Topological Clustering of Aphasic Brain Networks

Jiaying Yi
Dept. of Epidemiology and Biostatistics
University of South Carolina
Columbia, SC 29201
jiayingy@email.sc.edu

Jian Yin *
Dept. of Biostatistics
City University of Hong Kong
Hong Kong
jian.yin@my.cityu.edu.hk

Rahul Ghosal
Dept. of Epidemiology and Biostatistics
University of South Carolina
Columbia, SC 29201
rghosal@mailbox.sc.edu

Dirk Den Ouden
Dept. of Communication Sciences and Disorders
University of South Carolina
Columbia, SC 29201
denouden@sc.edu

Julius Fridriksson
Dept. of Communication Sciences and Disorders
University of South Carolina
Columbia, SC 29201
fridriks@mailbox.sc.edu

Rutvik Desai
Dept. of Psychology
University of South Carolina
Columbia, SC 29201
rutvik@sc.edu

Yuan Wang
Dept. of Epidemiology and Biostatistics
University of South Carolina
Columbia, SC 29201
wang578@mailbox.sc.edu

Abstract

Topological data analysis (TDA) is a powerful tool for detecting hidden structures in complex data like biological signals and networks. A key TDA algorithm, persistent homology (PH), captures multi-scale topological features in data robust to noise, as summarized by persistence diagrams (PDs). However, The non-Euclidean nature of PDs complicates traditional analysis. Recent topological inference methods use heat kernel (HK) expansion of PDs in multi-group permutation tests. Extending the topological inference methods, we develop a topological clustering framework based on HK expansion of PDs. This flexible framework allows incorporation of Euclidean covariates into topological clustering, as well as an automated data-driven selection procedure for identifying the optimal number of topological clusters and most significant covariates associated with them. We demonstrate our method's effectiveness in cluster detection with varying degrees of topological dissimilarity through simulations of signals and point clouds in comparison to state-of-the-art functional and topological clustering methods, as well as applications to subtyping and treatment outcome exploration in post-stroke aphasia.

Keywords: Topological Data Analysis; Topological Clustering ; Brain Network ; Aphasia Subtyping

*Equal contribution as first author

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

1 Introduction

Brain network modeling using electrophysiological and neuroimaging tools leverages the inherent graph structure of brain connectivity and has traditionally relied on single-scale covariance estimation [13] or graph-theoretic methods [24, 21]. While effective, these approaches fall short on capturing the complexity of brain networks, prompting a shift toward multi-scale models [2]. Topological data analysis (TDA) based on persistence homology has emerged as an effective approach for robustly identifying topological features across scales [9]. It avoids issues of arbitrary thresholding [8, 12] by tracking topological changes through filtrations, with the birth and death of homological structures summarized via persistence descriptors such as barcodes or persistence diagrams. Recent statistical and learning frameworks using graph filtrations have enabled scalable topological clustering—especially through 0- and 1-dimensional Betti functions [23, 6]. This study introduces a novel distance-based topological clustering framework using a heat-kernel representation of persistence diagrams derived from Rips filtrations, alongside algorithms for optimal cluster and covariate selection. Demonstrating greater robustness and sensitivity than existing methods, the framework is validated through simulations and real-world application to subtyping and treatment outcome prediction in post-stroke aphasia.

2 Methods

2.1 Brain network filtration and persistence descriptors

Brain networks are typically modeled as a weighted graph, with the edge weights given by a similarity measure between the measurements on the nodes of the network [1, 3]. Suppose we have a network represented by the weighted graph $G = (V, w)$ with the node set $V = \{1, \ldots, p\}$ and unique positive undirected edge weights $w = (w_{ij})$ constructed from a similarity measure such as the absolute value of the Pearson's correlation between the blood oxygen level dependent (BOLD) signals of the i-th and j-th region of interest (ROI). We define the binary network $G_\epsilon = (V, w_\epsilon)$ as a subgraph of G consisting of the node set V and the binary edge weights w_ϵ defined by

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

meaning that we connect nodes i a j with an edge when the edge weight w_{ij} is under the threshold ϵ. As we increase ϵ, which we call the *filtration value*, more edges are included in the binary network G_ϵ and so the size of the edge set increases. Since edges connected in the network do not get disconnected again, we observe a sequence of nested subgraphs $G_{\epsilon_0} \subset G_{\epsilon_1} \subset G_{\epsilon_2} \subset \cdots$, for any $\epsilon_0 \leq \epsilon_1 \leq \epsilon_2 \leq \cdots$. This sequence of nested subgraphs make up a *Rips filtration* where two nodes with a weight w_{ij} smaller than ϵ are connected, and the birth and death of *homological structures* in the form of cycles in different dimensions are tracked through the filtration [16, 17]. Here we focus on the 1-cycles (polygons formed with more than three edges in a network). We pair the birth and death times of the cycles as the coordinates of scatter points on a planar graph $\{(a_i, b_i)\}_{i=1}^L$ in the *persistence diagram* (PD). The persistence of cycles is measured by the drop from their corresponding points to the $y = x$ line on the PD. Long persistence indicates that the corresponding cluster or cycle is more likely to be an underlying feature in the network. Figure 1 shows a point that corresponds to a 1-cycle stands out with high persistence in the PD from the Rips filtration constructed on a 100-point point cloud sampled from a key shape with a hole.

2.2 Heat kernel representation of persistence diagram

Let $\mathcal{T}$ be the upper triangular region above $y = x$ where the scatter points $\{(a_i, b_i)\}_{i=1}^L$ are located. The heat kernel (HK) in $\mathcal{T}$ is $K_\sigma(p, q) = \sum_{r=0}^{\infty} e^{-\lambda_r \sigma} \boldsymbol{\psi}_r(p) \boldsymbol{\psi}_r(q)$ with respect to the eigenfunctions $\boldsymbol{\psi}_r$ of Laplace-Beltrami (LB) operator Δ satisfying $\Delta \boldsymbol{\psi}_r(p) = \lambda_r \boldsymbol{\psi}_r(p)$ for $p \in \mathcal{T}$. The first eigenvalue $\lambda_0 = 0$ corresponds to eigenfunction $\boldsymbol{\psi}_0 = \frac{1}{\sqrt{\mu(\mathcal{T})}}$, where $\mu(\mathcal{T})$ is the area of triangle $\mathcal{T}$ and σ is the bandwidth of the HK. Consider heat diffusion

$$\frac{\partial \mathbf{h}(\sigma, p)}{\partial \sigma} = \Delta \mathbf{h}(\sigma, p) \tag{2}$$

with the initial condition $\mathbf{h}(\sigma = 0, p) = \sum_{i=1}^L \delta_{(a_i, b_i)}(p)$, where $\delta_{(a_i, b_i)}$ is the Dirac-delta function at (a_i, b_i). The scatter points in the PD serve as the heat sources. A unique solution to (2) is thus

given by the HK expansion $\mathbf{h}(\sigma, p) = \int_{\mathcal{T}} K_\sigma(p,q)\mathbf{h}(\sigma = 0, q)\ d\mu(q) = \sum_{r=0}^{\infty} e^{-\lambda_r \sigma} f_r \boldsymbol{\psi}_r(p)$, where $f_r = \int_{\mathcal{T}} \mathbf{h}(\sigma = 0, q)\boldsymbol{\psi}_r(q)\ d\mu(q) = \sum_{i=1}^{L} \boldsymbol{\psi}_r(a_i, b_i)$ are the Fourier coefficients with respect to the the LB eigenfunctions. In practice, we include a finite number of terms for PD estimation:

$$\mathbf{h}_K(\sigma, p) = \sum_{r=0}^{R} e^{-\lambda_r \sigma} f_r \boldsymbol{\psi}_r(p), \tag{3}$$

with sufficiently large degree, e.g. $R = 1,000,000$ for convergence. As $\sigma \to 0$, we can completely recover the initial scatter points. As $\sigma \to \infty$, we are doing kernel density estimation with uniform kernel on $\mathcal{T}$. Figure 1 shows the HK-estimation of a PD with varied bandwidths σ.

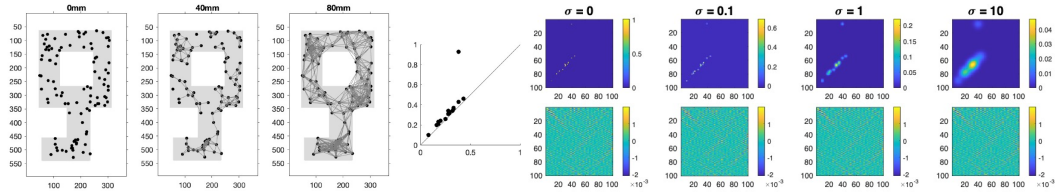


Figure 1: Left three: The evolving 1-skeleton of a 100-point point cloud sampled from a key shape with a distinct 1-cycle. Fourth: PD from the Rips filtration constructed on the 1-skeletons of the point cloud. The point in the PD that corresponds to the key cycle stands out with high persistence - much further away from the diagonal ($y = x$) line than the rest of the points. Right eight: Heat kernel (HK) smoothing of the PD from the Rips filtration through Laplace-Beltrami (LB) eigenfunctions with respect to the bandwidths $\sigma = 0$ (original PD), 0.1, 1, 10. Top: Smoothed PDs. Bottom: Corresponding Fourier coefficients with respect to the LB eigenfunctions presented in matrix form.

2.3 Topological clustering based on heat kernel representation of persistence diagrams

We start with a truncated HK expansion of a PD as $h_\sigma(p) = \sum_{r=0}^{R} e^{-\lambda_r \sigma} h_r \psi_r(p)$, where ψ_r are eigenfunctions of the Laplace-Beltrami (LB) operator, λ_r are the eigenvalues and h_r are the corresponding Fourier coefficients. Now the distance between two HK expansions $h^1_\sigma(p)$ and $h^2_\sigma(p)$ can be defined as: $d(h^1_\sigma(\cdot), h^2_\sigma(\cdot))^2 = ||h^1_\sigma(\cdot) - h^2_\sigma(\cdot)||^2_2 = \sum_{r=0}^{R} e^{-\lambda_r \sigma}(h^1_r - h^2_r)^2$. The Fourier coefficients stay the same at different diffusion scales σ. The HK smoothing can be shown to reduce the topological variability in the PD. Having defined this distance, given $\{h^i_\sigma(p)\}_{i=1}^n$, where $h^i_\sigma(p) = \sum_{k=0}^{K} e^{-\lambda_k \sigma} h^i_k \psi_k(p)$, we propose a K-means type topological clustering algorithm based on the HK representations of PD to partition them into K disjoint sets $S = \{S_1, S_2, \ldots, S_K\}$. Here, we define the functional centroid in each cluster by

$$\bar{h}_{S_j}(p) = \sum_{r=0}^{R} e^{-\lambda_r \sigma} \frac{\sum_{i \in S_j} h^i_r}{|S_j|} \psi_r(p) = \sum_{r=0}^{R} e^{-\lambda_r \sigma} \mu_{j,r} \psi_r(p) = \mathcal{M}_j,$$

where $\mu_{j,r} = \frac{\sum_{i \in S_j} h^i_r}{|S_j|}$. The cluster assignments are found to minimize the within-cluster sum of squares (WCSS)

$$W = \sum_{j=1}^{K} \sum_{h_\sigma(\cdot) \in S_j} ||h_\sigma(\cdot) - \bar{h}_{S_j}(\cdot)||^2_2. \tag{4}$$

The proposed topological clustering method is presented below in Algorithm 1. Since the objective function (WCSS) decreases after each iteration, the proposed algorithm is always guaranteed to converge, but not necessarily to the global optimum.

Selection of optimal number of clusters. The optimal number of clusters, K_{opt}, is chosen by maximizing the average silhouette score [20] over a predefined range of possible cluster counts, $[K_{min}, K_{max}]$ and repeating the procedure a large number of times, to select the one yielding the maximum average silhouette value. The silhouette score for a data point $h^i_\sigma(\cdot)$ belonging to cluster S_k (in a clustering with K_{clust} total clusters) is calculated as $S_{score}(K_{clust}) = \frac{1}{n} \sum_{i=1}^{n} s(h^i_\sigma(\cdot))$, where $s(h^i_\sigma(\cdot)) = (b(h^i_\sigma(\cdot)) - a(h^i_\sigma(\cdot)))/\max\{a(h^i_\sigma(\cdot)), b(h^i_\sigma(\cdot))\}$ (if $|S_k| > 1$; else $s(h^i_\sigma(\cdot)) = 0$). We

Algorithm 1 Topological K-means clustering for HK Representations

Given: Data $\{H_i\}_{i=1}^n$ (HK coefficients $H_i = (h_0^i, \ldots, h_R^i)$), K_{clust} clusters.
Parameters: Eigenvalues $\{\lambda_r\}$, scale σ.
Initialize centroids $\{\mathcal{M}_j^{(0)}\}_{j=1}^{K_{clust}}$, where $\mathcal{M}_j^{(t)} = (\mu_{j,0}^{(t)}, \ldots, \mu_{j,R}^{(t)})$. Let $t \leftarrow 0$.
repeat
▷ **Assignment Step**
    For each data point H_i:
        Assign H_i to cluster $S_{j^*}^{(t+1)}$, where
        $j^* \leftarrow \arg\min_{j \in \{1,\ldots,K_{clust}\}} \sum_{r=0}^{R} e^{-\lambda_r \sigma} (h_r^i - \mu_{j,r}^{(t)})^2$.
▷ **Update Step**
    For each cluster $j \in \{1, \ldots, K_{clust}\}$:
        Let $S_j^{(t+1)} = \{H_i \mid H_i \text{ assigned to cluster } j\}$.
    if $S_j^{(t+1)}$ is not empty **then**
            For $r = 0, \ldots, R$: $\mu_{j,r}^{(t+1)} \leftarrow \frac{1}{|S_j^{(t+1)}|} \sum_{H_i \in S_j^{(t+1)}} h_r^i$.
    else
            $\mathcal{M}_j^{(t+1)} \leftarrow \mathcal{M}_j^{(t)}$ ▷ Or re-initialize
    end if
    $t \leftarrow t + 1$.
until centroids converge or max iterations reached.
Return Clusters $\{S_j^{(t)}\}$ and Centroids $\{\mathcal{M}_j^{(t)}\}$.

denote $a(h_\sigma^i(\cdot))) = \frac{1}{|S_k|-1} \sum_{h_\sigma^j(\cdot)) \in S_k, j \neq i} d(h_\sigma^i(\cdot)), h_\sigma^j(\cdot)))$ to be the mean intra-cluster distance (if $|S_k| > 1$; else $a(h_\sigma^i(\cdot))) = 0$) and $b(h_\sigma^i(\cdot))) = \min_{m \neq k} \left\{ \frac{1}{|S_m|} \sum_{h_\sigma^j(\cdot)) \in S_m} d(h_\sigma^i(\cdot)), h_\sigma^j(\cdot))) \right\}$ to be the mean nearest-cluster distance. The whole procedure is repeated a large number (B) of times (with $K_{clust} \in [K_{min}, K_{max}]$), and K_{opt} is chosen as the value that achieves the highest silhouette score among the B repetitions.

Covariate selection. We extend the proposed topological clustering algorithm for the selection of influential covariates for the cluster structure, while discarding the ones that might mask such structure by adapting the VS-KM algorithm [5] based on the adjusted Rand index (ARI). Briefly, the algorithm resembles a forward selection process based on the ARI. Given the current set of included variables and the partition, a new variable is added, if the ARI (capturing agreement) between the old cluster and a new cluster using only the new variable is higher than a prespecified threshold, which suggests adding that variable would not mask the existing clustering structure. The topological clustering with the covariates proceeds in a similar fashion as in Algorithm 1 for a fixed K_{clust}, using the features $\{\sqrt{e^{-\lambda_r \sigma}}(h_r^i)\}_{r=1}^R$ and the selected covariates to define a modified distance $d(\{h_\sigma^1(\cdot), \mathbf{X}_1\}, \{h_\sigma^2(\cdot), \mathbf{X}_2\})$ based on (6). The detailed algorithm is presented as algorithm S1 in the supplementary material.

3 Simulations

Simulation studies were conducted to assess the performance of the methods. In each of the studies below, we generated multiple groups of signals or point clouds with underlying topological similarity or dissimilarity to be clustered with our proposed algorithm and state-of-the-art methods for comparison. Our simulation settings and assessment criteria are topological in nature and so may appear somewhat counterintuitive. We will explain them in detail under each setting.

3.1 Simulation studies based on signals

For the signal-based simulations, we compared the performance of the proposed topological clustering method with three state-of-the-art clustering methods from functional data analysis (FDA): 1) funFEM [4], a mixture model-based method, 2) FADPclust [19], a method based on an adaptive density peak

detection technique, and 3) fdakmeans [22], a K-means clustering algorithm for functional data. Signals were fed as functional data input to these methods, whereas HK coefficients were computed from PDs extracted from sublevel filtrations on signals [27] and fed as input into the proposed topological clustering algorithm. Note that the Rips filtration cannot be directly constructed on signals in its 1-dim. form. Two popular approaches to work around this are 1) a sublevel-set filtration that tracks time segments born at troughs and die at peaks of a signal as a vertical threshold dynamically filters through the amplitude of the signal [27]; 2) a Rips filtration on point clouds converted from the signals through a Taken's embedding [29]. Our method applies to the PDs generated from both. We opted for the former to simplify computation.

Study 1. Topological similarity. In each of 100 simulations, we generated four groups of frequency-scaled signals at time interval $0 \leq x_1 \leq \cdots \leq x_{1000} \leq 2\pi$: $y(x_i) = x_i \sin(\omega x_i)$, where ω varied between 1) $\omega = 4$, 2) $\omega = 6$, 3) $\omega = 8$ and 4) $\omega = 10$. For each ω, we simulated 50 noisy copies of the signals by adding independent Gaussian noise $\epsilon \sim N(0, 2^2)$ to $y(x_i)$. Examples of simulated signals for different groups are shown in Figure 1 of the Supplement. Rationale for why frequency-scaled signals up to a certain extent share underlying topological similarity is explained in detail in [27]. The range of optimal number of clusters was 2 to 6 for all methods under comparison. The results of the topological clustering approach in comparison with the FDA methods are summarized in Table 1. Note that the signals are supposed to be clustered as one group due to their underlying topological similarity. But we can only obtain two or more clusters through the clustering algorithms. So we applied the "next best" criterion of the optimal number of clusters being identified as two (random assignment of signals). An algorithm with the average accuracy closest to 0.5 and the percentage in all simulations of identifying two optimal clusters being closest to 100% were considered the best performer. Our method stood out with an average accuracy closest to 0.5, in comparison with the other methods, and by identifying the optimal number of clusters being two in 98% of 100 simulations, a much higher percentage than the other methods.

Study 2. Topological dissimilarity. In each of 100 simulations, we generated four groups of signals at time interval $0 \leq x_1 \leq \cdots \leq x_{1000} \leq 2\pi$:

$$y_j(x_i) = \begin{cases} x_i \sin(8x_i) - 20, & 0.4\pi < x_i \leq 0.8\pi \text{ and } j \geq 2 \\ x_i \sin(8x_i) + 10, & 0.8\pi < x_i \leq 1.2\pi \text{ and } j \geq 3 \\ x_i \sin(8x_i) - 30, & 1.2\pi < x_i \leq 1.6\pi \text{ and } j = 4 \\ x_i \sin(8x_i), & \text{else} \end{cases}$$

for $j = 1, \ldots, 4$. We simulated 50 noisy copies of the signals by adding independent Gaussian noise $\epsilon \sim N(0, 2^2)$ to $y_j(x_i)$ for $j = 1, \cdots, 4$. Examples of simulated signals for different groups are shown in Figure 1 of the Supplement. The range of optimal number of clusters was 2 to 6. For performance evaluation, we calculated the mean and standard deviation of accuracy and ARI, as well as the percentage of the optimal number of clusters selected in 100 simulations. Results summarized in Table 1 shows that our topological clustering method achieved the highest accuracy and ARI, as well as the highest success rate in identifying the correct number of clusters (four) compared to the FDA methods.

3.2 Simulation studies based on point clouds

In each simulation, we generated three groups of 50 (results consistent with 20) 200-point point clouds. For the first group, the 200 points in each point cloud were generated randomly from the rectangular image. For the second and third group, the 200 points in each point cloud were generated randomly with a percentage of 95% of points from the shape of a key and four cycles respectively and the rest 5% from the white space. Examples of simulated point clouds for different groups are shown in Figure 1 of the Supplement. Rips and graph filtrations were constructed on the pairwise Euclidean distance between points of each point cloud. Rips PDs from all point clouds were then fed into the proposed topological clustering algorithm, whereas Betti functions from the graph filtration were input for the two state-of-the-art topological clustering methods proposed by [6] and [23]. Results were compared with these two methods, which do not incorporate an optimal cluster number selection procedure so the number of clusters is pre-specified as three for these methods. The range of optimal

Signal Simulation							
Study 1. Topological Similarity (Frequency Scaling)							
Method	Accuracy		Percentage of Optimal Number of Clusters Selected				
			2	3	4	5	6
Our Method	0.5778 ± 0.0447		98.00%	2.00%	0.00%	0.00%	0.00%
funFEM	0.2500 ± 0.0000		0.00%	0.00%	1.00%	36.00%	63.00%
FADPclust	0.2615 ± 0.0498		0.00%	4.00%	95.00%	0.00%	1.00%
fdakmeans	0.3866 ± 0.1198		1.00%	47.00%	52.00%	0.00%	0.00%
Study 2. Topological Dissimilarity (Topological Tearing)							
Method	Accuracy	ARI	Percentage of Optimal Number of Clusters Selected				
			2	3	4	5	6
Our Method	0.9998 ± 0.0001	0.9997 ± 0.0020	0.00%	0.00%	**99.00%**	1.00%	0.00%
funFEM	0.8442 ± 0.0517	0.8685 ± 0.0435	0.00%	0.00%	**0.00%**	39.00%	61.00%
FADPclust	0.4368 ± 0.0775	0.2155 ± 0.1169	69.00%	11.00%	**3.00%**	5.00%	12.00%
fdakmeans	0.9925 ± 0.0429	0.9913 ± 0.0495	0.00%	3.00%	**97.00%**	0.00%	0.00%

Point Cloud Simulation			
	Our Method	Method [6]	Method [23]
Accuracy	0.9987 ± 0.0030	0.9160 ± 0.0286	0.8964 ± 0.1062
ARI	0.9962 ± 0.0089	0.8167 ± 0.0504	0.7746 ± 0.1624

Table 1: Summary of mean $\pm$ standard deviation of accuracy and ARI (the latter for topological dissimilarity only) of topological clustering and the other methods, and percentages of the optimal number of clusters selected by corresponding methods (for signal simulation only) in 100 simulations. The actual number of clusters is 4 in under the topological tearing setting.

number of clusters for the proposed method was 2 to 6. Results summarized in Table 1 show that the proposed topological clustering method achieved the highest accuracy and ARI. In addition, it consistently clustered the point clouds into three groups for all simulations.

4 Application

Stroke is the leading cause of severe adult disability in the United States [26]. A left-hemisphere stroke commonly leads to aphasia, a speech-language disorder traditionally studied through behavioral measures. To what extent brain networks constructed from neuroimaging tools provide insight into the disorder, particularly in subtyping and identifying factors contributing to treatment outcome, remains an active research question.

4.1 Data acquisition and preprocessing

Resting-state fMRI data were collected from 103 individuals with aphasia due to a single left-hemisphere ischemic or hemorrhagic stroke using a Siemens Prisma 3T scanner with a 20-channel head coil. All procedures were IRB-approved. Using the AAL atlas, 116 ROIs were defined and used as network nodes. Participants were scheduled for four visits over eight months: baseline, post-treatment 1 and 2, and six months post-treatment 2. Aphasia severity was assessed at baseline using the Aphasia Quotient from the Western Aphasia Battery-Revised (WAB-R) [15], along with WAB-R subscores for fluency, repetition, comprehension, and naming. Philadelphia Naming Test (PNT) scores were recorded at baseline and final visits. Resting-state functional brain networks were constructed using Pearson correlations of BOLD signals between the 116 ROIs. A Rips filtration was applied to each individual's correlation matrix, and the resulting 1-dimensional persistence diagrams (PDs) were smoothed via a heat kernel (HK) representation before being input into the methods.

4.2 Aphasia subtyping via topological clustering of functional brain networks

Traditional aphasia subtypes are based on binarized WAB-R subtest scores (fluency, comprehension, repetition), yielding eight categories such as Broca's and Wernicke's aphasia [15]. However, these labelings—derived from subjective, thresholded scores—have been widely criticized for poor clinical consistency [10, 7, 14, 11]. While behavioral clustering (e.g., K-means) has been used to refine subtypes [11], brain network–based clustering remains underexplored. Here, we apply topological clustering to baseline resting-state functional networks to uncover connectivity-driven aphasia subtypes, relating them to WAB-R profiles. Although persistent homology has been used in brain network studies [25, 6], this is the first applied to aphasia subtyping. Using 1D persistence diagrams from Rips filtrations, we repeated clustering 100 times (no covariates) and consistently identified three clusters via silhouette scores. Compared to K-means clusters on WAB-R subscores, lesion maps revealed that topological clusters were not confounded by lesion extent, unlike the baseline clusters. Only topological clusters showed distinct connectivity patterns. T-ANOVA [28] confirmed significant

topological differences across topological clusters (ratio statistics = 5.4728, $p = 0$), but not across baseline clusters ($p = 0.3051$). WAB-R subscore trends across topological clusters (see Supplement Figure 2) revealed low–medium–high medians for Repetition, Comprehension, and Naming, and a low–high–medium pattern for Fluency, suggesting nuanced behavioral profiles for the new subtypes.

4.3 Treatment outcome exploration with topological clustering of functional brain networks

We examined changes in topological clusters of resting-state functional brain connectivity in 54 participants at Visit 4 (down from 103 at Visit 1 due to dropouts) and the contributing baseline covariates. PDs from Visit 1 and Visit 4 were clustered separately with covariate selection based on demographic (age, sex, education), clinical (age at stroke, lesion volume, WAB-R subscores, AQ, PNT), graph-theoretic (degree, density, efficiency, modularity, clustering coefficient), and topological features (birth-death count, mean persistence, persistence entropy). Across six non-zero bandwidth settings (detail in Supplement), the optimal number of clusters was consistently two, indicating a strong but dynamic binary structure in brain connectivity clusters. Membership changes between visits suggest treatment-related reorganization. At Visit 4, both within- and between-cluster distances increased, with within-cluster distance showing stronger significance ($p = 0.0001$ vs. $p = 0.0537$) and between-cluster differences in connectivity maps also increasing over time (Figure 3 in the Supplement). At baseline, selected covariates included education, average degree, density, and PNT scores. At Visit 4, more heterogeneous variables were selected, including sex, fluency, comprehension, naming, baseline PNT, and topological features. These results highlight the evolving nature of post-stroke brain networks and may inform treatment assessment beyond behavioral scores.

5 Limitations and Solutions

In this study, we explored the longitudinal effect of treatment on aphasic brain networks with clustering applied at two time points separately. Our framework is not yet able to track individual changes, but we can incorporate random effects into the clustering framework in a follow-up study to achieve that. Also, we currently need to implement the selection of optimal number of clusters and covariates associated with them separately. But we can develop a model-based clustering framework as in [18] with an EM algorithm and BIC-type criterion to simultaneously select the number of clusters and scalar variables/covariates. Other important issues like automated bandwidth selection and scalability with respect to sample size (not network size, as the input to our algorithms are HK-smoothed PDs) will also be addressed in future studies. Options include generalized cross validation for bandwidth selection and identifying properties like monotonicity to ensure scalability.

Acknowledgments Funding support: NIHP50DC014664 (PI: Fridriksson, Project PI: Den Ouden), NIH R01DC017162 and R01DC01716202S1 (PI: RHD).

References

[1] D.S. Bassett. Adaptive reconfiguration of fractal small-world human brain functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, 203:19518–23, 2006.

[2] R.F. Betzel and D.S. Bassett. Multi-scale brain networks. *NeuroImage*, 160:73 – 83, 2017.

[3] J. Bien and R. Tibshirani. Hierarchical clustering with prototypes via minimax linkage. *Journal of the American Statistical Association*, 106:1075 – 1084, 2011.

[4] Charles Bouveyron, Etienne Come, and Julien Jacques. The discriminative functional mixture model for a comparative analysis of bike sharing systems. 2015.

[5] Michael J Brusco and J Dennis Cradit. A variable-selection heuristic for k-means clustering. *Psychometrika*, 66(2):249–270, 2001.

[6] M.K. Chung, C.G. Ramos, F.B. de Paiva, J. Mathis, V. Prabharakaren, V.A. Nair, E. Meyerand, B.P. Hermann, J.R. Binder, and A.F. Struck. Unified topological inference for brain networks in temporal lobe epilepsy using the wasserstein distance. *arXiv: 2302.06673*, 2023.

[7] M. A. Crary, R. T. Wertz, and J. L. Deal. Classifying aphasias: Cluster analysis of western aphasia battery and boston diagnostic aphasia examination results. *Aphasiology*, 6:29–36, 1992.

[8] M. Drakesmith, K. Caeyenberghs, A. Dutt, G. Lewis, A.S. David, and D.K. Jones. Overcoming the effects of false positives and threshold bias in graph theoretical analyses of neuroimaging data. *NeuroImage*, 118:313–333, 2015.

[9] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete Computational Geometry*, pages 511 – 533, 2002.

[10] J. M. Ferro and A. Kertesz. Comparative classification of aphasic disorders. *Journal of clinical and experimental Neuropsychology*, 9:365–375, 1987.

[11] Davida Fromm, Joel Greenhouse, Mitchell Pudil, Yichun Shi, and Brian MacWhinney. Enhancing the classification of aphasia: a statistical analysis using connected speech. *Aphasiology*, 36:1492–1519, 2022.

[12] K.A. Garrison, D. Scheinost, E.S. Finn, X. Shen, and R. T. Constable. The (in)stability of functional brain network measures across thresholds. *NeuroImage*, 118:651 – 661, 2015.

[13] S. Huang, J. Li, L. Sun, J. Ye, A. Fleisher, T. Wu, K. Chen, and E. Reiman. Learning brain connectivity of alzheimer's disease by sparse inverse covariance estimation. *NeuroImage*, 50:935 – 949, 2010.

[14] A. A. John, M. Javali, R. Mahale, A. Mehta, P. T. Acharya, and R. Srinivasa. Clinical impression and western aphasia battery classification of aphasia in acute ischemic stroke: Is there a discrepancy? *Journal of Neurosciences in Rural Practice*, 8:074–078, 2017.

[15] A. Kertesz. The western aphasia battery - revised. 2007.

[16] 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. *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 14:302 – 309, 2011.

[17] H. Lee, M.K. Chung, H. Kang, and D.S. Lee. Hole detection in metabolic connectivity of alzheimer's disease using k-laplacian. *International Conference on Medical Imaging Computing and Computer-Assisted Intervention (MICCAI). Lecture Notes in Computer Science (LNCS).*, 8675:297 – 304, 2014.

[18] Wei Pan and Xiaotong Shen. Penalized model-based clustering with application to variable selection. *Journal of machine learning research*, 8(5), 2007.

[19] Rui Ren, Kuangnan Fang, Qingzhao Zhang, and Xiaofeng Wang. Multivariate functional data clustering using adaptive density peak detection. *Statistics in Medicine*, 42(10):1565–1582, 2023.

[20] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[21] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*, 52:1059 – 1069, 2010.

[22] Laura M Sangalli, Piercesare Secchi, Simone Vantini, and Valeria Vitelli. K-mean alignment for curve clustering. *Computational Statistics & Data Analysis*, 54(5):1219–1233, 2010.

[23] Tananun Songdechakraiwut, Bryan M Krause, Matthew I Banks, Kirill V Nourski, and Barry D Van Veen. Fast topological clustering with wasserstein distance. In *International Conference on Learning Representations*, 2022.

[24] O. Sporns. Network analysis, complexity, and brain function. *Complexity*, 8:56 – 60, 2002.

[25] B. Stolz, H. Harrington, and M. Porter. Persistent homology of time-dependent functional networks constructed from coupled time series. *Chaos*, 27, 2017.

[26] Connie W Tsao, Aaron W Aday, Zaid I Almarzooq, Alvaro Alonso, Andrea Z Beaton, Marcio S Bittencourt, Amelia K Boehme, Alfred E Buxton, April P Carson, Yvonne Commodore-Mensah, et al. Heart disease and stroke statistics—2022 update: a report from the american heart association. *Circulation*, 145:e153–e639, 2022.

[27] Y. Wang, R. Behroozmand, L. Phillip Johnson, L. Bonilha, and J. Fridriksson. Topological signal processing and inference of event-related potential response. *Journal of Neuroscience Methods*, 363:109324, 2021.

[28] Yuan Wang, Jian Yin, and Rutvik H Desai. Topological inference on brain networks across subtypes of post-stroke aphasia. *arXiv*, pages arXiv–2311, 2023.

[29] Jian Yin and Yuan Wang. Topological inference and correlation of signals with application to electroencephalography in epilepsy. *Biomedical Signal Processing and Control*, 80:104396, 2023.

Supplement

Methods: Algorithm for topological clustering with variable selection

Algorithm. The algorithm for topological clustering with covariate selection is as follows.

Algorithm S1 Topological VS-KM Clustering with HK Representation

Given: Data $\{H_i\}_{i=1}^n$ with $H_i = (h_0^i, \ldots, h_R^i)$, covariates $\mathbf{X} \in \mathbb{R}^{n \times p}$
Parameters: $\{\lambda_r\}$, scale σ, $G_{\min}$, G_{fac}, ϵ_c, ϵ_w, T_{outer}
One-hot encode covariates $\mathbf{X}$ into $\tilde{\mathbf{X}} \in \mathbb{R}^{n \times C}$
Initialize weight matrix $W^{(0)} \leftarrow \mathbf{1}_{(R+1)\times C}$
for $t = 1$ to T_{outer} **do**
  for each $K \in \mathcal{K}$ **do**
    Run **Algorithm S1** using weighted distance:

$$d(H_i, \mathcal{M}_j) = \sum_{r=0}^{R} \sum_{c=1}^{C} W_{r,c}^{(t-1)} \cdot e^{-\lambda_r \sigma} (h_{r,c}^i - \mu_{j,r,c})^2$$

    Compute clustering $\{S_j^{(t)}\}$ and centroids $\{\mathcal{M}_j^{(t)}\}$ for current K
    Evaluate silhouette index $\mathbf{sil}_K$
  end for
  $K_{\text{opt}} \leftarrow \arg\max_K \mathbf{sil}_K$
  Re-run **Algorithm S1** with K_{opt} and weights $W^{(t-1)}$ to obtain $\{S_j^{(t)}\}$ and $\{\mathcal{M}_j^{(t)}\}$
  for $c = 1$ to C **do**
    Compute $\text{ARI}_c = \text{AdjustedRandIndex}(\{S_j^{(t)}\}, \tilde{X}_{:,c})$
  end for
  Let $\mathcal{S} = \{c : \text{ARI}_c \geq G_{\min}\}$
  Initialize $\omega^{(t)} \in \mathbb{R}^C$ as all ones
  for each $c \in \mathcal{S}$ **do**

$$\omega_c^{(t)} = 1 + \left(\frac{\text{ARI}_c}{\max_{j \in \mathcal{S}} \text{ARI}_j + \epsilon} \right) (G_{\text{fac}} - 1)$$

  end for
  Construct full weight matrix:

$$W_{r,c}^{(t)} = \omega_c^{(t)}, \quad \forall r = 0, \ldots, R,\ c = 1, \ldots, C$$

  Compute average weight change:

$$\Delta_t = \frac{1}{(R+1)C} \sum_{r=0}^{R} \sum_{c=1}^{C} \left| W_{r,c}^{(t)} - W_{r,c}^{(t-1)} \right|$$

  if $\Delta_t < \epsilon_w$ **then**
    break
  end if
end for
return Clustering $\{S_j^{(t)}\}$, centroids $\{\mathcal{M}_j^{(t)}\}$, optimal K_{opt}, final weights $W^{(t)}$, and ARI scores $\{\text{ARI}_c\}$

Simulations: Figure, parameter settings, stability analysis, and computational time

Examples of simulated signals and point clouds are shown in Fig 1.

Parameter settings. Table 1 summarizes the parameters used in the simulation section.

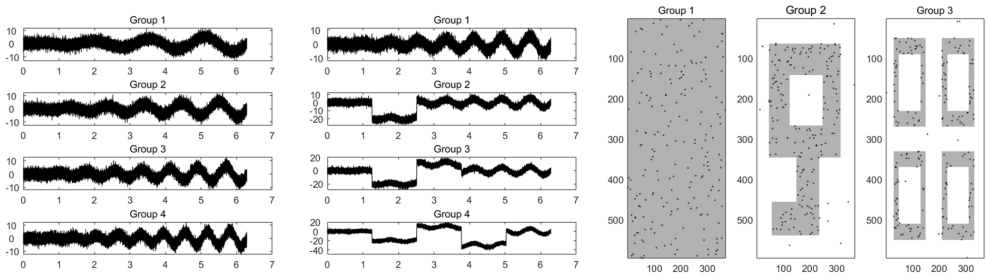


Figure 1: Examples of simulated signals after frequency scaling (left) and topological tearing (middle), and examples of simulated point clouds (right).

Parameter	Description	Value
σ	Bandwidth for heat kernel estimation	5
R	Degree for heat kernel estimation	10000
T_{max}	Max number of k-means iterations	100
ϵ	Convergence threshold for centroid updates	10^{-5}
B	Number of repetitions of algorithm	10
$\mathcal{K} = [K_{min}, K_{max}]$	Range of optimal numbers of clusters for algorithm	$[2, 6]$

Table 1: Parameter settings for simulations.

Stability analysis. We verified the stability of our method by increasing the number of signals/point clouds from 50 to 100 per group. The results summarized in Table 2 are aligned with those obtained with 50 signals/point clouds per group, consistently indicating the superior performance of our approach.

Signal Simulation							
Study 1. Topological Similarity (Frequency Scaling)							
Method	Accuracy		Percentage of Optimal Number of Clusters Selected				
			2	3	4	5	6
Our Method	0.5542 ± 0.0337		100.00%	0.00%	0.00%	0.00%	0.00%
funFEM	0.2500 ± 0.0000		0.00%	0.00%	0.00%	13.00%	87.00%
FADPclust	0.2503 ± 0.0035		0.00%	0.00%	100.00%	0.00%	0.00%
fdakmeans	0.3793 ± 0.1230		1.00%	44.00%	55.00%	0.00%	0.00%
Study 2. Topological Dissimilarity (Topological Tearing)							
Method	Accuracy	ARI	Percentage of Optimal Number of Clusters Selected				
			2	3	4	5	6
Our Method	0.9999 ± 0.0004	0.9997 ± 0.0013	0.00%	0.00%	**100.00%**	0.00%	0.00%
funFEM	0.8250 ± 0.0594	0.8548 ± 0.0503	0.00%	0.00%	**1.00%**	17.00%	82.00%
FADPclust	0.4254 ± 0.0785	0.2059 ± 0.1151	61.00%	14.00%	**6.00%**	7.00%	12.00%
fdakmeans	0.9825 ± 0.0641	0.9799 ± 0.0737	0.00%	7.00%	**93.00%**	0.00%	0.00%

Point Cloud Simulation			
	Our Method	Method [1]	Method [2]
Accuracy	0.9986 ± 0.0023	0.9164 ± 0.0263	0.8940 ± 0.1070
ARI	0.9957 ± 0.0068	0.8167 ± 0.0373	0.7718 ± 0.1587

Table 2: Summary of mean $\pm$ standard deviation of accuracy and ARI (the latter for topological dissimilarity only) of topological clustering and the other methods, and percentages of the optimal number of clusters selected by corresponding methods (for signal simulation only) in 100 simulations. The actual number of clusters is 4 in under the topological tearing setting.

Computational time. We summarized the computational time of the proposed topological clustering method in comparison with two other state-of-the-art topological clustering methods in Table 3. As we mentioned in Limitations, scalability was not the focal point of the paper but we will explore options to reduce computational complexity in a follow-up study.

Signal Simulation		
Method	50 signals per group	100 signals per group
Our Method	387.46 ± 8.38	1308.10 ± 18.06
funFEM	4.23 ± 1.63	8.24 ± 2.34
FADPclust	3.65 ± 0.19	11.58 ± 0.87
fdakmeans	12714.94 ± 431.53	40446.76 ± 367.43
Point Cloud Simulation		
Method	50 point clouds per group	100 point clouds per group
Our Method	131.19 ± 2.31	492.81 ± 23.21
Method [1]	9.26 ± 1.17	27.68 ± 3.18
Method [2]	70.27 ± 44.90	145.01 ± 86.25

Table 3: Summary of mean$\pm$standard deviation of time (in seconds) in 100 simulations.

Computational Resources. All simulation analyses were conducted in MATLAB R2023a on a local Windows machine running Microsoft Windows 10 with an Intel(R) Xeon(R) Gold 5218 CPU (32 processors, 64 logical processors, 2.30 GHz) and 64 GB RAM.

Application: Data processing, results, parameter settings, stability analysis, and computation

Processed resting-state functional magnetic resonance imaging (rs-fMRI) data used in the application have been made available on Open Neuro. As half of the authors on this paper are listed on the open-access data page on Open Neuro, we did not include a link here for review to avoid violation of anonymity. We included here the data processing details and the parameter settings that we used for the second part of the application (treatment outcome exploration) on the real data, as well as stability analysis results and computational times.

Data processing. The following imaging parameters of images were used: a multiband sequence (x2) with a 216×216 mm field of view, a 90×90 matrix size, and a 72-degree flip angle, 50 axial slices (2 mm thick with 20% gap yielding 2.4 mm between slice centers), repetition time TR =1650 ms, TE=35 ms, GRAPPA=2, 44 reference lines, interleaved ascending slice order. During the scanning process, the participants were instructed to stay still with eyes closed. A total of 370 volumes were acquired. The preprocessing procedures of the rs-fMRI data include motion correction, brain extraction and time correction using a novel method developed for stroke patients [4]. The Realign and Unwarp procedure in SPM12 with default settings was used for motion correction. Brain extraction was then performed using the SPM12 script pm_brain_mask with default settings. Slice time correction was also done using SPM12. The mean fMRI volume for each participant was then aligned to the corresponding T2-weighted image to compute the spatial transformation between the data and the lesion mask. The fMRI data were then spatially smoothed with a Gaussian kernel with FWHM= 6 mm. To eliminate artifacts driven by lesions, a pipeline proposed by [4] was applied on the the rs-fMRI. The FSL MELODIC package was used to decompose the data into independent components (ICs) and to compute the Z-scored spatial maps for the ICs. The spatial maps were thresholded at $p < 0.05$ and compared with the lesion mask for the participant. The Jaccard index, computed as the ratio between the numbers of voxels in the intersection and union, was used to quantify the amount of spatial overlap between the lesion mask and thresholded IC maps, both of which were binary. ICs corresponding to Jaccard index greater than 5% were deemed significantly overlapping with the lesion mask and then regressed out of the fMRI data using the fsl_regfilt script from the FSL package.

Results in more detail: Aphasia subtyping. Using PDs constructed on resting-state functional connectivity matrices from Visit 1, we repeated the topological clustering algorithm 100 times, without covariate selection, and checked for cluster consistency across repetitions through correlation plots. Three clusters had the overall best fit via Sillouette= indices. The overall lesion map and average absolute connectivity of three baseline clusters obtained through standard k-means and three topological clusters obtained through the proposed clustering algorithm with bandwidth 0 are shown in Figure 2 (left). The lesion map was created by augmenting stroke lesion damage in the brain of all subjects within each cluster. Note that the three baseline k-means clusters appear to be confounded by the overall lesion extent of the subjects as they show distinctly different lesion

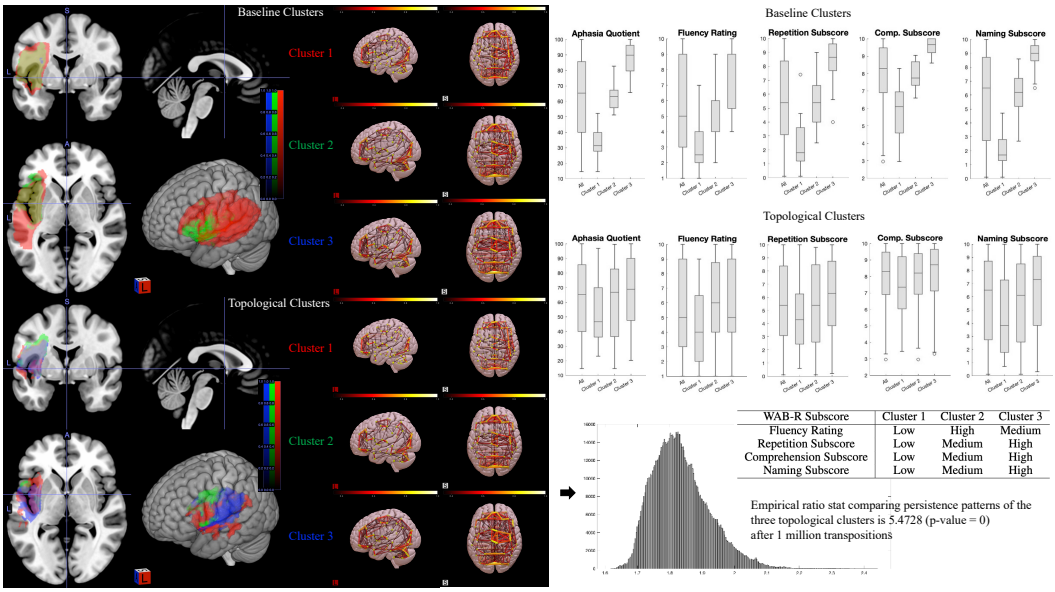
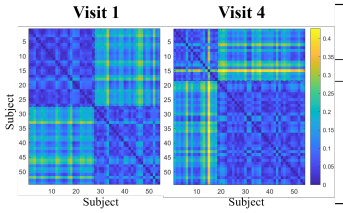


Figure 2: Left: The lesion and average connectivity maps of three topological and baseline k-means clusters. Top Right: Box plots of subscores for all participants and those in each of the topological and k-means clusters. Bottom Right: Empirical distribution of the ratio statistic based on L_2-distances of HK-smoothed 1-dimensional PDs within and between the three clusters over 1 million transpositions, and pattern of median and interquartile range of WAB-R subscores across three topological clusters/subtypes.

extent (Cluster 1 > Cluster 2 > Cluster 3). This is confirmed by the AQ and subscore distributions summarized in Figure 2 (top right), where the AQ score is known to positively correlate with lesion extent and the subscore distributions show a distinct monotone pattern consistent with that of AQ across clusters. On the other hand, the topological clusters do not appear to be confounded by lesion extent as the lesion extent do not vary significantly across the clusters and the subscore distributions do not follow a specific trend with reference to the AQ score. As of the average connectivity, we see different connectivity patterns in the three topological clusters, whereas the baseline clusters show similar connectivity pattens. To confirm that the topological clusters did capture significant statistical difference in brain networks, we also compared the brain networks across different clusters through the permutation-based topological ANOVA (T-ANOVA) test, proposed by [3], on their HK-smoothed PDs. Figure 2 (bottom right) shows the empirical distribution of the ratio statistic based on L_2-distances of HK-smoothed 1-dimensional PDs within and between the three clusters over 1 million transpositions. The observed value of the ratio statistic was 5.4728, yielding a p-value of 0 and the conclusion of significant topological difference between the 1-cycle presence in the three clusters of brain networks.

Results in more detail: Treatment response. We also explored the changes in topological clusters of resting-state functional brain connectivity of 54 participants at Visit 4 (down from original 103 at Visit 1 due to dropouts across the course of study) and significant contributing factors to these clusters. The PDs of resting-state functional brain networks at Visit 1 (baseline visit) and Visit 4 (6 months after treatment visit 2) underwent topological clustering separately, with covariate selection w.r.t. variables from the baseline visit: demographic variables - age, sex, and years of education; clinical variables - age at stroke, lesion volume, WAB-R subscores (fluency, repetition, comprehension, naming), aphasia quotient, average PNT score; graph-theoretic variables - average degree and density, efficiency, modularity, and clustering coefficient; topological variables - number of birth-death pairs and mean persistence in PD, and persistence entropy. We analyzed changes in covariate selection patterns, cluster membership, and connectivity strength. Across six parameter settings with non-zero bandwidths (Table 4), the optimal number of clusters selected was consistently two, with agreement in most cases — indicating a strong underlying binary grouping in the stroke patients. However, notable membership changes between visits show that this binary structure is not static, but rather reflects dynamic changes in brain connectivity and behavior over the course of treatment. Within-



Adjusted within-cluster sum of distances		Adjusted between-cluster sum of distances	
Visit 1	Visit 4	Visit 1	Visit 4
0.0690	0.0781	0.1642	0.1917
0.0373	0.0429	0.0907	0.1048
0.0690	0.0753	0.1642	0.1750
0.0373	0.0430	0.0907	0.0907
0.0690	0.0753	0.1642	0.1750
0.0373	0.0430	0.0907	0.0907
p-value = 0.0001		p-value = 0.0537	

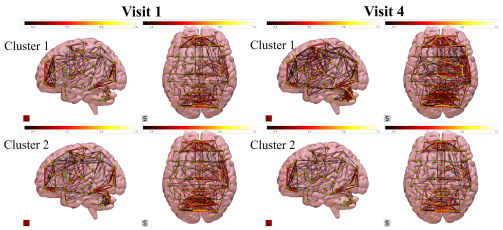


Figure 3: Left: Within- and between-cluster sums of distances (adjusted for cluster sizes) across six parameter settings and paired t-test p-values. Right: Average connectivity maps for each cluster at baseline and fourth visit.

and between-cluster distances after adjusting for cluster sizes both increased, with the former (p = 0.0001) being more significantly than the latter (p = 0.0537) at the fourth visit (Figure 3 left). Figure 3 (right) confirms the increase in the between-cluster distance change as average connectivity maps in the two clusters at Visit 4 does appear to be more different than Visit 1. Across the same parameter settings, our algorithm applied at baseline visit consistently selected education level, average degree, density and PNT Baseline scores, reflecting preposition, whereas results at the fourth visit showed more heterogeneity, selecting a broader set of variables, including sex, fluency rating, comprehension and naming subscores, average PNT at baseline, and number of birth-death pairs in PD. Our findings may potentially inform clinicians when assessing treatment outcome in post-stroke aphasia beyond behavioral scores and with regard to various factors.

Parameter settings. We applied the heat kernel estimation of persistence diagrams (PDs) and proposed topological clustering method under six parameter settings at Visit 1 and Visit 4. For each parameter setting, the proposed topological clustering method was applied to the same set of PDs 100 times. Table 4 and 5 respectively summarize the six parameter settings and the other input parameters for heat kernel estimation of PDs and proposed topological clustering method under each parameter setting.

Settings	σ	G_{min}	G_{fac}
1	5	0.01	0.9
2	10	0.005	0.9
3	5	0.01	0.9
4	10	0.005	0.9
5	5	0.01	1.5
6	10	0.005	1.5

* σ: bandwidth for the heat kernel estimation. * G_{min} and G_{fac}: thresholding factors controlling sparsity in covariate selection. Covariates with corresponding $\text{ARI}_c \geq G_{\text{min}}$ and $\text{ARI}_c \leq G_{\text{fac}}$ are selected.

Table 4: Six parameter settings of heat kernel estimation of persistence diagrams (PDs) and proposed topological clustering method in second part of the application section.

Parameter	Description	Value
R	Degree of heat kernel estimation of PDs	10000
T_{max}	Max number of k-means iterations with selected covariates	100
T_{outer}	Max number of VS-KM iterations	100
$\epsilon_{\text{centroid}}$	Convergence threshold for centroid updates	10^{-4}
ϵ_{weight}	Convergence threshold for weight updates	10^{-4}
B	Number of repetitions of algorithm	100
$\mathcal{K} = [K_{min}, K_{max}]$	Range of optimal numbers of clusters for algorithm	$\mathcal{K} = [2, 7]$

Table 5: The other input parameters for heat kernel estimation of PDs and proposed topological clustering under each parameter setting in Table 4.

Stability analysis. Table 6a summarizes the distribution of the optimal number of clusters (K) chosen across 100 replicates at Visit 1 and Visit 4. For Visit 1, all settings consistently select $K = 2$, suggesting highly stable clustering structure. In contrast, Visit 4 exhibits variability, with Settings 2, 4, and 6 producing a mixture of $K = 2$, 3, and 4. Covariate selection results are summarized in Table 6b. Visit 1 had perfect consistency of selected covariates across four settings, whereas Visit 4 had more varied subsets of covariates being selected across the settings.

Visit 1				Visit 4			
Parameter Setting	$K = 2$	$K = 3$	$K = 4$	Parameter Setting	$K = 2$	$K = 3$	$K = 4$
1	100	0	0	1	100	0	0
2	100	0	0	2	75	23	2
3	100	0	0	3	100	0	0
4	100	0	0	4	74	25	1
5	100	0	0	5	100	0	0
6	100	0	0	6	70	29	1

(a) The optimal number of clusters (K) over 100 runs for each parameter setting at Visit 1 and Visit 4.

Data of Visit	Covariates	Parameter Setting 1	Parameter Setting 2	Parameter Setting 3	Parameter Setting 4	Parameter Setting 5	Parameter Setting 6
Visit 1	Education level	100	100	100	100	100	100
	Density	–	–	100	100	100	100
	Average degree	–	–	100	100	100	100
	PNT baseline score	–	–	100	100	100	100
	# Runs with selection	100	100	100	100	100	100
Visit 4	Education level	0	18	0	25	0	29
	Density	0	18	94	50	93	58
	Average degree	0	18	94	50	93	58
	PNT baseline score	0	0	0	49	0	43
	Comprehension subscore	0	28	0	51	0	43
	Fluency rating	0	16	0	22	0	26
	Number of birth-death pairs	0	2	0	51	0	43
	Sex	0	3	0	4	0	4
	Repetition subscore	0	0	0	1	0	4
	Naming subscore	0	0	0	0	0	3
	# Runs with selection	0	47	94	99	93	99

(b) Covariate selection frequencies (out of 100 runs) across visits and parameter settings.

Table 6: Clustering results at Visit 1 and Visit 4.

Internal clustering stability was assessed using two agreement metrics—ARI and Normalized Mutual Information (NMI)—computed over 100 replicate runs for each parameter setting (restricted to runs with $K = 2$) in Table 7. At Visit 1, all settings yielded perfectly stable clustering solutions (ARI = 1.00, NMI = 1.00), whereas Visit 4 showed varying degrees of stability (ARI = 0.65–0.67; NMI = 0.66–0.69), with Parameter Settings 2 and 6 showing the lowest clustering consistency at Visit 4 (ARI and NMI), likely due to their relaxed variable selection thresholds. Setting 2 led to low selection frequency and high variability in covariate subsets across runs, while Setting 6 included a larger set of covariates (Table 6b). Both factors may have introduced additional noise and inconsistency into the clustering process.

Visit	Metrics	Parameter Setting 1	Parameter Setting 2	Parameter Setting 3	Parameter Setting 4	Parameter Setting 5	Parameter Setting 6
Visit 1	Number of runs with $K = 2$	100	100	100	100	100	100
	ARI	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
	NMI	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
Visit 4	Number of runs with $K = 2$	100	75	100	74	100	70
	ARI	0.84 ± 0.16	0.65 ± 0.27	0.86 ± 0.16	0.73 ± 0.28	0.86 ± 0.15	0.67 ± 0.29
	NMI	0.81 ± 0.19	0.66 ± 0.24	0.83 ± 0.18	0.74 ± 0.26	0.83 ± 0.18	0.69 ± 0.27

Table 7: Internal clustering stability metrics (ARI and NMI: mean $\pm$ standard deviation) over the runs with $K = 2$ for each parameter setting at Visit 1 and Visit 4.

Figure 4 shows the silhouette scores over 100 runs across visits and parameter settings. Each subplot shows the silhouette scores under different values of K from individual runs, with the best-performing run highlighted—selected as the one with the highest total silhouette score across all data points.

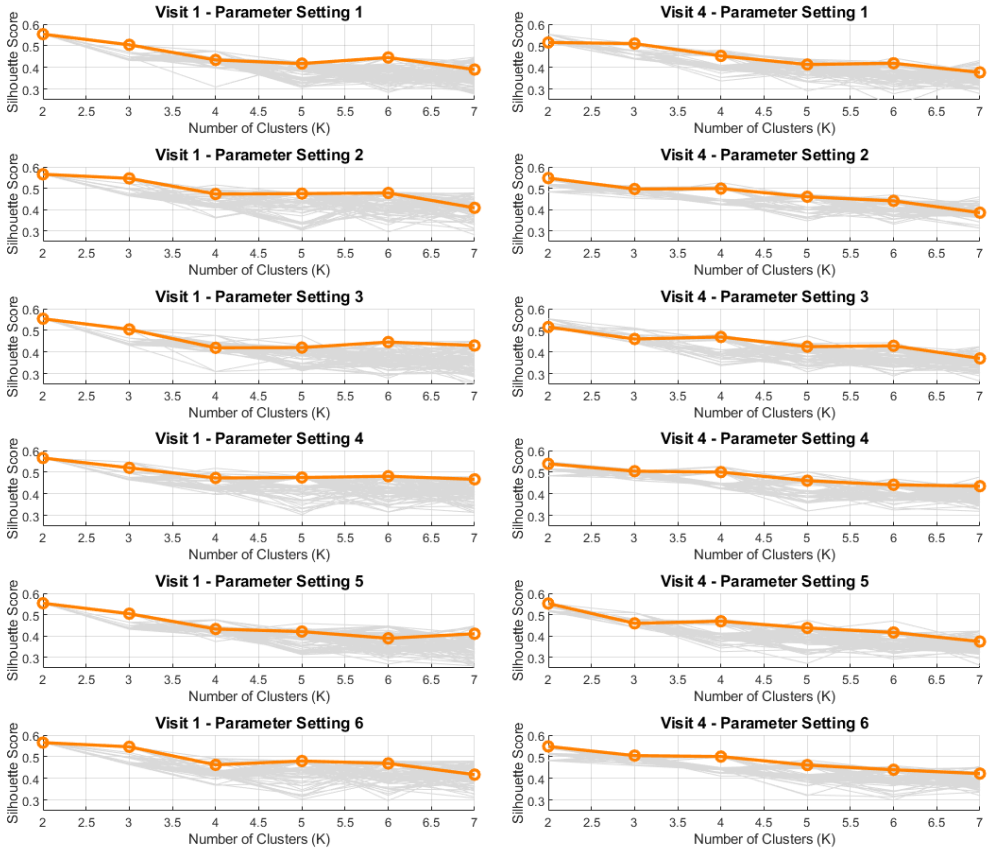


Figure 4: Silhouette scores under varying numbers of clusters (K) across 100 runs for each parameter setting at Visit 1 (left) and Visit 4 (right). Gray lines show individual runs; the highlighted curve indicates the best-performing run, selected based on the highest total silhouette scores.

Computational time. Table 8 includes the computational time averaged in 100 clustering runs with each setting for Visit 1 and Visit 4.

Parameter Setting	1	2	3	4	5	6
Visit 1	5.6397 ± 0.6862	6.5224 ± 0.5094	5.9230 ± 0.4791	6.1054 ± 0.5867	4.5677 ± 0.2991	4.8382 ± 0.4147
Visit 4	2.5554 ± 0.3581	20.7167 ± 20.8751	5.9926 ± 1.4001	21.5791 ± 20.4974	5.7110 ± 1.5227	25.5473 ± 17.9320

Table 8: Mean$\pm$ standard deviation of computational time (in seconds) over 100 runs for clustering at Visit 1 and Visit 4 under each parameter setting.

Computational Resources. All treatment outcome exploration application analyses were conducted in MATLAB R2023b on a local Windows machine running Microsoft Windows 11 Pro with an Intel(R) Core(TM) Ultra 7 165U CPU (12 cores, 14 logical processors, 1.70 GHz) and 32 GB RAM. Brain connectivity visualizations were realized through Surf Ice (6-October-2021 release).

References

[1] M.K. Chung, C.G. Ramos, F.B. de Paiva, J. Mathis, V. Prabharakaren, V.A. Nair, E. Meyerand, B.P. Hermann, J.R. Binder, and A.F. Struck. Unified topological inference for brain networks in temporal lobe epilepsy using the wasserstein distance. *arXiv: 2302.06673*, 2023.

[2] Tananun Songdechakraiwut, Bryan M Krause, Matthew I Banks, Kirill V Nourski, and Barry D Van Veen. Fast topological clustering with wasserstein distance. In *International Conference on Learning Representations*, 2022.

[3] Yuan Wang, Jian Yin, and Rutvik H Desai. Topological inference on brain networks across subtypes of post-stroke aphasia. *arXiv*, pages arXiv–2311, 2023.

[4] G. Yourganov, J. Fridriksson, B. Stark, and C. Rorden. Removal of artifacts from resting-state fmri data in stroke. *NeuroImage: Clinical*, 17:297–305, 2018.