

# A Comprehensive Cloud Framework for Scalable, Reliable, and Replicable Human Connectome Estimation and Meganalysis

Gregory Kiar<sup>†</sup>, William R. Gray Roncal<sup>†</sup>, Vikram Chandrashekhar, Eric W. Bridgeford,  
Disa Mhembere, Randal Burns, Joshua T. Vogelstein  
Johns Hopkins University & Child Mind Institute

## 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

Neuroimaging methods such as MRI are becoming increasingly more accessible and available for use across both clinical and research populations. With contrast highlighting connective tissue within the brain (i.e. white matter), Diffusion Weighted MRI (DWI) in particular enables the study of structural networks within the brain. As such, DWI data is being collected at an unprecedented rate, and publicly available datasets of both healthy and diseased populations are increasingly commonplace [1–3].

Generating consistently reliable estimates of brain connectivity graphs, or connectomes, has thus far been challenging [4]. Variation in data quality and properties across studies or collections makes the selection of globally robust parameters difficult, and “standard” processing methods operating on multimodal MRI (M3R) data have a history of reporting misleadingly low false positive rates [5], resulting in claims that may not generalize across datasets.

Existing tools such as PANDAS [6] and CMTK [7] tackle the challenge of robustness by enabling users to select hyper-parameters for their dataset. This freedom of course emphasizes the between-dataset differences in connectomes estimated with varying processing parameters, and cannot necessarily be directly compared.

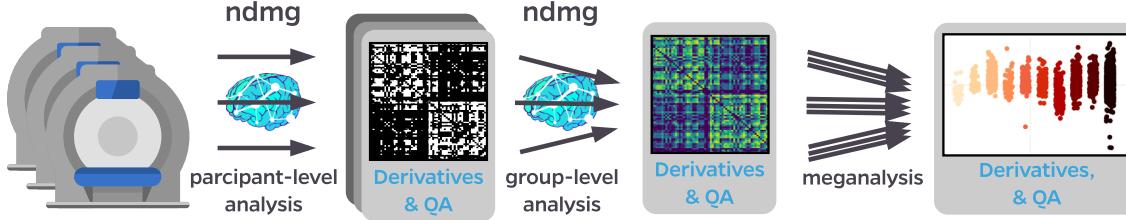
The MRCAP [8] and MIGRAINE [9] pipelines ab-

stract hyper-parameter selection from the user, ensuring that all data is processed identically. While these pipelines produced harmonized results, they have suffered in their ability to be repeatably deployed across hardware and computing environments. This limitation severely impacts the ability of researchers to generate connectomes for large cohorts or collections of data.

We present NDMG, a reliable and robust turn-key solution for structural connectomes estimation that can be deployed at scale either in the cloud or locally for cross-study analysis. Leveraging existing tools such as FSL [10–12], Dipy [13], the MNI152 brain atlas [14] and others, NDMG is a one-click pipeline that lowers the barrier to entry for connectomics. By virtue of harmonized processing, the NDMG pipeline enables scientific “meganalysis” in which data from multiple studies can be pooled, opening the door for more highly generalizable statistical analyses of the structure of the human brain. Connectomes from 2,861 DWI sessions and across 24 brain atlases have been generated with NDMG and have been released to the public.

## 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. The



**Figure 1: ndmg usage workflow.** The NDMG pipeline enables rapidly going from data collection to analysis at multiple levels: participant-, group-, and meganalysis. While participant-level analysis generates connectomes from diffusion MRI data, group-level analysis computes and plots a variety of graph statistics on the derived connectomes, both individually and as a group. The derivatives from all stages of NDMG are computed consistently and reliably across sessions and datasets, enabling the pooling of data for doing “meganalysis.”.

NDMG package and pipeline is open-source and available on Github<sup>1</sup>, and has been configured in several large public-facing platforms, including CBRAIN [15] and OpenNeuro [16], as well as the Amazon Web Services computing cloud [17; 18].

The NDMG pipeline enables three tiers of analysis: participant-level, group-level, and meganalysis. Figure 2 illustrates how NDMG can be used to perform computation for each of these analysis levels.

When data is collected as part of a study, it can be organized in accordance to the BIDS [19; 20] organization standard. Each session of data, consisting of a structural scan (T1w/MPRAGE), a diffusion scan (DWI), and the diffusion parameters files (b-values, b-vectors), can then be used as inputs to generate a connectome. As summarized in Figure 2, NDMG performs registration, tensor estimation, tractography, and graph generation for each session. Connectomes are generated at multiple resolutions based on neuroanatomical parcellations defined in MNI152 [14] space. Participant-level analysis in NDMG takes approximately 1-hour to complete using 1 CPU core and 12 GB of RAM.

Once participant-level analysis has completed for a cohort of data, NDMG group-level analysis can be performed on the generated graphs. At this stage, summary statistics are computed for all graphs generated within the dataset. These statistics are then plotted for additional quality assurance by the scientist, and saved for rigorous quantitative evaluation.

Since NDMG processes all participants and groups identically, it enables the pooling of data across co-

horts, a principle we here refer to as “meganalysis.” This enables scientists to expand the sample represented within their analyses and potentially improve their statistical power and ultimately the scientific impact of their findings.

## 2.1 Subject-Level Analysis

The participant level of NDMG has been developed by leveraging and interfacing existing tools, including FSL [10–12], Dipy [13], the MNI152 atlas [14], and a variety of parcellations defined in the MNI152 space [21–27]. All algorithms which required hyperparameter selection were initially set to the suggested parameters for each tool, and tuned to improve the quality, reliability, and robustness of the results.

Conceptually, this pipeline can be broken up into four key components: Registration, Tensor Estimation, Tractography, and Graph Generation. The NDMG pipeline has been validated through reliability on test-retest datasets. While each of these steps is described here, an in depth look at the specific models and algorithms implemented at each stage contained within Appendix A.

Each stage of NDMG is accompanied by quality assurance (QA) figures, enabling the user to easily detect whether or not the pipeline is producing expected results, both for intermediate derivatives as well as the resulting connectomes. Snapshots of these QA figures are shown throughout Figure 2.

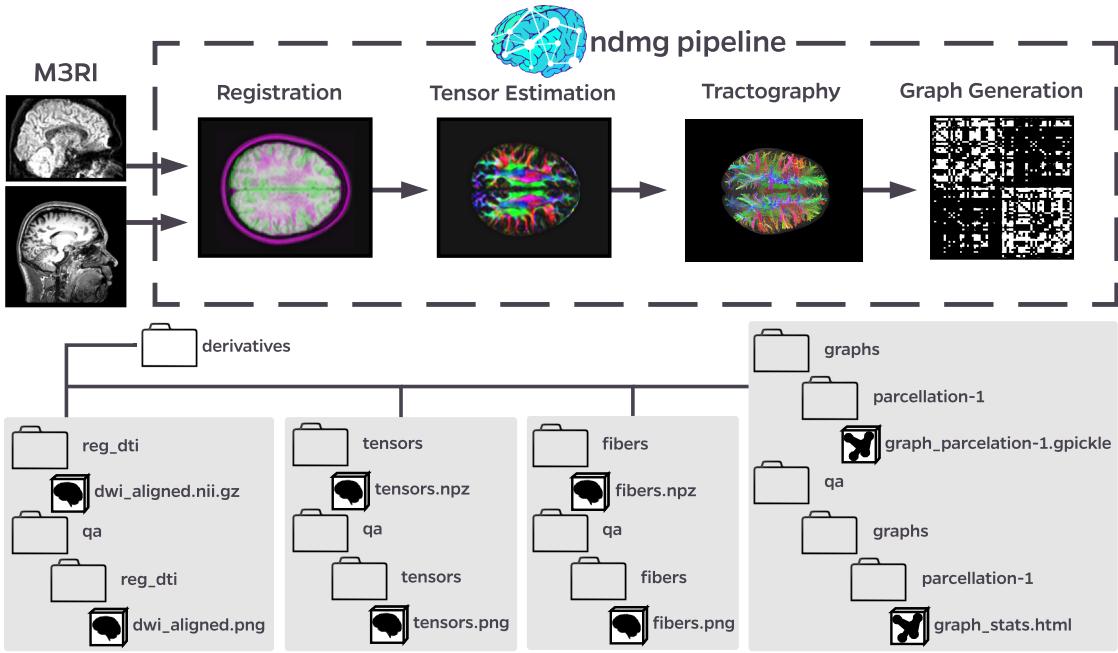


Figure 2: **ndmg participant-level analysis.** The participant-level of the NDMG pipeline transforms raw diffusion weighted MRI data into structural connectomes. Here, NDMG consists of four main steps: Registration, Tensor Estimation, Tractography, and Graph Generation . At each stage, NDMG produces quality assurance figures of the derivatives, as can be seen in as the icon for each stage above.

**Registration** In order for connectomes to be compared, they must be defined in the same space – such as that defined by the MNI152 [14] reference atlas. NDMG leverages FSL [10–12] for a series of linear registrations, the end result of which is the input diffusion weighted volume aligned to the MNI152 atlas. The registration pipeline implemented is “standard” when working with diffusion data and FSL’s tools.

The QA figure produced at this stage is several snapshots of an overlay of the inputted DWI image and the MNI152 template, allowing the user to verify that the brain boundary and higher-level structures of the input image have been closely aligned to the atlas.

**Tensor Estimation** Once the DWI image is aligned, the b-values and b-vectors are interpreted and enable the generation of a voxelwise tensor image from the DWI image stack. A simple 6-component tensor model from the Dipy [13] package is used.

Associated with the resulting tensor image is a fractional anisotropy map of the tensors, which can

be used for analysis and QA. This map allows the user to inspect the global structure and arrangement of the tensors, and whether they conform to the expected structure of the brain, such as a high density of contra-lateral connectivity through the corpus callosum.

**Tractography** Streamlines are generated from the tensors using Dipy’s EuDX [28], a deterministic tractography algorithm closely related to FACT [29]. Each voxel within the brain mask is used as a seed-point in EuDX, and fibers are then pruned based on their length.

NDMG provides a QA plot visualizing 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 fractional anisotropy map generated in the previous step.

**Graph Generation** Connectomes are created by tracing fibers through pre-defined parcellations. The parcellations used have been defined by neu-

roanatomists, such as the HarvardOxford cortical atlas [24], JHU [23], Talairach [25], Desikan [21], and AAL [22] atlases, or generated by segmentation algorithms, such as slab907 [26], Slab1068 [27], CC200 [30] and 16 downsampled (DS) parcellations [31] ranging from 70 to 72,783 nodes. An undirected edge is added to the graph for every pair of regions each a fiber, where the weight of the edge is the cumulative number of fibers between two regions.

An adjacency matrix summarizing the derived connectome is provided as a QA figure and enables users to verify the expected organization of edges and possible hemisphere structure within the graph.

**Validation** Results from NDMG have been validated using a statistically generalized form of test-retest reliability [32], and will be referred to as reliability for the remainder of this manuscript.

Reliability, as seen in Equation (1), describes the

probability that two observations within the same class are more similar to one another than to objects belonging to a different class:

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

In the context of validation in NDMG, this means that the each connectome,  $a_{ij}$ , in a test-retest dataset is first compared to other connectomes belonging to the same subject,  $a_{ij'}$ , and then to all connectomes belonging to other subjects,  $a_{i'j'}$ . A perfect reliability score is 1, meaning that for all observations within the dataset, each connectome is more alike to connectomes from the same subject than to others. The reliability score obtained if a pipeline produced random outputs is summarized in Equation (2), and is a function of the number of classes,  $k$ , the number of elements in each class,  $M_i$ , and the total number of observations,  $N$ :

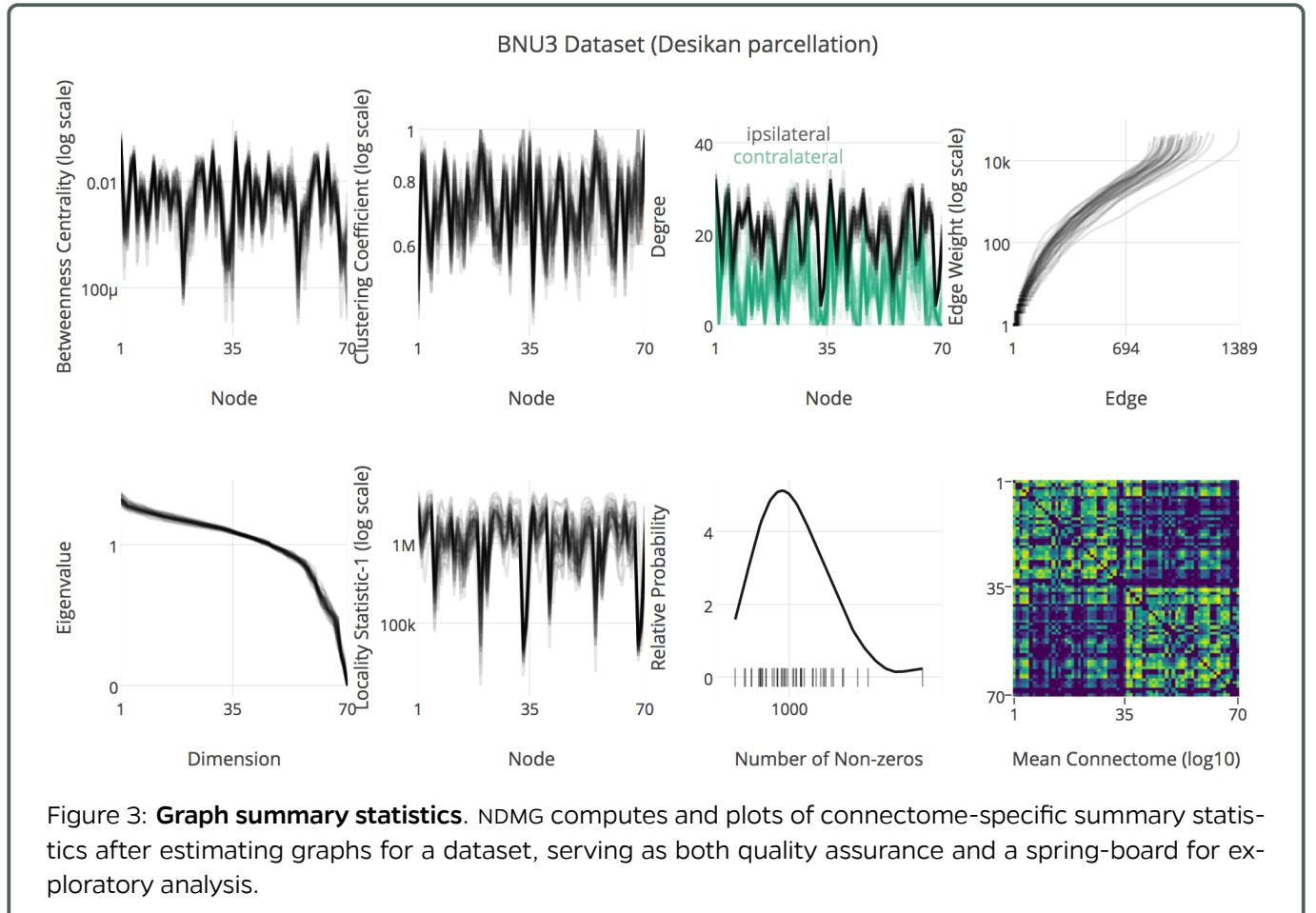


Figure 3: **Graph summary statistics.** NDMG computes and plots of connectome-specific summary statistics after estimating graphs for a dataset, serving as both quality assurance and a spring-board for exploratory analysis.

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

Optimizing NDMG with respect to reliability enables us to minimize the upper-bound on error for any general downstream inference task. Using reliability for optimization also prevents over-fitting to covariate-specific signal (i.e. in contrast to optimizing the pipeline for sex classification).

## 2.2 Group-Level Analysis

Once connectomes have been generated for a dataset, NDMG group-level analysis computes and plots graph summary statistics, as can be seen in Figure 3 for graphs generated with the Desikan [21] parcellation. The summary statistics computed have been chosen as they highlight features of connectomes that are expected to vary based on brain region. The features of the graphs which NDMG computes are [31], clockwise from top-left: between-

ness centrality, clustering coefficient, hemisphere-separated degree sequence, edge weight, eigen values, locality statistic-1, number of non-zero edges, and the mean connectome.

These statistics enable detailed quality assurance of the graphs; for instance, enabling the user to confirm that edge density is higher ipsi-laterally 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.

**Multi-Scale Analysis** As NDMG produces connectomes across a variety of parcellations, it enables multi-scale analysis and flexibility for users. Figure 4 shows the group-level summary statistics of connectomes belonging to same dataset over a range of parcellations. Though 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 4.

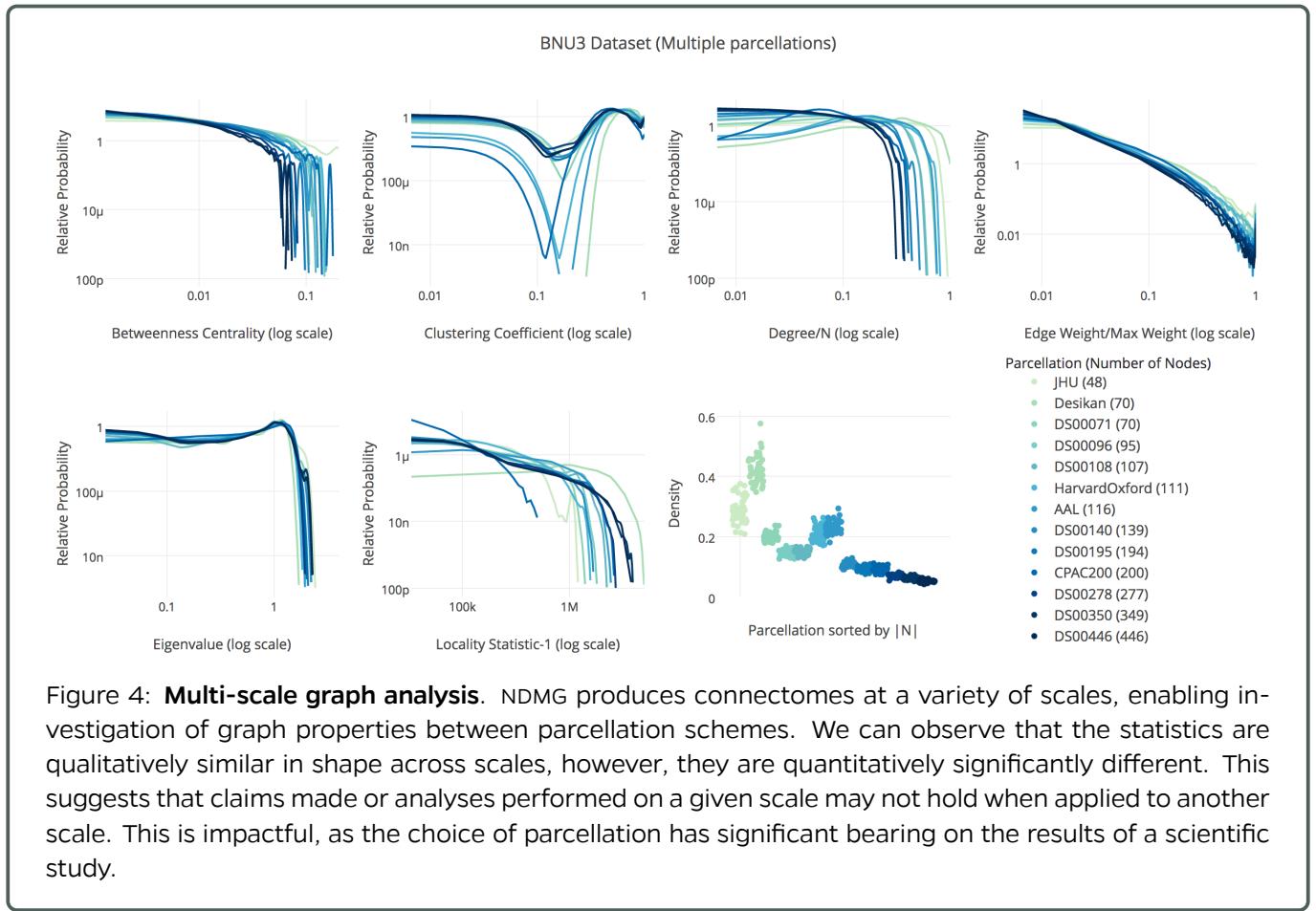


Figure 4: **Multi-scale graph analysis.** NDMG produces connectomes at a variety of scales, enabling investigation of graph properties between parcellation schemes. We can observe that the statistics are qualitatively similar in shape across scales, however, they are quantitatively significantly different. 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.

Table 1: **Processed public M3R datasets.** The derivatives from each dataset processed with NDMG are publicly available at <http://m2g.io>. Test-ReTest (TRT) datasets were evaluated using reliability, where 1 indicates perfectly reliable connectomes. The pooled reliability 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	TRT
BNU1 [2]	Siemens	30	23.0 $\pm$ 2.3	0.53	57	2	114	0.984
BNU3 [2]	Siemens	64	22.5 $\pm$ 2.1	0.50	48	1	47	-
HNU1 [2]	GE	33	24.4 $\pm$ 2.3	0.50	30	10	300	0.993
KKI2009 [33]	Philips	33	31.8 $\pm$ 9.4	0.52	21	2	42	1.0
MRN1313	-	70	-	-	1313	1	1299	-
NKI1 [2]	Siemens	137	34.4 $\pm$ 12.8	0.0	24	2	40	0.984
NKI-ENH [34]	Siemens	137	-	-	198	1	198	-
SWU4 [2]	-	93	20.0 $\pm$ 1.3	0.51	235	2	454	0.884
Templeton114	Siemens	70	21.8 $\pm$ 3.0	0.58	114	1	114	-
Templeton255	Siemens	150	-	-	255	1	253	-
<b>Pooled</b>					<b>2295</b>		<b>2861</b>	<b>0.979</b>

To compare the vertex statistics across scales, their values were normalized and densities were computed and plotted. As we can see in Figure 4, 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 a region-growing function [31]. However, graphs from the smaller DS parcellations look less similar to those from the neuroanatomically defined parcellations (JHU [23], Desikan [21], HarvardOxford [24], CC200 [30]). This suggests that the neuroanatomically defined parcellations are more similar to one another than they are to the downsampled parcellations.

### 2.3 Meganalysis

The NDMG pipeline has been used to process ten public datasets, upon which it demonstrated that it is highly reliable and robust. Table 1 summarizes the 2,861 scans processed with NDMG; each scan processed was used to generate connectomes across each of the 24 parcellations in NDMG, 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`, and can also be accessed through <http://m2g.io>.

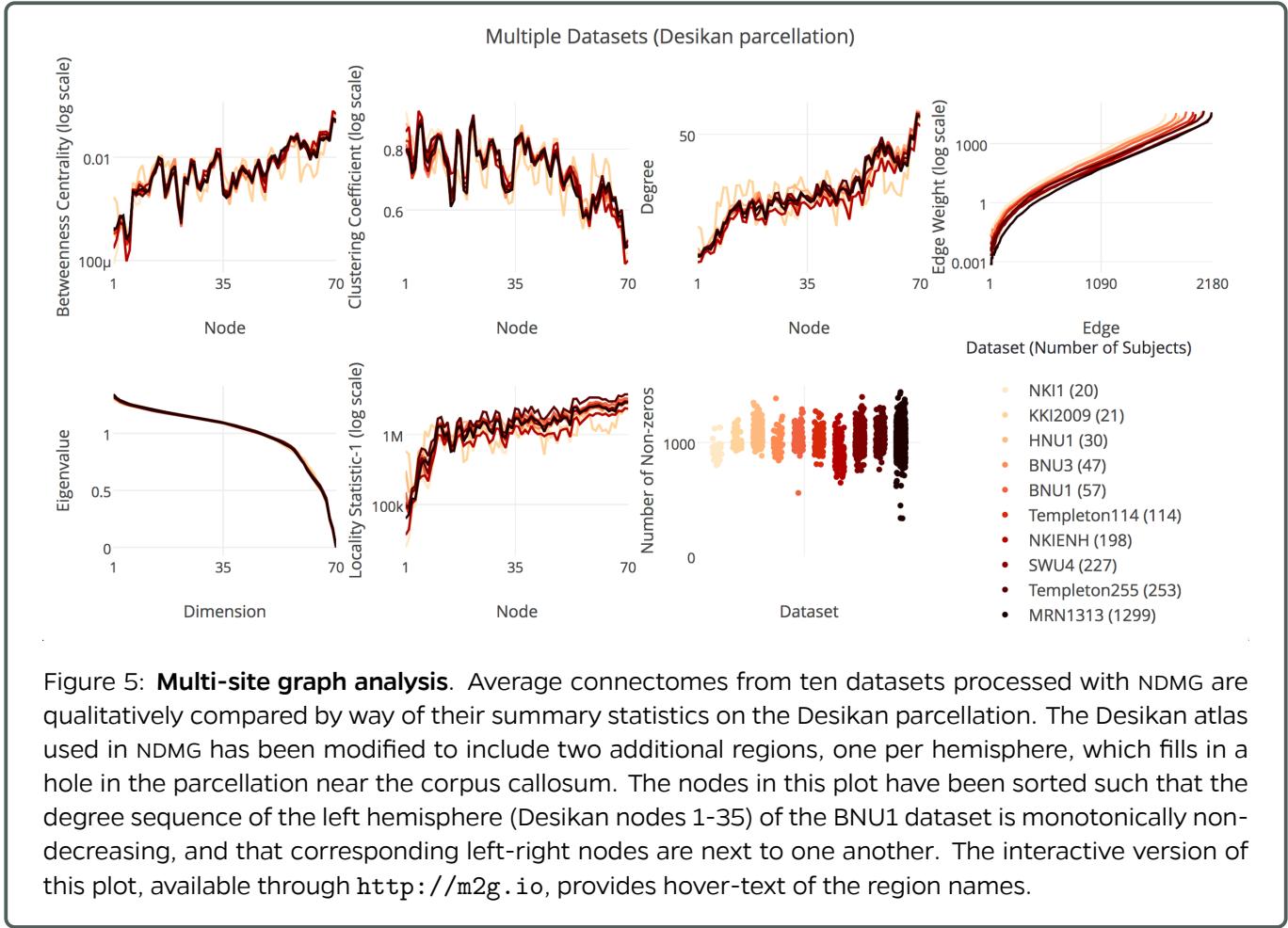
As NDMG harmonized connectome generation, we were able to perform meganalysis towards the evaluation of an essential pillar of scientific discovery: reproducibility of findings across datasets.

### Qualitative Similarity 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 5 shows a variety of uni- and multi-variate statistics of the average connectome from each of the datasets enumerated in Table 1 using the Desikan parcellation.

Each dataset largely appears to have similar trends across each of the statistics shown. Though the KKI2009 [33] 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 others. A likely cause of this difference is the scanner manufacturer, as the KKI2009 dataset was acquired with a Phillips scanner while all of the other datasets were acquired with either Siemens scanners, or in one case a GE scanner. While the remaining datasets have variation between them, they appear to trend more closely together.

Figure 6 shows the mean connectome computed from each dataset, as well as the weighted mega-mean and mega-standard deviation connectomes combining all datasets. As with the multivariate statistics, dataset means have very similar structures and intensity profiles, with minor noticeable differences predominantly visible in the contra-lateral (off-diagonal) blocks. Considering 2,861 sessions, we



**Figure 5: 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.

believe this mega-mean connectome to have the highest- $N$  of any mean connectome computed to date.

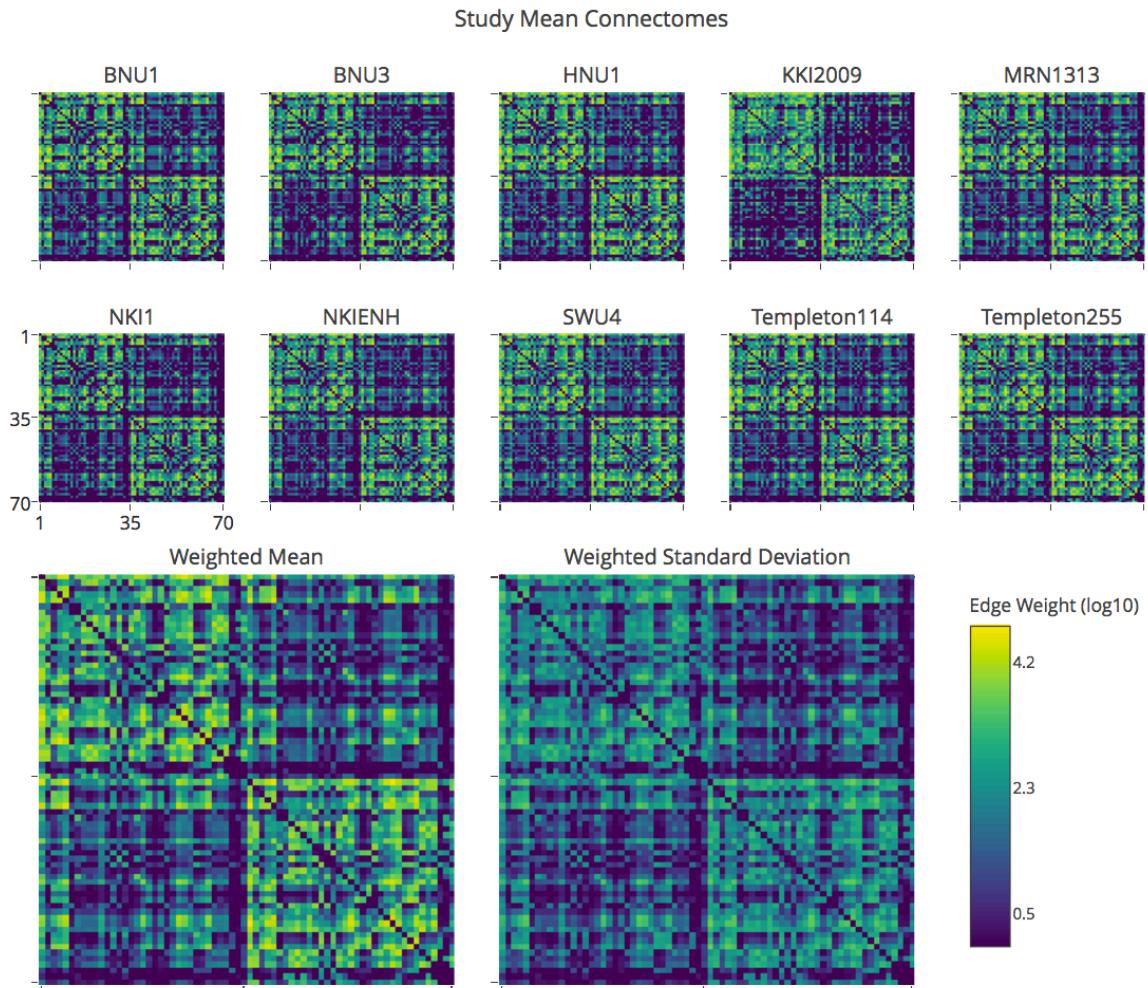
Several properties of the connectomes are present across all datasets, such as that they contain a high number of ipsi-lateral connections relative to a lower number of contra-lateral connections. The standard deviation connectome also shows that the ipsi-lateral connectivity is also more highly variable, suggesting that there exists a relationship between connection density and variability. The relationship between connection density and variability is consistent to that of an Erdős-Rényi random graph, a model previously used in connectomics [35]. 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<sup>2</sup>.

Figures 5 and 6 show similar structure and prop-

erties 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.

## 2.4 Quantitative Difference Across Datasets

Though qualitative analysis allows us to do basic quality assurance and verify that the structure and 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, leaving the generalizability of scientific findings up for debate. Through Reliability [32] we have been able to quantify the significance of batch effects in DWI data. We can use reliability to evaluate the similarity corresponding to



**Figure 6: 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.

connectomes within and across dataset; here, unlike during the pipeline optimization shown in Table 1, a low reliability is desirable for this meganalysis. Chance performance, indicating that connectomes were all equally similar regardless of dataset, can be calculated using Equation (2).

The reliability across the non-TRT selection of scans (2,295 sessions rather than 2,861) was computed, for which chance is 0.363. The computed re-

liability was 0.632, suggesting considerable dataset-specific signal is present at a significance of  $p < 0.0001$  when performing a permutation test. The differences between 20 randomly-selected graphs in each dataset is summarized in Figure 7.

As there is an obvious visual difference between KKI2009, acquired on a Philips scanner, and the other datasets, acquired predominantly on Siemens scanners and one using GE, the reliability was also

### Multi-Study Discriminability Using Desikan Atlas

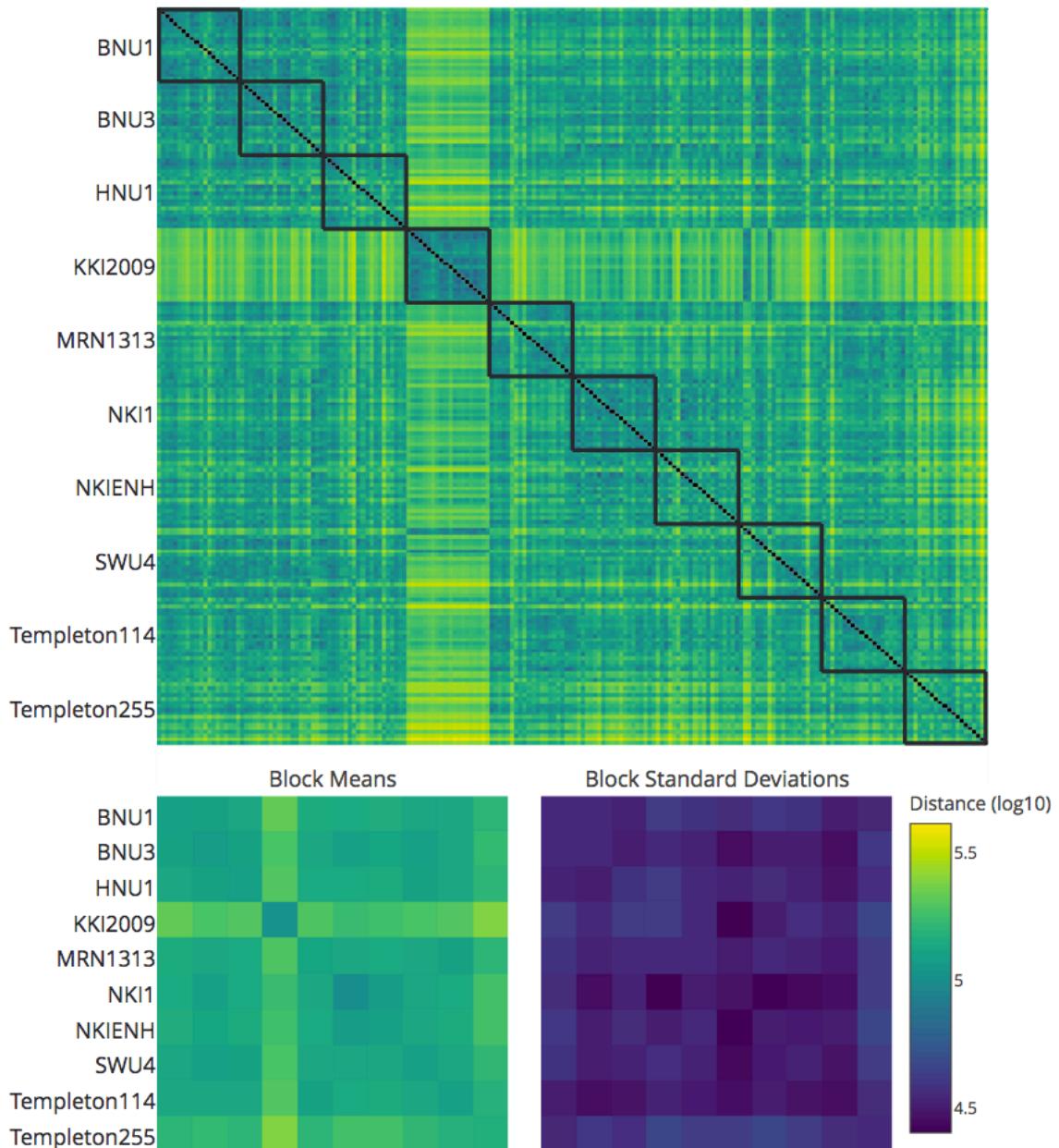


Figure 7: **Prevalence of batch effects.** Discriminability was computed across the ten processed datasets using the dataset-id as the class label. If no significant difference between datasets exists, the discriminability score would not be significantly different from chance, a score of 0.363. Here, 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.

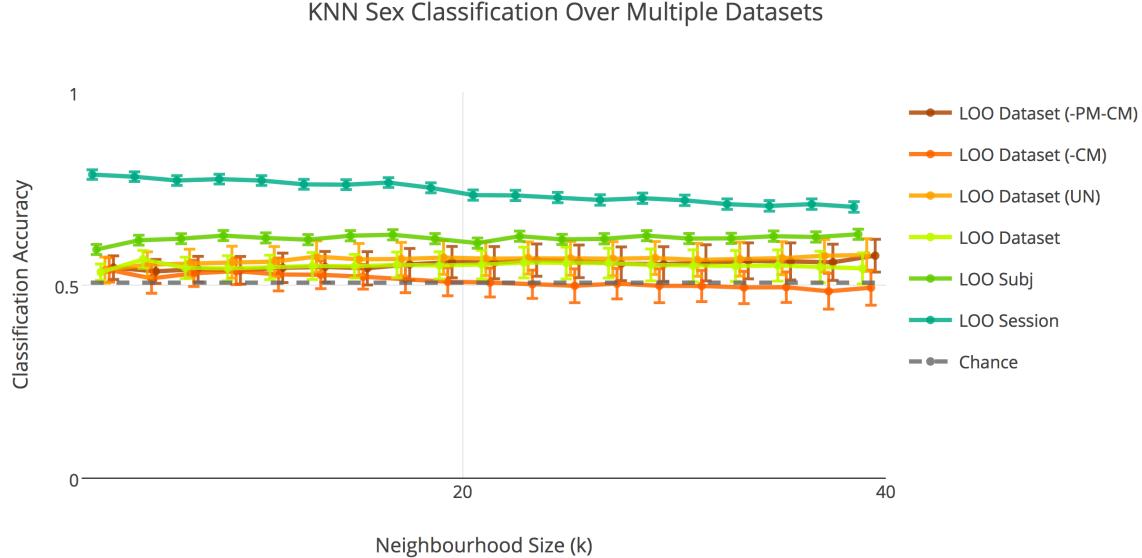


Figure 8: **Connectome Sex Classification.** Using K-Nearest Neighbours classification, several cross-validation attempts were made on “raw” and further post-processed connectomes to improve accuracy in sex classification. The cross-validation methods were leave-out-one (LOO): session, subject, dataset, dataset with unit normalization, dataset with subtracting cohort means, and dataset with subtracting cohort and population means. If the batch effects were insignificant, then we would expect all methods to have equivalent performance except for LOO session, which serves as an upper bound. We notice that LOO subject performs the greatest of these approaches, indicating that there is considerable batch effect, and that post-processing attempts did not effectively mitigate these differences.

computed when omitting the KKI2009 dataset which resulted in a score of 0.626, a value still significant with  $p < 0.0001$ . Also removing the sole dataset using a GE scanner, HNU1, we compute the reliability score considering only data from Siemens sites to be 0.627, which is also significant at  $p < 0.0001$ .

Figure 8 demonstrates the prevalence of this effect in a naive sex classification task. Several post-hoc normalization attempts were made to reduce the impact of batch effects, such as unit-normalizing connectomes, subtracting the cohort-mean, and subtracting the cohort-mean and then the population-mean. Classification attempts based on all leave-out-one (LOO) dataset cross-validation methods are within chance performance, where LOO subject performance is consistently higher. This suggests that connectome normalization is insufficient for addressing batch effects.

### 3 Discussion

The NDMG pipeline is a highly-reliable tool for structural connectome estimation with a low barrier to entry for neuroscientists, and has shown that it is capable of producing meaningful and consistent brain-graphs across scales and datasets. NDMG abstracts hyper-parameter selection from users by providing a default setting that has been used to demonstrate robustness across a variety of datasets, achieving matched or improved reliability when performing either single- or multi-dataset analysis compared to alternatives [4; 36]. Though this generalizability 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 and makes comparing graphs trivial across studies, avoiding overfitting of the pipeline to a specific dataset.

Though NDMG has been optimized with respect to reliability, an exhaustive 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 reliability would improve the reliability of NDMG even further. Though this 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 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 when deterministic or probabilistic tractography is a more reliable when estimating connectivity, and by how much.

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 qualitatively similar across cohorts, these batch effects are significant when performing quantitative analysis. A consequence of this is that clinical assays or scientific findings made on one group of data may not generalize across datasets, even in the case of harmonized processing.

Multi-site predictive tasks have been performed considering fMRI-derived connectomes from the ABIDE dataset [37], and performing similar analyses upon a multi-dataset collection of DWI-derived connectomes is an exciting avenue for future exploration. Additionally, a variety of studies propose methods for data harmonization upon either minimally pre-processed or raw M3R data [38–40] which could be explored within the context of the NDMG pipeline.

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 covariate-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 computing platforms such as OpenNeuro, CBRAIN, and Amazon makes it

an accessible choice for neuroscientists who are limited in their in-house compute infrastructure or experience deploying pipelines on high performance computing clusters.

Through the use and development of NDMG we have simultaneously lowered the barrier to entry for performing connectomics research and demonstrated a statistically-principled method for evaluating the reliability of pipelines and data. NDMG further democratizes human connectomics research, enabling users to produce statistically-validated connectomes at an unprecedented scale with ease and perform meganalysis across broad cohorts of data.

## Affiliation Information

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

<sup>1</sup>Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA.

<sup>2</sup>Center for Imaging Science, Johns Hopkins University, Baltimore, MD, USA.

<sup>3</sup>Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA.

<sup>4</sup>Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA.

<sup>5</sup>Institute for Computational Medicine, Johns Hopkins University, Baltimore, MD, USA.

<sup>6</sup>Child Mind Institute, New York, NY, 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] 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.
- [2] X.-N. Zuo, J. S. Anderson, P. Bellec, R. M. Birn, B. B. Biswal, J. Blautzik, J. C. Breitner, R. L. Buckner, V. D. Calhoun, F. X. Castellanos et al., “An open science resource for establishing reliability and reproducibility in functional connectomics,” *Scientific data*, vol. 1, p. 140049, 2014.
- [3] 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.

- [4] D. Petrov, A. Ivanov, J. Faskowitz, B. Gutman, D. Moyer, J. Villalon, N. Jahanshad, and P. Thompson, "Evaluating 35 methods to generate structural connectomes using pairwise classification," *arXiv preprint arXiv:1706.06031*, 2017.
- [5] 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.
- [6] 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>
- [7] 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>
- [8] W. R. Gray et al., "Magnetic Resonance Connectome Automated Pipeline," p. 5, nov 2011. [Online]. Available: <http://arxiv.org/abs/1111.2660>
- [9] W. Gray Roncal et al., "MIGRAINE: MRI Graph Reliability Analysis and Inference for Connectomics," *Global Conference on Signal and Information Processing*, 2013.
- [10] 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>
- [11] 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>
- [12] 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>
- [13] 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.
- [14] 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.
- [15] 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.
- [16] R. A. Poldrack, D. M. Barch, J. Mitchell, T. Wager, A. D. Wagner, J. T. Devlin, C. Cumba, O. Koyejo, and M. Milham, "Toward open sharing of task-based fmri data: the openfmri project," *Frontiers in neuroinformatics*, vol. 7, p. 12, 2013.
- [17] 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.
- [18] 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 ndmrg pipeline (sic:ndmrg)," 2017. [Online]. Available: <https://doi.org/10.5524/100285>
- [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] 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>
- [21] 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.
- [22] 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.
- [23] K. Oishi et al., *MRI atlas of human white matter*. Academic Press, 2010.
- [24] 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.
- [25] J. Lancaster, "The Talairach Daemon, a database server for Talairach atlas labels," *NeuroImage*, 1997.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] 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.

- [34] 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.
- [35] M. T. Gastner and G. Ódor, "The topology of large open connectome networks for the human brain," *Scientific reports*, vol. 6, p. 27249, 2016.
- [36] J. P. Owen, E. Ziv, P. Bukshpun, N. Pojman, M. Wakahiro, J. I. Berman, T. P. Roberts, E. J. Friedman, E. H. Sherr, and P. Mukherjee, "Test-retest reliability of computational network measurements derived from the structural connectome of the human brain," *Brain connectivity*, vol. 3, no. 2, pp. 160–176, 2013.
- [37] A. Abraham, M. P. Milham, A. Di Martino, R. C. Craddock, D. Samaras, B. Thirion, and G. Varoquaux, "Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example," *NeuroImage*, vol. 147, pp. 736–745, 2017.
- [38] 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.
- [39] 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.
- [40] 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.

## 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 9B1, 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

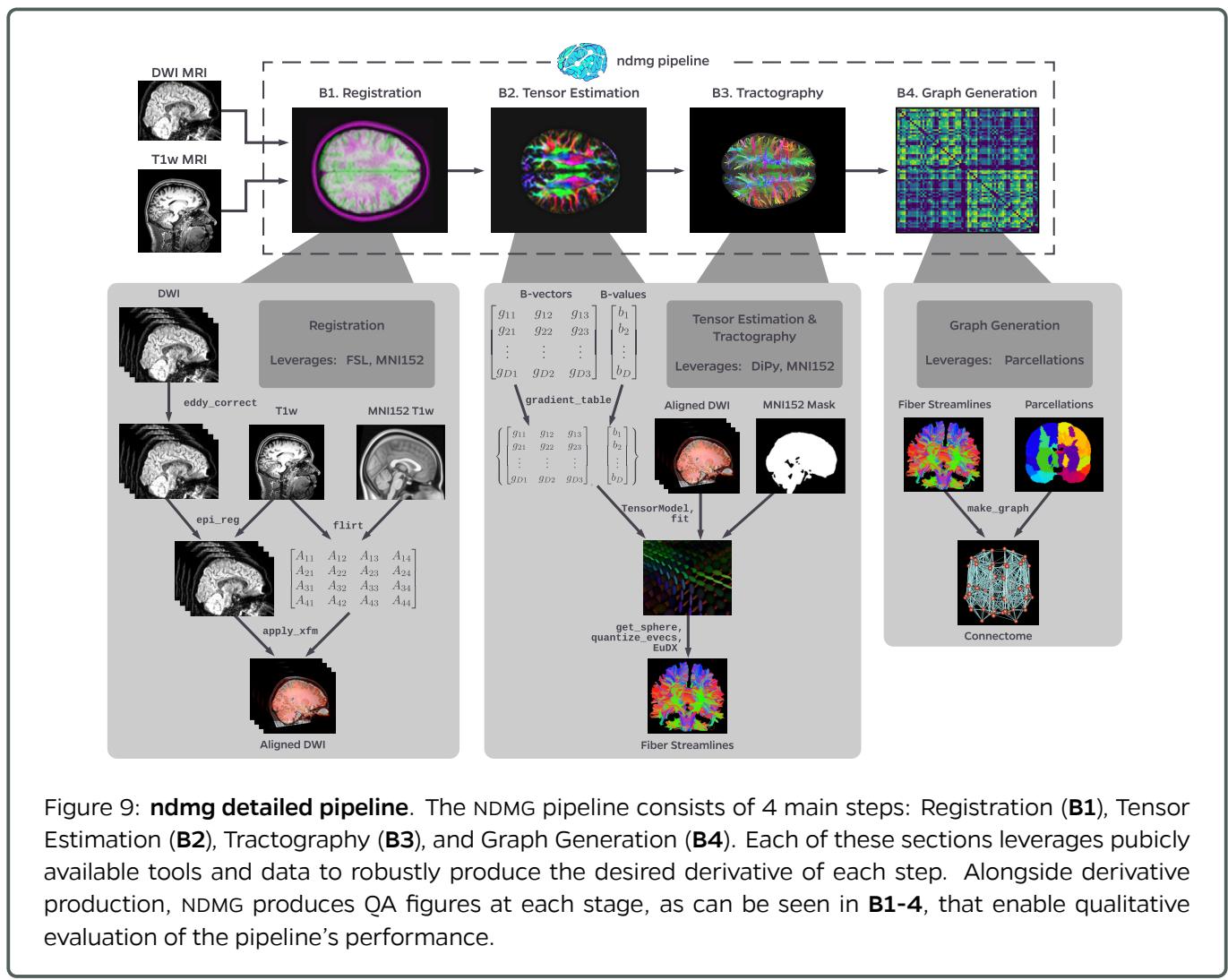


Figure 9: **ndmg detailed pipeline**. The NDMG pipeline consists of 4 main steps: Registration (**B1**), Tensor Estimation (**B2**), Tractography (**B3**), and Graph Generation (**B4**). Each of these sections leverages publicly available tools and data to robustly produce the desired derivative of each 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 website<sup>3</sup>, 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` module<sup>4</sup>. 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

<sup>1</sup><https://github.com/neurodata/ndmg>

<sup>2</sup><https://github.com/neurodata/ndmg-paper/tree/master/code/meanconnectome>

<sup>3</sup>[http://nipy.org/dipy/examples\\_built/reconst\\_dti.html](http://nipy.org/dipy/examples_built/reconst_dti.html)

<sup>4</sup>[https://github.com/neurodata/ndmg/blob/master/ndmg/stats/qa\\_graphs.py](https://github.com/neurodata/ndmg/blob/master/ndmg/stats/qa_graphs.py)