REVIEW

Connectomics and open science approaches to analyze brain connectivity

R. Cameron Craddock[1,2*], Rosalia L. Tungaraza[1] and Michael P. Milham[1,2]

*Correspondence: ccraddock@nki.rfmh.org
[1]Computational Neuroimaging Lab, Center for Biomedical Imaging and Neuromodulation, Nathan Kline Institute for Psychiatric Research, 140 Old Orangeburg Rd, 10962 Orangeburg, New York, USA
Full list of author information is available at the end of the article

Abstract

First part title: Text for this section.

Second part title: Text for this section.

Keywords: human connectome; functional MRI; brain graphs; open data; open science

1 Introduction

With its new emphasis on collecting larger datasets, data sharing, deep phenotyping, and multimodal integration, neuroimaging has become a data intensive science. This is particularly true for connectomics where grass-root initiatives (e.g.the 1000 Functional Connectomes Project (FCP)[1], the International Neuroimaging Data-sharing Initiative (INDI)[2]) and large-scale interational projects (the Human Connectome Project[**?**, 3], the Brainnetome[4], the Human Brain Project in EU known as CONNECT[5], PING, Philadelphia Neurodevelopmental Cohort, Brain Genomics Superstruct, NDAR) are collecting and openly sharing thousands of brain imaging scans, each of which consist of hundreds of observations of thousands of variables. Although this deluge of complex data promises to enable the investigation of neuroscientific questions that were previously inaccessible, it is quickly overwhelming the capacity of existing tools and algorithms to extract meaningful information from the data. This combined with a new focus on discovery science is creating a plethora of opportunities for data scientists from a wide range of disciplines such as computer science, engineering, mathematics, statistics, etc., to make substantial contributions to neuroscience. The goal of this review is to describe the state-of-the-art in connectomics research and enumerate opportunities for data scientists to contribute to the field.

The human connectome is a comprehensive map of the brain's circuitry, which consists of brain areas, their structural connections and their functional interactions. The connectome can be measured with a variety of different imaging techniques, but magnetic resonance imaging (MRI) is the most common primarily due to its near-ubiquity, non-invasiveness, and high spatial resolution. As measured by MRI: brain areas are patches of cortex (approximately $1cm^3$ volume) containing 1,000s of neurons, structural connections are long range fiber tracts that are inferred from the motion of water particles measured by diffusion weighted MRI (dMRI), and functional interactions are inferred from synchronized brain activity measured by

functional MRI (fMRI). Addressing the current state-of-the-art for both functional and structural connectivity is well beyond the scope of a single review. Instead, this review will focus on functional connectivity, which is particularly fast growing and offers many exciting opportunities for data scientists.

The advent of functional connectivity analyses has popularized the application of discovery science to brain function, which marks a shift in emphasis from hypothesis testing, to supervised and unsupervised methods for learning statistical relationships from the data. Since functional connectivity is inferred from statistical dependencies between physiological measures of brain activity (i.e. correlations between the dependent variables), it can be measured without an experimental manipulation. Thus, functional connectivity is most commonly measured from "resting state" fMRI scans, during which the study participant is lying quietly and not performing an experimenter specified task- when measured in this way, it is referred to as intrinsic functional connectivity (iFC). Once iFC is measured, data mining techniques can be applied to identify iFC patterns that covary with phenotypes, such as, indices of cognitive abilities, personality traits, or disease state, severity, and prognosis, to name a few. In a time dominated by skepticism about the ecological validity of psychiatric diagnoses, iFC analyses have become particularly important for identifying subgroups within patient populations by similarity in brain architecture, rather than similarity in symptom profiles. This new emphasis in discovery necessitates a new breed of data analysis tools that are equipped to deal with the issues inherent to functional neuroimaging data.

2 The Connectome Analysis Paradigm

In 2005 Sporn and Hagmann [6, 7] independently and in parallel coined the term *the human connectome*, which embodies the notion that the set of all connections within the human brain can be represented and understood as graphs. In the context of iFC, graphs provide a mathematical representation of the functional interactions between brain areas - nodes in the graph represent brain areas and edges indicate their functional connectivity. While general graphs can have multiple edges between two nodes, brain graphs tend to be simple graphs with a single undirected edge between pairs of nodes (i.e. the direction of influence between nodes is unknown). Additionally edges in graphs of brain function tend to be weighted - annotated with a value ¡¡¡¡¡¡¡ Local Changes ¡¡¡¡¡¡¡ Local Changes that indicates the similarity between nodes. Analyzing the functional connectivity ======= that indicates the similarity between nodes. Analyzing the functional connectome [CHANGE] ¿¿¿¿¿¿¿ External Changes ======= that indicates the similarity between nodes. Analyzing the functional connectome [CHANGE] ¿¿¿¿¿¿¿ External Changes involves 1) preprocessing the data to remove confounding variation and to make it comparable across datasets, 2) specification of brain areas to be used as nodes, 3) identification of edges from the iFC between nodes, and 4) analysis of the graph (i.e. the structure and edges) to identifiy relationships with inter- or intra- individual variability. All of these steps have been well covered in the literature by other reviews and repeating that information provides little value. Instead we will focus on a exciting areas in the functional connectomics literature that we beleive provide the greatest opportunities for data scientists in this quickly advancing field.

e many papers

2.1 Modeling connections

2.1.1 Static Connectivity

A variety of different bivariate and multivariate methods have been proposed for measuring the similarity between timecourses of brain areas [**?**, **?**]. Although these methods are well suited for identifying weighted edges for connectome graphs, they provide an incomplete picture of node interactions that limits the amount of neuroscientific information that can be extracted. Thus several different modeling techniques have been proposed to provide a more precise mathematical description of the relationship between brain areas. Model confirmatory approaches such as structural equation modeling (SEM) and dynamic causal modeling (DCM) offer fairly detailed descriptions of node relationships that have shown great promise in applications such as predicting the affected hemisphere from stroke data [**?**]. But, they rely on the prespecification of a model and are limited in the size of a network that can be modeled. Cross-validation methods have been proposed to systematically search for the best model, but simulations have shown that those methods do not necessarily converge to the correct model [**?**]. Granger causality has been particularly popular due to its promise of identifying causal relationships between nodes based on temporal lags between them []. But the assumptions underlying granger causality do not quite fit with fMRI data, where delays in the time-courses between regions may reflect some physiological phenomena, such as a perfusion deficit [**?**], rather than causal relationships between brain areas.

Perhaps the oldest model of functional connectivity represents the activity of a single brain areas or node as the weighted average of the activity measured in every other region of the brain [**?**, **?**]. This multivariate regression model provides a more complete picture than commonly used bivariate measures, because the estimated coefficients describe a precise mathematical relationship, albeit not causal, between brain areas. Additionally this model is primarily sensitive to direct, rather than indirect, interactions. Unfortunately due to the large number of brain areas in the connectome, and the few numbers of observations available standard resting state fMRI acquisitions, this model is underdetermined, and methods that rely on either dimensionality reduction [**?**] or regularization [**?**, **?**, **?**] must be employed to find a unique solution. These methods have yet to become very popular for modeling connections, perhaps due to the complexity (real or perceived) in their use. One interesting application of these multivariate regression approach, is that they can be applied to data from a different scanning session, experimental paradigm, or even a different subjet to measure how well the model generalizes to the new data [8].

2.2 Dynamic Connectivity

[?ALT: THE DYNAMIC CONNECTOME] [ALT: Consider starting with the idea that: Although functional interactions within the connectome are commonly represented using three dimensional graphs, much like the structural wiring diagram for the connectome, recent work has drawn attention to time as a potential fourth dimension that requires consideration].

Standard seed- and ICA- methods for mapping iFC assume that it is stationary, and derive connectivity patterns from the entirity of the available fMRI time course. Recent studies however, have demonstrated that connectivity between brain regions

change dynamically over time [**?**, **?**, **?**, 9, **?**]. A variety of investigations have dynamic iFC have already been performed, most of which measure connectivity withen small a window of the fMRI time course that is gradually moved forward along time []. Several problems must be overcome in order to reliably measure changing functional connectivity patterns, from the inherently slow and poorly sampled fMRI signal. First, the variance of correlation estimates increases with decreasing window size, meaning that unless proper statistical controls are utilized, the observed dynamics may arise solely from the increased variance []. This issue may be mitigated using the new higher speed imaging methods, which has already shown promise for extracting dynamic network modes using temporal ICA, although really large number of observations are still necessary [**?**].Node definition is another issue, as it is unclear whether brain areas defined from static iFC are appropriate for dynamic iFC, although initial work has shown that parcellations of at least some brain regions from dynamic iFC are consistent with what is found with static [**?**].

rify this

2.3 Comparing brain graphs

The ultimate goals of connectomics is to map the brain's functional architecture and to annotate that architecture with the cognitive or behavioral functions that they subtend. This latter pursuit is achieved by a group level analysis in which variations in the connectome are mapped to inter-individual variations in phenotype [**?**], or intra-individual responses to experimental perturbations [**?**]. Several different analyses have been proposed for accomplishing these goals, and they all require some mechanism for comparing brain graphs.

Approaches to comparing brain graphs can be differentiated based on how they treat the statistical relationships between edges. One such approach, referred to as *bag of edges*, is to treat each edge in the brain graph as a sample from some random variable. Thus, a set of N brain graphs each with M edges will have N observations for each of the M random variables. In this case, the adjacency (or similarity) matrix that describes the brain graphs can be flattened into a vector representation and any of the well explored similarity or dissimilarity metrics can be applied to the data [**?**]. One of the benefits of this representation is the ability to treat each edge as independent of all other edges and to compare graphs using mass univariate analysis, in which, a separate univariate statistical test (e.g. t-test, anova, or ancova) is performed at each edge. This will result in a very large number of comparisons and an appropriate correction for multiple comparisons, such as Network Based Statistic [**?**], Spatial Pairwise Clustering [10], Statistical Parametric Networks [**?**], or group-wise false discovery rate [], must be employed to control the number of false positives. Alternatively the interdependencies between edges can be modeled at the node level using multivariate distance multiple regression (MDMR) [**?**], or across all edges using machine learning methods [11, 12, 13].

Despite the successful application of this technique, a drawback of representing a brain graph as a bag of edges is that this representation throws away all information about the structure of the graph. Being able to retain these graph structures within an analysis commonly known as Frequent Subgraph Mining (FSM) has facilitated the discovery of features that better discriminated between different groups of graphs [**?**]. For instance, [14] were able to identify discriminative subgraphs from

functional connectivity graphs that had a high predictive power for high versus low learners given specific motor tasks. [15] outlines other approaches that take the graph structure into account e.g. the graph edit distance and a number of different graph kernels. All these methods are under active development and have not been widely adapted by the connectomics

Another approach for graph similarity using all the vertexes involves computing a set of *graph-invariants* such as node centrality, modality, global efficiency, etc. and using the values of these measures to represent the graph [**?**][**?**]. Depending on the invariant used, this approach may permit the direct comparison of graphs that are not aligned. Another advantage is that invariants substantially reduce the dimensionality of the graph comparison problem. On the other hand, representing the graph using its computed invariants throws away information about that graph's vertex labels [**?**]. Moreover, after computing these invariants it is often unclear how they can be interpreted biologically. It is important that the invariant used matches the relationships represented by the graph. Since, edges in functional brain graphs represent statistical dependencies between nodes and not anatomical connections, many of the path based invariants do not make sense, as indirect relationships are not interpretable []. For example, the relationships $A \leftrightarrow B$ and $B \leftrightarrow C$ do not imply that there is a path between nodes A and C, if a statistical relationship between A and C were to exist they would be connected directly.

2.3.1 Prediction

Resting state fMRI and iFC analyses are most commonly applied to studying clinical disorders and to this end, the ultimate goal is the identification of biomarkers of disease state, severity, and prognosis[**?**]. Prediction modelling has become a popular analysis method because it most directly addresses the question of biomarker efficacy[**?**, **?**, **?**]. Additionally, the prediction framework provides a principled means for validating multivariate models that more accurately deal with the statistical dependencies between edges than mass univariate techniques, all while obviating the need to correct for multiple comparisons.

The general predictive framework involves learning a relationship between a *training* set of brain graphs and a corresponding categorical or continuous variable. The features for the brain graphs can be (1) a set of topological properties from each brain graph [**?**, **?**], (2) a vector embedding of the brain graphs [**?**, **?**, 11], or (3) the result of passing the brain graphs through a graph kernel []. The learnt model is then applied to an independent *testing* set of brain graphs to decode or *predict* their corresponding value of the variable. These values are compared to their "true" values to estimate *prediction accuracy* - a measure of how well the model generalizes to know data. Several different strategies can be employed to split the data into training and testing datasets, although leave-one-out cross-validation has high variance and should be avoided [].

A variety of different machine learning algorithms have been applied to analyzing brain graphs in this manner, but by far the most commonly employed has been support vector machines[**?**]. Although these methods offer excellent prediction accuracy, they are often black boxes, for which the information that is used to make the predictions is not easily discernable. The extraction of neuroscientifically meaningful information from the learnt model cab be added by employing sparse methods

and feature selection to reduce the input variables to only those that are essential for prediction. There is still considerable work to be performed in improving the extraction of information from these models, for developing techniques that permit multiple labels to be considered jointly, and developing kernels for measuring distances between graphs.

2.3.2 specificity, better controls

As a quick aside, it is important to keep in mind a few common analytical and experimental decisions that limit the utility of the putative biomarkers learned through predictive modeling. Generalization ability is most commonly used to measure the quality of predictive models, but since this measure doesn't consider the prevalance of the disorder in the population, it doesn't provide an accurate picture of how well a clinical diagnostic based on the model would perform. Instead it is important to estimate positive and negative predictive values *Grimes, Altman* using disease prevalence information from resources such as Centers for Disease Control and Prevention Mortality and Morbidity Weekly Reports. Also, the majority of neuroimaging studies are designed to differentiate between an ultra-healthy cohort and a single severely-ill population, which further waters down estimates of specificity. Instead it is also important to validate a biomarker's ability to differentiate between several different disease populations - a very understudied area of connectomes research. Lastly, most predictive modeling based explorations of the connectome are classifier based, which is very sensitive to noisy labels. Methods which incoporate some measure of label uncertainty or are robust to noisy labels are needed to help deal with this confound.

2.3.3 dimensions

[ALT: A PSYCHIATRIC NOSOLOGY IN CRISIS: CHALLENGES AND POTENTIAL] With the growing uncertainty about the biological validity of classical categorizations of mental health disorders, there is a growing focus on symptoms that can be measured dimensionally. This Research Domain Criteria (RDoC) has become a major focus of the NIMH, and will no doubt engender a major shift in the manner in which connectomes experiments are performed. In the context of predicitive modeling this translates into change in focus toward regression mdoels, which to date have been under utilizied in analyses of connectomes. But this dissatisfaction with extant clinical categories, opens up a new broad opportunity for redefining clinical populations based on their biology rather than their symptomatology.

[CONSIDER COMBINING THE DIMENSIONS AND BLOBBING PIECES; THEY ARE BOTH SOLUTIONS TO THE CHALLENGE OF REDEFINING THE PSYCHIATRIC NOMENCLATURE. FOLKS WHO PUSH ON DIMENSIONALITY ONLY SUFFER FROM THE FACT THAT THEY HAVE NO VISION THAT THE BIOLOGY AND/OR BETTER PHENOTYPING COMBINED WITH BIG DATA CAN BE USED TO REDRAW CATEGORIES].

Most prediction modeling in connectomes research has focused on classifier problems, with few studies using regression frameworks.

2.4 blobbing

A very under explored area of study is distinguishing between different popl

Predictive models are typically validated

Several limitations of neuroimaging data, and the manner in which predicitive modeling analyses are commonly employed, limit the utility of the putative biomarkers that they learn. Probably the foremost issue is that they are typically validated using cross-validation generalization accuracy, which does not consider disease prevalence in their calculation, and thus is not informative about how well the classifier would perform as a clinical diagnostic. For example, given a mental disorder with a high prevalence (ADHD, 7.2%) the probability of receving a false positive on a test with 100% sensitivity and 90% specificity is .56, and for less common disorders (Autism, 1%) the probability of a false positive becomes almost .91 [**?**, **?**]. Valuable information on prevalence of different disorders can be found from Centers for Disease Control and Prevention Mortality and Morbidity Weekly Reports.

Most neuroimaging studies of disease populations acquire data from an ultra-healthy cohort that is compared to a severely-ill population. Although this strategy is ideal for maximizing power in inferential statistics, it

2.5 Turtles all the way down

- Controversies for preprocessing - Role of graph alignment - Batch effects - Dimensionality reduction - Lack of a gold standard

With the growing prevalence of openly shared data, the desire to compare data acquired on different scanners, with different scanning protocols, and at different sites is becoming common. Although early studies have found that the amount of biological variation between individuals outweighs between site variation[**?**], there is still a need to standardize datasets against batch effects[**?**]. Ideally, this will entail the acquisition of calibration data, but until an accepted standard on what this calibration data will include exists and its collection becomes commonplace, post-hoc statistical corrections will be essential. Understanding the nature of batch effects and developing adequate correction strategies is in its infancy and there is substantial need for contributions in this area.

No matter what type of method is being applied, iFC analyses are plagued by the curse of dimensionality and an inability to validate their results with a gold standard. Functional MRI datasets typically involve hundreds or thousands of scans, each of which consist of a time series of hundreds of 3D brain volumes, that each contain measures from hundreds of thousands of brain locations (variables). Whether performing an analysis within a scan, or across scans, the number of variables (N) is much larger than the number of observations (P). Different analyses have approached the dimensionality problems in different ways, but there is no consensus on the best algorithm for brain data. Indeed there isn't a consensus about the best methods to use for any step of iFC analyses, and this lack of agreement is due to a lack of a ground truth or gold standard for comparing different techniques. Although several simulations have been proposed, there is always some skepticism about their accuracy and comprehensiveness. In their place, cross validation techniques have been utilized to compare methods based on some measure of reproducibility and/or generalizability.

A considerable drawback of using data-driven approaches for defining brain regions, is that the number of regions in the clustering solution, or conversely their size, must be specified. The choice of node size has been shown to have a considerable effect on the resulting graph topology [**?**] and can also affect the outcome of an analysis[**?**]. There are a variety of mechanisms for optimizing the number of brain areas in a parcellation, many of which using cross-validation methods, but in application they do not tend to converge to a single optimum, and instead provide a range of parcellations that are suitable[16]. The choice of parcellation used will strictly depend on the problem to be solved by that analysis and the amount of error that one can allow when interpreting the results.

3 MAKING DATA SHARING ACCESSIBLE TO DATA SCIENTISTS

As with other big data problems, the analysis of the type of brain data described above will benefit significantly from open science and open data. By open science we mean the process of making intermediate and end data products available to researchers outside the lab that initiated the investigation and by open data we mean data sharing, data reuse, allowing groups other than the ones that acquired the data to further analyze freely available data [17]. Both open science and open data have the benefit of facilitating the creation of new questions to answer from a data set that the people who acquired the data didn't conceive. A great example of this is the bioinformatics community, which came into existence due to the availability of large genomics datasets that were publicly available and could be used by anybody who possessed the skills and knowledge to mine them[18]. This brought scientists and researchers from disparate fields (ranging from biology, statistics, to computer science and engineering) together. The same can be aspired for the brain imaging community by making data acquired in different labs publicly available. At the moment there are a few public databases such as the fMRI data center [1], LONI IDA [2], 1000 Functional Connectomes Project (FCP)[3], the International Neuroimaging Data-sharing Initiative (INDI)[4], ABIDE preprocessed[5], Human Connectome Project [6], ADHD 200 preprocessed [7],CORR [8]. Among these, there are a few such as the data from the Human Connectome Project and the Cameron data set, which contain clean and tidy data. The latter will speed up knowledge discovery and new method development by alleviating the need for newcomers to acquire the knowledge and skills required to properly clean raw MRI data.

4 Conclusion

Functional connectomics is now a data science. As highlighted in this review, classic data science challenges related to data characterization, data reduction, and data

[1] fmri datacenter webaddress
[2] LONI webaddress
[3] http://fcon_1000.projects.nitrc.org
[4] INDI webaddress
[5] http://preprocessed-connectomes-project.github.io/abide/
[6] HCP webaddress
[7] http://neurobureau.projects.nitrc.org/ADHD200/Introduction.html
[8] http://fcon_1000.projects.nitrc.org/indi/CoRR/html/index.html

classification are rapidly emerging as rate-limiting steps for this burgeoning field and its promises to transform the clinical realm. It is our hope that the recent augmentation of open science data-sharing initiatives with preprocessing efforts explicitly focussed on the removal of of common barriers to entry for data scientists (e.g., domain-specific knowledge, computational demands of image processing), will serve as an undeniable invitation and request for the involvement of the broader data science community.

Competing interests
The authors declare that they have no competing interests.

Author's contributions
Text for this section . . .

Acknowledgements
Text for this section . . .

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

References

1. Biswal, B.B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C., Smith, S.M., Beckmann, C.F., Adelstein, J.S., Buckner, R.L., Colcombe, S., Dogonowski, A.-M., Ernst, M., Fair, D., Hampson, M., Hoptman, M.J., Hyde, J.S., Kiviniemi, V.J., Kötter, R., Li, S.-J., Lin, C.-P., Lowe, M.J., Mackay, C., Madden, D.J., Madsen, K.H., Margulies, D.S., Mayberg, H.S., McMahon, K., Monk, C.S., Mostofsky, S.H., Nagel, B.J., Pekar, J.J., Peltier, S.J., Petersen, S.E., Riedl, V., Rombouts, S.a.R.B., Rypma, B., Schlaggar, B.L., Schmidt, S., Seidler, R.D., Siegle, G.J., Sorg, C., Teng, G.-J., Veijola, J., Villringer, A., Walter, M., Wang, L., Weng, X.-C., Whitfield-Gabrieli, S., Williamson, P., Windischberger, C., Zang, Y.-F., Zhang, H.-Y., Castellanos, F.X., Milham, M.P.: Toward discovery science of human brain function. Proceedings of the National Academy of Sciences of the United States of America **107**(10), 4734–9 (2010). doi:10.1073/pnas.0911855107
2. Mennes, M., Biswal, B.B., Castellanos, F.X., Milham, M.P.: Making data sharing work: the FCP/INDI experience. NeuroImage **82**, 683–91 (2013). doi:10.1016/j.neuroimage.2012.10.064
3. Van Essen, D.C., Ugurbil, K.: The future of the human connectome. NeuroImage **62**(2), 1299–310 (2012). doi:10.1016/j.neuroimage.2012.01.032
4. Jiang, T.: Brainnetome: a new -ome to understand the brain and its disorders. NeuroImage **80**, 263–72 (2013). doi:10.1016/j.neuroimage.2013.04.002
5. Assaf, Y., Alexander, D.C., Jones, D.K., Bizzi, A., Behrens, T.E.J., Clark, C.a., Cohen, Y., Dyrby, T.B., Huppi, P.S., Knoesche, T.R., Lebihan, D., Parker, G.J.M., Poupon, C., Anaby, D., Anwander, A., Bar, L., Barazany, D., Blumenfeld-Katzir, T., De-Santis, S., Duclap, D., Figini, M., Fischi, E., Guevara, P., Hubbard, P., Hofstetter, S., Jbabdi, S., Kunz, N., Lazeyras, F., Lebois, A., Liptrot, M.G., Lundell, H., Mangin, J.-F., Dominguez, D.M., Morozov, D., Schreiber, J., Seunarine, K., Nava, S., Riffert, T., Sasson, E., Schmitt, B., Shemesh, N., Sotiropoulos, S.N., Tavor, I., Zhang, H.G., Zhou, F.-L.: The CONNECT project: Combining macro- and micro-structure. NeuroImage **80**, 273–82 (2013). doi:10.1016/j.neuroimage.2013.05.055
6. Sporns, O., Tononi, G., Kötter, R.: The human connectome: A structural description of the human brain. PLoS computational biology **1**(4), 42 (2005). doi:10.1371/journal.pcbi.0010042
7. Hagmann, P.: From diffusion MRI to brain connectomics. PhD thesis (2005)
8. Craddock, R.C., Jbabdi, S., Yan, C.-G., Vogelstein, J.T., Castellanos, F.X., Di Martino, A., Kelly, C., Heberlein, K., Colcombe, S., Milham, M.P.: Imaging human connectomes at the macroscale. Nature methods **10**(6), 524–39 (2013). doi:10.1038/nmeth.2482
9. Fu, Z., Di, X., Chan, S.-C., Hung, Y.-S., Biswal, B.B., Zhang, Z.: Time-varying correlation coefficients estimation and its application to dynamic connectivity analysis of fMRI. In: Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE, pp. 2944–2947 (2013)
10. Zalesky, A., Cocchi, L., Fornito, A., Murray, M.M., Bullmore, E.: Connectivity differences in brain networks. NeuroImage **60**(2), 1055–1062 (2012). doi:10.1016/j.neuroimage.2012.01.068
11. Craddock, R.C., Holtzheimer, P.E., Hu, X.P., Mayberg, H.S.: Disease state prediction from resting state functional connectivity. Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine **62**(6), 1619–28 (2009). doi:10.1002/mrm.22159
12. Dosenbach, N.U.F., Nardos, B., Cohen, A.L., Fair, D.a., Power, J.D., Church, J.a., Nelson, S.M., Wig, G.S., Vogel, A.C., Lessov-Schlaggar, C.N., Barnes, K.A., Dubis, J.W., Feczko, E., Coalson, R.S., Pruett, J.R., Barch, D.M., Petersen, S.E., Schlaggar, B.L.: Prediction of individual brain maturity using fMRI. Science (New York, N.Y.) **329**(5997), 1358–61 (2010). doi:10.1126/science.1194144
13. Richiardi, J., Eryilmaz, H., Schwartz, S., Vuilleumier, P., Van De Ville, D.: Decoding brain states from fMRI connectivity graphs. NeuroImage **56**(2), 616–26 (2011). doi:10.1016/j.neuroimage.2010.05.081
14. Bogdanov, P., Dereli, N., Bassett, D.: Learning about Learning: Human Brain Sub-Network Biomarkers in fMRI Data. arXiv preprint arXiv: . . . (2014). arXiv:1407.5590v1
15. Richiardi, J., Achard, S., Bunke, H., Ville, D.V.D.: Machine Learning with Brain Graphs (April), 58–70 (2013)

16. Craddock, R.C., James, G.A., Iii, P.E.H., Hu, X.P., Mayberg, H.S.: A whole brain fMRI atlas generated via spatially constrained spectral clustering. Human Brain Mapping **33**(8) (2012). doi:10.1002/hbm.21333.A
17. Milham, M.P.: Open neuroscience solutions for the connectome-wide association era. Neuron **73**(2), 214–8 (2012). doi:10.1016/j.neuron.2011.11.004
18. Van Horn, J.D., Gazzaniga, M.S.: Why share data? Lessons learned from the fMRIDC. NeuroImage **82**, 677–82 (2013). doi:10.1016/j.neuroimage.2012.11.010