MIGRAINE: MRI Graph Reliability Analysis and Inference for Connectomics

William Gray Roncal*†, Zachary H. Koterba*, Disa Mhembere†,
Dean M. Kleissas* Joshua T. Vogelstein‡†§ Randal Burns† Anita R. Bowles¶
Dimitrios K. Donavos¶ Sephira Ryman‖ Rex E. Jung‖ Lei Wu‖ ** Vince Calhoun‖ ** and R. Jacob Vogelstein*†
*JHU Applied Physics Laboratory, Laurel Maryland 20723, USA. Email: willgray@jhu.edu
†Johns Hopkins University, 3400 N Charles Street, Baltimore, Maryland 21218, USA
‡Duke University, Durham, NC 27708, USA
§Child Mind Institute, 445 Park Avenue, New York, NY 10022, USA
¶University of Maryland, Center for Advanced Study of Language, 7005 52nd Avenue, College Park, MD 20742, USA
‖University of New Mexico, 1 University Blvd NE, Albuquerque, NM 87131, USA
**The Mind Research Network, 1101 Yale Blvd NE, Albuquerque, NM 87106, USA

Abstract—Currently, connectomes (e.g., functional or structural brain graphs) can be estimated in humans at $\approx 1\ mm^3$ scale using a combination of diffusion weighted magnetic resonance imaging, functional magnetic resonance imaging and structural magnetic resonance imaging scans. This manuscript summarizes a novel, scalable implementation of open-source algorithms to rapidly estimate magnetic resonance connectomes, using both anatomical regions of interest (ROIs) and voxel-size vertices. To assess the reliability of our pipeline, we develop a novel non-parametric non-Euclidean reliability metric. Here we provide an overview of the methods used, demonstrate our implementation, and discuss available user extensions. We conclude with results showing the efficacy and reliability of the pipeline over previous state-of-the-art.

Index Terms—connectomics, magnetic resonance imaging, network theory, pipeline

I. INTRODUCTION

The ability to estimate a connectome, i.e. a description of connectivity in the brain of an individual, promises advances in many areas from personalized medicine to learning and education, and even to intelligence analysis [1, 2]. A robust analysis of these brain-graphs is on the horizon due to recent efforts to collect large amounts of multimodal magnetic resonance (M^3R) imaging data [3, 4]. An ideal methodology would enable scalable computing of graphs and functionals thereof in a way that yields estimates that are reliable. Moreover, such a tool would be open source, and make the data it processes open source in a user friendly way.

Building such a tool, however, is challenging. The data for each subject consists of about 1 gigabyte (GB) of M^3R data. Converting from raw data to graphs and functions thereof requires daisy-chaining over 20 subroutines, each of which implements a different transformation of the data. MRCAP is an existing pipeline that we previously developed for inference of graphs [5]. However, MRCAP has robustness and scalability limitations; it requires about 10 hours per subject to generate a final output, and has scheduler constraints. Moreover, MRCAP only generates small graphs, with 70 vertices, rather than voxel-wise graphs and functions thereof. Other pipelines (for example, [6]) have similar problems.

The utility of the output brain-graphs is a function of its scientific meaningfulness. Because such data currently lack ground truth for the estimates, or even other gold standards, investigators are left to assess the quality of estimates using only the data itself. Test-retest datasets consist of multiple subjects, each of whom have been scanned multiple times. Most previous work on reliability has assessed parametric reliability of features of the data. For example, it is standard to compare the between and within variances of, say, the number of edges [7]. However, these approaches are limited because they make parametric assumptions, and test scalar functions of the graphs.

We present here an MRI Graph Reliability Analysis and INference for ConnEctomics (MIGRAINE). methodology and associated software package. In addition to satisfying the above two mentioned desiderata (scalability and reliability), our pipeline, and much of our data, are provided in accordance with open science. The tested data originate from a wide assortment of institutions and projects, demonstrating the robustness of MIGRAINE. We demonstrate the improved reliability of our pipeline over the previous state-of-the-art, in addition to its improved scalability. Moreover, we utilize a notion of reliability related to the mean reciprocal rank often used in the information retrieval community. Finally, a useful metric for assessing pipeline and graph reliability is constructed and demonstrated.

II. MIGRAINE PIPELINE

A. Raw Data

Table I provides summary statistics for the various datasets processed via MIGRAINE. For each dataset, we collected both diffusion and structural MRI (MPRAGE). The two test-retest datasets (KKI–42 and NKI–24) were used to assess reliability; the other datasets were processed because they contain interesting phenotypic information (covariates) that we will utilize in future studies.
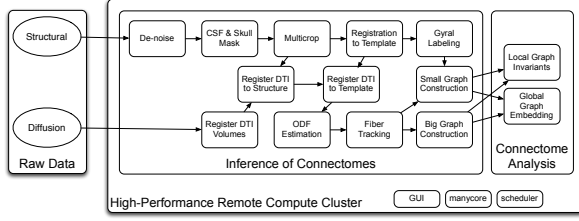
Fig. 1: MIGRAINE Pipeline Overview

TABLE I: Various datasets successfully processed via MI-GRAINE. Key for covariates: S=standard (sex, age, handedness), C=cognitive, B=behavioral, L=language, D=diagnostic (e.g., bipolar). The COINS column indicates whether the dataset will be made available via the Mind Research Network COINS data exchange (http://coins.mrn.org/dx). *=available at www.nitrc.org/projects/multimodal.

name	# subjects	covariates	ref	COINS
KKI–42	21	S	[8]	*
NKI–24	12	S	[9]	Y
MRN–111	111	C	N/A	Y
MRN–1313	1313	C,D	N/A	N
CASL–36	36	C,B,L	N/A	N

B. Inference of Brain-Graphs

1) LONI Processing Framework: The original Java Image Science Toolkit (JIST)-based [10] pipeline, consisting of 22 Java modules, is wrapped and integrated within the LONI pipeline framework [11]. Via swapping out some modules with improved functions, reducing I/O and communication, processing time is significantly reduced. MIGRAINE is flexible and can be modified using existing neuroimaging modules already incorporated in LONI (e.g., [12]), or custom code that is command-line executable.

2) Small Graph Generation: Our graph generation routines (summarized in Figure 1) consist largely of the following steps. Small graphs (e.g. 70 vertices and $\binom{70}{2} = 2415$ edges) are computed as detailed in [5], except for additional steps to register the structural and diffusion data for each subject to a common template (e.g., the MNI atlas [13]). The voxels in the template brain volume are labeled in accordance with the Desikan atlas regions [14]. We estimate tensors and perform deterministic tractography (FACT [15]) to estimate fibers in the brain (note that both of these functions can easily be swapped out for more sophisticated, but time consuming, options). Finally, an estimate of connectivity between each pair of regions is recorded (e.g., the number of times each region pair is connected by a fiber). Much of the connectome literature uses *region*-wise graphs, rather than *voxel*-wise [16]. These results therefore enable comparison with previous analysis, allow for the assessment of reliability between pipelines, and support existing classification methods.

3) Big Graph Generation: To generate big graphs, we utilize the Magnetic Resonance One-Click Pipeline (MROCP) code base as detailed in [17]. Initially a mask (e.g ROIs) is applied to the fiber streamlines created during small graph

estimation. Each surviving voxel becomes a vertex in the sparse column compressed graph. Here edges represent a single fiber connecting a pair of vertices within the bounds defined by the mask. Finally, we iterate over each fiber streamline, recording an edge between every two vertices that can be reached (i.e. that are connected) by a single fiber. The entire codebase to generate the big graphs was developed and written in Python.

All of these graphs are aligned, and for each MR scan we obtain a big graph with $\approx 10^7$ *aligned* vertices and $\approx 10^{10}$ edges. Because we are conservative with the mask, many of these voxels are noise. We therefore reduce these graphs to their largest connected component, which keeps essentially all white matter voxels, consisting of $\approx 10^5$ vertices and $\approx 10^8$ edges.

C. Analysis of Brain-Graphs

Computing analytics (i.e., multivariate glocal invariants [17]) on big graphs is a challenging endeavor due to the computational intensity associated with processing graphs with $\approx 10^8$ edges. Equivalent computational tasks are thus generally designated to specialized hardware like GPUs, graph processing engines like GraphLab [18], or distributed solutions like MapReduce. We utilize MROCP to efficiently compute several multivariate graph analytics, including Latent Position (LP-k), Number of Local 3-Cliques (NL-3), Clustering Coefficient (CC), Scan Statistic-1 (SS-1), Degree and Edge count (see [17] for details).

D. Graph Reliability

The literature on reliability focuses primarily on parametric reliability of scalar functions of the graphs. In other words, they (implicitly) assume that the graphs themselves are reliable, and then ask questions about particular features of the graphs. Because the data collection and graph generation processes are so noisy, we desire to assess the reliability of their composition. Thus, given a fixed pipeline, we can compare reliability of two different scanners or scanning sequences. Alternatively, given a fixed scanner and sequence, we can compare the reliability of two different pipelines.

Let $\xi\colon \Omega \times \mathcal{T} \to \Xi$ be a brain-valued random variable. In words, $\omega \in \Omega$ denotes a particular person, and $t \in \Omega$ denotes a particular time, and $\xi_t(\omega) \equiv \xi(\omega, t)$ denotes person ω's *actual* brain at time t. Let $\psi\colon \Xi \to \mathcal{X}$ be a particular M³R scanner and scanner sequence, so $x_t(\omega) = \psi(\xi_t(\omega)) \in \mathcal{X}$ is the output from the scanner, and the input to MIGRAINE. MIGRAINE converts x to graphs, $\phi\colon \mathcal{X} \to \mathcal{G}$. Therefore, the scanning and pipeline together form the composition $f = \phi \circ \psi\colon \Omega \times \mathcal{T} \to \mathcal{G}$, and we can test the reliability of either ϕ or ψ over time for a particular subject, as follows.

Let $\delta\colon \mathcal{G} \times \mathcal{G} \to \mathbb{R}$ be a graph-value metric. For vertex aligned graphs, we adopt the Frobenius norm of their adjacency matrices, $\delta(G, G') = \|A - A'\|_F$, due to its simplicity and its theoretical properties (in particular, under that metric, a k-nearest neighbor classifier is universally consistent for

graphs [19]). Thus, given n subjects, each with observations at two different times, we obtain $2n$ x's as input to MI-GRAINE, and we compute $\binom{2n}{2}$ distances (because distances are symmetric). For each scan $i \in [2n] = \{1, \ldots, 2n\}$, we rank all remaining $2n - 1$ scans using δ. Let i and i' denote the first and second scan of subject i, respectively, and let $r_i \in [2n - 1]$ denote the rank of i' relative to i. We define reliability of f as $R(f) = (2n - \sum_{i \in [2n]} r_i)/(2n-1) \in (0, 1)$, so $R = 1$ is maximally reliable and $R = 0$ is minimally reliable. Note that this notion of reliability is not limited to assessing f's or graphs – it is broadly applicable. Moreover, it is nonparametric and robust to outliers, and makes no distributional assumptions.

III. RESULTS

We successfully processed subjects from a variety of data corpora (both existing and new), totaling over 1500 subjects from multiple centers and acquisition paradigms. All of the resulting graphs and analytics are currently being used to develop classifiers, provide new insight into the way brains are wired and to determine which aspects of the network are informative in predicting cognitive properties.

A. Scalability

The current iteration of our software in the LONI Pipeline results in significant improvements to both scalability and processing time relative to the MRCAP baseline [5], which produces a small graph in approximately 10 hours on our small cluster (248 concurrent nodes, 1 TB total RAM). On average, the MIGRAINE baseline takes approximately 3 hours/subject to compute small graphs (i.e., the output from MRCAP), an additional 5 hours/subject to produce big graphs, and 3.5 hours/subject for graph invariants, for a total of 11.5 hours/subject. Much of this improvement is obtained by utilizing a common registration template, allowing for anatomical labels to be computed only once and then reused. Multi-node capabilities only contribute marginally for a single subject (in both pipelines) because the most intensive computations occur serially. However, there are significant efficiencies in scheduling when evaluating a large number of subjects, and the number of nodes is the limiting factor. Run time for each of our datasets is presented in Table II, including the results of a dataset with over 1000 subjects. A univariate measure of total fiber count per subject is shown in Figure 2.

TABLE II: Total and average run times for each dataset.

Dataset	# scans	Time (Hours)				
		Small Graphs	Big Graphs	Analytics	Total	Average/ Subject
KKI-42	**42**	2.9	4.7	3.9	11.5	11.2
CASL-36	**36**	3.1	4.6	3.7	11.3	11.0
NKI-TRT	**24**	3.2	5.2	3.4	11.8	11.6
MRN-111	**111**	1.5	6.1	3.0	10.5	9.8
MRN-1313	**1313**	9.9	37.3	18.3	65.5	10.2
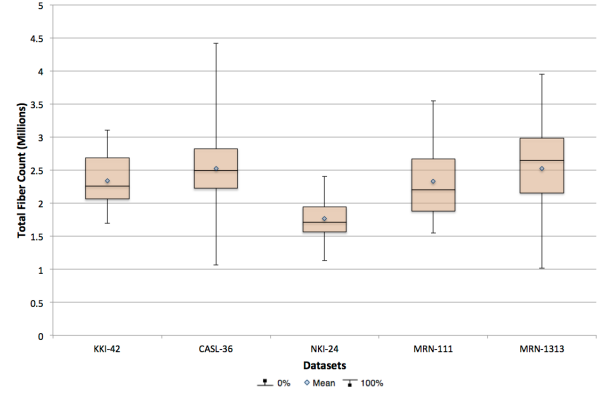
Fig. 2: Box plots for each data set, showing total fiber count for each dataset.

B. Reliability

A variety of tools have been developed to analyze intermediate pipeline products, such as matrix comparison tools and analysis of fiber counts. MIGRAINE leverages several algorithmic improvements versus MRCAP, including changes to data preprocessing and the registration to a common registration space as described previously. The results differed by 13% from the MRCAP baseline and produced better subject separability as shown in Table III.

TABLE III: Validation showing improved discrimination relative to MRCAP using the KKI–42 dataset [8].

Pipeline	Intra-Sub Diff	Inter-Sub Mean Diff	Closest Inter-Sub	# Matches
MRCAP	26032	51584	38451	40/42
MIGRAINE	20378	56126	42663	42/42

To validate that our graphs produce a repeatable signal, we used the KKI–42 Test-Retest Data [8] to analyze graph estimation reliability, similar to [20]. We demonstrated that the MIGRAINE pipeline produced a stable connectivity measurement across multiple scans of the same subject.

For all 42 graphs, the most closely related graph (as computed with the Frobenius norm) belonged to the same person, scanned at a different time. A visualization of the graphs for six test-retest pairs are shown in Figure 3, and the results of all individual subject comparisons are shown in Figure 4.

IV. CONCLUSIONS

MIGRAINE is robust and has been utilized to process over 1500 subjects from a variety of datasets in a rapid, extensible, automated framework. In addition to producing small graphs, we have demonstrated additional processing capability through the estimation of big graphs and analytics. The pipeline is scalable and has internal validation and packaging scripts to enable efficient analysis. Finally, we provide evidence of classification signal in the estimated brain graphs.
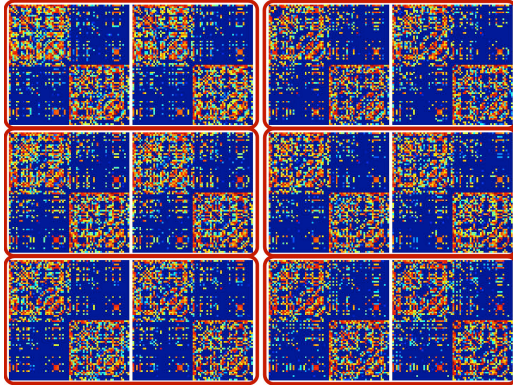
Fig. 3: Six Test-retest graphs. Top (L-R): Male, 25 years old (M25), F26, Middle: M25, F30, Bottom: M38, F61.
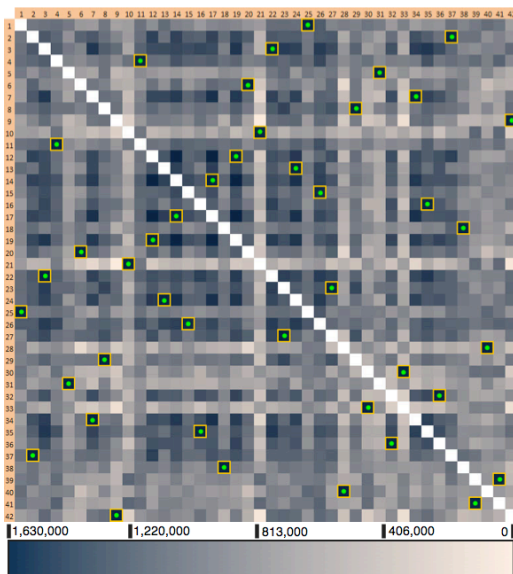
Fig. 4: KKI Test-Retest Results. Yellow boxes: Highest similarity, Green dots: True pairs, White: Self-comparison.

REFERENCES

[1] O. Sporns, "Networks of the Brain," *Learning*, no. August, p. 375, 2010.

[2] J. W. Lichtman and J. R. Sanes, "Ome sweet ome: what can the genome tell us about the connectome?" *Current opinion in neurobiology*, vol. 18, no. 3, Jun. 2008.

[3] D. C. Van Essen *et al.*, "The Human Connectome Project: a data acquisition perspective." *NeuroImage*, vol. 62, no. 4, pp. 2222–31, Oct. 2012.

[4] M. Mennes, B. B. Biswal, F. X. Castellanos, and M. P. Milham, "Making data sharing work: The FCP/INDI experience." *NeuroImage*, Oct. 2012.

[5] W. R. Gray *et al.*, "Magnetic Resonance Connectome Automated Pipeline," *Pulse*, no. APRIL, pp. 1–5, 2012.

[6] Z. Cui *et al.*, "PANDA: a pipeline toolbox for analyzing brain diffusion images." *Frontiers in human neuroscience*, vol. 7, no. February, p. 42, Jan. 2013.

[7] U. Braun *et al.*, "Test-retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures." *Neuroimage*, vol. 59, no. 2, pp. 1404–1412, 2012.

[8] B. A. Landman *et al.*, "Multi-parametric neuroimaging reproducibility: a 3-T resource study." *NeuroImage*, vol. 54, no. 4, pp. 2854–66, Mar. 2011.

[9] K. B. Nooner *et al.*, "The NKI-Rockland Sample: A Model for Accelerating the Pace of Discovery Science in Psychiatry," *Frontiers in Neuroscience*, vol. 6, no. 152, 2012.

[10] B. C. Lucas *et al.*, "The Java Image Science Toolkit (JIST) for rapid prototyping and publishing of neuroimaging software." *Neuroinformatics*, vol. 8, no. 1, pp. 5–17, Mar. 2010.

[11] I. D. Dinov *et al.*, "Efficient, Distributed and Interactive Neuroimaging Data Analysis Using the LONI Pipeline." *Frontiers in neuroinformatics*, vol. 3, p. 22, Jan. 2009.

[12] M. Jenkinson *et al.*, "FSL," *NeuroImage*, vol. 62, no. 2, pp. 782–790, 2012.

[13] G. Grabner *et al.*, "Symmetric Atlasing and Model Based Segmentation: An Application to the Hippocampus in Older Adults," in *Lecture Notes in Computer Science*, 2006, vol. 4191, pp. 58–66.

[14] R. S. Desikan *et al.*, "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest." *NeuroImage*, vol. 31, no. 3, pp. 968–80, Jul. 2006.

[15] S. Mori, B. Crain, V. Chacko, and P. Van Zijl, "Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging," *Annals of neurology*, vol. 45, no. 2, pp. 265–269, Feb. 1999.

[16] R. C. Craddock *et al.*, "Imaging human connectomes at the macroscale." *Nature methods*, vol. 10, no. 6, pp. 524–39, Jun. 2013.

[17] D. Mhembere *et al.*, "Computing Scalable Multivariate Glocal Invariants of Large (Brain-) Graphs," *IEEE GlobalSIP, Accepted*, 2013.

[18] J. E. Gonzalez, D. Bickson, and C. Guestrin, "PowerGraph : Distributed Graph-Parallel Computation on Natural Graphs," pp. 17–30, 2012.

[19] J. T. Vogelstein, R. J. Vogelstein, and C. E. Priebe, "Are mental properties supervenient on brain properties?" *Nature Scientific Reports*, p. 11, 2011.

[20] W. Gray *et al.*, "Magnetic Resonance Connectome Automated Pipeline and Repeatability Analysis," *Society for Neuroscience Abstract*, 2011.