

Topological Data Analysis of Functional MRI Connectivity in Time and Space Domains

Keri L. Anderson, Jeffrey S. Anderson, Sourabh Palande, and Bei Wang

University of Utah

Abstract. The functional architecture of the brain can be described as a dynamical system where components interact in flexible ways, constrained by physical connections between regions. Using correlation, either in time or in space, as an abstraction of functional connectivity, we analyzed resting state fMRI data from 1003 subjects. We compared several data preprocessing strategies and found that independent component-based nuisance regression outperformed other strategies, with the poorest reproducibility in strategies that include global signal regression. We also found that temporal vs. spatial functional connectivity can encode different aspects of cognition and personality. Topological analyses using persistent homology show that persistence barcodes are significantly correlated to individual differences in cognition and personality, with high reproducibility. Topological data analyses, including approaches to model connectivity in the time domain, are promising tools for representing high-level aspects of cognition, development, and neuropathology.

1 Introduction

Recent advances in the science of complex networks have utilized tools from topological data analysis, in particular, persistent homology [10,29]. This approach investigates connections between constituent parts of networks using powerful and flexible algorithms designed to encode and measure the persistence of relationships across multiple scales. As a discipline, topological data analysis combines algebraic topology and other tools from pure and applied mathematics to support a widening array of applications for studying the architecture of dynamic and complex networks. Although graph-theoretic connectivity, a relatively simple type of topological analysis, has found wide application in functional neuroimaging [23], more advanced uses of topological data analysis are emerging.

We evaluate the performance of graph-theoretic and persistent homology metrics on both time and space formulations of functional connectivity and use the Human Connectome Project dataset [26] to evaluate how the topological architecture of brain networks is related to cognitive function. Since image processing strategies traditionally used for resting state fMRI data have been optimized for conventional types of analysis, we also evaluate the effects of different preprocessing pipelines. Specifically, we measure the reproducibility of graph-theoretic metrics obtained from both time and space dimensions, following several preprocessing pipelines to assess the potential impact of processing strategies on the

ability of graph-theoretic metrics to identify individual differences in connectivity and related cognition, behavior and personality.

Our analyses demonstrate that: (1) Graph-theoretic and topological analyses performed on connectivity across time and space are correlated with distinct aspects of cognitive function. (2) The barcode obtained from the persistent homology of resting state fMRI data is a compact representation of information about individual differences in cognition and personality, with nearly all cognitive metrics tested showing significant correlation to topological features within brain regions known to be substrates for related cognitive functions. (3) Topological metrics show excellent reproducibility when applied to long-duration functional connectivity metrics (1 hour resting state fMRI per subject). Reproducibility was highest for FIX ICA processed data [15] and consistently poorer for preprocessing strategies that include global signal regression [20].

2 Related Work and Technical Background

Dynamic Functional Connectivity. Methods for dynamical analysis of the temporal information in fMRI image data have shown promise for extracting information not available from conventional functional connectivity approaches [18]. Several of these methods involve sliding window approaches [1], where synchronization between brain regions is estimated from short epochs of time. Windowed analyses, however, obtain estimates of connectivity from short time series, resulting in limited accuracy of individual measurements.

An additional approach to component analysis includes identifying point-process temporal co-activation patterns [17,5] by clustering time points that exhibit similar activation across the brain and examining temporal structure within the relative sequence of activation of these co-activation patterns. Point-process methods such as temporal co-activation patterns have been analyzed by clustering timepoints together with similar patterns of relative activation across the brain into composite spatial nodes or networks that show hierarchical similarity structure to each other with architecture that combines features of more familiar intrinsic connectivity networks [17,5]. Nevertheless, there is a symmetry where an fMRI time series can arbitrarily be seen as connectivity between time points across the brain, where the individual time points become the nodes in a graph, with edges reflected by spatial similarity across the brain.

Topological Data Analysis (TDA) of Networks. TDA [3,13] of networks goes beyond graph-theoretic analysis by utilizing tools from computational topology to describe the architecture of networks or data structures in more flexible ways. In particular, it encodes higher order (not just pairwise) interactions in the system and studies topological features of the brain network across all possible thresholds. Persistent homology, a main ingredient in topological data analysis, is an emerging tool in studying complex networks, including, for instance, collaboration [4] and brain networks [19]. Topological methods have also shown

promise in modeling transitions between brain states in functional imaging data using combined information in space and time [24].

Persistent Homology and Barcodes. Persistent homology studies the topological features of a point cloud at multiple scales; see [10] for seminal work on the topic and [3,13] for excellent surveys. We follow the illustrative example of persistent homology and metric space mapping in [27].

As illustrated in Fig. 1, we begin with a point cloud P , equipped with a (Euclidean) distance metric. For some $t \geq 0$, the union of balls of radius t , centered at the points of P , forms a topological space. As the radius t increases, we get a nested sequence of spaces referred to as a *filtration*. The radius t , which parametrizes the spaces in such filtration, is often viewed as time. Using persistent homology, we investigate the evolution in time of the topological features of spaces in the filtration.

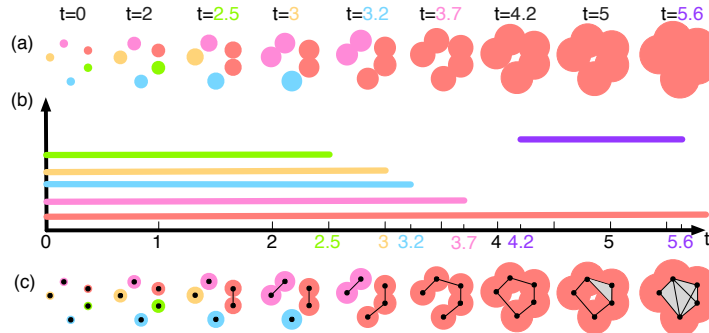


Fig. 1: Computing the persistent homology of a point cloud (image adapted from [27], Fig. 1).

As t increases, we focus on the important events when the topology of the space changes. This change occurs, for example, when components merge with one another to form larger components or tunnels. We track the birth and death times of each topological feature (a component or a tunnel). The lifetime of a feature in the filtration is called its *persistence*. In Fig. 1(a), at time $t = 0$, each colored point is *born* (appears) as an independent (connected) component. At $t = 2.5$, the green component merges into the red component and *dies* (disappears). Therefore, the green component has a persistence of 2.5. At $t = 3$, the orange component merges into the pink component and dies. Hence, it has a persistence of 3. Similarly, the blue component dies at $t = 3.2$ and the pink component dies at $t = 3.7$. At time $t = 4.2$, the collection of components forms a tunnel which has a persistence of 1.4 and disappears at $t = 5.6$. The red component born at time 0 never dies and thus it has a persistence of ∞ . In Fig. 1(b), we visualize the appearance (birth), the disappearance (death) and the persistence of these topological features in the filtration via the *barcode* [13], where each feature is summarized by a horizontal bar that begins at its birth time and ends at

its death time. Computationally, the above nested sequence of spaces can be combinatorially represented by a nested sequence of simplicial complexes (i.e., collections of vertices, edges and triangles) with a much smaller footprint, as illustrated in Fig. 1(c); see [9] for computational details.

3 Methods

Data Sources. Resting state fMRI data from 1003 participants (534 female, mean age = 29.45 ± 3.61 (SD); 469 male, mean age 27.87 ± 3.65) out of the 1200 Subjects Data Release of the Human Connectome Project (HCP) [26] were analyzed. The current study utilized both minimally preprocessed and FIX ICA cleaned BOLD resting-state data [15] acquired over four 15-minute multiband BOLD resting-state scans over 2 days. Only subjects completing all four resting-state scans are included in this analysis. The first 20 volumes of each run were excluded, yielding 1180 timepoints.

Region of Interest (ROI) Selection. Gray matter regions of interest consisted of 333 regions in the cerebral cortex [14], 14 subject-specific subcortical regions from FreeSurfer derived segmentation [12] (bilateral thalamus, caudate, putamen, amygdala, hippocampus, pallidum and nucleus accumbens) and 14 bilateral cerebellar representations of a 7-network parcellation [28]. This combined parcellation scheme incorporates gray matter ROIs totalling 361 regions.

Image Processing. A time series for each scan in each subject was extracted from FIX ICA cleaned and minimally preprocessed BOLD data. The minimally preprocessed BOLD data were also analyzed with head motion, white matter and CSF regression and these regressors plus global signal regression [25].

Functional Connectivity Calculation. Functional connectivity was calculated in both time and space domains (Supplement, Fig. 5). For the space domain, a 361×361 matrix was computed for each scan representing the Pearson correlation coefficient between the time series for each pair of gray matter regions. For the time domain, the matrix of 361×1180 time series was transposed and the correlation coefficient was analogously calculated between 1180×1180 pairs of timepoints. Given the large number of intercorrelated node pairs, full correlation was used because of the potential instability of partial correlation results.

Graph-Theoretic Metrics. We selected four graph-theoretic metrics for comparison of reproducibility of results across preprocessing strategies and between time and space connectivity measurements: modularity, characteristic path length, global efficiency and clustering coefficient. These were computed using the Brain Connectivity Toolbox software for Matlab [23].

Reproducibility Metrics. As a metric of reproducibility, the intraclass correlation coefficient (ICC) was calculated using the ICC.m function for Matlab using 1-k parameter across four scans for each of 1003 subjects for each measurement. This represents the expected ICC value that would be obtained for

four scans per subject (1 hour of resting state fMRI data). To interpret the reproducibility of an ICC score, we used the following guidelines – Poor - less than 0.4; Fair 0.4-0.59; Good 0.6-0.74; Excellent 0.74-1 – according to [6].

Persistent Homology Analysis. To apply persistent homology to brain networks, we map a given brain network to a point cloud in a metric space, where network nodes map to points and the measures of association between pairs of nodes map to distances between pairs of points [27]. In this paper, the association between two nodes u, v in the brain network is measured by their correlation coefficients $\text{corr}(u, v)$. The idea is to map this association to a distance measure such that higher correlations between nodes map to smaller distances. We use the mapping $d(u, v) = \sqrt{1 - \text{corr}(u, v)}$. Subsequently, a nested sequence of Vietoris-Rips complexes (a type of simplicial complex) is constructed in the metric space for persistent homology computation. Dimension 0 persistence barcodes were calculated with the R toolbox package TDA [11]. Specifically, connectivity matrices were converted into normalized distance matrices and used directly as input into function `ripsDiag` from the TDA library.

Cognitive and Personality Variables. Subject-level cognitive and personality scores used in this study contained scores from the 12 cognition domain measures included in the HCP battery of behavioral and individual difference measures - Cognition Domain [2] and 5 factor-level scores for personality from the NEO Five Factor Inventory (NEO FFI) [7]. We use corrected scoring of Agreeableness factor rather than the initial data supplied with the 1200 subjects release of the Human Connectome Project dataset [8]. Specific cognitive measures included: Episodic Memory (Picture Sequence Memory), Executive Function / Cognitive Flexibility (Dimensional Change Card Sort), Executive Function / Inhibition (Flanker Inhibitory Control and Attention Task), Fluid Intelligence (Penn Progressive Matrices), Language / Reading Decoding (Oral Reading Recognition), Language / Vocabulary Comprehension (Picture Vocabulary), Processing Speed (Pattern Comparison Processing Speed), Self-regulation / Impulsivity (Delay Discounting), Spatial Orientation (Variable Short Penn Line Orientation Test), Sustained Attention (Short Penn Continuous Performance Test), Verbal Episodic Memory (Penn Word Memory Test), Working Memory (List Sorting).

4 Results

Effects of Image Preprocessing on Reproducibility of Functional Connectivity. Functional connectivity was calculated by computing Pearson correlation coefficients for each pair of fMRI time series for 1003 subjects, each with four 15 minute scans, using 4 preprocessing strategies for each pair of 361 gray matter nodes. For each preprocessing strategy, reproducibility was calculated by computing the intraclass correlation coefficient (ICC) for each pair of gray matter regions, obtained from 1003 subjects and 4 scans per subject (Supplement, Fig. 4). Statistical analysis using 1-way ANOVA demonstrated that each

set of ICC values was significantly different for all four preprocessing strategies ($F < 1.0e^{-7}$), with markedly higher ICC values for Independent Component nuisance regression and lowest ICC values for preprocessing including global signal regression. For ICA-based nuisance regression, ICC was excellent (> 0.7) for almost all connections but weaker for connections involving the subcortex [25].

Reproducibility was also calculated for four graph-theoretic metrics and time and space topological persistent homology measures. For graph-theoretic and topological metrics obtained by functional connectivity over space and time, reproducibility was highest for independent component-based nuisance regression, shown in Fig. 2. For persistent homology measures in the spatial domain (Fig. 2, top right), the difference was striking, with markedly improved reproducibility using ICA-based regression compared to more limited preprocessing strategies, and the weakest reproducibility obtained when including global signal regression. Taken together, these results provided strong evidence that for functional connectivity, whether computed over space or time, and including both graph-theoretic and topological measures computed from functional connectivity, the highest reproducibility was obtained using ICA-based nuisance regression. For all remaining analyses in this report, we used only FIX ICA cleaned data.

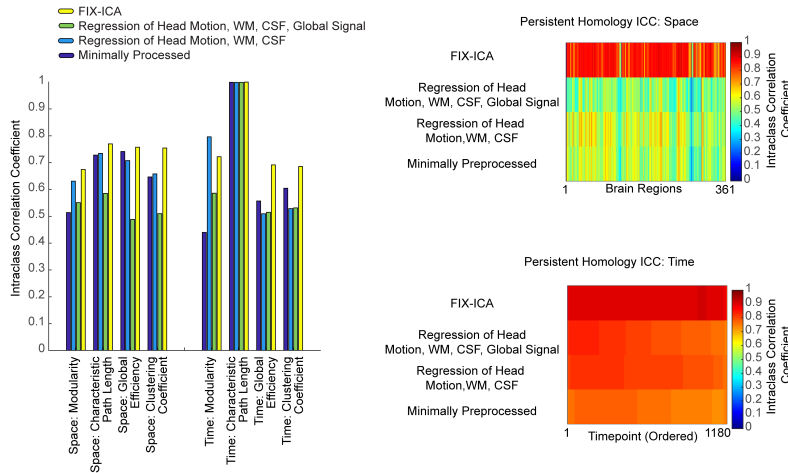


Fig. 2: Reproducibility of graph-theoretic and topological measures of functional connectivity by preprocessing strategy.

Differences in Results for Space and Time Connectivity. Reproducibility, as shown in Fig. 2, left, shows a comparable intraclass correlation for graph-theoretic measures obtained from connectivity with nodes representing spatial regions and with nodes representing timepoints. Reproducibility for global efficiency in time was higher than for other measures, which were all about 0.7.

Yet, when these graph-theoretic measures were compared across subjects to scores on 12 cognitive tests and 5 personality factors, only functional connectivity

obtained from the time domain (using timepoints as nodes and correlation across spatial regions as edges) showed significant partial correlation to cognitive tests. Partial correlations were computed for each cognitive and personality metric separately, with age, sex and mean head motion used as subject-level covariates in each case, false discovery rate corrected. In particular, modularity in the time connectivity graphs was correlated with inhibitory components of executive function, and the characteristic path length and median clustering coefficient in the time connectivity graphs were correlated with vocabulary and processing speed (Supplement, Fig. 6).

Differences in Reproducibility and Behavioral Correlations for Persistent Homology. Persistence barcodes were calculated for connectivity graphs in the time (correlation across spatial ROIs) and space (correlation across timepoints) domains by calculating the connectivity distance at which each node merged with another cluster as described in section 3. Each barcode consisted of a vector of 361 elements (space domain) or 1180 elements (time domain). Barcodes were reordered by persistence (length) for display in Fig. 3, left. Re-ordered barcodes were also used for analysis in the time domain, because timepoints are arbitrary and convey no consistent meaning across subjects or scans. In the space domain, barcodes were used without reordering, as each element of a barcode corresponds to a preserved brain region across subjects and scans.

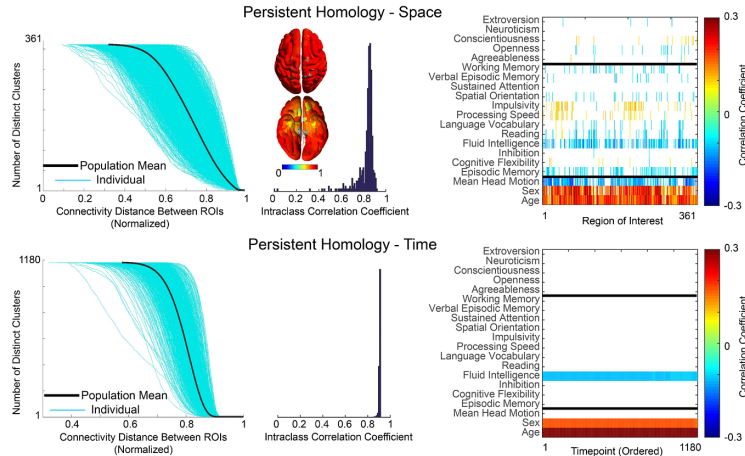


Fig. 3: Reproducibility and cognitive correlation of persistent homology barcode analyses.

The reproducibility of barcode results is shown in Fig. 3, center panels for space and time domains, yielding excellent (> 0.7) ICC values for almost all elements. This is shown graphically on a template brain for the space domain as an inset in the figure, demonstrating that regions of lower ICC are exclusively in the medial orbitofrontal and medial anterior temporal regions, areas that are

in close to brain/bone interfaces and are known to represent regions of high susceptibility artifact in the fMRI BOLD signal.

When comparing cognitive and personality metrics to persistence barcodes, also with partial correlation with age, sex and head motion as covariates, there were significant corrected correlations between persistent homology and fluid intelligence in both time and space domains, with less sensitivity to head motion when calculated in the time domain. For the spatial domain, fluid intelligence was correlated with barcode values in brain regions comprising association cortex of the frontal, parietal and temporal lobes, with weaker correlations in sensory and motor regions. Of 12 cognitive tests performed, 11 showed significant partial correlation with persistent homology in the space domain after correction for multiple comparisons, and of five personality factors, four showed significant corrected partial correlation with persistent homology in the space domain (Supplement, Fig. 8 to Fig. 21, spatial distribution of correlation between persistent homology barcodes and specific cognitive measures including episodic memory, cognitive flexibility, agreeableness, openness, etc.)

5 Discussion

Topological data analysis of resting state functional connectivity using persistence barcodes identified individual differences in cognition and personality in the Human Connectome Project sample. Whether examining functional connectivity in the time or space domain, fluid intelligence was predicted by persistence barcode values, and distinct patterns of spatial regions were significantly correlated with a wide array of cognitive performance scores. Persistent homology showed excellent reproducibility across scans for the same individual for functional connectivity over both space and time in the brain. Graph-theoretic values also demonstrated good to excellent reproducibility, with functional connectivity in time and space domains correlated with distinct aspects of cognitive performance. For all tests, reproducibility was highest when using a robust independent component analysis nuisance regression strategy with a less connected graph than for other cleaning pipelines, suggesting that the spurious artifactual correlation between brain regions had been reduced. Reproducibility was lowest when using a strategy that includes global signal regression, possibly representing effects of contaminating results by incorporating information from other brain regions [20].

Not only were persistence barcodes significantly correlated with a surprising number of cognitive and personality features, but the spatial distributions of regions correlated with each behavior were also informative. Brain regions showing a correlation between barcode values and fluid intelligence were located in the association cortex, particularly in frontoparietal attentional regions, but not the sensory and motor cortex (Supplement, Fig. 7). These frontoparietal regions have been favored in the literature as substrates for general intelligence [16].

Similarly, correlations between persistent homology and working memory were observed in the ventral attention network and prefrontal cortex, the core

neural substrates associated with pattern recognition, working memory and focused attention. A correlation between persistent homology and spatial attention specifically identified right hemispheric frontoparietal regions, consistent with well-known right dominance of lesions contributing to hemispatial neglect and lateralization of brain function to the right for spatial attention [21]. Both reading and language vocabulary scores were correlated with areas of the posterior temporal lobe associated with the Wernicke Area. Both episodic memory and verbal episodic memory scores were significantly correlated with areas of the posterior cingulate and medial temporal lobe, critical regions well known for their involvement in memory recall. Agreeableness was significantly correlated with persistent homology in the superior temporal sulcus, a core region of the social brain related to social empathy [22].

Although less intuitive than traditional functional connectivity between brain regions, our results suggest that functional connectivity between timepoints may offer new insights into aspects of cognition and neuropathology. Persistent homology, including potential higher dimensional topological features, may represent distinct aspects of brain function. These approaches may reflect dynamical aspects of connectivity such as the temporal duration and frequency of brain microstates or oscillations between metastable patterns of relative brain activity and provide new insights into brain network architecture or opportunities for the prediction of behavioral traits.

Acknowledgements. This work was supported by NIH grants R01EB022876, R01MH080826, and NSF grant IIS-1513616.

References

1. Allen, E., et al.: Tracking whole-brain connectivity dynamics in the resting state. *Cerebral Cortex* **24**(3), 663–676 (2009)
2. Barch, D., et al.: Function in the human connectome: task-fMRI and individual differences in behavior. *NeuroImage* **80**, 169–189 (2013)
3. Carlsson, G.: Topology and data. *Bulletin of the American Mathematical Society* **46**, 255–308 (2009)
4. Carstens, C.J., Horadam, K.J.: Persistent homology of collaboration networks. *Mathematical Problems in Engineering* **2013** (2013)
5. Chen, J.E., Chang, C., Greicius, M.D., Glover, G.H.: Introducing co-activation pattern metrics to quantify spontaneous brain network dynamics. *NeuroImage* **111**, 476–488 (2015)
6. Cicchetti, D.: Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment* **6**(4), 284–290 (1994)
7. Costa, P.T., McCrae, R.R.: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) manual. *Psychological Assessment Resources* (1992)
8. Dubois, J., Galdi, P., Paul, L.K., Adolphs, R.: A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *bioRxiv* (2018). <https://doi.org/10.1101/257865>

9. Edelsbrunner, H., Harer, J.: *Computational Topology: An Introduction*. American Mathematical Society, Providence, RI, USA (2010)
10. Edelsbrunner, H., Letscher, D., Zomorodian, A.J.: Topological persistence and simplification. *Discrete and Computational Geometry* **28**(4), 511–533 (2002)
11. Fasy, B.T., Kim, J., Lecci, F., Maria, C.: Introduction to the R package TDA. ArXiv: 1411.1830 (2015)
12. Fischl, B., et al.: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**(3), 341–355 (2002)
13. Ghrist, R.: Barcodes: The persistent topology of data. *Bulletin of the American Mathematical Society* **45**, 61–75 (2008)
14. Gordon, E., et al.: Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cerebral Cortex* **16**(1), 288–303 (2016)
15. Griffanti, L., et al.: ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *Neuroimage* **95**, 232–247 (2014)
16. Jung, R., Haier, R.: The parieto-frontal integration theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behavioral and Brain Sciences* **30**(2), 135–154 (2007)
17. Karahanoglu, F., Van De Ville, D.: Transient brain activity disentangles fMRI resting-state dynamics in terms of spatially and temporally overlapping networks. *Nature Communications* **6**, 7751 (2015)
18. Keilholz, S., Caballero-Gaudes, C., Bandettini, P., Deco, G., Calhoun, V.: Time-resolved resting-state functional magnetic resonance imaging analysis: Current status, challenges, and new directions. *Brain Connectivity* **7**(8), 465–481 (2017)
19. Lee, H., Chung, M.K., Kang, H., Kim, B.N., Lee, D.S.: Discriminative persistent homology of brain networks. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro* pp. 841–844 (2011)
20. Murphy, K., Fox, M.D.: Towards a consensus regarding global signal regression for resting state functional connectivity MRI. *NeuroImage* **154**, 169–173 (2017)
21. Nielsen, J.A., Zielinski, B.A., Ferguson, M.A., Lainhart, J.E., Anderson, J.S.: An evaluation of the left-brain vs. right-brain hypothesis with resting state functional connectivity magnetic resonance imaging. *PLoS One* **8**(8), e71275 (2013)
22. Pelphrey, K., Morris, J., McCarthy, G.: Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *Journal of Cognitive Neuroscience* **16**(10), 1706–1716 (2004)
23. Rubinov, M., Sporns, O.: Complex network measures of brain connectivity: uses and interpretations. *NeuroImage* **52**(3), 1059–1069 (2010)
24. Sagar, M., et al.: Towards a new approach to reveal dynamical organization of the brain using topological data analysis. *Nature Communications* **9**(1), 1399 (2018)
25. Shah, L.M., Cramer, J.A., Ferguson, M.A., Birn, R.M., Anderson, J.S.: Reliability and reproducibility of individual differences in functional connectivity acquired during task and resting state. *Brain and Behavior* **6**(5), e00456 (2016)
26. Van Essen, D., et al.: The WU-Minn human connectome project: an overview. *Neuroimage* **80**, 62–79 (2013)
27. Wong, E., et al.: Kernel partial least squares regression for relating functional brain network topology to clinical measures of behavior. *International Symposium on Biomedical Imaging* (2016)
28. Yeo, B., et al.: The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology* **106**(3), 1125–1165 (2011)
29. Zomorodian, A., Carlsson, G.: Computing persistent homology. *Discrete and Computational Geometry* **33**(2), 249–274 (2005)