The Waisman Laboratory
for Brain Imaging and Behavior

University of Wisconsin
SCHOOL OF MEDICINE AND PUBLIC HEALTH

Random Walks in the Permutation Group

Moo K. Chung

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

May 18, 2019, KSIAM

Abstract

The permutation test is an often used test procedure in brain imaging. Unfortunately, generating every possible permutation for large-scale brain image datasets such as HCP and ADNI with thousands images is not practical. Many previous attempts at speeding up the permutation test rely on various approximation strategies such as estimating the tail distribution with known parametric distributions. In this study, we show how to rapidly accelerate the permutation test without any type of approximate strategies by exploiting the underlying topological structure of the permutation group via random walks. The method is applied to large number of magnetic resonance images in two applications: (1) localizing the male and female differences and (2) localizing the regions of high genetic heritability in the sulcal and gyral graph patterns of the human cortical brain. The talk is based on arXiv:1812.06696.

Acknowledgement

Yixian Wang, Shih-Gu Huang, Andrey Grisenko,
Ross Luo, Nagesh Adluru, Andrew Alexander,
Richard Davidson, Hill Goldsmith
University of Wisconsin-Madison, USA

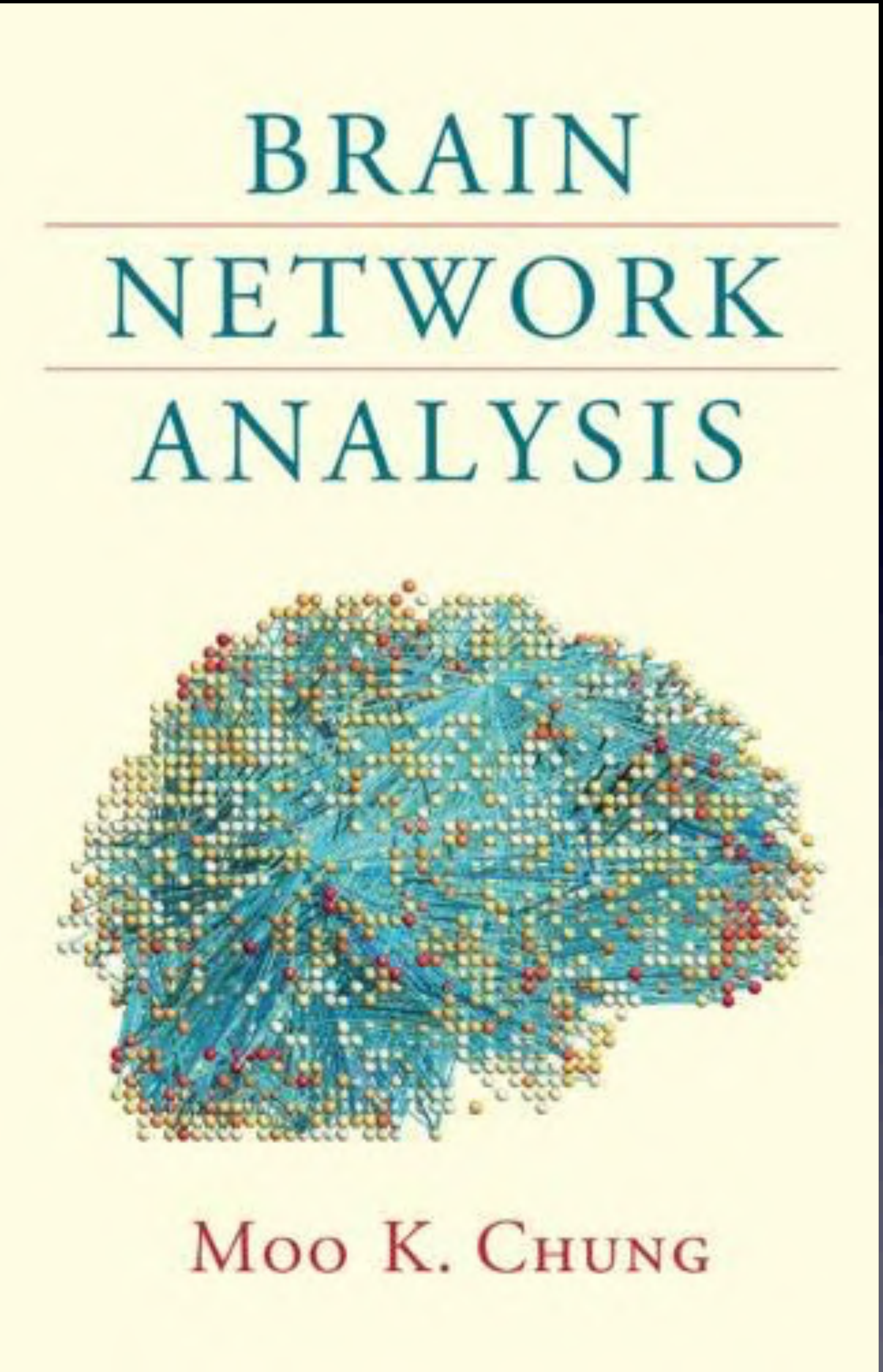
Ilwoo Lyu *Vanderbilt University*
Yuan Wang *University of South Carolina*
Hyekyung Lee *Seoul National University*
Li Shen *University of Pennsylvania*
Martin Lindquist *Johns Hopkins University*

NIH grants: R01 EB022856, R01 MH101504,
P30 HD003352, U54 HD09025

Full day course
Topological and Object Oriented Data Analysis in International Biometric Conference (IBC2020)
COEX Seoul, Korea

Sunday July 5, 2020

OODA: Steve Marron (UNC)
TDA: Yuan Wang (USC)
Computation: Moo K. Chung (UW-Madison)



Date: August 31, 2019

Publisher: Cambridge University Press

One chapter deals with TDA based network analysis

TDA MATLAB Codes, Data & Talk Slides

www.stat.wisc.edu/~mchung/TDA

What is permutation test?

$$\mathbf{x} = (x_1, x_2, \cdots, x_m)$$
$$\mathbf{y} = (y_1, y_2, \cdots, y_n)$$
$$(\mathbf{x}, \mathbf{y}) = (x_1, \cdots, x_m, y_1, \cdots, y_n)$$

$$\pi(\mathbf{x}, \mathbf{y}) \in \mathbb{S}_{m+n}$$

Permutation group of order *m*+*n*

$$p\text{-value} = \frac{1}{(m+n)!} \sum_{\pi \in \mathbb{S}_{m+n}} \mathcal{I}\Big(f(\mathbf{x}, \mathbf{y}) \geq f(\pi(\mathbf{x}, \mathbf{y}))\Big)$$

History of permutation test

Fisher 1935, The Design of Experiment

$$\binom{8}{4} = 70$$

Thompson et al. 2001, Nature Neuroscience

$$\binom{40}{20} = 1.34 \cdot 10^{11}$$

Supercomputer for 1million permutations

Nichols et al. 2002, Human Brain Mapping

4279 citations

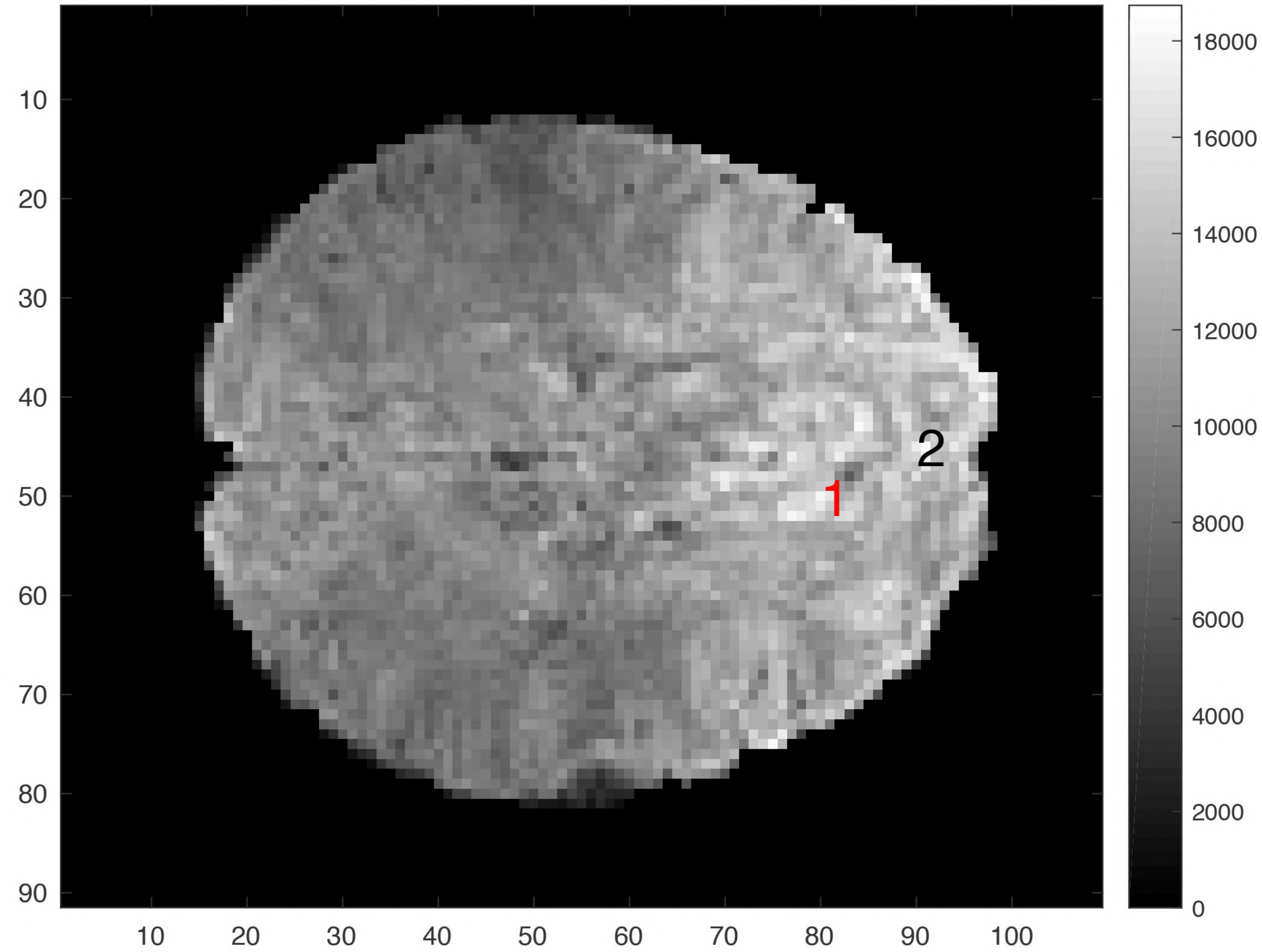
$$\binom{6}{3} = 20$$

Previous method: Exact topological inference

Chung et al. 2017 Information Processing in Medical Imaging

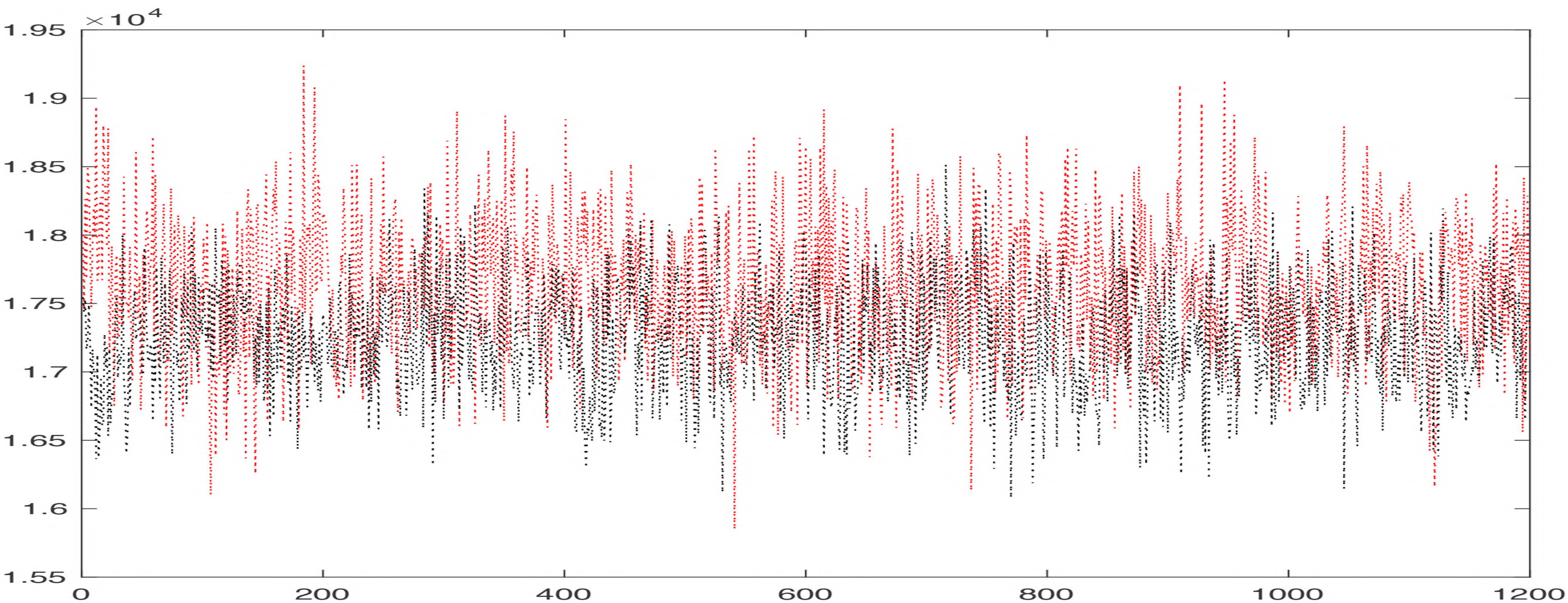
Chung et al. 2018 MICCAI

Chung et al. 2019 Network Neuroscience

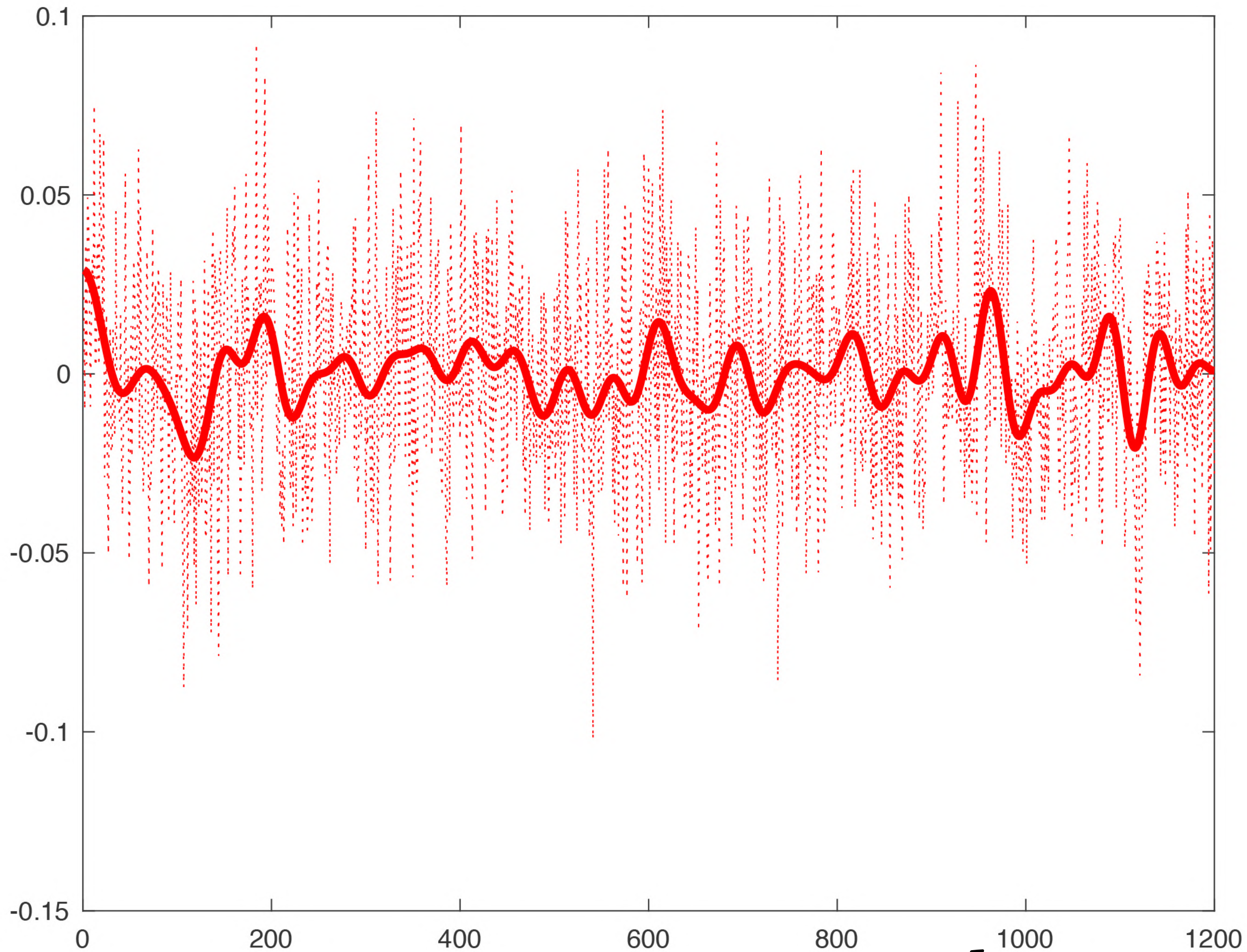


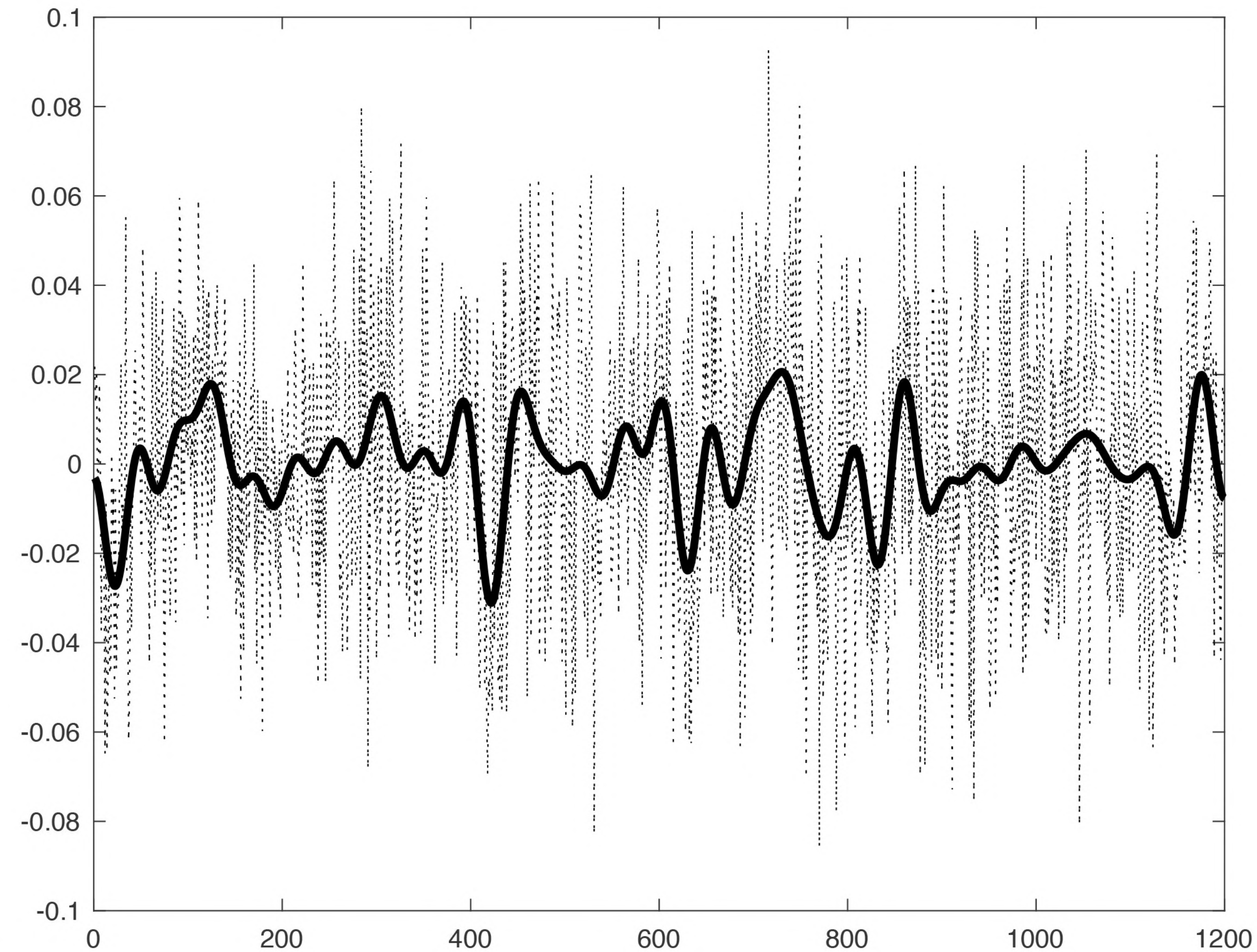
Resting-state functional-MRI at 1200 time measured over 14min 33 seconds.

416 subjects
= 131 Monozygotic (MZ) twins
77 Dizygotic (DZ) twins
= 2GB x 416 = 832GB data



Cosine Series Representation



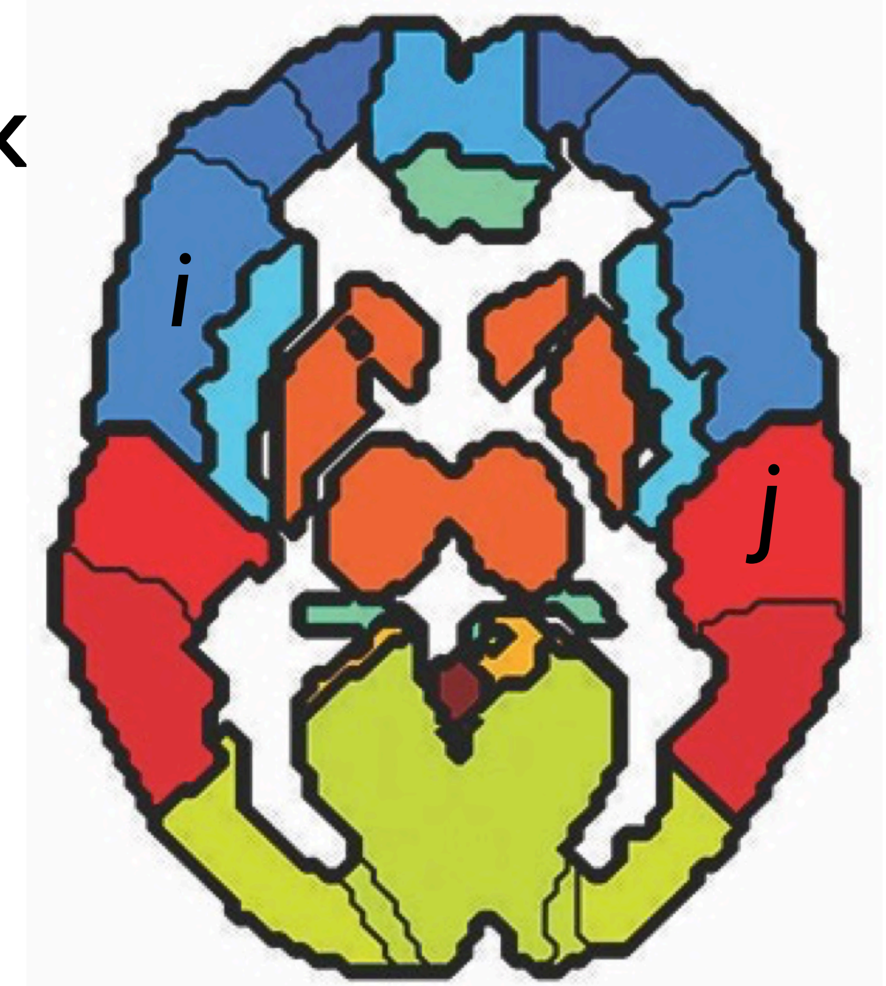


$$\zeta_i(t) = \sum_{l=0}^{k} d_{li}\psi_l(t),\ t \in [0,1]$$

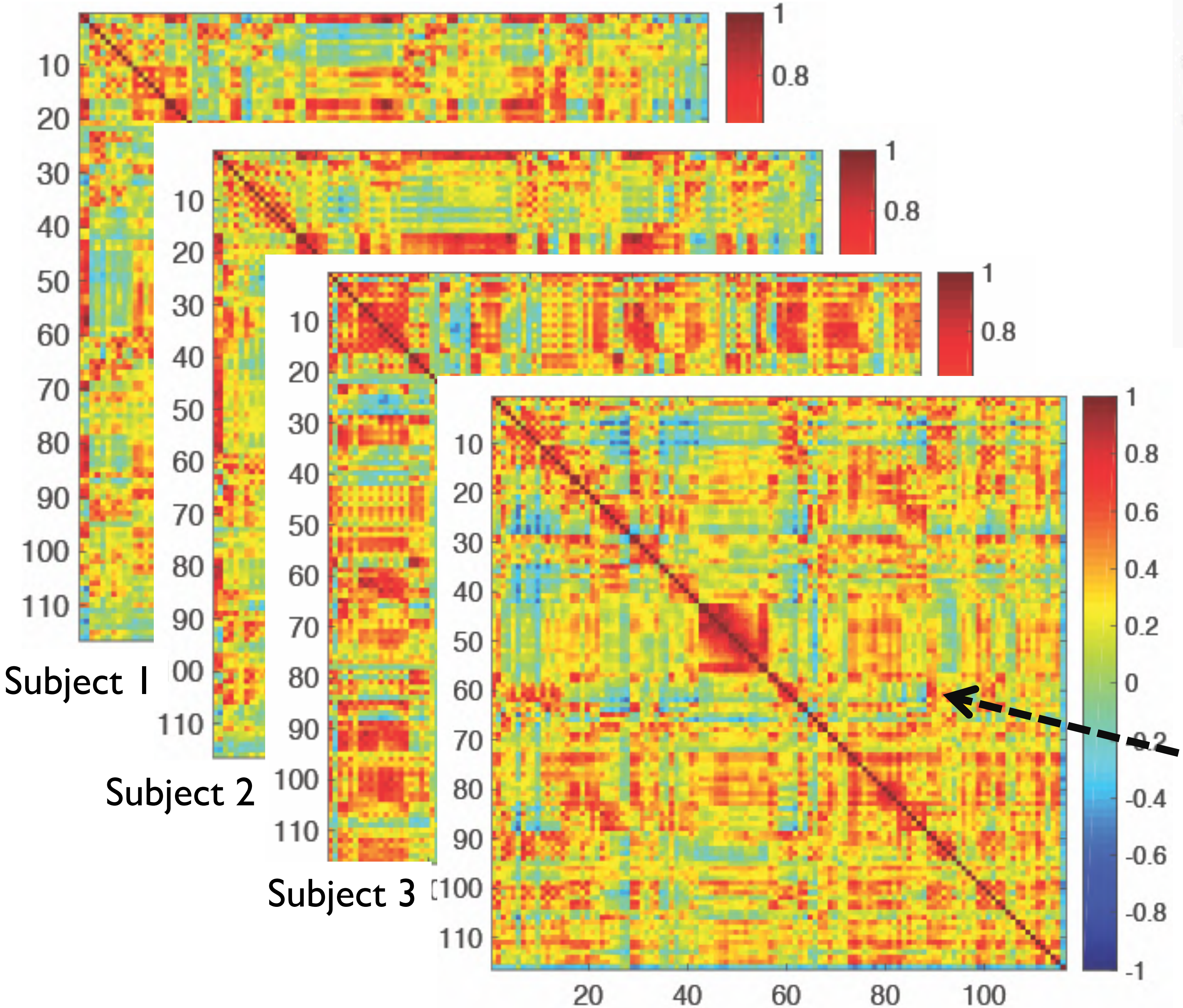
$$\psi_0(t) = 1, \psi_l(t) = \sqrt{2}\cos(l\pi t)$$

120 features ⟶ $\mathbf{d}_i = (d_{0i}, d_{1i}, \cdots, d_{ki})$

Subject level brain connectivity matrix

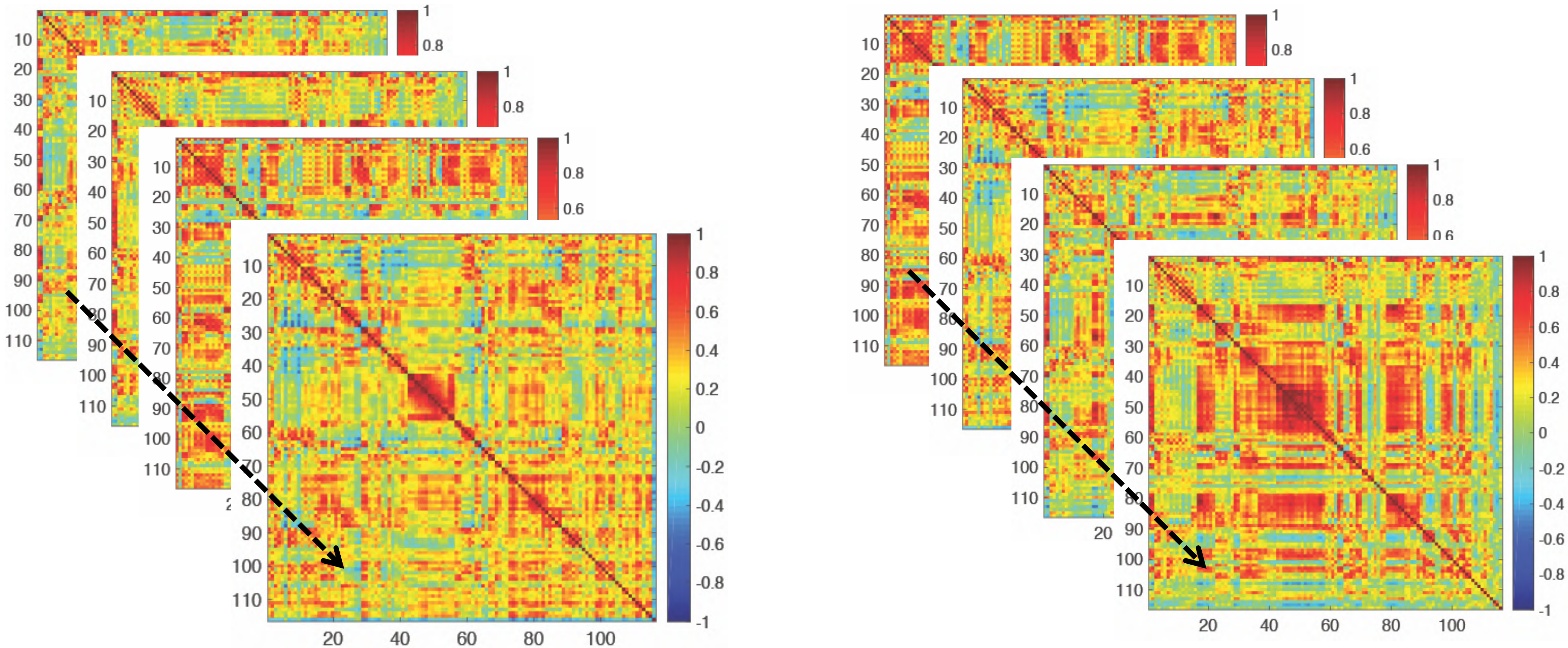


$$c_{ij} = corr(\mathbf{d}_i, \mathbf{d}_j)$$



Correlation of Fourier coefficients

Correlation (group) of correlation (subject)

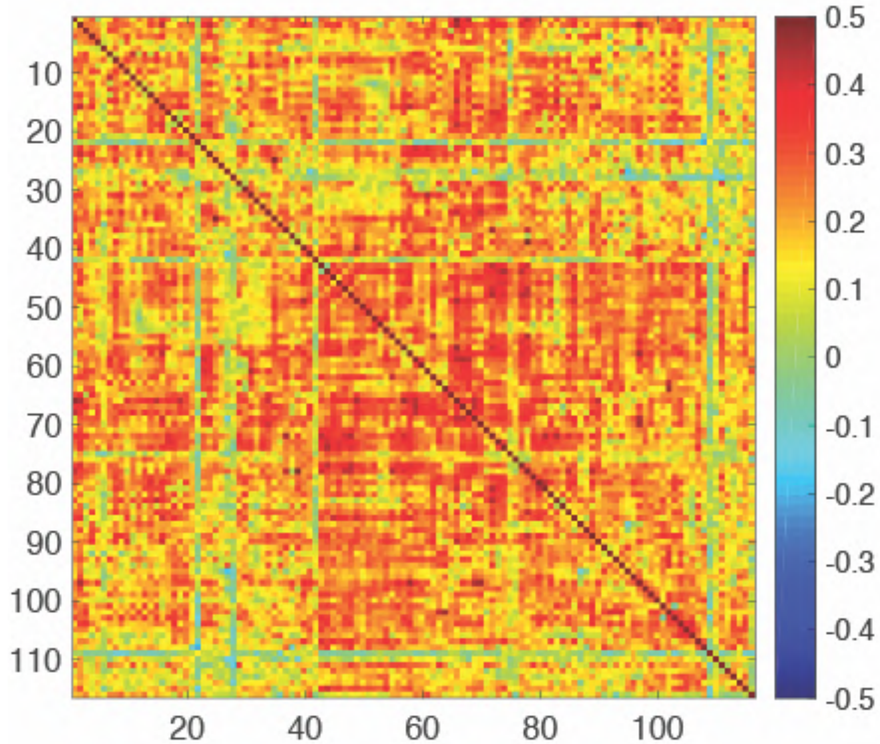


$$\mathbf{c}_{ij}^{1} = (c_{ij}^{11}, \cdots, c_{ij}^{1m}) \qquad \mathbf{c}_{ij}^{2} = (c_{ij}^{21}, \cdots, c_{ij}^{2m})$$

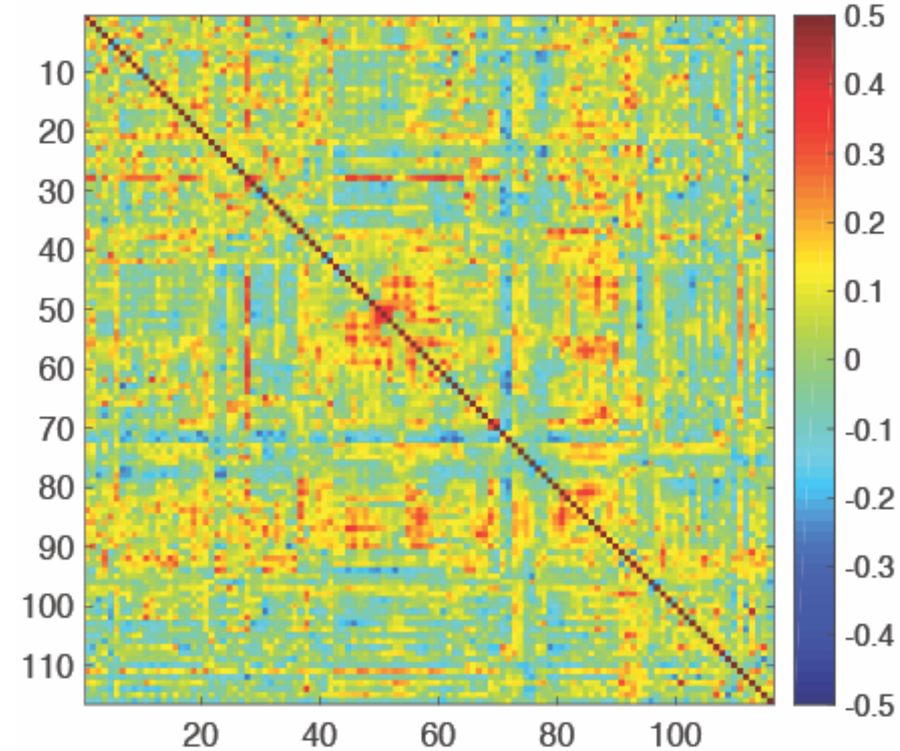
$$c_{ij}^{MZ} = corr(\mathbf{c}_{ij}^{1}, \mathbf{c}_{ij}^{2})$$

MZ- and DZ-twin correlation difference

c_{ij}^{MZ}

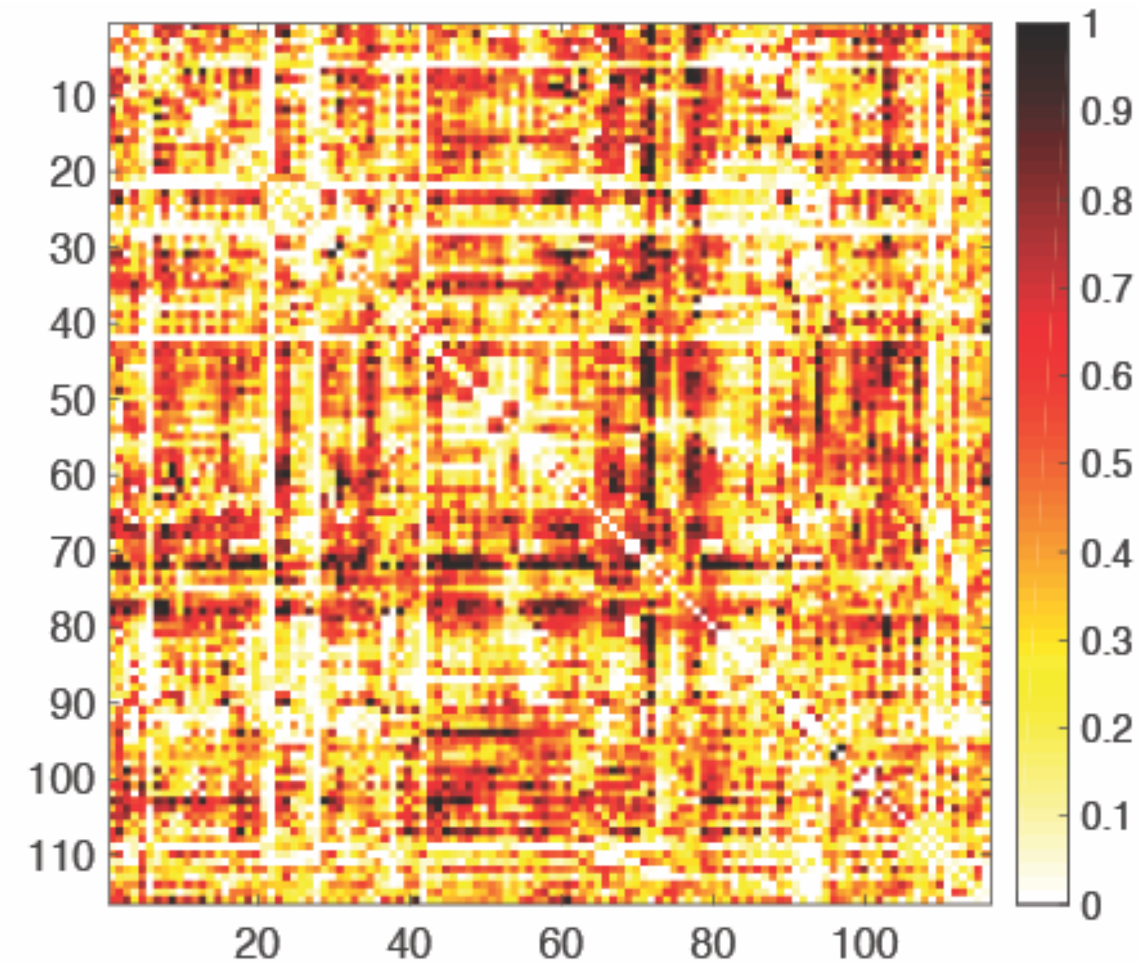


c_{ij}^{DZ}

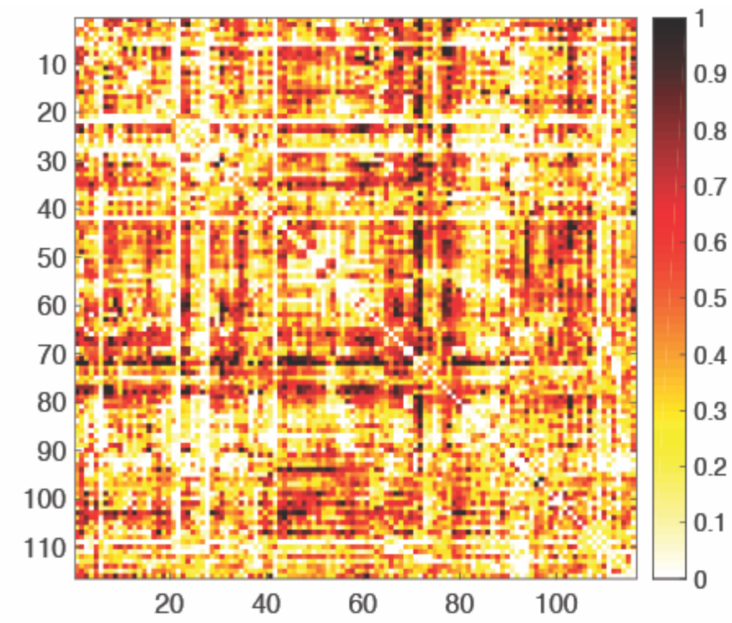


$$h_{ij} = 2(c_{ij}^{MZ} - c_{ij}^{CZ})$$

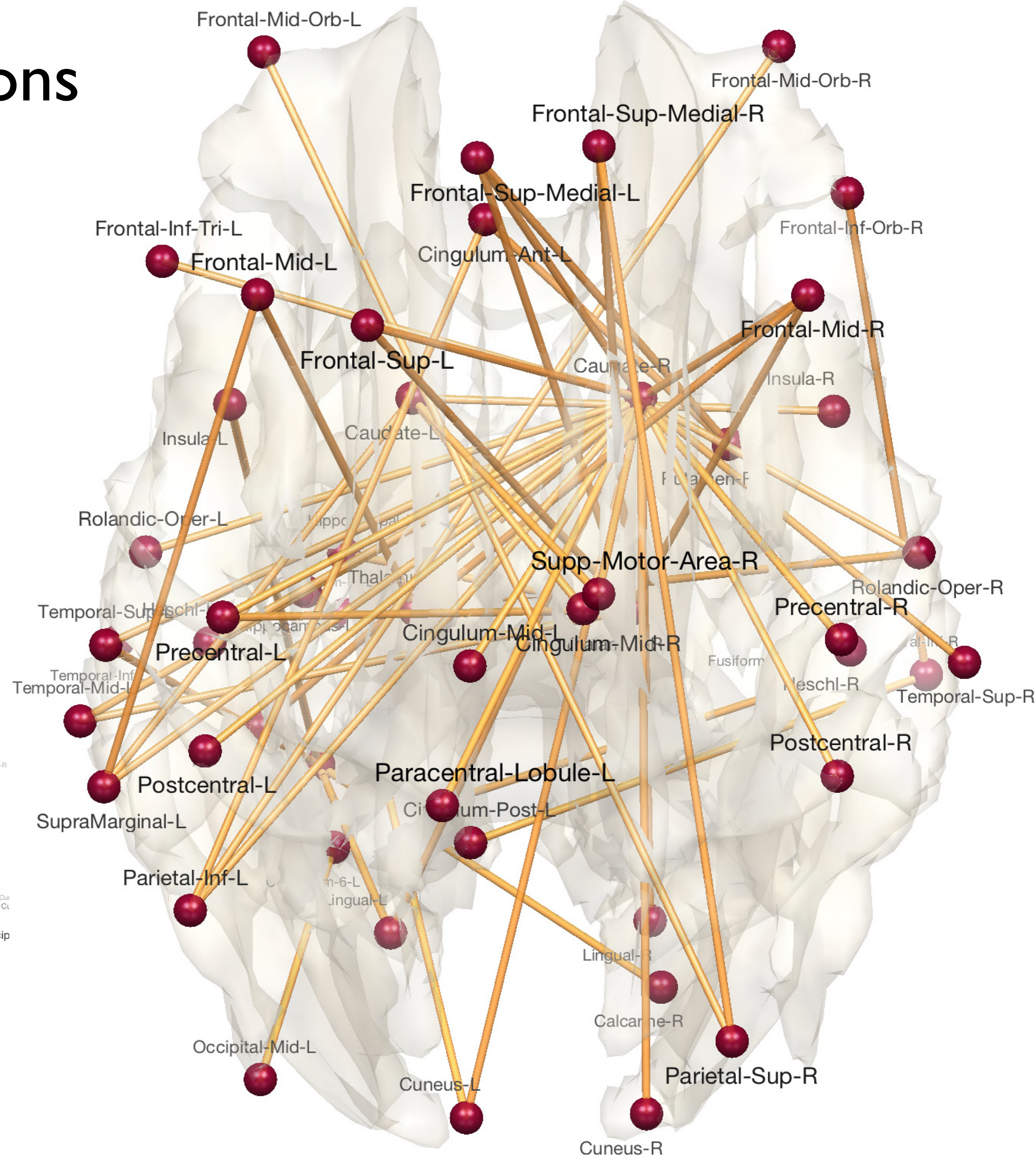
Heritability index = amount of genetic contribution (MZ- and DZ-twins)



Heritable brain regions

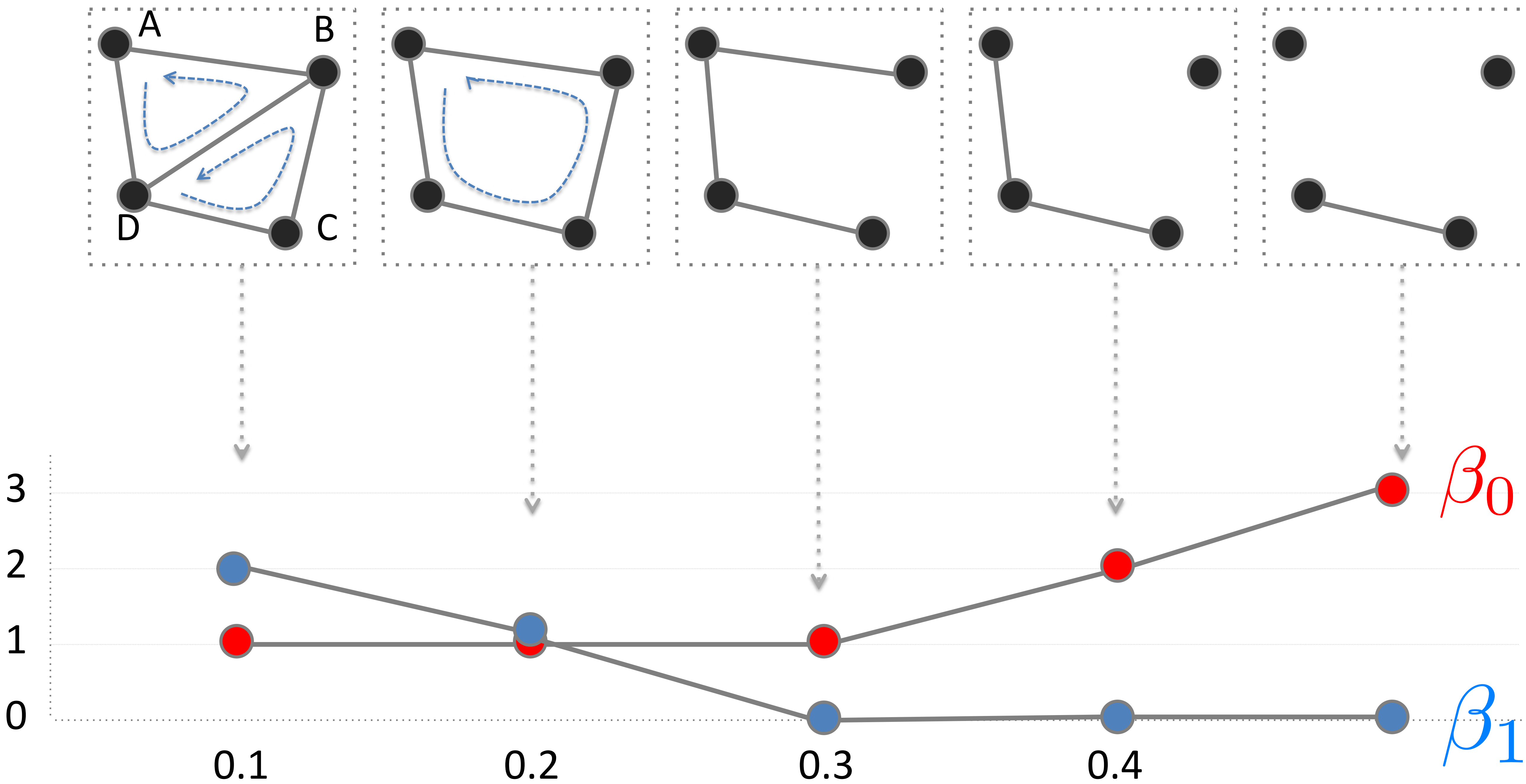


$$h_{ij} \geq 1$$



Statistical significance?

Betti-plots



filtraiton value = correlation

Theorem 1. β_0 and β_1 are monotone over graph filtration.

Monotonicity of β_0: Deletion of edge increases the the number of connected components by at most 1. β_0 increases by 0 or 1.

Monotonicity of β_1: Euler characteristic:

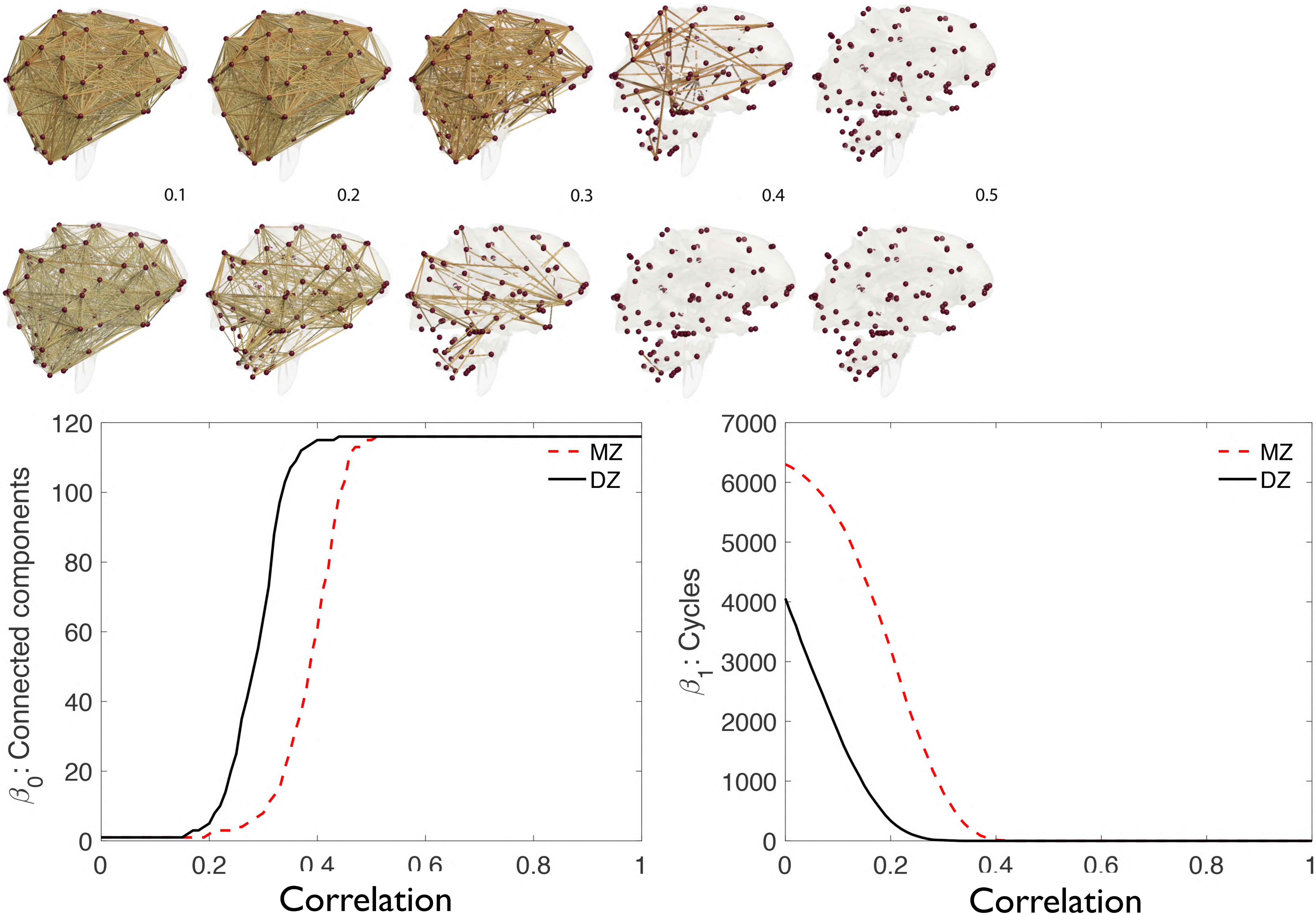
$$\chi = \beta_0 - \beta_1 = p - q$$

nodes (p) edges (q)

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

β_1: -1, 0 β_0: 0, +1 p: fixed q: -1

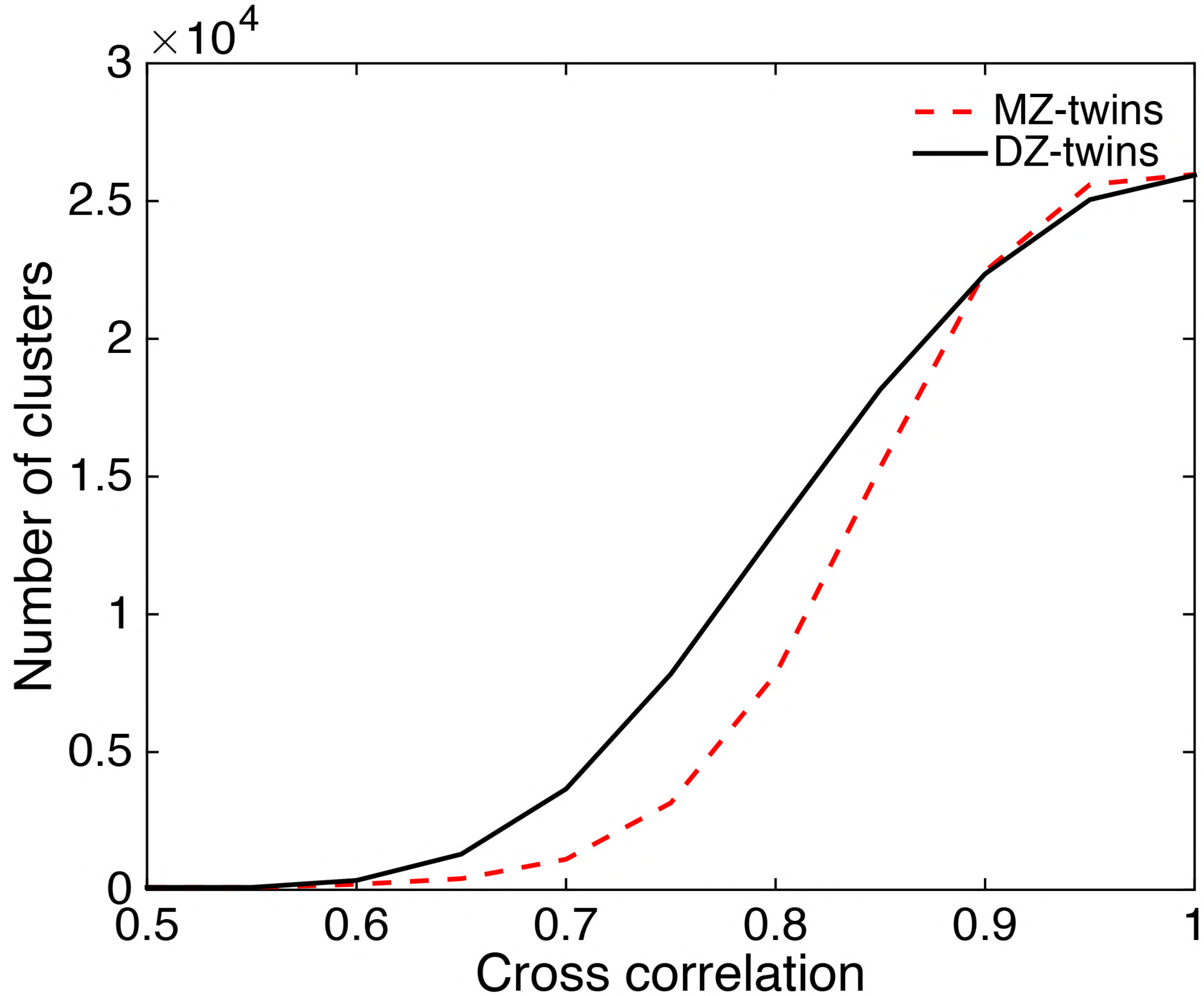
Betti-plots on graph filtration

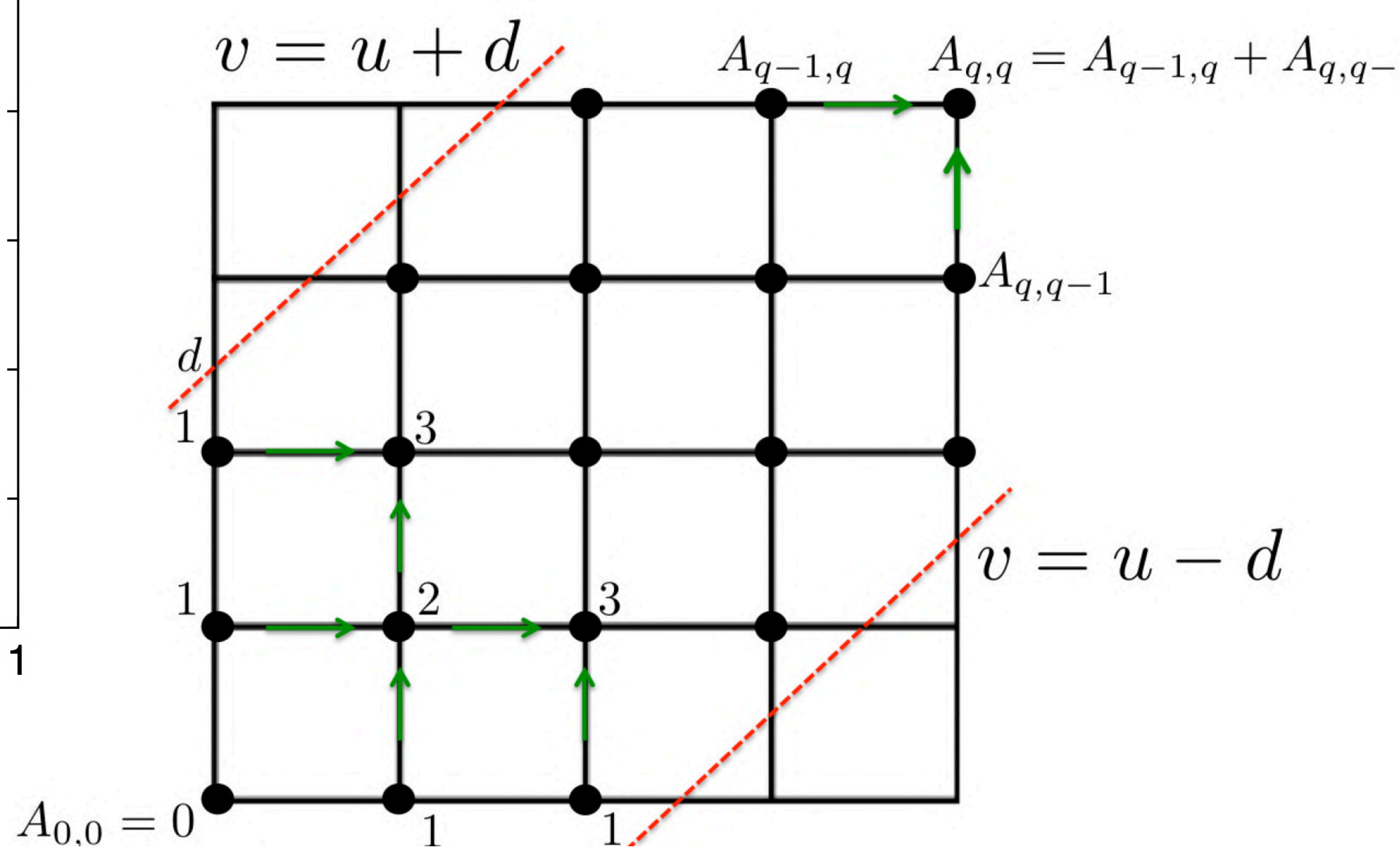


Exact topological inference

Theorem 2. $$D_q = \sup_{1 \le j \le q} \left| \beta_i(G^1_{\lambda_j}) - \beta_i(G^2_{\lambda_j}) \right|$$

$$P(D_q \ge d) = 1 - \frac{A_{q,q}}{\binom{2q}{q}}$$





$$\lim_{q \to \infty} P\left(D_q / \sqrt{2q} \ge d \right) = 2 \sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2 d^2}$$

New approach
Rapid accelation via slow random walks

Chung et al. 2019 arXiv:1812.06696

Random walk on the permutation group

$$\mathbf{x} = (x_1, x_2, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_m)$$

transpose *i*-th and *j*-th data

$$\mathbf{y} = (y_1, y_2, \cdots, y_{j-1}, y_j, y_{j+1}, \cdots, y_n)$$

$$\pi_{ij}(\mathbf{x}) = (x_1, x_2, \cdots, x_{i-1}, y_j, x_{i+1}, \cdots, x_m)$$

$$\pi_{ij}(\mathbf{y}) = (y_1, y_2, \cdots, y_{j-1}, x_i, y_{j+1}, \cdots, y_n)$$

Online computation of statistic over transposition

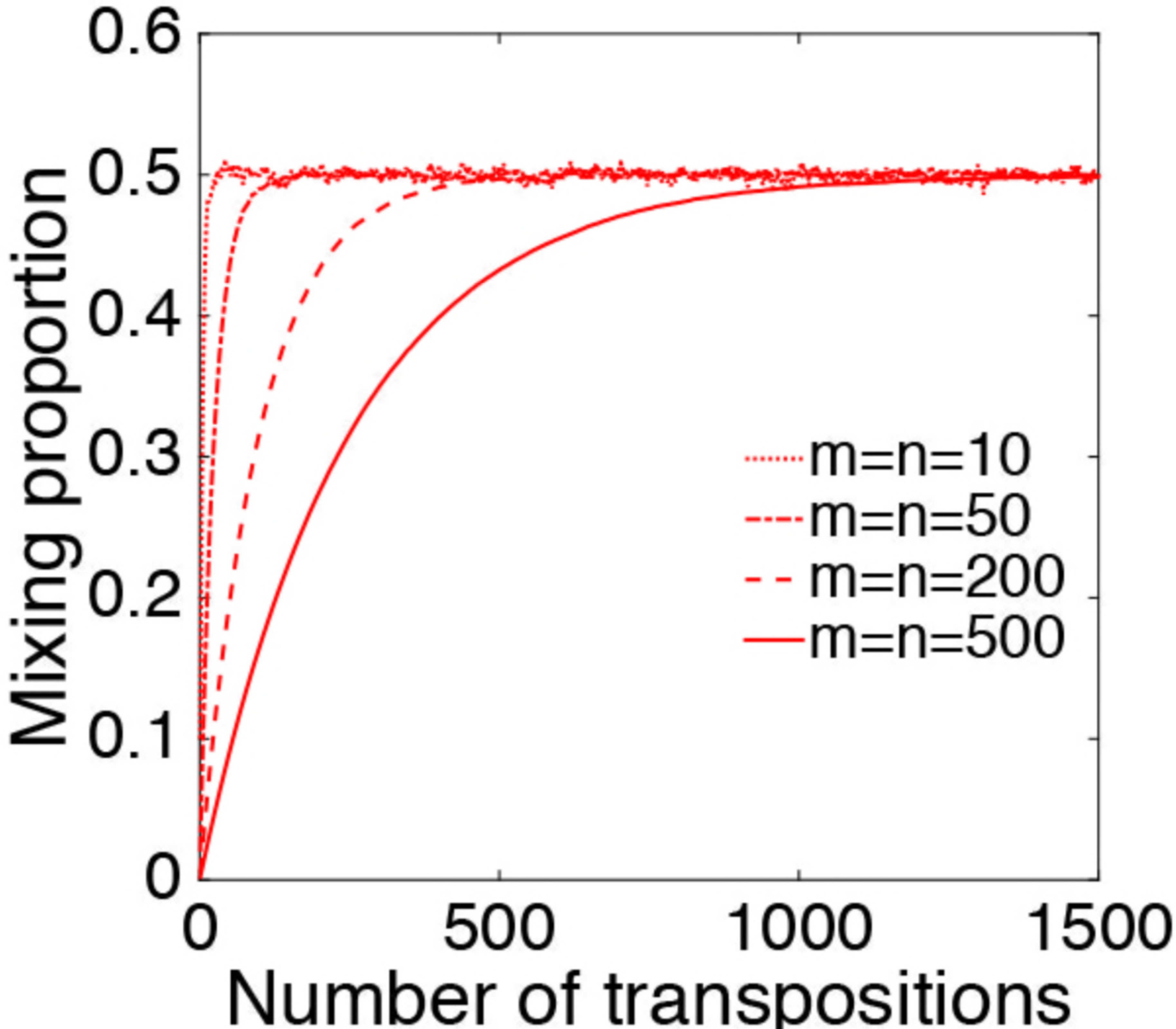
$$f(\mathbf{x}, \mathbf{y}) \dashrightarrow f(\pi_{ij}(\mathbf{x}), \pi_{ij}(\mathbf{y}))$$

Theorem 3. If the test statistic is algebraic, there exists a function g such that

$$f(\pi_{ij}(\mathbf{x}), \pi_{ij}(\mathbf{y})) = g(f(\mathbf{x}, \mathbf{y}), x_i, y_i)$$

where computational complexity of g is constant.

Theorem 4. Any permutation in $\mathbb{S}_{m+n}$ can be reachable by a sequence of walks.



Online computation

$$\mu(\mathbf{x}) = \frac{1}{m}\sum_{k=1}^{m} x_k$$

Transpose *i*-th and *j*-th ⟷

$$\mu(\mathbf{y}) = \frac{1}{m}\sum_{k=1}^{n} y_k$$

O(*n*)

$$m\mu(\pi_{ij}(\mathbf{x})) = m\mu(\mathbf{x}) + y_j - x_i$$

O(2)

$$\sigma^2(\mathbf{x}) = \frac{1}{m-1}\sum_{k=1}^{m}(x_k - \mu(\mathbf{x}))^2$$

Transpose *i*-th and *j*-th ↕

$$\sigma^2(\mathbf{y}) = \frac{1}{m-1}\sum_{k=1}^{n}(y_k - \mu(\mathbf{y}))^2$$

O(3*m*+2)

$$(m-1)\sigma^2(\pi_{ij}(\mathbf{x})) =$$

O(9)

$$(m-1)\sigma^2(\mathbf{x}) + \Big((m\mu(\mathbf{x}))^2 - (m\mu(\pi_{ij}(\mathbf{x})))^2\Big)/m + y_j^2 - x_i^2$$

Computational Complexity

$$T(\pi_{ij}(\mathbf{x}), \pi_{ij}(\mathbf{y})) = \frac{\left(\frac{\nu(\pi_{ij}(\mathbf{x}))}{m} - \frac{\nu(\pi_{ij}(\mathbf{y}))}{n}\right)}{\sqrt{\frac{\omega^2(\pi_{ij}(\mathbf{x}))+\omega^2(\pi_{ij}(\mathbf{y}))}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}}$$

T-stat computation per permutation

Standard method: $O(4m+4n+20)$

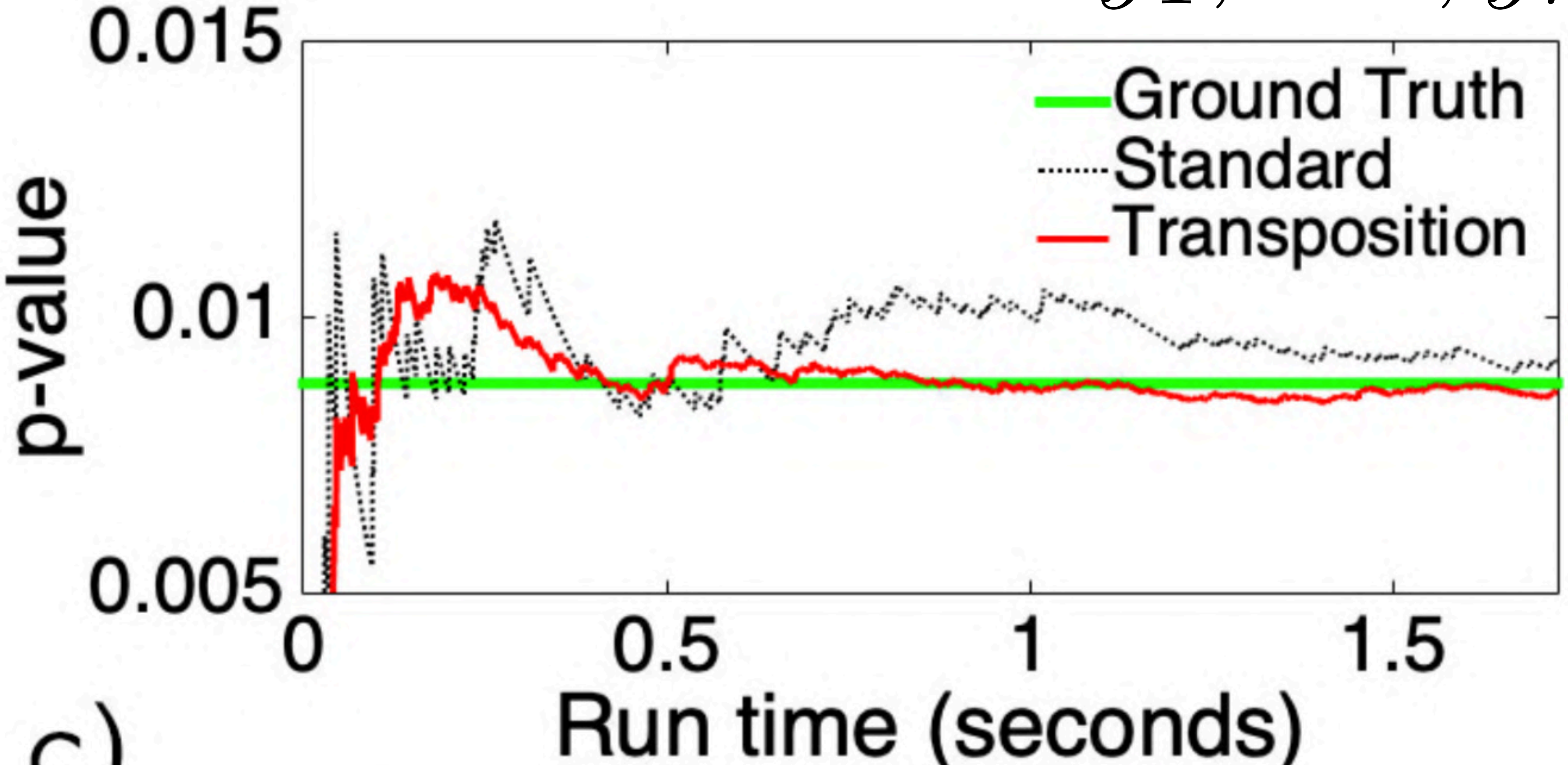
Our approach: $O(35)$

Simulation

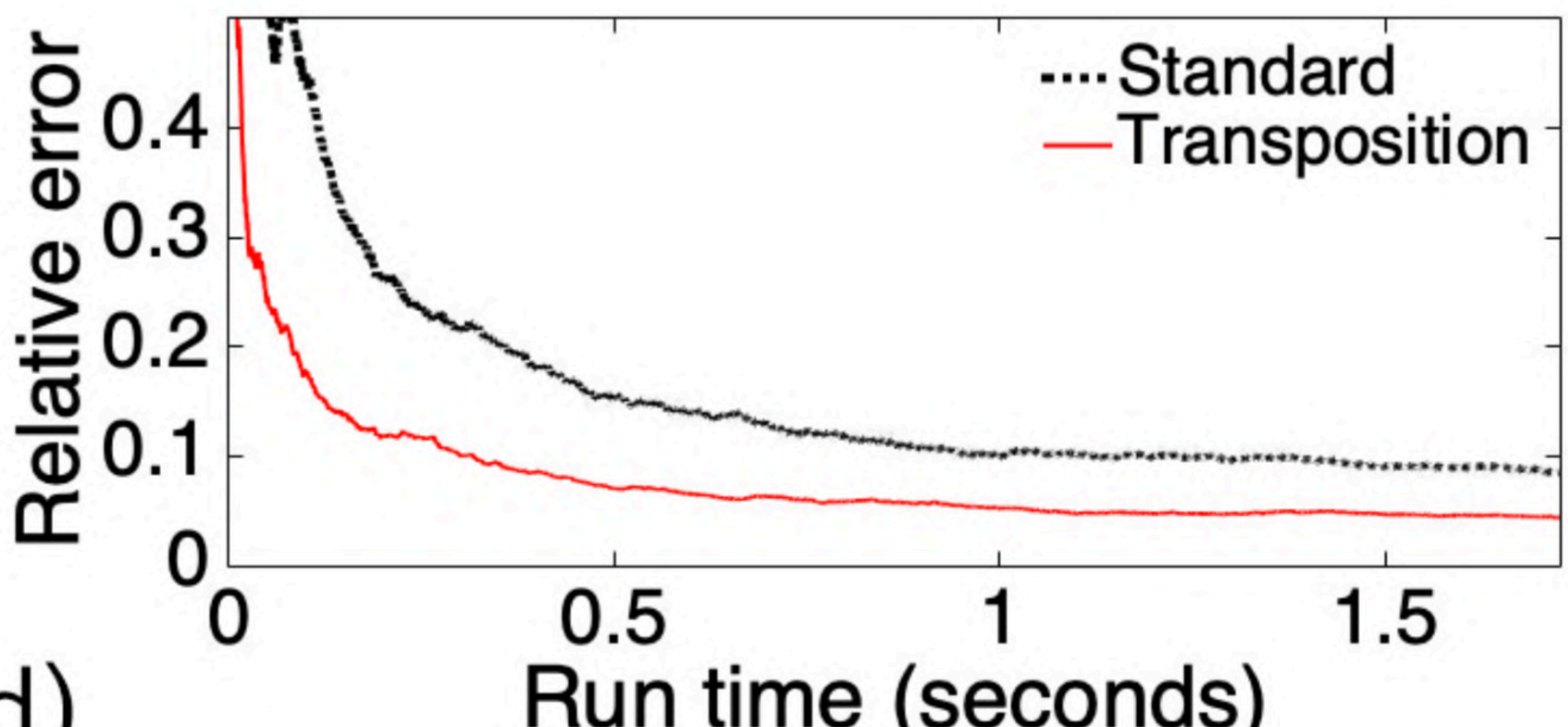
$$x_1, \cdots, x_m \sim N(0,1)$$

$$y_1, \cdots, y_n \sim 0.1 + N(0,1)$$

m=n=100



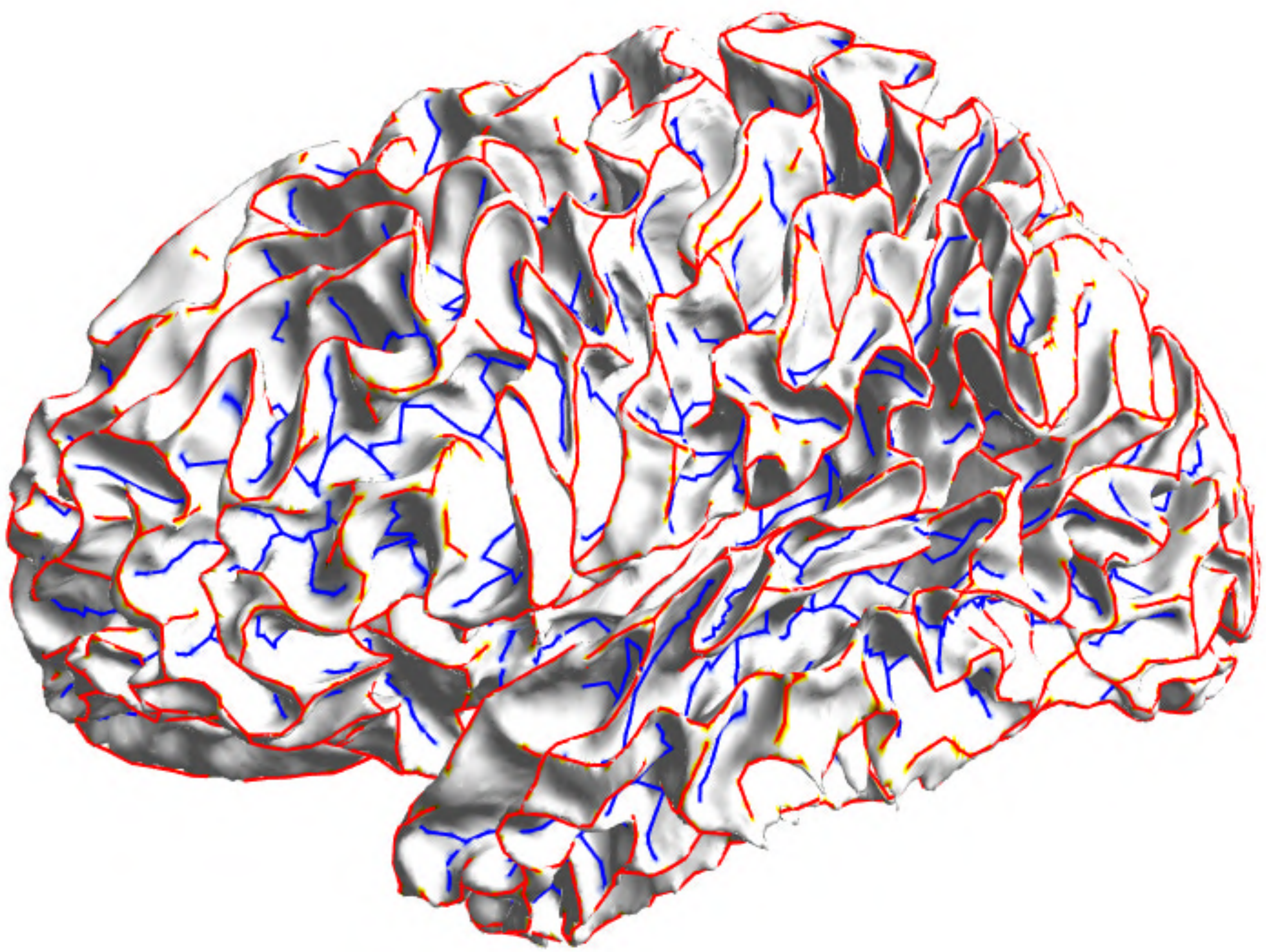
c)

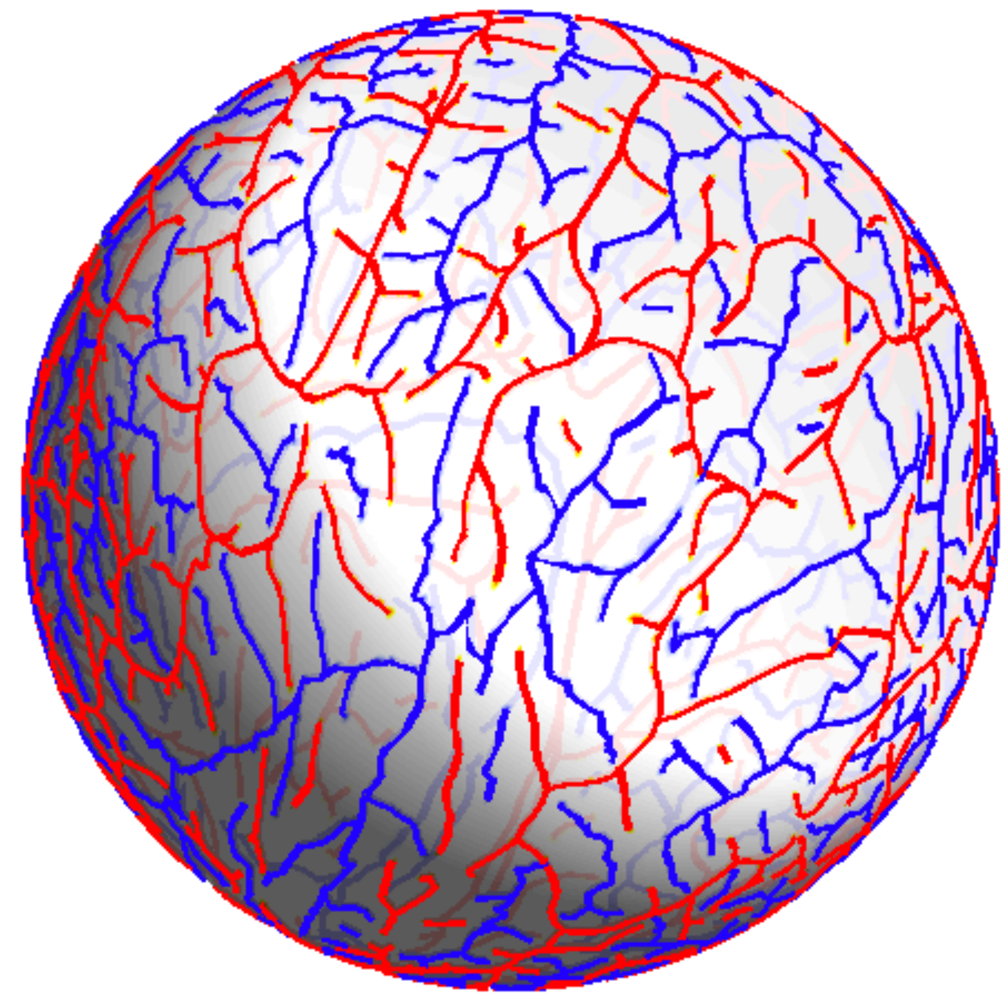


d)

Sulcal and gyral graphs on brain cortical surface

Sulci = mountain regions of the brain (red)
Gyri = valley regions of the brain (blue)



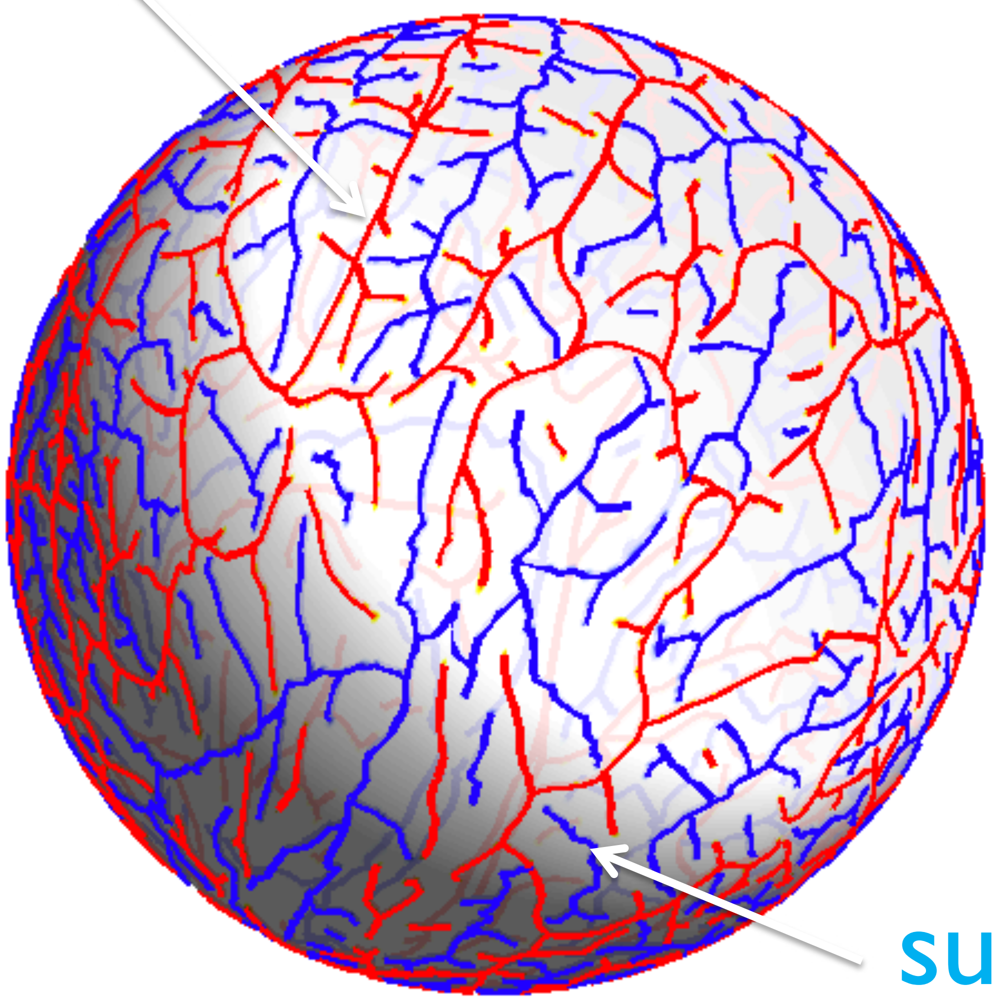


White matter surface

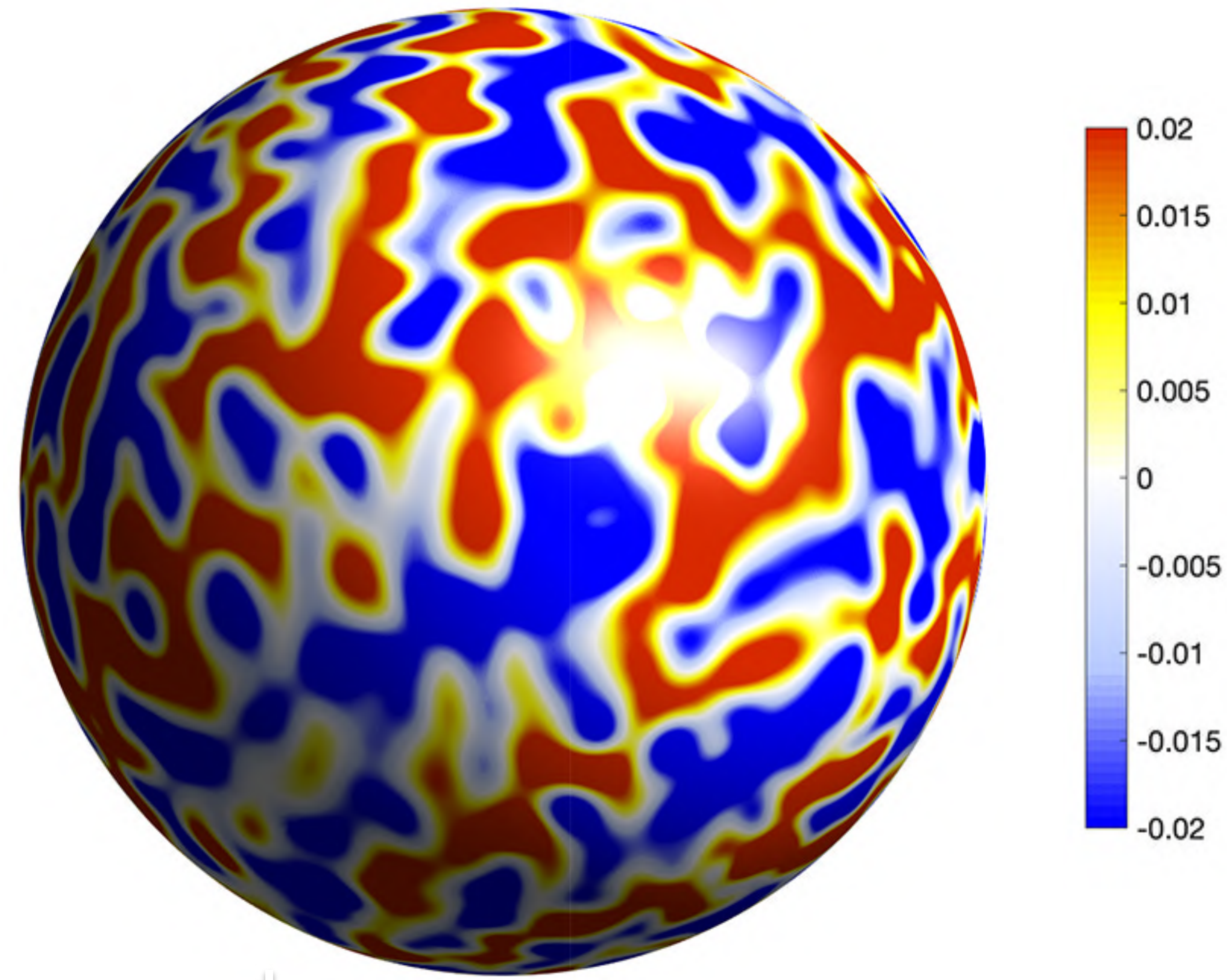
Sulcal/gyral graphs on manifold

Heat kernel smoothing (*Chung et al. 2007 IEEE TMI*)

gyri=+1

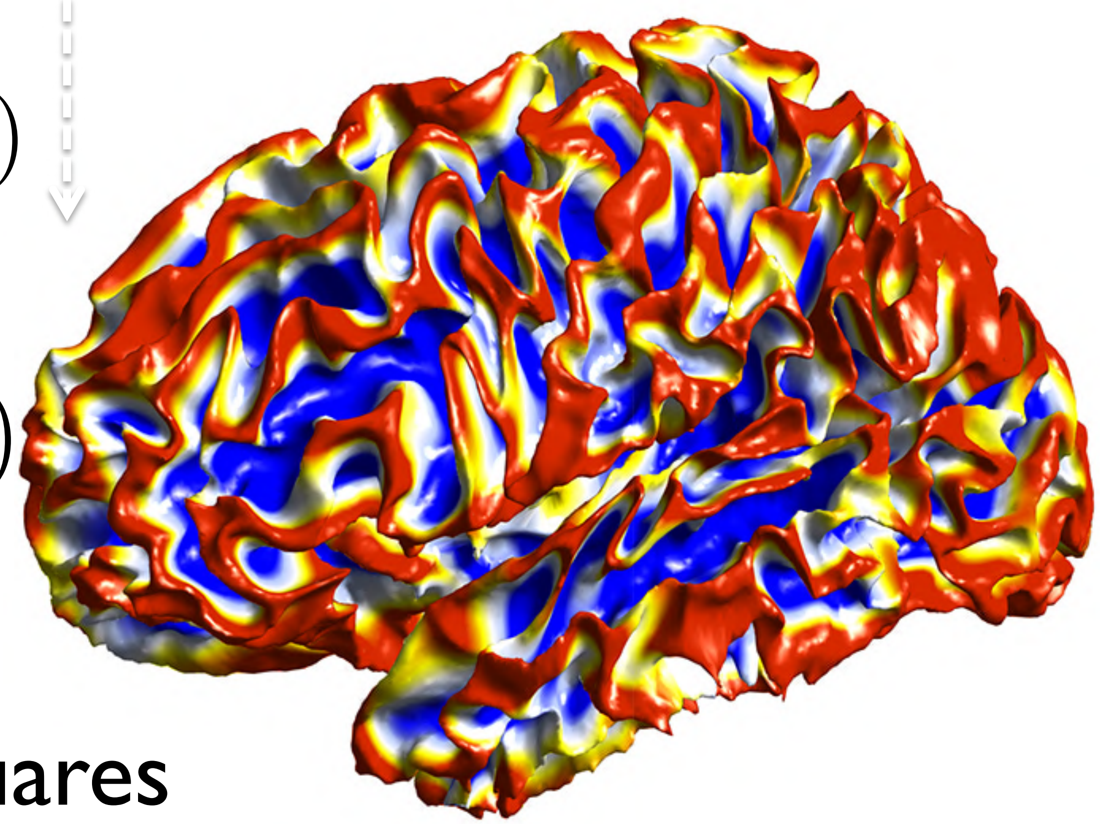


sulci = -1



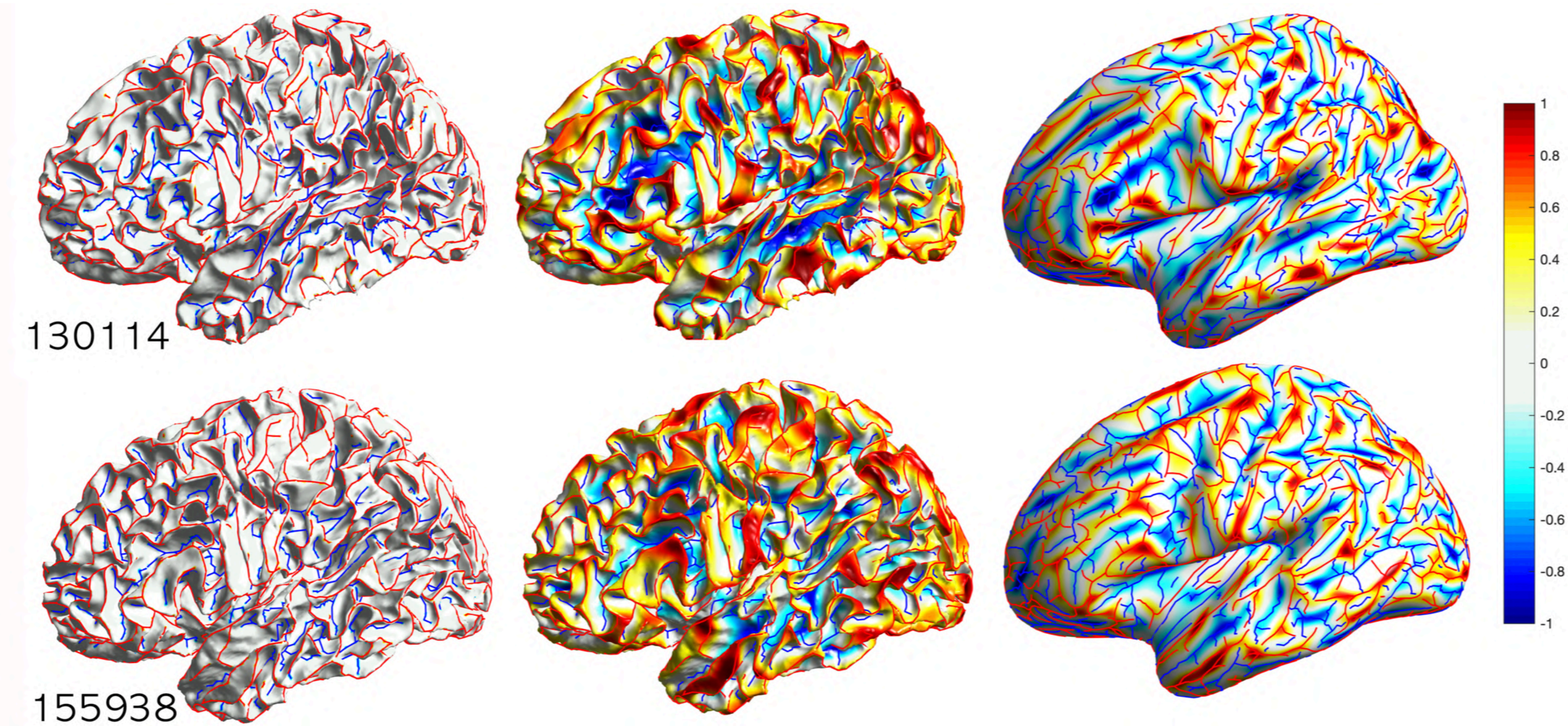
$$\frac{dg(p,t)}{dt} = \Delta_p g(p,t) \quad g(p,t=0) = f(p)$$

$$g(p,t) = \sum_{l=0}^{100} \sum_{m=-l}^{l} f_{lm} e^{-l(l+1)t} Y_{lm}(p)$$



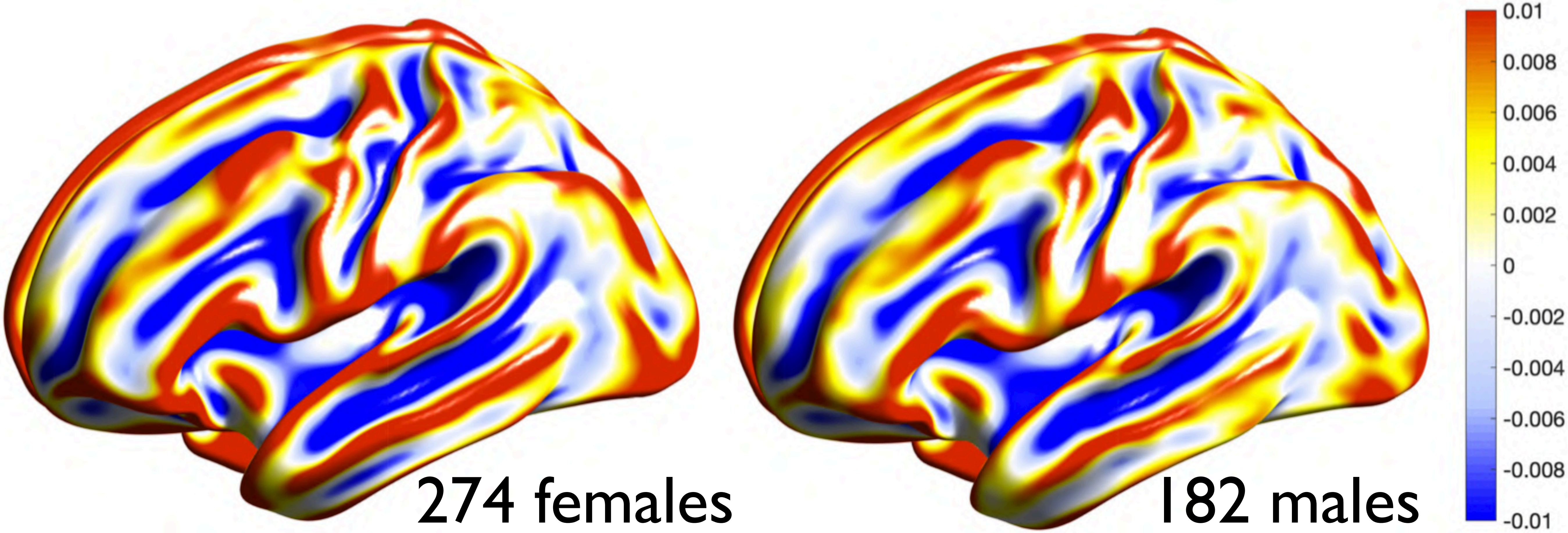
10201 parameters estimated via least squares

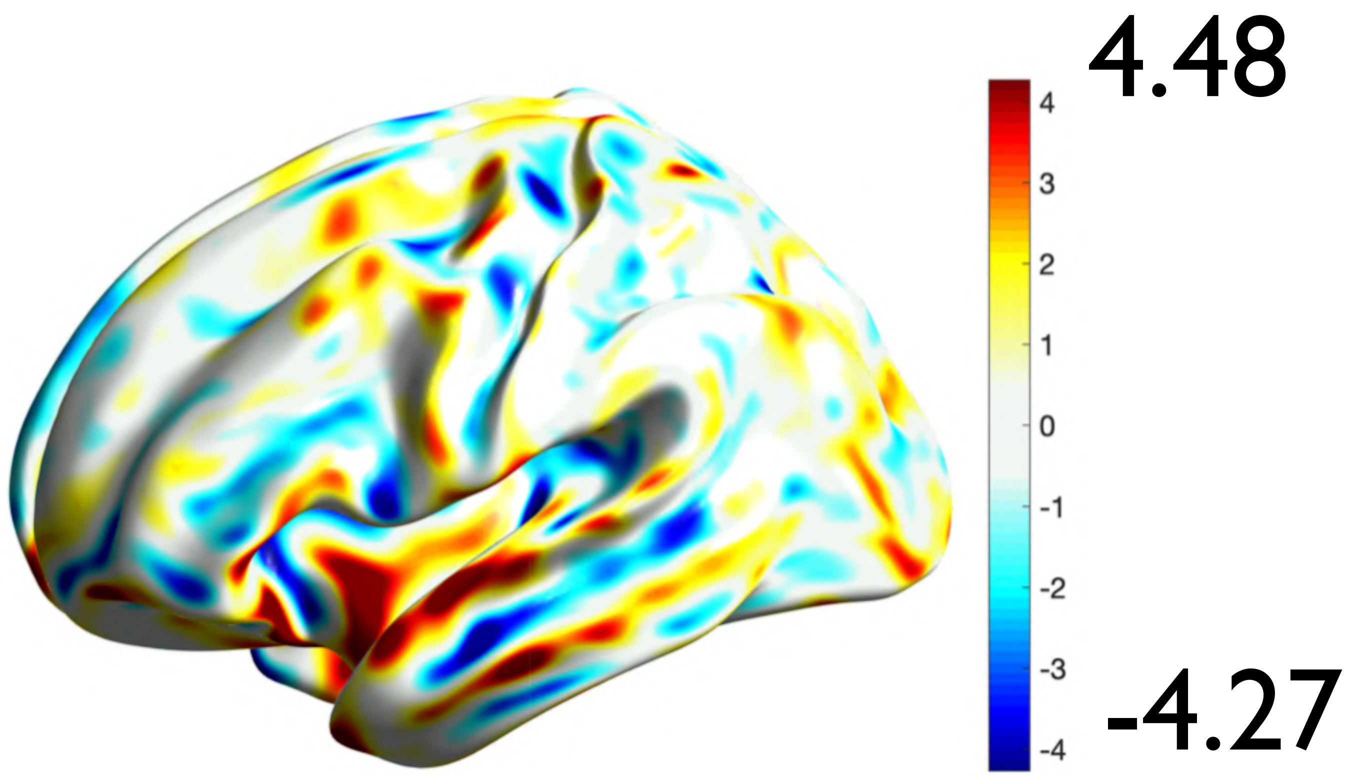
Heat kernel smoothing



Sulcal/gyral graphs Heat kernel smoothing Surface flattening

Group mean differences





t-statistic (thresholds at corrected pvalue of 0.05)

Our method: 40min.

Standard methods: 18 days

Thank you!

Question? mkchung@wisc.edu