Network Construction, Partitioning Methods, and Connectivity Measures for Analysis of Neural States Across Subjects

A Thesis
Submitted to the Faculty
in partial fulfillment of the requirements for the
degree of

Bachelor of Arts

in

Mathematics

by Rachel Matthew

Advised by Professor Peter Mucha
Dartmouth College
Hanover, NH

April 2024

Abstract

In this thesis, we survey Pearson correlation, Granger causality, and multivariate autoregressive measures of functional and effective connectivity to explore a novel application of network science—through the Leiden, CHAMP, and pruning algorithms—to whole-brain functional magnetic resonance imaging (fMRI) scans and to produce insights on a real data set centered on the task of cognitive regulation during exposure to aversive stimuli. Networks of increased connectivity during cognitive regulation of aversive stimuli are identified: intra-connectivity in medial superior frontal gyrus and inter-connectivity between the basal ganglia and cortex. We also introduce a new measure of functional connectivity, the Pearson correlation of the derivatives of blood-oxygen-level-dependent (BOLD) signals across regions, and with it identify two additional regions with increased inter-brain connectivity during regulation: the vagal nerve nuclei and the left parainsular area.

Acknowledgements

I will start, and likely end, by thanking my primary thesis and research advisor, Peter Mucha, for everything. I will elaborate on this to say thank you for talking to me when I showed up randomly at your office because Anne Gelb referred me, and for continuing to talk to me after that, for consistently expressing excitement at my ideas and progress, for believing in me, and for reading this thesis so thoroughly so late in the game. I can be very timid, and you have made it much easier to keep, stated simply, *doing things* which I otherwise might not have done.

This last statement might apply even more so to Peter Doyle, because his belief in me so far exceeds my own that I feel surprised yet honored with every email I receive containing a new opportunity or question. It is possible that I would not have had the confidence to enter Dr. Mucha's office had I not had the enriching experience of researching under you, and I want to give special thanks for your habit of reading everything I write, even when I didn't write it with the expectation of it being read. It makes me feel like my thoughts

are valuable.

I will further my thanks to Anne Gelb, Dimitris Giannakis, and Prasad Jayanti for believing in me, Dana Williams for helping me sort out this math-neuroscience-complex systems major, Šarūnas Burdulis for patiently resetting my doob password each time I asked, and Asher Auel for taking the time one class to be a teacher beyond just being an instructor by telling us that everyone, at some point in pursuing math, will eventually reach a point at which it is hard. Thank you to Richard Granger, for believing I could do good work back in 2019, before anyone else had the chance to, and to Charles Palmer for easing my mind about my future before I dove face-first into writing this thesis.

I will refrain from thanking classmates to encourage myself to thank them in person. Thank you once more to Peter Mucha, the third and most recent Peter at Dartmouth who has made me feel like I could be an impressive person, and to the other two Peters as well.

I could not imagine any of this a year ago.

Contents

Chapter 1

Introduction

In 2022, Tor Wager from the Department of Psychological and Brain Sciences, Marta Čeko, and three other authors published a paper [7] in *Nature Neuroscience* using functional magnetic resonance imaging (fMRI) scans of 55 subject's brains during a "multimodal aversiveness task," in which a subject is exposed to a sequence of twenty 6–10sec stimuli with pauses between each stimulus for the subject to rate the preceding stimulus's aversiveness and recover. Each subject performed this task in two different sessions, six runs—continuous periods of scanning during which a subject performs the task—per session. In 2023, Wager and Čeko granted me access to this data.

I chose to write my thesis on my analysis of this data out of a desire to pursue my interest in the brain, which led me to the Complex Systems modified major, and performed that analysis with methods that align with my interest in mathematics, which led me to choose

to major in the Mathematics Department. I consider the data from a network approach, in this case according to both the mathematical and neuroscientific definitions of the word, to evaluate neuronal connectivity but also assess the validity of applying novel (mathematical) network analysis techniques to fMRI data. My aim, in analyzing real data with this methodology, is to produce insights on both the data I analyzed and on workflows which could be generalized to future datasets.

1.1 Outline

In the following chapters, I will review the concepts of neuronal connectivity and network partitioning that informed my work, cover the methods I applied to the dataset, and present what I found. My connectivity analysis is limited to functional and effective connectivity paradigms—anatomical data is not used—and my partition analysis is limited to extensions of modularity-based methods. I employ machine learning and statistical methods to evaluate the results, and I conclude with the limitations of the work and avenues for its continuation in the future.

Chapter 2

Background

2.1 Neuronal Connectivity

"Functional connectivity" (FC) is the concept of temporal correlations between distinct regions of the brain. "Effective connectivity" (EC) is the concept of causal relationships between activity in distinct regions of the brain. Because both are concepts and not strict measures, there exist multiple methods by which FC and EC may be evaluated, and these methods are non-trivially different. The following subsections will detail those methods which I employ in this thesis.

2.1.1 Pearson Correlation

One of the most common methods in the neuroscience literature for evaluating functional connectivity between two brain regions is

the Pearson correlation [4]. For brain regions x and y, the Pearson correlation estimate of functional connectivity between the regions is given by:

$$\rho_{x,y} = \frac{\sum_{i=1}^{n}(\frac{X_i-\mu_X}{\sigma_X})(\frac{Y_i-\mu_Y}{\sigma_Y})}{n} = \frac{cov(X,Y)}{\sigma_X\sigma_Y}$$

where we use the convention of capital letters to denote time series in brain regions indicated by the corresponding lower-case letter, e.g. R denotes the time series of the average blood-oxygen-level-dependent (BOLD) signal across the voxels corresponding to region r, n is the length of that time series, μ_R is the mean of R, and σ_R is the standard deviation of R [8]. Note that, in fMRI analysis, the BOLD signal is assumed to be a proxy for neural activity. The limitations of this assumption are beyond the scope of this thesis.

Typically, Pearson correlation coefficients for fMRI data are acquired as described above; however, this method only represents certain aspects of the data, even in the domain of temporal correlation. Early in this project, my intuition in this domain was that, because the brain is a complex system in which many parts interact with many others simultaneously at every given moment, even region pairs which are very strongly coupled may appear more weakly coupled because so many *other* regions influence each, in a sense "drowning out" the correlation. More sophisticated methods for approximating connectivity in spite of this are included in the following

sections, but I also make a distinction in this thesis between "Type 1" (T1) and "Type 2" (T2) Pearson correlations, as follows.

T1, or "standard," Pearson correlations are computed as described above. A T2 Pearson correlation coefficient between any two regions x and y is the standard Pearson correlation of the *derivatives* of X and Y, i.e.:

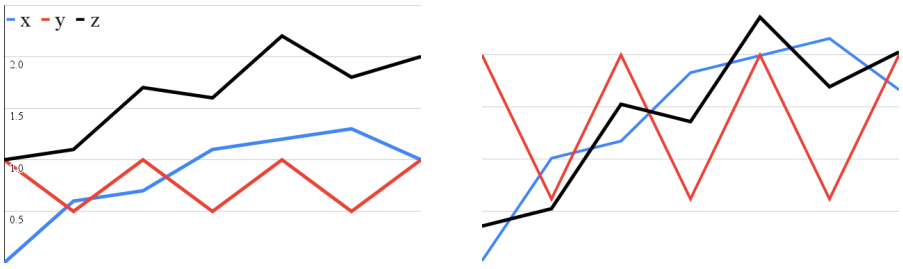
$$\rho'_{x,y} = \frac{cov(X', Y')}{\sigma_{X'}\sigma_{Y'}}$$

where $\rho'_{x,y}$ is the T2 Pearson correlation between regions x and y, and the R' notation denotes the sequence of differences between each non-starting entry and its previous entry in the R time series—simply put, the derivative of R (up to a multiple of the time step).

To properly illustrate both how these correlations differ and to provide a more refined intuition behind why it might be important to test both, I will use the following toy example (visualized in Figure 2.1). Let:

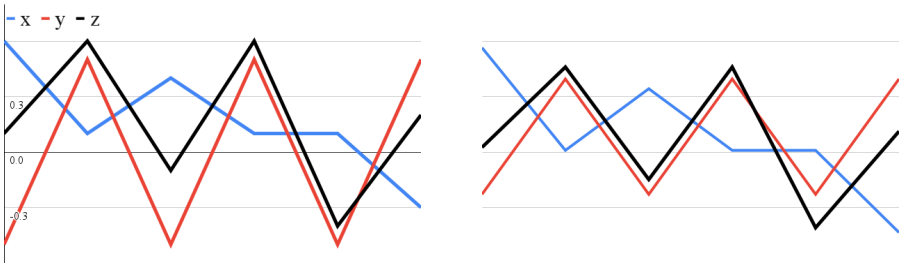
$$\begin{aligned} x(t) &= X_t \in X = (0, 0.6, 0.7, 1.1, 1.2, 1.3, 1), \\ y(t) &= Y_t \in Y = (1, 0.5, 1, 0.5, 1, 0.5, 1), \\ z(t) &= x(t) + y(t). \end{aligned}$$

Simply looking at the plotted sequences, it is clear that both x and y can be considered to "contribute" to z, with the global changes in the sum sequence z reflecting those in x but a local pattern of activity mirroring that in y. However, when the sequences are normalized,



(a) Sequences Plotted (default) (b) Sequences Plotted (normalized)

Figure 2.1: Visualization of the sequences. The black line indicates the sequence z, the blue x, and the red y.



(a) Sequence Derivatives Plotted (default) (b) Derivatives Plotted (normalized)

Figure 2.2: Visualization of the derivative sequences. Note that the increased similarity of the normalized z sequence has flipped from greater resemblance to x to greater resemblance to y.

z's adherence to the global pattern of x results in a high value of $\rho_{x,z}$ (R=0.821, p=0.024) and a negligible $\rho_{y,z}$ (R=0.272, p=0.555).

Compare this to the derivative sequences in Figure 2.2. The derivative sequence of z more closely captures the alternating pattern of y, and thus the correlation coefficients "switch," with $\rho'_{x,z}$ being low, and negative, (R=-0.209, p=0.691) while $\rho'_{y,z}$ is high (R=0.836, p=0.038).

Neither is necessarily a "better" measure than another in the *a priori* sense; each captures information the other neglects. Addi-

tionally, each measure has similar properties as a metric of functional connectivity: both $\rho_{x,y}$ and $\rho'_{x,y}$ are coefficients representing an undirected correlation between activity at region x and activity at region y. The higher the coefficient, the greater we interpret functional connectivity between those regions to be. Other methods, by contrast, are directed and attempt to capture a causal relationship. These are metrics of *effective* connectivity.

2.1.2 Granger Causality

Any two region time series X and Y may be evaluated for the Granger causality of X on Y and of Y on X. An assessment of Granger Causality provides two pieces of information. First, the binary response to the query: Does X Granger-cause Y? Second, to what degree is X informative in predicting Y? Numerically, the response to the second question can be presented as the corresponding F-statistic [10] because it corresponds to the improvement in predicting future values of Y when X is considered vs. when X is not. The more influential the activity of region x is on the future activity region y, the more a prediction of region y's future value will be improved by including region x's past values in the prediction. The converse—that the greater the improvement, the greater the influence—does not necessarily follow, but it can be the case in practice often enough to provide insight into the relationships between involved regions.

2.1.3 Multivariate Autoregression

Multivariate autoregressive (MAR) modeling can be applied to fMRI data. In this application, we fit the data to a model in which future values of each region's time series are a linear combination of all of the past m values in every region time series. Non-linear dynamics may be included by including additional time series in the input to each future-value prediction, eg. via the introduction of bilinear terms [12].

2.2 CHAMP and Pruning

Each of the above-discussed measures of functional connectivity can be combined into a network representation. For example, the correlation coefficients $\rho_{x,y}$ between regions x and y may be combined into a weighted network of connections between brain regions, with the correlation coefficient $\rho_{x,y}$ providing the weight (possibly after additional manipulation such as taking absolute values) of the edge connecting region x and y. "Networks" here are effectively synonymous for our purposes to the mathematical concepts of "graphs" and their generalizations to, possibly, weighted, directed, and/or multilayer situations. Because the language in the network neuroscience literature aligns more closely with that of network science, we will use that terminology throughout.

2.2.1 Leiden Algorithm

A common procedure for analyzing network data is to calculate the "community structure" of a network. The most commonly used methods for community detection partition the nodes of the network into assortative communities with greater connectivity within the communities compared to that between communities.

The Leiden algorithm [20] for community detection is a modularity-based method. Modularity, H, is computed for network N and a given partition on N as:

$$H = \frac{1}{2m} \sum_{c} \left(m_c - \gamma \frac{K_c^2}{2m} \right)$$

where m is the number of edges in N, c is a community in the given partition, m_c is the number of edges in c, K_c is the sum of the degrees of all nodes in c, and γ is the "resolution parameter." Alternately, in a weighted network like those analyzed in this thesis, m is the sum of the weights of every edge in N, m_c is the sum of the weights of all the edges in c, and K_c is the sum of the *strengths* of all nodes in c. A popular algorithm for optimizing the modularity of a network is the Louvain algorithm.

The Louvain algorithm starts by assigning each node in the network to its own community, then proceeds through the repetition of two phases. In the first phase, the algorithm iterates through each node and computes the change in modularity that would result from

swapping the node from its current community to the community of each of its neighbors. If one such swap would yield an increase in modularity, then the option with the greatest gain in modularity is chosen; otherwise, the node remains. This iteration repeats until no further increases are possible.

In the second phase, the communities themselves are made into a new network in which each node i is community c_i from the previous network and each edge e_{ij} is weighted as the sum of the weights of every edge between one node in community c_i and another in community c_j. The first phase is then repeated on this new network, then the second, and so forth until none of the nodes in the most recent community-network can be merged without a resulting loss in modularity [3].

The Leiden algorithm is a modification of the Louvain algorithm that aims to solve the problem of poorly connected communities, which Louvain may yield. It inserts a *refinement* phase between the first, node-moving phase and the final, aggregating phase. The refinement phase proceeds similarly to the node-moving phase, but with both more stochasticity and more guidance. Iteration through the nodes is limited to those in the neighborhoods of nodes which have been moved. There is an added rule: nodes that did not end up in the same community after the node-moving phase cannot be placed in the same community in the refinement phase. Accordingly, the output of the refinement phase is a partition with no less than the

number of communities resulting from the node-moving phase, but potentially more because rather than a node being moved greedily to the community with the greatest modularity gain, it is moved probabilistically—the greater the modularity gain, the more likely the move—to any community which would result in a modularity gain [20].

2.2.2 CHAMP

One notable feature of both the Louvain and Leiden algorithms is the stochasticity. Two runs of the Leiden algorithm on a sufficiently large and well-connected network are almost guaranteed to produce different results. Additionally, both the Python and R implementations of the Leiden algorithm retain a property in principle available to all modularity-based community detection methods: the resolution parameter γ. Changing γ can drastically alter the results of the Leiden algorithm. For any fully connected network with n nodes, the resolution parameter can be varied from low enough to result in only one community to high enough to result in n communities every time the algorithm is run. It is clear that a *useful* value of γ would fall between these extremes, but within that range of "valid" values, results can still differ meaningfully.

A method to account for both randomness in the Leiden algorithm and the question of how to select γ is the Convex Hull of Admissible Modularity Partitions (CHAMP) algorithm [24]. The

algorithm takes a network and a set of partitions as input, and it determines, across a specified range of γ-values, which of the provided partitions has the highest modularity at each γ. For example, for the range $0.0 \leq \gamma \leq 2.0$, one partition might be the "best"—highest in modularity—for $0.0 \leq \gamma < 0.97$ and another for $0.97 < \gamma \leq 2.0$. In this case, we would say there were two "admissible" partitions for this γ range. This can be computed quickly, as a manipulation of the modularity function demonstrates:

$$\begin{aligned} H &= \frac{1}{2m} \sum_c \left(m_c - \gamma \frac{K_c^2}{2m} \right) \\ &= \frac{1}{2m} \sum_c (m_c) - \gamma \sum_c \left(\frac{K_c^2}{4m^2} \right) \end{aligned}$$

Thus, for any given partition p, we may compute constants $a_p = \frac{1}{2m} \sum_c (m_c)$ and $b_p = \sum_c (\frac{K_c^2}{4m^2})$, such that:

$$H(\gamma) = a_p - \gamma b_p$$

Because this is true for every p in the partition set, determining the partition with the highest modularity at $\gamma = 0$ requires only identifying the partition with the highest a_p. Call this partition p_0. Then for every partition p_i, we may compute the value of γ where the modularity of p_i equals that of p_0 with:

$$\gamma(H(p_i) = H(p_0)) = \frac{a_{p_0} - a_{p_i}}{b_{p_0} - b_{p_i}}$$

Partition p_0 ceases to be the optimal partition at the minimum of the positive next γs calculated in this way, and the operation may be repeated for p_i to determine the next γ, and that the next, and so on until all partitions optimal on a pre-specified range of γ-values has been found.

Through this process, we obtain a subset of "admissible" partitions from an original set which ranges from tens to thousands of times smaller than the original [24]. This allows a larger number of γ-values to be sampled in generating partitions without creating an unwieldy quantity of data, and without losing high-quality partitions.

2.2.3 Pruning

The CHAMP algorithm remains subject to "researcher choice" regarding the *range* of values of gamma on which partitions are generated and, more significantly, evaluated. At $\gamma = 0$, for example, a fully connected graph will always have the highest modularity when every node is in one community. This one-community partition would thus be the result of the CHAMP algorithm, in spite of having minimal utility in analysis and relationship to the network's structure, unless the human individual performing the analysis chooses a higher value of gamma to start, and *which* value is chosen can vary the results. Then, even once the range of γ is selected, the CHAMP results only identify different intervals within

that range, with each interval corresponding to a single partition dominant in that interval. While this sometimes makes a preferred selection of a large interval and its corresponding partition, in many cases deciding which interval(s) to focus on is another researcher choice.

"Pruning" the output of the CHAMP algorithm is a way to reduce the impact of researcher choice in the numerical data analysis. It derives from Newman's observation [17] of the equivalence between the equation for modularity as we have presented it here and that of the log-likelihood of a planted partition degree-corrected stochastic block model, i.e. a stochastic block model in which the expected number of edges between any two nodes is the product of those two nodes' degrees, a constant corresponding to one of two values—θ_{in} if the nodes are in the same community, θ_{out} otherwise—and the constant term $\frac{1}{2m}$ *if* $\gamma = \frac{\theta_{in}-\theta_{out}}{\ln\theta_{in}-\ln\theta_{out}}$.

The parameters θ_{in} and θ_{out} can be estimated for a given partition of a network with:

$$\theta_{in} = \frac{2m_{in}}{\sum_c K_c^2/(2m)}, \qquad \theta_{out} = \frac{2m_{out}}{\sum_{c\neq g} K_c K_g/(2m)}$$

where m_{in} is the number of edges between nodes in the same community, m_{out} is the number of edges between nodes in different communities [17, 11].

Thus, from each partition a γ-value can be acquired that best

aligns the modularity maximization with the maximum likelihood fit of a planted partition degree-corrected stochastic block model, and furthermore, if CHAMP is run on a suitably broad range of values, there will be an admissible partition corresponding to that γ-value. This admissible partition has its own derivable γ-value, and so on until we reach a partition that either leads to itself with its γ-value or is part of a loop of partitions that point cyclically to each other. In this way, pruning—collecting only those partitions which are stable or periodic under this process—restricts partitions to a subset implied by the structure of the data (insofar as it is well matched by the assumed underlying planted partition model), and not by a researcher's preference.

Chapter 3

Methods

3.1 The Data

The fMRI data I worked with came preprocessed in the Neuroimaging Informatics Technology Initiative (NIfTI) file format. It comprises a total of 660 scans from 55 individuals, twelve scans per individual, each scan containing 387 to 950 "slices"—time-points separated by a repetition time of 0.46 seconds—with 72,017 voxels in each slice. Before the data could be fit, I had to perform one final preprocessing step of reslicing it; reslicing was done with SPM12. To reduce the data's size, both for the purpose of rendering it more workable and disregarding voxels outside of the brain, I applied the CANlab atlas [6] to the data to collapse the 72,017 points in 3D space into 489 regions of interest (ROIs). The typical scan, once processed in this manner, could be represented by a 920x489 ma-

trix.

Of each subject's twelve scans, six correspond to the "runs"—approximately seven-minute periods of continuous scan data—that were performed during the "Experience" session and six correspond to the runs of the "Regulate" session. The subjects are engaged in the 'multimodal aversiveness task' [7] in both sessions. During the task, they are presented with one of five types of stimuli: painful pressure, heat, unpleasant sounds, unpleasant images, or pleasant images. They are then asked to rate the aversiveness of the stimuli, time passes, and the next stimulus appears. Each run presents each type of stimuli a total of four times, with the order of presentation randomized for each run.

The significance of the "session" signifier is how the subjects were asked to respond to the task. In the six "Experience" runs, the subjects were told to simply experience the stimuli without special instruction. In the "Regulate" runs, subjects were instructed to perform a "cognitive self-regulation strategy" [7]. This change in strategy affected their ratings—the "Regulate" strategy resulted in lower evaluations of aversiveness—so the question follows: how does the "Regulate" strategy manifest in the brain? This is the real-data question I intend to tackle with the methods I describe below.

3.2 Network Construction

Each network N was constructed from one scan, with 489 nodes corresponding to the CANlab atlas's 489 ROIs. Each network N is undirected, dense, and weighted. The weights of the edges correspond to a measure of connectivity.

3.3 Verifying Utility

Because I compared many methods of assessing connectivity in this thesis, it was necessary to acquire a metric for their performance. In the following sections and chapters, I will refer often to this metric as "predictive quality" or "performance." Specifically, I assess the utility of a connectivity measure by its accuracy in a simple machine learning test. I train a basic multilayer perceptron classifier on the task of predicting the session signifier of a scan—"Experience" or "Regulate"—from its neural connectivity measures.

The classifier architecture was kept simple and constant regardless of the measure being evaluated. It takes the estimated connectivity between every unique undirected pair of distinct regions as its input vector, a vector of 119,316 elements, passes the data through three hidden layers, each of size 16 and activated by ReLu, then outputs a single float ranging from 0 to 1, with 0 corresponding to an "Experience" prediction and 1 to "Regulate." The classifier is trained on 75% of the data and tested on the remaining 25%, with

10% of the training data set aside for validation. It trains for a hard maximum of 100 epochs but stops early if the validation score does not increase by at least 1e-4 for 10 epochs in a row. In practice, the classifier never ran for the full 100 epochs. The optimizer used was Adam with a L2 regularization term of 1e-4.

No preprocessing was performed on the data prior to training for the sake of improving the classifier's performance, and I do not employ any sophisticated machine learning techniques to improve performance. The purpose of the "predictive quality" test is to determine how readily the cognitive states associated with a session signifier can be extracted from a measure of connectivity, not to perfect a machine learning model.

3.4 Measures of Functional Connectivity

Having outlined how connectivity measures were assessed, I proceed to present the connectivity measures I examined.

3.4.1 Pearson Correlation

"T1" and "T2" Pearson correlation coefficients were computed via the df.corr() function from the python package *pandas*. I surveyed the performance resulting from handling negative coefficients via setting them to zero, taking their absolute value, or squaring them, and I found that squared coefficients resulted in the best predic-

tive quality; therefore I only used squared coefficients in subsequent analysis.

3.4.2 Granger Causality

F-statistics for Granger Causality were computed via the "Granger Causality Test" app in MATLAB. Preliminary tests of predictive quality found that the performance of the measure didn't improve with an increase in maximum possible lags. Thus, subsequent analysis utilizes F-statistics acquired from Granger causality computations with a lag of one. To create an undirected network using the directed F-statistics, the weight for each pair of region i and region j was set to $e_{ij} = \frac{F_{ij}+F_{ji}}{2}$ where F_{ij} is the F-statistic for the Granger causality of region i on region j.

3.4.3 MAR

MAR was assessed via linear regression and the Lasso model using python's *scikit-learn* package. The Lasso model—a linear model regularized with the L1 prior—was run with an L1 regularization term of 5e-5 for a maximum of 1,000 iterations.

3.5 CHAMP & Pruning Parameters

For each measure, I ran the CHAMP algorithm on the corresponding network representation for 400 evenly-spaced values of γ ranging

from 0.5 to 2.5. The choice to start at $\gamma = 0.5$ instead of $\gamma = 0$ derives from the fact that, in the case of this particularly dense data, a γ-value of 0.5 was always a sufficiently low value to guarantee the generated partitions started at the single-community partition, therefore ensuring no low-γ partitions were excluded by virtue of γ starting too high. The choice to not generate beyond γ=2.5 similarly derives from an observed property of this dataset. Many partitions generated with a higher γ-value than 2.5 comprised predominantly singleton communities, and all of them pruned to a lower γ-value, thus ensuring no high-γ partitions were excluded by virtue of γ ending too low. The partitions were generated with the cluster_leiden function from R-igraph, and CHAMP and pruning were performed using R code written and refined in Professor Mucha's lab.[1]

3.6 T-Tests

A more straightforward method, but important for meaningful analysis nonetheless, that I employed was the statistical t-test. To search for differences in connectivity between the cognitive states of "Experience" and "Regulate," I ran a two-tailed t-test for each connectivity measure for each (undirected) pair of regions, comparing the set of coefficients obtained from all "Experience" scans to those from all "Regulate scans"—across runs and across subjects. Because the

[1] Additionally, as part of my project work in Mucha's lab, I developed R code for CHAMP and pruning of multilayer networks.

number of pairs to consider is large (119,316), it would be exceptionally improbable to *fail* to acquire many pairs with a P-value less than 0.05 for each measure, entirely by pure chance. To account for this, I only considered pairs whose t-test resulted in a P-value of 4.25e-7 of lower, the reason for this value being that $1 - (1 - 4.25e - 7)^{119,316} = 0.0494 < 0.05$, ergo the chance of acquiring even one result per connectivity measure with this degree of significance is less than five percent, thus matching a sufficiently rigorous test of statistical significance.

Chapter 4

Results

4.1 The Survey of Connectivity Measures

The "predictive quality" of each of the evaluated measures is displayed in Figure 4.1. For each measure, three metrics are displayed: average train, test, and "restricted test" accuracy. The first two are typical machine learning classifier measures, the latter being the more important of the two; train accuracy is the proportion of samples in the 75% of the set used for training and validation which the model correctly classifies as either "Experience" or "Regulate." Test accuracy is this proportion, but of the 25% of the data not shown to the classifier during training.

Average restricted test accuracy is the average of test accuracies only over the iterations where the classifier achieved a train accuracy above 90%. Thus, this third metric represents the ability of classi-

fiers that successfully learn a given input to *generalize*. The metric was employed following a qualitative observation that typically the worst train and test accuracy scores—those near 50%—co-occur within iterations, suggesting that the models producing their scores failed to fit to the data in the first place, and therefore the low test accuracy of such a model does not provide useful data on the ability of the model to generalize. Averages for each input set tested were collected from 100 iterations of randomly-generated train-test splits.

As can be observed from Figures 4.1 and 4.2, the measures of effective connectivity—Granger causality, linear regression, and Lasso model regression—fit best to the training data. However, in generalizing to novel input data, the measures that performed best were T2 Pearson and Granger Causality. The Lasso data performed similarly to these two measures in test accuracy, but when comparing restricted test accuracy, the measure's performance decreases, implying these data yielded higher test results because the model more frequently converged, not because they generalize as well as the T2 and Granger measures.

Linear regression in particular exhibited a pattern of over-fitting. The model reached an average value exceeding 90% on the training data, but less than 60% on the test set. T2 Pearson, Granger Causality, and Lasso, by contrast, performed consistently above 80% on every metric. In contrast to this, the T1 Pearson measure was *less* than 80% on every metric. Thus the T2 type of Pearson cor-

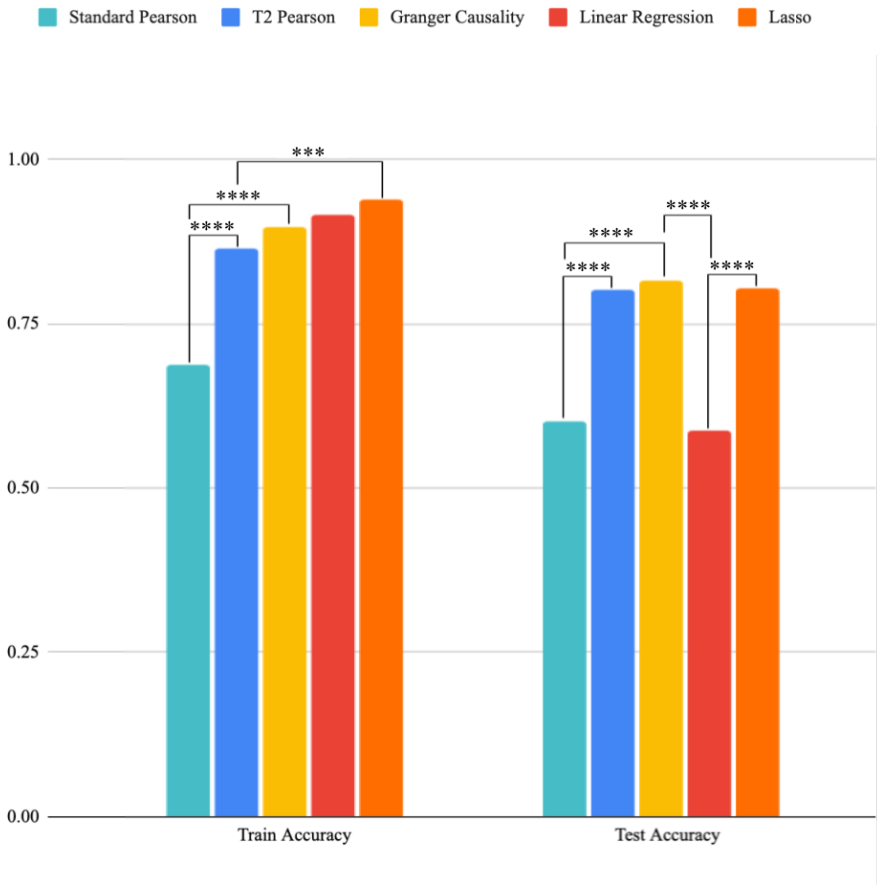


Figure 4.1: Averages of train and test accuracy scores. P-values are denoted as follows: * for p<0.05, ** for p<0.01, *** for p<0.001, and **** for p<1e-6.

relation out-performed the standard by every metric. This held for the "raw" data—not processed beyond the coefficients—and when pruned, as will be discussed in the following section (see Figure 4.3). I further discuss this difference in performance in the subsection *Out-performance of T2 over T1.*

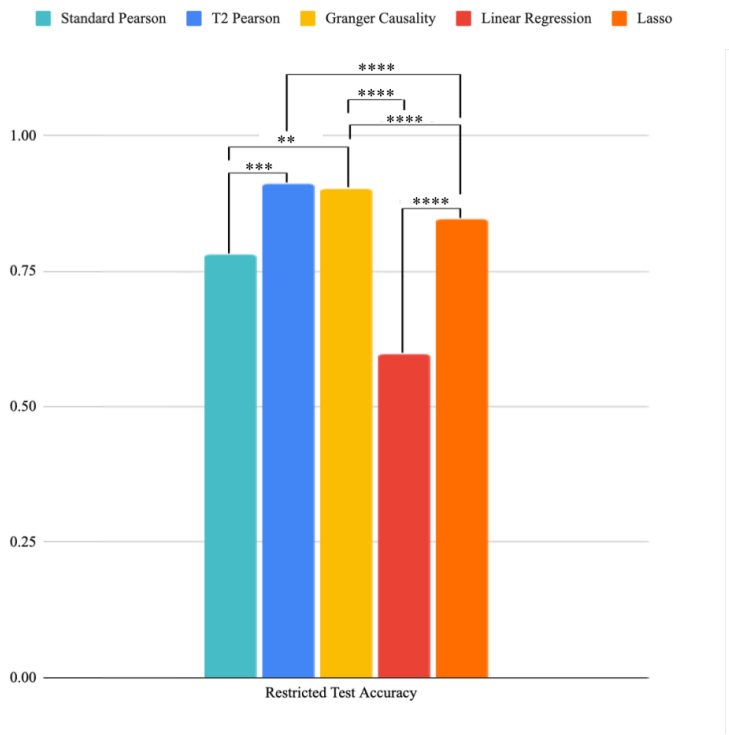
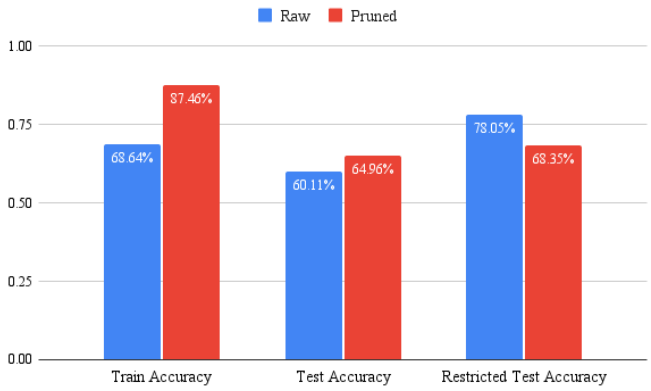


Figure 4.2: Averages of restricted test accuracy scores. P-values are denoted as follows: * for p<0.05, ** for p<0.01, *** for p<0.001, and **** for p<1e-6.

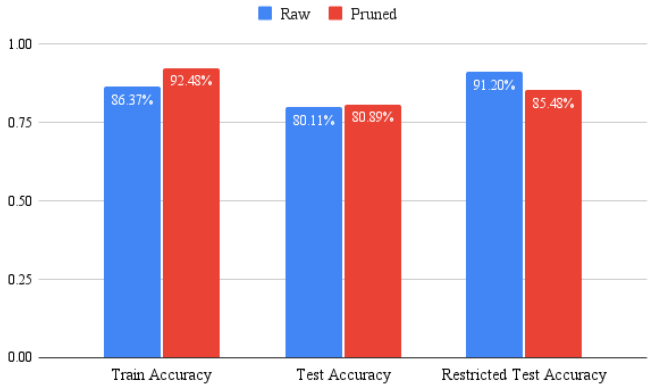
4.1.1 Effect of Pruning on Classifier Performance

Performance of pruning results was assessed, like "raw" results, with accuracy scores on the classification task. In the case of pruned data, the "coefficients" used as input to the classifier are computed as the frequency with which the regions in a pair are placed together in the pruned partitions. If pruning returned only one partition for a scan—which occurred for just over one third of all T1 Pearson-evaluated scans and almost half of all absolute-value T2 Pearson-evaluated scans—then all the coefficients for that input would be either 1 or 0: 1 if the regions in the corresponding pair are in the same community and 0 otherwise. If pruning returned three partitions, then the coefficients could be 0, $\frac{1}{3}$, $\frac{2}{3}$, or 1. This restricted the number of possible of values more for Pearson correlation data than Granger causality data as the latter tended to result in greater numbers of pruned partition per network (μ=4.805) compared to the former (μ=2.001).

Overall, the classifier fails to converge less often on pruned data, resulting in higher train accuracy scores, but it also generalizes to test data more poorly. An explanation for this result may emerge from the properties of the pruned data when input to a multilayer perception model. Pruned data can be said to be "cleaner" than raw data; the values within it range from 0.0 to 1.0, and many are exactly zero. Because the multilayer perceptron classifier acquires values for

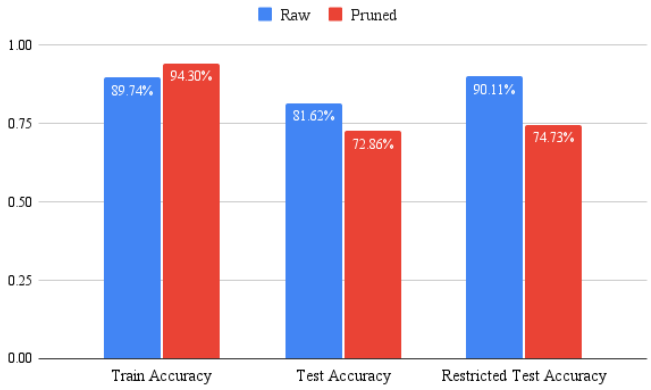


(a) T1 Pearson

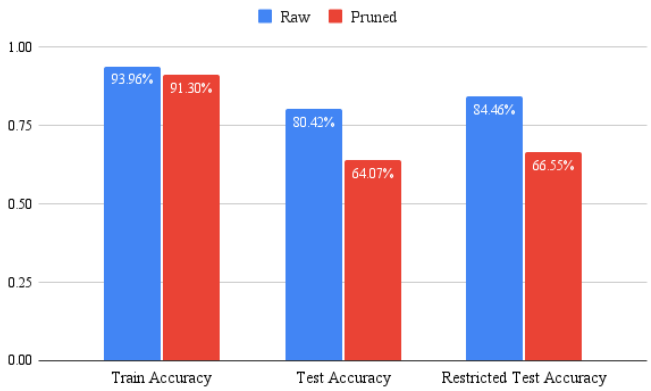


(b) T2 Pearson

Figure 4.3: A comparison of accuracy scores of "raw" to pruned data for the measures of functional connectivity surveyed. $p<0.005$ for each with the exception of the comparison between test accuracies for raw and pruned T2 data.



(a) 1-lag Granger Causality



(b) Lasso Regression

Figure 4.4: A comparison of accuracy scores of "raw" to pruned data for the measures of effective connectivity surveyed. $p<0.0001$ for all test and restricted test accuracy score comparisons, and $p<0.05$ for the Granger train accuracy score comparison.

each layer via matrix multiplication followed by the ReLu activation function, entry values of zero have no effect on the next layer. Thus an input with many entries equal to zero has only a subset of the model's weights actively contributing to the prediction. Over-fit happens more readily in such a case.

In support of this theory is the fact that, typically, the more entries with a zero value a pruned input set has, the greater the drop in test accuracy between the raw and pruned data tends to be. The average number of communities in a pruned partition resulting from a Pearson correlation is 9.22. Compare this to the average number of communities from a Granger or Lasso pruned partition, 36.18 and 22.72 respectively. When we match these differences to Figures 4.3 and 4.4, we can observe how T1 and T2, the data sets with fewer communities—therefore more region pairs in the same community, therefore more non-zero values in the pruned input data deriving from the data set—have an increase in test accuracy when pruned and a decrease in restricted test accuracy of approximately 5–10%, but the Granger and Lasso have a drop in test accuracy and decrease in restricted test accuracy by more than 15%.

4.1.2 Out-performance of T2 over T1

The significantly enhanced performance of the T2 Pearson correlation measure over the T1 is clearly illustrated in Figure 4.1. The T2 measure has greater predictive quality than the T1 in training, test-

ing, and restricted test accuracies, all to a statistically significant degree.

The reason for this difference in performance is difficult to extrapolate from this single instance. The Pearson correlation of the the derivatives of two region time series from an fMRI scan may be a useful measure in a general case, or it might simply correspond to this specific predictive task for this specific dataset.

Because T2 Pearson correlation is a cheap measure to compute and performs well on this dataset-task pair, I utilized it as the default measure of connectivity when extending my analysis. In one such extension, I created a set of "stimulus-split" networks. These were created with one network N per scan *per stimulus*. Each scan was split into the five types of stimuli presented—aversive pressure, heat, unpleasant sound, unpleasant image, and pleasant image [7]—through the use of "annotations." I processed .mat data I was given on the structure of the runs to mark, for each slice in each run, the type of stimulus that the subject either experienced at the time the slice was taken or, if no stimulus was present at that time, the type of the most recent prior stimulus. A stimulus-scan network was then obtained with nodes corresponding to regions, as I described above for all the networks in *Network Construction*, and edge weights computed as the squared T2 correlation coefficient of region-pairs *only over the slices annotated as the stimulus of interest*. Thus, I acquired five networks per scan instead of one, but none of the networks were

derived from overlapping data.

In terms of predictive accuracy, this stimulus-split T2 Pearson correlation set performed well. For a typical network structure with hidden layers of size 16, it averaged 93.02%, 91.96%, and 98.64% for training, testing, and restricted test accuracies, respectively.

4.2 Subject Specificity

Having acquired such high accuracy scores with the stimulus-split T2 Pearson correlation data, it naturally follows that the result becomes suspicious. I considered explanations for the result: either an increased number of data samples benefited training, given that splitting each scan according to stimulus type quintupled the sample number, or it could be the case of data leakage due to the fact that, although none of the data used in one network was used in another, generating the stimulus-split networks required using data from the same *file*, the same "run," in different networks. It was theoretically possible that each run possessed particular properties that would enable a machine learning algorithm to fit to those properties, and not those of the "Experience"/"Regulate" cognitive states. For example, one might imagine that seeing the "aversive sound" network from Subject 36's 3rd run in the "Experience" session taught the classifier what Subject 36's 3rd run of the "Experience" session looked like in general, contributing to an accurate classification of

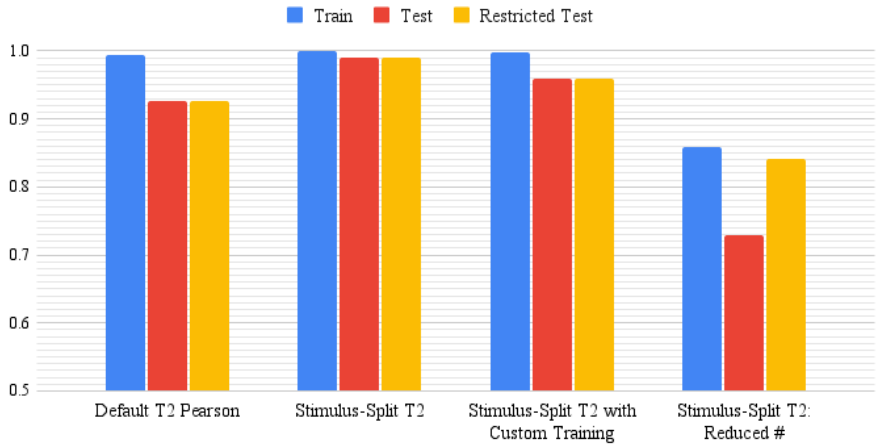


Figure 4.5: Train, test, and restricted test accuracy scores of the default T2 Pearson correlation data vs. different methods of training on the "Stimulus-split" T2 data. For this specific test, a classifier with three hidden layers of size 64, instead of 16, was employed. Training was otherwise the same. The "Custom Training" is as described in the document. The "Reduced #" randomly selected a number of samples from the Stimulus-Split set equal to the number of samples in the default T2 Pearson data.

the heat, pressure, unpleasant, and pleasant image networks of the same scan.

To determine which hypothesis holds, I subjected the classification test to a minor modification. Instead of using train_test_split from *scikit-learn*, I wrote a custom function to separate the data such that if any scan from subject S's i^{th} run appeared in the train set, then *every* scan from subject S's i^{th} run appeared in the train set, and vice versa. The results are displayed in Figure 4.5.

Given the minor difference between the performance of the classifier under the default split vs. the custom split—and the notable decrease in performance when the number of samples available in

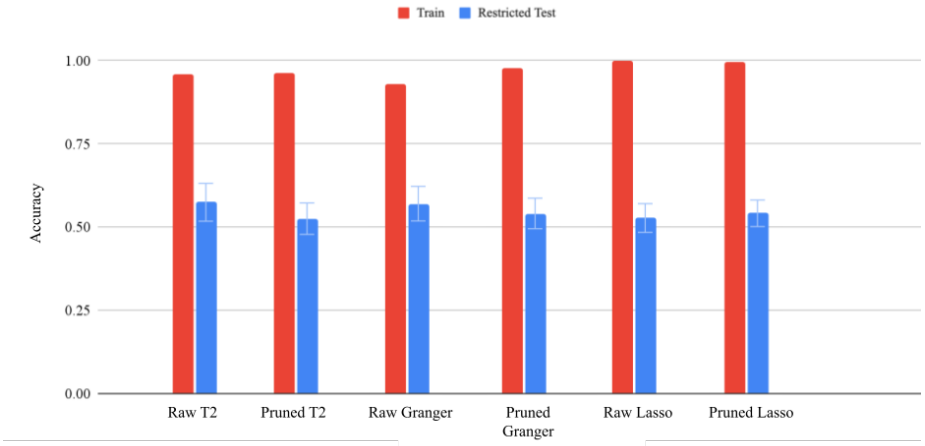


Figure 4.6: Restricted test accuracy scores vs. train accuracy scores when testing occurs on subjects held out of the training set.

the stimulus-split set was reduced to that resembling other input sets—the "increased sample size" hypothesis was more likely; however, the possibility that the classifier achieves high accuracy without learning features of the "Experience" vs. "Regulate" cognitive states remained. To test the classifier's ability to generalize, I modified the custom split function to write a *"subject split"* function. The function guarantees that if any scan of subject S appeared in the train set, then *every* scan of subject S appeared in the train set, and vice versa.

With these distinctions made, I re-ran the classification test on the data. The results appear in Figure 4.6.

While none of the restricted test accuracy scores collected under these conditions dropped below 50%, all dropped close to it and the average scores for Raw Lasso, and Pruned T2 and Granger, fell

within one standard deviation of 50%.

There are a number of reasons the classifier may fail to generalize across subjects. A clear risk of an input of size 119,316 is overfitting, even if the hidden layers are far smaller.[1] We must also consider the nature of the data's collection. Each subject came in for two sessions, with each session separated from the other by over a day, but every run per subject-session pair was recorded in close temporal proximity [7]. As a result, a sufficiently powerful machine learning model might learn to distinguish the connectivity patterns characterizing the runs in each subject-session pair, regardless of what led to the patterns. A subject's brain might look different during her "Regulate" session because she has a song stuck in her head, because she is more relaxed at this time of day and therefore shifting less in the scanner, because that day's higher temperature causes the scanner to record slightly differently, etc. It is possible that the classifier learns cognitive states, but that it learns 80 of them instead of 2. It is also possible that some of what it learns is not a cognitive state at all, but various conditions of the fMRI machine. Most likely, it is a combination of both.

[1] As a counterpoint, this problem of failure to generalize occurred even for significantly pared-down versions of the data, including PCA component coefficients in the hundreds instead of hundreds of thousands.

4.3 Statistical Analysis

A machine learning classifier of "Experience" vs. "Regulate" fMRI scans generalizes poorly to novel subjects. However, this does not mean that the input to the classifier lacks meaningful data. The excellent generalization of the stimulus-split T2 Pearson correlation input to novel runs demonstrates that the input *does* contain meaningful data, but the opacity of machine learning models prevents a human observer from extracting that meaning. A t-test is a simple and clear way to identify differences between groups. Performing and comparing t-tests across the measures assessed grants direct insights into the variable of interest: "Experience" vs. "Regulate."

As addressed in *Methods*, I discuss here, for each included connectivity measure, only the pairs of regions whose connectivities differed between "Experience" and "Regulate" scans with a probability of 4.25e-7 or lower of the difference being by chance. A brief overview of the number and qualities of such pairs for each connectivity measure considered is presented in Table 4.1.

Although there is some overlap between the regions that appear in these significantly different pairs under the metrics tested, they are by no means equivalent, and the pairs themsleves overlap very little between measures. Between the two raw Pearson measures, only connectivity between Bstem_Ponsrd_L (the left rostrodorsal pons) and nuc_ambiguus_R (the right nucleus ambiguus) is

Features of T-Test Pairs

Measure	N	MFR
T1 (Raw)	13	anterior medial SFG, visceral sensorimotor cranial nerve nuclei (eg. solitary tract nucleus, nucleus ambiguus, dorsal motor nucleus)
T1 (Pruned)	0	N/A
T2 (Raw)	44	**dorsal motor and solitary tract nuclei**, cerebellum, visual cortex
T2 (Pruned)	138	**parainsular area (L)**, amygdala, raphe nuclei, supramarginal gyrus, Brodmann area 6, brainstem, frontal operculum
Granger (Raw)	174	**ventral tegmental area (R)**, orbitofrontal complex (L), basal ganglia, cerebellum, brainstem, opercular cortex (especially lateral belt and parietal operculum)
Granger (Pruned)	0	N/A
Lasso (Raw)	0	N/A
Lasso (Pruned)	0	N/A

Table 4.1: N is the number of region pairs that reached the required degree of statistical significance. MFR, i.e. "Most Frequent Region(s)" is a brief, descriptive list of the brain regions which appeared most in the pairs that were significant. (L) indicates left side specifically, and (R) indicates right. If applicable, bolded text indicates the one to two regions that appeared the absolute most often, at least three times more than the second most frequent. (SFG=Superior Frontal Gyrus)

significantly different in both, with, for both measures, an average Pearson connectivity of approximately 0.02 in "Experience" and 0.05 in "Regulate" (p=0.0102 in T1, p=0.0008 in T2, where p is scaled up to $1-(1-p)^{119,316}$ to account for the 119316 pairs considered). Between raw T2 Pearson and Granger, only two pairs were shared: Ctx_PEF_L (left premotor eye field) with Bstem_Med_L (left medulla) and Ctx_PGs_L (the left parietal area G superior) with dmnx_nts_R (right dorsal motor and solitary nuclei). Both pairs exhibited greater connectivity in "Regulate" scans. No pairs that appeared in Granger appear in T1 Pearson, and none of the pairs found from the pruned T2 Pearson co-occurred in any of sets collected from the other metrics. Lasso resulted in zero pairs, and therefore also shared zero pairs with each other measure. This difference between different measures supports the idea that, although every measure represents "connectivity," meaningful distinction can be made between the data resulting from differents methods of evaluating connectivity.

4.3.1 Differences in Connectivity between Cognitive States

One overwhelming trend across all measures is that significantly different connectivities tended to be higher in the "Regulate" condition than in "Experience." Of the total 369 pairs across measures, in only three was the average "Experience" score higher than the "Regulate." From the pairs collected from raw T2 Pearson data,

Ctx_MIP_L (a section of left intraparietal sulcus involved in arm movement) with Cblm_Vermis_CrusI (cerebellar vermis and Crus I) and Ctx_Pir_L with Ctx_AAIC_L (the left piriform cortex and bordering anterior agranular insular cortex) were more strongly paired in "Experience" (0.0237 vs. 0.0129, corrected p=0.0257 for the former, 0.3324 vs. 0.2296, corrected p=0.0029 for the latter). From the pairs collected from pruned T2 Pearson data, only one pair was more connected in "Experience": Ctx_Pir_L and GPi_R (left piriform cortex and right globus pallidus internus)(0.4329 vs. 0.2172, corrected p=0.0281). We note that the left piriform cortex, a region heavily involved in olfactory processing and implicated in temporal lobe epilepsy [22, 19, 9], appears in multiple of these exceedingly rare examples of greater "Experience" connectivity. Region Ctx_Pir_L does appear with increased connectivity in "Regulate" in exactly two pairs collected from the Granger data, partnered with two regions in the right basal ganglia (STN_R and VTA_R), both with corrected p<0.0001. Because the right globus pallidus internus is also a right-side region of the basal ganglia, a universal interpretation regarding the "general" connectivity of the left piriform cortex between the "Experience" and "Regulate" is unclear; however, because Granger causality is a method of assessing effective connectivity and Pearson correlation a method of functional connectivity, we may conclude from the data that the piriform cortex is more weakly functionally paired with the basal ganglia when cognitively regulating aversive

stimuli compared to simply experiencing them, but more strongly causally affiliated. Note that the right piriform cortex appears only in the statistically significant pairs collected from the Granger measure, where it more strongly connected during "Regulate" with the subthalamic nucleus, right ventral tegmental area, and lobule IV of the cerebellar vermis.

Proceeding from this analysis of the rare instances where pairs were more strongly connected during the "Experience" cognitive state, I observe another difference between the effective and functional connectivity measures. While pairs from raw T1 and T2 involve left-side regions at slightly more than double the rate they involve right-side regions, in the Granger pairs, the reverse is true. Pruned T2 matches neither, with an approximately equal number of regions from both sides. The differences between measures is significant enough that it makes most sense organizationally to discuss each separately for the rest of this section.

Standard Pearson Pairs

At the smallest number of significantly different pairs, T1 Pearson is the simplest to analyze. The pairs predominantly involve two broader areas of the brain: the medial frontal lobe and the brainstem. Greater T1 Pearson functional intraconnectivity occurs within these regions when a person is in the "Regulate" cognitive state compared to "Experience," particularly amongst regions in the medial

superior frontal gyrus (mSFG), between regions in the mSFG and the left anteroinferior insula, and between the left pons and right nucleus ambiguus. The SFG, particularly the left—recall that T1 pairs are dominated by left-side regions—is involved in working memory [5] and in semantic processing [15], while the nucleus ambiguus is involved in the motor function of speech [18]. The pons plays a significant role in pain signaling [16].

T2 Pearson Pairs

Approximately two thirds of all raw T2 Pearson pairs are between a cortical region and either the left or right dorsal motor and solitary nuclei. Many of the remaining third involve either a cerebellar region or the left orbitofrontal complex. Two involve the right nucleus ambiguus.

In overview, the results of the t-test selected pairs from the T2 Pearson correlation data indicate that the vagal nerve nuclei—the dorsal motor nucleus, nucleus ambiguus, and the solitary nucleus—exhibit higher functional connectivity with many regions of the rest of the brain while the "Regulate" cognitive strategy is employed. The vagal nerve nuclei receive and send signals to the heart, stomach, glands, lungs, mouth, throat, and are responsible for coughing, the gag reflex, and slowing heart-rate and breathing during a parasympathetic nervous response [2]. Intuitively, one could understand how greater coupling between these nuclei and the rest of the

brain might occur during or aid in maintaining a regulatory cognitive state, where the subject exercises mindfulness and aims to keep calm while experiencing a stimulus, compared to when the same subject experiences the stimulus as it comes.

Granger Causality Pairs

The number of pairs yielded from the Granger causality data renders a concise analysis difficult. To simplify my overview of the insights provided, I will focus only on the pairs between regions which appear, between both regions in the pair, at least ten times within the pair set. This may exclude meaningful pairs between less frequently appearing regions from the overview, so I will also discuss regions known to be involved in pain regulation where they appear.

Forty-seven of the seventy-four pairs meeting the first criterion given contain one of the following three regions, all right-side: the ventral tegmental area (VTA), parabrachial pigmented area (a substructure of VTA), and subthalamic nucleus (STN). Both VTA and STN are part of the basal ganglia, and both have been implicated in responding to aversive stimuli [13, 23]. The VTA is, incidentally, one of the regions known to be involved in pain regulation which would be important to check when analyzing these results. In the pairs, VTA, the parabrachial pigmented area, and STN are more strongly connected with the lateral parietal lobe and various operculoinsular regions in the "Regulate" cognitive state. After VTA and STN, the

orbitofrontal complex also appears frequently.

Focusing on regions known to be involved in pain regulation, I check pairs containing the anterior cingulate cortex (ACC), amygdala, VTA, nucleus accumbens, periaqueductal gray (PAG), and dorsal raphe (DR). This is not a comprehensive list of every region involved in pain regulation, but it is a survey of many key ones. There are 54 Granger pairs containing one of the listed regions, the vast majority corresponding to those already discussed in the paragraph above. Excluding the ventral tegmental area from the search criteria still results in two pairs containing it, as right VTA has greater connectivity during "Regulate" with the left nucleus accumbens and the basolateral amygdala. Of the remaining six pairs, two connect the basolateral amygdala with the left and right anterior precuneus, two connect the superficial amygdala to regions in the right basal ganglia—the globus pallidus internus and the subthalamic nucleus—and the final two connect the PAG with the left ventral pallidum, another region in the basal ganglia, and with a right-side medial region in Brodmann area 5. As we saw when considering the pairs occurring between most frequently appearing regions, the Granger pairs show that, during cognitive regulation of pain, the basal ganglia are more effectively connected to many other regions in the brain, both cortical and basal. This may emerge due to greater communication happening with these regions in the process of coordinating a regulative cognitive response.

4.3.2 Effect of Pruning on Emergence of Differences

An interesting contrast between each included measure of connectivity emerged when the t-tests were run on their pruned versions. All measures except for T2 yielded no significantly different pairs, while the number of pairs significantly *increased* between the raw and pruned T2 sets. One explanation is that the "sparser" the data, the fewer significantly different pairs emerge. Lasso, even when raw, produced zero pairs, and Granger, which produced many when raw, produced none after pruning and, as observed previously, pruned Granger data contains far more zeros than pruned data from either T2 or T1 Pearson. This is not the whole picture—T1 also reduced in number of pairs, and the high number of pairs from raw Granger derived from some property of the data other than its sparsity—but it is likely a contributing factor.

In analyzing the pairs from the pruned T2 data, I employ the same methods I used for the Granger due to the larger number of pairs. Every single pair between regions occurring more than a total of ten times contains the left parainsular area (area PI). Area PI exhibited greater "Regulate" connectivity to regions in the operculum and frontal gyri. Both the operculum and frontal lobe of the brain are involved in a number of complex cognitive processes [14, 1], while the insula, to which the parainsular area belongs, plays a role in salience and attention [21].

Of the fourteen pairs containing regions implicated in pain regulation, four involve the centromedial amygdala, three the amygdalostriatal area, and one the basolateral amygdala, all generally connecting to regions, left or right, within the premotor, sensory, or motor cortices. Of the six remaining pairs, three included the ventral tegmental area, two the dorsal raphe, and one the PAG with the right ventral midbrain.

Due to the prominence of the insula and amygdala in these pairs, a prominence not observed in raw T2, the network of enhanced connectivity suggested by the pruned T2 pairs is distinct from the vagal nerve nuclei–centric network of enhanced functional connectivity indicated by the non-pruned data. The reason for, and significance behind, this difference is subject to future study.

4.3.3 Predictive Quality Generalization

I can check the utility of these pairs in the same way I checked the utility of the connectivity measures on their own: through their prediction quality. Particularly, because these pairs were extracted as meaningfully different features between "Experience" and "Regulate" cognitive states, I would expect that they would be useful for generalization. Thus, I tested the sets of pairs by restricting the input of the machine learning task—using the Stimulus-Split T2 data set as a base—to only connectivity between pairs collected from the t-tests. This reduces the size of the input data from 119,316 to a number in

		Accuracy (%)			
Pair Source	N	Train	Control Train	Test	Control Test
T1	13	67.25	57.24	59.23	54.66
T2 (Raw)	44	78.81	66.80	66.65	53.05
Granger	137	89.88	81.87	55.22	51.18
T2 (Pruned)	171	89.92	83.02	58.21	51.35
All	362	95.99	90.46	59.39	52.29

Table 4.2: Average accuracy scores for pair-reduced stimulus-split T2 data on session classification task, tested on subjects excluded from training data. N indicates the numbers of region pairs to which the input data was restricted. The p-values of all comparisons between pair-reduced and control randomly-reduced accuracies were less than 1e-6.

the hundreds or less. I perform this reduction and train on the result for each set of pairs acquired and for a "super-set" containing all pairs acquired across the different measures. For comparison, I also perform the classification task on the data set randomly reduced to a number of input features equal to that which the pair reduction produces. The results are displayed in Table 4.2.

The resulting improvements in performance are modest but statistically significant. It is also worth noting that, though patterns of over-fit do occur when subjects are held out for testing, the results of the predictive task, both in this section and the previous, provide utility. Patterns that hold between runs capture data pertaining to a cognitive state even when the complexity of the state's composition greatly exceeds the variable of interest. The patterns of connectivity identified through the t-tests, and the associated improvement in the predictive task, demonstrate that insight on the variable of interest may still be, and has still been, extracted.

Chapter 5

Conclusion, Discussion, and Future Work

Over the course of this thesis, I surveyed three distinct measures of functional and effective connectivity, introduced the "T2" variant of the Pearson correlation measure of functional connectivity, and built and analyzed networks with each. In exploring the differences across these measures, and under the effects of the Leiden, CHAMP, and pruning algorithms, between two cognitive states—"Experience" and "Regulate"—I identified four distinct hubs of increased connectivity during cognitive regulation of pain: the medial superior frontal gyrus, the vagal nerve nuclei, the basal ganglia, and the left parainsular area.

5.1 Limitations

The analysis I performed with pruning was constrained to single-layer networks. This was due to limitations in time and computing power—with larger multilayer networks running at rates that would require over eighteen years to reach completion—that could be overcome with enhanced computing power or with sufficiently expert alterations to the code. I suspect insights could be drawn from the data when viewed with a temporal dimension. Each run of fMRI scanning lasted for approximately seven minutes. An individual's mind may operate differently at the beginning of seven minutes of near-continuous aversive stimuli compared to the end, and this difference could relate to the cognitive strategy they employed during the elapsed minutes.

5.2 Future Directions

5.2.1 Modeling

In the domain of time, I performed several tests with modeling near the end of the spring of 2024. This was unfortunately too late to synthesize the work into the body of this thesis, but I believe further insights could be drawn from future analysis. The model took advantage of the fundamental differences between functional and effective connectivity measures with to combine both; for each scan,

for each pruned partition, I created a representation of the brain using "macro-regions," where each macro-region corresponds to a community of the partition, with an activity time series equal to the average of those of the community's constituent regions. I then used these macro-region time series, the annotations of the stimuli at each time-point, and noise data extracted during the fMRI scan preprocessing to train a linear model regularized with the L1 prior, a Lasso model, on the first 700 slices of the 920-slice scan. The model's predictions for neural activity in each macro-region were then tested on the 220 slices of the scan.

With this procedure, it was possible to generate a "projection" using the model, where the target macro-region's activity for the next hundred seconds is estimated entirely from the stimulus, noise, and *other* macro-regions, but no data from the target macro-region itself. Often, these predictions would be poor. The system of the mind, which in practice has significantly more nuance in its dynamics than even the 72,017 voxels of the original scan can capture, is harshly simplified in the model. However, the predictions were not always poor. For most runs, at least one partition would have at least one community whose activity could be more than 50% predicted ($R^2 > 0.5$). Figure 5.1 provides a visual reference of the projection for one such community.

Further exploration on any insights yielded by the model would require testing beyond what I completed, which was a comparison

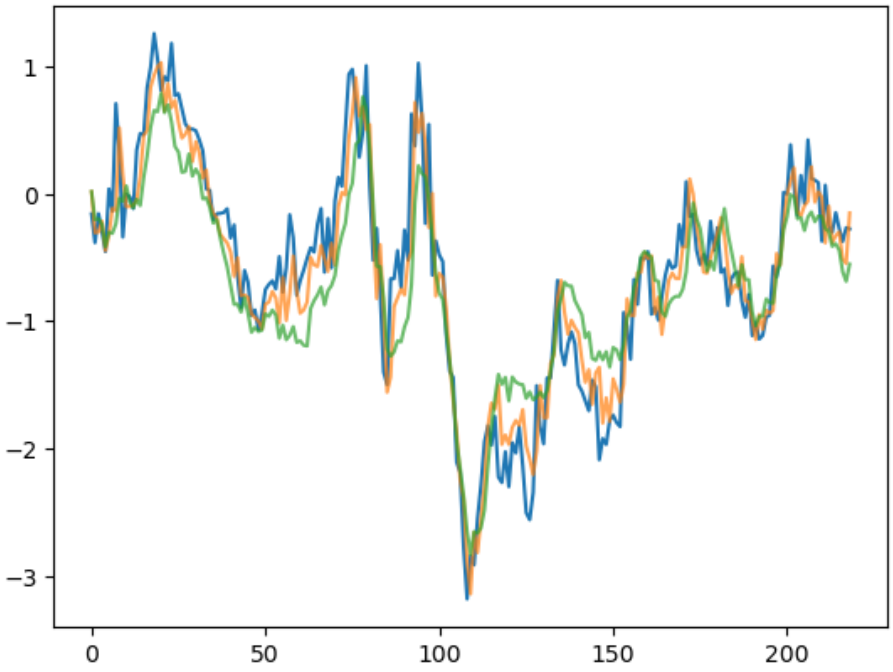


Figure 5.1: The true activation data is shown in blue, the "default fit" (where value x_t of the sequence is predicted using the true value of x_{t-1}) of the model in orange, and the projection (where x_t is predicted using the previous predicted value of x_{t-1}) in green. The value of R^2 for this projection was 0.627.

of the predictable macro-regions with the predictability of random time series generated to rudimentarily resemble the time series of macro-regions and with the predictability of artificial macro-regions generated by randomly grouping regions. The former comparison yielded no sufficiently predictable "macro-regions" while the latter yielded a 50% increase in the number of macro-regions, the former validating the significance of the occurrence of such predictable regions and the latter potentially implying that measures of connectivity may have more in common than the t-tests would imply, if we assume it becomes more difficult to predict one area when we cannot use the most similar areas as reference.

Interestingly, the regions that most often belonged to a predictable macro-region differed between the modeled and second randomized case, with the former including lobule X of the cerebellum, the spinal trigeminal, locus coeruleus, basal ganglia, and parainsular area—most of these being familiar given the results reported in this thesis, involved in pain regulation, or both—while the latter was dominated by cortical areas, particularly visual cortex and the lateral parietal lobe. I would have liked to continue this exploration further—examining the regions most strongly causal to the predictable macro-regions would have been the next step—but for the time being, it remains work for the future.

5.2.2 The T2 Pearson Measure

In addition, the T2 Pearson correlation metric of functional connectivity is an avenue for future work. As far as I have been able to find, there is no mention of assessing functional connectivity via a Pearson correlation of the derivatives of region time series, and given the measure's predictive quality here, I wonder if it might be promising on other data sets. The possibility remains that its performance in this case derived from a unique property of the multimodality aversiveness task or another particular feature of this data set, but verifying the measure's utility or lack thereof would require tests on a novel data set, and my work here was restricted to the data to which I had access. I am hopeful regarding the generalizability of the T2 measure, as it would not be difficult to apply to a set of fMRI data collected on scans for a different task, or even resting state, and it displayed among the highest predictive quality of all measures surveyed.

Bibliography

[1] Rami M El-Baba and Mark P Schury. "Neuroanatomy, frontal cortex". In: (2020).

[2] Elliott Baker and Forshing Lui. "Neuroanatomy, vagal nerve nuclei". In: (2019).

[3] Vincent D Blondel et al. "Fast unfolding of communities in large networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10 (Oct. 2008), P10008. DOI: 10.1088/1742-5468/2008/10/P10008. URL: https://dx.doi.org/10.1088/1742-5468/2008/10/P10008.

[4] Sebastian Bobadilla-Suarez et al. "Measures of neural similarity". In: *Computational brain & behavior* 3 (2020), pp. 369–383.

[5] Foucaud du Boisgueheneuc et al. "Functions of the left superior frontal gyrus in humans: a lesion study". In: *Brain* 129.12 (2006), pp. 3315–3328.

[6] *CANLAB - COMBINED ATLAS-2018*. URL: https://sites.google.com/dartmouth.edu/canlab-brainpatterns/brain-atlases-and-parcellations/2018-combined-atlas.

[7] Marta Čeko et al. "Common and stimulus-type-specific brain representations of negative affect". In: *Nature neuroscience* 25.6 (2022), pp. 760–770.

[8] David Freedman, Robert Pisani, and Roger Purves. "Statistics (international student edition)". In: *Pisani, R. Purves, 4th edn. WW Norton & Company, New York* (2007).

[9] M. Galovic et al. "Association of Piriform Cortex Resection With Surgical Outcomes in Patients With Temporal Lobe Epilepsy". In: *JAMA Neurol* 76.6 (June 2019), pp. 690–700.

[10] John Geweke. "Measurement of Linear Dependence and Feedback Between Multiple Time Series". In: *Journal of the American Statistical Association* 77.378 (1982), pp. 304–313. ISSN: 01621459. URL: http://www.jstor.org/stable/2287238 (visited on 05/21/2024).

[11] Ryan A Gibson and Peter J Mucha. "Finite-state parameter space maps for pruning partitions in modularity-based community detection". In: *Scientific Reports* 12.1 (2022), p. 15928.

[12] L Harrison, W.D Penny, and K Friston. "Multivariate autoregressive modeling of fMRI time series". In: *NeuroImage* 19.4 (2003), pp. 1477–1491. ISSN: 1053-8119. DOI: https://doi.org/10.1016/S1053-8119(03)00160-5. URL: https://www.sciencedirect.com/science/article/pii/S1053811903001605.

[13] Elizabeth N Holly and Klaus A Miczek. "Ventral tegmental area dopamine revisited: effects of acute and repeated stress". In: *Psychopharmacology* 233 (2016), pp. 163–186.

[14] Mihai-Dragoş Mălia et al. "Functional mapping and effective connectivity of the human operculum". In: *Cortex* 109 (2018), pp. 303–321.

[15] Mengjie Meng et al. *The Effect of Disturbance on the Neural Mechanisms of Learning Word Formation Rules in a Novel Language*. Mar. 2024. DOI: 10.21203/rs.3.rs-4015255/v1.

[16] Vitaly Napadow, Roberta Sclocco, and Luke A Henderson. "Brainstem neuroimaging of nociception and pain circuitries". In: *Pain Reports* 4.4 (2019), e745.

[17] M. E. J. Newman. "Equivalence between modularity optimization and maximum likelihood methods for community detection". In: *Phys. Rev. E* 94 (5 Nov. 2016), p. 052315. DOI: 10.1103/PhysRevE.94.052315. URL: https://link.aps.org/doi/10.1103/PhysRevE.94.052315.

[18] Bogdana Petko and Prasanna Tadi. *Neuroanatomy, Nucleus Ambiguus — ncbi.nlm.nih.gov*. July 2023. URL: https://www.ncbi.nlm.nih.gov/books/NBK547744/#:~:text=The%20nucleus%20ambiguus%20is%20the,formation%20in%20the%20medulla%20oblongata..

[19] D. Steinbart et al. "Automatic and manual segmentation of the piriform cortex: Method development and validation in patients with temporal lobe epilepsy and Alzheimer's disease". In: *Hum Brain Mapp* 44.8 (June 2023), pp. 3196–3209.

[20] Vincent A Traag, Ludo Waltman, and Nees Jan Van Eck. "From Louvain to Leiden: guaranteeing well-connected communities". In: *Scientific reports* 9.1 (2019), p. 5233.

[21] Lucina Q Uddin et al. "Structure and function of the human insula". In: *Journal of Clinical Neurophysiology* 34.4 (2017), pp. 300–306.

[22] D. N. Vaughan and G. D. Jackson. "The piriform cortex and human focal epilepsy". In: *Front Neurol* 5 (2014), p. 259.

[23] David B Weintraub and Kareem A Zaghloul. "The role of the subthalamic nucleus in cognition". In: *Reviews in the Neurosciences* 24.2 (2013), pp. 125–138.

[24] William Weir et al. “Post-Processing Partitions to Identify Domains of Modularity Optimization”. In: *Algorithms* 10.3 (Aug. 2017), p. 93. ISSN: 1999-4893. DOI: 10.3390/a10030093. URL: http://dx.doi.org/10.3390/a10030093.