

Topological Gradient of Structural Functional Brain Network in Frontotemporal Degeneration

Moo K. Chung, Jeffrey Phillips, James Gee

Abstract Our topological methods are applied to the structural brain networks of frontotemporal Degeneration (FTD) patients.

1 Methods

1.1 Birth-death decomposition of brain networks

A high dimensional object such as brain networks can be modeled as weighted graph $\mathcal{X} = (V, w)$ consisting of node set V indexed as $V = \{1, 2, \dots, p\}$ 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:

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

where $q \leq (p^2 - p)/2$. The subscript (\cdot) denotes the order statistic. In terms of sorted edge weight set $W = \{w_{(1)}, \dots, w_{(q)}\}$, we may also write the graph as $\mathcal{X} = (V, W)$. 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* \mathcal{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

$$\mathcal{X}_{\epsilon_0} \subset \mathcal{X}_{\epsilon_1} \subset \mathcal{X}_{\epsilon_2} \subset \dots$$

for $0 = \epsilon_0 < \epsilon_1 < \epsilon_2 < \dots$, 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

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

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.

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

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

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 $\epsilon > 0$. 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.

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 (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): $\beta_1(w_{(i)}) \geq \beta_1(w_{(i+1)})$. The method is applied to FTD subjects in showing the Betti curves for males and females (Figure 1).

During the graph filtration, when new components are born, they never die. 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 be partitioned into 0D birth values and 1D death values (Songdechakraiut et al., 2021):

Theorem 1 (Birth-death decomposition) (Songdechakraiut and Chung, 2023)
The edge weight set $W = \{w_{(1)}, \dots, w_{(q)}\}$ has the unique decomposition

$$W = W_b \cup W_d, \quad W_b \cap W_d = \emptyset \quad (1)$$

where birth set $W_b = \{b_{(1)}, b_{(2)}, \dots, b_{(q_0)}\}$ is the collection of 0D sorted birth values and death set $W_d = \{d_{(1)}, d_{(2)}, \dots, d_{(q_1)}\}$ 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 produce 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 $\{(b_{(1)}, \infty), \dots, (b_{(q_0)}, \infty)\}$. Ignoring ∞ , W_b is the 0D persistent diagram. The 1D persistent diagram of the graph filtration is given by $\{(-\infty, d_{(1)}), \dots, (-\infty, d_{(q_1)})\}$. Ignoring $-\infty$, W_d is the 1D persistent diagram. We can show that the birth set is the maximum spanning tree (MST) (Songdechakraiut and Chung, 2023).

1.2 Topological distance between brain networks

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. For this purpose we use the 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

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

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

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

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; Songdechakraiut and Chung, 2023; Songdechakraiut 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 r -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 .

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

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

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 (Aguech 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 \mathcal{X}_1 and \mathcal{X}_2 :

$$\mathcal{D}(\mathcal{X}_1, \mathcal{X}_2) = D_{W0}^2(P_1, P_2) + D_{W1}^2(P_1, P_2)$$

as the Wasserstein distance between graphs. However, inference procedure can be equally applicable if we use only one of distances such that

$$\mathcal{D}(\mathcal{X}_1, \mathcal{X}_2) = D_{W0}^2(P_1, P_2)$$

or

$$\mathcal{D}(\mathcal{X}_1, \mathcal{X}_2) = D_{W1}^2(P_1, P_2).$$

The method in the study is equally applicable regress of how $\mathcal{D}(\mathcal{X}_1, \mathcal{X}_2)$ is defined. The topological sum does not uniquely define networks in topological sense. The average of two graphs may 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

$$\mathcal{D}(c + \mathcal{X}_1, c + \mathcal{X}_2) = \mathcal{D}(\mathcal{X}_1, \mathcal{X}_2).$$

If we scale connectivity matrices by c , we have

$$\mathcal{D}(c\mathcal{X}_1, c\mathcal{X}_2) = c^2 \mathcal{D}(\mathcal{X}_1, \mathcal{X}_2).$$

Definition 1 The topological mean $\mathbb{E}\mathcal{X}$ of networks $\mathcal{X}_1, \dots, \mathcal{X}_n$ is the graph given by

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

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

Definition 2 The topological variance $\mathbb{V}\mathcal{X}$ of networks $\mathcal{X}_1, \dots, \mathcal{X}_n$ is the graph given by

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

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 $W_{ib} = \{b_{(1)}^i, \dots, b_{(q_0)}^i\}$ and death sets $W_{id} = \{d_{(1)}^i, \dots, d_{(q_1)}^i\}$. Then, there exists topological mean $\mathbb{E}\mathcal{X} = (V, w)$ with birth-death decomposition $W_b \cup W_d$ with $W_b = \{b_1, \dots, b_{q_0}\}$ and $W_d = \{d_1, \dots, d_{q_1}\}$ satisfying

$$b_j = \frac{1}{n} \sum_{i=1}^n b_{(j)}^i, \quad d_j = \frac{1}{n} \sum_{i=1}^n d_{(j)}^i. \quad (3)$$

1.3 Topological embedding via the birth-death decomposition

We can use the concept of topological mean in embedding each brain network as a scatter point in 2D plane. Consider topological mean $\mathbb{E}_0\mathcal{X}$ of networks $\mathcal{X}_1, \dots, \mathcal{X}_n$ computed using 0D distance $D_{W^{0^2}}$. The birth values of the $\mathbb{E}_0\mathcal{X}$ are given by (3) as

$$b_1 < b_2 < \dots < b_{q_0}.$$

Given sorted birth values $b_{(1)}^i, b_{(2)}^i, \dots, b_{(q_0)}^i$ for the i -th subject, we anchor the topological mean $\mathbb{E}_0\mathcal{X}$ to the x -coordinate

$$\mu_b = \frac{1}{q_0} \sum_{i=1}^{q_0} b_i.$$

μ_b is the average of birth values of all the subjects. The embedding x -coordinate of the k -th subject is then measured with respect to the topological mean $\mathbb{E}_0 \mathcal{X}$ as

$$\frac{1}{q_0} \sum_{i=1}^{q_0} (b_{(i)}^k - b_i).$$

Similarly, consider topological mean $\mathbb{E}_1 \mathcal{X}$ of networks $\mathcal{X}_1, \dots, \mathcal{X}_n$ computed using 1D distance D_{W1^2} . The death values of the $\mathbb{E}_1 \mathcal{X}$ are given by (3) as

$$d_1 < d_2 < \dots < d_{q_1}.$$

Given sorted death values $d_{(1)}^k, d_{(2)}^k, \dots, d_{(q_1)}^k$ for the k -th subject, we anchor the topological mean $\mathbb{E}_1 \mathcal{X}$ to the y -coordinate

$$\mu_d = \frac{1}{q_1} \sum_{i=1}^{q_1} d_i.$$

The embedding y -coordinate of the k -th subject is then measured with respect to the topological mean $\mathbb{E}_1 \mathcal{X}$ as

$$\frac{1}{q_1} \sum_{i=1}^{q_1} (d_{(i)}^k - d_i).$$

The embedding shows the relative topological distance with respect to the topological center of each group. The method is applied to FTD subjects in discriminating males and females (Figure 1-right).

1.4 Topological inference

There are very few studies that used the Wasserstein distance in brain imaging studies (Mi et al., 2018; Yang et al., 2020). The existing methods are mainly applied to geometric data without topological consideration. It is not obvious how to apply the method to perform statistical inference for a population study. We will present a new statistical inference procedure for testing the topological inference of two groups, the usual setting in brain network studies.

Consider a collection of graphs $\mathcal{X}_1, \dots, \mathcal{X}_n$ that are grouped into two groups C_1 and C_2 such that

$$C_1 \cup C_2 = \{\mathcal{X}_1, \dots, \mathcal{X}_n\}, \quad C_1 \cap C_2 = \emptyset.$$

We assume there are n_i graphs in C_i and $n_1 + n_2 = n$. In the usual statistical inference, we are interested in testing the null hypothesis of the equivalence of topological summary \mathcal{T} :

$$H_0 : \mathcal{T}(C_1) = \mathcal{T}(C_2).$$

Under the null, there are $\binom{n}{n_1}$ number of permutations to permute n graphs into two groups, which is an extremely large number and most computing systems including MATLAB/R cannot compute them exactly if the sample size is larger than 50 in each group. If $n_1 = n_2$, the total number of permutations is given asymptotically by Stirling's formula (Feller, 2008)

$$\binom{n}{n_1} \sim \frac{4^{n_1}}{\sqrt{\pi n_1}}.$$

The number of permutations *exponentially* increases as the sample size increases, and thus it is impractical to generate every possible permutation. In practice, up to hundreds of thousands of random permutations are generated using the uniform distribution on the permutation group with probability $1/\binom{n}{n_1}$. The computational bottleneck in the permutation test is mainly caused by the need to recompute the test statistic for each permutation. This usually cause a serious computational bottleneck when we have to recompute the test statistic for large samples for more than million permutations. We propose a more scalable approach.

Define the within-group distance \mathcal{L}_W as

$$2\mathcal{L}_W = \sum_{X_i, X_j \in C_1} \mathcal{D}(X_i, X_j) + \sum_{X_i, X_j \in C_2} \mathcal{D}(X_i, X_j).$$

The average within-group distance is then given by

$$\bar{\mathcal{L}}_W = \frac{\mathcal{L}_W}{n_1(n_1 - 1) + n_2(n_2 - 1)}.$$

The between-group distance \mathcal{L}_B is defined as

$$2\mathcal{L}_B = \sum_{X_i \in C_1} \sum_{X_j \in C_2} \mathcal{D}(X_i, X_j) + \sum_{X_i \in C_2} \sum_{X_j \in C_1} \mathcal{D}(X_i, X_j).$$

The average between-group distance is then given by

$$\bar{\mathcal{L}}_B = \frac{\mathcal{L}_B}{n_1 n_2}.$$

Note that the sum of within-group and between-group distance is the sum of all the pairwise distances:

$$2\mathcal{L}_W + 2\mathcal{L}_B = \sum_{i=1}^n \sum_{j=1}^n \mathcal{D}(X_i, X_j).$$

When we permute the group labels, the total sum of all the pairwise distances do not change and fixed. If the group difference is large, the between-group distance \mathcal{L}_B will be large and the within-group distance \mathcal{L}_W will be small. Thus, to measure the disparity between groups as the ratio (Songdechakriwut and Chung, 2023)

$$\phi_{\mathcal{L}} = \frac{\mathcal{L}_B}{\mathcal{L}_W}.$$

The ratio statistic is related to the elbow method in clustering and behaves like traditional F -statistic, which is the ratio of squared variability of model fits. If $\phi_{\mathcal{L}}$ is large, the groups differ significantly in network topology. If $\phi_{\mathcal{L}}$ is small, it is likely that there is no group differences.

Since the ratio is always positive, its probability distribution cannot be Gaussian. Since the distributions of the ratio $\phi_{\mathcal{L}}$ is unknown, the permutation test can be used to determine the empirical distributions. The p -value is the area of the right tail thresholded by the observed ratio $\phi_{\mathcal{L}}$ (dotted red line) in the empirical distribution. Since we only compute the pairwise distances only once and only shuffle each entry over permutations. The simple rearranging of rows and columns of entries and sum them in the block-wise fashion should be faster than the usual two-sample t test which has to be recomputed for each permutation.

To speed up the permutation further, we adapted the transposition test, the online version of permutation test (Chung et al., 2019b). In the transposition test, we only need to work out how \mathcal{L}_B and \mathcal{L}_W changes over a transposition, a permutation that only swaps one entry from each group. When we transpose k -th and l -th graphs between the groups (denoted as τ_{kl}), all the k -th and i -th rows and columns will be swapped. The within-group distance after the transposition τ_{kl} is given by

$$\tau_{kl}(\mathcal{L}_W) = \mathcal{L}_W + \Delta_W,$$

where Δ_W is the terms in the k -th and i -th rows and columns that are required to swapped. We only need to swap up to $O(2n)$ entries while the standard permutation test that requires the computation over $O(n^2)$ entries. Similarly we have incremental changes

$$\tau_{kl}(\mathcal{L}_B) = \mathcal{L}_B + \Delta_B.$$

The ratio statistic over the transposition is then sequentially updated over random transpositions. To further accelerate the convergence and avoid potential bias, we introduce one permutation to the sequence of 1000 consecutive transpositions.

2 Application

2.1 Data

We used 173 subjects frontotemporal Degeneration (FTD) patients collected at the University of Pennsylvania (Irwin et al., 2015). The fiber tractography is performed using the diffusion tensor imaging and connectivity matrices are built using the Brainnetome Atlas with 246 parcellations. The connectivity matrices of subjects with up to 3 repeat scans are averaged. FTD is a neurodegenerative disorder characterized by the progressive degeneration of the frontal and temporal lobes of the brain.

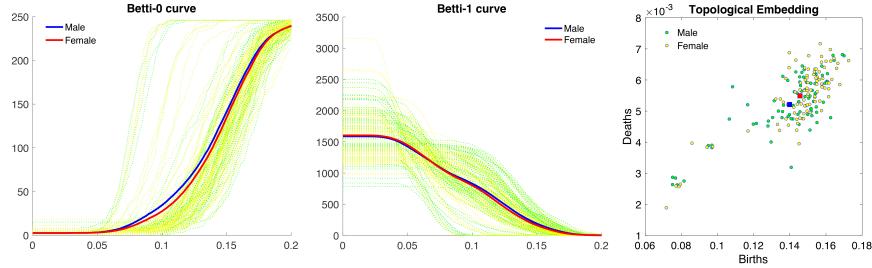


Fig. 1 The β_0 (left) and β_1 curves (middle) over thresholded edge weights (horizontal axis) for 173 subjects. Males are colored green while females are colored yellow. The thick lines are the mean curves. There is visibly consistent group separations particularly for β_0 (p -values = 0.0174) but less so in β_1 (p -values = 0.0172). Betti-1 curves show two different types of patterns (rapid vs. gradual declines) indicative of subpopulations with different pathology. Right: We also performed topological embedding to see if there is any visible group separation. Both 0D topology (horizontal) and 1D topology (vertical) can separate the groups easily. We can clearly see the spread of the topological means (blue square for male and red square for female).

Structural connectivity abnormalities are commonly observed in FTD and contribute to the clinical manifestations of the disease.

2.2 Topological outlier detection

The Betti curves can be easily used for detecting outliers in any type of imaging modalities including fMRI, MRI and DTI. We used the the Betti curves, which plots the topological invariants β_i over thresholded edge weights for identifying potential outliers (Figure 1). Since Betti curves are algebraically related to empirical cumulative distribution functions (CDF) through order statistics, the Kolmogorov-Smirnov (KS) distance can be used to quantify how much data is outlying from the average pattern (Chung et al., 2017a, 2019b, 2017b). Based on Betti curves, we did not detect significantly outlying data that diverge from the usual pattern of monotonic increase for β_0 -curve and monotonic decrease for β_1 -curve. However, compared to the usual healthy controls, there are significant variability in the β_1 -curve pattern demonstrating the heterogeneous nature of the population. it is likely there are two major subpopulations with distinct pathology.

2.3 Topological sex difference

We also tested the possible sex effect. There are total $n_M = 81$ males and $n_F = 92$ females in the dataset. Using the ratio statistic, we tested the statistical significance

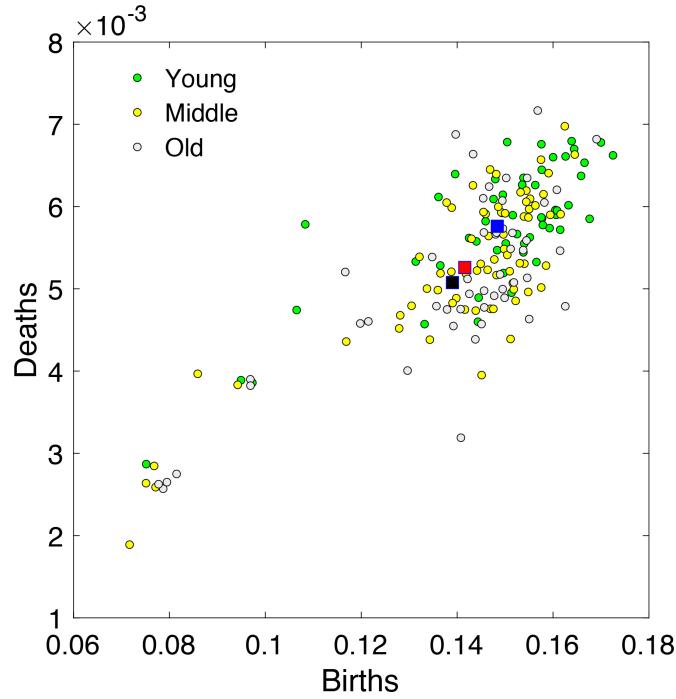


Fig. 2 173 subjects are binned into 3 age groups: young (ages below 40 years), middle (ages between 40 and 60 years) and old (ages above 60 years). Then the topological embedding is applied to each group. The blue, red and black squares are the topological centroids of young, middle and old groups. They show clear gradient of decreasing birth and death values over ages (p -value = 0.0034). Decreasing birth values implies there are more disconnected subnetworks at given edge thresholds. Decreasing death values implies cycles are disappearing with increasing age.

of the sex difference. For 0D topology, we obtained the p -value of 0.0174 with 10 million permutations indicating there is significant topological difference between the sexes. Females have more connected patterns across the whole brain compared to males. For 1D topology, we also obtained the p -value of 0.0172 with 10 million permutations indicating there is significant topological difference between the sexes. The topological embedding shows discernible group difference between males and females (Figure 1-right). The blue square for male and red square for female are the topological mean of males and females respectively. They show visible group separation.

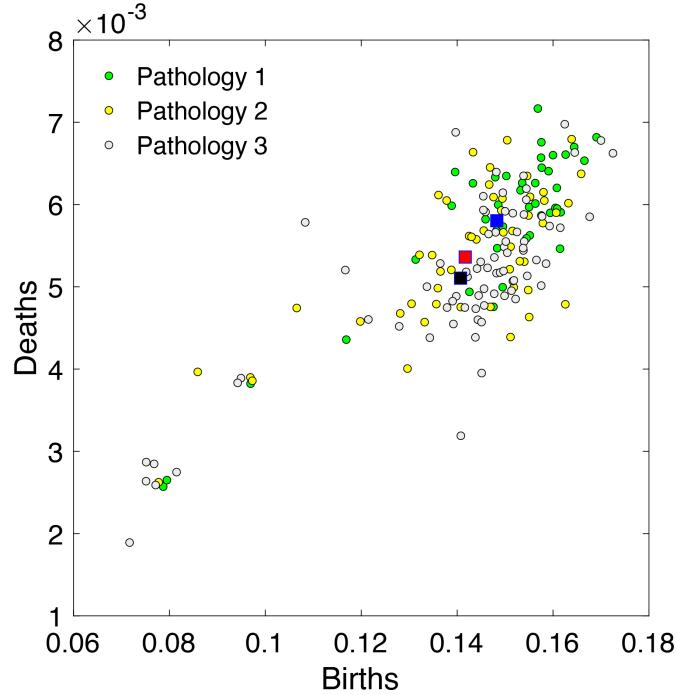


Fig. 3 The blue, red and black squares correspond to the topological centroid of three different pathologies. Pathologies 1 and 3 are topologically different (p -value = 0.0034).

2.4 Topological gradient over age

We binned 173 subjects into 3 age groups: Young (ages below 40 years), Middle (ages between 40 and 60 years) and Old (ages above 60 years). There are 53, 70 and 50 subjects in Young, Middle and Old groups respectively. Then the topological embedding is applied to each group (Figure 3). Subjects in Young are clustered at top right corner (high birth and death values) compared to older age groups. Subjects in Old are usually clustered in lower birth and death values. Subjects in Middle are scattered all over the range. The blue, red and black squares are the topological centroids of Young, Middle and Old. We tested the age effect of this trend using the proposed ratio test with 10 million transpositions. Between Young and Middle, p -value is 0.1803. Between Middle and Old, p -value is 0.3806. Between Young and Old, p -value is 0.0034. Given sufficient age range and samples, we should be able to detect significant topological gradient.

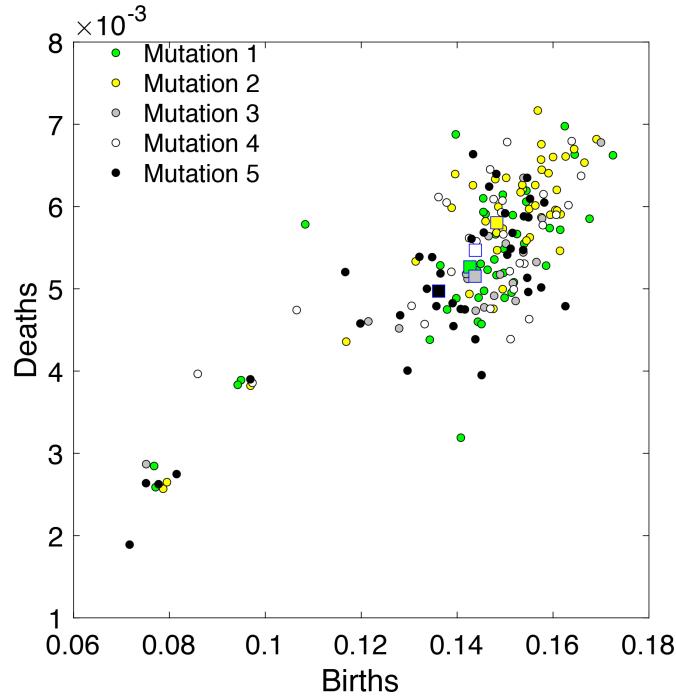


Fig. 4 The topological embedding of different mutations. Between Mutation 2 and 5, we have p -value 0.0018. Between Mutation 1 and 2, we have p -value 0.0326.

2.5 Topology over pathology

We tested if there is any topology difference between different pathology. There are three different pathologies that were blinded to Chung who performed the analysis. The same ratio test is used with 10 million transpositions. For testing difference in Pathologies 2 and 3, p -value is 0.3806. For testing difference in Pathologies 1 and 2, p -value is 0.1803. For testing difference in Pathologies 1 and 3, p -value is 0.0034.

2.6 Topology over mutations

We tested if there is any topology difference between different mutations. We performed the topological embedding (Figure 4) There are five different pathologies that were blinded to Chung who performed the analysis. The same ratio test is used with 10 million transpositions. The statistical significance is weak for testing topological difference between mutations. For instance, between Mutation 1 and 5, p -value is 0.3384. However, between Mutation 2 and 5, which has the largest topological

discrepancy, we have p -value 0.0018. Between Mutation 1 and 2, we have p -value 0.0326. If there is clear clustering pattern without much overlap in the topological embedding, we should detect group differences.

Acknowledgement

This study was supported by NIH EB028753 and NSF MDS-2010778.

References

- M. Aguech and G. Carlier. Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43:904–924, 2011.
- H.J. Aizenstein, M.A. Butters, M. Wu, L.M. Mazurkewicz, V.A. Stenger, P.J. Giannaros, J.T. Becker, C.F. Reynolds III, and C.S. Carter. Altered functioning of the executive control circuit in late-life depression: episodic and persistent phenomena. *The American Journal of Geriatric Psychiatry*, 17:30–42, 2009.
- J.J. Berwald, J.M. Gottlieb, and E. Munch. Computing wasserstein distance for persistence diagrams on a quantum computer. *arXiv:1809.06433*, 2018.
- G.D. Canas and L. Rosasco. Learning probability measures with respect to optimal transport metrics. *arXiv preprint arXiv:1209.1077*, 2012.
- A.E. Cavanna and M.R. Trimble. The precuneus: a review of its functional anatomy and behavioural correlates. *Brain*, 129:564–583, 2006.
- 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, 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, 2017a.
- M.K. Chung, V. Vilalta-Gil, H. Lee, P.J. Rathouz, B.B. Lahey, and D.H. Zald. Exact topological inference for paired brain networks via persistent homology. *Information Processing in Medical Imaging (IPMI), Lecture Notes in Computer Science (LNCS)*, 10265:299–310, 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, L. Xie, S.-G. Huang, Y. Wang, J. Yan, and L. Shen. Rapid acceleration of the permutation test via transpositions. *11848:42–53*, 2019b.
- M. Cuturi and A. Doucet. Fast computation of Wasserstein barycenters. In *International conference on machine learning*, pages 685–693. PMLR, 2014.
- I.J. Deary. Why do intelligent people live longer? *Nature*, 456(7219):175–176, 2008.

- I.J. Deary, S. Strand, P. Smith, and C. Fernandes. Intelligence and educational achievement. *Intelligence*, 35:13–21, 2007.
- 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.
- W. Feller. *An introduction to probability theory and its applications*, volume 2. John Wiley & Sons, 2008.
- R. Ghrist. Barcodes: The persistent topology of data. *Bulletin of the American Mathematical Society*, 45:61–75, 2008.
- 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.
- D.J. Irwin, N.J. Cairns, M. Grossman, C.T. McMillan, E.B. Lee, V.M. Van Deerlin, V. M.-Y. Lee, and J.Q. Trojanowski. Frontotemporal lobar degeneration: defining phenotypic diversity through personalized medicine. *Acta Neuropathologica*, 129: 469–491, 2015.
- S.C. Johnson. Hierarchical clustering schemes. *Psychometrika*, 32:241–254, 1967.
- R.H. Kaiser, J.R. Andrews-Hanna, T.D. Wager, and D.A. Pizzagalli. Large-scale network dysfunction in major depressive disorder: a meta-analysis of resting-state functional connectivity. *JAMA psychiatry*, 72:603–611, 2015.
- J. Kaufman, B. Birmaher, D. Brent, U. Rao, C. Flynn, P. Moreci, D. Williamson, and N. Ryan. Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry*, 36: 980–988, 1997.
- M.S. Korgaonkar, C. Antees, L.M. Williams, J.M. Gatt, R.A. Bryant, R. Cohen, R. Paul, R. O’Hara, and S.M. Grieve. Early exposure to traumatic stressors impairs emotional brain circuitry. *PLoS One*, 8:e75524, 2013.
- 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.
- J.D. Mayer, R.D. Roberts, and S.G. Barsade. Human abilities: emotional intelligence. *Annual Review of Psychology*, 59:507–536, 2008.

- 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.
- A.S. Mitchell, S.M. Sherman, M.A. Sommer, R.G. Mair, R.P. Vertes, and Y. Chudasama. Advances in understanding mechanisms of thalamic relays in cognition and behavior. *Journal of Neuroscience*, 34:15340–15346, 2014.
- 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.
- W.W. Seeley, V. Menon, A.F. Schatzberg, J. Keller, G.H. Glover, H. Kenna, A.L. Reiss, and M.D. Greicius. Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience*, 27:2349–2356, 2007.
- T. Songdechakriwut and M.K. Chung. Topological learning for brain networks. *Annals of Applied Statistics*, 17:403–433, 2023.
- T. Songdechakriwut, 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.
- K. Turner, Y. Mileyko, S. Mukherjee, and J. Harer. Fréchet means for distributions of persistence diagrams. *Discrete & Computational Geometry*, 52:44–70, 2014.
- S.S. Vallender. Calculation of the Wasserstein distance between probability distributions on the line. *Theory of Probability & Its Applications*, 18:784–786, 1974.
- C. van Heeringen, S. Bijtebier, and K. Godfrin. Suicidal brains: a review of functional and structural brain studies in association with suicidal behaviour. *Neuroscience and Biobehavioral Reviews*, 35:688–698, 2011.
- P. Vidal-Ribas, D. Janiri, G.E. Doucet, N. Pornpattananangkul, D.M. Nielson, S. Frangou, and A. Stringaris. Multimodal neuroimaging of suicidal thoughts and behaviors in a US population-based sample of school-age children. *American Journal of Psychiatry*, 178:321–332, 2021.
- A. Wiglesworth, C.A. Falke, M. Fiecas, M. Luciana, K.R. Cullen, and B. Klimes-Dougan. Brain signatures in children who contemplate suicide: learning from the large-scale ABCD study. *Psychological Medicine*, 53:2164–2173, 2023.
- Y. Wu, P. Besson, E.A. Azcona, S.K. Bandt, T.B. Parrish, H.C. Breiter, and A.K. Katsaggelos. A multicohort geometric deep learning study of age dependent cortical and subcortical morphologic interactions for fluid intelligence prediction. *Scientific reports*, 12:17760, 2022.
- 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.
- Y. Zemel and V.M. Panaretos. Fréchet means and procrustes analysis in Wasserstein space. *Bernoulli*, 25:932–976, 2019.
- A.J. Zomorodian. *Topology for computing*. Cambridge University Press, Cambridge, 2009.