Human Structural Connectomes are Heritable

Jaewon Chung [1]   Benjamin D. Pedigo [1]   Carey E. Priebe [2]   Joshua T. Vogelstein [1]

[1]Department of Biomedical Engineering, Johns Hopkins University   [2]Department of Applied Mathematics and Statistics, Johns Hopkins University

Summary

- Understanding how **brain connectivity is influenced by genetics** can improve our understanding of brain function and diseases.
- Existing methods ignore spatial arrangement of the brain or are not statistically justified.
- Our random graph theory and statistical approach enable formulation and testing of **different models of connectome heritability**.
- We show that brain connectivity is **more similar in monozygotic twins** compared to those of dizigotic

twins and siblings.
- Other frameworks (e.g. joint graph embeddings) can be used for vertex level analysis.

Data and Preprocessing

Human Connectome Project Young Adult study (HCP1200) [4].

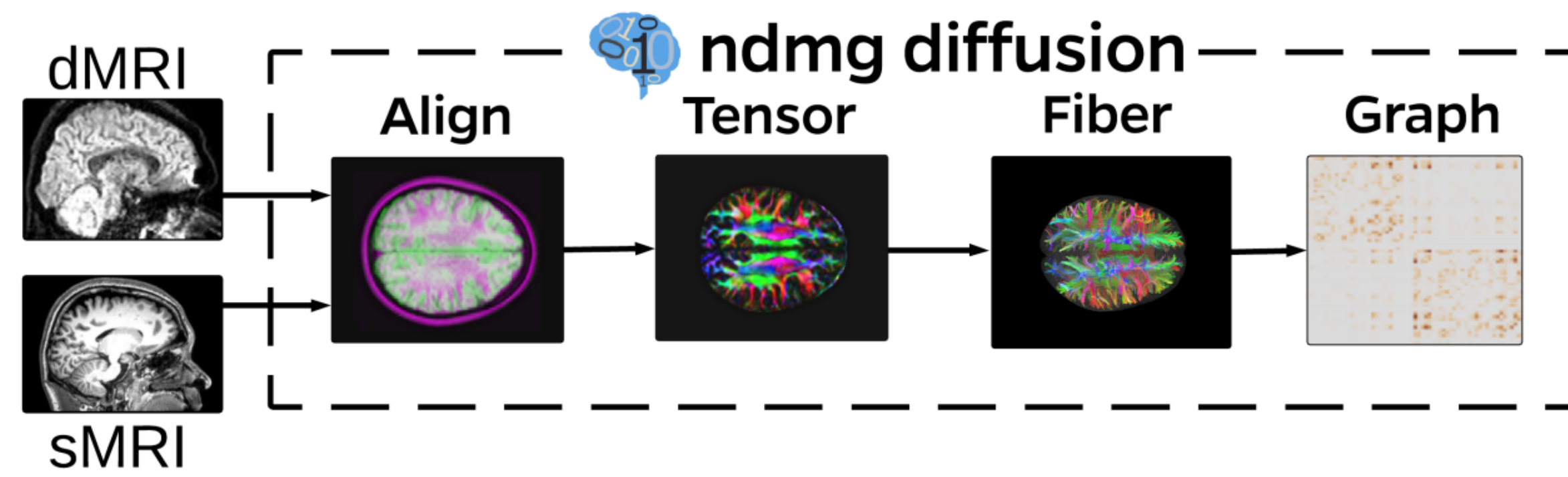| Zygosity	Monozygotic	Dizygotic	Non-Twin siblings
N	250	259	476
Sex	167 F, 83 M	140 F, 119 M	223 F, 248 M
Age	29.6 (3.3)	28.9 (3.4)	28.3 (3.9)
Age range	22-36	22-36	22-37

Table 1. HCP 1200 Participants and their demographics

Figure 1. Outline of the *ndmg* (https://ndmg.neurodata.io) pipeline. Image taken from [2].

Estimating Heritability

$$G, F \in \mathbb{R}^{n \times n} \qquad \hat{X}, \hat{Y} \in \mathbb{R}^{n \times d}$$
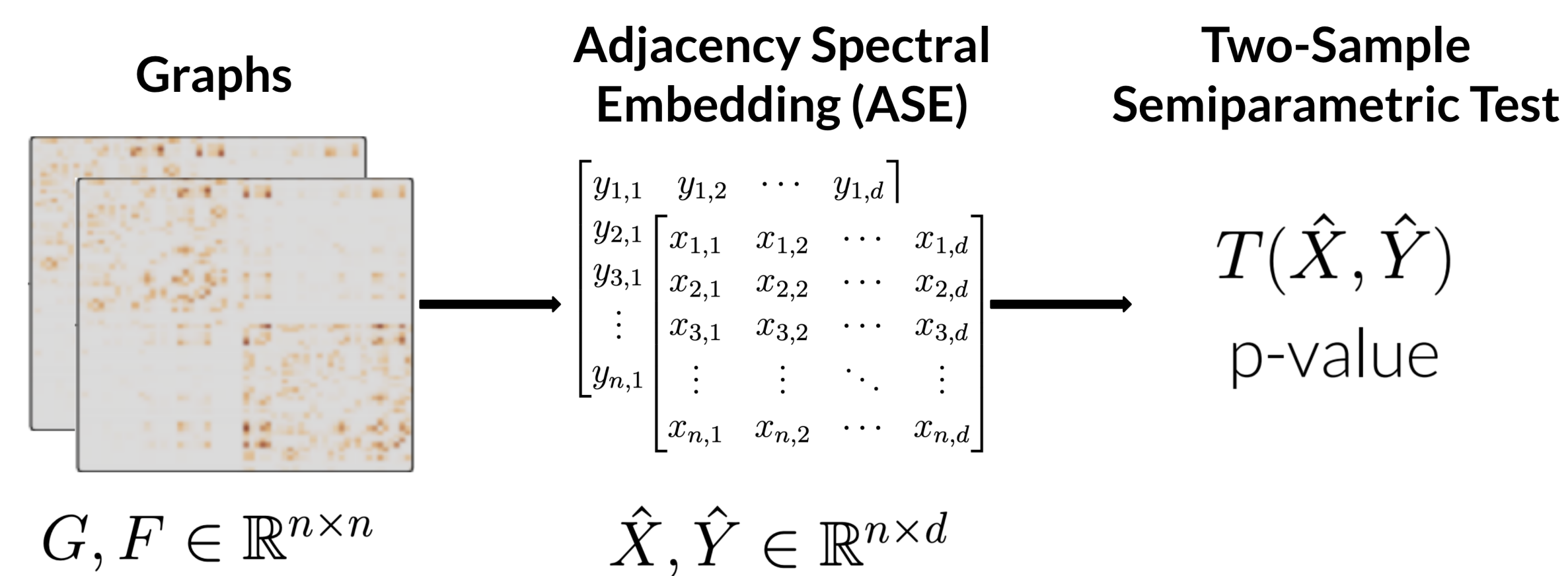
Figure 2. Outline of heritability estimation process.

For semiparametric testing, we consider the three **test cases**:

- Equality up to rotation:
  $H_0 : X = WY$ *against* $H_a : X \neq WY$ where $W \in \mathbb{R}^{n \times n}$ and $WW^T = I$
- Equality up to scaling:
  $H_0 : X = cWY$ for some $c > 0$ *against* $H_a : X \neq cWY$ for any $c > 0$
- Equality up to diagonal transformation:
  $H_0 : X = DWY$ for some diagonal D *against* $H_a : X \neq WDY$ for any D

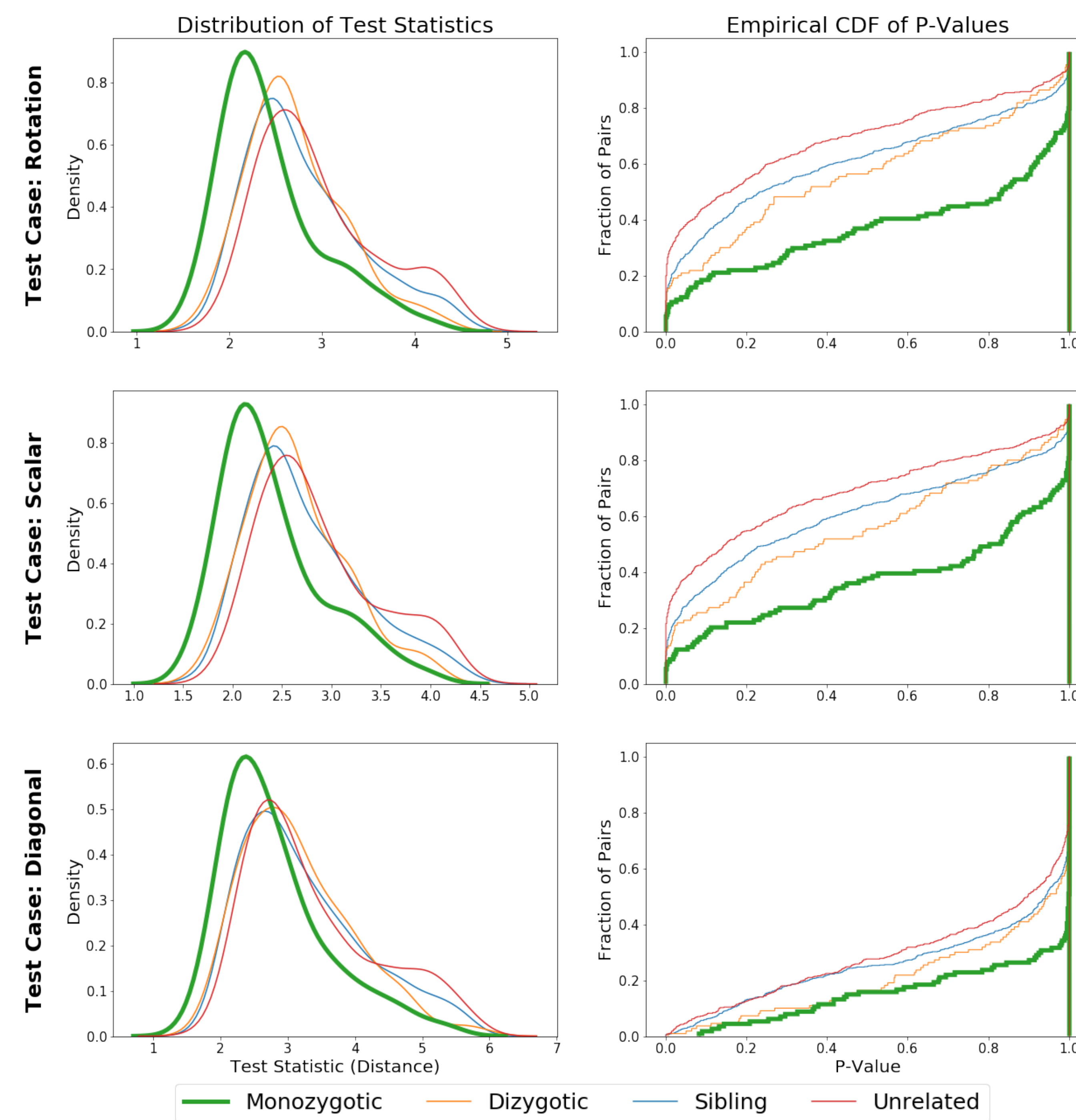## Two-Sample Semiparametric Test Results

Figure 3. Kernel density estimates (KDEs) of the test statistics is shown on the left column, and the empirical cumulative distribution (eCDF) of associated p-value give by null-distribution is shown on the right column.

1. Kernel density estimates (KDEs) for *monozygotic twins* is different from those of *dizygotic twins*, *sibling*, and *unrelated*.
2. Distribution of test statistics for *monozygotic twins* is more positively skewed compared to others.
3. Empirical cumulative distribution function (eCDF) for *monozygotic twins* is more concentrated near 1 compared to others.
4. This empirical finding is validated through six statistical models and tested via two sample Kolmogrov-Smirnoff test (Table 2 and 3).

Two-Sample Kolmogrov-Smirnov (KS) Test Results

Null hypothesis is that the two samples are drawn from the same distribution.

Alternate Hypothesis	Test Cases		
	Rotation	Scalar	Diagonal
Monozygotic < Dizygotic	< .001 ***	< .001 ***	< .001 ***
Monozygotic < Sibling	< .001 ***	< .001 ***	< .001 ***
Monozygotic < Unrelated	< .001 ***	< .001 ***	< .001 ***
Dizygotic ≠ Sibling	0.353	0.337	0.750
Dizygotic ≠ Unrelated	0.009 **	0.009 **	0.169
Sibling ≠ Unrelated	< .001 ***	< .001 ***	0.100

Table 2. P-values for KS tests on **distributions of test statistics** from semiparametric tests.

Alternate Hypothesis	Test Cases		
	Rotation	Scalar	Diagonal
Monozygotic > Dizygotic	< .001 ***	< .001 ***	0.0012 **
Monozygotic > Sibling	< .001 ***	< .001 ***	< .001 ***
Monozygotic > Unrelated	< .001 ***	< .001 ***	< .001 ***
Dizygotic ≠ Sibling	0.120	0.109	0.199
Dizygotic ≠ Unrelated	< .001 ***	< .001 ***	0.154
Sibling ≠Unrelated	< .001 ***	< .001 ***	0.025 *

Table 3. P-values for KS tests on **distributions of p-values** from semiparametric tests.

Significance levels are marked with * ($p < .05$), ** ($p < .01$), and *** ($p < .001$).

What's Next?

1. **Vertex importance** - which vertex confers the most heritability?
2. **Edge importance** - which edges are important?

Code and Data

All analysis was performed using an open-source package **GrasPy** (https://graspy.neurodata.io). All dMRI images are open-source and provided by the Human Connectome Project (https://www.humanconnectome.org/study/hcp-young-adult/overview).

References

[1] Avanti Athreya, Donniell E. Fishkind, Minh Tang, Carey E. Priebe, Youngser Park, Joshua T. Vogelstein, Keith Levin, Vince Lyzinski, Yichen Qin, and Daniel L Sussman.
Statistical inference on random dot product graphs: a survey.
Journal of Machine Learning Research, 18(226):1–92, 2018.

[2] Gregory Kiar, Eric Bridgeford, Will Gray Roncal, , Vikram Chandrashekhar, Disa Mhembere, Sephira Ryman, Xi-Nian Zuo, Daniel S Marguiles, R Cameron Craddock, Carey E Priebe, Rex Jung, Vince Calhoun, Brian Caffo, Randal Burns, Michael P Milham, and Joshua Vogelstein.
A high-throughput pipeline identifies robust connectomes but troublesome variability.
bioRxiv, 2018.

[3] Minh Tang, Avanti Athreya, Daniel L. Sussman, Vince Lyzinski, Youngser Park, and Carey E. Priebe.
A semiparametric two-sample hypothesis testing problem for random graphs.
Journal of Computational and Graphical Statistics, 26(2):344–354, 2017.

[4] David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy EJ Behrens, Essa Yacoub, Kamil Ugurbil, Wu-Minn HCP Consortium, et al.
The wu-minn human connectome project: an overview.
Neuroimage, 80:62–79, 2013.