

Heritability of Large-Scale Functional Brain Networks *via* Persistent Homology

Moo K. Chung¹, Victoria Vilalta-Gil², Paul J. Rathouz¹, Benjamin B. Lahey³, David H. Zald²

¹University of Wisconsin-Madison ²Vanderbilt University, ³University of Chicago

email: mkchung@wisc.edu

Introduction

The extent to which heritability influences functional brain networks is not clearly established. In this study, we propose to map the heritability of a large-scale brain networks at the voxel-level by taking each voxel as a network node.

Twin fMRI Dataset

The study consists of 11 monozygotic (MZ) and 9 same-sex dizygotic (DZ) twin pairs of 3T functional magnetic resonance images (fMRI). Subjects completed the *monetary incentive delay task* (Knutson et al., 2001) of 3 runs of 40 trials consisting of \$0, \$1 and \$5 rewards, which are randomly split into 3 runs. After fitting a general linear model at each voxel, the contrast map of testing the significance of delay for \$5 trials to the baseline (\$0 reward) is obtained (Figure 1). The contrast maps are used in constructing the functional network.

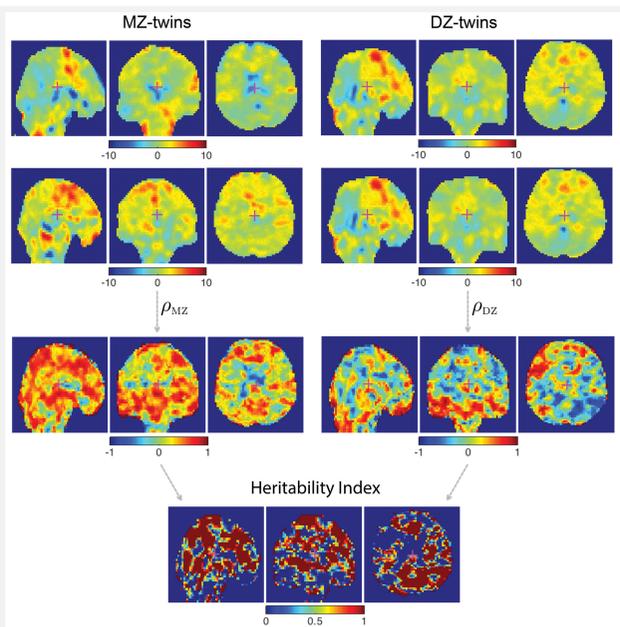


Figure 1. Representative MZ- and DZ-twin pairs of the contrast maps obtained from the first level analysis testing the significance of the delay in hitting the response button in \$5 reward in contrast to \$0 reward. Middle: The correlation of the contrast maps within twins. The heritability index map measures the difference between the correlations.

Sparse Cross-Correlations

Let $V = \{v_1, \dots, v_p\}$ be a node set with $p = 25672$ nodes in the template. Let $x_k(v_i)$ and $y_k(v_i)$ be the k -th paired image intensity values at voxel v_i . Denote $\mathbf{x}(v) = (x_1, \dots, x_n)'$ and $\mathbf{y}(v) = (y_1, \dots, y_n)'$. Center \mathbf{x} and \mathbf{y} such that

$$\sum_{k=1}^n x_k(v_i) = \sum_{k=1}^n y_k(v_i) = 0,$$

$$\|\mathbf{x}(v_i)\|^2 = \mathbf{x}'(v_i)\mathbf{x}(v_i) = \|\mathbf{y}(v_i)\|^2 = \mathbf{y}'(v_i)\mathbf{y}(v_i) = 1.$$

Set up a linear model between $\mathbf{x}(v_i)$ and $\mathbf{y}(v_j)$:

$$\mathbf{y}(v_j) = \beta_{ij} \mathbf{x}(v_i) + \mathbf{e},$$

where \mathbf{e} is the zero-mean error vector. The connectivity strength $\beta = (\beta_{ij})$ is estimated as

$$\hat{\beta}_{ij} = \mathbf{x}'(v_i)\mathbf{y}(v_j) = \arg \min_{\beta_{ij}} \sum_{i=1}^p \sum_{j=1}^p \|\mathbf{x}(v_i) - \beta_{ij} \mathbf{y}(v_j)\|^2, \quad (1)$$

which are the cross-correlations. The cross-correlation matrix (β_{ij}) requires computing 1.3 billion entries for each twin group and 5.2GB of computer memory per matrix (Figure 2). Our formulation takes less than 18 seconds in computing such matrix in a desktop computer. Since the cross-correlation matrices are very dense, it is difficult to visualize and provide biological interpretation. It is a necessity to reduce the number of connections by sparsifying cross-correlations.

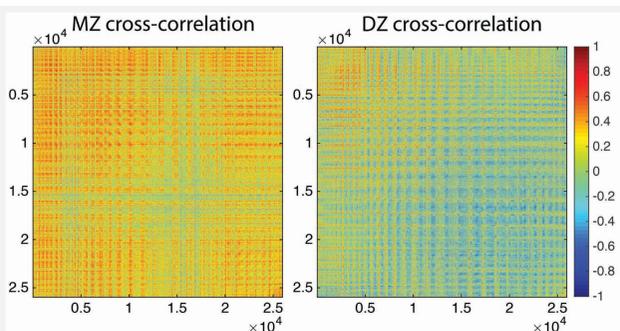


Figure 2. Cross-correlations for MZ- and DZ-twins for all 25972 voxels. Since the correlation matrices are too dense to visualize and interpret, it is necessary to reduce the filter out number of connections. The diagonal entries are correlations.

The sparse version of (1) is given by

$$F(\beta; \mathbf{x}, \mathbf{y}) = \frac{1}{2} \sum_{i=1}^p \sum_{j=1}^p \|\mathbf{x}(v_i) - \beta_{ij} \mathbf{y}(v_j)\|^2 + \lambda \sum_{i=1}^p \sum_{j=1}^p |\beta_{ij}|. \quad (2)$$

The *sparse cross-correlation* is then given by

$$\hat{\beta} = \arg \min_{\beta} F(\beta; \mathbf{x}, \mathbf{y}).$$

The direct numerical optimization of (2) for large p is computationally demanding. It can be shown that the minimization can be done algebraically (Chung et al., 2016) (Figure 3):

Theorem 1. For $\lambda \geq 0$, the minimizer of $F(\beta; \mathbf{x}, \mathbf{y})$ is given by

$$\hat{\beta}_{ij}(\lambda) = \begin{cases} \mathbf{x}'(v_i)\mathbf{y}(v_j) - \lambda & \text{if } \mathbf{x}'(v_i)\mathbf{y}(v_j) > \lambda \\ 0 & \text{if } |\mathbf{x}'(v_i)\mathbf{y}(v_j)| \leq \lambda \\ \mathbf{x}'(v_i)\mathbf{y}(v_j) + \lambda & \text{if } \mathbf{x}'(v_i)\mathbf{y}(v_j) < -\lambda \end{cases}$$

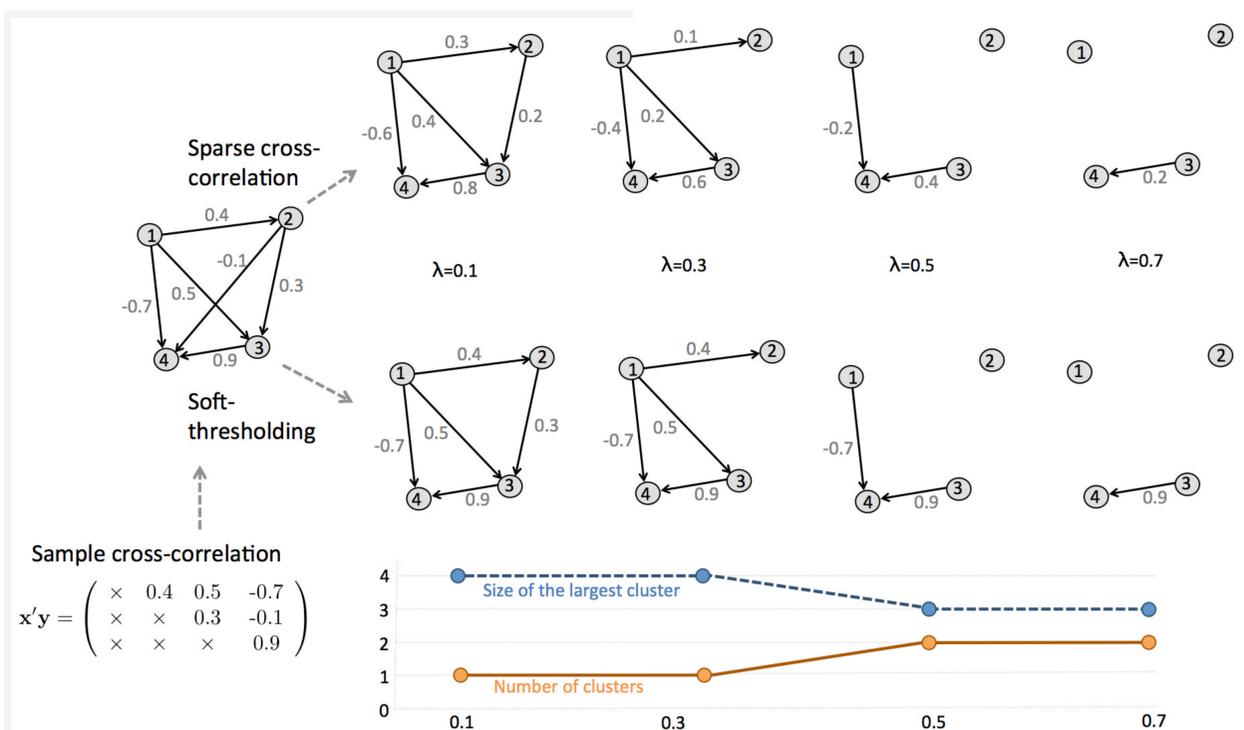


Figure 3. Schematic showing the equivalence of constructed binary graphs using existing LASSO (top) and the proposed soft-thresholding method (bottom) on the sample cross-correlations. Any edge whose absolute value is smaller than the sparse parameter λ is deleted. The number of clusters (#) and the size of the largest cluster (&) are tabulated over λ .

Graph Filtration

The adjacency matrix $A(\lambda) = (a_{jk})$ of sparse graph $\mathcal{G}_A(\lambda)$ is defined as

$$a_{jk}(\lambda) = \begin{cases} 1 & \text{if } \hat{\beta}_{jk} \neq 0; \\ 0 & \text{otherwise.} \end{cases}$$

Instead of determining the optimal parameter λ that may not be optimal in other studies and datasets, we will analyze $\mathcal{G}_A(\lambda)$ for every possible λ using *persistent homology*. We can show that

$$\mathcal{G}_A(\lambda_1) \supset \mathcal{G}_A(\lambda_2) \supset \mathcal{G}_A(\lambda_3) \supset \dots \text{ for any } 0 \leq \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \dots$$

This type of nested graphs is called *graph filtration*, a special case of Rips filtrations (Lee et al., 2012; Chung et al., 2015). We can show that $\mathcal{G}_A(\lambda)$ can be equivalently obtained by (Figure 3)

Theorem 2. (Soft-thresholding) Consider a binary graph $\mathcal{G}_B(\lambda)$ with the adjacency matrix $B = (b_{ij})$ given by

$$b_{jk}(\lambda) = \begin{cases} 1 & \text{if } |\mathbf{x}'(v_i)\mathbf{y}(v_j)| > \lambda; \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Then $\mathcal{G}_A(\lambda) = \mathcal{G}_B(\lambda)$ for all λ .

Many graph functions computed on graph filtrations have monotonic properties. Consider the number of disjoint clusters (#) and the size of the largest cluster (&).

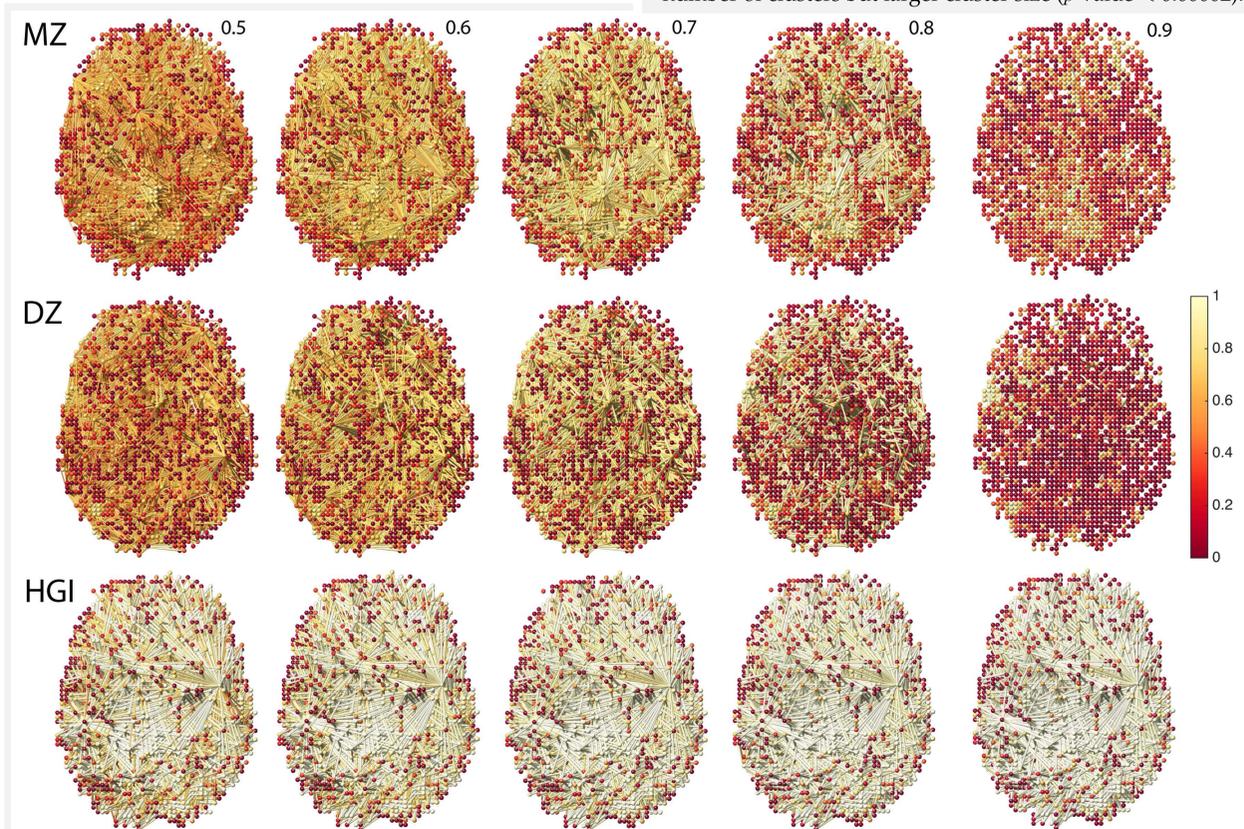


Figure 4. Top and Middle: Node colors are correlations of MZ- and DZ-twins. Edge colors are sparse cross-correlations at given sparsity λ . Bottom: node colors are heritability indices (HI) and edge colors are heritability graph index (HGI). Some low HI nodes show high HGI. Using only the voxel-level HI feature will fail to detect such subtle genetic effects on the functional network.

Heritability Graph Index

Heritability index (HI) determines the amount of variation due to genetic influence and estimated using Falconer's formula (Falconer and Mackay, 1995). At each voxel v_i , HI is defined as

$$HI_i = 2[\rho_{MZ}(v_i) - \rho_{DZ}(v_i)],$$

where ρ_{MZ} and ρ_{DZ} are the pairwise correlation within MZ- and same-sex DZ-twins. HI is a univariate feature that does not quantify if the change in one voxel is related to other voxels. We extend the concept of HI to the network level by defining the *heritability graph index* (HGI) (Figure 4):

$$HGI_{ij} = 2[\varrho_{MZ}(v_i, v_j) - \varrho_{DZ}(v_i, v_j)],$$

where ϱ_{MZ} and ϱ_{DZ} are the cross-correlations between voxels v_i and v_j within MZ- and DZ-twin pairs. The statistical significance of HGI is determined using Theorem 3 (Figure 5).

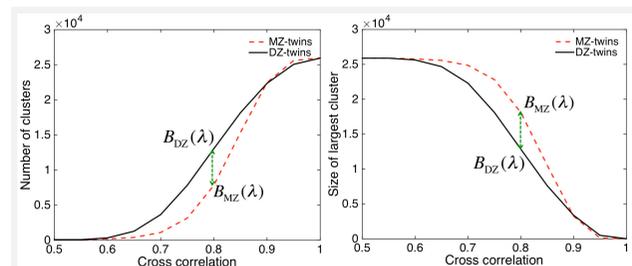


Figure 5. Inference of graph filtrations of MZ- and DZ-twins. The number of clusters (left) and the size of largest cluster (right) are plotted over the sparse parameter λ . MZ-twins have smaller number of clusters but larger cluster size (p -value < 0.00002).

References

- M.K. Chung, J.L. Hanson, Ye J., Davidson R.J., and Pollak S.D. Persistent homology in sparse regression and its application to brain morphometry. *IEEE Transactions on Medical Imaging*, 34: 1928–1939, 2015.
- M.K. Chung, V.G. Vilalta, P.J. Rathouz, B.B. Lahey, and D.H. Zald. Linear embedding of large-scale brain networks for twin fMRI. *arXiv preprint arXiv:1509.04771*, 2016.
- D. Falconer and T Mackay. *Introduction to Quantitative Genetics*, 4th ed. Longman, 1995.
- B. Knutson, C.M. Adams, G.W. Fong, and D. Hommer. Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *Journal of Neuroscience*, 21:RC159, 2001.
- H. Lee, H. Kang, M.K. Chung, B.-N. Kim, and D.S Lee. Persistent brain network homology from the perspective of dendrogram. *IEEE Transactions on Medical Imaging*, 31:2267–2277, 2012.

$$\# \mathcal{G}_A(\lambda_1) \leq \# \mathcal{G}_A(\lambda_2) \leq \# \mathcal{G}_A(\lambda_3) \leq \dots, \\ \& \mathcal{G}_A(\lambda_1) \geq \& \mathcal{G}_A(\lambda_2) \geq \& \mathcal{G}_A(\lambda_3) \geq \dots$$

Such monotonic features can be quantified using the test statistic (Chung et al., 2016)

$$D_p = \sup_{\lambda} |B_1(\lambda) - B_2(\lambda)|,$$

where B_i is the monotonic features of the i -th group (Figure 5). The p -value for test statistic D_p is determined from

$$\text{Theorem 3. } \lim_{p \rightarrow \infty} P(D_p / \sqrt{2(p-1)} \leq d) = 1 - 2 \sum_{i=1}^{\infty} (-1)^{i-1} e^{-2i^2 d^2}.$$