

Topological Embedding of Functional Human Brain Network in Children Exhibiting Suicidal Tendency

Moo K. Chung, Anqi Qiu

Abstract The method is applied to the resting-state functional magnetic resonance images (rs-fMRI) in ABCD study collected in multiple different sites. However, the topological method is robust to variations due to acquisition and sex, and thus there is no need to account for sex and sex as nuisance covariates. However, we are able to localize brain regions that contribute the most to suicidal ideation and correlate with topology.

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 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): $\beta_1(w_{(i)}) \geq \beta_1(w_{(i+1)})$.

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 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 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 $\{(b_{(1)}, \infty), \dots, (b_{(q_0)}, \infty)\}$. 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_{W0^2} . 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

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

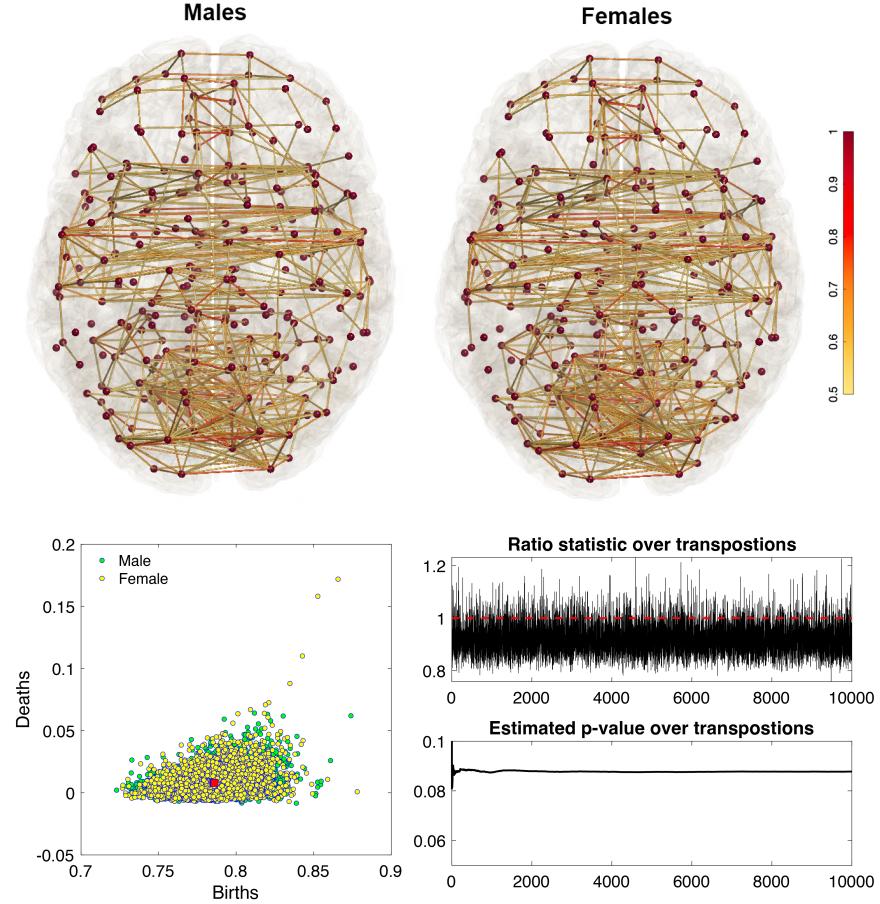


Fig. 1 Top: 7693 subjects are split into males and females. Bottom left: Topological embedding of 7693 subjects. There is no visible clustering difference between males and females. The topological centroid of each group (blue square for males and red squares for females) coincide. Bottom right: The observed ratio statistic is 1.0003, which is not strongly statistically significant (p -value = 0.088 over 100 million transpositions). Thus the sex effect will not be accounted for.

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.

We can clearly see more spread for TLE compared to HC. We also can see that 0D topology is the main topological discriminator separating the two groups while 1D topology may not be able to discriminate the groups. Similar observation was made in the huge β_0 -curve shape difference while there is almost no shape difference in β_1 -curve. The method is used to embed thousands networks for possible discrimination between male and female into a 2D plane (Figure 1). Unfortunately, there is no significant sex difference (p -value = 0.088).

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 X_1, \dots, X_n that are grouped into two groups C_1 and C_2 such that

$$C_1 \cup C_2 = \{X_1, \dots, X_n\}, \quad C_1 \cap C_2 = \emptyset.$$

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 (Songdechakraiut 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 be 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 description

We used the ongoing Adolescent Brain Cognitive Development (ABCD) study database — the largest long-term study of brain development and child health in US with more than 100 psychiatric and 11 cognitive measures (Huang et al., 2023). The brain imaging component of ABCD follows the HCP protocol and contains the resting-state fMRI and T1-weighted MRI Barch et al. (2018). Youth ($n = 11,875$) 9–11 years of age were recruited for the study. This age range is important as it is a period of development critical to an individual’s life trajectory. The incidence of psychiatric illnesses, such as attention deficit hyperactivity disorder (ADHD), anxiety, mood disorders, and psychosis, increases through adolescence (Paus et al., 2008) as well as cognitive variables such as the g-factor of general intelligence (Caspi et al., 2014; Parkes et al., 2021; Lahey et al., 2012) and fluid intelligence. The ABCD dataset is handy for understanding its risk factors and predictive models in the early onset of psychiatric disorders during adolescence. We will develop interpretable topology-based network models to predict general psychopathology and identify transdiagnostic patterns of clinical characteristics. In the study, we examined suicidal intent related variables available in ABCD.

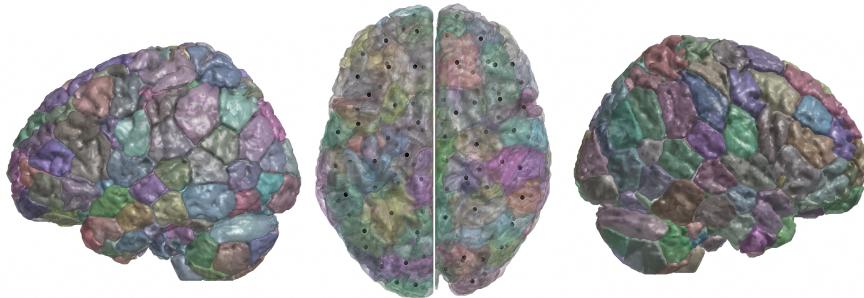


Fig. 2 Shen268 atals is used to parcellate the brain into 268 regions. The black dots are the centroid of each parcellation, where the brain network nodes are displayed.

This is highly significant in our society which is currently plagued with violent crimes. Studies have shown that offenders in school shooting incidents commonly show symptoms of depression and suicidal ideation (Vossekuil, 2002). It is reported that 61% of offenders had a history of depression and 78% had either attempted suicide or expressed suicidal thoughts prior to the incident (Vossekuil, 2002). In this project, we will fine-tune our proposed methods and use them to determine how genetics and childhood experiences affect suicidal ideation and identify biomarkers that will provide parents, teachers, health professionals and policymakers with practical information to promote the mental health of children.

2.2 Image preprocessing

During collection of resting state fMRI (rs-fMRI) data, children viewed a crosshair for 20 minutes (Vidal-Ribas et al., 2021). The connectivity matrices across different brain regions are computed using the Pearson correlation over whole time points using the Shen268 atlas. In the Shen268 atlas, the human brain is divided into 268 regions (134 per hemisphere) based on a group-wise spectral clustering algorithm using resting-state fMRI data from 45 subjects (Shen et al., 2013). The atlas contains 268 functional coherent nodes. Figures 2 displays the Shen268 parcellation where each region is color coded. The black dots are the mass center of parcellation where we are building 3D graph representation of brain networks.

2.3 No site effect

Since the images were collected in 18 different sites, we tested if there is any site effect based on the proposed ratio statistic on the Wasserstein distance. We compared brain connectivity difference across 18 sites. Figure 4 displays the p -values with one

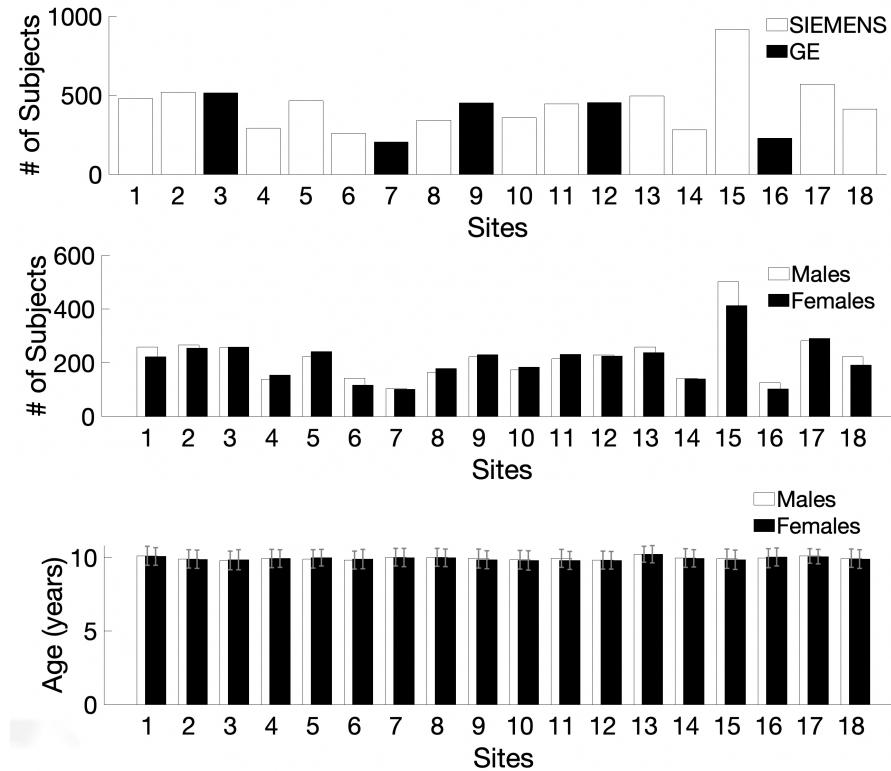


Fig. 3 Top: the total number of subjects scanned in each site. The subjects are scanned either Siemens or GE scanner. Middle: distribution of males and females in each site. Bottom: The age distribution in each site with mean and one standard deviation.

million transpositions across different sites plotted in $-\log_{10}$ scale. The smallest p -value is 0.007 between sites 1 and 12. However, none of p -values passed the multiple comparisons correction using the false discovery rate (FDR) at 0.01 or 0.05 levels. Thus, we conclude that there is no statistically significant site effect and we did not account for site effect in our analysis. The topological method does not penalize the geometric differences such as correlation differences but only topological differences and should be extremely robust for site differences.

2.4 No sex effect

We also tested sex effect. There are total $n_M = 3922$ males and $n_F = 3771$ females in the dataset. Using the ratio statistic, we tested the statistical significance of the sex difference. We obtained the p -value of 0.088 with 100 million transpositions indicating the topological method is not detecting strong sex effect. The topological

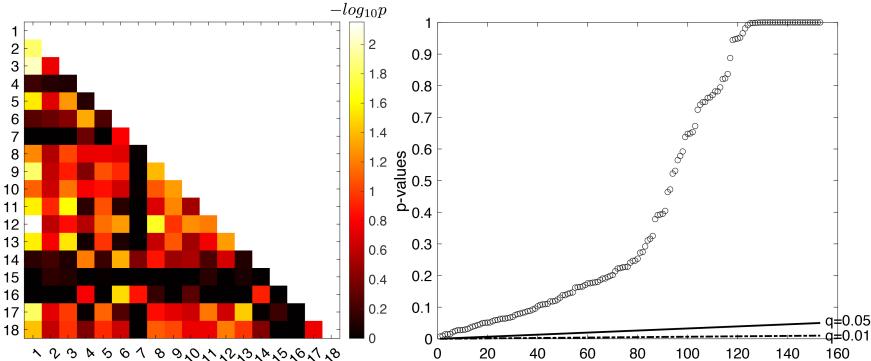


Fig. 4 Left: p -values testing for site effect plotted in $-\log_{10}$ scale. Right: sorted p -values and rejection lines at 0.05 and 0.01 level for false recovery rate (FDR). None of p -values are below the rejection lines. We conclude that there is no statistically significant site difference.

embedding shows almost no discernible group difference between males and females (Figure 1). Our topological method is very robust over sex differences. Thus, we did not account for sex effect in the subsequent analysis. Most of sex differences in the brain connectivity is likely due to connectivity strength differences that may not cause topological differences.

2.5 No age effect

We also tested age effect. We split subjects into age less than 115 months and age more than 125 months. The topological embedding is not showing any significant difference between the age groups. The observed ratio statistic is 0.98, which is not strongly statistically significant (p -value = 0.57 over 100 million transpositions). Thus the age effect will not be accounted for in the study. Even though this is age range for significant developmental changes, the change is likely to be scalar changes in the strength of connectivity without any major topological changes. In fact, the average correlation networks are almost identical (Figure 5).

2.6 The group-level resting-state brain connectivity

The individual functional connectivity is averaged at the edge level to obtain the group level average functional connectivity (Figure 6). The edges are thresholded at correlation 0.4. Most of connections are clustered within the 8 subnetworks of Shen268 atlas. The motor (blue) and visual I (yellow) are the most densely connected subnetworks along with medical frontal (teal) and cerebellum (red) areas. The node values are the mean correlation of all the connections at each node.

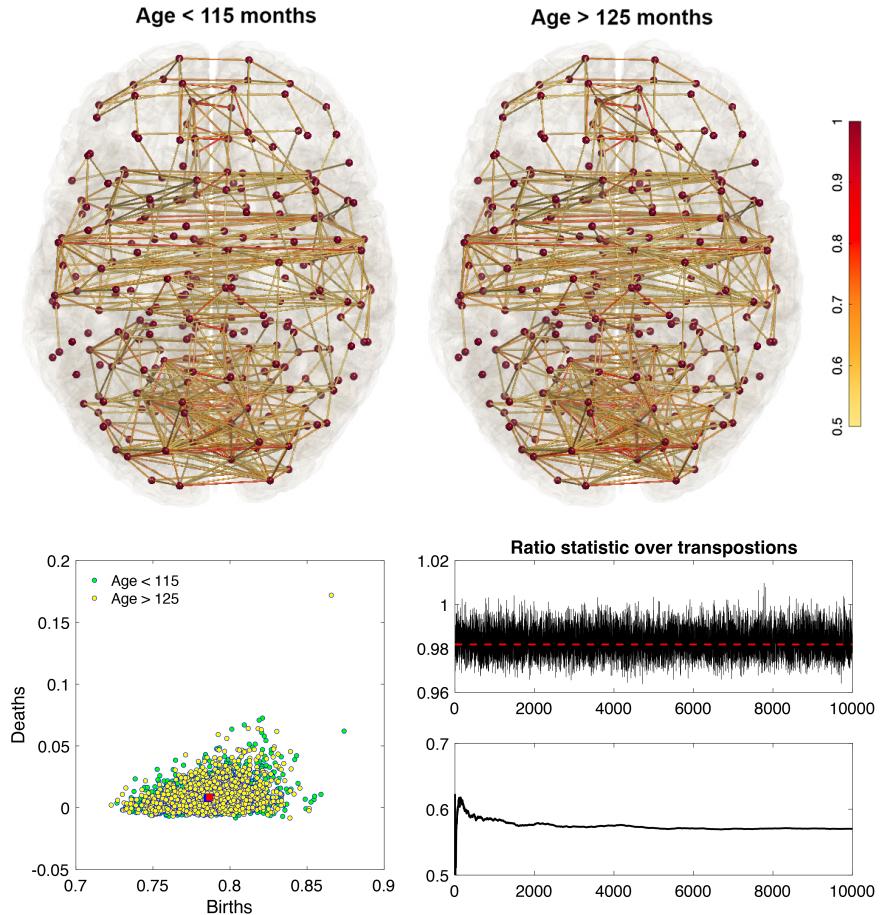


Fig. 5 Top: 7693 subjects are split into age less than 115 months and age more than 125 months. There is almost no cluster group differences. Bottom left: Topological embedding. There is no visible clustering difference between the groups. The topological centroid of each group coincide. Bottom right: The observed ratio statistic is 0.98, which is not strongly statistically significant (p -value = 0.57 over 100 million transpositions). Thus the age effect will not be accounted for.

2.7 Negative correlation of suicidal ideation and fluid intelligence

As part of this study, other measures are collected to better understand the factors that contribute to the development of mental health issues, including suicidal behavior. One of the variables collected in the ABCD study is **suicidality**, which assesses the frequency of suicidal ideation, suicidal plans, and suicide attempts (Wiglesworth et al., 2023). In the assessment of suicidality in children, both child-report and caregiver-report were obtained using the computerized Kiddie Schedule for Affective

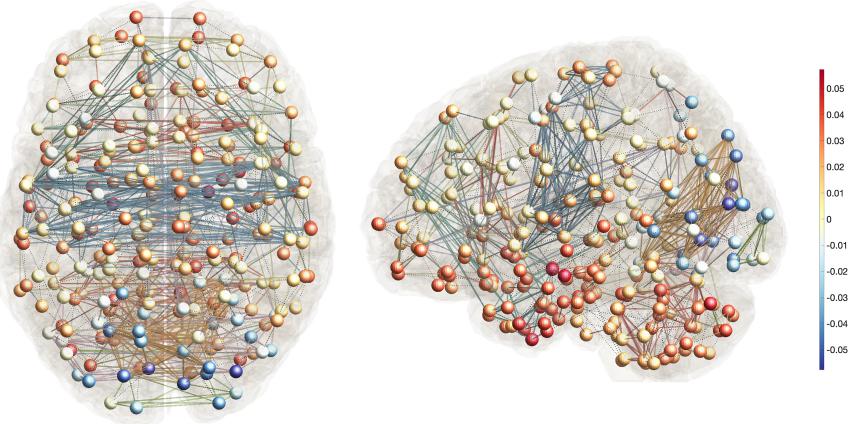


Fig. 6 The average functional brain networks of 7693 subjects. The average Pearson correlation is thresholded at 0.4 showing mostly connections within functional subnetworks, which are color coded as medial frontal (teal), frontoparietal (yellow), default mode (purple), subcortical and cerebebellum (red), motor (blue), visual I (orange), visual II (green), visual association (pink) networks. The colorbar is for mean correlation over all connections at each node. Occipital regions show negative correlations.

Disorders and Schizophrenia (KSADS) for DSM-5 (KSADS-5) (Kaufman et al., 1997). The KSADS-5 captures multiple dimensions of suicidality, including passive suicidal ideation, active but nonspecific suicidal ideation, suicidal ideation with a specific method, active suicidal ideation with intent, active suicidal ideation with a plan, preparatory actions toward imminent suicidal behavior, interrupted suicidal attempts, aborted suicidal attempts, and suicidal attempts (Vidal-Ribas et al., 2021). The suicidality variable is important because it can help researchers understand the prevalence and severity of suicidal behavior among adolescents, as well as the risk factors and protective factors that are associated with this behavior.

Among 7693 subjects, we have 6,638 subjects with the suicide ideation variable. Figure 7 displays the histogram of suicide ideation, which has exponentially decreasing distributions. There is no age effect on suicide ideation (p -value = 0.93). The ABCD Study uses a variety of tests and assessments to measure cognitive function, including fluid intelligence. We correlated suicide ideation with fluid intelligence, which measures the capacity to think logically and solve problems in new situations, independent of acquired knowledge (Wu et al., 2022). It involves the ability to identify patterns, use logic to solve problems, and think abstractly. This is different from crystallized intelligence, which involves the ability to use skills, knowledge, and experience and does not rely on the capacity to think flexibly or creatively in response to new and unique challenges. The fluid intelligence increases with age till it reaches its steady state during the third decade of life prior to a delayed declination (Figure 7-bottom right) (Wu et al., 2022). The fluid intelligence correlates with a vast number of cognitive activities, and is an important predictor of both educational

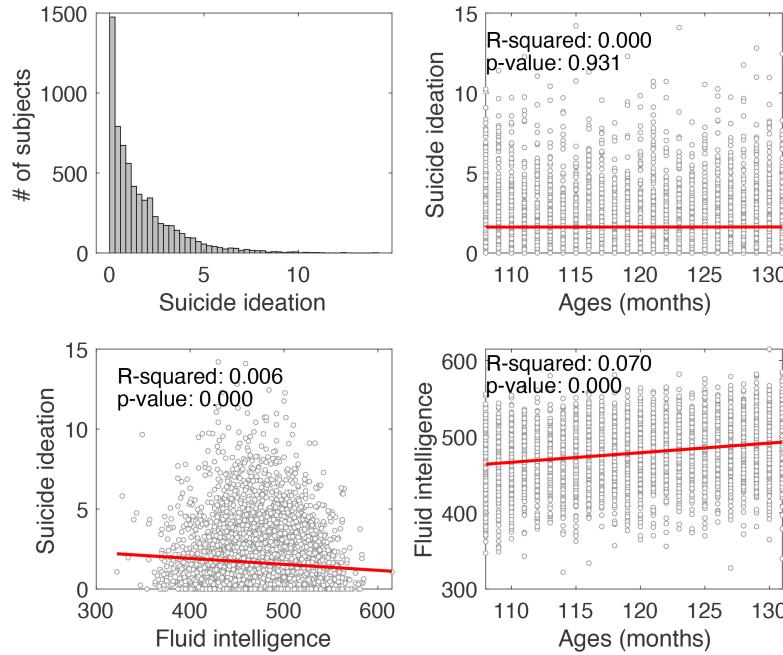


Fig. 7 Top left: the distribution of suicide ideation for 6,638 subjects. Top right: there is no age effect on suicide ideation. Bottom left: There is significant negative correlation between suicide ideation and fluid intelligence ($p\text{-value} = 1.36 \times 10^{-9}$). Bottom right: Fluid intelligence increases with ages in children ($p\text{-value} < 0.01 \times 10^{-10}$).

and professional success (Deary, 2008). We found a statistically significant negative correlation between suicidal ideation and fluid intelligence. This relationship was found to be remarkably strong suggesting that higher scores in fluid intelligence are associated with lower levels of suicidal ideation ($p\text{-value} = 1.4 \times 10^{-9}$). Higher levels of fluid intelligence might equip individuals with better problem-solving skills and coping strategies. This could potentially make them less likely to experience suicidal ideation when faced with stressors or challenges. Also higher fluid intelligence might be associated with better emotional regulation (Mayer et al., 2008). The ability to understand and manage one's emotions could contribute to lower levels of suicidal ideation. Further, higher fluid intelligence is often linked to better educational achievement, which could in turn be associated with more supportive social networks, higher self-esteem, and better access to mental health resources (Deary et al., 2007). These multiple factors might protect against suicidal ideation.

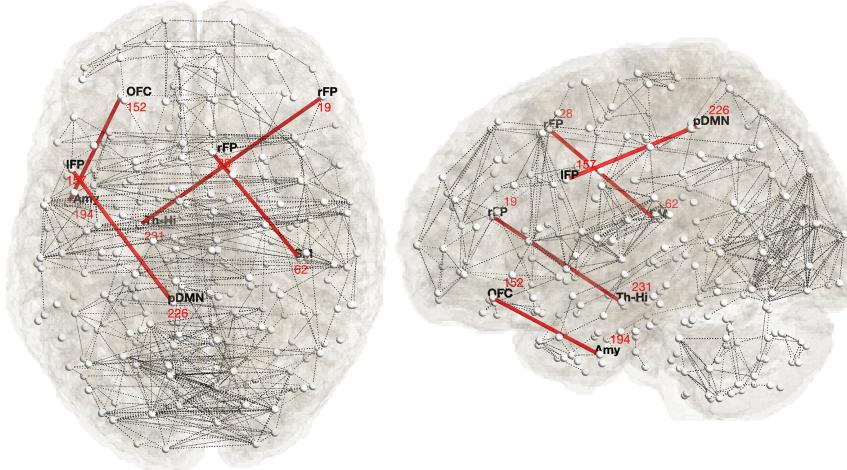


Fig. 8 Top four most significant connections that correlate with suicide ideation identified with node indices: 19-231, 28-62, 152-194, 157-226. Nodes are also identified with 14 functional subnetworks. Node 194 is the inferior temporal gyrus and belong to the Amygdala subnetwork.

2.8 Correlating rs-fMRI connectivity with suicidal ideation

[None of p-values below pass the FDR correction at 0.05 when we consider for every possible connections $268^2/2$. We need to reduce the potential number of connections since not every nodes will be connected to other 267 nodes in practice.]

We correlated correlation at each edge with suicide ideation by fitting a linear model

$$\text{connectivity} = c_1 \text{suicide} + c_2.$$

The parameters c_1 and c_2 are fitted in the traditional least squares fashion. We observed significant connection between the precuneus (part of posterior default mode network (pDMN), 226) and the left middle frontal gyrus (part of left fronto-parietal (IFP) network, 157), where the strength of this connectivity is negatively associated with suicide ideation at p -value 4.62×10^{-6} (Figure 9). The precuneus region is related to self-awareness, self-referential thinking and self-reflection, including autobiographical memory, envisioning the future, and taking the perspective of others (Cavanna and Trimble, 2006). Impairments in this region have been linked with depressive symptoms, which are often associated with suicidal ideation. The IFP, including the left middle frontal gyrus, is associated with executive functions such as decision-making, attention, working memory, and cognitive control. These functions are critical for managing thoughts and controlling impulses. Alterations in the functionality of this region have been linked to impulsive behaviors, a known risk factor for suicide (Seeley et al., 2007). If the connectivity between these regions decreases, it could be indicative of a disruption in the coordination between the

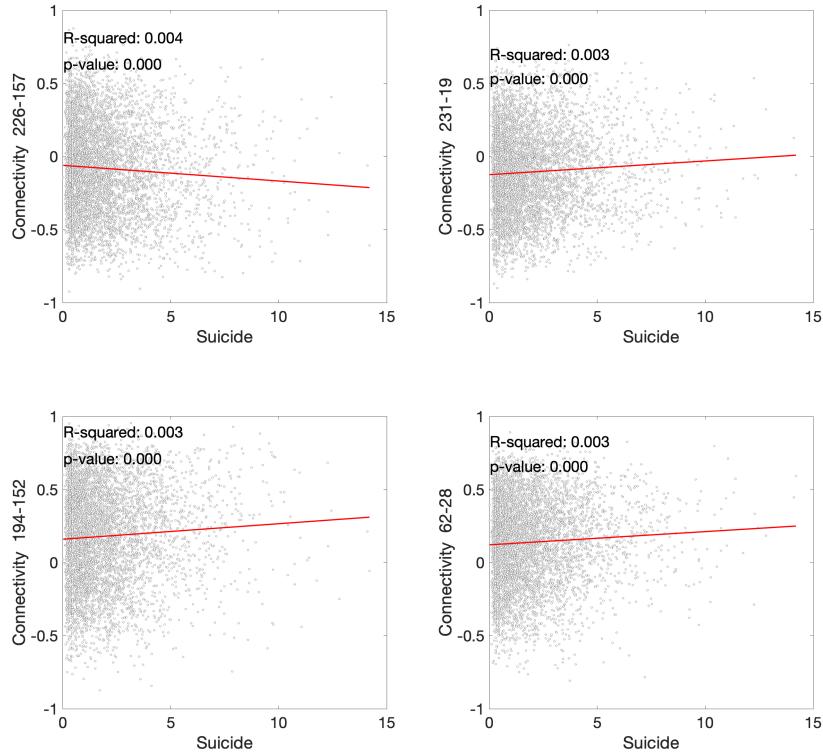


Fig. 9 The regression plots of fitting $\text{connectivity} = c_1\text{suicide} + c_2$ at each edge. The plots correspond to the four most significant connections that correlate with suicide ideation identified with node indices: 19-231, 28-62, 152-194, 157-226. It is interesting to observe that relation can be either positively or negatively correlated.

self-referential processing and executive control processes. This might lead to difficulty in managing negative thoughts, regulating emotions, and controlling impulses, which could, in turn, increase suicidal ideation.

We observed significant connection between the thalamus-hippocampus region (Th-Hi, 231) and right middle frontal gyrus (part of right fronto-parietal network (rFP), 19), where the strength of this connectivity is positively associated with suicide ideation at p -value 1.59×10^{-5} (Figure 9). The thalamus serves as a relay station for sensory information and plays a crucial role in various cognitive and sensory processes (Mitchell et al., 2014). The right middle frontal gyrus, part of the prefrontal cortex, is associated with higher-order cognitive functions, including decision-making, cognitive control, and emotional regulation (Seeley et al., 2007). Dysfunction or altered connectivity in this region has been implicated in various psychiatric conditions and psychological processes. The positive association between the strength of connectivity between TH-Hi and rFP and suicide ideation suggests

that individuals with increased connectivity in this pathway may have a heightened risk of experiencing suicidal thoughts.

We observed significant connection between regions between the inferior temporal gyrus (ITG) (part of amygdala network (Amy), 194) and orbitofrontal cortex (OFC, 152), where the strength of this connectivity is positively associated with suicide ideation (Figure 9). The ITG is involved in the processing and recognition of visual information and semantic memory. The alterations in temporal lobe structures might be associated with suicide risk, potentially due to their roles in memory and emotional processing. The OFC is involved in decision-making, reward processing, and emotion regulation. Alterations in the OFC have been implicated in psychiatric conditions associated with increased suicide risk, such as depression and borderline personality disorder (van Heeringen et al., 2011; Aizenstein et al., 2009). The OFC is thought to play a role in impulsivity and decision-making, both of which can be relevant to suicidal ideation and behavior. The alterations in how these regions process emotional and sensory information or make decisions might contributed to increased suicide ideation. Children might be interpreting or recalling emotional or negative visual stimuli more intensely involving the ITG, and this might affect their decision-making or emotional regulation processes involving the OFC.

We observed significant connection between the somatomotor (SM) (62) region and the right fronto-parietal (IFP) network (28), where the strength of this connectivity is positively associated with suicide ideation (Figure 9). The somatomotor (SM) region is associated with sensory processing and motor control. Although traditionally associated with movement and physical sensation, the somatomotor network may also play a role in emotional processing and has been linked to mood disorders (Korgaonkar et al., 2013; Kaiser et al., 2015). Increased connectivity between these regions could reflect an overactive emotional response paired with a lack of effective executive control to manage this response. This might lead to difficulty managing negative thoughts, regulating emotions, and controlling impulses, potentially increasing suicidal ideation.

2.9 Topological regression

In the previous section, where the regression model is fitted at each connectivity matrices at the edge level using the least squares. We show how to perform topological regression in this section. [This is work in progress.]

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.
- D.M. Barch, M.D. Albaugh, S. Avenevoli, L. Chang, D.B. Clark, M.D. Glantz, J.J. Hudziak, T.L. Jernigan, S.F. Tapert, and D. Yurgelun-Todd. Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: Rationale and description. *Developmental Cognitive Neuroscience*, 32: 55–66, 2018.
- 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. Caspi, R.M. Houts, D.W. Belsky, S.J. Goldman-Mellor, H.L. Harrington, S. Israel, M.H. Meier, S. Ramrakha, I. Shalev, and R. Poulton. The p factor: one general psychopathology factor in the structure of psychiatric disorders? *Clinical psychological science*, 2:119–137, 2014.
- 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, 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.
- J. Huang, M.K. Chung, and A. Qiu. Heterogeneous graph convolutional neural network via Hodge-laplacian for brain functional data. *arXiv preprint arXiv:2302.09323*, 2023.
- 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.
- Benjamin B Lahey, Brooks Applegate, Jahn K Hakes, David H Zald, Ahmad R Hariri, and Paul J Rathouz. Is there a general factor of prevalent psychopathology during adulthood? *Journal of abnormal psychology*, 121:971, 2012.
- 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.
- L. Parkes, T.M. Moore, M.E. Calkins, P.A. Cook, M. Cieslak, D.R. Roalf, D.H. Wolf, R.C. Gur, R.E. Gur, and T.D. Satterthwaite. Transdiagnostic dimensions of psychopathology explain individuals’ unique deviations from normative neurodevelopment in brain structure. *Translational psychiatry*, 11:1–13, 2021.

- T. Paus, M. Keshavan, and J.N. Giedd. Why do many psychiatric disorders emerge during adolescence? *Nature Reviews Neuroscience*, 9:947–957, 2008.
- 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.
- X. Shen, F. Tokoglu, X. Papademetris, and R.T. Constable. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *Neuroimage*, 82:403–415, 2013.
- 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.
- 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.
- B. Vossekuil. *The final report and findings of the Safe School Initiative: Implications for the prevention of school attacks in the United States*. Diane Publishing, 2002.
- 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.