Thesis Proposal.

A GENERALIZED TOOL FOR DERIVING CONNECTOMES IN SUPPORT OF COMPUTATIONAL NEUROSCIENCE

By

David Mattie

Submitted in partial fulfillment of the
requirements for the degree of
Masters of Science in Computer Science
at
Saint Francis Xavier University
Antigonish, Nova Scotia
August 2018

Saint Francis Xavier University
Department of
Mathematics, Statistics and Computer Science

The undersigned hereby certify that they have read a thesis entitled "**Thesis Proposal. A GENERALIZED TOOL FOR DERIVING CONNECTOMES IN SUPPORT OF COMPUTATIONAL NEUROSCIENCE**" by **David Mattie** in partial fulfillment of the requirements for the degree of **Masters of Science**.

Dated: _____

Supervisor: _____
Dr. Jacob Levman

St. Francis Xavier University

August 2018

Author:	David Mattie
Title:	Thesis Proposal. A GENERALIZED TOOL FOR DERIVING CONNEC
Department:	Mathematics, Statistics and Computer Science
Faculty:	Science
Convocation	August 2018

to Dad

Abstract

This research intends to establish a framework in which to perform connectomic measurements of tens of thousands of data points across all major brain regions. A user shell will be developed to first derive these connectomes, and then provide mechanisms to quickly extract and analyze the calculated data of interest. Machine learning algorithms will demonstrate the application of computational neuroscience by discovering patterns consistent with autism patients. The objective is to provide a generalized tool that can extract data for multiple uses.

Acknowledgements

I would like to express my gratitude to my advisor Dr. Jacob Levman PhD, for giving me the opportunity for this thesis. His sharing of knowledge across many facets of the material has been helpful and encouraging. I would also like to thank Dr. James Hughes for his knowledge of machine learning and data mining. Finally, Dr. Iker Gondra for his broad experience with image parcellation and artificial intelligence.

Contents

1 Introduction **1**

1.1 Hypothesis 2

1.2 Relevance 3

1.2.1 Tractography 3

1.2.2 Current Landscape of Connectomics 9

1.2.3 Machine Learning Applications 11

1.2.4 Preliminary Discussion and Results 12

1.2.5 Approach 13

Bibliography **14**

List of Tables

List of Figures

1.1 Collected from the literature [13] 5

Chapter 1

Introduction

The structure of the human brain is extraordinarily complicated, in part because of how interconnected neurons extend between brain regions. Patterns of neural development and neuronal migration make the assessment of fiber tracts in the developing brain particularly challenging. Studying the structure and synaptic connections within the brain has been in practice since the 1970s, however it has only recently become possible to acquire and process the unprecedented quantity of data and measurements produced at a macro level in the brains of real-world clinical patients. Analysis of the entire connectome (a representation of the physical connections apparent in a patient's brain) is within reach as computational power, acquisition technologies and storage capabilities mature.

In recent years there have been dozens of tools developed to extract measures from magnetic resonance imaging (MRI) data, such as FSL and TrackVis for conducting functional MRI (fMRI), MRI, diffusion tensor imaging (DTI), and fiber tract analyses. While diverse in functionality, these tools tend to be focused on specific functional capabilities, such as eddy_correct, a tool for correcting eddy currents from diffusion

data. When combined with other tools to form a workflow pipeline, it becomes possible to achieve complex renderings of imaging data that may play a key role in improving our understanding of healthy brain development, abnormal brain development and to thoroughly map the brain's physical connections.

This research intends to leverage existing tools while creating novel pipelines to accomplish two objectives: First, provide the capability to conduct a cartesian product fiber tract analysis among a wide range of regions of interest (ROIs). Second, to provide a mechanism for conducting a rapid extract of inter-ROI measurements from all sample data. The aim is to establish a tool whereby future studies can be conducted to understand relationships among ROIs, the degree of connectivity associated with any two region pairs and determining any possible correlation with physiological conditions such as gender, age, or disorders such as autism.

1.1 Hypothesis

It is possible to derive thousands of measures such as fiber tract length, fractional anisotrophy, fiber tract count and apparent diffusion coefficient across an exhaustive list of ROI pairs to support the neuroscientific analysis of brain development as well as to create the next generation of diagnostic and disease characterization technologies.

1.2 Relevance

1.2.1 Tractography

Tractography is the only tool we have today that allows us to noninvasively model white matter trajectories throughout the brain [4]. Tractography imaging has demonstrated itself to have considerable potential in neuroscientific analyses [5] [13] [6], however, tractography has yet to become a gold standard imaging technique relied upon clinically for the management of any given medical condition.

Although current tractography methods do not track axons directly (fiber tracking is based on inferring the tract's presence based on neighbouring voxel diffusion profiles) and have varying accuracy by technique, there is clear potential value from mapping fiber bundles in white matter towards better understanding the structural organization of white matter [3] [8]. These analytic technologies can allow neuroscientists to test hypotheses (such as whether the autistic exhibit abnormally reduced or increased fiber tract connections between two key brain regions relative to neurotypical contols) that could not previously have been accomplished in a living human.

Common approaches to tractography involve tracking fiber tracts based on adjacent water diffusion profiles with the least angular deviations between voxels. This voxel-based imaging method is typically performed within a single volume that measures a few cubic millimeters, which can, in turn, involve many thousands of axons. The scale disparity between our cubic millimeter macro-scale voxel based measurements and the number of axons involved in the MRI signals acquired results in assumptions and approximations being made when researchers correlate structural connectivity and function.

Previous research has demonstrated reasonable agreement among structural, functional, and historical fiber tracking results suggesting a simple model of direct anatomical connectivity between regions of interest in the brain is capable of explaining much of the observed correlations in neural activity[17], however, traditional diffusion tensor imaging based analyses have been reported to be unreliable[9][4]. Unfortunately, there is no ability to detect the presence of synapses or to determine whether a pathway is functional. False positives and negatives are inevitable given the spatial resolution, especially in regions of heavy fiber crossing or complexity [5].

1.2.1.1 Current technology landscape

There are a number of technologies actively used to perform tractography. There are currently three categories of automated tract segmentation:

- ROI based approach uses information from a common atlas space that is registered to the subject in order to perform ROI extraction. Streamlines can then be filtered according to their spatial relation to the ROI using white matter atlases[15].

- Clustering-based segmentation groups streamlines into spatially coherent clusters. These clusters are then either manually or automatically assigned to anatomically meaningful fiber tracts[3].

- Direct segmentation attempts to improve the efficiency and simplicity of tract segmentation by producing complete tract segmentation from the input images. While novel, the quality currently achieved by these approaches has been limited until recently. TractSeg has produced very accurate and efficient results using a Convolutional Neural Network to directly segment white matter tracts[16].

Each method approaches the problem differently, and so comparisons are challenging, however, an efficient way to assess segmentation performance is to use a Dice Score[13] as shown in Figure 1.1 . While it may make sense to choose the technology with the highest Dice Score and move on, there are advantages and disadvantages to each.

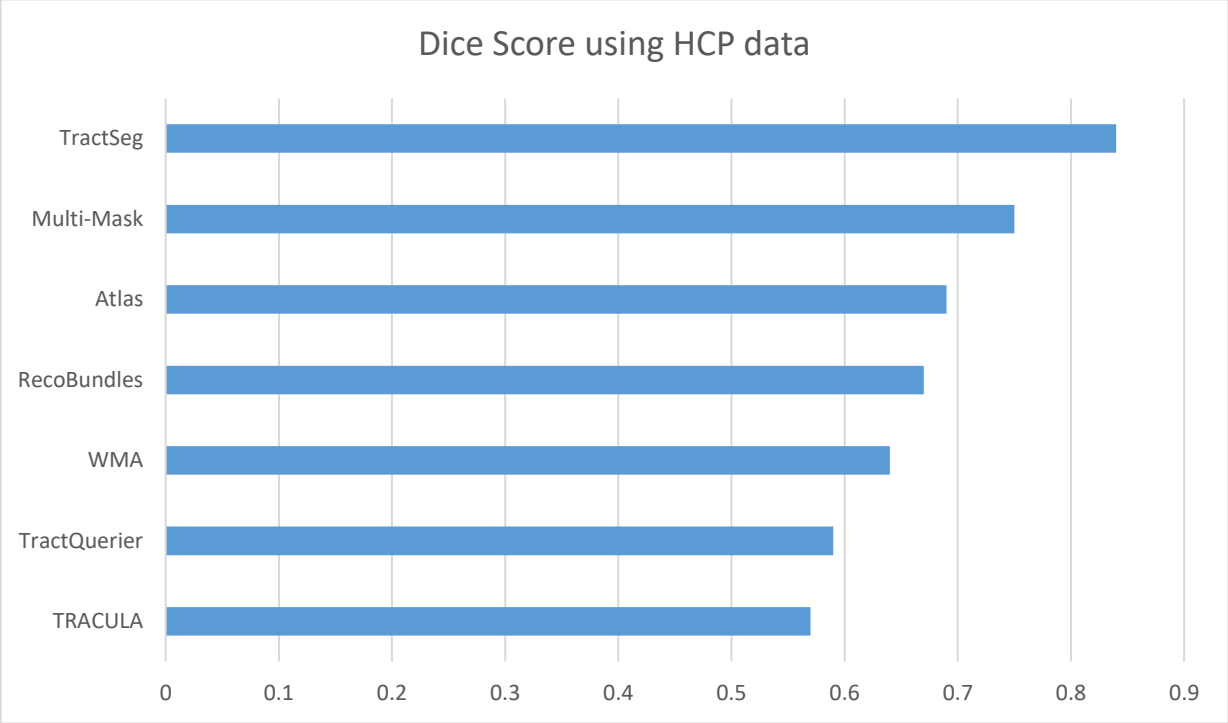

Figure 1.1: Collected from the literature [13]

1.2.1.1.1 TRACULA TRActs Constrained by Underlying Anatomy (TRACULA) uses probabilistic tracking and a predefined atlas of the underlying anatomy to segment tracts. This tool offers the opportunity to automatically reconstruct white matter bundles without requiring expert manual parcellation. This is valuable for its ability to derive white matter segments that deviate from standard atlases. For example, patients with brain injury such as hippocampal sclerosis where differences between the injured brain and healthy tract tissue characteristics can be studied[6]. Noticeably, TRACULA

provides the weakest Dice Score of comparable tractography technologies. The weaknesses could stem from the probabilistic approach to parecellation.

1.2.1.1.2 TractQuerier TractQuerier provides a White Matter Query Language (WMQL) designed to extract tracts based on where streamlines start, end, and pass through. This method makes it possible to anatomically label white matter anatomy across patients. A key differentiator of this technology is found in its ability to textually label anatomical structures that have not been catalogued. This is different from other common approaches to automatic extraction of white matter tracts because it does not rely on fixed sets of anatomical definitions [15].

1.2.1.1.3 WhiteMatterAnalysis (WMA) WhiteMatterAnalysis clusters streamlines across subjects and generates a cluster atlas. Clusters are assigned to anatomical tracts manually, and future anatomical tract delineation is made by registering new subjects to the atlas. A pre-trained cluster atlas is provided by the tool, however there is only a small number of tracts (10) within the atlas. There are two weaknesses in this tool: The low tract count (and significant manual effort required to expand the atlas), as well as the computational resources required to identify new clusters [8].

1.2.1.1.4 RecoBundles RecoBundles can be used to find streamlines in a subject based on tracts from a strong anatomical reference subject. One of the objectives of this tool is to reduce the number of invalid streamlines and white matter bundles to defend against biases in tractometry analysis [3]. RecoBundles is useful for detecting deformed and interrupted bundles going through or around tumor areas due to its ability to adapt to sharp changes or incomplete data. RecoBundles uses supervised machine learning,

and relies on a reliable bundle model as input to detect relevant bundles.

1.2.1.1.5 Atlas Registration Fractional Anisotrophy maps of several subjects can be averaged to an atlas to produce a probability map in which to base future subject segmentation. The pipeline steps needed to produce this atlas is complex, computationally intensive, and tedious to fine tune, however, it will produce a strong result if skilled anatomists are relied upon for quality checking to avoid propagation of subtle errors[16].

1.2.1.1.6 Multi-Mask An alternative approach to atlas registration is to register the masks of single training subjects to a test subject instead of against an averaged atlas. This can reduce the blurring of details to some extent. While this may improve the blurring effect that can occur when registering to an atlas based on group averages, the complexity and fine tuning required to achieve a reliable outcome still demands a complex pipeline and diligence[16].

1.2.1.1.7 TractSeg TractSeg is the newest and most promising technology depicted in Figure 1. TractSeg has been used to provide complete and accurate segmentations using fully convolutional neural networks to directly segment white matter tracts. Not only is this method efficient, it is also less affected by the reduction in resolution seen in ROI and clustering based analysis[16].

1.2.1.2 Technology Roadmap

While today's tractography methods are continually developing, they hold enormous potential value to neuroscience, even though we are still unable to identify and directly measure axon connections within the brain. This lack of fidelity certainly does not

mean fiber tract discovery will be abandoned, rather it is expected to improve with ever-evolving techniques. As techniques improve, the data collected is liable to play an important role in the assessment of neurological pathways in the brain.

The fundamental weaknesses in today's technology lies in both the processing time required to uncover individual patient fiber tracts, the resolution and accuracy of those tracts and a lack of gold-standard connectomics analytic techniques to extract, process and analyze the many neural fiber tracts detectable in a patient imaged with MRI. As the neuroscience community works to achieve these goals, we are seeing novel approaches to image parcellation as advances using machine-learning (ML) techniques mature into this field. An important dynamic with regards to ML approaches lies in the ability to reduce the complexity of tractography pipelines. It is possible that the strength of ML algorithms' ability to process and model diffusion data may one day result in performance levels that negate our need for a traditional tractographic-processing pipeline, possibly through future deep-learning architectures.

Ideally, tractography techniques will mature to the point where reliable and, efficient performance can be obtained at high-resolution. Ideally, neuroscientists can simply run a computational program to retrieve a data rich array of tracts in which to interrogate for possible abnormalities associated with a pathology under investigation. This could not only involve a reduction of complexity of tractography, but could also enhances interoperability of solutions, improves flexibility, and provides a foundation in which to implement new uses of tractographic data. At some point, it may be possible to consolidate solutions that exist globally into a single library of computer programs depending on the requirement. As solutions continue to evolve, a shared library or specific computer program will hopefully benefit the research community by encapsulating the

internal complexity of the brain.

1.2.2 Current Landscape of Connectomics

1.2.2.1 Human Connectome Project

The Human Connectome Project's (HCP) stated goal is to build a "network map" that will shed light on the anatomical and functional connectivity within the heathy human brain, as well as produce a body of data that will facilitate research into brain disorders such as dyslexia, autism, Alzheimer's Disease, and schizophrenia [14] [12]. This is a 5 year project sponsored by sixteen components of the National Institutes of Health, and split in to two consortia of research institutes. This consortium that aims to characterize human brain connectivity and function in a population of 1200 healthy adults and to enable detailed comparison between brain circuits, behavior, and genetics at the level of individual subjects. The imaging data acquired for this project include four data modalities and is being made open access.

- rfMRI, resting state functional magnetic resonance imaging
- tfMRI, task-based functional magnetic resonance imaging
- dMRI, diffusion magnetic resonance imaging, and
- sMRI, structural magnetic resonance imaging

This data can be used with standard fiber reconstruction techniques to map individual tractograms from patient samples. For example, probabilistic tractography has been applied to some of these datasets using FSL's existing probabilistic tractography approaches to generated connectomes[11].

1.2.2.2 Thousand Functional Connectome Project

There are a number of projects carrying out large scale brain MRI mapping on different populations, though none with the same aspirations as the Human Connectome Project. The Thousand Functional Connectome project has published their dataset after gathering R-fMRI images of 1414 volunteers demonstrating a universal architecture of positive and negative functional connections [1]. The Alzheimer's Disease Neuroimaging Initiative was initiated with the aim of advancing Alzheimer's Disease research with the distinct goals of studying each phase of Alzheimer's. Depending on the sub-study, there are between 600 and 2000 participant samples that have been imaged. In comparison, the HCP is unique in terms of the diversity of imaging modalities and the richness of the behavioral and genetic information being collected.

1.2.2.3 Clinical

Since 2007, Boston Children's Hospital (BCH) has been collecting high angular resolution diffusion data clinically with a consistent set of 3 Tesla MRI magnets. This data provides a unique opportunity to investigate the potential for studying the use of connectomics technologies in the context of what can be accomplished with clinical data. Existing connectomics projects target healthy brain development (as in the Human Connectome Project) or Alzheimer's Disease (as in the Thousand Functional Connectome project), however, clinical data provices unique opportunities for further analysis. Clinical data has been acquired at BCH for patients with a wide variety of medical conditions, including autism, attention-deficit hyperactivity disorder, cerebral palsy, multiple sclerosis etc. as well as including imaging of neurotypical patients. Thus creating connectomics technologies that can be applied to real-world BCH data will fa-

cilitate the investigation of possible fiber tract-based structural abnormalities associated with a variety of medical conditions and may represent a key component playing a role in the next generation of diagnostic clinical technologies.

1.2.3 Machine Learning Applications

Machine learning (ML) is a branch of artificial intelligence based on the idea that systems can learn from data by identifying patterns, making classifications, predictions, and ultimately making decisions about the data with minimal human involvement. ML solutions are typically organized into two categories: supervised and unsupervised learning. The core objective of a ML algorithm is to generalize from data (learn) a model that can be reapplied to future data. Applications are wide reaching, especially in data rich research fields such as neuroscience or bioinformatics in general.

There are a number of ways in which machine learning can be leveraged from a neuroscience perspective. With respect tractography, TractSeg (previously discussed) has shown exciting promise both with speed of execution and accuracy. TractSeg uses a form of supervised machine learning called neural networks to uncover fiber bundles and tracts in the brain[16]. Machine Learning has been leveraged to diagnose Parkinson's disease based on supervised learning to detect sensitive biomarkers based on MR images[10]. In this research, machine learning will be used to demonstrate the viability of supervised learning to create biomarkers indicative of autism.

1.2.3.1 Techology Landscape

There are currently two approaches to understanding the axonal connections found within the human brain. Anatomical studies (such as dissection or diffusion tensor

imaging) and electrostimulation mapping investigations are used to map subcortical pathways as a way to model the human brain. Understanding this axonal connectivity is crucial to build models of cognition, behavior, and even consciousness[2]. Once connectomes are produced, it is the study of the data that is most valuable. This research is based on the theory that connectomic data can drive innovation, find opportunities for detection and diagnosis and provide insights into brain disorders or brain disease.

There are a number of problems with the current state of related technology. Reliability, computational requirements and utility of the resultant technologies are a reflection of the broad and diverse expectations of neuroscience as we seek to learn new methods to improve accuracy and consistency of the known fiber tracts within the human brain. This research seeks to provide a connectome reporting user shell that offers a layer of abstraction from the myriad of technologies employed in biomarker extraction, including MRI scanners, data processing techniques and brain atlases. The goal of this work is to provide the framework for a computational interface in which data can be quickly extracted and manipulated in data science tools such as Matlab, R, Python, and SPSS regardless of the state of evolution of underlying tractography techniques and connectomic data available. This research goal also involves the development of connectomic processing and analysis software tools that work on real-world clinical data from BCH.

1.2.4 Preliminary Discussion and Results

In January 2018, [7] illustrated the investigative potential of performing connectomics-style analysis and demonstrated the feasibility of creating novel pipelines that use current technologies to perform regionally focused clinical connectivity studies. This research encouraged future functionality whereby clinical analysis of other major neural fiber

tracts throughout the brain could become possible. This capability when executed at scale would enable analysis across hundreds or thousands of patient samples as we seek to understand correlations and patterns among ROIs in the entire connectome. This will provide a mechanism to discover insights, make predictions, and ultimately may create new diagnostic and disease characterization technologies for patients with autism, attention deficit hyperactivity disorder, cerebral palsy and more.

1.2.5 Approach

A python library will be created to execute a workflow designed to collect a combined structural (T1) and high angular resolution diffusion imaging (HARDI) analysis of all major neural fiber tracts throughout the brain. Python will be used for its versatility and extensive application within the neuroimaging and data science communities. This analysis will produce thousands of metrics for each sample in a single pass, and store those metrics for future recall.

The same library will extract, process and report data related to regions of interest in future studies. The final output can be a narrowly focused report or a very wide data set (that includes many thousands of feature measurements), depending on the study. Tools such as MATLAB or R can leverage this data quickly and easily. As studies reveal new information, the same tool can be used to quickly extract related information without requiring a full recalculation. This agility in data discovery allows the researcher to focus on valuable data rather than become delayed with building new processing pipelines.

Bibliography

[1] Bharat B. Biswal. Toward discovery science of human brain function. *PNAS*, pages 4734–4739, 2010.

[2] Hafues Daffau. Neurobiology of language. pages 129–137, 2016.

[3] Eleftherios Garyfallidis, Mark-Alexandre Cote, Francois Rehault, Jasmine Sidhu, Janice Hau, Laurent Petit, David Fortin, Stephen Cunanne, and Maxime Descoteaux. Recognition of white matter bundles using local and global streamline-based registration and clustering. *NeuroImage*, pages 283–295, 2018.

[4] Saad Jbabdi and Heidi Johansen-Berg. Tractography: Where do we go from here? *Brain Connect*, pages 169–183, 2011.

[5] Heidi Johansen-Berg and Timothy E. J. Behrens. Just pretty pictures? what diffusion tractography can add in clinical neuroscience. *Current Opinions in Neurology*, pages 379–385, 2011.

[6] Barbara A.K. Kreilkamp, Bernd Webber, Mark P Richardson, and Simon S Keller. Automated tractography in patients with temporal lobe epilepsy using tracts constrained by underlying anatomy (tracula). *NeuroImage*, pages 67–76, 2017.

[7] Jacob Levman, Zihang Fang, Katarina Zumwalt, Liam Cogger, Lana Vasung, Patrick MacDonald, Ashley Lim, and Emi Takahashi. Asymmetric insular connectomics in a pediatric population: A clinical diffusion magnetic resonance imaging analysis of healthy development. *In review at Brain Connectivity*, 2018.

[8] Lauren J O'Donnell and Carl-Frederik Westin. Automatic tractography segmentation using a high-dimensional white matter atlas. *IEEE Transactions on Medical Imaging*, pages 1562–1575, 2007.

[9] Farquharson S, Tournier JD, Calamante F, Fabinyi G, Schneider-Kolsky M, Jackson GD, and Connelly A. White matter fiber tractography: why we need to move beyond dti. *Journal of Neurosurgery*, 2013.

[10] C. Salvatore, A. Cerasa, I. Castiglioni, F. Gallivanone, A. Augimeri, M. Lopez, G. Arabia, M. Morelli, M. C. Gilardi, and A. Quatrone. Machine learning on brain mri data for differential diagnosis of parkinson's disease and progressive supranuclear palsy. *Clinical Neuroscience*, pages 230–237, 2014.

[11] S. N. Sotiropoulos, S. Jbabdi, J. Xu, J. L. Andersson, S. Moeller, E. J. Auerbach, M. F. Glasser, M. Hernandez, G. Sapiro, M. Jenkinson, D. A. Feinberg, E. Yacoub, C. Lenglet, D. C. Van Essen, K. Ugurbil, and T.E.J. Behrens. Advances in diffusion mri acquisition and processing in the human connectome project. *NeuroImage*, pages 125–143, 2013.

[12] Olaf Sporns, Giulio Tononi, and Rolf Kötter. The human connectome: A structural description of the human brain. *PLoS Comput Biol*, 2005.

[13] Abdel Aziz Taha and Allan Hanbury. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*, pages 15–29, 2015.

[14] David C Van Essen, Stephen M. Smith, Deanna M. Barch, Timothy E.J. Behrens, Essa Yacoub, and Kamil Ugurbil. The wu-minn human connectome project: An overview. *NeuroImage*, pages 62–79, 2013.

[15] Demian Wasserman, Nikos Makris, Yogesh Rathi, Martha Shenton, Ron Kikinis, Marek Kubicki, and Carl-Frederik Westin. The white matter query language: a novel approach for describing human white matter anatomy. *Brain Structures and Function*, pages 4705–4721, 2016.

[16] Jakob Wasserthal, Peter Neher, and Klaus H. Maier-Hein. Tractseg - fast and accurate white matter tract segmentation. *NeuroImage*, pages 239–253, 2016.

[17] Dongyang Zhang, Abraham Z Snyder, Joshua S Shimony, Michael D Fox, and Marcus E Raichle. Noninvasive functional and structural connectivity mapping of the human thalamocortical system . *Cerebral Cortex*, pages 1187–1194, 2009.