fMRI preprocessing considerations for graph theory analysis

Paul McCarthy
Artificial Intelligence Research Group
Computer Science Department
University of Otago
Dunedin, Otago, New Zealand
pauld.mccarthy@gmail.com

ABSTRACT

This paper discusses the application of graph theory analysis to fMRI brain imaging data. In particular, the effects of standard fMRI preprocessing techniques upon graphs generated from fMRI volumes are explored.

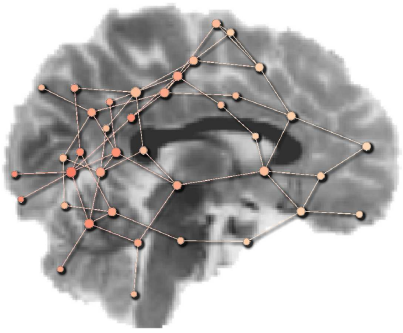
A case study on real data is presented, whereby the effects of preprocessing upon graphs generated from the data are analysed.

Categories and Subject Descriptors

G.2.2 [**Graph Theory**]: Analysis; J.3 [**Life and Medical Sciences**]: Neuroimaging

Keywords

fMRI, Preprocessing, Graph Theory



1. INTRODUCTION

Graph theory has become a popular technique for analysis of fMRI and other brain imaging data. Graph theory analysis provides a unique, global perspective on the function and structure of the brain, allowing insights into the processes of the brain at a higher level than afforded by more traditional techniques ([17]).

Preprocessing is an essential component in the analysis of fMRI data and, like other methods of analysis, the output of graph theoretical techniques may be adversely affected by poorly or improperly processed data. This paper presents a discussion of the effects that fMRI preprocessing can have on subsequent graph theory analysis. Four standard preprocessing tasks are covered:

- *Slice timing correction*, temporally aligning slices within every image of a fMRI volume.
- *Motion correction*, correcting for head movement during a fMRI volume.
- *Intra-subject registration*, aligning fMRI volumes of each subject to a corresponding structural MRI image.
- *Spatial normalisation*, aligning fMRI recordings of multiple subjects to a standard anatomical template, or atlas.

The effects of each of these tasks upon graphs generated from the data are discussed, and a case study presented which demonstrates these effects. The paper concentrates on the simplest definition of a graph, that is, unweighted and undirected, with graph nodes corresponding to voxels in a fMRI volume.

1.1 Functional Magnetic Resonance Imaging (fMRI)

fMRI is a technique for capturing high resolution 3D images of blood oxygen (BOLD) levels in the brain. Neural activity throughout the brain may be inferred from these BOLD levels ([20]). A typical fMRI image consists of 3D pixels, known as *voxels*, each representing an area as small as $\sim 1mm^3$. A collection of these images over time becomes a 4 dimensional data set, and is referred to as a *volume*.

Capturing a single fMRI image is a time consuming process, so while fMRI provides spatial resolution down to the millimetre, temporal resolution is typically on the order of seconds. Thus, the types of events which can be analysed using fMRI are those which occur at a relatively low frequency, below $1Hz$ ([11], [21]).

1.2 Crash course in graph theory

Graph theory is a branch of mathematics which is useful in modelling and simulating real world systems that display graph- or network-like characteristics. Graph theory is based upon the simple concept of nodes and edges. A node represents an element in the graph, and an edge represents a connection between two nodes[1].

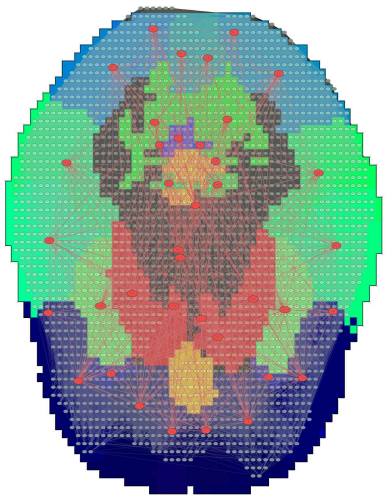
Some interesting measures can be calculated over such a graph.

- The *order* of a graph G is the number of nodes contained in G.
- The *density* of a graph is the ratio of existing edges to all possible edges ([5]).
- The *characteristic path length* ([24]) is the average shortest path length between all pairs of nodes in a graph. The average path length of a node or a graph provides insight into how efficiently the graph may process information, and how quickly nodes may communicate with each other.
- *Clustering coefficient* ([24]) is the ratio of the number of edges which are present between a node's neighbours to the number of possible edges. In other words, the clustering coefficient of a node is the density of the subgraph formed by the node's immediate neighbours, and the edges which exist between them. The clustering coefficient of a graph provides insight into the level to which neighbourhoods, or clusters, are present in the graph.
- The *small-world index* ([24], [10]) gives an indication of the small-world nature of a graph. For a graph to be considered *small-world*, it must exhibit both the high clustering coefficient of an ordered graph, and the low path length of a random graph. Small-world graphs are considered to be highly efficient ([2], [22]).

1.3 Applying graph theory to fMRI data

There are many ways to apply graph theory analysis to fMRI data; one typical approach proceeds as follows (e.g. [17], [4], [15]):

1. Define your nodes. Most studies segment the brain area into a small number (< 200) of *regions-of-interest* (ROI), corresponding to areas of importance or interest. Another, more computationally intensive, option, and the approach considered in this paper, is to treat every single voxel as a ROI (a standard resolution fMRI image may contain ~ 10000 active voxels).
2. Define your edges. Some measure of association is calculated between every pair of ROIs in the graph. The *Pearson Correlation Coefficient* ([14]) is commonly used (e.g. [6], [23]), and is calculated over the BOLD time series for each ROI in a fMRI volume.
3. Binarise your graph. To create an unweighted graph, some threshold must be applied to the association values. An edge is added to the graph for each association value above the threshold.



[1]Many variations of this simple model exist, but for the purposes of this paper, we will stick with these *unweighted, undirected* graphs.

2. FMRI PREPROCESSING

Preprocessing is universally acknowledged to be an essential step in fMRI data analysis, however the specific tools and techniques differ between studies ([18], [7]). Certain preprocessing steps may not be required for certain experiments. For example, an experiment which does not compare subject brains on a region-by-region basis may not require registration to a standard anatomical template.

This section discusses four core steps in a fMRI preprocessing schedule, and the effects that each may have on subsequent graph theory analysis.

2.1 Slice timing correction

fMRI images are captured in *slices*, which are 2D cross-sections of the brain. A single fMRI scan will capture a number of slices; these slices are then stiched together to form a single 3D image of the entire brain.

As mentioned earlier, fMRI image capture is a time consuming process - so much so that a significant period of time ($\sim 100ms$) may elapse between consecutive slices. This lag must be accounted for, as analyses will generally assume that each 3D image in a volume was captured instantaneously. Tools exist to perform this correction, known as *slice timing correction*, adjusting values to temporally align slices within each image of a volume ([12]).

When applying graph theory analysis to a fMRI volume, slice timing correction should be applied, as calculation of a measure of association between a pair of time series will generally require the corresponding samples within each series to have been taken at the same time.

As an example, take the time series from two ROIs, one from the first slice captured, and one from the last slice captured. Without slice timing correction, the temporal difference between corresponding pairs of samples from these

series could be up to 3 seconds. This is enough time for a significant change in BOLD level ([8]).

If we expand this scenario to every pair of ROIs in a fMRI volume, it is clear that the lag between slices could result in significant deviation in graph density, due to errors in association measure calculation, and large numbers of edges being incorrectly added to, or omitted from, the resulting graph.

2.2 Motion correction

A typical fMRI run may last several minutes and, whilst large errors can be minimised by immobilising the subject's head, there is bound to be some movement throughout the recording process ([18], [7]). A common early step in fMRI preprocessing, then, is to align the images within each volume so that they all line up in space, over time. This is accomplished with image transformation techniques, and is referred to as *registration* ([7]).

Motion correction is an important step in graph theory analysis. Extracting BOLD time series data from each ROI in a fMRI volume relies upon those ROIs being aligned over time. When movement has occurred, some ROIs may not be aligned, and the extracted time series will not be representative of the actual activity which occurred in those regions.

This error will have the same effect as for slice timing correction, in that the association measures calculated over pairs of time series will contain errors, and will affect the density and structure of the resulting graph.

2.3 Intra-subject registration

A standard fMRI study on a subject will generate more than one fMRI volume, for example, repeated recordings of some task. These volumes may then be averaged, with subsequent analyses performed upon the averaged volume ([7]).

Before this averaging occurs, it is essential that the volumes being averaged are spatially aligned, otherwise non-corresponding voxels will be averaged, and the resulting volume will be at best blurred, at worst useless. Again, image transformation techniques are used to resolve this issue. It is common for each fMRI volume of a subject to be aligned with a corresponding structural MRI image, taken at the same time ([7], [18]).

During a series of fMRI recordings, it is typical for the subject to be allowed to rest between these recordings (see e.g. [3]). A potential source of error is introduced if, during one of these rest periods, the subject adjusts his/her position, for example to stretch or make him/herself more comfortable. This exact scenario occurs in Figure 1.

Consider what would happen if these volumes were averaged, and a graph generated from the average. The number of active voxels in the average is greater than that in either of the individual volumes, so the resulting graph would contain an abnormally high number of nodes. ROIs are not aligned so will be averaged incorrectly - this would impact the association measure accuracy and, in turn, the graph density.

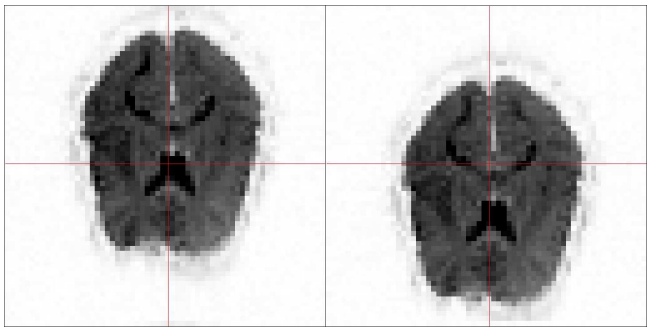


Figure 1: Corresponding slices from two fMRI images of the same brain, taken from consecutive recordings in a fMRI study ([3], [1]). The images were captured approximately 2 minutes apart; either the subject position, or the scanner configuration, changed during this time.

2.4 Spatial normalisation

Every brain is different, in size and, to some extent, in structure (see Figure 2 for an example). It is difficult, then, for direct anatomical comparisons to be made between subjects in a study. This is why it is common to apply further image transformation techniques, aligning every fMRI image in a study to a standard anatomical template, and thus allowing comparisons to be made from an anatomical perspective. Templates most commonly in use are the ICBM152 and ICBM452 atlases ([13]), which have replaced the Talairach Atlas ([19]) as the preferred standard.

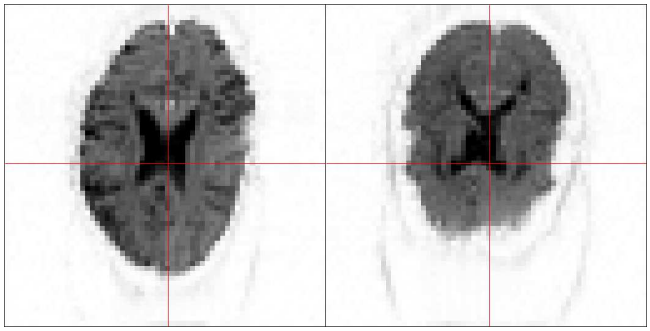


Figure 2: Slices from fMRI images of two brains, taken from the same study ([3], [1]).

The size of a subject's brain determines the number of active voxels in a corresponding fMRI image; this is equivalent to the order (number of nodes) of the resulting graph. Thus, without registration to a standard template, the order of graphs generated from different brains would vary widely, and any statistical analyses would merely reflect the brain size of the individual subjects[2].

3. CASE STUDY

To demonstrate the effects of fMRI preprocessing on resulting graphs, analysis upon a publicly available fMRI data set

[2]This problem would not occur if ROIs were used as graph nodes, although, without registration, the ROI approach would require a large amount of tedious work in manually defining the ROIs for each individual subject.

([3], [1]) was performed. The data were preprocessed, and graphs generated from both the raw, unprocessed data, and the preprocessed data, for every subject in the study.

The data set consists of fMRI recordings from 41 subjects, each categorised into one of three groups - *young*, *aged*, and *demented*. Various studies on this data set (e.g. [3], [9]) have sought differences between these three groups; we will not do so, as our primary concern is comparing graphs generated from raw data with those generated from preprocessed data.

3.1 The experiment

The study involved subjects completing a simple visual stimulus task. Each subject underwent four fMRI recording sessions, referred to as *runs*. Each run consisted of 128 fMRI images, with a TR (image acquisition) time of 2.68 seconds; this corresponds to nearly 6 minutes of recording per run.

During a single run, 15 *trials* were executed, each having a duration of 8 images (about 20 seconds). A trial consisted of either one or two visual stimuli. During a 'one-trial' condition, the stimulus was triggered at the start of the trial. During a 'two-trial' condition, the first stimulus was triggered at the start of the trial, and the second stimulus was triggered about 5 seconds after the first. The subjects were instructed to push a button upon the presentation of each stimulus.

3.2 Preprocessing

Preprocessing consisted of the steps described below. Slice timing correction was accomplished with the `slicetimer` command, provided with `FSL 4.1`[3] ([16]). Image registration tasks were performed with `AIR 5.2.6`[4] ([25]). Manipulation of `ANALYZE75` files was performed with the `analyzetools`[5] package. The first trial in each run began at image #5, and the last trial ended at image #125; thus, before any preprocessing, the first four and last four images from every run were discarded.

1. *Visual inspection*: Every volume in the data set was visually inspected to check for obvious anomalies. This step uncovered three suspect data sets; the volumes for subject #3 were discarded, due to the presence of significant errors throughout every run. The volumes for subjects #15 and #19 contained aliasing effects, which were manually corrected.

2. *Slice-timing correction*: Every volume was corrected for slice timing differences.

3. *Motion correction*: Motion correction was performed on every volume using a 6-parameter rigid body transformation.

4. *Intra-subject registration*: Every volume was aligned to the corresponding structural MRI image using a 12-parameter affine transformation.

5. *Spatial normalisation*: Every volume was then aligned to the ICBM452 atlas ([13]). This was accomplished by aligning the subject's structural MRI image to the atlas with a 12-parameter affine transformation, and then applying the same alignment parameters to the subject's fMRI volumes. This two step method is considered to achieve better results than aligning fMRI images directly to the atlas ([7], [18]). Finally, the fMRI images were downsampled back to the original resolution of $64 * 64 * 16$ voxels.

[3] `http://www.fmrib.ox.ac.uk/fsl/`
[4] `http://bishopw.loni.ucla.edu/air5/`
[5] `https://github.com/pauldmccarthy/analyzetools`

3.3 Graph generation

The data used for graph generation consists of a single *trial-averaged* volume for each subject. For each subject, every trial period was extracted from every run. These trial periods were then averaged over time, resulting in a single 8-image fMRI volume for each subject. These volumes formed the basis for subsequent graph generation. Correlation matrix generation, graph generation, and graph statistic calculation were accomplished with two software packages, `hnet`[6] and `cnet`[7].

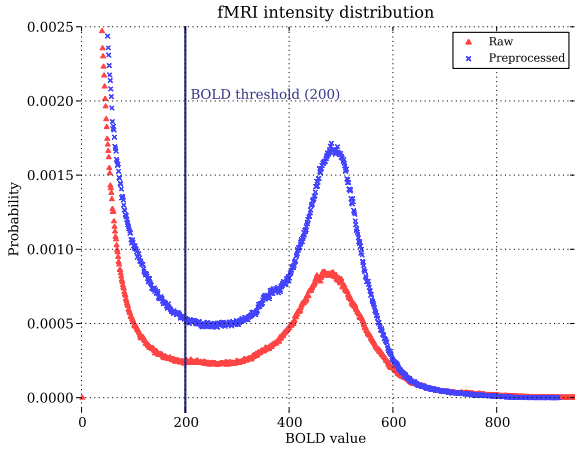


Figure 3: BOLD intensity distribution for raw and preprocessed fMRI volumes. The larger intensity for preprocessed data is due to the fact that, after spatial normalisation to the ICBM452 atlas, the active brain area is larger, hence there are more active voxels.

For each volume, all voxels for which the BOLD level did not exceed 200 over time were excluded from the graph generation. The value of 200 was chosen by plotting the intensity distributions for both raw and preprocessed data; these distributions are shown in Figure 3. The remaining *active* voxels became nodes in the graph. The Pearson Correlation Coefficient was calculated between the time series for every pair of active voxels. For all correlation values which fell above a correlation threshold value of 0.9, an edge was added between the corresponding nodes in the graph.

3.4 Results

Figures 4, 5, and 6 display the order, density, and small-world index respectively, of graphs generated from both the raw and preprocessed fMRI data sets. Statistics are given in

[6] `https://github.com/pauldmccarthy/hnet`
[7] `https://github.com/pauldmccarthy/cnet`

		Raw	*Preprocessed*
Order	Mean (SD)	**12324.5** (1203.21)	**22915.5** (1448.97)
Density	Mean (SD)	**0.01094** (**0.00504**)	**0.02264** (**0.01501**)
Small-world index	Mean (SD)	**23.4429** (8.71413)	**15.0261** (8.03610)

Table 1: Mean and standard deviations for statistics of the graphs generated from raw and preprocessed fMRI data. Values in bold were found to be significantly different.

Table 1. Quantile-quantile plots for each of the three parameters did not show siginficant departures from normality, so the F-test was used to compare variances, and Welch's T-test, (for equal or unequal variances, depending upon the outcome of the F-test), was used to compare means.

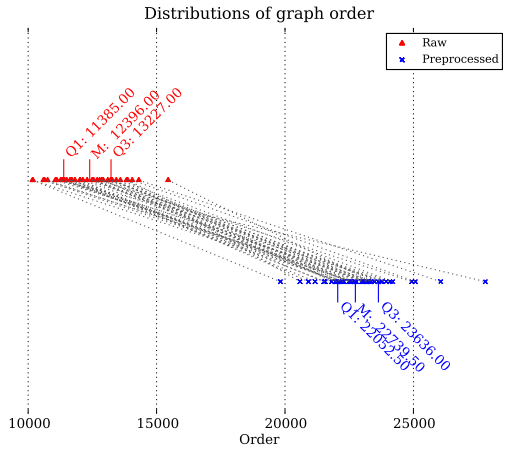


Figure 4: Order of graphs generated from raw and preprocessed fMRI data. The first quartile (Q1), median (M) and third quartile (Q3) are shown. Dotted lines between points indicate corresponding graphs from each of the data sets.

As mentioned earlier, spatial normalisation increased the number of active voxels (voxels which reach a BOLD level above 200) in each volume; this corresponds directly to the number of nodes in resulting graphs, so the mean order for preprocessed data is significantly higher ($p < 0.0001$). However, despite the effects of brain size upon graph order being removed via spatial normalisation, no significant difference in order variance is present ($p = 0.25$).

Perhaps the true variance in the number of active voxels between subjects is higher than that exhibited in the raw data sets, and preprocessing has highlighted this difference. It is also possible that some other preprocessing step has influenced the overall voxel activity.

Preprocessing had a large effect on graph density, with a significantly higher variance ($p < 0.0001$), and different mean ($p < 0.0001$). Small-worldness was also affected, with similar variances ($p = 0.62$), but different means ($p < 0.0001$). This is an interesting effect, and appears to be directly related to the different densities - the higher density of the preprocessed graphs leads to a corresponding decrease in the average path lengths of equivalent random graphs[8]. There is no decrease in the path lengths of the preprocessed graphs however; this leads to a reduction in the small-world index.

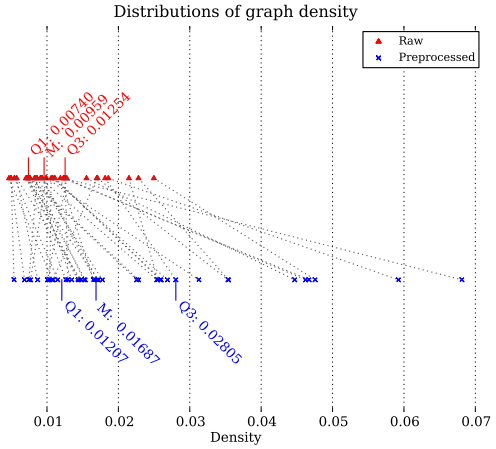


Figure 5: Density of graphs generated from raw and preprocessed fMRI data.

One trend apparent across each of the three parameters is the seemingly random reordering of graphs on each property, after preprocessing, as shown by the crossing over of the dotted lines between corresponding graphs in Figures 4, 5 and 6. This shows that the preprocessing steps affect the data from each individual subject in different ways.

Consider, for example, the different levels of warping which would be produced during spatial normalisation upon the two brains shown in Figure 2. Such effects manifest themselves in complex ways in the resulting graphs, as demonstrated by the differences between the two data sets in density and small-worldness.

4. CONCLUSION

This paper has provided an overview of the graph theoretical analysis of fMRI data, and the effects that standard fMRI preprocessing techniques have on such analyses. A case study was presented, demonstrating the importance of applying preprocessing to fMRI data before performing any graph theory analysis.

Results change significantly when preprocessing is applied; without preprocessing, any results derived from analyses may be biased for a number of reasons. Care must be taken

[8]The small-world index of a graph is calculated by comparing its clustering coefficient and path length to that of an equivalent random graph, that is, a randomly generated graph with the same order and density as the graph in question.

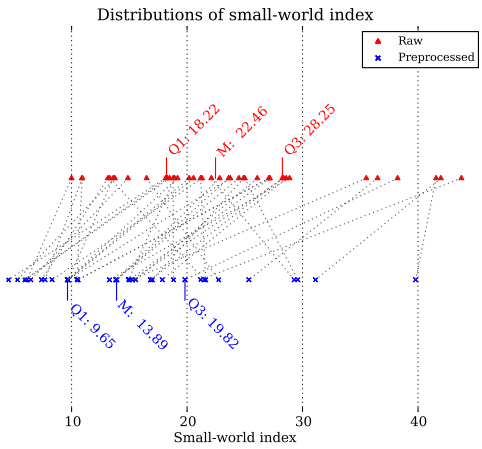


Figure 6: Small-world index of graphs generated from raw and preprocessed fMRI data.

that necessary corrections have been made to the data before performing analysis, and before drawing any conclusions from the results of analysis.

5. ACKNOWLEDGEMENTS

This research was funded by a University of Otago PhD Scholarship.

6. REFERENCES

[1] The fMRI Data Center. http://www.fmridc.org.

[2] D. S. Bassett and E. Bullmore. Small-world brain networks. *Neuroscientist*, 12(6):512–23, 2006.

[3] R. L. Buckner, A. Z. Snyder, A. L. Sanders, M. E. Raichle, and J. C. Morris. Functional Brain Imaging of Young, Nondemented, and Demented Older Adults. *Journal of Cognitive Neuroscience*, 12(Supplement 2):24–34, 2000.

[4] E. Bullmore and O. Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.*, 10(3):186–98, 2009.

[5] R. Diestel. *Graph Theory*. Springer-Verlag, 3rd edition, 2005.

[6] V. M. Eguiluz, D. R. Chialvo, G. A. Cecchi, M. Baliki, and A. V. Apkarian. Scale-Free Brain Functional Networks. *Physical Review Letters*, 94(1):018102–1 – 018102–4e, 2005.

[7] M. Filippi, editor. *fMRI Techniques and Protocols*, volume 41 of *Neuromethods*. Humana Press, 2009.

[8] K. Friston, P. Jezzard, and R. Turner. Analysis of Functional MRI Time Series. *Human Brain Mapping*, 1(2):153–171, 1994.

[9] M. D. Greicius, G. Srivastava, A. L. Reiss, and V. Menon. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: Evidence from functional MRI. *PNAS*, 101(13):4637–4642, 2004.

[10] M. D. Humphries, K. Gurney, and T. J. Prescott. The brainstem reticular formation is a small-world, not scale-free, network. *Proc. Biol. Sci.*, 273(1585):503–11, 2006.

[11] V. K. Jirsa and A. R. Mcintosh, editors. *Handbook of Brain Connectivity*. Springer Complexity. Springer, 2007.

[12] N. A. Lazar. *The Statistical Analysis of Functional MRI Data*. Statistics for Biology and Health. Springer, 2008.

[13] J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, T. Paus, G. Simpson, B. Pike, C. Holmes, L. Collins, P. Thompson, D. MacDonald, M. Iacoboni, T. Schormann, K. Amunts, N. Palomero-Gallagher, S. Geyer, L. Parsons, K. Narr, N. Kabani, G. Le Goualher, D. Boomsma, T. Cannon, R. Kawashima, and B. Mazoyer. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 356(1412):1293–322, 2001.

[14] J. L. Rodgers and W. A. Nicewander. Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician*, 42(1):59–66, 1988.

[15] M. Rubinov and O. Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–69, 2010.

[16] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. J. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J. M. Brady, and P. M. Matthews. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, 23 Suppl 1:S208–19, 2004.

[17] C. J. Stam and J. C. Reijneveld. Graph theoretical analysis of complex networks in the brain. *Nonlinear biomedical physics*, 1(1):3, 2007.

[18] S. C. Strother. Evaluating fMRI Preprocessing Pipelines. *IEEE Engineering in Medicine and Biology Magazine*, 25(2):27–41, 2006.

[19] J. Talairach and P. Tournoux. *Co-Planar Stereotaxic Atlas of the Human Brain: 3-D Proportional System: An Approach to Cerebral Imaging*. Thieme, 1988.

[20] S. Ulmer and O. Jansen, editors. *fMRI: Basics and Clinical Applications*. Springer, 2010.

[21] M. P. van den Heuvel and H. E. Hulshoff Pol. Exploring the brain network: a review on resting-state fMRI functional connectivity. *Eur Neuropsychopharmacol*, 20(8):519–34, 2010.

[22] M. P. van den Heuvel, C. J. Stam, M. Boersma, and H. E. Hulshoff Pol. Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. *Neuroimage*, 43(3):528–39, 2008.

[23] M. P. van den Heuvel, C. J. Stam, R. S. Kahn, and H. E. Hulshoff Pol. Efficiency of functional brain networks and intellectual performance. *J. Neurosci.*, 29(23):7619–24, 2009.

[24] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393:440–442, 1998.

[25] R. Woods, S. Cherry, and J. Mazziotta. Rapid automated algorithm for aligning and reslicing PET images. *Journal of Computer Assisted Tomography*, 16(4):620–633, 1992.