The Waisman Laboratory for Brain Imaging and Behavior

University of Wisconsin
SCHOOL OF MEDICINE AND PUBLIC HEALTH

Wasserstein distance on graphs

Moo K. Chung
University of Wisconsin-Madison
www.stat.wisc.edu/~mchung

POSTECH MINDS WORKSHOP July 8, 2021

Abstract

The first part of the tutorial (MATLAB based) introduces the graph filtration, the baseline filtration on graph data structure. The second part of the tutorial explains how to compute the Wasserstein distance without doing numerical optimization that is needed to find the optimal bijection. The tutorial is based on arXiv:2012.00675.

Matlab codes:
http://www.stat.wisc.edu/~mchung/dynamicTDA
Algorithms are simple (10 lines), no need for MATLAB

Acknowledgement

Tananun Songdechakraiwut, Zhan Luo, Ian Carroll,
Gregory Kirk, Andrew Alexander, Hill Goldsmith, Richard Davidson
Univ. of Wisconsin-Madison

Li Shen Univ. of Pennsylvania
Jingwen Yan Indiana University–Purdue University Indianapolis
Hernando Ombao, Chee-Ming Ting KAUST, Saudi Arabia
Hyekyung Lee Seoul National University, Korea
Anqi Qiu, Shih-Gu Huang National University of Singapore
Victor Solo University of New South Wales

Grants:

NIH R01 Brain Initiative EB022856, R01 EB028753, NSF DMS-2010778

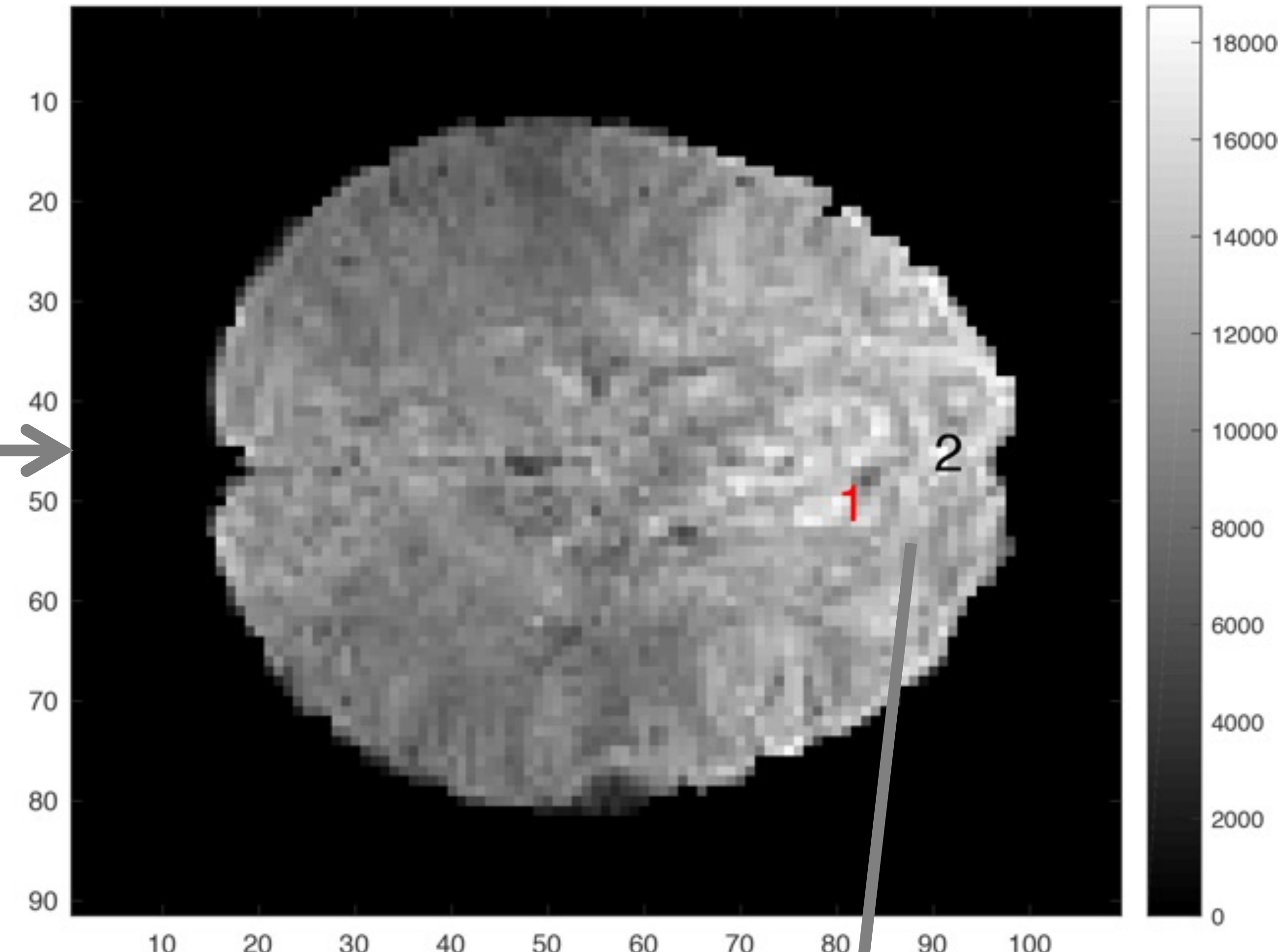
MATLAB demonstration

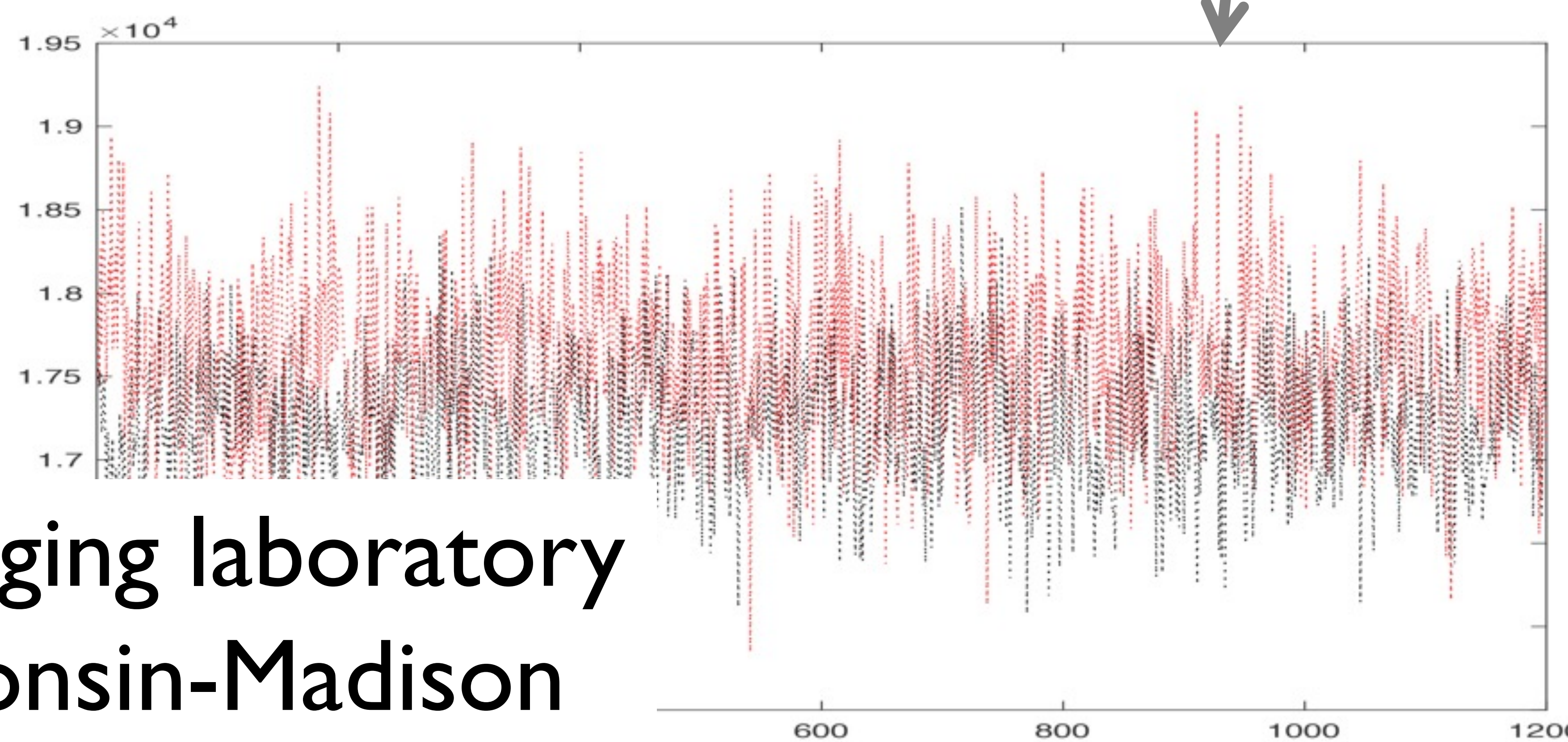
http://www.stat.wisc.edu/~mchung/dynamicTDA/

Other codes and brain imaging data:
http://www.stat.wisc.edu/~mchung/softwares.html
Many published papers have weblink for codes.
We have a library of +5000 custom functions.

Matlab can call Python libraries
directly inside Matlab.

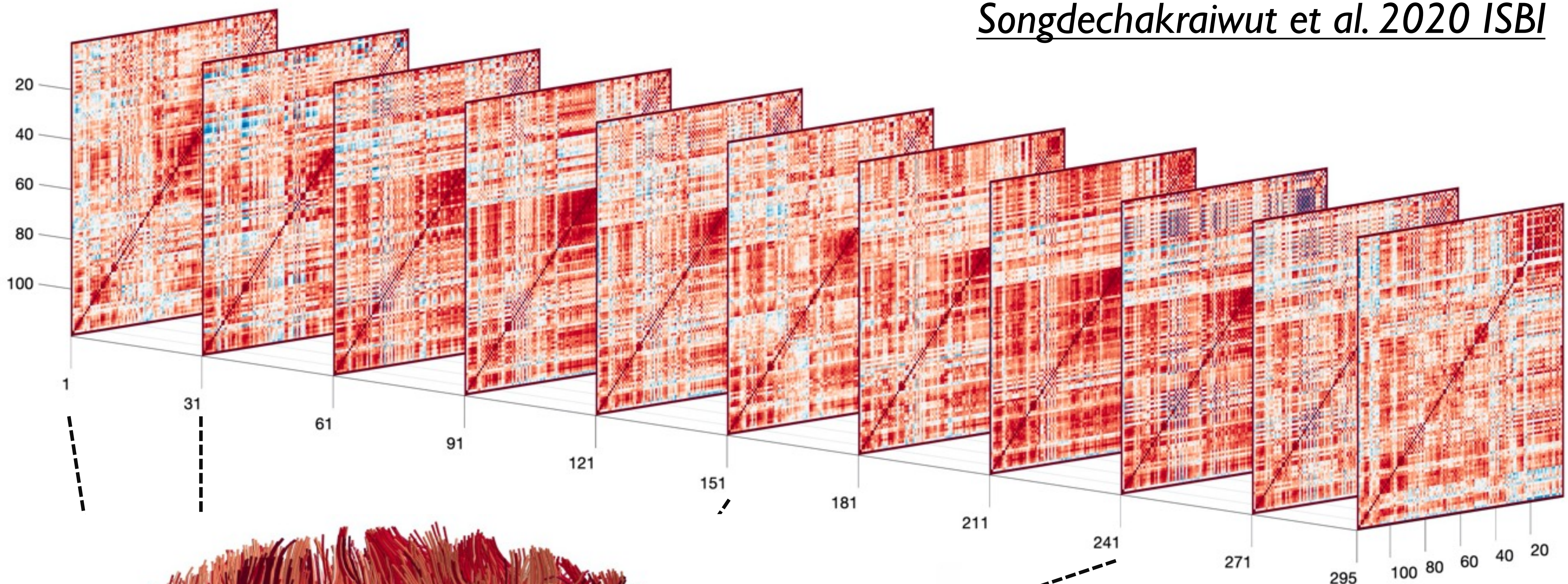
Resting-state functional magnetic resonance imaging (rs-fMRI)

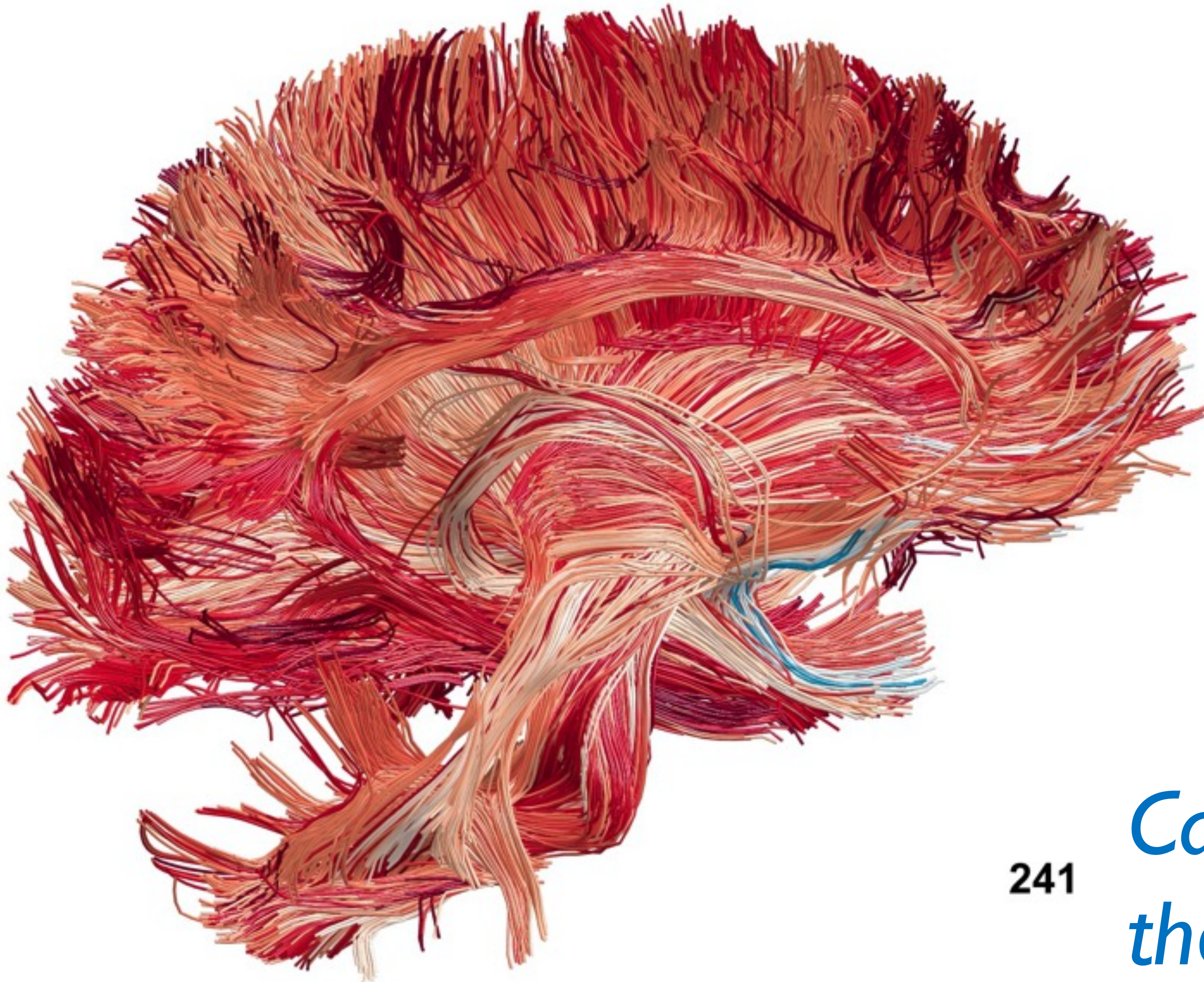




Waisman brain imaging laboratory
University of Wisconsin-Madison

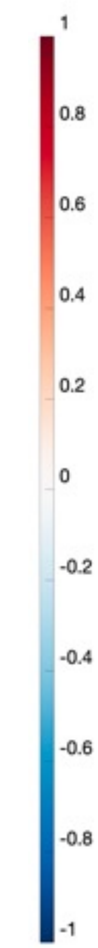
Songdechakraiwut et al. 2020 ISBI





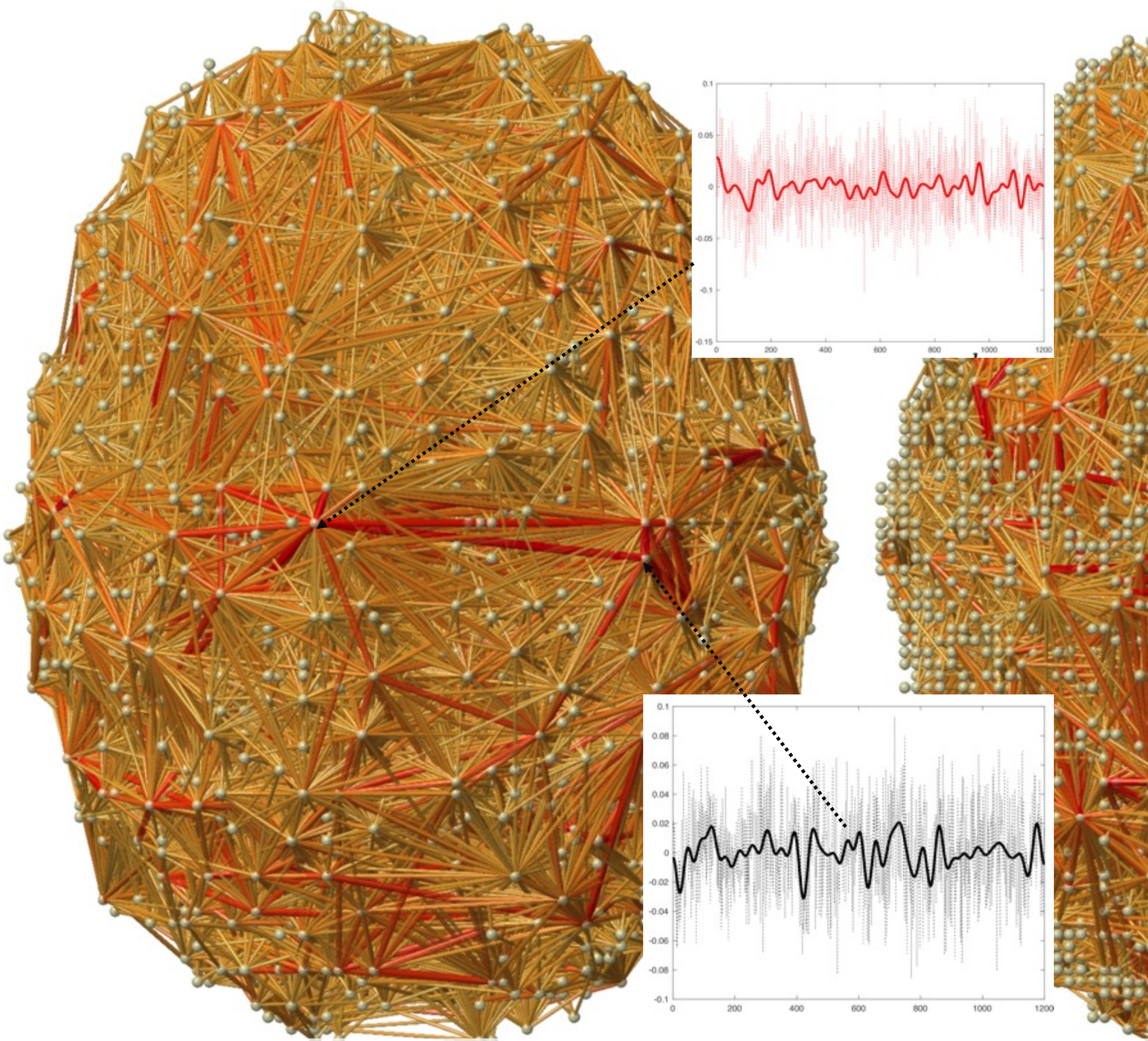
241

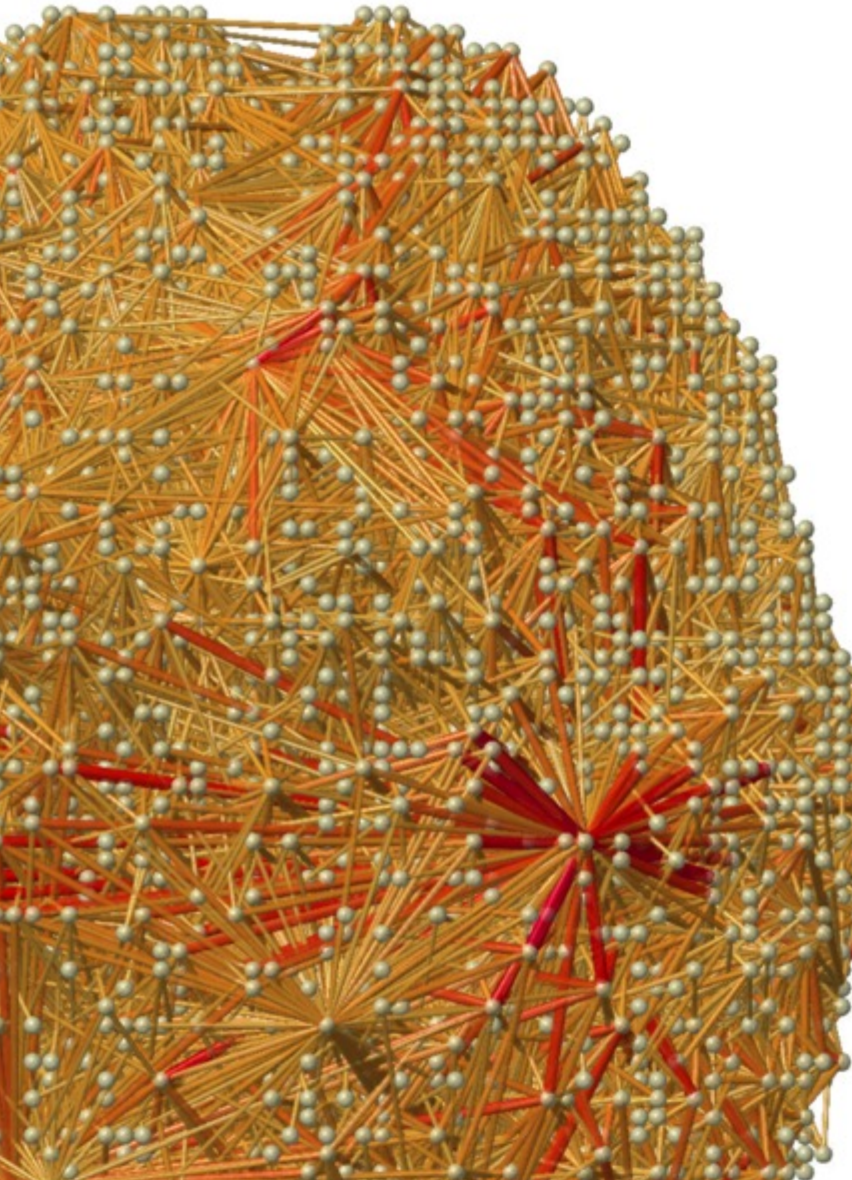
What is the common topology of brain network that does not change over time.



Can be easily addressed using the 10-line code given in tutorial

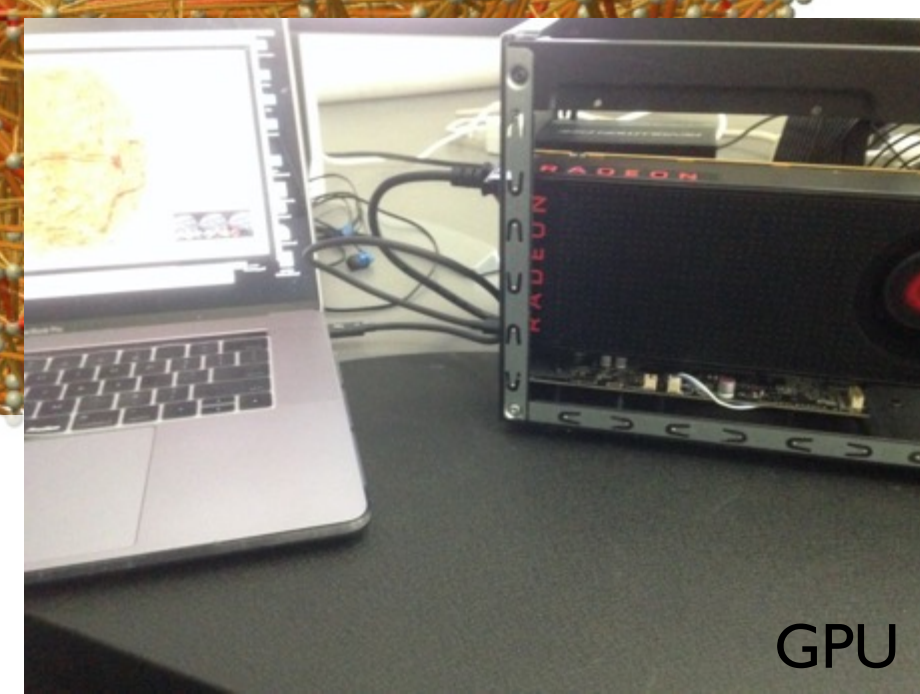
Correlation brain network at voxel level



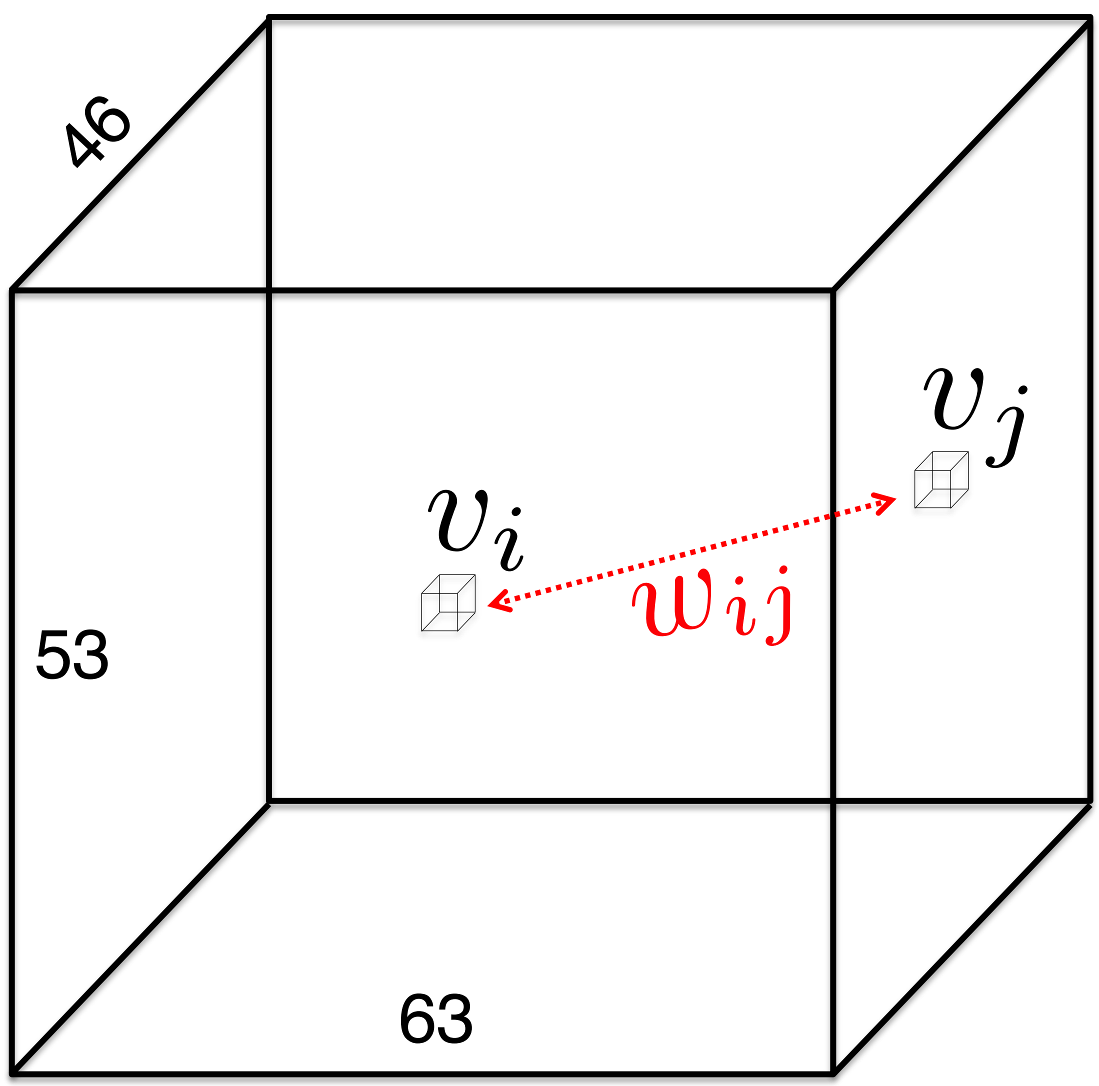


Correlation network of 300000 time series

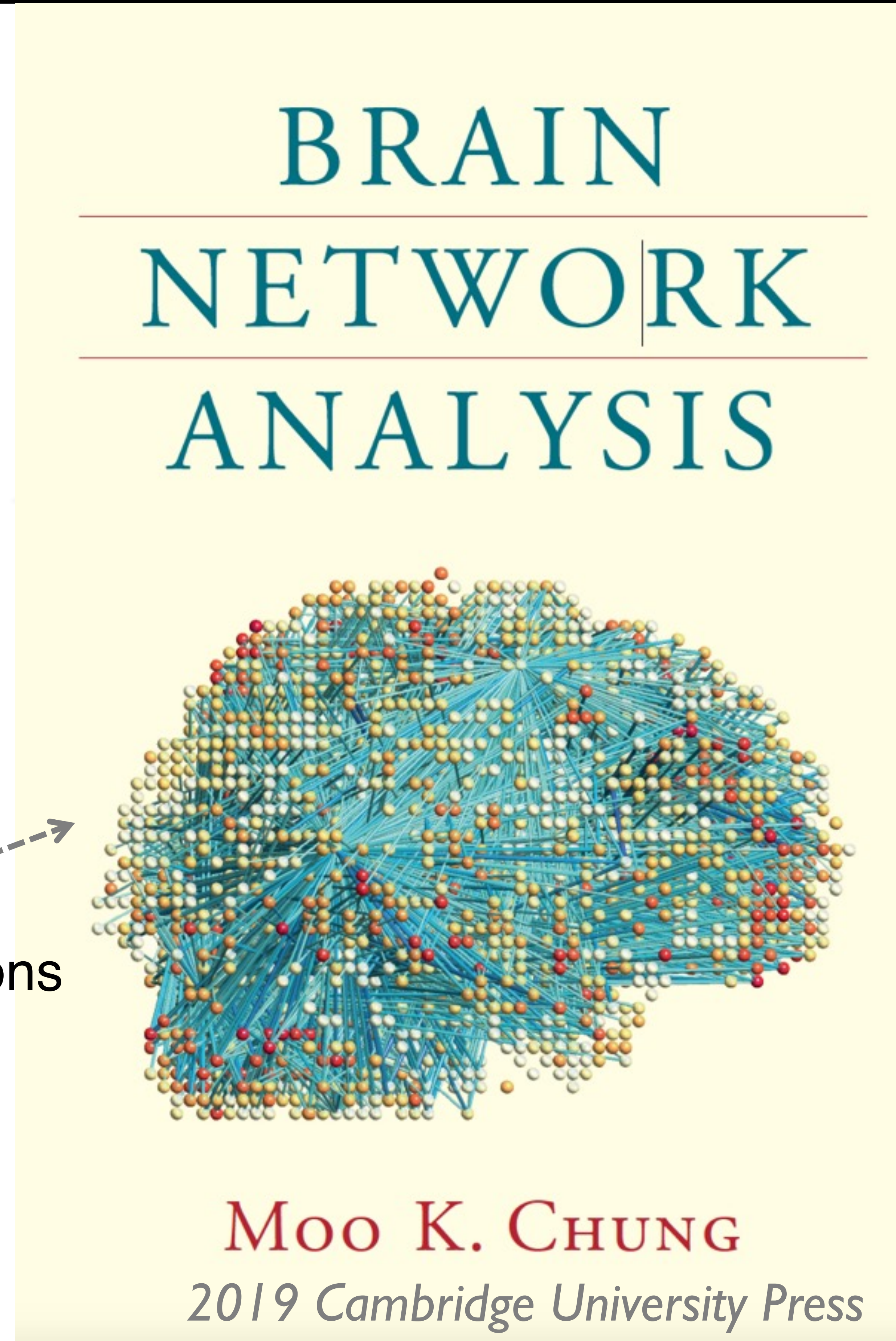
Complete graph with about $300000^2/2$ cycles.



How big is brain network data?



v_i



p=25972 voxels (3mm) in the brain
→ 25972 x 25972 = 0.67 billion connections
5.2GB memory

300000 voxels (1mm)
→ 90 billion connections
→ 700 GB memory

Persistent homology computation is slow:
$O(n^3)$ rank estimation via Gaussian elimination
3D problem

Scalable persistent homology

Graph filtrations

$O(n \log n)$ 1.5D problem

Concept first introduced in

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

Baseline filtration for (brain) network data

Computing the Shape of Brain Networks Using Graph Filtration and Gromov-Hausdorff Metric

Hyekyoung Lee[1,2,3], Moo K. Chung[2,6,7], Hyejin Kang[1,3],
Boong-Nyun Kim[5], and Dong Soo Lee[1,3,4]

[1] Department of Nuclear Medicine,
[2] Department of Brain and Cognitive Sciences,
[3] Institute of Radiation Medicine, Medical Research Center,
[4] WCU Department of Molecular Medicine and Biopharmaceutical Sciences,
[5] Department of Neuropsychiatry, Seoul National University,
College of Medicine, Seoul, Korea
[6] Department of Biostatistics and Medical Informatics,
[7] Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin,
Madison, WI 53706, USA
mkchung@wisc.edu

Lee et al. 2011 MICCAI 302-309

Graph Filtrations

Weighted complete graph

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

Node set

Edge weight

Binary graph

$$\mathcal{X}_\epsilon = (V, w_\epsilon)$$

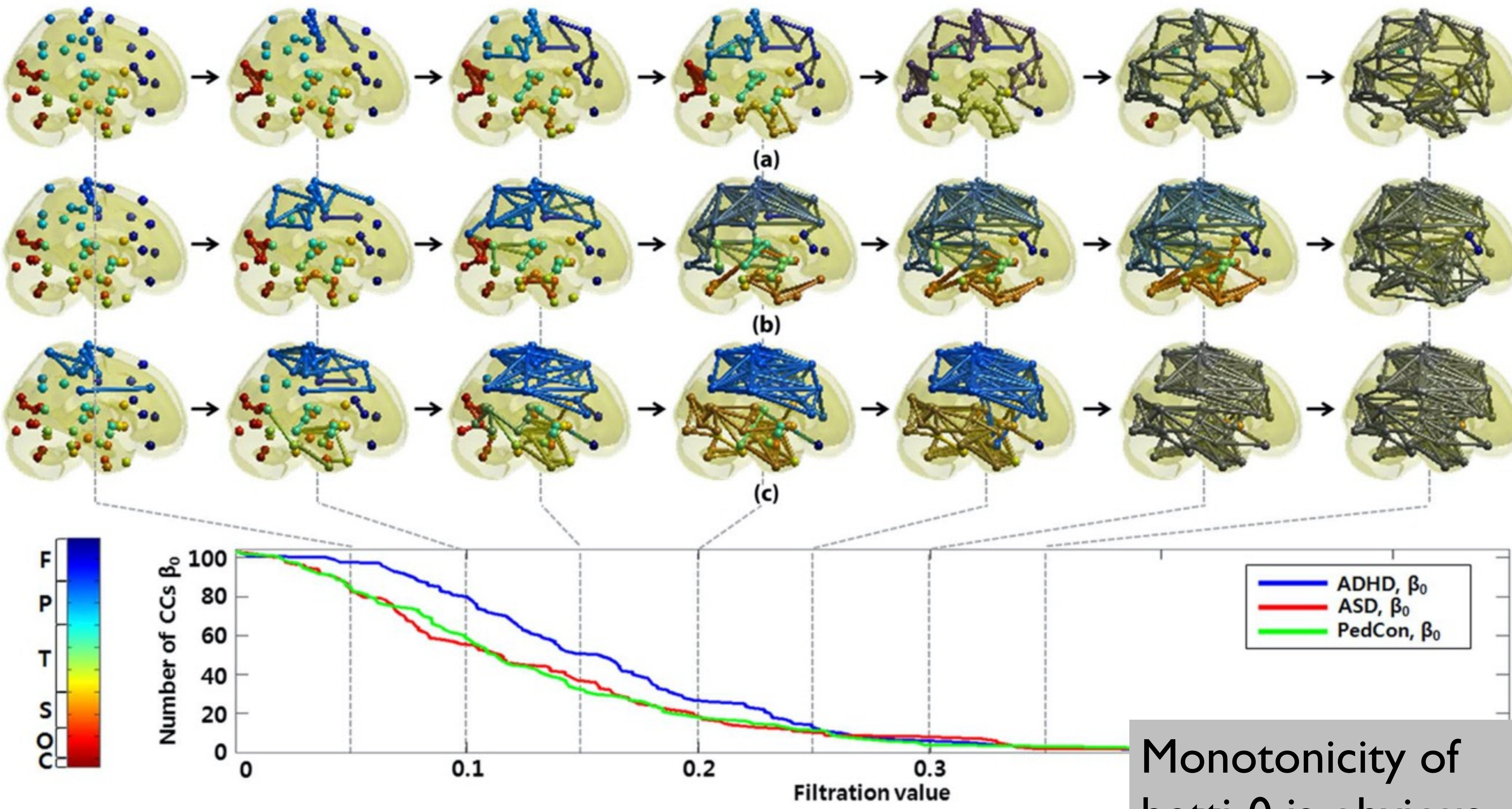
$$w_{\epsilon,ij} = \begin{cases} 1 & \text{if } w_{ij} > \epsilon; \\ 0 & \text{otherwise.} \end{cases}$$

Graph filtration

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

for increased edge weights

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

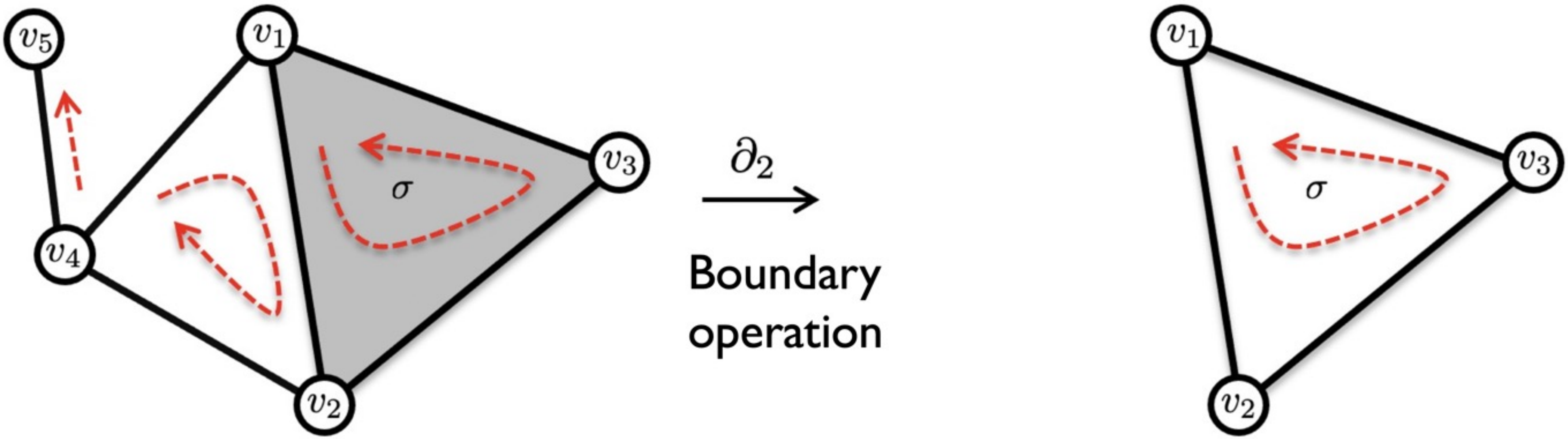


Monotonicity of betti-0 is obvious

Kruskal's algorithm
for minimum spanning tree
is used for graph filtration

Lee et al. 2012 IEEE Transactions on Medical Imaging 31:2267-2277

Simplical complex vs. Graph



High-order interaction

Pairwise interaction

Rips filtration vs. graph filtration

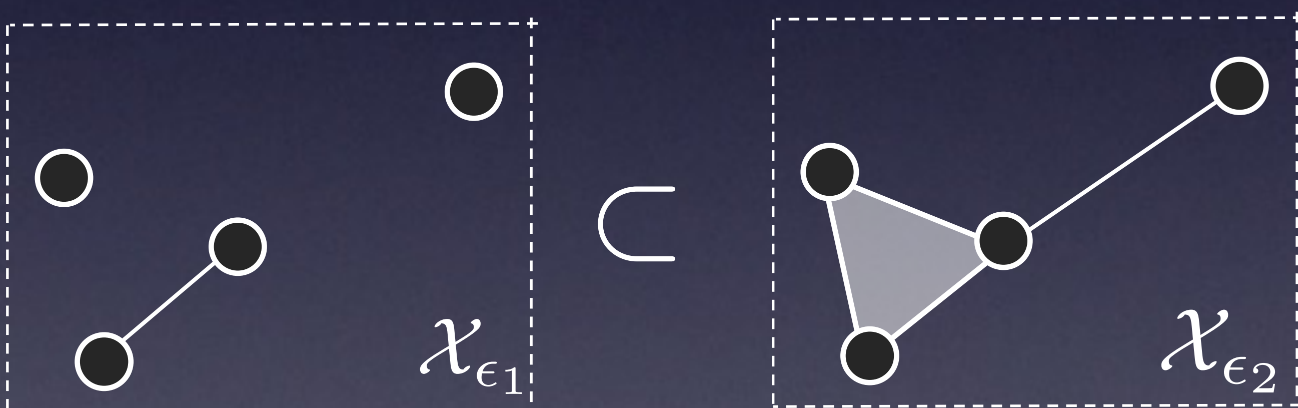
Metric space

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

Node set — Metric

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

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

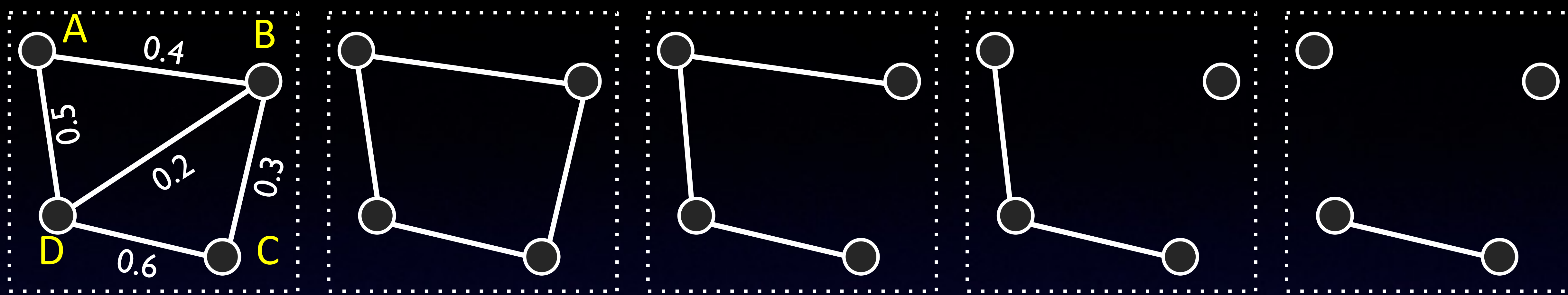
Graph filtration

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

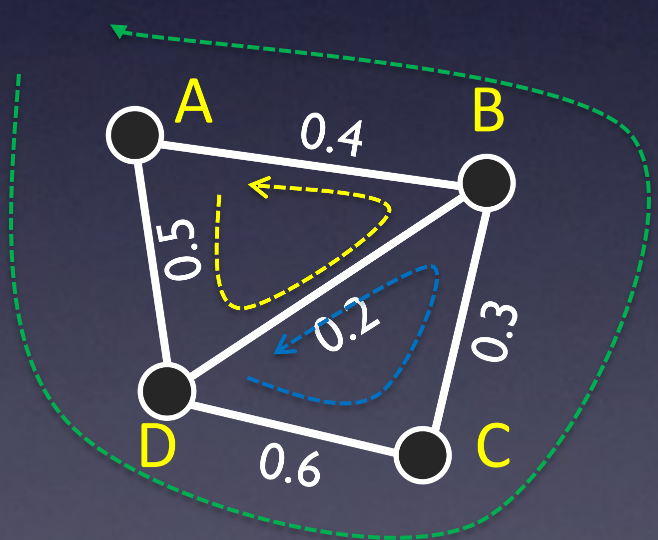
for increased edge weights

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

Graph filtration



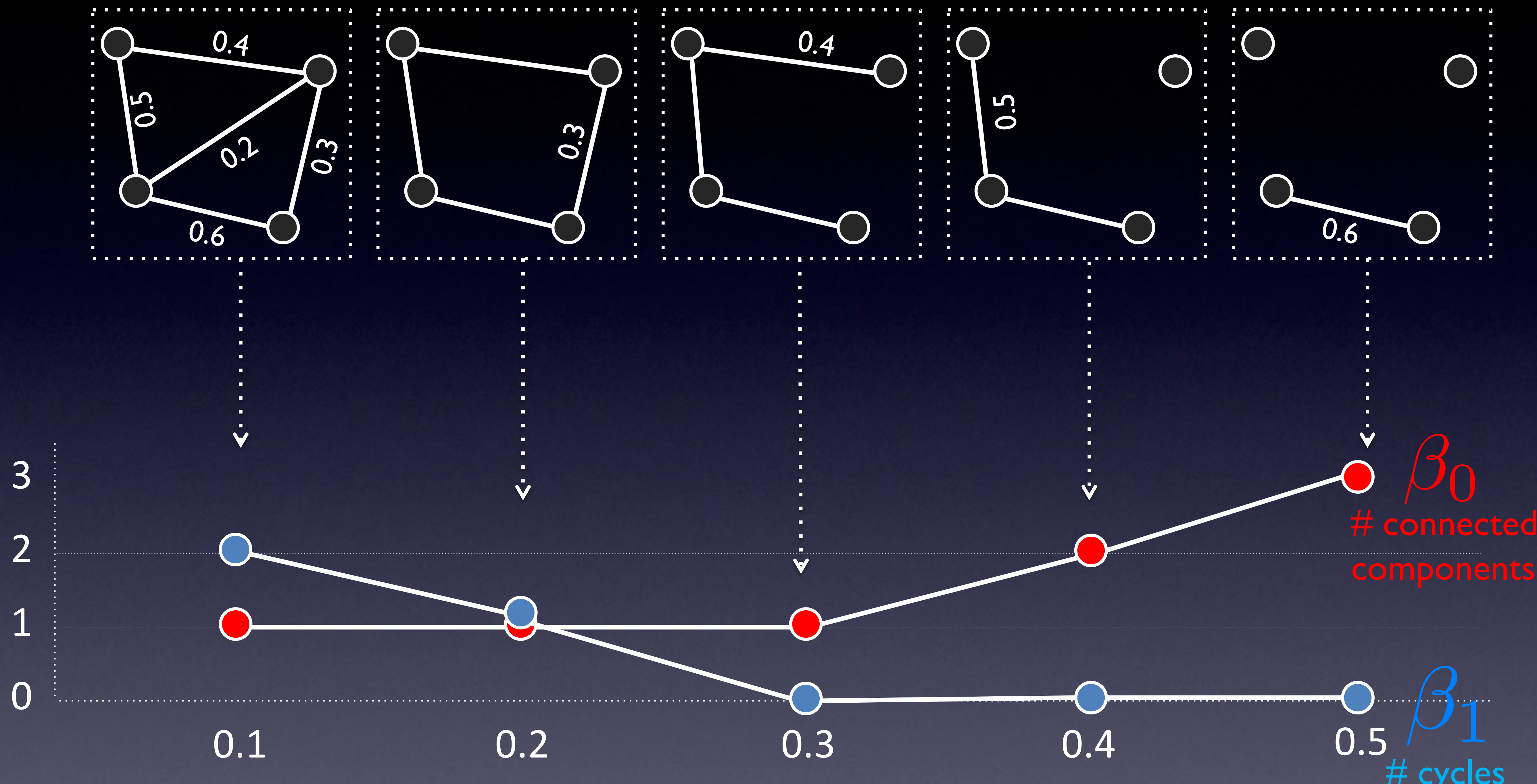
$G_0 \supset G_{0.2} \supset G_{0.3} \supset G_{0.4} \supset G_{0.5}$



ADCD = ADB + DCB → vector space

Can we have monotonicity of betti-1?

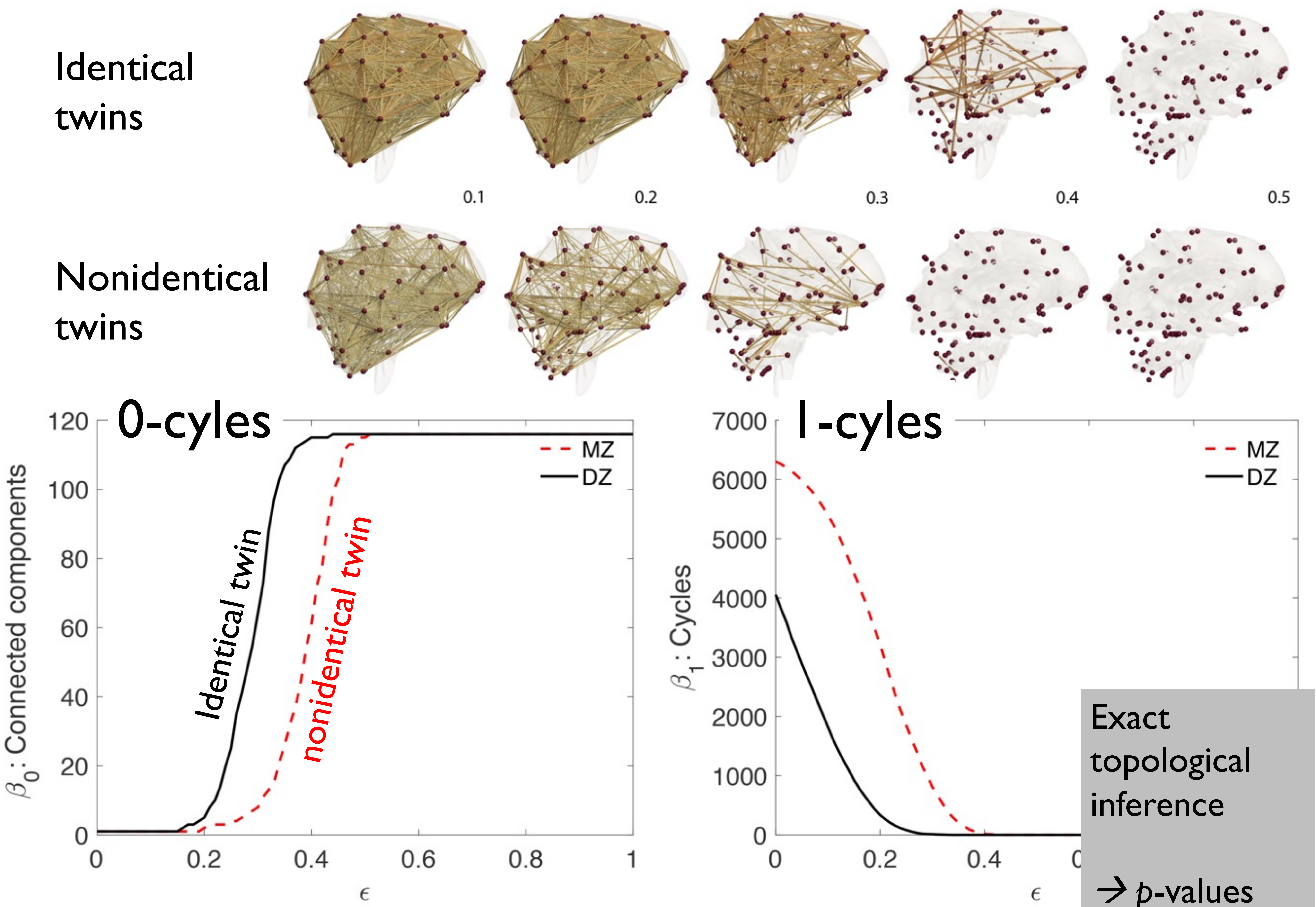
Theorem 1: *Monotonicity of Betti numbers over graph filtration*



Chung et al. 2019 ISBI Statistical inference on the number of cycles

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

Genetic difference in brain network

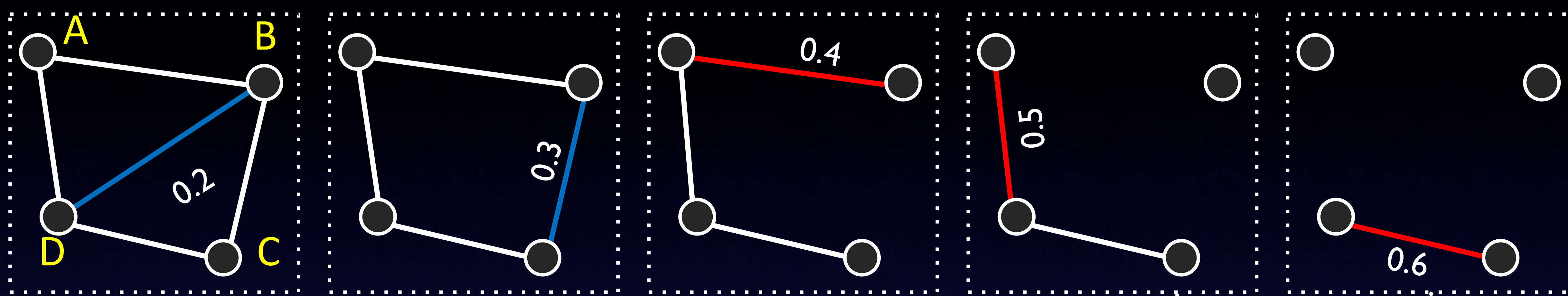


Exact topological inference

→ p-values

Theorem 2: Birth & death decomposition

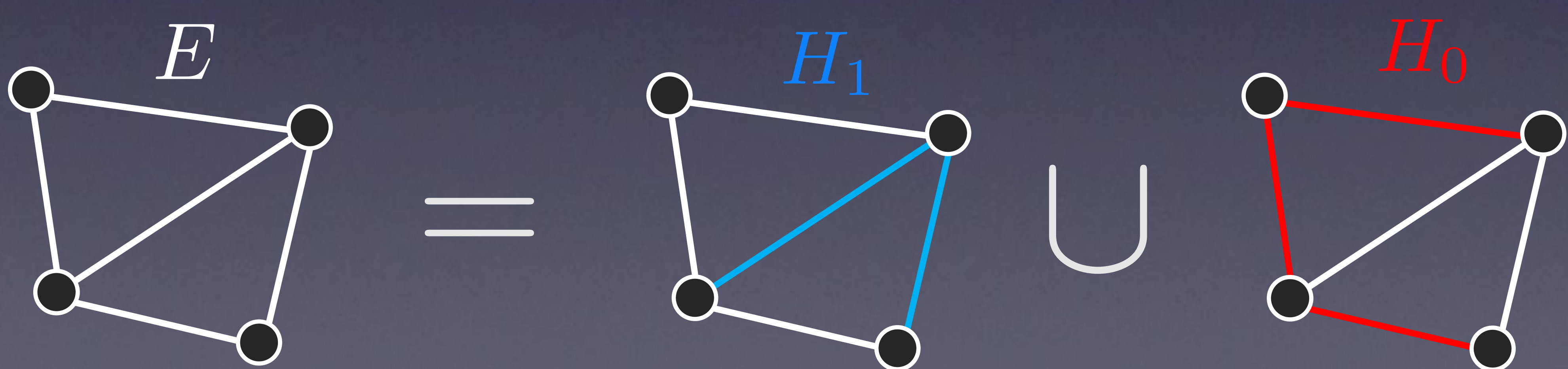
H_1 Edges destroy cycles H_0 Edges create components



$$\#(H_1) = 1 + \frac{|V|(|V|-3)}{2}$$

$$\#(H_0) = |V| - 1$$

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

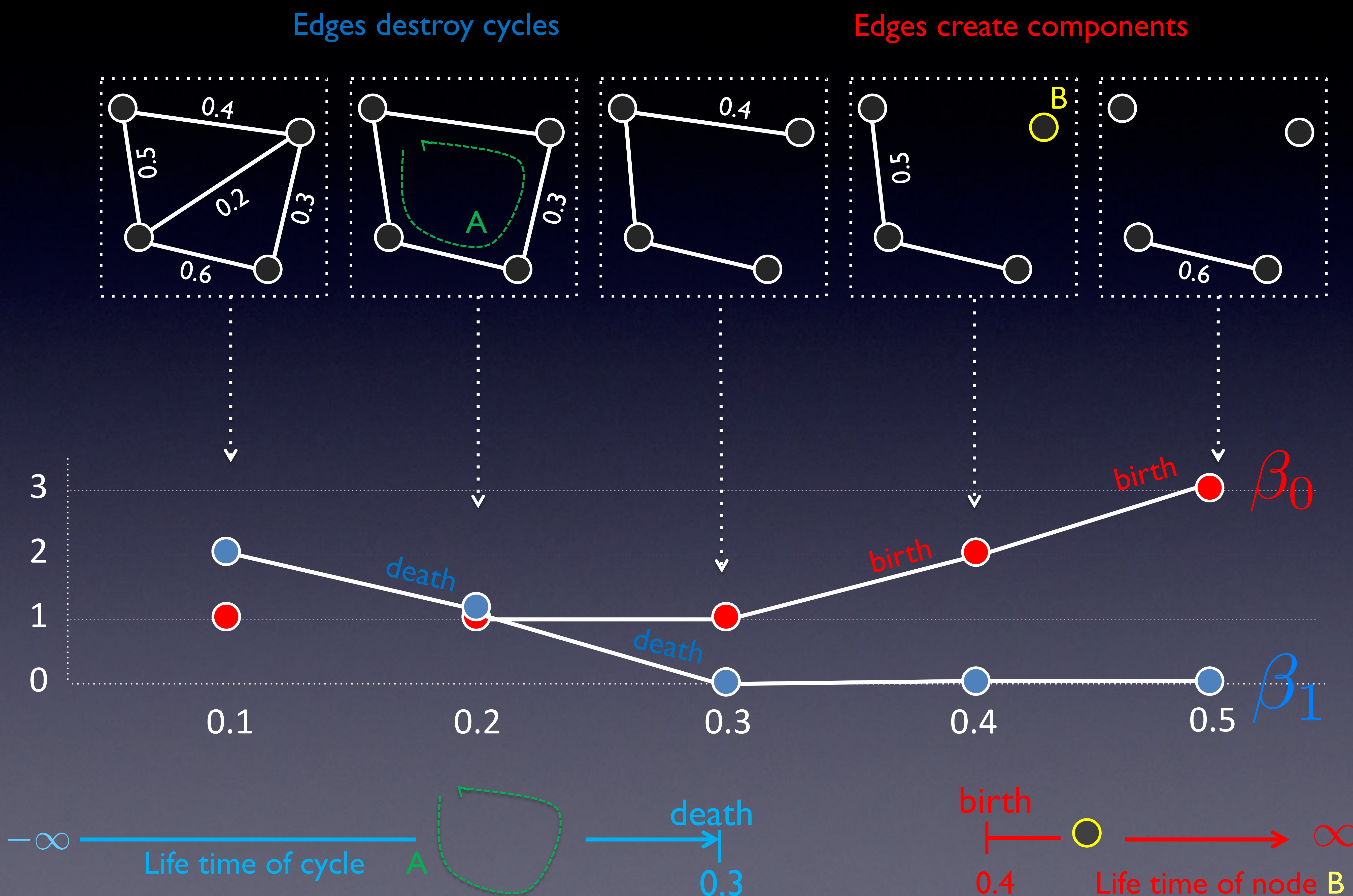


Maximum spanning tree

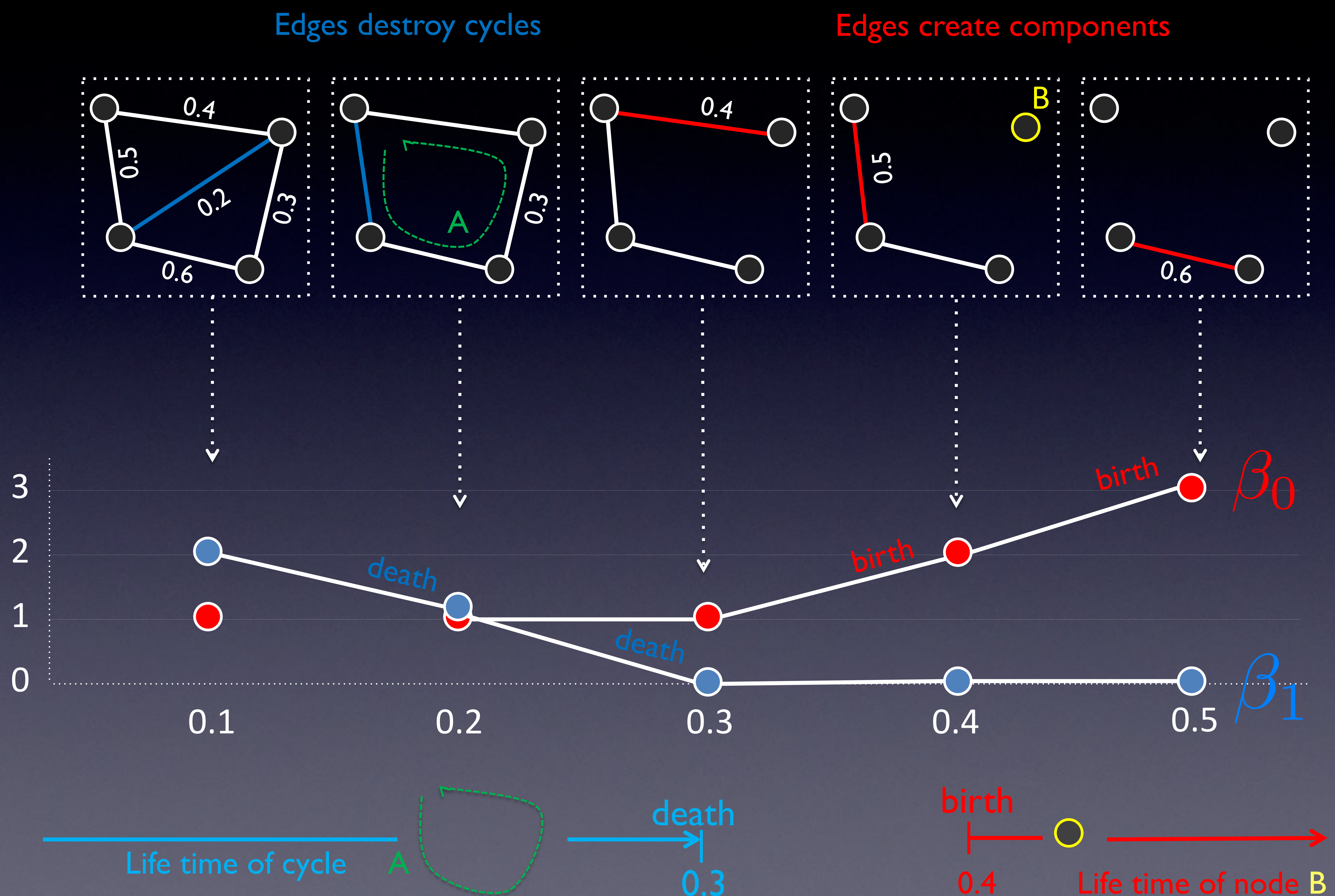
$O(|E| \log |V|)$

Songdechakraiwut & Chung 2020 arXiv: 2012.00675

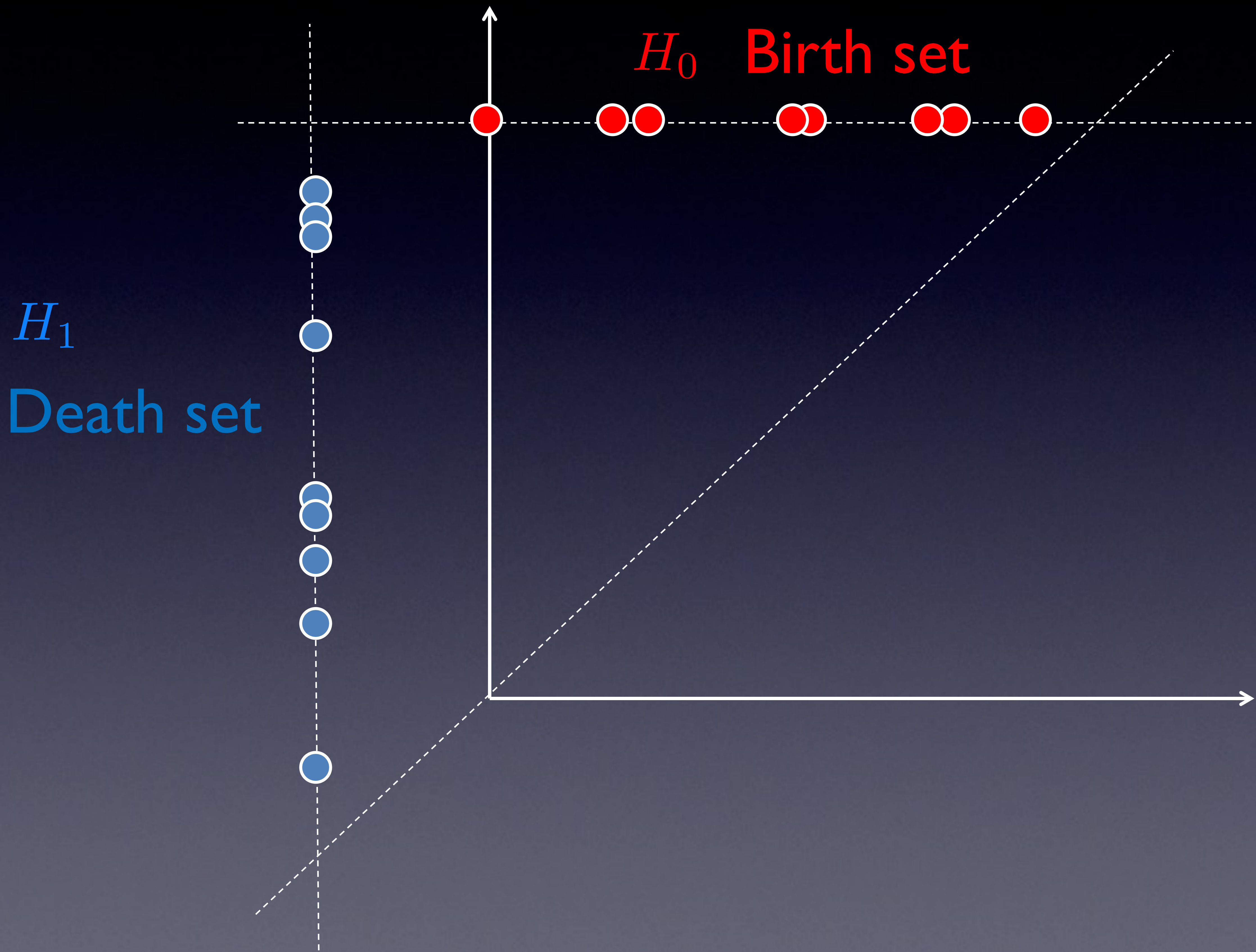
Persistence = Life time (death – birth) of a feature



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

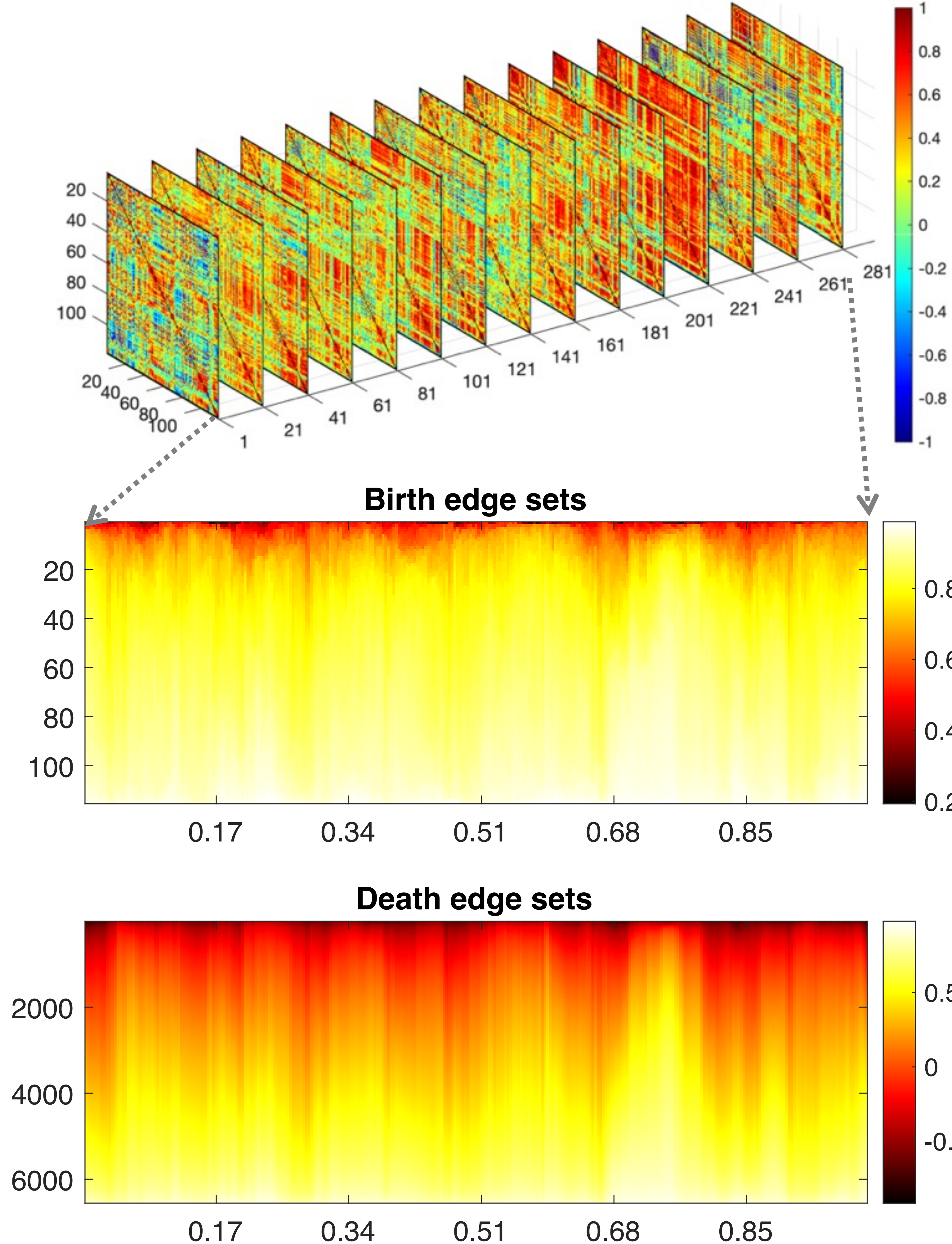


Persistence diagram for graph filtrations



Dynamically changing
correlation matrices
from functional brain network

Time series of birth
& death edge sets



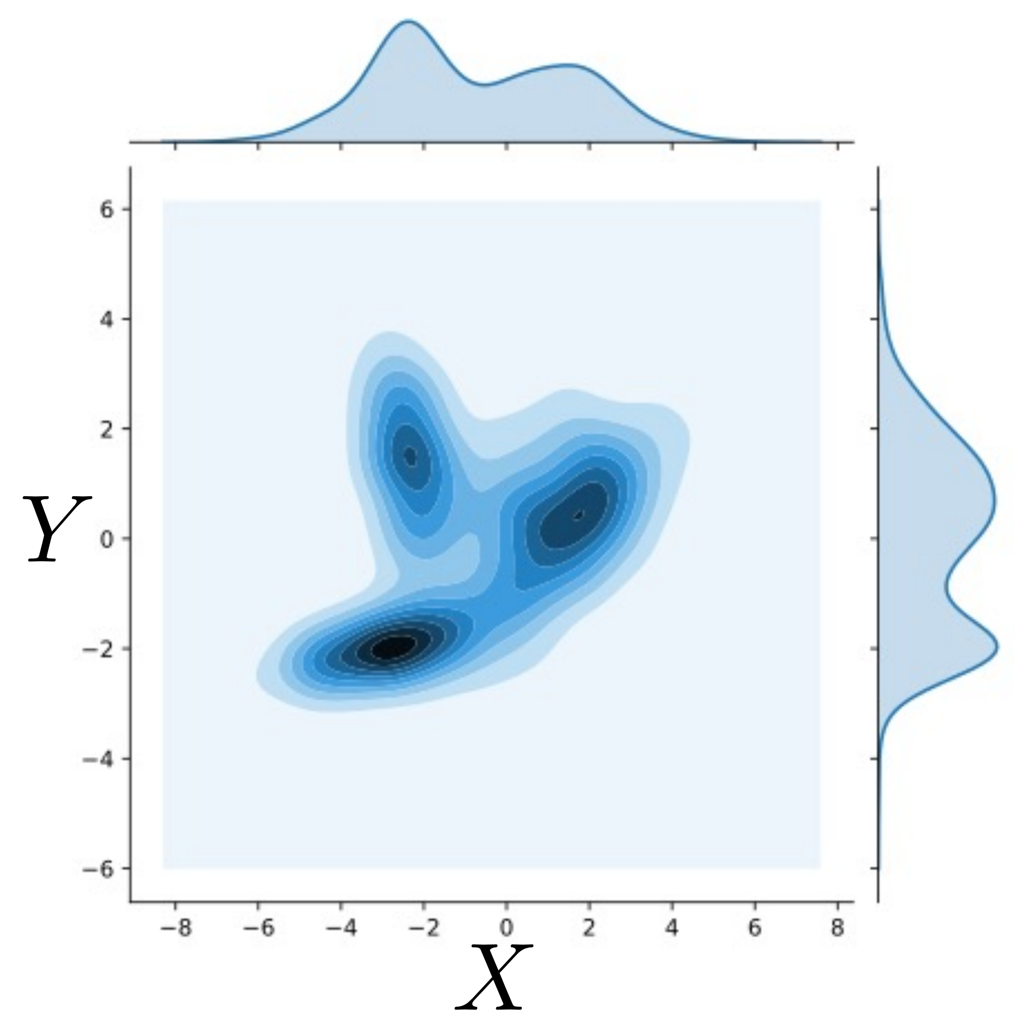
2-Wasserstein distance

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

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

Infinium taken over all possible joint distributions

Nonunique joint distribution



2-Wasserstein distance on persistent diagram

Persistent diagrams $P_1 = \{x_1, \cdots, x_q\}$ $P_2 = \{y_1, \cdots, y_q\}$

Empirical distributions

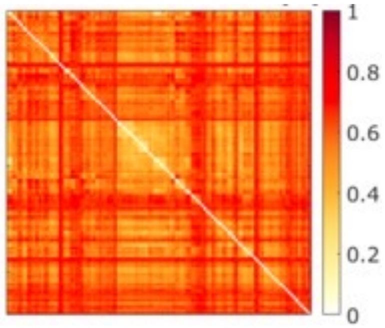
$$f_1(x) = \frac{1}{q}\sum_{i=1}^{q}\delta(x - x_i) \qquad f_2(y) = \frac{1}{q}\sum_{i=1}^{q}\delta(y - y_i)$$

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

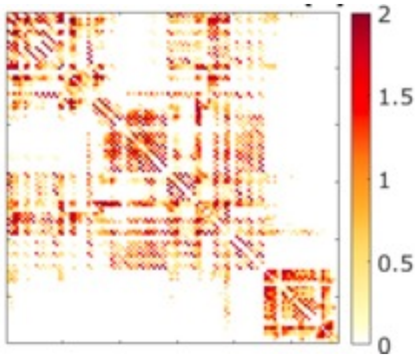
Assignment problem: Hungarian algorithm $\mathcal{O}(q^3)$

Wasserstein distance on the PD of graph filtration

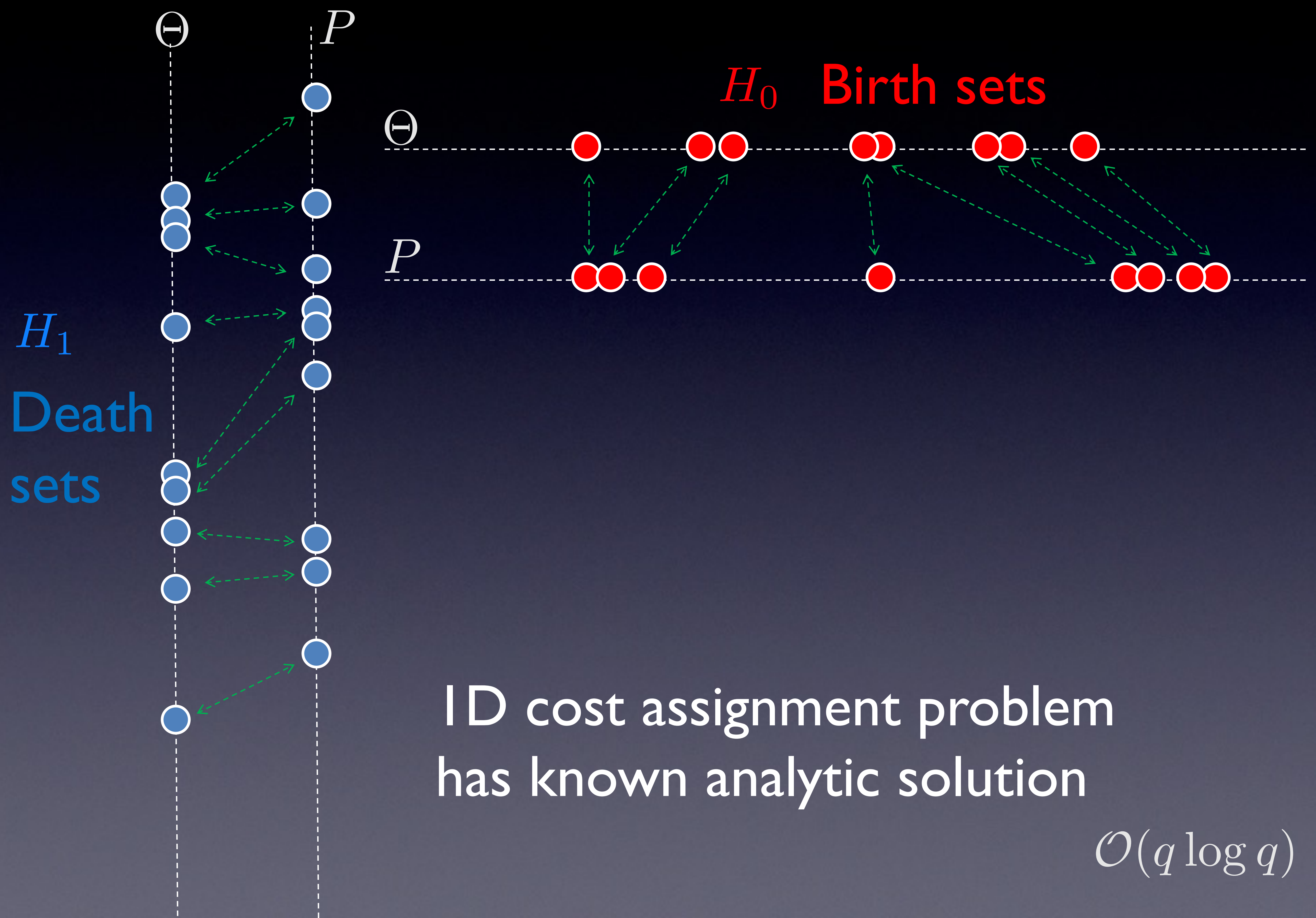
$$\Theta = (V^{\Theta}, w^{\Theta})$$



$$P = (V^{P}, w^{P})$$



Wasserstein distance for graph filtrations



1D cost assignment problem
has known analytic solution

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

Theorem 3 Wasserstein distance on graph filtrations

$$\mathcal{L}_{0D}(\Theta, P) = \min_{\tau} \sum_{b \in E_0} \left[b - \tau(b) \right]^2$$

$$= \sum_{b \in E_0} \left[b - \tau_0^*(b) \right]^2$$

τ_0^* *:The i-th smallest birth value to the i-th smallest birth value*

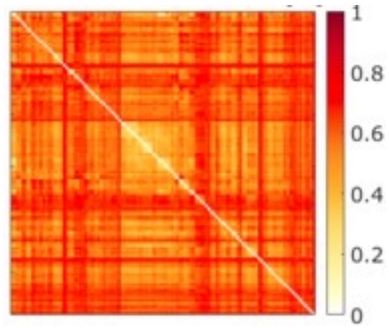
$$\mathcal{L}_{1D}(\Theta, P) = \min_{\tau} \sum_{d \in E_1} \left[d - \tau(d) \right]^2$$

$$= \sum_{d \in E_1} \left[d - \tau_1^*(d) \right]^2$$

τ_1^* *:The i-th smallest death value to the i-th smallest death value*

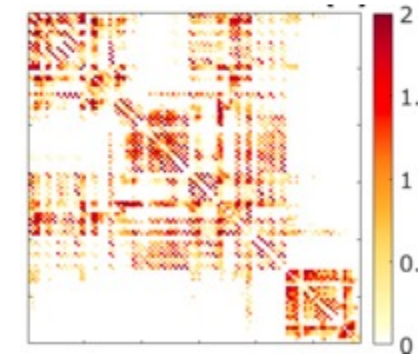
2-Wasserstein distance on graphs

$\Theta = (V^{\Theta}, w^{\Theta})$



bijection τ

$P = (V^{P}, w^{P})$

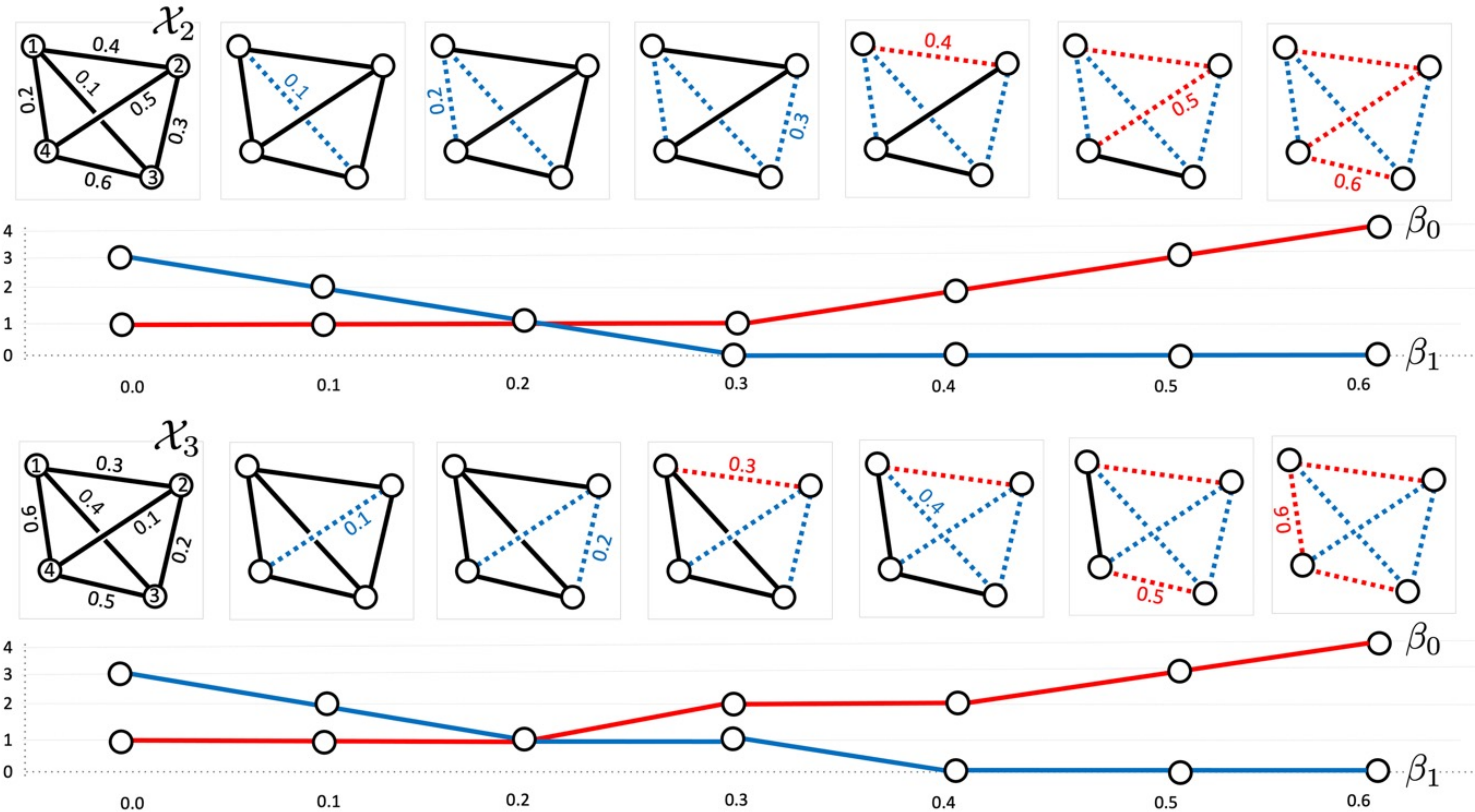


$$\mathcal{L}_{top}(\Theta, P) = \mathcal{L}_{0D}(\Theta, P) + \mathcal{L}_{1D}(\Theta, P)$$

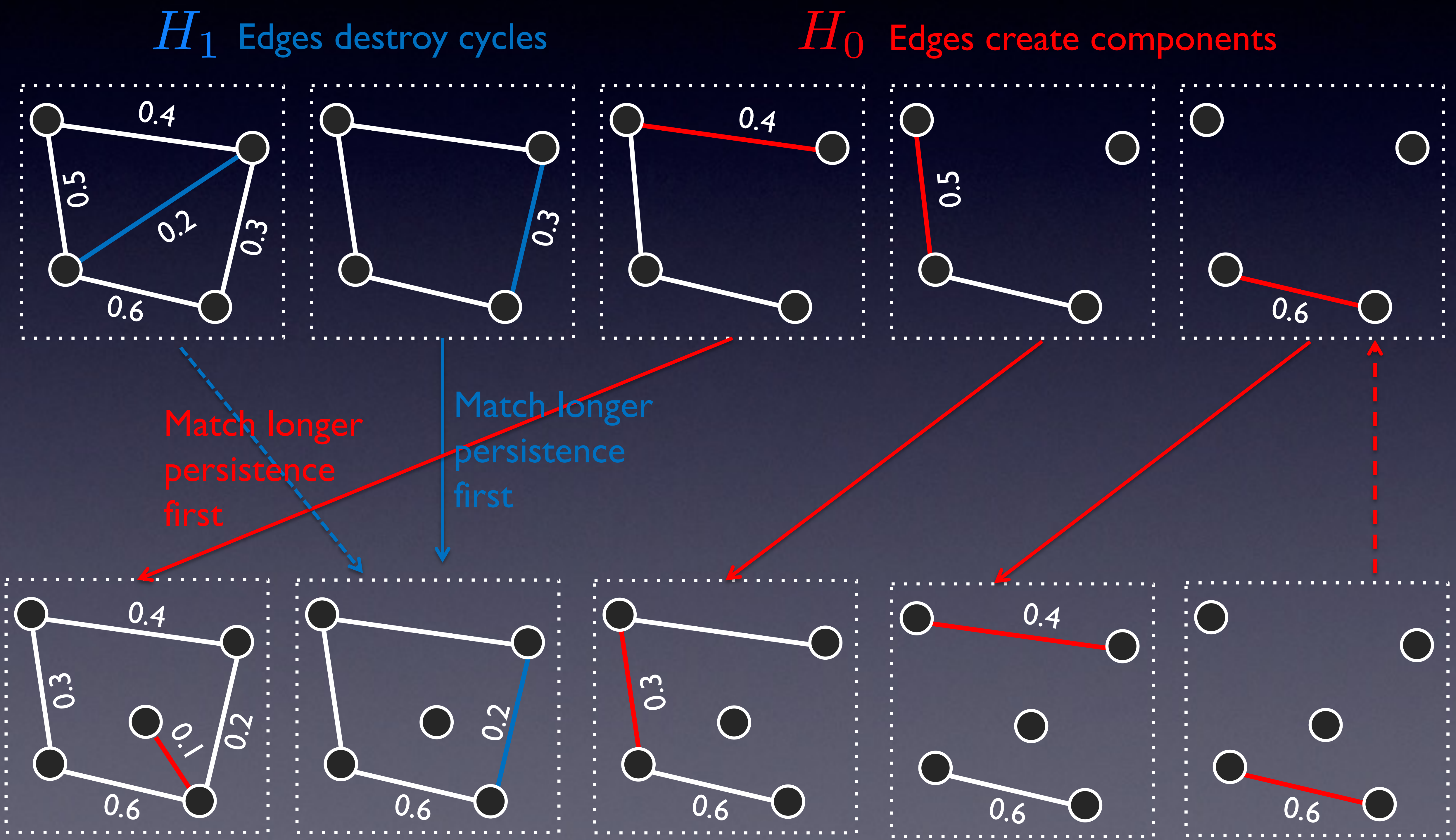
Gives bijection of edges between graphs

Wasserstein graph matching

Red edges = MST



Data augmentation for different network sizes

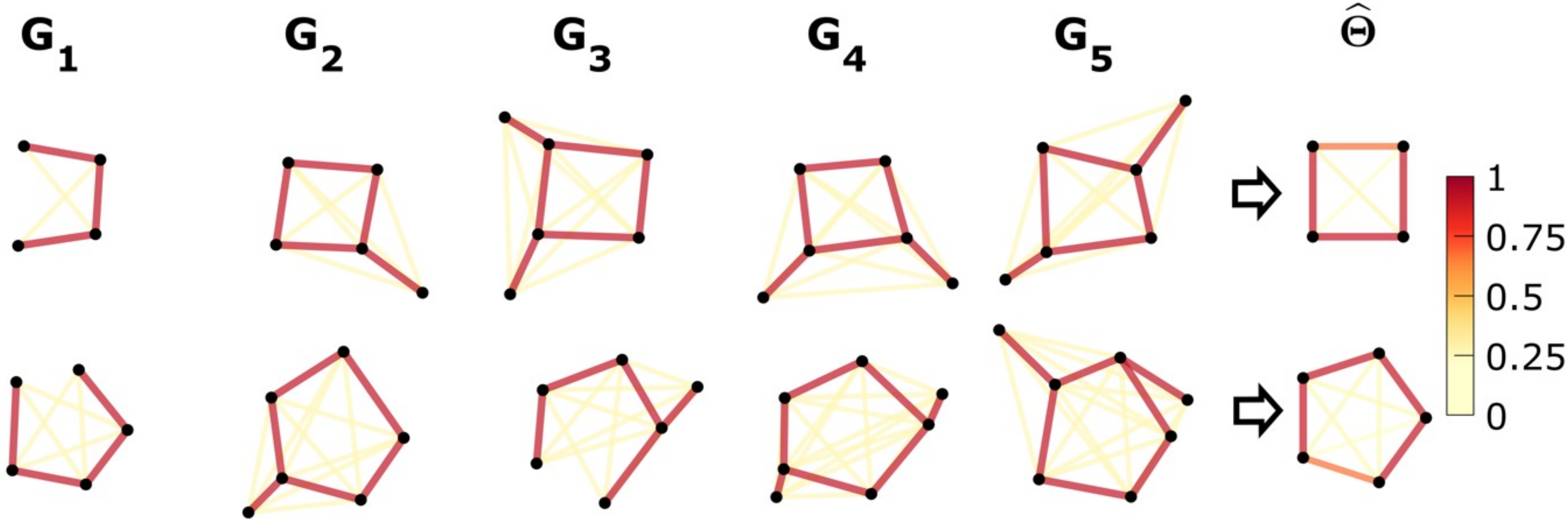


Wassertein graph mean

$$\widehat{\Theta} = \arg\min_{\Theta} \sum_{k=1}^{n} \mathcal{L}_{top}(\Theta, G_k)$$

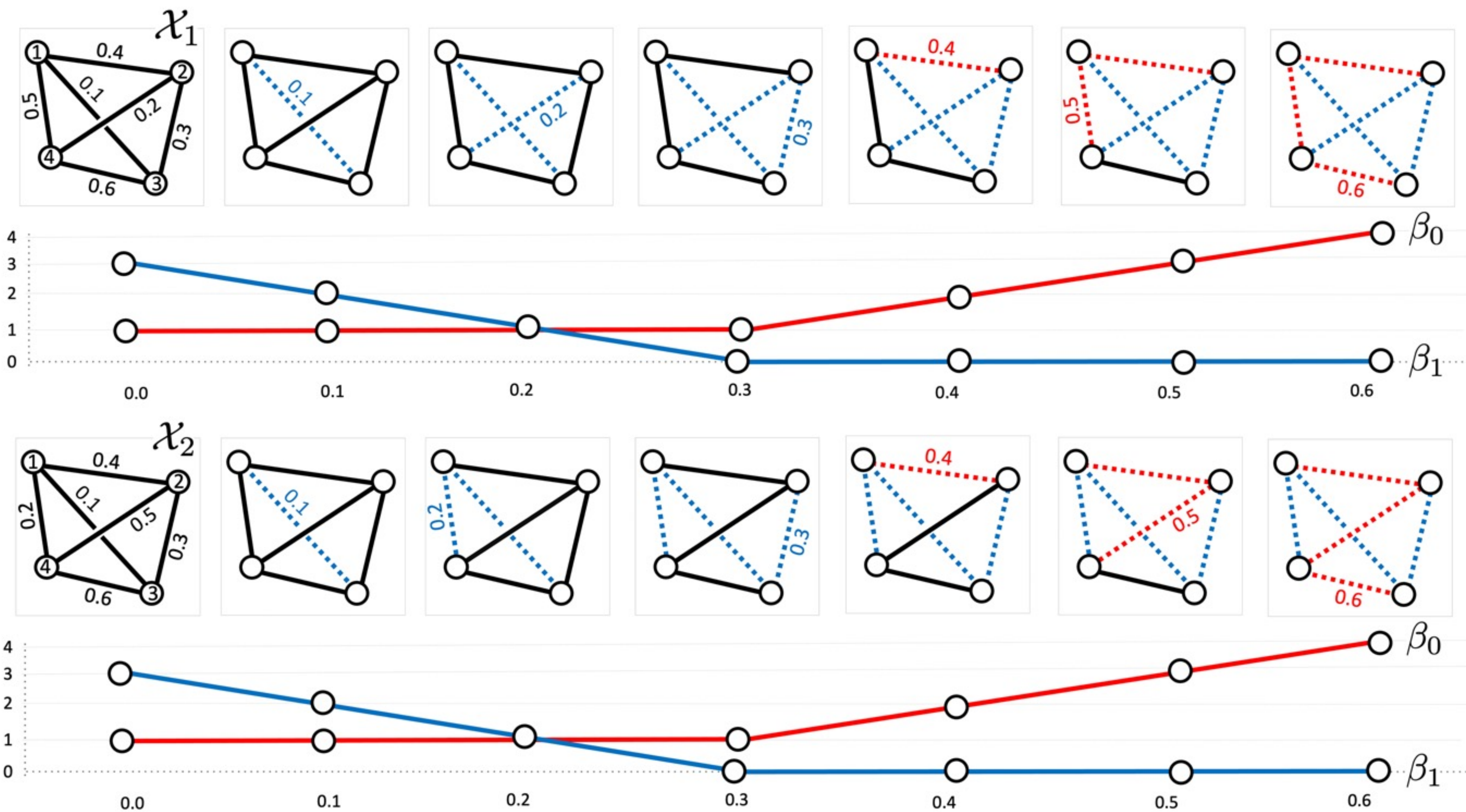
$$\mathcal{L}_{top}(\Theta, P) = \mathcal{L}_{0D}(\Theta, P) + \mathcal{L}_{1D}(\Theta, P)$$

Death values of Θ are given by averaging the sorted death values of all the networks G_k .



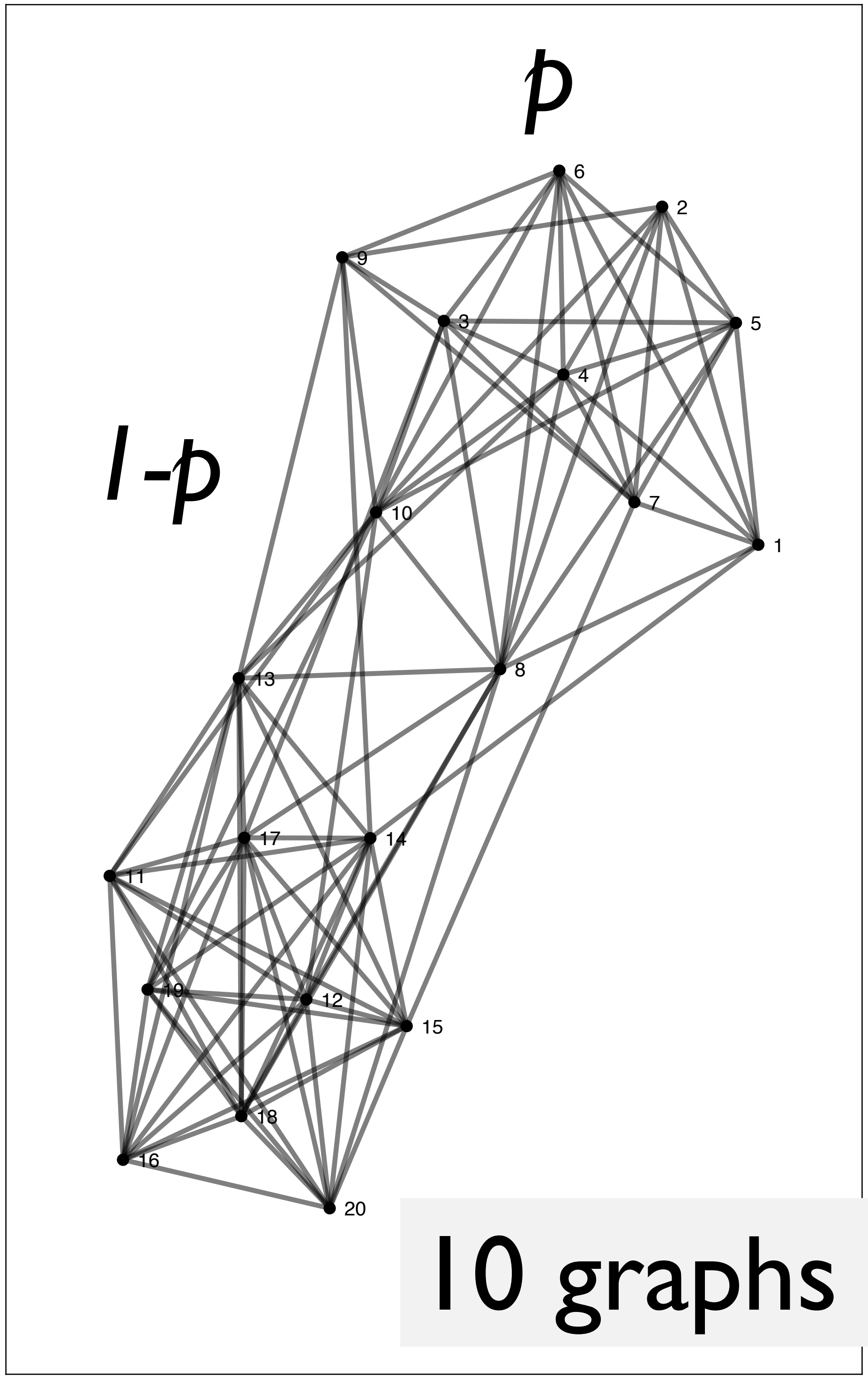
Graph mean is not uniquely defined

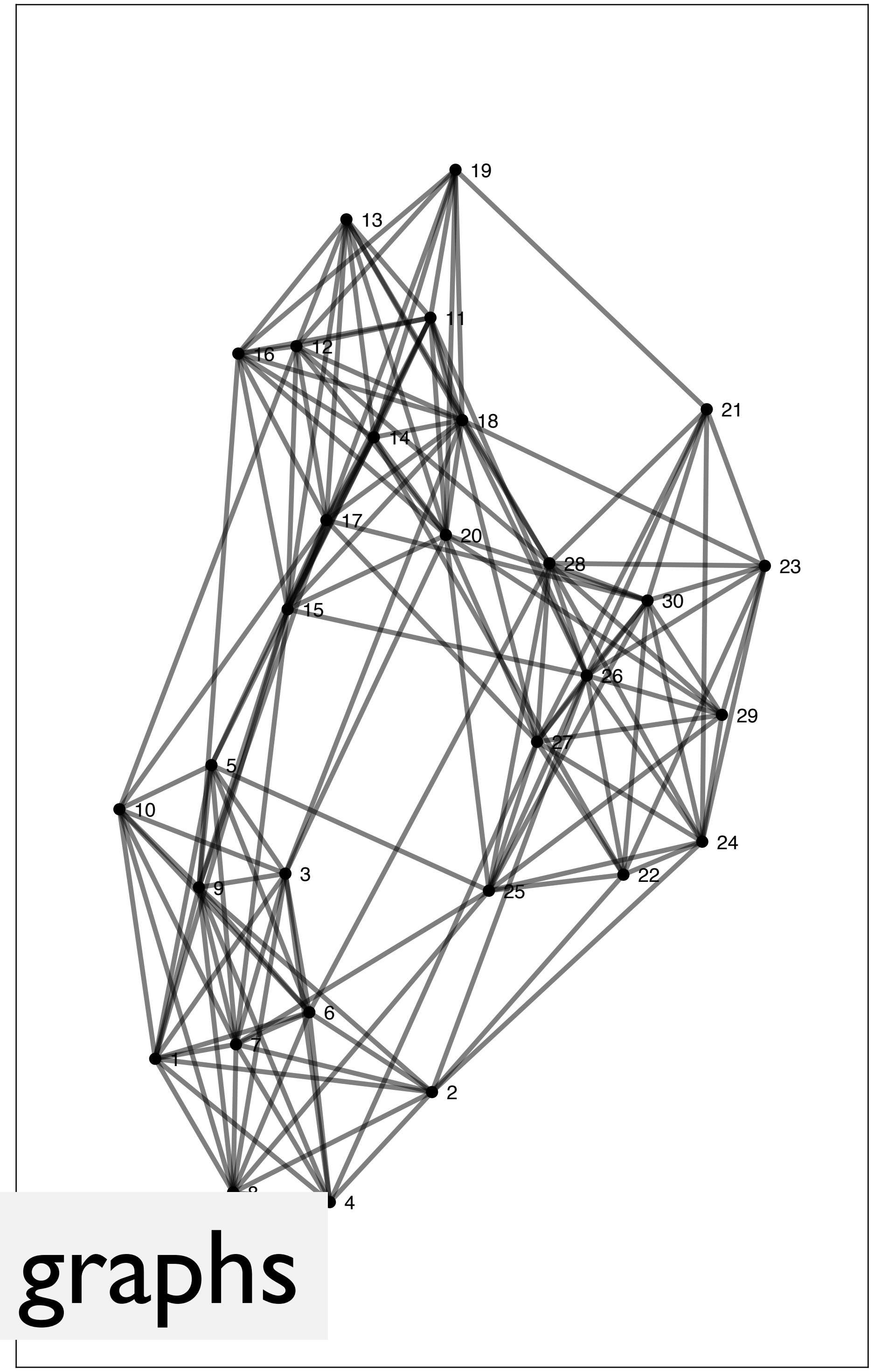
Visual proof: Non-uniqueness of Wasserstein graph mean



Simulation study

Within module connection probability p
Between module connection probability $1\text{-}p$





10 graphs vs. 10 graphs

Graph with 2 modules

Graph with 3 modules

Permutation test for Wasserstein distances

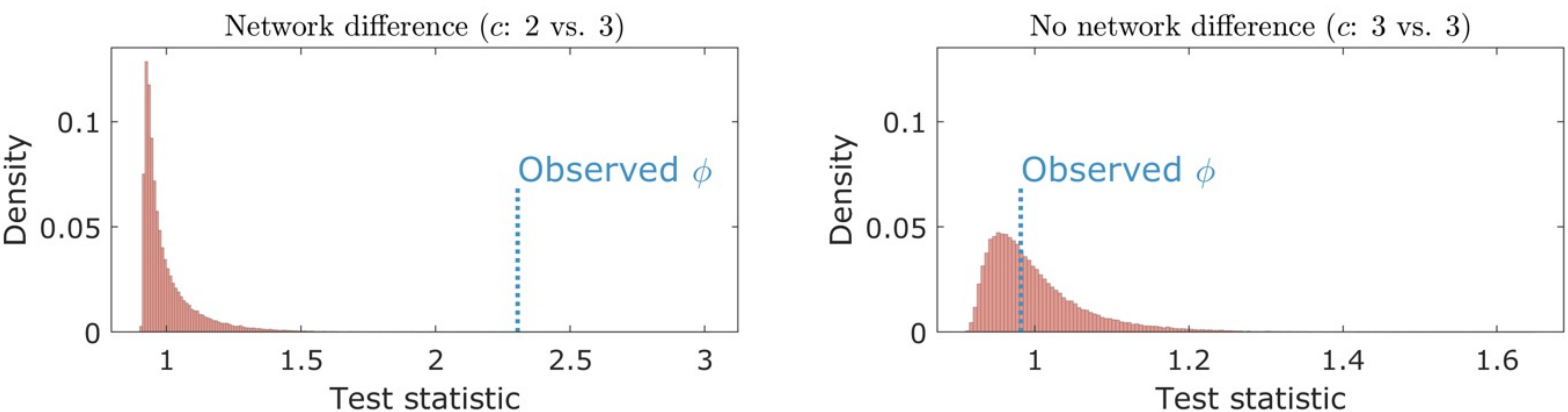
Between-group distance

$$\overline{\mathcal{L}}_B \propto \sum_{i \in C_1, j \in C_2} \mathcal{L}(G_i, G_j)$$

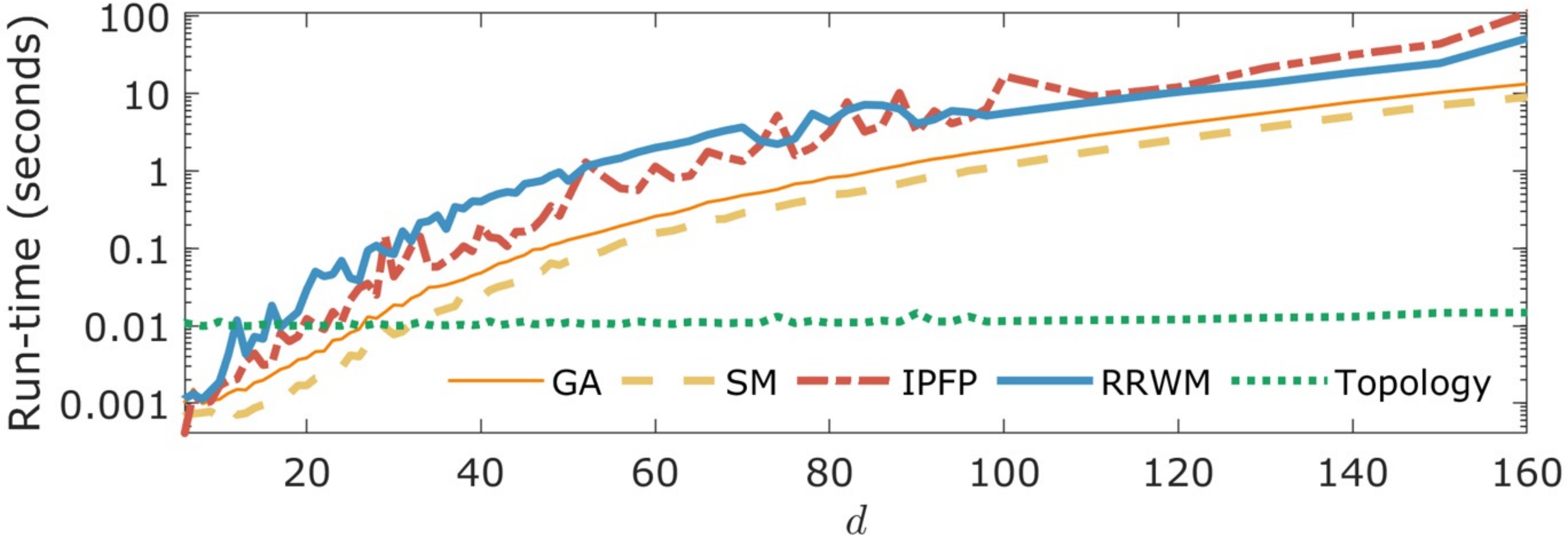
Within-group distance

$$\overline{\mathcal{L}}_W \propto \sum_k \sum_{i,j \in C_k} \mathcal{L}(G_i, G_j)$$

Statistic $\phi = \frac{\overline{\mathcal{L}}_B}{\overline{\mathcal{L}}_W}$



Permutation test *cannot* be applied to existing graph matching algorithms!

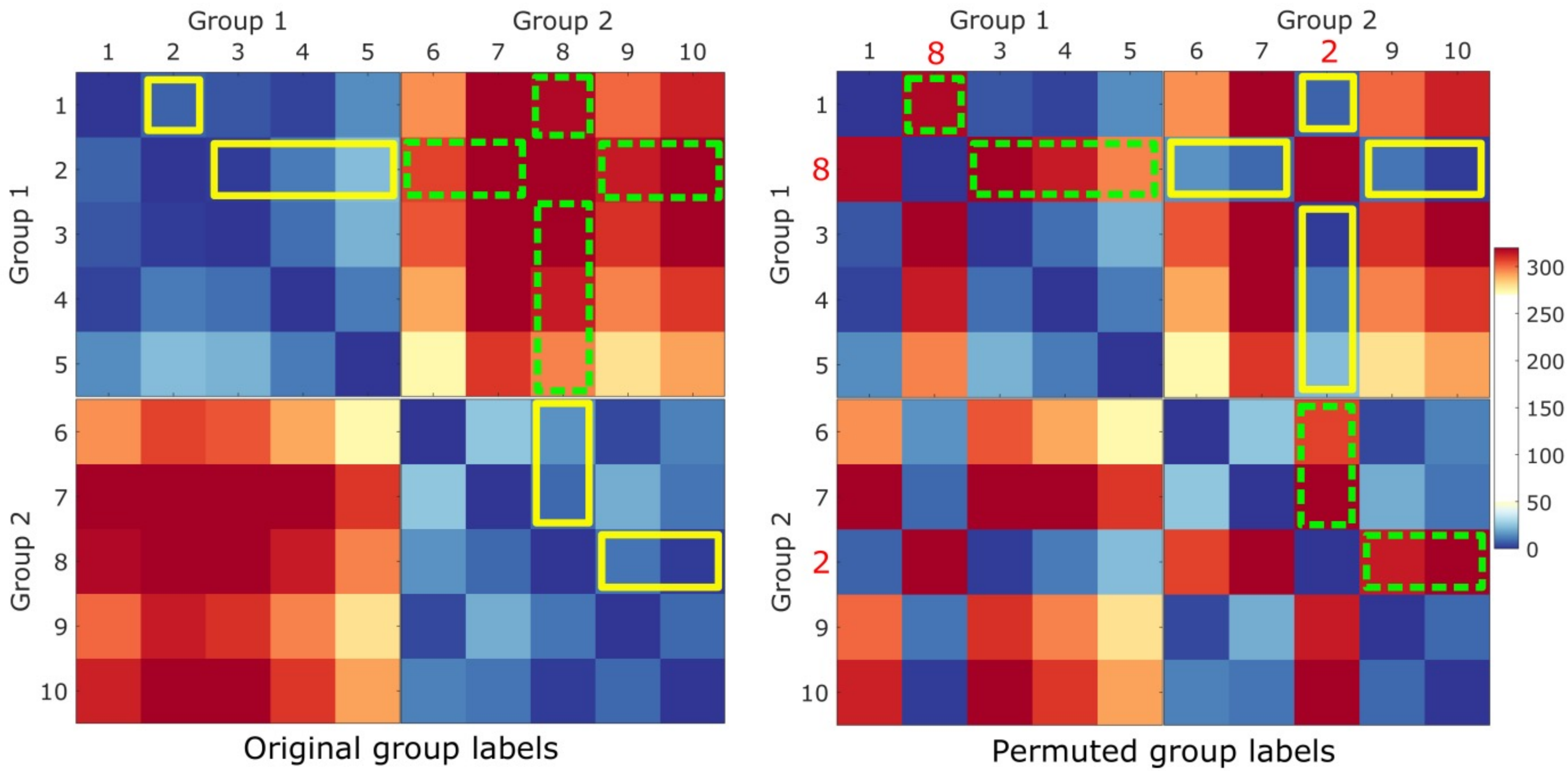


→ Rapid permutation test via random transpositions

Chung et al. 2019 Connectomics in NeuroImaging

Transposition test on Wasserstein distance

Subject 2 in group 1 transposed with subject 8 in group 2



Simply increment changes of loss functions over transposition

$$\overline{\mathcal{L}}_W \rightarrow \overline{\mathcal{L}}_W + \Delta(tranposition)$$

$$\overline{\mathcal{L}}_B \rightarrow \overline{\mathcal{L}}_B + \Delta(tranposition)$$

Average p-value in 50 independent simulations

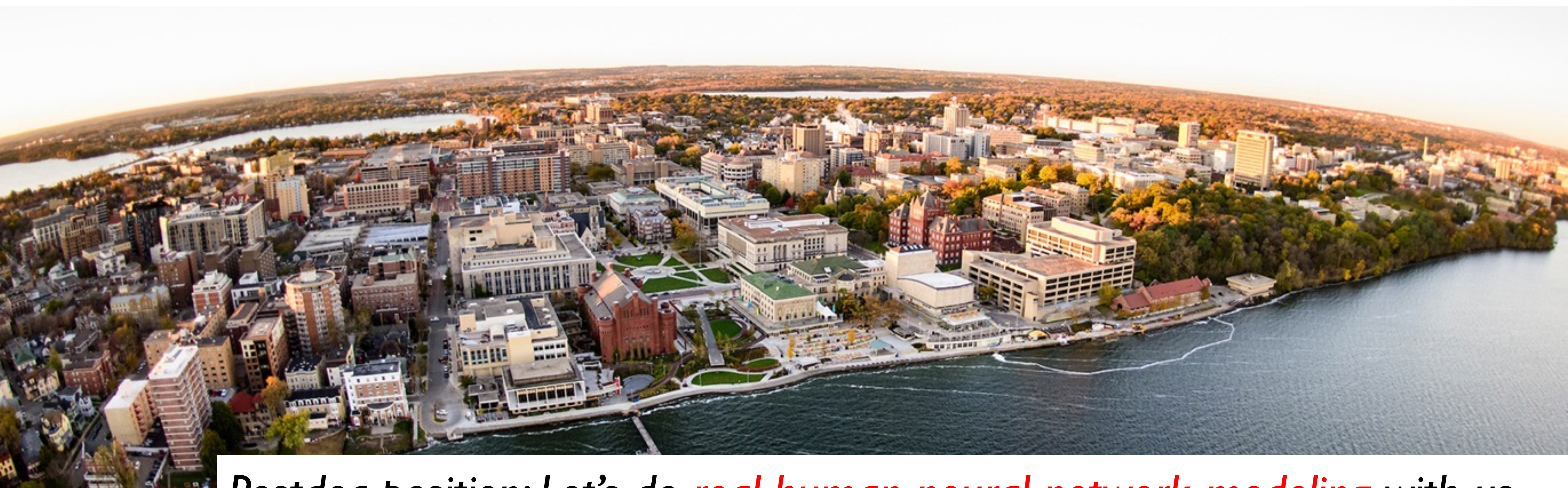
nodes	modules	p	GA (Graduated assignment)	SM (Spectral matching)	RRWM (Reweighted random walk matching)	IPFP (Integer projected fixed point)	$\mathcal{L}_{top}$
12 vs. 12	2 vs. 3	0.6	0.45 ± 0.27	0.48 ± 0.30	0.28 ± 0.31	0.34 ± 0.28	0.08 ± 0.16
		0.8	0.26 ± 0.24	0.30 ± 0.28	0.06 ± 0.12	0.28 ± 0.28	0.01 ± 0.03
	2 vs. 6	0.6	0.06 ± 0.10	0.17 ± 0.20	0.04 ± 0.13	0.23 ± 0.28	0.00 ± 0.00
		0.8	0.00 ± 0.01	0.01 ± 0.03	0.00 ± 0.00	0.02 ± 0.04	0.00 ± 0.00
	3 vs. 6	0.6	0.40 ± 0.29	0.35 ± 0.28	0.24 ± 0.26	0.35 ± 0.28	0.06 ± 0.13
		0.8	0.21 ± 0.23	0.28 ± 0.27	0.08 ± 0.14	0.26 ± 0.25	0.00 ± 0.01
18 vs. 18	2 vs. 3	0.6	0.25 ± 0.23	0.41 ± 0.26	0.26 ± 0.24	0.42 ± 0.28	0.01 ± 0.02
		0.8	0.12 ± 0.17	0.19 ± 0.22	0.00 ± 0.00	0.04 ± 0.05	0.00 ± 0.00
	2 vs. 6	0.6	0.02 ± 0.05	0.07 ± 0.17	0.00 ± 0.00	0.14 ± 0.20	0.00 ± 0.00
		0.8	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
	3 vs. 6	0.6	0.28 ± 0.24	0.37 ± 0.31	0.21 ± 0.24	0.37 ± 0.30	0.01 ± 0.01
		0.8	0.15 ± 0.22	0.13 ± 0.14	0.00 ± 0.01	0.16 ± 0.18	0.00 ± 0.00
24 vs. 24	2 vs. 3	0.6	0.23 ± 0.25	0.30 ± 0.26	0.14 ± 0.20	0.31 ± 0.28	0.00 ± 0.01
		0.8	0.06 ± 0.11	0.12 ± 0.19	0.00 ± 0.00	0.01 ± 0.05	0.00 ± 0.00
	2 vs. 6	0.6	0.00 ± 0.01	0.03 ± 0.06	0.00 ± 0.00	0.09 ± 0.13	0.00 ± 0.00
		0.8	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
	3 vs. 6	0.6	0.24 ± 0.26	0.29 ± 0.28	0.10 ± 0.13	0.37 ± 0.26	0.00 ± 0.00
		0.8	0.07 ± 0.12	0.13 ± 0.19	0.00 ± 0.01	0.12 ± 0.19	0.00 ± 0.00

Thank you! Ready for more TDA?

Sept. 27

First MICCAI workshop

http://sites.google.com/view/tda-for-medical-data



Postdoc position: Let’s do real human neural network modeling with us. You don’t want to be satisfied with artificial neural networks. :P