

Bayesian Community Detection in the Space of Group-Level Functional Differences

Archana Venkataraman*, Daniel Y.-J. Yang, Kevin A. Pelphrey, and James S. Duncan

Abstract—We propose a unified Bayesian framework to detect both hyper- and hypo-active communities within whole-brain fMRI data. Specifically, our model identifies dense subgraphs that exhibit population-level differences in functional synchrony between a control and clinical group. We derive a variational EM algorithm to solve for the latent posterior distributions and parameter estimates, which subsequently inform us about the afflicted network topology. We demonstrate that our method provides valuable insights into the neural mechanisms underlying social dysfunction in autism, as verified by the Neurosynth meta-analytic database. In contrast, both univariate testing and community detection via recursive edge elimination fail to identify stable functional communities associated with the disorder.

Index Terms—Bayesian modeling, community detection, functional magnetic resonance imaging, population analysis.

I. INTRODUCTION

UNIVARIATE analyses of fMRI activation patterns [1], [2] have allowed us to localize functional differences induced by neurological disorders, such as autism [3]–[5]. Despite the insights gained from univariate methods, there is increasing evidence that complex pathologies reflect distributed impairments across multiple brain systems [6]–[8]. These findings underscore the importance of network-based approaches for functional neuroimaging data.

The seminal works of [9]–[11] propose a series of large-scale network measures, which have precipitated the discovery of small-world functional architectures in the brain [12]. Disruptions of the small-world organization are believed to play a role in neurological disorders [13], [14]. However, these aggregate

properties (e.g., small-worldness, centrality and degree distribution) do not illuminate concrete etiological mechanisms. Recently, more sophisticated models, based on directed Bayesian networks, have been proposed for fMRI data [15], [16]. Unfortunately, due to the large number of parameters, causal interactions are typically estimated on a restricted set of nodes. Moreover, these methods have difficulty estimating the directionality of connection information due to noise and subject differences. In contrast, we present a hierarchical framework that captures the underlying topology of the altered functional pathways. We extract both hyper- and hypo-active communities while simultaneously accommodating population variability.

Community detection is the process of identifying highly interconnected subgraphs within a larger network; these nodes (i.e., community members) share common properties and are crucial to understanding the organization of complex systems [17], [18]. Traditional community detection algorithms include hierarchical cluster analysis [19], which constructs an agglomerative dendrogram based on pairwise similarity measures, and spectral clustering [20], [21], which leverages higher-order network information in the Laplacian eigenvectors. Despite their prevalence, both methods fail to detect stable communities in noisy graphs, or when the clusters are significantly mixed [17], [22]. Recently, there has been a shift towards divisive algorithms, which identify and remove inter-cluster edges. The seminal work of [23] recursively eliminates network edges based on three centrality measures, and most importantly, recalculates the edge scores after each iteration. Similar approaches have been developed using modularity [24]–[26], clustering coefficient [27], and information compression [28] as selection criteria. While promising, the above methods are heuristic and do not assume a principled model of community structure [18].

Modularity-based approaches have recently been applied to fMRI data [29]–[31]. Broadly, the works of [29], [30] estimate a hierarchical community organization independently within each subject and use a post-hoc cluster matching procedure to perform group-level statistical analysis. Unfortunately, the results depend heavily on preselected thresholds and subroutines (i.e., adjacency matrix construction, subgraph alignment). Moreover, since these approaches do not impose topological constraints on the group differences, the clinical findings are difficult to interpret. The alternative method reported in [31] identifies statistical differences in betweenness centrality and average path length, neither of which informs us about the network structure and organization. In contrast, our model automatically identifies the abnormal networks with minimal user input.

Manuscript received January 18, 2016; accepted February 22, 2016. Date of publication March 02, 2016; date of current version July 29, 2016. This work was supported in part by R01 NS035193 (NINDS) and by R01 MH100028 (NIMH). D. Yang is also supported by the Autism Speaks Meixner Postdoctoral Fellowship in Translational Research (#9284). *Asterisk indicates corresponding author.*

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the authors.

*A. Venkataraman is with the Department of Diagnostic Radiology, Yale School of Medicine, New Haven, CT 06520 USA (e-mail: archana.venkataraman@yale.edu).

D. Yang and K. Pelphrey are with the Center for Translational Developmental Neuroscience, Yale School of Medicine, New Haven, CT 06520 USA (e-mail: daniel.yj.yang@yale.edu; kevin.pelphrey@yale.edu).

J. Duncan is with the Department of Diagnostic Radiology, Yale School of Medicine, New Haven, CT 06520 USA and also with the Department of Biomedical Engineering, Yale University, New Haven, CT 06519 USA (e-mail: james.duncan@yale.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2016.2536559

TABLE I
RANDOM VARIABLES (TOP) AND NON-RANDOM PARAMETERS (BOTTOM) IN OUR GRAPHICAL MODEL SHOWN IN FIG. 1.

R_i	Multinomial variable that indicates the state (healthy or belonging to community k) for each brain region i
F_{ij}	Latent functional connectivity between regions i and j in the neurotypical 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 neurotypical population
$\bar{B}_{ij}^m$	Observed fMRI correlation between regions i and j in subject m of the clinical population
π^r	Prior for multinomial region indicator R_i ($K+1$ states)
π^f	Prior for multinomial functional connectivity F_{ij} (3 states)
η	Concentration of abnormal edges within the latent communities
ϵ	Probability of deviating from the idealized graph of aberrant functional connectivity
μ_s	Mean fMRI correlation given that $F_{ij} = s$ ($s = 0, 1, 2$)
σ_s^2	Variance of fMRI correlation given that $F_{ij} = s$ ($s = 0, 1, 2$)

Our approach goes beyond the prior literature in three crucial ways. First, we propose a unified generative model that describes the relationship between population templates and individual subject observations. Second, we perform community detection in the space of group-level functional *differences*, which allows us to identify both heightened and reduced synchrony. Finally, our framework can simultaneously detect multiple abnormal communities of varying type. We have previously used generative models to infer disease foci from resting-state fMRI data [32]–[34]. Our current framework identifies interconnected cliques within task fMRI data (different from an outward spreading model), and it distinguishes between hyper- and hypo-active subgraphs¹. These two properties describe a fundamentally different network topology, which has not been captured by prior Bayesian models. While an early version of this work was presented in conference form [35], the present effort provides detailed derivations of the model and estimation procedure, along with more extensive experimental evaluation of the method.

We demonstrate our approach on an fMRI study of social perception in autism. Autism Spectrum Disorder (ASD) is characterized by impaired social-emotional reciprocity, communication deficits, and stereotyped behavior [36]. Furthermore, many individuals with ASD exhibit enhanced perception of detail, which is related to a sharper focus of spatial attention, or “tunnel vision” [37]. Despite ongoing efforts, the etiological complexity and acute heterogeneity of ASD has greatly limited our understanding of its pathogenesis. One prevailing theory posits that ASD alters the developmental trajectory of inter-regional connections [7], [38], [39]. Accordingly, task fMRI studies reveal both hyper- and hypo-connectivity between the prefrontal cortex, superior temporal sulcus and limbic structures [7], [4]. However, such analyses are restricted to predefined neural systems and do not reflect whole-brain information. Likewise, network approaches have detected group differences in node-based measures [40], [41] but have not illuminated a specific neural pathways for ASD. Our method provides valuable insights into ASD by extracting stable communities that reflect either heightened or reduced brain synchrony. These subnetworks map onto key functional domains, in which autistic individuals exhibit

strengths and vulnerabilities, respectively. We demonstrate robustness of our results via bootstrapping and meta-analysis.

The remainder of this paper is organized as follows. We introduce our generative model in Section II and develop the corresponding inference algorithm in Section III. Section IV presents the methods used to empirically validate our framework. Sections V and VI report our experimental results based on synthetic and clinical data, respectively. Section VII discusses the behavior of our model, its advantages and drawbacks, and future directions of research. Finally, we conclude with a summary of contributions in Section VIII.

II. GENERATIVE MODEL OF ABNORMAL COMMUNITIES

Our graph representation assumes that each node is a region in the brain, and each edge describes the pairwise functional synchrony between nodes. The brain is partitioned into N regions based on a common atlas, which ensures that the region boundaries are consistently defined across subjects. Pairwise observations $\langle i, j \rangle$ are generated by correlating the fMRI signals in region i and region j . We assume that group-level functional differences can be represented by K interconnected subgraphs. In order to model both enhanced and impaired processes, we further stipulate that each subgraph reflects either a uniform increase or decrease relative to the neurotypical functional synchrony between regions. For simplicity, we assume that the K networks are disjoint; this constraint helps us to infer multiple network influences.

Fig. 1 outlines our hierarchical Bayesian framework. The white boxes correspond to *latent* variables, which specify a template organization in the brain that we cannot directly access. Rather, we observe noisy measurements of the hidden structure via subject-wise fMRI correlations (shaded boxes), which are generated stochastically from the population-level latent structure. As seen, the K communities are grouped into subsets of regions based on the indicator variables $\{R_i\}$. Each network is associated with a binary label Y_k that denotes either a hyper-active ($Y_k = 1$) or a hypo-active ($Y_k = 0$) subgraph. In this work Y_k is specified *a priori* by the user for greater control over the results. This level of input is akin to setting the number of clusters or dictionary components in an unsupervised learning scenario, and it enables us to explore the network evolution in the absence of ground truth. If desired, it is straightforward to model Y_k as an unknown Bernoulli random variable, as described in the Supplementary Materials.

¹In this work, the terms “hyper-active” and “hypo-active” refer to networks of differential functional synchrony, as opposed to the strength of response to an experimental stimulus. None of the methods presented in this paper rely on univariate activation strength, as defined via GLM [1].

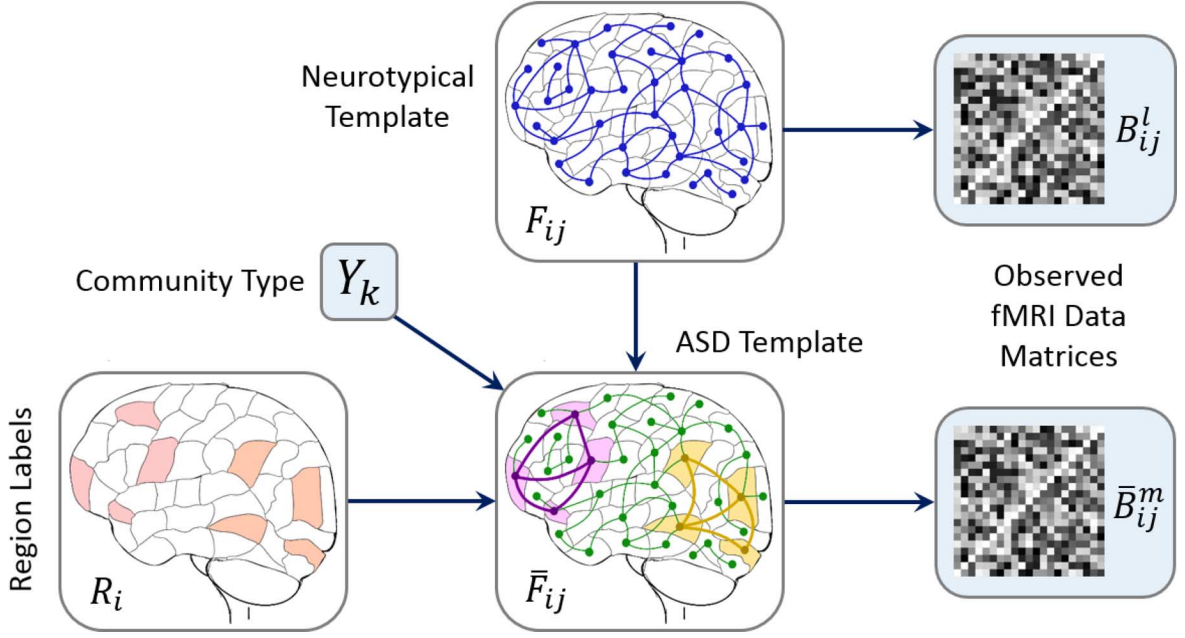


Fig. 1. Hierarchical model of community structure for $K = 2$. The label R_i indicates whether region i is healthy (white) or whether it belongs to one of the two abnormal communities (red). The binary variable Y_k denotes either a hyper-active (purple) or hypo-active (yellow) subgraph k . The neurotypical template $\{F_{ij}\}$ provides a baseline functional architecture for the brain, whereas the clinical template $\{\bar{F}_{ij}\}$ describes the latent organization of ASD. The green connections $\langle i, j \rangle$ are unchanged from baseline; the purple and yellow lines signify heightened and reduced synchrony, respectively. Each template generates a set of subject observations $\{B_{ij}^l\}$ and $\{\bar{B}_{ij}^m\}$ for the group.

We use multinomial variables F_{ij} and $\bar{F}_{ij}$ to denote the latent functional synchrony between regions i and j in the neurotypical and clinical groups, respectively. Empirically, we find that three states: low ($F_{ij} = 0$), medium ($F_{ij} = 1$), and high ($F_{ij} = 2$), adequately capture the dynamic range and variability of our fMRI dataset. The conditional relationships are defined such that $\bar{F}_{ij} > F_{ij}$ within a hyper-active community, and $\bar{F}_{ij} < F_{ij}$ within a hypo-active community.

The fMRI metrics B_{ij}^l for subject l and $\bar{B}_{ij}^m$ for subject m are noisy observations of the underlying latent structure. Table I summarizes the random variables and nonrandom parameters in our generative model. Although the variable definitions are similar to [33], our current framework specifies a multi-cluster community organization. Below we formalize the mathematical representations.

Node Selection: Let N represent the total number of regions in the brain. The multinomial random variable R_i ($i = 1, \dots, N$) indicates whether region i is healthy ($R_i = 0$), or whether it belongs to an abnormal community k ($R_i = k$). We assume an *i.i.d.* multinomial prior for R_i as follows:

$$P(R_i = k; \pi^r) = \pi_k^r, \quad k = 0, \dots, K. \quad (1)$$

The $(K + 1)$ unknown parameters π_k^r specify the *a priori* probabilities of a node assignment such that $\sum_k \pi_k^r = 1$. Our model assumes that these values are shared by all regions.

Neurotypical Connectivity: The latent functional synchrony F_{ij} denotes the co-activation between regions i and j in the neurotypical template. As previously described, F_{ij} is modeled as a tri-state random variable. For notational convenience, we represent F_{ij} as a length-three indicator vector, such that exactly one of its elements $[F_{ij0} \ F_{ij1} \ F_{ij2}]^T$ is

equal to one. The multinomial prior is *i.i.d.* across all pairwise connections, i.e.,

$$P(F_{ij}; \pi^f) = \prod_{s=0}^2 (\pi_s^f)^{F_{ij s}}. \quad (2)$$

Abnormal Network Topology: The clinical variable $\bar{F}_{ij}$ in Fig. 1 is also tri-state; its value depends on the control template F_{ij} and the abnormal communities defined by the region labels R and subgraph types Y . To derive the conditional distribution $P(\bar{F}_{ij} | F_{ij}, R, Y)$, we first define an auxiliary graph $\{T_{ij}\}$, which describes the idealized functional differences between member nodes of a community. Mathematically, T_{ij} is a binary random variable that indicates either a healthy ($T_{ij} = 0$) or an abnormal ($T_{ij} = 1$) edge between nodes i and j :

$$P(T_{ij} = 1 | F_{ij}, R_i, R_j, Y; \eta) = \begin{cases} 0, & R_i \neq R_j \text{ or } R_i = R_j = 0 \\ 0, & R_i = R_j > 0, Y_k = 1, F_{ij} = 2, \\ 0, & R_i = R_j > 0, Y_k = 0, F_{ij} = 0, \\ \eta, & R_i = R_j > 0, Y_k = 1, F_{ij} < 2 \\ \eta, & R_i = R_j > 0, Y_k = 0, F_{ij} > 0. \end{cases} \quad (3)$$

Notice that the first case in (3) specifies the community organization of our abnormal subnetworks. Specifically, if regions i and j are not part of the same subnetwork (i.e., they have different labels or are both normal), then the corresponding functional edge is modeled as healthy ($T_{ij} = 0$ deterministically). By extension, the functional edges that differ between groups must concentrate among member nodes of the same subnetwork; this defines a community topology.

TABLE II
CONDITIONAL DISTRIBUTION $P(\bar{F}_{ij}|F_{ij}, R, Y; \eta, \epsilon)$ FOR HYPER-ACTIVE (TOP) AND HYPO-ACTIVE (BOTTOM) COMMUNITIES.

$$R_i = R_j = k > 0 \text{ and } Y_k = 1$$

		$\bar{F}_{ij}$		
		0	1	2
F_{ij}	0	$(1 - \eta)(1 - \epsilon) + \eta\epsilon$	$(1 - \eta)(\epsilon/2) + \eta(\frac{1-\epsilon}{2})$	$(1 - \eta)(\epsilon/2) + \eta(\frac{1-\epsilon}{2})$
	1	$\epsilon/2$	$(1 - \eta)(1 - \epsilon) + \eta(\epsilon/2)$	$(1 - \eta)(\epsilon/2) + \eta(1 - \epsilon)$
	2	$\epsilon/2$	$\epsilon/2$	$1 - \epsilon$

$$R_i = R_j = k > 0 \text{ and } Y_k = 0$$

		$\bar{F}_{ij}$		
		0	1	2
F_{ij}	0	$1 - \epsilon$	$\epsilon/2$	$\epsilon/2$
	1	$(1 - \eta)(\epsilon/2) + \eta(1 - \epsilon)$	$(1 - \eta)(1 - \epsilon) + \eta(\epsilon/2)$	$\epsilon/2$
	2	$(1 - \eta)(\epsilon/2) + \eta(\frac{1-\epsilon}{2})$	$(1 - \eta)(\epsilon/2) + \eta(\frac{1-\epsilon}{2})$	$(1 - \eta)(1 - \epsilon) + \eta\epsilon$

On the flip side, when both regions are in the same community (i.e., $R_i = R_j > 0$), then we have two scenarios. First, the border conditions $Y_k = 1$, $F_{ij} = 2$ and $Y_k = 0$, $F_{ij} = 0$ represent hard constraints based on our variable definitions. Specifically, if $F_{ij} = 2$, then $\bar{F}_{ij}$ cannot assume a higher state, as indicated by the community type $Y_k = 1$, and likewise for the lower bound $F_{ij} = 0$ and $Y_k = 0$. In these cases, we assume that the latent functional synchrony is unaltered between the groups (i.e., $T_{ij} = 0$). Alternatively, if the border conditions do not apply (last two cases), then T_{ij} is Bernoulli with unknown parameter η . Notice that (3) is a special case of the planted ℓ -partition model [18], where η controls the density of edges within the abnormal communities.

Ideally, $\bar{F}_{ij} = F_{ij}$ for healthy edges. However, if $T_{ij} = 1$, then $\bar{F}_{ij} > F_{ij}$ for $Y_k = 1$, and $\bar{F}_{ij} < F_{ij}$ for $Y_k = 0$. However, to accommodate the complexities of fMRI data, we introduce a second level of noise, in which the scalar parameter ϵ controls the probability of deviating from the above rules. We marginalize the graph $\{T_{ij}\}$ to better estimate the community assignments R_i . Beginning with the first condition in (3), if regions i and j are not in the same community, then $T_{ij} = 0$ and we obtain the following baseline distribution:

$$P(\bar{F}_{ij}|F_{ij}, \{R_i \neq R_j | R_i, R_j = 0\}; \epsilon) = (1 - \epsilon)^{F_{ij}^T \bar{F}_{ij}} \left(\frac{\epsilon}{2}\right)^{1 - F_{ij}^T \bar{F}_{ij}}, \quad (4)$$

such that $F_{ij}^T \bar{F}_{ij} = 1$ if the functional synchrony is the same between groups, and $F_{ij}^T \bar{F}_{ij} = 0$ otherwise.

The conditional relationships between community members (i.e., $R_i = R_j > 0$) are given in Table II. Due to the distinction between hyper- and hypo-synchronous connections, the overall distribution cannot be represented compactly, as in our previous works [33], [42]. In practice, we expect the latent noise to be

small ($\epsilon < 0.5$) and the edge density to be moderate ($0.4 < \eta < 0.6$). Hence, the largest entries in Table II will involve the term $1 - \epsilon$. Notice that this term appears in the upper quadrant (i.e., $\bar{F}_{ij} \geq F_{ij}$) for the hyper-active case and in the lower quadrant (i.e., $\bar{F}_{ij} \leq F_{ij}$) for the hypo-active case. For notational convenience, we denote entries of Table II as $\mathcal{M}_{ss'}^{Y_k}(\eta, \epsilon)$, where $F_{ij} = s$, $\bar{F}_{ij} = s'$, and $Y_k \in \{1, 0\}$, for the top and bottom sub-tables.

Data Likelihood: The fMRI correlation B_{ij}^l between regions i and j in neurotypical subject l is a noisy realization of the latent functional template F_{ij} . We assume a gaussian distribution, with mean and variance controlled by F_{ij} :

$$P(B_{ij}^l|F_{ij} = s; \{\mu, \sigma^2\}) = \prod_{s=0}^1 [\mathcal{N}(B_{ij}^l; \mu_s, \sigma_s^2)]^{f_{ijs}}. \quad (5)$$

The likelihood for a clinical patient $\bar{B}_{ij}^m$ has the same functional form as (5) but relies on the clinical template $\bar{F}_{ij}$. We compute B_{ij}^l via pearson correlation coefficients. Our prior work [42] has demonstrated that Gaussian distributions can adequately capture the global fMRI statistics, while greatly simplifying the parameter estimation steps.

III. MODEL INFERENCE

Given the observed fMRI measurements $X = \{B, \bar{B}\}$, we employ a maximum likelihood (ML) framework to fit our hierarchical Bayesian model to the data. Let $\Theta = \{\pi^r, \pi^f, \eta, \epsilon, \mu, \sigma^2\}$ denote the collection of non-random model parameters, and let L and M be the number of subjects in the neurotypical and clinical populations, respectively. For convenience, we now switch to a binary vector representation for the region labels R_i . Specifically, $R_i = [R_{i0} \ R_{i1} \ \dots \ R_{iK}]^T$, such that the state $R_i = k$ implies that $R_{ik} = 1$ and the remaining entries are equal zero.

Combining the distributions in (1)–(5) with the expressions in Table II yields the following joint density for the latent and observed random variables:

$$\begin{aligned}
\log P(R, F, \bar{F}, X|Y; \Theta) = & \sum_{i=1}^N \sum_{k=0}^K R_{ik} \log(\pi_k^r) \\
& + \sum_{\langle i,j \rangle} \sum_{s=0}^2 F_{ijs} \log(\pi_s^f) \\
& + \sum_{\langle i,j \rangle} \left(1 - \sum_{k=1}^K R_{ik} R_{ij} \right) (F_{ij}^T \bar{F}_{ij} \log(1 - \epsilon)) \\
& + \sum_{\langle i,j \rangle} \left(1 - \sum_{k=1}^K R_{ik} R_{ij} \right) \left((1 - F_{ij}^T \bar{F}_{ij}) \log\left(\frac{\epsilon}{2}\right) \right) \\
& + \sum_{k=1}^K \sum_{\langle i,j \rangle} R_{ik} R_{jk} \sum_{s=0}^2 \sum_{s'=0}^2 F_{ijs} \bar{F}_{ijs'} (Y_k \cdot \log \mathcal{M}_{ss'}^1(\eta, \epsilon)) \\
& + \sum_{k=1}^K \sum_{\langle i,j \rangle} R_{ik} R_{jk} \sum_{s=0}^2 \sum_{s'=0}^2 F_{ijs} \bar{F}_{ijs'} ((1 - Y_k) \cdot \log \mathcal{M}_{ss'}^0(\eta, \epsilon)) \\
& + \sum_{\langle i,j \rangle} \sum_{s=0}^2 F_{ijs} \sum_{l=1}^L \log \mathcal{N}(B_{ij}^l; \mu_s, \sigma_s^2) \\
& + \sum_{\langle i,j \rangle} \sum_{s'=0}^2 \bar{F}_{ijs} \sum_{m=1}^M \log \mathcal{N}(\bar{B}_{ij}^m; \mu_{s'}, \sigma_{s'}^2). \quad (6)
\end{aligned}$$

We observe that the community structure is specified by the third and fourth lines of (6), which denote the natural logarithm of the baseline distribution in (4). Also, while (6) assumes that the subgraph types $\{Y_k\}$ are given, the contribution of Y_k in lines 5–6 is equivalent to that of a Bernoulli random variable. Therefore, by adding an *i.i.d.* Bernoulli prior: $\log P(Y_k; \pi^y) = Y_k \log(\pi^y) + (1 - Y_k) \log(1 - \pi^y)$, we can model these values as unknown. The incremental changes to our model are presented in the Supplementary Materials.

The community assignments R_i induce a complex coupling across pairwise connections $\langle i, j \rangle$. Therefore, we adopt a variational approximation [43] to infer the latent posterior distribution $Q(\cdot)$ and corresponding parameters $\hat{\Theta}$, which minimize the variational free energy

$$\mathcal{FE} = -E_Q[\log P(R, F, \bar{F}, X|Y; \Theta)] - \mathcal{H}(Q). \quad (7)$$

Notice that the quantity $-\mathcal{FE}$ is a lower bound to the marginal log-likelihood $P(X|Y; \Theta)$ in the original ML framework.

Our approximate posterior assumes the factorized form:

$$\begin{aligned}
Q(R, F, \bar{F}) = & \prod_{i=1}^N Q_i^r(R_i; \tilde{u}_i) \prod_{\langle i,j \rangle} Q_{ij}^c(F_{ij}, \bar{F}_{ij}; \tilde{v}_{ij}) \\
= & \prod_{i=1}^N \prod_{k=0}^K (\tilde{u}_{ik})^{R_{ik}} \prod_{\langle i,j \rangle} \prod_{s=0}^2 \prod_{s'=0}^2 (\tilde{v}_{ij}^{ss'})^{F_{ijs} \bar{F}_{ijs'}}. \quad (8)
\end{aligned}$$

As seen, $Q_i^r(\cdot)$ is a multinomial distribution with $K + 1$ states, parameterized by $\tilde{u}_i$. Likewise, $Q_{ij}^c(\cdot)$ is a multinomial distribution with 9 states, parameterized by $\tilde{v}_{ij}$, which account for

the 9 configurations in Table II. Essentially, $\tilde{u}_{ik}$ is the posterior probability that $R_i = k$, and $\tilde{v}_{ij}^{ss'}$ is the posterior probability that $F_{ij} = s$ and $\bar{F}_{ij} = s'$. The factorization in (8) preserves the region-to-connection dependencies and is scalable to accommodate a large number of parcels N .

Using (8), we can expand the entropy term $\mathcal{H}(Q)$ in (7) as shown below:

$$-\mathcal{H}(Q) = \sum_{i=0}^N \sum_{k=0}^K \tilde{u}_{ik} \log(\tilde{u}_{ik}) + \sum_{\langle i,j \rangle} \sum_{s=0}^2 \sum_{s'=0}^2 \tilde{v}_{ij}^{ss'} \log(\tilde{v}_{ij}^{ss'}). \quad (9)$$

The expectation with respect to $Q(\cdot)$ of the complete log-likelihood shares the same form as (6), but with the following substitutions:

$$\begin{aligned}
R_{ik} & \rightarrow \tilde{u}_{ik} \\
F_{ijs} \bar{F}_{ijs'} & \rightarrow \tilde{v}_{ij}^{ss'} \\
F_{ij}^T \bar{F}_{ij} & \rightarrow \sum_{s=0}^2 \tilde{v}_{ij}^{ss} \triangleq \tilde{v}_{ij}^{eq}.
\end{aligned}$$

In the next section, we describe a coordinate descent algorithm to estimate both the parameters $\{\tilde{u}_{ik}, \tilde{v}_{ij}^{ss'}\}$ of the variational posterior and the nonrandom model parameters Θ .

A. Variational EM Algorithm

1) *E-Step*: For a fixed setting of model parameters $\hat{\Theta}$, we iteratively update the elements of $Q^r(\cdot)$ and $Q^c(\cdot)$ to minimize the variational free energy. The iterations for $\tilde{v}_{ij}$ can be expressed in closed form given $\{\tilde{u}_{ik}\}$:

$$\begin{aligned}
\tilde{v}_{ij}^{ss'} \propto & \exp \left(\left[1 - \sum_{k=1}^K \tilde{u}_{ik} \tilde{u}_{jk} \right] \begin{cases} \log(1 - \epsilon), & s = s' \\ \log(\frac{\epsilon}{2}), & s \neq s' \end{cases} \right) \\
& + \log(\pi_s^f) + \sum_{k=1}^K \tilde{u}_{ik} \tilde{u}_{jk} Y_k \cdot \log \mathcal{M}_{ss'}^1(\eta, \epsilon) \\
& + \sum_{k=1}^K \tilde{u}_{ik} \tilde{u}_{jk} (1 - Y_k) \cdot \log \mathcal{M}_{ss'}^0(\eta, \epsilon) \\
& + \sum_{l=1}^L \log \mathcal{N}(B_{ij}^l; \mu_s, \sigma_s^2) + \sum_{m=1}^M \log \mathcal{N}(\bar{B}_{ij}^m; \mu_{s'}, \sigma_{s'}^2) \Big), \quad (10)
\end{aligned}$$

where the proportionality constant is computed such that $\sum_s \sum_{s'} \tilde{v}_{ij}^{ss'} = 1$. The first three lines in (10) leverage the prior and latent functional interactions, whereas the last line incorporates the data likelihood for both populations.

In contrast, differentiating $\mathcal{FE}$ with respect to $\tilde{u}_{ik}$ results in coupled update expressions

$$\begin{aligned}
\tilde{u}_{ik} \propto & \exp \left(\sum_{j \neq i} \tilde{u}_{jk} \left(\tilde{v}_{ij}^{eq} \log(1 - \epsilon) + (1 - \tilde{v}_{ij}^{eq}) \log\left(\frac{\epsilon}{2}\right) \right) \right) \\
& + \log(\pi_k^r) + \sum_{j \neq i} \tilde{u}_{jk} \sum_{s=0}^2 \sum_{s'=0}^2 \tilde{v}_{ij}^{ss'} Y_k \cdot \log \mathcal{M}_{ss'}^1(\eta, \epsilon) \\
& + \sum_{j \neq i} \tilde{u}_{jk} \sum_{s=0}^2 \sum_{s'=0}^2 \tilde{v}_{ij}^{ss'} (1 - Y_k) \cdot \log \mathcal{M}_{ss'}^0(\eta, \epsilon) \Big). \quad (11)
\end{aligned}$$

Since (11) depends on the other region posterior estimates $\tilde{u}_{jk}$, we perform an inner fixed-point iteration until convergence of the distribution $Q^r(R)$.

The estimates for $\hat{\Theta}$ in the following section rely on marginal probabilities of $Q_{ij}^c(F_{ij}, \bar{F}_{ij})$. We compute these quantities after convergence of the full variational posterior:

$$\tilde{f}_{ijs} = Q(F_{ij} = s | X, Y; \Theta) = \sum_{s'=0}^2 \tilde{v}_{ij}^{ss'} \quad (12)$$

$$\tilde{g}_{ijs'} = Q(\bar{F}_{ij} = s' | X, Y; \Theta) = \sum_{s=0}^2 \tilde{v}_{ij}^{ss'}. \quad (13)$$

2) *M-Step*: We update the model parameters $\hat{\Theta}$ by fixing the posterior distribution $Q(R, F, \bar{F})$ and setting the gradient of (7) with respect to $\hat{\Theta}$ equal to zero. The expressions for the multinomial prior parameters $\{\pi^r, \pi^f\}$ reduce to averages over the marginal posterior distributions:

$$\hat{\pi}_k^r = \frac{1}{N} \sum_{i=1}^N \tilde{u}_{ik} \quad (14)$$

$$\hat{\pi}_s^f = \frac{1}{C} \sum_{\langle i,j \rangle} \tilde{f}_{ijs} \quad (15)$$

where $C = N(N-1)/2$ is the total number of unique pairwise connections between N regions.

The fMRI likelihood parameters are estimated as weighted statistics of the data. The updates parallel those of a standard Gaussian mixture model:

$$\hat{\mu}_s = \frac{\sum_{\langle i,j \rangle} [\tilde{f}_{ijs} \sum_l B_{ij}^l + \tilde{g}_{ijs} \sum_m \bar{B}_{ij}^m]}{\sum_{\langle i,j \rangle} [L \cdot \tilde{f}_{ijs} + M \cdot \tilde{g}_{ijs}]}, \quad (16)$$

$$\hat{\sigma}_s^2 = \frac{\sum_{\langle i,j \rangle} [\tilde{f}_{ijs} \sum_l (B_{ij}^l - \hat{\mu}_k)^2 + \tilde{g}_{ijs} \sum_m (\bar{B}_{ij}^m - \hat{\mu}_k)^2]}{\sum_{\langle i,j \rangle} [L \cdot \tilde{f}_{ijs} + M \cdot \tilde{g}_{ijs}]}. \quad (17)$$

We jointly update the edge concentration η and latent noise ϵ via Newton's method. Details of this step are presented in the Supplementary Materials. We emphasize that given the community types Y , both the posterior distribution and the model parameters are estimated directly from the observed data. We do not tune any auxiliary parameters or thresholds.

B. Initialization

Like many gradient-based methods, the final solution to our model will depend on proper initialization. At a minimum, the variational EM algorithm in Section III-A requires us to initialize the non-random model parameters $\hat{\Theta}$ and the region posterior variables $\{\tilde{u}_{ik}\}$. The algorithm proceeds by updating the connectivity posteriors $\{\tilde{v}_{ij}^{ss'}\}$ in the E-step and iterating between the distributions until convergence. We then estimate the parameters $\hat{\Theta}$ in the M-step and repeat the outer loop.

We initialize the functional prior π^f and the Gaussian means μ and variances σ^2 based on statistics of the data. For regularization, we opt to center the correlation distributions of each subject and fix $\mu_1 = 0$ in subsequent analysis. This procedure effectively models the relative deviations

from the baseline functional synchrony for each subject. We set the initial latent noise parameter to $\epsilon = 0.01$, which encourages consistency between the region labels and the observed fMRI data, and we uniformly sample the edge concentration η from the interval $[0.2, 0.5]$. Larger values of η encourage more abnormal functional edges between community members. Likewise, we randomly initialize the region prior as $\pi^r = [(1-\alpha) \ (\alpha/K) \ \dots \ (\alpha/K)]^T$, such that $\alpha \in [0.2, 0.5]$. As α increases, the algorithm is biased towards larger communities during the first iteration.

Finally, we initialize the region posterior parameters $\tilde{u}_{ik}$ via a two step procedure. First, we average the fMRI correlations across subjects and cluster these values to approximate the discrete latent functional connectivity. Second, we separate the hyper- and hypo-synchronous connections and apply a clique detection algorithm to each grouping. The initial communities are based on the regions associated with largest interconnected cliques. In practice, the initial communities consist of 4–6 regions, which is significantly less than in the final solution.

We mitigate the influence of local minima by randomly initializing the variational EM algorithm 10 times and select the optimal $\mathcal{FE}$ solution. This procedure is applied to both synthetic and real-world experiments.

IV. MODEL EVALUATION

A. Estimating the Abnormal Network Topology

The idealized graph of functional differences $\{T_{ij}\}$ provides valuable insights into the topological properties of each abnormal community by removing the effects of the latent noise ϵ . T_{ij} also describes the subset of community edges that strongly differ between the control and clinical populations.

We can retrospectively approximate these variables based on the maximum *a posteriori* (MAP) solution for R and the parameter estimates $\hat{\Theta}$. Specifically, given R and Y , our model decouples by pairwise connection, so we can assign each T_{ij} independently. Furthermore, (3) implies that $T_{ij} = 0$ for $R_i \neq R_j$ and $R_i = R_j = 0$ (i.e., when the nodes belong to different communities and when they are healthy, respectively). For the remaining case ($R_i = R_j > 0$), we select the value $T_{ij} \in \{0, 1\}$ according to the following optimization:

$$\hat{T}_{ij} = \arg \max_T E_Q [\log P(T | F_{ij}, \bar{F}_{ij}, \hat{R}^{MAP}, Y; \hat{\Theta})]. \quad (18)$$

The right-hand side of (18) can be computed by multiplying the natural logarithm of entries in Table II with the corresponding variational posterior parameter $\tilde{v}_{ij}$ and summing across all nine configurations of $\{F_{ij}, \bar{F}_{ij}\}$. This computation weighs the influence of the edge density η , which controls the proportion of abnormal connections, against the latent noise ϵ , which encourages the functional templates to deviate from the rules outlined in (3).

B. Robustness of the Identified Communities

The marginal posterior statistics $\tilde{u}_{ik}$ inform us whether region i is healthy or if it belongs to one of the K abnormal communities. We evaluate the robustness of these region assignments via bootstrapping. In particular, we fit the model to random subsets of the data, such that the ratio of neurotypical controls to ASD

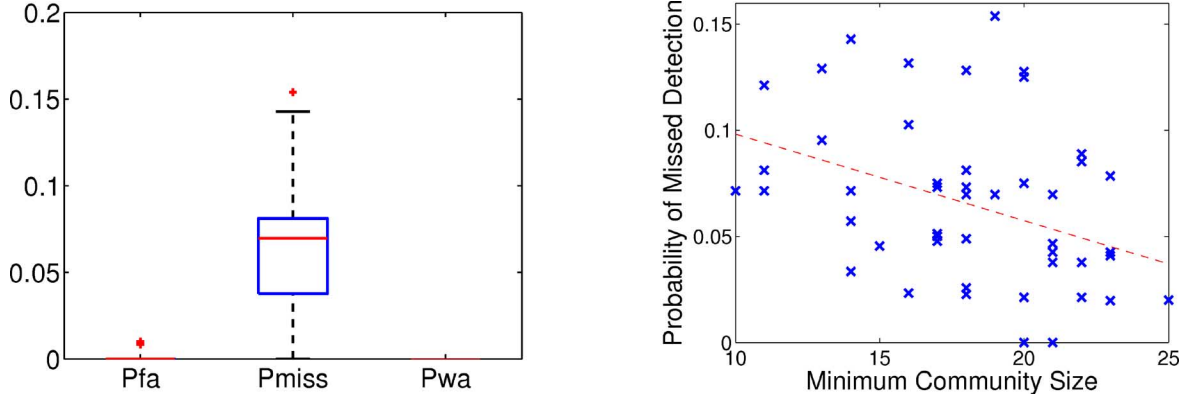


Fig. 2. Results of Experiment I across 50 synthetic trials. Left: Errors in region assignment corresponding to the probability of false alarm (Pfa), the probability of a miss (Pmiss) and the probability of an abnormal region being assigned to the wrong community (Pwa). Right: Inverse relationship between Pmiss and the minimum community size.

patients is preserved. Each subset contains only 80% of the total number subjects. We re-sample the data subsets 50 times and average the estimated posterior $\tilde{u}_{ik}$ across runs to quantify the influence of region i .

We focus on region stability for three reasons. First, due to the community organization built into our model, abnormal edges are constrained to lie between member nodes of the same subnetwork. Moreover, the sign of each edge (i.e., hyper- vs. hypo-synchronous) is given by the binary vector Y . Therefore, the topology of the group-level functional differences is largely specified by just the region assignments and community types. Second, 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. From a clinical perspective, we can use the region assignments to perform follow-up analyses (see Neurosynth below); but the connectivity pattern is much harder to interpret. Finally, our previous work [33], [44] suggests that connection-wise differences may not be as robust as region measures.

C. Baseline Methods

We compare our generative model with two alternative techniques: statistical testing and a well-known community detection procedure, which prunes edges from the original network. These comparisons evaluate the benefits of (1) an imposed community organization, and (2) explicitly differentiating between hyper- and hypo-active communities.

Univariate testing is one of the standard approaches used in population studies of functional connectivity. Specifically, the two sample t-test confirms or rejects the null hypothesis that the group means of a given fMRI correlation are equal. Here, we vary the significance threshold in order to determine whether univariate functional differences naturally exhibit an underlying community structure.

Our second comparison implements the seminal method of [23], which iteratively removes edges from the network in order to discover community structure. The edges are selected according to a *current-flow betweenness* measure, which favors connections that lie between communities rather than ones that span a single community. This procedure requires a sparse binary input graph $\mathbf{A}$ that describes the group-wise functional

differences. Rather than averaging and thresholding the fMRI correlation values, we construct $\mathbf{A}$ via the Bayesian connection model described in [42]. Specifically, $\mathbf{A}_{ij} = 1$ if the posterior probability that the latent functional connectivity templates differ exceeds 0.5. We take this approach because data thresholding introduces nonlinear dependencies into the analysis, which are known to bias network results [45]. Moreover, since the fMRI likelihood in [42] is the same as (5), we can directly evaluate the gain from our community priors.

D. Meta-Analysis via Neurosynth

The Neurosynth database (www.neurosynth.org) provides an unbiased and comprehensive evaluation of the functionality supported by each hyper- and hypo-active community. Specifically, the database aggregates both the MNI activation coordinates reported in prior fMRI studies along with a set of descriptive words and phrases pulled from the abstracts. The meta-analytic framework leverages the power of large datasets to calculate the posterior probability $P(\text{Feature}|\text{Coordinate})$ for an individual psychological feature (i.e., word or phrase) at a given spatial location [46]. Essentially, this reverse inference procedure identifies psychological constructs that have consistently been associated with a particular activation coordinate across a range of fMRI studies (varying paradigms, demographics and clinical statuses). Neurosynth has precomputed and stored 3,099 brain maps based on this posterior information; each map is associated with a given feature term. Additionally, the website creators have used Latent Dirichlet Allocation (LDA) [47] to identify a set of high-level topics from the original 3,099 words and phrases.

In this work, we correlate the spatial regions defined by each hyper- and hypo-active community with the statistical maps stored in the Neurosynth database; terms with a correlation above the default threshold ($r > 0.001$) were retained. We then apply the methods of [46], [48] to perform a whole-brain reverse inference for each LDA topic and rank them according to their association with each abnormal community.

V. RESULTS ON SYNTHETIC DATA

In this section, we use simulated experiments to demonstrate that our variational EM algorithm recovers the ground truth re-

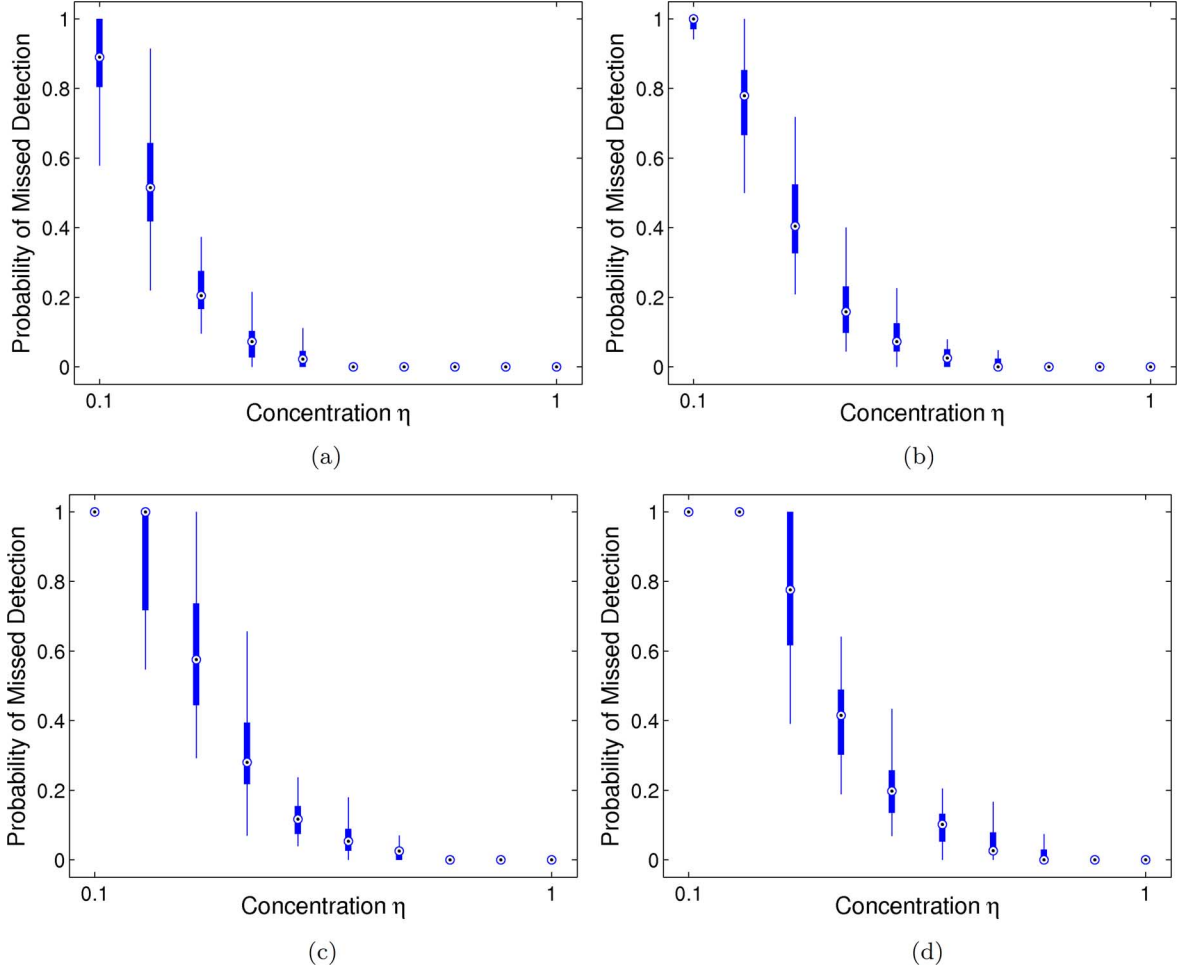


Fig. 3. Probability of missed detection (Type I error) for the **Good Data** parameterization across 50 random samplings of the observed data. Thick lines correspond to the interquartile interval, and the whiskers extend to the 10–90% range. (a) Latent noise $\epsilon = 0.01$ (b) Latent noise $\epsilon = 0.04$ (c) Latent noise $\epsilon = 0.07$ (d) Latent noise $\epsilon = 0.10$.

gion labels. We assume group differences can be explained by one hyper-active and one hypo-active community ($K = 2$). Subject correlations are generated according to Fig. 1 for different noise statistics and community organizations.

A. Experiment I: Preliminary Evaluation

As a baseline, we first generate subject observations using the same signal-to-noise ratio and community properties estimated by the maximum likelihood solution of our clinical dataset. In particular, we set $(\mu_1 - \mu_0) = 0.13$, $(\mu_2 - \mu_1) = 0.2$, $\sigma_0^2 = 0.07$ and $\sigma_1^2 = \sigma_2^2 = 0.06$ for the data likelihood. Notice that there is significant overlap in the correlation distributions for the three latent connectivity states, i.e., the standard deviation of each Gaussian is 2–3 times greater than the separation between the means. We mimic the organization of our clinical dataset by specifying an underlying network with 150 regions and sampling 50 subjects in each population. The latent concentration and noise parameters are fixed at $\eta = 0.5$ and $\epsilon = 0.03$.

During each trial, we uniformly sample the region prior π^r such that 11–16% of regions belong to each abnormal cluster. We generate the latent variables in the following order $R \rightarrow F \rightarrow \bar{F} \rightarrow B \rightarrow \bar{B}$, and we use the variational EM algorithm to infer the region posterior statistics $\tilde{u}_{ik}$. Fig. 2 (left) illustrates

the average errors across 50 random trials. As seen, the probability of a false alarm or wrong assignment is negligible. While the probability of a missed detection is slightly higher, the median error is only 7%, and the 75th percentile error is under 9%. Fig. 2 (right) suggests an inverse relationship between the detection error and the minimum community size. Intuitively, larger communities have more abnormal edges and are easier to identify, whereas smaller communities are more likely to be overwhelmed by noise.

B. Experiment II: Effect of Community Organization

Next, we evaluate the sensitivity and robustness of our algorithm with respect to the edge concentration η and the latent noise ϵ . We expect the region assignment errors to be positively correlated with ϵ and negatively correlated with η . This is because higher values of η increase the density of abnormal connections within each community, thus making them more distinctive. Conversely, larger ϵ results in more functional connections deviating from the ideal community organization (i.e., T_{ij}). Here, we present only the probability of a missed detection.

We sweep the parameters of interest across the range $\eta \in [0.1, 1]$ and $\epsilon \in [0.01, 0.1]$ in increments of 0.1 and 0.01, respectively. For each (η, ϵ) pair, we generate the latent variables

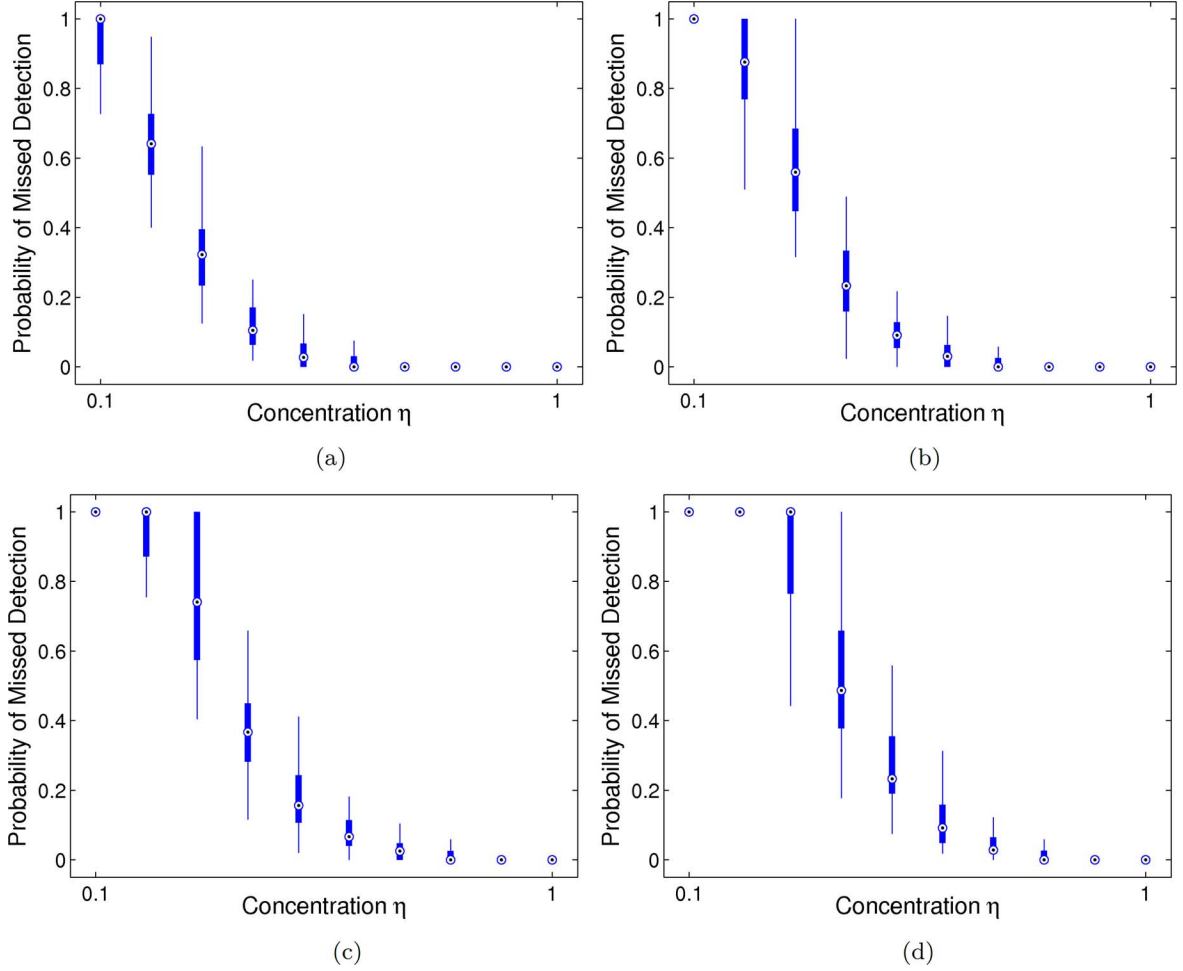


Fig. 4. Probability of missed detection (Type I error) for the **Noisy Data** parameterization across 50 random samplings of the observed data. Thick lines correspond to the interquartile interval, and the whiskers extend to the 10–90% range. (a) Latent noise $\epsilon = 0.01$ (b) Latent noise $\epsilon = 0.04$ (c) Latent noise $\epsilon = 0.07$ (d) Latent noise $\epsilon = 0.10$.

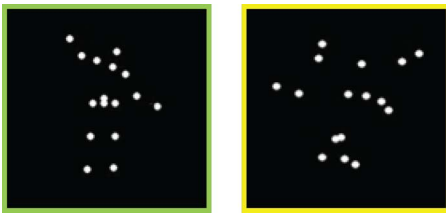


Fig. 5. Snapshot of the visual stimuli. The protocol alternates between coherent biological motion (left) and scrambled animations (right).

and observed data as described in Section V-A. Once again, the region prior π^r is sampled such that 11–16% of region belong to each community. We consider two likelihood parameterizations $\{\mu, \sigma^2\}$. The **Good Data** parameterization clearly separates the data distributions for each latent functional state. Specifically, we set $(\mu_1 - \mu_0) = (\mu_2 - \mu_1) = 0.35$ and $\sigma_s^2 = 0.05$ for $s = 0, 1, 2$. This configuration allows us to accurately infer the latent templates $\{F, \bar{F}\}$, which are then used to delineate abnormal communities. The **Noisy Data** parameterization uses the same likelihoods as Section V-A.

Figs. 3 and 4 report the proportion of missed community nodes (Type I error) across 50 synthetic trials for the

Good Data and **Noisy Data** parameterizations, respectively. These results highlight three particularly striking trends. First, the performance of our model and algorithm worsens with increasing ϵ . Intuitively, ϵ controls the proportion of both missing and spurious connections within the multi-class community organization, which hinders our ability to identify the abnormal subnetworks. Second, in accordance with our initial hypothesis, there is an inverse relationship between the concentration η and the Type I error, with an inflection point at $\eta = 0.5$. The parameter η controls the percentage of nonzero edges between community members and is closely linked to the probability p_{in} from the planted ℓ -partition model [49]. Hence, our model can only extract dense subgraphs. Finally, we observe that the error rates are similar for both likelihood parameterizations. This suggests that errors in community detection are primarily due to inconsistent functional differences rather than to noisy data observations.

VI. SOCIAL PERCEPTION IN AUTISM

A. Biopoint Dataset and fMRI Preprocessing

We demonstrate our methods on a clinical study of 72 ASD children and 43 age-matched ($p > 0.124$) and IQ-matched ($p >$

TABLE III
LATENT COMMUNITY PARAMETERS (TOP) AND DATA LIKELIHOOD (BOTTOM), ESTIMATED BY THE ML SOLUTION FOR THE CLINICAL DATA.

	π_1^r	π_2^r	π_3^r	π_0^f	π_1^f	π_2^f	η	ϵ	μ_0	μ_1	μ_2	σ_0^2	σ_1^2	σ_2^2
$Y_1 = 1$	0.094	—	—	0.34	0.44	0.22	0.65	0.041	-0.13	0	0.19	0.072	0.062	0.058
$Y_1 = 0$	0.25	—	—	0.34	0.44	0.22	0.45	0.034	-0.13	0	0.19	0.072	0.062	0.059
$Y_1 = 1, Y_2 = 0$	0.11	0.22	—	0.34	0.44	0.22	0.54	0.031	-0.13	0	0.19	0.072	0.062	0.059
$Y_1 = 1, Y_2 = Y_3 = 0$	0.12	0.11	0.18	0.34	0.44	0.22	0.52	0.029	-0.13	0	0.19	0.072	0.062	0.059

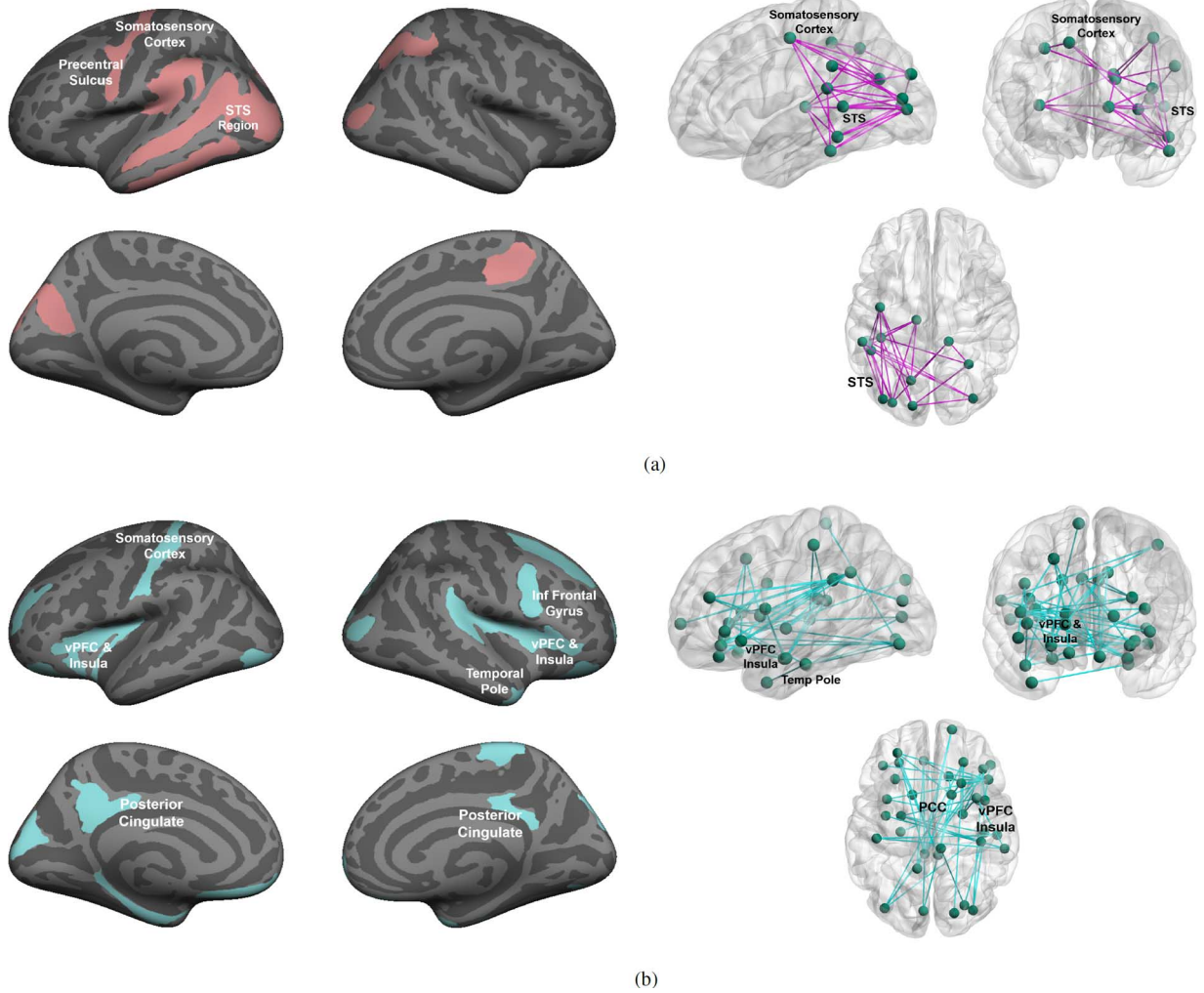


Fig. 6. Abnormal communities inferred by our Bayesian model for $K = 1$. Left: Region membership in each community. Red nodes indicate the hyper-active community, whereas blue areas delineate the hypo-active cluster. Right: Estimated network of abnormal functional synchrony. Each node corresponds to one of the predefined regions, and each edge represents an abnormal functional connection between two community members. Blue lines signify reduced functional synchrony in ASD across the paradigm; magenta lines denote increased functional synchrony in ASD. (a) Hyper-active community: $Y_1 = 1$ (b) Hypo-active community: $Y_1 = 0$.

0.122) neurotypical controls. For each subject, a T1-weighted scan (MPRAGE, $TR = 1900$ ms, $TE = 2.96$ ms, flip angle $= 9^\circ$, $res = 1$ mm³) and a task fMRI scan (BOLD, $TR = 2000$ ms, $TE = 25$ ms, flip angle $= 60^\circ$, $res = 3.44 \times 3.44 \times 4$ mm, 164 volumes) were acquired on a Siemens MAGNETOM Trio TIM 3T scanner.

The experimental paradigm features coherent and scrambled point-light animations created from motion capture data. The coherent biological motion depicts an adult male actor performing movements relevant to early childhood experiences [50]. The scrambled animations combine the trajectories of 16

randomly selected points from the coherent displays. Snapshots of these stimuli are illustrated in Fig. 5. Six biological motion clips and six scrambled motion clips were presented in an alternating-block design (24s per block).

We segment the anatomical images into 150 cortical and sub-cortical regions based on the Destrieux atlas in Freesurfer [51]. The fMRI data was preprocessed using FSL [52]. The pipeline was developed at Yale and consists of: 1) motion correction using MCFLIRT, 2) interleaved slice timing correction, 3) BET brain extraction, 4) spatial smoothing with FWHM 5mm, and 5) high-pass temporal filtering. The functional and anatomical

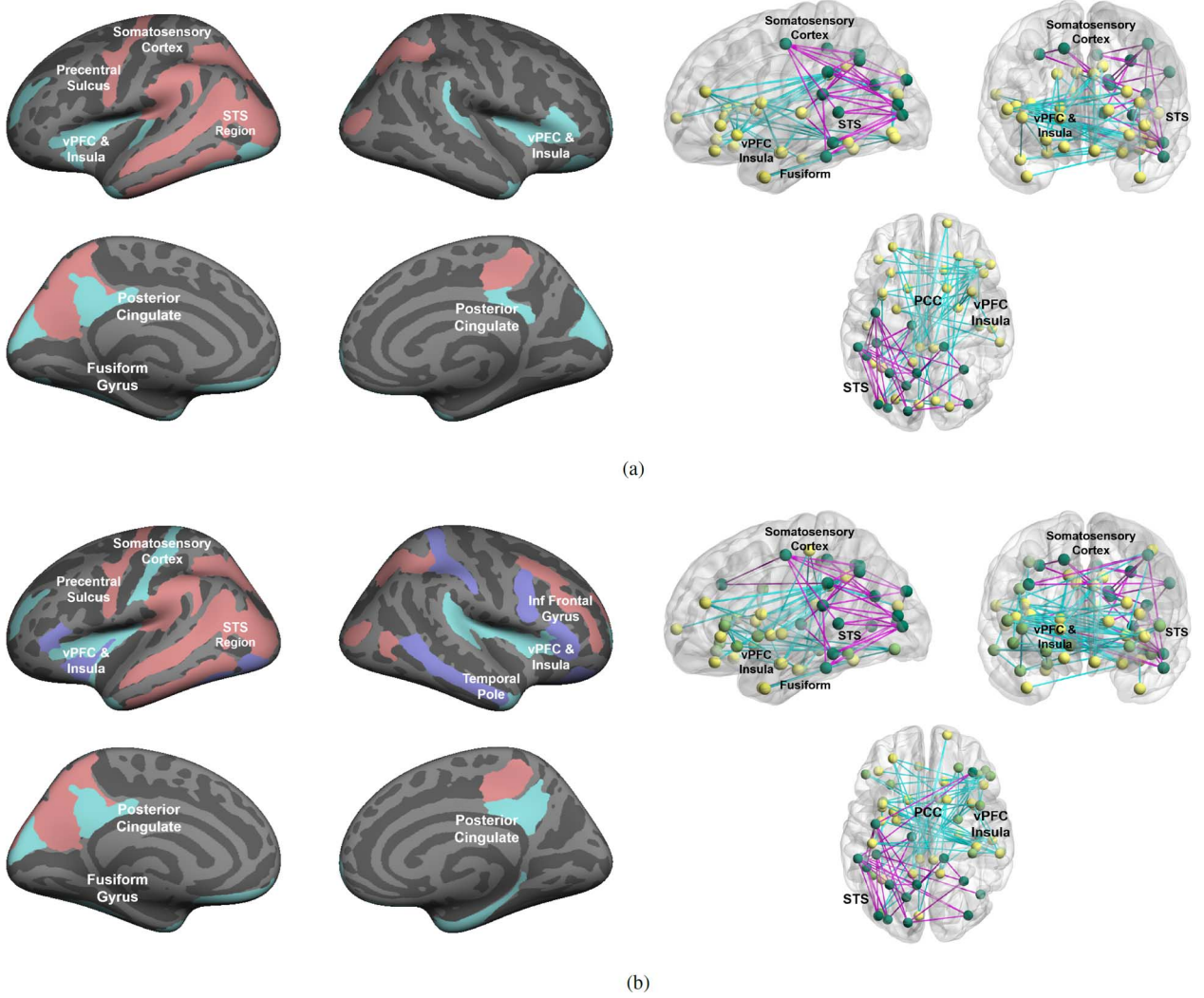


Fig. 7. Abnormal communities inferred by our Bayesian model for $K = 2$ (top) and $K = 3$ (bottom). Left: Region membership in each community. Red nodes indicate the hyper-active community, whereas blue areas delineate the hypo-active cluster. Right: Estimated network of abnormal functional synchrony. Each node corresponds to one of the predefined regions, and each edge represents an abnormal functional connection between two community members. Blue lines signify reduced functional synchrony in ASD across the paradigm; magenta lines denote increased functional synchrony in ASD. (a) One hyper- and one hypo-active community: $Y_1 = 1$ and $Y_2 = 0$ (b) One hyper-active and two hypo-active communities: $Y_1 = 1$ and $Y_2 = Y_3 = 0$.

data were registered to the MNI152 standard brain. The fMRI measure B_{ij}^l is computed as the Pearson correlation coefficient between the mean time courses of regions i and j . As previously stated, we center the correlation distribution of each subject and fix the likelihood mean μ_1 during the model estimation procedure.

B. Abnormal Communities

We vary the number and type of communities Y_k to explore the evolution of abnormal subnetworks. Recall that these subgraphs reflect atypical functional synchrony across the entire stimulus paradigm. Figs. 6 and 7 illustrates the community-based differences inferred by our model for $K = 1$ and $K > 1$, respectively. Table III reports the corresponding ML parameter values estimated by the variational EM algorithm.

Encouragingly, we observe a steady progression with the number of communities K in Figs. 6 and 7. In particular, as K increases, our model augments the previous community organization with an additional subnetwork of the required

type. Similarly, the functional prior π^f and data likelihood parameters $\{\mu, \sigma^2\}$ in Table III are identical between model realizations. Differences in community organization seem to be captured by the region prior π^r , the edge concentration η , and the noise level ϵ . This behavior indicates stability of the inferred latent functional templates, which are subsequently used to identify the abnormal communities. Further, these nested solutions provide empirical evidence for the validity of our model and inference algorithm.

Clinically, the hyper-active communities are depicted in red/pink and concentrate in the left superior temporal sulcus (STS), the visual cortex, and the somatosensory cortex. Conversely, the blue regions indicate hypo-activity. As seen, they localize to the bilateral ventral prefrontal cortex (vPFC), insula, and posterior cingulate. The hypo-active networks also includes subcortical activations in the caudate and amygdala, which are not shown on the cortical surface plots. The corresponding network diagrams [53] reveal increased synchrony between temporal and occipital nodes (magenta lines), along with reduced synchrony

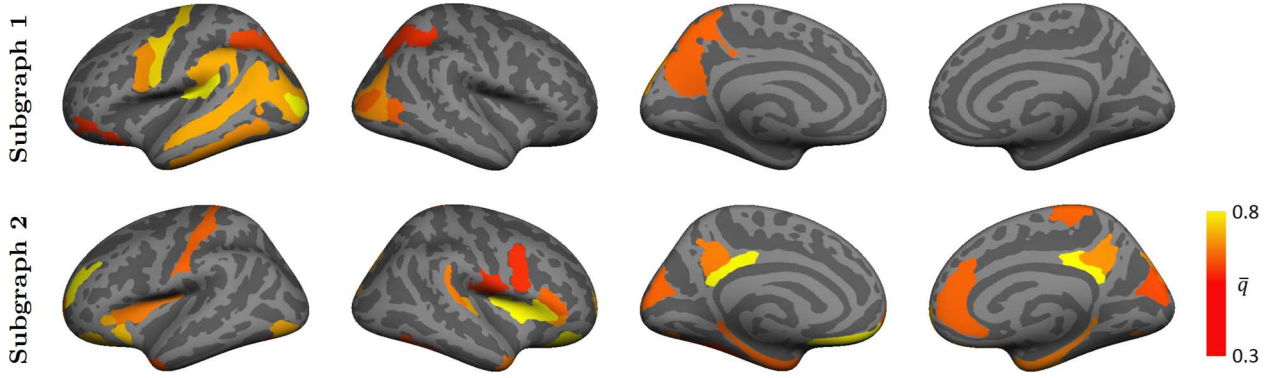


Fig. 8. Average marginal posterior probability $\bar{u}_i$ for each of the $K = 2$ communities across 50 random samplings of the fMRI data. Each sample includes 80% of the subjects, such that the ratio of ASD patients to neurotypical controls is preserved. The colorbar denotes the average posterior probability $\bar{u}_{ik}$ for each region i and community k .

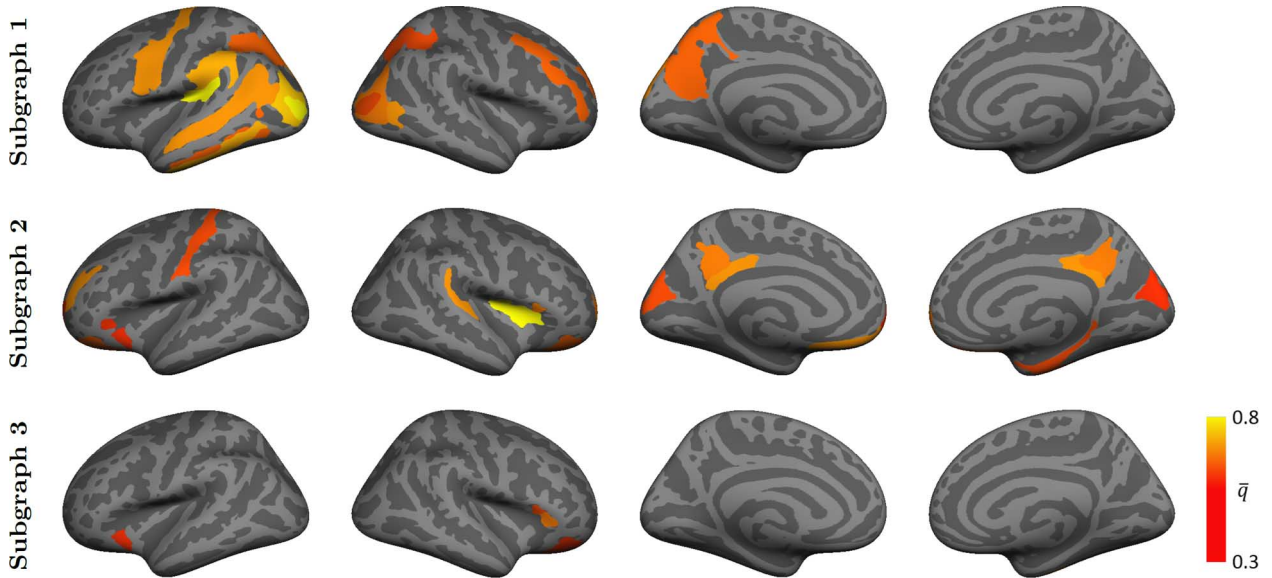


Fig. 9. Average marginal posterior probability $\bar{u}_i$ for each of the $K = 3$ communities across 50 random samplings of the fMRI data. Each sample includes 80% of the subjects, such that the ratio of ASD patients to neurotypical controls is preserved. The colorbar denotes the average posterior probability $\bar{u}_{ik}$ for each region i and community k . The clusters are aligned to the final solution in Fig. 7 to maintain correspondence of Y_k .

between the frontal and parietal nodes (blue lines). These results support a view of impaired social communication in ASD. Consistent with neurocognitive findings, Figs. 6 and 7 suggest hyper-connectivity, and perhaps hyper-functionality, in a visual perception network. Hypo-active circuits include two networks that are well known for their role in social cognition and the high-level interpretation of social stimuli [6], [7], [4].

C. Model Robustness

Figs. 8 and 9 report the average posterior probability $Q_i^r(\cdot)$ of each region across 50 resamplings of the data for $K = 2$ and 3. We display only the regions for which $\bar{u}_{ik} > 0.3$, thereby emphasizing the most prominent patterns.

Favorably, our bootstrapping analysis in Fig. 8 consistently recovers both the red hyper-active and blue hypo-active communities in Fig. 7(a). Additionally, subgraph 2 in Fig. 8 uncovers medial activations in the right hemisphere that are

not present in the full solution. This behavior suggests that while some regions are weakly implicated by the data, they are not included in the full variational EM solution. Despite the inconsistency, reproducibility across 80% subsets has not been demonstrated in prior fMRI studies of ASD. Instead, prior works employ leave-one-out techniques, which artificially impose a stronger bias between samples.

The $K = 3$ model in Fig. 9 is able to recover the hyper-active community (Subgraph 1) with the same frequency as in Fig. 8. However, the reproducibility of the two hypo-active networks is much lower. Empirically, we observe that the nodes are shuffled between communities during the bootstrapping experiment, which reduces the overall consistency between trials. At the same time, highly-implicated regions, such as the posterior cingulate and insula, manage to survive our selection criterion. We note that the weakened robustness for $K > 2$ provides an important direction for future refinements to our model and inference algorithm.

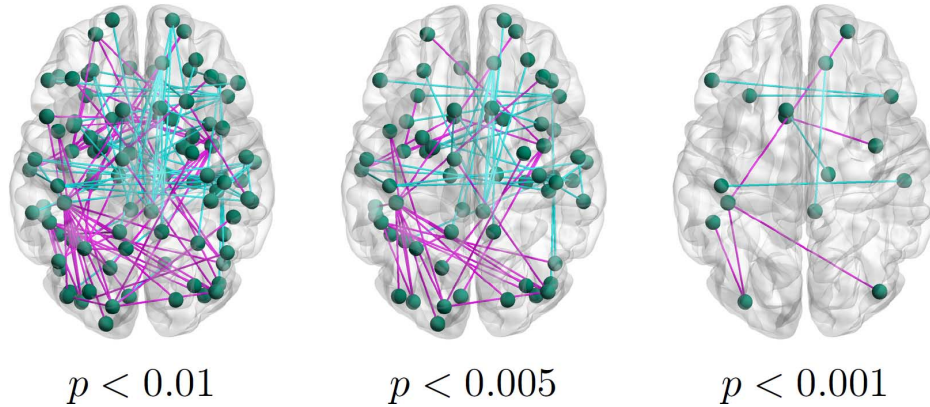


Fig. 10. Population differences in functional connectivity when varying the significance threshold. Blue lines denote a lower average correlation across ASD patient than for neurotypical controls. Magenta lines signify the opposite relationship.

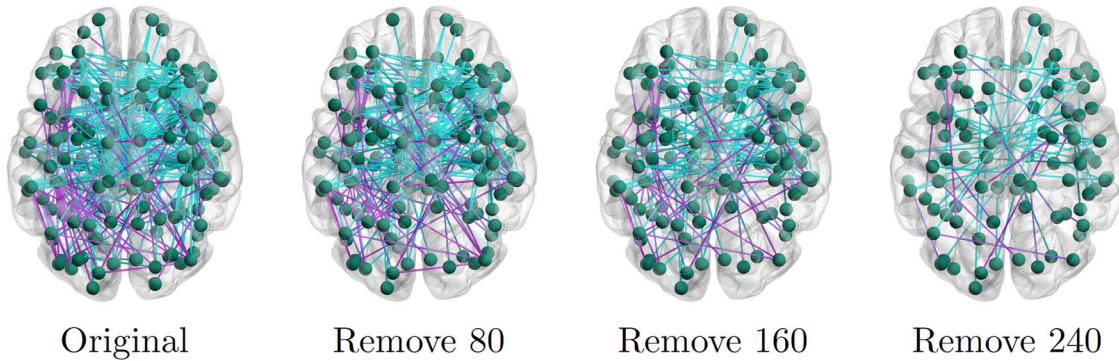


Fig. 11. Community detection based on current-flow betweenness. Left to right depicts the original network of pairwise functional differences (294 edges) and the iterative removal of 80, 160 and 240 edges. Blue and magenta lines correspond to reduced and increased functional synchrony in ASD, respectively.

D. Baseline Comparison

Figs. 10 and 11 illustrate the connectivity differences based on the two sample t-tests and the current-flow betweenness algorithm, respectively. For ease of visualization, we have colored each pairwise connection such that blue corresponds to reduced synchrony in ASD, and magenta denotes increased synchrony in ASD. However, we do not differentiate between heightened and reduced connections when applying the baseline methods. Hence, these results underscore the importance of our multi-cluster region labels and the community types Y_k .

In Fig. 10 we threshold the (uncorrected) p-value to enforce different levels of statistical significance. As an aside, none of the connections survive a false discovery rate (FDR) correction for multiple comparisons, which suggests that the univariate functional differences are relatively weak in our dataset. Similar to Fig. 7, we observe interconnected magenta lines in STS region and a high concentration of blue lines in the default mode and precentral cortices for $p < 0.01$. There is also a second magenta cluster in the left frontal regions. Reducing the threshold to $p < 0.005$ prunes some of the connections; however, the frontal areas no longer exhibit a clique-like organization. Finally, the result for $p < 0.001$ contains isolated groups of vertices rather than a unified network.

The community detection procedure was applied to an original network of 294 edges, which corresponds to connection-

wise latent functional differences, as computed via [42]. We notice that the initial graph is a superset of the connections identified by our model in Figs. 6 and 7. However, pruning the edges according to the current-flow betweenness measure does not reveal interconnected communities. For example, despite removing 80% of the edges in the rightmost figure, the overall number of nodes is fairly similar to the original network. Consequently, we cannot identify a natural stopping point for the algorithm. The observed behavior suggests that pairwise betweenness measures are insufficient to distinguish the core subnetworks associated with ASD.

E. Neurosynth Results

The polar plots in Fig. 12 illustrates the top nine meta-analytic constructs implied by the regions in each community, as inferred by our generative model. Recall that these topics reflect unbiased, aggregate findings across the fMRI literature. The functional dimensions in Fig. 12 expose a clear functional distinction between the two community types. Specifically, the hyper-active network maps onto visual perception and pattern recognition domains for both $K = 2$ and $K = 3$. Correspondingly, individuals with ASD tend to outperform their neurotypical peers in such tasks [54]. In contrast, the hypo-active networks implicate areas of clear deficit in ASD, such as emotion regulation and social communication, perception and reward. There is also considerable overlap in the functionality of the blue and green

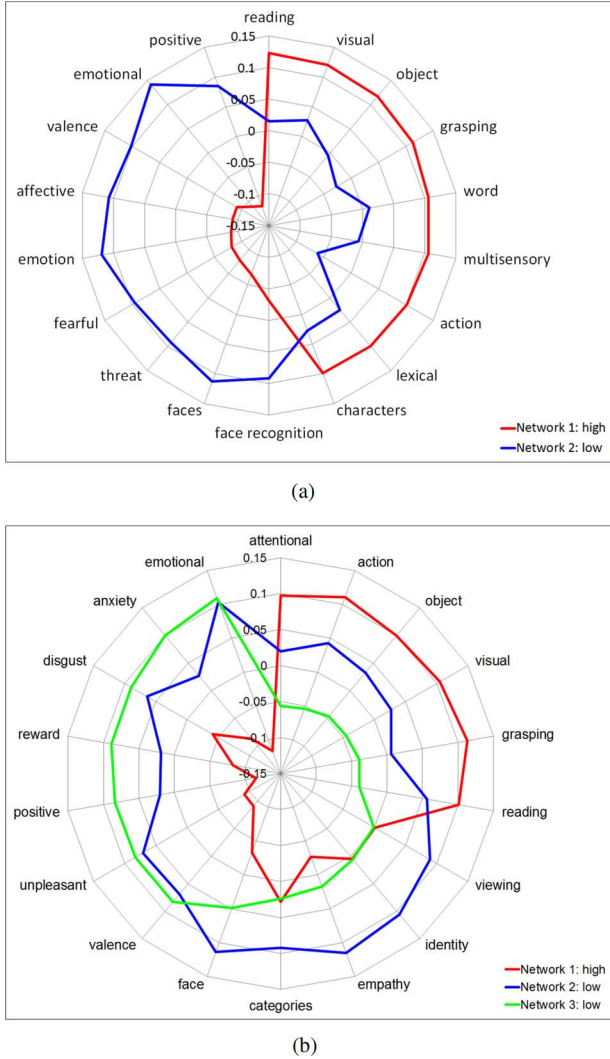


Fig. 12. Correlation values of the top 9 features for each network based on the Neurosynth database. These topics represent the specificity of neurocognitive functions derived from meta-analytic decoding. Red corresponds to the hyper-active community, and blue/green denotes the hypo-active subgraphs. These networks map onto functional domains that highlight both the relative strengths (basic visual processing) and vulnerabilities (interpreting social cues and emotional processing) of ASD, respectively. (a) $Y_1 = 1$ and $Y_2 = 0$ (b) $Y_1 = 1$ and $Y_2 = Y_3 = 0$.

networks in Fig. 12(b), likely driven by community membership of the bilateral STS, vPFC and insulae. At a high level, our Neurosynth results support the hypothesis that ASD children are more stimulated by the point-light animations but are failing to grasp the social context. This conclusion fits nicely both with behavioral observations when administering the fMRI paradigm, and with a prevailing theory that ASD includes areas of cognitive strengths amidst the social deficits [37], [55], [56].

VII. DISCUSSION

Community detection procedures enable us to identify important subcomponents of a complex network. As such, we have presented a unified Bayesian framework that models both hyper- and hypo-active communities based on group-level functional differences. We formulate a variational EM algorithm for

maximum likelihood estimation of the model parameters. The algorithm simultaneously infers the region labels and the graph of abnormal functional synchrony. Effectively, our approach highlights the core deficits of a given neurological disorder, and it facilitates a biological interpretation of the pathological brain circuitry.

We validate the model on synthetic data, and we apply it to a task fMRI study of social perception in ASD. Unlike conventional fMRI analysis, which identifies separate activation patterns for each category of stimuli, we compute subject-level correlations across the entire paradigm. We have previously mined this dataset for differences between ASD patients and neurotypical controls as a functional of stimulus condition [3]. Therefore, in this work, we are interested in functional subnetworks across both social (coherent motion) and nonsocial (scrambled) processing. We emphasize that our Bayesian framework can just as easily be applied to individual stimulus blocks within a block-design fMRI paradigm.

Figs. 6 and 7 depict the altered subnetworks associated with ASD when varying the number and type of communities. We observe a desirable nesting property, in which the inferred topologies for larger K are supersets of the solutions for smaller K values. At the same time, the latent functional prior and the likelihood parameters remain unchanged in Table III. These results suggest an evolution of hyper- and hypo-active subnetworks, such that additional clusters explain progressively weaker phenomena. Conversely, while the baseline methods detect altered functional edges in the vicinity of our model result, both p -value thresholds and iterative edge removal fail to identify core interconnected communities. Moreover, the distributed pattern in Figs. 10 and 11 are difficult to interpret *post hoc*, since our knowledge of the brain is organized around region properties (e.g., functional localization, tissue composition, cortical thickness) and not the potentially indirect connections between them. Our method overcomes such limitations by explicitly modeling the information flow from regions to connections. The resulting communities have a straightforward interpretation and can be used to design follow-up experiments for ASD.

Our approach reveals two pathways of reduced functional synchrony. Automatic decoding of these circuits by the Neurosynth meta-analytic framework in Fig. 12 implicates high-level topics linked to social and emotional processing. Key elements of the hypo-active communities include the ventral prefrontal cortex (vPFC), the insula and the posterior cingulate. The prefrontal cortex plays a vital role in executive functions, such as working memory, social development and language processing. Intuitively, this area reflects many of the core deficits of ASD and has been implicated in prior studies [39], [57]–[59]. It has also been hypothesized that genes associated with ASD influence the deep layer projection neurons in the prefrontal cortex during development [60]. Exploring the relationship between our hypo-active communities and the underlying genetic factors of ASD remains an important direction for future work. The insula is linked to emotional processing and empathy. It also acts as a hub, which mediates interactions between large-scale cognitive systems [61]. Aberrant striatal connectivity in ASD has been reported in both task and resting-state fMRI studies

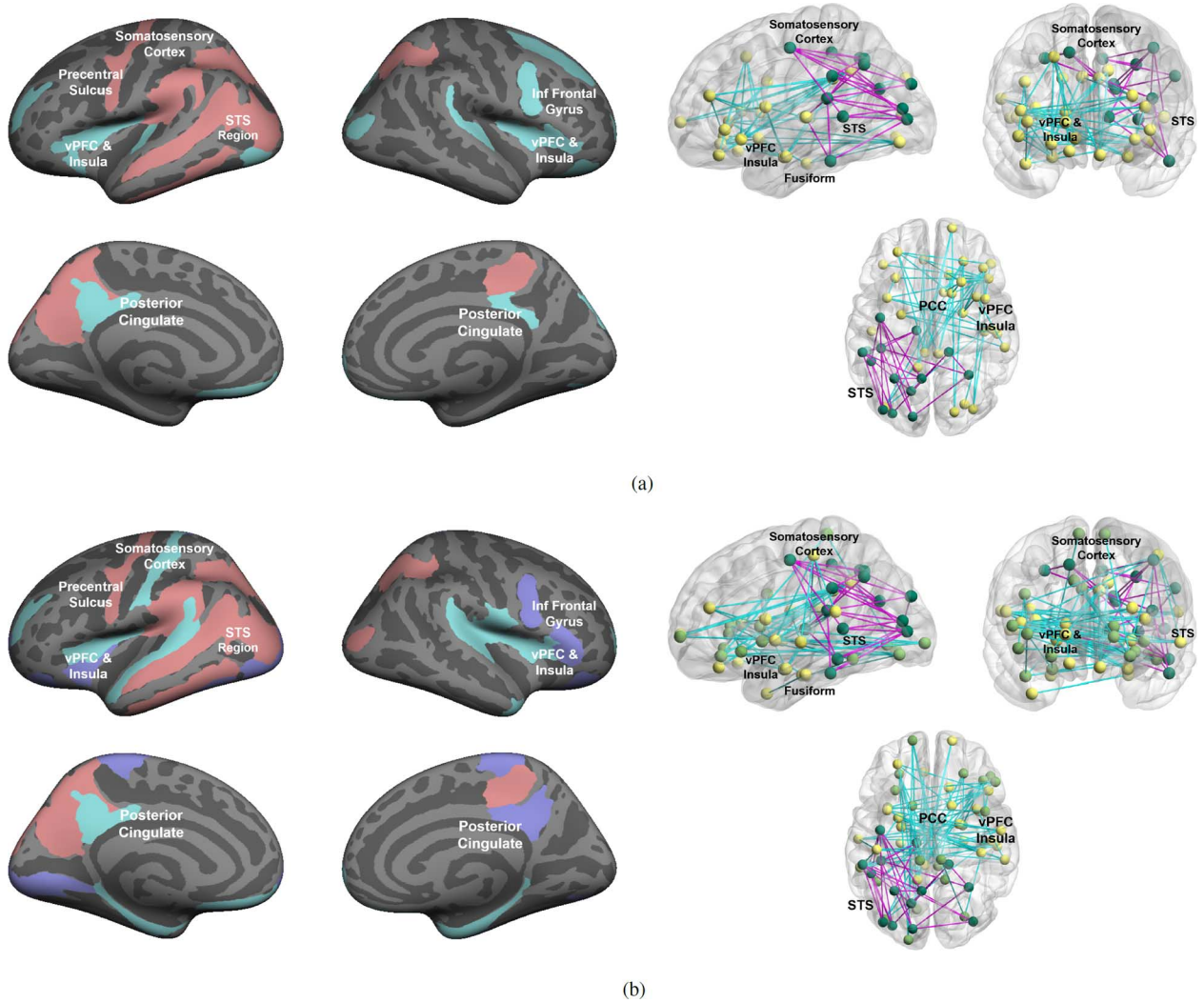


Fig. 13. Abnormal subnetworks inferred by our Bayesian model for $K = 2$ (top) and $K = 3$ (bottom) when estimating the community types Y_k . Left: Region membership in each community. Red nodes indicate the hyper-active community, whereas blue areas delineate the hypo-active cluster. Right: Estimated network of abnormal functional synchrony. Blue lines signify reduced functional synchrony in ASD across the paradigm; magenta lines denote increased functional synchrony in ASD. (a) Model solution for $K = 2$ (b) Model solution for $K = 3$.

[62]–[64]. Finally, the posterior cingulate is a key element of the default mode network. Individuals with ASD exhibit reduced long-range connectivity with the both posterior cingulate and limbic structures [7], [65], [4], which has been linked to higher-order integration processes, such as social, emotional and communication functions [5], [66].

Concurrent with the reduced clusters, our model identifies a hyper-active community for which the pairwise functional synchrony is greater in ASD patients, relative to neurotypical controls. Most studies and theories of ASD have focused on reductions in functional synchrony; our results are generally consistent with these prior findings and demonstrate system-specific (not globally) reduced coherence in circuits known to contribute to emotion regulation and social cognition/perception. However, interestingly, the pathways that exhibit over-connectivity in ASD involve reading the printed word, object recognition, visual memory, and visual processing. Strikingly, these are the very areas of functioning that, on average, are preserved or are even relative strengths of individuals with ASD [37], [54]–[56].

The main presentation of our Bayesian framework requires the user to specify the number and types of the abnormal communities via the binary vector $Y = [Y_1 \dots Y_K]^T$. For comparison, we implement the variational EM extension in the Supplementary Materials to infer the variables Y_k given just the total number of clusters K . Fig. 13 illustrates the model results when extracting $K = 2$ and $K = 3$ subgraphs. Encouragingly, the communities for $K = 2$ clusters are nearly identical to Fig. 7(a), and likewise for the $K = 3$ hyper-active network. The hypo-active clusters for $K = 3$ jointly contain similar regions when fixing and estimating Y_k , but the assignments between the two communities vary slightly. While the above results are promising, empirically, the reproducibility across bootstrapped samples is lower when estimating Y_k . Moreover, the algorithm occasionally returns degenerate solutions with one or more empty clusters. This behavior is highly dependent on initialization, particularly for $K = 3$. Unfortunately, we were unable to design an initialization scheme that completely avoids degenerate solutions without making strong *a priori* assump-

tions about the network organization. The robustness issues can be partially attributed to inter-subject variability and possibly weak differentiation between communities of the same type. Therefore, we recommend that the user start by fixing Y , which allows for greater control over the results.

We acknowledge that setting the community types *a priori* may be difficult in certain applications, and that this requirement is a potential drawback of our method. However, ASD is clinically described both in terms of deficits, which usually map onto hypo-activation and hyper-connectivity [5], [6], and in terms of isolated cognitive improvements, which often relate to hyper-activation and hyper-connectivity [54], [55]. This binary categorization of ASD fits nicely into our Bayesian framework. Moreover, we can formulate an initial guess for the number of hyper- and hypo-active communities based on existing theories of ASD. We also observe that many clustering techniques use model selection procedures, such as the Bayesian Information Criterion (BIC), to determine the appropriate number of clusters K . If desired, BIC and similar metrics can also be applied to our method to select the optimal Y_k .

Anatomical connectivity based on Diffusion Weighted Imaging (DWI) provides complimentary information about pairwise region interactions. Our prior work [32], [33] describes a straightforward approach to incorporating anatomical data into the Bayesian model. In the future, we might constrain the abnormal functional edges between community members according to the underlying white matter connections. This extension would focus the results on salient pathways that are consistent with the structural organization of the brain.

Finally, we recognize many areas of improvement for our model, particularly due to its simplicity. For example, we assume disjoint hyper- and hypo-active communities, which do not capture overlapping processes. In the case of ASD, the social and language circuitry in the brain contain many of the same regions. Future refinements to the model may draw on the matrix factorization and dictionary learning literatures. Another area of exploration is to leverage prior clinical knowledge. Such methods would enable us to identify more focused networks that pertain to a specific cognitive process. Nonetheless, the current modeling choices are deliberate on our part. Specifically, despite advancements in the field, the effects of ASD and of other cognitive disorders remain largely unknown. Therefore, we have formulated simplistic relationships between the region labels and latent functional synchrony templates. We also reduce the number of nonrandom parameters to avoid overfitting the data.

VIII. CONCLUSION

We have demonstrated that a unified Bayesian framework for abnormal community detection supports a binary view of ASD. Specifically, we observe hyper-synchrony between key visual processing areas of the brain, which corresponds to a relative strength (visual perception) in autistic individuals. Concurrently, we detect hypo-synchrony within social information processing and social motivation networks; this finding relates to the hallmark socioemotional impairments of ASD [6]. Unlike prior methods, we explicitly model *differences* in the global

functional organization of the brain, from latent region properties to the observed fMRI measures. We use bootstrapping to verify the robustness of our region assignments. Subsequent decoding via the Neurosynth meta-analytic database confirms the clinical validity of our results within the fMRI literature.

REFERENCES

- [1] K. Friston *et al.*, "Statistical parametric maps in functional imaging: A general linear approach," *Human Brain Map.*, vol. 2, pp. 189–210, 1995.
- [2] K. Friston, P. Jezzard, and R. Turner, "Analysis of functional mri time-series," *Human Brain Map.*, vol. 1, pp. 151–174, 1994.
- [3] M. Kaiser *et al.*, "Neural signatures of autism," *Proc. Nat. Acad. Sci. USA*, vol. 107, no. 49, pp. 21223–21228, 2010.
- [4] H. Koshino *et al.*, "Functional connectivity in an fmri working memory task in high-functioning autism," *Neuroimage*, vol. 24, pp. 810–821, 2005.
- [5] G. Rippon, J. Brock, C. Brown, and J. Boucher, "Disordered connectivity in the autistic brain: Challenges for the "new psychophysiology,"" *Int. J. Psychophysiol.*, vol. 63, pp. 164–172, 2007.
- [6] K. Pelphrey, D. Y.-J. Yang, and J. McPartland, "Building a social neuroscience of autism spectrum disorder," *Curr. Topics Behav. Neurosci.*, vol. 16, pp. 215–233, 2014.
- [7] M. Just, T. Keller, V. Malave, R. Kana, and S. Varma, "Autism as a neural systems disorder: A theory of frontal-posterior underconnectivity," *Neurosci. Biobehav. Rev.*, vol. 36, pp. 1292–1313, 2012.
- [8] R. Tandon and S. Sra, "Sparse nonnegative matrix approximation: New formulations and algorithms," *Max Planck Inst. Biol. Cybern., Tech. Rep.*, 2010.
- [9] O. Sporns, D. Chialvo, M. Kaiser, and C. Hilgetag, "Organization, development and function of complex brain networks," *TRENDS Cognitive Sci.*, vol. 8, pp. 418–425, 2004.
- [10] E. Bullmore and O. Sporns, "Complex brain networks: Graph theoretical analysis of structural and functional systems," *Nature Rev.*, vol. 10, pp. 186–198, 2009.
- [11] M. Rubinov and O. Sporns, "Complex network measure of brain connectivity: Uses and interpretation," *NeuroImage*, vol. 52, pp. 1059–1069, 2010.
- [12] D. Bassett and E. Bullmore, "Small-world brain networks," *Clin. Neurol.*, vol. 12, pp. 512–523, 2006.
- [13] Y. Liu *et al.*, "Disrupted small-world networks in schizophrenia," *Brain*, vol. 131, pp. 945–961, 2008.
- [14] E. Sanz-Arigata *et al.*, "Loss of small-world networks in alzheimer's disease: Graph analysis of fMRI resting-state functional connectivity," *PLOS One*, vol. 5, p. E13788, 2010.
- [15] S. Smith *et al.*, "Network modelling methods for fMRI," *NeuroImage*, vol. 54, pp. 875–891, 2011.
- [16] J. Mumford and J. Ramsey, "Bayesian networks for fMRI: A primer," *NeuroImage*, vol. 86, pp. 573–582, 2014.
- [17] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, pp. 75–174, 2010.
- [18] A. Lancichinetti and S. Fortunato, "Community detection algorithms: a comparative analysis," *Phys. Rev.*, vol. 80, pp. 1–12, 2010.
- [19] B. Everitt, *Cluster Analysis*. New York: Wiley, 1974.
- [20] W. Donath and A. Hoffman, "Lower bounds for the partitioning of graphs," *IBM J. Res. Develop.*, vol. 17, pp. 420–425, 1973.
- [21] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
- [22] M. Newman, "The structure and function of complex networks," *SIAM Rev.*, vol. 45, pp. 167–256, 2003.
- [23] M. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev.*, vol. 69, p. ArXiv:cond-mat/0 308 217, 2004.
- [24] M. Girvan and M. Newman, "Community structure in social and biological networks," *Proc. Nat. Acad. Sci.*, vol. 99, pp. 7821–7826, 2002.
- [25] R. Guimera and L. Amaral, "Functional cartography in complex metabolic networks," *Nature*, vol. 433, pp. 895–900, 2005.
- [26] V. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech.*, vol. ArXiv:0803.0476, pp. 1–6, 2008.
- [27] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, "Defining and identifying communities in networks," *Proc. Nat. Acad. Sci.*, vol. 101, pp. 2658–2663, 2004.

- [28] M. Rosvall and C. Bergstrom, "An information-theoretic framework for resolving community structure in complex networks," *Proc. Nat. Acad. Sci.*, vol. 104, pp. 7327–7331, 2007.
- [29] A. Alexander-Bloch *et al.*, "The discovery of population differences in network community structure: New methods and applications to brain functional networks in schizophrenia," *NeuroImage*, vol. 59, pp. 3889–3900, 2012.
- [30] J. Mumford *et al.*, "Detecting network modules in fMRI time series: A weighted network analysis approach," *NeuroImage*, vol. 52, pp. 1465–1476, 2010.
- [31] A. Lord, D. Horn, M. Breakspear, and M. Walter, "Changes in community structure of resting state functional connectivity in unipolar depression," *PLoS One*, vol. 7, p. E41282, 2012.
- [32] A. Venkataraman, M. Kubicki, and P. Golland, "From brain connectivity models to identifying foci of a neurological disorder," in *Proc. MICCAI*, 2012, pp. 697–704.
- [33] A. Venkataraman, M. Kubicki, and P. Golland, "From brain connectivity models to region labels: Identifying foci of a neurological disorder," *IEEE Trans. Med. Imag.*, vol. 32, no. 11, pp. 2078–2098, Nov. 2013.
- [34] A. Venkataraman, J. Duncan, D. Yang, and K. Pelphrey, "An unbiased bayesian approach to functional connectomics implicates social-communication networks in autism," *NeuroImage Clin.*, vol. 8, pp. 356–366, 2015.
- [35] A. Venkataraman, D. Yang, K. Pelphrey, and J. Duncan, "Community detection in the space of functional abnormalities reveals both heightened and reduced brain synchrony in autism," in *Bayesian Graph. Models Biomed. Imag.*, 2015, LNCS.
- [36] *American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders*. Arlington, VA: Am. Psychiatric, 2013.
- [37] C. Robertson, D. Kravitz, J. Freyberg, S. Baron-Cohen, and C. Baker, "Tunnel vision: Sharper gradient of spatial attention in autism," *J. Neurosci.*, vol. 33, pp. 6776–6781, 2013.
- [38] R. Melillo and G. Leisman, "Autistic spectrum disorders as functional disconnection syndrome," *Rev. Neurosci.*, vol. 20, pp. 111–131, 2011.
- [39] E. Courchesne and K. Pierce, "Why the frontal cortex in autism might be talking only to itself: Local over-connectivity but long-distance disconnection," *Curr. Op. Neurobiol.*, vol. 15, pp. 225–230, 2005.
- [40] V. Cherkassky, R. Kana, T. Keller, and M. Just, "Functional connectivity in a baseline resting-state network in autism," *NeuroReport*, vol. 17, pp. 1687–1690, 2006.
- [41] J. Paakki *et al.*, "Alterations in regional homogeneity of resting-state brain activity in autism spectrum disorders," *Brain Res.*, vol. 1321, pp. 169–179, 2010.
- [42] 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.
- [43] 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.
- [44] A. Venkataraman, T. Whitford, C.-F. Westin, P. Golland, and M. Kubicki, "Whole brain resting state functional connectivity abnormalities in schizophrenia," *Schizophrenia Res.*, vol. 139, pp. 7–12, 2012.
- [45] D. Scheinost *et al.*, "The intrinsic connectivity distribution: A novel contrast measure reflecting voxel level functional connectivity," *NeuroImage*, vol. 62, pp. 1510–1519, 2012.
- [46] T. Yarkoni, R. Poldrack, T. E. Nichols, D. V. Essen, and T. Wager, "Large-scale automated synthesis of human functional neuroimaging data," *Nature Methods*, vol. 8, pp. 665–670, 2011.
- [47] D. Blei, A. Ng, and M. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.
- [48] T. Yarkoni, D. Barch, J. Gray, T. Conturo, and T. Braver, "Bold correlates of trial-by-trial reaction time variability in gray and white matter: A multi-study fMRI analysis," *PLoS One*, vol. 4, p. E4257, 2009.
- [49] A. Condon and R. Karp, "Algorithms for graph partitioning on the planted partition model," *Random Structures Algorithms*, vol. 18, pp. 116–140, 2001.
- [50] A. Klin, D. Lin, P. Gorrindo, G. Ramsay, and W. Jones, "Two-year-olds with autism orient to non-social contingencies rather than biological motion," *Nature*, vol. 459, pp. 257–261, 2009.
- [51] B. Fischl *et al.*, "Sequence-independent segmentation of magnetic resonance images," *NeuroImage*, vol. 23, pp. 69–84, 2004.
- [52] S. Smith *et al.*, "Advances in functional and structural MR image analysis and implementation as FSL," *NeuroImage*, vol. 23, pp. 208–219, 2004.
- [53] M. Xia, J. Wang, and Y. He, "Brainnet viewer: A network visualization tool for human brain connectomics," *PLoS One*, vol. 8, p. E68910, 2013.
- [54] S. Baron-Cohen, *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press, 1995.
- [55] P. Turkeltaub *et al.*, "The neural basis of hyperlexic reading: An fMRI case study," *Neuron*, vol. 41, pp. 11–25, 2004.
- [56] T. Iuculano *et al.*, "Brain organization underlying superior mathematical abilities in children with autism," *Biol. Psychiatry*, vol. 75, pp. 223–230, 2014.
- [57] M. Just, V. Cherkassky, T. Keller, and N. Minshew, "Cortical activation and synch during sentence comprehension in high-functioning autism: Evidence of underconnectivity," *Brain*, vol. 127, pp. 1811–1821, 2004.
- [58] S. Baron-Cohen *et al.*, "Social intelligence in the normal and autistic brain: an fMRI study," *Eur. J. Neurosci.*, vol. 11, pp. 1891–1898, 1999.
- [59] S. Gilbert, G. Bird, R. Brindley, C. Frith, and P. Burgess, "Atypical recruitment of medial prefrontal cortex in autism spectrum disorders: An fMRI study of two executive function tasks," *Neuropsychologia*, vol. 46, pp. 2281–2291, 2008.
- [60] A. Tebbenkamp, A. Willsey, M. State, and N. Sestan, "The developmental transcriptome of the human brain: Implications for neurodevelopmental disorders," *Curr. Op. Neurol.*, vol. 27, pp. 1–8, 2014.
- [61] L. Uddin and V. Menona, "The anterior insula in autism: Under-connected and under-examined," *Neurosci. Biobehav. Rev.*, vol. 33, pp. 1198–1203, 2009.
- [62] M. Dapretto *et al.*, "Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorders," *Nature Neurosci.*, vol. 9, pp. 28–30, 2005.
- [63] G. Silani *et al.*, "Levels of emotional awareness and autism: An fMRI study," *Social Neurosci.*, vol. 3, pp. 97–112, 2008.
- [64] A. D. Martino *et al.*, "Aberrant striatal functional connectivity in children with autism," *Biol. Psychiatry*, vol. 69, pp. 847–856, 2011.
- [65] J. Wolff *et al.*, "Differences in white matter fiber tract development from 6 to 24 months in infants with autism," *Am. J. Psych.*, vol. 169, pp. 589–600, 2012.
- [66] L. Oberman and V. Ramachandran, "Preliminary evidence for deficits in multisensory integration in autism spectrum disorders: The mirror neuron hypothesis," *Social Neurosci.*, vol. 3, pp. 348–355, 2008.