

# Developing Methods Towards Understanding Populations of Networks

Jaewon Chung

(he/him) - [NeuroData lab](#)

Johns Hopkins University

Department of Biomedical Engineering

 [j1c@jhu.edu](mailto:j1c@jhu.edu)

 [@j1c \(Github\)](#)

 [@j1c \(Twitter\)](#)

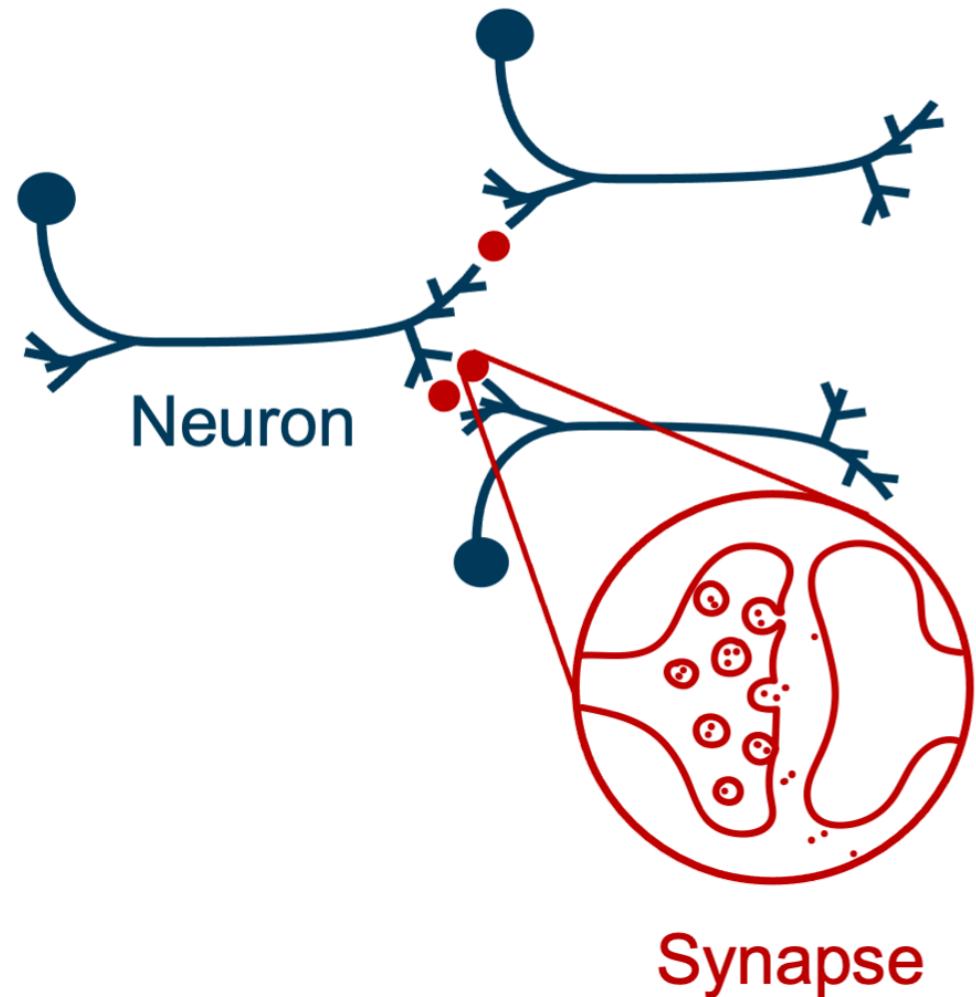


# Overview

- M.S. and Ph.D. in Biomedical Engineering @ Johns Hopkins University
- Focus:
  - i. developing statistical methods
    - for network data, specifically brain networks
    - for time-series data
    - for causal inference in observational studies
  - ii. developing open-source software

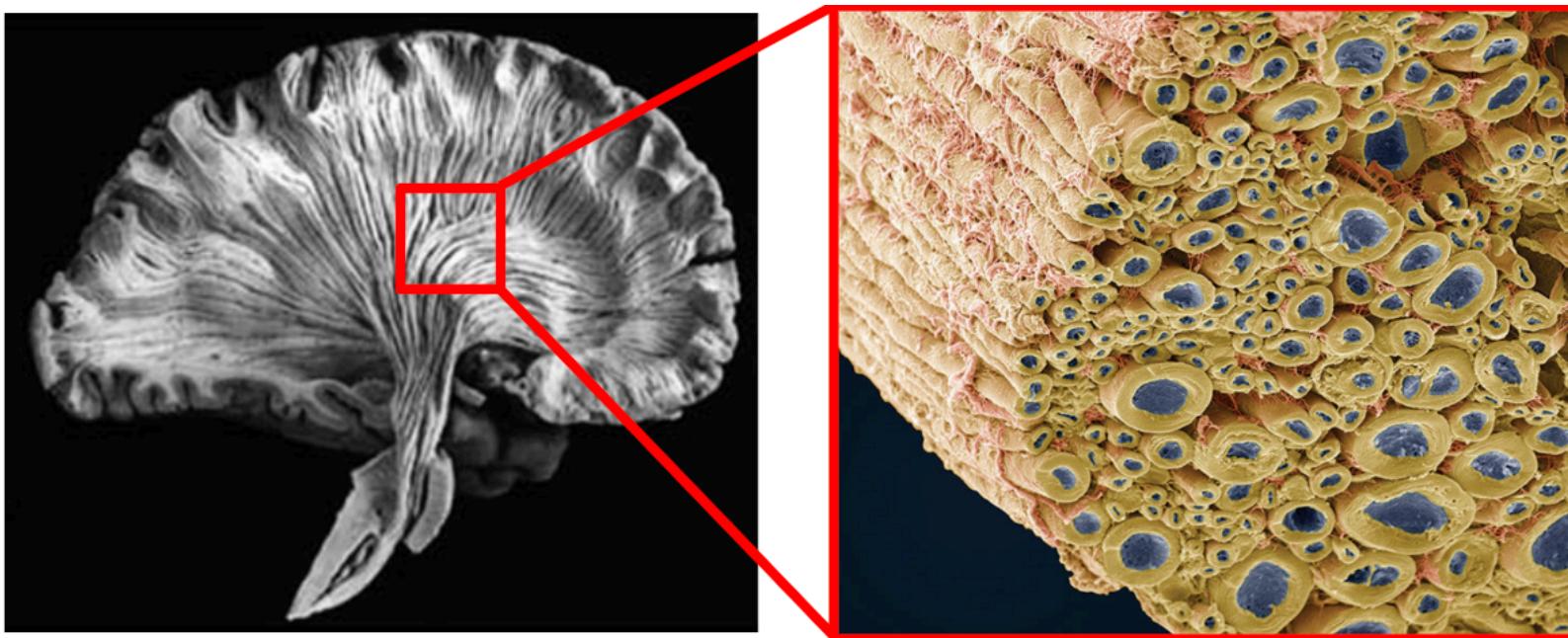
# Connectomes: maps of neural wiring

- Brains contain **neurons**, which carry information via electrical signals
- Neurons **connect** to each other via synapses, allowing neurons to "talk" to each other
- **Connectome** is a map of the structure of neurons and the connections between them
  - Shaped by evolution, experience, influences neural activity, behavior



# How do we measure connectomes in humans?

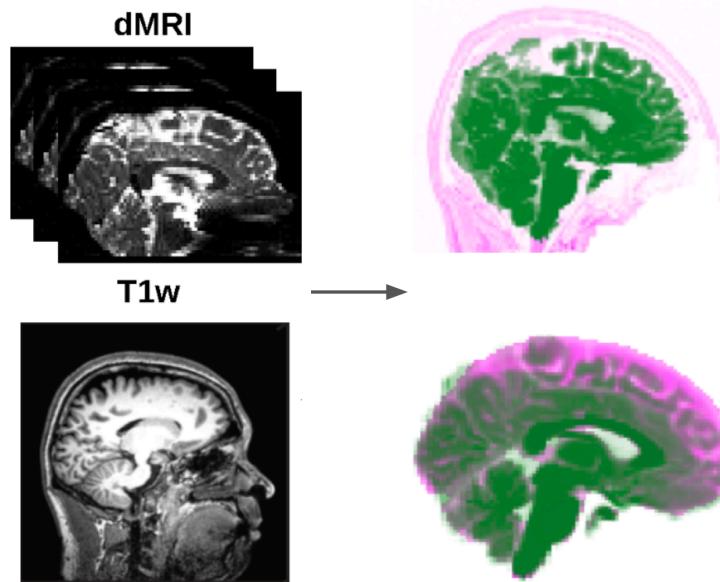
- Diffusion MRI (dMRI): *in vivo* imaging technique
- Exploits direction of water diffusion
  - Anisotropic in white matter tracts
  - Isotropic in other tissues



## MRI Scans



## Preprocessing



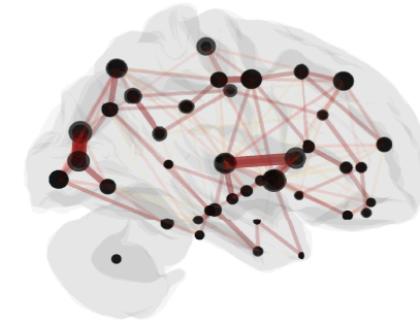
## Tractography



# Representing brains as networks

Networks (or graphs) are mathematical abstractions to represent relational data

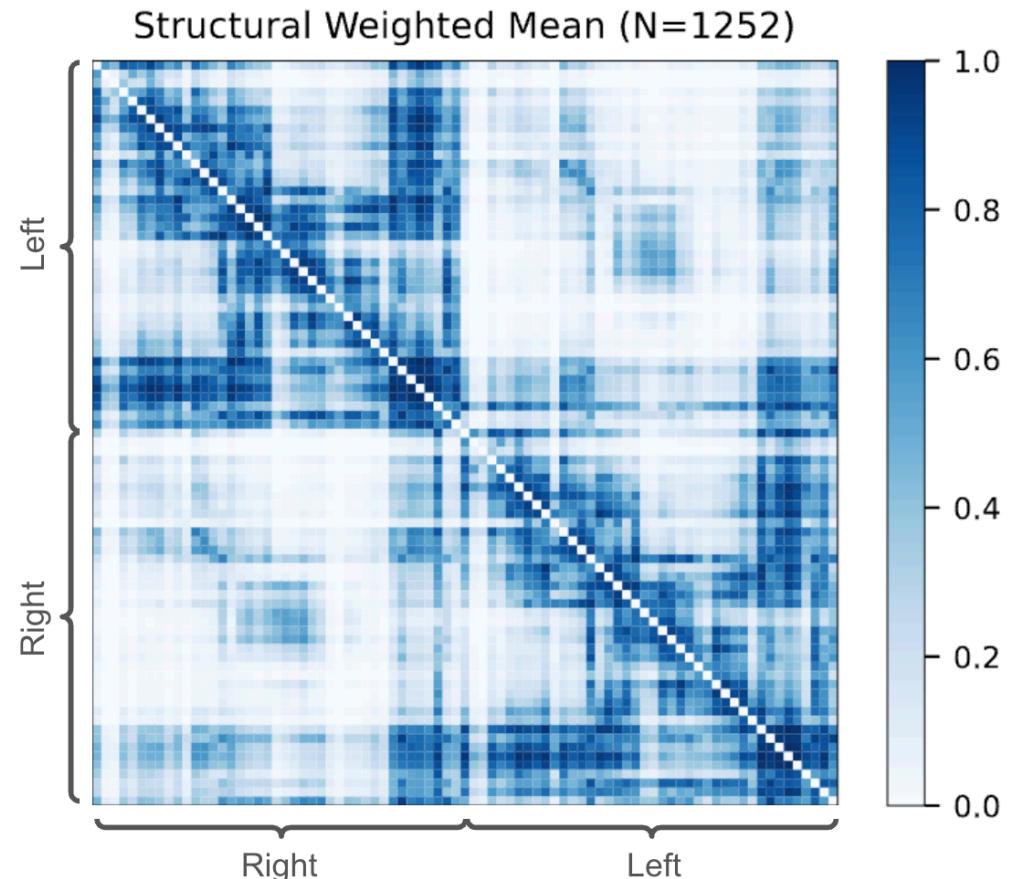
- **Vertices** - the set of objects (brain regions)
- **Edges** - the set of connections between those objects (brain regions)
  - E.g. region 1 connects to region 2 with 100 neural bundles



$$\begin{bmatrix} 0 & 3 & 0 & 0 & 1 & 0 \\ 3 & 0 & 2 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 6 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 6 & 0 & 0 \end{bmatrix}$$

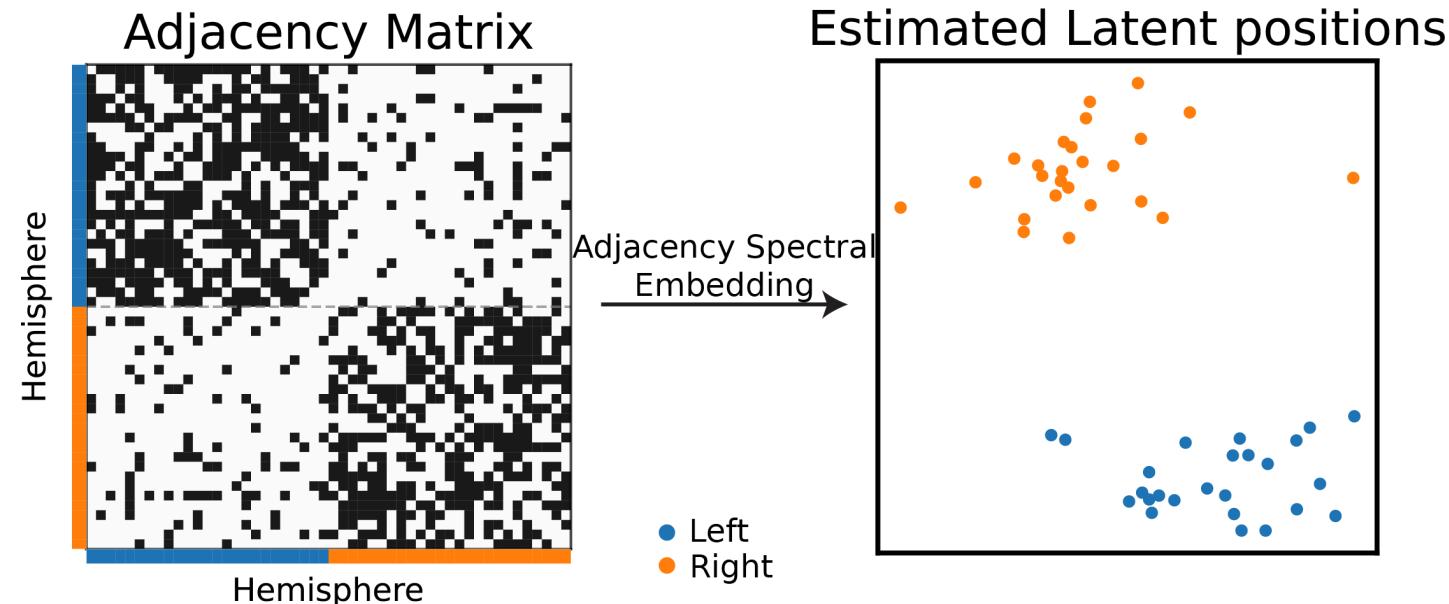
# MRI to graphs (m2g)

- Easy to use end-to-end pipeline
  - Input: MRI data
  - Output: Connectomes, QA measures, derivatives
- Reproduces biological properties
  - Stronger ipsilateral connections
- High discriminability
  - Same subjects' connectomes are more similar than different subjects'



# Statistical models for networks

- Random dot product graphs (RDPGs)
  - Each vertex has a low  $d$  dimensional latent position.
  - Estimate latent position matrix  $X$  via adjacency spectral embedding.
  - $P[i \rightarrow j] = \langle x_i, x_j \rangle$



# Two sample graph testing

- Suppose we have two networks
- Want to test if they are "same" or not

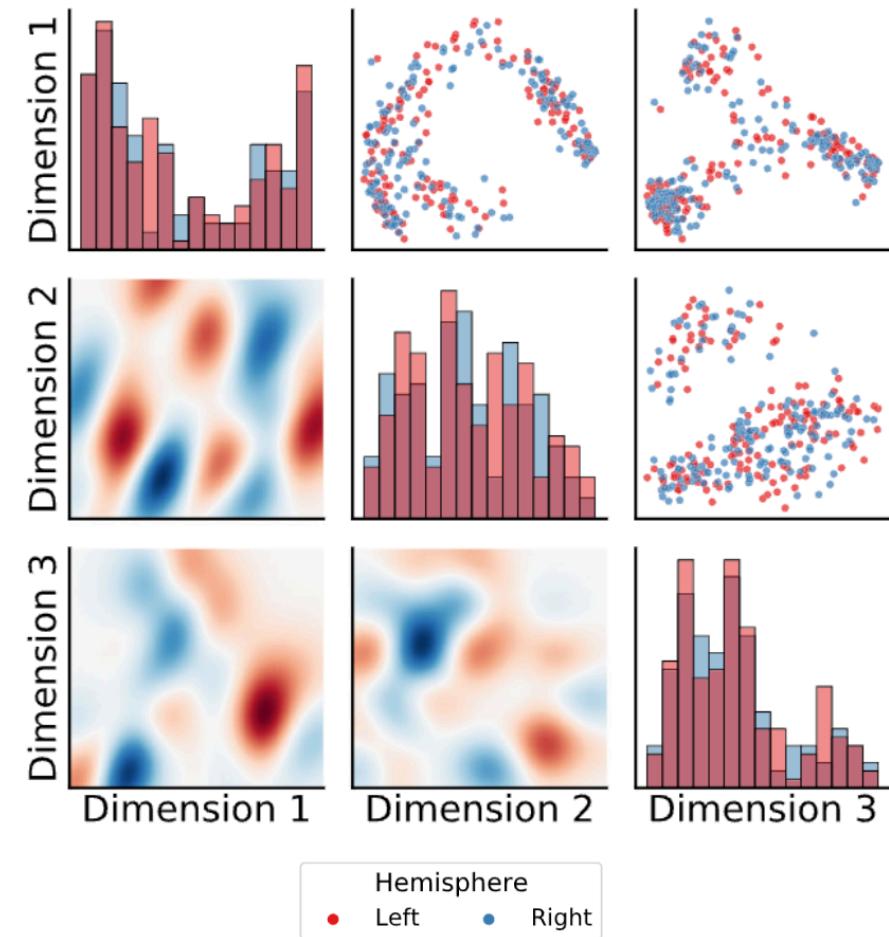
Hypothesis:

- $H_0 : F(\text{Network 1}) = F(\text{Network 2})$
- $H_A : F(\text{Network 1}) \neq F(\text{Network 2})$

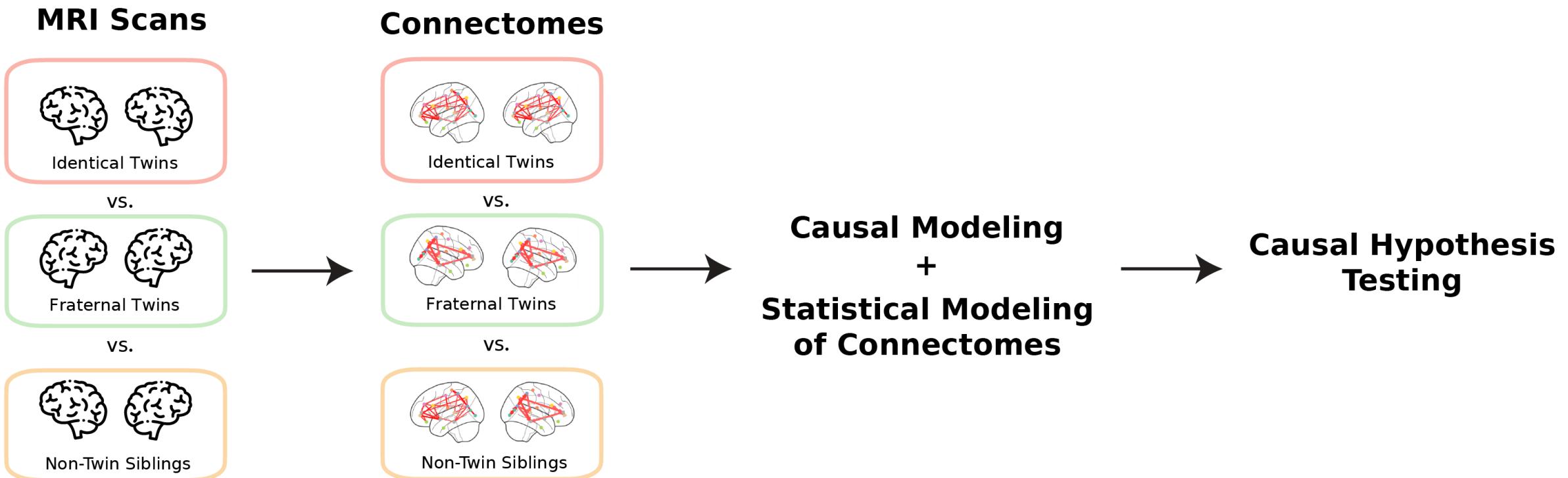
More precisely:

- $H_0 : F_X = F_Y \circ W$
- $H_A : F_X \neq F_Y \circ W$

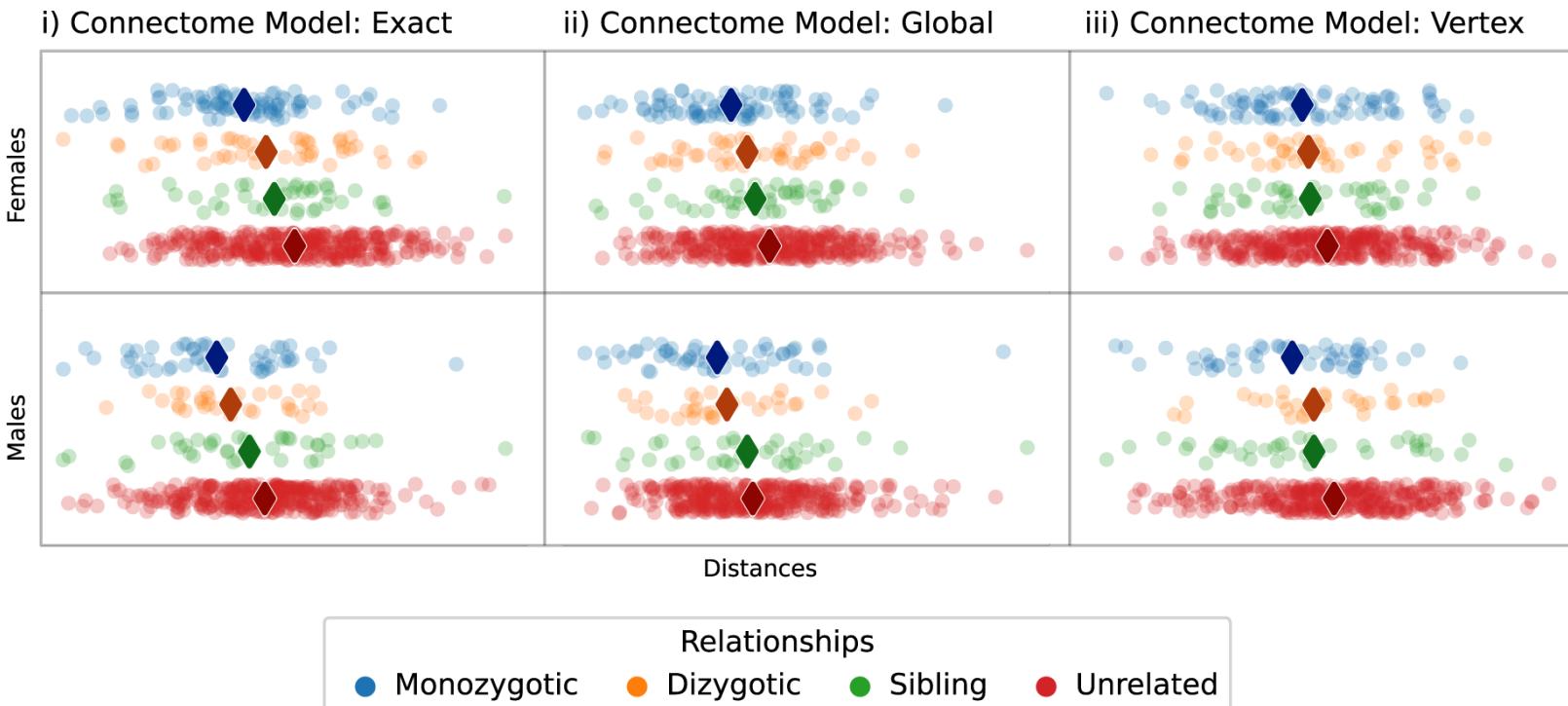
Drosophila Left vs Right Brain



# Connectomic heritability

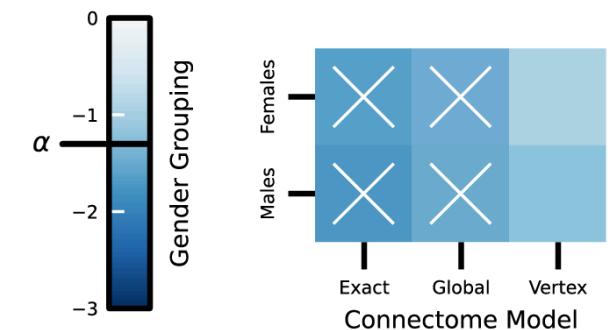


### A) Pairwise Distances Separated by Familial Relationships



### B) P-values From Tests for Heritability

#### Causal Hypothesis Test



# Other works

## 1. Independence Testing in Multivariate Time-series data

- Let  $(X_t, Y_t) \sim F_{XY}$ ,  $t \in [n] = \{1, 2, \dots, n\}$
- Maximum lag  $M$ .

$$H_0 : F_{X_t, Y_{t-j}} = F_{X_t} F_{Y_{t-j}} \text{ for each } j \in \{0, 1, \dots, M\}$$

$$H_A : F_{X_t, Y_{t-j}} \neq F_{X_t} F_{Y_{t-j}} \text{ for some } j \in \{0, 1, \dots, M\}$$

## 2. Hypothesis Testing for Causal Effects

$$H_0 : F_{\mathbf{y}_i, \mathbf{v}_i | \mathbf{x}_i} = F_{\mathbf{y}_i | \mathbf{x}_i} F_{\mathbf{v}_i | \mathbf{x}_i}$$

$$H_A : F_{\mathbf{y}_i, \mathbf{v}_i | \mathbf{x}_i} \neq F_{\mathbf{y}_i | \mathbf{x}_i} F_{\mathbf{v}_i | \mathbf{x}_i}$$

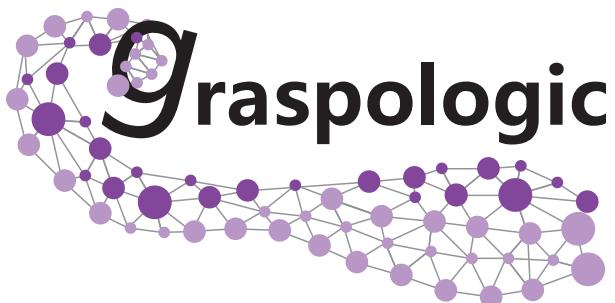
# Open-source software

## graspologic

downloads 273k

Stars 595

contributors 54



## hippo

downloads 184k

Stars 215

contributors 20

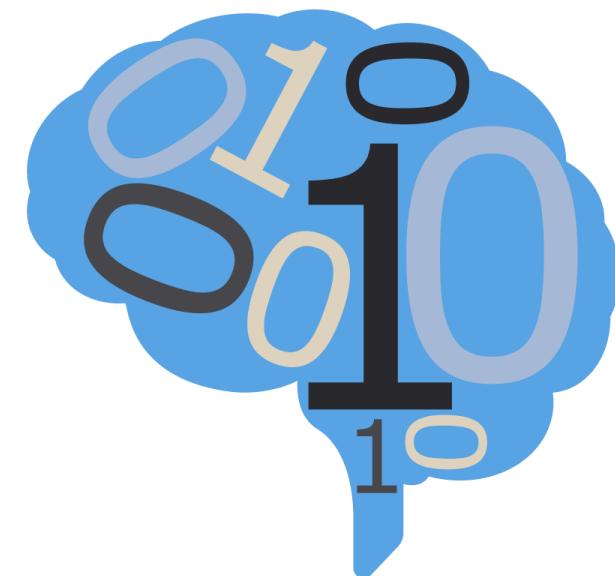


## m2g

downloads 5k

Stars 61

contributors 15

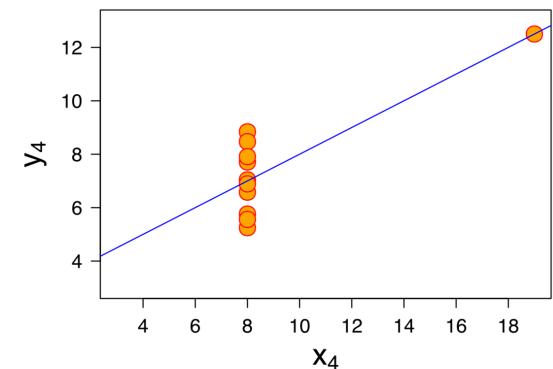
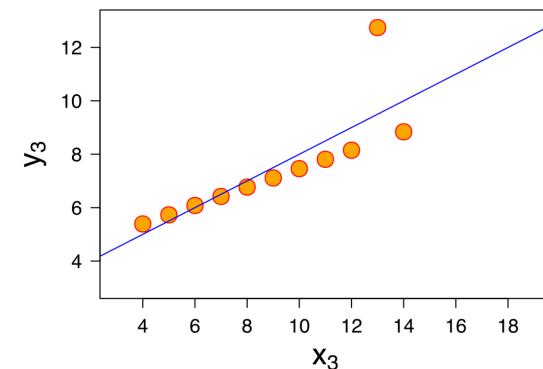
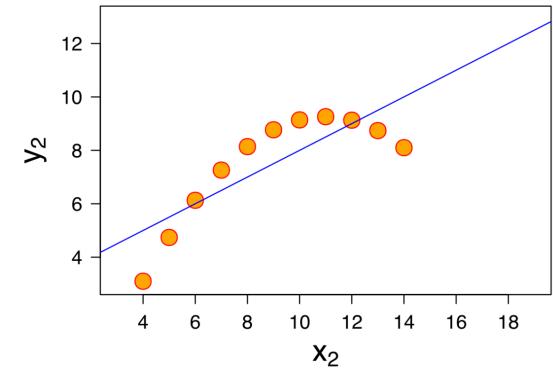
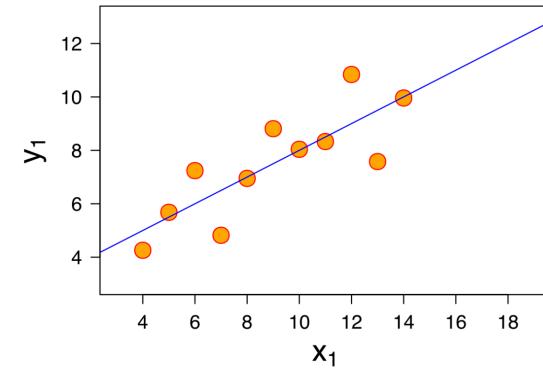


# **Additional Slides**

# Different data, same statistics (Ascombe's Quartet)

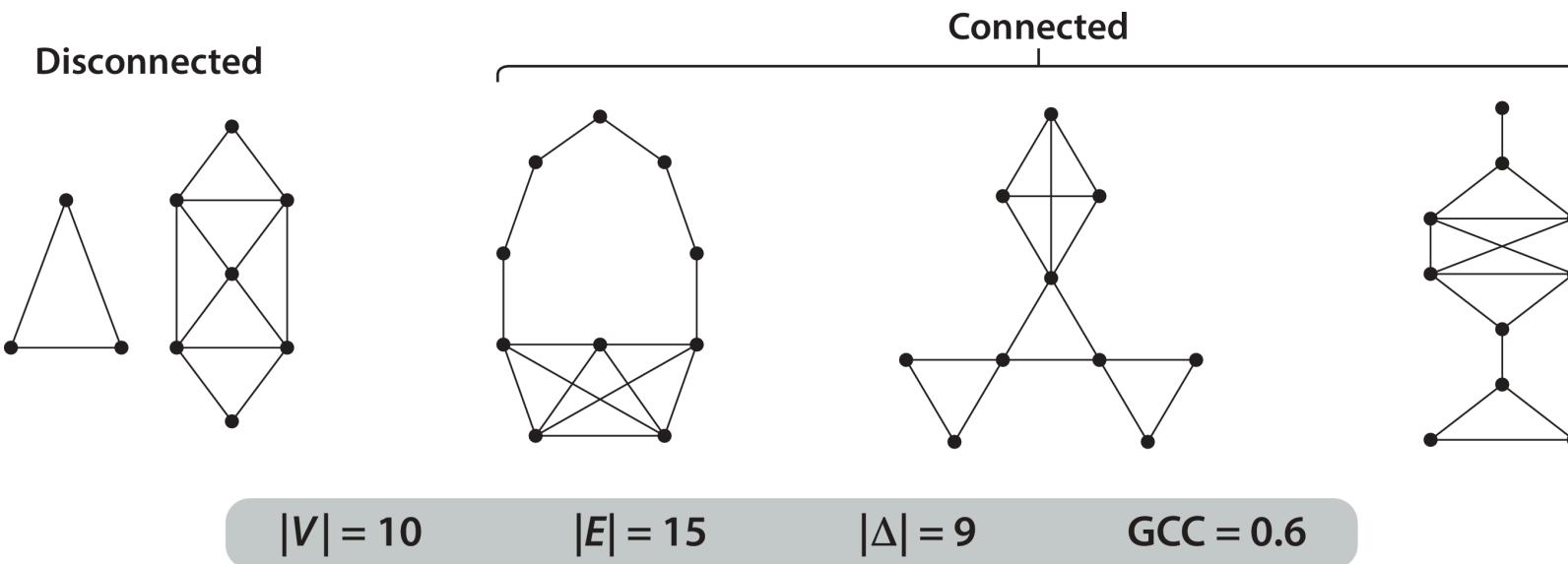
- These four datasets have same statistics!

- Mean ( $\bar{x}$ ): 9
- Variance ( $s_x^2$ ): 11
- Mean ( $\bar{y}$ ): 7.5
- Variance ( $s_y^2$ ): 4.12
- Correlation ( $\rho_{xy}$ ): 0.816



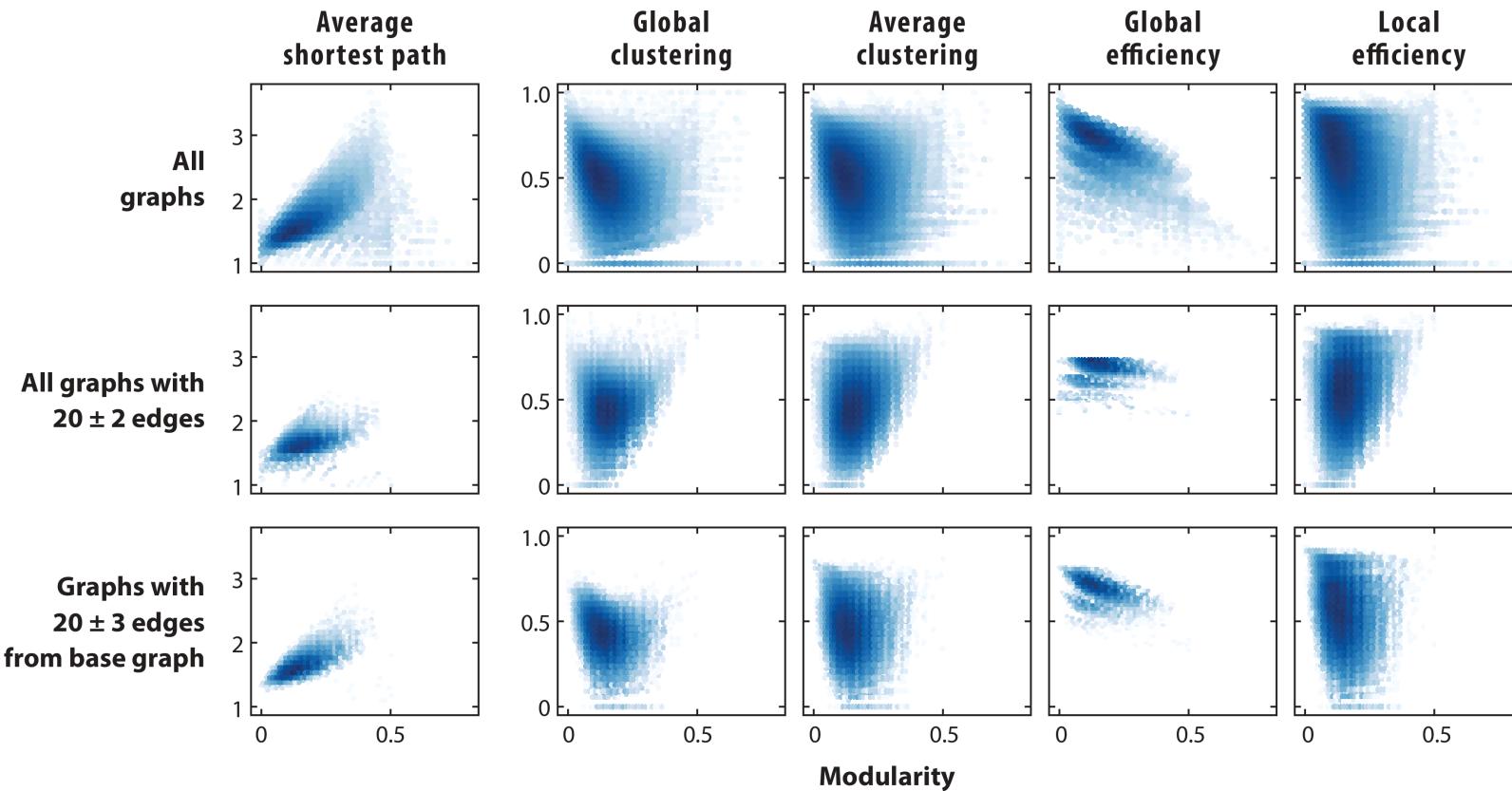
# Different networks, same statistics

- These four networks have same (network) statistics!



# Show the other figure of the same statistics

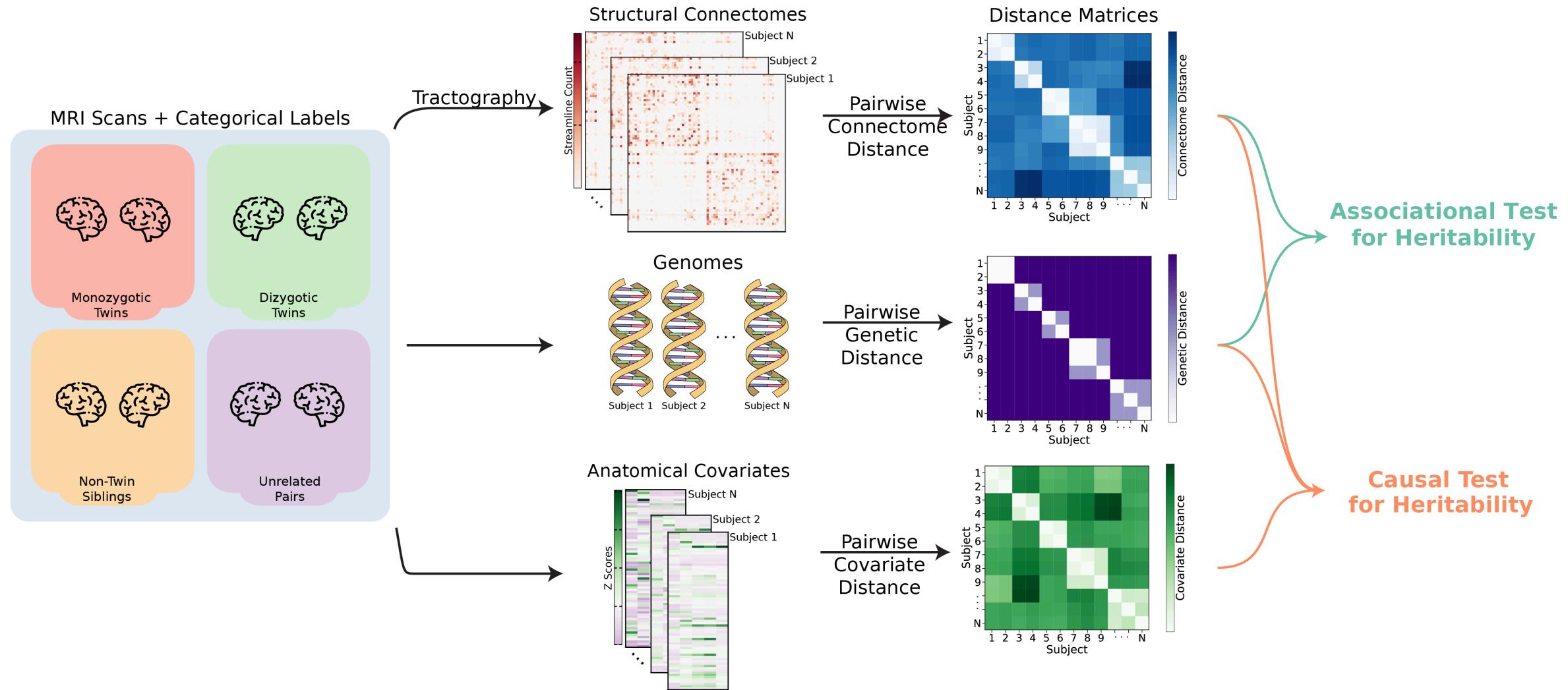
- Consider all non-isomorphic graphs with 10 vertices



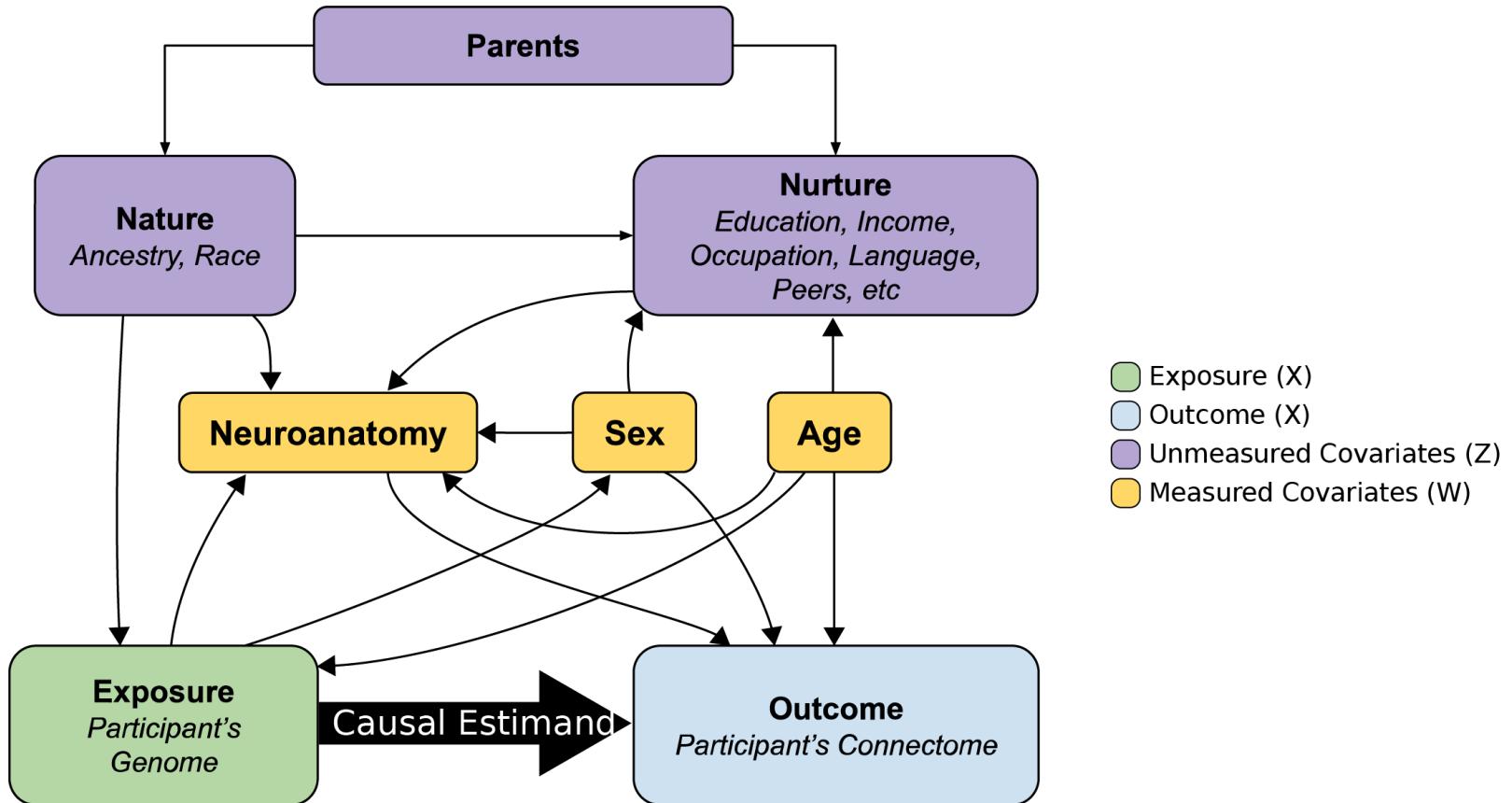
Chung J, et al. 2021  
Annu. Rev. Stat. Appl. 8:463–92

Chung et al. "Statistical connectomics." *Annual Review of Statistics and Its Application* (2021)

# Analysis Framework



# Causal Directed Acyclic Graph



# Causal Estimand

- $X$  denote exposure,  $Y$  denote outcome,  $W$  denote measured covariates,  $Z$  denote unmeasured covariates.
- Want to estimate the effect of different exposures on the outcome, which is quantified using the backdoor formula if  $W$  and  $Z$  close all backdoor paths.

$$f_{w,z}(y|x) = \int_{\mathcal{W} \times \mathcal{Z}} f(y|x, w, z) f(w, z) d(w, z)$$

- Above integrates over *all* measured and unmeasured covariates.

$$f(y|x) = \int_{-\mathcal{W} \times \mathcal{Z}} f(y|x, w, z) f(w, z|x) d(w, z)$$

- Averages the true outcome distribution over the *conditional* distribution of the measured and unmeasured covariates.

# Hypothesis Testing

## Associational Heritability

- We observe the pairs  $(x_i, y_i)$  for  $i \in [n]$ .
- Only be able to estimate the functions of  $(X, Y)$
- The corresponding hypothesis test is:

$$H_0 : f(y|x) = f(y) \quad \text{vs} \quad H_A : f(y|x) \neq f(y).$$

## Causal Heritability

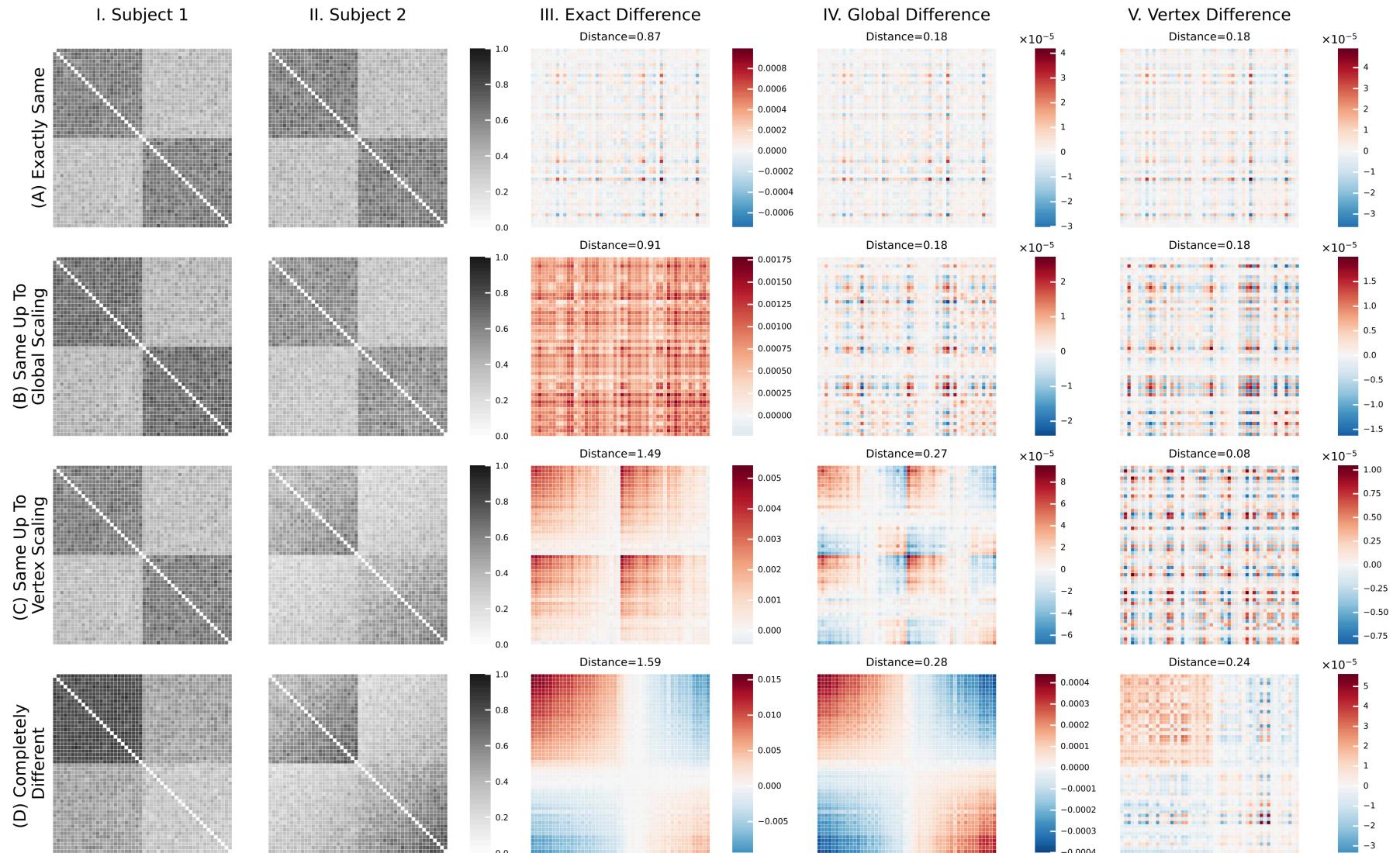
- We observe the triples  $(x_i, y_i, w_i)$  for  $i \in [n]$ .
- Only be able to estimate the functions of  $(X, Y, W)$
- The corresponding hypothesis test is:

$$H_0 : f(y|x, w) = f(y|w) \quad \text{vs} \quad H_A : f(y|x, w) \neq f(y|w).$$

# Connectome Models

- Exact model:** This model measures all differences between latent positions, with differences in the latent positions implying differences in the connectomes themselves.
- Global model:** This model examines whether the latent positions of one connectome are a scaled version of the other. For example, if the number of edges in male connectomes are consistently larger than those in females, we have no way of differentiating whether significant findings from the exact model are a result of differences in scaling or differences in the fundamental structure of the connectomes themselves.
- Vertex model:** This model is similar to the global model, but it allows for each vertex to be scaled differently. The idea behind this approach is that some vertices may have a greater impact on the overall network than others, so scaling them differently can provide a more accurate representation of the network. Consider the examples of how brain regions connect with each other. Regions in the same hemisphere are more likely to be connected than across hemispheres. Even within the same hemisphere, different regions may have distinct preferences for forming connections with other specific regions.

# Connectome Model Simulations



# Shortcomings - Model assumptions

- No interaction between genome and environment
- No epistasis
  - Effect of one gene is dependent on another
  - Ex: black hair and baldness
- No dominance effects
- Strong assumptions in genetic distances