

A Hierarchical Latent Space Network Model for Population Studies of Functional Connectivity

James D. Wilson

jdwilson4@usfca.edu

Department of Mathematics and Statistics

University of San Francisco

San Francisco, CA 94117-1080

Skyler Cranmer

cranmer.12@osu.edu

Department of Political Science

The Ohio State University

Columbus, OH 43210

Zhong-Lin Lu

zhonglin@nyu.edu

Division of Arts and Sciences

NYU Shanghai

1555 Century Avenue, Shanghai, 200122, China

Center for Neural Science and Department of Psychology

New York University

New York, NY 10003

NYU-ECNU Institute of Cognitive Neuroscience at NYU Shanghai

Shanghai, 200062, China

January 22, 2020

Abstract

A major challenge in network neuroscience lies in understanding the organizational principles of the brain at different spatial scales. The brain is highly modular, in that brain regions naturally divide into densely connected subnetworks, which often themselves contain densely connected subnetworks. Modeling these complex hierarchies is a major technical challenge currently inhibiting progress in the field. We develop the hierarchical latent space model (HLSM) that can capture hierarchy at both the individual and population levels, account for multiple predictors of functional connectivity, and account for individual heterogeneity that manifests over a population. We apply several specifications of our model to healthy and paranoid schizophrenia patients collected from the Center for Biomedical Research Excellence project. We find that for both healthy and patient groups, the spatial location of two regions, in hemisphere and functional subnetwork, strongly influence their propensity to connect. We also find that alone, the spatial distance between two regions is significantly and inversely related to their connection probability, but that it is no longer significant once hemisphere and subnetwork locations have been controlled for. The HLSM also identifies increased heterogeneity in the connectivity of the healthy individuals over the patient group, suggesting a difference in overall connectivity patterns between the two populations.

Keywords: latent space model, Markov chain Monte Carlo, stochastic block model, Bayesian hierarchical modeling

1 Introduction

Network analyses of functional connectivity of the brain have revealed that the brain demonstrates a hierarchical structure at both the individual and population levels. In particular, the brain demonstrates (i) *individual hierarchy*: the brain itself has been found to consist of densely connected functional subnetworks that are involved in specific cognitive functions [Meunier et al., 2010, Bressler and Menon, 2010]; and (ii) *population hierarchy*: the connectivity of individuals from a population differ according to similar traits and clinical diagnoses. It is well known, for example, that patients with schizophrenia show differences in brain structure (e.g., [de Souza Crippa et al., 2006, Heckers, 2001]), function during tasks (e.g., [Carter et al., 1997]), and cognitive qualities (e.g., [Luck and Gold, 2008, Wykes et al., 2000]), and that these differences have been discovered in network organization [Supekar et al., 2019]. Individual hierarchy has led to the identification of important subnetworks and their function (see Table 1). These subnetworks are consistent with those identified with data-driven clustering and community detection techniques [Power et al., 2011, Yeo et al., 2011] and are often the basis for analysis and prediction of brain illness or injury [Medaglia et al., 2015, Betzel and Bassett, 2016, Sporns, 2011, Bullmore and Sporns, 2009, Bassett and Sporns, 2017]. While much is known about the functional subnetworks of the brain, the principles that govern their organization remain unclear. Initial analyses in functional magnetic resonance imaging (fMRI) suggest both consistencies and differences when investigating network structure at different scales. Both the whole-brain and its functional subnetworks, for example, have been shown to display high clustering. Yet, while the whole-brain has been shown to exhibit a small-world structure, analyses of subnetworks of the brain reveal that locally, the brain displays a segregated highway architecture [Stillman et al., 2017, 2019].

Many of the challenges and seemingly contradictory findings in network neuroscience stem from a lack of statistical methodology able to deal with the complex and dynamic nature of the human brain. Because functional connectivity varies across time, across person, and across cognitive task [Betzel and Bassett, 2016] it is problematic that a majority of network neuroscience strategies consider only a static network representation of the brain. As a result, researchers who are interested in characterizing networks across (for instance) different people either analyze each individual network separately before comparing across analyses, or aggregate the connectivity across people. Such single-layered analyses neglect heterogeneity among individuals as well as their interdependencies (see [Wilson et al., 2017b] for a discussion). Multilayer network representations of the brain enable researchers to fully analyze the relationships within and between networks observed over time, person, and/or task [e.g., [Bassett et al., 2011, 2015].

A second statistical problem exists: Initial attempts at discerning population- and individual-level hierarchies have focused on the analyses of activation motifs like clustering, overall connectivity, and others [Sporns, 2011]. Many techniques analyze single activation motifs in isolation [Rubinov and Sporns, 2010]. Activation motifs, however, are often highly correlated with one another, and thus may have unexpected interactions that are only apparent when considering multiple motifs simultaneously. Second, current techniques are often unable to separate the influence of activation motifs from the influence of anatomical properties. For example, the distance between nodes is often strongly related to connection strength [Honey et al., 2009], and may confound attempts to quantify activation motifs. Finally, most existing techniques provide only a single point estimate of a given statistic, making it difficult to compare that value against chance. These challenges necessitate the use of generative network models that not only jointly characterize the effect of the activation motifs on the structure of the brain, but also provide a measure of statistical certainty to the observed trend. Recent work has called for the use of such generative network models in structural and functional analyses [Betzel et al., 2016, Betzel and Bassett, 2017].

In this manuscript, we introduce and investigate a generative network model, called the hierarchical latent space model (HLSM), that characterizes the hierarchical structure of functional connectivity networks at both the individual and population level. The HLSM is a joint probability distribution that models the generative mechanism of a collection of connectivity networks arising at the group level as a function of observable covariates, which are common across individuals, as well as unobserved latent information that describes individual and population-level hierarchical structures. Our model quantifies both the effect of activation motifs as well as anatomical properties on the generative mechanism describing the brain. We demonstrate the use of the HLSM through its application on functional images from a group of healthy individuals and a group of patients with schizophrenia that was collected from the Center for Biomedical Research Excellence.

Subnetwork	Function
Visual	Vision
Auditory	Hearing
Somatosensory/Motor	Bodily perceptions and motor movement
Subcortical	Thalamus, putamen, and brainstem
Dorsal Attention	Top-down influences on attention
Ventral Attention	Attention switching
Saliency	Directing attention to emotionally relevant external stimuli
Cingulo-Opercular Task Control	Initiating and sustaining task behavior
Fronto-Parietal Task Control	Top-down exertion of control and executive function
Default Mode	Internally generated cognition

Table 1: Brief description of the function of important subnetworks of the brain

Our analysis reveals several important aspects of these groups. First, for both healthy and patient groups, the hemisphere and subnetwork location of two regions strongly influence their propensity to connect. We also find that alone, spatial distance between two regions is significantly and inversely related to their connection probability, but that it is no longer significant once hemisphere and subnetwork locations have been controlled for. Within-network clustering is also identified by the HLSM, and when accounted for, systematically outperform other covariate models when clusters do not confound with the included covariates. Finally, we find increased heterogeneity in the connectivity of the healthy group over the patient group which may explain less resilient connectivity patterns in patients with schizophrenia.

The HLSM allows the researcher to account for the network structure of the brain, including its individual and population hierarchy, in a statistically principled way while also allowing for covariates, multiple effects, weighted edges, dynamic change in the network, and other complexities essential to modeling the human brain. The result is a powerful and stable model that radically expands the space of effects that can be used to model the brain, thus possessing the potential to solve old questions about competing and dynamic effects, and to pose new ones about the role that activation motifs play in human cognition.

1.1 Related Work

Hierarchical Bayesian models have been developed to factorize multi-dimensional model parameters into a series of conditional probabilities that depend on structure of the model [Kruschke, 2014, Woolrich, 2012]. Through the hierarchical structure, we can constrain the inference process using shared components of the model across subnetworks and subjects, reducing the impact of random noise on the estimated parameters. The approach has been used to impose spatial constraints on voxel-level analyses [Bowman et al., 2008] as well as task and subject constraints on ROI analyses in task fMRI research [Molloy et al., 2018, 2019]. Our development in this paper on functional connectivity analyses compliments those earlier developments in task fMRI.

The latent space network model from [Hoff et al., 2002] has since been adapted to networks beyond binary unweighted networks, including networks with clustering [Handcock et al., 2007, Krivitsky et al., 2009], as well as dynamic network sequences [Sarkar and Moore, 2005, Sewell and Chen, 2015, 2016]. Latent space approaches have been formulated for several application areas, including [Gormley and Murphy, 2007], who developed a model for voting data from an Irish election; [Sweet et al., 2013] who proposed a hierarchical LSM for education research. [Smith et al., 2017] provides a detailed review of the properties of different latent spaces themselves, which should help guide future methodological development and novel applications. Our model is most closely related to the latent cluster random effects model (LCREM) from [Krivitsky et al., 2009]. The LCREM incorporates both clustering structure of each vertex as structure (Equation (4)) and random effects for the vertex propensity to connect (Equation (5)). Our hierarchical model extends the LCREM to a population of networks that is made up of distinct subpopulations, or clusters, of networks.

The HLSM simultaneously models the community structure of each individual network, as well as the clustering of networks according to individual similarity. Community detection is a common and powerful exploratory technique that identifies densely connected subregions in a network. Community detection has been well-studied over the past two decades and many detection algorithms have been developed for

static, temporal, and multilayer networks (see for example, Mucha et al. [2010]). Community detection techniques have been widely applied in network neuroscience for both structural and functional connectivity networks [Bassett et al. 2011]. Network clustering, which focuses on clustering populations of networks into subpopulations of similar networks, is a relatively new problem for which research is in its nascent studies. Network clustering techniques have been largely motivated by studies in group-level fMRI [Durante et al. 2017, Mukherjee et al. 2017].

In the last decade, there has been substantial interest in the development of generative network models for structural and functional connectivity of the brain (see Betzel and Bassett [2017] for a recent review). Generative network models seek to characterize the (stochastic) mechanism that gives rise to the biological features of the brain. Examples of recent generative network models developed for functional connectivity networks include the exponential random graph model for unweighted networks [Simpson et al. 2011], which was later extended for groups [Simpson et al. 2012]. The correlation generalized exponential random graph model from [Stillman et al. 2017] was developed to directly model correlation networks arising in fMRI and applied to the default mode network and later to thirteen other well-known subnetworks in [Stillman et al. 2019].

There are three other recently proposed generative network models that model populations of networks, including the random effects stochastic block model [Paul and Chen 2018] and the multi-subject stochastic block model [Pavlovic et al. 2019], as well as the edge-based logistic model from [Simpson and Laurienti 2015, Simpson et al. 2019]. Given the subpopulation label of the individual, these three models assume independence of the edges within and across individuals. Like the HLSM, the random effects and multi-subject stochastic block models both capture individual hierarchy and model the community structure of the regions of each connectivity network. Both of these models characterize individual heterogeneity through possible differences in the community structure across individuals. The block models cannot, however, directly capture other activation motifs in the brain. The logistic model from [Simpson and Laurienti 2015, Simpson et al. 2019] models both activation motifs and individual variability of each individual directly like the HLSM. However, the logistic model assumes edges in the population are independent across pairs given network information, which is a stronger assumption than the conditional independence assumption of the HLSM.

2 A Hierarchical Latent Space Network Model for Population Level Functional Connectivity

We are interested in modeling the generative process of the hierarchical structures that give rise to the whole-brain network of an individual. We aim to develop a model that characterizes an individual while accounting for hierarchies that arise at two levels: (i) the whole-brain as a function of functional subnetworks, and (ii) the brain of an individual as a function of that individual’s subpopulation. We first describe network modeling of group-level fMRI and then describe our latent space model for group functional connectivity.

2.1 Network Modeling of Functional Connectivity

Network neuroscience takes on the perspective that the structural and functional characteristics of the brain manifest in network models of the brain. A common strategy for generating group-level functional connectivity networks, the focus of this paper, is described as follows. A toy demonstration is shown in Figure 1. For each individual, images of the whole brain are collected over time. For example in resting state fMRI, the individual is given no task but to stay awake under the fMRI machine and images are continually taken for several minutes. Each image represents the measured blood oxygen level dependence (BOLD) response across small spatial regions of the brain. Regions of interest (ROIs) are often defined either based on anatomical structures or functional clusters, and the average time series of all the voxels in each region is treated as the time series of BOLD responses. The granularity of the spatial resolution depends on the user and machine; high resolution images, for instance, measure BOLD responses at 2mm^3 cubes called voxels. Images are transformed to a multivariate time series of these BOLD levels, where each region is treated as a time series of BOLD responses. Network representations of these time series are created based on the temporal similarity between each pair of regions. In the final network model, regions are modeled as vertices



Figure 1: Example of generating a network representation for a collection of fMRI images for an individual. Scans are collected over time, mapped to a multivariate time series, and finally represented as a network.

and edges between incident regions represent the similarity of the two regions over time. Edges may be weighted to represent the strength of connection between incident regions based on the similarity metric used [Stillman et al., 2017, 2019], or unweighted to represent statistically significant relationships.

In the setting of group-fMRI, a group of individuals undergo an fMRI session with the same machine and for the same amount of time. Network representations for each individual are generated using the same strategy as described above, and standard preprocessing steps (motion correction, anatomical registration, spatial-temporal smoothing) are used for each individual to ensure that networks are registered across regions. Preprocessing is particularly important for group studies – network models must be appropriately aligned so that comparisons across individuals are valid. We provide details of the preprocessing we used in our application study in Section 5.

2.2 A Hierarchical Latent Space Model

We begin with a collection of L undirected functional connectivity networks $\mathbf{Y} = \{Y_1, Y_2, \dots, Y_L\}$ from a group of individuals, where Y_ℓ is a network (weighted or unweighted) that encodes the functional connectivity for individual ℓ . Suppose that the networks are each constructed on n regions. Then Y_ℓ is a random vector

whose entries $y_{ij,\ell} \in \mathcal{Y}$ are the edge weights that describe the connection strengths between regions i and j , for $i, j = 1, \dots, n$.

Our proposed model, the hierarchical latent space model, is a probability distribution function that characterizes the strength of a connection between regions i and j as a function of observable covariate information $X = \{x_{ij,\ell} : i < j, \ell = 1, \dots, L\}$ as well as unobserved latent variables for (i) the individual with the vector $W = \{w_\ell, \ell = 1, \dots, L\}$, (ii) for the latent position of region through the vector $Z = \{z_i, i = 1, \dots, n\}$, and (iii) for random effects of each region with the variable $\Delta = \{\delta_i, i = 1, \dots, n\}$. We assume that w_ℓ and $z_i \in \mathbb{R}^d, d \geq 1$, and that $\delta_i \in \mathbb{R}$. We now describe the model in more detail.

The observable covariates, X , capture properties of the whole-brain network at a vertex or dyad level. These features are well-suited to incorporate features that characterize homophily among regions as well as other structural motifs present in the network. For example, one can readily evaluate whether or not the hemisphere location of two regions – left, right, or midline – are positively related to the connection of two regions as expected from previous studies [Pandya and Rosene, 1985]. We assume that these features, and their effects, are common across individuals.

The latent variable w_ℓ quantifies individual random effects of individual ℓ on the overall connectivity of Y_ℓ . The variable z_i quantifies the latent position for region i , which are common across individuals. The variable δ_i models the random effects of region i on the propensity of connectivity for that region. Given the latent variables W , Z and Δ , edges are assumed to be conditionally independent across pairs of regions and across individuals. Therefore, the joint probability distribution of $\mathbf{Y}$ can be decomposed as:

$$\mathbb{P}(\mathbf{Y} \mid W, X, Z) = \prod_{\ell=1}^L \prod_{i < j} \mathbb{P}(y_{ij,\ell} \mid w_\ell, z_i, z_j, \delta_i, \delta_j, x_{ij,\ell}) \quad (1)$$

Suppose that $x_{ij,\ell}$ is a p -dimensional vector of covariates describing the connectivity of regions i and j for individual ℓ . We model Y_ℓ as a function of observed and unobserved effects through a random effects model over the population. Let $\beta \in \mathbb{R}^p$ and let $\mathbb{E}[y_{ij,\ell} \mid w_\ell, z_i, z_j, x_{ij,\ell}]$ denote the conditional expectation of $y_{ij,\ell}$ given its latent positions w_ℓ, z_i, z_j and covariate information $x_{ij,\ell}$. For a specified link function $\phi : \mathcal{Y} \rightarrow \mathbb{R}$, the HLSM relates the edge weight $y_{ij,\ell}$ to w_ℓ, z_i, z_j , and $x_{ij,\ell}$ as a generalized linear model:

$$\phi(\mathbb{E}[y_{ij,\ell} \mid w_\ell, z_i, z_j, x_{ij,\ell}]) = \beta^T x_{ij,\ell} - |z_i - z_j|_d + \delta_i + \delta_j + w_\ell, \quad i < j, \ell = 1, \dots, L. \quad (2)$$

Model (2) fits networks with edge values that are assumed to follow an exponential family probability distribution. Common distributions include the Poisson, binomial, and Gaussian distributions. In this way, model (2) is suitable for networks unweighted networks, for weighted networks with continuous-valued edge weights, as well as weighted networks with discrete-valued edge weights. Choices of ϕ depend on the nature of the connectivity network under consideration. In the simplest case when Y_ℓ is assumed to have binary edge weights (i.e., when each connectivity matrix is unweighted), one can specify ϕ as the logit link function. In this case, model (2) simplifies to:

$$\text{logit}(\mathbb{P}(y_{ij,\ell} \mid w_\ell, z_i, z_j, x_{ij,\ell})) = \beta^T x_{ij,\ell} - |z_i - z_j|_d + \delta_i + \delta_j + w_\ell, \quad i < j, \ell = 1, \dots, L.$$

Other common choices of link functions include the identity link for continuous-valued edge weights, and the natural log link function for discrete-valued edge weights. See [Sewell and Chen, 2016] for a recent discussion of link functions for weighted networks. Model (2) explicitly models the distribution of an edge between each pair of regions across all individuals in the population. It is through the latent variables W and Z that we model both the hierarchy of the whole-brain and the hierarchy of population of individuals being analyzed. We use model-based clustering to quantify the division of the population into subpopulations as well as the clustering of the whole brain into subnetworks. Let K denote the number of subnetworks present in the whole-brain, and let $M \leq L$ denote the number of subpopulations in the group under study. We model the latent variables as Gaussian mixture models:

$$w_\ell \sim \sum_{m=1}^M \lambda_m \text{MVN}_d(\mu_m, \sigma_m^2 \mathbb{I}_d), \quad \ell = 1, \dots, L; \quad (3)$$

$$z_i \sim \sum_{k=1}^K \gamma_k \text{MVN}_d(\nu_k, \tau_k^2 \mathbb{I}_d), \quad i = 1, \dots, n. \quad (4)$$

In model (3), λ_m specifies the probability that individual ℓ belongs to the m th subpopulation. The parameter γ_k from (4) is the probability that region i belongs to the k th subnetwork. In this way, model (3) represents clustering of the individuals into subpopulations; whereas, model (4) represents clustering of the network into subnetworks.

The parameters μ_m and σ_m^2 characterize the mean and variance of the overall propensity of connection in the m th subpopulation. Model (3) enables the direct quantification of individual variability in functional connectivity. The choice of M is flexible, and guides model granularity at the individual level. When $M = L$, estimates of the variability of each individual will be provided through σ_m^2 . In cases when subsamples of the group under study have differing clinical illnesses, one may instead wish to set M equal to the number of diagnoses in the group to capture the overall variability of each of the subpopulations. The parameters ν_k and τ_k^2 from model (4) characterize the mean and variance of the latent position of the k th subnetwork. Here, K generally represents the number of functional subnetworks in the parcellation under study though again the choice of this number is flexible with the extreme case of $K = n$ providing mean and variance estimators for each region in the network. Model-based clustering for Gaussian data was introduced in Banfield and Raftery [1993] and has been applied to model-based clustering of networks Handcock et al. [2007].

We incorporate random effects of region i , δ_i , in model (2.2) to capture the degree heterogeneity of each vertex in the population of networks. Degree heterogeneity is an important feature of real networks that manifests in social, biological, and functional connectivity networks. There has been an emphasis on the importance of modeling degree variability in network neuroscience as well as other network analysis tasks like community detection and parametric modeling Wilson et al. [2014, 2017b], Krivitsky et al. [2009], Karrer and Newman [2011]. As these effects are unobserved, we model δ_i as independent and identically distributed Gaussian random variables:

$$\delta_i \stackrel{iid}{\sim} N(0, \sigma_\delta^2), \quad (5)$$

where the parameter σ_δ^2 provides a measure of the overall variation of the degree distribution over the entire population.

Further discussion of the differences between the logistic model from Simpson and Laurienti [2015] and Simpson et al. [2019] and the HLSM is warranted. The original specification of the logistic model from Simpson and Laurienti [2015] is similar to the HLSM in that both models quantify the connection probability of two regions as a function of activation motifs and latent variables that represent individual heterogeneity. Furthermore, both models directly assess the effect of dyad-level covariates on connection probability. These two models differ, however, in their generalization to weighted networks as well as the inclusion of node-based network features like the clustering coefficient, and degree in the model. The logistic model can be used to directly model correlation networks via a Fisher transform on the observed logistic values. The HLSM, on the other hand, can model any weighted network from which the edge weights follow an exponential family distribution. Thus, the HLSM and logistic model are complementary. The HLSM also implicitly accounts for these features through the inclusion of the latent position of each region, represented by Z . The logistic model, on the other hand, explicitly models the effects of these terms at the population and individual level using node-level covariates. Furthermore, the HLSM directly estimates and accounts for the community structure of a connectivity network using a Gaussian mixture model. The logistic model, on the other hand, cannot directly estimate community structure in a network, though it can incorporate model terms that are functions of known communities. In practice, the logistic model from Simpson and Laurienti [2015] and Simpson et al. [2019] enables the direct quantification of the effect of node-level summaries; whereas, the HLSM implicitly captures these features through modeling latent positions and directly estimates community structure in functional connectivity networks.

Our proposed hierarchical model is a generalization of the latent space model (LSM), which was first introduced in Hoff et al. [2002]. The basic idea behind the LSM is that network dependencies can be viewed as the proximity of vertices in some unobserved social, physical, or other latent space. Vertices that are nearer to one another in latent space will be more likely to tie. In social relations, the intuition is easy to see: everyone is part of society's social fabric and those nearer to each other in that fabric are more likely to form

friendships, business relationships, romantic partnerships and other social interactions. In the context of functional connectivity networks, this latent space is not only physical, but rather it represents proximity in functional space, which may be correlated to, for example the physical distance or white matter connectivity of two regions.

3 Estimation of the Model

Once a model has been specified, fitting model (2.2) to a collection of functional connectivity networks provides estimates of the effects of observed dyad-level features, as well as unobserved individual and subnetwork effects. Fitting the full model from (2.2) – including clustering, individual and population hierarchies and random effects – requires estimating $3M + 3K + p + 2$ parameters summarized in Table 2 below. Here, we provide a schematic for estimation of these parameters.

Parameter(s)	Common across individuals?	Fixed or Latent?	Description
$\{\beta_j : j = 1, \dots, p\}$	Y	F	effect size of each observed covariate
$\{\gamma_k : k = 1, \dots, K\}$	Y	L	γ_k = probability of region belonging to subnetwork k
$\{\nu_k, \tau_k^2 : k = 1, \dots, K\}$	Y	L	means and variances of connection propensity for regional cluster k
σ_δ^2	Y	L	variance of the random effects of degree heterogeneity
$\{\lambda_m : m = 1, \dots, M\}$	N	L	λ_m = probability of individual belonging to subpopulation m
$\{\mu_m, \sigma_m^2 : m = 1, \dots, M\}$	N	L	means and variances of subpopulation connection propensity

Table 2: A summary of the parameters that can be estimated with the HLSM.

Given a collection of connectivity networks $\mathbf{Y}$ and observable covariate matrix X , the estimation of model (2) can be achieved in three steps. Suppose that Θ represents the collection of all parameters in model (2). We describe maximum likelihood estimation of these parameters; however, we note that Bayesian estimation (as in Krivitsky et al. 2009) can also be used. Maximum likelihood estimation of model (2) consists of three key steps, described as follows.

Step 1: First, the non-clustering parameters are estimated by fitting a latent space model, as described in Hoff et al. 2002. To begin, let α_ℓ represent an intercept term of individual ℓ . The parameter α_ℓ will serve as a nuance parameter that will be later used for estimation of the mixture model from (3). Estimators for α_ℓ , β , and δ are obtained by optimizing the joint log-likelihood function given by:

$$\hat{\Theta}_1 = \operatorname{argmax}_{\beta, \delta_i, \delta_j} \left\{ \sum_{\ell=1}^M \sum_{i < j} \log [\phi^{-1}(\beta^T x_{ij, \ell} - |z_i - z_j|_d + \delta_i + \delta_j + \alpha_\ell)] \right\}. \quad (6)$$

The log-likelihood in (6) is convex in the latent distances $|z_i - z_j|$, but not in the latent positions themselves. Thus, to solve (6), one first optimizes the log-likelihood over latent distances using multi-dimensional scaling methods. From these distances, a configuration of latent positions $Z = (z_1, \dots, z_n)$, who are centered at the origin, can be generated. Using this configuration as a starting point, one next solves for β , δ_i , and δ_j using a Metropolis-Hastings rejection algorithm update, as described in Section 3 of Hoff et al. 2002. This results in a configuration of latent positions $\hat{Z} = (\hat{z}_1, \dots, \hat{z}_n)$, and estimated parameters $\{\hat{\beta}, \hat{\alpha}_\ell, \hat{\delta}_i, \hat{\delta}_j : i = 1, \dots, n, \ell = 1, \dots, L\}$.

Step 2: The second step of maximum likelihood estimation consists of identifying the maximum likelihood estimator of the mixture model for the latent positions provided in model (4). Given $\hat{Z}$ and the estimated parameters from Step 1, we now solve

$$\hat{\gamma}_k, \hat{\nu}_k, \hat{\tau}_k^2 = \operatorname{argmax}_{\gamma_k, \nu_k, \tau_k^2} \left\{ \sum_{i=1}^n \sum_{k=1}^K p_{ik} (\log \gamma_k + \log N(z_i | \nu_k, \tau_k^2)) \right\}, \quad (7)$$

where p_{ik} is the probability that z_i belongs to the k th class, and $N(a|b, c)$ is the probability density function of a Normal random variable with mean b and variance c evaluated at a . Maximization of (7) can be done using the Expectation-Maximization algorithm (Dempster et al. 1977). The likelihood function for model (7) does not have a unique local maximum, and the local maximum that is found by the EM algorithm can

depend on the starting values. We use starting values from k-means with $k = K$ clustering of the estimated latent values $\hat{Z}$.

Step 3: The final step of model estimation involves the estimation of the clustering of the individuals according to the mixture model in model (3). Given the estimators from **Step 1** and **Step 2**, we solve:

$$\hat{\lambda}_m, \hat{\mu}_m, \hat{\sigma}_m^2 = \operatorname{argmax}_{\lambda_m, \mu_m, \sigma_m^2} \left\{ \sum_{\ell=1}^L \sum_{m=1}^M r_{\ell m} (\log \lambda_m + \log N(w_\ell | \mu_m, \sigma_m^2)) \right\}, \quad (8)$$

where $r_{\ell m}$ is the probability that w_ℓ belongs to the m th class. As in **Step 2**, the Expectation-Maximization algorithm can be used to provide an approximate solution to this mixture model. We use k-means clustering with $k = M$ on the estimated coefficients $\hat{\alpha}_1, \dots, \hat{\alpha}_L$ to provide starting values for the algorithm.

The estimation procedure outlined above can be readily implemented using the `latentnet` package in the R software language [Krivitsky and Handcock, 2008]. Estimation of the latent space model is both straightforward and numerically stable [Shortreed et al., 2006, Raftery et al., 2012]. The expectation maximization algorithm has been widely applied to Gaussian mixture models; thus, accomplishing Steps 2 and 3 above is straightforward with standard software in, for example, R, Python or Matlab.

4 Analysis of the COBRE Data Set

4.1 Description of the Data

We apply our latent space model to a subset of participants from the dataset contributed to the 1000 connectomes project by the Center for Biomedical Research Excellence (COBRE) (http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html). The original dataset includes 72 patients with schizophrenia and 75 healthy controls (HC). Exclusion criteria for both groups were history of neurological disorder, mental retardation, severe head trauma (more than 5 minutes of unconsciousness), and substance abuse or dependence within the last 12 months. Diagnostic information was collected using the Structured Clinical Interview for the Diagnostic and Statistical Manual, 4th edition [First et al., 2015]. We additionally removed participants who were diagnosed with disorders other than schizophrenia or schizoaffective disorders. The remaining population contained 71 healthy controls. The breakdown of subtypes of patients with schizophrenia were as follows: 24 paranoid type (8 female, mean age = 40 ± 14 , range 19 - 65), 2 disorganized type (0 female, mean age = 44 ± 23 , range 28 - 60), 9 residual (1 female, mean age = 36 ± 15 , range 19 - 57), 4 NOS (2 female, mean age = 26 ± 6 , range 20 - 33), and 3 schizoaffective disorder (0 female, mean age = 39 ± 14 , range 23 - 50).

IQ scores were collected by administering the Wechsler Abbreviated Scale of Intelligence (WASI). All patients in the study were taking some form of antipsychotic medication. The mean and standard deviations on the olanzapine total equivalent dose, as well as the IQ scores for performance, overall, and reading are reported in Table 3.

	IQ Scores			Olanzapine Equivalents
	Performance	Overall	Verbal	
HC	113.74(11.15)	111.40(11.34)	106.52(11.28)	—
disorganized	104.67(11.02)	98.67(19.22)	96.00(24.64)	126.73(150.05)
paranoid	101.92(17.77)	98.17(18.64)	96.97(18.34)	35.59(27.21)
residual	106.33(13.67)	103.33(12.23)	100.22(9.50)	16.20(10.36)
schizoaffective	110.00(23.30)	107.00(13.75)	104.67(3.79)	25.03(21.85)
schizophrenia NOS	104.33(20.13)	94.00(16.37)	85.33(10.26)	63.12(55.88)

Table 3: Mean and standard deviation of IQ scores and total olanzapine equivalents of participants

4.2 Preprocessing

The participants' structural MRI scans were completed using a multi-echo MPRAGE sequence with the following parameters: TR/TE/TI = 2530/[1.64, 3.5, 5.36, 7.22, 9.08]/900 ms, flip angle = 7 degree, FOV =

256x256 mm, Slab thickness = 176 mm, Matrix = 256x256x176, Voxel size = 1x1x1 mm, Number of echos = 5, Pixel bandwidth = 650 Hz, Total scan time = 6 min. Given five echoes the multi-echo MPRAGE’s TR, TI and time to encode partitions are similar to that of a conventional MPRAGE granting similar gray matter, white matter (WM), and cerebrospinal fluid (CSF) contrast. Slices were collected interleaved in the sagittal plane (multi-slice mode, single shot).

Resting state data was collected with single-shot full k-space echo-planar imaging (EPI) with ramp sampling correction using the intercommissural line (AC-PC) as a reference (TR: 2 s, TE: 29 ms, matrix size: 64x64, 32 slices, voxel size: 3x3x4 mm³). Slices were collected ascending in the axial plane (multi-slice mode, series interleaved). Scans were preprocessed using FSL (FMRIB Software Library, www.fmrib.ox.ac.uk/fsl) and FEAT (FMRI Expert Analysis Tool) Version 6.00. Scans were motion corrected using MCFLIRT [Jenkinson and Smith, 2001], non-brain stripping with BET [Smith, 2002], spatial smoothing using a Gaussian kernel of FWHM 6 mm, grand-mean intensity normalization of the entire 4D dataset using a single multiplicative factor, and high-pass temporal filtering (Gaussian-weighted least-squares straight line fitting, sigma = 45 s). Finally, nonlinear registration was completed to Montreal-Neurological Institute space using FNIRT [Jenkinson et al., 2002, Jenkinson and Smith, 2001].

Scans were preprocessed using FSL (FMRIB Software Library, www.fmrib.ox.ac.uk/fsl) and FEAT (FMRI Expert Analysis Tool) Version 6.00. Scans were motion corrected using MCFLIRT [Jenkinson and Smith, 2001], non-brain stripping with BET [Smith, 2002], spatial smoothing using a Gaussian kernel of FWHM 6 mm, grand-mean intensity normalization of the entire 4D dataset using a single multiplicative factor, and high-pass temporal filtering (Gaussian-weighted least-squares straight line fitting, sigma = 45 s). Finally, nonlinear registration was completed to Montreal-Neurological Institute space using FNIRT [Jenkinson et al., 2002, Jenkinson and Smith, 2001].

To construct adjacency matrices for each participant we use the Power atlas [Power et al., 2011], which specifies 264 8mm³ regions of interest (ROIs). Voxel time-series within each ROI were averaged, yielding 264 time-series per participant. From each time series we regressed out 6 motion parameters accounting for head movement, 4 parameters account for CSF, and 4 parameters accounting for WM in order to reduce bias and noise within the data [Ashby, 2011]. Finally, for each participant, we correlated each of the 264 time series with the others, resulting in a 264 x 264 correlation matrix for each participant. We retrieve an unweighted connectivity network for each participant by thresholding the correlation matrix so that edges represent correlations that are above 0.50. We investigate the effect of thresholding the networks on coefficient estimates in the Appendix.

4.3 Model Specification

We apply the hierarchical latent space model to the population of 71 healthy individuals and 24 patients with paranoid type schizophrenia from the COBRE data set. We consider three network-based features in our analysis: *hemisphere location*, *subnetwork location*, and *spatial distance*. The hemisphere location and spatial distance features describe the spatial relationship of the regions in the brain. The hemisphere location feature is a binary dyad-level variable that indicates whether or not a pair of nodes (regions) are located in the same hemisphere of the brain, either the left hemisphere, right hemisphere, or the midline. Given that regions within a hemisphere are typically more connected within rather than across hemisphere [Pandya and Rosene, 1985], we expected this covariate to be significantly positive. The subnetwork location is also a binary dyad-level variable that indicates whether or not two regions are located in the same known subnetwork, where the subnetworks considered are those described in Table 1. The spatial distance covariate describes the Euclidean distance between regions. We expect the spatial distance between two regions to be inversely related to their propensity to interact.

For our demonstration, we fit the following three models for each of the individuals in the study:

- **Cluster model:** We set $d = 2$ and $K = 3$ to identify three clusters that best partition the network. No other features were considered in this model.
- **Spatial distance model:** We included the spatial distance between pairs of nodes as well as random effects to model degree heterogeneity.

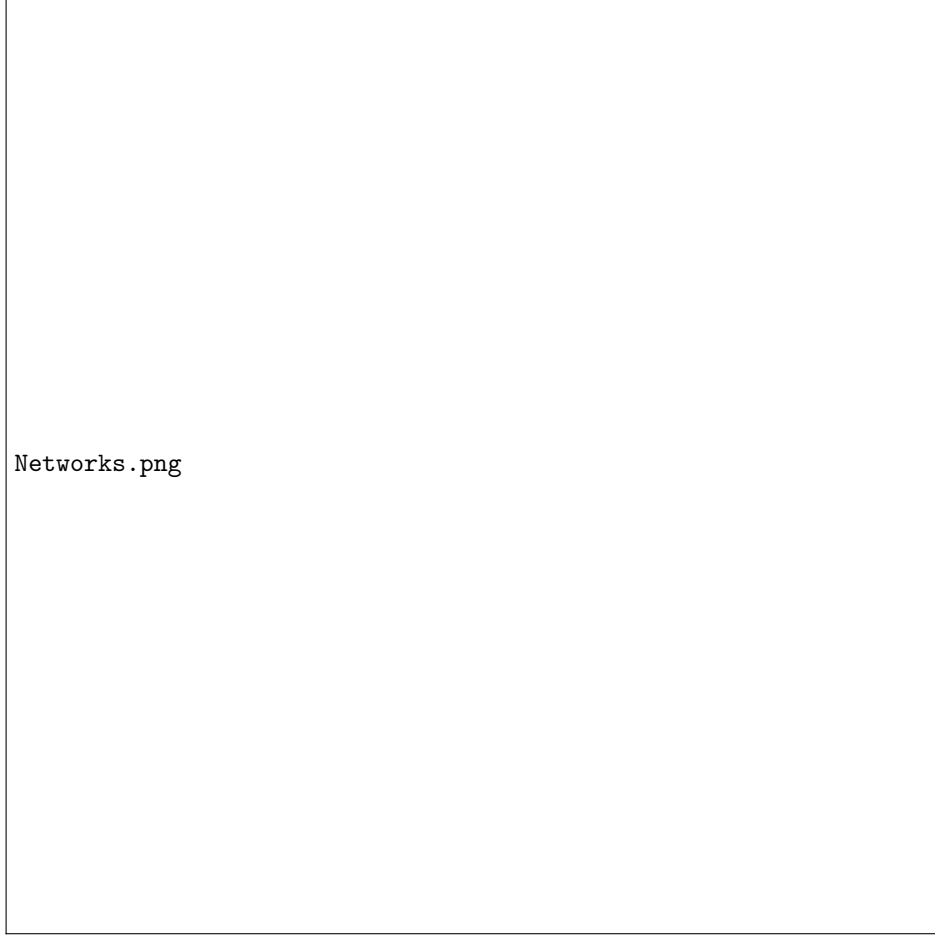


Figure 2: The connectivity networks from individual 1 from the (Left) healthy control group and the (Right) patient group. Each of these networks were drawn using a force-directed layout and nodes are colored according to functional subnetwork location.

- **Full covariate model (Full):** We included the spatial distance, subnetwork location, and hemisphere location in the model. Random effects were included to model degree heterogeneity. Furthermore, we included clusters and set $d = 2$ and $K = 3$ for clustering in each network.

In the present study, we set $M = L$ and thus consider population hierarchy at the individual level. Future work will further explore other specifications of the model on population studies like the one considered here.

5 Results

5.1 Exploratory Analysis

Before fitting the latent space model, we first analyzed the patterns of connectivity for individuals in the study. Figure 2 provides a force-directed layout of a functional connectivity network from each group. Although it provides just one example from the two groups, Figure 2 seems to suggest that the structures of these networks are similar. To further test this, we further calculated three summaries of these networks that may reveal differences. The first is network density, which measures the average strength of connections between regions in the network. The second summary is the tendency for two nodes connected to a third to be connected to one another. This is called “triadic closure” or “clustering,” and is found in many networks across disciplines and application areas. We next calculated the modularity of each network, which

quantifies the extent to which the regions cluster into tightly connected communities. The modularity score was calculated after the fast and greedy community detection algorithm in the `igraph` package in R was run to identify an approximate partitioning of the network (Clauset et al., 2004). We report the edge density, clustering coefficient, and modularity of the best partition identified by the fast and greedy community detection method in Figure 3.

Figure 3 provides further support that the structure of the connectivity networks is similar across groups. Indeed, independent Bayesian t-test comparisons between the groups were not statistically significant (Bayesian p-values greater than 0.585) for each summary statistic. Notably, Figure 3 reveals two other properties of these connectivity networks. First, the networks exhibit moderate clustering – the global clustering coefficient is about 0.50 for each network. Finally, a vast majority of the networks have a high modularity (> 0.7), suggesting that the connectivity networks in this study exhibit clear community structure.

5.2 Model Results

With the HLSM, we can formally test the effect of features like those discovered from the exploratory analysis on the likelihood of the observed network. Furthermore, we can make inferential statements about the differences between two groups. We apply the three hierarchical latent space models described in Section 5.3. For each model, we estimate the observed feature effects, as well as the latent variable means and variances when they were included in the model. Below, we investigate the goodness of fit of each model, the estimated covariate effects, and the degree heterogeneity for each group. We discuss these in turn and highlight the similarities and differences between groups.

5.2.1 Goodness of Fit

We begin by assessing the goodness of fit of each fitted model to compare models for each individual. We calculated the Bayesian information criterion (BIC) of each fitted model on the data. For the HLSM, the overall BIC provides a summary of both the likelihood misspecification on the observed covariates, as well as unexplained variation in the Gaussian models for random and clustering effects (see Krivitsky et al., 2009). In general, lower values of BIC indicate better fit of the model. To get an idea of comparing these models, we report the BICs for each model across the first 10 individuals in each group in Table 4.

Healthy Controls				Patients			
Individual	Cluster	Spatial	Full	Individual	Cluster	Spatial	Full
1	3351.4	3622.8	2970.4	1	4798.1	5878.1	5331.1
2	5649.7	6745.4	5999.2	2	3403.1	3553.4	3068.8
3	3329.6	3722.1	2844.9	3	1753.3	1600.5	1299.9
4	4244.3	4885.1	4284.8	4	3685.8	4970.4	4635.6
5	2937.5	3344.5	2969.3	5	4021.9	4845.8	4366.2
6	2424.3	2321.2	1850.6	6	2661.6	2797.7	2231.3
7	1875.6	1765.6	1377.6	7	1841.9	1775.0	1470.5
8	8273.9	11597.9	10540.5	8	3063.3	3724.7	3235.1
9	6861.2	9234.7	8228.8	9	2075.5	1977.3	1560.8
10	3236.7	3705.4	3163.6	10	2641.6	2982.3	2352.1

Table 4: The Bayesian information criterion (BIC) for each overall model across the first 10 individuals in each group. Lower values of BIC suggest better fit. Bolded values indicate which model had the lowest BIC for that individual.

Table 4 reveals that across all individuals either the cluster model or the full covariate model had the lowest BIC. Indeed, this trend extends to the entire population of each group as well. Over the entire population of healthy patients, the cluster model outperformed competing models in 35 of 71 individuals, and the full covariate model was best in the remaining 36 healthy individuals. In the population of 24 paranoid schizophrenia patients, the cluster model fit the network best in 12 of 24 individuals, while the full covariate model was best for the other 12. Furthermore, for every individual the spatial model always performed the worst among competing models.

To further understand the differences between the cluster and full models, we visualize two connectivity networks - one where the cluster model performed best and the other where the full model performed best - in Figure 4. The top left plot of Figure 4 shows the identified cluster labels of the regions when the cluster model outperformed the full model. The top right plot is drawn with the same spatial layout but with vertex colors representing subnetwork label. The bottom row of Figure 4 shows a side-by-side comparison of these vertex labels in an example where the full model outperformed the cluster model.

In comparing the top and bottom rows of the figure, we find that the 3-cluster model performs best when the identified clusters do not closely match the subnetwork location of the regions. That is to say, the cluster model performs better when the subnetwork location and cluster labels do not provide confounding information. We further support this finding by measuring the match between cluster label and subnetwork location for each network using the rand index [Rand, 1971]. The rand index is a value between 0 and 1 and provides a measure of similarity between two partitions. Large rand index values represent strong similarity. Though not shown here, we investigated the rand index against the binary variable of whether the full model was best and found a strong inverse relationship between the two variables (Spearman rank correlation = 0.72) as we expected.

Together, these results have two primary implications. First, models including only spatial distance do not capture enough variability in the network and should be substituted for latent models with clustering or models with more rich spatial information like the subnetwork and hemisphere location. Second, the cluster model performs better than the more complete covariate model when the clusters do not closely match the covariates included in the model.

5.2.2 Covariate Effects

We next investigate the effects of the observable covariates considered in this study. We plot the posterior mean estimators of each these coefficients in Figure 5. Figure 5 reveals several interesting characteristics of generative mechanisms describing the functional connectivity networks in both groups. First, the effects of the observable covariates considered in the study are similar across groups and models. This suggests that the effect of spatial location and homophily play an important and related role in the structure of the connectivity networks in both healthy and patient groups. Next, it is notable that in the spatial model, when the spatial distance is the only observable feature, the effect of spatial distance on the connectivity is significant and negative across 92 of the 95 individuals in the study. This agrees with previous empirical findings like those in [Honey et al., 2009]. When the hemisphere location and subnetwork location are controlled for in the model, as is done with the full model, the effect of spatial distance is no longer significant at a 0.95 level. This perhaps expected finding suggests that the hemisphere and subnetwork location confounds with, and is more important than, the spatial distance between regions.

Finally, Figure 5 reveals that the hemisphere location and subnetwork location of two regions significantly influence their propensity to connect. The effect of subnetwork location is particularly evident from this study, and was significant across all 95 individuals in the full model. The average effect size of subnetwork location was approximately 3.02, suggesting that the propensity two regions to connect statistically increases when the two regions are within the same functional subnetwork. The hemisphere location was significant in 76 of the 95 participants and had an average effect size of 0.35. Thus we find that when both hemisphere and subnetwork location are considered as covariates, subnetwork location tends to have a stronger and more consistent effect on the likelihood of two regions connecting.

5.2.3 Heterogeneity in Functional Connectivity

In the final part of our analysis, we investigate the heterogeneity in functional connectivity of the regions across groups. In our model, this heterogeneity is quantified by the random effect term for each node, δ_i . We plot the distributions of the estimated connectivity values of the two groups in Figure 6. From Figure 6 we see that the degree heterogeneity distributions across groups are different, and that the patient group exhibits less variability in connectivity. Indeed, for both the spatial and full model, we find that the group distributions are statistically different with a Kolmogorov Smirnov comparison test (p-value < 0.001 in both cases). This finding complements the recent work of [Alnæs et al., 2019], who found that schizophrenia is associated with anatomical heterogeneity in the brain, including heterogeneity in frontotemporal thickness and area and cortical, ventricle, and hippocampal volumes. Our analysis suggests that patients with paranoid

schizophrenia tend to have less variable connectivity patterns. That is, the latent degree of each region is more dispersed in healthy controls than in patients with paranoid schizophrenia. This finding might reflect the fact that we are comparing a specific subtype of schizophrenia, the paranoid type, with a broad classification of healthy individuals. An interesting area of future work might include heterogeneity of functional connectivity as a variable in the classification and identification of schizophrenia patients.

6 Discussion

In this paper, we introduced the hierarchical latent space network model (HLSM) as a means to characterize functional connectivity of the whole brain in a population. The HLSM is a probability distribution on the network describing functional connectivity within a population that characterizes both individual-level and group-level hierarchy. We investigated the utility of the model through an application to group comparisons of healthy individuals and patients with paranoid schizophrenia. With the HLSM, we were able to formally test the effects of network motifs on functional connectivity, and we found that across all individuals the subnetwork and hemisphere location of two regions significantly increased the probability of their connection. We also tested the effect of clustering within each network and identified when clustering variants of the HLSM outperform other covariate-based models. Finally, we identified intrinsic differences between the heterogeneity of functional connectivity between healthy and patient groups, and that in our study patients tended to exhibit lower variability than their healthy counterparts.

We emphasize the point that the findings in our study would not have been possible through typical separate analyses of activation motifs. Rather, joint probability models like the HLSM are required to fully characterize the marginal effects and interactions of activation motifs. Future research will investigate the development and implementation of other hierarchical models for functional connectivity studies. Beyond the brain, hierarchical structure frequently arises in systems of social interaction and movement. For example in US county-to-county migration data, it is often the case that migration occurs in some systematic structured manner among counties within the same state, but that the movement between states is much less likely. Social network analysis has a rich literature from which models can be adapted to address challenges in functional connectivity. A starting point includes the adaptation of more complex network models for weighted networks like the generalized exponential random graph model [Wilson et al., 2017a, Denny et al., 2017]. With generative models like the HLSM, we aim to fully understand the generative mechanisms describing commonly observed structures in the brain like community structure and the small-world property. We hope that the HLSM provides a starting-point for these endeavors.

7 Acknowledgements

JDW was supported by NSF grant DMS-1830547. SJC was supported by NSF grants SES-1357622, SES-1461493, and SES-1514750, NIH R-34 DA043079, and the Alexander von Humboldt Foundation’s Fellowship for Experienced Researchers. ZLL was supported by NSF- SMA 1533500. The authors have intellectual property considerations for using cGERGM in a medical context. SJC, JDW, and ZLL are co-founders and equity holders in Cerenetics, Inc., a start-up company commercializing the technology described in this paper.

References

- Dag Alnæs, Tobias Kaufmann, Dennis van der Meer, Aldo Córdova-Palomera, Jaroslav Rokicki, Torgeir Moberget, Francesco Bettella, Ingrid Agartz, Deanna M Barch, Alessandro Bertolino, et al. Brain heterogeneity in schizophrenia and its association with polygenic risk. *JAMA psychiatry*, 2019.
- G Ashby. *Statistical Analysis of fMRI Data*. MIT Press, 2011.
- Jeffrey D Banfield and Adrian E Raftery. Model-based gaussian and non-gaussian clustering. *Biometrics*, pages 803–821, 1993.
- Danielle S Bassett and Olaf Sporns. Network neuroscience. *Nature neuroscience*, 20(3):353–364, 2017.

- Danielle S Bassett, Nicholas F Wymbs, Mason A Porter, Peter J Mucha, Jean M Carlson, and Scott T Grafton. Dynamic reconfiguration of human brain networks during learning. *Proceedings of the National Academy of Sciences*, 108(18):7641–7646, 2011.
- Danielle S Bassett, Muzhi Yang, Nicholas F Wymbs, and Scott T Grafton. Learning-induced autonomy of sensorimotor systems. *Nature neuroscience*, 18(5):744–751, 2015.
- Richard F Betzel and Danielle S Bassett. Multi-scale brain networks. *NeuroImage*, 2016.
- Richard F Betzel and Danielle S Bassett. Generative models for network neuroscience: Prospects and promise. *Journal of The Royal Society Interface*, 14(136):20170623, 2017.
- Richard F Betzel, Andrea Avena-Koenigsberger, Joaquín Goñi, Ye He, Marcel A De Reus, Alessandra Griffa, Petra E Vértes, Bratislav Mišić, Jean-Philippe Thiran, Patric Hagmann, et al. Generative models of the human connectome. *Neuroimage*, 124:1054–1064, 2016.
- F DuBois Bowman, Brian Caffo, Susan Spear Bassett, and Clinton Kilts. A bayesian hierarchical framework for spatial modeling of fmri data. *NeuroImage*, 39(1):146–156, 2008.
- Steven L Bressler and Vinod Menon. Large-scale brain networks in cognition: emerging methods and principles. *Trends in cognitive sciences*, 14(6):277–290, 2010.
- Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009.
- Cameron S. Carter, Mark Mintun, Thomas Nichols, and Jonathan D. Cohen. Anterior cingulate gyrus dysfunction and selective attention deficits in schizophrenia: H2o pet study during single-trial stroop task performance. *American Journal of Psychiatry*, 154(12):1670–1675, 1997.
- Aaron Clauset, Mark EJ Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical review E*, 70(6):066111, 2004.
- José Alexandre de Souza Crippa, Antonio Waldo Zuairi, Geraldo F Busatto, Rafael Faria Sanches, Antonio Carlos Santos, David Araújo, Edson Amaro, Jaime Eduardo Cecílio Hallak, Virginia Ng, and Philip K McGuire. Cavum septum pellucidum and adhesio interthalamica in schizophrenia: an MRI study. *European psychiatry : the journal of the Association of European Psychiatrists*, 21(5):291–9, 2006.
- Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.
- MJ Denny, JD Wilson, SJ Cranmer, BA Desmarais, and S Bhamidi. Gergm: Estimation and fit diagnostics for generalized exponential random graph models. *R package version 0.11*, 2, 2017.
- Daniele Durante, David B Dunson, and Joshua T Vogelstein. Nonparametric bayes modeling of populations of networks. *Journal of the American Statistical Association*, 112(520):1516–1530, 2017.
- MB First, JBW Williams, RS Karg, and RL Spitzer. *STRUCTURED CLINICAL INTERVIEW FOR DSM-5*. American Psychiatric Association, Arlington, VA, 2015.
- Isobel Claire Gormley and Thomas Brendan Murphy. A latent space model for rank data. In Edoardo Airoldi, David M. Blei, Stephen E. Fienberg, Anna Goldenberg, Eric P. Xing, and Alice X. Zheng, editors, *Statistical Network Analysis: Models, Issues, and New Directions*, pages 90–102, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.
- Mark S Handcock, Adrian E Raftery, and Jeremy M Tantrum. Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2):301–354, 2007.
- Stephan Heckers. Neuroimaging studies of the hippocampus in schizophrenia. *Hippocampus*, 11(5):520–528, oct 2001. ISSN 1050-9631.

- Peter D. Hoff, Adrian E. Raftery, and Mark S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002.
- CJ Honey, O Sporns, Leila Cammoun, Xavier Gigandet, Jean-Philippe Thiran, Reto Meuli, and Patric Hagmann. Predicting human resting-state functional connectivity from structural connectivity. *Proceedings of the National Academy of Sciences*, 106(6):2035–2040, 2009.
- Mark Jenkinson and Stephen Smith. A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2):143–156, 2001.
- Mark Jenkinson, Peter Bannister, Michael Brady, and Stephen Smith. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, 17(2):825–841, 2002.
- Brian Karrer and Mark EJ Newman. Stochastic blockmodels and community structure in networks. *Physical review E*, 83(1):016107, 2011.
- Pavel N Krivitsky and Mark S Handcock. Fitting position latent cluster models for social networks with latentnet. *Journal of Statistical Software*, 24, 2008.
- Pavel N Krivitsky, Mark S Handcock, Adrian E Raftery, and Peter D Hoff. Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social networks*, 31(3):204–213, 2009.
- John Kruschke. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press, 2014.
- Steven J. Luck and JM Gold. The construct of attention in schizophrenia. *Biological psychiatry*, 64(1):34–39, 2008.
- John D Medaglia, Mary-Ellen Lynall, and Danielle S Bassett. Cognitive network neuroscience. *Journal of cognitive neuroscience*, 2015.
- David Meunier, Renaud Lambiotte, Alex Fornito, Karen D Ersche, and Edward T Bullmore. Hierarchical modularity in human brain functional networks. *Hierarchy and dynamics in neural networks*, 1:2, 2010.
- M Fiona Molloy, Giwon Bahg, Xiangrui Li, Mark Steyvers, Zhong-Lin Lu, and Brandon M Turner. Hierarchical bayesian analyses for modeling bold time series data. *Computational Brain & Behavior*, 1(2):184–213, 2018.
- M Fiona Molloy, Giwon Bahg, Zhong-Lin Lu, and Brandon M Turner. Individual differences in the neural dynamics of response inhibition. *Journal of Cognitive Neuroscience*, pages 1–21, 2019.
- Peter J Mucha, Thomas Richardson, Kevin Macon, Mason A Porter, and Jukka-Pekka Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980):876–878, 2010.
- Soumendu Sundar Mukherjee, Purnamrita Sarkar, and Lizhen Lin. On clustering network-valued data. In *Advances in neural information processing systems*, pages 7071–7081, 2017.
- Deepak N Pandya and Douglas L Rosene. Some observations on trajectories and topography of commissural fibers. In *Epilepsy and the corpus callosum*, pages 21–39. Springer, 1985.
- Subhadeep Paul and Yuguo Chen. A random effects stochastic block model for joint community detection in multiple networks with applications to neuroimaging. *arXiv preprint arXiv:1805.02292*, 2018.
- Dragana M Pavlovic, Bryan LR Guillaume, Emma K Towlson, Nicole MY Kuek, Soroosh Afyouni, Petra E Vertes, Thomas BT Yeo, Edward T Bullmore, and Thomas E Nichols. Multi-subject stochastic blockmodels for adaptive analysis of individual differences in human brain network cluster structure. *bioRxiv*, page 672071, 2019.

- Jonathan D Power, Alexander L Cohen, Steven M Nelson, Gagan S Wig, Kelly Anne Barnes, Jessica A Church, Alecia C Vogel, Timothy O Laumann, Fran M Miezin, Bradley L Schlaggar, et al. Functional network organization of the human brain. *Neuron*, 72(4):665–678, 2011.
- Adrian E. Raftery, Xiaoyue Niu, Peter D. Hoff, and Ka Yee Yeung. Fast inference for the latent space network model using a case-control approximate likelihood. *Journal of Computational and Graphical Statistics*, 21(4):901–919, 2012.
- William M Rand. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850, 1971.
- Mikhail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.
- Purnamrita Sarkar and Andrew W. Moore. Dynamic social network analysis using latent space models. *ACM SIGKDD Explorations*, 7(2):31–40, 2005.
- Daniel K. Sewell and Yuguo Chen. Latent space models for dynamic networks. *Journal for the American Statistical Association*, 110(512):1646–1657, 2015.
- Daniel K. Sewell and Yuguo Chen. Latent space models for dynamic networks with weighted edges. *Social Networks*, 44:105–116, 2016.
- Susan Shortreed, Mark S. Handcock, and Peter D. Hoff. Positional estimation within a latent space model for networks. *Methodology*, 2(1):24–33, 2006.
- Sean L Simpson and Paul J Laurienti. A two-part mixed-effects modeling framework for analyzing whole-brain network data. *NeuroImage*, 113:310–319, 2015.
- Sean L Simpson, Satoru Hayasaka, and Paul J Laurienti. Exponential random graph modeling for complex brain networks. *PLoS One*, 6(5):e20039, 2011.
- Sean L Simpson, Malaak N Moussa, and Paul J Laurienti. An exponential random graph modeling approach to creating group-based representative whole-brain connectivity networks. *Neuroimage*, 60(2):1117–1126, 2012.
- Sean L Simpson, Mohsen Bahrami, and Paul J Laurienti. A mixed-modeling framework for analyzing multitask whole-brain network data. *Network Neuroscience*, 3(2):307–324, 2019.
- Anna L Smith, Dena M Asta, and Catherine A Calder. The geometry of continuous latent space models for network data. *arXiv preprint arXiv:1712.08641*, 2017.
- Stephen M. Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17(3):143–155, nov 2002. ISSN 1065-9471.
- Olaf Sporns. *Networks of the Brain*. MIT press, 2011.
- Paul E Stillman, James D Wilson, Matthew J Denny, Bruce A Desmarais, Shankar Bhamidi, Skyler J Cranmer, and Zhong-Lin Lu. Statistical modeling of the default mode brain network reveals a segregated highway structure. *Scientific Reports*, 7(1):11694, 2017.
- Paul E Stillman, James D Wilson, Matthew J Denny, Bruce A Desmarais, Skyler J Cranmer, and Zhong-Lin Lu. A consistent organizational structure across multiple functional subnetworks of the human brain. *NeuroImage*, 197:24–36, 2019.
- Kaustubh Supekar, Weidong Cai, Rajeev Krishnadas, Lena Palaniyappan, and Vinod Menon. Dysregulated Brain Dynamics in a Triple-Network Saliency Model of Schizophrenia and Its Relation to Psychosis. *Biological Psychiatry*, 85(1):60–69, 2019. ISSN 18732402.
- Tracy M Sweet, Andrew C Thomas, and Brian W Junker. Hierarchical network models for education research: Hierarchical latent space models. *Journal of Educational and Behavioral Statistics*, 38(3):295–318, 2013.

- James D Wilson, Simi Wang, Peter J Mucha, Shankar Bhamidi, and Andrew B Nobel. A testing based extraction algorithm for identifying significant communities in networks. *The Annals of Applied Statistics*, 8(3):1853–1891, 2014.
- James D Wilson, Matthew J Denny, Shankar Bhamidi, Skyler J Cranmer, and Bruce A Desmarais. Stochastic weighted graphs: Flexible model specification and simulation. *Social Networks*, 49:37–47, 2017a.
- James D Wilson, John Palowitch, Shankar Bhamidi, and Andrew B Nobel. Community extraction in multilayer networks with heterogeneous community structure. *The Journal of Machine Learning Research*, 18(1):5458–5506, 2017b.
- Mark W Woolrich. Bayesian inference in fmri. *NeuroImage*, 62(2):801–810, 2012.
- Til Wykes, Clare Reeder, and Julia Corner. The prevalence and stability of an executive processing deficit, response inhibition, in people with chronic schizophrenia. *Schizophrenia Research*, 46(2-3):241–253, dec 2000. ISSN 0920-9964.
- BT Thomas Yeo, Fenna M Krienen, Jorge Sepulcre, Mert R Sabuncu, Danial Lashkari, Marisa Hollinshead, Joshua L Roffman, Jordan W Smoller, Lilla Zöllei, Jonathan R Polimeni, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, 106(3):1125–1165, 2011.

Appendix: The Effect of Thresholding on Coefficient Estimators

In the application study presented in Section 5, we tested the HLSM on unweighted networks that were thresholded so that binary edge values represented correlation values above 0.50. Here, we investigate the effect of the choice of threshold on the coefficient estimators that we obtain from the HLSM. To test this, we refit the **Full Model** unweighted HLSM on both groups, the 71 healthy individuals, and the 24 patients with paranoid schizophrenia across threshold values 0.3, 0.35, $\dots$, 0.70. For each threshold value, we fit the full model to each individual and report the coefficient estimators for the subnetwork effect, the hemisphere effect, and the spatial distance effect. These results are shown in Figure 7.

Figure 7 suggests two important patterns. First, across threshold values the coefficient estimators keep the same sign and statistical significance. This reveals that the interpretation discussed in Section 5 is robust across values of the threshold. Secondly, there are noticeable trends in the coefficient estimators across threshold values. For example, the subnetwork effect tends to increase as the threshold increases, suggesting that as the network becomes more sparse (and contains fewer edges), the subnetwork community structure of the network has a stronger effect on connection probability within the population. It also appears that as the threshold increases, the variability in the estimators for the hemisphere and spatial distance effects increase. This trend intuitively suggests that estimator variability is inversely related to the number of edges in the network.

This investigation supports the analysis and discussion of the application in Section 5 and provides some intuition as to the effect of thresholding on fitting the unweighted HLSM. In the future, a comparison between these unweighted models and the weighted HLSM for correlation networks should be studied. We plan to pursue this in future work.



Figure 3: Summary statistics of the connectivity networks considered in this study. For each network, the density, clustering coefficient, and modularity of the best partition identified by the fast and greedy community detection algorithm.

Colored by Cluster

Colored by Subnetwork

Figure 4: Clustering layouts of networks, where clusters were identified by the cluster model. The top row gives an example where the clustering model had a lower BIC than the full model. The bottom row gives an example where the clustering model had a higher BIC than the full model.

Spatial Model Results

spatial_coefficients.png

Full Model Results

full_model_coefficients.png

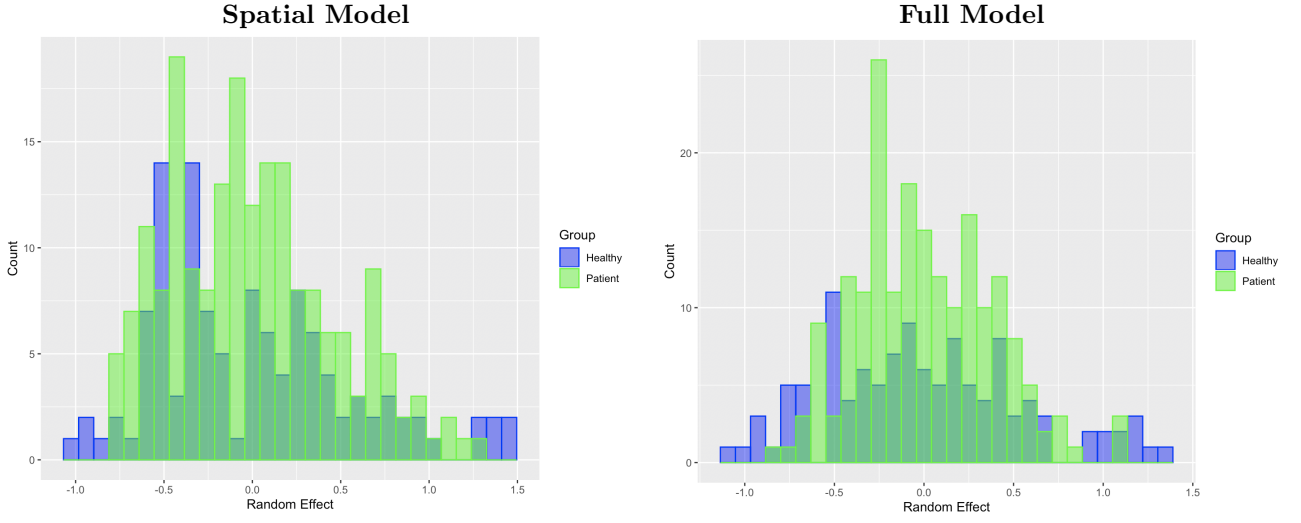


Figure 6: The distribution of estimated regional random effects across each group from the full and spatial distance models. Random effects quantify the latent degree heterogeneity of the regions in each network. In both models, healthy individuals have a significantly more variable latent heterogeneity than patients with paranoid schizophrenia (p-value < 0.001 in both cases).

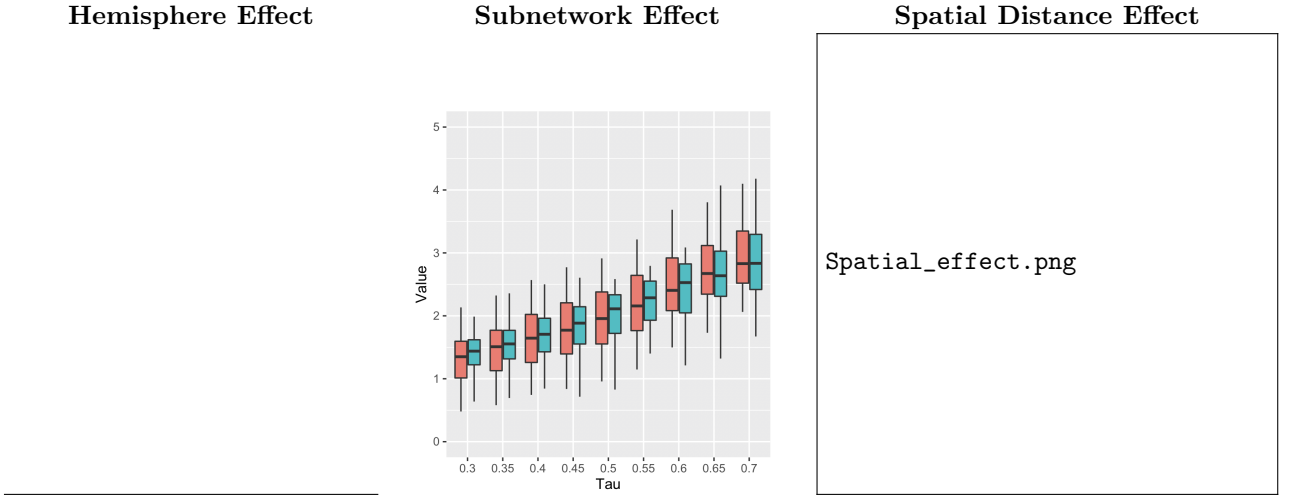


Figure 7: Coefficient estimator distributions for the schizophrenia and healthy controls group across threshold values. For each threshold value τ , patient correlation networks were first binarized by setting all correlations above τ to 1 and keeping the remaining values 0. The full model HLSM was subsequently fit to each network. The distribution of coefficient estimators across patients is shown for each threshold value.