

A High-Throughput Pipeline Identifies Robust Connectomes But Troublesome Variability

<https://www.biorxiv.org/content/10.1101/188706v6>

Abstract

Modern scientific discovery depends on large heterogeneous datasets. Different people have different usage. For the research of macroscale connectomics studies, it's required the complex pipelines. For individual studies, it's required customize pipelines to their needs. So in general, the purpose of using connectome is differ across studies. In order to solve this problem, we developed NeuroData's MRI to Graphs (NDMG) pipeline using several functional and discussion studies, including the Consortium for Reliability and Reproducibility, to estimate connectomes. Without any manual intervention or parameter tuning, NDMG ran on 25 different studies ($\approx 6,000$ scans) from 15 sites, with each scan resulting in a biologically plausible connectome (as assessed by multiple quality assurance metrics at each processing stage).

Introduction

To rigorously identify and quantify the sources of variability both within and across multi-modal neuroimaging requires (1) data and (2) a pipeline. The requisite data includes numerous datasets with multiple measurements per individual including data that both conserve and vary a number of different factors. The requisite pipeline must be able to fully process each sample and study, and analyze the results both within and across studies using a coherent statistical model.

The Consortium of Reliability and Reproducibility (CoRR) consists of about 30 different studies from nearly 20 different institutions around the world that provide the necessary data.

We developed an approach, "Neuro Data MRI to Graphs" (NDMG), that meets or exceeds standards along each of the above mentioned principles.

We validated our pipeline by running NDMG on 11 dMRI studies comprising 3,227 individuals with 4,347 scans, and 18 fMRI studies comprising 714 individuals with 1,646 scans. For each scan NDMG estimates a "connectome" (a functional or structural map of connectivity) at 24 different spatial resolutions—yielding a total of $> 100,000$ estimated

connectomes—all of which are publicly available from <http://m2g.io>. This is the largest open database of connectomes, and one of the largest mega-analyses (inference across studies) of multi-modal connectomics data to date.

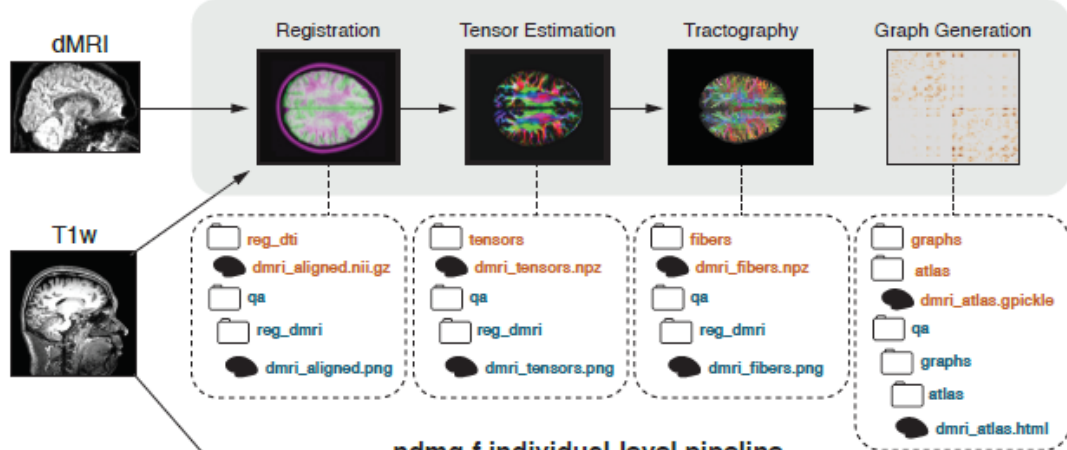
Table 1: Comparing M3R Processing Pipelines. NDMG is designed with both algorithmic and implementation principles in mind. This table compares existing pipelines along these principles, demonstrating that for each, NDMG performs at least as well as the current state of the art. A ✓ is given for pipelines that satisfy the respective desiderata, a ✓ for pipelines that partially satisfy the respective desiderata, and a ✗ is given for pipelines that do not satisfy the respective desiderata. [Appendix A](#) provides more details.

pipeline	Statistical Principles				Computational Principles				Connectome Principles		
	accurate	reliable	robust	expedient	scalable	portable	turn-key	open	dMRI	fMRI	raw-to-graph
NDMG	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
HCP [29]	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	✗
PANDA [6]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓
CMTK [7]	✓	✓	✓	✓	✓	✗	✗	✓	✓	✗	✓
CPAC [5]	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗
fmrprep [30]	✓	✓	✓	✗	✓	✓	✓	✓	✗	✓	✗
NIAX [31]	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗

Results

1. **Statistical Principles:** The statistical principles are designed to evaluate the empirical quality of the method on real data. Better quality suggests the pipeline can be trusted to be reliable for subsequent investigations.
2. **Robustness** quantifies accuracy and reliability across a wide range of studies with different properties, including different experimental design, measurement devices, etc. We, therefore, ran NDMG on 11 dMRI studies and 18 fMRI studies using different hardware and acquisition parameters (see Table 2 for details). In some of these studies samples were not filtered to discard outliers or samples with poor signal-to-noise properties. Nonetheless, for each study, NDMG’s QA suggested accuracy, and each study achieved a score of discriminability > 0.8.
3. **Computational Principles:** Adhering to the computational principles lowers the barrier for use. In practice, this means that both domain experts and researchers from other disciplines (such as machine learning and statistics), can more easily use the tools.

ndmg-d individual-level pipeline



ndmg-f individual-level pipeline

