Registration and Joint Identification of Cycles in Brain Networks

Sixtus Dakurah[1] and Moo K Chung[1]

University of Wisconsin-Madison, Madison WI 53715, USA
{sdakurah,mkchung}@wisc.edu

Abstract. In topological data analysis, matching topological signals across subjects is of paramount importance. We develop a novel procedure for registering and matching cycles in brain networks using tools from algebraic topology. This procedure will allow for the joint identification of cycles in brain networks. A modification of the Wasserstein distance that is computationally scalable is proposed to discriminate between persistence diagrams emanating from registered cycles. We showed that the registration scheme provides for a more topologically discriminative set of features. The method is applied in comparing cycles present in male and female functional brain networks.

Keywords: Topological data analysis · Cycles Registration · Persistent homology · Joint Identification · Hodge Laplacian

1 Introduction

The structure and function of the human brain network are characterized by synchrony in neurophysical activities [15]. The neuronal interactions underlying this synchrony transcend dyadic or pairwise interactions and are primarily driven by polyadic or higher-order interactions [15]. The *cycles* in the human brain network embeds these higher-order interactions [3, 15]. Further, the cycles in the brain are associated with the propagation of information and the accompanying feedback loop, as well as the redundancy and information bottleneck problems [11]. This posits the cycles as communication channels along which signals on shared or co-activated neural functions are propagated. A natural question is then how one identify the meaningful features underlying these signals while reducing the effect of noise? To explore this question, we propose to use tools and concepts from topological data analysis (TDA) [1, 6, 7, 13].

One of the main tools of TDA is persistent homology (PH), which allows for the computation of topological features of space at different spatial resolutions [7]. Previously, PH has been applied to establish algebraic and statistical properties of topological features [3]. The central theme in these studies is to develop inference procedures on topological features that persists over wide spatial scales, which are likely to present as signals. To study the cycles in brain networks from the construct of PH, one first casts the brain networks as simplicial complexes, which are the basic building blocks in TDA for representing complex data. Brain

networks are normally expressed as graphs. In TDA, graphs are 1D simplicial complexes [9]. The cycles in brain networks are topological invariants, and their homology can be probed through their simplicial complex representation.

To identify meaningful features underlying the cycles, one approach is to compare the homology generating the cycles using topological distances. Two common topological distances with established stability properties for comparing the homology of topological objects are the Bottleneck and the Wasserstein distances [4, 16]. The main inputs to these distances are the persistence diagrams, which are the 2D embedding of the persistent homology as scatter points. In the context of 1D simplicial complexes, this simplifies to comparing the sorted 1D scatter points constituting the persistent diagrams. However, this simple procedure conflates the effect of comparing topological features with similar or dissimilar topological features [13].

In this study we develop a novel procedure for registering and matching cycles in brain networks using tools from algebraic topology. The main idea is to isolate matching and non-matching cycles by comparing their homology generators, which we termed as the *cycle registration*. This allows for the joint identification of cycles across topological spaces. We use a modification of the Wasserstein distance that is computationally more scalable in discriminating persistence diagrams. We show that the matching process reduces spurious significant cycles in a discriminating process. The method is applied in comparing cycles present in male and female functional brain networks.

2 Preliminaries

Graphs as Simplicial Complexes A k-simplex $\sigma_k = (v_0, \cdots, v_k)$ is a k-dimensional polytope of nodes $v_0, \cdots, v_k$. A simplicial complex K is a finite set of simplices such that for any $\sigma_k^i, \sigma_k^j \in K$, $\sigma_k^i \cap \sigma_k^j$ is a face of both simplices; and a face of any $\sigma_k^i \in K$ is also a simplex in K [7]. A 0-skeleton is a simplicial complex consisting of only nodes, whiles a 1-skeleton consists of nodes and edges. A k-chain is a finite sum of simplices. For two successive chain groups $\mathcal{K}_k$ and $\mathcal{K}_{k-1}$, the boundary operator $\partial_k : \mathcal{K}_k \longrightarrow \mathcal{K}_{k-1}$ for each σ_k is given by

$$\partial_k(\sigma_k) = \sum_{i=0}^{k} (-1)^i (v_0, \cdots, \widehat{v_i}, \cdots, v_k), \tag{1}$$

where $(v_0, \cdots, \widehat{v_i}, \cdots, v_k)$ gives the k-1 faces of σ_k obtained by deleting node $\widehat{v_i}$. The matrix representation $\mathbb{B}_k = (\mathbb{B}_k^{ij})$ of the boundary operator has the form

$$\mathbb{B}_k^{ij} = \begin{cases} 1, & \text{if } \sigma_{k-1}^i \subset \sigma_k^j \text{ and } \sigma_{k-1}^i \sim \sigma_k^j \\ -1, & \text{if } \sigma_{k-1}^i \subset \sigma_k^j \text{ and } \sigma_{k-1}^i \nsim \sigma_k^j \\ 0, & \text{if } \sigma_{k-1}^i \not\subset \sigma_k^j \end{cases}, \tag{2}$$

where $\sim$ and $\nsim$ denote similar and dissimilar orientations respectively. Two important components of the boundary map (1) are its kernel $ker(\partial_k)$ and image

$img(\partial_{k+1})$, which are subspaces of $\mathcal{K}_k$. The elements of $ker(\partial_k)$ and $img(\partial_{k+1})$ are known as *k-cycles* and *k-boundaries* respectively [9]. Graphs are 1D simplicial complex, and hence are topological spaces [9]. For graphs, $img(\partial_2) = \emptyset$ and the first homology module $\mathcal{H}_1(\mathcal{X}) = \ker(\partial_1)$, whose elements are 1-cycles. The rank of $\mathcal{H}_1(\mathcal{X})$ gives the first Betti number β_1 [3, 14].

Graph Filtration & Birth-death Decomposition We consider a graph $\mathcal{X} = (V, E, w)$ with node set V, edge set E, and edge weight matrix $w = (w_{ij})$. The binary graph $\mathcal{X}_\epsilon = (V, E, w_\epsilon)$ is a graph with binary edge weights $w_\epsilon = (w_{\epsilon,ij})$, where $w_{\epsilon,ij} = 0$ if $w_{ij} > \epsilon$ and is 0 otherwise. A graph filtration for ordered thresholds $\epsilon_0 < ... < \epsilon_k$ is a collection of sequence of nested binary networks [3]

$$\mathcal{X}_{\epsilon_0} \supset ... \supset \mathcal{X}_{\epsilon_k}. \tag{3}$$

As ϵ_k increases, more edges are disconnected, increasing the number of connected components (0-cycles) (β_0), and decreasing the number of 1-cycles (β_1). β_0 and β_1 are monotone over filtration values [3]. The filtration tracks the persistence of the cycles [20]. The persistence of a cycle that appears at filtration value b_i and disappears at filtration value d_i is given by the interval $[b_i, d_i]$. A finite collection of $[b_i, d_i]$ can be summarized in the form of a *barcode*. During the filtration, 0-cycles are born at b_i and never dies (∞) whiles 1-cycles are born when the graph is formed ($-\infty$) and die at d_i. Ignoring $\pm\infty$, the collection of birth values $B(\mathcal{X})$

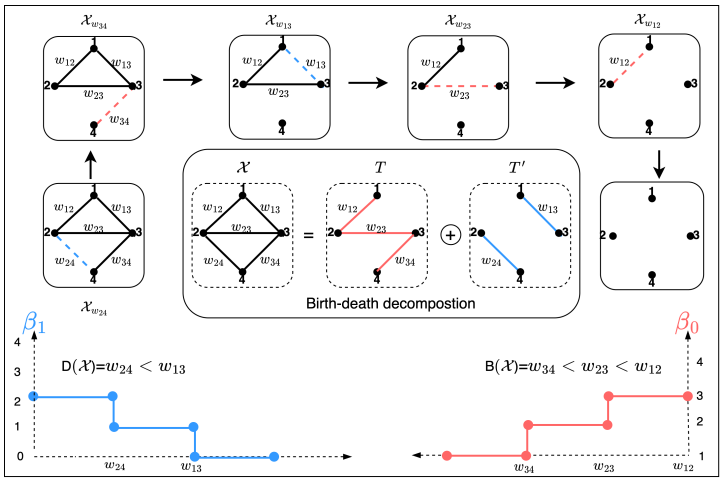

Fig. 1: Graph filtration on a four-node network. β_0 is monotonically increasing whiles β_1 is monotonically decreasing.

and death values $D(\mathcal{X})$ can be represented as the 0D and 1D *barcode*:

$$B(\mathcal{X}) = b_1 < b_2 < \cdots < b_p, \quad D(\mathcal{X}) = d_1 < d_2 < \cdots < d_q.$$

The sets $B(\mathcal{X})$ and $D(\mathcal{X})$ partition edge weight $\{w_{ij}\}$ such that $B(\mathcal{X}) \cup D(\mathcal{X}) = \{w_{ij}\}$ and $B(\mathcal{X}) \cap D(\mathcal{X}) = \emptyset$. Figure 1 illustrates this decomposition.

Hodge Laplacian The Hodge Laplacian $\mathcal{L}_k$ is a higher-dimensional generalization of the graph Laplacian for k-simplices [11]. It has the algebraic form:

$$\mathcal{L}_k = \mathbb{B}_{k+1}\mathbb{B}_{k+1}^\top + \mathbb{B}_k^\top \mathbb{B}_k. \tag{4}$$

The k-th homology group $\mathcal{H}_k$ is the kernel of $\mathcal{L}_k$, i.e., $\mathcal{H}_k = ker\mathcal{L}_k$ which is spanned by the eigenvectors corresponding to the zero eigenvalues of $\mathcal{L}_k$ [11]. The eigendecomposition of $\mathcal{L}_k$ is given by $\mathcal{L}_k\mathbf{U}_k = \mathbf{U}_k\boldsymbol{\Lambda}_k$, where $\boldsymbol{\Lambda}_k$ is a diagonal matrix of eigenvalues with corresponding eigenvectors in the columns of $\mathbf{U}_k$. For graphs or 1-skeletons, the boundary matrix $\mathbb{B}_2 = 0$ and $\mathcal{L}_1 = \mathbb{B}_1^\top \mathbb{B}_1$ [11].

Homology Basis The homology of a topological space is characterized by its k-dimensional cycles [7]. The basis of $\mathcal{H}_1(\mathcal{X})$ are the algebraically independent 1-cycles in $\mathcal{X}$. To find basis, consider the birth-death decomposition of $\mathcal{X}$ [1, 6]. The birth set forms the maximum spanning tree (MST) (T) and and the death set the non-MST part (T'). If the m-th edge σ_1^m in T' is added to T, a subgraph $\mathcal{X}^m$ with exactly one 1-cycle is formed. The Hodge Laplacian on $\mathcal{X}^m$ yields the eigendecomposition identifying the 1-cycle. The m-th 1-cycle $\mathcal{C}_m$ is given by

$$\mathcal{C}_m = \sum_{j=1}^{|T_1|} c_m^j, \quad \text{where} \quad c_m^j = \begin{cases} u_m^j \sigma_1^j, & \text{if } \sigma_1^j \in \mathcal{X}^m \\ 0, & \text{otherwise} \end{cases}. \tag{5}$$

Here, $\mathbf{u}_m$ is the m-th eigenvector (a column of $\mathbf{U}_1$) corresponding to a zero eigenvalue and u_m^j is the *j-th* entry in this vector. We can show that $\mathcal{H}_1 = ker\mathcal{L}_1$, and $\{\mathcal{C}_1, \cdots, \mathcal{C}_{|K'|}\}$ spans $ker\mathcal{L}_1$ [1, 6].

3 Cycle Registration and Joint Identification

Registration in Homology Groups Consider continuous functions f and g that maps between topological spaces $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$:

$$f : \mathcal{X} \to \mathcal{Z}, \qquad g : \mathcal{Y} \to \mathcal{Z}. \tag{6}$$

Let $\mathcal{H}_1(\mathcal{X})$, $\mathcal{H}_1(\mathcal{Y})$, $\mathcal{H}_1(\mathcal{Z})$ denote the homology groups associated with these spaces. Let f', g' be the induced homology maps. We assume further $\mathcal{X} \cup \mathcal{Y} \subset \mathcal{Z}$. For any inclusion relation ($\subset$) between 1-skeletons, the inclusion will be taken with reference to their binary representation as defined in (3). Then we have $\mathcal{H}_1(\mathcal{X}) \to \mathcal{H}_1(\mathcal{Z}) \leftarrow \mathcal{H}_1(\mathcal{Y})$, where $\leftarrow$ or $\rightarrow$ denotes a mapping in the respective direction [2, 13]. Two elements $\mathcal{C}^x \in \mathcal{H}_1(\mathcal{X})$ and $\mathcal{C}^y \in \mathcal{H}_1(\mathcal{Y})$ are said to match via space $\mathcal{Z}$ if

$$[f'(\mathcal{C}^x)] = [g'(\mathcal{C}^y)] \neq \emptyset.$$

This matched cycle is a non-trivial cycle in $\mathcal{Z}$ with generators in both mappings defined in (6). Repeating this process for all non-trivial cycles in $\mathcal{H}_1(\mathcal{Z})$ will generate a matching of all possible cycles between $\mathcal{H}_1(\mathcal{X})$ and $\mathcal{H}_1(\mathcal{Y})$ [2, 13]. The process described above matches cycles in the homology groups in an abstract sense. In what follows, we extend the concept to match persistent cycles [5, 13].

Image-Persistent Homology To extend the registration to the persistence of the homology generators, the image-persistent homology is adopted [5, 13]. We reformulate it under the framework of graph filtration. Let

$$\mathcal{F}_{\mathcal{X}} = \{\mathcal{X}_{\epsilon_k} : \epsilon_k < \epsilon_{k+1}, \forall k\}, \mathcal{F}_{\mathcal{Y}} = \{\mathcal{Y}_{\epsilon_k} : \epsilon_k < \epsilon_{k+1}, \forall k\}, \mathcal{F}_{\mathcal{Z}} = \{\mathcal{Z}_{\epsilon_k} : \epsilon_k < \epsilon_{k+1}, \forall k\}$$

be graph filtration on the spaces $\mathcal{X}, \mathcal{Y}$, and $\mathcal{Z}$ respectively. The inclusion map f_{ϵ_k} and g_{ϵ_k} are defined on the filtered topological spaces:

$$f_{\epsilon_k} : \mathcal{X}_{\epsilon_k} \to \mathcal{Z}_{\epsilon_k}, \quad g_{\epsilon_k} : \mathcal{Y}_{\epsilon_k} \to \mathcal{Z}_{\epsilon_k}. \tag{7}$$

The graph filtration embeds maps between subsequent filtrations of the form $\mathcal{X}_{\epsilon_{k+1}} \xrightarrow{\iota} \mathcal{X}_{\epsilon_k}, \mathcal{Y}_{\epsilon_{k+1}} \xrightarrow{\iota} \mathcal{Y}_{\epsilon_k}, \mathcal{Z}_{\epsilon_{k+1}} \xrightarrow{\iota} \mathcal{Z}_{\epsilon_k}$. Coupling with (7), we obtain the commutative diagram (Figure 2-left) and the induced commutative diagram between the homology groups (Figure 2-right). The commutative diagram for $\mathcal{Y}$ can be similarly drawn. The commutative property is essential to ensure that all fil-

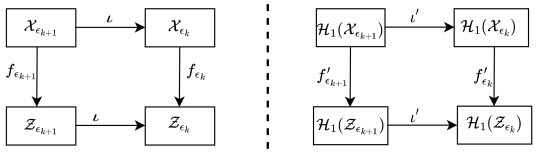

Fig. 2: Left: the nested structure of the graph filtration led to the observed commutative diagram. Right: The mapping between the topological spaces induces a homology map which extends the commutative property to the homology groups.

tered spaces at the same filtration leads to the same result. To characterize the induced persistence maps, we formalize the construction of the graph filtration as follows.

(i) (Stable homology) During the filtration, for any $\epsilon_k \notin \mathcal{E}$, we can always find small enough $\gamma > 0$ such that the map $f'_{\epsilon_k} : \mathcal{H}_1(\mathcal{X}_{\epsilon_k+\gamma}) \to \mathcal{H}_1(\mathcal{X}_{\epsilon_k-\gamma})$. The map f'_{ϵ_k} induces isomorphsim and there is no change in homology.
(ii) (Death of 1-cycles) For any $\epsilon_k \in \mathcal{E}$, there exists small enough $\gamma > 0$ such that $f'_{\epsilon_k} : \mathcal{H}_1(\mathcal{X}_{\epsilon_k+\gamma}) \to \mathcal{H}_1(\mathcal{X}_{\epsilon_k-\gamma})$ and $\beta_1(\mathcal{X}_{\epsilon_k+\gamma}) = \beta_1(\mathcal{X}_{\epsilon_k-\gamma}) - 1$. The homology change is the destruction of a single 1-cycle.

This construction allows us to define the images of the induced maps f'_{ϵ_k} as

$$Im(f'_{\epsilon_k}) = \{f'_{\epsilon_k}(\mathcal{C}) \in \mathcal{H}_1(\mathcal{X}_{\epsilon_k}) : \mathcal{C} \in \mathcal{H}_1(\mathcal{X}_{\epsilon_{k+1}})\}.$$

Note that these images are persistent homology groups, hence the name image-persistent homology [5]. The collection $\{Im(f'_{\epsilon_k})\}$ is a filtration satisfying

$$Im(f'_{\epsilon_0}) \supset Im(f'_{\epsilon_1}) \supset \cdots \supset Im(f'_{\epsilon_k}),$$

where $\epsilon_0 < \cdots < \epsilon_k$ are ordered thresholds. The elements of $Im(f'_{\epsilon_k})$ are the image-persistent cycles [5]. Computing the image-persistent homology under

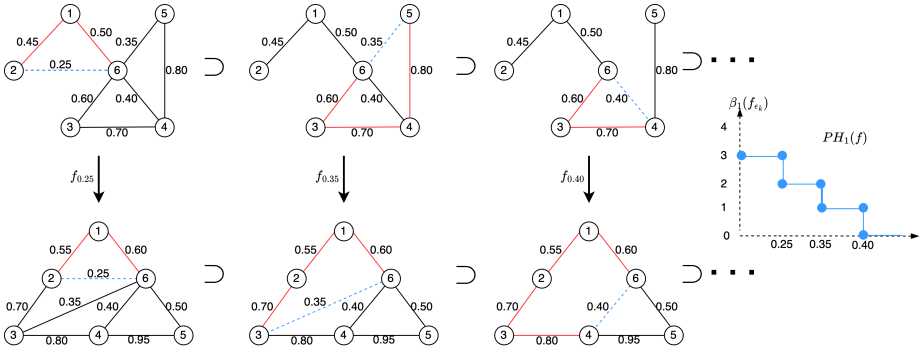

Fig. 3: Top: the filtration $\mathcal{F}_{\mathcal{X}}$ of the space $\mathcal{X}$. Bottom: the filtration $\mathcal{F}_{\mathcal{Z}}$ of the space $\mathcal{Z}$ with $\mathcal{X} \subset \mathcal{Z}$. The mapping f_{ϵ_k} between the two spaces induces a mapping f'_{ϵ_k} between the homology groups. At each ϵ_k, the blue dashed-line indicates the points that make up the barcodes. When combined with the solid red lines, it gives the matching elements in $\mathcal{H}_1(\mathcal{X}_{\epsilon_k})$ and $\mathcal{H}_1(\mathcal{Z}_{\epsilon_k})$. Right: the persistence diagram tracking the life-span of the image-persistent cycles.

graph filtration follows the same process as in Figure 1. Using the image persistence, we can now establish matching between persistence cycles.

Cycle Registration in Persistent Homology Using the image persistence, we now define a matching of cycles in persistent homology. Let $PH_1(\mathcal{X})$, and $PH_1(\mathcal{Y})$ be the persistence homology associated with $\mathcal{X}$ and $\mathcal{Y}$ respectively. Let $PH_1(f)$ and $PH_1(g)$ be the corresponding image-persistence homology. Observe that the persistence intervals are the basis elements of PH_1. Further, these persistence intervals characterized by their birth and death values generate the persistence cycles in $\mathcal{H}_1$. The cycles with their persistence (birth to death interval) given by $\mathcal{I}_x \in PH_1(\mathcal{X})$ and $\mathcal{I}_y \in PH_1(\mathcal{Y})$ are said to match via the third space $\mathcal{Z}$ if the following three conditions are satisfied:

$$\tilde{\mathcal{I}}_f(b) = \mathcal{I}_x(b), \quad \tilde{\mathcal{I}}_g(b) = \mathcal{I}_y(b), \quad \tilde{\mathcal{I}}_f(d) = \tilde{\mathcal{I}}_g(d). \tag{8}$$

Here $\tilde{\mathcal{I}}_f(b)$ and $\tilde{\mathcal{I}}_f(d)$ respectively index the birth value b and death value d in $\tilde{\mathcal{I}}_f$. Under graph filtration, persistent-cycle matching can be attained by checking only the last equality since the birth time is $-\infty$.

Metric for Comparing Cycles under Registration We modified the 2-Wasserstein metric for comparing cycles under registration. Let $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ be the persistence diagrams associated with the spaces $\mathcal{X}$ and $\mathcal{Y}$. The 2-Wasserstein distance $\mathcal{D}_2$ between $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$ under graph filtration is given by [14]

$$\left[\mathcal{D}_2\left(P_{\mathcal{X}}, P_{\mathcal{Y}}\right)\right]^2 = \sum_{i=1}^{q} |d_i^x - d_i^y|^2, \tag{9}$$

where $d_i^x \in P_{\mathcal{X}}$ is the i-th smallest death value. When there is a mismatch in the number of elements of $P_{\mathcal{X}}$ and $P_{\mathcal{Y}}$, data augmentation can be used [14].

However, 2-Wasserstein distance is not necessarily a good metric for topological spaces under registration. Consider the following example. Figure 4 gives two topological spaces $\mathcal{X}$ and $\mathcal{Y}$ and their topological union $\mathcal{Z} = \mathcal{X} \cup \mathcal{Y}$. The topological union is defined as the average of the two spaces. We perform the birth-death

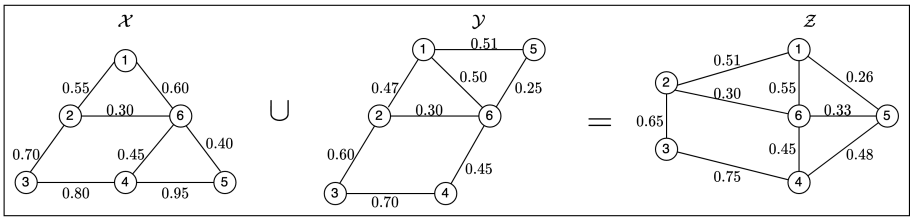

Fig. 4: Top: The two topological spaces $\mathcal{X}$ and $\mathcal{Y}$ and their topological union $\mathcal{Z}$. Bottom: The cycle spaces associated with the topological spaces $\mathcal{X}$ and $\mathcal{Y}$. The 2-Wasserstein distance between the two spaces is computed as $\mathcal{D}_2(P_{\mathcal{X}}, P_{\mathcal{Y}}) = 0.11$.

decomposition and obtain the persistent diagrams $P_{\mathcal{X}} = \{0.30, 0.40, 0.45\}$ and $P_{\mathcal{Y}} = \{0.25, 0.30, 0.45\}$. The 2-Wasserstein distance is computed as $\mathcal{D}_2(P_{\mathcal{X}}, P_{\mathcal{Y}}) = 0.11$. This estimate understates the true difference between the two spaces as will be shown in the next section. Observe that if two cycles match via the third space $\mathcal{Z}$, they do not contribute significantly to the difference between the two spaces. Hence by considering only the scatter points in differentiating the cycle space of $\mathcal{X}$ and $\mathcal{Y}$, we conflate the effect of matching cycles and non-matching cycles. In what follows, we incorporate the matching information in constructing the discriminating metric.

Modified Wasserstein Distance Let $d_x \sim d_y$ denotes that d_x and d_y generates non-trivial matching cycles over the space $\mathcal{Z}$, and $d_x \nsim d_y$ denotes that there does not exists matching non-trivial cycles generated by d_x and d_y over the space $\mathcal{Z}$. The modified 2-Wasserstein distance is defined as

$$\left[\widetilde{\mathcal{D}}_2(P_{\mathcal{X}}, P_{\mathcal{Y}})\right]^2 = \sum_{d_i^x \sim d_i^y} |d_i^x - d_i^y|^2 + \sum_{d_i^x \nsim d_i^y} |d_i^x - d_i^y|^2. \tag{10}$$

We note that the first half of this equation vanishes when $\mathcal{I}_z$ is in a discrete space. Now consider the example given in Figure 4. If we register the cycles in the two spaces $\mathcal{X}$ and $\mathcal{Y}$ over their topological union $\mathcal{Z}$ using our developed registration scheme, then the modified Wasserstein distance between the spaces evaluates to $\widetilde{\mathcal{D}}_2(P_{\mathcal{X}}, P_{\mathcal{Y}}) = 0.15$. The only difference between the two spaces is given by the cycle generated by $d_2^x = 0.4$ and $d_1^y = 0.25$.

4 Validation

Since there is no topological ground truth in real data, we demonstrate the matching scheme and the discriminatory power of the proposed Wasserstein distance through a simulation. The three topological spaces used in this simulation are modelled after the skeletons of Tropaeolum, Chardonnay and Cabernet (Figure 5-top). Eighteen points were specifically chosen at cordinates (Figure 5-

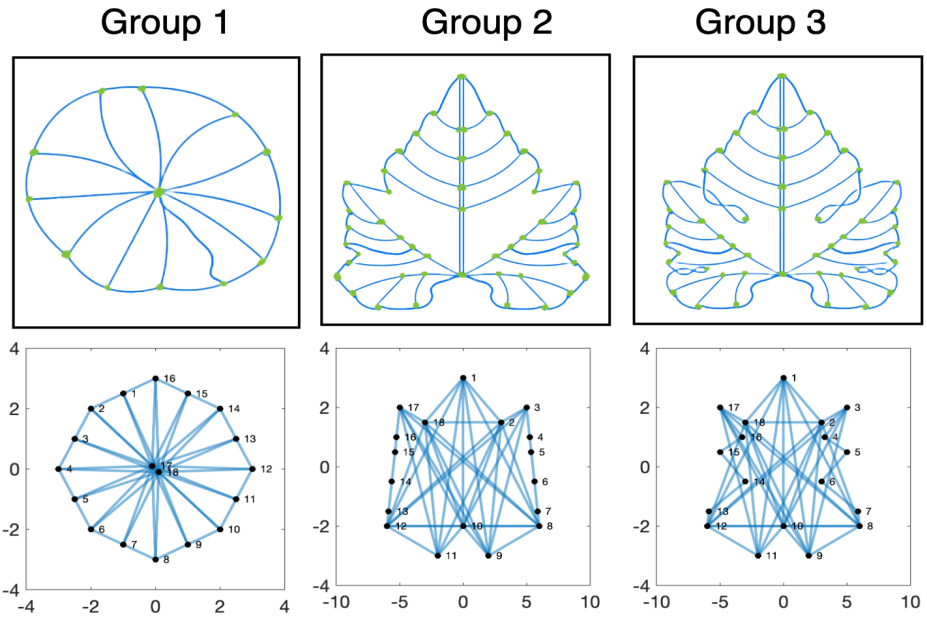

Fig. 5: The topological spaces used in the simulation. Group 1 (Tropaeolum) is topologically different from Group 2 and Group 3. Group 2 (Chardonnay) is topologically equivalent to Group 3 (Cabernet).

bottom) along these skeletons and perturbed with noise $N(0, 0.025)$, and seven networks were generated in each group. For two groups of networks $\mathcal{N}_1$, $\mathcal{N}_2$, we compute a ratio statistic as follows. Let $\mathfrak{L}_W$ denote the average of all within-

group Wasserstein distances

$$\mathfrak{L}_W = \Big(\sum_{\mathcal{X}\in\mathcal{N}_1,\mathcal{Y}\in\mathcal{N}_1} \widetilde{\mathcal{D}}_2(P_{\mathcal{X}}, P_{\mathcal{Y}}) + \sum_{\mathcal{X}\in\mathcal{N}_2,\mathcal{Y}\in\mathcal{N}_2} \widetilde{\mathcal{D}}_2(P_{\mathcal{X}}, P_{\mathcal{Y}}) \Big) / \Big(\binom{m}{2} + \binom{n}{2} \Big). \quad (11)$$

Let $\mathfrak{L}_B$ be the average of all between-group Wasserstein distances

$$\mathfrak{L}_B = \Big(\sum_{\mathcal{X}\in\mathcal{N}_1,\mathcal{Y}\in\mathcal{N}_2} \widetilde{\mathcal{D}}_2(P_{\mathcal{X}}, P_{\mathcal{Y}}) \Big) / (m \times n). \quad (12)$$

The ratio statistic is then constructed as [14]

$$\mathcal{R}(\mathcal{N}_1, \mathcal{N}_2) = \mathfrak{L}_B / \mathfrak{L}_W. \quad (13)$$

We used $\binom{14}{7} = 3432$ permutations to approximate the distribution of the test statistic (13) and compute the p-values. We validated our method against widely used geometric distances: $\mathfrak{L}_1-$, $\mathfrak{L}_2-$, $\mathfrak{L}_\infty$-norms, Bottleneck (BN), Gromov-Hausdorff (GH) as well as the standard Wasserstein distance (WS) distances [3]. The results are given in Table 1, where the average p-values of 10 independent simulations are reported. We also reported the false positive and false negative rates computed as the fraction of 10 simulations with p-values below 0.05 or above 0.05 [14]. We tested if our method can detect *topological equivalence* by comparing

Table 1: The performance results summarized as average p-values, false positive rates (first 4 rows) and false negative rates (last 2 rows). Group 2 and 3 have matching cycles and are topologically identical. Group 1 differs topologically from 2 and 3. Smaller false positive rates and false negative rates are preferred.

Groups	$\mathfrak{L}_1$	$\mathfrak{L}_2$	$\mathfrak{L}_\infty$	GH	BN	WS	$\mathcal{R}$
1 vs. 1	0.43 ± 0.25 (0.00)	0.43 ± 0.23 (0.00)	0.50 ± 0.31 (0.00)	0.45 ± 0.32 (0.10)	0.51 ± 0.29 (0.00)	0.47 ± 0.23 (0.1)	0.59 ± 0.21 (0.00)
2 vs. 2	0.50 ± 0.35 (0.10)	0.47 ± 0.36 (0.10)	0.41 ± 0.33 (0.20)	0.44 ± 0.26 (0.10)	0.36 ± 0.28 (0.20)	0.58 ± 0.41 (0.00)	0.65 ± 0.27 (0.00)
3 vs. 3	0.52 ± 0.33 (0.10)	0.53 ± 0.34 (0.10)	0.51 ± 0.38 (0.20)	0.42 ± 0.31 (0.10)	0.38 ± 0.33 (0.20)	0.53 ± 0.32 (0.10)	0.53 ± 0.27 (0.10)
2 vs. 3	0.00 ± 0.00 (1.00)	0.00 ± 0.00 (1.00)	0.00 ± 0.00 (1.00)	0.67 ± 0.13 (0.00)	0.02 ± 0.03 (0.90)	0.25 ± 0.19 (0.10)	0.43 ± 0.32 (0.00)
1 vs. 2	0.00 ± 0.00 (0.00)	± 0.00 (0.00)	0.00 ± 0.00 (0.00)	0.17 ± 0.00 (1.00)	0.00 ± 0.00 (0.00)	0.00 ± 0.00 (0.00)	0.00 ± 0.00 (0.00)
1 vs. 3	0.00 ± 0.00 (0.00)	± 0.00 (0.00)	0.00 ± 0.00 (0.00)	0.63 ± 0.04 (1.00)	0.00 ± 0.00 (0.00)	0.00 ± 0.00 (0.00)	0.00 ± 0.00 (0.00)

between groups (the first 4 rows in Table 1). The test should not detect signals and higher p-values and smaller false positive rates are preferred. The proposed ratio statistic $\mathcal{R}$ outperformed the other distances. We also tested if the proposed method can detect *topological difference* by comparing Groups 1 vs. 2,

1 vs. 3 (the last 2 rows in Table 1). The test procedures should detect signals and smaller p-values and smaller false negative rates are preferred. $\mathcal{R}$ performed comparable to the other distances.

This simulation shows that the matching procedure coupled with the Wasserstein metric performs extremely well in not producing topological false positives. When there is topological differences, there are always accompanying geometric signals so existing geometric methods usually perform reasonably.

5 Application

Dataset We applied the method to the resting state functional magnetic resonance imaging (rs-fMRI) data of 412 subjects obtained from the human connectome project [18, 17]. rs-fMRI are collected at 2 mm isotropic voxels and 1200 time points [17]. Data that went through the standard minimal preprocessing pipelines [8] was used. Volumes with framewise displacement larger than 0.5mm and their neighbors were scrubbed [18, 17]. Twelve subjects having excessive head movement were excluded. Subsequently, the Automated Anatomical Labeling (AAL) template is used to parcellate and average rs-fMRI spatially into 116 non-overlapping anatomical regions [10]. The final data is comprised of the fMRI of 400 subjects of which 232 are females and 168 are males.

Sexual Dimorphism in Cycles The networks are fully connected with $p = 116$ brain regions (nodes), hence we expect $q = p(p-1)/2 = 6555$ linearly independent 1-cycles [14] per network. We tested difference in the cycles space between the male and female networks using the Wasserstein distance. 100000 permutations were used to approximate the distribution of the statistic (13) and compute the p-value. The observed test statistic is 1.0609 and the p-value obtained is 0.1103. We can further identify the specific cycles contributing to this difference.

Most Discriminating Cycles We compute the matching score that quantifies each connections contribution to the observed difference based on the following steps. i) For any pair of networks, determine the cycle generators in each network. ii) Perform the cycle registration between the two networks according to (8). iii) For all such pairwise matching, compute the proportion of time a given connection generated a matching cycle. The proportion is termed the *matching score*. Connections with lower *matching score* will generate the most discriminating cycles across groups. Figure 6-middle shows the *matching score* between networks in the male and female groups.

We identify the six most discriminating cycles by choosing the cycle generators (connections) with the least matching score. Figure 6 shows the six most discriminating cycles corresponding to lowest values in the matching scores. These cycles contributes the most to the difference between the male and female group of networks. The Lobule X of vermis is present in five out of the six cycles. The

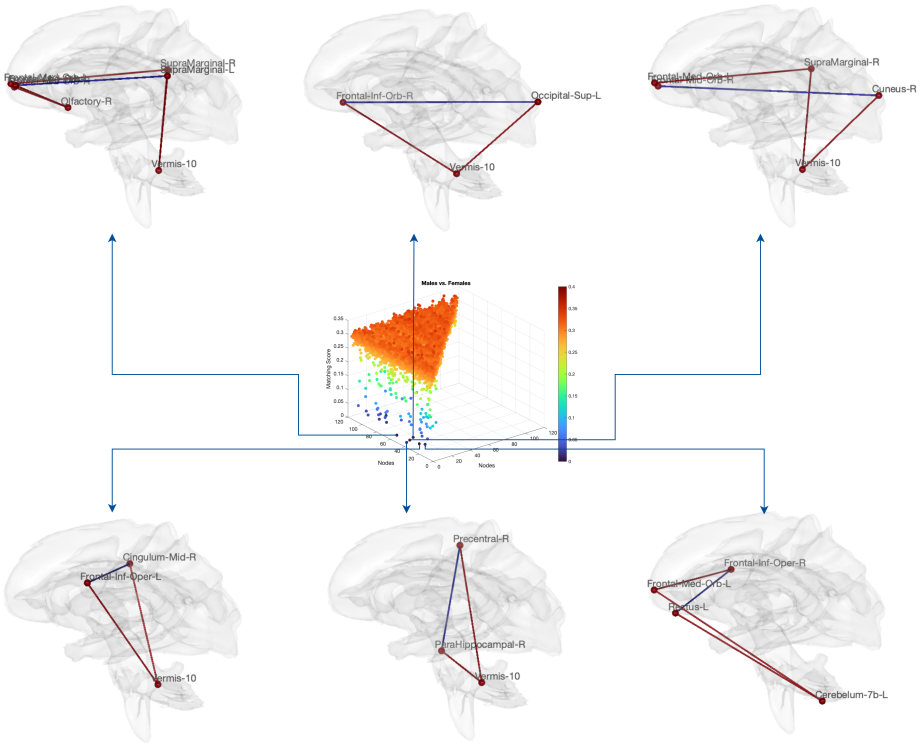

Fig. 6: From top-left, the six most discriminating cycles corresponding to lowest values in the matching scores. These cycles contributes the most to the significance difference between the male and female group of networks based on the modified Wasserstein distance. Middle: the plot of the matching score.

vermis propagates neural signals to three core brain regions and studies have shown a differential effect of sex on the vermis and its propagation of neural signals in brain networks [12, 19].

6 Conclusion

We proposed a novel framework for registering and jointly identifying cycles in brain networks. We showed that the registration scheme provides for a more topologically discriminative set of features. The six most discriminating 1-cycles identified through this registration contain brain regions: the Lobule X of vermis (vermis-10), the right olfactory cortex (Olfactory-R), and the supramarginal gyrus (SupraMarginal-L, SupraMarginal-R), and the amygdala (Amygdala-R, Amygdala-L) (Figure 6). This work serves as a baseline for the registration and joint identification of higher-dimensional topological features in brain networks, and is left as a future study.

References

1. Anand, D.V., Dakurah, S., Chung, M.K.: Hodge-laplacian of brain networks. arXiv preprint arXiv:2110.14599 (2021)
2. Carlsson, G., De Silva, V.: Zigzag persistence. Foundations of computational mathematics **10**(4), 367–405 (2010)
3. Chung, M.K., Huang, S.G., Gritsenko, A., Shen, L., Lee, H.: Statistical inference on the number of cycles in brain networks. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). pp. 113–116. IEEE (2019)
4. Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. In: Proceedings of the twenty-first annual symposium on Computational geometry. pp. 263–271 (2005)
5. Cohen-Steiner, D., Edelsbrunner, H., Harer, J., Morozov, D.: Persistent homology for kernels, images, and cokernels. In: Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms. pp. 1011–1020. SIAM (2009)
6. Dakurah, S., Anand, D.V., Chen, Z., Chung, M.K.: Modelling cycles in brain networks with the hodge laplacian. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 326–335. Springer (2022)
7. Edelsbrunner, H., Harer, J., et al.: Persistent homology-a survey. Contemporary mathematics **453**, 257–282 (2008)
8. 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.: The minimal preprocessing pipelines for the human connectome project. Neuroimage **80**, 105–124 (2013)
9. Hatcher, A., Press, C.U., of Mathematics, C.U.D.: Algebraic Topology. Algebraic Topology, Cambridge University Press (2002)
10. Huang, S.G., Samdin, S.B., Ting, C.M., Ombao, H., Chung, M.K.: Statistical model for dynamically-changing correlation matrices with application to brain connectivity. Journal of neuroscience methods **331**, 108480 (2020)
11. Lee, H., Chung, M.K., Kang, H., Lee, D.S.: Hole detection in metabolic connectivity of alzheimer's disease using k-laplacian. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 297–304. Springer (2014)
12. Raz, N., Dupuis, J.H., Briggs, S.D., McGavran, C., Acker, J.D.: Differential effects of age and sex on the cerebellar hemispheres and the vermis: a prospective mr study. American Journal of Neuroradiology **19**(1), 65–71 (1998)
13. Reani, Y., Bobrowski, O.: Cycle registration in persistent homology with applications in topological bootstrap. arXiv preprint arXiv:2101.00698 (2021)
14. Songdechakraiwut, T., Shen, L., Chung, M.: Topological learning and its application to multimodal brain network integration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 166–176. Springer (2021)
15. Sporns, O., Tononi, G., Kötter, R.: The human connectome: a structural description of the human brain. PLoS computational biology **1**(4), e42 (2005)
16. Su, Z., Zeng, W., Wang, Y., Lu, Z.L., Gu, X.: Shape classification using wasserstein distance for brain morphometry analysis. In: International Conference on Information Processing in Medical Imaging. pp. 411–423. Springer (2015)
17. Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.M.H., et al.: The wu-minn human connectome project: an overview. Neuroimage **80**, 62–79 (2013)

18. Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.E., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S.W., et al.: The human connectome project: a data acquisition perspective. Neuroimage **62**, 2222–2231 (2012)
19. Womer, F.Y., Tang, Y., Harms, M.P., Bai, C., Chang, M., Jiang, X., Wei, S., Wang, F., Barch, D.M.: Sexual dimorphism of the cerebellar vermis in schizophrenia. Schizophrenia research **176**(2-3), 164–170 (2016)
20. Zomorodian, A., Carlsson, G.: Computing persistent homology. Discrete & Computational Geometry **33**(2), 249–274 (2005)