Topological State-Space Estimation of Functional Human Brain Networks

Moo K. Chung, Shih-Gu Huang, Ian C. Carroll,Vince D. Calhoun, H. Hill Goldsmith

Abstract We present a new data driven topological approach for estimating state spaces in dynamically changing functional brain networks of human. Our approach penalizes the topological discrepancy between networks and cluster dynamically changing brain networks into topologically distinct states. Our method take into account the temporal dimension of the data through the Wasserstein distance between networks. Our method is shown to outperform the widely used k-means clustering often used in estimating the state space in brain networks. The method is applied in more accurate determination of the state spaces of dynamically changing functional brain networks. Sequentially we address the question of if the overall topology of brain networks is a heritable feature using the twin study design.

Moo K. Chung

Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, USA e-mail: mkchung@wisc.edu

Shih-Gu Huang
PPG/ECG Signal, Taiwan e-mail: shihgu@gmail.com

Ian C. Carroll
Department of Psychology, University of Wisconsin-Madison, USA e-mail: icarroll@wisc.edu

Vince D. Calhoun
Tri-institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State, Georgia Tech, Emory Georgia State University, Georgia, USA e-mail: vcalhoun@gsu.edu

H. Hill Goldsmith
Department of Psychology, University of Wisconsin-Madison, USA e-mail: hill.goldsmith@wisc.edu

1 Introduction

In standard graph theory based network analysis, network features such as node degrees and clustering coefficients are obtained from the adjacency matrices after thresholding weighted edges (Sporns, 2003; Wijk et al., 2010; Chung et al., 2017a). The final statistical analysis results change depending on the choice of threshold or parameter (Chung et al., 2013; Lee et al., 2012). There is a need to develop a multi-scale network analysis framework that provides consistent results and interpretation regardless of the choice of parameter. Persistent homology, a branch of algebraic topology, offers a novel solution to this multiscale analysis challenge (Edelsbrunner and Harer, 2010). Instead of examining networks at one fixed scale, persistent homology identifies persistent topological features that are robust under different scales (Petri et al., 2014; Sizemore et al., 2018). Unlike existing graph theory approaches that analyze networks at one different fixed scale at a time, persistent homology captures the changes of topological features over different scales and then identifies the most persistent topological features that are robust under noise perturbations. This robust performance under different scales is needed for *dynamic brain networks* that change over time.

Persistent homological network approaches are shown to be more robust and outperforming many existing graph theory measures and methods. In (Lee et al., 2011, 2012), persistent homology was shown to outperform eight existing graph theory features such as clustering coefficient, samll-worldness and modularity. In (Chung et al., 2017b, 2019a), persistent homology was shown to outperform various matrix norm based network distances. In (Wang et al., 2018), persistent homology was shown to outperform the power spectral density and local variance methods. In (Wang et al., 2017), persistent homology was shown to outperform topographic power maps. In (Yoo et al., 2017), center persistency was shown to outperform the network-based statistic and element-wise multiple corrections. However, the method has been mainly used on *static* networks or as a static summary of time varying networks (Bassett and Sporns, 2017). The dynamic pattern of persistent homology for time varying brain network was rarely investigated expect few (Yoo et al., 2016; Santos et al., 2019; Songdechakraiwut and Chung, 2020).

In this paper, we propose to develop the novel *dynamic persistent homology* framework for time varying network data. Our coherent scalable framework for the computation is based on the Wasserstein distance between persistent diagrams, which provides the topological profile of data into 2D scattered point. We directly establish the relationship between the Wasserstein distance and edge weights in networks making the method far more accessible and adaptable. We achieve $O(n \log n)$ run time in most graph manipulation tasks such as matching and averaging. Such scalable computation enables us to perform a computationally demanding task such as topological clustering with ease. The method is applied in the determination of the state space of dynamically changing functional brain networks obtained from the resting-state functional magnetic resonance imaging (rs-fMRI). We will show that the proposed method based on the Wasserstein distance can capture the topological patterns that are consistently observed across different time points. The Wasserstein

distance or Kantorovich–Rubinstein metric is originally defined between probability distributions (Vallender, 1974; Canas and Rosasco, 2012; Berwald et al., 2018). Due to the connection to the optimal mass transport, which enjoys various optimal properties, the Wasserstein distance has been applied to various imaging applications. However, there are not many applications of Wasserstein distance in network data. (Mi et al., 2018) used the Wasserstein distance in resampling brain surface meshes. (Shi et al., 2016; Su et al., 2015), used the Wasserstein distance in classifying brain cortical surface shapes. (Hartmann et al., 2018) used the Wassterstein distance in building generative adversarial networks. (Sabbagh et al., 2019) used the Wassterstein distance for manifold regression problem in the space of positive definite matrices for the source localization problem in EEG. (Xu et al., 2021) used the Wasserstein distance in predicting Alzheimer's disease progression in magnetoencephalography (MEG) brain networks. However, the Wassterstein distance in these applications are all geometric in nature.

To increase the reproducibility, MATLAB code for performing the method is provided in `https://github.com/laplcebeltrami/dynamicTDA`.

2 Birth-death decomposition of brain networks

2.1 Graphs as simplical complices

A high dimensional object such as brain networks can be modeled as weighted graph $X = (V, w)$ consisting of node set V indexed as $V = \{1, 2, \cdots, p\}$ and edge weights $w = (w_{ij})$ between nodes i and j. If we order the edge weights in the increasing order, we have the sorted edge weights:

$$\min_{j,k} w_{jk} = w_{(1)} < w_{(2)} < \cdots < w_{(q)} = \max_{j,k} w_{jk},$$

where $q \leq (p^2 - p)/2$. The subscript $_{( )}$ denotes the order statistic. In terms of sorted edge weight set $W = \{w_{(1)}, \cdots, w_{(q)}\}$, we may also write the graph as $X = (V, W)$. If we connect nodes following some criterion on the edge weights, they will form a simplicial complex which will follow the topological structure of the underlying weighted graph (Edelsbrunner and Harer, 2010; Zomorodian, 2009). Note that the k-simplex is the convex hull of $k + 1$ points in V. A simplicial complex is a finite collection of simplices such as points (0-simplex), lines (1-simplex), triangles (2-simplex) and higher dimensional counter parts.

The *Rips complex* X_ϵ is a simplicial complex, whose k-simplices are formed by $(k + 1)$ nodes which are pairwise within distance ϵ (Ghrist, 2008). While a graph has at most 1-simplices, the Rips complex has at most $(p - 1)$-simplices. The Rips complex induces a hierarchical nesting structure called the Rips filtration

$$X_{\epsilon_0} \subset X_{\epsilon_1} \subset X_{\epsilon_2} \subset \cdots$$

for $0 = \epsilon_0 < \epsilon_1 < \epsilon_2 < \cdots$, where the sequence of ϵ-values are called the filtration values. The filtration is quantified through a topological basis called *k-cycles*. 0-cycles are the connected components, 1-cycles are 1D closed paths or loops while 2-cycles are a 3-simplices (tetrahedron) without interior. Any k-cycle can be represented as a linear combination of basis k-cycles. The Betti numbers β_k counts the number of independent k-cycles. During the Rips filtration, the i-th k-cycle is born at filtration value b_i and dies at d_i. The collection of all the paired filtration values

$$P(\mathcal{X}) = \{(b_1, d_1), \cdots, (b_q, d_q)\}$$

displayed as 1D intervals is called the *barcode* and displayed as scatter points in 2D plane is called the *persistent diagram*. Since $b_i < d_i$, the scatter points in the persistent diagram are displayed above the line $y = x$ line by taking births in the x-axis and deaths in the y-axis.

For dynamically changing brain network $\mathcal{X}(t) = (V, w(t))$, we assume the node set is fixed while edge weights are changing over time t. If we build persistent homology at each fixed time, the resulting barcode is also time dependent:

$$P(\mathcal{X}(t)) = \{(b_1(t), d_1(t)), \cdots, (b_q(t), d_q(t))\}.$$

2.2 Graph filtrations

As the number of nodes p increases, the resulting Rips complex becomes very dense. As the filtration values increases, there exists an edge between every pair of nodes. At higher filtration values, Rips filtration becomes an ineffective representation of networks. To remedy this issue, graph filtration was introduced (Lee et al., 2011, 2012). Given weighted graph $\mathcal{X} = (V, w)$ with edge weight $w = (w_{ij})$, the binary network $\mathcal{X}_\epsilon = (V, w_\epsilon)$ is a graph consisting of the node set V and the binary edge weights $w_\epsilon = (w_{\epsilon,ij})$ given by

$$w_{\epsilon,ij} = \begin{cases} 1 & \text{if } w_{ij} > \epsilon; \\ 0 & \text{otherwise.} \end{cases}$$

Note w_ϵ is the adjacency matrix of $\mathcal{X}_\epsilon$, which is a simplicial complex consisting of 0-simplices (nodes) and 1-simplices (edges) (Ghrist, 2008). While the binary network $\mathcal{X}_\epsilon$ has at most 1-simplices, the Rips complex can have at most $(p-1)$-simplices. By choosing threshold values at sorted edge weights $w_{(1)}, w_{(2)}, \cdots, w_{(q)}$ (Chung et al., 2013), we obtain the sequence of nested graphs:

$$\mathcal{X}_{w_{(1)}} \supset \mathcal{X}_{w_{(2)}} \supset \cdots \supset \mathcal{X}_{w_{(q)}}.$$

The sequence of such a nested multiscale graph is called as the *graph filtration* (Lee et al., 2011, 2012). Note that $\mathcal{X}_{w_{(1)}-\epsilon}$ is the complete weighted graph for any $\epsilon > 0$.
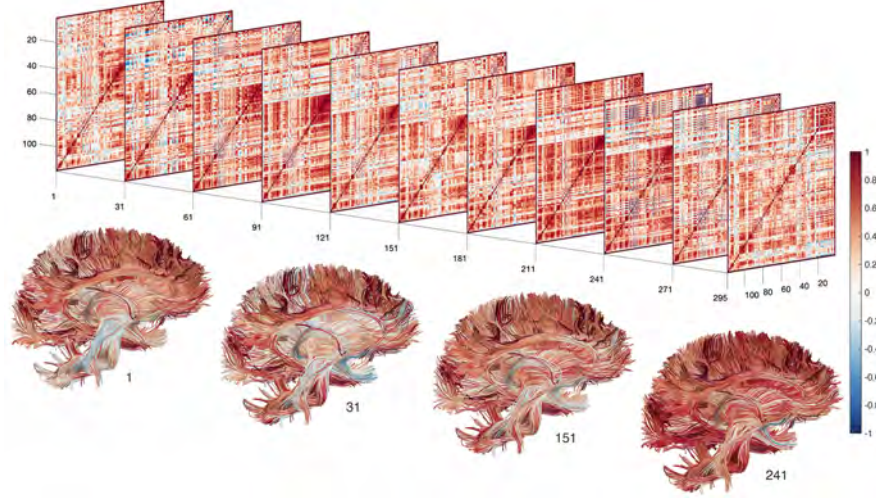
Fig. 1 Dynamically changing correlation matrices computed from rs-fMRI using the sliding window of size 60 for a subject. The constructed correlation matrices are superimposed on top of the white matter fibers in the MNI space and color coded based on correlation values.

On the other hand, $\mathcal{X}_{w_{(q)}}$ is the node set V. By increasing the threshold value, we are thresholding at higher connectivity so more edges are removed.

For dynamically changing brain networks (Figure 1), we can similarly build time varying graph filtrations at each time point $\{\mathcal{X}_\epsilon(t) : t \in \mathbb{R}^+\}$.

2.3 Birth-death decomposition

Unlike the Rips complex, there are no higher dimensional topological features beyond the 0D and 1D topology in graph filtration. The 0D and 1D persistent diagrams (b_i, d_i) tabulate the life-time of 0-cycles (connected components) and 1-cycles (loops) that are born at the filtration value b_i and die at value d_i. The 0th Betti number $\beta_0(w_{(i)})$ counts the number of 0-cycles at filtration value $w_{(i)}$ and shown to be non-decreasing over filtration (Figure 2) (Chung et al., 2019a): $\beta_0(w_{(i)}) \leq \beta_0(w_{(i+1)})$. On the other hand the 1st Betti number $\beta_1(w_{(i)})$ counts the number of independent loops and shown to be non-increasing over filtration (Chung et al., 2019a): $\beta_1(w_{(i)}) \geq \beta_1(w_{(i+1)})$.

During the graph filtration, when new components is born, they never dies. Thus, 0D persistent diagrams are completely characterized by birth values b_i only. Loops are viewed as already born at $-\infty$. Thus, 1D persistent diagrams are completely characterized by death values d_i only. We can show that the edge weight set W can
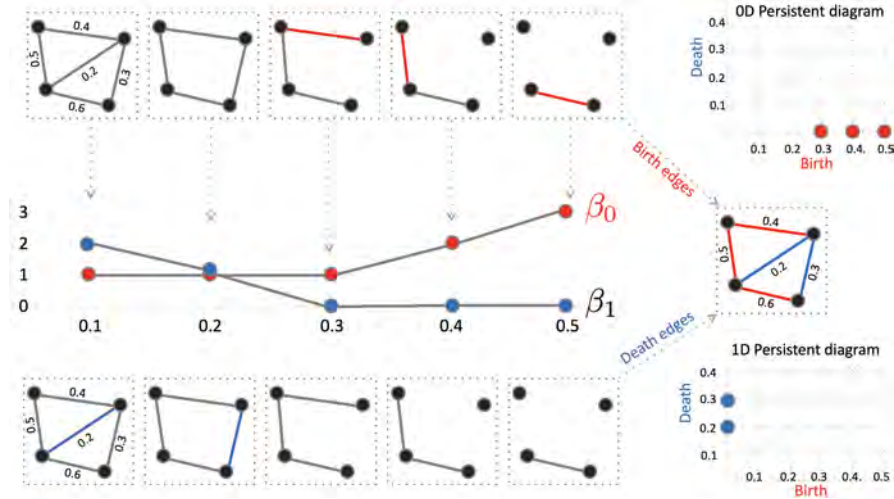
Fig. 2 The birth-death decomposition partitions the edge set into the birth and death edge sets. The birth set forms the maximum spanning tree (MST) and contains edges that create connected components (0D topology). The death set contains edges not belong to maximum spanning tree (MST) and destroys loops (1D topology).

be partitioned into 0D birth values and 1D death values (Songdechakraiwut et al., 2021):

Theorem 1 (Birth-death decomposition) *The edge weight set $W = \{w_{(1)}, \cdots, w_{(q)}\}$ has the unique decomposition*

$$W = W_b \cup W_d, \quad W_b \cap W_d = \emptyset \tag{1}$$

where birth set $W_b = \{b_{(1)}, b_{(2)}, \cdots, b_{(q_0)}\}$ is the collection of 0D sorted birth values and death set $W_d = \{d_{(1)}, d_{(2)}, \cdots, d_{(q_1)}\}$ is the collection of 1D sorted death values with $q_0 = p - 1$ and $q_1 = (p-1)(p-2)/2$. Further W_b forms the 0D persistent diagram while W_d forms the 1D persistent diagram.

In a complete graph with p nodes, there are $q = p(p-1)/2$ unique edge weights. There are $q_0 = p - 1$ number of edges that produces 0-cycles. This is equivalent to the number of edges in the maximum spanning tree of the graph. Thus, $q_1 = q - q_0 = \frac{(p-1)(p-2)}{2}$ number of edges destroy loops. The 0D persistent diagram is given by $\{(b_{(1)}, \infty), \cdots, (b_{(q_0)}, \infty)\}$. Ignoring ∞, W_b is the 0D persistent digram. The 1D persistent diagram of the graph filtration is given by $\{(-\infty, d_{(1)}), \cdots, (-\infty, d_{(q_1)})\}$. Ignoring $-\infty$, W_d is the 1D persistent digram. We can show that the birth set is the maximum spanning tree (MST) (Figure 2) (Songdechakraiwut and Chung, 2023).
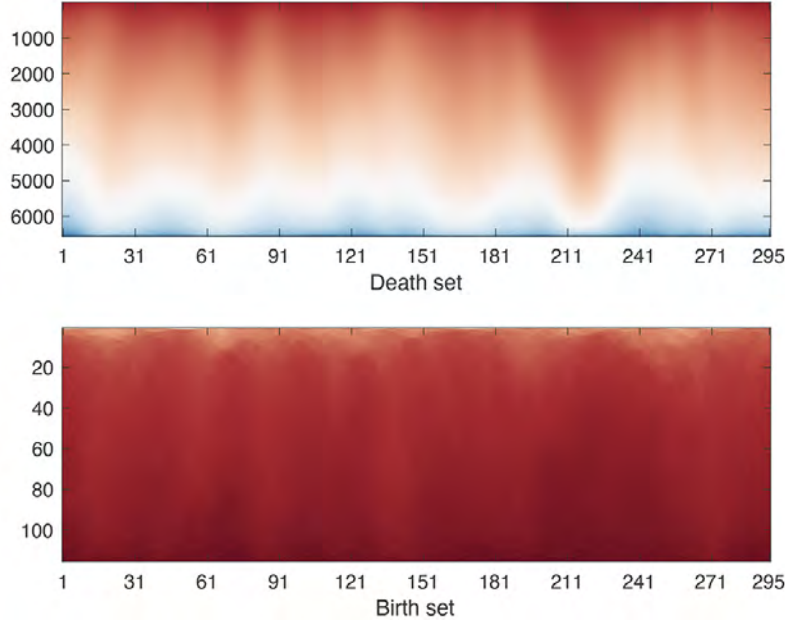
Fig. 3 The corresponding birth and death sets of dynamically changing correlation matrix shown in Figure 1. The horizontal axis is the time point. Columns are the sorted birth and death edge values at the time point.

Numerical implementation. The identification of W_b is based on the modification to Kruskal's or Prim's algorithm and identify the MST (Lee et al., 2012). Then W_d is identified as W/W_d. Figure 3 displays how the birth and death sets change over time for a single subject used in the study. Given edge weight matrix W as input, the Matlab function `WS_decompose.m` outputs the birth set W_b and the death set W_d.

3 Topological clustering using the Wasserstein distance

Just like the majority of clustering methods such as k-means and hierarchical clustering that uses geometric distances (Johnson, 1967; Hartigan and Wong, 1979; Lee et al., 2012), we propose to develop a topological clustering method using topological distances. The main difference between the geometric and topological distance is if the distance can discriminate in the presence of topological difference and not able to discriminate in the presence of topological indifference (Figure 4). For this purpose we use the Wasserstein distance.
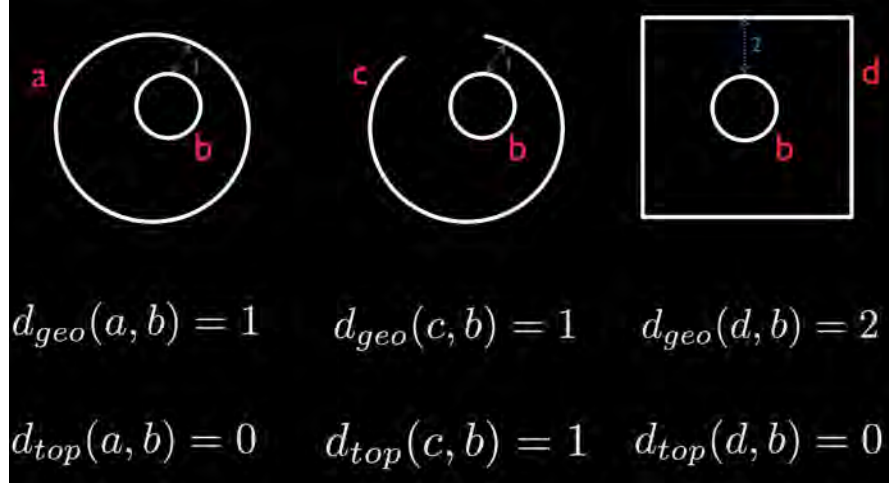
Fig. 4 Comparison between geometric distance d_{geo} and topological distance d_{top}. We used the shortest distance between objects as the geometric distance. The left and middle objects are topologically different while the left and right objects are topologically equivalent. The geometric distance cannot discriminate topologically different objects (left and middle) and produces false negatives. The geometric distance incorrectly discriminate topologically equivalent objects (left and right)and produces false positives.

3.1 Wasserstein distance

Given two probability distributions $X \sim f_1$ and $Y \sim f_2$, the *r-Wasserstein distance* D_W, which is the probabilistic version of the optimal transport, is defined as

$$D_W(f_1, f_2) = \left(\inf \mathbb{E}|X - Y|^r \right)^{1/r},$$

where the infimum is taken over every possible joint distributions of X and Y. The Wasserstein distance is the optimal expected cost of transporting points generated from f_1 to those generated from f_2 (Canas and Rosasco, 2012). There are numerous distances and similarity measures defined between probability distributions such as the Kullback-Leibler (KL) divergence and the mutual information (Kullback and Leibler, 1951). While the Wasserstein distance is a metric satisfying positive definiteness, symmetry, and triangle inequality, the KL-divergence and the mutual information are not metric. Although they are easy to compute, the biggest limitation of the KL-divergence and the mutual information is that the two probability distributions has to be defined on the same sample space. If the two distributions do not have the same support, it may be difficult to even define them. If f_1 is discrete while f_2 is continuous, it is difficult to define them. On the other hand, the Wasserstein distance can be computed for any arbitrary distributions that may not have the common sample space making it extremely versatile.

Consider persistent diagrams P_1 and P_2 given by

$$P_1 : x_1 = (b_1^1, d_1^1), \cdots, x_q = (b_q^1, d_q^1), \quad P_2 : y_1 = (b_1^2, d_1^2), \cdots, y_q = (b_q^2, d_q^2).$$

Their empirical distributions are given in terms of Dirac-Delta functions

$$f_1(x) = \frac{1}{q} \sum_{i=1}^q \delta(x - x_i), \quad f_2(y) = \frac{1}{q} \sum_{i=1}^q \delta(y - y_i).$$

Then we can show that the *r-Wasserstein distance* on persistent diagrams is given by

$$D_W(P_1, P_2) = \inf_{\psi : P_1 \to P_2} \left(\sum_{x \in P_1} \|x - \psi(x)\|^r \right)^{1/r} \tag{2}$$

over every possible bijection ψ, which is permutation, between P_1 and P_2 (Vallender, 1974; Canas and Rosasco, 2012; Berwald et al., 2018). Optimization (2) is the standard assignment problem, which is usually solved by Hungarian algorithm in $O(q^3)$ (Edmonds and Karp, 1972). However, for graph filtration, the distance can be computed *exactly* in $O(q \log q)$ by simply matching the order statistics on the birth or death values (Rabin et al., 2011; Songdechakraiwut and Chung, 2023; Songdechakraiwut et al., 2021):

Theorem 2 *The r-Wasserstein distance between the 0D persistent diagrams for graph filtration is given by*

$$D_{W0}(P_1, P_2) = \left[\sum_{i=1}^{q_0} (b_{(i)}^1 - b_{(i)}^2)^r \right]^{1/r},$$

where $b_{(i)}^j$ is the i-th smallest birth values in persistent diagram P_j. The 2-Wasserstein distance between the 1D persistent diagrams for graph filtration is given by

$$D_{W1}(P_1, P_2) = \left[\sum_{i=1}^{q_1} (d_{(i)}^1 - d_{(i)}^2)^r \right]^{1/r},$$

where $d_{(i)}^j$ is the i-th smallest death values in persistent diagram P_j.

3.2 Topological mean and variance

Given a collection of graphs $\mathcal{X}_1 = (V, w^1), \cdots, \mathcal{X}_n = (V, w^n)$ with edge weights $w^k = (w_{ij}^k)$, the usual approach for obtaining the average network $\bar{\mathcal{X}}$ is simply averaging the edge weigth matrices in an element-wise fashion

$$\bar{X} = \left(V, \frac{1}{n} \sum_{k=1}^{n} w_{ij}^k \right).$$

However, such average is the average of the connectivity strength. It is not necessarily the average of underlying topology. Such an approach is usually sensitive to topological outliers (Chung et al., 2019a). We address the problem through the Wasserstein distance. A similar concept was proposed in persistent homology literature through the Wasserstein barycenter (Agueh and Carlier, 2011; Cuturi and Doucet, 2014), which is motivated by Fréchet mean (Le and Kume, 2000; Turner et al., 2014; Zemel and Panaretos, 2019; Dubey and Müller, 2019). However, the method has not seen many applications in modeling graphs and networks.

In this study, we use the squared sum of 0D and 1D Wasserstein distances between networks X_1 and X_2:

$$\mathcal{D}(X_1, X_2) = D_{W0}^2(P_1, P_2) + D_{W1}^2(P_1, P_2)$$

as the Wasserstein distance between graphs. However, the sum does not uniquely define networks. Just like a toy example in Figure 4, we can have many topological equivalent brain networks that gives the identical distance. Thus, the average of two graphs is also not uniquely defined. The situation is analogous to Fréchet mean, which often does not yield the unique mean (Le and Kume, 2000; Turner et al., 2014; Zemel and Panaretos, 2019; Dubey and Müller, 2019). We define the *topological mean* of networks as the minimizer with respect to the Wasserstein distance. This is analogous to the situation where the sample mean is the minimizer of Euclidean distance. The squared Wasserstein distance is translation invariant such that

$$\mathcal{D}(c + X_1, c + X_2) = \mathcal{D}(X_1, X_2).$$

If we scale connectivity matrices by c, we have

$$\mathcal{D}(cX_1, cX_2) = c^2 \mathcal{D}(X_1, X_2).$$

Definition 1 The topological mean $\mathbb{E}X$ of networks $X_1, \cdots, X_n$ is the graph given by

$$\mathbb{E}X = \arg \min_{X} \sum_{k=1}^{n} \mathcal{D}(X, X_k).$$

Unlike the sample mean, we can have many different networks with identical topology that gives the minimum. Similarly we can define the *topological variance* $\mathbb{V}X$ as

Definition 2 The topological variance $\mathbb{V}X$ of networks $X_1, \cdots, X_n$ is the graph given by

$$\mathbb{V}X = \frac{1}{n} \sum_{k=1}^{n} \mathcal{D}(\mathbb{E}X, X_k).$$

The topological variance can be interpreted as the variability of graphs from the topological mean $\mathbb{E}\mathcal{X}$. To compute the topological mean and variance easily, we only need to identify a network with identical topology as the topological mean or the topological variance.

Theorem 3 *Consider graphs* $\mathcal{X}_i = (V, w^i)$ *with corresponding birth-death decompositions* $W_i = W_{ib} \cup W_{id}$ *with birth sets* $W_{ib} = \{b^i_{(1)}, \cdots, b^i_{(q_0)}\}$ *and death sets* $W_{id} = \{d^i_{(1)}, \cdots, d^i_{(q_1)}\}$. *Then, there exists topological mean* $\mathbb{E}\mathcal{X} = (V, w)$ *with birth-death decomposition* $W_b \cup W_d$ *with* $W_b = \{b_1, \cdots, b_{q_0}\}$ *and* $W_d = \{d_1, \cdots, d_{q_1}\}$ *satisfying*

$$b_j = \frac{1}{n}\sum_{i=1}^{n} b^i_{(j)}, \quad d_j = \frac{1}{n}\sum_{i=1}^{n} d^i_{(j)}.$$

3.3 Topological clustering

There are few studies that used the Wasserstein distance for clustering (Mi et al., 2018; Yang et al., 2020). The existing methods are mainly applied to geometric data without topological consideration or persistence. It is not obvious how to apply such geometric methods to cluster graph or network data. We propose to use the Wasserstein distance to cluster collection of graphs $\mathcal{X}_1, \cdots, \mathcal{X}_n$ into k clusters $C_1, \cdots, C_k$ such that

$$\cup_{i=1}^{k} C_i = \{\mathcal{X}_1, \cdots, \mathcal{X}_n\}, \quad C_i \cap C_j = \emptyset.$$

The total number of ways of partitioning n data points into k nonempty clusters is the *Stirling number of the second kind* $S_{n,k}$ (Lord et al., 2017; Rennie and Dobson, 1969)

$$S_{n,k} = \frac{1}{k!}\sum_{i=0}^{k}(-1)^i\binom{k}{i}(k-i)^n$$

satisfying recurrence

$$S_{n+1,k} = S_{n,k-1} + kS_{n,k}.$$

There are $S_{n,1} = 1$ 1-clusters, $S_{n,2} = 2^{n-1} - 1$ 2-clusters and $S_{n,n-1} = \frac{n(n-1)}{2}$ possible $(n-1)$-clusters out of n data points. Asymptotically $S_{n,k}$ increases exponentially as (Lord et al., 2017)

$$S_{n,k} \sim \frac{k^n}{k!}.$$

Brute-force approaches for searching for every possible clusters is not feasible for large n. We propose to a more scalable approach for clustering following k-means (Hartigan and Wong, 1979).

Let $C = (C_1, \cdots, C_k)$ be the collection of clusters. Let μ_j be the *topological cluster mean* within C_j given by

$$\mu_j = \arg\min_X \sum_{\mathcal{X}_k \in C_j} \mathcal{D}(X, \mathcal{X}_k).$$

The cluster mean is computed through Theorem 3. Just like Fréchet mean, the cluster mean is not unique in a geometric sense but only unique in a topological sense (Turner et al., 2014; Le and Kume, 2000; Zemel and Panaretos, 2019; Dubey and Müller, 2019). Let $\mu = (\mu_1, \cdots, \mu_k)$ be the cluster mean vector. The within-cluster distance from the cluster centers is given by

$$l_W(C; \mu) = \sum_{j=1}^{k} \sum_{X \in C_j} \mathcal{D}(X, \mu_j). \tag{3}$$

If we let $|C_j|$ to be the number of networks within cluster C_j, (3) can be written as

$$l_W(C; \mu) = \sum_{j=1}^{k} |C_j| \mathbb{V}_j \mathcal{X}, \tag{4}$$

with topological cluster variance

$$\mathbb{V}_j \mathcal{X} = \frac{1}{|C_j|} \sum_{X \in C_j} \mathcal{D}(X, \mu_j)$$

within cluster C_j. The optimal cluster is found by minimizing within-cluster distance $l_W(C; \mu)$ in (3) over every possible partitions of C.

If μ is given and fixed, the identification of clusters C can be done easily by assigning each network to the closest mean. Thus the topological clustering algorithm can be written as the two-step optimization similar to the expectation maximization (EM) algorithm often used in variational inferences and likelihood methods (Bishop, 2006). The first step computes the cluster mean. The second step minimizes the within-cluster distance. Just like k-means clustering, the two-step optimization is then iterated till convergence. Such process converges locally.

Theorem 4 *The topological clustering converges locally.*

Proof. The direct algebraic proof is fairly involving. So we use more intuitive geometric argument that relates to k-means clustering. Let $\mathbf{b}_i = (b_{(1)}^i, b_{(2)}^i \cdots, d_{(q_0)}^i)^\top$ be the vector of sorted birth values of network $\mathcal{X}_i$:

$$b_{(1)}^i < b_{(2)}^i < \cdots < b_{(q_0)}^i.$$

The 0D Wasserstein distance between connected component is given by

$$D_{W0}^2(P_i, P_j) = \sum_{k=1}^{q_0} (b_{(k)}^i - b_{(k)}^j)^2 = \mathbf{b}_i^\top \mathbf{b}_j.$$

Note that the sorted vector $\mathbf{b}_i$ is a point in the $(q_0 + 1)$-simplex $\mathcal{T}_0$ given by

$$\mathcal{T}_0 = \{(x_1, x_2, \cdots, x_{q_0}) | x_1 < x_2 < \cdots < x_{q_0}\} \subset \mathbb{R}^{q_0}.$$

Hence, the 0D Wasserstein distance is equivalent to Euclidean distance in the q-dimensional convex set $\mathcal{T}_0$. Similarly, let $\mathbf{d}_i = (d_{(1)}^i, d_{(2)}^i \cdots, d_{(q_1)}^i)^\top$ be the vector of sorted death values of network $\mathcal{X}_i$:

$$d_{(1)}^i < d_{(2)}^i < \cdots < d_{(q_1)}^i.$$

The 1D Wasserstein distance between two cycles is given by

$$D_{W1}^2(P_i, P_j) = \sum_{k=1}^{q_1} (d_{(k)}^i - d_{(k)}^j)^2 = \mathbf{d}_i^\top \mathbf{d}_j.$$

The sorted vector $\mathbf{d}_i$ is a point in the $(q_1 + 1)$-simplex $\mathcal{T}_1$ given by

$$\mathcal{T}_1 = \{(x_1, x_2, \cdots, x_{q_1}) | x_1 < x_2 < \cdots < x_{q_1}\} \subset \mathbb{R}^{q_1}.$$

Hence, the 1D Wasserstein distance is equivalent to Euclidean distance in the q-dimensional convex set $\mathcal{T}_1$.

Note the vector $(\mathbf{b}_i, \mathbf{d}_j) \in \mathbb{R}^{q_0 + q_1}$ is then a point in $\mathcal{T}_0 \otimes \mathcal{T}_1$, the Cartesian product of convex sets, which is again convex. Now consider our combined Wasserstein distance

$$\mathcal{D}(P_i, P_j) = \mathbf{b}_i^\top \mathbf{b}_j + \mathbf{d}_i^\top \mathbf{d}_j,$$

which is the squared Euclidean distance in the convex set $\mathcal{T}_0 \otimes \mathcal{T}_1$. Thus, any linear operation on the Wasserstein distance can be done as if it is the Euclidean distance within $\mathcal{T}_0 \otimes \mathcal{T}_1$. Thus, our topological clustering is equivalent to k-means clustering restricted to the convex set $\mathcal{T}_0 \otimes \mathcal{T}_1$. The convergence of topological clustering is then the direct consequence of the convergence of k-means clustering, which always converges in such a convex space. Thus, we can minimize (3) by replacing the Wasserstein distance with the 2-norm between sorted vectors of birth and death values in k-means clustering.

Numerical implementation. Just like k-means clustering algorithm that only converges to local minimum, there is no guarantee the topological clustering converges to the global minimum (Huang et al., 2020). This is remedied by repeating the algorithm multiple times with different random seeds and taking the smallest possible minimums. The method is implemented as the Matlab function `WS_cluster.m` which inputs the collection of networks and outputs the cluster labels and clustering accuracy. Different choice of initial cluster centers may lead to different results. So the algorithm may stuck in a local minimum and may not converge to global minimum. Thus, in actual numerical implementation, we used different initializations of centers. Then picked the best one clustering results with the smallest within cluster distance l_W.
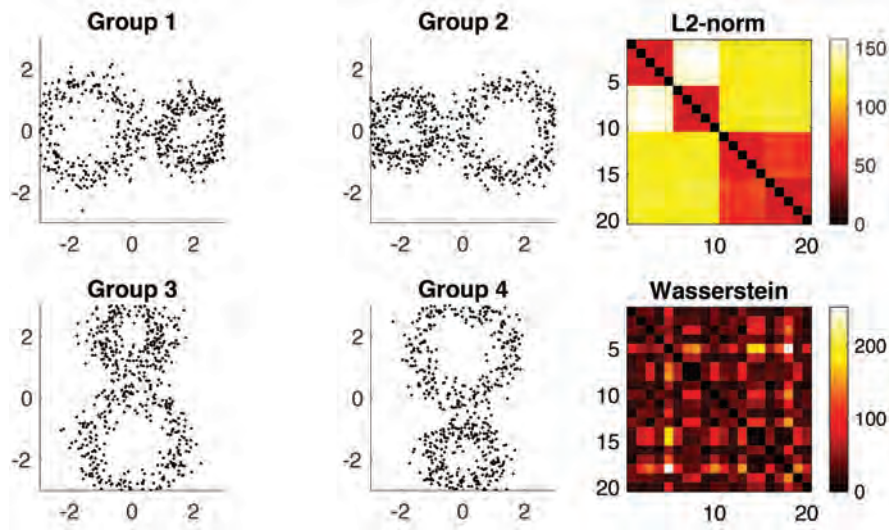
Fig. 5 Simulation study on topological equivalence. The correct clustering method should *not* be able to cluster them since they are all topologically equivalent. Right: the pairwise Euclidean distance (L_2-norm) is used in k-means and hierarchical clustering. The Wasserstein distance is used in topological clustering.

4 Validation

We validated the topological clustering in a simulation with the ground truth against k-means and hierarchical clustering (Lee et al., 2011). We generated 4 circular patterns of identical topology (Figure 5) and different topology (Figure 6). Along the circles, we uniformly sampled 60 nodes and added Gaussian noise $N(0, 0.3^2)$ on the coordinates. We generated 5 random networks per group. The Euclidean distance (L_2-norm) between randomly generated points are used to build connectivity matrices for k-means and hierarchical clustering. Figures 5 and 6 shows the superposition of nodes from 20 networks. For k-means and Wasserstein graph clustering, the average result of 100 random seeds are reported.

We tested for false positives when there is no topology difference in Figure 5, where all the groups are simply obtained from Group 1 by rotations. All the groups are topologically equivalent and thus we should not detect any topological difference. Any detected signals are all false positives. The k-means had 0.90 ± 0.15 while the hierarchical clustering had perfect 1.00 accuracy. Existing clusterng methods based on Euclidean distance are reporting significant false positives and should not be used in topological clustering task had the accuracy 0.43 ± 0.039. On the other hand, the Wasserstein graph clustering had low 0.53 ± 0.08 accuracy. We conclude that Wasserstein graph clustering are not reporting topological false positive like k-means and hierarchical clusterings.
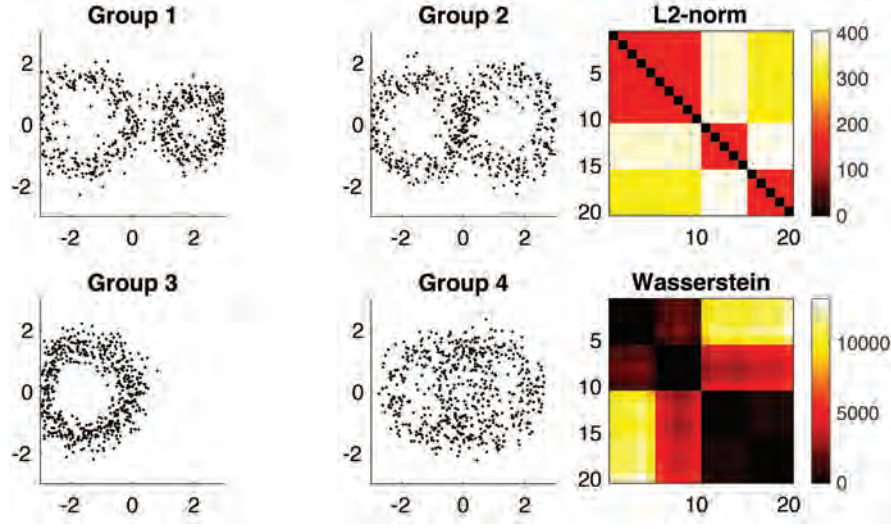
Fig. 6 Simulation study on topological difference. The correct clustering method should be able to cluster them since they are all topologically different. Right: the pairwise Euclidean distance (L_2-norm) is used in k-means and hierarchical clustering. The Wasserstein distance is used in topological clustering.

We also tested for false negatives when there is topology difference in Figure 6, where all the groups have different number of cycles. All the groups are topologically different and thus we should detect topological differences. The k-means clustering achieved 0.83 ± 0.16 accuracy. The hierarchical clustering is reporting perfect 1.00 accuracy. On the other hand, the topological clustering achieved respectable 0.98 ± 0.09 accuracy. It is extremely difficult to separate purely topological signals from geometric signals. Thus, when there is topological difference, it is expected to have geometric signal. Thus, all the methods are expected to perform well.

Existing clustering methods based on geometric distances will likely to produce significant amount of false positives and and not suitable for topological learning tasks. On the other hand, the proposed Wasserstein distance performed extremely well in both cases and not likely to report false positives or false negatives.

5 Application

The proposed method is applied in the accurate estimation of state spaces in dynamically changing functional brain networks. The 479 subjects resting-state functional magnetic resonance images (rs-fMRI) used in this paper were collected on a 3T MRI scanner (Discovery MR750, General Electric Medical Systems, Milwaukee, WI, USA) with a 32-channel RF head coil array. The 479 healthy subjects consist

of 231 males and 248 females ranging in age from 13 to 25 years were used. It also has 132 monozygotic (MZ) twin pairs and 93 same-sex dizygotic (DZ) twin pairs. The image preprocessing includes motion corrections and image alignment to the MNI template and exactly follows (Burghy et al., 2016; Jenkinson et al., 2002). The resulting rs-fMRI consist of $91 \times 109 \times 91$ isotropic voxels at 295 time points. We further parcellated the brain volume into 116 non-overlapping brain regions from a widely used AAL atlas (Tzourio-Mazoyer et al., 2002). The fMRI data were averaged across voxels within each brain region, resulting in 116 average fMRI signals with 295 time points for each subject. The rs-fMRI signals were then scaled to fit to unit interval [0, 1] and treated as functional data in [0, 1].

5.1 Weighted Fourier series representation

The most common approach in computing time-varying correlation in time series data is through SW, where correlations between brain regions are computed over the windows (Allen et al., 2014; Hutchison et al., 2013; Shakil et al., 2016; Mokhtari et al., 2019; Huang et al., 2020). However, the use of discrete windows can induce unnecessary high-frequency fluctuations in dynamic correlations (Oppenheim et al., 1999), though in some cases tapering can mitigate this effect (Allen et al., 2014). Further, correlation computation within windows is sensitive to outliers (Devlin et al., 1975).

To address these problems, we performed the Weighted Fourier series (WFS) representation that generalizes the cosine Fourier transform with the additional exponential weight that smooths out high frequency noises while reducing the Gibbs phenomenon (Chung et al., 2007; Huang et al., 2019a). WFS further avoids using sliding windows (SW) in computing correlations over time. For persistent homology method to work robustly across different subjects and time points, such signal denoising methods are needed. Consider arbitrary noise signal $f(t), t \in [0, 1]$ which will be denoised through diffusion.

Theorem 5 *The unique solution to 1D heat diffusion:*

$$\frac{\partial}{\partial s} h(t, s) = \frac{\partial^2}{\partial t^2} h(t, s) \tag{5}$$

on unit interval [0, 1] *with initial condition* $h(t, s = 0) = f(t)$ *is given by WFS:*

$$h(t, s) = \sum_{l=0}^{\infty} e^{-l^2 \pi^2 s} c_{fl} \psi_l(t), \tag{6}$$

where $\psi_0(t) = 1$, $\psi_l(t) = \sqrt{2} \cos(l\pi t)$ *are the cosine basis and* $c_{fl} = \int_0^1 f(t)\psi_l(t)dt$ *are the expansion coefficients.*

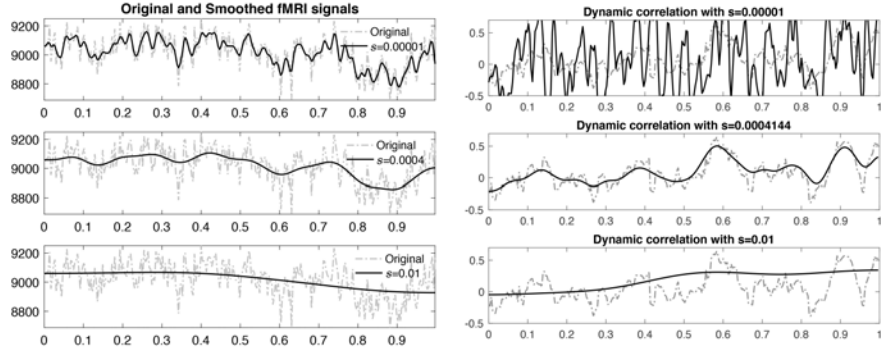The algebraic derivation is given in (Chung et al., 2007). Note the cosine basis is orthonormal

Fig. 7 Left: The original and smoothed fMRI time series using WFS with degree $L = 295$ and different heat kernel bandwidth s. The bandwidth 4.141×10^{-4} is used in this study approximately matches 20 TRs often used in the sliding window methods. Right: Doted gray lines are correlations computed over sliding windows. The solid black lines are correlations computed using WFS.

$$\langle \psi_l, \psi_m \rangle = \int_0^1 \psi_l(t) \psi_m(t) \; dt = \delta_{lm},$$

where δ_{lm} is Kroneker-detal taking value 1 if $l = m$ and 0 otherwise. We can rewrite (6) as a more convent convolution form

$$h(t, s) = \int_0^1 K_s(t, t') f(t') dt',$$

where heat kernel $K_s(t, t')$ is given by

$$K_s(t, t') = \sum_{l=0}^{\infty} e^{-l^2 \pi^2 s} \psi_l(t) \psi_l(t').$$

The diffusion time s is usually referred to as the kernel bandwidth and controls the amount of smoothing. Heat kernel satisfies $\int_0^1 K_s(t, t') \; dt = 1$ for any t' and s.

To reduce unwanted boundary boundary effects near the data boundary $t = 0$ and $t = 1$ (Huang et al., 2019a, 2020), we project the data onto the circle C with circumference 2 by the mirror reflection:

$$g(t) = f(t) \text{ if } t \in [0, 1], \quad g(t) = f(2 - t) \text{ if } t \in [1, 2].$$

Then perform WFS on the circle.

Theorem 6 *The unique solution to 1D heat diffusion:*

$$\frac{\partial}{\partial s} h(t, s) = \frac{\partial^2}{\partial t^2} h(t, s) \tag{7}$$

on the circle C with the initial periodic condition $h(t, s = 0) = f(t)$ if $t \in [0, 1]$, $h(t, s = 0) = f(2 - t)$ if $t \in [1, 2]$ is given by WFS:

$$h(t, s) = \sum_{l=0}^{\infty} e^{-l^2 \pi^2 s} c_{fl} \psi_l(t), \tag{8}$$

where $\psi_0(t) = 1$, $\psi_l(t) = \sqrt{2} \cos(l\pi t)$ are the cosine basis and $c_{fl} = \int_0^1 f(t)\psi_l(t)dt$ are the expansion coefficients.

Proof The cosine basis is defined on interval $[0, 1]$. We extend the domain of the basis by mirror reflection $\tilde{\psi}(t) = \psi(t)/\sqrt{2}$ in $[0, 1]$ and $\tilde{\psi}(t) = \psi(2 - t)/\sqrt{2}$ for $t \in [1, 2]$. Since $\tilde{\psi}(2) = \tilde{\psi}(0)$, the extended basis $\tilde{\psi}$ is a proper basis on circle C. The basis is scaled to have orthonormality:

$$\langle \tilde{\psi}_l, \tilde{\psi}_m \rangle = \int_0^1 \psi_l(t)\psi_m(t) \ dt + \int_1^2 \psi_l(2 - t)\psi_m(2 - t) \ dt = \delta_{lm}.$$

Subsequently, we can also extend the heat kernel as $\tilde{K}_s(t, t') = K_s(t, t')/2$ if $t' \in [0, 1]$ and $\tilde{K}_s(t, t') = K_s(t, 2 - t')/2$ if $t' \in [1, 2]$. The extended heat kernel satisfies

$$\int_0^2 K_s(t, t')dt' = \int_0^1 K_s(t, t')dt' + \int_1^2 K_s(t, 2 - t')dt' = 1.$$

Then, the solution to (7) is given by heat kernel convolution (Chung et al., 2007)

$$\begin{aligned} h(t, s) &= \int_0^2 \tilde{K}_s(t, t')h(t', s = 0) \ dt' \\ &= \int_0^1 \frac{1}{2} K_s(t, t')f(t')dt' + \int_1^2 \frac{1}{2} K_s(t, 2 - t')f(2 - t')dt' \\ &= \int_0^1 K_s(t, t')g(t')dt'. \end{aligned}$$

Hence, heat kernel smoothing on the circle with mirror reflection symmetry can be simply done by applying WFS in unit interval $[0, 1]$. $\square$

Numerical implementation. The cosine series coefficients c_{fl} are estimated using the least squares method by setting up a matrix equation (Chung et al., 2007). We set the expansion degree to equate the number of time points, which is 295. The window size of 20 TRs were used in most sliding window methods (Allen et al., 2014; Lindquist, 2014; Huang et al., 2020). We matched the full width at half maximum (FWHM) of heat kernel to the window size numerically. We used the fact that diffusion time s in heat kernel approximately matches to the kernel bandwidth of Gaussian kernel $e^{-t^2/2\sigma^2}$ as $\sigma = s^2/2$ (page 144 in (Chung, 2012)). 20 TRs is approximately equivalent to heat kernel bandwidth of about $4.144 \cdot 10^{-4}$ in terms of FWHM. Figure 7 displays the WFS representation of rsfMRI with different kernel bandwidths.

5.2 Dynamic correlation on weighted Fourier series

The weighted Fourier series representation provides a way to compute correlations dynamically without using sliding windows. Consider time series $x(t)$ and $y(t)$ with heat kernel $K_s(t, t')$. The mean and variance of signals with respect to the heat kernel are given by

$$\mathbb{E}x(t) = \int_0^1 K_s(t,t')x(t') \ dt'. \quad \mathbb{V}x(t) = \int_0^1 K_s(t,t')x^2(t') \ dt' - \big[\mathbb{E}x(t)\big]^2$$

Subsequently, the correlation $w(t)$ of $x(t)$ and $y(t)$ is given by

$$w(t) = \frac{\int_0^1 K_s(t,t')x(t')y(t') \ dt - \mathbb{E}x(t)\mathbb{E}y(t)}{\sqrt{\mathbb{V}x(t)}\sqrt{\mathbb{V}y(t)}}.$$

When the kernel is shaped as a sliding window, the correlation $w(t)$ exactly matches the correlation computed over the sliding window. The kernelized correlation generalizes the concept of integral correlations with the additional weighting term (Huang et al., 2019b). As $s \to \infty$, $w(t)$ converges to the Pearson correlation computed over the whole time points. Thus, the kernel bandwidth behaves like the length of sliding window.

Theorem 7 *The correlation $w(t)$ of time series $x(t)$ and $y(t)$ with respect to heat kernel $K_s(t, t')$ is given by*

$$w(t) = \frac{\sum_{l=0}^{\infty} e^{-l^2\pi^2 s}c_{xyl}\psi_l(t) - \mu_x(t)\mu_y(t)}{\sigma_x(t)\sigma_y(t)}, \tag{9}$$

with

$$\mu_x(t) = \sum_{l=0}^{\infty} e^{-l^2\pi^2 s}c_{xl}\psi_l(t), \quad \sigma_x^2(t) = \sum_{l=0}^{\infty} e^{-l^2\pi^2 s}c_{xxl}\psi_l(t) - \mu_x^2(t).$$

$$c_{xl} = \int_0^1 x(t)\psi_l(t)dt, \quad c_{yl} = \int_0^1 y(t)\psi_l(t)dt$$

are the cosine series coefficients. Similarly we expand $x(t)y(t)$, $x^2(t)$ and $y^2(t)$ using the cosine basis and obtain coefficients c_{xyl}, c_{xxl} and c_{yyl}.

The derivation follows by simply replacing all the terms with the WFS representation. Correlation (9) is the formula we used to compute the dynamic correlation in this study. Figure 7 displays the WFS-based dynamic correlation for different bandwidths. A similar weighted correlation was proposed in (Pozzi et al., 2012), where time varying exponential weights proportional to $e^{t/\theta}$ with exponential decay factor θ. However, our exponential weight term is related to the spectral decomposition of heat kernel in the spectral domain and invariant over time. The WFS based correlation is not related to (Pozzi et al., 2012).
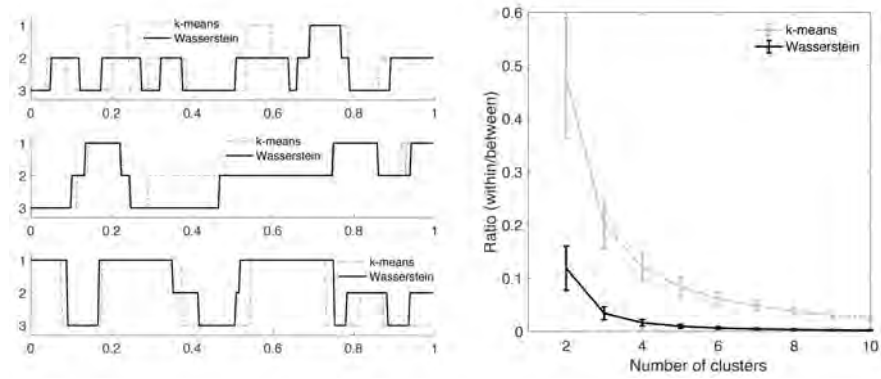
Fig. 8 Left: The time series of estimated state spaces using the topological clustering and k-means clustering for 3 subjects. The time is normalized into unit interval $[0, 1]$. Right: The ratio of within-cluster to between-cluster distances. Smaller the ratio, better the clustering fit is.

5.3 Topological state space estimation

For p brain regions, we estimated $p \times p$ dynamically changing correlation matrices $C_i(t)$ for the i-th subject using WFS. Let $\mathbf{C}_{ij}$ denote the vectorization of the upper triangle of $p \times p$ matrix $C_i(t_j)$ at time point t_j into $p^2 \times 1$ vector. For each fixed i, the collection of $\mathbf{C}_{ij}$ over $T = 295$ time points is then feed into topological clustering in identifying the recurring brain connectivity states that is common across subjects at the group level. We are clustering individually brain networks without putting any constraint on group or twin. We compared the proposed Wassertein clustering against the k-means clustering, which has been often used baseline method in the state space modeling (Allen et al., 2014; Huang et al., 2019a, 2020). After clustering, each correlation matrix $C_i(t_j)$ is assigned integers between 1 and k. These discrete states serve as the basis of investigating the dynamic pattern brain connectivity (Ting et al., 2018). For the convergence of both topological clustering and k-means clustering, the clusterings were repeated 10 times with different initial centroids and the best result (smallest within-cluster distance) is reported. Figure 8-left displays the result of the topological clustering against the k-means in few brain regions for a subject. 295 time points are rescaled to fit into unit interval $[0, 1]$.

The optimal number of cluster k was determined by the *elbow method* (Allen et al., 2014; Rashid et al., 2014; Ting et al., 2018; Huang et al., 2020). For each value of k, we computed the ratio of the within-cluster to between-cluster distances. The ratio shows the goodness-of-fit of the cluster model. The optimal number of clusters were determined by the elbow method, which gives the largest slope change in the ratio. $k = 3$ gives the largest slope in the both methods (Figure 8-right). At $k = 3$, the ratio is 0.034 ± 0.012 for 479 subjects for Wasserstein while it is 0.202 ± 0.047 for the k-means. The *six* times smaller ratio for the topological clustering demonstrates the superior model fit over k-means. Figure 9 shows the results of clustering for both
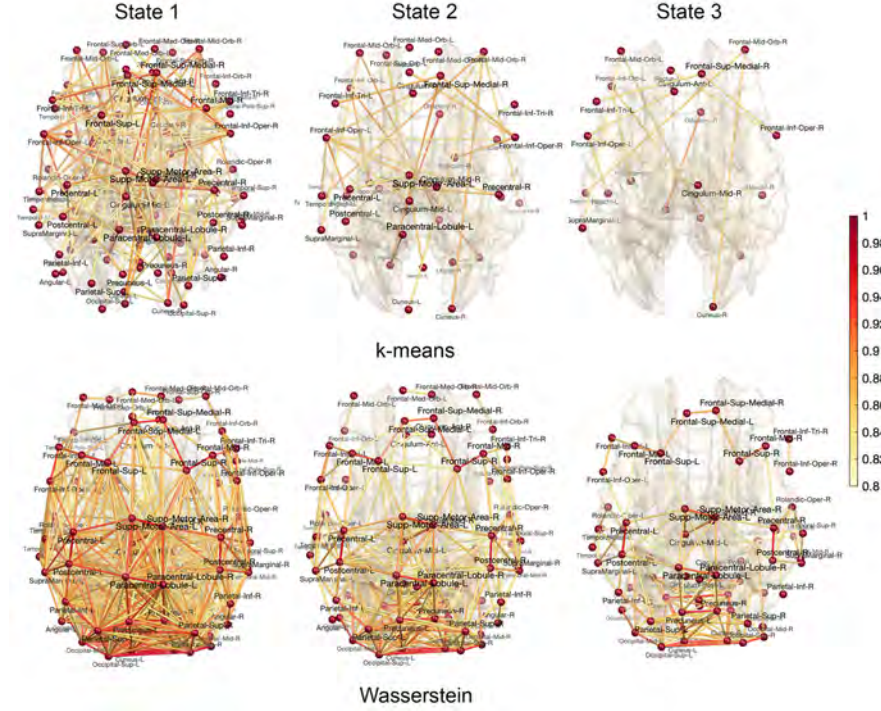
Fig. 9 The estimated state spaces of dynamically changing brain networks. The correlations are averaged over every time point and subject within each state for k-means clustering (top) and Wasserstein distance based topological clustering (bottom). In k-means clustering, the connectivity pattern of each state is somewhat random. In topological clustering, the connectivity pattern is highly symmetric even though we did not put any symmetry constraint in the clustering method.

methods. The k-means clustering results is based on averaging correlations of every time point and subject within each state. The resulting state is somewhat random without any biological interpretable pattern. The topological clustering computes the *topological mean* of every time point and subject within each state.

5.4 Twin correlations over transpositions

Using additional twin information in the data, we further we investigated if the state change pattern itself is genetically heritable for the first time. This is not yet reported in existing literature. This requires computing twin correlations. We assume there are m MZ- and n DZ-twins. For some feature, let $x_i = (x_{i1}, x_{i2})^\top$ be the i-th twin pair in MZ-twin and $y_i = (y_{i1}, y_{i2})^\top$ be the i-th twin pair in DZ-twin. They are represented as
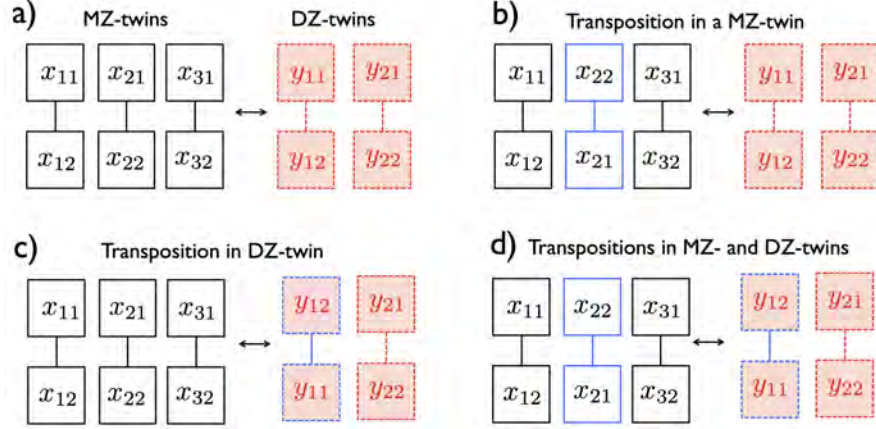
Fig. 10 The schematic of transpositions on 3 MZ- and 2 DZ-twins. a) One possible pairing. b) Transposition within a MZ-twin. c) Transposition within a DZ-twin. d) Transpositions in both MZ- and DZ-twins simulataneously. Any transposition will affect the heritability estimate so it is necessary to account for as many transpositions as possible.

$$\mathbf{x} = \begin{pmatrix} x_{11}, & \cdots & , x_{m1} \\ x_{12}, & \cdots & , x_{m2} \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_{11}, & \cdots & , y_{n1} \\ y_{12}, & \cdots & , y_{n2} \end{pmatrix}.$$

Let $\mathbf{x}_j$ be the j-th row of $\mathbf{x}$, i.e., $\mathbf{x}_j = (x_{1j}, x_{2j}, \cdots, x_{mj})$. Similarly let $\mathbf{y}_j = (y_{1j}, y_{2j}, \cdots, y_{nj})$. Then MZ- and DZ-correlations are computed as the sample correlation

$$\gamma^{MZ}(\mathbf{x}_1, \mathbf{x}_2) = corr(\mathbf{x}_1, \mathbf{x}_2)$$
$$\gamma^{DZ}(\mathbf{y}_1, \mathbf{y}_2) = corr(\mathbf{y}_1, \mathbf{y}_2).$$

However, there is no preference on the order of twins and we can *transpose* the i-th twin pair in MZ-twin such that

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

and obtain another twin correlation $\gamma^{MZ}(\tau_i(\mathbf{x}_1), \tau_i(\mathbf{x}_2))$ (Chen et al., 2018; Chung et al., 2019c). Ignoring symmetry, there are 2^m possible combinations in ordering the twins, which forms a permutation group. The size of the permutation group grows exponentially large as the sample size increases. Computing correlations over all permutations is not even computationally feasible for large m beyond 100. Figure 10 illustrates many possible transpositions within twins. Thus, we propose a new fast online computational strategy computing twin correlations.

Over transposition τ_i, the correlation changes

$$\gamma^{MZ}(\mathbf{x}_1, \mathbf{x}_2) \rightarrow \gamma^{MZ}(\tau_i(\mathbf{x}_1), \tau_i(\mathbf{x}_2))$$

incrementally. We will determine the exact increment over the transposition. The sample correlation between $\mathbf{x}_k$ and $\mathbf{x}_l$ involves the following functions.

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

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

The functions μ and ω are updated over transposition τ_i as

$$\nu(\tau_i(\mathbf{x}_k)) = \nu(\mathbf{x}_k) - x_{ik} + x_{il}$$

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

Then the MZ-twin correlation after transposition is updated as

$$\gamma^{MZ}(\tau_i(\mathbf{x}_1), \tau_i(\mathbf{x}_2)) = \frac{\omega(\tau_i(\mathbf{x}_1), \tau_i(\mathbf{x}_2))}{\sqrt{\omega(\tau_i(\mathbf{x}_1), \tau_i(\mathbf{x}_1))\omega(\tau_i(\mathbf{x}_2), \tau_i(\mathbf{x}_2))}}.$$

The time complexity for correlation computation is 33 operations per transposition, which is substantially lower than the computational complexity of directly computing correlations per permutation. In the numerical implementation, we sequentially apply random transpositions $\tau_{i_1}, \tau_{i_2}, \cdots, \tau_{i_J}$. This results in J different twin correlations, which are averaged. Let

$$\pi_1 = \tau_{i_1}, \pi_2 = \tau_{i_2} \circ \tau_{i_1}, \cdots, \pi_J = \tau_{i_J} \circ \cdots \circ \tau_{i_2} \circ \tau_{i_1}.$$

The average correlation $\overline{\gamma}_J^{MZ}$ of all J transpositions is given by

$$\overline{\gamma}_J^{MZ} = \frac{1}{J} \sum_{j=1}^{J} \gamma^{MZ}(\pi_{i_j}(\mathbf{x}_1), \pi_{i_j}(\mathbf{x}_2)).$$

In each sequential update, the average correlation can be updated iteratively as

$$\overline{\gamma}_J^{MZ} = \frac{J-1}{J}\overline{\gamma}_{J-1}^{MZ} + \frac{1}{J}\gamma^{MZ}(\pi_{i_J}(\mathbf{x}_1), \pi_{i_J}(\mathbf{x}_2)).$$

If we use enough number of transpositions, the average correlation $\overline{\gamma}_J^{MZ}$ converges to the true underlying twin correlation γ^{MZ} for sufficiently large J. DZ-twin correlation γ^{DZ} is estimated similarly.

In the widely used ACE genetic model, the heritability index (HI) h, which determines the amount of variation due to genetic influence in a population, is estimated using Falconer's formula (Falconer and Mackay, 1995; Chung et al., 2019b; Arbet et al., 2020). MZ-twins share 100% of genes while same-sex DZ-twins share 50% of genes on average. Thus, the additive genetic factor A and the common environmental factor C are related as
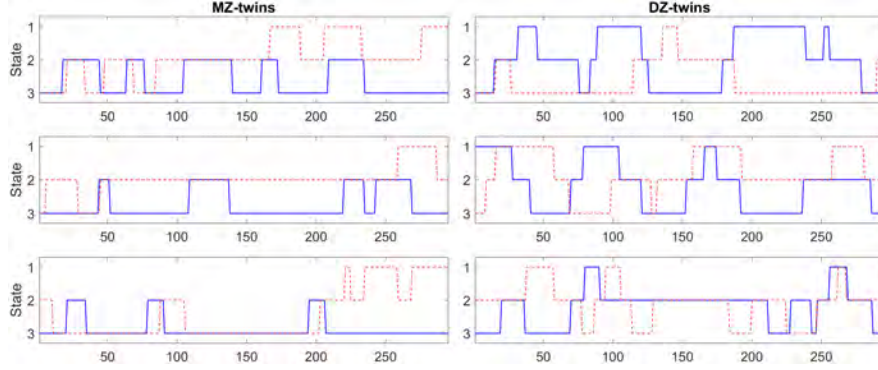
Fig. 11 State visits for 3 MZ-twins (left) and 3 DZ-twins (right) obtained from the baseline k-means clustering. We are interested in determining the heritability such state changed. Unfortunately, even within a twin, the time series of state change do not synchronize making the task extremely challenging.

$$\gamma^{MZ} = A + C,$$
$$\gamma^{DZ} = A/2 + C.$$

Thus HI h, which measures the contribution of A, is given by

$$h(\mathbf{x}, \mathbf{y}) = 2(\gamma^{MZ} - \gamma^{DZ}).$$

In numerical implementation, 100 million transpositions can be easily done in 100 seconds in a desktop. Similarly, we update the DZ-correlation over the transposition. The computer code for computing twin correlations is given as a function `transpose_corr` which inputs two input twin data matrices $\mathbf{x}$ and $\mathbf{y}$.

5.5 Heritability of the state space

The heritability estimation of state space is not a trivial task since the estimated state does not synchronize across twins making the task fairly difficult. Figure 11 displays the state visits in randomly selected 3 MZ- and 3 DZ-twins. However, the time series of state changes do not synchronize within twins. This is likely a reason there is almost no reported heritability of the state space in literature.

For each subject, we computed the average correlation of each state, where the average is taken over all time points within each state. The correlation matrices are then used as the input to the transposition based twin correlations (Chung et al., 2019b). Subsequently, we computed the MZ- and DZ-twin correlations within each state (Figure 12). The MZ-twin correlations (Figure 12-top) are densely observed in many connections while there is no DZ-twin correlations (Figure 12-middle) observed above 0.3. We then computed the heritability index (HI) of each state (Figure
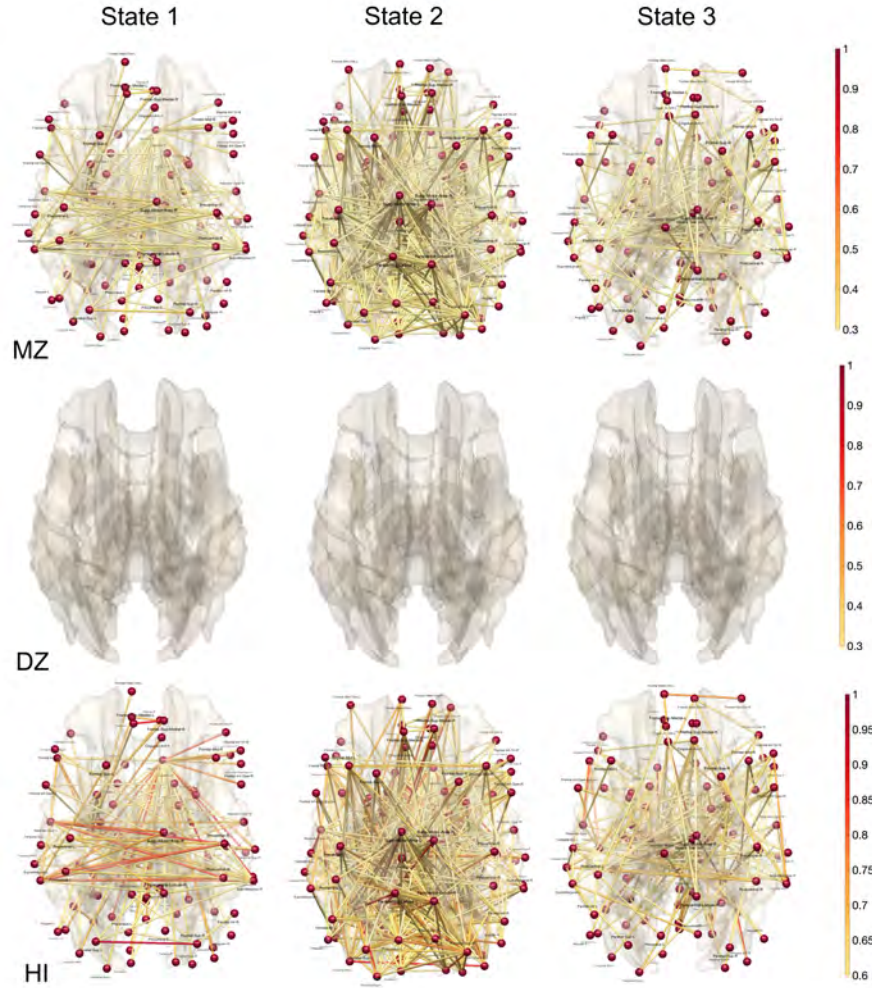
Fig. 12 MZ-correlation (top) and DZ-correlation (middle) in each state obtained through topological clustering in Figure 9. There is no MZ-correlation above 0.3 and not displayed. The heritability index (HI) is determined by the twice the difference in twin correlations. HI of each state shows extensive genetic contribution of dynamically changing states.

12-bottom). The heritability of the first state is characterized by strong lateralization of the hemisphere connections. The heritability of the second state is characterized by front and back connections. We believe the topological approach provides far more accurate and stable heritability index map for dynamically changing state that are biologically interpretable.

We reported 10 connections that give the highest HI values in all three states in Tables 1, 3 and 3. Although there are numerous studies reporting high heritability for anatomical features such as gray matter density, there are few rs-fMRI studies

reporting heritability of rs-fMRI (Glahn et al., 2010; Korgaonkar et al., 2014). Most of studies are reporting low HI compared to our high HI. (Glahn et al., 2010) reported HI of 0.104 in the left cerebellum, 0.304 in the right cerebellum, 0.331 in the left temporal parietal region, 0.420 in the right temporal parietal region. (Korgaonkar et al., 2014) reported HI of 0.41 in the connection between the posterior cingulate cortex and right inferior parietal cortex in the default mode network involving 79 MZ- and 46 same-sex DZ-twins. Other connections are all reporting very low HI below 0.24. We believe our topological method is clustering topological similar function network patterns and cluster them and significantly boosting genetic signal.

Regions	Regions	HI
Parietal-Sup-L	Parietal-Sup-R	0.96
Frontal-Sup-Medial-L	Frontal-Sup-Medial-R	0.90
Olfactory-R	Temporal-Mid-R	0.89
Precentral-R	Rolandic-Oper-L	0.88
Olfactory-R	Temporal-Inf-R	0.88
Olfactory-R	Fusiform-R	0.87
Olfactory-R	Cerebelum-4-5-L	0.87
Precentral-R	SupraMarginal-L	0.85
Rolandic-Oper-L	Postcentral-R	0.84
Olfactory-R	Lingual-L	0.84

Table 1 10 connections with the highest heritability index for state 1. Connections are sorted with respect to HI values.

Regions	Regions	HI
Hippocampus-L	Cerebelum-4-5-L	1.00
Olfactory-L	Fusiform-R	0.92
Precuneus-R	Cerebelum-Crus2-L	0.90
Occipital-Sup-L	Fusiform-L	0.89
Supp-Motor-Area-L	Cerebelum-Crus2-L	0.88
Occipital-Mid-L	Occipital-Mid-R	0.87
Thalamus-L	Cerebelum-9-L	0.86
Rolandic-Oper-L	Temporal-Sup-L	0.85
Paracentral-Lobule-L	Cerebelum-Crus2-L	0.85
Caudate-R	Cerebelum-Crus2-L	0.85

Table 2 10 connections with the highest heritability index for state 2. Connections are sorted with respect to HI values.

Regions	Regions	HI
Hippocampus-R	Cerebelum-3-R	1.00
Hippocampus-L	Cerebelum-4-5-L	0.93
Occipital-Mid-R	Cerebelum-Crus2-R	0.86
Olfactory-L	Cerebelum-3-L	0.81
Heschl-L	Temporal-Pole-Sup-L	0.81
Rolandic-Oper-L	Temporal-Pole-Sup-L	0.80
Caudate-R	Cerebelum-Crus1-L	0.79
Cerebelum-7b-R	Cerebelum-9-R	0.78
Cingulum-Ant-L	Cerebelum-3-R	0.78
Frontal-Mid-Orb-R	Frontal-Med-Orb-L	0.78

Table 3 10 connections with the highest heritability index for state 3. Connections are sorted with respect to HI values.

6 Discussion

In this study, the proposed the Wasserstein graph clustering for estimation and quantification of dynamic state changes in time varying networks. We developed a coherent statistical theory based on persistent homology and presented how such method is applied to the resting state fMRI data. The resting-state brain networks tend to remain in the same state for a long period before the transition to another state (Allen et al., 2014; Shakil et al., 2016; Calhoun and Adali, 2016). The average brain network in each state (Figure 9) does not follow similar connectivity patterns observed in the previous studies (Cai et al., 2018; Haimovici et al., 2017). Further research is needed for independent validation.

Intraclass correlation (ICC) has been often used as a reliability and reproducibility metric that gauges similarity in paired data when the ordering in pairing is not preserved (Chen et al., 2018). It has been a popular baseline metric for test and retest (TRT) reliability metric in various brain imaging applications along with Dice coefficients (Liao et al., 2013; Cole et al., 2014; Cousineau et al., 2017; Zhang et al., 2018). ICC is mainly computed through ANOVA statistical model, which can be fairly limited and inflexible in modeling. Over the years, various more complex statistical models mostly based on the mixed-effects model has been proposed for the better estimate of ICC (Chen et al., 2018). The proposed transposition based method can be used to compute ICC quickly.

Acknowledgement

References

M. Agueh and G. Carlier. Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43:904–924, 2011.

E.A. Allen, E. Damaraju, S.M. Plis, E.B. Erhardt, T. Eichele, and V.D. Calhoun. Tracking whole-brain connectivity dynamics in the resting state. *Cerebral cortex*, 24:663–676, 2014.

J. Arbet, M. McGue, and S. Basu. A robust and unified framework for estimating heritability in twin studies using generalized estimating equations. *Statistics in Medicine*, 2020.

D.S. Bassett and O. Sporns. Network neuroscience. *Nature neuroscience*, 20(3): 353–364, 2017.

J.J. Berwald, J.M. Gottlieb, and E. Munch. Computing wasserstein distance for persistence diagrams on a quantum computer. *arXiv:1809.06433*, 2018.

C.M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.

C.A. Burghy, M.E. Fox, M.D. Cornejo, D.E. Stodola, S.L. Sommerfeldt, C.A. Westbrook, C. Van Hulle, N.L. Schmidt, H.H. Goldsmith, R.J. Davidson, et al. Experience-driven differences in childhood cortisol predict affect-relevant brain function and coping in adolescent Monozygotic twins. *Scientific Reports*, 6:37081, 2016.

B. Cai, P. Zille, J.M. Stephen, T.W. Wilson, V.D. Calhoun, and Y.P. Wang. Estimation of dynamic sparse connectivity patterns from resting state fMRI. *IEEE Transactions on Medical Imaging*, 37:1224–1234, 2018.

V.D. Calhoun and T. Adali. Time-varying brain connectivity in fMRI data: whole-brain data-driven approaches for capturing and characterizing dynamic states. *IEEE Signal Processing Magazine*, 33:52–66, 2016.

G.D. Canas and L. Rosasco. Learning probability measures with respect to optimal transport metrics. *arXiv preprint arXiv:1209.1077*, 2012.

G. Chen, P.A. Taylor, S.P. Haller, K. Kircanski, J. Stoddard, D.S. Pine, E. Leibenluft, M.A. Brotman, and R.W. Cox. Intraclass correlation: Improved modeling approaches and applications for neuroimaging. *Human brain mapping*, 39:1187–1206, 2018.

M.K. Chung. *Computational Neuroanatomy: The Methods*. World Scientific, Singapore, 2012.

M.K. Chung, K.M. Dalton, L. Shen, A.C. Evans, and R.J. Davidson. Weighted Fourier representation and its application to quantifying the amount of gray matter. *IEEE Transactions on Medical Imaging*, 26:566–581, 2007.

M.K. Chung, J.L. Hanson, H. Lee, Nagesh Adluru, Andrew L. Alexander, R.J. Davidson, and S.D. Pollak. Persistent homological sparse network approach to detecting white matter abnormality in maltreated children: MRI and DTI multimodal study. *MICCAI, Lecture Notes in Computer Science (LNCS)*, 8149:300–307, 2013.

M.K. Chung, J.L. Hanson, L. Adluru, A.L. Alexander, R.J. Davidson, and S.D. Pollak. Integrative structural brain network analysis in diffusion tensor imaging. *Brain Connectivity*, 7:331–346, 2017a.

M.K. Chung, H. Lee, V. Solo, R.J. Davidson, and S.D. Pollak. Topological distances between brain networks. *International Workshop on Connectomics in Neuroimaging*, 10511:161–170, 2017b.

M.K. Chung, S.-G. Huang, A. Gritsenko, L. Shen, and H. Lee. Statistical inference on the number of cycles in brain networks. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 113–116. IEEE, 2019a.

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*, 3: 674–694, 2019b.

M.K. Chung, L. Xie, S.-G. Huang, Y. Wang, J. Yan, and L. Shen. Rapid acceleration of the permutation test via transpositions. 11848:42–53, 2019c.

J.H. Cole, R.E. Farmer, E.M. Rees, H.J. Johnson, C. Frost, R.I. Scahill, and N.Z. Hobbs. Test-retest reliability of diffusion tensor imaging in HuntingtonÕs disease. *PLoS Currents*, 6, 2014.

M. Cousineau, P.-M. Jodoin, E. Garyfallidis, M.-A. Côté, F.C. Morency, V. Rozanski, M. GrandÕMaison, B.J. Bedell, and M. Descoteaux. A test-retest study on Parkinson's PPMI dataset yields statistically significant white matter fascicles. *NeuroImage: Clinical*, 16:222–233, 2017.

M. Cuturi and A. Doucet. Fast computation of Wasserstein barycenters. In *International conference on machine learning*, pages 685–693. PMLR, 2014.

S.J. Devlin, R. Gnanadesikan, and J.R. Kettenring. Robust estimation and outlier detection with correlation coefficients. *Biometrika*, 62:531–545, 1975.

P. Dubey and H.-G. Müller. Fréchet analysis of variance for random objects. *Biometrika*, 106:803–821, 2019.

H. Edelsbrunner and J. Harer. *Computational topology: An introduction*. American Mathematical Society, 2010.

J. Edmonds and R.M. Karp. Theoretical improvements in algorithmic efficiency for network flow problems. *Journal of the ACM (JACM)*, 19:248–264, 1972.

D. Falconer and T Mackay. *Introduction to Quantitative Genetics, 4th ed.* Longman, 1995.

R. Ghrist. Barcodes: The persistent topology of data. *Bulletin of the American Mathematical Society*, 45:61–75, 2008.

D.C. Glahn, A.M. Winkler, P. Kochunov, L. Almasy, R. Duggirala, M.A. Carless, J.C. Curran, R.L. Olvera, A.R. Laird, and S.M. Smith. Genetic control over the resting brain. *Proceedings of the National Academy of Sciences*, 107:1223–1228, 2010.

A. Haimovici, E. Tagliazucchi, P. Balenzuela, and H. Laufs. On wakefulness fluctuations as a source of BOLD functional connectivity dynamics. *Scientific Reports*, 7:5908, 2017.

J.A. Hartigan and M.A. Wong. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (applied statistics)*, 28:100–108, 1979.

K.G. Hartmann, R.T. Schirrmeister, and T. Ball. Eeg-gan: Generative adversarial networks for electroencephalograhic (eeg) brain signals. *arXiv preprint arXiv:1806.01875*, 2018.

S.-G. Huang, M. K. Chung, I. C. Carroll, and H. H. Goldsmith. Dynamic functional connectivity using heat kernel. In *2019 IEEE Data Science Workshop (DSW)*, pages 222–226, 2019a. doi: 10.1109/DSW.2019.8755550.

S.-G. Huang, A. Gritsenko, M.A. Lindquist, and M.K. Chung. Circular pearson correlation using cosine series expansion. In *IEEE 16th International Symposium on Biomedical Imaging (ISBI)*, pages 1774–1777, 2019b.

S.-G. Huang, S.-T. Samdin, C.M. Ting, H. Ombao, and M.K. Chung. Statistical model for dynamically-changing correlation matrices with application to brain connectivity. *Journal of Neuroscience Methods*, 331:108480, 2020.

R.M. Hutchison, T. Womelsdorf, E.A. Allen, P.A. Bandettini, and V.D. et. al. Calhoun. Dynamic functional connectivity: promise, issues, and interpretations. *NeuroImage*, 80:360–378, 2013.

M. Jenkinson, P. Bannister, M. Brady, and S. Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17:825–841, 2002.

S.C. Johnson. Hierarchical clustering schemes. *Psychometrika*, 32:241–254, 1967.

M.S. Korgaonkar, K. Ram, L.M. Williams, J.M. Gatt, and S.M. Grieve. Establishing the resting state default mode network derived from functional magnetic resonance imaging tasks as an endophenotype: a twins study. *Human brain mapping*, 35: 3893–3902, 2014.

S. Kullback and R.A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22:79–86, 1951.

H. Le and A. Kume. The Fréchet mean shape and the shape of the means. *Advances in Applied Probability*, 32:101–113, 2000.

H. Lee, M.K. Chung, H. Kang, B.-N. Kim, and D.S. Lee. Computing the shape of brain networks using graph filtration and Gromov-Hausdorff metric. *MICCAI, Lecture Notes in Computer Science*, 6892:302–309, 2011.

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.

X.-H. Liao, M.-R. Xia, T. Xu, Z.-J. Dai, X.-Y. Cao, H.-J. Niu, X.-N. Zuo, Y.-F. Zang, and Y. He. Functional brain hubs and their test–retest reliability: a multiband resting-state functional mri study. *NeuroImage*, 83:969–982, 2013.

M Lindquist. Statistical and computational methods in brain image analysis. by Moo K. Chung. Boca Raton, Florida: CRC press. 2013. *Journal of the American Statistical Association*, 109:1334–1335, 2014.

E. Lord, M. Willems, F.-J. Lapointe, and V. Makarenkov. Using the stability of objects to determine the number of clusters in datasets. *Information Sciences*, 393:29–46, 2017.

L. Mi, W. Zhang, X. Gu, and Y. Wang. Variational wasserstein clustering. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 322–337, 2018.

F. Mokhtari, M.I. Akhlaghi, S.L. Simpson, G. Wu, and P.J. Laurienti. Sliding window correlation analysis: Modulating window shape for dynamic brain connectivity in resting state. *NeuroImage*, 189:655–666, 2019.

A.V. Oppenheim, R.W. Schafer, and J.R. Buck. *Discrete-time signal processing*. Upper Saddle River, NJ: Prentice Hall, 1999.

G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P.J. Hellyer, and F. Vaccarino. Homological scaffolds of brain functional networks. *Journal of The Royal Society Interface*, 11:20140873, 2014.

F. Pozzi, T. Di Matteo, and T. Aste. Exponential smoothing weighted correlations. *The European Physical Journal B*, 85:1–21, 2012.

J. Rabin, G. Peyré, J. Delon, and M. Bernot. Wasserstein barycenter and its application to texture mixing. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 435–446. Springer, 2011.

B. Rashid, E. Damaraju, G.D. Pearlson, and V.D. Calhoun. Dynamic connectivity states estimated from resting fMRI identify differences among schizophrenia, bipolar disorder, and healthy control subjects. *Frontiers in Human Neuroscience*, 8:897, 2014.

B.C. Rennie and A.J. Dobson. On stirling numbers of the second kind. *Journal of Combinatorial Theory*, 7:116–121, 1969.

D. Sabbagh, P. Ablin, G. Varoquaux, A. Gramfort, and D.A. Engemann. Manifold-regression to predict from meg/eeg brain signals without source modeling. *arXiv preprint arXiv:1906.02687*, 2019.

F.A.N. Santos, E.P. Raposo, M.D. Coutinho-Filho, M. Copelli, C.J. Stam, and L. Douw. Topological phase transitions in functional brain networks. *Physical Review E*, 100:032414, 2019.

S. Shakil, C.-H. Lee, and S.D. Keilholz. Evaluation of sliding window correlation performance for characterizing dynamic functional connectivity and brain states. *NeuroImage*, 133:111–128, 2016.

J. Shi, W. Zhang, and Y. Wang. Shape analysis with hyperbolic wasserstein distance. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5051–5061, 2016.

A.E. Sizemore, C. Giusti, A. Kahn, J.M. Vettel, R.F. Betzel, and D.S. Bassett. Cliques and cavities in the human connectome. *Journal of computational neuroscience*, 44:115–145, 2018.

T. Songdechakraiwut and M.K Chung. Dynamic topological data analysis for functional brain signals. *IEEE International Symposium on Biomedical Imaging Workshops*, 1:1–4, 2020.

T. Songdechakraiwut and M.K Chung. Topological learning for brain networks. *Annals of Applied Statistics*, 17:403–433, 2023.

T. Songdechakraiwut, L. Shen, and M.K. Chung. Topological learning and its application to multimodal brain network integration. *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 12902:166–176, 2021.

O. Sporns. *Graph Theory Methods for the Analysis of Neural Connectivity Patterns*, pages 171–185. Springer US, Boston, MA, 2003.

Z. Su, W. Zeng, Y. Wang, Z.-L. Lu, and X. Gu. Shape classification using wasserstein distance for brain morphometry analysis. In *International Conference on Information Processing in Medical Imaging*, pages 411–423. Springer, 2015.

C.-M. Ting, H. Ombao, S.B. Samdin, and S.-H. Salleh. Estimating dynamic connectivity states in fMRI using regime-switching factor models. *IEEE transactions on Medical imaging*, 37:1011–1023, 2018.

K. Turner, Y. Mileyko, S. Mukherjee, and J. Harer. Fréchet means for distributions of persistence diagrams. *Discrete & Computational Geometry*, 52:44–70, 2014.

N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in

spm using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15:273–289, 2002.

S.S. Vallender. Calculation of the Wasserstein distance between probability distributions on the line. *Theory of Probability & Its Applications*, 18:784–786, 1974.

Y. Wang, M.K. Chung, D. Dentico, A. Lutz, and R.J. Davidson. Topological network analysis of electroencephalographic power maps. In *International Workshop on Connectomics in NeuroImaging, Lecture Notes in Computer Science (LNCS)*, volume 10511, pages 134–142, 2017.

Y. Wang, H. Ombao, and M.K. Chung. Topological data analysis of single-trial electroencephalographic signals. *Annals of Applied Statistics*, 12:1506–1534, 2018.

B. C. M. Wijk, C. J. Stam, and A. Daffertshofer. Comparing brain networks of different size and connectivity density using graph theory. *PloS one*, 5:e13701, 2010.

Mengjia Xu, David Lopez Sanz, Pilar Garces, Fernando Maestu, Quanzheng Li, and Dimitrios Pantazis. A graph Gaussian embedding method for predicting Alzheimer's disease progression with MEG brain networks. *IEEE Transactions on Biomedical Engineering*, 68:1579–1588, 2021.

Z. Yang, J. Wen, and C. Davatzikos. Smile-GANs: Semi-supervised clustering via GANs for dissecting brain disease heterogeneity from medical images. *arXiv preprint*, arXiv:2006.15255, 2020.

J. Yoo, E.Y. Kim, Y.M. Ahn, and J.C. Ye. Topological persistence vineyard for dynamic functional brain connectivity during resting and gaming stages. *Journal of neuroscience methods*, 267:1–13, 2016.

K. Yoo, P. Lee, M.K. Chung, W.S. Sohn, S.J. Chung, D.L. Na, D. Ju, and Y. Jeong. Degree-based statistic and center persistency for brain connectivity analysis. *Human Brain Mapping*, 38:165–181, 2017.

Y. Zemel and V.M. Panaretos. Fréchet means and procrustes analysis in Wasserstein space. *Bernoulli*, 25:932–976, 2019.

Z. Zhang, M. Descoteaux, J. Zhang, G. Girard, M. Chamberland, D. Dunson, A. Srivastava, and H. Zhu. Mapping population-based structural connectomes. *NeuroImage*, 172:130–145, 2018.

A.J. Zomorodian. *Topology for computing*. Cambridge University Press, Cambridge, 2009.