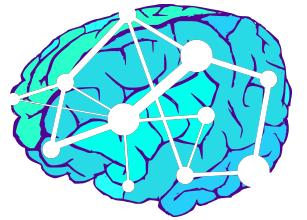


NDMG: A Scalable, Reliable, and Replicable Pipeline for Diffusion-MRI Cloudified Connectome Meganalysis



Gregory Kiar^{1,2†}, William R. Gray Roncal^{3,4†}, Vikram Chandrashekhar²,
Eric W. Bridgeford¹, Disa Mhembere⁴, Randal Burns⁴, Joshua T. Vogelstein^{1,2}

Abstract

The expansion of publicly available multimodal MR datasets enables analysis of the structure and function of the human brain at an unprecedented scale. Alongside development of tools and standards for this data, neuroinformatics studies regularly uncover evidence for patterns between behaviour and structure of the brain. As the field of data and processing tools grows, findings made across studies are becoming increasingly difficult to compare. We have developed a turn-key pipeline for reliable structural connectome estimation at scale, and used it to process and publicly release connectomes from over 2,000 subjects. By virtue of harmonized data processing, we show that while we can yield qualitatively similar graphs, results are meaningfully quantitatively different, making statistics such as p-values and clinical assays incomparable across datasets.

1 Introduction

Multimodal MR (M3R) Imaging enables a diverse look at the structure and function of the human brain. Processing M3R data is becoming an increasingly popular technique in psychology for linking behaviour or mental illness with structure or function of the brain [1]. As such, M3R data is being collected at an unprecedented rate, and publicly available datasets of both healthy and diseased populations are commonplace [2–4].

Processing of M3R data has been greatly helped by the development of several widely adopted tools, including FSL [5–7], SPM [8], CPAC [9], ANTs [10], and many others. While these tools each offer a wide range of processing paradigms and options, “standard” processing methods or pipelines which robustly generate reliable results from M3R data are lacking, and false positive rates can be reported to be misleadingly low [11].

In structural connectome estimation, prior art such as PANDAS [12], CMTK [13], MRCAP [14], and MIGRAINE [15] have each contributed significantly to the advancement of connectomics research. While these tools vary in terms of the algorithms or implementations that used in processing, or the parcellations used to define nodes in the resulting graphs, they have all been built upon the idea of reconstructing white-matter tracts through DWI imaging of the brain.

CMTK and PANDAS enable users to specify parameters for processing; while this increases the flexibility of the tools it ultimately limits the ability of researchers to compare estimated connectomes generated across parameter settings. While MRCAP and MIGRAINE operate with a fixed set of parameter settings, enabling comparison across graphs, connectomes are only able to be generated using a limited set of brain parcellations. Each of these tools has additionally suffered from aging dependencies and lack of scalability, making them difficult to install, run, and ultimately be inaccessible to new users.

We have built a reliable turn-key solution for computing structural connectomes from diffusion and structural MR images that can be deployed at scale either in the cloud or on physical resources. Leveraging existing tools such as FSL, Dipy [16], the MNI152 brain atlas [17] and others, the NDMG pipeline is a one-click pipeline that lowers the barrier to entry to connectomics research. NDMG has been used to process +2,500 brain scans and generate +60,000 connectomes across a range of scales. We have used these connectomes to explore the significance of batch effects in MR data, as well as how parcellations affect brain graphs.

2 Results

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 mul-

timodal MR, or M3R), and receive a robust estimate of connectivity in return. NDMG is integrated with the BIDS [18] organization standard, and the Ama-

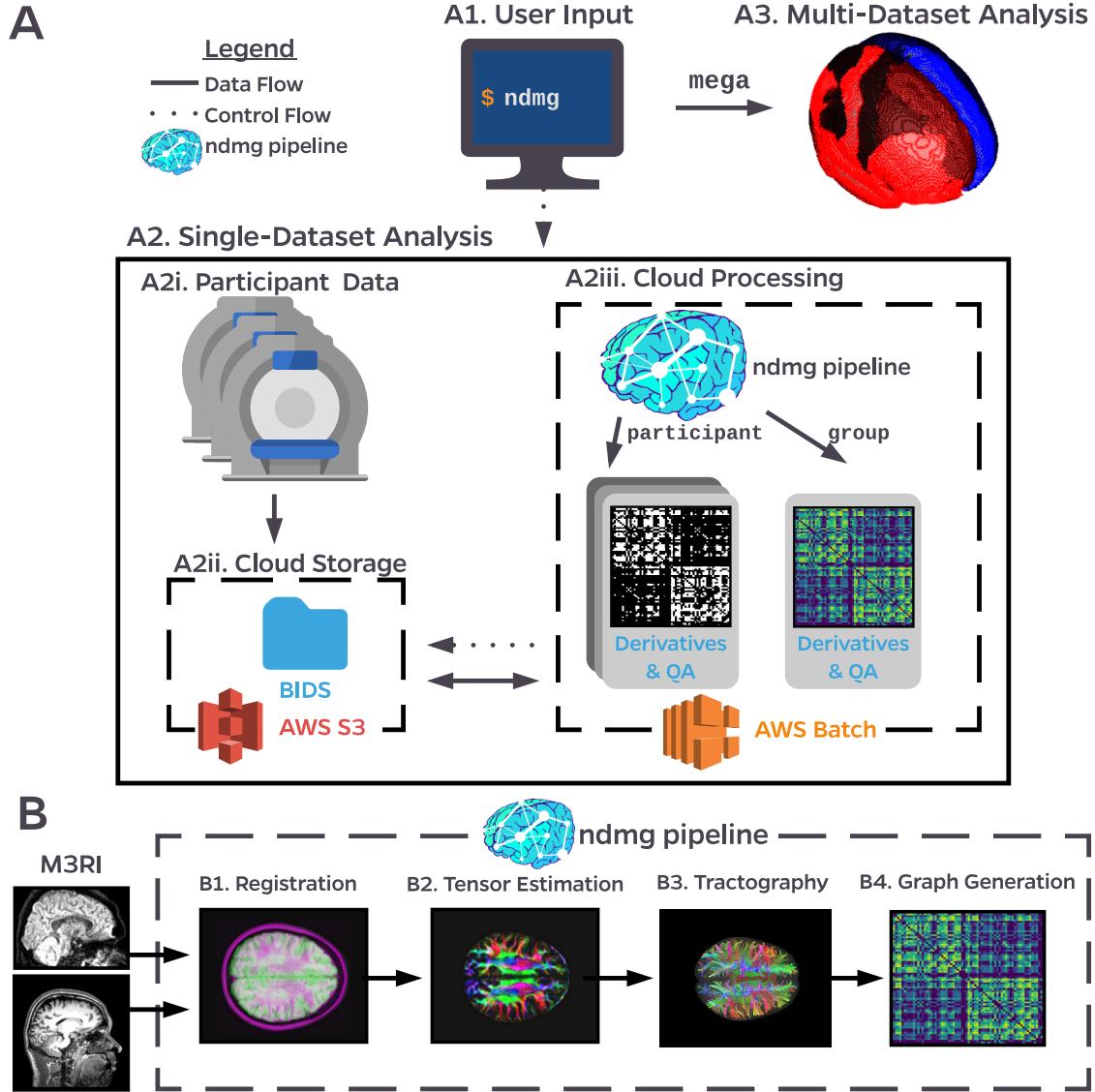


Figure 1: **ndmg usage schematic.** **A)** The NDMG pipeline enables rapidly going from data collection to analysis. NDMG can be launched in the cloud or locally (**A1**). Single-dataset analysis (**A2**) takes M3R data (**A2i**) stored in the cloud in accordance with the BIDS [18] specification (**A2ii**), and launches the pipeline in the cloud (**A2iii**) to perform either participant- or group-level analysis. The derivatives from NDMG are computed consistently and reliably across sessions and datasets, enabling “mega-analysis,” the investigation of scientific questions across multiple datasets (**A3**). **B)** The NDMG pipeline consists of 4 main steps: Registration (**B1**), Tensor Estimation (**B2**), Tractography (**B3**), and Graph Generation (**B4**). At each stage, NDMG produces quality assurance figures of the derivatives, as can be seen in **B1-4**. Each of these steps contains several routines which will be explored in Section 4. The NDMG package contains and leverages tools such as FSL [5-7], Dipy [16], MNI152 [17], and a variety of parcellations [19-25].

zon Web Services (AWS) computing cloud, allowing it to serve as a turn-key solution for connectome estimation regardless of available local computational resources.

2.1 One-Click Connectomics

NDMG is a reliable, accessible, turn-key solution to human connectome estimation at scale. The NDMG pipeline takes approximately 1 hour to generate connectomes across 24 parcellations for a single session, when provided with 1 CPU and 12 GB of RAM. 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 [18; 28] and the descriptive command-line schema Boutiques³. NDMG has been deployed in several large computing infrastructures including OpenfMRI [29] and CBRAIN [30], and has been integrated with the AWS computing cloud, enabling users to launch the pipeline in the cloud with a single command [31; 32]. The NDMG package and pipeline is open-source and available on Github⁴.

2.2 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 and intermediate derivatives are publicly available on our S3 bucket, `mrneurodata`. The data can 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 assurance plots (like those in Figure 6) would likely increase the score further.

Figure 3 shows a variety of uni- and multi-variate statistics of the average graphs produced by each of the datasets enumerated in Table 1.

2.3 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 2 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 72,783; only those with nodes under 500 nodes are shown in Figure 2. 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.

TODO: gk: define a narrative

2.4 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-id with dataset as the observation class label, we can compare the similarity between graphs derived within a dataset and across datasets. Shown in Figure 4, 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 is an equal probability of selecting each of the 10 datasets. However, we compute a discriminability score of 0.632, suggesting significant dataset-specific signal is present. There is an obviously difference between KKI2009, acquired on a Philips scanner, and the other datasets, acquired predominantly on Siemens scanners and one using GE; computing the discriminability here without KKI2009 yields a score of 0.626 which is still significant at $p < 1E^{-4}$. Removing the sole dataset using a GE scanner, HNU1, we compute the discriminability score considering only data from Siemens sites to be 0.627, which is also significant at the same confidence level. 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.



Table 1: **Processed public M3R datasets.** Each dataset processed with NDMG has connectomes ranging across 24 parcellations, resulting in over 65,000 total publicly available connectomes at <http://m2g.io>. For TRT datasets, discriminability is calculated and ranges from 1 (perfect) to 0, with chance at $1/k$. The total discriminability is the computed when all TRT datasets are analyzed as one.

Dataset	Scanner	Diff. Dirs	Avg Subj Age	Portion Male	Subjs (k)	Scans	Total	Disc
BNU1 [3]	Siemens	30	23.0 ± 2.3	0.53	57	2	114	0.984
BNU3 [3]	Siemens	64	22.5 ± 2.1	0.50	48	1	47	-
HNU1 [3]	GE	33	24.4 ± 2.3	0.50	30	10	300	0.993
KKI2009 [26]	Philips	33	31.8 ± 9.4	0.52	21	2	42	1.0
MRN1313	-	70	-	-	1313	1	1299	-
NKI1 [3]	Siemens	137	34.4 ± 12.8	0.0	24	2	40	0.984
NKI-ENH [27]	Siemens	137	-	-	198	1	198	-
SWU4 [3]	-	93	20.0 ± 1.3	0.51	235	2	454	0.884
Templeton114	Siemens	70	21.8 ± 3.0	0.58	114	1	114	-
Templeton255	Siemens	150	-	-	255	1	253	-
Total					2295		2861	0.979

2.5 Mean Connectome

We computed a mean connectome using the Desikan parcellation for each dataset processed in Ta-

ble 1, as well as a mean and standard deviation (SD) connectome across datasets, as shown in Figure 5. With $N=2861$, we believe this is the largest mean connectome computed to date. Properties of the brain

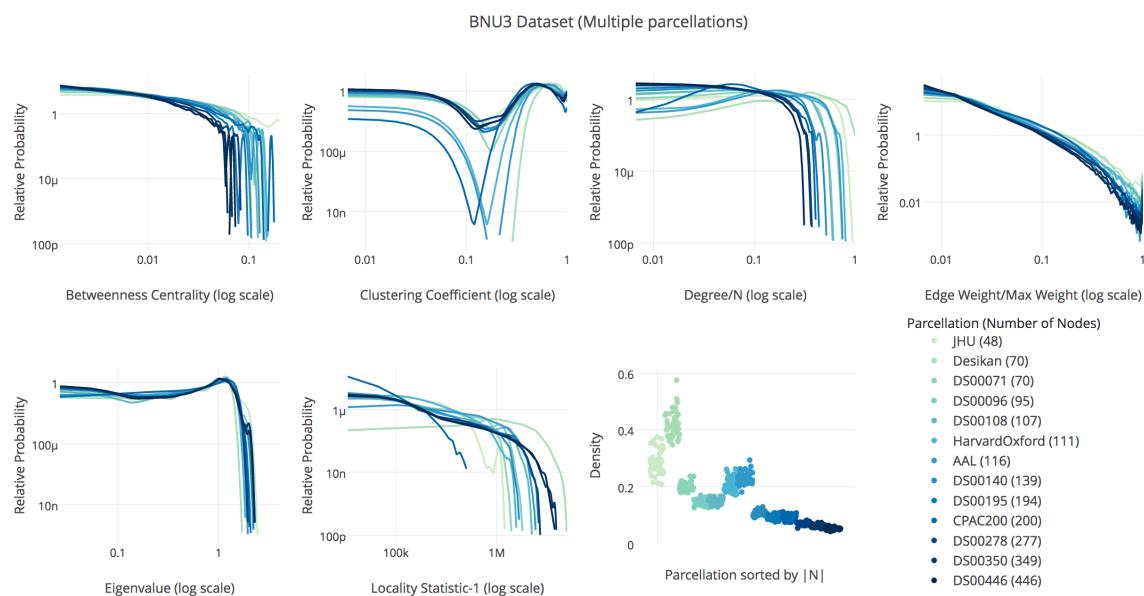


Figure 2: **Multi-scale graph analysis.** NDMG produces connectomes at a variety of scales, enabling investigation of graph properties between parcellation schemes. We can observe that the statistics are qualitatively similar in shape across scales, however, they are quantitatively significantly different. This suggests that claims made or analyses performed on a given scale may not hold when applied to another scale. This is impactful, as the choice of parcellation has significant bearing on the results of a scientific study.

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.

3 Discussion

The NDMG pipeline is a highly-reliable tool with a low barrier to entry for neuroscientists. NDMG has shown that it is capable of producing meaningful and consistent graphs across scales and datasets, and to be deployed across a wide range of computing infrastructures with limited dependencies or computational resources. While many existing tools for connectome estimation enable the user to easily select hyper-parameters for their data, NDMG abstracts this selection from users; though this means that NDMG may not use the optimal parameters for

a given dataset, it provides a consistent estimate of connectivity across a wide range of datasets, making it trivial to compare graphs derived from one study to another, and avoids overfitting of the pipeline to a specific dataset.

While NDMG has been verified to be highly reliable, a hyper-parameter sweep over all parameter selections has not yet been conducted. Performing such an optimization over a range of datasets and maximizing their joint discriminability would improve the reliability of NDMG even further. Though this complete optimization has not occurred, important contribution of NDMG is the illustration that highly sophisticated and computationally burdensome algorithms such as probabilistic tractography are not necessary to create highly reliable estimates of brain connectivity. Evaluating the quality of connectome estimation pipelines using probabilistic tractography with discriminability would enable a decisive answer to the question of whether deterministic or probabilistic tractography is more a reliable when estimating connectivity.

We show that batch effects are a significant and potentially troublesome artefact that accompany M3R datasets. Though this manuscript does not

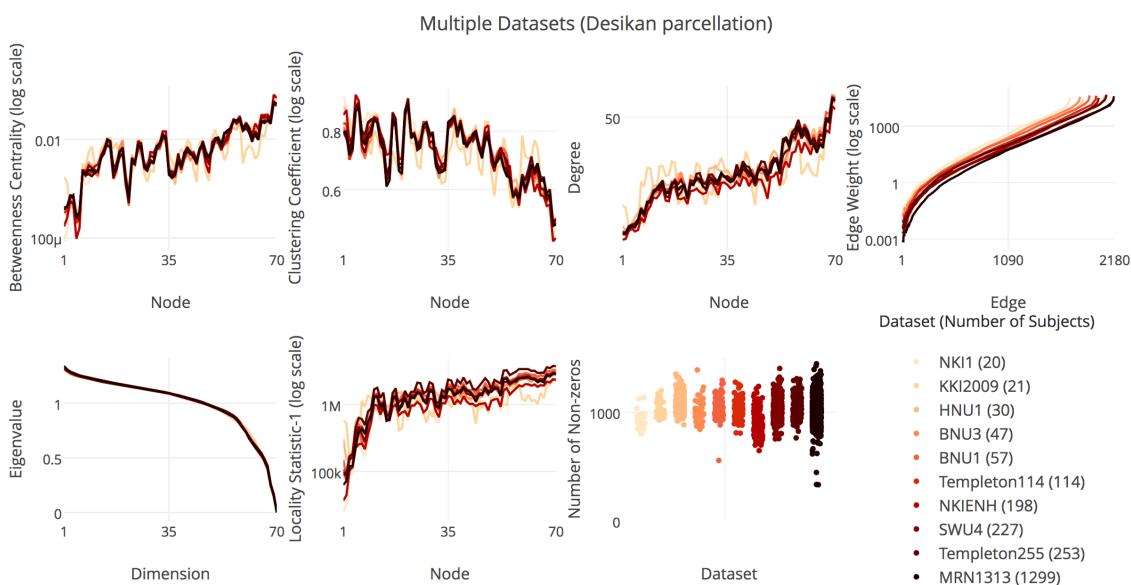


Figure 3: Multi-site graph analysis. NDMG has been used to process 10 publicly available datasets, using an identical version of the pipeline on the same hardware. We can see that though the underlying structure of the connectomes produced is similar by most statistics, there is high variance between data collected as a part of one study to another, even when averaging over the entire cohort.

Multi-Study Discriminability Using Desikan Atlas

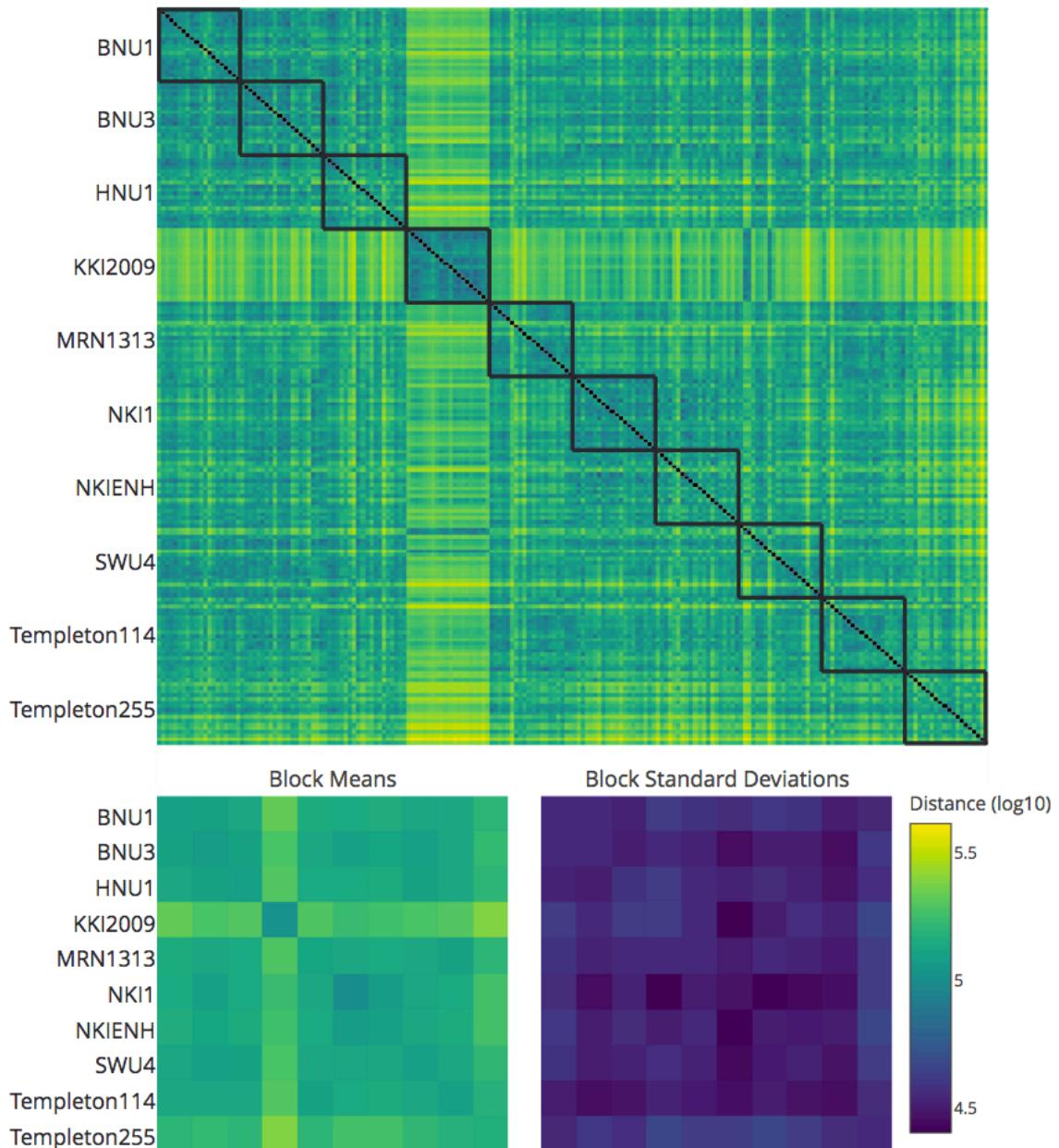


Figure 4: **Prevalence of batch effects.** Discriminability was computed across the ten processed datasets using the dataset-id as the class label. If no significant difference between datasets exists, the discriminability score would not be significantly different from chance, a score of 0.363. Here, the discriminability score was 0.632, which is significant with a p-value of less than 0.0001 when performing a permutation test, suggesting that there is significant dataset specific signal in the graphs.

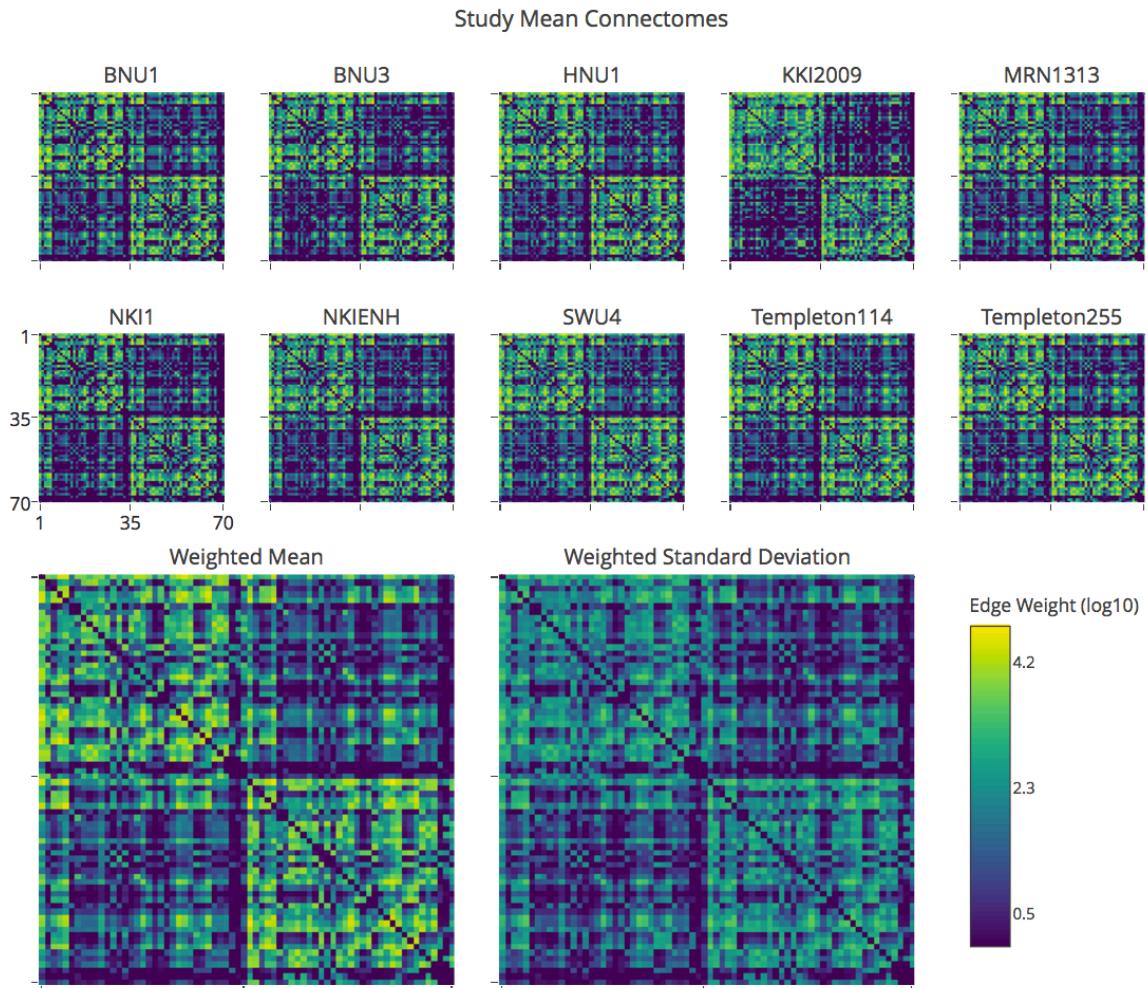


Figure 5: Multi-Study Mean Connectomes. Looking at 70-nodes graphs built upon the Desikan parcellation, we have computed the mean connectome from a variety of datasets. The mean-of-mean connectomes and standard deviation-of-mean connectomes were computed, and result in the largest known mean connectome 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.

present efforts to mitigate these effects, there are a variety of studies which propose methods for data harmonization upon either minimally pre-processed or raw M3R data [33–35]. It is our belief that these approaches may reduce the impact of batch effects on downstream derivatives and subsequent inference, and should be further investigated as possible pre-processing steps that researchers conducting quantitative analysis of M3R images add to their

pipelines.

The Human Connectome Project (HCP) has recently collected an extraordinary M3R dataset, and developed processing pipelines accordingly. It is our belief that NDMG is explicitly not in competition with the tools and resources produced by HCP. The HCP pipeline when run on the HCP data creates likely the most outstanding map of human brain connectivity to date, however, their pipeline relies on many ad-

ditional parameters and files (such as scanner bias fields, and multi-shell acquisitions of DWI data), making it unable to be deployed across the majority of existing DWI datasets. NDMG takes the approach of being a generally useful, reliable, and impactful tool, without requiring state-of-the-art acquisitions which are not commonly available outside of next-gen experimental paradigms. NDMG can be deployed on the HCP data, though it would not take advantage of the additional scans and parameters acquired. We believe that the HCP pipelines and NDMG complement one another, and comparing the two more closely is an exciting avenue to be explored in future work.

As NDMG is highly modular, it can also serve as a reference pipeline for new algorithms when performing connectome estimation. Through the use of discriminability, the NDMG pipeline can be modified with new algorithms and tested against the original. When algorithms consistently improve the reliability of NDMG, they may be integrated into the pipeline and both i) improve the quality of the NDMG reference pipeline, as well as ii) provide a strong statistical basis for the development and use of new algorithms.

Integration of NDMG to the AWS cloud and with services such as OpenfMRI and CBRAIN makes it an accessible choice for neuroscientists who are limited in their in-house compute infrastructure, and those with limited experience deploying pipelines in high performance computing environments.

The NDMG pipeline takes a step towards democratizing the field of human connectomics, enabling users to produce statistically-validated connectomes at an unprecedented scale with ease, enabling massive-scale analyses of the underlying structure of the human brain.

4 Methods

The NDMG pipeline has been developed by leveraging and interfacing existing tools, including FSL [5–7], Dipy [16], the MNI152 atlas [17], and a variety of parcellations defined in the MNI152 space [19–25]. 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 process 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 common space. To accomplish this, we transform the input M3R into the template MNI152 [17] 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 [5–7], 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 [16] a simple 6-component tensor model is used. EUDX [36] performs deterministic tractography, an algorithm inspired by FACT [37], and 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 the HarvardOxford cortical atlas [22], JHU [21], and Talairach [23] atlases are packaged with FSL, while others were found from publications directly, such as the Desikan [19], AAL [20], Slab907 [24], and Slab1068 [25] parcellations, while the remaining were generated from segmentation algorithms, including CC200 [9] and 16 downsampled (DS) parcellations [38] 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. See Appendix A.4 for more information.

4.1 Quality Assurance

As they are the final derivatives produced, connectomes, are far removed from the original imaging data. Importantly, we have developed 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 the main body of Figure 1. The QA plots have been 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 more easily detectable by the user.

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

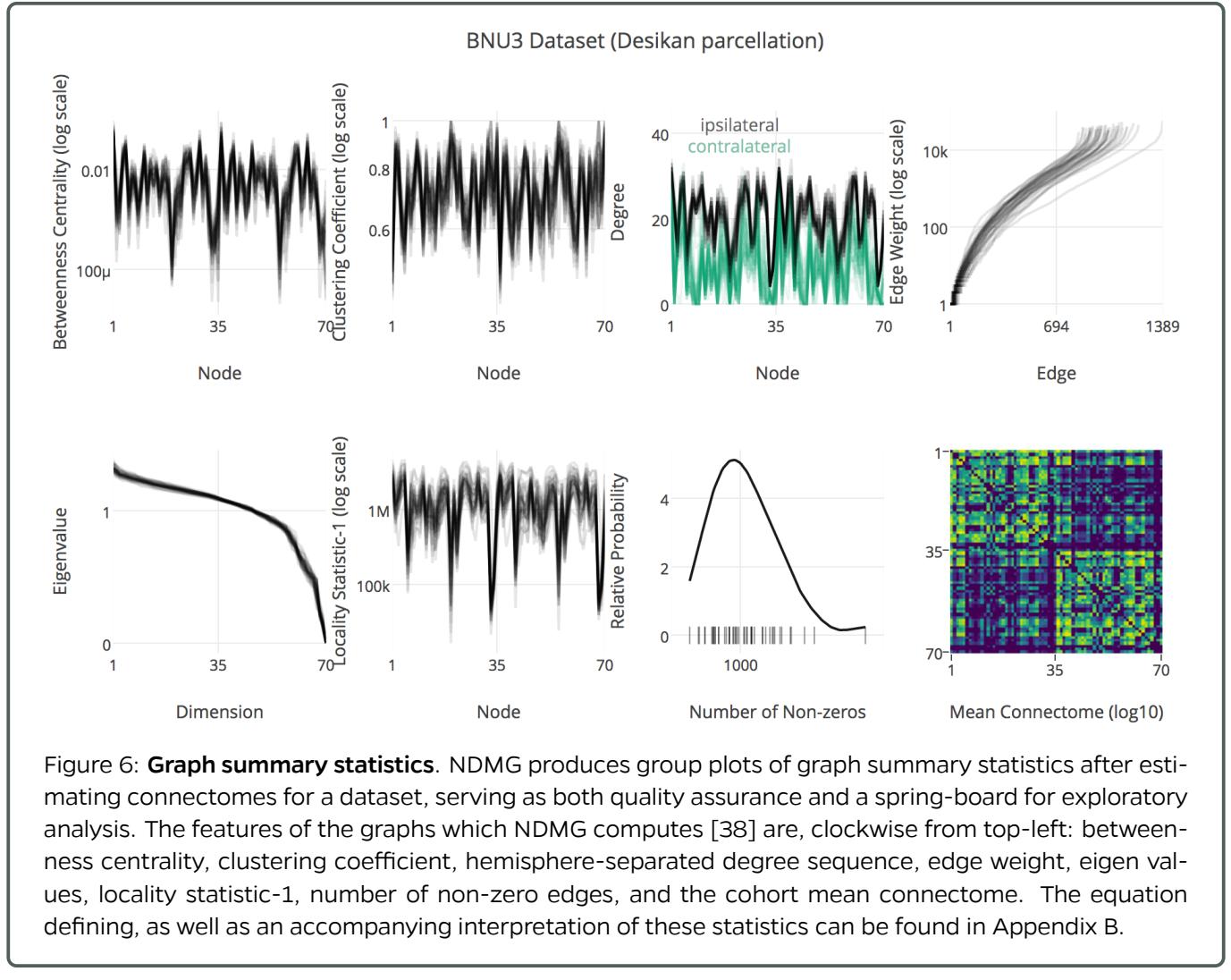


Figure 6: Graph summary statistics. NDMG produces group plots of graph summary statistics after estimating connectomes for a dataset, serving as both quality assurance and a spring-board for exploratory analysis. The features of the graphs which NDMG computes [38] are, clockwise from top-left: betweenness centrality, clustering coefficient, hemisphere-separated degree sequence, edge weight, eigen values, locality statistic-1, number of non-zero edges, and the cohort mean connectome. The equation defining, as well as an accompanying interpretation of these statistics can be found in Appendix B.

obtained brain graphs. Figure 6 shows the summary statistics plot for an exemplar dataset constructed with the Desikan [19] parcellation. The features of the graphs which NDMG computes are [38], 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 assurance of the graphs; for instance, enabling the user to confirm that edge density is higher ipsilaterally than contra-laterally, and thus serve as a preliminary exploratory analysis of the processed data.

4.2 Validation

Results from NDMG have been validated using a metric termed Discriminability [39]. Discriminability as seen in Equation (1) is a statistically generalized version of Test-ReTest (TRT) reliability:

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

Discriminability describes the probability in given a dataset consisting of multiple observations of objects within multiple classes, 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, $a_{i'j'}$. In this case, the class is subject-id and the observations are a session of scans. A perfect discriminability score is 1, meaning that the nearest-neighbour for any observation is always a different observation from the same subject, and the worst possible score is 0, meaning the opposite. 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).

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 Laboratory, Laurel, MD, USA.

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

[†]Authors contributed equally.

Declarations

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

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

References

- [1] M. Goodkind, S. B. Eickhoff, D. J. Oathes, Y. Jiang, A. Chang, L. B. Jones-Hagata, B. N. Ortega, Y. V. Zaiko, E. L. Roach, M. S. Korgaonkar et al., "Identification of a common neurobiological substrate for mental illness," *JAMA psychiatry*, vol. 72, no. 4, pp. 305–315, 2015.
- [2] J. A. Nielsen, B. A. Zielinski, P. T. Fletcher, A. L. Alexander, N. Lange, E. D. Bigler, J. E. Lainhart, and J. S. Anderson, "Multisite functional connectivity mri classification of autism: Abide results," 2013.
- [3] X.-N. Zuo, J. S. Anderson, P. Bellec, R. M. Birn, B. B. Biswal, J. Blautz, 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.
- [4] S. Das, A. P. Zijdenbos, D. Vins, J. Harlap, and A. C. Evans, "Loris: a web-based data management system for multi-center studies," *Frontiers in neuroinformatics*, vol. 5, p. 37, 2012.
- [5] 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>
- [6] 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>
- [7] 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>
- [8] S. B. Eickhoff, K. E. Stephan, H. Mohlberg, C. Grefkes, G. R. Fink, K. Amunts, and K. Zilles, "A new spm toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data," *NeuroImage*, vol. 25, no. 4, pp. 1325–1335, 2005.
- [9] R. C. Craddock, G. A. James, P. E. Holtzheimer, X. P. Hu, and H. S. Mayberg, "A whole brain fmri atlas generated via spatially constrained spectral clustering," *Human brain mapping*, vol. 33, no. 8, pp. 1914–1928, 2012.
- [10] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, "A reproducible evaluation of ants similarity metric performance in brain image registration," *NeuroImage*, vol. 54, no. 3, pp. 2033–2044, 2011.
- [11] A. Eklund, T. E. Nichols, and H. Knutsson, "Cluster failure: why fmri inferences for spatial extent have inflated false-positive rates," *Proceedings of the National Academy of Sciences*, p. 201602413, 2016.

- [12] Z. Cui et al., "PANDA: a pipeline toolbox for analyzing brain diffusion images." *Frontiers in human neuroscience*, vol. 7, p. 42, jan 2013. [Online]. Available: <http://journal.frontiersin.org/article/10.3389/fnhum.2013.00042/abstract>
- [13] A. Daducci et al., "The connectome mapper: an open-source processing pipeline to map connectomes with MRI." *PloS one*, vol. 7, no. 12, p. e48121, jan 2012. [Online]. Available: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0048121>
- [14] W. R. Gray et al., "Magnetic Resonance Connectome Automated Pipeline," p. 5, nov 2011. [Online]. Available: <http://arxiv.org/abs/1111.2660>
- [15] W. Gray Roncal et al., "MIGRAINE: MRI Graph Reliability Analysis and Inference for Connectomics," *Global Conference on Signal and Information Processing*, 2013.
- [16] E. Garyfallidis, M. Brett, B. Amirkhanian, 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.
- [17] 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.
- [18] 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."
- [19] 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.
- [20] 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.
- [21] K. Oishi et al., *MRI atlas of human white matter*. Academic Press, 2010.
- [22] N. Makris, J. M. Goldstein, D. Kennedy, S. M. Hodge, V. S. Caviness, S. V. Faraone, M. T. Tsuang, and L. J. Seidman, "Decreased volume of left and total anterior insular lobule in schizophrenia," *Schizophrenia research*, vol. 83, no. 2, pp. 155–171, 2006.
- [23] J. Lancaster, "The Talairach Daemon, a database server for Talairach atlas labels," *NeuroImage*, 1997.
- [24] C. S. Sriprada 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.
- [25] 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.
- [26] 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.
- [27] K. B. Nooner, 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.
- [28] 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/journal.pcbi.1005209>
- [29] R. A. Poldrack, D. M. Barch, J. Mitchell, T. Wager, A. D. Wagner, J. T. Devlin, C. Cumbo, O. Koyejo, and M. Milham, "Toward open sharing of task-based fmri data: the openfmri project," *Frontiers in neuroinformatics*, vol. 7, p. 12, 2013.
- [30] 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.
- [31] G. Kiar, K. J. Gorgolewski, D. Kleissas, W. Gray Roncal, B. Litt, B. Wandell, R. A. Poldrack, M. Wiener, R. Vogelstein, R. Burns, and J. T. Vogelstein, "Science in the cloud (sic): A use case in mri connectomics," *GigaScience*, vol. gix013, mar 2017.
- [32] G. Kiar, K. J. Gorgolewski, D. Kleissas, W. G. Roncal, B. Litt, B. Wandell, R. A. Poldrack, M. Wiener, R. J. Vogelstein, R. Burns, and J. T. Vogelstein, "Example use case of sic with the ndmg pipeline (sic:ndmg)," 2017. [Online]. Available: <https://doi.org/10.5524/100285>
- [33] H. Mirzaalian, L. Ning, P. Savadjiev, O. Pasternak, S. Bouix, O. Michailovich, S. Karmacharya, G. Grant, C. E. Marx, R. A. Morey et al., "Multi-site harmonization of diffusion mri data in a registration framework," *Brain Imaging and Behavior*, pp. 1–12, 2017.
- [34] J.-P. Fortin, D. Parker, B. Tunc, T. Watanabe, M. A. Elliott, K. Ruparel, D. R. Roalf, T. D. Satterthwaite, R. C. Gur, R. E. Gur et al., "Harmonization of multi-site diffusion tensor imaging data," *bioRxiv*, p. 116541, 2017.
- [35] E. Olivetti, S. Greiner, and P. Avesani, "Adhd diagnosis from multiple data sources with batch effects," *Frontiers in systems neuroscience*, vol. 6, p. 70, 2012.
- [36] 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.
- [37] 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.
- [38] D. Mhembe, 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.
- [39] S. Wang, Z. Yang, X.-N. Zuo, M. Milham, C. Craddock, G. Kiar, W. R. Gray Roncal, E. Bridgeford, CORR, C. E. Preibe, and J. T. Vogelstein, "Optimal decisions for discovery science via maximizing discriminability: Applications in neuroimaging," *Tech. Rep.*, 2017.

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

Registration in NDMG leverages FSL and the Nilearn Python package. The primary concern in development of NDMG was the reliability and robustness of each step. Additionally, a desired feature of the pipeline was that it could be run on non-specialized hardware in a timeframe that didn't significantly hinder the rate of progress of scientists who wish to use it. As such, NDMG uses linear registrations, as non-linear methods were found to have higher variability across datasets while simultaneously increasing the resource and time requirements of the pipeline.

As is seen in Figure 7B1, the first registration step is Eddy-current correction and DWI self-alignment to the volume-stack's BO volume. FSL's `eddy_correct` was used to accomplish this. The `eddy_correct` function was chosen over the newer `eddy` function as the `eddy` function, while providing more sophisticated denoising, takes significantly longer to run or relies on GPU acceleration, which would reduce the accessibility of NDMG.

Once the DWI data is self aligned, it is aligned to the same-subject T1w image through FSL's `epi_reg`. This

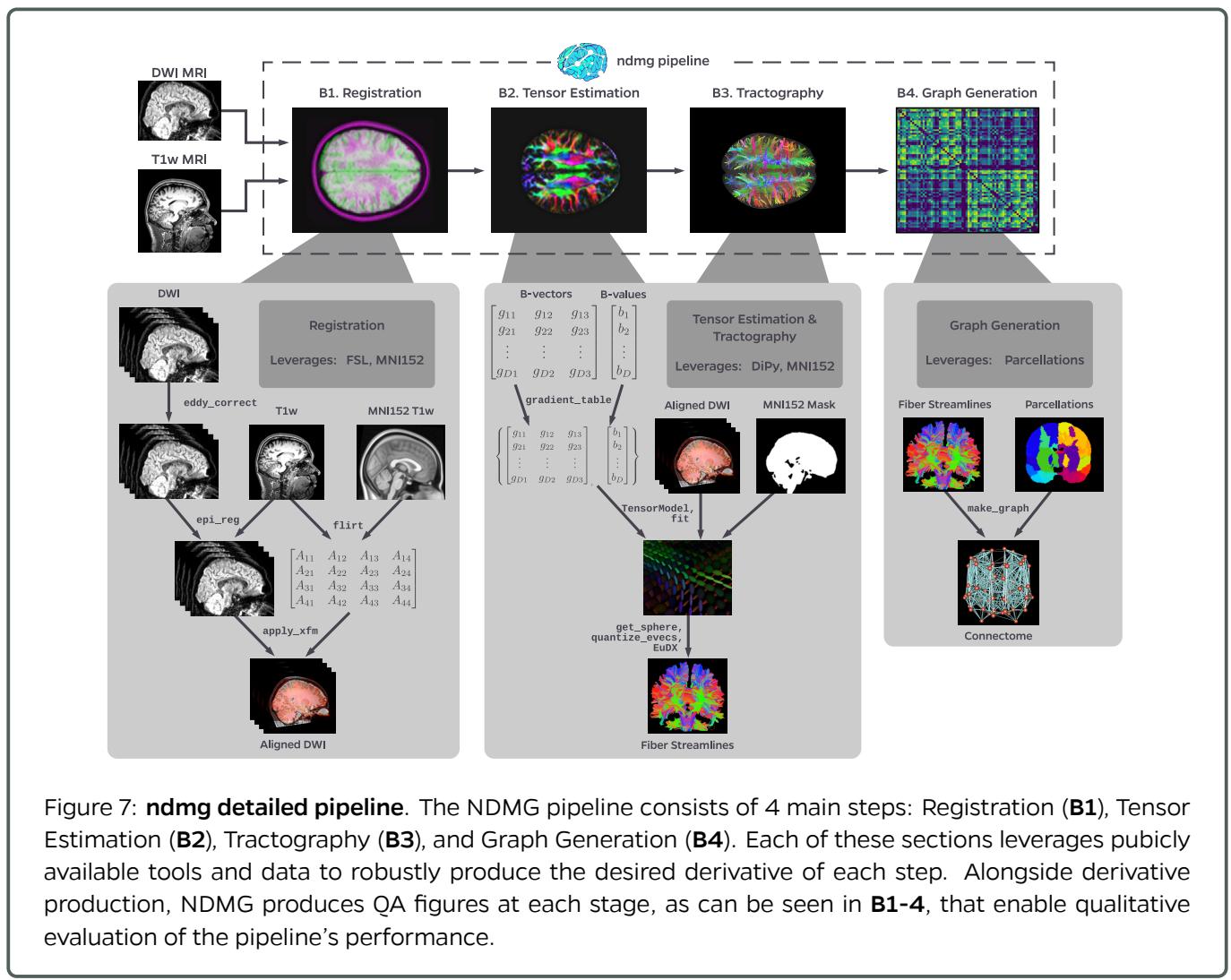


Figure 7: **ndmg detailed pipeline**. The NDMG pipeline consists of 4 main steps: Registration (**B1**), Tensor Estimation (**B2**), Tractography (**B3**), and Graph Generation (**B4**). Each of these sections leverages publicly available tools and data to robustly produce the desired derivative of each step. Alongside derivative production, NDMG produces QA figures at each stage, as can be seen in **B1-4**, that enable qualitative evaluation of the pipeline's performance.

tool performs a linear alignment between each image in the DWI volume-stack and the T1w volume.

The T1w volume is then aligned to the MNI152 template using linear registration computed by FSL's `flirt`. This alignment is computed using the 1mm MNI152 atlas, as this enables higher freedom in terms of the parcellations that may be used, such as near-voxelwise parcellations that have been generated at 1mm. FSL's non-linear registration, `fnirt`, is not used in NDMG as the performance was found to vary significantly based on the collection protocol of the T1w images, often resulting in either slightly improved or significantly deteriorated performance.

The transform mapping the T1w volume to the template is then applied to the DWI image stack, resulting in the DWI image being aligned to the MNI152 template in *real*-coordinate space. However, while `flirt` aligns the images in real space, it does not guarantee an overlap of the data in voxelspace. Using Nilearn's `resample`, NDMG ensures that images are aligned in both voxel- and real-coordinates so that all analyses can be performed equivalently either with or without considering the image affine-transforms mapping the data matrix to the real-coordinates.

Finally, NDMG produces a QA plot showing 3 slices of the first BO volume of the aligned DWI image overlaid on the MNI152 template in the 3 principle coordinate planes, providing 9 plots in total which enable qualitative assessment of the quality of alignment.

Appendix A.2 Tensor Estimation

Once the DWI volumes have been aligned to the template, NDMG begins doing diffusion-specific processing on the data. All diffusion processing in NDMG is performed using the Dipy Python package. The diffusion processing in NDMG is performed after alignment because in order to compare connectomes to one another they must be generated in the same space.

While high-dimensional diffusion models such as orientation distribution functions (ODFs) or qBall impressively allow reconstruction of crossing fibers and complex fiber trajectories, these methods have yet to be demonstrated as effective models when there are not a large number of diffusion volumes/directions for a given image. As NDMG was required to run robustly on as broad a range DWI datasets as possible, a lower-dimensional tensor model was used. The model, described in detail on Dipy's website⁶, computes a 6-component tensor for each voxel in the image, reducing the DWI image stack to a single image which can be used for tractography.

Once tensor estimation has been completed, a similar plot as to that produced for registration QA is produced, showing slices of the FA map derived from the tensors in 9 panels enabling visual inspection of the derivatives.

Appendix A.3 Tractography

In keeping with the theme of robust and widely-applicable methods, tractography was performed with Dipy's deterministic tractography algorithm, EuDX. Integration of tensor estimation and tractography methods was minimally complex with this tractography method, as it has been designed to operate on the tensors produced by Dipy in the previous step. The EuDX tracing algorithm has been shown to be a robust and computationally efficient algorithm for generating streamlines. Probabilistic tractography provides a probability distribution for points along streamlines, which has been shown to be beneficial when working with higher-dimensional diffusion representations. Ultimately, as fibers will be resolved into edges in a graph produced by NDMG, if probabilistic fibers were generated they would have to be thresholded or discretized, limiting the benefit of performing probabilistic tractography even if the diffusion model used were high enough order to benefit from it directly.

A subset of the resolved streamlines are visualized in an axial projection of the brain mask with the fibers contained, allowing the user to verify that streamlines are following expected patterns within the brain and do not leave the boundary of the mask.

Appendix A.4 Graph Estimation

The fiber streamlines produced in the previous step are used by NDMG to generate connectomes across multiple parcellations. The connectomes generated are graph objects, with nodes in the graph representing regions of interest (ROIs), and edges representing connectivity via fibers. Each streamline is traversed and the ROIs which it touches are recorded, ultimately with an edge being added to the graph for all corresponding pairs of ROIs along a streamline. As M3R imaging provides insufficient resolution to recover the direction of flow, an undirected edge is added. Edge weight is determined by the number of streamlines which pass through a given pair of regions - i.e. each fiber connecting regions A and B adds a weight of 1 between them in the graph. Other measures such as fiber length or mean FA along the streamline have not been implemented but could serve as replacements for fiber count, given that a mechanism for combining these values over multiple fibers or normalizing them by the fiber count through a region are also considered.

The parcellations used in NDMG were selected based on availability and use in the DWI processing community. Use of additional parcellations is trivial with NDMG, further enabling analysis to occur across a variety of scales and users to produce a range of connectomes that compliment one another and can be used either independently or for multiscale analyses.

Appendix B Graph Summary Statistics

When NDMG produces graph summary plots it computes eight node- or edge-wise statistics of the connectomes, and displays them for comparison across sessions within the analyzed cohort. The statistics computed were chosen based on their relevance to the field of connectomics and aim to shed light on properties of the graph in which researchers find relevance. The graph statistics are primarily computed with NetworkX and Numpy, and all implementations for NDMG live within the `graph_qa` module⁷. Below, for each statistic we provide a link to the code/documentation of the statistic as it was implemented.

Table 2: **Graph statistics**. Each of the graph statistics computed by NDMG.

Statistic	Implementation
Betweenness Centrality	NetworkX
Clustering Coefficient	NetworkX
Degree Sequence	NetworkX
Edge Weight Sequence	NetworkX
Eigen Values	NetworkX and Numpy
Locality Statistic-1	ndmg and NetworkX
Number of Non-Zero Edges	NetworkX
Cohort Mean Connectome	Numpy

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>

⁶http://nipy.org/dipy/examples_builtin/reconst_dti.html

⁷https://github.com/neurodata/ndmg/blob/master/ndmg/stats/qa_graphs.py