University of California
Santa Barbara

Reeb graphs for topological connectomics

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Electrical and Computer Engineering

by

S. Shailja

Committee in charge:

Professor B. S. Manjunath, Chair
Professor Scott T. Grafton
Professor Jefferson W. Chen, University of California Irvine
Professor Shiv Chandrasekaran
Professor Nina Miolane

June 2024

The Dissertation of S. Shailja is approved.

Professor Scott T. Grafton

Professor Jefferson W. Chen, University of California Irvine

Professor Shiv Chandrasekaran

Professor Nina Miolane

Professor B. S. Manjunath, Committee Chair

May 2024

Reeb graphs for topological connectomics

by

S. Shailja

Dedicated to Mummy, Papa, Didi, and Ayush.

Acknowledgements

Past 5 years of my PhD had some of the best days of my life! I have enjoyed doing research in Santa Barbara and I am grateful to my advisor, B. S. Manjunath, and to the members of my committee, Scott Grafton, Jefferson Chen, Nina Miolane, and Shiv Chandrasekaran for their guidance and support. I am thankful to Prof. Manjunath for guiding me throughout my PhD research. I am grateful to Prof. Grafton for shaping my brain connectome research. I appreciate Prof. Chen for his clinical insights in my research projects. I am grateful to Prof. Miolane for her invaluable discussions on brain connectivity during my early research and to Prof. Chandrasekaran for his insights on neural network ensembling methods during my internship at Mayachitra.

I couldn't have asked for a better advisor than Prof. Manjunath. I am grateful for his mentorship and encouragement to pursue my research interest. He is a very kind person at heart. He has always listened and addressed the challenges I faced during my journey even beyond academia. I always felt his trust and support in all my endeavours throughout my PhD journey. His expertise and teaching have been helpful for me to learn about machine learning and neural networks. In all my future research, I will apply his research principles to always keep in mind the bigger picture of the problem and answer the "Heilmeier Catechism". One of his most notable qualities is his extensive network of collaborations across various departments and beyond UCSB. I have been extremely lucky to work with some of the most exceptional collaborators.

First, I would like to thank Scott T. Grafton who has played an pivotal role in my PhD journey. I learned concepts behind MRI techniques from him. I am thankful how patiently he would discuss brain anatomy with me, especially as I come from the computational side of research. His support during my postdoctoral search was incredibly helpful. I thank him for his guidance and sharing his research vision with me. Later

during my PhD, I started collaborating with Jeff Chen. I truly appreciate this collaboration. I always wondered how Dr. Chen managed to meet with us every week. I am grateful for his dedication. I am proud of the clinical research that resulted from our collaboration. It gave me the sense of accomplishment for making actual impact in the field of neurosurgery.

Alongside my research, mentoring has been an important part of my journey. Many undergraduate and high school students have significantly contributed to this experience. I am particularly grateful to Vikram and Krithika for their substantial research contributions to both of my major journal publications. I also want to express my gratitude to high school students: Nisha, Claire, Alex, Kathy, Sanjay, Saanvi, and Ayush for being part of my journey. A special thanks to Lina Kim for her trust in me and giving me opportunities of being an instructor of record and a mentor in SRA and RMP program. Dr. Kim has always helped me get way better at teaching and mentoring! I am also thankful to Lisa Berry for her assistance in my completion of the CCUT certificate.

Going back to how my research journey started, I have been lucky to find a friend and now my partner, Ayush who introduced me to research during my undergraduate days. His belief in my research instincts and passion has led me to where I am today. I never thought of pursuing PhD until he motivated me and made me believe that I will excel in this. I always appreciate how he can read every single line of my research paper and give me feedback. I am happy how well we work together, especially proud of our ASEE work. Ayush is a role model for me. His positive outlook has always inspired me. I am thankful to Ayush for being there through all the highs and lows of my life!

Finally, my warmest thanks to my family. What my family has done for me and my dream are beyond what words can express here. I can never thank you enough mummy and papa for your continuous support and endless love. My mother is an inspiration to me! Thank you for encouraging me to be independent and to pursue my dreams even if

this meant I will be away from you and home. Thank you Papa for trusting me always and being so proud of me all the time! Your early morning calls would take all my stress away. I am also blessed to have two wonderful sisters. Their selfless love along with the joy brought by my adorable nieces Lado, Sado and Fudo brighten even my toughest days. My life is dedicated to all of you! I also want to thank my life-long friend, Kirti who has been like a sister to me. Thank you for being there for me always. I am grateful to Nishtha and Aneesha for being part of my UCSB life. It was fun filled first two years of my PhD because of you both!

I am glad to have met my fellow graduate students and researchers: Poyu, Oytun, Angela, Tom, Matt, Bowen, Connor, Chandrakanth, Satish, Raphael, Fei, Iftekhar, Devendra, Mike, Ekta, Amil, Austin, Fei, Joaquin, Kimia, and Umang. I am grateful to Poyu for mentoring me and giving me the opportunity to work with him in his brain tumor project. I appreciate Tom for his assistance in writing my first first-authored PhD paper. I am thankful to Angela for her help with my first MICCAI paper and for trusting me with her Ciona project. I appreciate Raphael and Fei for their insightful discussions on the COVID project. I am grateful to Kin, Bowen, Connor, and Chandrakanth for their discussion on the IARPA project. I am thankful to Satish and Arthur for TAing in my summer course. Finally, I thank the institutions and fellowships that supported the completion of my PhD: UC Santa Barbara, ECE Dissertation Fellowship, NSF-1664172, NIH-5R01NS103774-04, CHRI, and IARPA-140D0423C0057.

Curriculum Vitæ

S. Shailja

Education

2024	Ph.D., Electrical and Computer Engineering, University of California, Santa Barbara.
2020	M.S., Electrical and Computer Engineering, University of California, Santa Barbara.
2016	B.Tech., Electrical Engineering, Indian Institute of Technology, Kharagpur.

Peer-reviewed journal publications

Shailja, S., Christopher Nguyen, Krithika Thanigaivelan, Vikram Bhagavatula, Jefferson W. Chen, and B. S. Manjunath. "AI-based analysis of the shunt treatment in pre- and post-surgery computed tomography brain scans of iNPH patients". Neurosurgery, 2024.

Shailja, S., Vikram Bhagavatula, Matthew Cieslak, Jean M. Vettel, Scott T. Grafton, and B. S. Manjunath. "ReeBundle: a method for topological modeling of white matter pathways". IEEE Transactions on Medical Imaging 2023.

JJiang, Jiaxiang, Amil Khan, **S. Shailja**, Samuel A. Belteton, Michael Goebel, Daniel B. Szymanski, and B. S. Manjunath. "Segmentation, Tracking, and Sub-cellular Feature Extraction in 3D Time-Lapse Imagery ". Nature Scientific Reports 2023.

Zhang, Angela, **S. Shailja**, Cezar Borba, Yishen Miao, Michael Goebel, Raphael Ruschel, Kerrianne Ryan, William Smith, and B. S. Manjunath. "Automatic Detection and Neurotransmitter Prediction of Synapses in Electron Microscopy". Biological Imaging 2022.

Kao, P. Y., **Shailja, S.**, Jiang, J., Zhang, A., Khan, A., Chen, J. W., and Manjunath, B. S. "Improving patch-based Convolutional Neural Networks for MRI brain tumor segmentation by leveraging location information". Frontiers in Neuroscience 2020.

Peer-reviewed conference publications

Bowen Zhang*, **Shailja, S*.** (* equal contributors), Chandrakanth Gudavalli, Connor Levenson, Amil Khan, and B. S. Manjunath. "ReeSPOT: Reeb Graph Models Semantic Patterns of Normalcy in Human Trajectories". Submitted to 27^{th} International Conference on Pattern Recognition (ICPR), 2024.

Shailja, S., Jefferson W. Chen, Scott T. Grafton, and B. S. Manjunath. "ReTrace: Topological insight to evaluation of white matter tractography using Reeb graphs". CDMRI Workshop in the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) 2023.

Shailja, S., Angela Zhang, and B. S. Manjunath. "A computational geometry approach for modeling neuronal fiber pathways". International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) 2021.

Shailja, S*., Jiang, J*., and Manjunath, B. S. "Semi supervised segmentation and graph-based tracking of 3D nuclei in time-lapse microscopy". International Symposium on Biomedical Imaging (ISBI) 2021.

Honors & Awards

2024	UCSB Lancaster Dissertation Award
2024	UCSB Certificate in College and University Teaching (CCUT)
2024	ECE Department Dissertation Fellowship, UCSB
2023	Fiona and Michael Goodchild Graduate Mentoring Award
2023	NSF iREDEFINE Fellow 2022
2023	Travel Award: MICCAI 2022, 2023 (Academic Senate), IPAM 2022, 2024, NeurIPS 2021 (WiML)
2020	2nd Place in Cell Tracking Challenge, ISBI

Teaching & Community Service

2024	**Co-organiser:** Computational Diffusion MRI (CDMRI) Workshop at MICCAI, 2024.
2024	**Student Board Advisory:** International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) 2024, Marrakesh.
2023	**Teaching Associate:** Diagnostic AI-Research in STEM (INT 93LS)
2019	**Teaching Assistant:** Digital Image Processing (ECE178), Design and Analysis of Algorithms (CS130B), Operating System (CS170)
2022-24	**Reviewer:** International Journal of Computer Assisted Radiology and Surgery (IJCARS) 2024, International Conference on Information Processing in Computer-Assisted Interventions (IPCAI) 2024, Medical Physics Journal 2023, Medical Image Computing and Computer Assisted Intervention (MICCAI 2022, 2023, 2024), International Workshop on Medical Optical Imaging and Virtual Microscopy Image Analysis (MOVI 2022).

Abstract

Reeb graphs for topological connectomics

by

S. Shailja

Our brain consists of approximately 100 billion neurons that form functional neural networks across different brain regions. Brain functions involve complex interactions between these regions that are poorly understood and lack quantitative characterizations. Diffusion MRI of the brain generates millions of complex curvilinear fibers (streamlines) in 3D that exhibits the geometry of white matter pathways in the brain. It has been shown that a critical element in neurological and developmental disorders is the topological deterioration in streamlines. Despite this, most existing methods model the neural connections with connectivity matrices that overlook the topology of connections.

Towards addressing this shortcoming, we model neuronal fibers as context-aware geometrical objects in three-dimensional space. We introduce a novel Reeb graph-based method that efficiently encodes the topology and geometry of white matter fibers. Given the trajectories of neuronal fiber pathways of a neuroanatomical bundle, we re-bundle the streamlines by modeling their spatial evolution to capture geometrically significant events (akin to a fingerprint). Reeb graph parameters control the granularity of the model and handle the presence of improbable streamlines commonly produced by tractography. Our method meaningfully describes the bundle tracts without oversimplifying them into skeletons.

To quantify the difference between the connections, we introduce a new Reeb graph-based distance metric that quantifies the topological differences in bundle comparison. For the longitudinal repeated measures in the Cognitive Resilience and Sleep History

(CRASH) dataset, repeated scans of a given subject acquired weeks apart lead to provably similar Reeb graphs that differ significantly from other subjects, thus highlighting our method's potential for clinical fingerprinting of brain regions. Reeb graph and topological distance metric are sensitive to variations in brain structure, such as those observed in the developing brain and in the presence of tumors. This sensitivity is beneficial in the context of longitudinal studies of brain development and the topological evolution of the brain. This thesis highlights the potential utility of our topology-based distance metric in tracking Alzheimer's disease progression using the ADNI dataset, evaluating the effects of surgical interventions for brain tumor using the OpenNeuro brain tumor datasets and quantifying the effect of shunt surgery for NPH brain condition in collaboration with UCI Medical Center.

Our method is not limited to neuroscience and is general-purpose in its applicability. This is demonstrated in our application of Reeb graphs for structure discovery in spatio-temporal trajectories where we use the model for detecting anomalous trajectories. Human behavior typically follows a pattern of normalcy in day-to-day activities. This is marked by recurring activities within specific time periods. Our method models this behavior using Reeb graphs where any deviation from usual day-to-day activities is encoded as nodes in the Reeb graph.

In summary, this thesis build up the theoretical framework for modeling spatio-temporal trajectories in 2D and 3D space, with applications to brain connectome analysis. A promising direction for future research is to link the Reeb graph-based structural model presented here with the behavioral model derived from functional MRI. This multi-modal strategy is expected to yield novel insights into the structural and functional connectivity of the human brain.

Contents

Chapter 1

Introduction

The human brain is incredibly complex, posing one of neuroscience's greatest questions: how can we infer what is actually happening within the brain? Determining the relationship between structural changes in the brain and its functions is challenging. This thesis aims to advance our understanding of brain structure through data-driven mathematical modeling.

To model the brain, one might consider detailed examination of specific areas, meticulously mapping individual neurons to understand their precise interactions. However, focusing solely on individual neurons can overlook the broader context of their interactions. Conversely, attempting to model the entire brain might miss critical details of neuronal activity.

Our approach strikes a balance by creating a sparse yet structurally significant model of specific brain areas. This dissertation focuses on the mathematical modeling of brain connections, developing novel theory to represent neuronal fibers as geometrical objects in three-dimensional space. This quantitative modeling addresses the generic sub-trajectory clustering problem and we show how it can be applied to fields beyond neuroscience.

1.1 Networks are Everywhere

A key observation in this thesis is how several biological structures of interest, such as blood vessels, neurons, neural circuits, and cell consortia, can be abstracted as networks. This abstraction is not unique to biology, as we often represent spatiotemporal behavior of multiple objects dynamically over time with graphs and networks.

Recent advancements in high-throughput biomedical imaging techniques have further supported this perspective, enabling researchers to model biological processes as interconnected networks. Graph theoretical analysis of these networks, combined with data-driven learning approaches, can boost our understanding of various human disease pathways and underlying biological phenomena. For example, examining synaptic and neuronal networks aids in studying the brain, while analyzing cell networks helps in understanding the growth, death, and grouping of cells.

To study these networks, we formulate a general problem of modeling 3D spatiotemporal trajectories. We propose a novel theory of Reeb graphs and their implementation algorithms to capture connections in trajectories. The collective motion of moving cells, the transmission of information in neurons, and human movement are characterized by the trajectories they follow, and this work aims to provide new insights into these complex processes.

1.2 Imaging the human brain

Over the past two decades, advancements in computing and image acquisition technology have significantly advanced the field of neuroimaging. These technological improvements have deepened our understanding of brain mechanisms and enhanced our ability to identify the causes of impairments by classifying patients and healthy individ-

uals. One of the greatest achievements in neuroimaging is the development of various kinds of magnetic resonance imaging (MRI) techniques. MRI has enabled high resolution imaging of the human brain, the most elusive human organ. MRI scans are flexible as radiologists adjust different contrast or relaxation settings [1] to visualize different brain disorders in varying contexts. For example, the T1 relaxation setting is used in imaging the cystic components of a tumor core in the brain, while the T2 relaxation is used to image the whole tumor structure. Even further, scientists used the contrast dependent properties of water molecules and developed the Diffusion Weighted Imaging (DWI) [2], another type of MRI. DWI can be used to image the 3D structural information of brain connectivity, thus offering a different perspective to a physician. In similar vein, functional MRI (fMRI) was developed for spatiotemporal imaging of brain cognition. DWI and fMRI can "image" molecular activity and cellular function at a resolution that is impossible for conventional MRI. As a result of this progress in imaging devices, today there are many high-resolution, large-scale brain imaging data sets available (HCP [1], ABCD [2], ADNI [3]). However, the computational and signal processing tools are lagging behind. There are no frameworks yet that comprehensively analyze the brain images of different modalities to explain the brain networks.

With the progress in AI, computational researchers have made significant strides towards understanding and designing cyber-physical systems by applying deep learning methods to different imaging modalities. However, it would be fair to say that the premise of this success is the reliance on high-throughput data from well-behaved systems. Biological systems are highly context-dependent, noisy, and the data modalities are often low-throughput. The inputs and outputs in these methods are often disconnected from the mechanistic details. We note that most deep learning methods are

[1] `https://www.humanconnectome.org/`
[2] `https://nda.nih.gov/abcd`
[3] `https://adni.loni.usc.edu/`

incapable of providing interpretable guidance to clinicians about disease biomarkers or crucial biomolecular location information about pathology. Therefore, in this thesis we explore the topology and structure of network data using first-principles, motivated by one of the most elusive problems — neuronal connectivity in the human brain.

1.3 Summary of Contributions

This dissertation poses a central research question: Can we model the structural connections of a network in a scalable way to represent spatial information flow? Motivated by brain connectivity, this work introduces a novel Reeb graph-based method that efficiently encodes the topology and geometry of the brain's white matter pathways, as captured in the diffusion MRI data. By defining the evolution of level sets, we rebundle 3D trajectories of neuronal fiber pathways, capturing critical geometric events – akin to a brain fingerprint. This method proves valuable in detecting disease-related and age-dependent topology alterations in studies on Alzheimer's, brain tumors, and iNPH.

Beyond neuroscience, the approach has broader applications, demonstrated through the development of a general algorithm for discovering structure in spatiotemporal trajectories. This algorithm identifies anomalies in such spatiotemporal data. In summary, this thesis presents an extensive exploration of spatial Reeb graphs, detailing their implementation for spatiotemporal data and discussing their potential as general purpose tools for scalable trajectory modeling. The main contributions of this dissertation are listed below.

1. We develop an efficient method to model the white matter connectivity as a sparse Reeb graph. We parameterize our model to mitigate the effects of noise in streamlines. Noise in this context refers to geometrically implausible streamlines due to curvature overshoot or short length due to premature termination.

2. We introduce a new graph-based method to quantify the topological difference between two bundle tracts that can be used to compare brain connectivity between individuals or populations. We demonstrate the utility of our approach for visualization and tractograph evaluation using the International Society for Magnetic Resonance in Medicine (ISMRM) dataset [3].

3. We utilize the proposed model to fingerprint the human brain networks, to track Alzheimer's disease progression using the ADNI dataset, and evaluating the effects of surgical interventions for brain tumor using the OpenNeuro brain tumor dataset. Towards quantifying the impact of shunt surgery for iNPH patients we demonstrate the usage of CT scans and discuss how Reeb graph-based pipeline can help us in multi-modal quantification.

4. We extend our theory to spatiotemporal trajectories. As a case study of this algorithm, we propose a Reeb graph-based method to model the day-to-day activities of a given agent and detect anomalous behavior.

1.4 Organization of Thesis

An overview of the chapters of this thesis is shown in Figure. 1.1.

Chapter 2: We formulate the problem of spatial trajectory clustering, introducing the concept of Reeb graphs as a tool to model the bundling structure of high-dimensional data. We develop a Reeb graph-based algorithm to discover the high-level topological structure and the evolution of a collection of 3D trajectories. This algorithm computes a sparse graph representing the latent bundling and unbundling structure of trajectories in space. To account for the irregular noise in the data, we parameterize the algorithm with biophysical parameters that capture the topology of the data — ϵ (distance between

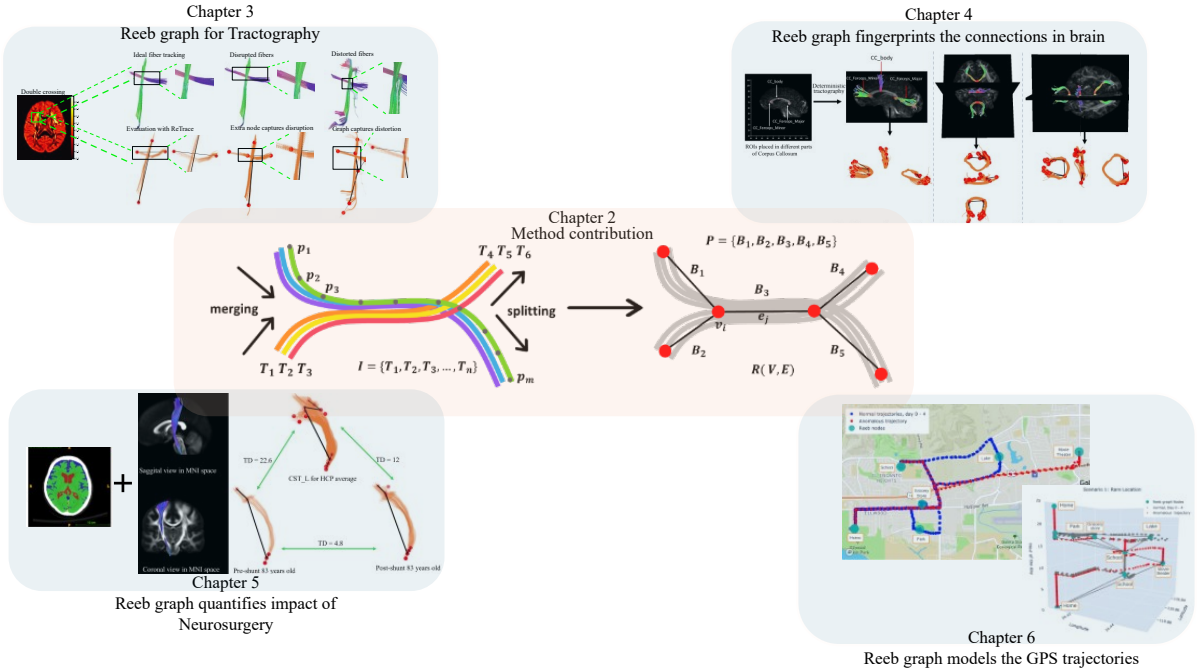

Figure 1.1: An overview of the thesis: Reeb graphs for topological connectomics.

a pair of trajectories in a bundle), α (spatial length of a bundle of trajectories), and δ (bundle thickness).

Chapter 3: Reeb graphs can guide the design of trajectory generation algorithms. Any trajectory analysis mainly depends on the quality of the trajectory data generated. As a desired consequence, the Reeb graph models of trajectories can also be tuned to be highly sensitive to the quality of the data. In neuroimaging, tractography is a trajectory-generation technique used to reconstruct white matter fiber pathways from diffusion magnetic resonance images (dMRIs). Therefore, in Chapter 3, we demonstrate the usage of Reeb graphs as an evaluation method for tractography algorithms. This method addresses the limitations of existing metrics by focusing on the topological accuracy of reconstructed pathways.

Chapter 4: We show the practical usage of the theory of Reeb graphs for brain: (1) For International Society for Magnetic Resonance in Medicine (ISMRM) dataset,

our method handles the morphology of the white matter tract configurations due to branching and local ambiguities in complicated bundle tracts like anterior and posterior commissures; (2) For the longitudinal repeated measures in the Cognitive Resilience and Sleep History (CRASH) dataset, repeated scans of a given subject acquired weeks apart lead to provably similar Reeb graphs that differ significantly from other subjects, thus highlighting the potential of Reeb graphs for clinical fingerprinting of brain regions. We develop a Reeb graph based pipeline for longitudinal tracking of neuroconditions such as Alzheimer's Disease and tumor.

Chapter 5: In this Chapter, we present a case study for using Reeb graphs for quantifying and localizing neurosurgery. We evaluate the effects of shunt surgery on the condition of idiopathic Normal Pressure Hydrocephalus (iNPH). We present an AI-based method for quantifying ventricular volume in iNPH patients pre- and post-surgery. With Reeb graphs, we develop a multi-modal analysis pipeline that uses segmented CT scans and diffusion MRI. We discuss how this analysis could improve our understanding of the disease by interpreting diagnostic symptoms associated with gait, cognition, and bladder function with respect to structural network changes in iNPH patients.

Chapter 6: The theoretical advances of the algorithm in this thesis are general-purpose and applicable to other big data contexts, including multi-omics, epidemiology, human GPS data, and other space-evolving datasets. In Chapter 6, we discuss one such problem of behavior modeling and anomaly detection in human mobility. Recently, there has been an increase in location-aware devices that use the Global Positioning System (GPS) for many applications such as finding efficient routes [4], fitness apps, understanding the progression of infectious diseases [5], and predicting demographic information [6]. This collection of movements, and thus vast amounts of raw trajectories, spotlights the need for a scalable representation of these trajectories that preserves and highlights the structurally and topologically important movement patterns. We use Reeb graphs to

extract contextual information for human behavior classification in this chapter. We discuss how future research in this direction could use the features extracted by Reeb graphs to create representative embeddings in a graph neural network model for anomaly prediction.

Chapter 7: We discuss some future directions and conclude this dissertation in this chapter.

In summary, this dissertation presents a new structure discovery method of brain connectomics using Reeb graphs. This offers a novel perspective on the structural intricacies of the human brain. By developing advanced computational models and graph-theoretic approaches, this work not only decodes the complex neural pathways but also establishes a method for their topological and geometric analysis. The proposed method in this thesis quantifies the topological differences in brain structures for localized comparisons for early diagnosis of neurological conditions. Furthermore, the application of these methods extends beyond neuroscience demonstrating their versatility in analyzing geo-spatial data. This interdisciplinary approach deepens our understanding of trajectories, whether in neural connections in the brain or in human movement trajectories. Reeb graphs hold promise in modeling such high dimensional and large scale systems effectively.

Chapter 2

Spatial Reeb Graphs

Understanding brain is a complex problem! As such, it requires a mix of approaches from computer vision, geometry, graph theory, and topology. In this chapter, we introduce brain connectome which describes the structural connections of the brain. Wiring formed by white matter fibers typically originate in functional regions, merge with nearby fibers into larger pathways for efficiency, and diverge as they approach their terminal functional regions. Therefore to understand a brain mathematically, we set up the subtrajectory clustering problem to model the structure of brain network. General-purpose graph-theoretical methods are computationally intractable and overlook crucial information about the physical proximity and bundling structure of fibers. In this chapter, we present a novel algorithm for addressing the mathematical modeling problem.

The content from this Chapter was published in the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) [7] and IEEE Transactions on Medical Imaging (TMI) [8].

2.1 Connectome

A *connectome* is a wiring diagram. In context with brain, a connectome is a comprehensive map of neural connections in the brain. A connectome is constructed by tracing the neuron in a nervous system and mapping where neurons are connected through synapses. Recent advances in non-invasive imaging technologies such as diffusion magnetic resonance imaging (dMRI) combined with tractography methods (discussed in Chapter 4) allows reconstruction of connectome in the human brain as shown in Figure. 2.1. The field of "Connectomics" focuses on the compilation and examination of connectome datasets, playing a crucial role in advancing our understanding of neural structures and functions [9]. This thesis centers around topological *connectomics* using Reeb graphs.

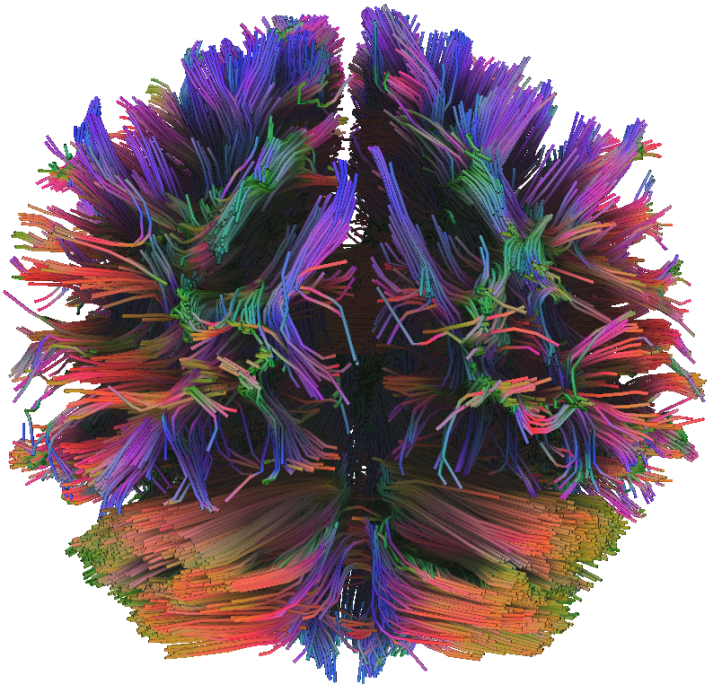

Figure 2.1: **Coronal view of human brain connectome.** This shows a dMRI image of brain where white matter fibers are tracked using the DSI Studio (`https://dsi-studio.labsolver.org/`). DSI studio is a tractography software tool that maps brain connections. The colors in this figure indicate the various directions of the fibers.

2.2 Significance of Topology

In the context of brain connectivity and white matter fibers, topology refers to the arrangement and organization of the fibers into structural networks, taking into account their arising/ending and branching behavior and the relationships between different regions of the brain. The topology of the white matter fibers is crucial for understanding the brain's structural/functional connectivity and how it is affected by various factors, such as aging, disease, and injury. We assume that the groups of trajectories that are spatially close to each other share similar properties. In this Chapter, we present our method, ReeBundle which is a Reeb graph-based model to encode the arising & ending and the merging & splitting behavior for groups of sub-trajectories.

Neurological disorders can cause two topologically relevant changes in the white matter fibers: physical disruptions and spatial distortions. The white matter fibers are physically disrupted in neurological disorders such as stroke or multiple sclerosis. On the other hand, diseases like brain tumor [10, 11] can cause spatial distortion in the white matter pathways as tumor mass occupies the brain space [12]. These topologically relevant changes caused by neurological disorders are illustrated in Figure. 2.2. Modeling

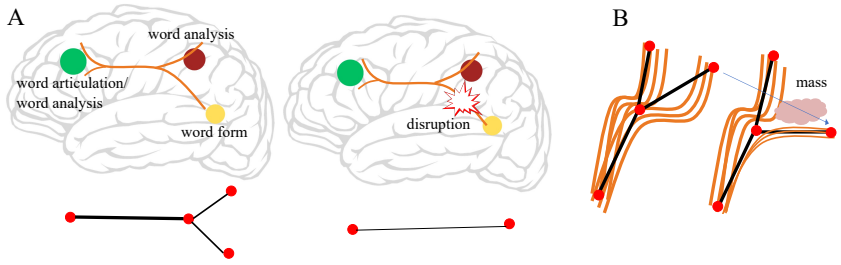

Figure 2.2: **Significance of topology in neurological disorders.** A) Branching disruption in Alexia without agraphia where the branching from word form to word analysis is disrupted, while the other branch remains intact. B) The pink-colored tumor mass occupies space in the brain and causes displacement and deviation of the surrounding white matter fibers. This spatial distortion of white matter fibers is reflected in the Reeb graphs, with displaced nodes indicating the distortion in white matter tracts caused by the tumor mass.

the branching behavior of the white matter fibers using ReeBundle can capture these topological changes and provide valuable information for understanding and detecting disease progression. For example, stroke [13] causes direct disruption of portions of classical tracts. It can lead to a selective disruption of the hand and face motor cortex in the superior spinal tract of the corticospinal tract (CST) or disrupt the classical language network that affects word analysis. Another example is Alexia without agraphia [14]. Here, the ability to read is lost while the ability to write remains intact. To analyze such neurological disorders, branching is important to consider as shown in Figure. 2.2A. We show the condition where the lower branch that connects the word form with word articulation stops functioning. Therefore, investigating brain topology by modeling the branching behavior of the bundle tracts using ReeBundle can provide useful biomarkers for the study of brain development and the detection of disease progression.

In this chapter, we establish the theoretical framework for the Reeb graph model.

2.3 Definitions

We have 3D Euclidean space with n trajectories and a total of N points. Given an input set of trajectories that correspond to a coherent streamline bundle, $\mathcal{I} = \{T_1, T_2, ..., T_n\}$, we compute the Reeb graph $\mathcal{R}(V, E)$, where edges represent the set of ϵ-step connected fibers and vertices represent the arise, disappear, split or merge behavior of trajectories as shown in Fig 2.3. $\mathcal{R}$ is an undirected, weighted graph that captures such critical points of the trajectories. The parameter ϵ controls the granularity of the grouping desired. Small values of ϵ allow only very dense sets of subtrajectories to be considered for grouping, while larger values of ϵ relax the groups and allow larger, sparser groups [15]. In the three dimensional Euclidean space $\mathbb{R}^3$, we define the following terms that help in setting up the problem.

***Trajectory*:** A trajectory T is as an ordered sequence of points in $\mathbb{R}^3$, $T = \{p_1, p_2, ..., p_m\}$, where m is the number of points in T and $p_i \in \mathbb{R}^3$.

***Subtrajectory*:** A subtrajectory U of a parent trajectory T starting at s and ending at t is represented by a consecutive subsequence of points $\{p_s, p_{s+1}, p_{s+2}, ..., p_t | p_i \in T\}$.

ϵ-(dis)connected points: For any pair of points p_1 and p_2 in $\mathbb{R}^3$, we define $d(p_1, p_2)$ as the Euclidean distance between the two points:

$$d(p_1, p_2) = \|p_1 - p_2\|_2,$$

where $\|\cdot\|_2$ represents the Euclidean norm. Two points p_1, p_2 are ϵ-connected if $d(p_1, p_2) \leq \epsilon$. Similarly, two points p_1, p_2 are ϵ-disconnected if $d(p_1, p_2) > \epsilon$.

ϵ-connected trajectories: Two trajectories T_i and T_j are ϵ-connected if every point in T_i is within ϵ of some point in T_j and vice versa. Note that the definition of ϵ-connected extends to subtrajectories as well.

Appear event: For each trajectory T, the initial point of its ordered sequence is labeled for the occurrence of the *appear event*. For example, for trajectory $T_1 = \{p_1, p_2, ..., p_m\}$, we observe the appear event at the point p_1.

Disappear event: For each trajectory T, the final point of its ordered sequence is labeled for the occurrence of the *disappear event*. For example, for trajectory $T_1 = \{p_1, p_2, ..., p_m\}$, we observe the disappear event at p_m.

(Dis)connect events: Consider two ϵ-connected subtrajectories U and U' belonging to the trajectories T and T',

$$U = \{p_1, p_2, ..., p_m\}, \quad U^{'} = \{p_1^{'}, p_2^{'}, ..., p_q^{'}\},$$

then a connect event for the pair $(T, T^{'})$ is defined by $(p_1, p_1^{'})$ and a disconnect event is

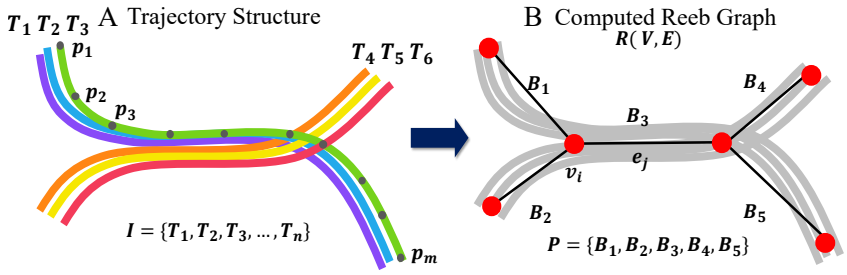

Figure 2.3: **Bundling structure of streamlines.** (A) An example showing the branching structure of the streamlines and (B) the Reeb graph where nodes encode the merge, split, and termination characteristics. Nodes of the Reeb graph $\mathcal{R}$ on the right-hand side are shown in red color and the edges are shown in black color throughout the thesis.

observed at (p_{m+1}, p'_{q+1}). If there is no such pair of points for connect event, it implies that T and T' are disjoint. Moreover, if T and T' are ϵ-connected and if T' and T^* are also ϵ-connected, then we say that T and T^* are *ϵ-step connected*.

Note that the trajectories estimated from dMRI tractography do not have a specific beginning or ending as dMRI is not sensitive to the direction of connections. So, reversing the order of the points in a streamline will produce similar results. The appear and disappear events are used for convenience in describing the algorithm and its implementation.

2.4 Bundling Structure Of Tractographs

White matter fibers in brain space are represented as a set of spatial trajectories in 3D Euclidean space such that each streamline is a trajectory $T = \{p_1, p_2, ..., p_m\}$ where $p_i \in \mathbb{R}^3$ is an ordered sequence of points and m is the total number of points in T. A common behavior of streamlines is that they start from one brain region, merge with other streamlines into bundles, and then split towards the end into other brain regions. Intuitively, we can assume that if a continuous portion of a set of fibers is close together then they share a common anatomical behavior. In this Chapter, we model

the topology of these streamlines by formulating a problem of computing a Reeb graph model $\mathcal{R}(V, E)$ given a set of trajectories $\mathcal{I} = \{T_1, T_2, ..., T_n\}$, where n is the number of trajectories. The vertices V of the Reeb graph encode the termination (*appear* and *disappear*) or branching (*connect* and *disconnect*) of trajectories and the set of edges E are the group of subtrajectories bundled together.

Bundling structure, as shown in Fig 2.3, partitions the set of trajectories into a group of subtrajectories to visualize the grouping aspects of the trajectories and to compare grouping across different input sets. Given $\mathcal{I}$ and a distance parameter ϵ, the bundling structure of a set of trajectories is modeled using a Reeb graph of the ϵ-connected trajectories. We define:

Bundle: A Bundle denoted by B is a set of subtrajectories that consists of at most one subtrajectory from each trajectory T. Two bundles B and B' are adjacent if $\exists\, U \in B$ that is continued as $U' \in B'$, where U and U' are the subtrajectories of T. B is ϵ-bundle if any two subtrajectories in B are ϵ-step connected i.e, connected by a sequence of ϵ-connected subtrajectories. An ϵ-bundle B is max-width if no other possible ϵ-bundle of $\mathcal{I}$ intersects B and contains a superset of the trajectories represented by B.

Max-width ϵ-bundle partition: Given a set of max-width ϵ-bundles, a max-width ϵ-bundle partition $\mathcal{P} = \{B_1, B_2, ..., B_l\}$ such that every point in $\mathcal{I}$ is assigned to a bundle.

Note that by the term bundle we mean the set of subtrajectories that are in close proximity, that is, max-width ϵ-step connected subtrajectories as shown in Figure. 2.4. A bundle shares the same spatial and shape characteristics and does not necessarily correspond to neuroanatomical bundle tracts.

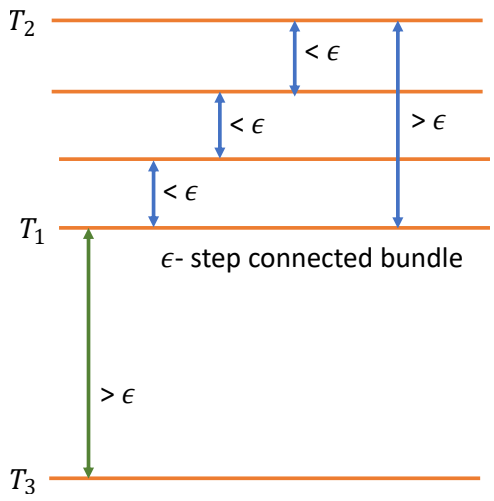

Figure 2.4: **Illustration of $\epsilon-$step connected trajectories.** Trajectories T_1 and T_2 are $\epsilon-$step connected while T_1 and T_3 are not $\epsilon-$step connected.

2.5 Construction of Reeb Graphs

The central part of solving the problem as stated above is to compute the bundling structure represented by Reeb Graph $\mathcal{R}$. Towards that end, we compute an undirected, weighted graph $\mathcal{R}$. In this section, we define the Reeb graph and then proceed to develop an algorithm that can compute this graph for a set of trajectories. Formally, a Reeb graph $\mathcal{R}$ is defined on a manifold $\mathcal{M} \in \mathbb{R}^3$ using the evolution of level sets L [16]. To adapt this definition of $\mathcal{R}$ for the case of neuronal fiber trajectory evolution, we define a manifold $\mathcal{M}$ in $\mathbb{R}^3$ as the union of all points in the tractogram. The set of points of trajectories at step k is the level set at k. The connected components in the level set at k correspond to the max-width ϵ-connected trajectories at step k. Unlike previous studies [17], here, any number of trajectories can become ϵ-(dis)connected at the same location. Reeb graphs are computed for spatially evolving level sets. Reeb graph, $\mathcal{R}$, describes the evolution of the connected components over sequential steps. At every step k, the changes in connected components are represented by vertices in $\mathcal{R}$.

In Section 2.3, for a given trajectory, we defined appear and disappear events. For a

pair of trajectories, we defined connect and disconnect events. These events describe the branching structure of the trajectories. To compute the Reeb graph, we process these events sequentially. We maintain a graph $G = (V^{'}, E^{'})$ where the vertices represent the trajectories. G is a graph that changes with steps representing the connect and disconnect relations between different trajectories.

Initialization: We spend $O(N^2)$ time to store the events for all pairs of the trajectories. This step is massively parallelizable. An initial level of simplification can be done by using QuickBundles [18] in $O(N)$ on average.

Handling *appear, disappear, split* and *merge* events: At each step k, we insert new nodes at appear events and delete nodes at disappear events. At connect events, we insert edges in G and at disconnect events, we delete edges. To handle a disconnect event of trajectories T_1 and T_2 at step k, we delete the edge (T_1, T_2) from G_k. Similarly, for a connect event of trajectories T_1 and T_2 at step k, we add the edge (T_1, T_2) to G_k. These steps are demonstrated in Figure. 2.5. At each step k, an edge (T_1, T_2) in G shows that T_1 and T_2 are directly connected. Therefore the max-width ϵ-connected trajectories correspond to the connected components in G at step k. We do this for all the connect and disconnect events as shown in Figure. 2.5.

Computing $\mathcal{R}$ from G and $\mathcal{P}$: At each step k, we query G_{k-1} and G_k to get the connected component C_{k-1} and C_k respectively. If a connected component $c_k \in C_k$ is present in C_{k-1}, we assign the same bundle id B_i to the points in step k for the trajectories in the connected component c_k. On the other hand, if c_k is not present in C_{k-1}, we create a new bundle id and assign it to the points in step k for the trajectories in the connected component c_k. This gives us the maximum bundle partition $\mathcal{P}$.

Finally, the Reeb graph $\mathcal{R}$ for $\mathcal{P}$ is an undirected graph where each edge represents a max-width ϵ-bundle B_i and each pair of edges e_i and e_j are connected with a vertex if B_i and B_j are adjacent bundles as shown in Figure. 2.3. The steps for computing Reeb graph

Algorithm 1 *Construction of Reeb Graph*

```
function ConstructReebGraph(I, set of events for all pairs of trajectories)
    for all steps k from 0 to maximum length of the trajectory in I do ▷ Event Handling
        if appear event of T then
            insert new node T to G_k
        if disappear event of T then
            delete node T from G_k
        if connect event between T_1 and T_2 then
            insert edge (T_1, T_2) to G_k
        if disconnect event of trajectories T_1 and T_2 then
            delete edge (T_1, T_2) from G_k
        P ← empty bundle partition                                  ▷ Bundle Partition
        Query G_{k-1} and G_k to get the connected components C_{k-1} and C_k respectively;
        for all connected component c_k ∈ C_k do
            if c_k ∈ C_{k-1} then
                assign the same bundle id B_i to the points for trajectories in c_k;
            else
                create a new bundle id B_{i+1} and assign it to the points for trajectories
                in c_k;
            Add B_{i+1} to P
    Construct Reeb graph R from P by connecting adjacent bundles with nodes and
    bundles as edges;
    return R
```

are shown in Algorithm 1. Note that the Reeb graph is monotonous with respect to the distance parameter ϵ i.e. if B is a bundle for a given ϵ, then for any $\epsilon' > \epsilon$, $\exists B' \supseteq B$. The monotonicity property ensures consistency between models at different levels of detail and allows tunable coarse-graining of models for the streamlines.

2.6 Parameter Selection For Reeb Graphs

The bundling structure discussed above is susceptible to noise and errors in the streamlines. It is known that tractography techniques can suffer from a large number of improbable streamlines due to the reconstruction of streamlines that do not correspond

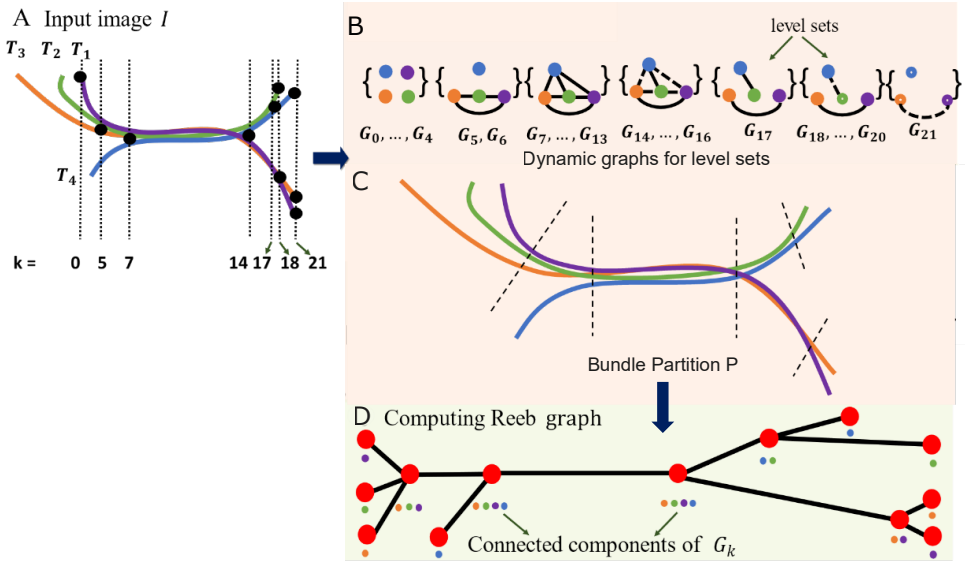

Figure 2.5: **Steps for ReeBundle method.** A) Input $\mathcal{I} = \{T_1, T_2, T_3, T_4\}$ exhibiting arising, merging, splitting, and ending behavior. Connect, disconnect, appear, and disappear events are marked by black circles. B) After processing the ordered sequence of points, $k = 0, 2, 5, ..., 21$, we modify the G_k respectively. Deleted nodes and edge queries are represented by dashed circles and dashed lines respectively in G_k. C) Max-width bundle partition for $\mathcal{I}$ is computed using the dynamic graphs. D) $\mathcal{R}$ is computed from G_k. Edges of $\mathcal{R}$ encode the maximal group of trajectories and vertices of $\mathcal{R}$ record the topologically significant points of the trajectories. The Reeb graph $\mathcal{R}$ for P is structured as an undirected graph where each bundle B_i in P is linked to an edge e_i in R. Two edges, e_i and e_j, are connected by a vertex in $\mathcal{R}$ if the bundles they correspond to are adjacent.

to real anatomical bundles [19]. Noise in this context refers to streamlines that are geometrically implausible due to curvature overshoot; short length because of premature termination, and spatially distant due to connection density biases. Hence, we propose a Reeb graph model to detect and discard outlier streamlines based on their geometrical properties such as length, density, size, and branching. Three key parameters in our method capture the geometry and topology of the streamlines (i) ϵ – the distance between a pair of streamlines in a bundle that defines its sparsity (ii) α – the spatial length of the bundle that introduces persistence and (iii) δ – the bundle thickness. Together, these parameters control the granularity of the model to provide a compact signature of

the tracts.

Minor interruptions of length < 2mm in the streamlines would be insignificant when modeling the bundling structures at the 20mm length scale. Hence, the amount of interruptions that may be considered insignificant is a latent parameter dependent on the modeling abstraction and the hypothesis underlying the model as shown in Figure. 2.6. We introduce a parameter $\alpha \geq 0$ to model this measure of topological persistence. Therefore, in the Reeb graph model, the precise events with length at most α are ignored. So, if a pair of trajectories are ϵ-(dis)connected for less than α interval then the associated event is ignored. Formally, given an input image $\mathcal{I}$ and its max-width bundle partition $\mathcal{P}$, we define:

α-relaxed ϵ-connected bundles: Two trajectories are α-relaxed and ϵ-connected at k if and only if they are ϵ connected at some $k' \in [k - \alpha/2, k + \alpha/2]$.

Finally, to control the thickness of the bundles, we introduce a parameter δ to model the minimum size (number of trajectories) of the bundles in the Reeb graph. This allows us to model the significant bundles whose size is more than δ trajectories and all other bundles are ignored. To sum up, we say that a bundle B is robust if it is a bundle according to the definition in Sec. 2.4 with its components replaced by α-relaxed ϵ-connected components and $|B| > \delta$. Note that the robust Reeb graph also manifests the monotonicity property in the parameters α and δ: a bundle B in an interval S remains a bundle in S on decreasing δ or α. Bundle ids in $\mathcal{B}$ are recomputed with respect to the new persistence (α) and size (δ) parameters.

Weighted Reeb Graph: Bundles in a given image are associated with edges in the Reeb graph and can be labeled with domain-specific information. Here, we use the normalized streamline count $\left(\frac{|B|}{n}\right)$ as the anatomical feature to represent the edge weight.

The optimal values for ϵ, α, and δ depend on the specific application being considered. For instance, in our study, we used a value of $\epsilon = 2.5$ for two main reasons: firstly, it

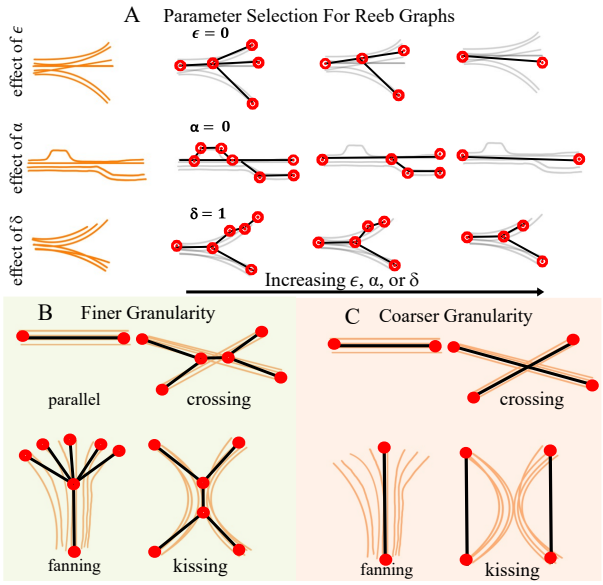

Figure 2.6: **Effect of Reeb graph parameters.** (A) Tuning ϵ, α, and δ obtains generalized views of the grouping structure and handles noise / false positives illustrated using toy trajectories. Common examples of white matter configuration that are observed in tractography and their Reeb graph models at (B) a finer granularity and (C) coarse granularity. For example, the *fanning* structure is finely encoded when the inter-trajectory distance, ϵ is less in (B). Similarly, the short interruptions while *crossing* and *kissing* can be captured with a smaller value of the persistence parameter α and can be ignored with a larger value of α.

is sufficiently close to the minimum dMRI resolution of 2mm, enabling us to capture the smallest reasonable groups. Secondly, it is large enough to ensure the identification of macroscopic groupings. The choice of parameters for our method is driven by the hypothesis underlying the Reeb graph construction. For example, a dense graph can be generated to compare and contrast tractography methods that generate streamlines, while a sparse graph can be computed for disease studies to suppress false positives and quantify disease-relevant features. For longitudinal studies, it may be useful to experiment with different parameter values to obtain the desired structure resolution and then fix the parameters for computing the Reeb graphs at subsequent time points.

2.7 Time Complexity Analysis For Reeb Graphs

We spend $O(N^2)$ time to store the events for all pairs of the trajectories. This step is parallelizable. An initial level of simplification can be done by using QuickBundles [18] in $O(N)$ on average.

Theorem 1 *For a given ϵ and set of trajectories, $\mathcal{I} = \{T_1, T_2, ..., T_n\}$ with a total of N points, the Reeb graph $\mathcal{R}$ of $\mathcal{I}$ can be computed in $O(N \log n)$ time.*

Proof: At any given step, G contains an edge (T_1, T_2) if and only if T_1 and T_2 are directly connected at that step. For a given ϵ and resolution of dMRI image, there is constant number of connections possible for a trajectory T_i. So, G forms a sparse graph with $O(n)$ nodes and edges. During event computation step, we know all indices of trajectories at which G changes in advance, we can use therefore maintain the connected components as the maximum weight spanning forest as an ST-tree [20, 21]. this will help in connectivity queries, inserts, and deletes, in $O(\log n)$ time. In the worst case, we modify and query the graph G to get the connected components for all the points in $\mathcal{I}$. Hence, the total time required for the construction of $\mathcal{R}$ is $O(N \log n)$. ■

Theorem 2 *For a given ϵ and a set of trajectories $\mathcal{I}$, where each trajectory consists of at most μ points, Reeb graph $\mathcal{R}$ can have $O(\mu n^2)$ vertices.*

Proof: Consider a pair of trajectories T and T'. T' is ϵ-connected to T during at most one spatial interval which yields two vertices in $\mathcal{R}$. A pair of trajectories produce $O(\mu)$ vertices when connected and disconnected for at most one interval. For all possible pairs n^2, this gives a total of $O(\mu n^2)$ vertices. This bound is tight in the worst case. ■

Theorem 3 *For a given ϵ, α, and a set of trajectories $\mathcal{I}$, where each trajectory consists of at most μ points, the Reeb graph $\mathcal{R}$ can have $O(\frac{\mu n^2}{\alpha})$ vertices.*

Proof: Consider two trajectories T and T'. T' is ϵ-connected to T during at most α spatial interval which yields to two vertices in $\mathcal{R}$. A pair of trajectories thus produces $O(\mu/\alpha)$ vertices. There are n^2 possible pairs giving a total of $O(\frac{\mu n^2}{\alpha})$ vertices. This bound is tight in the worst case. ■

2.8 Summary

In this Chapter, we present our algorithmic theory of Reeb graph-based method for subtrajectory clustering. We provide an implementation of the discussed method, ReeBundle as a computational framework which is available as a Python package. Users can interact with the 3D visualization of the model using the Jupyter notebooks provided as a part of the GitHub code: `https://github.com/UCSB-VRL/ReeBundle`. Reeb graph representations encode topological information that can be used as input in data-driven methods instead of directly using deep learning methods on limited available and noisy raw dMRI data. The algorithm's parameters control the granularity of the bundling process—small parameter values restrict grouping to very dense sets of subtrajectories, while larger values allow for the formation of larger and sparser groups.

Our main contributions in this Chapter are summarized below:

1. We develop an efficient method (ReeBundle) to model the white matter connectivity as a sparse Reeb graph. This tool can be used as a plug-and-play solution for brain connectivity analysis, this will be discussed in detail in future chapters.

2. We parameterize our model to mitigate the effects of noise in streamlines. Noise in this context refers to geometrically implausible streamlines due to curvature overshoot or short length due to premature termination.

Chapter 3

Reeb Graph Models for Tractography

Our proposed Reeb graph method in Chapter 2 is sensitive to the input set of fibers and thus can be used as a quality assurance measure to tweak the pre-processing steps or to decide the fiber tracking methods (tractography). In this Chapter, we utilize the proposed sparse Reeb graphs to rank and compare different tractography algorithms. We design novel Reeb graph matching algorithm to quantify the comparison and rank 96 tractography methods for given neuroanatomical bundles of interest. The existing evaluation metrics such as the f-score, bundle overlap, and bundle overreach fail to account for fiber continuity resulting in high scores even for broken fibers, branching artifacts, and mis-tracked fiber crossing. In contrast, we show that ReTrace effectively penalizes the incorrect tracking of fibers within bundles while concurrently pinpointing positions with significant deviation from the ground truth. Based on our analysis of ISMRM challenge data, we find that no single algorithm consistently outperforms others across all known white matter fiber bundles, highlighting the limitations of the current tractography methods. We also observe that deterministic tractography algorithms perform better in tracking the fundamental properties of fiber bundles, specifically merging and

The content from this Chapter was published in CDMRI Workshop at the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) [22].

splitting, compared to probabilistic tractography. We compare different algorithmic approaches for a given bundle to highlight the specific characteristics that contribute to successful tracking, thus providing topological insights into the development of advanced tractography algorithms.

3.1 Tractography

Tractography is a technique utilized in neuroimaging to reconstruct white matter fiber pathways from diffusion magnetic resonance images (dMRIs). The reconstructed fiber bundles provide valuable insights into the connectivity between different regions of the brain. They play a crucial role in understanding neuroanatomy and studying various brain disorders. For instance, tractography has been used to reveal brain abnormalities across a range of conditions [11, 23, 24], including multiple sclerosis, cognitive disorders, Parkinson's disease, brain trauma, tumors, and psychiatric conditions. To ensure accurate interpretation of the obtained tractography results, it is vital to evaluate the performance of tractography methods on these neuroanatomical bundles and select appropriate metrics for assessment. Evaluating the performance of tractography methods is a complex task due to the intricate nature of white matter pathways and the challenges associated with capturing their neuroanatomical topology [13, 25].

Empirical studies have examined the effectiveness of many tractography methods in investigating a range of neurodegenerative diseases like Alzheimer's [26], as well as in measuring patient outcomes following the utilization of tractography for tumor resection [27]. However, the comparison between tractography algorithms remains mainly qualitative for the most part. These studies do not provide a comprehensive evaluation of the tractography algorithms themselves, as they lack a ground truth reference for the reconstructed bundles. This limitation hinders the ability to quantitatively compare

standard tracking algorithms and obtain meaningful feedback about their performance.

3.2 Evaluation of Tractography

To quantify tractography, the FiberCup phantom dataset [28] is commonly used. The International Society for Magnetic Resonance in Medicine (ISMRM) organized a tractography challenge [19] on revised FiberCup dataset to establish a ground truth and provide score using the Tractometer [29]. Tractometer provides global connectivity metrics and facilitates extensive assessment of tracking outputs, fiber bundle detection accuracy, and incomplete fiber quantification. As bundle analysis is crucial in neurological studies, the tractograms were divided into 25 major bundles in the ISMRM dataset. To assess the performance of tractography method on bundles, bundle coverage metrics were proposed [19]. These metrics transform the fibers into voxel images, which results in the loss of fiber point-correspondence. They fail to account for many reconstruction errors, thereby yielding inflated and potentially misleading scores. For example, the topological complexity arising due to the geometrical structure and branching within valid bundles is often neglected in voxel-based metrics. In other words, the existing tractography assessment metrics do not answer “how" the fibers are connected but only analyze the connection percentages. Hence, there is a critical need for methods that can quantify the anatomical validity of fiber branching, given the complex fiber topology [30].

Topology pertains to the organization of white matter fibers into structural networks of brain. It considers the fibers' origination, termination, and branching, as well as their relationship to different brain regions. Neurological disorders can induce topological changes in white matter fibers, such as physical disruptions in cases like stroke, or spatial distortions as seen in brain tumors. To effectively assess the topology of the fiber reconstruction in bundle tracts, we discuss our method ReTrace [22] in this Chap-

ter. It is a novel graph matching algorithm that is based on the construction of a Reeb graph [7, 31, 32]. This graph matching algorithm enables a comprehensive quantitative analysis of topological connectivity patterns by considering both global and local network features. Further, for each quantitative assessment provided by ReTrace, a graph visualization of the bundle in 3D is also associated. This visualization is crucial in analyzing the efficacy of the tractography. Finally, ReTrace can also be tuned to explore the output of a given tractogram in different resolutions. The implementation code and interactive notebooks for utilizing ReTrace are publicly available on GitHub [1].

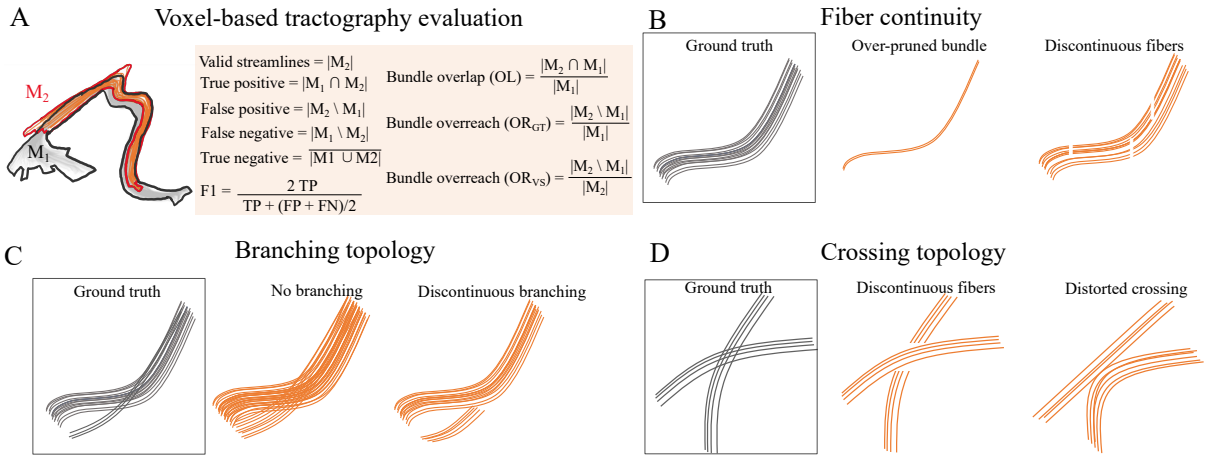

Figure 3.1: Traditional bundle coverage metrics and their limitations in tractography evaluation. (A) Mask M_1 illustrates the ground truth bundle (black outline), while Mask M_2 reveals valid fibers (red outline) produced by a specific tractography method. Bundle coverage evaluation uses these voxel masks. (B)-(D) display scenarios where the traditional metric falls short. Gray fibers represent the ground truth, while orange fibers depict the reconstructed fibers. (B) highlights potential errors in fiber continuity. (C) demonstrates potential inaccuracies in the branching topology of the tracked fibers, and (D) exposes potential misinterpretations in crossing topology, which despite yielding high scores in bundle coverage, might be anatomically implausible.

[1] `https://github.com/s-shailja/ReTrace`

3.3 Limitations of Existing Methods

Quantitative metrics such as global connectivity-dependent metrics in Tractometer [29] and connectivity matrix have been used to evaluate the fiber reconstruction [33]. They give information about the high-level connection of brain regions. However for the assessment of a given bundle tract, they offer a snapshot of valid bundle coverage with their ground truth counterparts and do not account for local topological features. We describe the commonly used metrics in Figure. 3.1A. While the existing metrics are valuable in assessing volumetric bundle coverage they fall short in capturing fiber continuity, branching, and crossing as highlighted in Figure. 3.1B, C, and D respectively in the bundle. Some of these limitations are:

1. Voxel-wise agreement between the algorithm's output and the ground truth is the main emphasis in computing the existing metrics. This overlooks the intricate tractography errors, such as, spurious fibers and incorrect connectivity patterns. Specifically, these metrics often neglect broken fibers as they could still cover the volume despite their discontinuity.

2. Spatial localization is not considered to evaluate the errors between the reconstructed bundles and the ground truth. Similarly, tractography errors cannot be attributed to specific inaccuracies in tracked fibers within the spatial context to guide the design of better algorithms.

3. Branching in the reconstructed pathways is largely inconsequential in the computation of voxel-based metrics. Typically, fiber bundles originate from functional regions, merge into larger pathways for efficiency, and branch out as they approach their respective terminal functional areas. Neuroanatomical studies [13, 31] have pointed out the importance of branching in tractography. But, the conventional

metrics (in Figure. 3.1A) may fail to detect errors related to this branching topology. As such, obscured connections within dense bundles could still achieve high scores despite inaccuracies.

4. Complex fiber orientations such as fiber crossing are not directly captured by the existing tractography evaluation metrics.

Our proposed method addresses these limitations for evaluating tractography methods. It incorporates higher-level tract analysis that considers the topological branching patterns and pathways. Additionally, our method is sensitive to spatial localization, taking into account the precise anatomical locations of inaccuracies within the reconstructed bundles. By addressing the limitations discussed above, our method provides a tunable and robust evaluation of tractography algorithms in terms of accuracy and anatomical fidelity.

3.4 Quantification of Tractography

ReTrace is an end-to-end tractography evaluation pipeline that starts by processing the given white matter fiber bundles to generate a Reeb graph model using our algorithm proposed in Chapter 2. Then, it evaluates the quality of fiber reconstruction by comparing two graphs taking into account many topological factors. The pipeline is illustrated in Figure. 3.2. It is based on the construction of a Reeb graph that provides a topological signature of the fibers. This graph representation makes the tractography evaluation amenable to graph and network theory methods.

3.4.1 Topology of Anatomical Bundles

A Reeb graph representation of white matter fibers has been discussed [7, 31, 32] in our previous Chapter. It provides a concise depiction of trajectory branching structures

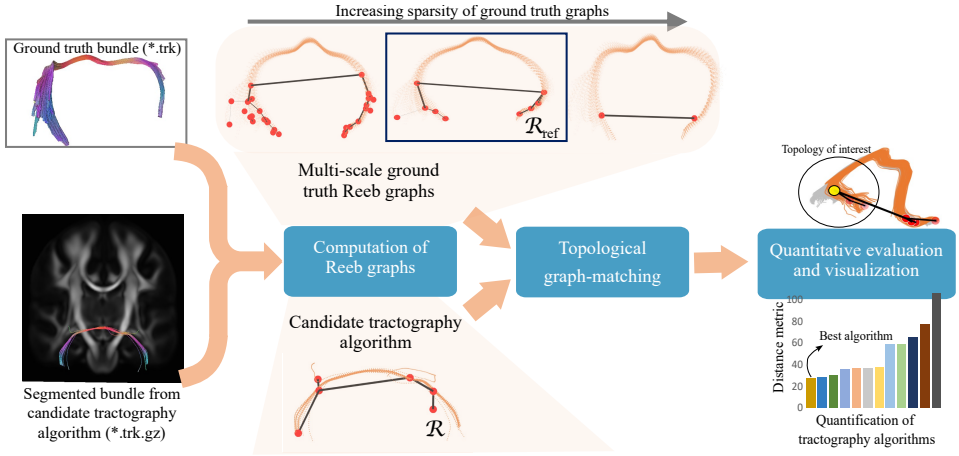

Figure 3.2: ReTrace evaluates tractography methods both qualitatively and quantitatively by comparing two Reeb graphs, $\mathcal{R}$ and $\mathcal{R}_{\text{ref}}$. By setting appropriate parameters, Reeb graphs of different resolutions can be computed from the ground truth data. Among the graphs of different sparsity, one is chosen based on the level at which tractography evaluation is desired. Similarly, for the candidate tractography algorithm under evaluation, a Reeb graph is computed from the bundle using the same Reeb graph parameters as the ground truth. The evaluation metric reflects the distance between the two graphs: a higher metric value indicates larger discrepancies, and the candidate algorithm with the lowest metric is assigned the first rank. For qualitative analysis, one can visually examine the Reeb graphs, locating nodes contributing to the distance.

by constructing undirected weighted graphs. The given bundle tracts (that is, a group of streamlines representing a major bundle) is represented as a graph. The vertices of the graph are critical points where the fibers appear, disappear, merge, or split. Edges of the graph connect these critical points as groups of sub-trajectories of the given bundle. The edge weight is the proportion of fibers that participate in the edge. Three key parameters capture the geometry and topology of fibers:

1. ϵ denotes the distance between a pair of fibers in a bundle, controlling its sparsity. Smaller ϵ values result in denser subtrajectory groups, while larger values allow for sparser groups,

2. α represents the spatial length of the bundle, introducing persistence and influencing the extent of bundling, and

3. δ determines the bundle thickness to shape the model's robustness and granularity.

By adjusting these parameters, we can explore the anatomical fiber structure at different scales of sparsity. For example, the *branching* structure is finely encoded when the inter-trajectory distance, ϵ is less. The interruption tolerance can be enhanced by increasing α. Finally, to ignore spurious fibers and bundles δ can be increased in the Reeb graph. In this Chapter, we analyze the reconstruction of anatomical bundles in different spatial resolutions, demonstrating the potential of graph-based tractography validation and evaluation.

Note: Throughout the thesis, we overlay the Reeb graphs on raw fibers (in orange and ground truth fibers in grey). In the graphs, the nodes are illustrated in red, while edges are shown in black.

3.4.2 Global Network Features

As discussed, representing the tracts as a graph allows the tractography evaluation to be compatible with existing network and graph theory metrics. For a given Reeb graph $\mathcal{R}(V, E)$, we compute the global network properties that provide a comprehensive view of the network's structure and behavior. Global features can be as simple as number of nodes ($|V|$), number of edges ($|E|$), or average degree ($2E/N$) to more sophisticated metrics such as diameter, assortativity, modularity [34], or transitivity depending on the specific application. These network properties provide valuable insights into the structural characteristics of anatomical bundles represented by the Reeb graph. For example, in the Reeb graph, diameter signifies the longest sequence of merging and splitting events along the trajectories. However, global network-based metrics lack specificity and can be challenging to trace back and localize. Therefore, we introduce a novel graph matching algorithm that uses spatial location and provides a quantified distance between the

ground truth and computed fiber bundles using local node features.

Algorithm 2 Topological Distance Computation

```
function OneWayDistance(R, R_ref, ϵ, γ)
    I ← Vref                                              ▷ Nodes to be inserted
    D ← ∅                                                 ▷ Nodes to be deleted
    d_edit ← 0                                            ▷ Score
    d_pos ← 0                                             ▷ Spatial score
    d_net ← 0                                             ▷ Network score
    for n_c ∈ V do
        c ← closest node in V_ref
        if d(n_c["pos"], n_r["pos"]) < 2ϵ then
            Remove c from I                               ▷ Successful correspondence
            if d(n_c["pos"], n_r["pos"]) < ϵ then
                continue                                  ▷ Equivalence (nothing added)
            d_pos ← d_pos + d(n_c["pos"], n_r["pos"])
                                                          ▷ Substitution
            d_net ← d_net + d(n_c["net"], n_r["net"])
        else
            Add c to D
    d_pos ← d_pos + |I| · 2ϵ(1 + γ)                       ▷ Insertion
    for n_i ∈ I do
        d_net ← d_net + d(n_i, 0)
    d_pos ← d_pos + |D| · 2ϵ(1 + γ)                       ▷ Deletion
    for n_d ∈ D do
        d_net ← d_net + d(n_d, 0)
    d_edit ← d_pos/|V_ref| + d_net
    return d_edit
function Distance(R, R_ref, ϵ, γ)
    s_cmp ← OneWayDistance(R, R_ref, ϵ, γ)
    s_ref ← OneWayDistance(R_ref, R, ϵ, γ)
    return 0.5 · (s_cmp + s_ref)
```

3.4.3 Topological Graph Matching

For any given node $v \in V$, we calculate two sets of features — spatial position-based features, denoted as “pos", and local network-level features, denoted as “net". Each node of the Reeb graph is linked to its 3D spatial location in the brain, so “pos" corresponds to the 3D coordinate yielding a 3-dimensional feature. On the other hand, the network features are computed using centrality metrics at the node level: degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality [35], producing a 4-dimensional feature.

Our proposed graph matching computation algorithm is adapted from Siminet [36], enhanced to accommodate additional parameters for robustness against noise in trac-

tography and to incorporate network metrics with spatial position as node features. It calculates an edit distance denoted by d_{edit} between two Reeb graphs, quantifying the cost of operations needed to transform one graph into another, as outlined in Algorithm 2.

The algorithm begins by iterating through the nodes of the comparison graph $\mathcal{R}$ and matching each node with its closest spatial counterpart within a search radius. The algorithm determines node correspondences by comparing the spatial distances ("pos") between nodes and their counterparts against a threshold ϵ (inter-fiber distance used in constructing the Reeb graph). The algorithm also accounts for insertions and deletions by penalizing them with a score γ proportional to the Euclidean distance between the centroids of node locations for $\mathcal{R}$ and $\mathcal{R}_{\text{ref}}$. The proposed metric d_{edit} is computed considering the Euclidean distance in node attributes: spatial positions and centrality metrics to compare graph-based representations of white matter bundles, ensuring that layout similarities correspond to similarities in brain regions.

Limitations: Note that the metric is not normalized, that is, it does not have the standard 0-1 limits. We opt to keep the distances between graphs unbounded to account for relative variations (in some cases, the distance could be very large). The ultimate aim of a new tractography design would be to minimize this distance. A potential limitation, or rather a feature, is that the user needs to select the resolution parameters based on the desired comparison of the bundle tracts. While this might appear a manual step, the preset parameters generally serve well as a great starting point for all the major bundle tracts. Also, the performance of all tractography evaluation metrics, both existing and our proposed metric, depends on the quality of bundle segmentation. Hence, further research on bundle segmentation could enhance metric performance.

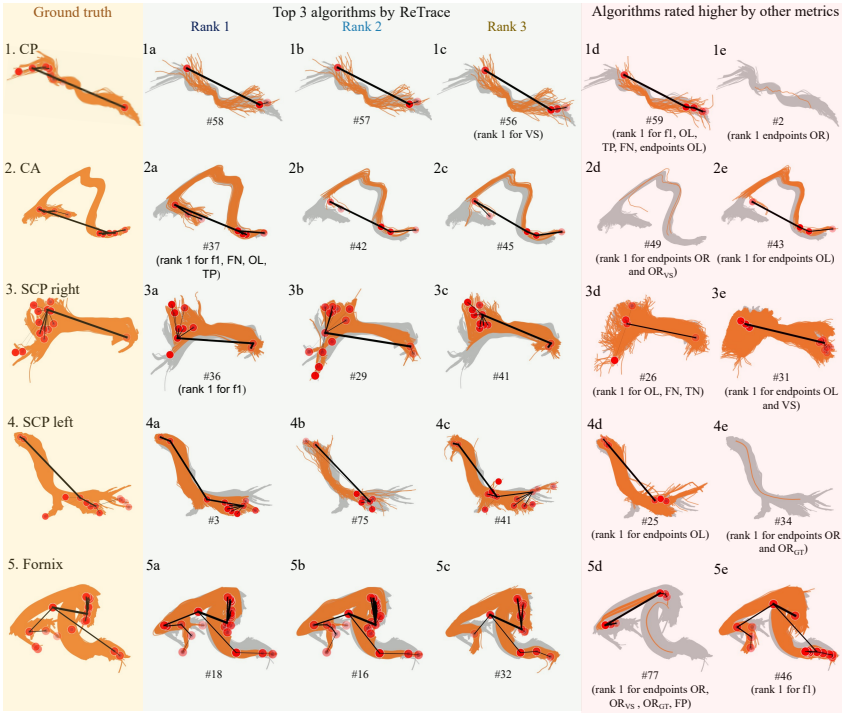

Figure 3.3: Evaluation of tractography algorithms using ReTrace for bundles in the ISMRM data. The leftmost column displays the ISMRM ground truth fiber bundles with associated Reeb graphs overlaid on the bundles. For the other columns, grey streamlines denote the ground truth, overlaid by orange streamlines representing the candidate tractography algorithm's results. On top of these, the Reeb graphs are overlaid with red nodes and black edges. The bundles illustrated in the figure are labeled from 1 to 5: CP, CA, $\text{SCP}_{\text{right}}$, SCP_{left}, and Fornix respectively. For each bundle (in a row), the next three columns (a-c) highlight the algorithms that achieved the highest rank according to our proposed metric. Notice that the algorithms that are ranked high on existing metrics (displayed in columns d-e), do not show the best tracked bundle. The reconstructions that do not capture the branches near the end of the bundles are ranked highly by other metrics. Similarly, fiber continuity and distorted fibers in the reconstructed output are given high ranks with existing metrics.

3.5 Results on ISMRM Dataset

3.5.1 The ISMRM 2015 Tractography Challenge

The challenge presented participants with a clinical-style dataset: a 2mm isotropic diffusion acquisition with 32 gradient directions and a b-value of 1000 s/mm^2. The task was to reconstruct fiber pathways using a realistically simulated replication of a whole-

brain diffusion-weighted MR image. The challenge resulted in 96 tractogram submissions, available publicly for download[2]. These submissions collectively represent a broad range of tractography pipelines, encompassing varied pre-processing, tractography, and post-processing algorithms. This provides a diverse platform for quantitative and qualitative analysis of tractography methods. We assign each algorithm a unique algorithm number, ranging from 1-96 (see Figure. 3.5 for the mapping of these ids to the original submission numbers).

For tractography evaluation, we compute the valid bundles for each submission using the ROI-based bundle segmentation system proposed by the challenge organizers [37]. An expert segmented ground truth bundle segmentation provided by the challenge allowed us to assess the submitted tractography methods based on traditionally computed metrics, establishing a baseline for comparison. The extracted bundles were then processed using the ReTrace pipeline. We can fine-tune the computation of Reeb graphs using the robustness parameters (ϵ, α, and δ), adjusting the granularity of the desired analysis. In this study, these are set to $\epsilon = 2.5$, $\alpha = 5$, and $\delta = 5$. The results for various bundles using our proposed metric compared with the traditional bundle coverage metrics are illustrated in Figure. 3.3. When the candidate tractography has less than 5 fibers, Reeb graphs are not constructed (automatically with the Reeb graph parameter $\delta = 5$) and the algorithm is rejected without rank assignment. For a more dense analysis δ may be set to 0. The comparison of existing bundle coverage metrics against our proposed d_{edit} metric is presented in Table 3.1.

All submissions faced difficulties reconstructing the smaller bundles, such as the anterior (CA) and posterior commissures (CP), which possess a cross-sectional diameter of no more than 2 mm. Due to the minimal branching owing to fewer recovered fibers, ReTrace performs similar to the existing methods on these bundles. However, as the

[2]https://zenodo.org/record/840086

Bundle	FN		FP		OL		OR_{GT}		OR_{VS}		TP		VS		Endpoints OL		Endpoints OR		f1		d_{edit}	
	#	val	#	val	#	val	#	val	#	val	#	val	#	val	#	val	#	val	#	val	#	val
CP	59	1134	75	38	59	0.29	75	0.02	32	0.42	59	474	56	34	59	14	42	1	59	0.33	58	40.1
CA	37	603	60	72	37	0.59	60	0.05	49	0.43	37	876	38	75	43	22	49	2	37	0.52	37	7.3
SCP right	26	471	44	6	26	0.88	44	0	44	0.08	26	3467	57	1768	57	242	44	1	36	0.57	36	10.6
SCP left	26	440	34	12	26	0.89	34	0	77	0.05	26	3638	57	2245	25	352	34	0	16	0.63	3	10.9
Fornix	26	2698	77	40	26	0.75	77	0	77	0.07	26	7971	25	2595	25	698	77	3	46	0.57	18	8.7

Table 3.1: Comparison of d_{edit} with existing bundle coverage metrics. The columns labeled as "#" represent the rank 1 algorithm number, while "val" denotes the respective metric values for each metric. Top-ranking methods according to d_{edit} differ from those determined by traditional metrics for most bundles. We demonstrate that the reconstruction ranked the best using d_{edit} effectively captures branching topology that is overlooked in evaluations using existing methods.

number of fibers increases for bundles like $\text{SCP}_{\text{right}}$, SCP_{left}, and Fornix (as shown in Figure. 3.3), the importance of accurately branching towards the end of the fibers became apparent. This leads to different algorithms being top-rank with our method as compared with existing metrics. As an example, while algorithm 36 and 29 achieve high ranks for the ReTrace evaluation in $\text{SCP}_{\text{right}}$, they do not perform as well according to the existing metrics. It is notable that simply increasing fiber count to cover the bundle volume without appropriate branching does not lead to a high ReTrace ranking. This is evident in algorithms 26 and 31 (Figure. 3.3d and 3.3e) as they are ranked high according to OL, FN, TN, endpoints OL, and VS metrics, but do not contribute significantly to the ReTrace ranking. The evaluation of all other bundles in the ISMRM dataset using d_{edit} along with their interactive visualizations are available on our GitHub repository. Beyond the topological distance that we calculate here, various other graph attributes such as variations in the number of nodes, edges, average degree, and diameter can also be computed and are available in our repository [3]. They collectively capture different facets of the graphs, and as a consequence, highlight various properties of the tractography reconstruction.

Different tractography methods were employed by different teams. The most im-

[3] `https://github.com/s-shailja/ReTrace`

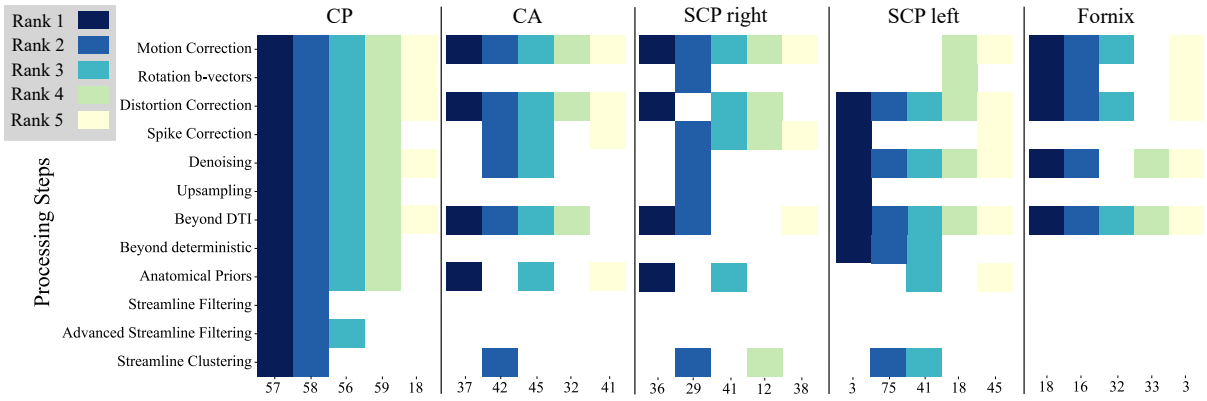

Figure 3.4: Correlation between processing steps and the successful reconstruction of the major bundles as assessed by ReTrace. Different colors represent the ranking of methods according to the ReTrace pipeline for five different bundles (rank 1 (dark blue) to rank 5 (yellow)). Y-axis indicate various steps of tractography: Preprocessing (from motion correction to upsampling); Tractography (diffusion modeling beyond DTI such as constrained spherical deconvolution, and tractography methods beyond deterministic approaches such as probabilistic tractography); Postprocessing (incorporation of anatomical priors to streamline clustering). The importance of each step is evident with how prevalent it is in the top ranked algorithms for each bundle.

portant preprocessing, postprocessing, and tractography steps among top-performing algorithms are shown in Figure. 3.4. For example, CP proved difficult to reconstruct for all algorithms, necessitating the use of extensive denoising and correction methods, as well as probabilistic methods for fiber tracking. Conversely, for other bundles, minimal preprocessing was required and deterministic fiber tracking was sufficient for accurate branching. The insights on the steps of the algorithms offer a topological perspective on designing advanced tractography algorithms and pipelines for future studies. We observe that none of the algorithms perform consistently well for all the bundles. So, an additional advantage of this exploration is that new tractography algorithms may be designed, tailored to specific bundles or neuroanatomical properties of interest.

The results presented in this Chapter are in contrast with the ranking of tractography algorithms discussed in the existing literature [19,37] but also confirm the findings of the importance of various steps in tractography [33]. The runtime of the pipeline depends

on the number of fibers and the desired resolution of the Reeb graph. After computing the graphs for a bundle, our evaluation method takes less than 120 seconds to obtain the results on an Intel Xeon CPU E5-2696 v4 @ 2.20 GHz. The mapping of the algorithm IDs in this Chapter to the original submission IDs in the Tractography Challenge submission can be found in Figure. 3.5

3.5.2 Evaluating tractography algorithms in fiber-crossing regions

Tractography's accuracy is influenced by the number of distinct fiber orientations per voxel, as fiber-crossing regions pose considerable challenges. A probabilistic white matter atlas highlights these areas with high fiber-crossing frequencies [30]. Conventional metrics may not fully capture an algorithm's ability to accurately resolve these complex regions, instead reflecting its capacity to generate copious or elongated fibers, regardless of their neurological plausibility or accuracy.

For example, a tractography algorithm might generate spurious fibers connecting distinct bundles in crossing fiber scenarios, falsely inflating the fiber count. This inflation could distort performance evaluations if they rely solely on fiber numbers. Endpoint-matching metrics, such as overlap or intersection between reconstructed and ground truth fibers, could also be misleading in the presence of crossing fibers: overlaps may not indicate accuracy if the algorithm incorrectly generates disrupted or distorted fibers due to crossing fibers. Spatial accuracy metrics, like the average or Hausdorff distance [38] between reconstructed and ground truth fibers, could also be skewed by crossing fibers. These regions may compromise the algorithm's capability to accurately resolve individual bundles, resulting in spatial inaccuracies or increased distances.

Submission ID	Algorithm ID
1_0	1
1_1	2
1_2	3
1_3	4
1_4	5
2_0	6
3_0	7
3_1	8
3_2	9
3_3	10
3_4	11
4_0	12
5_0	13
5_1	14
6_0	15
6_1	16
6_2	17
6_3	18
6_4	19
7_0	20
7_1	21
7_2	22
7_3	23
8_0	24
9_0	25
9_1	26
9_2	27
9_3	28
9_4	29
10_0	30
10_1	31
10_2	32
10_3	33
10_4	34
10_5	35
10_6	36
10_7	37
10_8	38
10_9	39
10_10	40
10_11	41
10_12	42
10_13	43
10_14	44
10_15	45
10_16	46
10_17	47
10_18	48
10_19	49
11_0	50
11_1	51
12_0	52
12_1	53
12_2	54
12_3	55
13_0	56
13_1	57
13_2	58
13_3	59
14_0	60
14_1	61
14_2	62
15_0	63
16_0	64
16_1	65
16_2	66
16_3	67
16_4	68
17_0	69
17_1	70
17_2	71
17_3	72
17_4	73
18_0	74
18_1	75
18_2	76
18_3	77
18_4	78
19_0	79
19_1	80
19_2	81
20_0	82
20_1	83
20_2	84
20_3	85
20_4	86
20_5	87
20_6	88
20_7	89
20_8	90
20_9	91
20_10	92
20_11	93
20_12	94
20_13	95
20_14	96

Figure 3.5: Mapping of algorithm IDs to the original submission IDs.

HCP Dataset

The efficacy of ReTrace is not limited to synthetic data alone. To demonstrate its applicability to real datasets and to show how Retrace handles fiber crossing, we use the average HCP 1065 template, constructed from the diffusion MRI data of 1065 subjects from the Human Connectome Project (HCP)[4]. We use the 1 mm population-averaged FIB file in the ICBM152 space for fiber tracking in DSI Studio.

ReTrace handles fiber crossings effectively, as demonstrated in Figure. 3.6. We select a small region of interest (a 2mm isotropic 3D region) from the probabilistic atlas in an area with a high probability (~ 0.9) of double-crossing fibers. We use the deterministic streamline tracking method implemented in DSI Studio[5] to compute the fibers with the parameters set (angular threshold, step size, min length, max length, terminate if seeds, iterations for topological pruning) to 35, 1, 70, 200, 1000, and 16, respectively. This allowed us to observe successful tracking without broken or distorted fibers. To mimic the spurious broken fibers that tractography methods may yield, we set the parameters to 35, 1, 0, 100, 1000, and 16. For observing the angular distortion where the fiber bends and follows a different path, we set the parameters as 90, 1, 70, 100, 1000, and 16. The resulting Reeb graphs clearly highlight how their nodes capture successful and unsuccessful tracking. Nodes formed near intersections indicate broken or bent fibers, as shown in Figure. 3.6. Whenever fibers travel together in a group, they form an edge in the graph. Any alteration within this group prompts a critical event, resulting in a node in the Reeb graph. Consequently, if a fiber breaks or diverges, the associated group changes, generating a node. This node, present at the merging point, contributes to a larger distance value. In the topological distance computation, d_{edit}, this node could be weighted more if the goal is to assess an algorithm's tracking ability in ambiguous fiber

[4] `https://brain.labsolver.org/hcp_template.html`
[5] `https://dsi-studio.labsolver.org/`

orientations. By providing a 3D location for our algorithm's attention, any discrepancy within that location can significantly affect the overall d_{edit} computation. The code is open source and can be tailored to specific needs.

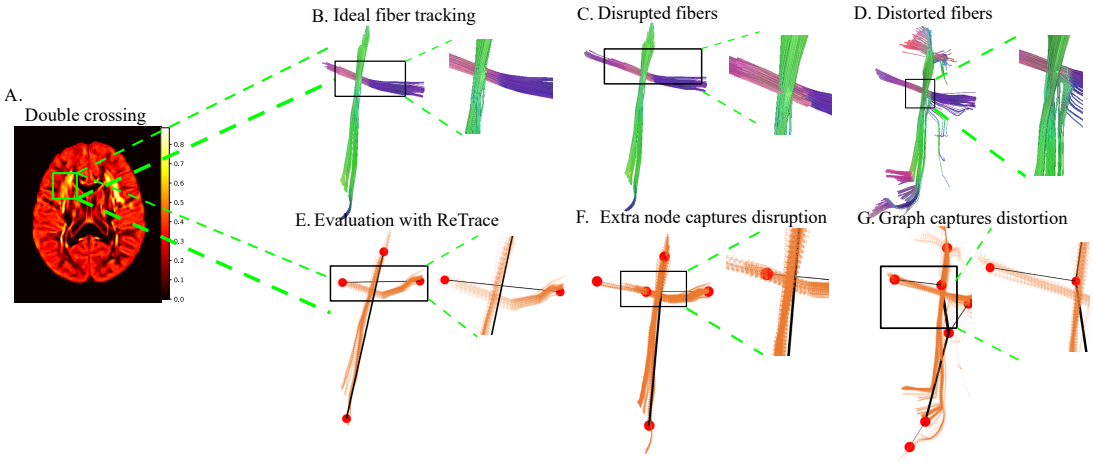

Figure 3.6: Evaluation of a tractography algorithm for fiber crossing. (A) shows the center slice of the probabilistic double crossing atlas. A small ROI with high double crossing fiber probability (shown in green) is selected. (B)-(D) demonstrate the reconstruction of fibers using the streamline method implemented in DSI Studio. (B) shows successful tracking of fiber crossings with ideal parameters. (C) shows fibers that terminate abruptly near the crossing as a result of reconstruction with different parameters. (D) shows bending or distortion of fibers when encountering multiple diffusion orientations. (E)-(G) present the constructed Reeb graphs overlaid on the computed fibers for each case, respectively (shown below each reconstruction). An additional node is observed in the Reeb graph that captures the fiber crossing or bending. A thicker edge in (G) indicates bending of the fibers, whereas ideally, the fiber should only cross without bending.

3.6 Summary

This chapter introduces ReTrace, an innovative evaluation method for tractography algorithms. This method addresses existing metrics' limitations by focusing on the topological accuracy of reconstructed pathways. We applied ReTrace to both synthetic and real-world datasets to demonstrate the branching fidelity and errors such as broken or bent fibers in tractography reconstruction. With the ISMRM dataset, we ranked 96 al-

gorithms on different neuroanatomical bundles. The rankings proposed by our method are in contrast with the rankings using the conventional voxel-based tractography metrics. We discuss this difference in ranking and performance of different algorithms by highlighting the topological features such as branching, fiber continuity, localization, and crossing. We demonstrated the utility of our method on the HCP dataset where we show the importance of reconstructed fiber crossing and discuss the performance of standard tractography algorithms. Our approach is disease-agnostic, does not require brain registration to an atlas, and works across different acquisition protocols. It is important to note that bundle segmentation is a common bottleneck in any evaluation metric. Therefore, advancements in segmentation research could greatly enhance these evaluations. The results on tractography comparison presented here could be extended to be used as a cost function for data-driven machine learning methods, like generative adversarial networks. With feedback from neuroscientists, we hope that the results in this Chapter will pave the way forward in improving existing tractography methods.

Chapter 4

Reeb Graph Models for Human Brain

Tractography as discussed in the previous Chapter can generate millions of complex curvilinear fibers (streamlines) in 3D that exhibit the geometry of white matter pathways in the brain. In this Chapter, we utilize the theory developed in Chapter 2 to efficiently encodes the topology and geometry of white matter fibers. Given the trajectories of neuronal fiber pathways (neuroanatomical bundle), we re-bundle the streamlines by modeling their spatial evolution to capture geometrically significant events (akin to a fingerprint). We demonstrate this using two datasets: (1) International Society for Magnetic Resonance in Medicine (ISMRM) dataset, ReeBundle handles the morphology of the white matter tract configurations due to branching and local ambiguities in complicated bundle tracts like anterior and posterior commissures; (2) For the longitudinal repeated measures in the Cognitive Resilience and Sleep History (CRASH) dataset, repeated scans of a given subject acquired weeks apart lead to provably similar Reeb graphs that differ significantly from other subjects, thus highlighting ReeBundle's potential for clinical fingerprinting of brain regions.

A critical element in neurological and developmental disorders is the topological dete-

The content from this Chapter was published in IEEE Transactions on Medical Imaging (TMI) [8].

rioration and irregularities in streamlines. Reeb graph and topological distance metric are sensitive to variations in brain structure, such as those observed in the developing brain and in the presence of tumors. This sensitivity is beneficial in the context of longitudinal studies of brain development and the topological evolution of the brain. In this Chapter, we have highlighted the potential utility of this metric in tracking Alzheimer's disease progression using the ADNI dataset, and evaluating the effects of surgical interventions for brain tumor using the OpenNeuro brain tumor dataset.

4.1 Existing Brain Network Models

It is important to characterize the underlying brain network structure to understand the relationship between changes or differences in network connectivity and clinical syndromes, disorders, and diseases. Existing research on modeling brain networks can be categorized into 1) synaptic networks mapped at a microscopic scale where nodes represent neurons and synapses represent edges, and 2) whole brain connectome at a macroscopic scale using diffusion magnetic resonance imaging (dMRI). Synaptic network modeling [39, 40] is desirable and informative but it is almost impossible to access for human brains. It is also challenging for the study of structural connectomics and analyses of neurological diseases due to the underlying anatomic complexity. On the other hand, connectome analysis [41] is a pragmatic strategy for human brain studies. Connectomics typically generates a connectivity matrix by partitioning the brain cortex into a limited set of regions derived from anatomical or computational brain atlases. The streamlines within each brain region are lumped into a single node in this representation to model their interconnections as shown in Figure. 4.1A. Hence, the dimension of a square connectivity matrix is equal to the number of regions of interest (usually $\sim$100). Connectivity matrices can provide measures of structural connectivity of the entire brain [42]. A key

limitation of the connectivity matrix is that the underlying topology of white matter bundles, as they traverse the brain, is lost in the dimensionality reduction process. The pros and cons of the two extremes for representing connectivity pose a question of whether a medium-scale representation is plausible. Can there be a medium-scale computational model of a brain connectivity network that is sparse but also preserves the topological information of the white matter connections?

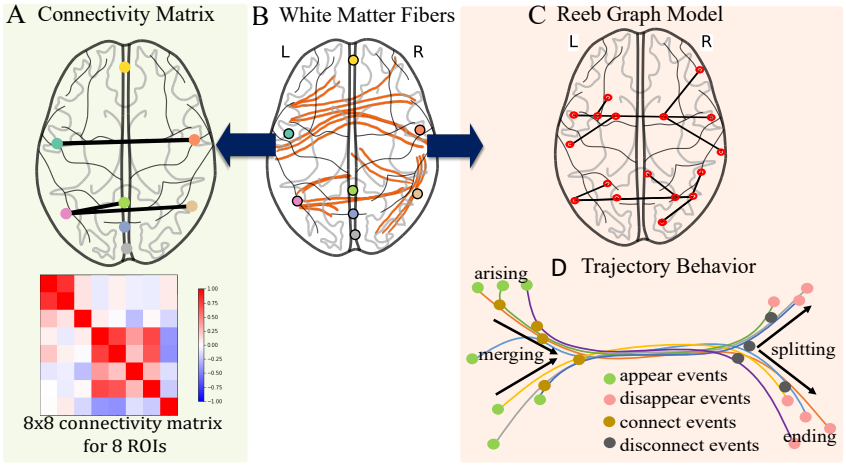

Figure 4.1: **Modeling of white matter fibers using dMRI.** A) Connectivity matrix represents interconnection between different regions of interest (ROIs). B) White matter fibers (streamlines) connecting different brain regions. C) Reeb graph allows the relevant geometric and topological structure of the streamlines to be recovered that is ignored in the connectivity matrix. D) A basic example of a set of trajectories displaying the arising, merging, splitting, and ending behaviors. These qualitative behaviors of a group of trajectories emerge due to the events (appear, connect, disconnect, and disappear) of individual trajectories that are encoded as nodes in the Reeb graph.

Tractometry and tract profiling methods are the ways towards medium-scale analysis that are based on bundle tracts as the unit of comparison. For example, DSI Studio's "Automatic Tractography" [43] uses the topology information from tractography streamlines to derive scalar values of their shape descriptors. However, properties that vary along the tract trajectory are typically averaged over the entire length of the tract. In contrast, tract profiling techniques like DIPY [44], BUAN [45], and PyAFQ [46] quantify

diffusion measures like fractional anisotropy, mean diffusivity, radial diffusivity, and axial diffusivity along bundle tract at multiple locations along the trajectory. These methods are generally used for statistically comparing groups but they do not consider the shape and geometry, especially the branching behavior of the tracts. To account for the shape of the tracts, envelope and contour-based skeleton methods [47, 48] have been proposed. These methods construct a single representative streamline that traverses the tract's entire center of mass. With this approach, tracts can be localized and shape features can be extracted for tract-based analyses. But, to model the skeletons of tracts, a carefully pruned input bundles is often required. Moreover, these methods produce a single core streamline that ignores the local branching structure of the tracts. In this Chapter, we discuss our proposed method, ReeBundle. It is a Reeb graph-based method that enables structure discovery of bundle tract without oversimplifying them into a single skeleton.

To computationally model brain structural connectivity at a medium scale, we propose a hypothesis-driven computational geometry approach. We model the geometry of the tract's trajectories using the concept of Reeb graphs [49]. These have been successfully used in a wide variety of applications in computational geometry and graphics, such as shape matching, topological data analysis, simplification, and segmentation.

We introduced ReeBundle as a method for characterizing white matter topology in Chapter 2. In this chapter, we will use Reeb graphs to quantify the topological differences between brains. We adapt the Siminet [36] graph distance metric to introduce a graph matching technique to compute the distance between two Reeb graphs. Topological quantification using Reeb graphs is important to model neurological disorders and fingerprint dMRI scans for an individual. We discuss these aspects in this Chapter. The source code for the data analysis and the ReeBundle Python tool are available on GitHub [50]. In Figure. 4.2, we present the proposed step-by-step pipeline to apply the ReeBundle method to any brain image diffusion MRI (dMRI) dataset.

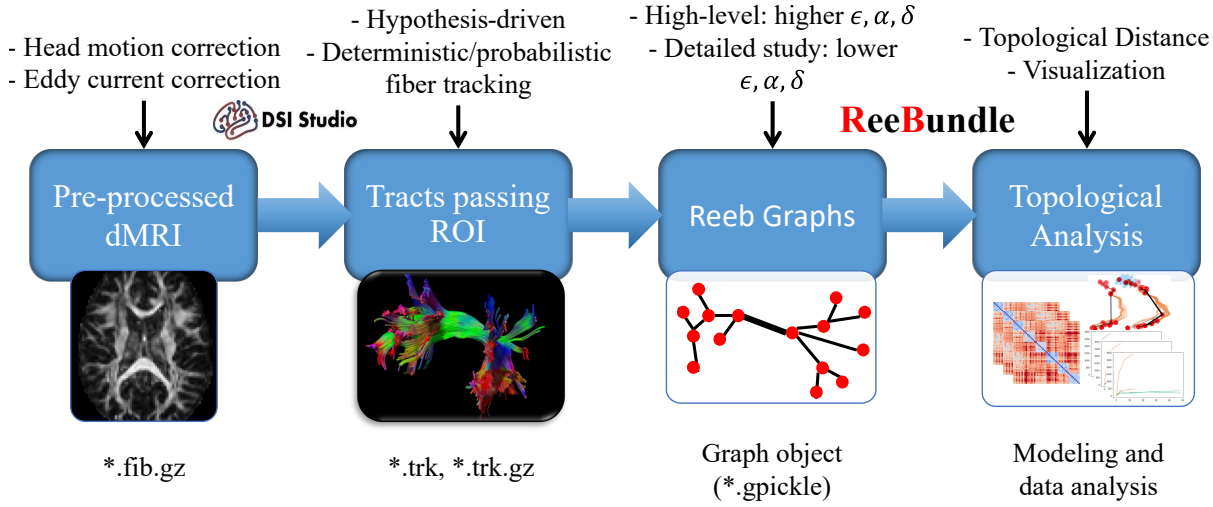

Figure 4.2: **Overview of the proposed pipeline for analyzing dMRI data.** Step 1: The proposed algorithm processes the dMRI data. Step 2: After standard pre-processing, fibers are tracked in a region of interest (ROI) chosen based on hypothesis using DSI Studio. Step 3: The proposed ReeBundle method is applied to the fiber bundle to discover the grouping structure of the trajectory in the form of Reeb graphs. Step 4: The resulting Reeb graphs are compared and analyzed to study and track disease.

4.2 Topological Distance Between Neuroanatomical Bundles

A Reeb graph representation of white matter fibers allows for the application of graph theory algorithms in order to assess the statistical and topological properties of networks reconstructed from dMRI data. Applications of these algorithms have revealed many properties about brain networks [51]. Any graph distance metric [52] can be adapted for the purpose of Reeb graph distance computation but our proposed algorithm is motivated by Siminet [36] since it takes into account the node location. This is important, especially in the context of the intended application to compare graph-based representations of white matter pathways — if two graphs are similar in layout but correspond to completely different locations in the brain, then they should not be considered similar. The key difference is that the original Siminet algorithm did not consider any additional

parameters for robustness against noise, while the one described here does.

The edit distance denoted as d_{edit}, intakes two graphs – the comparison graph $R(V, E)$ and the reference graph $R_{\text{ref}}(V_{\text{ref}}, E_{\text{ref}})$. It measures the overall cost of the operations (node shifting) to transform the comparison Reeb graph (R) into the reference Reeb graph (R_{ref}). The graph matching distance denoted by $d_{\text{edit}}(R, R_{\text{ref}})$ returns a scalar representing the *node score*, which is computed based on node attributes, here, spatial position. Node attributes may include the degree or other desired network or anatomical properties of the node.

The node score is computed by forming unique node correspondences between R and R_{ref}, on the basis of proximity as shown in Figure. 4.3. That is, the algorithm iterates through nodes in R, attempting to match each node with its closest spatial counterpart in R_{ref} within some search radius (also referred to as the *substitution radius*) of size 2ϵ. Any R_{ref} node that is outside the search radius cannot be considered as a potential counterpart for a given R node. A node-node "correspondence" indicates that a node in R_{ref} can be obtained by spatially shifting its counterpart in R by some distance. This is similar to the notion of "node substitution" used by the Siminet algorithm. For example, if R consisted of only two nodes $p_0 = (0, 0, 0)$ and $p_1 = (0, 0, 5)$, and R_{ref} had two nodes at $q_0 = (-1, 0, 0)$ and $q_1 = (0, 0, 10)$, then p_0 would correspond to q_0 as it is its closest option in R_{ref}. Similarly, p_1 would correspond to q_1. Ideally, if R and R_{ref} are the same graphs, then every node in R matches with itself.

The limited search radius for counterparts as well as the uniqueness criteria means that some nodes may not have counterparts. Nodes in one graph get to be associated with nodes in another on a "first come, first served" basis. So, the order of visitation of nodes potentially affects the score. For example, if we have a single node in R_{ref} that spatially sits in the middle between two nodes in R, and all other nodes in R_{ref} are very far away, then the node that we visit first in R gets to take the single R_{ref} node as its

counterpart. Any node in R that fails to obtain a counterpart is marked for *deletion*. On the other hand, any node in R_{ref} that fails to get selected as a counterpart is marked for *insertion*. The issue with the order of visitation is that there is a risk of "outlier nodes" in R. That is, nodes in R that are very far away from nodes in R_{ref} (relative to other nodes in R) latching onto nodes in R_{ref} as counterparts. This situation drives up the node score. However, the limited search radius can prevent this to an extent, restricting the potential counterparts for these outlier nodes to those that are very close to them. Additionally, if we have the scenario where several nodes in R surround a single node in R_{ref} which falls in their search radii, then the order of visitation could lead to a fluctuation in node score, and potentially drive it up. The small search radius yet again mitigates this issue, as the area of overlap between search regions is much smaller than the search regions themselves.

Ultimately, the process of finding node counterparts leads to the creation of several sets – the deletion set D for nodes in R that cannot find a counterpart, the insertion set I for nodes in R_{ref} that cannot find a counterpart, and the set of node correspondences $\mathcal{C} \subseteq V \times V_{\text{ref}}$. The total node score is computed as a sum of values derived from these three sets. For each node correspondence, the spatial distance between a node and its counterpart is compared against ϵ and the substitution radius. This helps to determine the value of the accumulated node score. If the spatial distance is less than ϵ, then the nodes are considered *equivalent* and the node score remains unchanged. This helps guard against small deviations in spatial positions. If the distance is greater than ϵ, then the spatial distance itself is added to the node score, indicating that the node in R can be shifted by this amount to obtain its R_{ref} counterpart. This scoring mechanism is described by the following function on node-node correspondences ($c \in \mathcal{C}$) for two nodes $n_c \in V$ and $n_r \in V_{\text{ref}}$, where $d(n_c, n_r)$ denotes Euclidean distance between the location of n_c and n_r:

$$c(n_c, n_r) = \begin{cases} 0, & d(n_c, n_r) < \epsilon \\ d(n_c, n_r), & d(n_c, n_r) \geq \epsilon, \end{cases}$$

$$d(n_c, n_r) = \|n_c - n_r\|_2,$$

where $\|\cdot\|_2$ represents the Euclidean norm. Total graph edit distance consists of the total substitution score ($d_{\text{substitution}}$), total deletion score (d_{deletion}), and total insertion score ($d_{\text{insertion}}$):

$$d_{\text{substitution}} = \sum_{n_c \in V, n_r \in V_{\text{ref}}} c(n_c, n_r),$$

$$d_{\text{deletion}} = |D| \cdot 2\epsilon(1+\gamma),$$

$$d_{\text{insertion}} = |I| \cdot 2\epsilon(1+\gamma),$$

$$d_{\text{edit}}(R, R_{\text{ref}}) = d_{\text{substitution}} + d_{\text{deletion}} + d_{\text{insertion}},$$

where D is the set of nodes to be deleted and I is the set of nodes to be inserted. γ in the insertion/deletion score is proportional to the Euclidean distance between the centroids of node locations for R and R_{ref}. If the centroids of V and V_{ref} are g and g_{ref} respectively, then $\gamma = \frac{1}{30} d(g, g_{\text{ref}})$. Here, γ helps in penalizing insertion/deletion more than substitution and at the same time takes into account the physical distance between the location of the Reeb graphs R and R_{ref}. In order to compare R and R_{ref}, we measure the topological distance (TD) between the two as the average edit distance when R is matched to R_{ref} and when R_{ref} is matched to R:

$$TD = \frac{1}{2}(d_{\text{edit}}(R, R_{\text{ref}}) + d_{\text{edit}}(R_{\text{ref}}, R))$$

Figure. 4.3 summarizes the algorithm. TD score is zero when we compare the same

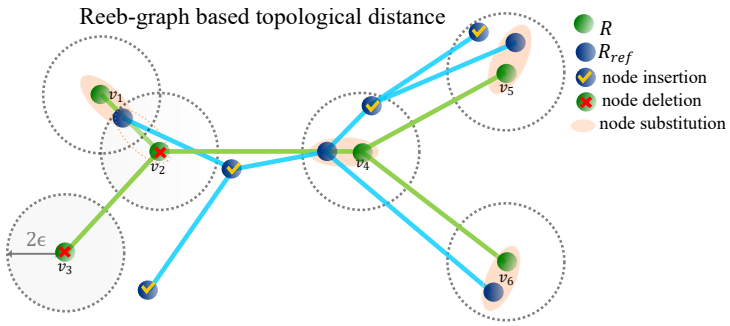

Figure 4.3: **Computation of topological distance.** $\mathcal{R}$ and $\mathcal{R}_{\text{ref}}$ are two Reeb graphs. We compute the topological distance TD $(\mathcal{R}, \mathcal{R}_{\text{ref}})$ by finding the node-correspondence and the nodes that will be inserted or deleted. A dotted gray outline around a node in $\mathcal{R}$ represents the search area of radius 2ϵ for the nodes corresponding to $\mathcal{R}_{\text{ref}}$. The nodes in $\mathcal{R}$ that need to be deleted are shown with red crosses. The nodes that need to be inserted from $\mathcal{R}_{\text{ref}}$ are shown with yellow ticks. The nodes of $\mathcal{R}$ are numbered in the order of traversal. The node correspondence is shown as a solid orange oval.

bundle tracts or eventually the same Reeb graphs. TD score increases when the bundle tracts are different representing the difference in the critical points encoded as nodes of the Reeb graphs. Later, in Sec. 4.3, we show the sensitivity analysis of the TD score with respect to the parameters introduced in Sec. 2.6.

Theorem 4 *For a given pair of graphs – R_{ref} which has M_{ref} nodes and a R which has M nodes – the total time required to run the comparison algorithm on these two graphs is $O(M_{ref}M)$, while the total space needed is $O(M)$.*

Proof: For each node in R, the potential counterparts are found by filtering R_{ref} for nodes that are within the substitution radius and are not in the counterparts set taking $O(M_{\text{ref}})$ time for each node. Once a R node selects a R_{ref} node as its counterpart, it adds it to the set of counterparts. For each correspondence, a substitution/deletion score is computed and added to the total substitution score. To compute the insertion scores, we iterate through the R_{ref} nodes and check if they are members of the counterparts set. This takes $O(M_{\text{ref}}M)$ time, and $O(M)$ space. ■

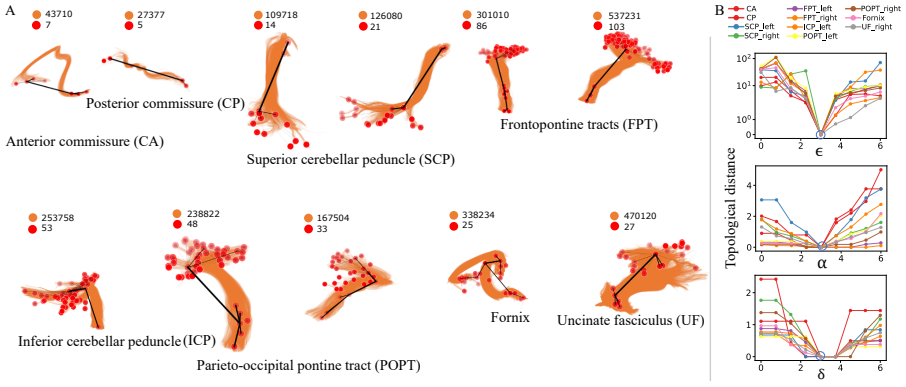

Figure 4.4: **Reeb graphs identify anatomically correct structures for ISMRM dataset.** A) Streamlines from the ISMRM dataset are represented in orange color and their Reeb graphs are overlayed on top. The nodes of the Reeb graphs are in red and the edges are black color. We observe that only a few points shown by nodes of the Reeb graphs are required to represent the branching behavior of the trajectories. B) We compute the base graph R_{ref} at $\epsilon_0 = 3, \alpha_0 = 3,$ and $\delta_0 = 3$ represented by blue circles in the plots. We compute the R by changing ϵ between 0 to 6 mm. Topological distance is computed between R_{ref} and R with varying ϵ. The same step is repeated We observe that the topological distance is more sensitive to the change in ϵ and least sensitive to the change in δ.The distances computed by varying ϵ are plotted on log scale for better visualization .

To demonstrate that the Reeb graph model is practical and captures the bundling behavior of the trajectories, we use two types of data sets to demonstrate the use of ReeBundle as a visualization tool and a fingerprint for longitudinal study: ISMRM [53] and CRASH dataset [54, 55]. ReeBundle detects and handles the morphology of the white matter tract configurations due to branching and local ambiguities in complicated bundle tracts. ISMRM helps in the qualitative validation of the ReeBundle approach. CRASH helps in the quantitative validation of the Reeb graph as a fingerprint for the bundle tract that can be used for longitudinal clinical diagnosis.

4.3 Reeb Graph as a Visualization Tool: ISMRM Dataset

We illustrate the behavior of our method by providing qualitative verification of the bundles that our algorithm re-bundles in the presence of noise and false positives in

Figure. 2.6. We account for the noisy streamlines by making one fundamental assumption that fibers are naturally organized as bundles in the brain. Therefore, curvature overshoot and short length streamlines are eliminated using the persistence parameter α while the isolated streamlines are excluded using the bundle size parameter δ as shown in Figure. 2.6A. Several complex fiber configurations are observed in tractograms [56] where streamlines cross, diverge, and turn resulting in crossing, fanning, and kissing characteristics as illustrated in Figure. 2.6. Such confounding patterns can be included or excluded in our model using the latent parameters. α is set to the length of the small encounters to handle the crossing and kissing patterns while δ is set to the size of the sparse bundle at the end to account for the fanning configuration. ISMRM [3] dataset is publicly available and consists of major bundle tracts that are representative of the challenges in human brain imaging *in vivo*. Tracts span different shapes, lengths, and sizes that help us in validating the output of our proposed method and analyze the effect of granularity parameters (ϵ, α, and δ). Figure. 4.4 shows the Reeb graph representation of the well-known neurological structures from the ISMRM dataset. The Reeb graph model detects and handles the morphology of the white matter tract configurations due to branching and local ambiguities such as crossing, kissing, and fanning. We visually inspect the maximal groups identified by our algorithm and verify the resulting model by using domain expertise. As an illustration, in all examples of fiber tracts shown in Figure. 4.4, no interruption is expected at the middle segments while a branching structure is expected at the end. Our method captures these branching structures. Furthermore, our model discovers the known termination regions and assigns them nodes (and hence, locations in brain space) in the Reeb graph. The run time on Intel CPU @ 2.20 GHz for CA ($N = 43710$) is $3.03s$, for CP ($N = 27377$) is $3.89s$, for SCP_R ($N = 109718$) is $22.7s$, and for SCP_L ($N = 126080$) is $13.9s$.

We present a sensitivity analysis of the topological distance of Reeb graphs to estimate

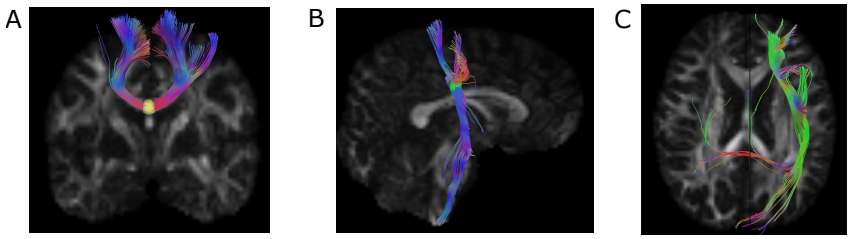

Figure 4.5: **Regions of interest and fiber tracts.** Isotropic sphere of radius 6 mm is placed for three bundle tracts as shown in (A) yellow sphere represents ROI for Corpus callosum (CC_body), B) green sphere represents ROI for Corticospinal tract (CST) and C) red sphere represents ROI for Inferior fronto-occipital fasciculus (IFOF).

the robustness under parametric uncertainties. This analysis helps in giving an idea of how an increase or decrease in the parameter's value affects the topological distance. We define the sensitivity of the distance with respect to a parameter and a base reference graph $\mathcal{R}_{\text{ref}}$ as $|\Delta\text{TD}|/|\Delta\epsilon|$, $|\Delta\text{TD}|/|\Delta\alpha|$, and $|\Delta\text{TD}|/|\Delta\delta|$. We observe that the sensitivity of the topological distance with δ is low whereas it is high for ϵ. The results are shown in Figure. 4.4B. Note that the sensitivity of the parameters depends on the dataset as well.

4.4 Topological Fingerprinting: CRASH Dataset

The CRASH longitudinal dataset comprises of 200 Diffusion Spectrum Imaging (DSI) scans. Twenty-five participants between the ages of 20 and 35 years old (mean = 22.63; SD=3.16; 52% male) were recruited from the greater Santa Barbara area as part of the Cognitive Resilience and Sleep History (CRASH) research study. DSI scans were acquired at bi-weekly experimental sessions over the course of 16 weeks (8 scans per subject). For a given 19 minute scan, a DSI scheme was used, and a total of 257 diffusion samples were acquired and a total of 20 regularly interspersed b0 volumes were included. The maximum b-value was 4985 s/mm^2. The in-plane resolution and slice thickness were 2mm. Data were preprocessed using ANTs [57] that included head motion correction with

b-vector rotation [54, 55]. Fiber tracking for original non-rotated data was qualitatively compared to the rotated data.

The HCP1065 tractography atlas [58] consisting of 64 volumetric regions was used to localize fiber tracts. These 64 NIFTI volumes record the population probability of each white matter tract aggregated from the tractography of 1065 subjects. We used spherical regions placed within three of these volumes: Corpus callosum (CC_body), Corticospinal tract (CST), and Inferior fronto-occipital fasciculus (IFOF) as shown in Figure. 4.5. Fiber tracking through these spheres was performed in DSI Studio [59]. The diffusion data were reconstructed using generalized q-sampling imaging [60] with a diffusion sampling length ratio of 1.25. A deterministic fiber tracking algorithm [61] was used. The ROI was a deformed sphere (sphere ROI in MNI space is warped using nonlinear registration) for each tract in the subject space. A seeding region was placed on the whole brain. The anisotropy threshold was randomly selected. The angular threshold was 35 degrees. The step size was 1 mm. Tracts with lengths shorter than 20 or longer than 300 mm were discarded. A total of 100000 seeds were placed.

All the ReeBundle parameters are kept the same throughout our experiments. To show that the parameters are good for all the DSI scans, we hypothesize that the Reeb graphs for the same subject should be similar even with different sessions. So, we fix the parameters for session 1 and solve an optimization problem using Markov Chain Monte Carlo (MCMC) sampling to compute the parameters' values for other sessions. In Figure. 4.6, we illustrate Corner plots [62] showing all the 2D projections of the posterior probability distribution of the granularity parameters $\epsilon, \alpha,$ and δ. We show that the initial parameters represented by blue lines are good enough to capture the topological characteristics of the bundle tracts. That is beneficial because we can potentially have a single ϵ, α, and δ that work across all the tracts in a given longitudinal study. Our comparison pipeline is sensitive to bundle extraction and can be used as a quality assurance

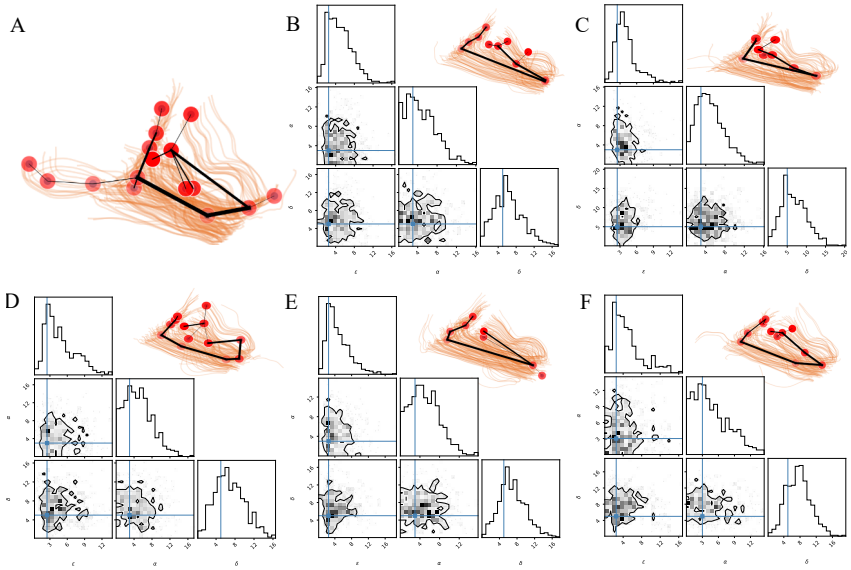

Figure 4.6: **Bayesian optimization validates the choice of parameters for the longitudinal study.** A) shows the Reeb graph of session 1 for $\epsilon_0 = 2.5, \alpha_0 = 3$, and $\delta_0 = 5$. B-F show the marginalized distribution for each parameter solving the optimization problem to minimize the topological distance between the Reeb graph in (A) and the Reeb graphs computed at sampled parameter's value. Blue lines represent ϵ_0, α_0, and δ_0 and the corresponding Reeb graphs are also illustrated respectively at the top right in each plot.

measure to tweak the pre-processing steps or to decide the fiber tracking methods.

We perform the topological analysis using our proposed ReeBundle method for the three bundle tracts: CC, CST, and IFOF of all the 200 DSI scans. An example Reeb graph and effects of ReeBundle parameters for the CC bundle can be visualized in .mp4 files [1]. The difference between two bundle tracts is equivalent to computing the topological distance between two Reeb graphs as explained in Sec. 4.2. The first step is to compute the Reeb graphs ($\mathcal{R}$) for all the scans. Figure. 4.7 A-C show the computed Reeb graphs for CC, CST, and IFOF respectively. Once we have the Reeb graphs, we compute the distances between all pairs of graphs. We also compute the distances of the Reeb graph with itself, which is 0 since the graphs are identical. We refer to the distances

[1] `https://ieeexplore.ieee.org/ielx7/42/10336247/10223239/supp2-3306049.mp4?arnumber=10223239`

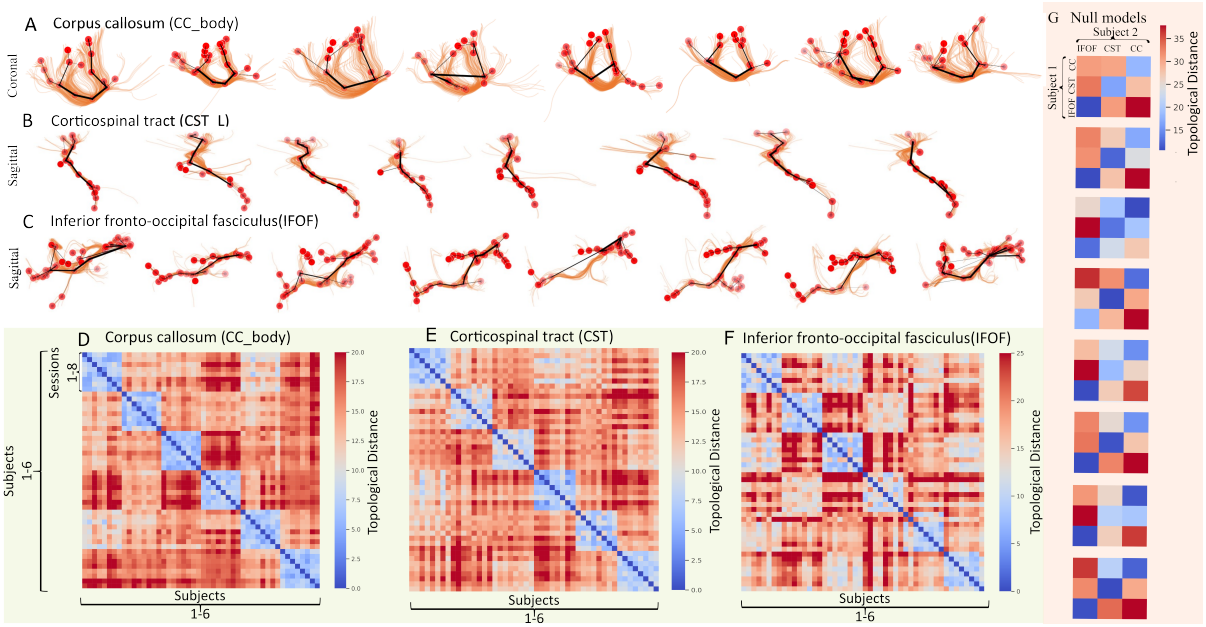

Figure 4.7: **Reeb graph as a fingerprint for repeated scans in longitudinal study.** Reeb graphs for A) CC B) CST C) IFOF. Heatmaps for the topological distance between the Reeb graphs for 6 subjects with respect to D) CC E) CST F) IFOF. The blue color in the heatmaps shows similar topological characteristics (relatively less value of TD) for the bundle tracts while the orange color shows higher values for TD. Heatmaps show low relative topological distances within sessions of the same subject. D-F shows the streamlines and corresponding Reeb graphs for $\epsilon = 2.5, \alpha = 3$, and $\delta = 5$ models. G) Heatmaps show large relative topological distances for CC vs. CST, CC vs. IFOF, and CST vs. IFOF.

computed between different subjects as *inter-distance* and the distance computed within the sessions of the same subject as *intra-distance*. The differences between the *inter-* and *intra-* distances are statistically significant with p ~ 0 calculated using the T-test.

Figure. 4.7 D-F shows the heatmaps representing the topological distances between the graphs. Consecutive groups of 8 rows in the heatmap represent 8 sessions of the same subject. The blue color of the heatmaps shows that the bundles are similar (less distance) while the orange color shows a larger difference between the bundles. The dark blue diagonal means that the distance is zero when the Reeb graph is compared to itself. This implies that topological organization for all the bundles is highly consistent across the sessions of the subjects in the follow-up scans. Reeb graph also plays the role of

Table 4.1: Comparison of different graph distance methods [52]

Methods	*Intra-distance*	*Inter-distance*	TD ($\mathcal{R}_{CST}, \mathcal{R}_{IFOF}$)
DegreeDivergence	0.12	0.13	0.12
IpsenMikhailov	0.37	0.41	0.37
LaplacianSpectral	0.54	0.55	0.54
PortraitDivergence	0.56	0.58	0.67
dkSeries	0.57	0.60	0.62
OnionDivergence	0.78	0.80	0.78
JaccardDistance	0.99	0.99	1.00
NetSimile	9.02	9.69	8.20
NetLSD	47.29	55.31	47.93
Our method	9.23	15.62	31.86

"fingerprint" as seen from the darker orange colors in the heatmaps. All the analyses were done in the subject space avoiding the impact of registration in the MNI space. Topological distance allows us to compare bundle tracts. The relative distance of bundle tracts is the least when compared within subjects while it increases when compared to different subjects as shown in Figure. 4.7 for 6 subjects (the code to generate heatmaps of all 25 subjects and the heatmaps as .png files can be found in the GitHub [50]). This difference further increases when two different bundle tracts are compared as shown in Figure. 4.7G. This is expected and validates our distance computation. Our comparison method does not require registration to the common MNI space but can be used to compare registered bundle tracts as well.

Our method produces a graph object at the end. So, any graph distance metric can be adapted for Reeb graph comparison. Table 4.1 compares the *intra-distance* and *inter-distance* measures for ten different methods of graph comparison, including our proposed method. All methods are evaluated based on their ability to distinguish between 25 subjects using CST bundle tract. The values in the table represent the distance measures between the scans of the same subjects with itself (*intra-distance*), between graphs of different subjects (*inter-distance*), and the distance between two different graphs from

two different bundle tracts, namely IFOF and CST using $\mathrm{TD}(\mathcal{R}_{CST}, \mathcal{R}_{IFOF})$. For all methods, $\mathrm{TD}(\mathcal{R}, \mathcal{R}) = 0$, that is, the distance between graphs that are identical is equal to zero. The ideal distance metric will minimize the *intra-distance* within each subject's scans, maximize the *inter-distance* among different subjects, and will exhibit a high value for tracts that are different. The results show that our method performs well compared to existing methods, particularly in distinguishing between different subjects as demonstrated by the relatively high *inter-distance* values and fingerprinting the subjects by showing relatively lower *intra-distance*s. We obtain similar results when using any other tract for comparison.

4.4.1 Comparison With Tractometry Methods

To the best of our knowledge, there are no existing methods in the literature for modeling the branching behavior of the brain fibers that can be used to compare our method directly. In Sec. 4.1, we discuss the major differences between the connectivity matrix representation and Reeb graph models. Analyzing the topological organization of white matter fibers using ReeBundle provides a more specific investigation than considering a region of interest or tract-averaged metric.

We compare our method with shape analysis introduced in [43] to compute length, area, volume, and shape metrics from tractography. Specifically, we used number of tracts, mean length, span, curl, elongation, diameter, volume, trunk volume, branch volume, total surface area, total radius of end regions, total area of end regions, irregularity, area of end regions, radius of end regions, irregularity of end regions, quantitative anisotropy (qa), normalized quantitative anisotropy (nqa), generalized fractional anisotropy (gfa), and isotropic measures (iso). These metrics are calculated using the DSI studio and saved as .csv file [50]. The *inter-* and *intra-* distances were computed for

each subject by simply finding the difference between the values. The statistical significance of the difference between *inter-* and *intra-* distances based on these tractometry metrics is negligible ($p \sim .99$) whereas ReeBundle provides a graph representation that visualizes, quantifies, and localizes the significant topological differences.

We also compare our method with BUAN [45] and observe a similar trend of *intra-* versus *inter-* distances. However, since the BUAN metric is normalized, the similarity scores for *inter-* subjects were approximately 1 which was the same even when different bundle tracts were compared while our proposed method shows relative distances. BUAN compares the same types of bundles while our distance computation can be used to visualize and compare different sets of bundles.

4.5 Tracking Longitudinal Changes: Alzheimer's Disease

Reeb graphs detect disease-related and age-dependent topology alterations. We apply the step-by-step pipeline (Figure. 4.2) on two example datasets: a) ADNI dataset "for the Alzheimer's Disease Neuroimaging Initiative"[2], the Reeb graphs were computed

[2]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [63]. As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at [64]., and b) OpenNeuro brain tumor dataset [65]. While the pipeline has the potential to be used in group-wise brain comparison studies, it is particularly well-suited for tracking changes in an individual subject's brain. This is because the variability caused by changes within an individual's brain remains consistent across scans. As a result, for a given subject, we can obtain coherent Reeb graphs with nodes and edges that can be tracked consistently. Therefore, the applications of ReeBundle include tracking disease progression and evaluating the effects of therapeutic and surgical interventions.

We applied the ReeBundle pipeline to the ADNI dataset to track topological changes with increasing age. We selected two subjects with ages ranging from 75–85 years. Both subjects have the same imaging protocols. After pre-processing, a deterministic fiber tracking algorithm was used to compute fiber tracts passing through a small ROI in the corpus callosum area decided from the HCP tractography probabilistic map (step 1 and step 2). For same choice of parameters as for the CRASH data analysis, $\epsilon = 2.5$, $\alpha = 3$, and $\delta = 5$

for all scans of the subjects (step 3). Finally, the topological distance was computed between each pair of graphs (step 4). In Figure. 4.8A, we show that the topological distances significantly increases with time for Alzheimer's patients. However, in case of normal subjects, the increase in the topological distance is less pronounced even when the scans were taken at greater age differences. This demonstrates how Reeb graphs can be computed at different time stamps to track and quantify changes related to Alzheimer's disease progression. Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early Alzheimer's disease.

4.6 Quantifying the Post-operative Structural Changes: Brain Tumor

Similarly, we use the pipeline to study pre- and post-operative patients with Meningioma I in the OpenNeuro brain tumor dataset. We computed the Inferior Fronto Occipital Fasciculus bundle tracts for a patient (PAT02). We chose this bundle tract since it is one of the tracts that pass through the segmented tumor mask. The post-operative Reeb graph for the patient has a marked increase in the edge weight when compared to the pre-operative Reeb graph edge weight. The edge weights are shown in bold in Figure. 4.8B and the difference in weights is captured by the thickness of graph edges. Note that even though the fiber count can be captured by the connectivity matrix, the

branching ends (see highlighted nodes in the figure) recovered post-surgery cannot be visualized or quantified in any other methods to the best of our knowledge.

Reeb graphs help in modeling the dynamics of how tumors affect the spatial orientation of the white matter fibers and how surgical interventions improve these distortions. In Figure. 4.8B, we observe that the fiber tracts were recovered post-surgery, which were distorted in the pre-operative patient. We obtained similar results for patient (PAT01) with the fiber tracts computed for the Cingulum Frontal Parahippocampal R region. For these ROIs, we observed similar results with scans for another Meningioma I patient as well. No such distinctions were observed for normal control subjects who did not undergo the surgery but had similar living conditions as the patients who underwent the tumor surgery. The preliminary results for brain tumor demonstrates that the change in the Reeb graph provides insights into the structural changes in white matter fiber tracts post-surgery. The change in the number of fibers and branching patterns in the Reeb graph after surgery indicates the successful resection of the tumor.

4.7 Discussion

In this Chapter, we discuss unique way to model white matter fibers as a Reeb graph network that is sparse and preserves underlying geometrical information. ReeBundle is robust to noise and is a computationally efficient framework to characterize white matter fibers from a topological point of view. We demonstrate that the proposed approach captures the fiber branching behavior in the presence of noise and is further validated on synthetic and real datasets. ReeBundle parameters ϵ, α, and δ control the granularity of the bundling desired – small values allow only very dense sets of streamlines for grouping, while larger values of ϵ relax the groups and allow longer, larger, and sparser groups to form. We note that the tunability of the parameters can be used to apply to a wider set

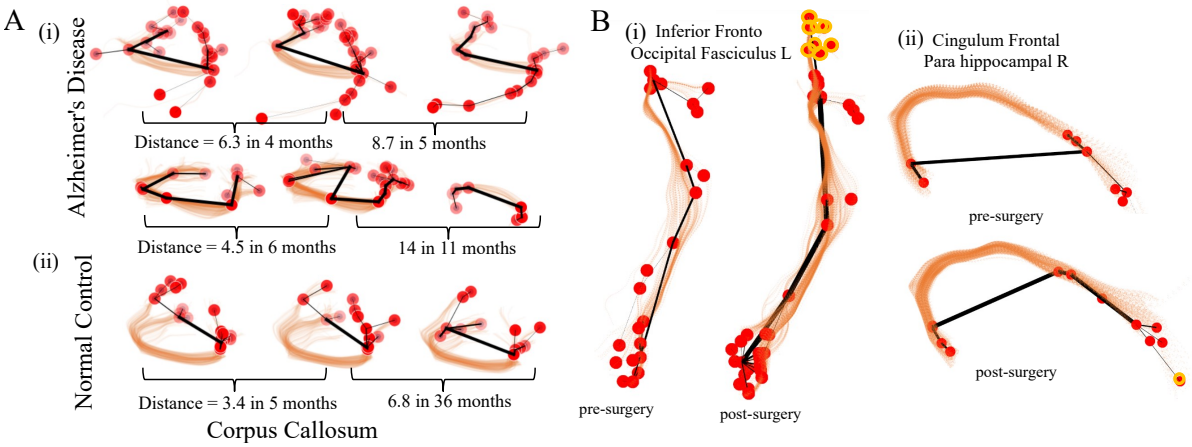

Figure 4.8: **ReeBundle Application to Neurological Disorders:** (A) Tracking Alzheimer's Disease Progression using the ADNI dataset. Reeb graphs overlaid on the streamlines for subjects with (i) Alzheimer's Disease and (ii) a normal subject at different ages. The topological distance significantly increases with time for Alzheimer's patients. In contrast, for normal subjects, the increase in topological distance is less pronounced even when the scans were taken at greater age differences. (B) Effects of brain tumor resection on the (i) Inferior Fronto-Occipital Fasciculus and (ii) Cingulum Frontal-Parahippocampal. Distortion in branching is recovered in post-operation shown by highlighted nodes towards the ends and the edge weights of the Reeb graph capture changes in the number of fibers.

of problems, from comparing tractography algorithms to studying neurological disorders. For neuroscience applications, we recommend using the default parameters as an initial set to begin the study. For general analysis, we recommend setting the ϵ parameter to around 3 mm, which is close to the minimum dMRI resolution of 2 mm. Sensitivity analysis indicates the algorithm's greater sensitivity to ϵ, so, fixing the other parameters and varying ϵ should be sufficient for most applications.

Our method does not require brain parcellation or registration to model the streamlines. Reeb graph and distance metric are sensitive to variations in brain structure, such as those observed in the developing brain and in the presence of tumors. This sensitivity is beneficial in the context of longitudinal studies of brain development and the topological evolution of the brain. We have highlighted the potential utility of this metric in tracking Alzheimer's disease progression using the ADNI dataset, and evaluating the ef-

fects of surgical interventions for brain tumor using the OpenNeuro brain tumor dataset. As a continuation of this research, we plan to investigate the impact of NPH condition on the CST bundle and its relationship to motor function of patients in the next Chapter.

Chapter 5

AI and Reeb Graphs in Neurosurgery

In the previous chapters, we explored the application of our Reeb graph methods for visualization, fingerprinting, and comparison of human brains using diffusion MRI. Reeb graphs and the topological distance metric are highly sensitive to variations in brain structure, such as those observed in the developing brain and in the presence of tumors. This sensitivity is particularly advantageous for longitudinal studies of brain development and the topological evolution of neuronal connections. Previously, in Chapter 4 we have demonstrated the potential utility of this metric in tracking Alzheimer's disease progression using the ADNI dataset and in evaluating the effects of surgical interventions for brain tumors using the OpenNeuro brain tumor dataset. In this chapter we explore application to Normal Pressure Hydrocephalous (NPH), a neurological condition that affects elderly population.

We propose using Reeb graphs to represent critical white matter bundles and analyze the topological differences between pre- and post-surgery scans to measure the impact of shunt surgery for NPH. This approach, combined with AI-based segmentation of ventricular regions, allows us to model and quantify structural distortions caused by enlarged

The content from this Chapter is accepted for publication in Neurosurgery [66].

ventricles and correlate these changes with observed symptoms. Our methodology utilizes a population-averaged model from the Human Connectome Project [1] and ReeBundle pipeline to provide a robust framework for clinical diagnostics and treatment evaluation.

Diffusion MRI has proven effective in predicting surgical outcomes in patients with iNPH by identifying neural distortions and examining tractography [67, 68]. This technique is valuable for assessing the potential reversibility of white matter injuries, which may correlate with NPH symptoms. Prior research has shown that dMRI can differentiate between NPH, Alzheimer's Disease, and healthy controls based on various measures, providing insights into microstructural changes in white matter that explain cognitive and physical impairments in NPH patients.

However, such diffusion MRI studies are rarely done in a clinical setting. To address this, we explore using the AI-segmented ventricular regions as regions to avoid (ROA) when tracking fibers affected by iNPH condition based on an average connectome from the HCP. Reeb graph models from pre- and post-surgery scans of the patient give localized and quantified information about structure changes in pre- vs post- as well as normal vs NPH.

In the following we first describe a quantitative method that distinguishes pre- and post-surgery ventricular volumes, establishing for the first time a direct correlation between ventricular volume reduction to observed improvement. We then provide our Reeb graph based analysis.

[1] `https://www.humanconnectome.org/`

5.1 Case Study: Idiopathic Normal Pressure Hydrocephalus

Idiopathic Normal Pressure Hydrocephalus (iNPH) is a neurodegenerative condition characterized by the accumulation of cerebrospinal fluid (CSF) in the brain, which results in ventriculomegaly. An estimate of more than 700,000 Americans have iNPH, but less than 20% receive an appropriate diagnosis [69]. The typical symptoms of iNPH are gait disturbance, urinary incontinence, and progressive dementia [70].

Ventriculo-peritoneal shunt surgery is an effective treatment for iNPH that diverts CSF from the ventricles to the peritoneal cavity [70]. The surgically treatable nature of iNPH highlights the recent advancements in diagnostics and treatments that have improved patient outcomes [71]. Follow-up studies of patients have shown positive results in gait followed by bladder and cognitive improvements [72].

5.2 AI-based Method for iNPH Shunt Surgery

The clinical improvement after shunt surgery in patients with iNPH is shown to be associated with a slight reduction of ventricular size. However, this change in the ventricular size is too small to visually detect from scans. Manual or semi-automatic methods are commonplace to measure this change [73–76]. Consequently, the computed CSF volume in ventricles cannot be used as a reliable radiographic marker in an automated manner. We propose a fully automated method with a user-friendly web interface for computing the ventricular volume metric, enabling the diagnosis and tracking of iNPH after shunt surgery.

Artificial intelligence (AI) has been used for automatic ventricle segmentation for brain CT scans and MRIs for various neuro-disorders studies. Prior research has devel-

oped AI-based methods [77, 78] to segment cerebral ventricles from T2-weighted MRIs and calculate ventricular volume in pediatric patients with obstructive hydrocephalus. However, analysis of change in ventricular volume after the shunt surgery remains unexplored. CT scans of the head are routinely performed on iNPH patients with VP shunts because they are faster than MRI and free from artifacts. Automated analysis of brain CT scans in relation to the subject's patency and functioning of the VP shunt is crucial for effective surgical planning and prognosis. In this section, we examine whether quantifiable changes can be detected in ventricular volume in iNPH patients that undergo ventriculo-peritoneal shunt procedures.

- We propose a novel AI-based ventricular metric (CVV) derived from ventricular volume in brain CT scans and validate the efficacy of this metric by tracking ventricular volume changes after shunt placement (Figure. 5.1). CVV shows significant correlation between changes in iNPH symptomatology and our metric before and after the shunt surgery.

- We provide a user-friendly web interface that automates the computation of CVV for convenient analysis of post-shunt surgery brain. We also provide the pre- and post-shunt brain scans publicly through our web-interface for further study.

5.2.1 Ethics approval and consent to participate

This retrospective study was conducted with all images de-identified by the University of California, Irvine medical center as specified by the IRB agreement. IRB waived the requirement of patient's consent as the data was de-identified and none of the information can be mapped back to the patient.

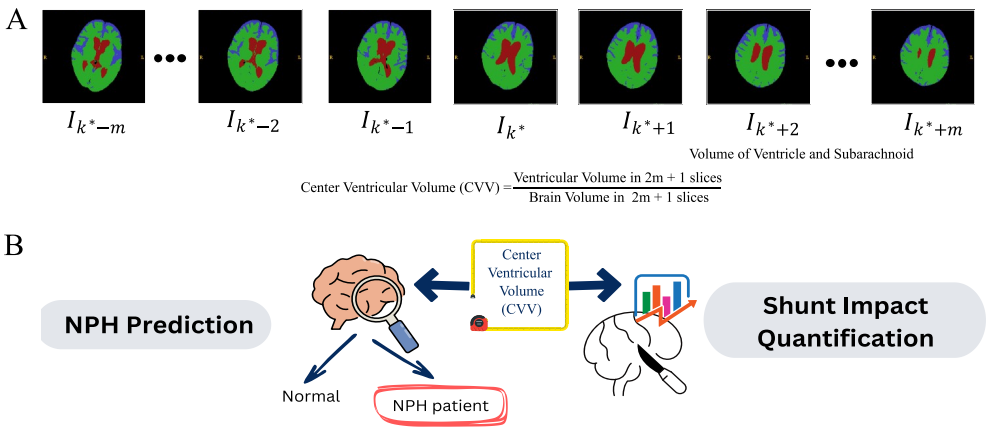

Figure 5.1: **Computation of the Center Ventricular Volume (CVV) and its usage.** (A) CVV is computed as the ratio of ventricular volume in $2m+1$ slices to the brain volume in those $2m+1$ slices using the segmented scans. m slices are chosen before and after slice k^* with the maximum ventricular volume. (B) CVV is used to distinguish between normal and iNPH subjects and to quantify the effect of shunt treatment by analyzing the segmented scans before and after surgery.

5.2.2 Retrospective review of study populations

The CT scans of adult patients who had a clinical and radiological diagnosis of iNPH from January 2015 to December 2022 were selected from the University of California, Irvine medical center NPH clinic. 47 scans from 15 patients (7 males and 8 females) that had undergone a ventriculo-peritoneal (VP) shunt ranging in age from 70s to 90s (mean = 78.5; SD = 5.2) were utilized to compare the pre-operative and post-operative ventricular volumes. At least one pre-operative CT scan of the brain was taken. Multiple post-surgery scans were taken in addition to the immediate post-operative scan. The additional scans were taken as part of routine care usually after a change in the shunt programming or when the patient had any new neurological complaints or concerns. This ensured that there was no development of a subdural hematoma or subdural hygroma.

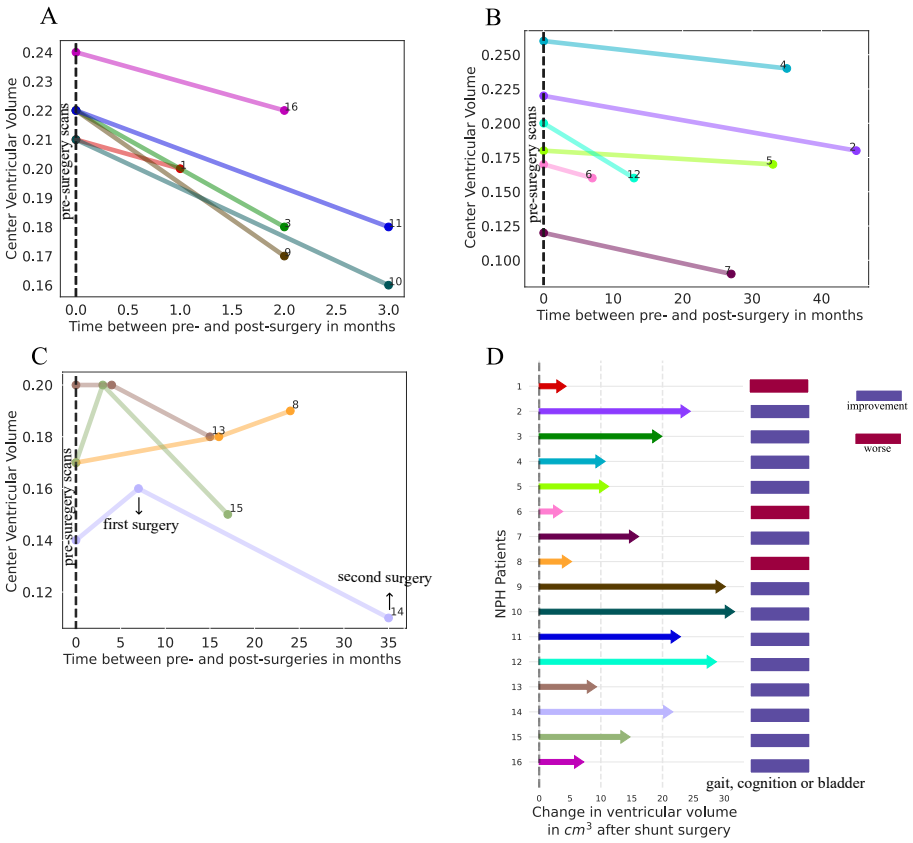

Figure 5.2: **Decrease in ventricular volume after shunt surgery.**(A) Pre- and post-surgery ventricular volume within three months of the shunt surgery. (B) Pre- and post-surgery ventricular volume after three months of the shunt surgery for a distinct set of subjects. (C) Ventricular volume reduction in iNPH patients not responding to the first shunt surgery. After additional time, shunt was adjusted to augment CSF flow, therefore subsequent CT scans show a decrease in the ventricular volume. (D) Absolute change in CSF for central ventricular volume in cc after shunt surgery for all the subjects, showing the correlation between improvement in patients and changes in ventricular volume.

5.2.3 Ventriculoperitoneal Shunt Surgery

Ventriculoperitoneal shunts were placed via a burr hole in the right frontal region at Kocher's point. The shunt tubing was tunneled posteriorly with a post-auricular

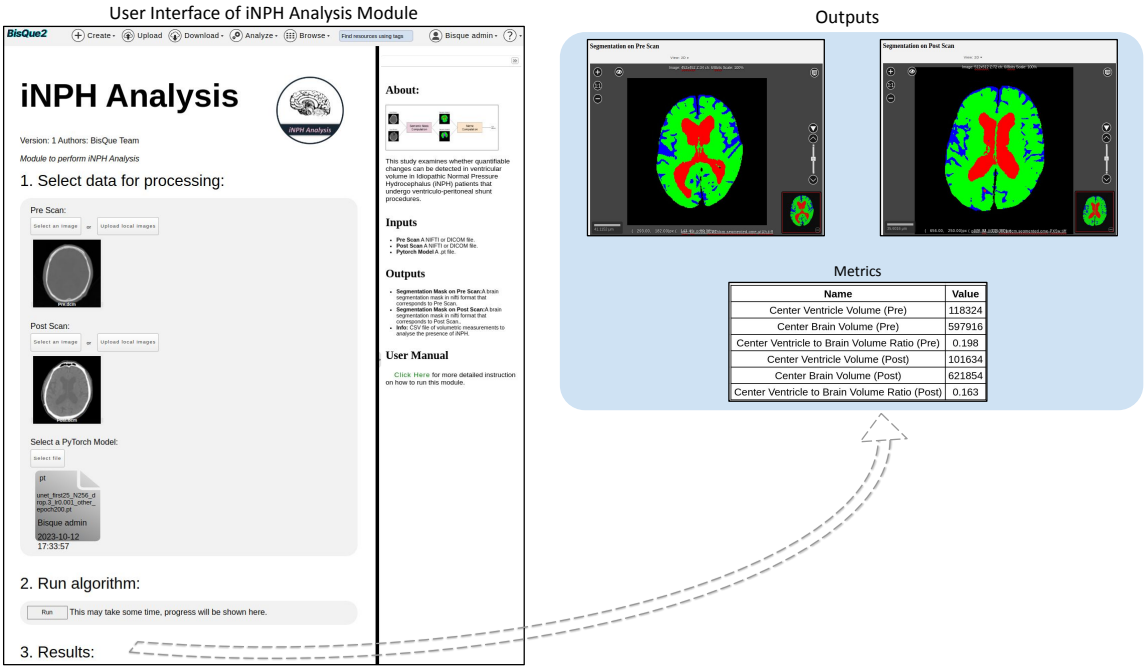

Figure 5.3: **The user interface of the iNPH Analysis module hosted on BisQue.** The module processes both input scans to produce brain image segmentations and then calculates the CVV metric using the segmented results.

intervening incision to travel across the clavicle and the abdomen where it was placed intraperitoneally using a general surgery laparoscopic technique. The Catheter was placed into the frontal horn using neuronavigation (Medtronics-Axiem EM system, Minneapolis, Minnesota). The shunt valve was placed in a subcutaneous pocket posterior to the ventricular entry site. The shunt used was the Codman-Certas programmable valve with Bactoseal proximal and distal catheters (Integra life sciences, Princeton, New Jersey). A post-operative CT scan of the head was done within 6 hours of the surgery to confirm the placement of the catheter and to ensure that there were no immediate intracranial complications. Patients were mobilized and the majority (90%) were discharged on the first post-operative day.

All patients had a pre-operative CT scan which was used for navigation purposes to help with the placement of the ventricular catheter. They subsequently had a post-

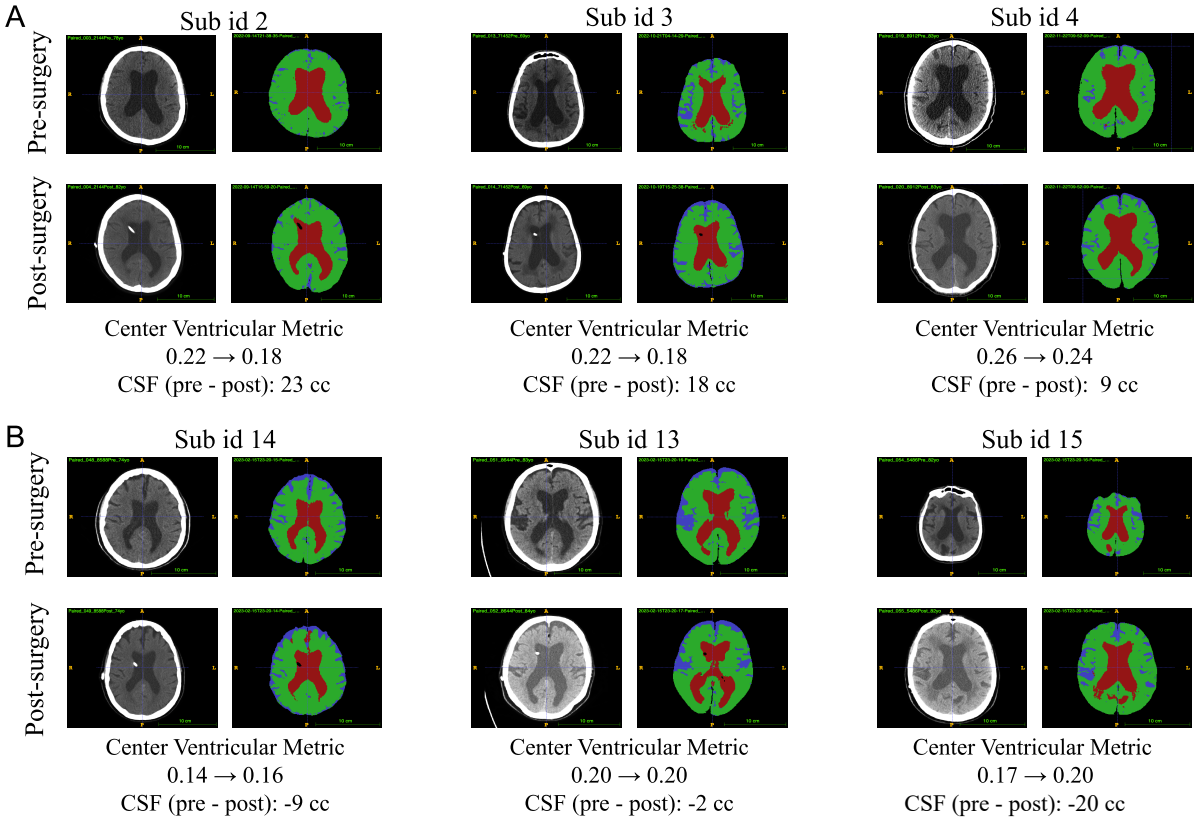

Figure 5.4: **Ventricular Volume Analysis for iNPH Patients.**(A) Decrease in CVV and ventricular volume shows positive response to shunt surgery. (B) Increase in CVV and ventricular volume indicating worsening iNPH conditions. The iNPH shunt valve was adjusted for these patients. There is decrease in the CVV and absolute ventricular volume after second shunt surgery in non-responders to the first surgery. This demonstrates the potential utility of the ventricular volume metric in determining the need for shunt adjustment. It can be challenging to accurately evaluate the condition based solely on CT scan visualization.

operative CT scan of the head within 24 hours of surgery as >95% of patients were discharged to home on post-operative day #1. This was to ensure there was no track related hemorrhage or subdural hematoma and to document the position of the catheter. Patients had a follow-up CT scan of the head 6 weeks after surgery just prior to the first shunt adjustment and one 3 months after surgery when they had had their shunt adjustments and were felt to be optimized clinically. Additional CT scans were done if the patients had a fall or injury, and we were concerned about the development of

subdural hematoma. They were also done if the patient or family were concerned about a deterioration in neurological status. These additional CT scans were done at the discretion of the attending neurosurgeon (Prof. Chen from UCI Medical Center).

Patients' responses to shunting were obtained from retrospective chart reviews. Improvement in any of the clinical indices of iNPH (gait, cognition, urinary function) was given a score of +1. Worsening of symptoms was scored as -1. No change was scored as 0. If the patient had a normal gait, cognition, or urinary function before or after the shunt surgery, it was still scored 0. Because of referral patterns and insurance restraints, CT scans were performed on different machines from other radiology centers. As CT scans are reported in different orientations (coronal, sagittal, axial), the images selected for our study are those in the axial-oblique orientation.

5.2.4 Automated pipeline for tracking the impact of shunt surgery

We propose a new metric, the Center Ventricular Volume (CVV) to effectively quantify the ventricular volume, which is a key feature affected in iNPH. CVV quantifies and tracks the effect of shunt treatment by analyzing the segmented scans before and after surgery in an automated way. We first segment the axial-oblique scan into ventricles, subarachnoid, and gray/white matter. We use a UNet-based artificial neural network developed by the authors at University of California, Santa Barbara [11, 79] to segment the scan. This neural network model is trained on a different dataset than used in this study. This ensures that our proposed method is broadly applicable. We can get the ventricular volume from each segmented slice of the scan by counting the voxels labeled as ventricles. To calculate the CVV, we find the 2D slice of the segmented brain volume that shows the largest volume of the ventricles based on the maximum number of voxels labeled as ventricles. Then, keeping this slice as center slice, we take the slices around

this to account for 35 mm thickness axially as shown in Figure 5.1.

We define the CVV as the normalized ventricular volume within these representative slices relative to the overall brain volume. Thus, an increase in the CVV corresponds to an increase in ventricular volume. The total volume achieved in the center slices used in this study was chosen to encompass the areas that were most representative (80%) of the ventricular volume of the brain and 10% of the entire brain volume. Specifically, the posterior fossa and 4th ventricle are not included. This was intentional, as these areas are hard to capture consistently on the axial CT scans and there are some variations. There are also more CT artifacts in the posterior fossa.

5.2.5 Results

The dataset consists of a total of 47 CT scans from pre- and post-surgery of 15 patients. We compute the CVV for all the 47 CT scans and compute the difference between the corresponding pre- and post-un-normalized ventricular volume. From the multiple pre- and post-scans available for a given subject, the last pre-scan and the first available post-scan were used to quantify the impact of the shunt surgery. A clear trend towards a decrease in ventricular volume and CVV in post-surgery scans is observed that significantly correlates with the qualitative improvement in the patient's iNPH condition as shown in Table 5.1.

All subjects with iNPH showed a positive response to shunt surgery and thus smaller CVV metric post-surgery. Figure 5.2A shows the subjects with scans collected within 3 months of surgery. In Figure 5.2B, we show the subjects with scans collected after 3 months with time differences between scans ranging up to 40 months. This shows the longer-term impact of shunt treatment and the quantitative validation of this improvement by the CVV metric. It is commonly perceived that the response to shunt surgery

Table 5.1: Computation of Evan's Index, CVV metric, and difference in ventricular volume in cc for subjects with pre- and post-surgery CT scans.

Anonymized Sub_id	Diagnosis	Shunt surgery	EI: Evan's Index	Ventricular Volume (CVV)	Change in Center Ventricle Volume (cc) (Pre – post)
1	worsening of cog fxn (could be attributed to covid or fall)	Pre	0.352	0.21	
	improved gait function, improved cognition has reverted to pre-op levels	Post	0.349	0.2	3
2	Freezing of gait at start, urinary urgency, mild cog impairment	Pre	0.35	0.22	
	continued gait problems + bradykinesia	Post	0.314	0.18	23
3	gait issues progressed to ataxic, slow, unbalanced; some memory/cog issues	Pre	0.345	0.22	
	gait has much improved	Post	0.32	0.18	18
4	progressively worsening gait	Pre	0.421	0.26	
	memory+thinking improved; gait also improved	Post	0.398	0.24	9
5	slow gait, sway side to side; some memory deficit	Pre	0.384	0.18	
	feeling of gait improvement, more fluent speech	Post	0.376	0.17	10
6	wheelchair bound at this time, incontinent	Pre	0.43	0.17	
	some/slow improvement	Post	0.428	0.16	2
7	<restricted access to medical record>	Pre	0.319	0.12	
	Improved bladder, overall improvement	Post	0.298	0.09	15
9	worsening gait; urinary urgency/incontinence	Pre	0.334	0.22	
	no more incontinence	Post	0.31	0.17	29
10	magnetic gait	Pre	0.338	0.21	
	able to walk independently, still magnetic/ataxic gait; improved gait fxn	Post	0.326	0.16	30
11	wide-based, slow, ataxic gait	Pre	0.33	0.22	
	improved gait, cognitive and urinary fxn are the same	Post	0.32	0.18	21
12	cognitive/memory impairment; urinary incontinence	Pre	0.33	0.2	
	made gain with gait and continence; cognitive has not increased or decreased	Post	0.307	0.16	27
13	progressive gait disturbance since 2015, severe balance problems; increased urinary frequency; memory is being impacted	Pre	0.33	0.196	
	still ataxic and wide-based, but can walk without walker; slight improvement with continence	Post	0.316	0.199	-2
	still ataxic and wide-based, but is fluid; shunt adjusted from 4 to 3	Post (after shunt adjustment)	0.308	0.18	10
14	some memory impairment; some urinary incontinence	Pre	0.306	0.14	
	fall; still has incontinence; shunt level 4	Post	0.298	0.16	-9
	slow, but fluid gait, not magnetic, another fall; worsening memory; shunt adjusted from 4.5 to 5	Post (after shunt adjustment)	0.3	0.11	29
15	gait difficulty; urinary incontinence; mild memory loss	Pre	0.306	0.17	
	no improvement in incontinence; can stand for 2 mins but not walk yet; codman certa shunt level 5 to 4	Post	0.303	0.2	-20
	can take small steps but gait is still ataxic and slow; memory is stable; uses adult briefs; current shunt setting is two	Post (after shunt adjustment)	0.3	0.15	33
16	slow, wide-based gait + ataxic; incontinence,	Pre	0.43	0.24	
	some improvement in gait, continence greatly improved	Post	0.42	0.22	6

for iNPH is of limited duration. But, to the contrary, we show here that even after 3 months, the ventricular volume is lower than pre-surgery scans.

A few subjects (sub_id = 13, 14, 15 in Table 1) do not positively respond to the first shunt surgery (as shown in Figure 5.2C). However, after additional time and shunt adjustments to augment CSF flow, subsequent CT scans show a decrease in the ventricular volume. This highlights the importance of the continued clinical evaluation and shunt adjustments to optimize the efficacy of the shunt. The ventricular volume computed after the first surgery shows no response or deterioration after shunt surgery. This is accurately captured and quantified with our method in the CVV computation. So, the metric effectively helps neurosurgeons in key decision-making regarding shunt valve adjustment. This enables more objective comparison of scans from a radiological point of view.

In Figure 5.2D, we observe the quantity of CSF drained after shunt surgery, represented by the size of the arrows for all subjects. Smaller arrows correspond to a smaller quantity of CSF drained. Among these, the two smallest arrows ($<$ 4cc CSF drained) are associated with worsening conditions in one of the symptoms of gait, cognition, or bladder control, as indicated by the red block on the right. This observation establishes a correlation between the quantified volume of CSF drained and qualitative diagnoses.

5.3 Diagnosis of iNPH

We also use CVV to see if it is reliable for differentiating between normal individuals and those with iNPH (24 diagnosed with iNPH and 21 normal elderly individuals). Our proposed metric achieves high accuracy (0.95), precision (0.96), and recall (0.96) in distinguishing between normal and iNPH subjects. The details and results for this experiment are included in the following section and Table 5.2.

Table 5.2: Comparison of methods to distinguish between Normal vs NPH

	Methods	Threshold (computed from ROC)	Accuracy	Precision	Recall
Manual	Evan's Index	0.3	0.86	0.79	1
Fully automated (UNet)	Center Ventricular Volume Metric	0.1	0.95	0.96	0.96

5.3.1 Dataset used for training UNet

Another set of 45 brain CT scans from 45 subjects (24 iNPH patients and 21 normal elderly controls) were acquired as part of the treatment process. This dataset is independent of that used in the pre- and post-surgery study. This same set of datasets is used to demonstrate the diagnostic value of the method in distinguishing iNPH from normal control subjects. The subjects were aged between 60s and 90s (mean=75; SD=7.7). The CT scans of adult patients who had a clinical and radiological diagnosis of iNPH from January 2015 to December 2022 are selected from the Neurosurgery NPH clinic. This retrospective study is conducted with all images de-identified by the Center for Artificial Intelligence in Diagnostic Medicine (CAIDM) team at UCI as specified by the IRB agreement.

5.3.2 CVV metric distinguishes iNPH patients from normal subjects

We compute the CVV metric for the above-mentioned dataset that consists of 45 CT scans. The metric is calculated for automatically segmented scans using the UNet neural network trained on this dataset. Additionally, Evan's Index is manually computed for all the scans. To classify the scans into iNPH and Normal, we compute the threshold of each metric and method using a Receiver Operating Characteristic (ROC) curve. The ROC

curve allows us to evaluate the performance of each metric and method by plotting the true positive rate against the false positive rate at various threshold values. By examining the curve, we determine the optimal threshold that balances sensitivity and specificity for each metric and method. The threshold for EI is 0.3 which is also the standard threshold to determine the iNPH condition. The threshold for CVV is 0.1. Table 5.2 displays the computed threshold and provides accuracy, precision, and recall metrics for EI and CVV. Our results indicate that the CVV performs better than EI for the classification task as shown in the Table 5.2.

The whole pipeline as shown in Figure 5.1 has been containerized and deployed as a module on our web platform BisQue13. To access and utilize the module, users are simply required to register an account on the BisQue platform using this link `https://bisque2.ece.ucsb.edu/`. Once registered, they can initiate and execute the entire pipeline directly from their web browser, analyzing both pre and post CT scans. The BisQue module is designed to accept scans in either DICOM or NIFTI file formats. Upon submission, the module processes these scans to generate brain image segmentations, subsequently calculating the CVV metric based on the segmented scans, as shown in Figure 5.3. Detailed instructions on accessing the module and executing them on a sample pair of scans can be found in the BisQue Online Documentation14 and video 1. We provide free access to all de-identified pre- and post-surgery CT scans in this study through the BisQue platform.

5.4 Reeb Graphs for Multimodal Analysis of iNPH

As noted previously, diffusion MRI can be used to predict surgical outcomes in patients with idiopathic Normal Pressure Hydrocephalus (NPH). It has proven effective in identifying neural distortions in NPH patients and is valuable for examining tractography

and the potential reversibility of white matter injuries, which may correlate with NPH symptoms [68]. In this prior research, dMRI was utilized to gain insights from scans of individuals with NPH, Alzheimer's Disease (AD), and healthy controls. Analysis shows that dMRI can differentiate between these groups based on measures like axial diffusion, axial kurtosis, and axonal water fraction, which vary significantly among them [67]. The subtle microstructural changes that are observed in white matter might explain the cognitive and physical impairments seen in NPH patients, supporting the use of dMRI to assess the likelihood of improvement with shunt treatment. Diffusion MRI also distinguishes between patients with NPH and those with AD by identifying changes in the microstructures of the corticospinal tract (CST) [80].

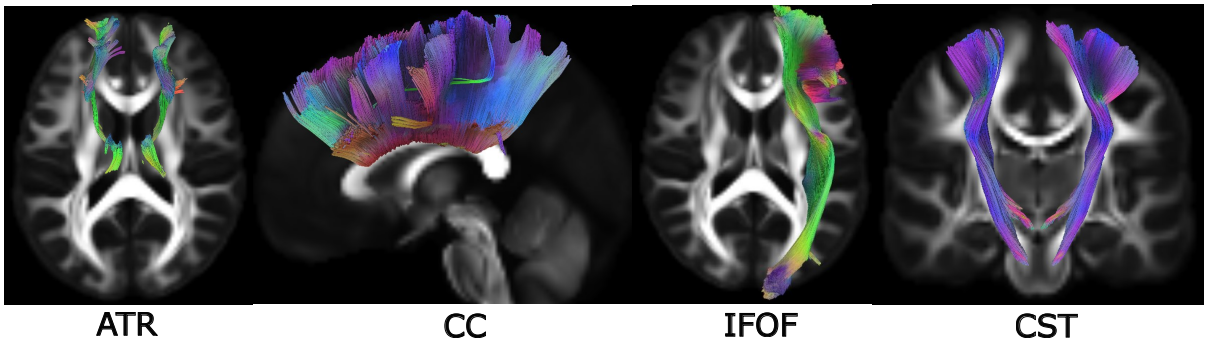

Figure 5.5: **White matter regions of interest (ROIs) for iNPH.** White matter tracts represented in the context of normal ventricular size: anterior thalamic radiation (ATR), corpus callosum (CC), inferior fronto-occipital fasciculus (IFOF), corticospinal tract (CST). These are the fibers generated from the group-average MRI data for the HCP S1200 young adult subjects using DSI studio.

Researchers have investigated the layout and potential reversibility of white matter injuries in NPH before and shortly after shunting procedures [68]. They redefined regions of interest (ROIs) into an 'at-risk' model for white matter injuries in NPH, encompassing six key tracts: the genu and body of the corpus callosum (GCC and BCC), the inferior longitudinal fasciculus (ILF), the anterior thalamic radiation (ATR), the conjunction of

the inferior fronto-occipital and uncinate fasciculi (IFO/UNC), and the posterior limb of the internal capsule (PLIC). These are the white matter tracts of interest near ventricular region of the brain as illustrated in Figure 5.5. We hypothesize that Reeb graphs can be used to represent these critical white matter bundles, with nodes indicating major areas of distortion due to ventricular enlargement. The topological differences between the Reeb graphs for pre- and post-surgery can be analyzed to measure the impact of shunt surgery, along with CSF volume obtained from CT scans. Changes in the location of Reeb graph nodes before and after the surgery can help clinicians in understanding and correlating the improved symptoms post-operation with the associated brain regions.

5.4.1 Comparative analysis using our proposed Reeb graph based pipeline

For Reeb graph based analysis, we utilize the HCP-1065 Young Adult Fiber Template. This template is a population-averaged model derived from the diffusion MRI data of 1065 subjects (575 female, 490 male), aged 22 to 37 years, from the Human Connectome Project's 1200-subject release in 2017. The mean age is 28.74 years with a first quartile of 26, a median of 29, and a fourth quartile of 32. The imaging used a multishell diffusion scheme with b-values of 1000, 2000, and 3000 s/mm^2, each with 90 diffusion sampling directions. Both in-plane resolution and slice thickness are 1.25 mm. The data were processed in MNI space using q-space diffeomorphic reconstruction (QSDR) to map the spin distribution function, with a diffusion sampling length ratio of 1.7 and an output resolution of 1 mm. We perform the fiber tracking using DSI Studio, utilizing the ready-to-track data stored in FIB files (`https://brain.labsolver.org/hcp_template.html`).

In this section, we focus our analysis of Cortico-spinal tract (CST_L) bundle tract since one of the most common symptom of NPH is gait impairment, which is known to

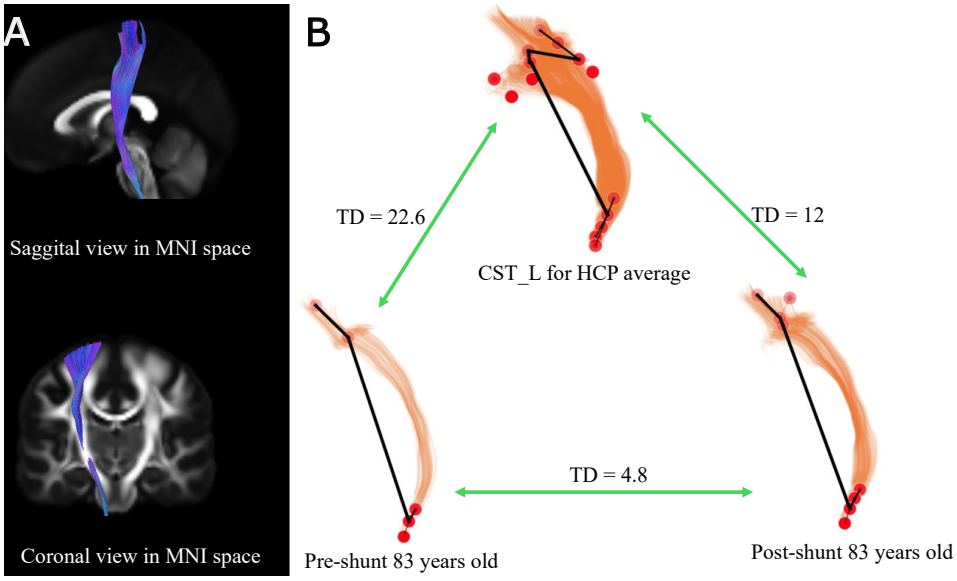

Figure 5.6: **Reeb graph modeling of CST_L for pre- vs post- shunt surgery in iNPH.** A) CST_L bundle tract visualized in DSI studio. B) Reeb graphs overlayed on CST_L tract for HCP average, pre-shunt 83 years old, and post-shunt 83 years old. ROA is extracted from the segmented ventricles from subject's CT scans. Topological distances (TD) are calculated among all the three tracts. This shows how the distance is larger when HCP average is compared to pre-shunt surgery while it is smaller when TD is computed against post-shunt surgery tract. This helps us in quantifying the positive response to the shunt surgery by the given subject.

correlate with this bundle tract. The tracked bundle tract is visualized in DSI studio as shown in Figure. 5.6. Here, we analyze a 83 years old subject pre- and post shunt surgery. The main symptom of NPH is the enlarged ventricle volume which affects the bundle tracts surrounding it. So, we first segment the ventricle using our trained UNet neural network [78]. We follow this step for both pre and post scan. Now, we apply the MNI transofrmation to the raw CT scan and save the resulting affine transformation matrix. Then, we apply that affine transform to segmented ventricle volume to get the ventricular region in the MNI space. Now this ventricular region in the MNI space acts as a region of avoidance (ROA) when we track the CST fibers. This simulates the fibers that are impacted by the increased ventricular volume. Finally, we compute the Reeb graph and the topological distances among these three sets of CST bundles: HCP normal

case, normal ventricular volume as ROA, and NPH ventricular volume as ROA. Changes in the nodes and edges of the Reeb graphs localizes the effect of enlarged ventricle pre and post surgery. This can be correlated to the symptoms seen in the pre- and post-surgery.

Topological distances (TD) between the three tracts are presented in Figure 5.6. The analysis reveals that the TD is greater when compared with the young adult brain scan from the Human Connectome Project (HCP) average prior to shunt surgery, and smaller when compared to the tract following post-shunt surgery. This variation in TD quantitatively demonstrates the subject's positive response to the shunt surgery.

One of the limitation of this method is that the enlarged ventricle should create a distortion effect. But due to the use of a region of avoidance the affected fibers have disappeared in our analysis. Therefore, we call this a simulated analysis, but with the ventricle segmentation calibrated using real CT scans.

In Section 5.2.5, we demonstrated the utility of the CVV metric for both evaluating the effectiveness of shunt surgery and distinguishing between normal and NPH conditions. Building on this, we conduct a comparative analysis using three sets of CST tracts: the HCP average, a normal 81-year-old subject, and an 81-year-old subject with NPH. This analysis aims to highlight how Reeb graphs can differentiate normal aging from NPH pathology. An age-matched normal subject serves as an effective control. ROA is derived from ventricles segmented from the subjects' CT scans. Topological distances (TD) are calculated among the three tracts, indicating that TD is larger when compared to the NPH subject and smaller against the normal subject. The Reeb graph's finer edges (edge density shows number of white matter fibers bundled together) in the NPH subject also visually represent the impact of fiber loss due to ventricular enlargement, as depicted in Figure 5.7.

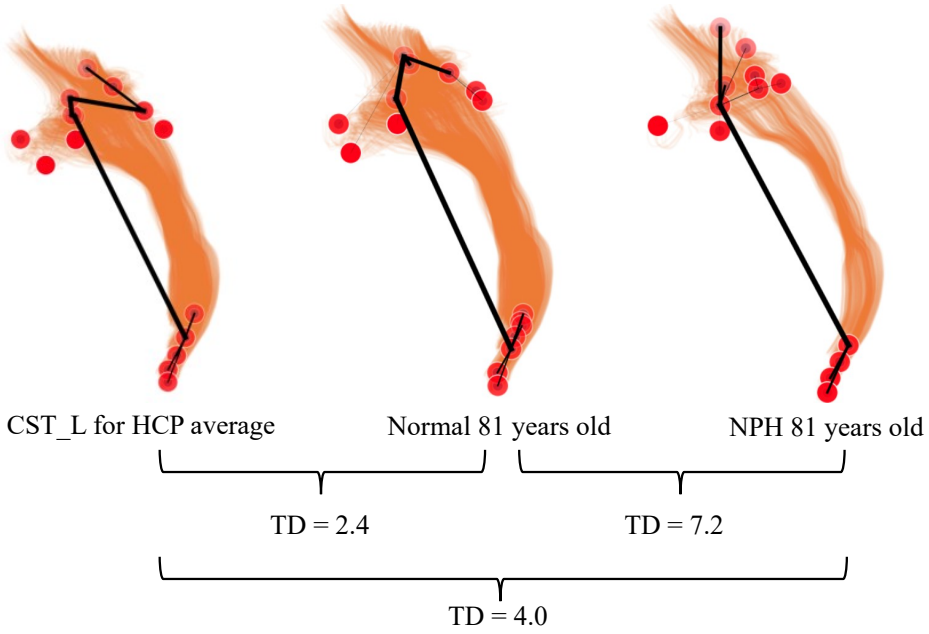

Figure 5.7: **Reeb graph modeling of CST_L to distinguish Normal from iNPH.** Reeb graphs overlayed on CST_L tract for HCP average, 81 years old normal subject, and 81 years old NPH subject. ROA is extracted from the segmented ventricles from subject's CT scans. Topological distances (TD) are calculated among all the three tracts. This shows how the distance is larger when HCP average is compared to NPH while it is smaller when TD is computed against Normal subject. The thinner edges of the Reeb graph for NPH subject also helps us visualize the effect of the lost fibers due to enlarged ventricle.

5.5 Discussion

Our proposed AI-based method in this Chapter has the potential to serve as a tool to assess shunt function which is challenging from CT scans as shown in Figures 5.4A and 5.4B. The ability to demonstrate a decrease in ventricular volume is significant to verify that there is patency and functioning of the VP shunt. The clinical significance is that with iNPH, the lack of improvement or the deterioration in symptoms may not necessarily be from shunt dysfunction. Demonstration that the decrease in ventricular volume is maintained directs the clinician to look for alternative explanations. Periodic quantitative evaluation of the ventricular volume can provide assurance of continued shunt functioning.

MRI is an important diagnostic modality in iNPH patients, but it is not common for routine longitudinal analysis. CT scans of the head are routinely performed on iNPH patients with VP shunts because they are quickly done and provide anatomical information that neurosurgeons may use in their treatment paradigms. MRIs may offer superior resolution and diagnostic capabilities, but post-surgery MRIs were are not available for the subjects in this study at multiple time points. Since our primary focus was on the differential analysis of shunt impact, we use CT scans instead of MRIs. The advantage of a CT scan over an MRI is its rapid acquisition. Additionally, CT scans are free from the shunt magnetic artifact that occurs in MRIs.

In this study, the CT scans that were used were derived from different CT scanners (GE, Siemens, Toshiba) at different institutions. This is the nature of the care of these patients. Often patients have the CT scans done at their home institute because of proximity or insurance restraints. If the patient or family perceive a possible neurological problem, they will go to the closest Emergency Department which is not necessarily our institute. The number of slices of scans in our analysis varied ranging from 30-57 and slice thickness resolution ranged from 3-5 mm. We were able to analyze these scans from various scanners and obtain meaningful quantitative evaluations. The analysis is done using the axial-oblique orientation to line-up the scans, which is similar across institutions and reduces the effect of tilting. The proposed method utilizes the resolution of the image to do all the computations in metric units rather than using a number of voxels. Thus, our automated analysis pipeline is agnostic to the scanner or CT scan resolution.

The lack of visible decrease in the ventricular size after VP shunting has traditionally been attributed to a decreased brain compliance in iNPH. As iNPH is more common in patients over the age of 60, there are a myriad of age-related factors that affect brain compliance including arterial and venous calcifications and aqueductal narrowing or stenosis. This is the first automated method to show that the ventricular volume changes with

shunting and correlates with improvement. This suggests that the CSF egress via the shunt may affect the overall brain compliance. Since the CVV metric is a ratio of ventricular volume to brain, the decrease in CVV with shunting suggests that there is both a decrease in ventricular volume and/or increase in brain volume.

To conclude, we prersented two approaches: 1) an AI-based method for quantifying ventricular volume, and 2) Reeb graph-based multimodal analysis using segmented CT scans and diffusion MRI. Reeb graph-based method help in quantifying the structural changes associated with iNPH condition. These multimodal analyses could enhance our understanding by interpreting diagnostic symptoms related to gait, cognition, and bladder function concerning structural network changes in iNPH patients. By utilizing diffusion MRI and AI-based segmentation of ventricular regions, we model and quantify structural distortions caused by enlarged ventricles. Our analysis highlights how the topological distances (TD) between Reeb graphs of pre- and post-shunt surgery tracts can serve as an indicator of surgical efficacy, with a reduction in TD correlating with symptomatic improvements. Furthermore, comparing Reeb graphs between normal aging subjects and those with iNPH reveals distinct patterns of fiber loss and structural changes, providing valuable insights into the pathology of iNPH. This method underlines the importance of integrating advanced computational techniques in clinical diagnostics and treatment evaluation, offering a robust framework for future research and clinical applications in neuroimaging and neurosurgery.

Chapter 6

Spatio-temporal Reeb Graphs

The Reeb graph method proposed in the previous chapters can be called spatial Reeb graph. The Reeb graph was defined and proposed on spatial trajectories. The problem was motivated from the spatial connections and wiring of the brain. However, even the spatial trajectories have ordered set of points (which can be considered as temporal) but they do not have the concept of global time points as can be in trajectories from time based acquisition such as sampling features with respect to time. Reeb graph model for connectome proposed in this thesis can be adapted for general-purpose trajectory analysis. Connectomes refer to the wiring diagram of the trajectories. Any object's movement, for example, human movement creates a large scale GPS data which forms a wiring diagram of their movement pattern just like Connectomes that we have been discussing in this thesis so far.

In this Chapter, we generalize the Reeb graph-based method to model patterns of a set of trajectories over time. Given a set of recurring patterns represented by data collected over time for a given object, we propose a general algorithm to model such data in a linear time complexity over time. Any deviation from usual normal pattern

The content from this Chapter is submitted for publication [81].

is encoded as nodes in the Reeb graph. We demonstrate the usage of Reeb graphs and how it captures the critically significant spatial and temporal deviations using the nodes of the Reeb graph. For example, in our case studies, we discuss how Reeb graphs can be utilized to model pattern-of-life in human trajectories (akin to a fingerprint). Human behavior typically follows a pattern of normalcy in day-to-day activities. Case studies discussed in this Chapter includes realistic human movement scenarios: visiting uncommon locations, taking odd routes at infrequent times, uncommon time visits, and uncommon stay durations. We analyze the Reeb graph to interpret the topological structure of the spatial trajectories over time. Potential applications of the algorithms introduced in this Chapter include urban planning, security surveillance, and behavioral research.

6.1 Definitions and Problem Formulation

A trajectory T is defined as a dictionary (key: value) containing an ordered sequence of time points and their associated spatial coordinates:

$$T = \{t_0 : p_0, t_1 : p_1, t_2 : p_2, \ldots, t_m : p_m\}, \tag{6.1}$$

where m is chosen according to the desired resolution to sample the pattern of the object. Here m denotes the total number of points in a given trajectory T. The frequency of data sampling decides m. For example, to model the weekdays of an object's activities, the raw data is sampled every second, giving us $m = 86400$ which is the total number of seconds in a day. Similarly, if the data is sampled every hour, then $m = 24$ points per day. We define n as the total number of trajectories for a given object. For example, to model month-long data, $n = 30$ and for weekdays, $n = 5$. The common setting used throughout this Chapter for our problem definition is $m = 24$ and $n = 5$. Each time point

t_i corresponds to a spatial coordinate p_i representing the position of the object at time t_i. $p_i = (\text{lat}_i, \text{lon}_i)$, where lat_i represents the latitude and long_i represents the longitude. The Euclidean distance between two spatial coordinates p_i and $p_{i'}$ is calculated at time t_i as follows:

$$d(p_i, p_i') = \sqrt{(\text{lat}_i - \text{lat}_i')^2 + (\text{lon}_i - \text{lon}_i')^2}, \tag{6.2}$$

where lat_i and lon_i are the latitude and longitude of the first point, and lat_i' and lon_i' are those of the second point. $d(p_i, p_i')$ gives the 2-norm distance between two points on the Euclidean plane. This approximates the geographic distance of the points. The algorithm is defined with respect to a distance threshold ϵ within which the points are considered sufficiently close together i.e. within a small geographical area. This is the inter-trajectory distance that guides the granularity of the Reeb graphs according to the problem definition.

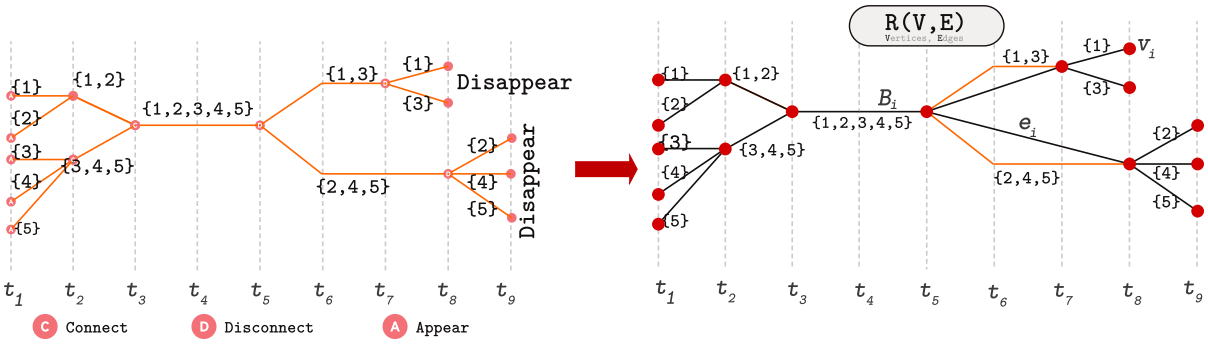

Figure 6.1: Reeb Graph Construction Over Time. We show the construction of Reeb graphs $R(V, E)$ for a set of five trajectories. The *appear, disappear, connect,* and *disconnect* events are shown on the left-hand side. Changes in the grouping of trajectories due to these events are encoded as nodes on the right-hand side. Nodes of the Reeb graph $\mathcal{R}$ on the right-hand side are shown in red color and the edges are shown in black color throughout the thesis.

Human behavior typically follows a pattern of normalcy in day-to-day activities. This is marked by recurring activities within specific time periods. In order to discover the

large-scale spatio-temporal patterns, we represent the bundling structure of trajectories as a *Reeb graph* $R(V, E)$. Nodes of the Reeb graph will pinpoint critical spatial points of the object's pattern. Intuitively, if a continuous portion of a behavior of the object happens at the same time and within the same spatial distance (ϵ) every day then they present a pattern of normalcy. We formalize this by introducing the concept of "bundles" to characterize normal behavior through consistent daily subtrajectory events. Each trajectory begins with an *appear* event at the first index and concludes with a *disappear* event at the last index of T. Deviations from this norm by more than ϵ are classified as *disconnect* events, while a return to the norm is labeled a *connect* event. Formally, for a given ϵ and $m = 23$ i.e. sampled every hour, let's take two trajectories T and T':

- At time t_0: p_0 and p'_0 are the *appear* events.
- At time t_{23}: p_{23} and p'_{23} are the *disappear* events.
- If $d(p_0, p'_0) \leq \epsilon, (p_1, p'_1) \leq \epsilon, \ldots, d(p_k, p'_k) \leq \epsilon$, but $d(p_{k+1}, p'_{k+1}) > \epsilon$, then t_{k+1} represents a *disconnect* event between T and T'.

Algorithm 3 Find *connect* and *disconnect* events

Input: Trajectories T and T', threshold ϵ
Output: Dictionary of *connect*/*disconnect* events, $events_{T,T'}$
Initialize $events_{T,T'}$ as an empty dictionary
Initialize $k \leftarrow 0$
Initialize $connect\_flag \leftarrow$ False
while $k < m$ **do**
    if $d(T[t_k], T'[t_k]) < \epsilon$ **then**
        $events_{T,T'}[t_k] \leftarrow connect$
        $connect\_flag \leftarrow$ True
        while $k < m$ and $d(T[t_k], T'[t_k]) < \epsilon$ **do**
            $k \leftarrow k + 1$
        if $k < m$ **then**
            $events_{T,T'}[t_k] \leftarrow disconnect$
    $k \leftarrow k + 1$
return $events_{T,T'}$

Algorithm 4 Construction of Reeb Graph

```
function ConstructReebGraph(set of events for all pairs of trajectories (E))
    for all steps k from 0 to |E| do                          ▷ Dynamic Graphs
        if appear event of T then
            insert new node T to G_k
        if disappear event of T then
            delete node T from G_k
        if connect event between T_x and T_y then
            insert edge (T_x, T_y) to G_k
        if disconnect event of trajectories T_x and T_y then
            delete edge (T_x, T_y) from G_k
        P ← empty bundle partition                            ▷ Bundle Partition
        Query G_{k-1} and G_k to get the connected components C_{k-1} and C_k respectively;
        for all connected component c_k ∈ C_k do
            if c_k ∈ C_{k-1} then
                assign the same bundle id B_i to the points for trajectories in c_k;
            else
                create a new bundle id B_{i+1} and assign it to the points for trajectories
                in c_k;
            Add B_{i+1} to P
    Construct Reeb graph R from P by connecting adjacent bundles with nodes and
    bundles as edges;                                         ▷ Construct Reeb graph
    return R
```

- If $d(p_0, p'_0) > \epsilon, (p_1, p'_1) > \epsilon, \ldots, d(p_k, p'_k) > \epsilon$, but $d(p_{k+1}, p'_{k+1}) \leq \epsilon$, then t_{k+1} represents a *connect* event between T and T'.

6.2 Parameter Selection

In cases where raw trajectories form extended groups, the nodes of the Reeb graph might be significantly spaced apart. Particularly along extended curves, the nodes and edges of the Reeb graph may fail to accurately reflect the true topological structure of these trajectories. This issue can be addressed by imposing a time constraint on temporal trajectories and a spatial length constraint on spatial trajectories. By doing so, we ensure that each bundle in the Reeb graph does not exceed a length of τ, as illustrated

in Figure 6.2.

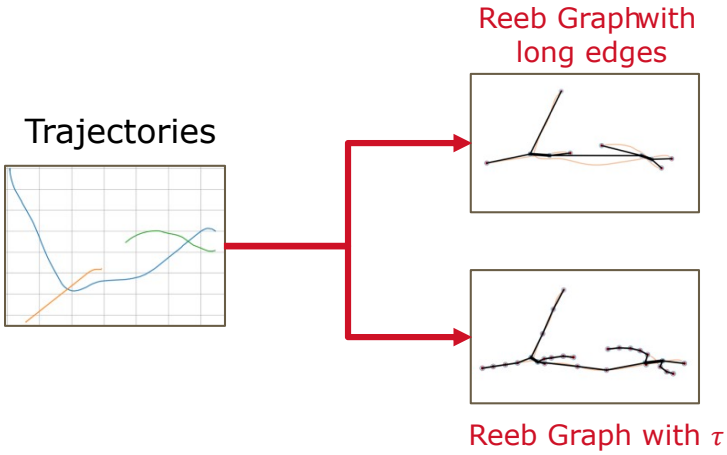

Figure 6.2: Reeb graphs with τ parameter to preserve the topological structure.

6.3 Time Complexity Analysis

Reeb graph construction (illustrated in Figure. 6.1) can be divided into the following major steps: event computation, construction of dynamic graphs (Gs), connectivity query in the dynamic graph for bundle partition (P), and construction of the Reeb graphs (R) from bundles partition as shown in Figure. 6.1. The first step of Reeb graph construction involves computing the *connect* and *disconnect* events. Algorithm 3 outlines the steps of computing events. The event computation takes $\mathcal{O}(m)$ time, where m represents the number of time points in the trajectories T and T'. At each time point, the algorithm looks for $\mathcal{O}(5 \times 5)$ possibilities of potential events. The second step of the Reeb graph involves handling the events to construct dynamic graph Gs. The nodes of G represent the daily trajectories and the edges of the G represent the ϵ-connectivity between them. The total number of nodes in G is 5 representing one trajectory for each day of the object. All the nodes of G represents the trajectories and those corresponding points of the trajectories form the level set at that given time. The connected component of the G will

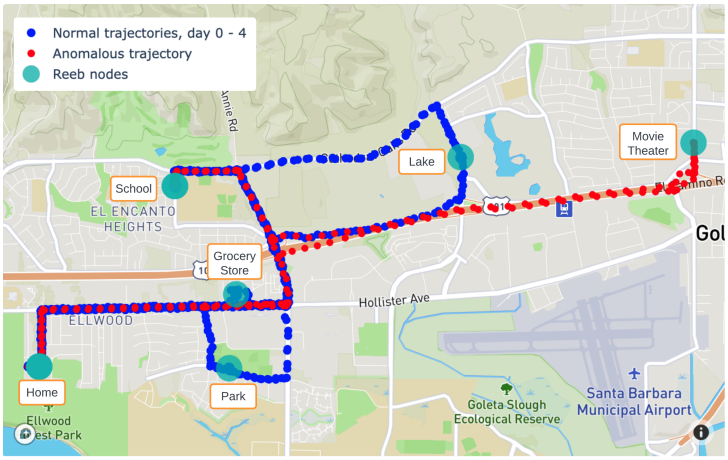

Figure 6.3: Map overlay of normal and anomalous trajectories from scenario 2 of the case study, annotated with semantic labels for points of interest (POIs).

give us the $\epsilon-$step bundle partition of subtrajectories denoted by $P = \{B1, B2, \ldots, Bk\}$ such that every segment in T_0, T_1, T_2, T_3, T_4 is uniquely assigned to exactly one bundle. The final step is to construct the Reeb graph from these bundles. Reeb graph R can be constructed from P by connecting adjacent bundles with nodes and bundles as edges similar to the described construction in [8]. So, the time complexity of the Reeb graph construction step would be $\mathcal{O}(m)$ because in the worst case, all the time points will have events. At each time, the connectivity query to the dynamic graph with 5 nodes takes constant time. The more detailed steps can be found in the Algorithm 4.

6.4 Case Study

6.4.1 Data generation

We model the pattern of life of a single object over different trajectories. Each trajectory is simulated using the SUMO software package [82] and represents realistic behavior and movement patterns over the course of one week. We construct the Reeb graph for each trajectory and show how it sufficiently represents the trajectory's information with

significantly fewer nodes.

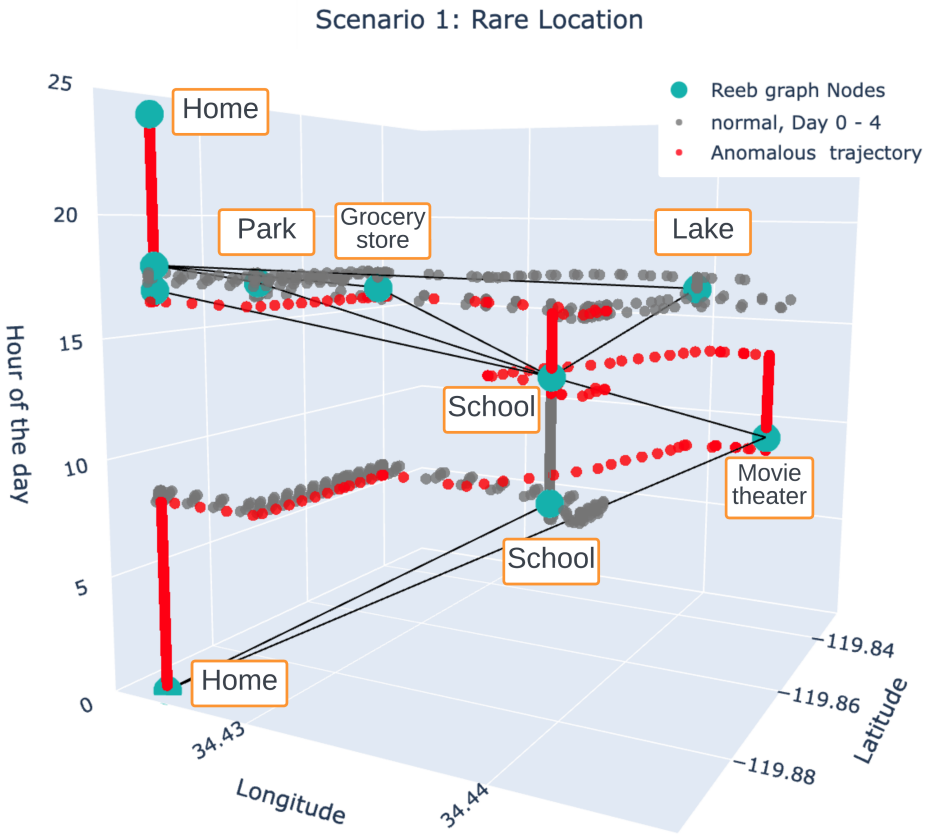

Figure 6.4: 3D trajectory plots with computed Reeb graph nodes for scenario 1 in Section 6.4, where day 0 to day 4 are normal trajectories, and the anomalous trajectory is in red.

In this case study, we analyze the behavioral patterns of a simulated high-school student from the city of Santa Barbara, California (Figure 6.3), using trajectory data that includes multiple points of interest (POIs), such as the student's home, school, park, grocery store, and lake. The student's daily routine typically consists of attending school from approximately 8:00 AM to 9:00 AM, concluding at around 4:00 PM to 5:00 PM, followed by visits to recreational sites before returning home. To thoroughly investigate both normal and anomalous behavioral patterns, we generated five days of normal trajectory data, complemented by additional days tailored to each specific scenario described

earlier. Each trajectory entry is recorded with timestamps, latitude, and longitude coordinates. Figure. 6.3 displays the student's trajectories across different POI locations for the rare location scenario, illustrating the distribution of both routine and deviant movements. Figure. 6.4 displays the same data as a 3D plot, providing a clear spatio-temporal visualization of the student's stay locations, duration, and revisit frequencies.

6.4.2 Definition of anomalous behavior

We define L as a set of normal POIs and their corresponding time points,

$$L = \{(lat_1, lon_1, t_1), (lat_2, lon_2, t_2), \ldots, (lat_n, lon_n, t_n)\}$$

where (lat_i, lon_i) represents the geographic coordinates with $lat_i \in [-90, 90]$ and $lon_i \in [-180, 180]$, and t_i is the time at which these coordinates were recorded. Relative to this definition, all the anomaly behaviors for a given object are defined as follows:

Scenario 1 (S1): Rare Location Anomaly Rare location anomaly refers to a scenario when an object visits a new location $(lat^*, lon^*, t_i) \notin L$. (lat^*, lon^*) is spatially different from their normal spatial geographical points of interest such as school or work. Reeb graph will encode this rare location by creating a new node localizing the abnormality.

Scenario 2 (S2): Rare Route Visit Anomaly In this scenario, the object visits the same POI locations multiple times but utilizes a uniquely different route on a single journey. This introduces *disconnect* event from their normal movement pattern, resulting in a new node in the Reeb graph. More formally, if $(lat^*, lon^*, t_{k:l}) \notin (lat, lon, t_{1:k-1})$ and $\notin (lat, lon, t_{l+1:m})$, then nodes v_k and v_l will be added to R.

Scenario 3 (S3): Uncommon Time Visit This is a case of time violation where the object visits a familiar location at an uncommon time t^* i.e, $(lat_i, lon_i, t^*) \neq (lat_i, lon_i, t_i)$

Algorithm 5 Trajectory Generation

Inputs:
  POIs – List of Points of Interest as coordinates on a map.
  $TimeList_n$ – Dictionary mapping each POI to normal visit times.
  $TimeList_a$ – Dictionary mapping each POI to abnormal visit times.
  Road Network – Road network graph for route generation.
Output:
  T – A list of normal trajectories of an object visiting specified POIs.
  T^* – A list of abnormal trajectories of an object visiting specified POIs.
Initialize Trajectories list
for each POI in the POIs list **do**
    Select $TimeList$ based on a decision rule (normal vs abnormal)
    for each time in $TimeList$ **do**
        Generate a starting point for the object
        Use duarouter to calculate the shortest path from the starting point to the POI at the given time
        Pass the list of edges to SUMO for movement simulation
        Collect the output trajectory from SUMO
        Append to T or T^* based on decision rule
return T, T^*

Scenario 4 (S4): Uncommon Stay Duration Anomaly In this scenario the object stays for an abnormal duration (Δ) at a specific location $(lat^*, lon^*, t_{i+\Delta})$. This results in a *disconnect* event for the object's trajectory from the normal pattern of life at t_i.

6.4.3 Reeb Graph Generation

We use a down-sampling rate of one hour for Reeb graphs. This setting helps us to monitor changes in location grouping states at each hour. The threshold ϵ for spatial connect and disconnect events is set to 0.0005 GPS degrees (5.56 meters). Initially, we construct a Reeb graph from the normal activity trajectories of days 0 to 4 to model the student's typical pattern of life.

As depicted in Figure. 6.3 and Figure. 6.4, Reeb graph successfully identifies all normal

POIs as a part of the Reeb graph nodes, demonstrating its efficacy in reflecting the spatial distribution of the student's activities. Notably, an anomalous scenario depicted in Figure. 6.3 and Figure. 6.4 shows the student visiting a movie theater during school hours which is defined as a deviation from the normal. This is captured by a new Reeb graph node, highlighting its potential for identifying critical spatial anomalies.

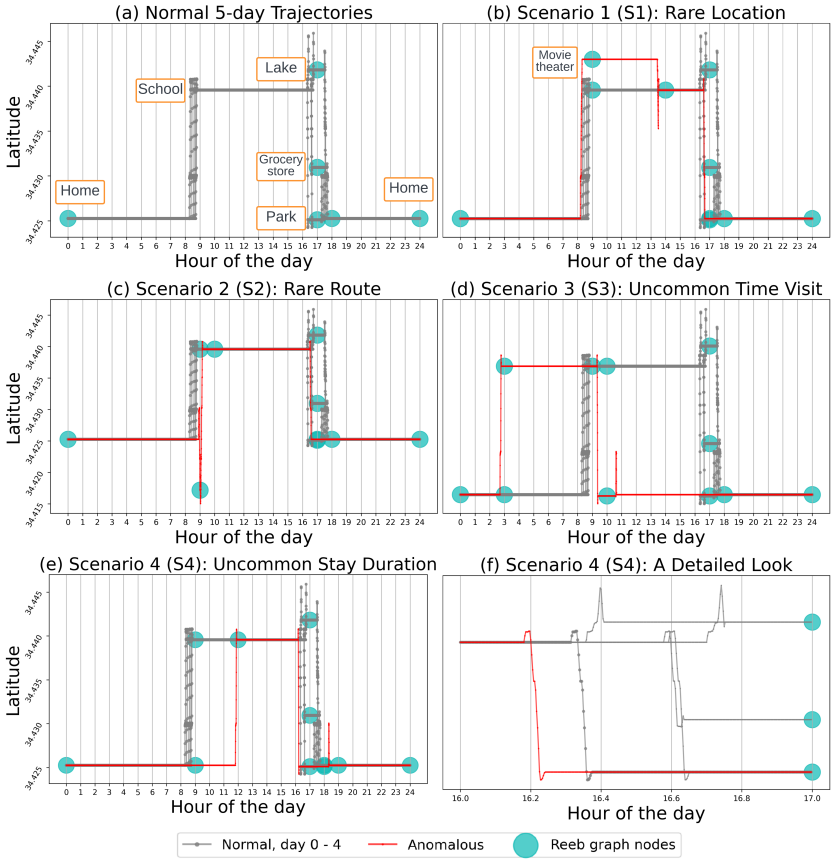

Figure 6.5: 2D Trajectory plots displaying time and latitude dimensions alongside computed Reeb graph nodes. These plots illustrate both normal and anomalous scenarios as outlined in Section 6.4.2. The detailed discussions on node generation and behavioral analysis can be found in Section 6.4.4.

6.4.4 Analysis and interpretation of scenarios using Reeb graphs

To better understand the formation of Reeb graph nodes and demonstrate the utility of the Reeb graph across all six scenarios, we generated time-latitude plots (Figure 6.5). These plots, with the hour of day on the x-axis and latitude on the y-axis, include trajectory points sampled every 10 seconds alongside Reeb graph nodes. Each plot provides a visual representation of different behavioral patterns and anomalies and illustrates Reeb graph's effectiveness in capturing anomalous trajectories for all scenarios. We explain the scenarios one by one below:

- **Figure 6.5(a)** illustrates the student's normal routine pattern, with stays at home, school, and visits to various recreational spots. Notable events include *appear* and *disappear* at the beginning and end of each day. There are three *disconnect* events around hour 17 which indicates divergences to different locations after school.*Connect* event shows trajectories getting merged back on the way home at hour 18.

- **Figure 6.5(b) for S1** depicts a rare location (lat^*, lon^*) where we visualize an abnormal visit to the movie theater, showing three additional Reeb nodes and altered connectivity events at hour 9 and 14.

- **Figure 6.5(c) for S2** captures an alternative route to school. At hour 9, instead of following the normal route, the student deviates towards a direction with a lower latitude and then returns to school. This deviation is captured by the bottom Reeb graph node at hour 9. Additionally, a *disconnect* event occurs at 9, followed by a *connect* event at hour 10 when all trajectories converge at the school.

- **Figure 6.5(d) for S3** reveals an uncommon time anomaly, where the student attends school at hour 2 and travels to the park at around hour 10, significantly

deviating from the typical schedule, but with the same POIs.

- **Figure 6.5(e) for S4** shows another time-related anomaly with a prolonged stay at home until almost hour 12, and similarly, 3 new nodes appear for the reeb graph because of *disconnect* event from the usual trajectory.

- **Figure 6.5(f) for S4** presents a detailed look at scenario 4, from hour 16 to hour 17. Since the reeb graph sample rate is one hour, the reeb graph nodes appear at hour 17 to represent the *disconnect* events in the past hour.

6.4.5 Reeb graph iteratively detects anomalous behavior of an object

In the context of detecting anomalous trajectories within real-life data (test dataset), we iteratively construct Reeb graphs on the test dataset to identify daily anomalous trajectories. An initial Reeb graph is constructed using training data with all normal trajectories. Subsequently, for each daily trajectory in the test dataset, the Reeb graph is iteratively updated day by day. To detect anomalous behaviors effectively, we compute the distance between the existing Reeb graph and every updated version that includes the additional daily trajectory. The subsequent section details our methodology for calculating this distance and presents the results derived from our case study.

6.5 Scalability of Reeb Graphs

We successfully applied Reeb graphs to a simulated dataset that is closer to a real-life distribution. This data is an extended version of the data that we described in this Chapter for proof-of-concept. Here, instead of modeling weekdays of data sampled every hour, we model the patterns over a month sampled at every 15-second interval. This

results in $m = 5760$ and $n = 30$. For this dataset, Reeb graphs models the patterns of daily activities for a simulated population of 800,000 objects. Each object is processed independently, and the Reeb graphs for the entire dataset were constructed within 7.2 hours, parallel processed across 384 CPU cores (AMD EPYC 9654 @ 3.7 GHz). We also implemented the spatial Reeb graph, ReeBundle as proposed in [8] but the quadratic time complexity with respect to m made it computationally challenging. More specifically, for $n = 7$ and $m = 5760$, the Reeb graph construction took around 4 minutes for an object. Reeb graphs is linear with respect to m and thus for the same problem setting it was able to construct Reeb graphs in approximately 12 seconds on one CPU core. This is an important advantage over spatial Reeb graphs which helps us to apply our method on large-scale datasets. Multi-processing across 384 cores enabled us to construct Reeb graphs in less than 8 hours. We also tested Reeb graphs on medium-sized data with 10,000 objects over a period of one week, Reeb Graphs were computed in approximately 5.5 minutes. The above experiments show the applicability of Reeb graphs in modeling object's data at different resolutions (weekly, monthly, yearly) and also emphasize the scalability of the proposed algorithm.

6.6 Discussion

In this Chapter, we proposed a general Reeb graph-based approach to model the patterns of normalcy using day-to-day human trajectory data. The proposed Reeb graphs abstract large-scale spatio-temporal data into a comprehensible topological construct. We design distinct real-life anomalous scenarios, develop trajectory generation methods, and provide a thorough interpretation of Reeb graph results. Our main contributions are summarized below:

- We present a novel Reeb graph-based approach to model the day-to-day activities

of a given agent.

- We discuss the algorithm and its time complexity demonstrating the scalability of the proposed method.
- We design normal and anomalous scenarios, describe the methods for trajectory generation and present detailed experiments on the interpretation and analysis of Reeb graphs.

Chapter 7

Conclusion and Future Work

This thesis presents the theory of Reeb graphs for modeling spatial trajectories, the implementation algorithms, analysis of computational complexity, and applications. Reeb graphs provide a unique way to model the human brain connectome as a graph network that is sparse and preserves underlying geometrical information. We presented the ReeBundle pipeline to model the diffusion MRI data that captures the white matter fiber connectivity in the brain. Reebundle is robust to noise and is a computationally efficient framework to characterize white matter fibers from a topological point of view. We demonstrate that the proposed approach captures the fiber branching behavior in the presence of noise and is further validated on synthetic and real dMRI datasets. ReeBundle parameters ϵ, α, and δ control the granularity of the bundling desired – small values allow only very dense sets of streamlines for grouping, while larger values of ϵ relax the groups and allow longer, larger, and sparser groups to form. These parameters can be tuned to apply the pipeline to a wider set of problems, from comparing tractography algorithms to studying neurological disorders. For neuroscience applications, we recommend using the default parameters as an initial set to begin the study. For general analysis, we recommend setting the ϵ parameter to around 3 mm, which is close to the minimum

dMRI resolution of 2 mm. Sensitivity analysis indicates the algorithm's greater sensitivity to ϵ, so, fixing the other parameters and varying ϵ should be sufficient for most applications. We also introduce the Reeb graph-based topological distance score that quantifies the relative topological difference between two bundle tracts. One limitation of the score is that it is not normalized to range from 0-1, which implies that it is not ideal for a standardized comparison across subjects and atlases. Thus, it does not give an absolute metric of the distance between two Reeb graphs. However, this is by choice since the distances between two graphs can be very large in some cases. We summarize the Reeb graph applications presented in this thesis below along with some future directions.

7.1 Tractography Evaluation

Our ReeBundle pipeline depends on the quality of streamlines and at the same time can be used for quality assessment of the tracts. Therefore, we use Reeb graphs to develop a new topological evaluation and validation method for tractography algorithms. We present ReTrace, a novel graph matching-based topological evaluation and validation method for tractography algorithms. ReTrace uses a Reeb graph whose nodes and edges capture the topology of white matter fiber bundles. We evaluate the performance of 96 algorithms from the ISMRM Tractography Challenge and the standard algorithms implemented in DSI Studio for the population-averaged Human Connectome Project (HCP) dataset. The existing evaluation metrics such as the f-score, bundle overlap, and bundle overreach fail to account for fiber continuity resulting in high scores even for broken fibers, branching artifacts, and mis-tracked fiber crossing. In contrast, we show that our approach effectively penalizes the incorrect tracking of fibers within bundles while concurrently pinpointing positions with significant deviation from the ground truth. The ultimate goal of the analysis pipeline is to guide tractography algorithm design

choices. The future work in this direction is clear: can one design a topologically-aware tractography algorithm that uses Reeb graphs in a feedback loop to iteratively improve the quality of tracking?

7.2 Early Diagnosis and Tracking of Neurological Disorders

Fingerprinting the human brain. Our method meaningfully describes the bundle tracts without oversimplifying them into skeletons. For the International Society for Magnetic Resonance in Medicine (ISMRM) dataset, Reeb graph handles the morphology of the white matter tract configurations due to branching and local ambiguities in complicated bundle tracts like anterior and posterior commissures. To quantify the difference between the connections, we introduce a new Reeb graph-based distance metric that quantifies the topological differences in bundle comparison. For the longitudinal repeated measures in the Cognitive Resilience and Sleep History (CRASH) dataset, repeated scans of a given subject acquired weeks apart lead to provably similar Reeb graphs that differ significantly from other subjects, thus highlighting our method's potential for clinical fingerprinting of brain regions. Reeb graph and topological distance metric are sensitive to variations in brain structure, such as those observed in the developing brain and in the presence of tumors. This sensitivity is beneficial in the context of longitudinal studies of brain development and the topological evolution of the brain. This thesis highlights the potential utility of this metric in tracking Alzheimer's disease progression using the ADNI dataset, and evaluating the effects of surgical interventions for brain tumor using the OpenNeuro brain tumor datasets.

Quantification of iNPH Shunt Surgery Efficacy In this thesis, we discussed

how different imaging modalities can be used together to solve neurological disorders. In particular, we discussed how the impact of shunt surgery for iNPH subjects can be quantified using CT scans. Diffusion MRI is collected along with CT scans for such patients. Future applications could include Reeb graphs for dMRI combined with CT to aid in interpreting and localizing diagnostic symptoms related to gait, cognition, and bladder function.

In our study, we demonstrated that quantifiable changes can be detected in ventricular volume in iNPH patients that undergo ventriculo-peritoneal shunt procedures. Our study developed an AI tool to provide automated measurement of cerebral ventricular volume over time for iNPH patients before and after the shunt surgery. We showed that our proposed metric significantly correlates with the subjective clinical diagnosis over time. The metric also accurately models non-responders who show significant CSF drainage and improved symptoms post shunt adjustment. The iNPH AI tool offers a simple drag-and-drop style interface to a computationally fast method, taking just 2-3 minutes for brain segmentation and less than 10 seconds for volumetric computation. This serves as a valuable clinical tool for enhancing patient care. However, our study may be limited by the absence of patients with fixed valves who required shunt removal due to severe over-drainage symptoms. Our patient selection was biased as we selectively included patients who showed improvement after shunt surgery. Additionally, as iNPH typically affects elderly patients with co-morbidities such as Parkinson's or Alzheimer's disease, the findings may not be generalizable to those populations.

In the light of these results, we anticipate many different research directions stemming from application of our proposed method to neurological conditions. For iNPH, an interesting research direction is to set a definitive threshold for the decrease in VP shunt surgery that correlates with "successful" cerebrospinal fluid drainage, akin to established metrics like Evan's Index. From the representative data used in this study, few patients

had multiple pre- surgery scans where we observe that the iNPH condition deteriorates before surgery resulting in increase in CVV metric with time until surgery (see Supplementary Table 1 [66]). This will be further investigated in future. This direction will require collecting multiple scans before surgery and after surgery.

7.3 Whole Brain Analysis

ReeBundle is intended for tract-specific analysis of individual white matter bundles or sub-pathways within a specific region of interest. However, an integration of Reebundle with tract segmentation methods can be used for whole brain analysis as well. White matter pathways in the human brain typically consist of different sub-pathways (bundle tracts). These sub-pathways could include the different geometry shapes, e.g., CC and cingulum bundle. ReeBundle is intended for tract-specific analysis of individual white matter bundles or sub-pathways within a specific region of interest. However, an integration of Reebundle with tract segmentation methods can be used for whole brain analysis as well. A possible extension of our method for unified brain analysis is illustrated in Figure. 7.1.

Another interesting direction for future research would be to explore the structure-behavior relationships, particularly for assessing changes over time. In this thesis, we show the visualization, quantification, and the analysis of neurological disorders. It is yet unexplored how these structural models of streamlines link to the behavioral models of functional MRI. New insights into structural and functional connectivity relationships in the brain can be explored by integrating these approaches. The critical points captured in Reeb graphs can serve as potential biomarkers in diseases of the nervous system that alter white matter morphology.

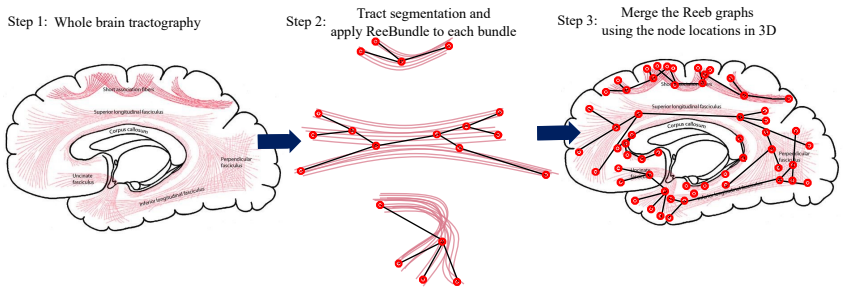

Figure 7.1: **Unified topology analysis of the whole brain tractogram.** The first step in whole brain analysis with Reebundle would be to obtain tract segmentation of the whole brain tractograph. Then, using each anatomical bundle as an input, as demonstrated in the ISMRM datasets, the Reeb graph can be extracted and analyzed. Finally, the results can be combined either through quantitative analysis to obtain an overall score of the topological distance or by merging the Reeb graphs to obtain a unified Reeb graph that can then be compared.

7.4 Reeb Graphs Beyond Neuroscience

Our method is not limited to neuroscience and is general-purpose in its applicability. Towards that, we presented a general algorithm for structure discovery in spatio-temporal trajectories. Human behavior typically follows a pattern of normalcy in day-to-day activities. This is marked by recurring activities within specific time periods. We model this behavior using Reeb graphs where any deviation from usual day-to-day activities is encoded as nodes in the Reeb graph. This approach helps us model patterns of life in human GPS trajectories (akin to a fingerprint). We propose a Reeb graph-based approach to model the patterns of normalcy using day-to-day human trajectory data from GPS sensors. The proposed Reeb graphs abstract large-scale spatio-temporal data into a comprehensible topological construct. We design distinct real-life anomalous scenarios, develop trajectory generation methods, and provide a thorough interpretation of Reeb graph results. Similar to the neuroscience applications, the parameters of Reeb graphs can control the granularity of the model according to different problem settings. But, false positives can impact the accuracy of our model. One explanation for this is the inherent stochasticity of general human behavior.

Another application is a quantifiable sanity check for raw trajectory data such as teleports. Our experiment settings un this thesis are based on the assumption that each agent is independent and the activities conducted by one agent are not related to the other. However, agents in a given population influence the behavior of each other. Such correlations could serve as additional features to our existing model. With Reeb graphs, we have the flexibility to introduce more parameters and features to robustly support the data abstractions by redefining level sets. For example, level sets can be defined on geo-foundation points of interests as shown in Figure 7.2. Geo-foundational features describe the nature of each location that the agent visited such as residential, commercial, recreational, etc. Nodes of the Reeb graphs can be labeled with such domain-specific information. Such representations can be used as an input to data-driven methods instead of directly using deep learning methods on raw GPS trajectories.

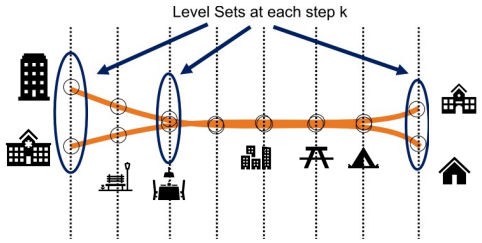

Figure 7.2: **Redefine level sets on the semantic points of interests.** Reeb graph has the flexibility to introduce more parameters and features to robustly support the data abstraction by redefining level sets. Geo-foundational features describe the nature of each location the agent visited such as residential, commercial, recreational, etc. Nodes of the Reeb graphs can be labeled with such domain-specific information.

7.5 Concluding Thoughts

Big data-driven analytics has played a crucial and transformative role in many applications, including health, education, transportation, and environmental sciences. Much of this data is unstructured and in differest resolution. This vast pool of routinely col-

lected unstructured multimodal data represents a wealth of unexploited opportunities. If we can successfully link and analyze this time or space evolving data across different modalities, spatial and temporal resolutions, it would yield new insights for addressing complex problems. Reeb graphs can provide a powerful means to represent and visualize dynamics of multimodal data, ultimately aiding in clustering and structure or semantic discovery. Such comprehensive data analysis is essential for realizing the full potential of multimodal data towards new scientific discoveries. Many interesting future research directions branch from the work presented in this thesis. These efforts can help in developing localized topological context that helps in generalizing the methods to new and unseen domains.

Bibliography

[1] S. Alzola-Aldamizetxebarria, L. Fernández-Méndez, D. Padro, J. Ruíz-Cabello, and P. Ramos-Cabrer, *A comprehensive introduction to magnetic resonance imaging relaxometry and contrast agents*, *ACS omega* **7** (2022), no. 42 36905–36917.

[2] D. Le Bihan, *Diffusion MRI: what water tells us about the brain*, *EMBO molecular medicine* **6** (2014), no. 5 569–573.

[3] K. Maier-Hein, P. Neher, J.-C. Houde, E. Caruyer, A. Daducci, T. Dyrby, B. Stieltjes, and M. Descoteaux, *Tractography challenge ismrm 2015 data*, May, 2015.

[4] N. Ta, Y. Zhao, and Y. Chai, *Built environment, peak hours and route choice efficiency: An investigation of commuting efficiency using gps data*, *Journal of Transport Geography* **57** (2016) 161–170.

[5] M. Hast, K. M. Searle, M. Chaponda, J. Lupiya, J. Lubinda, J. Sikalima, T. Kobayashi, T. Shields, M. Mulenga, J. Lessler, *et. al.*, *The use of gps data loggers to describe the impact of spatio-temporal movement patterns on malaria control in a high-transmission area of northern zambia*, *International Journal of Health Geographics* **18** (2019) 1–18.

[6] L. Wu, L. Yang, Z. Huang, Y. Wang, Y. Chai, X. Peng, and Y. Liu, *Inferring demographics from human trajectories and geographical context*, *Computers, Environment and Urban Systems* **77** (2019) 101368.

[7] S. Shailja, A. Zhang, and B. Manjunath, *A computational geometry approach for modeling neuronal fiber pathways*, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 175–185, Springer, 2021.

[8] S. Shailja, V. Bhagavatula, M. Cieslak, J. M. Vettel, S. T. Grafton, and B. Manjunath, *Reebundle: a method for topological modeling of white matter pathways using diffusion MRI*, *IEEE Transactions on Medical Imaging* (2023).

[9] P. Hagmann, *From diffusion mri to brain connectomics*, tech. rep., EPFL, 2005.

[10] C. A. Clark, T. R. Barrick, M. M. Murphy, and B. A. Bell, *White matter fiber tracking in patients with space-occupying lesions of the brain: a new technique for neurosurgical planning?*, *NeuroImage* **20** (2003), no. 3 1601–1608.

[11] P.-Y. Kao, S. Shailja, J. Jiang, A. Zhang, A. Khan, J. W. Chen, and B. Manjunath, *Improving patch-based convolutional neural networks for MRI brain tumor segmentation by leveraging location information*, *Frontiers in neuroscience* **13** (2020) 1449.

[12] M. Lazar, A. Alexander, P. Thottakara, B. Badie, and A. Field, *White matter reorganization after surgical resection of brain tumors and vascular malformations*, *American journal of neuroradiology* **27** (2006), no. 6 1258–1271.

[13] J. Gillard, N. Papadakis, K. Martin, C. Price, E. Warburton, N. Antoun, C. L. Huang, T. Carpenter, and J. Pickard, *Mr diffusion tensor imaging of white matter tract disruption in stroke at 3 t*, *The British Journal of Radiology* **74** (2001), no. 883 642–647.

[14] S. H. Greenblatt, *Alexia without agraphia or hemianopsia anatomical analysis of an autopsied case*, *Brain* **96** (1973), no. 2 307–316.

[15] J. Sun, M. Cieslak, S. Grafton, and S. Suri, *A reeb graph approach to tractography*, in *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2015.

[16] H. Doraiswamy and V. Natarajan, *Efficient algorithms for computing reeb graphs*, *Computational Geometry* **42** (2009) 606–616.

[17] K. Buchin, M. Buchin, M. v. Kreveld, B. Speckmann, and F. Staals, *Trajectory grouping structure*, in *Workshop on Algorithms and Data Structures*, pp. 219–230, Springer, 2013.

[18] E. Garyfallidis, M. Brett, M. M. Correia, G. B. Williams, and I. Nimmo-Smith, *Quickbundles, a method for tractography simplification*, *Frontiers in neuroscience* **6** (2012) 175.

[19] K. H. Maier-Hein, P. F. Neher, J.-C. Houde, M.-A. Côté, *et. al.*, *The challenge of mapping the human connectome based on diffusion tractography*, *Nature Communications* **8** (2017), no. 1 1–13.

[20] S. Parsa, *A deterministic o (m log m) time algorithm for the reeb graph*, *Proceedings of the 28th Annual Symposium on Computational Geometry* (2012) 269–276.

[21] D. D. Sleator and R. E. Tarjan, *A data structure for dynamic trees*, in *Proceedings of the thirteenth annual ACM symposium on Theory of computing*, pp. 114–122, 1981.

[22] S. Shailja, J. W. Chen, S. T. Grafton, and B. Manjunath, *Retrace: Topological evaluation of white matter tractography algorithms using reeb graphs*, in *International Workshop on Computational Diffusion MRI*, pp. 177–191, Springer, 2023.

[23] D. Arienzo, A. Leow, J. A. Brown, L. Zhan, J. GadElkarim, S. Hovav, and J. D. Feusner, *Abnormal brain network organization in body dysmorphic disorder*, *Neuropsychopharmacology* **38** (2013), no. 6 1130–1139.

[24] M. García-Gomar, L. Concha, S. Alcauter, J. S. Abraham, J. Carrillo-Ruiz, G. C. Farfan, and F. V. Campos, *Probabilistic tractography of the posterior subthalamic area in Parkinson's disease patients*, *Journal of Biomedical Science and Engineering* (2013).

[25] D. K. Jones, *Challenges and limitations of quantifying brain connectivity in vivo with diffusion MRI*, *Imaging in Medicine* **2** (2010), no. 3 341.

[26] L. Zhan, J. Zhou, Y. Wang, Y. Jin, N. Jahanshad, G. Prasad, T. M. Nir, C. D. Leonardo, J. Ye, P. M. Thompson, *et. al.*, *Comparison of nine tractography algorithms for detecting abnormal structural brain networks in alzheimer's disease*, *Frontiers in Aging Neuroscience* **7** (2015) 48.

[27] J.-S. Wu, L.-F. Zhou, W.-J. Tang, Y. Mao, J. Hu, Y.-Y. Song, X.-N. Hong, and G.-H. Du, *Clinical evaluation and follow-up outcome of diffusion tensor imaging-based functional neuronavigation: a prospective, controlled study in patients with gliomas involving pyramidal tracts*, *Neurosurgery* **61** (2007), no. 5 935–949.

[28] P. Fillard, M. Descoteaux, A. Goh, S. Gouttard, B. Jeurissen, J. Malcolm, A. Ramirez-Manzanares, M. Reisert, K. Sakaie, F. Tensaouti, *et. al.*, *Quantitative evaluation of 10 tractography algorithms on a realistic diffusion MR phantom*, *Neuroimage* **56** (2011), no. 1 220–234.

[29] M.-A. Côté, G. Girard, A. Boré, E. Garyfallidis, J.-C. Houde, and M. Descoteaux, *Tractometer: towards validation of tractography pipelines*, *Medical Image Analysis* **17** (2013), no. 7 844–857.

[30] L. J. Volz, M. Cieslak, and S. Grafton, *A probabilistic atlas of fiber crossings for variability reduction of anisotropy measures*, *Brain Structure and Function* **223** (2018) 635–651.

[31] S. Shailja, S. T. Grafton, and B. Manjunath, *A robust reeb graph model of white matter fibers with application to alzheimer's disease progression*, *bioRxiv* (2022).

[32] S. Shailja, V. Bhagavatula, M. Cieslak, J. M. Vettel, S. T. Grafton, and B. S. Manjunath, *Reebundle: a method for topological modeling of white matter pathways using diffusion MRI*, *IEEE Transactions on Medical Imaging* (2023) 1–1.

[33] T. Sarwar, K. Ramamohanarao, and A. Zalesky, *Mapping connectomes with diffusion MRI: deterministic or probabilistic tractography?*, *Magnetic Resonance in Medicine* **81** (2019), no. 2 1368–1384.

[34] J. Reichardt and S. Bornholdt, *Statistical mechanics of community detection*, *Physical Review E* **74** (2006), no. 1 016110.

[35] A. Hagberg, P. Swart, and D. S Chult, *Exploring network structure, dynamics, and function using networkx*, tech. rep., Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

[36] A. Mheich, M. Hassan, M. Khalil, V. Gripon, O. Dufor, and F. Wendling, *Siminet: a novel method for quantifying brain network similarity*, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **40** (2017), no. 9 2238–2249.

[37] E. Renauld, A. Théberge, L. Petit, J.-C. Houde, and M. Descoteaux, *Validate your white matter tractography algorithms with a reappraised ISMRM 2015 Tractography Challenge scoring system*, *Scientific Reports* **13** (2023) 2347.

[38] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, *Comparing images using the Hausdorff distance*, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **15** (1993), no. 9 850–863.

[39] R. Y. K. Chang, N. Etheridge, A. S. Nouwens, and P. R. Dodd, *Swath analysis of the synaptic proteome in alzheimer's disease*, *Neurochemistry International* **87** (2015) 1–12.

[40] A. Zhang, S. Shailja, C. Borba, Y. Miao, M. Goebel, R. Ruschel, K. Ryan, W. Smith, and B. Manjunath, *Automatic classification and neurotransmitter prediction of synapses in electron microscopy*, *Biological Imaging* **2** (2022) e6.

[41] A. Zalesky, L. Cocchi, A. Fornito, M. M. Murray, and E. Bullmore, *Connectivity differences in brain networks*, *NeuroImage* **60** (2012), no. 2 1055–1062.

[42] M. John, T. Ikuta, and J. Ferbinteanu, *Graph analysis of structural brain networks in alzheimer's disease: beyond small world properties*, *Brain Structure and Function* **222** (2017), no. 2 923–942.

[43] F.-C. Yeh, *Shape analysis of the human association pathways*, *NeuroImage* **223** (2020) 117329.

[44] E. Garyfallidis, M. Brett, B. Amirbekian, A. Rokem, S. Van Der Walt, M. Descoteaux, I. Nimmo-Smith, and D. Contributors, *Dipy, a library for the analysis of diffusion mri data*, *Frontiers in Neuroinformatics* **8** (2014) 8.

[45] B. Q. Chandio, S. L. Risacher, F. Pestilli, D. Bullock, F.-C. Yeh, S. Koudoro, A. Rokem, J. Harezlak, and E. Garyfallidis, *Bundle analytics, a computational framework for investigating the shapes and profiles of brain pathways across populations*, *Scientific Reports* **10** (2020), no. 1 1–18.

[46] J. D. Yeatman, R. F. Dougherty, N. J. Myall, B. A. Wandell, and H. M. Feldman, *Tract profiles of white matter properties: automating fiber-tract quantification*, *PloS One* **7** (2012), no. 11 e49790.

[47] M. Chamberland, S. St-Jean, C. M. Tax, and D. K. Jones, *Obtaining representative core streamlines for white matter tractometry of the human brain*, in *Computational Diffusion MRI: International MICCAI Workshop, Granada, Spain*, pp. 359–366, Springer, 2019.

[48] W. Li and X. Hu, *Robust tract skeleton extraction of cingulum based on active contour model from diffusion tensor mr imaging*, *PloS one* **8** (2013), no. 2 e56113.

[49] Y. Shinagawa, T. L. Kunii, and Y. L. Kergosien, *Surface coding based on morse theory*, *IEEE Computer Graphics and Applications* **11** (1991), no. 5 66–78.

[50] "`https://github.com/UCSB-VRL/ReeBundle`."

[51] D. S. Bassett, E. Bullmore, B. A. Verchinski, V. S. Mattay, D. R. Weinberger, and A. Meyer-Lindenberg, *Hierarchical organization of human cortical networks in health and schizophrenia*, *Journal of Neuroscience* **28** (2008) 9239–9248.

[52] Distances, *https://netrd.readthedocs.io/en/latest/distance.html*, 2022.

[53] K. Maier-Hein, P. Neher, J.-C. Houde, M.-A. Côté, Garyfallidis, *et. al.*, *Tractography-based connectomes are dominated by false-positive connections*, 2016.

[54] S. M. Thurman, N. Wasylyshyn, H. Roy, G. Lieberman, J. O. Garcia, Asturias, *et. al.*, *Individual differences in compliance and agreement for sleep logs and wrist actigraphy: a longitudinal study of naturalistic sleep in healthy adults*, *PloS One* **13** (2018) e0191883.

[55] N. Lauharatanahirun, K. Bansal, S. M. Thurman, J. M. Vettel, B. Giesbrecht, S. Grafton, Elliott, *et. al.*, *Flexibility of brain regions during working memory curtails cognitive consequences to lack of sleep*, *arXiv:2009.07233* (2020).

[56] C. R. Buchanan, *Structural brain networks from diffusion mri: methods and application*, 2015.

[57] B. B. Avants, N. Tustison, G. Song, *et. al.*, *Advanced normalization tools (ants)*, *Insight j* **2** (2009), no. 365 1–35.

[58] F.-C. Yeh, *Population-based tract-to-region connectome of the human brain and its hierarchical topology*, *Nature Communications* **13** (2022), no. 1 4933.

[59] "`https://dsi-studio.labsolver.org/`."

[60] F.-C. Yeh, V. J. Wedeen, and W.-Y. I. Tseng, *Generalized q-sampling imaging*, *IEEE Transactions on Medical Imaging* **29** (2010), no. 9 1626–1635.

[61] F.-C. Yeh, T. D. Verstynen, Y. Wang, J. C. Fernández-Miranda, and W.-Y. I. Tseng, *Deterministic diffusion fiber tracking improved by quantitative anisotropy*, *PloS One* **8** (2013), no. 11 e80713.

[62] D. Foreman-Mackey *et. al.*, *corner. py: Scatterplot matrices in python.*, *Journal of Open Source Software* **1** (2016), no. 2 24.

[63] S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, and L. Beckett, *The alzheimer's disease neuroimaging initiative*, *Neuroimaging Clinics* **15** (2005), no. 4 869–877.

[64] "`https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI\_Acknowledgement_List.pdf`."

[65] H. Aerts and D. Marinazzo, *Openneuro BTC_preop*, 2022.

[66] S. Shailja, C. Nguyen, K. Thanigaivelan, V. Bhagavatula, J. W. Chen, and B. Manjunath, *Ai-based analysis of the shunt treatment in pre-and post-surgery computed tomography brain scans of inph patients*, *medRxiv* (2023) 2023–06.

[67] B. Ades-Aron, S. Yeager, N. Miskin, E. Fieremans, A. George, and J. Golomb, *Diffusional kurtosis along the corticospinal tract in adult normal pressure hydrocephalus*, *American Journal of Neuroradiology* **39** (2018), no. 12 2218–2223.

[68] N. C. Keong, A. Pena, S. J. Price, M. Czosnyka, Z. Czosnyka, E. E. DeVito, C. R. Housden, B. J. Sahakian, and J. D. Pickard, *Diffusion tensor imaging profiles reveal specific neural tract distortion in normal pressure hydrocephalus*, *PloS one* **12** (2017), no. 8 e0181624.

[69] M. A. Williams and J. Malm, *Diagnosis and treatment of idiopathic normal pressure hydrocephalus*, *Continuum: Lifelong Learning in Neurology* **22** (2016), no. 2 579–599.

[70] K. A. Peterson, G. Savulich, D. Jackson, C. Killikelly, J. D. Pickard, and B. J. Sahakian, *The effect of shunt surgery on neuropsychological performance in normal pressure hydrocephalus: a systematic review and meta-analysis*, *Journal of neurology* **263** (2016) 1669–1677.

[71] A. M. Isaacs, M. A. Williams, and M. G. Hamilton, *Current update on treatment strategies for idiopathic normal pressure hydrocephalus*, *Current treatment options in neurology* **21** (2019) 1–24.

[72] S. Pujari, S. Kharkar, P. Metellus, J. Shuck, M. A. Williams, and D. Rigamonti, *Normal pressure hydrocephalus: long-term outcome after shunt surgery*, *Journal of Neurology, Neurosurgery & Psychiatry* **79** (2008), no. 11 1282–1286.

[73] G. Sohn, M. J. Bae, J. Park, and S. E. Kim, *Semi-quantitative analysis of periventricular gray-white matter ratio on ct in patients with idiopathic normal pressure hydrocephalus*, *Journal of Clinical Neuroscience* **101** (2022) 16–20.

[74] J. Neikter, S. Agerskov, P. Hellström, M. Tullberg, G. Starck, D. Ziegelitz, and D. Farahmand, *Ventricular volume is more strongly associated with clinical improvement than the evans index after shunting in idiopathic normal pressure hydrocephalus*, *American Journal of Neuroradiology* **41** (2020), no. 7 1187–1192.

[75] J. Virhammar, K. Laurell, K. G. Cesarini, and E.-M. Larsson, *Increase in callosal angle and decrease in ventricular volume after shunt surgery in patients with idiopathic normal pressure hydrocephalus*, *Journal of neurosurgery* **130** (2018), no. 1 130–135.

[76] J. L. Quon, M. Han, L. H. Kim, M. E. Koran, L. C. Chen, E. H. Lee, J. Wright, V. Ramaswamy, R. M. Lober, M. D. Taylor, *et. al.*, *Artificial intelligence for automatic cerebral ventricle segmentation and volume calculation: a clinical tool for the evaluation of pediatric hydrocephalus*, *Journal of neurosurgery: Pediatrics* **27** (2020), no. 2 131–138.

[77] F. Yepes-Calderon and J. G. McComb, *Eliminating the need for manual segmentation to determine size and volume from mri. a proof of concept on segmenting the lateral ventricles*, *Plos one* **18** (2023), no. 5 e0285414.

[78] A. Zhang, A. Khan, S. Majeti, J. Pham, C. Nguyen, P. Tran, V. Iyer, A. Shelat, J. Chen, and B. Manjunath, *Automated segmentation and connectivity analysis for normal pressure hydrocephalus*, *BME frontiers* (2022).

[79] P.-Y. Kao, S. Shailja, J. Jiang, A. Zhang, A. Khan, J. W. Chen, and B. Manjunath, *Improving patch-based convolutional neural networks for mri brain tumor segmentation by leveraging location information*, *Frontiers in neuroscience* **13** (2020) 1449.

[80] T. Hattori, T. Yuasa, S. Aoki, R. Sato, H. Sawaura, T. Mori, and H. Mizusawa, *Altered microstructure in corticospinal tract in idiopathic normal pressure hydrocephalus: comparison with alzheimer disease and parkinson disease with dementia*, *American journal of neuroradiology* **32** (2011), no. 9 1681–1687.

[81] B. Zhang, S. Shailja, C. Gudavalli, C. Levenson, A. Khan, and B. S. Manjunath, *Reespot: Reeb graph models semantic patterns of normalcy in human trajectories*, 2024.

[82] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wießner, *Microscopic traffic simulation using sumo*, in *The 21st IEEE International Conference on Intelligent Transportation Systems*, IEEE, 2018.