

UNIVERSITY OF CALGARY | Program for Undergraduate Research Experience (PURE)

---

## FINAL ASSIGNMENT

*“CORTICAL FUNCTIONAL NETWORK ASSIGNMENT USING  
TEMPLATE PRIORS”*

**Rylan Marianchuk**

Project Duration: May 1 – Aug 31<sup>st</sup>, 16 weeks

Date of Submission: September 20, 2021

Supervisor Name: Dr. Signe Bray

Supervisor Signature:

X Signe Bray

# 1 Abstract

“ACTS GET PERFORMED; THE PERSON IS A PERFORMER OF ACTS.”

- Martin Heidegger, *Being and Time*

Variance exists in the proclivities for humans to act. This is manifest by the disparate array of human endeavor and its specialization, from art, music, science, and physical sport. The trait-like behaviors that give rise to this diversity are known to have a biological basis which neuroscience looks to characterize. More specifically, this basis is thought to be found in the functional network organization of the human brain. To know precisely how these networks vary across or within subjects, there must be an assignment method that takes minimally preprocessed magnetic resonance images of the brain *in vivo*, where the human sits in the scanner, and derives a model of their brain that outlines the distribution of well known networks across their cortex. This report attempts to explicate the specific methodological steps required to achieve this using a previously proposed technique that matches individual functional connectivity with an externally provided template that contains network labeled regions of high group consensus. The task requires onerous computation, and was developed using Python along with a Cython binding to allow for C++ multithreading speedup. Code was made freely available on github for future use across the neuroimaging community.

## 1.1 Index of terms

**Cython** : a binding produced for the multi-purpose programming language Python, where a function call may be reduced to C code for enhanced execution speed.

**Multithreading** : A computing paradigm where code blocks are run in tandem using ‘threads’ – the most basic unit of code execution. Even with initial overhead scaffolding to distribute necessary resources to all threads, execution speed up is usually observed over single threaded code if implemented prudently.

**Voxel** : A volumetric pixel. The most basic container of data in a volume or three dimensional array. A voxel holds an instance value of a datatype, like a pixel does in a two dimensional image. Every voxel holds a unique integer triple as an index: usually denoted by  $i, j, k$ .

**MNI space** : (M)ontreal (N)eurological (I)nstitute three dimensional coordinate mapping of a standardized brain template.

**Network** : a mathematical object consisting of nodes and associations between them, analogous to the function and structure of the human brain. Structurally, nuclei of neural bodies are regionally clustered (nodes) with their axons stretching (connections) to other regions at various densities. Functionally, statistical dependencies in co-activations between nodal regions are considered edges, or associations, or connections.

**Dice Coefficient** : A float value measuring the degree of overlap, or intersection, between two finite sets.

**ROIs** (Regions of interest) : coordinates defined in MNI space that host a notable characteristic within its spatial vicinity.

**BOLD time course** : (B)lood (O)xygen (L)evel (D)ependent measurement fluctuations in neural tissue over a scan time for a collection of small square cubic spaces in the brain (voxels).

**fMRI** : (f)unctional (M)agnetic (R)esonance (I)maging uses magnetic gradients to orient hydrogen nuclei so that it can be morphed into an image characterizing its tissue structure by measuring a BOLD response. It is ‘functional’ because the image is acquired over contiguous time, observing the brain in function.

**Gray Matter** : neural tissue that comprises mainly of the soma, or body of neurons, in comparison to the axonal wrapping of myelin which manifests white color (coined white matter). The exterior of the human cerebral cortex hosts most prominently gray matter.

**Gray Matter Mask** : a volume or three dimensional array that contains binary data (yes/no, 0/1) indicating whether a given voxel is a member of the brain’s gray matter.

**Functional Connectivity** : the measure of temporally correlated neurophysiological events is disparate brain regions.

**Correlated** : a quality or measure denoting the degree to which two variables are similar; that is a change in one variable sees a matching change (either decrease or increase) in the second variable. Given  $n$  points in  $\mathbb{R}^2$ , the pearson coefficient is given by  $r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$

$\mathcal{T}$  : For a given voxel,  $\mathcal{T}$  is a vector holding the (T)op correlations (sorted) across all gray matter voxels with the one given.

$\mathcal{R}$  : Symmetric Voxel-wise correlation matrix.  $\mathcal{T}$  may assume any arbitrary row or column.

$\mathcal{A}$  : (A)ssignment function, required to compute across all gray matter voxels for the purposes of the project.

## 2 Research Findings and Synthesis

### 2.1 Introduction

Humans have very similar absolute sized sensory and motor networks to primates [1]. Yet we evidently know a vast gap exists between the human and primate brain. A major asymmetry is thought to exist in corticocortical functional coupling, with humans possessing exponentially more complex network associations across the cortex. Activity across the cortex can be observed in BOLD-fMRI without any external stimuli, suggesting resting state networks exists as a part of the bespoke corticocortical functional coupling. This aligns with the conception of the brain as an amalgam of networks [2]. Modern neuroscience still has not finalized the association between network variance and behaviour variance, although there are suspicions of its validity [3]. It is therefore of paramount importance to lay rigorous methodological steps to obtain a map, or measurement, of an individuals networks across their cortex. The goal of this report is to achieve careful elucidation of how a semi-technical person in computation could implement such a technique, previously introduced at [4].

### 2.2 Methods

Emphasis on methods in this report is a result of the nature of the project and my academic background in computer science. A computation task lies within a tension of execution speed and development speed. It is rare for significant tasks to achieve both minimal execution time and minimal development time. The programming language Python [5] is ubiquitous for its readability and speedy development, but that necessarily comes with slow execution for large scientific datasets. An amelioration of this is to use Cython [6], a C/C++ binding for Python code which allows a function to be executed closer to the hardware (C level speed). In what follows I outline my steps taken with Cython and multithreading to achieve adequate execution time for the required onerous computation task: cortical functional network assignment using prior empirically derived group templates.

The methodology of the task is outlined at [4] and may be formally described by assigning a network to each voxel  $i, j, k$ , which requires computing the function  $\mathcal{A}$  across all gray matter voxels of the brain:

$$\mathcal{A} : (i, j, k) \longrightarrow \{0, 1, \dots, 13\}$$

Where each integer from 0 to 13 corresponds to a network in the following order: fronto-parietal, parieto-medial, dorsal attention, medial temporal, somatomotor-dorsal, parieto-occipital, cingulo-opercular, visual, auditory, language, salience, somatomotor-lateral, default mode. More precisely, for a given voxel, dice coefficients between (template) network regions and functionally correlated voxels are calculated for each network, and the maximal one is assigned to the voxel:

$$\mathcal{A}(i, j, k) = \arg \max_{n \in \{0, 1, \dots, 13\}, \mathcal{T}} \text{dice}(n, \mathcal{T})$$

In English this may be read as assigning the network with the highest dice coefficient (or greatest region overlap) between the template regions and the top correlated areas to voxel  $i, j, k$ . Broadly, the dice coefficient is defined as the degree of overlap between two sets [7]:

$$\text{dice}(n, \mathcal{T}) = \frac{|\text{rois}(n) \cap \mathcal{T}|}{|\text{rois}(n)| + |\mathcal{T}|}$$

It may be noted that  $\text{rois}(n)$  is not a set in space but a collection of points in  $\mathbb{R}^3$ .  $\text{rois}(n)$  can be considered a set of the unions of 7mm spheres with centroids as the provided template regions in MNI space. An intersection is counted when any point in  $\mathcal{T}$  falls within one of the 7mm spheres.  $|\text{rois}(n)|$  is equal to the number of centroids given for the network  $n$ , given by the prior at [4].

The last step is to describe the acquisition of  $\mathcal{T}$ , an array of the top five percent of the correlates across the scan.  $\mathcal{T}$  must be calculated for all gray matter voxels masked, hence the task reduces to computing the matrix  $\mathcal{R}$ , where  $\mathcal{T}$  is any row (or column due to symmetry), and  $B_x, B_y$  are the BOLD time series for given voxels  $x$  and  $y$  masked:

$$\mathcal{R}_{x,y} = \text{correlation}(B_x, B_y)$$

The approach at computing the entirety of entries in  $\mathcal{R}$  was to distribute equal amounts of matrix rows to all available threads. This is synonymous with assigning a list of index pairs to each thread where the two indices denoted the time series of voxels to correlate (Pearson). It may be noted that this algorithm is redundant, since it does not exploit the symmetry of  $\mathcal{R}$  and therefore is twice as slow as a potentially optimal computation.

## 2.3 Results

### 2.3.1 Execution Format

Code for this project may be found at [github.com/BrayNeuroimagingLab/cortical-template-matching](https://github.com/BrayNeuroimagingLab/cortical-template-matching) and its execution is setup to be completed on the University of Calgary research computing clusters. Execution of implemented code requires two files. (a) the fMRI image, a 4D .nii file capturing the BOLD fluctuations of neural activity in space, over time. (b) that image’s associated gray matter mask. After specifying these file names in the `params.csv` located in the working directory, a simple `sbatch run.sh` executes the assignment on the high performance clusters, using 40 cpus and 80GB of memory. The resulting assignment is saved as a 3D .nii file, with each voxel holding an integer, -1 denoting no assignment.

### 2.3.2 Multithreaded Execution time

The methodology described above was implemented in Python with a C++ multithreaded binding to compute the voxelwise correlation  $\mathcal{R}$  matrix. Initial tests found that invoking the Python single threaded `np.corrcoef()` function (that returns this desired  $\mathcal{R}$  matrix) executed for over an hour without finishing and was therefore terminated in hopes of ameliorating such a time burden. The computation of the  $\mathcal{R}$  matrix was the first bottleneck because of its size (anywhere between 91k and 50k square, depending on the gray matter mask). Hence a manual C++ correlation implementation was completed as described. This function’s correctness was confirmed at smaller sizes so that execution could finish, and it was found that the C++ implementation produced an identical  $\mathcal{R}$  matrix to `np.corrcoef()`. Its execution speed was timed at  $\approx 50$  seconds on 40 threads to populate a 91k square  $\mathcal{R}$  matrix, garnering astonishing brevity in comparison to single threaded Python.

Subsequent steps requiring the  $\mathcal{R}$  matrix for calculating dice overlap was also found to be very slow. The library `Ray` allows the programmer to define a remote multi-thread safe function that can share large amounts of memory (notably the  $\mathcal{R}$  matrix which is upwards of 16 GB) during parallel execution using a multitude of cpus. This final step takes the longest, with  $\approx 6700$  seconds on 40 concurrent cpus.

### 2.3.3 Visual Correctness

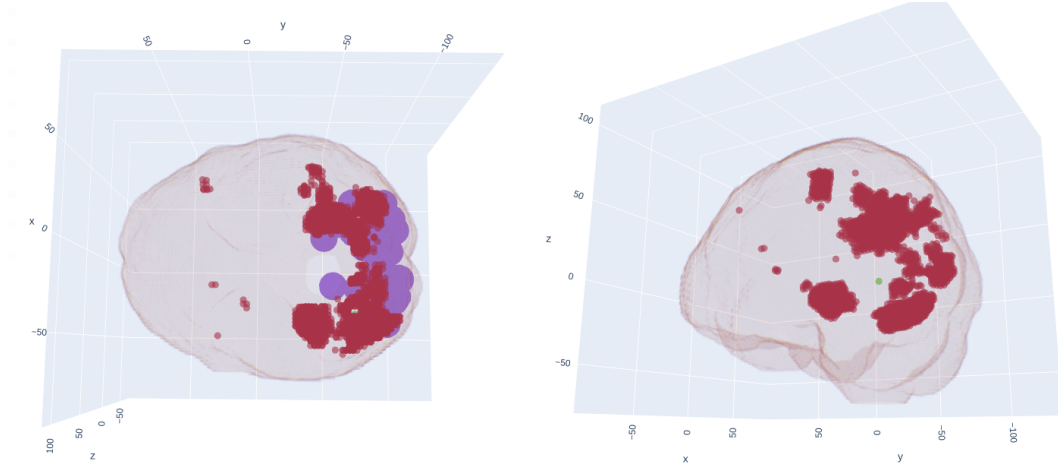


Figure 1: **Left** (Transverse plane): The 7mm diameter rois for the Visual network (purple) overlaid with the top 5% of functionally correlated voxels (red) to the selected green voxel on the **Right** (Sagittal plane of the same image, red voxels exactly equal in both). The green voxel was assigned to the Visual network due to its maximal dice coefficient with the purple rois. Note, as per [8], voxels within 20mm of the green were omitted from the dice overlap to alleviate the bias of local connectivity, as can be seen on the **Right**.

### 3 References

1. Thomas Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3), 1125–1165. <https://doi.org/10.1152/jn.00338.2011>
2. Bressler, S. L. and Menon, V. (2010) ‘Large-scale brain networks in cognition: emerging methods and principles’, *Trends in cognitive sciences*, 14(6), pp. 277–290. <https://doi.org/10.1016/j.tics.2010.04.004>.
3. Seitzman, B. A., Gratton, C., Laumann, T. O., Gordon, E. M., Adeyemo, B., Dworetsky, A., Kraus, B. T., Gilmore, A. W., Berg, J. J., Ortega, M., Nguyen, A., Greene, D. J., McDermott, K. B., Nelson, S. M., Lessov-Schlaggar, C. N., Schlaggar, B. L., Dosenbach, N. U. F., & Petersen, S. E. (2019). Trait-like variants in human functional brain networks. *Proceedings of the National Academy of Sciences*, 116(45), 22851–22861. <https://doi.org/10.1073/pnas.1902932116>
4. Dworetsky, A., Seitzman, B. A., Adeyemo, B., Neta, M., Coalson, R. S., Petersen, S. E., & Gratton, C. (2020). Probabilistic mapping of human functional brain networks identifies regions of high group consensus. *Cold Spring Harbor Laboratory*. <http://dx.doi.org/10.1101/2020.09.28.313791>
5. Python Software Foundation. Python Language Reference, version 3.8. Available at [www.python.org](http://www.python.org)
6. Behnel, S., Bradshaw, R., Citro, C., Dalcin, L., Seljebotn, D.S., & Smith, K. (2011). Cython: The Best of Both Worlds. *Computing in Science and Engineering*, 13, 31-39, <https://doi.org/10.1109/MCSE.2010.118>
7. Dice, L. R. (1945) ‘Measures of the Amount of Ecologic Association Between Species’, *Ecology (Durham)*, 26(3), pp. 297–302. <https://doi.org/10.2307/1932409>.
8. Gordon, E. M., Laumann, T. O., Adeyemo, B., Gilmore, A. W., Nelson, S. M., Dosenbach, N. U. F., & Petersen, S. E. (2017). Individual-specific features of brain systems identified with resting state functional correlations. *NeuroImage*, 146, 918–939. <https://doi.org/10.1016/j.neuroimage.2016.08.032>

## 4 Reflection

The impetus behind my initial decision to reach out to Dr. Signe Bray was due to an interest more broadly in the impact of cell phones on our lives. It is a simple but often overlooked question: why do our screens grab attention more potently than other objects? Can screen experiences afforded by these devices dysregulate our nervous systems? Attention and reward are cognitive features that neuroscience has been seeking to characterize for decades, and from my suspicion, lie at the bottom of these questions.

Dr. Bray believed I had the potential to contribute to the lab for data analysis, so she accepted me as a student. I learned that in neuroimaging, participant image acquisition is not a downstream response to specific questions sought, instead a general paradigm is proposed and data acquisition is first a priority, where many specific studies may follow from the same dataset. This is a consequence of the expense, both in time and cost, of recruiting many research participants to undertake functional Magnetic Resonance Imaging of their brain.

The main study The Bray Lab is executing is called PreciseKIDS. Sought participants are parents with children around age six, where both the child and parent will have lengthy scans of their brains, while looking across different screen groups, differentiated by their media type. We were unfortunately unable to look at reward circuit activation in our data because its acquisition was delayed significantly due to the COVID-19 pandemic.

In the interim Dr. Bray directed me to look at Connectome Predictive Modelling, a new paradigm of data analysis in neuroscience that uses brain connectivity features to establish brain-behaviour relationships. I recall before the start of my summer research term viewing papers on exactly this method, but found its comprehension out of reach. After truly sitting through the tension of reading these papers once they were assigned to me, I grew an enhanced understanding of the exact steps behind this type of predictive modelling.

It is a prudent goal to develop abilities that are not prevalent in the majority of students. This marks your time with great monetary value. As such, I looked to press hard into a interdisciplinary field of neuro-informatics: how can we make sense of the observations extracted from brain images? Massive amounts of data calls for expertise in computer science. I learned over the summer the ideal environment for scientific computing. Python should be defaulted to. By taking advantage of its ease of development and comprehensive package manager Anaconda, many common scientific and statistical computations can be visualized with haste. However this necessarily comes with poor execution speed when the data scales. The fix here is to use a low level binding called Cython, where any python function may be written instead in the C++ language to garner its speed benefits. C++ also holds excellent capability of multithreading, a method to parallelize execution and further enhance speed. Over the summer I was able to work diligently in this environment, my familiarity with C++ multithreading, a more challenging programming paradigm, was greatly improved.