The Waisman Laboratory for Brain Imaging and Behavior

University of Wisconsin SCHOOL OF MEDICINE AND PUBLIC HEALTH

Topological inference and learning for graphs

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

Penn Statistics in Imaging and Visualization Endeavor Center, January 18, 2022 12:30pm

Acknowledgement

Vijay Anand D., Soumya Das, Sixtus Darkurah, Tananun Songdechakraiwut, Zhan Luo, Ian Carroll, Andrew Alexander, Richard Davidson, Hill Goldsmith Univ. of Wisconsin-Madison

Yuan Wang University of South Carolina
Li Shen Univ. of Pennsylvania
Anass El Yaagoubi Bourakna, Hernando Ombao, KAUST, Saudi Arabia
Hyekyung Lee , Dong Soo Lee Seoul National University, Korea
Shih-Gu Huang, Anqi Qiu National University of Singapore
Victor Solo University of New South Wales

Grants:

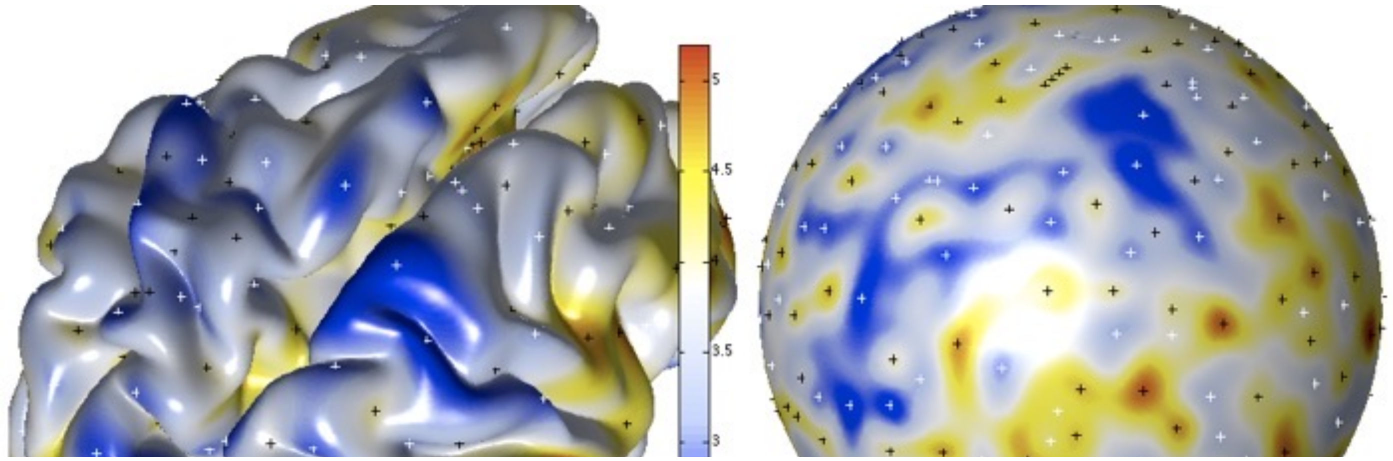
NIH R01 EB022856, R01 EB028753, UL1 TR002373, NSF DMS-2010778

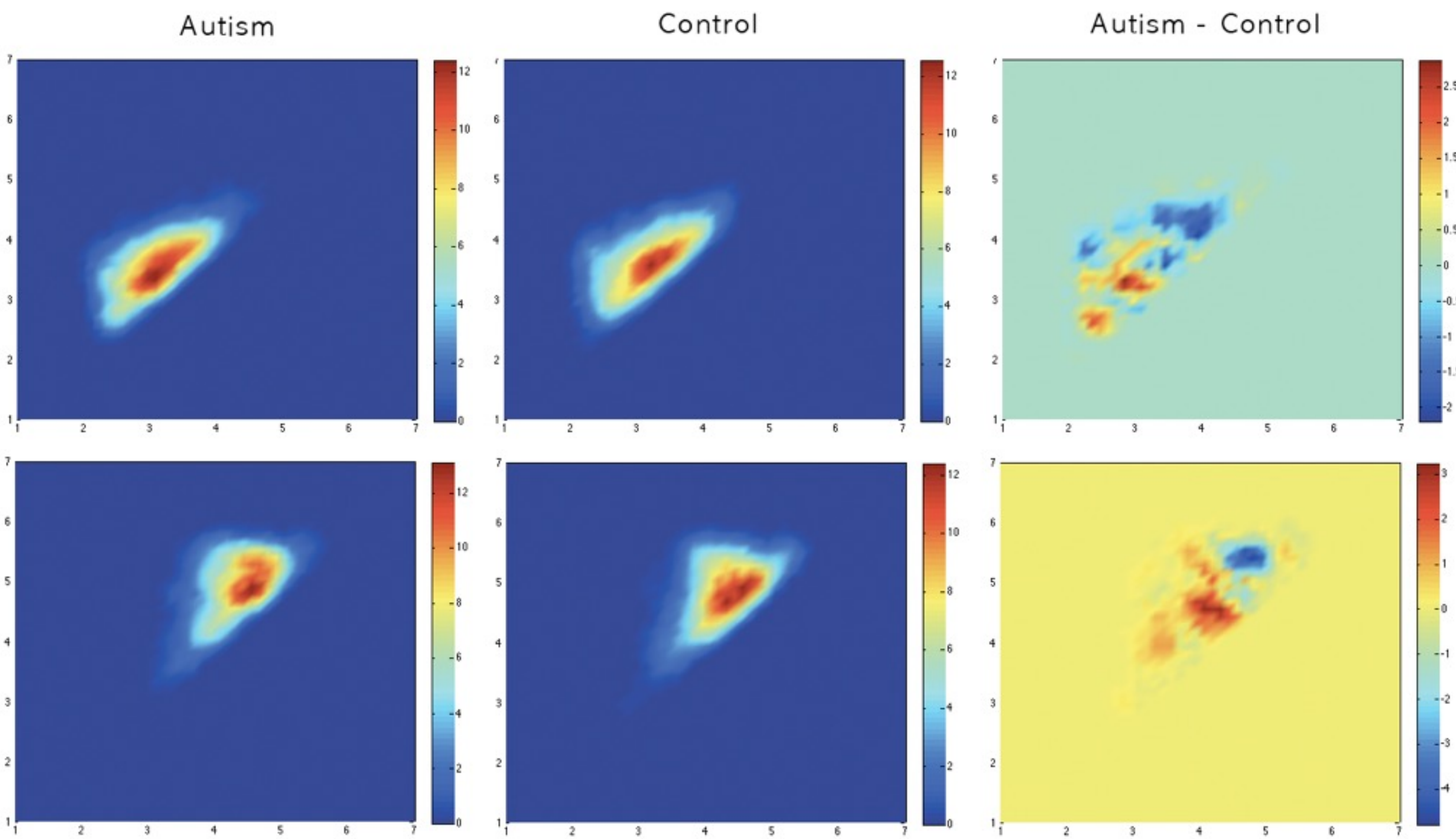
Abstract

Many previous studies on networks have mainly focused analyzing graph theory features that are often parameter dependent. Persistent homology provides a more coherent mathematical framework that is invariant to the choice of parameters. Instead of looking at networks at a fixed scale, persistent homology charts the topological changes of networks over every possible parameter. In doing so, it reveals the most persistent topological features that are robust to parameter changes. In this talk, we present novel topological inference and learning frameworks that can integrate networks of different sizes, topology or modalities through the Wasserstein distances. The use of Wasserstein distances bypasses the intrinsic computational bottleneck associated with persistent homology. It is now possible to perform various graph computations including matching in O(n log n). We demonstrate the versatility of the proposed method through the twin brain imaging study where we determine the extent to which brain networks are genetically heritable. The talk is based on preprints: Songdechakraiwut et al. 2021 (arXiv:2012.0067), Anand et al. 2021 (arXiv:2110.14599) and Chung et al. 2022 (arXiv:2201:00087).

Persistence Diagrams of Cortical Surface Data

Moo K. Chung[1,2], Peter Bubenik[3], and Peter T. Kim[4]

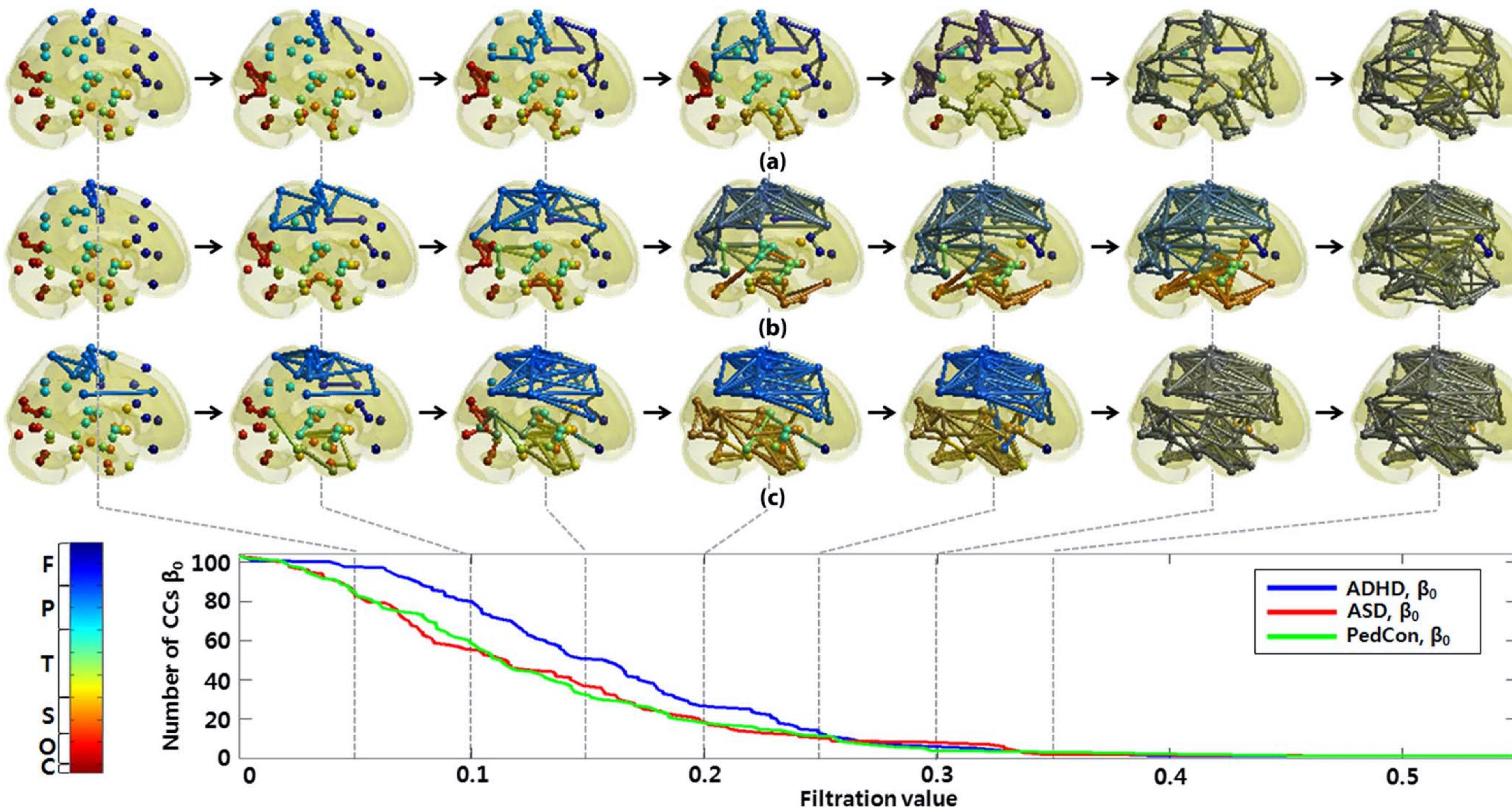




Chung et al., 2009
Information Processing in Medical Imaging (IPMI) 5636:386-397.

First persistent homology paper in medical imaging

Baseline TDA data representation technique



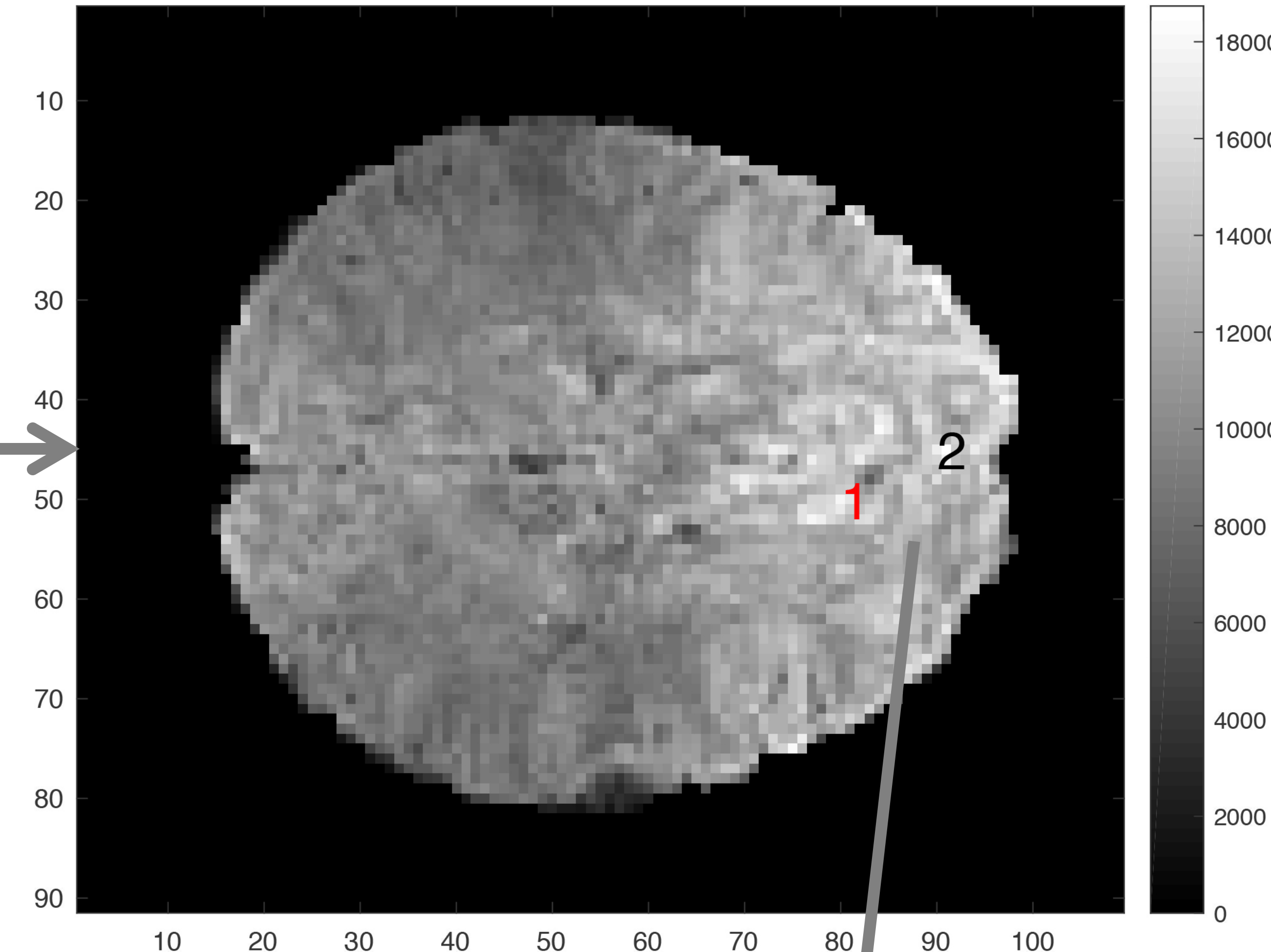
Lee et al. (2011) ISBI

First persistent homology paper in brain network analysis

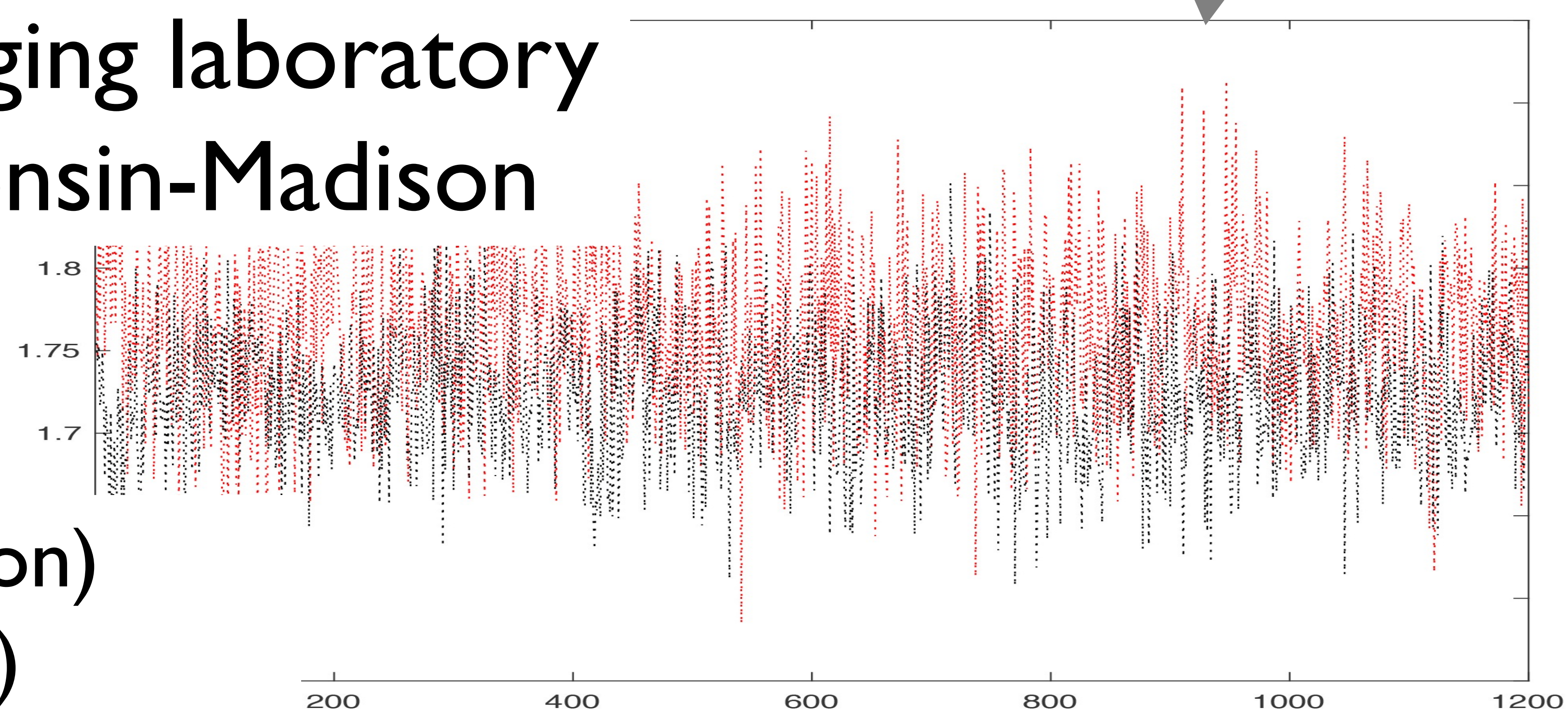
Lee et al. 2012 IEEE Transactions on Medical Imaging 31:2267-2277

Brain Network Data

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



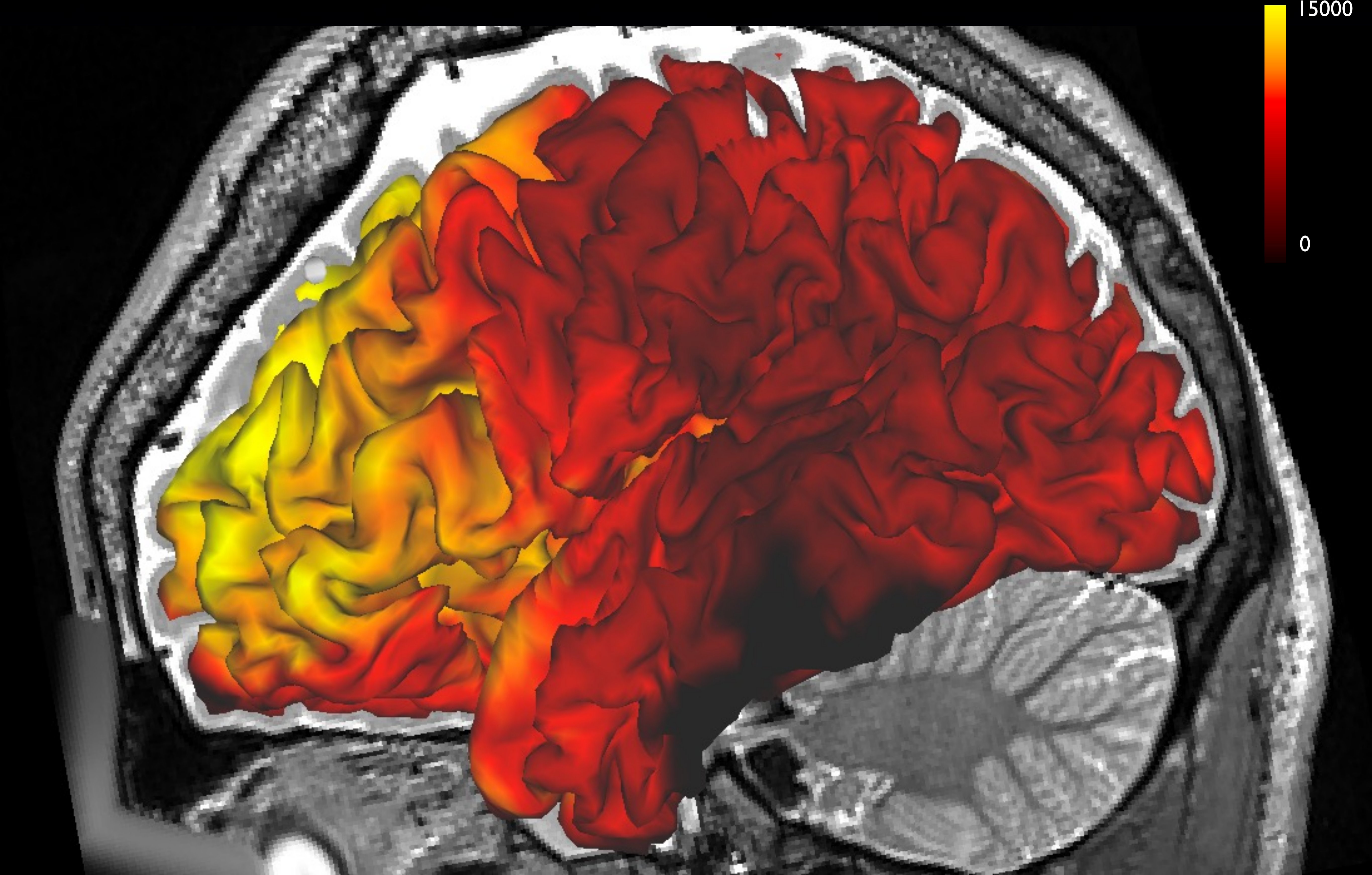
Waisman brain imaging laboratory
University of Wisconsin-Madison



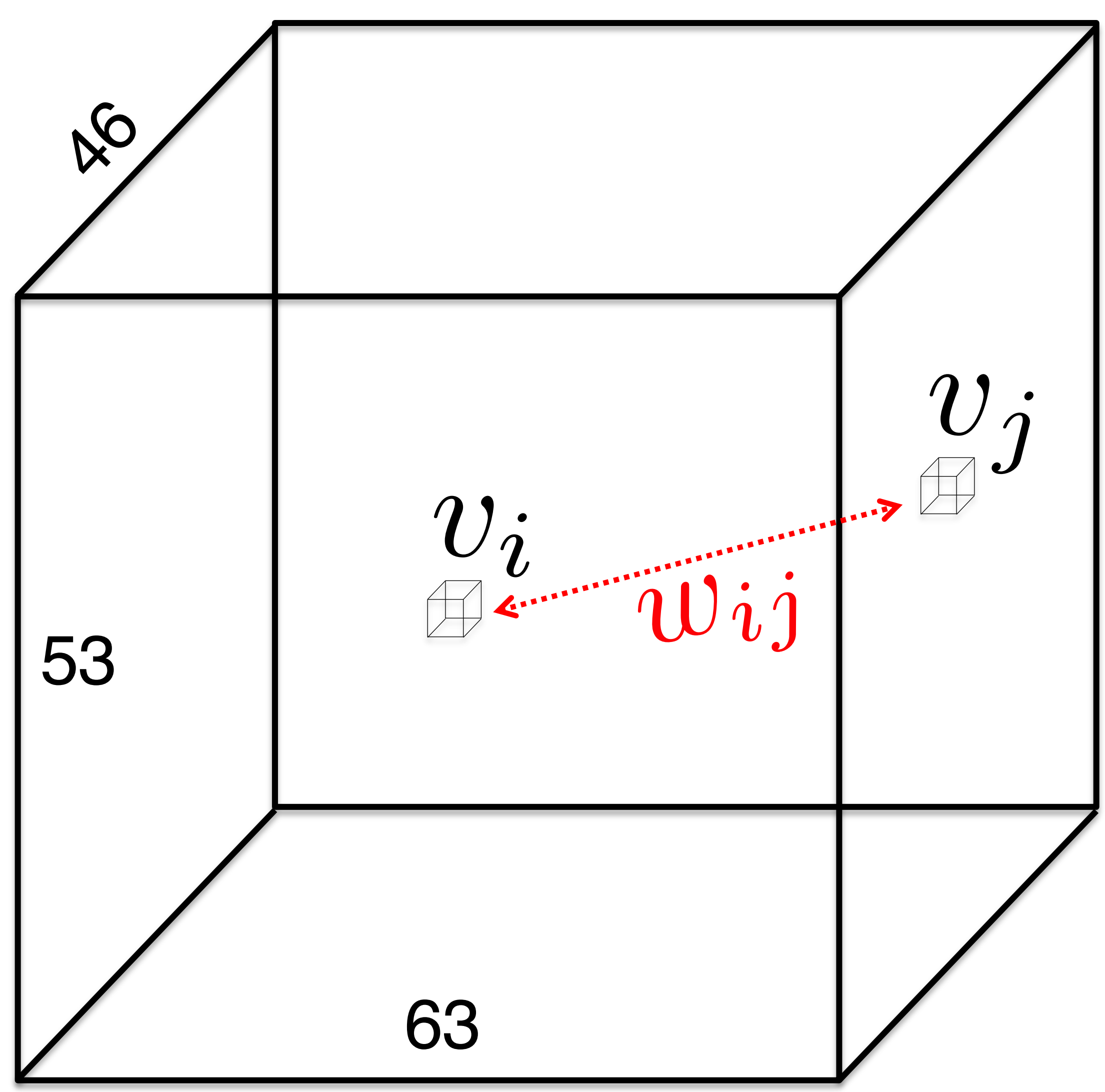
295 time points (Madison)
1200 time points (HCP)

Resting state fMRI (every 30 seconds)

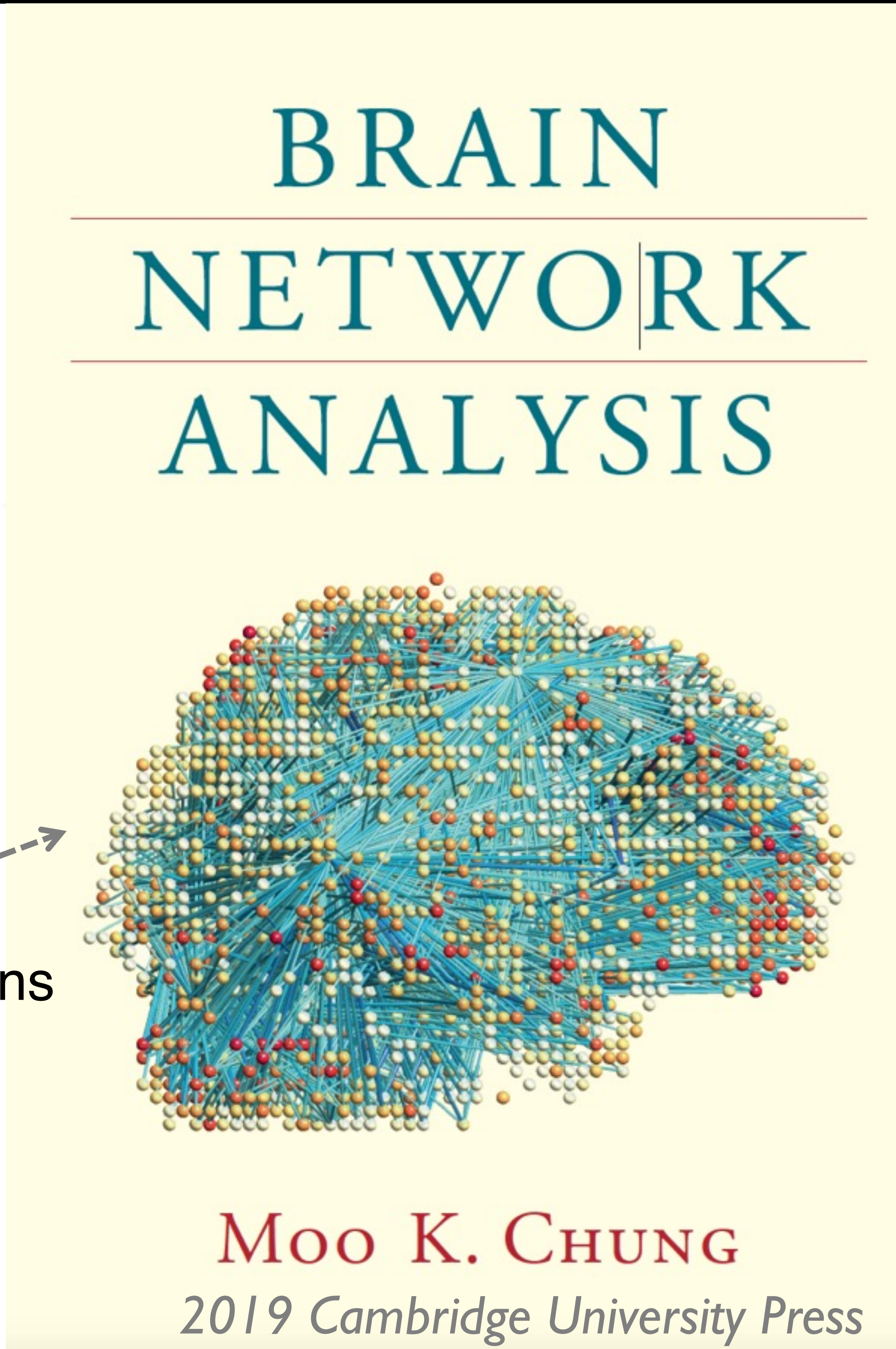
1200 time points x 3D volume



How big brain network data is?



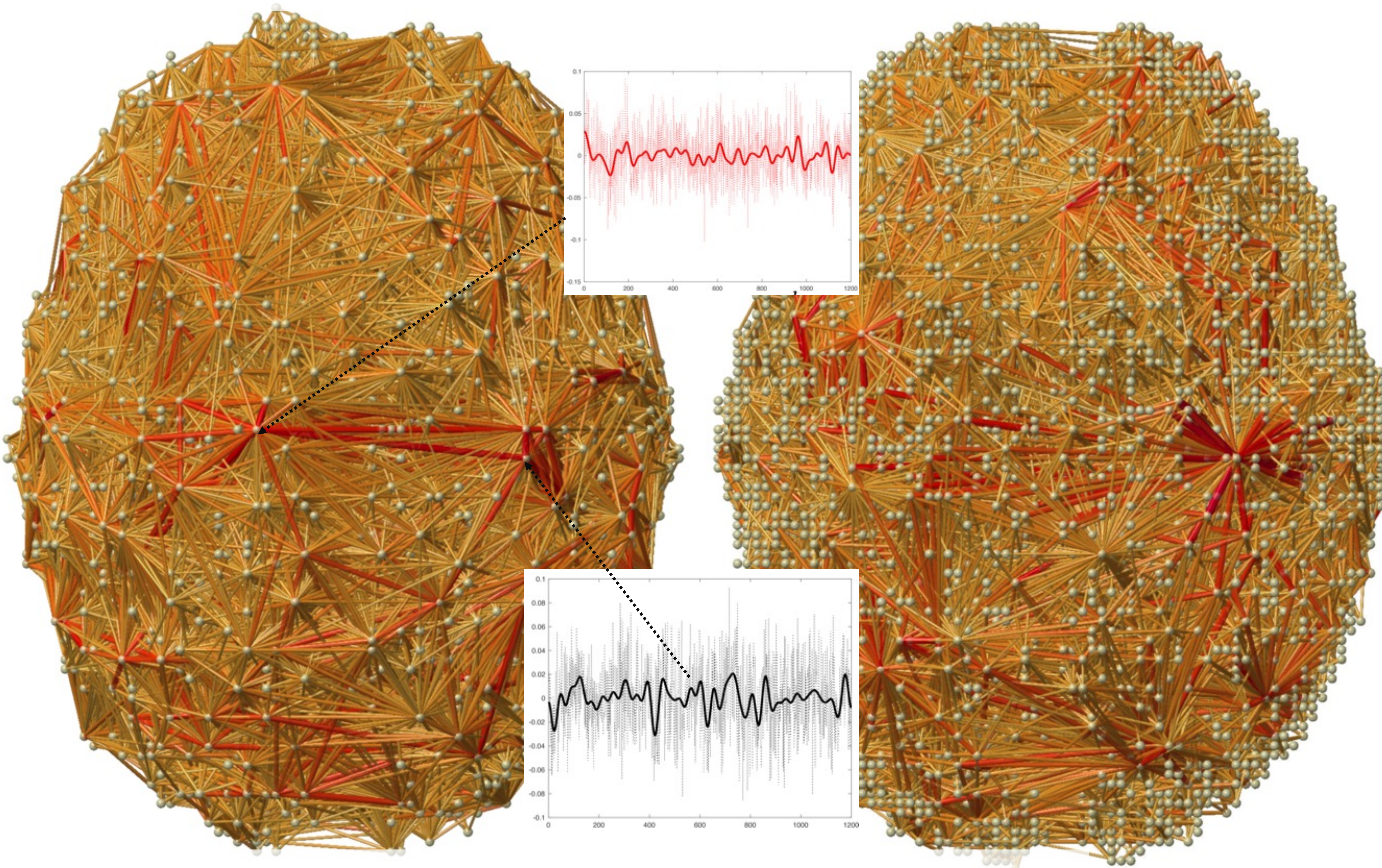
v_i



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

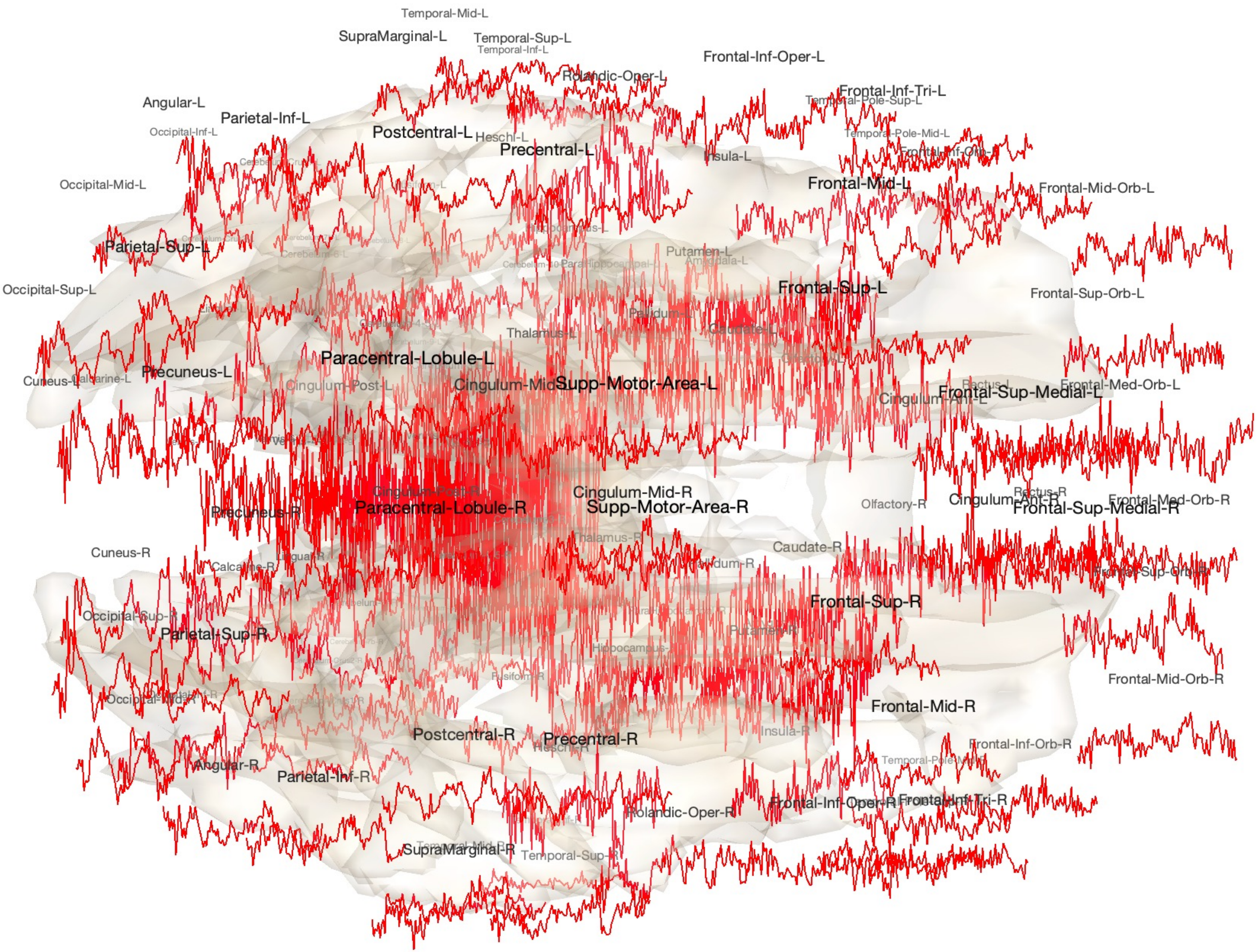
300000 voxels (1mm)
→ 90 billion connections
→ 700 GB memory

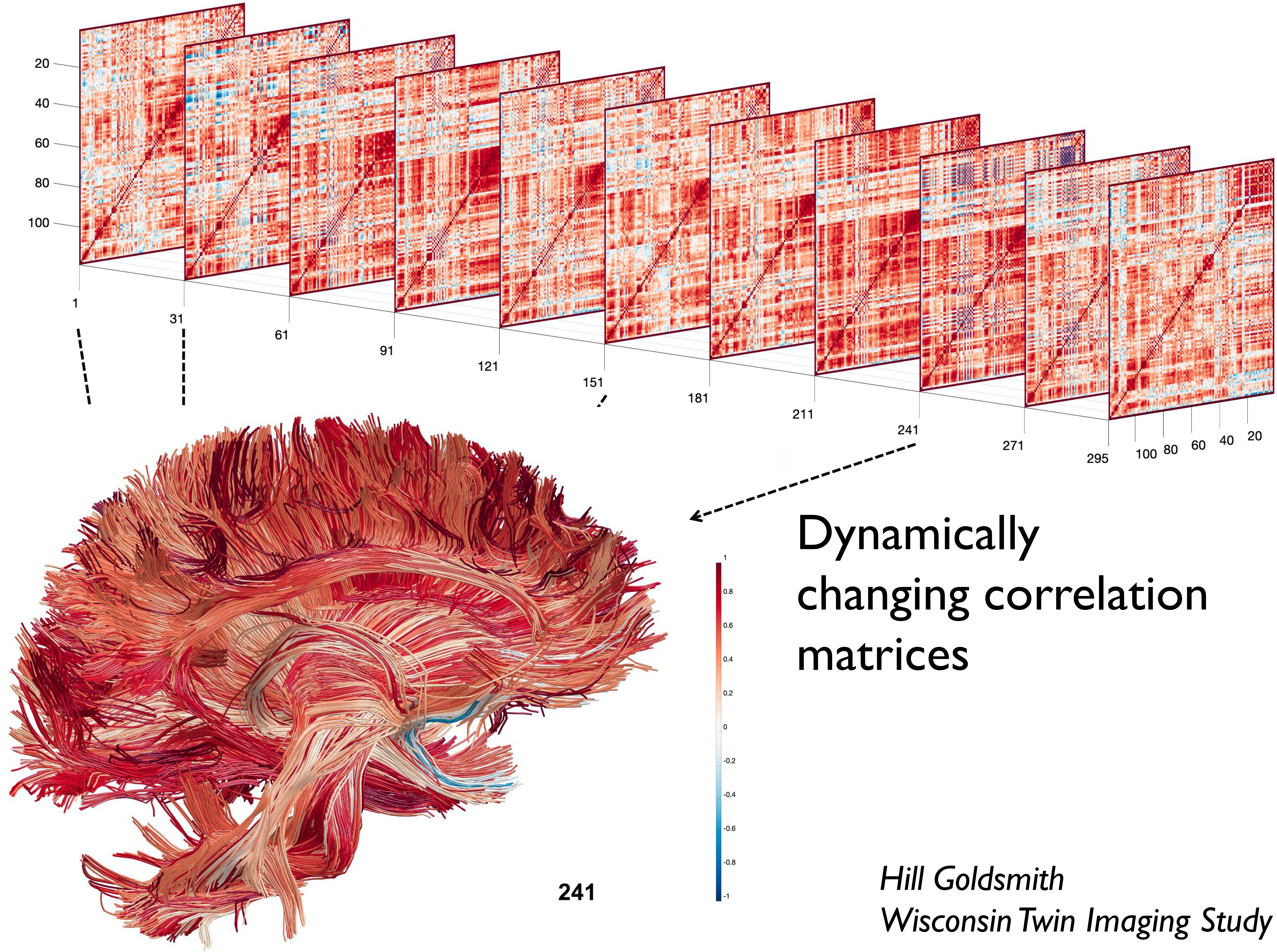
Correlation brain network at voxel level



Correlation network of 300000 time series
Complete graph with about $300000^2/2$ cycles.

Time series averaged into 116 brain regions

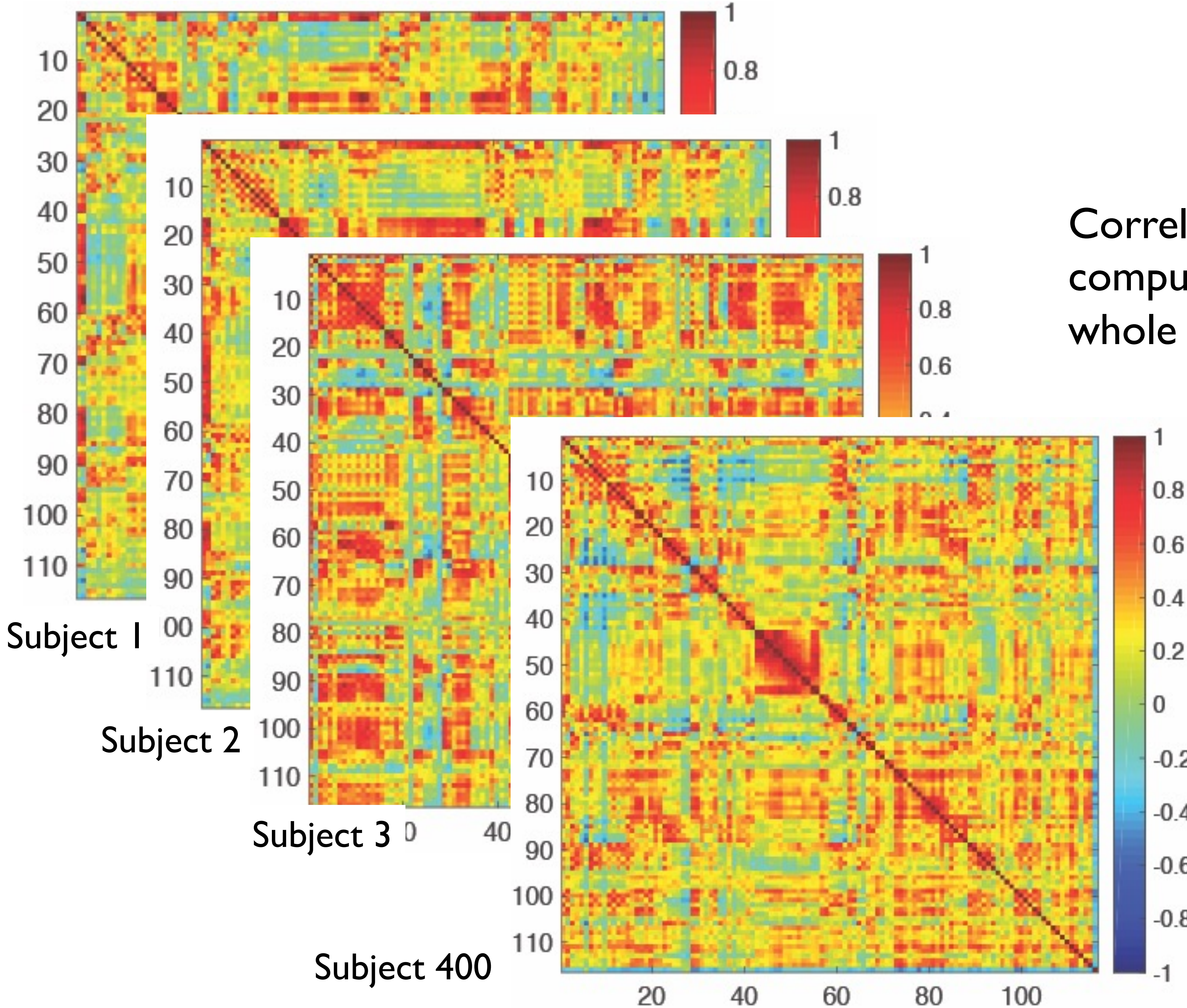




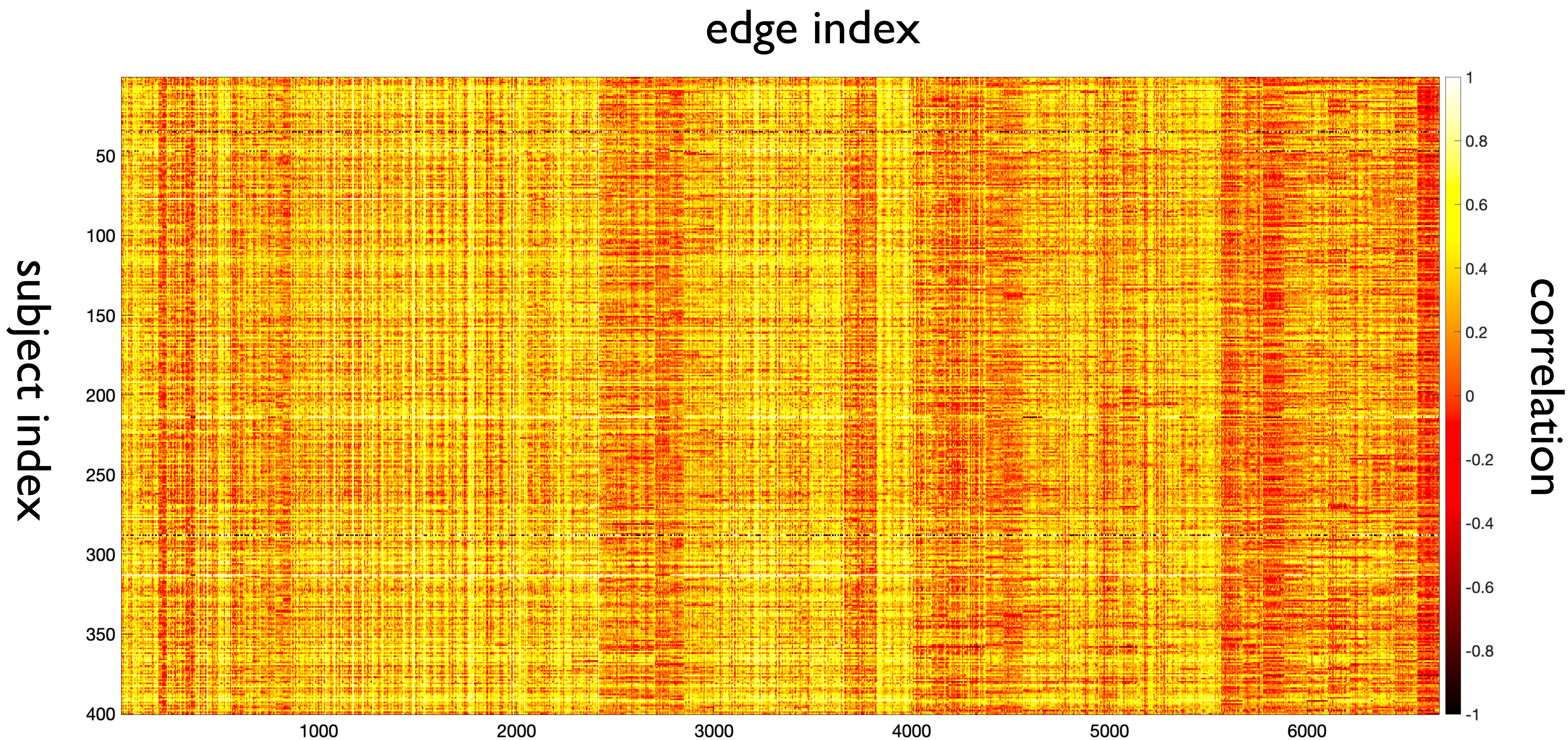
Dynamically changing correlation matrices

Hill Goldsmith
Wisconsin Twin Imaging Study

Subject level brain connectivity matrix



Correlation computed over whole time points



116 nodes networks
6670 edges per network

168 males, 232 females
124 MZ-twins, 70 DZ-twins

54 Subjects Multimodal Brain Network Data

http://github.com/laplcebeltrami/maltreated

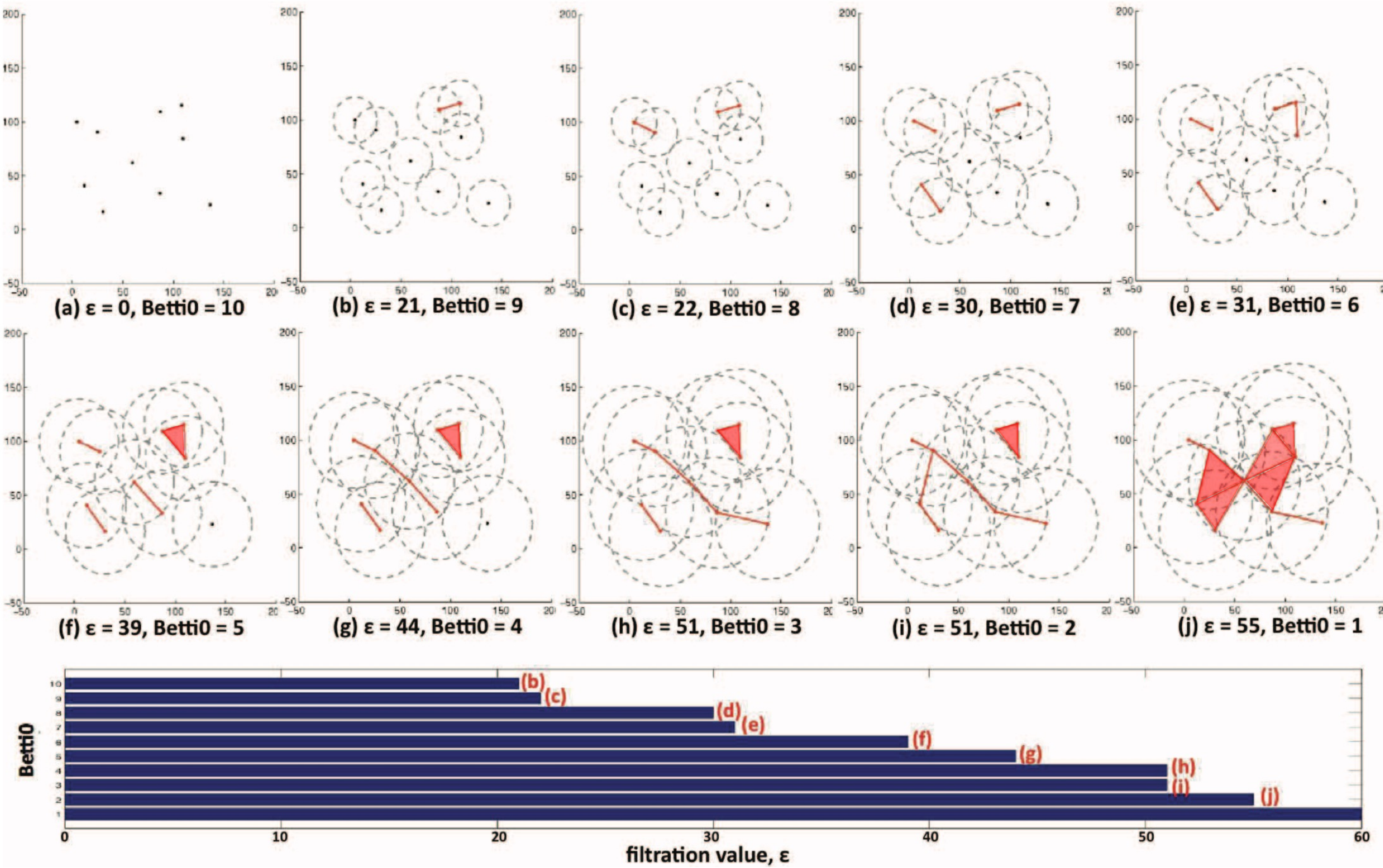
Chung, M.K., Hanson, J.L., Ye, J., Davidson, R.J. Pollak, S.D. 2015 Persistent homology in sparse regression and its application to brain morphometry. *IEEE Transactions on Medical Imaging*, 34:1928-1939

Chung, M.K., Hanson, J.L., Lee, H., Adluru, N., Alexander1, A.L., Davidson, A.L., Pollak, S.D. 2013. Persistent homological sparse network approach to detecting white matter abnormality in maltreated children: MRI and DTI multimodal study, *MICCAI* 8149:300-307

Chung, M.K., Hanson, J.L., Adluru, Aleander, A.L., Davidson, R.J., Pollak, S.D. 2017 Integrative structural brain network analysis in diffusion tensor imaging, *Brain Connectivity* 7:331-346

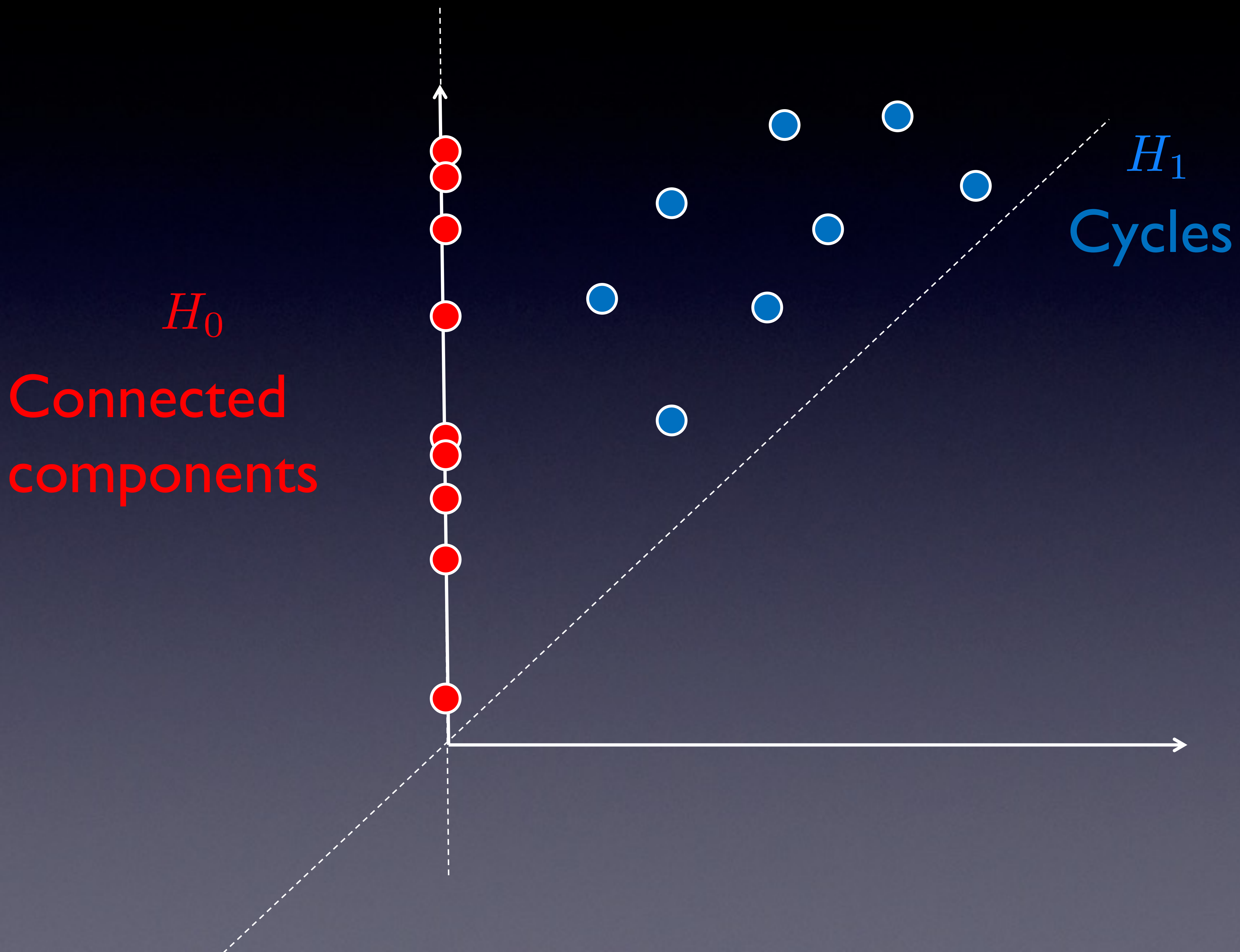
Persistent Homology on Graph filtrations

Rips filtration



Lee et al. 2011 ISBI 841-844

Persistence diagram for Rips filtrations



Graph Filtrations

Weighted complete graph

$$\mathcal{X} = (\underbrace{V}_{\text{Node set}}, \underbrace{w}_{\text{Edge weight}}) \quad w = (w_{ij})$$

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

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

Rips filtration vs. graph filtration

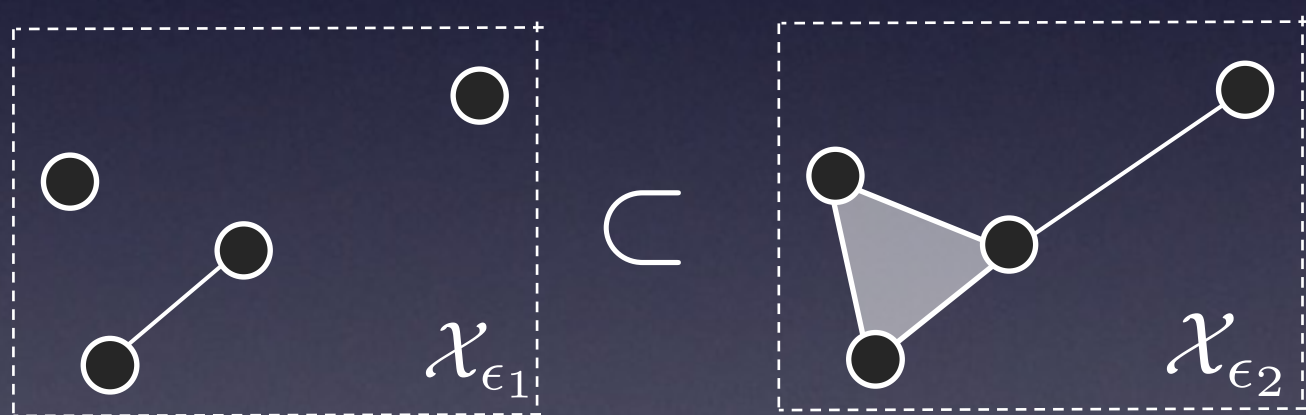
Metric space

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

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

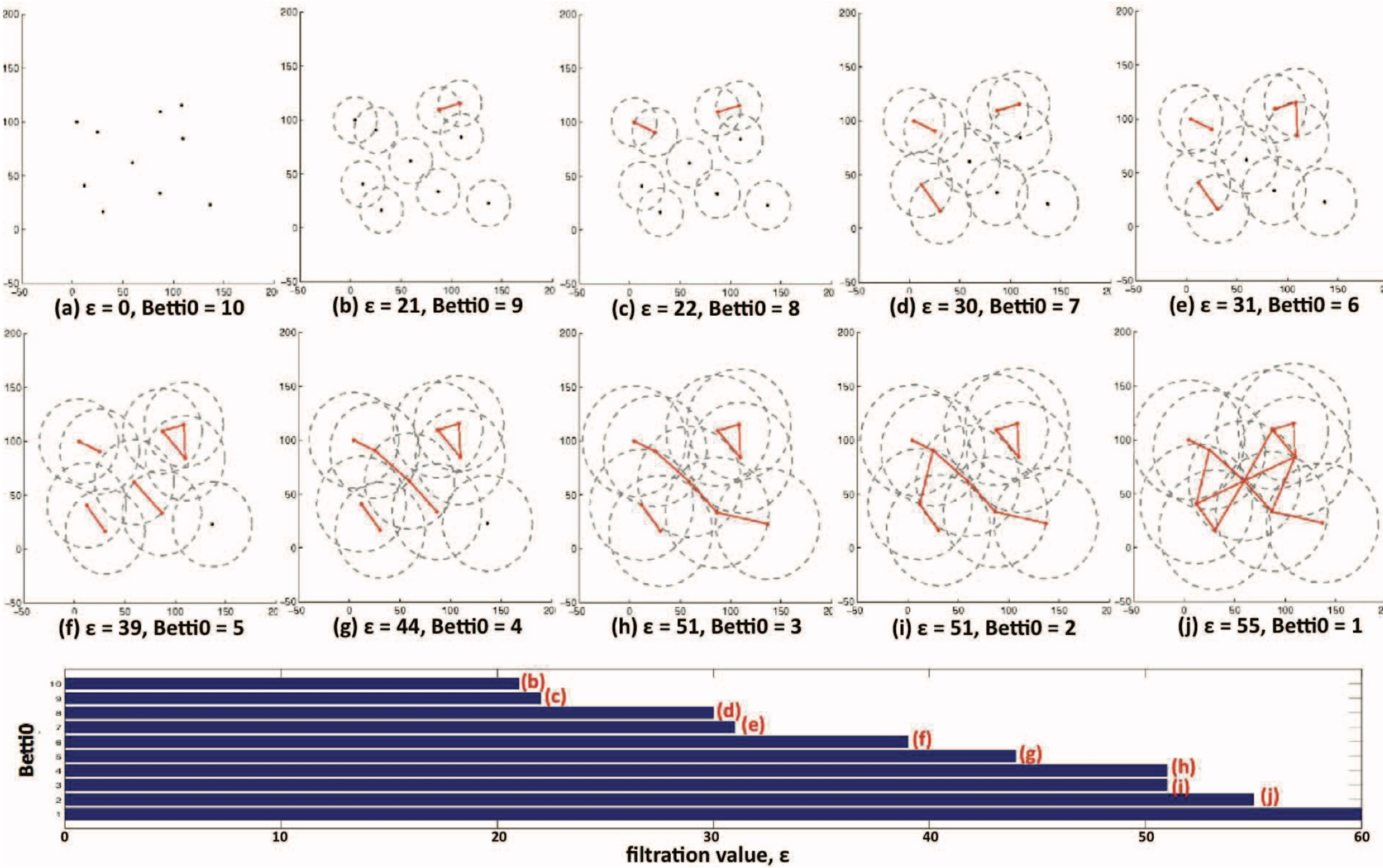
Graph filtration

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

for increased edge weights

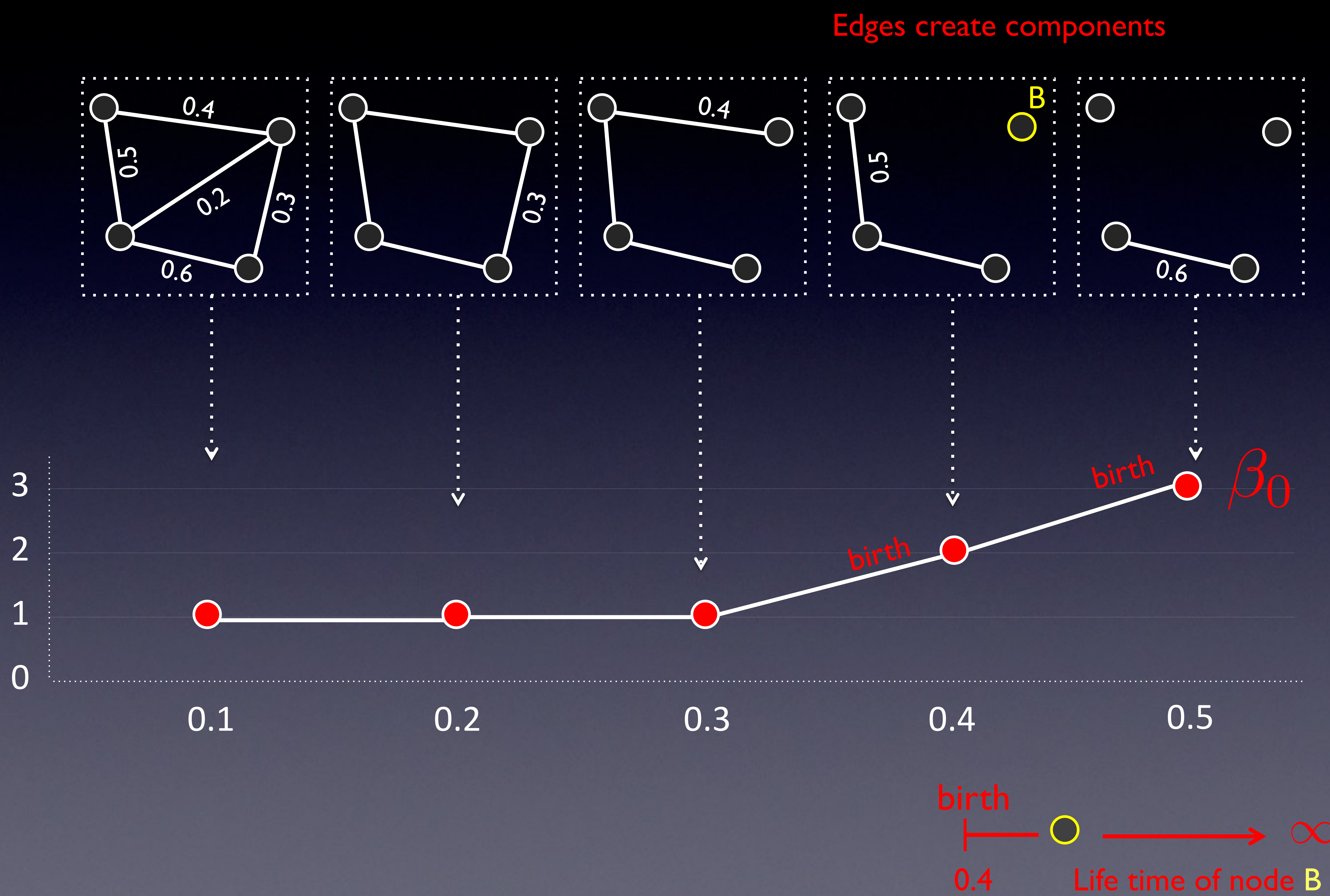
$$\epsilon_0 < \epsilon_1 < \epsilon_2 < \cdots$$

Graph filtration

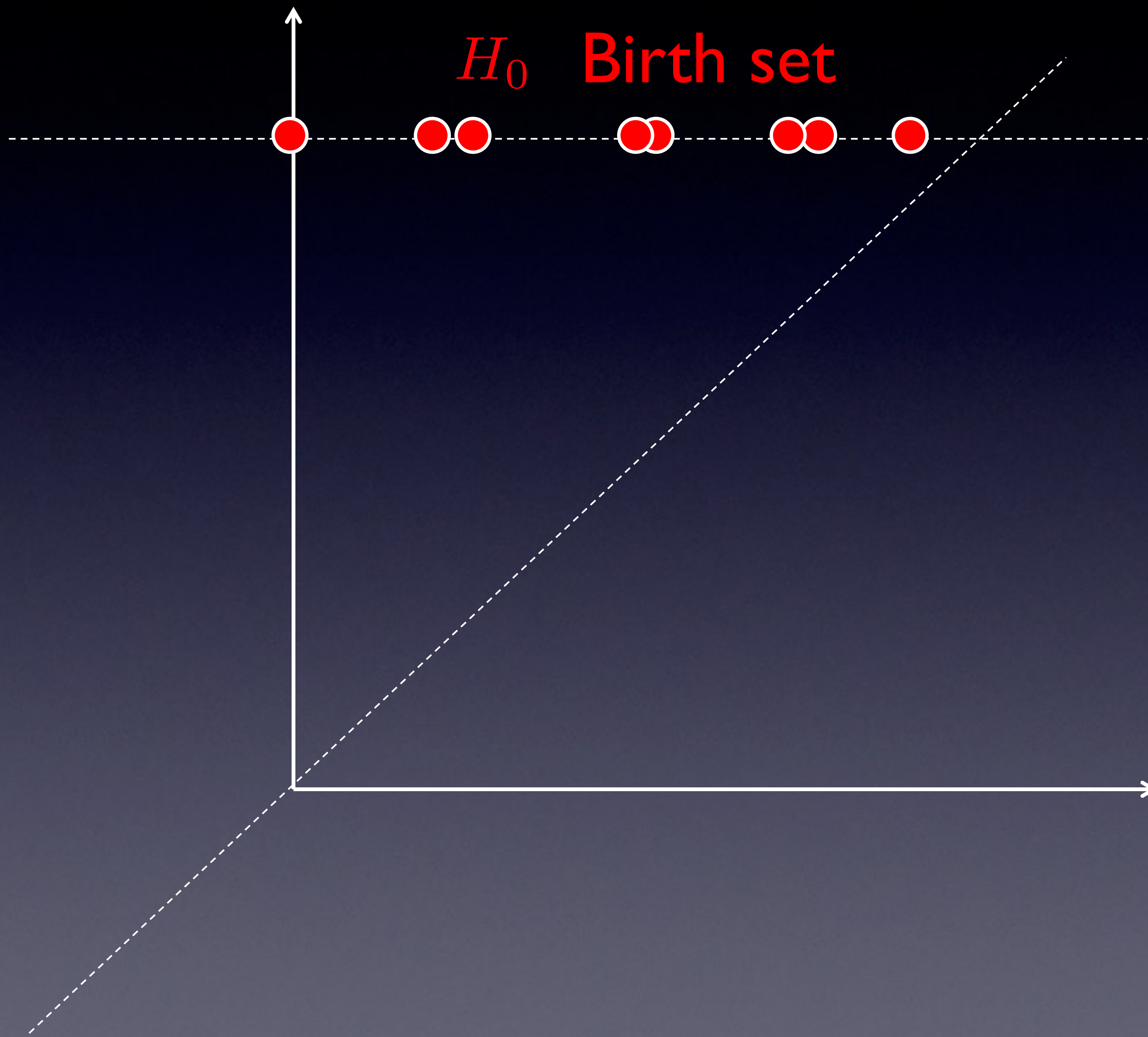


Lee et al. 2011 ISBI 841-844

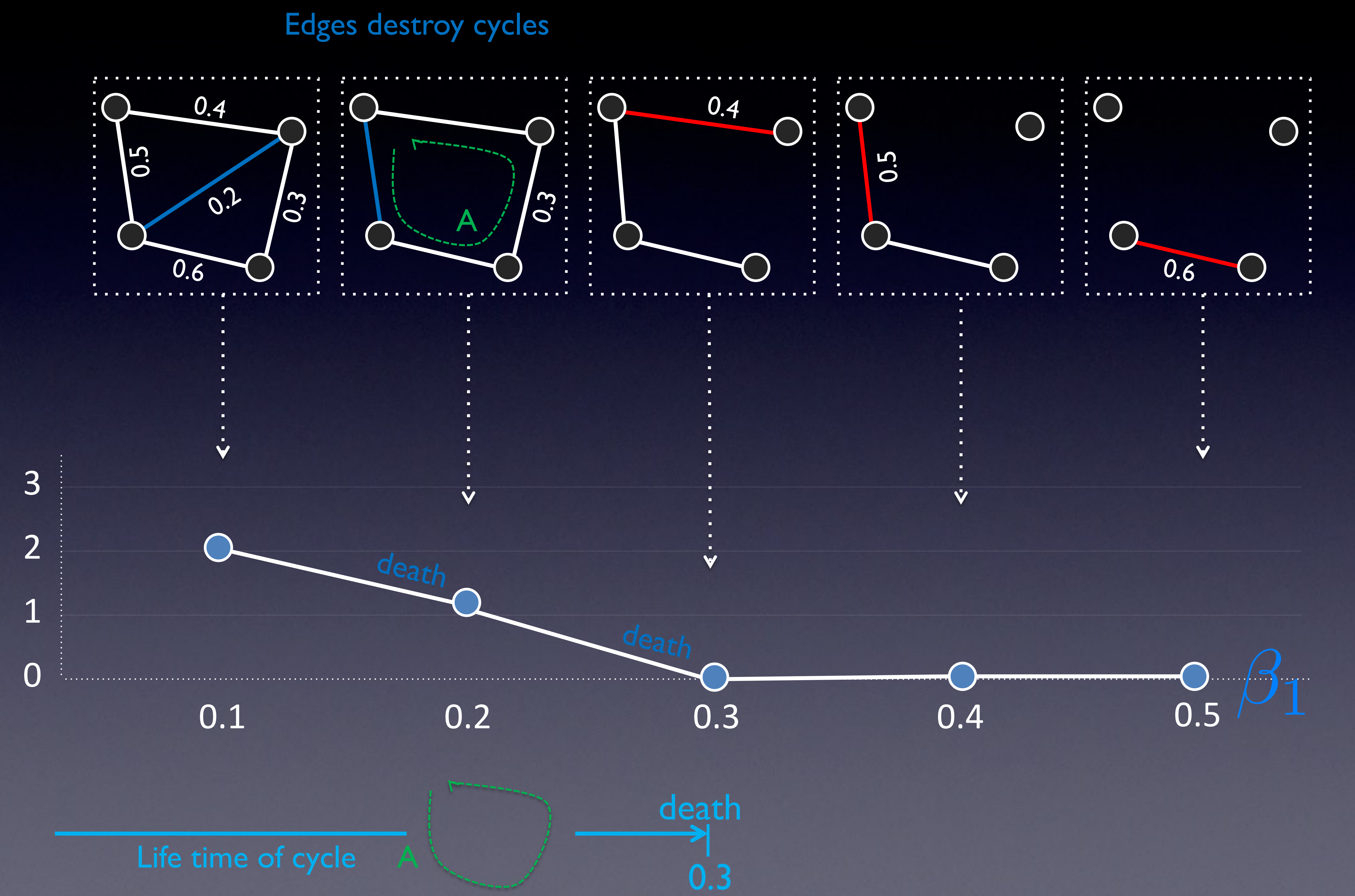
0D Persistence = Life time (death – birth) of 0-cycle



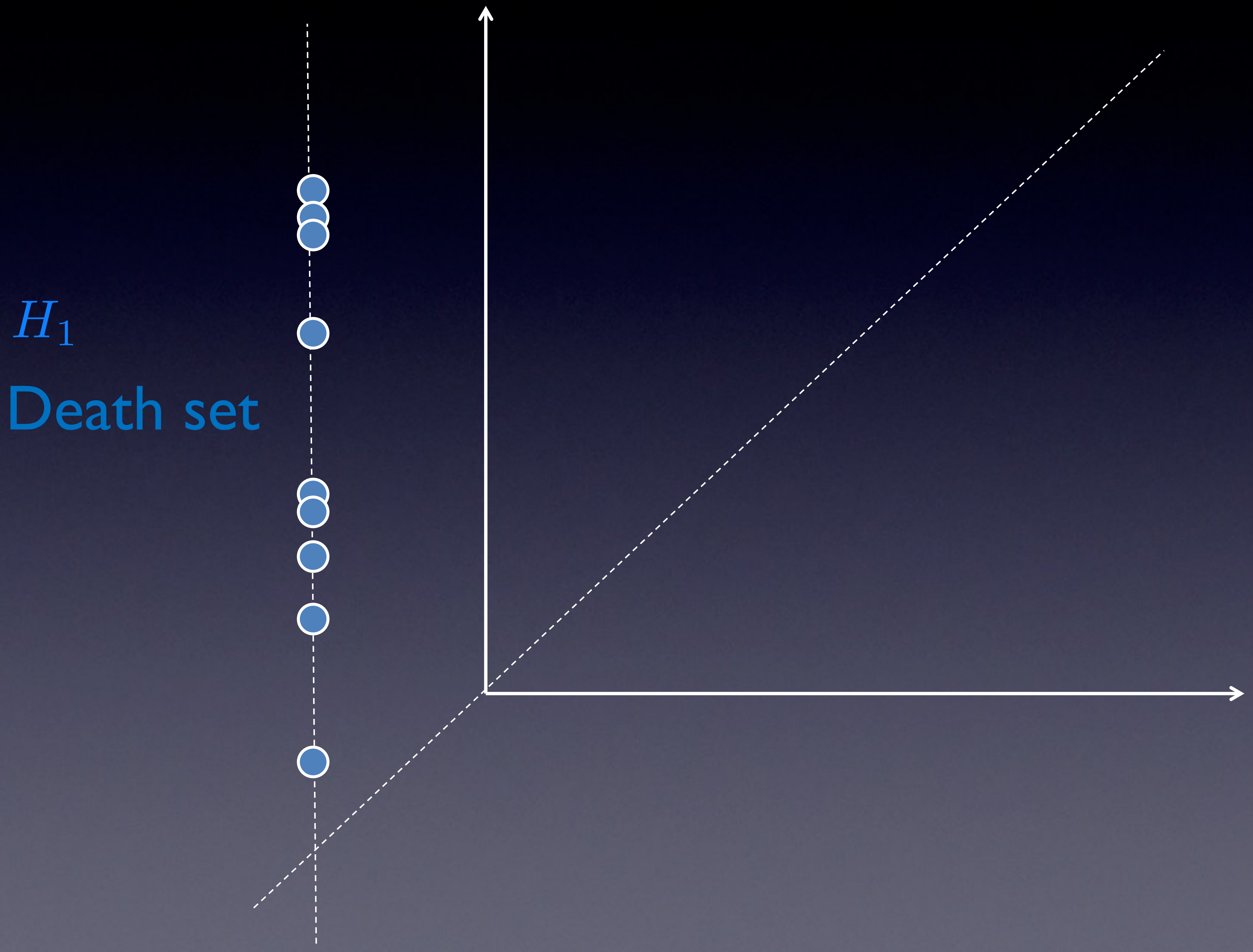
Persistence diagram for graph filtrations



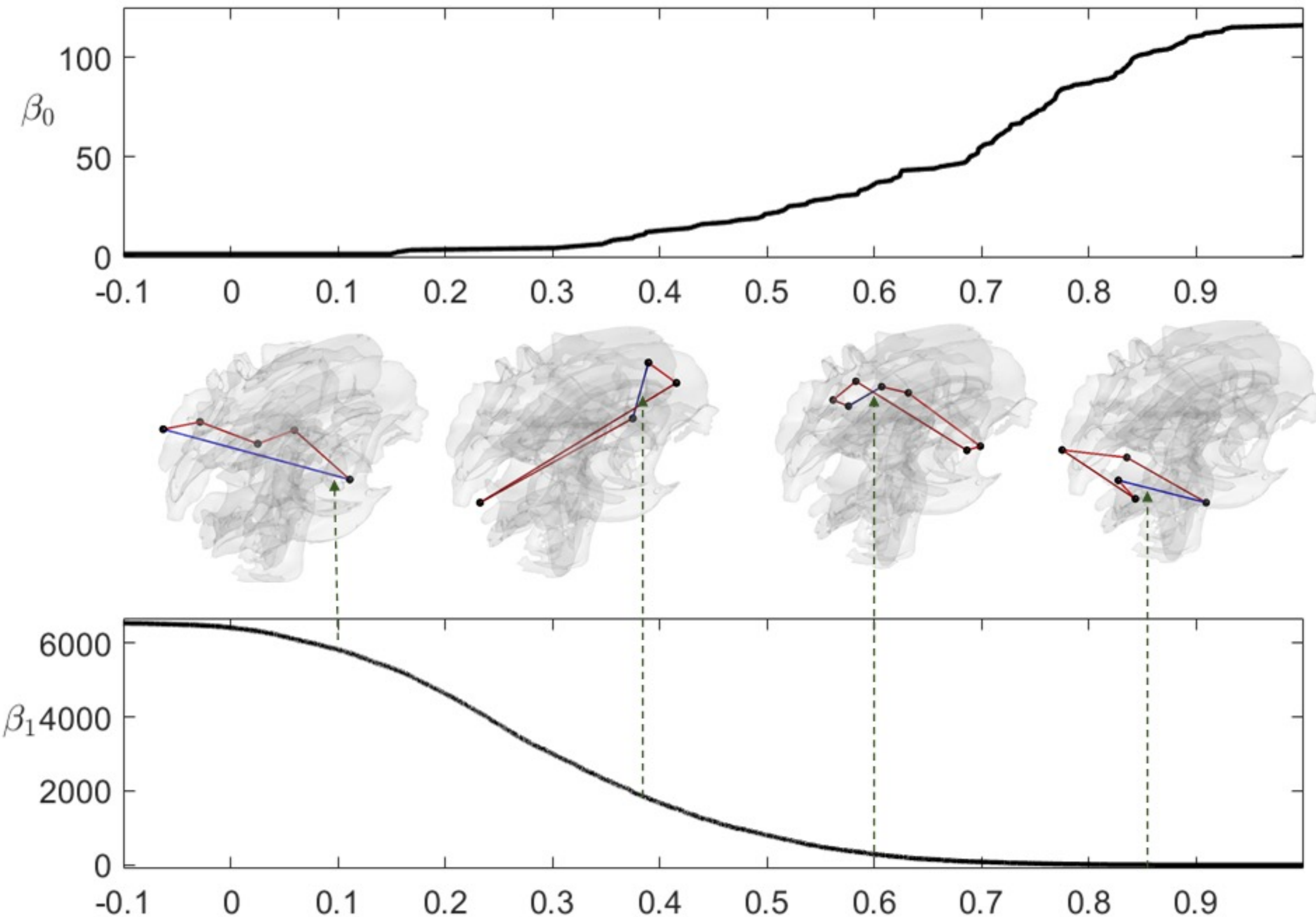
1D Persistence = Life time (death – birth) of 1-cycle



Persistence diagram for graph filtrations

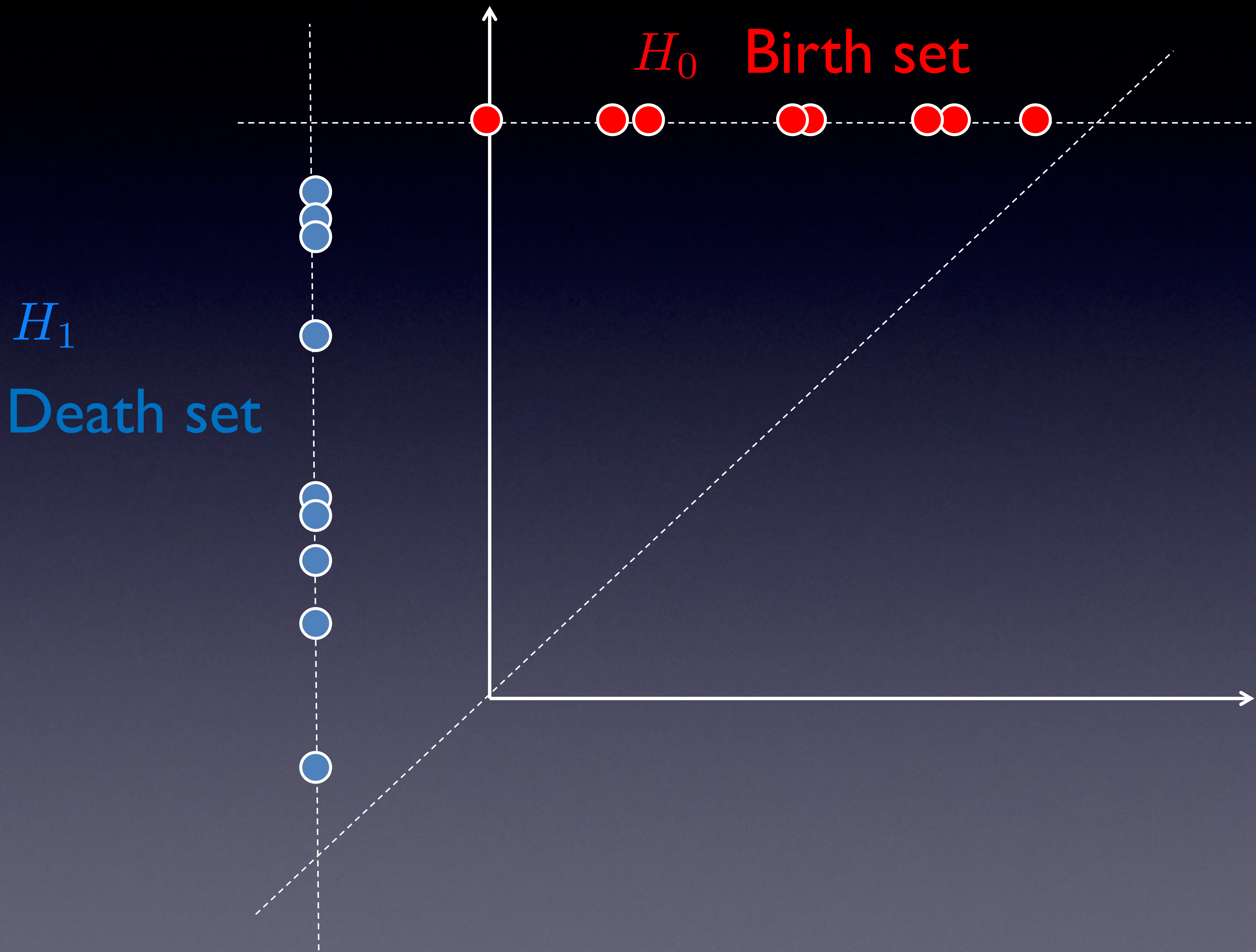


Betti curves for average network of 400 subjects



Anand et al. 2021, arXiv:2110.14599 Review in IEEE Transactions on Medical Imaging

Persistence diagram for graph filtrations



Theorem: *Birth & death decomposition*

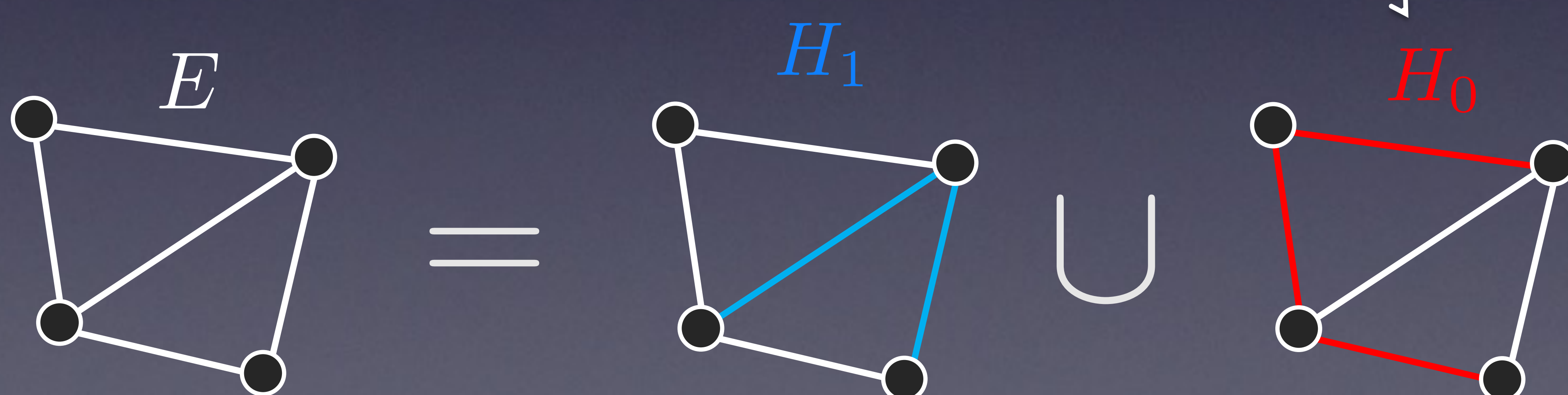
H_1 Edges destroy cycles

H_0 Edges create components

$$\#(H_1) = 1 + \frac{|V|(|V|-3)}{2} \quad + \quad \#(H_0) = |V| - 1$$

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

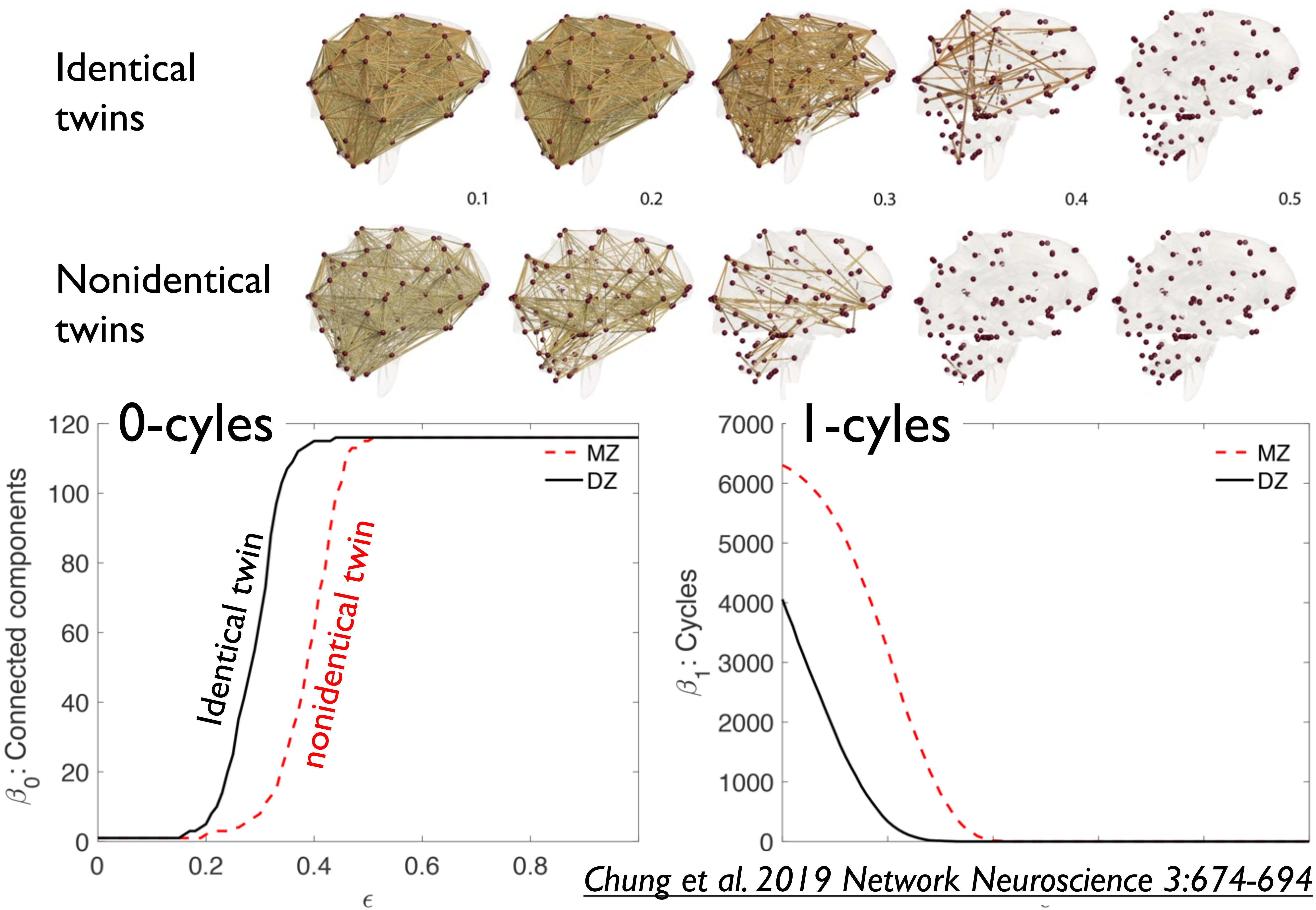
Maximum spanning tree



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

Songdechakraiwut et al. 2021, MICCAI

Genetic difference in brain network in HCP



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

Wasserstein distance on graph filtrations

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

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

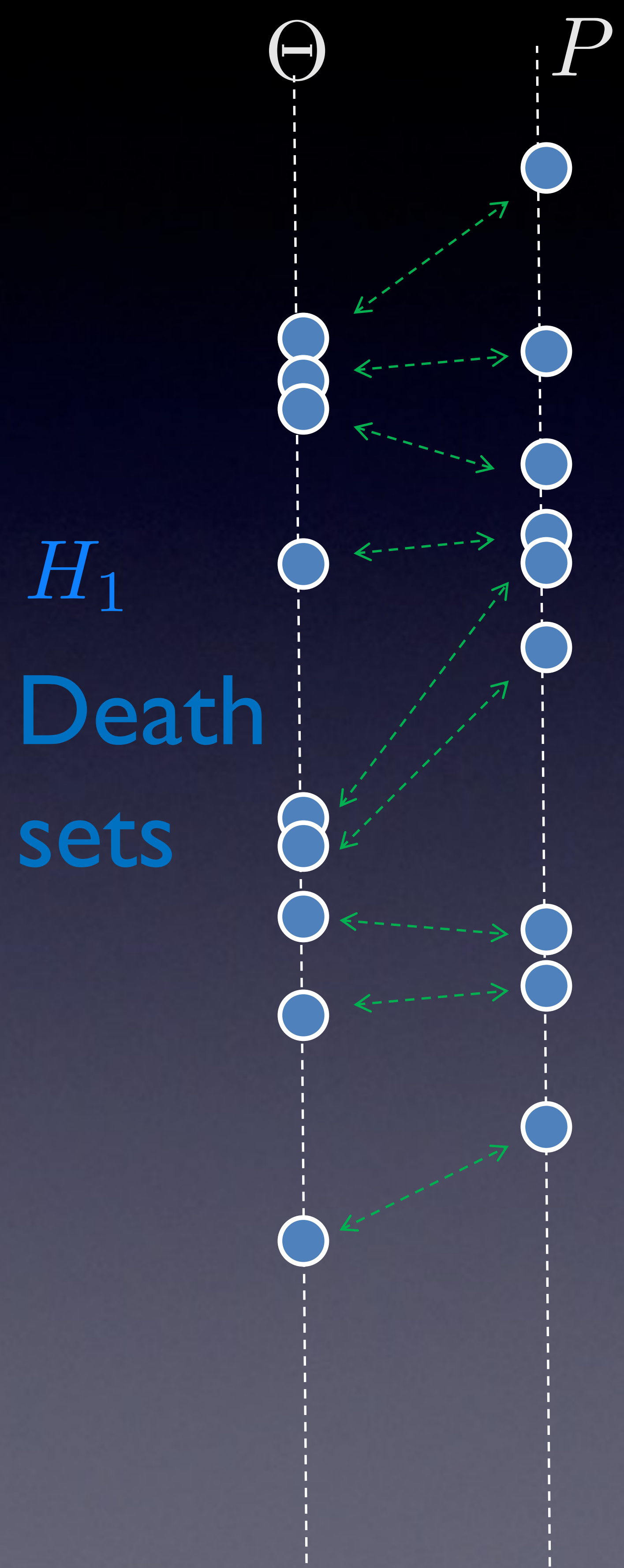
Empirical distributions

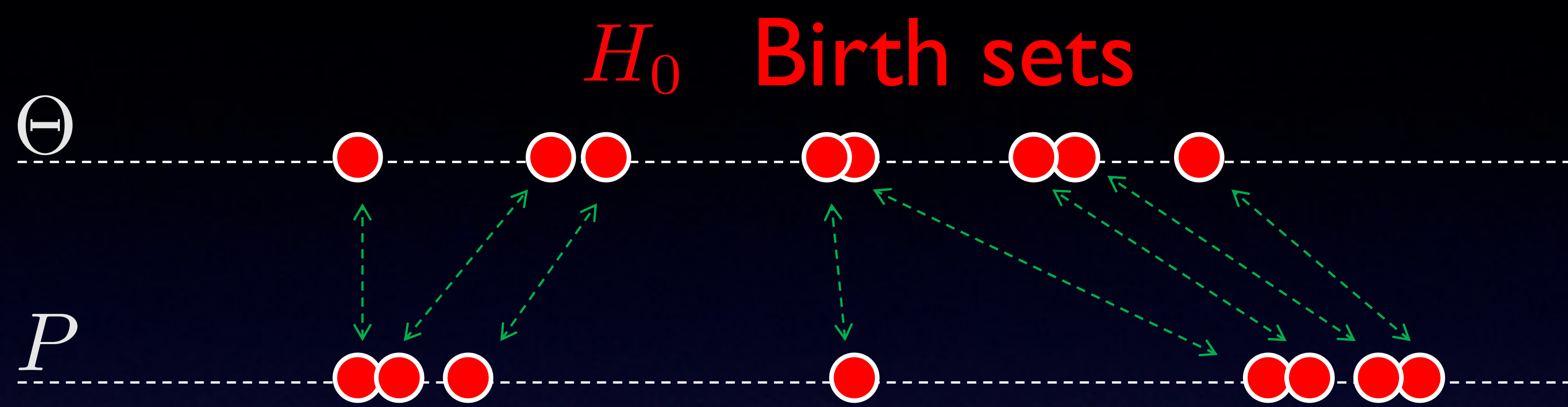
$$f_1(x) = \frac{1}{q}\sum_{i=1}^{q} \delta(x - x_i) \qquad f_2(y) = \frac{1}{q}\sum_{i=1}^{q} \delta(y - y_i)$$

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

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

Wasserstein distance for graph filtrations





1D assignment problem
has known analytic solution

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

Theorem: Wasserstein distance on graph filtrations

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

Birth set

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

τ_0^* *:The i-th smallest <u>birth value</u> to the i-th smallest <u>birth value</u>*

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

Death set

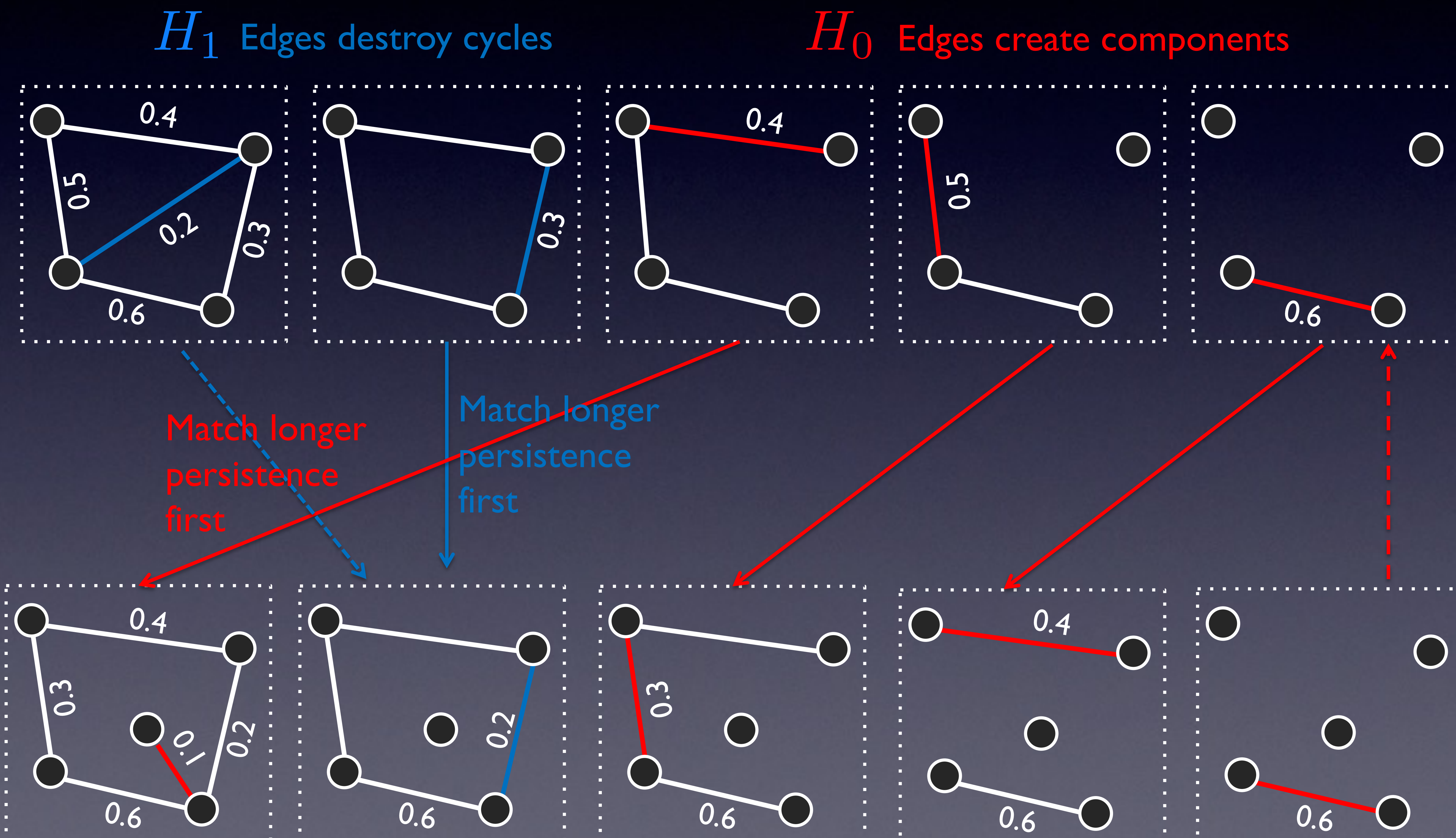
$$= \sum_{d \in E_1} \left[d - \tau_1^*(d) \right]^2$$

τ_1^* *:The i-th smallest <u>death value</u> to the i-th smallest <u>death value</u>*

<u>Songdechakraiwut et al. 2021, MICCAI</u>

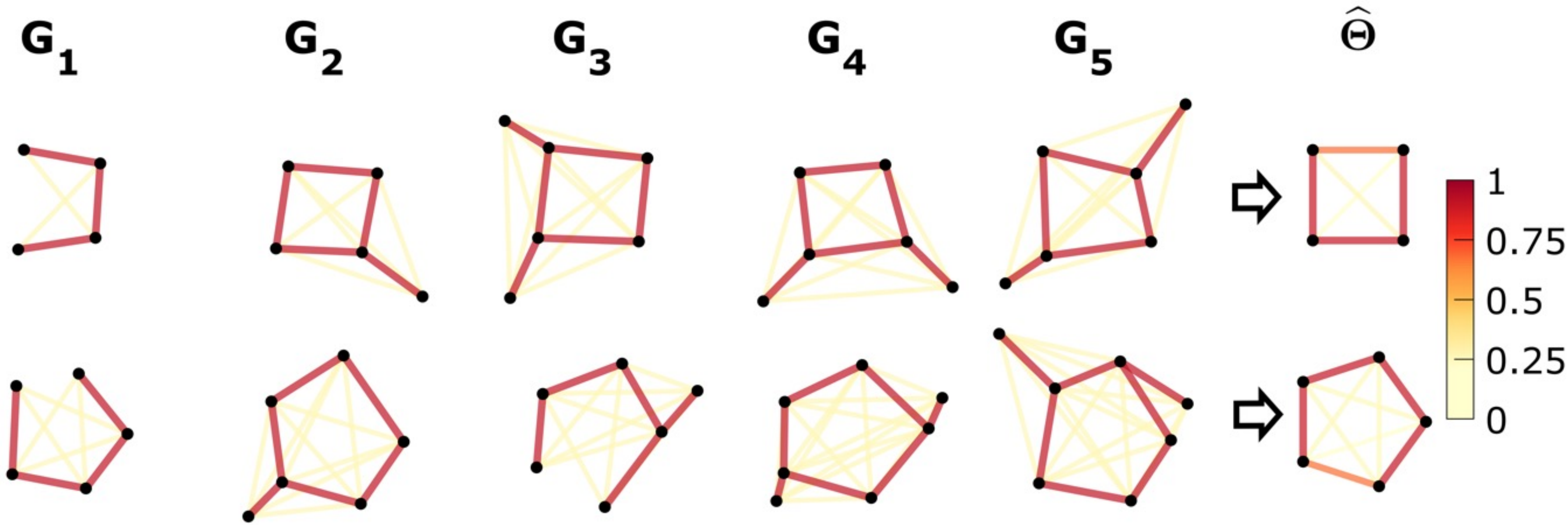
Graph matching by minimizing Wasserstein distances

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



Graph matching → Topological mean of graphs

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



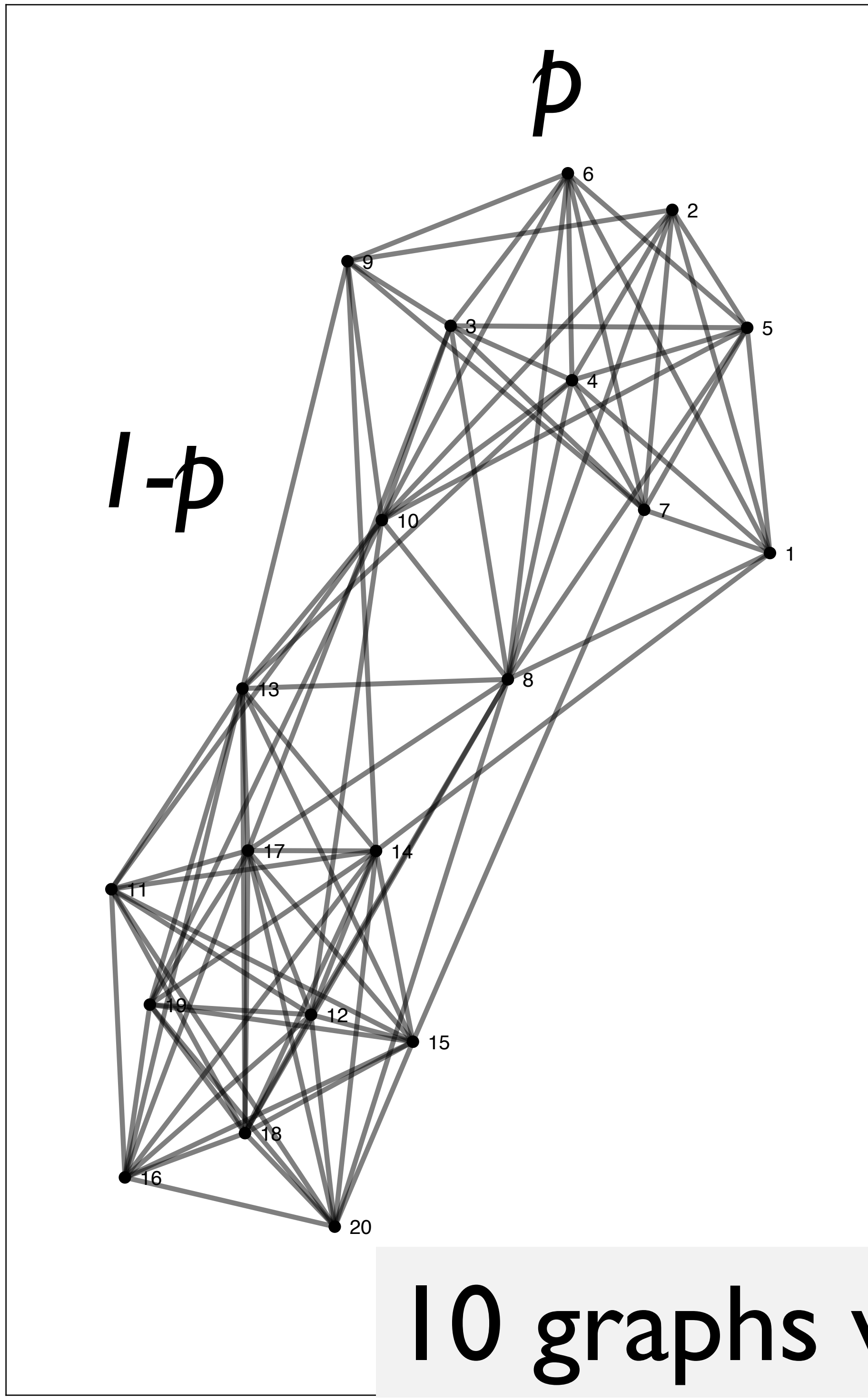
Topological averaging: $G_1 + G_2 + \cdots + G_n$

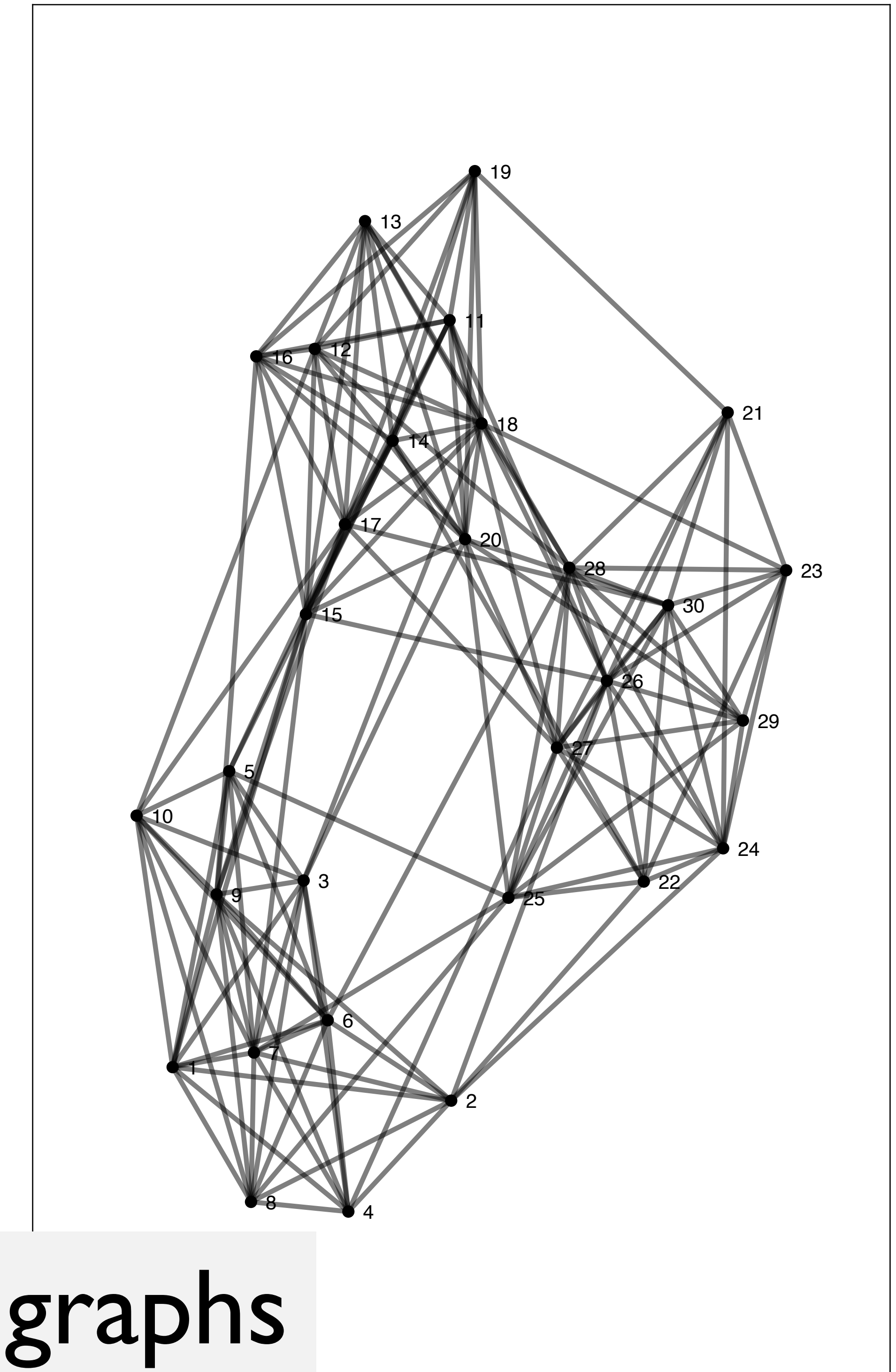
Songdechakraiwut et al. 2022 Annals of Applied Statistics arXiv: 2012.00675

Validation

Random graph model

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

Ratio statistic for Wasserstein distances

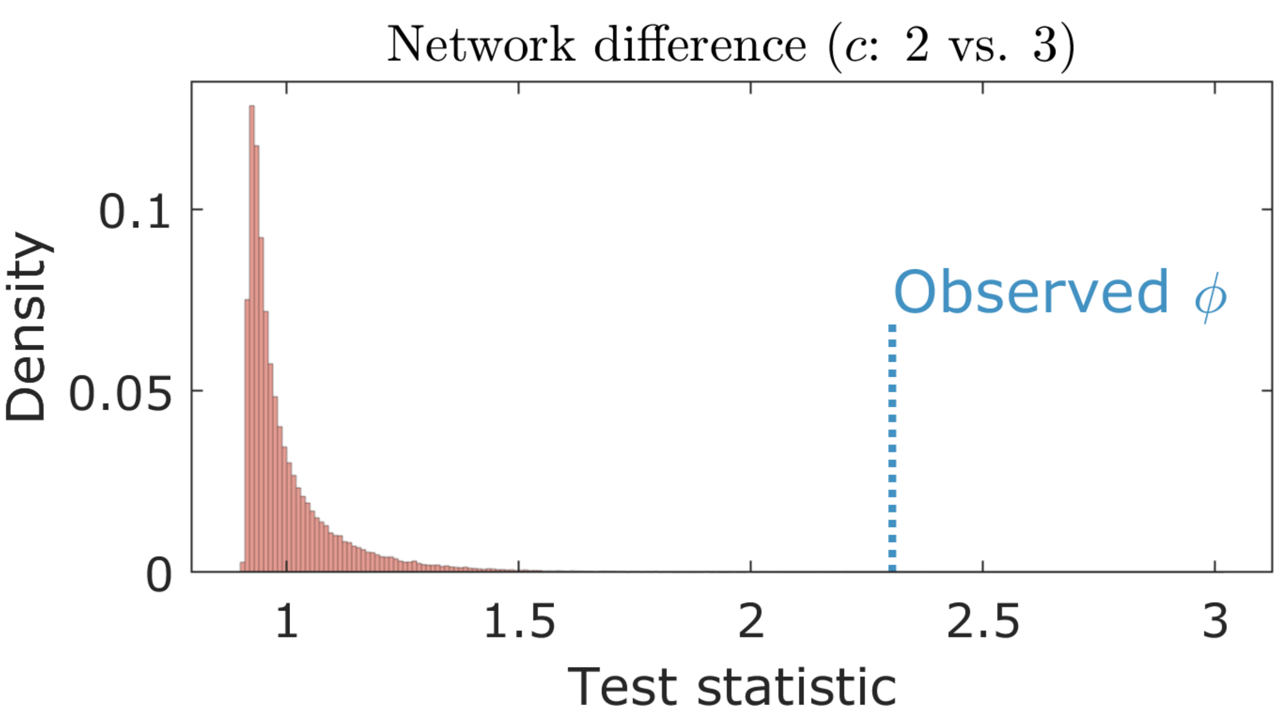
Between-group distance

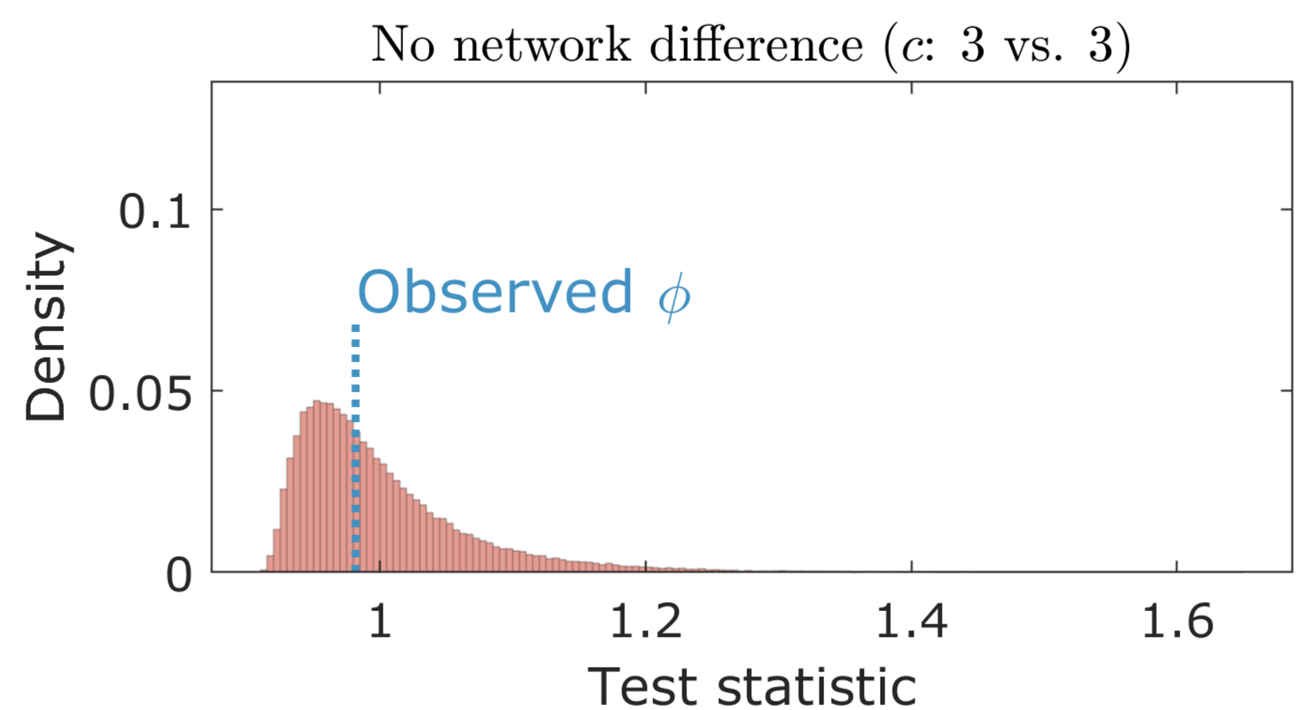
$$l_B \propto \sum_{i \in C_1, j \in C_2} \mathcal{L}(G_i, G_j)$$

Within-group distance

$$l_W \propto \sum_{k} \sum_{i,j \in C_k} \mathcal{L}(G_i, G_j)$$

Statistic $\phi = \frac{l_B}{l_W}$

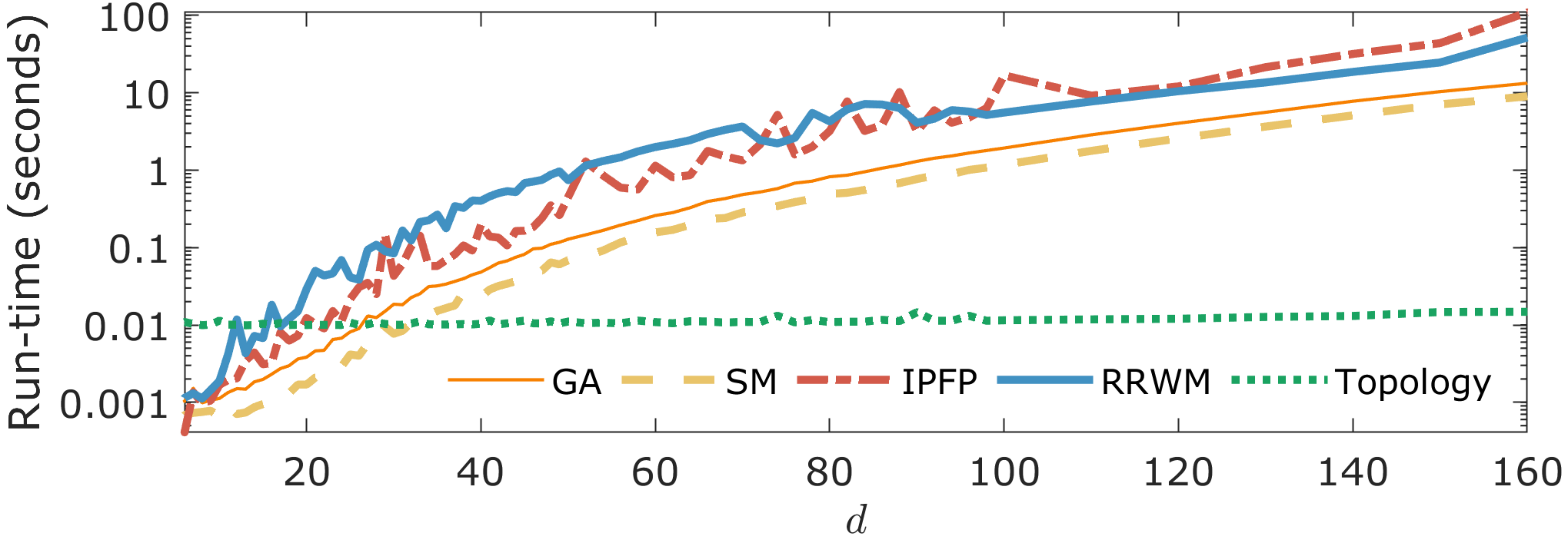




Performance: Average p-value in 50 simulations

nodes	modules	p	Graduated assignment GA	Spectral matching SM	Reweighted random walk matching RRWM	Integer projected fixed point IPFP	$\mathcal{L}_{top}$
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

Fastest possible graph matching algorithm



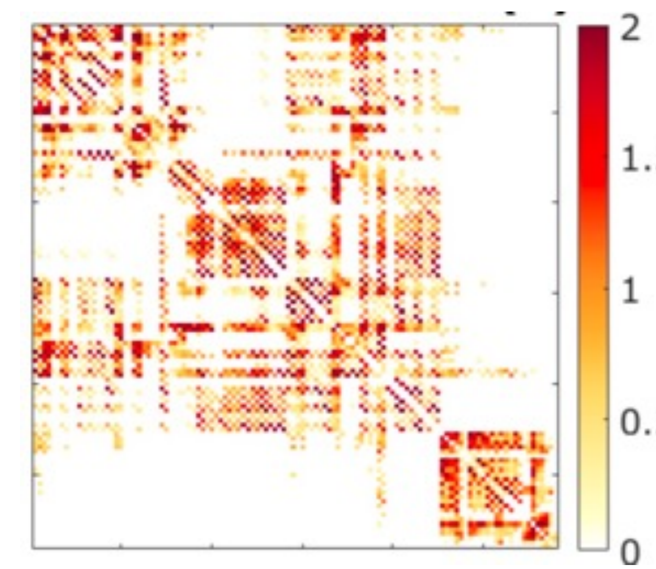
Graduated assignment (GA)
Spectral matching (SM)
Integer projected fixed point method (IPFP)
Re-weighted random walk matching (RRWM)

Topological Regression

Subject-level learning

Structural network template

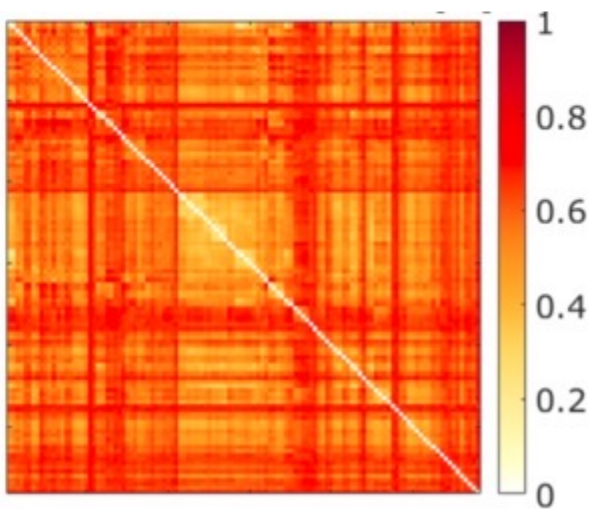
P



Sparse trees

Functional network of subject *k*

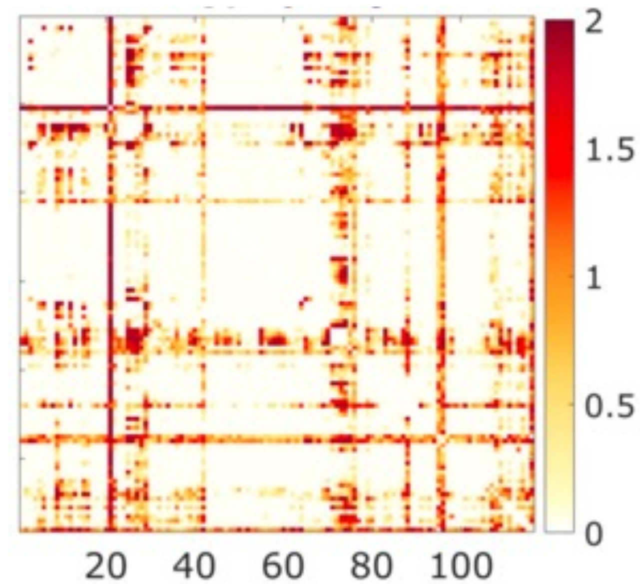
$G_k = (V, w^k)$



Dense cycles

Graph matching

Estimated model



$$\widehat{\Theta}_k = arg \min_{\Theta} \mathcal{L}_F(\Theta, G_k) + \lambda_k \mathcal{L}(\Theta, P)$$

Frobenius norm
Goodness-of-fit

Control amount
of topology

Topological loss

Topological gradient descent

$$\widehat{\Theta}_k = arg\min_{\Theta} \mathcal{L}_F(\Theta, G_k) + \lambda_k \mathcal{L}_{top}(\Theta, P)$$

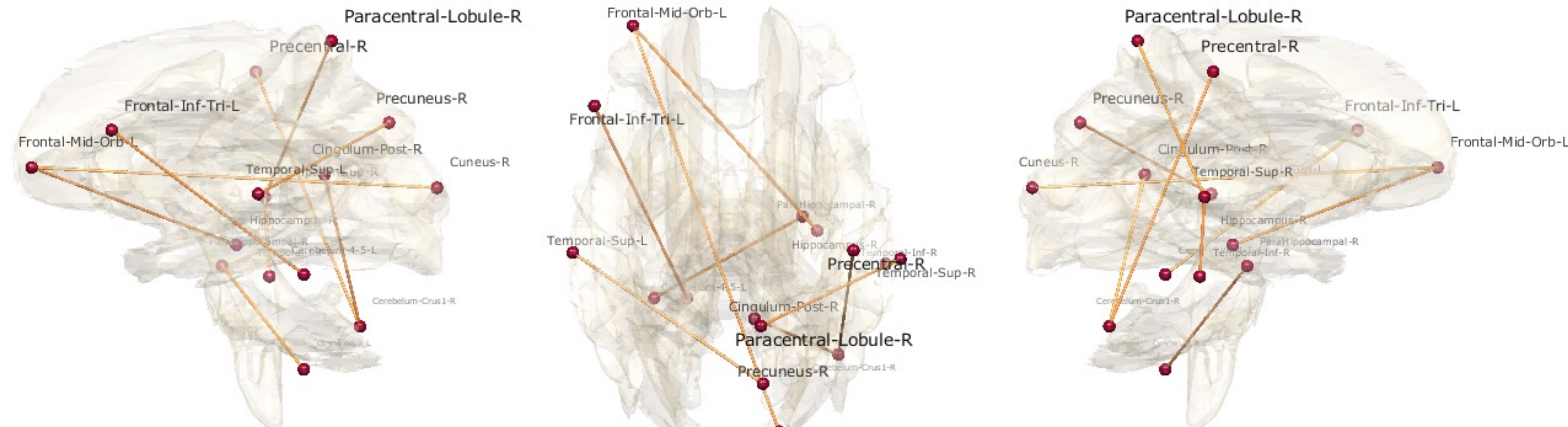
<u>Topological gradient</u> by matching sorted birth and death values

$$\frac{\partial \mathcal{L}_{top}(\Theta, P)}{\partial w_{ij}^{\Theta}} = \begin{cases} 2\big[w_{ij}^{\Theta} - \tau_{0*}(w_{ij}^{\Theta})\big] & \text{if } w_{ij}^{\Theta} \in E_0; \\ 2\big[w_{ij}^{\Theta} - \tau_{1*}(w_{ij}^{\Theta})\big] & \text{if } w_{ij}^{\Theta} \in E_1 \end{cases}$$

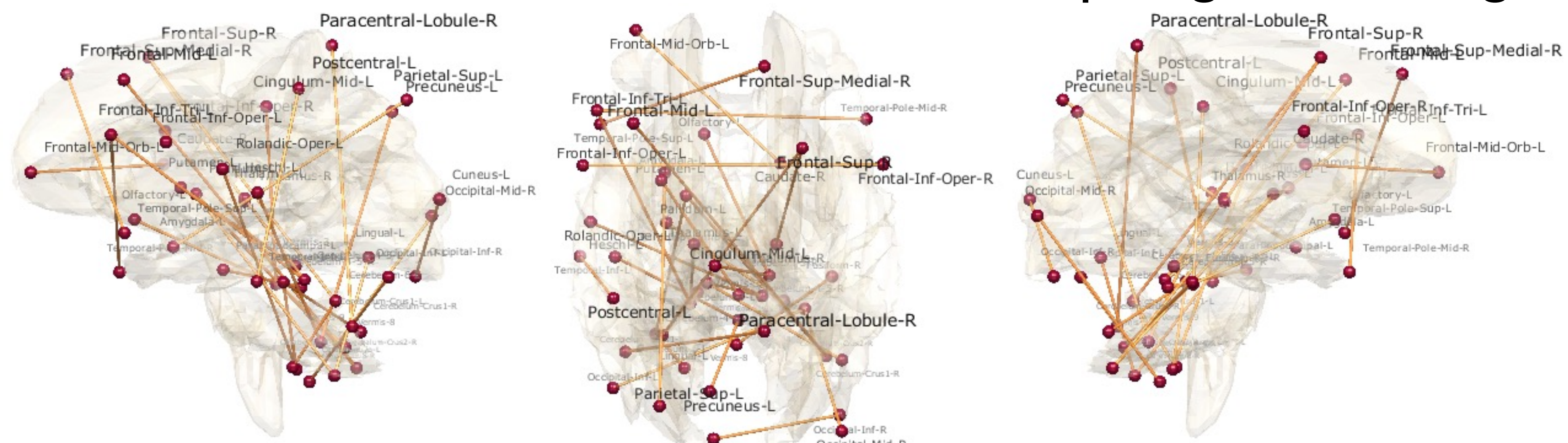
Run time $O(|E| \log |V|)$

Heritability index above 1.00 through ACE-model

Original Pearson correlation



After topological learning



Songdechakraiwut et al. 2021, MICCAI 116-176

Topological clustering

Wasserstein graph clustering

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

L2-norm on sorted birth values

L2-norm on sorted death values

$$\cup_{i=1}^{k} C_i = \{G_1, \cdots, G_n\}, \quad C_i \cap C_j = \emptyset$$

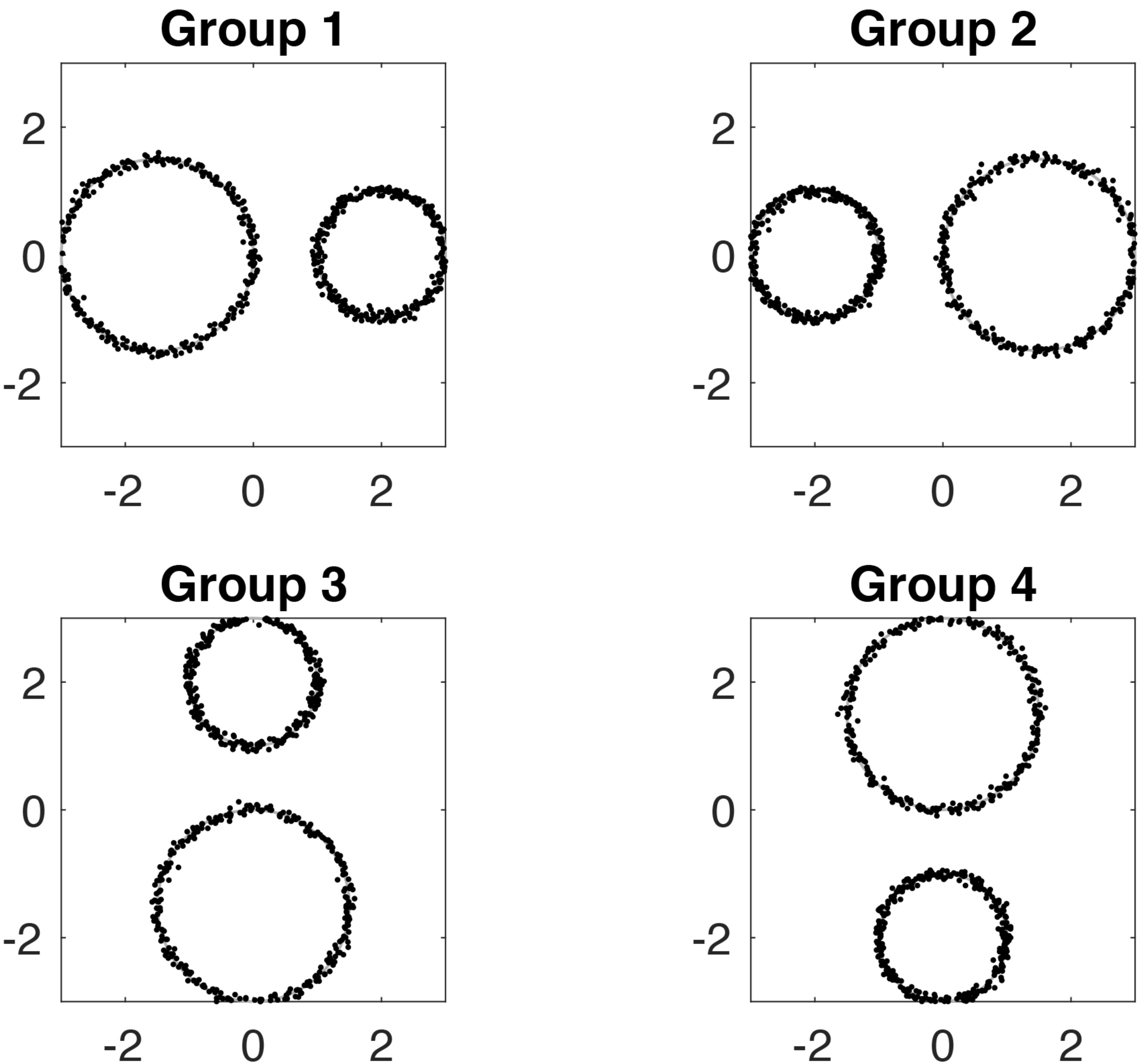
$$l_W(C; \mu) = \sum_{j=1}^{k} \sum_{X \in C_j} \mathcal{L}_{top}(X, \mu_j)$$

$$\mu_j = \frac{1}{|C_j|} \sum_{X \in C_j} X$$

Proof of convergence

Chung et al. 2022 arXiv:2201.00087

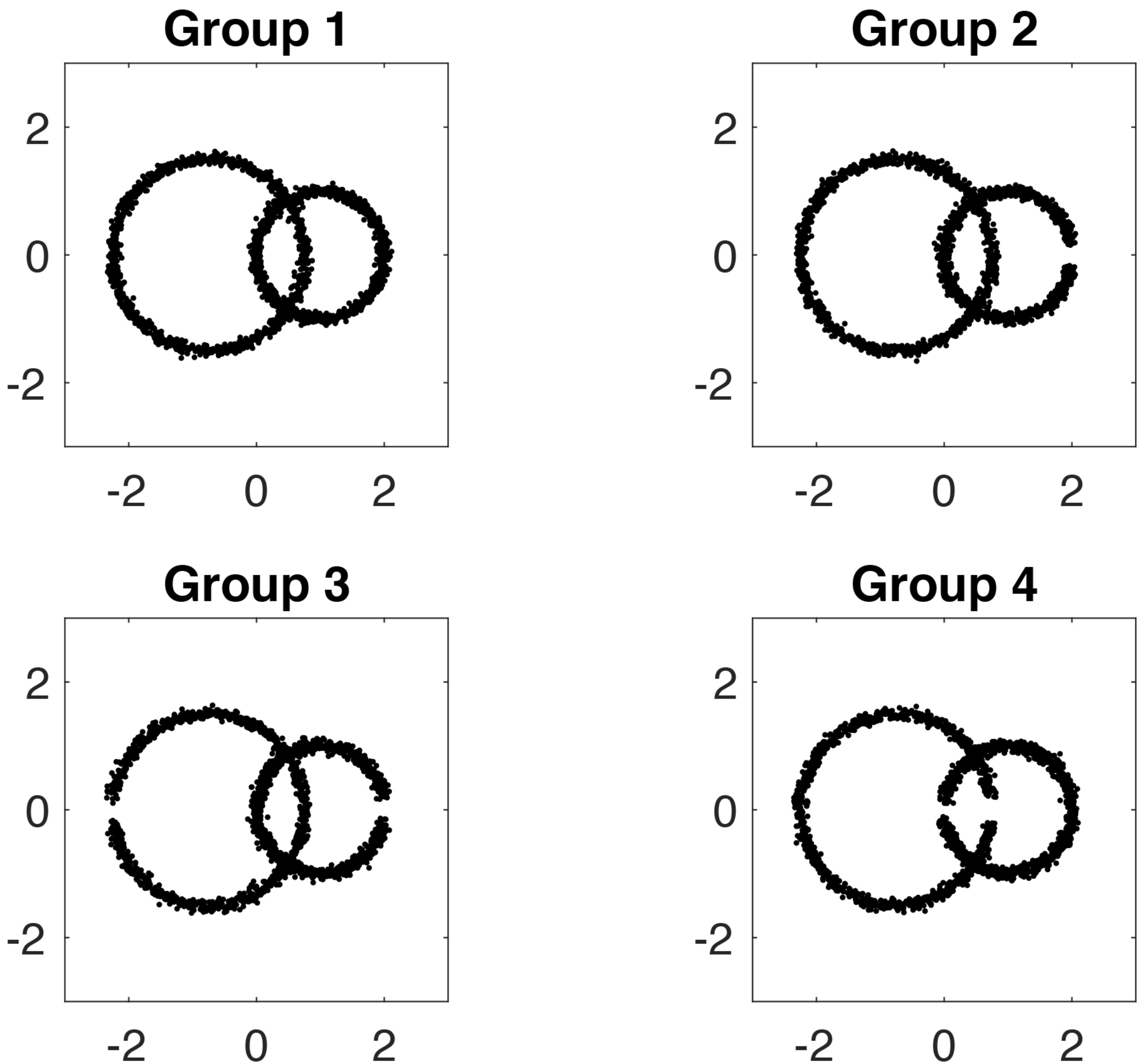
No topological difference



K-means clustering
1.00 +- 0.04

Wasserstein Graph clustering
0.53 +- 0.08

Topological difference

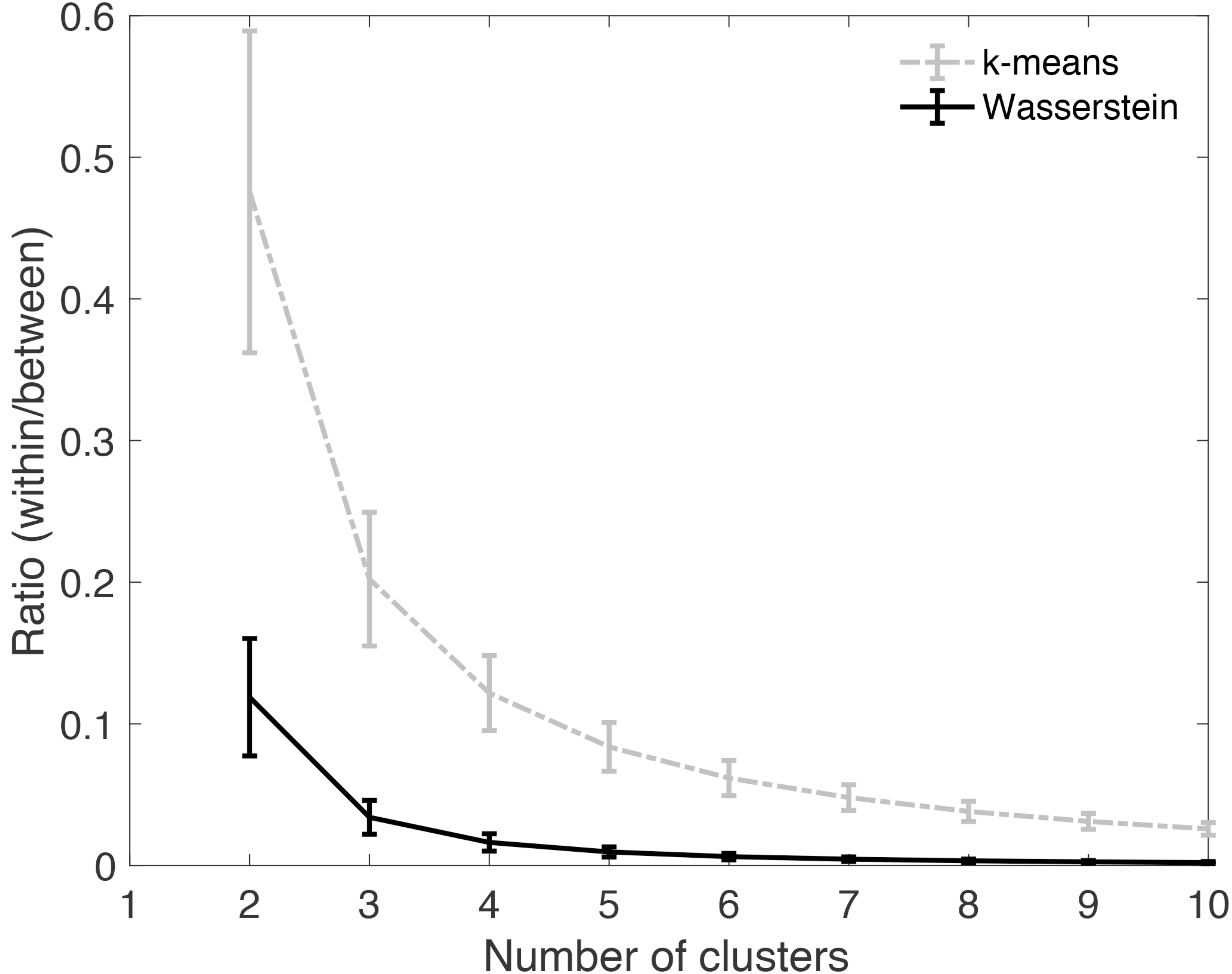


K-means clustering
0.83 +- 0.10

Wasserstein Graph clustering
0.96 +- 0.10

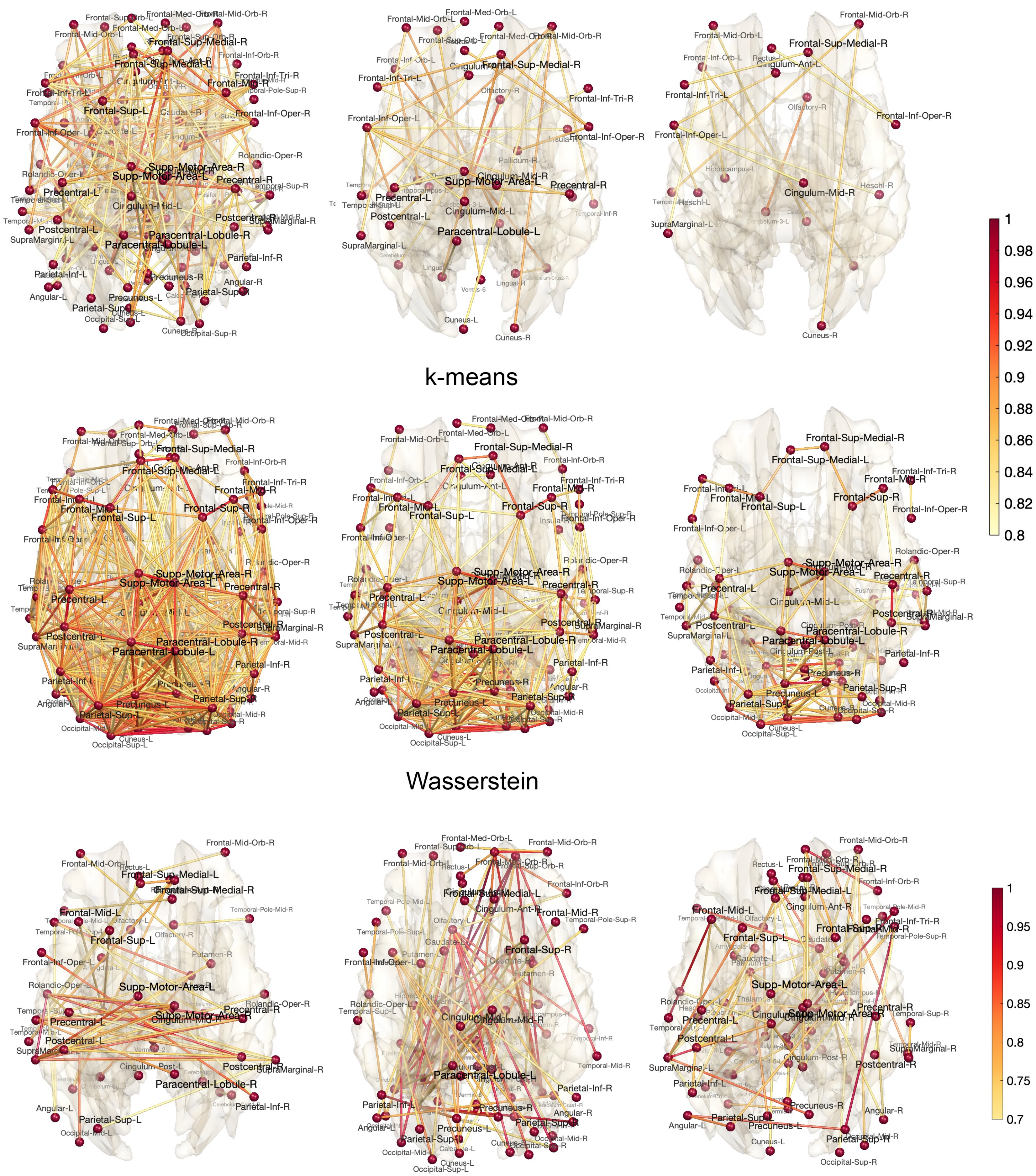
16% improvement

$$\frac{1}{\phi} = \frac{l_W}{l_B}$$

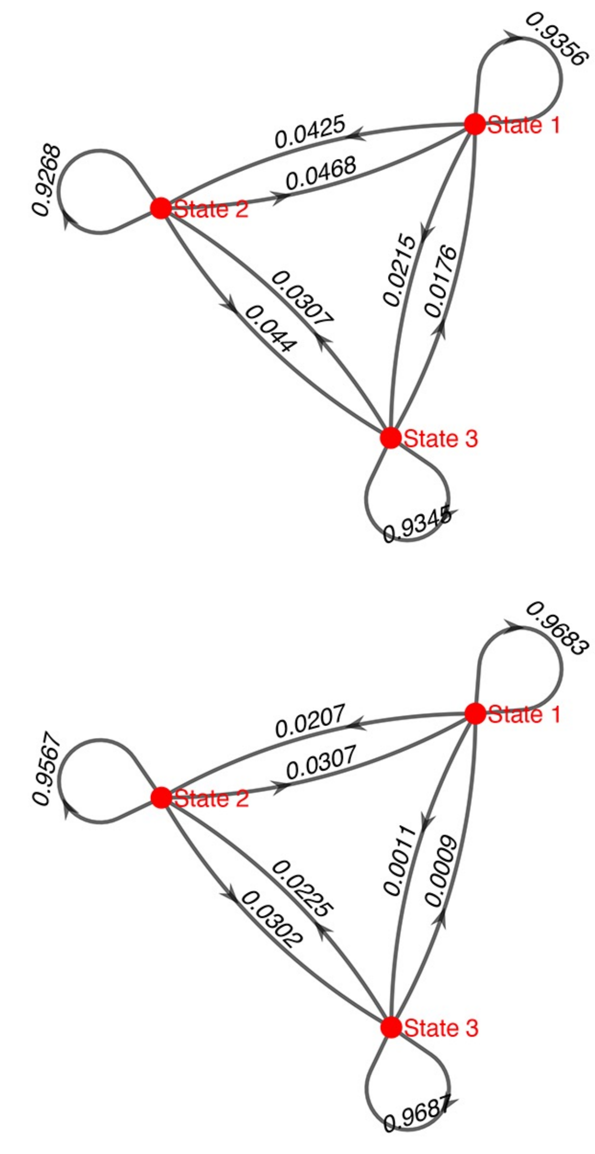


The within cluster variance of Wassertein clustering is *6 times* smaller than k-means clustering.

State 1 State 2 State 3



Heritability Index



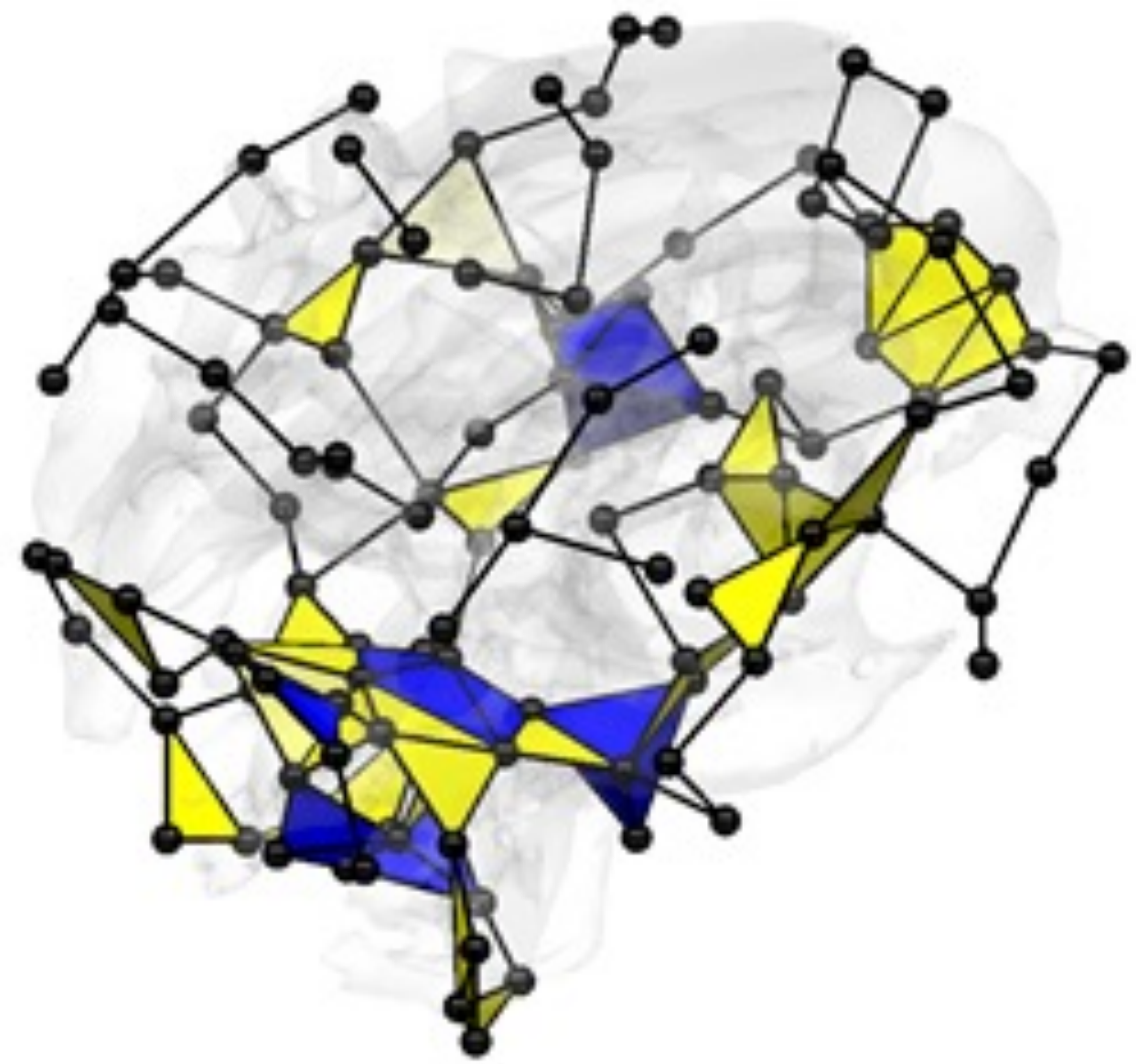
Markov chains

Heritability index
for UW-Madison twin
study – 200 twin pairs

Explicit modeling of 1-cycles: higher order connectivity

Graph *vs.* simplicial complex

Pairwise interaction



Higher order interaction

Yellow = 3 nodes (2 simplices)

Blue = 4 nodes (3 simplices)

Anand et al. 2021, arXiv:2110.14599

k-th Hodge Laplacian

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

For graph,

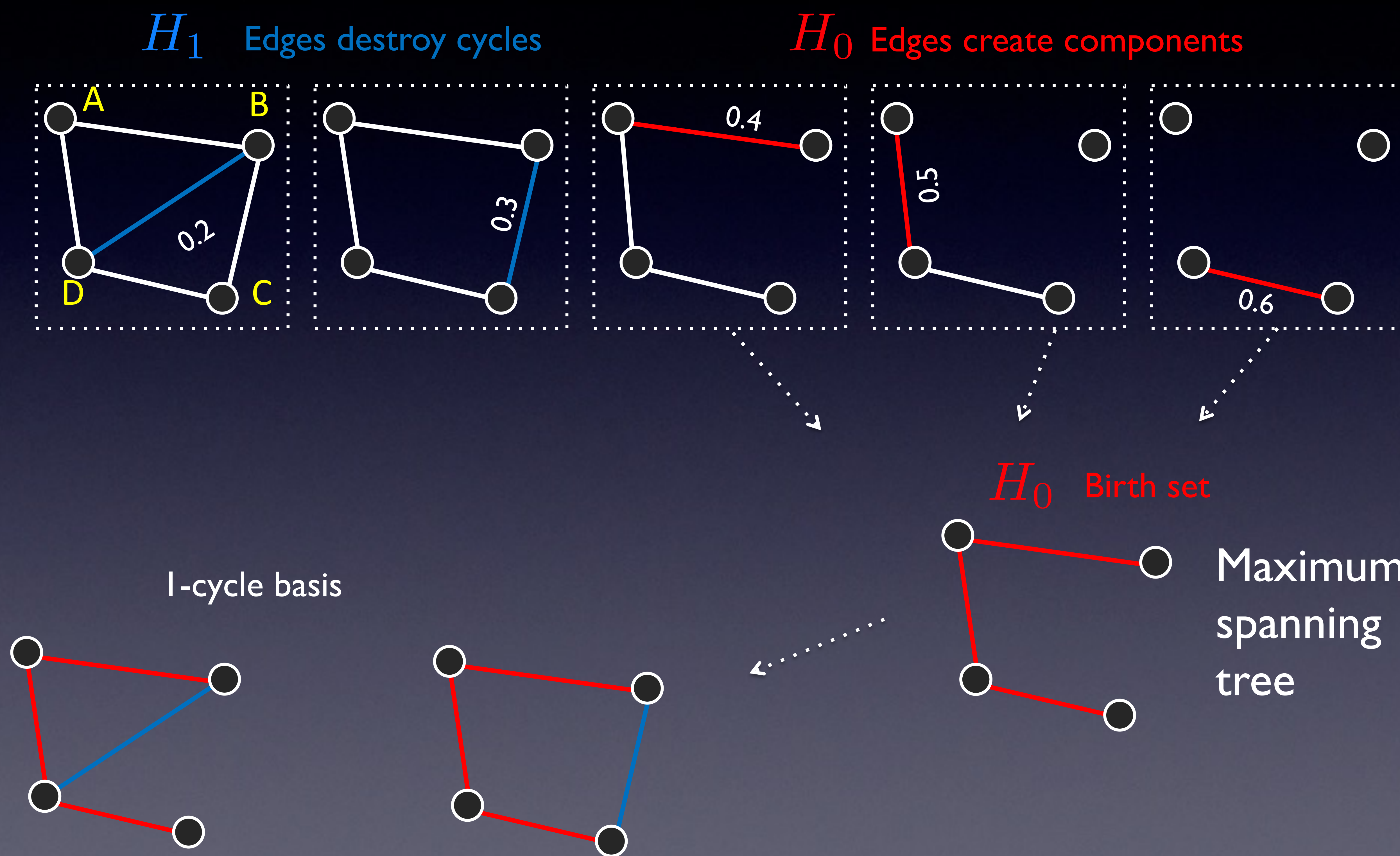
$$\Delta_0 = \partial_1 \partial_1^\top$$

$ker \Delta_0$ is spanned by the eigenvectors of zero eigenvalues

# of zero eigenvalues
= # of k-cycles

Theoren (Identification of 1-cycle basis)

Anand et al. 2021, arXiv:2110.14599



Compute the eigenvector of zero eigenvalue of the 1st Hodge Laplacian

Basis expansion using 1-cycle basis

Anand et al. 2021, arXiv:2110.14599

1-cycle basis: $\mathcal{C}^1, \cdots, \mathcal{C}^Q$

vector of 0's and 1's

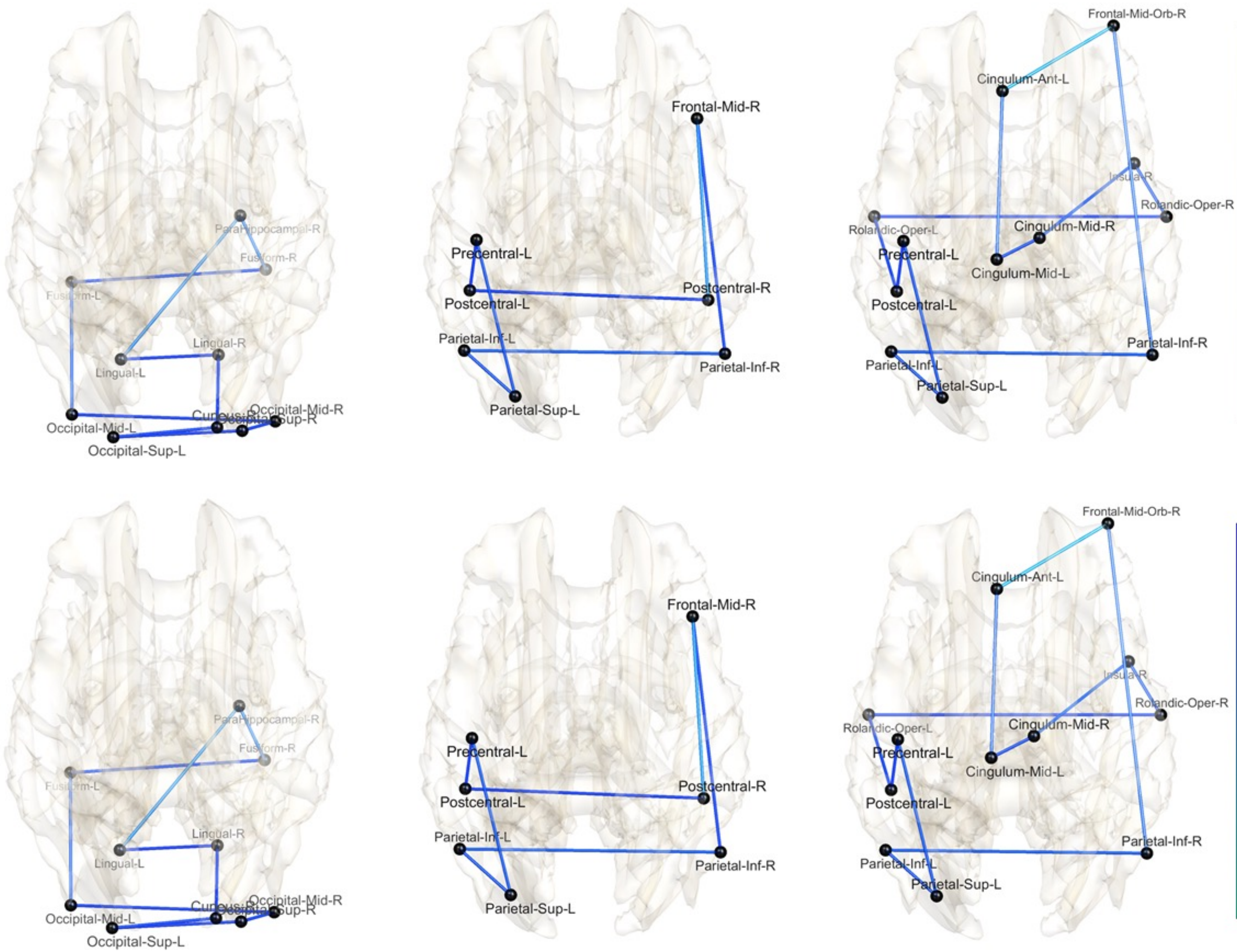
Correlation matrix = $\displaystyle\sum_{j=1}^{Q} \alpha_j \mathcal{C}^j$

Least squares estimation

Test statistic:

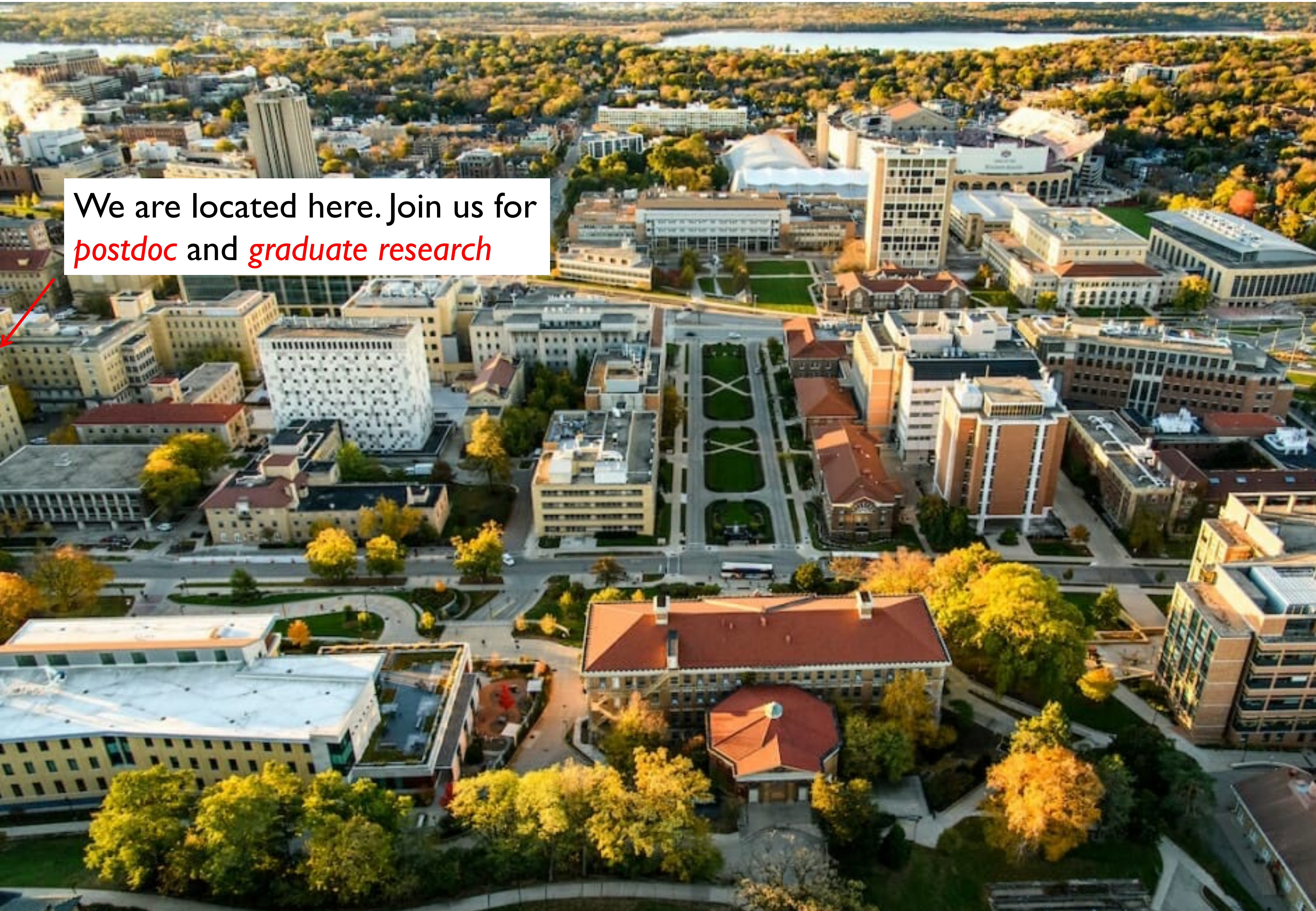
$$T(Group1, Group2) = \max_{1 \le j \le Q} |\bar{\alpha}_j^{Group1} - \bar{\alpha}_j^{Group2}|$$

Three biggest cycle differences in male vs. female in HCP



p-value = 0.007

Thank you.


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