Persistent Homology Distances for Comparing Disease-Filtered Structural Connectomes

Hamza Chaudhry* HCHAUDHRY@G.HARVARD.EDU
Harvard University

Vineet Tiruvadi* VTIRUVADI@BWH.HARVARD.EDU
Harvard Medical School

Cengiz Pehlevan CPEHLEVAN@SEAS.HARVARD.EDU
Harvard University

Michael D. Fox MDFOX@BWH.HARVARD.EDU
Harvard Medical School

Kanaka Rajan KANAKA_RAJAN@HMS.HARVARD.EDU
Harvard Medical School

Abstract

Whole-brain circuit mapping yields wiring diagrams of axon bundles, or *connectomes*, that improve both scientific understanding and clinical care of brain disorders. However, the lack of direct, objective metrics to compare disease-relevant connectomes potentially hinders progress. Here, we introduce early steps towards building such metrics with a parcellation-free, geometry-aware method for analyzing disease-specific streamline bundles using persistent homology and distributions within bundles. We use persistent homology for multiscale comparison of streamlines in bundles with Wasserstein distance distributions to characterize topological distances in disease-specific, or *filtered*, connectomes. We observed hierarchical similarity hemisphere < disease < connectome. This measure may serve as a foundation to guide costly next-generation connectome generation while also optimizing disease-specific connectome *atlases* for neural implant design, implantation, and programming.

Keywords: Structural Connectomics, Diffusion MRI, Deep Brain Stimulation, Persistent Homology, Neural Implants, Neuroengineering

1. Introduction

Detailed maps of the brain are improving our ability to treat neurological and psychiatric disorders (Smith and Lee, 2022; Horn, 2021; Riva-Posse et al., 2018) (1a). Mapping axon bundles with diffusion MRI (dMRI) has yielded wiring diagrams that promise to explain disease (Doe and Wang, 2021), improve medical therapy (Anderson and Chen, 2021), and advance neural implant engineering more broadly (Kumar and Miller, 2022). Efforts to collect larger-scale *connectomes* and isolate disease-filtered circuits (Figure 1b), or *subatlases*, are growing but expensive - their marginal utility and overall cost-benefit are difficult to assess objectively (Johnson and Patel, 2021). Current approaches to comparing connectomes assume improved acquisition resolution is better, or rely on indirect assessments through noisy clinical predictions (Hollunder et al., 2024; Mehta et al., 2025; Johnson and Patel, 2021; Thompson and Rodriguez, 2021). More objective metrics are needed, and topological data analyses (TDA) provide approaches that are particularly robust to the sources of

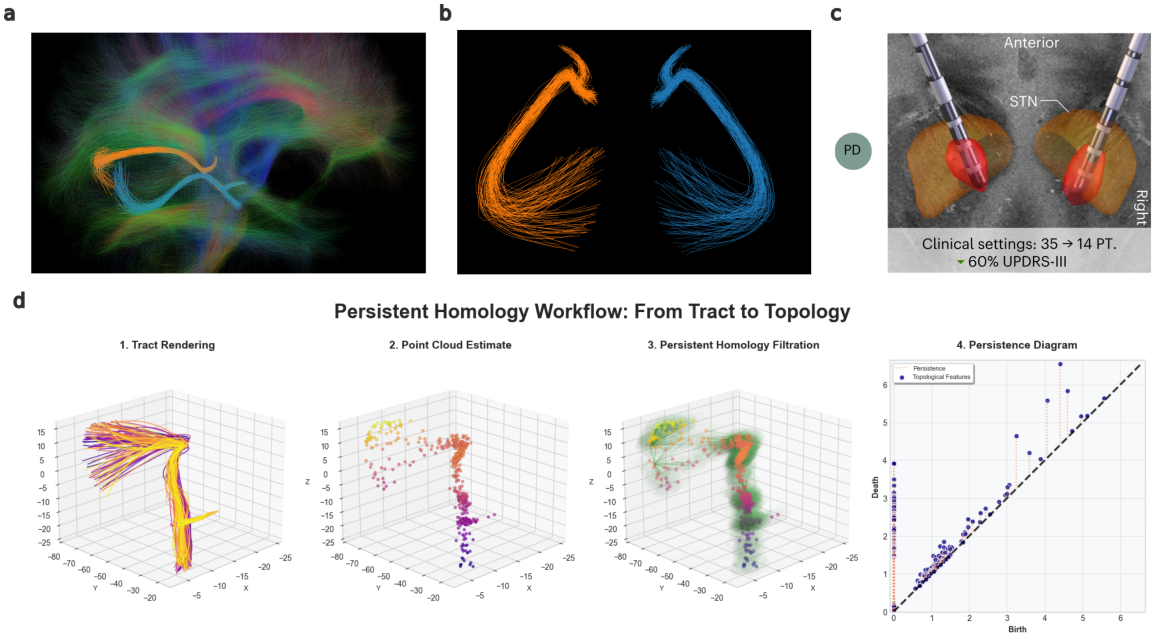

Figure 1: **Disease Subatlas Targeting and Workflow** a) Schematic of whole brain connectome (transparent) and isolated disease-filtered circuit (blue - left, orange - right). b) Similarities and differences in fibers of subatlas are visualized, but hard to quantify (above). Different connectomes will filter into different subatlases, but whether these differences matter when doing DBS targeting (below) is unclear. c) Methods workflow for this paper. (1) Rendering of top 100 streamlines of Petersen PD subatlas. (2) Point cloud sampling of streamlines. (3) Rips filtration of point cloud showing green epsilon balls around each point, drawing simplicial connections between intersecting balls. (4) Persistence diagram showing birth-death times of topological features, where distance from diagonal indicates feature significance.

noise intrinsic to dMRI-based tractography (Bhattacharya et al., 2024). This is especially important as distinct connectomes implicate distinct fibers for therapeutic neural implants, like *deep brain stimulation* (DBS; Figure 1c).

Inspired by previous works in topological data analysis (Petri et al., 2014), we develop a toolset to compute distance-like metrics on disease-specific streamline sets using topological data analysis (TDA), with the goal of providing meaningful distributions and summary statistics that reflect structural shape and topology. We investigate whether persistence summaries are stable within a connectome, whether they detect disease contrasts, and how symmetric disease-relevant streamlines are across hemispheres.

2. Methods

2.1. Data and Preprocessing

Template white matter tracts for HCPx were imported from the dipy data repository (). Human subjects data was derived from publicly available structural connectomes (Petersen,

MGH, HCP) (Petersen et al., 2019; Wang et al., 2021; Van Essen et al., 2013). Briefly, the Petersen is an expert-anatomist drawn connectome. MGH and HCP are dMRI derived, with MGH coming from a single human subject, and HCP averaged from 985 subjects.

2.2. Fiber Filtering

This method was used to perform fiber filtering on Parkinson's Disease (PD) and Dystonia (DYT) DBS clinical improvement data (Hollunder et al., 2024; Horn et al., 2015; Horn, 2021). Briefly, this involved correlating the presence of a streamline inside a simulated *volume of tissue activated* under DBS with improvement to clinical scores reflecting disease state. In some analyses, the *volume of tissue activated* (VTA) at each hemisphere's DBS site is mirrored to the other side and used for fiber filtering, yielding a more symmetric subatlas for cross-hemisphere comparison.

2.3. Streamline Preprocessing

The .trk files for the top 100 predictive streamlines from fiber filtering—the **disease subatlas**—were loaded and rendered through the DIPY library (Garyfallidis et al., 2014), with each connectome represented as a set of streamlines in the MNI common 3D coordinate system. Only the first 100 streamlines are analysed per bundle/subatlas to avoid cardinality effects and ensure computational tractability. These streamlines are then converted to point clouds and systematically downsampled (every 100th point) to balance topological fidelity with computational efficiency.

2.4. Persistent Homology Computation

We compute persistent homology on the point clouds using a Vietoris–Rips (VR) filtration under the Euclidean metric, implemented through the GUDHI library (Carlsson, 2009; Maria et al., 2014). The Rips complex is constructed up to homology dimension 2, capturing connected components (H_0), loops (H_1), and voids (H_2) (Figure 1d). Persistence diagrams are computed from the resulting simplex tree, encoding the birth and death scales of topological features across the filtration (Edelsbrunner et al., 2002). For robustness assessment, we employ a jackknife resampling strategy (leave one out, no replacement) over streamlines: 100 calculations over all N sets of (N-1) streamlines (Efron, 1992). A persistence diagram is computed for each jackknife subset, yielding a distribution reflective of bundle properties and potential outliers.

2.5. Distance, Distributions, and Comparisons

Topological differences between persistence diagrams are quantified using the 2-Wasserstein distance with standard diagonal matching, implemented via the GUDHI Wasserstein module (Maria et al., 2014; Skraba and Turner, 2020). Distances are computed on the full set of birth–death pairs aggregated across homology dimensions (no dimension-wise separation and no explicit removal of infinite-death pairs). For each condition, we compute all pairwise 2-Wasserstein distances among the 100 jackknife diagrams, producing $\binom{100}{2} = 4{,}950$ intra-tract distances that characterize within-condition variability. Summary distances are

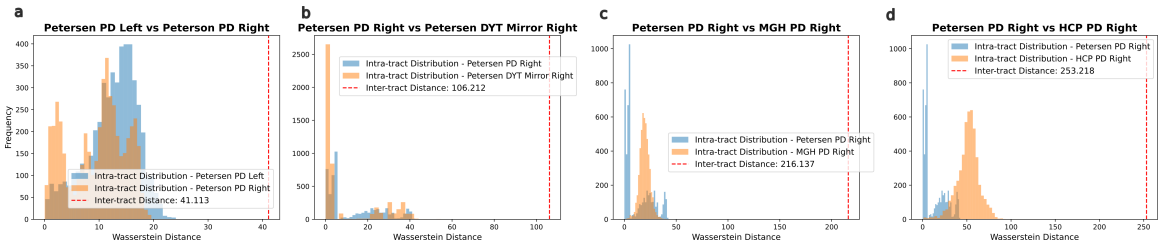

Figure 2: **Comparison Distributions** a) Comparison of PD subatlases in left and right hemispheres. b) Comparison of right hemispheres for PD and DYT subatlases. c) Petersen PD Right vs MGH PD Right. d) Petersen PD Right vs HCP PD Right . Dashed red lines mark single-run cross-tract distance.

computed between full (all-100 streamline) persistence diagrams using the 1-in-50 decimation: inter-hemisphere distances compare PD Left and PD Right within a connectome, inter-disease distances compare PD and DYT within the same site, and inter-connectome distances compare identical conditions across different sites (e.g., Petersen vs MGH vs HCP). Bundle size (100 streamlines) and filtration settings are held constant across datasets.

3. Results

Across all comparisons, we observe a consistent ordering of topological separations: hemisphere < disease < connectome. Within Petersen, the inter-hemisphere distance between PD Right and PD Left is modest (**41**). Disease contrasts within the same connectome are larger (Petersen PD Right vs Petersen DYT Mirror Right = **106**). As seen in Figure 2, the largest gaps arise when holding disease and hemisphere fixed but changing connectome identity (Petersen PD Right vs MGH PD Right = **216**; Petersen PD Right vs HCP PD Right = **253**). In Figure 5, Hemisphere checks show larger left–right distances in HCP (**131**) and MGH (**155**), yet these remain well below the cross-connectome distance (**222**) in Figure 6.

4. Discussion

Our persistence-based approach to comparing disease-specific connectomes reveals a consistent ordering of topological separations across datasets and comparison types: hemisphere < disease < connectome, indicating stability within a site and sensitivity to disease, while cross-site acquisition/reconstruction imposes dominant fingerprints.

Practically, topological distances offer a parcellation-free, geometry-aware metric for evaluating disease-targeted bundles and methodological choices. Within a single connectome, small hemisphere differences and tight jackknife distributions support using PH-derived metrics as internal quality control for acquisition parameters, tractography, and DBS seeding. In contrast, the marked cross-connectome separations caution against naïve pooling or benchmarking without harmonization. This dominance of acquisition effects over disease signals implies that unharmonized multi-site analyses risk characterizing scanner artifacts rather than the underlying pathology. For DBS optimization, stable within-site

topology summaries can validate disease subatlases and compare candidate targets or contacts under a common metric.

Limitations include our use of downsampling for tractability, which may obscure fine-scale features. Vietoris–Rips filtrations may be sensitive to subsampling density and numerical tolerances, with variability across implementation libraries and subsampling density.Leakage is possible if disease-informed streamline selection is not strictly segregated from evaluation.

Further work will build richer representations (e.g., persistence landscapes and higher-order homology (Bubenik, 2015)) and more stable measures (e.g. Rigorous comparison of libraries and alternatives to VR filtration - including alpha complexes, witness complexes, and flood complexes (Graf et al., 2025)). Extended analysis with systematic ablations, resampling, filtration, and Wasserstein orders will be guided by clinical utility and feedback. Linking to previously reported clinical outcomes is planned Hollunder et al. (2024) along with ANN classifiers for clinical tasks and identification of informative PH features. Further validation with clinical studies will be critical in calibrating our metrics and identifying clinically optimal connectomes and/or subatlases.

References

D Vijay Anand, Zhenyu Meng, Kelin Xia, and Yuguang Mu. Weighted persistent homology for osmolyte molecular aggregation and hydrogen-bonding network analysis. *Scientific reports*, 10(1):9685, 2020.

Peter Anderson and Li Chen. Structural connectivity predicts deep brain stimulation outcomes. *Brain*, 144(2):356–370, 2021.

Debanjali Bhattacharya, Ninad Aithal, Manish Jayswal, and Neelam Sinha. Analyzing brain tumor connectomics using graphs and persistent homology. In *International Workshop on Topology-and Graph-Informed Imaging Informatics*, pages 33–42. Springer, 2024.

Peter Bubenik. Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research*, 16(3):77–102, 2015. Submitted July 2014; published January 2015.

Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46 (2):255–308, 2009.

Jane Doe and Michael Wang. Connectomics and the mechanistic understanding of brain disorders. *Neuron*, 109(9):1452–1470, 2021.

Edelsbrunner, Letscher, and Zomorodian. Topological persistence and simplification. *Discrete & computational geometry*, 28(4):511–533, 2002.

Bradley Efron. Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics: Methodology and distribution*, pages 569–593. Springer, 1992.

Eleftherios Garyfallidis, Matthew Brett, Bagrat Amirbekian, Ariel Rokem, Stefan Van Der Walt, Maxime Descoteaux, Ian Nimmo-Smith, and Dipy Contributors. Dipy, a library for the analysis of diffusion mri data. *Frontiers in neuroinformatics*, 8:8, 2014.

Florian Graf, Paolo Pellizzoni, Martin Uray, Stefan Huber, and Roland Kwitt. The flood complex: Large-scale persistent homology on millions of points. *arXiv preprint arXiv:2509.22432*, 2025.

Barbara Hollunder, Jill L Ostrem, Ilkem Aysu Sahin, Nanditha Rajamani, Simón Oxenford, Konstantin Butenko, Clemens Neudorfer, Pablo Reinhardt, Patricia Zvarova, Mircea Polosan, et al. Mapping dysfunctional circuits in the frontal cortex using deep brain stimulation. *Nature Neuroscience*, 27(3):573–586, 2024.

Andreas Horn. *Connectomic deep brain stimulation*. Academic Press, 2021.

Andreas Horn, Martin Reich, Johannes Vorwerk, Nan Li, Gregor Wenzel, Qianqian Fang, Tanja Schmitz-Hübsch, Rainer Nickl, Andreas Kupsch, Jens Volkmann, et al. Lead-dbs: A toolbox for deep brain stimulation electrode localization and connectivity. *NeuroImage*, 107:127–135, 2015.

Emily Johnson and Arjun Patel. Large-scale efforts in human connectomics: Progress and pitfalls. *Science*, 374(6573):89–95, 2021.

Rajesh Kumar and Sarah Miller. Engineering neural implants guided by structural connectivity. *Nature Biomedical Engineering*, 6(7):711–725, 2022.

Clément Maria, Jean-Daniel Boissonnat, Marc Glisse, and Mariette Yvinec. The gudhi library: Simplicial complexes and persistent homology. In *International Congress on Mathematical Software*, pages 167–174. Springer, 2014.

Ketan Mehta, Angela M Noecker, and Cameron C McIntyre. Comparison of structural connectomes for modeling deep brain stimulation pathway activation. *NeuroImage*, 312: 121211, 2025.

Mikkel V Petersen, Jeffrey Mlakar, Suzanne N Haber, Martin Parent, Yoland Smith, Peter L Strick, Mark A Griswold, and Cameron C McIntyre. Holographic reconstruction of axonal pathways in the human brain. *Neuron*, 104(6):1056–1064, 2019.

Giovanni Petri, Paul Expert, Federico Turkheimer, Robin Carhart-Harris, David Nutt, Peter J Hellyer, and Francesco Vaccarino. Homological scaffolds of brain functional networks. *Journal of The Royal Society Interface*, 11(101):20140873, 2014.

Patricio Riva-Posse, KS Choi, Paul E Holtzheimer, Andrea L Crowell, Steven J Garlow, Justin K Rajendra, Cameron C McIntyre, Robert E Gross, and Helen S Mayberg. A connectomic approach for subcallosal cingulate deep brain stimulation surgery: prospective targeting in treatment-resistant depression. *Molecular psychiatry*, 23(4):843–849, 2018.

Primoz Skraba and Katharine Turner. Wasserstein stability for persistence diagrams. *arXiv preprint arXiv:2006.16824*, 2020.

John Smith and Anna Lee. Therapeutic applications of connectomics in psychiatry and neurology. *Nature Reviews Neurology*, 18(4):200–212, 2022.

Alice Thompson and Miguel Rodriguez. Challenges and critiques of diffusion mri connectome reconstructions. *NeuroImage*, 245:118716, 2021.

David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy EJ Behrens, Essa Yacoub, Kamil Ugurbil, Wu-Minn HCP Consortium, et al. The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79, 2013.

Fuyixue Wang, Zijing Dong, Qiyuan Tian, Congyu Liao, Qiuyun Fan, W Scott Hoge, Boris Keil, Jonathan R Polimeni, Lawrence L Wald, Susie Y Huang, et al. In vivo human whole-brain connectom diffusion mri dataset at 760 μm isotropic resolution. *Scientific data*, 8(1):122, 2021.

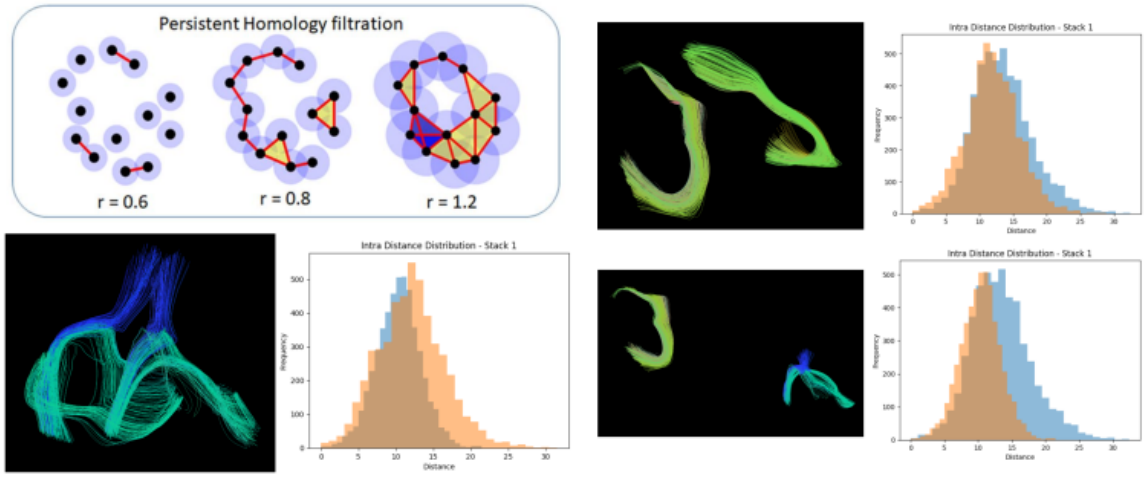

Figure 3: **Persistent Homology for Template Fibers**. Depiction of persistent homology filtration at distinct radii. Adapted from (Anand et al., 2020). Example distributions from various tracts, both matches across hemispheres (top right, bottom left) and across tracts (bottom right). Distributions represent intra-tract variability, while difference in distribution moments align with tract identity.

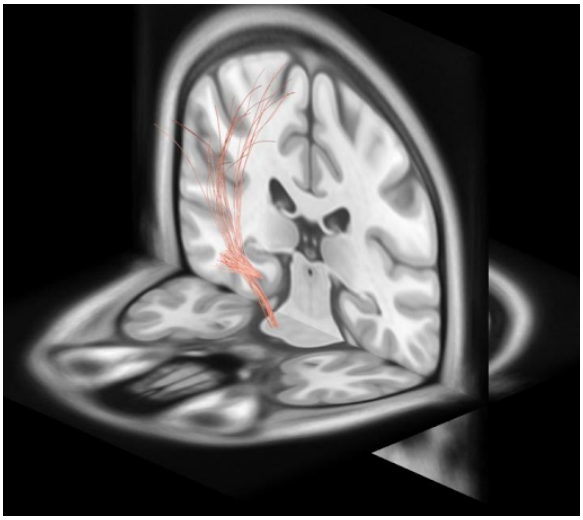

Figure 4: Rendering of Petersen PD Top 100 streamlines in Lead-DBS (Horn et al., 2015).

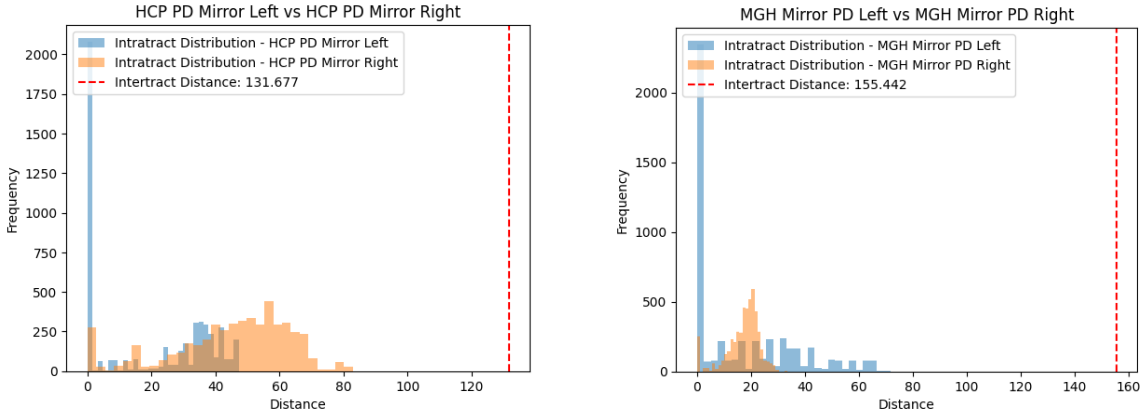

Figure 5: **Cross-Hemisphere Comparisons.** Left: Comparison of PD subatlases in left and right hemispheres of HCP (**131**). Right: Comparison of PD subatlases in left and right hemispheres of MGH (**155**).

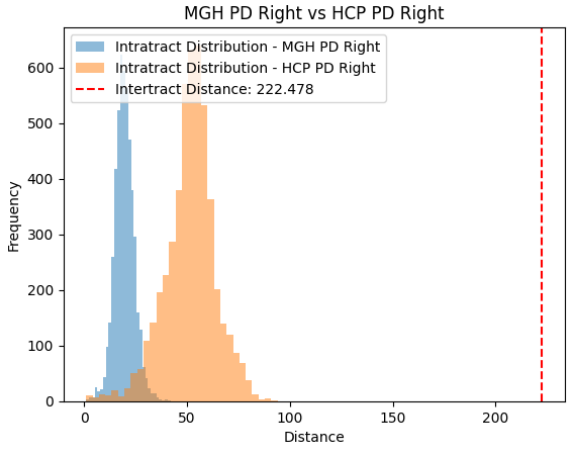

Figure 6: **Cross-Connectome Comparison.** MGH PD Right vs HCP PD Right (**222**). MGH connectome is from single subject, while HCP is an ensemble average of 900+ subjects. Further study needed to directly link distributions to connectome properties.