WHOLE BRAIN GROUP NETWORK ANALYSIS USING NETWORK BIAS AND VARIANCE PARAMETERS

Alireza Akhondi-Asl[1]*,Arne Hans*[1]*, Benoit Scherrer*[1]*, Jurriaan M. Peters*[1]*, Simon K. Warfield*[1]

[1]Computational Radiology Laboratory, Childrens Hospital Boston, and Harvard Medical School
300 Longwood Ave. Boston MA 02115 USA

ABSTRACT

The disruption of normal function and connectivity of neural circuits is common across many diseases and disorders of the brain. This disruptive effect can be studied and analyzed using the brain's complex functional and structural connectivity network. Complex network measures from the field of graph theory have been used for this purpose in the literature. In this paper we have introduced a new approach for analyzing the brain connectivity network. In our approach the true connectivity network and each subject's bias and variance are estimated using a population of patients and healthy controls. These parameters can then be used to compare two groups of brain networks. We have used this approach for the comparison of the resting state functional MRI network of pediatric Tuberous Sclerosis Complex (TSC) patients and healthy subjects. We have shown that a significant difference between the two groups can be found. For validation, we have compared our findings with three well known complex network measures .

***Index Terms*—** Functional connectivity, Resting state fMRI, Connectivity graph, Parcellation, Tuberous Sclerosis Complex

1. INTRODUCTION

The human brain can be considered as a complex functional and structural network [1]. Graph theory has been widely utilized for the analysis and characterization of the brain connectivity network. To this end, connectivity measures between different parts of the brain are calculated and used to create a functional or structural connectivity matrix [2]. Recently, group analysis has been utilized in several studies to analyze the effect of diseases on the brain network [3]. For this purpose, a series of complex network measures has been used to analyze the functional and structural connectivity networks of the brain. Using these measures, the effect of different diseases on integration, segregation, centrality, and resilience of each node and also the whole brain network has been studied [4]. In this paper, we introduce a new approach to analyze brain networks by using the Expectation-Maximization (EM) algorithm to estimate the true brain network, and bias and variance parameters. The estimated bias and variance parameters are used for the group analysis to compare the group of controls and patients. We have used our new measure to compare the functional network of a group of pediatric Tuberous Sclerosis Complex (TSC) patients with age matched healthy subjects. TSC is a neurologic disorder and patients present with severe epilepsy, cognitive impairment and neurobehavioral abnormalities, particularly autism [5]. In several studies abnormalities in the white matter of TSC patients including dys-myelination in the white matter tracts have been reported [6]. However, there is limited knowledge on functional and structural connectivity in pediatric TSC patients.

2. MATERIAL AND METHODS

Assume that the connectivity matrix W_j is generated for each one of the patients and healthy subjects in a population with J members. For each network j, w_{mnj} indicates the weight of the link that connects the regions m and n. We assume that there are L nodes (regions) in each one of the networks. In the literature, complex networks measures have been used to analyze the differences of the networks between patients and controls. For this purpose, using each one of the measures, global or local organization of the networks are characterized and then populations of patients and healthy subjects are compared.

2.1. True Brain Network

All of the above mentioned measures use graph features to compare different networks. However, in our approach, we consider each one of the connectivity matrices as a variation of the true brain network. Following the approach in [7, 8], we model the variation in the following form:

$$w_{mnj} = \tau_{mn} + \beta_j + \epsilon_{mnj} \quad (1)$$

In this equation τ is the true brain network and τ_{mn} is the weight of the link between the nodes m and n in the true brain network. Also, β is the vector of bias of different networks in the population where β_j shows the bias of the j-th network, and ϵ_{mnj} denotes the error in the weight of the link

between nodes m and n. It is assumed that the error has an uncorrelated normal distribution $\epsilon_{mnj} \sim N(0, \sigma_j^2)$. Thus, we characterize the j-th brain network in our population with a bias β_j from the true brain network and a variance σ_j^2 which models the errors. We assume that the joint distribution of the weights given the default network and each network's parameters have the following form:

$$Pr(w|\tau, \sigma, \beta) = \prod_{j=1}^{J} \prod_{m=1}^{L} \prod_{n=1, n<m}^{L} \phi\left(\frac{w_{mnj} - (\tau_{mn} + \beta_j)}{\sigma_j}\right) \quad (2)$$

where $\phi\{.\}$ is the pdf with normal distribution $N(0, 1)$. We assume that brain networks of different subjects in the population are independent. In addition, we also assume that the link weights in the network are independent. Because of the symmetrical form of the network, the elements of the matrix are not independent, and therefore we need to use the lower triangular or upper triangular part of the network matrices. The true network is not known and maximization of the complete data likelihood can not be used to estimate the bias and variance of each one of the networks in the population. Thus, the EM algorithm is used to estimate the true brain network, bias, and variance. It should be mentioned that there is no assumption about the weight values. For example, one of the problems of resting state fMRI (rsfMRI) analysis are the negative correlations, and many complex network measures work on either positive or negative weights. Thus, in many methods either the negative values are eliminated or two different networks are considered for negative and positive connections [9].

Our aim is to find the bias and variance of each one of the networks. Using the EM algorithm at each iteration t, the expectation steps σ and τ are computed using the following two equations:

$$\frac{1}{(\sigma^2)^{(t)}} = \sum_{j=1}^{J} \frac{1}{(\sigma_j^2)^{(t)}}$$
$$\tau_{mn}^{(t)} = \frac{\sum_{j=1}^{J} (w_{mnj} - \beta_j^{(t)})/(\sigma_j^2)^{(t)}}{1/(\sigma^2)^{(t)}} \quad (3)$$

Then, in the maximization step using the results given in Eq. 3, the estimation of β_j for each one the networks is updated using the following equation:

$$\beta_j^{(t+1)} = \frac{1}{L(L-1)/2} \sum_{n=1, n<m}^{n=L, m=L} (w_{mnj} - \tau_{mn}^{(t)}) \quad (4)$$

Moreover, the estimation of σ_j for each one the networks is updated using the following equation:

$$(\sigma_j^2)^{(t+1)} = (\sigma^2)^{(t)} + \frac{1}{L(L-1)/2} \sum_{n=1, n<m}^{n=L, m=L} \left(w_{mnj} - \beta_j^{(t+1)} - \tau_{mn}^{(t)}\right)^2 \quad (5)$$

Using this framework, the parameters are updated iteratively until convergence is obtained which is guaranteed by using the EM algorithm. Last but not least, we initialize the bias and variance parameters to zero and one, respectively.

2.2. Distance Calculation

After the calculation of the bias and variance of each network, these parameters will be used for the analysis of the networks. In this work, we focus on the application of the framework for group analysis. Without loss of generality it can be assumed that the networks of subjects $j \in \{1, ..., J_1\}$ and $j \in \{J_1 + 1, ..., J\}$, indicate healthy subjects and patient, respectively. It is possible to compare bias and variance of two groups independently, however, we are more interested in using both bias and variance parameters in the comparison of the groups. To this end, for the controls (C) and patients (P), the average bias is computed using the following equations:

$$\bar{\beta}_P = \frac{1}{J_1} \sum_{j=1}^{J_1} \beta_j$$
$$\bar{\beta}_C = \frac{1}{J_2 - J_1} \sum_{j=J_1+1}^{J} \beta_j \quad (6)$$

where $\bar{\beta}_C$ and $\bar{\beta}_P$ denote the average bias of controls and patients, respectively. Moreover, using the following equations, the average variance of controls $\bar{\sigma}_C^2$, and the average variance of patients $\bar{\sigma}_P^2$, can be computed:

$$\bar{\sigma}_C^2 = \frac{1}{J_1} \sum_{1}^{J_1} \left(\sigma_j^2 + \left(\bar{\beta}_C - \beta_j\right)^2\right)$$
$$\bar{\sigma}_P^2 = \frac{1}{J_2 - J_1} \sum_{j=J_1+1}^{J} \left(\sigma_j^2 + \left(\bar{\beta}_P - \beta_j\right)^2\right) \quad (7)$$

Now it is possible to directly compare the Gaussian probability distribution of the controls and patients using any probability distance measure. In this paper we use the symmetrized Kullback-Leibler divergence (SKLD) for the comparison of the two groups which can be defined as in [10]:

$$S_D = KLD(N_P||N_C) + KLD(N_C||N_P) \quad (8)$$

where $KLD(N_1||N_2)$, the Kullback-Leibler divergence (KLD) of Gaussian probability distribution of group one and two is:

$$KLD(N_1||N_2) = \frac{1}{2}\left(\log\frac{(\bar{\sigma}_2^2)}{(\bar{\sigma}_1^2)} + \frac{(\bar{\sigma}_1^2)}{(\bar{\sigma}_2^2)} + \frac{(\bar{\beta}_2 - \bar{\beta}_1)^2}{\bar{\sigma}_2^2} - 1\right) \quad (9)$$

2.3. Complex Network Measures

There are different types of complex network measures that can be used for the analysis of the brain network. In this paper, we compared our findings with 3 well known measures that are usually considered for this purpose. The considered measures are: Total connection strength (K), overall weighted clustering coefficient (C), and overall weighted transitivity (T) [4]. For each subject, K_j, C_j, and T_j can be computed using the following set of equations:

$$K_j = \frac{1}{L}\sum_{m,n} w_{mnj}$$
$$C_j = \frac{1}{L}\sum_{m} \frac{\sum_{n,k}(w_{mnj}w_{mkj}w_{nkj})^{(\frac{1}{3})}}{(\sum_n w_{mnj})(\sum_n w_{mnj}-1)}$$
$$T_j = \frac{\sum_{m,n,k}(w_{mnj}w_{mkj}w_{nkj})^{(\frac{1}{3})}}{\sum_m(\sum_n w_{mnj})(\sum_n w_{mnj}-1)} \quad (10)$$

Similar to our approach, for each one of these complex network measures, and also for the group of patients and controls, the average can be computed and their difference can be used for the analysis of the group differences. We define these differences as K_D, C_D and T_D.

2.4. Statistics

After finding a difference between the two groups using any measure, this difference should be examined using a statistical test. In this paper we have used the permutation test for the analysis of the group differences. First, the distance between the two groups is computed using each one of the measures. Then, subjects are randomly divided into two groups with J_1 and $J-J_1$ subjects and the difference between the two groups is computed for each one of the measures. The procedure is repeated R times to find S_D^r, K_D^r,C_D^r and T_D^r for each $r \in \{1,\ldots,R\}$. Finally, we calculate the p-value of measure X using $p_X = \frac{R-\sum_r H(X_D - X_D^r)}{R}$ where H is the step function. In this equation the number of times that the difference of the measures between two randomly generated groups is larger than the difference between the controls and the TSC patients is used to estimate the p-value.

2.5. Data Acquisition and Pre-processing

Structural MRI and rsfMRI were carried out in 22 subjects with TSC (age range 3-24 years, mean age 11.4), and in 18 age-matched controls on a 3T Siemens scanner. For rsfMRI, sequences with TR ranging from 2400ms to 3000ms were used. T1-weighted MPRAGE images of each subject were automatically segmented by label fusion into 128 cortical/sub-cortical structures using the IBSR datasets [11] as a template. The fusion algorithm is an extension of the STAPLE algorithm [12]. Figure 1(a) shows the parcellation and segmentation of an axial slice through the brain based on 114 cortical and sub-cortical grey matter structures. A series of pre-processing steps were applied to the rsfMRI data of each subject. Head motion was corrected by rigid registration of each volume to the average of all volumes, and each motion corrected volume was spatially smoothed using a 8-mm full-width half-maximum (FWHM) Gaussian kernel. T1-weighted images and their segmentation were registered to the average of the head motion corrected rsfMRI images. Using a regression model, linear and quadratic trends, the averaged signal over the whole brain, the averaged signal over the ventricles, and the averaged signal over the deep white matter were removed [3]. Finally, the time series were band pass filtered by retaining frequencies between 0.01-0.08Hz.

3. RESULTS

For each one of the cortical/sub-cortical grey matter structures the average pre-processed time signal was utilized to construct a weighted connectivity graph for each subject, based on Pearson's correlation. In this paper we focus on positive connectivity to be able to compare our findings with the results of other complex network measures. We have used our approach to estimate the true connectivity matrix of the brain network which is shown in Figure 1(b). In addition, Figures 1(c) and 1(d) show a sample connectivity matrix of a TSC patient and a healthy subject. It can be seen that the connectivity matrix of the healthy subject and the estimated true connectivity matrix are similar. We applied a group difference analysis using the permutation test for our introduced method and also for each one of the complex network measures. For the permutation test and for each one of the measures, we have used 10,000 random permutations and a significance threshold of 0.05. Statistical analysis using the permutation test shows a significant difference between patients and healthy subjects using our approach ($p_S < 0.05$). In addition, the permutation test shows a significant difference between TSC patients and controls for the total connection strength and overall weighted clustering coefficients, ($p_C, p_K < 0.05$). However, the difference between the two groups is not significant based on the overall weighted transitivity ($p_T > 0.08$). These findings show that our approach can be used for connectivity network analysis.

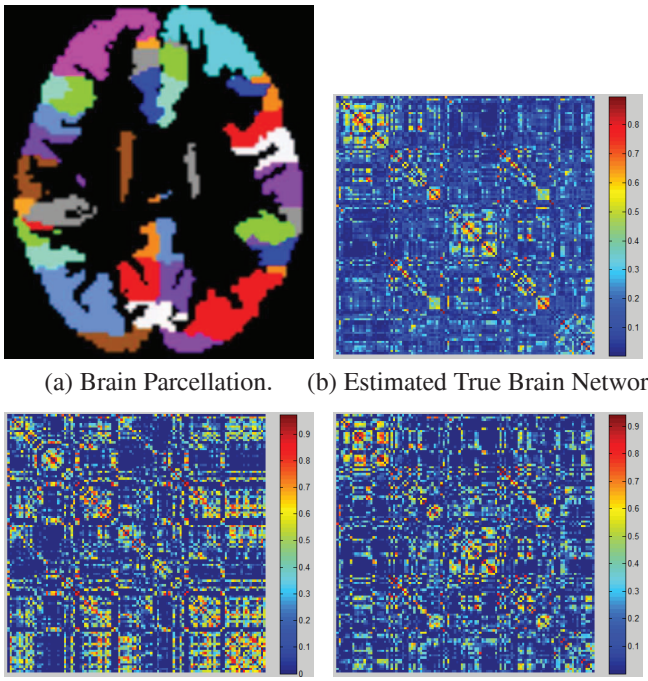
4. CONCLUSIONS

We have introduced a measure for the comparison of the connectivity matrices of a group of patients and healthy subjects. In our approach we estimated the true brain network, bias, and variance of each one of the subjects which can be used for group analysis. We performed resting state functional connectivity group analysis of pediatric TSC patients and controls. The statistical analysis using a permutation test shows a significant difference between the networks ($p < 0.05$). In

addition, we have used three well established complex network measures for the analysis of the same subjects. Two of the complex network measures show a significant difference between functional connectivity in TSC patients compared to controls ($p < 0.05$). It should be pointed out that we have used our method for the global network analysis, however, it is also possible to use our method to find the local differences between the patient population and healthy subjects. The differences in connectivity that we found could help explain the neurological phenotype in patients with TSC, as decreased long-range connectivity is thought to be associated with autism spectrum disorders.

5. ACKNOWLEDGMENTS

This investigation was supported in part by NIH grants R01 RR021885, R01 EB008015, R03 EB008680 and R01 LM010033.



(a) Brain Parcellation. (b) Estimated True Brain Network.

(c) Sample TSC Network. (d) Sample Healthy Network.

Fig. 1. Brain parcellation and connectivity matrices. (a). Parcellation and segmentation of an axial slice based on 114 cortical and sub-cortical grey matter structures. (b). Estimated true brain network. (c). Network of a patient with TSC. (d). Network of a healthy subject.

6. REFERENCES

[1] A. Zalesky, A. Fornito, and E.T. Bullmore, "Network-based statistic: identifying differences in brain networks," *Neuroimage*, vol. 53, no. 4, pp. 1197–1207, 2010.

[2] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186–198, 2009.

[3] M.P. van den Heuvel, R.C.W. Mandl, C.J. Stam, R.S. Kahn, and H.E.H. Pol, "Aberrant frontal and temporal complex network structure in schizophrenia: a graph theoretical analysis," *The Journal of Neuroscience*, vol. 30, no. 47, pp. 15915–15926, 2010.

[4] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.

[5] P. Curatolo and R. Bombardieri, "Tuberous sclerosis," *Handbook of Clinical Neurology*, vol. 87, pp. 129–151, 2007.

[6] Y.J. Choi, A. Di Nardo, I. Kramvis, L. Meikle, D.J. Kwiatkowski, M. Sahin, and X. He, "Tuberous sclerosis complex proteins control axon formation," *Genes & development*, vol. 22, no. 18, pp. 2485, 2008.

[7] S.K. Warfield, K.H. Zou, and W.M. Wells, "Validation of image segmentation by estimating rater bias and variance," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 366, no. 1874, pp. 2361–2375, 2008.

[8] O. Commowick and S.K. Warfield, "A continuous staple for scalar, vector, and tensor images: An application to dti analysis," *Medical Imaging, IEEE Transactions on*, vol. 28, no. 6, pp. 838–846, 2009.

[9] M. Rubinov and O. Sporns, "Weight-conserving characterization of complex functional brain networks," *Neuroimage*, 2011.

[10] S. Kullback, *Information theory and statistics*, Dover Pubns, 1997.

[11] "Internet brain segmentation repository," *http://www.cma.mgh.harvard.edu/ibsr/data.html*.

[12] S.K. Warfield, K.H. Zou, and W.M. Wells, "Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation," *Medical Imaging, IEEE Transactions on*, vol. 23, no. 7, pp. 903–921, 2004.