NeuroImage 284 (2023) 120436

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Unified topological inference for brain networks in temporal lobe epilepsy using the Wasserstein distance

Moo K. Chung [a,*], Camille Garcia Ramos [b], Felipe Branco De Paiva [b], Jedidiah Mathis [c], Vivek Prabharakaren [d], Veena A. Nair [d], Elizabeth Meyerand [e], Bruce P. Hermann [b], Jeffrey R. Binder [c], Aaron F. Struck [b]

[a] Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, USA
[b] Department of Neurology, University of Wisconsin-Madison, USA
[c] Department of Neurology, Medical College of Wisconsin, USA
[d] Department of Radiology, University of Wisconsin-Madison, USA
[e] Departments of Medical Physics & Biomedical Engineering, University of Wisconsin-Madison, USA

ARTICLE INFO

Dataset link: https://github.com/laplcebeltrami/PH-STAT

ABSTRACT

Persistent homology offers a powerful tool for extracting hidden topological signals from brain networks. It captures the evolution of topological structures across multiple scales, known as filtrations, thereby revealing topological features that persist over these scales. These features are summarized in persistence diagrams, and their dissimilarity is quantified using the Wasserstein distance. However, the Wasserstein distance does not follow a known distribution, posing challenges for the application of existing parametric statistical models. To tackle this issue, we introduce a unified topological inference framework centered on the Wasserstein distance. Our approach has no explicit model and distributional assumptions. The inference is performed in a completely data driven fashion. We apply this method to resting-state functional magnetic resonance images (rs-fMRI) of temporal lobe epilepsy patients collected from two different sites: the University of Wisconsin-Madison and the Medical College of Wisconsin. Importantly, our topological method is robust to variations due to sex and image acquisition, obviating the need to account for these variables as nuisance covariates. We successfully localize the brain regions that contribute the most to topological differences. A MATLAB package used for all analyses in this study is available at https://github.com/laplcebeltrami/PH-STAT.

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 that measure brain connectivity (Chung et al., 2017a; Sporns, 2003; Wijk et al., 2010). However, the final statistical analysis results change depending on the choice of threshold or parameter (Chung et al., 2013; Lee et al., 2012; Zalesky et al., 2010). There is a need to develop a multiscale network analysis framework that provides consistent results and interpretation regardless of the choice of parameter (Li et al., 2020; Kuang et al., 2020). Persistent homology, an algebraic topology method in topological data analysis (TDA), 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. Persistent homological network approaches have been shown to be more robust and to outperform many existing graph theory measures and methods (Bassett and Sporns, 2017; Yoo et al., 2016; Santos et al., 2019; Songdechakraiwut and Chung, 2020). In Lee et al. (2012), persistent homology was demonstrated to outperform eight existing graph theory features, such as the clustering coefficient, small-worldness, and modularity. Similarly, in Chung et al. (2017b, 2019a), persistent

* Corresponding author.
E-mail addresses: mkchung@wisc.edu (M.K. Chung), garciaramos@wisc.edu (C.G. Ramos), fbpaiva@neurology.wisc.edu (F.B. De Paiva), jmathis@mcw.edu (J. Mathis), prabhakaran@wisc.edu (V. Prabharakaren), vnair@uwhealth.org (V.A. Nair), memeyerand@wisc.edu (E. Meyerand), hermann@neurology.wisc.edu (B.P. Hermann), jbinder@mcw.edu (J.R. Binder), struck@neurology.wisc.edu (A.F. Struck).

https://doi.org/10.1016/j.neuroimage.2023.120436
Received 15 May 2023; Received in revised form 14 September 2023; Accepted 30 October 2023
Available online 4 November 2023

homology was found to outperform various matrix-norm-based network distances.

Starting with the first TDA application in brain imaging in Chung et al. (2009), where Morse filtration was used to characterize the cortical thickness of autistic children, there have been numerous applications of TDA in brain imaging. Lee et al. (2011b,a) demonstrated the use of persistent homology in modeling functional brain networks for the first time. In recent years, TDA has garnered increasing interest in the neuroimaging community, with various applications in different modalities and disorders. For instance, TDA has also been employed in the analysis of resting-state fMRI data (Petri et al., 2014), where TDA was used to identify topological changes in brain networks under the influence of the psychedelic compound psilocybin. Lord et al. (2016) applied TDA to investigate the topological properties of resting-state networks, contrasting them against graph theory approaches. Giusti et al. (2016) proposed using simplicial homology to model higher-order brain connectivity. In Wang et al. (2017, 2018), persistent homology was shown to outperform topographic power maps, power spectral density, and local variance methods in an EEG study. In Yoo et al. (2017), center persistency was demonstrated to outperform both the network-based statistic and element-wise multiple corrections. Saggar et al. (2018) applied TDA to task fMRI to track within- and between-task transitions. Stolz et al. (2021) used TDA to characterize the task-based fMRI of schizophrenia patients.

TDA has also been applied to analyze structural brain connectivity, starting with Chung et al. (2011), where Rips filtration was employed to model white matter fiber tracts. Reimann et al. (2017) utilized Betti numbers to model synaptic neural networks. Sizemore et al. (2018) employed TDA to investigate the topological cavities that exist in structural connectivity across subjects, using diffusion spectrum imaging. Although persistent homology has been applied to a wide range of brain network studies, its primary role has been as an exploratory data analysis tool, providing anecdotal evidence for network differences. One factor limiting its more widespread adoption is the lack of transparent interpretability. The lack of transparent interpretability is one factor hampering more widespread use of topological approaches. A method that directly ties topology to brain function and structure is needed to understand the origin of topological differences. The methods proposed in this study aim to address these deficits.

The Wasserstein distance is a popular metric for comparing persistence diagrams. However, its application in statistical inference for brain network studies has been limited, largely due to computational constraints and scalability issues. Notable exceptions exist in the literature (Abdallah et al., 2023; Kumar et al., 2023; Robinson and Turner, 2017; Salch et al., 2021; Songdechakraiwut et al., 2021). Instead, researchers have turned to the vectorization of persistence diagrams as a more practical and efficient alternative for statistical inference. Vectorization involves transforming a persistence diagram into a vector representation, making it more amenable to standard machine learning and statistical techniques. Chung et al. (2009) vectorized the persistence diagram into images by counting the number of scatter points in the unit squares. Bubenik (2015) vectorized the persistence diagram into a sequence of tent functions, known as the persistence landscape. Adams et al. (2017) converted the persistence diagram into a discrete, grid-based representation referred to as the persistence image. In this paper, we demonstrate the feasibility of developing a coherent scalable statistical inference framework based on the Wasserstein distance for differentiating brain networks in a two-sample comparison setting. Our method simply bypasses the need for vectorization of persistence diagrams.

The Wasserstein distance or the Kantorovich–Rubinstein metric, was originally defined for comparing probability distributions (Vallender, 1974; Canas and Rosasco, 2012; Berwald et al., 2018). Due to its connection to optimal mass transport, which enjoys various optimal properties, the Wasserstein distance has found applications in various imaging domains. However, its use in brain imaging and network data has been limited. Mi et al. (2018) employed the Wasserstein distance in resampling brain surface meshes. Shi et al. (2016), Su et al. (2015) utilized the Wasserstein distance for classifying brain cortical surface shapes. Hartmann et al. (2018) leveraged the Wasserstein distance in building generative adversarial networks. Sabbagh et al. (2019) applied the Wasserstein distance to a manifold regression problem in the space of positive definite matrices for source localization in EEG. Xu et al. (2021) used the Wasserstein distance for predicting the progression of Alzheimer's disease in magnetoencephalography (MEG) brain networks. However, the Wasserstein distance in these applications is purely geometric in nature, and no TDA is performed.

We present a coherent scalable framework for the computation of *topological* distance on graphs through the Wasserstein distance. We directly build the Wasserstein distance using the edge weights in graphs making the method far more accessible and adaptable. We achieve $\mathcal{O}(n \log n)$ run time in most graph manipulation tasks such as matching and averaging. The method is applied in building a unified inference framework for discriminating networks topologically. Compared to existing graph theory feature based methods and other topological distances, the method provides more robust performance against false positives while increasing sensitivity when subtle topological signals are present. The method is applied in characterizing the brain networks of temporal lobe epilepsy patients obtained from the resting-state functional magnetic resonance imaging (rs-fMRI) without model specification or statistical distributional assumptions. We will show that the proposed method based on the Wasserstein distance can capture the topological patterns that are consistently observed across different subjects.

2. Methods

2.1. Preliminary: graphs as simplicial complexes

A high-dimensional object, such as brain networks, can be modeled as a weighted graph $\mathcal{X} = (V, w)$, consisting of a node set V indexed as $V = \{1, 2, \dots, p\}$ and edge weights $w = (w_{ij})$ between nodes i and j (Fig. 1). Although the method is applicable to arbitrary edge weights, in this study, the edge weights will be derived from correlations across different brain regions in rs-fMRI. Determining optimal methods to threshold the weight adjacency matrix w in order to quantify brain connectivity remains a persistent challenge in graph theory (Adamovich et al., 2022). Unfortunately, choices in the threshold parameter can significantly impact the results, hindering both reproducibility and cross-study comparisons, as well as the biological interpretation of results. The persistent homology approaches are able to overcome this limitation of graph theory through the process of filtration (Lee et al., 2011b,a). If we connect nodes based on a certain criterion and index them over increasing filtration values, they form a simplicial complex that captures the topological structure of the underlying weighted graph (Edelsbrunner and Harer, 2010; Zomorodian, 2009). During the filtration, a *topological latent space* is created, spanning from the minimum to maximum weights of the weighted adjacency matrix. Various metrics have been proposed to quantify the differences in structural and functional connectivity between different brain states or conditions within this latent space (Sizemore et al., 2018, 2019; Songdechakraiwut and Chung, 2023).

The Rips filtration is most commonly used in the literature. The *Rips complex* $\mathcal{R}_\epsilon$ is a simplicial complex, where k-simplices are formed by $(k+1)$ nodes that are pairwise within distance ϵ (Ghrist, 2008). While a graph has at most 1-simplices, the Rips complex can have up to $(p-1)$-simplices. The Rips complex induces a hierarchical nesting structure known as the Rips filtration

$$\mathcal{R}_{\epsilon_0} \subset \mathcal{R}_{\epsilon_1} \subset \mathcal{R}_{\epsilon_2} \subset \cdots$$

for $0 = \epsilon_0 < \epsilon_1 < \epsilon_2 < \cdots$, where the sequence of ϵ-values are called the filtration values. The filtration is characterized through a topological

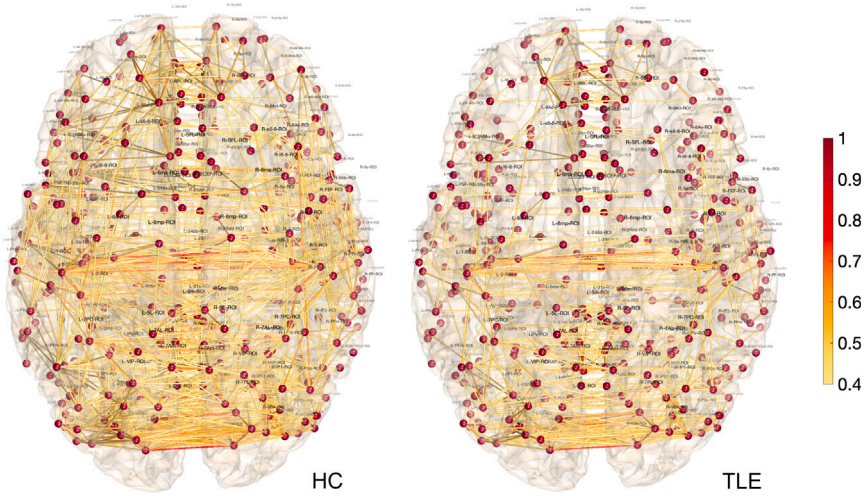

Fig. 1. The average correlation brain networks of 50 healthy controls (HC) and 101 temporal lobe epilepsy (TLE) patients. They are overlaid on top of the gray matter boundary of the MNI template. The brain network of TLE is far sparse compared to that of HC. The sparse TLE network is also consistent with the plot Betti-0 curve where TLE networks are more disconnected than HC networks. It demonstrates the global dysfunction of TLE and the breakdown of typical brain connectivity.

basis known as *k-cycles*. 0-cycles correspond to connected components, 1-cycles represent 1D closed paths or loops, and 2-cycles are 3-simplices (tetrahedra) without an interior. Any k-cycle can be represented as a linear combination of basis k-cycles. The Betti number β_k counts the number of independent k-cycles. During the Rips filtration, the ith k-cycle is born at filtration value b_i and dies at d_i. The collection of all paired filtration values

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

displayed as 1D intervals is called the *barcode*, and when displayed as scatter points in a 2D plane, it is called the *persistence diagram*. Since $b_i < d_i$, the scatter points in the persistence diagram are displayed above the line $y = x$ by placing births on the x-axis and deaths on the y-axis. Any k-cycle can be represented as a linear combination of basis k-cycles. The Betti number β_k counts the number of independent k-cycles. During the Rips filtration, the ith 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 *persistence diagram*. Since $b_i < d_i$, the scatter points in the persistence diagram are displayed above the line $y = x$ line by taking births in the x-axis and deaths in the y-axis.

2.2. Graph filtration

As the number of nodes p grows, the resulting Rips complex becomes overly dense, making it less effective for representing brain networks at higher filtration values. The computational complexity of Rips filtration grows exponentially with the number of nodes making it impractical for large datasets (Topaz et al., 2015; Solo et al., 2018). To address this, graph filtration, a special case of Rips filtration, was first introduced (Lee et al., 2011a, 2012).

Consider weighted graph $\mathcal{X} = (V, w)$ with edge weight $w = (w_{ij})$. 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 $q \le (p^2 - p)/2$. The subscript $_{(\,)}$ denotes the order statistic. In terms of sorted edge weight set $W = \{w_{(1)}, \dots, w_{(q)}\}$, we may also write the graph as $\mathcal{X} = (V, W)$.

We define binary network $\mathcal{X}_\epsilon = (V, w_\epsilon)$ 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 $w_{(1)}, w_{(2)}, \dots, w_{(q)}$ (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., 2011a, 2012). Fig. 2 illustrates a graph filtration in a 4-nodes example. 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.

Graph filtration is a special case of Rips filtration

$$\mathcal{X}_{w_{(j)}} = \mathcal{R}_{w_{(q)}-w_{(j)}}$$

restricted to 1-skeletons (Chung et al., 2013). Fig. 3 compares the two filtrations. Both utilize Euclidean distance as edge weights and have monotone β_0-curve. However, only the graph filtration has a monotone β_1-curve making it more suitable for scalable Wasserstein distance computations (Chung et al., 2019b; Songdechakraiwut and Chung, 2023).

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 persistence 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 (Fig. 2) (Chung et al., 2019a): $\beta_0(w_{(i)}) \le \beta_0(w_{(i+1)})$. 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 (Fig. 2) (Chung et al., 2019a): $\beta_1(w_{(i)}) \ge \beta_1(w_{(i+1)})$. Fig. 4 displays the Betti curves plotting β_0 and β_1 values over filtration values.

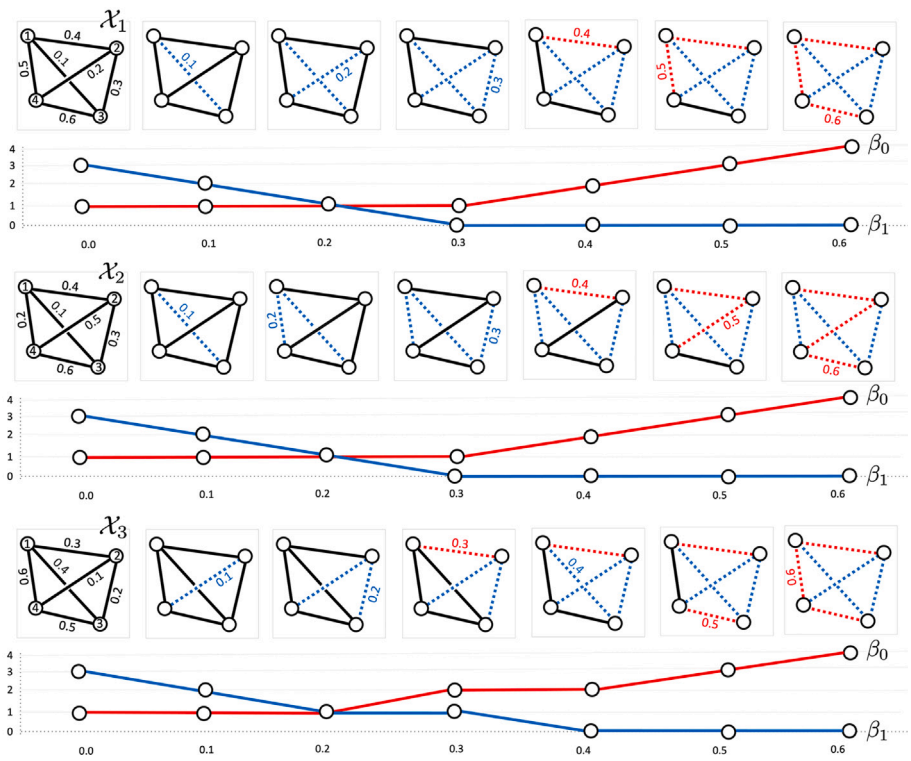

Fig. 2. Graph filtrations are obtained by sequentially thresholding graphs in increasing edge weights. The 0th Betti number β_0 (number of connected components) and the first Betti number β_1 (number of cycles) are then plotted over the filtration values. The Betti curves are monotone over graph filtrations. However, different graphs (top vs. middle) can yield identical Betti curves. As the number of nodes increases, the chance of obtaining the identical Betti curves exponentially decreases. The edges that increase β_0 (red) forms the birth set while the edge that decrease β_0 (blue) forms the death set. The birth and death sets partition the edge set.

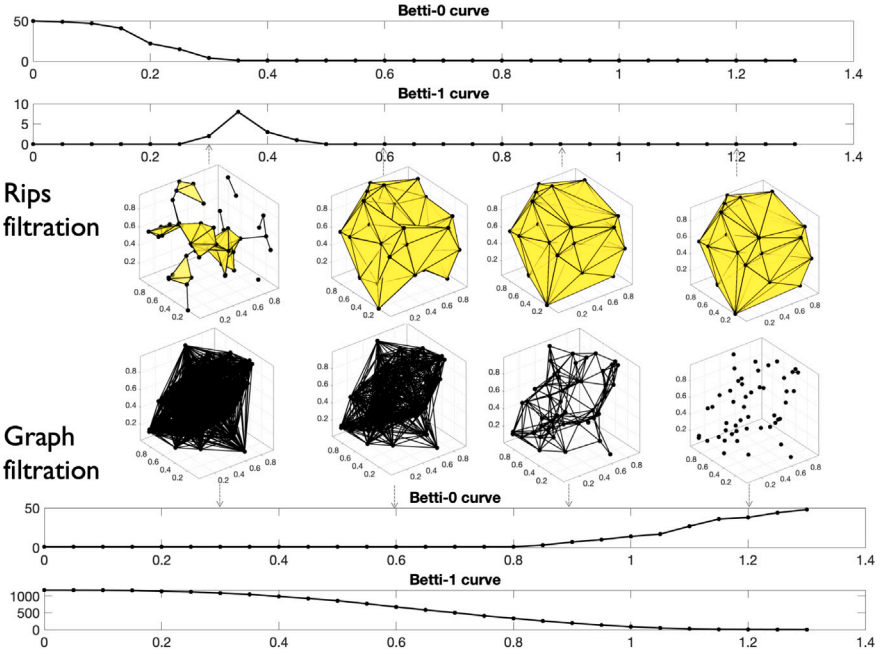

Fig. 3. The comparison between the Rips and graph filtrations performed on 50 scatter points randomly sampled in a unit cube. The Euclidean distance between points are used as edge weights. Unlike Rips filtrations, β_0 and β_1 curves for graph filtrations are always monotone making the subsequent statistical analysis far more stable.

During the graph filtration, when new components are born, they never die. Thus, 0D persistence diagrams are completely characterized by birth values b_i only. Loops are viewed as already born at $-\infty$. Thus, 1D persistence 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 (Songdechakraiwut et al., 2021):

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

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

where the birth set $W_b = \{b_{(1)}, b_{(2)}, \dots, b_{(q_0)}\}$ *is the collection of 0D sorted birth values, and the death set* $W_d = \{d_{(1)}, d_{(2)}, \dots, d_{(q_1)}\}$ *is the collection of*

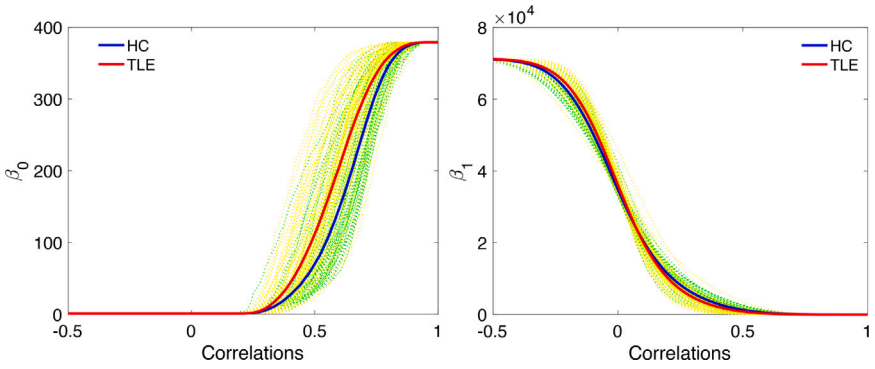

Fig. 4. Betti-0 and Betti-1 curves obtained in graph filtrations on 50 healthy controls (HC) and 101 temporal lobe epilepsy (TLE) patients. TLE has more disconnected subnetworks (β_0) compared to HC while having compatible higher order cyclic connectivity (β_1). The statistical significance of Betti curve shape difference is quantified through the proposed Wasserstein distance.

1D sorted death values, with $q_0 = p-1$ and $q_1 = \frac{(p-1)(p-2)}{2}$. Furthermore, W_b forms the 0D persistence diagram, while W_d forms the 1D persistence diagram.

Proof. During the graph filtration, when an edge is deleted, either a new component is born or a cycle dies (Chung et al., 2019a). These events are disjoint and cannot occur simultaneously. We prove the claim by contradiction. Assume, contrary to the claim, that both events happen at the same time. Then β_0 increases by 1, while β_1 decreases by 1. As an edge is deleted, the number of nodes p remains fixed, while the number of edges q is reduced to $q-1$. Thus, the Euler characteristic $\chi = p-q$ of the graph increases by 1. However, the Euler characteristic can also be expressed as an alternating sum $\chi = \beta_0 - \beta_1$ (Adler et al., 2010). As a result, the Euler characteristic would increase by 2, contradicting our previous computation. Therefore, both events cannot occur at the same time, establishing the decomposition $W = W_b \cup W_d, W_b \cap W_d = \emptyset$.

In a complete graph with p nodes, there are $q = \frac{p(p-1)}{2}$ unique edge weights. There are $q_0 = p-1$ edges that produce 0-cycles, equivalent to the number of edges in the maximum spanning tree (MST) of the graph. Since W_b and W_d partition the set, there are

$$q_1 = q - q_0 = \frac{(p-1)(p-2)}{2}$$

edges that destroy 1-cycles. The 0D persistence diagram of the graph filtration is given by $\{(b_{(1)}, \infty), \dots, (b_{(q_0)}, \infty)\}$. Ignoring ∞, W_b is the 0D persistence diagram. The 1D persistence diagram of the graph filtration is given by $\{(-\infty, d_{(1)}), \dots, (-\infty, d_{(q_1)})\}$. Ignoring $-\infty$, W_d is the 1D persistence diagram. □

Numerical implementation. The algorithm for decomposing the birth and death set is as follows. As the corollary of Theorem 1, we can show that the birth set is the maximum spanning tree (MST). 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. Fig. 2 displays graph filtration on 2 different graphs with 4 nodes, where the birth sets consists of 3 red edges and the death sets consist of 3 blue edges. Fig. 5 displays how the birth and death sets for 151 brain networks used in the study. Given edge weight matrix W as an input, Matlab function `WS_decompose.m` outputs the birth set W_b and the death set W_d. Fig. 6 displays the MST of healthy controls (HC) and temporal epilepsy (TLE) patients. 0D topology (topology of MST) is mainly characterized by the left and right hemisphere connections.

2.4. Wasserstein distance between graph filtrations

The Wasserstein Distance provides a method to quantify the similarity between brain networks. Consider persistence diagrams P_1 and P_2 given by 2D scatter points

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

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 2-*Wasserstein distance* on persistence diagrams is given by

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

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

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

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

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

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

where $d_{(i)}^j$ is the ith smallest death values in persistence diagram P_j.

Proof. 0D persistence diagram is given by $\{(b_{(1)}, \infty), \cdots, (b_{(q_0)}, \infty)\}$. Ignoring ∞, the 0D Wasserstein distance is simplified as

$$D_{W0}^2(P_1, P_2) = \min_{\psi} \sum_{i=1}^{q_0} |b_i^1 - \psi(b_i^1)|^2,$$

where the minimum is taken over every possible bijection ψ from $\{b_1^1, \dots, b_{q_0}^1\}$ to $\{b_1^2, \dots, b_{q_0}^2\}$. Note $\sum_{i=1}^{q_0} |b_i^1 - \psi(b_i^1)|^2$ is minimum only if $\sum_{i=1}^{q_0} b_i^1 \psi(b_i^1)$ is maximum. Rewrite $\sum_{i=1}^{q_0} b_i^1 \psi(b_i^1)$ in terms of the order statistics as $\sum_{i=1}^{q_0} b_{(i)}^1 \psi(b_{(i)}^1)$. Now, we prove by *induction*. When $q = 2$, there are only two possible bijections:

$$b_{(1)}^1 b_{(1)}^2 + b_{(2)}^1 b_{(2)}^2 \quad \text{and} \quad b_{(1)}^1 b_{(2)}^2 + b_{(2)}^1 b_{(1)}^2.$$

Since $b_{(1)}^1 b_{(1)}^2 + b_{(2)}^1 b_{(2)}^2$ is larger, $\psi(b_{(i)}^1) = b_{(i)}^2$ is the optimal bijection. When $q_0 = k$, assume $\psi(b_{(i)}^1) = b_{(i)}^2$ is the optimal bijection. When $q_0 = k+1$,

$$\max_{\psi} \sum_{i=1}^{k+1} b_{(i)}^1 \psi(b_{(i)}^2) \le \max_{\psi} \sum_{i=1}^{k} b_{(i)}^1 \psi(b_{(i)}^1) + \max_{\psi} b_{(k+1)}^1 \psi(b_{(k+1)}^1).$$

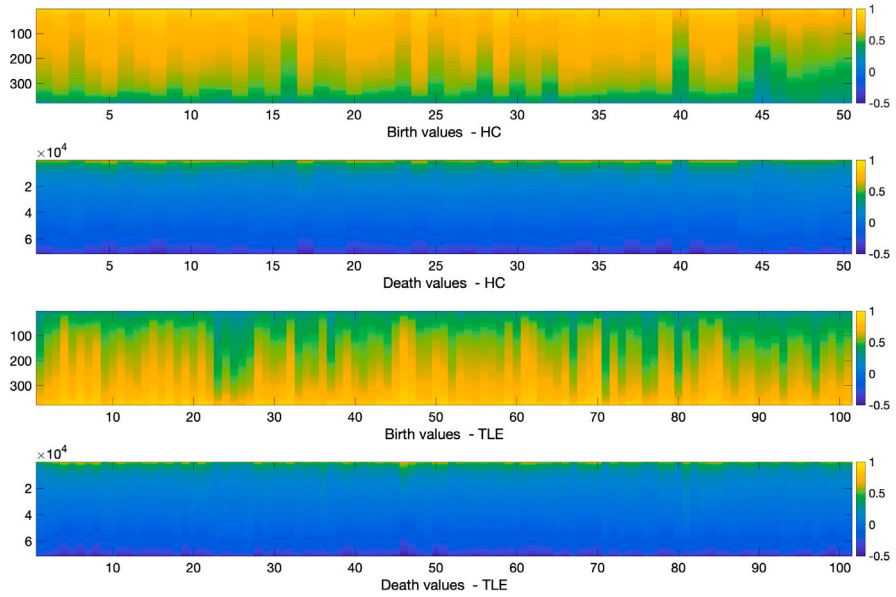

Fig. 5. The birth and death sets of 50 healthy controls (HC) and 101 temporal lobe epilepsy (TLE) patients. The Wasserstein distance between the birth sets measures 0D topology difference while the Wasserstein distance between the death sets measures 1D topology difference.

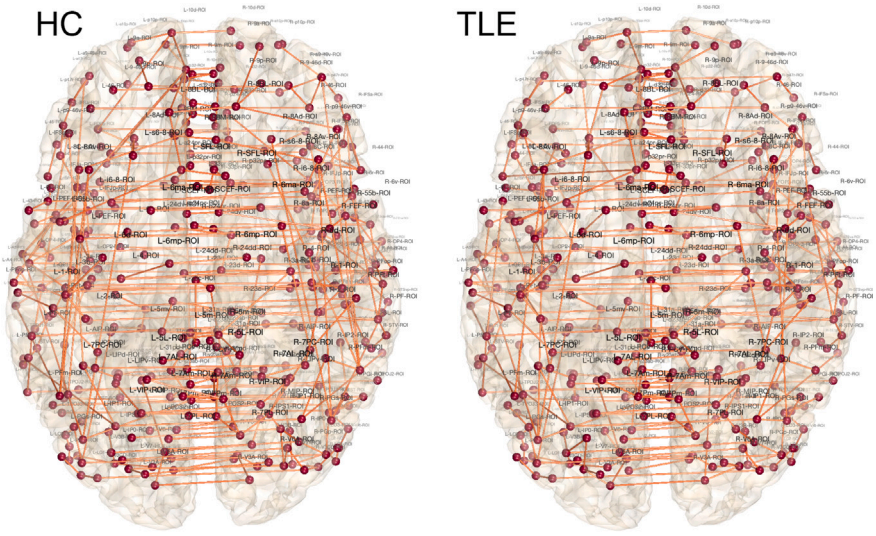

Fig. 6. Maximum spanning trees (MST) of the average correlation of HC and TLE. MST are the 0D topology while none-MST edges not shown here are 1D topology. MST forms the birth set. MST of rs-fMRI is mainly characterized by the left–right connectivity.

The first term is maximized if $\psi(b^1_{(i)}) = b^2_{(i)}$. The second term is maximized if $\psi(b^1_{(k+1)}) = b^2_{(k+1)}$. Thus, we proved the statement.

1D persistence diagram of graph filtration is given by $\{(-\infty, d_{(1)}), \ldots, (-\infty, d_{(q)})\}$. Ignoring $-\infty$, the Wasserstein distance is given by

$$D^2_{W1}(P_1, P_2) = \min_{\psi} \sum_{i=1}^{q_1} |d^1_i - \psi(d^1_i)|^2.$$

Then we follow the similar inductive argument as the 0D case. □

Using the Wasserstein distance between two graphs, we can match graphs at the edge level. In the usual graph matching problem, the node labels do not have to be matched and thus, the problem is different from simply regressing brain connectivity matrices over other brain connectivity matrices at the edge level (Becker et al., 2018; Surampudi et al., 2018). Existing geometric graph matching algorithms have been previously used in matching and averaging heterogeneous tree structures (0D topology) such as brain artery trees and neuronal trees (Guo and Srivastava, 2020; Zavlanos and Pappas, 2008; Babai and Luks, 1983). But rs-fMRI networks are dominated by 1-cycles (1D topology) and not necessarily perform well in matching 1D topology.

Suppose we have weighted graphs $\mathcal{X}_1 = (V_1, w^1)$ and $\mathcal{X}_1 = (V_2, w^2)$, and corresponding 0D persistence diagrams P^0_1 and P^0_2 and 1D persistence diagrams P^1_1 and P^1_2. We define the Wasserstein distance between graphs $\mathcal{X}_1$ and $\mathcal{X}_2$ as the Wasserstein distance between corresponding persistence diagrams P_1 and P_2:

$$D_{Wj}(\mathcal{X}_1, \mathcal{X}_2) = D_{Wj}(P^j_1, P^j_2).$$

The 0D Wasserstein distance matches birth edges while the 1D Wasserstein distance matches death edges. We need to use both distances together to match graphs. Thus, we use the squared sum of 0D and 1D Wasserstein distances

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

as the Wasserstein distance between graphs in the study. Then we can show the distance is translation and scale invariant in the following

sense:

$$D(c + \mathcal{X}_1, c + \mathcal{X}_2) = D(\mathcal{X}_1, \mathcal{X}_2),$$
$$\frac{1}{c^2} D(c\mathcal{X}_1, c\mathcal{X}_2) = D(\mathcal{X}_1, \mathcal{X}_2).$$

Unlike existing computationally demanding graph matching algorithms, the method is scalable at $\mathcal{O}(q \log q)$ run time. The majority of runtime is on sorting edge weights and obtaining the corresponding maximum spanning trees (MST).

2.4.1. Gromov-Hausdorff distance

In comparison to the Wasserstein distance, the Gromov–Hausdorff (GH) and bottleneck distances have previously been used for inference on brain networks. The GH distance for brain networks was introduced in Lee et al. (2011a, 2012). The GH distance measures the difference between networks by embedding each network into an ultrametric space that represents the hierarchical clustering structure of the network (Carlsson and Mémoli, 2010).

The Single Linkage Distance (SLD) is defined as the shortest distance between two connected components that contain nodes i and j. SLD is an ultrametric, satisfying the stronger form of the triangle inequality $s_{ij} \leq \max(s_{ik}, s_{kj})$ (Carlsson and Mémoli, 2010). Thus, a dendrogram can be represented as an ultrametric space, which is also a metric space. The GH-distance between networks is then defined through the GH-distance between corresponding dendrograms as

$$D_{GH}(\mathcal{X}_1, \mathcal{X}_2) = \max_{i,j} |s^1_{ij} - s^2_{ij}|. \tag{3}$$

The edges that yield the maximum SLD is the GH-distance between the two networks.

2.4.2. Bottleneck distance

The bottleneck distance for graph filtration reduces to matching sorted birth values or sorted death values. Given networks $\mathcal{X}_i = (V, w^i)$, the corresponding persistence diagrams P_i are obtained through the graph filtration (Lee et al., 2012; Chung et al., 2019b). The bottleneck distance between persistence diagrams P_i and P_j is given by

$$D_B(P_i, P_j) = \inf_{\psi} \sup_{x_i \in P_i} \|x_i - \psi(x_i)\|_\infty, \tag{4}$$

where and ψ is a bijection from P_i to P_j (Cohen-Steiner et al., 2007; Edelsbrunner and Harer, 2008). Since 0D persistence diagram consists of sorted birth values $b^i_{(k)}$ and $b^j_{(k)}$, we have the 0D bottleneck distance

$$D_{B0}(P_i, P_j) = \max_k \left| b^i_{(k)} - b^j_{(k)} \right|,$$

the largest gap in the order statistics difference (Das et al., 2022). Similarly, for 0D persistence diagram consists of sorted death values $d^i_{(k)}$ and $d^j_{(k)}$, we have the 1D bottleneck distance

$$D_{B1}(P_i, P_j) = \max_k \left| d^i_{(k)} - d^j_{(k)} \right|.$$

2.5. Topological inference

There are a few studies that used the Wasserstein distance (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.

2.5.1. Topological mean of graphs

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

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

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.

With Theorem 6 (Appendix), we define the *Wasserstein graph sum* of graphs $\mathcal{X}_1 = (V, w^1)$ and $\mathcal{X}_2 = (V, w^2)$ as $\mathcal{X}_1 + \mathcal{X}_2 = (V, w)$ with the birth–death decomposition $W_b \cup W_d$ satisfying

$$W_b \cup W_d = (W_{1b} + W_{2b}) \cup (W_{1d} + W_{2d}).$$

with

$$w = \mathcal{F}(W_b \cup W_d).$$

However, the sum is not uniquely defined. 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). However, this is not an issue since their topology is uniquely defined and produces identical persistence diagrams. Now, we define the *topological mean of graphs* $\mathbb{E}\mathcal{X}$ of $\mathcal{X}_1, \dots, \mathcal{X}_n$ as

$$\mathbb{E}\mathcal{X} = \frac{1}{n} \sum\nolimits_{k=1}^n \mathcal{X}_k. \tag{5}$$

The topological mean of graphs is the minimizer with respect to the Wasserstein distance, which is analogous to the sample mean as the minimizer of Euclidean distance. However, the topological mean of graphs is not unique in geometric sense. It is only unique in topological sense.

Theorem 3. *The topological mean of graphs $\mathcal{X}_1, \dots \mathcal{X}_n$ is the graph given by*

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

Proof. Since the cost function is a linear combination of quadratic functions, the global minimum exists and unique. Let $X = (V, W_b \cup W_d)$ be the birth–death decomposition with $W_b = \{b_{(1)}, \dots, b_{(q_0)}\}$ and $W_d = \{d_{(1)}, \dots, d_{(q_1)}\}$. From Theorem 2,

$$\sum_{k=1}^n D(X, \mathcal{X}_i) = \sum_{k=1}^n \left[\sum_{i=1}^{q_0} (b_{(i)} - b^k_{(i)})^2 + \sum_{i=1}^{q_1} (d_{(i)} - d^k_{(i)})^2 \right].$$

This is quadratic so the minimum is obtained by setting its partial derivatives with respect to $b_{(i)}$ and $d_{(i)}$ equal to zero:

$$b_{(i)} = \frac{1}{n} \sum_{k=1}^n b^k_{(i)}, \quad d_{(i)} = \frac{1}{n} \sum_{k=1}^n d^k_{(i)}.$$

Thus, we obtain

$$W_b = \frac{1}{n} \sum_{k=1}^n W_{kb}, \quad W_d = \frac{1}{n} \sum_{k=1}^n W_{kd}.$$

This is identical to the birth–death decomposition of $\frac{1}{n}\sum_{k=1}^n \mathcal{X}_k$ and hence proves the statement. □

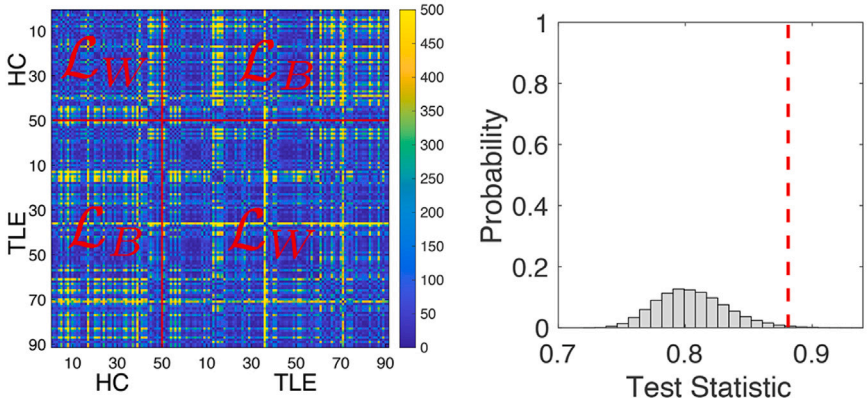

Fig. 7. Pairwise Wasserstein distance between 50 healthy controls (HC) and 101 temporal lobe epilepsy (TLE) patients. There are subtle pattern difference in the off-diagonal patterns (between group distances $\mathcal{L}_B$) compared to diagonal patterns (within group distances $\mathcal{L}_W$). The permutation test with 100 million permutations was used to determine the statistical significance using the ratio statistic. The red line is the observed ratio. The histogram is the empirical null distribution obtained from the permutation test.

The *topological variance of graphs* $\mathbb{V}\mathcal{X}$ is defined in a similar fashion:

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

which is interpreted as the variability of graphs from the Wasserstein graph mean $\mathbb{E}\mathcal{X}$. We can rewrite the topological variance of graphs as

$$\begin{aligned}\mathbb{V}\mathcal{X} &= \frac{1}{n}\sum_{k=1}^{n} D\Big(\frac{1}{n}\sum_{j=1}^{n}\mathcal{X}_j, \mathcal{X}_k\Big)\\ &= \frac{1}{n^2}\sum_{j,k=1}^{n} D(\mathcal{X}_j, \mathcal{X}_k). \qquad (6)\end{aligned}$$

The formulation (6) compute the variance using the pairwise distances without the need for computing the topological mean of graphs.

2.5.2. Distance-based topological inference

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 topological 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 (Chung et al., 2018). This usually cause a serious computational bottleneck when we have to recompute the test statistic for large samples when more than million permutations are needed. We propose a more scalable approach.

Define the within-group distance $\mathcal{L}_W$ as (Songdechakraiwut et al., 2021)

$$2\mathcal{L}_W = \sum_{\mathcal{X}_i,\mathcal{X}_j\in C_1} D(\mathcal{X}_i, \mathcal{X}_j) + \sum_{\mathcal{X}_i,\mathcal{X}_j\in C_2} D(\mathcal{X}_i, \mathcal{X}_j).$$

The within-group distance corresponds to the sum of all the pairwise distances in the block diagonal matrices in Fig. 7. The average within-group distance is then given by

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

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

$$2\mathcal{L}_B = \sum_{\mathcal{X}_i\in C_1}\sum_{\mathcal{X}_j\in C_2} D(\mathcal{X}_i, \mathcal{X}_j) + \sum_{\mathcal{X}_i\in C_2}\sum_{\mathcal{X}_j\in C_1} D(\mathcal{X}_i, \mathcal{X}_j).$$

The between-group distance corresponds to the off-diagonal block matrices in Fig. 7. The average between-group distance is then given by

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

The sum of within-group (or within-cluster) and between-group (or between-cluster) distances is the sum of all pairwise distances:

$$2\mathcal{L}_W + 2\mathcal{L}_B = \sum_{i=1}^{n}\sum_{j=1}^{n} D(\mathcal{X}_i, \mathcal{X}_j) = 2c$$

for some constant c that is invariant over permutations of group labels. When we permute the group labels, the total sum of all the pairwise distances remains unchanged. 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, we measure the disparity between groups as the ratio (Songdechakraiwut and Chung, 2023),

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

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. The ratio statistic is used in topological inference in the two-sample test but can be easily extended to k-sample tests. 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. The p-value is then computed as the probability

$$p\text{-value} = P\big(\phi_{\mathcal{L}} > \phi_{\mathcal{L}}^{\text{observed}}\big) = P\Big(\mathcal{L}_W < \frac{c}{1+\phi_{\mathcal{L}}^{\text{observed}}}\Big)$$

with the observed statistic value $\phi_{\mathcal{L}}^{\text{observed}}$.

Since the ratio is always positive, its probability distribution cannot follow Gaussian and unknown, the permutation test can be used to determine its empirical distributions. Fig. 7-right displays the empirical distribution of $\phi_{\mathcal{L}}$. The p-value is the area of the right tail, thresholded by the observed ratio $\phi_{\mathcal{L}}^{\text{observed}}$ (dotted red line) in the empirical distribution. In the permutation test, we compute the pairwise distances only once and shuffle each entry across permutations. This is equivalent

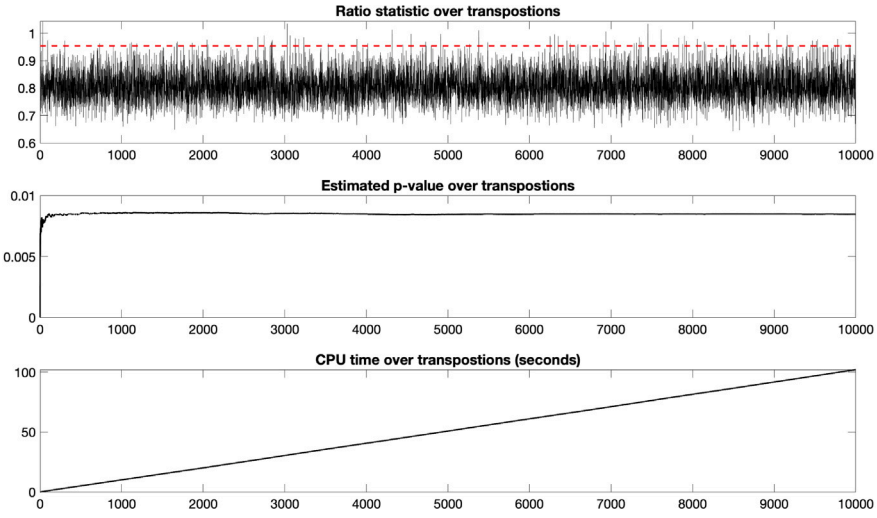

Fig. 8. The plot of ratio statistic $\phi_{\mathcal{L}}$ (top) over 100 million transpositions in testing the topological difference between HC and TLE. The plot is only shown at every 10000 transposition. The redline is the observed ratio static 0.9541. The estimated p-value (middle) converges to 0.0086 after 100 million transpositions. The CPU time (bottom) is linear and takes 102 s for 100 million transpositions.

to rearranging the rows and columns of entries corresponding to the permutations, as shown in Fig. 7. The permutation test is applicable to various two-sample comparison settings where group labels are permutable, such as the t-statistic. Simple rearrangement of rows and columns, followed by block-wise summation, should generally be faster than performing the permutation test on the conventional two-sample t-statistic, which requires recalculation for each permutation (Chung et al., 2018; Nichols and Holmes, 2002).

To speed up the permutation further, we adapted the transposition test, the online version of permutation test (Chung et al., 2019c). 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 kth and lth graphs between the groups (denoted as τ_{kl}), all the kth and ith 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 kth and ith rows and columns that are required to swapped. We only need to swap up to $\mathcal{O}(2n)$ entries while the standard permutation test that requires the computation over $\mathcal{O}(n^2)$ 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. Fig. 8 displays the convergence plot of the transposition test. Our procedure does not rely on distributional assumptions about the test statistic, making it robust to varying levels of variance between groups. The transposition test, akin to the standard permutation test, approximates the sampling distribution of the test statistic under the null hypothesis of equal distributions (Bullmore et al., 1999; Chung et al., 2018; Hayasaka et al., 2004; Nichols and Holmes, 2002). The method then quantifies how the observed data deviate from this null distribution. Therefore, the method is expected to be robust even when the groups have unequal variances.

3. Validation

We validate the proposed topological distances in simulations with the ground truth in a clustering setting. The Wasserstein distance was previously used for clustering for *geometric objects* without topology in $\mathcal{O}(q^3)$ (Mi et al., 2018; Yang et al., 2020). The proposed topological method builds the Wasserstein distances on persistence diagrams in $\mathcal{O}(q \log q)$ making our method scalable.

Consider a collection of graphs $\mathcal{X}_1, \dots, \mathcal{X}_n$ that will be clustered into k clusters $C = (C_1, \dots, C_k)$. Let $\mu_j = \mathbb{E}C_j$ be the topological mean of C_j computing using the Wasserstein distance. Let $\mu = (\mu_1, \dots, \mu_k)$ be the cluster mean vector. The within-cluster Wasserstein distance is given by

$$l_W(C;\mu) = \sum_{j=1}^{k} \sum_{X \in C_j} D(X, \mu_j) = \sum_{j=1}^{k} |C_j| \mathbb{V}C_j$$

with the topological variance $\mathbb{V}C_j$ of cluster C_j. The within-cluster Wasserstein distance generalizes the within-group distance defined on two groups to k number of groups (or clusters). When $k = 2$, we have $l_W(C;\mu) = 2\mathcal{L}_W$.

The topological clustering through the Wasserstein distance is then performed by minimizing $l_W(C)$ over every possible C. The Wasserstein graph clustering algorithm can be implemented as the two-step optimization often used in variational inferences (Bishop, 2006). The algorithm follows the proof below.

Theorem 4. *Topological clustering with the Wasserstein distance converges locally.*

Proof. (1) Expectation step: Assume C is estimated from the previous iteration. In the current iteration, the cluster mean μ corresponding to C is updated as $\mu_j \leftarrow \mathbb{E}C_j$ for each j. From Theorem 3, the cluster mean gives the lowest bound on distance $l_W(C;\nu)$ for any $\nu = (\nu_1, \dots, \nu_k)$:

$$l_W(C;\mu) = \sum_{j=1}^{k} \sum_{X \in C_j} D(X, \mu_j) \leq \sum_{j=1}^{k} \sum_{X \in C_j} D(X, \nu_j) = l_W(C;\nu). \tag{7}$$

(2) We check if the cluster mean μ is changed from the previous iteration. If not, the algorithm simply stops. Thus we can force $l_W(C;\nu)$ to be strictly decreasing over each iteration. (3) Minimization step: The clusters are updated from C to $C' = (C'_{J_1}, \dots, C'_{J_k})$ by reassigning each graph $\mathcal{X}_i$ to the closest cluster C_{J_i} satisfying $J_i = \arg\min_j D(\mathcal{X}_i, \mu_j)$. Subsequently, we have

$$l_W(C';\mu) = \sum_{J_i=1}^{k} \sum_{X \in C'_{J_i}} D(X, \mu_{J_i}) \leq \sum_{j=1}^{k} \sum_{X \in C_j} D(X, \mu_j) = l_W(C;\mu). \tag{8}$$

From (7) and (8), $l_W(C;\mu)$ strictly decreases over iterations. Any bounded strictly decreasing sequence converges. □

Just like k-means clustering that converges only to local minimum, there is no guarantee the Wasserstein distance based clustering converges to the global minimum (Huang et al., 2020). This is remedied by repeating the algorithm multiple times with different random seeds and identifying the cluster that gives the minimum over all possible seeds.

3.1. Topological clustering as a linear assignment problem

Let y_i be the true cluster label for the ith data. Let $\hat{y}_i$ be the estimate of y_i we determined from topological clustering. Let $y = (y_1, \dots, y_n)$ and $\hat{y} = (\hat{y}_1, \dots, \hat{y}_n)$. There is no direct association between true clustering labels and predicted cluster labels and they are independent. Given k clusters $C_1, \dots, C_k$, its permutation $\pi(C_1), \dots, \pi(C_k)$ is also a valid cluster for $\pi \in \mathbb{S}_k$, the permutation group of order k. There are $k!$ possible permutations in $\mathbb{S}_k$ (Chung et al., 2019c). The clustering accuracy $A(y, \hat{y})$ is then defined as

$$A(\hat{y}, y) = \frac{1}{n} \max_{\pi \in \mathbb{S}_k} \sum_{i=1}^{n} \mathbf{1}(\pi(\hat{y}) = y).$$

This is a modification to an assignment problem and can be solved using the Hungarian algorithm in $\mathcal{O}(k^3)$ run time (Edmonds and Karp, 1972). Let $F(\hat{y}, y) = (f_{ij})$ be the confusion matrix of size $k \times k$ tabulating the correct number of clustering in each cluster. The diagonal entries show the correct number of clustering while the off-diagonal entries show the incorrect number of clusters. To compute the clustering accuracy, we need to sum the diagonal entries. Under the permutation of cluster labels, we can get different confusion matrices. For large k, it is prohibitive expensive to search for all permutations. Thus we need to maximize the sum of diagonals of the confusion matrix under permutation:

$$\frac{1}{n} \max_{Q \in \mathbb{S}_k} \mathrm{tr}(QC) = \frac{1}{n} \max_{Q \in \mathbb{S}_k} \sum_{i,j} q_{ij} f_{ij}, \tag{9}$$

where $Q = (q_{ij})$ is the permutation matrix consisting of entries 0 and 1 such that there is exactly single 1 in each row and each column. This is a linear sum assignment problem (LSAP), a special case of linear assignment problem (Duff and Koster, 2001; Lee et al., 2018b).

In random assignment, each subject has an equal chance $\frac{1}{k}$ of being placed in any of the k clusters. This is true for both y_i and $\hat{y}_i$. Therefore, the expected clustering accuracy for each subject is

$$\mathbb{E}\mathbf{1}(\pi(\hat{y}_i) = y_i) = \frac{1}{k}.$$

Then the expected clustering accuracy in the random assignment is

$$\mathbb{E}[A(\hat{y}, y)] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\mathbf{1}(\pi(\hat{y}_i) = y_i) = \frac{1}{k}. \tag{10}$$

3.2. Relation to topological inference

The topological clustering used in validation is directly related to topological inference. A larger ratio statistic (or equivalently, a smaller $\mathcal{L}_W$) implies a smaller p-value. Thus, p would be a decreasing function of $\phi_{\mathcal{L}}$:

$$\frac{dp}{d\phi_{\mathcal{L}}} \leq 0,$$

or equivalently $\frac{d\phi_{\mathcal{L}}}{dp} \leq 0$. On the other hand, topological clustering is performed by minimizing $\mathcal{L}_W$ over cluster (group) labels, which is equivalent to maximizing the ratio statistic $\phi_{\mathcal{L}}$. Thus, an increase in the ratio statistic corresponds to an increase in clustering accuracy A:

$$\frac{dA}{d\phi_{\mathcal{L}}} \geq 0.$$

Subsequently,

$$\frac{dA}{dp} = \frac{dA}{d\phi_{\mathcal{L}}} \cdot \frac{d\phi_{\mathcal{L}}}{dp} \leq 0,$$

and we conclude that a decrease in p-value directly corresponds to an increase in clustering accuracy. Thus, there exists a monotonically decreasing function f satisfying $p\text{-value} = f(A)$.

3.3. Simulation with the ground truth

The proposed method is validated in a random network simulation with the ground truth against k-means and hierarchical clustering (Lee et al., 2011a). We generated 4 circular patterns of identical topology (Fig. 9-top) and different topology (Fig. 9-bottom). Along the circles, we uniformly sampled 400 (200 nodes per circle or arc) nodes and added Gaussian noise $N(0, \sigma^2)$ on the coordinates. We generated 5 random networks per group. The Euclidean distance between randomly generated points are used in the clustering task. Fig. 9 shows the superposition of nodes of all 5 networks.

We compared the proposed Wasserstein distance against two established topological distances: bottleneck and Gromov–Hausdorff (GH). The bottleneck distance is perhaps the most often used distance in persistent homology (Cohen-Steiner et al., 2007; Edelsbrunner and Harer, 2008). The Gromov–Hausdorff (GH) distance is possibly the most popular distance that is originally used to measure distance between metric spaces (Tuzhilin, 2016). It was later adapted to measure distances in persistent homology, dendrograms (Carlsson and Memoli, 2008; Carlsson and Mémoli, 2010; Chazal et al., 2009) and brain networks (Lee et al., 2011a, 2012). The distances are used in two different clustering tasks with the known ground truths. In the first task, we determined if the distance can incorrectly discriminate topologically equivalent patterns (Fig. 9-top). In the second task, we determined if the distance can correctly discriminate topologically different patterns (Fig. 9-bottom). We compared distances in three different noise settings ($\sigma = 0.1, 0.2, 0.3$). The average result of 10 independent clustering task is then reported in Table 1.

3.3.1. False positives

We conducted tests for false positives (Fig. 9-top), where all groups were generated from the Group 1 pattern through rotations. Since all these groups are topologically invariant, no topological differences should be detected; any signals identified would be false positives. In all noise settings, both the bottleneck distances and our proposed Wasserstein distance reported the lowest clustering accuracy, making them the best-performing methods (Table 1). In contrast, k-means and GH-distance incorrectly clustered the groups too well. The k-means clustering, which uses Euclidean distance, fails to recognize topological equivalence. GH distance, based on the single linkage matrix (Carlsson and Mémoli, 2010; Lee et al., 2011a, 2012), tends to cluster indiscriminately, thereby overfitting and inflating the rate of false positives.

In an ideal scenario where cluster assignments are random, the best performance would correspond to a 25% clustering accuracy. Therefore, the average false positive (FP) error rate is calculated as the mean of clustering accuracies minus 0.25. This serves as an overall performance metric for the method across different noise settings.

3.3.2. False negatives

We also tested for false negatives when there is a topological difference (Fig. 9-bottom), where all the groups have different numbers of connected components or cycles. All the groups are topologically different, and thus we should detect these differences. However, k-means clustering and bottleneck distances do not necessarily perform as well as the GH- and Wasserstein distances. GH-distance is related to hierarchical clustering, which always cluster regardless if there are clusters or not. Therefore, GH-distance is always expected to perform well if there are topological differences. Conversely, GH-distance may do not perform well when there is no topological difference. Bottleneck distances are only aware of 0D or 1D topology but not both at the same time, and their performance begins to suffer as the level of noise

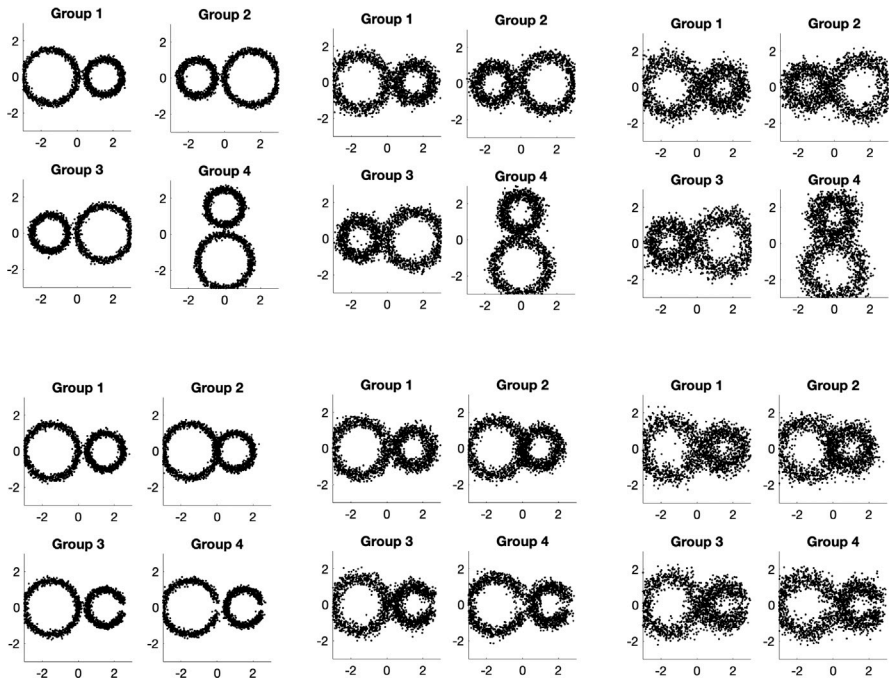

Fig. 9. Simulation study testing topological equivalence (top) and difference (bottom) in three noise settings: σ = 0.1 (left), 0.2 (middle), 0.3 (right). We should not cluster topologically equivalent patterns while we should cluster topologically different patterns well.

Table 1
Clustering accuracy in false positive (top five rows) and false negative (bottom five rows) settings. We used Euclidean distance (k-means), Bottleneck distance for 0D topology (Bottleneck-0) and 1D topology (Bottleneck-1), Gromov–Hausdorff (GH), and Wasserstein distances. The averages of 10 independent simulation results in three different noise settings are reported. Lower clustering accuracy is better for the false positive setting, while higher clustering accuracy is preferred for the false negative setting. The best-performing methods are marked in bold. Overall, the Wasserstein distance showed the best performance.

Noise σ	k-means	Bottleneck-0	Bottleneck-1	GH	Wasserstein
Small 0.1	0.75 ± 0.02	0.42 ± 0.04	0.46 ± 0.05	0.81 ± 0.08	**0.45 ± 0.04**
Medium 0.2	0.74 ± 0.01	0.44 ± 0.05	0.47 ± 0.06	0.75 ± 0.07	**0.44 ± 0.02**
Large 0.3	0.74 ± 0.01	0.43 ± 0.09	0.42 ± 0.05	0.78 ± 0.07	**0.44 ± 0.01**
FP error rate	0.49 ± 0.01	0.18 ± 0.06	0.20 ± 0.05	0.53 ± 0.07	**0.19 ± 0.19**
Small 0.1	0.78 ± 0.02	0.85 ± 0.11	0.88 ± 0.12	1.00 ± 0.00	**0.97 ± 0.02**
Medium 0.2	0.71 ± 0.02	0.63 ± 0.10	0.63 ± 0.09	0.97 ± 0.04	**0.89 ± 0.03**
Large 0.3	0.65 ± 0.02	0.56 ± 0.09	0.59 ± 0.09	0.88 ± 0.08	**0.80 ± 0.05**
FN error rate	0.29 ± 0.02	0.32 ± 0.10	0.30 ± 0.10	0.05 ± 0.04	**0.11 ± 0.03**
Total error	0.78 ± 0.02	0.50 ± 0.08	0.50 ± 0.08	0.58 ± 0.06	**0.30 ± 0.11**

increases. For false negative tests, higher clustering accuracy is better. Thus, the average false negative (FN) error rate is computed as one minus the average of clustering accuracies, which measures the overall performance of the method across different noise settings. We then computed the total error, measuring the sum of FP and FN error rates. The total error rate for the Wasserstein distance is 20% smaller than that for the bottleneck distance and 28% smaller than that for the GH distance. In summary, the proposed Wasserstein distance performed the best across most test settings. Compared to other distances, the Wasserstein distance is less likely to report false positives or false negatives.

A simulation study on the performance of topological inference using the Wasserstein distance, compared against other topological distances (GH, bottleneck), and graph theory features (Q-modularity and betweenness), is presented in Anand and Chung (2023). Our results align with these findings, demonstrating that approaches based on the Wasserstein distance consistently outperform those based on graph theory features and other topological distances. A similar conclusion is reached in Songdechakraiwut and Chung (2023), where the Wasserstein distance is compared against L_1, L_2, L_∞ matrix norms, GH-distance, and bottleneck distance, as well as various graph matching algorithms. While certain distances may outperform others in specific simulations, the Wasserstein distance consistently outperforms other metrics across a range of simulations, on average. One key factor contributing to this performance is the multi-scale nature of the Wasserstein distance in measuring topological discrepancies across filtrations. This makes it particularly robust against large perturbations and noise, outperforming existing uni-scale approaches. Furthermore, the Wasserstein distance is expected to have higher discriminative power compared to the bottleneck distance or other topological metrics. This is due to the fact that the bottleneck and GH-distances are all upper bounds for the Wasserstein distance (Dłotko and Hellmer, 2023; Chung et al., 2019b).

4. Application

4.1. Dataset

The method is applied the functional brain networks of 151 subjects in the Epilepsy Connectome Project (ECP) database (Hwang et al., 2020). We used 50 healthy controls (mean age 31.78 ± 10.32 years) and 101 chronic temporal lobe epilepsy (TLE) patients (mean age 40.23 ± 11.85). The resting-state fMRI were collected on 3T General Electric 750 scanners at two institutes (University of Wisconsin-Madison and Medical College of Wisconsin). T1-weighted MRI were acquired using MPRAGE (magnetization prepared gradient echo sequence, TR/TE = 604 ms/2.516 ms, TI = 1060.0 ms, flip angle =

8°, FOV = 25.6 cm, 0.8 mm isotropic) (Hwang et al., 2020). Resting-state functional MRI (rs-fMRI) were collected using SMS (simultaneous multi-slice) imaging (Moeller et al., 2010) (8 bands, 72 slices, TR/ TE = 802 ms/33.5 ms, flip angle = 50°, matrix = 104 . 104, FOV = 20.8 cm, voxel size 2.0 mm isotropic) and a Nova 32-channel receive head coil. The participants were asked to fixate on a white cross at the center of a black screen during the scans (Patriat et al., 2013). 40 healthy controls (HC) were scanned at the University of Wisconsin-Madison (UW) while 10 healthy controls were scanned at the Medical College of Wisconsin (MCW). 39 TLE patients were scanned at the University of Wisconsin-Madison while 62 TLE patients were scanned at the Medical College of Wisconsin.

MRIs were processed following Human Connectome Project (HCP) minimal processing pipelines (Glasser et al., 2013). Additional preprocessing was performed on the rs-fMRI using AFNI (Cox, 1996) and included motion regression using 12 motion parameters and band-pass filtering (0.01–0.1 Hz) (Hwang et al., 2020). We used 360 Glasser parcellations (Glasser et al., 2016) and additional 19 FreeSurfer subcortical regions (Fischl et al., 2002) in computing pairwise Pearson correlation between brain regions over the whole time points. This results in 379 by 379 connectivity matrix per subject. 180 brain regions reported in Glasser et al. (2013) are indexed between 1 to 180 for the left hemisphere and 181 to 360 for the right hemisphere. The 19 subcortical structures from FreeSurfer are indexed between 361 to 379. Fig. 1 displays the average connectivity in HC and TLE. TLE shows far sparse more fractured network topology compared to HC.

4.2. Topological difference in temporal lobe epilepsy

The Wasserstein distance provides a method to quantify the similarity between networks. The inferential ratio statistic $\phi_{\mathcal{L}}$ uses the Wasserstein distance to quantify within-group versus between-group likelihood. This methodology employed a meaningful statistical framework and proved reliably characterize differences between the patterns of rs-fMRI connectivity in TLE versus HC. As the topological latent space is only dependent on the relative strength of connections of nodes or loops via rank-order, it is potentially more robust to scanner and institutional differences. There is no need to account for site as a confounding factor (Jovicich et al., 2006; Gunter et al., 2009).

Since the images were collected in two different sites, we tested if there is any site effect. Using the proposed ratio statistic on the Wasserstein distance, we compared 40 healthy controls from UW and 10 healthy controls from MCW. We obtained the p-value of 0.62 with one million transpositions indicating there is no site effect observed in HC. We also compared 39 TLE from UW and 62 TLE from MCW. We obtained the p-value of 0.58 with one million transpositions indicating there is no site effect observed in TLE as well. Thus, we did not account for site effects in comparing healthy controls and TLE. The topological method does not penalize the geometric differences such as correlation differences but only topological differences and should be very robust for site differences.

We also tested for a sex effect. There are 25 males and 25 females in HC. We obtained a p-value of 0.70 with one million transpositions, indicating that *no* sex effect was observed in HC. There are 39 males and 62 females in TLE. We obtained a p-value of 1.00 with one million transpositions, indicating that *no* sex effect was observed in TLE. Thus, we did not account for the sex effect when comparing HC and TLE. Our topological method appears to be very robust to sex differences. Since older TLE patients have been suffering from TLE for a longer duration compared to younger TLE patients, it is unclear whether the age effect is due to the duration of exposure to TLE or to actual age differences. Thus, we did not test for an age effect.

The proposed method is subsequently applied in comparing 50 healthy and 101 TLE patients. The pairwise distance within TLE is 114.62 ± 147.67 while the pairwise distance within HC is 110.65 ± 124.78. The average pairwise distance within a group is not too different between TLE and HC. What separates TLE and HC is the between-group distance which measures the sum of all possible pairwise distance between a TLE subject and a HC subject. From the ratio statistic of the between-group over within-group distance, we obtained the p-value of 0.0086 after 100 million transpositions for 100 s computation in a desktop (Fig. 7).

This can be easily interpreted if we spread 151 subject as scatter points in topological embedding (Fig. 10-left). The figure displays the spread of each subject with respect to the group topological mean (blue square for HC and red square for TLE), where the x-axis shows the spread with respect to the 0D topology and the y-axis shows the spread with respect to the 1D topology. Given sorted birth values $b^k_{(1)}, b^k_{(2)}, \dots, b^k_{(q_0)}$ for the kth subject, the x-coordinates of the group topological mean is given by

$$\mu_b = \frac{1}{q_0}\sum_{k=1}^{q_0} b^k_{(i)}.$$

The y-coordinates of the group topological mean is obtained similarly using death values. The embedding x-coordinate of the kth subject is then

$$\frac{1}{q_0}\sum_{k=1}^{q_0}(b^k_{(i)} - \mu_b).$$

The embedding y-coordinate of the kth subject is similarly given using the death values. 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.

From the ratio statistic of the between-group over within-group distance, we obtained the p-value of 0.0086 after 100 million transpositions for 100 s computation in a desktop (Fig. 7). The sample size is significantly larger so need more transpositions for the converging result (Fig. 8). The primary topological differences between TLE and HC were found in the Betti-0 (nodes) as compared to the Betti-1 (loops) (Fig. 4). The TLE patients had more sparse connections (Fig. 1). Essentially there were weaker connections between nodes in TLE patients. This result was most evident in the regions associated with the epileptic region. This result is anticipated from more traditional global graph theory approaches to rs-fMRI (Struck et al., 2021; Mazrooyisebdani et al., 2020; Liao et al., 2010; Lopes et al., 2017).

We compared our findings against our previous graph theory-based analysis in Garcia-Ramos et al. (2022), where Q-modularity and global efficiency were used to quantify the TLE network (Fig. 10). Q-modularity determines the strength of the division of a network into modules (Newman and Girvan, 2004). Networks with high modularity have dense connections within modules but sparse connections across modules. The TLE networks have smaller Q-modularity, indicating that modules are somewhat fractured, with fewer connections across different modules, which is also evident in Fig. 1. Global efficiency measures the efficiency of a network in terms of the average inverse shortest path length (Latora and Marchiori, 2001). A higher global efficiency indicates a more interconnected, tightly knit network, where information can flow rapidly and directly between nodes. The TLE networks have higher global efficiency but lower Q-modularity compared to the HC networks. A network with high modularity might have lower global efficiency if the separation into distinct modules leads to longer paths between nodes in different modules. This inverse relationship between modularity and global efficiency in TLE networks aligns with findings reported in our previous study (Garcia-Ramos et al., 2022). We conclude that there is a strong topological difference between HC and TLE that is also consistent with findings from graph theory features.

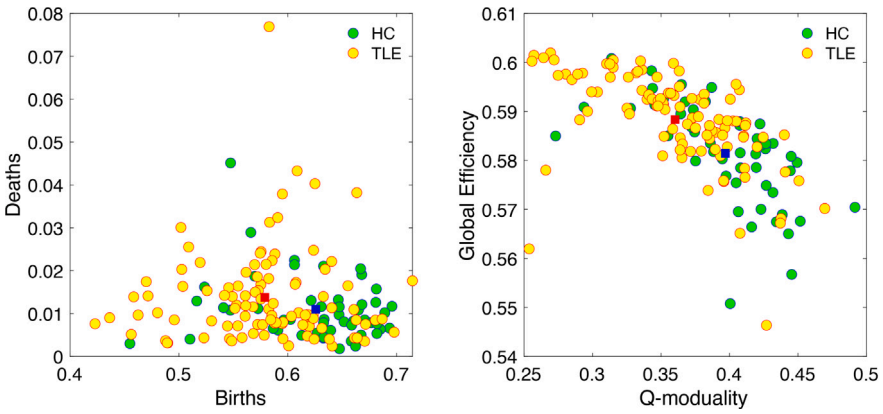

Fig. 10. Left: Topological embedding of 151 subjects. Green circles are HC and yellow circles are TLE. The blue square is the topological center of HC while the red square is the topological center of TLE. The horizontal axis represents 0D topology (connected components) through birth values while the vertical axis represents 1D topology (circles) through death values. The embedding shows that the topological separation is mainly through 0D topology. Right: Embedding through graph theory features global efficiency at vertical and Q-modularity at horizontal axes.

4.2.1. Localizing topological signals

In traditional Topological Data Analysis (TDA), it is often challenging to identify the specific brain regions responsible for observed topological differences. In contrast, our method enables the localization of such differences. Utilizing a *node attack* strategy (Bullmore and Sporns, 2012; Lee et al., 2018a), we assessed the impact of excluding each node on the ratio statistic $\phi_{\mathcal{L}}$ (Fig. 11). Specifically, we computed the difference in $\phi_{\mathcal{L}}$ with and without each node, denoted as $\Delta\phi_{\mathcal{L}}$. A larger $\Delta\phi_{\mathcal{L}}$ indicates a more discriminative subnetwork without the node. The 20 most influential brain regions, as reflected by their effect on $\Delta\phi_{\mathcal{L}}$, are illustrated in Fig. 11 and highlighted in teal in Fig. 12.

Among 20 regions, ten regions that increase the ratio statistic most are listed here: the left fundus of the superior temporal visual area (L-Area FST), brain stem, right frontal operculum area 1 (R-Frontal OPercular Area 1), right subgenual anterior cingulate cortex s32 (Right-Area s32), left temporal gyrus dorsal (L-Area TG dorsal), the middle of the left primary auditory cortex (L-Area TE1 Middle), the posterior of the right auditory cortex TE2 (R-Area TE2 posterior), right superior temporal area (R-Area FST), left subgenual anterior cingulate cortex s32 (L-Area s32), right temporo-parieto-occipital junction (R-Area TemporoParietoOccipital Junction 3). These regions are 10 most influential brain regions that are responsible for the topological difference against HC. These regions are all associated with the extended network of temporal lobe epilepsy. These regions are within the bilateral temporal regions or in close proximity, both structurally and functionally (right frontal operculum area 1, right subgenual anterior cingulate cortex or right temporal-parieto-occipital junction). The one potentially surprising result was the brainstem. The brainstem is integral to the process of losing awareness during a temporal lobe seizure and is thereby implicated in the symptomogenic zone of TLE (Mueller et al., 2018). The work of Blumenfeld, Englot and others highlights the importance of examining the connectivity of the brainstem in TLE, a result that is only starting to be explored by the larger epilepsy community (Englot et al., 2008, 2009, 2018). Finding the importance of the brainstem in fMRI connectivity difference between TLE and controls is an exciting result from this topological approach to brain networks.

5. Discussion

5.1. Unaddressed challenges in Wasserstein distance

In this study, we proposed the unified topological inference framework for discriminating the topological difference between healthy controls (HC) and temporal lobe epilepsy (TLE) patients. The method is based on computing the Wasserstein distance, the probabilistic version of optimal transport, which can measure the topological discrepancy in persistence diagrams. We developed a coherent statistical framework based on persistent homology and presented how such method is applied to the resting state fMRI data in localizing the brain regions affecting topological difference in TLE. An alternative approach for localizing the brain regions in persistent homology is to use ∞-Wasserstein distance which is the bottleneck distance given by

$$\mathcal{L}_{\infty 0}(P_1, P_2) = \max_i |b^1_{(i)} - b^2_{(i)}|$$

for 0D topology and

$$\mathcal{L}_{\infty 1}(P_1, P_2) = \max_i |d^1_{(i)} - d^2_{(i)}|$$

for 1D topology (Das et al., 2022). Due to the birth–death decomposition, the ith largest birth edges and death edges that optimize the ∞-Wasserstein distance can be easily identifiable. This is left as a future study.

The Wasserstein distance can also be used as a metric for unsupervised machine-learning to characterize latent phenotypes. In simulated data, it performed better than k-means clustering and hierarchical clustering in not detecting false positives. Although we did not explore the problem of determining optimal number of clusters, the Wasserstein distance can handle such a problem through the *elbow method* (Allen et al., 2014; Rashid et al., 2014; Ting et al., 2018; Huang et al., 2020). For each cluster number k, we compute the ratio ψ_l of the within-cluster l_W to between-cluster distance l_B, i.e.,

$$\psi_l = \frac{l_W}{l_B}.$$

The within-cluster distance generalizes the within-group distance $\mathcal{L}_W$ between two groups to k groups while the between-cluster distance generalizes the between-group distance $\mathcal{L}_B$ between two groups to k groups. Thus, when $k = 2$, we have the inverse relation with the ratio static we used in the two group discrimination task

$$\psi_l = \frac{1}{\phi_{\mathcal{L}}}.$$

The ratio shows the goodness-of-fit of the cluster model. Fig. 13 plots the ratio over different number of k for a controlled experiment. The optimal number of clusters were determined by the elbow method, gives the largest slope change in the ratio in the plot. $k = 3$ gives the largest slope in the both methods and we determine $k = 3$ is the optimal number of clusters. The performance of the elbow method is well understood in traditional k-means clustering, but its performance using the Wasserstein distance has not yet been quantified. Other methods such as the gap-statistic, silhouette or graphical methods are possible. This approach could be fruitful as TLE is a heterogeneous disease with varying etiologies, differing responses to anti-seizure medications, differing propensity to secondary generalized tonic clonic

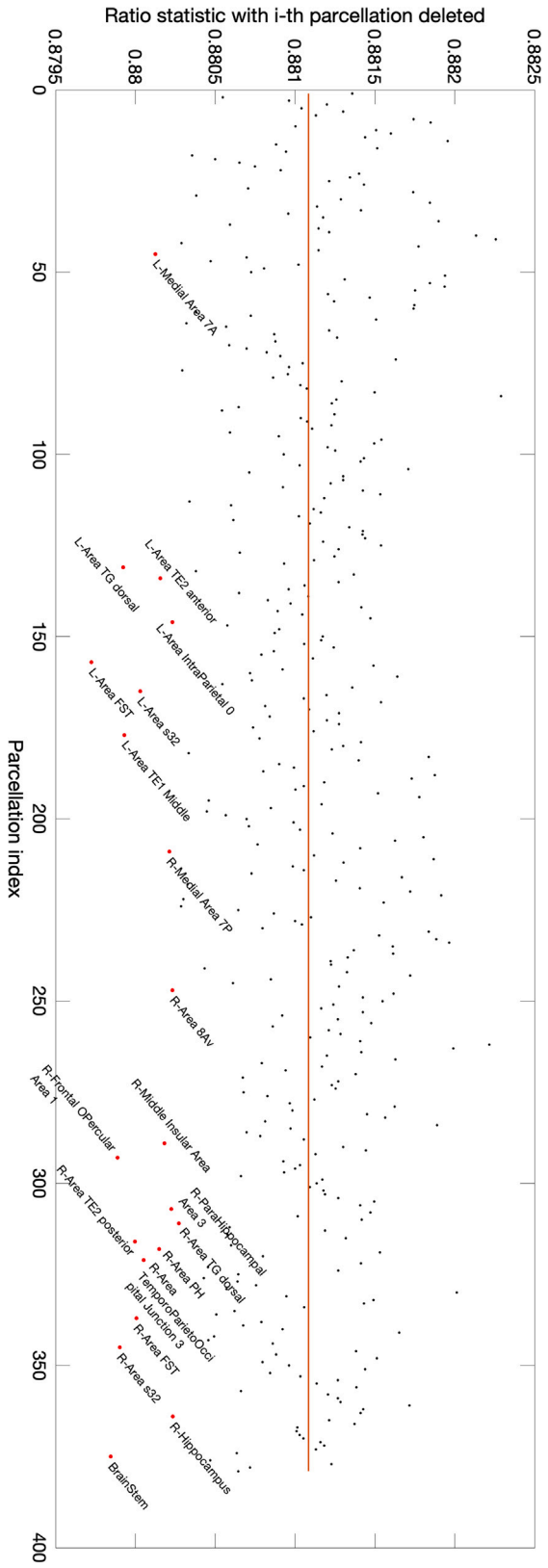

Fig. 11. The plot of the Wasserstein distance based ratio static $\phi_{\mathcal{L}}$ under the node attack. The red line is the ratio statistic of the whole brain network without any node attack. After deleting each parcellation under the node attack, we recomputed the ratio statistic (black dots). The biggest drop in the ratio statistic corresponds to the biggest topological difference for TLE. Listed 20 regions that decrease the ratio statistic and in turn decreases the discrimination power the most.

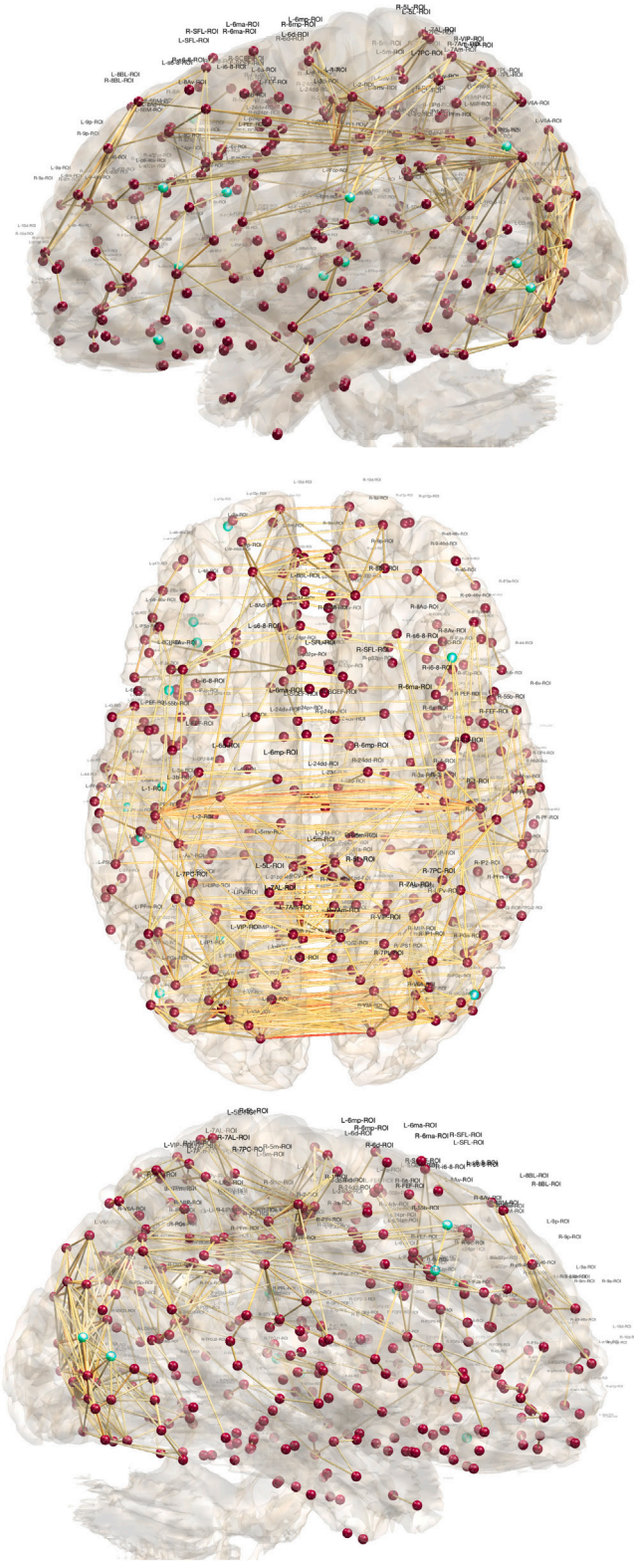

Fig. 12. 20 localized brain regions (teal color) identified under node attack on the ratio statistic $\phi_{\mathcal{L}}$ displayed over Glasser pacellated brain regions. The results are overlaid on top of average correlation map of TLE patients.

seizures, laterality, and psychosocial outcomes including cognition and psychopathology (Garcia-Ramos et al., 2021; Hermann et al., 2020, 2021). Further uses of $\Delta\psi_{\mathcal{L}}$ could be to find the regions that drive the differences between latent TLE phenotypes or as metrics for supervised

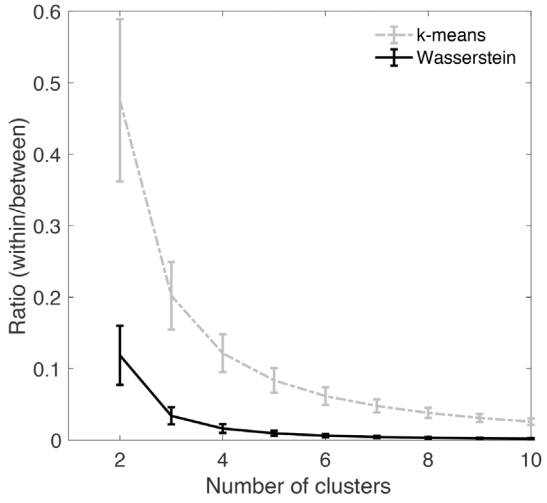

Fig. 13. The ratio ψ_l of the within-cluster distance over the between-cluster distance. The topological approach using the Wasserstein distance usually gives far smaller ratio compared to the traditional k-means clustering. In the elbow method, the largest slope change occurs at $k = 3$ and we determine $k = 3$ is the optimal number of clusters.

machine learning classification problems and regional association with cognitive or disease variables of interest, both undertakings for future studies.

5.2. Sex differences might be geometric but not topological

As supported by a substantial body of literature, sex differences in brain networks are well-established (Ingalhalikar et al., 2014; Jahanshad et al., 2011; Rubin et al., 2017). Similarly, the influence of scanner (or site) differences in brain imaging studies has long been a significant consideration (Gunter et al., 2009; Jovicich et al., 2006). However, our topological method did not detect effects related to either sex or site. If such differences are primarily geometric in nature, our topological approach would be robust against them, thereby not detecting variations attributable to sex or site. The nature of these differences, whether geometric or topological, remains unclear, as comprehensive research on this topic is limited. Most existing studies that report sex or site differences rely on geometric methods, such as traditional t-tests and ANOVA. We *hypothesize* that sex differences are primarily geometric. It is likely that the strength of connectivity in specific brain regions varies between males and females, without significant topological differences. Unlike biological sex differences, scanner differences are physical; therefore, the signal differences are somewhat artificial, and the nature of these differences remains very unclear. These issues warrant further study.

5.3. Topological characterization in focal epilepsy

Further development of topological approaches is needed to replicate some of the other key findings in rs-fMRI in focal epilepsy. It is proposed that the seizure generating region is often internally hyperconnected. Independent component analysis (Boerwinkle et al., 2016) and graph theory measure like the rich club (Lopes et al., 2017) have been used to support this hypothesis. Further methodological and empiric work is needed to developed the topological equivalents. Techniques to define the expected functional networks in topological space like the default and attention networks are needed to measure the impact of epilepsy on these networks. Additional areas for further investigation are to apply these techniques in task-related, morphological, and DWI connectivity matrices and further exploration in EEG bands for multimodal network analysis. Further the clinical, cognitive, and psychological consequences of the differing functional topology should be explored. The potential benefits of topological approaches warrant further methodological development and clinical investigation on epilepsy patients.

CRediT authorship contribution statement

Moo K. Chung: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition, Project administration. **Camille Garcia Ramos:** Data curation, Resources, Validation. **Felipe Branco De Paiva:** Writing – review & editing. **Jedidiah Mathis:** Data curation. **Vivek Prabharakaren:** Funding acquisition, Project administration. **Veena A. Nair:** Data curation, Writing – review & editing. **Elizabeth Meyerand:** Funding acquisition, Project administration. **Bruce P. Hermann:** Funding acquisition, Project administration. **Jeffrey R. Binder:** Funding acquisition, Project administration. **Aaron F. Struck:** Writing – review & editing, Funding acquisition, Project administration.

Declaration of competing interest

The study was approved and follows the University of Wisconsin-Madison and Medical College of Wisconsin IRB protocols. The study was conducted ethically following the IRB protocol. The epilepsy data is not available due to the IRB protocol.

Data availability

We made the core MATLAB computer code available at https://github.com/laplcebeltrami/PH-STAT.

Acknowledgments

This study was supported by NIH, United States U01NS093650, NS117568, EB028753, MH133614 and National Science Foundation, United States MDS-2010778. We also like to thank Sixtus Dakurah, Tahmineh Azizi, Soumya Das, and Tananun Songdechakraiwut of University of Wisconsin-Madison for discussion on the Wasserstein distance. During the preparation of this work, Chung used Chat GPT-4 in order to check mostly English usage, grammar and references. After using this tool, Chung reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Appendix A. Algebra on birth–death decompositions

We cannot build coherent statistical inference framework if we cannot even compute the sample mean and variance. Thus, we need to define valid algebraic operations on the birth–death decomposition and check if they are even valid operations. Here addition + is defined in an element-wise fashion in adding matrices while ∪ is defined for the birth–death decomposition.

Consider graph $\mathcal{X} = (V, w)$ with the birth–death decompositions $W = W_b \cup W_d$:

$$W_b = \{b_{(1)}, \dots, b_{(q_0)}\}, \quad W_d = \{d_{(1)}, \dots, d_{(q_1)}\}.$$

Let $\mathcal{F}(W) = w$ be the function that maps each edge in the ordered edge set W back to the original edge weight matrix w. $\mathcal{F}^{-1}(w) = W$ is the function that maps each edge in the edge weight matrix to the birth death decomposition. Such maps are one-to-one. Since W_b and W_d are disjoint, we can write as

$$\mathcal{F}(W_b \cup W_d) = \mathcal{F}(W_b) + \mathcal{F}(W_d).$$

Define the *scalar multiplication* on the ordered set W as

$$cW = (cW_b) \cup (cW_d) = \{cb_{(1)}, \dots, cb_{(q_0)}\} \cup \{cd_{(1)}, \dots, cd_{(q_1)}\}$$

for $c \in \mathbb{R}$. Then we have $\mathcal{F}(cW) = c\mathcal{F}(W)$ for $c \geq 0$. The relation does not hold for $c < 0$ since it is not order preserving. Define the *scalar addition* on the ordered set W as

$$c + W = (c + W_b) \cup (c + W_d) = \{c + b_{(1)}, \dots, c + b_{(q_0)}\} \cup \{c + d_{(1)}, \dots, c + d_{(q_1)}\}$$

for $c \in \mathbb{R}$. Since the addition is order preserving, $\mathcal{F}(c + W) = c + \mathcal{F}(W)$ for all $c \in \mathbb{R}$.

Define scalar multiplication of c to graph $\mathcal{X} = (V, w)$ as $c\mathcal{X} = (V, c\mathcal{F}(W))$. Define the scalar addition of c to graph $\mathcal{X}$ as $c + \mathcal{X} = (V, c + \mathcal{F}(W))$. Let $c = c_b \cup c_d$ be an ordered set with $c_b = (c^b_{(1)}, \dots, c^b_{(q_0)})$ and $c_d = (c^d_{(1)}, \dots, c^d_{(q_1)})$. Define the *set addition* of c to the ordered set W as

$$c + W = (c_b + W_b) \cup (c_d + W_d)$$

with $c_b + W_b = \{c^b_{(1)} + b_{(1)}, \dots, c^b_{(q_0)} + b_{(q_0)}\}$ and $c_d + W_d = \{c^d_{(1)} + d_{(1)}, \dots, c^d_{(q_1)} + d_{(q_1)}\}$. Then we have the following decomposition.

Theorem 5. *For graph $\mathcal{X} = (V, w)$ with the birth–death decompositions $W = W_b \cup W_d$ and positive ordered sets c_b and c_d, we have*

$$\mathcal{F}((c_b + W_b) \cup W_d) = (c_b + \mathcal{F}(W_b)) + \mathcal{F}(W_d) \tag{11}$$

$$\mathcal{F}(W_b \cup (c_d - c_\infty + W_d)) = \mathcal{F}(W_b) + \mathcal{F}(c_d - c_\infty + W_d), \tag{12}$$

where c_∞ is a large number bigger than any element in c_d.

Proof. Note $c_b + W_b$ is order preserving. W_b is the MST of graph $\mathcal{X}$. The total edge weights of MST does not decrease if we change all the edge wights of MST from W_b to $c_b + W_b$. Thus $c_b + W_b$ will be still MST and $\mathcal{F}(c_b + W_b) = c_b + \mathcal{F}(W_b)$. The death set W_d does not change when the edges in MST increases. This proves (11).

The sequence $(a_1, \dots, a_{q1}) = c_d - c_\infty$ with $a_i = c^d_{(i)} - c_\infty < 0$ is increasing. Adding $(a_1, \dots, a_{q1})$ to W_d is order preserving. Decreasing edge weights in W_d will not change the total edge weights of MST. Thus the birth set is still identical to W_b. Then the death set is $c_d - c_\infty + W_d$. This proves (12). □

The decomposition (12) does not work if we simply add an arbitrary ordered set to W_d since it will change the MST. Numerically the above algebraic operations are all linear in run time and will not increase the computational load. So far, we demonstrated what the valid algebraic operations are on the birth–death decompositions. Now we address the question of *if the birth–death decomposition is additive.*

Given graphs $\mathcal{X}_1 = (V, w^1)$ and $\mathcal{X}_2 = (V, w^2)$ with corresponding birth–death decompositions $W_1 = W_{1b} \cup W_{1d}$ and $W_2 = W_{2b} \cup W_{2d}$, define the sum of graphs $\mathcal{X}_1 + \mathcal{X}_2$ as a graph $\mathcal{X} = (V, w)$ with birth–death decomposition

$$W_b \cup W_d = (W_{1b} + W_{2b}) \cup (W_{1d} + W_{2d}). \tag{13}$$

However, it is unclear if there even exists a unique graph with decomposition (13). Define *projection* $\mathcal{F}(W_1|W_2)$ as the projection of edge values in the ordered set W_1 onto the edge weight matrix $\mathcal{F}(W_2)$ such that the birth values W_{1b} are sequentially mapped to the W_{2b} and the death values W_{1d} are sequentially mapped to the W_{2d}. Trivially, $\mathcal{F}(W_1|W_1) = \mathcal{F}(W_1)$. In general, $\mathcal{F}(W_1|W_2) \neq \mathcal{F}(W_2|W_1)$. The projection can be written as

$$\mathcal{F}(W_1|W_2) = \mathcal{F}(W_{1b}|W_{2b}) + \mathcal{F}(W_{1d}|W_{2d}).$$

Theorem 6. *Given graphs $\mathcal{X}_1 = (V, w^1)$ and $\mathcal{X}_2 = (V, w^2)$ with corresponding birth–death decompositions $W_1 = W_{1b} \cup W_{1d}$ and $W_2 = W_{2b} \cup W_{2d}$, there exists graph $\mathcal{X} = (V, w)$ with birth–death decomposition $W_b \cup W_d$ satisfying*

$$W_b \cup W_d = (W_{1b} + W_{2b}) \cup (W_{1d} + W_{2d}).$$

with

$$w = \mathcal{F}(W_b \cup W_d) = \mathcal{F}(W_{1b} + W_{2b}|W_{1b}) + \mathcal{F}(W_{1d} + W_{2d}|W_{1d}).$$

Proof. We prove by the explicit construction in a sequential manner by applying only the valid operations.

(1) Let c_∞ be some fixed number larger than any edge weights in w^1 and w^2. Add c_∞ to the decomposition $W_{1b} \cup W_{1d}$ to make all the edges positive:

$$c_\infty + (W_{1b} \cup W_{1d}) = (c_\infty + W_{1b}) \cup (c_\infty + W_{1d}). \tag{14}$$

The edge weight matrix is given by

$$\mathcal{F}((c_\infty + W_{1b}) \cup (c_\infty + W_{1d})) = c_\infty + \mathcal{F}(W_1).$$

(2) We add the ordered set W_{2b} to decomposition (14) and obtain

$$c_\infty + (W_{1b} + W_{2b}) \cup W_{1d} = (c_\infty + W_{1b} + W_{2b}) \cup (c_\infty + W_{1d}). \tag{15}$$

We next determine how the corresponding edge weight matrix changes when the birth–death decomposition changes from (14) to (15). Increasing birth values from $c_\infty + W_{1b}$ to $c_\infty + W_{1b} + W_{2b}$ increases the total edge weights in the MST of $c_\infty + \mathcal{X}_1$. Thus, $c_\infty + W_{1b} + W_{2b}$ is still MST. The death set does not change from $c_\infty + W_{1d}$. The edge weight matrix is then given by

$$\begin{aligned} &\mathcal{F}((c_\infty + W_{1b} + W_{2b}) \cup (c_\infty + W_{1d})) \\ &= \mathcal{F}(c_\infty + W_{1b} + W_{2b}|W_{1b}) + \mathcal{F}(c_\infty + W_{1d}). \end{aligned} \tag{16}$$

(16) can be also derived from (11) in Theorem 5 as well.

(3) Add ordered set $W_{2d} - c_\infty$ to the death set in the decomposition (15) and obtain

$$(c_\infty + W_{1b} + W_{2b}) \cup (W_{1d} + W_{2d}). \tag{17}$$

Decreasing death values from $c_\infty + W_{1d}$ to $W_{1d} + W_{2d}$ does not affect the total edge weights in the MST of (16). There is no change in MST. The birth set does not change from $c_\infty + W_{1b} + W_{2b}$. Thus,

$$\begin{aligned} &\mathcal{F}((c_\infty + W_{1b} + W_{2b}) \cup (W_{1d} + W_{2d})) \\ &= \mathcal{F}(c_\infty + W_{1b} + W_{2b}|W_{1b})\mathcal{F}(W_{1d} + W_{2d}|W_{1d}) \\ &= (c_\infty + \mathcal{F}(W_{1b} + W_{2b}|W_{1b})) + \mathcal{F}(W_{1d} + W_{2d}|W_{1d}) \end{aligned} \tag{18}$$

Since edge weights in $W_{2d} - c_\infty$ are all negative, we can also obtain the above result from Theorem 5.

(4) Finally we subtract c_∞ from the birth set in (17) and obtain the projection of sum onto W_1.

$$\mathcal{F}(W_{1b} + W_{2b}|W_{1b}) + \mathcal{F}(W_{1d} + W_{2d}|W_{1d}). \quad \square \tag{19}$$

Remark. Theorem 6 does not guarantee the uniqueness of edge weight matrices. Instated of projecting birth and death values onto the first graph, we can also project onto the second graph

$$\mathcal{F}(W_{1b} + W_{2b}|W_{2b}) + \mathcal{F}(W_{1d} + W_{2d}|W_{2b}).$$

or any other graph. Different graphs can have the same birth–death sets. Fig. 14 shows two different graphs with the identical birth and death sets.

Appendix B. Matlab implementation

We made MATLAB codes used in the study as part of PH-STAT (Statistical Inference on Persistent Homology): https://github.com/laplcebeltrami/PH-STAT. Simply run Matlab live script `SCRIPT.mlx`.

Graph filtration

Graph filtration is performed using the function `PH_betti.m` that inputs connectivity matrix `C` and the range of filtration values `thresholds`. It outputs Betti-curves as structured array `beta.zero` and `beta.one`, which can be displayed using `PH_betti_display.m`:

```
thresholds=[0:0.01:1];
beta = PH_betti(w, thresholds);
PH_betti_display(beta,thresholds)
```

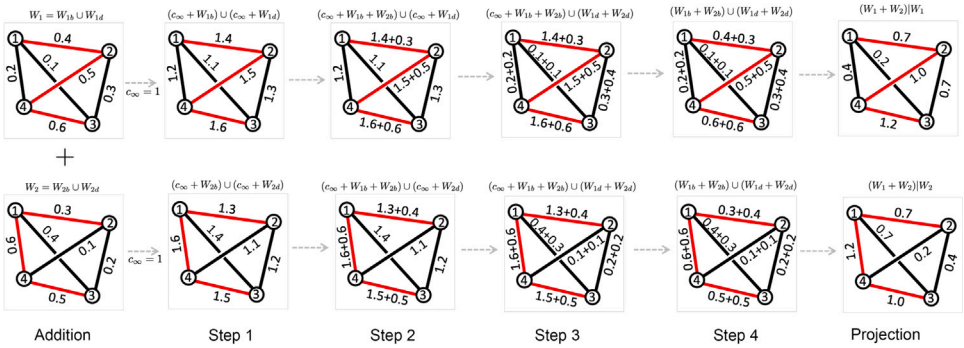

Fig. 14. Schematic of Theorem 6 with 4-nodes examples. Each step of operations yield graphs with valid birth–death decompositions. The first row is the construction of sum operation by projecting to W_1. The second row is the construction of sum operation by projecting to W_2. Red colored edges are the maximum spanning trees (MST). Each addition operation will not change MST. Eventually, we can have two different graphs with the identical birth–death decomposition.

Wasserstein distance

The 2-Wasserstein distances are computed using `WS_pdist2.m`, which inputs a collection of connectivity matrices `con_i` of size $p \times p \times m$ and `con_j` of size $p \times p \times m$ (p number of nodes and m and n samples). Then the function outputs structured array `lossMtx`, where `lossMtx.D0`, `lossMtx.D1` and `lossMtx.D01` are $(m+n) \times (m+n)$ pairwise distance matrix for 0D distance D^2_{W0}, 1D distance D^2_{W1}, combined distance $D = D^2_{W0} + D^2_{W1}$ respectively:

```
lossMtx = WS_pdist2(con_i,con_j);
WS_pdist2_display(con_i,con_j)
```

`WS_pdist2_display.m` displays the comparison between the Euclidean distance and the Wasserstein distances.

Topological inference

The observed ratio statistic $\phi_{\mathcal{L}}$ is computed using `WS_ratio.m`, which inputs one of distance matrices, such as `lossMtx.D0` or `lossMtx.D01`, and sample size in each group:

```
observed = WS_ratio(lossMtx.D01, nGroup_i, nGroup_j);
```

The transposition test on the ratio statistic is implemented as `WS_transpositions.m` and takes less than one second in a desktop computer for million permutations. The function inputs one of distance matrices such as `lossMtx.D01`, sample size in each group, number of transpositions and the number of permutations that are interjected into transpositions:

```
nTrans = 1000000;
permNo = 1000;
[transStat, ~] = WS_transpositions(lossMtx.D01,
nGroup_i, nGroup_j, nTrans, permNo);
```

In this example, we are intermixing 1000 permutations (`permNo`) in 1000000 transpositions (`nTrans`). This produces a sequence of ratio statistics`transStat` that is updated over transpositions. p-value `pval` is then computed in an iterative fashion:

```
transPval = online_pvalues(transStat', ratio);
pval = transPval(end)
```

References

Abdallah, H., Regalski, A., Kang, M.B., Berishaj, M., Nnadi, N., Chowdury, A., Diwadkar, V.A., Salch, A., 2023. Statistical inference for persistent homology applied to simulated fMRI time series data. Found. Data Sci. 5, 1–25.

Adamovich, T., Zakharov, I., Tabueva, A., Malykh, S., 2022. The thresholding problem and variability in the EEG graph network parameters. Sci. Rep. 12, 18659.

Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., Chepushtanova, S., Hanson, E., Motta, F., Ziegelmeier, L., 2017. Persistence images: A stable vector representation of persistent homology. J. Mach. Learn. Res. 18, 218–252.

Adler, R.J., Bobrowski, O., Borman, M.S., Subag, E., Weinberger, S., 2010. Persistent homology for random fields and complexes. In: Borrowing Strength: Theory Powering Applications–a Festschrift for Lawrence D. Brown. Institute of Mathematical Statistics, pp. 124–143.

Agueh, M., Carlier, G., 2011. Barycenters in the wasserstein space. SIAM J. Math. Anal. 43, 904–924.

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

Anand, D.V., Chung, M.K., 2023. Hodge-Laplacian of brain networks. IEEE Trans. Med. Imaging 42, 1563–1573.

Babai, L., Luks, E.M., 1983. Canonical labeling of graphs. In: Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing. pp. 171–183.

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

Becker, C.O., Pequito, S., Pappas, G.J., Miller, M.B., Grafton, S.T., Bassett, D.S., Preciado, V.M., 2018. Spectral mapping of brain functional connectivity from diffusion imaging. Sci. Rep. 8, 1–15.

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

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer.

Boerwinkle, V.L., Wilfong, A.A., Curry, D.J., 2016. Resting-state functional connectivity by independent component analysis-based markers corresponds to areas of initial seizure propagation established by prior modalities from the hypothalamus. Brain Connect 6, 642–651.

Bubenik, P., 2015. Statistical topological data analysis using persistence landscapes. J. Mach. Learn. Res. 16, 77–102.

Bullmore, E., Sporns, O., 2012. The economy of brain network organization. Nat. Rev. Neurosci. 13, 336–349.

Bullmore, E.T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., Brammer, M.J., 1999. Global, voxel, and cluster tests, by theory and permutation, for difference between two groups of structural MR images of the brain. IEEE Trans. Med. Imaging 18, 32–42.

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

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

Carlsson, G., Mémoli, F., 2010. Characterization, stability and convergence of hierarchical clustering methods. J. Mach. Learn. Res. 11, 1425–1470.

Chazal, F., Cohen-Steiner, D., Guibas, L.J., Mémoli, F., Oudot, S.Y., 2009. Gromov-Hausdorff stable signatures for shapes using persistence. In: Computer Graphics Forum, Vol. 28. pp. 1393–1403.

Chung, M.K., Adluru, N., Dalton, K.M., Alexander, A.L., Davidson, R.J., 2011. Scalable brain network construction on white matter fibers. In: Proc. of SPIE, Vol. 7962. p. 79624G.

Chung, M.K., Bubenik, P., Kim, P.T., 2009. Persistence diagrams of cortical surface data. In: Proceedings of the 21st International Conference on Information Processing in Medical Imaging. IPMI, In: Lecture Notes in Computer Science (LNCS), vol. 5636, pp. 386–397.

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

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

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

Chung, M.K., Lee, H., DiChristofano, A., Ombao, H., Solo, V., 2019b. Exact topological inference of the resting-state brain networks in twins. Network Neurosci. 3, 674–694.

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

Chung, M.K., Wang, Y., Huang, S.-G., Lyu, I., 2018. Rapid acceleration of the permutation test via slow random walks in the permutation group. arXiv preprint arXiv:1812.06696.

Chung, M.K., Xie, L., Huang, S.-G., Wang, Y., Yan, J., Shen, L., 2019c. Rapid acceleration of the permutation test via transpositions. In: International Workshop on Connectomics in Neuroimaging, Vol. 11848. pp. 42–53.

Cohen-Steiner, D., Edelsbrunner, H., Harer, J., 2007. Stability of persistence diagrams. Discrete Comput. Geom. 37, 103–120.

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.

Cuturi, M., Doucet, A., 2014. Fast computation of wasserstein barycenters. In: International Conference on Machine Learning. PMLR, pp. 685–693.

Das, Soumya, Ombao, Hernando, Chung, Moo K., 2022. Topological data analysis for functional brain networks. arXiv preprint arXiv:2210.09092.

Dłotko, P., Hellmer, N., 2023. Bottleneck profiles and discrete Prokhorov metrics for persistence diagrams. Discrete Comput. Geom. 1–34.

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

Duff, I.S., Koster, J., 2001. On algorithms for permuting large entries to the diagonal of a sparse matrix. SIAM J. Matrix Anal. Appl. 22 (4), 973–996.

Edelsbrunner, H., Harer, J., 2008. Persistent homology - a survey. Contemp. Math. 453, 257–282.

Edelsbrunner, H., Harer, J., 2010. Computational Topology: An Introduction. American Mathematical Society.

Edmonds, J., Karp, R.M., 1972. Theoretical improvements in algorithmic efficiency for network flow problems. J. ACM 19, 248–264.

Englot, D.J., Gonzalez, H.F.J., Reynolds, B.B., Konrad, P.E., Jacobs, M.L., Gore, J.C., Landman, B.A., Morgan, V.L., 2018. Relating structural and functional brainstem connectivity to disease measures in epilepsy. Neurology 91, e67–e77.

Englot, D.J., Mishra, A.M., Mansuripur, P.K., Herman, P., Hyder, F., Blumenfeld, H., 2008. Remote effects of focal hippocampal seizures on the rat neocortex. J. Neurosci. 28, 9066–9081.

Englot, D.J., Modi, B., Mishra, A.M., DeSalvo, M., Hyder, F., Blumenfeld, H., 2009. Cortical deactivation induced by subcortical network dysfunction in limbic seizures. J. Neurosci. 29, 13006–13018.

Feller, W., 2008. An Introduction To Probability Theory and Its Applications, Vol. 2. John Wiley & Sons.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., K., Van Der, et al., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33 (3), 341–355.

Garcia-Ramos, C., Nair, V., Maganti, R., Mathis, J., Conant, L.L., Prabhakaran, V., Binder, J.R., Meyerand, B., Hermann, B., Struck, A.F., 2022. Network phenotypes and their clinical significance in temporal lobe epilepsy using machine learning applications to morphological and functional graph theory metrics. Sci. Rep. 12, 14407.

Garcia-Ramos, C., Struck, A.F., Cook, C., Prabhakaran, V., Nair, V., Maganti, R., Binder, J.R., Meyerand, M., Conant, L.L., Hermann, B., 2021. Network topology of the cognitive phenotypes of temporal lobe epilepsy. Cortex 141, 55–65.

Ghrist, R., 2008. Barcodes: The persistent topology of data. Bull. Amer. Math. Soc. 45, 61–75.

Giusti, C., Ghrist, R., Bassett, D.S., 2016. Two's company, three (or more) is a simplex. J. Comput. Neurosci. 41, 1–14.

Glasser, M.F., Smith, S.M., Marcus, D.S., Andersson, J.L.R., Auerbach, E.J., Behrens, T.E.J., Coalson, T.S., Harms, M.P., Jenkinson, M., Moeller, S., 2016. The human connectome project's neuroimaging approach. Nature Neurosci. 19, 1175.

Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., et al., 2013. The minimal preprocessing pipelines for the human connectome project. NeuroImage 80, 105–124.

Gunter, J.L., Bernstein, M.A., Borowski, B.J., Ward, C.P., Britson, P.J., Felmlee, J.P., Schuff, N., Weiner, M., Jack, C.R., 2009. Measurement of MRI scanner performance with the ADNI phantom. Med. Phys. 36, 2193–2205.

Guo, X., Srivastava, A., 2020. Representations, metrics and statistics for shape analysis of elastic graphs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 832–833.

Hartmann, K.G., Schirrmeister, R.T., Ball, T., 2018. EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals. arXiv preprint arXiv:1806.01875.

Hayasaka, S., Phan, K.L., Liberzon, I., Worsley, K.J., Nichols, T.E., 2004. Nonstationary cluster-size inference with random field and permutation methods. NeuroImage 22, 676–687.

Hermann, B., Conant, L.L., Cook, C.J., Hwang, G., Garcia-Ramos, C., Dabbs, K., Nair, V.A., Mathis, J., Bonet, C.N.R., Allen, L., 2020. Network, clinical and sociodemographic features of cognitive phenotypes in temporal lobe epilepsy. NeuroImage: Clinical 27, 102341.

Hermann, B.P., Struck, A.F., Dabbs, K., Seidenberg, M., Jones, J.E., 2021. Behavioral phenotypes of temporal lobe epilepsy. Epilepsia Open 6, 369–380.

Huang, S.-G., Samdin, S.-T., Ting, C.M., Ombao, H., Chung, M.K., 2020. Statistical model for dynamically-changing correlation matrices with application to brain connectivity. J. Neurosci. Methods 331, 108480.

Hwang, G., Hermann, B., Nair, V.A., Conant, L.L., Dabbs, K., Mathis, J., Cook, C.J., Rivera-Bonet, C.N., Mohanty, R., Zhao, G., et al., 2020. Brain aging in temporal lobe epilepsy: Chronological, structural, and functional. NeuroImage: Clinical 25, 102183.

Ingalhalikar, M., Smith, A., Parker, D., Satterthwaite, T.D., Elliott, M.A., Ruparel, K., Hakonarson, H., Gur, R.E., Gur, R.C., Verma, R., 2014. Sex differences in the structural connectome of the human brain. Proc. Natl. Acad. Sci. 111, 823–828.

Jahanshad, N., Aganj, I., Lenglet, C., Joshi, A., Jin, Y., Barysheva, M., McMahon, K.L., De Zubicaray, G., Martin, N.G., Wright, M.J., Toga, A.W., Sapiro, G., Thompson, P.M., 2011. Sex differences in the human connectome: 4-Tesla high angular resolution diffusion imaging (HARDI) tractography in 234 young adult twins. In: Biomedical Imaging: From Nano To Macro, 2011 IEEE International Symposium on. pp. 939–943.

Jovicich, J., Czanner, S., Greve, D., Haley, E., van Der Kouwe, A., Gollub, R., Kennedy, D., Schmitt, F., Brown, G., MacFall, J., et al., 2006. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. NeuroImage 30, 436–443.

Kuang, L., Jia, J., Zhao, D., Xiong, F., Han, X., Wang, Y., Alzheimer's Disease Neuroimaging Initiative, et al., 2020. Default mode network analysis of APOE genotype in cognitively unimpaired subjects based on persistent homology. Front. Aging Neurosci. 188.

Kumar, S., Shovon, A.R., Deshpande, G., 2023. The robustness of persistent homology of brain networks to data acquisition-related non-neural variability in resting state fMRI. Hum. Brain Mapp. 44, 4637–4651.

Latora, V., Marchiori, M., 2001. Efficient behavior of small-world networks. Phys. Rev. Lett. 87, 198701.

Le, H., Kume, A., 2000. The Fréchet mean shape and the shape of the means. Adv. Appl. Probab. 32, 101–113.

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

Lee, H., Chung, M.K., Kang, H., Kim, B.-N., Lee, D.S., 2011b. Discriminative persistent homology of brain networks. In: IEEE International Symposium on Biomedical Imaging. ISBI, pp. 841–844.

Lee, H., Kang, H., Chung, M.K., Kim, B.-N., Lee, D.S, 2012. Persistent brain network homology from the perspective of dendrogram. IEEE Trans. Med. Imaging 31, 2267–2277.

Lee, M.-H., Kim, D.-Y., Chung, M.K., Alexander, A.L., Davidson, R.J., 2018a. Topological properties of the brain network constructed using the epsilon-neighbor method. IEEE Trans. Biomed. Eng. 65, 2323–2333.

Lee, M., Xiong, Y., Yu, G., Li, G.Y., 2018b. Deep neural networks for linear sum assignment problems. IEEE Wirel. Commun. Lett. 7, 962–965.

Li, J., Bian, C., Chen, D., Meng, X., Luo, H., Liang, H., Shen, L., 2020. Effect of APOE $\varepsilon 4$ on multimodal brain connectomic traits: a persistent homology study. BMC Bioinf 21, 1–18.

Liao, W., Zhang, Z., Pan, Z., Mantini, D., Ding, J., Duan, X., Luo, C., Lu, G., Chen, H., 2010. Altered functional connectivity and small-world in mesial temporal lobe epilepsy. PLoS One 5, e8525.

Lopes, M.A., Richardson, M.P., Abela, E., Rummel, C., Schindler, K., Goodfellow, M., Terry, J.R., 2017. An optimal strategy for epilepsy surgery: Disruption of the rich-club? Plos Computat. Biol. 13, e1005637.

Lord, L.-D., Expert, P., Fernandes, H.M., Petri, G., Van H., T.J., Vaccarino, F., Deco, G., Turkheimer, F., Kringelbach, M.L., 2016. Insights into brain architectures from the homological scaffolds of functional connectivity networks. Front. Syst. Neurosci. 10, 85.

Mazrooyisebdani, M., Nair, V.A., Garcia-Ramos, C., Mohanty, R., Meyerand, E., Hermann, B., Prabhakaran, V., Ahmed, R., 2020. Graph theory analysis of functional connectivity combined with machine learning approaches demonstrates widespread network differences and predicts clinical variables in temporal lobe epilepsy. Brain Connect. 10, 39–50.

Mi, L., Zhang, W., Gu, X., Wang, Y., 2018. Variational wasserstein clustering. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 322–337.

Moeller, S., Yacoub, E., Olman, C.A., Auerbach, E., Strupp, J., Harel, N., Uğurbil, K., 2010. Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. Magn. Reson. Med. 63, 1144–1153.

Mueller, S.G., Nei, M., Bateman, L.M., Knowlton, R., Laxer, K.D., Friedman, D., Devinsky, O., Goldman, A.M., 2018. Brainstem network disruption: a pathway to sudden unexplained death in epilepsy? Human Brain Mapp. 39, 4820–4830.

Newman, M.E.J., Girvan, M., 2004. Finding and evaluating community structure in networks. Phys. Rev. E 69, 026113.

Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: A primer with examples. Human Brain Mapp. 15, 1–25.

Patriat, R., Molloy, E.K., Meier, T.B., Kirk, G.R., Nair, V.A., Meyerand, M.E., Prabhakaran, V., Birn, R.M., 2013. The effect of resting condition on resting-state fMRI reliability and consistency: a comparison between resting with eyes open, closed, and fixated. NeuroImage 78, 463–473.

Petri, G., Expert, P., Turkheimer, F., Carhart-Harris, R., Nutt, D., Hellyer, P.J., Vaccarino, F., 2014. Homological scaffolds of brain functional networks. J. R. Soc. Interface 11, 20140873.

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

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

Reimann, M.W., Nolte, M., Scolamiero, M., Turner, K., Perin, R., Chindemi, G., Dłotko, P., Levi, R., Hess, K., Markram, H., 2017. Cliques of neurons bound into cavities provide a missing link between structure and function. Front. Comput. Neurosci. 11, 48.

Robinson, A., Turner, K., 2017. Hypothesis testing for topological data analysis. J. Appl. Comput. Topol. 1, 241–261.

Rubin, L.H., Yao, L., Keedy, S.K., Reilly, J.L., Bishop, J.R., Carter, C.S., Pournajafi-Nazarloo, H., Drogos, L.L., Tamminga, C.A., Pearlson, G.D., 2017. Sex differences in associations of arginine vasopressin and oxytocin with resting-state functional brain connectivity. J. Neurosci. Res. 95, 576–586.

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

Saggar, M., Sporns, O., Gonzalez-Castillo, J., Bandettini, P.A., Carlsson, G., Glover, G., Reiss, A.L., 2018. Towards a new approach to reveal dynamical organization of the brain using topological data analysis. Nature Commun. 9, 1399.

Salch, A., Regalski, A., Abdallah, H., Suryadevara, R., Catanzaro, M.J., Diwadkar, V.A., 2021. From mathematics to medicine: A practical primer on topological data analysis (TDA) and the development of related analytic tools for the functional discovery of latent structure in fMRI data. PLoS One 16, e0255859.

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

Shi, J., Zhang, W., Wang, Y., 2016. Shape analysis with hyperbolic wasserstein distance. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5051–5061.

Sizemore, A.E., Giusti, C., Kahn, A., Vettel, J.M., Betzel, R.F., Bassett, D.S., 2018. Cliques and cavities in the human connectome. J. Comput. Neurosci. 44, 115–145.

Sizemore, A.E., Phillips-Cremins, J.E., Ghrist, R., Bassett, D.S., 2019. The importance of the whole: topological data analysis for the network neuroscientist. Netw. Neurosci. 3, 656–673.

Solo, V., Poline, J.B., Lindquist, M., Simpson, S.L, Bowman, D., Chung, M.K., Cassidy, B., 2018. Connectivity in fMRI: a review and preview. IEEE Trans. Med. Imaging 37, 1537–1550.

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

Songdechakraiwut, T., Chung, M.K, 2023. Topological learning for brain networks. Ann. Appl. Stat. 17, 403–433.

Songdechakraiwut, T., Shen, L., Chung, M.K., 2021. Topological learning and its application to multimodal brain network integration. Med. Image Comput. Comput. Assist. Intervent. (MICCAI) 12902, 166–176.

Sporns, O., 2003. Graph theory methods for the analysis of neural connectivity patterns. In: Kötter, R. (Ed.), Neuroscience Databases: A Practical Guide. Springer US, Boston, MA, pp. 171–185.

Stolz, B.J., Emerson, T., Nahkuri, S., Porter, M.A., Harrington, H.A., 2021. Topological data analysis of task-based fMRI data from experiments on Schizophrenia. J. Phys. Complexity 2, 035006.

Struck, A.F., Boly, M., Hwang, G., Nair, V., Mathis, J., Nencka, A., Conant, L.L., DeYoe, E.A., Raghavan, M., Prabhakaran, V., 2021. Regional and global resting-state functional MR connectivity in temporal lobe epilepsy: Results from the epilepsy connectome project. Epilepsy Behav. 117, 107841.

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

Surampudi, S.G., Naik, S., Surampudi, R.B., Jirsa, V.K., Sharma, A., Roy, D., 2018. Multiple kernel learning model for relating structural and functional connectivity in the brain. Sci. Rep. 8, 1–14.

Ting, C.-M., Ombao, H., Samdin, S.B., Salleh, S.-H., 2018. Estimating dynamic connectivity states in fMRI using regime-switching factor models. IEEE Trans. Med. Imag. 37, 1011–1023.

Topaz, C.M., Ziegelmeier, L., Halverson, T., 2015. Topological data analysis of biological aggregation models. PLoS One e0126383.

Turner, K., Mileyko, Y., Mukherjee, S., Harer, J., 2014. Fréchet means for distributions of persistence diagrams. Discrete Comput. Geom. 52, 44–70.

Tuzhilin, A.A., 2016. Who invented the Gromov-Hausdorff distance? arXiv preprint arXiv:1612.00728.

Vallender, S.S., 1974. Calculation of the wasserstein distance between probability distributions on the line. Theory Probab. Appl. 18, 784–786.

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

Wang, Y., Ombao, H., Chung, M.K., 2018. Topological data analysis of single-trial electroencephalographic signals. Ann. Appl. Stat. 12, 1506–1534.

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

Xu, Mengjia, Sanz, David Lopez, Garces, Pilar, Maestu, Fernando, Li, Quanzheng, Pantazis, Dimitrios, 2021. A graph Gaussian embedding method for predicting alzheimer's disease progression with MEG brain networks. IEEE Trans. Biomed. Eng. 68, 1579–1588.

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

Yoo, J., Kim, E.Y., Ahn, Y.M., Ye, J.C., 2016. Topological persistence vineyard for dynamic functional brain connectivity during resting and gaming stages. J. Neurosci. Methods 267, 1–13.

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

Zalesky, A., Fornito, A., Harding, I.H., Cocchi, L., Yücel, M., Pantelis, C., Bullmore, E.T., 2010. Whole-brain anatomical networks: Does the choice of nodes matter? NeuroImage 50, 970–983.

Zavlanos, M.M., Pappas, G.J., 2008. A dynamical systems approach to weighted graph matching. Automatica 44, 2817–2824.

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

Zomorodian, A.J., 2009. Topology for Computing. Cambridge University Press, Cambridge.