Unified Topological Inference for Brain Networks in Temporal Lobe Epilepsy Using the Wasserstein Distance

Moo K. Chung, Camille Garcia Ramos, Aaron F. Struck

Abstract Persistent homology can extract hidden topological signals present in brain networks. Persistent homology summarizes the changes of topological structures through over multiple different scales called filtrations. Doing so detect hidden topological signals that persist over multiple scales. However, a key obstacle of applying persistent homology to brain network studies has always been the lack of coherent statistical inference framework. To address this problem, we present a unified topological inference framework based on the Wasserstein distance. The method is applied to the resting-state functional magnetic resonance images (rs-fMRI) of the temporal lobe epilepsy patients. We made MATLAB packag available at `https://github.com/laplcebeltrami/dynamicTDA` that was used to perform all the analysis in this study.

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.,

Moo K. Chung
Department of Biostatistics and Medical Informatics, Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin-Madison, USA e-mail: `mkchung@wisc.edu`

Camille Garcia Ramos
Department of Neurology, University of Wisconsin-Madison, USA e-mail: `garciaramos@wisc.edu`

Aaron F. Struck
Department of Neurology, University of Wisconsin-Madison, USA e-mail: `struck@neurology.wisc.edu`

2012). There is a need to develop a multiscale network analysis framework that provides consistent results and interpretation regardless of the choice of parameter. Persistent homology, a branch of algebraic topology, offers a novel solution to this multiscale analysis challenge (Edelsbrunner and Harer, 2010). Instead of examining networks at one fixed scale, persistent homology identifies persistent topological features that are robust under different scales (Petri et al., 2014; Sizemore et al., 2018). Unlike existing graph theory approaches that analyze networks at one different fixed scale at a time, persistent homology captures the changes of topological features over different scales and then identifies the most persistent topological features that are robust under noise perturbations.

Persistent homological network approaches are shown to be more robust and outperforming many existing graph theory measures and methods (Bassett and Sporns, 2017; Yoo et al., 2016; Santos et al., 2019; Songdechakraiwut and Chung, 2020) starting with (Lee et al., 2011a,b) in year 2011. In (Lee et al., 2011b, 2012), persistent homology was shown to outperform eight existing graph theory features such as clustering coefficient, samll-worldness and modularity. In (Chung et al., 2017b, 2019a), persistent homology was shown to outperform various matrix norm based network distances. In (Wang et al., 2018), persistent homology was shown to outperform the power spectral density and local variance methods. In (Wang et al., 2017), persistent homology was shown to outperform topographic power maps in an EEG study. In (Yoo et al., 2017), center persistency was shown to outperform the network-based statistic and element-wise multiple corrections. Even though persistent homology has been applied to numerous brain network studies (Lee et al., 2012; Petri et al., 2014; Sizemore et al., 2018; Chung et al., 2019b), the method has been mainly applied as an exploratory data analysis tool providing anecdotal evidence for network differences. The method still suffers the lack of more coherent statistical inference framework.

In this paper, we present a unified topological inference framework for differentiating brain networks in a two-sample comparision setting based on the Wasserstein distance. We will show that the proposed method based on the Wasserstein distance can capture the topological patterns that are consistently observed across different subjects. The Wasserstein distance or Kantorovich–Rubinstein metric is originally defined between probability distributions (Vallender, 1974; Canas and Rosasco, 2012; Berwald et al., 2018). Due to the connection to the optimal mass transport, which enjoys various optimal properties, the Wasserstein distance has been applied to various imaging applications. However, the Wasserstein distance has not seen many applications in brain imaging and network data. (Mi et al., 2018) used the Wasserstein distance in resampling brain surface meshes. (Shi et al., 2016; Su et al., 2015) used the Wasserstein distance in classifying brain cortical surface shapes. (Hartmann et al., 2018) used the Wassterstein distance in building generative adversarial networks. (Sabbagh et al., 2019) used the Wassterstein distance for manifold regression problem in the space of positive definite matrices for the source localization problem in EEG. (Xu et al., 2021) used the Wasserstein distance in predicting Alzheimer's disease progression in magnetoencephalography (MEG) brain networks. However, the Wassterstein distance in these applications are all geometric in nature.

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 $O(n \log n)$ run time in most graph manipulation tasks such as matching and averaging. The method is applied in the building a unified inference framework for discriminating networks topologically. Compared to existing graph theory feature based methods and other topological distances, the method provide more robust performance against false positives while increased sensitivity when subtle 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).

2 Methods

2.1 Graphs as simplical complexes

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

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

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

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

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

for $0 = \epsilon_0 < \epsilon_1 < \epsilon_2 < \cdots$, where the sequence of ϵ-values are called the filtration values. The filtration is quantified through a topological basis called *k-cycles*. 0-cycles are the connected components, 1-cycles are 1D closed paths or loops while 2-cycles are a 3-simplices (tetrahedron) without interior. Any k-cycle can be represented as a linear combination of basis k-cycles. The Betti numbers β_k counts the

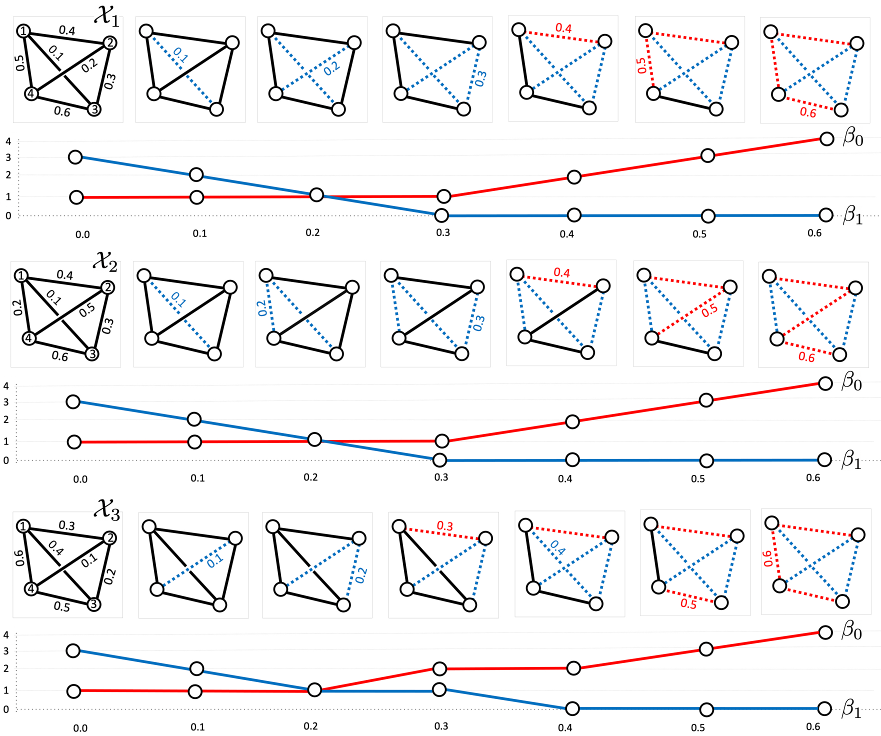

Fig. 1 Graph filtrations are obtained by sequentially thresholding graphs in increasing edge weights. The 0-th Betti number β_0 (number of connected components) and the first Betti number β_1 (number of cycles) are then plotted over the filtration values. The Betti curves are monotone over graph filtrations. 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.

number of independent k-cycles. During the Rips filtration, the i-th k-cycle is born at filtration value b_i and dies at d_i. The collection of all the paired filtration values

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

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

As the number of nodes p increases, the resulting Rips complex becomes very dense. As the filtration values increases, there exists an edge between every pair of nodes. At higher filtration values, Rips filtration becomes an ineffective representation of networks. To remedy this issue, graph filtration was introduced (Lee et al.,

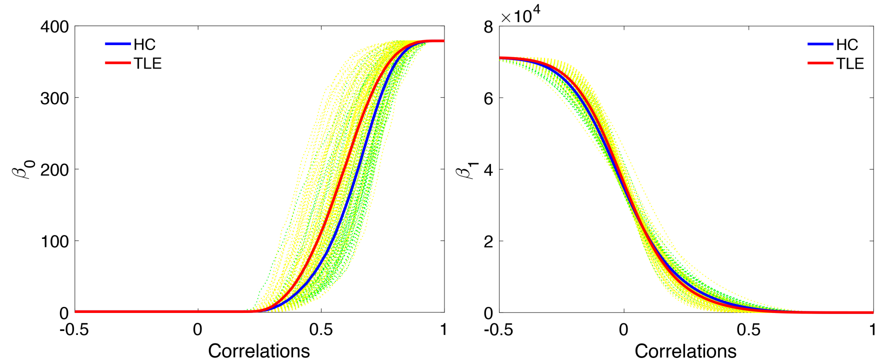

Fig. 2 Betti-0 and Betti-1 curves obtained in graph filtrations on 50 healthy controls (HC) and 101 temporal lobe epilepsy (TLE) patients. TLE has more disconnected subnetworks (β_0) compared to HC while having compatible higher order cyclic connectivity (β_1). The statistical significance of Betti curve shape difference is quantified through the proposed Wasserstein distance.

2011b, 2012). Given weighted graph $\mathcal{X} = (V, w)$ with edge weight $w = (w_{ij})$, the binary network $\mathcal{X}_\epsilon = (V, w_\epsilon)$ is a graph consisting of the node set V and the binary edge weights $w_\epsilon = (w_{\epsilon,ij})$ given by

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

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

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

The sequence of such a nested multiscale graph is called as the *graph filtration* (Lee et al., 2011b, 2012). Figure 1 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.

2.2 Birth-death decomposition

Unlike the Rips complex, there are no higher dimensional topological features beyond the 0D and 1D topology in graph filtration. The 0D and 1D persistent diagrams (b_i, d_i) tabulate the life-time of 0-cycles (connected components) and 1-cycles

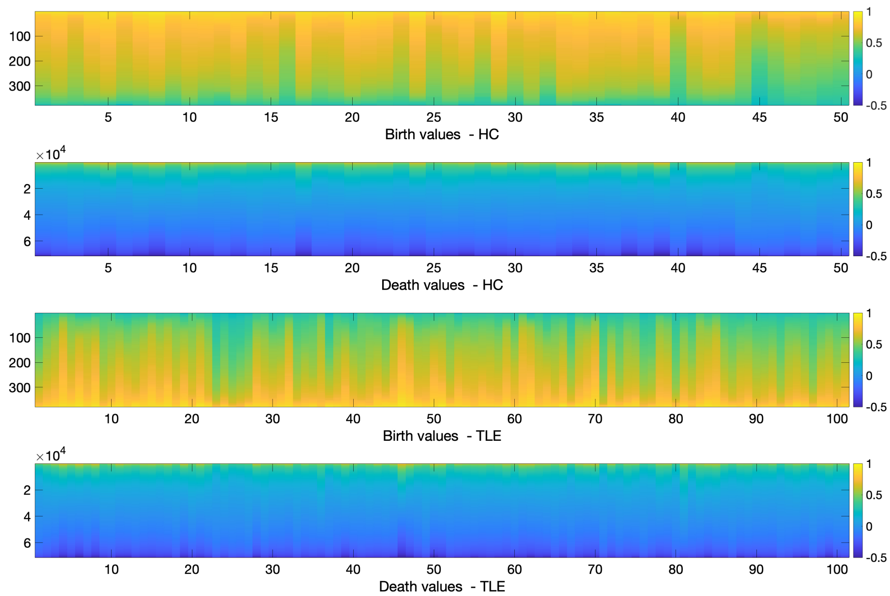

Fig. 3 The birth and death sets of 50 healthy controls (HC) and 101 temporal lobe epilepsy (TLE) patients. The Wasserstein distance between the birth sets measures 0D topology difference while the Wasserstein distance between the death sets measures 1D topology difference.

(loops) that are born at the filtration value b_i and die at value d_i. The 0th Betti number $\beta_0(w_{(i)})$ counts the number of 0-cycles at filtration value $w_{(i)}$ and shown to be non-decreasing over filtration (Figure 1) (Chung et al., 2019a): $\beta_0(w_{(i)}) \leq \beta_0(w_{(i+1)})$. On the other hand the 1st Betti number $\beta_1(w_{(i)})$ counts the number of independent loops and shown to be non-increasing over filtration (Figure 1) (Chung et al., 2019a): $\beta_1(w_{(i)}) \geq \beta_1(w_{(i+1)})$. Figure 2 displays the Betti curves plotting β_0 and β_1 values over filtration vales.

We implemented the procedure as the matlab function `PH_betti.m` inputs a connectivity matrix `C` and the range of filtration values `thresholds` where the graph filtration will be performed and outputs Betti-curves as `beta_0` and `beta_1`. During the graph filtration, when new components is born, they never dies. Thus, 0D persistent diagrams are completely characterized by birth values b_i only. Loops are viewed as already born at $-\infty$. Thus, 1D persistent diagrams are completely characterized by death values d_i only. We can show that the edge weight set W can be partitioned into 0D birth values and 1D death values (Songdechakraiwut et al., 2021):

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

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

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

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 do not happen at the same time. The claim is proved by contradiction. Assume the both events happen at the same time in contrary. Then β_0 increases by 1 while β_1 decreases by 1. When an edge is deleted, the number of nodes p is 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. The Euler characteristic can be also given by an alternating sum $\chi = \beta_0 - \beta_1$ (Adler et al., 2010). Subsequently, the Euler characteristic increases by 2, which contradict the previous computation. Thus, both events cannot occur at the same time. This establishes the decomposition $W = W_b \cup W_d, W_b \cap W_d = \emptyset$.

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

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

number of edges that destroy 1-cycles.

The 0D persistent diagram of the graph filtration is given by $\{(b_{(1)}, \infty), \cdots, (b_{(q_0)}, \infty)\}$. Ignoring ∞, W_b is the 0D persistent digram. The 1D persistent diagram of the graph filtration is given by $\{(-\infty, d_{(1)}), \cdots, (-\infty, d_{(q_1)})\}$. Ignoring $-\infty$, W_d is the 1D persistent digram. □

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. Figure 1 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. Figure 3 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.

2.2.1 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 $W = W_b \cup W_d$:

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

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

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

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

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

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 $c_b = (c^b_{(1)}, \cdots, c^b_{(q_0)})$ and $c_d = (c^d_{(1)}, \cdots, c^d_{(q_1)})$. Define the *set addition* of c to the ordered set W as

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

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

Theorem 2 *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*

$$\mathcal{F}((c_b + W_b) \cup W_d) = (c_b + \mathcal{F}(W_b)) + \mathcal{F}(W_d) \tag{2}$$
$$\mathcal{F}(W_b \cup (c_d - c_\infty + W_d)) = \mathcal{F}(W_b) + \mathcal{F}(c_d - c_\infty + W_d), \tag{3}$$

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 (2).

The sequence $(a_1, \cdots, a_{q1}) = c_d - c_\infty$ with $a_i = c^d_{(i)} - c_\infty < 0$ is increasing. Adding $(a_1, \cdots a_{q1})$ 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 (3). $\square$

The decomposition (3) 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 addictive.*

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

$$W_b \cup W_d = (W_{1b} + W_{2b}) \cup (W_{1d} + W_{2d}). \tag{4}$$

However, it is unclear if there even exists a unique graph with decomposition (4). 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 3 *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:

$$c_\infty + (W_{1b} \cup W_{1d}) = (c_\infty + W_{1b}) \cup (c_\infty + W_{1d}). \tag{5}$$

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 (5) and obtain

$$c_\infty + (W_{1b} + W_{2b}) \cup W_{1d} = (c_\infty + W_{1b} + W_{2b}) \cup (c_\infty + W_{1d}). \tag{6}$$

We next determine how the corresponding edge weight matrix changes when the birth-death decomposition changes from (5) to (6). Increasing birth values from

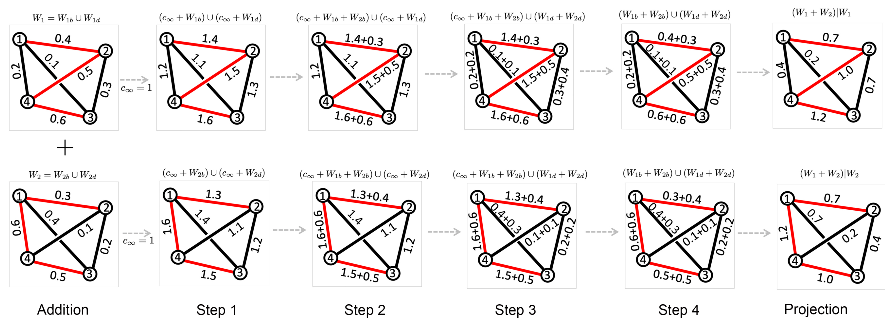

Fig. 4 Schematic of Theorem 3 with 4-nodes examples. Each step of operations yield graphs with valid birth-death decompositions. The first row is the construction of sum operation by projecting to W_1. The second row is the construction of sum operation by projecting to W_2. Red colored edges are the maximum spanning trees (MST). Each addition operation will not change MST. Eventually, we can have two different graphs with the identical birth-death decomposition.

$c_\infty + W_{1b}$ 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

$$\begin{aligned} &\mathcal{F}((c_\infty + W_{1b} + W_{2b}) \cup (c_\infty + W_{1d})) \\ &= \mathcal{F}(c_\infty + W_{1b} + W_{2b}|W_{1b}) + \mathcal{F}(c_\infty + W_{1d}). \end{aligned} \tag{7}$$

(7) can be also derived from (2) in Theorem 2 as well.

3) Add ordered set $W_{2d} - c_\infty$ to the death set in the decomposition (6) and obtain

$$(c_\infty + W_{1b} + W_{2b}) \cup (W_{1d} + W_{2d}). \tag{8}$$

Decreasing death values from $c_\infty + W_{1d}$ to $W_{1d} + W_{2d}$ does not affect the the total edge weights in the MST of (7). There is no change in MST. The birth set does not change from $c_\infty + W_{1b} + W_{2b}$. Thus,

$$\begin{aligned} &\mathcal{F}((c_\infty + W_{1b} + W_{2b}) \cup (W_{1d} + W_{2d})) \\ &= \mathcal{F}(c_\infty + W_{1b} + W_{2b}|W_{1b})\mathcal{F}(W_{1d} + W_{2d}|W_{1d}) \\ &= (c_\infty + \mathcal{F}(W_{1b} + W_{2b}|W_{1b})) + \mathcal{F}(W_{1d} + W_{2d}|W_{1d}) \end{aligned} \tag{9}$$

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

4) Finally we subtract c_∞ from the brith set in (8) and obtain the projection of sum onto W_1.

$$\mathcal{F}(W_{1b} + W_{2b}|W_{1b}) + \mathcal{F}(W_{1d} + W_{2d}|W_{1d}). \tag{10}$$

$\square$

Remark. Theorem 3 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. Figure 5 shows two different graphs with the identical birth and death sets.

2.3 Wasserstein distance on graph filtrations

Consider persistent diagrams P_1 and P_2 given by 2D scatter points

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

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

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

Then we can show that the 2-*Wasserstein distance* on persistent diagrams is given by

$$D_W(P_1, P_2) = \inf_{\psi: P_1 \to P_2}\Big(\sum_{x \in P_1}\|x - \psi(x)\|^2\Big)^{1/2} \tag{11}$$

over every possible bijection ψ between P_1 and P_2 (Vallender, 1974). Optimization (11) is the standard assignment problem, which is usually solved by Hungarian algorithm in $O(q^3)$ (Edmonds and Karp, 1972). However, for graph filtration, the distance can be computed in $O(q \log q)$ by simply matching the order statistics on birth or death sets (Rabin et al., 2011; Songdechakraiwut et al., 2021):

Theorem 4 *The 2-Wasserstein distance between the 0D persistent 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 i-th smallest birth values in persistent diagram P_j. The 2-Wasserstein distance between the 1D persistent diagrams for graph filtration is given by

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

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

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

$$D^2_{W0}(P_1, P_2) = \min_{\psi} \sum_{i=1}^{q_0} |b^1_i - \psi(b^1_i)|^2,$$

where the minimum is taken over every possible bijection ψ from $\{b^1_1, \cdots, b^1_{q_0}\}$ to $\{b^2_1, \cdots, b^2_{q_0}\}$. Note $\sum_{i=1}^{q_0} |b^1_i - \psi(b^1_i)|^2$ is minimum only if $\sum_{i=1}^{q_0} b^1_i \psi(b^1_i)$ is maximum. Rewrite $\sum_{i=1}^{q_0} b^1_i \psi(b^1_i)$ in terms of the order statistics as $\sum_{i=1}^{q_0} b^1_{(i)} \psi(b^1_{(i)})$. Now, we prove by *induction*. When $q = 2$, there are only two possible bijections:

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

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

$$\max_{\psi} \sum_{i=1}^{k+1} b^1_{(i)} \psi(b^2_{(i)}) \leq \max_{\psi} \sum_{i=1}^{k} b^1_{(i)} \psi(b^1_{(i)}) + \max_{\psi} b^1_{(k+1)} \psi(b^1_{(k+1)}).$$

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 persistent diagram of graph filtration is given by $\{(-\infty, d_{(1)}), \cdots, (-\infty, d_{(q)})\}$. 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. □

2.4 Graph matching via the Wasserstein distance

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 heterogenous 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 persistent diagrams P_1^0 and P_2^0 and 1D persistent diagrams P_1^1 and P_2^1. We define the Wasserstein distance between graphs $\mathcal{X}_1$ and $\mathcal{X}_2$ as the Wasserstein distance between corresponding persistent diagrams P_1 and P_2:

$$D_{Wj}(\mathcal{X}_1, \mathcal{X}_2) = D_{Wj}(P_1^j, P_2^j).$$

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

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

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

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

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

2.5 Wasserstein graph mean

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

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

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). However, the method has not seen many applications in modeling graphs and networks.

With Theorem 3, 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 persistent diagrams. Now, we define the *Wasserstein graph mean* $\mathbb{E}\mathcal{X}$ of $\mathcal{X}_1, \cdots, \mathcal{X}_n$ as

$$\mathbb{E}\mathcal{X} = \frac{1}{n}\sum_{k=1}^{n} \mathcal{X}_k. \tag{12}$$

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

Theorem 5 *The Wasserstein graph mean is the graph given by*

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

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 $W_b = \{b_{(1)}, \cdots, b_{(q_0)}\}$ and $W_d = \{d_{(1)}, \cdots, d_{(q_1)}\}$. From Theorem 4,

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

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

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

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. ∎

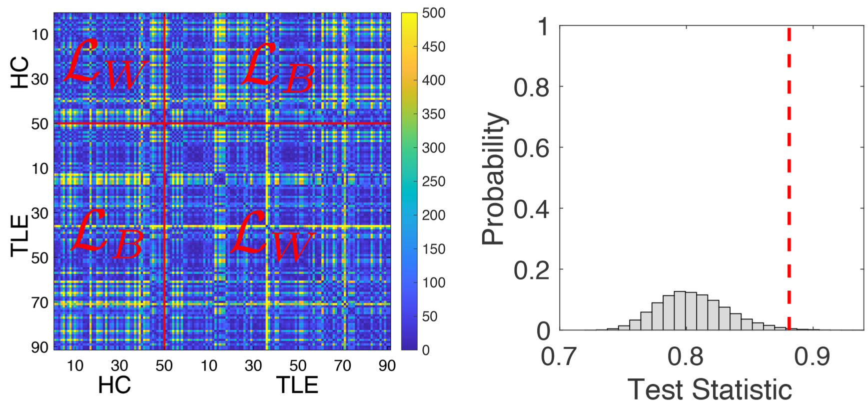

Fig. 5 Pairwise Wasserstein distance between 50 healthy controls (HC) and 101 temporal lobe epilepsy (TLE) patients. There are subtle pattern difference in the off-diagonal patterns (between group distances) compared to diagonal patterns (within group distances). The permutation test with half-million permutations was used to determine the statistical significance using the ratio ststatistic. The red line is the observed ratio. The histrogram is the empirical null distribution obtained from the permutation test.

The *Wasserstein graph variance* $\mathbb{V}\mathcal{X}$ is defined in a similar fashion:

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

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

$$\begin{aligned} \mathbb{V}\mathcal{X} &= \frac{1}{n}\sum_{k=1}^{n} \mathcal{D}\Big(\frac{1}{n}\sum_{j=1}^{n} \mathcal{X}_j, \mathcal{X}_k\Big) \\ &= \frac{1}{n^2}\sum_{j,k=1}^{n} \mathcal{D}(\mathcal{X}_j, \mathcal{X}_k). \end{aligned} \tag{13}$$

The formulation (13) compute the variance using the pairwise distances without the need for computing the Wasserstein graph mean.

2.6 Topological inference

There are not 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. Consider a collection of graphs $\mathcal{X}_1, \cdots, \mathcal{X}_n$ that are grouped into two groups C_1 and C_2 such that

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

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

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

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

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

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

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

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

The within-group distance corresponds to the sum of all the pairwise distances in the block diagonal matrices in Figure 5. The between-group distance $\mathcal{L}_B$ is defined as

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

The between-group distance corresponds to the off-diaognal block matrices in Figure 5. Note that the sum of within-group and between-group distance is the sum of all the pairwise distances in Figure 5:

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

When we permute the group labels, the total sum of all the pairwise distances do not change and fixed. If the group difference is large, the between-group distance $\mathcal{L}_B$ will be large and the within-group distance $\mathcal{L}_W$ will be small. Thus, to measure the disparity between groups as the ratio (Songdechakraiwut and Chung, 2022)

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

The ratio statistic is related to the elbow method in clustering and behaves like traditional F-statistic, which is the ratio of squared variability of model fits. If $\phi_{\mathcal{L}}$ is large, the groups differ significantly in network topology. If $\phi_{\mathcal{L}}$ is small, it is likely that there is no group differences. Since the ratio is always positive, its probability distribution cannot be Gaussian. Since the distributions of the ratio $\phi_{\mathcal{L}}$ is unknown, the permutation test can be used to determine the empirical distributions. Figure 5-right displays the empirical distribution of $\phi_{\mathcal{L}}$. The p-value is the area of the right tail thresholded by the observed ratio $\phi_{\mathcal{L}}$ (dotted red line) in the empirical distribution. Since we only compute the pairwise distances only once and only shuffle each entry over permutations. This is equivalent tot to rearranging rows and columns of entries corresponding to the permutations in Figure 5. The simple rearranging of rows and columns of entries and sum them in the block-wise fashion should be faster than the usual two-sample t test which has to be recomputed for each permutation.

To speed up the permutation further, we adapted the transposition test, the online version of permutation test (Chung et al., 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 k-th and l-th graphs between the groups (denoted as τ_{kl}), all the k-th and i-th rows and columns will be swapped. The within-group distance after the transposition τ_{kl} is given by

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

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

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

The ratio statistic over the transposition is then sequentially updated over random transpositions. To further accelerate the convergence and avoid potential bias, we introduce permutation to the sequence of 1000 consecutive transpositions. The whole procedure is implemented as a Matlab function `WS_transpositions.m` in `https://github.com/laplcebeltrami/dynamicTDA` and takes less than one second in a desktop computer for half million permutations.

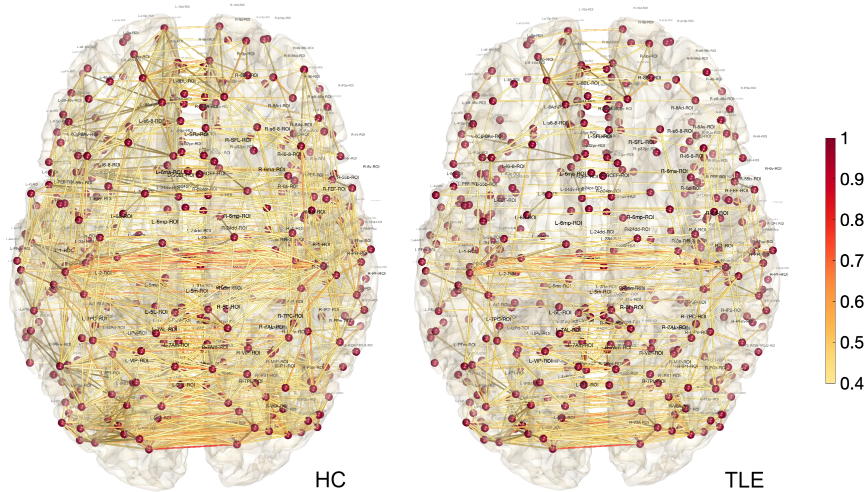

Fig. 6 The average correlation brain networks of 50 healthy controls (HC) and 101 temporal lobe epilepsy (TLE) patients. They are overlaid on top of the white matter boundary of the MNI template. The brain network of TLE is far sparse compared to that of HC. The sparse TLE network is also consistent with the plot Betti-0 curve where TLE networks are more disconnected than HC networks.

3 Application

3.1 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 control (mean age 31.78 ± 10.32 years) and 101 chronic temporal lobe epilepsy (TLE) patients (mean age 40.23 ± 11.85). The resting-state fRMI 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.) (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

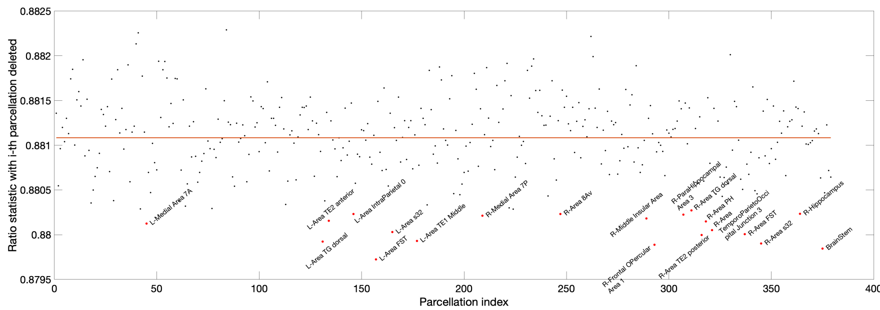

Fig. 7 The plot of the Wasserstein distance baed ratio static $\phi_{\mathcal{L}}$ under node attack. After deleting each parcellations, we computed the ratio statistic. Listed 20 regions that decrease the ratio statistic and in turn decreases the discrimination power.

Wisconsin-Madison while 62 TLE patients were scanned at the Medical College of Wisconsin.

MRI 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 397 connectivity matrix per subject. 180 brain regions reported (Glasser et al., 2013) is 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. Figure 6 displays the average connectivity in HC and TLE. TLE shows far sparse more fractured network topology compared to HC.

3.2 Topological difference in temporal lobe epilepsy

Since the images were collected in two different sites, we tested if there is any site effects. 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 0.64 with half million permutations indicating there is no site effect observed in healthy controls. We also compared 39 TLE from UW and 62 TLE from MCW. We obtained the p-value of 0.51 with half million permutations indicating there is no site effect observed in healthy controls. Thus, we did not account for site effects in comparing health controls and TLE. The topological method does not penalize the geometric differences such as correlation differences but only topological differences.

The proposed method is applied in subsequently applied comparing 50 healthy and 101 TLE patients. We obtained the p-value of 0.0065 after half million transpositions.

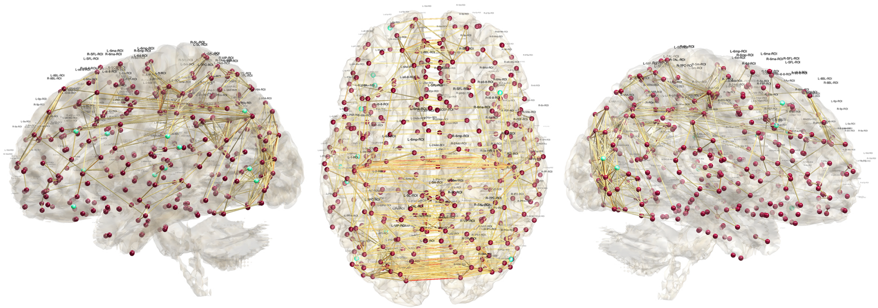

Fig. 8 20 localized brain regions (teal color) identified under node attack on the ratio statistic $\phi_{\mathcal{L}}$. The results are overlaid on top of average correlation map of TLE patients.

We concluded there is strong topological difference between the brain networks. Unlike existing topological data analysis that cannot extract local connections that is responsible for topological differences, our method can localize the source of topological differences.

In traditional TDA, it is difficult if not impossible to localize the brain regions responsible for topological difference. However, using the node attack (Lee et al., 2018), we determine the contribution of each node to the overall contribution of ratio statistic $\phi_{\mathcal{L}}$ (Figure 7). The larger the ratio statistic is, it is more discriminative of the groups. The 20 regions that decrease the ratio statistic most are listed in Figure 7. 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. The results are identified as teal colored nodes in Figure 8.

4 Conclusion

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 persistent 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 i-th largest birth edges and death edges that optimize the ∞-Wasserstein distance can be easily identifiable. This is left as a future study.

Acknowledgement

This study was supported by NIH R01 EB022856 and NSF MDS-2010778. We also like to thank Sixtus Dakurah, Tahmineh Azizi, Soumya Das, and Tananun Songdechakraiwut of University of Wisconsin-Madison for discussion on Wasserstein distances.

References

R.J. Adler, O. Bobrowski, M.S. Borman, E. Subag, and S. Weinberger. Persistent homology for random fields and complexes. In *Borrowing strength: theory powering applications–a Festschrift for Lawrence D. Brown*, pages 124–143. Institute of Mathematical Statistics, 2010.

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

L. Babai and E.M. Luks. Canonical labeling of graphs. In *Proceedings of the fifteenth annual ACM symposium on Theory of computing*, pages 171–183, 1983.

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

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

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

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

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

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

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

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

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

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

R.W. Cox. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, 29:162–173, 1996.

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

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

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

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

W. Feller. *An introduction to probability theory and its applications*, volume 2. John Wiley & Sons, 2008.

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

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

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

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

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

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

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

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

H. Lee, M.K. Chung, H. Kang, B.-N. Kim, and D.S. Lee. Discriminative persistent homology of brain networks. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 841–844, 2011a.

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

H. Lee, H. Kang, M.K. Chung, B.-N. Kim, and D.S Lee. Persistent brain network homology from the perspective of dendrogram. *IEEE Transactions on Medical Imaging*, 31:2267–2277, 2012.

M.-H. Lee, D.-Y. Kim, M.K. Chung, A.L. Alexander, and R.J. Davidson. Topological properties of the brain network constructed using the epsilon-neighbor method. *IEEE Transactions on Biomedical Engineering*, 65:2323–2333, 2018.

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

S. Moeller, E. Yacoub, C.A. Olman, E. Auerbach, J. Strupp, N. Harel, and journal=Magnetic Resonance in Medicine volume=63 pages=1144–1153 year=2010 Uğurbil, K. 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.

R. Patriat, E.K. Molloy, T.B. Meier, G.R. Kirk, V.A. Nair, M.E. Meyerand, V. Prabhakaran, and R.M. Birn. 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, 2013.

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

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

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

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

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

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

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

T. Songdechakraiwut and M.K Chung. Topological learning for brain networks. *Annals of Applied Statistics*, in press:arXiv:2012.00675, 2022.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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