

# From Brain Connectivity Models to Identifying Foci of a Neurological Disorder

Archana Venkataraman<sup>1</sup>, Marek Kubicki<sup>2</sup>, and Polina Golland<sup>1</sup>

<sup>1</sup> MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA

<sup>2</sup> Psychiatry Neuroimaging Laboratory, Harvard Medical School, Boston, MA

**Abstract.** We propose a novel approach to identify the foci of a neurological disorder based on anatomical and functional connectivity information. Specifically, we formulate a generative model that characterizes the network of abnormal functional connectivity emanating from the affected foci. We employ the variational EM algorithm to fit the model and to identify both the afflicted regions and the differences in connectivity induced by the disorder. We demonstrate our method on a population study of schizophrenia.

## 1 Introduction

Aberrations in functional connectivity inform us about neuropsychiatric disorders. Functional connectivity is measured via temporal correlations in resting-state functional Magnetic Resonance Imaging (fMRI) data [1]. Univariate tests and random effects analysis are commonly used in population studies of connectivity [2]. This approach relies on a statistical score, computed independently for each functional correlation, to determine significantly different connections within a clinical population. Multi-pattern analysis of functional connectivity has also been explored for clinical applications [3–5]. Although these studies identify functional connections affected by the disease, connectivity results are difficult to interpret and validate. Specifically, the bulk of our knowledge about the brain is organized around regions (i.e., functional localization, tissue properties, morphometry) and not the connections between them. Moreover, it is nearly impossible to design non-invasive experiments that target a particular connection between two brain regions.

In contrast to prior work, we propose a novel framework that pinpoints regions, which we call “foci”, whose functional connectivity patterns are the most disrupted by the disorder. Using a probabilistic setting, we define a latent (hidden) graph that characterizes the network of abnormal functional connectivity emanating from the affected brain regions. This generates population differences in the observed fMRI correlations. We employ the variational EM algorithm to fit the model to the observed data. Our algorithm jointly infers the regions affected by the disease and the induced connectivity differences.

We use neural anatomy as a substrate for modeling functional connectivity. In particular, we rely on Diffusion Weighted Imaging (DWI) tractography to

estimate the underlying white matter fibers in the brain. The latent anatomical connectivity inferred from these fibers constrains the graph of aberrant functional connections. Previous work in joint modeling of resting-state fMRI and DWI data [4, 6–8] suggests that a direct anatomical connection between two regions predicts a higher functional correlation; however, multi-stage pathways may explain some of the functional effects. Since neural communication in the brain is constrained by white matter fibers, we hypothesize that the strongest effects of a disorder will occur along direct anatomical connections. Hence, we model whole-brain functional connectivity but only use functional abnormalities between anatomically connected regions to identify the disease foci. We emphasize that our model can be readily applied to the complete graph of pairwise functional connections and need not incorporate anatomy.

We demonstrate that our method learns a stable set of afflicted regions on a population study of schizophrenia. Schizophrenia is a poorly understood disorder marked by impairments in widely-distributed functional and anatomical networks [2, 9]. Accordingly, we apply our model to whole-brain connectivity information. Our results identify the posterior cingulate and superior temporal gyri as most affected regions in schizophrenia.

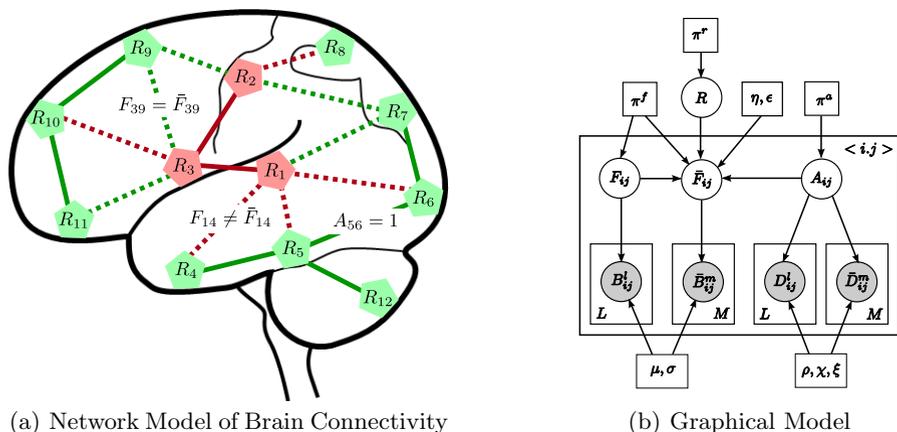
## 2 Generative Model and Inference

The basic assumption of our model is that impairments of the disorder localize to a small subset of brain regions, which we call foci, and affect the neural signaling along pathways associated with these regions. Fig. 1 presents a network diagram of the brain and the corresponding graphical model.

The nodes in Fig. 1(a) correspond to regions in the brain. The green nodes are healthy, and the red nodes are diseased. The edges denote neural connections, which are captured by latent anatomical connectivity  $A_{ij}$ . Specifically, the presence or absence of edge  $\langle i, j \rangle$  in the network is governed by the value of  $A_{ij}$ . The anatomical network structure is shared between the control and clinical populations. The regions in this work correspond to (large) Brodmann areas. Prior results in the field suggest that the anatomical differences between schizophrenia patients and normal controls are very small in this case.

Based on the region assignments, aberrant functional connectivity along anatomical pathways is defined using a simple set of rules: (1) a connection between two diseased regions is always abnormal (solid red lines in Fig. 1(a)), (2) a connection between two healthy regions is never abnormal (solid green lines), and (3) a connection between a healthy and a diseased region is abnormal with probability  $\eta$  (dashed lines). We use latent functional connectivity variables  $F_{ij}$  and  $\bar{F}_{ij}$  to model the neural synchrony between two regions in the control and clinical populations, respectively. Ideally,  $\bar{F}_{ij} \neq F_{ij}$  for abnormal connections and  $\bar{F}_{ij} = F_{ij}$  for healthy connections. However, due to noise, we assume that the latent templates can deviate from the above rules with probability  $\epsilon$ .

The observed DWI measurements  $D_{ij}^l$  and fMRI correlations  $B_{ij}^l$  provide noisy information about the latent network structure.



**Fig. 1.** (a) A network model of connectivity. The nodes correspond to regions in the brain, and the lines denote anatomical connections between them. The green nodes and edges are normal. The red nodes and red edges specify pathways of abnormal functional connectivity. The solid lines are deterministic given the region labels; the dashed lines are probabilistic. (b) Graphical model representation. Vector  $R$  specifies diseased regions.  $A_{ij}$  and  $F_{ij}$  represent the latent anatomical and functional connectivity, respectively, between regions  $i$  and  $j$ .  $D_{ij}^l$  and  $B_{ij}^l$  are the observed DWI and fMRI measurements in the  $l^{th}$  subject. Variables associated with the diseased population are identified by an overbar. Boxes denote non-random parameters; circles indicate random variables; shaded variables are observed.

**Disease Foci.** The random variable  $R = [R_1, \dots, R_N]$  is a binary vector that indicates the state, healthy ( $R_i = 0$ ) or diseased ( $R_i = 1$ ), for each brain region  $i$ . We assume an *i.i.d.* Bernoulli prior for the elements of  $R$ :

$$P(R_i; \pi^r) = (\pi^r)^{R_i} (1 - \pi^r)^{1-R_i}, \tag{1}$$

where  $\pi^r$  is an unknown parameter shared by all nodes in the network.

**Latent Connectivity.** We model anatomical connectivity  $A_{ij}$  as a binary random variable with *a priori* probability  $\pi^a$  that a connection is present:

$$P(A_{ij}; \pi^a) = (\pi^a)^{A_{ij}} (1 - \pi^a)^{1-A_{ij}}. \tag{2}$$

Latent functional connectivity  $F_{ij}$  is modeled as a tri-state random variable drawn from a multinomial distribution with parameter  $\pi^f$ . These states represent little or no functional co-activation ( $F_{ij} = 0$ ), positive functional synchrony ( $F_{ij} = 1$ ), and negative functional synchrony ( $F_{ij} = -1$ ). For convenience, we represent  $F_{ij}$  as a length-three indicator vector with exactly one of its elements  $[F_{ij,-1} \ F_{ij0} \ F_{ij1}]$  equal to one:

$$P(F_{ij}; \pi^f) = \prod_{k=-1}^1 (\pi_k^f)^{F_{ijk}}. \tag{3}$$

The latent functional connectivity  $\bar{F}_{ij}$  of the clinical population is also tri-state and is based on  $F_{ij}$  and the healthy/diseased indicator vector  $R$ :

$$P(\bar{F}_{ij}|F_{ij}, R_i, R_j, A_{ij}) = \begin{cases} (1 - \epsilon)^{F_{ij}^T \bar{F}_{ij}} \left(\frac{\epsilon}{2}\right)^{1 - F_{ij}^T \bar{F}_{ij}}, & A_{ij} = 1, R_i = R_j = 0, \\ \epsilon^{F_{ij}^T \bar{F}_{ij}} \left(\frac{1 - \epsilon}{2}\right)^{1 - F_{ij}^T \bar{F}_{ij}}, & A_{ij} = 1, R_i = R_j = 1, \\ \epsilon_1^{F_{ij}^T \bar{F}_{ij}} \left(\frac{1 - \epsilon_1}{2}\right)^{1 - F_{ij}^T \bar{F}_{ij}}, & A_{ij} = 1, R_i \neq R_j, \\ \mathcal{M}(\pi^f), & A_{ij} = 0, \end{cases} \quad (4)$$

such that  $\mathcal{M}(\cdot)$  is a multinomial distribution and  $\epsilon_1 = \eta\epsilon + (1 - \eta)(1 - \epsilon)$ . The first condition in Eq. (4) states that if there exists a latent anatomical connection ( $A_{ij} = 1$ ) and if both regions are healthy ( $R_i = R_j = 0$ ), then the edge  $\langle i, j \rangle$  is healthy. Consequently, the functional connectivity of the clinical population is equal to that of the control population with probability  $1 - \epsilon$ , and it differs with probability  $\epsilon$ . The second term is similarly obtained by replacing  $\epsilon$  with  $1 - \epsilon$ . The probability  $\epsilon_1$  in the third condition reflects the coupling between  $\eta$  and  $\epsilon$  when the region labels differ. The final term of Eq. (4) implies that  $\bar{F}_{ij}$  is drawn from the prior  $\pi^f$ , irrespective of  $F_{ij}$  and  $R$ , if there is no anatomical connection between the regions  $i$  and  $j$ .

**Data Likelihood.** The DWI measurement  $D_{ij}^l$  for the  $l^{th}$  subject in the control population is a noisy observation of the latent anatomical connectivity  $A_{ij}$ :

$$P(D_{ij}^l|A_{ij}; \{\rho, \chi, \xi^2\}) = \mathcal{P}_0(D_{ij}^l; \{\rho, \chi, \xi^2\})^{1 - A_{ij}} \cdot \mathcal{P}_1(D_{ij}^l; \{\rho, \chi, \xi^2\})^{A_{ij}}, \quad (5)$$

where  $\mathcal{P}_k(D_{ij}) = \rho_k \delta(D_{ij}) + (1 - \rho_k) \mathcal{N}(D_{ij}; \chi_k, \xi_k^2)$  for  $k = 0, 1$ .  $\rho_k$  represents the probability of failing to find a tract between two regions, which corresponds to  $D_{ij}^l = 0$ . Otherwise,  $D_{ij}^l$  is drawn from a Gaussian distribution with mean  $\chi_k$  and variance  $\xi_k^2$  ( $k = 0, 1$ ). The data  $\bar{D}_{ij}^m$  for the clinical population follows the same likelihood.

The BOLD fMRI correlation  $B_{ij}^l$  is a noisy observation of the functional connectivity  $F_{ij}$ :

$$P(B_{ij}^l|F_{ij}; \{\mu, \sigma^2\}) = \prod_{k=-1}^1 \mathcal{N}(B_{ij}^l; \mu_k, \sigma_k^2)^{F_{ijk}}. \quad (6)$$

We fix  $\mu_0 = 0$  to center the parameter estimates. The likelihood for the clinical population  $\bar{B}_{ij}^m$  has the same functional form and parameter values as Eq. (6) but uses the clinical template  $\bar{F}_{ij}$  instead of the control template  $F_{ij}$ .

Using histograms of the data, we verified that the Gaussian distributions in Eqs. (5-6) provide reasonable approximations for the DWI and fMRI data. Pragmatically, they greatly simplify the learning/inference steps.

**Variational EM.** We employ a maximum likelihood (ML) framework to fit the model to the data. The region variable  $R$  induces a complex coupling between pairwise connections. Therefore, we use a variational approximation [10] for the

latent posterior probability distribution when deriving the ML solution. Our variational posterior assumes the following form:

$$Q(R, A, F, \bar{F}) = Q^r(R) \cdot Q^c(A, F, \bar{F}) = Q^r(R) \prod_{\langle i,j \rangle} Q_{ij}^c(A_{ij}, F_{ij}, \bar{F}_{ij}), \quad (7)$$

where  $Q_{ij}^c(\cdot)$  is an 18-state multinomial distribution corresponding to all configurations of anatomical and functional connectivity. This factorization yields a tractable inference algorithm and also preserves the dependency between  $A_{ij}$ ,  $F_{ij}$ , and  $\bar{F}_{ij}$  given the region indicator vector  $R$ .

For a fixed setting of model parameters, we obtain the distribution  $Q(\cdot)$  that minimizes the variational free energy by alternatively updating  $Q^r(R)$  and  $Q^c(A, F, \bar{F})$  until convergence. Due to space constraints, we omit the update rules. We emphasize that both the posterior distribution  $Q(\cdot)$  and the model parameters are estimated directly from the observed data without tuning any auxiliary parameters.

**Model Evaluation.** The marginal posterior distribution  $Q^r(R_i = 1)$  tells us how likely region  $i$  is to be diseased given the observed connectivity data. We estimate the marginal posterior by averaging across Gibbs samples  $\mathcal{S}$ :

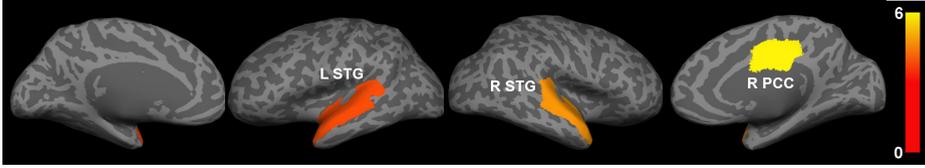
$$q_i = Q^r(R_i = 1) = \frac{1}{S} \sum_{s=1}^S R_i^s. \quad (8)$$

We evaluate the significance of our model through non-parametric permutation tests. To construct the null distribution for  $\{q_i\}$ , we randomly permute the subject diagnosis labels (NC vs. SZ) 1,000 times. For each permutation, we fit the model and compute the statistic in Eq. (8). The significance of each region is equal to the proportion of permutations that yield a larger value of  $q_i$  than the true labeling. These uncorrected p-values confirm that a particular region is rarely selected. Since our inference algorithm estimates the joint posterior distribution over the entire vector  $R$ , and since none of the permutations return the same set of disease foci, it is unclear that element-wise correction provides any additional insight.

Based on the MAP estimate of each  $R_i$  and the ML estimates of the model parameters, we can also construct the graph of functional connectivity differences to gain insight into the behavior of individual connections.

Our framework enables us to estimate all unknown parameters. However, we further explore the solution space by specifying the expected number of diseased regions via the prior  $\pi^r$ . In particular, the evolution of disease foci across a range of prior values (in this work  $\pi^r \in [0, 0.5]$ ) illustrates the stability of our model in explaining the data. Moreover, tuning  $\pi^r$  is an intuitive way to inject clinical knowledge into our framework and may be useful in certain applications.

Finally, we have run extensive simulations on synthetic data, which demonstrate that our model recovers the ground truth region labels. However, due to space constraints, we focus on real data in this paper.



**Fig. 2.** Significant regions based on permutation tests ( $q_i > 0.5$ , uncorrected  $p < 0.044$ ). The colorbar corresponds to the negative log p-value. We present the lateral and medial viewpoints for each hemisphere. The highlighted regions are the posterior cingulate (R PCC) and the superior temporal gyrus (L STG & R STG).

### 3 Results

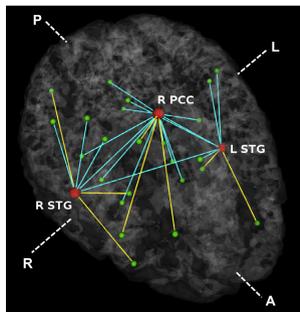
We demonstrate our model on a study of 19 male patients with chronic schizophrenia and 19 male healthy controls. For each subject, an anatomical scan (SPGR,  $TE = 3ms$ ,  $res = 1mm^3$ ), a diffusion-weighted scan (EPI,  $TE = 78ms$ ,  $res = 1.66 \times 1.66 \times 1.7mm$ , 51 gradient directions with  $b = 900s/mm^2$ , 8 baseline) and a resting-state functional scan (EPI-BOLD,  $TR = 3s$ ,  $TE = 30ms$ ,  $res = 1.875 \times 1.875 \times 3mm$ ) were acquired using a 3T GE Echospeed system.

We segment the anatomical images into 77 cortical and sub-cortical regions using FreeSurfer [11]. The DWI data is corrected for eddy-current distortions, and two-tensor tractography [12] is used to estimate the white matter fibers. The DWI measure  $D_{ij}^l$  is computed by averaging FA along all detected fibers between regions  $i$  and  $j$ .  $D_{ij}^l$  is set to zero if no tracts are found. We discard the first five fMRI time points and perform motion correction by rigid body alignment and slice timing correction using FSL [13]. The data is spatially smoothed using a  $5mm$  kernel, temporally low-pass filtered with  $0.08Hz$  cutoff, and motion corrected via linear regression. We also remove global contributions from the white matter, ventricles and the whole brain. The fMRI measure  $B_{ij}^l$  is the Pearson correlation coefficient between the mean time courses of regions  $i$  and  $j$ .

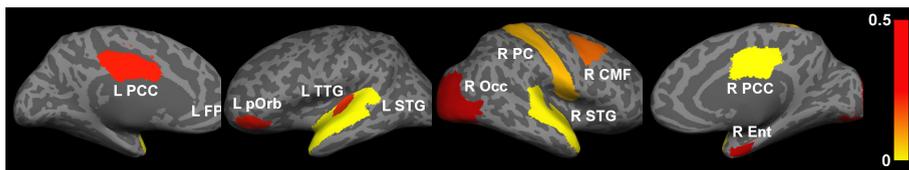
**Significant Regions.** Our method identifies three foci such that  $q_i > 0.5$ ; the uncorrected p-value of each region is less than 0.044. Fig. 2 displays the significant regions, which we color according to  $-\log(p)$ . Our results implicate the right posterior cingulate ( $q_i = 1, p < 0.004$ ), the right superior temporal gyrus ( $q_i = 1, p < 0.014$ ), and the left superior temporal gyrus ( $q_i = 1, p < 0.044$ ).

Prior studies have found abnormalities in the superior temporal gyri in schizophrenia [14]. These impairments correlate with clinical measures of auditory hallucination and attentional deficits. The default network has been implicated in resting-state fMRI studies [2]. Reduced connectivity in the posterior cingulate correlate with both positive and negative symptoms of schizophrenia.

In Fig. 3, we observe that functional abnormalities originating in the posterior cingulate project to the midbrain and frontal lobe, whereas abnormalities stemming from the right and left superior temporal gyri tend to span their



**Fig. 3.** Estimated graph of functional connectivity differences. The red nodes indicate the disease foci. Blue lines indicate reduced functional connectivity and yellow lines indicate increased functional connectivity in the schizophrenia population.



**Fig. 4.** Evolution of the disease foci when varying the region prior  $\pi^r$ . The color corresponds to the smallest value of  $\pi^r$  such that  $q_i > 0.2$ . The highlighted regions are the posterior cingulate (L PCC & R PCC), the superior temporal gyrus (L STG & R STG), the postcentral gyrus (R PC), the frontal pole (L FP), the caudal middle frontal gyrus (R CMF), the transverse temporal gyrus (L TTG), the pars orbitalis (L pOrb), the entorhinal cortex (R Ent) and the lateral occipital cortex (R LOcc).

respective hemispheres. Overall, schizophrenia patients exhibit reduced functional connectivity. Of notable exception are connections to the frontal lobe. This phenomenon has been reported in prior studies of schizophrenia [9] and is believed to interfere with perception by misdirecting attentional resources.

**Effects of the Region Prior.** Fig. 4 illustrates the results of varying the prior  $\pi^r$  for the region indicator vector  $R$ . Empirically, we observe that sets of regions affected by the disease form a nested structure as  $\pi^r$  increases. We color each of the selected regions according to the smallest value of  $\pi^r$  such that the marginal posterior  $q_i > 0.2$ . The yellow regions are always identified as foci, whereas the orange/red regions are selected for larger values of the prior  $\pi^r$ .

The nesting property is a highly desirable feature of our model. It suggests an initial set of disease foci, identical to the significant regions in Fig. 2. We can then tune a single scalar to progressively include regions that exhibit some functional abnormalities but are not as strongly implicated by the data.

## 4 Conclusion

We proposed a novel probabilistic framework for multimodal analysis of fMRI and DWI data that integrates population differences in connectivity to isolate

foci of a neurological disorder. This is achieved by defining a network of abnormal connectivity emanating from the affected regions. We demonstrate that our method identifies a stable set of schizophrenia foci consisting of the right posterior cingulate and the right and left superior temporal gyri. Prior clinical studies have linked these regions to the effects of schizophrenia. Moreover, we uncover additional regions by adjusting the prior on the number of disease foci. These results establish the promise of our approach for aggregating connectivity information to localize region effects.

**Acknowledgments.** This work was supported by the National Alliance for Medical Image Computing (NIH NIBIB NIMIC U54-EB005149), the Neuroimaging Analysis Center (NIH NCRR NAC P41-RR13218 and NIH NIBIB NAC P41-EB-015902) and NSF CAREER Grant 0642971. A. Venkataraman is supported in part by the NIH Advanced Multimodal Neuroimaging Training Program.

## References

1. Fox, M., Raichle, M.: Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature* 8, 700–711 (2007)
2. Bluhm, R., et al.: Spontaneous low-frequency fluctuations in the bold signal in schizophrenic patients: Abnormalities in the default network. *Schizophrenia Bulletin*, 1–9 (2007)
3. Jafri, M., et al.: A method for functional network connectivity among spatially independent resting-state components in schiz. *NeuroImage* 39, 1666–1681 (2008)
4. Venkataraman, A., et al.: Joint modeling of anatomical and functional connectivity for population studies. *IEEE TMI* 31, 164–182 (2012)
5. Venkataraman, A., et al.: Robust feature selection in resting-state fmri connectivity based on population studies. In: *MMBIA*, pp. 63–70 (2010)
6. Greicius, M., et al.: Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cerebral Cortex* 19, 72–78 (2008)
7. Koch, M., et al.: An investigation of functional and anatomical connectivity using magnetic resonance imaging. *NeuroImage* 16, 241–250 (2002)
8. Deligianni, F., Varoquaux, G., Thirion, B., Robinson, E., Sharp, D.J., Edwards, A.D., Rueckert, D.: A Probabilistic Framework to Infer Brain Functional Connectivity from Anatomical Connections. In: Székely, G., Hahn, H.K. (eds.) *IPMI 2011*. LNCS, vol. 6801, pp. 296–307. Springer, Heidelberg (2011)
9. Gabrieli-Whitfield, S., et al.: Hyperactivity and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia. *PNAS* 106, 1279–1284 (2009)
10. Jordan, M., et al.: An introduction to variational methods for graphical models. *Machine Learning* 37, 183–233 (1999)
11. Fischl, B., et al.: Sequence-independent segmentation of magnetic resonance images. *NeuroImage* 23, 69–84 (2004)
12. Malcolm, J., et al.: A filtered approach to neural tractography using the watson directional function. *NeuroImage* 14, 58–69 (2010)
13. Smith, S., et al.: Advances in functional and structural mr image analysis and implementation as fsl. *NeuroImage* 23(51), 208–219 (2004)
14. Lee, K., et al.: Increased diffusivity in superior temporal gyrus in patients with schizophrenia: A diffusion tensor imaging study. *Schiz Research* 104, 33–40 (2009)