Exact Topological Inference for Paired Brain Networks *via* Persistent Homology

Moo K. Chung[1(✉)], Victoria Villalta-Gil[2], Hyekyoung Lee[3], Paul J. Rathouz[1],
Benjamin B. Lahey[4], and David H. Zald[2]

[1] University of Wisconsin-Madison, Madison, USA
mkchung@wisc.edu
[2] Vanderbilt University, Nashville, USA
[3] Seoul National University, Seoul, South Korea
[4] University of Chicago, Chicago, USA

Abstract. We present a novel framework for characterizing paired brain networks using techniques in hyper-networks, sparse learning and persistent homology. The framework is general enough for dealing with any type of paired images such as twins, multimodal and longitudinal images. The exact nonparametric statistical inference procedure is derived on testing monotonic graph theory features that do not rely on time consuming permutation tests. The proposed method computes the exact probability in quadratic time while the permutation tests require exponential time. As illustrations, we apply the method to simulated networks and a twin fMRI study. In case of the latter, we determine the statistical significance of the heritability index of the large-scale reward network where every voxel is a network node.

1 Introduction

There are many studies related to paired images: longitudinal studies with two repeat scans [7], multimodal imaging study involving PET and MRI [13] and twin imaging studies [15]. The paired images are usually analyzed by relating voxel measurements that match across two images in a mass univariate fashion at each voxel. Compared to the paired image setting, paired brain networks have not been often analyzed possibly due to the lack of problem awareness and analysis frameworks.

In this paper, we present a new unified statistical framework that can integrate paired networks in a holistic fashion by pairing every possible combination of voxels across two images. This is achieved using a hyper-network that connects multiple smaller networks into a larger network. Although hyper-networks are frequently used in machine learning [16], the concept has not been often used in medical imaging. Jie et al. used the hyper-network framework from the resting-state fMRI in classifying MCI from AD in a machine learning framework [9]. Bezerianos et al. constructed the hyper-network from the coupling of EEG activity of pilots and copilots operating an aircraft albeit using ad-hoc procedures without an explicit model specification [1]. Motivated particularly by

the science of [1], we rigorously formulate the problem of characterizing paired networks as a single hyper-network with physically nonexistent hyper-edges connecting between two existing networks.

Graph theory is often used framework for analyzing brain networks mainly due to easier accessibility and interpretation. Since we do not know the exact statistical distribution of many graph theory features, resampling techniques such as the permutation tests have been mainly used in estimating the null distribution and computing p-values. The availability of inexpensive and fast computers made the permutation tests a natural choice for computing p-values when underlying distributions are unknown. However, even with fast computers, permutation test is still extremely slow even for small-scale problems. There is a strong need for fast inference procedures. We propose a new exact inference procedure for graphs and networks. The method speeds up the computation by counting the number of permutations combinatorially instead of empirically computing them by numerically generating large number of permutations.

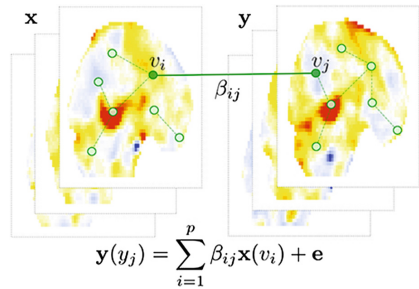Our main contributions are as follows.

1. The new formulation of the hyper-network approach for paired images. We show that hyper-networks can be effectively used as a baseline model for paired twin fMRI. The proposed holistic framework is applied to a twin fMRI study in determining the statistical significance of the brain network differences and the heritability of the reward network.
2. Showing that the topological structure of the sparse hyper-network has a monotonic nestedness. This is used to define monotonic graph features and subsequently in computing *topologically aware* distance between networks.
3. The derivation of the probability distribution of the combinatorial test procedure on graph theory feature vectors that do not rely on resampling techniques such as the permutation tests. While the permutation tests require exponential run time, our exact combinatorial approach requires quadratic run time.

2   Sparse Hyper-Network

Consider a collection of n paired images

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n).$$

We assume x_k and y_k are related either genetically (twin or sibling images), scanwise (multimodal images of the same subject) or longitudinally. We further assume (x_k, y_k) are independently but identically distributed for different i. Let $\mathbf{x} = (x_1, \ldots, x_n)^\top$ and $\mathbf{y} = (y_1, \ldots, y_n)^\top$ be the vectors of images. We set up a hyper-network by relating the paired vectors at different voxels v_i and v_j:

$$\mathbf{y}(y_j) = \sum_{i=1}^{p} \beta_{ij}\mathbf{x}(v_i) + \mathbf{e}$$

Fig. 1. The schematic of hyper-network construction on paired image vectors $\mathbf{x}$ and $\mathbf{y}$. The image vectors $\mathbf{y}$ at voxel v_j is modeled as a linear combination of the first image vector $\mathbf{x}$ at all other voxels. The estimated parameters β_{ij} give the hyper-edge weights.

$$\mathbf{y}(v_j) = \sum_{i=1}^{p} \beta_{ij} \, \mathbf{x}(v_i) + \mathbf{e} \tag{1}$$

for some zero-mean noise vector $\mathbf{e}$ (Fig. 1). The parameters $\beta = (\beta_{ij})$ are the weights of the hyper-edges between voxels v_i and v_j that have to be estimated. We are constructing a *physically nonexistent artificial network* across different images. For fMRI, (1) requires estimating over billions of connections, which is computationally challenging. In practice however, each application will likely to force β to have a specific structure that may reduce the computational burden. For this study, we will consider a reduced model relevant to our genetic imaging application:

$$\mathbf{y}(v_j) = \beta_{ij} \, \mathbf{x}(v_i) + \mathbf{e}. \tag{2}$$

The scientific motivation for using the reduced model will be explained in Sect. 6. Without loss of generality, we can center and scale $\mathbf{x}$ and $\mathbf{y}$ such that

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

$$\|\mathbf{x}(v_i)\|^2 = \mathbf{x}^{\top}(v_i)\mathbf{x}(v_i) = \|\mathbf{y}(v_i)\|^2 = \mathbf{y}^{\top}(v_i)\mathbf{y}(v_i) = 1 \tag{3}$$

for all voxels v_i.

In many applications, we have significantly more number of voxels (p) than the number of images (n), so model (2) is an under-determined system and belongs to the *small-n large-p problem* [5]. It is necessary to regularize the model using a sparse penalty:

$$\widehat{\beta}(\lambda) = \arg\min_{\beta} \frac{1}{2} \sum_{i,j=1}^{p} \| \, \mathbf{y}(v_j) - \beta_{ij} \, \mathbf{x}(v_i) \, \|^2 + \lambda \sum_{i,j=1}^{p} |\beta_{ij}|, \tag{4}$$

where sparse parameter $\lambda \geq 0$ modulates the sparsity of the hyper-edges. The estimation $\widehat{\beta}$ is a function of λ. When $\lambda = 0$, (4) is a least-squares problem and we obtain $\widehat{\beta}_{ij}(0) = \mathbf{x}^{\top}(v_i)\mathbf{y}(v_j)$, which is referred to as the *cross-correlation*. The cross-correlation is invariant under the centering and scaling.

3   Distance Between Networks

To study the topological characteristic of the hyper-network, the estimated edge weights $\widehat{\beta}_{ij}(\lambda)$ are binarized by assigning value 1 to any nonzero weight and 0 otherwise. Let G_λ be the resulting binary graph, i.e., unweighted graph, with adjacency matrix $A_\lambda = (a_{ij}(\lambda))$:

$$a_{ij}(\lambda) = \begin{cases} 1 & \text{if } \widehat{\beta}_{ij}(\lambda) \neq 0; \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

As λ increases, G_λ becomes smaller and nested in a sense that every edge weight gets smaller. This is *not* true for the binarization of other sparse network models. Our sparse hyper-network has this extra layer of additional structure that guarantees the nestedness.

Theorem 1. *The binary graph G_λ obtained from sparse model (4) satisfies*

$$G_{\lambda_1} \supset G_{\lambda_2} \supset \cdots \supset G_{\lambda_q} \tag{6}$$

for any $0 \le \lambda_1 \le \lambda_2 \le \ldots \le \lambda_q \le 1$.

Proof. By solving $d\widehat{\beta}(\lambda)/d\lambda = 0$ for $\lambda \ne 0$ and considering $\widehat{\beta}(0)$ separately, it can be algebraically shown that the solution of the optimization problem (2) is given by [5]

$$\widehat{\beta}_{ij}(\lambda) = \begin{cases} \mathbf{x}^\top(v_i)\mathbf{y}(v_j) - \lambda & \text{if } \mathbf{x}^\top(v_i)\mathbf{y}(v_j) > \lambda \\ 0 & \text{if } |\mathbf{x}^\top(v_i)\mathbf{y}(v_j)| \le \lambda \\ \mathbf{x}^\top(v_i)\mathbf{y}(v_j) + \lambda & \text{if } \mathbf{x}^\top(v_i)\mathbf{y}(v_j) < -\lambda \end{cases} \tag{7}$$

From (3) and the Cauchy-Schwarz inequality, we have $|\mathbf{x}^\top(v_i)\mathbf{y}(v_j)| \le 1$. Thus, if $\lambda > 1$, $\widehat{\beta}(\lambda) = 0$. To avoid the trivial case of zero edge weights everywhere, we need $0 \le \lambda \le 1$. From (7), we have $|\widehat{\beta}_{ij}(\lambda_j)| \ge |\widehat{\beta}_{ij}(\lambda_{j+1})|$ for $0 \le \lambda_j \le \lambda_{j+1} \le 1$. Subsequently, from (5), $a_{ij}(\lambda_j) \ge a_{ij}(\lambda_{j+1})$. Therefore, we have $G_{\lambda_j} \supset G_{\lambda_{j+1}}$. □

The sequence of nested multi-scale graph structure (6) is called the *graph filtration* in persistent homology [11]. The graph filtration can be quantified using monotonic function B satisfying

$$B(G_{\lambda_1}) \le B(G_{\lambda_2}) \le B(G_{\lambda_3}) \le \ldots \le B(G_{\lambda_q}). \tag{8}$$

The number of connected components, which is the zero-th Betti number $\boldsymbol{\beta}_0$ and the most often used topological invariant in persistent homology, satisfies this condition. The size of the largest connected component satisfies a similar but opposite relation of monotonic decrease.

Given two different graph filtrations $\mathbf{G}^1 = \{G_\lambda^1 : 0 \le \lambda \le 1\}$ and $\mathbf{G}^2 = \{G_\lambda^2 : 0 \le \lambda \le 1\}$, define the distance between them as

$$D(\mathbf{G}^1, \mathbf{G}^2) = \sup_{\lambda \in [0,1]} \left| B(G_\lambda^1) - B(G_\lambda^2) \right|,$$

which can be discretely approximated as

$$D_q = \sup_{1 \le j \le q} \left| B(G_{\lambda_j}^1) - B(G_{\lambda_j}^2) \right|.$$

If we choose enough number of q such that λ_j are all the sorted edge weights, then $D(\mathbf{G}^1, \mathbf{G}^2) = D_q$. This is possible since there are only up to $p(p-1)/2$ number of unique edge weights in a graph with p nodes and $B(G_\lambda^1)$ and $B(G_\lambda^2)$

increase discretely. In practice, $\rho_j = j/q$ is chosen uniformly in $[0, 1]$ or a divide-and-conquer strategy can be used to do adaptively grid the unit interval.

D satisfies all the axioms of a metric except identity. $D(\mathbf{G}^1, \mathbf{G}^2) = 0$ does not imply $\mathbf{G}^1 = \mathbf{G}^2$. Thus, D is *not* a metric. However, it will be shown in Sect. 4 that probability $P(D(\mathbf{G}^1, \mathbf{G}^2) = 0) = 0$ showing such event rarely happens in practice. Thus, D can be treated as a metric-like in applications without much harm.

4   Exact Topological Inference

We are interested in testing the null hypothesis H_0 of the equivalence of two monotonic graph feature functions:

$$H_0 : B(G_\lambda^1) = B(G_\lambda^2) \text{ for all } \lambda \in [0, 1]$$

$$vs.$$

$$H_1 : B(G_\lambda^1) \neq B(G_\lambda^2) \text{ for some } \lambda \in [0, 1].$$

As a test statistic, we propose to use D_q. The test statistic takes care of the multiple comparisons by the use of supremum. Under the null hypothesis, we can derive the probability distribution of D_q combinatorially without numerically permuting images.

Fig. 2. $A_{u,v}$ are computed within the boundary (dotted red line). The red numbers are the number of paths from $(0, 0)$. (Color figure online)

Theorem 2.

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

where $A_{u,v}$ satisfies $A_{u,v} = A_{u-1,v} + A_{u,v-1}$ with the boundary condition $A_{0,q} = A_{q,0} = 1$ within band $|u - v| < d$.

Proof. The proof is similar to the combinatorial construction of Kolmogorov-Smirnov (KS) test [2,5,8]. Combine two monotonically increasing vectors

$$(B(G_{\rho_1}^1), \ldots, B(G_{\rho_q}^1)), \quad (B(G_{\rho_1}^2), \ldots, B(G_{\rho_q}^2))$$

and arrange them in increasing order. Represent $B(G_{\rho_j}^1)$ and $B(G_{\rho_j}^2)$ as $\uparrow$ and $\rightarrow$ respectively. For example, $\uparrow\uparrow\rightarrow\uparrow\rightarrow\rightarrow\cdots$. There are exactly q number of $\uparrow$ and q number of $\rightarrow$ in the sequence. Treat the sequence as walks on a Cartesian grid. $\rightarrow$ indicates one step to the right and $\uparrow$ indicates one step up. Thus the walk starts at $(0, 0)$ and ends at (q, q) (Fig. 2). There are total $\binom{2q}{q}$ possible number of paths, which forms the sample space. This is also the total number of possible permutations between the elements of the two vectors. The null assumption is that the two vectors are identical and there is no preference to one vector element to another. Thus, each walk is equally likely to happen in the sample space.
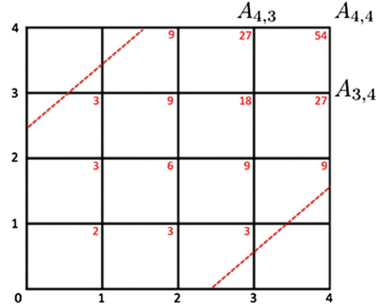
The values of $B(G^1{}_{\rho_j})$ and $B(G^2{}_{\rho_j})$ satisfying the condition $\sup_j |B(G^1{}_{\rho_j}) - B(G^2{}_{\rho_j})| < d$ are the integer coordinates of (u, v) on the path satisfying $|u - v| < d$. These are integer grid points within the boundary $v = u \pm d$ (dotted red lines in Fig. 2). Thus the probability can be written as

$$P(D_q \geq d) = 1 - P(D_q < d) = 1 - \frac{A_{q,q}}{\binom{2q}{q}}, \tag{9}$$

where $A_{u,v}$ be the total number of passible paths from $(0,0)$ to (u, v) within the boundary. Since there are only two paths (either $\uparrow$ or $\rightarrow$), $A_{u,v}$ can be computed recursively as

$$A_{u,v} = A_{u-1,v} + A_{u,v-1}.$$

within the boundary. On the boundary, $A_{0,q} = A_{q,0} = 1$ since there is only one path. □

Theorem 2 provides the exact probability computation for any number of nodes p. For instance, probability $P(D_4 \geq 2.5)$ is computed iteratively as follows. We start with computing $A_{1,1} = A_{0,1} + A_{1,0} = 2, A_{2,1} = A_{1,1} + A_{1,0} = 3, \cdots, A_{4,4} = A_{4,3} + A_{3,4} = 27 + 27 = 54$ (red numbers in Fig. 2). Thus the probability is $P(D_4 \geq 2.5) = 1 - 54/\binom{8}{4} = 0.23$. Few other examples that can be computed easily are

$$P(D_q = 0) = 0, \ P(D_q \geq 1) = 1, \ P(D_q \geq q) = 2/\binom{2q}{q}, \ P(D_q \geq q+1) = 0.$$

Computing $A_{q,q}$ iteratively requires at most q^2 operations while permuting two samples consisting of q elements each requires $\binom{2q}{q}$ operations. Thus, our method can compute the p-value exactly substantially faster than the permutation test that is approximate and exponentially slow.

The asymptotic probability distribution of D_q can be also determined for sufficiently large q without computing iteratively as Theorem 2.

Theorem 3. $\lim_{q \to \infty} P\left(D_q/\sqrt{2q} \geq d\right) = 2\sum_{i=1}^{\infty}(-1)^{i-1}e^{-2i^2d^2}.$

The proof is not given here but the result follows from [8,14]. From Theorem 3, p-value under H_0 is computed as

$$p - \text{value} = 2e^{-d_o^2} - 2e^{-8d_o^2} + 2e^{-18d_o^2}\cdots,$$

where d_o is the *least integer greater than* $D_q/\sqrt{2q}$ in the data (Fig. 3-left). For any large observed value $d_0 \geq 2$, the second term is in the order of 10^{-14} and insignificant. Even for small observed d_0, the expansion converges quickly and 5 terms are sufficient.
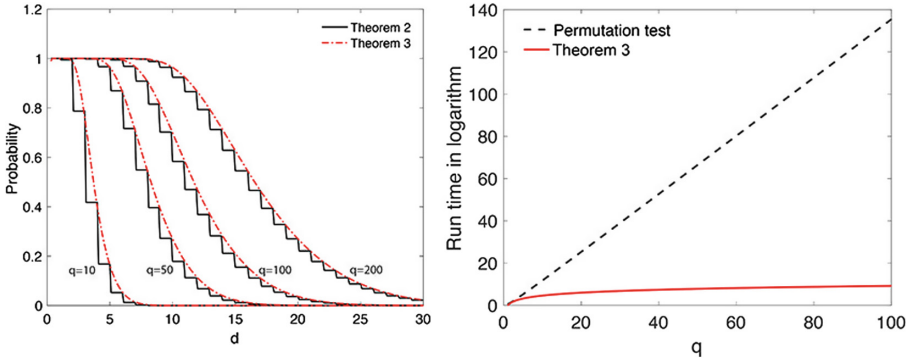
Fig. 3. Left: Convergence of Theorem 2 to 3 for $q = 10, 50, 100, 200$. Right: Run time of permutation test (dotted black line) *vs.* the combinational method in Theorem 2 (solid red line) in logarithmic scale. (Color figure online)

5 Validation and Comparison

The proposed method was validated and compared in two simulation studies against Gromov-Hausdorff (GH) distance. GH-distance was originally used to measure distance between two metric spaces. It was later adapted to measure distances in persistent homology, dendrograms [3] and brain networks [11]. Following [11], the computation of GH-distance is done done on graphs with1-correlation as edge weights. GH-distance is the difference in the L^{∞}-norm of corresponding single linkage matrices.

No Network Difference. The simulations were performed 1000 times and the average results were reported. There are three groups and the sample size is $n = 5$ in each group and the number of nodes are $p = 40$. In Groups I and II, data $x_k(v_i)$ at each node v_i was simulated as standard normal, i.e., $x_k(v_i) \sim N(0, 1)$. The paired data was simulated as $y_k(v_i) = x_k(v_i) + N(0, 0.02^2)$ for all the nodes (Fig. 4). Using the proposed combinatorial method, we obtained the p-values of 0.6568 ± 0.3367 and 0.2830 ± 0.3636 for $\boldsymbol{\beta}_0$ and the size of the largest connected component demonstrating that the method did not detect network differences as expected.

Using GH-distance, we obtained the p-value of 0.4938 ± 0.2919. The p-value was computed using the permutation test. Since there are 5 samples in each group, the total number of permutation is $\binom{10}{5} = 272$ making the permutation test exact and the comparison fair. Both methods seem to perform reasonably well. However, GH-distance method took about $950\,\mathrm{s}$ ($16\,\mathrm{min}$) while the combinatorial method took about $20\,\mathrm{s}$ in a computer.

Network Difference. Group III was generated identically and independently like Group I but additional dependency was added by letting $y_k(v_j) = x_k(v_1)/2$ for ten nodes indexed by $j = 1, 2, \ldots, 10$. This dependency gives high connectivity
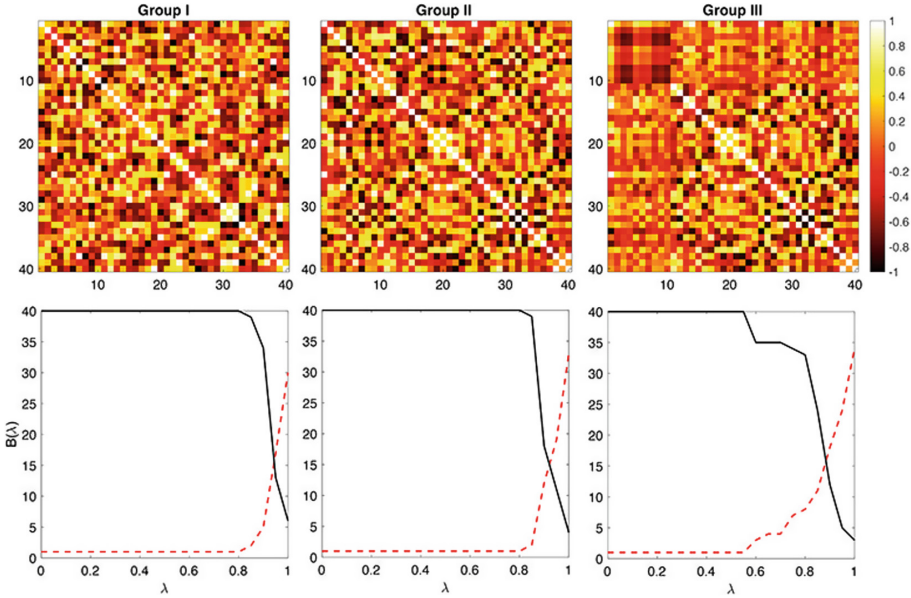
Fig. 4. Simulation studies. Group I and Group II are generated independently and identically. The resulting $B(\lambda)$ plots are similar. The dotted red line is β_0 and the solid black line is the size of the largest connected component. No statistically significant differences can be found between Groups I and II. Group III is generated independently but identically as Group I but additional dependency is added for the first 10 nodes (square on the top left corner). The resulting $B(\lambda)$ plots are statistically different between Groups I and II. (Color figure online)

differences between Groups I and III (Fig. 4). Using the proposed combinatorial method, we obtained the p-values of 0.0422 ± 0.0908 and 0.0375 ± 0.1056 for β_0 and the size of the largest connected component. On the other hand, we obtained 0.1145 ± 0.1376 for GH-distance. The proposed method seems to perform better in the presence of signal. The `MATLAB` code for performing these simulations is given in http://www.cs.wisc.edu/~mchung/twins.

6   Twin fMRI Study

The study consists of 11 monozygotic (MZ) and 9 same-sex dizygotic (DZ) twin pairs of 3 T functional magnetic resonance images (fMRI) acquired in Intera-Achiava Phillips MRI scanners with a 32 channel SENSE head coil. Subjects completed monetary incentive delay task of 3 runs of 40 trials [10]. A total 120 trials consisting of 40 $0, 40 $1 and 40 $5 rewards were pseudo randomly split into 3 runs.

fMRI data went through spatial normalization to the template following the standard SPM pipeline. The fMRI dimensions after preprocessing are $53 \times 63 \times 46$ and the voxel size is 3 mm cube. There are a total of $p = 25972$ voxels in the template. After fitting a general linear model at each voxel, we obtained the
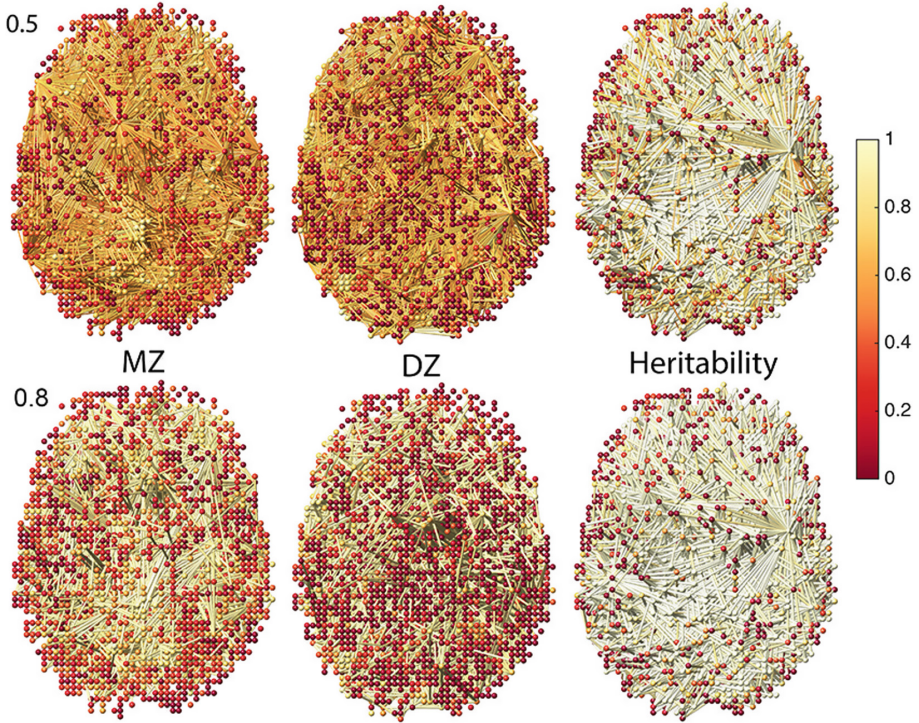
Fig. 5. Left, middle: Node colors are correlations of MZ- and DZ-twins. Edge colors are sparse cross-correlations at sparsity $\lambda = 0.5, 0.8$. Right: Heritability index (HI) at nodes and edges. MZ-twins show higher correlations compared to DZ-twins. Some low HI nodes show high HI edges. Using only the voxel-level HI feature, we may fail to detect such high-order genetic effects on the functional network.

contrast maps testing the significance of activity in the delay period for \$5 trials relative to the control condition of \$0 reward (Fig. 1). The paired contrast maps were then used as the initial images vectors $\mathbf{x}$ and $\mathbf{y}$ in the starting model (2).

We are interested in knowing *the extent of the genetic influence on the functional brain network* of these participants while anticipating the high reward as measured by activity during the delay that occurs between the reward cue and the target, and its statistical significance. The hyper-network approach is applied in extending the voxel-level univariate genetic feature called *heritability index* (HI) into a network-level multivariate feature. HI determines the amount of variation (in terms of percentage) due to genetic influence in a population. HI is often estimated using Falconer's formula [6] as a baseline. MZ-twins share 100% of genes while DZ-twins share 50% of genes. Thus, the additive genetic factor A, the common environmental factor C for each twin type are related as

$$\rho_{\mathrm{MZ}}(v_i) = A + C, \tag{10}$$

$$\rho_{\mathrm{DZ}}(v_i) = A/2 + C, \tag{11}$$

where ρ_{MZ} and ρ_{DZ} are the pairwise correlation between MZ- and and same-sex DZ-twins at voxel v_i. Solving (10) and (11), we obtain the additive genetic factor, i.e., HI given by

$$\mathcal{H}(v_i) = 2[\rho_{\mathrm{MZ}}(v_i) - \rho_{\mathrm{DZ}}(v_i)].$$

HI is a univariate feature measured at each voxel and does not quantify how the changes in one voxel is related to other voxels. We can determine HI across voxels v_i and v_j as

$$\mathcal{H}(v_i, v_j) = 2[\varrho_{\mathrm{MZ}}(v_i, v_j) - \varrho_{\mathrm{DZ}}(v_i, v_j)],$$

ϱ_{MZ} and ϱ_{DZ} are the symmetrized cross-correlations between voxels v_i and v_j for MZ- and DZ-twin pairs:

$$\varrho_{\mathrm{MZ}}(v_i, v_j) = (\widehat{\beta}_{ij}^{\mathrm{MZ}} + \widehat{\beta}_{ji}^{\mathrm{MZ}})/2, \quad \varrho_{\mathrm{DZ}}(v_i, v_j) = (\widehat{\beta}_{ij}^{\mathrm{DZ}} + \widehat{\beta}_{ji}^{\mathrm{DZ}})/2.$$

$\widehat{\beta}_{ij}^{\mathrm{MZ}}$ and $\widehat{\beta}_{ij}^{\mathrm{DZ}}$ are the estimated cross-correlations from model (4). Note that

$$\varrho_{\mathrm{MZ}}(v_i, v_i) = \rho_{\mathrm{MZ}}(v_i), \quad \varrho_{\mathrm{DZ}}(v_i, v_i) = \rho_{\mathrm{DZ}}(v_i), \quad \mathcal{H}(v_i, v_i) = \mathcal{H}(v_i).$$

In Fig. 5-left and -middle, the nodes are $\rho_{\mathrm{MZ}}(v_i)$ and $\rho_{\mathrm{DZ}}(v_i)$ while the edges are $\varrho_{\mathrm{MZ}}(v_i, v_j)$ and $\varrho_{\mathrm{DZ}}(v_i, v_j)$ for $\lambda = 0.5, 0.8$. In Fig. 5-right, we have $\mathcal{H}(v_i) = \mathcal{H}(v_i, v_i)$. The network visualization shows high HI values for almost everywhere along the edges. We are interested in testing the statistical significance of the estimated HI for all edges. Testing for $\mathcal{H}(v_i, v_j) = 0$ is equivalent to testing $\varrho_{\mathrm{MZ}}(v_i, v_j) = \varrho_{\mathrm{DZ}}(v_i, v_j)$. Thus, we test for hyper-network differences between twins using the test statistic D_q ($q = 100$ is used) (Fig. 6). For $\boldsymbol{\beta}_0$ and the size of the largest connected component p-values are less than 0.00002 and 0.00001 respectively indicating very strong significance of MZ- and DZ- network difference and HI of the whole brain network.

7   Discussion

Hyper-Network. In this paper, we presented a unified statistical framework for for analyzing paired images using a hyper-network. Although we applied the framework in twin fMRI, the method can be easily applied to other paired image settings such as longitudinal and multimodal studies. The method can be further extended to triple images. Then we should start with the following hyper-network model

$$\mathbf{y}(v_j) = \sum_{i=1}^{p} \beta_{ij} \mathbf{x}(v_i) + \sum_{i=1}^{p} \gamma_{ij} \mathbf{z}(v_i) + \mathbf{e}. \tag{12}$$

Since we usually have more number of nodes p than the number of images n, it is necessary to introduce two separate sparse parameters in (12). Then, instead of building graph filtration over 1D, we need to build it over 2D sparse parameter space. This is related to multidimensional persistent homology [4, 12]. Extending the proposed method to triple image setting is left as a future study.
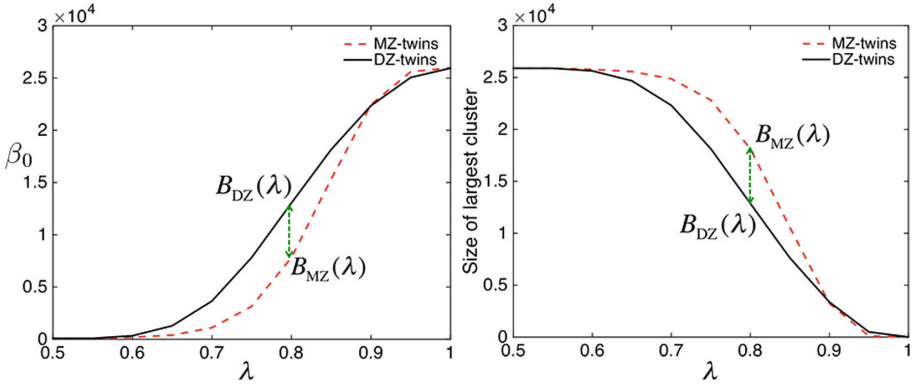
Fig. 6. The result of graph filtrations on twin fMRI data. The number of connected components (left) and the size of the largest connected component (right) are plotted over the sparse parameter λ. For each λ, MZ-twins tend to have smaller number of connected components but larger connected component. The dotted green arrow (D_q) where the maximum group separation occurs. (Color figure online)

Exact Topological Inference. We presented a combinatorial method for computing the probability that avoids time consuming permutation tests. Theorem 2 is the exact nonparametric procedure and does not assume any statistical distribution on graph features other than that they has to be monotonic. The technique is very general and applicable to any monotonic graph features such as node degree. Based on Stirling's formula, $q! \sim \sqrt{2\pi q}(\frac{q}{e})^q$, the total number of permutations in permuting two vectors of size q each is $\binom{2q}{q} \sim \frac{4^q}{\sqrt{2\pi q}}$. This is substantially larger than the quadratic run time q^2 needed in our method (Fig. 3). Even for small $q = 10$, more than tens of thousands permutations are needed for the accurate estimation the *p*-value. On the other hand, only up to 100 iterations are needed in our combinatorial method.

References

1. Bezerianos, A., Sun, Y., Chen, Y., Woong, K.F., Taya, F., Arico, P., Borghini, G., Babiloni, F., Thakor, N.: Cooperation driven coherence: brains working hard together. In: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 4696–4699 (2015)
2. Böhm, W., Hornik, K.: A Kolmogorov-Smirnov test for r samples. Institute for Statistics and Mathematics. Research report Series: Report, 105 (2010)

3. Carlsson, G., Memoli, F.: Persistent clustering and a theorem of J. Kleinberg. arXiv preprint arXiv:0808.2241 (2008)

4. Carlsson, G., Singh, G., Zomorodian, A.: Computing multidimensional persistence. In: Dong, Y., Du, D.-Z., Ibarra, O. (eds.) ISAAC 2009. LNCS, vol. 5878, pp. 730–739. Springer, Heidelberg (2009). doi:10.1007/978-3-642-10631-6_74

5. Chung, M.K., Hanson, J.L., Ye, J., Davidson, R.J., Pollak, S.D.: Persistent homology in sparse regression and its application to brain morphometry. IEEE Trans. Med. Imaging **34**, 1928–1939 (2015)

6. Falconer, D., Mackay, T.: Introduction to Quantitative Genetics, 4th edn. Longman, Harlow (1995)

7. Freeborough, P.A., Fox, N.C.: Modeling brain deformations in Alzheimer disease by fluid registration of serial 3D MR images. J. Comput. Assist. Tomogr. **22**, 838–843 (1998)

8. Gibbons, J.D., Chakraborti, S.: Nonparametric Statistical Inference. Chapman & Hall/CRC Press, Boca Raton (2011)

9. Jie, B., Shen, D., Zhang, D.: Brain connectivity hyper-network for MCI classification. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8674, pp. 724–732. Springer, Cham (2014). doi:10.1007/978-3-319-10470-6_90

10. Knutson, B., Adams, C.M., Fong, G.W., Hommer, D.: Anticipation of increasing monetary reward selectively recruits nucleus accumbens. J. Neurosci. **21**, 159 (2001)

11. Lee, H., Chung, M.K., Kang, H., Kim, B.-N., Lee, D.S.: Computing the shape of brain networks using graph filtration and Gromov-Hausdorff metric. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011. LNCS, vol. 6892, pp. 302–309. Springer, Heidelberg (2011). doi:10.1007/978-3-642-23629-7_37

12. Lee, H., Kang, H., Chung, M.K., Lim, S., Kim, B.-N., Lee, D.S.: Integrated multimodal network approach to PET and MRI based on multidimensional persistent homology. Hum. Brain Mapp. **38**, 1387–1402 (2017)

13. Pichler, B.J., Kolb, A., Nägele, T., Schlemmer, H.-P.: PET/MRI: paving the way for the next generation of clinical multimodality imaging applications. J. Nuclear Med. **51**, 333–336 (2010)

14. Smirnov, N.V.: Estimate of deviation between empirical distribution functions in two independent samples. Bull. Moscow Univ. **2**, 3–16 (1939)

15. Thompson, P.M., Cannon, T.D., Narr, K.L., van Erp, T., Poutanen, V.P., Huttunen, M., Lonnqvist, J., Standertskjold-Nordenstam, C.G., Kaprio, J., Khaledy, M.: Genetic influences on brain structure. Nat. Neurosci. **4**, 1253–1258 (2001)

16. Zhang, B.-T.: Hypernetworks: a molecular evolutionary architecture for cognitive learning and memory. IEEE Comput. Intell. Mag. **3**, 49–63 (2008)