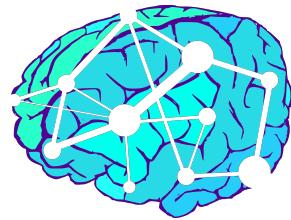


ndmg: a reliable one-click pipeline for M3R connectome estimation



Gregory Kiar^{1,2†}, William R. Gray Roncal^{3†}, Vikram Chandrashekhar²,
Eric W. Bridgeford¹, Disa Mhembere⁴, Randal Burns⁴, Joshua T. Vogelstein^{1,2}

Abstract

We have produced a 1-click pipeline for reliable connectome estimation at scale and used it to process and release 1,000s of graphs across a range of scales to the public. Subsequently, we computed a mean and variance connectome from over 2200 individuals and 2800 sessions.

1 Introduction

just so I can see where non listed text will be

- O: mental health is a big problem. connectopathies are thought to be a subset of neurological disorders which derive from problems in connectivity within the brain
- S: estimating connectomes allows for the investigation of connectopathies and development of the understanding of their underlying causes, eventually leading to mechanisms for treatment
- C: many connectome estimation tools exist, either as partial- or complete-solutions. a current issue, however, is that there is no "standard" processing method, and not all methods produce equally reliable results; claims made from data processed one way may not be consistent with the claims made from the same data when processed differently.
- G: a consistent and reliable one-click connectome estimation pipeline which abstracts parameter selection and overfitting to their data has not been developed in the community, resulting in incongruent results across the community, and causes connectomics to suffer from a lack of reproducibility and repeatability.
- A: we have developed a turn-key solution for structural connectome estimation at scale and packaged it to ship across any (Windows, Linux, OSx) modern platform.

- R: Our open-source scalable pipeline ndmg has been used to process +2,500 brain scans and generate +50,000 connectomes across a range of scales. We use these connectomes to illustrate the presence of batch effects in MRI, and have computed the largest known mean-and variance-connectomes to date.

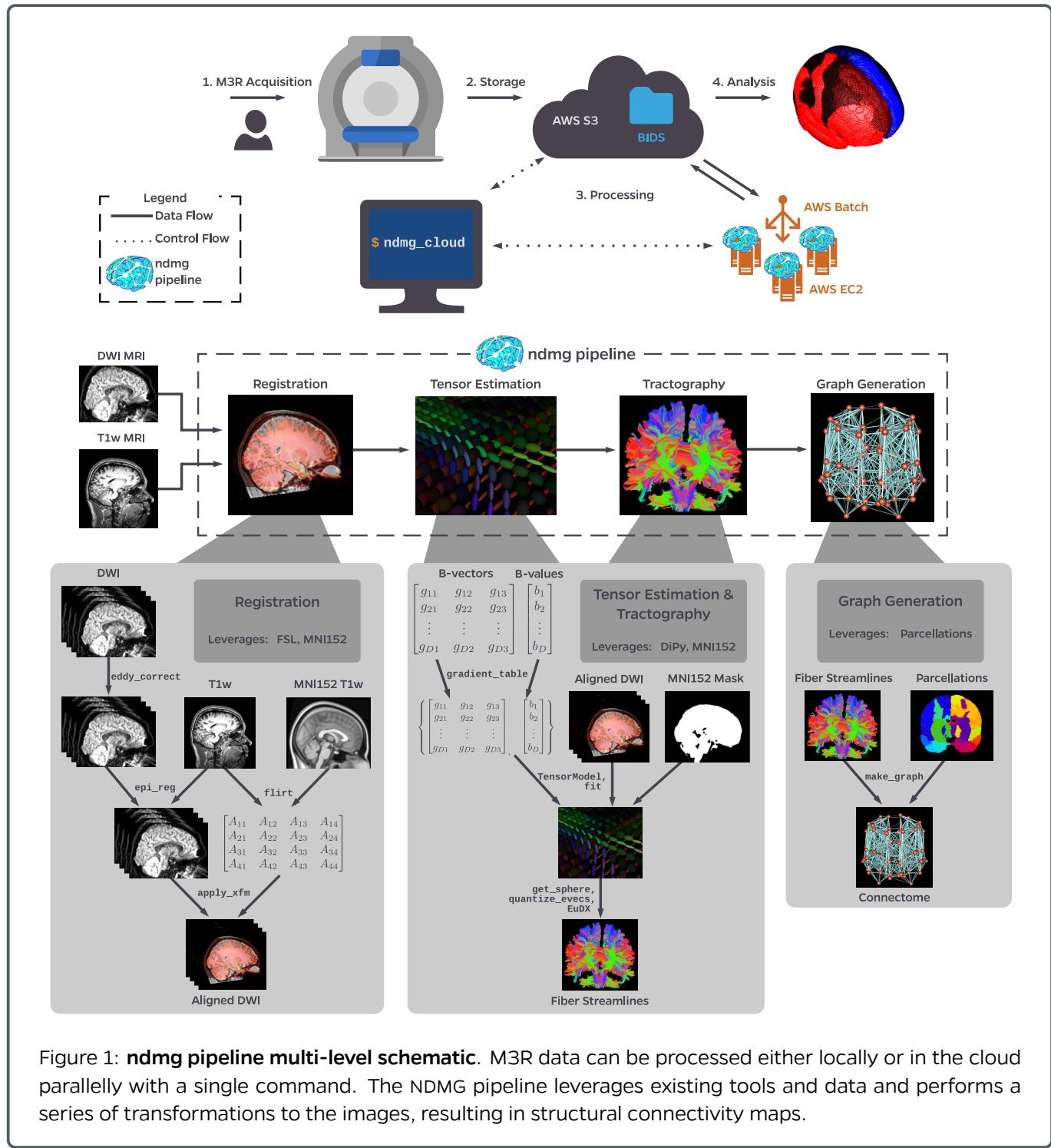
The multi-level schematic of NDMG usage and operation can be seen in Figure 1. In the system-level diagram, we illustrate that NDMG is designed such that users are able to input minimally-processed DWI and T1w MR images (furthermore referred to as multimodal MR, or M3R), and receive a robust estimate of connectivity in return. NDMG is integrated with the BIDS [1] organization standard, and the Amazon Web Services (AWS) computing cloud, allowing it to serve as a turn-key solution for connectome estimation regardless of available local computational resources.

2 Methods

The NDMG pipeline has been developed by leveraging and interfacing existing tools, including FSL [2-4], DiPy [5], the MNI152 atlas [6], and a variety of parcellations defined in the MNI152 space [7-13]. All algorithms which required hyper-parameter selection were initially set to the suggested parameters for each tool, and tuned to improve the quality and robustness of the results. Turning again to Figure 1, the lower-level schematic provides a glimpse at the inner-workings of NDMG, breaking the pipeline into four key components: Registration, Tensor Estimation, Tractography, and Graph Generation. The pro-

cess for each of these steps is described below, with an in depth look at the specific models and algorithms implemented at each stage contained within Appendix A.

Registration When generating connectomes, it is important that the derivatives from one scan can be compared to that of another. In order for the connectomes to align, they must be defined in a com-



mon space. To accomplish this, we transform the input M3R into the template MNI152 [6] space. Transforming the input data into the atlas space as opposed to the reverse enables connectomes that are acquired in different spaces or resolutions to be compared to one another reliably. Registration is done through FSL [2–4], and more detail can be found in Appendix A.1.

Tensor Estimation & Tractography Once the M3R data is aligned, we can use the gradient vectors and b-values accompanying the DWI scan to generate a voxelwise tensor image from the DWI image stack. Using DiPy [5] a simple 6-component tensor model is used. Using EuDX [14], deterministic tractography algorithm inspired by FACT [15], streamlines are then traced from the tensors. For more information on tensor estimation and tractography, see Appendix A.2 and Appendix A.3, respectively.

Graph Generation Connectomes are created by mapping fibers through pre-made parcellations of the brain. These parcellations can be defined by neuroanatomists or generated by segmentation algorithms. Many of the parcellations, such as HarvardOxford cortical mask [10], JHU [9], and Talairach [11] were obtained from FSL, while others were found from publications, such as Desikan [7], AAL [8], Slab907 [12], and Slab1068 [13], while the remaining were generated from segmentation algorithms, including CPAC200 [16] and 16 downsampled (DS) parcellations [17] ranging from 70 to 72,783 nodes. Tracing streamlines with each parcellation, an edge is added to the corresponding graph for every pair of nodes along a fiber. The edges are undirected, and their weight is the total number of edges added between two nodes.

2.1 Quality Assurance

As the final derivatives produced, connectomes, are far removed from the original imaging data, it is important to have quality assurance (QA) figures, enabling the user to easily detect whether or not the pipeline is producing expected results. QA for the intermediate derivatives in NDMG are shown in Figure 2. The QA plots were designed to highlight likely

problems that may occur within the pipeline. For registration, NDMG produces an overlay of the inputted DWI image and the MNI152 T1w template, which allows the user to verify that the brain boundary and higher-level structures of the input image have been correctly aligned to the atlas. For tensor estimation, NDMG produces a fractional anisotropy (FA) map of the tensors, which allow the user to verify that the gradient table and the diffusion image are properly co-aligned. For tractography, NDMG visualizes a subset of the generated streamlines within a mask of the MNI152 brain, so that the user can verify that no fibers leave the brain and that their structure resembles that of the FA map generated in the previous step. Together, these figures provide a thorough inspection of the derivatives as they are being produced, so that if an error were to occur it would be obvious to the user.

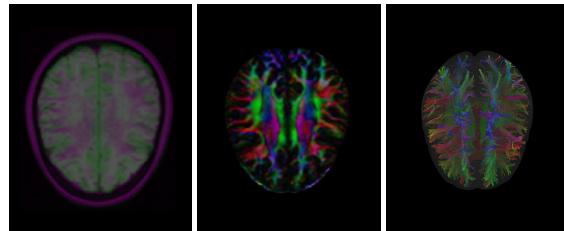


Figure 2: **Intermediate quality assessment outputs from `ndmg`.** NDMG generates registration, tensor, and fiber quality assessment images after each is produced during pipeline operation.

Once connectomes have been produced for an entire dataset, the user may then use NDMG to generate a plot of graph summary statistics to develop further understanding of the structure of their newly obtained brain graphs. Figure 3 shows the summary statistics plot for an exemplar dataset constructed with the Desikan [7] parcellation. The features of the graphs which NDMG computes are [17], clockwise from top-left: betweenness centrality, clustering coefficient, hemisphere-separated degree sequence, edge weight, eigen values, locality statistic-1, number of non-zero edges, and the cohort mean connectome. These statistics enable detailed quality assessment of the graphs; for instance, enabling the user

to confirm that edge density is higher ipsi-laterally than contra-laterally, and serve as a preliminary exploratory analysis of the processed data.

2.2 Validation

Results from NDMG have been validated using a metric termed Discriminability [18]. Discriminability, or Mean Normalized Rank (MNR), as seen in Equation (1) is a statistically generalized version of Test-ReTest (TRT) reliability:

$$\text{MNR} = p(||a_{ij} - a_{ij'}|| \leq ||a_{ij} - a_{i'j'}||) \quad (1)$$

Discriminability describes the probability in given a dataset consisting of multiple observations of objects within multiple classes (i.e. scans as observations of each subject being a class), that the nearest-neighbour for an arbitrary observation, a_{ij} , is another observation of the same class, $a_{ij'}$, and not an observation of another class, $a_{i'j'}$. A perfect discrim-

inability score is 1 (i.e. always selecting the same subject from one of their scans), and the worst score is 0. Optimizing NDMG with respect to discriminability enables us to minimize the upper-bound on error for any general downstream inference task, without overfitting to covariate-specific signal (i.e. optimizing the pipeline for sex classification).

3 Results

3.1 One-Click Connectomics

NDMG is a reliable, accessible, turn-key solution to human connectome estimation at scale. NDMG has been developed with a low barrier to entry in terms of input data quality, computational resources, and computational expertise. NDMG is installable for Python via the PyPi package repository¹, or a container image made available through DockerHub². NDMG has been integrated with a variety of platforms and standards including the BIDS data organization standard [1; 19] and the descriptive command-

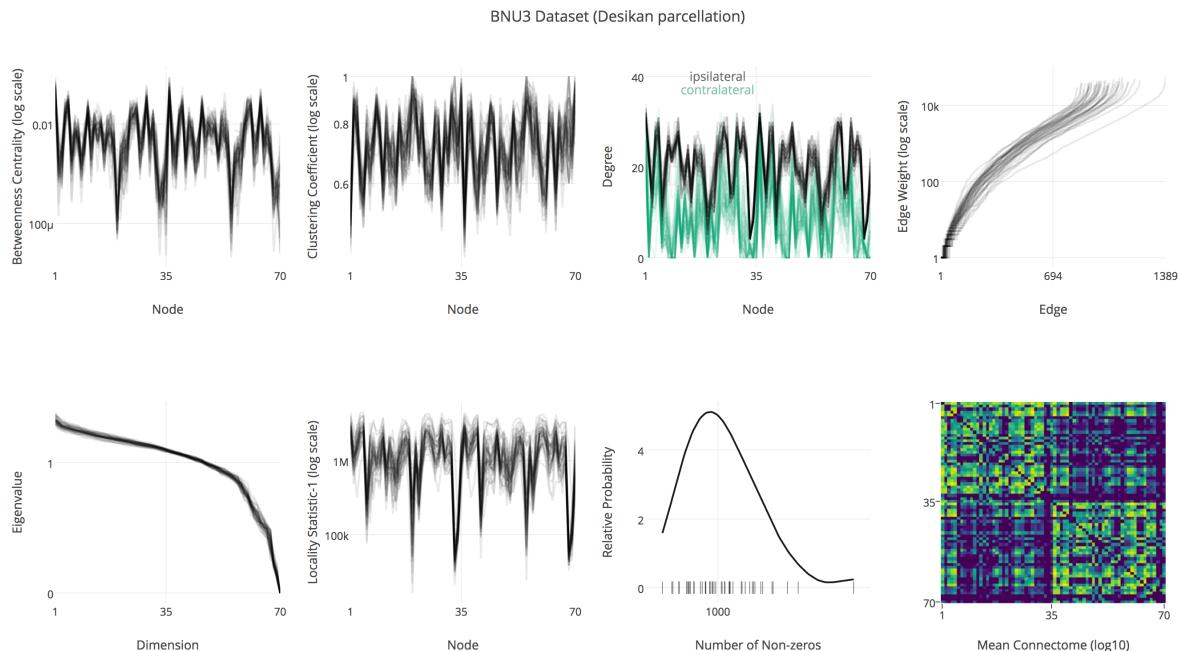


Figure 3: **Graph summary statistics.** Once a dataset is processed through NDMG, group analysis plots summary statistics of the graphs, serving as both quality assurance and a spring-board for exploratory analysis

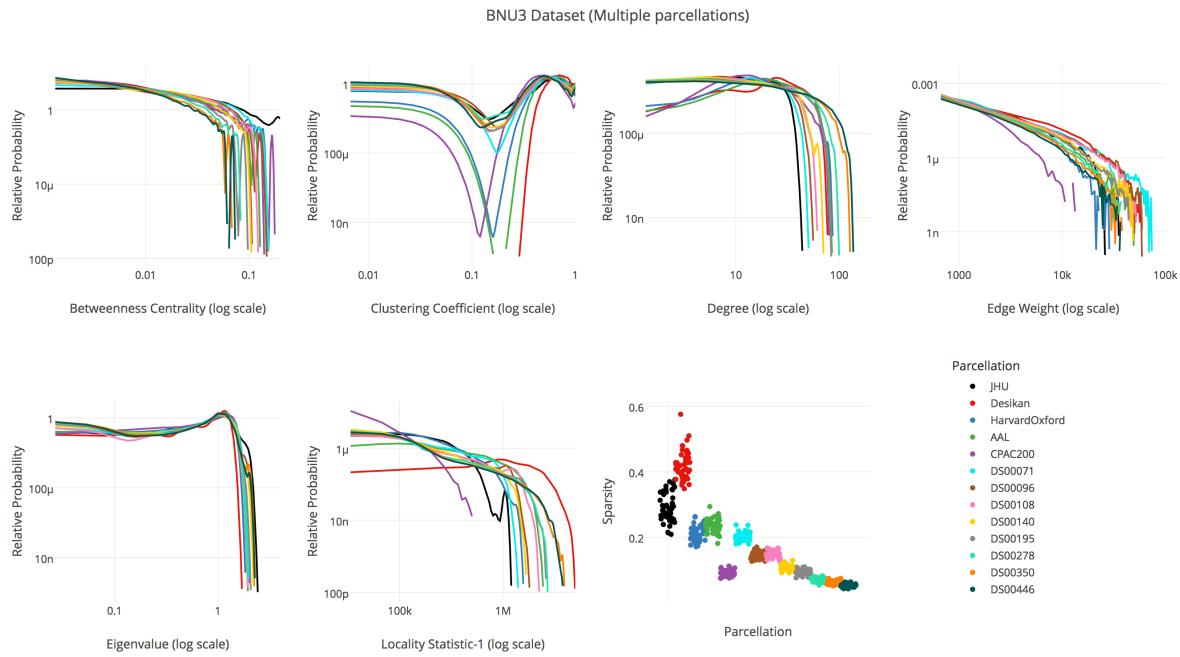


Figure 4: **Multi-scale graph analysis.** NDMG produces connectomes at a variety of scales, enabling investigation of graph properties between any desired parcellation schemes.

line schema Boutiques³. NDMG has been deployed in several large computing infrastructures including OpenfMRI [20] and CBRAIN [21], and has been integrated with the AWS computing cloud, enabling users to launch the pipeline in the cloud with a single command. The NDMG package and pipeline is open-source and available on Github⁴.

3.2 Processed Data

The NDMG pipeline has been run on a variety of public datasets, upon which it proves to be highly discriminable. Table 1 summarizes the 2,861 scans that have been processed with NDMG; each scan processed has been parcellated using 24 different atlases, resulting in 68,664 total graphs. All of the derived graphs, intermediate derivatives, and quality control figures are publicly available on our S3 bucket, `mrnurodata`. The data can also be accessed through our website, <http://m2g.io>.

We can see from Table 1 that NDMG has near-perfect discriminability for many of the TRT datasets evaluated. No outlier removal was done prior

to computing discriminability, and eliminating errorful-graphs such as those which could be detected by quality assessment plots (like those in Figure 2 and Figure 3) would likely increase the score further.

3.3 Multi-scale Analysis

As NDMG produces connectomes at a variety of scales, it enables either multi-scale analysis or flexibility in terms of the parcellation a researcher wishes to use for a given task. Figure 4 shows various statistics of graphs generated from the same dataset over a subset of the 24 parcellations used to estimate connectomes. The parcellations used in NDMG range in number of nodes from 48 to 72,000; only those with nodes under 500 nodes are shown in Figure 4. Comparing graphs generated with the same fiber across various parcellations relative to one another may be valuable and can inform the optimal scale upon which to perform specific analyses.

Table 1: **Processed public M3R datasets.** Each dataset has been processed with 24 parcellations, resulting in over 65,000 total publicly available connectomes.

Dataset	Subjs (k)	#	Total	MNR
BNU1 [22]	57	2	114	0.984
BNU3 [22]	48	1	47	-
HNU1 [22]	30	10	300	0.993
KKI2009 [23]	21	2	42	1.0
MRN1313	1313	1	1299	-
NKI1 [22]	24	2	40	0.984
NKI-ENH [24]	198	1	198	-
SWU4 [22]	235	2	454	0.884
Templeton114	114	1	114	-
Templeton255	255	1	253	-
Total	2295		2861	0.994

A perfect MNR for each TRT dataset is 1 and random-chance is $1/k$.

3.4 Batch Effects in M3RI

The datasets summarized in Table 1 were all processed using an identical version of NDMG, enabling “mega-analysis” between datasets to be conducted. Here, we investigate the prevalence of batch effects in M3R data. Using discriminability as described as in Equation (1), and substituting subject label with dataset as the observation’s class-label, we can compare the similarity between graphs derived within a dataset and across datasets. Shown in Figure 5, we sampled 20 sessions from each dataset, ensuring no subject in the TRT datasets was selected twice, and computed the discriminability between datasets. If there existed no dataset-specific signal, then we would expect a discriminability score of $1/10$, as there are 10 datasets. However, we compute a discriminability score of 0.7227, suggesting significant dataset-specific signal is present. This analysis was also conducted across all subjects within each dataset, and an interactive figure for both of these analyses is available on our website⁵.

3.5 Mean Connectome

We computed a mean connectome using the Desikan parcellation for each dataset processed in Table 1, as well as a mean and standard deviation (SD)

connectome across datasets, as shown in Figure 6. With $N=2861$, we believe this is the largest mean connectome computed to date. Properties of the brain such as high number of ipsi-lateral connections relative to a lower number of contra-lateral connections are present across all datasets. However, the SD connectome also shows that the ipsi-lateral connectivity is also more highly variable, suggesting that there exists a relationship between connection density and variability. We also notice that the ipsi-lateral connectivity within left (nodes 1-35) and right (nodes 36-70) hemispheres, respectively, are very similar in structure. An interactive version of this figure can be found on our website.

4 Discussion

- our pipeline is highly reliable, and has a low barrier to entry.
- a hyperparameter search optimizing individual parameters for each step of the pipeline has not been conducted
- our pipeline serves as a reliable pipeline for all dwi datasets, rather than an optimal pipeline for more sophisticated datasets such as HCP. Pipelines such as the HCP pipelines certainly outperform our pipeline for datasets containing all the additional information, but cannot be run on datasets without it. Our pipeline produces uniform coverage across datasets.
- our pipeline shows that highly sophisticated and computationally burdensome algorithms such as probabilistic tractography are not necessary to create highly reliable estimates of brain connectivity.
- deployment of the pipeline to the cloud, and integration with services such as openfmri and cbrain, makes our tool very accessible.
- As ndmg is reliable and highly modular, it can act as a reference pipeline for connectome estimation, in which new algorithms are evaluated and implemented when they out perform the standard.

Affiliation Information

Corresponding Author: Joshua T. Vogelstein <jovo@jhu.edu>

¹Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA.

²Center for Imaging Science, Johns Hopkins University, Baltimore, MD, USA.

³Johns Hopkins University Applied Physics Lab, Laurel, MD, USA.

⁴Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA.

[†]Authors contributed equally.

Availability of supporting source code and requirements

Project name M3RI to Graphs (m2g)

Project home page <http://m2g.io>

Operating system(s) Platform independent

Programming languages Python, Docker, Bash

Other requirements M3RI Data, Python or Docker

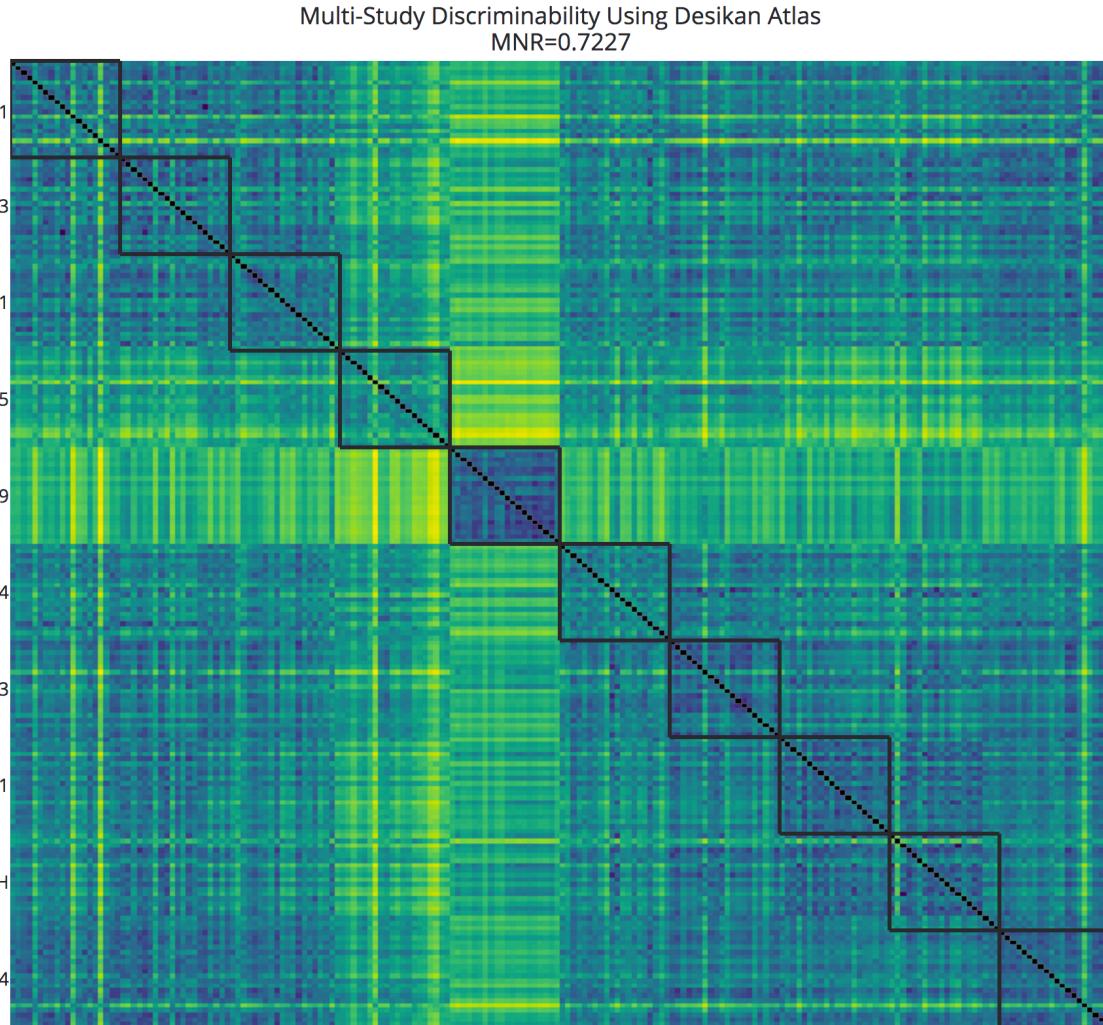


Figure 5: **Prevalence of batch effects.** Ten datasets have been processed with NDMG, and the discriminability across datasets was computed to quantify the prevalence of batch effects in M3RI data. If no batch effect were present, a discriminability of 0.1 would be expected. A discriminability score of 0.7227 as seen here suggests that there is considerable dataset-specific signal in the graphs.

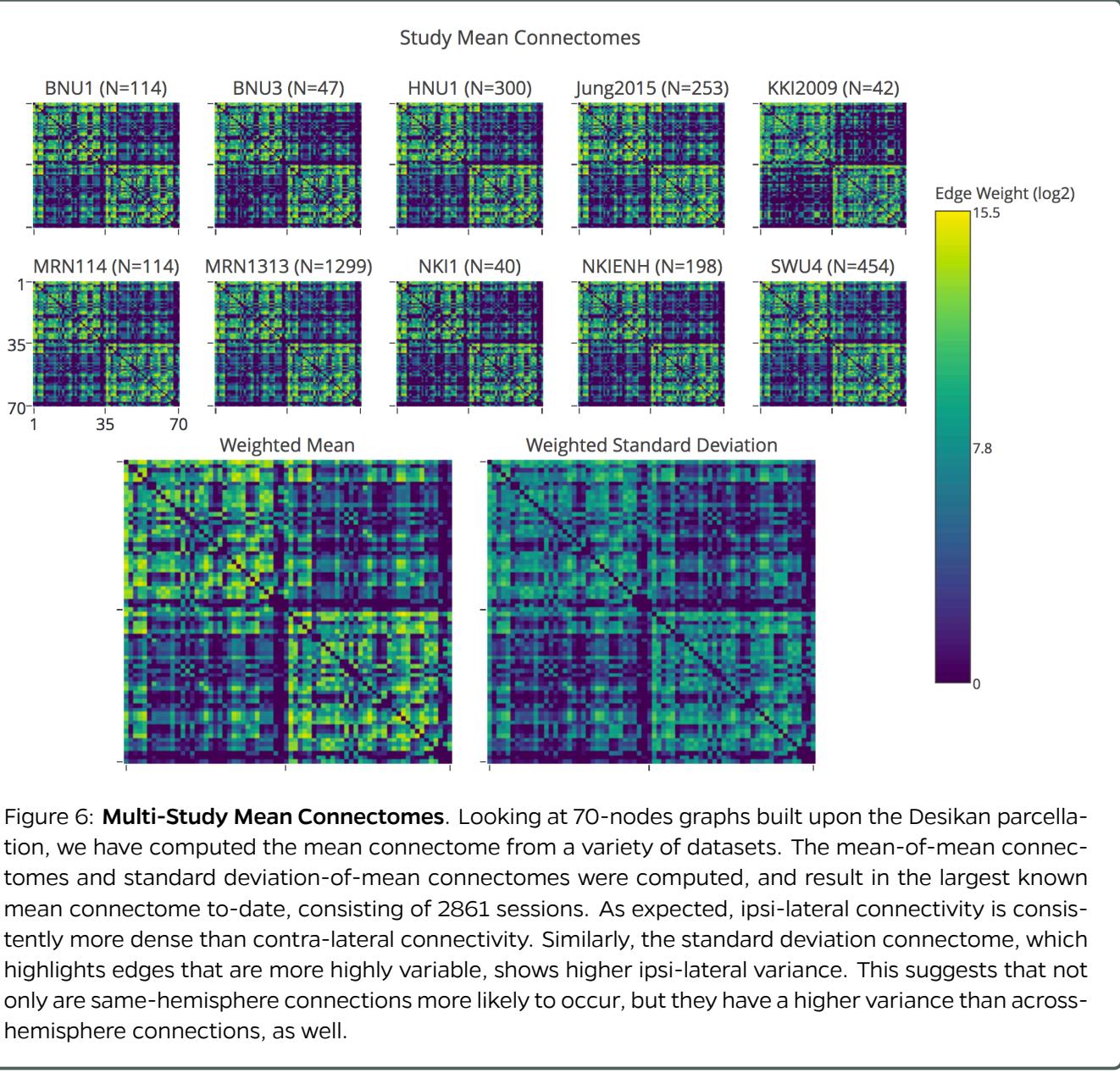


Figure 6: Multi-Study Mean Connectomes. Looking at 70-nodes graphs built upon the Desikan parcellation, we have computed the mean connectome from a variety of datasets. The mean-of-mean connectomes and standard deviation-of-mean connectomes were computed, and result in the largest known mean connectome to-date, consisting of 2861 sessions. As expected, ipsi-lateral connectivity is consistently more dense than contra-lateral connectivity. Similarly, the standard deviation connectome, which highlights edges that are more highly variable, shows higher ipsi-lateral variance. This suggests that not only are same-hemisphere connections more likely to occur, but they have a higher variance than across-hemisphere connections, as well.

License Apache 2.0

Declarations

Competing Interests The authors declare no competing interests in this manuscript.

Abbreviations Multimodal Magnetic Resonance Imaging (M3RI), NeuroData's MRI Graph Pipeline (NDMG), M3RI to Graphs (m2g), Mean Normalized Rank (MNR), Brain Imaging Data Structure (BIDS), Amazon Web Services (AWS).

Funding The authors would like to graciously thank: NIH, NSF, DARPA, IARPA, Johns Hopkins University, and the Kavli Foundation for their support. Specific award information can be found at <https://neurodata.io/about>.

References

- [1] K. Gorgolewski, T. Auer, V. Calhoun, C. Craddock, S. Das, E. Duff, G. Flandin, S. Ghosh, T. Glatard, Y. Halchenko et al., “The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments.”

- [2] S. M. Smith et al., "Advances in functional and structural MR image analysis and implementation as FSL." *NeuroImage*, vol. 23 Suppl 1, pp. S208–19, jan 2004. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/15501092>
- [3] M. W. Woolrich et al., "Bayesian analysis of neuroimaging data in FSL." *NeuroImage*, vol. 45, no. 1 Suppl, pp. S173–86, mar 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1053811908012044>
- [4] M. Jenkinson et al., "FSL." *NeuroImage*, vol. 62, no. 2, pp. 782–90, aug 2012. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21979382>
- [5] E. Garyfallidis, M. Brett, B. Amirbekian, A. Rokem, S. Van Der Walt, M. Descoteaux, and I. Nimmo-Smith, "Dipy, a library for the analysis of diffusion mri data," *Frontiers in neuroinformatics*, vol. 8, p. 8, 2014.
- [6] J. Mazziotta et al., "A four-dimensional probabilistic atlas of the human brain," *Journal of the American Medical Informatics Association*, vol. 8, no. 5, pp. 401–430, 2001.
- [7] 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*, 2006.
- [8] N. Tzourio-Mazoyer et al., "Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain," *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002.
- [9] K. Oishi et al., *MRI atlas of human white matter*. Academic Press, 2010.
- [10] 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. 968–980, 2006.
- [11] J. Lancaster, "The Talairach Daemon, a database server for Talairach atlas labels," *NeuroImage*, 1997.
- [12] C. S. Sripada et al., "Lag in maturation of the brain's intrinsic functional architecture in attention-deficit/hyperactivity disorder," *Proceedings of the National Academy of Sciences*, vol. 111, no. 39, pp. 14 259–14 264, 2014.
- [13] D. Kessler et al., "Modality-spanning deficits in attention-deficit/hyperactivity disorder in functional networks, gray matter, and white matter," *The Journal of Neuroscience*, vol. 34, no. 50, pp. 16 555–16 566, 2014.
- [14] E. Garyfallidis, M. Brett, M. M. Correia, G. B. Williams, and I. Nimmo-Smith, "Quickbundles, a method for tractography simplification," *Frontiers in neuroscience*, vol. 6, p. 175, 2012.
- [15] S. Mori et al., "Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging," *Annals of neurology*, vol. 45, no. 2, pp. 265–269, 1999.
- [16] S. Sikka et al., "Frontiers: Towards Automated Analysis of Connectomes: The Configurable Pipeline for the Analysis of Connectomes (C-PAC)," *Frontiers in neuroscience*, 2012. [Online]. Available: http://www.frontiersin.org/10.3389/conf.fninf.2014.08.00117/event_{_}abstract
- [17] D. Mhembere, W. G. Roncal, D. Sussman, C. E. Priebe, R. Jung, S. Ryman, R. J. Vogelstein, J. T. Vogelstein, and R. Burns, "Computing scalable multivariate glocal invariants of large (brain-) graphs," in *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*. IEEE, 2013, pp. 297–300.
- [18] S. Wang, Z. Yang, X.-N. Zuo, M. Milham, C. Craddock, G. Kiar, W. R. Gray Roncal, E. Bridgeford, CORR, C. E. Priebe, and J. T. Vogelstein, "Optimal decisions for discovery science via maximizing discriminability: Applications in neuroimaging," *Tech. Rep.*, 2017.
- [19] K. J. Gorgolewski, F. Alfaro-Almagro, T. Auer, P. Bellec, M. Capotă, M. M. Chakravarty, N. W. Churchill, A. L. Cohen, R. C. Craddock, G. A. Devenyi, A. Eklund, O. Esteban, G. Flandin, S. S. Ghosh, J. S. Guntupalli, M. Jenkinson, A. Keshavan, G. Kiar, F. Liem, P. R. Raamana, D. Raffelt, C. J. Steele, P.-O. Quirion, R. E. Smith, S. C. Strother, G. Varoquaux, Y. Wang, T. Yarkoni, and R. A. Poldrack, "Bids apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods," *PLOS Computational Biology*, vol. 13, no. 3, pp. 1–16, 03 2017. [Online]. Available: <http://dx.doi.org/10.1371%2Fjournal.pcbi.1005209>
- [20] R. A. Poldrack, D. M. Barch, J. Mitchell, T. Wager, A. D. Wagner, J. T. Devlin, C. Cumba, O. Koyejo, and M. Milham, "Toward open sharing of task-based fmri data: the openfmri project," *Frontiers in neuroinformatics*, vol. 7, p. 12, 2013.
- [21] T. Sherif, P. Rioux, M.-E. Rousseau, N. Kassis, N. Beck, R. Adalat, S. Das, T. Glatard, and A. C. Evans, "Cbrain: a web-based, distributed computing platform for collaborative neuroimaging research," *Recent Advances and the Future Generation of Neuroinformatics Infrastructure*, p. 102, 2015.
- [22] X.-N. Zuo, J. S. Anderson, P. Bellec, R. M. Birn, B. B. Biswal, J. Blautzik, J. C. Breitner, R. L. Buckner, V. D. Calhoun, F. X. Castellanos et al., "An open science resource for establishing reliability and reproducibility in functional connectomics," *Scientific data*, vol. 1, p. 140049, 2014.
- [23] B. A. Landman, A. J. Huang, A. Gifford, D. S. Vikram, I. A. L. Lim, J. A. Farrell, J. A. Bogovic, J. Hua, M. Chen, S. Jarso et al., "Multi-parametric neuroimaging reproducibility: a 3-t resource study," *Neuroimage*, vol. 54, no. 4, pp. 2854–2866, 2011.
- [24] K. B. Noonan, S. Colcombe, R. Tobe, M. Mennes, M. Benedict, A. Moreno, L. Panek, S. Brown, S. Zavitz, Q. Li et al., "The nki-rockland sample: a model for accelerating the pace of discovery science in psychiatry," *Frontiers in neuroscience*, vol. 6, p. 152, 2012.

Appendix A Processing Pipeline

Here we take a deep-dive into each of the modules of the NDMG pipeline. We will explain algorithm and parameter choices that were implemented at each step, and the justification for why they were used over alternatives.

Appendix A.1 Registration

words

Appendix A.2 Tensor Estimation

words

Appendix A.3 Tractography

words

Notes

¹<https://pypi.python.org/pypi/ndmg>

²<https://hub.docker.com/r/bids/ndmg/>

³<http://boutiques.github.io/>

⁴<https://github.com/neurodata/ndmg>

⁵<https://github.com/neurodata/ndmg-paper>