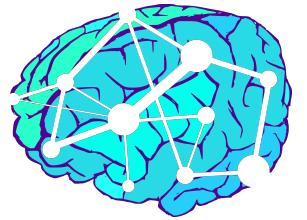


NDMG: A Scalable, Reliable, and Replicable Pipeline for Diffusion-MRI Cloudified Connectome Meganalysis



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 2,861 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 limits the ability of researchers to compare connectomes generated across parameter settings. While MRCAP and MIGRAINE operate with a fixed set of parameters enabling consistent comparison across graphs, connectome generation is limited to the Desikan parcellation [16] or voxelwise graphs. 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 locally. Leveraging existing tools such as FSL, Dipy [17], the MNI152 brain atlas [18] and others, the NDMG pipeline is a one-click pipeline that lowers the barrier to entry for connectomics research. NDMG has been used to process 2,861 brain scans across a range of scales, enabling investigation of the effect of various parcellations on the structure of connectomes and conduct a meganalysis exploring qualitative and quantitative differences across 10 datasets.

2 Results

The NDMG pipeline was developed to provide robust and reliable connectome estimation parallelly

in the cloud, with a low barrier to entry for users. NDMG enables three tiers of analysis: participant-level, group-level, and “meganalysis”, a term we

use here to describe analyses performed across cohorts of data. Figure 1A illustrates how NDMG can be used to perform computation for each of these

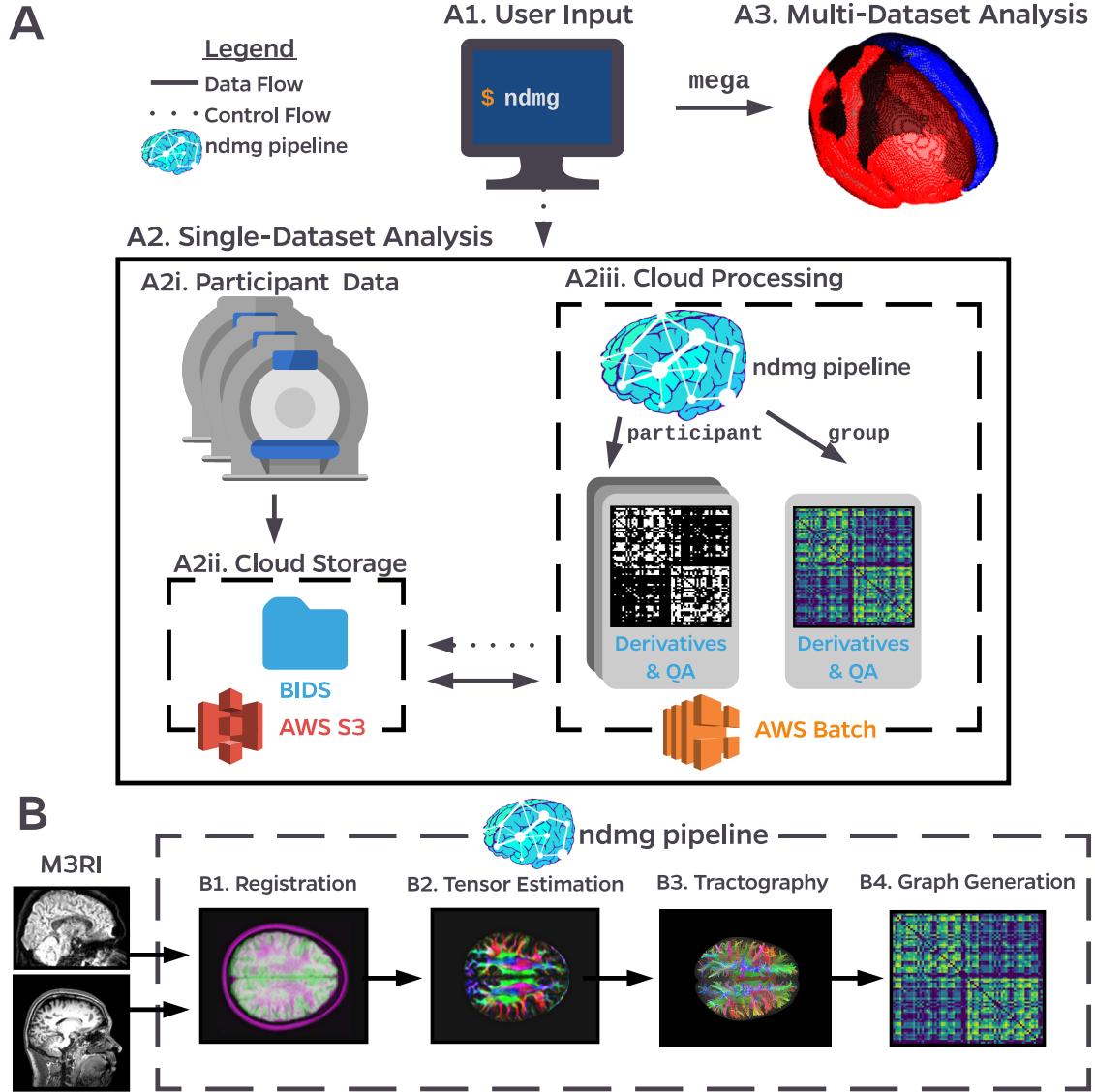


Figure 1: **ndmg usage schematic.** **A)** The NDMG pipeline enables rapidly going from data collection to analysis. NDMG can be launched in the cloud or locally (**A1**). Single-dataset analysis (**A2**) takes M3R data (**A2i**) stored in the cloud in accordance with the BIDS [19] specification (**A2ii**), 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 “meganalysis,” 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 [17], MNI152 [18], and a variety of parcellations [16; 20–25].

analysis levels, and Figure 1B provides an overview of the steps in the core NDMG connectome estimation pipeline. NDMG is integrated with the BIDS [19; 29] 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 [30; 31]. The derivatives produced by NDMG can be easily read into Python or R, allowing for subsequent analysis accessibly.

2.1 Multi-Scale Connectomics at Scale

NDMG can be robustly applied to M3R datasets as a one-click human connectome estimation solution at scale. NDMG 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 the descriptive command-line schema Boutiques³, enabling repeatable deployment across environments. NDMG has been deployed on several large computing infrastructures including OpenfMRI [32] and CBRAIN [33], increasing its accessibility for users of those popular platforms. The NDMG package and pipeline is open-source and available on Github⁴.

After connectome estimation, NDMG optionally computes and plots graph summary statistics for a dataset, as can be seen in Figure 2. The summary statistics computed are explained in detail in Section 4.1, and have been chosen as they highlight features of connectomes that are expected to vary

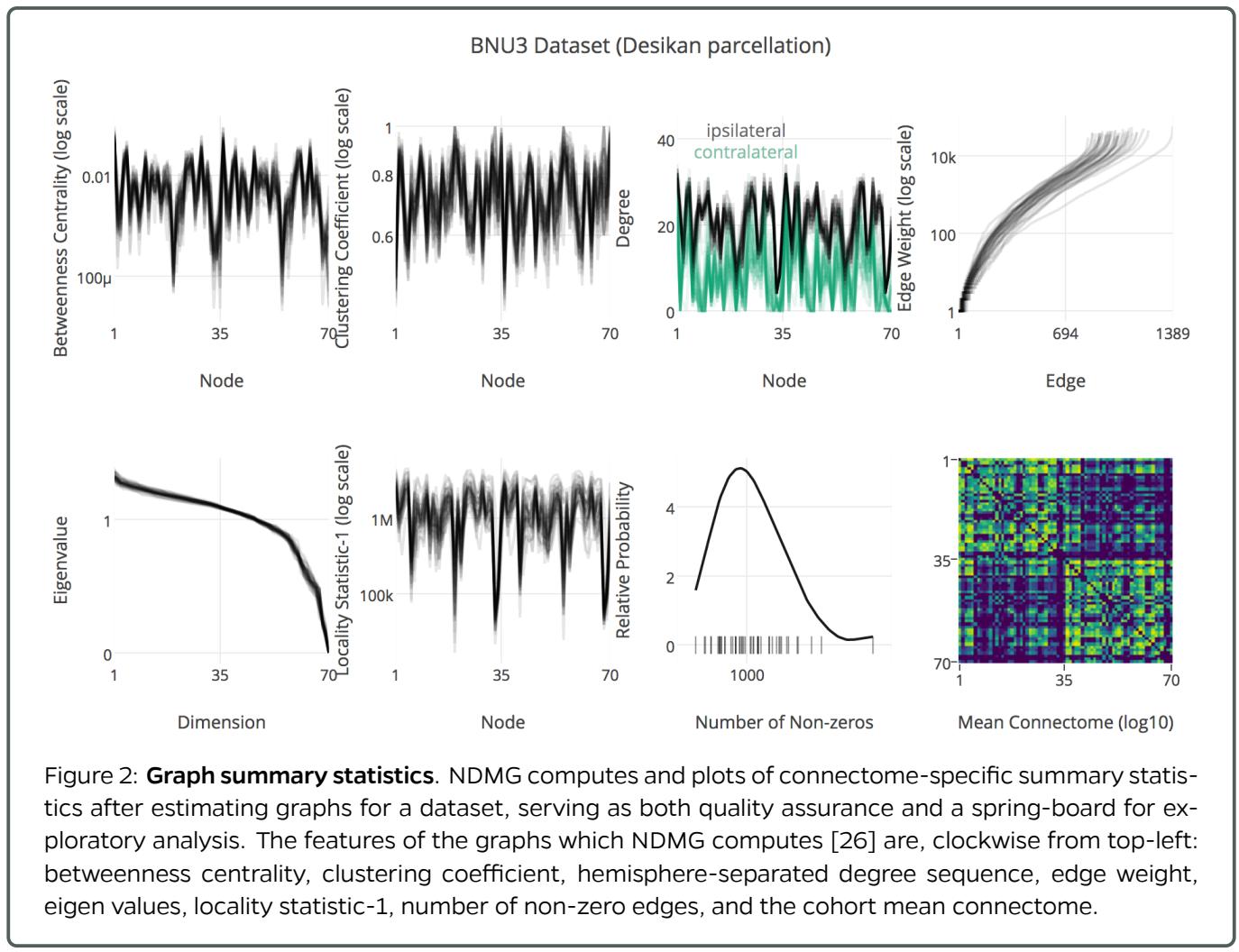


Figure 2: **Graph summary statistics.** NDMG computes and plots of connectome-specific summary statistics after estimating graphs for a dataset, serving as both quality assurance and a spring-board for exploratory analysis. The features of the graphs which NDMG computes [26] are, clockwise from top-left: 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.

Table 1: **Processed public M3R datasets.** The derivatives from each dataset processed with NDMG are publicly available at <http://m2g.io>. For Test-ReTest (TRT) datasets, discriminability evaluates the reliability of NDMG; the score ranges from 0 to 1 (perfect), with chance at $1/k$, where k is the number of subjects. The pooled discriminability is the computed when all TRT datasets are analyzed together. Age is reported as the dataset mean \pm standard deviation.

Dataset	Scanner	Diff. Dirs	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 [27]	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 [28]	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	-
Pooled					2295		2861	0.979

based on brain region, such as the centrality of a node within the graph. These statistics provide a baseline for quality assurance of the connectomes directly, and can inspire further exploratory analysis.

Processed Data The NDMG pipeline has been used to process ten public datasets, upon which it proved to be highly reliable, as is assessed through the discriminability statistic (elaborated upon in Section 4.2). Table 1 summarizes the 2,861 scans processed with NDMG; each scan processed was used to generate connectomes across 24 atlases, resulting in 68,664 total graphs. All of the derived graphs and intermediate derivatives have been made publicly available on our Amazon S3 bucket, `mrneurodata`. The data can also be accessed through <http://m2g.io>.

Multi-Scale Analysis As NDMG produces connectomes at a variety of scales, it enables multi-scale analysis and flexibility in terms of the parcellation a researcher wishes to use for a given analysis task. Figure 3 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 3.

In order to compare the vertex statistics across scales, their values were normalized where necessary (degree and edge weight) and densities were

computed and plotted. As we can see in Figure 3, the shape of the distributions for each statistic are relatively similar across scales. In particular, we notice that graphs from the the downsampled block-atlases (DS) appear to be scaled versions of one another, as is expected as they are related to one-another by region-growing. Similarly, however, graphs from the smaller DS parcellations look similar to those from the neuroanatomically defined parcellations (JHU [21], Desikan [16], HarvardOxford [22], CPACC200 [9]), suggesting that connectomes generated with one parcellation may be considered as scaled versions of another, enabling analysis to be performed across a range of both neuroanatomical and dowsampled parcellations.

2.2 Qualitative Similarity Across Datasets

By way of harmonized processing, we were able to perform meganalysis to begin evaluation of an essential pillar of scientific discovery: reproducibility of findings across datasets. Though inter-subject variability is expected, observing summary statistics or graphs of the average connectome in a dataset are expected to be relatively robust to this variability when studying healthy populations, suggesting that each dataset will look similar. Figure 4 shows a variety of uni- and multi-variate statistics of the average connectome from each of the datasets enumerated in Table 1 on the Desikan parcellation.

We note that datasets largely appear to have similar trends across each of the statistics shown.

Though the KKI2009 [27] dataset appears to be consistent with the others when investigating the edge weight or number of non-zeros plots, the degree sequence or clustering coefficient plots in particular highlight that the KKI2009 dataset may be an outlier compared to the rest. While the remaining datasets have variation between them, they appear to trend more closely together.

Figure 5 shows the mean connectome computed from each dataset, as well as the weighted mean and standard deviation (SD) connectomes combining all datasets. Considering 2,861 scans, we believe this is the largest computed mean connectome to date. The means from each dataset have very similar structures and intensity profiles, with minor noticeable differences predominantly visible in the contra-lateral (off-diagonal) blocks.

Properties of the connectomes such as that they contain a high number of ipsi-lateral connections relative to a lower number of contra-lateral connections are present across all datasets. Additionally, 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, as expected. 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 also be found on our website.

Figures 4 and 5 show similar structure and properties of connectomes across datasets. Though the datasets deviate from one another, this perceived “batch effect” appears to be relatively minor based on commonly assessed properties of connectomes, suggesting that either qualitative analysis is insufficient for detecting batch effects or that batch effects do not play a significant role in this data.

2.3 Quantitative Difference Across Datasets

Though qualitative analysis allows us to do basic quality assurance and verify that the structure and

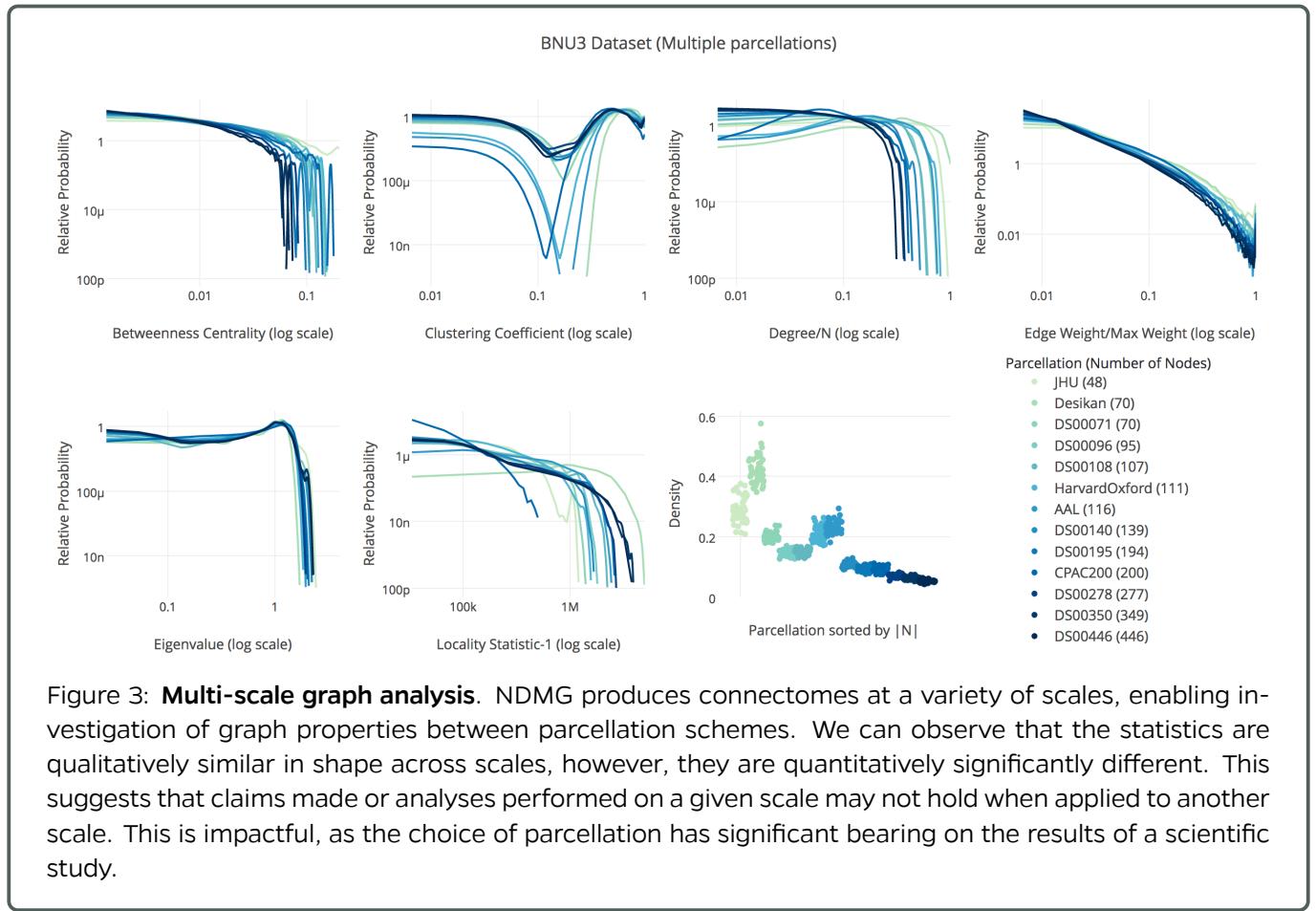


Figure 3: Multi-scale graph analysis. NDMG produces connectomes at a variety of scales, enabling investigation of graph properties between parcellation schemes. We can observe that the statistics are qualitatively similar in shape across scales, however, they are quantitatively significantly different. 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.

general properties of our connectomes are consistent across datasets, it is insufficient if we wish to quantify the similarity of our datasets and by extension the claims made from them. Discriminability [34], a statistic which summarizes cross-class similarity (described in Section ?? and Equation (1)), we have been able to quantify the significance of batch effects in M3RI data.

Using the corresponding dataset as each graph's class label, we computed the discriminability across the non-TRT selection of scans (i.e. 2,295 sessions rather than 2,861). Chance performance, indicating that connectomes were all equally similar regardless of dataset, here results in a discriminability score of 0.363. This is calculated by Equation (2) in Section 4.2. The discriminability score computed here was 0.632, suggesting dataset-specific signal is present at a significance of $p < 0.0001$ when performing a permutation test. The differences between a subset of graphs in each dataset is summa-

rized in Figure 6.

There is an obvious 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 with $p < 0.0001$. 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 significant at the same confidence level.

While the connectomes have appeared similar according to previous qualitative analyses, quantitative analysis indicates that derivatives produced from one dataset can not be trusted to be quantitatively similar to those from another dataset. A consequence of this is that clinical assays or scientific findings made on one group of data may not generalize across datasets, even if the same processing methods are used.

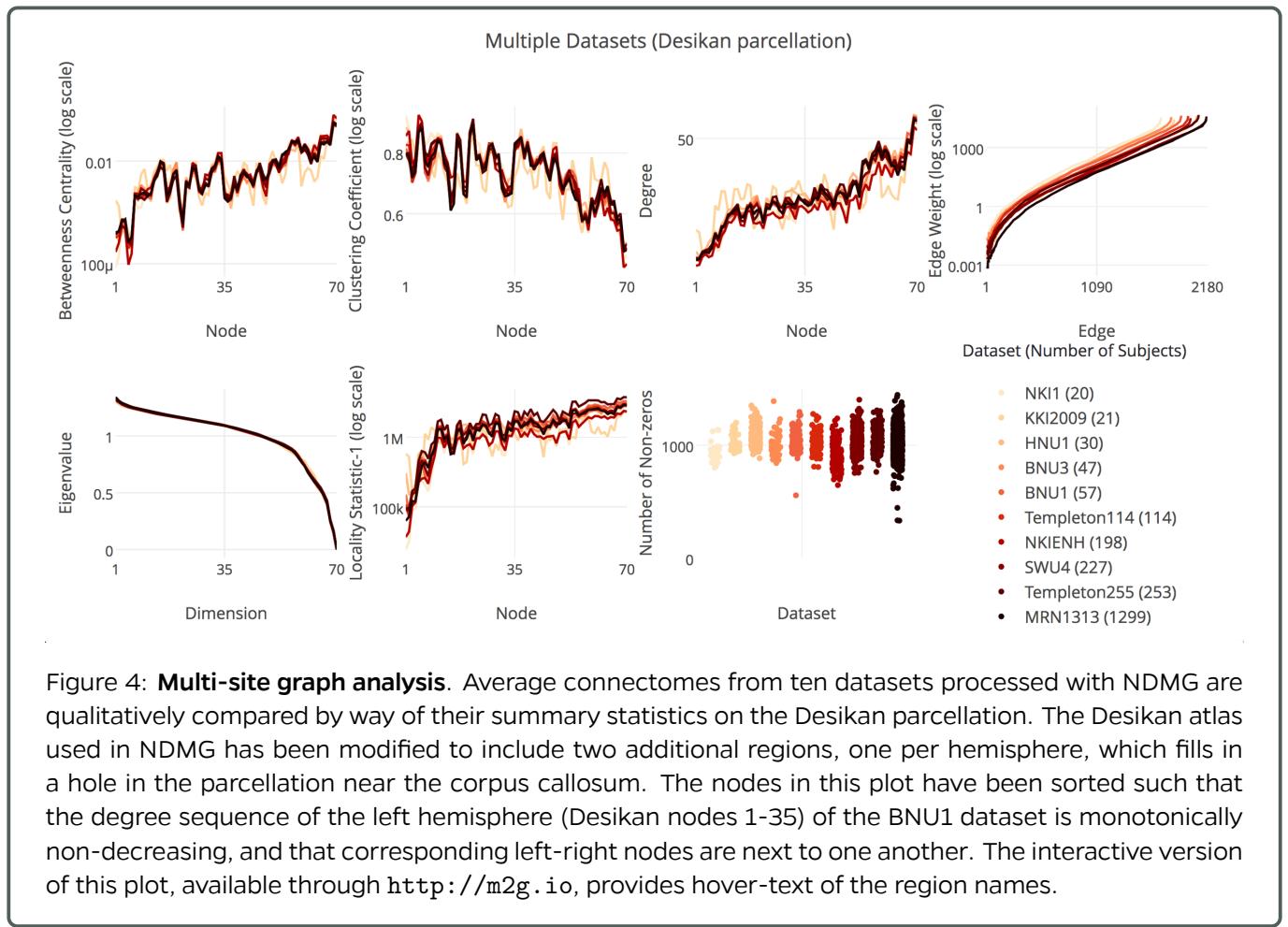


Figure 4: Multi-site graph analysis. Average connectomes from ten datasets processed with NDMG are qualitatively compared by way of their summary statistics on the Desikan parcellation. The Desikan atlas used in NDMG has been modified to include two additional regions, one per hemisphere, which fills in a hole in the parcellation near the corpus callosum. The nodes in this plot have been sorted such that the degree sequence of the left hemisphere (Desikan nodes 1-35) of the BNU1 dataset is monotonically non-decreasing, and that corresponding left-right nodes are next to one another. The interactive version of this plot, available through <http://m2g.io>, provides hover-text of the region names.

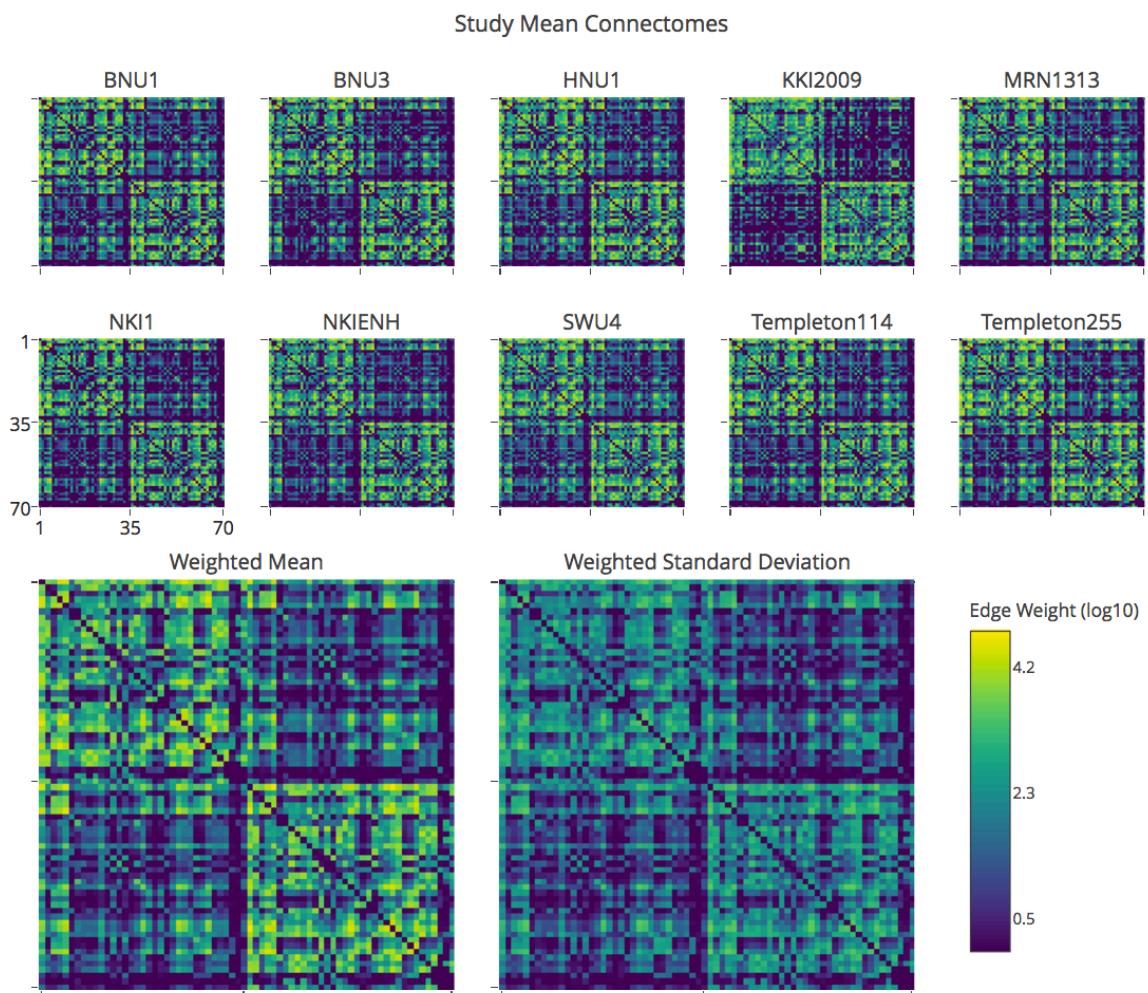


Figure 5: Multi-Study Mean Connectomes. Dataset-mean connectomes and a combined mean-of-mean and standard deviation-of-mean connectomes were computed from the Desikan labeled graphs produced by NDMG, resulting in the largest known mean connectome to-date, consisting of 2,861 sessions. Datasets appear qualitatively similar to one another, with minor deviations particularly visible in the contra-lateral regions of the graphs. 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 these connections more likely to occur, but they have a higher variance, as well.

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.

Multi-Study Discriminability Using Desikan Atlas

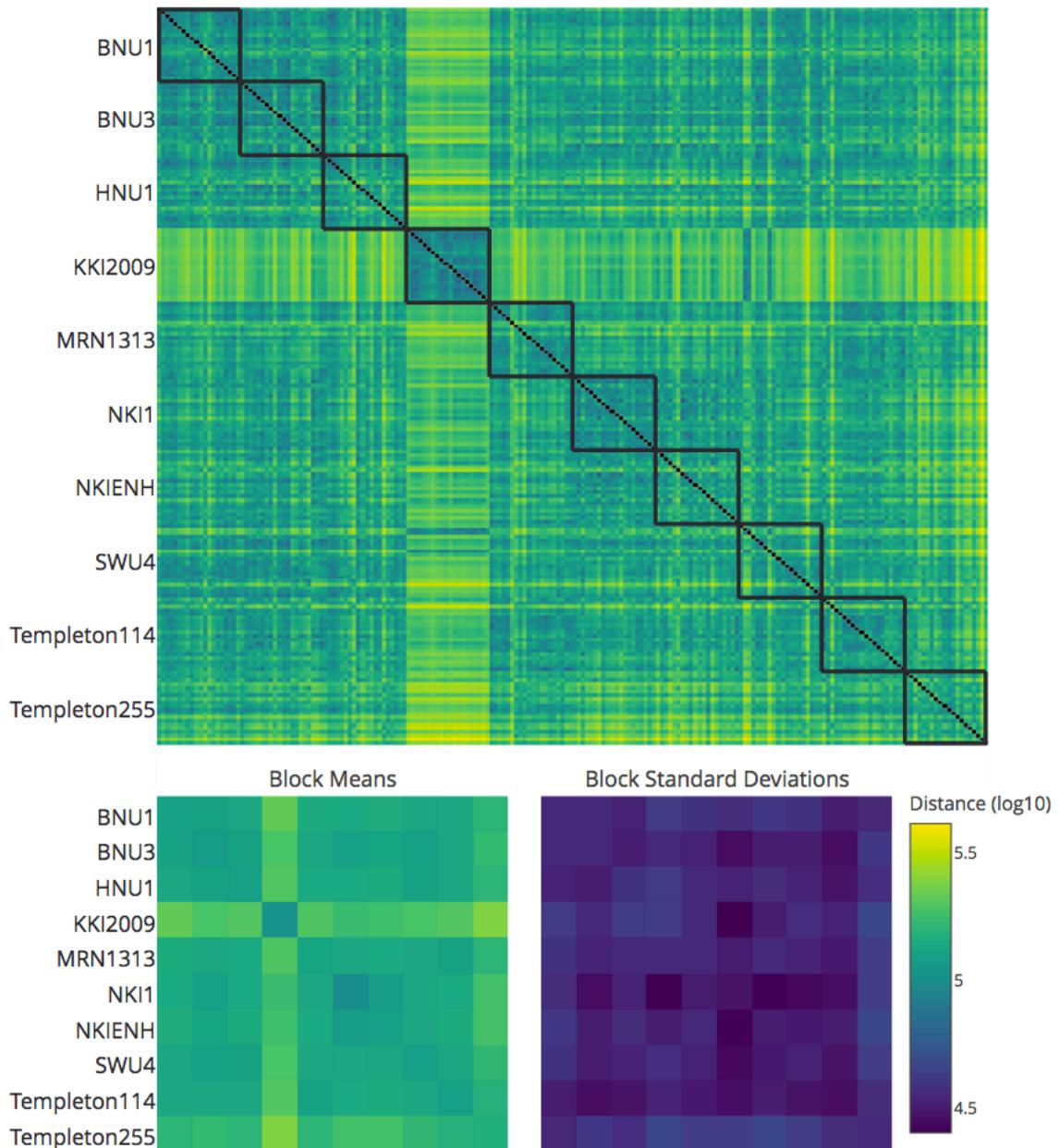


Figure 6: **Prevalence of batch effects.** Discriminability was computed across the ten processed datasets using the dataset-id as the class label. If no significant difference between datasets exists, the discriminability score would not be significantly different from chance, a score of 0.363. Here, 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.

Though NDMG has shown its reliability, a hyper-parameter sweep optimizing each parameter selection in NDMG 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 optimization has not occurred, an 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 reliable when estimating connectivity.

By harmonizing data processing we have been able to remove processing variation and in a statistically-principled manner quantitatively compare connectomes across datasets. We show that while derivatives are highly qualitatively similar across cohorts, these batch effects are significant when performing quantitative analysis. Though the existence of batch effects in M3R data is widely acknowledged, analysis has not previously been conducted to our knowledge that quantified their significance or potential impact on downstream quantitative analyses. Our finding suggests that assays developed from single-cohorts of M3R data can not be reliably considered generalizable, absent a data harmonization technique.

Though this manuscript does not 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 [35–37]. 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 preprocessing steps that researchers conducting quantitative analysis of M3R images add to their pipelines.

As NDMG is highly modular, it can serve as a reference pipeline for data harmonization techniques or specific processing algorithms used in connectome estimation. Through the use of discriminability, the NDMG pipeline can be modified and tested against the original, both to observe the pipeline’s reliability and the prevalence of dataset-specific signal.

When algorithms robustly improve 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, publication, and use of new algorithms.

Integration of NDMG with 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.

Through the use and development of NDMG we have simultaneously lowered the barrier to entry for big-data neuroscience and demonstrated a statistically-principled method for quantifying the prevalence of batch effects in M3R data.

4 Methods

The NDMG pipeline has been developed by leveraging and interfacing existing tools, including FSL [5–7], Dipy [17], the MNI152 atlas [18], and a variety of parcellations defined in the MNI152 space [16; 20–25]. 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 1B, the lower-level schematic provides a glimpse at the inner-workings of NDMG. We break 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 [18] 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 [17] a simple 6-component tensor model is used. EuDX [38] performs deterministic tractography, an algorithm inspired by FACT [39], 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 [16], AAL [20], Slab907 [24], and Slab1068 [25] parcellations, while the remaining were generated from segmentation algorithms, including CC200 [9] and 16 downsampled (DS) parcellations [26] 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

Connectomes generated 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 for intermediate derivatives as well as the resulting connectomes. QA for the intermediate derivatives in NDMG are shown in Figure 1B. 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 ver-

ify that the properties of the brain highlighted by the diffusion imaging under the implemented tensor model conform to the expected structure. 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 obtained brain graphs. Figure 2 shows the summary statistics plot for an exemplar dataset constructed with the Desikan [16] parcellation. The features of the graphs which NDMG computes are [26], 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. These statistics are tabulated and links to their implementations made available in Appendix B.

4.2 Validation

Results from NDMG have been validated using a metric termed Discriminability [34]. Discriminability, as seen in Equation (1) is a statistically generalized version of Test-ReTest (TRT) reliability:

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

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 the case of pipeline validation, the class label was subject-id and the observations were single sessions of scans. In the case of meganalysis, in which datasets are compared

to one another, the class label was the dataset-id, and the observations were again single sessions of scans. This can trivially be altered for whatever class label is desired for a particular analysis.

A perfect discriminability score is 1, meaning that the nearest-neighbour for any observation is always a different observation from the same class, 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, by ensuring maximum similarity between observations which are believed to be maximally similar. Additionally, this method prevents overfitting the pipeline to covariate-specific signal (i.e. optimizing the pipeline for sex classification).

Computing the chance probability for discriminability can be done by Equation (2). Chance indicates the expected discriminability score if class assignments were made at random. From the law of total probability we can simplify this to the sum over each class $i \in K$ of the square number of elements in the class M_i^2 normalized by the square of the total number of observations N^2 :

$$C = \frac{\sum_{i \in K} M_i^2}{N^2}. \quad (2)$$

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] 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.
- [17] 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.
- [18] 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.
- [19] 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."
- [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. Sripatha 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] 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.
- [27] 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.
- [28] 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.
- [29] 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>
- [30] 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.
- [31] 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>
- [32] 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.
- [33] 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.
- [34] 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.
- [35] 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.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] 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.

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

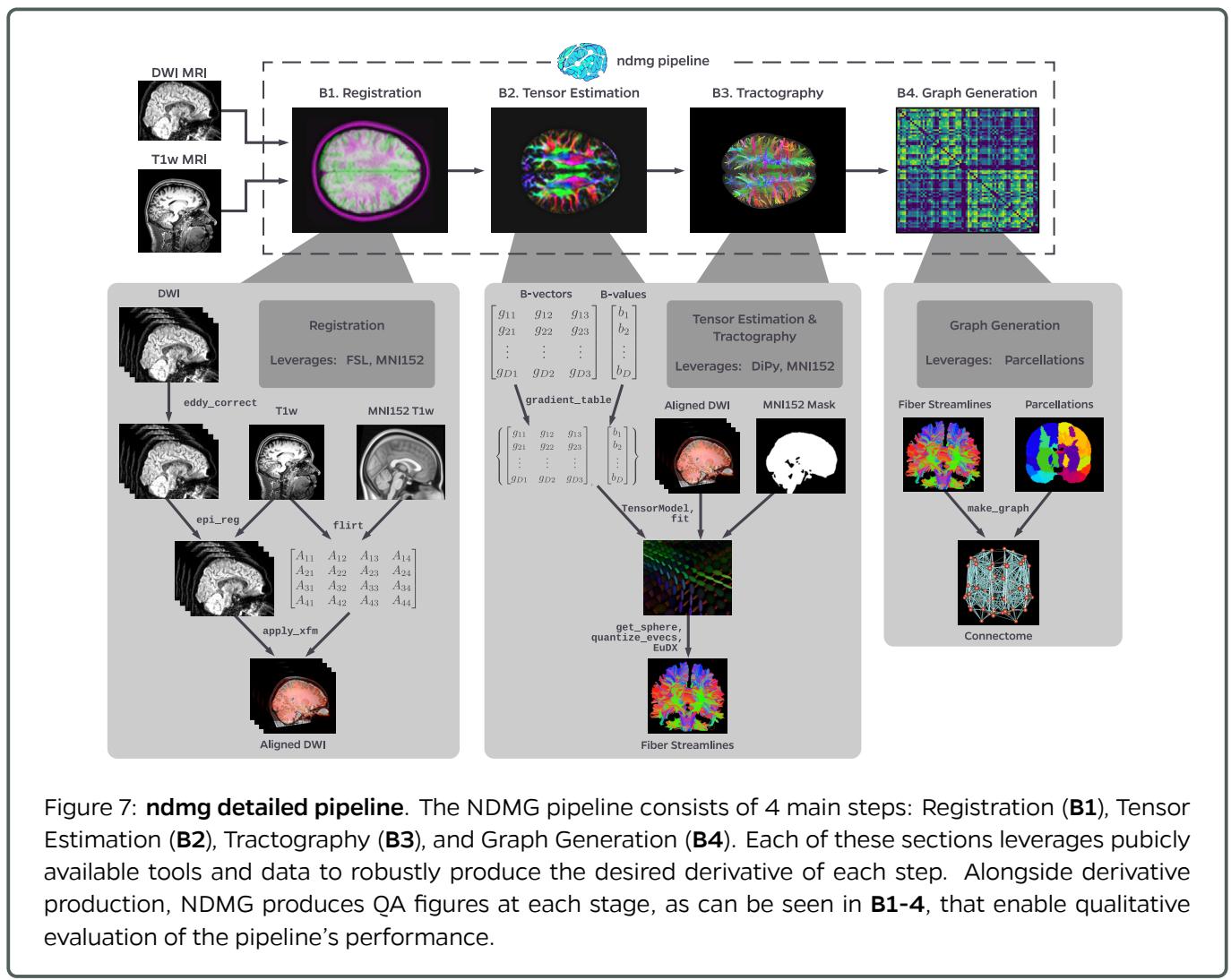


Figure 7: **ndmg detailed pipeline**. The NDMG pipeline consists of 4 main steps: Registration (**B1**), Tensor Estimation (**B2**), Tractography (**B3**), and Graph Generation (**B4**). Each of these sections leverages publicly available tools and data to robustly produce the desired derivative of each 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 website⁵, 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` module⁶. 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>

⁵http://nipy.org/dipy/examples_builtin/reconst_dti.html

⁶https://github.com/neurodata/ndmg/blob/master/ndmg/stats/qa_graphs.py