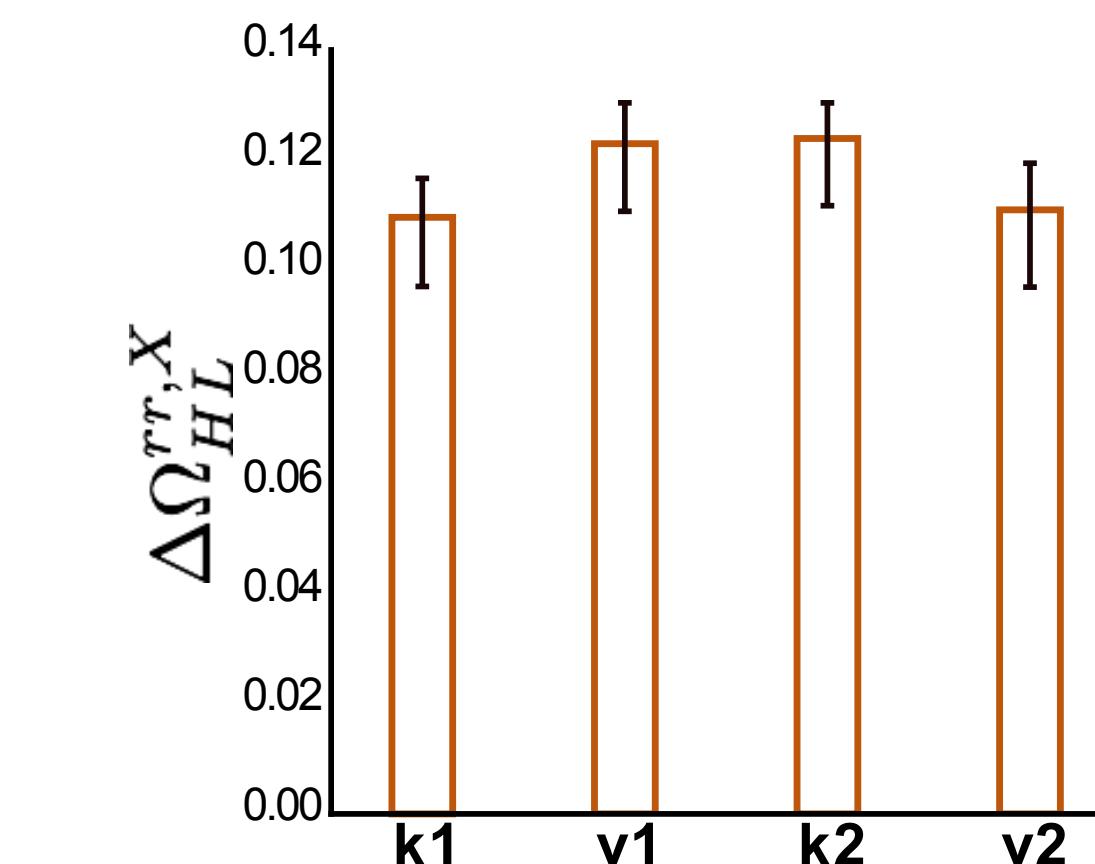
## Task accuracy

proxy for FE



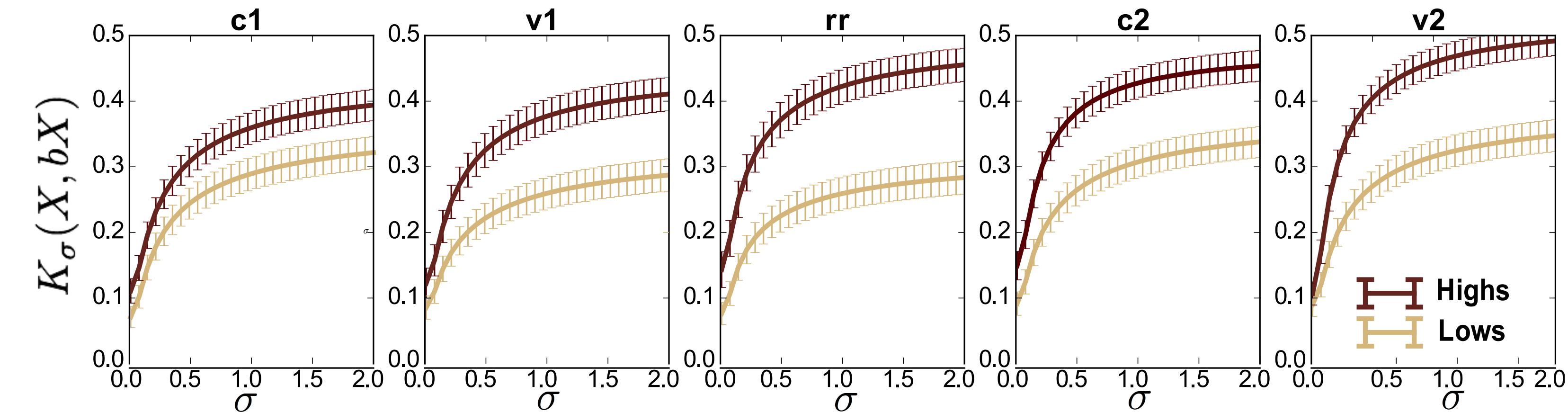
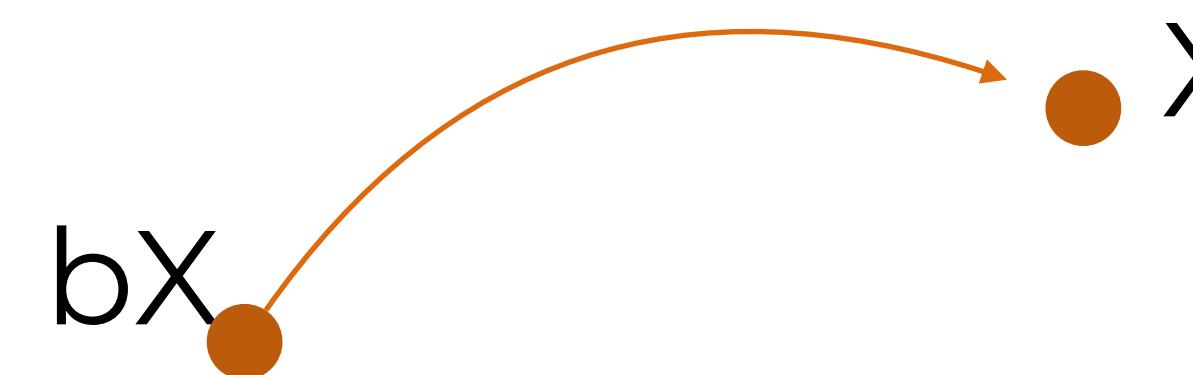
Are they more accurate?  
**YES!**

higher vividness in Highs  
(group level)

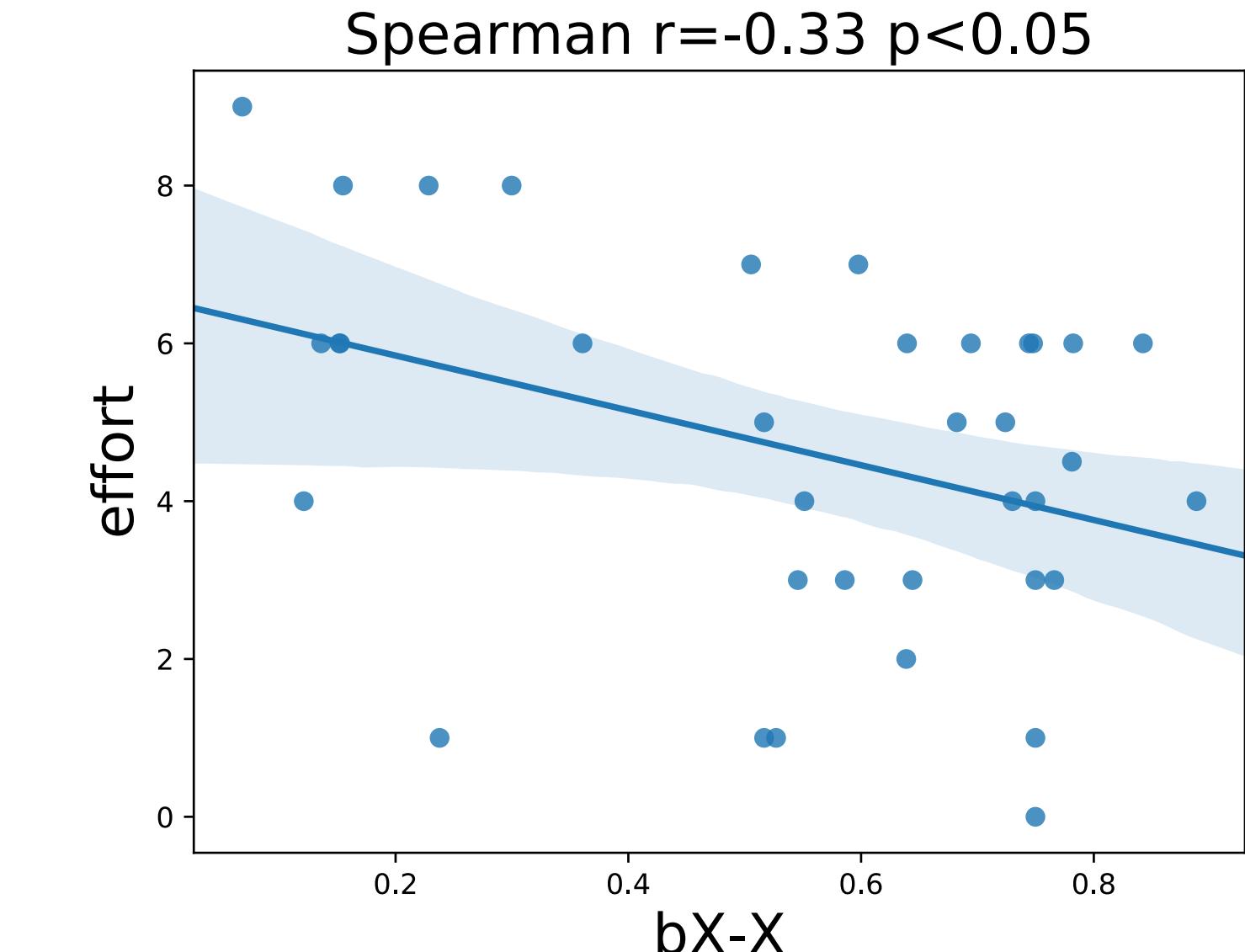


# Do homological shapes capture transitions?

## Basal deviation



- Do *highs* reconfigure less?  
**YES!**
- In all sensorimotor and imaginative tasks *highs* show higher topological similarity with basal conditions than *lows*
- Lower reported **effort** in *highs*

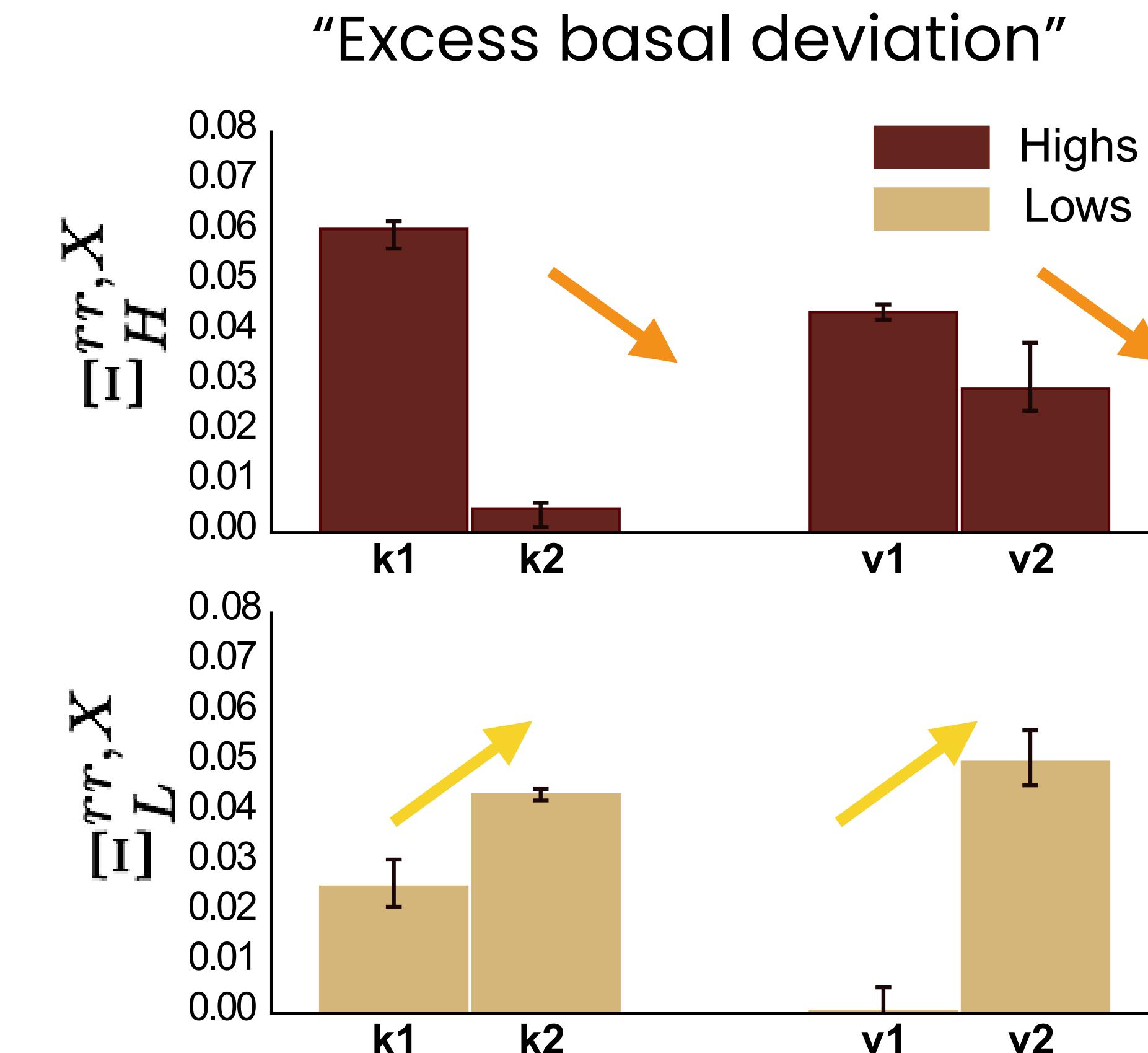


# Do homological shapes capture transitions?

## Learning

Do they learn?

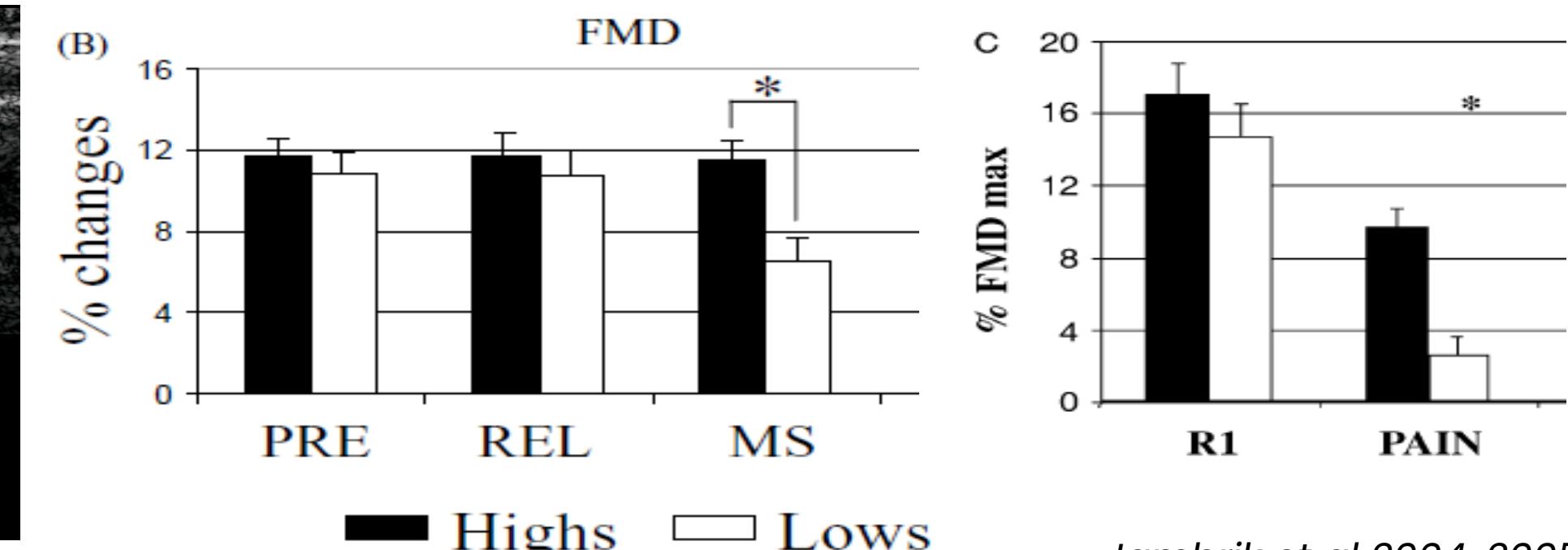
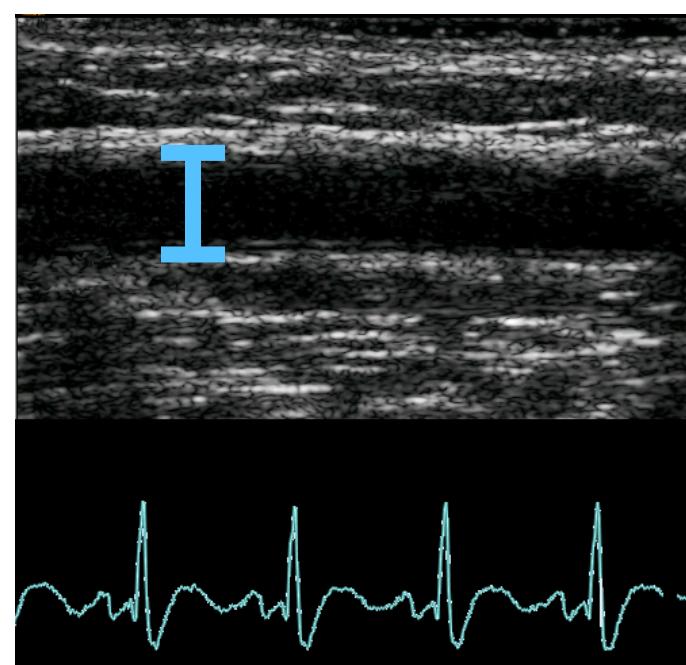
- Highs
  - Reduced effort in Highs
  - same vividness and accuracy
- Lows
  - reduced vividness, same accuracy
  - increased effort
- No measurable improvement in accuracy but “in easiness”



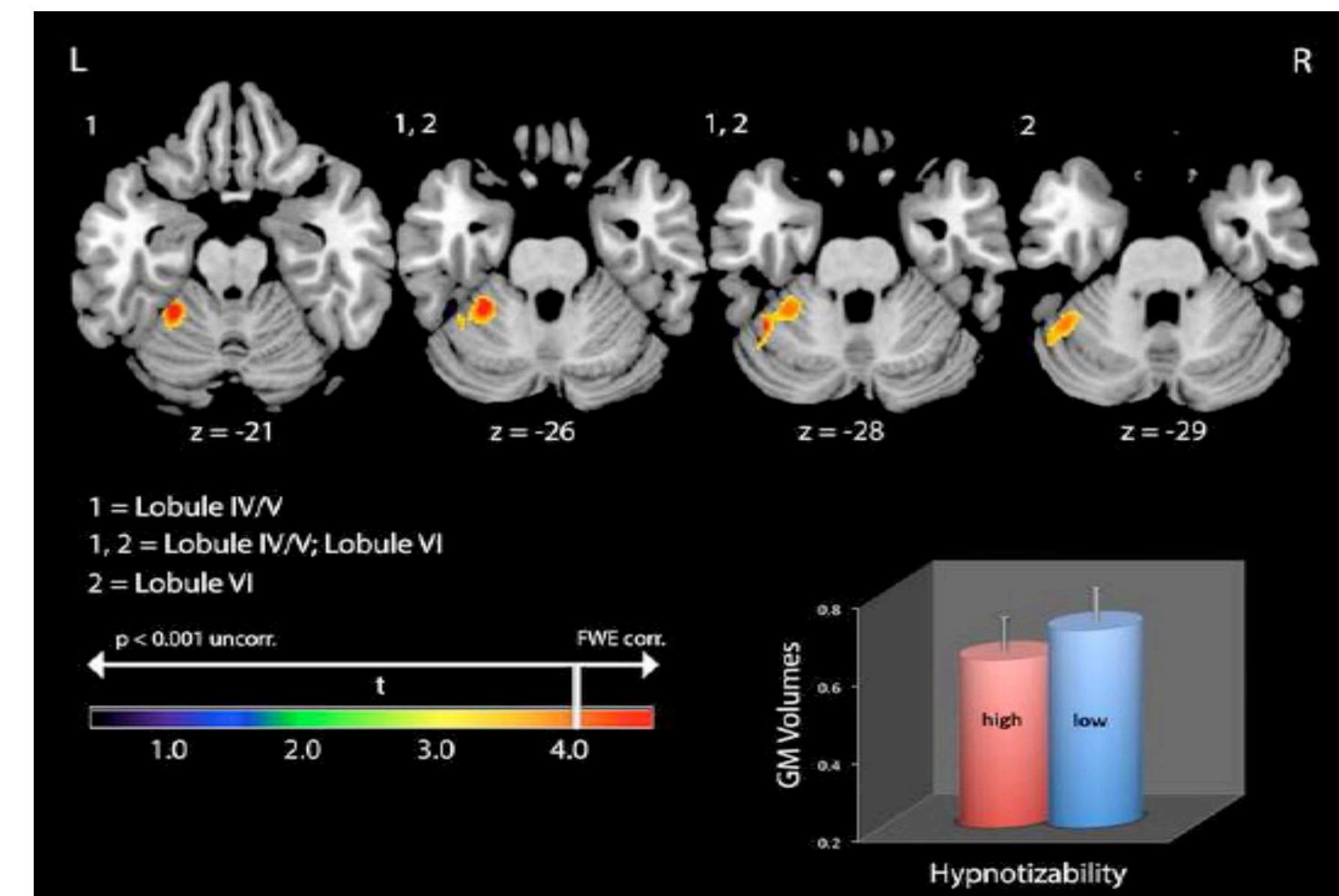
# Do homological shapes capture transitions?

## Potential mechanism

- larger FMD in Highs depends on larger release of endothelial nitric oxide
- reduced cerebellar volume
- hence, reduced inhibition of the cerebral cortex by cerebellum causing:
  - stronger FE + highly distributed information processing
  - Paradoxical control of pain after cerebellar anodal stimulation (*Bocci et al, 2017*)



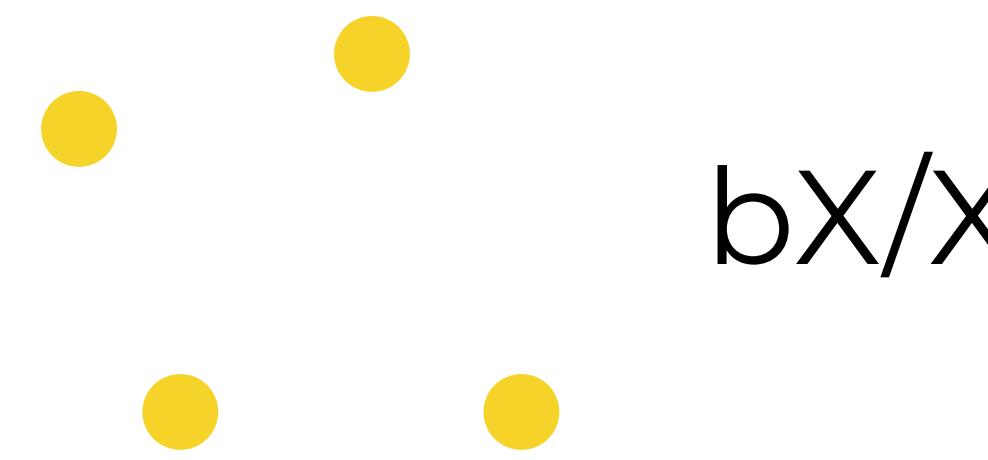
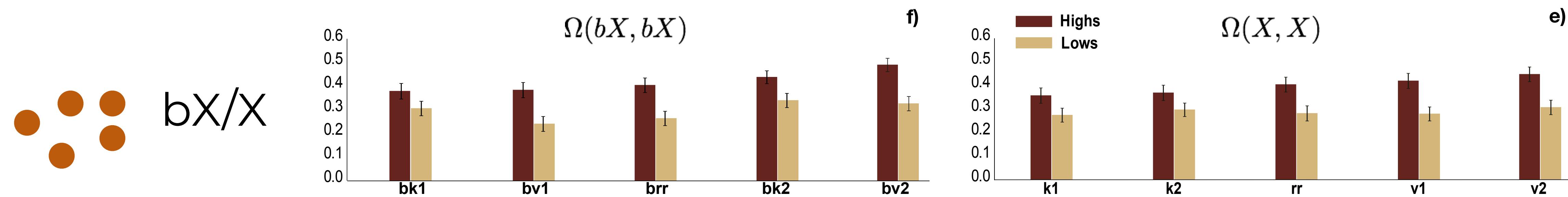
Jambrik et al 2004, 2005



# Do homological shapes capture transitions?

One more thing...

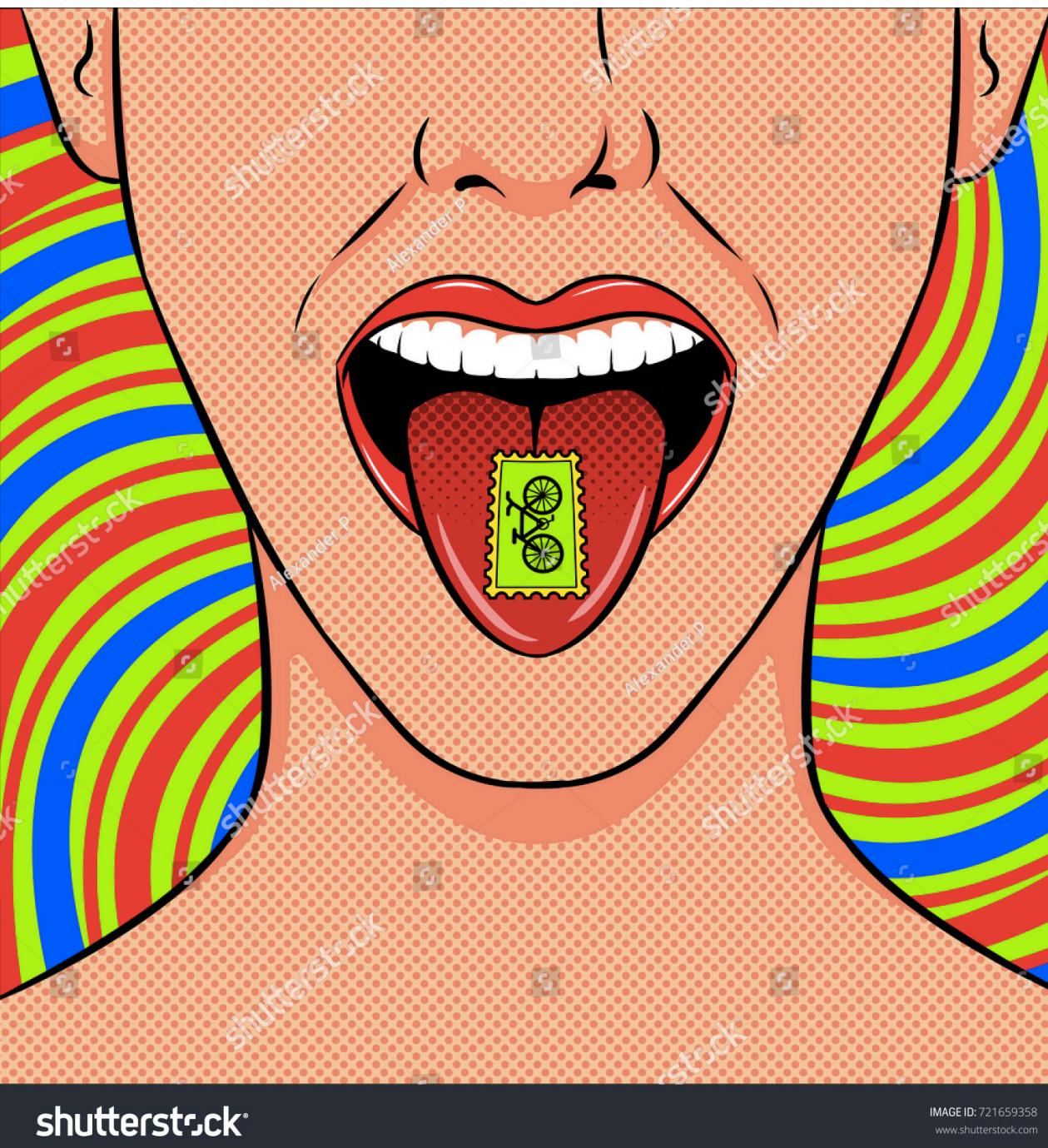
## Group homogeneity



- scarce topological changes could be the effect of largely distributed information processing (*Sporns, 2013*)
- highly distributed information could be induced by ascending neuro-modulatory systems (*Shine et al. 2016; Bell and Shine 2016*)
- NO increases the brain noradrenaline and dopamine availability (*Ghasemi et al., 2018*)

# Can we detect altered states?

## Datasets



rs-fMRI  
15 subjects, 2 sessions  
3 recording conditions

Carhart-Harris, Robin L., et al. "Neural correlates of the LSD experience revealed by multimodal neuroimaging." *Proceedings of the National Academy of Sciences* 113.17 (2016): 4853-4858.



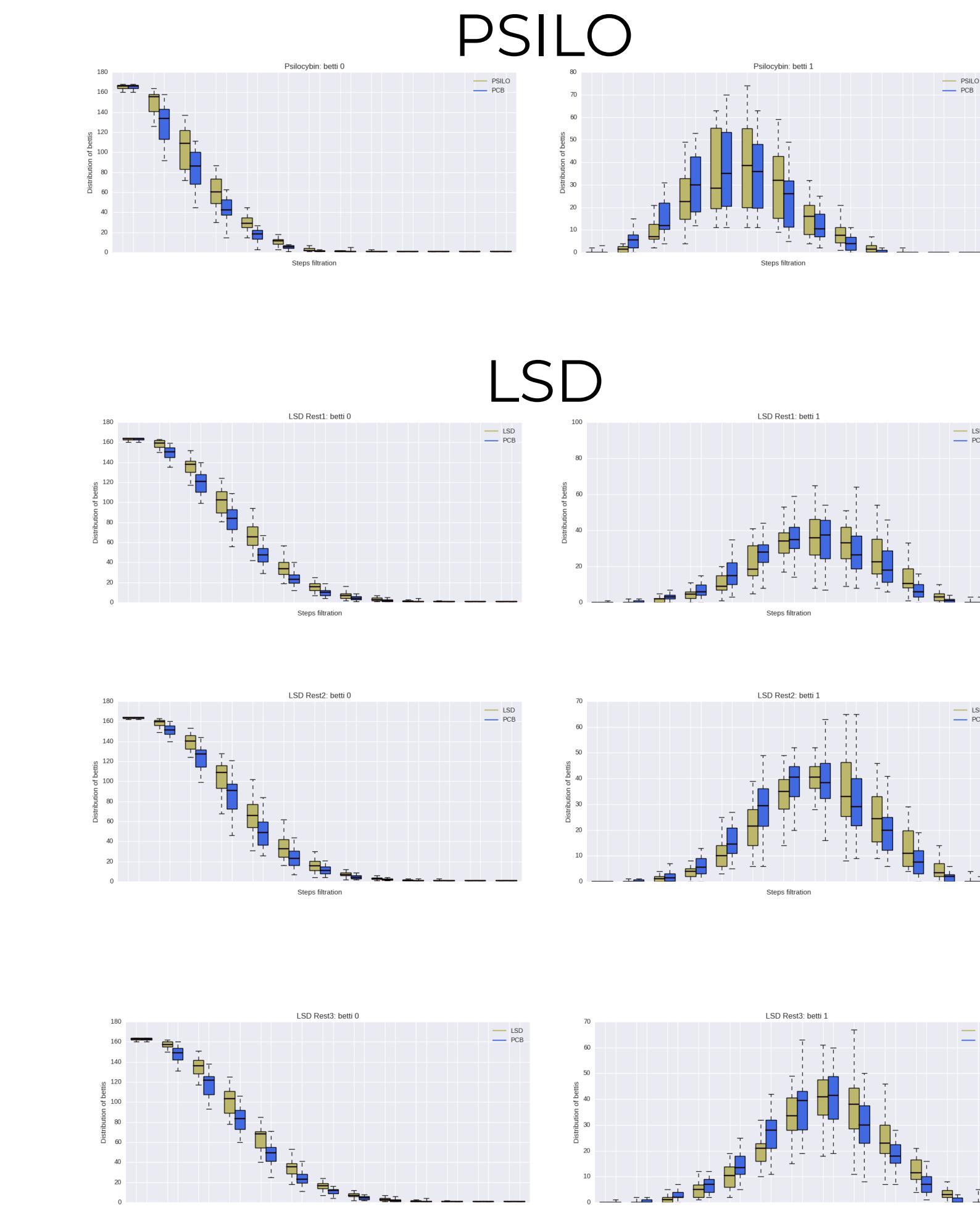
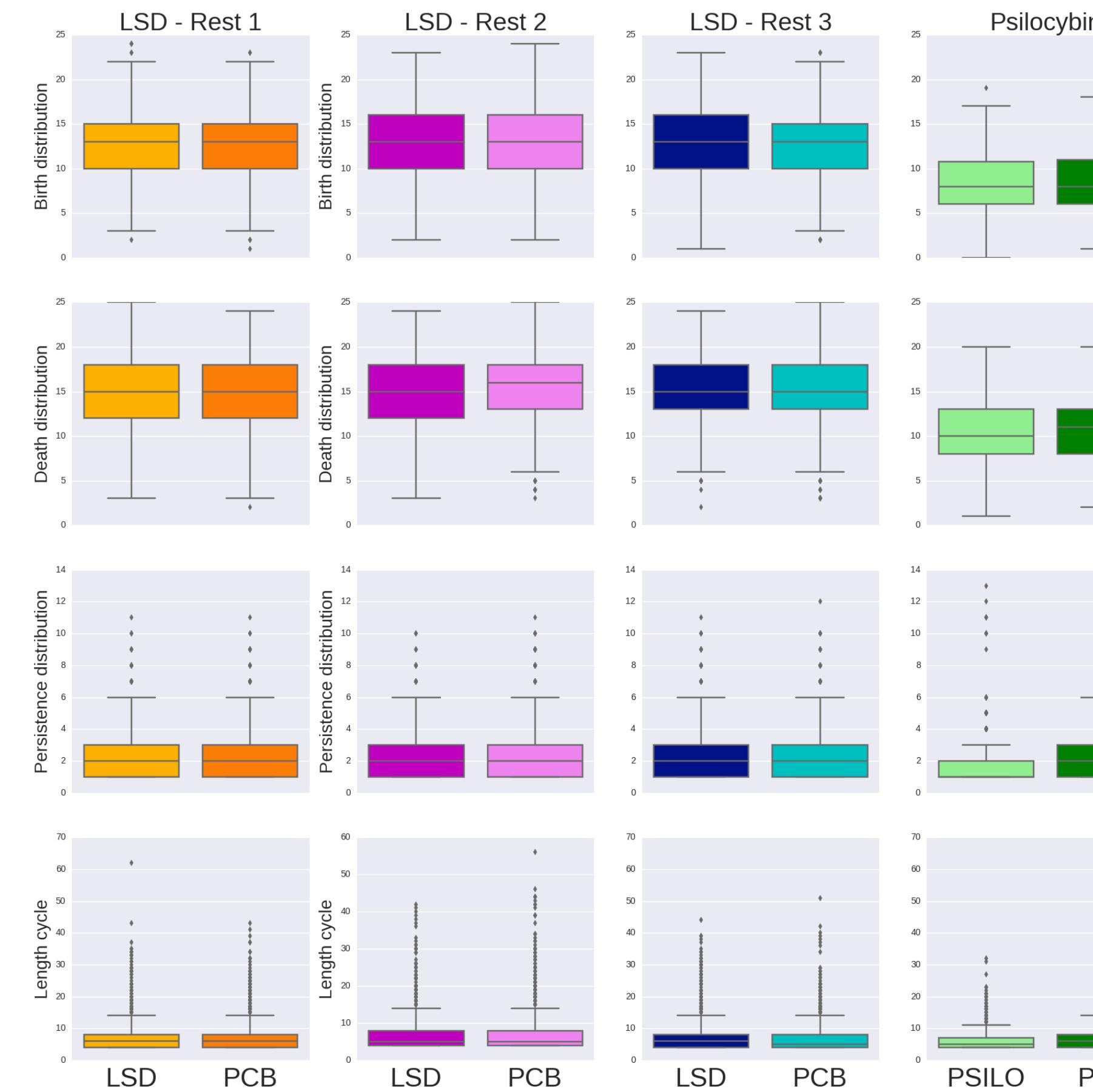
rs-fMRI  
15 subjects, 2 sessions  
1 recording condition

Carhart-Harris, Robin L., et al. "Neural correlates of the psychedelic state as determined by fMRI studies with psilocybin." *Proceedings of the National Academy of Sciences* 109.6 (2012): 2138-2143.

# Can we detect altered states?

No changes in the:

- persistences
- length of cycles
- birth/death distribution
- abundance of cycles

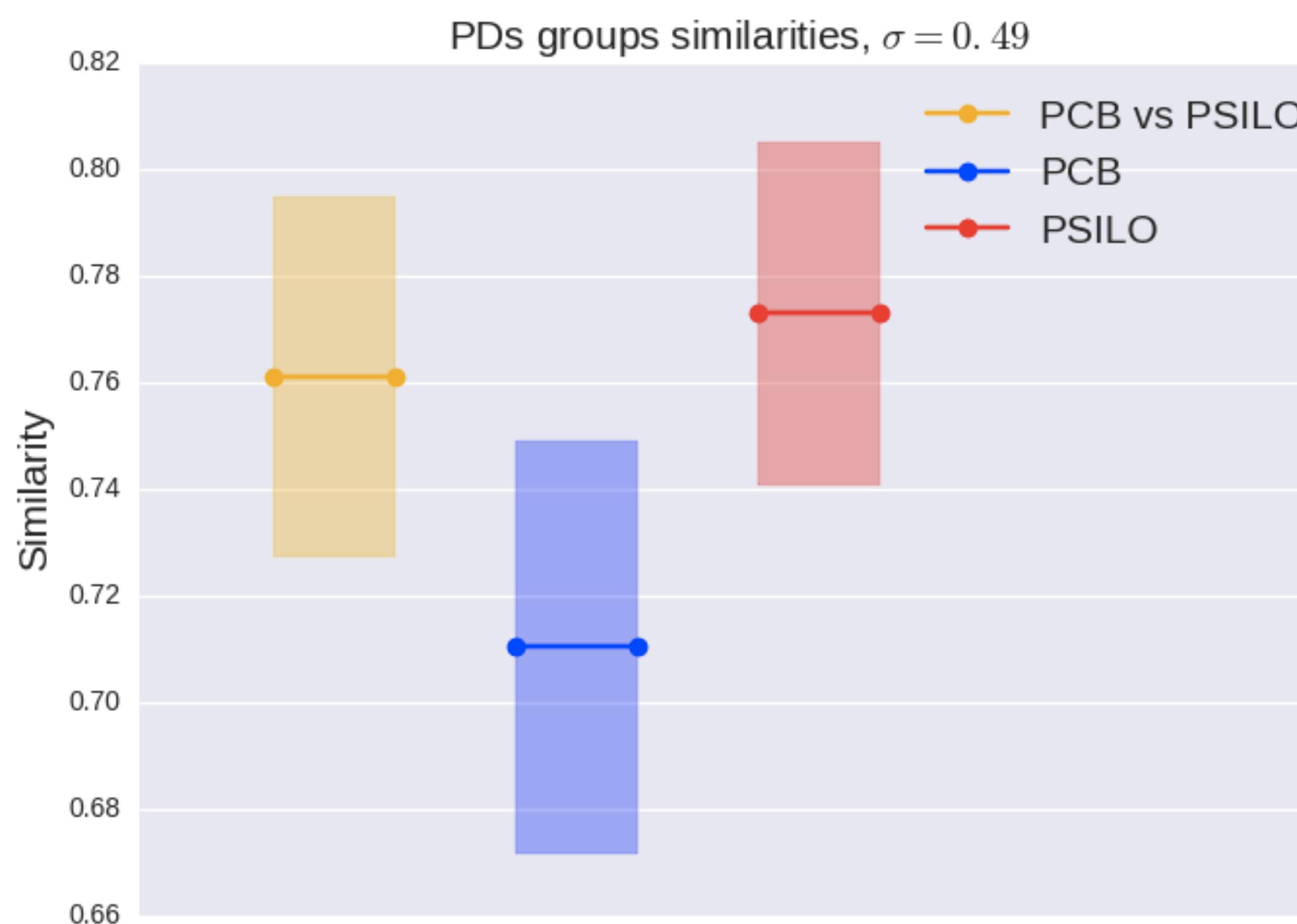


# Can we detect altered states?

**Opposite results!**

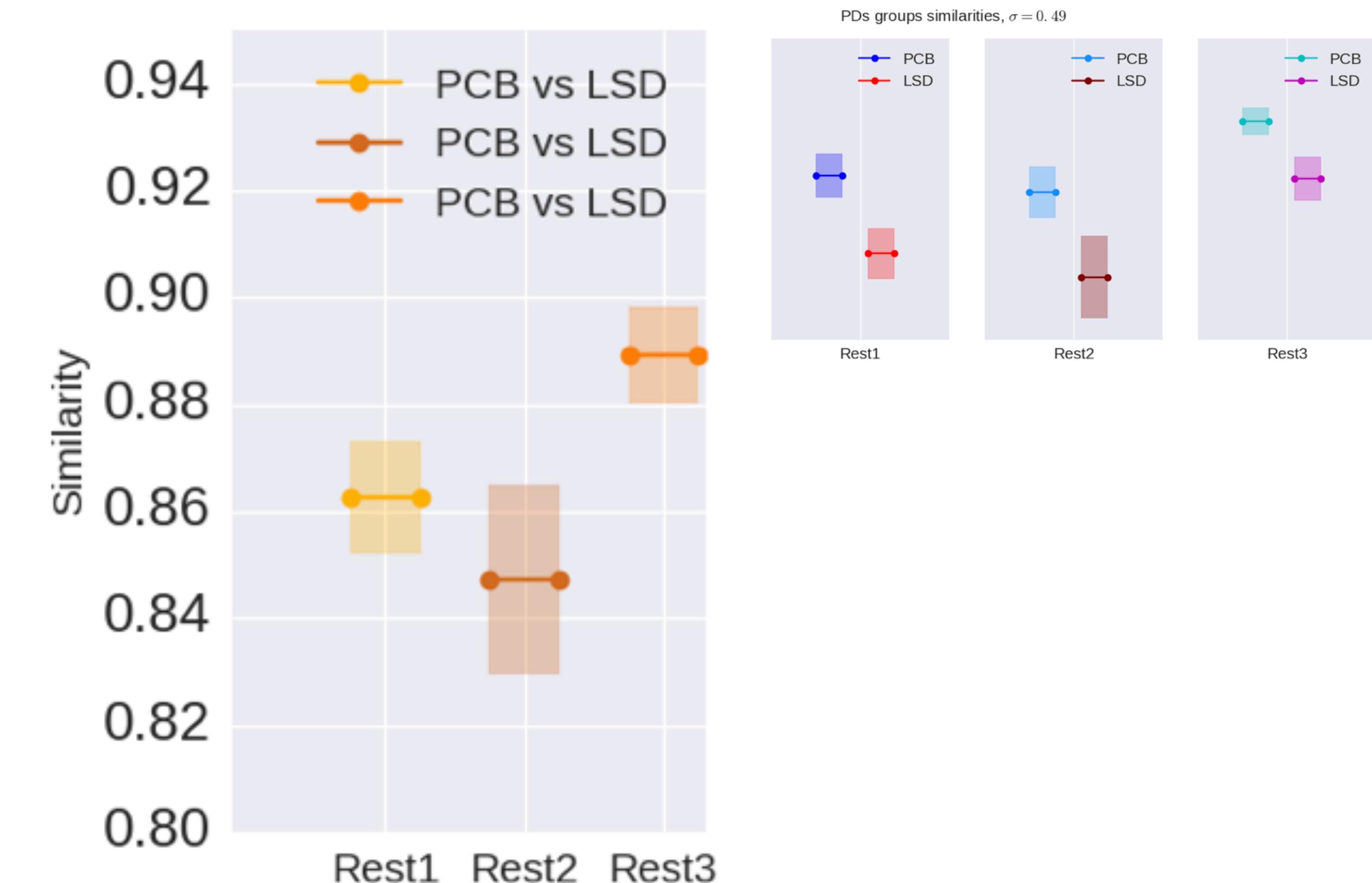
## PSILO

- Topological changes are large
- PSILO group more uniform than PCB



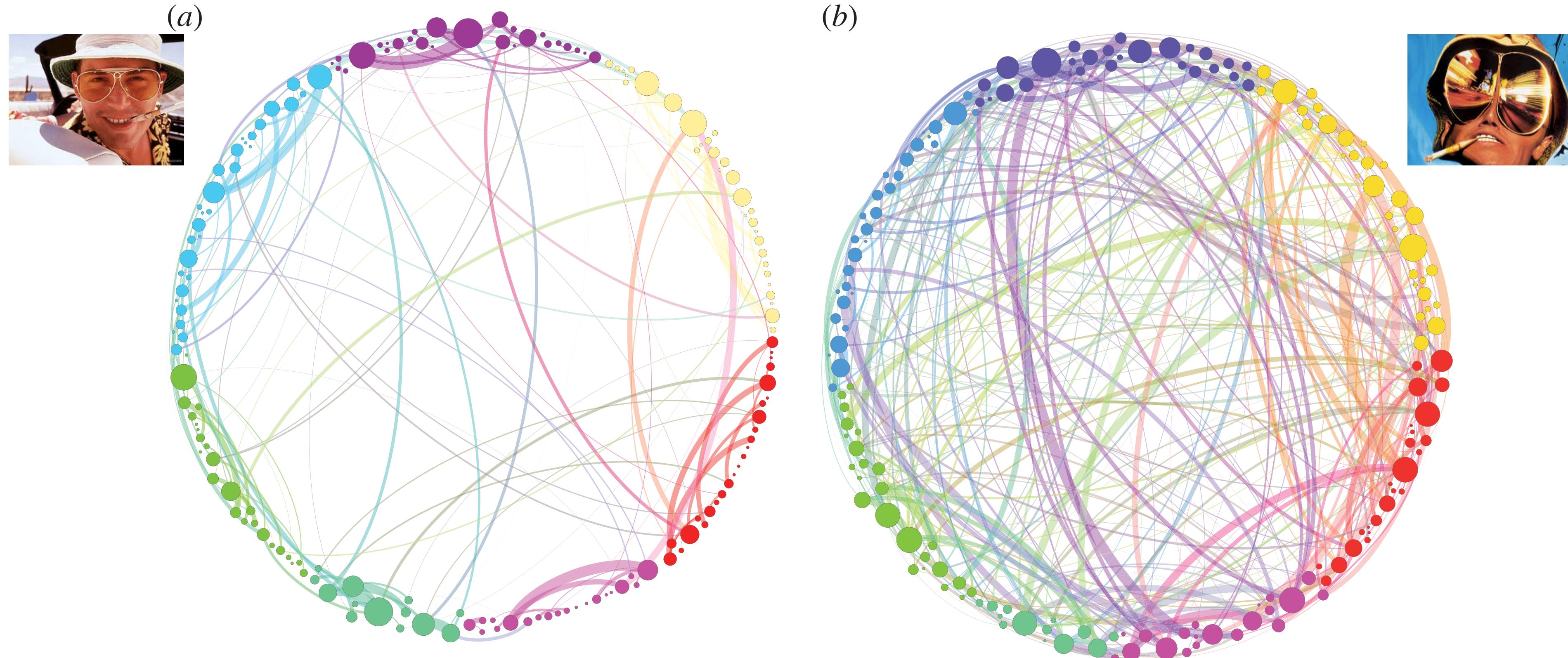
## LSD

- Topological changes are small overall
- LSD group less uniform than PCB



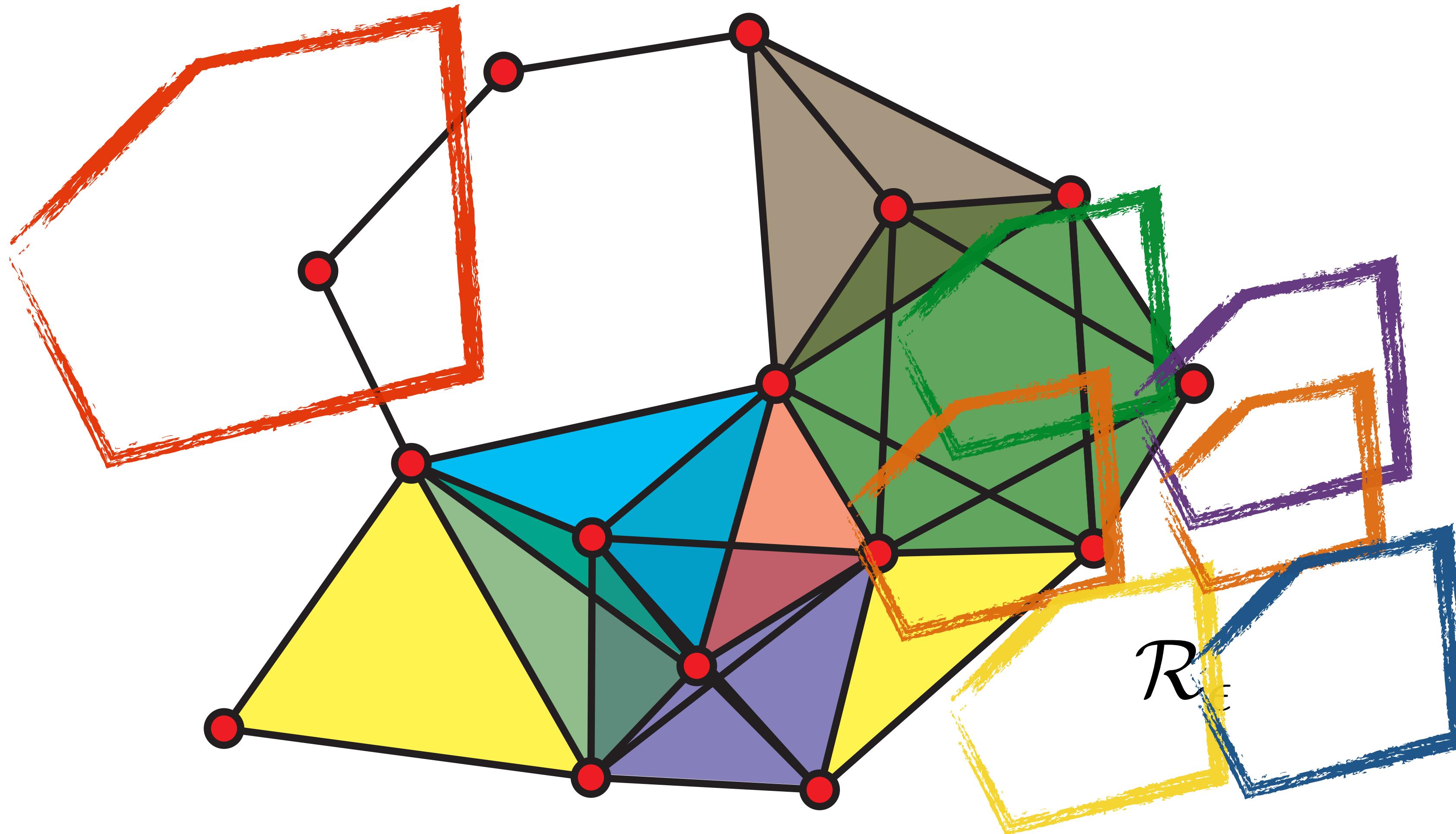
# Is it really topology?

## Psilocybin scaffolds



# Is it really topology?

Holes as shape  
indicators?



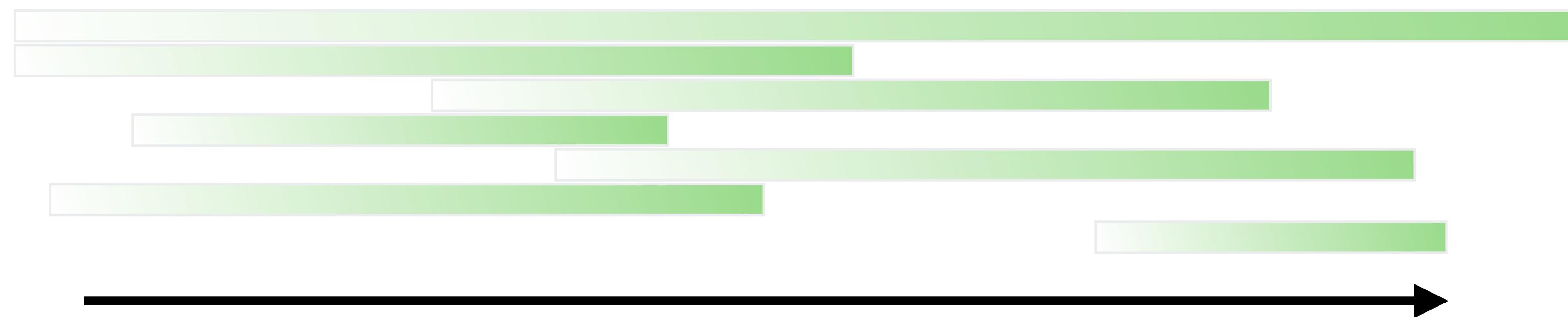
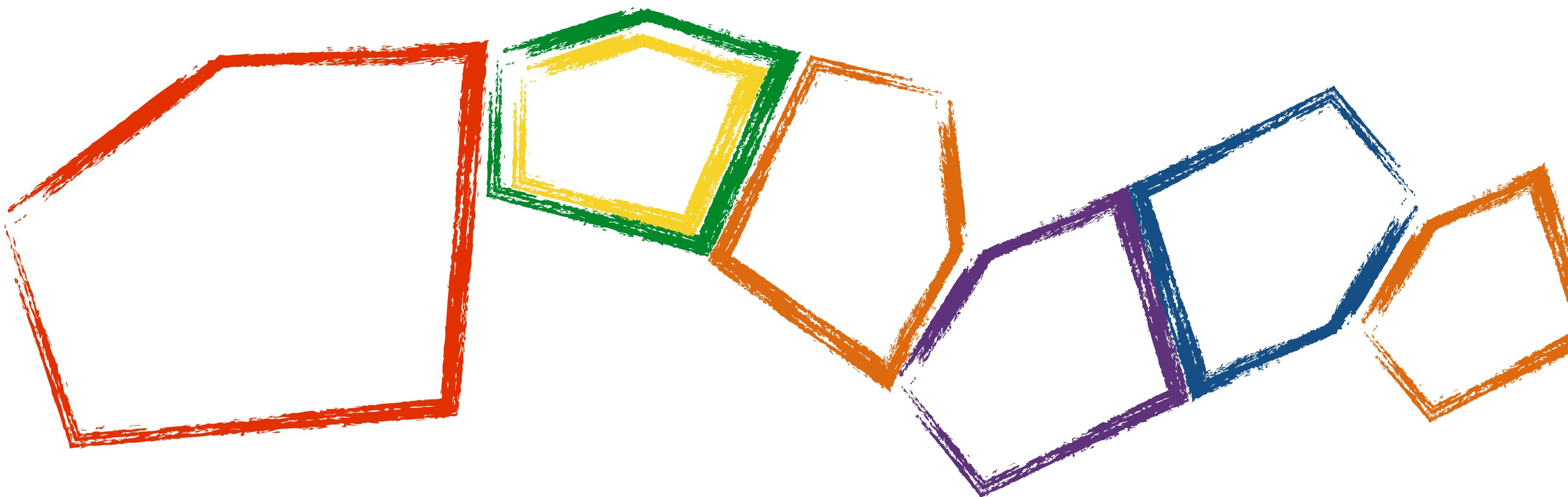
# Is it really topology?

**How does a \*man build a scaffold?**



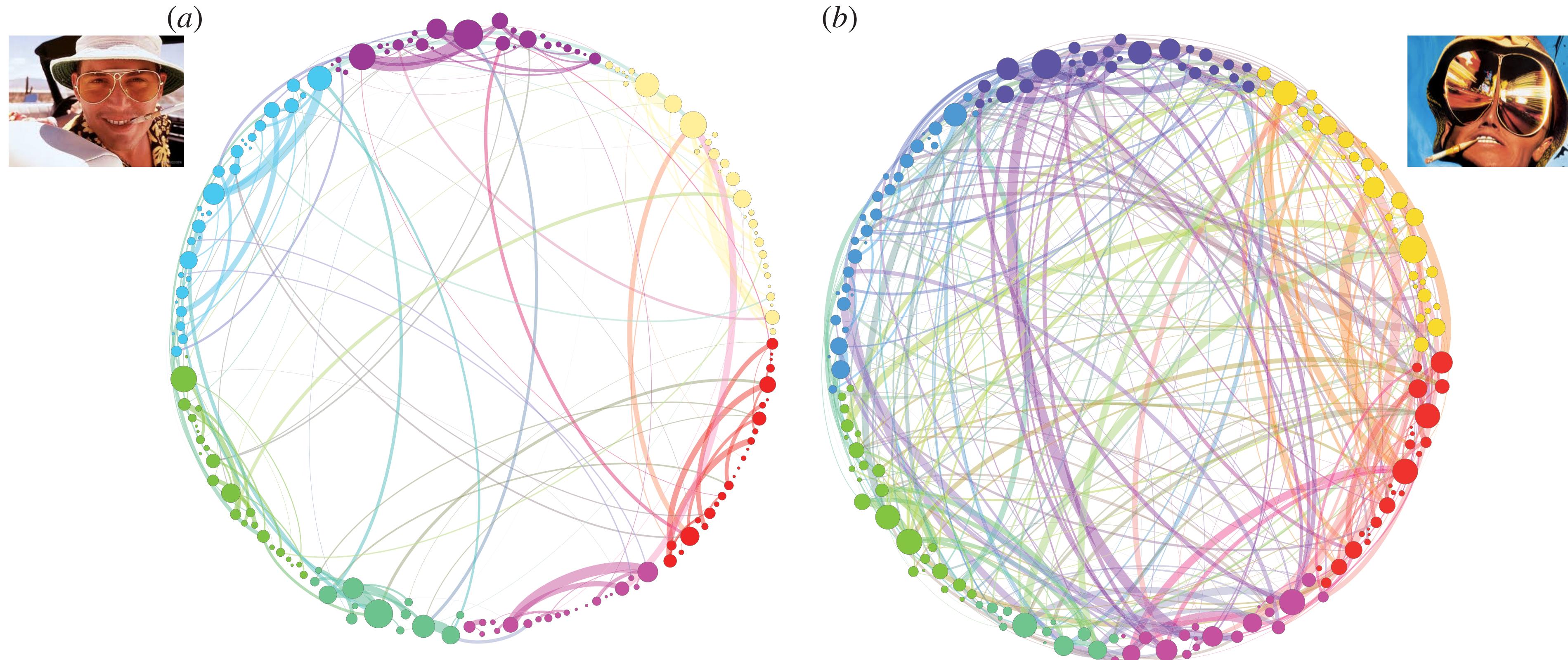
# Is it really topology?

**How does a \*man build a scaffold?**



# Is it really topology?

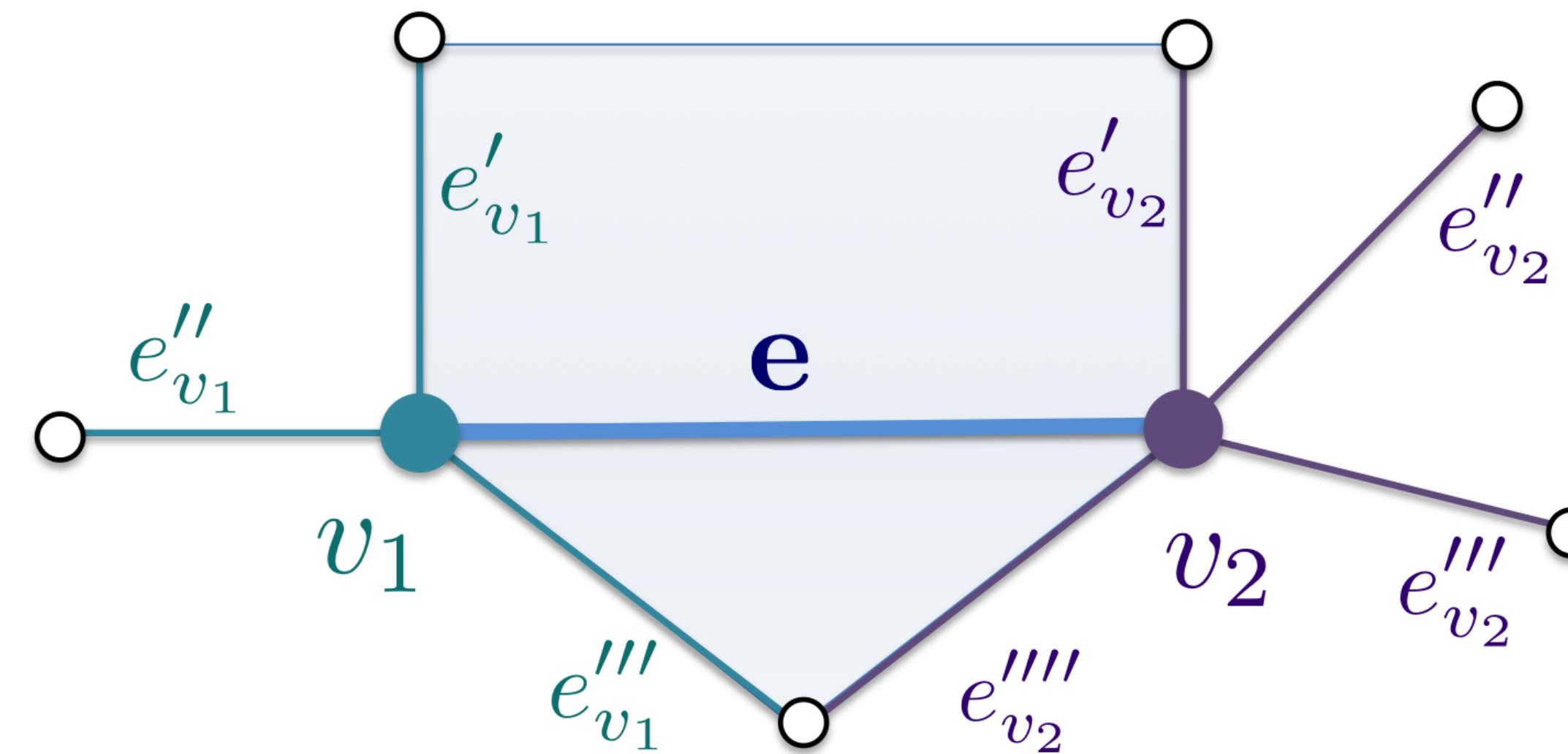
## Psilocybin scaffolds



# Is it really topology?

## Ricci-Foreman Curvature

After all these are correlations!

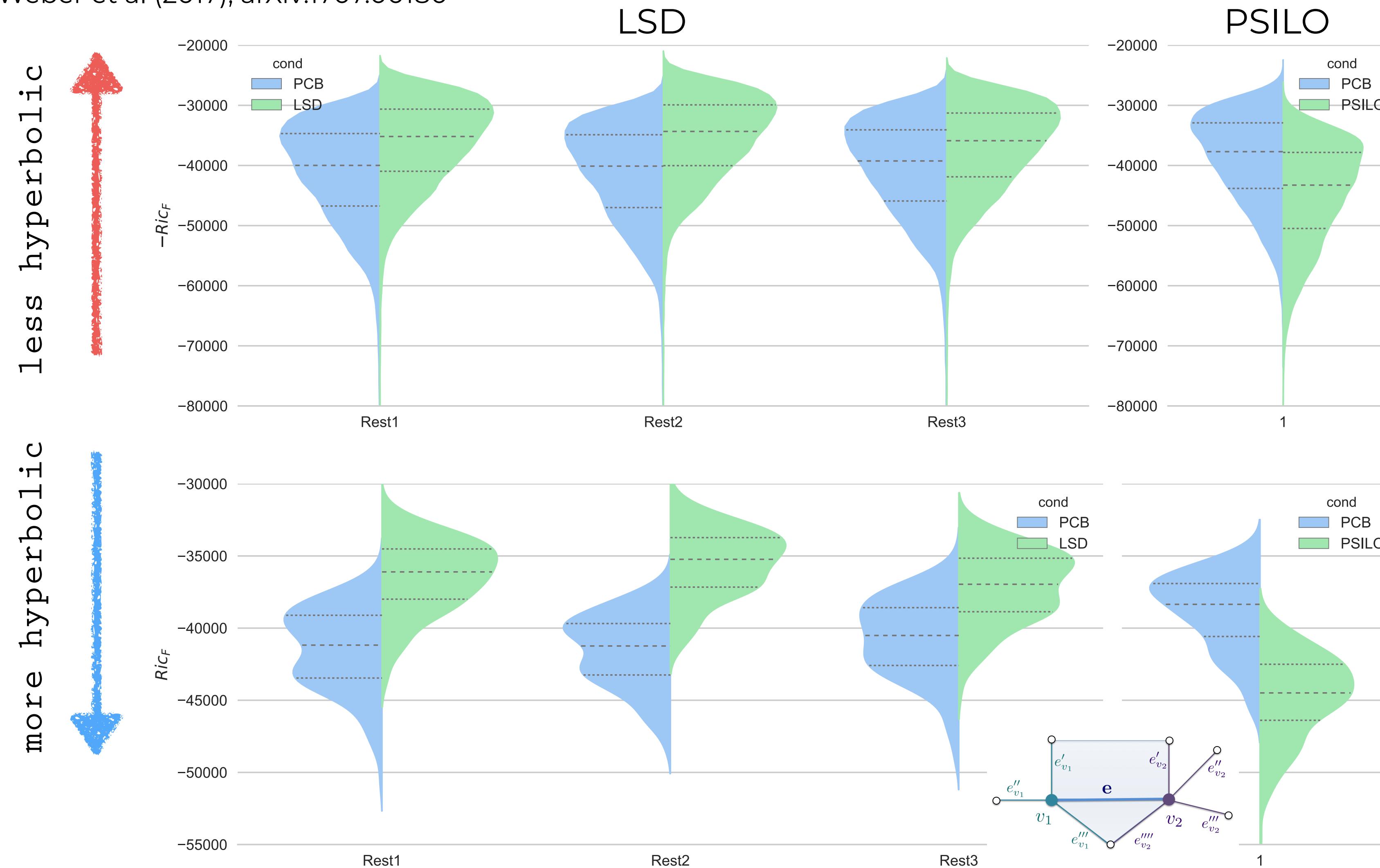


$$Ric_F(e) = w(e) \left( \frac{\omega(v_1)}{\omega(e)} + \frac{\omega(v_2)}{\omega(e)} - \sum_{e_{v_1} \sim e, e_{v_2} \sim e} \left[ \frac{\omega(v_1)}{\sqrt{\omega(e)\omega(e_{v_1})}} + \frac{\omega(v_2)}{\sqrt{\omega(e)\omega(e_{v_2})}} \right] \right)$$

# Is it really topology?

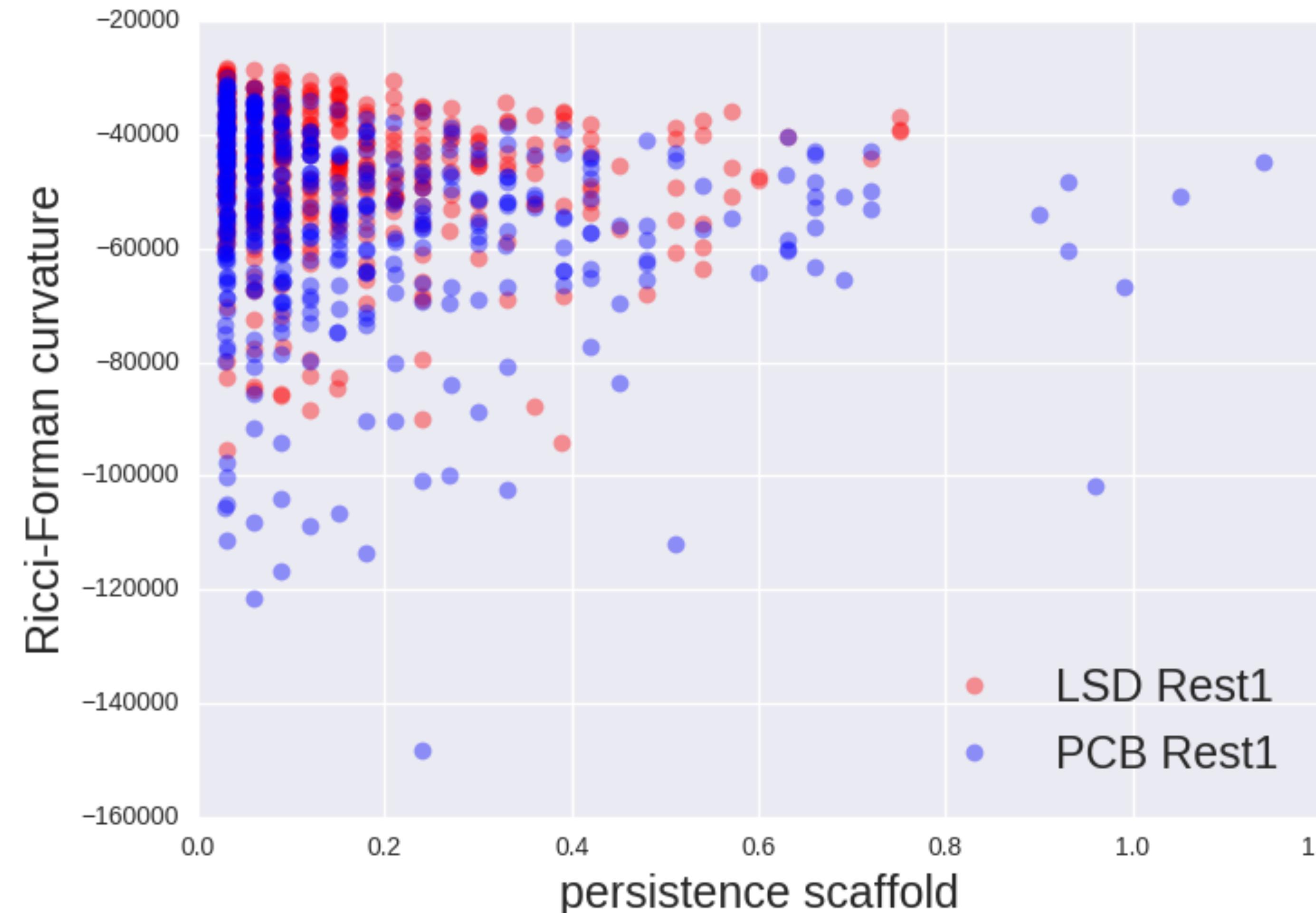
After all these are correlations!  
look at Foreman-Ricci curvature

Weber et al (2017), arXiv:1707.00180



# Is it really topology?

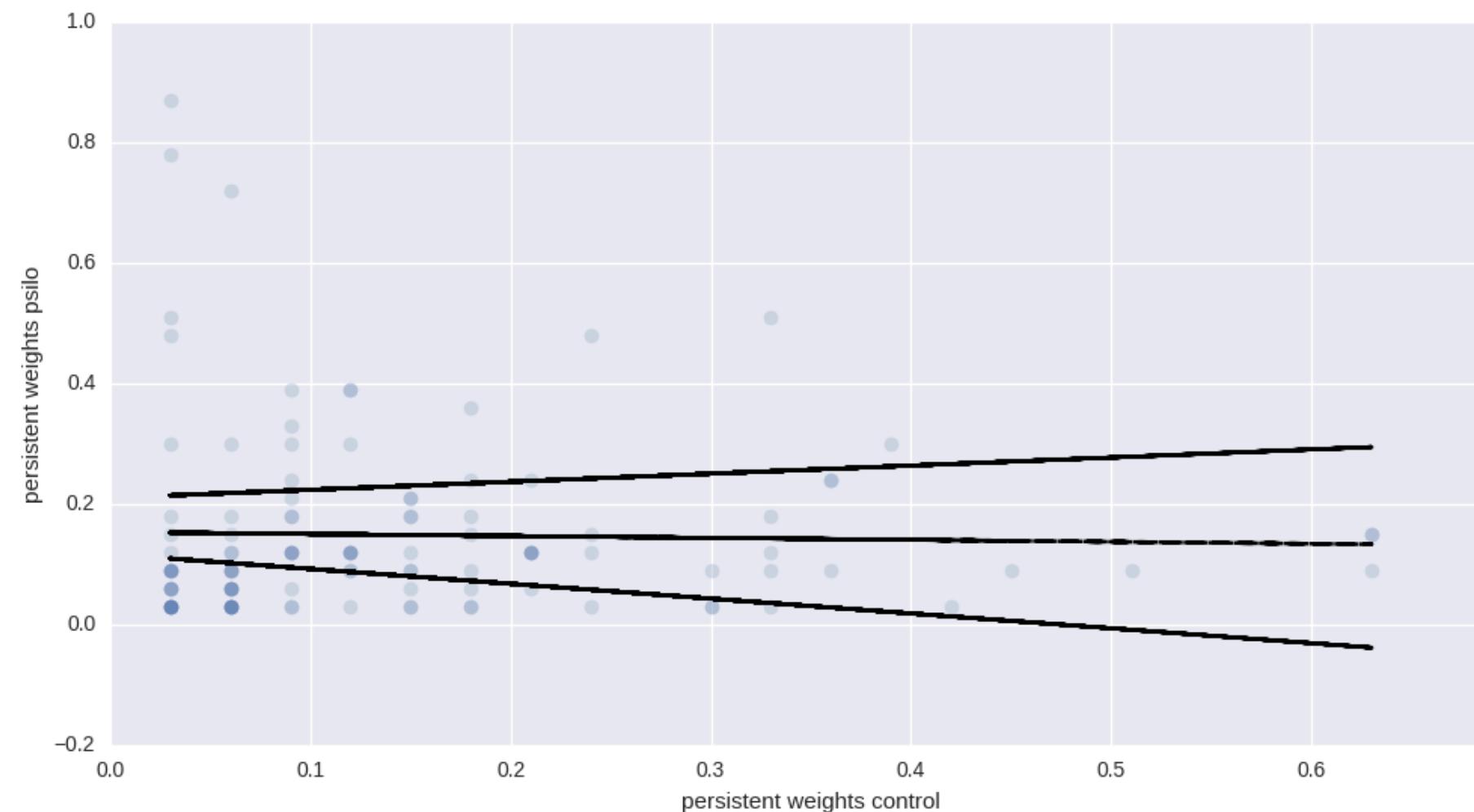
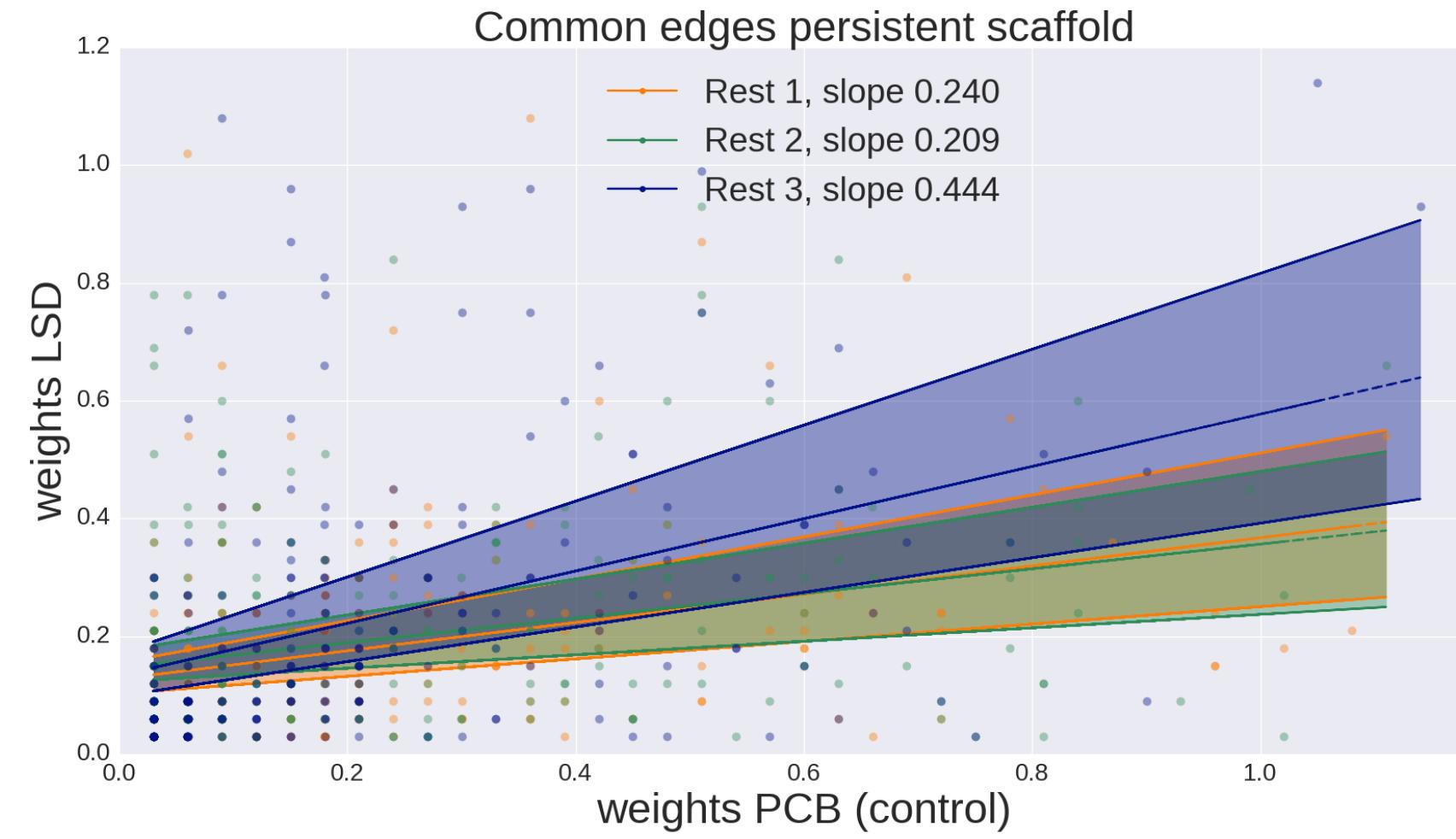
**Genuine topology!**



# Modulation

## vs disruption of topology

# Is it really topology?



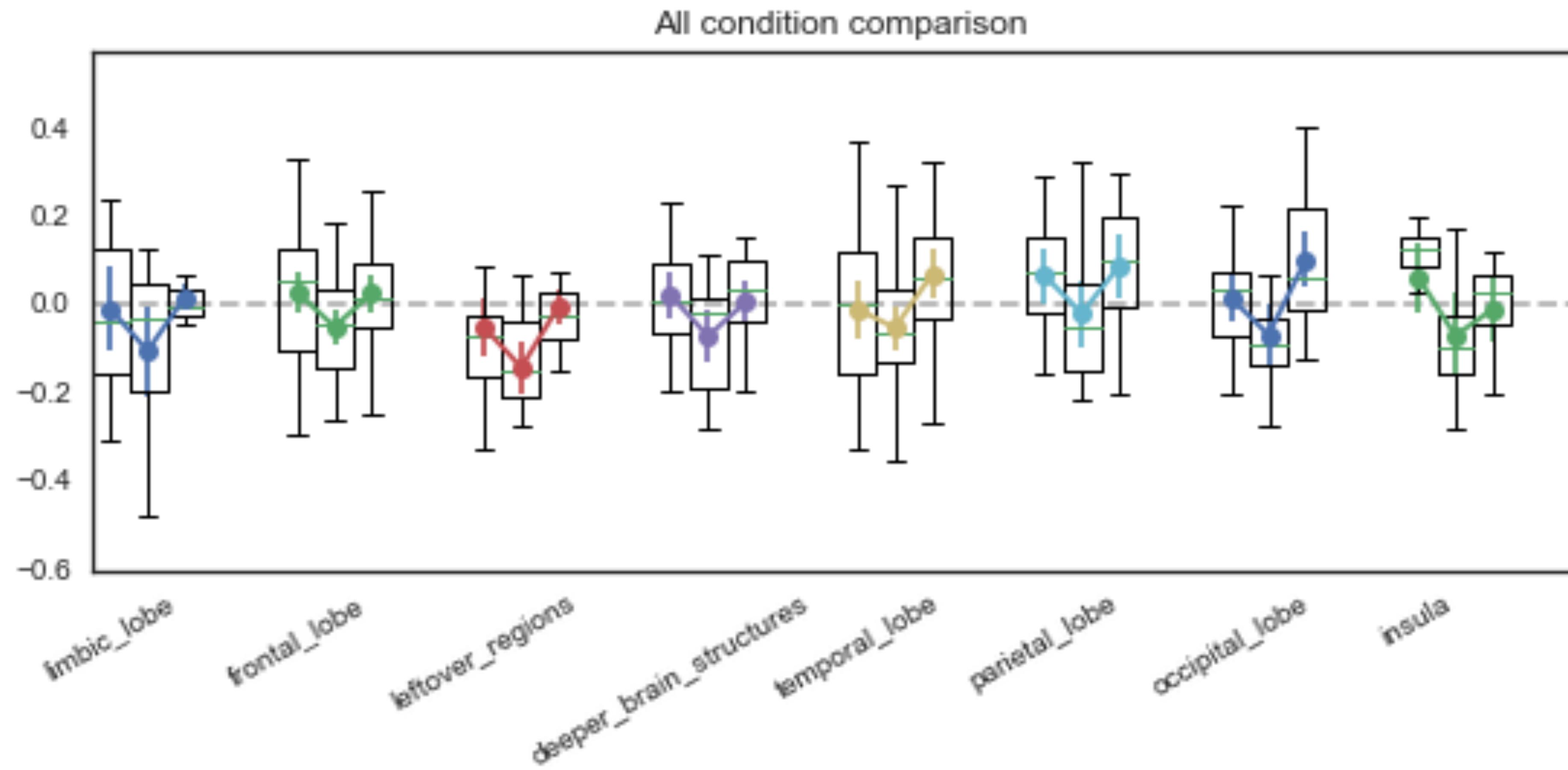
## LSD

- positive correlation
- modulation of the slope
- core of shared edges between drug and condition

## PSILO

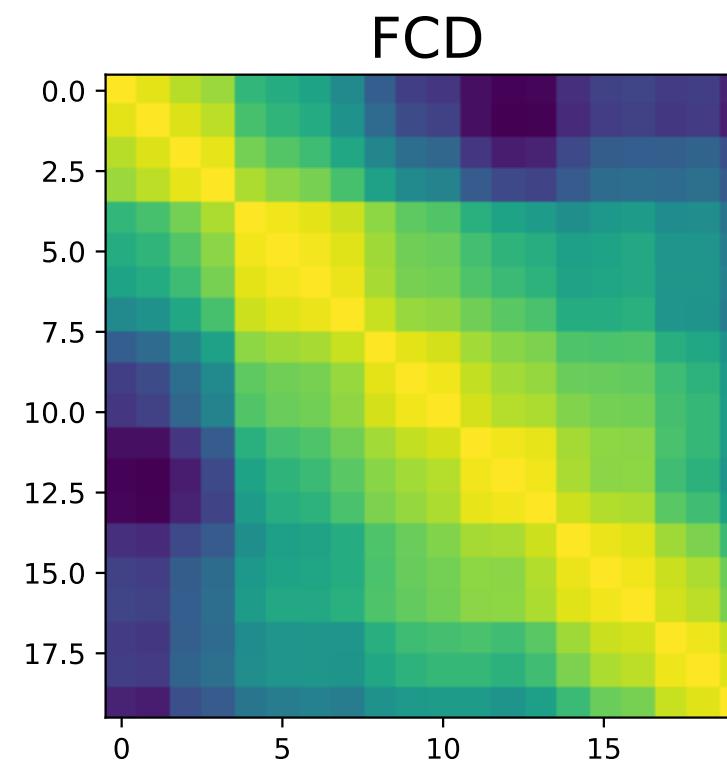
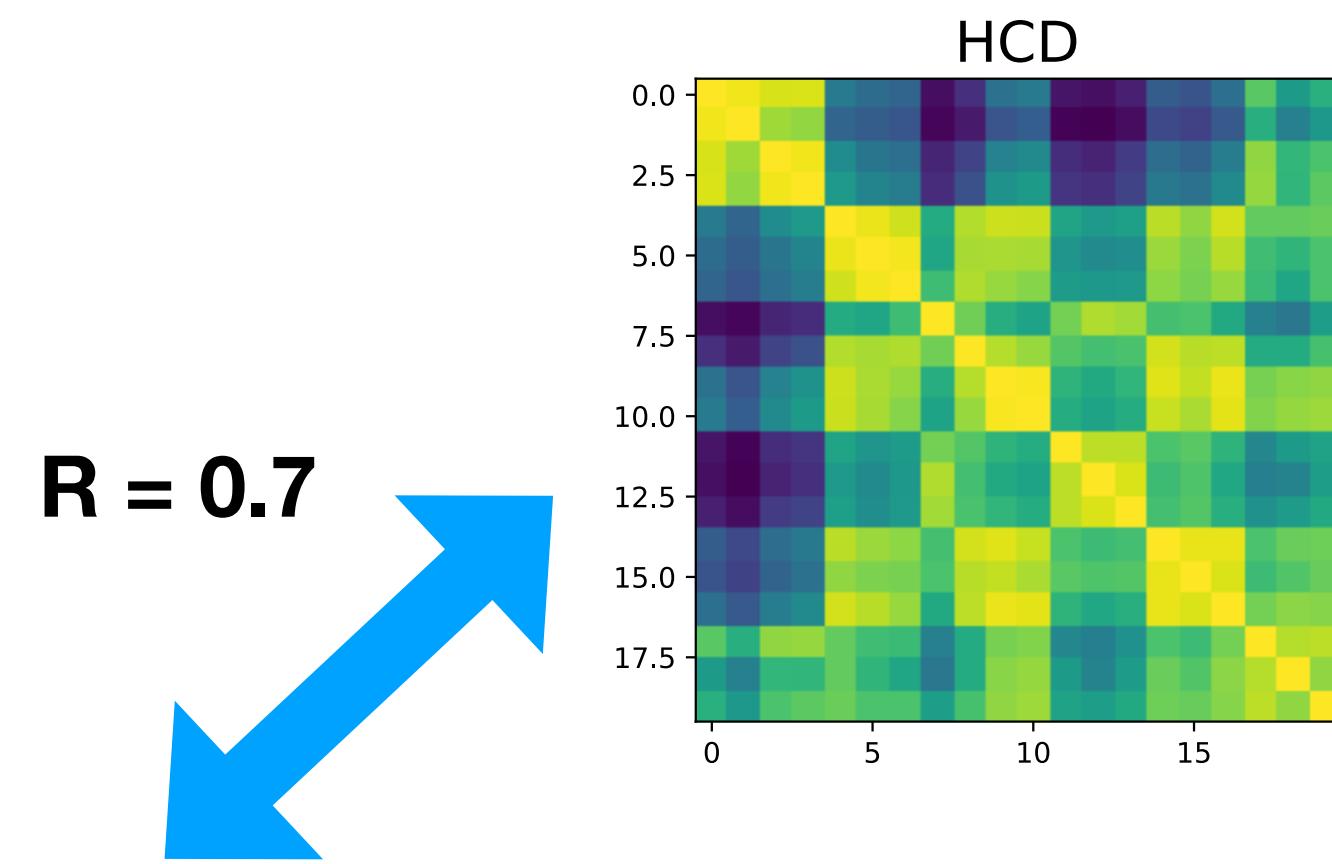
- zero/small correlation
- no overlap in edges between drug and condition
- LARGE disruption

# Is it really topology?

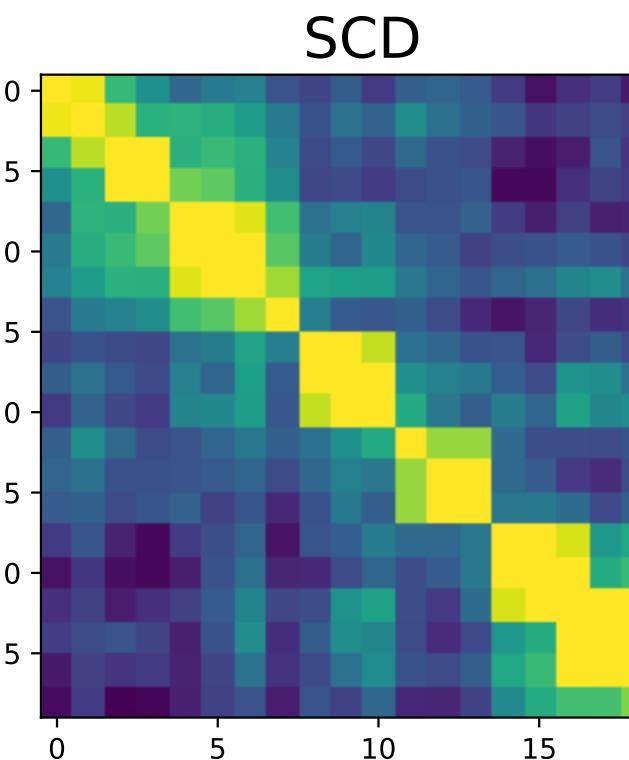


# What about temporal evolution?

Homological distances



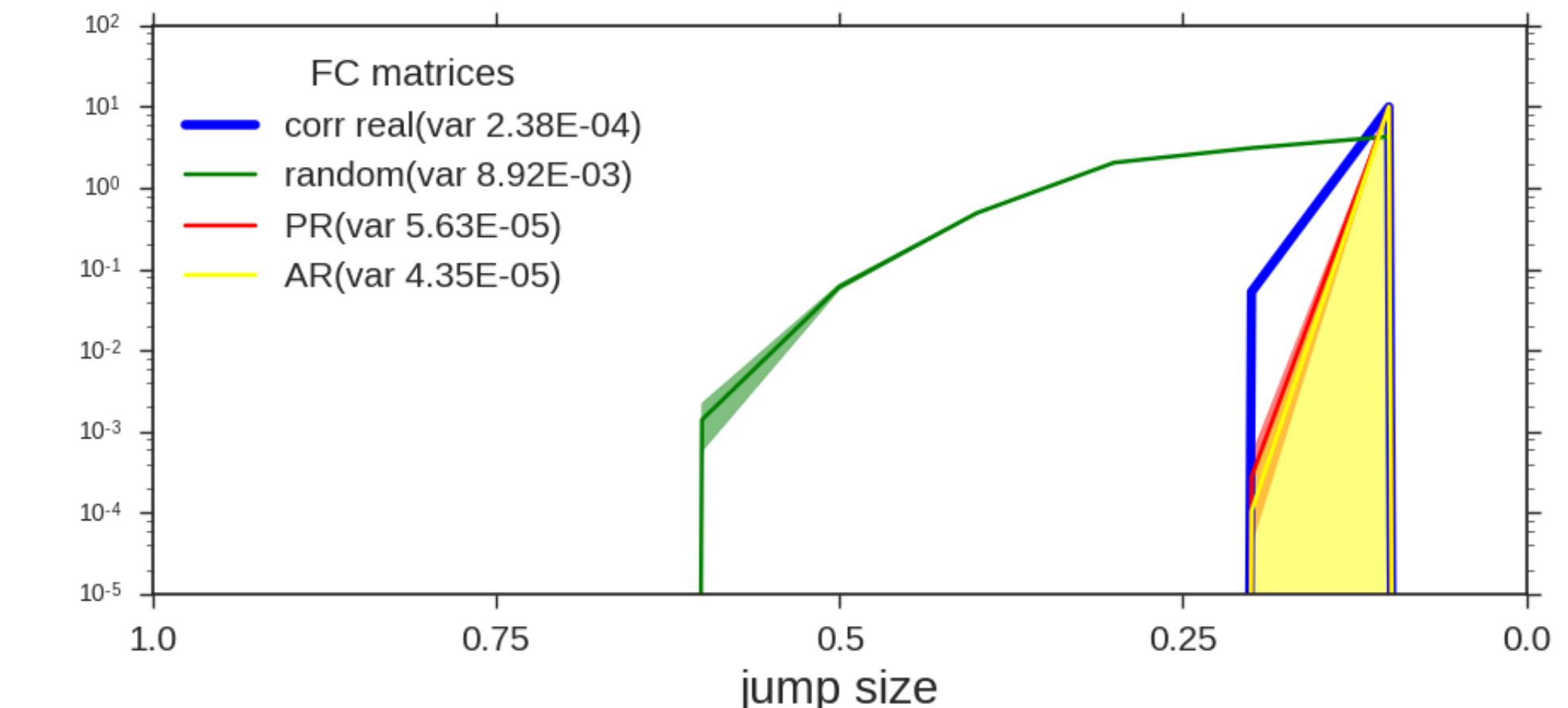
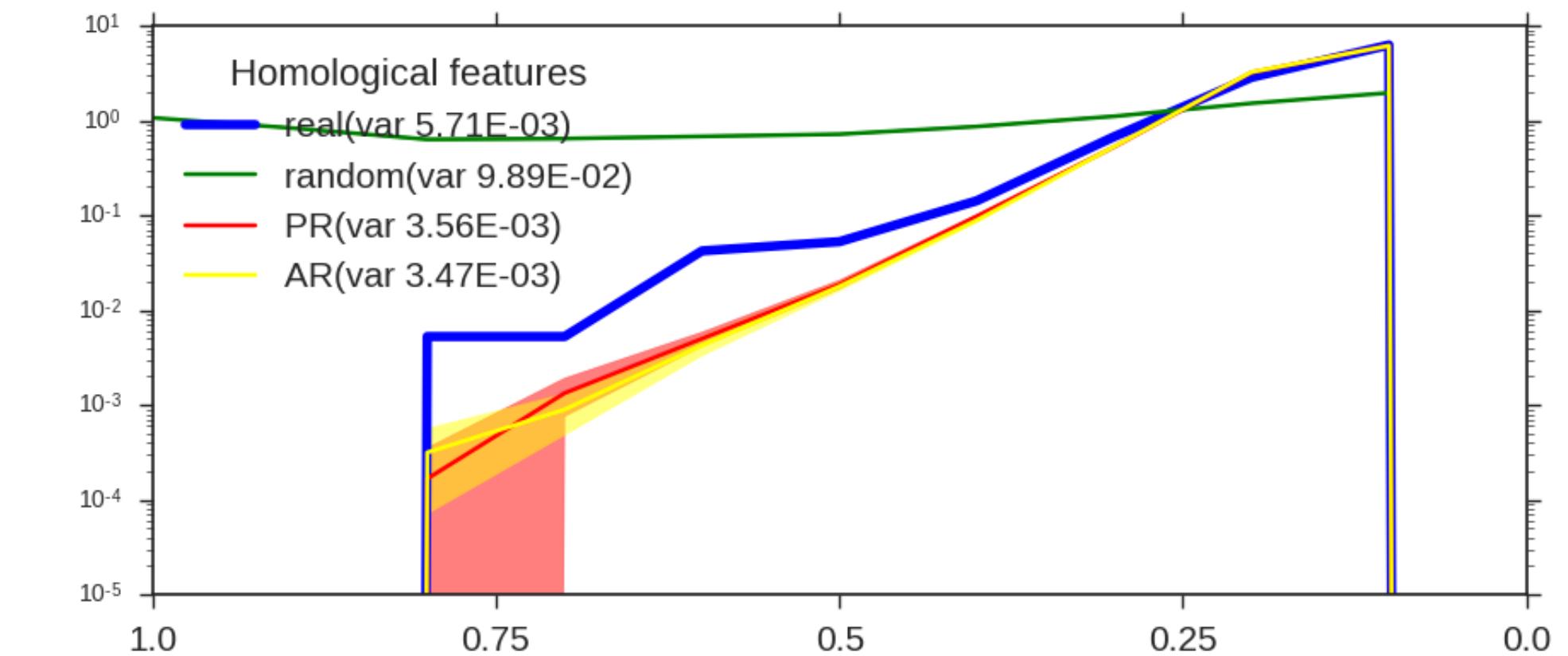
$R = 0.76$



Functional network distances

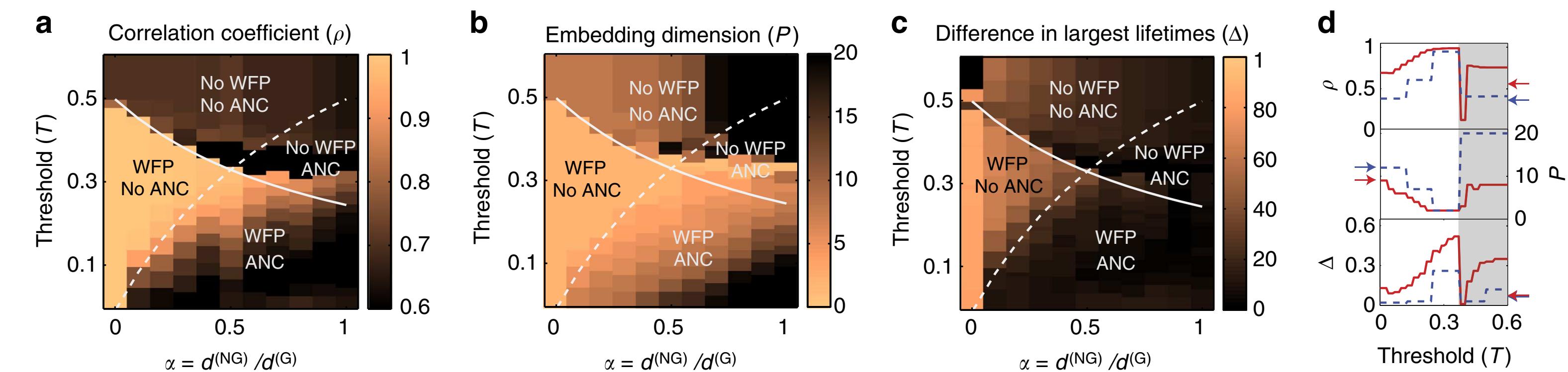
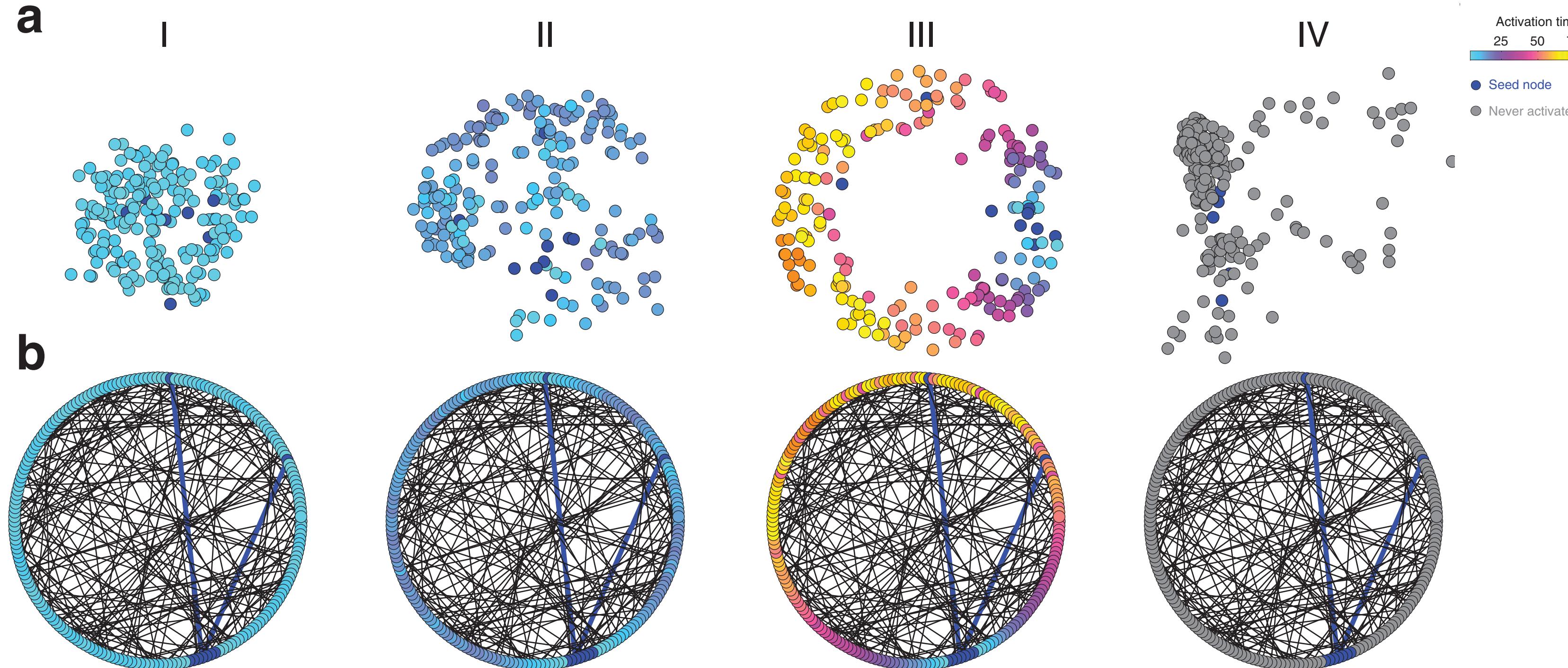
Scaffold distances

Temporal jumps distribution ( $\delta = 400, \text{shift}:40$ )



# Does the shape of activity match the structure?

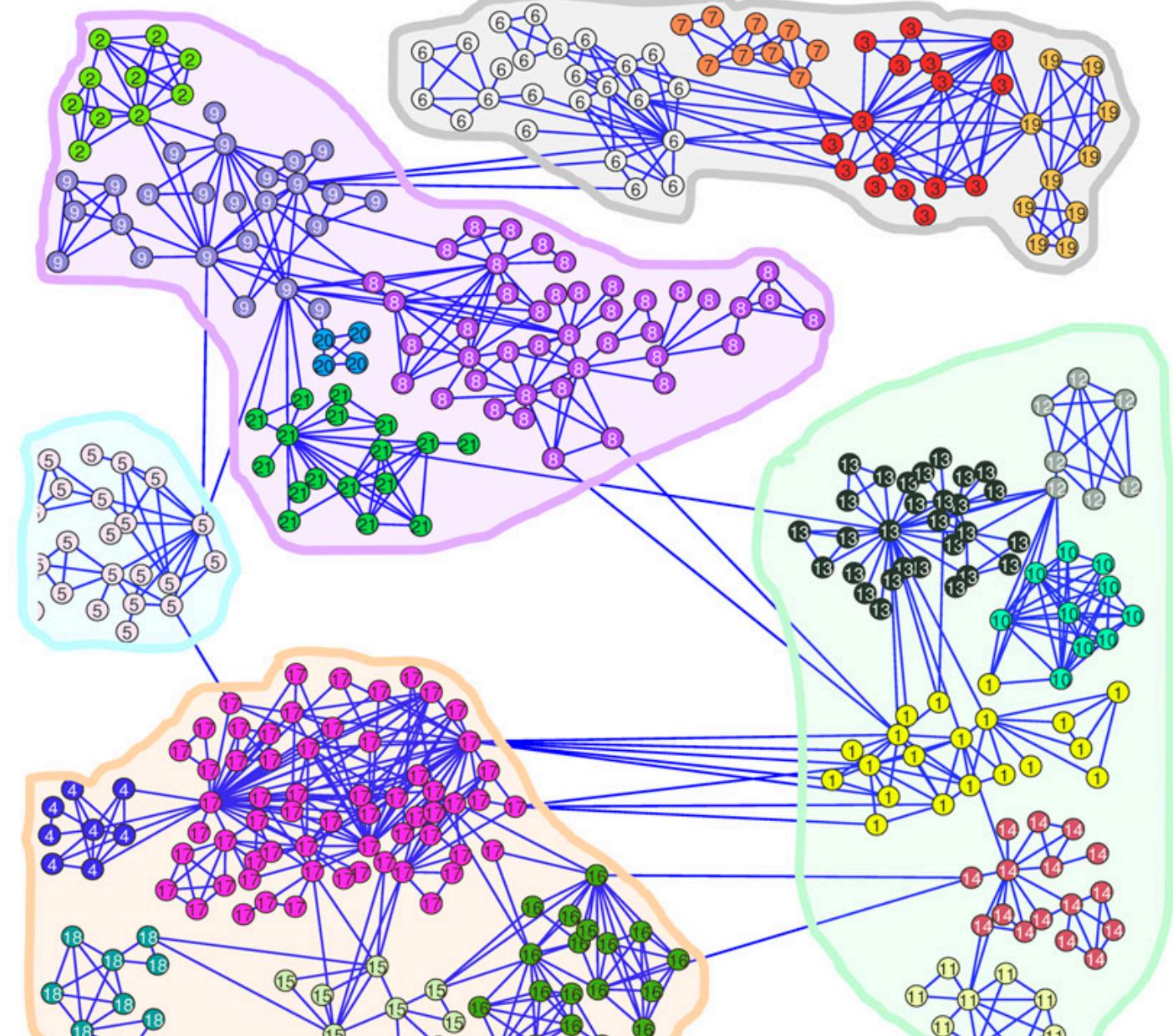
Taylor, Dane, et al. "Topological data analysis of contagion maps for examining spreading processes on networks." *Nature communications* 6 (2015): 7723.



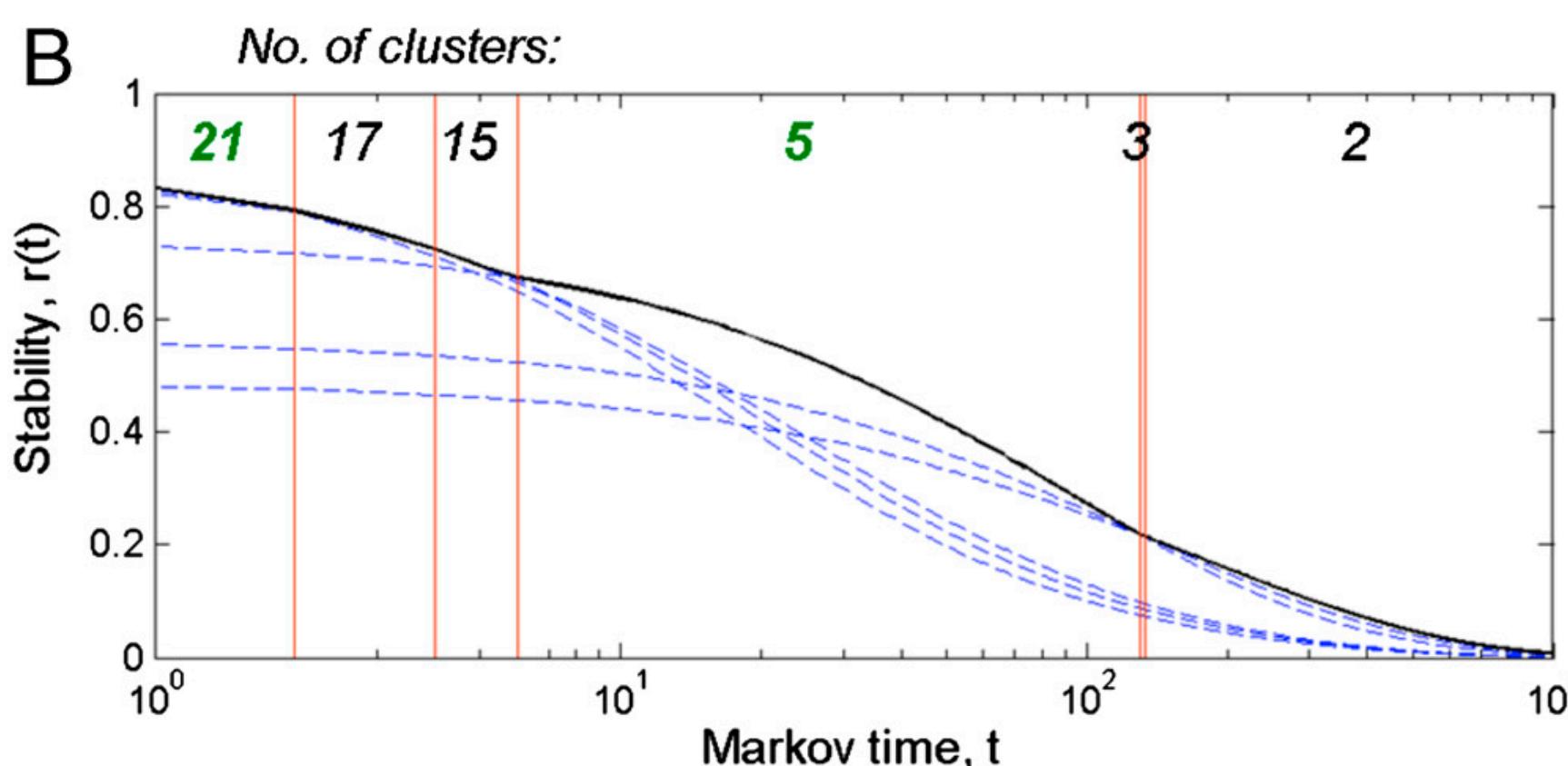
# Does the shape of activity match the structure?

Tran, Quoc Hoan, Van Tuan Vo, and Yoshihiko Hasegawa. "Scale-variant topological information for characterizing complex networks." *arXiv preprint arXiv:1811.03573* (2018).

A



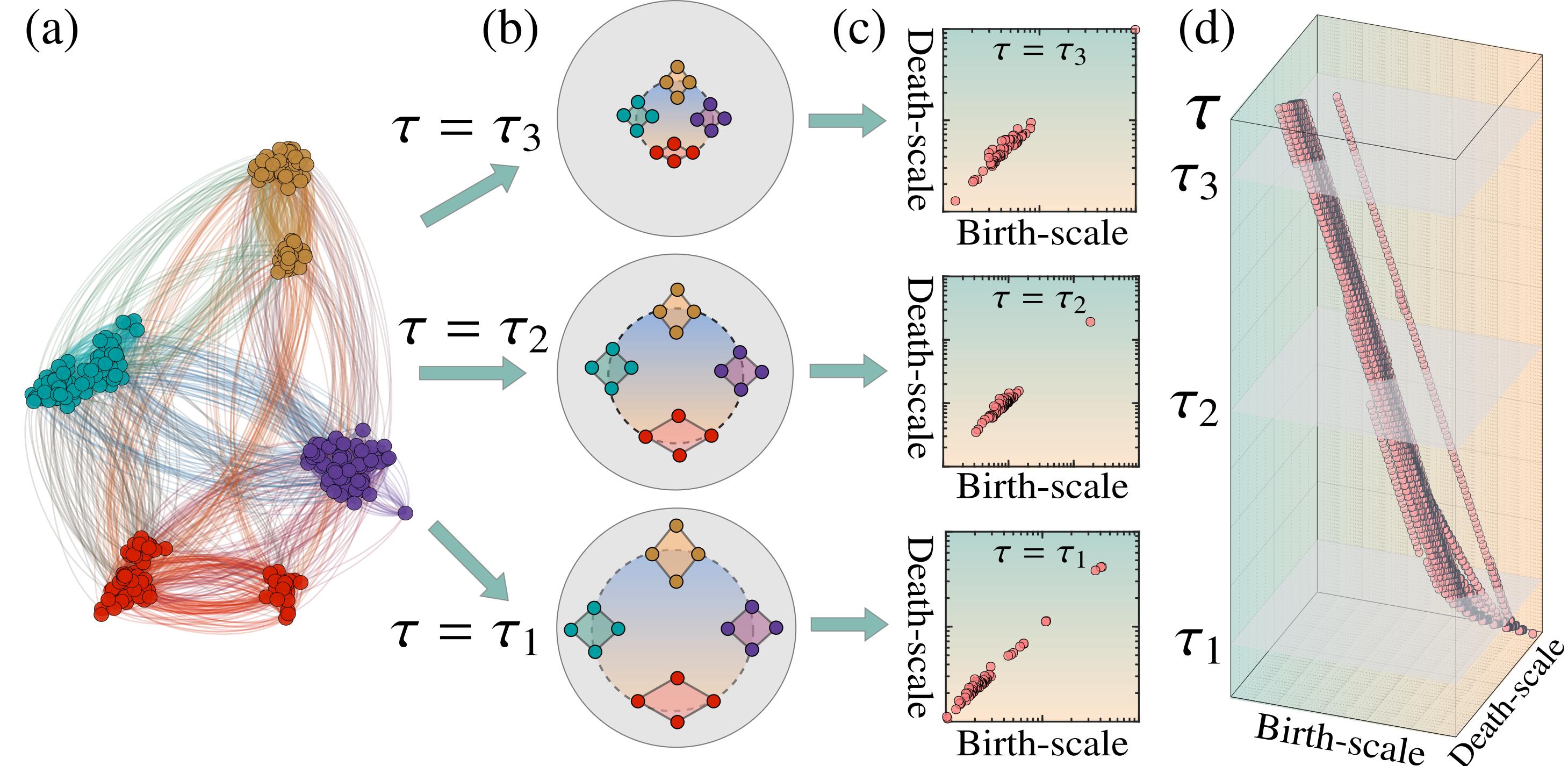
B



Stability of graph communities across time scales

J.-C. Delvenne<sup>a</sup>, S. N. Yaliraki<sup>a,b</sup>, and M. Barahona<sup>a,c,1</sup>

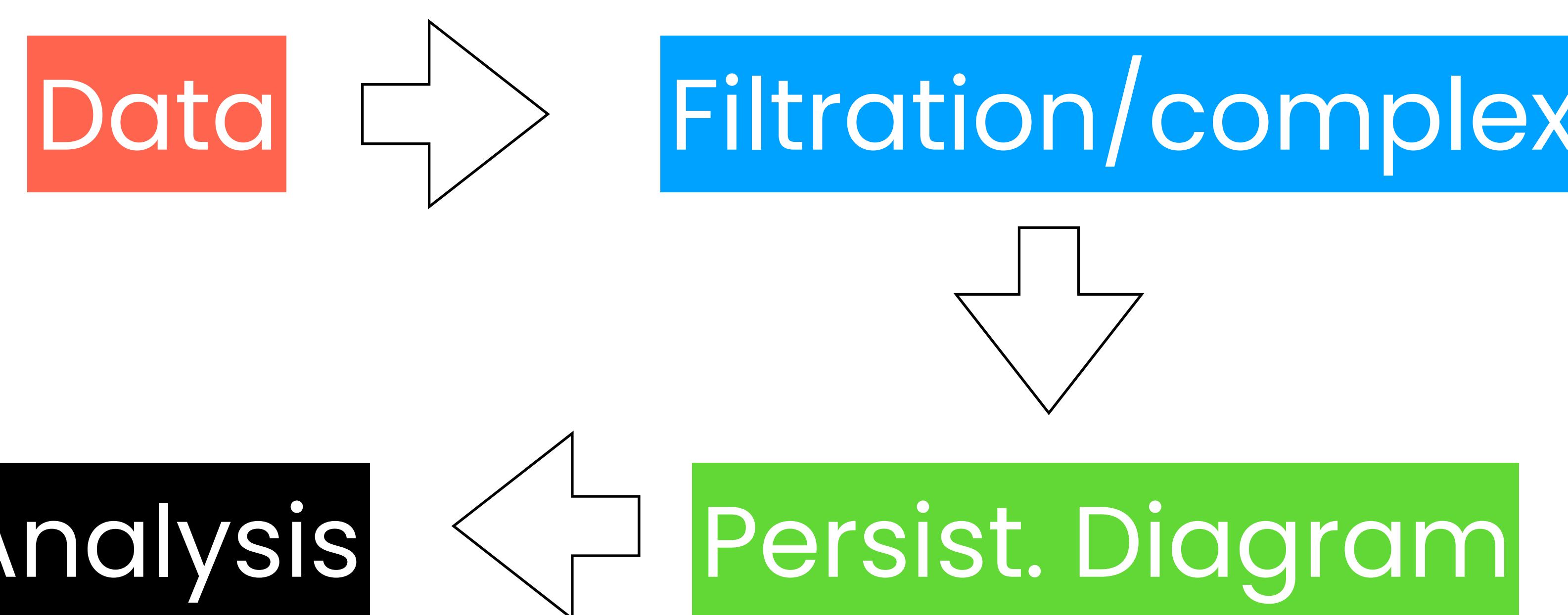
(a)



# Randomization

+

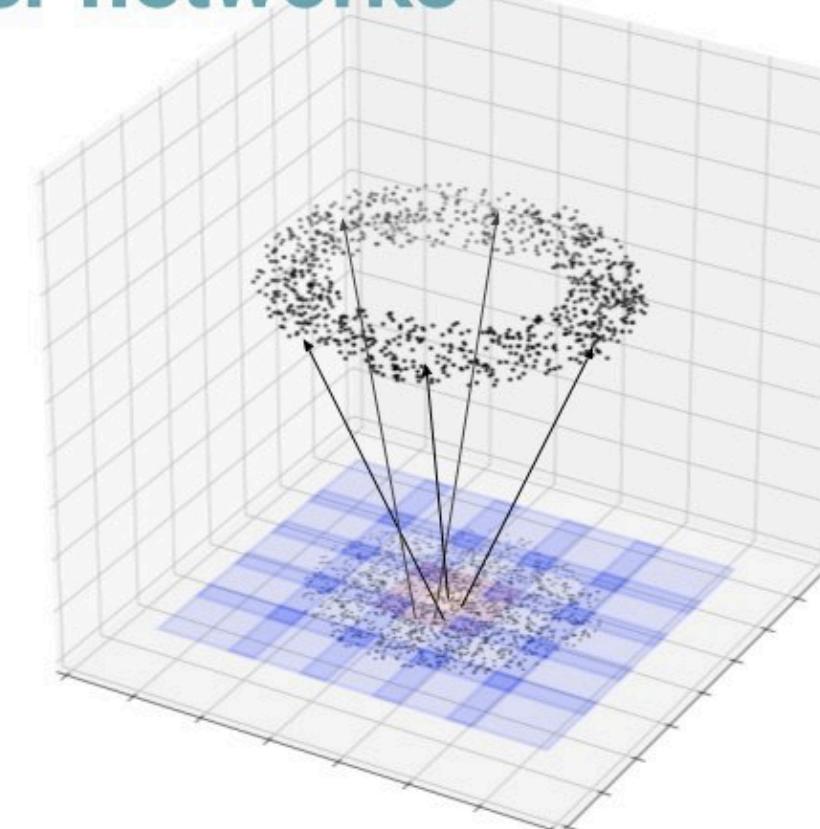
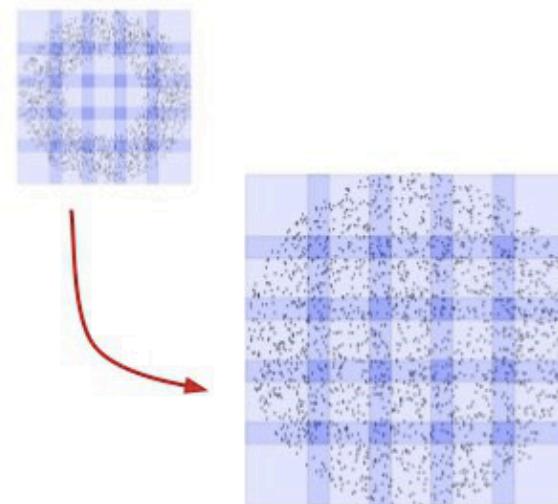
## validation



# Randomization?

## Randomizing Mapper networks

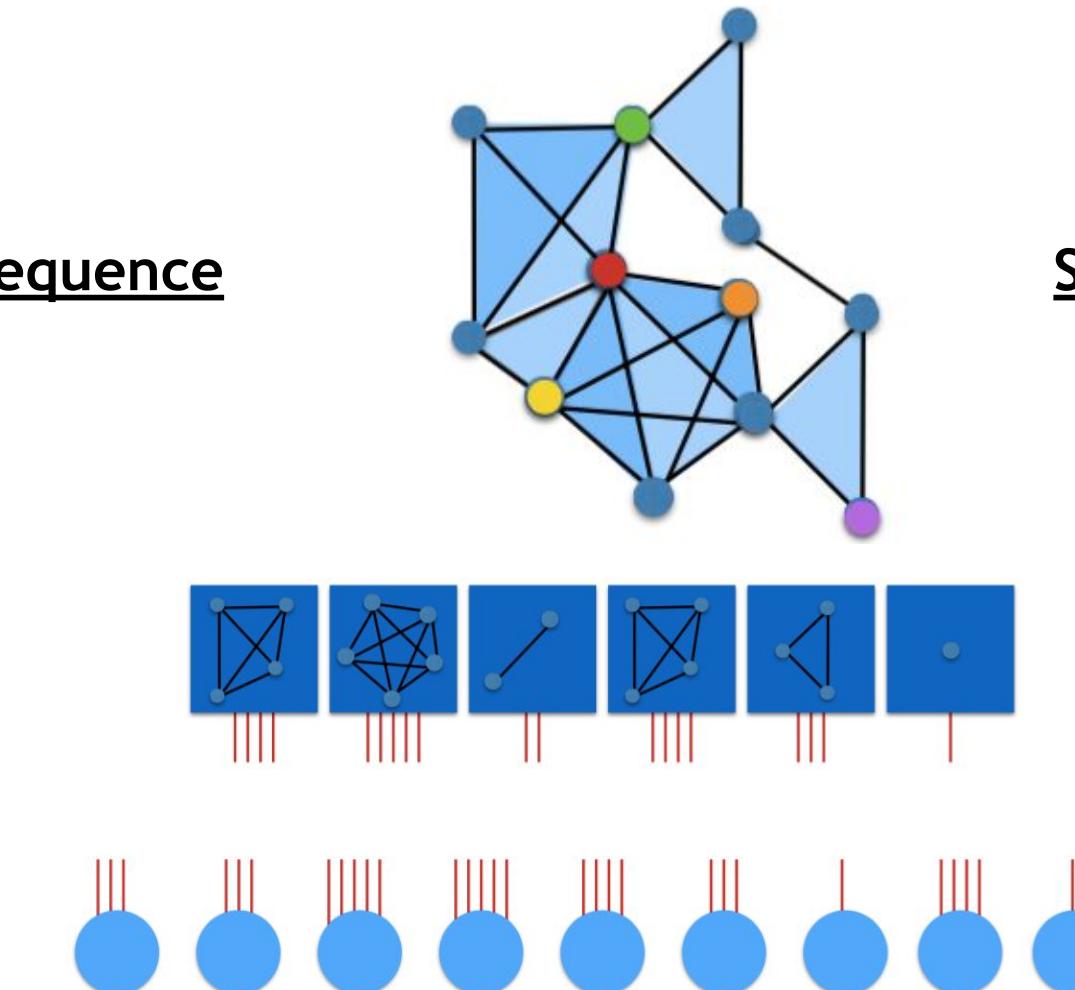
Randomizing the filter parameters:



Filter  
randomization

## Facets size sequence

$s_j$
4
5
3
3
3
2

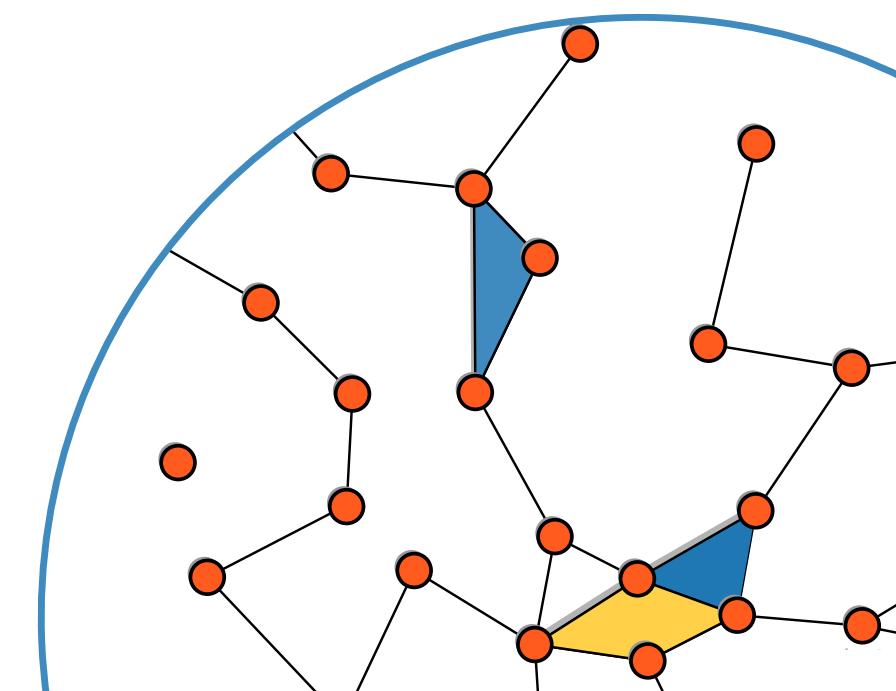


## Simplicial degree sequence

$k_i$
1
3
1
2
2

Simplicial  
randomization

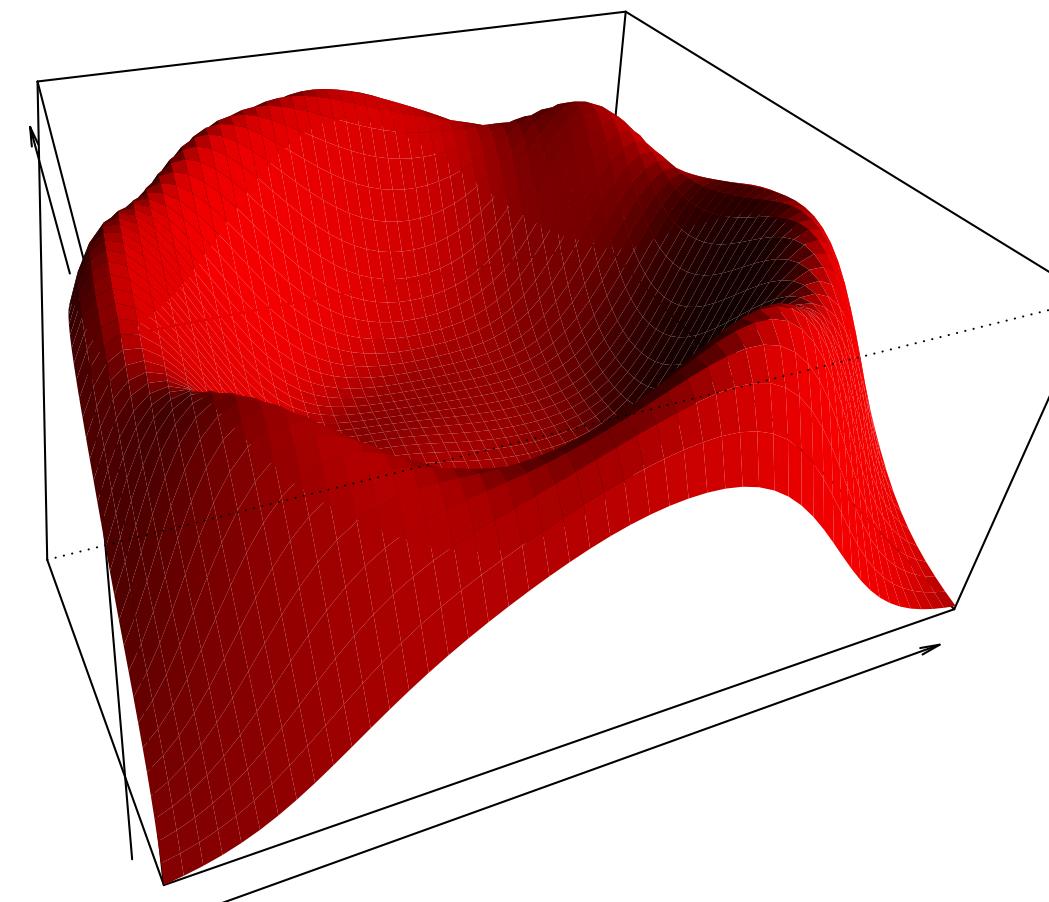
Joint work with JG Young and A. Patania  
Reference : arxiv.org/1705.10298  
Software : [github.com/jg-you/scm](https://github.com/jg-you/scm)



# Confidence sets

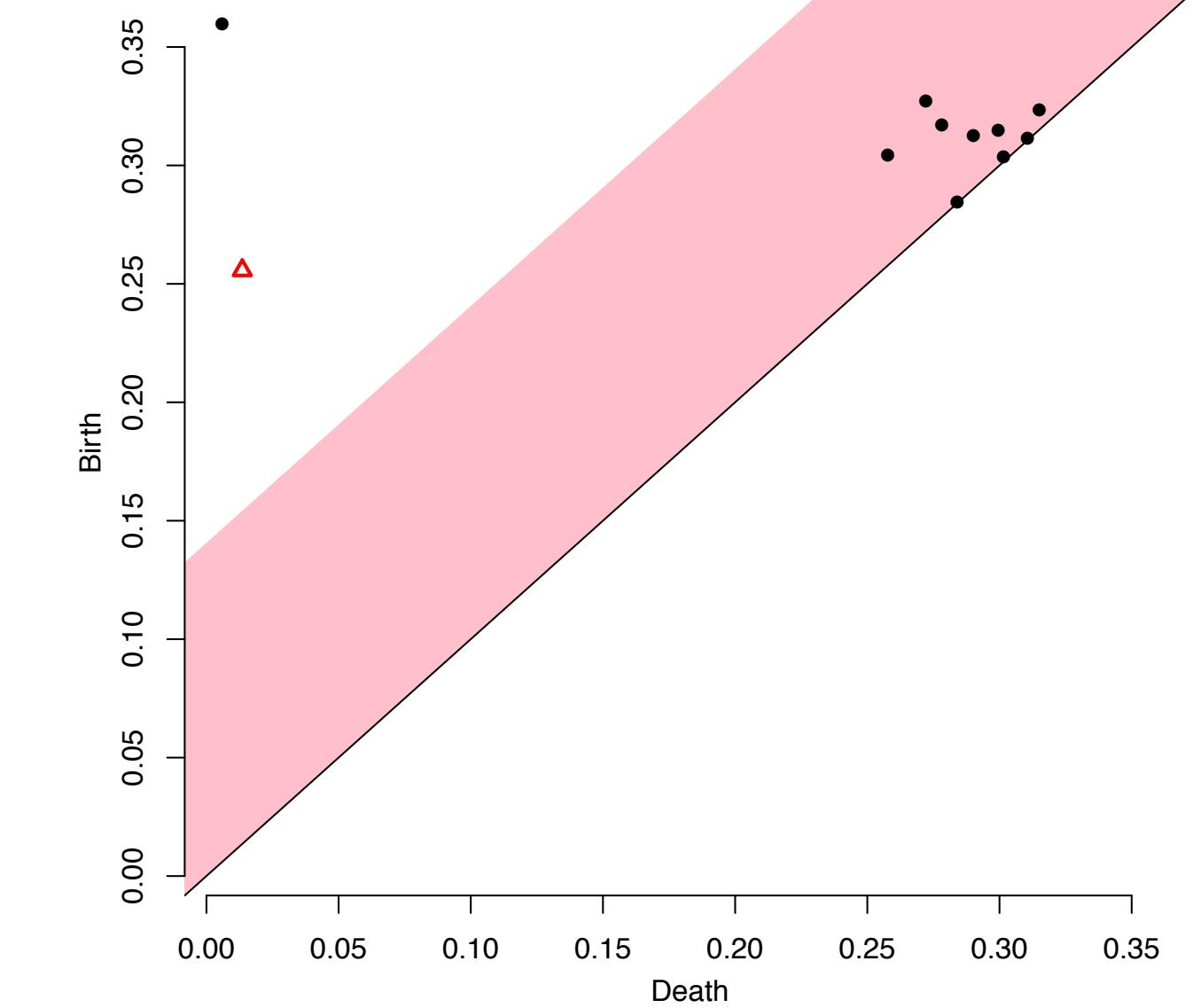
Fasy et al. (2014b); Chazal et al. (2014a)  
suggest the following method. Let

$$F(t) = P(\sqrt{n} \delta_\infty(\hat{D}, D) \leq t)$$



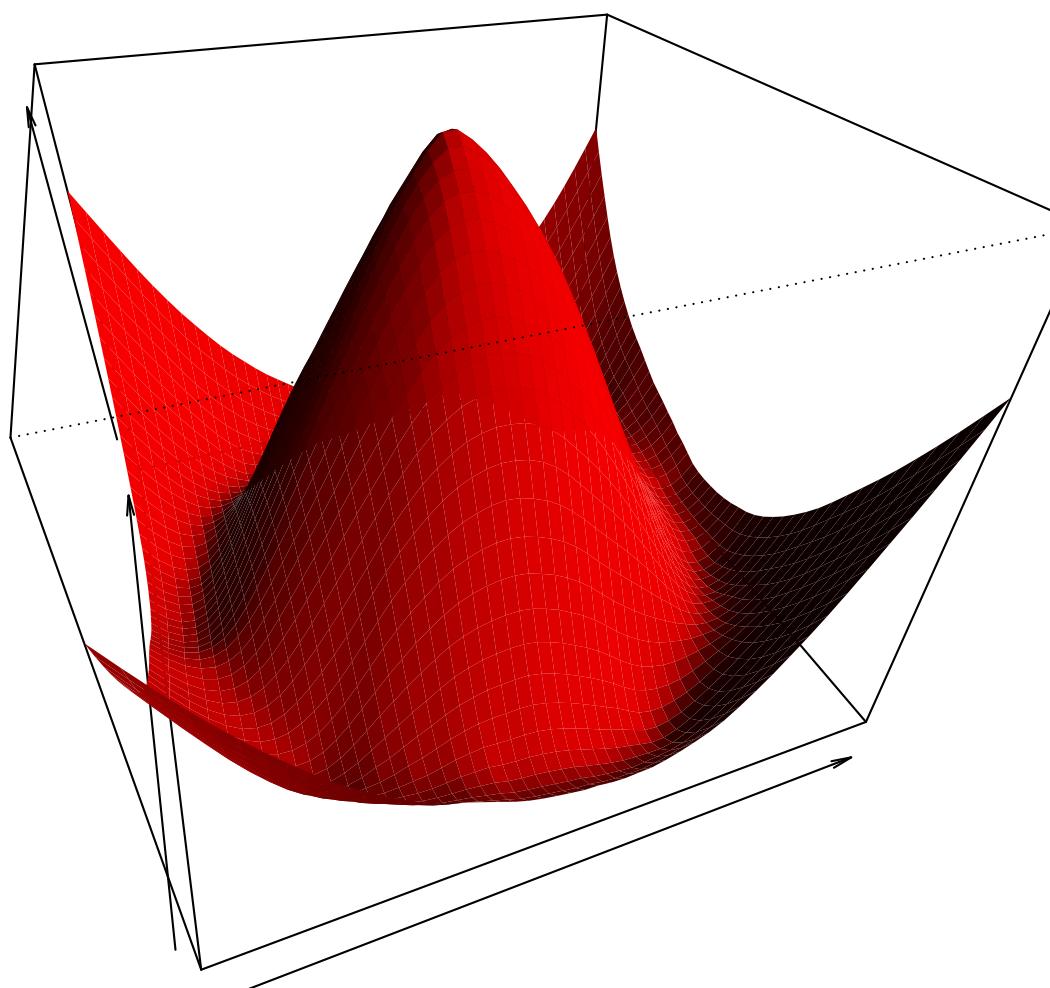
Estimated by bootstrap

$$\hat{F}(t) = \frac{1}{B} \sum_{j=1}^B I(\sqrt{n} d_\infty(\hat{D}_j^*, \hat{D}) \leq t)$$



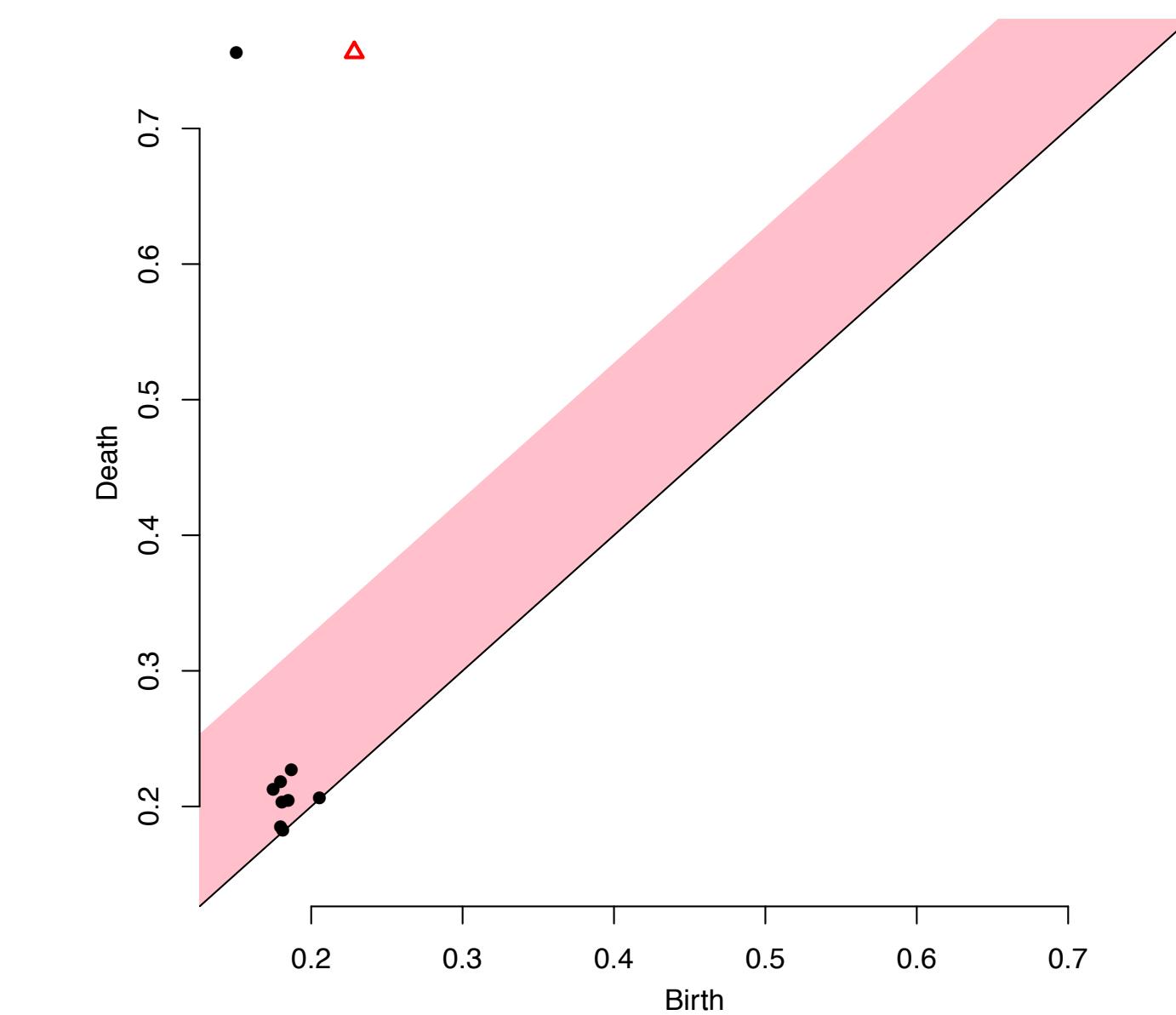
Significance threshold

$$t_\alpha = F^{-1}(1 - \alpha)$$



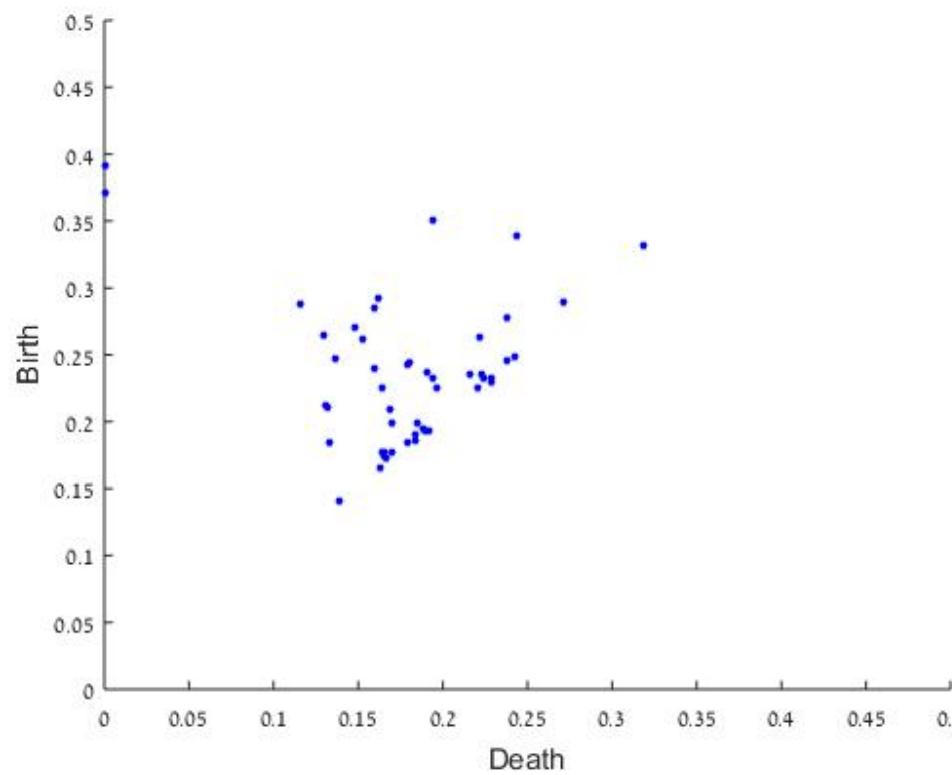
**Note:** Works for metrical cases.

**Tip:** try this in the TDA r-library

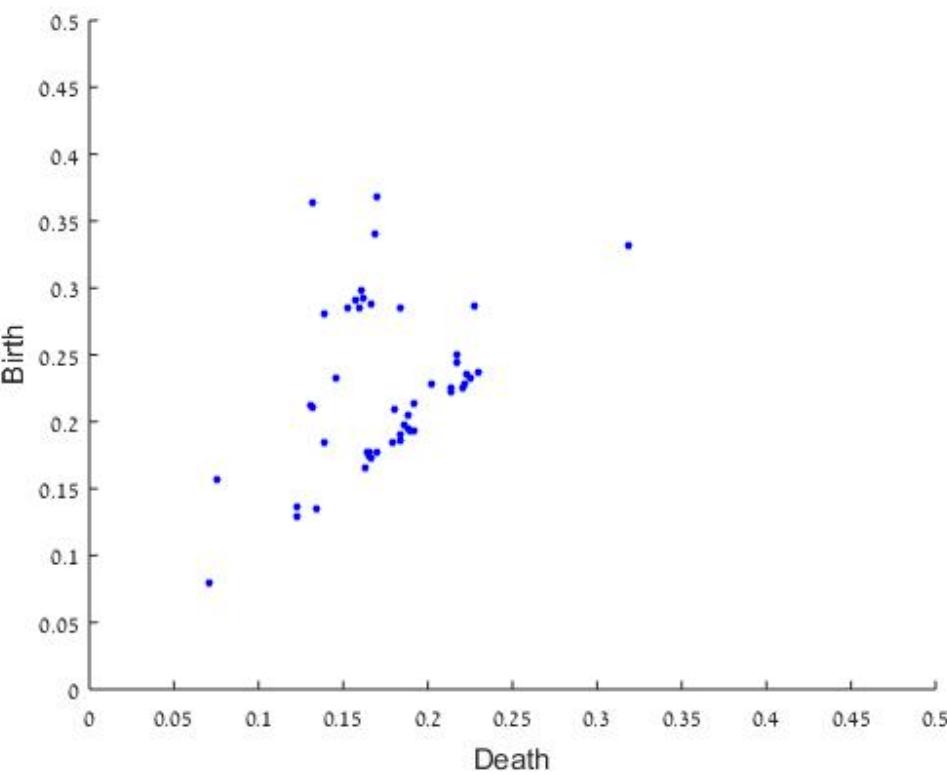


# PD “multiplication”

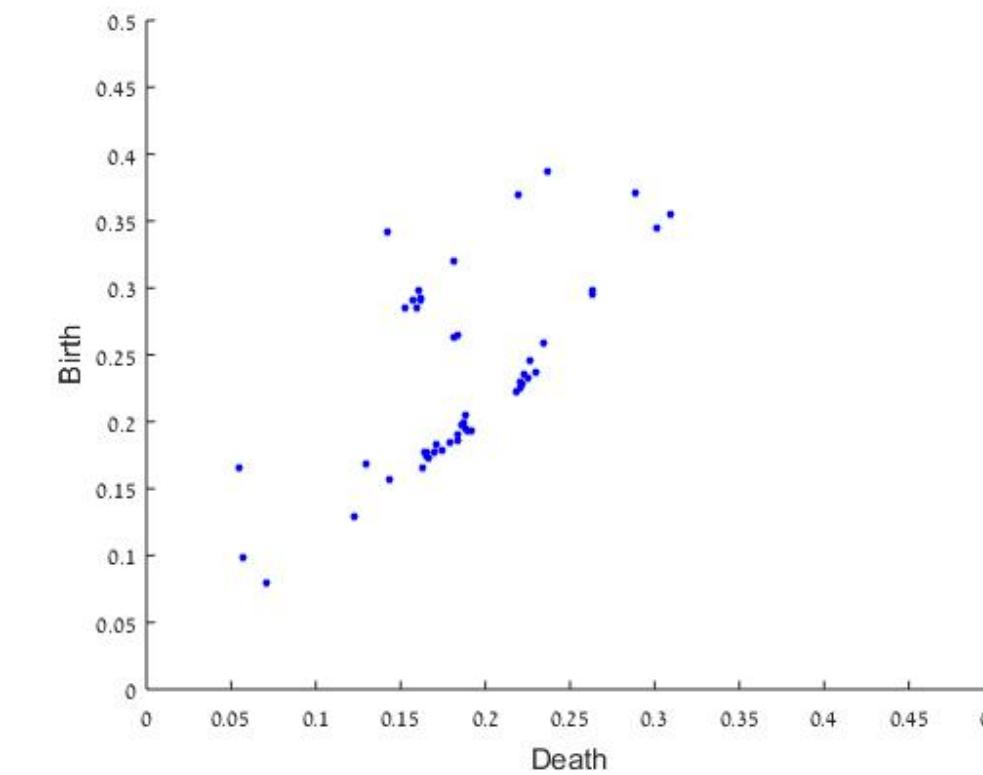
- Compute your PD
- Fit a Gibbs Model with Hamiltonian
- Use it to produce synthetic PDs



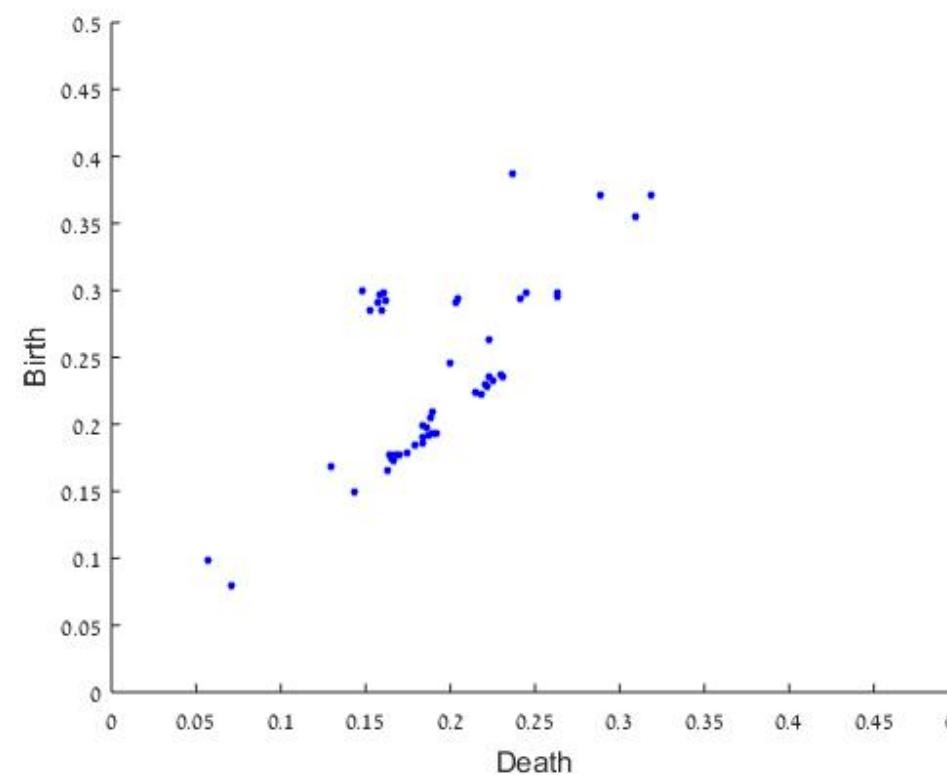
(a)



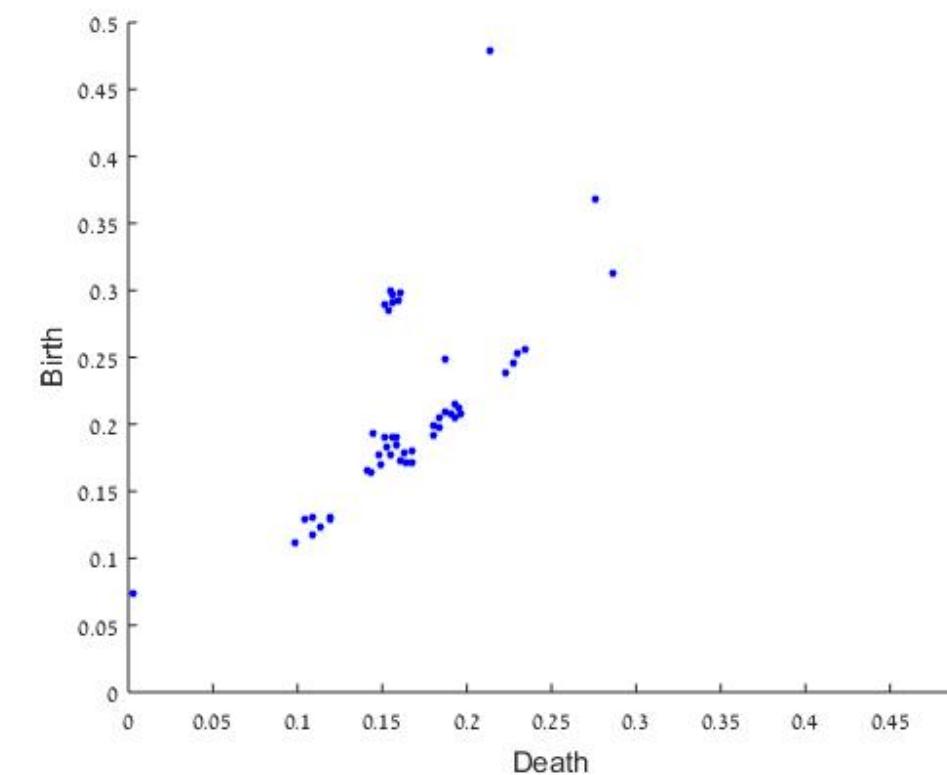
(b)



(c)



(d)



(e)

$$\varphi_{\Theta}(\tilde{x}_N) = \frac{1}{Z_{\Theta}} \exp(-H_{\Theta}(\tilde{x}_N)),$$

$$H_{\delta, \Theta}^2(\tilde{x}_N) = \theta_H \sigma_H^2 + \theta_V \sigma_V^2 + \sum_{k=1}^3 \delta^{-2} \theta_k \mathcal{L}_{\delta, k}(\tilde{x}_N),$$



# Notebook

## 07-08-09

**Thanks!**