Topological data analysis of task-based fMRI data from experiments on schizophrenia

Journal of Physics: Complexity

OPEN ACCESS

PAPER

Topological data analysis of task-based fMRI data from experiments on schizophrenia

Bernadette J Stolz[1], **Tegan Emerson**[2], **Satu Nahkuri**[3], **Mason A Porter**[4] and **Heather A Harrington**[1]

[1] Mathematical Institute, University of Oxford, Oxford, United Kingdom
[2] Pacific Northwest National Laboratory, Seattle, WA, United States of America
[3] F Hoffmann-La Roche AG, Basel, Switzerland
[4] Department of Mathematics, University of California, Los Angeles, Los Angeles, CA, United States of America

E-mail: stolz@maths.ox.ac.uk, tegan.emerson@pnnl.gov, satu.nahkuri@roche.com, mason@math.ucla.edu and harrington@maths.ox.ac.uk

RECEIVED
2 June 2020

REVISED
25 August 2020

ACCEPTED FOR PUBLICATION
2 September 2020

PUBLISHED
19 May 2021

Keywords: persistent homology, networks, fMRI, persistence landscapes, persistence images, functional brain networks

Abstract

We use methods from computational algebraic topology to study functional brain networks in which nodes represent brain regions and weighted edges encode the similarity of functional magnetic resonance imaging (fMRI) time series from each region. With these tools, which allow one to characterize topological invariants such as loops in high-dimensional data, we are able to gain understanding of low-dimensional structures in networks in a way that complements traditional approaches that are based on pairwise interactions. In the present paper, we use persistent homology to analyze networks that we construct from task-based fMRI data from schizophrenia patients, healthy controls, and healthy siblings of schizophrenia patients. We thereby explore the persistence of topological structures such as loops at different scales in these networks. We use persistence landscapes and persistence images to represent the output of our persistent-homology calculations, and we study the persistence landscapes and persistence images using *k*-means clustering and community detection. Based on our analysis of persistence landscapes, we find that the members of the sibling cohort have topological features (specifically, their one-dimensional loops) that are distinct from the other two cohorts. From the persistence images, we are able to distinguish all three subject groups and to determine the brain regions in the loops (with four or more edges) that allow us to make these distinctions.

1. Introduction

Schizophrenia is a chronic psychiatric disorder that affects more than 21 million people worldwide [1]. Up to 80% of the risk factors appear to be genetic, although it is difficult to identify the specific genes that are involved in the disease [2]. The disease usually commences in early adulthood, and symptoms range from hallucinations and avolition to cognitive deficits (such as impaired working memory) [1, 3]. It is believed that the cognitive deficits arise from compromised functional integration between neural subsystems [3–6].

There can be significant differences in the properties of the time series from imaging measurements of healthy versus schizophrenic individuals, although different studies have found seemingly contradictory results when comparing functional magnetic resonance imaging (fMRI) time series from two distinct brain regions in a schizophrenia patient and a healthy control. The majority of studies have concluded that schizophrenia patients have less-similar time series than healthy controls across different brain regions [7]. Zalesky *et al* [8] suggested that such reduced similarity may arise from an altered coupling between brain regions and local decoherence within brain regions in schizophrenia patients. However, some studies have observed that schizophrenia patients have more-similar series than healthy controls across brain regions. For a detailed discussion of these seemingly contradictory findings, see [9]. In some cases, methodological steps in fMRI analyses

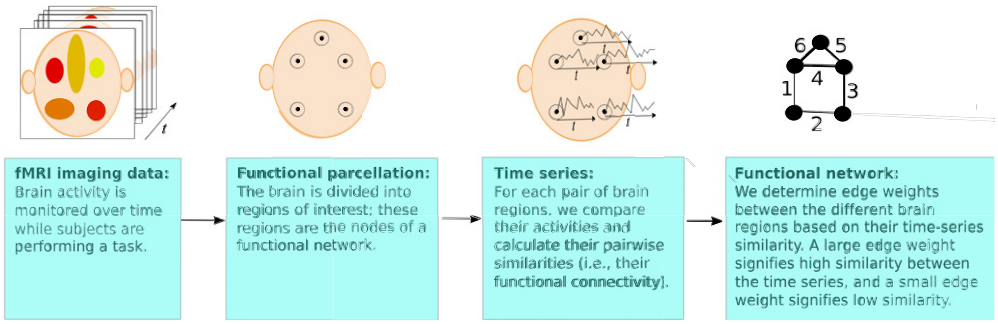


Figure 1. Pipeline to construct functional networks from imaging data (e.g., fMRI data).

seem to yield increases in these similarities, but abnormal neurodevelopment or drug treatment may play a role in increasing them in other cases [9].

One approach for studying the human brain is to construct a (possibly time-dependent) neuronal network based on experimental data and then analyze the network's structure and dynamics to gain insights into its properties [10–17]. One can form a so-called *functional network* [10, 11, 18–20], in which each node represents a brain region and one weights the edges between them based on some measure of the similarity between the nodes' fMRI time series. (Researchers also employ time series from other imaging modalities to construct functional networks.) In figure 1, we show a pipeline of how to construct a functional network from fMRI time-series data. When interpreting functional networks in fMRI studies, it is very important to consider the cautionary notes in [21].

Studies of functional networks of schizophrenia patients have revealed that such networks differ significantly from the functional networks of healthy controls [6, 7, 22–26]. For example, schizophrenia patients can have rather different community structure than controls [24, 27]. In one study, Alexander-Bloch *et al* [24] observed significant differences (due to a small number of brain regions) in the community structures in schizophrenia patients, whereas the community structures in healthy subjects appeared to be consistent with each other. Moreover, in some studies, the maximum modularity of functional networks was smaller for schizophrenia patients than for healthy controls [24, 28]. Two recent papers, Flanagan *et al* [27] and Towlson *et al* [29], compared the network structures of schizophrenia patients and healthy controls under the effects of different drugs and a placebo.

An increasingly popular approach for the analysis of functional networks is to use ideas from computational algebraic topology, as these approaches allow one to characterize topological invariants (such as connectedness and loops) in high-dimensional data [30–36]. In contrast to standard methods of network analysis [37], employing computational topology allows one to explicitly go beyond pairwise connections; this is helpful for gaining global understanding of low-dimensional structures in networks. Although one can also use frameworks such as hypergraphs [38] to study higher-order network structures (see, e.g., [39]), such a formalism does not by itself give direct information about the shape or scale of mesoscale features in networks. By contrast, *persistent homology* (PH), the most prominent approach in topological data analysis, allows one to explore the persistence of features (such as connectedness and loops) in data sets [40, 41]. PH has led to interesting insights in a variety of fields (for examples, see [42–48]), and it has been used increasingly in studies of neuronal networks, leading to several promising insights [20, 49–64].

In the present paper, we construct functional networks using fMRI data from schizophrenia patients, healthy controls, and siblings of schizophrenia patients. We create a nested sequence of networks in which we add edges, one by one, to the networks in order from the largest edge weights to the smallest. (In the unlikely case of two edges having the exact same weight, we add both edges simultaneously in one step.) We then construct a weight rank clique filtration (WRCF) [51] by determining cliques and tracking their changes in each step of the network sequence. We then compute PH and Betti numbers [34, 65] of the WRCF and examine the results by applying tools from statistics and machine learning, respectively, on the persistence landscapes (PLs) and persistence images (PIs) that result from our computation of PH. We compare our findings from these two approaches. We focus on loops (with four or more edges)[5] in the networks in our nested sequence, rather than on connected components, because one can also study the latter using more conventional approaches (such as by computing the spectrum of the combinatorial graph Laplacian matrix [37, 38]). Although it is interesting

[5] We use the term 'loop' to refer to at least four edges in a network that are connected in a way that forms a cycle. Conventionally, loops (other than self-loops) in undirected graphs must have at least 3 edges, and loops in directed graphs must have at least 2 edges. In our paper, we adapt this terminology for the salient topological features that we examine in simplicial complexes.

to also consider higher-dimensional topological features in functional networks, the computational cost of PH is very high [40] (especially in higher dimensions), and we therefore focus on the analysis of loops.

Our paper proceeds as follows. We introduce the data set and the mathematical methods in an intuitive way in section 2, present our findings in section 3, and discuss our comparisons in the context of current biological research in section 4. We give some additional details about a few results and report the results of our computation of Betti curves in appendix A.

2. Methods

2.1. Data set: fMRI data of schizophrenia patients, siblings of schizophrenia patients, and healthy controls

We use a data set that consists of time series from blood oxygen level-dependent (BOLD) fMRI data that was collected from 281 subjects (encompassing 54 schizophrenia patients, 50 healthy siblings of schizophrenia patients, and 177 healthy controls) with 120 time steps (where the length of 1 time step corresponds to $\Delta t = 2\,\mathrm{s}$). The brain regions were determined according to the Montreal neurological institute template [66]. Prior to obtaining the time series, the fMRI data were corrected for head motion, and they were normalized and smoothed with a Gaussian filter. The voxel-wise signal intensities were normalized to the whole-brain global mean. The data set was acquired by Bertolino, Blasi, and their collaborators as part of a larger fMRI data set over a period of approximately 10 years. Subsets of the data set have been studied previously [67–69], although these previous studies of the data did not include the data for siblings.

The experimentalists obtained fMRI images while subjects were performing a block paradigm of a so-called 'n-back task'. During an n-back task, subjects are presented with a sequence of stimuli (such as numbers). In each step m of the sequence, subjects are first shown a number and then asked to recall the number from step $m - n$ of the sequence. For example, during a 2-back task, subjects are shown a sequence $\{\ldots, x_{i-1}, x_i, x_{i+1}, x_{i+2}, \ldots\}$ and are asked to recall number x_{i-1} while being shown number x_{i+1}, recall number x_i while being shown number x_{i+2}, and so on. For the present data set, the block paradigm consisted of alternating blocks of four 0-back tasks and four 2-back tasks.

We preprocess the data to remove noise that arises due to (1) contributions from brain white matter [70] and cerebrospinal fluid [70, 71] (in these areas, one does not expect a response that is related to neuronal processes), (2) spontaneous global signal fluctuations [70, 72, 73], and (3) signal mismatches between images from the head motion of subjects [74]. For each subject and time step, we calculate the mean signal for white-matter brain regions, the mean signal for regions that consist of cerebrospinal fluid, and the mean of the global signal. In addition to these mean values, we also include the squares and cubes of the global-signal means, as well as head-motion parameters (3 translation and 3 rotation parameters), as rows in our 11×120 subject-specific design matrices. We then perform linear regression for each time series using MATLAB's command (PINV()) for the Moore–Penrose pseudoinverse[6]; we exclude brain regions without gray matter from our calculations. We then use the residuals from the regression as our time series for the 120 brain regions that we list in tables A1–A5. Such preprocessing steps are common when working with fMRI data, but they are not uncontroversial. In particular, the effects of global signal regression can alter correlations between time series [9, 75, 76].

2.2. Functional connectivity

We construct functional networks from the fMRI time series for each subject by using the 120 distinct brain regions (see tables A1–A5) as the nodes of the networks and calculating Pearson correlations[7] (without a time lag) between the nodes' time series as a measure of pairwise functional connectivity. The values of the pairwise functional connectivity give the edge weights between the brain regions (i.e., nodes) in the functional networks.

In our key computations in the present paper, we consider four contiguous time regimes of 30 time points each; this yields four functional networks per subject. (The lone exception to this approach is in appendix A.1, in which we use each subject's full time series, which consists of 120 time points, to construct a single functional network for each subject.) Although the four time regimes each overlap temporally with times during which subjects performed one 0-back and one 2-back task, our separation into time regimes is motivated by an interest in potential developments in the dynamics over time, rather than in relating the fMRI response to the task. We represent each functional network using an adjacency matrix $A = A(\text{subject, time regime})$, where the element A_{ij} (for each i and j) encodes the edge weight between node i and node j. Due to the high

[6] Because some of the matrices are ill-conditioned, the resulting networks differ across different runs of the preprocessing. However, in our observations, the matrices differ by only up to 0.2% of their elements between two separate realizations of preprocessing.

[7] There are numerous ways to measure functional connectivity [11, 77, 78]. For a discussion in the context of schizophrenia research, see [7].

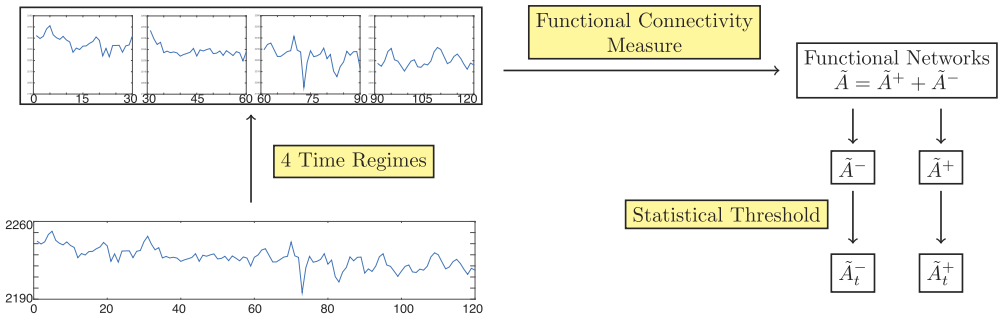


Figure 2. Steps that we perform on the preprocessed time series of each brain region to construct a functional network for each subject during each of four time regimes. We study the positive parts of the resulting networks using PH.

(a) Swiss (Emmental) Cheese (b) Mozzarella Cheese

Figure 3. An example of two topologically-different objects. Homology highlights topological differences by counting the numbers of holes in the cheeses.

computational cost of PH calculations [40], we reduce the number of edges in the networks that we analyze to enable computations. We apply a statistical threshold, which is described in [79], to the weighted adjacency matrices without modifying the remaining edge weights. To obtain the thresholded adjacency matrices, we estimate p-values for the correlations using the MATLAB function CORRCOEF and retain only those elements whose p-value is less than 0.05. Using this type of thresholding, we retain at most 44% of the edges in a network and retain a mean of 20%–30% of the edges in each subject group. We then separate each adjacency matrix into a positive and a negative part, $A = A^+ + A^-$, and study only the positive part A^+ of the adjacency matrix[8]. In discarding negative edge weights, we retain a mean of slightly more than 50% of the elements of our thresholded adjacency matrices. In figure 2, we show a diagram of the steps that we perform to construct our functional networks. Although we consider the four time regimes separately, we treat all subjects and all time regimes together as one data set.

2.3. Persistent homology

Persistent homology is a technique from topological data analysis, which aims to understand the 'shape' of data [40]. PH is based on the topological concept of *homology*, which is used to study the shape of objects in a way that disregards changes from stretching and bending.

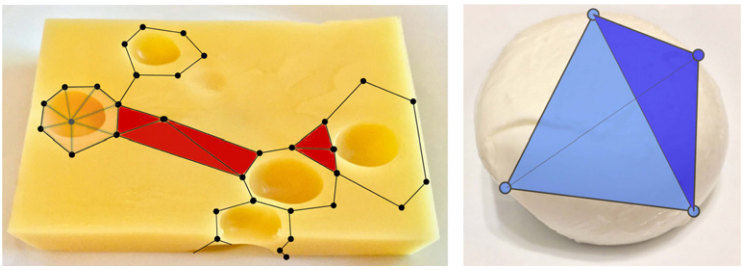
We motivate our use of PH for brain networks by considering different types of cheese and how they differ in their homology. Calculating homology allows one to differentiate between the shape of a stereotypical Swiss cheese (of the Emmental sort) with holes and the shape of a mozzarella cheese by providing information about the presence or absence of holes in the cheeses. (See figure 3 for examples of the aforementioned cheeses.) One can thereby consider the space that surrounds the holes; these are the so-called *loops*. However, homology does not give information about the geometry of the cheeses; for example, it does not 'see' that the Swiss cheese is a cube or that the mozzarella cheese is a sphere (unless it happens to be hollow), as it only characterizes differences in the numbers of holes.

We now give a brief intuitive introduction to a few concepts behind homology and PH for network data. For more mathematical introductions, see [20, 31, 32, 34, 40, 65].

2.3.1. Simplicial complexes

To study the characteristics of topological spaces [80], such as the Swiss cheese and the mozzarella cheese, we consider small pieces ('morsels'), on which we can perform computations more easily. When reassembled, the

[8] We choose to include only positive edge weights to avoid the need to interpret negative correlations between time series.



(a) Swiss (Emmental) Cheese (b) Mozzarella Cheese

Figure 4. Simplicial complexes approximate topological spaces and capture their properties. (In the cross section of the Emmental cheese, all holes are topologically the same in the original object. However, we only visualize parts of the simplicial complex in which one of the holes is tiled with triangles but the others are merely surrounded by edges.)

morsels carry the same overall topological information as the original space. We begin building these morsels (i.e., 'spaces', to be more formal) using a discrete set of points, which we call 'nodes'. We then add 'edges' to connect pairs of nodes; 'triangles', which consist of three nodes, three edges, and a face; 'tetrahedra'; and so on. Formally, each of these elements is called a 'k-simplex', where k indicates the dimension of the simplex. A point is a 0-simplex, an edge is a 1-simplex, a triangle is a 2-simplex, and a tetrahedron is a 3-simplex.

We can combine different simplices to capture different aspects of a topological space. For example, to capture the holes in the Emmental cheese, we glue together a collection of triangles and edges around the holes; we enclose the same number of holes as in the original cheese. Note that we can only capture the holes that are enclosed inside the cheese (using the triangles), as one can deform the visible holes on the surface into a smooth surface of the cheese. For the purpose of demonstration, we therefore assume that the Emmental cheese in figure 3 is a cross section of a larger cheese that encloses the holes that are visible in the image.

One can combine simplices to obtain a *simplicial complex* Σ, and we take the *dimension* of Σ to be the dimension of its highest-dimensional simplex. We show examples of simplicial complexes in figure 4, where we again note that we are assuming that the Emmental cheese is a cross section of a larger hunk of cheese.

2.3.2. Homology and Betti numbers

Homology assigns a family of vector spaces (called *homology groups* in more general settings) to a simplicial complex. For a given dimension, the vector spaces capture the topological features in that dimension. For example, for dimension 0, homology gives a vector space whose elements are connected components; for dimension 1, homology gives a vector space that has loops as its elements. The dimensions of these vector spaces are called *Betti numbers*, where β_D denotes the Betti number for dimension D. The first three Betti numbers (β_0, β_1, and β_2) count, respectively, the number of connected components, the number of one-dimensional (1D) holes (i.e., loops), and the number of two-dimensional (2D) holes (i.e., voids) in a simplicial complex.

2.3.3. Weight rank clique filtration (WRCF)

Similarly to being able to distinguish between two types of cheese, we are interested in whether we can use homology (and specifically PH) to distinguish between functional networks of schizophrenia patients, siblings of schizophrenia patients, and healthy controls.

In a network, we take a loop to consist of a sequence of four or more nodes and edges that begins and ends at the same node. If two loops surround the same hole and can be deformed into one another in the space without tearing open either of the loops, then one counts the loops only once, and we interpret them to be different *representatives* (i.e., *generators*) of a loop.

To obtain simplicial complexes from a weighted network, we construct a so-called *filtration*. A filtration is a sequence of embedded simplicial complexes that starts with the empty complex:

$$\emptyset = \Sigma_0 \subseteq \Sigma_1 \subseteq \Sigma_2 \subseteq \cdots \subseteq \Sigma_{\max} = \Sigma\,.$$

One can obtain a filtration from data in various ways [81]. When given data in the form of a weighted network, the easiest method to construct a filtration is to filter by weights [82]. In the first filtration step, one includes all nodes and the edge(s) with the largest weight in the simplicial complex. In the second step of the filtration, one adds the edge(s) with the second-largest weight to the simplicial complex from step one, and so on. In this way, one obtains a sequence of embedded simplicial complexes that satisfies the properties of a filtration. To construct a WRCF [51], one performs one additional step: whenever three edges in a simplicial complex of a filtration form a triangle, one fills in the associated face and interprets the triangle as a 2-simplex. Similarly, when four nodes are all connected pairwise by edges, the nodes form a (filled) tetrahedron (i.e., a

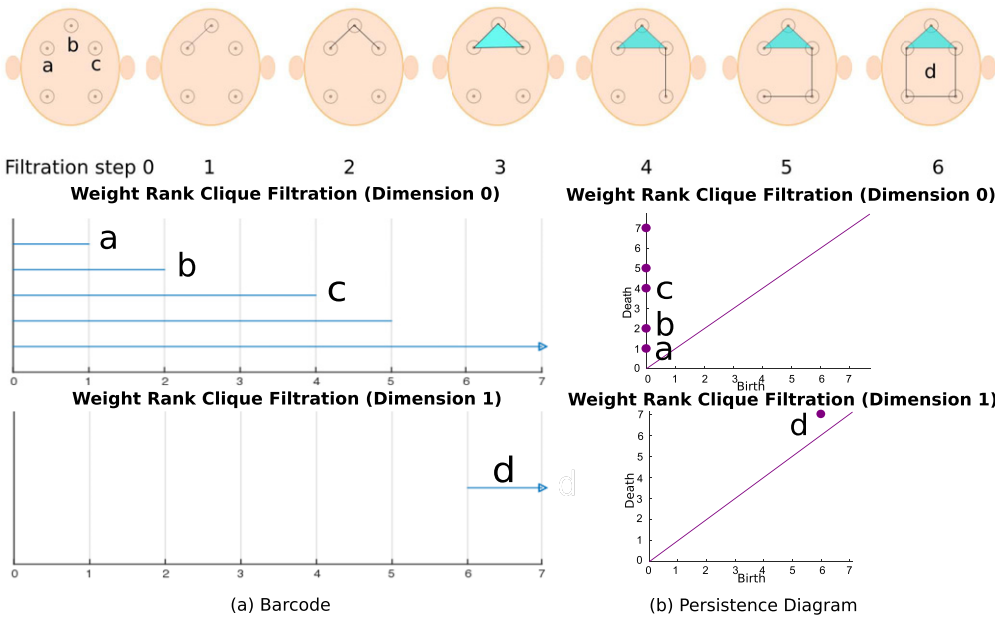


Figure 5. Example of a WRCF of a neuronal network and the corresponding (a) barcodes and (b) PDs in dimensions 0 and 1. The neuronal network consists of different brain regions (indicated by circles), which we interpret as the nodes (indicated by dots) of a network, and weighted edges between the nodes. To construct the filtration, we add the nodes in step 0, followed by the edge with the largest weight in step 1, the edge with the second-largest weight in step 2, and so on. As soon as three nodes are all connected pairwise by edges, we cover the resulting region with a triangle. When four nodes are all connected pairwise, we fill in a tetrahedron. In a dimension-0 barcode, we represent each connected component by a bar that starts when the component is born and ends when it dies (e.g., when two components combine with each other). In a dimension-1 barcode, each bar represents a loop, which consists of 4 or more edges and starts and ends at the same node. In a PD, one represents topological features by points, rather than by bars. The distance of a point to the diagonal (the purple line) indicates the persistence of the corresponding feature in the filtration.

3-simplex). We use the WRCF to analyze our weighted networks. The WRCF has been applied to weighted neuronal networks in several previous studies, including [20, 51, 53, 83].

One can use homology to study topological features, such as loops, in every step of a filtration and determine the extent to which a feature persists with respect to the filtration [40]. We say that a topological feature h in a given dimension is *born* at filtration step m if the homology group of Σ_m is the first homology group of a simplicial complex in the filtration to include that feature. Similarly, we say that a topological feature *dies* at filtration step n if it is present in the homology group of Σ_{n-1} but not in the homology group of Σ_n. The lifetime of a feature in a filtration is the *persistence* p, which is defined by the equation

$$p = n - m\,. \tag{1}$$

If a feature persists until the last filtration step, we say that it has *infinite persistence*. Persistence was first used as a measure to rank topological features based on their lifetimes in a filtration in [30].

Ideally, one performs a WRCF on a fully connected functional network. However, because of the high computational cost of doing so, this is often impossible in practice. We avoid this issue by thresholding our weighted networks before analyzing them.

2.4. Representations of PH

There are multiple ways to represent the output of PH calculations and to visualize the persistence of topological features that one obtains from a filtration. The most common representations are barcodes and persistence diagrams (PDs). In recent years, a desire to leverage the output of PH computations for machine-learning and data-mining tasks has resulted in the development of alternative representations to both barcodes and PDs [40]. Two of these alternative representations are persistence landscapes (PLs) [84, 85] and persistence images (PIs) [86]. In the following subsubsections, we describe barcodes, PDs, PLs, and PIs.

2.4.1. Barcodes

A common representation of the output of PH calculations is a *barcode* [32, 87]. See figure 5 for an example. A D-dimensional barcode is a plot of a collection of filtration parameter intervals $\{[\text{birth}, \text{death})_l\}_{l=1}^{t}$ that indicate the births and deaths of topological features of dimension D. The horizontal axis represents the filtration steps, and each D-dimensional topological feature in a filtration is represented by a bar that starts at the filtration step at which the feature is born and ends at the filtration step at which it dies. In a dimension-0 barcode,

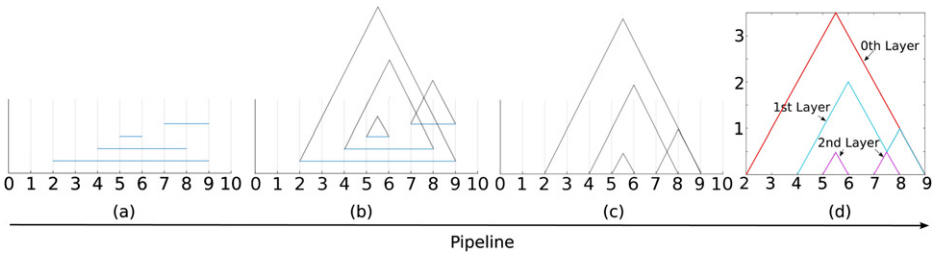


Figure 6. Schematic illustration of the steps for converting a barcode into a persistence landscape (PL). We use an example based on a WRCF in dimension 1. (a) Example barcode. (b) One defines peak functions on the bars of the barcode. (c) One collapses the peak functions onto the horizontal axis. (d) The PL consists of the collection of layers q (with $q = 0$, $q = 1$, and $q = 2$ in this example), which indicate the qth-largest values of the collection of peak-function values. To visualize the third dimension, we show the different layers using different colors. (This figure is a modified version of a figure in [20].)

each bar corresponds to a connected component, and the length of a bar indicates how long a particular component is disconnected from other components in a simplicial complex. Similarly, in a dimension-1 barcode, each bar corresponds to a loop in a simplicial complex.

2.4.2. Persistence diagrams

As an alternative to a barcode, one can use a *persistence diagram* (PD) [88], which is a planar representation of a barcode that conveys the same information. One maps each [birth, death) interval in a barcode to birth–death coordinates, where the horizontal coordinate of a point represents the birth time of a feature in the associated filtration and its vertical coordinate represents the death time of that feature. Alternatively, one can use a birth–persistence coordinate system, which is particularly useful when examining PIs (which we will discuss shortly). Points that are farther away from the diagonal identity line represent more-persistent topological features in a filtration. We show an example of a PD in figure 5. As with barcodes, one can treat PDs as mathematical objects, and one can endow the space of PDs with a distance.

2.4.3. Persistence landscapes

A persistence landscape (PL) [84, 85], which we often will call simply a 'landscape', is a sequence of piecewise-linear functions that one can use to visualize and analyze the information in a barcode or PD. Instead of using a bar and its length to represent a feature and its persistence, one now interprets each topological feature as a peak, whose height is determined by the feature's persistence and whose location corresponds to the feature's location in the filtration. In contrast to a barcode or a PD, a PL has three dimensions. As in a barcode, the horizontal axis represents the filtration step. The other two dimensions of a PL are the persistence of a feature and the different layers of the PL.

To create a PL from a barcode, one first defines a peak function for each bar. For a given [birth, death) interval in a barcode, one constructs the function

$$f_{[\text{birth},\text{death})}(x) = \begin{cases} 0\,, & \text{if } x \notin (\text{birth}, \text{death}) \\ x - \text{birth}\,, & \text{if } x \in \left(\text{birth}, \dfrac{\text{birth} + \text{death}}{2}\right] \\ -x + \text{death}\,, & \text{if } x \in \left(\dfrac{\text{birth} + \text{death}}{2}, \text{death}\right). \end{cases} \tag{2}$$

One then collapses the collection of peak functions onto the horizontal axis of the barcode. For a barcode that consists of the collection $\{[\text{birth}, \text{death})_l\}_{l=1}^{t}$ of intervals, the qth layer (with $q \geqslant 0$) of the PL (i.e., the qth PL) is the following set of functions:

$$\begin{aligned} &\lambda_q : \mathbb{R} \to \mathbb{R}\,, \\ &\lambda_q(x) = q\text{th-largest value of } \left\{f_{[\text{birth},\text{death})_l}(x)\right\}_{l=1}^{t}. \end{aligned} \tag{3}$$

If the qth-largest value does not exist, then $\lambda_q(x) = 0$. The 0th layer of a PL consists of the maximum function values among the collection of functions that one evaluates across a filtration. Similarly, the 1st layer of a PL consists of the second-largest values of the collection of functions that one evaluates across a filtration. One defines other layers in an analogous way. Formally, the *persistence landscape* λ of a barcode $\{[\text{birth}, \text{death})_l\}_{l=1}^{t}$ is defined as the sequence $\{\lambda_q\}$ of the functions λ_q. We illustrate the pipeline to go from a barcode to a PL in figure 6.

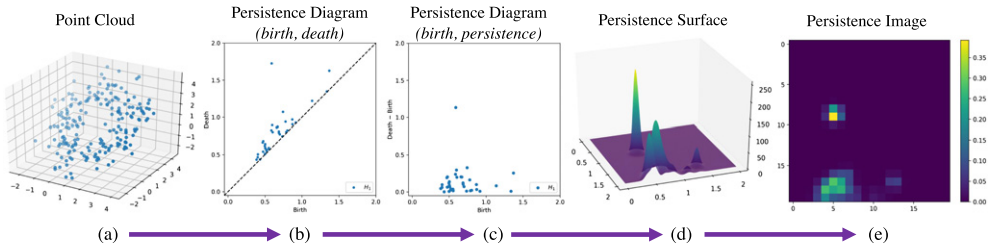


Figure 7. Schematic illustration of the primary steps for converting a PD to a PI. (a) A sample point cloud in $\mathbb{R}^3$. (b) A PD in birth–death coordinates (i.e., the standard choice); we also show the diagonal identity line. (c) A PD in birth–persistence coordinates. (d) The 'persistence surface' that is generated by centering 2D Gaussian distributions at each point in panel (c). (e) One generates a PI by summing the volumes under the 2D Gaussian distributions on the pixels (specifically, on the squares in a uniformly-spaced overlaid grid).

An advantage of PLs is that one can construct a mean PL for a set of landscapes. A mean landscape no longer corresponds to a barcode or a PD. However, one can define pairwise distances between two or more mean landscapes and use them to quantify the difference between two sets of barcodes. We use the L_2 distance. One can also use a variety of statistical tools on PLs [84]. Such calculations have been used for applications like conformational changes in protein binding sites [89], the origin of seizures in electroencephalographic (EEG) data from epileptic patients [90], phase separation in binary metal alloys [91], brain geometry in neurodegenerative diseases [92], audio signals in music [93], and motor learning in humans [20].

2.4.4. Persistence images

Another representation of topological features in PH calculations are *persistence images* (PIs), which are based on PDs and take the form of real-valued vectors[9] that one can use as an input to a variety of machine-learning approaches. The transformation from a PD to a PI is stable with respect to the 1-Wasserstein distance and maintains a clear and interpretable connection to the original PD [86]. In recent work, PIs were used to classify different types of neurons [94, 95].

We show a schematic illustration of the mapping from a PD to a PI in figure 7, which depicts the various stages that are involved in the transformation. Recall that the output of a PH computation is a set of points (or intervals) that correspond to the birth and death times of each topological feature for a specified homological dimension. In figure 7(b), we show a PD for the point cloud in figure 7(a). In figure 7(c), we show the PD in the birth–persistence coordinate system. In figure 7(d), we show an overlay of the surface that is generated by centering a 2D Gaussian on each point in the PD in figure 7(c). Finally, in figure 7(e), we show an example PI that is produced by computing the volume under the surface in figure 7(d) over a uniformly-spaced grid. (We set the resolution so that we have a 20×20 grid of elements.) One can then reshape this final PI into a vector by stacking the columns (or, alternatively, the rows), as is often done in image processing. As described in [86], the generation of a PI involves the choice of (1) a 2D probability density function to center at each point in the birth–persistence PD, (2) a resolution, and (3) a weighting function. The role of the weighting function is, when necessary, to suppress points in a PD that lie very close to the diagonal and are often construed as 'noisy' features. For all of the PIs that we examine in the present paper, we use the default settings for the code: 2D Gaussian probability density functions, a linear weighting function, and a 50×50 grid of elements. We choose additional parameters that are associated with these choices (e.g., the variance of the Gaussians) according to the defaults in [96]. When analyzing and comparing multiple PIs, there is an additional pair of values that one must choose based on the data; these are the maximum birth time and maximum persistence value. We will see in section 3.3 that this pair of values can influence our results.

2.4.5. Software employed

For our PH calculations, we implement MATLAB code that we construct using JAVAPLEX [97], a software package for PH[10]. For a given filtration of a simplicial complex, JAVAPLEX can output [birth, death) barcode intervals, representatives for each topological feature, and PDs. It outputs PDs in standard birth–death coordinates, from which one computes birth–persistence coordinates as (birth, death − birth). For the WRCF, we also use a maximal clique-finding algorithm (that is based on the Bron–Kerbosch algorithm [98]) from the Mathworks library [99]. For the analysis and interpretation of our barcodes, we use the PERSISTENCE LANDSCAPE TOOLBOX [85]. We create PIs using the code at [96] with the default parameters.

[9] We use the term 'vectorization' for the production of such a vector from a PD. One PD produces one PI, which yields one vector after it is reshaped.

[10] For an overview of available PH software and additional references, see [40].

2.5. Clustering methods from data mining and network analysis

Given the output of PH calculations, it can be insightful to use clustering methods to compare the PHs of different networks. There are myriad ways to proceed. In the present paper, we use a few different approaches. First, we apply the k-means clustering algorithm and community detection to examine whether we can separate the three subject groups based on the topological features of their functional networks. Second, we apply a linear sparse support vector machine (SSVM) to identify pixels in PIs to discriminate between the subject groups and examine which brain regions are generators of loops that help discriminate between groups. We describe these techniques in the following subsubsections.

2.5.1. Employing k-means clustering for subject-group separation

The method of k-means clustering aims to produce a partition of a metric space into k clusters of points [100]. Suppose that there are μ data points in a metric space. One selects k of the μ points of a data set as 'centers' and assigns all other points of the data set into clusters based on their closest center point. The 'score' of such a clustering is the sum of the distances from each point to its nearest center. The desired output of k-means clustering is an assignment of points to clusters with the minimum clustering score. However, an exhaustive search for a global minimum is often prohibitively expensive. A typical approach to search for a global minimum is to choose a large selection of k initial centers uniformly at random, iteratively improve each selection of centers until the clustering score stabilizes, and then return the identified final clustering with the lowest score for each initialization. One iteratively updates the centers by setting each new center to be the mean of the points that are assigned to the associated center in the current iteration. One can apply k-means clustering either to a distance matrix (which one can calculate for the set of PDs or for the set of PLs) or to a set of input vectors (such as those that one obtains from a PI).

2.5.2. Community detection for PL classification

Community detection is a method from network analysis that attempts to partition a network into sets (called 'communities') of nodes that are more densely connected to themselves than to other sets of nodes [37, 101, 102]. One can detect communities in either weighted or unweighted networks. In a weighted network, the sums of the edge weights are larger within communities than between them.

One can also use community detection to partition data (e.g., for classification) by studying a distance matrix of data objects (such as mean PLs). We interpret the n PLs as n nodes of a network and convert the pairwise distances between them into edge weights, where a large edge weight signifies a short distance between two PLs and a small edge weight signifies a long distance between two PLs. We convert the distance $d(i,j)$ between landscapes i and j into an edge weight A_{ij} between nodes i and j with the following formula:

$$A_{ij} = 1.01 - \frac{d(i,j)}{\max_{i,j\in\{1,\ldots,n\}}\{d(i,j)\}}\,. \tag{4}$$

This yields an adjacency matrix A with elements A_{ij}. There are many choices for converting from pairwise distances to pairwise weights, and one has to be careful about how that influences community structure and other network computations.

There are numerous methods that one can use for community detection in networks [102]. One approach for decomposing a network into communities (i.e., for performing a 'hard partitioning') is to seek a partition that maximizes an objective function Q. The objective function that we use is modularity

$$Q = \sum_{i,j}[A_{ij} - \gamma P_{ij}]\delta(g_i, g_j)\,, \tag{5}$$

where P (with elements P_{ij}) is a null-model matrix (which specifies the expected edge weight between nodes i and j), the resolution parameter γ is a factor that determines how much weight one gives to the null model, and $\delta(g_i, g_j) = 1$ if nodes i and j are in the same community (i.e., if $g_i = g_j$) and $\delta(g_i, g_j) = 0$ otherwise [101, 102].

For our computations, we use the GenLouvain package [103, 104], which maximizes Q using a variant of the Louvain algorithm [105] to algorithmically detect communities in the adjacency matrix that we construct from our mean PLs. We vary the resolution parameter γ to compare results for different values of γ.

2.5.3. Linear SSVMs for discriminatory feature selection

The 1-norm, regularized, linear support vector machine (i.e., the SSVM) classifies data by generating a separating hyperplane between data points in a way that depends on very few input features [106–108]. A hyperplane is a flat surface that cuts an ambient space into two parts. One can use an SSVM to identify discriminatory features between different groups of data points. One implements linear SSVM feature selection on data points in the form of vectors, so we can use it on our PIs to select 'distinguishing pixels' during classification. In a PI, a

distinguishing pixel is a bounded region in the birth–persistence coordinate system. For clarity, we use the term 'distinguishing pixel' to signify a region that is selected by SSVM and a 'feature' to refer to a topological feature from a PH computation. During the analysis of our results (see section 3.3), we aim to match distinguishing pixels to their corresponding features.

We apply a 'one-against-all' SSVM to PIs that we generate from dimension-1 output of WRCF computations from each subject to identify pixels in the PIs that can discriminate between the subject groups. In a one-against-all SSVM, there is one binary SSVM for each group to separate members of that group from the members of all other groups. In our case, this amounts to defining three hyperplanes: one that separates patients from controls and siblings, one that separates siblings from patients and controls, and one that separates controls from patients and siblings. We use a 5-fold cross-validated SSVM. We specify an optimal separating hyperplane by a normal vector, and we use the term 'SSVM weights' for the values of the elements of the normal vector. We select distinguishing pixels for each classifier by retaining the vector elements (which are pixels in this application) with nonzero SSVM weights, ordering the nonzero SSVM weights by decreasing magnitude, and discarding SSVM weights when the ratio of successive SSVM weights drops below a user-specified tolerance. For details, see [109].

Given a set of distinguishing pixels, we can see for each subject whether its associated functional networks have any loops that are born and persist in the corresponding bounded PI regions. If there are loops in a bounded PI region, we can identify a set of representative brain regions for each of these loops in a subject's network. We are thereby able to leverage PIs to obtain biologically interpretable information about the involvement of different brain regions in the task (as measured with fMRI) for different subject types.

3. Results

We now present the results of our PH computations to examine loops in functional brain networks. We focus exclusively on topological features in dimension 1. We perform our computations on all four time regimes as part of one data set, rather than separating the data for each time regime. We run our PH computations on four functional networks per subject. From the PH output, we create either PLs or PIs. We then perform our computations either on (1) the full data set of PLs or PIs of 281 subjects and four time regimes (which gives 1124 landscapes or PIs, respectively, for the data set) or on (2) the 12 subject-group means of the landscapes or PIs (from three subject groups with four time regimes each). We indicate which of the two cases we are examining in the relevant subsections.

For both PLs and PIs, we find that there seem to be differences in the topological features of the functional networks between subject groups, although we only observe these for PLs when examining means across groups. To illustrate limitations of the methods, we also discuss results in which we were unable to discern differences between subject groups.

3.1. Results of *k*-means clustering on PLs

By applying *k*-means clustering to the mean PLs, we are able to separate siblings of schizophrenia patients from controls and patients. Recall that we use all four time regimes for each of the 12 mean landscapes.

We construct mean PLs from the dimension-1 barcodes (i.e., the barcodes that represent loops in the networks) for each time regime and each subject group. We obtain 12 mean landscapes and exclude infinitely-persisting bars, because all of our landscapes include such features and these tend to dominate the first several layers of the landscapes. Other researchers have excluded layers of landscapes (e.g., the first twenty) to filter out 'topological noise' [110]. Although we threshold our weighted networks prior to analyzing them, this does not necessarily imply that we lose significant information by disregarding the infinitely-persisting features. Additionally, such features do not necessarily correspond to the most-persistent features in barcodes, as even features that are born in the last filtration steps are infinitely persisting if they do not die during the filtration. In our case, the presence of infinitely-persisting features prevented us from discriminating between landscapes based on the pairwise distances between them. When we examined the distributions of the infinitely-persisting features in each subject group without the other features, we did not observe any noticeable differences between the three subject groups.

We calculate a pairwise L_2 distance matrix of the mean landscapes, and we then perform *k*-means clustering on this landscape distance matrix (which has 12×12 elements). For $k = 3$, we obtain the expected division of the mean landscapes into patients, controls, and siblings. Although the fact that one can separate the three cohorts based on fMRI data is not a new finding—see, for example, [6, 7, 22–26] for patients versus controls and [111] for patients versus siblings—the novelty of our calculation is that *k*-means clustering successfully distinguishes between the three different cohorts based on topological information (in the form of loops) in the functional networks.

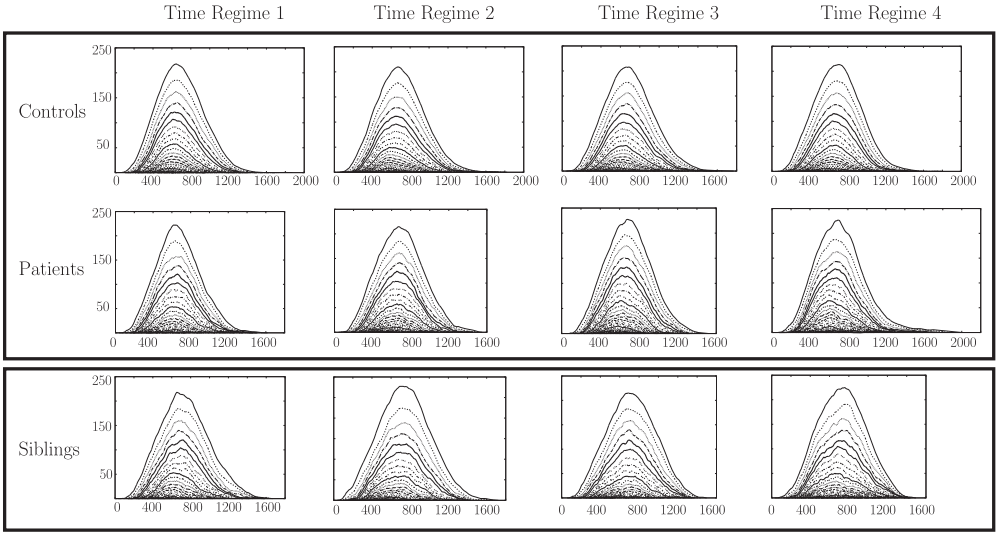


Figure 8. Mean PLs for each of the four time regimes and subject groups. Using k-means clustering with $k = 2$ on the set of 12 PLs (which consists of all subject-group means and time regimes as one data set) assigns patients and controls to one group.

Table 1. Using a permutation test, we calculate p-values for the mean landscape distances between the three subject groups in each time regime.

	Controls versus patients	Controls versus siblings	Patients versus siblings
Time regime 1	0.302	0.200	0.051
Time regime 2	0.460	0.009	0.052
Time regime 3	0.477	0.102	0.270
Time regime 4	0.736	0.110	0.229

We also perform k-means clustering with $k = 2$. Surprisingly, we find that the patients and controls are grouped into one cluster and that the siblings are in a separate cluster. We show the mean landscapes and clusters in figure 8. For $k \geqslant 4$, we do not observe a clear subject-group separation.

To compare our results with ones from other clustering methods, we also apply average linkage clustering to the mean landscape distance matrix and perform community detection on a network that we construct from it (see section 2.5.2). We obtain the same qualitative result for these two methods as we did for k-means clustering. For community detection, we observe a clear separation for resolution-parameter values of $\gamma = 0.82, 0.83, \ldots, 1.14$ into two communities (the siblings versus the patients and controls). These results appear to support our prior observation that the sibling cohort is particularly distinct from the other two cohorts, as compared to any other pairwise comparison of the three cohorts, with respect to the PH of the loops in their functional networks.

We also perform a permutation test (as suggested in [85]) on the mean PLs for each time regime to determine the significance of the landscape distances. In this permutation test, we regroup the individual landscapes into three groups uniformly at random, create a new mean landscape for each of the new groups, and calculate the pairwise L_2 distances between the mean landscapes of the regrouped subjects. We then count how many of these L_2 distances are larger than the ones that we observe when using the mean landscapes of the three subject groups. We use 10 000 permutations to obtain our results, which we summarize in table 1.

Interestingly, for time regimes 1 and 2, we find almost significant distances (i.e., the p-values are slightly larger than 0.05) between the mean landscapes of the patients and those of the siblings, whereas the p-values for time regime 3 and 4 suggest that the distance is not significant (even though the p-values are small in comparison to the p-values that we observe for controls versus patients). The distance between the mean landscapes of the controls and the siblings appears to be significant for time regime 2, but this does not appear to be the case for the other time regimes, although the p-values are again much smaller than they are for the distances between the mean landscapes of the patients and those of the controls. For the controls and the patients, there are many other divisions into two groups that lead to more extreme distances between the mean landscapes than what one obtains by simply assigning them to a control group and a patient group.

To see if we can further support our result from k-means clustering with $k = 2$, we artificially group the controls and patients into one group to create a mean landscape and again perform a permutation test to verify

Table 2. Using a permutation test, we calculate p-values for the controls-and-patients mean landscape versus the mean landscape of the siblings.

Time regime 1	Time regime 2	Time regime 3	Time regime 4
0.112	0.008	0.092	0.110

Table 3. Number of subjects from each subject group that are assigned to the two communities that we obtain using modularity maximization of a network that we construct using a distance matrix from individual PLs.

Subject group	Number of subjects in community 1	Number of subjects in community 2
Patients	122	94
Controls	418	290
Siblings	93	107

whether the distance between the mean landscapes for the two groups is significant. In table 2, we show the p-values that we obtain with 10 000 permutations.

For time regime 2, we obtain a significant distance, but the p-values for time regimes 1, 3, and 4 are about 0.1. Given the artificial grouping of the two subject groups, we interpret these values as small, although they are not statistically significant.

3.2. Results of community detection using a distance matrix from individual PLs

We construct PLs from each of the dimension-1 barcodes, which we calculate by examining each subject in each of the four time regimes, and we calculate the L_2 distance matrix for the resulting 1124 PLs. We again use the distance matrix to construct a network between the PLs, and we detect communities in this network by maximizing modularity. For $\gamma = 0.92, 0.93, \ldots, 1$, we obtain a separation into two communities. The partition that is closest to what we observe for 2-means clustering of the mean-landscape distance matrix occurs for the resolution-parameter value $\gamma = 0.93$. We summarize our results in table 3.

We also apply k-means clustering and average linkage clustering to the distance matrix from the individual PLs (results not shown). Of all of the classification methods that we use for these distance matrices, community detection appears to perform best at 'separating' the subject groups, although we do not observe a very clear separation.

3.3. Results from our analysis of PIs

We find that PIs can identify discriminatory topological features across the three subject groups that we consider. We generate PIs for each of the subjects for each of the four time regimes for the dimension-1 PDs. We set the resolution, probability density function, and weighting function to the defaults in the PI code at [96].

There are additional values—the maximum birth times and maximum persistence values—that one must choose from the data that one is analyzing using PIs. These values determine the discretization of the pixel boundaries in the PIs once one sets the resolution. Possibilities include taking the maximum birth time and maximum persistence value across all PDs or normalizing each PD relative to its individual maxima. The original paper on PIs [86] took maximum values across all PDs under consideration, although no theoretical rationale was provided for this choice. We were unable to obtain clear results using either of the above two approaches for calculating the maximum birth times and maximum persistence values. For example, in the left image of figure 9, we show the mean vectorized PI for each subject group when we generate the PIs using the maximum birth time and maximum persistence value across all subjects. (We create the mean vectorized PI for each subject group by taking the mean of each vector element. We take the mean across all four time regimes to ensure that we have enough data to make a meaningful comparison.) Observe that the means look very similar, aside from slight differences in their amplitudes. For each group, the mean PI is the mean of the vectorized PIs for all of the individuals in the group. In the top row of figure 10, we show the mean PIs (in image form) when we select the maximum birth time and maximum persistence value across all subjects.

Alternatively, if we use *a priori* knowledge of subject-group membership and fix the maximum birth times and maximum persistence values separately for each subject group (based on the collection of PDs that we compute separately for each subject group), we can discriminate between the three subject groups. This provides a first interesting observation from the PIs: there appears to be nontrivial information in the maximum birth time that corresponds—or almost corresponds, in exceptional cases in which multiple edges have exactly the same weight—to the number of pairs of regions in the brain with positive functional connectivity. (Recall

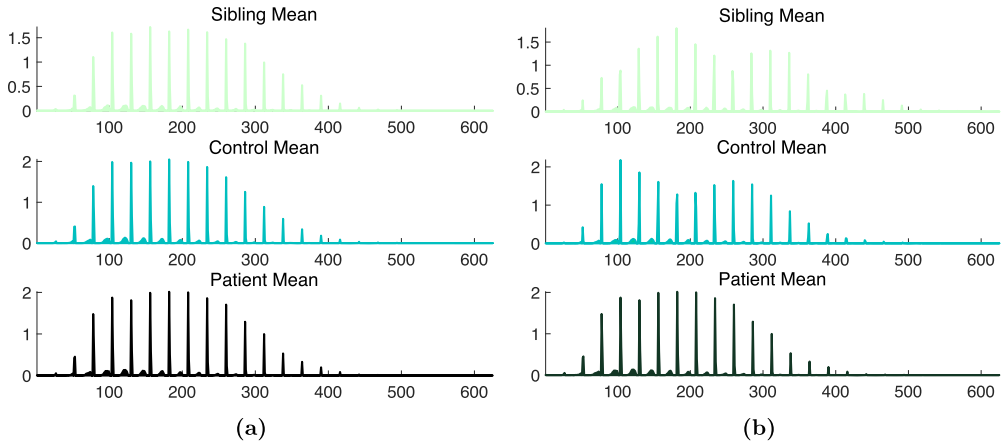


Figure 9. (a) The mean vectorized PI that we generate for each subject group when we use the maximum birth time and maximum persistence value across all subjects to create the PIs. We then calculate the means of the PIs in each group. The horizontal axis corresponds to individual pixels in the PIs, and the vertical axis indicates their intensity values. (b) The mean vectorized PI that we generate for each subject group using the maximum birth times and maximum persistence values that we determine by calculating the maxima for each of the three groups separately. After calculating these maxima, we use this group-specific information to create the PIs for each subject (with one PI for each time regime) within its group. We then calculate the means of the PIs in each group.

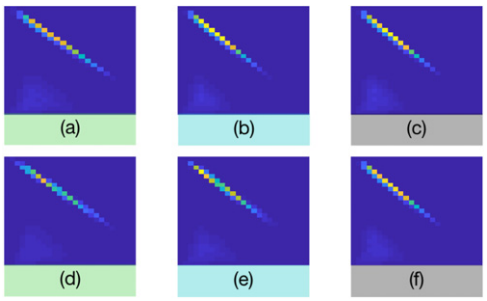


Figure 10. The mean PI for each subject group. We generate panels (a)–(c) using the maximum birth time and maximum persistence value across all subjects to compute all PIs before calculating the depicted means of the PIs in each subject group. We generate panels (d)–(f) using the maximum birth times and maximum persistence values by calculating the maxima separately for each subject group to compute PIs within each group before calculating the depicted means of the PIs in each group. The color axis is the same across rows. From left to right in each row, we show the mean PIs for siblings, controls, and patients.

that we do not add edges that correspond to negative Pearson correlations.) In figure 9(b), we show the mean vectorized PI for each subject group, where we calculate the maximum birth times and maximum persistence values separately for each subject group (instead of setting the maximum birth time and the maximum persistence value to be the same for all subjects). Observe that the sibling and control means both have two humps, whereas the patients have one hump that is clearly discernible. Similarly, in figure 10, we observe two patches along the prominent diagonal with high intensity for the means of the siblings and the controls. However, in the bottom row, we only observe one clear (and elongated) hot spot for the mean of the patients. Therefore, there are multiple, smaller regions in which loops often occur in the filtrations of the functional networks of the siblings and controls, whereas there is seemingly a single, larger region of loops in the filtrations of the networks of the patients.

It is also worth commenting on the locations of the local maxima for each subject group. Relative to the maximum values across each subject type, groupings of loops occur at different locations. From the values of the vector elements, we see that the controls and patients have more similar maximum magnitudes than do the patients and their siblings. Therefore, we conclude that we are able to accurately separate the populations using PIs. Surprisingly, despite the pronounced difference in SSVM performance based on the PIs when we use different maximum values for each subject group, the distributions of the maximum birth times and maximum persistence values for each subject type are not statistically-significantly different from each other. In figure 11, we show Gaussian fits to the sets of maximum birth times and maximum persistence values for each subject type. Observe the strong similarity across all subject groups and the especially close similarity between the control and patient distributions. Because the maximum values are linked closely to the preprocessing of the data, it is important to conduct further research into how to account for these observations. One can also normalize PIs in other ways, such as by normalizing each PI individually by its own maximum value. However,

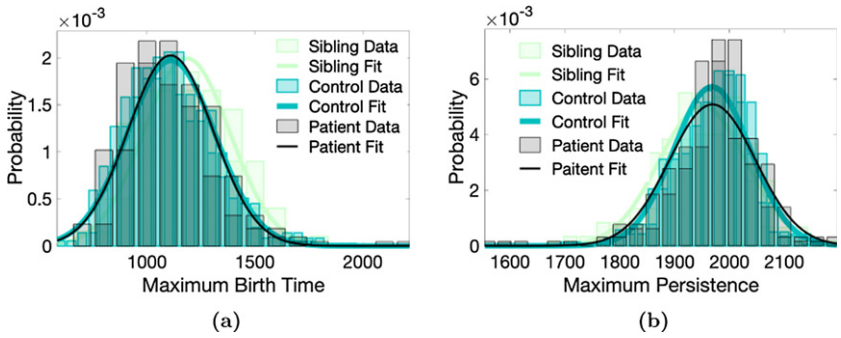


Figure 11. The distribution of (a) the maximum birth times across all samples for each subject type and (b) the maximum persistence values across all samples for each subject type.

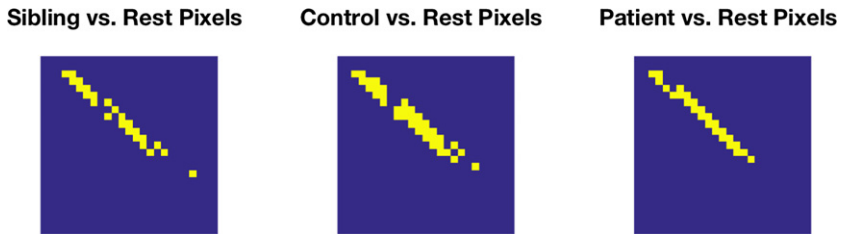


Figure 12. The set of 41 distinguishing pixels (of 625 pixels in total) that we determine via SSVM to be critical for obtaining 100% classification accuracy on the testing set of PIs.

we find that such individual normalization of PIs obscures information (specifically, the maximum birth time of each subject group) that appears to be relevant. The results that we now present are based on the PIs that we generated using the maximum values that we determined based on subject-group membership.

As we discussed in section 2.5.3, it is possible to apply a linear SSVM to the set of PIs to identify distinguishing pixels to help us interpret our classification results. Using a one-against-all SSVM with 5-fold cross-validation, we obtain a 100% classification accuracy. In figure 12, we show the distinguishing pixels from each of the three binary classifiers. By taking the union of these pixels, we obtain 41 distinct pixels from the total of 625 pixels in the PIs. We again emphasize that each of these pixels corresponds to a bounded region in the birth–persistence plane.

Interpreting the distinguishing pixels requires discussing their relationships with particular regions of the brain. We make these connections as follows. For each subject, it is possible to determine whether a topological feature (in our case, a loop) in a filtration of a network exists in the bounded region of the birth–persistence plane that corresponds to a particular distinguishing pixel. If a loop does exist, one can identify a set of brain regions that comprise the loop (i.e., representatives of this loop, as we discuss briefly in section 4). We are particularly interested in brain regions that are consistently involved in the generation of particular loops across subjects. We identify the set of nodes, which we call *top node(s)*[11], that are involved in the generation of one or more loops in each distinguishing pixel across all four time regimes for each subject. We then create histograms of the union of the nodes that we select in this fashion to examine the relative importances of top nodes across each subject type.

In figure 13, we give the relative importances of different brain regions for each pixel. In figure 13(a), we show the top nodes for each subject type based the proportion of the subject types for which that top node is involved in the generation of a loop in the distinguishing-pixel region. In figure 13(b), we show the proportion of the subjects for which the top node(s) is (are) present. The vertical gaps in each plot signify that there are no nodes that are consistently involved in loops for that distinguishing pixel. We make several observations from figure 13(a). First, there are only five distinguishing pixels for which we find top nodes for the patients. Therefore, we are unable to infer which brain regions are involved in loops in the functional networks during the given task for schizophrenia patients. By contrast, there are many distinguishing pixels for which we find top nodes for the siblings. The control group lies between the other two in terms of its number of distinguishing pixels with top nodes, but there are still few top nodes in comparison to the number of distinguishing pixels that have top nodes. In tables 4–7 (see also figures A2–A4 of appendix A), we indicate which brain regions (as

[11] One can interpret our calculation of top nodes in a similar spirit as calculations of node centralities [37].

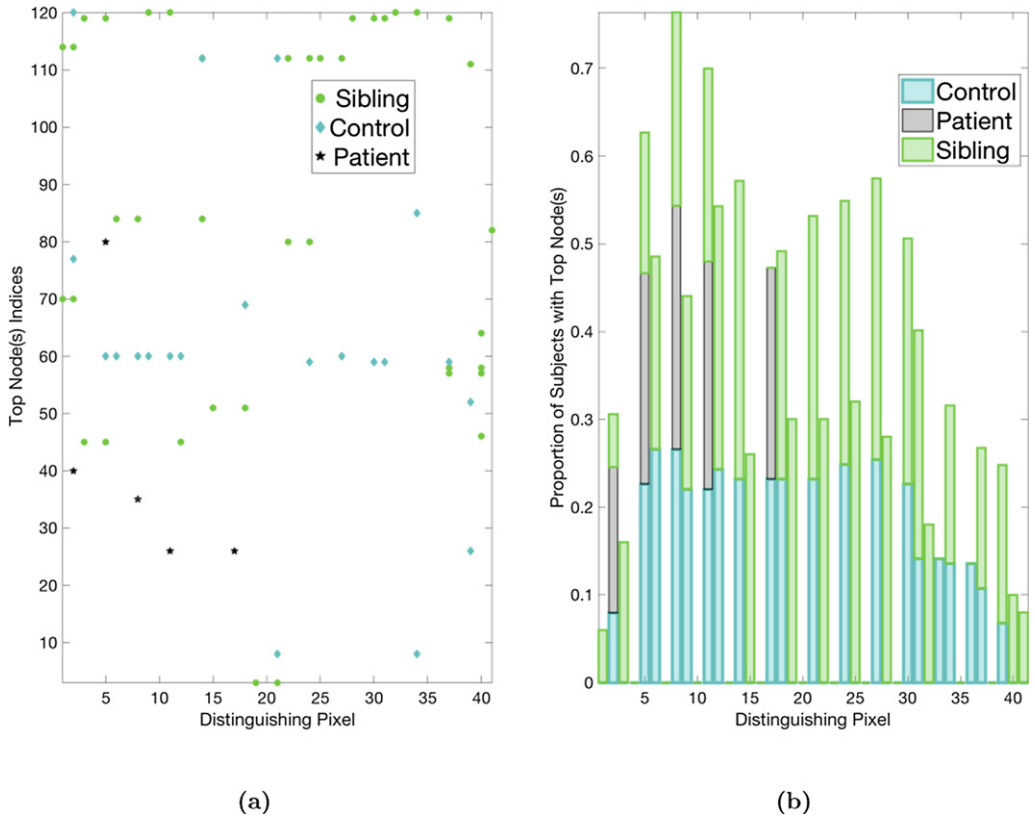


Figure 13. (a) The index (indices) of the top node(s) that are associated with each distinguishing pixel that we determine via SSVM. (b) A stacked bar graph of the proportion of each subject type for which there is a topological feature that includes at least one top node. Pale green indicates siblings, greenish blue indicates controls, and gray indicates patients.

well as their locations) we identify as top nodes. We include only the distinguishing pixels for which top nodes exist within a cohort.

An equivalent way to identify a top node is to calculate the percentage of a given subject group that has a topological feature in the corresponding pixel region (see the bar graph in figure 13) and determine if a specific node is in the group of representative generators for all of the subjects that have a topological feature in the pixel region. We identify a node as a top node if it occurs in the list of representative generators of a topological feature for every subject within the group for which we detect a topological feature in the pixel region. Therefore, when considering tables 4–7, it is possible for the same brain regions to be listed for more than one distinguishing-pixel index. This is also reflected in figure 13 by the occurrence of multiple markers along the same horizontal line.

4. Discussion

We applied methods from persistent homology to analyze loops in functional brain networks of schizophrenia patients, siblings of schizophrenia patients, and healthy controls. We constructed both persistence landscapes and persistence images for these networks, and we compared them to each other using several clustering techniques.

We observed topological differences in the functional brain networks of schizophrenia patients, siblings of schizophrenia patients, and healthy controls with respect to the loops in their networks. We also found that PLs and PIs have different practical advantages and disadvantages when applied to the same data set; these insights may be useful for interpreting the results of PH computations in networks in diverse applications.

Computing PLs gave interesting results when comparing the mean PLs of the cohorts but not when comparing the individual PLs of the subjects. Using mean PLs, we were able to separate the sibling cohort from the other two subject groups in each of the four time regimes. This is supported by the p-values that we obtained for the distances between the mean landscapes of the sibling cohort versus those that we obtained for the control and patient cohorts, although not all of our p-values are statistically significant. The shape of the mean PLs seems to suggest that loops that occur in the functional brain networks of siblings are more persistent on average than those in the functional networks of controls or patients. This may imply either that (1) loops in

Table 4. Top nodes that are involved in loop representatives within the bounds of distinguishing pixel birth–persistence regions (part I). We include only distinguishing pixels for which there is (are) top node(s) within a cohort. 'Left' and 'Right' refer to the hemispheres of the brain. We use the following abbreviations: superior frontal gyrus medial segment (MSFG), superior temporal gyrus (STG), opercular part of the inferior frontal gyrus (OpIFG), transverse temporal gyrus (TTG), frontal operculum (FO), gyrus rectus (GRe), middle frontal gyrus (MFG), orbital part of the inferior frontal gyrus (OrIFG), precuneus (PCu), cuneus (CC), amygdala (Amyg.), anterior insula (AIns), superior parietal lobule (SPL), lingual gyrus (LiG), cerebellum exterior (CE), parahippocampal gyrus (PHG), medial frontal cortex (MFC), medial orbital gyrus (MOrG), and posterior cingulate gyrus (PCgG).

Pixel index	Siblings		Controls		Patients	
	Node	Location	Node	Location	Node	Location
1	70	Left MSFG	—	—	—	—
	114	Left STG	—	—	—	—
2	70	Left MSFG	77	Left OpIFG	40	Left FO
	114	Left STG	120	Left TTG	—	—
3	45	Right GRe	—	—	—	—
	119	Right TTG	—	—	—	—
5	45	Right GRe	60	Left MFG	80	Left OrIFG
	119	Right TTG	—	—	—	—
6	84	Left PCu	60	Left MFG	—	—
8	84	Left PCu	60	Left MFG	35	Right CC
9	120	Left TTG	60	Left MFG	—	—
11	120	Left TTG	60	Left MFG	26	Left AIns
12	45	Right GRe	60	Left MFG	—	—

Table 5. Top nodes that are involved in loop representatives within the bounds of distinguishing pixel birth–persistence regions (part II). We include only distinguishing pixels for which there is (are) top node(s) within a cohort. 'Left' and 'Right' refer to the hemispheres of the brain. We use the following abbreviations: superior frontal gyrus medial segment (MSFG), superior temporal gyrus (STG), opercular part of the inferior frontal gyrus (OpIFG), transverse temporal gyrus (TTG), frontal operculum (FO), gyrus rectus (GRe), middle frontal gyrus (MFG), orbital part of the inferior frontal gyrus (OrIFG), precuneus (PCu), cuneus (CC), amygdala (Amyg.), anterior insula (AIns), superior parietal lobule (SPL), lingual gyrus (LiG), cerebellum exterior (CE), parahippocampal gyrus (PHG), medial frontal cortex (MFC), medial orbital gyrus (MOrG), and posterior cingulate gyrus (PCgG).

Pixel index	Siblings		Controls		Patients	
	Node	Location	Node	Location	Node	Location
14	84	Left PCu	112	Left SPL	—	—
	112	Left SPL	—	—	—	—
15	51	Right LiG	—	—	—	—
17	—	—	112	Left SPL	26	Left AIns
18	51	Right LiG	69	Right MSFG	—	—
19	3	Right Amyg.	—	—	—	—
21	3	Right Amyg.	8	Left CE	—	—
	—	—	112	Left SPL	—	—
22	80	Left OrIFG	—	—	—	—
	112	Left SPL	—	—	—	—
24	80	Left OrIFG	59	Right MFG	—	—
	112	Left SPL	—	—	—	—
25	112	Left SPL	—	—	—	—

the networks of siblings tend to be longer or that (2) one of the three edges between three mutually-adjacent nodes has a substantially smaller edge weight than the other two. Because of the second possible scenario, we see that the occurrence of three brain regions with a large pairwise Pearson correlation between one region and each of the two other regions does not necessarily imply that there is a large correlation between the other two brain regions; such an edge facilitates the formation of a loop structure in a filtration. (Recall that we need at least four nodes for our loops.) To examine this issue further, it may be useful to analyze cross-links in the functional networks, as in [112]. For the above computations and our interpretations of them, we need to take into account that we did not include infinitely-persisting loops (which persist until the end of a filtration). We also include only positive edge weights in our networks, so we only analyzed loops that arise from brain regions with positive pairwise Pearson correlations.

Table 6. Top nodes that are involved in loop representatives within the bounds of distinguishing pixel birth–persistence regions (part III). We include only distinguishing pixels for which there is (are) top node(s) within a cohort. 'Left' and 'Right' refer to the hemispheres of the brain. We use the following abbreviations: superior frontal gyrus medial segment (MSFG), superior temporal gyrus (STG), opercular part of the inferior frontal gyrus (OpIFG), transverse temporal gyrus (TTG), frontal operculum (FO), gyrus rectus (GRe), middle frontal gyrus (MFG), orbital part of the inferior frontal gyrus (OrIFG), precuneus (PCu), cuneus (CC), amygdala (Amyg.), anterior insula (AIns), superior parietal lobule (SPL), lingual gyrus (LiG), cerebellum exterior (CE), parahippocampal gyrus (PHG), medial frontal cortex (MFC), medial orbital gyrus (MOrG), and posterior cingulate gyrus (PCgG).

Pixel index	Siblings		Controls		Patients	
	Node	Location	Node	Location	Node	Location
27	112	Left SPL	60	Left MFG	—	—
28	119	Right TTG	—	—	—	—
30	119	Right TTG	59	Right MFG	—	—
31	119	Right TTG	59	Right MFG	—	—
32	120	Left TTG	—	—	—	—
33	—	—	59	Right MFG	—	—
34	120	Left TTG	8	Left CE	—	—
	—	—	85	Right PHG	—	—
36	—	—	8	Left CE	—	—
	—	—	85	Right PHG	—	—

Table 7. Top nodes that are involved in loop representatives within the bounds of distinguishing pixel birth–persistence regions (part IV). We include only distinguishing pixels for which there is (are) top node(s) within a cohort. 'Left' and 'Right' refer to the hemispheres of the brain. We use the following abbreviations: superior frontal gyrus medial segment (MSFG), superior temporal gyrus (STG), opercular part of the inferior frontal gyrus (OpIFG), transverse temporal gyrus (TTG), frontal operculum (FO), gyrus rectus (GRe), middle frontal gyrus (MFG), orbital part of the inferior frontal gyrus (OrIFG), precuneus (PCu), cuneus (CC), amygdala (Amyg.), anterior insula (AIns), superior parietal lobule (SPL), lingual gyrus (LiG), cerebellum exterior (CE), parahippocampal gyrus (PHG), medial frontal cortex (MFC), medial orbital gyrus (MOrG), and posterior cingulate gyrus (PCgG).

Pixel index	Siblings		Controls		Patients	
	Node	Location	Node	Location	Node	Location
37	57	Right MFC	59	Right MFG	—	—
	58	Left MFC	—	—	—	—
	119	Right TTG	—	—	—	—
39	111	Right SPL	26	Left AIns	—	—
	—	—	52	Left LiG	—	—
40	46	Left GRe	—	—	—	—
	57	Right MFC	—	—	—	—
	58	Left MFC	—	—	—	—
	64	Left MOrG	—	—	—	—
41	82	Left PCgG	—	—	—	—

Although we were able to obtain interesting insights using mean PLs, we did not find interpretable results from comparing individual landscapes, and using only the mean landscapes reduces the amount of information that one can obtain from this approach. By contrast, using individual PIs and SSVMs allowed us to separate the entire set of subjects (with 100% accuracy) in each of the four time regimes. In previous work, Anderson and Cohen [113] obtained 65% accuracy for schizophrenia classification by applying machine-learning techniques to functional brain networks. It is important to note, however, that our results are based on using *a priori* knowledge of group membership (specifically, by including the maximum birth times of loops within subject groups). These birth times seem to include nontrivial information that is important to examine further in future studies. Moreover, such *a priori* knowledge is tied closely to the choice of statistical thresholding when preprocessing fMRI data. Consequently, developing a statistical model that can classify a novel subject based on a PI representation also requires further explorations of how to choose such a threshold.

Computing PIs also allowed us to identify brain regions with consistent involvement in loops in the functional networks within subject cohorts. Of the three cohorts, we found that siblings have the most consistent brain-region involvement across the four time regimes when performing the mental task in this study. That is, brain regions that are involved in loops for siblings in one of the time regimes are more likely to also be

involved in loops in other time regimes than is the case for patients or controls. It is particularly noteworthy that siblings have more brain regions than controls with consistent involvement in loops that are relevant for our SSVM classification of the three cohorts. We view heterogeneous involvement of brain regions in loops as a notion of neurological 'flexibility'. Various works have studied concepts of brain flexibility using community structure [79, 114]. In those studies, flexibility was defined differently—based on how often a brain region (i.e., a node) changes its allegiance to a community of nodes over time, so it does not use loops directly—than in the above characterization, but it is noteworthy that Braun *et al* [114] observed that relatives of schizophrenia patients tend to have larger brain flexibilities than healthy controls. In our work, we found that a specific group of brain regions leads to the separation of the three subject groups when using PIs and observed for the schizophrenia patients that the regions that lead to a separation consistently in each of the four time regimes are fewer in number than for the siblings and the controls. Braun *et al* [114] reported that there is larger node flexibility in the network architecture of schizophrenia patients than in healthy controls. Additionally, Siebenhühner *et al* [115] observed a greater variability in temporal networks that were constructed from magnetoencephalographic (MEG) data of schizophrenia patients than in those that were constructed from the MEG data of healthy controls.

We observed four time regimes, which each consist of fMRI signals that were recorded during one block of a 0-back task and a 2-back task. For time regime 2, we obtained very small and statistically significant p-values in our mean PL computations when comparing siblings to controls and when comparing siblings to a group that consists of all patients and all controls. We did not observe this for any of the other time regimes. We conclude that time regime 2 appears to capture significant changes in the persistence and/or appearance of loops in the networks of siblings during the working-memory task. It will be useful to conduct further laboratory experiments to draw biological conclusions from such computational findings.

Schizophrenia has a high genetic determinism, so siblings of schizophrenia patients have a significant genetic risk of developing the disease themselves [116], and it has been demonstrated that they have abnormalities in their structural neuronal networks [117]. Although our results that functional brain networks that are constructed from fMRI measurements of siblings differ both from those of patients and from those of healthy controls do not agree completely with the current scientific literature, other studies have also reported that the features of fMRIs of siblings of schizophrenia patients differ both from schizophrenia patients and from controls. For example, Callicott *et al* [118] observed in an fMRI study that there was no difference in task performance between healthy siblings of schizophrenia patients and healthy controls, yet they detected a physiological similarity between the sibling cohort and the schizophrenia patients in the associated fMRI data. Similarly, Sepede *et al* [111] observed using fMRI data from a different data set that healthy siblings of schizophrenia patients exhibit differences in brain function from schizophrenia patients, although they did not differ significantly in task performance. Additionally, it was demonstrated recently that schizophrenia patients undergo a cortical normalization process over the course of the disease [119]. However, one needs further phenotypic information to assess whether it is possible to directly connect the results of such a study to observations like ours.

Because our results are somewhat inconsistent with prior observations, it is also possible that our data set contains experimental noise that is beyond our control. When we used standard network-analysis techniques (specifically, community detection) on the functional brain networks, we did not observe meaningful differences between the three subject groups. Nevertheless, we believe that our comparison of PLs to PIs (along with our examination of the different results from these techniques) is a useful illustration of topological data analysis in the study of functional brain networks. To give another cautionary note, one needs to take into account that there are difficulties when interpreting the information about node participation in loops from computations of PH, as the software that is used for such computations (including, specifically, JAVAPLEX, which is what we used) only finds representatives of the loops. These representatives are not determined in an optimal way, and they need not be 'geometrically nice' [120]. For example, in such calculations, one often encounters double loops or even triple loops as generators for one loop in a functional network. Selecting a basis of homology generators that behaves in a biologically representative way corresponds mathematically to solving a problem known as the 'optimal homology-basis problem', which is difficult (and is NP-hard in the worst case) [121]. Despite these difficulties, our list of discriminating nodes provides a useful starting point for further investigations into neuronal abnormalities in functional networks of schizophrenia patients.

Another important issue is that we preprocessed the data for our study. This is very common when working with fMRI data, but such steps are not uncontroversial; studies of functional connectivity in schizophrenia patients have found contradictory results, which depend on whether one performs global signal correction [9, 76]. It is also relevant to keep in mind that the choice of functional connectivity measure can influence results [77]. We used the Pearson correlation because of its simplicity and the fact that it is a widely used measure of functional connectivity [122, 123]. Many other choices are also available.

Finally, we chose to threshold our networks and removed edges whose weights were below a certain amount. However, it has been observed previously [123] that edges with small weights can be important when comparing functional brain networks of schizophrenia patients to those of healthy controls. However, such missing edges—depending on their locations—can result in loops in a network, our analysis indirectly includes some of this information. In future work, it seems interesting to consider only the parts of the functions networks that are below the threshold value that we employed in our present study and analyze them with PH and examine the ensuing classification of controls, patients, and siblings of patients.

Data availability statement

No new data were created or analysed in this study.

Acknowledgments

We thank Alessandro Bertolino, Fabio Sambataro, and the Bari psychiatric neuroscience group for permission to study their data. All rights to the data lie with their research groups; we are unable to release the data. We also thank Pawel Dłotko for his help with the PERSISTENCE LANDSCAPE TOOLBOX and for providing us with new versions of the software during our project. We are also grateful to Danielle Bassett, Peter Bubenik, Carina Curto, Parker Edwards, and several referees for helpful comments. We thank Florian Lipsmeier and Franziska Mech from Roche for useful discussions. We also acknowledge Advanced Research Computing (ARC) at University of Oxford for resources that we used in carrying out this work. BJS thanks the EPSRC and MRC (Grant No. EP/G037280/1) and F Hoffmann-La Roche AG for funding her doctoral studies. HAH acknowledges funding from an EPSRC Fellowship (EP/K041096/1) and a Royal Society University Research Fellowship. BJS and HAH are members of the Center for Topological Data Analysis, which is funded by an EPSRC grant (EP/R018472/1).

Appendix A

We now give some additional details about a few of our calculations and results.

A.1. Betti curves

We also study *Betti curves*, which were introduced in [53] and describe how Betti numbers change across a filtration. We use each subject's full time series, which consists of 120 time points, to construct a single functional network for each subject (i.e., one time regime, rather than four separate ones). In all other respects, we construct the functional networks as we described in section 2.2. We compute the mean and standard deviation across the Betti numbers for dimension 1 (i.e., we compute the number of loops) for each cohort in each filtration step. Aside from a slightly larger standard deviation in the patient cohort, we find that the Betti curves of the three groups look essentially the same. We show our results of computing Betti curves in figure A1.

A.2. Top brain regions in the distinguishing-pixel birth–persistence bounds

In this section, we illustrate the top nodes (i.e., top brain regions) within the bounds of the distinguishing-pixel birth–persistence regions for the three cohorts. Recall that each pixel in the birth–persistence plane corresponds to a bounded region of the original PD (i.e., in the birth–death plane). We show results for siblings in figure A2, controls in figure A3, and patients in figure A4.

A.3. Supplementary tables

In tables A1–A5, we give the numbering of the brain regions and their corresponding IDs.

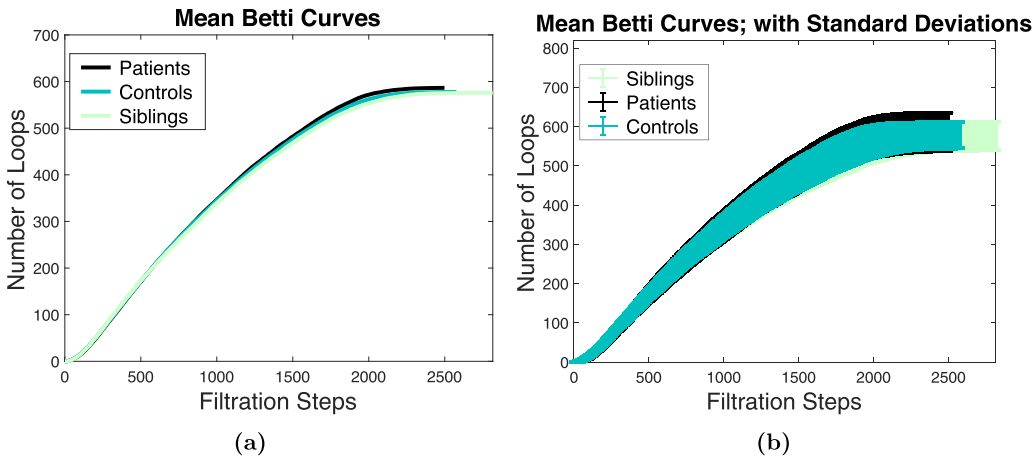


Figure A1. (a) Mean Betti curves for the patients, controls, and siblings. (b) Mean Betti curves and their standard deviations for patients, controls, and siblings.

Figure A2. The top nodes in representatives of the loops that are present within the birth–persistence bounds of the distinguishing pixels for siblings.

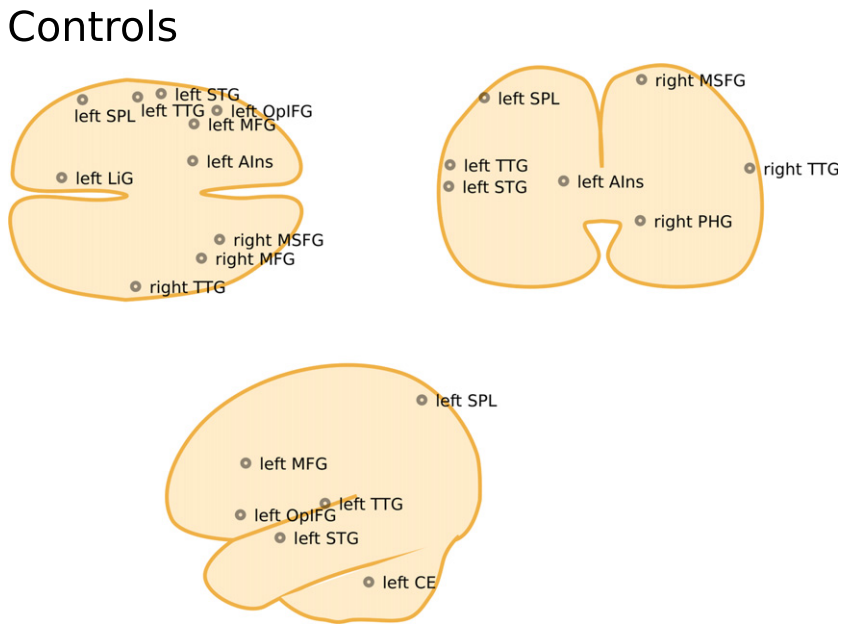


Figure A3. The top nodes in representatives of the loops that are present within the birth–persistence bounds of the distinguishing pixels for controls.

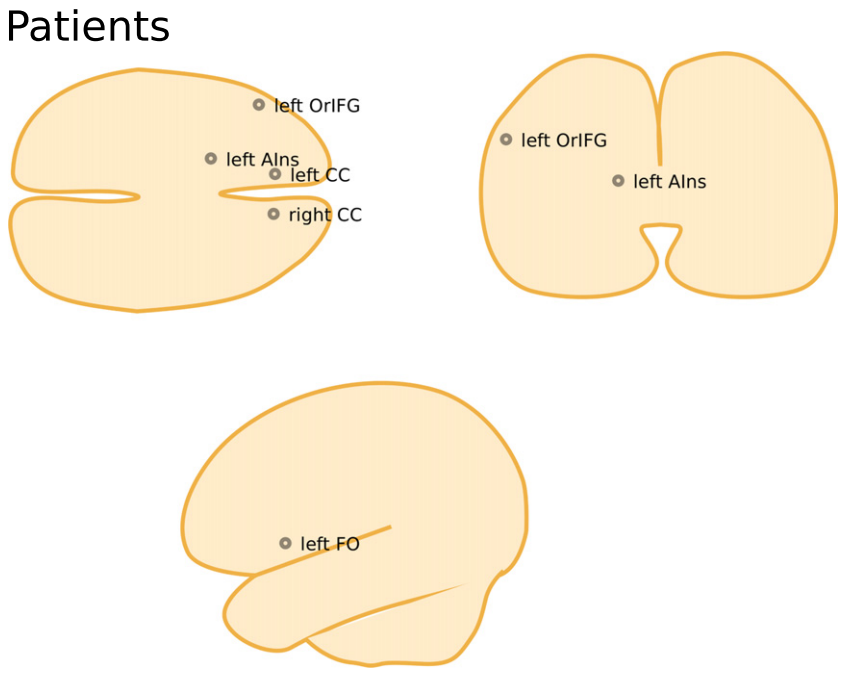


Figure A4. The top nodes in representatives of the loops that are present within the birth–persistence bounds of the distinguishing pixels for patients.

Table A1. Node numbers (NNs) of brain regions (BRs) and their corresponding IDs (part I).

NN	1	2	3	4	5	6
ID	23	30	31	32	36	37
BR	Right accumbens area	Left accumbens area	Right amygdala	Left amygdala	Right caudate	Left caudate
NN	7	8	9	10	11	12
ID	38	39	47	48	55	56
BR	Right cerebellum exterior	Left cerebellum exterior	Right hippocampus	Left hippocampus	Right pallidum	Left pallidum
NN	13	14	15	16	17	18
ID	57	58	59	60	61	62
BR	Right putamen	Left putamen	Right thalamus proper	Left thalamus proper	Right ventral diencephalon	Left ventral diencephalon
NN	19	20	21	22	23	24
ID	71	72	75	76	100	101
BR	Cerebellar vermal lobules I–V	Cerebellar vermal lobules VI–VII	Left basal forebrain	Right basal forebrain	Right anterior cingulate gyrus	Left anterior cingulate gyrus
NN	25	26	27	28	29	30
ID	102	103	104	105	106	107
BR	Right anterior insula	Left anterior insula	Right anterior orbital gyrus	Left anterior orbital gyrus	Right angular gyrus	Left angular gyrus

Table A2. Node numbers (NNs) of brain regions (BRs) and their corresponding IDs (part II).

NN	31	32	33	34	35	36
ID	108	109	112	113	114	115
BR	Right calcarine cortex	Left calcarine cortex	Right central operculum	Left central operculum	Right cuneus	Left cuneus
NN	37	38	39	40	41	42
ID	116	117	118	119	120	121
BR	Right entorhinal area	Left entorhinal area	Right frontal operculum	Left frontal operculum	Right frontal pole	Left frontal pole
NN	43	44	45	46	47	48
ID	122	123	124	125	128	129
BR	Right fusiform gyrus	Left fusiform gyrus	Right gyrus rectus	Left gyrus rectus	Right inferior occipital gyrus	Left inferior occipital gyrus
NN	49	50	51	52	53	54
ID	132	133	134	135	136	137
BR	Right inferior temporal gyrus	Left inferior temporal gyrus	Right lingual gyrus	Left lingual gyrus	Right lateral orbital gyrus	Left lateral orbital gyrus

Table A3. Node numbers (NNs) of brain regions (BRs) and their corresponding IDs (part III).

NN	55	56	57	58	59	60
ID	138	139	140	141	142	143
BR	Right middle cingulate gyrus	Left middle cingulate gyrus	Right medial frontal cortex	Left medial frontal cortex	Right middle frontal gyrus	Left middle frontal gyrus
NN	61	62	63	64	65	66
ID	144	145	146	147	148	149
BR	Right middle occipital gyrus	Left middle occipital gyrus	Right medial orbital gyrus	Left medial orbital gyrus	Right postcentral gyrus medial segment	Left postcentral gyrus medial segment
NN	67	68	69	70	71	72
ID	150	151	152	153	154	155
BR	Right precentral gyrus medial segment	Left precentral gyrus medial segment	Right superior frontal gyrus medial segment	Left superior frontal gyrus medial segment	Right middle temporal gyrus	Left middle temporal gyrus
NN	73	74	75	76	77	78
ID	156	157	160	161	162	163
BR	Right occipital pole	Left occipital pole	Right occipital fusiform gyrus	Left occipital fusiform gyrus	Right opercular part of the inferior frontal gyrus	Left opercular part of the inferior frontal gyrus

Table A4. Node numbers (NNs) of brain regions (BRs) and their corresponding IDs (part IV).

NN	79	80	81	82	83	84
ID	164	165	166	167	168	169
BR	Right orbital part of the inferior frontal gyrus	Left orbital part of the inferior frontal gyrus	Right posterior cingulate gyrus	Left posterior cingulate gyrus	Right precuneus	Left precuneus
NN	85	86	87	88	89	90
ID	170	171	172	173	174	175
BR	Right parahipp-ocampal gyrus	Left parahipp-ocampal gyrus	Right post-erior insula	Left post-erior insula	Right parietal operculum	Left parietal operculum
NN	91	92	93	94	95	96
ID	176	177	178	179	180	181
BR	Right postcentral gyrus	Left postcentral gyrus	Right posterior orbital gyrus	Left posterior orbital gyrus	Right planum polare	Left planum polare
NN	97	98	99	100	101	102
ID	182	183	184	185	186	187
BR	Right precentral gyrus	Left precentral gyrus	Right planum temporale	Left planum temporale	Right subcallosal area	Left subcallosal area

Table A5. Node numbers (NNs) of brain regions (BRs) and their corresponding IDs (part V).

NN	103	104	105	106	107	108
ID	190	191	192	193	194	195
BR	Right superior frontal gyrus	Left superior frontal gyrus	Right supplementary motor cortex	Left supplementary motor cortex	Right supramarginal gyrus	Left supramarginal gyrus
NN	109	110	111	112	113	114
ID	196	197	198	199	200	201
BR	Right superior occipital gyrus	Left superior occipital gyrus	Right superior parietal lobule	Left superior parietal lobule	Right superior temporal gyrus	Left superior temporal gyrus
NN	115	116	117	118	119	120
ID	202	203	204	205	206	207
BR	Right temporal pole	Left temporal pole	Right triangular part of the inferior frontal gyrus	Left triangular part of the inferior frontal gyrus	Right transverse temporal gyrus	Left transverse temporal gyrus

ORCID iDs

Bernadette J Stolz https://orcid.org/0000-0001-8658-8666
Tegan Emerson https://orcid.org/0000-0002-1898-5923
Mason A Porter https://orcid.org/0000-0002-5166-0717

References

[1] World Health Organization 4 October 2019 Schizophrenia. Available at http://who.int/mental_health/management/schizophrenia/en/
[2] Bertolino A and Blasi G 2009 *Neuroscience* **164** 288–99
[3] Dawson N, Xiao X, McDonald M, Higham D J, Morris B J and Pratt J A 2014 *Cereb. Cortex* **24** 452–64
[4] Bullmore E T, Frangou S and Murray R M 1997 *Schizophrenia Research* **28** 143–56
[5] Peled A, Geva A B, Kremen W S, Blankfeld H M, Esfandiarfard R and Nordahl T E 2001 *Int. J. Neurosci.* **106** 47–61
[6] Bassett D S, Bullmore E, Verchinski B A, Mattay V S, Weinberger D R and Meyer-Lindenberg A 2008 *J. Neurosci.* **28** 9239–48
[7] Fornito A, Zalesky A, Pantelis C and Bullmore E T 2012 *NeuroImage* **62** 2296–314
[8] Zalesky A, Fornito A, Egan G F, Pantelis C and Bullmore E T 2012 *Hum. Brain Mapp.* **33** 2535–49
[9] Fornito A and Bullmore E T 2015 *Curr. Opin. Neurobiol.* **30** 44–50
[10] Bullmore E and Sporns O 2009 *Nat. Rev. Neurosci.* **10** 186–98
[11] Bullmore E T and Bassett D S 2011 *Annu. Rev. Clin. Psycho.* **7** 113–40
[12] Sporns O 2014 *Nat. Neurosci.* **17** 652–60
[13] Papo D, Zanin M, Pineda-Pardo J A, Boccaletti S and Buldú J M 2014 *Phil. Trans. R. Soc.* B **369** 20130525
[14] Papo D, Buldú J M, Boccaletti S and Bullmore E T 2014 *Phil. Trans. R. Soc.* B **369** 20130520
[15] Betzel R F and Bassett D S 2017 *NeuroImage* **160** 73–83
[16] Bassett D S and Sporns O 2017 *Nat. Neurosci.* **20** 353–64
[17] Bassett D S, Zurn P and Gold J I 2018 *Nat. Rev. Neurosci.* **19** 566–78
[18] Sporns O 2015 Graph-theoretical analysis of brain networks *Brain Mapping: An Encyclopedic Reference* vol 1 ed A W Toga (Cambridge, MA: Academic) pp 629–33
[19] Petersen S E and Sporns O 2015 *Neuron* **88** 207–19
[20] Stolz B J, Harrington H A and Porter M A 2017 *Chaos* **27** 047410
[21] Eklund A, Nichols T E and Knutsson H 2016 *Proc. Natl Acad. Sci. USA* **113** 7900–5
[22] Lynall M-E, Bassett D S, Kerwin R, McKenna P J, Kitzbichler M, Muller U and Bullmore E 2010 *J. Neurosci.* **30** 9477–87
[23] Rubinov M and Bullmore E 2013 *Dialogues Clin. Res.* **15** 339–49
[24] Alexander-Bloch A, Lambiotte R, Roberts B, Giedd J, Gogtay N and Bullmore E 2012 *NeuroImage* **59** 3889–900
[25] Liu Y *et al* 2008 *Brain* **131** 945–61
[26] Singh M and Bagler G 2016 (arXiv:1602.01191)
[27] Flanagan R, Lacasa L, Towlson E K, Lee S H and Porter M A 2019 *J. Complex Netw.* **7** 932–60
[28] Alexander-Bloch A F, Gogtay N, Meunier D, Birn R, Clasen L, Lalonde F, Lenroot R, Giedd J and Bullmore E T 2010 *Front. Syst. Neurosci.* **4** 147
[29] Towlson E K, Vértes P E, Müller U and Ahnert S E 2019 *Front. Psychiatry* **10** 611
[30] Edelsbrunner H, Letscher D and Zomorodian A 2002 *Discrete Comput. Geom.* **28** 511–33
[31] Edelsbrunner H and Harer J L 2008 Persistent homology—A survey *Surveys on Discrete and Computational Geometry: Twenty Years Later* (*Contemporary Mathematics* vol 453) ed J E Goodman, J Pach and R Pollak (Providence, RI: American Mathematical Society) pp 257–82
[32] Ghrist R 2008 *Bull. Am. Math. Soc.* **45** 61–75
[33] Carlsson G 2009 *Bull. Amer. Math. Soc.* **46** 255–308
[34] Edelsbrunner H and Harer J L 2010 *Computational Topology* (Providence, RI: American Mathematical Society)
[35] Sizemore A E, Phillips-Cremins J E, Ghrist R and Bassett D S 2019 *Netw. Neurosci.* **3** 656–73
[36] Battiston F, Cencetti G, Iacopini I, Latora V, Lucas M, Patania A, Young J-G and Petri G 2020 *Phys. Rep.* **874** 1

[37] Newman M E J 2018 *Networks* 2nd edn (Oxford: Oxford University Press)
[38] Bollobás B 1998 *Modern Graph Theory* (Heidelberg: Springer)
[39] Bassett D S, Yang M, Wymbs N F and Grafton S T 2015 *Nat. Neurosci.* **18** 744–51
[40] Otter N, Porter M A, Tillmann U, Grindrod P and Harrington H A 2017 *Eur. Phys. J. — Data Sci.* **6** 17
[41] Patania A, Vaccarino F and Petri G 2017 *Eur. Phys. J. — Data Sci.* **6** 7
[42] Kramár M, Goullet A, Kondic L and Mischaikow K 2013 *Phys. Rev.* E **87** 042207
[43] Taylor D, Klimm F, Harrington H A, Kramár M, Mishchaikow K, Porter M A and Mucha P J 2015 *Nat. Commun.* **6** 7723
[44] Bhattacharya S, Ghrist R and Kumar V 2015 *IEEE Trans. Robot.* **31** 578–90
[45] Topaz C M, Ziegelmeier L and Halverson T 2015 *PLoS One* **10** e0126383
[46] Bendich P, Marron J S, Miller E, Pieloch A and Skwerer S 2016 *Ann. Appl. Stat.* **10** 198–218
[47] Feng M and Porter M A *Phys. Rev. Res.* **2** 033426
[48] Byrne H M, Harrington H A, Muschel R, Reinert G, Stolz B J and Tillmann U 2019 *Math. Today* **55** 206–10
[49] Curto C and Itskov V 2008 *PLoS Comput. Biol.* **4** e1000205
[50] Dabaghian Y, Mémoli F, Frank L and Carlsson G 2012 *PLoS Comput. Biol.* **8** e1002581
[51] Petri G, Scolamiero M, Donato I and Vaccarino F 2013 *PLoS One* **8** e66506
[52] Lee H, Chung M K, Kang H, Kim B N and Lee D S 2011 Discriminative persistent homology of brain networks *IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro* 841–4
[53] Giusti C, Pastalkova E, Curto C and Itskov V 2015 *Proc. Natl Acad. Sci. USA* **112** 13455–60
[54] Spreemann G, Dunn B, Botnan M B and Baas N A 2018 *Phys. Rev.* E **97** 032313
[55] Curto C 2017 *Bull. Am. Math. Soc.* **54** 63–78
[56] Giusti C, Ghrist R and Bassett D S 2016 *J. Comput. Neurosci.* **41** 1–14
[57] Reimann M W *et al* 2017 *Front. Comput. Neurosci.* **11** 48
[58] Babichev A and Dabaghian Y 2017 Persistent memories in transient networks *Emergent Complexity from Nonlinearity, in Physics, Engineering and the Life Sciences* (Heidelberg: Springer) pp 179–88
[59] Lee H, Chung M K, Choi H, Kang H, Ha S, Kim Y K and Lee D S 2019 Harmonic holes as the submodules of brain network and network dissimilarity *Computational Topology in Image Context. CTIC 2019* (*Lecture Notes in Computer Science* vol 11382) ed R Marfil, M Calderón, F Díaz del Río, P Real and A Bandera (Cham: Springer International Publishing) pp 110–22
[60] Bardin J-B, Spreemann G and Hess K 2019 *Netw. Neurosci.* **3** 725–43
[61] Chung M K, Lee H, DiChristofano A, Ombao H and Solo V 2019 *Netw. Neurosci.* **3** 674–94
[62] Babichev A, Morozov D and Dabaghian Y 2019 *Netw. Neurosci.* **3** 707–24
[63] Geniesse C, Sporns O, Petri G and Saggar M 2019 *Netw. Neurosci.* **3** 763–78
[64] Ibáñez-Marcelo E, Campioni L, Phinyomark A, Petri G and Santarcangelo E L 2019 *NeuroImage* **200** 437–49
[65] Croom F H 1978 *Basic Concepts of Algebraic Topology* (Heidelberg: Springer)
[66] Talairach J and Tournoux P 1988 *Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System: An Approach to Cerebral Imaging* (New York: Thieme Medical Publishers)
[67] Bertolino A *et al* 2010 *PLoS One* **5** e9348
[68] Sambataro F *et al* 2010 *Neuropsychopharmacology* **35** 904–12
[69] Rampino A *et al* 2014 *PLoS One* **9** e99892
[70] Weissenbacher A, Kasess C, Gerstl F, Lanzenberger R, Moser E, Windischberger C, Moser E and Windischberger C 2009 *NeuroImage* **47** 1408–16
[71] Dagli M S, Ingeholm J E and Haxby J V 1999 *NeuroImage* **9** 407–15
[72] Birn R M, Diamond J B, Smith M A and Bandettini P A 2006 *NeuroImage* **31** 1536–48
[73] Fox M D and Raichle M E 2007 *Nat. Rev. Neurosci.* **8** 700
[74] Friston K J, Williams S, Howard R, Frackowiak R S J and Turner R 1996 *Magn. Reson. Med.* **35** 346–55
[75] Murphy K, Birn R M, Handwerker D A, Jones T B and Bandettini P A 2009 *NeuroImage* **44** 893–905
[76] Fox M D, Zhang D, Snyder A Z and Raichle M E 2009 *J. Neurophysiol.* **101** 3270–83
[77] Smith S M, Miller K L, Salimi-Khorshidi G, Webster M, Beckmann C F, Nichols T E, Ramsey J D and Woolrich M W 2011 *NeuroImage* **54** 875–91
[78] Zhou D, Thompson W K and Siegle G 2009 *NeuroImage* **47** 1590–607
[79] Bassett D S, Wymbs N F, Porter M A, Mucha P J, Carlson J M and Grafton S T 2011 *Proc. Natl Acad. Sci. USA* **108** 7641–6
[80] Kosniowski C 1980 *A First Course in Algebraic Topology* (Cambridge: Cambridge University Press)
[81] Feng M and Porter M A 2021 *SIAM Rev.* **63** 67–99
[82] Lee H, Kang H, Chung M K, Kim B N and Lee D S 2012 Weighted functional brain network modeling via network filtration *Advances in Neural Information Processing Systems (NIPS) Workshop On Algebraic Topology And Machine Learning*
[83] Petri G, Expert P, Turkheimer F, Carhart-Harris R, Nutt D, Hellyer P J and Vaccarino F 2014 *J. R. Soc. Interface* **11** 20140873
[84] Bubenik P 2015 *J. Mach. Learn. Res.* **16** 77–102
[85] Bubenik P and Dłotko P 2017 *J. Symb. Comput.* **78** 91–114
[86] Adams H *et al* 2017 *J. Mach. Learn. Res.* **8** 1–35
[87] Carlsson G, Zomorodian A, Collins A and Guibas L J 2005 *Int. J. Shape Model.* **11** 149–87
[88] Cohen-Steiner D, Edelsbrunner H and Harer J 2007 *Discrete Comput. Geom.* **37** 103–20
[89] Kovacev-Nikolic V, Bubenik P, Nikolic D and Heo G 2016 *Stat. Appl. Genet. Mol. Biol.* **15** 1–27
[90] Wang Y, Ombao H and Chung M K 2015 Topological seizure origin detection in electroencephalographic signals *IEEE 12th Int. Symp. on Biomedical Imaging (ISBI 2015)* pp 351–4
[91] Dłotko P and Wanner T 2016 *Physica* D **334** 60–81
[92] Garg A, Lu D, Popuri K and Beg M F 2017 Brain geometry persistent homology marker for Parkinson's disease *IEEE 14th Int. Symp. on Biomedical Imaging (ISBI 2017)* pp 525–8
[93] Liu J Y, Jeng S K and Yang Y H 2016 Applying topological persistence in convolutional neural network for music audio signals (arXiv:1608.07373)
[94] Kanari L *et al* 2019 *Cerebral Cortex* **29** 1719–35
[95] Kanari L, Dłotko P, Scolamiero M, Levi R, Shillcock J, Hess K and Markram H 2018 *Neuroinformatics* **16** 3–13
[96] Adams H *et al* 2016 Persistence images. Software available at https://github.com/CSU-TDA/PersistenceImages

[97] Adams H, Tausz A and Vejdemo-Johansson M 2014 JAVAPLEX: *A Research Software Package for Persistent (Co)homology* (*Mathematical Software—ICMS 2014*) (*Lecture Notes in Computer Science* vol 8592) ed H Hong and C Yap (Heidelberg: Springer-Verlag) pp 129–36. Software available at http://javaplex.github.io/
[98] Bron C and Kerbosch J 1973 *Commun. ACM* **16** 575–7
[99] Wildmann J 2011 Bron–Kerbosch maximal clique-finding algorithm. Software available at https://www.mathworks.com/matlabcentral/fileexchange/30413-bron-kerbosch-maximal-clique-finding-algorithm
[100] Gan G, Ma C and Wu J 2007 *Data Clustering: Theory, Algorithms, and Applications* (Philadelphia, PA: Society for Industrial and Applied Mathematics)
[101] Porter M A, Onnela J-P and Mucha P J 2009 *Notices of the American Mathematical Society* **56** 1082–97, 1164–66
[102] Fortunato S and Hric D 2016 *Phys. Rep.* **659** 1–44
[103] Jeub L G S, Bazzi M, Jutla I S and Mucha P J 2011–2016 A generalized Louvain method for community detection implemented in MATLAB, *Version 2.0.* Software available at https://github.com/GenLouvain/GenLouvain
[104] Mucha P J, Richardson T, Macon K, Porter M A and Onnela J-P 2010 *Science* **328** 876–8
[105] Blondel V D, Guillaume J-L, Lambiotte R and Lefebvre E 2008 *J. Stat. Mech.* P10008
[106] Bradley P S and Mangasarian O L 1998 Feature selection via concave minimization and support vector machines *Proc. Fifteenth Int. Conf. Mach. Learn. (ICML 1998)* pp 82–90
[107] Zhu J, Rosset S, Hastie T and Tibshirani R 2004 1-norm support vector machines *NIPS '03: 16th Int. Conf. on Neural Information Processing Systems (NIPS)* pp 49–56
[108] Zhang L and Zhou W 2010 *Neural Netw.* **23** 373–85
[109] Chepushtanova S, Gittins C and Kirby M 2014 Band selection in hyperspectral imagery using sparse support vector machines *Proc. SPIE* **9088** 90881F
[110] Patrangenaru V, Bubenik P, Paige R L and Osborne D 2018 (arXiv:1804.10255)
[111] Sepede G *et al* 2010 *NeuroImage* **49** 1080–90
[112] Bassett D S, Wymbs N F, Porter M A, Mucha P J and Grafton S T 2014 *Chaos* **24** 013112
[113] Anderson A and Cohen M S 2013 *Front. Hum. Neurosci.* **7** 520
[114] Braun U *et al* 2016 *Proc. Natl Acad. Sci. USA* **113** 12568–73
[115] Siebenhühner F, Weiss S A, Coppola R, Weinberger D R and Bassett D S 2013 *PLoS One* **8** e72351
[116] Bertolino A *et al* 2009 *J. Neurosci.* **29** 1224–34
[117] Collin G, Kahn R S, de Reus M A, Cahn W and van den Heuvel M P 2014 *Schizophr. Bull.* **40** 438–48
[118] Callicott J H, Egan M F, Mattay V S, Bertolino A, Bone A D, Verchinksi B and Weinberger D R 2003 *Am. J. Psychiatry* **160** 709–19
[119] Guo S, Palaniyappan L, Liddle P F and Feng J 2016 *Psychol. Med.* **46** 2201–14
[120] Adams H and Tausz A 2015 JAVAPLEX tutorial. Available at www.math.colostate.edu/~adams/research/javaplex_tutorial.pdf
[121] Erickson J 2012 Combinatorial optimization of cycles and bases *Advances in Applied and Computational Topology* vol 70 ed A Zomorodian (Providence, RI: American Mathematical Society) pp 195–228
[122] Wang L, Metzak P D, Honer W G and Woodward T S 2010 *J. Neurosci.* **30** 13171–9
[123] Bassett D S, Nelson B G, Mueller B A, Camchong J and Lim K O 2012 *NeuroImage* **59** 2196–207