

From Connectivity Models to Region Labels: Identifying Foci of a Neurological Disorder

Archana Venkataraman*, Marek Kubicki, and Polina Golland

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. This allows us to aggregate pairwise connectivity changes into a region-based representation of the disease. We employ the variational expectation-maximization algorithm to fit the model and subsequently identify both the afflicted regions and the differences in connectivity induced by the disorder. We demonstrate our method on a population study of schizophrenia.

Index Terms—Brain connectivity, diffusion weighted imaging (DWI), functional magnetic resonance imaging (fMRI), population analysis.

I. INTRODUCTION

ABERRATIONS in functional connectivity are often correlated with neuropsychiatric disorders. Functional connectivity is typically measured via temporal correlations in resting-state functional magnetic resonance imaging (fMRI) data [1], [2]. Univariate tests and random effects analysis are commonly used in population studies of connectivity [3]–[6]. This approach relies on a statistical score, computed independently for each functional correlation, to determine connections that differ between a clinical population and normal controls. Multi-pattern analysis of functional connectivity has also been explored for clinical applications [7]–[10]. Although the above studies identify functional connections influenced 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 noninvasive

experiments that target a particular connection between two brain regions. In this paper, we propose and demonstrate an approach to pinpoint regions, which we call “foci,” whose connectivity patterns are most disrupted by the disorder. We identify the disease foci from a set of predefined regions, as specified by an atlas.

Our method effectively translates differences in connectivity between a control and a clinical population into estimates of the regions associated with the disease. Using a probabilistic setting, we define a *latent* or *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 expectation-maximization (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. To the best of our knowledge, ours is the first stochastic model to relate connectivity information to region labels.

We present two versions of the model. The first variant considers the complete graph of pairwise functional connections. The second model uses 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 suggests that a direct anatomical connection between two regions predicts a higher functional correlation [9], [11]–[13]; however, many functional effects can be attributed to complex multi-stage pathways. Since neural communication between brain regions 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. This paper extends our prior work [9] on the joint estimation of anatomical and functional connectivity; aggregating the latent connectivity information to predict region effects is the novel contribution presented here. A preliminary version of this work was presented at the International Conference on Medical Image Computing and Computer Assisted Intervention [14]. In this paper we provide more detailed derivations of the model and estimation procedure and include more extensive experimental evaluation of the methods.

We demonstrate that our methods learns a stable set of afflicted regions in a population study of schizophrenia.

Manuscript received May 29, 2013; accepted July 04, 2013. Date of publication July 10, 2013; date of current version October 28, 2013. This work was supported in part by the National Alliance for Medical Image Computing under Grant NIH NIBIB NAC P41-RR13218 and Grant NIH NIBIB NAC P41-EB-015902, and in part by the National Science Foundation CAREER Grant 0642971 and NIH R01MH074794. The work of A. Venkataraman was supported in part by the NIH Advanced Multimodal Neuroimaging Training Program. *Asterisk indicates corresponding author.*

*A. Venkataraman is with the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: pegas85@csail.mit.edu).

M. Kubicki is affiliated with the Psychiatry Neuroimaging Laboratory, Harvard Medical School, Boston, MA 02215 USA (e-mail: kubicki@bwh.harvard.edu).

P. Golland are affiliated with the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: polina@csail.mit.edu).

Digital Object Identifier 10.1109/TMI.2013.2272976

Schizophrenia is a poorly understood disorder marked by widespread cognitive difficulties affecting intelligence, memory, and executive attention. These impairments are not localized to a particular cortical region; rather, they reflect abnormalities in widely-distributed functional and anatomical networks [15], [16]. Accordingly, we apply our model to whole-brain connectivity information. Our results identify the posterior cingulate, the superior temporal gyri and the transverse temporal gyri as the most affected regions in schizophrenia.

The remainder of this paper is organized as follows. Section II summarizes prior research on functional connectivity and multimodal analyses. We also review clinical findings of schizophrenia. We introduce our generative model in Section III and develop the corresponding inference algorithm in Section IV. Section V presents the framework used for the empirical validation of our approach. Sections VI and VII report experimental results based on synthetic and clinical data, respectively. Section VIII discusses the behavior of our model, its advantages and drawbacks, and future directions of research. We conclude with a summary of our contributions in Section IX.

II. PRIOR WORK

A. Functional Connectivity for Clinical Applications

fMRI studies can be divided into two broad categories. *Task-based* studies measure the response to a given experimental paradigm in order to localize brain function [17]. In contrast, *resting-state* fMRI captures spontaneous, low-frequency oscillations. Temporal correlations between these signals reflect the intrinsic functional connectivity between brain regions [1], [2]. Resting-state fMRI is particularly attractive for clinical populations, since patients are not required to perform challenging experimental paradigms.

Univariate tests and random effects analysis are, to a great extent, the standard in population studies of connectivity [3], [4], [18]. These methods identify significantly different connections using a statistical score that is computed independently for each functional correlation. Consequently, the analysis ignores important *networks* of connectivity within the brain.

Prior work has also explored multi-pattern analysis for functional connectivity [7]–[10]. For example, [7], [8] employ group independent component analysis to represent the fMRI data as a set of spatially-independent regions with associated time courses. In [7], group functional connectivity is computed as the maximum lagged correlation between the estimated time courses; the two-sample t-test is used to identify significant population differences. In [8], a neural network is constructed for patient classification of first-episode schizophrenia. Similarly, the method of [10] uses a metric called Gini Importance [19] to summarize multivariate patterns of interaction. When trained on these measures, a classifier for a clinical population exhibited superior accuracy than when trained on univariate statistics. Finally, [9] presents a probabilistic framework for connectivity analysis. Differences between two populations are explained via changes in the latent anatomical and functional connectivity graphs.

Despite the progress made to robustly identify functional connectivity patterns associated with a disease, the results are difficult to validate and to interpret. For example, due to variations in preprocessing and region definitions, relatively few functional connections are consistently reported in clinical studies. Moreover, the relationship between functional activation and functional connectivity is poorly-understood; hence, it is challenging to incorporate connectivity results into the knowledge gained from task-based fMRI studies. Finally, short of direct stimulation, we do not know how to design *in vivo* experiments that target a particular connection between two brain regions. In contrast to prior work, we propose a framework that consolidates population changes in functional connectivity to localize hubs of a disease. The results can be easily compared and integrated with other sources of information about the detected regions.

B. Multimodal Analysis of Connectivity

In addition to purely functional analysis, we explore the relationship between functional connectivity and anatomical connectivity, as measured by DWI tractography. DWI captures the anisotropic diffusion of water throughout the brain and is often used to estimate the underlying white matter bundles via tractography. Common measures of anatomical connectivity include the probability of diffusion between two brain regions, the number of fibers linking two regions, and the mean fractional anisotropy (FA) along the tracts [20].

Early work in multimodal analysis computed statistics of the fMRI and DWI signals (such as fMRI correlations, fractional anisotropy values, etc.) and searched for correspondences between these metrics *a posteriori* [21], [22], [12]. This method has yielded important insights into the nature of connectivity in the brain. For example, it has been shown that while a high degree of anatomical connectivity predicts higher functional correlations, the converse does not always hold [21]. In particular, strong functional correlations can be found between spatially distributed locations in the brain, whereas one is more likely to identify white matter tracts connecting nearby regions. A notable exception is the recently demonstrated approach in [23] where the authors construct cortical connection graphs based on histological data of the macaw brain and simulate the corresponding functional correlations using a dynamical system. However, this method has not been demonstrated on *in vivo* human data.

Recent studies explicitly model the interaction between resting-state fMRI and DWI data by attempting to predict functional information based on anatomy [13], [24]. The work of [24] explores how well the anatomical network structure explains large-scale properties of functional systems. The findings are verified using a computational model of the brain. The method of [13] uses a sparse multivariate autoregressive model and multivariate linear regression to determine which anatomical connections contribute to a particular functional correlation. Our alternative methodology in [9] infers latent connectivity differences based on the fMRI and DWI values.

TABLE I
RANDOM VARIABLES (TOP) AND NONRANDOM PARAMETERS (BOTTOM) IN OUR GRAPHICAL MODELS SHOWN IN FIGS. 1 AND 2

R	Binary vector that indicates the state (healthy vs. diseased) for each brain region i
A_{ij}	Latent anatomical connectivity between regions i and j
T_{ij}	Edge (i, j) in the latent graph of abnormal functional connectivity
F_{ij}	Latent functional connectivity between regions i and j in the control population
\bar{F}_{ij}	Latent functional connectivity between regions i and j in the clinical population
B_{ij}^l	Observed fMRI correlation between regions i and j in subject l of the control population
\bar{B}_{ij}^m	Observed fMRI correlation between regions i and j in subject m of the clinical population
D_{ij}^l	Observed DWI connectivity measure between regions i and j in subject l of the control population
\bar{D}_{ij}^m	Observed DWI connectivity measure between regions i and j in subject m of the clinical population
π^r	Prior for binary region indicator R_i
π^f	Prior for multinomial functional connectivity F_{ij}
π^a	Prior for binary anatomical connectivity A_{ij}
η	Probability of a diseased connection between a healthy and diseased node
ϵ	Probability of deviating from the latent graph of aberrant functional connectivity
μ_k, σ_k^2	Mean and variance of fMRI correlation given that $F_{ij} = k$ ($k = -1, 0, 1$)
ρ_k	Probability of failing to find a white matter tract given that $A_{ij} = k$ ($k = 0, 1$)
χ_k, ξ_k^2	Mean and variance of DWI connectivity measure if there is a white matter tract given that $A_{ij} = k$ ($k = 0, 1$)

Specifically, we use anatomy to inform the functional connectivity graph but do not try to merge the population differences within the two modalities. In this paper, we carry the analysis one step further and infer properties of individual brain regions from connectivity data.

C. Schizophrenia: Findings and Hypotheses

Schizophrenia is a neuropsychiatric disorder characterized by gross distortions in the perception of reality. Despite generating considerable interest within the neuroscience community, the origins and expression of the disease are still poorly understood [25]. For example, structural findings only weakly and inconsistently correlate with the clinical and cognitive symptoms of schizophrenia [26]. Similarly, functional experiments report deficits in many cognitive domains, most notably memory and attention, but do not consistently identify clinical correlates [27].

At present, the cognitive impairments of schizophrenia are thought to reflect underlying abnormalities in distributed brain networks. In particular, schizophrenia may compromise neural communication between multiple cortical regions [28]. Recent studies have also focused on the degeneration of anatomical connectivity [29], motivated in part by post mortem evidence of myelination anomalies in patients with schizophrenia.

Findings from resting-state fMRI data include reduced connectivity in the brain's default network [30], [31], dorsolateral prefrontal cortex [18], and a widespread reduction in connectivity throughout the brain [4]. The superior temporal gyrus has been implicated using diffusivity measures [32] and volume changes [26]. Our method bridges the gap between connectivity differences and region effects in schizophrenia.

To summarize, prior work in connectivity analysis has focused on properties of connections and provides little information about regions in the brain. This makes it difficult to interpret results across different imaging techniques. In the next section, we present a novel approach that translates differences in connectivity between a control and a clinical population into estimates of the regions associated with the disorder.

III. GENERATIVE MODEL

We assume that the disorder is characterized by impairments in a small subset of brain regions, which we designate as *foci*. The impairments affect neural signaling along pathways associated with the diseased regions. We use a probabilistic framework to represent the interaction between regions and the effects of the disease. In particular, *latent* variables specify a template organization of the brain, which we cannot directly access. Instead, we observe noisy measurements of the hidden structure via resting-state fMRI correlations and DWI tractography. The fMRI and DWI signals vary across subjects; we assume they are generated stochastically from a group-wise latent template shared by all subjects.

We first develop the model for functional data. This formulation serves as a foundation for incorporating anatomical connectivity, as presented later in the section. Table I summarizes our notation in this paper.

A. Functional Model

Fig. 1 presents a network diagram of the brain and the corresponding graphical model for the functional connectivity data. The nodes in Fig. 1(a) denote regions in the brain, and the edges correspond to pairwise functional connections between them. The green nodes/edges are healthy and the red nodes/edges are diseased.

Based on the region assignments, we define a binary graph T of aberrant functional connectivity using a simple set of rules: 1) a connection between two diseased regions is always abnormal [$T_{ij} = 1$, solid red lines in Fig. 1(a)], 2) a connection between two healthy regions is always healthy ($T_{ij} = 0$, solid green lines), and 3) a connection between a healthy and a diseased region is abnormal with probability η (dashed lines). We use the 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 and intersubject variability, we assume that the latent templates can deviate from the above rules with (small but unknown) probability ϵ .

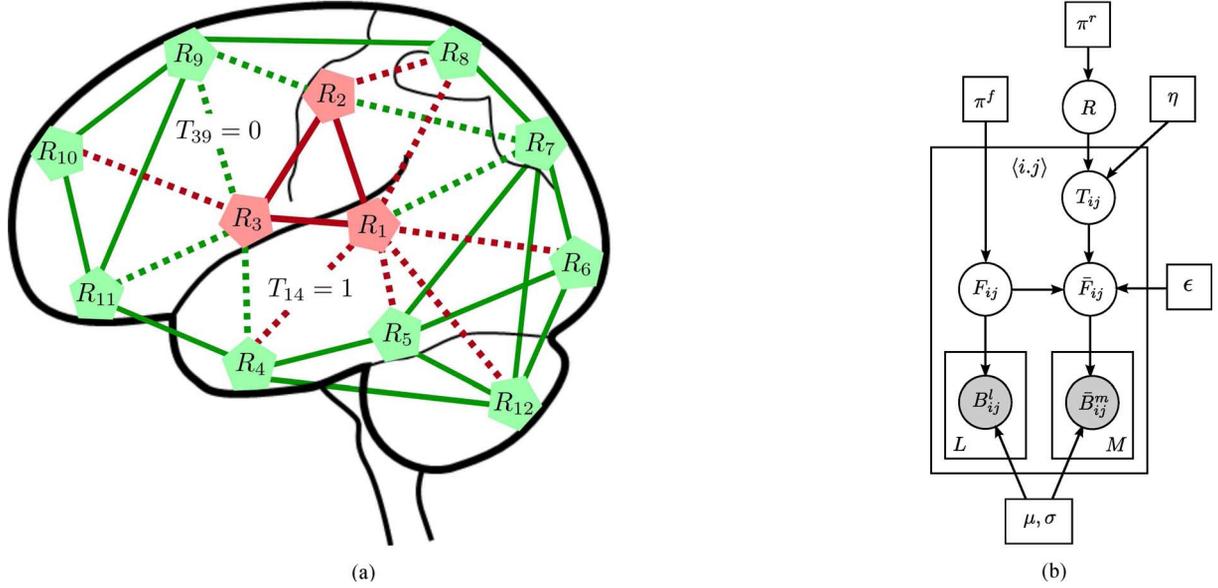


Fig. 1. (a) Network model of connectivity for the functional data. The nodes correspond to regions in the brain, and the lines denote pairwise functional connections between them. Only a subset of edges is shown; the model is defined on the full graph of pairwise connections. The green nodes and edges correspond to the normal regions and connections, respectively. The red nodes are foci of the disease, and the red edges specify pathways of abnormal functional connectivity. The solid lines are deterministic given the region labels; the dashed lines are probabilistic. (b) The corresponding graphical model. Vector R specifies diseased regions. F_{ij} denotes the latent functional connectivity between regions i and j . B_{ij}^l is the observed fMRI measurements in the l th subject. Variables associated with the diseased population are identified by an overbar. Boxes denote nonrandom parameters; circles indicate random variables; shaded variables are observed.

The observed fMRI correlations B_{ij}^l provide noisy information about the latent network structure. Below, we formalize our generative model.

a) *Disease Foci*: Let N be the total number of regions in the brain. 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 ($i = 1, \dots, N$). We assume an *i.i.d.* Bernoulli prior for the elements of R :

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

where the scalar parameter π^r specifies the *a priori* probability that a region is diseased. The prior is shared by all nodes in the network.

b) *Graph of Abnormal Connectivity*: The binary graph T represents the abnormal functional connectivity emanating from the disease foci. Each edge T_{ij} is generated independently given the labels of regions i and j

$$P(T_{ij} | R_i, R_j; \eta) = \begin{cases} \delta_T(T_{ij}), & R_i = R_j = 0 \\ 1 - \delta_T(T_{ij}), & R_i = R_j = 1 \\ \eta^{T_{ij}} (1 - \eta)^{1-T_{ij}}, & R_i \neq R_j \end{cases} \quad (2)$$

where $\delta_T(\cdot)$ is an indicator function that equals to one if and only if its argument is zero, and η is the scalar parameter that represents the probability of a connection between a healthy and a diseased region being altered.

c) *Latent Functional Connectivity*: We model the latent functional connectivity F_{ij} of the control population as a tri-state random variable drawn from a multinomial distribution with parameter π^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 notational convenience, we represent F_{ij} as a length-three indicator vector with exactly one of its elements $[F_{ij,-1} \ F_{ij,0} \ F_{ij,1}]$ equal to one, i.e.,

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

The latent functional connectivity \bar{F}_{ij} of the clinical population is also tri-state and is based on F_{ij} and the graph T_{ij} . If the edge $\langle i, j \rangle$ is healthy ($T_{ij} = 0$), the functional connectivity of the clinical population is equal to that of the control population with probability $1 - \epsilon$, and it differs with probability ϵ . Conversely, if the edge $\langle i, j \rangle$ is diseased ($T_{ij} = 1$), then the functional connectivity of the clinical population differs from the control population with probability $1 - \epsilon$, and it is equal with probability ϵ . Formally,

$$P(\bar{F}_{ij} | F_{ij}, T_{ij}; \epsilon) = \begin{cases} (1 - \epsilon)^{F_{ij}^T \bar{F}_{ij}} \left(\frac{\epsilon}{2}\right)^{1-F_{ij}^T \bar{F}_{ij}}, & T_{ij} = 0 \\ \epsilon^{F_{ij}^T \bar{F}_{ij}} \left(\frac{1-\epsilon}{2}\right)^{1-F_{ij}^T \bar{F}_{ij}}, & T_{ij} = 1. \end{cases} \quad (4)$$

d) *fMRI Likelihood*: Let L be the number of subjects in the control population and M be the number of subjects in the clinical population. The BOLD fMRI correlation B_{ij}^l between regions i and j in the l th subject of the control population is a noisy observation of the functional connectivity indicator F_{ij} . In particular, B_{ij}^l is a Gaussian random variable whose mean and variance depend on the value of F_{ij}

$$P(B_{ij}^l | F_{ij}; \{\mu, \sigma^2\}) = \prod_{k=-1}^1 [\mathcal{N}(B_{ij}^l; \mu_k, \sigma_k^2)]^{F_{ijk}} \quad (5)$$

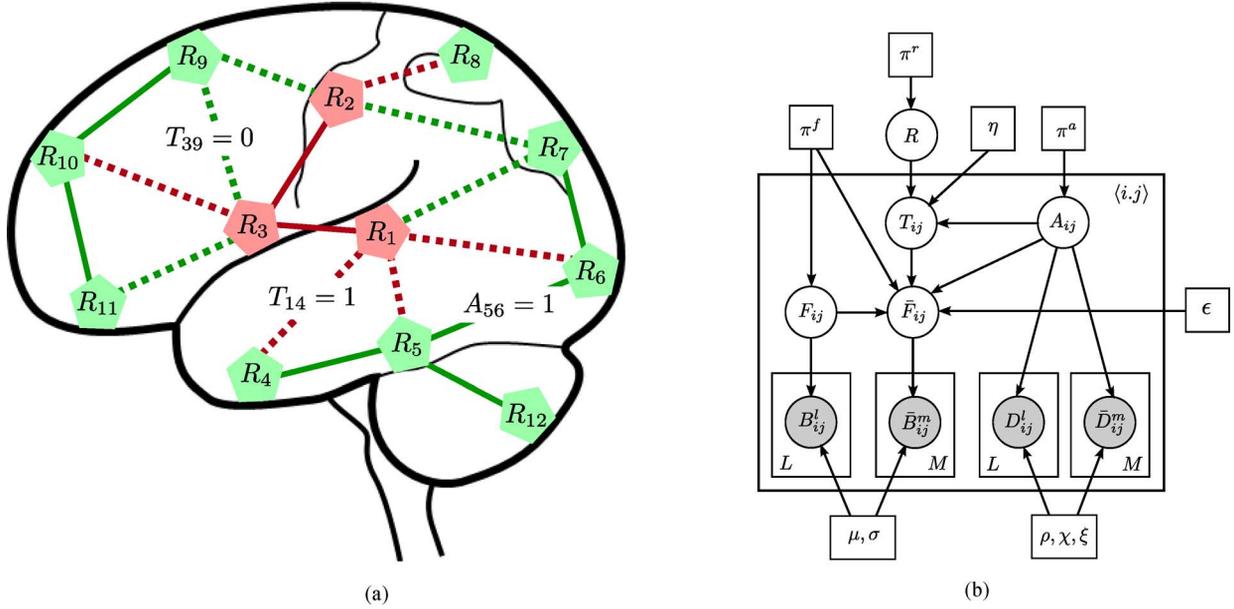


Fig. 2. (a) Network model of connectivity. The nodes correspond to regions in the brain, and the lines denote anatomical connections between them. The green nodes and edges correspond to the normal regions and connections, respectively. The red nodes are foci of the disease, and the red edges specify pathways of abnormal functional connectivity. The solid lines are deterministic given the region labels; the dashed lines are probabilistic. (b) Corresponding graphical model. Vector R specifies diseased regions. A_{ij} represents the latent anatomical connectivity between regions i and j . F_{ij} denotes the corresponding latent functional connectivity. D_{ij}^l and B_{ij}^l are the observed DWI and fMRI measurements, respectively, in the l th subject. Variables associated with the diseased population are identified by an overbar. Boxes denote nonrandom parameters; circles indicate random variables; shaded variables are observed. (a) Network model of brain connectivity. (b) Graphical model.

where $\mathcal{N}(\cdot; \mu, \sigma^2)$ denotes a Gaussian distribution with mean μ and variance σ^2 . 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 (5) but uses the clinical template \bar{F}_{ij} instead of the control template F_{ij} . In this work, we compute B_{ij}^l using Pearson correlation coefficients. Our empirical analysis in [9] suggests that the Gaussian likelihood in (5) reasonably approximates the data distribution. We describe these experiments in detail in Section III-B.

B. Multimodal Analysis

Since functional communication in the brain is constrained by neural axons, our second model assumes that the salient effects of a disorder occur along anatomical pathways. This extension is illustrated in Fig. 2. The edges in Fig. 2(a) correspond to neural connections, which are captured by latent anatomical connectivity A_{ij} . Specifically, the presence or absence of an edge $\langle i, j \rangle$ in the network is governed by the binary 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 [9]. Once again, the observed DWI measurements D_{ij}^l and fMRI correlations B_{ij}^l provide noisy information about the latent network structure.

a) *Latent Anatomical Connectivity*: The latent anatomical connectivity variable A_{ij} indicates the presence or absence of a *direct* anatomical pathway between regions i and j . It does not quantify the number or trajectory of the underlying neural

fibers. We model A_{ij} as a binary random variable with *a priori* probability π^a that a connection is present

$$P(A_{ij}; \pi^a) = (\pi^a)^{A_{ij}} (1 - \pi^a)^{1 - A_{ij}}. \quad (6)$$

b) *Graph of Abnormal Connectivity*: The binary graph T of aberrant functional connectivity is now defined along latent anatomical pathways. Therefore, we modify the rules from Section III-A and generate the edge T_{ij} between regions i and j as follows:

$$P(T_{ij} | R_i, R_j; \eta) = \begin{cases} \delta_T(T_{ij}), & A_{ij} = 0 \\ \delta_T(T_{ij}), & A_{ij} = 1, R_i = R_j = 0 \\ 1 - \delta_T(T_{ij}), & A_{ij} = 1, R_i = R_j = 1 \\ \eta^{T_{ij}} (1 - \eta)^{1 - T_{ij}}, & A_{ij} = 1, R_i \neq R_j. \end{cases} \quad (7)$$

In particular, we assume that $T_{ij} = 0$ when the corresponding anatomical connection is absent.

c) *Functional Connectivity of the Clinical Population*: We adapt the distribution for the latent functional connectivity \bar{F}_{ij} in (4) to reflect the anatomical constraint

$$P(\bar{F}_{ij} | F_{ij}, T_{ij}, A_{ij}; \epsilon) = \begin{cases} (1 - \epsilon)^{F_{ij}^T \bar{F}_{ij}} \left(\frac{\epsilon}{2}\right)^{1 - F_{ij}^T \bar{F}_{ij}}, & T_{ij} = 0, A_{ij} = 1 \\ \epsilon^{F_{ij}^T \bar{F}_{ij}} \left(\frac{1 - \epsilon}{2}\right)^{1 - F_{ij}^T \bar{F}_{ij}}, & T_{ij} = 1, A_{ij} = 1 \\ \prod_{k=-1}^1 \left(\pi_k^f\right)^{\bar{F}_{ij}^k}, & A_{ij} = 0. \end{cases} \quad (8)$$

If there is a latent anatomical connection between regions i and j ($A_{ij} = 1$), then \bar{F}_{ij} is generated according to (4). If there is no anatomical connection ($A_{ij} = 0$), then the final term in (8)

implies that \bar{F}_{ij} is drawn from the prior π^f , irrespective of F_{ij} and T_{ij} .

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

$$P(D_{ij}^l | A_{ij}; \{\rho, \chi, \xi^2\}) = \mathcal{P}_0(D_{ij}^l; \{\rho, \chi, \xi^2\})^{1-A_{ij}} \cdot \mathcal{P}_1(D_{ij}^l; \{\rho, \chi, \xi^2\})^{A_{ij}} \quad (9)$$

where $\mathcal{P}_k(D_{ij}) = \rho_k \delta(D_{ij}) + (1 - \rho_k) \mathcal{N}(D_{ij}; \chi_k, \xi_k^2)$ for $k = 0, 1$. The data \bar{D}_{ij}^m of the clinical population follows the same likelihood. The parameter ρ_k represents the probability of failing to find a tract between two regions, which corresponds to $D_{ij}^l = 0$. Ideally, $\rho_0 = 1$ and $\rho_1 = 0$, i.e., a white matter tract should be found if and only if there is an underlying anatomical connection. However, detection via tractography is imperfect. Consequently, our observation model explicitly accounts for missing tracts between anatomically connected regions and spurious tracts between isolated ones by allowing $\rho_0, \rho_1 \in [0, 1]$. If we identify one or more white matter fibers between regions i and j , the value of D_{ij}^l is modeled as a Gaussian random variable whose mean χ and variance ξ^2 depend on anatomical connectivity A_{ij} . In this work we use the average FA along white matter as our DWI measure. As demonstrated in [9], the Gaussian distributions in (5) and (9) adequately capture the empirical data distributions. To generate these results, we first approximate the discrete latent connectivity templates \hat{A} and \hat{F} from the data. For anatomical connectivity, we threshold the proportion of subjects that exhibit white matter tracts between region i and region j to obtain individual values \hat{A}_{ij} . For functional connectivity, we compute the average fMRI correlation across subjects for each pair of regions. We cluster these values across connections to obtain the labels \hat{F}_{ij} . Once we have the latent assignments, we fit a Gaussian distribution to the observed fMRI and DWI data under each configuration of latent anatomical and functional connectivity. Qualitatively, we observe that variabilities in the data across connections and subjects are reasonably approximated via Gaussian distributions.

We emphasize that our model can be readily extended to accommodate other measures of connectivity by redefining the data likelihood term.

IV. INFERENCE

Since we are primarily interested in the region labels R , we opt to marginalize out the graph structure T . This simplifies the relationship between R and the observed data.

The only term which is affected by the marginalization is the conditional distribution of the clinical template \bar{F}_{ij} , which now depends on R_i and R_j . Specifically, we have

$$P(\bar{F}_{ij} | F_{ij}, R_i, R_j; \eta, \epsilon) = \begin{cases} (1 - \epsilon)^{F_{ij}^T \bar{F}_{ij}} \left(\frac{\epsilon}{2}\right)^{1 - F_{ij}^T \bar{F}_{ij}}, & R_i = R_j = 0 \\ \epsilon^{F_{ij}^T \bar{F}_{ij}} \left(\frac{1 - \epsilon}{2}\right)^{1 - F_{ij}^T \bar{F}_{ij}}, & R_i = R_j = 1 \\ \epsilon_1^{F_{ij}^T \bar{F}_{ij}} \left(\frac{1 - \epsilon_1}{2}\right)^{1 - F_{ij}^T \bar{F}_{ij}}, & R_i \neq R_j \end{cases} \quad (10)$$

for the functional model and

$$P(\bar{F}_{ij} | F_{ij}, R_i, R_j, A_{ij}; \eta, \epsilon) = \begin{cases} (1 - \epsilon)^{F_{ij}^T \bar{F}_{ij}} \left(\frac{\epsilon}{2}\right)^{1 - F_{ij}^T \bar{F}_{ij}}, & A_{ij} = 1, R_i = R_j = 0 \\ \epsilon^{F_{ij}^T \bar{F}_{ij}} \left(\frac{1 - \epsilon}{2}\right)^{1 - F_{ij}^T \bar{F}_{ij}}, & A_{ij} = 1, R_i = R_j = 1 \\ \epsilon_1^{F_{ij}^T \bar{F}_{ij}} \left(\frac{1 - \epsilon_1}{2}\right)^{1 - F_{ij}^T \bar{F}_{ij}}, & A_{ij} = 1, R_i \neq R_j \\ \prod_{k=-1}^1 (\pi^f)^{\bar{F}_{ij}^k}, & A_{ij} = 0 \end{cases} \quad (11)$$

for the joint model, where $\epsilon_1 = \eta\epsilon + (1 - \eta)(1 - \epsilon)$. It is easy to see that ϵ_1 reflects the coupling between the graph prior η and latent noise variable ϵ when the region labels differ.

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 forcing us to adopt a variational approximation [33] for the posterior probability distribution when deriving the EM algorithm for nonrandom parameter estimation.

A. Functional Model

Let $Y = \{B, \bar{B}\}$ and $\Theta = \{\pi, \eta, \epsilon, \mu, \sigma^2\}$ denote the fMRI observations and the set of model parameters, respectively. Our variational posterior assumes the following form:

$$Q(R, F, \bar{F}) = Q^r(R) \cdot Q^c(F, \bar{F}) = Q^r(R) \prod_{\langle i, j \rangle} Q_{ij}^c(F_{ij}, \bar{F}_{ij}) \quad (12)$$

where $Q^r(\cdot)$ is a distribution over the length- N binary vector R and $Q_{ij}^c(\cdot)$ is a nine-state multinomial distribution corresponding to all configurations of latent functional connectivity. This factorization yields a tractable inference algorithm while preserving the dependency between F_{ij} , and \bar{F}_{ij} given the region indicator vector R .

We use a variational EM formulation [34] to obtain the posterior distribution $Q(\cdot)$ and model parameters Θ which minimize the variational free energy

$$\mathcal{F}\mathcal{E} = -E_Q[\log P(R, F, \bar{F}, Y; \Theta)] - \mathcal{H}(Q) \quad (13)$$

where the joint log-likelihood of all hidden and observed variables is obtained by combining the prior and likelihood distributions from Section III-A with (10)

$$\begin{aligned} \log P(R, F, \bar{F}, Y; \Theta) &= \log(\pi^r) \sum_{i=1}^N R_i + \log(1 - \pi^r) \sum_{i=1}^N (1 - R_i) \\ &+ \sum_{\langle i, j \rangle} \sum_{k=-1}^1 F_{ijk} \log(\pi_k^f) \\ &+ \sum_{\langle i, j \rangle} (1 - R_i)(1 - R_j) \\ &\times \left(F_{ij}^T \bar{F}_{ij} \log(1 - \epsilon) + (1 - F_{ij}^T \bar{F}_{ij}) \log\left(\frac{\epsilon}{2}\right) \right) \end{aligned}$$

$$\begin{aligned}
& + \sum_{\langle i,j \rangle} (R_i R_j) \left(F_{ij}^T \bar{F}_{ij} \log(\epsilon) + (1 - F_{ij}^T \bar{F}_{ij}) \log\left(\frac{1-\epsilon}{2}\right) \right) \\
& + \sum_{\langle i,j \rangle} (R_i(1 - R_j) + (1 - R_i)R_j) (F_{ij}^T \bar{F}_{ij} \log(\epsilon_1)) \\
& + \sum_{\langle i,j \rangle} (R_i(1 - R_j) + (1 - R_i)R_j) \\
& \times \left((1 - F_{ij}^T \bar{F}_{ij}) \log\left(\frac{1-\epsilon_1}{2}\right) \right) \\
& + \sum_{\langle i,j \rangle} \sum_{k=-1}^1 F_{ijk} \sum_{l=1}^L \log \mathcal{N}(B_{ij}^l; \mu_k, \sigma_k^2) \\
& + \sum_{\langle i,j \rangle} \sum_{k=-1}^1 \bar{F}_{ijk} \sum_{m=1}^M \log \mathcal{N}(\bar{B}_{ij}^m; \mu_k, \sigma_k^2). \quad (14)
\end{aligned}$$

E-Step: For a fixed setting of model parameters $\hat{\Theta}$, the free energy in (13) can be expanded as follows:

$$\begin{aligned}
\mathcal{FE} & = - \sum_R Q^r(R) \sum_{F, \bar{F}} \log P(R, F, \bar{F}, Y; \hat{\Theta}) \\
& + \sum_R Q^r(R) \log Q^r(R) - \mathcal{H}(Q^c) \\
& = - \sum_R Q^r(R) \cdot E_{\hat{Q}^c} \left[\log P(R, F, \bar{F}, Y; \hat{\Theta}) \right] \\
& + \sum_R Q^r(R) \log Q^r(R) - \mathcal{H}(Q^c). \quad (15)
\end{aligned}$$

We can define the (normalized) probability distribution $\tilde{P}(R; \hat{\Theta})$ as

$$\tilde{P}(R; \hat{\Theta}) \propto \exp \left\{ E_{\hat{Q}^c} \left[\log P(R, F, \bar{F}, Y; \hat{\Theta}) \right] \right\}. \quad (16)$$

By substituting (16) into (15), it is trivial to show that

$$\mathcal{FE} = \text{KL} \left(Q^r(R) \parallel \tilde{P}(R; \hat{\Theta}) \right) + \text{const.} \quad (17)$$

where $\text{KL}(p \parallel q)$ is the Kullback–Leibler (KL) divergence from the distribution $p(\cdot)$ to the distribution $q(\cdot)$, and the additional constants do not depend on R .

Using a similar expansion, we can also show that

$$\mathcal{FE} = \text{KL} \left(Q^c(F, \bar{F}) \parallel \tilde{P}(F, \bar{F}; \hat{\Theta}) \right) + \text{const.} \quad (18)$$

where $\tilde{P}(F, \bar{F}; \hat{\Theta}) \propto \exp \{ E_{\hat{Q}^r} [\log P(R, F, \bar{F}, Y; \hat{\Theta})] \}$ and the additional constants do not depend on F, \bar{F} .

Since the KL divergence is nonnegative, (17) and (18) give us the following fixed-point equations for the variational posterior $\hat{Q}(\cdot)$:

$$\hat{Q}^c(F, \bar{F}) = \tilde{P}(F, \bar{F}; \hat{\Theta}) \propto \exp \left\{ E_{\hat{Q}^r} \left[\log P(R, F, \bar{F}, Y; \hat{\Theta}) \right] \right\} \quad (19)$$

$$\hat{Q}^r(R) = \tilde{P}(R; \hat{\Theta}) \propto \exp \left\{ E_{\hat{Q}^c} \left[\log P(R, F, \bar{F}, Y; \hat{\Theta}) \right] \right\}. \quad (20)$$

We alternate between updating $\hat{Q}^r(R)$ and updating $\hat{Q}^c(F, \bar{F})$, according to the above expressions, until convergence. Specifically, we employ Gibbs sampling to obtain samples $\mathcal{S} = \{R^s\}$ from $\hat{Q}^r(R)$. Based on the joint log-likelihood in (14), the right-hand side of (19) can be expressed in terms of the first- and second-order statistics of R

$$\hat{q}_{ij}^{00} \triangleq E_{\hat{Q}^r} [(1 - R_i)(1 - R_j)] \quad (21)$$

$$\hat{q}_{ij}^{11} \triangleq E_{\hat{Q}^r} [R_i R_j] \quad (22)$$

$$\hat{q}_{ij}^{10} \triangleq E_{\hat{Q}^r} [R_i(1 - R_j) + (1 - R_i)R_j]. \quad (23)$$

We approximate these quantities using averages of R_i and $R_i R_j$ over the elements of \mathcal{S} .

To update $\hat{Q}^c(\cdot)$, we evaluate the right-hand side of (19) for each configuration $F_{ij} = k, \bar{F}_{ij} = k' (k, k' \in \{-1, 0, 1\})$ and normalize over all nine combinations of k, k' to obtain a valid probability distribution.

According to the joint log-likelihood in (14), the right-hand side of (20) is given in terms of $E_{\hat{Q}^c} [F_{ij}^T \bar{F}_{ij}]$. Since F_{ij} and \bar{F}_{ij} are indicator variables, this quantity can be evaluated as

$$\hat{p}_{ij} \triangleq E_{\hat{Q}^c} [F_{ij}^T \bar{F}_{ij}] = \sum_{k=-1}^1 \hat{Q}_{ij}^c(F_{ij} = k, \bar{F}_{ij} = k). \quad (24)$$

The model parameter estimates $\hat{\Theta}$ in the following section rely on marginal probabilities of $\hat{Q}^c(F, \bar{F})$. We compute these quantities after convergence of the variational posterior distribution $\hat{Q}(\cdot)$

$$\hat{s}_{ijk} = \hat{P}(F_{ij} = k \mid Y; \Theta) = \sum_{\bar{F}_{ij}} \hat{Q}_{ij}^c(F_{ijk} = 1, \bar{F}_{ij}) \quad (25)$$

$$\hat{u}_{ijk} = \hat{P}(\bar{F}_{ij} = k \mid Y; \Theta) = \sum_{F_{ij}} \hat{Q}_{ij}^c(F_{ij}, \bar{F}_{ijk} = 1). \quad (26)$$

M-Step: We fix the posterior probability estimates $\hat{Q}(R, F, \bar{F})$ and update the model parameter estimates $\hat{\Theta}$ by differentiating (13) with respect to each element of Θ and setting the gradient equal to zero.

The update for π^r involves averaging the proportion of diseased regions across Gibbs samples

$$\hat{\pi}^r = \frac{1}{NS} \sum_{i=1}^N \sum_{s=1}^S R_i^s. \quad (27)$$

The multinomial prior reduces to an average over the marginal posterior distribution

$$\hat{\pi}_k^f = \frac{1}{C} \sum_{\langle i,j \rangle} \hat{s}_{ijk} \quad (28)$$

where C is the total number of pairwise connections.

The fMRI likelihood parameter estimates are computed as weighted statistics of the data

$$\mu_k = \begin{cases} \frac{\sum_{\langle i,j \rangle} [\hat{s}_{ijk} \sum_l B_{ij}^l + \hat{u}_{ijk} \sum_m \bar{B}_{ij}^m]}{\sum_{\langle i,j \rangle} [L \cdot \hat{s}_{ijk} + M \cdot \hat{u}_{ijk}]}, & k = \pm 1 \\ 0, & k = 0 \end{cases} \quad (29)$$

$$\sigma_k^2 = \frac{\sum_{\langle i,j \rangle} [\hat{s}_{ijk} \sum_l (B_{ij}^l - \hat{\mu}_k)^2 + \hat{u}_{ijk} \sum_m (\bar{B}_{ij}^m - \hat{\mu}_k)^2]}{\sum_{\langle i,j \rangle} [L \cdot \hat{s}_{ijk} + M \cdot \hat{u}_{ijk}]} \quad (30)$$

where we have fixed $\mu_0 = 0$ for the component that represents zero functional synchrony to center the parameter estimates and regularize the model.

We use Newton's method to jointly update $\hat{\eta}$ and $\hat{\epsilon}$. The details of this step are provided in the appendix.

B. Joint Model

The variational EM algorithm can be easily extended to incorporate anatomical connectivity. Below, we let $Y = \{B, \bar{B}, D, \bar{D}\}$ denote the observed fMRI and DWI measurements, respectively, and we let $\Theta = \{\pi, \eta, \epsilon, \mu, \sigma^2, \rho, \chi, \xi^2\}$ be the set of model parameters. Since A_{ij} is binary and F_{ij} and \bar{F}_{ij} are tri-state, the variational posterior is

$$Q(R, A, F, \bar{F}) = Q^r(R) \cdot Q^c(A, F, \bar{F}) \\ = Q^r(R) \prod_{\langle i,j \rangle} Q_{ij}^c(A_{ij}, F_{ij}, \bar{F}_{ij}) \quad (31)$$

where $Q^r(\cdot)$ is a distribution over the length- N binary vector R and $Q_{ij}^c(\cdot)$ is an 18-state multinomial distribution corresponding to all configurations of latent anatomical and functional connectivity.

E-Step: For a fixed setting of model parameters $\hat{\Theta}$, we alternate updates for $\hat{Q}^r(R)$ and $\hat{Q}^c(A, F, \bar{F})$ according to the following expressions:

$$\hat{Q}^r(R) \propto \exp \left\{ E_{\hat{Q}^c} [\log P(R, A, F, \bar{F}, Y; \Theta)] \right\} \quad (32)$$

$$\hat{Q}^c(A, F, \bar{F}) \propto \exp \left\{ E_{\hat{Q}^r} [\log P(R, A, F, \bar{F}, Y; \Theta)] \right\}. \quad (33)$$

Once again, we use Gibbs sampling to obtain samples $\mathcal{S} = \{R^s\}$ from (32) and then evaluate $\hat{Q}^c(A, F, \bar{F})$ using averages of R_i and $R_i R_j$ over the elements of \mathcal{S} . We update $\hat{Q}^c(\cdot)$ by evaluating the right-hand side of (33) for all 18 configurations of $\{A_{ij}, F_{ij}, \bar{F}_{ij}\}$ and normalizing. $\hat{Q}^r(R)$ is given in terms of $E_{\hat{Q}^c}[A_{ij} \cdot F_{ij}^T \bar{F}_{ij}]$ and $E_{\hat{Q}^c}[A_{ij}(1 - F_{ij}^T \bar{F}_{ij})]$, which are evaluated similar to (24).

M-Step: Similar to the construction for the functional variables, we define the marginal posterior probability for latent anatomical connectivity

$$\hat{\alpha}_{ij} = \sum_{F_{ij}, \bar{F}_{ij}} Q_{ij}^c(A_{ij} = 1, F_{ij}, \bar{F}_{ij}).$$

Additionally, we let L_{ij}^0 be the number of control subjects for whom $D_{ij}^l = 0$ and M_{ij}^0 be the number of schizophrenia patients for whom $\bar{D}_{ij}^m = 0$.

The updates for $\hat{\pi}^r$ and the fMRI likelihood parameters remain unchanged. The prior estimate for $\hat{\pi}^a$ is an intuitive average of marginal probabilities

$$\hat{\pi}^a = \frac{1}{C} \sum_{\langle i,j \rangle} \hat{\alpha}_{ij} \quad (34)$$

where C is the total number of pairwise connections.

The prior π^f interacts with A, F and \bar{F} due to (3) and (8). Minimizing the free energy with respect to π_k^f results in the following update equation:

$$\hat{\pi}_k^f = \frac{\sum_{\langle i,j \rangle} [\hat{s}_{ijk} + \sum_{F_{ij}} \hat{Q}_{ij}^c(A_{ij} = 0, F_{ij}, \bar{F}_{ij} = k)]}{C + \sum_{\langle i,j \rangle} \sum_{F_{ij}} \hat{Q}_{ij}^c(A_{ij} = 0, F_{ij}, \bar{F}_{ij} = k)}. \quad (35)$$

The probability $\hat{\rho}_1$ is the empirical likelihood of not finding a white matter tract between two regions given an underlying anatomical connection

$$\hat{\rho}_1 = \frac{\sum_{\langle i,j \rangle} \hat{\alpha}_{ij} (L_{ij}^0 + M_{ij}^0)}{\sum_{\langle i,j \rangle} \hat{\alpha}_{ij} (L + M)}. \quad (36)$$

The Gaussian likelihood parameters for the DWI measurements are given by the weighted empirical mean and empirical variance over all *nonzero* values

$$\hat{\chi}_1 = \frac{\sum_{\langle i,j \rangle} \hat{\alpha}_{ij} (\sum_{l \in \mathcal{I}_{ij}^l} D_{ij}^l + \sum_{m \in \bar{\mathcal{I}}_{ij}^m} \bar{D}_{ij}^m)}{\sum_{\langle i,j \rangle} \hat{\alpha}_{ij} (L - L_{ij}^0 + M - M_{ij}^0)} \quad (37)$$

$$\hat{\xi}_1^2 = \frac{\sum_{\langle i,j \rangle} \hat{\alpha}_{ij} (\sum_{l \in \mathcal{I}_{ij}^l} (D_{ij}^l - \hat{\chi}_1)^2 + \sum_{m \in \bar{\mathcal{I}}_{ij}^m} (\bar{D}_{ij}^m - \hat{\chi}_1)^2)}{\sum_{\langle i,j \rangle} \hat{\alpha}_{ij} (L - L_{ij}^0 + M - M_{ij}^0)} \quad (38)$$

where \mathcal{I}_{ij}^l and $\bar{\mathcal{I}}_{ij}^m$ denote the subset of control subjects and patients, respectively, that exhibit white matter tracts between regions i and j . The parameter updates for $\{\rho_0, \chi_0, \xi_0^2\}$ are trivially obtained from these expressions by replacing $\hat{\alpha}_{ij}$ with $(1 - \hat{\alpha}_{ij})$.

Similar to the previous algorithm, we update $\hat{\eta}$ and $\hat{\epsilon}$ using a Newton's method iteration. We omit the expressions for the first and second derivatives, as they do not provide additional insight into the algorithm.

C. Implementation Details

In this section we describe the optimization choices in our implementation of the variational EM algorithm.

1) *Initialization:* Like many hill-climbing methods, the quality of our results depends on proper initialization. For the variational algorithm, it suffices to initialize the model parameters $\Theta = \{\pi, \eta, \epsilon, \mu, \sigma^2, \rho, \chi, \xi^2\}$ and the marginal posterior statistics for R , $E_{\hat{Q}^r}[R_i]$ and $E_{\hat{Q}^r}[R_i R_j]$. The algorithm proceeds by computing the joint posterior distribution $\hat{Q}^c(\cdot)$ in

the E-step and updating $\hat{Q}^r(R)$ and $\hat{Q}^c(\cdot)$ until convergence. We then estimate the model parameters (M-step) and iterate.

We initialize the prior parameters π^a, π^f , the probability of not detecting white matter fibers ρ , and the Gaussian variances σ^2, ξ^2 using statistics of the data. We also set the initial value of the noise parameter $\epsilon = 0.01$, which encourages consistency between the region labels and the observed connectivity data. Perturbations in these values do not seem to impact our final solution. We sample the initial values for the Bernoulli region prior π^r and for the graph parameter η from a uniform distribution over the interval $[0.2, 0.5]$. Larger values of π^r and η encourage the algorithm to select more regions as foci during the first iteration.

The initial values of the Gaussian means $\{\mu, \chi\}$ largely determine the initial latent connectivity assignments, and hence, have the biggest influence on the final solution. Empirically, our model finds sparse solutions for the region label vector R . This means that if the initial connectivity data is too similar between the populations, then the algorithm will converge to a suboptimal solution (with respect to the free energy) in which none of the regions are diseased. Therefore, we initialize $\{\mu, \chi\}$ to exaggerate the relevant functional connectivity differences. In particular, we sample each of these values from a uniform distribution over specific interval such that: 1) the initial distribution of latent functional connectivity is roughly uniform, and 2) the initial graph of latent anatomical connectivity is fairly dense. These choices improve our chances of finding the global optimum. We emphasize that our initialization is still fairly naïve and that we do not place strong *a priori* assumptions on the model. Rather, our initialization provides enough flexibility for the algorithm to efficiently traverse the parameter space.

Finally, we initialize the posterior statistics by computing the mean fMRI correlation across subjects and clustering these values for each connection. We select regions with the highest number of connections with different cluster assignments between the two populations as the set of disease foci. We sample $E_{\hat{Q}^r}[R_i]$ from a uniform distribution over the interval $[0.8, 1]$ for each selected focus region and sample $E_{\hat{Q}^r}[R_i] \in [0, 0.2]$ for nonfoci. The pairwise statistics are computed as $E_{\hat{Q}^r}[R_i R_j] = E_{\hat{Q}^r}[R_i] \cdot E_{\hat{Q}^r}[R_j]$.

Empirically, we find that the final region posterior distribution is fairly stable within the above parameter ranges. We run the algorithm five times for the functional model and ten times for the joint model to sample the solution space; we then select the solution with the lowest free energy.

2) *Gibbs Sampling*: In the E-step, we sample the region indicator vector R from the posterior distribution $\hat{Q}^r(R)$. Specifically, for each region i , we sample the value $R_i \in \{0, 1\}$ while fixing the other region assignments. The regions are updated in random order. To speed up computation, we run Gibbs sampling simultaneously on four processors and combine the resulting samples. In each case, the first 500 iterations are used for burn-in; we collect 50 samples spaced 100 iterations apart. Here, one iteration refers to updating all elements of the vector R .

3) *Convergence and Runtime*: Convergence of our algorithms was based on a relative change in free energy of less than 10^{-4} between consecutive iterations. On average, both algorithms converge in less than 10 iterations (E-step/M-step

updates). All simulations were performed using MATLAB on a modern quad processor workstation.

The runtime of each variational EM iteration depends linearly on the number of subjects. As shown in Section IV, the nonrandom parameter updates are based on simple statistics of the data, summed across subjects. Similarly, the updates for the latent posterior distributions in (19) and (20) rely on the data log-likelihood, once again summed across subjects. Fortunately, these operations are inexpensive to compute, even for large populations.

The complexity of our algorithm is sensitive to the number of regions. First, the number of connections scales quadratically with the number of regions, thus increasing the number of terms $Q_{ij}^c(F_{ij}, \bar{F}_{ij})$ in the variational posterior distribution. Second, the runtime of each Gibbs sampling iteration increases linearly with the number of regions. Given that we require several thousand iterations to robustly estimate the region posterior $Q^r(R)$, increasing the number of regions will slow down the algorithm substantially. For reference, our current implementation requires 30 min for the functional model and 15 min for the joint model. Although the runtime can be improved by using more parallel computation, additional approximations are needed for more than 100–200 regions.

V. MODEL EVALUATION

A. Identifying Disease Foci

The marginal posterior distribution $\hat{Q}^r(R_i = 1)$ informs us about the disease foci. Let \hat{q}_i denote the marginal probability that region i is diseased. We estimate this quantity by averaging across Gibbs samples \mathcal{S}

$$\hat{q}_i \triangleq \hat{Q}^r(R_i = 1) = \frac{1}{S} \sum_{s=1}^S R_i^s. \quad (39)$$

The joint distributions in our method are non-Gaussian due to multiplicative interactions between latent variables and the effects of unknown nonrandom parameters. Therefore, we evaluate the significance of the resulting estimates through nonparametric permutation tests. To construct the null distribution for \hat{q}_i , we randomly permute the subject diagnoses (NC versus SZ) 1000 times. For each permutation, we fit the model and compute the statistic in (39). The significance of each region is equal to the proportion of permutations that yield a larger value of \hat{q}_i than is obtained under the true labeling.

B. Graph of Abnormal Connectivity

The graph of connectivity differences T in Section III provides insight into the behavior of individual connections. Although we marginalize this random variable prior to inference, we can approximate T based on the *max a posteriori* (MAP) estimate of each R_i and the parameter estimates $\hat{\Theta}$.

Given R , our models decouple by pairwise connection, so we can assign each T_{ij} independently. Recall that $T_{ij} = 0$ indicates a healthy edge and $T_{ij} = 1$ denotes a diseased connection. Based on our construction in Section III, many of the values T_{ij}

TABLE II
LIKELIHOOD PARAMETERIZATIONS USED TO GENERATE SYNTHETIC DATA

	μ_{-1}	μ_0	μ_1	σ_{-1}^2	σ_0^2	σ_1^2	ρ_0	ρ_1	χ_0	χ_1	ξ_0^2	ξ_1^2
Good Data	-0.35	0	0.35	0.050	0.050	0.050	0.70	0.10	0.45	0.35	0.0050	0.0050
Noisy Data	-0.18	0	0.36	0.050	0.058	0.072	0.67	0.10	0.41	0.34	0.0050	0.0026

TABLE III
PARAMETERS OF THE FUNCTIONAL MODEL IN FIG. 1(B) AND THE JOINT MODEL IN FIG. 2(B) ESTIMATED FROM THE CLINICAL DATA

	π^r	π_{-1}^f	π_0^f	π_1^f	η	ϵ	μ_{-1}	μ_0	μ_1	σ_{-1}^2	σ_0^2	σ_1^2
Functional	0.039	0.33	0.46	0.21	0.16	0.030	-0.18	0	0.36	0.050	0.059	0.073
Joint	0.040	0.36	0.45	0.19	0.17	0.025	-0.18	0	0.36	0.050	0.058	0.073

	π^a	ρ_0	ρ_1	χ_0	χ_1	ξ_0^2	ξ_1^2
Functional	—	—	—	—	—	—	—
Joint	0.34	0.67	0.10	0.41	0.34	0.0050	0.0026

are deterministic. For example, (2) of the functional model implies that $T_{ij} = 0$ if $R_i = R_j = 0$ and $T_{ij} = 1$ if $R_i = R_j = 1$. For connections $\langle i, j \rangle$ such that $R_i \neq R_j$, we select the value $T_{ij} \in \{0, 1\}$ that satisfies

$$\hat{T}_{ij} = \arg \max_T E_{\hat{Q}^c} \left[\log P(T | F_{ij}, \bar{F}_{ij}, \hat{R}^{\text{MAP}}; \hat{\Theta}) \right].$$

Equation (2) further specifies that if the region labels differ, the prior on T_{ij} is a Bernoulli distribution with parameter η . Additionally, if $T_{ij} = 0$ (the edge $\langle i, j \rangle$ is healthy), then the functional connectivity is the same in both populations with probability $1 - \epsilon$, and it differs with probability ϵ . Likewise, if $T_{ij} = 1$, then the functional connectivity differs between the populations with probability $1 - \epsilon$ and is the same with probability ϵ . After some algebraic manipulations we arrive at the decision rule for the functional model

$$\begin{aligned} \log(\hat{\eta}) + \hat{p}_{ij} \log(\hat{\epsilon}) + (1 - \hat{p}_{ij}) \log\left(\frac{1 - \hat{\epsilon}}{2}\right) \\ \hat{T}_{ij=1} \geq \log(1 - \hat{\eta}) + \hat{p}_{ij} \log(1 - \hat{\epsilon}) + (1 - \hat{p}_{ij}) \log\left(\frac{\hat{\epsilon}}{2}\right) \\ \hat{T}_{ij=0} \end{aligned} \quad (40)$$

where \hat{p}_{ij} is defined in (24) and the parameters $\hat{\eta}$, $\hat{\epsilon}$ are estimated via the fixed-point algorithm in the Appendix.

The joint decision rule is similarly derived by incorporating the anatomical constraints in (7) and (8).

C. Varying the Region Prior π^r

Although our framework enables us to estimate all unknown parameters, we further explore the solution space by specifying the expected number of diseased regions via the prior π^r . In particular, the evolution of disease foci across a range of prior π^r (in this work $\pi^r \in [0, 0.5]$) illustrates the stability of our model in explaining the data. Moreover, tuning π^r is an intuitive way to inject clinical knowledge into our framework and may be useful in certain applications. Fixing π^r does not affect the update equations in Section IV.

D. Model Robustness

We evaluate the robustness of our approach by fitting the models to random subsets of the data. Specifically, we withhold W subjects from each population. The value of W is varied

from $W = 1, \dots, 10$. This corresponds to leaving out between 5%–52% of the subjects. We resample the data 20 times for each value of W and consider the region posterior statistic \hat{q}_i , as given in (39), averaged across all runs.

VI. EXPERIMENTAL RESULTS—SYNTHETIC DATA

We first evaluate the robustness and sensitivity of our algorithms using synthetic data. Our primary focus is on the effect of the parameters η and ϵ on identifying the disease focus. We expect the performance to improve with increasing η and worsen with increasing ϵ . This is because higher values of η raise the number of functional connectivity differences associated with each disease foci. Consequently, the algorithms can better detect these regions. In contrast, larger values of ϵ increase the number of functional differences involving healthy regions, which negatively impacts the final quality of estimation.

We sweep the parameter values across the ranges $\eta \in [0.1, 0.5]$ and $\epsilon \in [0, 0.05]$: for each (η, ϵ) pair, we generate the latent connectivity templates and observed data according to the generative models in Fig. 1(b) and Fig. 2(b). We fit the data using the algorithms presented in Section IV and compute the false-negative (Type I) and false-positive (Type II) errors based on the MAP estimates $\hat{R}_i \in \{0, 1\}$ for each region i .

We mimic the organization of our clinical dataset by specifying a template with 78 regions (39 per hemisphere) and with two disease foci in each hemisphere. Throughout this section, we fix the functional prior π^f to the value inferred from the clinical experiments. We also sample the latent anatomical connectivity A such that the intra- and inter-hemisphere statistics match those of our clinical data.

We consider two likelihood parameterizations for $\{\mu, \sigma^s, \rho, \chi, \xi^2\}$, as shown in Table II. The Good Data parameterization assumes a clear separation between the data distributions for different latent connectivity values. In this case, we can accurately infer the connectivity templates $\{A, F, \bar{F}\}$, which are then used for region assignments \hat{R} . The Noisy Data parameterization uses the ML parameter estimates $\hat{\Theta}$ from the clinical experiments to generate the observed synthetic measurements. As reported in Table III, there is a significant overlap in the ML data distributions. Hence, we observe the effects of noise on the estimated latent connectivity and region assignments.

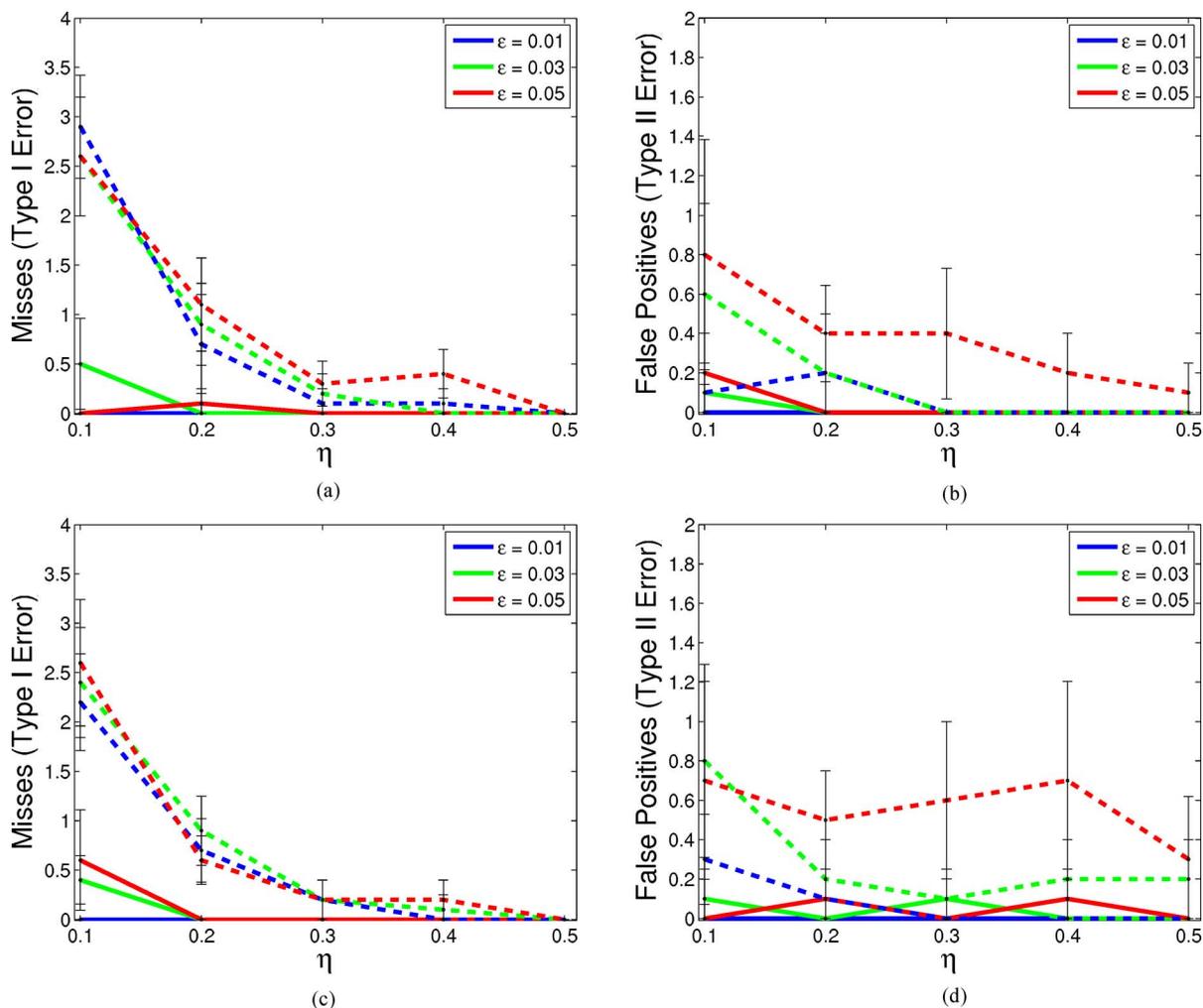


Fig. 3. Average number of mislabeled region assignments when sampling from the functional model. The solid lines correspond to fitting the functional model, and the dashed lines represent the joint model results. The error bars denote one standard deviation. Type I error corresponds to the number of disease foci that were missed by our algorithm. Similarly, Type II error denotes the number of healthy regions that were incorrectly identified as diseased. (a) *Good Data*, Missed disease foci. (b) *Good Data*, False positive region assignments. (c) *Noisy Data*, Missed disease foci. (d) *Noisy Data*, False positive region assignments.

A. Sampling From the Functional Model

Given the region labels R , we sample the graph structure T , the latent functional templates F, \bar{F} and the observed fMRI correlations $\{B, \bar{B}\}$ according to (2)–(5). In order to fit the joint model, we independently generate the latent anatomical connectivity A and the observed DWI measures $\{D, \bar{D}\}$ via (6) and (9), respectively. We resample the latent connectivity templates and observed data 10 times to collect error statistics.

Fig. 3 reports the errors in determining the region labels across 10 samples of the latent connectivity templates and corresponding observed data. Unsurprisingly, the functional model achieves uniformly lower Type I and Type II errors. This is because the functional model exploits all pairwise connectivity information when determining the region labels, whereas the joint model must rely on a random subset of connections, specified by A . Nonetheless, the detection accuracy of the joint model improves significantly for larger values of η . The parameter η controls the density of nonzero edges in the graph T . Hence, as η increases, we are more likely to observe functional connectivity differences along the randomly generated anatomical template A .

The parameter ϵ influences the rate of false-positive assignments, particularly for the joint model. Intuitively, higher values of ϵ produce a greater number of (spurious) functional connectivity differences involving healthy regions. Therefore, the algorithm is more likely to incorrectly label one of these regions as diseased.

Despite the large variability in Type II error in Fig. 3(b) and (d), on average fewer than two (of 74) healthy regions are labeled as disease foci. This behavior suggests an implicit regularization in our framework. Specifically, labeling a region as diseased permits the associated functional connections to differ between groups, which can lower the free energy. However, connections to all other foci are automatically diseased, which may come with a cost. Our algorithm balances these competing influences by identifying a sparse set of disease foci.

Finally, we observe that the error rates are similar for both the *Good Data* and the *Noisy Data* likelihood parameterizations. This indicates that errors in region assignments are primarily due to functional differences that are inconsistent with the underlying disease foci rather than to noisy data observations.

B. Sampling From the Joint Model

We now evaluate the model in a situation when the functional effects of a disease are restricted to direct anatomical pathways. Given the region labels R , we generate the control template F , the latent anatomical connectivity A and the graph structure T according to the model in Fig. 2(b). However, we modify the construction of the clinical template \bar{F} . Since the joint model does not impose any relationship between the values F_{ij} and \bar{F}_{ij} in the absence of an anatomical connection, the latent functional templates differ dramatically when $A_{ij} = 0$. The functional model assumes all connections are equally important. Consequently, it cannot detect the true disease foci amid the overwhelming number of unrelated connectivity differences. For this reason, we sample \bar{F} using (4), repeated below for convenience

$$P(\bar{F}_{ij} | F_{ij}, T_{ij}; \epsilon) = \begin{cases} (1 - \epsilon)^{F_{ij}^T \bar{F}_{ij}} \left(\frac{\epsilon}{2}\right)^{1 - F_{ij}^T \bar{F}_{ij}}, & T_{ij} = 0 \\ \epsilon^{F_{ij}^T \bar{F}_{ij}} \left(\frac{1 - \epsilon}{2}\right)^{1 - F_{ij}^T \bar{F}_{ij}}, & T_{ij} = 1. \end{cases}$$

Since $T_{ij} = 0$ if $A_{ij} = 0$, we omit the multinomial prior when there is no underlying anatomical connection. Instead, we encourage the latent functional connectivity templates to be the same in the control and clinical populations. Although not fully consistent with the joint model, the above equation enables us to fit the functional model with some degree of accuracy. The observed data $\{B, \bar{B}, D, \bar{D}\}$ is generated according to (5) and (9). We repeat the experiment 10 times to collect error statistics.

Fig. 4 illustrates the error in region assignments across 10 instantiations of the latent connectivity templates and observed data measures. Despite modifying the sampling procedure to accommodate the functional model, it exhibits significantly worse detection accuracy than the joint model for nearly all (η, ϵ) values. The performance reduction can be attributed to the anatomical constraint, which reduces the effective number of connections, and subsequently the number functional differences, associated with each region by 40%–60%. Since the functional model treats all connections equally, the reduced number of functional differences is insufficient to pinpoint the disease foci. In contrast, the joint model adjusts the number of connectivity differences associated with a given region by the number of anatomical connections. Hence, the algorithm can isolate the diseased regions based on fewer differences. Despite the poor detection performance, the functional model demonstrates lower Type II error. This suggests that it produces sparser estimates of the disease foci than the joint model.

We also observe similarities between our synthetic results in Figs. 3 and 4. As expected, the detection accuracy improves with increasing η , as it results in a greater number of functional differences associated with each diseased region. In addition, the Type II error variance is high, but on average, relatively few healthy regions are mislabeled. Finally, the error rates are similar for both likelihood parameterizations. Once again, this suggests that noise in the latent structure has a greater impact than the observation noise.

In summary, each model can robustly identify diseased regions if the data is sampled accordingly. In Fig. 3, the joint

model exhibits slightly worse detection accuracy than the functional model; however, Fig. 4 reports a considerable drop in performance of the functional model when applied to the joint data. Both models exhibit an intrinsic regularization and infer sparse sets of foci with few false positive assignments.

VII. EXPERIMENTAL RESULTS—CLINICAL DATA

A. Image Acquisition and Preprocessing

We demonstrate our model on a study of 19 male patients with chronic schizophrenia and 19 healthy male controls. The control participants were group matched to the patients on age, handedness, parental socioeconomic status, and an estimated pre-morbid IQ. For each subject, an anatomical scan (SPGR, TR = 7.4 s, TE = 3 ms, FOV = 26 cm², res = 1 mm³), a diffusion-weighted scan (EPI, TR = 17 s, TE = 78 ms, FOV = 24 cm, res = 1.66 × 1.66 × 1.7 mm, 51 gradient directions with $b = 900$ s/mm², eight baseline scans with $b = 0$ s/mm²), and a resting-state functional scan (EPI-BOLD, TR = 3 s, TE = 30 ms, FOV = 24 cm, res = 1.875 × 1.875 × 3 mm) were acquired using a 3T GE Echospeed system.

We segmented the structural images into 77 anatomical regions with Freesurfer [35]. The DWI data was corrected for eddy-current distortions using the FSL FLIRT algorithm [36]. A two-tensor tractography was used to estimate the white matter fibers [37]. We computed the DWI observation D_{ij}^l in subject l by averaging FA along all fibers that connect regions i and j . If no tracts were found, D_{ij}^l was set to zero.

We discarded the first five fMRI time points and performed motion correction by rigid body alignment and slice timing correction using FSL [36]. The data was spatially smoothed using a Gaussian filter, temporally low-pass filtered with 0.08 Hz cutoff, and motion corrected via linear regression. Finally, we removed global contributions to the time courses from the white matter, ventricles and the whole brain. We computed the fMRI measurement B_{ij}^l as the Pearson correlation coefficient between the mean time courses of regions i and j in subject l .

B. Significant Regions

Fig. 5 illustrates the detected disease foci ($\hat{q}_i > 0.5$) for the functional and joint models, respectively. We color each region according to $-\log(p\text{-value})$ such that red corresponds to low significance and yellow indicates high significance. Each method identified three disease foci, all of which are significant. The functional model implicated the left posterior cingulate ($\hat{q}_i = 1, p < 0.008$), the right posterior cingulate ($\hat{q}_i = 1, p < 0.017$) and the left transverse temporal gyrus (Heschl's gyrus) ($\hat{q}_i = 1, p < 0.021$). The joint mode implicates a different subset of regions, namely, the right posterior cingulate ($\hat{q}_i = 1, p < 0.004$), the right superior temporal gyrus ($\hat{q}_i = 1, p < 0.014$), and the left superior temporal gyrus ($\hat{q}_i = 1, p < 0.044$).

Both models identify significant foci in the default network and in the temporal lobes of the brain. Interestingly, we observe symmetry in region assignments across the hemispheres, as evident for the posterior cingulate (PCC) and the superior temporal

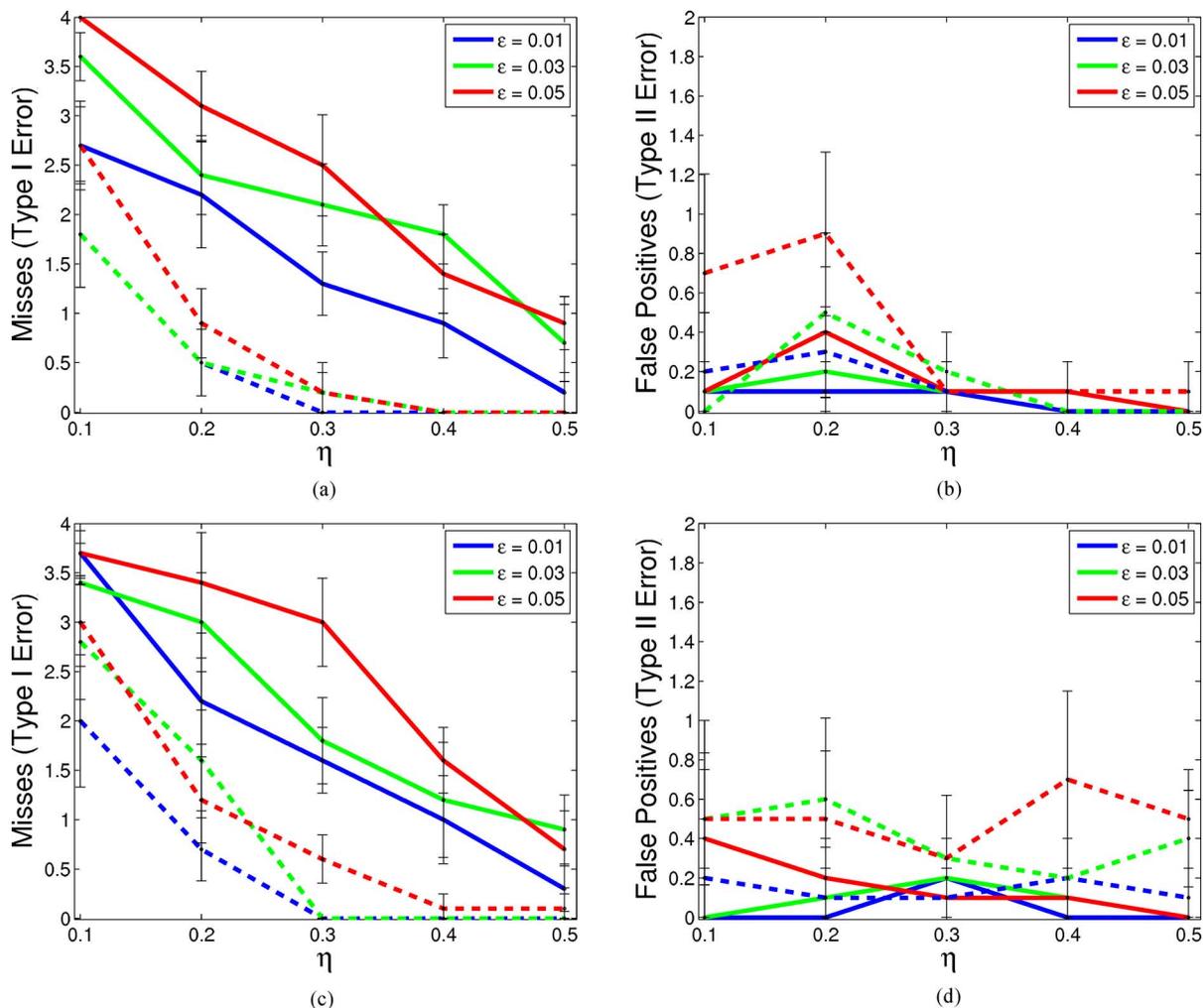


Fig. 4. Average number of mislabeled region assignments when sampling from the joint model. The solid lines are obtained when fitting the functional model, and the dashed lines correspond to the joint model results. The error bars denote one standard deviation. Type I error corresponds to the number of disease foci that were missed by our algorithm. Similarly, Type II error denotes the number of healthy regions that were incorrectly identified as diseased. (a) *Good Data*, Missed disease foci. (b) *Good Data*, False positive region assignments. (c) *Noisy Data*, Missed disease foci. (d) *Noisy Data*, False positive region assignments.

gyri (STG). This phenomenon may arise from the well-documented symmetry found in resting-state fMRI correlations [38]. We discuss the differences between the functional and joint results in Section VIII.

Table III reports the parameters inferred by our algorithms. We observe that the fMRI likelihood parameters are almost identical for both algorithms. This suggests that the difference between the two results in Fig. 5 is driven by the hierarchical structure from connections to region assignments rather than by the inference of latent functional connectivity from the data. Additionally, we observe consistency in parameter estimates across random subject relabelings in the permutation procedure (not shown). This implies that the main effects of permuting the subject diagnoses are reflected in the latent assignments rather than in the data likelihood.

C. Differences in Functional Connectivity

Fig. 6 displays the estimated graph of anomalous functional connectivity for each model. The functional model identifies abnormal connections distributed throughout the brain. For the joint model, 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 respective hemispheres. This difference in organization is explained by the constraint in Fig. 2(a) that functional anomalies should occur along anatomical pathways.

Both models detect an overall reduction in functional connectivity for schizophrenia patients. Of notable exception are connections to the frontal lobe. This phenomenon has been reported in prior studies of schizophrenia [15] and is believed to interfere with perception by misdirecting attentional resources.

D. Effect of Region Prior

Fig. 7 illustrates the results of varying the prior π^r of the region indicator vector R for the functional and joint models, respectively. We color each of the selected regions according to the smallest value of π^r such that the marginal posterior of the region i being a focus is greater than 0.2 (i.e., $\hat{q}_i > 0.2$). The yellow regions are always identified as foci, whereas the orange and red regions are only selected for larger prior values.

We observe that the functional model identifies a stable set of disease foci with an additional region for large values of π^r . In

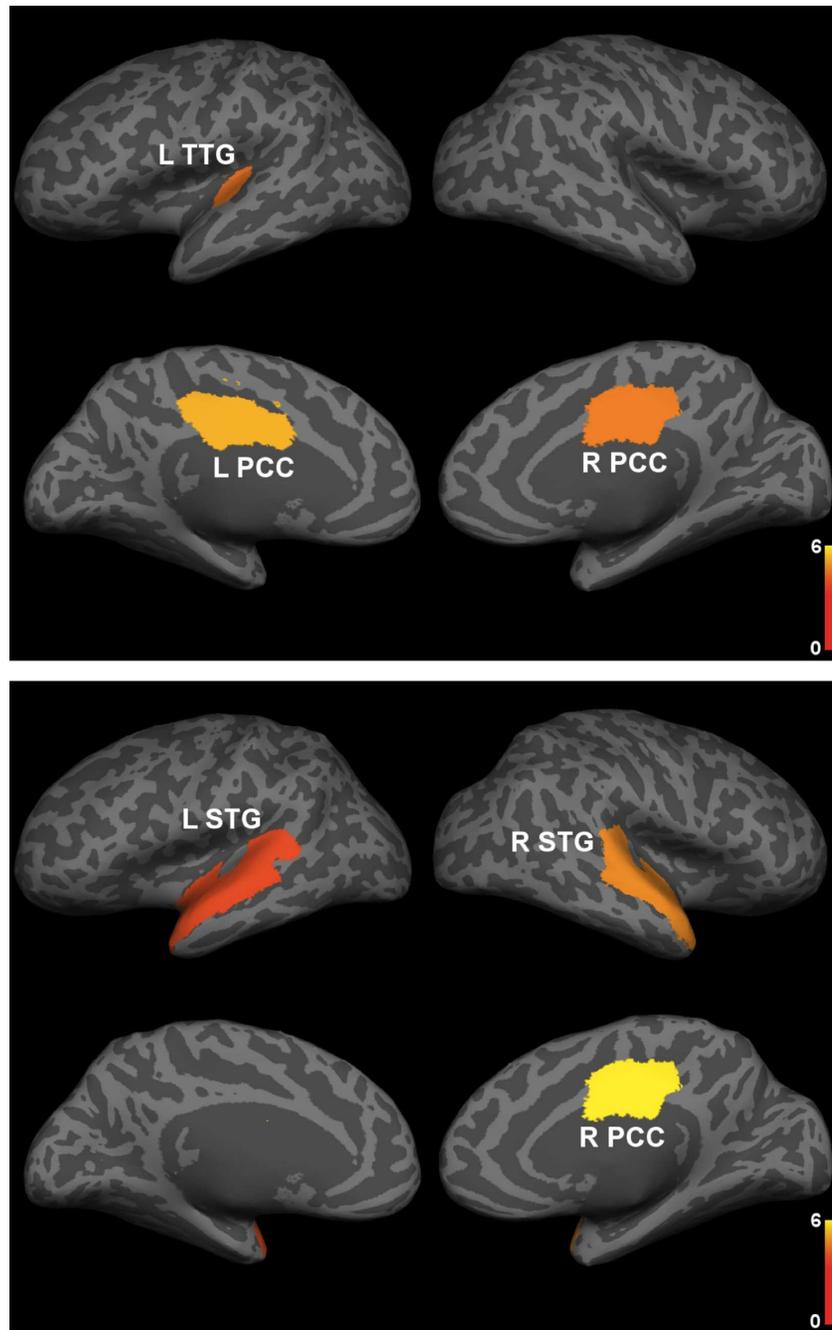


Fig. 5. Significant regions based on permutation tests ($q_i > 0.5$, uncorrected $p < 0.044$) identified by the functional model (top) and the joint model (bottom). The colorbar corresponds to the negative log p-value. We present the lateral and medial viewpoints for each hemisphere. The highlighted regions are the posterior cingulate (L PCC & R PCC), the transverse temporal gyrus (L TTG), and the superior temporal gyrus (L STG & R STG).

contrast, the sets of affected regions in the joint model form a nested substructure as π^r increases. It suggests an initial set of disease foci, identical to the significant regions in Fig. 5. For increasing π^r , the algorithm progressively includes regions that exhibit some functional abnormalities but are not as strongly implicated by the data. This extended set of regions is a superset of those identified by the functional model. We elaborate on the differences between the two results in Fig. 7 in the following section.

E. Model Robustness

Figs. 8 and 9 depict the average posterior probability of each region being a focus, for the functional model and the joint

model, respectively. We report only the regions for which the average probability is greater than 0.1; this allows us to focus on the most prominent patterns. The colorbar indicates the average posterior probability, such that yellow corresponds to the strongest disease foci and red denotes the weakest regions.

We observe that the regions with the highest posterior probabilities correspond to the significant disease foci in Fig. 5. Furthermore, these regions are consistently identified by our models when omitting $W = 1, \dots, 5$ subjects from each population. This is true for both the functional and the joint models. Intuitively, as we withhold more subjects (top row to bottom row in Figs. 8 and 9), the algorithm is less consistent across random re-

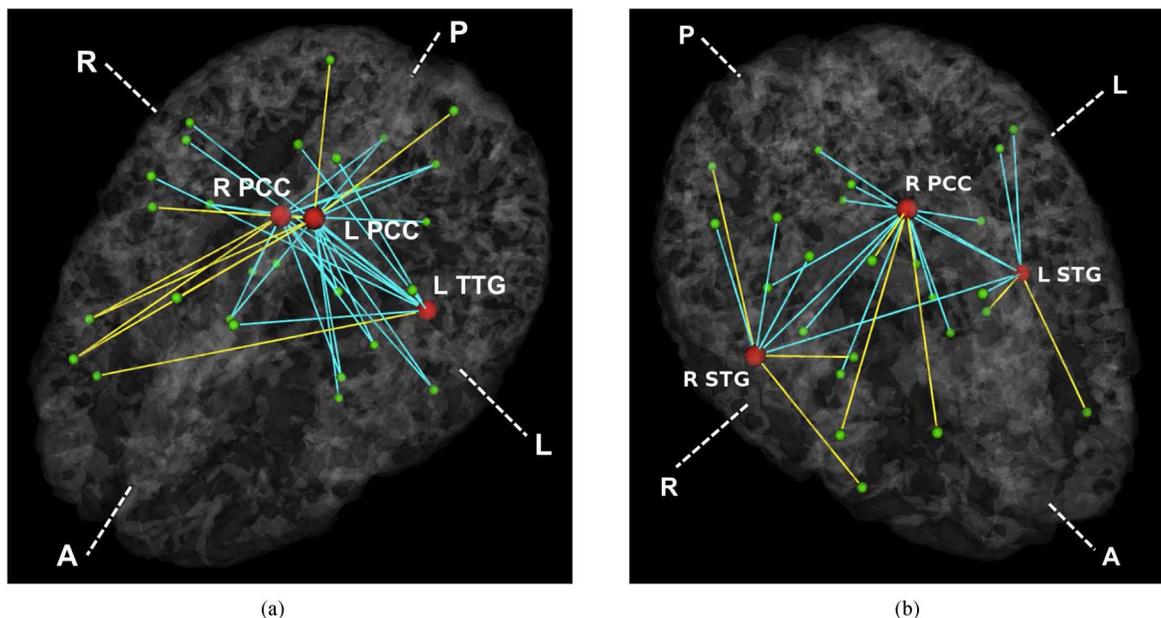


Fig. 6. Estimated graph of functional connectivity differences. The red nodes indicate the disease foci. Blue lines indicate reduced functional connectivity and yellow lines indicate increased functional connectivity in the schizophrenia population. (a) Functional model. (b) Joint model.

samplings of the data. Specifically, the average posterior probabilities of the significant foci decrease as W increases, i.e., these regions are less frequently selected by the algorithm. Although the model pinpoints additional regions that are not among the significant foci in Fig. 5, the average posterior probabilities of these new regions are low and nearly all of them are discovered when sweeping the region prior π^r in Fig. 7.

Finally, our holdout experiments for $W = 6, \dots, 10$ (not shown) are increasingly initialization-driven. Specifically, the algorithm rarely adds or removes regions from the initial set of disease foci. Consequently, as W increases, the solutions are less consistent across initializations. Such behavior may indicate that the training sets are insufficient to robustly estimate the region labels; hence, our variational algorithm is more likely to get caught in local minima.

To summarize, our results are consistent across reasonable perturbations in the dataset. This gives us confidence that the corresponding regions are relevant to schizophrenia and merit further exploration (e.g., follow-up anatomical or task fMRI studies). Clearly, our algorithm requires an adequate number of subjects to robustly estimate the population differences. Empirically, we find that the results are stable given 15 subjects in each group. Nonetheless, population size is an important consideration in future applications of this model.

VIII. DISCUSSION

We present a unified approach to infer regions associated with a disorder based on population differences in connectivity. Our first model operates on the complete graph of pairwise functional connections. Our second model incorporates anatomical constraints into this basic framework. We formulate a variational EM algorithm for maximum likelihood estimation of the model parameters. The algorithm simultaneously infers the posterior distribution over the region labels and the graph of abnormal functional connectivity.

Fig. 5 depicts the diseased regions implicated by each model. The main difference between the two results is that the functional model labels the transverse temporal gyrus as a disease focus, whereas the joint model pinpoints the superior temporal gyrus as relevant for schizophrenia. This discrepancy is partially explained by the size difference between these regions. As seen in Fig. 10, we identify significantly more neural connections involving the (large) STG than for the TTG. Hence, we are more likely to detect functional abnormalities associated with the STG that occur along direct anatomical pathways. This is reflected in Fig. 6(a). The majority of abnormal functional connections emanating from the TTG are inter-hemispheric, and hence, do not coincide with latent anatomical connections. Fig. 10 suggests that the quality of the joint model is largely dependent on the detection power of tractography. This underscores the need for advanced tractography algorithms that reliably identify long-range connections.

The TTG, or Heschl Gyrus, plays crucial role in auditory perception and language processing. Reduction in TTG volume, especially in the left hemisphere, has been associated with hallmark schizophrenia symptoms, such as auditory hallucinations, delusions and thought disorder [26]. Heschl's gyrus has also been linked to disease progression [39], suggesting its crucial role in schizophrenia pathophysiology.

The STG connects with heteromodal neocortical regions and temporolimbic areas. Electrophysiology and PET/fMRI studies in humans highlight the STG's role in the interpretation, production and self-monitoring of language. There is also evidence for structural and functional abnormalities of the STG in schizophrenia, which may be associated with formal thought disorder and auditory hallucinations [32], [40].

The PCC is one of the key structures in the default mode network. Recent functional schizophrenia studies [41] reported altered temporal frequency and spatial location of the default mode network. This suggests that the default network may be under- or overmodulated by key regions, including the anterior

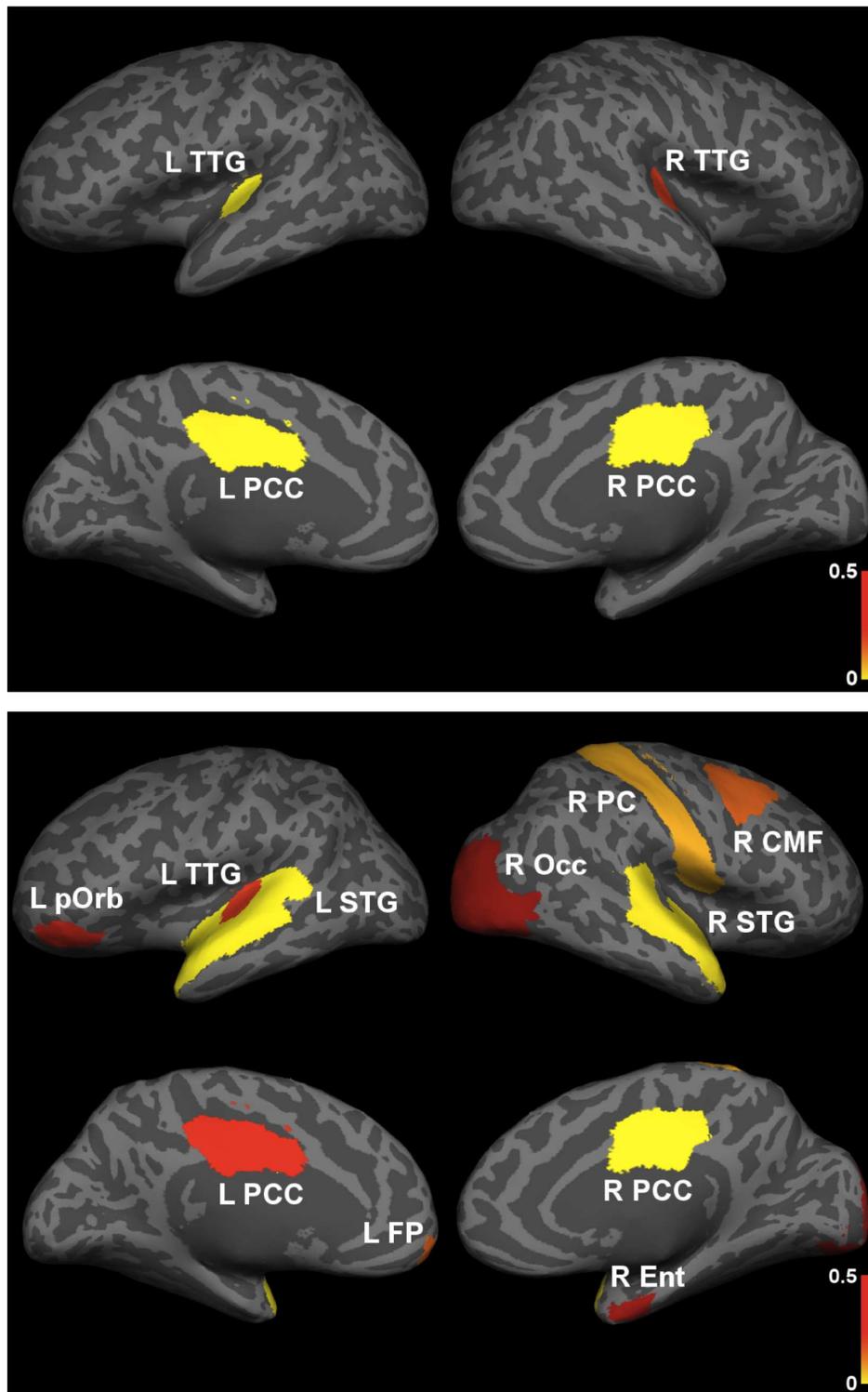


Fig. 7. Evolution of the disease foci when varying the region prior π^r for the functional model (top) and the joint model (bottom). The colorbar corresponds to the smallest value of π^r such that $q_i > 0.2$. The highlighted regions correspond to the posterior cingulate (L PCC & R PCC), the transverse temporal gyrus (L TTG & R TTG), 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).

and the posterior cingulate cortex. Our results confirm this hypothesis, further illustrating how such modulation can affect functional connectivity, leading to decreased connectivity between PCC and posterior parietal and temporal regions and increased connectivity between PCC and occipital and frontal lobes reported in Fig. 6. Reduced connectivity in the posterior

cingulate has been shown to correlate with both positive and negative symptoms of schizophrenia [30].

The role of anatomy is also evident in the graphs of aberrant functional connectivity depicted in Fig. 6. The functional results are distributed across the brain with little high-level organization. In contrast, the connections identified by the joint model

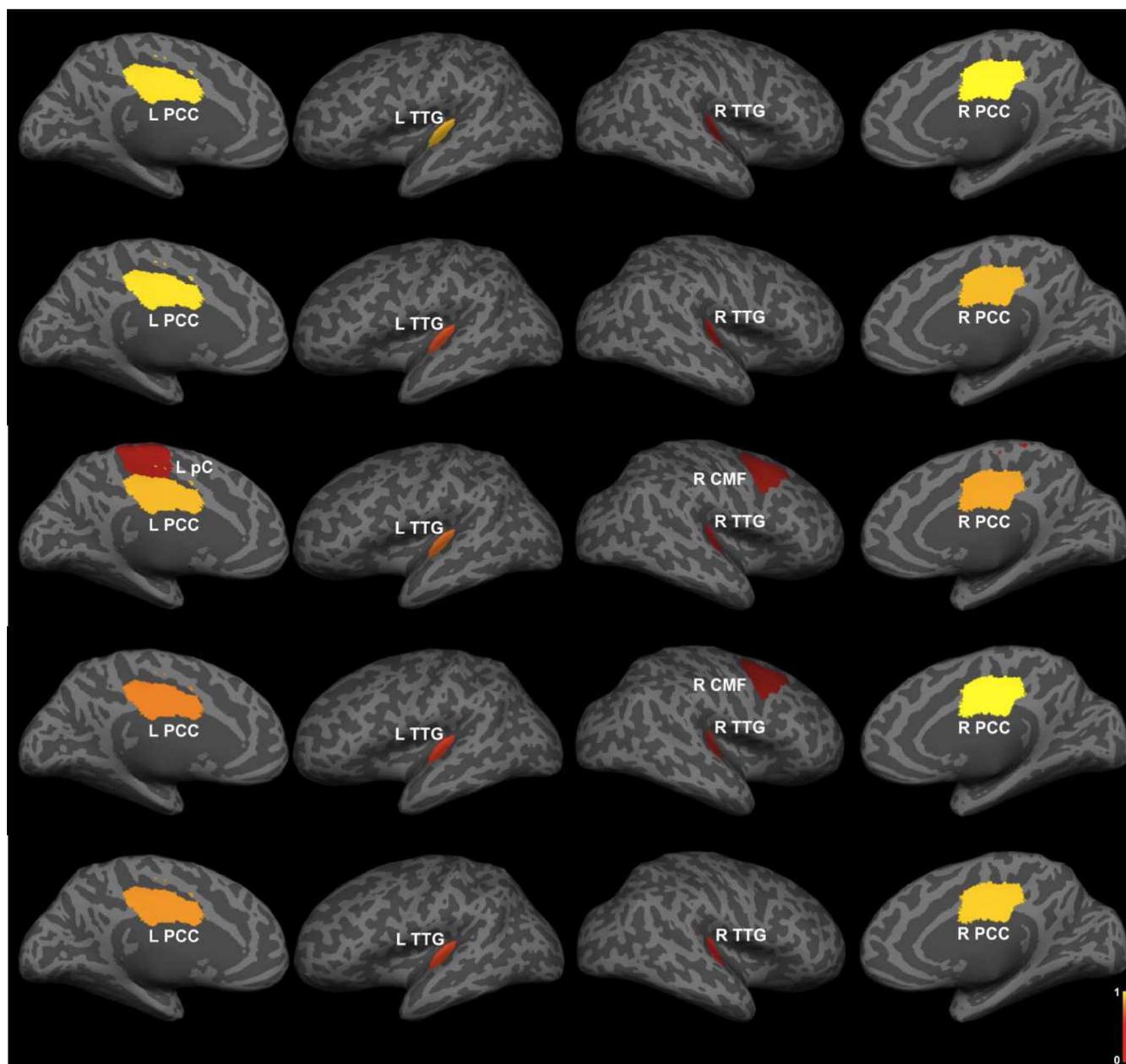


Fig. 8. Average marginal posterior probability across 20 resamplings of the data based on the functional model. The results are displayed for $W = 1$ subjects omitted from each population (top row) to $W = 5$ omitted subjects (bottom row). The colorbar indicates the average posterior probability. The regions correspond to the posterior cingulate (L PCC & R PCC), the transverse temporal gyrus (L TTG & R TTG), the precentral gyrus (L pC), and the caudal middle frontal gyrus (R CMF).

are largely separated by hemisphere and seem consistent with estimated white matter tracts. Despite their differences, both models detect a similar global pattern, which may reveal underlying neurological changes induced by schizophrenia. Specifically, we observe increased functional connectivity to the frontal lobe and reduced functional connectivity between the parietal/posterior cingulate region and the temporal lobe in the clinical population.

Increased connectivity between the default network and the medial frontal lobe, both at rest and during task, has been reported in schizophrenia [15], [42]. It is believed to interfere with perception of the external world by misdirecting attentional resources. Interestingly, decreased connectivity within

the default network has been described as well [18], [30]. The latter study reported decreased functional connectivity between the posterior cingulate gyrus and the hippocampus, which is consistent with our findings. The relationship between disruptions in functional connectivity and the integrity of the fornix has also been suggested. Along with prior findings, our results suggest an inverse relationship between connectivity in the temporal and frontal parts of the default network. Such “anticorrelations” have been previously described between the default and task-related networks. Two connections along white matter tracts in Fig. 6(b) have been implicated in schizophrenia [43]: the connection between the left and the right STG, provided by corpus callosum, and the connection between

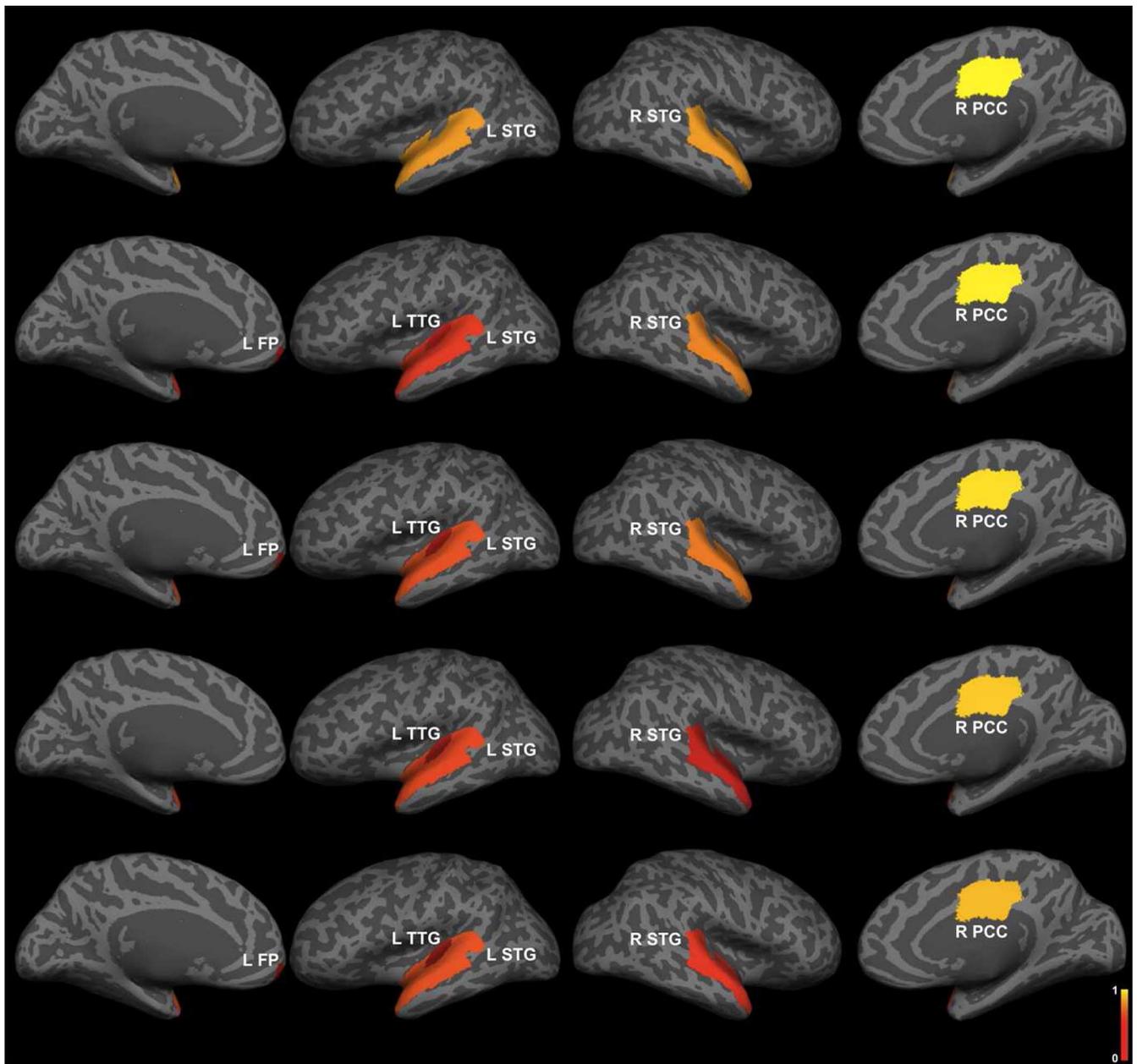


Fig. 9. Average marginal posterior probability across 20 resamplings of the data based on the joint model. The results are displayed for $W = 1$ subjects omitted from each population (top row) to $W = 5$ omitted subjects (bottom row). The colorbar indicates the average posterior probability. The regions denote the superior temporal gyrus (L STG & R STG), the posterior cingulate (R PCC), and the frontal pole (L FP).

posterior and anterior CG, provided by cingulum bundle. These two white matter tracts suggest a direct, causal relationship between anatomical and functional connectivity disruptions in schizophrenia.

Tuning the region prior parameter π^r enables us to explore the solution space. Once again, we observe differences between the two models. In particular, the functional results are consistent across a large range of prior values. In contrast, the joint model localizes nested subsets of disease foci as π^r increases. This suggests that the anatomical constraint increases the sensitivity of the joint model. Specifically, the effective number of connections to each region is reduced to the number of direct anatomical pathways. Hence, the joint model selects diseased regions based on fewer functional connectivity differ-

ences. Since many regions are weakly implicated by the data (i.e., associated with merely a few abnormal connections), biasing the algorithm through the region prior π^r causes them to be selected as foci.

The question remains: which model should we use? Presently, there is no standard technique to integrate anatomical and functional connectivity in order to pinpoint region impairments. Therefore, we argue that this is largely a philosophical issue based on a set of assumptions one makes about the brain. This work presents two different viewpoints. Clearly, if we assume that impairments of a neurological disorder affect functional synchrony between any two brain regions equally, then Fig. 3 suggests that we should fit the functional model. Similarly, if we assume that the most salient effects of a disorder occur

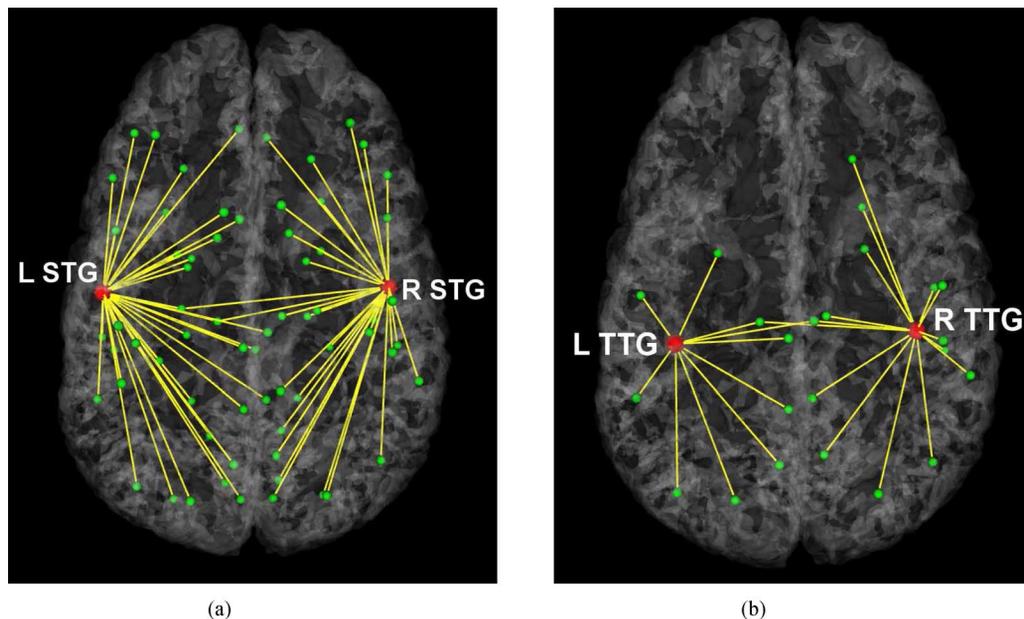


Fig. 10. Estimated neural pathways to the superior temporal gyri and the transverse temporal gyri. (a) Superior Temporal Gyri. (b) Transverse Temporal Gyri.

along direct anatomical connections, then Fig. 4 encourages us to choose the joint model. If we are unsure, then our synthetic results suggest that, on average, we are better off using the joint model. This is because the joint model achieves higher detection accuracy on data sampled from the functional model than *vice versa*. In the absence of latent anatomical connectivity, the joint model compares aggregate statistics of the templates F and \bar{F} . Therefore, data sampled according to the functional model in Fig. 1(b) is fairly consistent with the assumptions of the joint model. In contrast, the functional model cannot be fitted accurately to data sampled from the joint model in Fig. 2(b). A future extension of this work may consider all two-stage anatomical pathways as being relevant for disease localization. This can be achieved by incorporating the pairwise terms $A_{ik}A_{kj}$ into the distribution for the functional template \bar{F} of the clinical population in (8).

Encouragingly, both models in our current formulation localize similar disease foci. In fact, the joint model recovers both posterior cingulate regions as well as the left transverse temporal gyrus when we vary the region prior parameter π^r . The increased sensitivity of the joint model may prove beneficial, as it identifies a larger set of candidate regions (bottom image in Fig. 7). The effects of a complex disorder like schizophrenia are often subtle. Hence, the functional model, which only identifies the strongest functional differences, may not find all relevant disease foci.

The results may be influenced by our selection of regions. If the regions are too small, the variability in DWI tractography across subjects makes it difficult to infer the template anatomical connectivity and group-level parameters [24]. However, larger regions smooth out important functional connectivity information. In this work, we rely on Brodmann regions identified by Freesurfer [35]. Brodmann areas provide anatomically meaningful correspondences across subjects that roughly correspond to functional divisions within the brain. Moreover,

these regions are large enough to ensure stable tractography results. We emphasize that our framework applies readily to any set of ROIs that are defined consistently across subjects.

Finally, our model is designed to capture population differences to better understand the connectivity patterns induced by a disorder. The insights gained from our framework can subsequently be integrated with other types of data in order to build a comprehensive picture of a given neurological condition. Furthermore, once we learn the model parameters, we can compute the likelihood of a new subject belonging to each group. This score can later be correlated with behavioral and cognitive measures for patient-specific analyses.

We recognize the limitations of our generative models, especially those related to simplicity. For example, our joint model considers only direct anatomical connections and places a binary constraint on the graph of functional aberrations; our functional model ignores all anatomical information. Furthermore, we model latent connectivity via discrete random variables, which may marginalize subtle variations between groups. Finally, we assume a single set of disease foci that share mutually abnormal connectivity. In reality, neurological disorders can arise from several impairments in the brain, and the relationship between these diseased regions is unknown. One can even imagine a collection of independent disease clusters that do not interact directly.

These choices are deliberate on our part. Despite advancements in the field, the effects of schizophrenia on brain connectivity are neither well understood nor well characterized. In this work we formulate a simple relationship between region assignments and latent connectivity. Furthermore, given the potentially large amounts of inter-subject variability and external noise, we intentionally reduce the number of model parameters to avoid over-fitting. These limitations provide ample opportunities for future exploration.

IX. CONCLUSION

We propose a novel probabilistic framework that integrates population differences in connectivity to isolate foci of a neurological disorder. We present two variations of the model. The first considers functional connectivity only, as inferred from resting-state fMRI data. The second uses anatomical connectivity information from DWI tractography to constrain the functional effects. We demonstrate that our method identifies a stable set of schizophrenia hubs in the default network and in the temporal area of the brain. Prior clinical studies have linked these regions to the effects of schizophrenia. 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 isolate region effects.

APPENDIX

This appendix derives the Newton's method update for η, ϵ based on the functional model. The parameters η, ϵ are tied through (4); the only term of the free energy objective that depends on either parameter is

$$\begin{aligned}
& E_{\hat{Q}} \left[\log P(\bar{F} | F, R; \hat{\Theta}) \right] \\
&= \sum_{\langle i,j \rangle} \hat{q}_{ij}^{00} \left(\hat{p}_{ij} \log(1 - \epsilon) + (1 - \hat{p}_{ij}) \log\left(\frac{\epsilon}{2}\right) \right) \\
&+ \sum_{\langle i,j \rangle} \hat{q}_{ij}^{11} \left(\hat{p}_{ij} \log(\epsilon) + (1 - \hat{p}_{ij}) \log\left(\frac{1 - \epsilon}{2}\right) \right) \\
&+ \sum_{\langle i,j \rangle} \hat{q}_{ij}^{10} \left(\hat{p}_{ij} \log(\epsilon_1) + (1 - \hat{p}_{ij}) \log\left(\frac{1 - \epsilon_1}{2}\right) \right)
\end{aligned} \tag{41}$$

where we have substituted the definitions from (21)–(24) into the expression. The joint update for η, ϵ uses the following fixed point iteration:

$$\begin{bmatrix} \hat{\epsilon}_{n+1} \\ \hat{\eta}_{n+1} \end{bmatrix} = \begin{bmatrix} \hat{\epsilon}_n \\ \hat{\eta}_n \end{bmatrix} - \begin{bmatrix} \frac{\partial^2 \mathcal{F}\mathcal{E}}{\partial \epsilon^2} & \frac{\partial^2 \mathcal{F}\mathcal{E}}{\partial \epsilon \partial \eta} \\ \frac{\partial^2 \mathcal{F}\mathcal{E}}{\partial \epsilon \partial \eta} & \frac{\partial^2 \mathcal{F}\mathcal{E}}{\partial \eta^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial \mathcal{F}\mathcal{E}}{\partial \epsilon} \\ \frac{\partial \mathcal{F}\mathcal{E}}{\partial \eta} \end{bmatrix} \Bigg|_{(\hat{\epsilon}_n, \hat{\eta}_n)}$$

where the first and second derivatives of (41) with respect to η, ϵ are given by

$$\begin{aligned}
\frac{\partial \mathcal{F}\mathcal{E}}{\partial \epsilon} &= \frac{1}{\epsilon} \sum_{\langle i,j \rangle} [\hat{q}_{ij}^{11} \cdot \hat{p}_{ij} + \hat{q}_{ij}^{00}(1 - \hat{p}_{ij})] \\
&- \frac{1}{1 - \epsilon} \sum_{\langle i,j \rangle} [\hat{q}_{ij}^{11} \cdot \hat{p}_{ij} + \hat{q}_{ij}^{00} \cdot (1 - \hat{p}_{ij})] \\
&+ \frac{2\eta - 1}{\epsilon_1} \sum_{\langle i,j \rangle} \hat{q}_{ij}^{10} \cdot \hat{p}_{ij} - \frac{2\eta - 1}{1 - \epsilon_1} \sum_{\langle i,j \rangle} \hat{q}_{ij}^{10}(1 - \hat{p}_{ij}) \\
\frac{\partial \mathcal{F}\mathcal{E}}{\partial \eta} &= \frac{2\epsilon - 1}{\epsilon_1} \sum_{\langle i,j \rangle} \hat{q}_{ij}^{10} \cdot \hat{p}_{ij} - \frac{2\epsilon - 1}{1 - \epsilon_1} \sum_{\langle i,j \rangle} \hat{q}_{ij}^{10}(1 - \hat{p}_{ij})
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 \mathcal{F}\mathcal{E}}{\partial \epsilon^2} &= -\frac{1}{\epsilon^2} \sum_{\langle i,j \rangle} [\hat{q}_{ij}^{11} \cdot \hat{p}_{ij} + \hat{q}_{ij}^{00}(1 - \hat{p}_{ij})] \\
&- \frac{1}{(1 - \epsilon)^2} \sum_{\langle i,j \rangle} [\hat{q}_{ij}^{11} \cdot \hat{p}_{ij} + \hat{q}_{ij}^{00}(1 - \hat{p}_{ij})] \\
&- \left(\frac{2\eta - 1}{\epsilon_1} \right)^2 \sum_{\langle i,j \rangle} \hat{q}_{ij}^{10} \cdot \hat{p}_{ij} \\
&- \left(\frac{2\eta - 1}{1 - \epsilon_1} \right)^2 \sum_{\langle i,j \rangle} \hat{q}_{ij}^{10}(1 - \hat{p}_{ij}) \\
\frac{\partial^2 \mathcal{F}\mathcal{E}}{\partial \eta^2} &= -\left(\frac{2\epsilon - 1}{\epsilon_1} \right)^2 \sum_{\langle i,j \rangle} \hat{q}_{ij}^{10} \cdot \hat{p}_{ij} \\
&- \left(\frac{2\epsilon - 1}{1 - \epsilon_1} \right)^2 \sum_{\langle i,j \rangle} \hat{q}_{ij}^{10}(1 - \hat{p}_{ij}) \\
\frac{\partial^2 \mathcal{F}\mathcal{E}}{\partial \eta \partial \epsilon} &= \frac{1}{\epsilon_1^2} \sum_{\langle i,j \rangle} \hat{q}_{ij}^{10} \cdot \hat{p}_{ij} - \frac{1}{(1 - \epsilon_1)^2} \sum_{\langle i,j \rangle} \hat{q}_{ij}^{10}(1 - \hat{p}_{ij}).
\end{aligned}$$

ACKNOWLEDGMENT

The authors would like to thank Y. Rathi and C.-F. Westin for their assistance in acquiring and processing the diffusion data.

REFERENCES

- [1] R. L. Buckner and J. L. Vincent, "Unrest at rest: Default activity and spontaneous network correlations," *NeuroImage*, vol. 37, pp. 1091–1096, 2007.
- [2] M. D. Fox and M. E. Raichle, "Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging," *Nature*, vol. 8, pp. 700–711, 2007.
- [3] M. D. Greicius, B. H. Flores, V. Menon, G. H. Glover, H. B. Solvason, H. Kenna, A. L. Reiss, and A. F. Schatzberg, "Resting-state functional connectivity in major depression: Abnormally increased contributions from subgenual cingulate cortex and thalamus," *Biol. Psychiatry*, vol. 62, pp. 429–437, 2007.
- [4] M. Liang, Y. Zhou, T. Jiang, Z. Liu, L. Tian, H. Liu, and Y. Hao, "Widespread functional disconnectivity in schizophrenia with resting-state functional magnetic resonance imaging," *Neuroreport Brain Imag.*, vol. 17, pp. 209–213, 2006.
- [5] M. Ke, X. Huang, H. Shen, Z. Zhou, X. Chen, and D. Hu, "Combined analysis for resting state fMRI and DTI data reveals abnormal development of function-structure in early-onset schizophrenia," *Lecture Notes Artif. Intell.*, vol. 5009, pp. 628–635, 2008.
- [6] Y. Zhou, N. Shu, Y. Liu, M. SONG, Y. Hao, H. Liu, C. Yu, Z. Liu, and T. Jiang, "Altered resting-state functional connectivity and anatomical connectivity of hippocampus in schizophrenia," *Schizophrenia Res.*, vol. 100, pp. 120–132, 2008.
- [7] M. J. Jafri, G. D. Pearson, M. Stevens, and V. D. Calhoun, "A method for functional network connectivity among spatially independent resting-state components in schizophrenia," *Neuroimage*, vol. 39, pp. 1666–81, 2008.
- [8] M. Jafri and V. Calhoun, "Functional classification of schizophrenia using feed forward neural networks," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2006, pp. 6631–6634.
- [9] A. Venkataraman, Y. Rathi, M. Kubicki, C.-F. Westin, and P. Golland, "Joint modeling of anatomical and functional connectivity for population studies," *IEEE Trans. Med. Imag.*, vol. 31, no. 2, pp. 164–182, Feb. 2012.
- [10] A. Venkataraman, M. Kubicki, C.-F. Westin, and P. Golland, "Robust feature selection in resting-state fMRI connectivity based on population studies," in *Proc. IEEE Comput. Soc. Workshop Math. Methods Biomed. Image Anal.*, 2010, pp. 63–70.

- [11] M. D. Greicius, K. Supekar, V. Menon, and R. F. Dougherty, "Resting-state functional connectivity reflects structural connectivity in the default mode network," *Cerebral Cortex*, vol. 19, pp. 72–78, 2008.
- [12] M. A. Koch, D. G. Norris, and M. Hund-Georgiadis, "An investigation of functional and anatomical connectivity using magnetic resonance imaging," *Neuroimage*, vol. 16, pp. 241–250, 2002.
- [13] F. Deligianni, G. Varoquaux, B. Thirion, E. Robinson, D. Sharp, A. Edwards, and D. Rueckert, "A probabilistic framework to infer brain functional connectivity from anatomical connections," in *IPMI: Information Processing in Medical Imaging*. New York: Springer, 2011, pp. 296–307.
- [14] A. Venkataraman, M. Kubicki, and P. Golland, "From brain connectivity models to identifying foci of a neurological disorder," in *Proc. Int. Conf. Med. Image Comput. Computer Assist. Intervent.*, 2012, pp. 697–704, LNCS 7510.
- [15] S. Gabrieli-Whitfield, H. W. Thermenos, Z. Milanovic, M. T. Tsuang, S. V. Faraone, R. W. McCarley, M. E. Shenton, A. I. Green, A. Nieto-Castanon, P. LaViolette, J. Wojcik, J. D. Gabrieli, and L. J. Seidman, "Hyperactivity and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia," *Nat. Acad. Sci. USA*, pp. 1279–1284, 2009.
- [16] J. Burns, D. Job, M. E. Bastin, H. Whalley, T. MacGillivray, E. C. Johnstone, and S. M. Lawrie, "Structural disconnectivity in schizophrenia: A diffusion tensor magnetic resonance imaging study," *Br. J. Psychiatry*, vol. 182, pp. 439–443, 2003.
- [17] K. Friston, A. Holmes, K. Worsley, J.-P. Poline, C. Frith, and R. Frackowiak, "Statistical parametric maps in functional imaging: A general linear approach," *Human Brain Mapp.*, vol. 2, pp. 189–210, 1995.
- [18] Y. Zhou, M. Liang, T. Jiang, L. Tian, Y. Liu, Z. Liu, H. Liu, and F. Kuang, "Functional dysconnectivity of the dorsolateral prefrontal cortex in first-episode schizophrenia using resting-state fMRI," *Neuro Lett.*, vol. 417, pp. 297–302, 2007.
- [19] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
- [20] P. Bassler and C. Pierpaoli, "Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI," *J. Magn. Reson.*, vol. 111, pp. 209–219, 1996.
- [21] M. D. Greicius *et al.*, "Resting-state functional connectivity in the default mode network," *Cerebral Cortex*, vol. 19, pp. 72–78, 2009.
- [22] C. J. Honey, R. Kotter, M. Breakspear, and O. Sporns, "Network structure of cerebral cortex shapes functional connectivity on multiple time scales," *Proc. Nat. Acad. Sci.*, vol. 104, pp. 10240–10245, 2007.
- [23] O. Sporns, G. Tononi, and G. M. Edelman, "Theoretical neuroanatomy: Relating anatomical and functional connectivity in graphs and cortical connection matrices," *Cerebral Cortex*, vol. 10, pp. 127–41, 2000.
- [24] C. J. Honey, O. Sporns, L. Cammoun, X. Gigandet, J. P. Thiran, R. Meuli, and P. Hagmann, "Predicting human resting-state functional connectivity from structural connectivity," *Proc. Nat. Acad. Sci.*, vol. 106, pp. 2035–2040, 2009.
- [25] R. Tandon, M. S. Keshavan, and H. A. Nasrallah, "Schizophrenia, 'just the facts': What we know in 2008, Part 1: Overview," *Schizophrenia Res.*, vol. 100, pp. 4–19, 2008.
- [26] M. E. Shenton, C. C. Dickey, M. Frumin, and R. W. McCarley, "A review of MRI findings in schizophrenia," *Schizophrenia Res.*, vol. 49, pp. 1–52, 2001.
- [27] R. L. Mitchell, R. Elliott, and P. W. Woodruff, "fMRI and cognitive dysfunction in schizophrenia," *Trends Cogn. Sci.*, vol. 5, pp. 71–81, 2001.
- [28] K. J. Friston and C. D. Frith, "Schizophrenia: A disconnection syndrome?," *Clin. Neurosci.*, vol. 3, pp. 89–97, 1995.
- [29] M. Kubicki, R. McCarley, C.-F. Westin, H.-J. Park, S. Maier, R. Kikinis, F. A. Jolesz, and M. E. Shenton, "A review of diffusion tensor imaging studies in schizophrenia," *J. Psychiatric Res.*, vol. 41, pp. 15–30, 2007.
- [30] R. L. Bluhm, J. Miller, R. A. Lanius, E. A. Osuch, K. Boksman, R. W. Neufeld, J. Théberge, B. Schaefer, and P. Williamson, "Spontaneous low-frequency fluctuations in the bold signal in schizophrenic patients: Abnormalities in the default network," *Schizophrenia Bull.*, pp. 1–9, 2007.
- [31] R. L. Buckner, J. R. Andrews-Hanna, and D. L. Schacter, "The brain's default network anatomy, function, and relevance to disease," *Ann. NY Acad. Sci.*, vol. 1124, pp. 1–38, 2008.
- [32] K. Lee, T. Yoshida, M. Kubicki, S. Bouix, C.-F. Westin, G. Kindlmann, M. Niznikiewicz, A. Cohen, R. McCarley, and M. Shenton, "Increased diffusivity in superior temporal gyrus in patients with schizophrenia: A diffusion tensor imaging study," *Schizophrenia Res.*, vol. 108, pp. 33–40, 2009.
- [33] M. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul, "An introduction to variational methods for graphical models," *Mach. Learn.*, vol. 37, pp. 183–233, 1999.
- [34] A. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *J. R. Stat. Soc.*, vol. 39, pp. 1–38, 1977.
- [35] B. Fischl, D. H. Salat, A. J. van der Kouwe, N. Makris, F. Ségonne, B. T. Quinn, and A. M. Dale, "Sequence-independent segmentation of magnetic resonance images," *Neuroimage*, vol. 23, pp. 69–84, 2004.
- [36] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. Behrens, H. Johansen-Bern, P. R. Bannister, M. D. Luca, I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. D. Stefano, J. M. Brady, and P. M. Matthews, "Advances in functional and structural MR image analysis and implementation as FSL," *Neuroimage*, vol. 23, pp. 208–219, 2004.
- [37] J. G. Malcolm, M. E. Shenton, and Y. Rathi, "Neural tractography using an unscented Kalman filter," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, 2009, vol. 21, pp. 126–138.
- [38] K. V. Dijk, T. Hedden, A. Venkataraman, K. Evans, S. Lazar, and R. Buckner, "Intrinsic functional connectivity as a tool for human connectomics: Theory, properties and optimization," *J. Neurophysiol.*, vol. 103, pp. 297–321, 2010.
- [39] K. Kasai, M. Shenton, D. Salisbury, Y. Hirayasu, T. Onitsuka, M. Spencer, D. Yurgelun-Todd, R. Kikinis, F. Jolesz, and R. McCarley, "Progressive decrease of left Heschl gyrus and planum temporale gray matter volume in first-episode schizophrenia, a longitudinal magnetic resonance imaging study," *Arch. General Psychiatry*, vol. 60, pp. 766–775, 2003.
- [40] G. Pearlson, P. Barta, R. Powers, R. Menon, S. Richards, E. Aylward, E. Federman, G. Chase, R. Petty, and A. Tien, "Median and superior temporal gyral volumes and cerebral asymmetry in schizophrenia versus bipolar disorder," *Biol. Psychiatry*, vol. 41, pp. 1–14, 1997.
- [41] A. G. Garrity, G. D. Pearlson, K. McKiernan, D. Lloyd, K. A. Kiehl, and V. D. Calhoun, "Aberrant 'default mode' functional connectivity in schizophrenia," *Am. J. Psychiatry*, vol. 164, pp. 450–457, 2007.
- [42] Y. Zhou, M. Liang, T. Jiang, L. Tian, Y. Liu, Z. Liu, H. Liu, and F. Kuang, "Functional dysconnectivity of the dorsolateral prefrontal cortex in first-episode schizophrenia using resting-state fMRI," *Neurosci. Lett.*, vol. 417, pp. 297–302, 2007.
- [43] E. Melonakos, M. Shenton, Y. Rathi, D. Terry, S. Bouix, and M. Kubicki, "Voxel-based morphometry (VBM) studies in schizophrenia—Can white matter changes be reliably detected with VBM," *Psychiatry Res.*, vol. 193, pp. 65–70, 2011.