Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

__________________________ __________________________

Xinyi Yang Date

A Novel Network Connectivity Measure with Application to Multimodal Brain Imaging Study

By

Xinyi Yang

Master of Science in Public Health

Department of Biostatistics and Bioinformatics

Ying Guo, Ph.D.

Committee Chair

Benjamin Risk, Ph.D.

Committee Member

A Novel Network Connectivity Measure with Application to Multimodal Brain Imaging Study

By

Xinyi Yang

B.S., Nankai University, 2016

Thesis Committee Chair: Ying Guo, Ph.D.

An abstract of

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health,

in Department of Biostatistics and Bioinformatics

2019

Abstract

A Novel Network Connectivity Measure with Application to
Multimodal Brain Imaging Study

By Xinyi Yang

Background: Network-oriented analysis on functional connectivity from functional magnetic resonance imaging (fMRI) data is not easy due to the large scale of fMRI data. The traditionally used Pearson correlation is not ideal because it is sensitive to indirect connectivity. The recently proposed partial correlation also has shortcomings because of both the loss of the "indirect" connectivity information and the requirement of a prescribed tuning parameter. Thus, a novel measure for functional network connectivity is proposed, in order to better discribe functional connevtivity and get rid of the predetermined tuning parameter.

Methods and Analysis: The distance derived from CLIME is used to construct the global structure in the brain. To overcome the problem of tuning parameter selection, we integrate out the effect of the tuning parameter. The intraclass correlation coefficient is calculated on Kirby21 data. Also, a multimodal brain imaging analysis is conducted based on the new functional connectivity measure for fMRI and DTI data from the PNC study with 515 subjects.

Results and Conclusion: We found that the novel measure of functional connectivity has a stronger relationship with structural connectivity than conventional Pearson correlation functional connectivity, which suggests our approach more effectively reveals neurologically related activity in functional magnetic resonance imaging. The intraclass correlation coefficient shows that the robustness of this new measure are at a similar level compared to the existing measures Pearson correlation and partial correlation. Thus, we see that the novel functional connectivity measure not only shares similarity with existing popular functional connectivity measures, but also performs better in terms of representativeness in multimodal analysis.

A Novel Network Connectivity Measure with Application to Multimodal Brain Imaging Study

By

Xinyi Yang

B.S., Nankai University, 2016

Thesis Committee Chair: Ying Guo, Ph.D.

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health,

in Department of Biostatistics and Bioinformatics

2019

Acknowledgement

First and for most, I would like to thank my thesis advisor, Dr. Guo, for her great help on not only her mentoring on the thesis, but also her guidence on how to start a research when having an idea, and how to promote the research when having some stageable outcomes. The creative thinking and the rigorous attitude on research I got are valuable assets in my research life. I would also like to thank my thesis reader, Dr. Risk, for his precious comments to my thesis. It is his advice that makes my thesis more detailed and convincing. Last but certainly not least, I am very grateful to Yikai Wang, a Ph.D. candidate in Center for Biomedical Imaging Statistics in Rollins School of Public Health. He offered so many suggestions on promoting the research. And I really learned a lot like some common treatments in preparing for data. Thank you all for your great help!

Also, I would like to thank my parants. Thank them for their greatly support, their consolation and their life experience. And thank my friends, Ying, Xisha, Huaying and Xi, for their kindly accompanying and suggestions in my daily life when pursuing the MSPH degree. Thank you!

Xinyi Yang

Contents

1 Introduction

Network-oriented analysis has played a main role in brain imaging. By conducting network-oriented analysis, we can better understand brain organization structurally or functionally. However, when studying functional connectivity from functional magnetic resonance imaging (fMRI) data in the view of network, it is not that easy due to the large scale of fMRI data.

Traditionally, Pearson correlation is often used to measure connectivity between a pair of nodes (Seeley et al., 2009). However, if using Pearson correlation to measure the functional connectivity of two nodes, the direct connectivity of the two nodes and the interference from other nodes cannot be distinguished. That may lead to unconvincing results (Smith et al., 2011). Wang et al. (2016) proposed a method for better estimating partial correlation between two nodes and used partial correlation to measure functional connectivity. Nevertheless, there are also shortcomings for partial correlation to be used in network-oriented analysis. That makes some information of the network lost. On the other hand, the estimation of partial correlation is under an assumption for sparsity of the network, thus requiring a prescribed tuning parameter, of which the selection needs broad background knowledge. That increases the complexity for using this method. In fact, the most commonly used methods for constructing a brain network, such as CLIME (Cai et al., 2011) and graphical lasso (Friedman et al., 2008), all require a sparsity assumption. That is one of the major difficulties in constructing brain functional networks.

The large scale of the data and the existing methods' requirement of prior knowledge call for easier-to-use and more straightforward measures. Therefore, in the present paper, a novel measure for functional network connectivity is proposed, which gets rid of the predetermined tuning parameter. We make use of the distance derived from CLIME (Cai et al., 2011) and construct the global structure in the brain. To overcome the problem of tuning parameter selection, we integrate out the effect of the tuning parameter. The intraclass correlation coefficient is calculated, which shows that the robustness of this new measure is at a similar level compared to the existing measures Pearson correlation and partial correlation. Also, a multimodal brain imaging analysis is conducted based on the new functional connectivity measure for fMRI and DTI data from the Philadelphia Neurodevelopmental Cohort (PNC) study with 515 subjects, showing a better representativeness than Pearson correlation and partial correlation.

In the rest of this paper, in Section 2, the novel functional network connectivity is introduced. We use the PNC and Kirby21 to do analysis. The statistical properties for this measure are analyzed in terms of robustness and representativeness, in section 3. In Section 4, the results on the statistical properties for this measure are presented. The conclusions derived from the results are given in Section 5. In Section 6, we give further discussions on this project.

2 Methods

2.1 Notation

Let $Y(t)$ denote the V dimensional fMRI data at time t, where $t = 1,2,...,T$. Let Ω_k denote the precision matrix given tuning parameter λ_k. Let $P(\lambda) = \{P_{ij}(\lambda); i, j = 1,2,...,V\}$ be the partial correlation matrix when the tuning parameter is λ. Denote $\mathscr{V}$ as the set of nodes, and denote $\mathscr{E}^{(f,\lambda)}$ as the set of all edges, which are defined by the transformation rule $f : P \to \mathscr{E}$ given λ. The network with vertex (node) set $\mathscr{V}$ and edge set $\mathscr{E}^{(f,\lambda)}$ is denoted as $(\mathscr{V}, \mathscr{E}^{(f,\lambda)})$. We also denote $D_{ij}(\lambda)$ as the distance between node i and node j in network

$(\mathscr{V}, \mathscr{E}^{(f,\lambda)})$, given λ. Denote λ_{ij}^L as the last point where $D_{ij}(\lambda)$ is finite as λ is increased, and λ^L is the largest λ_{ij}^L, $i, j = 1, 2, ..., V$.

2.2 Pathway-embedded functional connectivity measure

Given a tuning parameter λ, we can estimate $P(\lambda)$ from the precision matrix, by first estimating the sparse inverse covariance matrices for a grid of values for lambda by using "CLIME", and evaluating the density levels of the precision matrices by using "DensParcorr", and then making binarization (Cai et al., 2011, Wang et al., 2016). Since we are interested in positive correlation, f is defined as: edge $E_{ij} \in \mathscr{E} \iff P_{ij} > e > 0$, where e is set as 0.001. Then D_{ij} is defined as the number of edges in the shortest pathway between them. If there is no pathway between node i and node j, then D_{ij} is set as $+\infty$. According to the property of tuning parameter λ, when λ goes from 0 to 1, the whole network gets sparser and sparser. For any node pair, when λ is large enough, the distance between them goes to $+\infty$. Then λ_{ij}^L is defined as:

$$\lambda_{ij}^L = \sup\{\lambda : \frac{1}{D_{ij}(\lambda)} > 0\} \tag{1}$$

Since larger distance means less connectivity, we can measure the functional connectivity by transforming D_{ij} with a monotone decreasing function $g(\cdot)$. Thus, the novel functional connectivity measure is defined as:

$$FC_{i,j} = \frac{1}{\lambda^L} \int_0^{\lambda_{ij}^L} g\Big(D_{ij}(\lambda)\Big) d\lambda \tag{2}$$

Where $g(\cdot)$ can take the form of $g(x) = x^{-\alpha}, \alpha > 0$.

2.3 Algorithm

As discussed earlier, for a pair of nodes (i, j), the novel functional connectivity measure is defined as the integral of the transformation of distance between Node i and Node j. And given a series of λ, we can construct the corresponding adjacency matrices by deriving partial correlation matrices. Then we use AUC (AUC $= \int g(\lambda)d\lambda$), to estimate the novel connectivity measure.

Algorithm 1: Estimate subject-specific functional connectivity matrix

---

Data: Subject-specific covariance matrices $\Sigma_n, n = 1, 2, ..., N$.
Initialize: a series of monotonically decreasing tuning parameters
$\Lambda = \{\lambda_k : k = 1, 2, ..., K\}$ ranging from 1 to 0.
Step 1: Estimate $D_{ij}(\lambda)$:
for $\lambda_k \in \Lambda$ **do**
  Estimate Ω_k based on CLIME;
  Calculate $P_k = P(\lambda_k) = \{P_{ij,k}\}_{V \times V}$ based on DensParcorr;
  Construct the adjacency matrix $A_k = \{A_{ij,k}\}_{V \times V}$;
  if $P_{ij,k} > \delta$ **then**
    $A_{ij,k} = 1$;
  else
    $A_{ij,k} = 0$;
  end
  Estimate $D_k = \{D_{ij,k}\}_{V \times V}$, where $D_{ij,k}$ =length of shortest path in Graph A_k.
end
Step 2: Estimate novel functional connectivity using AUC:
Identify λ_{ij}^L, with order K_L;
Calculate AUC of $\{g(D_{ij,1}), g(D_{ij,2}), ..., g(D_{ij,K_L})\}$:

$$AUC_{ij} = \sum_{k=K_L}^{K-1} \frac{1}{2}\Big(g(D_{ij,k} + g(D_{ij,k+1})\Big)\Big(\lambda_k - \lambda_{k+1}\Big) \quad (3)$$

$$FC_{ij} = \frac{AUC_{ij}}{\lambda^L} \quad (4)$$

---

3 Analysis

In the following sections, we choose $\Lambda = \{0.001, 0.005, 0.01 \text{ to } 0.6, \text{ by } 0.01\}$, to do the analysis.

3.1 The Choice of Transformation Function

For the transformation function $g(\cdot) = x^{-\alpha}$, different values of α would lead to different results. Here we use Shannon's entropy (Hausser & Strimmer, 2009) as the criteria for the selection of α. The definition of Shannon's entropy is:

$$S = -\sum_i P_i \log P_i \tag{5}$$

And we select the value of α from $\{1,2,3,4,5\}$ for the PNC dataset, by using the with the smallest entropy.

3.2 Reproducibility

In order to investigate the reproducibility of the novel FC measure, we use the intraclass correlation coefficient (Koo & Li, 2016), which is computed from a two-way mixed effects model. The ICC is computed as:

$$ICC = \frac{MS_B - MS_E}{MS_R + (k-1)MS_E} \tag{6}$$

where MS_B is the mean square between targets, MS_E is the mean square error, and k is the number of raters/measurements (Shrout & Fleiss, 1979).

Since there is only one scan for each subject in the PNC dataset, here we use Kirby21 dataset, in which each subject has two scans, to analyze reproducibility. First, for each pair of nodes (edge), the Pearson connectivity and novel FC of Kirby21 data are calculated seperately. Then we compare the ICC so as to compare their reproducibility. For the transformation function $g(\cdot) = x^{-\alpha}$, we choose $\alpha = 1 to see the per formance of novel FC based on them.$

3.3 Multimodality

In order to compare the representativeness between the novel FC and current FC measure, Pearson connectivity, we do multimodal analysis between FC and structural connectivity (SC). Here the SC is estimated using diffusion tensor imaging (DTI) data (Kemmer et al., 2018). We use Pearson correlation, Spearman's rank correlation, and mutual information (Kraskov et al., 2004) to quantify the relationship between FC and SC.

Using PNC dataset, we make comparisons among novel FC when $\alpha = 1,2,3,4$, and 5, and Pearson connectivity.

4 Results

4.1 Visualization

The mean FC across subjects of each pair of nodes (edge) when α =1, 2, 3, 4, and 5 are calculated (Figure 1). The density plot of FC for all edges in each subject is also made, with the density plot of mean FC across subjects for all edges (Figure 2). Since by the definition of the novel FC, its range is $(0,1)$, we also try logit transformation on FC and calculate the mean across subjects of each pair of nodes (edge) when α =1, 2, 3, 4, and 5 (Figure 3). So is the density plot of logit-transformed FC (Figure 4). In terms of the novel FC, the within-region edges tend to have larger functional connectivity, while the between-region edges tend to have smaller functional connectivity.

Compared with mean Pearson connectivity (Figure 5), the difference between within-region edges and between-region edges tend to be larger in the result of novel FC. Nevertheless, more quantitative comparisons are needed in order to compare between novel FC and Pearson connectivity.

Density plot shows the distribution of the novel FC. Since the logit transformation extends the range from (0,1) to $(-\infty, \infty)$, it is suggested to refer to the density plot for $logit(FC)$ to see more details. In addition, there is a right shift for the density plot of mean FC across subjects, compared to the density plot of FC for each single subject. That is possibly because the smaller FC appear on a different set of edges in different subjects.

4.2 The Choice of Transformation Function

When α changes from 1 to 5, the entropy of the novel FC in each subject becomes smaller and smaller, and the entropy of the mean FC across subjects of all edges also becomes smaller and smaller (Figure 6). Since $\alpha = 5$ has the lowest entropy, we suggest selecting $\alpha = 5$ when calculating the novel FC of the PNC dataset.

Additionally, from the current line chart of entropy, the rate of decrease tends to be smaller and smaller. Still, entropy of mean FC across subjects behaves different from mean entropy of FC across subjects. That is, $S(\overline{FC}) \neq \overline{S}(FC)$.

4.3 Reproducibility

For the Kirby21 dataset, ICC of the novel FC (when $\alpha = 1$) is slightly lower than that of Pearson connectivity (Figure 7). While for the within-region edges, ICC of the novel FC (when $\alpha = 1$) and ICC of Pearson connectivity are similar (Figure 8)That means the reproducibility of the novel FC is almost comparable to that of Pearson connectivity.

Additionally, when looking at the ICC of all edges, neither ICC of Pearson connectivity nor that of novel FC shows obvious distinction (Figure 8).

4.4 Multimodality

For the result of Pearson correlation, the novel FC shows larger positive correlation with SC (around 0.6), compared with the correlation between Pearson connectivity and SC (0.325) (Table 1). That means the novel FC has better linear correlation with SC. Additionally, when α gets larger, the linear correlation between novel FC and SC also becomes larger (from 0.534 to 0.613).

For the result of Spearman's rank correlation, the Spearman's rank correlation between Pearson connectivity and SC is 0.305, while those between novel FC and SC (when $\alpha = 1, 2, 3, 4$,and 5) are all larger than or equal to 0.360. That shows novel FC and SC become closer to being perfectly positively monotone than Pearson connectivity and SC.

For the result of MI, the mutual information between Pearson connectivity and SC is 4.899, while the mutual information between novel FC and SC (when $\alpha = 1, 2, 3, 4$,and 5) are all larger than or equal to 4.907. Thus, the mutual information between novel FC and SC is larger than that between Pearson connectivity and SC.

All of the Pearson correlation, Spearman's rank correlation, and mutual information reveal the novel FC to have better agreement with SC, compared with the agreement between Pearson connectivity and SC.

5 Discussion

We found that the novel measure of functional connectivity has a stronger relationship with structural connectivity than conventional Pearson correlation functional connectivity, which suggests our approach more effectively reveals neurologically related activity in functional magnetic resonance imaging. Specifically, in the multimodality analysis between SC and FC, all of the Pearson correlation, Spearman's rank correlation, and mutual information suggest that the novel FC is a better functional connectivity measure than Pearson correlation. Additionally, when visualizing the novel FC of the PNC dataset, we found that the novel FC shares some characteristics with the Pearson correlation. For instance, for node pairs between the medial visual network, occipital pole visual network, and lateral visual network, the novel FC tends to be larger than other between-region edges. Plus, the reproducibility of the novel FC is comparable to that of Pearson connectivity. Thus, we see that the novel FC is a measure that not only shares similarity with existing popular functional connectivity measures, but also performs better in terms of discrimination and in multimodal analysis.

However, there are limitations in this novel FC measure. First and foremost, when calculating AUC for estimating the FC, there is a series of functional distances under different tuning parameters that need to be calculated. That leads to a much larger computational complexity, compared with the existing popular measures. Additionally, when constructing the adjacency matrix so as to calculate the distance between nodes, we binarize the adjacency matrix. We may lose some information or characteristics because of the binarization.

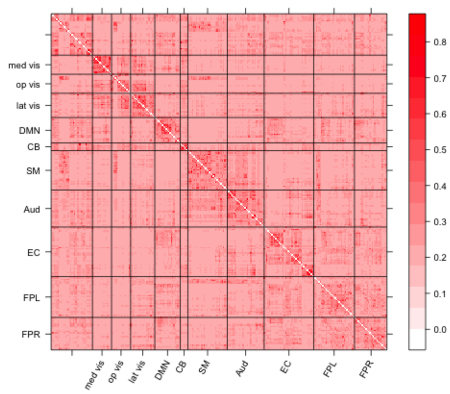
In the future, we are considering the following explorations. Firstly, we are trying to study the reliability of the novel FC using related measures like Image Intraclass Correlation Coefficient (I2C2)(Shou et al., 2013). We are also going to further study the reproducibility of the novel FC by analysing the ICC of more network measures. Secondly, we can choose more α in the transformation function $g(\cdot) = x^{-\alpha}$, such as $\alpha = 6, 7, 8...$, to see if we can know more about the trend and get better characteristics. In the long run, when constructing the adjacency matrix so as to calculate the distance between nodes, we can think of a way that doesn't need to binarize the adjacency matrix. Additionally, when visualizing the novel FC of PNC dataset, we found that the density plot for mean FC across subjects behaves different from the mean density plot across subjects of FC. That is, $\overline{\mu}(p) \neq \mu(\overline{p})$, where $\mu(\cdot)$ is the density function. It may be an interesting topic to explore the relationship between that difference and the structure (or the characteristics) of the network.

Appendix: Figures and Tables

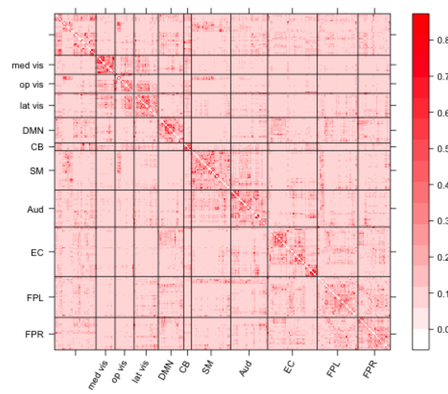
1. Visualization of Novel FC

Figure 1: Mean FC across subjects under $\alpha = 1, 2, 3, 4$, and 5 — PNC data)

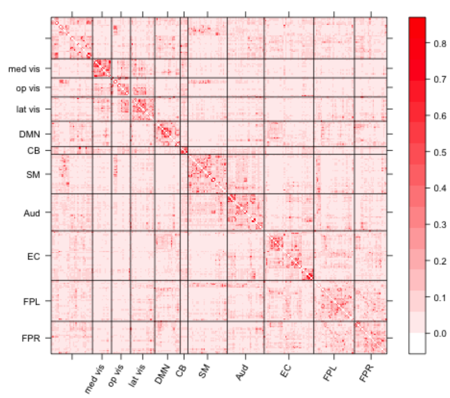
(a) $\alpha = 1$



(b) $\alpha = 2$



(c) $\alpha = 3$



(d) $\alpha = 4$

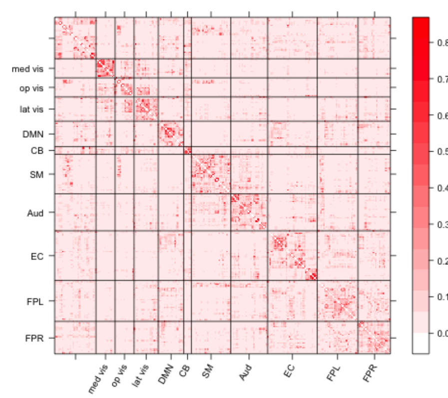


Figure 1: Mean FC across subjects under $\alpha = 1, 2, 3, 4,$ and 5 — PNC data (cont.)

(e) $\alpha = 5$

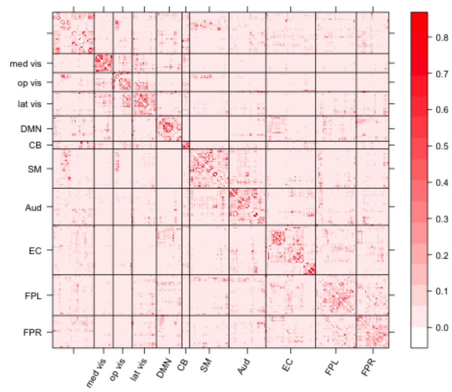
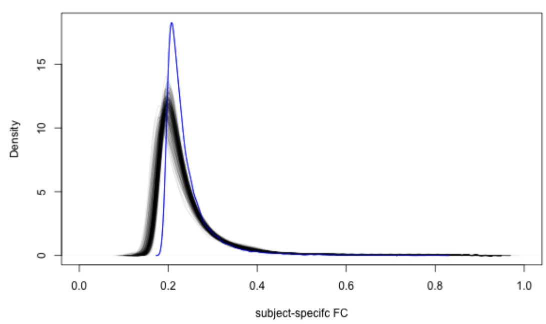
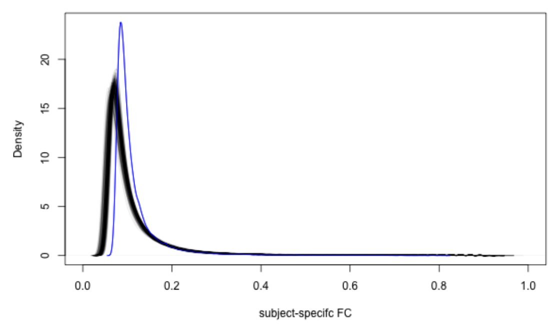


Figure 2: Density plot for FC of each subject under $\alpha = 1, 2, 3, 4,$ and 5 — PNC data (Blue line: mean FC across subjects)

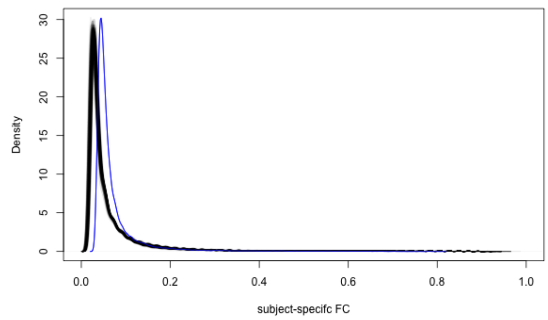


(a) $\alpha = 1$

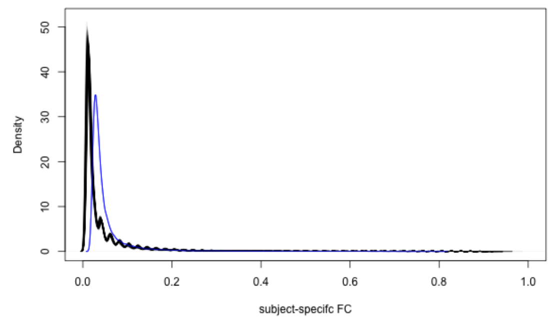


(b) $\alpha = 2$

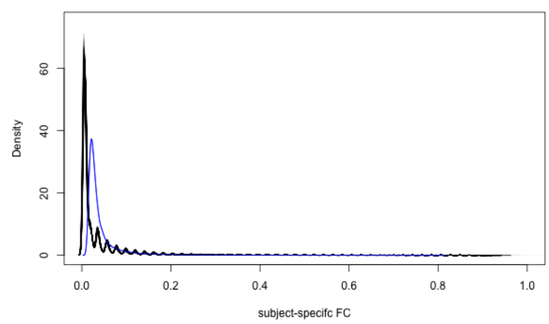
Figure 2: Density plot for FC of each subject under $\alpha = 1, 2, 3, 4$, and 5 — PNC data (Blue line: mean FC across subjects) (cont.)



(c) $\alpha = 3$

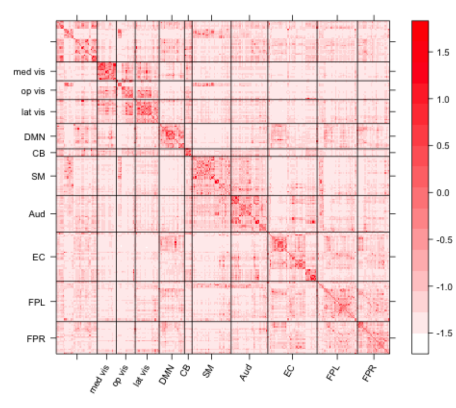


(d) $\alpha = 4$

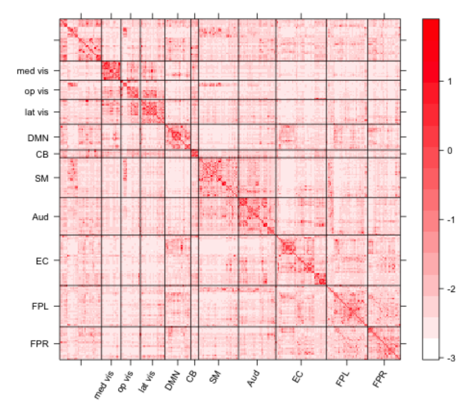


(e) $\alpha = 5$

Figure 3: Mean $logit(FC)$ across subjects under $\alpha = 1, 2, 3, 4$, and 5 — PNC data

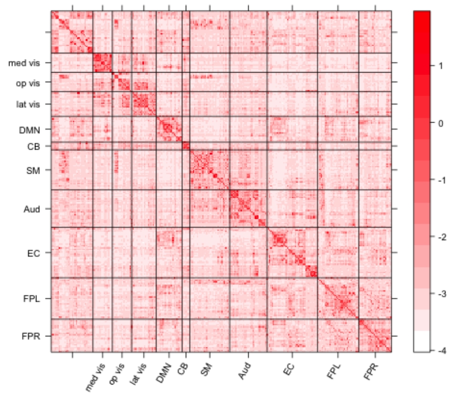


(a) $\alpha = 1$

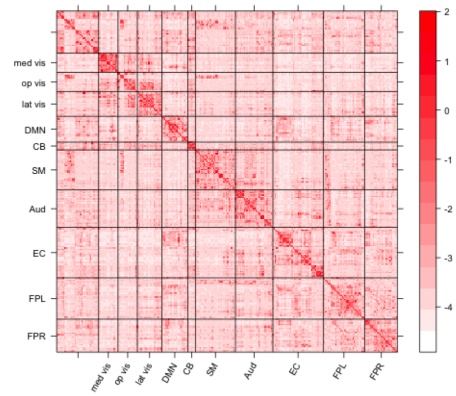


(b) $\alpha = 2$

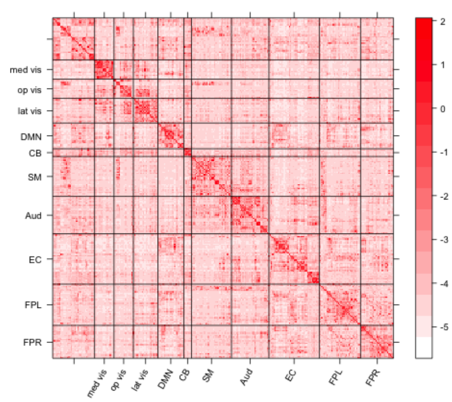
Figure 3: Mean $logit(FC)$ across subjects under α = 1, 2, 3, 4, and 5 (cont.)



(c) $\alpha = 3$

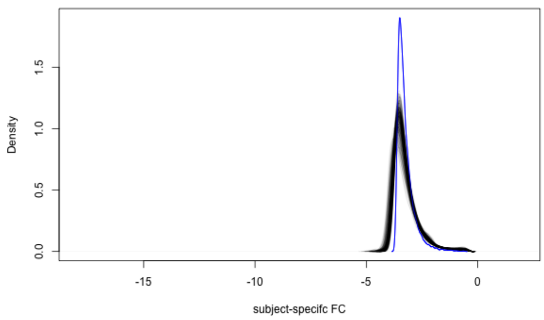


(d) $\alpha = 4$

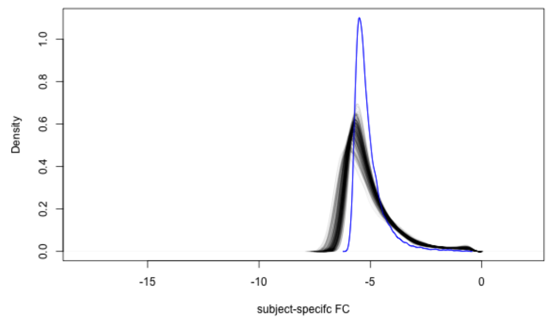


(e) $\alpha = 5$

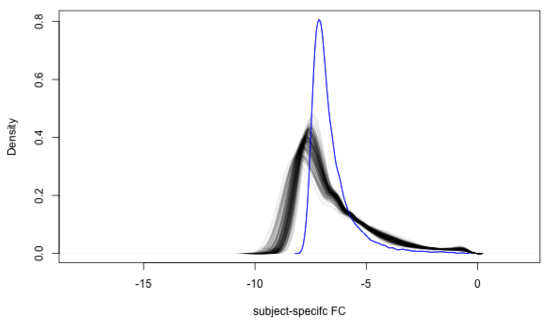
Figure 4: Density plot for logit(FC) of each subject under α = 1, 2, 3, 4, and 5 — PNC data (Blue line: mean logit(FC) across subjects)



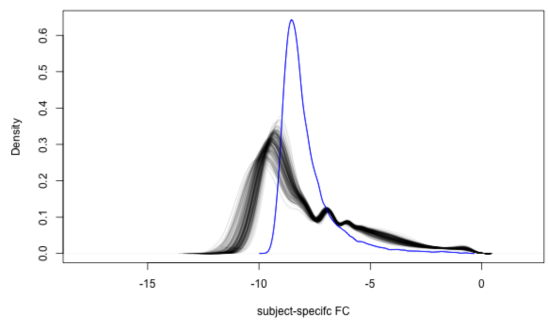
(a) $\alpha = 1$



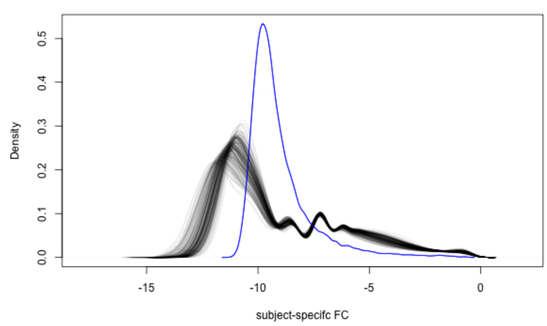
(b) $\alpha = 2$



(c) $\alpha = 3$

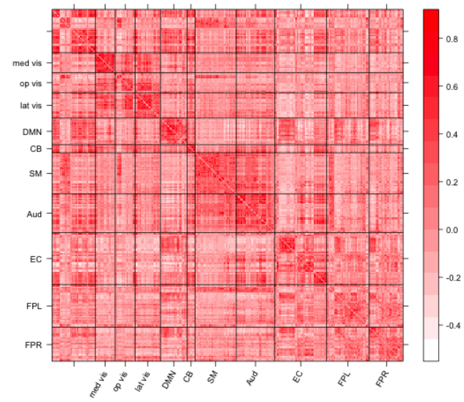


(d) $\alpha = 4$



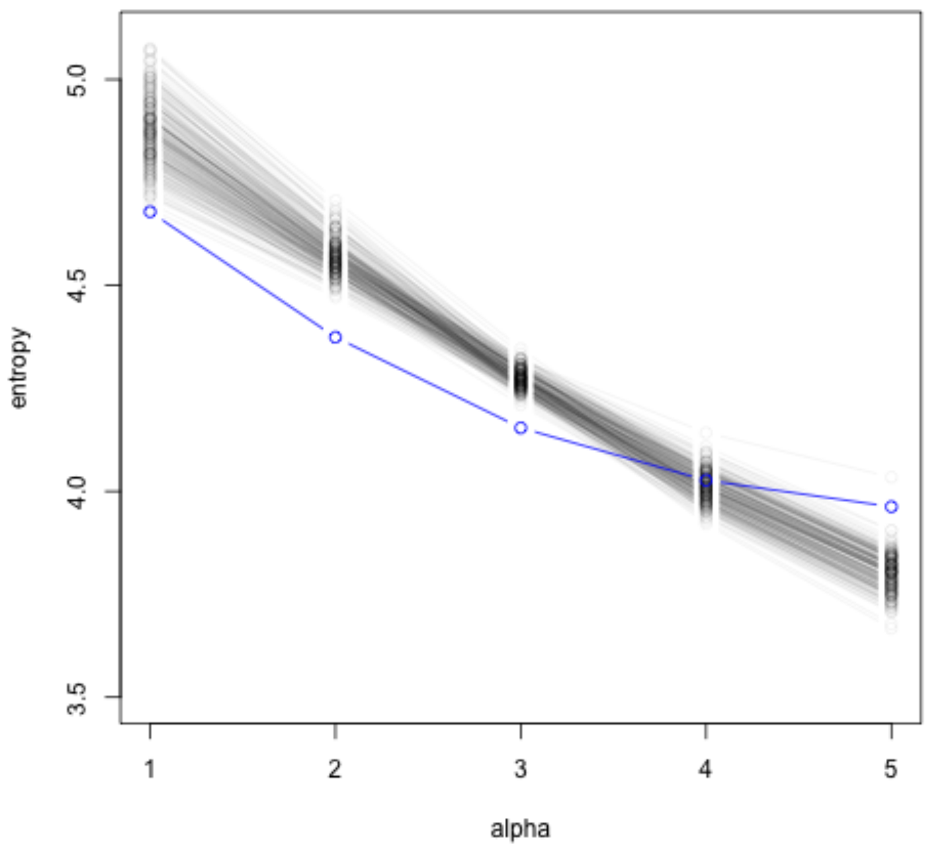
(e) $\alpha = 5$

Figure 5: Mean of Pearson connectivity across subjects — PNC data



2. Selecting Transformation Function — Entropy

Figure 6: Entropy of novel FC under different transformation functions — PNC data (Blue line: entropy of mean FC across subjects)



3. Reproducibility – ICC

Figure 7: Histograms for ICC of Pearson connectivity and novel FC — Kirby21 data

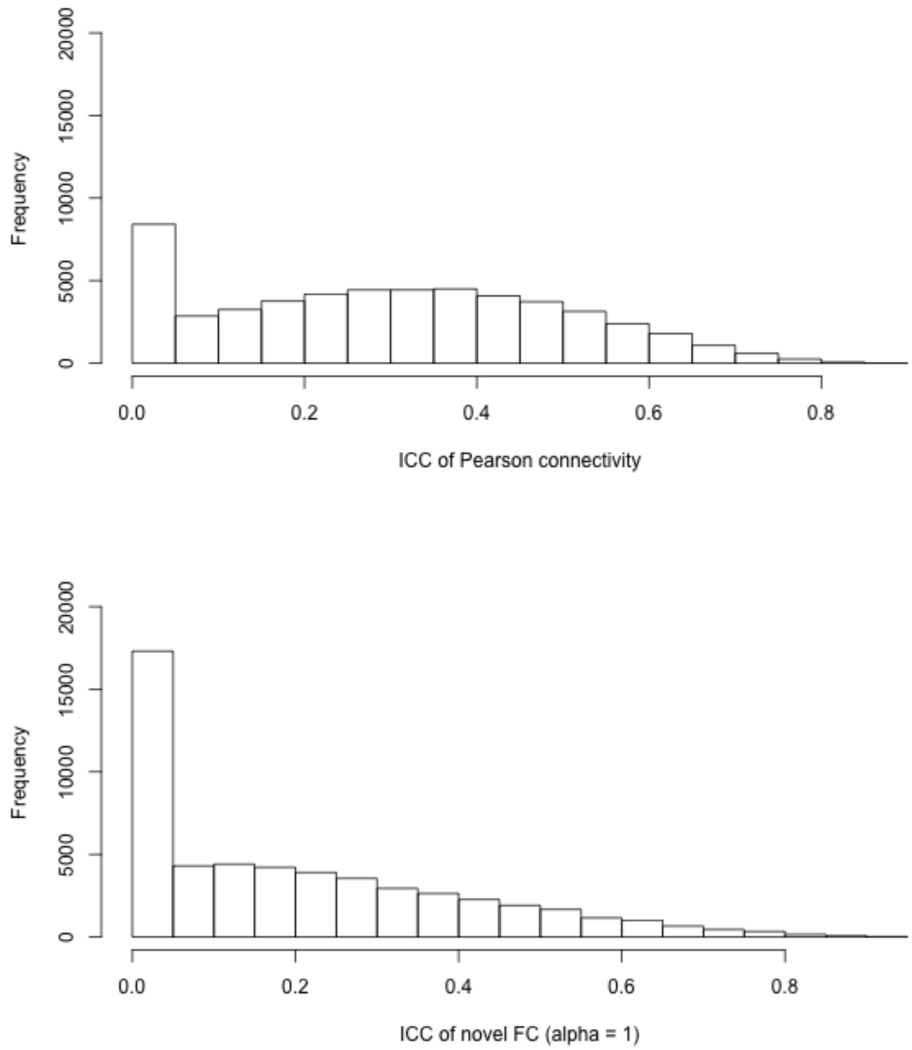
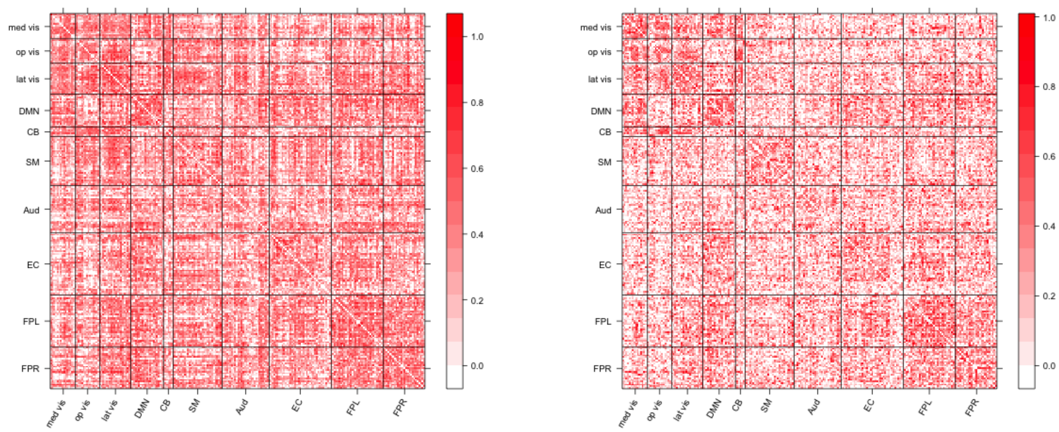


Figure 8: ICC of Pearson connectivity and novel FC — Kirby21 data



(a) ICC of Pearson connectivity

(b) ICC of novel FC ($\alpha = 1$)

4. Multimodal Analysis

Table 1: Pearson correlation, Spearman's rank correlation, and mutual information (MI) between FC and SC — PNC data

	Pearson correlation	Spearman's rank correlation	MI
Pearson connectivity	0.325	0.305	4.899
novel connectivity ($\alpha = 1$)	0.534	0.360	4.907
novel connectivity ($\alpha = 2$)	0.578	0.366	4.920
novel connectivity ($\alpha = 3$)	0.598	0.367	4.913
novel connectivity ($\alpha = 4$)	0.608	0.365	4.914
novel connectivity ($\alpha = 5$)	0.613	0.363	4.917

References

Cai, T., Liu, W., & Luo, X. (2011). A constrained l1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, *106*(494), 594–607.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432–441.

Hausser, J., & Strimmer, K. (2009). Entropy inference and the james-stein estimator, with application to nonlinear gene association networks. *Journal of Machine Learning Research*, *10*(Jul), 1469–1484.

Kemmer, P. B., Wang, Y., Bowman, F. D., Mayberg, H., & Guo, Y. (2018). Evaluating the strength of structural connectivity underlying brain functional networks. *Brain Connectivity*, *8*(10), 579–594.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, *15*(2), 155–163.

Kraskov, A., Stögbauer, H., & Grassberger, P. (2004). Estimating mutual information. *Physical review E*, *69*(6), 066138.

Seeley, W. W., Crawford, R. K., Zhou, J., Miller, B. L., & Greicius, M. D. (2009). Neurodegenerative diseases target large-scale human brain networks. *Neuron*, *62*(1), 42–52.

Shou, H., Eloyan, A., Lee, S., Zipunnikov, V., Crainiceanu, A., Nebel, M., … Crainiceanu, C. M. (2013). Quantifying the reliability of image replication studies: the image intraclass correlation coefficient (i2c2). *Cognitive, Affective, & Behavioral Neuroscience*, *13*(4), 714–724.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, *86*(2), 420.

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., ... Woolrich, M. W. (2011). Network modelling methods for fmri. *Neuroimage*, *54*(2), 875–891.

Wang, Y., Kang, J., Kemmer, P. B., & Guo, Y. (2016). An efficient and reliable statistical method for estimating functional connectivity in large scale brain networks using partial correlation. *Frontiers in neuroscience*, *10*, 123.