Optimized implementations of voxel-wise degree centrality and local functional connectivity density mapping in AFNI

Project URL: http://github.com/ccraddock/afni

R. Cameron Craddock¹,²* and Daniel J. Clark²

1 Introduction
Degree centrality (DC) [1] and local functional connectivity density (lFCD) [2] are statistics calculated from brain connectivity graphs that measure how important a brain region is to the graph. DC (a.k.a. global functional connectivity density [2]) is calculated as the number of connections a region has with the rest of the brain (binary DC), or the sum of weights for those connections (weighted DC) [1]. lFCD was developed to be a surrogate measure of DC that is faster to calculate by restricting its computation to regions that are spatially adjacent [2]. Although both of these measures are popular for investigating inter-individual variation in brain connectivity, efficient neuroimaging tools for computing them are scarce. The goal of this Brainhack project was to contribute optimized implementations of these algorithms to the widely used, open source, AFNI software package [3].

2 Approach
Tools for calculating DC (3dDegreeCentrality) and lFCD (3dLFCD) were implemented by modifying the C source code of AFNI’s 3dAutoTcorrelate tool. 3dAutoTcorrelate calculates the voxel × voxel correlation matrix for a dataset and includes most of the functionality we require, including support for OpenMP [4] multithreading to improve calculation time, the ability to restrict the calculation using a user-supplied or auto-calculated mask, and support for both Pearson’s and Spearman correlation.

3dDegreeCentrality: Calculating DC is straightforward and is quick when a correlation threshold or is used. In this scenario, each of the \( \frac{5 \times N_{\text{vox}} \times (N_{\text{vox}} - 1)}{2} \) unique correlations are calculated, and if they exceed a user specified threshold (default threshold = 0.0) the binary and weighted DC value for each of the voxels involved in the calculation are incremented. The procedure is more tricky if sparsity thresholding is used, where the top \( P\% \) of connections are included in the calculation. This requires that a large number of the connections be retained and ranked - consuming substantial memory and computation. We optimize this procedure with a histogram and adaptive thresholding. If a correlation exceeds threshold it is added to a 50-bin histogram (array of linked lists). If it is determined that the lowest bin of the histogram is not needed to meet the sparsity goal, the threshold is increased by the bin-width and the bin is discarded. Once all of the correlations have been calculated, the histogram is traversed from high to low, incorporating connections into binary and weighted DC until a bin is encountered that would push the number of retained connections over the desired sparsity. This bin’s values are sorted into a 100-bin histogram that is likewise traversed until the sparsity threshold is met or exceeded. The number of bins in the histograms affects the computation time and determine the precision with which ties between voxel values are broken. A greater number of bins allow the sparsity threshold to be determined more precisely but will take longer to converge. Fewer bins will result in faster computation but will increase the tendency of the algorithm to return more voxels than requested. The chosen parameters enable ties to be broken with a precision of \( 1.0/(50 \times 100) \), which in our experience
offers quick convergence and a good approximation of the desired sparsity.

3dLFC: IFCD was calculated using a region growing algorithm in which face-, side-, and corner-touching voxels are iteratively added to the cluster if their correlation with the target voxel exceeds a threshold (default threshold = 0.0). Although IFCD was originally defined as the number of voxels locally connected to the target, we also included a weighted version.

Validation: Outputs from the newly developed tools were benchmarked to Python implementations of these measures from the Configurable Pipeline for the Analysis of Connectomes (C-PAC) [5] using in the publically shared Intrinsic Brain Activity Test-Retest (IBATRT) dataset from the Consortium for Reliability and Reproducibility[6].

3 Results
AFNI tools were developed for calculating IFCD and DC from functional neuroimaging data and have been submitted for inclusion into AFNI. IFCD and DC maps from the test dataset (illustrated in Fig. 1) are highly similar to those calculated using C-PAC (spatial concordance correlation [7] $\rho \geq 0.99$) but required substantially less time and memory (see Table 1).

4 Conclusions
Optimized versions of IFCD and DC achieved 4× to 10× decreases in computation time compared to C-PAC’s Python implementation and decreased the memory footprint to less than 1 gigabyte. These improvements will dramatically increase the size of Connectome analyses that can be performed using conventional workstations. Making this implementation available through AFNI ensures that it will be available to a wide range of neuroimaging researchers who do not have the wherewithal to implement these algorithms themselves.

Availability of Supporting Data
More information about this project can be found at: http://github.com/ccraddock/afni. Further data and files supporting this project are hosted in the GigaScience repository doi:10.5524/100219.

Competing interests
None

Author’s contributions
RCC and DJC wrote the software, DJC performed tests, and DJC and RCC wrote the report.

Acknowledgements
The authors would like to thank the organizers and attendees of Brainhack MX and the developers of AFNI. This project was funded in part by a Educational Research Grant from Amazon Web Services.

Author details
1. Computational Neuroimaging Lab, Center for Biomedical Imaging and Neuro modulation, Nathan Kline Institute for Psychiatric Research, Orangeburg, 140 Old Orangeburg Rd, 10962, New York, USA. 2. Center for the Developing Brain, Child Mind Institute, New York, 445 Park Ave, 10022, New York, USA.

References
1. Rubinov, M., Sporns, O.: Complex network measures of brain connectivity: uses and interpretations. Neuroimage 52(3), 1059–1069 (2010). doi:10.1016/j.neuroimage.2009.10.003
2. Tomasi, D., Volkow, N.D.: Functional connectivity density mapping. Proc. Natl. Acad. Sci. U.S.A. 107(21), 9885–9890 (2010). doi:10.1073/pnas.1001414107
3. Cox, R.W.: AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29(3), 162–173 (1996)
4. Dagum, L., Menon, R.: Openmp: an industry standard api for shared-memory programming. Computational Science & Engineering, IEEE 5(1), 46–55 (1998)
5. Craddock, C., Sikka, S., Cheung, B., Khanuja, R., Ghosh, S.S., Yan, C., Li, Q., Lurie, D., Vogelstein, J., Burns, R., Colcombe, S., Mennes, M., Kelly, C., Di Martino, A., Castellanos, F.X., Milham, M.: Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (c-pac). Frontiers in Neuroinformatics (42) (2013). doi:10.3389/conf.fninf.2013.09.00042
6. Zuo, X.-N., Anderson, J.S., Bellec, P., Birn, R.M., Biswal, B.B., Blauzitk, J., Breitner, J.C.S., Buckner, R.L., Calhoun, V.D., Castellanos, F.X., Chen, A., Chen, B., Chen, J., Chen, X., Colcombe, S.J., Courtney, W., Craddock, R.C., Martino, A.D., Dong, H.-M., Fu, X., Gong, Q., Gorgolewski, K.J., Han, Y., He, Y., He, Y., Ho, E., Holmes, A., Hou, X.-H., Huckins, J., Jiang, T., Jiang, Y., Kelley, W., Kelly, C., King, M., LaConte, S.M., Lainhart, J.E., Lei, X., Li, H.J., Li, K., Li, K., Lin, Q., Liu, D., Liu, J., Liu, X., Liu, Y., Lu, G., Lu, J., Luna, B., Luo, J., Lurie, D., Mao, Y., Margulies, D.S., Mayer, A.R., Meindl, T., Meyerand, M.E., Nan, W., Nielsen, J.A., O’Connor, D., Paulsen, D., Prabhakaran, V., Qi, Z., Qiu, J., Shao, C., Shehzad, Z., Tang, W., Villinger, A., Wang, H., Wang, K., Wei, D., Wei, G.-X., Weng, X.-C., Wu, X., Xu, T., Yang, N., Yang, Z., Zang, Y.-F., Zhang, L., Zhang, Q., Zhang, Z., Zhang, Z., Zhao, K., Zou, Y., Zhu, X.-T., Milham, M.P.: An open science resource for establishing reliability and reproducibility in functional connectomics. Scientific Data 1, 140049 (2014). doi:10.1038/sdata.2014.49
Table 1 Comparison of the time and memory required by the Python and C implementations to calculate DC (sparsity and correlation threshold) and IFCD on the first resting state scan of the first scanning session for all 36 participants’ data in the IBATRT dataset. Values are averaged across the 36 datasets and presented along with standard deviations in parenthesis. Impl: Implementation, Thr: Number of threads used to process a single dataset, Mem: average (standard deviation) memory in gigabytes used to process a single dataset, \( T_D \): the average (standard deviation) time in seconds to process a dataset. These statistics were collected on a C3.xlarge Amazon Web Services Elastic Compute Cloud node with 8 hyperthreads and 15 GB of RAM.

| Impl | Thr | Mem (GB) | \( T_D \) (s) | Mem (GB) | \( T_D \) (s) | Mem (GB) | \( T_D \) (s) |
|------|-----|----------|----------------|----------|----------------|----------|----------------|
| Python | 1 | 2.17 (0.078) | 67.7 (3.90) | 5.62 (0.176) | 342.2 (12.25) | 2.16 (0.082) | 88.3 (6.40) |
| C | 1 | 0.84 (0.003) | 62.6 (9.23) | 0.85 (0.002) | 86.3 (13.83) | 0.86 (0.003) | 8.8 (1.27) |
| C | 2 | 0.86 (0.002) | 39.0 (4.62) | 0.86 (0.003) | 38.2 (0.55) | 0.86 (0.003) | 5.1 (0.25) |
| C | 4 | 0.86 (0.003) | 18.2 (1.93) | 0.87 (0.003) | 19.0 (0.45) | 0.87 (0.003) | 4.3 (0.23) |
| C | 8 | 0.87 (0.002) | 11.2 (0.25) | 0.87 (0.000) | 11.3 (0.31) | 0.87 (0.000) | 4.1 (0.15) |

7. Lange, N., Strother, S.C., Anderson, J.R., Nielsen, F.A., Holmes, A.P., Kolenda, T., Savoy, R., Hansen, L.K.: Plurality and resemblance in fMRI data analysis. Neuroimage 10(3 Pt 1), 282–303 (1999)