INVESTIGATING HERITABILITY ACROSS RESTING STATE BRAIN NETWORKS VIA HEAT KERNEL SMOOTHING ON PERSISTENCE DIAGRAMS

Arman P. Kulkarni[1], *Moo K. Chung*[1], *Barbara B. Bendlin*[1], *Vivek Prabhakaran*[1]

[1]University of Wisconsin, Madison, USA

ABSTRACT

The brain's heritable topological differences in resting state functional connectivity (rsfc) measured via resting state fMRI (rsfMRI) provide important insight into brain function and dysfunction. Current techniques investigating heritability are limited by arbitrary rsfc threshold selection and reduction of otherwise detailed brain topological properties into summary measures. Topological Data Analysis (TDA) is a novel tool for addressing these limitations by analyzing how the topological properties of data vary without arbitrary threshold and summary metric construction. TDA applies a filtration to the data and constructs a persistence diagram (PD). Therefore, the purpose of this study was to compute PDs to determine TDA-based heritability of static brain network topological features. To this end, we calculated a robust heritability index map across smoothed PDs derived from twin rsfMRI data.

***Index Terms*—** topological data analysis, heat kernel, twin study, heritability, transpositions, resting state fMRI

1. INTRODUCTION

The role of genetic factors on brain disease and behavior has great implications for networks worthy of targeted intervention and prognostic outcomes. Studies have elucidated brain resting state functional connectivity (rsfc), measured via resting state fMRI (rsfMRI), and related resting state networks as potential endophenotypes of disease and disorder [1]. One criterion essential for defining an endophenotype is heritability, which is calculated by leveraging monozygotic (MZ) and dizygotic (DZ) twin pairs [1]. Given MZ twins and same-sex DZ twins share 100% and 50% of genes, respectively, Falconer's Model decomposes the variance due to genetic influence in a population based on these twin correlations, and computes a heritability index (HI) [2].

However, current techniques investigating heritability are limited by either 1) a variety of thresholding values and summary metrics in graph theory analysis, or 2) performing analyses directly on collapsed spatial regions of the brain [3, 4, 5]. The choice of varying thresholds is arbitrary in nature, and may lead to differences in reproducibility, and summary measures may not elucidate the full detailed picture of the brain's topological properties [3].

Notably, Topological Data Analysis (TDA) addresses these concerns by applying a filtration to the data and constructing a persistence diagram (PD) [6]. One such TDA filtration, a graph filtration, overcomes the inherent limitation of arbitrary threshold values by applying a multi-threshold transformation from singular graph nodes to features based on all possible threshold values [5, 6]. Furthermore, TDA is robust to noise, and has been applied to determine how topological properties vary in brain studies [5, 6]. Therefore, TDA provides the robust framework needed for more accurate estimates of heritability than otherwise possible.

Nevertheless, PDs have proven difficult to compare across subjects due to the lack of correspondence between features. Such quantification is made possible by heat kernel smoothing the PDs [7]. Moreover, by utilizing a novel adaptation of permutations via transpositions in calculating twin correlations, more robust HI map estimates may be derived from these smoothed PDs when leveraging Falconer's Model [8].

Therefore, the purpose of this study was to compute TDA-based heritability of static brain network topological features by utilizing the Human Connectome Project's (HCP) data [9]. The corresponding major contributions are thus: 1) robust PD-based HI map measures calculated by leveraging TDA properties, 2) overcoming the subject-level feature correspondence problem across PDs by employing heat kernel smoothing, and 3) more robust HI estimates from transpositions within twin pairs than typical permutation tests allow.

2. METHODS & APPLICATION

2.1. Overview & Pipeline

The pipeline in Fig.1 was followed to obtain an HI map of TDA-derived features. Each of the following steps will be explained in detail: a) acquisition and preprocessing of rsfMRI signal, b) rsfc correlation matrices derived from rsfMRI signal, c) PDs produced via graph filtration on the rsfc data, d) heat kernel smoothing applied to each PD to reduce topological noise and match PDs across subjects, e) twin pair zygosity based correlation matrices from the resultant smoothed PDs, and f) an HI map derived from the PD correlation matrices.

Correspondence should be sent to Arman Kulkarni (email: apkulkarni2@wisc.edu). For funding, see acknowledgements section.

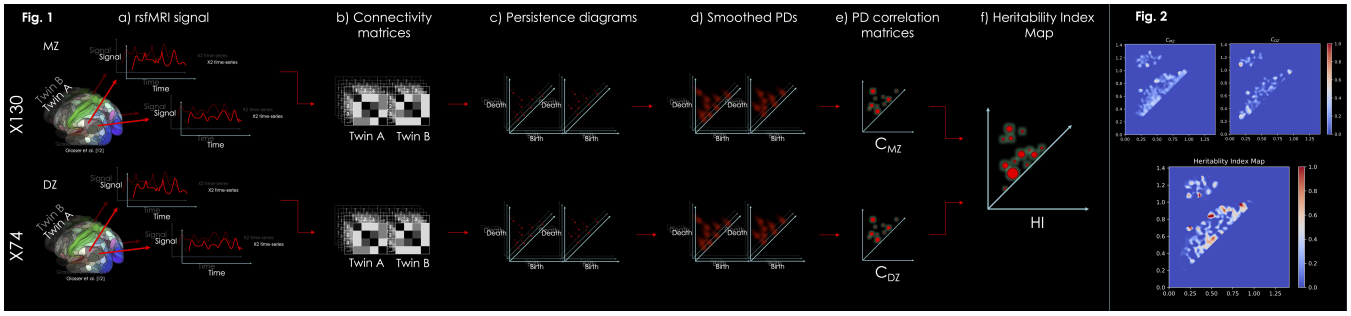

Fig. 1. Step-by-step pipeline (a-f) of the proposed method to determine HI via heat kernel smoothing on PDs based on rsfMRI data. **Fig. 2**. HI map identifying the most heritable clusters (red) with associated C_{MZ} and C_{DZ} maps (fixed range=[0,1]). The C_{MZ} and C_{DZ} maps were computed after a total of 1.2 million transpositions and 200 permutations.

2.2. Data Collection & Preprocessing

The HCP is well known for its collection of rsfMRI scans in young healthy adults, including MZ and DZ twins [9]. The HCP's rsfMRI data were acquired using a multiband (72 slice, 8 band) gradient-echo echo-planar imaging (EPI) sequence (simultaneous multi-slice, TR=720ms, TE=33.1ms, FOV=208x180mm (ROxPE), resolution=2.0mm isotropic, flip angle=52 degrees) [9]. From these data, 130 MZ twin pairs' (56M, 29.2±3.3yrs) and 74 DZ twin pairs' (30M, 29.1±3.5yrs) rsfMRI scans were leveraged. Gender, education and age were all matched across twin groups ($p > 0.15$).

Two scans (runs) for each twin, acquired with right-to-left and left-to-right phase encoding directions across the first session, were selected. Each scan underwent both the HCP minimal preprocessing pipeline as well as Independent-component-analysis-based X-noisifier (ICA-FIX) [10, 11]. The minimal preprocessing pipeline notably removed spatial distortions, realigned subject volumes to account for motion, and registered the fMRI volumes to the anatomical MRI, while ICA-FIX removed spatially specific structured noise inclusive of cardiac, motion and respiratory effects [10, 11].

Each scan was then demeaned and normalized by the standard deviation of each run to account for bias across runs. Subsequently, to average brain regions based on functionally and anatomically suitable criteria, both HCP's Glasser Parcellation, consisting of 360 parcels (regions), and the subcortical FreeSurfer parcellation scheme, consisting of 19 parcels, for a total of 379 parcels, were employed (subsequently shown in 2.3) [12, 13]. The resulting two parcellated scans were concatenated for every twin to increase the reliability of calculated rsfc estimates (Fig.1a) [14].

2.3. Connectivity Matrices & Graph Filtration

Resting state functional connectivity correlation matrices were subsequently computed across each parcellated time series, $p_i(t) = \boldsymbol{z}_i$ for each twin as $\rho_{ij} = corr(p_i(t), p_j(t))$, for the aforementioned 379 (parcels) nodes i and j (Fig.1b). Each correlation matrix was transformed into the distance metric $w_{ij} = \sqrt{1-\rho_{ij}}$ to satisfy the triangle inequality, $w_{ij} \leq w_{ik} + w_{kj}$, and form the metric space with weighted network $X = (V, w_{ij})$, node set $V = \{1, 2..., n = 379\}$ and edge weight $w = (w_{ij})$ [6]. An upward graph filtration was then applied to each twin's transformed data, and PDs were constructed (Fig.1c) with binary network, $X_\epsilon = (V, w_\epsilon)$, where the binary edge weight w_ϵ was determined by

$$w_\epsilon = (w_{ij,\epsilon}) = \begin{cases} 1 & \text{if } w_{ij} < \epsilon \\ 0 & \text{otherwise.} \end{cases}$$

PDs demonstrate the births and deaths of connected components composed of individual nodes. The first occurrence (birth) of the connected component to when it ceases to exist (death) for a given filtration determines features found in PDs [6]. For construction of the PDs, any identical edge weights were offset by incrementally small random perturbations to enable discrete sorting. All resulting PD values were interpolated onto a matrix $G = (g_{rs})$ of rectangular grid size (300,300) per twin with pixel coordinates r and s. This interpolation was performed to further process the data.

2.4. Heat Kernel Smoothing

In order to smooth out topological noise and match PDs across twins, heat kernel smoothing was applied (Fig.1d). The resulting PD $G(p)$ was defined on the upper triangle above the line $y = x$ [7]. Based on the eigensystem containing the Laplace-Beltrami operator Δ defined on the upper triangle, the expression $\Delta\psi_j = -\lambda_j\psi_j$ was then solved for the Laplace-Beltrami operator such that the eigenvalues $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq ...$ and corresponding eigenfunctions $\psi_0, \psi_1, \psi_2, ...$ were ordered from smallest to largest. Subsequently, the Laplace-Beltrami eigenvalues λ_j and eigenfunctions denoted as $\psi_j(p)$ and $\psi_j(q)$ were used to construct the heat kernel

$$K_\sigma(p, q) = \sum_{j=0}^{\infty} e^{-\lambda_j\sigma}\psi_j(p)\psi_j(q),$$

where σ denotes the kernel bandwidth. Therefore it follows that the heat kernel smoothing operation applied on the predefined PD is given by

$$G_s(p) = K_\sigma(p) * G(p) = \sum_{j=0}^{\infty} e^{-\lambda_j \sigma} \beta_j \psi_j(p),$$

where $\beta_j = \langle G, \psi_j \rangle$ represents Fourier coefficients estimated using the least squares method. In this study, bandwidth $\sigma = 10$ and expansion degree 300 were used to smooth each PD.

2.5. PD Correlation Matrices & HI Map

A twin-specific rapid permutation procedure via transpositions was applied to the resultant smoothed PDs to calculate twin correlations within zygosity, and to ultimately compute heritability by using Falconer's Formula [2, 8]. Given the arbitrary ordering of twins in pairing twins, twin correlations should be computed over as many possible permutations within twins. This transposition method has constant run time which allows for more permutations than the traditional permutation technique [8].

2.5.1. Transpositions Within Zygosity Per Twin Pair

Consider the original twin data at a given pixel of the smoothed PD per zygosity as

$$\boldsymbol{x} = \begin{pmatrix} x_{11}, ..., x_{m1} \\ x_{12}, ..., x_{m1} \end{pmatrix},$$

where $x_i = (x_{i1}, x_{i2})^T$ represents the i-th twin pair's data across MZ twins. Let $\boldsymbol{x}_k$ be the k-th row vector of $\boldsymbol{x}$. Specifically, by denoting $\boldsymbol{x}_k^{MZ} = (x_{1k}, x_{2k}, ..., x_{mk})$, the MZ correlations after a transposition can be calculated as follows. Let τ_i be the transposition within the i-th MZ twin pair such that the k-th row transposition, $\tau_i(\boldsymbol{x}_k)$ which exchanges data across the i-th twin pair, can be used to calculate a resulting permuted correlation among MZ twins:

$$\tau_i(\boldsymbol{x}_1) = (x_{11}...x_{i-1,1}, x_{i2}, x_{i+1,1}, ...x_{m1})$$
$$\tau_i(\boldsymbol{x}_2) = (x_{12}...x_{i-1,2}, x_{i1}, x_{i+1,2}, ...x_{m2})$$

Then for the resulting k-th row transposition $\tau_i(\boldsymbol{x}_k)$, let the respective unnormalized mean across the row be denoted as

$$\nu(\boldsymbol{x}_k) = \sum_{l=1}^{m} x_{lk}$$

and the respective unnormalized row covariance follows as

$$\omega(\boldsymbol{x}_k, \boldsymbol{x}_l) = \sum_{r=1}^{m} (x_{rk} - \nu(\boldsymbol{x}_k)/m)(x_{rl} - \nu(\boldsymbol{x}_l)/m)$$

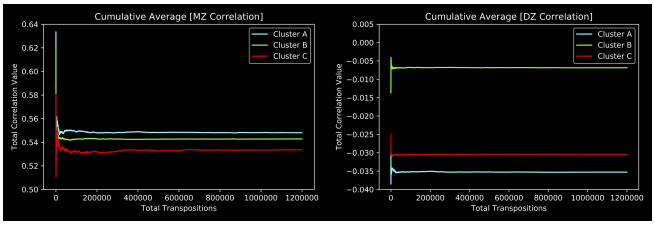

Fig. 3. Convergence of select pixels of MZ and DZ correlations values from three distinct clusters of highest HI (HI≥1). For every 6000 transpositions that occurred, one permutation was intermixed.

Subsequently, ν and ω are updated over the transposition as

$$\nu(\tau_i(\boldsymbol{x}_k)) = \nu(\boldsymbol{x}_k) + x_{il} - x_{ik}$$
$$\omega(\tau_i(\boldsymbol{x}_k), \tau_i(\boldsymbol{x}_l)) = \omega(\boldsymbol{x}_k, \boldsymbol{x}_l) + x_{il}^2 - x_{ik}^2$$
$$+(\nu(\boldsymbol{x}_k)\nu(\boldsymbol{x}_l) - \nu(\tau_i(\boldsymbol{x}_k))\nu(\tau_i(\boldsymbol{x}_l)))/m$$

The MZ twin correlation after the transposition is then

$$\rho(\tau_i(\boldsymbol{x}_1), \tau_i(\boldsymbol{x}_2)) = \frac{\omega(\tau_i(\boldsymbol{x}_1), \tau_i(\boldsymbol{x}_2))}{\sqrt{\omega(\tau_i(\boldsymbol{x}_1), \tau_i(\boldsymbol{x}_1))\omega(\tau_i(\boldsymbol{x}_2), \tau_i(\boldsymbol{x}_2))}}$$

The same concept can be extended to DZ twin correlations via the transposition $\tau_i(\boldsymbol{y}_k)$ across DZ twin pair vector $\boldsymbol{y}_i$.

2.5.2. Computation of PD Correlation Matrices & HI Map

Therefore, the MZ and DZ correlation maps were respectively computed from MZ and DZ twin pairs' smoothed PDs as

$$C_{MZ} = (g_{lm}^{MZ}) = \rho_{lm}(\tau_i(\boldsymbol{x}_1), \tau_i(\boldsymbol{x}_2))$$
$$C_{DZ} = (g_{lm}^{DZ}) = \rho_{lm}(\tau_i(\boldsymbol{y}_1), \tau_i(\boldsymbol{y}_2))$$

where l and m denote the PD pixel location for $\boldsymbol{x}_k$ and $\boldsymbol{y}_k$, the k-th row twin vectors to undergo transpositions. Based on the proposed method, 1.2 million MZ and DZ twin correlations were computed from 200 iterations of one permutation followed by 6000 sequential transpositions and averaged (with undefined and infinite correlations set to zero). The average outputs resulted in the stable MZ-twin and DZ-twin correlation maps, C_{MZ}, and C_{DZ}, respectively (Fig.1e, Fig.2). Finally, the HI map at the matrix level was computed as

$$HI = 2(C_{MZ} - C_{DZ})$$

based on Falconer's formula (Fig.1f) [2]. The robust HI estimate consisted of 1.2 million transpositions and 200 permutations from the PD correlation matrices, C_{MZ} and C_{DZ}. The resulting thresholded HI map (with values less than zero set to zero and values greater than one set to one) is seen in Fig.2. Convergence of several high C_{MZ} and corresponding C_{DZ} pixel values related to high HI clusters is shown in Fig.3.

3. DISCUSSION

This study presented a pipeline to investigate HI on smoothed PDs derived from a graph filtration across rsfc. The analysis found certain clusters of smoothed connected components heritable from the heat kernel smoothed PDs. The average HI map of TDA-based smoothed features serves as a robust estimate of the overlying heritability from 1.2 million transpositions.

While further analysis is required to validate rsfc heritability at the brain network level, this study takes advantage of TDA-based multi-scale properties which may not be inherent in other techniques. Additionally, it serves as one of the few studies to investigate PDs across subjects, accounting for the difficulty in analysis of subject-level variability of PD features. Finally, it implements the transposition method of permutations across twin data, which may be used in lieu of intra-class correlation used in twin studies.

These findings may serve as further points of interest in 1) backprojection of the data for interpretation of heritable brain networks, and 2) extension as a novel and robust means for investigating heritability across dynamic rsfc. The overall HI map may be backprojected to robustly determine the most heritable rsfc patterns. In doing so, it may be possible to identify TDA as providing potential endophenotypic information for rsfMRI, and to extend this information to specific brain networks. Additional parameter investigation of the heat kernel bandwidth should be performed to determine any impact on the result. This question is left as a future study.

Acknowledgements Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. The research was supported under NIH award TL1 TR002375, as well as NIH grants R01 EB022856, R01 EB028753, R01 NS105646, R01 EB027087, & UF1 AG051216.

4. REFERENCES

[1] Y. Fu, Z. Ma, C. Hamilton, Z. Liang, X. Hou, and et. al., "Genetic influences on resting-state functional networks: A twin study," *Human brain mapping*, vol. 36, pp. 3959–3972, 2015.

[2] D. Falconer and T. Mackay, *Introduction to Quantitative Genetics*, Longman, 4th edition, 1995.

[3] M.N. Hallquist and F.G. Hillary, "Graph theory approaches to functional network organization in brain disorders: A critique for a brave new small-world," *Network Neuroscience*, vol. 3, pp. 1–26, 2018.

[4] K.A. Garrison, D. Schenoist, E.S. Finn, X. Shen, and R.T. Constable, "The (in)stability of functional brain network measures across thresholds," *NeuroImage*, vol. 118, pp. 651–661, 2015.

[5] M. Saggar, O. Sporns, J. Gonzalez-Castillo, P.A. Bandettini, G. Carlsson, G. Glover, and A.L. Reiss, "Towards a new approach to reveal dynamical organization of the brain using topological data analysis," *Nature communications*, vol. 9, pp. 1–14, 2018.

[6] M.K. Chung, H. Lee, A. DiChristofano, H. Ombao, and V. Solo, "Exact topological inference of the resting-state brain networks in twins," *Network Neuroscience*, vol. 3, pp. 674–694, 2019.

[7] S. Seo, M.K. Chung, and H.K. Vorperian, "Heat kernel smoothing using Laplace-Beltrami eigenfunctions," *MICCAI*, 2010, pp. 505–512.

[8] M.K. Chung, L. Xie, S.G. Huang, Y. Wang, J. Yan, and L. Shen, "Rapid acceleration of the permutation test via transpositions," *International Workshop on Connectomics in Neuroimaging*, 2019, pp. 42–53, Springer.

[9] S.M. Smith, C.F. Beckmann, J. Andersson, E.J. Auerbach, J. Bijsterbosch, and et. al., "Resting-state fMRI in the Human Connectome Project," *NeuroImage*, vol. 80, pp. 144–168, 2013.

[10] G. Salimi-Khorshidi, G. Douaud, C.F. Beckmann, M.F. Glasser, L. Griffanti, and S.M. Smith, "Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers," *NeuroImage*, vol. 90, pp. 449–468, 2014.

[11] M.F. Glasser, S.N. Sotiropoulos, J.A. Wilson, T.S. Coalson, B. Fischl, J.L. Andersson, J. Xu, S. Jbabdi, M. Webster, J.R. Polimeni, and D.C. Van Essen, "The minimal preprocessing pipelines for the Human Connectome Project," *NeuroImage*, vol. 80, pp. 105–124, 2013.

[12] M.F. Glasser, T.S. Robinson, E.C. Hacker, C.D. Harwell, J. Yacoub, and et. al., "A multi-modal parcellation of human cerebral cortex," *Nature*, vol. 536, pp. 171–178, 2016.

[13] A.R. Khan, L. Wang, and M.F. Beg, "Freesurfer-initiated fully-automated subcortical brain segmentation in MRI using large deformation diffeomorphic metric mapping," *NeuroImage*, vol. 41, pp. 735–746, 2008.

[14] R.M. Birn, E.K. Molloy, R. Patriat, T. Parker, T.B. Meier, G.R. Kirk, V.A. Nair, M.E. Meyerand, and V. Prabhakaran, "The effect of scan length on the reliability of resting-state fMRI connectivity estimates," *NeuroImage*, vol. 83, pp. 550–558, 2013.